Mathematical modelling and analysis of communication networks:

Transient characteristics of traffic processes and models for end-to-end delay and delay-jitter

by

Olav Østerbø

Telenor Research and Development

Submitted to NTNU-Department of Telematics

in partial fulfilment of the requirements

for the degree of Dr. Philos.

To Silje, Sunniva and
Håkon Olav

Abstract

The first part of the thesis (Part I) is devoted to find methods to describe transient behaviour of traffic processes, where the main emphasis is put on the description and analysis of excess periods and excess volumes of quite general stochastic processes. By assuming that traffic changes on different time scales, the transient characteristics such as excess periods could be important measures to describe periods of congestion on a communication link and moreover, the corresponding excess volume will represent lost information during such periods. Although the results obtained are of rather general nature, they provide some rather fundamental insight into transient characteristics of traffic processes. The distributions of the length of excess periods may then be expressed it terms of some excess probabilities that are related to the minimum of the process in the time interval considered. Similar relations for the excess volumes are harder to obtain and require the joint probability of the arrived volume and the minimum of the process in the same time interval.

We put particular emphasis on Gaussian traffic models and we propose an approximative method to get the distributions of excess times and excess volumes. The main idea is to approximate the excess probabilities by multinormal integrals. We also consider the Ornstein-Uhlenbeck (O-U) process mainly because the O-U process is a special case of a Gaussian process and could therefore be used as a test case for the proposed approximations, but also because the O-U process may be obtained as a limit of a large numbers of ON/OFF sources (with exponentially distributed ON- and OFF-times). For the O-U process we have given the Laplace transforms for the first passage times and the corresponding volumes. These Laplace transforms are inverted by the locating the residues yielding infinite series. Asymptotic expansions for small arguments are also found.

We have also considered excess times and excess volumes for semi-Markov processes. The main results obtained are general expressions for the Laplace transforms and distribution functions of the excess times and the excess volumes in terms of the generator matrices. For birth-death semi-Markov processes the generator matrices simplifiy and the transforms may be found recursively.

The second part (Part-II) deals with models to obtain end-to-end queueing delay for networks deploying statistical multiplexing. The first model is based on the assumption (approximation) that the end-to-end delay may be found by convolution, where the key assumption is that the parts of the end-to-end delay stemming from the different nodes are independent stochastic variables. As model for each node we take the ordinary M/G/1 queue.

If in addition the nodes are identical i.e. the convolution consists of the waiting times of a fixed numbers of identical M/G/1 queues, the evaluation may be substantially simplified. It turns out the convolutions may be found by taking some partial derivatives with respect to the load parameter. The same technique may be generalised in various directions for instance it is possible to extend the result to the case with two groups of queues where the queues in each group are identical.

For M/D/1 queues with identical service times we find explicit closed form results for the convolutions. We also generalize this result to consider two groups of M/D/1 queues having different service times, and this is a particularly interesting case since it may be used as model for end-to-end delay also including access links with low capacity. Similar results are also found for end-to-end queueing models with priority.

A different approach is obtained by assuming a slotted model. The main idea is to capture the disturbance of a packet stream as it passes through a series of multiplexers. Even though the output process from a multiplexer is non-renewal, we get the distribution between two consecutive departures, and approximate the process with a renewal stream. This stream is then feed into the next multiplexer (together with other crossing traffic). In this way we obtain recursive relations for the jitter and the end-to-end delay.

Preface

The writing of this thesis, which is the result of several years of research, has finally come to its end. The work has been carried out during the years 1998-2003 partly as ordinary research projects, but also a lot of spare time, late evenings and weekends have been spent at office and at home to finalise the thesis.

During the year 1998-1999 I was privileged to work full time doing basic research, and I would like to thank may employer Telenor R&D giving me the opportunity to start the research activity. Without that year of research there would definitive not been any thesis. I would like to thank Terje Ormhaug and Harald Pettersen for their support when I applied for sabbatical year and Mette Røhne for encouraging me to do so, at that time when she was finalising her own doctoral thesis and I was her assistant supervisor.

There are several other persons to whom I am grateful. Among them I would like to mention my superior Nils Flaarønning for encouraging me to finalise the dissertation and not imposing me with any “extra” tasks during the last year. Inge Svinnset, my project manager and (train travelling companion), for his tolerance and for making room for plenty of my research in his projects. Ralph Lorentzen, for helping me with the proof of theorem A.1. Thor Gunnar Eskedal for having time for talks both professional and personal. Last but not least, Terje Jensen for always being interested in discussing technical questions, giving constructive criticism to various proposals, and for reading the manuscript.

With NTNU Department of Telematics I have benefit from good co-operation for a number of years, and I would particular thank Prof. Peder J. Emstad and Prof. Bjarne E. Helvik for several stimulating discussions.

I would also like to thank my family for their support and patience during these years, and I would dedicate the thesis to my children Silje, Sunniva, and Håkon Olav.

Olav Østerbø

Strømmen ,18th August. 2003

TABLE OF CONTENTS

Acronyms

ADSL	Asynchronous Digital Subscriber Line
ATM	Asynchronous Transfer Mode
a.s.	almost sure
BS	Background stream
CAC	Call Acceptance Control
CDF	Complementary Distribution Function
DF	Distribution Function
DiffServ	Differentiated Services
DSL	Digital Subscriber Line
F	First
FCFS	First Come First Serve
FIFO	First In First Out
FS	Foreground stream
HOL	Head Of Line
IntServ	Integrated Services
IP	Internet Protocol
L	Last
LD	Large Deviation
LST	Laplace-Stieltjes Transform
MPLS	Multi Protocol Label Switching
O-U	Orstein-Uhlenbeck
NA	Normal Approximation
PDF	Probability Density Function
PHB	Per Hop Behaviour
QoS	Quality of Service
R	Random
RT	Real Time
SLA	Service Level Agreement
TCP	Transmission Control Protocol
STM	Synchronous Transport Module
UAA	Uniform Asymptotic Approximation
VDSL	Very high speed Digital Subscriber Line

1

Overall introduction to the thesis

1.1 Background and motivation

There is an increasing need to address and understand some fundamental teletraffic issues in today's and forthcoming communication network, where the current trend is a change-over towards heterogeneous network types where the end-to-end communication may involve more than one operator and the QoS (Quality of Service) provisioning is not currently satisfactorily solved. New switching techniques emerge, capacity increases, and new services are introduced, causing a steady growth in the traffic. The competition in the telecom market is hard, the revenue is squeezed, so the slogan "throwing bandwidth at the problems", will not be a winning strategy for the operators. On the other hand too scarce network resources could result in degradation in the quality of the services offered, leading to discontented customers with the subsequent consequences that may cause. So there is a strong need for network optimization and performance modelling.

By the emerging of IP (Internet Protocol) multiservice networks a lot of new teletraffic challenges emerge. Among those we would particularly mention:

- traffic models for different services or flows and also models for aggregation of flows,
- differentiation between classes of services, i.e. scheduling and buffer management,
- traffic control, i.e. SLA (Service Level Agreement), CAC (Call Acceptance Control) and policing,
- QoS provisioning end-to-end,
- dimensioning models.

Thus, the need for performance and dimensioning modelling of today's communication networks is sustained and is definitive not of less importance than for former types of networks.

1.2 Main achievements

In this thesis we have focused on a few of the issues mentioned above; namely models to describe transient behaviour of rate processes and models to get the end-to-end delay in packet networks. As such, we are confident that these are important models; providing insight into important aspects of networking, as congestion periods and information loss and

the delay and delay variation for a packet flow. The main objective have been to obtain analytical results based on mathematical modelling, but where we always take the applied viewpoint, where the goals have been to provide models that are numerical feasible and provide numerical examples that are interesting from the perspective of network performance. In the analysis we have derived results on the basis of several applied mathematical fields such as: probability theory and queueing analysis, asymptotic expansions of integrals, differential equations and a few results from real analysis. We may summarise the main achievements obtained as:

- Give some general results on level crossing and excess distribution for stationary rate processes.
- Suggest approximations of the excess times and excess volumes distributions for general stationary Gaussian process.
- Give expansions and asymptotic formulae for the PDF (Probability Density Function) and CDF (Complementary Distribution Function) of first passage times and corresponding volumes for the Orstein-Uhlenbeck (O-U) process.
- Give some general results for level crossing and excess distributions for semi-Markov processes where we for more specific models as birth-death process obtain the LST (Laplace-Stieltjes Transforms) recursively due to the special structure of corresponding generator matrices.
- Give an effective method to obtain the PDF and DF (Distribution Function) of the convolution of a given number of waiting times of identical M/G/1 queues.
- Extend the result on convolutions to cover cases where not all the service times may be identically distributed, and also to cover HOL (Head Of Line) priority queueing.
- Provide a slotted queueing model and using generating function techniques to obtain the output distribution of particular packet stream and apply this model recursively to obtain both end-to-end delay and the evolution of the jitter for a deterministic packet stream through a series of nodes.

All the numerical examples are obtained by Mathematica programming except for the examples in chapter 5 which is obtained by FORTRAN routines. Many of the numerical results are checked against similar (but different) models and also against asymptotic expansions.

1.3 Overall organization of the thesis

This thesis is divided into two parts, Part I and Part II, but where each chapter more or less is self-contained with an introductory chapter. A quite large number of more technical details are put in separate appendices. The use of symbols throughout the thesis is not stringent, for instance is the symbol B used to denote a rate process in Part I while in Part II the same symbol is used as a symbol of the number of packet arrivals from a background stream during a slot. However, within each chapter the notation should be “consistent”.

Part I is devoted to find methods to describe transient behaviour of traffic processes, where the main emphasis is placed on the description and analysis of excess periods and excess

volumes of quite general stochastic processes, and the organisation is described in section 2.3.

Part II provides models to obtain traffic dependent end-to-end queueing delay and give methods to calculate evolution of the jitter through out a network. The organisation of Part II is given in section 6.4.

PART-I

Some results on level crossing, excess distributions and first passage times for stationary rate processes

2

Introduction

Traffic models are needed as an input for dimensioning and performance evaluation of telecommunication systems. They will also be important in the process of designing and structuring networks. In this context we consider “fresh” traffic, i.e. before it has entered some network elements where it could be disturbed by other traffic streams. By “fresh” traffic we here shall mean the raw bit rate generated from various traffic sources. However, we must to some extent also include the effect from different protocol layers adding on overhead and framing the bit stream into packets. We shall also consider traffic models that are a superposed or an “enveloped” traffic stream, formed as a collection of individual sources, where the gross bit rate is taken as the sum the instantaneous bit rates from the sources as if there where no other constraints or limitations, e.g. buffers, capacity etc.

2.1 Traffic modelling and scaling phenomena

The presence of scaling phenomena in some type of data traffic is well documented in the literature. The area of analysis, modelling and characterization of traffic in communication networks has quickly evolving since the seminal paper by Leland et al. [Lela93] in 1993. This paper showed that network traffic in many cases has properties characterized by long-range dependence and variability at a wide range of time scales, and it introduced the notion of self-similarity to communication networks. Later on, these properties have been shown to hold also for a much wider range of experimental environments [Paxs95], [Crov96].

The evidence of traffic being long range dependent has certain implication on the behaviour of the autocorrelation function for large arguments, resulting in a power law behaviour with exponent between zero and unity, where the exponent is expressed in terms of the Hurst-parameter by $2H-2$. We therefore aim to look at models that have long-range dependence. On the other hand the behaviour of the autocorrelation for small arguments will determine the behaviour of the processes at micro-level (at least for processes that have continuous sample paths) indicating that one possibly should use “different models” on the micro level than on the “macro level”.

2.2 Transient characteristic of traffic processes

By assuming that traffic change on different time scales, the transient characteristics of the traffic processes could be an important measure to describe periods of congestion on a com-

munication link (by assuming that the relaxation time for the link buffer is of order less than the typical length of an excess period). Moreover, the corresponding excess volume will represent lost information during such periods. For traffic model acting on a (virtual) link (of fixed capacity), it is of great interest to be able to answer to some of the following questions:

- How often will the excess periods occur?
- What is the distribution of the length of the excess periods?
- What is the distribution of the corresponding excess volumes?

Although the results obtained are of rather general nature, they give some rather fundamental insight into transient characteristic of traffic processes. The aim has been to provide some results concerning the duration of excess periods and the corresponding excess volumes. It turns out that the up- and down-crossing rate is an important measure, and will answer the first question above provided that the limit is finite. The distributions of the length of an excess period may then be expressed it terms of some excess probabilities that is related to the minimum of the process in the time interval considered. Similar relations for the excess volume is harder to obtain and requires the joint probability of the arrived volume and the minimum of the process in the same time interval. Unless for some very special models the exact excess probability is difficult to find expressions for and therefore some approximations will be needed to get results that are numerical feasible.

2.3 The organisation of PART-I of the thesis

PART-I of the thesis is a collection of three chapters. The main focus through these chapters is the use of level crossing to describe transient phenomena as excess times and excess volumes for bit rate processes. As mentioned these periods may represent periods of congestion for bufferless multiplexing.

Chapter 3 deals with fundamental questions concerning level crossings for stationary stochastic processes where we discuss some of the basic properties. It is known that level crossing is a rather tricky matter, and put strong limitations of the class of processes, especially for processes with continuous sample paths. The chapter is logically divided into two parts where we in the first part give some fundamental results and the only assumption is that the process is stationary, while we in the second part consider processes that are continuous in time and space and we describe a method that makes it possible to also include the excess volumes into the analysis.

In the first part we define the crossing rate and deduce that if this rate is finite it is given as negative derivative of the excess probability. Then we discuss the relation between the excess probabilities and the distribution of the excess periods. In the second part we consider processes that are continuous in time and space. The claim of having finite crossing rate for such processes will put rather strong implications on the behaviour of the autocorrelation near the origin. For processes with continuous sample paths we also give relations between some joint excess probabilities and the joint distribution of the first passage times and corresponding volumes. Similar relations are also found for the joint distribution of the excess times and corresponding excess volumes, but they are more tricky to obtain.

In chapter 4 we consider Gaussian traffic models. In the first part we propose an approximative method to obtain the distributions of excess times and excess volumes. The main idea is to approximate the excess probabilities by multinormal integrals. Based on these ideas we express both first passage times and corresponding volumes and the excess times and corresponding excess volumes in terms of multinormal integrals. In a separate appendix (Appendix B) we have given many interesting properties of such type of integrals, among them the result that makes it possible to calculate the multinormal integrals by calculating a multiple integral with only half the dimension.

In the second part we consider the Ornstein-Uhlenbeck (O-U) process. The main motivation was firstly because the O-U process is a special case of Gaussian process and could therefore be used as a test process for the proposed approximations. Secondly, since the O-U process may be obtained as a limit of a large numbers of ON/OFF sources (with exponential distributed ON- and OFF- times) the results is important in its own. For the O-U process we have given the Laplace transforms for the first passage times and the corresponding volumes. These Laplace transforms are inverted by locating the residues yielding infinite series. Asymptotic expansions for small arguments are also found.

In a series of numerical examples we first tested the approximation by applying multinormal integral of dimension five or six with the exact first passage time distributions for the O-U process. Unfortunately the correspondence was not as good as we hoped, however, the proposed approximation seems to yield an upper bound for the distribution functions. In a second series of examples we chose a process with typical long rang dependence. We conclude that the corresponding (approximative) excess time seems to have long tails.

In chapter 5 we consider the case when the bit rate process is a semi-Markov process. This type of process is not limited by the claim on the behaviour of autocorrelation function near the origin as found for processes with continuous sample paths.

The main results obtained are general expressions for the Laplace transforms and distribution functions of the excess times and the excess volumes in terms of the generator matrices. For birth-death semi-Markov processes the generator matrices simplify and the transforms may be found recursively. For ordinary Markov processes the excess distributions may be obtained by finding the eigenvalues to the corresponding rate matrices, and finally for birth-death processes these eigenvalues may be effectively found by applying the method of bisection.

3

Some transient characteristics of traffic analysed by methods of level crossing and excess distributions

3.1 Introduction

If statistical multiplexing is allowed in broadband networks it may happen that the load form the ongoing communications (or connections) may exceed the capacity of a particular link. Due to statistical fluctuations this situations may occur even though the network is well dimensioned. This may lead to periods with excessive information loss and thereby possible degradation of the QoS. The time scale of such variations may typical be that of an ON/OFF activity period of a frame duration for a video source. At this level the discrete nature of the transmission e.g. packets (or cells) are negligible, and we consider a more or less continuous bit stream with different characteristics. Thus rather than considering events where arrivals of packets or cells occur we take the fluid approach where we observe a continuous bit stream representing the traffic under consideration.

In the literature statistical fluctuations and level crossing initially appeared in the field of statistical communication theory and analog signal processing. The first result on level crossing is due to Rice [Rice45], and goes actually back to 1936 where he gave the classic formula on the average rate of level crossing for Gaussian processes. In his context the main focus was the ability to detect levels of a signal that was influenced by random noise. (See also [Rice48].) Along with other authors in the same field there exists a quite large numbers of papers considering level crossings for Gausian processes, however the main focus has been the crossing of the zero level, whereas we are mainly interesting in crossing of levels having small probabilities. In some quite early papers by McFadden [McFa56] and [McFa58] some quite general results are given for axis crossing. Similar results where also the distributions between successive zeros are discussed are found in the papers [Long58] and [Long62] but with the assumption of a Gaussian process. In the book of Leadbetter et al. [Lead83] a quite large number of results on level crossing are given mainly for Gaussina processes, but also some basic results concerning the crossing intensity for general processes that have continuous sample paths are given. Also in several other textbooks as [Lars79b], [Papo65] and [Midd60] the topic of axis crossing for Gaussian processes have been treated.

Our aim in this chapter is not to give a complete mathematical treatment of the different topics, but rather take the perspective of an engineer and give some general results that are important for performance measures characterising rate processes. A more thorough mathematical treatment will require techniques that are beyond the scope of this thesis.

The usual way of characterising dependencies in stochastic processes is to introduce the second order statistics; that is the covariance or autocorrelation function. The level-crossing description introduced below will be more appropriate to get the performance measures needed for instance considering bufferless multiplexing. This fact will be evident when we derive formulas for the second order moments of the excess volumes that may quite easily be obtained by using the level-crossing description. Also the new achievements obtained by using large deviations seem to fit well into this description.

3.2 Some general results concerning the excess times and excess volumes for stationary stochastic processes

By taking a general starting point we let $\{\boldsymbol{B}_t\}$ be a (non-negative) stochastic process representing the instantaneous bit rate (load) on communication link. The main assumption we put on the bit rate process is that it is stationary (in the strict sense), which means that any group $\{\boldsymbol{B}_{t_1}, \ldots, \boldsymbol{B}_{t_n}\}$ has the same distribution as $\{\boldsymbol{B}_{\tau + t_1}, \ldots, \boldsymbol{B}_{\tau + t_n}\}$ for all choices of τ. It follows that the behaviour of the process is independent of the staring point of the observations (which we in most cases choose to be $t = 0$). We shall also limit ourselves to consider only time continuous processes. For discrete time processes a similar development is possible but such models will not be discussed in this thesis.

In the following we let $m = E[\boldsymbol{B}_0]$ denote the *Mean* bit rate and $\sigma^2 = E[\boldsymbol{B}_0^2] - m^2$ be the corresponding *Variance* and we let $\rho(t) = \dfrac{E[\boldsymbol{B}_0 \boldsymbol{B}_t] - m^2}{\sigma^2}$ denote the *Autocorrelation* function of the process. It turns out that the behaviour of the Autocorrelation function near the origin will provide the necessary information to determine whether the up and down crossing intensities are finite, and therefore determine when the description below is fruitful or not. (See for instance the introductory textbooks of stochastic processes [Cinl75] [Cox70] [Fell68a] [Fell68b].)

Assume that we for a given level (link capacity) C may identify up and down crossing instants $\{U_k\}$ and $\{D_k\}$ such that $\boldsymbol{B}_t - C > 0$ in the interval (U_k, D_k) and $\boldsymbol{B}_t - C \leq 0$ in the interval (D_k, U_{k+1}) (see figure 3.1). The possible up and down crossings intervals (U_k, D_k) and (D_k, U_{k+1}) will describe periods of congestion and non-congestion for bufferless multiplexing, or period of buffer filling or buffer emptying in a fluid queue.

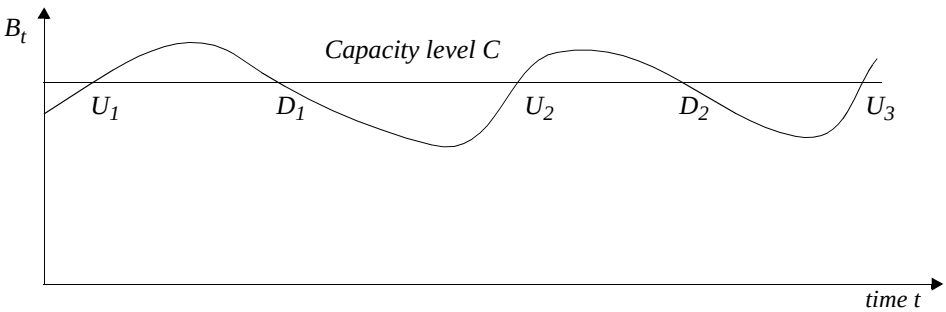

Figure 3.1: Definition of up and down crossing instance.

By assuming that it is possible to identify the random sequences $\{U_k\}$ and $\{D_k\}$ of the up and down crossing instants, we define the excess times and excess volumes:

$$T_k = D_k - U_k \text{ and } A_k = \int_{U_k}^{D_k} (\boldsymbol{B}_\tau - C)d\tau \tag{3.1}$$

and similar periods of normal load and corresponding volumes:

$$S_k = U_{k+1} - D_k \text{ and } V_k = \int_{D_k}^{U_{k+1}} (C - \boldsymbol{B}_\tau)d\tau \tag{3.2}$$

For a loss system A_k will describe the amount of information lost during a congestion period but for the fluid model A_k and V_k describe the net increase and decrease in buffer content during the intervals (U_k, D_k) and (D_k, U_{k+1}). We also define the total volume arriving during two consecutive down crossings as $W_k = \int_{D_k}^{D_{k+1}} \boldsymbol{B}_\tau d\tau$.

The main contribution in this chapter will be to describe a general framework to get the distributions (and moments) of the length of these intervals and the corresponding volumes. If it is possible to obtain these distributions they will give interesting performance measures such as the length of overload periods and the time between them. By considering the excess volumes it is possible to estimate the information loss for bufferless multiplexing and especially the losses in the period of overload.

3.2.1 Some general remarks on level crossing

It turns out that one need to be some careful when considering up and down crossing for stochastic processes. This is specially seen in processes having continuous sample paths where the term up and down crossing can be defined slight different. For instance in the book of M. R. Leadbetter et al. [Lead83] it is defined both the term up crossing and strict up crossing where a so called non strict up crossing is allowing for infinite many up crossings in a small interval. Since we consider processes that have both continuous and piece-wise continuous sample paths we shall take the following definition:

A function $f(t)$ (which we assume to be piece-wise continuous) is said to have an up crossing of the level C at a point $t = t_0$ if for some $\varepsilon > 0$ and every $\eta > 0$, then $f(t) \le C$ for all t in the interval $(t_0 - \varepsilon, t_0)$ and $f(t) > C$ for some t in the interval $(t_0, t_0 + \eta)$. (In a similar way we also define down crossing.)

In the following we shall derive some quite general expressions for the excess distribution on the basis of the basic knowledge of the bit rate process $\{\boldsymbol{B}_t\}$. It turns out that the crossing intensities may be expressed through the functions (excess probabilities):

$$\psi_C(t) = \boldsymbol{P}\{Inf_{\tau \in (0,\, t)} \boldsymbol{B}_\tau > C\} \text{ and} \tag{3.3}$$

$$\phi_C(t) = \boldsymbol{P}\{Sup_{\tau \in (0,\, t)} \boldsymbol{B}_\tau \le C\} \tag{3.4}$$

The functions $\psi_C(t)$ (and $\phi_C(t)$) are the probabilities that the process either is above (or below) the level C (and does not crosses that level) in an interval of length t. It will be convenient to approximate the process $\{\boldsymbol{B}_t\}$ by a sequence $\{\boldsymbol{B}_t^n\}$ taking the value of $\{\boldsymbol{B}_t\}$ at points $t_i^n = \frac{i}{2^n} t$ $(i = 0, 1, \ldots, 2^n)$ and let $\{\boldsymbol{B}_t^n\}$ being linear between such points. (With this type of partition we obtain the $(n+1)$'th from the n'th by halving each interval and therefore doubling the number of points.) We may now approximate $m_t = Inf_{\tau \in (0,\, t)} \boldsymbol{B}_\tau$ by the corresponding 2^n-point approximation $m_t^n = Min_{0 \le i \le 2^n} \boldsymbol{B}_{t_i^n}$ and we define:

$$\psi_C^n(t) = \boldsymbol{P}\{Min_{0 \le i \le 2^n} \boldsymbol{B}_{t_i^n} > C\} = \boldsymbol{P}\{\boldsymbol{B}_{t_0^n} > C, \boldsymbol{B}_{t_1^n} > C, \ldots, \boldsymbol{B}_{t_n^n} > C\} \tag{3.5}$$

Similar we also define the 2^n-point approximation $M_t^n = Max_{0 \le i \le n} \boldsymbol{B}_{t_i^n}$ for the maximum $M_t = Sup_{\tau \in (0,\, t)} \boldsymbol{B}_\tau$ and we define

$$\phi_C^n(t) = \boldsymbol{P}\{Max_{0 \le i \le 2^n} \boldsymbol{B}_{t_i^n} \le C\} = \boldsymbol{P}\{\boldsymbol{B}_{t_0^n} \le C, \boldsymbol{B}_{t_1^n} \le C, \ldots, \boldsymbol{B}_{t_n^n} \le C\} \tag{3.6}$$

By choosing the special form of partitioning the interval we secure that $\{m_t^n\}$ is a decreasing sequence and $\{M_t^n\}$ is an increasing sequence, and it follows that both converge a.s. (almost sure) to the limits $\{m_t\}$ and $\{M_t\}$ respectively, and furthermore the approximative excess probabilities $\psi_C^n(t)$ and $\phi_C^n(t)$ constitute decreasing sequences and will therefore converge (pointivise) to the (desired) excess probabilities:

$$\lim_{n \to \infty} \psi_C^n(t) = \psi_C(t) \text{ and} \tag{3.7}$$

$$\lim_{n \to \infty} \phi_C^n(t) = \phi_C(t) \tag{3.8}$$

We shall define the following up and down crossing rates:

$$\Delta_C^{uc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 \le C < \boldsymbol{B}_t\} \text{ and} \tag{3.9}$$

$$\Delta_C^{dc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\} \tag{3.10}$$

These rates are simply the probabilities that there has been at least one up or down crossing in an interval of length t divided by the length of the interval. For a stationary process the up and down crossing rate must be equal: (The proof of this statement is given in Appendix A by theorem A.1)

$$\Delta_C(t) = \Delta_C^{uc}(t) = \Delta_C^{dc}(t) \tag{3.11}$$

where we define $\Delta_C(t)$ as the crossing rate (either up or down crossing) intensity for an interval of length t. It turns out that the crossing rate has some nice properties that makes the limit $t \to 0$ easier to examine. In Appendix A by theorem A.1 it is also shown that the crossing rate satisfies the nice inequality:

$$\Delta_C(t) \le \gamma\Delta_C(\gamma t) + (1-\gamma)\Delta_C((1-\gamma)t) \text{ for all } t \text{ and } 0 \le \gamma \le 1 \tag{3.12}$$

It follows that $\Delta_C(t) \le \Delta_C\left(\frac{t}{2}\right)$ (by choosing $\gamma = \frac{1}{2}$) for all t and by continuously sub-dividing the interval we get an increasing sequence $\Delta_C(t) \le \Delta_C\left(\frac{t}{2}\right) \le \Delta_C\left(\frac{t}{2^2}\right) \le \ldots$. It follows that $\lim_{n \to \infty} \Delta_C\left(\frac{t}{2^n}\right)$ exists for every t and it is shown that if $\lim_{t \to 0} \sup \Delta_C(t) < \infty$ then the instantaneous crossing rate exists and is finite (and given by):

$$\Delta_C = \lim_{t \to 0} \Delta_C(t) = \lim_{n \to \infty} \Delta_C\left(\frac{t}{2^n}\right) \tag{3.13}$$

It turns out that the instantaneous crossing rate will play an important part in the effort to find expressions for the different excess distributions. Before we go further to find the different moments and distributions of the different excess times we shall first show the following important results which relate crossing rate Δ_C to the derivative of the excess functions $\psi_C(t)$ (and$\phi_C(t)$) at $t = 0$:

$$\Delta_C = -\psi_C{}'(0) = -\phi_C{}'(0) = -\psi_C^n{}'(0) = -\phi_C^n{}'(0) \tag{3.14}$$

To prove (3.14) we start with the obvious inequality $\psi_C(t) \le \boldsymbol{P}\{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\} = \boldsymbol{P}\{\boldsymbol{B}_0 > C\} - \boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\}$ implying that $-\frac{\psi_C(t) - \psi_C(0)}{t} \ge \Delta_C(t)$. The last inequality shows that if $\Delta_C(t) \to \infty$ then $-\frac{\psi_C(t) - \psi_C(0)}{t} \to \infty$ when $t \to 0$.

Next we assume that $\Delta_C = \lim_{t \to 0} \Delta_C(t)$ is finite. Then we also have $\psi_C^n(t) \le \boldsymbol{P}\{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\} = \boldsymbol{P}\{\boldsymbol{B}_0 > C\} - \boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\}$ for $n = 0, 1, 2, \ldots$ (with equality for $n = 0$) giving $-\frac{\psi_C^n(t) - \psi_C(0)}{t} \ge \Delta_C(t)$. By writing out the difference $\psi_C(0) - \psi_C^n(t)$ as follows we find:

$$\psi_C(0) - \psi_C^n(t) = \boldsymbol{P}\{\boldsymbol{B}_0 > C\} - \boldsymbol{P}\{\boldsymbol{B}_{t_0^n} > C, \boldsymbol{B}_{t_1^n} > C, \ldots, \boldsymbol{B}_{t_n^n} > C\} =$$

$$\sum_{i=1}^{2^n} \left[\boldsymbol{P}\left\{\boldsymbol{B}_{t_0^n} > C, \boldsymbol{B}_{t_1^n} > C, \ldots, \boldsymbol{B}_{t_{i-1}^n} > C\right\} - \boldsymbol{P}\left\{\boldsymbol{B}_{t_0^n} > C, \boldsymbol{B}_{t_1^n} > C, \ldots, \boldsymbol{B}_{t_i^n} > C\right\} \right] = \sum_{i=1}^{2^n} \left[\boldsymbol{P}\left\{\boldsymbol{B}_{t_0^n} > C, \boldsymbol{B}_{t_1^n} > C, \ldots, \boldsymbol{B}_{t_{i-1}^n} > C, \boldsymbol{B}_{t_i^n} \le C\right\} \right]$$

Each term in the last sum is obviously bounded by $\boldsymbol{P}\{\boldsymbol{B}_{t_{i-1}^n} > C, \boldsymbol{B}_{t_i^n} \le C\} = \boldsymbol{P}\left\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_{\frac{t}{2^n}}\right\} = \frac{t}{2^n}\Delta_C\left(\frac{t}{2^n}\right)$ (where the equality is due to the assumption of stationarity of the stochastic process). We therefore have $\frac{\psi_C(0) - \psi_C^n(t)}{t} \le \Delta_C\left(\frac{t}{2^n}\right)$ for $n = 0, 1, 2, \ldots$ (with equality for $n = 0$). Combining the two inequalities above gives:

$$\Delta_C\left(\frac{t}{2^n}\right) \geq -\frac{\psi_C^n(t) - \psi_C(0)}{t} \geq \Delta_C(t) \text{ (with equality for } n = 0) \tag{3.15}$$

If we now fix t and let $n \to \infty$ in (3.15) then we get:

$$\Delta_C \geq -\frac{\psi_C(t) - \psi_C(0)}{t} \geq \Delta_C(t) \tag{3.16}$$

and the result (3.14) follows now by letting $t \to 0$ in (3.15) and (3.16).

As a consequence of the inequality we get the following bounds on the function $\psi_C(t)$ and $\psi_C^n(t)$ (for small t):

$$\boldsymbol{P}\{\boldsymbol{B}_0 > C\} - t\Delta_C \leq \psi_C(t) \leq \boldsymbol{P}\{\boldsymbol{B}_0 > C\} - t\Delta_C(t) \text{ and} \tag{3.17}$$

$$\boldsymbol{P}\{\boldsymbol{B}_0 > C\} - t\Delta_C\left(\frac{t}{2^n}\right) \leq \psi_C^n(t) \leq \boldsymbol{P}\{\boldsymbol{B}_0 > C\} - t\Delta_C(t) \tag{3.18}$$

The proof for $\phi_C(t)$ is similar and is therefore omitted, but the corresponding inequalities (3.15) and (3.16) yield by replacing $\psi_C(t)$ with $\phi_C(t)$ and $\psi_C^n(t)$ with $\phi_C^n(t)$ and further the bounds (3.17) and (3.18) will read:

$$\boldsymbol{P}\{\boldsymbol{B}_0 \leq C\} - t\Delta_C \leq \phi_C(t) \leq \boldsymbol{P}\{\boldsymbol{B}_0 \leq C\} - t\Delta_C(t) \text{ and} \tag{3.19}$$

$$\boldsymbol{P}\{\boldsymbol{B}_0 \leq C\} - t\Delta_C\left(\frac{t}{2^n}\right) \leq \phi_C^n(t) \leq \boldsymbol{P}\{\boldsymbol{B}_0 \leq C\} - t\Delta_C(t) \tag{3.20}$$

(We should mention here that when we derive (3.14) by (3.15) and (3.16) we only assume that subintervals are of equal lengths, which means the results also yield for the case where the we divide the interval into subinterval of equal lengths $\frac{t}{n}$)

It is also of interest to find the probability of having more than one crossing (of the level C) in an interval of length t when Δ_C is finite. We let $\Lambda_C(t)$ denote this probability and find the following result:

If the crossing intensity Δ_C is finite then $\lim_{t \to 0} \frac{\Lambda_C(t)}{t} = 0$ or $\Lambda_C(t) = o(t)$ as $t \to 0$.

To show this result we define the probability of having an even number of crossings in an interval of length t by $\Lambda_C^{even}(t)$. Since in this case either both the starting point and the end point are above or below the level C we must have

$\Lambda_C^{even}(t) = \psi_C^0(t) - \psi_C(t) + \phi_C^0(t) - \phi_C(t)$. Similar we also define the probability of having an odd number (greater than one) of crossings in an interval of length t by $\Lambda_C^{odd}(t)$. To have an odd number (greater than one) of crossings in an interval of length t we must have an even number of crossing in the interval $(0, \tau)$, a single crossing in $(\tau, \tau + d\tau)$ and no crossing in $(\tau + d\tau, t)$. Therefore $\Lambda_C^{odd}(t)$ will be bounded by the integral of $\Lambda_C^{even}(t)$, that is we have $\Lambda_C^{odd}(t) \leq \int_{\tau = 0}^{t} \Lambda_C^{even}(\tau) d\tau = t\Lambda_C^{even}(t_1)$ for some $0 \leq t_1 \leq t$. Then by (3.15) (where we have equality for $n = 0$) and (3.16) we get $0 \leq \frac{\Lambda_C(t)}{t} = \frac{\Lambda_C^{even}(t) + \Lambda_C^{odd}(t)}{t} \leq 2(\Delta_C - \Delta_C(t)) + 2t(\Delta_C - \Delta_C(t_1))$ for $0 \leq t_1 \leq t$. If we let $t \to 0$ the result follows.

By applying the nice property above we may neglect the probability of having more than one level crossing in a small interval (of say length dt since we will have $\Lambda_C(dt) = dt \cdot o(dt)$) and this will heavily simplify the derivation of the different excess distributions below where we always shall assume that the crossing intensity is finite.

3.2.2 Distribution and moments of the excess times

In this section we shall discuss a general framework to get the excess time distributions T_k and S_k for a general stationary stochastic process. Sometimes we also want to get the time to the first down crossing (first passage time) conditioning on the bit rate process (when we are inside an excess period). We therefore also define T^x to be the time to the first down crossing for the process $\{\boldsymbol{B}_t\}$ when we start the observation in an excess period with $\boldsymbol{B}_0 = x$.

It turns out that the first passage time may be expressed in terms of what we call envelope probabilities (or excess probabilities) defined by:

$$F_C(x, y, t) = \boldsymbol{P}\{\boldsymbol{B}_t > y, Inf_{\tau \in (0, t)} \boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = x\} \text{ for } x \geq C,\ y \geq C \text{ and} \tag{3.21}$$

$$G_C(x, y, t) = \boldsymbol{P}\{\boldsymbol{B}_t \leq y, Sup_{\tau \in (0, t)} \boldsymbol{B}_\tau \leq C | \boldsymbol{B}_0 = x\} \text{ for } x \leq C,\ y \leq C \tag{3.22}$$

and where we also denote the corresponding densities $f_C(x, y, t) = \frac{\partial}{\partial y} F_C(x, y, t)$ and $g_C(x, y, t) = \frac{\partial}{\partial y} G_C(x, y, t)$.

The CDF (Complementary Distribution Function) of the first passage time T^x may be obtained from (3.21) by setting $y = C$, so we get

$$P(T^x > t) = \boldsymbol{P}\{Inf_{\tau \in (0,t)} \boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = x\} = F_C(x, C, t) \tag{3.23}$$

Below we will show that we, on the basis of the characteristics (3.21) and (3.22) and the stationary distribution, are able to derive the excess time distribution above, $\psi_C(t)$ (or below $\phi_C(t)$), the level C for the bit rate process. Based on the definitions above we get the following relations to the excess probabilities defined by (3.3) and (3.4):

$$\psi_C(t) = \int_{x=C}^{\infty} F_C(x, C, t) d\Phi(x) \text{ and} \tag{3.24}$$

$$\phi_C(t) = \int_{x=0}^{C} G_C(x, C, t) d\Phi(x) \tag{3.25}$$

where $\Phi(x)$ is the stationary CDF of $\boldsymbol{B}_t$ i.e. $\Phi(x) = P\{\boldsymbol{B}_0 \le x\}$. (If the distribution function is differentiable we have $d\Phi(x) = \varphi(x)dx$ where $\varphi(x)dx = P\{\boldsymbol{B}_0 \in (x, x + dx)\}$ is the PDF (Probability Density Function) of $\boldsymbol{B}_t$.)

The functions $\psi_C(t)$ (and $\phi_C(t)$) are the probabilities that the process either is above (or below) the level C and does not cross that level in an interval of length t. Based on these key probabilities we may proceed to obtain the distribution of the excess times T_k. Now, to have an up crossing in a small interval $(0, dt)$ we must have $B_0 \le C$ and $B_{dt} > C$ (see figure 3.2).

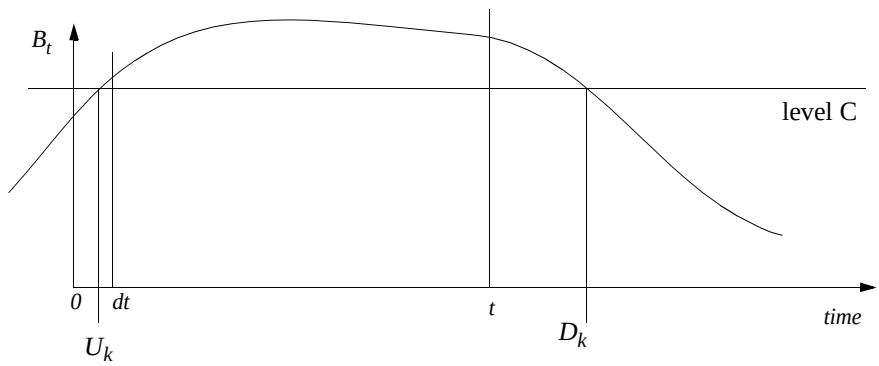

Figure 3.2: The excess time distribution based on up and down crossing instances of the bit rate process.

Conditioning on this event and observing that the event

$\{\boldsymbol{B}_0 \le C, Inf_{\tau \in (dt,\, t)} \boldsymbol{B}_\tau > C\} = \{Inf_{\tau \in (dt,\, t)} \boldsymbol{B}_\tau > C\} - \{Inf_{\tau \in (0,\, t)} \boldsymbol{B}_\tau > C\}$ we may express the excess time distribution as

$P(T_k > t) = \lim_{dt \to 0} P(\boldsymbol{B}_0 \le C, Inf_{\tau \in (dt,\, t)} \boldsymbol{B}_\tau > C | \boldsymbol{B}_0 \le C, \boldsymbol{B}_{dt} > C)$; and by rewriting the event above and using the assumption of having a stationary process the last expression may be rewritten as:

$$F_{T_k}(t) = P(T_k > t) = \lim_{dt \to 0} \frac{\psi_C(t - dt) - \psi_C(t)}{\psi_C(0) - \psi_C(dt)} = \frac{-\psi_C{}'(t)}{\Delta_C} \tag{3.26}$$

provided that Δ_C is finite. Hence, these results state that the CDF of an excess period for a general stationary stochastic process is given as the normalized derivative of the excess probability (3.24). The result (3.26) requires that crossing intensity Δ_C (or the derivative at time $t = 0$) exists and is finite. Unfortunately this is not the case for some rather interesting classes of continuous processes. For the Ornstein-Uhlenbeck (U-O) process the derivative $\psi_C{}'(t)$ will become arbitrary large for small values of t so the limit (3.26) will become zero for all t. This is in accordance with the well known rapid oscillations for the Wiener process and the O-U process described in many textbooks for stochastic processes [Cox70], [Karl66]. (We shall discuss the O-U process in section 4.4.)

An alternative to consider the excess time distribution defined above (which does not always exists) we may consider a somewhat simpler variable taken to be the first passage time T_C given that the bit rate process is above the level C. The corresponding CDF may be found by integrating (3.23) by the conditional distribution of $\boldsymbol{B}_0$ given that $\boldsymbol{B}_0 > C$:

$$P(T_C > t) = \int_{x = C}^{\infty} F_C(x, C, t) \frac{d\Phi(x)}{P(\boldsymbol{B}_0 > C)} = \frac{\psi_C(t)}{P(\boldsymbol{B}_0 > C)} \tag{3.27}$$

To obtain the CDF of the time distribution of the “normal load” period; that is the distribution of S_k we proceed as for the “overload” case, and we get:

$$F_{S_k}(t) = P(S_k > t) = \lim_{dt \to 0} \frac{\phi_C(t - dt) - \phi_C(t)}{\phi_C(0) - \phi_C(dt)} = \frac{-\phi_C{}'(t)}{\Delta_C} \tag{3.28}$$

On the basis of (3.26) and (3.28) it is straight forward to find the first two moments of the excess times:

$$\boldsymbol{E}[T_k] = \frac{P(\boldsymbol{B}_0 > C)}{\Delta_C} \text{ and } \boldsymbol{E}[S_k] = \frac{P(\boldsymbol{B}_0 \le C)}{\Delta_C} \text{ and} \tag{3.29}$$

$$\boldsymbol{E}[T_k^{\,2}] = \frac{2\int_0^\infty \psi_C(t)\,dt}{\Delta_C} \quad \text{and} \quad \boldsymbol{E}[S_k^{\,2}] = \frac{2\int_0^\infty \phi_C(t)\,dt}{\Delta_C}. \tag{3.30}$$

In words it is possible to express the mean excess times as the stationary probability that the bit rate is above (or below) the capacity limit divided by the rate at which the process cross that level.

We may also obtain the expected time between two consecutive up or down crossings:

$$\boldsymbol{E}[U_{k+1} - U_k] = \boldsymbol{E}[S_k] + \boldsymbol{E}[T_k] = \frac{1}{\Delta_C} \tag{3.31}$$

Thus, the mean excess times above (or below) a given capacity level could as well be derived by direct arguments; as the portion of time the process is above (or below) a given level decided by the up or down crossing rate.

3.2.3 Moments of the excess volumes

By taking the expectation of the stochastic integrals (3.2) we may express the mean values of the excess volumes as:

$$\boldsymbol{E}[A_k] = \boldsymbol{E}[T_k]\boldsymbol{E}[\boldsymbol{B}_0 - C|\boldsymbol{B}_0 > C] = \frac{\int_C^\infty P(\boldsymbol{B}_0 > x)\,dx}{\Delta_C} \quad \text{and} \tag{3.32}$$

$$\boldsymbol{E}[V_k] = \boldsymbol{E}[S_k]\boldsymbol{E}[C - \boldsymbol{B}_0|\boldsymbol{B}_0 \le C] = \frac{\int_0^C P(\boldsymbol{B}_0 \le x)\,dx}{\Delta_C} \tag{3.33}$$

Based on the results above we may estimate the overall information loss p_{loss} as the ratio of the mean excess volume and the mean traffic volume in a cycle:

$$p_{loss} \approx \frac{\int_C^\infty P(\boldsymbol{B}_0 > x)\,dx}{m} \tag{3.34}$$

and the information loss in an overload period q_{loss} as the ratio of the mean excess volume and the mean traffic volume in an excess period:

$$q_{loss} \approx \frac{\int_C^\infty P(\boldsymbol{B}_0 > x)dx}{mP(\boldsymbol{B}_0 > C)} \tag{3.35}$$

We observe that the crossing rate is not included in the estimates (3.34) and (3.35) and they may as well be used for processes where crossing rate does not exist. To justify this we may consider a sampled version of the process (where we take the sampled version linear between samples). If we choose a sample interval of length θ the corresponding crossing rate is $\Delta_C(\theta)$ and by choosing θ small (but not $\theta = 0$) we will get the estimates (3.34) and (3.35).

To obtain the second order moments of the excess volumes we define the conditional covariances

$$\gamma_C(t) = \boldsymbol{E}[(\boldsymbol{B}_t - C)(\boldsymbol{B}_0 - C)\mathbf{1}_{\{Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C\}}] = \int_{x=C}^{\infty}\int_{y=C}^{\infty}(x-C)(y-C)\varphi(x)f_C(x,y,t)dxdy \tag{3.36}$$

$$\beta_C(t) = \boldsymbol{E}[(C - \boldsymbol{B}_t)(C - \boldsymbol{B}_0)\mathbf{1}_{\{Sup_{\tau \in (0,t)}\boldsymbol{B}_\tau \le C\}}] = \int_{x=0}^{C}\int_{y=0}^{C}(C-x)(C-y)\varphi(x)g_C(x,y,t)dxdy \tag{3.37}$$

First we calculate the conditional moment $\boldsymbol{E}[A_k^2 \big| T_k]$. After some manipulations we find:

$$\boldsymbol{E}[A_k^2 \big| T_k] = 2\int_{\xi=0}^{T_k}(T_k - \xi)\frac{\gamma_C(\xi)}{\psi_C(\xi)}d\xi \tag{3.38}$$

By the theorem of double expectation (see for instance [Cinl75]) we get from (3.38):

$$\boldsymbol{E}[A_k^2] = \boldsymbol{E}[\boldsymbol{E}[A_k^2 \big| T_k]] = \frac{2\int_{t=0}^{\infty}\gamma_C(t)dt}{\Delta_C} \tag{3.39}$$

By proceeding in the same way for V_k we finally obtain:

$$\boldsymbol{E}[V_k^2] = \frac{2\int_{t=0}^{\infty}\beta_C(t)dt}{\Delta_C} \tag{3.40}$$

3.3 Further results for bit rate processes that are continuous in time and space

The rest of this chapter will be devoted to processes that have continuous sample paths and are absolute continuous (which means that the PDF for the process exists and is a continuous function). In the literature there exists sufficient criteria for a stationary process having continuous sample paths. We refer to the textbook of H. J. Larson and B. O. Shubert [Lars79b] where it is stated that if it possible to find constants such that

$$1-\rho(t) \le at^{\gamma} \text{ for } 0 < t \le l \tag{3.41}$$

where $a > 0$ and $\gamma > 1$ then the process is sample path continuous. If the process is Gaussian one can relax the demand on γ to have $\gamma > 0$. In the following we shall assume that these criteria are fulfilled to be sure that the corresponding processes are sample path continuous.

3.3.1 Up and down crossing intensity

For absolute continuous processes it is possible to write the crossings intensity as an integral. To do this we define $\boldsymbol{dB}_t = \dfrac{\boldsymbol{B}_t - \boldsymbol{B}_0}{t}$ which is the differential process scaled by $1/t$ (when t is small the scaled differential process will be close to the derivative of $\boldsymbol{B}_t$ provided that the derivative exists).

By conditioning on $\boldsymbol{B}_0$ we may write the up and down crossing intensity as:

$$\Delta_C^{uc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 \le C < \boldsymbol{B}_t\} = \int_{y=0}^{\infty} P(\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = C - ty)\varphi(C - ty)dy \text{ and} \tag{3.42}$$

$$\Delta_C^{dc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\} = \int_{y=0}^{\infty} P(-\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = C + ty)\varphi(C + ty)dy \tag{3.43}$$

where $\varphi(z)$ is the PDF of $\boldsymbol{B}_0$. If the functions $F_{1_t}(y,z) = P(\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = z)\varphi(z)$ and $F_{2_t}(y,z) = P(-\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = z)\varphi(z)$ are uniform continuous with respect to z for all y then one may take the limit $t \to 0$ under the integral sign in (3.42) and (3.43) giving:

$$\Delta_C = \frac{1}{2}\lim_{t\to 0}(\Delta_C^{uc}(t) + \Delta_C^{dc}(t)) = \frac{\varphi(C)}{2}\lim_{t\to 0}\int_{y=0}^{\infty}(P(\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = C) + P(-\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = C)dy)$$

$$= \frac{\varphi(C)}{2}\lim_{t\to 0}\boldsymbol{E}[|\boldsymbol{dB}_t| \,|\, \boldsymbol{B}_0 = C] \tag{3.44}$$

In Appendix A we have given a more rigorous proof of (3.44) where we assume that $F_{1_t}(y, z)$ and $F_{2_t}(y, z)$ satisfy the *condition A*:

$$\left|F_{1_t}(y, C) - F_{1_t}(y, x)\right| \le M_t(y)|C - x| \quad \text{and} \quad \left|F_{2_t}(y, C) - F_{2_t}(y, x)\right| \le M_t(y)|C - x| \quad (\text{for } y \ge 0)$$

where $M_t = \int_{y=0}^{\infty} yM_t(y)dy$ exists and further $\lim_{t \to 0} tM_t = 0$.

In practice the hard part in this approach is to find the function $M_t(y)$ used in the *condition A*, and this task will often be as difficult as to find each of the limits in (3.44) direct.

It would be nice to try to link the existence of the crossing intensity to the second order statistics for the process, that is the behaviour of the autocorrelation function for small t. We have $\boldsymbol{E}[\boldsymbol{dB}_t^{\,2}] = \frac{2\sigma^2}{t^2}(1 - \rho(t))$ and by [Fell68b] (see page 155), we have that the mean scaled drift is bounded by the inequality:

$$\boldsymbol{E}[\left|\boldsymbol{dB}_t\right|] \le \sqrt{\boldsymbol{E}[\boldsymbol{dB}_t^{\,2}]} = \frac{\sigma}{t}\sqrt{2(1 - \rho(t))}. \tag{3.45}$$

Now it is clear that to have the limit $\boldsymbol{E}[\boldsymbol{dB}_t^{\,2}]$ to exists for $t \to 0$ we must have

$$\rho(t) = 1 - at^2 + o(t^2) \tag{3.46}$$

for small t where a is a positive constant.

If we integrate the crossing intensity $\Delta_C(t)$ over all the crossing levels we get (by applying both (3.42) and (3.43)):

$\int_{C=-\infty}^{\infty} \Delta_C(t)dC = \frac{1}{2}\int_{C=-\infty}^{\infty}\left(\int_{y=0}^{\infty}[P(\boldsymbol{dB}_t > y|\boldsymbol{B}_0 = C - ty)\varphi(C - ty) + P(-\boldsymbol{dB}_t > y|\boldsymbol{B}_0 = C + ty)\varphi(C + ty)]dy\right)dC$, and by changing the integration we get:

$$\int_{C=-\infty}^{\infty} \Delta_C(t)dC = \frac{1}{2}\int_{y=0}^{\infty} P(\left|\boldsymbol{dB}_t\right| > y)dy = \frac{1}{2}\boldsymbol{E}[\left|\boldsymbol{dB}_t\right|] \tag{3.47}$$

We may now state the following result: If we assume $\rho(t)$ to be on the form (3.46) then we have by (3.45) that $\lim_{t \to 0}\boldsymbol{E}[\left|\boldsymbol{dB}_t\right|] = \gamma$ with γ finite and

$\gamma = \lim_{t \to 0} \boldsymbol{E}[|\boldsymbol{dB}_t|] \le \lim_{t \to 0} \sqrt{\boldsymbol{E}[\boldsymbol{dB}_t^2]} = \sigma\sqrt{2a}$ and furthermore by the monotone convergence theorem [Royd68] $\lim_{n \to \infty} \Delta_C\left(\frac{t}{2^n}\right) = \lim_{t \to 0} \Delta_C(t) = \Delta_C$ is finite a.s. and further

$$\lim_{t \to 0} \boldsymbol{E}[|\boldsymbol{dB}_t|] = 2 \int_{C=-\infty}^{\infty} \Delta_C dC \tag{3.48}$$

Of course (3.46) put a quite strict limitation on the class of autocorrelation functions for which it is meaningful to use the notion of up and down crossing intervals even though the processes may have continuous sample paths. Thus, the requirement of having continuous sample paths combined by requirements of having finite up and down crossing intensities will limit the class of processes to those having autocorrelation function of type (3.46). This is in strict contrast to processes with jumps, for instance stationary Markov processes where the up and down crossing intensities will exists even though the autocorrelation function is on the form $\rho(t) = 1 - at + o(t^2)$ for small t where a is a positive constant. (Processes with jumps are discussed in detail in chapter 5 in this thesis.)

As a side results, by applying the inequality above, we find the integral over the up or down crossing intensity is bounded by $\sigma\sqrt{\frac{a}{2}}$ that is:

$$\int_{C=-\infty}^{\infty} \Delta_C dC \le \sigma\sqrt{\frac{a}{2}} = \sigma\sqrt{-\rho''(0)} \tag{3.49}$$

Summarising the discussion above we have shown that sufficient condition that the up and down crossing intensities exist and are finite a.s. is that the autocorrelation is on the form (3.46). Whether this also is a necessary condition will not tried to be answered in this thesis, however, for a Gaussian processes this is the case. As a side result of the discussion we obtain a bound on the integral of the crossing rates and provide a measure of the variability in the process that is proportional with the standard deviation and the square root of the second order derivative of the autocorrelation (at the origin) with negative sign.

3.3.2 Joint distribution of the first passage time and the corresponding volume

The general formulae above give a framework to find the first two moments of the excess volumes. However, in many performance questions the interested part is the tail of the distributions, which generally is much harder to obtain. In the following the aim is also to include the volume in the analysis and to do so we must also include the volume in the distribution (3.21). In the succeeding we shall assume that the bit rate process is continuous in time and space, and we limit ourselves to consider the most interesting case, the excess volumes when the process is above the capacity level C. (A similar analysis is possible to perform for “the normal case” when the process is below the level C.) We let

$A_t = \int_0^t (\boldsymbol{B}_\tau - C)d\tau$ be the excess volume up to a certain time t, and we assume that the following time dependant probability distribution is known:

$$F_C(x, y, z, t) = \boldsymbol{P}\{\boldsymbol{B}_t > y, A_t > z, Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = x\} \text{ for } x \geq C,\ y \geq C \tag{3.50}$$

with the corresponding PDF:

$$f_C(x, y, z, t) = \frac{\partial^2}{\partial y \partial z} F_C(x, y, z, t) \tag{3.51}$$

Based on these time dependant functions we will first derive the joint PDF of the first passage time T^x and the corresponding excess volume $A^x = A_{T_x} = \int_0^{T^x} (\boldsymbol{B}_\tau - C)d\tau$. We let $F_{A^x T^x}(x, z, t) = P(A^x > z, T^x > t)$ be the joint CDF and $f_{A^x T^x}(x, z, t) = \frac{\partial^2}{\partial t \partial z} F_{A^x T^x}(x, z, t)$ the corresponding joint PDF. The event "down crossing in the interval $(t, t + dt)$", that is $T^x \in (t, t + dt)$, is equivalent with the event $\{Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C, \boldsymbol{B}_{t+dt} \leq C\}$. This event may be written as the difference $\{Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C\} - \{Inf_{\tau \in (0,t+dt)}\boldsymbol{B}_\tau > C\}$ (see figure 3.3).

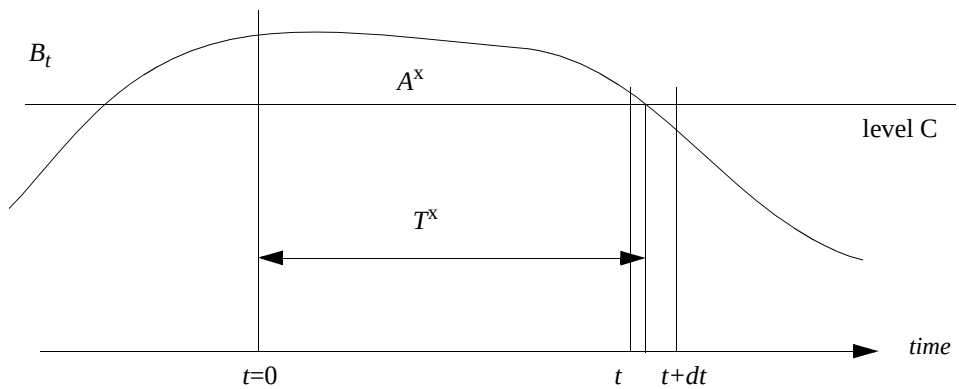

Figure 3.3: The excess time and excess volume and down crossing instances of the bit rate process.

From the definition of T^x and A^x we have

$$P(A^x \in (z, z + dz), T^x \in (t, t + dt)) = P(A_t \in (z, z + dz), Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C, \boldsymbol{B}_{t+dt} \leq C) =$$

$$P(A_t \in (z, z+dz), Inf_{\tau \in (0,t)} \boldsymbol{B}_\tau > C) - P(A_t \in (z, z+dz), Inf_{\tau \in (0,t+dt)} \boldsymbol{B}_\tau > C)$$

The first part of this expression is given by

$P(A_t \in (z, z+dz), Inf_{\tau \in (0,t)} \boldsymbol{B}_\tau > C) = \int_{y=C}^{\infty} f_C(x, y, z, t)\, dy dz$. The second part is some more difficult to obtain (expressed in terms of the function $f_C(x, y, z, t)$) because it involves the changes in the volume in the interval $(t, t+dt)$. Some informally we have $dA_t = (\boldsymbol{B}_t - C)dt$ which implies $A_{t+dt} = A_t + (\boldsymbol{B}_{t+dt} - C)dt$, so to have $A_t \in (z, z+dz)$ we must have $A_{t+dt} \in (z + (\boldsymbol{B}_{t+dt} - C)dt, z + (\boldsymbol{B}_{t+dt} - C)dt + dz)$. Conditioning on $\boldsymbol{B}_{t+dt} = y$ and integrating we then get:

$$P(A_t \in (z, z+dz), Inf_{\tau \in (0,t+dt)} \boldsymbol{B}_\tau > C) = \int_{y=C}^{\infty} f_C(x, y, z + (y-C)dt, t+dt)\, dy dz$$

By expanding for small dt we finally get:

$$f_{T^x A^x}(x, z, t) = -\int_{y=C}^{\infty} \left((y-C)\frac{\partial f_C}{\partial z} + \frac{\partial f_C}{\partial t} \right) dy \tag{3.52}$$

The main restrictions we put on the bit rate process $\boldsymbol{B}_t$ to make the derivation above correct are mainly that the sample paths of $\boldsymbol{B}_t$ must be continuous. With this assumption we have $A_{t+dt} - A_t = \int_{\tau=t}^{t+dt} (\boldsymbol{B}_\tau - C)d\tau = (\boldsymbol{B}_{t+dt} - C)dt + \int_{\tau=t}^{t+dt} (\boldsymbol{B}_{t+dt} - \boldsymbol{B}_\tau)d\tau$ and due to the continuity of the sample paths the last integral will of order δdt where δ can be made arbitrarily small (depending on dt). Unfortunately the method fails if the sample paths of the process contain some kind of jumps. In this case the evolution of the volume (in time) will depend on the bit rate both before and after the jumps. This is in contrast to the results derived for the excess times in section 3.2.2 where no particular assumption is made about the continuity of the sample paths. (In chapter 5 we consider the case where the bit rate process is a semi-Markov process and therefore containing jumps.)

To find the time dependent CDF (3.50) and PDF (3.51) for specific models is of cause the hard part to find explicit expressions of the joint excess distributions. However, these functions must have some specific initial and boundary conditions to make the proposed description meaningful. Firstly, the initial conditions obtained by letting $t \to 0$ in the definition of f_C, imply that:

$$f_C(x, y, z, 0) = \delta(y - x)\delta(z) . \tag{3.53}$$

where $\delta(u)$ denotes Diracs delta function. Secondly, this assumption and the claim that $f_{A^x T^x}(x, z, t)$ is a proper probability density function impose the following boundary condition on f_C:

$$f_C(x, y, 0, t) = 0 \text{ for } t > 0 , x > C, y > C. \tag{3.54}$$

We recognize that this is the appropriate boundary condition at $z = 0$ since the volume A_t cannot be zero for positive time.

The joint CDF $F_{A^x T^x}(x, z, t) = P(A^x > z, T^x > t | B_0 = x)$ can be expressed in terms of (3.50) and (3.51) by using (3.52):

$$F_{T^x A^x}(x, z, t) = F_C(x, C, z, t) + \int_{y=C}^{\infty} (y - C) \int_{\tau=t}^{\infty} f_C(x, y, z, \tau) d\tau dy \tag{3.55}$$

The marginal PDF and CDF for the excess volume A^x is easily obtained from (3.52) and (3.55):

$$f_{A^x}(x, z) = \delta(z) - \int_{y=C}^{\infty} (y - C) \int_{t=0}^{\infty} \frac{\partial f_C}{\partial z}(x, y, z, t) dt dy \text{ and} \tag{3.56}$$

$$F_{A^x}(x, z) = P(A^x > z \,|\, \boldsymbol{B}_0 = x) = \int_{y=C}^{\infty} (y - C) \int_{t=0}^{\infty} f_C(x, y, z, t) dt dy \tag{3.57}$$

Due to the initial impuls of the joint density $f_C(x, y, z, t)$ (at the origin) one must be careful with the limit of integration at $t = 0$ (and $z = 0$).

The functional relation (3.52) with the initial and boundary condition deduced above will become more apparent if we introduce the double Laplace transform $\hat{f}_{T^x A^x}(x, s, \zeta) = E[e^{-sT^x - \zeta A^x} | \boldsymbol{B}_0 = x]$. From (3.52) we get the following functional relation:

$$\hat{f}_{T^x A^x}(x, \zeta, s) = 1 - \int_{y=C}^{\infty} ((y - C)\zeta + s) \hat{f}_C(x, y, \zeta, s) dy \tag{3.58}$$

where $\hat{f}_C(x, y, \zeta, s)$ is the double LST (Laplace-Stieltjes Transform) of f_C in the variable z and t defined by:

$$\hat{f}_C(x, y, \zeta, s) = \int_{z=0}^{\infty} \int_{t=0}^{\infty} e^{-st-\zeta z} f_C(x, y, z, t) dz dt \tag{3.59}$$

3.3.3 Joint distribution of the excess times and excess volumes

We shall now proceed with a quite similar analysis as above to get the joint PDF of the pair (T_k, A_k), bearing in mind that this pair of variable is determined through a second order limiting procedure, that is we both require up crossing in a small interval $(0, dt_1)$ and also down crossing in a second small interval $(t, t + dt_2)$. Before entering the analysis we define the following joint density:

$g_C(x, y, z, t) dx dy dz =$

$P(\boldsymbol{B}_0 \in (x, x + dx), \boldsymbol{B}_t \in (y, y + dy), A_t \in (z, z + dz), Inf_{\tau \in (0, t)} \boldsymbol{B}_\tau > C)$, then by (3.50) and (3.51)

$$g_C(x, y, z, t) = \varphi(x) f_C(x, y, z, t). \tag{3.60}$$

We let $F_{A_k T_k}(z, t) = P(A_k > z, T_k > t)$ be the joint CDF of the pair (T_k, A_k) and $f_{A_k T_k}(z, t) = \frac{\partial^2}{\partial t \partial z} F_{A_k T_k}(z, t)$ the corresponding joint density function. Some informal we have:

$$f_{A_k T_k}(z, t) dz dt_2 = \lim_{dt_1 \to 0} \frac{P(A_k \in (z, z + dz), T_k \in (t, t + dt_2) \text{and upcrossing in}(0, dt_1))}{P(\text{upcrossing in}(0, dt_1))} =$$

$$\lim_{dt_1 \to 0} \frac{P(A_t \in (z, z + dz), \boldsymbol{B}_0 \le C, Inf_{\tau \in (dt_1, t)} \boldsymbol{B}_\tau > C, \boldsymbol{B}_{t + dt_2} \le C)}{P(\boldsymbol{B}_0 \le C, \boldsymbol{B}_{dt_1} > C)} \tag{3.61}$$

provided that the limit exists. By using a similar approach as we applied for the conditional excess times and excess volumes it is possible to expand the nominator to second order for small dt_1 and dt_2. These rather technical details are placed in Appendix A where we find:

$P(A_t \in (z, z + dz), \boldsymbol{B}_0 \le C, Inf_{\tau \in (dt_1, t)} \boldsymbol{B}_\tau > C, \boldsymbol{B}_{t + dt_2} \le C) =$

$\int_{y=C}^{\infty}\int_{x=C}^{\infty}\left((y-C)(x-C)\frac{\partial^2 g_C}{\partial z^2}+(x+y-2C)\frac{\partial^2 g_C}{\partial z\partial t}+\frac{\partial^2 g_C}{\partial t^2}\right)dxdy\cdot dzdt_1dt_2$ for small dt_1 and dt_2.

If is possible to expand the denominator in (3.61) to first order for small dt_1 the joint density function for (T_k, A_k) will be given by:

$$f_{A_kT_k}(z,t) = \frac{1}{\Delta_C}\int_{y=C}^{\infty}\int_{x=C}^{\infty}\left((y-C)(x-C)\frac{\partial^2 g_C}{\partial z^2}+(x+y-2C)\frac{\partial^2 g_C}{\partial z\partial t}+\frac{\partial^2 g_C}{\partial t^2}\right)dxdy \quad (3.62)$$

An alternative way of writing (3.62) is found by applying the conditional PDF $f_{A^xT^x}(x,z,t)$ given by (3.52) and then define:

$$h_C(x,z,t) = \varphi(x)f_{A^xT^x}(x,z,t). \quad (3.63)$$

giving the following alternative way of writing $f_{A_kT_k}(z,t)$:

$$f_{A_kT_k}(z,t) = \frac{1}{\Delta_C}\int_{x=C}^{\infty}\left((x-C)\frac{\partial h_C}{\partial z}+\frac{\partial h_C}{\partial t}\right)dx \quad (3.64)$$

Unfortunately the restriction we have put on the bit rate process will somehow limit the usefulness of the last derived formula. As for the conditional down crossing the analysis is limited by the assumption that the sample paths of the bit rate process have continuous sample paths. Secondly, and perhaps more restrictive, is the claim on behaviour of the excess probability $\psi_C(t)$ for small t. As stated in section 3.3.1 it is sufficient that the autocorrelation function is on the form $\rho(t) = 1-at^2+o(t^2)$ for small t where a is a positive constant.

The functional relation given by (3.63) and (3.64) is on the same form as (3.52) and therefore it is easy to write down the corresponding LST. Before doing this we must examine the conditional PDF at the boundaries $f_{A^xT^x}(x,0,t)$ and $f_{A^xT^x}(x,z,0)$ since these specific values will be included in the transform. Some informally it is clear that if T^x lies in the interval $(t, t+dt)$ for $t>0$ the volume A^x must be positive and therefore we must have $f_{A^xT^x}(x,0,t) = 0$ for $t>0$. Then T^x lies in the interval $(0, dt)$ for small dt the conditional density function of the volume A^x given T^x must have an “impulse” like shape. We must therefore have $f_{A^xT^x}(x,z,0) = h(x)\delta(z)$ for some function $h(x)$.

If we let $\hat{f}_{T_kA_k}(s, \zeta) = E[e^{-sT_k - \zeta A_k}]$ denote the double LST for the stochastic variables (T_k, A_k) we get from the equation (3.64):

$$\hat{f}_{T_kA_k}(\zeta, s) = \frac{1}{\Delta_C} \int_{x=C}^{\infty} \varphi(x)[h(x) - ((x-C)\zeta + s)\hat{f}_{T^xA^x}(x, s, \zeta)]dx \tag{3.65}$$

Since we claim $\hat{f}_{T_kA_k}(0, 0) = 1$ it follows that $\int_{x=C}^{\infty} \varphi(x)h(x)dx = \Delta_C$. Then by inserting for $\hat{f}_{T^xA^x}(x, \zeta, s)$ from (3.58) in (3.65) we finally end up with the following expression:

$$\hat{f}_{T_kA_k}(\zeta, s) = \frac{1}{\Delta_C}\left(\Delta_C - s\int_{x=C}^{\infty} \varphi(x)dx - \zeta\int_{x=C}^{\infty} (x-C)\varphi(x)dx + \int_{x=C}^{\infty}\int_{y=C}^{\infty} ((x-C)\zeta + s)((y-C)\zeta + s)\varphi(x)\hat{f}_C(x, y, s, \zeta)dydx\right) \tag{3.66}$$

where $\hat{f}_C(x, y, \zeta, s)$ is the double LST of $f_C(x, y, z, t)$ defined in (3.59).

We are pleased to note that formula (3.66) contains all the previous formulae for the first and second order moments of the excess times and excess volumes. By direct differentiation it is easy to verify that the first and second order moments of the excess time and excess volume coincide with the formulae, (3.30) and (3.32), (3.40). We may also find the correlation between the excess time and excess volume by first evaluating:

$$\boldsymbol{E}[T_kA_k] = \frac{\partial^2}{\partial\zeta\partial s}\hat{f}_{T_kA_k}(0, 0) = \frac{2\int_{t=0}^{\infty} \alpha_C(t)dt}{\Delta_C} \tag{3.67}$$

$$\alpha_C(t) = \int_{x=C}^{\infty}\int_{x=C}^{\infty}\int_{z=0}^{\infty} (x + y - 2C)\varphi(x)f_C(x, y, z, t)dxdydt \tag{3.68}$$

We may also find the LSTs for the marginal distributions $\hat{f}_{T_k}(s) = E[e^{-sT_k}] = \hat{f}_{T_kA_k}(0, s)$ and $\hat{f}_{A_k}(\zeta) = E[e^{-\zeta A_k}] = \hat{f}_{T_kA_k}(\zeta, 0)$ from the result above. For the excess time we get:

$$\hat{f}_{T_k}(s) = \frac{1}{\Delta_C}\left(\Delta_C - s\int_{x=C}^{\infty} \varphi(x)dx + s^2\int_{x=C}^{\infty}\int_{y=C}^{\infty} \varphi(x)\hat{f}_C(x, y, 0, s)dydx\right) \tag{3.69}$$

and for the excess volume the similar result is:

$$\hat{f}_{A_k}(\zeta) = \frac{1}{\Delta_C}\left(\Delta_C - \zeta \int_{x=C}^{\infty} (x-C)\varphi(x)dx + \zeta^2 \int_{x=C}^{\infty}\int_{y=C}^{\infty} (x-C)(y-C)\varphi(x)\hat{f}_C(x, y, \zeta, 0)dydx\right) \tag{3.70}$$

3.4 Some concluding remarks

In this chapter we have shown that it is possible to relate many important performance characteristics to some basic fundamental properties which essential is described by the joint probability (3.50). This function will therefore be a natural starting point when studying specific models. The main hindrance in finding this probability is of cause the claim that the process shall have no down crossing in the interval up to the given time t. Unless for some few specific type of processes it is extremely difficult to obtain closed form expression for probabilities involving the minimum $m_t = Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau$ and maximum $M_t = Sup_{\tau \in (0,t)}\boldsymbol{B}_\tau$. By relaxing on the assumption on the minimum could hopefully give reasonable accurate results for small t but will surely give inaccurate results for larger values of t. One possible way to improve such an approach is to divide the interval $(0, t)$ into say m points $t_0 = 0 < t_1 < \ldots < t_{m-1} < t_m = t$ and then calculate the $(m+1)$-dimensional probability:

$$G_C^m(x, y, z, t) = \boldsymbol{P}\{\boldsymbol{B}_0 > x, \boldsymbol{B}_{t_1} > C, ..., \boldsymbol{B}_{t_{m-1}} > C, \boldsymbol{B}_t > y, A_t > z\} \text{ for } x \geq C,\ y \geq C \tag{3.71}$$

A natural choice will be to divide the interval by equally spacing; that is $t_i = \frac{i}{m}t$ (for $i = 0, 1, \ldots, m$). We must, however, be aware that this partition (of the interval) not necessary will lead to a decreasing sequence in m as the 2^n partition in section 3.2.1 would have given.

In principle, it could thereby be possible to obtain approximation to n'th order of the excess probabilities by using the corresponding approximative function for (3.50) defined by

$$F_C^n(x, y, z, t) = -\frac{\partial}{\partial x}(G_C^n(x, y, z, t))/\varphi(x) \tag{3.72}$$

as starting point for an approximative analysis. In the next chapter in section 4.3 we have analysed the m-point approximation for Gaussian processes and developed methods that enable obtaining the approximative distributions for the excess times for up to $m = 6$, and where we also have compared with corresponding exact results for the O-U process. As discussed in chapter 4 the differences between the exact and approximations are pronounced, also when the numbers of points are taken as high as $m = 6$. This example indicates that the convergence by (3.71) and (3.72) may be quit slow.

4

Transient behaviour of Gaussian traffic models through level crossing

4.1 Introduction

Gaussian modelling has been widely used as a powerful and successful tool in applied science. The different application areas vary from statistical communication theory to different areas in physics. Traditionally, in traffic theory, where the models usually have been of discrete nature, the continuous state models have been devoted less attention. Quite recently however, different Gaussian models have turned out to constitute an important analytical framework to describe newly observed phenomena in traffic streams. For example the self-similar behaviour observed for some type of Internet traffic may be described and analysed by applying fractional Brownian motion as arrival process. Many interesting new results for self-similar traffic may be found in the book *Self-similar Network Traffic and performance Evaluation* edited by K. Park and W. Willinger [Park00].

For Gaussian models level crossing have been studied for a rather long period. Best known are perhaps the early works of O. C. Rice, [Rice45] and [Rice48], where the famous Rice's formula on the crossing rate for Gaussian processes is given. He also gives some preliminary results on the distribution between two successive zeros. These results have been extended by J. McFadden and M. S. Longuet-Higgens in later works [McFa56], [McFa58] and [Long58], [Long62] where different approximations to get the distribution between successive zeros are discussed. These papers are all based on problems within the field of random noise and are not direct applicable to traffic models where we are interested in deviations of the processes having small probabilities rather than variations around the mean value.

There is one main concern when applying Gaussian models to describe network traffic. This is due to the irregular sample paths for such models. It turns out that the autocorrelation must have specific behaviour near the origin to have finite up and down crossing intensities [Lead83]. This limitation fits rather badly with the possibility of having so called *long-range dependence* where the autocorrelation behaves as

$$\rho(t) \sim ct^{2H-2} \text{ as } t \to \infty \text{ for } 1/2 < H < 1 \tag{4.1}$$

where as the requirement of having finite up and down crossing intensities requires [Lead83]:

$$\rho(t) \sim 1 - at^2 \text{ as } t \to 0 \tag{4.2}$$

If we for instance consider an autocorrelation function on the form $\rho(t) = \dfrac{1}{1 + at^{2-2H}}$, then (4.1) is fulfilled for $1/2 < H < 1$ but to get (4.2) this requires $H = 0$ which gives a process that is not long-range dependent. In a paper by A. Barbe [Barb92] a method is described to relax condition (4.2). This can be done by considering a sampled version of the process (and let the process be linear between samples) and only counting the crossings of the sampled process. For the sampled process the crossing intensities will be finite. There is however, a problem to choose the appropriate size of the sampling interval. We shall use the first passage time and the corresponding volume as an alternative measure when the crossing intensities are infinite since these distributions are possible to obtain independent of the crossing rates.

4.2 Gaussian traffic models

In the following we shall consider a stationary Gaussian (normal) random process $\{\boldsymbol{B}_t\}$ with Mean value m and Standard deviation σ, and with autocorrelation function $\rho(t)$. For a given capacity level C we also let $A_t = \int_0^t (\boldsymbol{B}_\tau - C)d\tau$ be the excess volume. In the succeeding subsections we shall work with scaled variables defined by: $\boldsymbol{B}_t^* = \dfrac{\boldsymbol{B}_t - m}{\sigma}$ and $A_t^* = \dfrac{A_t}{\sigma} = \int_0^t (\boldsymbol{B}_\tau^* - C^*)d\tau$ where we also have introduced the scaled capacity by $C^* = \dfrac{C - m}{\sigma}$. We let also $\tilde{A}_t = \int_0^t \boldsymbol{B}_\tau^* d\tau = A_t^* + C^* t$ be the normalized arrived volume of the process. (Below we shall omit the $*$ but remember that through the rest of this chapter we are working with normalized variables as defined above.)

For Gaussian (normal) processes it is possible to relax the requirements on the autocorrelation function to secure that the process has continuous sample path [Lead83]. It is sufficient that the autocorrelation is bounded by $1 - \rho(t) \le at^\gamma$ for some $a > 0$ and $\gamma > 0$.

For a (standard) stationary Gaussian (normal) process it is well known that a necessary and sufficient condition for the up and down crossing intensity

$$\Delta_C = \lim_{t \to 0} \Delta_C^{uc}(t) = \lim_{t \to 0} \frac{\boldsymbol{P}\{\boldsymbol{B}_0 < C < \boldsymbol{B}_t\}}{t} \tag{4.3}$$

to exists and that the limit is finite, is that the autocorrelation takes the specific form:

$$\rho(t) = 1 - at^2 + o(t^2) \text{ as } t \to 0 \tag{4.4}$$

and moreover the up and down crossing intensity Δ_C is given by the famous Rice's formula [Lead83]:

$$\Delta_C = \frac{\sqrt{2a}}{2\pi} e^{-\frac{C^2}{2}} = \sqrt{\frac{a}{\pi}} \varphi(C) \tag{4.5}$$

where $\varphi(C) = \frac{1}{\sqrt{2\pi}} e^{-\frac{C^2}{2}}$ is the standard normal density function, and further we also denote $\phi(C) = \int_{t=C}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ as the standard normal integral.

It is possible to obtain these results quite easily by applying some of the properties for multinormal integrals given in Appendix B. By applying (B.63) we have that the partial derivative of the probability $I(2, C, C, \rho) = \boldsymbol{P}\{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\}$ with respect to ρ is:

$$\frac{\partial I}{\partial \rho}(2, C, C, \rho) = \frac{e^{-\frac{C^2}{1+\rho}}}{2\pi\sqrt{1-\rho^2}} \tag{4.6}$$

Integrating we find $\boldsymbol{P}\{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\} = \phi(C) + \int_{\xi=1}^{\rho} \frac{e^{-\frac{C^{*2}}{1+\xi}}}{2\pi\sqrt{1-\xi^2}} d\xi$ so we get:

$\Delta_C^{uc}(t) = \frac{\boldsymbol{P}\{\boldsymbol{B}_0 < C < \boldsymbol{B}_t\}}{t} = -\frac{1}{2\pi t} \int_{\xi=1}^{\rho} \frac{e^{-\frac{C^2}{1+\xi}}}{\sqrt{1-\xi^2}} d\xi$. Then for instance by applying l'Hopitals rule for limit of a fraction this gives:

$$\Delta_C = \lim_{t \to 0} \Delta_C^{uc}(t) = \frac{e^{-\frac{C^2}{2}}}{2\pi} \left\{ \lim_{t \to 0} \frac{-\rho'(t)}{\sqrt{2(1-\rho(t))}} \right\} \tag{4.7}$$

It is now obvious that to have the limit to exist we must have the autocorrelation on the form (4.4) and in that case we get (4.5).

We may now write down the mean excess times and excess volumes as found in chapter 3 by the equations (3.29), (3.30), (3.32) and (3.33) for a stationary Gaussian process:

$$\boldsymbol{E}[T_k] = \sqrt{\frac{\pi}{a}}\frac{\phi(C)}{\varphi(C)} \text{ and } \boldsymbol{E}[S_k] = \sqrt{\frac{\pi}{a}}\frac{1-\phi(C)}{\varphi(C)} \text{ and} \tag{4.8}$$

$$\boldsymbol{E}[A_k] = \sqrt{\frac{\pi}{a}}\left(1 - C\frac{\phi(C)}{\varphi(C)}\right) \text{ and } \boldsymbol{E}[V_k] = \sqrt{\frac{\pi}{a}}\left(\frac{C}{\varphi(C)} - \left(1 - C\frac{\phi(C)}{\varphi(C)}\right)\right) \tag{4.9}$$

where we also have used the following expressions for the integrals:

$$\int_C^{\infty} \phi(x)dx = \varphi(C) - C\phi(C) \text{ and } \int_{-\infty}^{C} (1-\phi(x))dx = C - (\varphi(C) - C\phi(C)).$$

We may also find the overall information loss p_{loss} given by (3.34) as:

$$p_{loss} = \varphi(C) - C\phi(C) \tag{4.10}$$

and the information loss in an overload period q_{loss} given by (3.35) as:

$$q_{loss} = \frac{\varphi(C)}{\phi(C)} - C \tag{4.11}$$

Note that the information losses do not depend on the parameter a but only on the capacity level C. Asymptotics for large values of C are easily found by using asymptotic expansion for the normal integral $\phi(C)$ for large C, cf. [Abra70] page 932 formula 26.2.10, giving:

$$\boldsymbol{E}[T_k] \sim \sqrt{\frac{\pi}{a}}\frac{1}{C} \text{ and } \boldsymbol{E}[A_k] \sim \sqrt{\frac{\pi}{a}}\frac{1}{C^2} \text{ and further} \tag{4.12}$$

$$p_{loss} \sim \frac{\varphi(C)}{C^2} \text{ and } q_{loss} \sim \frac{1}{C} \tag{4.13}$$

Thus, even though the overall information losses may be well limited, the losses in an overload period will be significant as shown by the asymptotics (4.13).

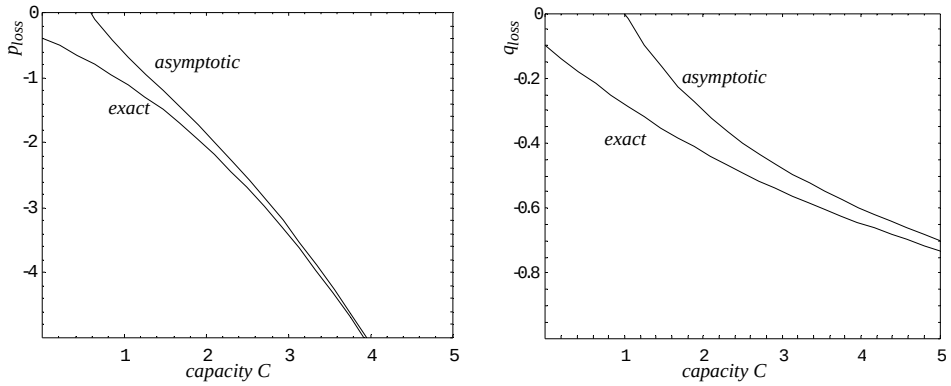

Figure 4.1: Logarithmic plot of the overall loss probability (left) and loss probability in overload periods (right) as function of the capacity.

In the figure 4.1 we have plotted the overall information loss probability based on equation (4.10) (left and marked "exact") and the loss probability in overload periods based on equation (4.11) (right and marked "exact") together with the asymptotics given by (4.13). If we for instance would like to keep the losses in the range 10^{-2}-10^{-4} this gives the interesting parameter value of the scaled capacity to be in the range 2.0-3.2.

4.3 The *n*-point approximation for Gaussian processes

Unless for some few cases it is difficult to obtain exact expressions for the excess probabilities involving the minimum $m_t = Inf_{\tau \in (0,t)} \boldsymbol{B}_\tau$ and maximum $M_t = Sup_{\tau \in (0,t)} \boldsymbol{B}_\tau$ of the process. In chapter 3 we saw that the 2^n-point approximation (obtained by continuously bisecting each interval) lead to approximations that were monotone and therefore had nice properties to secure convergence of the corresponding approximations to the "exact" probabilities. However, due to the difficulties to calculate multinormal integrals we shall apply the n-point approximation by dividing the interval into sub-intervals of equal lengths, (where we keep in mind that the n-point approximation coincides with the bisecting method for $n = 2, 4, 8$.)

More generally, and due to the property of a stationary Gaussian random process, we have that for every sequence of n succeeding points, say $t_1 = 0 < t_2 < \ldots < t_{n-1} < t_n = t$, that the ensemble $\{\boldsymbol{B}_{t_1}, \boldsymbol{B}_{t_2}, \ldots, \boldsymbol{B}_{t_n}\}$ is Multivariate Gaussian distributed with zero mean and Covariance matrix $M = (\rho_{ij})$ given by

$$\rho_{ij} = \boldsymbol{E}[\boldsymbol{B}_{t_i}\boldsymbol{B}_{t_j}] = \rho(|t_i - t_j|) \tag{4.14}$$

The joint distribution function for a Multivariate normal process is well known and is expressed in terms of the inverse $M^{-1} = (M_{ij}^{-1})$ of the covariance matrix M. (See (B.1) in Appendix B.) With the notation introduced in Appendix B (see equation (B.2)) we may express the n-point approximation of the excess probability as:

$$\psi_C^n(t) = I(n, \boldsymbol{C}, M) \tag{4.15}$$

where M is given by (4.14) and the vector giving the integration limits $\boldsymbol{C} = (C, ..., C)$ and where C is the scaled capacity. We shall apply the results derived in Appendix B to get more explicit expressions for the excess time distribution. By theorem B.3 in Appendix B (equation (B.65)) this integral may be written as:

$$\psi_C^n(t) = \phi(C)^n + \sum_{1 \le k < l \le n} \rho_{kl} \int_{\xi = 0}^{1} \frac{e^{-\frac{C^2}{1 + \xi\rho_{kl}}}}{2\pi\sqrt{1 - (\xi\rho_{kl})^2}} I(n-2, \boldsymbol{C}_\xi^{k,l}, M_\xi^{k,l}) d\xi \tag{4.16}$$

where $\boldsymbol{C}_\xi^{k,l}$ and $M_\xi^{k,l}$ is given as $\boldsymbol{C}^{k,l}$ and $M^{k,l}$ in theorem B.2 and corollary B.1 in Appendix B and are explicitly given below in (4.30) and (4.31) but with replacing ρ_{ij} with $\xi\rho_{ij}$ for all i, j, and further $\phi(x)$ is the standard normal integral.

The main achievements by writing the integral (4.15) of the form (4.16) is the fact that the dimension of the integral in the integrand is reduced to $n-2$ which means that the numbers of possible numerical integrations is reduced by 1. Continuing this process we obtain the remarkable result which shows that it is possible to calculate the n-dimensional multinormal integral by only performing $n/2$ (if n is even) or $n/2-1$ (if n is odd) successive numerical integrations. This very particular property of the multinormal integrals is widely applied and enable us to calculate multinormal probabilities of dimension seven by performing only three numerical integrations.

In the following we shall apply the n-point approximation for to find approximations of the various excess distributions defined in chapter 3. A natural choice of the points t_k would be to take them equally spaced, that is:

$$t_k = \frac{k-1}{n-1} t \text{ for } k = 1, \ldots, n \tag{4.17}$$

$$\text{which gives } \rho_{kl} = \rho(t_l - t_k) = \rho\left(\frac{l-k}{n-1} t\right) \text{ for } k < l. \tag{4.18}$$

4.3.1 *n*-point approximation of the distribution of excess times and first pas-

sage times for a stationary Gaussian process

We start by first finding the n-point approximation for the distribution of the first passage time T^x defined in (3.23). By applying the results for multinormal integrals (theorem B.1 and theorem B.2 in Appendix B) and conditional CDF (by conditioning on $\boldsymbol{B}_0 = x$) we get:

$$F^n_{T^x}(t) = -\frac{1}{f_1(x)}\frac{\partial}{\partial x}I(n, \boldsymbol{C}_x, M) = \frac{1}{f_1(x)}I^1(n, \boldsymbol{C}_x, M) = I(n-1, \boldsymbol{C}^1_x, M^1) \tag{4.19}$$

where $\boldsymbol{C}_x = (x, C, ..., C)$ and $f_1(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the (standard) normal density. By theorem B.2 ((B.57) and (B.58)) we also find the corresponding parameters for the conditional normal integral of dimension $n-1$ with limit vector $\boldsymbol{C}^1_x$ with elements:

$$c^1_i = \frac{C - x\rho_{1i}}{\sqrt{1-\rho_{1i}^2}} \text{ for } i = 2, ..., n \tag{4.20}$$

and M^1 is the conditional (symmetric) $(n-1)x(n-1)$ correlation matrix with elements:

$$\rho^1_{ij} = \frac{\rho_{ij} - \rho_{1i}\rho_{1j}}{\sqrt{1-\rho_{1i}^2}\sqrt{1-\rho_{1j}^2}} \text{ for } 2 \le i \le j \le n \tag{4.21}$$

We have $\rho_{1i} = \rho(t_i)$ implying that $\rho(t_i) \to 1$ when $t \to 0$, so for $x > C$ we have $\boldsymbol{C}^1_i \to -\infty$ for all $i = 2, ..., n$. Then by theorem B.8 (in section B.4 Appendix B) we get the desired result

$$\lim_{t\to 0} F^n_{T^x}(t) = \lim_{t\to 0} I(n-1, \boldsymbol{C}^1_x, M^1) = 1 \tag{4.22}$$

The corresponding result when $x = C$ is not that simple. We find $\boldsymbol{C}^1_i \to 0$ for all $i = 2, ..., n$. If the autocorrelation have the form (4.4) we find that all the $\rho^1_{ij} \to 1$ and we get $\lim_{t\to 0} F^n_{T^x}(t) = \phi(0) = \frac{1}{2}$. If the autocorrelation function $\rho(t)$ does not have the form (4.4), then the off diagonal elements of $M^1(0)$ are strictly smaller than unity and moreover a large part of the off diagonal elements will become small as the number of points increases. It is therefore reasonable to believe that $I(n-1, \mathbf{0}, M^1(0)) \to 0$ as $n \to \infty$. This example clearly demonstrates the difference in behaviour of the normal processes depending on the actual form of the correlation function for small values.

The 2-points and 3-points approximation yield:

$$F^2_{T^x}(t) = \phi\left(\frac{C - x\rho(t)}{\sqrt{1-\rho(t)^2}}\right) \tag{4.23}$$

$$F^3_{T^x}(t) = I\left(2, \frac{C - x\rho(t/2)}{\sqrt{1-\rho(t/2)^2}}, \frac{C - x\rho(t)}{\sqrt{1-\rho(t)^2}}, \rho(t/2)\sqrt{\frac{1-\rho(t)}{(1+\rho(t))(1-\rho(t/2)^2)}}\right) \tag{4.24}$$

The corresponding PDF may be found by differentiating the integral (4.19), giving:

$$f^n_{T^x}(t) = \frac{1}{f_1(x)}\frac{\partial^2}{\partial t \partial x} I(n, \boldsymbol{C}_x, M) = \frac{1}{f_1(x)}\frac{\partial}{\partial x}\left\{\frac{\partial I}{\partial t}(n, \boldsymbol{C}_x, M)\right\} \tag{4.25}$$

By first differentiating with respect to the time, applying corollary B.4 and theorem B.6 (Appendix B) we find the following expression for the n-point approximation of the PDF of the first passage time T^x:

$$f^n_{T^x}(t) = -\frac{1}{f_1(x)}\left\{\sum_{k=2}^{n}\frac{d\rho_{1k}}{dt}\frac{x - C\rho_{1k}}{1-\rho^2_{1k}} I^{1,k}(n, \boldsymbol{C}_x, M) + \sum_{2 \le k < l \le n}\left[\frac{d\rho_{kl}}{dt} - \frac{d\rho_{1k}}{dt}\frac{\rho_{1l} - \rho_{1k}\rho_{kl}}{1-\rho^2_{1k}} - \frac{d\rho_{1l}}{dt}\frac{\rho_{1k} - \rho_{1l}\rho_{kl}}{1-\rho^2_{1l}}\right] I^{1,k,l}(n, \boldsymbol{C}_x, M)\right\} \tag{4.26}$$

where the $I^{1,k}(n, \boldsymbol{C}_x, M)$ and $I^{1,k,l}(n, \boldsymbol{C}_x, M)$ are integrals of type (B.50) and may be written as products of bi- and tri-normal distributions and standard multinormal integrals of dimension $n-2$ and $n-3$ by applying conditional distributions as given in Appendix B by equation (B.51).

To find the n-point approximation of the excess time distribution we first examine the behaviour of $\psi^n_C(t)$ for small t. By the general results in chapter 3 equation (3.18) we find the following bounds for $\psi^n_C(t)$:

$$(n-1)\psi^2_C\left(\frac{t}{n-1}\right) - (n-2)\phi(C) \le \psi^n_C(t) \le \psi^2_C(t) \tag{4.27}$$

It follows now since $\psi^n_C(0) = \phi(C)$ that $\frac{\psi^2_C\left(\frac{t}{n-1}\right) - \psi^2_C(0)}{\frac{t}{n-1}} \le \frac{\psi^n_C(t) - \psi^n_C(0)}{t} \le \frac{\psi^2_C(t) - \psi^2_C(0)}{t}$. Then by letting $t \to 0$ in the last inequality we find that if $\Delta_C = \lim_{t \to 0} \Delta^{uc}_C(t) = -\psi^{2\prime}_C(0)$ is finite then $\Delta_C = -\psi^{2\prime}_C(0) = -\psi^{n\prime}_C(0)$. We therefore conclude that the n-point approximation will lead to the correct up crossing intensity for all values of n and we find the approximative distribution that may be defined as in chapter 3 by:

$$F^n_{T_k}(t) = \lim_{dt \to 0} \frac{\psi^n_C(t-dt) - \psi^n_C(t)}{\psi^n_C(0) - \psi^n_C(dt)} = \frac{-\psi^n_C{}'(t)}{\Delta_C} \tag{4.28}$$

We have $\psi^n_C{}'(t) = \sum_{1 \le k < l \le n} \frac{\partial I}{\partial \rho_{kl}}(n, \boldsymbol{C}, M) \frac{d}{dt}\rho_{kl}(t)$. By applying theorem B.2 and corollary B.3 (in Appendix B) we get:

$$\psi^n_C{}'(t) = \sum_{1 \le k < l \le n} \frac{d}{dt}\rho_{kl}(t) \frac{e^{-\frac{C^2}{1+\rho_{kl}}}}{2\pi\sqrt{1-\rho_{kl}{}^2}} I(n-2, \boldsymbol{C}^{k,l}, M^{k,l}) \text{ where} \tag{4.29}$$

and where the $\boldsymbol{C}^{k,l}$-vector is given by:

$$C^{k,l}_i = C \frac{(1+\rho_{kl}-\rho_{ki}-\rho_{li})}{\sqrt{\frac{(1+\rho_{kl})(1-\rho^2_{kl}-\rho^2_{ik}-\rho^2_{il}+2\rho_{kl}\rho_{ik}\rho_{il})}{1-\rho_{kl}}}} \text{ for } i = 1, ..., n,\ i \ne k, l \tag{4.30}$$

and the correlation matrix $M^{k,l}$ is given by:

$$\rho^{k,l}_{ij} = \frac{\rho_{ij}(1-\rho^2_{kl}) - \rho_{ik}\rho_{jk} - \rho_{il}\rho_{jl} + \rho_{kl}(\rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk})}{\sqrt{1-\rho^2_{kl}-\rho^2_{ik}-\rho^2_{il}+2\rho_{kl}\rho_{ik}\rho_{il}}\sqrt{1-\rho^2_{kl}-\rho^2_{jk}-\rho^2_{jl}+2\rho_{kl}\rho_{jk}\rho_{jl}}} \tag{4.31}$$

for $i, j = 1, ..., n,\ i, j \ne k, l$

The specific choice of the points t_k makes it possible to rewrite equation (4.29) by grouping factors for which $l - k = s$ and this somehow simplifies the expression for $\psi^n_C{}'(t)$:

$$\psi^n_C{}'(t) = \sum_{s=1}^{n-1} \frac{s}{n-1}\rho'(t_{s+1}) \frac{e^{-\frac{C^2}{1+\rho(t_{s+1})}}}{2\pi\sqrt{1-\rho(t_{s+1})^2}} \left\{ \sum_{k=1}^{n-s} I(n-2, \boldsymbol{C}^{k,k+s}, M^{k,k+s}) \right\} \tag{4.32}$$

(4.32) represents the limit of how far it is possible to analyse the n-point approximation for general correlation function. The main difficulty to get a proper excess distribution is the behaviour of the approximation for small t due to the square root in the denominator. To examine the behaviour for small t we shall examine the two lowest approximation separately; namely the case $n = 2$ and $n = 3$. We have:

$$\psi^2_C{}'(t) = \rho'(t) \frac{e^{-\frac{C^2}{1+\rho(t)}}}{2\pi\sqrt{1-\rho(t)^2}} \text{ and} \tag{4.33}$$

$$\psi_C^3{}'(t) = \rho'(t)\frac{e^{-\frac{C^2}{1+\rho(t)}}}{2\pi\sqrt{1-\rho(t)^2}}\phi\left(C\left(\frac{1+\rho(t)-2\rho(t/2)}{\sqrt{\frac{1+\rho(t)}{1-\rho(t)}D^3(t)}}\right)\right) + \rho'(t/2)\frac{e^{-\frac{C^2}{1+\rho(t/2)}}}{2\pi\sqrt{1-\rho(t/2)^2}}\phi\left(C\left(\frac{1-\rho(t)}{\sqrt{\frac{1+\rho(t/2)}{1-\rho(t/2)}D^3(t)}}\right)\right) \quad (4.34)$$

where $D^3(t)$ is the determinant of the $3x3$ covariance matrix:

$$D^3(t) = (1-\rho(t))(1+\rho(t)-2\rho(t/2)^2) \quad (4.35)$$

An alternative to consider the excess time distribution above we may consider the *n*-point approximation of the CDF of the first passage time T_C given that the rate process is above the level C (defined by (3.27)). By applying (4.19) we find:

$$F_{T_C}^n(t) = -\frac{1}{\phi(C)}\int_{x=C}^{\infty}\frac{\partial}{\partial x}I(n, \boldsymbol{C}_x, M)dx = \frac{\psi_C^n(t)}{\phi(C)} \quad (4.36)$$

where $\psi_C^n(t) = I(n, \boldsymbol{C}, M)$ is the *n*-point approximation of the excess probability given as multinormal integral of dimension n. This is fully in accordance with the corresponding definition in chapter 3 (given by equation (3.27)). The corresponding density is then given by

$$f_{T_C}^n(t) = -\frac{\psi_C^n{}'(t)}{\phi(C)} \quad (4.37)$$

where the derivative $\psi_C^n{}'(t)$ is given by (4.32).

4.3.2 Joint *n*-point approximation of the distribution of excess times and excess volumes and first passage times and the corresponding volume for a stationary Gaussian process

Generally, it is not difficult to incorporate the volume in the analysis because adding the volume will not alter the fact that we are dealing with multivariate normal distributions. Therefore the volume will just add an extra dimension in the analysis, and $\{\boldsymbol{B}_{t_1}, \boldsymbol{B}_{t_2}, \ldots, \boldsymbol{B}_{t_n}, \tilde{A}_t\}$ (where $\tilde{A}_t = \int_0^t \boldsymbol{B}_\tau d\tau$) also will be Multivariate normally distributed with zero mean. We denote the Covariance matrix $\tilde{M} = (\tilde{\rho}_{ij})$ for this ensemble and we find:

$$\tilde{\rho}_{ij} = \boldsymbol{E}[\boldsymbol{B}_{t_i}\boldsymbol{B}_{t_j}] = \rho(|t_i - t_j|) \quad i, j = 1, \ldots, n \text{ and} \quad (4.38)$$

$$\tilde{\rho}_{n+1j} = \tilde{M}_{jn+1} = \boldsymbol{E}[\boldsymbol{B}_{t_j}\tilde{A}_t] = \alpha(t-t_i) + \alpha(t_i) \quad j = 1, \ldots, n \quad \text{with}$$

$$\alpha(t) = \int_{\tau=0}^{t} \rho(\tau)d\tau \text{ and} \tag{4.39}$$

$$\tilde{\rho}_{n+1n+1} = \boldsymbol{E}[\tilde{A}_t^{\,2}] = \beta(t) \text{ with } \beta(t) = 2\int_{\tau=0}^{t} (t-\tau)\rho(\tau)d\tau \tag{4.40}$$

To find the covariance matrix on standard form we scale the volume by its standard deviation $\sqrt{\beta(t)}$ by defining $A_t^s = \dfrac{A_t + Ct}{\sqrt{\beta(t)}}$. With this scaling we get $\left\{\boldsymbol{B}_{t_1}, \boldsymbol{B}_{t_2}, \ldots, \boldsymbol{B}_{t_n}, A_t^s\right\}$ as standard multinormal distributed of dimension $n+1$ with correlation matrix $M^A = (\rho_{ij}^A)$ given by:

$$\rho_{ij}^A = \rho(|t_i - t_j|), \; i, j = 1, \ldots, n \tag{4.41}$$

$$\rho_{n+1i}^A = \rho_{in+1}^A = \gamma_i(t), \text{ with } \gamma_i(t) = \frac{\alpha(t_n - t_i) + \alpha(t_i)}{\sqrt{\beta(t)}} \quad i = 1, \ldots, n \tag{4.42}$$

(and $\rho_{n+1n+1}^A = 1$)

By integrating over all chosen points $t_1 = 0 < t_2 < \ldots < t_{n-1} < t_n = t$ we may write the n-point approximation of the excess probability $G_C^n(x, y, z, t) = \boldsymbol{P}\{\boldsymbol{B}_0 > x, \boldsymbol{B}_{t_2} > C, ..., \boldsymbol{B}_{t_{n-1}} > C, \boldsymbol{B}_t > y, A_t > z\}$ suggested by (3.71) as a standard normal integral of dimension $n+1$:

$$G_C^n(x, y, z, t) = I(n+1, \boldsymbol{C}_{x,y,z,t}^A, M^A) \tag{4.43}$$

where the limit vector $\boldsymbol{C}_{x,y,z,t}^A$ is given by:

$$C_1^A = x, \; C_2^A = ... = C_{n-1}^A = C, \; C_n^A = y \text{ and } C_{n+1}^A = \frac{z + Ct}{\sqrt{\beta(t)}} \tag{4.44}$$

and the covariance matrix M^A is defined by (4.41) and (4.42). By (4.43) we have linked the n-point approximations for the excess probability to the standard normal integral of dimension $n+1$ and we shall show below that it is possible to obtain an expression for the corresponding joint excess distribution defined by (3.52) and (3.62). Some informally we then have the corresponding PDF as:

$$g_C^n(x, y, z, t) = -\frac{\partial^3}{\partial x \partial y \partial z} I(n+1, \boldsymbol{C}_{x, y, z, t}^A, M^A) = \frac{1}{\sqrt{\beta(t)}} I^{1, n, n+1}(n+1, \boldsymbol{C}_{x, y, z, t}^A, M^A) \tag{4.45}$$

It is possible to write equation (4.45) as a product of tri-normal distribution and a standard normal integral of dimension $n-2$ by applying conditional distributions as given in Appendix B by equation (B.51).

By applying (4.45) as a staring point we may now find the n-point approximation for the joint CDF of the first passage time and the corresponding volume by (3.52). By introducing the different types of integrals defined in Appendix B it is possible to write the integrals $\int_{y=C}^{\infty} g_C^n(x, y, z, t)dy = \frac{\partial^2}{\partial x \partial z} I(n+1, \boldsymbol{C}_{x, z, t}^A, M^A)$ and $\int_{y=C}^{\infty} (y-C) g_C^n(x, y, z, t)dy = \frac{\partial^2}{\partial x \partial z} J_n(n+1, \boldsymbol{C}_{x, z, t}^A, M^A)$. By applying these expressions we may write the joint PDF in the following way:

$$f_{T^x A^x}^n(x, z, t) = -\frac{1}{\varphi(x)} \frac{\partial^2}{\partial x \partial z} (\boldsymbol{H}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A)) \text{ where} \tag{4.46}$$

$$\boldsymbol{H}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) = \frac{\partial}{\partial z} J_n(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) + \frac{\partial}{\partial t} I(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) \tag{4.47}$$

and where we have redefined the limit vector $\boldsymbol{C}_{x, z, t}^A$ by: $C_1^A = x$, $C_2^A = ... = C_n^A = C$ and $C_{n+1}^A = \frac{z + Ct}{\sqrt{\beta(t)}}$ and where the integral of type $J_i(n, \boldsymbol{C}, M)$ is defined by (B.79) in (Appendix B). If we now let $\tilde{f}_{T^x A^x}^n(x, z, t) = \int_{\zeta=z}^{\infty} f_{T^x A^x}^n(x, \zeta, t)d\zeta$ then we readily get by integrating over z-variable in (4.46):

$$\tilde{f}_{T^x A^x}^n(x, z, t) = \frac{1}{\varphi(x)} \frac{\partial}{\partial x} (\boldsymbol{H}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A)) \tag{4.48}$$

To obtain more explicit expression of the n-point approximations which are suitable for numerical calculations we must perform the differentiations in (4.46) and (4.48). This may be done by applying the "algebra of multi-normal integrals" developed in Appendix B through the equations (B.79) to (B.88). We find:

$$\boldsymbol{H}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) = \sum_{1 \le k < l \le n} I^{k, l}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) \frac{d\rho_{kl}}{dt} + \sum_{k=2}^{n} I^{k, n+1}(n+1, \boldsymbol{C}_{x, z, t}^A, M^A) \frac{\rho_{1k} - \rho_{nk}}{\sqrt{\beta}} t_k' \tag{4.49}$$

By differentiating this expression with respect to x and z enable us to find explicit formulae for the joint PDF in terms of multinormal integrals. We get:

$$f^{n}_{T^xA^x}(x,z,t) = -\frac{1}{f_1(x)\sqrt{\beta}}\left\{\sum_{k=2}^{n}\left\{D_1^{1,k,n+1}\left(x,C,\frac{z+C^*t}{\sqrt{\beta}}\right)\frac{d\rho_{1k}}{dt}+D_3^{1,k,n+1}\left(x,C,\frac{z+C^*t}{\sqrt{\beta}}\right)\frac{\rho_{1k}-\rho_{nk}}{\sqrt{\beta}}t_k'\right\}I^{1,k,n+1}(n+1,\boldsymbol{C}^A_{x,z,t},M^A)\right.$$

$$+\sum_{2\le k<l\le n}\left\{\frac{d\rho_{kl}}{dt}-D_1^{1,k,n+1}(\rho_{1l},\rho_{kl},\gamma_l)\frac{d\rho_{1k}}{dt}-D_1^{1,l,n+1}(\rho_{1k},\rho_{kl},\gamma_k)\frac{d\rho_{1l}}{dt}\right. \tag{4.50}$$

$$\left.\left.-D_3^{1,k,n+1}(\rho_{1l},\rho_{kl},\gamma_l)\frac{\rho_{1k}-\rho_{nk}}{\sqrt{\beta}}t_k'-D_3^{1,l,n+1}(\rho_{1k},\rho_{kl},\gamma_k)\frac{\rho_{1l}-\rho_{nl}}{\sqrt{\beta}}t_l'\right\}I^{1,k,l,n+1}(n+1,\boldsymbol{C}^A_{x,z,t},M^A)\right\}$$

where we also have defined the auxiliary functions (defined by 3x3 determinants):

$$D_1^{1,k,n+1}(a,b,c) = \frac{a(1-\gamma_1^2)+b(\gamma_1\gamma_k-\rho_{1k})+c(\rho_{1k}\gamma_k-\gamma_1)}{1-\rho_{1k}^2-\gamma_1^2-\gamma_k^2+2\rho_{1k}\gamma_1\gamma_k} \text{ and} \tag{4.51}$$

$$D_3^{1,k,n+1}(a,b,c) = \frac{a(\rho_{1k}\gamma_k-\gamma_1)+b(\gamma_1\rho_{1k}-\gamma_k)+c(1-\rho_{1k}^2)}{1-\rho_{1k}^2-\gamma_1^2-\gamma_k^2+2\rho_{1k}\gamma_1\gamma_k} \tag{4.52}$$

We may also get the corresponding expansion for the function $\tilde{f}^{n}_{T^xA^x}(x,z,t)$ (which is not as complex as the joint density function above):

$$\tilde{f}^{n}_{T^xA^x}(x,z,t) = -\frac{1}{f_1(x)}\left\{\sum_{k=2}^{n}\frac{d\rho_{1k}}{dt}\frac{x-C^*\rho_{1k}}{1-\rho_{1k}^2}I^{1,k}(n+1,\boldsymbol{C}^A_{z,t},M^A)+\sum_{k=2}^{n}\left[\frac{\rho_{1k}-\rho_{nk}}{\sqrt{\beta}}t_k'-\frac{\gamma_1-\rho_{1k}\gamma_k}{1-\rho_{1k}^2}\right]I^{1,k,n+1}(n+1,\boldsymbol{C}^A_{z,t},M^A)\right.$$

$$\left.+\sum_{2\le k<l\le n}\left[\frac{d\rho_{kl}}{dt}-\frac{d\rho_{1k}}{dt}\frac{\rho_{1l}-\rho_{1k}\rho_{kl}}{1-\rho_{1k}^2}-\frac{d\rho_{1l}}{dt}\frac{\rho_{1k}-\rho_{1l}\rho_{kl}}{1-\rho_{1l}^2}\right]I^{1,k,l}(n+1,\boldsymbol{C}^A_{z,t},M^A)\right\} \tag{4.53}$$

Similar as for the n-point approximation of the joint CDF of the first passage time and the corresponding volume we get the n-point approximation of the joint CDF of the excess time and excess volume by (3.62). By applying the different types of integrals defined in Appendix B it is possible to express joint PDF as follows:

$$f^{n}_{A_kT_k}(z,t) = \frac{-1}{\Delta_C}\frac{\partial}{\partial z}(\boldsymbol{G}(n+1,\boldsymbol{C}^A_{z,t},M^A)) \text{ where} \tag{4.54}$$

$$\boldsymbol{G}(n+1,\boldsymbol{C}^A_{z,t},M^A) = \frac{\partial^2 I}{\partial t^2}(n+1,\boldsymbol{C}^A_{z,t},M^A)+\frac{\partial^2}{\partial z\partial t}(J_1(n+1,\boldsymbol{C}^A_{z,t},M^A)+J_n(n+1,\boldsymbol{C}^A_{z,t},M^A))+\frac{\partial^2}{\partial z^2}J_{1,n}(n+1,\boldsymbol{C}^A_{z,t},M^A) \tag{4.55}$$

and where the integrals of type $J_i(n,\boldsymbol{C},M)$ and $J_{i,j}(n,\boldsymbol{C},M)$ are defined by (B.79) and (B.80) in (Appendix B), and where we have taken the limit vector $\boldsymbol{C}^A_{z,t}$ as: $C_i^A = C$

$i = 1, \ldots, n$, and $C^A_{n+1} = \dfrac{z + Ct}{\sqrt{\beta(t)}}$. If we (as above) let $\tilde{f}^n_{T_k A_k}(x, z, t) = \int_{\zeta = z}^{\infty} f^n_{T_k A_k}(x, \zeta, t) d\zeta$ then we readily get by integrating over z-variable in (4.54):

$$\tilde{f}^n_{T_k A_k}(x, z, t) = \frac{1}{\Delta_C}(\boldsymbol{G}(n+1, \boldsymbol{C}^A_{z,t}, M^A)) \tag{4.56}$$

By the results obtained in Appendix B we have found the rules that relate integrals of type $J_i(n, \boldsymbol{C}, M)$ and $J_{i,j}(n, \boldsymbol{C}, M)$ to sums of integrals of the standard multinormal of type $I(n, \boldsymbol{C}, M)$ (given by the equations (B.81) and (B.82)). By the general results found for the derivative with respect to both elements in the limit vectors $\boldsymbol{C}$ (given by (B.86)) and with respect to the elements in the correlation matrix M (given by (B.88)) enable relating all the derivatives in the expressions above to integrals of type $I(n, \boldsymbol{C}, M)$. Performing the differentiation we can therefore write the n-point approximation (4.54) and (4.56) as a sum where only integrals of type $I(n, \boldsymbol{C}, M)$ are included. (It is necessary to define $I(n, \boldsymbol{C}, M) = 1$ for $n = 0$, and $I(n, \boldsymbol{C}, M) = 0$ for $n = -1, -2, \ldots,$.)

4.3.3 Some numerical examples

One of the main intension by the developments in the previous subsections were to find approximative methods to obtain the transient characteristics (described in section 3) for a general Gaussian process by applying the properties of multinormal integrals given in Appendix B. As the "test" case we have chosen the autocorrelation function to be on the form:

$$\rho(t) = e^{-a|t|} \tag{4.57}$$

for which it is well known that the corresponding process is the Ornstein-Uhlenbeck (O-U) process which turns out to be the only stationary Gaussian process that is Markovian [Fell68b]. For the O-U process the exact distribution functions are given by (4.89) (for the first passage time given that the process is in an excess period) and (4.82) (for the conditional first passage time). The corresponding n-point approximations are calculated by (4.36) and (4.19) respectively. For the conditional first passage time (4.36) the approximation is given as a multinormal integral of dimension $n-1$ which gives us the possibility to calculate the approximative distribution function for $n = 2, 3, 4, 5, 6$ by the method described in Appendix B. For the first passage time given that the process is in an excess period (4.19), the approximation is a multinormal integral of dimension n allowing us to calculate the approximative distribution function for $n = 2, 3, 4, 5$.

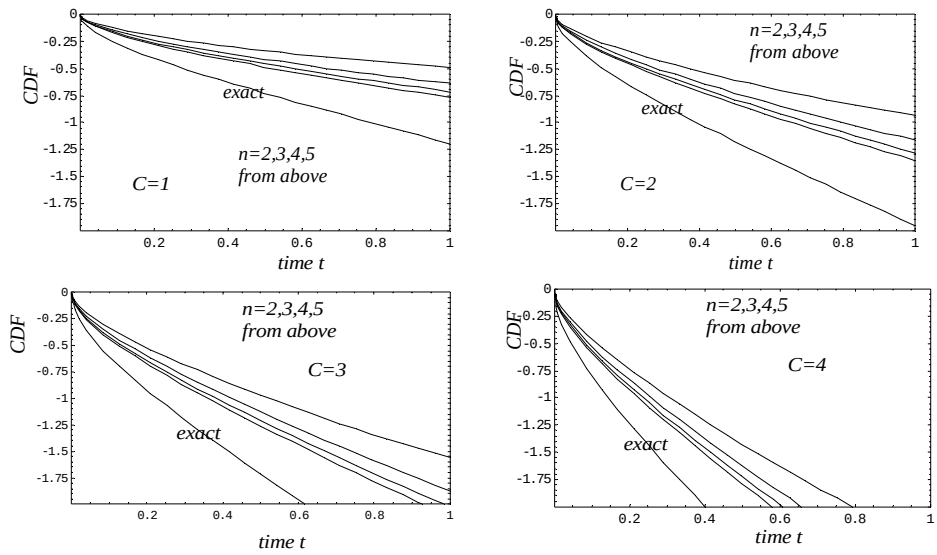

Figure 4.2: Logarithmic plot of the CDF of the first passage time given that the process is in an excess period for exact and different approximations (n=2,3,4,5) and different scaled capacities as function of time.

In figure 4.2 we have given plots of the exact and approximative CDFs of the first passage times given that the process is in an overload (excess) period. These CDFs will give an overall impression of the length of a congestion period for the chosen (scaled) capacity level. The main observation is that the convergence of the approximative solution is quite slow also for small values of the time where we had hoped that the accordance would have been better, however, the actual form of the curves are similar. It seems that to get a tight approximation one will need a rather huge number of points that will lead to multinormal integrals of dimension that are impossible to calculate by the methods described in Appendix B. Another observation is that the relative difference seems to be independent of the (scaled) capacity levels, however it is clear that larger capacities will lead to excess periods that are significant smaller than for lower capacities.

The results for the conditional CDFs given in figures 4.3-4.5 are quite similar. We observe however, that the accuracy will depend on the difference between the starting value of the process and the capacity level. This difference is taken to be 0.1 in figure 4.3, 0.3 in figure 4.4 and 0.5 in figure 4.5, and it is clear that the approximations are more accurate as this difference increases. For the difference equal to 0.1 (figure 4.3) the approximations must say to perform rather badly.

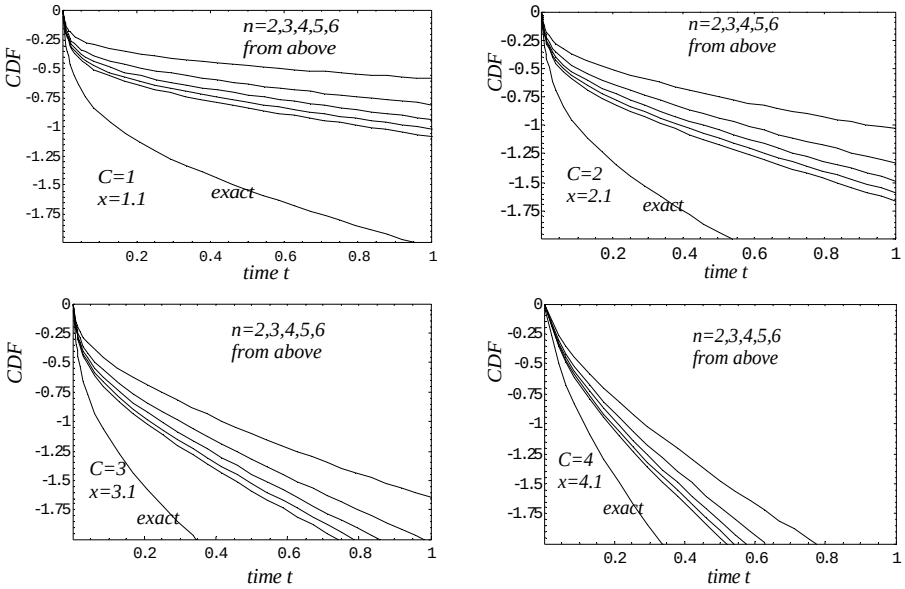

Figure 4.3: Logarithmic plot of the CDF of the conditional first passage time for exact and different approximations (n=2,3,4,5,6), different scaled capacities and different starting values as function of time.

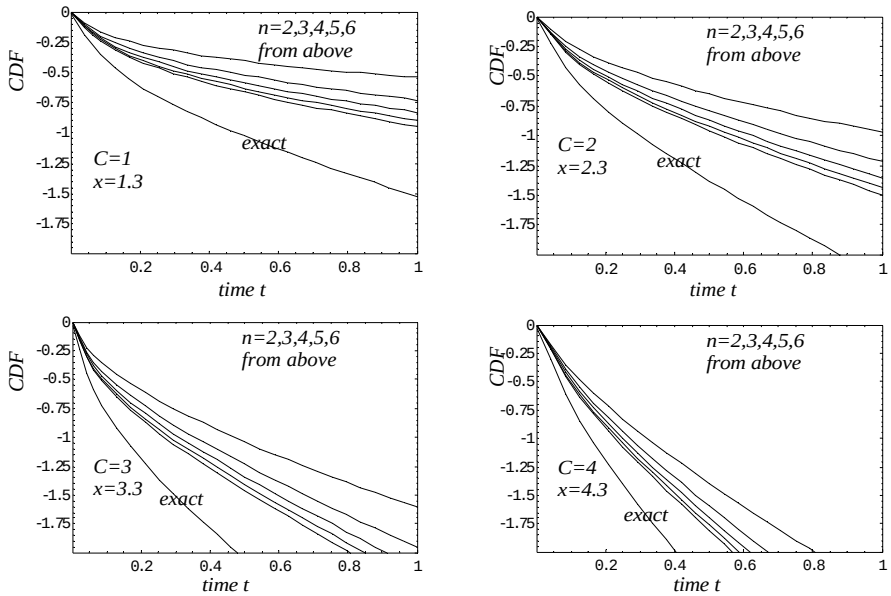

Figure 4.4: Logarithmic plot of the CDF of the conditional first passage time for exact and different approximations (n=2,3,4,5,6), different scaled capacities and different starting values as function of time.

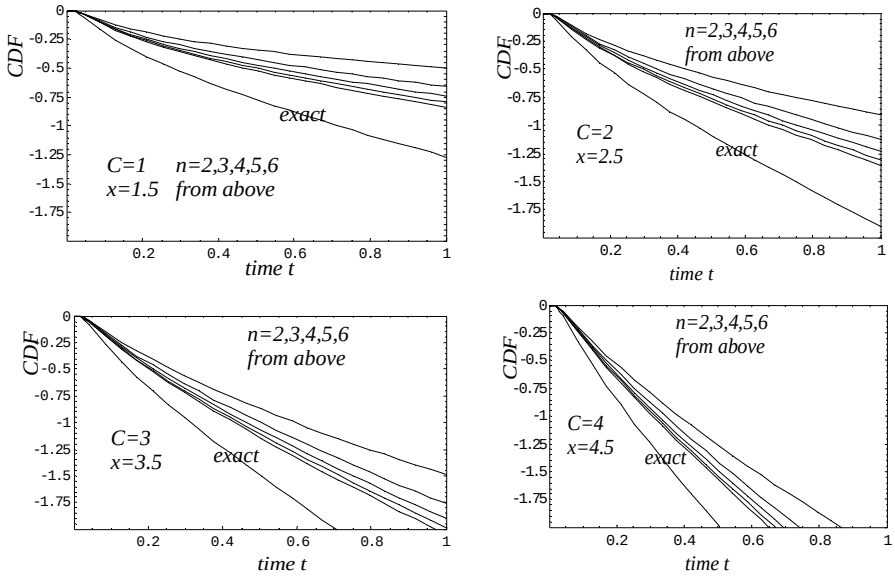

Figure 4.5: Logarithmic plot of the CDF of the conditional first passage time for exact and different approximations (n=2,3,4,5,6), different scaled capacities and different starting values as function of time.

As a common remark on the discussion of the approximations and, despite the disappointment of the accuracy, we conclude that we have found approximations that will provide upper bounds for the exact CDFs and will have the same form, but will over-estimate the lengths of the excess periods to some extent.

As a second example we have chosen a Gaussian process which exhibits long-rage dependence with autocorrelation on the form:

$$\rho(t) = \frac{1}{1 + (at)^{2-2H}} \text{ for } 1/2 < H < 1 \tag{4.58}$$

where we scale the time and take $a = 1$ in the numerical examples below. We have $\rho(t) \sim ct^{2H-2}$ with $c = a^{2-2H}$ as $t \to \infty$ for $1/2 < H < 1$ where H is the Hurst-parameter describing the degree of self-similarity in the process.

In figure 4.6 we have plotted the curves for the conditional CDFs of the excess time (first passage) for the different approximations (n=2, 3, 4, 5, 6) to view the "speed" of convergence for the case with autocorrelation given by (4.58) (i.e. with long range dependence). In the left figure we have chosen that scaled capacity to be $C = 1$ and the starting value quite close to C with value x=1.1, where as in the right figure the corresponding parameters are $C = 3$ and x=3.5. In both cases the Hurst parameter is set to H=0.7. The rate of "convergence" for this case looks very similar to what was observed for the O-U process, (figures

4.3-4.5 above). It seems that the "convergence" gets worse when the process starts out very close to the excess level C. (And this is also what is expected from the way this type of approximation is constructed.) Nevertheless, we are quite confident that the given approximation will provide valuable insight to get the typical form of the CDFs, and it should be worthwhile to conduct calculations using the largest n-value (5 or 6) for a broader set of parameter values.

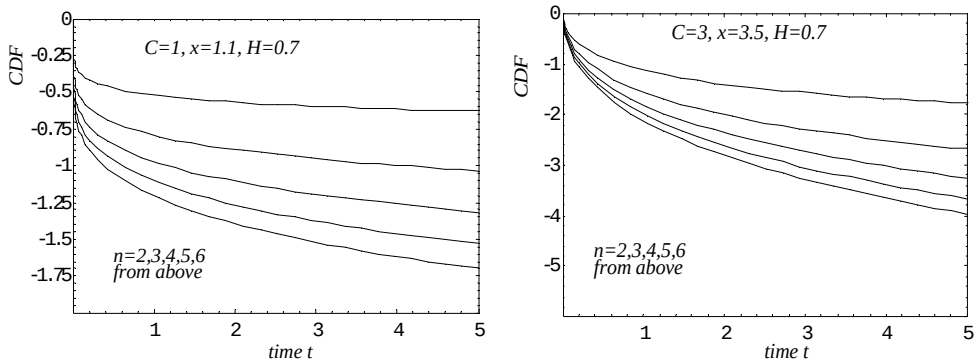

Figure 4.6: Logarithmic plot of the CDF of the conditional first passage time based on different approximations (n=2,3,4,5,6), and scaled capacity C=1 and starting value x=1.1 (left), and scaled capacity C=3 and starting value x=3.5 (right) for a process with long-range dependence with H=0.7 as function of time.

In figure 4.7 we have given plots of the approximative CDFs of the first passage times by using $n = 5$ intervals, given that the process starts in an overload (excess) period for some different capacity levels and some different Hurst parameters. The striking evidence in figure 4.7 is that all the curves having equal H seems nearly to have the same form plotted in a logarithmic scale (the only difference is the scaling of the y-axis.) Written out mathematically this seems to imply that for two capacity levels C_1 and C_2 the follow relation holds $\frac{\log P(T^5_{C_2} > t)}{\log P(T^5_{C_1} > t)} \approx \frac{C_2}{C_1}$. (This formula should not be thought of as yielding for all values of C but only in the range from 1 to 4 which is observed in the figures.) If, for instance, the given distribution is decreasing as power law for large values of t that is, $P(T^5_C > t) \sim ct^{-\alpha_C}$, for some positive constants c and α_C, then the observed relations will make implications on the relations between the exponents by $\frac{\alpha_{C_2}}{\alpha_{C_1}} \approx \frac{C_2}{C_1}$.

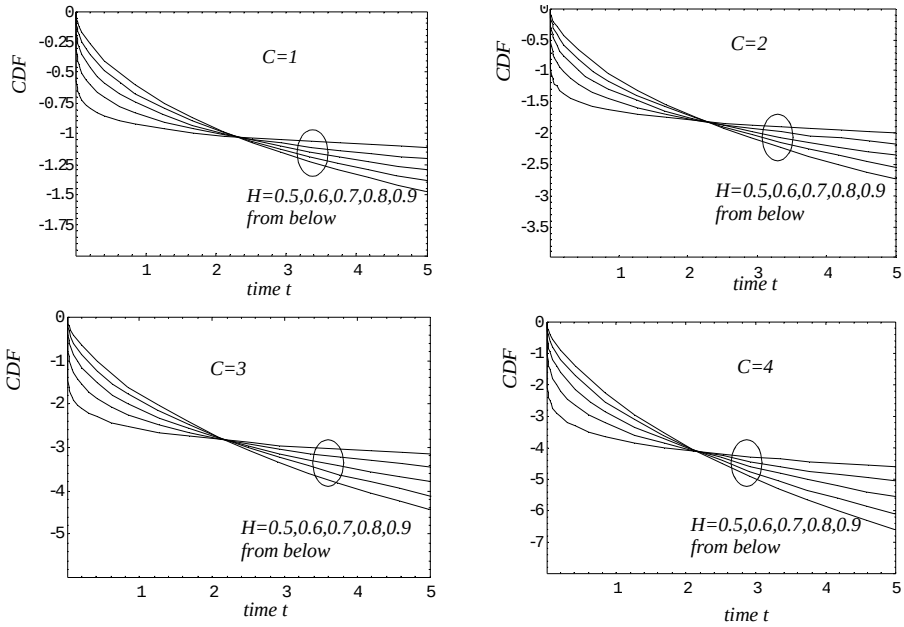

Figure 4.7: Logarithmic plot of the CDF of the first passage time given that the process is in an excess period based on approximation with n=5 intervals and different scaled capacities as function of time for a process with long-range dependence.

In figures 4.8-4.10 we have given plots of the approximative CDFs of the conditional first passage times with $n = 6$ intervals, for some different capacity levels with some different chosen starting values for the process and some different Hurst parameters. Although the curves look very similar to the case above there are some differences that should be mentioned:

- The scaling formulae found above are not so accurate especially for the case where the starting value x is close to the capacity level C. This is seen in figure 4.8 where $x - C = 0.1$.
- The curves for $H = 0.9$ and $x - C = 0.1$ $C = 1, 2, 3, 4$) are very flat.
- The intersection point which in figure 4.7 is located at $t \approx 2.3$ has move outwards to around $t \approx 3$ with some minor variations.

We may conclude these examples by mentioning that the obtained distributions seem to have "heavy" tails at least for the range of time parameter up to $t = 5$, and we have demonstrated that the form of the excess time distributions will heavily depend on the Hurst-parameter.

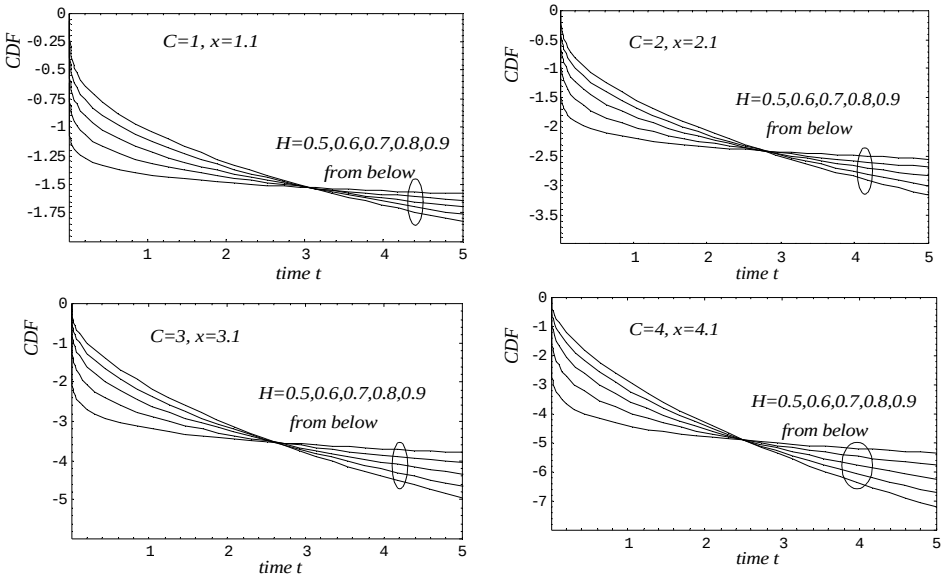

Figure 4.8: Logarithmic plot of the CDF of the conditional first passage time based on approximation with n=6 intervals and different scaled capacities and different starting values as function of time for a process with long-range dependence.

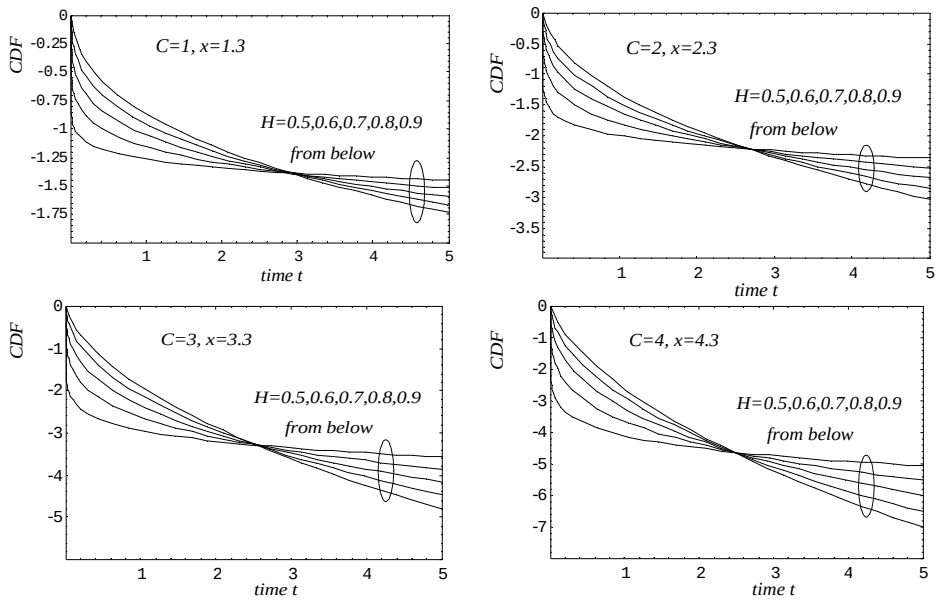

Figure 4.9: Logarithmic plot of the CDF of the conditional first passage time based on approximation with n=6 intervals and different scaled capacities and different starting values as function of time for a process with long-range dependence.

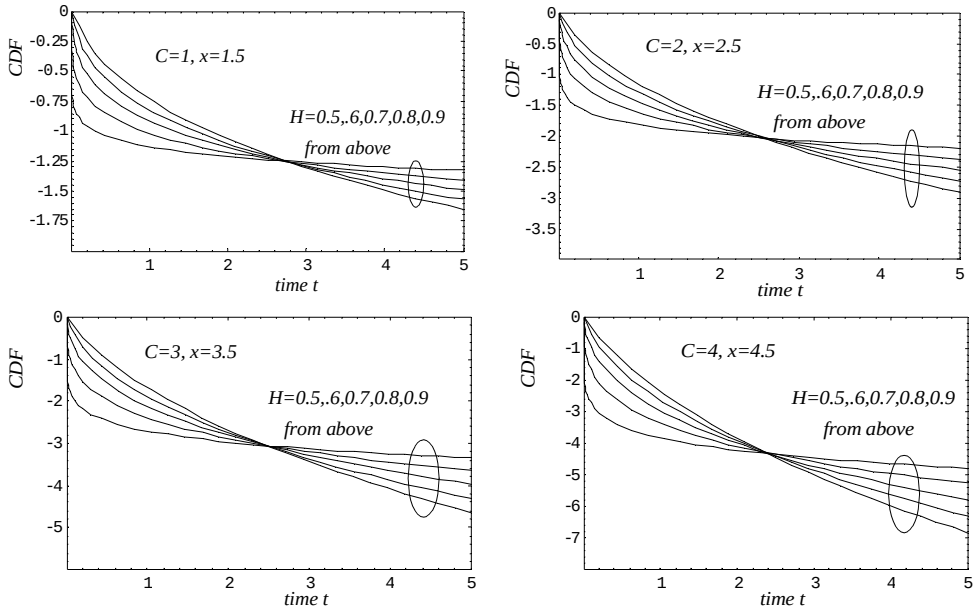

Figure 4.10: Logarithmic plot of the CDF of the conditional first passage time based on approximation with n=6 intervals and different scaled capacities and different starting values as function of time for a process with long-range dependence.

4.4 Distribution of the first passage times and the corresponding volumes for the Ornstein-Uhlenbeck process

In this section we shall consider the classical diffusion process known as the Ornstein-Uhlenbeck (O-U) process. It is well known [Fell68b] that this is the only stationary Gaussian (normal) process that is Markovian. We shall consider the O-U process in more depth, not only for the sake of its “famous” properties that are well known in the literature, but rather to analyse this process as an example where it is possible to obtain exact results (for the first passage times and the corresponding volumes) and use this particular process to test the approximations proposed in the preceding sections. It is well known that the O-U process is a diffusion process and that the free space properties may be found by from the solution of the corresponding diffusion equation. Thus, the excess probabilities will satisfy the same diffusion equation as the free space probabilities, but where the special requirements for the excess probabilities are expressed through extra boundary conditions. Such methods have been applied by Hagan et al. in the paper [Haga89].

Secondly, the studying of the O-U process may also be motivated by the fact that this process may be obtained as the limiting behaviour of a large numbers of on/off sources (with exponentially distributed on- and off-times) in the heavy traffic regime. As pointed out in [Knes91] the asymptotics (leading to the O-U process) are obtained by assuming:

$$j_C - A = O(\sqrt{N}) \text{ as } N \to \infty \tag{4.59}$$

where $A = N\frac{\lambda}{\lambda + \mu}$ is the offered traffic and N the number of sources; μ^{-1} and λ^{-1} are the mean on- and off-times and j_C is capacity scaled by the peak bit rate for a source. As the number of sources increases, then by (4.59) the load is:

$$r = \frac{A}{j_C} \sim 1 - O\left(\frac{1}{\sqrt{N}}\right) \text{ as } N \to \infty \tag{4.60}$$

The same asymptotic regime will also apply for an M/M/∞ queueing system as pointed out in [Guill96] (A is then the total input rate to the system.)

In the succeeding we shall also work with scaled (normalized) variables defined in the introductory part of this chapter, and we consider the pair $\{\boldsymbol{B}_t, A_t\}$ where the evaluation given $\{\boldsymbol{B}_0, A_0\}$ is described by the differential system

$$d\boldsymbol{B}_t = -a\boldsymbol{B}_t dt + W_t \text{ and} \tag{4.61}$$

$$dA_t = (\boldsymbol{B}_t - C)dt \tag{4.62}$$

where W_t denotes the standard Wiener process. It is well known that the pair $\{\boldsymbol{B}_t, A_t\}$ given by (4.61) and (4.62) constitutes a pair of normal stochastic variables where $\{\boldsymbol{B}_t\}$ is an Ornstein-Uhlenbeck process with correlation function:

$$\rho(t) = e^{-a|t|} \tag{4.63}$$

Since $\rho(t) = 1 - a|t| + o(t))$ for small t (which is not on the form (4.4)) and it follows that the up and down crossing intensities do not exist (are infinite). This is due to the irregular elapse of the sample paths of this process, which, in spite of being continuous everywhere, are not differentiable in any point, (see for instance [Cox70]).

Due to the specific form of the autocorrelation function we shall also scale the time according to $t' = at$ (and in the succeeding we drop the marks).

4.4.1 First passage time distribution for the Ornstein-Uhlenbeck process

We shall start by finding the first passage time for the Ornstein-Uhlenbeck process. The Laplace transform of the first passage time is known, but we shall develop along a line for which it is quite easy to also include the excess volume in the analysis. The function of interested to determine the first passage time is the conditional probability:

$$F_C(x, y, t) = \boldsymbol{P}\{\boldsymbol{B}_t > y, Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = x\} \tag{4.64}$$

and we define the PDF $f_C(x, y, t) = -\frac{\partial}{\partial y}F_C(x, y, t)$ and also the excess function $\psi_C(t) = \boldsymbol{P}\{Inf_{\tau \in (0, t)}\boldsymbol{B}_\tau > C\}$ where we have $\psi_C(t) = \int_{x=C}^{\infty}\int_{y=C}^{\infty} \varphi(x)f_C(x, y, t)dydx$. We may now write the Fokker-Plank equation for the PDF $f_C(x, y, t)$ together with the appropriate initial and boundary conditions:

$$\frac{\partial^2 f_C}{\partial y^2} + \frac{\partial}{\partial y}[yf_C] = \frac{\partial f_C}{\partial t} \text{ for } x > C,\ y > C \text{ and } t > 0 \tag{4.65}$$

$$f_C(x, y, 0) = \delta(y - x) \text{ for } x > C,\ y > C \text{ and} \tag{4.66}$$

$$f_C(x, C, t) = 0 \text{ for } x > C \text{ and } t > 0 \tag{4.67}$$

Given the govern equation (4.65), the task will then be to solve the diffusion equation above with the given initial and boundary equations. To get rid of the initial condition it is convenient to introduce the LST:

$$\hat{f}_C(x, y, s) = \int_{t=0}^{\infty} e^{-st}f_C(x, y, t)dt \text{ giving} \tag{4.68}$$

$$\frac{\partial^2}{\partial y^2}\hat{f}_C + \frac{\partial}{\partial y}\hat{f}_C + (1 - s)\hat{f}_C = -\delta(y - x) \tag{4.69}$$

with the boundary condition $\hat{f}_C(x, C, s) = 0$. We shall sketch how it is possible to find the solution of the (ordinary) differential equation (4.69). Two linear independent solutions of the corresponding homogenous equation may be written in terms of parabolic cylinder functions $D_{-s}(y)$ [Grad94], (9.255 page 1095) as:

$$f_+(y, s) = e^{-\frac{y^2}{4}}D_{-s}(y) \text{ and } f_-(y, s) = e^{-\frac{y^2}{4}}D_{-s}(-y). \tag{4.70}$$

Further the corresponding Wronski determinant is [Abra70], (page 687):

$$W(y, s) = Wr(f_+(y, s), f_-(y, s)) = e^{-\frac{y^2}{2}}Wr(D_{-s}(y), D_{-s}(-y)) = \frac{\sqrt{2\pi}}{\Gamma(s)}e^{-\frac{y^2}{2}} \tag{4.71}$$

where $\Gamma(s)$ is the Gamma Function. It is now possible to obtain a particular solution of the non-homogenous differential equation on the following form [Codd55] (page 87):

$$\hat{f}_H(x,y,s) = -\int_{\xi=-\infty}^{y} \frac{f_+(\xi,s)f_-(y,s)-f_-(\xi,s)f_+(y,s)}{W(\xi,s)}\delta(\xi-x)d\xi = \begin{cases} \dfrac{f_-(x,s)f_+(y,s)-f_+(x,s)f_-(y,s)}{W(x,s)} & ;\ y>x \\ 0 & ;\ y<x \end{cases} \tag{4.72}$$

giving the general solution of (4.69) on the form $\hat{f}_C(x,y,s) = a_1 f_+(y,s) + a_2 f_-(y,s) + \hat{f}_H(x,y,s)$. The boundary condition is fulfilled if $a_2 = -a_1 \dfrac{f_+(C,s)}{f_-(C,s)}$. If we let $y>x$ and set $b = \dfrac{a_1}{f_-(C,s)}$ then we have $\hat{f}_C(x,y,s) = \left\{ bf_-(C,s) + \dfrac{f_-(x,s)}{W(x,s)} \right\} f_+(y,s) - \left\{ bf_+(C,s) + \dfrac{f_+(x,s)}{W(x,s)} \right\} f_-(y,s)$. Now to have a bounded solution as $y \to \infty$ we must have vanishing coefficient in front of the solution $f_-(y,s)$ since this function is unlimited when $y \to \infty$. Thus $b = -\dfrac{f_+(x,s)}{W(x,s)f_+(C,s)}$ and the solution of the boundary problem is found to be:

$$\hat{f}_C(x,y,s) = \frac{\Gamma(s)}{\sqrt{2\pi}} e^{\frac{x^2}{2}} \begin{cases} \left[f_-(x,s) - \dfrac{f_-(C,s)}{f_+(C,s)} f_+(x,s) \right] f_+(y,s) & ;\ y \geq x \\ \left[f_-(y,s) - \dfrac{f_-(C,s)}{f_+(C,s)} f_+(y,s) \right] f_+(x,s) & ;\ y \leq x \end{cases} \tag{4.73}$$

To obtain the corresponding LST of the first passage time we have to integrate (4.73). If $F_{T^x}(x,t) = \boldsymbol{P}(T^x > t)$ and $\hat{F}_{T^x}(x,s)$ denote the corresponding LST, then $\hat{F}_{T^x}(x,s) = \int_{y=C}^{\infty} \hat{f}_C(x,y,s)dy$. Using the fact that both f_+ and f_- satisfy the (non-homogeneous) differential equation gives $\int f(y)dy = \dfrac{f' + yf}{s}$ and then by integrating (4.73) we find:

$$\hat{F}_{T^x}(x,s) = \frac{1}{s}\left(1 - e^{-\left(\frac{C^2}{2} - \frac{x^2}{2}\right)} \frac{f_+(x,s)}{f_+(C,s)} \right) = \frac{1}{s}\left(1 - e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \frac{D_{-s}(x)}{D_{-s}(C)} \right) \tag{4.74}$$

the corresponding LST for the density function $\hat{f}_{T^x}(x,s)$ is readily found:

$$\hat{f}_{T^x}(x,s) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \frac{D_{-s}(x)}{D_{-s}(C)} \tag{4.75}$$

We find the LST of the excess probability $\hat{\psi}_C(s)$ from (4.74) by multiplying with the standard normal density $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and integrate:

$\hat{\psi}_C(s) = \int_{x=C}^{\infty} \varphi(x) \hat{F}_{T^x}(x, s) dx = \frac{1}{s}\left(\phi(C) - \frac{\varphi(C)}{f_+(C, s)} \int_{x=C}^{\infty} f_+(x, s) dx \right)$. By integrating and using the relation between parabolic cylinder functions [Grad94], (9.247 page 1094) gives:

$$\hat{\psi}_C(s) = \frac{1}{s}\left(\phi(C) - \varphi(C) \frac{D_{-s-1}(C)}{D_{-s}(C)} \right) \tag{4.76}$$

If we let T_C be the first passage time given that the process is in an overload period (that is $\boldsymbol{B}_0 \geq C$), we may find the corresponding LST by integrating (4.75) with the conditional stationary PDF given $\boldsymbol{B}_0 \geq C$. We let $F_{T_C}(t) = \boldsymbol{P}(T_C > t)$ and $\hat{F}_{T_C}(s)$ the corresponding LST then:

$$\hat{F}_{T_C}(s) = \int_{x=C}^{\infty} \frac{\varphi(x)}{\phi(C)} \hat{F}_{T^x}(x, s) dx = \frac{\hat{\psi}_C(s)}{\phi(C)} = \frac{1}{s}\left(1 - \frac{\varphi(C)}{\phi(C)} \frac{D_{-s-1}(C)}{D_{-s}(C)} \right). \tag{4.77}$$

The corresponding Laplace transform for the PDF yields:

$$\hat{f}_{T_C}(s) = \frac{\varphi(C)}{\phi(C)} \frac{D_{-s-1}(C)}{D_{-s}(C)}. \tag{4.78}$$

We shall also sketch how it is possible to invert the LST (4.74), (4.75) and (4.76), (4.77). It turns out that the denominator $D_{-s}(C)$ (in all these expressions) has oscillating behaviour on the negative real axis as function of the variable s. The corresponding zeros will therefore be poles for the Laplace transforms. Based on the asymptotics for the Parabolic Cylinder Function one has [Abra70], (19.9.4 page 689):

$$D_{-s}(C) \sim \frac{\Gamma\left(\frac{1-s}{2}\right)}{2^{s/2} \pi^{1/2}} e^{\frac{C^2}{16p^2}} \cos\left(\frac{\pi}{2} s + pC - \frac{C^3}{24p} \right) \text{ when } s \to -\infty \tag{4.79}$$

and where $p = \sqrt{\frac{1}{2} - s}$. By (4.79) it is obvious that the number of zeros is infinite (but countable) and further they will not be limited. If we denote the zeros by $r_k(C)$ in descending order, then these zeros will be poles of first order for the Laplace transforms above. For

large values of k it is possible to obtain asymptotics for the roots by applying (4.79). We find $r_k(C) \sim \frac{1}{2} - p_k(C)^2$ where

$$p_k(C) = \frac{C}{\pi} + \sqrt{\left(2k-\frac{1}{2}\right) + \frac{C^2}{\pi^2} + \frac{C^3}{12\pi}\left(\sqrt{\left(2k-\frac{1}{2}\right) + \frac{C^2}{\pi^2}} - \frac{C^2}{\pi^2}\right)\left(2k-\frac{1}{2}\right)^{-1}} \tag{4.80}$$

as $k \to \infty$ (for moderate values of C)

It fact (4.80) is a second order approximation and turns out to be very accurate also for small and moderate values of k.

If we denote $S_k(C) = \frac{d}{ds} D_{-s}(C)\Big|_{s = r_k(C)}$ then we may invert the LST by applying the residue theorem, and we find the following series expansion for the PDF of the first passage time:

$$f_{T^x}(x, t) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \sum_{k=1}^{\infty} R_k(x, C) e^{r_k(C)t} \quad \text{where} \tag{4.81}$$

$R_k(x, C) = \frac{D_{-r_k(C)}(x)}{S_k(C)}$ is the corresponding residue. The corresponding series for the CDF yields:

$$F_{T^x}(x, t) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \sum_{k=1}^{\infty} \frac{R_k(x, C)}{-r_k(C)} e^{r_k(C)t} \tag{4.82}$$

For large k it will be convenient to find asymptotics for the residue $R_k(x, C)$. This can be done by applying (4.79) and the approximation (4.80) for the corresponding root. We find:

$$R_k(x, C) \sim (-1)^{k-1} \frac{e^{\frac{x^2 - C^2}{16 p_k(C)^2}} \cos\left(\frac{\pi}{4} - \frac{\pi}{2} p_k(C)^2 + p_k(C)x - \frac{x^3}{24 p_k(C)}\right)}{\frac{\pi}{2} - \frac{C}{2p_k(C)}\left(1 + \frac{C^2}{24 p_k(C)^2}\right)} \tag{4.83}$$

By several numerical computations we have found that the asymptotic formula (4.83) is very accurate for quite moderate values of k and we have used this formula for the residue in the numerical computations for $k \geq 100$ mainly because of the fact that beyond this value the exact roots are difficult to find, and we observe that for small values of t the series

(4.81) and (4.82) are slowly converging, and therefore by applying the asymptotic residue we may perform the summation to quite large value of k.

To complete the picture, we have also considered the asymptotics of the LST for large positive values of s. It is well known that the behaviour of the LST for large s determine course of the distribution for small t. In Appendix E we find the following asymptotics for small t:

$$f_{T^x}(x,t) \sim e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \frac{x-C}{2\sqrt{\pi t^3}} e^{-\frac{(x-C)^2}{4t}} \left\{ \left(1 + t\left(1 - \frac{x^2+xC+C^2}{12}\right)\right) \right\} \text{ and} \tag{4.84}$$

$$F_{T^x}(x,t) \sim 1 - e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \left(Erfc\left[\frac{x-C}{2\sqrt{t}}\right]\left(1 + \frac{1}{2}(x-C)\left(\frac{x^2+xC+C^2}{12}\right)\right) - \frac{x-C}{2}\sqrt{\frac{t}{\pi}} e^{-\frac{(x-C)^2}{4t}} \left(\frac{x^2+xC+C^2}{12}\right) \right) \tag{4.85}$$

as $t \to 0$.

The asymptotics (4.84) and (4.85) act more or less like a boundary layer where the course of the density function suddenly changes (for very small values of t) and forcing the density to approach zero in a very short interval near zero.

Special cases occur for instance when we consider crossings across the median; that is $C = 0$. In this case the poles are located at $s = -(2k+1)$ and the residues may be evaluated explicitly. We find that the PDF of the first passage time of the median may be expressed as:

$$f_{T^x}(x,t) = e^{\frac{x^2}{4}} \sum_{k=0}^{\infty} \frac{(-1)^k D_{2k+1}(x)}{2^k k!} e^{-(2k+1)t} \tag{4.86}$$

with the corresponding CDF

$$F_{T^x}(x,t) = e^{\frac{x^2}{4}} \sum_{k=0}^{\infty} \frac{(-1)^k D_{2k+1}(x)}{(2k+1)2^k k!} e^{-(2k+1)t} \tag{4.87}$$

By applying the same method as above we may find the residue series for the CDF of the variable T_C, the first passage time of the level C given that the process is in an overload period as:

$$F_{T_C}(t) = \frac{\varphi(C)}{\phi(C)} \sum_{k=1}^{\infty} \frac{R_k(C)}{-r_k(C)} e^{r_k(C)t} \tag{4.88}$$

and the PDF is

$$f_{T_C}(t) = \frac{\varphi(C)}{\phi(C)} \sum_{k=1}^{\infty} R_k(C) e^{r_k(C)t} \tag{4.89}$$

where the corresponding residue $R_k(C) = \dfrac{D_{-r_k(C)-1}(C)}{S_k(C)}$. Further the excess probability $\psi_C(t)$ is found by (4.77) as the relation between the LST implying:

$$\psi_C(t) = \phi(C) F_{T_C}(t) \tag{4.90}$$

For large values of k we find the following asymptotic formula for the residue $R_k(C)$

$$R_k(C) \sim \frac{(-1)^{k-1} e^{-\frac{C^2}{16-(1-2p_k(C))^2}} \cos\left(\frac{\pi}{4} - \frac{\pi}{2} q_k(C)^2 + C q_k(C) - \frac{C^3}{24 q_k(C)}\right)}{\left(\sqrt{p_k(C)^2 - \frac{3}{2}}\right)\left(\frac{\pi}{2} - \frac{C}{2p_k(C)}\left(1 + \frac{C^2}{24 p_k(C)^2}\right)\right)} \tag{4.91}$$

where we have defined $q_k(C) = \sqrt{p_k(C)^2 - 1}$ and $p_k(C)$ is given by (4.80).

In Appendix E we have found the following asymptotics for small t:

$$f_{T_C}(t) \sim \frac{\varphi(C)}{\phi(C)}\left(\frac{1}{\sqrt{\pi t}} - \frac{C}{2}\right) \text{ and} \tag{4.92}$$

$$F_{T_C}(t) \sim 1 - \frac{\varphi(C)}{\phi(C)}\left(2\sqrt{\frac{t}{\pi}} - \frac{Ct}{2}\right) \tag{4.93}$$

As above the cases $C = 0$ simplify the residues and we find for this case:

$$f_{T_C}(t) = \frac{2}{\pi}\left\{ e^{-t} + \sum_{k=1}^{\infty} \frac{1 \cdot 3 \cdot \ldots (2k-1)}{2 \cdot 4 \cdot \ldots 2k} e^{-(2k+1)t} \right\} \text{ and} \tag{4.94}$$

$$F_{T_C}(t) = \frac{2}{\pi}\left\{ e^{-t} + \sum_{k=1}^{\infty} \frac{1 \cdot 3 \cdot \ldots (2k-1)}{2 \cdot 4 \cdot \ldots (2k)(2k+1)} e^{-(2k+1)t} \right\} \tag{4.95}$$

Generally the residue expansion will converge slowly for small t (and actually the series for $f_{T_C}(t)$ will be divergent for $t = 0$. This is seen from (4.92) showing that the limit does not exist as $t \to 0$, explaining the well known fact that the crossing intensity of the O-U process is infinite since $\psi_C'(t) \sim \varphi(C)\left(\frac{1}{\sqrt{\pi t}} - \frac{C}{2}\right)$ as $t \to 0$.

4.4.2 Joint distribution of the first passage time and the corresponding volume for the Ornstein-Uhlenbeck process

By also introducing the excess volume into the analysis we hope it will be quite easy to obtain the corresponding results as above (in section 4.4.1). In the analysis of the joint distribution we shall apply the general relations obtained in section 3.3.2. The function fully describing the state of both the excess time and volume is (given by (3.50)):

$F_C(x, y, y, t) = \boldsymbol{P}\{\boldsymbol{B}_t > y, A_t > z, Inf_{\tau \in (0,\, t)} \boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = x\}$, and we define the joint density $f_C(x, y, z, t) = \frac{\partial^2}{\partial y \partial z} F_C(x, y, z, t)$. We also let $g_C(x, y, z, t) = \varphi(x) f_C(x, y, z, t)$ be the corresponding density function without the condition on $\boldsymbol{B}_0$ (defined by (3.60)).

The density function (also including the volume) $f_C(x, y, z, t)$ will obey the following "extended" Fokker-Plank equation together with appropriate initial and boundary conditions:

$$\frac{\partial^2 f_C}{\partial y^2} + \frac{\partial}{\partial y}[y f_C] - (y - C)\frac{\partial f_C}{\partial z} = \frac{\partial f_C}{\partial t} \text{ for } x > C,\ y > C,\ z > 0 \text{ and } t > 0 \tag{4.96}$$

$$f_C(x, y, z, 0) = \delta(y - x)\delta(z) \text{ for } x > C,\ y > C,\ z > 0 \text{ and} \tag{4.97}$$

$$f_C(x, C, z, t) = 0 \text{ for } x > C,\ z > 0 \text{ and } t > 0 \text{ and} \tag{4.98}$$

$$f_C(x, y, 0, t) = 0 \text{ for } x > C,\ y > C,\ t > 0 \tag{4.99}$$

Given the partial differential equation above, the main task will be to solve this diffusion equation with the given initial and boundary equations. To get rid of the initial condition and the last boundary condition it is convenient to introduce the double LST:

$$\hat{f}_C(x, y, \zeta, s) = \int_{z=0}^{\infty} \int_{t=0}^{\infty} e^{-st - \zeta z} f_C(x, y, z, t)\, dz\, dt$$

The transformed problem is reduced to the following non-homogenous ordinary differential equation

$$\frac{\partial^2}{\partial y^2}\hat{f}_C + \frac{\partial}{\partial y}\hat{f}_C + (1 - s - (y - C)\zeta)\hat{f}_C = -\delta(y - x) \tag{4.100}$$

with the boundary condition $\hat{f}_C(x, C, \zeta, s) = 0$. The corresponding homogenous differential equation is

$$\frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial y} + (1 - s - (y - C)\zeta)f = 0 \tag{4.101}$$

The substitution $f = e^{-\frac{y^2}{4}}h$ in (4.101) gives $\frac{\partial^2 h}{\partial y^2} + \left(\frac{1}{2} + \zeta^2 + C\zeta - s - (y + 2\zeta)^2\right)h = 0$ and this is the differential equation for the Parabolic Cylinder Function with parameter $\zeta^2 + C\zeta - s$ and argument $y + 2\zeta$ [Grad94], (9.255 page 1095). Thus, two linear independent solution of (4.101) may be written in terms of Parabolic Cylinder Functions as:

$$f_+(y, \zeta, s) = e^{-\frac{y^2}{4}} D_{\zeta^2 + C\zeta - s}(y + 2\zeta) \text{ and } f_-(y, \zeta, s) = e^{-\frac{y^2}{4}} D_{\zeta^2 + C\zeta - s}(-(y + 2\zeta)) . \tag{4.102}$$

Further the corresponding Wronski determinant is [Abra70], (page 687):

$$W(y, \zeta, s) = Wr(f_+((y, \zeta, s), f_-(y, \zeta, s))) = e^{-\frac{y^2}{2}} Wr(D_{\zeta^2 + C\zeta - s}((y + 2\zeta), D_{\zeta^2 + C\zeta - s}(-(y + 2\zeta)))) = \frac{\sqrt{2\pi}}{\Gamma(s - (\zeta^2 + C\zeta))} e^{-\frac{y^2}{2}} \tag{4.103}$$

where $\Gamma(s)$ is the Gamma Function. By knowing the Wronsky-determinant and two linear independent solutions of the homogenous equation (4.101) we can write down the solution of the boundary problem directly as (4.73):

$$\hat{f}_C(x, y, \zeta, s) = \frac{\Gamma(s - (\zeta^2 + C\zeta))}{\sqrt{2\pi}} e^{\frac{x^2}{2}} \begin{cases} \left[f_-(x, \zeta, s) - \frac{f_-(C, \zeta, s)}{f_+(C, \zeta, s)} f_+(x, \zeta, s)\right] f_+(y, \zeta, s) & ; \ y \geq x \\ \left[f_-(y, \zeta, s) - \frac{f_-(C, \zeta, s)}{f_+(C, \zeta, s)} f_+(y, \zeta, s)\right] f_+(x, \zeta, s) & ; \ y \leq x \end{cases} \tag{4.104}$$

As a side result of the special form of the differential equation (4.101) we have $\int (s + \zeta(y - C))f(y)dy = f' + yf$ for both f_+ and f_-. To find the (double) LST of joint distribution of the first passage time and the corresponding volume, we start with evaluating the integral $\int_{y=C}^{\infty} (s + \zeta(y - C))\hat{f}_C(x, y, \zeta, s)dy$. By dividing this integral into two parts depending on whether $y \leq x$ or $y \geq x$ and using the property above we find:

$$\int_{y=C}^{\infty} (s+\zeta(y-C))\hat{f}_C(x,y,\zeta,s)dy = 1 - e^{\left(\frac{x^2}{2}-\frac{C^2}{2}\right)} \frac{f_+(x,\zeta,s)}{f_+(C,\zeta,s)} \tag{4.105}$$

We let $F_{A^xT^x}(x,z,t) = P(A^x > z, T^x > t)$ be the joint CDF of the first passage time T^x and the corresponding excess volume $A^x = A_{T_x} = \int_0^{T^x} (\boldsymbol{B}_\tau - C)d\tau$ and let $f_{A^xT^x}(x,z,t) = \frac{\partial^2}{\partial t \partial z} F_{A^xT^x}(x,z,t)$ denote the corresponding joint PDF, and further we let $\hat{f}_{T^xA^x}(x,s,\zeta) = E[e^{-sT^x-\zeta A^x}\big|\boldsymbol{B}_0 = x]$ be the corresponding double LST. By the relation (3.58) we have:

$$\hat{f}_{T^xA^x}(x,\zeta,s) = e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \frac{D_{\zeta^2+C\zeta-s}(x+2\zeta)}{D_{\zeta^2+C\zeta-s}(C+2\zeta)} \tag{4.106}$$

From (4.106) we find the LST of the density function for the excess volume A^x by:

$$\hat{f}_{A^x}(x,\zeta) = \hat{f}_{T^xA^x}(x,\zeta,0) = e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \frac{D_{\zeta^2+C\zeta}(x+2\zeta)}{D_{\zeta^2+C\zeta}(C+2\zeta)} \tag{4.107}$$

and the LST for the corresponding distribution function

$$\hat{F}_{A^x}(x,\zeta) = \frac{1}{\zeta}\left(1 - e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \frac{D_{\zeta^2+C\zeta}(x+2\zeta)}{D_{\zeta^2+C\zeta}(C+2\zeta)}\right) \tag{4.108}$$

To compare with the general expressions of the LSTs found in chapter 3 (cf. (3.65) and (3.66)) we would like to find the integral:

$$\hat{H}_C(\zeta,s) = \int_{x=C}^{\infty} (s+\zeta(x-C))\varphi(x)\hat{f}_{T^xA^x}(x,\zeta,s)dx. \tag{4.109}$$

By (4.106) we get $\hat{H}_C(\zeta,s) = \frac{\varphi(C)}{f_+(C,\zeta,s)} \int_{x=C}^{\infty} (s+\zeta(x-C))f_+(x,\zeta,s)dx$. Integrating and using the relation between derivatives of parabolic cylinder functions [Grad94], (9.247 page 1094) we find:

$$\hat{H}_C(\zeta, s) = \varphi(C)\left[\zeta - (\zeta^2 + C\zeta - s)\frac{D_{\zeta^2 + C\zeta - s - 1}(C + 2\zeta)}{D_{\zeta^2 + C\zeta - s}(C + 2\zeta)}\right] \tag{4.110}$$

To compare with the general expressions for the LSTs found in chapter 3 (cf. (3.65) and (3.66)), we would also like to find the following integral:

$$\hat{G}_C(\zeta, s) = \int_{x=C}^{\infty}\int_{y=C}^{\infty}(s + \zeta(x - C))(s + \zeta(y - C))\varphi(x)\hat{f}_C(x, y, \zeta, s)dydx. \tag{4.111}$$

Then by (4.105) we get $\hat{G}_C(\zeta, s) = \int_{x=C}^{\infty}(s + \zeta(x - C))\left(\varphi(x) - \varphi(C)\frac{f_+(x, \zeta, s)}{f_+(C, \zeta, s)}\right)dx$, and we find:

$$\hat{G}_C(\zeta, s) = s\phi(C) + \zeta(\varphi(C) - C\phi(C)) + \hat{H}_C(\zeta, s) \tag{4.112}$$

The transforms (4.110) and (4.112) are in full accordance with the results found by (3.65) and (3.66), however due to the behaviour of the corresponding inverse $H(z, t)$ (of $\hat{H}_C(\zeta, s)$) for small values of z and t, the corresponding crossing intensity Δ_C is infinite. We may use (4.111) together with the transforms (4.110) and (4.112) to obtain the LST of the excess probability by:

$$\hat{\psi}_C(s) = \frac{\hat{G}_C(0, s)}{s^2} = \int_{x=C}^{\infty}\int_{y=C}^{\infty}\varphi(x)f_C(x, y, 0, s)dydx = \frac{1}{s}\left(\phi(C) - \varphi(C)\frac{D_{-s-1}(C)}{D_{-s}(C)}\right) \tag{4.113}$$

The corresponding result for the LST of the excess volume $\hat{\Psi}_C(\zeta)$ is found similarly by setting $s = 0$ in (4.110) and (4.112):

$$\hat{\Psi}_C(\zeta) = \frac{\hat{G}_C(\zeta, 0)}{\zeta^2} = \int_{x=C}^{\infty}\int_{y=C}^{\infty}(x - C)(y - C)\varphi(x)\hat{f}_C(x, y, \zeta, 0)dydx =$$

$$\frac{1}{\zeta}\left(\varphi(C) - C\phi(C) - \varphi(C)\left(1 - (\zeta + C)\frac{D_{\zeta^2 + C\zeta - 1}(C + 2\zeta)}{D_{\zeta^2 + C\zeta}(C + 2\zeta)}\right)\right) \tag{4.114}$$

If we let A_C be the excess volume for the corresponding first passage time T_C given that the process is in an overload period (that is $\boldsymbol{B}_0 \geq C$) (defined in section 4.4.1), and let $F_{A_C}(z) = \boldsymbol{P}(A_C > z)$ and let $\hat{F}_{A_C}(\zeta)$ denote the corresponding LST, we may find this Laplace transform from the excess volume (4.114) by a simple scaling so that we get a proper distribution:

$$\hat{F}_{A_C}(\zeta) = \frac{1}{\zeta}\left(1 - \frac{\varphi(C)}{\varphi(C) - C\phi(C)}\left(1 - (\zeta + C)\frac{D_{\zeta^2 + C\zeta - 1}(C + 2\zeta)}{D_{\zeta^2 + C\zeta}(C + 2\zeta)}\right)\right) \tag{4.115}$$

The LST of the corresponding PDF is then:

$$\hat{f}_{A_C}(\zeta) = \frac{\varphi(C)}{\varphi(C) - C\phi(C)}\left(1 - (\zeta + C)\frac{D_{\zeta^2 + C\zeta - 1}(C + 2\zeta)}{D_{\zeta^2 + C\zeta}(C + 2\zeta)}\right) \tag{4.116}$$

It turns out that the LST (4.115) also could have been obtained from (4.107) by performing the integral $\int_{x=C}^{\infty} \varphi^*(x)\hat{F}_{A^*}(x, \zeta)dx$ where $\varphi^*(x) = \frac{(x-C)\varphi(x)}{\varphi(C) - C\phi(C)}$ for $x \geq C$ and where we observe that $\varphi^*(y + C)$; $y \geq 0$ is the residual density for the shifted variable $\boldsymbol{B}_0 - C$ conditioned on $\boldsymbol{B}_0 \geq C$ (having the density $\frac{\varphi(y + C)}{\phi(C)}$ for $y \geq 0$).

Below we shall describe how it is possible to invert the LST (4.107), (4.108) and (4.115), (4.116). It turns out that the key to find the inversion of the given Laplace transforms is the behaviour of the denominator $f_C(\zeta) = D_{\zeta^2 + C\zeta}(C + 2\zeta)$ in the negative half-plane. This function will have infinite many zeros on the negative real axis, and we denote the zeros by $u_k(C)$ in descending order and it follows that these zeros will be poles of first order for the Laplace transforms above.

By the asymptotic formula (E.15) (and (E.9)) Appendix E we have:

$$f_C(-\eta) \sim \frac{(2\pi)^{\frac{1}{2}}}{3^{2/3}\Gamma(2/3)}\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}} e^{-\frac{1}{2}\left(\eta - \frac{C}{2}\right)^2 + (\zeta^2 + C\zeta - 1)\log\left(\eta - \frac{C}{2}\right)} g_C(\eta) \text{ where} \tag{4.117}$$

$$g_C(\eta) = \cos(\pi(\eta^2 - C\eta))\left(1 + \frac{3^{1/3}\Gamma(2/3)}{\Gamma(1/3)}\left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\right) - \sqrt{3}\sin(\pi(\eta^2 - C\eta))\left(1 - \frac{3^{1/3}\Gamma(2/3)}{\Gamma(1/3)}\left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\right)$$

$$\text{when } \eta \to \infty \text{ (and } Re(\eta) \geq 0\text{).} \tag{4.118}$$

and where we also have used the results [Abra70] (10.4.4 and 10.4.5 page 446) for the values of the Airy functions for $z = 0$. For large values of k it is possible to obtain asymptotics for the roots by finding the zeros of (4.118). We find $u_k(C) \sim -v_k(C)$ where

$$v_k(C) = \frac{C}{2} + \sqrt{\frac{C^2}{4} + \left(k - \frac{5}{6}\right) + \frac{3^{5/6}\Gamma(2/3)}{2\pi\Gamma(1/3)}\left(\frac{C^2}{4} + \left(k - \frac{5}{6}\right)\right)^{-\frac{1}{3}}\left(\frac{1}{2} - \frac{C^2}{4}\right)} \tag{4.119}$$

as $k \to \infty$ (for moderate values of C). It fact (4.119) is a second order approximation and turns out to be very accurate also for small and moderate values of k.

If we denote $U_k(C) = \frac{d}{d\zeta} D_{\zeta^2 + C\zeta}(C + 2\zeta)\Big|_{\zeta = u_k(C)}$ we may invert the LSTs by applying the residue theorem, and we get the following infinite series for the PDF of the excess volume of the first passage time:

$$f_{A^x}(x, z) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \sum_{k=1}^{\infty} V_k(x, C) e^{u_k(C) z} \quad \text{where} \tag{4.120}$$

$V_k(x, C) = \frac{D_{u_k(C)^2 + C u_k(C)}(x + 2u_k(C))}{U_k(C)}$ is the corresponding residue. The corresponding series for the CDF yields:

$$F_{A^x}(x, z) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \sum_{k=1}^{\infty} \frac{V_k(x, C)}{-u_k(C)} e^{u_k(C) z} \tag{4.121}$$

Based on the asymptotics for parabolic cylinder functions in Appendix E we may find the find asymptotic expressions for the residue $V_k(x, C)$ in terms of Airy functions. We get:

$$V_k(x, C) \sim \frac{T(x, v_k(C), C)}{N(x, v_k(C), C)} \tag{4.122}$$

where

$$T(x, \eta, C) = Ai\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) - \left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}} Ai'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) \tag{4.123}$$

$$-\frac{1}{\sqrt{3}}\left(Bi\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) + \left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\left[\frac{2 \cdot 3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} Bi\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) - Bi'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right)\right]\right)$$

and

$$N(x, \eta, C) = \frac{8\pi}{3^{\frac{7}{6}}\Gamma\left(\frac{2}{3}\right)}\left(\eta - \frac{C}{2}\right)\left(1 + \frac{3^{\frac{2}{3}}\Gamma\left(\frac{2}{3}\right)^2}{\Gamma\left(\frac{1}{3}\right)^2}\left(\frac{1}{2} - \frac{C^2}{4}\right)^2\left(\eta - \frac{C}{2}\right)^{-\frac{4}{3}}\right) \tag{4.124}$$

We have experienced (by several numerical computations) that the asymptotic formula for the residue is quite accurate also for moderate values of k. Due to numerical difficulties to calculate the parabolic cylinder functions for large arguments and large parameter, we are not able to calculate the corresponding roots when k is larger than approximately 27. Therefore, we have used the asymptotics (4.122)-(4.124) for the residue in the numerical computations for $k \geq 27$, and we observe as for the first passage time distributions that the series (4.120) and (4.121) are slowly converging for small values of z, and therefore by using the asymptotic residue we may perform the summation to quite large k-values. This is mainly due to the fact that the Airy functions (in (4.123)) for negative values are oscillating and may easily be computed also for large arguments.

An efficient alternative to obtain the distributions for small z is to apply the asymptotics obtained by considering the Laplace transforms for large (positive) arguments. In Appendix E we find the following asymptotics for small z:

$$f_{A^x}(x,z) \sim \frac{3^{\frac{7}{6}}\Gamma\left(\frac{2}{3}\right)}{2\pi} e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} e^{-\frac{(x-C)^3}{18z}} \left(\frac{z^{-\frac{2}{3}}}{(x-c)} W_{\frac{2}{3},\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) + \frac{C}{6}(x-c)^{\frac{1}{2}} z^{-\frac{1}{6}} W_{\frac{1}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) + \right.$$

$$\left. \left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\frac{z^{-\frac{1}{6}}}{(x-c)^{\frac{1}{2}}} W_{\frac{1}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) - \frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \frac{1}{(x-c)} W_{0,\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) \right) \right) \quad \text{and} \tag{4.125}$$

$$F_{A^x}(x,z) \sim 1 - \frac{3^{\frac{7}{6}}\Gamma\left(\frac{2}{3}\right)}{2\pi} e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} e^{-\frac{(x-C)^3}{18z}} \left(\frac{z^{\frac{1}{3}}}{(x-c)} W_{-\frac{1}{3},\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) + \frac{C}{6}(x-c)^{\frac{1}{2}} z^{\frac{5}{6}} W_{-\frac{5}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) + \right.$$

$$\left. \left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\frac{z^{\frac{5}{6}}}{(x-c)^{\frac{1}{2}}} W_{-\frac{5}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) - \frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \frac{z}{(x-c)} W_{-1,\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) \right) \right) \tag{4.126}$$

as $z \to 0$ and where $W_{\kappa,\mu}(y)$ is the second Whittakers's function ([Abra70] (9.22-9.23 page 1086).

As for the distribution of the first passage time the asymptotics (4.125) and (4.126) act like boundary layer solutions where the density function suddenly changes (for very small values of z) and forcing the density to approach zero in a very short interval near zero.

The form of the asymptotics for the first passage time ((4.84) and (4.85)) and the corresponding result for the volume ((4.125) and (4.126)) are quite similar, however, while

we for the time distribution have $\frac{(x-C)^2}{4t}$ as the "local variable", the corresponding "boundary" variable for the volume is $\frac{(x-C)^3}{9z}$. This suggests that these variable should be introduced in the differential equations to obtain the corresponding boundary equations.

We shall also give the residue expansion for the CDF of the volume A_C, as defined above, on the basis of the first passage time of the level C given that the process is in an overload period. We find:

$$F_{A_C}(z) = \frac{\varphi(C)}{\varphi(C)-C\phi(C)} \sum_{k=1}^{\infty} \frac{V_k(C)}{-u_k(C)} e^{u_k(C)z} \tag{4.127}$$

and the corresponding PDF

$$f_{A_C}(z) = \frac{\varphi(C)}{\varphi(C)-C\phi(C)} \sum_{k=1}^{\infty} V_k(C) e^{u_k(C)z} \tag{4.128}$$

where the corresponding residue $V_k(C) = \frac{-(u_k(C)+C)D_{u_k(C)^2+Cu_k(C)-1}(C+2u_k(C))}{U_k(C)}$.

By applying the asymptotic formulas (E.15) and (E.16) in Appendix E we find the following asymptotic formula for the residue $V_k(C)$ for large values of k:

$$V_k(C) \sim \frac{3^{\frac{5}{6}}\Gamma\left(\frac{2}{3}\right)}{4\pi\Gamma\left(\frac{1}{3}\right)v_k(C)\left(v_k(C)-\frac{C}{2}\right)^{\frac{2}{3}}}\left(1+\frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)}\left(\frac{1}{2}-\frac{C^2}{4}\right)\left(v_k(C)-\frac{C}{2}\right)^{-\frac{2}{3}}\right) \tag{4.129}$$

where and $v_k(C)$ is the asymptotics of the roots given by (4.119).

In Appendix E we have found the following asymptotics for small z:

$$f_{A_C}(z) \sim \frac{\varphi(C)}{\varphi(C)-C\phi(C)}\left(\frac{3^{\frac{1}{3}}}{\Gamma\left(\frac{1}{3}\right)}z^{-\frac{1}{3}}-\frac{C}{2}\right) \text{ and} \tag{4.130}$$

$$F_{A_C}(z) \sim 1-\frac{\varphi(C)}{\varphi(C)-C\phi(C)}\left(\frac{3^{\frac{4}{3}}}{2\Gamma\left(\frac{1}{3}\right)}z^{\frac{2}{3}}-\frac{Cz}{2}\right) \tag{4.131}$$

We observe that the density function of the A_C grows by $z^{-1/3}$ for small z whereas the corresponding density function for the time T_C behaves as $t^{-1/2}$. Thus it seems that the time distribution is “less regular” than the corresponding volume, however, both density functions are unlimited near the origin.

4.4.3 Some numerical examples

As mentioned above the numerical computations are based on the residue series both for the distribution of the first passage times and for the distribution of the corresponding volume. For the lowest residues the corresponding roots were calculated numerically and the corresponding values of the residues were calculated by means of numerical derivation. However, due to difficulties when calculating Cylinder Functions for large arguments we failed to calculate the residues for larger than approximately:

- 100 for the excess time distributions and
- 27 for the excess volume distributions

and these numbers were far too small to obtain the residue expansion to converge for small arguments. This is due to the fact that these expansions converge very slowly in this region. However, by means of the asymptotics derived for the residues, which turn out to be very accurate also for relative small numbers (in the series), we were able to obtain accurate numerical values also for small arguments. In the example below we therefore used the asymptotic residues from:

- 100 to 8000 for the excess time distributions and
- 27 to 16000 for the excess volume distributions.

In addition the numerical calculations were compared with the asymptotic expansions for small arguments, and we were pleased to observe that for both the excess times and excess volume, calculated by using the residue series, give results that are very close to curves obtained by the asymptotic expansion for small arguments. (See figures 4.11, 4.12, 4.13 and 4.15.)

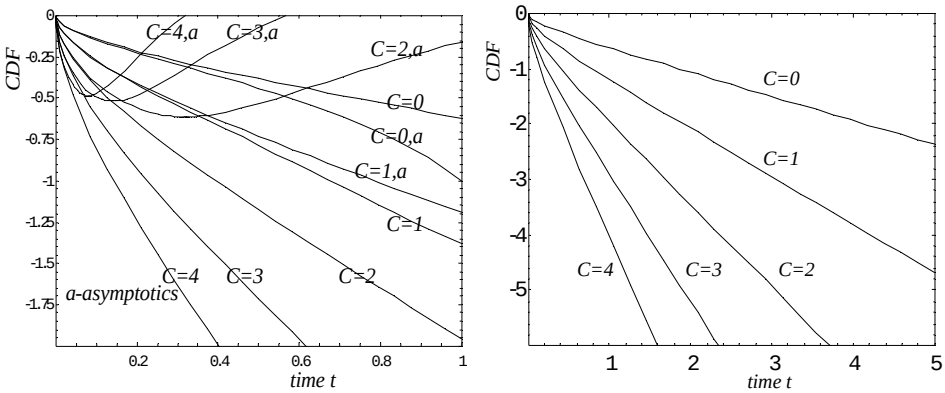

Figure 4.11: Logarithmic plot of the CDF of the first passage time given that the process is in an excess period for the O-U process for different values of the scaled capacity (C=0,1,2,3,4) as function of time.

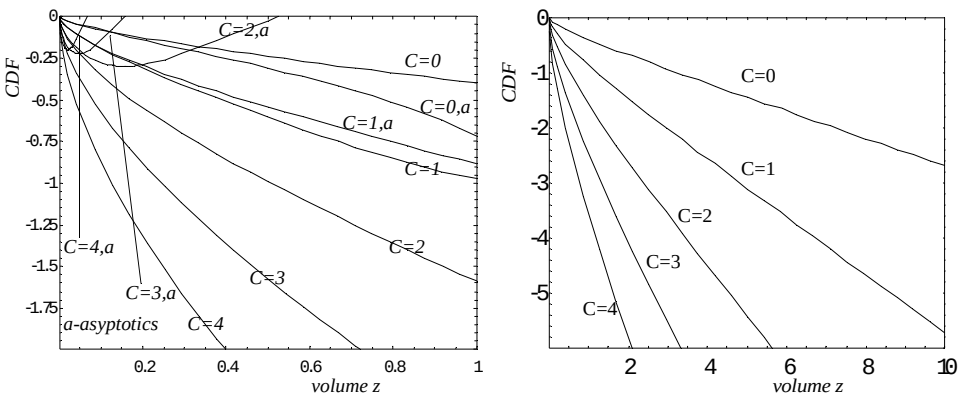

Figure 4.12: Logarithmic plot of the CDF of the excess volume during the first passage time given that the process is in an excess period for the O-U process for different values of the scaled capacity (C=0,1,2,3,4) as function of volume.

In figure 4.11 we have given plots of the CDF of the first passage time given that the process starts in an excess period for different values of the capacity level. As expected this distribution will depend on the actual level C. As pointed out in section 4.2 the interesting values of C will be in the range 2.0-3.2. It is obvious that these distributions will have exponential tail since we get nearly straight lines as the argument increases.

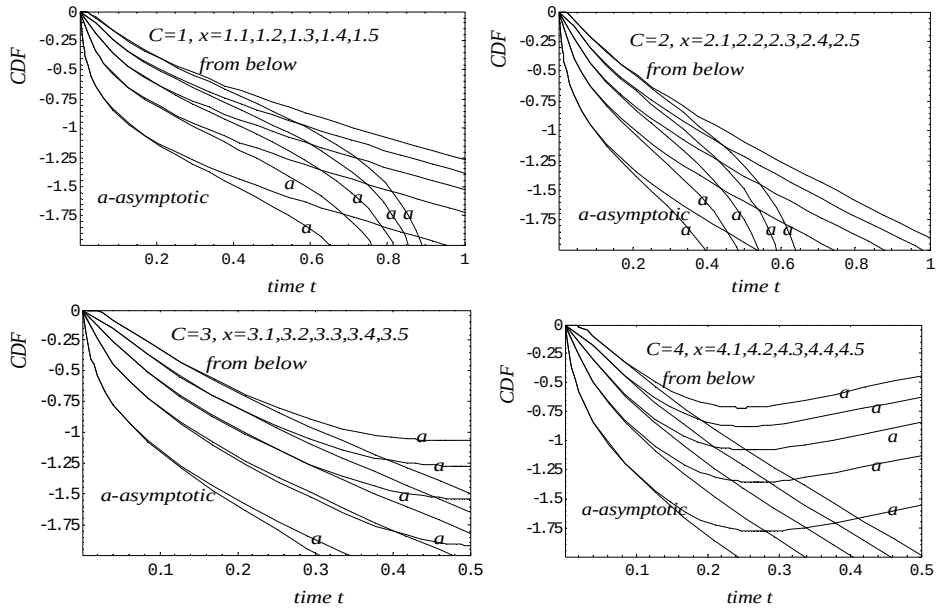

Figure 4.13: Logarithmic plot of the CDF of the conditional first passage time for the O-U process for different values of the scaled capacity (C=1,2,3,4) and different starting values as function of time.

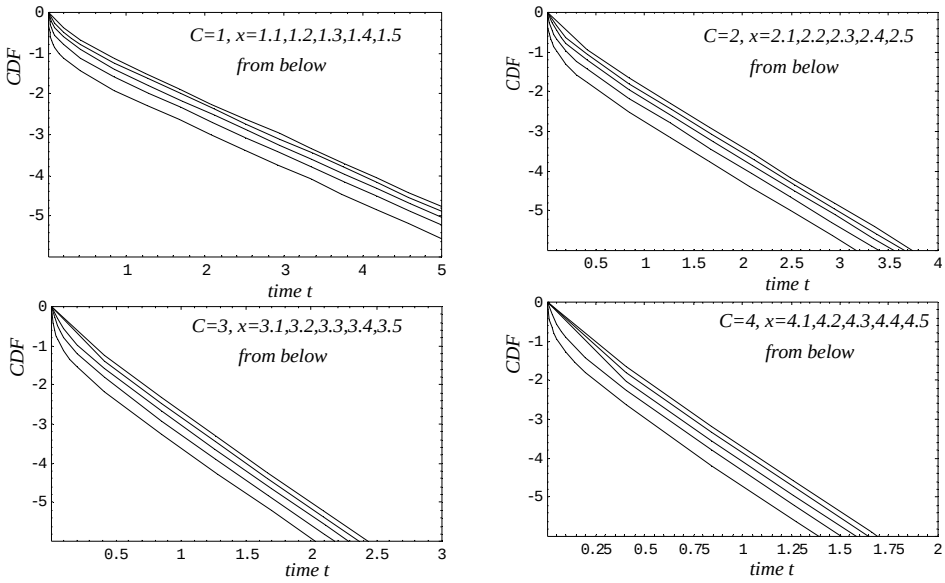

Figure 4.14: Logarithmic plot of the CDF of the conditional first passage time for the O-U process for different values of the scaled capacity (C=1,2,3,4) and different starting values as function of time.

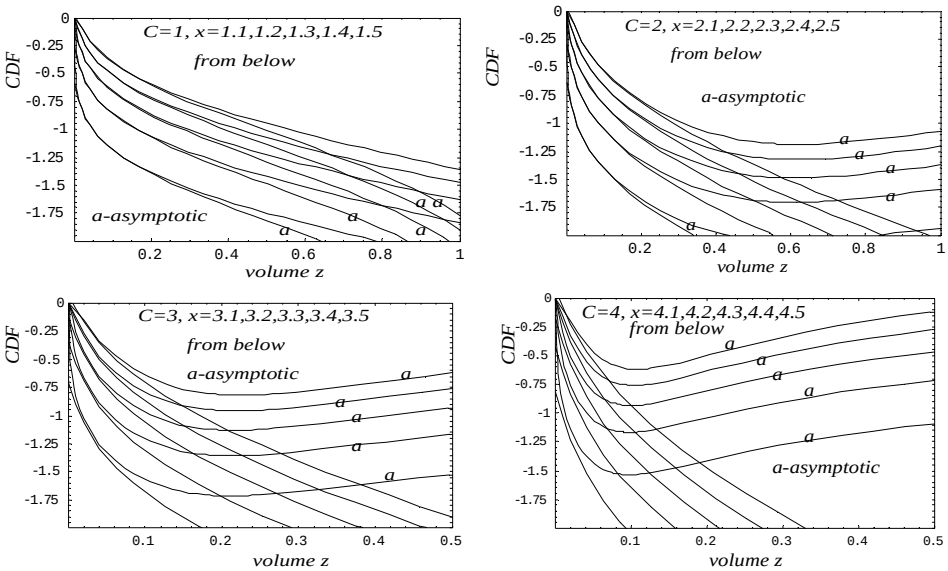

Figure 4.15: Logarithmic plot of the CDF of the conditional excess volume during the first passage time for the O-U process for different values of the scaled capacity (C=1,2,3,4) and different starting values as function of volume.

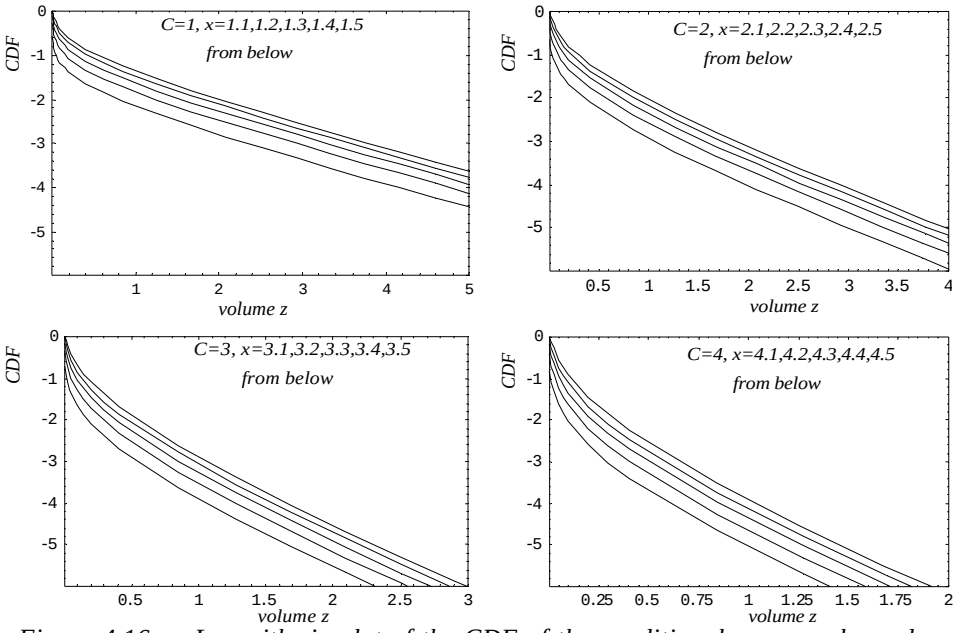

Figure 4.16: Logarithmic plot of the CDF of the conditional excess volume during the first passage time for the O-U process for different values of the scaled capacity (C=1,2,3,4) and different starting values as function of volume.

When it comes to the corresponding excess volume (figure 4.12) the curves are similar but we observe that the curves are not complete straight lines especially for higher values of the capacity level as $C = 3$ and $C = 4$.

For the conditional excess (first passage) time (figures 4.13 and 4.14) and excess volume (4.15 and 4.16) we see more variety in the curves. This is especially observed for small values of the argument where we see that the starting value of the process is an important parameter for the excess distributions. If the process starts close to the actual capacity level C we see that the distributions will drop suddenly and this is due to the fact that with high probability such an excess period will end very rapidly. On a longer scale however as seen by figures 4.14 and 4.16 we end up with exponential tails since all the curves become straight lines for large arguments. This fact will also be obvious from the obtained residue expansion where the first term in the series will dominate for large value of the arguments.

In this section we have demonstrated that it is possible to obtain by analytical means both the Laplace transforms for the first passage times and the time dependant excess probabilities for the O-U process. The results for the distribution of the excess volumes are of special interest since this can estimate the amount of information loss for communication systems during congestion periods. It turns out that it is possible to invert the Lapace transforms by finding the poles. These are all located on the negative real axis and we obtain asymptotic formulas for the location of these poles correct to second order. Furthermore we also obtain asymptotics for the corresponding residue also correct to second order. By means of these asymptotics the residue series may be calculated for large numbers terms and we obtain numerical results also near the origin where the residue series is slowly converging. The numerical values for small arguments are also found to be in accordance with the distribution functions obtained by the asymptotics found for small arguments.

5

Some results on excess times and excess volume for semi-Markov processes

5.1 Introduction

In chapter 3 we discussed a general framework to determine the distribution of excess times and excess volumes for a stationary stochastic process. In that chapter we focused on some rather general considerations, although in section 3.3 where we consider the joint distribution of excess volumes and excess times we assumed that the processes are continuous (in time and space). As pointed out processes with continuous sample paths do not fulfil the requirements for level crossing unless the autocorrelation has quite specific behaviour near the origin. Discrete processes, however, will not need these restrictions, and we know that such processes have been widely used as models to describe traffic load on communication links. Especially, some specific types of Markov processes have been analysed in light of investigating transient phenomena as duration of excess periods and the corresponding excess volumes. As mentioned the distribution of such periods may be viewed as periods with overload on a communication link and the excess volumes may represent the amount of information loss in bufferless multiplexing.

For a Markov process the excess time distribution has been derived by Buzacott used in reliability analysis to find up and down times for repairable systems [Buza70]. Other authors have used the classical $M/M/\infty$ system as a model for bufferless multiplexing of large number of data sources on a communication link, and explicit expressions and various asymptotics are found for the different transient performance measures [Guill95], [Guill96], [Dupu97].

In our analysis we have taken a more general approach by assuming that the rate process is a general semi-Markov process. By applying semi-Markov processes we may also analyse sources that have autocorrelations that are different from exponentially decaying ones obtained from ordinary Markov modelling. In the following we shall start with a general semi-Markov model, but we will also include a lot of results for processes that are particular cases of the general model, such as ordinary Markov processes, or even the simpler birth-death processes. There is, however, one main drawback with the general semi-Markov models that need to be mentioned: the class of sami-Markov processes is not closed under “addition”.

5.2 Some general properties for semi-Markov processes

In this chapter we shall assume that the bit rate process $\{\boldsymbol{B}_t\}$ is a semi-Markov process, where the bit rate takes the possible values jb where b is the unit changes in the bit rate and $j \in E$ where E is a countable set of numbers. We shall not go into any further discussion on the theory of semi-Markov processes (or Markov Renewal processes), but we refer to the textbook of Cinlar [Cinl75] for basic properties.

To describe the excess times and excess volume for a semi-Markov process we need to divide the state space into two disjoint sets: E^u and E^l where $jb > C$ for $j \in E^u$ which we call the "overload states" and $jb \le C$ for $j \in E^l$ which we take as the "normal load states", and we let $j_C = \left\lfloor \frac{C}{b} \right\rfloor$ be the limiting state. In the succeeding we will make the following convention:

For a vector $\boldsymbol{a} = (a_j)$ and a matrix $\boldsymbol{A} = (A_{ij})$; $i, j \in E$ we use the notation $\boldsymbol{a}^u$, $\boldsymbol{A}^u$ and $\boldsymbol{a}^l$, $\boldsymbol{A}^l$ for the overload part and normal load part of $\boldsymbol{a}$ and $\boldsymbol{A}$; i.e. $\boldsymbol{a}^u = (a_j)$, $j \in E^u$, $\boldsymbol{A}^u = (A_{ij})$, $i, j \in E^u$ and $\boldsymbol{a}^l = (a_j)$, $j \in E^l$, $\boldsymbol{A}^l = (A_{ij})$, $i, j \in E^l$. To complete the partitioning of the matrix $\boldsymbol{A}$ we also define the sub-matrices $\boldsymbol{A}^{l,u} = (A_{ij})$, $i \in E^l$, $j \in E^u$, and $\boldsymbol{A}^{u,l} = (A_{ij})$, $i \in E^u$ $j \in E^l$.

We let $\{T_n\}$ be the sequence of jump instants for the bit rate process $\{\boldsymbol{B}_t\}$ and let $B_n = \boldsymbol{B}_{T_n}$ be the bit rate at jump instants. The evolution of the process over time is described by the generator $\boldsymbol{P}(t) = (P_{ij}(t))$ which we take to be:

$$P_{ij}(t) = \boldsymbol{P}\{B_{n+1} = jb, T_{n+1} - T_n > t | B_n = ib\} \tag{5.1}$$

Based on the generator matrix $\boldsymbol{P}(t)$ it is possible to find the different characteristics for the process. For example the conditional CDF between two succeeding jumps is found from the generator by adding the states

$$H_i(t) = \boldsymbol{P}\{T_{n+1} - T_n > t | B_n = ib\} = \sum_{j \in E} P_{ij}(t) \tag{5.2}$$

The steady state distributions at jump instants $\pi = (\pi_i)$ are given as the solution of the equation $\pi \cdot \boldsymbol{P}(0) = \pi$ together with the relation $\sum_{i \in E} \pi_i = 1$.

Also the steady state distribution $\boldsymbol{p} = (p_i)$ is known to be proportional to $m_i \pi_i$, where

$$m_i = \boldsymbol{E}[T_{n+1} - T_n | B_n = ib] = \int_{t=0}^{\infty} H_i(t)dt \tag{5.3}$$

is the conditional mean between two consecutive changes in the bit rate. Alternative the steady state probabilities $\boldsymbol{p}$ can be determined from the equation $\boldsymbol{p} \cdot \boldsymbol{Q} = \boldsymbol{0}$ and the relation $\sum_{i \in E} p_i = 1$, where the matrix $\boldsymbol{Q}$ is given by:

$$Q_{ij} = \frac{P_{ij}(0) - \delta_{ij}}{m_i} \tag{5.4}$$

As we shall discuss later, the $\boldsymbol{Q}$-matrix defined by (5.4) will give the corresponding generator matrix when the bit rate process is an ordinary Markov process.

To obtain the excess time distributions (defined in section 3.2) for the semi-Markov process $\{\boldsymbol{B}_t\}$ we define the upper and lower (Markov) renewal kernels:

$$\boldsymbol{R}^u(t) = \boldsymbol{I}^u\delta(t) + \Pi^u(t) + \{\Pi^u(t)\}^{*(2)} + ... \text{ and} \tag{5.5}$$

$$\boldsymbol{R}^l(t) = \boldsymbol{I}^l\delta(t) + \Pi^l(t) + \{\Pi^l(t)\}^{*(2)} + ... \tag{5.6}$$

where $\{\ \}^{*(n)}$ denotes n-times convolutions of $\{\ \}$ and $\Pi(t) = -\frac{d\boldsymbol{P}}{dt}(t)$ is the (negative) derivative of the generator matrix. We consider the process which is in equilibrium at time $t = 0$, and we define the conditional the excess probabilities (the discrete version of (3.21) and (3.22)) by:

$F_C(i, j, t) = \boldsymbol{P}\{\boldsymbol{B}_t = jb; Inf_{\tau \in (0,t)}\boldsymbol{B}_\tau > C | \boldsymbol{B}_0 = ib\}$ for $i > j_C$, $j > j_C$ and

$G_C(i, j, t) = \boldsymbol{P}\{\boldsymbol{B}_t = jb; Sup_{\tau \in (0,t)}\boldsymbol{B}_\tau \le C | \boldsymbol{B}_0 = ib\}$ for $i \le j_C$, $j \le j_C$.

It is possible to express these excess probabilities in terms of the upper and lower renewal kernels (defined in equation (5.5) and equation (5.6)) in the following way:

$$F_C(i, j, t) = \frac{1}{m_i}\left(\delta_{ij}\int_{\tau=t}^{\infty} H_j(\tau)d\tau + \{[P^u(t)*R^u(t)]_{ij}\}*\{H_j(t)\}\right);\ i > j_C,\ j > j_C \text{ and} \tag{5.7}$$

$$G_C(i, j, t) = \frac{1}{m_i}\left(\delta_{ij}\int_{\tau=t}^{\infty} H_j(\tau)d\tau + \{[P^l(t)*R^l(t)]_{ij}\}*\{H_j(t)\}\right);\ i \le j_C,\ j \le j_C \tag{5.8}$$

We shall briefly sketch the derivation of the excess probabilities (5.7) (and (5.8)), and we refer to figure 5.1 below.

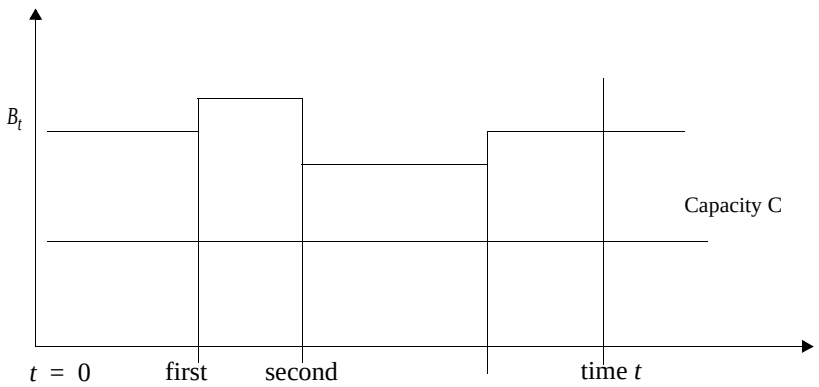

Figure 5.1: Distribution of the excess times for a semi-Markov process.

Starting the observation at time $t = 0$ in state $\boldsymbol{B}_0 = ib$ when the process is in steady state, the time to the first jump T_1 and next state for the bit rate B_1 is distributed according to the residual time for semi-Markov processes (see Cinlar [Cinl75])

$$\tilde{P}_{ij}(t) = \boldsymbol{P}\{B_1 = jb, T_1 > t | \boldsymbol{B}_0 = ib\} = \frac{1}{m_i} \int_{\tau = t}^{\infty} P_{ij}(\tau) d\tau . \tag{5.9}$$

The first part in formula (5.7) corresponds to the case where there is no state change in the process up to time t and we obtain this part by summing over the possible states for B_1 in (5.9). The second part corresponds to the case where one or more jumps have occurred before the time t, and in this case the joint density for the time to the first jump is the negative derivative of (5.9); i.e. $\tilde{\Pi}_{ij}(t) = -\frac{d}{dt}\tilde{P}_{ij}(t) = \frac{1}{m_i} P_{ij}(t)$. By convoluting these probabilities by the matrix $\{\Pi^u(t)\}^{*(n-1)}$; and adding for $n = 1, 2, 3, ,....$ and then convoluting with the conditional distribution (5.2) we get (5.7).

On the basis of (5.7) and (5.8) we obtain the excess probabilities (3.24) and (3.25) for the semi-Markov process:

$$\psi_C(t) = \sum_{i > j_C} \sum_{j > j_C} p_i F_C(i, j, t) \text{ and} \tag{5.10}$$

$$\phi_C(t) = \sum_{i \le j_C} \sum_{j \le j_C} p_i G_C(i, j, t) \tag{5.11}$$

Likewise we also obtain the conditional covariances defined in (3.36) and (3.37):

$$\gamma_C(t) = \sum_{i > j_C} \sum_{j > j_C} (ib - C)(jb - C) p_i F_C(i, j, t) \text{ and} \tag{5.12}$$

$$\beta_C(t) = \sum_{i \le j_C} \sum_{j \le j_C} (C - ib)(C - jb) p_i G_C(i, j, t) \tag{5.13}$$

It is interesting to note that all the important characteristics to describe the excess behaviour of a semi-Markov process can be expressed in terms of the generator matrix $\boldsymbol{P}(t)$ and the steady state probability vector $\boldsymbol{p}$. To find the explicit expressions for different models, the hard part is obviously to find the upper and lower renewal kernels defined through infinite sums of convolutions. One possible way to simplify the expressions may be to take Laplace transforms, which will result in a power series and where the sum can be expressed as the inverse of known matrices. We may therefore take advantage of the powerful tools of matrix algebra. Before we enter into any further discussion concerning the actual distribution of the excess times and volumes, we shall first show that the two first moments for both the excess times and excess volumes are quite straight forward to obtain from the expressions above by using the general formulae derived in chapter 3.

5.3 The first two moments for the excess times and excess volumes for semi-Markov processes

By differentiating and integrating the equation (5.7) (and (5.8)) we find:

$$F_C(i, j, 0) = \delta_{ij};\ \frac{d}{dt} F_C(i, j, 0) = Q_{ij} \text{ and } \int_{t=0}^{\infty} F_C(i, j, t)\,dt = \frac{1}{2}\chi_i \delta_{ij} - [\boldsymbol{M}^u (\boldsymbol{Q}^u)^{-1}]_{ij} \text{ where}$$

$\chi = (\chi_i)$ and $\boldsymbol{M} = (M_{ij})$ are given as the following (form) factors:

$$\chi_i = \frac{m_i^{(2)}}{m_i} \text{ where } m_i^{(2)} = \boldsymbol{E}[(T_{n+1} - T_n)^2 \big| B_n = ib] = 2 \int_{t=0}^{\infty} t H_i(t)\,dt \text{ and} \tag{5.14}$$

$$M_{ij} = \frac{m_{ij}}{m_i} \text{ where } m_{ij} = \boldsymbol{E}[(T_{n+1} - T_n)\mathbf{1}_{\{B_{n+1} = jb\}} \big| B_n = ib] = \int_{t=0}^{\infty} P_{ij}(t)\,dt \tag{5.15}$$

The results for $G_C(i, j, t)$, that is for the "normal loaded" case, are similar to the expressions above, and obtained by simply replacing $\boldsymbol{M}^u(\boldsymbol{Q}^u)^{-1}$ with $\boldsymbol{M}^l(\boldsymbol{Q}^l)^{-1}$. By using the general formulae (3.29), (3.30), and (3.32), (3.33), together with (3.39), (3.40) derived in chapter 3 we are now in the position to write down the two first moments for the excess times and excess volumes. These moments will be given in terms of the up and down crossing intensity given as the following sum:

$$\Delta = -\psi_C'(0) = -\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u = -\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l = \sum_{j > j_C} \sum_{i \le j_C} p_i Q_{ij} \tag{5.16}$$

where $\boldsymbol{e}$ denotes a column-vector with ones (with the desired dimension). In (5.16) we have used the steady state equation $\boldsymbol{p} \cdot \boldsymbol{Q} = 0$ and the relation $\boldsymbol{Q} \cdot \boldsymbol{e} = 0$.

Further for the moments of the excess times we obtain:

$$\boldsymbol{E}[T_k] = \frac{1}{\Delta} \sum_{j > j_C} p_j \text{ and } \boldsymbol{E}[S_k] = \frac{1}{\Delta} \sum_{j \le j_C} p_j \text{ and further} \tag{5.17}$$

$$\boldsymbol{E}[T_k^{\,2}] = \frac{1}{\Delta} \left\{ \sum_{j > j_C} p_j \chi_j - 2 \sum_{i > j_C} \sum_{j > j_C} p_i [\boldsymbol{M}^u (\boldsymbol{Q}^u)^{-1}]_{ij} \right\} \text{ and} \tag{5.18}$$

$$\boldsymbol{E}[S_k^{\,2}] = \frac{1}{\Delta} \left\{ \sum_{j \le j_C} p_j \chi_j - 2 \sum_{i \le j_C} \sum_{j \le j_C} p_i [\boldsymbol{M}^l (\boldsymbol{Q}^l)^{-1}]_{ij} \right\}. \tag{5.19}$$

The moments for the excess volumes may be written:

$$\boldsymbol{E}[A_k] = \frac{1}{\Delta} \sum_{j > j_C} (jb - C) p_j \text{ and } \boldsymbol{E}[V_k] = \frac{1}{\Delta} \sum_{j \le j_C} (C - jb) p_j \text{ and further} \tag{5.20}$$

$$\boldsymbol{E}[A_k^{\,2}] = \frac{1}{\Delta} \left\{ \sum_{j > j_C} (jb - C)^2 p_j \chi_j - 2 \sum_{i > j_C} \sum_{j > j_C} (jb - C)(ib - C) p_i [\boldsymbol{M}^u (\boldsymbol{Q}^u)^{-1}]_{ij} \right\}$$

and (5.21)

$$\boldsymbol{E}[V_k^{\,2}] = \frac{1}{\Delta} \left\{ \sum_{j \le j_C} (C - jb)^2 p_j \chi_j - 2 \sum_{i \le j_C} \sum_{j \le j_C} (C - jb)(C - ib) p_i [\boldsymbol{M}^l (\boldsymbol{Q}^l)^{-1}]_{ij} \right\}. \tag{5.22}$$

The formulae (5.16)-(5.22) are expressions which yield for general semi-Markov processes. For the first moments the results are identical (in form) with the corresponding results for ordinary Markov process when we define the $\boldsymbol{Q}$-matrix by (5.4) [Buza70]. To find the second order moments we require the first and second order conditional moments of the time between jumps in the semi-Markov process and also that the moment matrix $\boldsymbol{M}$ defined above to be known. The demanding part in computing the moments (5.16)-(5.22) is of cause to obtain the inverse of the upper- and lower- part of the $\boldsymbol{Q}$-matrix.

As we mentioned in the introduction the class of semi-Markov processes is a quite broad class of processes which contains a number of interesting subclasses. For some of these subclasses the expressions above will be simplified. Below we shall briefly discuss some of these cases.

5.3.1 The case when the time spent in a state is independent of the next state

If the time spent in a state in independent of the next state, i.e. we have a generator on a somewhat simpler form

$$P_{ij}(t) = P_{ij}H_i(t) \tag{5.23}$$

where $H_i(t)$ is the distribution of the time spent in state i and P_{ij} are the transition probabilities for the corresponding jump Markov-chain (with $H_i(0) = 1$ and $\sum_{i \in E} P_{ij} = 1$).

In this case we may eliminate the matrix $\boldsymbol{M}$ since we have $M_{ij} = P_{ij} = m_i Q_{ij} + \delta_{ij}$ and we get $\int_{t=0}^{\infty} F_C(i,j,t)dt = \frac{1}{2}\kappa_i\delta_{ij} - [(\boldsymbol{Q}^u)^{-1}]_{ij}$ where we define

$$\kappa_i = \chi_i - 2m_i = \frac{\boldsymbol{Var}[T_{n+1} - T_n | B_n = ib]}{\boldsymbol{E}[T_{n+1} - T_n | B_n = ib]} - \boldsymbol{E}[T_{n+1} - T_n | B_n = ib] \tag{5.24}$$

The moments for this case are therefore given through (5.16)-(5.22) by replacing χ_i by κ_i and $[\boldsymbol{M}^u(\boldsymbol{Q}^u)^{-1}]_{ij}$ by $[(\boldsymbol{Q}^u)^{-1}]_{ij}$ and $[\boldsymbol{M}^l(\boldsymbol{Q}^l)^{-1}]_{ij}$ by $[(\boldsymbol{Q}^l)^{-1}]_{ij}$.

5.3.2 Markov processes

For Markov processes the time spent in a state is negative exponentially distributed so this is a special case covered in section 5.3.1 with $\kappa_i = 0$ and the moments are given through (5.16)-(5.22) by replacing χ_i by $\kappa_i = 0$ and $[\boldsymbol{M}^u(\boldsymbol{Q}^u)^{-1}]_{ij}$ by $[(\boldsymbol{Q}^u)^{-1}]_{ij}$ and $[\boldsymbol{M}^l(\boldsymbol{Q}^l)^{-1}]_{ij}$ by $[(\boldsymbol{Q}^l)^{-1}]_{ij}$.

5.3.3 Birth-death semi-Markov processes

For a birth-death process we only allow jumps to the neighbouring states which means that the generator is on the form

$$\boldsymbol{P}(t) = \begin{bmatrix} 0 & H_0(t) & 0 & 0 & \dots & 0 & 0 & 0 \\ d_1H_1(t) & 0 & b_1H_1(t) & 0 & \dots & 0 & 0 & 0 \\ 0 & d_2H_2(t) & 0 & b_2H_2(t) & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & d_{N-1}H_{N-1}(t) & 0 & b_{N-1}H_{N-1}(t) \\ 0 & 0 & 0 & 0 & \dots & 0 & H_N(t) & 0 \end{bmatrix} \tag{5.25}$$

where $b_i + d_i = 1$, $i = 1, ..., N-1$ and we obtain the $\boldsymbol{Q}$-matrix on the well known form:

$$\boldsymbol{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & ... & 0 & 0 & 0 \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & ... & 0 & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & \mu_{N-1} & -(\lambda_{N-1}+\mu_{N-1}) & \lambda_{N-1} \\ 0 & 0 & 0 & 0 & ... & 0 & \mu_N & -\mu_N \end{bmatrix} \tag{5.26}$$

where $\lambda_0 = \dfrac{1}{m_0}$ and $\mu_i = \dfrac{d_i}{m_i}$, $\lambda_i = \dfrac{b_i}{m_i}$, $i = 1, ..., N-1$ and $\mu_N = \dfrac{1}{m_N}$

For birth-death processes, upper and lower $\boldsymbol{Q}$-matrix may be inverted explicitly and the second order moments greatly simplified. We obtain

$$\boldsymbol{E}[T_k^{\,2}] = \frac{1}{\Delta}\left\{ \sum_{j>j_C} p_j\kappa_j + 2\sum_{i>j_C} \frac{1}{p_i\mu_i}\left(\sum_{j\geq i} p_i\right)^2 \right\} \text{ and} \tag{5.27}$$

$$\boldsymbol{E}[S_k^{\,2}] = \frac{1}{\Delta}\left\{ \sum_{j\leq j_C} p_j\kappa_j + 2\sum_{i\leq j_C} \frac{1}{p_i\lambda_i}\left(\sum_{j\leq i} p_i\right)^2 \right\}, \text{ and} \tag{5.28}$$

$$\boldsymbol{E}[A_k^{\,2}] = \frac{1}{\Delta}\left\{ \sum_{j>j_C} (jb-C)^2 p_j\kappa_j + 2\sum_{i>j_C} \frac{1}{p_i\mu_i}\left(\sum_{j\geq i} (jb-C)p_i\right)^2 \right\} \text{ and} \tag{5.29}$$

$$\boldsymbol{E}[V_k^{\,2}] = \frac{1}{\Delta}\left\{ \sum_{j\leq j_C} (C-jb)^2 p_j\kappa_j + 2\sum_{i\leq j_C} \frac{1}{p_i\lambda_i}\left(\sum_{j\leq i} (C-jb)p_i\right)^2 \right\}. \tag{5.30}$$

Also the up crossing intensity for this case reduces to:

$$\Delta = \mu_{j_C+1} p_{j_C+1} = \lambda_{j_C} p_{j_C}. \tag{5.31}$$

Finally we also mention that the steady state probability for this particular case is the product solution:

$$p_i = p_0 \prod_{s=1}^{i} \left(\frac{\lambda_{s-1}}{\mu_s}\right) \text{ and } p_0 = \frac{1}{1 + \sum_{i\geq 1}\prod_{s=1}^{i}\left(\frac{\lambda_{s-1}}{\mu_s}\right)} \tag{5.32}$$

As a remark we observe that for birth-death processes with negative exponentially distributed sojourn times in the different states; that is $\kappa_j = 0$ in formulae (5.27)-(5.30), then the

excess times and excess volumes will have hyper-exponential characteristics. For example by comparing the first and second moment of the excess time (above the capacity level) we find:

$\boldsymbol{E}[T_k^{\,2}] = \frac{2}{\Delta}\left\{\sum_{i>j_C}\frac{1}{p_i\mu_i}\left(\sum_{j\geq i}p_i\right)^2\right\} > \frac{2}{\Delta}\left\{\frac{1}{p_{j_C+1}\mu_{j_C+1}}\left(\sum_{j\geq j_C+1}p_i\right)^2\right\} = 2(\boldsymbol{E}[T_k])^2$. Similarly we also get $\boldsymbol{E}[S_k^{\,2}] > 2(\boldsymbol{E}[S_k])^2$, and for the volume we have: $\boldsymbol{E}[A_k^{\,2}] > 2(\boldsymbol{E}[A_k])^2$ and $\boldsymbol{E}[V_k^{\,2}] > 2(\boldsymbol{E}[V_k])^2$.

5.4 Distribution of the excess times and excess volumes

Although we have been able to obtain the two lowest moments for the excess distributions, we are not quite satisfied with only these results. In many applications in communication systems it is often the tail of the distributions that gives the best performance measure. As we have pointed out above it seems that the excess distributions are quite “long tailed” (or heavy tailed”). It is therefore of general interest to obtain the full distributions of the excess periods and the volume of information that may be lost during excess periods. By the framework described in chapter 3 and the results derived above, we may express these distributions in terms of the steady state probabilities, the $\boldsymbol{Q}$-matrix and the upper (and lower) renewal kernels for the semi-Markov process.

5.4.1 General formulae for semi-Markov processes

Differentiating equation (5.7) gives $\frac{dF_C}{dt}(i,j,t) = \{[\boldsymbol{Q}^u\boldsymbol{R}^u(t)]_{ij}\}*\{H_j(t)\}$ and by multiplying with the steady state probabilities and summing over the appropriate states we may write the CDF of the excess time as:

$$F_{T_k}(t) = P(T_k > t) = \frac{\boldsymbol{p}^u\cdot\boldsymbol{Q}^u\cdot\boldsymbol{R}^u(t)*\boldsymbol{H}^u(t)}{\boldsymbol{p}^u\cdot\boldsymbol{Q}^u\cdot\boldsymbol{e}^u} \tag{5.33}$$

where $\boldsymbol{H}(t)$ is a column-vector with elements $H_j(t)$ and $\boldsymbol{e}$ is a column-vector with ones.

The CDF of the length of the “normal load” periods is obtained by just substituting for the lower part of the matrices:

$$F_{S_k}(t) = P(S_k > t) = \frac{\boldsymbol{p}^l\cdot\boldsymbol{Q}^l\cdot\boldsymbol{R}^l(t)*\boldsymbol{H}^l(t)}{\boldsymbol{p}^l\cdot\boldsymbol{Q}^l\cdot\boldsymbol{e}^l} \tag{5.34}$$

We get the PDF for T_k and S_k by differentiating (5.33) and (5.34):

$$f_{T_k}(t) = -\frac{dF_{T_k}}{dt}(t) = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{R}^u(t) * \Pi^{u,l}(t) \cdot \boldsymbol{e}^l}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ and} \tag{5.35}$$

$$f_{S_k}(t) = -\frac{dF_{S_k}}{dt}(t) = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{R}^l(t) * \Pi^{l,u}(t) \cdot \boldsymbol{e}^u}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l} \tag{5.36}$$

Sometimes it is desired to introduce the LST to obtain specific results. This may often be effective to determine the tail of the distributions. Since the upper and lower renewal kernels are a sum of convolutions, the LSTs are easy to get from the expressions (5.33), (5.34), (5.35), and (5.36), and (5.5) and (5.6). We find:

$$\hat{F}_{T_k}(s) = \int_{t=0}^{\infty} e^{-st} P(T_k > t)dt = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [\boldsymbol{I} - \hat{\Pi}^u(s)]^{-1} \cdot \hat{\boldsymbol{H}}^u(s)}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u}, \tag{5.37}$$

$$\hat{F}_{S_k}(s) = \int_{t=0}^{\infty} e^{-st} P(S_k > t)dt = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot [\boldsymbol{I} - \hat{\Pi}^l(s)]^{-1} \cdot \hat{\boldsymbol{H}}^l(s)}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l}, \tag{5.38}$$

$$\hat{f}_{T_k}(s) = \int_{t=0}^{\infty} e^{-st} f_{T_k}(t)dt = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [\boldsymbol{I} - \hat{\Pi}^u(s)]^{-1} \cdot \hat{\Pi}^{u,l}(s) \cdot \boldsymbol{e}^l}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ and} \tag{5.39}$$

$$\hat{f}_{S_k}(s) = \int_{t=0}^{\infty} e^{-st} f_{S_k}(t)dt = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot [\boldsymbol{I} - \hat{\Pi}^l(s)]^{-1} \cdot \hat{\Pi}^{l,u}(s) \cdot \boldsymbol{e}^u}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l}, \tag{5.40}$$

where $\hat{\Pi}(s)$ is the LST of $\Pi(t)$ and $\hat{\boldsymbol{H}}(s)$ is the LST of $\boldsymbol{H}(t)$. Since we have $\hat{\Pi}(s) = \boldsymbol{P}(0) - s\hat{\boldsymbol{P}}(s)$ where $\hat{\boldsymbol{P}}(s)$ is the Laplace transform of the generator $\boldsymbol{P}(t)$ we find:

$$\hat{\boldsymbol{H}}(s) = \hat{\boldsymbol{P}}(s) \cdot \boldsymbol{e} = \frac{1}{s}(1 - \hat{\Pi}(s) \cdot \boldsymbol{e}) \tag{5.41}$$

When we consider the excess volume we use of the fact that the bit rate is constant between to consecutive jumps in the rate process, and therefore we may quite easily obtain the distribution of the excess volume between two subsequent jumps. We let

$U_n^u = \int_{T_n}^{T_{n+1}} [\boldsymbol{B}_t - C]dt = (\boldsymbol{B}_t - C)(T_{n+1} - T_n)$ when $\boldsymbol{B}_t > C$ in the interval (T_n, T_{n+1}) and

$U_n^l = \int_{T_n}^{T_{n+1}} [C - \boldsymbol{B}_t]dt = (C - \boldsymbol{B}_t)(T_{n+1} - T_n)$ when $\boldsymbol{B}_t \le C$ in the interval (T_n, T_{n+1}), and we define the following joint probabilities:

$$U_{ij}^u(x) = \boldsymbol{P}\{U_n^u > x, B_{n+1} = jb | B_n = ib\} \text{ for } i > j_C \text{ and} \tag{5.42}$$

$$U_{ij}^l(x) = \boldsymbol{P}\{U_k^l > x, B_{n+1} = jb | B_n = ib\} \text{ for } i \le j_C \tag{5.43}$$

From the these relation we obtain the following joint probabilities:

$$U_{ij}^u(x) = \boldsymbol{P}\{(ib - C)(T_{n+1} - T_n) > x, B_{n+1} = jb | B_n = ib\} = P_{ij}\left(\frac{x}{ib - C}\right) \tag{5.44}$$

for $i > j_C$ and

$$U_{ij}^l(x) = \boldsymbol{P}\{(C - ib)(T_{n+1} - T_n) > x, B_{n+1} = jb | B_n = ib\} = P_{ij}\left(\frac{x}{C - ib}\right) \tag{5.45}$$

for $i \le j_C$. (If $C - j_C b = 0$ then (5.45) will give $U_{ij_C}^l(x) = 0$ so this case will be included in the “below” states if we just skip that state in the matrices.)

By considering the evolution of the U_k^u or U_k^l in either an excess period or a normal load period we will obtain the desired distributions of the excess volume by consider a parallel semi-Markov process with the generator $\boldsymbol{U}(x)$ by:

$$U_{ij}(x) = \begin{cases} U_{ij}^u(x) & for \;\; i > j_C \\ U_{ij}^l(x) & for \;\; i \le j_C \end{cases} \tag{5.46}$$

Since $\boldsymbol{U}(0) = \boldsymbol{P}(0)$ the two semi-Markov processes will have the same $\boldsymbol{Q}$-matrix and the steady state distributions. We are therefore in a position to write the CDFs and PDFs for the excess volumes by using the corresponding results as for the excess times:

$$F_{A_k}(x) = P(A_k > x) = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{S}^u(x) * \boldsymbol{G}^u(x)}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \tag{5.47}$$

$$F_{V_k}(x) = P(V_k > x) = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{S}^l(x) * \boldsymbol{G}^l(x)}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l} \tag{5.48}$$

$$f_{A_k}(x) = -\frac{dF_{A_k}}{dx}(x) = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{S}^u(x) * \Xi^{u,l}(x) \cdot \boldsymbol{e}^l}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ and} \tag{5.49}$$

$$f_{V_k}(x) = -\frac{dF_{V_k}}{dx}(x) = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{S}^l(x) * \Xi^{l,u}(x) \cdot \boldsymbol{e}^u}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l} \tag{5.50}$$

where $\boldsymbol{S}^u(x)$ and $\boldsymbol{S}^l(x)$ are the corresponding upper and lower renewal kernels:

$$\boldsymbol{S}^u(x) = \boldsymbol{I}^u \delta(x) + \Xi^u(x) + \{\Xi^u(x)\}^{*(2)} + ... \text{ and} \tag{5.51}$$

$$\boldsymbol{S}^l(x) = \boldsymbol{I}^l \delta(x) + \Xi^l(x) + \{\Xi^l(x)\}^{*(2)} + ... \tag{5.52}$$

$\Xi(x) = -\frac{d\boldsymbol{U}}{dx}(x)$ and $\boldsymbol{G}(x) = \boldsymbol{U}(x) \cdot \boldsymbol{e}$. From (5.44) and (5.45) get the following relation between the between $\Xi(x)$ and $\Pi(t)$, and $\boldsymbol{G}(x)$ and $\boldsymbol{H}(t)$:

$$\Xi_{ij}^u(x) = \frac{1}{ib - C}\Pi_{ij}\left(\frac{x}{ib - C}\right) \text{ for } i > j_C,\ j > j_C,\ \Xi_{ij}^{u,l}(x) = \frac{1}{ib - C}\Pi_{ij}\left(\frac{x}{ib - C}\right) \text{ for } i > j_C,\ j \le j_C, \tag{5.53}$$

$$\text{and } G_i^u(x) = H_i\left(\frac{x}{ib - C}\right) \text{ for } i > j_C \text{ and} \tag{5.54}$$

$$\Xi_{ij}^l(x) = \frac{1}{C - ib}\Pi_{ij}\left(\frac{x}{C - ib}\right) \text{ for } i \le j_C,\ j \le j_C,\ \Xi_{ij}^{l,u}(x) = \frac{1}{C - ib}\Pi_{ij}\left(\frac{x}{C - ib}\right) \text{ for } i \le j_C,\ j > j_C, \tag{5.55}$$

$$\text{and } G_i^l(x) = H_i\left(\frac{x}{C - ib}\right) \text{ for } i \le j_C \tag{5.56}$$

Finally we may write the LSTs of the excess volumes on the following form:

$$\hat{F}_{A_k}(\zeta) = \int_{x=0}^{\infty} e^{-\zeta x} P(A_k > x)dx = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [\boldsymbol{I} - \hat{\Xi}^u(\zeta)]^{-1} \cdot \hat{\boldsymbol{G}}^u(\zeta)}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u}, \tag{5.57}$$

$$\hat{F}_{V_k}(\zeta) = \int_{x=0}^{\infty} e^{-\zeta x} P(V_k > x)dx = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot [\boldsymbol{I} - \hat{\Xi}^l(\zeta)]^{-1} \cdot \hat{\boldsymbol{G}}^l(\zeta)}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l}, \tag{5.58}$$

$$\hat{f}_{A_k}(\zeta) = \int_{t=0}^{\infty} e^{-\zeta x} f_{A_k}(x)dx = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [\boldsymbol{I} - \hat{\Xi}^u(\zeta)]^{-1} \cdot \hat{\Xi}^{u,l}(\zeta) \cdot \boldsymbol{e}^l}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ and} \tag{5.59}$$

$$\hat{f}_{V_k}(\zeta) = \int_{t=0}^{\infty} e^{-\zeta x} f_{V_k}(x)dx = \frac{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot [\boldsymbol{I} - \hat{\Xi}^l(\zeta)]^{-1} \cdot \hat{\Xi}^{l,u}(\zeta) \cdot \boldsymbol{e}^u}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l}. \tag{5.60}$$

By using the functional relation between $\Xi(x)$ and $\Pi(t)$, and $\boldsymbol{G}(x)$ and $\boldsymbol{H}(t)$ given above by (5.53) and (5.55) we obtain:

$$\hat{\Xi}^u_{ij}(\zeta) = \hat{\Pi}_{ij}((ib-C)\zeta) \text{ for } i > j_C,\ j > j_C,\ \hat{\Xi}^{u,l}_{ij}(\zeta) = \hat{\Pi}_{ij}((ib-C)\zeta) \text{ for } i > j_C,$$
$$j \le j_C, \tag{5.61}$$

$$\text{and } \hat{G}^u_i(\zeta) = (ib-C)\hat{H}_i((ib-C)\zeta) \text{ for } i > j_C \text{ and} \tag{5.62}$$

$$\hat{\Xi}^l_{ij}(\zeta) = \hat{\Pi}_{ij}((C-ib)\zeta) \text{ for } i \le j_C, j \le j_C, \hat{\Xi}^{l,u}_{ij}(\zeta) = \hat{\Pi}_{ij}((C-ib)\zeta) \text{ for } i \le j_C, j > j_C \tag{5.63}$$

$$\text{and } \hat{G}^l_i(\zeta) = (C-ib)\hat{H}_i((C-ib)\zeta) \text{ for } i \le j_C. \tag{5.64}$$

The derived formulae for the excess times (5.33), (5.34), (5.35) and (5.36) and the formulae (5.47), (5.48), (5.49) and (5.50) for the excess volumes together with the corresponding LST (5.37), (5.38), (5.39), (5.40) and (5.57), (5.58), (5.59), (5.60) constitute to our knowledge new development in the effort to describe transient phenomena for quite general traffic models as semi-Markov processes. Knowing the increasing variety of new services expected in future networks, the modelling and understanding of transient behaviour will be important especially in connection with congestion phenomena. To carry the analysis any further one has to be more specific about the processes, this could for instance be done by assuming the broad class of models where the time spent in the different states are a mixture of exponential distributions (phase type) and then looking for possible poles in the expressions for the LSTs. (We will skip such an analysis here, it would be merely technical, - although it logically would be straight forward to perform.)

5.4.2 Spectral decomposition of the distributions

To perform a spectral decomposition of some of the distributions above we must assume that it is possible to diagonalize the corresponding matrix. Below we shall just sketch the possible analysis. We pick the excess time T_k as the distribution under consideration. (The decomposition for the other variable will be entirely similar.) It is possible to rewrite the Laplace transform (5.37) in the following way:

$$\hat{F}_{T_k}(s) = \int_{t=0}^{\infty} e^{-st} P(T_k > t)dt = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [\hat{\Gamma}^u(s)]^{-1} \cdot \boldsymbol{e}^u}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ where} \tag{5.65}$$

the matrix $\hat{\Gamma}^u(s)$ is given as:

$$\hat{\Gamma}^u(s) = Diag\left[\frac{1}{\hat{H}_i(s)}\right] \cdot (\boldsymbol{I} - \hat{\Pi}^u(s)) \tag{5.66}$$

In the succeeding we assume that it is possible to diagonalize the matrix $\hat{\Gamma}^u(s)$, which means that we may write:

$$\hat{\Gamma}^u(s) = \boldsymbol{Y}^u(s) \cdot Diag[\gamma_i^u(s)] \cdot \boldsymbol{X}^u(s) \text{ where} \tag{5.67}$$

$\gamma_i^u(s)$ are the eigenvalues of $\hat{\Gamma}^u(s)$ and the matrices $\boldsymbol{X}^u(s)$ and $\boldsymbol{Y}^u(s)$ constitute the left and right eigenvectors:

$\boldsymbol{X}^u(s) = \begin{bmatrix} \boldsymbol{x}_1^u(s) \\ \boldsymbol{x}_2^u(s) \\ . \end{bmatrix}$ where the row-vectors $\boldsymbol{x}_i^u(s)$ are the solution of the linear equations

$\boldsymbol{x}_i^u(s) \cdot \hat{\Gamma}^u(s) = \gamma_i^u(s)\boldsymbol{x}_i^u(s)$, and $\boldsymbol{Y}^u(s) = [\boldsymbol{y}_1^u(s), \boldsymbol{y}_2^u(s), ...]$ where the column-vectors $\boldsymbol{y}_i^u(s)$ are the solutions of $\hat{\Gamma}^u(s) \cdot \boldsymbol{y}_i^u(s) = \gamma_i^u(s)\boldsymbol{y}^u{}_i(s)$. We also assumes the eigenvectors are normalised so that $\boldsymbol{x}_i^u(s) \cdot \boldsymbol{y}_i^u(s) = 1$. With these rather technical details we may express the brackets in the Laplace transform (5.65) as:

$$[\hat{\Gamma}^u(s)]^{-1} = \boldsymbol{Y}^u(s) \cdot Diag\left[\frac{1}{\gamma_i^u(s)}\right] \cdot \boldsymbol{X}^u(s). \tag{5.68}$$

It is possible to carry the analysis further if we assume that singularities of the Laplace transform are simple poles. By (5.68) we have that the poles of the Laplace transform (5.37) located as roots of the equations

$$\gamma_i^u(s) = 0 \tag{5.69}$$

and we denote $s = \lambda_i^l$, $l = 1, 2, ...$ for these roots which we for simplicity assume to be distinct. The inversion of the Laplace transform can therefore be carried out by the residue theorem giving the following expression for the CDF for T_k:

$$F_{T_k}(t) = P(T_k > t) = \sum_i \sum_{l=1} R_i(\lambda_i^l) e^{\lambda_i^l t} \text{ where} \tag{5.70}$$

$R_i(\lambda_i^l)$ is the residue of the transform taken at $s = \lambda_i^l$, and may be written (in terms of the eigenvalues and eigenvectors) by:

$$R_i(s) = \frac{\boldsymbol{x}_i^u(s) \cdot \boldsymbol{e}^u}{\gamma_i^{u\prime}(s)} \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{y}_i^u(s)}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \tag{5.71}$$

In Appendix C we have shown that $Re(\lambda_i^l) < 0$ for all the roots, which guarantees the proper behaviour of the distribution for t is large.

One important outcome of the spectral decomposition is the asymptotic behaviour for large t. If we let $s = \lambda_1^1$ be the dominant root (largest real part) of (5.69). Then we get from (5.70):

$$F_{T_k}(t) = P(T_k > t) \sim R_1(\lambda_1^1) e^{\lambda_1^1 t} \quad \text{for} \quad t \to \infty \tag{5.72}$$

knowing that $Re(\lambda_1^1) < 0$ and where $R_1(\lambda_1^1)$ is the residue (5.71) taken at $s = \lambda_1^1$.

It may be worth to mention some of the problems that the spectral decomposition may cause. If the number of states is infinite the diagonalization procedure may fail. This is more or less caused by the fact that the number of eigenvalues also will be infinite and the spacing of the roots is difficult to predict. In some cases this implies that the sum in (5.70) will be an integral. Another problem that may occur is of cause numerical difficulties to find the left and right eigenvectors and to find the derivative of all the eigenvalues. Even more difficult is it to invert the corresponding LSTs if other forms of singularities occur for instance branch cuts. For such cases one has to apply specific methods and will not be discussed in this thesis.

5.4.3 Birth-death semi-Markov processes

Birth-death semi-Markov processes impose some simplifying qualities which may be worth to take into account when considering the excess distributions, and enable carrying the analysis some further. This is because of the special structure of the generator matrix which will be tri-diagonal. Taking this special structure into account it is possible to invert the matri-

ces in the Laplace transforms of the excess times (5.39) and (5.40). After some manipulations we obtain:

$$\hat{f}_{T_k}(s) = d_{j_C+1}\frac{D^u_{j_C+2}(s)}{D^u_{j_C+1}(s)} \text{ where} \tag{5.73}$$

$D^u_j(s)$ is the determinant obtained from the j-th upper-sub matrix of $\boldsymbol{I}-\hat{\Pi}(s)$:

$$D^u_j(s) = Det\begin{bmatrix} \frac{1}{\hat{f}_j(s)} & -b_j & 0 & \dots & 0 & 0 & 0 \\ -d_{j+1} & \frac{1}{\hat{f}_{j+1}(s)} & -b_{j+1} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -d_{N-1} & \frac{1}{\hat{f}_{N-1}(s)} & -b_{N-1} \\ 0 & 0 & 0 & \dots & 0 & -1 & \frac{1}{\hat{f}_N(s)} \end{bmatrix} \text{ and} \tag{5.74}$$

$$\hat{f}_{S_k}(s) = b_{j_C}\frac{D^l_{j_C-1}(s)}{D^l_{j_C}(s)} \text{ where} \tag{5.75}$$

$D^l_j(s)$ is the determinant obtained from the j-th lower-sub matrix of $\boldsymbol{I}-\hat{\Pi}(s)$:

$$D^l_j(s) = Det\begin{bmatrix} \frac{1}{\hat{f}_0(s)} & -1 & 0 & \dots & 0 & 0 & 0 \\ -d_1 & \frac{1}{\hat{f}_1(s)} & -b_1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -d_{j-1} & \frac{1}{\hat{f}_{j-1}(s)} & -b_{j-1} \\ 0 & 0 & 0 & \dots & 0 & -d_j & \frac{1}{\hat{f}_j(s)} \end{bmatrix} \text{ where} \tag{5.76}$$

$\hat{f}_j(s)$ is the LST of the sojourn time PDFs $f_j(t) = -\frac{dH_j}{dt}(t)$ in state j. The effectiveness of the formulae above is due to the tri-diagonal form of the involved matices, which allows for recursive evaluations. Expanding the determinants (after the last row for $D^u_j(s)$ and the first row for $D^l_j(s)$) we find the following iterative expressions:

$$D_j^u(s) = \frac{1}{\hat{f}_j(s)} D_{j+1}^u(s) - b_j d_{j+1} D_{j+2}^u(s) \text{ for } j = N-1, ..., j_C + 1 \tag{5.77}$$

starting by defining $D_{N+1}^u(s) = 1$ and $D_N^u(s) = \frac{1}{\hat{f}_N(s)}$ and taking $d_N = 1$ and

$$D_j^l(s) = \frac{1}{\hat{f}_j(s)} D_{j-1}^u(s) - b_{j-1} d_j D_{j-2}^l(s) \text{ for } j = 1, ..., j_C \tag{5.78}$$

starting by defining $D_{-1}^l(s) = 1$ and $D_0^l(s) = \frac{1}{\hat{f}_0(s)}$ and taking $b_0 = 1$.

The corresponding results for the excess volumes are:

$$\hat{f}_{A_k}(\zeta) = d_{j_C+1} \frac{\Delta_{j_C+2}^u(\zeta)}{\Delta_{j_C+1}^u(\zeta)} \text{ where } \Delta_j^u(s) \text{ is the determinant:} \tag{5.79}$$

$$\Delta_j^u(\zeta) = Det \begin{bmatrix} \frac{1}{\varphi_j^u(\zeta)} & -b_j & 0 & ... & 0 & 0 & 0 \\ -b_{j+1} & \frac{1}{\varphi_{j+1}^u(\zeta)} & -b_{j+1} & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -b_{N-1} & \frac{1}{\varphi_{N-1}^u(\zeta)} & -b_{N-1} \\ 0 & 0 & 0 & ... & 0 & -1 & \frac{1}{\varphi_N^u(\zeta)} \end{bmatrix} \tag{5.80}$$

where $\varphi_j^u(\zeta) = \hat{f}_j((jb - C)\zeta)$; $j > j_C$ and

$$\hat{f}_{V_k}(\zeta) = b_{j_C} \frac{\Delta_{j_C-1}^l(\zeta)}{\Delta_{j_C}^l(\zeta)} \text{ where } \Delta_j^l(s) \text{ is the determinant:} \tag{5.81}$$

$$\Delta_j^l(\zeta) = Det \begin{bmatrix} \frac{1}{\varphi_0^l(\zeta)} & -1 & 0 & ... & 0 & 0 & 0 \\ -d_1 & \frac{1}{\varphi_1^l(\zeta)} & -b_1 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -d_{j-1} & \frac{1}{\varphi_{j-1}^l(\zeta)} & -b_{j-1} \\ 0 & 0 & 0 & ... & 0 & -d_j & \frac{1}{\varphi_j^l(\zeta)} \end{bmatrix} \tag{5.82}$$

where $\varphi_j^l(\zeta) = \hat{f}_j((C-jb)\zeta)\,;\, j \le j_C$

The corresponding recursion formulae for the determinants in expressions (5.80) and (5.82) for the excess volumes are:

$$\Delta_j^u(\zeta) = \frac{1}{\varphi_j^u(\zeta)}\Delta_{j+1}^u(\zeta) - b_j d_{j+1}\Delta_{j+2}^u(\zeta) \text{ for } j = N-1, ..., j_C+1 \tag{5.83}$$

starting by defining $\Delta_{N+1}^u(s) = 1$ and $\Delta_N^u(\zeta) = \dfrac{1}{\varphi_N^u(\zeta)}$ and taking $d_N = 1$ and

$$\Delta_j^l(\zeta) = \frac{1}{\varphi_j^l(\zeta)}\Delta^u{}_{j-1}(\zeta) - b_{j-1} d_j \Delta_{j-2}^l(\zeta) \text{ for } j = 1, ..., j_C \tag{5.84}$$

starting by defining $\Lambda_{-1}^l(s) = 1$ and $\Delta_0^l(\zeta) = \dfrac{1}{\varphi_0^l(\zeta)}$ and taking $b_0 = 1$.

Based on the recursions above it is easy to calculate the LST for the excess times and excess volumes even for large dimensions. However, often forced by the difficult problems of numerical inversion of LSTs, one will often try to invert the transforms analytically. This is usually done by locating the singularity (poles) of the transforms. In this case we must evaluate the corresponding residue at the singular points. Let us for simplicity assume that the singularity is a simple pole. To calculate the residue we need also to calculate the derivative of the nominator in the transforms. This can be done by taking the derivative of the recursion formulae (5.77), (5.78), (5.83) or (5.84). If we choose $D_j^u(s)$ we find

$$D_j^u{}'(s) = -\frac{\hat{f}_j{}'(s)}{\hat{f}_j(s)^2} D_{j+1}^u(s) + \frac{1}{\hat{f}_j(s)} D_{j+1}^u{}'(s) - b_j d_{j+1} D_{j+2}^u{}'(s) \tag{5.85}$$

for $j = N-1, ..., j_C+1$ starting with $D_{N+1}^u{}'(s) = 0$ and $D_N^u{}'(s) = -\dfrac{\hat{f}_N{}'(s)}{\hat{f}_N(s)^2}$

Together with (5.77), (5.85) will give the pair $(D_j^u(s), D_j^u{}'(s))$ for $j = N, ..., j_C+1$.

It is possible to rewrite the formulae for the LSTs for birth-death semi-Markov processes given above. This can be done by rewriting the determinants (5.74), (5.76), (5.80) and (5.82), so that the matrices are symmetric. This can be done by pre- and post-multiplying by a given diagonal matrix and its inverse. We have put the analysis in Appendix D where we obtain the corresponding determinants as the product of all the eigenvalues and moreover the required eigenvalues may be found by the powerful method of bisection by applying the “so-called” Sturm sequence property of leading principal minor of the symmetric

matrices. This gives an alternative method to find the LSTs to those described by the recursion formulas (5.77), (5.78) and (5.83), (5.84). The latter is especially effective if we consider birth-death processes with exponential distributed sojourn times. In this case we also find the corresponding residue and we show that the excess distribution fully is determined by the eigenvalues of the principal minor of order $N - j_C - 1$ and $N - j_C - 2$ for the excess distributions and the principal minor of order j_C and $j_C - 1$ for the time and volume below the capacity C.

5.4.4 Markov processes

As mentioned earlier a Markov process is a special case of a semi-Markov process where the time spent in a particular state is negative exponentially distributed. For this case it is possible to carry out the analysis in section 5.4.1 further. Recall that for a Markov process we have the generator (5.1) on the form:

$$P_{ij}(t) = P_{ij} e^{-\gamma_i t} \tag{5.86}$$

where $m_i = \dfrac{1}{\gamma_i}$ is the mean sojourn time in state i and P_{ij} are the transition probabilities for the corresponding jump Markov-chain, where we also assume that $P_{ii} = 0$. (See Cinlar for a discussion of related topics [Cinl75]). The $\boldsymbol{Q}$-matrix for the Markov process will therefore be on the form:

$$Q_{ii} = -\gamma_i \text{ and } Q_{ij} = \gamma_i P_{ij} \text{ for } j \neq i \tag{5.87}$$

Working with the LSTs we obtain: $\hat{\Pi}_{ij}(s) = \dfrac{Q_{ij}}{s + \gamma_i}$ for $j \neq i$ $(\Pi_{ii}(s) = 0)$ and $H_i(s) = \dfrac{1}{s + \gamma_i}$, Inserting these simplifications for example in (5.37) we get:

$$\hat{F}_{T_k}(s) = \int_{t=0}^{\infty} e^{-st} P(T_k > t) dt = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot [s\boldsymbol{I} - \boldsymbol{Q}^u]^{-1} \cdot \boldsymbol{e}^u}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \tag{5.88}$$

which we recognize as the LST of the following matrix expression:

$$F_{T_k}(t) = P(T_k > t) = \frac{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot Exp[\boldsymbol{Q}^u t] \cdot \boldsymbol{e}^u}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u}. \tag{5.89}$$

(The result for S_k is completely similar we just change the u-upper to the l-lower matrices.)

By the relations (5.61) and (5.63) we have:

$\hat{\Xi}^u_{ij}(\zeta) = \dfrac{Q_{ij}}{(ib-C)\zeta+\gamma_i}$ and $\hat{G}^u_i(\zeta) = \dfrac{(ib-C)}{(ib-C)\zeta+\gamma_i}$ for $i > j_C$ and

$\hat{\Xi}^l_{ij}(\zeta) = \dfrac{Q_{ij}}{(C-ib)\zeta+\gamma_i}$ and $\hat{G}^l_i(\zeta) = \dfrac{(C-ib)}{(C-ib)\zeta+\gamma_i}$ for $i \le j_C$. Inserting in (5.57) and (5.58) we get:

$$\hat{F}_{A_k}(\zeta) = \int_{x=0}^{\infty} e^{-\zeta x} P(A_k > x)dx = \frac{\boldsymbol{p}^{*u} \cdot \boldsymbol{Q}^{*u} \cdot [\zeta \boldsymbol{I} - \boldsymbol{Q}^{*u}]^{-1} \cdot \boldsymbol{e}^u}{\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ and} \tag{5.90}$$

$$\hat{F}_{V_k}(\zeta) = \int_{x=0}^{\infty} e^{-\zeta x} P(V_k > x)dx = \frac{\boldsymbol{p}^{*l} \cdot \boldsymbol{Q}^{*l} \cdot [\zeta \boldsymbol{I} - \boldsymbol{Q}^{*l}]^{-1} \cdot \boldsymbol{e}^l}{\boldsymbol{p}^l \cdot \boldsymbol{Q}^l \cdot \boldsymbol{e}^l} \tag{5.91}$$

where we have redefined the $\boldsymbol{Q}$-matrices in the brackets by:

$$Q^{*u}_{ij} = \frac{Q_{ij}}{ib-C} \text{ for } i > j_C \text{ and } Q^{*l}_{ij} = \frac{Q_{ij}}{C-ib} \text{ for } i \le j_C, \text{ and} \tag{5.92}$$

$\boldsymbol{p}^{*u}$ is a row-vector with elements $p^{*u}_j = (jb-C)p_j$, $j > j_C$ and $\boldsymbol{p}^{*l}$ as a row-vector with elements $p^{*l}_j = (C-jb)p_j$, $j \le j_C$.

The similarity between the time distributions and the volume distributions are striking. We shall, however, notice that the latter will give quite another type of generating equation if one applies the method of generating functions to find the inverse of the matrix $[\zeta \boldsymbol{I} - \boldsymbol{Q}^{*u}]^{-1}$ compared with $[s\boldsymbol{I} - \boldsymbol{Q}^u]^{-1}$.

For Markov processes the spectral decomposition of the distributions greatly simplifies. We choose the excess time T_k. (The decomposition for the other variable will be similar.) Suppose that all the eigenvalues of $\boldsymbol{Q}^u$ are distinct. By this assumption we may write $\boldsymbol{Q}^u$ as: $\boldsymbol{Q}^u = \boldsymbol{Y}^u Diag[\gamma^u_i] \cdot \boldsymbol{X}^u$ where γ^u_i are the eigenvalues of $\boldsymbol{Q}^u$ and the matrices $\boldsymbol{X}^u$ and $\boldsymbol{Y}^u$ constitute the left and right eigenvectors: $\boldsymbol{X}^u = \begin{bmatrix} \boldsymbol{x}^u_1 \\ \boldsymbol{x}^u_2 \\ . \end{bmatrix}$ where the row-vectors $\boldsymbol{x}^u_i$ are the solution of the linear equations $\boldsymbol{x}^u_i \cdot \boldsymbol{Q}^u = \gamma^u_i \boldsymbol{x}^u_i$, and $\boldsymbol{Y}^u = [\boldsymbol{y}^u_1, \boldsymbol{y}^u_2 .....]$ where the column-vectors $\boldsymbol{y}^u_i$ are the solution of $\boldsymbol{Q}^u \cdot \boldsymbol{y}^u_i = \gamma^u_i \boldsymbol{y}^u_i$. We also assume that

the eigenvectors are normalised, that is $\boldsymbol{x}_i^u \cdot \boldsymbol{y}_i^u = 1$. With these rather technical details we may express the brackets in the Laplace transform (5.88) as:

$$\boldsymbol{Q}^u \cdot [s\boldsymbol{I} - \boldsymbol{Q}^u]^{-1} = \boldsymbol{Y}^u \cdot Diag\left[\frac{\gamma_i^u}{s - \gamma_i^u}\right] \cdot \boldsymbol{X}^u \text{ which} \tag{5.93}$$

shows that the poles of the distribution are located at $s = \gamma_i^u$. By evaluating the corresponding residues we obtain:

$$F_{T_k}(t) = P(T_k > t) = \sum_i R_i e^{\gamma_i^u t} \text{ where } R_i \text{ is the residue:} \tag{5.94}$$

$$R_i = \frac{(\boldsymbol{p}^u \cdot \boldsymbol{y}_i^u)(\boldsymbol{x}_i^u \cdot \boldsymbol{e}^u)}{-\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u}\gamma_i^u \tag{5.95}$$

and $Re(\gamma_i^u) < 0$ by the result in Appendix C.

For Markov processes it is possible to find the moments in terms of the $\boldsymbol{Q}$-matrices by expanding the Laplace transforms. We obtain

$$\boldsymbol{E}[T_k^n] = \frac{n!\boldsymbol{p}^u \cdot [-\boldsymbol{Q}^u]^{-(n-1)} \cdot \boldsymbol{e}^u}{-\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \tag{5.96}$$

For the excess volume the corresponding moments are:

$$\boldsymbol{E}[A_k^n] = \frac{n!\boldsymbol{p}^{*u} \cdot [-\boldsymbol{Q}^{*u}]^{-(n-1)} \cdot \boldsymbol{e}^u}{-\boldsymbol{p}^u \cdot \boldsymbol{Q}^u \cdot \boldsymbol{e}^u} \text{ where} \tag{5.97}$$

(The results for $\boldsymbol{E}[S_k^n]$ and $\boldsymbol{E}[V_k^n]$ are on the same form as (5.96) and (5.97), we just substitute the u with l.)

For birth-death processes the moments may be calculated recursively. This can be shown by defining the row-vector $\xi_m^u = \boldsymbol{p}^u \cdot [-\boldsymbol{Q}^u]^{-m}$, which implies $\xi_m^u = \xi_{m-1}^u \cdot [-\boldsymbol{Q}^u]^{-1}$ or $\xi_m^u[-\boldsymbol{Q}^u] = \xi_{m-1}^u$ for $m = 1, 2, ,....$. From the last equation we may solve explicitly for ξ_m^u in terms of ξ_{m-1}^u, we find:

$$\xi_{im}^u = p_i \sum_{l = j_C + 1}^{i} \frac{1}{\mu_l p_l} \sum_{s \geq l} \xi_{sm-1}^u \text{ for } i = j_C + 1, j_C + 2, ..., \; m = 1, 2, ,.... \tag{5.98}$$

Then by starting with $\xi_{i0}^{u} = p_i$ for $i = j_C + 1, j_C + 2, ...$ and evaluating ξ_{im}^{u} recursively by (5.98), the n-th moment of the excess time distribution is given by:

$$\boldsymbol{E}[T_k^n] = \frac{n!}{\mu_{j_C+1} p_{j_C+1}} \sum_{i > j_C} \xi_{in-1}^{u} \tag{5.99}$$

The corresponding result for the n-th moment of the excess volume is obtained by substituting for $\boldsymbol{p}^{*u}$ and $\boldsymbol{Q}^{*u}$ in the evaluation above giving:

$$\boldsymbol{E}[A_k^n] = \frac{n!}{\mu_{j_C+1} p_{j_C+1}} \sum_{i > j_C} \varsigma_{in-1}^{u} \text{ where } \varsigma_m^{u} \text{ is given recursively by:} \tag{5.100}$$

$$\varsigma_{im}^{u} = p_i \sum_{l = j_C+1}^{i} \frac{lb - C}{\mu_l p_l} \sum_{s \geq l} \varsigma_{sm-1}^{u} \text{ for } i = j_C + 1, j_C + 2, ..., \; m = 1, 2, ,...., \tag{5.101}$$

starting with $\varsigma_{i0}^{u} = (ib - C)p_i$ for $i = j_C + 1, j_C + 2, ...$

The results for $\boldsymbol{E}[S_k^n]$ and $\boldsymbol{E}[V_k^n]$ are similar and we find:

$$\boldsymbol{E}[S_k^n] = \frac{n!}{\lambda_{j_C} p_{j_C}} \sum_{i \leq j_C} \xi_{in-1}^{l} \quad \text{where } \xi_m^{l} \text{ is given recursively by} \tag{5.102}$$

$$\xi_{im}^{l} = p_i \sum_{l = i}^{j_C} \frac{1}{\mu_l p_l} \sum_{s \leq l} \xi_{sm-1}^{l} \text{ for } i = 0, 1, ..., j_C, \; m = 1, 2, ,.... \tag{5.103}$$

starting with $\xi_{i0}^{l} = p_i$ for $i = j_C, j_C - 1, ...$, and

$$\boldsymbol{E}[V_k^n] = \frac{n!}{\lambda_{j_C} p_{j_C}} \sum_{i \leq j_C} \zeta_{in-1}^{l} \quad \text{where } \xi_m^{l} \text{ is given recursively by} \tag{5.104}$$

$$\zeta_{im}^{l} = p_i \sum_{l = i}^{j_C} \frac{C - lb}{\mu_l p_l} \sum_{s \leq l} \zeta_{sm-1}^{l} \text{ for } i = 0, 1, ..., j_C, \; m = 1, 2, ... \tag{5.105}$$

starting with $\zeta_{i0}^{l} = (C - ib)p_i$ for $i = j_C, j_C - 1, ...$.

5.5 Some numerical examples

As an example we shall consider bufferless multiplexing of a given number of N identical ON/OFF sources on a communication link of capacity C. The source model is as follows:

- the ON/OFF periods are mutually independent and exponentially distributed with mean on- and off-time given μ^{-1} and λ^{-1} respectively and
- in ON state the source emits data at rate b

It is well known that superposing a fixed number of this type of sources constitutes a birth-death process with rates:

$$\lambda_i = (N-i)\lambda \text{ and } \mu_i = i\mu \tag{5.106}$$

The limiting state (for which information losses will occur) is $j_C = \left\lfloor \frac{C}{b} \right\rfloor$. We shall apply the framework described in Appendix D for general birth-death processes to find the CDFs for both the excess times and corresponding volumes. All the results are described in terms of the eigenvalues for symmetric tridiagonal matrices through the formulae (D.38)-(D.41). The eigenvalues are computed by applying the method of bisection which turns out to be very effective, and makes it feasible to study system of quite large dimension.

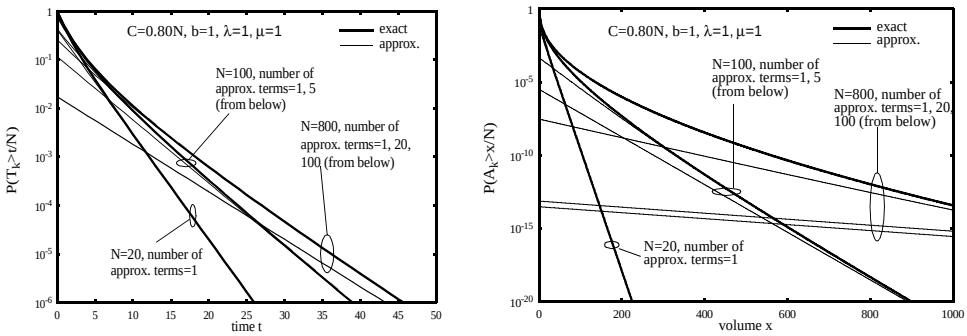

Figure 5.2: CDF of excess times and excess volumes with different numbers of terms in the expansion.

In the first examples given in figures 5.2-5.6 we have chosen $b = \lambda = \mu = 1$ which give the average rate from each source to 0.5.

In figure 5.2 we have tested the numbers of terms that are needed in the eigenvalue expansion of the different CDFs. For systems with relative small number of sources we see that the number of terms needed is small. Actually it seems to be sufficient with only the first term in the expansion both for the time variable and for the volume. However, for larger systems the number of terms to get an accurate approximation becomes quite large and for the case with 800 sources it seems that one needs at least more than 100 terms to get accurate approximations. Thus, it seems that the pure exponential approximation with only the

first term (in the expansion) fails to give accurate approximations in the interested range of the parameters.

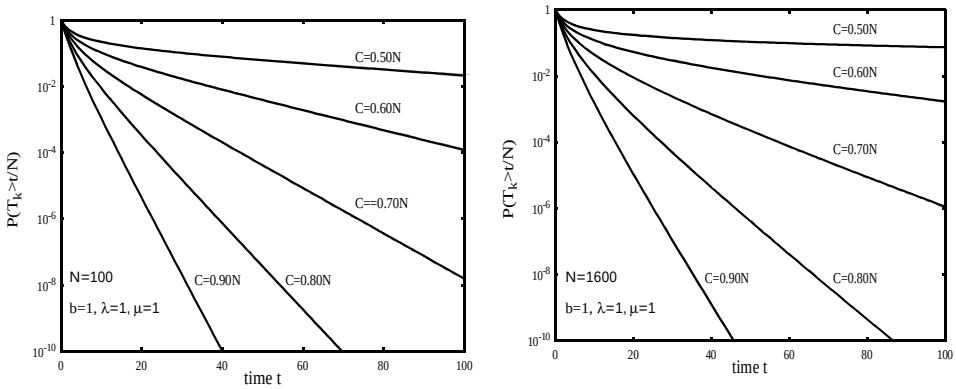

Figure 5.3: CDF of excess times for some different choice of the parameters as function of time.

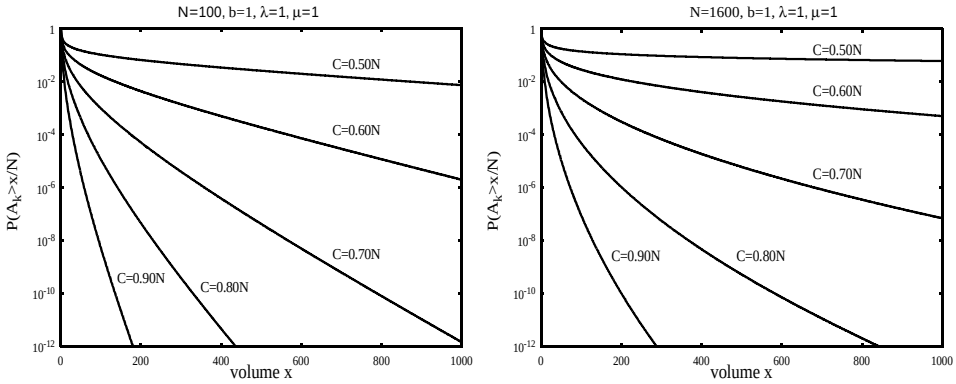

Figure 5.4: CDF of excess volumes for some different choice of the parameters as function of volume.

In figure 5.3 and figure 5.4 we test how the CDF of the excess time depends on the different parameters in the model. As the capacity increases while keeping the numbers of sources constant we observe the changes in the different curves. The change when going from $N = 100$ (left) to $N = 1600$ is not very pronounced but clearly visible. Another observation is that the CDFs for the excess volumes seem to be more curved then the corresponding excess times which actually means the excess volumes are “less exponential” then the corresponding time distributions.

In figure 5.5 and figure 5.6 we demonstrate the convergence of the CDFs by keeping the load constant and increasing the number of sources in both cases the convergence is reached at approximately 400 sources.

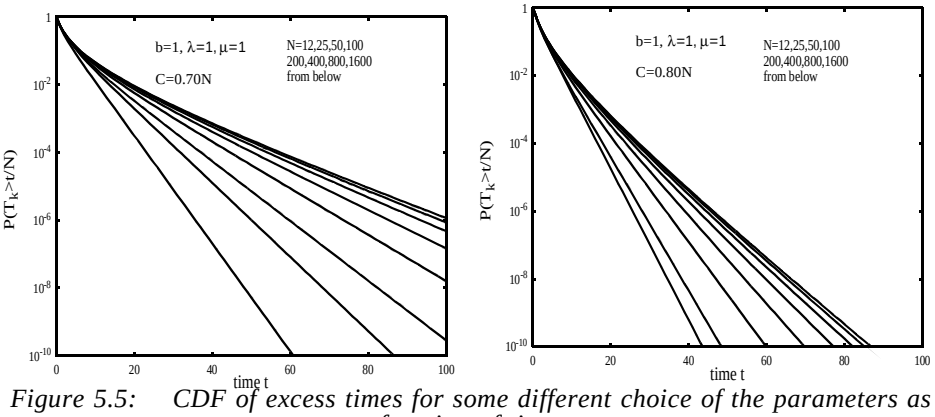

Figure 5.5: CDF of excess times for some different choice of the parameters as function of time.

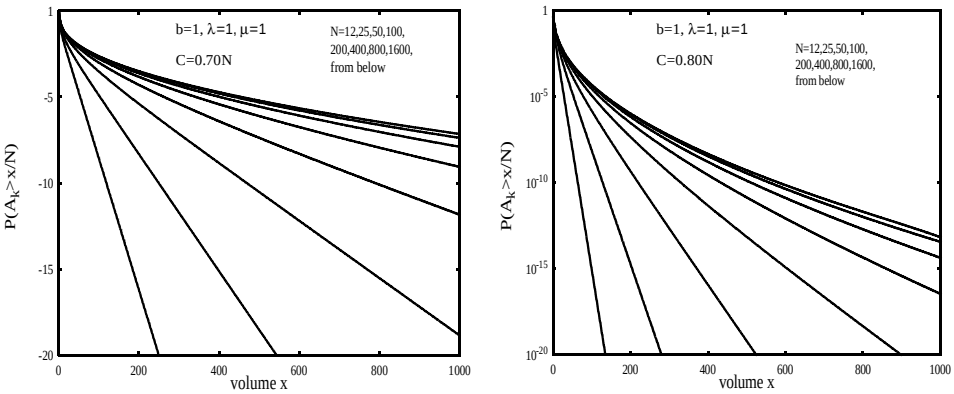

Figure 5.6: CDF of excess volumes for some different choice of the parameters as function of volume.

As a final example we consider sources with parameters as in [Dupu97] where the peak bit rate b=2 Mbps and μ^{-1}=1 ms and λ^{-1}=4.5 ms giving a mean bit rate of 364 Kbps. This source could for instance be a typical ADSL user having access rate of 2 Mbps. The question in mind will then be to try to find out how many sources of this type that can be multiplexed on a high capacity link, and find the typical values of the performance measures described in this section. In the example we have chosen a link with capacity 600 Mbps, and the typical values for number of sources will then be 1200. (We have also included a second example where the link capacity is taken as large as 2.4 Gbps.) In figures 5.7-5.9 we have depicted the excess time CDFs for this example. In the first two graphs we vary both the number of sources and the capacity to see the effect on the different CDFs.

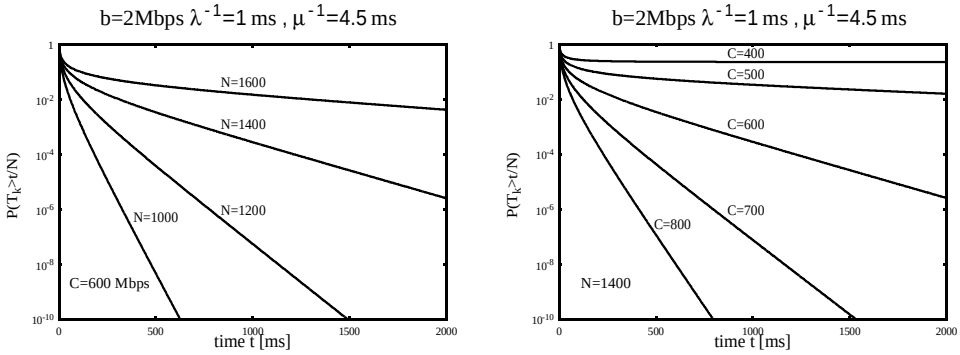

Figure 5.7: CDF of excess times for some different choice of the parameters as function of time.

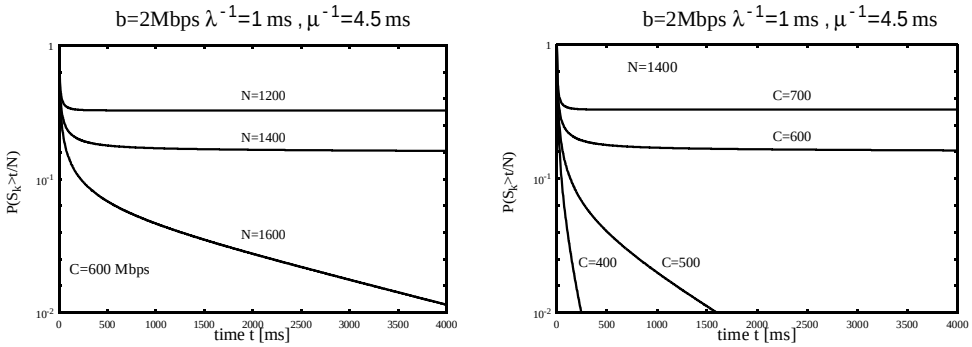

Figure 5.8: CDF of "normal loaded" periods for some different choice of the parameters as function of time.

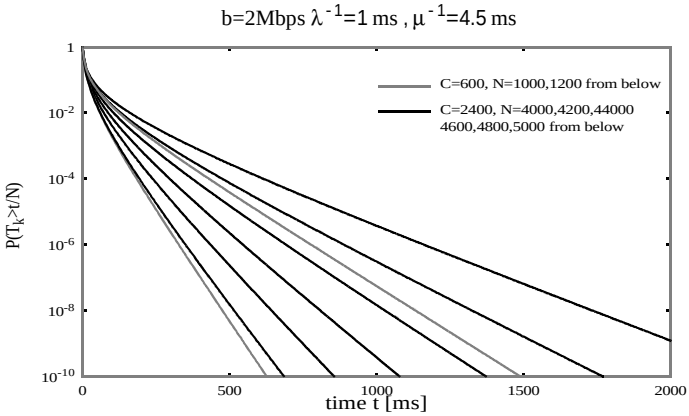

Figure 5.9: CDF of excess times for some different choice of the parameters as function of time.

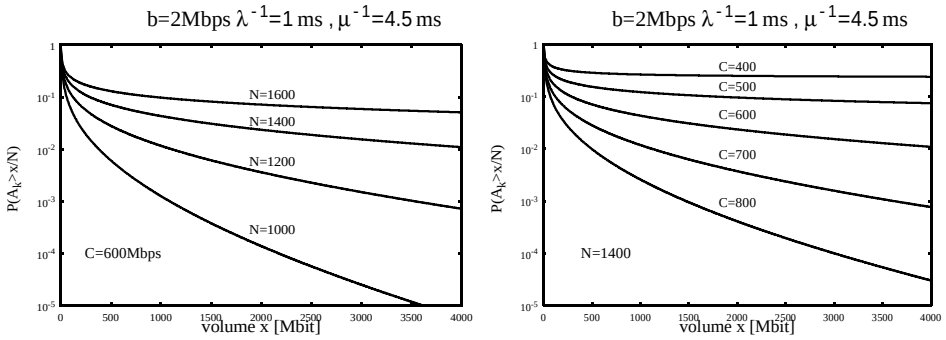

Figure 5.10: CDF of excess volumes for some different choice of the parameters as function of volume.

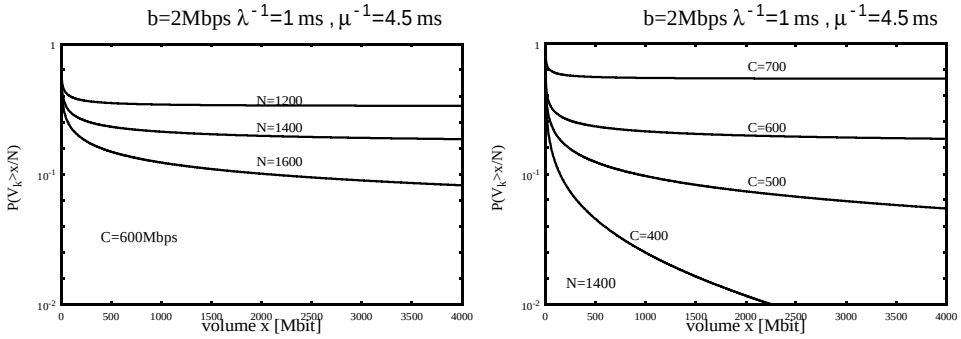

Figure 5.11: CDF of "normal loaded" volumes for some different choice of the parameters as function of volume.

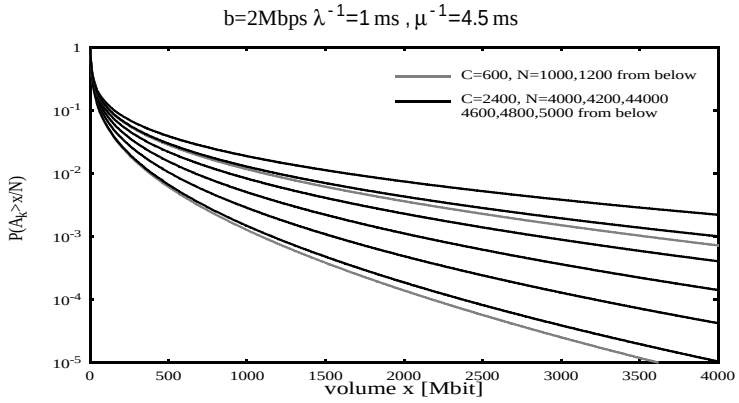

Figure 5.12: CDF of excess volumes for some different choice of the parameter as function of volume.

In figure 5.8 we have also given the CDFs for the “normal loaded” periods and we observe that these curves are extremely flat.

In figures 5.10-5.12 we have also depicted the CDF of the excess volume for this example. In the first two graphs we vary both the number of source and the capacity and we see that the curves for the volumes are curved and diverge more from a straight line than the corresponding curves for the time distributions. Thus the volume is more likely to have “long tails” than the corresponding excess time.

If we, as an example, look at the 10^{-4} quantile for the excess time, we find that with 1400 sources the corresponding time is approximately 0.75 ms, which is not a very long time. For the excess volume the picture is somewhat different. If we consider the 10^{-2} quantile, we find that with 1400 sources the excess volume is approximately 2.5 Mbit which is a quite large loss (even on a 600 Mbps link) and represents a typical loss of approximately 2 Kbit per source if the total loss is equally spread.

We shall close this chapter by consider the Ornstein-Uhlenbeck process as an approximation for the given example above. We have the following parameters (mean bit rate and standard deviation):

$$m = \frac{\lambda N}{\lambda + \mu} b \text{ and } \sigma = \frac{\sqrt{N\lambda\mu}}{\lambda + \mu} b \tag{5.107}$$

and the autocorrelation:

$$\rho(t) = e^{-(\lambda + \mu)t} \tag{5.108}$$

With the appropriate scalings (see [Knes91]) we have the following asymptotics based on the first passage times and corresponding volumes for the O-U process:

$$P(T_k > t) \approx F_{T^x}(x', t') \text{ and } P(A_k > z) \approx F_{A^x}(x', z') \tag{5.109}$$

where $F_{T^x}(x', t')$ and $F_{A^x}(x', z')$ are the CDFs of the first passage times and the corresponding volumes for the O-U process with (scaled) capacity (found in section 4.4):

$$C' = \frac{C - \frac{\lambda N}{\lambda + \mu} b}{\frac{\sqrt{N\lambda\mu}}{\lambda + \mu} b} = \frac{j_C - \frac{\lambda N}{\lambda + \mu}}{\frac{\sqrt{N\lambda\mu}}{\lambda + \mu}} \tag{5.110}$$

(where we for simplicity assumes that $j_C = \frac{C}{b}$ is an integer) and

$$x' = \frac{j_C + 1 - \frac{\lambda N}{\lambda + \mu}}{\frac{\sqrt{N\lambda\mu}}{\lambda + \mu}}, \; t' = (\lambda + \mu)t \text{ and } z' = \frac{(\lambda + \mu)^2}{\sqrt{N\lambda\mu}} \frac{z}{b} \tag{5.111}$$

is the corresponding dimensionless variable for the O-U process. As pointed out in [Knes91] the asymptotic formula yields when the scaled capacity C' remains constant as the number of sources N increases that is:

$$j_C - \frac{\lambda N}{\lambda + \mu} = O(\sqrt{N}) \quad \text{as } N \to \infty \tag{5.112}$$

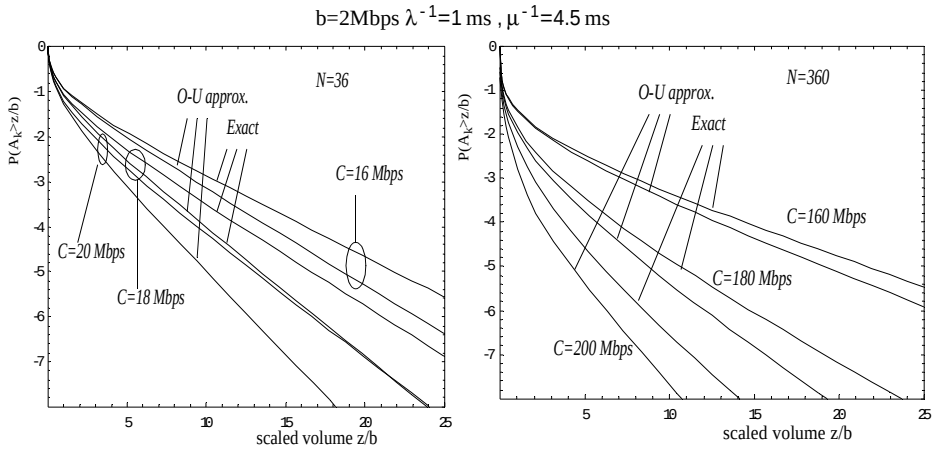

Figure 5.13: CDF of excess volumes based on the ON/OFF model and the corresponding O-U approximation for some different choice of the parameters as function of volume (scaled by peak bit rate).

In a couple of numerical examples (figure 5.13) we have tested the O-U approximation described above for the CDF of the excess volume based on the ON/OFF source model. In the right figure we have taken an example with a rather small number of sources ($N = 36$). For so few sources the O-U approximation underestimates the CDF of the excess volume. In the right figure we have depicted the corresponding cases by increasing both the number of sources and the capacity by a decade. In this case the O-U approximation improves especially for the high load case ($N = 360$ and $C = 160$ Mbps), but also for those cases the O-U model underestimates the CDF of the excess volume.

For birth-death models the method of bisection provides a very effective way of calculating the eigenvalues also for systems of large dimensions. The corresponding PDFs (and CDFs) of the excess times and volumes are calculated by (D.38) and (D.40) and it turns out to be far more effective in terms of computer time than the rather slowly converging series of the PDFs and (the CDFs) of the first passage times and the corresponding volumes for the O-U process (given by (4.81) and (4.82), and (4.120) and (4.121)).

PART-II

Models for calculating end-to-end delay and delay-jitter in packet networks

6

Introduction

It is reason to believe that real time services will be a significant part of the traffic offered in future multi service networks. Real time services require a quite regular bit stream delivered at the receiver's site to maintain the necessary quality for the various applications. It is well known that networks based on statistical multiplexing (like IP-networks) will introduce certain disturbance (jitter) in the bit stream mainly due to queuing in routers (or switches). These disturbances will add on along the path from the sender to the receiver, explaining the necessity of some kind of de-jitter buffer at the receiver site to compensate for these variations. The end-to-end delay is therefore an important parameter not only for dimensioning the de-jitter buffers, but also for providing some upper bounds of the total network delay for particular services and it is among the most important QoS (Quality of Service) parameter in networks deploying statistical multiplexing.

The Internet has traditionally offered a service that is commonly characterised as a best-effort service. The network tries to deliver the IP packets at their destination but no guarantees are given. Many applications (e.g. email) happily run using this traditional best-effort service model. Some areas of the Internet may be heavily congested and, consequently, a considerable fraction of packets is discarded by the network. Usually, additional higher layer protocols (e.g. TCP for error detection, retransmission and flow control) compensate for the lost packets. At the application level this is then noticed as a reduced throughput, which is unacceptable for many type of real time services without destroying the quality.

By introducing high capacity Internet, with differentiated QoS, it will be possible to offer services with highly variable characteristics in one common network and thereby reducing the cost compared with operating a number of more or less specialized networks for each type of services. The success of such a scenario will strongly depend on the ability to perform the necessary differentiation among services.

To provide IP transport with QoS guarantees for throughput critical and delay critical applications, the IP community has realised that in IP routers packets of delay critical flows need to be forwarded differently from other packets, e.g. by applying some kind of priority mechanisms. The specific details on how to realise the necessary difference in packet forwarding, for example on how to recognise flow classes, has led to different service models: the Integrated Services (IntServ) approach and the Differentiated Services (DiffServ) approach. The most relevant transport service descriptions for delay critical applications are the IntServ inspired Guaranteed Service [RFC1633], the DiffServ inspired Expedited For-

warding behaviour [RFC2475] and the more generic Dedicated Bandwidth IP transfer capability [ITU02b]. Though differing in approach and in detail, a common factor in all these descriptions is that some kind of differentiation (in the packet handling) is introduced.

6.1 Addressing the QoS

The end-to-end QoS is realised through the contributions from the different domains and the QoS guarantees end-to-end will be realised through different SLAs between the customer and the access network and/or between different core network domains. For each domain it will be important to estimate the contribution to the QoS parameters since each administrative domain will be responsible for their own contribution through the SLA. The most important parameters will typically be delay, jitter and information losses due to buffer overflow. It will be important for a network operator to be able to estimate the QoS parameters in his own domain to set the appropriate parameters in the SLAs. In addition it will be of vital importance to implement the necessary control structures that make it possible to maintain the guarantees especially in his own domain and thereby preventing degradation in QoS which is often seen in best effort networks of today.

To realise many type of services the traffic flow will have to cross one or more administrative domains with their own SLAs. Such domains could for instance typically be

- access networks with rather low capacity based on different wireless or DSL technologies
- core networks with high capacity links but with large numbers of routers

The access network will encompass a variety of different access technologies that are currently available. These can be divided according to

- fixed access, or
- mobile access.

With the recent advances in access technology the fixed access may be a mixture of one or more different types as Asymmetric Digital Subscriber Line (ADSL), Very high speed Digital Subscriber Line (VDSL), Coax and optical fibre, all having very different physical characteristics. The logical structure of the access network may therefore be very different. Traditionally there has been quite a strict distinction between the access network and the core (transit) network, where the access is defined to be the part of the network from the subscribers to the local exchange. By increasing the line speed by introducing different active components these definitions of where the access network ends and where the core (transit) network starts are not direct valuable any more. In IP networks the definition seems to be more flexible on the basis of more functional distinctions. Usually one will define the core network as the part of the network where DiffServ and/or MPLS are deployed. By the increase of the line speeds it is however an interesting question to find out how 'far ' out in the 'old access network' the DiffServ model (and possible MPLS) is effective.

6.2 Performance issues

6.2.1 IP-multiplexing for low capacity links

When traffic of different types is multiplexed on the IP level this may cause delay and jitter problems if these traffic types share a link with rather low capacity. The main cause for this delay and jitter is the variation in the packet lengths for the different traffic types. While typical real time traffic like voice will emit packets of a small fixed size, the typical data application may generate packets that are quite long. Due to this mismatch in packet size between different applications the queueing delay for typical real time traffic may increase over the limit resulting in a degradation of the quality. This negative multiplexing effect will add for each router along the path from the sender to the receiver. However, for high capacity links this queueing delay will be more or less negligible, leaving the main delay contribution to low capacity links in the access network.

One could hope that deploying the DiffServ model with traffic classification and PHB priority scheduling would overcome this problem. This is however not the case unless there is some kind of fragmentation of the long IP packets on lower layers. This means that although most of the DiffServ implementations (in routers) have implemented priority among different traffic classes these priority mechanisms are all non-preemptive. With this type of priority mechanism a high priority packet cannot interrupt an ongoing transmission of a packet of lower priority. This means that the packet length distribution of the lower priority traffic classes will have an impact on the delay for the high priority traffic.

The only way to get around the multiplexing problem for low capacity links is to have some kind of fragmentation of the long IP packet, making it possible to interleave small real time IP packets. By this option the maximum waiting time due to lower priority traffic classes will just be the transmission time for a single fragment. This fragmentation will be possible if IP is transported over ATM, and in this case the maximum disturbance of the high priority traffic due to lower priority is limited to one ATM cell.

6.2.2 Addressing the end-to-end queueing delay

In PART-II of the thesis we focus on the delay and delay variation experienced in an IP domain for which some form of delay (variation) commitment is intended. The commitments may vary in strictness, ranging from a strict guarantee to a more loosely defined objective. In any case, it is relevant to have some kind of estimates, in particular before the design and rollout of new services, of the expected delay and delay variation.

The least known factor in the end-to-end delay of an IP packet is the delay contribution due to queuing in the network elements. This contribution is also important because, in most cases, the delay variation introduced by the network is to be removed by the receiving application (de-jittering), thus introducing additional de-jitter delay which is necessarily at least as large as the maximum (or suitable quantile) of the delay variation. Other factors, such as the contribution of the propagation delay and the variation in router's routing lookup latency are expected to either be much easier to assess or to be negligible in comparison to the queueing delay due to statistical multiplexing in the routers.

We therefore address some methods to calculate, the queuing delay in a network where several interactive IP flows are mixed with best-effort IP flows and where the interactive packets may have strict priority over best effort traffic. The objective is to provide both a suitable model and suitable assessment techniques (e.g. calculation methods) to arrive at a suitable estimate of the queuing delay in a given situation. With 'suitable' we mean a model which is sufficiently close to the real world to have a practical meaning and, at the same time, is sufficiently easy to allow the calculations to be carried out. The results are targeted on network providers to allow them to assess the expected behaviour and the delay commitments in their own network as well as on standardisation bodies to assess the expected behaviour over various different networks of loosely co-operating providers (e.g. network performance objectives).

6.3 Reference configuration

The reference configuration for the modelling of the queuing behaviour consists in an upstream access network part, a multi-hop core network and a downstream access network part as illustrated in figure 6.1

.

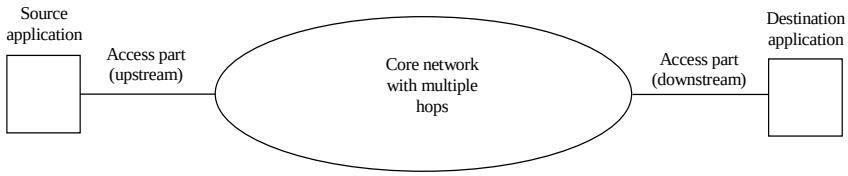

Figure 6.1: Overview of the reference configuration: upstream access network part, core network part (multi-hop) and downstream access part between the source and destination application.

The access network part is designated separately from the core network part because the parameters in these network parts (e.g. link rate, number of flows) may differ considerably. In addition, it is expected that the users at the source and destination have much more control over the traffic aggregate on 'their' access part (e.g. in case of a dedicated ADSL link) than over the traffic aggregate in the core network.

The reference configuration for the router in each network hop consists in an high priority queue for the interactive flows and a low priority queue for the best-effort flows as shown in figure 6.2. Each of these queues is served as FCFS (First Come First Serve) and the high priority queue is served with non-preemptive priority over the best-effort queue. In this part of the thesis we focus on the high priority traffic and neglecting the influence from the low priority (best-effort) queue.

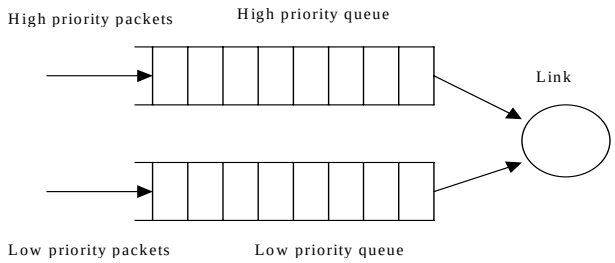

Figure 6.2: The router output buffer for the reference configuration.

The relevant modelling parameters for each router hop are the following.

- The capacity of the outgoing links available for the transport of IP packets (e.g. 149.76 Mbit/s on an 155 Mbit/s STM-1 link but where we may also include one or more relative low capacity access link with capacity typical that of an ADSL modem.
- The size (in byte) of the high priority best-effort IP packets. To keep the number of parameters at a manageable level, all best-effort IP packets are assumed to have the same size; for practical results a value of 1500 byte is used.
- The size of the low priority interactive IP packets (in byte or as a fraction of the best-effort size) and the load (as a percentage of the outgoing link rate) of the interactive flows. To keep the number of parameters at a manageable level, all interactive IP packets are assumed to have the same size. For voice applications a size between 150 byte and 300 byte (i.e. 1/10th and 1/5th a 1500 byte best-effort packet) is expected to be common. The arrival process of the interactive IP packets is assumed to be Poisson.

6.4 The organisation of PART-II of the thesis

PART-II of the thesis consists of three more or less self-contained chapters. The “red thread” through these chapters is the modelling of end-to-end queueing delay for networks deploying statistical multiplexing.

In chapter 7 we have consider some models to calculate the end-to-end delay distributions for packet networks based on the assumption that the end-to-end delay may be found by convolutions, where the key assumption is that the parts of the end-to-end delay stemming from the different nodes are independent stochastic variables. As the model for each node we take the ordinary M/G/1 queue. If in addition the nodes are identical, i.e. the convolution consists of the waiting times of a fixed numbers of identically M/G/1 queues, the evaluation may be substantially simplified. In this case we show that the convolutions may be found by taking some partial derivatives with respect to the load parameter. This result is shown to yield for the Laplace-transforms and will therefore also yields for the correspond-

ing distribution functions and the corresponding densities. The applied technique may be generalised in various directions for instance it is possible to extend the result to the case with two groups of queues where the queues in each group are identical.

For the M/D/1 model we give explicit closed form results for the convolutions of a given number queues having identical waiting times distributions. We also generalize this result to consider two groups of M/D/1 queues having different (but constant) service times, and this is a particular interesting case since it may be used as model for end-to-end delay also including access links with low capacity. Some approximations are also given based on large deviation techniques. Part of the results will be presented at ITC-18 [Øste03c] and an extended version of the paper is found in [Øste03b].

In chapter 8 we have extended the results found in chapter 7 to also include queueing models with priority. This is an important extension since new service models of IP flows have been introduced (DiffServ) where different treatment of flows in routers is assumed to provide QoS guarantees for the different QoS classes. As a model for the delay in such nodes we take the M/G/1 non-preemptive priority model as the basis, where we primarily are interested in the delay for the high priority class. By using the method described in chapter 7 we find a method to express the convolution in terms of the convolutions of waiting times for the M/G/1 queue (without any priority). For the case with deterministic service times, (M/D/1 model), we find explicit formula for the desired convolution. Numerical examples show that end-to-end delays for rather large chains of nodes may be analysed without numerical difficulties. Most of the material in this section is found in [Øste03a].

In chapter 9 we describe a different approach to obtain the end-to-end queueing delay in packet networks. The main idea is to try to capture the disturbance of a packet stream as it passes through a series of multiplexers. Even though the output process from a multiplexer surely is non-renewal, we get the distribution between two consecutive departures, and approximate the process with a renewal stream. This stream is then fed into the next multiplexer (together with other crossing traffic). In this way we obtain recursive relations for the jitter and the end-to-end delay. In the analyse we use a slotted model rather than a time continuous one. The reason is that the slotted model is easier to analyse and we use generating function techniques and apply the theory of complex analysis rather than Laplace transforms for the continuous time counterpart. Numerical examples show good accordance with the convolution approach of chapter 7. The material in this chapter is yet unpublished.

7

Convolution of a given number of waiting times of M/G/1queues having identical service time distributions

7.1 Some preliminary considerations

In the following we consider a model to calculate the end-to-end delays in a large scale IP-network. The aim is to calculate the distribution of the end-to-end delay for a particular path consisting of a series of routers. We assume that all the nodes in the end-to-end path are *statistically independent;* this is a key assumption to obtain the end-to-end delay by convolution. The condition under which this independence assumption applies is not considered in this chapter. We shall therefore take the M/G/1 queue as the model to obtain the waiting time distribution in each node and then apply the convolution to obtain the end-to-end waiting time distribution.

We consider a path consisting of a given number of nodes (say K). Each of the nodes is taken to be a M/G/1 queue with load ρ_k, and where we assume that each queue has service times that are identically distributed given by a PDF (Probability Density Function) $b_k(t)$, and LST (Laplace-Stieltjes Transform) denoted by $B_k(s)$; $k = 1,..., K$.

Further we let W_k denote the waiting time in queue k, and we denote the corresponding PDF (Probability Density Function) $w(x, \rho_k)$ and the corresponding DF (Distribution Function) $W(x, \rho_k)$. (Where we have indicated that the waiting time will also depend upon the parameter ρ_k and of course on the service time distribution.)

The LST of the waiting time for one particular queue is given by the well-known *Pollaczek-Khinchin* (*P-K*) formula (see for instance Kleinrock [Klei76a]):

$$\tilde{W}(s, \rho_k) = \frac{1 - \rho_k}{1 - \rho_k \hat{B}_k(s)} \tag{7.1}$$

where $\hat{B}_k(s)$ is the LST of the remaining service time and is given by

$$\hat{B}_k(s) = \frac{1 - B_k(s)}{s b_k} \tag{7.2}$$

where $b_k = E[B_k]$ is the mean service time; $k = 1,...,K$.

We are interested in the sum of waiting times in a series of K queues and we denote the sum $W = W_1 + \ldots + W_K$. If all the waiting times may be taken to be independent the PDF of the sum yields the convolution of the waiting times in each queue. The LST of the convolution (of waiting times for all the K queues) yields the product:

$$\tilde{W}(s, \rho_1,..., \rho_K) = \prod_{k=1}^{K} W(s, \rho_k) = \prod_{k=1}^{K} \frac{1 - \rho_k}{1 - \rho_k \hat{B}_k(s)} \tag{7.3}$$

Generally it is possible to obtain the end-to-end distributions above by inverting the transform (7.3) numerically. Such methods are described in the literature. For instance the DF of the end-to-end queueing times may be written by the inversion integral as:

$$W(t, \rho_1,..., \rho_K) = P\{W \le t\} = \frac{1}{2\pi i} \int_{\gamma} \frac{e^{st}}{s} W(s, \rho_1,..., \rho_K) ds \tag{7.4}$$

where the integration line parallel with the imaginary axis $\gamma = \{s | s = a + iy\}$ and where $a > 0$ is a constant and $y \in (-\infty, \infty)$. Abate & Whitt have given a method to calculate the inversion integral based on Poisson's summation formula. The result yields an alternating series that may be difficult to use to determine the tail of the distribution and thereby obtain the desired quantiles (see [Abat92]).

7.2 Convolution of waiting times in M/G/1 queues all having identically distributed service times

In the following we assume that all the nodes have identically distributed service times, that is, we assume $b_k(t) = b(t)$ also implying $\hat{B}_k(s) = \hat{B}(s)$; $k = 1,..., K$. In this case it is possible to obtain substantial simplification of the convolution (7.3) (and also on the inversion integral (7.4)). If the loads of the different M/G/1 queues all are distinct, that is $\rho_i \neq \rho_j$ for all $i, j = 1,..., K$, then the LST of the convolution can be written as a weighted sum of the individual LST for each queue as follows:

$$\tilde{W}(s, \rho_1,..., \rho_K) = \sum_{k=1}^{K} c_k \tilde{W}(s, \rho_k) \text{ where} \tag{7.5}$$

the coefficients c_k only depend on the loads in the different queues and are given by:

$$c_k = \prod_{l=1, l \neq k}^{K} \frac{1 - \rho_l}{1 - \rho_l / \rho_k} \tag{7.6}$$

The result (7.5) and (7.6) enables obtaining corresponding convolution of the waiting times by inverting the LST above. We may express this result as follows: Let

$$w(x,\rho_1,...,\rho_K) = w(x,\rho_1) * .... * w(x,\rho_K) \tag{7.7}$$

be the convolution of waiting times, then we have:

$$w(x,\rho_1,...,\rho_K) = \sum_{k=1}^{K} c_k w(x,\rho_k) \tag{7.8}$$

where $w(x,\rho_k)$ is the PDF for the waiting time for the k'th queue. Similar for the DF of the convolution we have the corresponding result

$$W(x,\rho_1,...,\rho_K) = \sum_{k=1}^{K} c_k W(x,\rho_k) \tag{7.9}$$

where $W(x,\rho_k)$ is the DF for the k'th queue. All the results above follow from the identity

$$\frac{(1-\rho_1)....(1-\rho_K)}{(1-\rho_1\hat{B})....(1-\rho_K\hat{B})} = \sum_{k=1}^{K} c_k \frac{1-\rho_k}{1-\rho_k\hat{B}} \tag{7.10}$$

and we can find c_k by multiplying the identity by $\frac{1-\rho_j\hat{B}}{1-\rho_j}$ and we find the following expression for c_j:

$c_j = \prod_{l=1,l\neq j}^{K} \frac{1-\rho_l}{1-\rho_l\hat{B}} - \left(\sum_{k=1,k\neq j}^{K} c_k \frac{1-\rho_k}{1-\rho_k\hat{B}}\right)\frac{1-\rho_j\hat{B}}{1-\rho_j}$. By taking the limit $\hat{B} \to \frac{1}{\rho_j}$ we obtain (7.6).

Often we are interested in the case where the loads on the different queues are equal. This result is possible to obtain relatively easily from (7.5) and (7.6) by letting $\rho_k \to \rho$ for all $k = 1,...,K$. It is possible to rewrite $\tilde{W}(s,\rho_1,...,\rho_K)$ as follows:

$$\tilde{W}(s,\rho_1,...,\rho_K) = \left(\prod_{i=1}^{K}(1-\rho_i)\right)\sum_{k=1}^{K}\left(\prod_{l=1,l\neq k}^{K}\frac{1}{\rho_k-\rho_l}\right)\left\{\frac{\rho_k^{K-1}}{1-\rho_k}\tilde{W}(s,\rho_k)\right\} \tag{7.11}$$

By taking the limits $\rho_k \to \rho$ for all $k = 1,...,K$ we obtain

$$\tilde{W}^K(s,\rho) = \left(\tilde{W}(s,\rho)\right)^K = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho}\tilde{W}(s,\rho)\right\} \tag{7.12}$$

and further in the time domain the corresponding results also yield:

$$w^K(x,\rho) = \left(w(x,\rho)\right)^{*(K)} = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho}w(x,\rho)\right\} \text{ and} \tag{7.13}$$

$$W^K(x,\rho) = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho}W(x,\rho)\right\} \tag{7.14}$$

In fact it is also easy to show (7.11) directly. Inserting for $\tilde{W}(s,\rho)$ we have

$$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho}\tilde{W}(s,\rho)\right\} = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho\hat{B}(s)}\right\} =$$

$$\frac{(1-\rho)^K}{(K-1)!}\sum_{l=0}^{K-1}\binom{K-1}{l}\frac{\partial^l}{\partial\rho^l}\left(\frac{1}{1-\rho\hat{B}(s)}\right)\frac{\partial^{K-l-1}}{\partial\rho^{K-l-1}}\{\rho^{K-1}\} = \left(\frac{1}{1-\rho\hat{B}(s)}\right)^K\sum_{l=0}^{K-1}\frac{(K-1)!}{l!(K-l-1)!}(\rho\hat{B}(s))^l(1-\rho\hat{B}(s))^{K-l-1} =$$

$$\left(\frac{1-\rho}{1-\rho\hat{B}(s)}\right)^K\left(\rho\hat{B}(s)+1-\rho\hat{B}(s)\right)^{K-1} = \left(\tilde{W}(s,\rho)\right)^K$$

We can now state the more general result where we consider the case where only some of the queues may be equally loaded. In this case we have LST of the convolution on the following (general) form:

$$\tilde{W}(s,\rho_1,...,\rho_N,n_1,...,n_N) = \prod_{j=1}^{N}\left(\frac{1-\rho_j}{1-\rho_j\hat{B}(s)}\right)^{n_j} \tag{7.15}$$

where we have N groups of queues of size n_j equally loaded and with distinct loads between the groups, that is $\rho_i \neq \rho_j$ for all $i,j = 1,...,N$ and $K = n_1 + ... + n_N$.

Then the LST given by (7.15) may be written:

$$\tilde{W}(s,\rho_1,...,\rho_N,n_1,...,n_N) = \sum_{j=1}^{N}\sum_{i=1}^{n_j}c_{ij}\left(\frac{1-\rho_j}{1-\rho_j\hat{B}(s)}\right)^i \tag{7.16}$$

where the coefficients c_{ij} only depend on the loads in the different queues and are given by:

$$c_{ij} = \frac{(-1)^{n_j-i}}{(n_j-i)!}\left(\frac{1-\rho_j}{\rho_j}\right)^{n_j-i}\frac{d^{n_j-i}}{dx^{n_j-i}}\left\{\prod_{l=1,l\neq j}^{N}\left(\frac{1-\rho_l}{1-\rho_l x}\right)^{n_l}\right\}_{x=\rho_j^{-1}} \tag{7.17}$$

for $i = 1,...,n_j$ and $j = 1,..,N$. The PDF and DF of the convolution may be obtained by inverting (7.16):

$$w(x,\rho_1,...,\rho_N,n_1,...,n_N) = \sum_{j=1}^{N}\sum_{i=1}^{n_j} c_{ij} w^i(x,\rho_j) \tag{7.18}$$

$$W(x,\rho_1,...,\rho_N,n_1,...,n_N) = \sum_{j=1}^{N}\sum_{i=1}^{n_j} c_{ij} W^i(x,\rho_j) \tag{7.19}$$

where $w^i(x,\rho)$ and $W^i(x,\rho)$ are given by (7.12) and (7.13) respectively. It is sufficient to prove equation (7.16) since (7.18) and (7.19) follow directly by applying (7.12) and (7.13). By partial expansion of the fraction (7.15) (by taking $\hat{B}(s)$ as the free variable) it is possible to obtain the expansion (7.16). One way to obtain the coefficients is given as follows. We pick a particular group, say the j'th one, and we get from (7.15) and (7.16):

$$c_{n_j j} + c_{n_j-1 j} y + ... + c_{2j} y^{n_j-2} + c_{1j} y^{n_j-1} = \prod_{l=1,l\neq j}^{N} \left(\frac{1-\rho_l}{1-\frac{\rho_l}{\rho_j} + \frac{\rho_l(1-\rho_j)}{\rho_j} y} \right)^{n_l} + y^{n_j} F(y)$$

where we have set $y = \frac{1-\rho_j \hat{B}(s)}{1-\rho_j}$ and $F(y) = \sum_{l=1,l\neq j}^{N} \sum_{i=1}^{n_j} c_{il} \left(\frac{1-\rho_l}{1-\frac{\rho_l}{\rho_j} + \frac{\rho_l(1-\rho_j)}{\rho_j} y} \right)^{i}$. Considered as a function of y we have that $F(y)$ is analytical at $y = 0$. By differentiating the relation above $n_j - i$ times and setting $y = 0$ we get: $c_{ij} = \frac{1}{(n_j - i)!} \frac{d^{n_j-i}}{dy^{n_j-i}} \left\{ \prod_{l=1,l\neq j}^{N} \left(\frac{1-\rho_l}{1-\frac{\rho_l}{\rho_j} + \frac{\rho_l(1-\rho_j)}{\rho_j} y} \right)^{n_l} \right\}_{y=0}$ which can be written as (7.17) by the translation $x = \frac{1}{\rho_j} - \frac{(1-\rho_j)}{\rho_j} y$.

As a side result we obtain for the interesting case with only two groups of queues, $N = 2$ (with equal load in each group), that the coefficients in equation (7.17) may be found explicitly since the differentiation may be carried out. We get the following expansions:

$$w(x,\rho_1,\rho_2,n_1,n_2) = \sum_{i=1}^{n_1} c_{i1} w^i(x,\rho_1) + \sum_{i=1}^{n_2} c_{i2} w^i(x,\rho_2) \text{ and} \tag{7.20}$$

$$W(x,\rho_1,\rho_2,n_1,n_2) = \sum_{i=1}^{n_1} c_{i1} W^i(x,\rho_1) + \sum_{i=1}^{n_2} c_{i2} W^i(x,\rho_2) \tag{7.21}$$

where $w^i(x,\rho)$ and $W^i(x,\rho)$ are given by (7.13) and (7.14) and

$$c_{i1} = (-1)^{n_1-i}\binom{n_1+n_2-i-1}{n_2-1}\left(\frac{(1-\rho_2)\rho_1}{\rho_1-\rho_2}\right)^{n_2}\left(\frac{(1-\rho_1)\rho_2}{\rho_1-\rho_2}\right)^{n_1-i} \quad \text{for } i=1,\ldots,n_1 \tag{7.22}$$

and

$$c_{i2} = (-1)^{n_2-i}\binom{n_1+n_2-i-1}{n_1-1}\left(\frac{(1-\rho_1)\rho_2}{\rho_2-\rho_1}\right)^{n_1}\left(\frac{(1-\rho_2)\rho_1}{\rho_2-\rho_1}\right)^{n_2-i} \quad \text{for } i=1,\ldots,n_2 \tag{7.23}$$

As a comment to the results derived above we have shown that it is quite easy to obtain a convolution of waiting times in a series of M/G/1 queues if the service times are identically distributed in all the queues. The main result follows by taking partial fractions expansion of the LST of the convolution given by the product of the LSTs of the waiting time for each queue, and thereby making it possible to write the LST of the convolution as a weighted sum of the LST of the individual queues. Since the result is obtained for the LST of the convolution, the same result will also apply for the DF and PDF. If all the queues in addition have equal load the result is obtained simply by taking partial derivatives with respect to the load as given by (7.11) (and (7.12) and (7.13)).

7.3 Convolution of waiting times in M/G/1 queues having different service times

In the general case the results derived so far require that all the service times are identically distributed. This will impose a rather strong restriction on the scenarios where the previous results are applicative. For instance, a path that also includes rather slow access links could not be modelled well with these results. An alternative could be to find the delay distributions for each part; the access and core, and then perform numerical convolution to find the total end-to-end delay. It is however possible to extend some of the results to cover convolutions between equally loaded groups (with different service time distributions in each group) if it is possible to find the convolution obtained by taking one single waiting time in each group. For simplicity we consider the case with two groups (and use the same notation as above) and we let

$$w(x,\rho_1,\rho_2) = w(x,\rho_1) * w(x,\rho_2) \tag{7.24}$$

be the convolution of PDFs of two waiting times distributions for M/G/1 queues, one from each group, and let $W(x,\rho_1,\rho_2)$ denote the corresponding DF. Then we may get the PDF and DF of the convolution of n_1 waiting times from group 1 and n_2 waiting times from group 2 as:

$$w(x,\rho_1,\rho_2,n_1,n_2) = \frac{(1-\rho_1)^{n_1}}{(n_1-1)!}\frac{(1-\rho_2)^{n_2}}{(n_2-1)!}\frac{\partial^{n_1+n_2-2}}{\partial\rho_1^{n_1-1}\partial\rho_2^{n_2-1}}\left\{\frac{\rho_1^{n_1-1}}{1-\rho_1}\frac{\rho_2^{n_2-1}}{1-\rho_2}w(x,\rho_1,\rho_2)\right\} \text{ and} \tag{7.25}$$

$$W(x,\rho_1,\rho_2,n_1,n_2) = \frac{(1-\rho_1)^{n_1}}{(n_1-1)!}\frac{(1-\rho_2)^{n_2}}{(n_2-1)!}\frac{\partial^{n_1+n_2-2}}{\partial\rho_1^{n_1-1}\partial\rho_2^{n_2-1}}\left\{\frac{\rho_1^{n_1-1}}{1-\rho_1}\frac{\rho_2^{n_2-1}}{1-\rho_2}W(x,\rho_1,\rho_2)\right\} \tag{7.26}$$

(7.25) and (7.26) follow directly from (7.12) and (7.13) and the fact that $w(x,\rho_1,\rho_2)$ is the convolution (7.24). To apply the last result one needs first to find the convolution $w(x,\rho_1,\rho_2)$ which may be difficult to find unless for specific models. Below we shall show that for M/D/1 queues this convolution is possible to obtain in closed forms, and whence it is possible to apply the result above to find the convolution where the service times are different.

7.3.1 Convolution of the waiting time distribution for a given number of M/D/1 queues all with equal service times

In the following we shall apply the results on specific models. Of main interested is the case with constant service times since this often will be the case for many applications.

Without losing generality we scale the service time to unity. Then it is well known that the DF for the waiting time of the M/D/1 queue is given by [Robe92](page 391):

$$q(x,\rho) = (1-\rho)\sum_{k=0}^{\lfloor x \rfloor}\frac{[\rho(k-x)]^k}{k!}e^{-\rho(k-x)} \tag{7.27}$$

Below we shall apply (7.13) to find explicit expression for the convolution of the waiting time distributions for given numbers of equally loaded M/D/1 queues.

If all of the queues are equally loaded with equal service times we get the following expression for the convolution the DF of a series of K waiting times of identical M/D/1 queues:

$$q^K(x,\rho) = (1-\rho)^K\sum_{k=0}^{\lfloor x \rfloor}\sum_{l=0}^{K-1}\frac{(-1)^l}{l!k!}\binom{K+k-1}{k+l}\big(\rho(k-x)\big)^{k+l}e^{-\rho(k-x)} \tag{7.28}$$

It is quite easy to show (7.28) by applying (7.13). We have:

$$q^K(x,\rho) = W^K(x,\rho) = \frac{(1-\rho)^K}{(K-1)!}\sum_{k=0}^{\lfloor x \rfloor}\frac{(k-x)^k}{k!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K+k-1}e^{-\rho(k-x)}\right\} \tag{7.29}$$

Differentiation gives

$$\frac{1}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K+k-1}e^{-\rho(k-x)}\right\} = \sum_{l=0}^{K-1}\frac{1}{(K-1)!}\binom{K-1}{l}\frac{\partial^l}{\partial\rho^l}\left\{e^{-\rho(k-x)}\right\}\frac{\partial^{K-l-1}}{\partial\rho^{K-l-1}}\left\{\rho^{K+k-1}\right\} =$$

$\sum_{l=0}^{K-1}(-1)^l\frac{(K+k-1)!}{(k+l)!l!(K-l-1)!}\rho^{k+l}(k-x)^l e^{-\rho(k-x)}$. Inserting this result in (7.29) we get the expression (7.28).

It is also possible to find corresponding formula for the PDF of the convolution. We find for one single M/D/1 queue

$$w(x,\rho)=\frac{d}{dx}W(x,\rho)=\rho\big(q(x,\rho)-H(x-1)q(x-1,\rho)\big) \tag{7.30}$$

where $q(x,\rho)$ is given by (7.27) and $H(x)$ is the unit step function. We find that the PDF of the convolution of K waiting times from identical M/D/1 queues may be written as the difference:

$$w^K(x,\rho)=\rho\big(q_1^K(x,\rho)-H(x-1)q_1^K(x-1,\rho)\big) \text{ where} \tag{7.31}$$

$$q_1^K(x,\rho)=(1-\rho)^K\sum_{k=0}^{\lfloor x\rfloor}\sum_{l=0}^{K-1}\frac{(-1)^l}{l!k!}\binom{K+k}{k+l+1}\big(\rho(k-x)\big)^{k+l}e^{-\rho(k-x)} \tag{7.32}$$

We obtain (7.31) and (7.32) by applying (7.13) on (7.30) and we find $w^K(x,\rho)=\rho\big(q_1^K(x,\rho)-H(x-1)q_1^K(x-1,\rho)\big)$ where

$$\rho q_1^K(x,\rho)=\frac{(1-\rho)^K}{(K-1)!}\sum_{k=0}^{\lfloor x\rfloor}\frac{(k-x)^k}{k!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K+k}e^{-\rho(k-x)}\right\} \tag{7.33}$$

Differentiation gives

$$\frac{1}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K+k}e^{-\rho(k-x)}\right\}=\sum_{l=0}^{K-1}\frac{1}{(K-1)!}\binom{K-1}{l}\frac{\partial^l}{\partial\rho^l}\left\{e^{-\rho(k-x)}\right\}\frac{\partial^{K-l-1}}{\partial\rho^{K-l-1}}\left\{\rho^{K+k}\right\}=$$

$\sum_{l=0}^{K-1}(-1)^l\frac{(K+k)!}{(k+l+1)!l!(K-l-1)!}\rho^{k+l+1}(k-x)^l e^{-\rho(k-x)}$. Inserting this result in (7.33) we obtain expression (7.32).

As for $q(x,\rho)$ given by formula (7.27) (for the single M/D/1 queue) the formulae for $q^K(x,\rho)$ and $q_1^K(x,\rho)$ given by (7.28) and (7.32) are not effective to calculate the DF and the PDF of the convolution for large values of x. In Appendix F we give an alternative way of writing these formulae that provides stable numerical calculations of the convolutions for quite large values of x, and up to 20 queues. It is therefore possible to study end-to-end delay in rather large networks by the convolution approach.

7.3.2 Convolution of waiting times in M/D/1 queues having different service times

For the M/D/1 model it is possible to find the convolution of two DF of waiting times with different service times. Then by (7.25) and (7.26) more general convolutions may be obtained by plain differentiations with respect to the different loads in the different groups. In the following we consider two M/D/1 queues with load ρ_i and service time $b_i, i=1,2$, and we denote $W(t,\rho_i,b_i)=q(t/b_i,\rho_i)$ the DF of the waiting time, $i=1,2$. Then the DF of the convolution is given by the following sums:

$$W(t,\rho_1,\rho_2,b_1,b_2)=(1-\rho_1)\sum_{k=0}^{\lfloor \frac{t}{b_1}\rfloor}\sum_{j=0}^{\lfloor \frac{t-kb_1}{b_2}\rfloor}\zeta p_{k,j}(\zeta,\eta)\left[q(\frac{t-kb_1-jb_2}{b_2},\rho_2)-q(\frac{t-kb_1-jb_2}{b_2}-1,\rho_2)\right]$$

$$+(1-\rho_2)\sum_{k=0}^{\lfloor \frac{t}{b_2}\rfloor}\sum_{j=0}^{\lfloor \frac{t-kb_2}{b_1}\rfloor}\eta p_{k,j}(\zeta,\eta)\left[q(\frac{t-kb_1-jb_2}{b_1},\rho_1)-q(\frac{t-kb_1-jb_2}{b_1}-1,\rho_1)\right] \tag{7.34}$$

where $p_{k,j}(\zeta,\eta)=\binom{k+j}{j}\eta^k\zeta^j$ and $\zeta=\frac{y}{y-x}$ and $\eta=\frac{x}{x-y}$; where we have defined $x=\frac{\rho_1}{b_1}$ and $y=\frac{\rho_2}{b_2}$; and where we have $\zeta+\eta=1$. (By writing (7.34)) it is understood that $q(x,\rho)=0$ for $x<0$.) To show (7.34) we let

$W_2(t,\rho_1,\rho_2,b_1,b_{2)})=\int_{x=0}^{t}W(x,\rho_1,b_1)W(t-x,\rho_2,b_2)dx=\int_{x=0}^{t}q(\frac{x}{b_1},\rho_1)q(\frac{t-x}{b_2},\rho_2)dx$, then the DF of the desired convolution is the time derivative of W_2:

$$W(t,\rho_1,\rho_2,b_1,b_{2)})=\frac{d}{dt}W_2(t,\rho_1,\rho_2,b_1,b_{2)}). \tag{7.35}$$

By introducing some different scaling we may write W_2 as:

$$W_2(t,\rho_1,\rho_2,b_1,b_{2)})=b_2 I(\frac{t}{b_2},\frac{b_2}{b_1},\rho_1,\rho_2) \text{ where} \tag{7.36}$$

$$I(t,\beta,\rho_1,\rho_2)=\int_{x=0}^{t}q(x\beta,\rho_1)q(t-x,\rho_2)dx \tag{7.37}$$

We have $q(x\beta,\rho_1)=(1-\rho_1)\sum_{k=0}^{\infty}H(x\beta-k)\frac{(-1)^k\rho_1^k}{k!}(x\beta-k)^k e^{-\rho_1(k-x\beta)}$. Inserting in (7.37) and interchange summation and integration give:

$$I(t,\beta,\rho_1,\rho_2)=(1-\rho_1)\sum_{k=0}^{\infty}H(t\beta-k)\frac{(-1)^k\rho_1^k}{k!}\int_{x=0}^{t-k\beta^{-1}}(x\beta)^k e^{\rho_1\beta x}q(t-k\beta^{-1}-x,\rho_2)dx$$

$=(1-\rho_1)\sum_{k=0}^{\lfloor t\beta\rfloor}\frac{(-1)^k(\rho_1\beta)^k}{k!}I_k(t-k\beta^{-1},-\rho_1\beta,\rho_2)$ where $I_k(t,\mu,\rho)$ is given in the Appendix F by equation (F.18). Inserting in the expression above we find:

$$I(t,\beta,\rho_1,\rho_2)=\frac{(1-\rho_1)}{\rho_2-\rho_1\beta}\sum_{k=0}^{\lfloor t\beta\rfloor}\sum_{l=0}^{\lfloor t-k\beta^{-1}\rfloor}(-1)^k\frac{(k+l)!}{k!l!}\frac{(\rho_1\beta)^k\rho_2^l}{(\rho_2-\rho_1\beta)^{k+l}}q(t-l-k\beta^{-1},\rho_2)$$

$$-\frac{(1-\rho_1)(1-\rho_2)}{\rho_2-\rho_1\beta}\sum_{l=0}^{\lfloor t\rfloor}\sum_{k=0}^{\lfloor (t-l)\beta\rfloor}\sum_{i=0}^{k}(-1)^{k-i}\frac{(k+l-i)!}{i!l!(k-i)!}\frac{(\rho_1\beta)^{k-i}\rho_2^l}{(\rho_2-\rho_1\beta)^{k+l-i}}\left(\rho_1(k-\beta(t-l))\right)^i e^{-\rho_1(k-\beta(t-l))}$$

In the last sum we change the summation by letting $j = k - i$ so that

$\sum_{k=0}^{\lfloor (t-l)\beta \rfloor} \sum_{i=0}^{k} = \sum_{j=0}^{\lfloor (t-l)\beta \rfloor} \sum_{i=0}^{\lfloor (t-l)\beta \rfloor - j}$ and we find that the last sum may be rewritten as:

$-\frac{(1-\rho_2)}{\rho_2 - \rho_1\beta} \sum_{l=0}^{\lfloor t \rfloor} \sum_{j=0}^{\lfloor (t-l)\beta \rfloor} (-1)^j \frac{(j+l)!}{j!l!} \frac{(\rho_1\beta)^j \rho_2^l}{(\rho_2 - \rho_1\beta)^{j+l}} q(\beta(t-l) - j, \rho_1)$. Inserting in the expression above we get:

$$I(t,\beta,\rho_1,\rho_2) = \frac{(1-\rho_1)}{\rho_2 - \rho_1\beta} \sum_{k=0}^{\lfloor t\beta \rfloor} \sum_{j=0}^{\lfloor t - k\beta^{-1} \rfloor} (-1)^k \frac{(k+j)!}{k!\,j!} \frac{(\rho_1\beta)^k \rho_2^j}{(\rho_2 - \rho_1\beta)^{k+j}} q(t - j - k\beta^{-1}, \rho_2)$$

$$-\frac{(1-\rho_2)}{\rho_2 - \rho_1\beta} \sum_{j=0}^{\lfloor t \rfloor} \sum_{k=0}^{\lfloor (t-j)\beta \rfloor} (-1)^k \frac{(j+k)!}{k!\,j!} \frac{(\rho_1\beta)^k \rho_2^j}{(\rho_2 - \rho_1\beta)^{j+k}} q(\beta(t-j) - k, \rho_1)$$

By (7.36) we get W_2 as:

$$W_2(t,\rho_1,\rho_2,b_1,b_2) = \frac{(1-\rho_1)}{y-x} \sum_{k=0}^{\lfloor \frac{t}{b_1} \rfloor} \sum_{j=0}^{\lfloor \frac{t-kb_1}{b_2} \rfloor} p_{k,j}(\zeta,\eta) q(\frac{t - kb_1 - jb_2}{b_2}, \rho_2) + \frac{(1-\rho_2)}{x-y} \sum_{k=0}^{\lfloor \frac{t}{b_2} \rfloor} \sum_{j=0}^{\lfloor \frac{t-kb_2}{b_1} \rfloor} p_{k,j}(\zeta,\eta) q(\frac{t - kb_1 - jb_2}{b_1}, \rho_1)$$

where $p_{k,j}(\zeta,\eta) = \binom{k+j}{j} \eta^k \zeta^j$ and $\zeta = \frac{y}{y-x}$ and $\eta = \frac{x}{x-y}$; where we have defined $x = \frac{\rho_1}{b_1}$ and $y = \frac{\rho_2}{b_2}$; and where we also have $\zeta + \eta = 1$.

Then by direct differentiation of W_2 and using (7.30) we get the desired convolution (7.34).

We now move to the interesting case with two groups of queues with different service times in each group**.** Specifically we consider two groups of M/D/1 queues of size n_i, each with load ρ_i and service time $b_i, i = 1,2$, and we denote $W(t,\rho_i,b_i) = q(t/b_i, \rho_i)$ the DF of the waiting time in each queue, $i = 1,2$. Then the DF of the convolution is given by the following sums:

$$W(t,\rho_1,\rho_2,b_1,b_2,n_1,n_2) = \tag{7.38}$$

$$(1-\rho_1)^{n_1} \sum_{k=0}^{\lfloor \frac{t}{b_1} \rfloor} \sum_{j=0}^{\lfloor \frac{t-kb_1}{b_2} \rfloor} \sum_{s=0}^{n_2-1} (1-\rho_2)^{n_2-s-1} \zeta p_{n_1-1,n_2-s-1k,j}(\zeta,\eta) \left[q^{s+1}(\frac{t-kb_1-jb_2}{b_2}, \rho_2) - q^{s+1}(\frac{t-kb_1-jb_2}{b_2} - 1, \rho_2) \right] +$$

$$(1-\rho_2)^{n_2} \sum_{k=0}^{\lfloor \frac{t}{b_2} \rfloor} \sum_{j=0}^{\lfloor \frac{t-kb_2}{b_1} \rfloor} \sum_{s=0}^{n_1-1} (1-\rho_1)^{n_1-s-1} \eta p_{n_1-s-1,n_2-1,k,j}(\zeta,\eta) \left[q^{s+1}(\frac{t-kb_1-jb_2}{b_1}, \rho_1) - q^{s+1}(\frac{t-kb_1-jb_2}{b_1} - 1, \rho_1) \right]$$

where we have

$$p_{K_1,K_2,k,j}(\zeta,\eta)=\binom{K_1+K_2}{K_2}\binom{k+j}{j}\zeta^{K_1}\eta^{K_2}\sum_{l=\max[0,j-K_1]}^{\min[K_2+j,k+j]}(-1)^{j-l}\binom{K_2+j}{l}\binom{K_1+k}{k+j-l}\zeta^{l}\eta^{k+j-l} \tag{7.39}$$

and where we define $\zeta=\dfrac{y}{y-x}$ and $\eta=\dfrac{x}{x-y}$; with $x=\frac{\rho_1}{b_1}$ and $y=\frac{\rho_2}{b_2}$; (and where we also have $\zeta+\eta=1$).

To show (7.38) we use (7.26) on equation (7.34) and by using equation (F.23) in Appendix F, we get the following expression for $W(t,\rho_1,\rho_2,b_1,b_2,n_1,n_2)$:

$$W(t,\rho_1,\rho_2,b_1,b_2,n_1,n_2)= \tag{7.40}$$

$$(1-\rho_1)^{n_1}\sum_{k=0}^{\left\lfloor\frac{t}{b_1}\right\rfloor}\sum_{j=0}^{\left\lfloor\frac{t-kb_1}{b_2}\right\rfloor}\sum_{s=0}^{n_2-1}(1-\rho_2)^{n_2-s-1}G^2_{n_1-1,n_2-s-1,k,j}(\rho_1,\rho_2)\left[q^{s+1}(\frac{t-kb_1-jb_2}{b_2},\rho_2)-q^{s+1}(\frac{t-kb_1-jb_2}{b_2}-1,\rho_2)\right]$$

$$+(1-\rho_2)^{n_2}\sum_{k=0}^{\left\lfloor\frac{t}{b_2}\right\rfloor}\sum_{j=0}^{\left\lfloor\frac{t-kb_2}{b_1}\right\rfloor}\sum_{s=0}^{n_1-1}(1-\rho_1)^{n_1-s-1}G^1_{n_1-s-1,n_2-1,k,j}(\rho_1,\rho_2)\left[q^{s+1}(\frac{t-kb_1-jb_2}{b_1},\rho_1)-q^{s+1}(\frac{t-kb_1-jb_2}{b_1}-1,\rho_1)\right]$$ where we have

$$G^2_{K_1,K_2,k,j}(\rho_1,\rho_2)=\frac{\rho_2\dfrac{d^{K_1}}{d\rho_1^{K_1}}\dfrac{d^{K_2}}{d\rho_2^{K_2}}\left\{\rho_1^{K_1}\rho_2^{K_2-1}\zeta p_{k,j}(\zeta,\eta)\right\}}{K_1!K_2!}=\zeta F_{K_1,K_2,k,j}(x,y)=\zeta p_{K_1,K_2,k,j}(\zeta,\eta) \text{ and} \tag{7.41}$$

$$G^1_{K_1,K_2,k,j}(\rho_1,\rho_2)=\frac{\rho_1\dfrac{d^{K_1}}{d\rho_1^{K_1}}\dfrac{d^{K_2}}{d\rho_2^{K_2}}\left\{\rho_1^{K_1-1}\rho_2^{K_2}\eta p_{k,j}(\zeta,\eta)\right\}}{K_1!K_2!}=\eta F_{K_1,K_2,k,j}(x,y)=\eta p_{K_1,K_2,k,j}(\zeta,\eta) \tag{7.42}$$

where we define

$$p_{K_1,K_2,k,j}(\zeta,\eta)=F_{K_1,K_2,k,j}(x,y)=(-1)^k\binom{k+j}{j}\frac{(y-x)}{K_1!K_2!}\frac{d^{K_1}}{dx^{K_1}}\frac{d^{K_2}}{dy^{K_2}}\left\{\frac{x^{k+K_1}y^{j+K_2}}{(y-x)^{k+j+1}}\right\}, \tag{7.43}$$

and where we as above have taken $\zeta=\dfrac{y}{y-x}$ and $\eta=\dfrac{x}{x-y}$; with $x=\frac{\rho_1}{b_1}$ and $y=\frac{\rho_2}{b_2}$; and where we have $\zeta+\eta=1$. In Appendix F the actual form of $F_{K_1,K_2,k,j}(x,y)$ is found (see equation (F.37)) and then by inserting for ζ and η we find $p_{K_1,K_2,k,j}(\zeta,\eta)$ as given by (7.39).

With the result above we have a tool to analyse end-to-end delay for realistic scenarios in large-scale IP networks, also allowing to include the access part of the network. A typical (realistic) scenario would be to include two (or more) low capacity queues (links) together with a number of rather high capacity queues (links) representing the core. This will provide a realistic estimate of this important QoS parameter for large networks. The basic building blocks in the end-to-end delay distribution is given in terms of convolutions $q^k(x,\rho)$, which are easily obtained by equation (7.28).

7.3.3 Asymptotic approximations

Below we shall consider the general case of convolution of M/G/1 queues where we consider the convolution of a chain consisting of N groups of queues of size n_j, where all the queues in each group are identical, but allowing for having different service time distributions between the groups, ($j = 1,..., N$). In this case the convolution will have the following LST:

$$\tilde{W}(s) = \prod_{j=1}^{N} \left(\frac{1-\rho_j}{1-\rho_j \hat{B}_j(s)} \right)^{n_j} \tag{7.44}$$

where we have a total of N groups of queues of size n_j and where ρ_j and $\hat{B}_j(s)$ are the load and the LST of the remaining service time in group $j\,(=1,..., N)$. As before we let $K = n_1 + ... + n_N$ be the total number of queues in the chain.

To this end some different types of approximations exist for sums of independent random variables in general. The first one is the normal approximation quoting that a sum of identically independent random variables approaches a normal distribution when the number of variables increases. Since the normal distribution is characterized by its two first (lowest) moments, this leads to the following approximation for the PDF and DF of the convolution:

$$w^{NA}(t) = \frac{1}{\sigma} \varphi(\frac{t-m}{\sigma}) \text{ and} \tag{7.45}$$

$$W^{NA}(t) = 1 - \phi(\frac{t-m}{\sigma}) \text{ where} \tag{7.46}$$

m and σ^2 are the mean and the variance of the total end-to-end queueing delay for the chain given by

$$m = \sum_{j=1}^{N} n_j m_j \text{ and } \sigma^2 = \sum_{j=1}^{N} n_j \sigma_j^{\,2} \tag{7.47}$$

where $m_j = E[W_j]$ and $\sigma_j^2 = E[W_j^2] - m_j^2$ is the mean and variance of the queueing delay in the corresponding single server M/G/1 queue given in terms of the three first moments of the service time distribution (and also the load) through:

$$m_j = E[W_j] = \frac{E[B_j^2]}{2E[B_j]} \frac{\rho_j}{1-\rho_j} \text{ and } \sigma_j^2 = E[W_j^2] - m_j^2 = \frac{E[B_j^3]}{3E[B_j]} \frac{\rho_j}{1-\rho_j} + m_j^2 \tag{7.48}$$

further $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the standard normal density and $\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$ is the standard normal integral.

The second and quite different approximation (though also involving normal distributions) is based on LD (Large Deviation) theory. This can for instance be reflected through the inversion integral of the LST given by

$$w(t) = \frac{1}{2\pi i} \int_{\gamma} e^{st} \tilde{W}(s) ds \tag{7.49}$$

Formally this approximation is obtained by considering the asymptotic behaviour of $w(Kx)$ for large values of K while x is fixed. We also assume that $n_j = \gamma_j K$ and γ_j are bounded away from zero as K increases. With these substitutions the inversion integral with argument Kx may be written

$$w(Kx) = \frac{1}{2\pi i} \int_{\gamma} e^{-Kg(s,x)} ds \text{ where} \tag{7.50}$$

$$g(s,x) = \sum_{j=1}^{N} \gamma_j \left[\log(1 - \rho_j \hat{B}_j(s)) - \log(1 - \rho_j) \right] - sx \tag{7.51}$$

Approximations of the contour integral above may be found by the method of Steepest Descent (or the Saddle Point method). (See for instance [Won89] for a thorough description of the method.) This is done by choosing the contour so that it passes through real axis where the maximum value for $g(s,x)$ is attained, which may be found by setting the derivative with respect to s equal zero. The corresponding value of $s = s^*(x)$, called the saddle point, is then the solution of the equation:

$$\sum_{j=1}^{N} \gamma_j \frac{\rho_j \hat{B}_j'(s)}{1 - \rho_j \hat{B}_j(s)} + x = 0 \tag{7.52}$$

The corresponding approximation is found by expanding the exponent (7.51) to second order in s (remembering that first order contribution vanishing due to equation (7.52)), and then factorising out the constant part, and transforming the contour integral to an exponential integral by neglecting higher order terms. One gets:

$$\frac{1}{2\pi i} \int_{\gamma} e^{-Kg(s,x)} ds \approx \frac{1}{2\pi} e^{-Kg(s^*(x),x)} \int_{y=-\infty}^{\infty} e^{-Kg''(s^*(x),x)\frac{y^2}{2}} dy = \frac{1}{\sqrt{-2\pi K g''(s^*(x),x)}} e^{-Kg(s^*(x),x)}$$

The corresponding approximation yields $w^{LD}(t) = f(\frac{t}{K})$ where

$$f(x) = \frac{1}{\sqrt{2\pi K h(x)}} e^{-Kg(s^*(x),x)} \tag{7.53}$$

where we find by differentiating (7.51):

$$h(x) = -g''(s^*(x),x) = \sum_{j=1}^{N} \gamma_j \left[\left(\frac{\rho_j \hat{B}_j'(s^*(x))}{1-\rho_j \hat{B}_j(s^*(x))} \right)^2 + \frac{\rho_j \hat{B}_j''(s^*(x))}{1-\rho_j \hat{B}_j(s^*(x))} \right] \tag{7.54}$$

The corresponding expression for the DF as an inversion integral is:

$$W(Kx) = \frac{1}{2\pi i} \int_{\gamma} \frac{e^{-Kg(s,x)}}{s} ds \tag{7.55}$$

If we apply the Saddle Point method directly on (7.55) we must make a distinction between two cases depending on the sign of the saddle point $s^*(x)$. It turns out that $s^*(x)$ is positive for small values of x but becomes negative as x increases. In the latter case we have to pick up the residue at $s = 0$ and we get the following approximation by combining the two cases:

$W^{LD}(t) = 1 - F(\frac{t}{K})$ where

$$F(x) = H(s^*(x)) - \frac{1}{s^*(x)\sqrt{2\pi K h(x)}} e^{-Kg(s^*(x),x)} \tag{7.56}$$

where $H(t)$ is the unit step function.

It turns out that the "ordinary" Saddle Point method given by (7.56) fails to approximate the integral (7.55) when the root $s^*(x)$ is close to the pole at $s = 0$. It is, however, possible to extract that pole and obtain a Uniform Asymptotic Approximation (UAA) that is uniform with respect to x and also yields in the area where $s^*(x)$ is close to zero. The actual method is described in the book by Wong [Won89]. We find the following approximation:

$W^{UAA}(t) = 1 - G(\frac{t}{K})$ where

$$G(x) = \phi(-\mathrm{sgn}(s^*(x))\sqrt{2Kg(s^*(x),x)}) - e^{-Kg(s^*(x),x)} \left\{ \frac{1}{s^*(x)\sqrt{2\pi K h(x)}} - \frac{\mathrm{sgn}(s^*(x))}{2\sqrt{\pi K g(s^*(x),x)}} \right\}$$

(7.57)

It is quite easy to see that the function defined by $G(x)$ does not have any singularities for values of x giving saddle point close to zero. By expanding the brackets to third order in $s^*(x)$ we find:

$$G(x) = \phi(-\mathrm{sgn}(s^*(x))\sqrt{2Kg(s^*(x),x)}) + \frac{h_3(x)}{6h(x)} \frac{e^{-Kg(s^*(x),x)}}{\sqrt{2\pi Kh(x)}} \text{ when } s^*(x) \text{ is close zero}$$

and where $h_3(x) = -g'''(s^*(x),x)$.

The main complexity in applying the Saddle Point methods described above is to locate the saddle points $s^*(x)$, for which we have to solve equation (7.52) for each value of x and this could limit the value of this type of approximations due to the computation time required to solve the equation numerically. When it comes to the accuracy it seems that the UAA will give results very close to the exact convolutions for a very broad range of parameters also including cases where the asymptotic is not fulfilled, i.e. for chains of relatively small sizes. The "classical" LD approximation, however, gives strict upper bound on the convolution and can therefore be a desirable method to apply for this reason.

7.4 Some numerical examples

End-to-end delay is one of the most important QoS parameters for real time services like voice and video. In an all IP-network the end-to-end delay for a particular stream will be the sum of the delay obtained in a cascade of routers (from the sender to the receiver). The total end-to-end delay will then consist of the waiting times in each node plus service times (transmission times onto the links). In the examples below we shall consider a particular chain of routers in a packet network and we assume that the routers have output buffers with no extra internal delay due to processing of the packets. The network model is shown in figure 7.1.

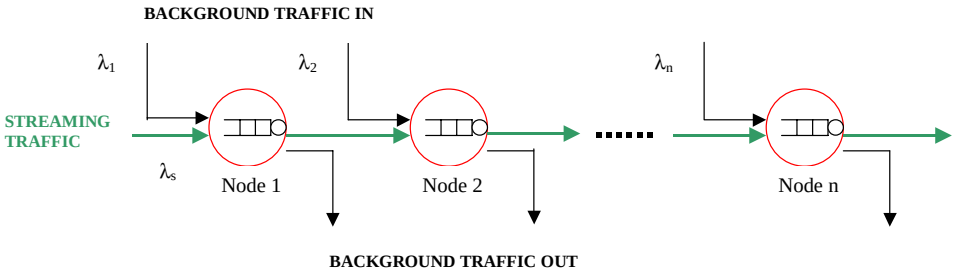

Figure 7.1: The queueing model of a particular packet-stream traversing n-nodes.

We have made the following assumptions:

- The queueing discipline is FIFO for all the queues.
- The background traffic enters (and leaves) node i according to a Poisson process with rate λ_i.
- The streaming traffic enters node 1 according to a Poisson process with rate λ_s and leaves node n.
- All packets (both background and streaming traffic) have constant (and equal) packet lengths of p_L and all the links have capacity of C_i.

The corresponding parameters used by the convolution approximation are the service times per packets, which are constant and equal to $b_i = p_L / C_i$, and the load on the different nodes given by $\rho_i = (\lambda_s + \lambda_i)b$.

7.4.1 Cases with identical nodes that are equally loaded

Based on the simplicity of the way the convolutions are performed if all the nodes are identically makes it feasible to evaluate end-to-end convolutions for rather large numbers of queues. The numerical algorithms derived for the cases with constant service times (M/D/1 model) make it possible to calculate the corresponding CDF (Complementary Distribution Function) for quite large paths containing up to 20 queues in series and will therefore cover paths that are of “real size” in networks of today.

In figure 7.2-figure 7.4 we demonstrate how the PDF of the end-to-end queueing delay “converges” when waiting time is scaled by the total service times (end-to-end). Typically up to a series of 15 queues the distinction is pronounced but for larger number of queues than that the difference between the curves seems to be small. Another important observation is that it seems that all the distributions have a common intersection point approximately around 0.1 and for that point on the distributions are bounded “from above” by the curves for smaller chains. In practice this means that it is sufficient to calculate the distributions for chains up to say 15 queues and then use the scaled result (for 15 queues) as an approximation for larger chains.

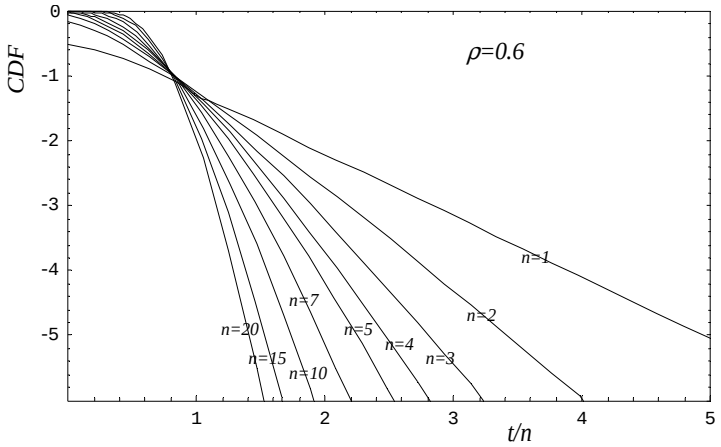

Figure 7.2: Logarithmic plot of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.6 scaled by the total service times (end-to-end) for different size of the series n=1,2,3,4,5,7,10,15 and 20.

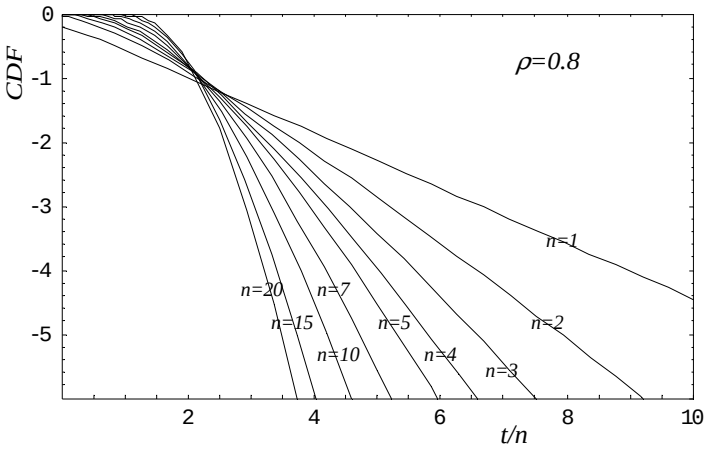

Figure 7.3: Logarithmic plot of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.8 scaled by the total service times (end-to-end) for different size of the series n=1,2,3,4,5,7,10,15 and 20.

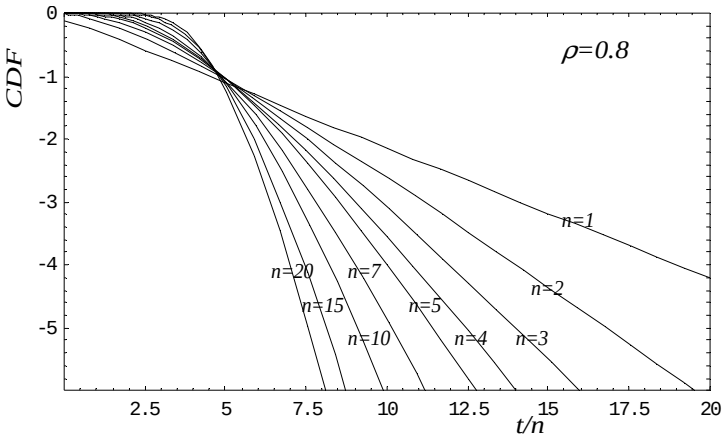

Figure 7.4: Logarithmic plot of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.9 scaled by the total service times (end-to-end) for different size of the series n=1,2,3,4,5,7,10,15 and 20.

Below we have made a comparison of the different approximations considered in section 7.3.3: NA (Normal Approximation), LD (Large Deviation) and UAA (Uniform Asymptotic Approximation) with the convolution method. In figure 7.5-figure 7.7 we have plotted the different approximations of the CDF (with the corresponding parameters as figure 7.2-figure 7.4). We find that the UAA gives an excellent approximation of the CDF that yields uniformly for all values of the argument. If the size of the chain is greater than 5 we find that the relative error is less than 0.25% and the difference is not visible in the graphs. Also for the single queue case we find that the UAA gives quite accurate estimates especially in the tail of the distribution. The maximum relative error in this case is 12% for a load of 0.6 and decreases to 2% for a load of 0.9. Although the UAA gives very accurate results it does not bound the CDF obtained by convolution.

For the LD approximation we observe:

- it always bounds the actual distribution
- it only applies in the tail of the distribution
- it is fairly good also for relatively small chains, typically from the size of 5.

On the contrary the Normal Approximation will typically have nearly opposite properties:

- it does not bound the actual distribution especially not in the tail
- it applies also near the origin
- from the chain size of 15 on it gives fairly good approximations except for the far tail (where this type of approximation fails).

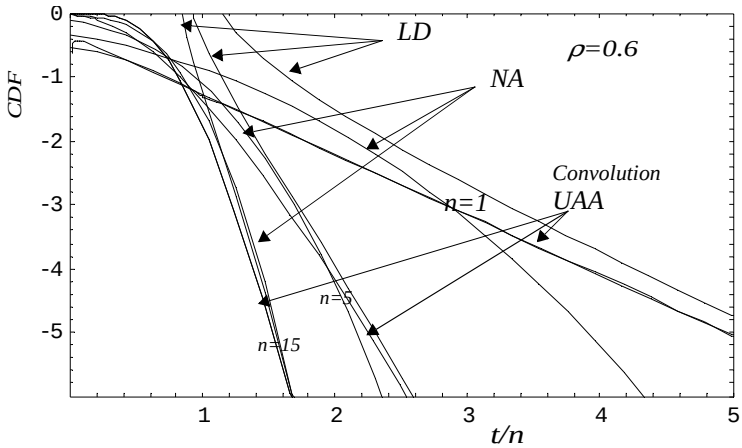

Figure 7.5: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.6 scaled by the total service times (end-to-end) for different sizes of the series n=1,5 and 15. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

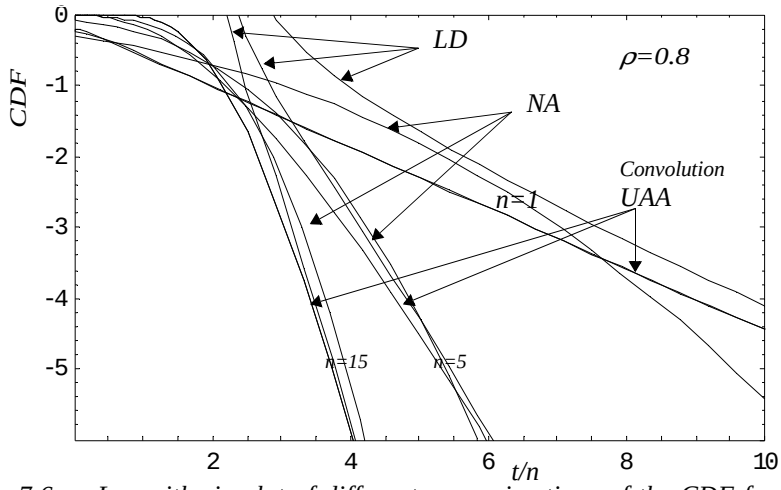

Figure 7.6: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.8 scaled by the total service times (end-to-end) for different sizes of the series n=1,5 and 15. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

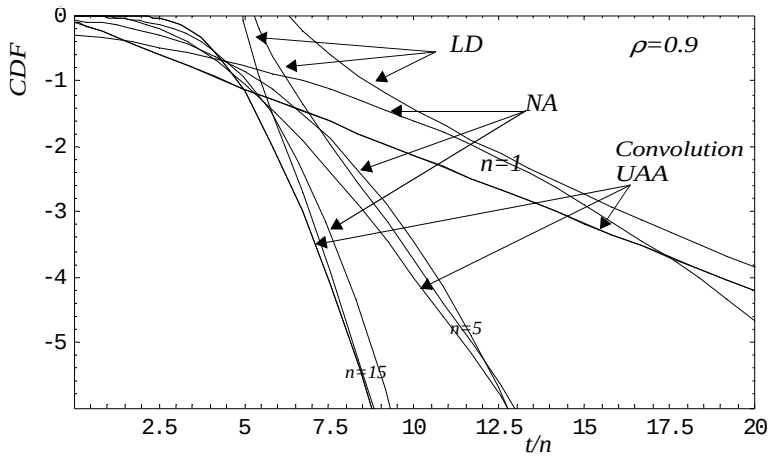

Figure 7.7: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of equally loaded queues with load equal to 0.9 scaled by the total service times (end-to-end) for different sizes of the series n=1,5 and 15. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

In network engineering it is important to make some statement of the guarantee of the end-to-end delay. This guarantee is often given in terms of probabilities, for instance that the delay shall not exceed a particular target value by some small probability. So we would like to find the $\alpha = 1 - \bar{\beta}$ quantile for small values of $\bar{\beta}$. We therefore have to find the value $t = t_n^{\beta}(\rho)$ that solves the equation

$$q^n(t,\rho) = 1 - \beta \text{ where} \tag{7.58}$$

$q^n(t,\rho)$ is the DF of the convolution. In figure 7.8-figure 7.10 we have given a logarithmic plot of the quantiles as a function of the load for different values of size of the chains ranging from 1- 20 queues, and for three guarantee levels 0.1, 0.01 and 0.01 respectively. (In all the figures the quantiles are scaled in units of one queue service time.) As an example suppose that the end-to-end QoS requirement says that only 1% of the packets shall have an end-to-end delay longer than 50 packet transmission times. (For 2 Mbit/s links and constant packet lengths of 200 bytes this corresponds to end-to-end delay of 40 ms.) Then by figure 7.11 we find the following load limits (as a function of the size of the chains):

- $\rho_{\max} \approx 0.74$ if the traffic traverses $n = 20$ nodes
- $\rho_{\max} \approx 0.79$ if the traffic traverses $n = 15$ nodes
- $\rho_{\max} \approx 0.85$ if the traffic traverses $n = 10$ nodes and
- $\rho_{\max} \approx 0.89$ if the traffic traverses $n = 5$ nodes

This simple example shows that the load limits imposed on the routers in a network should to some extent depend on the actual size of the network. A large network containing long paths should operate at slightly lower load than a corresponding network with shorter paths.

These load limits could also be found from equation (7.58) by solving for the load while keeping the quantile fixed. In figure 7.11 we have plotted this load limit as a function of the guarantee level β given in logarithmic scale for two chain sizes, 5 and 10, and for four values of quantile of the end-to-end delay. (In this figure the quantiles are scaled by the total service time end-to-end.)

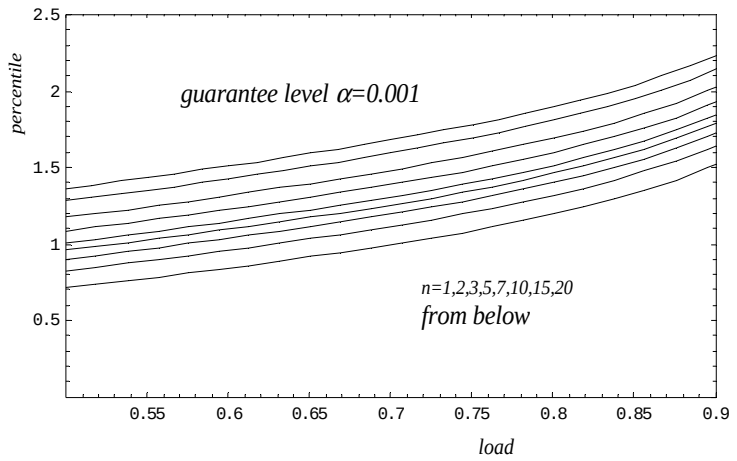

Figure 7.8: Logarithmic plot of the 0.001 percentile of the end-to-end waiting time for chains of n=1,2,3,4,5,7,10,15 and 20 queues as a function of the load. The percentile is scaled to one packet transmission time.

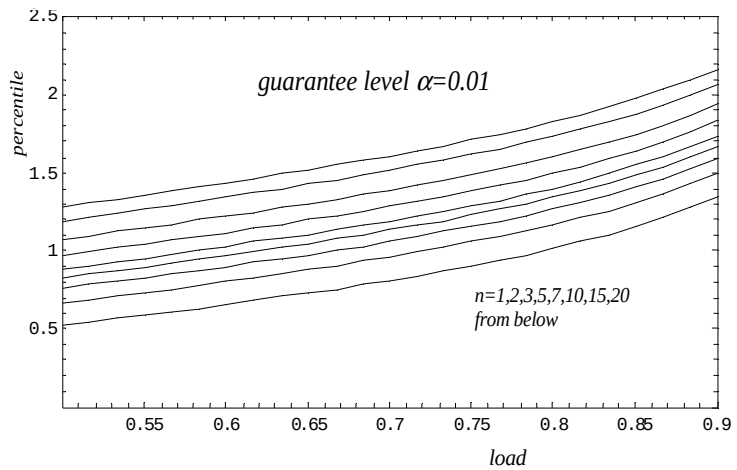

Figure 7.9: Logarithmic plot of the 0.01 percentile of the end-to-end waiting time for chains of n=1,2,3,4,5,7,10,15 and 20 queues as a function of the load. The percentile is scaled to one packet transmission time.

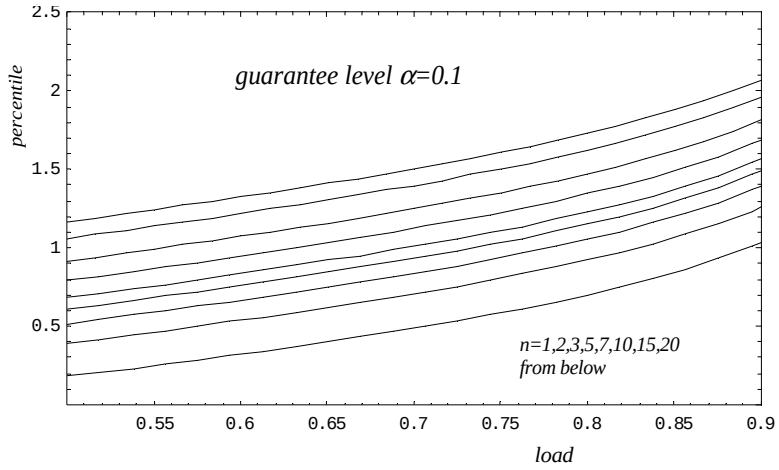

Figure 7.10: Logarithmic plot of the 0.1 percentile of the end-to-end waiting time for chains of n=1,2,3,4,5,7,10,15 and 20 queues as a function of the load. The percentile is scaled to one packet transmission time.

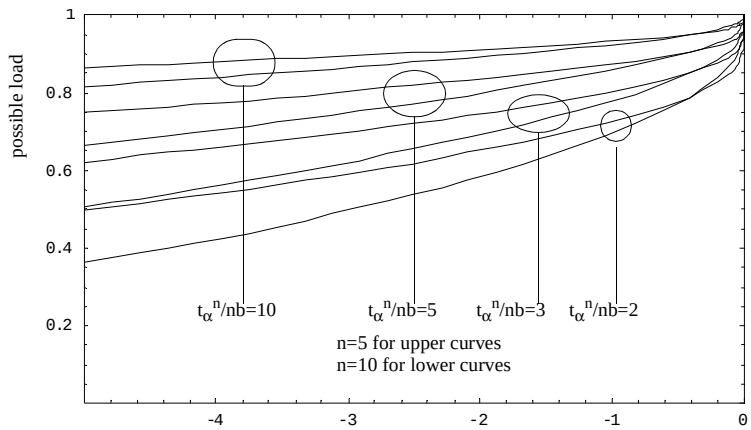

Figure 7.11: The maximal possible load for a chain of equally loaded queues as a function of the guarantee level (in logarithmic scale), where the corresponding percentile for the end-to-end queueing delay, scaled by the total service times (end-to-end) equals 2,3,5 and 10, and the number of queues (in the chain) equals 5 and 10.

7.4.2 Numerical examples including one or more low capacity access links

An end-to-end path in an IP-network will typically include one or more low capacity access links that are well below the capacity deployed in the core networks. On the other hand the core part of a path will typically consist of a rather large number of hops and the core part could therefore contribute to the end-to-end delay by having a large number of hops. Since the users observe their QoS on an end-to-end basis it is important to have models that include both low capacity access parts as well as the high capacity core networks that may have considerable diameter in terms of hops. In section 7.3 we have given the end-to-end queueing delay for the convolution of two groups of M/D/1 queues where we may have different loads in each group, and more important, also allowing for having different capacity (service times) in each groups.

As the final example we consider a typical example where we have a path consisting of an upstream access part, a core network with multiple hops and eventually a downstream access part. In the example below we have taken the following parameters:

- The access part consists of one ore two low capacity links (with the same capacity).
- The core part consists of five or ten links all with the same capacity.
- The access link capacity is 1/10 of the corresponding core link capacity, giving for instance the access capacity of approximately 15 Mbit/s if the core links are STM-1 link at approximately 150 Mbit/s.

This example could for instance represent the case of a typical DSL (Digital Subscriber Line) access line that is connected to a core network with minimum STM-1 links (or higher). The CDF of the end-to-end waiting times are plotted for some typical load levels in figure 7.12-figure 7.15 and some quantiles are given in table 7.1, all scaled by the packet transmission time for the low capacity link. The main influence on the end-to-end performance for this particular example will come from the access part if the network elements are more or less equally loaded. This is easily seen from the figures below. The difference between the case $n_2 = 0$ (i.e. neglecting the influence from the core network) and $n_2 = 10$ is limited. In the example where we assume that the access is less loaded, as in figure 7.15, the situation is different and the core will contribute to a significant part of the end-to-end delay.

In figure 7.16- figure 7.19 we have also given plots of some of the cases given in figure 7.12-figure 7.15 for the approximations described in section 7.3.3. For this example with two groups of queues with only one or two queues in the first group we would not expect that the asymptotic would be very accurate. This is indeed the case for the NA. For the LD the curves are well above the corresponding obtained by the convolution approach. When it comes to the UAA this approximation is surprisingly good for nearly all the cases considered. There is a small region close to point where the actual saddle point is close to zero, where the UAA fails. These difficulties could however be ruled out by expanding the expression for assuming that the saddle point is close to zero. (See section 7.3.3 for closer explanation.) The relative error is found to be less than 3% in the regions outside this region for all the cases considered. Another observation is that the saddle points seem to be locat-

ed at 0.5 quantile of the end-to-end queueing delay. By considering smaller quantiles we will therefore be well away from the critical area giving saddle points close to zero.

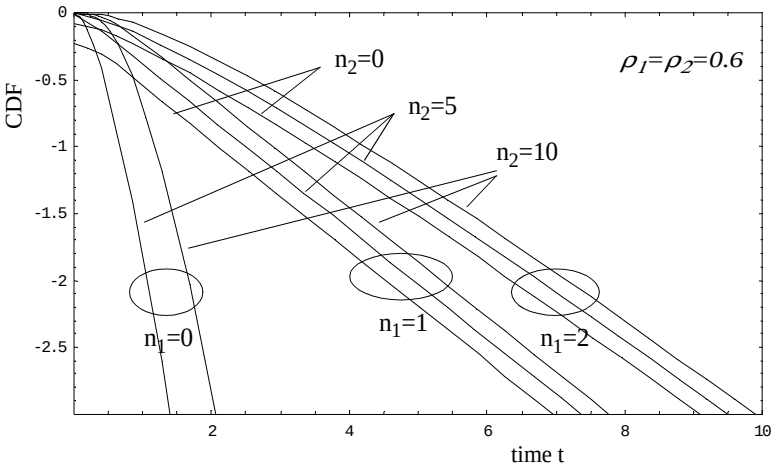

Figure 7.12: Logarithmic plot of the CDF for end-to-end queueing delay for a series of two groups of size n_1=0,1,2 and n_2=0,5,10 that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.6 and the time unit is scaled to one packet transmission time for the low capacity group.

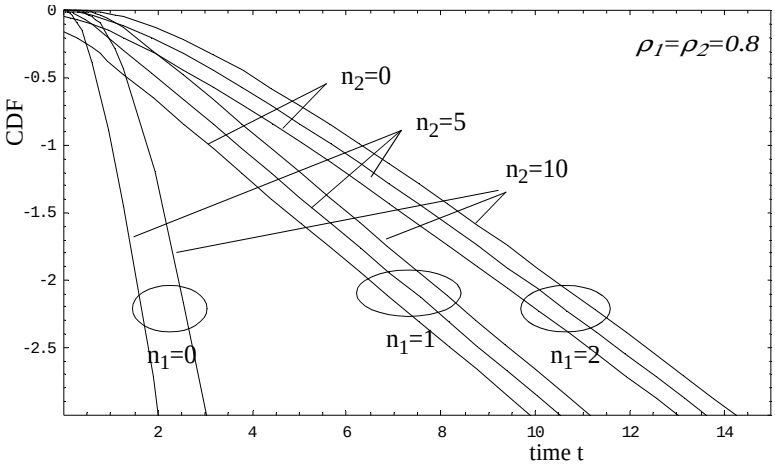

Figure 7.13: Logarithmic plot of the CDF for end-to-end queueing delay for a series of two groups of size n_1=0,1,2 and n_2=0,5,10 that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.7 and the time unit is scaled to one packet transmission time for the low capacity group.

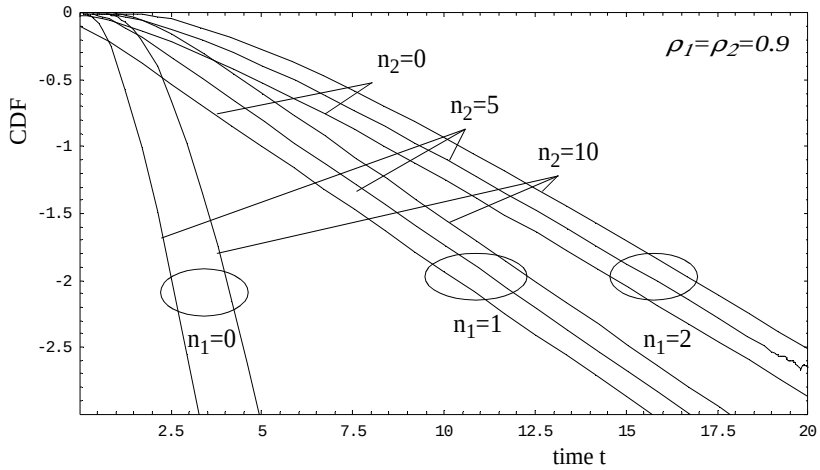

Figure 7.14: Logarithmic plot of the CDF for end-to-end queueing delay for a series of two groups of size n_1=0,1,2 and n_2=0,5,10 that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.8 and the time unit is scaled to one packet transmission time for the low capacity group.

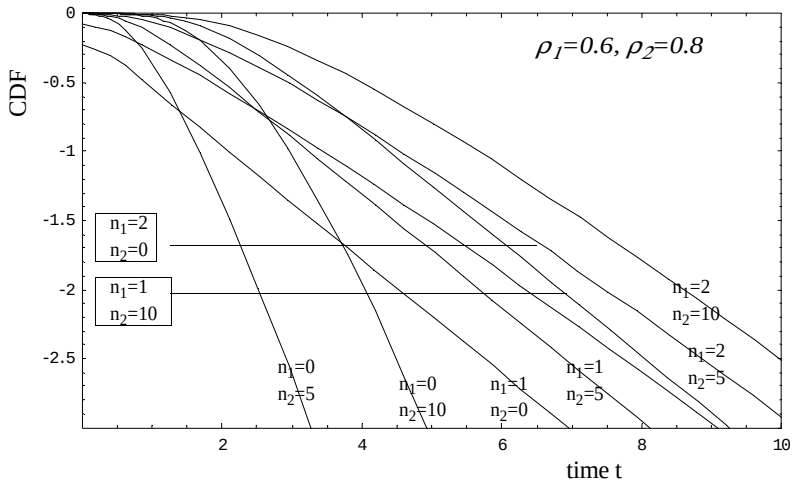

Figure 7.15: Logarithmic plot of the CDF for end-to-end queueing delay for a series of two groups of size n_1=0,1,2 and n_2=0,5,10 that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in the first group is 0.6 while the load in the second group is 0.8 and the time unit is scaled to one packet transmission time for the low capacity group.

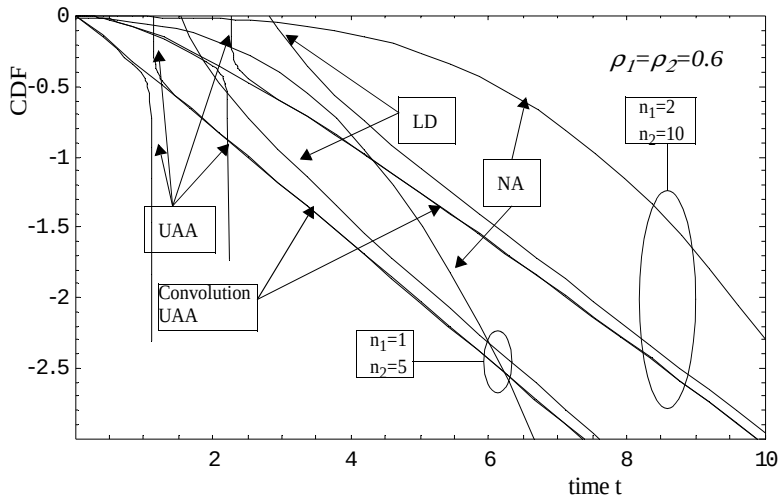

Figure 7.16: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of two groups of size $n_1=1$, $n_2=5$ and $n_1=2$, $n_2=10$ that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.6 and the time unit is scaled to one packet transmission time for the low capacity group. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation).

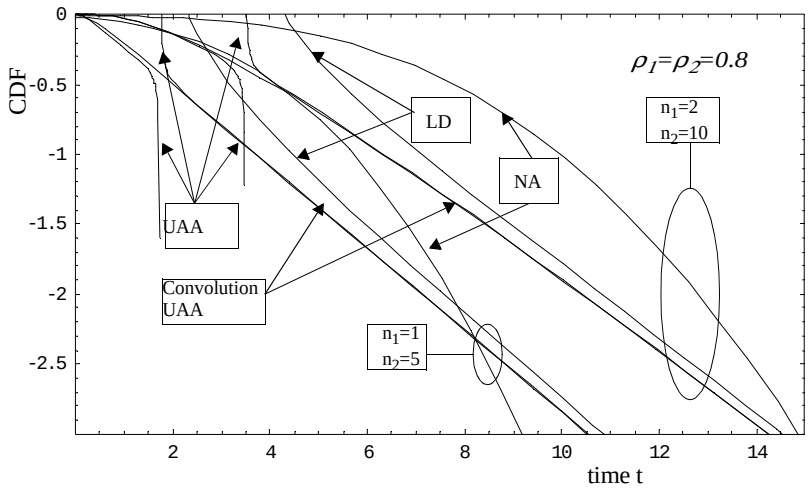

Figure 7.17: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of two groups of size $n_1=1$, $n_2=5$ and $n_1=2$, $n_2=10$ that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.7 and the time unit is scaled to one packet transmission time for the low capacity group. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

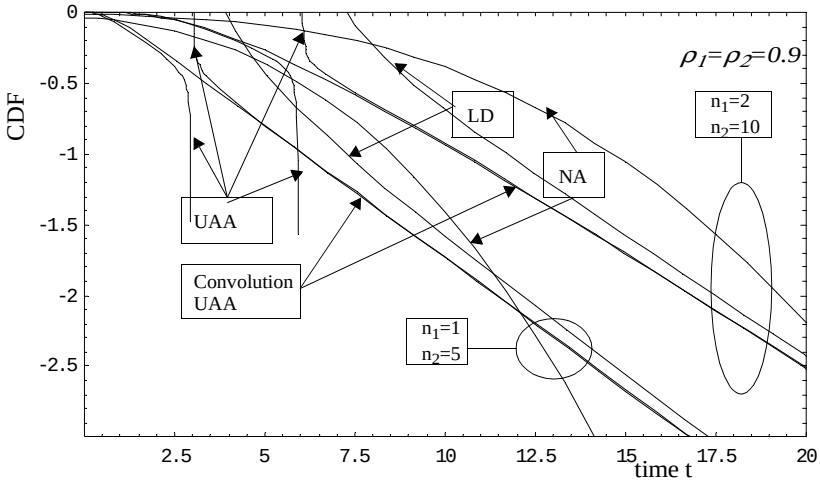

Figure 7.18: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of two groups of size $n_1=1$, $n_2=5$ and $n_1=2$, $n_2=10$ that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in both groups is 0.7 and the time unit is scaled to one packet transmission time for the low capacity group. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

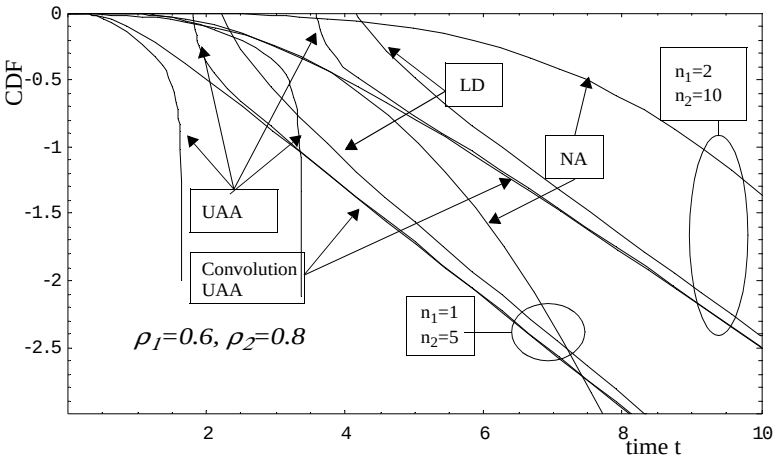

Figure 7.19: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for a series of two groups of size $n_1=1$, $n_2=5$ and $n_1=2$, $n_2=10$ that are equally loaded. The capacity in the first group is 1/10 of that in the second group. The load in the first group is 0.6 while the load in the second group is 0.8 and the time unit is scaled to one packet transmission time for the low capacity group. (NA-Normal Approximation, LD -Large Deviation, UAA -Uniform Asymptotic Approximation)

In table 7.1 we have given some quantiles for the example in the discussion on the basis of the UAA model. The actual numerical values are checked against the graphs obtained by the convolution approach in figure 7.13 and figure 7.14 and we conclude that the accuracy is satisfactory.

Table 7.1: The different quantiles for the end-to-end queueing delay for the example above with two groups of queues where the capacity in the first group is 1/10 of that in the second group and all the queues are equally loaded and the time unit is scaled to one packet transmission time for the low capacity group.

n_1	n_2	ρ=0.7		ρ=0.8	
		β=0.01	β=0.001	β=0.01	β=0.001
0	5	1.76	2.23	2.93	3.66
0	10	2.67	3.23	4.46	5.35
1	0	6.50	9.92	10.38	15.74
1	5	7.12	10.54	11.44	16.80
1	10	7.75	11.16	12.51	17.86
2	0	9.18	13.02	14.76	20.78
2	5	9.79	13.64	15.81	21.84
2	10	10.41	14.26	16.87	22.90

One of the questions in mind for the given scenario is the following: What would be the proper guarantee for the end-to-end queueing delay (not including other delay components which must be added) for such a scenario? If we assume that the packet lengths are limited to 1500 bytes the corresponding transmission times (service times in queueing terminology) are approximately 0.8 ms for the access links and 0.08 ms for the high capacity links. By the table above we have for instance the 0.999 quantile ($\beta = 0.001$) for a network loaded at 0.8 to be approximately 17.9 ms for the case with two access links and ten core links. The corresponding result with the slightly "looser" 0.99-quantile ($\beta = 0.01$) is 13.5 ms.

7.5 Concluding remarks

The end-to-end delay is an important QoS parameter for real time services. In IP networks deploying statistical multiplexing this parameter will depend on several parameters like traffic pattern, background traffic, number of hops, the network load, etc. The method proposed gives an effective way of calculating the end-to-end delay distribution. It is shown that load limit will depend on the size of the network indicating that a larger network should be slightly less loaded than a small network provided that the links have the same capacity.

The first method proposed in this chapter applies only for links with equal capacity, for instance the core part of an IP network. We also give the corresponding results for a chain containing two groups of links with different capacity in each group. This is a particularly interesting case and makes it possible to model a path in an IP network that includes both access and core links. The latter model is however far more complex and requires more computing effort to obtain the desired results.

8

Convolution of a given number of waiting times of M/G/1 non-preemptive priority queues having identical service time distributions

8.1 Some preliminary considerations

In the following chapter we shall consider a DiffServ scenario for the end-to-end delay for typical RT (Real Time) traffic in a large scale IP-network. We consider a path in the network consisting of a given number of (say) K nodes and the aim is to calculate the CDF and the quantiles of the queueing delay for that particular path. We assume that each node may be considered as a non-preemptive priority queueing system with two priority classes where the RT traffic is scheduled as highest priority and the Best Effort (BE) type traffic is scheduled as lower (second) priority.

To calculate the delay of a particular path we make the same assumption (approximation) as in chapter 7: *All nodes in the end-to-end path are statistically independent.* This is the key assumption for the model and makes it possible to obtain the end-to-end delay by convolution. Under which conditions the independent assumption applies is not quite clear, but it seems to be reasonable for rather thin streams where the aggregate flows split at each node and are mixed with traffic from different nodes.

We take the M/G/1 non-preemptive queueing system as the model to obtain the waiting time distribution (for the high priority RT packets) in each node and then apply convolution to get the end-to-end waiting time distribution. If we let W_k^H denote the waiting time in the k'th node for the RT-packets, then the total delay may be written $W_{NP}^H = W_1^H + ... + W_K^H$, and the LST of the sum is found as the product of the LSTs of the waiting times in the individual nodes:

$$\tilde{W}_{NP}^H(s) = \prod_{k=1}^{K} \tilde{W}_k(s) \tag{8.1}$$

Where $W_k(s)$ is the LST of the waiting time for the highest priority packets in an M/G/1 non-preemptive queueing model, and is given by the *Pollaczek-Khinchin* formula (with a slight modification). (See for instance [Taka91] or [Klei76b]):

$$\tilde{W}_k(s) = \frac{1-\rho_k^H}{1-\rho_k^H \hat{B}_k^H(s)}(1-p_k + p_k \hat{B}_k^L(s)) \tag{8.2}$$

where ρ_k^H and ρ_k^L are the loads and $\hat{B}_k^H(s)$ and $\hat{B}_k^L(s)$ are the LSTs of the remaining service times for high and low priority packets respectively and further $p_k = \frac{\rho_k^L}{1-\rho_k^H}$. (The relation between the LST of the remaining service times and the "ordinary" service times is given as $\hat{B}(s) = \frac{1-\tilde{B}(s)}{sb}$ where $b = E[B]$ is the mean service time.)

It may be convenient to relate the end-to-end waiting time distribution based on the LST (8.1) to the corresponding model without any priority. We may write $W_{NP}^H = W^T + B_{NP}^T$ where W^T and B_{NP}^T are independent and W^T represents the end-to-end queueing delay for the corresponding path without any priority and B_{NP}^T is the extra delay due to the influence from the lower priority packets. Consequently, the distribution of W_{NP}^H may therefore be found by convoluting the distributions of W^T and B_{NP}^T. We may therefore re-write the LST:

$$\tilde{W}_{NP}^H(s) = \tilde{W}^T(s)\tilde{B}_{NP}^T(s) \text{ with} \tag{8.3}$$

$$\tilde{W}^T(s) = \prod_{k=1}^{K} \frac{1-\rho_k^H}{1-\rho_k^H \hat{B}_k^H(s)} \text{ and} \tag{8.4}$$

$$\tilde{B}_{NP}^T(s) = \prod_{k=1}^{K}(1-p_k + p_k \hat{B}_k^L(s)) \tag{8.5}$$

The expressions above will also apply for saturated system. In this case there will always be low priority packets present in the low priority queue and this corresponds to the case with $p_k = 1$ (or $\rho_k^H + \rho_k^L = 1$).

The rest of this section will be devoted to find the DF of the end-to-end waiting time $W_{NP}^H(t) = P\{W_{NP}^H \le t\}$, based on the LST (8.1) and (8.2) or (8.3), (8.4) and (8.5). Especially we are interested in the tail of the CDF of end-to-end delay to get the desired quantiles.

As for the case without priority queueing (in chapter 7) it is possible to obtain the distributions above by inverting the transform numerically. For instance the DF of the end-to-end queueing times may be written by the inversion integral as:

$$W_{NP}^{H}(t) = P\{W_{NP}^{H} \le t\} = \frac{1}{2\pi i}\int_{\gamma} \frac{e^{st}}{s}\tilde{W}_{NP}^{H}(s)ds \tag{8.6}$$

where the integration line is parallel with the imaginary axis $\gamma = \{s | s = a + iy\}$ and where $a > 0$ is a constant and $y \in (-\infty, \infty)$. Some problems with such inversion are mentioned in chapter 7.

8.2 Exact results when all the nodes are identical

For the case where all the nodes are identical it is possible to carry the analysis significantly further without introducing any approximations by applying similar approach as used in chapter 7. This is due to the fact that it is possible to obtain the LST of the convolution through partial derivatives of the load for the LST of the waiting times in a single M/G/1 queue.

By (7.14) we have

$$W^{T}(t,\rho) = \frac{(1-\rho)^{K}}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{K-1}}{1-\rho}W(t,\rho)\right\} \tag{8.7}$$

$W(t,\rho)$ denotes the DF of the waiting time in an M/G/1 queue with load ρ and with LST

$$\tilde{W}(s,\rho) = \frac{1-\rho}{1-\rho\hat{B}^{H}(s)}. \tag{8.8}$$

The DF of the end-to-end queueing delay can therefore be written as the sum obtained by inverting (8.3)-(8.5):

$$W_{NP}^{H}(t) = (1-p)^{K}W^{T}(t,\rho^{H}) + \sum_{r=1}^{K} b_{r}(p,K)W^{T}(t,\rho^{H})(*)\hat{b}^{L}(t)^{*(r)} \tag{8.9}$$

where $b_{r}(p,K) = \binom{K}{r}p^{r}(1-p)^{K-r}$ is the binomial probabilities with parameters $p = \frac{\rho^{L}}{1-\rho^{H}}$ and K and $\hat{b}^{L}(t)^{*(r)}$ is the r-times convolution of the PDF of the remaining service times for the low priority packets ($(*)$ denotes convolution). The expression (8.9) represents a general expression without any specific assumptions on the actual service time distributions. To carry the analysis any further specific choices on the service time distributions therefore have to be made.

In the following we shall assume that both the high priority packets have constant service times given by b^{H}. In this case we have the DF on the form:

$$W(t,\rho^H) = q(\frac{t}{b^H},\rho^H) \tag{8.10}$$

where $q(x,\rho)$, given by (7.27), is the DF of the waiting time in an M/D/1 queue with service times scaled to unity. The K-fold convolution of $W(t,\rho^H)$ is found in section 7.3.2:

$$W^T(t,\rho^H) = q^K(\frac{t}{b^H},\rho^H) \tag{8.11}$$

where $q^K(x,\rho)$ is given by (7.28).

8.2.1 Deterministic service times for low priority packets

In the following we shall assume that the low priority packets have constant service times given by b^L. It follows that the remaining service times for the low priority packets are uniformly distributed over the interval $(0,b^L)$, and we find the r-time convolution $\hat{b}^L(t)^{*(r)}$ on the following form:

$$\hat{b}^L(t)^{*(r)} = \frac{r}{(b^L)^r}\sum_{m=0}^{r}\frac{(-1)^m}{m!(r-m)!}H(t-mb^L)(t-mb^L)^{r-1} \tag{8.12}$$

where $H(x)$ is the unit step function. We find that the convolution $W^T(t,\rho^H)(*)\hat{b}^L(t)^{*(r)}$ may be written as:

$$W^T(t,\rho^H)(*)\hat{b}^L(t)^{*(r)} = r\left(\frac{b^H}{b^L}\right)^r\sum_{m=0}^{r}\frac{(-1)^m}{m!(r-m)!}H(t-mb^L)I^{K,r-1}(\frac{t-mb^L}{b^H},\rho^H)$$

where $I^{K,i}(t,\rho)$ are the following integrals:

$I^{K,i}(t,\rho) = \int_{x=0}^{t}(t-x)^i q^K(x,\rho)dx$ for $i = 0,1,...,K-1$. These integrals may be evaluated in terms of some auxiliary functions defined in Appendix F by equations (F.33), (F.34) and (F.35). Collecting the different terms we finally find the following expression for the DF of the end-to-end queueing delay:

$$W_{NP}^H(t) = (1-p)^K q^K(\frac{t}{b^H},\rho^H) +$$

$$\sum_{r=1}^{K} b_r(p,K)\left(\frac{b^H}{b^L}\right)^r\sum_{k=0}^{\left\lfloor\frac{t}{b^H}\right\rfloor}\sum_{m=0}^{\left\lfloor\frac{t-kb^H}{b^L}\right\rfloor}(-1)^m\binom{r}{m}\binom{k+r-1}{r-1}q^{K,r-1}(\frac{t-kb^H-mb^L}{b^H},\rho^H) \tag{8.13}$$

where the auxiliary functions $q^{K,i}(x,\rho)$; $i = 1,2...,K-2$ are found from (F.34) and (F.35):

$$q^{K,i}(x,\rho)=\frac{(1-\rho)^K}{\rho^{i+1}}\sum_{k=0}^{\lfloor x\rfloor}\sum_{l=0}^{K-1}\frac{(-1)^l}{l!k!}\binom{K+k-i-2}{K-l-1}\big(\rho(k-x)\big)^{k+l}e^{-\rho(k-x)} \tag{8.14}$$

For $i=K-1$ we must add an extra term to get $q^{K,K-1}(x,\rho)$:

$$q^{K,K-1}(x,\rho)=\frac{(1-\rho)^K}{\rho^K}\left(\sum_{k=0}^{\lfloor x\rfloor}\sum_{l=0}^{K-1}\frac{(-1)^l}{l!k!}\binom{k-1}{K-l-1}\big(\rho(k-x)\big)^{k+l}e^{-\rho(k-x)}+(-1)^K\right) \tag{8.15}$$

(In the expression above we define the binomial coefficient as $\binom{n}{m}=\frac{n(n-1)...(n-m+1)}{m!}$ also allowing for negative n and implying $\binom{n}{m}=0$ for $m>n$.) The expression (8.13) for the DF of the end-to-end queueing delay gives stable numerical results for at least up to $K=20$ identical nodes. The numerical accuracy depends heavily on the fact that the auxiliary functions $q^{K,i}(x,\rho)$ may be calculated by introducing "local" variables (see Appendix F section F.2) and thereby avoiding summation of alternating series.

Substantial simplification yields for special choices of the parameters. We shall mention these cases below:

A. *The service times for the low priority packets are exactly an integer times the service times of the high priority packets, that is* $b^L=lb^H$ *with integer* l.

In this case we can simplify the summation giving:

$$W_{NP}^H(t)=(1-p)^K q^K\left(\frac{t}{b^H},\rho^H\right)+\sum_{r=1}^{K}b_r(p,K)\sum_{k=0}^{\left\lfloor\frac{t}{b^H}\right\rfloor}c_r(k,l)q^{K,r-1}\left(\frac{t}{b^H}-k,\rho^H\right) \tag{8.16}$$

where $$c_r(k,l)=l^{-r}\sum_{m=0}^{\min\left\{r,\left\lfloor\frac{k}{T}\right\rfloor\right\}}(-1)^m\binom{r}{m}\binom{r+k-lm-1}{r-1} \tag{8.17}$$

B. *The service times for low and high priority packets are equal, that is* $b^L=b^H$ *(*$l=1$ *in the case above).*

In this case we have $c_r(0,1)=1$ and $c_r(k,1)=0$ for $k=1,2,..$ giving:

$$W_{NP}^H(t)=(1-p)^K q^K\left(\frac{t}{b^H},\rho^H\right)+\sum_{r=1}^{K}b_r(p,K)q^{K,r-1}\left(\frac{t}{b^H},\rho^H\right) \tag{8.18}$$

C. *The queueing system is saturated*, that is $\rho^H+\rho^L=1$ implying $p=1$ in the expression above giving:

$$W_{NP}^{H}(t)=\left(\frac{b^H}{b^L}\right)^K\sum_{k=0}^{\left\lfloor\frac{t}{b^H}\right\rfloor}\sum_{m=0}^{\left\lfloor\frac{t-kb^H}{b^L}\right\rfloor}(-1)^m\binom{K}{m}\binom{k+K-1}{K-1}q^{K,K-1}\left(\frac{t-kb^H-mb^L}{b^H},\rho^H\right) \tag{8.19}$$

D. *Saturated system and the fraction between high and low priority service times are an integer*.

We find:

$$W_{NP}^{H}(t)=\sum_{k=0}^{\left\lfloor\frac{t}{b^H}\right\rfloor}c_K(k,l)q^{K,K-1}\left(\frac{t}{b^H}-k,\rho^H\right) \tag{8.20}$$

E. *Saturated system and equal service times for low and high priority packets*.

We get:

$$W_{NP}^{H}(t)=q^{K,K-1}\left(\frac{t}{b^H},\rho^H\right) \tag{8.21}$$

8.2.2 Exponentially distributed service times for low priority packets

In this case we have that the remaining service times also are negative exponentially distributed and $\hat{b}^L(t)=\mu^L e^{-\mu^L t}$ where $b^L=1/\mu^L$ is the mean service times. Further it follows that the r-times convolution of the PDF of the remaining service times for the low priority packets $\hat{b}^L(t)^{*(r)}$ is Erlang-r distributed given as:

$$\hat{b}^L(t)^{*(r)}=\frac{\mu^L\left(\mu^L t\right)^{r-1}}{(r-1)!}e^{-\mu^L t} \tag{8.22}$$

The convolution of the DF of the K-folded waiting time for an M/D/1 queue (all with service times scaled to unity) with the PDF of an Erlang-r distributed variable with parameter μ are given in Appendix F by equations (F.22). Applying these results we may write the DF of the end-to-end queueing delay as:

$$W_{NP}^{H}(t)=(1-p)^K q^K\left(\frac{t}{b^H},\rho^H\right)+\sum_{r=1}^{K}b_r(p,K)F_{K,r}\left(\frac{t}{b^H},\mu^L b^H,\rho^H\right) \tag{8.23}$$

where $F_{k,r}(t,\mu,\rho)$ is given by (F.22).

Special case:

A. *The queueing system is saturated*, that is $\rho^H+\rho^L=1$ implying $p=1$ in the expression above giving

$$W_{NP}^{H}(t) = F_{K,K}(\frac{t}{b^H}, \mu^L b^H, \rho^H) \tag{8.24}$$

8.3 Approximative methods

As for the case without priority it is possible to obtain approximations by assuming that the corresponding stochastic variable converges to normal distributions. Since the normal distribution is characterized by its two first (lowest) moments, this leads to the following approximation for the PDF and CDF of the convolution:

$$w_{NP}^{HNA}(t) = \frac{1}{\sqrt{K}\sigma}\varphi(\frac{t-Km}{\sqrt{K}\sigma}) \text{ and} \tag{8.25}$$

$$W_{NP}^{HNA}(t) = 1-\phi(\frac{t-Km}{\sqrt{K}\sigma}) \text{ where} \tag{8.26}$$

$m = m^H + m^L$ and $\sigma^2 = \sigma^{H2} + \sigma^{L2}$. Further $m^H = E[W^H]$ and $\sigma^{H2} = E[W^{H2}] - m^{H2}$ are the mean and variance of the queueing delay in the corresponding single server M/G/1 queue with only high priority traffic present, given in terms of the three first moments of the service time distribution (and also the load) through:

$$m^H = E[W^H] = \frac{E[B^{H2}]}{2E[B^H]}\frac{\rho^H}{1-\rho^H} \text{ and } \sigma^{H2} = \frac{E[B^{H3}]}{3E[B^H]}\frac{\rho^H}{1-\rho^H} + m^{H2} \tag{8.27}$$

The influence from the low priority traffic is given through the remaining service times for a low priority packet by:

$$m^L = p\frac{E[B^{L2}]}{2E[B^L]} \text{ and } \sigma^{L2} = p\frac{E[B^{L3}]}{3E[B^L]} - m^{L2} \tag{8.28}$$

(Further $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ is the standard normal density and $\varphi(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt$ is the standard normal integral.)

The second and quite different approximation (though also involving normal distributions) is based on LD (Large Deviation) theory. This can for instance be reflected through the inversion of the LST of the end-to-end delay given by the inversion integral

$$w_{NP}^{H}(t) = \frac{1}{2\pi i}\int_\gamma e^{st}[\tilde{W}(s,\rho)(1-p-p\hat{B}^L(s))]^K ds \tag{8.29}$$

As for the case without any priority (in chapter 7) we may find asymptotics by the Saddle Point method. Formally this approximation is given as the asymptotic behaviour of

$w_{NP}^{H}(Kx,\rho)$ for large values of K and fixed x. With this substitution the inversion integral may be written

$$w_{NP}^{H}(Kx,\rho)=\frac{1}{2\pi i}\int_{\gamma}e^{-Kg(s,x)}ds \quad \text{with} \tag{8.30}$$

$$g(s,x)=\log(1-\rho^{H}\hat{B}^{H}(s))-\log(1-p+p\hat{B}^{L}(s))-sx-\log(1-\rho^{H}) \tag{8.31}$$

We get $w_{NP}^{HLD}(t)=f(\frac{t}{K})$; where

$$f(x)=\frac{1}{\sqrt{2\pi Kh(x)}}e^{-Kg(s^{*}(x),x)} \tag{8.32}$$

and where $s=s^{*}(x)$ is the solution of the equation

$$\frac{\rho^{H}\hat{B}^{H'}(s)}{1-\rho^{H}\hat{B}^{H}(s)}+\frac{p\hat{B}^{L'}(s)}{1-p+p\hat{B}^{L}(s)}+x=0 \tag{8.33}$$

and

$$h(x)=-g''(s,x)=\left(\frac{\rho^{H}\hat{B}^{H'}(s)}{1-\rho^{H}\hat{B}^{H}(s)}\right)^{2}-\left(\frac{p\hat{B}^{L'}(s)}{1-p+p\hat{B}^{L}(s)}\right)^{2}+\frac{\rho^{H}\hat{B}^{H''}(s)}{1-\rho^{H}\hat{B}^{H}(s)}+\frac{p\hat{B}^{L''}(s)}{1-p+p\hat{B}^{L}(s)} \tag{8.34}$$

The corresponding expression for the DF as an inversion integral is:

$$W_{NP}^{H}(Kx) = \frac{1}{2\pi i}\int_{\gamma}\frac{e^{-Kg(s,x)}}{s}ds \tag{8.35}$$

If we apply the Saddle Point method directly on (8.35) we must make distinction between two cases depending on the sign of saddle point $s^{*}(x)$. It turns out that $s^{*}(x)$ is positive for small values of x but becomes negative as x increases. In the latter case we have to pick up the residue at $s=0$ and we obtain the following approximation by combining the two cases:

$$W_{NP}^{H\ LD}(t)=1-F(\frac{t}{K}) \quad \text{where} \tag{8.36}$$

$$F(x)=H(s^{*}(x))-\frac{1}{s^{*}(x)\sqrt{2\pi Kh(x)}}e^{-Kg(s^{*}(x),x)} \tag{8.37}$$

where $H(t)$ is the unit step function.

It turns out that the "ordinary" Saddle Point method given by (8.36) and (8.37) fails to approximate the integral (8.35) when the root $s^*(x)$ is close to the pole at $s=0$. It is, however, possible to extract that pole and obtain a Uniform Asymptotic Approximation (UAA) that is uniform with respect to x and also yields in the area where $s^*(x)$ is close to zero. We find the following approximation:

$$W_{NP}^{H\ UAA}(t) = 1 - G(\frac{t}{K}) \text{ where} \tag{8.38}$$

$$G(x) = \phi(-\mathrm{sgn}(s^*(x))\sqrt{2Kg(s^*(x),x)}) - e^{-Kg(s^*(x),x)}\left\{\frac{1}{s^*(x)\sqrt{2\pi Kh(x)}} - \frac{\mathrm{sgn}(s^*(x))}{2\sqrt{\pi Kg(s^*(x),x)}}\right\} \tag{8.39}$$

It is quite easy to see that the function defined by $G(x)$ does not have any singularities for values of x giving saddle point close to zero. By expanding the brackets to third order in $s^*(x)$ we find:

$$G(x) = \phi(-\mathrm{sgn}(s^*(x))\sqrt{2Kg(s^*(x),x)}) + \frac{h_3(x)}{6h(x)}\frac{e^{-Kg(s^*(x),x)}}{\sqrt{2\pi Kh(x)}} \tag{8.40}$$

when $s^*(x)$ is close to zero and where $h_3(x) = -g'''(s^*(x),x)$. The accuracy of the UAA seems to be good and gives numerical values very close to the exact convolutions for a broad range of parameters also including cases where the asymptotic is not fulfilled, i.e. for chains of relatively small sizes. The "classical" LD approximation, however, gives strict upper bound on the convolution and can therefore be a desirable method to apply for this reason.

8.4 Examples

In the following we shall give some numerical examples by applying the models described in the previous section where we focus on some typical scenarios. We assume a network with two priority classes and HOL scheduling and we focus on the end-to-end delay for the high priority traffic. The high priority class will typically be real time traffic like voice and video that will have constraints on the maximum end-to-end delay. Under the assumption that the load from the high priority traffic is limited we would like to find out the effect the low priority traffic will have on the performance of the high priority real time classes. This is a typical situation in IP-networks deploying DiffServ. In an IP-network the end-to-end delay for a particular stream will be the sum of the delay obtained in a cascade of routers (from the sender to the receiver). The total end-to-end delay will then consist of the waiting times in each node plus service times (transmission times onto the links). In the examples below we shall consider a particular chain of routers in a packet network and we as-

sume that the routers have output buffers with two priority classes with no extra internal delay due to processing of the packets.

To apply the results in section 8.2 we must assume that all the nodes in the chain have identical parameters:

- The link capacity is equal for all the routers.
- Packets for the two priority classes arrive according to Poisson processes with parameters that are equal in each router.
- The packet lengths for the high priority class is constant with mean P_H . (All the numerical results are scaled according to the transmission time for a high priority packet.)
- The packet lengths for the low priority class are either constant or exponentially distributed with mean P_L .
- The load from the high priority traffic class is ρ_H and we shall assume that routers are saturated, this means that there will always be low priority packets to be transmitted, implying that the low priority load $\rho_L \geq 1 - \rho_H$.

By the last assumption we may use the somewhat simplified formula given by the equations (8.19) and (8.24) in section 8.2 to obtain the DF of the end-to-end delay distribution. With these definitions above we find mean service times for packets $b_H = P_H / C$, $b_L = \mu_L^{-1} = P_L / C$ where C is the link capacity. In the examples below we have chosen scenarios among the following parameter values:

- The ratio between low and high priority packets P_L / P_H is either 1,5 or 10.
- The load from high priority traffic ρ_H is either 0.4 or 0.6 .
- The number of hops K is either 5, 10 or 15 .

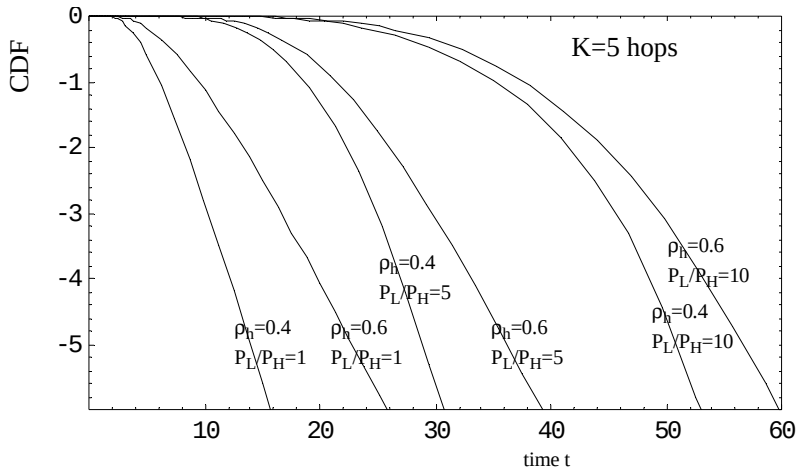

Figure 8.1: Logarithmic plot of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 5 and the ratio between low and high priority packet lengths is 1, 5 and 10 and the low priority packets are constant.

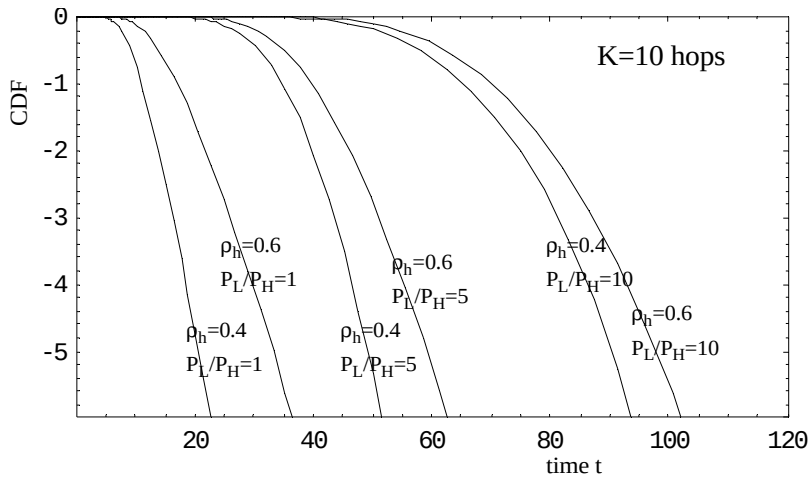

Figure 8.2: Logarithmic plot of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 10 and the ratio between low and high priority packet lengths is 1, 5 and 10 and the low priority packets are constant.

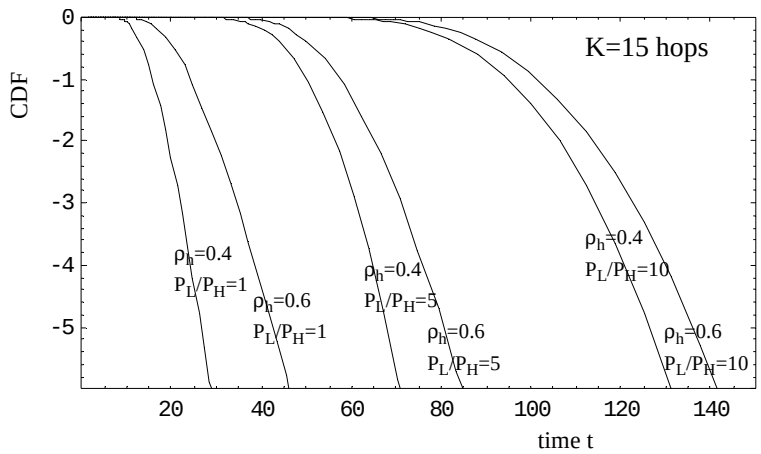

Figure 8.3: Logarithmic plot of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 15 and the ratio between low and high priority packet lengths is 1, 5 and 10 and the low priority packets are constant.

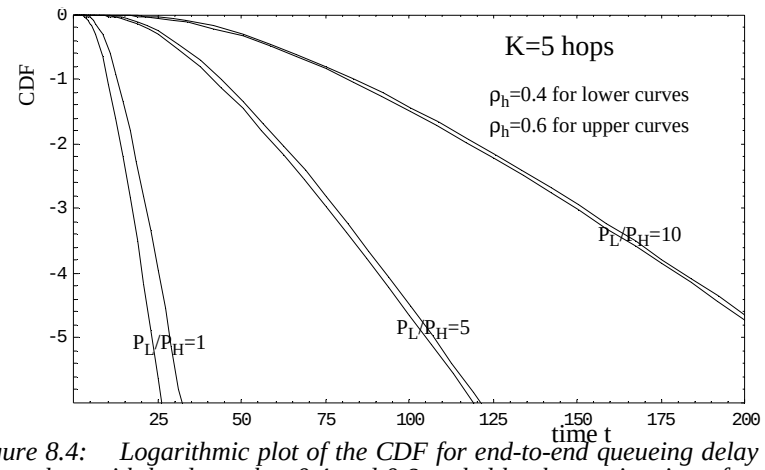

Figure 8.4: Logarithmic plot of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 5 and the ratio between low and high priority packet lengths is 1, 5 and 10 and the low priority packets are exponentially distributed.

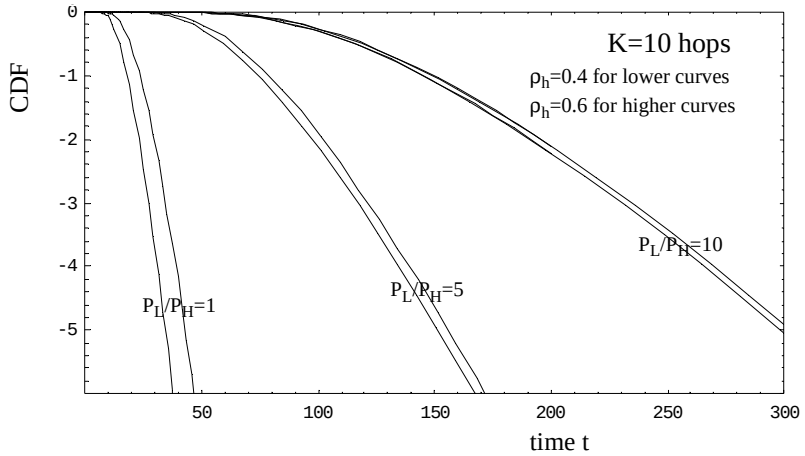

Figure 8.5: Logarithmic plot of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 10 and the ratio between low and high priority packet lengths is 1, 5 and 10 and the low priority packets are exponentially distributed.

In figure 8.1-figure 8.3 we have depicted some results for the case when the low priority packet lengths are assumed to be constant. This case could represent the case when we have the packet length more or less limited by an Ethernet frame of 1500 bytes and in addition by assuming rather short real time packet lengths of around 200 bytes. We observe rather strong impact from the ratio of the low and high priority packet lengths. The influence of the load from the high priority traffic is not that strong and this is more or less expected since we assume that the high priority load is limited to say less than 60%, which seems to be reasonable keeping in mind the need to reserve some part of the capacity also for low priority traffic. If we for instance take an example with STM-1 links of approximately 150 Mbit/s and assume that the real time packet lengths are 200 bytes, this will give packet transmission time of around 10 μsec. By assuming a path of 15 hops and assuming packet length ratio of 10, then figure 8.3 provide us with the appropriate quantile. If we take the $1-10^{-3}$ quantile for the highest load we find the appropriate value to be around 125 (high priority packet transmission time), and this leaves us with a value of 1.25 ms for this particular case. This tells us that if a core network deploying DiffServ is properly engineered so that the high priority load is limited to say 60% then the end-to-end queueing delay will be limited to a few milliseconds. (One has to add the contributions from the access part of the particular path to get the complete picture, and this contribution could be larger due to slower links in the access network.)

In figure 8.4 and figure 8.5 we have given corresponding results for the case with exponentially distributed packet lengths for low priority packets. We see that the exponential distribution of the packet lengths gives considerably worse performance. This is due to the tail in the exponential distribution compared with a rectangularly distributed variable. In this case we also observe a very small influence of the high priority load (as long as it is well limit-

ed to say 60%), and we conclude that the actual performance is determined by the number of hops and the ratio between the mean packet lengths between low and high priority packets. The corresponding $1\text{-}10^{-3}$ quantile for 10 hops is found to be approximately 250, which is twice that of the same quantile with 15 hops and constant packet lengths.

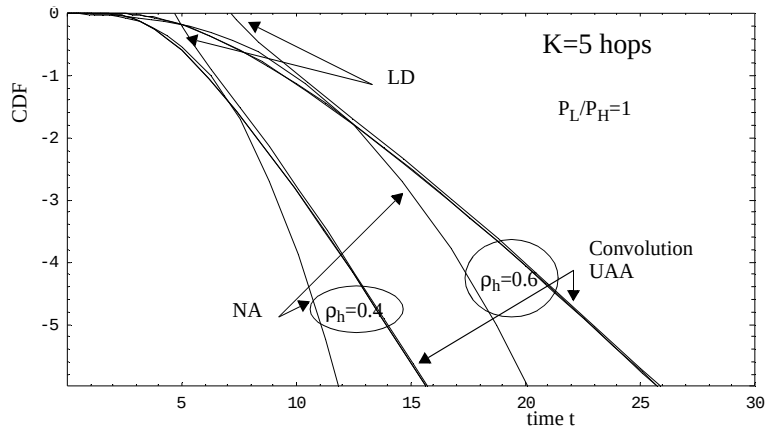

Figure 8.6: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 5 and the ratio between low and high priority packet lengths is 1 and the low priority packets are constant. (NA Normal Approximation, LD- Large Deviation, UAA- Uniform Asymptotic Approximation,

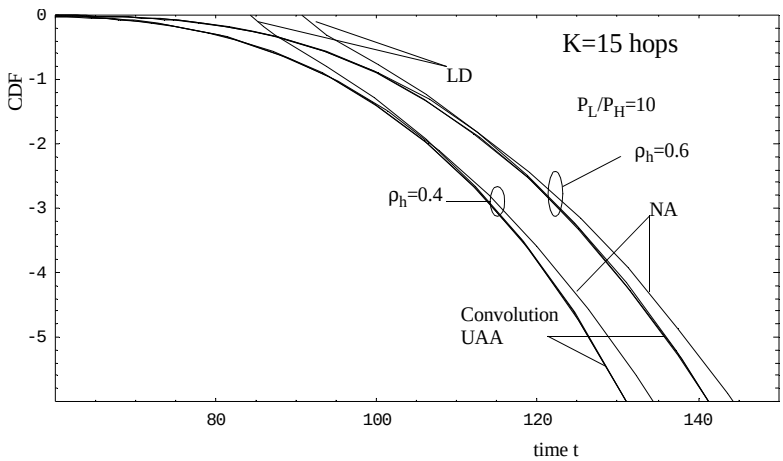

Figure 8.7: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 and 0.6 scaled by the service times for a high priority packet. The number of hops is 15 and the ratio between low and high priority packet lengths is 10 and the low priority packets are constant. (NA-Normal Approximation, LD- Large Deviation, UAA- Uniform Asymptotic Approximation)

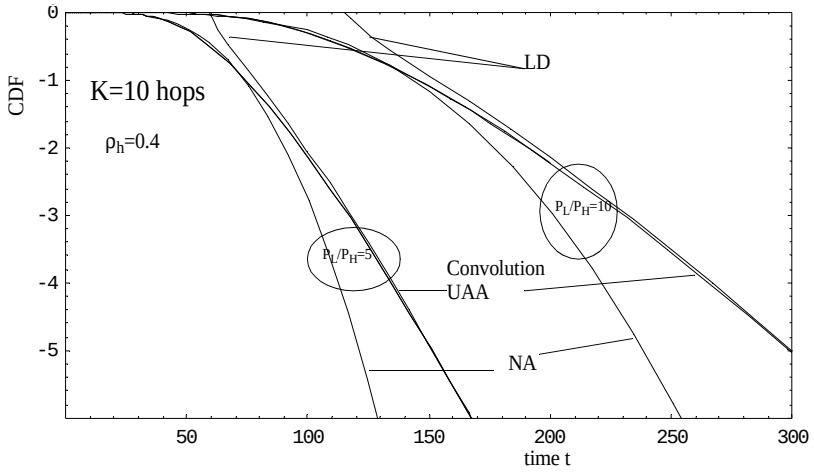

Figure 8.8: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.4 scaled by the service times for a high priority packet. The number of hops is 10 and the ratio between low and high priority packet lengths is 5 and 10 and the low priority packets are exponentially distributed. (NA-Normal Approximation, LD- Large Deviation, UAA- Uniform Asymptotic Approximation)

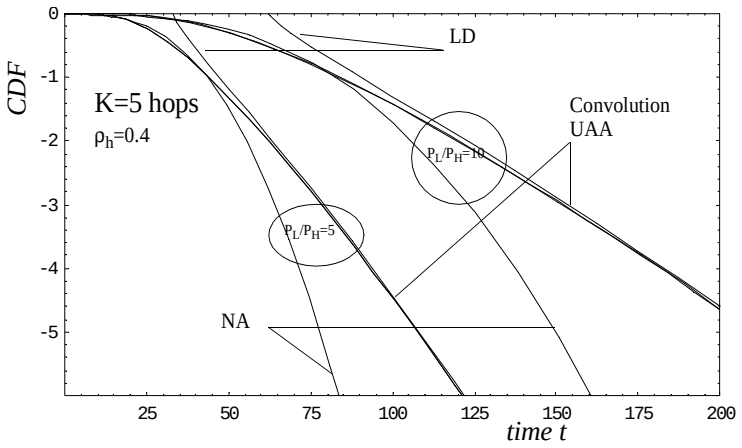

Figure 8.9: Logarithmic plot of different approximations of the CDF for end-to-end queueing delay for high priority packets with load equal to 0.6 scaled by the service times for a high priority packet. The number of hops is 5 and the ratio between low and high priority packet lengths is 5 and 10 and the low priority packets are exponentially distributed. (NA-Normal Approximation, LD- Large Deviation, UAA- Uniform Asymptotic Approximation)

In figure 8.6-figure 8.9 we have tested the different approximation described in section 8.3 by applying some different values of the parameters. We may draw the following conclusions:

- The NA does not behave well in the tail of the distribution. And often it will underestimate the actual probabilities. On the other hand it is very simple and easy to apply.
- The LD approach gives quite reasonable results for a broad range of the distribution, especially in the tail. In addition it also provides us with an upper bound. To apply this method one has to locate the saddle point for each value one would calculate the distribution function of.
- The UAA is an excellent approximation and gives a uniform approximation of the distribution over the whole range of the distribution function. The relative error is very small in all the cases we have considered (less than 3%) and it is nearly impossible to make the distinction in the graphs. It also seems to give accurate estimates for values where the asymptotic is not fulfilled, i.e. for chains consisting of only one or two queues.

8.5 Concluding remarks

In this chapter we have discussed and given some new methods to calculate the end-to-end queueing delay in a packet network where real time traffic has strict priority over other classes of traffic. The described method could for instance be applied to estimate typical end-to-end delay in a core network deploying DiffServ. The proposed methods are tested against known approximation such as the saddle point method. Especially the UAA (Uniform Asymptotic Approximation) gives very accurate results. Compared with the exact methods proposed in this chapter the UAA requires that the corresponding saddle points have to be located for each single value under consideration.

We have also demonstrated by the numerical examples that by deploying DiffServ in a core network with STM-1 links or links with higher bit-rate and by limiting the load from the real time traffic to less than 60% it is possible to guarantee the corresponding end-to-end queueing delay to just a few milliseconds with very high probability (e.g. $1\text{-}10^{-3}$ quantile).

9

Discrete time queueing models

9.1 Introduction

In the previous chapters we have focused on models to analyse the traffic dependent end-to-end waiting times for a path in a network deploying statistical multiplexing. The aim was to obtain models with limited complexity so that they could be used for dimensioning purposes. The "critical assumption" for those models is of cause the *independent assumption* needed to obtain the corresponding Laplace transform on product form, and we did not try to find the actual condition for which this approximation is applicable.

One approach would be to analyse a particular traffic stream (flow) as it traverses a multiplexer and try to capture the characteristics of that particular traffic process (flow) at the output. This particular output process will then be mingled with other traffic streams and will constitute the input to the next multiplexer in the chain under consideration. By this approach we are able to trace a particular stream (flow) describing the distortion as it passes through a particular path through the network. Similar approach to study end-to-end behaviour is well documented in the literature. (See [Matr94a], [Matr94b].)

It turns out that the discrete time queueing model is easier to analyse than the corresponding continuous time counterpart, and this is mainly due the discrete nature of the corresponding models; where the corresponding analyse tool will be based on generating function techniques rather than Laplace transforms (often applied for continuous time models). Nevertheless, it is well known that slotted queueing models could be regarded as approximations of continuous time models and in this perspective the discrete time models will be of interest to analyse. In the following we shall consider a discrete (slotted) queueing model where we will put the main emphasis on the possibility of tracing a particular traffic stream as it passes through a multiplexer where it will be disturbed by crossing packet streams (background traffic). We are particularly interested in describing the output process of that particular stream which then will be part of the input traffic to the next multiplexer.

9.2 A discrete time queueing model with a renewal foreground and a batch background stream as input

The queueing model taken as basis of the analysis is depicted in figure 9.1 below. It is a single server, infinite capacity queue operating in discrete (or slotted) time with two classes of customers. Any activity in the system, e.g. arrivals, departures, etc., is assumed to occur at the slot boundaries.

The arrival process is formed by superposing of a discrete time renewal process, *foreground stream* **(FS)**, and a (discrete) batch arrival process (e.g. Poisson or Bernoulli process), *background stream* **(BS)**. We let the slots be successively numbered $k = 0, 1, \ldots$ and we assume that the batch size in slot k, generated by the **BS**, B_k is independent and follows a general (discrete) distribution $b(i) = P(B_k = i)$ with generating function $B(z)$. The **FS** renewal process is characterized by the distribution of the numbers of slots between arrivals $A_n = T_{n+1} - T_n$ where T_n is the slot number for the n'th arrival of the **FS**; $n = 0, 1, \ldots$, and we assume that A_n is independent (of n and of **BS**) and follows a general (discrete) distribution $a(i) = P(A_n = i)$ with generating function $A(z)$.

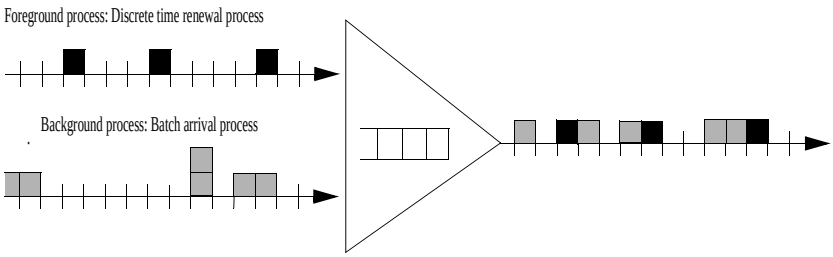

Figure 9.1: The queueing model for the packet multiplexers.

The total load on the multiplexer is $\rho = \rho_{FS} + \rho_{BS}$ where $\rho_{FS} = \dfrac{1}{A'(1)}$ is the load from **FS,** and $\rho_{BS} = B'(1)$ is the load from **BS,** and we shall assume that $\rho < 1$ to secure stability for the queueing system.

We observe the queue size at the end of each slot, and we define the following stochastic variables: Q_n-the number of packets in the queue at the end of the slot just prior to the n'th arrival of a packet from the **FS,** and conditioning on $A_n = T_{n+1} - T_n = k$ we let Q_n^i-the numbers of packets in the queue in the end of the slot $T_n + i$ $(i = 1, \ldots, k)$. If we let B_n^i denote the numbers of packets arriving from **BS** during slot $T_n + i$, then we have the following relation between the queue lengths in the different slots:

$$Q_n^1 = Q_n + B_n^1 \tag{9.1}$$

$$Q_n^i = [Q_n^{i-1} + B_n^i - 1]^+ \quad \text{for } i = 2, 3, \ldots, k \tag{9.2}$$

We obviously also have:

$$Q_{n+1} = Q_n^k \tag{9.3}$$

9.2.1 Transient queueing analysis

The equations (9.1)-(9.3) describe the evolution of the queue length when it is combined with the arrival instants T_n of the **FS**. In the following we shall define a joint generating function taking both the evolution of the queue content and the arrival of the **FS** into account by defining:

$$Q_n(z, x) = \boldsymbol{E}[z^{Q_n} x^{T_n}] \text{ and} \tag{9.4}$$

$$Q_n^i(z, x) = \boldsymbol{E}\left[z^{Q_n^i} x^{T_n} \middle| T_{n+1} - T_n = k\right] \text{ for } i = 2, 3, \ldots, k \tag{9.5}$$

By the relations (9.1) and (9.2) we find the following recursions:

$$Q_n^1(z, x) = Q_n(z, x) B(z) \text{ and} \tag{9.6}$$

$$Q_n^i(z, x) = Q_n^{i-1}(z, x)\frac{B(z)}{z} + \left(1 - \frac{1}{z}\right) q_n^{i-1}(x) \text{ for } i = 2, 3, \ldots, k \text{ where} \tag{9.7}$$

$q_n^i(x) = \boldsymbol{E}\left[x^{T_n} \mathbf{1}_{\{Q_n^i + B_n^{i+1} = 0\}} \middle| T_{n+1} - T_n = k\right]$ is the boundary transform, taking into account the probability of having an empty queue in slot $T_n + i + 1$ ($i = 1, \ldots, k-1$). Solving (9.7) recursively we obtain $Q_n^k(z, x)$ as function of $Q_n^1(z, x)$:

$$Q_n^k(z, x) = Q_n^1(z, x)\left(\frac{B(z)}{z}\right)^{k-1} + \left(1 - \frac{1}{z}\right) \sum_{l=1}^{k-1} q_n^l(x)\left(\frac{B(z)}{z}\right)^{k-l-1} \text{ for } k = 1, 2, \ldots \tag{9.8}$$

For $n = 0$ we shall make the conditions that $T_0 = 0$ and $Q_0^1 = m$ implying that:

$$Q_0^k(z, x) = z^m\left(\frac{B(z)}{z}\right)^{k-1} + \left(1 - \frac{1}{z}\right) \sum_{l=1}^{k-1} q_0^l(x)\left(\frac{B(z)}{z}\right)^{k-l-1} \text{ and by (9.6)} \tag{9.9}$$

$$Q_n^k(z, x) = zQ_n(z, x)\left(\frac{B(z)}{z}\right)^{k} + \left(1 - \frac{1}{z}\right) \sum_{l=1}^{k-1} q_n^l(x)\left(\frac{B(z)}{z}\right)^{k-l-1} \text{ for } n = 1, 2, \ldots. \tag{9.10}$$

Since we have $\boldsymbol{E}[z^{Q_{n+1}}x^{T_{n+1}}\big|T_{n+1}-T_n=k] = x^k\boldsymbol{E}\Big[z^{Q_n^k}x^{T_n}\big|T_{n+1}-T_n=k\Big] = x^kQ_n^k(z,x)$ implying

$$Q_{n+1}(z,x) = \sum_{k=1}^{\infty} a(k)x^kQ_n^k(z,x) \tag{9.11}$$

By applying (9.11) on (9.9) and (9.10) we find:

$$Q_1(z,x) = z^mA\Big(x\frac{B(z)}{z}\Big)\frac{z}{B(z)} + \Big(1-\frac{1}{z}\Big)\sum_{k=0}^{\infty}\tilde{q}_0^k(x)\Big(\frac{B(z)}{z}\Big)^k \text{ and} \tag{9.12}$$

$$Q_{n+1}(z,x) = zQ_n(z,x)A\Big(x\frac{B(z)}{z}\Big) + \Big(1-\frac{1}{z}\Big)\sum_{k=0}^{\infty}\tilde{q}_n^k(x)\Big(\frac{B(z)}{z}\Big)^k \text{ for } n = 1,2\ldots \tag{9.13}$$

where we also have defined the quantities $\tilde{q}_n^k(x) = \sum_{l=1}^{\infty} a(k+l+1)x^{k+l+1}q_n^l(x)$. To combine the equations (9.12) and (9.13) we introduce generating functions

$$Q^m(z,x,s) = \sum_{n=1}^{\infty} s^{n-1}Q_n(z,x) \text{ and } \tilde{q}^k(x,s) = \sum_{n=1}^{\infty} s^n\tilde{q}_n^k(x) \tag{9.14}$$

(where we have indicated that we have the condition $Q_0^1 = m$). By multiplying (9.13) by s^n and summing and combining with (9.12) we may solve for $Q(z,x,s)$ by the fraction:

$$Q^m(z,x,s) = \frac{z^mA\Big(x\frac{B(z)}{z}\Big)\frac{z}{B(z)} + \Big(1-\frac{1}{z}\Big)\sum_{k=0}\tilde{q}^k(x,s)\Big(\frac{B(z)}{z}\Big)^k}{1-szA\Big(x\frac{B(z)}{z}\Big)} \tag{9.15}$$

It remains to determine the unknown coefficients $\tilde{q}^k(x,s)$ in (9.15). To do so we shall assume that there is a maximum number k_{max} so that $a(i) = 0$ for $i > k_{max}$. (This restriction we put on the discrete distribution is quite weak; since we always may approximate an infinite (countable) discrete distribution by a finite one by simply choosing k_{max} large enough.) This means that $A(z) = \sum_{i=1}^{k_{max}} a(i)z^i$ is a polynomial of degree k_{max}. With this assumption we may apply the powerful method, often used to determine unknown coeffi-

cients in generating functions given as a fraction, by locating the zeros of the dominator inside the unit disc and claiming analytical behaviour of the transforms in the same domain. If we set $K = k_{max} - 1$ we find from the definition that $\tilde{q}^k(x, s) = 0$ for $k > K - 1 = k_{max} - 2$. Next by applying the famous Rouche's theorem we have that for $|s| < 1$ and $|x| \le 1$ or $|s| \le 1$ and $|x| < 1$ that the equation:

$$1 - szA\left(x\frac{B(z)}{z}\right) = 0 \tag{9.16}$$

will have exactly K distinct roots $r_j = r_j(x, s)$; $j = 1, \ldots, K$ inside the unit disc $|z| < 1$. Moreover, by letting $z \to r_j$ we must have $r_j^m A\left(x\frac{B(r_j)}{r_j}\right)\frac{r_j}{B(r_j)} + \left(1 - \frac{1}{r_j}\right)\sum_{k=0}^{K-1}\tilde{q}^k(x, s)\left(\frac{B(r_j)}{r_j}\right)^k = 0$ or

$$\sum_{k=0}^{K-1}\tilde{q}^k(x, s)\left(\frac{B(r_j)}{r_j}\right)^k = -\frac{r_j^m}{1 - \frac{1}{r_j}}A\left(x\frac{B(r_j)}{r_j}\right)\frac{r_j}{B(r_j)} \quad \text{for } j = 1, \ldots, K \tag{9.17}$$

The equations (9.17) are linear and determine the unknown coefficients $\tilde{q}^k(x, s)$ uniquely. By exploiting the specific form of this linear system we find by applying (G.5) in Appendix G:

$$Q^m(z, x, s) = \frac{z^m A\left(x\frac{B(z)}{z}\right)\frac{z}{B(z)} - \sum_{j=1}^{K} r_j^m \frac{1 - \frac{1}{z}}{1 - \frac{1}{r_j}} A\left(x\frac{B(r_j)}{r_j}\right)\frac{r_j}{B(r_j)} \prod_{l=1, l \ne j}^{K} \frac{\frac{B(z)}{z} - \frac{B(r_l)}{r_l}}{\frac{B(r_j)}{r_j} - \frac{B(r_l)}{r_l}}}{1 - szA\left(x\frac{B(z)}{z}\right)} \tag{9.18}$$

The transform (9.18) gives the transient behaviour of the queueing model seen at instants just prior to the n'th arrival of the **FS** and will be a good starting point to obtain the distortion (or colouring) of the **FS** as it passes the multiplexer queue. For small values of the parameter K (9.18) will be well suited to obtain the desired transforms, however, for larger values of K it will be more efficient to transform the expression into a contour integral. In Appendix G (by (G.16) and (G.17)) we have found the product:

$$\prod_{l=1}^{K}\left(\frac{z}{B(z)} - \frac{r_l}{B(r_l)}\right) = \left(\frac{z}{B(z)}\right)^K\left[1 - szA\left(x\frac{B(z)}{z}\right)\right]\exp[I(z, x, s)] \tag{9.19}$$

where $I(z, x, s)$ is the contour integral:

$$I(z, x, s) = \frac{1}{2\pi i}\int_{C_r}\frac{\log\left[1 - s\zeta A\left(x\frac{B(\zeta)}{\zeta}\right)\right]}{\frac{z}{B(z)} - \frac{\zeta}{B(\zeta)}}\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)d\zeta \tag{9.20}$$

and where C_r is the disc $|\zeta| \le r$ where we can choose $1 < r < r_{K+1}$ and where r_{K+1} is the root of (9.16) outside the unit disc with the smallest modulo. (For more information see Appendix G)

From (9.19) we have $\prod_{l=1,\, l\neq j}^{K} \left(\frac{z}{B(z)} - \frac{r_l}{B(r_l)}\right) = \left(\frac{z}{B(z)}\right)^K \frac{h(z)}{\frac{z}{B(z)} - \frac{r_j}{B(r_j)}} \exp[I(z,x,s)]$. By taking the limit $z \to r_j$ gives $\prod_{l=1,\, l\neq j}^{K} \left(\frac{r_j}{B(r_j)} - \frac{r_l}{B(r_l)}\right) = \left(\frac{r_j}{B(r_j)}\right)^K \frac{h'(r_j)}{\frac{1}{B(r_j)} - \frac{r_j B'(r_j)}{B(r_j)^2}} \exp[I(r_j,x,s)]$ where we have set $h(z) = 1 - szA\left(x\frac{B(z)}{z}\right)$. Then by inserting for the products in (9.18) and simplifying, the expression for $Q^m(z,x,s)$ may be written as:

$$Q^m(z,x,s) = \frac{z}{B(z)}\left[\frac{z^m A\left(x\frac{B(z)}{z}\right)}{1 - szA\left(x\frac{B(z)}{z}\right)} + \sum_{j=1}^{K} r_j^{\,m} \frac{\left(1-\frac{1}{z}\right) r_j^{\,m} A\left(x\frac{B(r_j)}{r_j}\right)}{\left(1-\frac{1}{r_j}\right)\left(\frac{r_j}{B(r_j)} - \frac{z}{B(z)}\right)} \frac{\left(\frac{1}{B(r_j)} - \frac{r_j B'(r_j)}{B(r_j)^2}\right)}{h'(r_j)} \exp[I(z,x,s) - I(r_j,x,s)]\right] \tag{9.21}$$

We recognize (9.21) as (the residue expansion of) the following contour integral:

$$Q^m(z,x,s) = \frac{1}{2\pi i}\frac{z}{B(z)} \int_{C_u} \frac{\left(1-\frac{1}{z}\right) \zeta^m A\left(x\frac{B(\zeta)}{\zeta}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)\left(1 - s\zeta A\left(x\frac{B(\zeta)}{\zeta}\right)\right)} \exp[I(z,x,s) - I(\zeta,x,s)]d\zeta \tag{9.22}$$

where C_u is the disc $|\zeta| \le u$ with $u < 1$ and where the radius u is chosen so large that z and all the roots r_j (but not $z = 1$) are inside the disc C_u.

By the transforms (9.22) and (9.20) or (9.18) the transient course of the queueing model is fully described in terms of both the queue length and the time between arrivals of the **FS**. In the succeeding we shall use the integral form above, mainly because it is far more easy to obtain the limit $s \to 0$, needed to get the queuing behaviour between two succeeding arrivals of the **FS.** That is $Q_1(z,x) = \lim_{s\to 0} Q(z,x,s)$ which is easily obtained from (9.22) since $\lim_{s\to 0} I(z,x,s) = 0$ (from (9.20)). First, however, we need to find the stationary queueing distribution.

9.2.2 Stationary queue length distribution

Since we have expressions for the transient transforms, it is quite easy to get the corresponding steady state transforms. Under the assumption that $\rho = \rho_{FS} + \rho_{BS} < 1$ the station-

ary distribution exists and may for instance be found from the transforms above by applying Tauberian theorem [Fell68b]. We have

$$Q_0(z) = \boldsymbol{E}[z^{Q_0}] = \lim_{n \to \infty} Q_n(z, 1) = \lim_{s \to 1} (1-s) Q^m(z, 1, s) \tag{9.23}$$

To find the limit we may use either (9.18) or (9.21). From (9.16) with $x = 1$ we see that $z = 1$ is a root in the equation for $s = 1$ and therefore for one of the roots $r_j = r_j(1, s)$ we have $r_j(1, s) \to 1$ when $s \to 1$. We denote this root as r_1 and we find $\frac{dr_1}{ds}(1, 1) = \frac{1}{A'(1)(1-\rho)}$. By applying (9.18) we find the limit (9.23):

$$Q_0(z) = \frac{A'(1)(1-\rho)\left(1-\frac{1}{z}\right)}{1 - zA\left(\frac{B(z)}{z}\right)} \prod_{l=2}^{K} \frac{\frac{B(z)}{z} - \frac{B(r_l^*)}{r_l^*}}{1 - \frac{B(r_l^*)}{r_l^*}} \quad \text{with } r_l^* = r_l(1, s) \tag{9.24}$$

If we use (9.21) then the corresponding stationary transform is found to be:

$$Q_0(z) = (1 - B'(1)) \frac{(z-1)}{z - B(z)} \exp[I^*(z) - I^*(1)] \quad \text{with} \tag{9.25}$$

$$I^*(z) = I(z, 1, 1) = \frac{1}{2\pi i} \int_{C_r} \frac{\log\left[1 - \zeta A\left(\frac{B(\zeta)}{\zeta}\right)\right]}{\frac{z}{B(z)} - \frac{\zeta}{B(\zeta)}} \left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right) d\zeta \tag{9.26}$$

For numerically calculations of the transform above, used for finding the stationary distribution, it is obvious that for small value of K (9.24) will be preferable since a small numbers of roots may be easily found numerically. However, for large values of K the product in (9.24) will become numerically unstable and it is better to evaluate the integral $I^*(z)$ numerically and apply (9.25). In (9.25) we also recognize the queue length transform with only the **BS** present (without the **FS**)

$$Q_{BS}(z) = (1 - B'(1)) \frac{(z-1)}{z - B(z)}, \tag{9.27}$$

so the exponential term $\exp[I^*(z) - I^*(1)]$ represents the "add on" due the **FS**. It seems also obvious that if the mean time between arrivals in the **FS** is large, that is $\rho_{FS} = \frac{1}{A'(1)}$ is small, then $\exp[I^*(z) - I^*(1)] \approx 1$ and $Q_{BS}(z)$ will be a reasonable approximation of $Q_0(z)$.

9.3 Delay and delay jitter for the FS

Based on the transient transforms (9.18) or (9.22) and the corresponding stationary transforms (9.24) or (9.25) we may find the characteristics of the output process for the **FS** when the queueing system is in steady state. Although we consider an FCFS (First-Come First-Serve) there still remain question in which order an **FS** packet is served when it arrives in a slot with arrivals of (possible several) **BS** packets. Below we have analysed three possible orderings of an **FS** packet when it arrives in a slot with (possible several) **BS** packets:

- ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets
- ***R***-The **FS** packet and possible **BS** packets is served at *random*
- ***L***-The **FS** packet is always served *last* when it arrives together with **BS** packets

Among these three orderings the random will be the most important one since this requires no special treatment of any packets (from any arrival streams). We let D_n denotes the delay for the n'arrival from the **FS**. Then we have

$$D_n = Q_n + U_n \text{ where} \tag{9.28}$$

U_n is the delay for a **FS** packet due to the possible arrivals of **BS** packets in the same slot. We are interested in the joint distribution of $D_n - D_0$ and T_n and we define the transforms:

$$W_n(z, x) = \boldsymbol{E}[z^{D_n - D_0} x^{T_n}] \text{ and} \tag{9.29}$$

$$W(z, x, s) = \sum_{n=1}^{\infty} s^{n-1} W_n(z, x) \tag{9.30}$$

Below we shall analyse the three cases described above separately to obtain the transform $W(z, x, s)$ and we shall use (***F***-*first*, ***R***-*random* and ***L***-*last*) as superscript to indicate (and distinguish) the different cases.

9.3.1 The FS packet is always served *first* when it arrives together with BS packets

In this case we have $D_n = Q_n + 1$ so $D_n - D_0 = Q_n - Q_0$. By conditioning on $Q_0 = k$ and $Q_0^1 = m$ we find:

$$\boldsymbol{E}[z^{D_n - D_0} x^{T_n}] = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \boldsymbol{E}[z^{Q_n - Q_0} x^{T_n} \big| Q_0 = k, Q_0^1 = m] P(Q_0 = k, Q_0^1 = m) =$$

$$\sum_{k=0}^{\infty} \sum_{m=0}^{\infty} z^{-k} Q_n(z, x) P(Q_0 = k, Q_0^1 = m) \text{ or}$$

$W_F(z,x,s) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} z^{-k}Q^m(z,x,s)P(Q_0 = k, Q_0^1 = m)$. Since $Q_0^1 = B_0^1 + Q_0$ we find the joint transform:

$$\boldsymbol{E}\left[z_1^{Q_0}z_2^{Q_0^1}\right] = \boldsymbol{E}[(z_1z_2)^{Q_0}]\boldsymbol{E}\left[z_2^{B_0^1}\right] = B(z_2)Q_0(z_1z_2) \tag{9.31}$$

By using (9.22) for $Q^m(z,x,s)$ and applying (9.31) with $z_1 = z^{-1}$ and $z_2 = \zeta$ we get:

$$W_F(z,x,s) = \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u}\frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(x\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)}-\frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)}-\frac{z}{B(z)}\right)\left(1-s\zeta A\left(x\frac{B(\zeta)}{\zeta}\right)\right)}\exp[I(z,x,s)-I(\zeta,x,s)]d\zeta \tag{9.32}$$

9.3.2 The FS packet and possible BS packets arriving in the same slot are served at *random*

In this case we have $D_n = Q_n + U_n$ where U_n is the number of **BS** packets arriving in the same slot as an **FS** packet and is placed prior to the **FS** packet when the mutual position among them is chosen at random. Then we have $D_n - D_0 = U_n + Q_n - D_0$ and by conditioning on $D_0 = k$ and $Q_0^1 = m$ we find:

$$\boldsymbol{E}[z^{D_n-D_0}x^{T_n}] = E[z^{U_n}]\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\boldsymbol{E}[z^{Q_n-D_0}x^{T_n}|D_0 = k, Q_0^1 = m]P(D_0 = k, Q_0^1 = m) =$$

$$E[z^{U_n}]\sum_{k=0}^{\infty}\sum_{m=0}^{\infty} z^{-k}Q_n(z,x)P(D_0 = k, Q_0^1 = m) \text{ or}$$

$W_R(z,x,s) = E[z^{U_n}]\sum_{k=0}^{\infty}\sum_{m=0}^{\infty} z^{-k}Q^m(z,x,s)P(D_0 = k, Q_0^1 = m)$. Since $Q_0^1 = B_0^1 + Q_0$ and $D_0 = U_0 + Q_0$ we find the joint transform:

$$\boldsymbol{E}\left[z_1^{D_0}z_2^{Q_0^1}\right] = \boldsymbol{E}\left[z_1^{U_0}z_2^{B_0^1}\right]\boldsymbol{E}[(z_1z_2)^{Q_0}] = U(z_1,z_2)Q_0(z_1z_2) \tag{9.33}$$

In Appendix G we have found the z-transform of the joint distribution of "extra" delay U_n and the number of arrivals from the **BS** B_n^1 as, (see (G.19)):

$$U(z_1,z_2) = \boldsymbol{E}\left[z_1^{U_n}z_2^{B_n^1}\right] = \frac{z_1}{(1-z_1)z_2}[BI(z_2) - BI(z_1z_2)] \tag{9.34}$$

where $BI(z) = \int_1^z B(x)dx$ is the integral of the z-transform of the **BS**.

By using (9.22) for $Q^m(z, x, s)$ and applying (9.33) and (9.34) with $z_1 = z^{-1}$ and $z_2 = \zeta$ we find:

$$W_R(z, x, s) = \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u}\frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(x\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)\left(1 - s\zeta A\left(x\frac{B(\zeta)}{\zeta}\right)\right)}\exp[I(z, x, s) - I(\zeta, x, s)]d\zeta \tag{9.35}$$

9.3.3 The FS packet is always served *last* when it arrives together with BS packets

In this case we have $D_n = Q_n + B_n^1 + 1$ so $D_n - D_0 = B_n^1 + Q_n - Q_0^1$. Conditioning on $Q_0^1 = m$ we obtain:

$$\boldsymbol{E}[z^{D_n - D_0}x^{T_n}] = E\left[z^{B_n^1}\right]\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\boldsymbol{E}\left[z^{Q_n - Q_0^1}x^{T_n}\middle|Q_0^1 = m\right]P(Q_0^1 = m) =$$

$B(z)\sum_{m=0}^{\infty} z^{-m}Q_n(z, x)P(Q_0^1 = m)$ or $W(z, x, s) = \sum_{m=0}^{\infty} z^{-m}Q^m(z, x, s)P(Q_0^1 = m)$. Since $Q_0^1 = B_0^1 + Q_0$ we find the transform:

$$\boldsymbol{E}\left[z^{Q_0^1}\right] = \boldsymbol{E}[(z)^{Q_0}]\boldsymbol{E}\left[z_2^{B_0^1}\right] = B(z)Q_0(z) \tag{9.36}$$

By using (9.22) for $Q^m(z, x, s)$ and applying (9.36) with $z \to \frac{\zeta}{z}$ gives:

$$W_L(z, x, s) = \frac{1}{2\pi i}\int_{C_u}\frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(x\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1 - \frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)\left(1 - s\zeta A\left(x\frac{B(\zeta)}{\zeta}\right)\right)}\exp[I(z, x, s) - I(\zeta, x, s)]d\zeta \tag{9.37}$$

9.3.4 Inter-departure time, jitter and queueing delay distributions

The joint transforms (9.32), (9.35) and (9.37), one for each of queueing discipline defined, makes it easy to find the inter-departure time distribution, jitter and queueing delay. We have that the jitter J is the difference of the delay for two succeeding packets from the **FS**:

$$J = D_1 - D_0 \text{ and} \tag{9.38}$$

the corresponding z-transform is simply:

$$J(z) = \boldsymbol{E}[z^{D_1 - D_0}] = W(z, 1, 0) \tag{9.39}$$

Similarly the inter-departure time between succeeding packets G from the **FS** is simply the difference between the departure times:

$$G = T_1 + D_1 - T_0 - D_0 = D_1 - D_0 + T_1 \text{ and} \tag{9.40}$$

the z-trasform is found by:

$$G(z) = \boldsymbol{E}[z^{D_1 - D_0 + T_1}] = W(z, z, 0) \tag{9.41}$$

We also denote $D(z) = \boldsymbol{E}[z^{D_0}]$ the z-tranform of the delay for a packet from the **FS.**

It is now easy to find the transforms for the three different cases defined above by using (9.32), (9.35) and (9.37):

a. ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets

$$J^F(z) = \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.42}$$

$$G^F(z) = \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \quad \text{and} \tag{9.43}$$

$$D^F(z) = zQ_0(z) \tag{9.44}$$

b. ***R***-The **FS** packet and possible **BS** packets are served at *random*

$$J^R(z) = \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.45}$$

$$G^R(z) = \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \quad \text{and} \tag{9.46}$$

$$D^R(z) = \frac{z}{(z-1)}BI(z)Q_0(z) \tag{9.47}$$

c. ***L***-The FS packet is always served last when it arrives together with BS packets

$$J^L(z) = \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.48}$$

$$G^L(z) = \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \text{ and} \tag{9.49}$$

$$D^L(z) = zB(z)Q_0(z) \tag{9.50}$$

In all the cases above the steady state z-tranform of the queue length distribution $Q_0(z)$ is given by either (9.24) (the root representation) or (9.25) and (9.26) (the integral representation). Further the contour C_u is chosen as a circle $\{|\zeta| = u\}$ so that $\zeta = z$ is inside but not $\zeta = 1$, that is $|z| < u < 1$.

9.3.5 Some variants of the inter-departure time and jitter z-transforms

The representation of the z-transforms above in terms of complex integrals is quite beneficial since it is possible to transform the contour by picking up the poles. This can be done for the poles at $\zeta = z$ and $\zeta = 1$. By letting the contour C_u be chosen so that both $\zeta = z$ and $\zeta = 1$ is outside we get by picking up the pole $\zeta = z$:

a. ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets

$$J^F(z) = zA\left(\frac{B(z)}{z}\right) + \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.51}$$

$$G^F(z) = zA(B(z)) + \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \text{ and} \tag{9.52}$$

b. ***R***-The **FS** packet and possible **BS** packets are served at *random*

$$J^R(z) = \frac{z}{B(z)}\left(\frac{BI(z)}{z-1}\right)^2 A\left(\frac{B(z)}{z}\right) + \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.53}$$

$$G^R(z) = \frac{z}{B(z)}\left(\frac{BI(z)}{z-1}\right)^2 A(B(z)) + \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.54}$$

c. ***L***-The FS packet is always served last when it arrives together with BS packets

$$J^L(z) = zA\left(\frac{B(z)}{z}\right) + \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.55}$$

$$G^L(z) = zA(B(z)) + \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.56}$$

where, in all of the integrals (9.51)-(9.56), the contour C_u is chosen as the circle $\{|\zeta| = u\}$ with $u < |z|$.

By letting the contour C_u be chosen so that both $\zeta = z$ and $\zeta = 1$ are inside we get by including the pole $\zeta = z$:

a. ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets

$$J^F(z) = D^F_{SB}(z)D^F\left(\frac{1}{z}\right) + \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.57}$$

$$G^F(z) = D^F_{SB}(z)D^F\left(\frac{1}{z}\right)A(z) + \frac{1}{2\pi i}\frac{z}{B(z)}\int_{C_u} \frac{\left(1-\frac{1}{z}\right)B(\zeta)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \text{ and} \tag{9.58}$$

b. ***R***-The **FS** packet and possible **BS** packets are served at *random*

$$J^R(z) = D^R_{SB}(z)D^R\left(\frac{1}{z}\right) + \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.59}$$

$$G^R(z) = D^R_{SB}(z)D^R\left(\frac{1}{z}\right)A(z) + \frac{1}{2\pi i}\frac{zBI(z)}{(z-1)B(z)}\int_{C_u} \frac{\left(BI(\zeta) - BI\left(\frac{\zeta}{z}\right)\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{(\zeta-1)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.60}$$

c. ***L***-The FS packet is always served last when it arrives together with BS packets

$$J^L(z) = D^L_{SB}(z)D^L\left(\frac{1}{z}\right) + \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.61}$$

$$G^L(z) = D^L_{SB}(z)D^L\left(\frac{1}{z}\right)A(z) + \frac{1}{2\pi i}\int_{C_u} \frac{(z-1)B\left(\frac{\zeta}{z}\right)A\left(z\frac{B(\zeta)}{\zeta}\right)Q_0\left(\frac{\zeta}{z}\right)\left(\frac{1}{B(\zeta)} - \frac{\zeta B'(\zeta)}{B(\zeta)^2}\right)}{\left(1-\frac{1}{\zeta}\right)\left(\frac{\zeta}{B(\zeta)} - \frac{z}{B(z)}\right)} d\zeta \tag{9.62}$$

where $D^F(z)$, $D^R(z)$ and $D^L(z)$ are the z-transforms of the delays given by (9.44), (9.47) and (9.50) respectively, and $D^F_{SB}(z)$, $D^R_{SB}(z)$ and $D^L_{SB}(z)$ are the corresponding z-transforms without the **FS** (when the load from the **FS** is set to zero, i.e. obtained by replacing $Q_0(z)$ with the simpler $Q_{SB}(z)$ given by (9.27)). Further, in all the integrals (9.57)-(9.62) the contour C_u is chosen as the circle $\{|\zeta| = u\}$ with $u > 1$.

The first integrals (initial results) for the jitter and inter-departure time given by (9.42) and (9.43), (9.45) and (9.46) and (9.48) and (9.49) have the main drawback that it we can not take the limit $z \to 1$ since we claim that z is inside a circle with radius less than unity. By the results given in section 9.3.5 this limitation is removed and we observe that all the transforms given by (9.51)-(9.62) have the limit unity when $z \to 1$. Moreover, these results are also suitable to obtain the moments by differentiation and taking the limit $z \to 1$. We shall omit such calculations here since we mainly are interested in the distributions. The mean time between packets from the **FS** should however remain unchanged as the stream passes the multiplexer (and this have of cause been checked by numerical examples).

9.4 Heavy and light traffic analysis

Another striking observation by considering the representation (9.51)-(9.62) is that "the non integral part" of the expressions represents distributions which turns out to yields for heavy and light traffic. This can be argued as follows: If the load is close to one i.e. $\rho_{BS} \to 1-\rho_{FS}$, the queueing system will never be empty and the boundary transforms $q^i_n(x) = \boldsymbol{E}\left[x^{T_n}\mathbf{1}_{\{Q^i_n + B^{i+1}_n = 0\}}\middle| T_{n+1} - T_n = k\right]$ will be zero, implying that $\tilde{q}^k_n(x)$ and $\tilde{q}^k(x,s)$ are zero giving (9.22) on the form $Q^m(z,x,s) = \frac{z}{B(z)}\left[\frac{z^m A\left(x\frac{B(z)}{z}\right)}{1 - szA\left(x\frac{B(z)}{z}\right)}\right]$. By applying the methods described in the previous sections to find the inter-departure time and jitter distributions we obtain the following heavy traffic approximations (obtained simply by setting the integrals in the expressions (9.51)-(9.56) equal to zero):

a. ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets

$$J^F(z) = zA\left(\frac{B(z)}{z}\right) \tag{9.63}$$

$$G^F(z) = zA(B(z)) \text{ and} \tag{9.64}$$

b. ***R***-The **FS** packet and possible **BS** packets are served at *random*

$$J^R(z) = \frac{z}{B(z)}\left(\frac{BI(z)}{z-1}\right)^2 A\left(\frac{B(z)}{z}\right) \tag{9.65}$$

$$G^R(z) = \frac{z}{B(z)}\left(\frac{BI(z)}{z-1}\right)^2 A(B(z)) \tag{9.66}$$

c. ***L***-The FS packet is always served last when it arrives together with **BS** packets

$$J^L(z) = zA\left(\frac{B(z)}{z}\right) \tag{9.67}$$

$$G^L(z) = zA(B(z)) \tag{9.68}$$

To find the light traffic approximation i.e. $\rho_{BS} \to 0$ we may argue as follows: If we simultaneously consider the queueing process just prior to the arrival of an **FS** packet and the arrival process of the **FS** packets, these two processes will not affect each other, and they may therefore be analysed separately as if *one* of them was switched off. If we also may assume that the relaxation time to reach the steady state for the reduced queueing model (with only the **BS** present) is much shorter than the time interval to the arrival of the next **FS** packet, then the two processes may be treated as independent and we may write $Q^m(z,x,s) = Q_{BS}(z)\frac{A(x)}{1-sA(x)}$ where $Q_{BS}(z)$ is the z-transform for the reduced queueing system (with the **FS** switched off) and is given by (9.27)). By applying the methods described in the previous sections to find the inter-departure time and jitter distributions we find the following light traffic approximations (obtained simply by setting the integrals in the expressions (9.57)-(9.62) equal to zero):

a. ***F***-The **FS** packet is always served *first* when it arrives together with **BS** packets

$$J^F(z) = D^F_{SB}(z)D^F\left(\frac{1}{z}\right) \tag{9.69}$$

$$G^F(z) = D^F_{SB}(z)D^F\left(\frac{1}{z}\right)A(z) \text{ and} \tag{9.70}$$

b. ***R***-The **FS** packet and possible **BS** packets are served at *random*

$$J^R(z) = D^R_{SB}(z)D^R\left(\frac{1}{z}\right) \tag{9.71}$$

$$G^R(z) = D^R_{SB}(z)D^R\left(\frac{1}{z}\right)A(z) \tag{9.72}$$

c. ***L***-The FS packet is always served last when it arrives together with BS packets

$$J^L(z) = D^L_{SB}(z)D^L\left(\frac{1}{z}\right) \tag{9.73}$$

$$G^L(z) = D^L_{SB}(z)D^L\left(\frac{1}{z}\right)A(z) \tag{9.74}$$

The key assumption for this approximation is that the queueing system with only the **BS** present shall reach steady state in the interval between two successive arrivals from the **FS**. It is clear that when the mean time between two such arrivals is small, say just some few slots, then the corresponding load ρ_{BS} from the **BS** must be very small to reach the steady state (in such a short interval). On the other hand, if the mean time between two successive arrivals from the **FS** is large, then the load from the **BS** may be moderate (but not close to one). So actually the requirements for the light traffic approximation are:

a. ρ_{FS} is moderate and $\rho_{BS} \to 0$ or

b. $\rho_{FS} \to 0$ and ρ_{FB} is moderate

In case b, we also may use the simpler $D_{SB}(z)$ for $D(z)$ in (9.69)-(9.74) since we assume $\rho_{FS} \to 0$.

9.5 End-to-end delay and jitter evaluation for a stream traversing a series of queueing nodes

The main objective in this chapter was to develop analytical models which were possible to extend to also cover end-to-end analysis that goes beyond the traditional models based on convolutions described in chapter 7 and chapter 8. With these models the distortion (colouring) of a particular packet stream as it passes a multiplexer is neglected. A more exact approach will be to consider a particular packet stream as it passes through a network and try to describe the change in the stream as it traverses the nodes where it will be disturb by other (background) traffic.

By the slotted model described in the this chapter we may analyse in detail the output process for a particular packet stream given that the same process at input is a renewal process. In particular we have analysed the distribution of the time between two successive departures. If we approximate the output stream with a renewal stream (which is fully characterized by the distribution between two successive renewals), we may take this renewal stream as the input to the next node and thereby apply the queueing model recursively to get an end-to-end description. By this method we may “track” a given packet stream from source to destination as it crosses a multiple of nodes. In figure 9.2 we have depicted the key idea behind the end-to-end model. It is, however, well known that the output process of the queueing model described in section 9.2 will not be exactly renewal. Nevertheless, simulation studies [Matr94b] indicate that this type approximation indeed is very good if we only consider the marginal distribution of the output processes. Especially the evolution of the jitter, but also the end-to-end delay will therefore be analysed more accurately than the convolutions given in chapter 7. However, the resulting model will be much more complicated and the results are given recursively in terms of the results found in section 9.3.

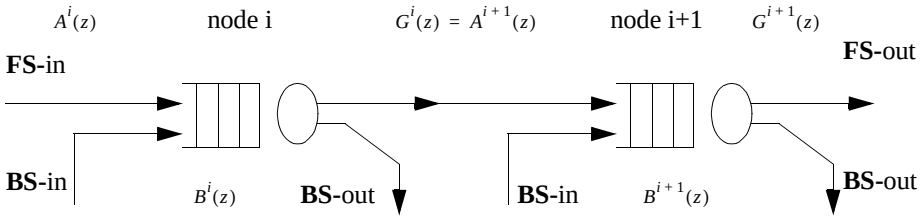

Figure 9.2: The tandem queueing model for the end-to-end modelling.

In the following we shall consider a chain of n queueing nodes. We shall make the following assumptions (similar to the assumptions in section 7.4):

- The streaming traffic (**FS)** enters node 1 according to a discrete time renewal process with distribution between arrivals given by $a_1(m)$ and generating function $A_1(z)$
- The background traffic (**BS)** at node i enters (and leaves the node) according to a batch process with distribution $b_i(m)$ and generating function $B_i(z)$
- The queueing discipline is FIFO for all the queues and the possible orderings when a streaming packet (**FS)** arrives in a slot with (possible several) background packets (**BS)** is described in section 9.3

If we let $d_i(m)$ denote distribution of the delay with generating function $D_i(z)$ and $g_i(m)$ denote distribution of the jitter with generating function $G_i(z)$ at node i we may write the functional relations (for $i = 1, \ldots, n$):

$$D_i(z) = F_I(A_i(z), B_i(z)) \tag{9.75}$$

$$G_i(z) = F_{II}(A_i(z), B_i(z)) \text{ and} \tag{9.76}$$

$$A_{i+1}(z) = G_i(z) \tag{9.77}$$

where the functional entities F_I and F_{II} relate the delay distribution and inter-departure distribution as “functions” of the arrival processes. The actual form of F_I is given by the z-transform of the steady state queue length distribution $Q_0(z)$ (as given by (9.24) or (9.25) and (9.26)) and the relations (9.44), (9.47) or (9.50) depending on the scheduling models chosen, and further F_{II} is either the representations (9.52) or (9.54), (9.56) or (9.58), (9.60) or (9.62) depending on the scheduling models chosen and the integration path chosen in the corresponding contour integrals.

Finally, if we denote $d^n(m)$ the distribution of the end-to-end delay for a chain of n successive nodes, then we find $d^n(m)$ by taking convolutions of the delay distributions at each node. Written recursively we have:

$$d^i(m) = \sum_{l=0}^{m} d^{i-1}(l) d_i(m-l) \text{ for } i = 2, \ldots, n \tag{9.78}$$

The corresponding z-transform is the product of the z-transforms in (9.75):

$$D^n(z) = \prod_{i=1}^{n} D_i(z) \tag{9.79}$$

9.6 Some comments on the numerical procedure to calculate the end-to-end delay and jitter

Although the recursions given by (9.75)-(9.77) are analytical in nature, the corresponding procedure is highly numerical and contains some key assumptions: In each iteration both $D_i(z)$ and $G_i(z)$ are evaluated by calculating the corresponding distributions numerically, and we truncate the distributions when the probabilities are less than some quoted accuracy. The distributions $d_i(m)$ and $g_i(m)$ are calculated by (Cauchys theorem):

$$d_i(m) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} D_i(e^{i\theta}) e^{-im\theta} d\theta \text{ for } m = 1, 2, \ldots, N_{D_i}^{max} \text{ and} \tag{9.80}$$

$$g_i(m) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} G_i(e^{i\theta}) e^{-im\theta} d\theta \text{ for } m = 1, 2, \ldots, N_{G_i}^{max} \text{ and} \tag{9.81}$$

then we take $D_i(z) = D_i^{num}(z) = \sum_{m=1}^{N_{D_i}^{max}} d_i(m) z^m$ and $G_i(z) = G_i^{num}(z) = \sum_{m=1}^{N_{G_i}^{max}} g_i(m) z^m$ and we take $A_{i+1}(z) = G_i^{num}(z)$ and since $G_i^{num}(z)$ is a polynomial we can apply the results in the previous sections (9.2 and 9.3) to calculate the next iteration in the chain.

To calculate the integral $I^*(z)$ given by (9.26) the circle (contour) C_r has to be chosen so that $r > 1$ but smaller than the first root of the function $f_i(\zeta) = 1 - \zeta A_i\left(\frac{B_i(\zeta)}{\zeta}\right)$ outside the unit circle. It is possible to shown that for a stable queue, $f_i(\zeta)$ has exactly one real root $\zeta = r_i^{out}$ outside the unit circle and that all other roots are outside a circle of radius equal

to that particular root. Therefore we must choose $1 < r < r_i^{out}$. A possible choice is to take the point where the function $f_i(x)$ attains its maximum i.e. choose r as the solution of the equation $f_i'(x) = 0$.

Since $G_i(z)$ is calculated for $z = e^{i\theta}$ i.e. for $|z| = 1$ by (9.81) we must use the expression in section 9.3.5 in the calculations. If the expressions (9.52), (9.54) or (9.56) are used, the circle (contour) C_u must be chosen so that $u < 1$ (and $u = 1/2$ seems to be a natural choice). On the other hand, if we choose the expressions (9.58), (9.60) or (9.62) then we must take $u > 1$ and we may choose $u = r$ as above.

9.7 Some numerical examples

In the numerical examples below we have taken the following input streams:

- The **FS** is either deterministic or geometrically distributed with mean time between arrivals equal $\frac{1}{\rho_{FS}}$.
- The **BS** is a Poisson stream with parameter ρ_{BS}.

For most of the examples we have taken the ***R***(random) queueing discipline where there is a random selection of all packets arriving in the same slot, but we have also given a few examples with the other two disciplines **F**(first) and **L**(last), more ore less to check out the numerical results and also see how sensitive the end-to-end delay and the evolution of the jitter are to the particular scheduling choice. We also assume that the load from the **FS** is relatively low, and we have taken the mean time between arrivals of the FS to be 10 or $\rho_{FS} = 0.1$ for most of the examples, but we also have a few examples with mean time between arrivals of the FS to be 5 or $\rho_{FS} = 0.2$.

We have two main goals with the examples:

a. Firstly to compare this rather heavy numerical approach with the convolution approach (given in chapter 7) and
b. secondly to investigate the evolution of the jitter distribution as the **FS** traverses a chain of queues.

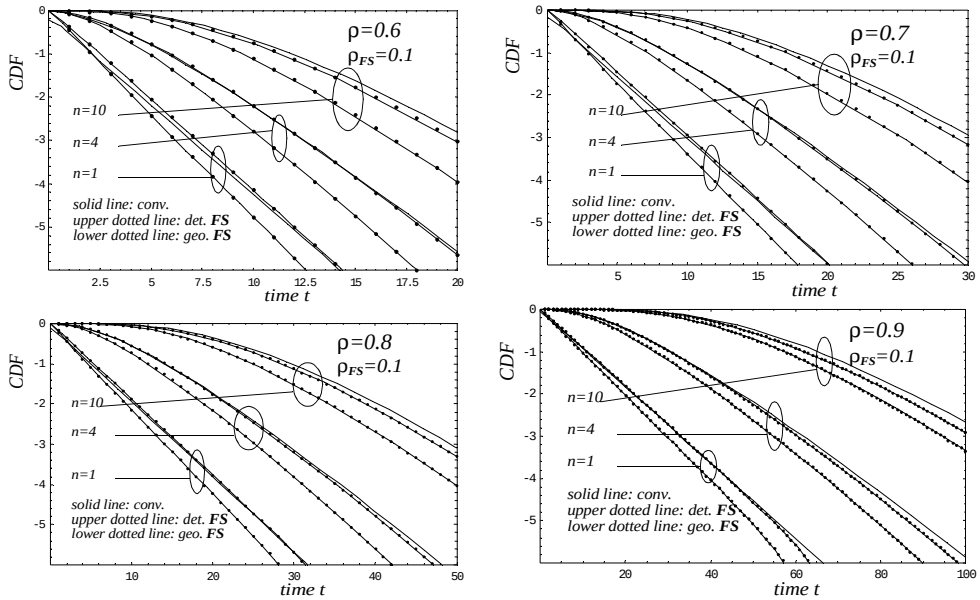

*Figure 9.3: Logarithmic plot of the end-to-end delay for **R**(random)-queueing discipline and some different parameters as function of time (in slots).*

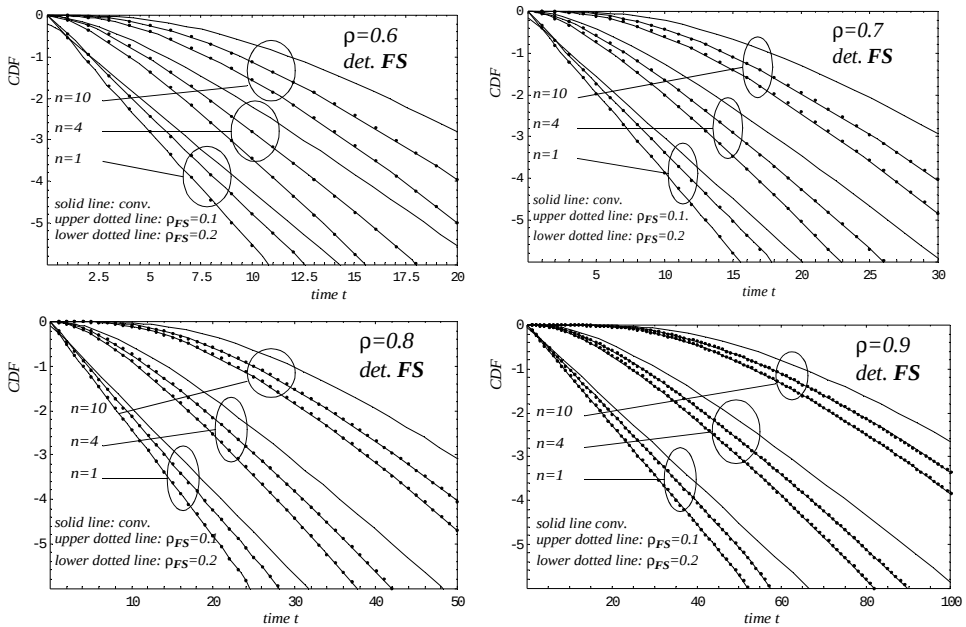

*Figure 9.4: Logarithmic plot of the end-to-end delay for **R**(random)-queueing discipline and some different parameters as function of time (in slots).*

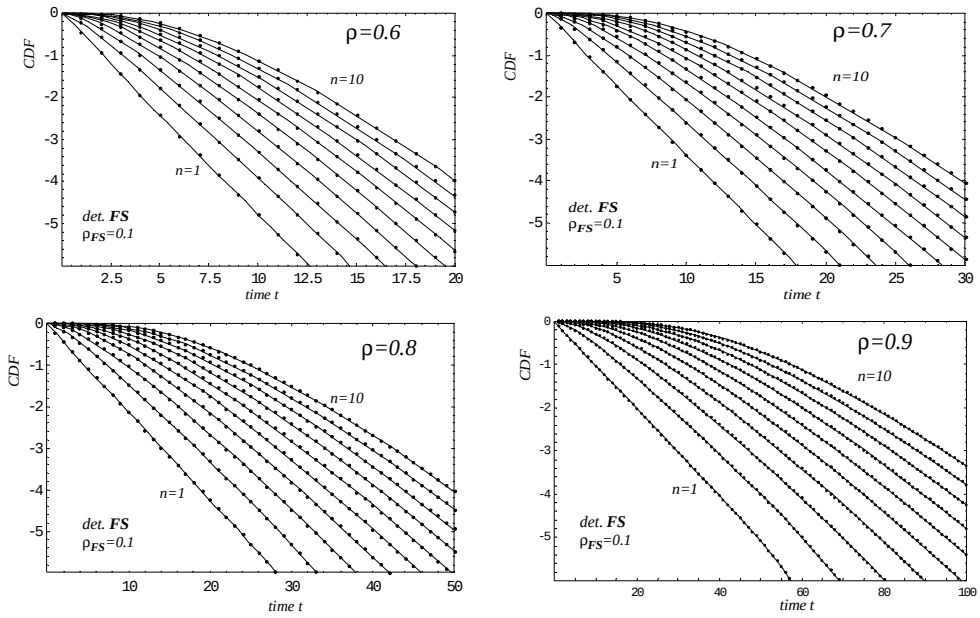

*Figure 9.5: Logarithmic plot of the end-to-end queuing delay for **R**(random)-queueing discipline and some different parameters as function of time (in slots).*

9.7.1 End-to-end delay

In figures 9.3-9.8 we have depicted the CDF of the end-to-end delay for various parameter choices and where we also have plotted the corresponding results obtained by the convolution approach. In the first graphs (figure 9.3) we have compared cases where the **FS** is deterministic with the case where **FS** is geometrically distributed, and where we have chosen the ***R***(andom) queueing discipline (with random selection of all packets arriving in the same slot). We observe that for all these cases the deterministic **FS** gives the best performance (when it comes to end-end-delay) but the difference is not very large, and it seems that the actual difference is decreasing slightly as the load increases. We also observe that the convolution approach and the case with FS being geometrically distributed nearly coincide, and this is expected since the "sum of" a thin geometrical stream and a Poisson stream will more or less also be a Poisson stream.

In figure 9.4 we have compared the cases where we halve the mean time between arrivals of the **FS** (while not changing the total load). (In this example we have deterministic **FS** and ***R***(random) queueing discipline.) As expected the effect of increasing the load from a deterministic stream while keeping the total load constant will lead to a stream with less variance; and hence the queueing performance will improve.

Figure 9.5 shows how the end-to-end delay evolves as the FS passes through the series of queues in the chain. The shape of the curves seems to be quite similar for different loads, however, we must bear in mind that the axis is scaled differently in the four cases.

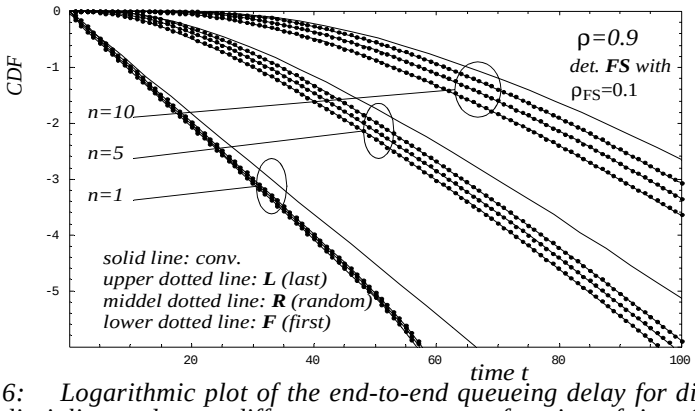

Figure 9.6: Logarithmic plot of the end-to-end queueing delay for different queueing discipline and some different parameters as function of time (in slots).

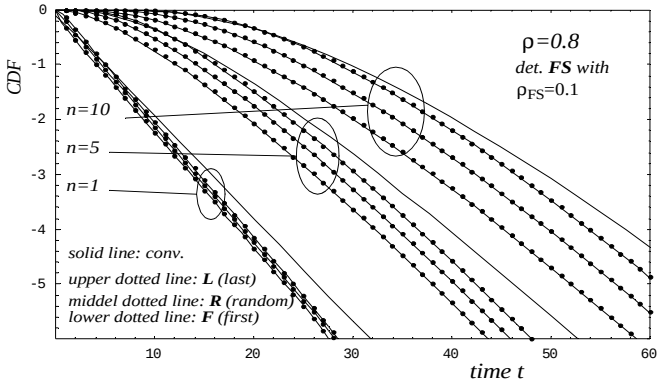

Figure 9.7: Logarithmic plot of the end-to-end queueing delay for different queueing discipline and some different parameters as function of time (in slots).

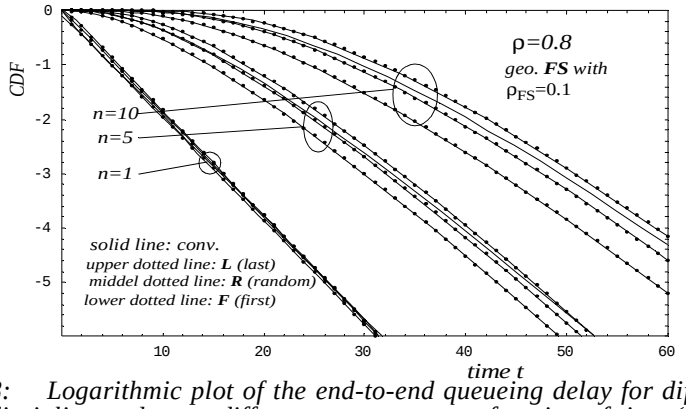

Figure 9.8: Logarithmic plot of the end-to-end queueing delay for different queueing discipline and some different parameters as function of time (in slots).

In figures 9.6-9.8 we have studied the effects of having different scheduling of the packets from the two streams arriving in the same slot. In figures 9.6 and 9.7 we have deterministic **FS** while in 9.8 geometrical **FS** is used. It seems that for deterministic **FS** all the three scheduling give end-to-end delay that are below that of the convolution. It seems also that the difference between the three scheduling principle will decrease as the load increases, while maintaining the other parameters.

From figure 9.8 we get the only case where the slotted model gives worse performance than the convolution. This occurs when we have geometrical **FS** and choose the **L**(last) queueing principles where the **FS** packet is placed behind all the **BS** packets arriving in the same slot.

As a conclusion to the numerical examples for the end-to-end delay we have seen that the convolution approach for all, but except on particular case, will give an upper bound the end-to-end delay compared with the slotted model considered in this chapter. Whether this is a result of general validity will not tried to be answered here.

9.7.2 Evolution of the jitter

The jitter a packet stream is inflicted will be an important measure for the QoS in a communication network. For real time services the jitter will decide the dimension of the dejitter buffer needed to obtain a regular bit stream at the receiver site. Generally the jitter is difficult to analyse since it represents a difference between two variables that are not independent.

In figures 9.9-9.16 we have depicted a series of examples for the evolution of the jitter for the **FS.** We have put main emphasis in the node-to-node evolution as the stream passes through a chain of nodes. It is of main interest to examine the disturbance of a regular stream as it passes through a network and we therefore mainly consider cases where **FS** is deterministic.

By figure 9.9 we have plotted the PDF of jitter where we look into the different scheduling strategies **R**(random), **F**(first) and **L**(last) for a deterministic **FS**. In these examples the load is set to 0.7 and the mean inter-arrival time for the **FS** is taken to be 10. As expected the scheduling **F**(first) gives the most narrow jitter, and in between is curves for **R**(random) scheduling, while the **L**(last) scheduling gives the broadest jitter (density function). This is most evident at the first queues in the chain. As the number of passed queues increases the jitter get broader and the difference becomes less visible. Even though the jitter seems to be symmetrical for small numbers of queues, we observe that after passing the 10’th queue the jitter is not completely symmetrical any more.

We have also considered a case where PDF of the jitter evolution for a geometric **FS** is compared with that for a deterministic input stream, (see figure 9.10). The changes in the jitter for the **FS** are very small and it looks as if this type of stream remains unchanged as it passes through the multiplexers. The reason is that this type stream is very similar to a Poisson stream and will also be nearly Poisson at the output of a multiplexer.

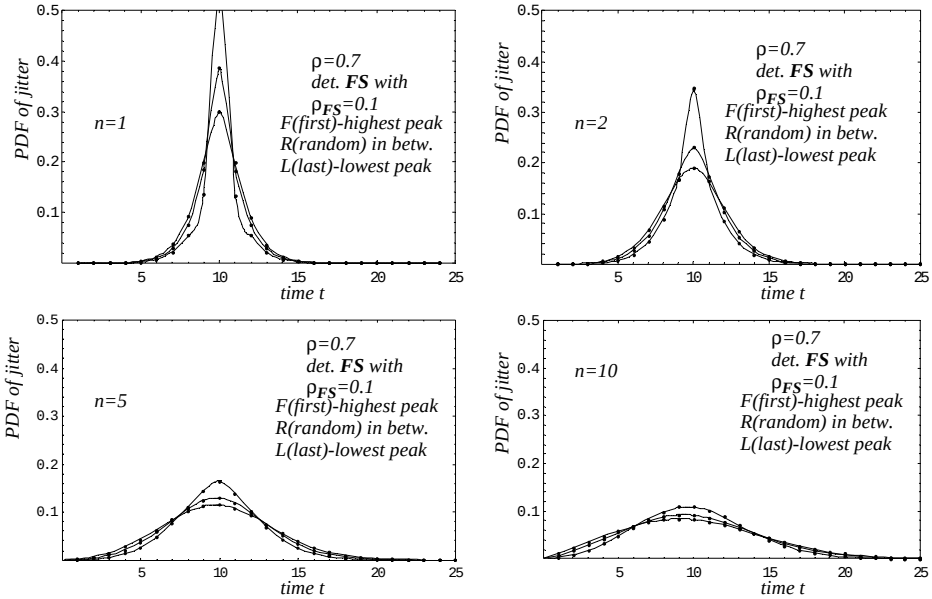

*Figure 9.9: PDF of the jitter as function of time for increasing number of nodes for the different queueing disciplines, **F**(first), **R**(random) and **L**(last).*

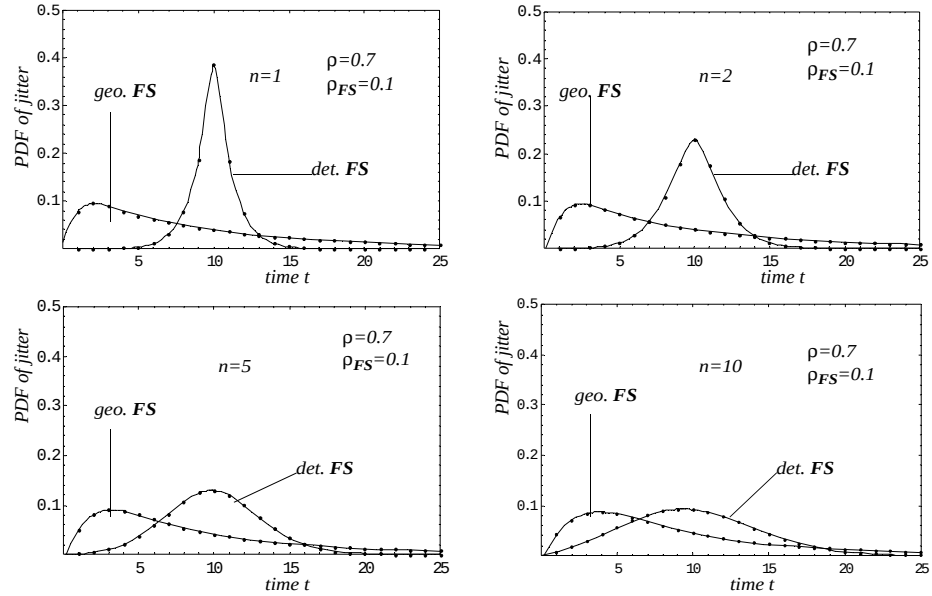

*Figure 9.10: PDF of the jitter as function of time for increasing number of nodes for geometric and deterministic distributed **FS** and **R**(random) queueing discipline.*

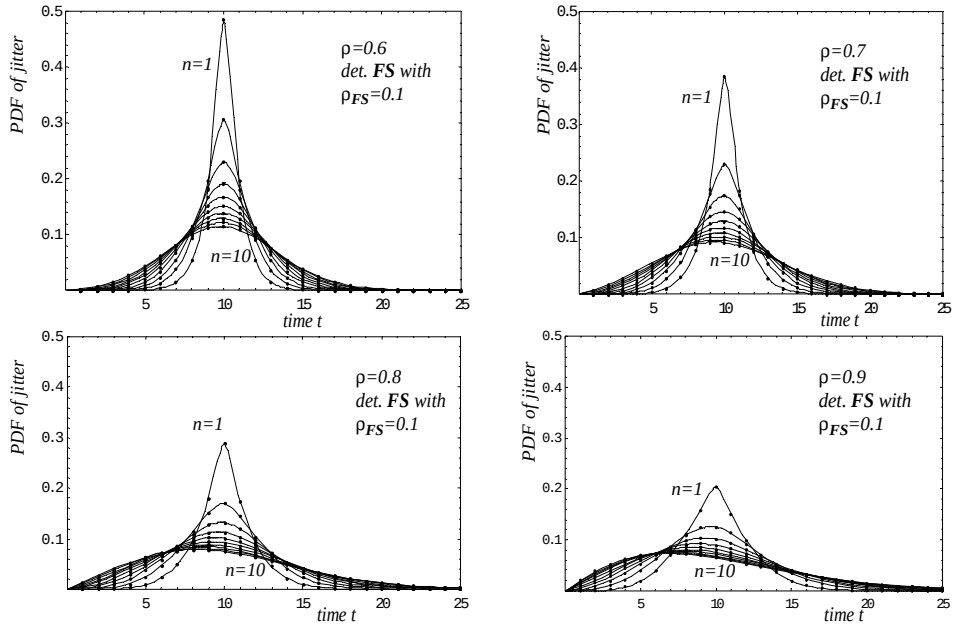

*Figure 9.11: PDF of the jitter as function of time for increasing number of nodes and deterministic distributed **FS** and **R**(random) queueing discipline.*

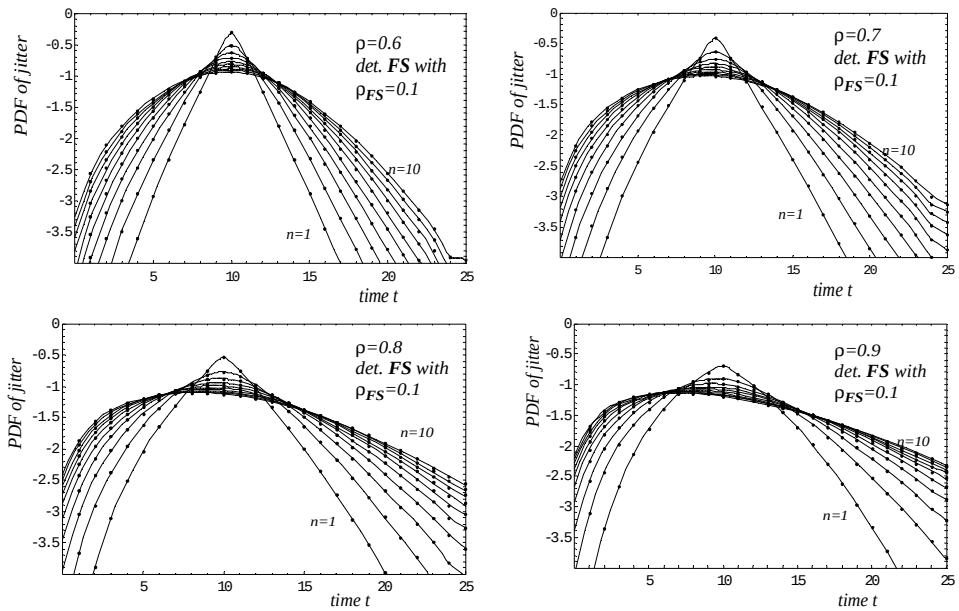

*Figure 9.12: Logarithmic plot of PDF of the jitter as function of time for increasing number of nodes and deterministic distributed **FS** and **R**(random) queueing discipline.*

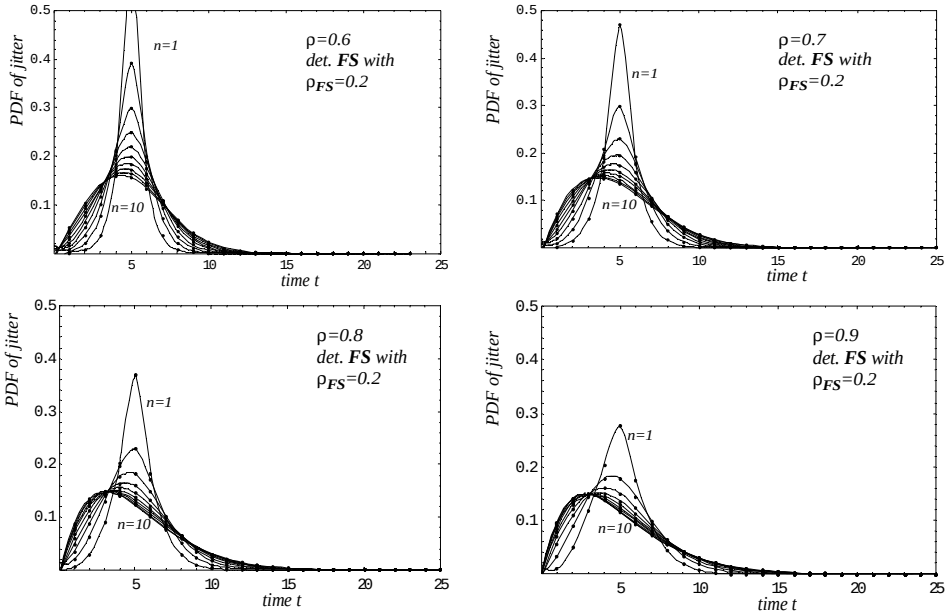

*Figure 9.13: PDF of the jitter as function of time for increasing number of nodes and deterministic distributed **FS** and **R**(random) queueing discipline.*

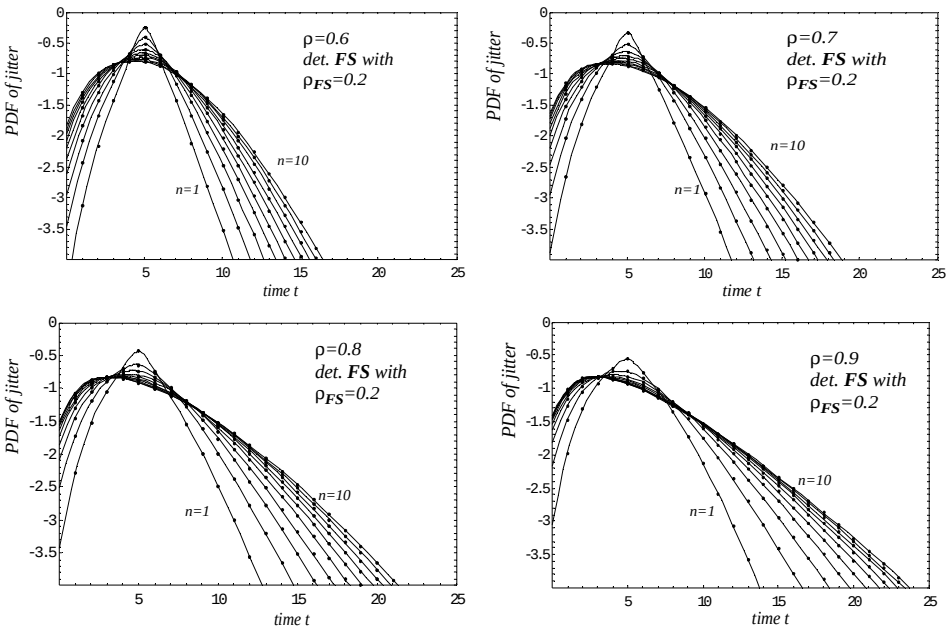

*Figure 9.14: Logarithmic plot of PDF of the jitter as function of time for increasing number of nodes and deterministic distributed **FS** and **R**(random) queueing discipline.*

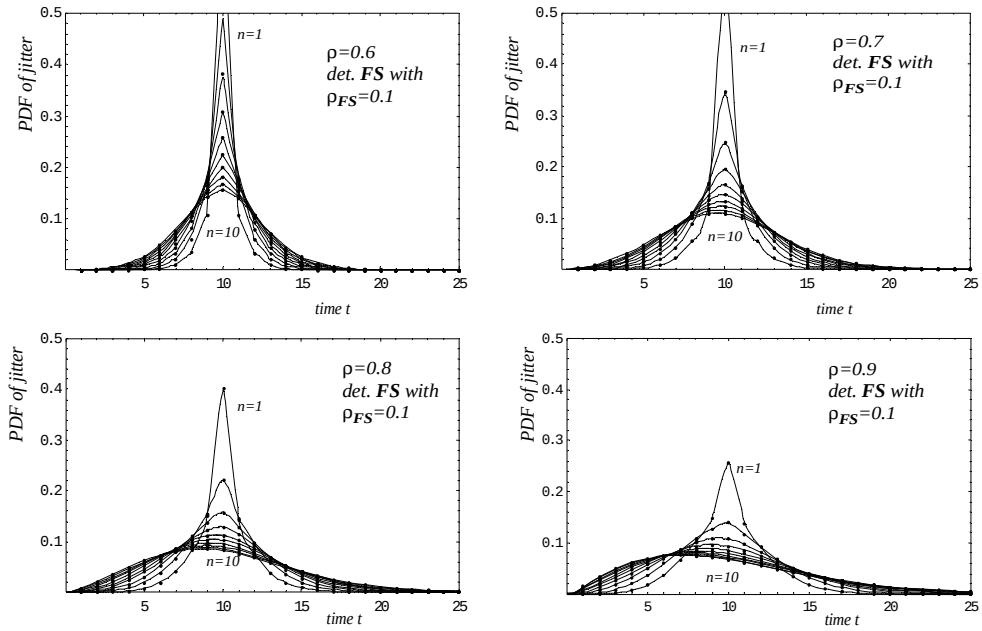

*Figure 9.15: PDF of the jitter as function of time for increasing number of nodes and deterministic distributed **FS** and **F**(first) queueing discipline.*

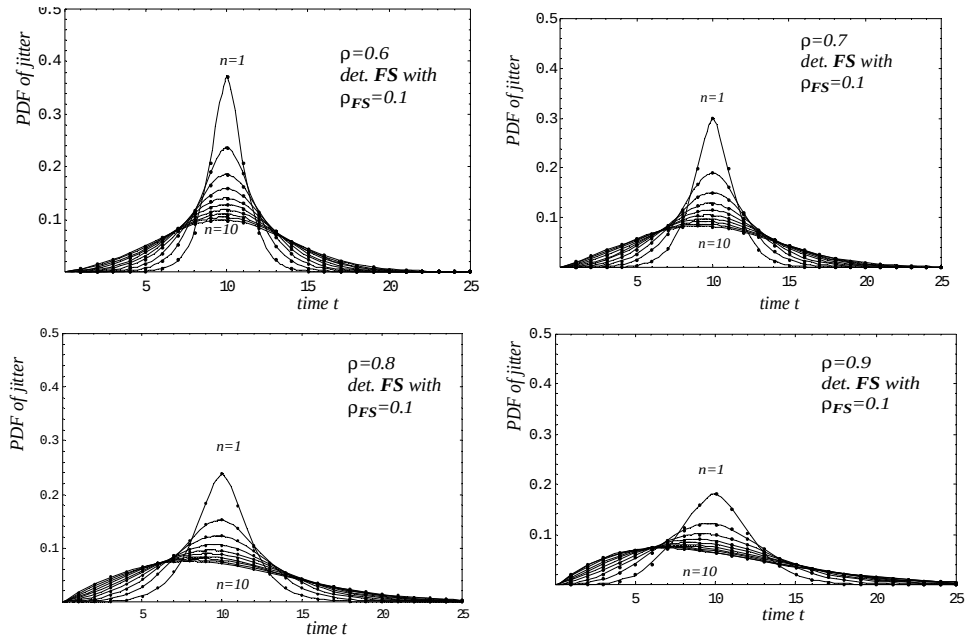

*Figure 9.16: PDF of the jitter as function of time for increasing number of nodes for deterministic distributed **FS** and **L**(last) queueing discipline.*

In figure 9.11 we have depicted all the PDFs of a deterministic **FS** with load 0.1**,** from node 1 up to the exit on node 10, using the **R**(random) queueing discipline. It is interesting to observer the relative strong impact from the load. For low load the jitter is quite narrow for just a few nodes but it gets broader as more nodes are passed. Another interesting observation is that it seems that the PDFs will converge to a limiting distribution as the numbers of node increases. This is already well known for results for chains of saturated queues, (see [Robe96] where such models are discussed). The convergence is especially visible in the logarithmic plots of figure 9.12.

If we increase the load from the **FS,** the mean inter-arrival time will decrease and the corresponding jitter will be more asymmetrical. This effect is clearly seen in figure 9.13 and the logarithmic counterpart figure 9.14. Also in this case the convergence seems to be quite rapid, say at around 10 nodes.

In figure 9.15 and figure 9.16 we have depicted the PDF of the jitter for the **F**(first) and the **L**(last) queueing discipline for deterministic **FS** of load 0.1. For the **F**(first) queueing discipline the jitter with low load is very narrow and it broadens slowly. In this case there will be only minor disturbance from the **BS** and in this case the jitter is also quite narrow at the first few nodes also for loads up to load 0.7. For the **L**(last) queueing discipline, however, the jitter is quite broad also at 0.6 when the stream has passed two or more nodes.

9.8 Concluding remarks

The methods proposed in this chapter show that it is possible to obtain analytical results for quite complicated models, even more important, to obtain numerical results from them. The aim has been to go beyond the assumption of product form solutions that were proposed in chapter 7. The proposed models have the advantage that it is recursive, the output from a queueing node constitutes the input to the next one, and in this way the end-to-end view is kept, and the changes of a stream from the input to the output are an important part of the analysis.

Based on the numerical results we feel confident that the convolution approach of chapter 7 provides a real upper bound for the end-to-end delay for sources that emit a deterministic packet stream. Secondly when it comes to the end-to-end delay the differences in the results between these two types of models seem to be minor.

By this model we have analysed the evolution of the jitter as a deterministic packet stream passes through a series of queues. If all the nodes are identical we have also demonstrated that the jitter will converge to a given probability distribution.

Bibliography

[Abat92] J. Abate, W. Whitt, The Fourier-series method for inverting transform of probability distributions, *Queueing System* 10, 1992, 5-88.

[Abra70] M. Abramowitz, I. A. Stegun eds., *Handbook of mathematical functions*, Dover Publications, Inc., New York, 1970.

[Anic82] D. Anick, D. Mitra, M. M. Sondhi, Stochastic theory of a data-handling system with multiple sources, *The Bell System Technical Journal*, 61(8), 1982, 1871-1894.

[Barb92] A. Barbe, A level-crossing-based scaling dimensionality transform applied to stationary Gaussian processes, *IEEE trans. on information theory* 38(2), 1992, 814-823.

[Bert92] D. Bertsekas, R. Gallager, *Data networks*, Prentice-Hall, Inc., 1992.

[Blan89] J. L. Blanchard, H. Newman, Numerical evaluation of parabolic cylinder function, *IEEE trans. on antennas and propagation*, 37(4), 1989, 519-523.

[Bric97] F. Brichet, L. Massoulié, J. Roberts, Stochastic ordering and the notion of negligible CDV, in *Proceedings ITC 15*, Washington DC, USA, June 22-27, 1997, 1433-1444.

[Buza70] J. A. Buzacott, Markov approach to finding failure times of repairable systems, *IEEE trans. on Reliability*, 19(4), 1970, 128-133.

[Chow88] Y. S. Chow, H. Teicher, *Probability theory, independence, interchangeability, martingales*, Springer-Verlag, 1988.

[Cinl75] E. Cinlar, *Introduction to stochastic processes*, Prentice-Hall, Inc., 1975.

[Codd55] E. A.Coddington, N. Levinson, *Theory of ordinary differential equations*, McGraw-Hill Book Company, Inc., 1955.

[Cohe82] J. W. Cohen, *The single server queue*, North-Holland Publishing Company, 1982.

[Coop72] R. B. Cooper, Introduction to queueing theory, The Macmillan Company, Collier-Macmillan, Ltd., 1972.

[Cox70] D. R. Cox, H. D. Miller, *The theory of stochastic processes*, Willam Clowes and Sons, Ltd., 1970.

[Cram46] H. Cramer, *Mathematical methods of statistics*, Princeton University Press, 1946.

[Crov96] Crovella M. E., Bestavros, A., Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. In *Proc. 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, May 1996.

[Dupu97] A. Dupuis, F. Guillemin, B. Sericola, Asymptotic results for the superposition of large number of data connections on an ATM link, *Queuing Systems* 26, 1997, 121-150.

[Fell68a] W. Feller, *An introduction to probability theory and its applications, Vol. I*, John Wiley & Sons, 1968.

[Fell68b] W. Feller, *An introduction to probability theory and its applications, Vol. II*, John Wiley & Sons, 1968.

[Gned68] B. V. Gnedenko, I. N. Kovalenko, *Introduction to queueing theory*, Israel Program for Scientific Translations, 1968.

[Grad94] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products, Fifth Edition*, Academic Press, Inc. San Diego, CA., 1994.

[Guill95] F. Guillemin, A. Simonian, Transient characteristics of an M/M/∞ system, *Adv. Appl. Prob.* 27, 1995, 862-888.

[Guill96] F. Guillemin, R. Mazumdar, A. Simonian, On heavy traffic approximations for transient characteristics of M/M/∞, *J. Appl. Prob.* 33, 1996, 490-506.

[Guill98] F. Guillemin, D. Pinchon, On the area swept under the occupation process of an M/M/1 queue in busy period, *Queuing Systems* 29 ,1998, 383-398.

[Guil99] F. Guillemin, D. Pinchon, On a random variable associated with excursions in an M/M/∞ system, *Queuing Systems* 32, 1999, 305-318.

[Haga89] P. S. Hagan, C. R. Doering, C. D. Levermore, Mean exit time for particles driven by weakly colored noice, *Siam J. Appl. Math.* 49(5), 1989, 1480-1513.

[Jais68] N. K. Jaiswal, *Priority queues*, Academic Press, Inc., 1968.

[Jean98] A. Jean-Marie, Z. Liu, P. Nain, D. Towsley, Computational aspects of the workload distribution in the MMPP/GI/1 queue, *IEEE J. on selected areas in communications*, 16(5) 1998, 640-652.

[Kala94] V. V. Kalashnikov, Mathematical methods in ueueing theory, Kluwer Academic Publishers, 1994.

[Karl66] S. Karlin, *A first course in stochastic processes*, Academic Press, Inc., 1966.

[Klam73] G. Klambauer, *Real analysis*, American Elsevies Publishing Company, Inc., 1973.

[Klei76a] L. Kleinrock, *Queueing Systems, Volume I: Theory*, New John Wiley & Sons, New York, 1976.

[Klei76b] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications*, New John Wiley & Sons, New York, 1976.

[Knes91] C. Knessl, J. A. Morrison, Heavy-traffic analysis of a data-handling system with many sources, *Siam J. appl. math.*, 51(1), 1991, 187-213.

[Kost73] L. Kosten, *Stochastic theory of service systems*, Pergamon Press, 1973.

[ITU02a] ITU-T, *Network Performance Objectives for IP-based Services*, ITU-T Recommendation Y.1541, Geneva, February 2002, (02/2002).

[ITU02b] ITU-T Rec. G.114, *One-way transmission time*, June 2002.

[Lars79a] H. J. Larson, B. O. Shubert, *Probabilistic models in engineering science, Volume I, random variables and stochastic processes*, John Wileys & Sons, 1979.

[Lars79b] H. J. Larson, B. O. Shubert, *Probabilistic models in engineering science, Volume II, random noise, signals and dynamic systems*, John Wileys & Sons, 1979.

[Lead83] R. M. Leadbetter, G. Lindgren, H. Rootzen, *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York, 1983, 146-172.

[Long58] M. S. Longuet-Higgins, On the intervals between successive zeros of a random function, *Proc. Roy. Soc. A*, 246, 1958, 99-118.

[Long62] M. S. Longuet-Higgins, The distribution of intervals between zeros of a stationary random function, *Phil. Trans. Roy. Soc. A*, 256, 1962, 557-599.

[Lela93] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the self-similar nature of Ethernet traffic, In *Proc. ACM SIGCOMM'93*, 183-193,1993.

[Lela94] W. E. Leland, M. S.Taqqu, W. Willinger, D. V. Wilson, On the self-similar nature of Ethernet traffic (extended version), *IEEE/ACM Trans. on Networking*, 2, 1994, 1-15.

[Mand99] M. Mandjes, K. van der Wal, R. Kooij, H. Bastiaansen, End-to-end delay models for interactive services in a large-scale IP network, in *Seventh IFIP workshop on Performance Modelling and Evaluation of ATM Networks: IFIP ATM'99*, Antwerp, Belgium, June 28-30, 1999.

[Marc77] M. B. Marcus, Level crossing of a stochastic process with absolutely continuous sample paths, *Annals of Prob.*, 5(1), 1977, 52-71¨

[Matr94a] W. Matragi, K. Sohraby, C. Bisdikian, Jitter calculus in ATM Networks: Single node case, in *Proc. IEEE INFOCOM'94*, Toronto, Ont., Canada, June 12-16, 1994.

[Matr94b] W. Matragi, K. Sohraby, C. Bisdikian, Jitter calculus in ATM Networks: Multiple node case, in *Proc. IEEE INFOCOM'94*, Toronto, Ont.,Canada, June 12-16, 1994.

[McFa56] J. A. McFadden, The axis-crossing intervals of random functions, *IRE Trans. on Information Theory*, 2, 1956, 146-150.

[McFa58] J. A. McFadden, The axis-crossing intervals of random functions-II, *IRE Trans. on Information Theory*, 4, 1958, 14-24

[Medh91] J. Medhi, *Stochastic models in queueing theory*, Academic Press, Inc., 1991.

[Midd60] D. Middleton, *An Introduction to statistical communication theory*, McGraw-Hill Book Company, Inc., 1960

[Neut81] M. F. Neuts, *Matrix-geometric solutions in sthochastic models*, The Johns Hopkins University Press, 1981,

[Olve59] F. W. J. Olver, Uniform asymptotic expansions for Weber parabolic cylinder functions of large orders, *J. of research of the National Bureau of Standars-B. Mathematics and Mathematical Physics*, 63B(2), 1959 131-169.

[Paxs95] Paxson, V., Floyd, S. Wide Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Trans. on Networking*, 3 (3), 1995 226-244.

[Park00] K. Park, W. Willinger, eds., *Self-similar network traffic and performance evaluation*, John Wiley & Sons, Inc., 2000.

[Papo65] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw-Hill, Inc., 1965.

[Rain87] A. J. Rainal, First and second passage times of Rayleigh processes, *IEEE trans. on information theory*, 33(3), 1987, 419-425.

[Rain90] A. J. Rainal, Passage times of Gaussian noise crossing a time-varying boundary, *IEEE trans. on information theory*, 36(5), 1990, 1179-1183.

[Ren98] Q. Ren, H. Kobayashi, Diffusion approximation modelling for Markov modulated bursty traffic and its applications to bandwidth allocation in ATM networks, *IEEE J. on selected areas in communications* 16(5), 1998, 679-691.

[Royd68] H. R. Royden, *Real analysis*, The Macmillan Company, Collier-Macmillan Ltd., London 1968.

[Simo91a] A. Simonian, Stationary analysis of a fluid queue with input rate varying as an Ornstein-Uhlenbeck process, *Siam J. appl. math.*, 51(3), 1991, 828-842.

[Simo91b] A. Simonian, J. Virtamo, Transient and stationary distributions for fluid queues and input processes with a density *Siam J. appl. math.*, 51(6), 1991, 1732-1739.

[RFC1633] RFC 1633, *Integrated Service in the Internet Architecture: an Overview*, IETF, 1994.

[RFC2212] RFC 2212, *Specification of Guaranteed Quality of Service*, IETF, September 1997.

[RFC2475] RFC 2475, *An Architecture for Differntiated Service*, IETF, 1998.

[RFC2598] RFC 2598, *An Expedited Forwarding PHB*, IETF, June 1999.

[Rice45] S. O. Rice, Mathematical analysis of random noice, *The Bell System Technical Jornal*, 24, 1945, 47-156.

[Rice48] S. O. Rice, Statistical properties of a sine wave plus random noice, *The Bell System Technical Jornal*, 27, 1948, 109-157.

[Robe96] J, Roberts, U. Mocci, J. Virtamo eds., *Broadband Network Teletraffic. Final Report of Action COST 242*, Springer-Verlag, Berlin, 1996.

[Robe92] J, Roberts, ed., *Performance evaluation and design of multiservice network, Final Report of Action COST 224,* Commision of the European Communities, 1992.

[Schw87] M. Schwartz, *Telecommunication networks, protocols, modelling and anlysis*, Addison-Wesley Publishing Company, 1987.

[Saat61] T. L. Saaty, *Elements of queueing theory with applications*, McGraw-Hill Book Company, 1961.

[Saue81] C. H. Sauer, K. M. Chandy, *Computer systems performance modelling*, Prentice-Hall, Inc., 1981.

[Shal84] M. Shalmon, M. A. Kaplan, A tandem network of queues with deterministic service and intermediate arrivals, *Oper. Research* 32(4), 1984, 753-773.

[Shwa95] A. Shwartz, A. Weiss, *Large deviations for performance analysis*, Chapman & Hall, 1995.

[Ster91] T. E. Stern, A. I. Elwalid, Analysis of separable Markov-modulated rate models for information-handling systems, *Adv. Appl. Prob.* 23, 1991, 105-139.

[Taka91] H. Takagi, *Queueing Analysis, Volume 1: Vacation and Priority Systems, Part 1,* North-Holland, Amsterdam, 1991.

[Vlee99] D. De Vleeschauwer, J. Janssen, G. H. Petit, Voice over IP in Access Networks, in *Seventh IFIP workshop on Performance Modelling and Evaluation of ATM Networks: IFIP ATM'99*, Antwerp, Belgium, June 28-30, 1999.

[Vlee00] D. De Vleeschauwer, G. H. Petit, B. Steyaert, S. Wittevrongel, H. Bruneel, An accurate closed-form formula to calculate the dejittering delay in packetised voice transport, in *Networking 200, IFIP-TC6*, Paris, France, May 14-19, 2000.

[Wilk65] J. H. Wilkinson, *The algebraic eigenvalue problem*, Oxford University Press, 1965, 299-307.

[Won89] R. Wong, *Asymptotic Approximations of Integrals*, Academic Press, Inc. San Diego, CA, 1989.

[Zahl84] U. Zahle, A general Rice formula, Palm measures, and horizontal-window conditioning for random fields, *Stochastic Processes and their Applications* 17, 1984, 265-283.

[Øste91] O. Østerbø, Duration of heavy load states in an ATM network, in *Proceedings ITC 13 workshop*, Copenhagen, Denmark, June 6-10, 1991, 461-470.

[Øste94] O. Østerbø, The effect of shaping or phasing of indevidual VCs in ATM networks, in *Proceedings ITC 14*, Antibes Juan-les-Pins, France, June 19-26, 1994, 91-95.

[Øste95a] O. Østerbø, Some important queueing models for ATM, *Telektronikk 2/3*, 1995, 208-219.

[Øste95b] O. Østerbø, *Some important queueing models for ATM*, Telenor Research Document serie, R&D R 22/1995.

[Øate01a] O. Østerbø, IP-multiplexing for low capacity links?, *Telektronikk 2/3*, 2001, 159-169.

[Øate01b] O. Østerbø, *IP-multiplexing for low capacity links?*, Telenor Research Document serie, R&D R 35/2001.

[Øste02] O. Østerbø, *An approximative method to calculate the distribution of end-to-end delay in packet network*, Telenor Research Document serie, R&D R 3/2002.

[Øste03a] O. Østerbø, *End-to-end queuing models with priority*, Telenor Research Document serie, R&D R 3/2003.

[Øste03b] O. Østerbø, *Models for end-to-end delay in packet networks*, Telenor Research Document serie, R&D R 4/2003.

[Øste03c] O. Østerbø, Models for calculated end-to-end delay in packet networks, in *Proceedings ITC 15*, Berlin, Germany, August 31-September 5, 2003.

Appendix A

Crossing intensities and the joint probability of the excess volume and the excess time that starts in $(0,dt_1)$ and ends in $(t,t+dt_2)$

A.1 Some general results on the level crossing intensity for stationary stochastic processes

Theorem A.1. Let $\{\boldsymbol{B}_t\}$ be a stationary stochastic process and let $\Delta_C^{uc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 \le C < \boldsymbol{B}_t\}$ and $\Delta_C^{dc}(t) = \frac{1}{t}\boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\}$ be the up- and down crossing rates:

$$\text{Then } \Delta_C(t) = \Delta_C^{uc}(t) = \Delta_C^{dc}(t) \tag{A.1}$$

and the following inequality yields:

$$\Delta_C(t) \le \gamma\Delta_C(\gamma t) + (1-\gamma)\Delta_C((1-\gamma)t) \text{ for all } t \text{ and } 0 \le \gamma \le 1 \tag{A.2}$$

Further if $\lim_{t\to 0}\sup\Delta_C(t) < \infty$ then the up- and down-crossing intensity $\Delta_C = \lim_{t\to 0}\Delta_C(t)$ exists and is finite.

Proof: We have the following relation between the events:

$$\{\boldsymbol{B}_0 \le C < \boldsymbol{B}_t\} = \{\boldsymbol{B}_0 \le C\} - \{\boldsymbol{B}_0 \le C, \boldsymbol{B}_t \le C\} = \{\boldsymbol{B}_t > C\} - \{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\}$$

and similar

$$\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\} = \{\boldsymbol{B}_0 > C\} - \{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\} = \{\boldsymbol{B}_t \le C\} - \{\boldsymbol{B}_0 \le C, \boldsymbol{B}_t \le C\}.$$

We must therefore have:

$$\boldsymbol{P}\{\boldsymbol{B}_0 \le C < \boldsymbol{B}_t\} = P(\boldsymbol{B}_0 \le C) - P(\boldsymbol{B}_0 \le C, \boldsymbol{B}_t \le C) = P(\boldsymbol{B}_t \le C) - P(\boldsymbol{B}_0 \le C, \boldsymbol{B}_t \le C) = \boldsymbol{P}\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_t\}$$

showing that $\Delta_C^{uc}(t) = \Delta_C^{dc}(t)$, since $P(\boldsymbol{B}_0 \le C) = P(\boldsymbol{B}_t \le C)$ due to the stationary assumption.

Let $0 \le \gamma \le 1$. To show (A.2) we start with the obvious inequality:

$P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C, \boldsymbol{B}_t > C\} \le P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_t > C\}$ giving

$$t\Delta_C(t) \le P\{\boldsymbol{B}_0 > C\} - P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C, \boldsymbol{B}_t > C\} =$$

$$P\{\boldsymbol{B}_0 > C\} - P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C\} + P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C\} - P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C, \boldsymbol{B}_t > C\} =$$

$$P\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_{\gamma t}\} + P\{\boldsymbol{B}_0 > C, \boldsymbol{B}_{\gamma t} > C, \boldsymbol{B}_t \le C\} \le P\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_{\gamma t}\} + P\{\boldsymbol{B}_{\gamma t} > C, \boldsymbol{B}_t \le C\}$$

$P\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_{\gamma t}\} + P\{\boldsymbol{B}_0 > C \ge \boldsymbol{B}_{(1-\gamma)t}\} = (t\gamma\Delta_C(\gamma t) + (1-\gamma)t\Delta_C((1-\gamma)t))$; or by dividing by t we get $\Delta_C(t) \le \gamma\Delta_C(\gamma t) + (1-\gamma)\Delta_C((1-\gamma)t)$.

To show the last part we set $M = \limsup_{t \to 0} \Delta_C(t)$ and let $m = \liminf_{t \to 0} \Delta_C(t)$. If $M < \infty$ then we shall show that $m = M$. By choosing $\gamma = \frac{1}{2}$ it follows that $\Delta_C(t) \le \Delta_C\left(\frac{t}{2}\right)$ for all t and by continuously sub-dividing the interval $\Delta_C(t) \le \Delta_C\left(\frac{t}{2}\right) \le \Delta_C\left(\frac{t}{2^2}\right) \le \ldots$. By the monotone convergence theorem $\lim_{n \to \infty} \Delta_C\left(\frac{t}{2^n}\right)$ exists for every t and are all finite (by the assumption). We shall assume $m < M$ and show that this assumption will lead to a contradiction. We choose $0 < \varepsilon < \frac{M-n}{12}$ and $\dot{t}$ so that $\Delta_C(t) < M + \varepsilon$ for $t < \dot{t}$ and a $t' < \dot{t}$ so that $\Delta_C(t') < M - \varepsilon$. Now we choose $\tilde{t} < \frac{t'}{8}$ so that $\Delta_C(\tilde{t}) < m + \varepsilon$, and we choose an integer n such that $\frac{t'}{2^{n+1}} < \tilde{t} \le \frac{t'}{2^n}$ and $t = \frac{t'}{2^{n-1}}$. Since $\Delta_C\left(\frac{t'}{2^n}\right)$ is an increasing sequence (in n), we have $\frac{t}{4} < \tilde{t} \le \frac{t}{2}$ and $\Delta_C(t) < M - \varepsilon$. We now choose γ so that $\gamma t = \tilde{t}$ and $\frac{t}{4} < \gamma t \le \frac{t}{2}$ or $\frac{1}{4} < \gamma \le \frac{1}{2}$. Now by using the inequality $\Delta_C(t) \le \gamma\Delta_C(\gamma t) + (1-\gamma)\Delta_C((1-\gamma)t)$, we have

$M - \varepsilon \le \gamma(m + \varepsilon) + (1-\gamma)\Delta_C((1-\gamma)t)$ or

$\Delta_C((1-\gamma)t) \geq \dfrac{M-\varepsilon-\gamma(m+\varepsilon)}{1-\gamma} = M+\nu\Big)$, where (since $\frac{1}{4} < \gamma \leq \frac{1}{2}$)

$\nu = \dfrac{\gamma(M-n)}{1-\gamma} - \dfrac{1+\gamma}{1-\gamma}\varepsilon \geq \dfrac{M-n}{3} - 3\varepsilon > \varepsilon$ which is a contradiction.

QED.

Theorem A.2. We let $\{\boldsymbol{B}_t\}$ be a stationary stochastic process with continuous sample paths and we assume that the distribution of $\boldsymbol{B}_0$ is absolute continuous (which means that the probability density function exists and is a continuous function) and let $\boldsymbol{dB}_t = \dfrac{\boldsymbol{B}_t - \boldsymbol{B}_0}{t}$ be the differential process scaled by $1/t$ where we consider crossings of a given level C. Set $F_{1_t}(y,z) = P(\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = z)\varphi(z)$ and $F_{2_t}(y,z) = P(-\boldsymbol{dB}_t > y \,|\, \boldsymbol{B}_0 = z)\varphi(z)$ where $\varphi(z)$ is the probability density function of $\boldsymbol{B}_0$. We shall make the following assumptions:

A. We assume that $F_{1_t}(y,z)$ and $F_{2_t}(y,z)$ satisfy the following conditions:

$\left|F_{1_t}(y,C) - F_{1_t}(y,x)\right| \leq M_t(y)|C-x|$ and $\left|F_{2_t}(y,C) - F_{2_t}(y,x)\right| \leq M_t(y)|C-x|$ (for $y \geq 0$)

where $M_t = \int_{y=0}^{\infty} yM_t(y)dy$ exists and further $\lim_{t\to 0} tM_t = 0$

If the limit $\Delta_C = \lim_{t\to 0}\Delta_C^{uc}(t) = \lim_{t\to 0}\Delta_C^{dc}(t)$ is finite, then the up and down crossing intensity is given as:

$$\Delta_C = \lim_{t\to 0}\Delta_C^{uc}(t) = \lim_{t\to 0}\Delta_C^{dc}(t) = \frac{\varphi(C)}{2}\lim_{t\to 0}\boldsymbol{E}[\,|\boldsymbol{dB}_t|\,|\boldsymbol{B}_0 = C] \tag{A.3}$$

Proof: We have $\Delta_C^{uc}(t) = \Delta_C^{dc}(t) = \frac{1}{2}(\Delta_C^{uc}(t) + \Delta_C^{dc}(t))$ where the up crossing intensity $\Delta_C^{uc}(t) = \frac{1}{t}P(\boldsymbol{B}_0 \leq C < \boldsymbol{B}_t)$ and down crossing intensity $\Delta_C^{dc}(t) = \frac{1}{t}P(\boldsymbol{B}_0 \geq C > \boldsymbol{B}_t)$. By conditioning on $\boldsymbol{B}_0$ we may write the up and down crossing intensities as:

$$\Delta_C^{uc}(t) = \int_{y=0}^{\infty} F_{1_t}(y, C-ty)dy \text{ and} \tag{A.4}$$

$$\Delta_C^{dc}(t) = \int_{y=0}^{\infty} F_{2_t}(y, C+ty)dy. \tag{A.5}$$

We also define the integrals $J_C^{uc}(t) = \int_{y=0}^{\infty} F_{1_t}(y, C)dy$ and $J_C^{dc}(t) = \int_{y=0}^{\infty} F_{2_t}(y, C)dy$ and we have:

$$J_C^{uc}(t) + J_C^{dc}(t) = \varphi(C) \int_{y=0}^{\infty} P(|\boldsymbol{dB}_t| > y | \boldsymbol{B}_0 = C)dy = \varphi(C)\boldsymbol{E}[|\boldsymbol{dB}_t| | \boldsymbol{B}_0 = C] \qquad \text{(A.6)}$$

By applying (A.4), (A.5) and (A.6) we obtain

$$\left|\Delta_C^{uc}(t) - \frac{1}{2}\varphi(C)\boldsymbol{E}[|\boldsymbol{dB}_t| | \boldsymbol{B}_0 = C]\right| = \frac{1}{2}\left|(\Delta_C^{uc}(t) - J_C^{uc}(t)) + (\Delta_C^{dc}(t) - J_C^{dc}(t))\right| \leq$$

$\frac{1}{2}\left|\int_{y=0}^{\infty} |F_{1_t}(y, C - ty) - F_{1_t}(y, C)|dy + \int_{y=0}^{\infty} |F_{2_t}(y, C + ty) - F_{2_t}(y, C)|dy\right| \leq tM_t$. Since by assumption $\lim_{t \to 0} tM_t = 0$ it follows that $\lim_{t \to 0} \Delta_C^{uc}(t) = \frac{\varphi(C)}{2} \lim_{t \to 0} \boldsymbol{E}[|\boldsymbol{dB}_t| | \boldsymbol{B}_0 = C]$. QED

A.2 Evaluating and expanding the joint probability of the excess volume and the excess time that starts in (0,dt$_1$) and ends in (t,t+dt$_2$)

In this section we derive an expression for the joint probability of the event that the excess volume that lies in the interval $(y, y + dy)$ and that the excess time starts in $(0, dt_1)$ and ends in $(t, t + dt_2)$ for small dt_1 and dt_2. By decomposing this event into four parts we may write this probability as:

$$P(A_t \in (z, z + dz), \boldsymbol{B}_0 \leq C, Inf_{\tau \in (dt_1, t)}\boldsymbol{B}_\tau > C, \boldsymbol{B}_{t + dt_2} \leq C) =$$
$$P(A_t \in (z, z + dz), Inf_{\tau \in (dt_1, t)}\boldsymbol{B}_\tau > C) - P(A_t \in (z, z + dz), Inf_{\tau \in (0, t)}\boldsymbol{B}_\tau > C)$$
$$-P(A_t \in (z, z + dz), Inf_{\tau \in (dt_1, t + dt_2)}\boldsymbol{B}_\tau > C) + P(A_t \in (z, z + dz), Inf_{\tau \in (0, t + dt_2)}\boldsymbol{B}_\tau > C)$$
$$= I_1 - I_2 - I_3 + I_4.$$

We use a similar approach to evaluate each of these parts as we used for the conditional excess times and excess volume described in chapter 3. For I_2 we find:

$$I_2 = P(A_t \in (z, z + dz), Inf_{\tau \in (0, t)}\boldsymbol{B}_\tau > C) = \int_{y=C}^{\infty} \int_{x=C}^{\infty} g_C(x, y, z, t)dxdy \cdot dz \qquad \text{(A.7)}$$

To evaluate $I_1 = P(A_t \in (z, z + dz), Inf_{\tau \in (dt_1, t)}\boldsymbol{B}_\tau > C)$ we relates the volume A_t to

$A_t^{dt_1} = \int_{\tau = dt_1}^{t} (\boldsymbol{B}_\tau - C)d\tau = A_t - (\boldsymbol{B}_{dt_1} - C)dt_1$ for small dt_1 so therefore the event $A_t \in (z, z + dz)$ gives $A_t^{dt_1} \in (z - (\boldsymbol{B}_{dt_1} - C)dt_1, z - (\boldsymbol{B}_{dt_1} - C)dt_1 + dz)$. Then by conditioning on $\boldsymbol{B}_{dt_1} = x$ and $\boldsymbol{B}_t = y$ and integrating, and applying the assumption that the rate process $\{\boldsymbol{B}_t\}$ is stationary we find that I_1 may be written:

$$I_1 = \int_{y = C}^{\infty} \int_{x = C}^{\infty} g_C(x, y, z - (x - C)dt_1, t - dt_1)dxdy \cdot dz \tag{A.8}$$

Continuing by evaluating $I_3 = P(A_t \in (z, z + dz), Inf_{\tau \in (dt_1, t + dt_2)} \boldsymbol{B}_\tau > C)$ we relate the volume A_t to $A_{t + dt_2}^{dt_1} = \int_{\tau = dt_1}^{t + dt_2} (\boldsymbol{B}_\tau - C)d\tau = A_t - (\boldsymbol{B}_{dt_1} - C)dt_1 + (\boldsymbol{B}_{dt_2} - C)dt_2$ for small dt_1 and dt_2. Therefore the event $A_t \in (z, z + dz)$ gives $A_{t + dt_2}^{dt_1} \in (z - (\boldsymbol{B}_{dt_1} - C)dt_1 + (\boldsymbol{B}_{t + dt_2} - C)dt_2, z - (\boldsymbol{B}_{dt_1} - C)dt_1 + (\boldsymbol{B}_{t + dt_2} - C)dt_2 + dz)$. By conditioning on $\boldsymbol{B}_{dt_1} = x$ and $\boldsymbol{B}_{t + dt_2} = y$ and integrating, and applying the assumption that the rate process $\{\boldsymbol{B}_t\}$ is stationary we find that I_3 may be written:

$$I_3 = \int_{y = C}^{\infty} \int_{x = C}^{\infty} g_C(x, y, z - (x - C)dt_1 + (y - C)dt_2, t - dt_1 + dt_2)dxdy \cdot dz \tag{A.9}$$

Finally to evaluate $I_4 = P(A_t \in (z, z + dz), Inf_{\tau \in (0, t + dt_2)} \boldsymbol{B}_\tau > C)$ we relate the volume A_t by $A_{t + dt_2}^0 = \int_{\tau = 0}^{t + dt_2} (\boldsymbol{B}_\tau - C)d\tau = A_t + (\boldsymbol{B}_{dt_2} - C)dt_2$ for small dt_2. Therefore the event $A_t \in (z, z + dz)$ gives $A_{t + dt_2}^0 \in (z + (\boldsymbol{B}_{t + dt_2} - C)dt_2, z + (\boldsymbol{B}_{t + dt_2} - C)dt_2 + dz)$. By conditioning on $\boldsymbol{B}_0 = x$ and $\boldsymbol{B}_{t + dt_2} = y$ and integrating, we find that I_4 may be written:

$$I_4 = \int_{y = C}^{\infty} \int_{x = C}^{\infty} g_C(x, y, z + (y - C)dt_2, t + dt_2)dxdy \cdot dz \tag{A.10}$$

Then by expanding the four integrals above to second order for small dt_1 and dt_2 and collecting we get:

$P(A_t \in (z, z + dz), \boldsymbol{B}_0 \le C, Inf_{\tau \in (dt_1, t)} \boldsymbol{B}_\tau > C, \boldsymbol{B}_{t + dt_2} \le C) =$

$$\int_{y=C}^{\infty}\int_{x=C}^{\infty}\left((y-C)(x-C)\frac{\partial^2 g_C}{\partial z^2}+(x+y-2C)\frac{\partial^2 g_C}{\partial z\partial t}+\frac{\partial^2 g_C}{\partial t^2}\right)dxdy\cdot dzdt_1dt_2 \qquad \text{(A.11)}$$

Appendix B

Some important properties for multinormal integrals

This appendix is devoted to integrals over multinormal distributions. Such types of integrals will show up during the study of normal stochastic processes, and will provide an important tool in the effort to gain knowledge of important properties of such processes. The main findings are that multinormal integrals have some nice properties that reduce the complexity in numerical computations. By applying these results we are able to obtain expressions that reduce the number of numerical integrations to **half** the dimension of the integrals. Thus by applying this method we are able to calculate a five dimension multinormal integral by performing two numerical integrations. These simplifications are obtained by taking various partial derivatives with respect to the parameters involved, especially the elements in the Covariance matrix.

Throughout this appendix we shall work with multinormal distributions and we take the following assumptions: We consider n-dimensional multinormal distributed variables $\{\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_n\}$ with zero mean, and covariance matrix $M = (\rho_{ij})$ where we assume that $\rho_{ii} = 1$. (This means that the $\boldsymbol{B}_j$'s are all standard (normalized) normal variables with zero mean and variance equal unity.) If we let $M^{-1} = (M_{ij}^{-1})$ denote the inverse covariance matrix M, then the joint density function for $\{\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_n\}$ is given by:

$$f_n(\xi_1,\ldots,\xi_n;M) = \frac{\sqrt{Det[M^{-1}]}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{n} \xi_i \xi_j M_{ij}^{-1}\right] \tag{B.1}$$

In the succeeding we shall investigate different types of integral derived from (B.1) and see that these integrals can be related by integrals of lower dimensions. When writing (B.1) we should also be aware of the fact that f_n can be considered as a function of the $\frac{n(n-1)}{2}$ covariances ρ_{ij}.

The main contributions throughout this appendix are on integrals based on the multinormal distribution. The various types of integrals considered will be brought into variants of integral of the (standard) form:

$$I = I(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} f_n(\xi_1,...,\xi_n;M) d\xi_n ... d\xi_1 \tag{B.2}$$

where $f_n(\xi_1,...,\xi_n;M)$ is given by (B.1), where we also indicate the dependencies of all the parameters involved:

n-the dimension of the integral

$\boldsymbol{C} = (C_1, ..., C_n)$-is the vector consisting of the integration limits

$M = (\rho_{ij})$-is the covariance matrix which is symmetric and with $\rho_{ii} = 1$

Before giving the main results (on multinormal integrals) we first need some preliminary results mainly on linear algebra and determinants. We put these results in a separate section below.

B.1 Some preliminary results

We first start by giving some preliminary results that we shall apply later on in the analysis.

Lemma B.1. Let $M = (\rho_{ij})$ be a symmetric nonsingular nxn matrix with $\rho_{ii} = 1$ and let $M^{-1} = (M_{ij}^{-1})$ be the corresponding inverse matrix. Then M^{-1} also is symmetric and the partial derivative of the elements in M^{-1} with respect to the elements in M is given by the relation:

$$\frac{\partial M_{ij}^{-1}}{\partial \rho_{rs}} = -(M_{ir}^{-1} M_{js}^{-1} + M_{jr}^{-1} M_{is}^{-1}) \tag{B.3}$$

(If $i = j$, (B.3) reduces to $\frac{\partial M_{ii}^{-1}}{\partial \rho_{rs}} = -2M_{ir}^{-1} M_{is}^{-1}$).

Further if $\rho_{ij} = \rho_{ij}(t)$ are functions of t then

$$\frac{dM_{ij}^{-1}}{dt} = \sum_{r,s=1}^{n} M_{ir}^{-1} M_{js}^{-1} \frac{d\rho_{rs}}{dt} \tag{B.4}$$

Proof: By taking partial derivatives of the relation $M \cdot M^{-1} = I$ we find:

$\frac{\partial}{\partial \rho_{rs}}(M \cdot M^{-1}) = \frac{\partial M}{\partial \rho_{rs}} \cdot M^{-1} + M \cdot \frac{\partial}{\partial \rho_{rs}}(M^{-1}) = 0$ and solving for $\frac{\partial M^{-1}}{\partial \rho_{rs}}$ gives

$\frac{\partial M^{-1}}{\partial \rho_{rs}} = -M^{-1} \cdot \frac{\partial M}{\partial \rho_{rs}} \cdot M^{-1}$. In component form this gives $\frac{\partial M_{ij}^{-1}}{\partial \rho_{rs}} = \sum_{l=1}^{n} \sum_{m=1}^{n} M_{il}^{-1} \frac{\partial \rho_{lm}}{\partial \rho_{rs}} M_{mj}^{-1}$.

The only non zero contribution to the last sum comes when $l = r, m = s$ and $l = s, m = r$ for which the partial derivatives equal unity and this gives (B.3).

If $\rho_{ij} = \rho_{ij}(t)$ are functions of t, then we obtain by applying (B.3):

$\frac{dM_{ij}^{-1}}{dt} = \sum_{1 \le r < s \le n} \frac{\partial}{\partial \rho_{rs}}(M_{ij}^{-1}) \frac{d\rho_{rs}}{dt} = -\sum_{1 \le r < s \le n} (M_{ir}^{-1} M_{js}^{-1} + M_{jr}^{-1} M_{is}^{-1}) \frac{d\rho_{rs}}{dt}$. By using the fact that $\frac{d\rho_{rr}}{dt} = 0$ and that the matrix M is symmetric we get (B.4). QED

When working with multidimensional normal distributions and integrals it is often necessary to change the integration variable, and this may lead to different transformations of the matrices involved. One such transformation we use is to transform the inverse M^{-1} so that the diagonal elements are unity. We denote this transformation by the matrix $\Theta = (\Theta_{ij})$ where

$$\Theta_{ij} = \frac{M_{ij}^{-1}}{\sqrt{M_{ii}^{-1}}\sqrt{M_{jj}^{-1}}} \quad i, j = 1, \ldots, n \tag{B.5}$$

Lemma B.2. The transformation (B.5) is fully symmetric. That is:

$$\Theta_{ii}^{-1} = M_{ii}^{-1} \text{ and } \rho_{ij} = \frac{\Theta_{ij}^{-1}}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} \tag{B.6}$$

Proof: In matrix notation (B.5) reads:

$\Theta = Diag\left(\frac{1}{\sqrt{M_{ii}^{-1}}}\right) \cdot M^{-1} \cdot Diag\left(\frac{1}{\sqrt{M_{ii}^{-1}}}\right)$ which implies

$\Theta^{-1} = Diag(\sqrt{M_{ii}^{-1}}) \cdot M \cdot Diag(\sqrt{M_{ii}^{-1}})$ or in component form

$$\Theta_{ij}^{-1} = \sqrt{M_{ii}^{-1}} \cdot \rho_{ij} \cdot \sqrt{M_{jj}^{-1}}. \tag{B.7}$$

The diagonal elements $i = j$ give $\Theta_{ii}^{-1} = M_{ii}^{-1}$, and by pre and post multiplying with the inverse of the diagonal matrices gives us M as:

$$M = Diag\left(\frac{1}{\sqrt{M_{ii}^{-1}}}\right) \cdot \Theta^{-1} \cdot Diag\left(\frac{1}{\sqrt{M_{ii}^{-1}}}\right) = Diag\left(\frac{1}{\sqrt{\Theta_{ii}^{-1}}}\right) \cdot \Theta^{-1} \cdot Diag\left(\frac{1}{\sqrt{\Theta_{ii}^{-1}}}\right) \tag{B.8}$$

which is the matrix form of (B.6). QED.

Lemma B.3. We have the following relation between the partial derivatives:

$$\frac{\partial \rho_{kl}}{\partial \Theta_{ij}} = \sqrt{M_{ii}^{-1}}\sqrt{M_{jj}^{-1}}[(\rho_{ik}\rho_{jk} + \rho_{il}\rho_{jl})\rho_{kl} - (\rho_{ik}\rho_{jl} + \rho_{jk}\rho_{il})] \tag{B.9}$$

Proof: We have $\frac{\partial \rho_{kl}}{\partial \Theta_{ij}} = \frac{\partial}{\partial \Theta_{ij}}\left(\frac{\Theta_{kl}^{-1}}{\sqrt{\Theta_{kk}^{-1}}\sqrt{\Theta_{ll}^{-1}}}\right) = -\frac{1}{2}\left(\frac{\frac{\partial \Theta_{kk}^{-1}}{\partial \Theta_{ij}}}{\Theta_{kk}^{-1}} + \frac{\frac{\partial \Theta_{ll}^{-1}}{\partial \Theta_{ij}}}{\Theta_{ll}^{-1}}\right)\rho_{kl} + \frac{\frac{\partial \Theta_{kl}^{-1}}{\partial \Theta_{ij}}}{\sqrt{\Theta_{kk}^{-1}}\sqrt{\Theta_{ll}^{-1}}}$. By using (B.3) we get $\frac{\partial \rho_{kl}}{\partial \Theta_{ij}} = \left(\frac{\Theta_{ki}^{-1}\Theta_{kj}^{-1}}{\Theta_{kk}^{-1}} + \frac{\Theta_{li}^{-1}\Theta_{lj}^{-1}}{\Theta_{ll}^{-1}}\right)\rho_{kl} - \frac{\Theta_{ki}^{-1}\Theta_{lj}^{-1} + \Theta_{kj}^{-1}\Theta_{li}^{-1}}{\sqrt{\Theta_{kk}^{-1}}\sqrt{\Theta_{ll}^{-1}}}$, and by applying (B.6) and (B.7) we find $\frac{\partial \rho_{kl}}{\partial \Theta_{ij}} = \sqrt{M_{ii}^{-1}}\sqrt{M_{jj}^{-1}}[(\rho_{ik}\rho_{jk} + \rho_{il}\rho_{jl})\rho_{kl} - (\rho_{ik}\rho_{jl} + \rho_{jk}\rho_{il})]$. QED

Lemma B.4. We have

$$\frac{\partial}{\partial \Theta_{ij}} Det[\Theta] = 2(-1)^{i+j} Det[\Theta^{(i,j)}] \tag{B.10}$$

where $\Theta^{i,j}$ is obtained from Θ by deleting row i and column j and further

$$\frac{\frac{\partial}{\partial \Theta_{ij}} Det[\Theta]}{Det[\Theta]} = 2\Theta_{ij}^{-1} \tag{B.11}$$

Proof: The proof relays on the assumption that Θ is symmetric. By expanding the determinant of Θ of the j-th row we obtain: $Det[\Theta] = \sum_{k=1}^{n} \Theta_{jk}(-1)^{j+k} Det[\Theta^{(j,k)}]$ and further by

expanding $Det[\Theta^{(j,k)}]$ by the i'th row $(i<j)$ we have factored out the two variable elements Θ_{ij} and $\Theta_{ji} = \Theta_{ij}$:

$Det[\Theta] = \sum_{\substack{k,l=1 \\ l\neq k}}^{n} \Theta_{jk}\Theta_{il}(-1)^{i+j+k+l-1}Det[\Theta^{(ij,kl)}]$, where $\Theta^{(ij,kl)}$ is obtained from Θ by deleting row i and j and column k and l. Differentiating the last equation with respect to Θ_{ij} we get a single contributions when $k = i$ and $l\neq j$ and $l = j$ and $k\neq i$ and a square contribution when $k = i$ and $l = j$. Collecting the different parts we find:

$$\frac{\partial}{\partial\Theta_{ij}}Det[\Theta] = (-1)^{i+j}\left(\sum_{\substack{l=1 \\ l\neq i,\, l\neq j}}^{n}\Theta_{il}(-1)^{i+l-1}Det[\Theta^{(ij,il)}] + \sum_{\substack{k=1 \\ k\neq j,\, k\neq i}}^{n}\Theta_{jk}(-1)^{k+j-1}Det[\Theta^{(ij,kj)}] + 2\Theta_{ij}(-1)^{i+j-1}Det[\Theta^{(ij,ij)}]\right)$$

$= (-1)^{i+j}\left(\sum_{\substack{l=1 \\ l\neq i}}^{n}\Theta_{il}(-1)^{i+l-1}Det[\Theta^{(ij,il)}] + \sum_{\substack{k=1 \\ k\neq j}}^{n}\Theta_{kj}(-1)^{k+j-1}Det[\Theta^{(ij,kj)}]\right)$. The first sum we recognize as $Det[\Theta^{(j,i)}]$ expanded after the i-th row, and second sum is $Det[\Theta^{(i,j)}]$ expanded after the j-th row. Since Θ is symmetric, it follows that $Det[\Theta^{(j,i)}] = Det[\Theta^{(i,j)}]$ and this gives $\frac{\partial}{\partial\Theta_{ij}}Det[\Theta] = 2(-1)^{i+j}Det[\Theta^{(i,j)}]$. Equation (B.11) follows from the fact that the inverse of a matrix may be expressed by its cofactors: $\Theta_{ij}^{-1} = \frac{(-1)^{i+j}Det[\Theta^{(j,i)}]}{Det[\Theta]}$.QED.

We shall end this introductory section by showing that the multinormal distribution has some remarkable properties which will be important when we consider integrals like (B.2).

Lemma B.5. The multinormal distribution (B.1) has the following properties:

$$\frac{\partial f_n}{\partial\xi_r} = -\left(\sum_{j=1}^{n}\xi_j M_{jr}^{-1}\right)f_n \text{ and further} \tag{B.12}$$

$$\frac{\partial f_n}{\partial\rho_{rs}} = \frac{\partial^2 f_n}{\partial\xi_r\partial\xi_s} = \left[\left(\sum_{j=1}^{n}\xi_j M_{jr}^{-1}\right)\left(\sum_{i=1}^{n}\xi_i M_{ir}^{-1}\right) - M_{rs}^{-1}\right]f_n \tag{B.13}$$

Proof: The first part is obvious. To prove the second part, we find by differentiation of f_n with respect to the parameter ρ_{rs}

$\frac{\partial f_n}{\partial \rho_{rs}} = \left[Det[M]^{1/2} \frac{\partial}{\partial \rho_{rs}} (Det[M]^{-1/2}) - \frac{1}{2} \sum_{i,j=1}^{n} \xi_i \xi_j \frac{\partial}{\partial \rho_{rs}} (M_{ij}^{-1}) \right] f_n = \left[-\frac{1}{2} \frac{\frac{\partial}{\partial \rho_{rs}} Det[M]}{Det[M]} - \frac{1}{2} \sum_{i,j=1}^{n} \xi_i \xi_j \frac{\partial}{\partial \rho_{rs}} (M_{ij}^{-1}) \right] f_n$. By using (B.3) and (B.11) we have $\frac{\partial f_n}{\partial \rho_{rs}} = \left[-M_{rs}^{-1} + \frac{1}{2} \sum_{i,j=1}^{n} \xi_i \xi_j (M_{ir}^{-1} M_{js}^{-1} + M_{jr}^{-1} M_{is}^{-1}) \right] f_n = \left[-M_{rs}^{-1} + \left(\sum_{i=1}^{n} \xi_i M_{ir}^{-1} \right) \right] \left(\sum_{j=1}^{n} \xi_j M_{js}^{-1} \right) f_n$ which also equals $\frac{\partial^2 f_n}{\partial \xi_r \partial \xi_s}$. QED

B.2 Conditional distributions of multinormal variables

We shall make use of some results obtained from conditional distributions. It is well known that conditional distribution derived from multinormal distributions also is multinormal, though the conditional variable will now longer have zero mean. The proof of the general statement on conditional distributions is done by rewriting the quadratic form $\sum_{i,j=1}^{n} \xi_i \xi_j M_{ij}^{-1}$ in different ways. We include the general theorem because we shall apply the results with different dimensions later on in the appendix. It is possible to find the following result in the literature but we shall state the general case here for the sake of completeness.

Lemma B.6. The conditional distribution of $\{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{s_1-1}, \boldsymbol{B}_{s_1+1}, \ldots, \boldsymbol{B}_{s_k-1}, \boldsymbol{B}_{s_k+1}, \ldots, \boldsymbol{B}_n\}$ given $\boldsymbol{B}_{s_1}, \ldots, \boldsymbol{B}_{s_k}$ (where $1 \le s_1 < s_2 \ldots < s_k \le n$) is multinormal of dimension $n-k$ and with parameters:

$$\boldsymbol{E}[\boldsymbol{B}_i | \boldsymbol{B}_{s_1} \ldots, \boldsymbol{B}_{s_k}] = \frac{\sum_{l=1}^{k} Det[M_{s_1, \ldots, s_k, ;li}] \boldsymbol{B}_{s_l}}{Det[M_{s_1, \ldots, s_k}]} \quad i = 1, \ldots, n, \; i \ne s_1, \ldots, s_k, \tag{B.14}$$

and where the covariance matrix given by:

$$\boldsymbol{Cov}[\boldsymbol{B}_i, \boldsymbol{B}_j | \boldsymbol{B}_{s_1} \ldots, \boldsymbol{B}_{s_k}] = \frac{Det \begin{bmatrix} M_{s_1, \ldots, s_k} & \overline{\rho_{s_1, \ldots, s_k;j}} \\ \rho_{s_1, \ldots, s_k;i} & \rho_{ij} \end{bmatrix}}{Det[M_{s_1, \ldots, s_k}]} \tag{B.15}$$

$i, j = 1, \ldots, n$, $i, j \ne s_1, \ldots, s_k$ where $M_{s_1, \ldots, s_k}$ is the k-rowed minor of the correlation covariance matrix M giving by the rows and columns $s_1, \ldots, s_k$, (that is $M_{s_1, \ldots, s_k} = (\rho_{s_i s_j})$ $i, j = 1, \ldots, k$, $\rho_{s_i s_i} = 1$ and $\rho_{s_i s_j} = \rho_{s_j s_i}$) and $\rho_{s_1, \ldots, s_k;i}$ is a k-dimensional (row-)vector

with elements $\rho_{s_r i}$ $r = 1, \ldots, k$; that is $\rho_{s_1, \ldots, s_k;i} = (\rho_{s_1 i}, \rho_{s_2 i}, \ldots, \rho_{s_k i})$ $i = 1, ..., n$, and $\overline{\rho_{s_1, \ldots, s_k;i}}$ is the corresponding column-vector (transposed). The matrix $M_{s_1, \ldots, s_k, ;li}$ is obtained from $M_{s_1, \ldots, s_k}$ by replacing column l, $\overline{\rho_{s_1, \ldots, s_k;s_l}}$ by $\overline{\rho_{s_1, \ldots, s_k;i}}$, that is $M_{s_1, \ldots, s_k;li} = [\overline{\rho_{s_1, \ldots, s_k;s_1}}, \ldots, \overline{\rho_{s_1, \ldots, s_k;s_l - 1}}, \overline{\rho_{s_1, \ldots, s_k;i}}, \overline{\rho_{s_1, \ldots, s_k;s_l + 1}}, \ldots, \overline{\rho_{s_1, \ldots, s_k;s_k}}]$,
$l = 1, \ldots, k$, $i = 1, ..., n$, $i \neq s_1, \ldots, s_k$.

Further the inverse of the Covariance matrix may be found from M^{-1} by deleting row and column $s_1, s_2 \ldots, s_k$.

Proof: Without losing generality we choose $s_1 = 1, ..., s_k = k$, (since it is always possible to re-arrange the rows and columns in the original correlation matrix M so that this will be the case) and we denote M^k the k-rowed minor of M given by the rows and columns $1, \ldots, k$. The proof is greatly simplified by introducing matrix formulation. We have

$$f(\xi_{k+1},...,\xi_n | \xi_1,...,\xi_k) = \frac{\sqrt{Det[M^{-1}]}}{(2\pi)^{(n-k)/2}\sqrt{Det[M^{k-1}]}} \exp\left[-\frac{1}{2}E\right] \text{ where} \tag{B.16}$$

$E = \xi M^{-1}\overline{\xi} - \xi_L M^{k-1}\overline{\xi_L}$ and $\xi = (\xi_1,...,\xi_n)$ and $\xi_L = (\xi_1,...,\xi_k)$. The rest of the proof is now devoted to rewrite the quadratic form E. We divide matrices M and M^{-1} into the following sub matrices $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ and $M^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ where $M_{11} (= M^k)$ and A_{11} are $k \times k$ matrices, M_{12} and A_{12} are $k \times (n-k)$ matrices, M_{21} and A_{21} are $(n-k) \times k$ matrices, and M_{22} and A_{22} are $(n-k) \times (n-k)$ matrices. (Since both M and M^{-1} are symmetric matrices, the same will apply for M_{11}, M_{22}, A_{11} and A_{22}, and the transposed of M_{12} and A_{12} equals M_{21} and A_{21} respectively.) If we also let $\xi = (\xi_L, \xi_H)$ where $\xi_H = (\xi_{k+1},...,\xi_n)$, the quadratic form E may be written as follows:

$$E = \xi_H A_{22}\overline{\xi_H} + \xi_L A_{12}\overline{\xi_H} + \xi_H A_{21}\overline{\xi_L} + \xi_L (A_{11} - M^{k-1})\overline{\xi_L} \tag{B.17}$$

The exponent E will represent a multinormal distribution if and only if we can find a (covariance)-matrix $M_{n|k}$ of dimension $(n-k) \times (n-k)$ and a matrix B of dimension $k \times (n-k)$ so that E may be re-written as:

$$E = (\xi_H - \xi_L B) M_{n|k}^{-1} (\overline{\xi_H} - \overline{B}\overline{\xi_L}) = \xi_H M_{n|k}^{-1}\overline{\xi_H} - \xi_L B M_{n|k}^{-1}\overline{\xi_H} - \xi_H M_{n|k}^{-1}\overline{B}\overline{\xi_L} + \xi_L B M_{n|k}^{-1}\overline{B}\overline{\xi_L} \tag{B.18}$$

Comparing (B.17) and (B.18), this can be done by choosing $M_{n|k}^{-1} = A_{22}$ giving $M_{n|k} = A_{22}^{-1}$. Secondly we must have $BM_{n|k}^{-1} = -A_{12}$ which gives $B = -A_{12}M_{n|k} = -A_{12}A_{22}^{-1}$. From the relation $M \cdot M^{-1} = I$ we get four matrix equations $M_{11}A_{11} + M_{12}A_{21} = I$, $M_{21}A_{11} + M_{22}A_{21} = 0$, $M_{11}A_{12} + M_{12}A_{22} = 0$ and $M_{21}A_{12} + M_{22}A_{22} = I$. From the second and third of these equations we deduce: $A_{21}A_{11}^{-1} = -M_{22}^{-1}M_{21}$ and $A_{12}A_{22}^{-1} = -M_{11}^{-1}M_{12}$, giving

$$B = -A_{12}A_{22}^{-1} = M_{11}^{-1}M_{12} \tag{B.19}$$

From the first matrix equation we find (by pre-multiplying by M_{11}^{-1}): $A_{11} - M^{k-1} = A_{11} - M_{11}^{-1} = -M_{11}^{-1}M_{12}A_{21} = A_{12}A_{22}^{-1}A_{21}$. Then inserting for $A_{12} = -BM_{n|k}^{-1}$ and $A_{21} = \overline{A_{12}} = -M_{n|k}^{-1}\overline{B}$ we get $A_{11} - M^{k-1} = BM_{n|k}^{-1}A_{22}^{-1}M_{n|k}^{-1}\overline{B} = BM_{n|k}^{-1}\overline{B}$. Thus by choosing $M_{n|k}^{-1} = A_{22}$ and $B = -A_{12}M_{n|k} = -A_{12}A_{22}^{-1}$, (B.17) and (B.18) are identically, and represent an $n-k$ dimensional multinormal distribution with covariance matrix $M_{n|k} = A_{22}^{-1}$. Since (B.16) represents a joint probability density function (in the variable $\xi_H = (\xi_{k+1},...,\xi_n)$, and the integral over these variables equals unity), it follows that the relation between the determinants yields:

$$Det[M_{n|k}^{-1}] = \frac{Det[M^{-1}]}{Det[M^{k-1}]}. \tag{B.20}$$

(The relation (B.20) could as well be proved directly for instance by applying the so called Jacobi's theorem which relates the r-rowed minors of an $n \times n$ matrix with the corresponding $(n-r)$-rowed minors of the corresponding matrix of co-factors, so by applying this theorem we also get: $Det[A_{22}] = \frac{Det[M^k]}{Det[M]}$. See [Grad94] page 1142.)

By inserting for $A_{12} = -M_{11}^{-1}M_{12}A_{22}$ in the fourth matrix equation above we find:

$(-M_{21}M_{11}^{-1}M_{12} + M_{22})A_{22} = I$ and it follows that the conditional correlation matrix, equals the inverse of A_{22}, may be written:

$$M_{n|k} = A_{22}^{-1} = M_{22} - M_{21}M_{11}^{-1}M_{12} \tag{B.21}$$

Collecting the results above and writing them component wise we get:

$$\boldsymbol{E}[\boldsymbol{B}_i|\boldsymbol{B}_1 = \xi_1, \ldots, \boldsymbol{B}_k = \xi_k] = \sum_{l=1}^{\kappa} B_{li}\xi_l, \; i = k+1, \ldots, n \text{ where} \tag{B.22}$$

$$B_{li} = \sum_{r=1}^{k} M_{lr}^{k\,-1}\rho_{ri} = \frac{\sum_{r=1}^{k}(-1)^{l+r} Det[M^{k(r,l)}]\rho_{si}}{Det[M^k]} = \frac{Det[M_{li}^k]}{Det[M^k]}, \tag{B.23}$$

for $i = k+1, \ldots, n$, $l = 1, \ldots, k$ and

$$\boldsymbol{Cov}[\boldsymbol{B}_i, \boldsymbol{B}_j|\boldsymbol{B}_1 = \xi_1, \ldots, \boldsymbol{B}_k = \xi_k] = [M_{n|k}]_{ij} = \rho_{ij} - \sum_{l=1}^{k}\sum_{r=1}^{k} M_{lr}^{k\,-1}\rho_{il}\rho_{rj}$$

$$= \frac{Det[M^k]\rho_{ij} - \sum_{l=1}^{k}\sum_{r=1}^{k}(-1)^{l+r} Det[M^{k(r,l)}]\rho_{il}\rho_{sj}}{Det[M^k]} = \frac{Det\begin{bmatrix} M^k & \overline{\rho_j^k} \\ \rho_i^k & \rho_{ij} \end{bmatrix}}{Det[M^k]} \text{ for } i, j = k+1, \ldots, n \tag{B.24}$$

In (B.23) and (B.24) we have written the inverse of a matrix by using the corresponding co-factors and the matrix $M^{k(r,l)}$ is obtained from M^k by deleting row s and column l. Further $Det[M^{k(r,l)}] = Det[M^{k(l,r)}]$ since M^k is a symmetric matrix, and $M_{li}^k = [\overline{\rho_1^k}, \ldots, \overline{\rho_{l-1}^k}, \overline{\rho_i^k}, \overline{\rho_{l+1}^k}, \ldots, \overline{\rho_k^k}]$, $l = 1, \ldots, k$, $i = k+1, \ldots, n$ where and ρ_i^k is a k-dimensional (row-)vector $\rho_i^k = (\rho_{1i}, \rho_{2i}, \ldots, \rho_{ki})$, $i = 1, \ldots, n$, and $\overline{\rho_i^k}$ is the corresponding column-vector (transposed). QED

From (B.15) we find the conditional standard deviations as:

$$\sigma_i^{s_1, \ldots, s_k} = (\boldsymbol{Var}[\boldsymbol{B}_i|\boldsymbol{B}_{s_1} \ldots, \boldsymbol{B}_{s_k}])^{1/2} = \sqrt{\frac{Det[M_{s_1, \ldots, s_k, i}]}{Det[M_{s_1, \ldots, s_k}]}} \; i = 1, \ldots, n, \; i \neq s_1, \ldots, s_k. \tag{B.25}$$

We may then find the conditional $(n-k)\times(n-k)$ correlation matrix $M^{s_1, \ldots, s_k} = (\rho_{ij}^{s_1, \ldots, s_k})$ as:

$$\rho_{ij}^{s_1, \ldots, s_k} = \frac{Det\begin{bmatrix} M_{s_1, \ldots, s_k} & \overline{\rho_{s_1, \ldots, s_k; j}} \\ \rho_{s_1, \ldots, s_k; i} & \rho_{ij} \end{bmatrix}}{\sqrt{Det[M_{s_1, \ldots, s_k, i}]Det[M_{s_1, \ldots, s_k, j}]}} = \frac{\rho_{ij} - \sum_{l=1}^{k}\sum_{r=1}^{k}[M_{s_1, \ldots, s_k}^{-1}]_{lr}\rho_{is_r}\rho_{s_lj}}{\sigma_i^{s_1, \ldots, s_k}\sigma_j^{s_1, \ldots, s_k}} \tag{B.26}$$

for $i, j = 1, \ldots, n$, $i, j \neq s_1, \ldots, s_k$

If we denote $M^{s_1, \ldots, s_k -1} = (M_{ij}^{s_1, \ldots, s_k -1})$ the inverse of the correlation matrix $M^{s_1, \ldots, s_k}$, we can write the joint conditional PDF (B.16) in a standard way:

$$f(\xi_1,\ldots,\xi_{s_1-1}, \xi_{s_1+1}, \ldots, \xi_{s_k-1}, \xi_{s_k+1}, \ldots, \xi_n | \xi_{s_1}, \ldots, \xi_{s_k}) =$$

$$\frac{\sqrt{Det\left[M^{s_1, \ldots s_k -1}\right]}}{(2\pi)^{(n-k)/2} \prod\limits_{\substack{i=1 \\ i \neq s_1, \ldots, s_k}}^{n} \sigma_i^{s_1, \ldots, s_k}} \exp\left[-\frac{1}{2}\left(\sum\limits_{\substack{i,j=1 \\ i,j \neq s_1, \ldots, s_k}}^{n} \frac{\xi_i - \sum\limits_{l=1}^{k} a_{il}^{s_1, \ldots, s_k} \xi_{s_l}}{\sigma_i^{s_1, \ldots, s_k}} \frac{\xi_j - \sum\limits_{l=1}^{k} a_{jl}^{s_1, \ldots, s_k} \xi_{s_l}}{\sigma_j^{s_1, \ldots, s_k}} M_{ij}^{s_1, \ldots s_k -1}\right)\right] \tag{B.27}$$

where we have defined the coefficients:

$$a_{il}^{s_1, \ldots, s_k} = \frac{Det[M_{s_1, \ldots, s_k; li}]}{Det[M_{s_1, \ldots, s_k}]} = \sum_{r=1}^{k} [M_{s_1, \ldots, s_k}^{-1}]_{lr} \rho_{s_r i} \quad i = 1, \ldots, n, \; i \neq s_1, \ldots, s_k \tag{B.28}$$

and $l = 1, \ldots, k$

$$\rho_{ij}^{s_1, \ldots, s_k} = \frac{Det\begin{bmatrix} M_{s_1, \ldots, s_k} & \overline{\rho_{s_1, \ldots, s_k; j}} \\ \rho_{s_1, \ldots, s_k; i} & \rho_{ij} \end{bmatrix}}{\sqrt{Det[M_{s_1, \ldots, s_k, i}] Det[M_{s_1, \ldots, s_k, j}]}} = \frac{\rho_{ij} - \sum\limits_{l=1}^{k} \sum\limits_{r=1}^{k} [M_{s_1, \ldots, s_k}^{-1}]_{lr} \rho_{is_r} \rho_{s_l j}}{\sigma_i^{s_1, \ldots, s_k} \sigma_j^{s_1, \ldots, s_k}}$$

Below we apply the general lemma above and give the results for some specific cases which we use later in the analysis.

Corollary B.1. The conditional distribution of $\{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{k-1}, \boldsymbol{B}_{k+1}, \ldots, \boldsymbol{B}_n\}$ given $\boldsymbol{B}_k$ is multinormal of dimension $n-1$ and with parameters:

$$\boldsymbol{E}[\boldsymbol{B}_i | \boldsymbol{B}_k] = \rho_{ik} \boldsymbol{B}_k \quad i = 1, \ldots, n, \; i \neq k \text{ and covariance matrix given by:} \tag{B.29}$$

$$\boldsymbol{Cov}[\boldsymbol{B}_i, \boldsymbol{B}_j | \boldsymbol{B}_k] = \rho_{ij} - \rho_{ik}\rho_{kj} \quad i,j = 1, \ldots, n, \; i \neq k, j \neq k \tag{B.30}$$

Further the inverse of the covariance matrix is obtained from M^{-1} by deleting row and column k.

The conditional standard deviations are found from (B.30):

$$\sigma_i^k = (\boldsymbol{Var}[\boldsymbol{B}_i | \boldsymbol{B}_k])^{1/2} = \sqrt{1 - \rho_{ik}^2} \quad i = 1, \ldots, n, \; i \neq k \text{ and} \tag{B.31}$$

the conditional $(n-1) \times (n-1)$ correlation matrix $M^k = (\rho_{ij}^k)$ is found to be:

$$\rho_{ij}^{k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{1-\rho_{ik}^2}\sqrt{1-\rho_{jk}^2}} \quad i, j = 1, ..., n, \; i, j \neq k. \tag{B.32}$$

Corollary B.2. The conditional distribution of $\boldsymbol{B}_1$ given $\boldsymbol{B}_2, ..., \boldsymbol{B}_n$ is normal distributed with parameters:

$$\boldsymbol{E}[\boldsymbol{B}_1 | \boldsymbol{B}_2 \ldots, \boldsymbol{B}_n] = -\sum_{j=1}^{n} \frac{M_{1j}^{-1}}{M_{11}^{-1}} \boldsymbol{B}_k \tag{B.33}$$

and the variance is given by:

$$\boldsymbol{Var}[\boldsymbol{B}_1 | \boldsymbol{B}_2 \ldots, \boldsymbol{B}_n] = \frac{1}{M_{11}^{-1}}, \text{ and} \tag{B.34}$$

further the corresponding PDF is give as:

$$f(\xi_1 | \xi_2, \ldots, \xi_n) = \frac{\sqrt{M_{11}^{-1}}}{(2\pi)^{1/2}} \exp\left[-\frac{M_{11}^{-1}}{2} \left(\xi_1 + \sum_{j=1}^{n} \frac{M_{1j}^{-1}}{M_{11}^{-1}} \xi_k \right)^2 \right] \tag{B.35}$$

Obviously the conditional distribution may be taken as a (new) starting point when conditioning on additional variable. This is clear since the normalised (conditional) variables

$$\boldsymbol{B}_i^{s_1, \ldots, s_k} = \frac{\boldsymbol{B}_i - \sum_{l=1}^{k} a_{il}^{s_1, \ldots, s_k} \boldsymbol{B}_{s_l}}{\sigma_i^{s_1, \ldots, s_k}} \quad i = 1, ..., n, \; i \neq s_1, \ldots, s_k \tag{B.36}$$

are standard multinormal distributed (of dimension $n-k$) given $\boldsymbol{B}_{s_l}$, $l = 1, ..., k$, with covariance (correlation) matrix $M^{s_1, \ldots, s_k}$. This makes it possible to obtain relations between the parameters of any sets of conditional multinormal variables. We shall use this property to find recursion formula for the parameters based on the number of variables that we are conditioning on; k. We start with the result for $k-1$. By the observation above we have $\boldsymbol{B}_i^{s_1, \ldots, s_{k-1}} = \frac{\boldsymbol{B}_i - \sum_{l=1}^{k-1} a_{il}^{s_1, \ldots, s_{k-1}} \boldsymbol{B}_{s_l}}{\sigma_i^{s_1, \ldots, s_{k-1}}}$ $i = 1, ..., n, \; i \neq s_1, \ldots, s_{k-1}$ are standard multinormal distributed (with dimension $n-(k-1)$) given $\boldsymbol{B}_{s_l}$, $l = 1, ..., k-1$. Thus, since conditioning on an extra variable $\boldsymbol{B}_{s_k}$ is the same as conditioning on $\boldsymbol{B}_{s_k}^{s_1, \ldots, s_{k-1}}$ given $\boldsymbol{B}_{s_l}$,

$l = 1, ..., k-1$ we obtain by applying corollary B.1 above that $\left\{ \boldsymbol{B}_i^{s_1, \ldots, s_{k-1}} \right\}$ is multinormal distributed with dimension $n-(k-1)-1 = n-k$ and with parameters:

$$\boldsymbol{E}[\boldsymbol{B}_i^{s_1, \ldots, s_{k-1}} \big| \boldsymbol{B}_{s_1}, \ldots, \boldsymbol{B}_{s_{k-1}}, \boldsymbol{B}_{s_k}] = \rho_{is_k}^{s_1, \ldots, s_{k-1}} \boldsymbol{B}_{s_k}^{s_1, \ldots, s_{k-1}} \text{ for} \tag{B.37}$$

$i = 1, ..., n$, $i \neq s_1, \ldots, s_k$ and covariance matrix given by:

$$\boldsymbol{Cov}[\boldsymbol{B}_i^{s_1, \ldots, s_{k-1}}, \boldsymbol{B}_j^{s_1, \ldots, s_{k-1}} \big| \boldsymbol{B}_{s_1}, \ldots, \boldsymbol{B}_{s_{k-1}}, \boldsymbol{B}_{s_k}] = \rho_{ij}^{s_1, \ldots, s_{k-1}} - \rho_{is_k}^{s_1, \ldots, s_{k-1}} \rho_{s_k j}^{s_1, \ldots, s_{k-1}} \tag{B.38}$$

$i, j = 1, ..., n$, $i, j \neq s_1, \ldots, s_k$.

Now the conditional normalised variables of the $\left\{ \boldsymbol{B}_i^{s_1, \ldots, s_{k-1}} \right\}$'s given $\boldsymbol{B}_{s_l}$, $l = 1, ..., k$ may be obtained by using (B.37) and (B.38) and must equal (B.36), and this gives the following important relations:

$$\boldsymbol{B}_i^{s_1, \ldots, s_k} = \frac{\boldsymbol{B}_i^{s_1, \ldots, s_{k-1}} - \rho_{is_k}^{s_1, \ldots, s_{k-1}} \boldsymbol{B}_{s_k}^{s_1, \ldots, s_{k-1}}}{\sqrt{1-(\rho_{is_k}^{s_1, \ldots, s_{k-1}})^2}} \quad i = 1, ..., n,\ i \neq s_1, \ldots, s_k. \tag{B.39}$$

Since the correlation does not alter by translations like (B.36), we may express the conditional correlation between $\boldsymbol{B}_i$ and $\boldsymbol{B}_j$ by applying (B.38):

$$\rho_{ij}^{s_1, \ldots, s_k} = \frac{\rho_{ij}^{s_1, \ldots, s_{k-1}} - \rho_{is_k}^{s_1, \ldots, s_{k-1}} \rho_{js_k}^{s_1, \ldots, s_{k-1}}}{\sqrt{1-(\rho_{is_k}^{s_1, \ldots, s_{k-1}})^2}\sqrt{1-(\rho_{js_k}^{s_1, \ldots, s_{k-1}})^2}} \quad i, j = 1, ..., n,\ i, j \neq s_1, \ldots, s_k. \tag{B.40}$$

Finally, by applying equation (B.39) and inserting expression (B.36) (on the left hand side with counting parameter k and on the left hand side with $k-1$) we obtain an identity in the variable $\boldsymbol{B}_i$ and $\boldsymbol{B}_{s_l}$ which requires the following relations to be fulfilled:

$$\sigma_i^{s_1, \ldots, s_k} = \sigma_i^{s_1, \ldots, s_{k-1}} \sqrt{1-(\rho_{is_k}^{s_1, \ldots, s_{k-1}})^2} \text{ for } i = 1, ..., n,\ i \neq s_1, \ldots, s_k \text{ and} \tag{B.41}$$

$$a_{il}^{s_1, \ldots, s_k} = a_{il}^{s_1, \ldots, s_{k-1}} - \frac{\rho_{is_k}^{s_1, \ldots, s_{k-1}} \sigma_{s_k}^{s_1, \ldots, s_{k-1}}}{\sigma_i^{s_1, \ldots, s_{k-1}}} a_{s_k l}^{s_1, \ldots, s_{k-1}} \tag{B.42}$$

for $l = 1, ..., k-1$ and $i = 1, ..., n$, $i \neq s_1, \ldots, s_k$ and

$$a_{ik}^{s_1,\ldots,s_k} = \frac{\rho_{is_k}^{s_1,\ldots,s_{k-1}}\sigma_{s_k}^{s_1,\ldots,s_{k-1}}}{\sigma_i^{s_1,\ldots,s_{k-1}}} \text{ for } i = 1,\ldots,n,\ i \neq s_1,\ldots,s_k \tag{B.43}$$

The recursion formulae (B.40)-(B.43) may also be proved more directly by applying the determinant expressions (B.25), (B.26) and (B.28) by applying Jacobi's theorem [Grad94], (14.16 page 1142) to obtain relations between determinants of different dimensions.

When considering multinormal integrals like (B.2) we will often use specific values of the variable $\xi_i = C_i$ (for instance as integration limits, and we denote the corresponding vector $\boldsymbol{C} = (C_1,\ldots,C_n)$). Recursively (by induction) we may define integrals of dimension $n-s$ based on the conditional distribution (B.27) (where the conditioning is done with respect to the variables $\boldsymbol{B}_{k_1} = C_{k_1},\ldots,\boldsymbol{B}_{k_s} = C_{k_s}$). The corresponding multinormal integral of dimension $n-k$ will then be of type $I(n-k,\boldsymbol{C}^{s_1,\ldots,s_k},M^{s_1,\ldots,s_k})$ where the corresponding integration limit vector $\boldsymbol{C}^{s_1,\ldots,s_k}$ (with elements $C_i^{s_1,\ldots,s_k}$ of dimension $n-s$) is found by inserting $\xi_i = C_i$ in the exponent of the conditional distribution (B.27):

$$C_i^{s_1,\ldots,s_k} = \frac{C_i - \sum_{l=1}^{k} a_{il}^{s_1,\ldots,s_k} C_{s_l}}{\sigma_i^{s_1,\ldots,s_k}} \quad i = 1,\ldots,n,\ i \neq s_1,\ldots,s_k \tag{B.44}$$

This equation is identically with (B.36) and therefore the recursion (B.39) yields. We summarise the results above in the following lemma:

Lemma B.7. Let $M^{s_1,\ldots,s_k} = (\rho_{ij}^{s_1,\ldots,s_k})$; $(i,j = 1,\ldots,n\ \ i,j \neq s_1,\ldots,\ i,j \neq s_k)$ be the correlation matrix and further $\boldsymbol{C}^{s_1,\ldots,s_k} = (C_i^{s_1,\ldots,s_k})$; $(i = 1,\ldots,n\ \ i \neq s_1,\ldots,\ i \neq s_k)$ be the vector of the corresponding integration limits for the integral $I(n-s,\boldsymbol{C}^{k_1,\ldots,k_s},M^{k_1,\ldots,k_s})$; based on the conditional multinormal distribution (B.27) of dimension $n-k$ given the k variable $\boldsymbol{B}_{s_1} = C_{s_1},\ldots,\boldsymbol{B}_{s_k} = C_{s_k}$; $(k<n)$. Then these parameters are given by the determinant expressions (B.26) and (B.44) by (B.28) and (B.25), and satisfy the following recursion formulae:

$$C_i^{s_1,\ldots,s_k} = \frac{C_i^{s_1,\ldots,s_{k-1}} - \rho_{is_k}^{s_1,\ldots,s_{k-1}} C_{s_k}^{s_1,\ldots,s_{k-1}}}{\sqrt{1-(\rho_{is_k}^{s_1,\ldots,s_{k-1}})^2}} \tag{B.45}$$

for $i = 1,\ldots,n$ and $i \neq s_1,\ldots,\ i \neq s_k$ and

$$\rho_{ij}^{s_1,\ldots,s_k} = \frac{\rho_{ij}^{s_1,\ldots,s_{k-1}} - \rho_{is_k}^{s_1,\ldots,s_{k-1}}\rho_{js_k}^{s_1,\ldots,s_{k-1}}}{\sqrt{1-(\rho_{is_k}^{s_1,\ldots,s_{k-1}})^2}\sqrt{1-(\rho_{js_k}^{s_1,\ldots,s_{k-1}})^2}} \tag{B.46}$$

for $i, j = 1, \ldots, n$ and $i, j \neq s_1, \ldots, i, j \neq s_k$

The recursion may be started for $k = 1$ by calculating $\boldsymbol{C}^{s_1} = (C_i^{s_1})$ and $M^{s_1} = (\rho_{ij}^{s_1})$:

$$C_i^{s_1} = \frac{C_i - \rho_{is_1}C_{s_1}}{\sqrt{1-\rho_{is_1}^2}} \quad \text{for } i = 1, \ldots, n \text{ and } i \neq s_1 \text{ and} \tag{B.47}$$

$$\rho_{ij}^{s_1} = \frac{\rho_{ij} - \rho_{is_1}\rho_{js_1}}{\sqrt{1-\rho_{is_1}^2}\sqrt{1-\rho_{js_1}^2}} \quad \text{for } i, j = 1, \ldots, n \text{ and } i, j \neq s_1 . \tag{B.48}$$

B.3 Multinormal integrals

In the rest of this appendix we shall examine different types of integrals over the multinormal distribution (e.g. given by (B.1)). The origin of these types of integrals comes from the n-point approximation of the excess time and volume distribution of a stationary Gaussian process. We find that these type of integrals have some particular properties which we shall take advantage of in numerical computations.

As a starting point we first consider the (standard) multinormal integral of the form:

$$I(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \ldots \int_{\xi_n = C_n}^{\infty} f_n(\xi_1,\ldots,\xi_n;M)\,d\xi_n \ldots d\xi_1 \tag{B.49}$$

where $f_n(\xi_1,\ldots,\xi_n;M)$ is given by (B.1), $\boldsymbol{C} = (C_1, \ldots, C_n)$-is the vector consisting of the integration limits, $M = (\rho_{ij})$-is the covariance matrix (which is symmetric and with $\rho_{ii} = 1$) and n is the dimension of the integral. Below we frequently also will have integrals of type

$$I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \ldots \int_{\xi_n = C_n}^{\infty} f_n(\xi_1,\ldots,\xi_{s_1-1}, C_{s_1}, \xi_{s_1+1}, \ldots, \xi_{s_k-1}, C_{s_k}, \xi_{s_k+1}, \ldots, \xi_n;M)\,d\xi_n \ldots d\xi_1 \tag{B.50}$$

where the integral is of dimension $n-k$ and does not involve the variables $\xi_{s_1}, \ldots, \xi_{s_k}$. By applying the conditional distribution (B.27) we find that this type of integral may be written as:

$$I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) = f_k(C_{s_1},\ldots,C_{s_k};M_{s_1,\ldots,s_k})\,I(n-k, \boldsymbol{C}^{s_1,\ldots,s_k}, M^{s_1,\ldots,s_k}) \tag{B.51}$$

where $f_k(\xi_1,...,\xi_k;M_{s_1,\ldots,s_k})$ is the k dimensional standard multinormal distribution with correlation matrix $M_{s_1,\ldots,s_k}$ which is the k-rowed minor of the correlation covariance matrix M giving by the rows and columns $s_1,\ldots,s_k$, that is $M_{s_1,\ldots,s_k} = (\rho_{s_i s_j})$ $i,j = 1,\ldots,k$, $\rho_{s_i s_i} = 1$ and $\rho_{s_i s_j} = \rho_{s_j s_i}$. (The corresponding integration limits $\boldsymbol{C}^{s_1,\ldots,s_k}$ are given by (B.44) and the correlation matrix $M^{s_1,\ldots,s_k}$ is give by (B.26).)

By using the conditional distribution (B.27) it is also possible to factor out some variables in $I(n,\boldsymbol{C},M)$ and integrate first over these variables. By this approach it is possible to write (B.39) as follows:

$$I(n,\boldsymbol{C},M) = \int_{\xi_1 = C_{s_1}}^{\infty} \cdots \int_{\xi_2 = C_{s_k}}^{\infty} f_k(\xi_1,...,\xi_k;M_{s_1,\ldots,s_k}) I(n-k,\boldsymbol{C}^{s_1,\ldots,s_k}(\xi_1,...,\xi_k),M^{s_1,\ldots,s_k}) d\xi_k \ldots d\xi_1 \tag{B.52}$$

where we indicate that the limit vector $\boldsymbol{C}^{s_1,\ldots,s_k}(\xi_1,...,\xi_k)$ are functions of the variable $(\xi_1,...,\xi_k)$ and the elements are given by

$$C_i^{s_1,\ldots,s_k}(\xi_1,...,\xi_k) = \frac{C_i - \sum_{l=1}^{k} a_{il}^{s_1,\ldots,s_k}\xi_l}{\sigma_i^{s_1,\ldots,s_k}} \quad i = 1,...,n,\ i \neq s_1,\ldots,s_k \tag{B.53}$$

Below we shall show that integrals of type $I(n,\boldsymbol{C},M)$ have some remarkable properties mainly because of the results in lemma B.5. By applying these properties we may relate the partial derivatives of the integral with respect to the parameters.

Theorem B.1. For the integral (B.49) we have:

$$\frac{\partial I}{\partial C_k} = -I^k(n,\boldsymbol{C},M) \text{ for } 1 \le k \le n \text{ and further} \tag{B.54}$$

$$\frac{\partial I}{\partial \rho_{kl}} = I^{k,l}(n,\boldsymbol{C},M) \text{ for } 1 \le k < l \le n \tag{B.55}$$

Proof: The first part of the theorem is obvious. The second part is a direct result of lemma B.5 equation (B.13) giving $\frac{\partial I}{\partial \rho_{kl}} = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} \frac{\partial^2 f_n}{\partial \xi_k \partial \xi_l}(\xi_1,...,\xi_n;M) d\xi_n \ldots d\xi_1 = I^{k,l}(n,\boldsymbol{C},M)$. QED

By exploiting the properties of the conditional density by applying the results from lemma B.6 and lemma B.7 we are able to relate differentiation with respect to the integration lim-

its C_k and differentiation with respect to the covariances ρ_{kl} of n-dimensional multinormal integral by corresponding $n-1$ and $n-2$ dimensional normal integrals.

Theorem B.2. We have

$$\frac{\partial}{\partial C_k} I(n, \boldsymbol{C}, M) = -f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) \tag{B.56}$$

where $f_1(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ is the standard normal density and the vector $\boldsymbol{C}^k$ is obtained from $\boldsymbol{C}$ by deleting element k. The other elements are given by (B.47):

$$C_i^k = \frac{C_i - \rho_{ik} C_k}{\sqrt{1-\rho_{ik}^2}} \text{ for } i = 1, ..., n,\ i \neq k \tag{B.57}$$

Further M^k is obtained from M by deleting row k and column k and having elements given by (B.48):

$$\rho_{ij}^k = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{1-\rho_{ik}^2}\sqrt{1-\rho_{jk}^2}} \text{ for } i, j = 1, ..., n,\ i, j \neq k \tag{B.58}$$

Further

$$\frac{\partial I}{\partial \rho_{kl}}(n, \boldsymbol{C}, M) = f_2(C_k, C_l, \rho_{kl}) I(n-2, \boldsymbol{C}^{k,l}, M^{k,l}) \tag{B.59}$$

where

$$f_2(x_1, x_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right] \tag{B.60}$$

is the standard bivariate normal density function with correlation ρ, and further the vector $\boldsymbol{C}^{k,l}$ may be obtained from $\boldsymbol{C}^k$ by deleting element l (keeping in mind that row k and is already deleted from $\boldsymbol{C}^k$, $k < l$) and its elements may be found for instance from lemma B.7 by using (B.45) with two iterations:

$$C_i^{k,l} = \frac{C_i^k - \rho_{il}^k C_l^k}{\sqrt{1-(\rho_{il}^k)^2}} = \frac{C_i - \dfrac{\rho_{ki} - \rho_{kl}\rho_{li}}{1-\rho_{kl}^2} C_k - \dfrac{\rho_{li} - \rho_{kl}\rho_{ki}}{1-\rho_{kl}^2} C_l}{\sqrt{\dfrac{1-\rho_{kl}^2 - \rho_{ik}^2 - \rho_{il}^2 + 2\rho_{kl}\rho_{ik}\rho_{il}}{1-\rho_{kl}^2}}} \text{ for } i = 1, ..., n,\ i \neq k, l \tag{B.61}$$

Further $M^{k,l}$ may be obtained from M^k by deleting row l and column l (keeping in mind that row k and column k are already deleted from M^k, $k < l$) and with elements obtained by using (B.46) with two iterations:

$$\rho_{ij}^{k,l} = \frac{\rho_{ij}^k - \rho_{il}^k \rho_{jl}^k}{\sqrt{1-(\rho_{il}^k)^2}\sqrt{1-(\rho_{jl}^k)^2}} = \frac{\rho_{ij}(1-\rho_{kl}^2) - \rho_{ik}\rho_{jk} - \rho_{il}\rho_{jl} + \rho_{kl}(\rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk})}{\sqrt{1-\rho_{kl}^2-\rho_{ik}^2-\rho_{il}^2+2\rho_{kl}\rho_{ik}\rho_{il}}\sqrt{1-\rho_{kl}^2-\rho_{jk}^2-\rho_{jl}^2+2\rho_{kl}\rho_{jk}\rho_{jl}}} \tag{B.62}$$

for $i, j = 1, ..., n$, $i, j \neq k, l$

Proof: The theorem follows directly form theorem B.1 and equation (B.51). QED

Corollary B.3. For the special case where $C_i = C$ $i = 1, ..., n$ we find:

$$\frac{\partial I}{\partial \rho_{kl}} = \frac{e^{-\frac{C^2}{1+\rho_{kl}}}}{2\pi\sqrt{1-\rho_{kl}^2}} I(n-2, \boldsymbol{C}^{k,l}, M^{k,l}) \tag{B.63}$$

and where the $\boldsymbol{C}^{k,l}$-vector is given by:

$$C_i^{k,l} = C\frac{(1+\rho_{kl}-\rho_{ki}-\rho_{li})}{\sqrt{\frac{(1+\rho_{kl})(1-\rho_{kl}^2-\rho_{ik}^2-\rho_{il}^2+2\rho_{kl}\rho_{ik}\rho_{il})}{1-\rho_{kl}}}} \quad \text{for } i = 1, ..., n,\ i \neq k, l \tag{B.64}$$

and the correlation matrix $M^{k,l}$ is given by (B.62).

Since we know all the partial derivatives of the integral $I(n, \boldsymbol{C}, M)$ as a function of the $\frac{n(n-1)}{2}$ correlation ρ_{ij} it possible to calculate it by applying standard contour integration.

Theorem B.3. We have:

$$\begin{aligned} I(n, \boldsymbol{C}, M) &= \prod_{i=1}^{n} \phi(C_i) + \sum_{1 \le k < l \le n} \int_{\xi=0}^{1} \rho_{kl} \frac{\partial I}{\partial \rho_{kl}}(n, \boldsymbol{C}, \xi M) d\xi \\ &= \prod_{i=1}^{n} \phi(C_i) + \sum_{1 \le k < l \le n} \rho_{kl} \int_{\xi=0}^{1} f_2(C_k, C_l, \xi\rho_{kl}) I(n-2, \boldsymbol{C}_\xi^{k,l}, M_\xi^{k,l}) d\xi \end{aligned} \tag{B.65}$$

where the matrix $\xi M = (\xi\rho_{ij})$ and $\boldsymbol{C}_{\xi}^{k,l}$ and $M_{\xi}^{k,l}$ are given as in theorem B.2 but with replacing ρ_{ij} with $\xi\rho_{ij}$ for all i, j, and further $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{t=x}^{\infty} e^{-t^2/2} dt$ is the normal integral.

Proof: This is application of the standard theorem of line integral of a gradient where we also make use of the fact that for all $\rho_{ij} = 0$, the variables are all independent, and the integral is the product of (standard) normal integrals. We therefore take the straight line from $(0, ..., 0)$ to $(\rho_{12}, ..., \rho_{n-1n})$ as the path of integration. In the last part of the theorem we just apply theorem B.2. QED.

Theorem B.3 prescribes an effective way of calculating integrals of multinormal distributions of rather high dimensions. For instance an integral of dimension $n = 2$ and $n = 3$ we only need one single (numerical) integration. By using the theorem recursively, integral of dimension $n = 4$ and $n = 5$ only need two numerical integrations. We shall write down the results for the two first cases ($n = 2$ and $n = 3$) explicitly as a reference:

$$I(2, C_1, C_2, \rho_{12}) = \phi(C_1)\phi(C_2) + \rho_{12} \int_{\xi=0}^{1} f_2(C_1, C_2, \xi\rho_{12}) d\xi \text{ and} \tag{B.66}$$

$$I(3, C_1, C_2, C_3, \rho_{12}, \rho_{13}, \rho_{23}) = \phi(C_1)\phi(C_2)\phi(C_3) + \rho_{12} \int_{\xi=0}^{1} f_2(C_1, C_2, \xi\rho_{12})\phi(C_3^{1,2}(\xi\rho_{12}, \xi\rho_{13}, \xi\rho_{23})) d\xi +$$

$$\rho_{13} \int_{\xi=0}^{1} f_2(C_1, C_3, \xi\rho_{13})\phi(C_2^{1,3}(\xi\rho_{12}, \xi\rho_{13}, \xi\rho_{23})) d\xi + \rho_{23} \int_{\xi=0}^{1} f_2(C_2, C_3, \xi\rho_{23})\phi(C_1^{2,3}(\xi\rho_{12}, \xi\rho_{13}, \xi\rho_{23})) d\xi$$

with: (B.67)

$$C_3^{1,2}(\rho_{12}, \rho_{13}, \rho_{23}) = \left(C_3 - C_1 \frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2} - C_2 \frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2}\right) \sqrt{\frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}} \text{ and}$$

$$C_2^{1,3}(\rho_{12}, \rho_{13}, \rho_{23}) = \left(C_2 - C_1 \frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{13}^2} - C_3 \frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{13}^2}\right) \sqrt{\frac{1 - \rho_{13}^2}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}} \text{ and}$$

$$C_1^{2,3}(\rho_{12}, \rho_{13}, \rho_{23}) = \left(C_1 - C_2 \frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{23}^2} - C_3 \frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{23}^2}\right) \sqrt{\frac{1 - \rho_{23}^2}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}}$$

In the literature the main focus have been on multinormal integrals where the lower integration limits is zero. For this special case it is possible to obtain some more explicit expressions. We shall mention some of these cases. First we observe that the integrals (B.66) and (B.67) may be expressed in terms of Arcsin functions. We find:

$$I(2, 0, 0, \rho_{12}) = \frac{1}{4}\left(1 + \frac{2}{\pi} Arc\sin(\rho_{12})\right) \quad \text{and} \tag{B.68}$$

$$I(3, 0, 0, 0, \rho_{12}, \rho_{13}, \rho_{23}) = \frac{1}{8}\left(1 + \frac{2}{\pi}[Arc\sin(\rho_{12}) + Arc\sin(\rho_{13}) + Arc\sin(\rho_{23})]\right) \tag{B.69}$$

In the general cases one could hope that (B.68) and (B.69) could be extended. This is possible only for some special case. By considering (B.59), (B.60) and (B.61) we have:

$$\frac{\partial I}{\partial \rho_{kl}}(n, \mathbf{0}, M) = \frac{1}{2\pi\sqrt{1-\rho_{kl}^2}} I(n-2, \mathbf{0}, M^{k,l}) \tag{B.70}$$

If for instance all the matrices $M^{k,l}$ equal the identity matrix (this requires that $\rho_{ij}^{k,l} = 0$, defined by (B.62), for each k, l, $(1 \le k < l \le n)$ and each (i, j, $1 \le i < j \le n$ and $i, j \ne k, l$), then $I(n-2, \mathbf{0}, M^{k,l}) = \frac{1}{2^{n-2}}$ and we may perform the integration in (B.65) explicitly by applying (B.70). We therefore have the following lemma:

Lemma B.8. If $\boldsymbol{B}_i$ and $\boldsymbol{B}_j$ conditions on each pair $\{\boldsymbol{B}_k, \boldsymbol{B}_l\}$, $k \ne l$ are independent for each $i \ne k, l$ and $j \ne i, k, l$, then $\boldsymbol{Cov}[\boldsymbol{B}_i, \boldsymbol{B}_j | \{\boldsymbol{B}_k, \boldsymbol{B}_l\}] = 0$, and we have:

$$I(n, \mathbf{0}, M) = \frac{1}{2^n}\left(1 + \frac{2}{\pi}\left[\sum_{1 \le k < l \le n} Arc\sin(\rho_{kl})\right]\right) \tag{B.71}$$

It is possible to extend the result above by assuming that $I(n-2, \mathbf{0}, M^{k,l})$ has the property of lemma B.8 (i.e. is on the form (B.71)) and thereby including the "next" contribution to the integral above. We shall, however, not carry the analysis any further because the result will include terms that are given as integrals that will be difficult to find explicit expressions for in the general case.

We shall also consider multinormal integrals of type:

$$I_i(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} \xi_i f_n(\xi_1, \ldots, \xi_n; M) d\xi_n \ldots d\xi_1 \quad \text{and} \tag{B.72}$$

$$I_{i,j}(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} \xi_i \xi_j f_n(\xi_1, \ldots, \xi_n; M) d\xi_n \ldots d\xi_1 \quad \text{for } i < j \tag{B.73}$$

We shall show that these integrals may be written as an integral of type (B.49).

Theorem B.4. We have

$$I_i(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \rho_{ik} I^k(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \rho_{ik} f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) \tag{B.74}$$

where $f_1(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ is the standard normal density and the vector $\boldsymbol{C}^k$ and the matrix M^k is given in theorem B.2.

Proof: We have $\dfrac{\partial}{\partial \xi_k} f_n(\xi_1,...,\xi_n;M) = -f_n(\xi_1,...,\xi_n;M)\left(\sum_{i=1}^{n} \xi_i M_{ik}^{-1}\right)$. Integrating this relation gives: $\sum_{i=1}^{n} I_i(n, \boldsymbol{C}, M) M_{ik}^{-1} = -\int_{\xi_1 = C_1}^{\infty} ... \int_{\xi_n = C_n}^{\infty} \dfrac{\partial}{\partial \xi_k} f_n(\xi_1,...,\xi_n;M) d\xi_n ... d\xi_1 = I^k(n, \boldsymbol{C}, M)$.
The last relation can be viewed as a linear system for I_i. Then by post multiplying by ρ_{ki} and summing we get: $I_i(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \rho_{ik} I^k(n, \boldsymbol{C}, M)$. The result follows now by using (B.51). QED.

The corresponding result for integrals of type (B.73) is more difficult to obtain. We shall use the transformation (B.5) to obtain the result.

Theorem B.5. We have:

$$I_{i,j}(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} [\rho_{ik}(\rho_{jl} - \rho_{kl}\rho_{jk})] I^{k,l}(n, \boldsymbol{C}, M) + \sum_{k=1}^{n} C_k \rho_{ik} \rho_{jk} I^k(n, \boldsymbol{C}, M) + \rho_{ij} I(n, \boldsymbol{C}, M) \tag{B.75}$$

$$= \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} [\rho_{ik}(\rho_{jl} - \rho_{kl}\rho_{jk})] f_2(C_k, C_l, \rho_{kl}) I(n-2, \boldsymbol{C}^{k,l}, M^{k,l}) + \sum_{k=1}^{n} C_k (\rho_{ik}\rho_{jk} f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) + \rho_{ij} I(n, \boldsymbol{C}, M)) \tag{B.76}$$

for $i < j$, where $\boldsymbol{C}^k$ and M^k, and $\boldsymbol{C}^{k,l}$ and $M^{k,l}$ are defined in theorem B.2.

Proof: We use a different approach to prove theorem B.5 than we used for theorem B.4. We shall prove the theorem by applying the transformation (B.5) for the integrals $I(n, \boldsymbol{C}, M)$ and $I_{i,j}(n, \boldsymbol{C}, M)$. By changing the integration variable to $\xi_i^* = \sqrt{M_{ii}^{-1}}\xi_i = \sqrt{\Theta_{ii}^{-1}}\xi_i$ it is possible to rewrite the integrals $I(n, \boldsymbol{C}, M)$ and $I_{i,j}(n, \boldsymbol{C}, M)$ in the following way:

$$I(n, \boldsymbol{C}, M) = \frac{\sqrt{Det[\Theta]}}{(2\pi)^{n/2}} \int_{\xi_1 = C_1\sqrt{\Theta_{11}^{-1}}}^{\infty} \cdots \int_{\xi_n = C_n\sqrt{\Theta_{nn}^{-1}}}^{\infty} \exp\left[-\frac{1}{2}\left(\sum_{k=1}^{n} \xi_k^2 + 2\sum_{1 \le k < l \le n} \xi_k \xi_l \Theta_{kl}\right)\right] d\xi_n \ldots d\xi_1 \quad \text{and} \qquad \text{(B.77)}$$

$$I_{i,j}(n, \boldsymbol{C}, M) = \frac{\sqrt{Det[\Theta]}}{(2\pi)^{n/2}\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} \int_{\xi_1 = C_1\sqrt{\Theta_{11}^{-1}}}^{\infty} \cdots \int_{\xi_n = C_n\sqrt{\Theta_{nn}^{-1}}}^{\infty} \xi_i \xi_j \exp\left[-\frac{1}{2}\left(\sum_{k=1}^{n} \xi_k^2 + 2\sum_{1 \le k < l \le n} \xi_k \xi_l \Theta_{kl}\right)\right] d\xi_n \ldots d\xi_1$$

Differentiating $I(n, \boldsymbol{C}, M)$ given as the integral (B.77) with respect to Θ_{ij} gives us:

$\frac{\partial}{\partial \Theta_{ij}} I(n, \boldsymbol{C}, M) = \frac{\frac{\partial}{\partial \Theta_{ij}}\sqrt{Det[\Theta]}}{\sqrt{Det[\Theta]}} I(n, \boldsymbol{C}, M) - \sum_{k=1}^{n} C_k \frac{\frac{\partial}{\partial \Theta_{ij}}\sqrt{\Theta_{kk}^{-1}}}{\sqrt{\Theta_{kk}^{-1}}} I^k(n, \boldsymbol{C}, M) - \sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}} I_{i,j}(n, \boldsymbol{C}, M)$. Then solving for $I_{i,j}(n, \boldsymbol{C}, M)$ and simplifying gives:

$$I_{i,j}(n, \boldsymbol{C}, M) = \frac{1}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}}\left(-\frac{\partial}{\partial \Theta_{ij}} I(n, \boldsymbol{C}, M) - \sum_{k=1}^{n} C_k I^k(n, \boldsymbol{C}, M) \frac{\frac{\partial \Theta_{kk}^{-1}}{\partial \Theta_{ij}}}{2\Theta_{kk}^{-1}} + \frac{\frac{\partial}{\partial \Theta_{ij}} Det[\Theta]}{2Det[\Theta]} I(n, \boldsymbol{C}, M)\right) \qquad \text{(B.78)}$$

By using the results from lemma B.1 we have $\frac{\partial \Theta_{kk}^{-1}}{\partial \Theta_{ij}} = -2\Theta_{ki}^{-1}\Theta_{kj}^{-1}$ and therefore $-\frac{1}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} \frac{\frac{\partial \Theta_{kk}^{-1}}{\partial \Theta_{ij}}}{2\Theta_{kk}^{-1}} = \frac{1}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} \frac{\Theta_{ki}^{-1}\Theta_{kj}^{-1}}{\Theta_{kk}^{-1}} = \rho_{ik}\rho_{jk}$ by lemma B.2. Further by lemma B.4 $\frac{\frac{\partial}{\partial \Theta_{ij}} Det[\Theta]}{2Det[\Theta]} = \Theta_{ij}^{-1}$ so $\frac{1}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}}\left(\frac{\frac{\partial}{\partial \Theta_{ij}} Det[\Theta]}{2Det[\Theta]}\right) = \frac{\Theta_{ij}^{-1}}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} = \rho_{ij}$ by (B.5). Further

$$-\frac{1}{\sqrt{\Theta_{ii}^{-1}}\sqrt{\Theta_{jj}^{-1}}} \frac{\partial}{\partial \Theta_{ij}} I(n, \boldsymbol{C}, M) = -\sum_{1 \le k < l \le n} \frac{\partial}{\partial \rho_{kl}} I(n, \boldsymbol{C}, M) \frac{1}{\sqrt{M_{ii}^{-1}}\sqrt{M_{jj}^{-1}}} \frac{\partial \rho_{kl}}{\partial \Theta_{ij}}$$

$= \sum_{1 \le k < l \le n} \frac{\partial}{\partial \rho_{kl}} I(n, \boldsymbol{C}, M)[(\rho_{ik}\rho_{jl} + \rho_{jk}\rho_{il}) - \rho_{kl}(\rho_{ik}\rho_{jk} + \rho_{il}\rho_{jl})]$ by lemma B.3. Inserting these results in (B.78) yields:

$$I_{i,j}(n, \boldsymbol{C}, M) = \sum_{1 \le k < l \le n} [\rho_{ik}\rho_{jl} + \rho_{jk}\rho_{il} - \rho_{kl}(\rho_{ik}\rho_{jk} + \rho_{il}\rho_{jl})] \frac{\partial}{\partial \rho_{kl}} I(n, \boldsymbol{C}, M) + \sum_{k=1}^{n} C_k \rho_{ik}\rho_{jk} I^k(n, \boldsymbol{C}, M) + \rho_{ij} I(n, \boldsymbol{C}, M)$$

The result follows now from theorem B.1 and theorem B.2 (where we also use the fact that $I^{k,l} = I^{l,k}$) QED.

We shall also consider multinormal integral of type:

$$J_i(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} (\xi_i - C_i) f_n(\xi_1, \ldots, \xi_n; M) d\xi_n \ldots d\xi_1 \text{ and} \tag{B.79}$$

$$J_{i,j}(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} (\xi_i - C_i)(\xi_j - C_j) f_n(\xi_1, \ldots, \xi_n; M) d\xi_n \ldots d\xi_1 \tag{B.80}$$

for $i < j$.

These integrals may be given as a sum of multinormal integrals on standard form by applying theorem B.4 and theorem B.5. We find:

$$J_i(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \rho_{ik} I^k(n, \boldsymbol{C}, M) - C_i I(n, \boldsymbol{C}, M) = \sum_{k=1}^{n} \rho_{ik} f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) - C_i I(n, \boldsymbol{C}, M) \tag{B.81}$$

and

$$\begin{aligned} J_{i,j}(n, \boldsymbol{C}, M) &= \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} [\rho_{ik}(\rho_{jl} - \rho_{kl}\rho_{jk})] I^{k,l}(n, \boldsymbol{C}, M) - \sum_{k=1}^{n} (C_i \rho_{jk} + C_j \rho_{ik} - C_k \rho_{ik} \rho_{jk}) I^k(n, \boldsymbol{C}, M) \\ &\quad + (\rho_{ij} + C_i C_j) I(n, \boldsymbol{C}, M) \end{aligned} \tag{B.82}$$

$$\begin{aligned} &= \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} [\rho_{ik}(\rho_{jl} - \rho_{kl}\rho_{jk})] f_2(C_k, C_l, \rho_{kl}) I(n-2, \boldsymbol{C}^{k,l}, M^{k,l}) - \sum_{k=1}^{n} (C_i \rho_{jk} + C_j \rho_{ik} - C_k \rho_{ik} \rho_{jk}) f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) \\ &\quad + (\rho_{ij} + C_i C_j) I(n, \boldsymbol{C}, M) \end{aligned}$$

In many applications involving multinormal integrals the parameters will be functions of different variables. In chapter 4 we frequently apply the following corollary which follows directly from theorem B.1 and theorem B.2.

Corollary B.4. Suppose $C_i = C_i(t, z)$, $i = 1, \ldots, n$ are functions of z and t and $\rho_{ij} = \rho_{ij}(t)$, $1 \leq i < j \leq n$ all are functions of t, then the partial derivative of $f(t, z) = I(n, \boldsymbol{C}, M)$ is found as:

$$\frac{\partial f}{\partial z} = -\sum_{k=1}^{n} I^k(n, \boldsymbol{C}, M) \frac{\partial}{\partial z} C_k(t, z) = -\sum_{k=1}^{n} f_1(C_k) I(n-1, \boldsymbol{C}^k, M^k) \frac{\partial}{\partial z} C_k(t, z) \text{ and} \tag{B.83}$$

$$\frac{\partial f}{\partial t} = -\sum_{k=1}^{n} I^{k}(n, \boldsymbol{C}, M)\frac{\partial}{\partial t}C_k(t, z) + \sum_{1 \le k < l \le n} I^{k,l}(n, \boldsymbol{C}, M)\frac{d}{dt}\rho_{kl}(t)$$

$$= -\sum_{k=1}^{n} I^{k}(n, \boldsymbol{C}, M)\frac{\partial}{\partial t}C_k(t, z) + \frac{1}{2}\sum_{k=1}^{n}\sum_{\substack{l=1 \\ l \neq k}}^{n} I^{k,l}(n, \boldsymbol{C}, M)\frac{d}{dt}\rho_{kl}(t) \tag{B.84}$$

$$= -\sum_{k=1}^{n} f_1(C_k)I(n-1, \boldsymbol{C}^k, M^k)\frac{\partial}{\partial t}C_k(t, z) + \frac{1}{2}\sum_{k=1}^{n}\sum_{\substack{l=1 \\ l \neq k}}^{n} f_2(C_k, C_l, \rho_{kl})I(n-2, \boldsymbol{C}^{k,l}, M^{k,l})\frac{d}{dt}\rho_{kl}(t) \tag{B.85}$$

It is possible to obtain similar results as theorem B.1 for higher order partial derivatives of the parameters by applying theorem B.2 recursively. This leads to integrals of type (B.50) (or (B.51)) and we may use Theorem B.2 to find the derivative. These results can therefore be expressed in terms of partial derivatives of integrals of type $I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M)$ (given by (B.50)) with respect to the parameters. Below we only give the general result for the partial derivative with respect to the elements in the limit vector $\boldsymbol{C}$.

Theorem B.6. If $l \neq s_1, \ldots, s_k$, then

$$\frac{\partial}{\partial C_l} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = -I^{s_1, \ldots, s_k, l}(n, \boldsymbol{C}, M) \text{ and} \tag{B.86}$$

if $l = s_i$ for $1 \le i \le k$, then

$$\frac{\partial}{\partial C_{s_i}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = -\left(\sum_{j=1}^{k} C_{s_j}[M^{-1}_{s_1, \ldots, s_k}]_{ji}\right) I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) + \sum_{\substack{j=1 \\ j \neq s_1, \ldots, s_k}}^{n} a_{ji}^{s_1, \ldots, s_k} I^{s_1, \ldots, s_k, j}(n, \boldsymbol{C}, M) \tag{B.87}$$

Proof: The first part is obvious. We therefore take $l = s_i$ (for $1 \le i \le k$). By (B.51) we have:

$$\frac{\partial}{\partial C_{s_i}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = I(n-k, \boldsymbol{C}^{s_1, \ldots, s_k}, M^{s_1, \ldots, s_k})\frac{\partial}{\partial C_{s_i}} f_k(C_{s_1}, \ldots, C_{s_k}; M_{s_1, \ldots, s_k}) + f_k(C_{s_1}, \ldots, C_{s_k}; M_{s_1, \ldots, s_k})\frac{\partial}{\partial C_{s_i}} I(n-k, \boldsymbol{C}^{s_1, \ldots, s_k}, M^{s_1, \ldots, s_k})$$

By lemma B.5 (equation (B.12)) we have $\frac{\partial}{\partial C_{s_i}} f_k(C_{s_1}, \ldots, C_{s_k}; M_{s_1, \ldots, s_k}) = -\left(\sum_{j=1}^{k} C_{s_j}[M^{-1}_{s_1, \ldots, s_k}]_{ji}\right) f_k(C_{s_1}, \ldots, C_{s_k}; M_{s_1, \ldots, s_k})$ and multiplying with $I(n-k, \boldsymbol{C}^{s_1, \ldots, s_k}, M^{s_1, \ldots, s_k})$ and using (B.51) gives the first part of expression (B.87). By applying corollary B.2 (equation (B.83)) we have

$$\frac{\partial}{\partial C_{s_i}} I(n-k, \boldsymbol{C}^{s_1,\ldots,s_k}, M^{s_1,\ldots,s_k}) = -\sum_{\substack{j=1 \\ j \neq s_1,\ldots,s_k}}^{n} f_1(C_j^{s_1,\ldots,s_k}) I(n-k-1, \boldsymbol{C}^{s_1,\ldots,s_k,j}, M^{s_1,\ldots,s_k,j}) \left(\frac{\partial}{\partial C_{s_i}} C_j^{s_1,\ldots,s_k}\right).$$ Further we have

$$f_k(C_{s_1},\ldots,C_{s_k};M_{s_1,\ldots,s_k}) f_1(C_j^{s_1,\ldots,s_k}) = \sqrt{\frac{Det[M_{s_1,\ldots,s_k,j}]}{Det[M_{s_1,\ldots,s_k}]}} f_{k+1}(C_{s_1},\ldots,C_{s_k}, C_j;M_{s_1,\ldots,s_k,j}) = \sigma_j^{s_1,\ldots,s_k} f_{k+1}(C_{s_1},\ldots,C_{s_k}, C_j;M_{s_1,\ldots,s_k,j})$$

where $M_{s_1,\ldots,s_k,j} = \begin{bmatrix} M_{s_1,\ldots,s_k} & \overline{\rho_{s_1,\ldots,s_k;j}} \\ \rho_{s_1,\ldots,s_k;j} & 1 \end{bmatrix}$ is obtained from $M_{s_1,\ldots,s_k}$ by adding the row-vector $\rho_{s_1,\ldots,s_k;j} = (\rho_{s_1 j}, \ldots, \rho_{s_k j})$ (and corresponding column-vector $\overline{\rho_{s_1,\ldots,s_k;j}}$), and where we also apply (B.25). By (B.44) we also have $\frac{\partial}{\partial C_{s_i}} C_j^{s_1,\ldots,s_k} = -\frac{a_{ji}^{s_1,\ldots,s_k}}{\sigma_j^{s_1,\ldots,s_k}}$. By inserting the different expressions and applying (B.44) we then obtain the second part of (B.87). QED.

The corresponding result for the partial derivative of the multinormal integral with respect to the elements in the correlation matrix M is also possible to obtain by applying the following quite remarkable theorem for these types of integrals. The following theorem is more or less a direct implication of the results in lemma B.5.

Theorem B.7. For integrals of type $I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M)$ (given in (B.50)) we have the following result:

$$\frac{\partial}{\partial \rho_{lm}} I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) = \frac{\partial^2}{\partial C_l \partial C_m} I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) \text{ for } 1 \le l < m \le n \tag{B.88}$$

Proof: Differentiating (B.50) with respect to the correlations we have

$$\frac{\partial}{\partial \rho_{lm}} I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) = \int_{\xi_1 = C_1}^{\infty} \cdots \int_{\xi_n = C_n}^{\infty} \frac{\partial f_n}{\partial \rho_{lm}}(\xi_1,\ldots,\xi_{s_1-1}, C_{s_1}, \xi_{s_1+1}, \ldots, \xi_{s_k-1}, C_{s_k}, \xi_{s_k+1}, \ldots, \xi_n;M) d\xi_n \ldots d\xi_1 \tag{B.89}$$

where the integral is of dimension $n-k$ and does not involve the variable $\xi_{s_1}, \ldots, \xi_{s_k}$. Then by applying lemma B.5 (equation (B.13)) we have depending on the values of l and m, the following cases:

(i): Both $l, m \neq s_1, \ldots, s_k$ giving $\frac{\partial f_n}{\partial \rho_{lm}} = \frac{\partial^2 f_n}{\partial \xi_l \partial \xi_m}$. Integrating twice gives the result for this case since $\frac{\partial}{\partial \rho_{lm}} I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M) = I^{s_1,\ldots,s_k,l,m}(n, \boldsymbol{C}, M) = \frac{\partial^2}{\partial C_l \partial C_m} I^{s_1,\ldots,s_k}(n, \boldsymbol{C}, M)$.

(ii): If either $l = s_i$ and $m \neq s_1, \ldots, s_k$ or $m = s_j$ and $j \neq s_1, \ldots, s_k$ then $\frac{\partial f_n}{\partial \rho_{s_i m}} = \frac{\partial^2 f_n}{\partial \xi_m \partial C_{s_i}}$ or $\frac{\partial f_n}{\partial \rho_{l s_j}} = \frac{\partial^2 f_n}{\partial C_{s_j} \partial \xi_l}$. Integrating once we get either $\frac{\partial}{\partial \rho_{s_i m}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = -\frac{\partial}{\partial C_{s_i}} I^{s_1, \ldots, s_k, m}(n, \boldsymbol{C}, M) = \frac{\partial^2}{\partial C_{s_i} \partial C_m} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M)$ or

$$\frac{\partial}{\partial \rho_{l s_j}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = -\frac{\partial}{\partial C_{s_j}} I^{s_1, \ldots, s_k, l}(n, \boldsymbol{C}, M) = \frac{\partial^2}{\partial C_l \partial C_{s_j}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M).$$

(iii): The last case occurs when $l = s_i$ and $m = s_j$. In this case $\frac{\partial f_n}{\partial \rho_{s_i s_j}} = \frac{\partial^2 f_n}{\partial C_{s_i} \partial C_{s_j}}$ and therefore $\frac{\partial}{\partial \rho_{s_i s_j}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M) = \frac{\partial^2}{\partial C_{s_j} \partial C_{s_j}} I^{s_1, \ldots, s_k}(n, \boldsymbol{C}, M)$. QED

B.4 Some limits for Multinormal integrals

In this section we shall prove some limits that we quite frequently have applied in chapter 4. We consider an integral $I(n, \boldsymbol{C}, M)$ of dimension n where the limits $C_i = C_i(t)$, $i = 1, ..., n$ and the covariances $\rho_{ij} = \rho_{ij}(t)$, $1 \leq i < j \leq n$ all are functions of t.

Theorem B.8. If all the $C_i(t) \to -\infty$ when $t \to 0$ for $i = 1, ..., n$, then

$$\lim_{t \to 0} I(n, \boldsymbol{C}, M) = 1 \tag{B.90}$$

Proof: This is obvious since the integral is over a multidimensional probability density function.

Theorem B.9. If there is one k for which the $C_k(t) \to \infty$ when $t \to 0$, then

$$\lim_{t \to 0} I(n, \boldsymbol{C}, M) = 0 \tag{B.91}$$

Proof: By applying (B.52) we may write $I(n, \boldsymbol{C}, M)$ on the form:

$$I(n, \boldsymbol{C}, M) = \int_{\xi_k = C_k(t)}^{\infty} f_1(\xi_k) I(n-1, \boldsymbol{C}^k(\xi_k), M^k) d\xi_k \tag{B.92}$$

where the integration limits are given as $C_i^k(\xi_k) = \frac{C_i(t) - \rho_{ik}(t)\xi_k}{\sqrt{1-\rho_{ik}(t)^2}}$ for $i = 1, ..., n$, $i \neq k$ and the covariance matrix M^k is given by (B.58). Since $\left|I(n-1, \boldsymbol{C}^k(\xi_k), M^k)\right| \leq 1$ it follows that $|I(n, \boldsymbol{C}, M)| \leq \int_{\xi_k = C_k(t)}^{\infty} f_1(\xi_k) d\xi_k = \phi(C_k(t))$ where $\phi(x)$ is the standard normal integral. The result follows now by applying the asymptotic expansion of $\phi(x)$ for large x; $\phi(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ as $x \to \infty$.

Theorem B.10. If all the $C_i(t) \to C$, $i = 1, ..., n$ and $\rho_{ij}(t) \to 1$, $1 \leq i < j \leq n$ when $t \to 0$, then

$$\lim_{t \to 0} I(n, \boldsymbol{C}, M) = \phi(C) \tag{B.93}$$

Proof: Applying (B.92) we have:

$$I(n, \boldsymbol{C}, M) - \phi(C) = \int_{\xi_k = C_k(t)}^{\infty} f_1(\xi_k)(I(n-1, \boldsymbol{C}^k(\xi_k), M^k) - 1) d\xi_k + \phi(C) - \phi(C_k(t)).$$

If we for instance choose ε_1 sufficient small (and fixed), then we can write this difference as:

$$\int_{\xi_k = C+\varepsilon_1}^{\infty} f_1(\xi_k)(I(n-1, \boldsymbol{C}^k(\xi_k), M^k) - 1) d\xi_k + \int_{\xi_k = C_k(t)}^{C+\varepsilon_1} f_1(\xi_k) I(n-1, \boldsymbol{C}^k(\xi_k), M^k) d\xi_k + \phi(C) - \phi(C+\varepsilon),$$ giving

$$|I(n, \boldsymbol{C}, M) - \phi(C)| \leq \int_{\xi_k = C+\varepsilon_1}^{\infty} f_1(\xi_k) \left|I(n-1, \boldsymbol{C}^k(\xi_k), M^k) - 1\right| d\xi_k + \phi(C_k(t))) - \phi(C+\varepsilon_1) + (\phi(C) - \phi(C+\varepsilon_1))$$

For the integration limits $C_i^k(\xi_k)$ we have $C_i^k(\xi_k) = \frac{C_i(t) - \rho_{ik}(t)\xi_k}{\sqrt{1-\rho_{ik}(t)^2}} < \frac{C_i(t) - \rho_{ik}(t)(C+\varepsilon_1)}{\sqrt{1-\rho_{ik}(t)^2}}$, so therefore $\lim_{t \to 0} C_i^k(\xi_k) < -\varepsilon_1 \lim_{t \to 0} \frac{1}{\sqrt{1-\rho_{ik}(t)^2}} = -\infty$ for all $i = 1, ..., n$, $i \neq k$ when $\xi_k \geq (C+\varepsilon_1)$. By theorem B.8 we have $\lim_{t \to 0} I(n-1, \boldsymbol{C}^k(\xi_k), M^k) = 1$ when $\xi_k \geq (C+\varepsilon_1)$. We may therefore choose a second ε_2 so that $\left|(I(n-1, \boldsymbol{C}^k(\xi_k), M^k) - 1)\right| < \varepsilon_2$ for all $\xi_k \geq (C+\varepsilon_1)$ and t sufficient small. Similar we may also choose a third ε_3 so that $|C_k(t) - C| < \varepsilon_3$ for t sufficient small. Then by collecting the results above we finally may derive the following inequality:

$$|I(n, \boldsymbol{C}, M) - \phi(C)| < \varepsilon_2 \int_{\xi_k = C + \varepsilon_1}^{\infty} f_1(\xi_k) d\xi_k + \phi(C_k(t))) - \phi(C + \varepsilon_1) + (\phi(C) - \phi(C + \varepsilon_1)) \le \varepsilon_2 + \frac{1}{\sqrt{2\pi}}(\varepsilon_3 + 2\varepsilon_1) .$$

The result follows since all the three epsilons may be chosen arbitrary small (by also choosing t sufficient close to 0) QED.

The result will be quite different if some of the correlations $\rho_{ij}(t)$ tend to -1. In chapter 4 we need the following result:

Theorem B.11. If we have two integer n_1 and n_2 such that $1 \le n_1 < n_2 < n$ or $1 < n_1 < n_2 \le n$ and

$C_i(t) \to C$, $1 \le i \le n_1$ and $n_2 \le i \le n$ and $C_i(t) \to -C$, $n_1 \le i \le n_2$, if further

$\rho_{ij}(t) \to 1$, $1 \le i, j \le n_1$, $1 \le i \le n_1$ $n_2 \le i \le n$, $n_1 \le i, j \le n_2$, $n_2 \le i \le n$ $1 \le j \le n_1$ and $n_2 \le i, j \le n$ and

$\rho_{ij}(t) \to -1$, $1 \le i \le n_1$ $n_1 \le j \le n_2$, $n_1 \le i \le n_2$ $1 \le j \le n_2$, $n_1 \le i \le n_2$ $1 \le j \le n_1$ and $n_2 \le i \le n$ $n_1 \le j \le n_2$ when $t \to 0$, then

$$\lim_{t \to 0} I(n, \boldsymbol{C}, M) = 0 \tag{B.94}$$

(See figure B.1 for the structure of the correlation matrix $M = \lim_{t \to 0} \rho_{ij}(t)$.)

$$M = \begin{array}{c} \\ n_1 \\ \\ n_2 \\ \\ \end{array} \begin{array}{c} \begin{array}{ccc} n_1 & & n_2 \end{array} \\ \left[\begin{array}{c|c|c} 1 & -1 & 1 \\ \hline -1 & 1 & -1 \\ \hline 1 & -1 & 1 \end{array}\right] \end{array}$$

Figure B.1: The structure of the correlation matrix $M = \lim_{t \to 0} \rho_{ij}(t)$ *in theorem*

Proof: We pick one k for which $C_k(t) \to C$ (i.e. $1 \le i \le n_1$ or $n_2 \le i \le n$), then by (B.92) we have:

$I(n, \boldsymbol{C}, M) = \int_{\xi_k = C_k(t)}^{\infty} f_1(\xi_k) I(n-1, \boldsymbol{C}^k(\xi_k), M^k) d\xi_k$. If we for instance choose ε_1 sufficient small (and fixed) we split the last integral into two parts

$\int_{\xi_k = C+\varepsilon_1}^{\infty} f_1(\xi_k) I(n-1, \boldsymbol{C}^k(\xi_k), M^k) d\xi_k + \int_{\xi_k = C_k(t)}^{C+\varepsilon_1} f_1(\xi_k) I(n-1, \boldsymbol{C}^k(\xi_k), M^k) d\xi_k$, bounding $I(n, \boldsymbol{C}, M)$ by

$|I(n, \boldsymbol{C}, M)| \le \int_{\xi_k = C+\varepsilon_1}^{\infty} f_1(\xi_k) \left| I(n-1, \boldsymbol{C}^k(\xi_k), M^k) \right| d\xi_k + \phi(C_k(t)) - \phi(C+\varepsilon_1)$. We now choose one i such that $\rho_{ik}(t) \to -1$ (By the structure of the limiting matrix this will always be possible, by choosing $n_1 \le i \le n_2$, see figure B.1.) We then have $C_i(t) \to -C$ and for this particular integration limit $C_i^k(\xi_k)$ we have $C_i^k(\xi_k) = \frac{C_i(t) - \rho_{ik}(t)\xi_k}{\sqrt{1-\rho_{ik}(t)^2}} > \frac{C_i(t) - \rho_{ik}(t)(C+\varepsilon_1)}{\sqrt{1-\rho_{ik}(t)^2}}$. Therefore $\lim_{t \to 0} C_i^k(\xi_k) > \varepsilon_1 \lim_{t \to 0} \frac{1}{\sqrt{1-\rho_{ik}(t)^2}} = \infty$ when $\xi_k \ge (C+\varepsilon_1)$. By Theorem B.9 we have $\lim_{t \to 0} I(n-1, \boldsymbol{C}^k(\xi_k), M^k) = 0$ when $\xi_k \ge (C+\varepsilon_1)$. We may therefore choose a second ε_2 so that $\left| I(n-1, \boldsymbol{C}^k(\xi_k), M^k) \right| < \varepsilon_2$ for all $\xi_k \ge (C+\varepsilon_1)$ and t sufficient small. Similar we may also choose a third ε_3 so that $|C_k(t) - C| < \varepsilon_3$ for t sufficient small. Then by collecting the results above we finally may derive the following inequality:

$$|I(n, \boldsymbol{C}, M)| < \varepsilon_2 \int_{\xi_k = C+\varepsilon_1}^{\infty} f_1(\xi_k) d\xi_k + \phi(C_k(t))) - \phi(C+\varepsilon_1) < \varepsilon_2 + \frac{1}{\sqrt{2\pi}}(\varepsilon_3 + \varepsilon_1).$$

The result follows since all the three epsilons may be chosen arbitrary small (by also choosing t sufficient close to 0) QED.

Appendix C

The sign of the real part of the poles of the Laplace transforms in section 5.4

In this appendix we shall show that all the poles of the LST (5.37) for the excess times must have negative real part.

Theorem C.1. If $s = \kappa$ is a pole of (5.37), then $Re(\kappa) < 0$.

Proof: We take the form (5.65) of the Laplace transform as the starting point. If $s = \kappa$ is a pole in (5.65), then by (5.66):

$$Det[\hat{\Gamma}^u(\kappa)] = \prod_{i \geq j_C} \left[\frac{1}{\hat{H}_i(\kappa)}\right] Det[\boldsymbol{I} - \hat{\Pi}^u(\kappa)] = 0 \tag{C.1}$$

If $s = \kappa$ corresponds to a pole of $H_i(t)$, that is $\frac{1}{\hat{H}_i(\kappa)} = 0$ then $Re(\kappa) < 0$ since

$\left|\hat{H}_i(s)\right| = \int_{t=0}^{\infty} e^{-Re(s)t} H_i(t)dt \leq \int_{t=0}^{\infty} H_i(t)dt = m_i$ for $Re(s) \geq 0$ which gives that

$\left|\frac{1}{\hat{H}_i(s)}\right| \geq \frac{1}{m_i}$ for $Re(s) \geq 0$.

If $s = \kappa$ corresponds to a pole where $Det[\boldsymbol{I} - \hat{\Pi}^u(\kappa)] = 0$, this means that $\gamma = 1$ is an eigenvalue of the matrix $\hat{\Pi}^u(\kappa)$. Let $\boldsymbol{x}_\kappa^u = (x_{\kappa j_C+1}^u, x_{\kappa j_C+2}^u .....)$, $(\boldsymbol{x}_\kappa^u \neq \boldsymbol{0})$ be the corresponding left eigenvector; that is,

$$\boldsymbol{x}_\kappa^u = \boldsymbol{x}_\kappa^u \cdot \hat{\Pi}^u(\kappa) \tag{C.2}$$

Suppose that $Re(\kappa) \geq 0$; then by (C.2) we have

$$\sum_{j \geq j_C+1} \left| x^u_{\kappa j} \right| = \sum_{j \geq j_C+1} \left| \sum_{i \geq j_C+1} x^u_{\kappa i} \hat{\Pi}_{ij}^{\ u}(\kappa) \right| \leq \sum_{i \geq j_C+1} \left| x^u_{\kappa i} \right| \sum_{j \geq j_C+1} \left| \hat{\Pi}_{ij}^{\ u}(\kappa) \right| \tag{C.3}$$

From the definition of $\hat{\Pi}_{ij}^{\ u}(\kappa)$ and the assumption $Re(\kappa) \geq 0$ we have:

$$\left| \hat{\Pi}_{ij}^{\ u}(\kappa) \right| = \left| \int_{t=0}^{\infty} \Pi_{ij}(t) e^{-\kappa t} dt \right| \leq \int_{t=0}^{\infty} \Pi_{ij}(t) e^{-Re(\kappa)t} dt \leq \int_{t=0}^{\infty} \Pi_{ij}(t) dt = P_{ij}(0) \tag{C.4}$$

We shall assume that there are at least one state $j^* \leq j_C$ such that $P_{i^* j^*}(0) > 0$ for some $i^* \geq j_C + 1$. By this assumption and (C.3) and (C.4) we find:

$$\sum_{j \geq j_C+1} \left| x^u_{\kappa j} \right| \leq \sum_{i \geq j_C+1} \left| x^u_{\kappa i} \right| \sum_{j \geq j_C+1} P_{ij}(0) < \sum_{i \geq j_C+1} \left| x^u_{\kappa i} \right| \sum_{j \geq 0} P_{ij}(0) = \sum_{i \geq j_C+1} \left| x^u_{\kappa j} \right| .$$

The assumption $Re(\kappa) \geq 0$ therefore leads to the contradiction:

$\sum_{j \geq j_C+1} \left| x^u_{\kappa j} \right| < \sum_{i \geq j_C+1} \left| x^u_{\kappa j} \right|$, which implies that if $s = \kappa$ is a pole where $Det[\boldsymbol{I} - \hat{\Pi}^u(\kappa)] = 0$, then $Re(\kappa) < 0$. QED.

The same result will also apply for the poles of the LST of the excess volume (5.57). The location of possible poles are given by the corresponding equation to (C.1):

$$\prod_{i \geq j_C} \left[\frac{1}{\hat{G}_i^{\ u}(\kappa)} \right] Det[\boldsymbol{I} - \hat{\Xi}^u(\kappa)] = 0 \tag{C.5}$$

Recall that the Matrix $\hat{\Xi}^u(\zeta)$ is given by $\hat{\Xi}^u_{ij}(\zeta) = \hat{\Pi}_{ij}((ib - C)\zeta)$ for $i > j_C$ and $j > j_C$, and $\hat{G}^u_i(\zeta) = (ib - C)\hat{H}_i((ib - C)\zeta)$ for $i > j_C$. We can therefore perform exactly the same proof as above by observing that (C.4) also yields for $\hat{\Xi}^u(\zeta)$.

We may also apply the same method on the LSTs (5.38) and (5.58) by considering the matrices $\hat{\Pi}^l(s)$ and $\hat{\Xi}^l(\zeta)$.

Appendix D

Alternative expressions for the Laplace transforms in section 5.4.3

It is possible to rewrite the formulae for the LSTs for birth-death semi-Markov processes given in section 5.4.3. This can be done by rewriting the determinants (5.74), (5.76), (5.80) and (5.82), where we also make the corresponding matrices symmetric by pre multiplying by a given diagonal matrix and post multiplying by the corresponding inverse. We find:

$$\hat{f}_{T_k}(s) = \mu_{j_C+1} \frac{E^u_{j_C+2}(s)}{E^u_{j_C+1}(s)} \tag{D.1}$$

where $E^u_j(s) = Det[\boldsymbol{E}^u_j(s)]$ and the matrix $\boldsymbol{E}^u_j(s)$ is the symmetric tri-diagonal matrix:

$$\boldsymbol{E}^u_j(s) = \begin{bmatrix} \frac{1}{m_j \hat{f}_j(s)} & -\sqrt{\lambda_j \mu_{j+1}} & 0 & \dots & 0 & 0 & 0 \\ -\sqrt{\lambda_j \mu_{j+1}} & \frac{1}{m_{j+1} \hat{f}_{j+1}(s)} & -\sqrt{\lambda_{j+1} \mu_{j+2}} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\sqrt{\lambda_{N-2} \mu_{N-1}} & \frac{1}{m_{N-1} \hat{f}_{N-1}(s)} & -\sqrt{\lambda_{N-1} \mu_N} \\ 0 & 0 & 0 & \dots & 0 & -\sqrt{\lambda_{N-1} \mu_N} & \frac{1}{m_N \hat{f}_N(s)} \end{bmatrix} \text{ and} \tag{D.2}$$

$$\hat{f}_{S_k}(s) = \lambda_{j_C} \frac{E^l_{j_C-1}(s)}{E^l_{j_C}(s)} \tag{D.3}$$

where $E^l_j(s) = Det[\boldsymbol{E}^l_j(s)]$ and the matrix $\boldsymbol{E}^l_j(s)$ is the symmetric tri-diagonal matrix:

$$\boldsymbol{E}_j^l(s) = \begin{bmatrix} \frac{1}{m_0\hat{f}_0(s)} & -\sqrt{\lambda_0\mu_1} & 0 & \dots & 0 & 0 & 0 \\ -\sqrt{\lambda_0\mu_1} & \frac{1}{m_1\hat{f}_1(s)} & -\sqrt{\lambda_1\mu_2} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\sqrt{\lambda_{j-2}\mu_{j-1}} & \frac{1}{m_{j-1}\hat{f}_{j-1}(s)} & -\sqrt{\lambda_{j-1}\mu_j} \\ 0 & 0 & 0 & \dots & 0 & -\sqrt{\lambda_{j-1}\mu_j} & \frac{1}{m_j\hat{f}_j(s)} \end{bmatrix} \tag{D.4}$$

where $\lambda_0 = \frac{1}{m_0}$ and $\mu_i = \frac{d_i}{m_i}$, $\lambda_i = \frac{b_i}{m_i}$, $i = 1, ..., N-1$ and $\mu_N = \frac{1}{m_N}$.

We may also obtain corresponding recursion formulae by expanding the determinants as (5.77) and (5.78):

The LSTs for the excess volumes may also be rewritten as follows:

$$\hat{f}_{A_k}(\zeta) = \mu^{*u}_{j_C+1}\frac{\Gamma^u_{j_C+2}(\zeta)}{\Gamma^u_{j_C+1}(\zeta)} \tag{D.5}$$

where $\Gamma^u_j(s) = Det[\Gamma^u_j(s)]$ and the matrix $\Gamma^u_j(s)$ is the symmetric tri-diagonal matrix:

$$\Gamma^u_j(\zeta) = \begin{bmatrix} \frac{1}{\phi^u_j(\zeta)} & -\sqrt{\lambda^{*u}_j\mu^{*u}_{j+1}} & 0 & \dots & 0 & 0 & 0 \\ -\sqrt{\lambda^{*u}_j\mu^{*u}_{j+1}} & \frac{1}{\phi^u_{j+1}(\zeta)} & -\sqrt{\lambda^{*u}_{j+1}\mu^{*u}_{j+2}} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\sqrt{\lambda^{*u}_{N-2}\mu^{*u}_{N-1}} & \frac{1}{\phi^u_{N-1}(\zeta)} & -\sqrt{\lambda^{*u}_{N-1}\mu^{*u}_N} \\ 0 & 0 & 0 & \dots & 0 & -\sqrt{\lambda^{*u}_{N-1}\mu^{*u}_N} & \frac{1}{\phi^u_N(\zeta)} \end{bmatrix} \tag{D.6}$$

where $\phi^u_j(\zeta) = (jb-C)m_j\hat{f}_j((jb-C)\zeta)$, $\lambda^{*u}_j = \frac{\lambda_j}{jb-C}$ and $\mu^{*u}_j = \frac{\mu_j}{jb-C}$; $j > j_C$ and

$$\hat{f}_{V_k}(\zeta) = \lambda^{*l}_{j_C}\frac{\Gamma^l_{j_C-1}(\zeta)}{\Gamma^l_{j_C}(\zeta)} \tag{D.7}$$

where $\Gamma^l_j(s) = Det[\Gamma^l_j(s)]$ and the matrix $\Gamma^l_j(s)$ is the symmetric tri-diagonal matrix:

$$\Gamma_j^l(\zeta) = \begin{bmatrix} \frac{1}{\phi_0^l(\zeta)} & -\sqrt{\lambda^{*u}_0\mu^{*u}_1} & 0 & \dots & 0 & 0 & 0 \\ -\sqrt{\lambda^{*u}_0\mu^{*u}_1} & \frac{1}{\phi_1^l(\zeta)} & -\sqrt{\lambda^{u}_1\mu^{*u}_2} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\sqrt{\lambda^{*u}_{j-2}\mu^{*u}_{j-1}} & \frac{1}{\phi_{j-1}^l(\zeta)} & -\sqrt{\lambda^{*u}_{j-1}\mu^{*u}_j} \\ 0 & 0 & 0 & \dots & 0 & -\sqrt{\lambda^{*u}_{j-1}\mu^{*u}_j} & \frac{1}{\phi_j^l(\zeta)} \end{bmatrix} \tag{D.8}$$

where $\phi_j^l(\zeta) = (C-jb)m_j\hat{f}_j((C-jb)\zeta)$; $\lambda^{*}{}_j^{\,l} = \frac{\lambda_j}{C-jb}$ and $\mu^{*}{}_j^{\,l} = \frac{\mu_j}{C-jb}$; $j \le j_C$

The reason for dealing with symmetric tri-diagonal matrices is that this type of matrices has very nice mathematical properties which make it easy to calculate the eigenvalues. We refer to textbooks in linear algebra for a thorough treatment of the various topics, but we just mention some of the properties which we shall exploit and apply to the matrices above. (See for instance [Wilk65] for treatment of the topic.)

By using the results for tri-diagonal matrices we may write (D.1) as:

$$\hat{f}_{T_k}(s) = \mu_{j_C+1}\frac{\prod_{i=j_C+2}^{N}\chi_{ui}^{j_C+2}(s)}{\prod_{i=j_C+1}^{N}\chi_{ui}^{j_C+1}(s)} \tag{D.9}$$

where $\chi_{ui}^{j_C+2}(s)$, $i = j_C+2, \dots, N$ are the eigenvalues of the matrix $\boldsymbol{E}_{j_C+2}^u(s)$ and $\chi_{ui}^{j_C+1}(s)$, $i = j_C+1, \dots, N$ are the eigenvalues of the matrix $\boldsymbol{E}_{j_C+1}^u(s)$. Moreover, these eigenvalues are strictly separated that is:

$$\chi_{uN}^{j_C+1}(s) < \chi_{uN-1}^{j_C+2}(s) < \chi_{uN-1}^{j_C+1}(s) < \dots < \chi_{uj_C+2}^{j_C+1}(s) < \chi_{uj_C+2}^{j_C+2}(s) < \chi_{uj_C+1}^{j_C+1}(s) \tag{D.10}$$

The leading principal minor of order $N-j$ of $\boldsymbol{E}_{j_C+1}^u(s) - \chi\boldsymbol{I}$, that is

$$E_j^u(s,\chi) = Det[\boldsymbol{E}_j^u(s) - \chi\boldsymbol{I}], \tag{D.11}$$

satisfies the following recursion starting by defining $E_{N+1}^u(s,\chi) = 1$ and

$$E_N^u(s,\chi) = \frac{1}{m_N \hat{f}_N(s)} - \chi \tag{D.12}$$

$$E_j^u(s,\chi) = \left(\frac{1}{m_j \hat{f}_j(s)} - \chi\right) E_{j+1}^u(s,\chi) - \lambda_j \mu_{j+1} E_{j+2}^u(s,\chi) \text{ for } j = N-1, ..., j_C + 1 \tag{D.13}$$

Similarly by applying corresponding results for the tri-diagonal matrices give

$$\hat{f}_{S_k}(s) = \lambda_{j_C} \frac{\prod_{i=0}^{j_C-1} \chi_{li}^{j_C-1}(s)}{\prod_{i=0}^{j_C} \chi_{li}^{j_C}(s)} \tag{D.14}$$

where $\chi_{li}^{j_C-1}(s)$, $i = 0, ...j_C-1$ are the eigenvalues of matrix $\boldsymbol{E}_{j_C-1}^l(s)$ and $\chi_{li}^{j_C}(s)$, $i = 0, ...j_C$ are the eigenvalues of matrix $\boldsymbol{E}_{j_C}^l(s)$. Moreover, these eigenvalues are strictly separated that is:

$$\chi_{l0}^{j_C}(s) < \chi_{l0}^{j_C-1}(s) < \chi_{l1}^{j_C}(s) < .... < \chi_{lj_C-1}^{j_C}(s) < \chi_{lj_C-1}^{j_C-1}(s) < \chi_{lj_C}^{j_C}(s), \tag{D.15}$$

The leading principal minor of order j of $\boldsymbol{E}_{j_C}^l(s) - \chi \boldsymbol{I}$, that is

$$E_j^l(s,\chi) = Det[\boldsymbol{E}_j^l(s) - \chi \boldsymbol{I}] \tag{D.16}$$

satisfies the following recursion starting by defining $E_{-1}^l(s,\chi) = 1$ and then

$$E_0^l(s,\chi) = \frac{1}{m_0 \hat{f}_0(s)} - \chi \tag{D.17}$$

$$E_j^l(s,\chi) = \left(\frac{1}{m_j \hat{f}_j(s)} - \chi\right) E_{j-1}^l(s,\chi) - \lambda_{j-1} \mu_j E_{j-2}^u(s,\chi) \text{ for } j = 1, ..., j_C \tag{D.18}$$

For sake of completeness we also write down the corresponding results for the excess volumes. We find:

$$\hat{f}_{A_k}(\zeta) = \mu^*{}^u_{j_C+1} \frac{\prod_{i=j_C+2}^{N} \omega_{ui}^{j_C+2}(\zeta)}{\prod_{i=j_C+1}^{N} \omega_{ui}^{j_C+1}(\zeta)} \tag{D.19}$$

where $\omega_{ui}^{j_C+2}(\zeta),\ i = j_C+2, ..., N$ are the eigenvalues of the matrix $\Gamma^u_{j_C+2}(\zeta)$ and $\omega_{ui}^{j_C+1}(\zeta),\ i = j_C+1, ..., N$ are the eigenvalues of the matrix $\Gamma^u_{j_C+1}(\zeta)$. Moreover, these eigenvalues are strictly separated, that is:

$$\omega_{uN}^{j_C+1}(\zeta) < \omega_{uN-1}^{j_C+2}(\zeta) < \omega_{uN-1}^{j_C+1}(\zeta) < .... < \omega_{uj_C+2}^{j_C+1}(\zeta) < \omega_{uj_C+2}^{j_C+2}(\zeta) < \omega_{uj_C+1}^{j_C+1}(\zeta) \tag{D.20}$$

The leading principal minor of order $N-j$ of $\Gamma^u_{j_C+1}(\zeta) - \omega \boldsymbol{I}$, that is

$$\Gamma^u_j(\zeta, \omega) = Det[\Gamma^u_j(\zeta) - \omega \boldsymbol{I}] \tag{D.21}$$

satisfies the following recursion starting by defining $\Gamma^u_{N+1}(\zeta, \omega) = 1$ and

$$\Gamma^u_N(\zeta, \omega) = \frac{1}{\phi^u_N(\zeta)} - \omega \tag{D.22}$$

$$\Gamma^u_j(\zeta, \omega) = \left(\frac{1}{\phi^u_j(\zeta)} - \omega\right)\Gamma^u_{j+1}(\zeta, \omega) - \lambda^*{}^u_j \mu^*{}^u_{j+1} \Gamma^u_{j+2}(\zeta, \omega) \text{ for}$$

$$j = N-1, ..., j_C+1 \tag{D.23}$$

Similarly by applying corresponding results for the tri-diagonal matrices give

$$\hat{f}_{V_k}(\zeta) = \lambda^*{}^l_{j_C} \frac{\prod_{i=0}^{j_C-1} \omega_{li}^{j_C-1}(\zeta)}{\prod_{i=0}^{j_C} \omega_{li}^{j_C}(\zeta)} \tag{D.24}$$

where $\omega_{li}^{j_C-1}(\zeta),\ i = 0, ...j_C-1$ are the eigenvalues of the matrix $\Gamma^l_{j_C-1}(s)$ and $\omega_{li}^{j_C}(\zeta)$, $i = 0, ...j_C$ are the eigenvalues of the matrix $\Gamma^l_{j_C}(s)$. Moreover, these eigenvalues are strictly separated, that is:

$$\omega_{l0}^{j_C}(\zeta) < \omega_{l0}^{j_C-1}(\zeta) < \omega_{l1}^{j_C}(\zeta) < .... < \omega_{lj_C-1}^{j_C}(\zeta) < \omega_{lj_C-1}^{j_C-1}(\zeta) < \omega_{lj_C}^{j_C}(\zeta) \tag{D.25}$$

The leading principal minor of order j of $\Gamma_{j_C}^{l}(\zeta) - \omega \boldsymbol{I}$, that is

$$\Gamma_j^l(\zeta, \omega) \;=\; Det[\Gamma_j^l(s) - \omega \boldsymbol{I}] \tag{D.26}$$

satisfies the following recursion starting by defining $\Gamma_{-1}^{l}(\zeta, \omega) \;=\; 1$ and then

$$\Gamma_0^l(\zeta, \omega) \;=\; \frac{1}{\phi_0^l(\zeta)} - \omega \tag{D.27}$$

$$\Gamma_j^l(\zeta, \omega) \;=\; \left(\frac{1}{\phi_j^l(\zeta)} - \omega\right)\Gamma_{j-1}^l(\zeta, \omega) - \lambda^*{}_{j-1}^{l}\mu^*{}_j^l\Gamma_{j-2}^u(\zeta, \omega) \;\text{ for } j \;=\; 1, ..., j_C \tag{D.28}$$

Numerically all the different eigenvalues may be calculating by using the method of bisection [Wilk65] by applying the Sturm sequence property of the sequences

$$\left\{E_N^u(s, \chi), E_{N-1}^u(s, \chi), ..., E_{j_C+1}^u(s, \chi)\right\}, \quad \left\{\Gamma_N^u(\zeta, \omega), \Gamma_{N-1}^u(\zeta, \omega), ..., \Gamma_{j_C+1}^u(\zeta, \omega)\right\} \quad \text{and}$$

$$\left\{E_0^l(s, \chi), E_1^l(s, \chi), ..., E_{j_C}^l(s, \chi)\right\}, \left\{\Gamma_0^l(\zeta, \omega), \Gamma_1^l(\zeta, \omega), ..., \Gamma_{j_C}^l(\zeta, \omega)\right\} \text{ for fixed } s \text{ and } \zeta.$$

Due to the continuity of the eigenvalues (as functions of s and ζ), the strict separation by (D.10), (D.15), (D.20) and (D.25) yields for all values of s and ζ. We may exploit these results in various directions to make statements concerning the location of the zeros of the eigenvalues.

One important implication is that we may find the dominating root in the various transforms by looking at the largest root with the smallest eigenvalue. That is we must look for the largest roots of $\chi_{uN}^{j_C+1}(s)$, $\omega_{uN}^{j_C+1}(\zeta)$, and $\chi_{l0}^{j_C}(s)$, $\omega_{l0}^{j_C}(\zeta)$.

D.1 Birth-death processes with exponential sojourn times

The main reason for rewriting the transforms on the specific form by the matrices (D.2), (D.4), (D.6) and (D.8) is that it will greatly simplify the expressions for birth-death process with exponentially distributed sojourn times in the different states. In this case we have $m_j \;=\; \dfrac{1}{\lambda_j + \mu_j}$ and $\hat{f}_i(s) \;=\; \dfrac{\lambda_j + \mu_j}{s + \lambda_j + \mu_j}$. We obtain:

$\frac{1}{m_j \hat{f}_j(s)} = \lambda_j + \mu_j + s$ and $\frac{1}{\phi_j^u(\zeta)} = \lambda^*{}_j^u + \mu^*{}_j^u + \zeta$ and $\frac{1}{\phi_j^l(\zeta)} = \lambda^*{}_j^l + \mu^*{}_j^l + \zeta$. By these substitutions in the various parts above we see that the (s, χ) and (ζ, ω) dependencies only appear through the differences $\chi - s$ and $\omega - \zeta$ and the corresponding eigenvalues will therefore be constants. We find:

$$\chi_{ui}^{j_C+1}(s) = \gamma_{ui}^{j_C+1} + s \tag{D.29}$$

where $\gamma_{ui}^{j_C+1}$, $i = j_C + 1, ..., N$ are the eigenvalues of the tri-diagonal matrix

$$\boldsymbol{F}_{j_C+1}^u = \begin{bmatrix} \lambda_{j_C+1} + \mu_{j_C+1} & -\sqrt{\lambda_{j_C+1}\mu_{j_C+2}} & 0 & ... & 0 & 0 & 0 \\ -\sqrt{\lambda_{j_C+1}\mu_{j_C+2}} & \lambda_{j_C+2} + \mu_{j_C+2} & -\sqrt{\lambda_{j_C+2}\mu_{j_C+3}} & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -\sqrt{\lambda_{N-2}\mu_{N-1}} & \lambda_{N-1} + \mu_{N-1} & -\sqrt{\lambda_{N-1}\mu_N} \\ 0 & 0 & 0 & ... & 0 & -\sqrt{\lambda_{N-1}\mu_N} & \lambda_N + \mu_N \end{bmatrix}, \tag{D.30}$$

$$\chi_{li}^{j_C}(s) = \gamma_{li}^{j_C} + s \tag{D.31}$$

where $\gamma_{li}^{j_C}$, $i = 0, ...j_C$ are the eigenvalues of the tri-diagonal matrix

$$\boldsymbol{F}_{j_C}^l = \begin{bmatrix} \lambda_0 + \mu_0 & -\sqrt{\lambda_0\mu_1} & 0 & ... & 0 & 0 & 0 \\ -\sqrt{\lambda_0\mu_1} & \lambda_1 + \mu_1 & -\sqrt{\lambda_1\mu_2} & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -\sqrt{\lambda_{j_C-2}\mu_{j_C-1}} & \lambda_{j_C-1} + \mu_{j_C-1} & -\sqrt{\lambda_{j_C-1}\mu_{j_C}} \\ 0 & 0 & 0 & ... & 0 & -\sqrt{\lambda_{j_C-1}\mu_{j_C}} & \lambda_{j_C} + \mu_{j_C} \end{bmatrix} \tag{D.32}$$

The different eigenvalues may be obtained by the method of bisection applying the Sturm sequence property of the sequences of leading minors [Wilk65]. The leading principal minor of order $N-j$ of $\boldsymbol{F}_{j_C+1}^u - \chi\boldsymbol{I}$ given by $F_j^u(\chi) = Det[\boldsymbol{F}_j^u - \chi\boldsymbol{I}]$ satisfies the following recursion starting by defining $F_{N+1}^u(\chi) = 1$ and

$$F_N^u(\chi) = \lambda_N + \mu_N - \chi \tag{D.33}$$

$$F_j^u(\chi) = (\lambda_j + \mu_j - \chi)F_{j+1}^u(\chi) - \lambda_j\mu_{j+1}F_{j+2}^u(\chi) \text{ for } j = N-1, ..., j_C + 1 \tag{D.34}$$

Further the leading principal minor of order j of $\boldsymbol{F}_{j_C}^l - \chi\boldsymbol{I}$ given by $F_j^l(\chi) = Det[\boldsymbol{F}_j^l - \chi\boldsymbol{I}]$ satisfies the following recursion starting by defining $F_{-1}^l(\chi) = 1$ and

$$F_0^l(\chi) = \lambda_0 + \mu_0 - \chi \tag{D.35}$$

$$F_j^l(\chi) = (\lambda_j + \mu_j - \chi)F_{j-1}^l(\chi) - \lambda_{j-1}\mu_j F_{j-2}^l(\chi) \quad \text{for } j = 1, ..., j_C \tag{D.36}$$

The corresponding results for the excess volume are:

$$\omega_{ui}^{j_C+1}(\zeta) = \gamma^{*j_C+1}_{ui} + \zeta \text{ and } \omega_{li}^{j_C}(\zeta) = \gamma^{*j_C}_{li} + \zeta \tag{D.37}$$

where $\gamma^{*j_C+1}_{ui}$, $i = j_C + 1, ..., N$ and $\gamma^{*j_C}_{li}$, $i = 0, ...j_C$ are eigenvalues of the matrices $\boldsymbol{F}^{*u}_{j_C+1}$ obtained from (D.30) by simply replacing λ_j by λ^{*u}_j and μ_j by μ^{*u}_j and $\boldsymbol{F}^{*l}_{j_C}$ obtained from (D.32) by replacing λ_j by λ^{*l}_j and μ_j by μ^{*l}_j.

We end this appendix by writing down the final result obtained inverting the transforms when we know all the eigenvalues, we find for the different PDFs for the different excess variables (by taking partial fractions expansion of the LSTs):

$$f_{T_k}(t) = \mu_{j_C+1} \sum_{j=j_C+1}^{N} a_{uj}^{j_C+1} e^{-\gamma_{uj}^{j_C+1} t} \text{ where } a_{uj}^{j_C+1} = \frac{\prod_{i=j_C+2}^{N} \left(\gamma_{ui}^{j_C+2} - \gamma_{uj}^{j_C+1}\right)}{\prod_{i=j_C+1, i\neq j}^{N} \left(\gamma_{ui}^{j_C+1} - \gamma_{uj}^{j_C+1}\right)} \tag{D.38}$$

and

$$f_{S_k}(t) = \lambda_{j_C} \sum_{j=0}^{j_C} a_{lj}^{j_C} e^{-\gamma_{lj}^{j_C} t} \text{ where } a_{lj}^{j_C} = \frac{\prod_{i=0}^{j_C-1} \left(\gamma_{li}^{j_C-1} - \gamma_{lj}^{j_C}\right)}{\prod_{i=0, i\neq j}^{j_C} \left(\gamma_{li}^{j_C} - \gamma_{lj}^{j_C}\right)} \tag{D.39}$$

and further for the excess volume:

$$f_{A_k}(x) = \mu^*_{j_C+1} \sum_{j=j_C+1}^{N} a^{*j_C+1}_{uj} e^{-\gamma^{*j_C+1}_{uj} x} \text{ where}$$

$$a^{*j_C+1}_{uj} = \frac{\prod_{i=j_C+2}^{N} \left(\gamma^{*j_C+2}_{ui} - \gamma^{*j_C+1}_{uj}\right)}{\prod_{i=j_C+1,\, i\neq j}^{N} \left(\gamma^{*j_C+1}_{ui} - \gamma^{*j_C+1}_{uj}\right)} \tag{D.40}$$

and

$$f_{V_k}(x) = \lambda^*_{j_C} \sum_{j=0}^{j_C} a^{*j_C}_{lj} e^{-\gamma^{*j_C}_{lj} x} \text{ where } a^{*j_C}_{lj} = \frac{\prod_{i=0}^{j_C-1} \left(\gamma^{*j_C-1}_{li} - \gamma^{*j_C}_{lj}\right)}{\prod_{i=0,\, i\neq j}^{j_C} \left(\gamma^{*j_C}_{li} - \gamma^{*j_C}_{lj}\right)} \tag{D.41}$$

and where the eigenvalues are ordered so that

$$0 < \gamma^{j_C+1}_{uN} < \gamma^{j_C+2}_{uN} < \gamma^{j_C+1}_{uN-1} < .... < \gamma^{j_C+1}_{uj_C+2} < \gamma^{j_C+2}_{uj_C+2} < \gamma^{j_C+1}_{uj_C+1},$$

$$0 < \gamma^{j_C}_{l0} < \gamma^{j_C-1}_{l0} < \gamma^{j_C}_{l1} < .... < \gamma^{j_C}_{lj_C-1} < \gamma^{j_C-1}_{lj_C-1} < \gamma^{j_C}_{lj_C}$$

and

$$0 < \gamma^{*j_C+1}_{uN} < \gamma^{*j_C+2}_{uN} < \gamma^{*j_C+1}_{uN-1} < .... < \gamma^{*j_C+1}_{uj_C+2} < \gamma^{*j_C+2}_{uj_C+2} < \gamma^{*j_C+1}_{uj_C+1}$$

$$0 < \gamma^{*j_C}_{l0} < \gamma^{*j_C-1}_{l0} < \gamma^{*j_C}_{l1} < .... < \gamma^{*j_C}_{lj_C-1} < \gamma^{*j_C-1}_{lj_C-1} < \gamma^{*j_C}_{lj_C}.$$

Appendix E

Asymptotics of the excess distributions and first passage times for the U-O process

E.1 Asymptotics for the Parabolic Cylinder Function $D_{\zeta^2 + C\zeta}(x + 2\zeta)$ for large ζ in terms of Airy functions

We have [Abra70], (19.5.1 page 687) that $D_a(z) = U\left(-\frac{1}{2} - a, z\right)$ where the $U(\alpha, z)$ may be written by the following integral:

$$U(\alpha, z) = \frac{\Gamma\left(\frac{1}{2} - \alpha\right)}{2\pi i} e^{-\frac{1}{4}z^2} \int_H e^{zs - \frac{1}{2}s^2} s^{a - \frac{1}{2}} ds \tag{E.1}$$

Where H is the so called Hankel contour (see figure E.1)

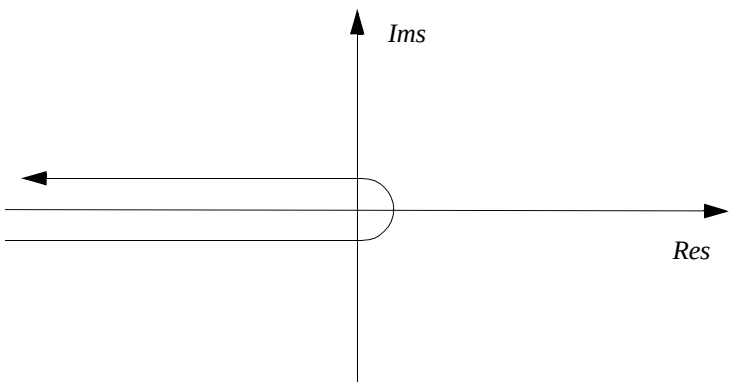

Figure E.1: The Hankel contour.

We shall find asymptotics for the functions

$$f_1(\zeta, x, C) \;=\; D_{\zeta^2 + C\zeta}(x + 2\zeta) \;=\; U\left(-\frac{1}{2} - C\zeta - \zeta^2, x + 2\zeta\right) \text{ and} \tag{E.2}$$

$$f_2(\zeta, x, C) \;=\; D_{\zeta^2 + C\zeta - 1}(x + 2\zeta) \;=\; U\left(\frac{1}{2} - C\zeta - \zeta^2, x + 2\zeta\right) \tag{E.3}$$

as $\zeta \to \infty$. We shall follow the line in [Haga89] for the expansion.

By introducing $z \;=\; x - C$ and $\lambda \;=\; \zeta + \frac{C}{2}$ and $f_1(\zeta, x, C) \;=\; g_1\left(\zeta + \frac{C}{2}, x - C, C\right)$ we obtain

$$g_1(\lambda, y, C) \;=\; \frac{\Gamma\left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)}{2\pi i} e^{-\frac{1}{4}(y + 2\lambda)^2} \int_H e^{(y + 2\lambda)s - \frac{1}{2}s^2} s^{-\left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)} ds \tag{E.4}$$

Following [Haga89] we introduce the following substitutions:

$z \;=\; \lambda^{\frac{1}{3}} y$ and $s \;=\; \lambda + \lambda^{\frac{1}{3}} t$ and $g_1(\lambda, y, C) \;=\; h_1\left(\lambda, \lambda^{\frac{1}{3}} y, C\right)$ giving:

$$h_1(\lambda, z, C) \;=\; \frac{\Gamma\left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)}{2\pi i} \lambda^{\frac{1}{3}} e^{-\frac{1}{4}\left(\lambda^{-\frac{1}{3}} z + 2\lambda\right)^2} \int_{H'} e^{F(\lambda, z, t)} dt \text{ with the exponent} \tag{E.5}$$

$F(\lambda, z, t) \;=\; \left(\lambda^{-\frac{1}{3}} z + 2\lambda\right)\left(\lambda + \lambda^{\frac{1}{3}} t\right) - \frac{1}{2}\left(\lambda + \lambda^{\frac{1}{3}} t\right)^2 - \left(\log\lambda + \log\left(1 + \lambda^{-\frac{2}{3}} t\right)\right)\left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)$, (and where H' also is a Hankel contour). Expanding the exponent to order $\lambda^{-\frac{2}{3}}$ for large λ gives:

$F(\lambda, z, t) \;=\; z\lambda^{\frac{2}{3}} + \frac{3}{2}\lambda^2 - \left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)\log(\lambda) + zt - \frac{t^3}{3} + \left(\frac{t^4}{4} - \left(1 - \frac{C^2}{4}\right)t\right)\lambda^{-\frac{2}{3}} + o\left(\lambda^{-\frac{2}{3}}\right)$. Collecting terms we find:

$$g_1(\lambda, y, C) \;=\; \frac{G(\lambda, C)}{2\pi i} \int_{H'} e^{\left(zt - \frac{t^3}{3}\right) + \lambda^{-\frac{2}{3}}\left\{\left(\frac{t^4}{4} - \frac{z^4}{4}\right) - \left(1 - \frac{C^2}{4}\right)t\right\} + o\left(\lambda^{-\frac{2}{3}}\right)} ds \tag{E.6}$$

where $G(\lambda, C) \;=\; \Gamma\left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)\lambda^{\frac{1}{3}} e^{\frac{1}{2}\lambda^2 - \left(\lambda^2 + 1 - \left(\frac{C}{2}\right)^2\right)\log\lambda}$. Expanding further gives

$$g_1(y, \lambda, C) \;=\; \frac{G(\lambda, C)}{2\pi i} \int_{H'} e^{\left(zt - \frac{t^3}{3}\right)} \left(1 + \lambda^{-\frac{2}{3}}\left\{\left(\frac{t^4}{4} - \frac{z^4}{4}\right) - \left(1 - \frac{C^2}{4}\right)t\right\} + o\left(\lambda^{-\frac{2}{3}}\right)\right) ds \;=\; G(\lambda, C)\left\{Ai(z) + \lambda^{-\frac{2}{3}}\left(I(z) + \left(1 - \frac{C^2}{4}\right)Ai'(z)\right) + o\left(\lambda^{-\frac{2}{3}}\right)\right\}$$

where $Ai(z) = \frac{1}{2\pi i}\int_{H'} e^{\left(zt-\frac{t^3}{3}\right)} ds$ is the Airy function. (This is seen by transforming the contour into the imaginary axis giving $Ai(z) = \frac{1}{\pi}\int_0^\infty \cos\left(zt+\frac{t^3}{3}\right) ds$ which is an integral representation of the Airy function by [Abra70], (10.4.33 page 447). Further the integral

$$I(z) = \frac{1}{2\pi i}\int_{H'}\left(\frac{t^4}{4}-\frac{z^4}{4}\right)e^{\left(zt-\frac{t^3}{3}\right)} ds = -\frac{1}{2\pi i}\int_{H'}(z(z-t^2))e^{\left(zt-\frac{t^3}{3}\right)} ds - \frac{1}{2\pi i}\int_{H'}(t^2(z-t^2))e^{\left(zt-\frac{t^3}{3}\right)} ds$$

Then integrating by parts we find $I(z) = \frac{1}{2}\frac{1}{2\pi i}\int_{H'} te^{\left(zt-\frac{t^3}{3}\right)} ds = \frac{1}{2}Ai'(z)$. Collecting the results above (and transforming back to the original parameters gives:

$$D_{\zeta^2+C\zeta}(x+2\zeta) = H_1(\zeta, C)\left\{Ai\left(\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}}(x-C)\right) - \left(\left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\zeta+\frac{C}{2}\right)^{-\frac{2}{3}}Ai'\left(\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}}(x-C)\right) + o\left(\left(\zeta+\frac{C}{2}\right)^{-\frac{2}{3}}\right)\right)\right\} \tag{E.7}$$

where $H_1(\zeta, C) = G\left(\zeta+\frac{C}{2}, C\right) = \Gamma(\zeta^2+C\zeta+1)\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}} e^{\frac{1}{2}\left(\zeta+\frac{C}{2}\right)^2-(\zeta^2+C\zeta+1)\log\left(\zeta+\frac{C}{2}\right)}$. (E.8)

By applying the asymptotic formula for the Gamma Function we find that

$$H_1(\zeta, C) \sim (2\pi)^{\frac{1}{2}}\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}} e^{-\frac{1}{2}\left(\zeta+\frac{C}{2}\right)^2+(\zeta^2+C\zeta)\log\left(\zeta+\frac{C}{2}\right)} \quad \text{as } \zeta\to\infty \tag{E.9}$$

By applying exactly the same asymptotic procedure for $f_2(x,\zeta) = D_{\zeta^2+C\zeta-1}(x+2\zeta)$ we find:

$$D_{\zeta^2+C\zeta-1}(x+2\zeta) = H_2(\zeta, C)\left\{Ai\left(\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}}(x-C)\right) + \left(\frac{1}{2}+\frac{C^2}{4}\right)\left(\zeta+\frac{C}{2}\right)^{-\frac{2}{3}}Ai'\left(\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}}(x-C)\right) + o\left(\left(\zeta+\frac{C}{2}\right)^{-\frac{2}{3}}\right)\right\} \tag{E.10}$$

where $H_2(\zeta, C) = \Gamma(\zeta^2+C\zeta)\left(\zeta+\frac{C}{2}\right)^{\frac{1}{3}} e^{\frac{1}{2}\left(\zeta+\frac{C}{2}\right)^2-(\zeta^2+C\zeta)\log\left(\zeta+\frac{C}{2}\right)}$ (E.11)

and with the following asymptotics:

$$H_2(\zeta, C) \sim (2\pi)^{\frac{1}{2}}\left(\zeta + \frac{C}{2}\right)^{\frac{1}{3}} e^{-\frac{1}{2}\left(\zeta + \frac{C}{2}\right)^2 + (\zeta^2 + C\zeta - 1)\log\left(\zeta + \frac{C}{2}\right)} \quad \text{as } \zeta \to \infty \tag{E.12}$$

It turns out that the asymptotic expansions (E.7) and (E.10) yield in the whole ζ excluding the negative real axis. The reason for this is that it is not possible to express the asymptotics by a single Airy-function as in (E.7) and (E.10). However this is a particular interesting case since the poles of the LST are located along the negative real axis. We shall follow along the course as described in the paper of Olver [Olve59], and find the desired asymptotic expansion by applying the connection formulae for the Parabolic Cylinder functions [Grad94] (9.248 page 1094) giving (where we also have used the formula [Grad94], (8.334 page 946) $D_p(z) = \cos(\pi p)D_p(-z) - \sin(\pi p)\frac{\Gamma(1+p)}{\sqrt{2\pi}}\left(e^{-i\frac{\pi}{2}(p+1)}D_{-p-1}(iz) + e^{i\frac{\pi}{2}(p+1)}D_{-p-1}(-iz)\right)$. By letting $\eta = -\zeta$ in (E.2) and (E.3) we are interested in the asymptotics of

$$u_1(\eta, x, C) = f_1(-\eta, x, C) = D_{\eta^2 - C\eta}(x - 2\eta) \text{ and} \tag{E.13}$$

$$u_2(\eta, x, C) = f_2(-\eta, x, C) = D_{\eta^2 - C\eta - 1}(x - 2\eta) \tag{E.14}$$

for $|\eta| \to \infty$ and $\arg(\eta) \le \frac{\pi}{2}$. Using the connection formula above gives:

$$u_1(\eta, x, C) = \cos(\pi(\eta^2 - C\eta))D_{\eta^2 - C\eta}(-x + 2\eta)$$

$$-\sin(\pi(\eta^2 - C\eta))\frac{\Gamma(\eta^2 - C\eta + 1)}{\sqrt{2\pi}}\left(e^{-i\frac{\pi}{2}(\eta^2 - C\eta + 1)}D_{-\eta^2 + C\eta - 1}(ix - 2i\eta) + e^{i\frac{\pi}{2}(\eta^2 - C\eta + 1)}D_{-\eta^2 + C\eta - 1}(-ix + 2i\eta)\right).$$

Written in terms of complex arguments this may be given as:

$$u_1(\eta, x, C) = \cos(\pi(\eta^2 - C\eta))f_1(\eta, -x, -C)$$

$$-\sin(\pi(\eta^2 - C\eta))\frac{\Gamma(\eta^2 - C\eta + 1)}{\sqrt{2\pi}}\left(e^{-i\frac{\pi}{2}(\eta^2 - C\eta + 1)}f_2(-i\eta, ix, iC) + e^{i\frac{\pi}{2}(\eta^2 - C\eta + 1)}f_2(i\eta, -ix, -iC)\right).$$

By direct substitution in the given asymptotic formulae (E.7) and (E.10) and then expanding gives:

$$\frac{\Gamma(\eta^2 - C\eta + 1)}{\sqrt{2\pi}}e^{-i\frac{\pi}{2}(\eta^2 - C\eta + 1)}f_2(-i\eta, ix, iC) =$$

$$H_1(\eta, -C)e^{-i\frac{\pi}{6}}\left\{Ai\left(-e^{-i\frac{2\pi}{3}}\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) - e^{-i\frac{2\pi}{3}}\left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}Ai'\left(-e^{-i\frac{2\pi}{3}}\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) + o\left(\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\right)\right\}$$

Similar for $\frac{\Gamma(\eta^2 - C\eta + 1)}{\sqrt{2\pi}}e^{i\frac{\pi}{2}(\eta^2 - C\eta + 1)}f_2(i\eta, -ix, -iC)$ yields the complex conjugate. Then by applying the following relation between the Airy functions [Abra70], (10.4.6 page 446)

$Bi(z) = e^{-i\frac{\pi}{6}} Ai\left(e^{-i\frac{2\pi}{3}} z\right) + e^{i\frac{\pi}{6}} Ai\left(e^{i\frac{2\pi}{3}} z\right)$ and its differentiated form, we obtain the following asymptotic expansion:

$$u_1(\eta, x, C) = H_1(\eta, -C)\left\{\cos(\pi(\eta^2 - C\eta))Ai\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) - \left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}} Ai'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right)\right.$$

$$\left. -\sin(\pi(\eta^2 - C\eta))\left(Bi\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) - \left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}} Bi'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right)\right) + o\left(\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\right)\right\} \quad \text{(E.15)}$$

The corresponding result for the second function $u_2(\eta, x, C) = D_{\eta^2 - C\eta - 1}(x - 2\eta)$ is found similarly and we get:

$$u_2(\eta, x, C) = H_2(\eta, -C)\left\{\cos(\pi(\eta^2 - C\eta - 1))\left(Ai\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) + \left(\frac{1}{2} + \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}} Ai'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right)\right)\right.$$

$$\left. -\sin(\pi(\eta^2 - C\eta - 1))\left(Bi\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right) + \left(\frac{1}{2} + \frac{C^2}{4}\right)\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}} Bi'\left(-\left(\eta - \frac{C}{2}\right)^{\frac{1}{3}}(x - C)\right)\right) + o\left(\left(\eta - \frac{C}{2}\right)^{-\frac{2}{3}}\right)\right\} \quad \text{(E.16)}$$

e.2 Asymptotics for the conditional excess volume for small arguments

We shall investigate the conditional density function $f_{A^x}(x, z)$ for small t. Recall from (4.107) that the LST is given as

$$\hat{f}_{A^x}(x, \zeta) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \frac{D_{\zeta^2 + C\zeta}(x + 2\zeta)}{D_{\zeta^2 + C\zeta}(C + 2\zeta)} \quad \text{(E.17)}$$

Expanding to second order by using (E.7) for $Re(\zeta) \gg 1$ we find:

$$\hat{f}_{A^x}(x, \zeta) \sim 3^{\frac{2}{3}}\Gamma\left(\frac{2}{3}\right) e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)}\left\{Ai\left(\zeta^{\frac{1}{3}} y\right) + \frac{Cy}{6} \cdot \zeta^{-\frac{2}{3}} Ai'\left(\zeta^{\frac{1}{3}} y\right) - \left(\frac{1}{2} - \frac{C^2}{4}\right)\left(\frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \zeta^{-\frac{2}{3}} Ai\left(\zeta^{\frac{1}{3}} y\right) + \zeta^{-\frac{2}{3}} Ai'\left(\zeta^{\frac{1}{3}} y\right)\right)\right\} \quad \text{(E.18)}$$

where we have set $y = x - C$. By [Abra70], (10.4.14 and 10.4.16 page 447) the Airy functions may be written in terms of modified Bessel functions of third kind through:

$$Ai\left(\zeta^{\frac{1}{3}}y\right) = \frac{1}{3^{1/2}\pi}y^{\frac{1}{2}}\zeta^{\frac{1}{6}}K_{1/3}\left(\frac{2}{3}y^{\frac{3}{2}}\zeta^{\frac{1}{2}}\right) \text{ and } Ai'\left(\zeta^{\frac{1}{3}}y\right) = -\frac{1}{3^{1/2}\pi}y\zeta^{\frac{1}{3}}K_{2/3}\left(\frac{2}{3}y^{\frac{3}{2}}\zeta^{\frac{1}{2}}\right). \quad \text{(E.19)}$$

We now make use of the following integral found in [Grad94], (7.629 page 872):

$$\int_0^\infty \left\{ t^{-k} e^{-\frac{a}{2t}} W_{k,\mu}\left(\frac{a}{t}\right)\right\} e^{-st}dt = s^{k-\frac{1}{2}} K_{2\mu}(2\sqrt{as}) \quad \text{(E.20)}$$

Substituting for $\lambda = 2\sqrt{a}$, $n = k - \frac{1}{2}$ and $r = 2\mu$ it follows that the inverse of the Laplace transform $\hat{g}(s) = s^n K_r(\lambda\sqrt{s})$ is $g(t) = t^{-\left(n+\frac{1}{2}\right)} e^{-\frac{\lambda^2}{8t}} W_{n+\frac{1}{2},\frac{r}{2}}\left(\frac{\lambda^2}{4t}\right)$ where $W_{\kappa,\mu}(y)$ is the second Whittakers's function ([Grad94] 9.22-9.23 page 1087). By using (E.19) and the result above we may find the inverse of the functions;

$$\hat{g}_1(\zeta, y, k) = \zeta^{-\frac{k}{3}} Ai\left(\zeta^{\frac{1}{3}}y\right) \text{ is } g_1(z, y, k) = \frac{3^{1/2}}{2\pi} y^{-1} z^{\frac{k-2}{3}} e^{-\frac{y^3}{18z}} W_{-\frac{k-2}{2},\frac{1}{6}}\left(\frac{y^3}{9z}\right) \text{ and } \quad \text{(E.21)}$$

$$\hat{g}_2(\zeta, y, k) = \zeta^{-\frac{k}{3}} Ai'\left(\zeta^{\frac{1}{3}}y\right) \text{ is } g_2(z, y, k) = -\frac{3^{1/2}}{2\pi} y^{-\frac{1}{2}} z^{\frac{2k-5}{6}} e^{-\frac{y^3}{18z}} W_{-\frac{2k-5}{6},\frac{1}{3}}\left(\frac{y^3}{9z}\right) \quad \text{(E.22)}$$

By inverting (E.18) term by term and applying (E.21) and (E.22) we obtain for $z \ll 1$:

$$f_{A^x}(x, z) \sim \frac{3^{\frac{7}{6}}\Gamma\left(\frac{2}{3}\right)}{2\pi} e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} e^{-\frac{(x-C)^3}{18z}} \left(\frac{z^{-\frac{2}{3}}}{(x-c)} W_{\frac{2}{3},\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) \div \frac{C}{6}(x-c)^{\frac{1}{2}} z^{-\frac{1}{6}} W_{\frac{1}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) + \right.$$

$$\left. \left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\frac{z^{-\frac{1}{6}}}{(x-c)^{\frac{1}{2}}} W_{\frac{1}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) - \frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \frac{1}{(x-c)} W_{0,\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right)\right)\right) \quad \text{(E.23)}$$

Based on (E.18) we also find the asymptotics for $\hat{F}_{A^x}(x, \zeta)$ as:

$$\hat{F}_{A^x}(x, \zeta) \sim \frac{1}{\zeta} - 3^{\frac{2}{3}}\Gamma\left(\frac{2}{3}\right) e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} \left\{ \zeta^{-1} Ai\left(\zeta^{\frac{1}{3}}y\right) + \frac{Cy}{6}\cdot\zeta^{-\frac{5}{3}} Ai'\left(\zeta^{\frac{1}{3}}y\right) - \left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \zeta^{-\frac{5}{3}} Ai\left(\zeta^{\frac{1}{3}}y\right) + \zeta^{-\frac{5}{3}} Ai'\left(\zeta^{\frac{1}{3}}y\right)\right)\right\} \quad \text{(E.24)}$$

Inverting this transform term by term by applying (E.21) and (E.22) we find:

$$F_{A^x}(x,z) \sim 1 - \frac{3^{\frac{7}{6}}\Gamma\left(\frac{2}{3}\right)}{2\pi} e^{\left(\frac{x^2}{4}-\frac{C^2}{4}\right)} e^{-\frac{(x-C)^3}{18z}} \left(\frac{z^{\frac{1}{3}}}{(x-c)} W_{-\frac{1}{3},\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) + \frac{C}{6}(x-c)^{\frac{1}{2}} z^{\frac{5}{6}} W_{-\frac{5}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) + \right.$$

$$\left. \left(\frac{1}{2}-\frac{C^2}{4}\right)\left(\frac{z^{\frac{5}{6}}}{(x-c)^{\frac{1}{2}}} W_{-\frac{5}{6},\frac{1}{3}}\left(\frac{(x-C)^3}{9z}\right) - \frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \frac{z}{(x-c)} W_{-1,\frac{1}{6}}\left(\frac{(x-C)^3}{9z}\right) \right) \right) \text{ for } z \ll 1 \tag{E.25}$$

The LST for the PDF for the variable A_C is found in (4.116):

$$\hat{f}_{A_C}(\zeta) = \frac{\varphi(C)}{\varphi(C) - C\phi(C)} \left(1 - (\zeta + C) \frac{D_{\zeta^2 + C\zeta - 1}(C + 2\zeta)}{D_{\zeta^2 + C\zeta}(C + 2\zeta)} \right) \tag{E.26}$$

By applying the asymptotics (E.7) and (E.10) we expand the transform to second order for $Re(\zeta) \gg 1$:

$$\hat{f}_{A_C}(\zeta) \sim \frac{\varphi(C)}{\varphi(C) - C\phi(C)} \left(\frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \zeta^{-\frac{2}{3}} - \frac{C}{2\zeta} \right) \tag{E.27}$$

Inverting we find the following asymptotics for small z:

$$f_{A_C}(z) \sim \frac{\varphi(C)}{\varphi(C) - C\phi(C)} \left(\frac{3^{\frac{1}{3}}}{\Gamma\left(\frac{1}{3}\right)} z^{-\frac{1}{3}} - \frac{C}{2} \right) \text{ and} \tag{E.28}$$

The asymptotics for $\hat{F}_{A_C}(\zeta)$ is then found as:

$$\hat{F}_{A_C}(\zeta) \sim \frac{1}{\zeta} - \frac{\varphi(C)}{\varphi(C) - C\phi(C)} \left(\frac{3^{\frac{1}{3}}\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} \zeta^{-\frac{5}{3}} - \frac{C}{2\zeta^2} \right). \tag{E.29}$$

Inverting term by term yields for small z:

$$F_{A_C}(z) \sim 1 - \frac{\varphi(C)}{\varphi(C) - C\phi(C)} \left(\frac{3^{\frac{4}{3}}}{2\Gamma\left(\frac{1}{3}\right)} z^{\frac{2}{3}} - \frac{Cz}{2} \right) \tag{E.30}$$

e.3 Asymptotics for the first passage time for small arguments

We shall also investigate the conditional PDF for the first passage time $f_{T^x}(x, z)$ for small t. Recall from (4.75) that the LST is given as:

$$\hat{f}_{T^x}(x, s) = e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \frac{D_{-s}(x)}{D_{-s}(C)} \tag{E.31}$$

For large values of s and moderate x and C, we may apply the asymptotic formula [Abra70] (19.9.1 page 689) and we find the following asymptotic expansion:

$$\hat{f}_{T^x}(x, s) \sim e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} e^{-(x-C)\sqrt{s})} \left(1 + \left\{\frac{1}{4}(x - C) - \frac{1}{24}(x^3 - C^3)\right\} s^{-\frac{1}{2}}\right) \tag{E.32}$$

Inverting term by term yields for small t:

$$f_{T^x}(x, t) \sim e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \frac{x - C}{2\sqrt{\pi t^3}} e^{-\frac{(x-C)^2}{4t}} \left\{\left(1 + \frac{t}{2}\left(1 - \frac{x^2 + xC + C^2}{6}\right)\right)\right\} \tag{E.33}$$

Based on (E.32) we find the following asymptotics for $\hat{F}_{T^x}(x, s)$ as:

$$\hat{F}_{T^x}(x, s) \sim \frac{1}{s} - e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} e^{-(x-C)\sqrt{s})} \left(\frac{1}{s} + \left\{\frac{1}{4}(x - C) - \frac{1}{24}(x^3 - C^3)\right\} s^{-\frac{3}{2}}\right) \tag{E.34}$$

Inverting term by term yields for small t:

$$F_{T^x}(x, t) \sim 1 - e^{\left(\frac{x^2}{4} - \frac{C^2}{4}\right)} \left(Erfc\left[\frac{x - C}{2\sqrt{t}}\right]\left(1 - \frac{(x-C)^2}{4}\left(1 - \frac{x^2 + xC + C^2}{6}\right)\right) - \frac{(x-C)}{2}\sqrt{\frac{t}{\pi}} e^{-\frac{(x-C)^2}{4t}} \left(1 - \frac{x^2 + xC + C^2}{6}\right)\right) \tag{E.35}$$

The LST for the PDF for the variable T_C is found in (4.78)

$$\hat{f}_{T_C}(s) = \frac{\varphi(C)}{\phi(C)} \frac{D_{-s-1}(C)}{D_{-s}(C)} \tag{E.36}$$

By applying the asymptotic formula [Abra70], (19.9.1 page 689) we find the following asymptotic expansion for large s:

$$\hat{f}_{T_C}(s) \sim \frac{\varphi(C)}{\phi(C)}\left(s^{-\frac{1}{2}} - \frac{C}{2s}\right) \tag{E.37}$$

Inverting term by term gives the following asymptotics for small t:

$$f_{T_C}(t) \sim \frac{\varphi(C)}{\phi(C)}\left(\frac{1}{\sqrt{\pi t}} - \frac{C}{2}\right) \tag{E.38}$$

For $\hat{F}_{T_C}(s)$ we then find for large s:

$$\hat{F}_{T_C}(s) \sim \frac{1}{s} + -\frac{\varphi(C)}{\phi(C)}\left(s^{-\frac{3}{2}} - \frac{C}{2s^2}\right) \tag{E.39}$$

and by inverting term by term we find for small t:

$$F_{T_C}(t) \sim 1 - \frac{\varphi(C)}{\phi(C)}\left(2\sqrt{\frac{t}{\pi}} - \frac{Ct}{2}\right) \tag{E.40}$$

Appendix F

Some technical details in chapter 7 and chapter 8

F.1 Numerical algorithms for calculating the convolution of the waiting time distributions for a given number of M/D/1 queues

It is well known that a direct implementation of the DF of the waiting time in an M/D/1 queue given by (7.27) may cause numerical problems for large x due to the appearance of large alternating nearly cancelling terms in the sum. The same problem will then also yield for the convolutions (7.28) and (7.32). It is possible to avoid this problem by introducing "local coordinates" by writing $q(x, \rho)$ as: ([Robe96], (page 391).

$$q(x,\rho) = (1-\rho)e^{\rho x}\sum_{i=0}^{\lfloor x \rfloor} a_i^{\lfloor x \rfloor}(x-\lfloor x \rfloor)^i \tag{F.1}$$

where the coefficients a_i^n may be calculated recursively by:

$a_0^0 = 1$ and for $n = 1,2,...$

$$a_0^n = \sum_{i=0}^{n-1} a_i^{n-1} \tag{F.2}$$

$$\text{for } i = 1,...,n \quad a_i^n = -\frac{1}{i}\rho e^{-\rho} a_{i-1}^{n-1} \tag{F.3}$$

By applying (7.14) and applying expression (F.1) we find $q^K(x, \rho)$ (defined in chapter 7 by (7.28)) on the form:

$$q^K(x,\rho) = (1-\rho)^K e^{\rho x}\sum_{l=0}^{K-1}\sum_{i=0}^{\lfloor x \rfloor}\frac{b_{l,i}^{\lfloor x \rfloor}}{l!(K-l-1)!}(x-\lfloor x \rfloor)^{K+i-l-1} \quad \text{where} \tag{F.4}$$

$$b_{l,i}^n = e^{-\rho n}\frac{\partial^l}{\partial \rho^l}\left\{e^{\rho n} a_i^n \rho^{K-1}\right\} \tag{F.5}$$

By differentiation and using the recursion (F.2) and (F.3) we find the following recursion for $b_{l,i}^{\lfloor x \rfloor}$:

for $l = 0,..., K-1$

$$b_{l,0}^{0} = (K-1)(K-2)...(K-l)\rho^{K-l-1} \tag{F.6}$$

for $n = 1,2.....$ $$b_{l,0}^{n} = \sum_{i=0}^{n-1}\sum_{s=0}^{l}\binom{l}{s} b_{s,i}^{n-1} \tag{F.7}$$

for $i = 1,2,...,n$ $$b_{0,i}^{n} = -\frac{1}{i}\rho e^{-\rho} b_{0s,i-1}^{n-1} \text{ and } b_{l,i}^{n} = -\frac{1}{i} e^{-\rho}(\rho b_{l,i-1}^{n-1} + l b_{l-1,i-1}^{n-1}) \text{ for } l \geq 1 \tag{F.8}$$

The function $q_1^K(x, \rho)$ has similar recursions as $q^K(x, \rho)$. By applying (7.13) on (7.32) and using expression (F.1) for $q(x, \rho)$ we may write $\rho q_1^K(x, \rho)$ as follows:

$$\rho q_1^K(x, \rho) = (1-\rho)^K e^{\rho x} \sum_{l=0}^{K-1}\sum_{i=0}^{\lfloor x \rfloor} \frac{d_{l,i}^{\lfloor x \rfloor}}{l!(K-l-1)!}(x - \lfloor x \rfloor)^{K+i-l-1} \text{ where} \tag{F.9}$$

$$d_{l,i}^{n} = e^{-\rho n}\frac{\partial^l}{\partial \rho^l}\left\{e^{\rho n} a_i^n \rho^K\right\} \tag{F.10}$$

By comparing (F.10) and (F.5) we see that $d_{l,i}^{\lfloor x \rfloor}$ obeys the same recursion formulae as $a_{l,i}^{\lfloor x \rfloor}$, but with K replaced with $K+1$ in (F.6)-(F.8).

F.2 Numerical algorithms for calculating the auxiliary functions $q^{K,j}(x, \rho)$

By applying expression (F.4) for $q^K(x, \rho)$ we may write $q^{K,j}(x, \rho)$ (defined in chapter 8 by (8.14) and (8.15)) on a similar form: (The expression also yields for $q^K(x, \rho)$ since we have $q^{K,-1}(x, \rho) = q^K(x, \rho)$.)

$$q^{K,j}(x, \rho) = (1-\rho)^K e^{\rho x}\sum_{l=0}^{K-1}\sum_{i=0}^{\lfloor x \rfloor}\frac{b_{l,i}^{j,\lfloor x \rfloor}}{l!(K-l-1)!}(x - \lfloor x \rfloor)^{K+i-l-1}\ ;\ j = -1,0,..., K-2 \text{ and}$$

$$q^{K,K-1}(x, \rho) = (1-\rho)^K e^{\rho x}\sum_{l=0}^{K-1}\sum_{i=0}^{\lfloor x \rfloor}\frac{b_{l,i}^{K-1,\lfloor x \rfloor}}{l!(K-l-1)!}(x - \lfloor x \rfloor)^{K+i-l-1} + (-1)^K\left(\frac{1-\rho}{\rho}\right)^K$$

where $b_{l,i}^{j,n} = e^{-\rho n} \frac{\partial^l}{\partial \rho^l} \left\{ e^{\rho n} a_i^n \rho^{K-j-2} \right\}$ for $l = 0,...,K-1$; $j = -1,0,...,K-1$ (F.11)

By differentiation and using the recursion for a_i^n we find the following recursion for $b_{l,i}^{j,n}$ for $l = 0,...,K-1$, $j = -1,0,...,K-1$:

$$b_{l,0}^{j,0} = (K-j-2)(K-j-3)...(K-j-l-1)\rho^{K-l-j-2} \tag{F.12}$$

for $n = 1,2.....$: $b_{l,0}^{j,n} = \sum_{i=0}^{n-1} \sum_{s=0}^{l} \binom{l}{s} b_{s,i}^{j,n-1}$ for $i = 1,2,...,n$: $b_{0,i}^{j,n} = -\frac{1}{i} \rho e^{-\rho} b_{0,i-1}^{j,n-1}$ (F.13)

and $b_{l,i}^{j,n} = -\frac{1}{i} e^{-\rho} (\rho b_{l,i-1}^{j,n-1} + l b_{l-1,i-1}^{j,n-1})$ for $l \geq 1$ (F.14)

F.3 Convolution of the waiting time of an M/D/1 queue with an exponentially distribution

We shall use the expression $q(x,\rho) = (1-\rho) \sum_{k=0}^{\lfloor x \rfloor} \frac{[\rho(k-x)]^k}{k!} e^{-\rho(k-x)}$ for the normalised waiting time for the M/D/1 queue to find the convolution

$$F(t,\mu,\rho) = \mu \int_{x=0}^{t} e^{-\mu(t-x)} q(x,\rho) dx \tag{F.15}$$

To get this convolution we write $F(t,\mu,\rho) = \mu e^{-\mu t} G(t,\mu,\rho)$ where

$$G(t,\mu,\rho) = \int_{x=0}^{t} e^{\mu x} q(x,\rho) dx = (1-\rho) \int_{x=0}^{t} \sum_{k=0}^{\infty} e^{\mu k} H(x-k) \frac{[\rho(k-x)]^k}{k!} e^{-(\rho+\mu)(k-x)} dx$$

$$= (1-\rho) \sum_{k=0}^{\infty} H(t-k) \int_{x=k}^{t} e^{\mu k} \frac{[\rho(k-x)]^k}{k!} e^{-(\rho+\mu)(k-x)} dx = (1-\rho) \sum_{k=0}^{\lfloor t \rfloor} \int_{x=k}^{t} e^{\mu k} \frac{[\rho(k-x)]^k}{k!} e^{-(\rho+\mu)(k-x)} dx$$

$$= -\frac{1-\rho}{\mu+\rho} \sum_{k=0}^{\lfloor t \rfloor} \left(\frac{\rho}{\mu+\rho} \right)^k e^{\mu k} \int_{\xi=0}^{(\mu+\rho)(k-t)} \frac{\xi^k}{k!} e^{-\xi} d\xi$$

Integrating (and collecting) we obtain:

$$G(t,\mu,\rho) = -\frac{1-\rho}{\mu+\rho} \left(\sum_{k=0}^{\lfloor t \rfloor} \left(\frac{\rho}{\mu+\rho} \right)^k e^{\mu k} - \sum_{k=0}^{\lfloor t \rfloor} \left(\frac{\rho}{\mu+\rho} \right)^k e^{\mu k} \sum_{i=0}^{k} \frac{((\rho+\mu)(k-t))^i}{i!} e^{-(\rho+\mu)(k-t)} \right).$$ The corresponding result for $F(t,\mu,\rho)$ is then:

$F(t,\mu,\rho)=\frac{\mu(1-\rho)}{\mu+\rho}\left(\sum_{k=0}^{\lfloor t\rfloor}\left(\frac{\rho}{\mu+\rho}\right)^k\sum_{i=0}^{k}\frac{\left((\rho+\mu)(k-t)\right)^i}{i!}e^{-\rho(k-t)}-\sum_{k=0}^{\lfloor t\rfloor}\left(\frac{\rho}{\mu+\rho}\right)^k e^{\mu(k-t)}\right)$. Introducing the new summing variable $j=k-i$ in the first (double) sum we have $\sum_{k=0}^{\lfloor t\rfloor}\sum_{i=0}^{k}=\sum_{j=0}^{\lfloor t\rfloor}\sum_{i=0}^{\lfloor t\rfloor-j}$ and the corresponding expression may be written as:

$$(1-\rho)\sum_{k=0}^{\lfloor t\rfloor}\sum_{i=0}^{k}\left(\frac{\rho}{\mu+\rho}\right)^k\frac{\left((\rho+\mu)(k-t)\right)^i}{i!}e^{-\rho(k-t)}=(1-\rho)\sum_{j=0}^{\lfloor t\rfloor}\sum_{i=0}^{\lfloor t\rfloor-j}\left(\frac{\rho}{\mu+\rho}\right)^{i+j}\frac{\left((\rho+\mu)(i-(t-j))\right)^i}{i!}e^{-\rho(i-(t-j))}$$

$=(1-\rho)\sum_{j=0}^{\lfloor t\rfloor}\left(\frac{\rho}{\mu+\rho}\right)^j\sum_{i=0}^{\lfloor t\rfloor-j}\frac{\left(\rho(i-(t-j))\right)^i}{i!}e^{-\rho(i-(t-j))}=\sum_{j=0}^{\lfloor t\rfloor}\left(\frac{\rho}{\mu+\rho}\right)^j q(t-j,\rho)$. Then collecting the results we finally get:

$$F(t,\mu,\rho)=\frac{\mu}{\mu+\rho}\sum_{k=0}^{\lfloor t\rfloor}\left(\frac{\rho}{\rho+\mu}\right)^k\left(q(t-k,\rho)-(1-\rho)e^{\mu(k-t)}\right) \tag{F.16}$$

It follows that the integral

$$\int_{x=0}^{t}e^{-\mu(t-x)}q(x,\rho)dx=\sum_{k=0}^{\lfloor t\rfloor}\frac{\rho^k}{(\rho+\mu)^{k+1}}\left(q(t-k,\rho)-(1-\rho)e^{\mu(k-t)}\right) \tag{F.17}$$

F.4 The convolution of the DF of the waiting time in an M/D/1 queue (with unit service times) and the PDF of an Erlang-i variable with parameter μ.

Differentiation of equation (F.17) i-times with respect to the parameter μ gives:

$$I_k(t,\mu,\rho)=\int_{x=0}^{t}(t-x)^i e^{-\mu(t-x)}q(x,\rho)dx=(-1)^i\frac{\partial^i}{\partial\mu^i}\left\{\sum_{k=0}^{\lfloor t\rfloor}\frac{\rho^k}{(\rho+\mu)^{k+1}}\left(q(t-k,\rho)-(1-\rho)e^{\mu(k-t)}\right)\right\}$$

performing the differentiation gives:

$$I_k(t,\mu,\rho)=\int_{x=0}^{t}(t-x)^i e^{-\mu(t-x)}q(x,\rho)dx= \tag{F.18}$$

$$\sum_{k=0}^{\lfloor t\rfloor}\left\{\frac{(k+i)!}{k!}\frac{\rho^k}{(\rho+\mu)^{k+i+1}}q(t-k,\rho)-(1-\rho)\rho^k\sum_{l=0}^{i}\binom{i}{l}\frac{(k+i-l)!}{k!}\frac{(t-k)^l}{(\rho+\mu)^{k+i-l+1}}e^{\mu(k-t)}\right\}.$$

If we let

$$F_j(t,\mu,\rho) = \mu \int_{x=0}^{t} \frac{\left(\mu(t-x)\right)^{j-1} e^{-\mu(t-x)}}{(j-1)!} q(x,\rho)dx \tag{F.19}$$

be the convolution of the DF of the waiting time in an M/D/1 queue (with unit service times) and the PDF of an Erlang i variable with parameter μ, then we obtain from the relation (F.18):

$$F_j(t,\mu,\rho) = \tag{F.20}$$

$$\sum_{k=0}^{\lfloor t \rfloor} \left\{ \binom{k+j-1}{k} \frac{\mu^j \rho^k}{(\rho+\mu)^{k+j}} q(t-k,\rho) - (1-\rho) \sum_{l=0}^{j-1} \frac{1}{l!} \binom{k+j-l-1}{k} \frac{\mu^{j-l} \rho^k}{(\rho+\mu)^{k+j-l}} (\mu(t-k))^l e^{\mu(k-t)} \right\}$$

F.5 The convolution of the DF of the K-folded waiting time for an M/D/1 queue (all with service times scaled to unity) with the PDF of an Erlang-i distributed variable with parameter μ.

If we further let

$$F_{K,j}(t,\mu,\rho) = \mu \int_{x=0}^{t} \frac{\left(\mu(t-x)\right)^{j-1} e^{-\mu(t-x)}}{(j-1)!} q^K(x,\rho)dx \tag{F.21}$$

be the convolution of the DF of the K-folded waiting time for an M/D/1 queue (all with service times scaled to unity) with the PDF of an Erlang-i distributed variable with parameter μ then we have by applying relation (7.14):

$$F_{K,j}(t,\mu,\rho) = \frac{(1-\rho)^K}{(K-1)!} \frac{\partial^{K-1}}{\partial \rho^{K-1}} \left\{ \frac{\rho^{K-1}}{1-\rho} F_j(t,\mu,\rho) \right\}.$$

By performing the differentiation by using expressions (F.24) and (F.25) below we obtain the following expression for the convolution (F.21):

$$F_{K,j}(t,\mu,\rho) = \sum_{k=0}^{\lfloor t \rfloor} \frac{\mu^j \rho^k}{(\rho+\mu)^{k+j}} \left\{ \binom{k+j-1}{k} \sum_{s=0}^{K-1} (1-\rho)^{K-s-1} f_{K-s-1,K+k-s-2,k+j}\left(\frac{\rho}{\rho+\mu}\right) q^{s+1}(t-k,\rho) - \right.$$

$$\left. (1-\rho)^K \sum_{l=0}^{j-1} \frac{1}{l!} \binom{k+j-l-1}{k} f_{K-1,K+k-1,k+j-l}\left(\frac{\rho}{\rho+\mu}\right) \left((\rho+\mu)(t-k)\right)^l e^{\mu(k-t)} \right\} \tag{F.22}$$

and where $f_{l,m,n}(x)$ is given below by (F.26).

F.6 Some frequently used expressions

We shall also frequently apply expressions of the type:

$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}f(\rho)\frac{q(x,\rho)}{1-\rho}\right\}$ where $f(\rho)$ is some function of the parameter ρ. Differentiation yields:

$$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}f(\rho)\frac{q(x,\rho)}{1-\rho}\right\}=\sum_{l=0}^{K-1}\frac{(1-\rho)^{K-l-1}}{(K-l-1)!}\frac{\partial^{K-l-1}}{\partial\rho^{K-l-1}}\left\{\rho^{K-1}f(\rho)\right\}q_1^{l+1}(x,\rho)$$ where

$q_1^{l+1}(x,\rho)$ is given by (72). Inserting for $q_1^{l+1}(x,\rho)$ by (74), we find:

$$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}f(\rho)\frac{q(x,\rho)}{1-\rho}\right\}=\sum_{s=0}^{K-1}\frac{\rho(1-\rho)^{K-s-1}}{(K-s-1)!}q^{s+1}(x,\rho)\left[\sum_{l=0}^{K-s-1}\binom{K-s-1}{l}(s+1)...(s+l)\rho^{-(l+s+1)}\frac{\partial^{K-s-1-l}}{\partial\rho^{K-s-1-l}}\left\{\rho^{K-1}f(\rho)\right\}\right]$$

$$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}f(\rho)\frac{q(x,\rho)}{1-\rho}\right\}=\sum_{s=0}^{K-1}\frac{\rho(1-\rho)^{K-s-1}}{(K-s-1)!}q^{s+1}(x,\rho)\left[\sum_{l=0}^{K-s-1}\binom{K-s-1}{l}(s+1)...(s+l)\rho^{-(l+s+1)}\frac{\partial^{K-s-1-l}}{\partial\rho^{K-s-1-l}}\left\{\rho^{K-1}f(\rho)\right\}\right]$$

The sum in the brackets may be found to be:

$$[\ ]=\sum_{l=0}^{K-s-1}\binom{K-s-1}{l}\frac{\partial^l}{\partial\rho^l}\rho^{-(s+1)}\frac{\partial^{K-s-1-l}}{\partial\rho^{K-s-1-l}}\left\{\rho^{K-1}f(\rho)\right\}=\frac{\partial^{K-s-1}}{\partial\rho^{K-s-1}}\left\{\rho^{K-s-2}f(\rho)\right\}$$

Collecting these results we find:

$$\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}f(\rho)\frac{q(x,\rho)}{1-\rho}\right\}=\sum_{s=0}^{K-1}\frac{1}{(K-s-1)!}q^{s+1}(x,\rho)(1-\rho)^{K-s-1}\rho\frac{\partial^{K-s-1}}{\partial\rho^{K-s-1}}\left\{\rho^{K-s-2}f(\rho)\right\}\quad\text{(F.23)}$$

Applying (F.23) on the expression

$$G_{K,j,k}(x,\rho,\mu)=\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{\rho^{k+K-1}}{(\rho+\mu)^{k+j}}\frac{q(x,\rho)}{1-\rho}\right\}\quad\text{(F.24)}$$

we obtain

$$G_{K,j,k}(x,\rho,\mu)=\frac{\rho^k}{(\rho+\mu)^{k+j}}\sum_{s=0}^{K-1}q^{s+1}(x,\rho)(1-\rho)^{K-s-1}f_{K-s-1,K+k-s-2.k+j}\left(\frac{\rho}{\rho+\mu}\right)\quad\text{(F.25)}$$

where

$$f_{l,m,n}(x)=\sum_{s=0}^{l}(-1)^s\binom{m}{l-s}\binom{n+s-1}{s}x^s\quad\text{(F.26)}$$

is defined so that we have $\frac{1}{l!}\frac{\partial^l}{\partial\rho^l}\left(\frac{\rho^m}{(\rho+\mu)^n}\right)=\frac{\rho^{m-l}}{(\rho+\mu)^n}f_{l,m,n}\left(\frac{\rho}{\rho+\mu}\right)$.

F.7 Relation between the convolution and some auxiliary function

In various expressions we come up with the slightly modified versions of the convolution of a given number of waiting times of identical M/D/1 queues. We have that the convolution of the waiting time of K identical M/D/1 queues may be written in terms of the partial derivative of the load parameter ρ as follows:

$$q^K(x,\rho) = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-1}\frac{q(x,\rho)}{1-\rho}\right\} \tag{F.27}$$

The auxiliary expressions occur when the power ρ^{K-1} is omitted in the brackets:

$$q_1^K(x,\rho) = \frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\frac{q(x,\rho)}{1-\rho}\right\} \tag{F.28}$$

We have the following relations between $q_1^K(x,\rho)$ and $q^K(x,\rho)$:

$$q^K(x,\rho) = \sum_{l=0}^{K-1}\binom{K-1}{l}\rho^l(1-\rho)^{K-l-1}q_1^{l+1}(x,\rho) \tag{F.29}$$

and the inverse

$$q_1^K(x,\rho) = \frac{1}{\rho^{K-1}}\sum_{l=0}^{K-1}\binom{K-1}{l}(-1)^{K-l-1}(1-\rho)^{K-l-1}q^{l+1}(x,\rho) \tag{F.30}$$

The first part (F.29) is found by direct differentiation. The inverse (F.30) is easiest proven by induction (on K). From (F.29) we have $q^K(x,\rho) = \sum_{l=0}^{K-2}\binom{K-1}{l}\rho^l(1-\rho)^{K-l-1}q_1^{l+1}(x,\rho) + \rho^{K-1}q_1^K(x,\rho)$ or

$q_1^K(x,\rho) = \frac{1}{\rho^{K-1}}\left(q^K(x,\rho) - \sum_{l=0}^{K-2}\binom{K-1}{l}\rho^l(1-\rho)^{K-l-1}q_1^{l+1}(x,\rho)\right)$. Inserting for $q_1^{l+1}(x,\rho)$ for $l = 0,...,K-2$ (by the induction assumption) we find:

$$q_1^K(x,\rho) = \frac{1}{\rho^{K-1}}\left(q^K(x,\rho) - \sum_{l=0}^{K-2}\binom{K-1}{l}(1-\rho)^{K-l-1}\sum_{s=0}^{l}\binom{l}{s}(-1)^{l-s}(1-\rho)^{l-s}q^{s+1}(x,\rho)\right)$$

$= \frac{1}{\rho^{K-1}}\left(q^K(x,\rho) - \sum_{s=0}^{K-2}(1-\rho)^{K-s-1}q^{s+1}(x,\rho)\sum_{l=s}^{K-2}(-1)^{l-s}\binom{l}{s}\binom{K-1}{l}\right)$. Now we have

$\sum_{l=s}^{K-2}(-1)^{l-s}\binom{l}{s}\binom{K-1}{l}=\binom{K-1}{s}\sum_{l=s}^{K-2}(-1)^{l-s}\binom{K-s-1}{l-s}=\binom{K-1}{s}\sum_{l=0}^{K-s-2}(-1)^{l}\binom{K-s-1}{l}$. Further it is known that the sum $\sum_{l=0}^{K-s-1}(-1)^{l}\binom{K-s-1}{l}=0$ giving $\sum_{l=0}^{K-s-2}(-1)^{l}\binom{K-s-1}{l}=(-1)^{K-s}$. Inserting we obtain $q_1^K(x,\rho)=\frac{1}{\rho^{K-1}}\left(q^K(x,\rho)-\sum_{s=0}^{K-2}(-1)^{K-s}\binom{K-1}{s}(1-\rho)^{K-s-1}q^{s+1}(x,\rho)\right)$

$=\sum_{s=0}^{K-1}(-1)^{K-s-1}\binom{K-1}{s}(1-\rho)^{K-s-1}q^{s+1}(x,\rho)$ which is in accordance with the induction assumption.

It is also possible to express the auxiliary functions $q^{K,i}(x,\rho)=\frac{(1-\rho)^K}{(K-1)!}\frac{\partial^{K-1}}{\partial\rho^{K-1}}\left\{\rho^{K-i-2}\frac{q(x,\rho)}{1-\rho}\right\}$ in terms of convolutions $q^{s+1}(x,\rho)$ $s=0,1,...,K-1$. Differentiating gives

$$q^{K,i}(x,\rho)=\frac{(1-\rho)^K}{(K-1)!}\sum_{s=0}^{K-1}\binom{K-1}{s}\frac{\partial^{K-s-1}}{\partial\rho^{K-s-1}}\left\{\rho^{K-i-2}\right\}\frac{\partial^s}{\partial\rho^s}\left\{\frac{q(x,\rho)}{1-\rho}\right\}$$

$=\sum_{s=0}^{K-1}\frac{(1-\rho)^{K-s-1}}{(K-s-1)!}\frac{\partial^{K-s-1}}{\partial\rho^{K-s-1}}\left\{\rho^{K-i-2}\right\}q_1^{s+1}(x,\rho)$. Then inserting for $q_1^{s+1}(x,\rho)$ from (F.30) we find:

$$q^{K,i}(x,\rho)=\sum_{r=0}^{K-1}(1-\rho)^{K-r-1}q^{r+1}(x,\rho)\left[\sum_{s=r}^{K-1}(-1)^{s-r}\binom{s}{r}\rho^{-s}\frac{1}{(K-s-1)!}\frac{\partial^{K-s-1}}{\partial\rho^{K-s-1}}\left\{\rho^{K-i-2}\right\}\right]$$

Evaluating the bracket in the last expression we obtain:

$[\ ]=\frac{\rho}{(K-r-1)!}\frac{\partial^{K-r-1}}{\partial\rho^{K-r-1}}\left\{\rho^{K-r-i-3}\right\}$ and collecting terms we find:

$$q^{K,i}(x,\rho)=\rho^{-i-1}\sum_{r=0}^{K-1}(-1)^{K-r-1}\binom{i+1}{K-r-1}(1-\rho)^{K-r-1}q^{r+1}(x,\rho)$$

F.8 The convolution of the DF of the K-folded waiting time for an M/D/1 queue with the PDF of an r-folded rectangular variable over $(0,b^L)$

The form of an r-folded rectangular variable $\bar{b}^L(t)^{*(r)}$ over $(0,b^L)$ is given by (8.12) and we find that the convolution $W^T(t,\rho^H)(*)\bar{b}^L(t)^{*(r)}$ may be written as:

$$W^T(t,\rho^H)(*)\hat{b}^L(t)^{*(r)} = r\left(\frac{b^H}{b^L}\right)^r \sum_{m=0}^{r} \frac{(-1)^m}{m!(r-m)!} H(t-mb^L) I^{K,r-1}\left(\frac{t-mb^L}{b^H},\rho^H\right)$$

where $I^{K,i}(t,\rho)$ are the following integrals:

$I^{K,i}(t,\rho) = \int_{x=0}^{t} (t-x)^i q^K(x,\rho)dx$ for $i = 0,1,\ldots,K-1$. These integrals may be evaluated in terms of some auxiliary functions given below by equations (F.33), (F.34) and (F.35).

By applying the relation (F.18) we shall find expressions for the integrals

$$I^{K,i}(t,\rho) = \int_{x=0}^{t} (t-x)^i q^K(x,\rho)dx \quad \text{for } i = 0,1,\ldots,K-1. \tag{F.31}$$

Taking the limit $\mu \to 0$ in (F.18) we obtain:

$$\int_{x=0}^{t} (t-x)^i q(x,\rho)dx = \tag{F.32}$$

$$\sum_{k=0}^{\lfloor t \rfloor} \left\{ \frac{(k+i)!}{k!} \rho^{-i-1} q(t-k,\rho) - (1-\rho) \sum_{l=0}^{i} (-1)^l \binom{i}{l} \frac{(k+i-l)!}{k!} \rho^{l-i-1} (t-k)^l \right\}$$

Multiplying the last relation by $\frac{\rho^{K-1}}{1-\rho}$ and taking partial derivatives with respect to the parameter ρ, $K-1$ times, and then multiplying with $\frac{(1-\rho)^K}{(K-1)!}$ gives

$$I^{K,i}(t,\rho) = \sum_{k=0}^{\lfloor t \rfloor} \frac{(k+i)!}{k!} q^{K,i}(t-k,\rho) \quad \text{for } i = 0,1,\ldots,K-2 \tag{F.33}$$

where we define the auxiliary functions

$$q^{K,i}(x,\rho) = \frac{(1-\rho)^K}{(K-1)!} \frac{\partial^{K-1}}{\partial \rho^{K-1}} \left\{ \rho^{K-i-2} \frac{q(x,\rho)}{1-\rho} \right\}. \tag{F.34}$$

For $i = K-1$ we get an extra term from the second part when $l = 0$ in (F.33) and we find that (F.34) also yields for $i = K-1$ if we define:

$$q^{K,K-1}(x,\rho) = \frac{(1-\rho)^K}{(K-1)!} \frac{\partial^{K-1}}{\partial \rho^{K-1}} \left\{ \rho^{-1} \frac{q(x,\rho)}{1-\rho} \right\} + (-1)^K \left(\frac{1-\rho}{\rho}\right)^K \tag{F.35}$$

F.9 Evaluation of the function $F_{K_1,K_2,k,j}(x,y)$

We shall derive an expression for the function

$$F_{K_1,K_2,k,j}(x,y) = (-1)^k \binom{k+j}{j} \frac{(y-x)}{K_1!K_2!} \frac{d^{K_1}}{dx^{K_1}} \frac{d^{K_2}}{dy^{K_2}} \left\{ \frac{x^{k+K_1} y^{j+K_2}}{(y-x)^{k+j+1}} \right\} \tag{F.36}$$

Introducing new variables by

$\xi = y - x$ and $\eta = x$ we find:

$$F_{K_1,K_2,k,j}(x,y) = G_{K_1,K_2,k,j}(\xi,\eta) = (-1)^k \binom{k+j}{j} \frac{\xi}{K_1!K_2!} \left(\frac{d}{d\eta} - \frac{d}{d\xi} \right)^{K_1} \frac{d^{K_2}}{d\xi^{K_2}} \left\{ \eta^{k+K_1} (\xi+\eta)^{j+K_2} \xi^{-(k+j+1)} \right\}$$

Expanding gives:

$$G_{K_1,K_2,k,j}(\xi,\eta) = (-1)^k \binom{k+j}{j} \frac{\xi}{K_1!K_2!} \sum_{s=0}^{K_1} \sum_{r=0}^{K_2+j} (-1)^{K_1-s} \binom{K_1}{s} \binom{K_2+j}{r} \frac{d^s}{d\eta^s} \left\{ \eta^{k+r+K_1} \right\} \frac{d^{K_1+K_2-s}}{d\xi^{K_1+K_2-s}} \left\{ \xi^{K_2-k-r-1} \right\}$$

We find:

$$\frac{d^s}{d\eta^s} \left\{ \eta^{k+r+K_1} \right\} \frac{d^{K_1+K_2-s}}{d\xi^{K_1+K_2-s}} \left\{ \xi^{K_2-k-r-1} \right\} = \frac{(-1)^{K_1+K_2-s}}{\xi} (K_1+K_2)! \binom{K_1+k+r}{K_1+K_2} \left(\frac{\eta}{\xi} \right)^{K_1+k+r-s}$$. Inserting gives:

$$G_{K_1,K_2,k,j}(\xi,\eta) = (-1)^{k+K_2} \binom{K_1+K_2}{K_2} \binom{k+j}{j} \sum_{r=\max[K_2-k,0]}^{K_2+j} \binom{K_2+j}{r} \binom{K_1+k+r}{K_1+K_2} \left(\frac{\eta}{\xi} \right)^{k+r} \left[\sum_{s=0}^{K_1} \binom{K_1}{s} \left(\frac{\eta}{\xi} \right)^{K_1-s} \right]$$

The last sum in the brackets is: $[\ ] = \sum_{s=0}^{K_1} \binom{K_1}{s} \left(\frac{\eta}{\xi} \right)^{K_1-s} = \left(\frac{\xi+\eta}{\xi} \right)^{K_1}$ which gives:

$$G_{K_1,K_2,k,j}(\xi,\eta) = (-1)^{k+K_2} \binom{K_1+K_2}{K_2} \binom{k+j}{j} \left(\frac{\xi+\eta}{\xi} \right)^{K_1} \sum_{r=\max[K_2-k,0]}^{K_2+j} \binom{K_2+j}{r} \binom{K_1+k+r}{K_1+K_2} \left(\frac{\eta}{\xi} \right)^{k+r}$$

giving

$$F_{K_1,K_2,k,j}(x,y) = (-1)^{k+K_2} \binom{K_1+K_2}{K_2} \binom{k+j}{j} \left(\frac{y}{y-x} \right)^{K_1} \sum_{r=\max[K_2-k,0]}^{K_2+j} \binom{K_2+j}{r} \binom{K_1+k+r}{K_1+K_2} \left(\frac{x}{y-x} \right)^{k+r}$$

By changing the summation and expanding we find

$$F_{K_1,K_2,k,j}(x,y) = (-1)^{k+K_2} \binom{K_1+K_2}{K_2} \binom{k+j}{j} \frac{y^{K_1} x^{K_2+k+j}}{(y-x)^{K_1+K_2+k+j}} \sum_{l=0}^{\min[K_2+j,k+j]} (-1)^{-l} \left(\frac{y}{x} \right)^l \left[\sum_{r=l}^{\min[K_2+j,k+j]} (-1)^r \binom{r}{l} \binom{K_2+j}{r} \binom{K_1+K_2+k+j-r}{K_1+K_2} \right]$$

It is possible to rewrite the last bracket as follows:

$$[\ ]=\binom{K_2+j}{l}\left[\sum_{r=l}^{\min[K_2+j,k+j]}(-1)^r\binom{K_2+j-l}{K_2+j-r}\binom{K_1+K_2+k+j-r}{K_1+K_2}\right]=(-1)^{k+j}\binom{K_2+j}{l}\left[\sum_{r=\max[0,k-K_2]}^{k+j-l}(-1)^r\binom{K_2+j-l}{k+j-l-r}\binom{K_1+K_2+r}{K_1+K_2}\right]$$

To find the last sum we apply the following result on the sum of binomial coefficients [Grad94] (0.156 page 5):

$\sum_{k=0}^{p}\binom{n}{k}\binom{m}{p-k}=\binom{m+n}{p}$ or letting $n\to -n$ gives $\sum_{k=0}^{p}(-1)^k\binom{m}{p-k}\binom{n-1-k}{n-1}=\binom{m-n}{p}$. Now by taking $p=k+j-l$, $m=K_2+j-l$ and $n=K_1+K_2+1$ we obtain:

$$[\ ]=(-1)^{k+j}\binom{K_2+j}{l}\binom{j-K_1-l-1}{k+j-l}=(-1)^l\binom{K_2+j}{l}\binom{K_1+k}{k+j-l}$$

Inserting we then finally obtain:

$$F_{K_1,K_2,k,j}(x,y)=(-1)^{k+K_2}\binom{K_1+K_2}{K_2}\binom{k+j}{j}\frac{y^{K_1}x^{K_2}}{(y-x)^{K_1+K_2+k+j}}\sum_{l=\max[0,\,j-K_1]}^{\min[K_2+j,k+j]}\binom{K_2+j}{l}\binom{K_1+k}{k+j-l}y^l x^{k+j-l}\,. \qquad \text{(F.37)}$$

Appendix G

Some technical details in chapter 9

G.1 Solution of the linear equations (9.17)

We shall find the solution of the linear equations

$$\sum_{k=0}^{K-1} \tilde{q}^k (\lambda_j)^k = h_j \text{ for } j = 1, \ldots, K \text{ where} \tag{G.1}$$

$\lambda_j = \dfrac{B(r_j)}{r_j}$ and $h_j = -\dfrac{r_j^m}{1-\frac{1}{r_j}} A\left(x\dfrac{B(r_j)}{r_j}\right)\dfrac{r_j}{B(r_j)}$. If we let $A = (a_{kj})$ be the $K \times K$-matrix with elements: $a_{kj} = (\lambda_j)^k$ and let $A^{-1} = (a_{ik}^{-1})$ be the corresponding inverse, then $\sum_{k=0}^{K-1} a_{ik}^{-1}(\lambda_j)^k = \delta_{ij}$.

We define the generating functions: $G_i(x) = \sum_{k=0}^{K-1} a_{ik}^{-1} x^k$, then $G_i(x)$ will have roots λ_j for $j = 1, \ldots, K$; $j \neq i$. Thus, $G_i(x) = C_i \prod_{l=1, l\neq i}^{K} (x - \lambda_l)$ and since $G_i(\lambda_i) = C_i \prod_{l=1, l\neq i}^{K} (\lambda_i - \lambda_j) = 1$, we obtain $C_i = \dfrac{1}{\prod_{l=1, l\neq i}^{K} (\lambda_i - \lambda_l)}$ giving

$$G_i(x) = \prod_{l=1, l\neq i}^{K} \frac{x - \lambda_l}{\lambda_i - \lambda_l} \tag{G.2}$$

The solution of (G.1) is

$$\tilde{q}^k = \sum_{j=1}^{K} h_j a_{jk}^{-1} \text{ and further} \tag{G.3}$$

$$\sum_{k=0}^{K-1} \tilde{q}^k x^k = \sum_{k=0}^{K-1} \left[\sum_{j=1}^{K} h_j a_{jk}^{-1} \right] x^k = \sum_{j=1}^{K} h_j \prod_{l=1,\, l \neq j}^{K} \frac{x - \lambda_l}{\lambda_j - \lambda_l} \tag{G.4}$$

If we take $x = \frac{B(z)}{z}$ and insert for h_j and λ_j, we obtain:

$$\sum_{k=0}^{K-1} \tilde{q}^k \left(\frac{B(z)}{z}\right)^k = -\sum_{j=1}^{K} \frac{r_j^{\,m}}{1 - \frac{1}{r_j}} A\left(x \frac{B(r_j)}{r_j}\right) \frac{r_j}{B(r_j)} \prod_{l=1,\, l \neq j}^{K} \frac{\frac{B(z)}{z} - \frac{B(r_l)}{r_l}}{\frac{B(r_j)}{r_j} - \frac{B(r_l)}{r_l}} \tag{G.5}$$

G.2 Evaluation of the product (9.19)

We shall sketch the deduction of the product

$$P = \prod_{l=1}^{K} \left(\frac{z}{B(z)} - \frac{r_l}{B(r_l)} \right) \tag{G.6}$$

where $r_j = r_j(x, s)$; $j = 1, \ldots, K$ are the roots of the equation:

$$h(z) = 1 - szA\left(x \frac{B(z)}{z}\right) = 0 \tag{7.7}$$

inside the unit disc $|z| < 1$, and $A(z) = \sum_{i=1}^{K+1} a(i) z^i$ is a polynomial of degree $K + 1$, (where we assume that $a(K + 1) \neq 0$).

By introducing the new variable $\zeta = \zeta(z) = \frac{z}{B(z)}$, we have

$$P = \prod_{l=1}^{K} (\zeta - \zeta_l) \tag{G.8}$$

where ζ_l is the corresponding root of $g(\zeta) = 1 - s\zeta A\left(\frac{x}{\zeta}\right) B(z(\zeta)) = 0$ where $z = z(\zeta)$ is the inverse of $\zeta = \zeta(z) = \frac{z}{B(z)}$.

We denote $S = \sum_{k=1}^{K} \log(\zeta - \zeta_k)$ (where we use the principal value of the logarithm). The contour-integral

$$\frac{1}{2\pi i}\int_{\Gamma} \log(\varsigma - \zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma = \sum_{k=1}^{K} \log(\zeta - \zeta_k) - K\log\zeta \tag{G.9}$$

where Γ is a contour containing all the roots $\zeta_1, \zeta_2, ..., \zeta_K$ of $g(\varsigma)$ and also contains $\varsigma = 0$ (which is a pole of multiplicity K for $g(\varsigma)$). Hence

$$S = \frac{1}{2\pi i}\int_{\Gamma} \log(\varsigma - \zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma + K\log\zeta\,. \tag{G.10}$$

Depending on the location of ζ we choose different contours. We let C be the circle $|\varsigma| = r$ containing all the roots $\zeta_1, \zeta_2, ..., \zeta_K$.

If $|\zeta| > r$, we choose the circle C as the contour Γ (see figure G.1)

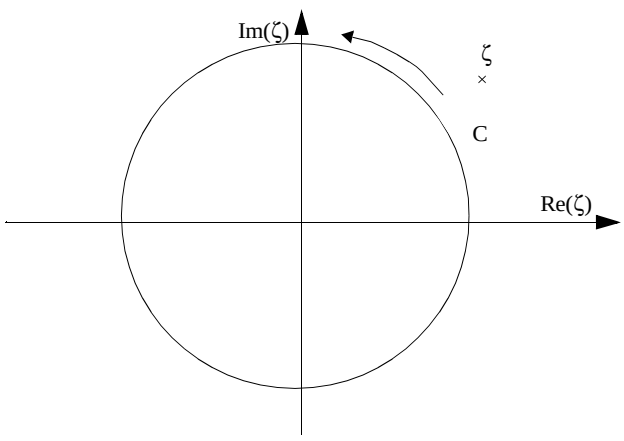

Figure G.1: The contour Γ when $|\zeta| > r$.

Integrating by parts we find

$$\frac{1}{2\pi i}\int_{C} \log(\varsigma - \zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma = \frac{1}{2\pi i}[\log(\varsigma - \zeta)\log g(\varsigma)]_C - \frac{1}{2\pi i}\int_{C} \frac{\log g(\varsigma)}{\varsigma - \zeta}d\varsigma \tag{G.11}$$

When ς is moving around C the argument of $\log g(\varsigma)$ returns to its initial value. Therefore

$[\log(\varsigma-\zeta)\log g(\varsigma)]_C = 0$. Collecting the results above gives:

$$S = \log\zeta^K + \frac{1}{2\pi i}\int_C \frac{\log g(\varsigma)}{\zeta-\varsigma}d\varsigma \tag{G.12}$$

If $|\zeta| < r$, we choose the contour $\Gamma = C \cup L_1 \cup L_2 \cup C_\varepsilon$ (see figure G.2)

On C by integrating by parts we obtain (G.11). However when ς moves along C from ζ_0, the argument of $\log(\varsigma - x)$ is increased by 2π, while the argument of $\log g(\varsigma)$ returns to its starting value so

$$\frac{1}{2\pi i}[\log(\varsigma-\zeta)\log g(\varsigma)]_C = \log g(\zeta_0)$$

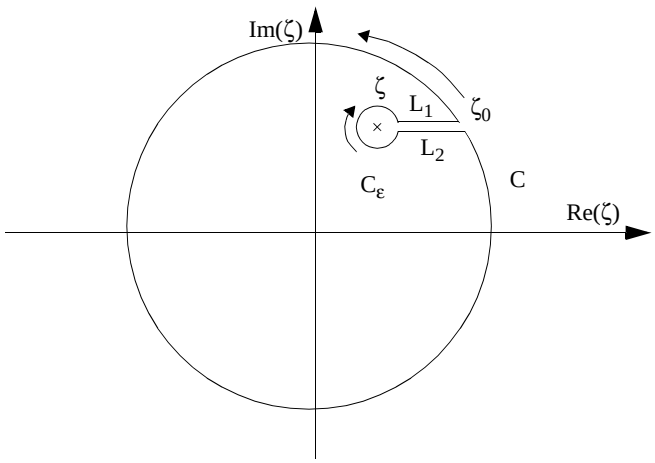

Figure G.2: The contour Γ when $|\zeta| < r$.

On $L_1 \cup L_2$ we have

$$\frac{1}{2\pi i}\int_{L_1\cup L_2}(\ )d\varsigma = \frac{1}{2\pi i}\int_{\zeta+\varepsilon}^{\zeta_0}\log(\varsigma-\zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma + \frac{1}{2\pi i}\int_{\zeta_0}^{\zeta+\varepsilon}\log(\varsigma-\zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma.$$

On L_2 the argument of $\log(\varsigma-\zeta)$ has increased by 2π so that

$$\frac{1}{2\pi i}\int_{L_1\cup L_2}(\ )d\varsigma = -\frac{2\pi i}{2\pi i}\int_{\zeta+\varepsilon}^{\zeta_0}\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma = -\log g(\zeta_0) + \log g(\zeta+\varepsilon) \tag{G.13}$$

On C_ε we have $\varsigma = \zeta + \varepsilon e^{i\theta}$ where θ moves from 2π to 0. (We also assume that $\zeta \neq \zeta_i$, $i = 1, \ldots, K$.) When $\varepsilon \to 0$ we have

$$\frac{1}{2\pi i}\int_{C_\varepsilon} \log(\varsigma - \zeta)\frac{g'(\varsigma)}{g(\varsigma)}d\varsigma \sim \frac{g'(\zeta)}{g(\zeta)}\frac{1}{2\pi}\int_{2\pi}^{0} \log(\varepsilon e^{i\theta})(\varepsilon e^{i\theta})d\theta \to 0 \tag{G.14}$$

Collecting the results above when $\varepsilon \to 0$ we get:

$$S = \log(\zeta^K) + \log g(\zeta) + \frac{1}{2\pi i}\int_C \frac{\log g(\varsigma)}{\zeta - \varsigma}d\varsigma \tag{G.15}$$

By (G.12) and (G.15) we find (by inserting for $z = \frac{\zeta}{B(\zeta)}$ and changing the integration variable $\varsigma = \frac{\eta}{B(\eta)}$):

$$P = \prod_{l=1}^{K}\left(\frac{z}{B(z)} - \frac{r_l}{B(r_l)}\right) = \begin{cases} \left(\frac{z}{B(z)}\right)^K \exp[I(z, x, s)] & \textit{for} \quad z \textit{ outside } C_r \\ \left(\frac{z}{B(z)}\right)^K \left[1 - szA\left(x\frac{B(z)}{z}\right)\right]\exp[I(z, x, s)] & \textit{for} \quad z \textit{ inside } C_r \end{cases} \tag{G.16}$$

where $I(z, x, s)$ is the contour integral:

$$I(z, x, s) = \frac{1}{2\pi i}\int_{C_r} \frac{\log\left[1 - s\eta A\left(x\frac{B(\eta)}{\eta}\right)\right]}{\frac{z}{B(z)} - \frac{\eta}{B(\eta)}}\left(\frac{1}{B(\eta)} - \frac{\eta B'(\eta)}{B(\eta)^2}\right)d\eta \tag{G.17}$$

and where we may choose C_r as the disc $|\eta| \leq r$ where $1 < r < r_{K+1}$, and r_{K+1} is the root of $h(z) = 1 - szA\left(x\frac{B(z)}{z}\right)$ outside the unit circle with the smallest modulo.

Comment: The mapping of the circle C by changing the integration variable $\varsigma = \frac{\eta}{B(\eta)}$ will be a closed contour Γ which contains all the roots r_j; $j = 1, \ldots, K$. This contour may be transformed to a circle as described above without changing the value of the integral.

G.3 The joint distribution of "extra" delay for an FS packet and the number of arrivals from the BS in a slot when the ordering of packets is chosen at random

When the ordering between the **FS** packet and possible arrivals of the **BS** packets are chosen at random we have that the **FS** packet is placed among the **BS** packets with equal prob-

ability. So we have: $P(U_n = i | B_n^1 = j) = \frac{1}{j+1}$ for $i = 1, \ldots, j+1$. Un-conditioning gives the joint distribution:

$$u(i,j) = P(U_n = i, B_n^1 = j) = \frac{1}{j+1} b(j) \text{ for } i = 1, \ldots, j+1 \text{ and } j = 0, 1, \ldots \quad \text{(G.18)}$$

The corresponding joint z-transform is found:

$$U(z_1, z_2) = \boldsymbol{E}\left[z_1^{U_n} z_2^{B_n^1}\right] = \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} z_1^i z_2^j \frac{b(j)}{j+1} = \frac{z_1}{z_2(1-z_1)}[BI(z_2) - BI(z_1 z_2)] \quad \text{(G.19)}$$

where $BI(z)$ is the integral of the z-transform of the batch process of the **BS**, i.e.

$$BI(z) = \int_1^z B(x)dx \quad \text{(G.20)}$$

The marginal z-transform is found from (G.19) (by taking $z_2 = 1$):

$$U(z) = \boldsymbol{E}[z^{U_n}] = U(z, 1) = \frac{z}{z-1} BI(z). \quad \text{(G.21)}$$