

Frode Drevland, Kjell Austeng og Olav Torp

Usikkerhetsanalyse  
Modellering, estimering  
og beregning

Concept rapport Nr 11



concept

Frode Drevland, Kjell Austeng og  
Olav Torp

# Usikkerhetsanalyse - Modellering, estimering og beregning

Teoretisk grunnlag

Concept rapport Nr 11



© Concept-programmet 2005

Concept rapport nr. 11

## **Usikkerhetsanalyse - Modellering, estimering og beregning**

Teoretisk grunnlag

*Frode Drevland , Kjell Austeng, og Olav Torp*

ISSN: 0803-9763 (papirversjon)

ISSN: 0804-5585 (nettversjon)

ISBN 978-82-92506-28-8 (papirversjon)

ISBN 978-82-92506-29-5 (elektronisk versjon)

### *Sammendrag:*

Rapporten starter med en kortfattet beskrivelse av nødvendig statistisk teori og noen sentrale sannsynlighetsfordelinger. Denne teoridelen følges opp med en beskrivelse av utvalgte metoder for estimering og beregning i forbindelse med usikkerhetsanalyser. Både analytisk modell og simulering er belyst. Det er gjort beregninger av formelverket for matematisk/statistisk beregning av tallgrunnlaget for analysene. Det konkluderes her med at de beregningene som foretas i for eksempel Trinnavis kalkulasjon og andre liknende analytiske metoder er svært robuste i forhold til avvik fra den virkelige sannsynlighetsfordelingen for de enkelte kostnadselementene. Et sentralt kapittel i rapporten er en analyse av hvor følsomme analyseresultatene er for feil i de estimerte inngangsdataene. Konklusjonene er at muligheten for og konsekvensene av grove feil er mye større når det gjelder estimering av inngangsdata enn de er når det gjelder antakelser om skjevhet i sannsynlighetsfordelingen, og også mye større enn de feil som gjøres i selve beregningsmodellen, enten man benytter analytisk modell eller simulering. Rapporten avsluttes med å beskrive grunnlaget for noen hjelpemidler som kan brukes som støtte i usikkerhetsanalyser.

*Dato: 13.12.2005*

### *Utgiver:*

*Concept-programmet*

*Institutt for bygg, anlegg og transport*

*Norges teknisk- naturvitenskapelige universitet*

*Høgskoleringen 7A*

*7491 NTNU - Trondheim*

*Tel. 73594640*

*Fax. 73597021*

*<http://www.concept.ntnu.no>*

Ansvaret for informasjonen i rapportene som produseres på oppdrag fra Concept-programmet ligger hos oppdragstaker. Synspunkter og konklusjoner står for forfatterens regning og er ikke nødvendigvis sammenfallende med Concept-programmets syn.

---

# Forord

Denne rapporten er nummer to i en serie på fem innenfor delprosjekt Usikkerhetsanalyser som en del av Concept-programmet. Concept er et forskningsprogram finansiert av Finansdepartementet, og har som formål å utvikle ny kunnskap om planlegging og gjennomføring av store offentlige prosjekter.

Innføring av ekstern kvalitetssikring av store statlige investeringer har bidratt til en utvikling i retning av en viss standardisering av usikkerhetsanalysene, og derav økt fokus på det teoretiske grunnlaget disse bygger på.

Rapporten søker å verifisere holdbarheten i de formelverk som benyttes ved analytiske beregninger, og gjør blant annet en analyse over følsomheten for feil i de forskjellige inngangsdata som benyttes.

Målgruppen for rapporten er mennesker som enten deltar i praktiske usikkerhetsanalyser i forbindelse med prosjektarbeid, eller som har behov for resultatene fra slike analyser.

Forfatterne vil rette en stor takk til førsteamanuensis Mette Langaas ved Institutt for matematiske fag, NTNU for hjelp med avklaring av deler av statistikkens mysterium.

Trondheim/NTNU 01.12.2005

Kjell Austeng

(Prosjektleder for Concept delprosjekt "Usikkerhetsanalyser")

# Innhold

Forord .....	3
Innhold .....	4
Figuroversikt .....	6
Tabelloversikt .....	8
Sammendrag .....	9
Summary .....	10
1. Innledning .....	11
2. Statistikk .....	12
2.1 Frekventativ statistikk .....	12
2.2 Bayesisk statistikk .....	13
2.3 Sannsynlighetsteori .....	14
2.4 Sannsynlighetsfordeling .....	15
2.5 Regning med usikre størrelser .....	19
2.6 Uavhengighet og korrelasjon .....	21
3. Noen sentrale sannsynlighetsfordelinger .....	22
3.1 Binomisk .....	22
3.2 Uniform .....	25
3.3 Trekant .....	27
3.4 Normal .....	29
3.5 Gamma- og Erlangfordeling .....	32
3.6 Eksponential .....	38
4. Estimering av inngangsdata til usikkerhetsanalyser .....	40
4.1 Tripplestimat .....	40
4.2 Grunnlag for estimering i prosjekter .....	40
4.3 Ulike tilnærminger i estimeringsprosessen .....	41
4.4 Usikkerhetsanalyser og bruken av subjektive vurderinger .....	46
5. Beregningsmetoder .....	47
5.1 Trinnvis kalkulasjon .....	48
5.2 Simulering .....	72
5.3 Simulering versus analytiske metoder .....	80
5.4 Resultater av beregninger .....	82
6. Følsomhet for feil i inngangsdata .....	84
6.1 Følsomhet for feil nedre anslag (n) .....	85
6.2 Følsomhet for feil anslag for mest sannsynlig verdi (s) .....	88
6.3 Følsomhet for feil øvre anslag ( $\theta$ ) .....	91
6.4 Konklusjoner om følsomhet for feil i inngangsdata .....	93
7. Noen spesialiteter .....	94
7.1 Følsomhetsanalyser .....	94
7.2 Regresjonsanalyser .....	97
7.3 Stokastisk tidstillegg .....	102

---

Litteratur	105
Vedlegg-A Gjennomsnitt - mål på sentraltendens .....	107
Vedlegg-B Verktøy for regresjonsanalyser .....	112
Vedlegg-C Hvordan finne Alfa, Beta og forskyvning fra trippelanslag ...	115
C.1 Algoritme for å finne Alfa.....	115
C.2 Regne ut skjevhetsforholdet fra alfa og beta .....	117
C.3 Algoritme for å finne beta .....	117
C.4 Algoritme for å finne forskyvning.....	118

# Figuroversikt

Figur 2-1	Eksempel på stolpediagram .....	12
Figur 2-2	Eksempel på diskret sannsynlighetsfordeling .....	12
Figur 2-3	Sannsynlighetsfordeling med forventningsverdi og mode .....	16
Figur 2-4	%-kvantilen, .....	17
Figur 2-5	Sammenhengen mellom tetthetsfunksjon og kumulativ f ordeling .....	18
Figur 2-6	Absolutt standardavvik knyttet til faktorer .....	20
Figur 3-1	Sannsynlighetsfordeling for antall oppdrag av fire mulige når vi har en binomisk fordeling, og sannsynligheten for oppdrag er 0,25 ved hvert forsøk. ....	23
Figur 3-2	Uniformfordelingen .....	25
Figur 3-3	Trekantfordeling .....	27
Figur 3-4	Normalfordelingen .....	29
Figur 3-5	Variasjon av Gammafordelingens formparameter. ....	32
Figur 3-6	Variasjon av Gammafordelingens skalarparameter. ....	33
Figur 3-7	Forskyvning av Gammafordelingen .....	33
Figur 3-8	Gammafordelingen og dens speilede variant .....	35
Figur 3-9	Eksponentialfordeling .....	38
Figur 4-1	Ulike alternativer for fremskaffelse av inngangsdata .....	42
Figur 4-2	Valg av beregningsmetode i en estimeringsprosess .....	45
Figur 5-1	Fordeling av skjevhetsforhold for Statens vegvesen og konsulenter .....	50
Figur 5-2	Typiske relative standardavvik for kostnadsanslag i Statens vegvesen .....	51
Figur 5-3	Theta optimal for 10/90 uttrykt ved Alfa .....	53
Figur 5-4	Theta optimal for 10 / 90 gitt ved skjevhetsforhold .....	54
Figur 5-5	Feil på forventningsverdi for Theta mellom 0,42 og 0,43 .....	55
Figur 5-6	Feil på forventningsverdi ved bruk av Theta = 0,42 .....	56
Figur 5-7	Feil på forventningsverdi ved bruk av Theta = 0,42, plottet mot Alfa .....	57
Figur 5-8	Feil på forventningsverdi ved bruk av Theta = 1 .....	58
Figur 5-9	Feil på forventningsverdi ved bruk av Theta = 1, plottet mot Alfa .....	58
Figur 5-10	Theta = 1 og Theta = 0,42 plottet mot hverandre .....	59
Figur 5-11	Feil på forventningsverdi for Theta = 1 for ulike skjevhetsforhold ved ekstreme relative standardavvik .....	60
Figur 5-12	Feil på forventningsverdi for ulike Theta verdier for ulike skjevhetsforhold ved ekstreme relative standardavvik .....	61
Figur 5-13	Zeta optimal for 10 / 90 for ulike verdier av Alfa .....	63
Figur 5-14	Zeta optimal for 10 / 90 for ulike skjevhetsforhold .....	64
Figur 5-15	Angitt kurve .....	65
Figur 5-16	Kurve som faktisk beregnes .....	65
Figur 5-17	Feil på standardavviket ved bruk av Zeta = 2,53 .....	66

Figur 5-18	Feil på forventningsverdi ved bruk av dynamisk Theta og Theta = 0,42 .....	68
Figur 5-19	Feil på standardavvik ved bruk av dynamisk Zeta og Zeta = 2,53 .....	68
Figur 5-20	S-kurve.....	73
Figur 5-21	Overordnet arbeidsgang i en simulering.....	76
Figur 5-22	Gangen i kjøring av en simulering.....	78
Figur 5-23	Eksempel på anbefalt tornadodiagram.....	82
Figur 6-1	Forventningsverdiens følsomhet for feil nedre anslag (n) for ulike standardavvik ved Alfa = 10.....	85
Figur 6-2	Forventningsverdiens følsomhet for feil nedre anslag (n) for ulike Alfa ved 10% standardavvik.....	86
Figur 6-3	Standardavvikets følsomhet for feil nedre anslag (n) for ulike Alfa .....	87
Figur 6-4	Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike standardavvik ved Alfa = 10.....	88
Figur 6-5	Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa ved 10% standardavvik.....	89
Figur 6-6	Standardavvikets følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa .....	89
Figur 6-7	Helt like i form, men speilvendt .....	90
Figur 6-8	Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike standardavvik ved Alfa = 10.....	91
Figur 6-9	Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike Alfa ved 10% standardavvik.....	92
Figur 6-10	Standardavvikets følsomhet for feil øvre anslag (ø) for ulike Alfa .....	92
Figur 7-1	Hvor følsom er internrenten for endringer i investeringskostnad? .....	94
Figur 7-2	Eksempel på “spiderdiagram”. .....	95
Figur 7-3	Hvilken merinvestering kan forsvares for å unngå 10% overskridelse av byggetid? .....	95
Figur 7-4	Hvilken merinvestering kan tillates for å oppnå 10% kortere byggetid? .....	96
Figur 7-5	Plotting av data for prosjektets størrelse opp mot prosentvis avvik. ....	98
Figur 7-6	Regresjonslinje for eksemplet.....	100
Figur 7-7	Stokastisk tidstillegg .....	103
Figur A-1	Forventningsverdi .....	108
Figur A-2	Mest sannsynlig verdi - Mode .....	109
Figur A-3	Sammenhengen mellom de tre målene for sentraltendens ....	110
Figur A-4	Relativ posisjon på høyreskjev fordeling .....	110
Figur A-5	Relativ posisjon på en s-kurve for en høyreskjev fordeling ...	111
Figur B-1	Forutsetninger og prosjektdata .....	112
Figur B-2	Plottet datasett med tilhørende trendlinje .....	113
Figur B-3	Analyse ved inndeling av erfaringsmaterialet i 10 delintervaller .....	113
Figur B-4	Regresjonsanalyse med 2 delintervaller .....	114



# Tabelloversikt

Tabell 4.1	Fordeler og ulemper ved formalistiske ekspertvurderinger kontra uformelle ingeniørmessige vurderinger .....	43
Tabell 5.1	Korrelasjonsmatrise .....	71
Tabell 5.2	Varians-kovariansmatrise .....	71
Tabell 5.3	Eksempel på tilfeldig-tall-generator .....	75
Tabell 5.4	Forskjeller mellom simulering og analytisk metode.....	80
Tabell 7.1	Erfaringsdata fra byggeprosjekter.....	98

---

# Sammendrag

*Frode Drevland, Kjell Austeng og Olav Torp*

## **Usikkerhetsanalyse - Modellering, estimering og beregning av**

Teoretisk grunnlag

Concept rapport nr. 11

Rapporten starter med en kortfattet beskrivelse av nødvendig statistisk teori og noen sentrale sannsynlighetsfordelinger.

Denne teoridelen følges opp med en beskrivelse av noen sentrale metoder for estimering og beregning i forbindelse med usikkerhetsanalyser. Både analytisk modell og simulering er belyst. I forbindelse med disse metodene er det gjort et sett med beregninger av formelverket for matematisk/statistisk beregning av det estimerte tallgrunnlaget for analysene.

Det konkluderes her blant annet med at de beregningene som foretas i for eksempel Trinnvis kalkulasjon og andre liknende analytiske metoder er svært robuste i forhold til avvik fra den virkelige sannsynlighetsfordelingen for de enkelte kostnadselementene.

Et sentralt kapittel i rapporten er en analyse av hvor følsomme analyseresultatene er for feil i de estimerte inngangsdataene. I flertallet av de analysemodeller som er i bruk i Norge blir sannsynlighetsfordelingen for enkeltestimater bestemt ved å fastlegge tre punkter på tetthetskurven. (Se Conceptrapport nr. 12 om metoder for usikkerhetsanalyser). Disse punktene er gjerne 10%-kvantilen (eller 1%-kvantilen), moden (eller forventningsverdien) og 90%-kvantilen (eller 99%-kvantilen). Det er gjort en analyse av konsekvensene av uriktige anslag for disse verdiene under forskjellige forutsetninger. Disse konsekvensene er allmenngyldige og gjelder enten man benytter analytiske beregninger eller simuleringer.

Konklusjonene er, for det første, at muligheten for og konsekvensene av grove feil er mye større når det gjelder estimering av inngangsdata enn de er når det gjelder antakelser om skjevhet i sannsynlighetsfordelingen, og også mye større enn de feil som gjøres i selve beregningsmodellen, enten man benytter analytisk modell eller simulering. Unøyaktigheter i formelverket gir også små utslag i denne sammenlikningen.

For det andre viser analysene at feil i input gir større utslag hvis relativ usikkerhet i den virkelige fordelingen er stor. Det samme gjelder hvis skjevheten i den virkelige fordelingen er stor. Ved bruk av 10%- og 90%-kvantilene som yttergrenser i anslagene, gir feil i estimeringen av disse større utslag enn feil i estimering av moden eller forventningsverdien.

Rapporten avsluttes med å beskrive grunnlaget for noen hjelpemidler som kan brukes som støtte i usikkerhetsanalyser.

# Summary

*Frode Drevland, Kjell Austeng and Olav Torp*

## **Uncertainty analysis - Modelling, estimation and calculation**

Theoretical foundation

Concept Report No. 11

The report starts with a short description of necessary statistical theory and some important probability distributions.

This theoretical part is followed up by a description of some central methods for estimation and calculation used in uncertainty analysis. Both analytical models and simulation is discussed. Related to this we have performed a set of calculations of the formulas for mathematical/statistical calculation of the estimated basis figures for the analyses.

One conclusion from the calculations is that Successive Calculation and other similar analytical methods are very robust in relation to deviations from the true probability distribution for the individual cost elements.

We have also looked at how sensitive the analysis results are for errors in the estimated input data. In the majority of analysis models that are being employed in Norway the probability distribution is set by determining three points on the probability density curve. (Se Concept Report No. 12 about methods for uncertainty analysis). These points are often the 10th percentile (or the 1th percentile), the mode and the 90th percentile (or the 99th percentile). An analysis has been made of the consequences of incorrect estimates for these values under different assumptions. These consequences are universal and apply whether ones uses analytical calculations or simulations.

The most important conclusion of the report is that the possibility for and the consequences of serious error are much larger when it comes to estimation of input data than for assumptions of skewness in the probability function, and also a lot larger than the errors that are made in the calculation model itself. This is regardless whether one uses an analytical model or simulation. Errors that occur due to inaccuracies in the formulas also are significantly smaller than errors caused by faulty estimates.

Further on the analyses show that the effects of errors in the input data are large if the relative uncertainty in the real probability distribution is high. The same applies if the skewness in the real distribution is large. When using the 10th and 90th percentile as outer boundaries in the estimates, errors in the estimation of these will yield larger errors than errors in the estimation of the mode or expected value.

The report finishes off by describing the basis for some tools that can be used as support in uncertainty analysis.

# 1. Innledning

Rapporten har fått tittelen “Modellering, estimering og beregning”. Denne tittelen er mer et signal om hva rapporten er ment å understøtte enn en naturlig oppdeling av stoffet den omhandler. Blant annet vil nok mange lesere føle at modelleringsdelen er dårligere dekket her enn hva tittelen kunne skape forventninger om. For mere om modellering vil vi henvise til Conceptrapportene nr. 10, 12 og 13.

Kapitlene 2 og 3 består av statistisk grunnkunnskap. Tanken bak dette er at de aller fleste som naturlig vil lese denne rapporten i utgangspunktet er kjent med stoffet, men har glemt detaljene. Stoffet er derfor forsøkt bearbeidet på en slik måte at det skal være en repetisjon som grunnlag for å få med seg det etterfølgende uten å måtte lete seg fram til andre kilder for å lese seg opp på grunnlagsstatistikken.

Kapittel 4, 5 og 6 dreier seg om estimering og metoder for tallbehandling med formål å etablere prosess eller elementdata (tid eller kostnad) som sannsynlighetsfordelinger, samt modellens følsomhet for feil i inngangsdataene.

Kapittel 7 og 8 omhandler grunnlaget for nyttige analyser for å finne sammenhenger mellom forskjellige typer data eller å finne følsomheten for endringer i parameterverdiene.

Når det gjelder valg av temaer som er tatt inn i rapporten har hovedretningslinjen vært å knytte dette opp mot metoder som er i praktisk bruk, og på et nivå som er tilpasset “problemeierne” mere enn analytikerne. Bakgrunnen for dette er vår erfaring om at usikkerhetsanalyser gjør størst nytte der hvor prosjektets folk er tungt involvert, og hvor profesjonelle analytikere er støtteapparat i den praktiske gjennomføringen.

I kapittel 5 om beregningsmetoder diskuteres både trinnvis kalkulasjon og simulering. Vi har i denne rapporten lagt mer vekt på trinnvis kalkulasjon på grunn av at det finnes langt mindre litteratur på emnet enn for simulering, og at det har vært endel usikkerhet og diskusjon i det norske prosjektmiljøet angående hva som er korrekte formler i trinnvis kalkulasjon.

I tillegg til selve rapporten med vedlegg er det lagt ut et regneark for regresjonsanalyse som Excel-fil og et program for analyse av Gammafordelinger på Concept sine hjemmesider: [www.concept.ntnu.no](http://www.concept.ntnu.no).

## 2. Statistikk

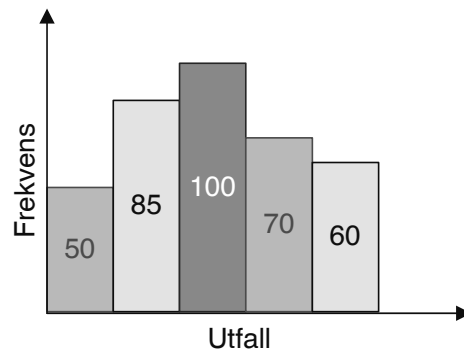
Bruken av statistikk og statistiske metoder er avhengig av mengden av data. Er mengden av data stor kan det lages modeller som beskriver den virkelighet vi ønsker å modellere med tilsvarende stor grad av nøyaktighet. De fleste rapporter fra Statistisk Sentralbyrå er eksempler på dette.

Situasjonen når det for eksempel gjelder prosjekter og prosjektusikkerhet er at mengden pålitelige data, og gjenbrukbarheten av disse dataene, er begrenset. Her er man da henvist til å bygge modeller ut fra en blanding av få pålitelige data, generell erfaring og mer eller mindre fornuftige resonneringer på denne bakgrunn.

Bevisstheten om denne avhengigheten av data har ført til at det er utviklet flere statistiske tilnæringsmåter for å kunne gjøre vurderinger om fremtiden. Vi skal i det etterfølgende si litt om de to viktigste.

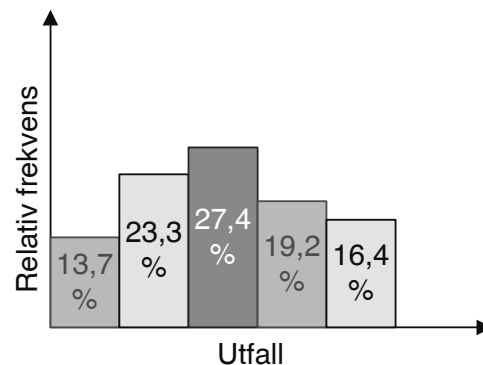
### 2.1 Frekventativ statistikk

Frekventativ statistikk går ut på å finne ut hvor ofte en hendelse skjer eller et forhold opptrer. Hvor ofte et forhold opptrer for eksempel pr tidsenhet kalles frekvens. Et resultat av et statistisk forsøk kalles et utfall. Alle mulige utfall kalles forsøkets utfallsrom. Hvis man gjør et forsøk mange ganger vil de forskjellige utfall opptre i forskjellig antall. Dette kan illustreres ved for eksempel et stolpediagram som vist i figur 2-1. I statistisk sammenheng kalles da dette gjerne et frekvensdiagram.



Figur 2-1 *Eksempel på stolpediagram*

Hvis et bestemt forhold som vi ønsker å undersøke i forsøket opptrer, kalles det et gunstig utfall. Sannsynligheten uttrykkes som forholdstallet mellom antall gunstige utfall og antall mulige utfall. Hvis vi da i stedet for et frekvensdiagram lager et diagram over relativ frekvens; dvs. vi deler alle tallene med antall mulige utfall (størrelsen på utfallsrommet), ser vi at vi har laget et diagram over de sannsynlighetene som opptrer. Vi har skapt en sannsynlighetsfordeling.



Figur 2-2 *Eksempel på diskret sannsynlighetsfordeling*

Sannsynlighet har betegnelsen P (probability).

Sannsynlighet kan sammenlignes med en mengde av størrelse 1, som på en eller annen måte skal fordeles på de muligheter som finnes (utfallsrommet). Denne fordelingen kalles sannsynlighetsfordeling. Vi ser at for-

delingen av  $P(A)$  fremkommer ved å ta frekvensdiagrammet ( $g$ ) og fordele det relativt i forhold til ( $N$ ).

$$1) \quad P(A) = \lim_{n \rightarrow \infty} \frac{g}{n}$$

hvor  $n$  er antall forsøk og  $g$  er antall gunstige utfall for  $A$  fra forsøkene

Sannsynligheten for en umulig hendelse er 0, og sannsynligheten for en sikker hendelse er 1.

Hvis utfallsrommet består av et begrenset antall muligheter (kast med mynt eller terning eller sågar spill i "Lotto"), sier vi at det foreligger en diskret sannsynlighetsfordeling (som den vi så i figur 2-2).

Hvis vi opererer innenfor et intervall hvor alle verdier er med i utfallsrommet, sier vi at vi har en kontinuerlig sannsynlighetsfordeling. I kontinuerlige fordelinger er utfallsrommet uendelig stort. Vi er derfor ikke i stand til å fastslå sannsynligheten for et bestemt utfall, men er henvist til å angi sannsynligheten som  $P(X \leq a)$ ,  $P(X \geq a)$  eller  $P(a \leq X < b)$ .

## 2.2 Bayesisk statistikk

I en frekventativ tilnærming tar vi, som vi har sett, utgangspunkt i at det forhold vi for eksempel skal anslå størrelsen på er ukjent, og at estimatet blir avledet fra mer eller mindre planlagte observasjoner som foretas. Estimering ut fra (mange) nøytrale observasjoner kalles *objektive metoder*.

En annen tilnærming, som går under navnet *Bayesisk estimering*, er å anta at de som skal gjøre vurderinger om et fenomen, har med seg kunnskaper om og erfaringer med fenomenet. Dette betyr ikke at man nødvendigvis må ha sett nøyaktig det samme før, og heller ikke at erfaringsbasen trenger å være veldig stor. Kombinasjonen kunnskap, erfaring og evnen til fornuftig resonnement danner basis for kvalifisert gjetning om størrelsesorden og tilhørende usikkerhet på for eksempel en kostnadspost.

Innen litteraturen om Bayesisk statistikk snakkes det om "grad av tillit" som måles i det som på engelsk blir kalt "likelihood" eller "the likely value", mens det i frekventistisk statistikk tilsvarende snakkes om "statistisk bevist" og "sannsynlighet" (probability).

For å bestemme sannsynlighet kan vi noen ganger ty til teoretiske betraktninger. Vi kan f.eks. uten å gjøre et eneste kast, fastslå at sannsynligheten for å få en 6-er i terning må være  $1/6$ .

Som oftest er vi imidlertid avhengig av et statistisk underlag som vi får ved å gjøre et stort antall forsøk, eller på annen måte samle inn statistisk materiale.

Svært ofte er det statistiske grunnlaget bare delvis dokumentert, eller ikke dokumentert i det hele tatt, men foreligger bare som erfaring, kunnskap og innsikt hos fagfolk innenfor det emne som omhandles.

Dette skaper grunnlag for å anta sannsynlighet ved hjelp av *subjektive vurderinger*. Subjektive vurderinger er nærmere beskrevet i kapittel 4 om estimering.

## 2.3 Sannsynlighetsteori

$A \cup B$  betyr at hendelsen A eller hendelsen B eller begge opptrer.  $A \cap B$  betyr at A og B opptrer samtidig.

Hvis A og B er uavhengige hendelser gjelder:

$$2) \quad P(A \cap B) = P(A) \cdot P(B)$$

$$3) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Hvis hendelsene A og B utelukker hverandre, (vi sier at de er disjunkte begivenheter)

er  $P(A \cap B) = 0$ .

Hvis  $\bar{A}$  betyr "ikke A" er  $P(\bar{A}) = 1 - P(A)$ . Vi kaller A og  $\bar{A}$  for komplementære begivenheter.

$P(A|B)$  betyr sannsynligheten for at hendelsen A skal inntreffe under forutsetning av at hendelsen B inntreffer. Dette kalles betinget sannsynlighet.

$$4) \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

### Stokastiske variabler

Før vi går over til å se på sannsynlighetsfordeling skal vi omtale begrepet stokastisk variabel. Det er en størrelse som varierer tilfeldig innenfor utfallsrommet.

Stokastiske variabler angis oftest med store bokstaver.

F.eks. X er antall øyne ved kast med terning, eller T er tiden det tar deg å kjøre på jobb om morgenen.

$P(X=2)$  betyr da sannsynligheten for å få en toer i for eksempel et terningkast.

$P(T < 40 \text{ min})$  betyr tilsvarende sannsynligheten for at du skal bruke mindre enn 40 min. på å komme deg på jobb.

## 2.4 Sannsynlighetsfordeling

Senere skal vi ta for oss noen sentrale sannsynlighetsfordelinger. Her i statistikkapitlet vil vi imidlertid repetere om sannsynlighetsfordeling generelt, og noen begreper fra statistikken og sannsynlighetsregningen.

Vi har, som nevnt i kapittel 2.1, to hovedtyper sannsynlighetsfordelinger. Den ene typen er fordelinger av sannsynlighet hvor utfallsrommet består av avgrensede størrelser, for eksempel  $\{1.0, 1.5, 2.0, 2.5 \text{ osv.}\}$ . Dette kalles *diskret* sannsynlighetsfordeling. I en rekke forhold hvor sannsynligheten er diskret fordelt er det lett å regne ut hvor stor del som faller på hvert mulig utfall. For eksempel kan vi, uten å ha gjort et eneste forsøk, fastslå at sannsynligheten for å få summen 4 ved å kaste to terninger må være  $1/12$ .

Regnestykket er som følger: Summen 4 fremkommer ved tre mulige utfall; kombinasjonene  $(1,3)$ ,  $(2,2)$  og  $(3,1)$ . Hvert av utfallene har sannsynligheten  $(1/6)^2 = 1/36$ . Tre mulige kombinasjoner gir  $3 \cdot \frac{1}{36} = \frac{1}{12}$ .

Den andre hovedtypen sannsynlighetsfordeling dekker et uendelig utfallsrom som for eksempel et åpent eller lukket tallintervall. Denne typen kalles *kontinuerlige* fordelinger. Når det gjelder kontinuerlige fordelinger er det ikke like enkelt å regne seg fram til sannsynligheter. Her er vi avhengig av å finne en standard sannsynlighetsfordeling som dekker vårt problem med stor nok nøyaktighet. I kapittel 3 vil vi beskrive noen standard sannsynlighetsfordelinger som er mye brukt.

### 2.4.1 Begreper

For å beskrive en fordeling har vi tre viktige karakteristika:

- Beliggenhet
- Spredning
- Formen på fordelingsdiagrammet.



## Forventningsverdi og varians

Beliggenheten angis oftest ved den såkalte forventningsverdien,  $E(X)$ , som er x-verdien til tyngdepunktet av fordelingen.

For diskrete fordelinger gjelder:

$$5) \quad E(X) = \sum_0^n p(x) \cdot x$$

Spredningen ut fra dette tyngdepunktet angis ved variansen ( $VAR(X)$ ):

$$6) \quad VAR(X) = \sum_0^n (x-E(x))^2 \cdot p(x)$$

For kontinuerlige fordelinger gjelder:

Beliggenhetsmålet

$$7) \quad E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (\text{jfr. figur 2-3})$$

Spredningsmålet

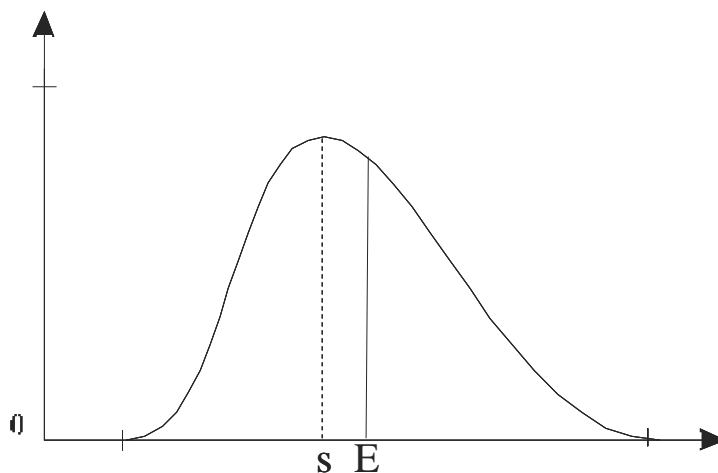
$$8) \quad VAR(x) = \int_{-\infty}^{\infty} [(x-E(x))^2 \cdot f(x)] dx \quad .$$

For begge typer fordelinger gjelder at standardavviket:

$$9) \quad \sigma = \sqrt{VAR(X)}$$

Standardavviket er det mest brukte målet for spredningen av fordelingen; altså om kurven er høy og trang eller lav og vid. Standardavviket angis som en absolutt størrelse med samme benevnelse som forventningsverdien, eller som en ubenevnt relativ størrelse:

$$10) \quad \sigma_{rel} = \frac{\sigma}{E}$$



Figur 2-3 Sannsynlighetsfordeling med forventningsverdi og mode

## Mode - mest sannsynlig verdi

Figur 2-3 viser også et annet typisk beliggenhetsmål for kontinuerlige fordelinger, nemlig moden, eller også kalt mest sannsynlig verdi, ( $s$ ). Moden er x-koordinaten for toppunktet på fordelingskurven. Tilsvarende har diskrete fordelinger det såkalte typetallet, som er det utfallet med sannsynlighetsmassen 1,0 eller 100% er representert ved det totale arealet mellom fordelingskurven og x-aksen.

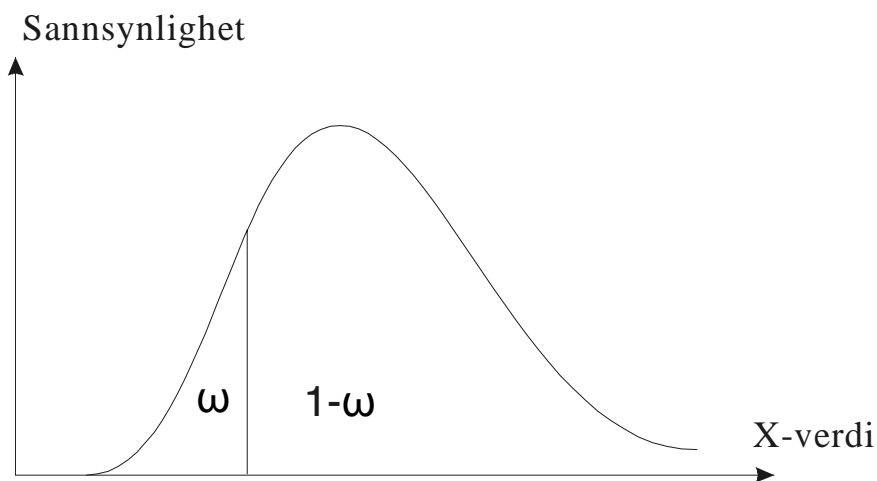
Figuren gir et eksempel på en sannsynlighetsfordeling som vil dukke opp mange ganger i den etterfølgende del av rapporten. Fordelingen er usymmetrisk slik at tyngdepunktet av sannsynlighetsmassen ligger til høyre for toppen av kurven. Vi sier at vi har en høyreskjev fordeling.

## Kvantiler og median

Kvantiler er en annen type beliggenhetsmål. De angir hvor på sannsynlighetskalaen en verdi befinner seg.

En  $\omega$ %-kvantil er den verdien av den stokastiske variable som er slik at det er  $\omega$ % sannsynlig at den *ikke* blir overskredet. Se figur 2-4.

$$11) \quad P(X \leq x_{\omega}) = \omega$$



Figur 2-4  $\omega$ %-kvantilen,  $x_{\omega}$

For eksempel er 20%-kvantilen den x-verdien som har 20% av sannsynlighetsmasser til venstre,

$$P(X \leq x_{0,2}) = 0,2$$

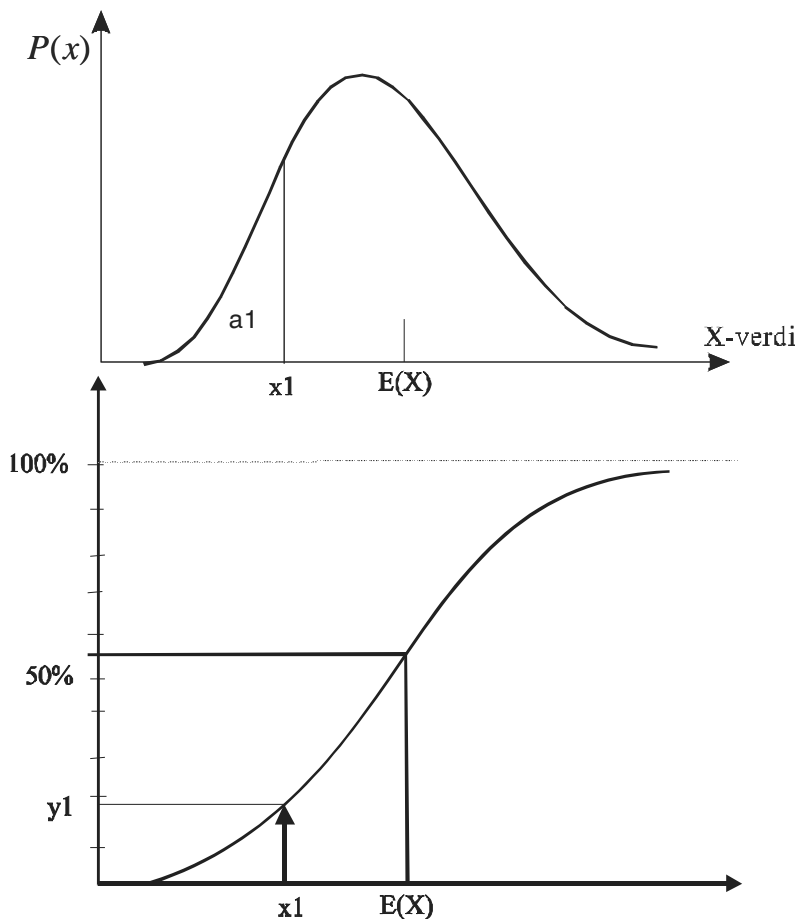
$$P(X > x_{0,2}) = 1 - P(X \leq x_{0,2}) = 0,8$$

50%-kvantilen kalles median, og 25%- og 75%-kvantilene kalles h.h.v. nedre og øvre kvartil.

## 2.4.2 Kumulative sannsynlighetsfordelinger

Fordeling av sannsynlighet består, som vi har sett, i å tilegne sannsynlighet til visse verdier på x-aksen. Sannsynligheten for en viss størrelse måles ved mengden sannsynlighet som er plassert over størrelsen. Å fastlegge sannsynligheten for en bestemt verdi i en kontinuerlig fordeling er derfor enkel; den er nemlig 0. Forklaringen er at alle bestemte verdier i et kontinuerlig intervall er punkter uten utstrekning, og på et slikt punkt er det følgelig ikke plass for noen sannsynlighetsmengde. Vi er henvist til å snakke om sannsynlighet for at vår stokastiske variabel er mindre enn eller større enn en gitt verdi, eller sannsynlighet for at den ligger innenfor et intervall;  $P(X \in [a, b])$  eller  $P(a \leq X \leq b)$ .

For å illustrere dette tyr vi til en fremstillingsform som er enklere og mer anvendelig enn tetthetsfunksjonen; nemlig en kumulativ fordeling. Den kumulative fordelingen fremkommer ved at man starter på den laveste verdien i sannsynlighetsintervallet og kontinuerlig summerer sannsynlighet mens man vandrer oppover (fra venstre mot høyre). Kurven for den kontinuerlige fordelingen vil dermed gå fra 0 (eller nær 0) til 100% (eller nær 100%). Se figur 2-5.



Figur 2-5 Sammenheng mellom tetthetsfunksjon og kumulativ fordeling

Høyden  $y_1$  i den kumulative kurven tilsvarer arealet  $a_1$  i tetthetsfordelingen.  $P(X \leq x_1) = y_1$ .

Den kumulative kurven er naturlig nok brattest der hvor sannsynligheten er tettest; dvs. toppunktet på tetthetsfunksjonen. Dette fører til at når tettheten beskrives av en klokkeformet kurve vil den kumulative kurven få en typisk S-form; derav navnet *S-kurve*.

## 2.5 Regning med usikre størrelser

For å få fram en prosjektkostnad med tilhørende usikkerhet gjelder noen enkle regneregler:

- Forventet prosjektkostnad,  $E_{\text{tot}}$ , består av summen av forventningsverdiene til kostnadselementene.

$$12) \quad E_{\text{tot}} = E_1 + E_2 + \dots + E_n$$

- Spredningen til prosjektkostnaden uttrykkes ved variansen  $\text{Var}(X_{\text{tot}})$ . Variansen til en sum (eller differens) er summen av variansene forutsatt at alle postene er statistisk uavhengige. (Se kapittel 2.6 om uavhengighet og korrelasjon)

$$13) \quad \text{Var}(X_{\text{tot}}) = \text{Var}(x_1) + \text{Var}(x_2) + \dots + \text{Var}(x_n)$$

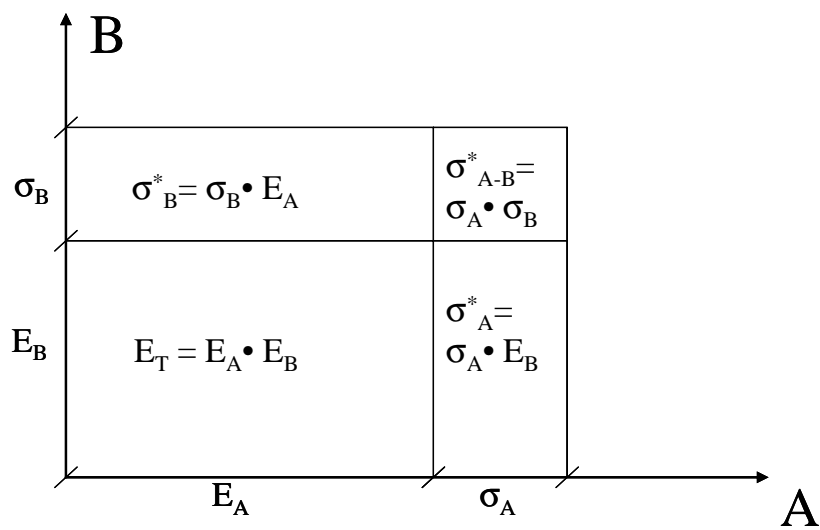
- Det er mengden av kostnadselementer og størrelsen på usikkerheten i dem, som avgjør usikkerheten på totalcostnaden, ikke fortegnet på elementene.

Svært ofte i tilknytning til kalkyle med tilhørende usikkerhetsvurdering har man det forhold at det kan være usikkerhet i mengdeanslagene og samtidig usikkerhet i prisanslagene; altså man må multiplisere to (eller flere) usikre forhold.

Problem:

- Kostnad (T) = Mengde (A) · enhetskostnad (B), hvor A og B er faktorer med forventningsverdi henholdsvis ( $E_A$ ) og ( $E_B$ ), og standardavvik henholdsvis ( $\sigma_A$ ) og ( $\sigma_B$ ).
- Hva blir middelvei ( $E_T$ ) og standardavvik ( $\sigma_T$ ) for produktet T?

Regnereglene er forklart ved hjelp av nedenstående figur 2-6.



Figur 2-6 *Absolutt standardavvik knyttet til faktorer*

Og svaret er:

Middelvei for produktet T:

- $E_T = E_A \cdot E_B$

Usikkerheten til produktet:\*

- $VAR(T) = (\sigma_A^*)^2 + (\sigma_B^*)^2 + (\sigma_{A-B}^*)^2$

## 2.6 Uavhengighet og korrelasjon

Utgangspunktet for de fleste analysemetoder, og trinnvismetoden spesielt, er at en posts sannsynlighetsfordeling er statistisk uavhengige av alle andre posters sannsynlighetsfordelinger. Det vil for eksempel si at hvis en post skulle vise seg i virkeligheten å havne i 90% kvantilen av hva som var anslått, så skal det fortsatt være (teoretisk) mulig for samtlige andre poster å ende opp på 10% kvantilen. Det er klart at for at dette skal være mulig må man ha urelaterede årsaksforhold. Om en post får en kostnad større enn forventet av en gitt årsak, kan det ikke også føre til at en eller flere andre poster får kostnader større eller mindre enn forventet. Mer presist kan vi si at to hendelser A og B er uavhengige hvis og bare hvis den enes opptreden (eller ikke opptreden) ikke har noen innvirkning på den andre hendelsen.

Matematisk definerer vi dette som at to tilfeldige variabler  $X=f(x)$  og  $Y=f(y)$  er uavhengig hvis den felles sannsynlighetsfordelingen  $f(x,y) = f(x)f(y)$ .

Hvis man har to poster som ikke er statistisk uavhengige, sier vi at de er korrelerte. Graden av korrelasjon uttrykkes som regel ved hjelp av en korrelasjonskoeffisient som ligger mellom -1 og 1. Korrelasjon har ingen betydning for postenes samlede forventningsverdi, men det har stor betydning for postenes samlede varians og standardavvik. For to uavhengige poster er det like sannsynlig at de nuller hverandre ut (den ene verdien blir lav og den andre høy) som at de forsterker hverandre (begge postene får enten høye eller lave verdier). Hvis man har korrelerte poster vil man ha økt sannsynlighet for at totalutslaget blir mindre eller at det blir større, ved henholdsvis negativ og positiv korrelasjonsfaktor, som dermed fører til en øking eller minking av postenes samlede varians. Denne tilleggsvariansen kalles kovarians.

Den totale kovariansen som følge av to posters korrelasjon kan uttrykkes som følger:

$$14) \quad Cov(X, Y) = 2Cor(X, Y) \cdot \sigma_X \cdot \sigma_Y$$

der  $Cor(X,Y)$  er korrelasjonsfaktoren og  $Cov(X,Y)$  er kovariansen.

## 3. Noen sentrale sannsynlighetsfordelinger

Som vi har anført tidligere har vi to hovedtyper av sannsynlighetsfordelinger; diskrete og kontinuerlige. Diskrete fordelinger har liten plass i usikkerhetsanalyser for prosjekter, hvor det stort sett dreier seg om tid, kostnad og nytte. I det nedenstående er tatt med ett eksempel på en fordeling som kan tenkes å være til hjelp i de få tilfellene hvor forholdet er diskret fordelt.

### 3.1 Binomisk

Dette er den eneste diskrete fordelingen vi vil ta med blant beskrivelsene av viktige standardfordelinger.

Binomisk betyr to-leddet, og i statistisk betydning dekker det forhold eller fenomener som hver gang de opptrer har to mulige utfall; dvs. utfallsrommet er av typen  $\{0,1\}$ ,  $\{\text{Ja, Nei}\}$ ,  $\{\text{Av, På}\}$  osv.

En binomisk fordeling får vi når slike fenomener opptrer flere ganger, og vi søker å finne sannsynligheten for de forskjellige antall ganger ett av utfallene kan forekomme.

Eksempel:

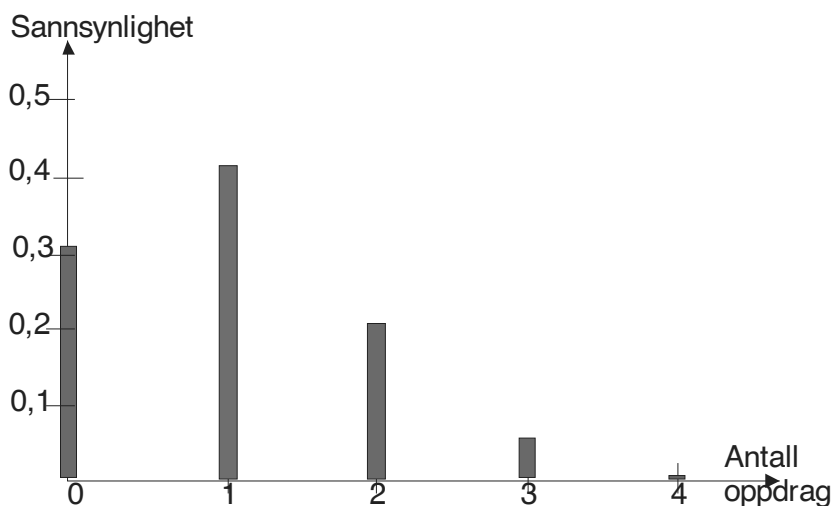
En entreprenør leverer jevnlig tilbud i et bestemt marked. Sannsynligheten for å få jobben er 25%. Han leverte fire tilbud nesten samtidig, og vil finne sannsynligheten for at han får to oppdrag. Ved å sette 0 = ikke oppdrag og 1 = oppdrag, og telle binært, får vi følgende  $2^4 = 16$  kombinasjoner:

0000  
0001  
0010  
0011 OK  
0100  
0101 OK  
0110 OK  
0111  
1000  
1001 OK  
1010 OK  
1011  
1100 OK  
1101  
1110  
1111

Vi ser at seks av de 16 kombinasjonene gir to oppdrag. Sannsynligheten for hver av disse kombinasjonene er  $(0,25)^2(1-0,25)^2$ , og sannsynligheten for å få akkurat to oppdrag er da:  $P(X = 2) = 6 \cdot (0,25)^2 \cdot (0,75)^2$ ; altså 21%.

Vi ser også at sannsynligheten for ingen oppdrag (kombinasjonen 0000) er  $1 \cdot (0,75)^4 = 0,32$ .

Sannsynlighetsfordelingen blir som vist i figur 3-1.



Figur 3-1 *Sannsynlighetsfordeling for antall oppdrag av fire mulige når vi har en binomisk fordeling, og sannsynligheten for oppdrag er 0,25 ved hvert forsøk.*

Vi så i det ovenstående at sannsynligheten fremkom ved å multiplisere antall gunstige kombinasjoner med sannsynligheten for kombinasjonen. Generelt er formelen som følger:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad \text{når } x = 1, 2, 3, \dots, n.$$

$p$  er sannsynligheten for et gunstig utfall hver gang,  $x$  er antall gunstige utfall, og  $n$  er antall mulige utfall, og

$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!} \quad (\text{i vårt tilfelle } \frac{4!}{2! \cdot 2!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 1 \cdot 2} = 6)$$

Forventningsverdien  $E(X) = np$ , og spredningen  $\text{VAR}(X) = np(1-p)$ .

Brukt på vårt eksempel blir  $E = 4 \cdot 0,25 = 1$ . Dvs. vi forventer å få treff på hvert fjerde anbud.  $\text{VAR} = 4 \cdot 0,25(1 - 0,25) = 0,75$ , og standardavviket  $\sigma = 0,87$ .



## Formler for Binominalfordelingen

---

### Parametre

---

$n$  - diskret "antall"-parameter  $n > 0$

$p$  - kontinuerlig suksessparameter  $0 < p < 1$

---

### Forventningsverdi

---

$$E(x) = np$$

---

### Varians

---

$$\sigma^2 = np(1 - p)$$

---

### Mode (Mest sannsynlige verdi)

---

Ikke entydig definert

---

### Tetthetsfunksjon

---

$$f(x) = \binom{n}{x} p^x (1 - p)^{(n - x)}$$

---

### Kumulativ tetthetsfunksjon

---

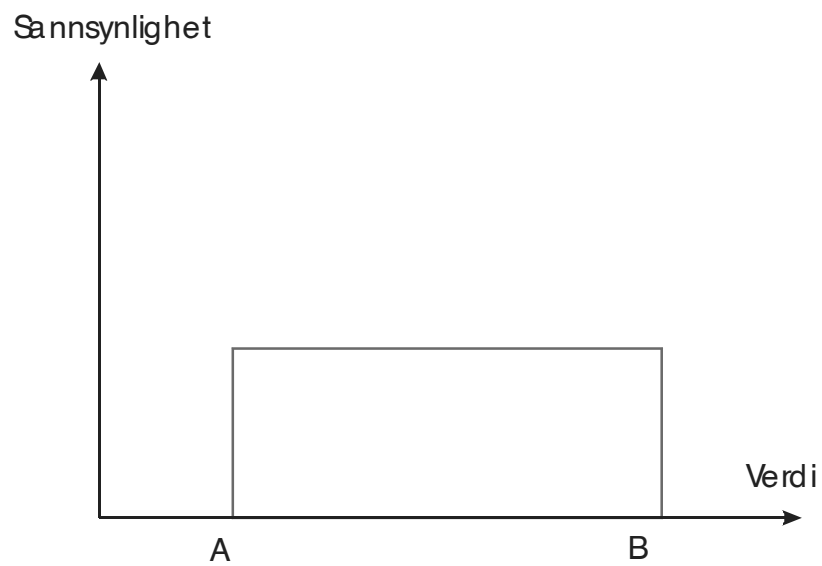
$$F(x) = \sum_{i=0}^x \binom{n}{i} p^i (1 - p)^{(n - i)}$$

## 3.2 Uniform

Den enkleste kontinuerlige fordelingen er uniform fordelingen. Denne er begrenset av en minimumsverdi og en maksimumsverdi, og alle verdiene fra og med minimumsverdien til og med maksimumsverdien er like sannsynlig.

Ofte brukt som en første modell for en mengde som man antar varier mellom de to ytterpunktene, men som man ellers vet lite om.

I forbindelse med simulering er denne fordelingen essensiell for å generere verdier for alle andre sannsynlighetsfordelinger.



Figur 3-2 *Uniformfordelingen*

## Formler for Uniformfordelingen

---

### Parametre

---

$a$  - kontinuerlig grenseparameter  $a < b$

$b$  - kontinuerlig grenseparameter

---

### Forventningsverdi

---

$$E(x) = \frac{b - a}{2}$$

---

### Varians

---

$$\sigma^2 = \frac{(b - a)^2}{12}$$

---

### Mode (Mest sannsynlige verdi)

---

Ikke entydig definert

---

### Tetthetsfunksjon

---

$$f(x) = \frac{1}{b - a}$$

---

### Kumulativ tetthetsfunksjon

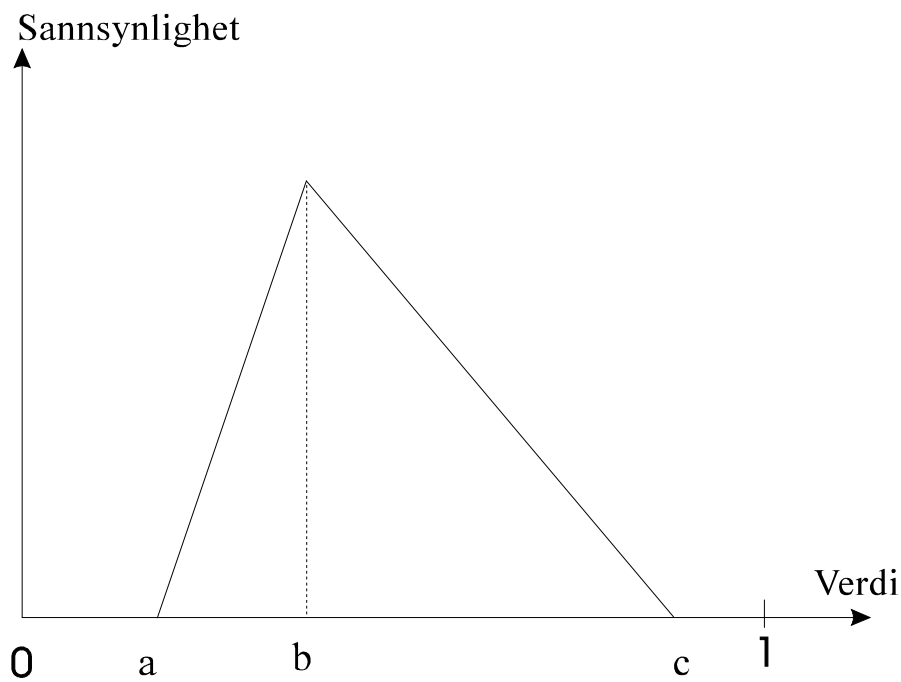
---

$$F(x) = \frac{x - a}{b - a}$$

### 3.3 Trekant

I noen tilfeller har vi forhold som er underlagt tilfeldighetenes lov, men har klart definerte grenser både nedover og oppover. Et eksempel på dette er anslag over sannsynlighet. Her vil verdiene være i intervallet  $[0,1]$ . Innenfor et anslått sannsynlighetsintervall vil det normalt være minst sannsynlighet for ytterverdiene.

En stokastisk modell som dekker disse forhold kan beskrives ved trekantfordelinger.



Figur 3-3 *Trekantfordeling*

## Formler for Trekantfordelingen

---

### Parametre

---

$a$  - kontinuerlig nedre grenseparameter  $a < c$

$b$  - kontinuerlig parameter for mest sannsynlig verdi  $a \leq b \leq c$

$c$  - kontinuerlig øvre grenseparameter

---

### Forventningsverdi

---

$$E(x) = \frac{a+b+c}{3}$$

---

### Varians

---

$$\sigma^2 = \frac{1}{18}[(c-a)^2 + (c-b)(a-b)]$$

---

### Mode (Mest sannsynlige verdi)

---

$$s = b$$

---

### Tetthetsfunksjon

---

$$f(x) = \frac{2(x-a)}{(b-a)(c-a)} \quad a \leq x \leq b$$

$$f(x) = \frac{2(c-x)}{(c-b)(c-a)} \quad b \leq x \leq c$$

---

### Kumulativ tetthetsfunksjon

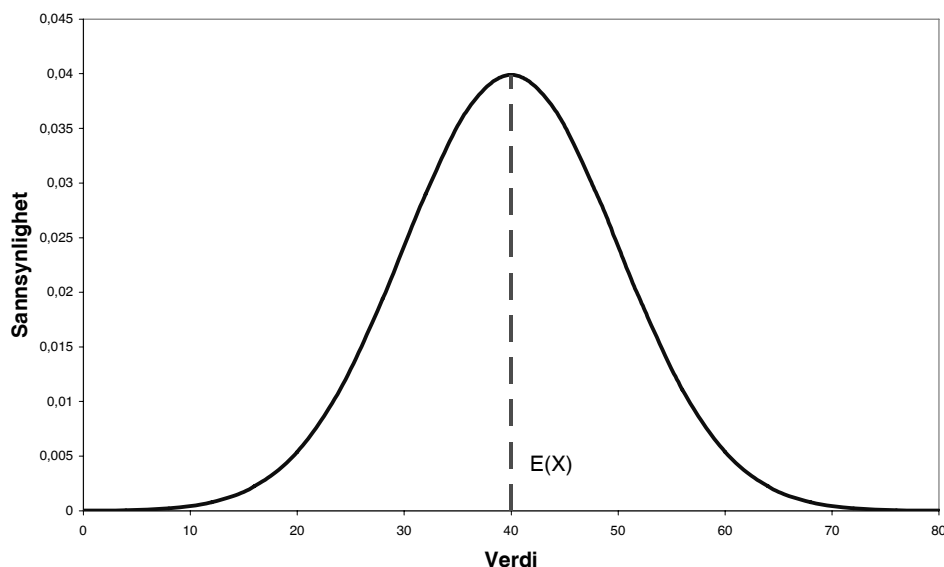
---

$$F(x) = \frac{(x-a)^2}{(b-a)(c-a)} \quad a \leq x \leq b$$

$$F(x) = 1 - \frac{(c-x)^2}{(c-b)(c-a)} \quad b \leq x \leq c$$

### 3.4 Normal

Normalfordelinger forekommer i mange situasjoner. Hvis man observerer størrelser som i utgangspunktet skulle være like, f.eks. tidsforbruket på et bestemt arbeidsstykke, vil man se at tidsforbruket varierer rundt en eller annen middelværdi med tilnærmet like mange observasjoner under middelværdien som over. De fleste observasjoner ligger i nærheten av middelværdien, men noen få avviker relativt sterkt fra denne. Etter hvert som man øker antall observasjoner, vil hyppighetsdiagrammet få mer og mer preg av en hatteform.



Figur 3-4 *Normalfordelingen*

Fordelingen er symmetrisk om middelværdien,  $E(X)$ , som dermed også blir mest sannsynlig verdi og 50%-kvantilen.

Fordelingen beskrives ved hjelp av parametrene  $\mu$  og  $\sigma^2$ , hvor  $\mu = E(X)$  og  $\sigma^2 = VAR(X)$ , og vi sier at en stokastisk variabel  $X$  er normalfordelt  $N(\mu, \sigma^2)$ .

For alle normalfordelinger gjelder:

$$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0,68$$

## Sentralgrenseteoremet

Sentralgrenseteoremet uttrykkes gjerne på en av to måter:

- Summen av et stort antall stokastiske verdier med tilfeldige fordelinger tenderer mot å være normalfordelt. Begrensningen er at verdiene er uavhengig variabler, og at ingen av verdiene er svært dominerende i forhold til de andre.
- Gjennomsnittet av et stort antall verdier med samme vilkårlige sannsynlighetsfordeling tenderer mot å være normalfordelt.

Vi ser at det ikke er noe krav om at grunnverdiene skal være normalfordelte, eller ha samme fordeling.

“Stort antall” er gjerne satt til  $N \geq 30$ , men sentralgrenseteoremet gjelder med bra tilnærming selv for  $n = 10$ .

Hvor fort tilnærmingen skjer til et nøyaktighetsnivå vi kan leve med er avhengig av fordelingen til de enkelte postene, innbyrdes størrelse mellom postene og nøyaktighetskravet.

Høyland kaller tendensen til normalfordeling for asymptotisk normalitet (Høyland 1979). I det ligger erkjennelsen av at tilnærmingen blir bedre og bedre med antall elementer i summen, men at vi aldri vil oppnå en eksakt normalfordeling. All erfaring fra praktisk usikkerhetsanalyse tilsier imidlertid at sammenliknet med usikkerheten i input til analysen er de relativt små unøyaktighetene i tilpasning til den fordelingsfunksjonen som benyttes av underordnet betydning.

---

## Formler for Normalfordelingen

---

### Parametre

---

$\mu$  - kontinuerlig beliggenhetsparameter

$\sigma$  - kontinuerlig skaleingsparameter

---

### Forventningsverdi

---

$$E(x) = \mu$$

---

### Varians

---

$$\sigma^2 = \sigma^2$$

---

### Mode (Mest sannsynlige verdi)

---

$$s = \mu$$

---

### Tetthetsfunksjon

---

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

---

### Kumulativ tetthetsfunksjon

---

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Hvor  $\Phi$  er Error-funksjonen



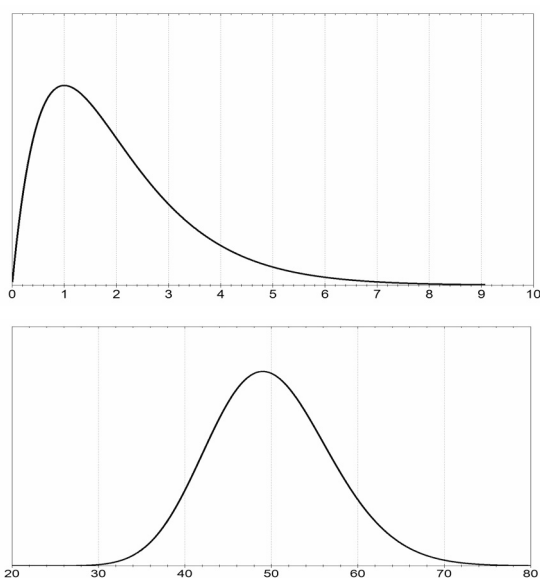
## 3.5 Gamma- og Erlangfordeling

Disse fordelingene er her beskrevet i langt mer detalj enn de andre fordelingene i dette kapitlet. Dette er på grunn av at store deler av den videre rapporten på forhold ved nettopp disse fordelingene.

### 3.5.1 Gammafordelingen

Gammafordelingen er en høyreskjev fordeling gitt av to parametere; Alfa og Beta. Alfa avgjør kurvens form, det vil si hvor høyreskjev den er, mens Beta er en skalarparameter som avgjør tallenes størrelse.

#### Form



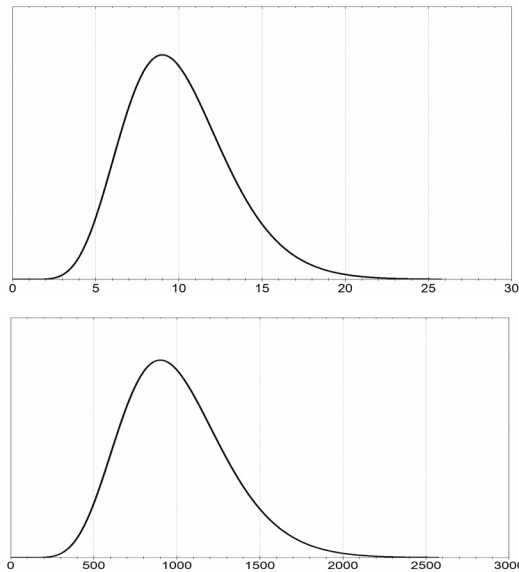
**Figur 3-5** *Variasjon av Gammafordelingens formparameter. Øverst: Alfa = 2, Beta = 1. Nederst: Alfa = 50, Beta = 1*

Et viktig fakta om gammafordelingen er at den er per definisjon høyreskjev, det vil si at fordelingsens mest sannsynlige verdi ligger til venstre for fordelingsens median. Som vi kan se av den øverste kurven i figur 3-5 er fordelingen svært høyreskjev ved lave verdier av Alfa. Ved større verdier av Alfa vil kurven bevege seg mot en normalfordeling, som vi kan se av den nederste kurven. Hvis Alfa går mot uendelig vil kurvens form gå mot en normalfordeling, men den vil alltid være litt høyreskjev selv om det er snakk om svært små verdier som det vil være umulig å se det med det blotte øye. Dette kan dog ha konsekvenser om man bruker gammafordelinger i simuleringer og enkelte analytiske metoder. Ved bruk av tilnæringsformler av den typen Statens Vegvesen benytter er ikke dette noe problem.

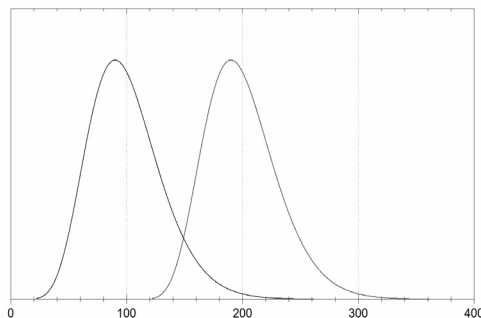
---

*TIPS! Ved bruk av simuleringsverktøy er det bedre å bytte ut gammafordelingen med en normalfordeling hvis man har et symmetrisk fordelt element.*

---



**Figur 3-6** *Variasjon av Gammafordelingens skalarparameter.*  
 Øverst:  $\text{Alfa} = 10$ ,  $\text{Beta} = 1$ .  
 Nederst:  $\text{Alfa} = 10$ ,  $\text{Beta} = 100$



**Figur 3-7** *Forskyvning av Gammafordelingen*  
 $\text{Alfa} = 10$ ,  $\text{Beta} = 10$ ,  
 Forskyvning = 0 og 100

### Skalering

Både Alfa og Beta bidrar til å skalere fordelingen som vi kan se i figur 3-5 til 3-8. Men at kurven blir skalert som følge av at Alfa øker er kun en bi effekt. Alfas funksjon er å bestemme kurvens form. Beta derimot bidrar kun til å skalere kurven. Som vi ser av figur 3-6 er kurvene identisk i form, men tallene på x-aksen har blitt skalert opp på den nederste kurven med en faktor på hundre som følge av at beta er endret fra 1 til 100.

### Plassering

Gammafordelingen starter per definisjon i null. Men man kan forskyve fordelingen slik man vil ved å legge til et vilkårlig tall til alle avleste eller beregnede verdier.

Hvis vi ser disse tre parameterne i forhold til trinnvisparametrene  $n$ ,  $s$  og  $\emptyset$  beskrevet i neste kapittel, kan vi si at Beta bestemmer avstanden mellom  $n$  og  $\emptyset$ , Alfa bestemmer hvor  $s$  ligger relativt til  $n$  og  $\emptyset$  og forskyvningsparameteren bestemmer fordelings nullpunkt og er altså da et tall som blir lagt til flatt til alle verdier som hentes ut av fordelingen (det vil si  $n$ ,  $s$ ,  $\emptyset$  og forventningsverdi etc.).

## 3.5.2 Erlangfordelingen

Erlangfordelingen er et spesialtilfelle av gammafordelingen. Hvis vi har en gammafordeling der Alfa er et heltall får vi Erlangfordelingen. Erlangfordelingens formparameter kalles  $k$  i stedet for Alfa. Denne har den fordelen at den har noe enklere matematisk formel. De opprinnelig Trinnvis-formlene som er omtalt i neste kapittel er basert på Erlangfordelingen. Det synes å ha vært flere ulike grunner til at man opprinnelig valgte å benytte Erlangfordelingen, men slik trinnvis-formlene brukes per i dag er det Gamma-fordelingen som er den korrekte fordelingen å forholde seg til, så Gamma-fordelingen blir brukt i det følgende i denne rapporten.

### 3.5.3 Noen formler for en forskjøvet Gammafordeling

---

#### Parametre

---

$\alpha$  - kontinuerlig formparameter  $\alpha > 0$

$\beta$  - kontinuerlig skalarparameter  $\beta > 0$

$\varphi$  - forskyvningsparameter

---

#### Forventningsverdi

---

$$E(x) = \alpha\beta + \varphi$$

---

#### Varians

---

$$\sigma^2 = \alpha\beta^2$$

---

#### Mode (Mest sannsynlige verdi)

---

$$s = \beta(\alpha - 1) + \varphi$$

---

#### Tetthetsfunksjon

---

$$f(x) = \begin{cases} \frac{\beta^\alpha (x - \varphi)^{\alpha - 1} e^{-(x - \varphi)/\beta}}{\Gamma(\alpha)} & \text{hvis } (x - \varphi) > 0 \\ 0 & \text{ellers} \end{cases}$$

hvor  $\Gamma(\alpha)$  er Gammafunksjonen definert som  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$  for alle reelle tall  $\alpha > 0$

---

#### Kumulativ tetthetsfunksjon

---

$$F(x) = \begin{cases} \frac{\Gamma(\alpha, (x - \varphi)/\beta)}{\Gamma(\alpha)} & \text{hvis } (x - \varphi) > 0 \\ 0 & \text{ellers} \end{cases}$$

hvor  $\Gamma$  er den Inkomplette gammafunksjonen definert som  $\Gamma(z, a) = \int_a^\infty t^{z-1} e^{-t} dt$  for alle reelle tall  $z > 0$

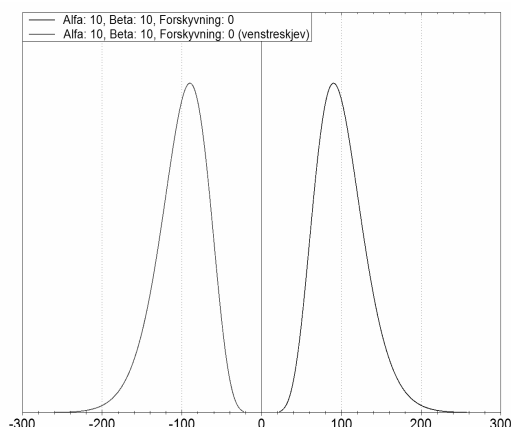
### 3.5.4 Venstreskjeve gammafordelinger

Gammafordelingen er per definisjon høyreskjev, men det vil ofte hende at man ønsker å beregne et kostnadselement man antar har en venstreskjev fordeling. Dette avsnittet tar for seg hvordan ved å benytte en speilet gammafordeling kan oppnå dette. Dette er plassert her for å ha teorien om gammafordelingen samlet på et sted, men det forutsetter at man er kjent med konseptet av trippelanslag og trinnvisformlene som omtales i neste kapittel.

#### Definisjonen av en venstreskjev gammafordeling

En uforskjøvet gammafordeling går fra null til uendelig. Vi definerer en venstreskjev fordeling til å være den speilede varianten av dette, det vil si en fordeling som går fra null til minus uendelig. Figur 3-8 viser eksempel på en vanlig høyreskjev gammafordeling og en speilet venstreskjev variant.

På parametersiden definerer vi skalarparameteren Beta til å være negativ for venstreskjeve kurver. Dette gjør at vi kan bruke formlene for forventningsverdi og varians som beskrevet i faktabeskrivelsen på forrige side.



Figur 3-8 *Gammafordelingen og dens speilede variant*

For å kunne bruke formelen for tetthetsfunksjonene må vi gjøre et par triks. Vi bruker formelen for tetthetsfunksjonene som normalt, men vi snur fortegnet på Beta og  $x$  i formelen. For den vanlige tetthetsfunksjonen er et vilkårlig kurvepunkt for tetthetsfunksjonen for en venstreskjev kurve gitt som  $(x, f(-x))$  der Beta i  $f(x)$  er positiv. Når det gjelder den kumulative tetthetsfunksjonen må man i tillegg snu på resultatet ellers vil S-kurven gå gal vei. Et vilkårlig punkt på den kumulative tetthetskurven er da gitt ved  $(x, 1-f(-x))$  der Beta i  $f(x)$  er positiv.

## Transformasjon av trippelanslag

De færreste vil ha behov for å befatte seg med formlene fra forrige avsnitt for speiling av tetthetsfunksjonene, men det kan være tilfeller der man har behov for speile trippelanslag. Hvis man bruker de lineære tilnæringsformlene kjent fra trinnvis kalkulasjon er ikke venstreskjeve trippelanslag noe problem, men hvis man skulle ønske å bruke mer nøyaktige analytiske modeller eller simulering står man ovenfor et problem. Gammafordelingen er per definisjon høyreskjev. Men dette betyr ikke at det er umulig å få beregnet verdier for et venstreskjevt anslag. Man må bare speile fordelingen slik at den fremstår som høyreskjev og trikse litt. I det følgende er en oppskrift for hvordan man går fram.

$n$  = nedre anslag,  $s$  = mest Sannsynlig Anslag,  $\emptyset$  = øvre Anslag

Ved et venstre skjevt anslag  $n,s,\emptyset$  gjøres følgende:

1. Snu fortegnene på alle tre anslagene.
2. Snu på rekkefølgen

Summert opp:

Transformert  $N,S$  og  $\emptyset$  blir i forhold til det opprinnelige trippelanslagt uttrykt som

$$n_{\text{Trans}} = - \emptyset$$

$$s_{\text{Trans}} = - s$$

$$\emptyset_{\text{Trans}} = - n$$

3. Regn ut forventningsverdien av trippelanslaget og snu fortegnet på resultatet

$$E(x) = - E(x)_{\text{Trans}}$$

## Eksempel på utregning av en venstreskjev fordeling

Trippelanslag: 60-100-110

$$n_{\text{Trans}} = -\phi = -110$$

$$s_{\text{Trans}} = -s = -100$$

$$\phi_{\text{Trans}} = -n = -60$$

Nytt trippelanslag = -110, -100, -60

$$E(x)_{\text{Trans}} = \frac{-110 + (0,42(-100)) - 60}{2,42} = 87,6$$

$$E(x) = -E(x)_{\text{Trans}} = -(-87,6) = 87,6$$

Regner ut opprinnelig trippelanslag direkte:

$$E(x) = \frac{60 + 0,42 \cdot 100 + 110}{2,42} = 87,6$$

Som vi ser av beregningsresultatene gir metodene for speiling samme resultatet som å regne ut verdien direkte ved hjelp av trinnvisformelen for forventningsverdien. Tilfeller hvor det er nødvendig å gjøre denne øvelsen er beskrevet senere i rapporten.

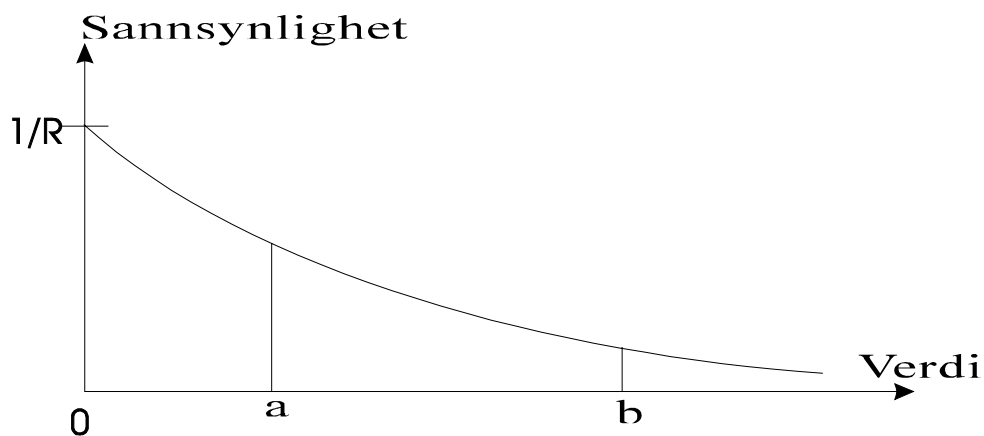
## 3.6 Eksponential

En rekke hendelser er av den art at sannsynligheten for at de skal opptre i løpet av et tidsintervall er konstant og uavhengig av når hendelsen sist opptrådte. Sannsynligheten for slike hendelser kan beskrives ved funksjonen:

$$f(t) = \frac{1}{R} e^{-\frac{t}{R}} \text{ for } t \geq 0 \text{ hvor } E(T) = R \text{ og } VAR(T) = R^2$$

$$F(t) = P(T \leq R) = \int_0^R \frac{1}{R} e^{-\frac{t}{R}} = \left[ -e^{-\frac{t}{R}} \right]_0^R = 1 - \frac{1}{e} = 0,63$$

T er den stokastiske variable: "Tiden mellom hver gang hendelsen inntreffer"



Figur 3-9 Eksponentialfordeling

Funksjonen har den egenskap at  $P(T \leq b | T \geq a) = P(T \leq b - a)$  for  $0 \leq a \leq b$

Dette er karakteristisk for en rekke fenomener i naturen. F.eks. vil sannsynligheten for at en 100-års flom vil opptre før det har gått 50 år siden sist, gitt at det alt har gått 30 år, altså innen 20 år fra nå, være den samme som at det vil komme en 100-års flom i løpet av de første 20 år etter at det nettopp har vært en slik flom.

Nedenfor følger et eksempel på bruk av eksponentialfordeling, kanskje med et litt overraskende resultat.

Hva er sannsynligheten for at en syklisk hendelse som nettopp har skjedd skal opptre én gang til i løpet av et tidsintervall tilsvarende returperioden?

$$F(t) = P(T \leq R) = \int_0^R \frac{1}{R} e^{-\frac{t}{R}} = \left[ -e^{-\frac{t}{R}} \right]_0^R = 1 - \frac{1}{e} = 0,63$$

Det er altså 63% sannsynlig at 100-årsflommen skal komme igjen i løpet av de første 100 år.

---

## Formler for Eksponentialfordelingen

---

### Parametre

---

$\beta$  - kontinuertlig skalarparameter  $\beta > 0$

---

### Forventningsverdi

---

$$E(x) = \beta$$

---

### Varians

---

$$\sigma^2 = \beta^2$$

---

### Mode (Mest sannsynlige verdi)

---

$$s = 0$$

---

### Tetthetsfunksjon

---

$$f(x) = \frac{e^{-x/\beta}}{\beta}$$

---

### Kumulativ tetthetsfunksjon

---

$$F(x) = 1 - e^{-x/\beta}$$



## 4. Estimering av inngangsdata til usikkerhetsanalyser

Som i andre typer analyser er en i usikkerhets- og risikoanalyser avhengig av god input for å få et riktig bilde av usikkerheten i prosjektet. En har ulike tilnærminger til å få fram slik input. Det benyttes gruppeprosesser, intervju med eksperter, etc. Alt dette må i denne sammenheng kunne kalles ekspertvurderinger. Ekspertvurderinger benyttes for å “karakterisere” innsamlingen av kvantitative data i mangel av statistiske data (Øien, m.fl. 1996). Disse ekspertvurderingene kan i mer eller mindre grad bygge på statistiske data.

Dette kapitlet gir en oppsummering over ulike tilnærminger til bruk av ekspertvurderinger i estimeringssammenheng. For en mer dyptgående gjennomgang av disse metodene og alternative estimeringsmetoder henvises til Concept rapport nr. 12, “*Usikkerhetsanalyse - Metoder*”.

### 4.1 Trippelestimat

Her i Norge er tripplestimat omtrent enerådende som inngangsdata til usikkerhetanalyser. Et tripplestimat, eller trepunktsestimat, består som navnet tilsier, av tre estimater. Vanligvis anslår man en nedre verdi, den mest sannsynlige verdien og en øvre verdi. Den nedre og den øvre verdien er alltid symmetriske kvantiler på fordelingen.

Det er per i dag to ulike “kvantilpar” som brukes i usikkerhetsanalyse, 1/99 og 10/90. De har hver sine fordeler og ulemper avhengig av hva man sitter på av erfaringstall og hva domeneekspertene best til å forholde seg til. For eksempel kan 1/99 være lettere å forholde seg til fordi man snakker om bortimot best tenkelig og verst tenkelige tilfelle, mens 10/90 typisk er lettere å finne relevante erfaringstall. Rent teknisk metodisk er ingen av dem overlegne.

### 4.2 Grunnlag for estimering i prosjekter

Et kjennetegn ved et prosjekt er at det er unikt, noe du gjør en gang (Rolstadås 1997). Da vil det ofte være begrenset hvilket statistisk grunnlag en har for eksempel som input til kostnadsestimering i prosjekter.

Det bygges mange meter vegger og mange kvadratmeter bygg pr. år, og basert på innsamlede erfaringsdata fra slike prosjekter kan en få fram statistikk på hvor mye en meter veg eventuelt en kvadratmeter bygg koster. Dette benyttes som input i en estimeringsprosess for fremtidige prosjekter. Eksempel på slik statistikk finnes i Holteprosjekt kalkulasjonsnøkkelen (se [www.holteprosjekt.no](http://www.holteprosjekt.no) for informasjon om denne statistikken). Statens vegvesen er i ferd med å etablere en kostnadsbank for å ta vare på og kunne analysere erfaringstall fra vegprosjekter. Andre prosjekteiere og utførere har tilsvarende statistikk som benyttes inn i estimering, ved blant annet etablering av kostnadsoverslag, kalkulasjon av anbud/tilbud og usikkerhetsanalyser.

Et nasjonalt operabygg er det ikke så ofte en bygger, og ved estimering av kostnader for et slikt prosjekt vil det være veldig begrenset hva slags statistikk en har å bygge på. En kan finne noen byggtypen som kan ligne på operahus, men gyldigheten av disse tallene benyttet på et operahus vil være begrenset. Videre i dette kapitlet diskuteres ulike tilnærminger til estimeringen avhengig av hvilket grunnlag en har å bygge på.

### 4.3 Ulike tilnærminger i estimeringsprosessen

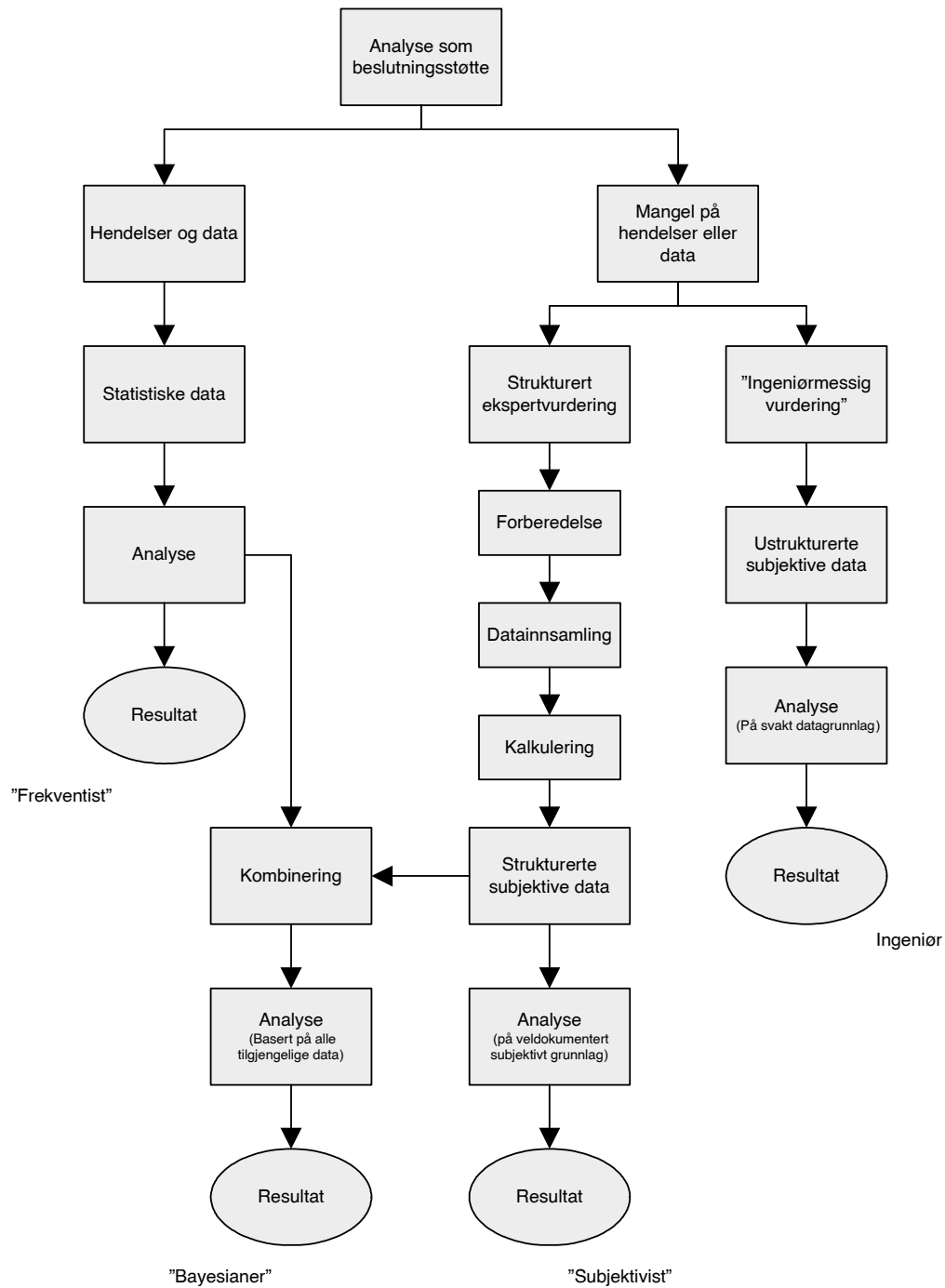
(Øien m.fl. 1996) stiller en del krav til metoder for ekspertvurderinger:

1. Dokumentasjon
2. Objektivitet
3. Empirisk kontroll (Estimatene bør kunne kontrolleres empirisk)
4. Kompletthet
5. Enkelthet

For å gjøre troverdige vurderinger som er etterprøvbare, er det sentralt at ekspertvurderinger tilfredsstillende kravene 1 - 4. Krav 5 har med anvendelse av metodene å gjøre. For at ting skal bli benyttet i det praktiske er det behov for enkle metoder. Avveininger i metodene som skal benyttes blir da å finne så enkle metoder som mulig, samtidig som de tilfredsstillende kravene 1 - 4.

Den ideelle input til analyser vil være dersom en har mange observerte registreringer på tilsvarende situasjoner. Da har man statistiske data, og disse kan benyttes inn i analysen. Disse vil i stor grad oppfylle kravene 1 - 5. Dette er benevnt "frekventistenes" alternativ. I motsatt ende av skalaen vil situasjonen være at det ikke finnes noen statistiske data å legge til grunn. En er avhengig av subjektive data, som kanskje også er innsamlet på ustrukturert og tilfeldig vis. Dette er referert til som "Ingeniørmessige" vurderinger. Ofte benyttes tommelfingerregler. Metodene vil i høyeste grad tilfredsstillende krav 5, men en vil ofte i liten grad kunne si at vurderingene tilfredsstillende kravene 1 - 4, som er sentralt for å kunne virke troverdige.

Dersom en har samlet inn subjektive data på en systematisk og strukturert måte kan en oppnå å gjøre de subjektive dataene så objektive som mulig, og dermed kunne benyttes subjektive data på en slik måte at kriteriene 1 - 5 på beste mulige måte blir tilfredsstillende. Dette vil øke titroen til datasettet. Dette vil være "subjektivistens" alternative tilnærming. Har en noen statistiske data, men ikke tilstrekkelig til å basere en analyse på kun rene statistiske data, vil en kombinere dette datagrunnlaget med ekspertvurderinger. Dette kalles "Bayesianerens" alternativ. Tradisjonelt benyttes ekspertvurderinger, som oppdateres etter hvert som statistiske data samles inn.



Figur 4-1 Ulike alternativer for fremskaffelse av inngangsdata (Øien m.fl., 1996)

Øien beskriver fordeler og ulemper ved formalistiske ekspertvurderinger i forhold til ingeniørmessige vurderinger som input til kvantitative verdier i modeller og beregninger (Øien m.fl. 1996) . Dette er gjengitt i tabell 4.1.

**Tabell 4.1** *Fordeler og ulemper ved formalistiske ekspertvurderinger kontra uformelle ingeniørmessige vurderinger*

Faktorer	Formell ekspertvurdering	Ingeniørmessig vurdering
Struktur	Systematisk og strukturert fremgangsmåte/prosess.	Usystematisk og ustrukturert prosess. "Diskusjon over bordet"
Spesifisering av informasjon	Godt spesifisert. Kun informasjon gitt som svar på vel definerte spørsmål.	Upresist. Antagelser ikke spesifisert.
Dokumentasjon	Alle trinn i prosedyren er god dokumentert.	Dårlig eller ingen dokumentering.
Omfang av innsamlet informasjon	Begrenset. Kun den som fås gjennom forhåndsdefinerte spørsmål.	Bred. Kan dekke mange aspekter ved temaet, også via oppfølgingsspørsmål.
Evaluering av eksperter	"Objektive" regler for evaluering og eventuell vektning av ekspertene.	Tiltroen til en bestemt ekspert bedømmes subjektivt av analytiker.
Enkelthet	Omfattende og kostbar	Veldig enkel. Utføres uten forberedelse.

Situasjonen i prosjekter er ofte mangel på statistiske data. Funderinger, brainstorming, gjettinger, og spekulasjoner utført av eksperter og brukt som underlag i strukturerte beslutningsprosesser er av relativt ny dato. Beskrevne metoder er blant andre scenariometoden og Delphi-metoden.

### 4.3.1 Delphi-metoden

Delphi-metoden er den mest kjente metoden for å fremskaffe og behandle eksperter sine meninger (Øien m.fl. 1996). Den kan summeres opp i 8 trinn:

1. Et "observasjonsteam" (analytikere) definerer målsetninger og velger ut respondenter (eksperter). Respondentene er vanligvis anonym for hverandre og svarene er anonyme.
2. Observasjonsteamet forbereder et foreløpig spørreskjema som de sender ut til respondentene for kommentarer.
3. Svarene gjennomgås og et endelig spørreskjema etableres.
4. Respondentene svarer på spørreskjemaene
5. Observasjonsteamet analyserer svarene og beregner medianverdier og interkvartilspredningen (25% og 75% estimatene).
6. Resultatene returneres til respondentene som blir spurt om de ønsker å endre på svarene deres. De som da fortsatt havner utenfor interkvartilene blir bedt om å gi argumenter for deres anslag.
7. Reviderte anslag blir prosessert nok engang og argumenter for "utstikkerne" blir beskrevet. Denne informasjonen blir sendt tilbake til respondentene og prosessen blir gjentatt 3 - 4 ganger.
8. Median-verdiene etter siste runde blir brukt som beste anslag. Generelt er spredningen i siste runde mindre enn i første runde, noe som menes å indikere en grad av konsensus.

### 4.3.2 Gjennomføring av ekspertvurderinger

Forskningsprogrammet Prosjektstyring år 2000 etablerte en håndbok for gjennomføring av ekspertvurderinger (Øien m.fl 1996). Håndboken beskriver et prosedyreverk for hvordan en bør samle inn ekspertvurderinger. Prosedyreverket er delt inn i tre faser:

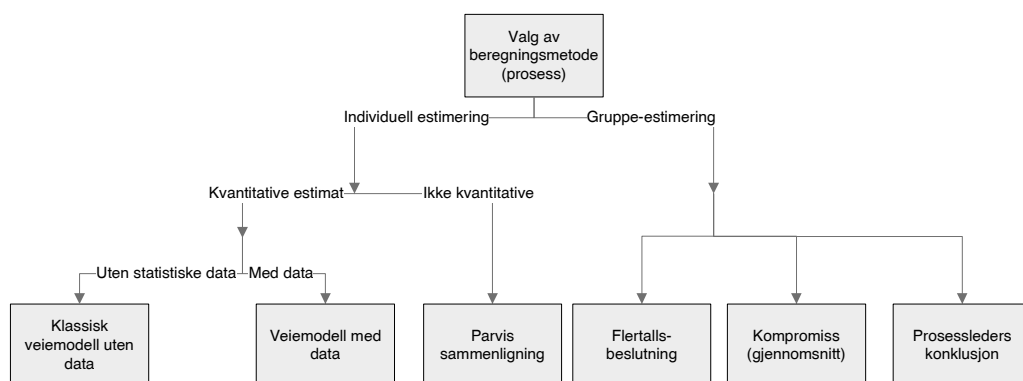
FASE I: Forberedelsesfasen

FASE II: Utspørringsfasen

FASE III: Beregningsfasen

Forberedelsesfasen kan på mange måter sammenlignes med planleggingsfasen i et prosjekt. Hensikten er å forberede og organisere den forestående ekspertvurderingen slik at best mulig resultat kan oppnås med de ressurser man har til rådighet. Fasen omfatter beskrivelse av problem, evaluering av tilgjengelige ressurser, valg av eksperter, evaluering av eksperter, valg av beregningsmetode og etablering av intervjuformat. I usikkerhetsanalyse er de 6 H'ene benyttet som beskrivelse av forberedelsesfasen (Hvem, Hva, Hvordan, Hvormed, Hvortid, og

Hvorfor), se (Husby m.fl 2000), Valg av beregningsmetode er anbefalt gjort etter prinsipper angitt i figur Valg av beregningsmetode i en estimeringsprosess (Øien m.fl. 1996).



Figur 4-2 Valg av beregningsmetode i en estimeringsprosess (Øien m.fl. 1996)

Utspøringsfasen vil trekke ut så mye relevant og objektiv kunnskap som mulig ut av ekspertene. Innholdet i fasen går på følgende: Informasjon til ekspertene, Gjennomføring av utspørring, Etablering av 3-punkts estimater, Etablere kunnskap for vektning av ekspertene, Oppsummering/avsluttende diskusjon, og evt. endring av estimatene. Utspørring av ekspertene kan gjøres individuelt eller i gruppe. Ved individuell utspørring må en sørge for at ekspertene får gjennomført estimeringen uten at de forstyrres/påvirkes av de andre ekspertene. Ved utspørring i gruppe må en sørge for at prosessleder styrer diskusjonen slik at en enkelt ekspert ikke får dominere for mye (Klakegg 1993). Det kan være en fordel at ekspertene gjør et individuelt estimat som de noterer ned før diskusjonen om felles estimat.

Det kan være nødvendig å foreta en vektning av ekspertene. Uten et rasjonelt grunnlag for å vekte ekspertene bør imidlertid ikke dette gjøres. En eventuell vektning kan gjøres på følgende måter (Øien m.fl. 1996):

- Kontrollspørsmål i utspørringen som blandes med de reelle spørsmål, og som av ekspertene vil bli besvart på samme måte som de reelle spørsmålene.
- Gjensidig ekspertvektning, hvor ekspertene gjør en vurdering av hverandre ved å fylle ut et skjema/matrise for gjensidig vurdering.
- Ekspertene fyller inn skjema med opplysninger om deres kunnskapsprofil, og deretter legges størst vekt på ekspertene med mest relevant kunnskap.

Gjennom beregningsfasen skal en få frem de resultater/estimater som legges frem for beslutningstaker basert på de estimater som ekspertene gir enkeltvis eller i gruppe. Innholdet i fasen er Kalibrering av ekspertene, Vektning av ekspertene, Beregning av felles estimat, Dokumentasjon, Kommunikasjon/presentasjon, Etterarbeid. Kalibrering gjennomføres hvis det er klare indikasjoner på at eksperten systematisk over/underestimerer. (Øien m.fl 1996) beskriver kriterier og eksepler på kalibrering av eksperter i en ekspertvurdering. Måter å vekte de ulike ekspertene på er også beskrevet i Øien.

Ved individuell estimering vil det være en jobb med å etablere et felles estimat. Det felles estimat vil basere seg på ekspertenes individuelle estimater/uttalelser/begrunnelser, men det kan være foretatt kalibrering og ulike vektninger av disse individuelle estimater. I en gruppe-

estimering er som oftest allerede det felles estimat etablert gjennom gruppeprosessen i utspøringsfasen.

Prosedyreverket bygger på eksisterende litteratur, samt erfaringer fra gruppen som utførte prosjektet. Mesteparten av litteraturen som omhandler ekspertvurderinger tar for seg utvalgte problemstillinger, samt at de er innrettet mot svært omfattende ekspertvurderinger med bruk av store ressurser (Øien m.fl. 1996). Usikkerhetsanalyser vil i liten grad passe inn i denne beskrivelsen. Ressursene vil være begrenset, og analysene skal gjøres i begrenset tidsrom og med begrensninger i antall eksperter som en involvert. Det er imidlertid sentralt å få kvalitet på input. Derfor er det jobbet mye med metoder for å sikre at input i usikkerhetsanalyser har tilstrekkelig kvalitet.

## 4.4 Usikkerhetsanalyser og bruken av subjektive vurderinger

Ulike tilnæringer til usikkerhetsanalyser er beskrevet i litteraturen. Disse benytter i stor grad subjektive vurderinger som input i estimeringsprosessen. Trinnavis-prosessen (Klakegg 1993) beskriver en fremgangsmåte for usikkerhetsanalyser. Metoden er basert på gruppearbeid. Estimeringen baseres på subjektive vurderinger. Subjektive vurderinger kan beskrives som *kvalifisert gjetning basert på kompetente persons fagkunnskap, erfaringer, intuisjon og beste skjønn*. I trinnavis-prosessen vil subjektive vurderinger være basert på og støttet av statistikk fra tidligere gjennomførte prosjekter, og erfaringer som ekspertene har med seg inn i estimeringen, utnyttet på en strukturert og systematisk måte.

Subjektive vurderinger utført av kompetente mennesker (eksperter) er benyttet som tilnærming til usikkerhetsanalyser i mange år (Lichtenberg 2000). Forskning viser at (1) de fleste av oss er lite dyktige til å gjøre subjektive vurderinger, og (2) at vi gjennom læring og trening kan bli betydelig bedre. Dersom estimatene i tillegg gjøres av grupper med eksperter er en i stand til å nøytralisere fallgruvene ved subjektive vurderinger. Dette gjelder spesielt når gruppene er ledet av en erfaren fasilitator/prosessleder (Klakegg 1993) med kunnskaper innen evalueringsteknikker (Lichtenberg 2000).

Analysegruppen bør bestå av fire til femten eksperter. Personer som er en del av prosjektteamet i det prosjektet som skal analyseres er oftest positive til prosjektet, og kanskje ikke i så stor grad i stand til å se risikoene i prosjektet. Det er derfor viktig å ha med personer (minst en) utenfra prosjektet, og gjerne utenfra organisasjonen. Som grunnlag for estimeringen i gruppeprosessen må en være veldig nøye med å beskrive de enkelte elementene som skal estimeres, slik at gruppen har den samme og riktig oppfatning av elementene før estimeringen foretas.

Klakegg beskriver en prosess (Klakegg 1993), bestående av teknikker som er benyttet for å gjøre subjektive vurderinger mer troverdige, og mer i stand til å tilfredsstille kravene til ekspertvurderinger, beskrevet tidligere i kapitlet. Det er lagt vekt på å tilfredsstille kravene 1 - 5 nevnt tidligere i kapitlet. Ulike metoder legger vekt på ulike sider ved en slik prosess, og bygger modeller og gjør beregninger på forskjellige måter. Likevel er input, enten det er samlet inn ved hjelp av ekspertintervju eller ved bruk av gruppearbeid, i stor grad basert på subjektive vurderinger.

## 5. Beregningsmetoder

Det finnes to prinsipielle måter å beregne stokastiske kostnadsanslag; ved bruk av analytiske metoder (for eksempel suksessive/trinnvis kalkulasjon) eller ved bruk av simuleringmetoder (for eksempel Monte Carlo simulering).

Ved bruk av analytiske metoder beskrives kostnadene for hvert element ved et matematisk uttrykk. Hver av kostnadene regnes så i hop til en nytt uttrykk som gir fordelingen av mulige utfall. Å modellere og regne en kostnadsanalyse ved bruk av eksakte matematiske sannsynlighetsfordelinger og beregningsmetoder er både meget tidkrevende og krever betydelige matematiske ferdigheter. Trinnvis kalkulasjon er en analytisk metode "light". I stedet for å regne eksakt benytter man lineære tilnæringsformler som gir et svar som er tilstrekkelig nøyaktig. Trinnvis kalkulasjon ble opprinnelig utviklet som et regneverktøy for bruk i trinnvis metoden, derav navnet trinnvis kalkulasjon. Men det er viktig å ha klart for seg at trinnvis kalkulasjon og trinnvis metoden ikke er det samme. Trinnvis metoden er verktøyuavhengig og man kan i den like gjerne bruke Monte Carlo simulering som trinnvis kalkulasjon eller andre analytiske metoder.

Monte Carlo simulering er mer en "rå-makt"-metode. I stedet for å regne gjennom kalkylen en gang regner man gjennom den fra noen hundre til flere tusen ganger. Det vil si, man får naturligvis en datamaskin til å gjøre det for seg. For hver gjennomregning trekkes en tilfeldig verdi ut fra kostnadssannsynlighetsfordelingene og benyttes i beregningen. Med andre ord for hver gjennomregning "kaster" datamaskinen en "terning" for å finne ut hvilken verdi den skal bruke for en gitt kostnad. Datamaskinen vil for hver gjennomregning ta vare på de tallene man er interessert i. Etter at simuleringen er kjørt ferdig vil den da ha et statistisk grunnlag for utarbeide en sannsynlighetsfordeling for sluttsummen og andre tall vi måtte ønske å få ut.

Begge tilnærmingene skal i prinsippet gi det samme resultatet når de arbeider med den samme innputten, det vil si når dataene og kalkylestrukturen er lik.

Når det gjelder analytiske metoder begrenser vi oss i det videre til å se på Trinnvis-kalkulasjon, samt å presentere er ny metode hvor vi i stedet for å bruke lineære tilnæringsformler bruker trippelanslaget til å regne ut Gammafordelingens parametere og, videre brukere Gammafordelingens formler for forventningsverdi og standardavvik som presentert i kapittel 3-5.

Dette kapitlet inneholder en god del mer om trinnvis kalkulasjon og analytiske metoder enn om simulering. Dette skyldes at vi parallelt med arbeidet på denne rapporten gjorde noen utredninger på formelverket for trinnvis kalkulasjon som følge av at det var en del uenigheter om dette i det norske prosjektstyringsmiljøet, og vi syntes det var på sin plass å presentere disse resultatene i denne rapporten.



## 5.1 Trinnvis kalkulasjon

Dette kapitlet tar for seg formelverket som benyttes i Trinnvis-kalkulasjon. Vi har sett på hva størrelsen på formlenes sentrale parametere bør være samt hvor nøyaktige variantene som brukes av disse formlene per i dag er samt hvor nøyaktig de potensielt kan bli.

### 5.1.1 Trinnvis formlene

De opprinnelige formlene for trinnvis eller suksessiv kalkulasjon ble utarbeidet av Steen Lichtenberg på 70-tallet. Ifølge Lichtenberg kan sannsynlighetsfordelingen til et kostnadselement i et prosjekt uttrykkes ved hjelp av en Erlangfordeling. Formlene Lichtenberg utarbeidet var basert på at man hadde  $k = 10$  for fordelingen. I sine bøker skriver han at dette representerer den mest typiske skjevheten. Videre skriver han at selv om den faktiske skjevheten skulle være noe annet så ville formlene gi en tilstrekkelig god tilnærming, og at det selv med rimelig ekstreme forskjeller i skjevhet mellom formelens ideal og inngangsdata, vil man kun få et par promille feil for forventningsverdien. I bøkene hans er det så vidt vi kan finne ikke ført form for bevis for dette eller på andre måter dokumentert.

Formlene (se boksen til høyre) tar utgangspunkt i at man gjør et tripelanslag for et en kostnad der man anslår verdiene for 1% og 99% kvartilene, samt mest sannsynlige verdi.

Trinnvis metoden og Lichtenbergs formler ble første gang brukt her i landet for ca 25 år siden. Rundt omkring årtusenskiftet gikk man etter hvert over til å anslå verdien for 10% og 90% kvartilene i stedet for 1% og 99% kvartilene. Reviderte versjoner av formlene dukket opp i forbindelse med denne overgangen. Ulike aktører kom opp med ulike varianter. Felles for dem alle er at de har hatt ingen eller manglende teoretisk forankring. Den mest riktige revisjonen av formlene som har vært i bruk er de som er vist midt i rammen til høyre.

Vi ser at forskjellen mellom formlene for forventningsverdien mellom 1/99 og 10/90 ligger i hva vektningen av mest sannsynlige verdi skal være. På generell form kan vi skrive disse formlene som visst nederst i rammen til høyre. Denne "vektfaktoren" har vi gitt navnet Theta. Og på tilsvarende vis gir vi Zeta som navn for standardavvikets delingsfaktor.

### 5.1.2 Forutsetninger for vurdering av formelverket

Det er noen viktige forutsetninger vi må ha klart for oss over når det vi skal se på hvor godt formelverket for Trinnvis kalkulasjon er.

1. Det finnes en virkelig sannsynlighetsfordeling for kostnadselementet som anslås.

#### Trinnvis formler

##### Opprinnelige for 1/99

$$E = \frac{n + 2,95s + \emptyset}{4,95}$$

$$\sigma = \frac{\emptyset - n}{4,6}$$

##### Revidert for 10/90

$$E = \frac{n + 0,42s + \emptyset}{2,42}$$

$$\sigma = \frac{\emptyset - n}{2,53}$$

##### Generelle formler

$$E = \frac{n + \theta s + \emptyset}{\theta}$$

$$\sigma = \frac{\emptyset - n}{\zeta}$$

##### Parametre

$E$  = forventningsverdi

$\sigma$  = standardavvik

$n$  = nedre anslag

$s$  = mest sannsynlige verdi

$\emptyset$  = øvre anslag

$\theta$  = Theta

Vekting av sannsynlig verdi ved utregning av forventningsverdi

$\zeta$  = Zeta

Delingsfaktor ved utregning av standardavviket

2. Denne virkelige sannsynlighetsfordelingen er gitt ved en gammafordeling, eventuelt forskjøvet og/eller speilet
3. De tallene man legger inn i formlene stemmer med den virkelige sannsynlighetsfordelingen.

Om disse forutsetningene er holdbare i alle sammenhenger er heller tvilsomt. Spesielt når det gjelder det siste punktet. Men disse forutsetningene er nå engang hva som ligger til grunn for formelverket.

Forutsetningen 2, om at den virkelige sannsynlighetsfordelingen er gitt ved en gammafordeling, trenger vi for å ha en fasit å vurdere formelverket opp mot. Det er ikke dermed sagt at virkeligheten faktisk forholder seg til en Gammafordeling.

Hvis man skal gjøre en vurdering av hvor godt formelverket er uten at man legger disse til grunn vil være en oppgave av gigantiske dimensjoner.

I kapittel 6 ser vi nærmere på hva som skjer hvis den siste forutsetningen viser seg å ikke holde.

### 5.1.3 Typiske skjevhetsforhold og relative standardavvik

Vi skal videre i denne rapporten se på hvor stor feilene kan bli som følge av feil i formlene og feil i inngangsdata. Hvor store disse feilene blir avhenger i stor grad av to faktorer, hvor skjeve fordelingene er og hvor store relative standardavvik de har. Resultatene senere i rapporten presenteres i stor grad i form av kurver plottet mot disse to parameterene. Det er derfor på sin plass at vi først ser på hva som er typiske verdier for disse slik at vi har en referanse å forholde oss til.

For å kartlegge disse typiske verdiene hentet vi inn tre ulike usikkerhetsanalyser fra Statens vegvesen og et tilsvarende antall fra hver av tre konsulentfirmaer som driver med ekstern kvalitetssikring. Alle analyser er utført av ulike prosessledere og resurssgrupper. Vi gikk inn i hver analyse og så på skjevhetsforholdet og relativt standardavvik for hvert enkelt trippe-lanslag og laget en statistikk på grunnlag av dette. Resultatet presenteres nedenfor.

#### Skjevhetsforhold

Kurvens grad av skjevhet kan uttrykkes ved bruk av Gammafordelingens forparameter Alfa. Men dette er en ikke-lineær parameter, og derfor ikke særlig enkel å forholde seg til. Vi innfører derfor i denne rapporten begrepet skjevhetsforhold.

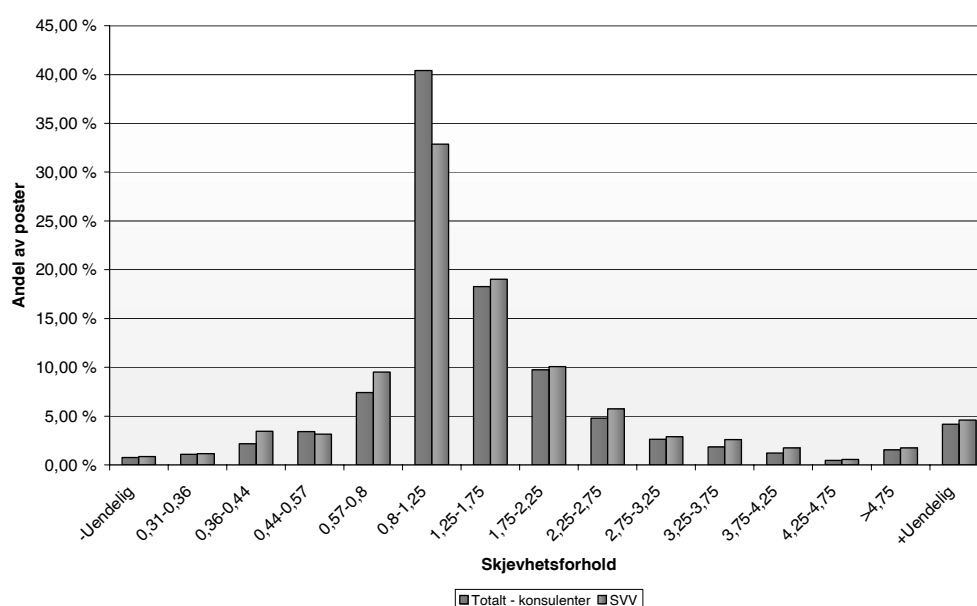
Vi definerer skjevhetsforholdet til å være differansen mellom øvre verdi og mest sannsynlige verdi delt på differansen av nedre og mest sannsynlige verdi:

$$Skf = \frac{\theta - s}{s - n}$$

Merk at slik vi har definert skjevhetsforholdet er det lineært for høyreskjeve fordelinger, mens den for venstreskjeve fordelinger blir ikke-lineært.

Skjevhetsforhold er interessant for oss av to grunner. For det første er formlene vi benytter lineære tilnæringsformler. Som vi skal nærmer på senere i rapporten, er disse formlene tilpasset et gitt skjevhetsforhold, og jo mer vi avviker fra dette jo større blir feilen.

Den andre grunnen til at skjevhetsforholdet er interessant for oss er når vi skal se på feil i inngangparameterene. En høyreskjev kurve har en lang hale til høyre. Jo skjevere kurven er jo lengre er halen (i forhold til resten av kurven). En lang hale betyr at det er større avstand mellom prosentkvantilene ute på halen. Det vil si at å gå fra 80%-kvantilet til 90%-kvantilet har mye mer å si i kroner å øre enn for en symmetrisk kurve. Det betyr at konsekvensen av å anslå feil prosentkvantil blir mye større.



Figur 5-1 Fordeling av skjevhetsforhold for Statens vegvesen og konsulenter

Figur 5-1 viser statistikken som ble utarbeidet for postenes skjevhetsforhold. Tallene som er lavere enn null på x-aksen kan virke litt merkelige. Det som er viktig å huske på her er at disse representerer venstreskjeve kurver og skjevhetsforholdet da ikke er lineært stigende. Grunnen til at ikke mer runde tall ble valgt for kategoriseringen er at et betydelig antall av trippelanslagene i grunnlagsmaterialet hadde runde skjevhetsforhold, og det ville gitt et statistisk feil bilde å brukt runde tall på avgrensningene.

Vi regner med at 70-75% av alle trippelanslag vil ha et skjevhetsforhold mellom 0,57 og 1,75. Med andre ord ikke særlig skjeve. Bare et fåtall vil ha skjevhetsforhold større enn fem. Man skal likevel være oppmerksom på er at dette fåtallet poster som har store skjevhetsforhold som regel er store poster i en analyse. Det er typisk der hvor man har nedre og mest sannsynlig verdi tilnærmet lik at man ser store skjevhetsforhold. Dette forekommer først og fremst på postene for indre og ytte påvirkningsfaktorer og andre prosentpåslagsposter som totalt sett utgjør en stor andel av forventningsverdien for et kostnadsanslag.

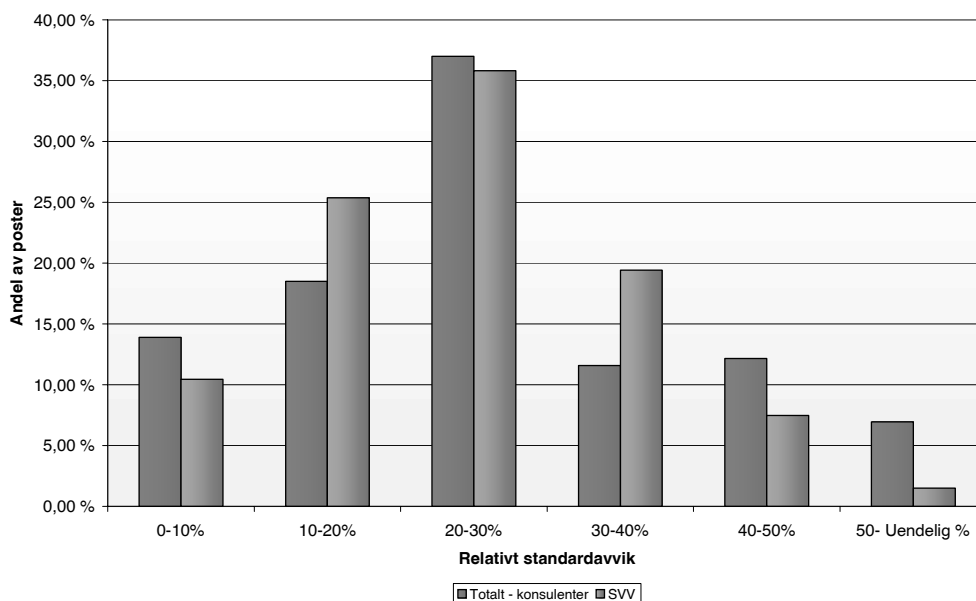
## Relativt standardavvik

Relativt standardavvik er et mye brukt begrep. Det er definert som standardavviket delt på forventningsverdien. Eller sagt med andre ord, hvor stor spredningen er i forhold til forventningsverdien.

$$\sigma_{\text{rel}} = \frac{\sigma}{E(x)}$$

Det relative standardavviket har mye si for hvor stor feilene blir. Det kommer av at vi benytter en *forskjøvet* Gammafordeling. Det vil si at benytter en fast tallverdi pluss en Gammafordeling. For standardavviket er denne tallverdien uinteressant. Standardavviket sier kun noe om spredningen på verdiene, ikke noe om hvor store de er. Men for forventningsverdien er det absolutt interessant. Når man forskyver fordelingen langs x-aksen blir alle verdier hentet ut fra kurven tilsvarende større. Hvis vi for forskyver kurven med 10, så øker verdien for 10%-kvantilet med 10, og likeså for alle andre kvantiler samt mest sannsynlige verdi og forventningsverdien. Trinnsformelen for forventningsverdien anslår verdien direkte for en forskjøvet gammafordeling. Men vi kunne like gjerne ha delt opp beregningen i to trinn. Først regnet ut forventningsverdien av en ufor-skjøvet Gammafordeling og deretter lagt forskyningsstillegget. Resultatet hadde blitt det samme. Men det er klart at i dette tilfellet må eventuelle feil i formlene slå til på i utregningen av forventningsverdien for Gammafordelingen, det andre er jo bare å legge til en gitt tallverdi. Det er også klart jo større denne "sikre" tallverdien er, jo mindre blir en eventuell beregningsfeil i forhold til den totale forventningsverdien (med forskyvning). Samtidig vet vi at dette sikre tallet ikke har noe å si for standardavviket, så om vi øker denne tallverdien øker vi altså forventningsverdien men ikke standardavviket. Dette får da naturlig nok den følgen at det relative standardavviket synker. Poenget her er da at konsekvensen av gjøre feil med lave relative standardavvik er mye lavere enn ved høye som følge av at det ligger et "sikkert" tall i bunn.

Figur 5-2 viser hvor stort det relative standardavviket vil være statistisk sett for en post i et kostnadsoverslag utført av Statens vegvesen. Man kan ha i bakhodet når man leser denne rapporten at det typiske relative standardavviket ligger på ca. 25%.



Figur 5-2 Typiske relative standardavvik for kostnadsanslag i Statens vegvesen

### 5.1.4 Theta optimal

Theta er, som beskrevet i forrige avsnitt, en vektfaktor som angir hvor mye  $s$  (mest sannsynlige verdi) skal vektlegges ved beregningen av forventningsverdien for en sannsynlighetsfordeling. Theta optimal er den verdien av Theta som ved bruk av suksessivformlene vil gi en forventningsverdi lik den matematiske korrekte verdien for sannsynlighetsfordelingen. Hva Theta optimal vil være for en gitt sannsynlighetsfordelingen vil avhenge av både hvilke kvantiler det er som benyttes når man anslår nedre og øvre verdier i trippelanslaget og hvilken alfa sannsynlighetsfordelingen har. På bakgrunn av den generelle formelen for forventningsverdien visst i forrige avsnitt kan vi ved hjelp av kjente formler for gammafordelingen sette opp et generelt uttrykk hva Theta optimal vil være.

#### Utledning av generell formel

Forventningsverdien av en gammafordeling kan finnes ved hjelp av en lineær tilnæringsformel på formen:

$$1) \quad E(x) = \frac{n + \theta s + \emptyset}{2 + \theta}$$

Der  $E(x)$  er forventningsverdien,  $n$  er nedre anslag,  $s$  er mest sannsynlige verdi og  $\emptyset$  er øvre anslag.

Noen formler som gjelder gammafordelingen er:

Mode eller mest sannsynlige verdi

$$2) \quad s = \beta(\alpha - 1)$$

Forventningsverdi:

$$3) \quad E(x) = \alpha\beta$$

Ved å skrive om 2) får vi et uttrykk for Beta 4):

$$4) \quad \beta = \frac{s}{\alpha - 1}$$

Vi setter så 4) inn i 3) og får et uttrykk for forventningsverdien uttrykt ved  $s$  og Alfa 5):

$$5) \quad E(x) = \alpha\beta = \frac{\alpha s}{\alpha - 1}$$

Vi setter så det nye uttrykket 5) for forventningsverdien likt vår antatte formel 1) og får ut et uttrykk for Theta gitt ved  $n$ ,  $s$ ,  $\emptyset$  og Alfa.

$$6) \quad \frac{n + \theta s + \emptyset}{2 + \theta} = \frac{\alpha s}{\alpha - 1}$$

$$7) \quad \theta = \frac{\alpha n - n + \alpha \emptyset - \emptyset - 2\alpha s}{s}$$

Dette uttrykket 7) for Theta er generelt (foren uforskjøvet gammafordeling), i neste avsnitt går vi videre og bruker dette uttrykket til å finne hva Theta optimal vil være for ulike alfa ved bruk av 10/90 kvantilene.

## Theta optimal for 10/90

Med utgangspunkt i den generelle formelen for Theta optimal fra forrige avsnitt kan vi regne ut hva Theta optimal vil være ved bruk av spesifikke prosentkvantiler. I dette avsnittet ser vi på hva Theta optimal vil være ved bruk av 10/90 kvantilene som basis for øvre og nedre verdier i et trippelanslag.

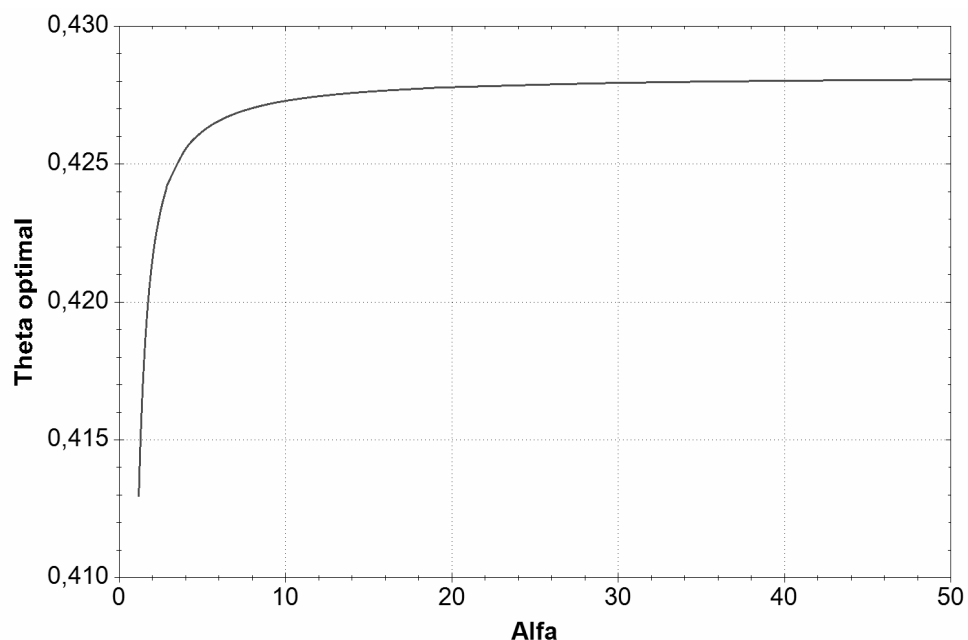
Den generelle formelen for Theta optimal er:

$$7) \quad \theta = \frac{\alpha n - n + \alpha \phi - \phi - 2\alpha s}{s}$$

Vi ser at i formelen for Theta optimal er alfa en parameter. Som beskrevet tidligere beskriver Alfa kurvens form. Det dette betyr er at Theta vil være avhengig av hvor skjev kurven er. Det betyr at det ikke er nok å regne ut en verdi for Theta optimal, vi er nødt til å regne ut en serie verdier for å se hvor mye Theta optimal varierer med skjevheten.

En gammafordeling bestemmes entydig ved hjelp av en tre parametere. Dette kan være Alfa, Beta og Forskyvning som beskrevet i avsnitt 3.5, men det er også gjøres ved å fastslå tre punkter på kurven som for eksempel n, s og  $\phi$  eller kombinasjon av disse to måtene. For å generere de tallsett med verdiene vi trenger (Alfa, n, s og  $\phi$ ) er det enklest å benytte Alfa, Beta og Forskyvning og regne ut n, s, og  $\phi$  fra disse.

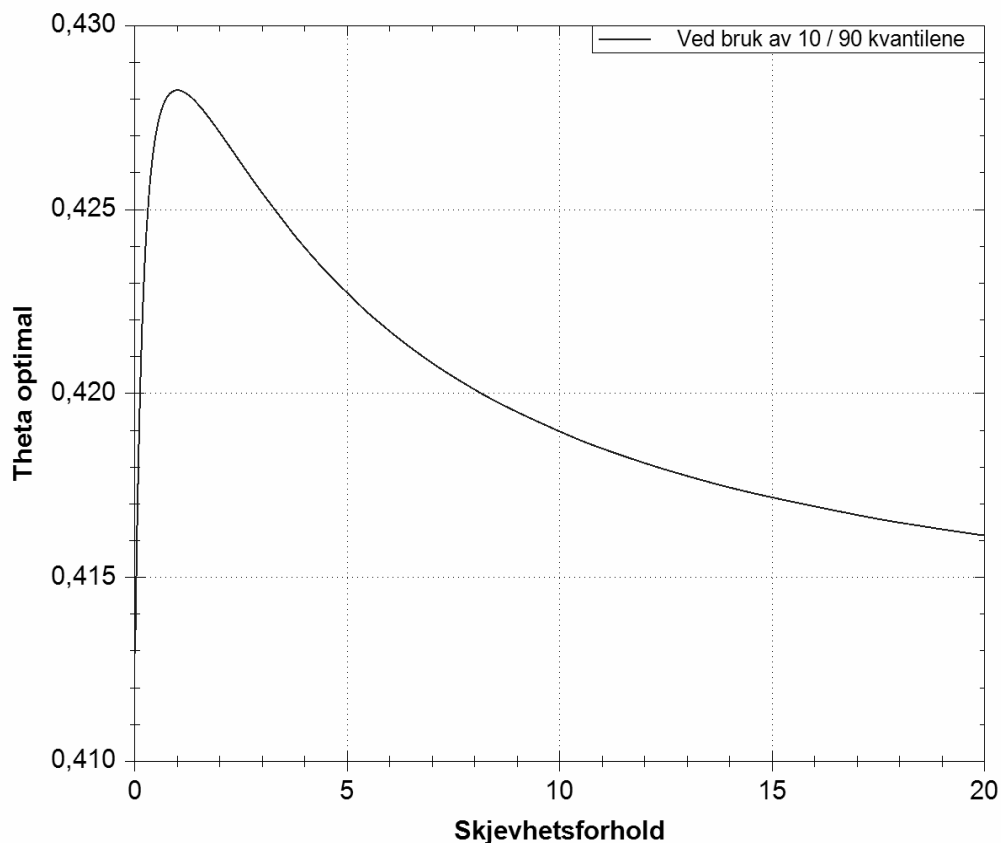
Vi gjør dette ved at Beta settes lik 10, Forskyvning settes lik 0 og Alfa varieres. Verdi for s regnes ut ved hjelp av tidligere beskrevet formel 2). n og  $\phi$  regnes ut ved å ta den inverse funksjonen for gammafordelingen av Alfa, Beta og den aktuelle prosentkvantilen (10 og 90). (for eksempel funksjon GAMMAINV i Excel.) Resultatet blir da som vist i figur 5-3.



Figur 5-3 *Theta optimal for 10/90 uttrykt ved Alfa*

Vi ser av figuren at det er svært liten variasjon av hva som er optimal Theta verdi for ulike verdier av Alfa. For Alfa tilnærmet lik 1 vil optimal Theta være 0,408. Mens det for Alfa lik 50 vil optimal Theta være 0,428. Altså en forskjell på 0,02.

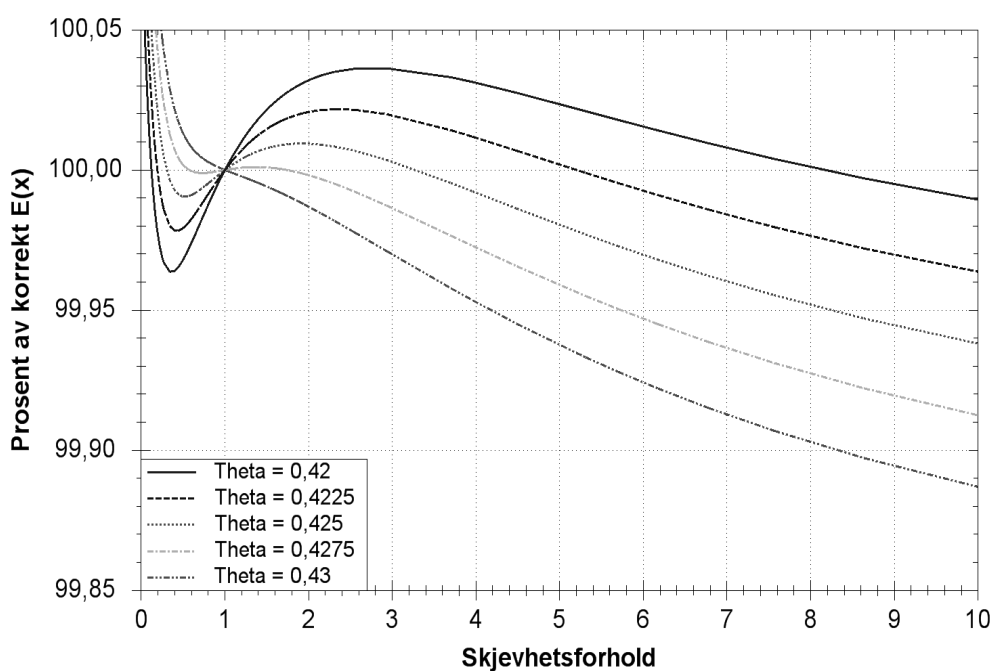
De færreste har noe særlig forhold til Alfa, spesielt med tanke på at kurvens form endres ikke-lineært med Alfa. Det er nok enklere for de fleste å forholde seg til skjevhetsforholdet. Vi definerer skjevhetsforholdet til å være differansen mellom øvre verdi og mest sannsynlige verdi delt på differansen av nedre og mest sannsynlige verdi  $((\theta-s)/(n-s))$ . Figur 5-4 viser hva optimal Theta vil være for et gitt skjevhetsforhold.



Figur 5-4 *Theta optimal for 10 / 90 gitt ved skjevhetsforhold*

Figur 5-4 viser hva Theta optimal vil være for skjevhetsforhold både større og mindre enn en. Altså både høyre og venstreskjeve fordelinger. Merk at verdien for Theta optimal blir akkurat det samme for en venstreskjev fordeling, men som følge av at den blir plottet mot en ulik skala ser det forskjellig ut på kurven. Skjevhetsforholdet for en venstreskjev kurve er likt  $1/\text{skjevhetsforholdet}$  for en høyreskjev kurve.

Det vi kan konkludere ut fra dette er at ved bruk av trinnvis formlene for 10/90 må Theta ligge et sted mellom 0,42 og 0,43 for de skjevhetsforholdene som vi kan regne med at det er realistisk at vi kommer over i praksis. Figur 5-5 viser hva ulike Theta verdier gir av feil på forventningsverdien for ulike skjevhetsforhold. Vi ser at en Theta verdi på rundt 0,4225 gir minst feil ved lave skjevhetsforhold. En Theta verdi på 0,425 er antageligvis bedre egnet i praksis der man må regne med at inngangsdataene har en viss skjevhet. I alle fall er her kun snakk om et par tidels promille i forskjell på forventningsverdien.



Figur 5-5 Feil på forventningsverdi for Theta mellom 0,42 og 0,43

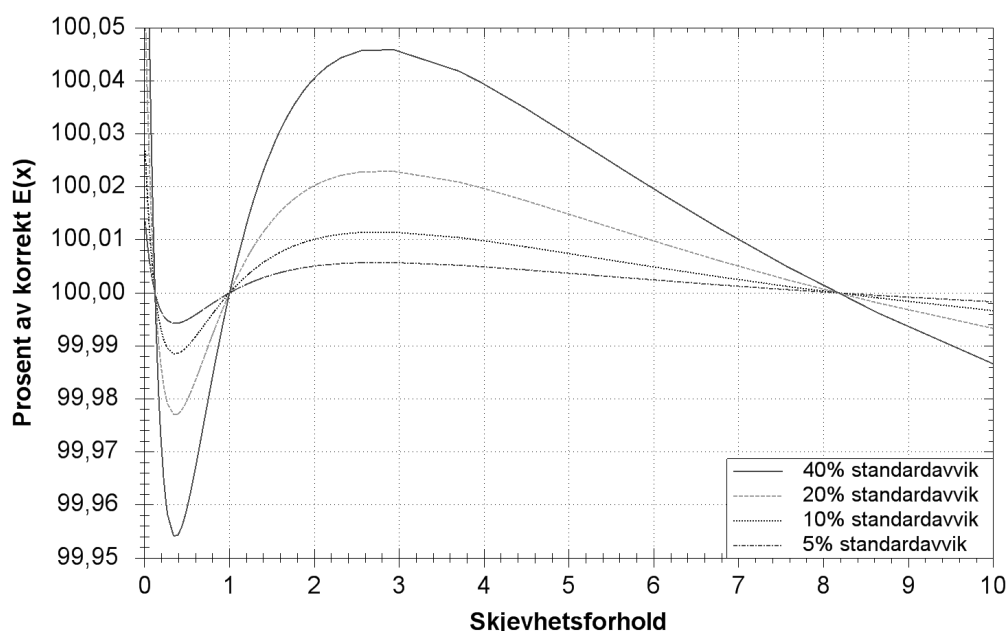
I stedet for å bruke statiske verdier for Theta vil en bedre metode i alle fall være å bruke dynamiske verdier for Theta, noe som er beskrevet senere i avsnitt 4.3.



## Feil på forventningsverdi ved bruk av $\Theta = 0,42$

Statens Vegvesen benytter i sitt beregningsprogram, Anslag, en verdi for  $\Theta$  på 0,42. Figur 5-6 viser hvor stort avviket på forventningsverdien blir ved bruke av denne  $\Theta$ -verdien for ulike skjevhetsforholdstall. Som vi ser av figuren er maks avvik mindre enn en halv promille for trippelanslag med standardavvik på 40%, det vil si for tilfeller der standardavvikets størrelse er 40% av forventningsverdiens størrelse. Relativt standardavvik høyere enn dette er uvanlig at man finner på vanlige kalkyleposter. På generelle forhold (indre og ytre påvirkningsfaktorer) derimot kan man finne relative standardavvik som er betraktelig høyere som igjen gir større feil på forventningsverdien. I verste fall kan man ved bruk av  $\Theta = 0,42$  få en feil på tre promille.

Feilen på forventningsverdien kunne blitt mindre ved å finjustere  $\Theta$  som visst i forrige avsnitt, men feilen her er uansett å betrakte som neglisjerbart med tanke på den usikkerheten som ligger i inngangsdataene. (Se kapittel 6).



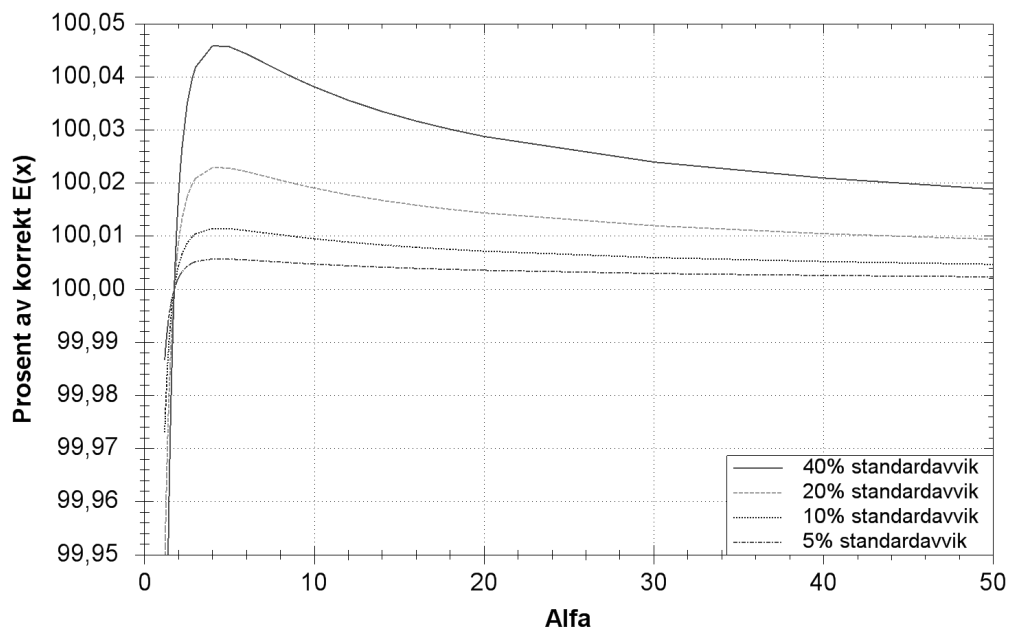
Figur 5-6 Feil på forventningsverdi ved bruk av  $\Theta = 0,42$

Figur 5-6 har tegnet inn flere kurver for ulike relative standardavvik og vi ser at jo større det relative standardavviket er jo større blir feilen. Det relative standardavviket avhenger i stor grad av fordelings forskyvningsparameter. Det relative standardavviket er standardavviket, som er uavhengig av forskyvingen, delt på forventningsverdien som er avhengig av forskyvningen, noe som gjør at jo større forskyvningsparameteren er jo mindre blir det relative standardavviket.

Vi kan se på dette som om at et kostnadsanslag har en deterministisk del og en stokastisk del. Der forskyvningen utgjør den deterministiske delen og en uforstykket gammafordeling gitt ved Alfa og Beta utgjør den stokastiske. Det er da naturlig at jo større den deterministiske delen er, jo mindre blir den totale feilen som følge av at den stokastiske delen, og eventuelle feil der, får mindre og si for helheten.

Grunnen til at kurvene i figur 5-6 får den formen den får her er at  $\Theta = 0,42$  er optimalverdien for en svært skjev fordeling med et skjevhetsforhold på 8 (se figur 5-5). Dette gjør at forventningsverdien blir mest korrekt jo nærmere man er dette skjevhetsforholdet. Vi ser også at feilen avtar når skjevhetsforholdet går ned mot 1. Dette skyldes at fordelingen da etter hvert blir så symmetrisk at mest sannsynlig verdi ikke får noe å si. Et skjevhetsforhold på 1 betyr at vi har en normalfordeling og forventningsverdien vil være lik snittet av nedre og øvre anslag.

Figur 5-7 viser det samme som figur 5-6, men med feilen på forventningsverdien plottet mot Alfa fremfor skjevhetsforholdet. Grunnen til at utslaget på y-aksen er større på denne figuren enn på den forrige er at figur 5-7 er plottet fra og med  $\text{Alfa} = 1,18$ . Dette tilsvarer et skjevhetsforhold på 158,65, Noe som er ca 16 ganger større enn hva som er vist i figur 5-6. Som nevnt tidligere i kapitlet er det i praksis sjelden man har et skjevhetsforhold større enn 5, men kurven viser at rinnvisformlene ved bruk av  $\Theta = 0,42$  er svært robuste, selv ved ekstremt skjeve trippelanslag.

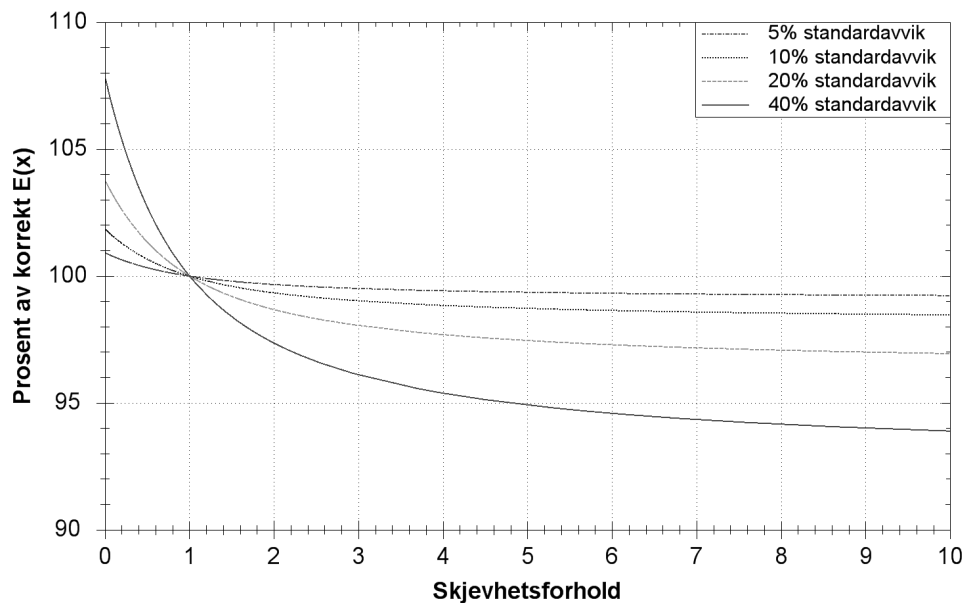


Figur 5-7 Feil på forventningsverdi ved bruk av  $\Theta = 0,42$ , plottet mot Alfa

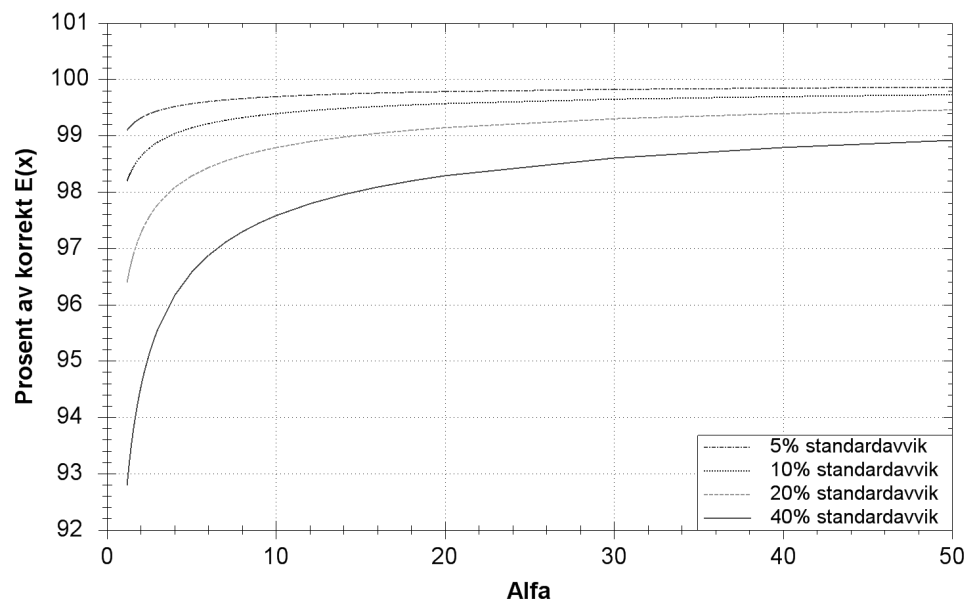
## Feil på forventningsverdi ved bruk av for stor Theta

Bruken av  $\Theta = 1$  i Trinnvis-formlene er ikke spesielt utbredt. Men vi har sett nærmere på den som et eksempel på hva det medfører å bruke en  $\Theta$ -verdi som er større enn hva vi tidligere har bestemt  $\Theta$  optimal til å være.

Figur 5-8 viser hva feilen blir ved bruk av  $\Theta = 1$ . Som vi ser av kurvene blir feilene betydelige. Feilene er her opp mot seks prosent i de mest ekstreme tilfellene der det for  $\Theta = 0,42$  maks var en halv promille. Hvis vi beveger oss lengre bort fra  $\Theta$  optimal vil feilen bli større. For eksempel ville en kurve plottet for  $\Theta = 2,95$  gitt oss opp mot 14% feil. (2,95 er  $\Theta$  som brukes når input er 1% kvantilet - mest sannsynlig verdi - 99% kvantilet).



Figur 5-8 Feil på forventningsverdi ved bruk av  $\Theta = 1$

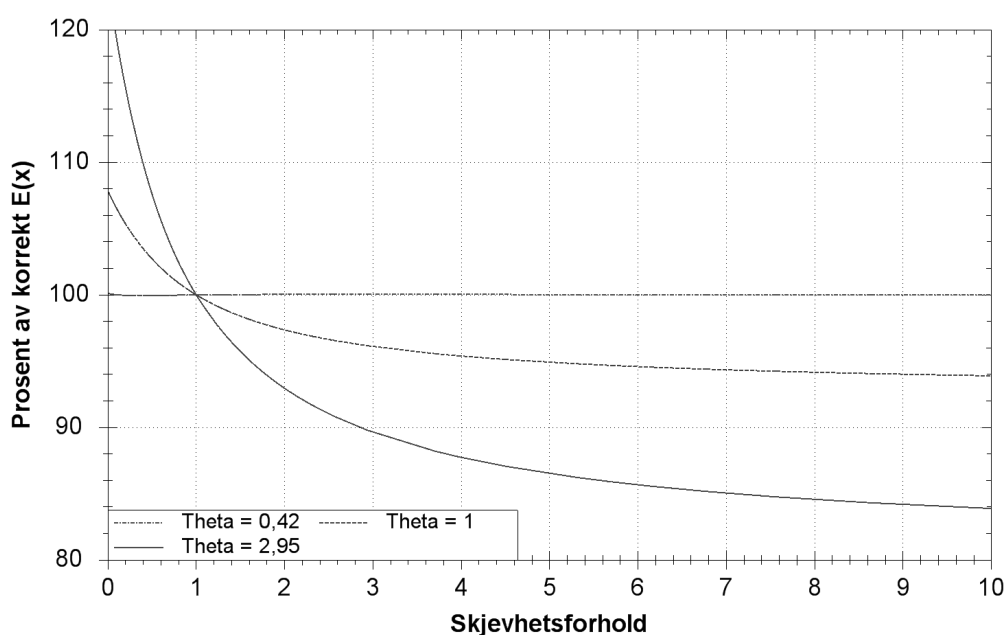


Figur 5-9 Feil på forventningsverdi ved bruk av  $\Theta = 1$ , plottet mot Alfa

### Gap mellom $\Theta = 1$ , $\Theta = 0,42$ og $\Theta = 2,95$

Figur 14 viser forskjellen i feil for  $\Theta = 1$ ,  $\Theta = 0,42$  og  $\Theta = 2,95$  plottet mot hverandre. Vi ser at når disse kurvene plottes inn på samme skala blir feilkurven for  $\Theta = 0,42$  tilnærmet rettlinjet i forhold til feilkurvene for  $\Theta = 1$  og  $\Theta = 2,95$ .

Denne figuren forklarer delvis et fenomen man har observert i Statens Vegvesen. Ved overgang fra Anslag versjon 3.0.3 (som hadde feil formel lagt inn med  $\Theta = 2,95$ ) til versjon 3.0.5 har man sett at forventningsverdiene på kostnadsoverslagene har gått opp rundt 2%, men i enkelte tilfeller mye mer. For eksempel økte kostnadsoverslaget for "Rv 659 Nordøyvegen" med hele 8,6% kun som følge av at man gikk fra den ene til den andre versjonen av Anslag. Det viste seg her at en god del av trippelanslagene for dette prosjektet var svært høyreskjeve. Og som vi ser av figuren blir avviket større jo større skjevheten er.



Figur 5-10  $\Theta = 1$  og  $\Theta = 0,42$  plottet mot hverandre

Men dette var ikke hele forklaringen på hvorfor feilen ble såpass stor som 8,6%. Selv om enkelte av trippelanslagene var svært høyreskjeve, var det også mange som var normalfordelte eller bare litt skjeve. Den gjennomsnittlige skjevheten på kalkylepostene var ikke alene nok til å gi et såpass stort utslag. Hvorfor ble da feilen såpass stor? Svaret ligger i multiplikasjonsposter. Hvis vi har en post som har en enhetspris og en mengde som begge har en feil i forventningsverdien på 2% vil posten ha en feil på 4,04%.

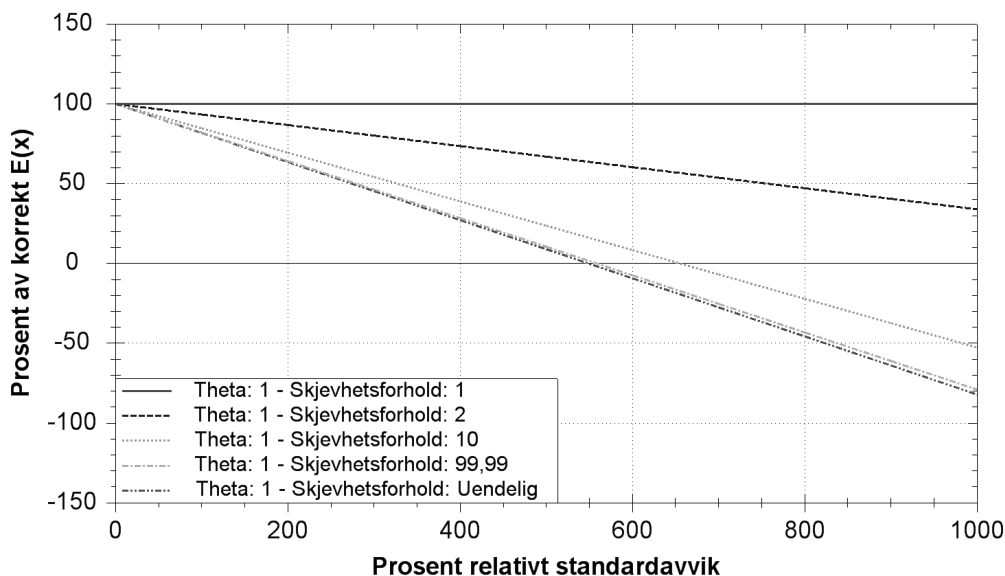
Særlig tydelig blir effekten av multiplisering av feil på indre og ytre påvirkningsfaktorer, spesielt fordi feilen der i utgangspunktet som regel er mye større.

## Feil på indre og ytre påvirkningsfaktorer ved bruk av uriktig Theta

De feilkurvene vi har sett på hittil har vært for standardavvik opp til 40%. Som tidligere nevnt er det uvanlig at vi finner relative standardavvik større enn det for vanlige kalkylepos-ter. For indre og ytre påvirkningsfaktorer er det derimot meget vanlig å ha relative standardavvik som ligger langt over dette. Trippelanslagene for indre og ytre påvirkningsfaktorer angis vanligvis som påslagsfaktorer omkring 1,0, men beregningsmessig ligger de sentrert rundt 0. Noe som betyr at vi får en forventningsverdi i nærheten av 0 og en relativt sett stor spredning omkring denne.

Figur 5-11 viser hva feilen på forventningsverdien blir for ulike skjevhetsforhold ved bruk av  $\Theta = 1$  for ekstremt høye relative standardavvik. Vi ser at ved et relativt standardavvik på 0 er feilen på forventningsverdien lik 0 og at den deretter øker lineært med at det relative standardavviket øker. En situasjon der man har et relativt standardavvik på null tilsvarer at man har et fullstendig deterministisk anslag, ergo er det irrelevant hvor godt formlene regner ut forventningsverdien av den stokastiske delen.

Vi ser også av figur 5-11 at feilen på forventningsverdien øker ikke-lineært med skjevhetsforholdet. For eksempel er det en prosentvis mindre økning av feilen på forventningsverdien ved å gå fra et skjevhetsforhold på 2 til et på 10, enn det er å gå fra et skjevhetsforhold på 10 til et på 99,99. Dette stemmer overens de foregående figurene som er “omvendt” plottet. Det vil si de har en kurve for hvert relative standardavvik og skjevhetsforholdet langs x-aksen.

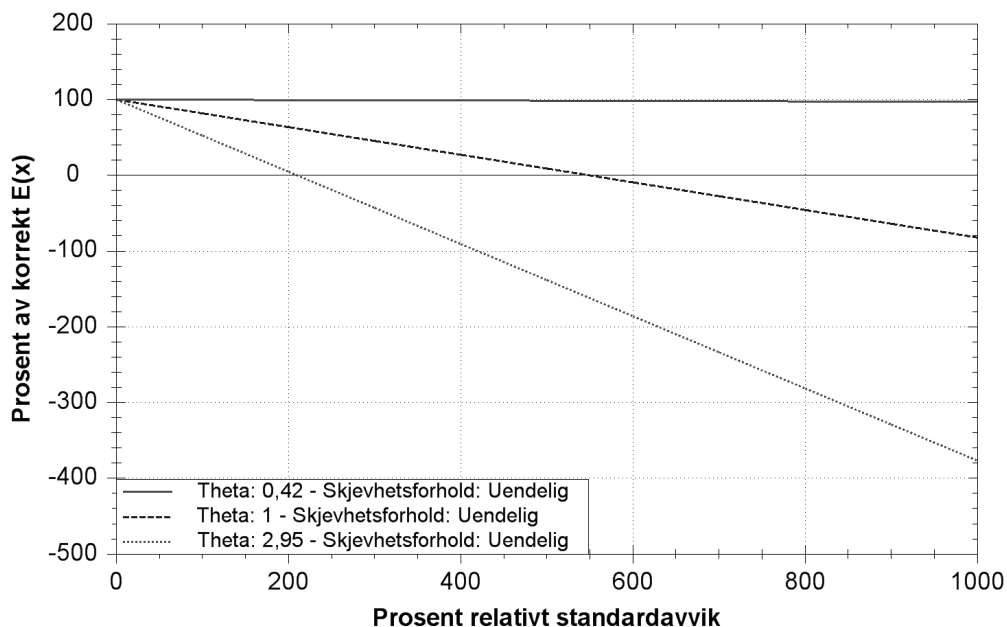


Figur 5-11 *Feil på forventningsverdi for  $\Theta = 1$  for ulike skjevhetsforhold ved ekstreme relative standardavvik*

Vi ser av figur 5-11 at det i ved svært høye skjevhetsforhold og relative standardavvik faktisk er mulig at man får en forventningsverdi som er såpass mye lavere enn det den skulle ha vært at den ender opp med motsatt fortegn. Det skal riktignok et veldig sært trippelanslag til for å komme i denne situasjonen. For eksempel vil trippelanslaget -10, -10, 20 gi et uendelig stort skjevhetsforhold og et relativt standardavvik på bortimot 500%.

Den største feilen man i praksis kan regne med å møte er når mest sannsynlig verdi er lik 0. Trinnvisformlene for  $\Theta = 1$  og  $\Theta = 2,95$  har en den interessante egenskapen at de for tilfeller der mest sannsynlig verdi er lik 0 bestandig gir en forventningsverdi som kun er henholdsvis 80,7% og 48,9% av det den korrekte verdien for høyreskjeve trippelanslag.

Figur 5-12 viser de mest ekstreme kurvene for alle de tre ulike verdiene av  $\Theta$  som er omtalt her. Vi ser at ved 1000% relativt standardavvik ligger forventningsverdien på ca -380% av korrekt for  $\Theta = 2,95$ , ca -80% for  $\Theta = 1$  og ca 97% for  $\Theta = 0,42$ . Det vi kan si ut i fra dette er at selv om denne situasjonen som sagt er svært lite sannsynlig i praksis viser den at  $\theta = 0,42$  gir en formel som er svært robust selv ved ekstreme trippelanslag.



Figur 5-12 *Feil på forventningsverdi for ulike Theta verdier for ulike skjevhetsforhold ved ekstreme relative standardavvik*

### 5.1.5 Zeta optimal

Zeta optimal er den verdien som må settes inn under brøkstreken i Trinnvisformlene for standardavviket for at den beregnede verdien for standardavviket skal bli matematisk korrekt. Tilsvarende som for Theta kan vi også finne et uttrykk for Zeta optimal ved hjelp av kjente formler for gammafordelingen.

#### Utredning av generell formel

Varians av en gammafordeling:

$$8) \quad \sigma^2 = \alpha\beta^2$$

Standardavvik av en gammafordeling

$$9) \quad \sigma = \sqrt{\alpha\beta^2}$$

Generalisert trinnvisformel for standardavvik

$$10) \quad \sigma = \frac{\theta - n}{\zeta}$$

Hvis vi setter sammen 9 og 10 og snur litt på uttrykket får vi:

$$11) \quad \zeta = \frac{\theta - n}{\sqrt{\alpha\beta^2}}$$

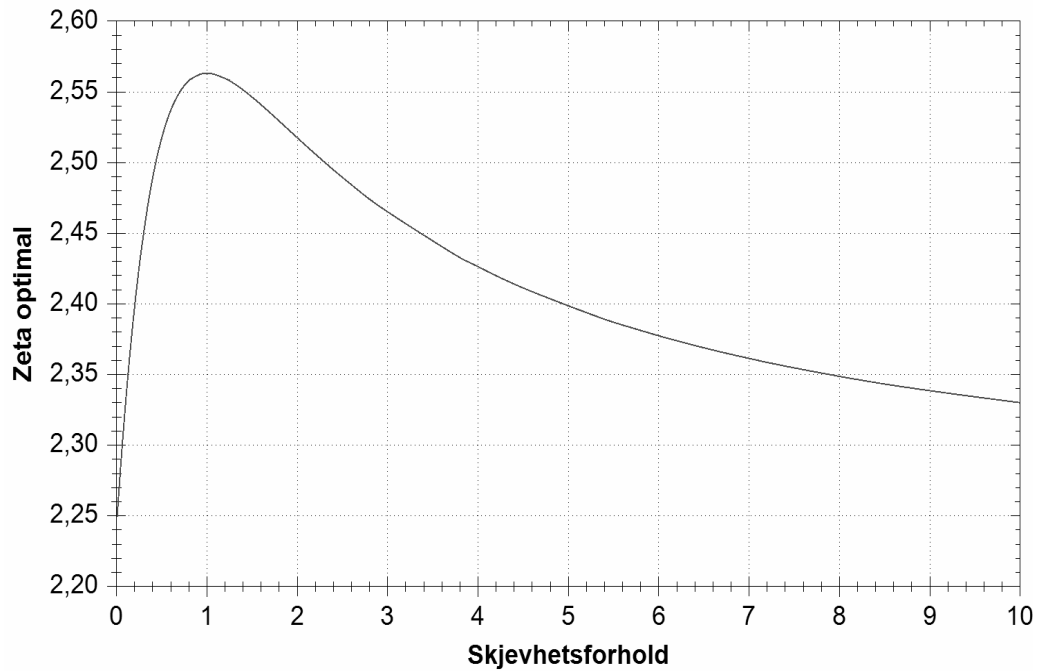
Dette er det generelle uttrykket for Zeta optimal. I neste avsnitt ser vi på hva Zeta optimal bør være ved bruk av 10/90 kvantilene.

### Zeta optimal for 10/90

Det generelle uttrykket for Zeta optimal er:

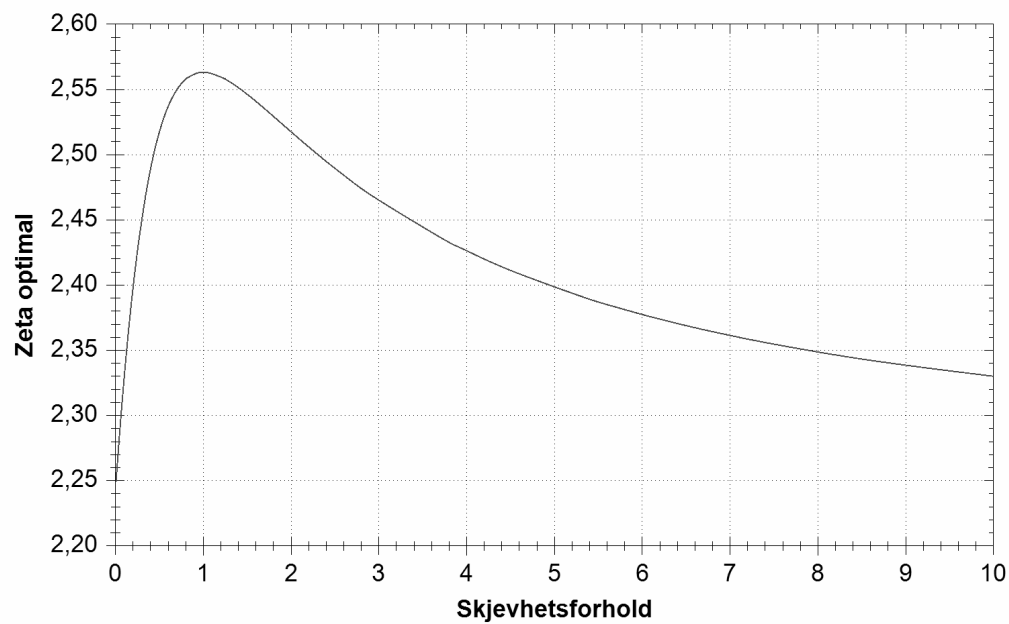
$$11) \quad \zeta = \frac{\theta - n}{\sqrt{\alpha\beta^2}}$$

Vi tar utgangspunkt i tallsettene generert tidligere for beregningen av Theta optimal og regner ut verdier for Zeta optimal for 10/90 ved hjelp av disse. Dette gir oss resultatet vist i Figur 17 og Figur 18.



Figur 5-13 *Zeta optimal for 10 / 90 for ulike verdier av Alfa*





Figur 5-14 *Zeta optimal for 10 / 90 for ulike skjevhetsforhold*

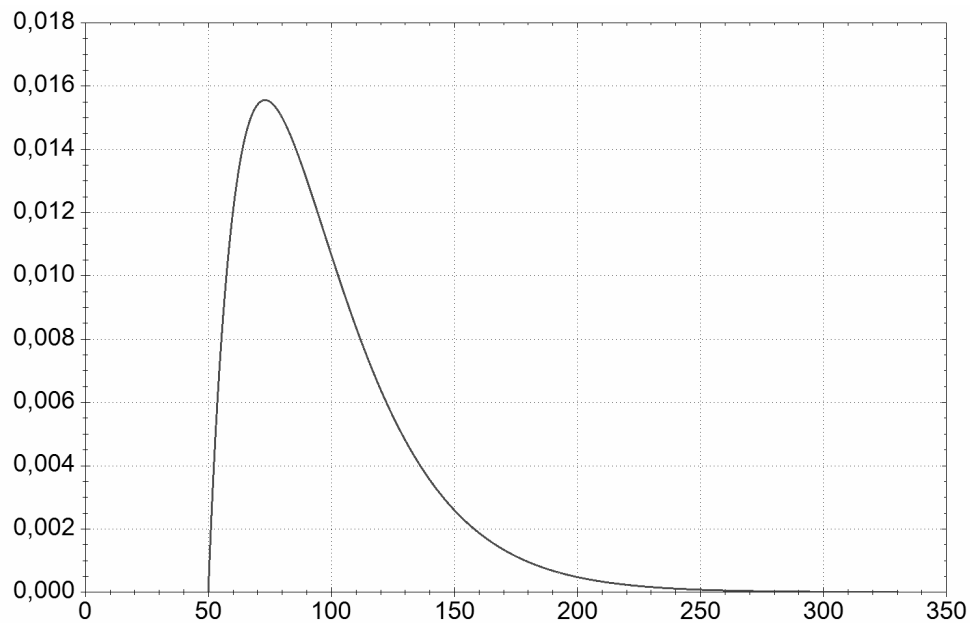
Vi ser her at spennet i Zeta optimal vist i kurvene på forrige side er mye større enn for Theta optimal vist i kurvene i avsnitt 4.2.2. Grunnen til dette er at mest sannsynlig verdi,  $s$ , ikke inngår i trinnvisformelen for standardavviket.

$$10) \quad \sigma = \frac{\theta - \mu}{\zeta}$$

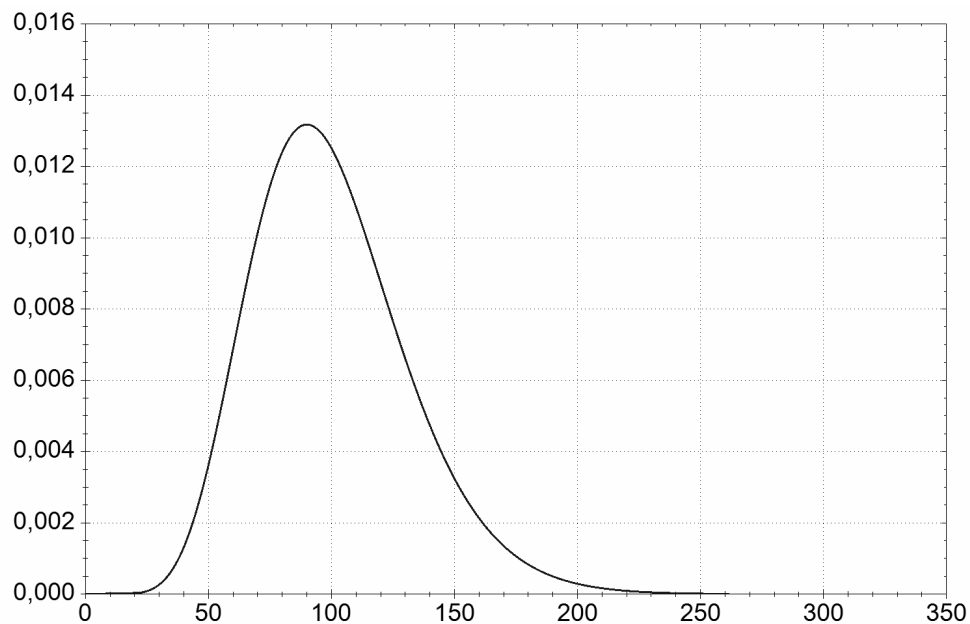
Dette betyr en gitt Zeta verdi tilsier et gitt skjevhetsforhold. Zeta = 2,53 er den som er vanlige å bruke ved anslag av 10/90 kvantilene. Denne verdien er korrekt for et skjevhetsforhold på 1,87 (Alfa = 10), så vi kan si at formelen er finstilt for dette skjevhetsforholdet. Dette betyr i praksis at man forskyver mest sannsynlige verdi til 41,26%-kvantilet, som er hvor mest sannsynlig ligger for gammafordelinger med Alfa = 10.

### Følgen av at Zeta ikke tas hensyn til i trinnvisformlen for standardavvik

La oss anta at vi skal beregne standardavviket for fordelingen vist i figur 5-15. Denne fordelingen har et skjevhetsforhold på 6,17 (Alfa=2). Ved bruk av formel 11 med Zeta = 2,53 vil fordelingen som vi faktisk beregner standardavviket for være den som er vist i figur 5-16.



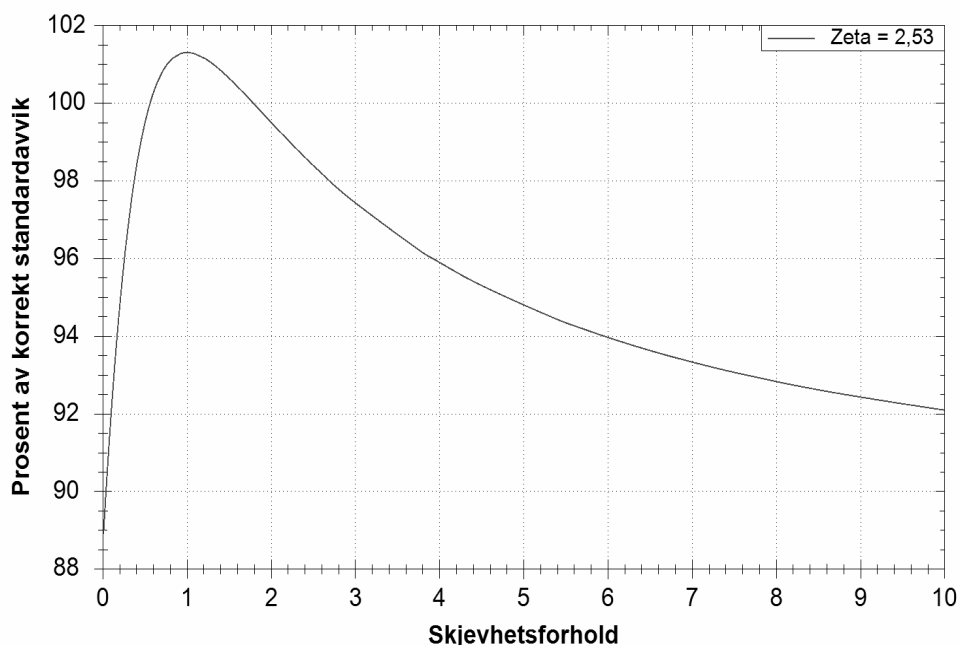
Figur 5-15 *Angitt kurve*



Figur 5-16 *Kurve som faktisk beregnes*

I tilfellet over vil det beregnede standardavviket være 6% for lavt. Figur 5-17 viser hva feilen blir for ulike skjevhetsforhold ved bruk av  $\zeta = 2,53$ . Hvis vi begrenser oss til de skjevhetsforholdene vi vanligvis kan forvente å finne i en usikkerhetskalkyle, mellom 0,5 og 4, ser vi at feilen er på opp mot 4%. En annen ting vi kan se er at symmetriske fordelte trippelanslag gir en feil på bortimot 1,5% på standardavviket.

Vi ser at kurven at fordelinger som er mer symmetriske enn “idealet” vil får for høye standardavvik, mens de som er mer skjeve får for lave beregnede standardavvik.



Figur 5-17 Feil på standardavviket ved bruk av  $\zeta = 2,53$

Det vi kan konkludere med er at trinnsformelen for standardavviket er mye mer unøyaktig enn formelen for forventningsverdi, og i motsetning til Theta er det ikke mulig å finne noen god allmenngyldig verdi for Zeta, skal man bruke en statisk verdi for Zeta må man ta et valg og finjustere formelen for et gitt skjevhetsforhold. Det vil dog være en mye bedre løsning å benytte dynamiske verdier for Zeta som presentert i neste kapittel.

### 5.1.6 Dynamisk Theta og Zeta

Trinnsformlene benytter seg av statiske verdier for Theta og Zeta. Som vi har sett er variasjonene i hva som er optimale verdier for disse for ulike alfa svært små når det gjelder Theta og ikke full så små når det gjelder Zeta. Men formlene gir i alle felle en rimelig god tilnærming til korrekte verdier med tanke på den usikkerheten som ligger i inngangsdataene. Hvis man ønsker en større grad av nøyaktighet vil en mulighet være å bruke dynamisk Theta og Zeta. Det vil si at man ved utregning av et trippelanslag bruker de optimale verdiene for Theta og Zeta for det aktuelle skjevhetsforholdet. Dette krever at det er mulig å beregne disse på en grei måte. En grei måte vil si at det enkelt og greit kan implementeres i dataprogrammer for beregning av usikkerhetsanalyser. Å skulle bruke en slik metode i forbindelse med håndregning vil være dødfødt.

Hvis vi legger en trendlinje på den høyreskjeve delen av kurvene vist i figur 5-3 "Theta optimal for 10/90 uttrykt ved Alfa" på side 53 og i figur 5-13 "Zeta optimal for 10 / 90 for ulike verdier av Alfa" på side 63 vil vi se at de er tilnærmet parabelformet. Vi antar da at vi kan uttrykke optimal Theta og Zeta ved hjelp av polynomuttrykk av skjevhetsforholdet.

Hvis  $\phi$  er skjevhetsforholdet

$$\phi = \frac{\phi - s}{s - n} \quad \phi - s > s - n$$

$$\phi = \frac{s - n}{\phi - s} \quad \phi - s < s - n$$

$$\phi = 1 \quad \phi - s = s - n$$

så vil et generelt uttrykk for Theta og Zeta være:

$$a\phi^3 + b\phi^2 + c\phi + d$$

der  $a$ ,  $b$  og  $c$  er konstanter. Disse bestemmes ved å gjennomføre regresjon på datasettet til kurvene i figur 5-3 og figur 5-13. Ved å gjøre dette får vi følgende formler for dynamisk Theta og Zeta.

$$\theta = 0,000003975312628\phi^3 + (0,000009459489819\phi^2) - 0,001602543687\phi + 0,4301263456$$

$$\zeta = -0,0001693580287\phi^3 + 0,005991846477\phi^2 - 0,07392170812\phi + 2,640746721$$

Disse formlene vil ikke gi korrekte verdier for svært høye skjevhetsforhold (over ca 20). I og med at dette sjelden forekommer, kan man i disse tilfellene gå over på en gitt forhåndsdefinert verdi. Det vil riktignok si at ved høye skjevhetsforhold vil få noen unøyaktigheter. Men for det første er så høye skjevhetsforhold svært sjeldne og for det andre vil det fortsatt gi et mer korrekt svar enn dagens formler gir i og med at man kan velge en Theta-verdi optimalisert for et smalere område.

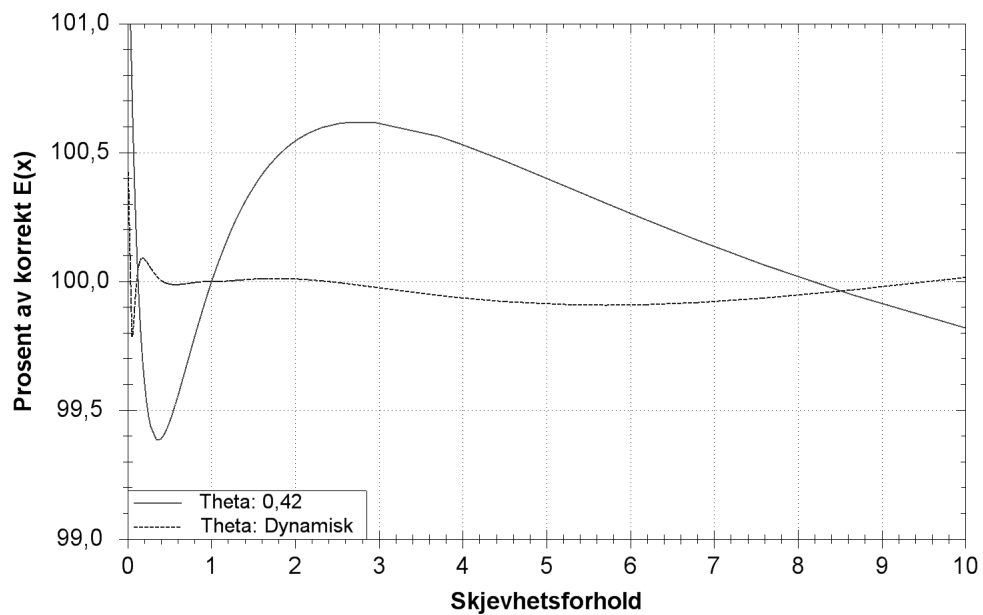
Ved bruk av disse formlene finnes to spesialtilfelle som må håndteres. Det ene er når nedre eller øvre anslag er lik mest sannsynlig anslag og det andre er når alle tre anslagene er like. Begge disse to tilfellene vil føre til feil i et dataprogram i og med at de forårsaker en divisjon med null hvis man prøver å regne ut skjevhetsforholdet for fordelingen.

Tilfellet der nedre eller øvre anslag er lik mest sannsynlig anslag tilsvarer et uendelig stort skjevhetsforhold. For dette tilfellet er Theta-optimal lik 4,85 og Zeta optimal er lik 2,24.

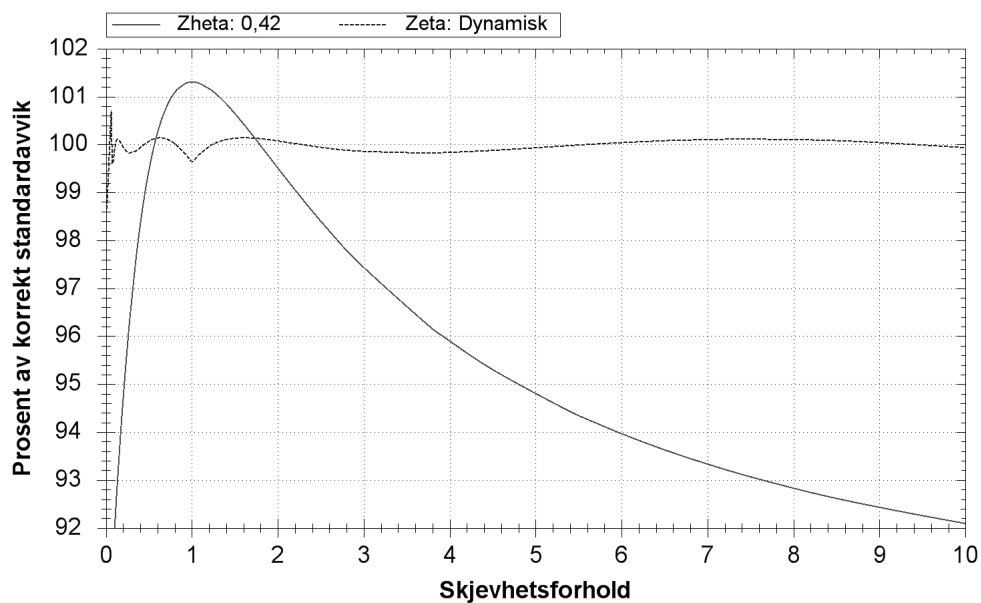
For tilfellet der alle der alle tre anslagene er like spiller det ingen rolle hvilke Theta- og Zeta-verdier en bruker. Resultatet blir det samme uansett.

Alt i alt er det derfor ingenting i veien for at dette kan implementeres i et dataprogram på en slik måte at det blir abstrahert vekk for brukeren. Det trenger ikke bli mer komplisert å vurdere eller gi innputt til analysen.

Å bruke disse formlene for å beregne en egen Theta og Zeta for hvert trippelanslag vil et nøyaktiger resultat enn bruk av dagens formler. Figur 5-18 og 5-19 viser henholdsvis sammenligninger av hva feilen av forventningsverdien og standardavviket blir for de gamle formlene og den nye dynamiske varianten.



Figur 5-18 Feil på forventningsverdi ved bruk av dynamisk Theta og  $\Theta = 0,42$



Figur 5-19 Feil på standardavvik ved bruk av dynamisk Zeta og  $Zeta = 2,53$

### 5.1.7 Regne ut Alfa, Beta og forskyvning

En alternativ og mer nøyaktig metode til å bruke Trinnvis-formlene til å regne ut forventingsverdi og standardavvik for et trippelanslag er å bruke Gammafordelingens egne formler for dette (se avsnitt 3.5 ‘Gamma- og Erlangfordeling’ på side 32). Dette krever at vi først klarer å omsette et trippelanslag til Gammafordelingens parametere Alfa, Beta og forskyvning. Dette er i utgangspunktet ikke en enkel sak og er i praksis umulig å gjøre uten å ta datamaskiner til hjelp. For å muliggjøre dette har vi utviklet et sett algoritmer som beregner disse parametrene ut i fra et gitt trippelanslag. Disse ligger vedlagt i vedlegg C.

Et problem ved bruk av denne metoden er at den er betydelig mye tregere enn å benytte enten de vanlige Trinnvis-formlene eller de dynamiske variantene. Den er dog ikke tregere enn at en vanlig datamaskin klarer å regne ut disse parametrene på et brøkdels sekund. Så når det gjelder beregning av en enkeltpost i et kalkyleprogram vil det ikke være noen merkbar forskjell for brukeren, men det kan by på problemer når man skal regne i hop postene, multipliserer inn faktorer etc. Dette vil være avhengig av hvordan programmet er bygd opp. I dette ligger også en annen ulempe med denne metoden kontra dynamisk utregning av Theta og Zeta. Den er mye vanskeligere å implementere. Selv med formlene i vedlegg C. tilgjengelig er det ikke tilrådelig å prøve å implementere dette med mindre man enten har svært gode kunnskaper om Gammafordelingen og de nødvendige statistiske beregningsmetodene eller et hav av tid for å sette seg inn i dette.

Selv om det er visse utfordringer i å implementere denne metoden bør man absolutt vurdere den ved utvikling av kalkulasjonsprogrammer. Ikke bare er den så å si 100% nøyaktig, med denne metoden kan også potensielt ivareta skjevheten i inngangsdataene ved summering av poster.

### 5.1.8 Korrelasjon i analytiske metoder

Modellering av korrelasjon har tradisjonelt ikke vært benyttet i forbindelse med Trinnvis kalkulasjon.

Parvis korrelasjon, det vil si der vi kun har to poster som er korrelert med hverandre, er enkelt sak å beregne. Kovariansen blir da som vist i avsnitt 2.6:

$$\text{Cov}(X, Y) = 2 \cdot \text{Cor}(X, Y) \cdot \sigma_X \cdot \sigma_Y$$

der  $\text{Cor}(X, Y)$  er korrelasjonsfaktoren og  $\text{Cov}(X, Y)$  er kovariansen.

Summen av variansen for to korrelerte poster A og B blir da:

$$\sigma_{\text{TOT}}^2 = \sigma_A^2 + \sigma_B^2 + 2 \cdot \text{Cor}(A, B) \cdot \sigma_A \cdot \sigma_B$$

Dette kan utvides til flere poster. For eksempel blir summen av variansen for tre poster A, B og C da:

$$\sigma_{\text{TOT}}^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + 2 \cdot \text{Cor}(A, B) \cdot \sigma_A \cdot \sigma_B + 2 \cdot \text{Cor}(A, C) \cdot \sigma_A \cdot \sigma_C + 2 \cdot \text{Cor}(B, C) \cdot \sigma_B \cdot \sigma_C$$

## Korrelasjonsmatriser

For alle andre tilfeller enn parvis korrelasjon vil det mest hensiktsmessige være å sette opp en korrelasjonsmatrise og tilhørende varians-kovarians matrise som vist i henholdsvis tabell 5.1 og tabell 5.2. Summen av variansen for postene blir da summen av alle verdiene i varians-kovariansmatrisen.

Tabell 5.1 *Korrelasjonsmatrise*

	A	B	C
A	1	Cor(A,B)	Cor(A,C)
B	Cor(A,B)	1	Cor(B,C)
C	Cor(A,C)	Cor(B,C)	1

Tabell 5.2 *Varians-kovariansmatrise*

	A	B	C
A	$\sigma_A^2$	Cor(A,B) $\sigma_A\sigma_B$	Cor(A,C) $\sigma_A\sigma_C$
B	Cor(A,B) $\sigma_A\sigma_B$	$\sigma_B^2$	Cor(B,C) $\sigma_B\sigma_C$
C	Cor(A,C) $\sigma_A\sigma_C$	Cor(B,C) $\sigma_B\sigma_C$	$\sigma_C^2$

Materialet omkring korrelasjonsmatriser er hentet fra finansiell usikkerhetsstyring og har, så vidt vi vet, ikke vært prøvd brukt i analytisk beregning av usikkerhetsanalyser. Det er i realiteten en del spørsmål som må avklares før at denne metoden skal være særlig brukbar til vårt formål.

For det første benyttes slike korrelasjonsmatriser på flate strukturer som aksjeporteføljer. Det er usikkert hvorvidt det vil være mulig å bruke korrelasjonsmatriser inn en hierarkisk kostnadsstruktur slik som de man har i usikkerhetsanalyser.

En annen ting som må gjøres er at man må ha inn metoder for å sørge for at matrisen blir positivt definert. La oss si at A er 100% korrelert med B og 100% korrelert med C. I dette tilfellet må nødvendigvis B og C også være 100% korrelerte.

De som er interessert i bruk av korrelasjonsmatriser anbefales å søke innen litteratur om finansiell matematikk og usikkerhetsstyring.



## 5.2 Simulering

Simulering går ut på å få datamaskiner til å imitere, eller *simulere*, ulike typer operasjoner for virkelige systemer. For å klare å imitere et system må vi som regel gjøre et sett antakelser om hvordan det fungerer. Disse antagelsene, som vanligvis tar form som matematiske eller logiske forhold, utgjør til sammen en modell som man bruker for å forstå hvordan det korresponderende virkelige systemet oppfører seg.

Å vurdere en slik modell kan gjøres både analytisk og ved hjelp av simulering.

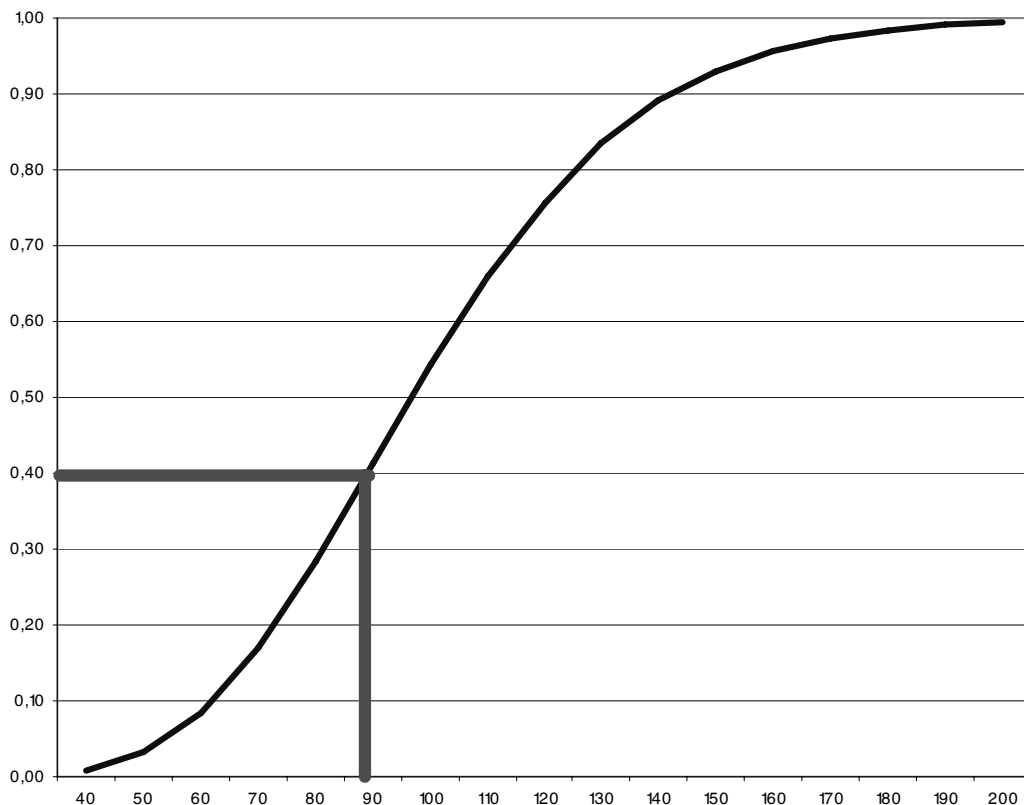
Det en datamaskin gjør for å evaluere en slik modell kjører den gjentatte ganger. Hver gjennomkjøring, eller *iterasjon* tilsvarer en mulig virkelighet for det aktuelle systemet. Ved å sammenstille dataene fra et stort antall gjennomkjøringer får vi en statistikk som er representativ for virkeligheten.

### 5.2.1 Om tilfeldig tallgeneratorer

I hjertet av ethvert simuleringsprogram finner man en tilfeldig tallgenerator. Som navnet tilsier er en tilfeldig tallgenerator noe som genererer tilfeldige tall. En tilfeldig tallgenerator som de fleste har et forhold til er en vanlig sekssidet terning. Hvis man kaster en terning er det tilfeldig hvilket av de seks mulige tallene som kommer opp. "Lotto"-maskiner der man har kuler med tall på er også en form for tilfeldig tall generator. Men i simuleringssammenheng trenger vi tusenvis om ikke millionvis av ulike tilfeldige tall, så å bruke fysiske metoder som dette er uaktuelt. Det er i det følgende beskrevet hvordan en datamaskin går frem for å generere "tilfeldige"-tall for å bruke i simuleringer. "Tilfeldig" er satt i hermetegn for som vi skal se er tallene egentlig ikke helt tilfeldige.

## Med utgangspunkt i U[0,1]

Uavhengig av hvilken fordeling man er ute etter å trekke tilfeldige tall fra går de aller fleste simuleringsverktøy veien om en uniform fordeling mellom 0 og 1 (U[0,1]). Det vil si at alle tall mellom 0 og 1 har en like stor sannsynlighet, mens tall mindre eller 0 eller større eller 1 aldri vil dukke opp. Har vi en tilfeldig tall fra en slik fordeling kan vi regne dette om til et tilfeldig tall fra en hvilken som helst annen fordeling, så sant vi har en formel for den kumulative sannsynlighetsfordelingen for denne. Nedenfor er vist et eksempel med Gammafordelingen.



Figur 5-20 S-kurve

S-kurven ovenfor i Figur 34 er den kumulative sannsynlighetsfordelingen for en Gammafordeling med Alfa = 10 og Beta = 10. Vi ser på kurven at den kumulative sannsynligheten strekker seg fra 0 til 1. Hvis en tilfeldig tallgenerator for eksempel gir oss 0,4 kan vi gå inn på grafen og se at dette tilsvarer en verdi av ca 88 for den gitte fordelingen. (Datamaskinen leser ikke av grafer, men bruker den matematiske formelen for kurven til å regne seg frem til det samme.)

## U[0,1] generatorer

En  $U[0,1]$  generator er en tilfeldig tall generator som spytter ut tilfeldig tall jevnt fordelt mellom 0 og 1. Denne rekken som kommer ut kalles vanligvis en tallstrøm.

Hvis man ønsker å generere tilfeldige tall for bruk i spill som Lotto må tallene virkelig være helt tilfeldige. Når det gjelder simulering er ikke det så nøye. Det som er nødvendig for å få et simuleringsresultat som gjenspeiler virkeligheten er at vi har en tallstrøm som er statistisk representativ og som ikke på noe tidspunkt “looper”, det vil si at tallstrømmen midtveis i simuleringen begynner å gjenta seg selv fra begynnelsen av.

Tvert om er det fordelaktig at tilfeldig tall generatorer ikke er virkelig tilfeldig. For at simuleringen skal kunne være etterprøvable for andre er det ønskelig at det er mulig å reproducere tallstrømmen. I tillegg vil man også i enkelte tilfeller for å oppnå større presisjon ønske å bruke identiske tilfeldige tall ved simulering av ulike systemer. Muligheten til å reproducere en tallstrøm er umulig med sanne tilfeldige tall generatorer. Du vil for eksempel aldri klare å gjennomføre en serie med terningkast og deretter få eksakt samme resultat hvis du forsøker å gjenta det.

Spørsmålet som gjenstår er hvordan en datamaskin klarer å generere slike strømmer av kvasi-tilfeldige tall. I praksis gjøres dette ved hjelp av en såkalt aritmetisk generator. Neste avsnitt viser et eksempel på hvordan disse fungerer.

### Eksempel på aritmetisk tilfeldig tall generator

Dette eksempelet viser en av de første aritmetiske generatorene som ble utviklet. Den ble foreslått av Neumann og Metropolis i førtiårene.

Start med et firesifret positivt heltall  $Z_0$  og opphøy det i andre. Dette resulterer i et heltall med opptil 8 siffer. Hvis nødvendig legg til nuller til venstre for tallet for å få eksakt 8 siffer. De fire midterste sifrene i serien utgjør  $Z_1$ . Sett et komma foran dette tallet for å få det første tallet i tallstrømmen. Ta deretter  $Z_1$ , opphøy det i andre og så videre for å få neste tall i tallstrømmen.

Denne metoden er ikke noen god metode. Den ender nokså raskt på null der den forblir til evig tid. Metodene som brukes i dag er langt mer sofistikert enn dette. Men prinsippet er det samme. Man starter med et tall, et såttall, setter dette inn en formel for å generere første "tilfeldige" tall, som igjen settes inn i samme formel for å generere neste tall i rekken.

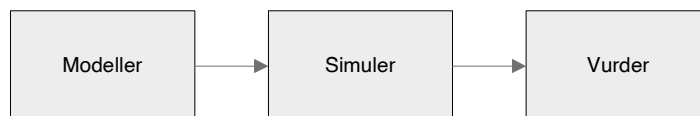
Tabell 5.3 *Eksempel på tilfeldig-tall-generator*

$n$	$Z_n$	$U_n$	$Z_n^2$
0	1234		01522756
1	5227	0,5227	27321529
2	3215	0,3215	10336225
3	3362	0,3362	11303044
4	3030	0,303	09180900
5	1809	0,1809	03272481
6	2724	0,2724	07420176
7	4201	0,4201	17648401
8	6484	0,6484	42042256
9	0422	0,0422	00178084

## 5.2.2 Gangen i en simulering

En simulering har grovt sett tre steg

1. Modellering - Lage en modell av virkeligheten
2. Simulering - Kjøre modellen
3. Vurdere resultatene fra kjøringen



Figur 5-21 Overordnet arbeidsgang i en simulering

### Modellering

Det første man må gjøre ved bruk av simuleringstøytøy er å lage en modell for systemet vi ønsker å simulere. Denne modellen vil bestå av en rekke matematiske formler og forhold som til sammen representerer virkeligheten på det abstraksjonsnivået vi opererer på. For eksempel vil en modell som tar for seg produksjonsanlegget i en fabrikk vil være meget detaljert og inkluderer ting som en enkelt gjenstands ferd langs samlebåndet, mens en kostnadsberegning for et byggeprosjekt vil foregå på et langt mer overordnet nivå. Dette henger selvsagt sammen med hva som finnes av tilgjengelige inndata og systemenes kompleksitet.

### Ulike type modeller

Vi kan skille mellom to ulike hovedtyper modeller; statiske og dynamiske. En statisk simuleringmodell er en representasjon av et system ved et gitt tidspunkt, eller et system der tiden ikke spiller noen rolle, for eksempel et kostnadsanslag. En dynamisk simuleringmodell representerer et system som utfolder seg med tiden, slikt som et samlebånd i en fabrikk eller fremdriftsplanen for et prosjekt.

Monte-Carlo simuleringer er av typen statiske simuleringer der tiden ikke spiller noen rolle.

Alle dynamiske simuleringmodeller har tilstandsvariabler som er tidsavhengige og som utvikler seg over tid. Eksempler på tilstandsvariabler kan være hastigheten til en bil eller tilstanden til en aktivitet. Disse to eksemplene representerer to undertyper av dynamiske simuleringmodeller. Vi skiller mellom diskret-hendelse simulering og kontinuerlig-tid simulering. Kontinuerlig simulering brukes til å simulere systemer der tilstandsvariablene endres kontinuerlig med tiden, som eksempelet med hastigheten på en bil, og disse variablene er gjerne uttrykt i modellen som differensialligninger. Diskret-hendelse simulering derimot befatter seg med modeller der endringen i tilstandsvariablene endres øyeblikkelig fra en tilstand til en annen. Eksempelvis en aktivitet i en fremdriftsplan endrer tilstand fra påbegynt til ferdig. Enhver simulering som kan modelleres som en diskret-hendelse simulering kan også modelleres som en kontinuerlig-tid simulering. Men en kontinuerlig-tid simulering er mye mer ressurskrevende med tanke på datakraft. I prosjektstyringssammenheng har kontinuerlig-tid simulering ingen plass. Der det er nødvendig med dynamiske simulering,

som ved simulering av fremdriftsplaner, er det tilstrekkelig å benytte diskret-hendelse simuleringer. Kontinuerlig-tid simuleringer er derfor ikke behandlet i det følgende.

Selv om vi ovenfor brukte fremdriftsplaner som et eksempel på dynamiske simuleringer er det også fullt mulig å beregne vanlige fremdriftsplaner ved hjelp av statiske simuleringer. Det er først nødvendig å bruke dynamiske simuleringer når man ønsker å modellere inn et ikke deterministisk hendelsesforløp. Det vil si at det ikke er gitt hvilke aktiviteter som skal gjennomføres og i hvilken rekkefølge.

## Modellering av simuleringer

I og med at tiden ikke spiller noen rolle i Monte-Carlo simulering er modelleringen her relativt enkelt. Kortversjonen er at man først setter opp et regnestykke for kalkylen (på samme måte som man ville gjort ved bruk av en deterministisk kalkyle), deretter går inn og spesifiserer hvilke verdier som er usikre, hvilken statistisk fordeling de skal ha, og til slutt sier noe om eventuelle sammenhenger (korrelasjon) mellom inngangsdataene.

Hvordan regnestykket for kalkylen settes opp avhenger av simuleringstøyet man bruker. De to mest utbredte programmene, @Risk og Crystal Ball, kjører på toppen av Excel. I disse verktøyene setter man opp kalkylen som et hvilket som helst annet regneark.

I enkelte verktøyer velger man hvilke verdier som er usikre og hvilken sannsynlighetsfordeling de skal benytte i det man bygger opp modellen. I andre, som for eksempel de tidligere nevnte Excel-baserte, kan dette gjøres etter at man har satt opp "grunnkalkylen". Her velger man regnearkcellen til den usikre verdien spesifiserer hvilken fordeling den skal ha.

Siste steg her er å eventuelt oppgi om noen av de usikre inngangsverdiene er korrelert med hverandre. Se kapittel 2.6 for nærmere forklaring hva korrelasjon er.

Modellering av diskret-hendelse simuleringer er en god del mer komplisert enn modellering av Monte Carlo simuleringer. På grunn av at det er såpass komplisert som det er og at det har et begrenset bruk for vår del, kommer vi ikke til å gå nærmere inn på dette her.

### 5.2.3 Kjøring av simuleringsmodellen

Dette kapittelet tar for seg hva datamaskinen gjør under kjøring av en Monte Carlo simuleringmodell. Det forutsettes at leseren har lest kapittel 5.2.1 om tilfeldig-tallgeneratorer.

Før simuleringen kjøres må man sette hvor mange iterasjoner man ønsker. Man må sørge for å kjøre nok iterasjoner til at simuleringen konvergerer. Hvor mange iterasjoner som er nødvendig er helt avhengig av modellen. Jo mer kompleks og jo flere inngangsvariable som man har, jo flere iterasjoner vil være nødvendig. Noen verktøyer har innebygde funksjoner for å si noe om når simuleringen konvergerer. I andre tilfeller er en sjekk å kjøre simuleringen to eller flere ganger med ulike såttall. Hvis det er stort sprik i resultatet mellom simuleringene har man ikke brukt tilstrekkelig antall iterasjoner.

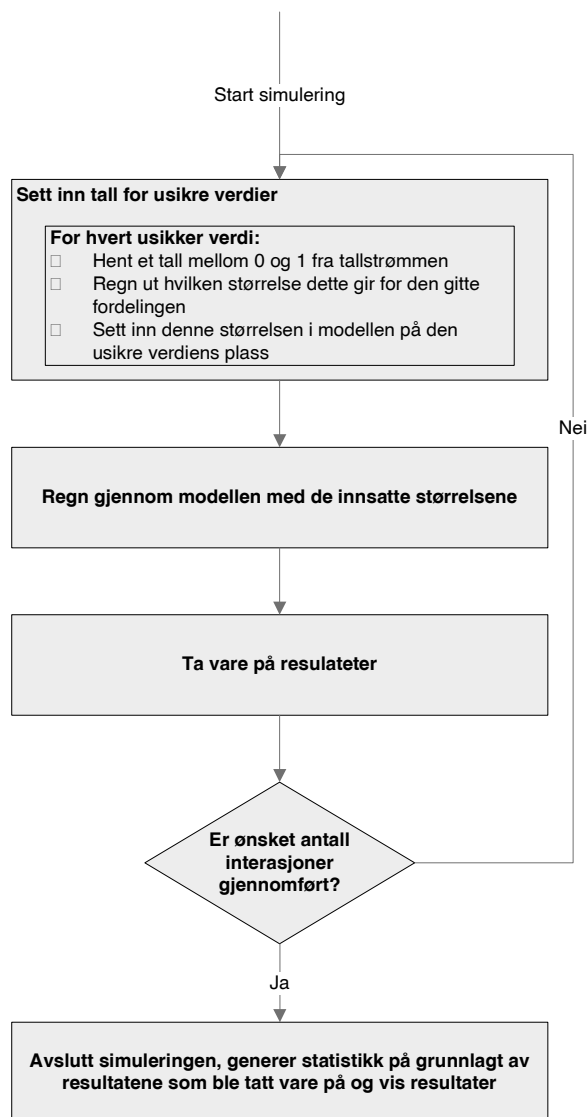
I de fleste simuleringstøyer som er aktuelle å bruke i usikkerhetsanalyser brukes et tilfeldig såttall hver gang simuleringen kjøres, men mange har muligheten til å gå inn og spesifisere et såttall. Hvis simuleringen benytter et tilfeldig såttall for hver gjennomkjøring er det ikke bød-

vendig å manuelt sette såtallene hvis man benytter den ovennevnte metoden for å sjekke om simuleringen konvergerer.

I mange simuleringsverktøy er det også mulighet for å sette såtall og antall simuleringer.

Figur 5-22 nedenfor viser stegene i kjøringen av simuleringen. Etter at brukeren har startet simuleringen bytter datamaskinen ut alle modellenes usikre verdier med absolutte tall tilfeldig trukket fra de sannsynlighetsfordelingene som hver størrelse er oppgitt til å ha. Deretter gjennomfører den alle beregninger vi har angitt og tar vare på de tallstørrelsene vi ønsker statistikk for.

Dette gjentas helt fram til vi har fått gjennomført det ønskede antall iterasjoner. Når det ønskede antall iterasjoner er gjennomført avsluttes simuleringen og simuleringsprogrammet genererer statistikker på grunnlag av de tallstørrelsene som ble tatt vare på underveis, som brukeren senere kan velge å fremstille som s-kurver etc.



Figur 5-22

*Gangen i kjøring av en simulering*

### 5.2.4 Korrelasjon i simulering

For enkelte sannsynlighetsfordelinger er det mulig å generere korrelerte variabler under selve kjøringen av modellen. For andre, som Gammafordelingen, er dette umulig. Og det er i alle fall umulig å generere korrelerte variabler for ulike typer sannsynlighetsfordelinger under kjøring av simuleringen. Simuleringsprogrammene kommer seg rundt dette ved å jukse litt. Det som typisk gjøres er at før selve kjøringen av modellen starter, trekker simuleringsprogrammet ut alle tilfeldige tall som trenges for å kjøre hele simuleringen (samtlige iterasjoner) for alle korrelerte fordelinger og sorterer dem slik at de blir korrelert. Disse settene av korrelerte verdier trekkes så senere inn i simuleringen ved hver iterasjon.

### 5.2.5 Trippelanslag og simulering

Hvis man ønsker å bruke trippelanslag av den typen man bruker ved Trinnvis-kalkulasjon inn i en simuleringsmodell står man ovenfor et problem. Det er i utgangspunktet umulig. I @Risk og Crystal Ball har man valget mellom å bruke gammafordelingens normale parametre Alfa, Beta og Forskyvning (kalt Shift i engelskspråklige programmer) eller bytte ut en eller flere av disse med en vilkårlig prosentkvantil, men det er ikke mulig på noe vis å direkte legge inn mest sannsynlige verdi. I Definitive Scenario virker det ut ifra dialogene for innlegging som om den midtre verdien man legger inn er mest sannsynlige verdi. Men hvis man leser programmets dokumentasjon fremgår det at det faktisk er 50% kvantilen som legges inn. Sistnevnte program har for så vidt heller ikke mulighet for å benytte gammafordelingen, man er nødt til å benytte en lognormalfordeling. Et annet problem med disse simuleringsprogrammene er at de heller ikke lar en direkte benytte venstreskjeve gammafordelinger.

Det er mulig å komme seg rundt disse to manglene ved å transformere de venstreskjeve fordelingene og transformere tilbake resultatet som beskrevet i avsnitt 3.5.4 og benytte algoritmen beskrevet i vedlegg C for å finne Alfa.



## 5.3 Simulering versus analytiske metoder

Både simulering og analytiske metoder er mulig å benytte for å beregne de typer kostnads-kalyer vi snakker om i denne rapporten, men det kan være greit å ha for seg noen fordeler og ulemper ved de to metodene. Tabell 5.4 summerer kort opp hva fordelene og ulempene er for simulering og Trinnvis eller tilsvarende metode.

Tabell 5.4 *Forskjeller mellom simulering og analytisk metode*

	Simulering	Trinnvis kalkulasjon per i dag	Potensiale for analytiske metoder
Struktur og muligheter for komplekse modeller	Fullstendig fri struktur. Trivielt å modellere kompliserte modeller (rent teknisk sett).	Krever en viss rigiditet i strukturen	Teoretisk sett mulig å ha like fri struktur og komplekse modeller som i simulering, men det vil kreve modeller og beregningsmetoder som er langt mer avanserte og kompliserte enn det som finnes per i dag.
Nøyaktighet	Svært nøyaktig forutsatt gode inndata og tilstrekkelig antall iterasjoner	Forventningsverdi tilnærmet helt nøyaktig. En del feil i standardavviket.	100% nøyaktig.
Korrelasjon	Korrelasjon modelleres fritt	Ingen muligheter for å modellere korrelasjon	Stort forbedringspotensial over dagens analytiske metoder. Men det er uvisst om det lar seg gjøre like bra som simulering
Typer fordelinger	Flere enn man trenger	Begrenset til Gamma-/Erlangfordeling	Teoretisk mulig å få inn hvilke som helst fordelinger
Tilbakemelding på endringer	Simuleringen må kjøres for å se effekten av endringer i modellen	Beregningene vises umiddelbart etter at en verdi er endret	Beregningene vises umiddelbart etter at en verdi er endret

Tabell 5.4 Forskjeller mellom simulering og analytisk metode

	Simulering	Trinnvis kalkulasjon per i dag	Potensiale for analytiske metoder
Bruker- terskel	Stiller i utgangspunktet store krav til brukerens kompetanse, både på selve programmet og ren statistikk. Friheten i strukturen gjør at man er nødt til å ha kontroll over hva man gjør. Mulig å lage spesialtilpassede verktøy som har lavere brukerterskel.	Krever mindre teoretiske forkunnskaper enn simulering, men har i realiteten en høy brukerterskel på grunn av at all tilgjengelig programvare har horrible brukergrensesnitt.	Stort forbedringspotensial i brukergrensesnittet. Lavest brukerterskel av alle.
Innlegging av data	Vanskelig å legge inn tradisjonelle trippelanslag	Trivielt	Trivielt
Gjennom- siktighet	Vanskelig å følge de enkelte data gjennom beregningen	Rimelig enkelt å følge de enkelte data gjennom beregningen. Det kommer frem mange flere pedagogiske poeng med enkle analytiske modeller enn i simulering.	Noe vanskeligere å følge de enkelte data som følge av at man går over til dynamiske formler som ikke lar seg kontrollere 100% ved håndregning.
Muligheter for komplekse modeller	Trivielt å modellere (rent teknisk sett)		

## 5.4 Resultater av beregninger

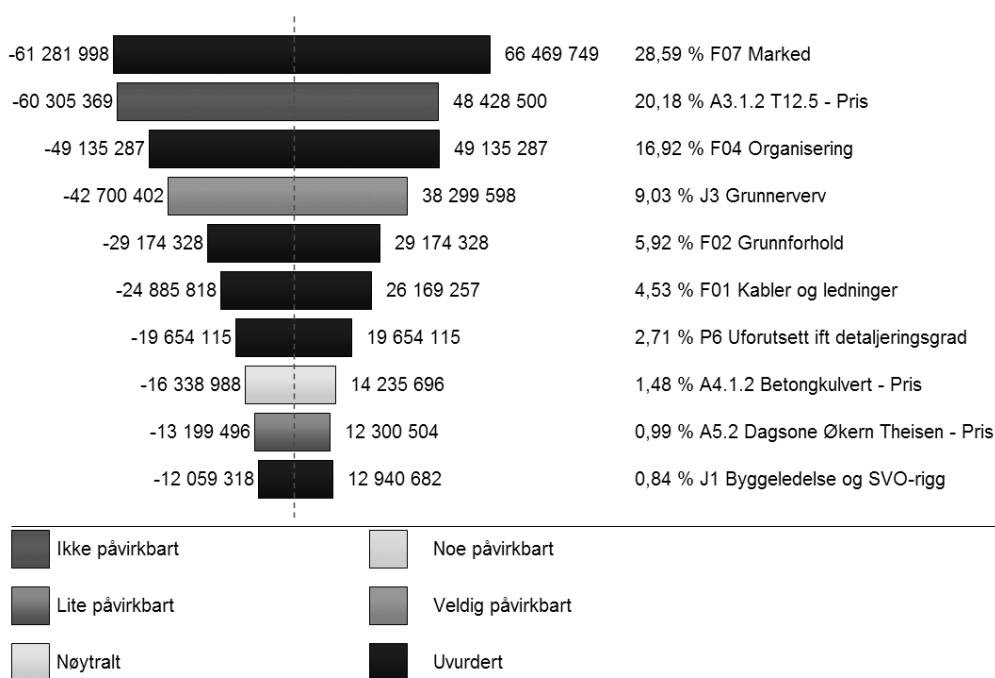
### 5.4.1 S-kurve

Å benytte en s-kurve (kumulativ sannsynlighetsfordeling) er den vanligste formen å fremstille resultatet fra en usikkerhetsberegning. S-kurven kan brukes til å lese med hvor stor sannsynlighet man vil havne innenfor en gitt verdi.

Kumulative sannsynlighetsfordelinger er tidligere beskrevet i kapittel 2.4.2 og vi går av den grunn ikke mer i detalj på det her.

### 5.4.2 Tornadodiagram

Tornadodiagram er en måte å fremstille hvilke poster i kalkylen som bidrar mest til usikkerheten i prosjektet. Diagrammet har fått navnet av at det i formen ligner på en Tornado. Det er noen ulike varianter ute og går, men figur 5-23 viser et eksempel på hva vi mener et Tornadodiagram bør inneholde.



Figur 5-23 Eksempel på anbefalt tornadodiagram

Angivelse av den relative størrelsen på usikkerheten

Det er en del ting ved tornadodiagrammet i figur 5-23 som er verdt og merke seg. For det første er skjevheten i usikkerheten angitt, man kan se at ikke alle postene er symmetrisk om nullinjen. Hvis man kun benytter seg av postenes varians når man tegner diagrammet vil man ikke klare å fremstille dette. Det vi har gjort her er at vi regnet ut hvor mye større og mindre summen av prosjektkalkylen blir hvis vi treffer ytterpunktene i anslagene for hver

post. Dette gir det samme resultatet størrelsesmessig på grafen som om vi hadde plottet variansen direkte, men vi får altså forskjøvet barene noe for å vise i hvilken retning usikkerheten tenderer.

Tallene som er plottet ved siden av barene viser hvor mye større eller mindre summen av prosjektkalkylen blir i absolutte termer hvis vi treffer ytterpunktene på den angitte post.

Dette diagrammet viser også hva påvirkbarheten av den enkelte post er. Å få med dette er viktig for det videre arbeid med usikkerhetsanalysen.

## 6. Følsomhet for feil i inngangsdata

Det vi har sett på så langt er hvorvidt selve formelverket er korrekt under forutsetning om at alle inngangsdata er korrekt. Men hva om de ikke er det? Hva om vi bommer når vi anslår en verdi? Hva skjer da?

For å finne ut av det gjennomførte vi følsomhetsanalyser for trippelanslag. Måten dette er gjort på er ved at to av verdiene i trippelanslaget er holdt konstant mens den tredje varieres.

De to andre forutsetningene som er gitt i avsnitt 5.1.2 antas fortsatt å være gyldige. Det vil si vi antar at kostnadselementet vi anslår har en virkelig sannsynlighetsfordeling, som er ukjent for oss, men som kan uttrykkes ved hjelp av en gammafordeling.

Hvor stor feilen blir på forventningsverdien er til dels avhengig av den virkelige kurvens skjevhet og av hvor stor spredning det er i verdiene i forhold til forventningsverdien, altså det relative standardavviket til den virkelige kurven.

For standardavviket er det kun den virkelige kurvens skjevhet som har noe å si.

Grafene som er visst i det følgende er for høyreskjeve fordelinger. For venstreskjeve kurver vil det bli de samme grafene bortsett fra at de vil være speilet, utslaget på y-aksen vil gå andre veien og kurvene for  $n$  vil være kurven for  $\infty$  og vice versa.

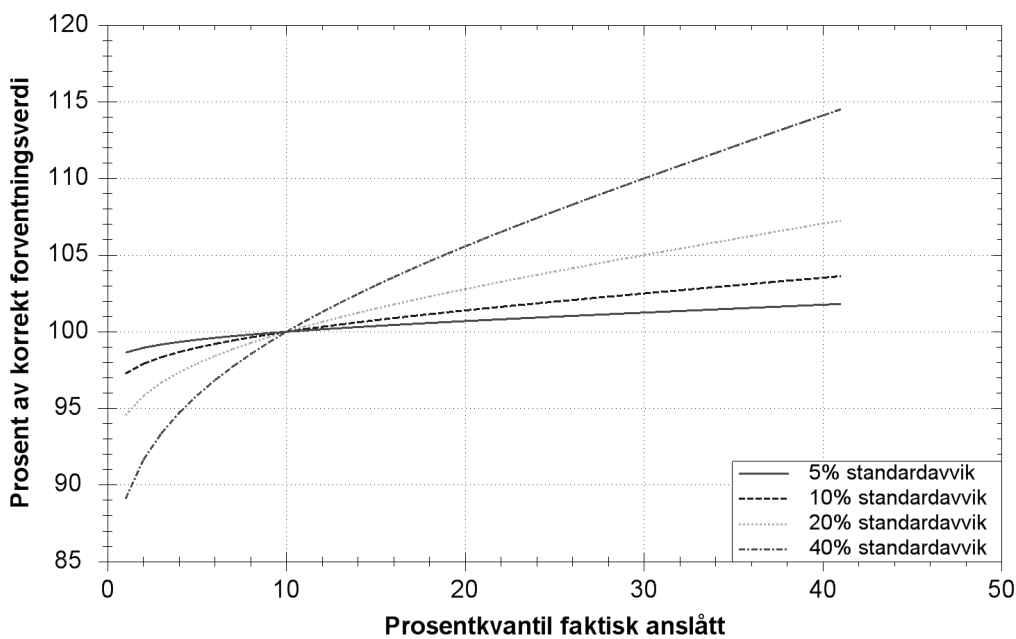
Det er tatt utgangspunkt i trippelanslag ved bruk av 10/90.

Vi har kun sett på følsomheten for feil inngangsdata for enkeltelementer. Det vil si der det er angitt verdier som standardavvik etc. er det verdiene for det enkelte element som er angitt, ikke totalt av en kostnads kalkyle.

## 6.1 Følsomhet for feil nedre anslag (n)

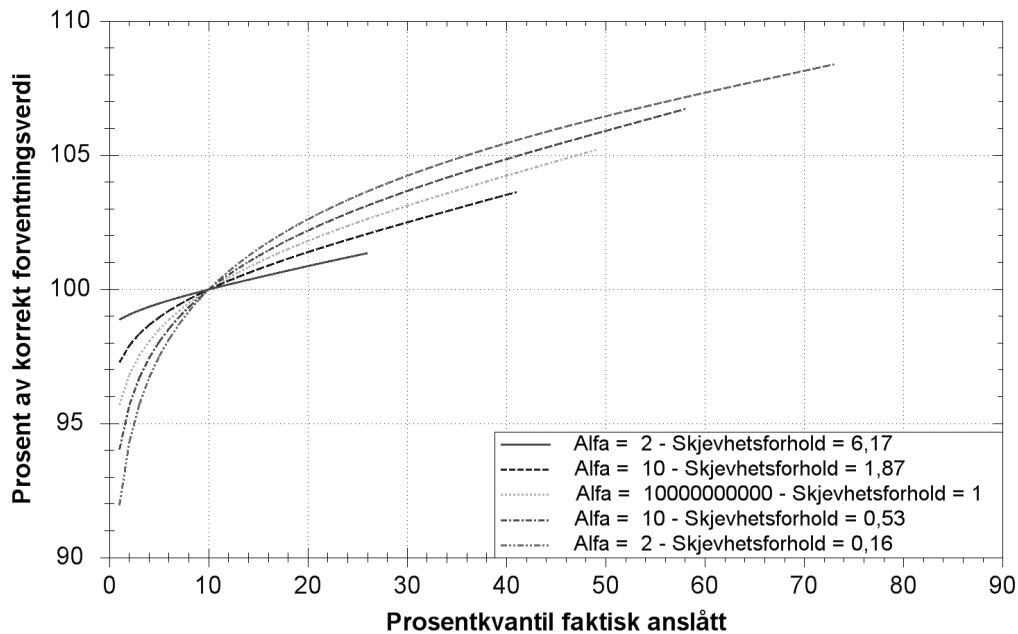
Figur 6-1 viser hvor stor feilen på forventningsverdien blir i prosent ved feil nedre anslag  $n$ . Den anslåtte  $n$  er her uttrykt ved hvilket prosentkvantil på den virkelige fordelingen vi faktisk anslår. Verdiene som vises på x-aksen er altså “virkelighetens prosentkvantiler”. Det er forutsatt at vi ikke kan anslå verdier som er utenfor den virkelige fordelingen. I og med at en høyreskjev gammafordeling er åpen på høyre side er dette en rimelig grei forutsetning. 100% kvantilet for fordelingen svarer til en verdi som er uendelig stor, og det er vel ganske så usannsynlig at vi anslår noe er større enn dette. Den virkelige fordelingen vil dog ha et absolutt minimumspunkt, men i praksis er det ikke spesielt sannsynlig at vi anslår en nedre verdi som er hinsides mye lavere en hva den virkelige verdien er.

Følsomheten for  $n$  er i figuren vist fra 1%-kvantilet til 41%-kvantilet. Grunnen til den øvre grensen er at mest sannsynlig verdi, som vi forutsetter at vi har anslått korrekt, for en gammafordeling med Alfa = 10 ligger på 41,26%-kvantilet. Vi går her ut fra at vi ikke anslår nedre verdi  $n$  til å være større enn mest sannsynlige verdi (selv om dette er teoretisk mulig ved ekstremt høyreskjeve fordelinger).



Figur 6-1 Forventningsverdiens følsomhet for feil nedre anslag ( $n$ ) for ulike standardavvik ved Alfa = 10

Figur 6-2 viser hva feilen blir på forventningsverdi for ulike Alfa/skjevhetsforhold ved 10% standardavvik. Vi ser her at feilen på forventningsverdien blir større jo mer venstreskjev den virkelige fordelingen er. Dette kommer av at  $n$  her i utgangspunktet ligger lengre unna kurvens tyngdepunkt (tenk vektstangprinsippet) og at hvert prosentkvantil utgjør et større hopp i absolutte verdier. Å anslå  $n$  feil for svært høyreskjeve kurver gir små utslag, mens det for svært venstreskjeve kurver gir meget store utslag.

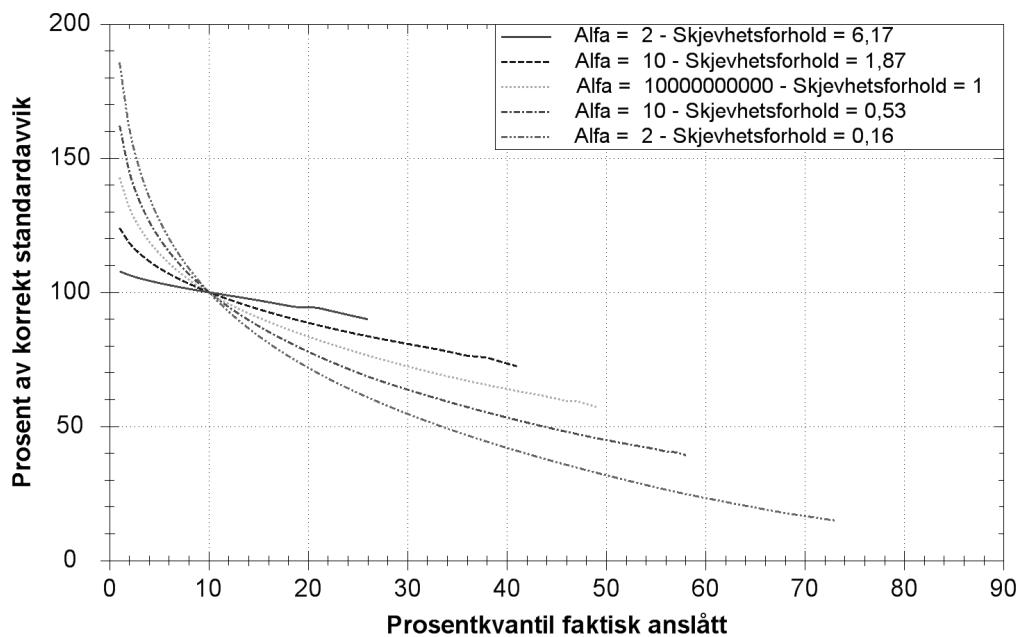


Figur 6-2 Forventningsverdiens følsomhet for feil nedre anslag ( $n$ ) for ulike Alfa ved 10% standardavvik

---

*TIPS! Jo skjevare fordelingen er, jo større er konsekvensen av å bomme. Hvis man har et skjevt kostnadselement er det svært viktig å anslå den mest ekstreme verdien riktig.*

---



Figur 6-3 Standardavvikets følsomhet for feil nedre anslag ( $n$ ) for ulike Alfa

Figur 6-3 viser standardavvikets følsomhet for feil nedre anslag. Vi ser her at hvis vi anslår  $n$  for lavt øker standardavviket. Dette er naturlig i og med at vi med å gjøre dette øker avstanden fra  $n$  til  $\emptyset$ . Feilen på standardavviket blir også størst for venstreskjeve fordelinger på grunn av at hvert prosentkvantil representerer et større gap i absolutte størrelser for disse.

---

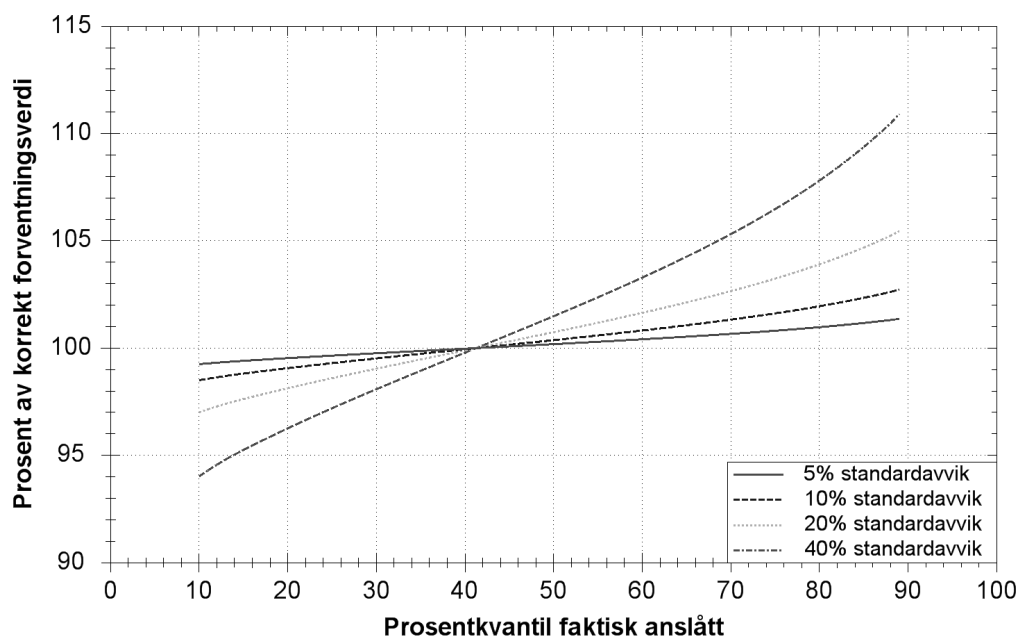
*I PRAKSIS: Anslår du nedre verdi for lavt vil forventningsverdien gå ned og standardavviket vil øke*

---



## 6.2 Følsomhet for feil anslag for mest sannsynlig verdi (s)

Figur 6-4 viser følsomheten for feil anslag av mest sannsynlig verdi for ulike relative standardavvik ved Alfa = 10. I og med at vi antar at vi har anslått nedre og øvre verdi korrekt er kurven avgrenset ved 10% og 90% kvantilene. Vi ser at feilen man gjør ved å anslå feil her er mye mindre enn hvis man anslår nedre verdi feil. Og som vi skal senere gjelder det samme i forhold til anslag av øvre verdi. Konsekvensen av å bomme på mest sannsynlig verdi er altså mye mindre enn å bomme på ytterpunktene



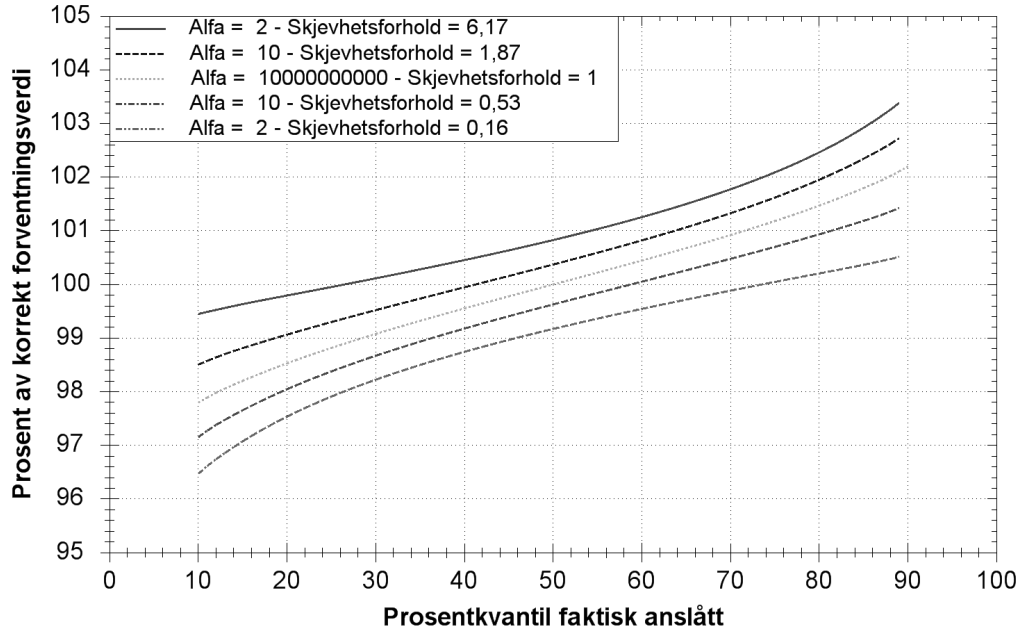
Figur 6-4 Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike standardavvik ved Alfa = 10

---

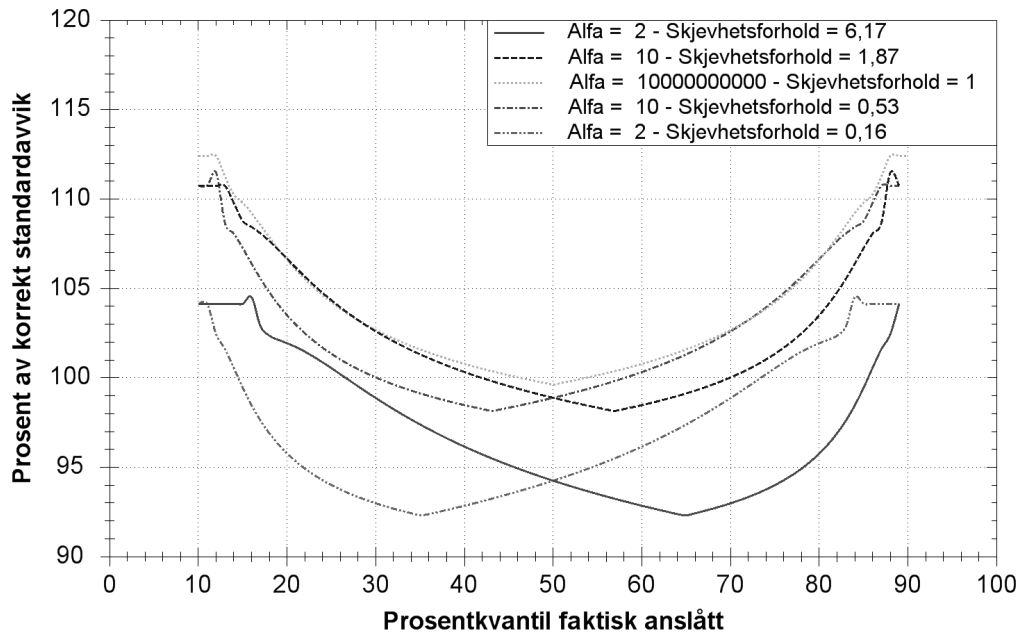
*TIPS! Anslå ytterpunktene først for du anslår mest sannsynlig verdi.*

---

Figur 6-5 viser forventningsverdiens følsomhet for feil mest sannsynlig verdi. I og med at verdien av Alfa bestemmer i hvilket prosentkvantil mest sannsynlig verdi ligger krysser kurvene 100% linja ved ulike punkter. Det samme gjelder for figur 6-6.

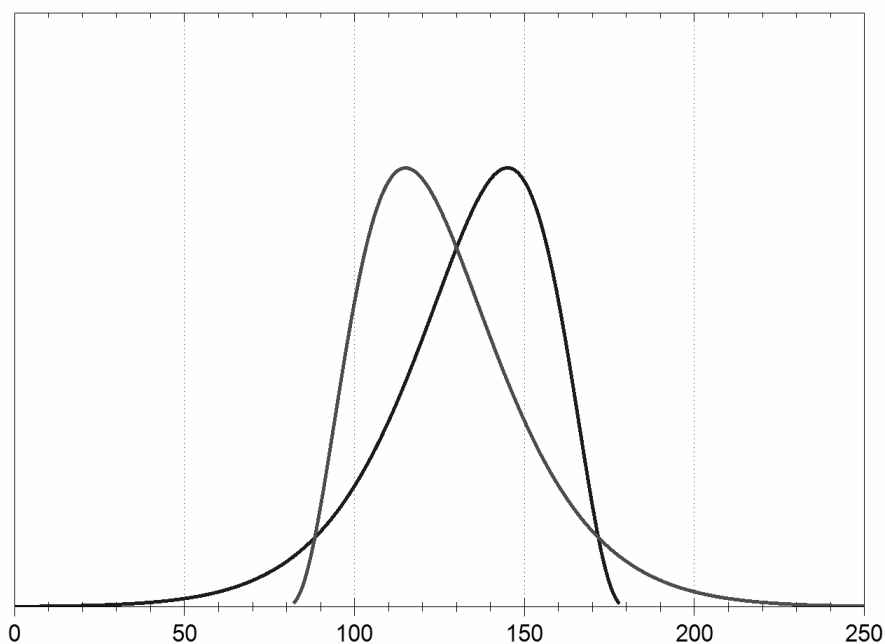


Figur 6-5 Forventningsverdiens følsomhet for feil mest sannsynlig anslag ( $s$ ) for ulike Alfa ved 10% standardavvik



Figur 6-6 Standardavvikets følsomhet for feil mest sannsynlig anslag ( $s$ ) for ulike Alfa

Det man også kan observere for figur 6-6 er at standardavviket er korrekt på to punkter. Standardavviket blir korrekt hvis man anslår mest sannsynlig verdi til å ligge i det korrekte prosentkvantilet, men også hvis man anslår verdien til å ligge i 100% minus det korrekte kvantilet, med andre ord en kurve som er helt lik i formen, men speilvendt, forhold til den korrekte som vist i figur 6-7. Hvis vi ser på formlene for standardavviket fra avsnitt 3.3 og definisjonen av en venstreskjev gammafordeling på side 35 i kapittel 3 ser vi lett hvorfor standardavviket her bli det samme.



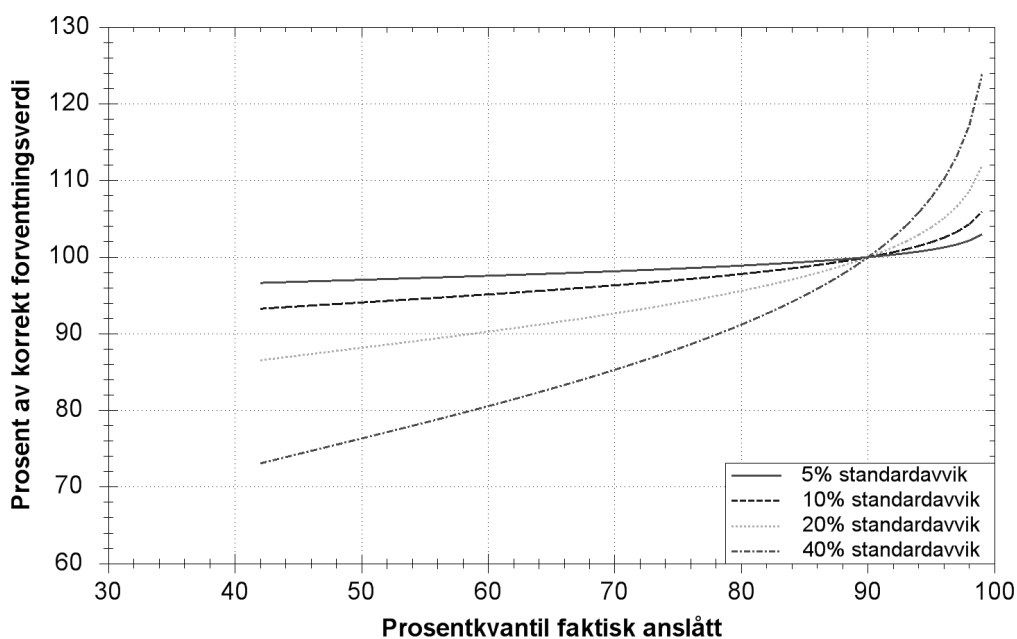
Figur 6-7 *Helt like i form, men speilvendt*

$\sigma = \sqrt{\alpha\beta^2}$  og vi definerer en venstreskjev gammafordeling til å ha negativ Beta.

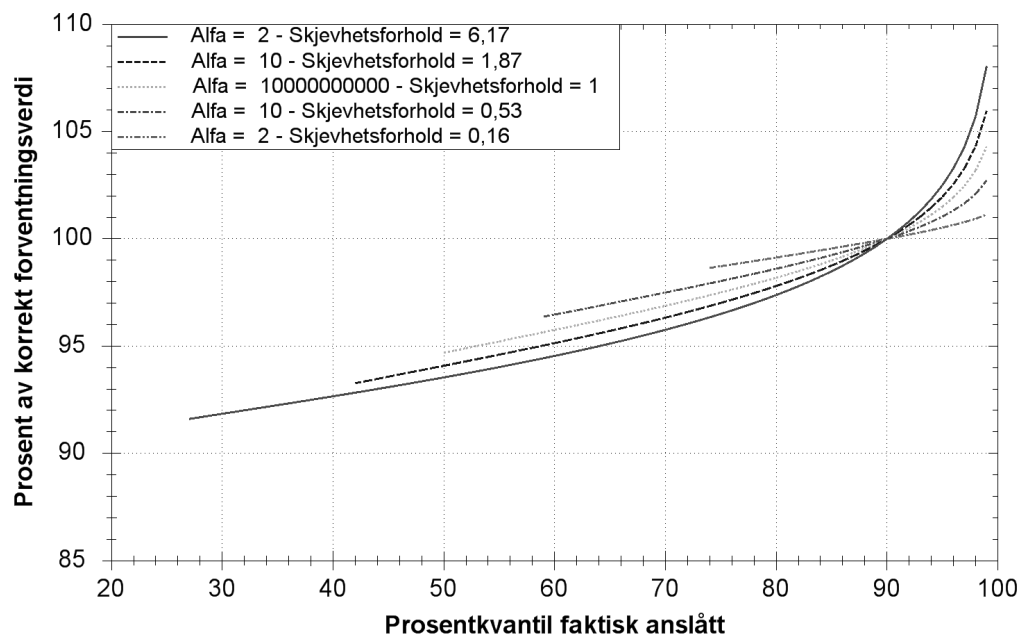
I og med at Beta kvadreres i formlene får ikke det negative fortegnet noe å si og standardavviket blir det samme. (Merk at det ikke blir slik for forventningsverdien for her kvadreres ikke Beta og forskyvningsparameteren som også spiller inn vil ikke være den samme.)

### 6.3 Følsomhet for feil øvre anslag ( $\emptyset$ )

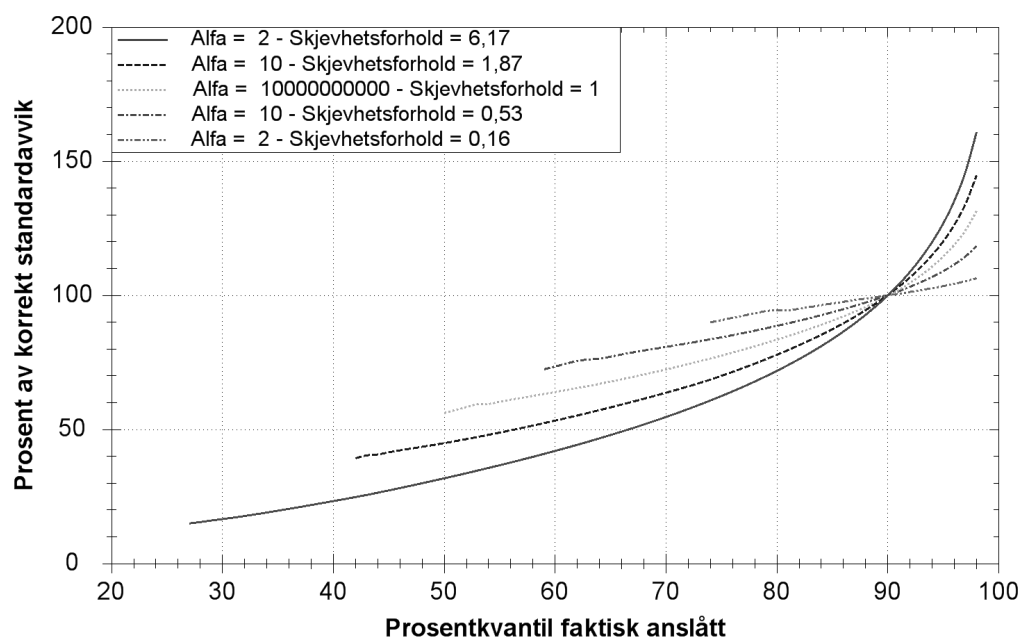
Figur 6-8, 6-9 og 6-10 viser det samme for feil øvre anslag som figurene som ble vist i avsnitt 6.1 for feil nedre anslag. Utslagene blir større for øvre anslag i og med at vi har tatt utgangspunkt i en høyreskjev fordeling med Alfa = 10. Hadde vi valgt en venstreskjev fordeling hadde resultatet blitt motsatt. Det vil si kurvene blir like i formen, men speilet om 50%-kvantilet og 100%-korrekt linjen.



Figur 6-8 Forventningsverdiens følsomhet for feil øvre anslag ( $\emptyset$ ) for ulike standardavvik ved Alfa = 10



Figur 6-9 *Forventningsverdiens følsomhet for feil øvre anslag ( $\theta$ ) for ulike Alfa ved 10% standardavvik*



Figur 6-10 *Standardavvikets følsomhet for feil øvre anslag ( $\theta$ ) for ulike Alfa*

## 6.4 Konklusjoner om følsomhet for feil i inngangsdata

Noe av det vi har sett her er ikke direkte overraskende. Jo større den relative usikkerheten er på den virkelige fordelingen, jo større blir konsekvensen av å anslå feil. Det som kanskje er mest interessant her er å se hvor liten konsekvensen av å anslå feil *mest sannsynlig verdi* er i forhold til å anslå ytterpunktene feil. Dett understreker viktigheten av å i en analysesammenheng å først bestemme disse før man går over til å bestemme den mest sannsynlige verdien.

Det vi også kan se er at feil anslag av inngangsverdiene fører til langt større feil i resultatene enn det som finnes av unøyaktigheter i Trinnsvis-formlene. Trinnsvis-formelen for forventningsverdien er såpass nøyaktig at enhver feil i inngangsdataene vil gi større utslag enn en eventuell feil som følge av formelen. Formlene for standardavviket er som sagt mer unøyaktig og feilen der er avhengig av hvor skjev fordelingen er. For de skjevhetsforholdene man typisk finner i praksis i usikkerhetsanalyser blir feilene på standardavviket større enn hva den potensielle formelfeilen er hvis vi plasser et anslag 1 prosentkvantil feil. For eksempel, hvis vi anslår 10%-kvantilet til å være verdien som faktisk er 11%-kvantilet, vil feilen på forventningsverdien være større enn en eventuell feil som følge av unøyaktighet i formlene. (Dette gjelder forutsatt at vi bruker en korrekt verdi for Theta.)

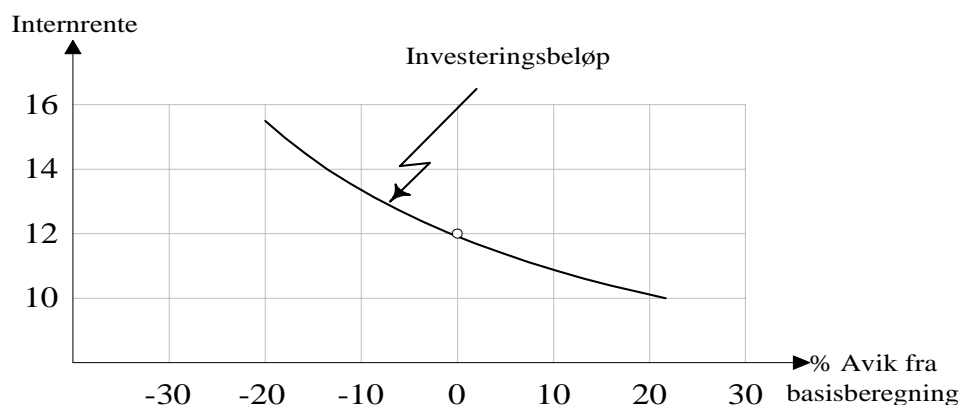
## 7. Noen spesialiteter

### 7.1 Følsomhetsanalyser

Et av resultatene fra en usikkerhetsanalyse er en oversikt over de mest usikre elementene; en rangering av elementene etter absolutt størrelse på spredningen (målt i standardavvik eller varians). Dette brukes til å prioritere hva vi bør sørge for å få mer og sikrere kunnskap om.

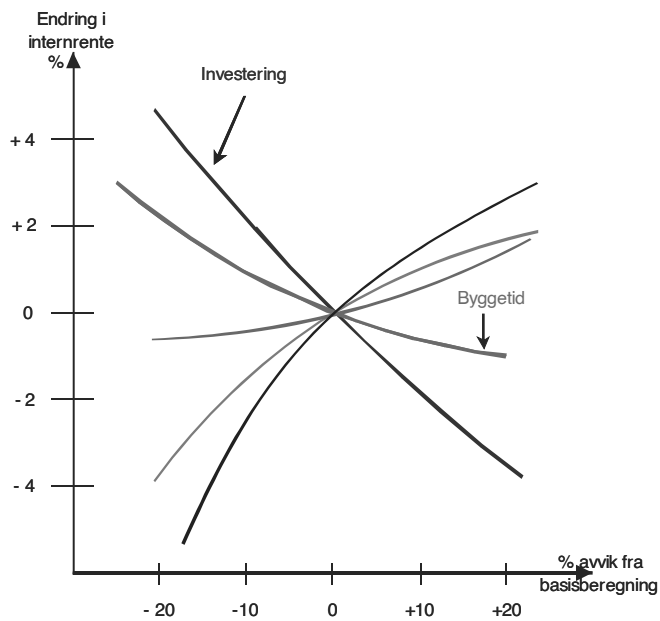
En annen viktig indikator på hvor vi bør rette innsats for å få mer kunnskap er i hvilken grad sluttresultatet er avhengig av variasjon i de enkelte elementer, eller sagt med andre ord; hvor følsomt er resultatet vi måler for variasjoner i inngangsparameterne?

For å få oversikt over dette benytter vi en såkalt følsomhetsanalyse. Vi gjør forutsetninger om de enkelte parametere og stipulerer for eksempel et forventet økonomisk resultat ut fra disse forutsetningene. Så endres en og en parameter nedover og oppover mens de andre holdes faste. Endringsskalaen er relativ i forhold til forutsatt størrelse på parameteren; for eksempel prosent ned og prosent opp. Figur 7-1 viser hvordan internrenten endrer seg som følge av relative endringer av investeringskostnad.



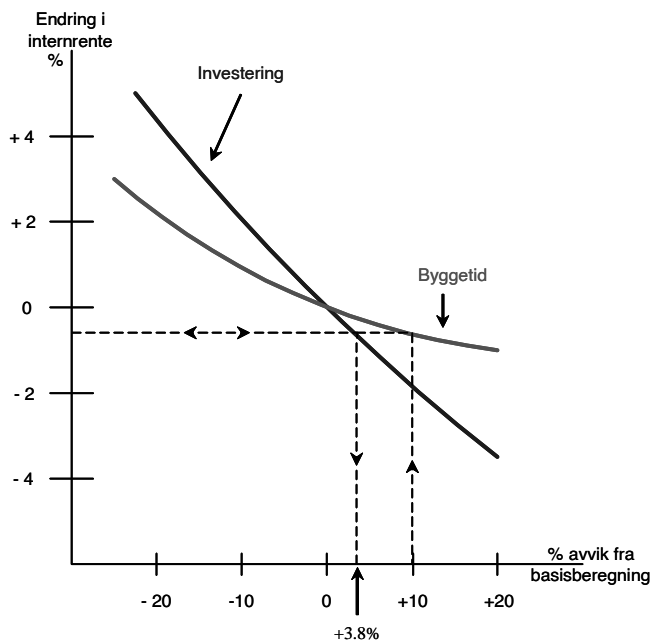
Figur 7-1 *Hvor følsom er internrenten for endringer i investeringskostnad? [Rolstadås]*

Figur 7-2 viser resultatet av en analyse av alle parametere. Av åpenbare grunner går denne framstillingen under navnet "Spiderdiagram". De parametere hvor endringer har størst virkning på resultatet; (disse sies, kanskje ikke helt semantisk korrekt, å ha størst følsomhet), og følgelig de det er naturlig å gi størst oppmerksomhet, har de bratteste kurvene.



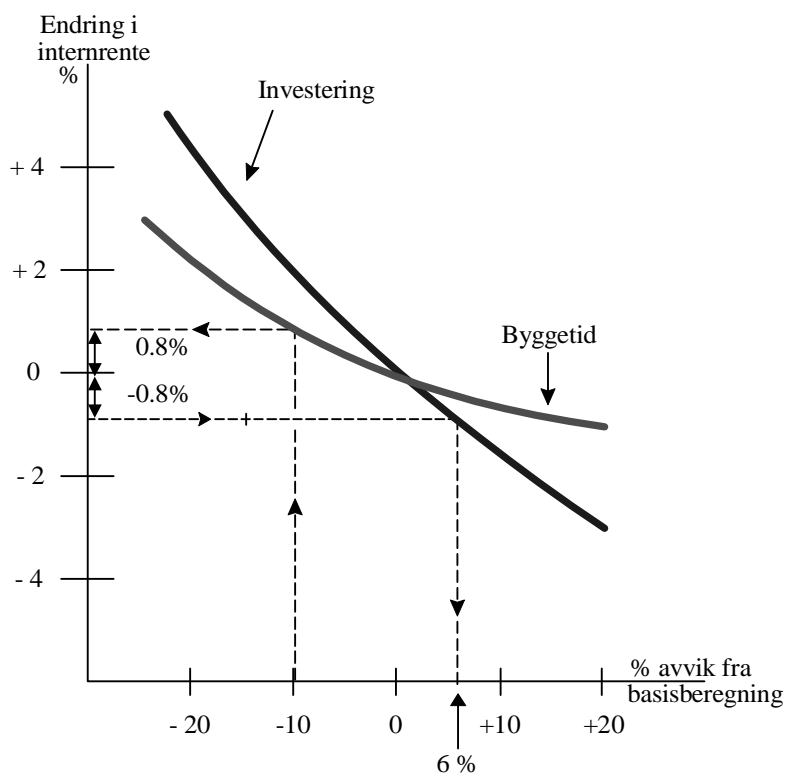
Figur 7-2 Eksempel på "spiderdiagram".

Resultatene fra en følsomhetsanalyse er gode og svært anvendelige som beslutningsstøtte i valg av løsninger, og også som styringssignaler i prosjektets gjennomføringsfase. Figur 7-3 og figur 7-4 er eksempler på det siste.



Figur 7-3 Hvilken merinvestering kan forsvares for å unngå 10% overskridelse av byggetid? [Rølstadås]





Figur 7-4 *Hvilken merinvestering kan tillates for å oppnå 10% kortere byggetid? [Rolstadås]*

Følsomhetsanalyser viser klart hvordan modellen er avhengig av den informasjonen som til enhver tid foreligger. Den viser hvor robust prosjektet er for endringer av de enkelte parametre, og kan således bidra til at det etableres forskjellige akseptnivåer for hvor usikre enkelte parametre får lov å være.

Følsomhetsanalyser er på denne måten nært knyttet til usikkerhetsstyring, som har som mål å kombinere resultater fra usikkerhetsanalysen med nødvendige tiltak.

## 7.2 Regresjonsanalyser

Dette kapitlet beskriver formålet med metoden regresjonsanalyse benyttet i forbindelse med usikkerhetsanalyser. For å visualisere og konkretisere forklaringene benyttes et eksempel.

### Generelt

En kan ofte være stilt ovenfor problemstillinger som involverer flere sett med variable, og der en ønsker å undersøke i hvilken grad det er sammenheng mellom variablene. En har observasjoner for hver variabel, og ønsker å modellere sammenhengen mellom variablene. Den statistiske utfordringen blir da å finne den beste fremstillingen av denne sammenhengen (Walpole, Myers and Myers, 1998). Situasjonen kan være slik at det er en avhengig variabel,  $Y$ , som er avhengig av en eller flere uavhengige variable (kalt regresjonsvariable). Videre i dette kapitlet vil vi se nærmere på en enkel lineær regresjon hvor kun en uavhengig variabel påvirker den avhengige variabelen. Et eksempel kan være sammenhengen mellom antall sykedager pr. år for ansatte i en bedrift og alder. En kan anta at det er mange parametere (variable) som kan påvirke antall sykedager, men en av dem antas å være alder. Det finnes observasjoner om sykefravær for hver person i for eksempel en bedrift, og en kan da studere sammenhengen mellom antall sykedager og alder for å prøve å se en sammenheng.

### Lineær regresjonsanalyse benyttet i sammenheng med usikkerhetsanalyser av prosjekter

Formålet med en regresjonsanalyse er i denne sammenheng å studere sammenhengen mellom to variable. Regresjonsanalyse kan benyttes til analyse av ulike faktorer i forbindelse med prosjekter og usikkerhet i prosjekter. En kan for eksempel tenke seg at en ønsker å analysere faktorer som påvirker prosjektkostnad for en portefølje av byggeprosjekter, og en antar at en av dem er byggeår. En kan ha et ønske om å finne ut hva som påvirker %-vis avvik mellom kostnadsoverslag og faktisk kostnad, og at en antar at prosjektets størrelse kan virke inn på dette. Det siste kan tenkes benyttet dersom en ønsker å danne seg et bilde av hva som er den "normale" kostnadsutviklingen, basert på erfaringstall, fra planstadiet til ferdigstillelse for et prosjekt av den aktuelle type og størrelse. Dette skal være til hjelp for å på forhånd utarbeide tiltak for å unngå kostnadsoverraskelser i prosjekter.

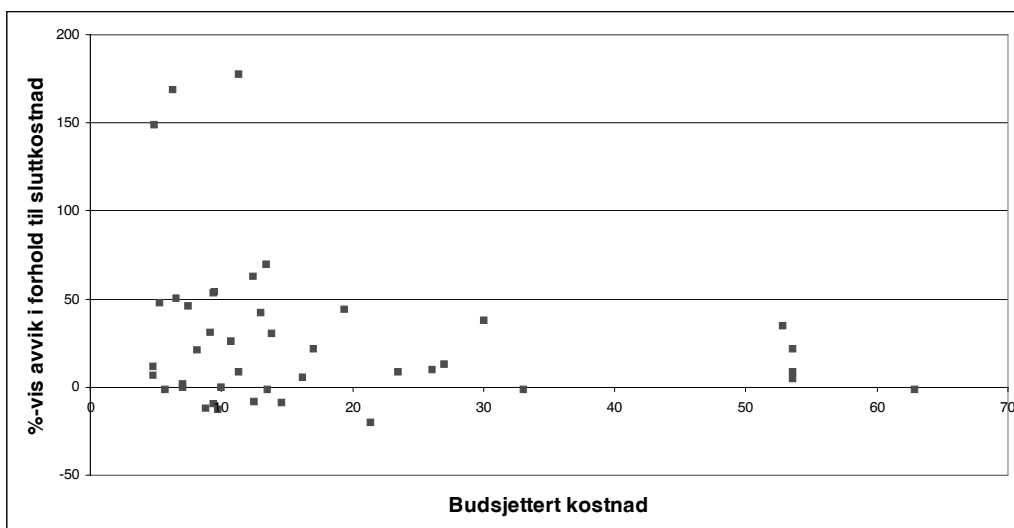
### Regresjonsanalyse

Noen sitter på planer for et prosjekt og ønsker å gjennomføre en regresjonsanalyse for sammenlignbare prosjekter man sitter med erfaringstall fra. En ønsker å undersøke om størrelse på prosjektet har betydning for avvik mellom faktisk kostnad og kostnadsoverslag i%. Erfaringstallene er vist i tabell 7.1. Tallene er hentet fra en empirisk undersøkelse av gjennomførte byggeprosjekter for en norsk kommune utført av Institutt for bygg, anlegg og transport.

Tabell 7.1 Erfaringsdata fra byggeprosjekter.

Kostnadsoverslag (MNOK)	Faktisk kostnad (MNOK)	Avvik %	Kostnadsoverslag (MNOK)	Faktisk kostnad (MNOK)	Avvik %
27	30,5	12,8	16,2	17,1	5,3
53,6	58,4	8,9	53,6	65,2	21,6
9,7	8,5	- 12,3	17	20,7	21,6
12,4	20,1	62,6	7,5	11	46,3
13,4	22,7	69,5	9,4	14,4	53,3
13	18,5	42	6,5	9,8	50,5
13,5	13,2	-1,6	9,1	12	31,3
30	41,3	37,8	9,5	14,7	54,3
10,7	13,5	25,9	23,5	25,5	8,8
12,5	11,4	- 8,4	6,3	16,9	168,9
52,8	71,2	34,8	13,8	18	30,5
8,1	9,8	20,9	4,9	12,3	148,9
11,3	12,3	8,5	5,7	5,6	- 1,3
7	7,1	1,6	5,3	7,8	47,8
19,4	27,9	44,0	53,6	56,2	4,8
26,1	28,6	9,6	9,4	8,5	-9,2
8,8	7,2	-11,9	4,8	5,1	6,7
4,8	5,4	11,6	62,9	62,3	- 1,1
14,6	13,3	- 8,9	33	32,6	-1,3
21,4	17,1	-20,0	11,3	31,4	177,6
10,0	9,9	-0,30	7,0	7,0	0

En plotting av kostnadsoverslag mot prosentvis avvik gir en fremstilling som i figur 7-5.



Figur 7-5 Plotting av data for prosjektets størrelse opp mot prosentvis avvik.

Det er mulig å regne ut en korrelasjonskoeffisient ( $r$ ) for å beskrive sammenhengen mellom de to variablene. Det søkes dermed her etter å finne sammenhengen (korrelasjonen) mellom utføre observasjoner av de to variable, å avdekke korrelasjon i et datasett i motsetning til å anslå korrelasjon mellom to parametere, som er omtalt i kapittel x. Størrelsen på  $r$  skal være mellom -1 og 1, der ytterpunktene beskriver en perfekt lineær sammenheng mellom de to variablene (at alle punktene ligger på linjen). Verdien -1 angir et perfekt lineær sammenheng

mellom variablene med negativ helning. Verdien 1 angir det samme, men med positiv helning. En kan si at verdier for korrelasjonskoeffisienten i nærheten av 1 i størrelse angir god korrelasjon, eller lineær sammenheng mellom X og Y. Dersom  $r=0$  er det ingen lineær korrelasjon mellom variablene. Korrelasjonskoeffisienten beregnes som følger (Walpole, Myers & Myers 1998):

$$r = \frac{S_{xy}}{\sqrt{(S_{xx} \cdot S_{yy})}}$$

$S_{xx}$  er summen av det kvadrerte avvik mellom den enkelte observasjon av X og middelerdien til  $X(\bar{X})$ .  $S_{yy}$  er det samme med hensyn på y.

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Differansen mellom den enkelte observasjon av X og  $\bar{X}$ , multiplisert med differansen mellom den tilhørende observasjonen av Y og  $\bar{Y}$ , sier noe om sammenhengen mellom størrelsen på X og størrelsen på Y for den enkelte observasjon. Om en gjør dette for hver av observasjonene, og summerer de multipliserte verdiene, får en  $S_{xy}$ .

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Korrelasjonskoeffisienten for dataene i tabell 1 er beregnet til å være - 0,21. Dette viser en sammenheng (regresjonslinje) med negativ helning. Kostnadsavvik for store prosjekt synes å være mindre i prosent enn for små prosjekt. Korrelasjonen er imidlertid ikke særlig sterk, i forhold til det at 0 angir ingen korrelasjon og 1 viser full korrelasjon.

Vi ønsker så å beskrive sammenhengen mellom variablene i form av regresjonsanalyse. Så fremt det er sammenheng mellom variablene som betraktes som lineær, kan den avhengige (Y) beskrives som en funksjon av den uavhengige (X). Prosentvis avvik for sluttkostnaden kan beskrives som en funksjon av størrelsen på kostnadsoverslaget. Den lineære funksjonen beskrives som følger:

$$Y = a + bX + e$$

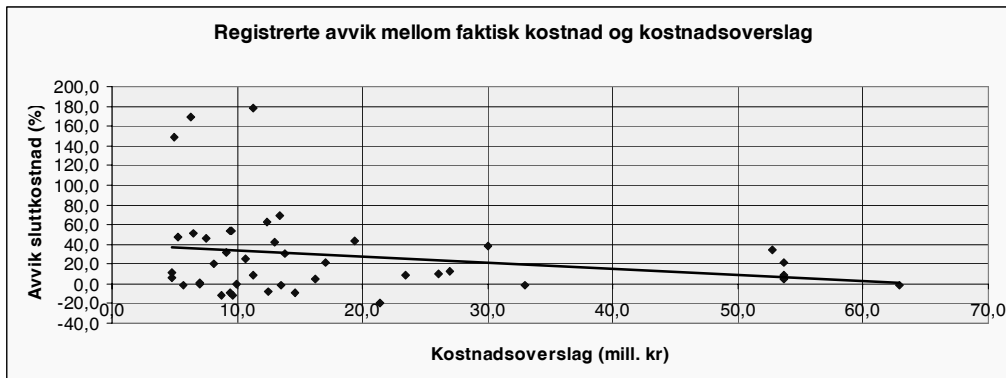
a er konstantleddet, som angir Y's verdi når  $x = 0$ .

b er stigningsforholdet til regresjonslinjen, og viser eksempelvis hvor mye den prosentvise kostnadsoverskridelsen stiger/synker med økende størrelse på prosjektet.

e: prediksjonsfeilen: antyder at regresjonslinjen ikke fanger opp den totale variasjonen i Y.

Den best tilpassede regresjonslinjen finner man ved å sette opp en ligning som minimerer den vertikale kvadrerte avstanden mellom de observerte og de predikerte verdiene. Hvert punkt i diagrammet (hver observasjon) har et visst vertikalt avvik fra linjen (avstanden opp/ ned til linjen). Dersom disse avvikene kvadreres og summeres får man en viss kvadratsum. Den beste linjen er den som gir den minste kvadratsummen. Når dette gjøres, minimeres også den totale summen av prediksjonsfeilene.

I figur 7-6 er det tegnet inn en regresjonslinje for datasettet i figur 7-5. Denne kan benyttes til å si noe om hva en kan forvente av størrelse på prosentvis kostnadsoverskridelse ved en gitt størrelse på prosjektet, forutsatt at kostnadsoverslaget har fremkommet på basis av faste rutiner og at prosjektene gjennomføres etter standard metode. For et prosjekt av størrelse 30 mill. kr. Kan en forvente en kostnadsoverskridelse på ca. 20%. I dette tilfellet kan en si at datagrunnlaget ikke er godt nok til å gjøre slike vurderinger. Det er ikke mange observasjoner på prosjekter i området 30 mill.kr. I tillegg er spredningen på observasjonene så stor at slike konklusjoner ikke med stor sikkerhet kan tas basert på det eksisterende datagrunnlaget.



Figur 7-6 Regresjonslinje for eksemplet.

Som nevnt tidligere er korrelasjonskoeffisienten til datasettet beregnet til å være  $-0,21$ . Som nevnt tidligere så viser dette at helningen på regresjonskurven er negativ, noe en kan se av figur 7-6. Sammenhengen er slik at negativ korrelasjonskoeffisient gir negativ helning på regresjonslinjen. Dette kommer av at  $S_{xy}$  er telleren både i korrelasjonskoeffisienten og i beregningen av stigningen på regresjonslinjen, og nevneren i korrelasjonskoeffisienten og stigningen på regresjonslinjen er positive størrelser. Korrelasjonskoeffisienten viser styrken på korrelasjonen, og det er ingen direkte sammenheng mellom helningen på regresjonslinjen og størrelsen på korrelasjonskoeffisienten. Når helningen på regresjonslinjen er 0, er imidlertid også korrelasjonen 0, uavhengig av om alle punkter ligger på en rett linje.

Tidligere i dette kapitlet ble korrelasjonskoeffisient  $r$  forklart og beregnet.  $r^2$  er et tall som forklarer andel av den totale variasjon i variabel  $Y$  som kan forklares med en lineær sammenheng med verdiene til variabel  $X$  (Walpole, Myers and Myers, 1998).  $r^2$  angir dermed andelen av totalvariasjon som regresjonsmodellen forklarer. I eksemplet er  $r^2 = 0,04$ . Dette betyr at 4% av den totale variasjonen til  $Y$  kan forklares ved en lineær sammenheng med variabelen  $X$ . En god tommelfingerregel er å si at dersom denne prosentatsen synker under 0,5, blir samvariasjonen tvilsom (Jessen 2000). I eksemplet kan en si at størrelsen på kostnadsover-

slaget har veldig liten betydning på prosentvis avvik på sluttkostnaden ut fra de observasjonene som foreligger. Det kan forutsatt være at det er en sammenheng mellom parametrene, men som ikke er lineær. Det er imidlertid vanskelig å se en slik annen sammenheng ut fra figur 7-5.

Institutt for bygg, anlegg og transport har utviklet et regneark for regresjonsanalyser av Kostnadsoverslag mot prosentvis avvik for sluttkostnaden. Verktøyet er beskrevet i vedlegg B, og kan lastes ned fra hjemmesidene til Concept, [www.concept.ntnu.no](http://www.concept.ntnu.no).

### **Bruk av resultatene**

Resultatene fra analysene kan tenkes benyttet til å utarbeide metoder for å redusere sprik mellom planlagt og virkelig kostnad på planlagt gjennomførte prosjekter. Dette skal være til hjelp for på forhånd å utarbeide tiltak for å unngå kostnadsoverraskelser i prosjektene. For eksempel kan bildet vise at det kan være aktuelt å legge til noe for å fange opp tendenser til overskridelser. I en organisasjon som sitter med en portefølje av prosjekter, og hvor det stadig kommer til nye prosjekter, vil dette kunne vise hva en kan forvente av prosentvise overskridelser for et prosjekt av en viss størrelse, under forutsetning av at kostnadsoverslaget har fremkommet på basis av faste rutiner, og prosjektet gjennomføres etter standard metode. Vi ser av resultatene fra eksemplene at en ikke har en indikasjon på at størrelse på prosjektet har noen stor betydning for prosentvis kostnadsoverskridelse. I eksemplet kan det se ut som at det er en trend at prosentvise kostnadsoverskridelser for store prosjekt er mindre enn for små prosjekt, selv om regresjonsanalysen ikke bekrefter dette i særlig grad. I figur 7-5 vises imidlertid en klar trend til at prosjektene havner over budsjett.

### 7.3 Stokastisk tidstillegg

Hvis starten på en aktivitet er avhengig av at flere aktiviteter som gjennomføres samtidig blir ferdige, vil forventet starttidspunkt være senere enn forventet sluttidspunkt på selv den lengste av de foregående aktivitetene. Dette gapet kalles stokastisk tidstillegg og kommer av at det alltid vil være en mulighet for at en av de parallelle aktivitetene som har kortere forventet tid faktisk kan bli så forsinket at den blir kritisk. Den vanligste metoden for å håndtere dette rent analyseteknisk er ved hjelp av simuleringer. Det er ikke vanskelig å forestille seg den situasjonen at vi setter opp en simuleringsmodell for et sett av parallelle aktiviteter som alle har en stokastisk gjennomføringstid angitt som en sannsynlighetsfordeling. Simuleringen vil da bestå i et tilfeldig trekk fra hver aktivitet, hvorpå det seneste tidspunktet velges som tidligst start for etterfølgende aktivitet. Denne trekkserien gjentas for eksempel tusen ganger, og vi får en sannsynlighetsfordeling for når etterfølgende aktivitet kan starte. Fordi vi hele tiden har gjort det seneste tidspunktet gjeldende vil forventningsverdien i denne fordelingen bli senere enn forventet sluttidspunkt for selv den seneste aktiviteten.

For å finne størrelsen på det stokastiske tidstillegget er, som vi tidligere har vært inne på, simuleringer eller simuleringsbaserte metoder som for eksempel Møllers metode og Suksessiv tidplanlegging (Klakegg, 1994) å anbefale.

Beregninger ved hjelp av statistiske regneregler blir fort svært kompliserte og tidkrevende. Teorien bak er imidlertid enkel. Den dekkes under det som populært kalles ekstremstatistikk, og regnereglene er av statistikerne betegnet som “vakre”.

Hvis vi har  $n$  parallelle aktiviteter hvor gjennomføringstiden for hver enkelt er en stokastisk variabel  $T$  og hvor tetthetsfunksjonen er  $f(t)$ , så er den kumulative fordelingsfunksjonen

$$F(t) = P(T \leq t) = \int_{t_{min}}^t f(t) dt$$

Den kumulative fordelingen for når alle aktivitetene er ferdig er

$$F_{maks}(t) = [P(T \leq t)]^n$$

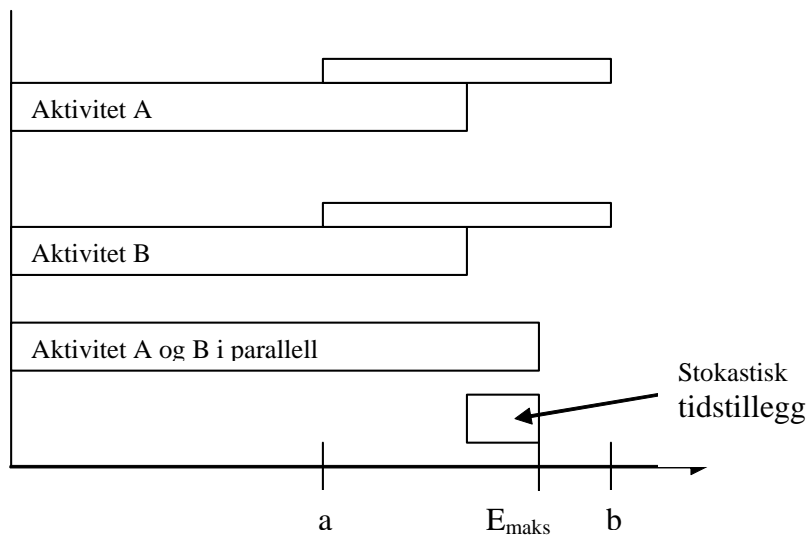
og tilsvarende tetthetsfunksjon er

$$f_{maks}(t) = \frac{d(F_{maks}(t))}{dt}$$

Forventet tid fra start og fram til alle aktivitetene er ferdig er da

$$E_{maks} = \int_{t_{min}}^{t_{maks}} t(f_{maks}(t))dt$$

For å vise den teoretisk bakgrunnen vil vi her gi et (svært) enkelt eksempel med to parallelle aktiviteter med samme sannsynlighetsfordeling og samme forventningsverdi. Sluttidspunktet for hver av aktivitetene er gitt som en uniform fordeling mellom tidspunkt a ( $t_{min}$ ) og tidspunkt b ( $t_{maks}$ ).



Figur 7-7 *Stokastisk tidstillegg*

Tetthetsfunksjonen for sluttidspunkt for aktiviteten er:

$$f(t) = \frac{1}{b-a} \text{ for } t \in [a, b]$$

Kumulativ tetthet blir da:

$$F(t) = P(T \leq t) = \frac{1}{b-a} \int_a^t dx = \frac{1}{b-a}(t-a)$$

For å lette fremstillingen velger vi å sette  $a = 70$  og  $b = 130$ . Forventet tid for hver av aktivitetene blir da  $E_A = 100$ .

Kumulativ tetthet for sluttiden for begge aktivitetene blir:



$$F_{maks}(t) = \left( \frac{1}{130 - 70}(t - 70) \right)^2 = \frac{1}{3600}(t - 70)^2$$

Tetthetsfunksjoner for  $T_{maks}$  blir da:

$$f_{maks}(t) = \frac{d(F_{maks}(t))}{dt} = \frac{2}{3600}(t - 70)$$

Forventet sluttidspunkt for begge aktivitetene blir:

$$E_{maks} = \frac{2}{3600} \int_{70}^{130} t(t - 70) dt = \frac{2}{3600} \int_{70}^{130} (t^2 - 70t) dt = \frac{2}{3600} \left[ \frac{1}{3}t^3 - \frac{70}{2}t^2 \right]_{70}^{130} = 110$$

Stokastisk tidstillegg er  $E_{maks} - E_A$ , og blir i dette tilfellet, som vi ser, 10.

Det finnes noen fordelinger som er slik at det er overkommelig å regne seg fram til forventet maksimaltid ut fra modellen. Eksempel på slike fordelinger er Gumbel, Freshe og "Weibull for maksimum". Disse fordelingene sies å være lukket under maksimum. Det vil si at hvis  $X_1$  og  $X_2$  er for eksempel Gumbel-fordelt så er  $\max(X_1, X_2)$  også Gumbelfordelt

# Litteratur

- Austeng, Kjell & Hugsted, Reidar (1995): *Trinnvis kalkulasjon*. Institutt for bygg- og anleggsteknikk, NTH.
- Dougherty, Edward R. (1990): *Probability and Statistics for the Engineering, Computing and Physical Sciences*. Prentice Hall International Editions. ISBN 0-13-715913-7
- Husby, Kilde, Klakegg, Torp, Berntsen og Samset (1999): *Usikkerhet som gevinst - Styring av usikkerhet i prosjekter*. Prosjektstyring år 2000, Trondheim, .
- Høyland, Arnljot (1979): *Sannsynlighetsregning og statistiske metodelære*. Tapir
- Jessen, Svein Arne (2000): *Prosjektarbeid og konsekvensanalyser*. Universitetsforlaget, Norge.
- Klakegg, Ole Jonny. (1993): *Trinnvis-prosessen*. Institutt for bygg og anleggsteknikk.
- Klakegg, Ole Jonny. (1994): *Tidsplanlegging under usikkerhet*. Institutt for bygg og anleggsteknikk.
- Kreyszig, Erwin (1999): *Advanced engineering mathematics*, 8th edition. John Wiley & Sons. ISBN 0-471-15496-2
- Law, Averill M. & Kelton, David W. (2000): *Simulation Modelling and Analysis, third edition*. McGraw-Hill Higher Education. ISBN 0-07-059292-6
- Lichtenberg, Steen. (1978): *Prosjekt planlegging - i en foranderlig verden*. Polyteknisk Forlag Lyngby.
- Lichtenberg, Steen (2000): *Proactive Management of Uncertainty using the Successive Principle*. Polyteknisk Press, Danmark.
- Lindgren, G., Leadbetter, M.R. and Rootzén H. (1983): *Extremes and related properties of stationary sequences and processes*. Springer, New York
- Palisade Corporation. (2002): *A Concise Summary of @RISK Probability Distribution Functions*. Palisade Corporation.
- Reibnitz, Ute von. (1988): *Scenario techniques*. McGraw Hill.
- Ringland, Gill. (2002): *Scenarios in Business*. John Wiley & Sons, Chichester, Storbritannia.
- Rolstadås, Asbjørn. (2001): *Praktisk Prosjektstyring*. Tapir Akademisk Forlag . ISBN 82-519-1652-6.
- Rolstadås, Asbjørn. (1997): *Praktisk Prosjektstyring*. Tapir, Trondheim.

Spiegel, Murray R. & Stephens, Larry J. (1999): *Schaum's Outline of Theory and Problems of Statistics Third Edition*. McGraw-Hill. ISBN 0-07-060281-6

Walpole, Ronald E, Myers, Raymond H. & Myers, Sharon L. (1998): *Probability and Statistics for Engineers and Scientists*. Prentice Hall. ISBN 0-13-095246-X

Øien, Klakegg, Hokstad og Rosness. (1996): *Inngangsdata til LCP og bruk av ekspertvurderinger*. Håndbok for gjennomføring av ekspertvurderinger, Prosjektstyring år 2000, Trondheim.

# Vedlegg-A Gjennomsnitt - mål på sentraltendens

Et *gjennomsnitt* er en verdi som er typisk, eller representativ, for et sett med data. Siden slike typiske verdier tenderer til å ligge sentralt innen et sett med data ordnet etter størrelse, gjennomsnitt kalles også mål på *sentraltendens*.

Flere ulike typer gjennomsnitt kan defineres. De mest vanlig er den aritmetiske middelværdien (arithmetic mean), medianen (median) og mest sannsynlig verdi (mode)

## **Aritmetisk middelværdi = forventningsverdi (Arithmetic Mean = Expected Value)**

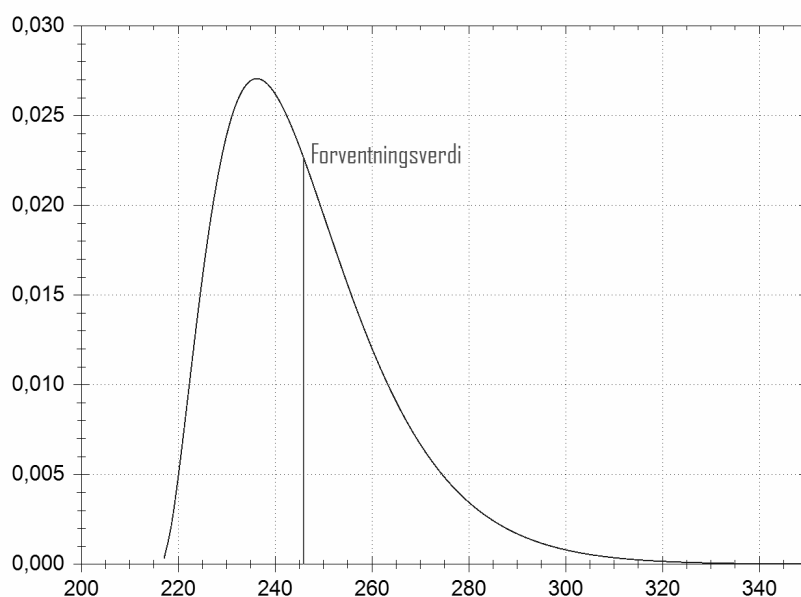
Den aritmetiske middelværdien til et sett med verdier finner man ved å legge sammen alle verdiene og dele det på antall verdier. La oss si at vi har fem forskere som er henholdsvis 170, 170, 180, 190 og 220 cm høye. Den aritmetiske middelværdien for disse fem blir da

$$\mu = \frac{170 + 170 + 180 + 190 + 220}{5} = 186$$

Men det vi har satt opp her er statistikk, behandling av innsamlet tallmateriale. Hva med sannsynlighetsregning? OK, la oss så snu litt på flisa og gjøre statistikk om til sannsynlighet. La oss anta at de fem forskerne er høydemessig representativ for alle forskere i hele verden. Det er da henholdsvis 40, 20, 20 og 20 prosent sjanse for at en forsker er henholdsvis 170, 180, 190 og 210 cm høy. Ut ifra dette kan vi regne ut den forventningsverdien for en forskers høyde.

$$E(x) = 170 \times 40\% + 180 \times 20\% + 190 \times 20\% + 220 \times 20\% = 186$$

Med andre ord den aritmetiske middelværdien og forventningsverdien er det samme. Forskjellen er at det ene begrepet bruker man for statistikk og det andre begrepet bruker man for sannsynlighetsregning. Forventningsverdien kan ses på som summen av alle tenkelige utfall, hvor hver av dem er vektet med sine respektive sannsynligheter.



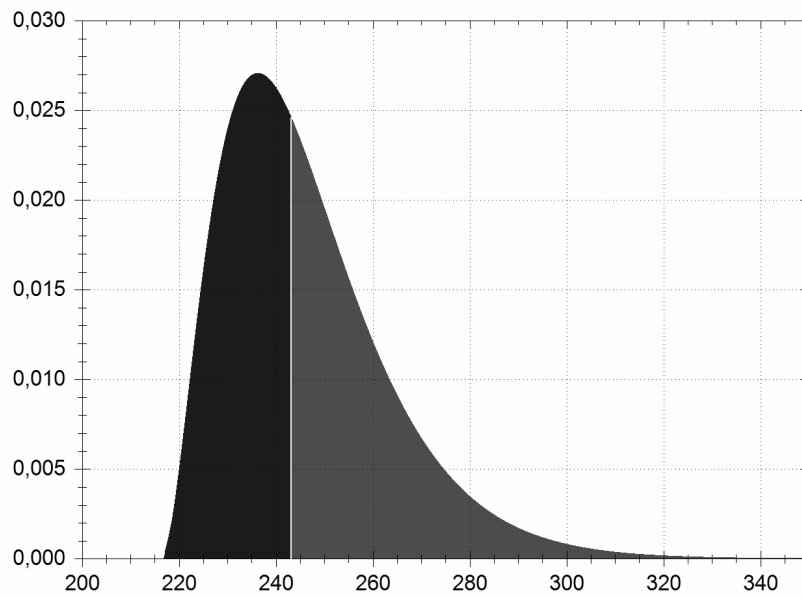
Figur A-1 *Forventningsverdi*

Hvis vi ser på kurven av en sannsynlighetsfordeling så er forventningsverdien lik x-verdien av kurvens tyngdepunkt som tegnet inn på figur A-1.

## Medianen

Medianen er midtverdien i et sett av sorterte verdier. Hvis vi stiller våre fem forskere opp på en rekke, sortert i stigende rekkefølge etter høyden, er midterstemann 180 cm høy. Ergo er medianen for høyden av forskerne 180 cm.

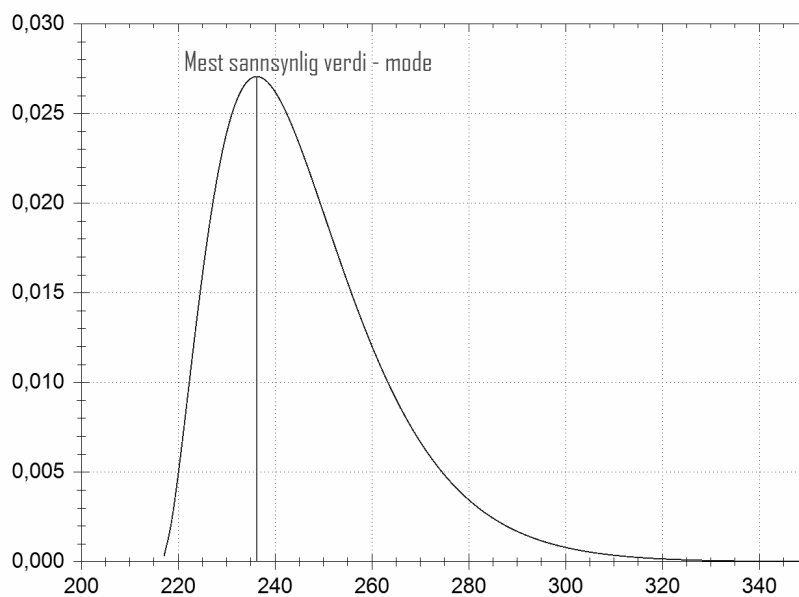
For sannsynlighetsregning blir det samme. La oss si at våre fem forskere avholder et lotteri, det er da like sannsynlig at en person som er over 180 cm høy skal vinne lotteriet som en som er under 180 cm høy. Medianen er altså den verdien hvor det er like stor sannsynlighet at man får en større eller mindre verdi. For en kontinuerlig sannsynlighetsfordeling vil det si det punktet på kurven der halvparten av arealet til tetthetskurven ligger til venstre og den andre halvparten av arealet ligger til høyre som vist i figur A-2. Det vil med andre ord si at medianen er synonymt med 50%-kvantilet.



**Figur A-2**    *Mest sannsynlig verdi - Mode*

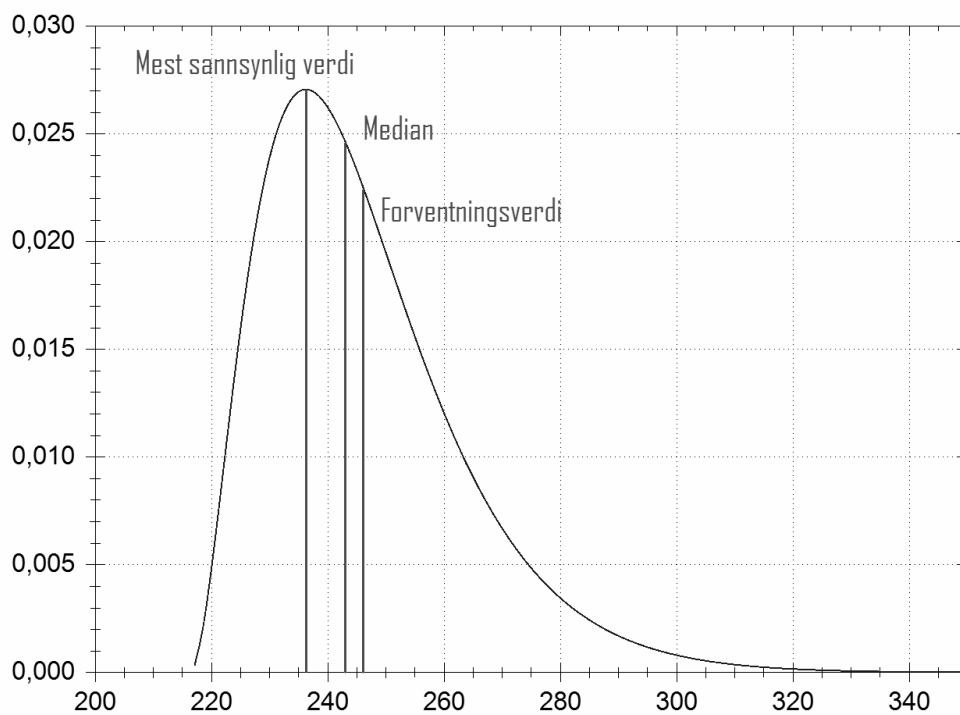
Den mest sannsynlige verdi eller mode er den verdien som forekommer oftest i et utvalg. Når det gjelder våre forskere har vi to som er 170 cm høye, mens alle de andre har forskjellige verdier. Det vil at den moden for forskeres høyde er 170 cm. Hvis vi igjen snur litt på det og lager sannsynlighet ut av statistikk, ved å anta at utvalget er representativt for alle verdens forskere, kan vi si at hvis vi møter en tilfeldig forsker på gaten så er det større sannsynlighet for at han er 170 cm høy en hvilken som helst annen høyde. Og hvis vi drar på en forskerkongress vil vi sannsynligvis se at det er en opphopning av 170 cm høye forskere.

Hvis vi ser på kurver for sannsynlighetsfordelinger er mest sannsynlige verdi alltid kurvens toppunkt som vist i figur A-3.

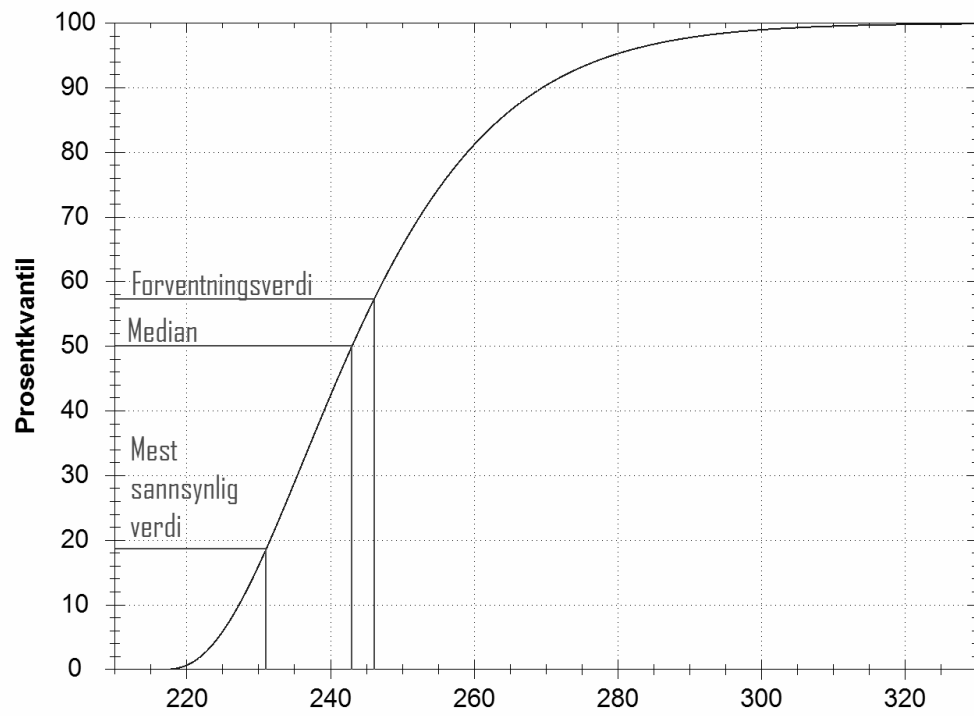


Figur A-3 *Sammenheng mellom de tre målene for sentral tendens*

Figur A-4 viser den relative posisjonen for mest sannsynlig verdi, median og forventningsverdi ved høyreskjeve kurver. For venstreskjeve kurver vil rekkefølgen bli motsatt og for symmetriske kurver vil alle tre være sammenfallende. Figur A-5 viser s-kurven for den samme fordelingen



Figur A-4 *Relativ posisjon på høyreskjev fordeling*



Figur A-5 *Relativ posisjon på en s-kurve for en høyreskjev fordeling*



# Vedlegg-B Verktøy for regresjonsanalyser

Verktøyet forklares best ved å bruke et eksempel til støtte. Vi tenker oss at noen sitter på planer for et vegprosjekt med et kostnadsoverslag på 164 mill. kr, og ønsker å gjennomføre en regresjonsanalyse for sammenlignbare prosjekter man sitter med erfaringstall fra. I utgangspunktet sitter man kanskje med kostnadsdata for prosjekter over mange år og med stor spredning i kostnadsoverslagene. Det kan da være tjenlig å gruppere disse prosjektene mht tidsperioder og kostnadsintervaller. Som et eksempel kan vi tenke oss at man sitter inne med kostnadsdata for prosjekter som er avsluttet i tidsperioden 1990 til 2005. Man mener imidlertid at det er utviklingen de senere år som gir det beste bildet av situasjonen, og ønsker å analysere trender med utgangspunkt i prosjekter avsluttet mellom 2000 og 2005 (femårsperiode). I tillegg bør prosjektene være av omtrent samme størrelsesorden som det prosjektet vi nå planlegger å gjennomføre, derfor ser vi kun på prosjekter med opprinnelig kostnadsoverslag mellom 100 og 200 mill. kr. Antall prosjekter det foreligger erfaringstall på innenfor dette intervallet er 15 stk.

Ut fra dette skrives det inn følgende tall i cellene for *Årstall/tidsperiode*, *Kostnadsintervall* og *Antall prosjekter* (jmf. figur B-1), deretter klikkes det på knappen *Generer tabell og diagram* (velg "Ja" på dialogboksen med spørsmål om du ønsker å endre tabellen). Tabellen for prosjektdata genereres:

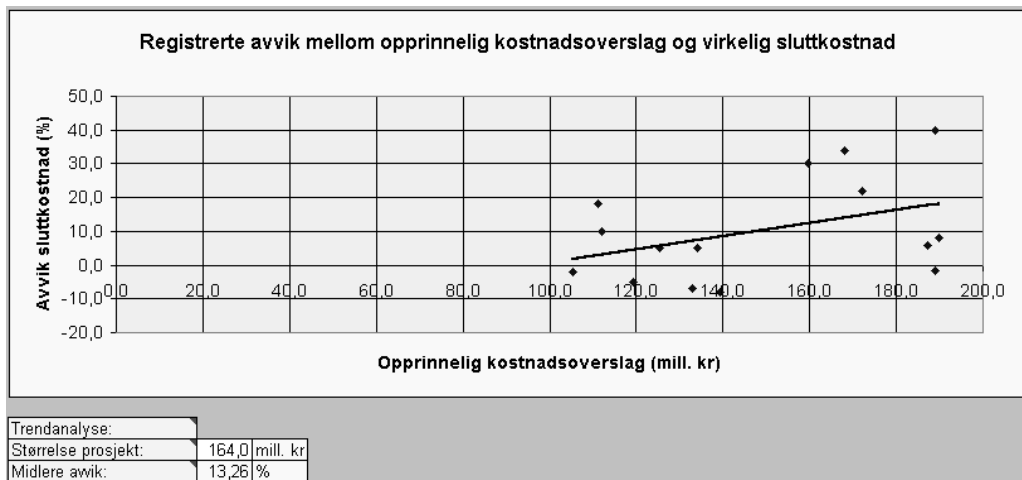
Regresjonsanalyse av gjennomførte vegprosjekter															
Analyse av avvik mellom opprinnelig kostnadsoverslag og endelig sluttkostnad															
Årstall/tidsperiode:	1994	til	1998												
Kostnadsintervall:	100	til	200	mill. kr											
Antall prosjekter =	15 stk			Generer tabell og diagram											
				Prosjektdata skrives inn som:											
				<input checked="" type="radio"/> Sluttkostnad <input type="radio"/> %-vis avvik											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Endelig sluttkostnad:	103,2	131,1	123,2	113,2	131,8	123,7	140,7	128,2	207,6	225,1	210,0	198,5	264,5	186,1	204,9
Oppr. kostnadsoverslag:	105,3	111,1	112,0	119,2	125,5	133,0	134,0	139,3	159,7	168,0	172,1	187,3	188,9	188,9	189,7
Avvik sluttkostnad (%):	-2,0	18,0	10,0	-5,0	5,0	-7,0	5,0	-8,0	30,0	34,0	22,0	6,0	40,0	-1,5	8,0

Figur B-1 Forutsetninger og prosjektdata

## Prosjektdata

Man sitter i dette eksempelet på kostnadsdata i form av *opprinnelig kostnadsoverslag* og *endelig sluttkostnad*, derfor velges "Sluttkostnad" i alternativvalget over tabellen (alternativt hadde man sittet på data for "%-vis avvik" og fått generert endelig sluttkostnad). Kostnadsdata for *opprinnelig kostnadsoverslag* og *endelig sluttkostnad* skrives inn, og er i dette tilfellet slik som vist i figur B-1.

Etter hvert som kostnadsdataene skrives inn, plottes *datasettet* automatisk inn i diagrammet under tabellen, med tilhørende regresjonslinje, jmf. figur 2.



Figur B-2 Plottet datasett med tilhørende trendlinje

Som nevnt sitter man med planer for et prosjekt med kostnadsestimat på 164 mill. kr. Det gjennomsnittlige avvik for et slik prosjekt, med utgangspunkt i de erfaringstall man sitter inne med, finnes ved å skrive inn 164 i cellen *Størrelse prosjekt* under diagrammet. Regnearket finner da avviket mellom x-aksen og trendlinjen ved x-verdien 164 mill. kr. Som vi ser er det et gjennomsnittlig avvik på vel 13% i dette tilfellet.

## Analyse av intervaller

Vi kan dele kostnadsintervallet opp i ytterligere delintervaller. Det finnes ingen eksakte krav til hvor mange delintervaller man bør dele opp i, det vil variere og må vurderes spesielt, evt. må man prøve seg frem. Vi forsøker her å dele opp fint, og velger 10 stk delintervall. Skriver derfor inn 10 i cellen *Antall delintervaller* (jmf. fig. B-3), trykker "Enter" og klikker på knappen "Generer avviksanalyse".

Beregning av **standardavvik** ut fra at hovedintervallet de opprinnelige kostnadsestimatene ligger innenfor deles inn i **delintervaller**. Standardavvikene innenfor hvert delintervall beregnes ut fra at alle verdier og avvik som ligger i delintervallet samles i to nye verdier lik delintervallets vektete middelveidier for verdi og avvik.

Avviksanalyse:	
Antall delintervaller:	10 stk.
Generer avviksanalyse	

	Prosjektverdi			Prosjektavvik	
	Ant. proj.	Sum verdi	middel verdi	middel verdi	Std. avvik
Delintervall 1:	fra o.m.	100,0	t.o.m.	110,0	mill. kr
	1	105,3	105,3	-2,0	0,0
Delintervall 2:	fra	110,0	t.o.m.	120,0	mill. kr
	3	342,3	114,1	7,7	11,7
Delintervall 3:	fra	120,0	t.o.m.	130,0	mill. kr
	1	125,5	125,5	5,0	0,0
Delintervall 4:	fra	130,0	t.o.m.	140,0	mill. kr
	3	406,3	135,4	-3,3	7,2
Delintervall 5:	fra	140,0	t.o.m.	150,0	mill. kr
	0	0,0	0,0	0,0	0,0
Delintervall 6:	fra	150,0	t.o.m.	160,0	mill. kr
	1	159,7	159,7	30,0	0,0
Delintervall 7:	fra	160,0	t.o.m.	170,0	mill. kr
	1	168,0	168,0	34,0	0,0
Delintervall 8:	fra	170,0	t.o.m.	180,0	mill. kr
	1	172,1	172,1	22,0	0,0
Delintervall 9:	fra	180,0	t.o.m.	190,0	mill. kr
	4	754,8	188,7	13,1	18,4
Delintervall 10:	fra	190,0	t.o.m.	200,0	mill. kr
	0	0,0	0,0	0,0	0,0

Figur B-3 Analyse ved inndeling av erfaringsmaterialet i 10 delintervaller

Basert på statistiske gjennomsnittsverdier vil da gjennomsnittlig avvik mellom opprinnelig kostnadsoverslag og sluttkostnad og spredningen på dette innenfor hvert delintervall beregnes.

Vi ser at det ikke finnes prosjekter med erfaringstall innen delintervall 5 og 10, og i delintervallene 1, 3, 6, 7 og 8 er det kun ett prosjekt i hver. Det betyr at det ikke er grunnlag for å beregne avviksanalyser innenfor disse delintervallene, det må minst eksistere 2 prosjekter for at en avviksanalyse skal gi noen mening. De eneste interessante delintervallene er altså 2, 4 og 9, men også her er det egentlig for lite prosjekter til at en avviksanalyse kan sies å være noenlunde pålitelig. Statistisk behandling av tall blir sikrere jo større datasettene er. Derfor bør en i dette tilfellet prøve å dele inn i færre delintervaller, vi prøver med kun 2 stk:

Avviksanalyse:					Generer avviksanalyse						
Antall delintervaller:		2 stk.									
					Prosjektverdi			Prosjektavvik			
					Ant. prosj.	Sum verdi	middel verdi	middel verdi	Std. avvik		
Delintervall 1:		fra o.m.	100,0	t.o.m.	150,0	mill. kr	8	979,4	122,4	2,0	9,1
Delintervall 2:		fra	150,0	t.o.m.	200,0	mill. kr	7	1254,6	179,2	19,8	15,8

**Figur B-4** Regresjonsanalyse med 2 delintervaller

Vi ser at vi får hhv 8 og 7 prosjekter innenfor hvert delintervall, og det er et mer fornuftig antall for å få en avviksanalyse som er mer til å stole på. Hvert delintervall strekker seg imidlertid over et større spenn (50 mill. kr).

# Vedlegg-C Hvordan finne Alfa, Beta og forskyvning fra trippelanslag

For å kunne benytte algoritmene som er oppgitt her må man tilgang på en funksjon for å regne ut den inverse kumulative tetthetsfunksjonen til en gammafordeling, slik som funksjonen GAMMAINV i Excel. Denne er dog ikke særlig egnet den ikke klarer verdier for alfa større enn ca 300. Eller finnes kildekode til funksjoner som regner ut den inverse kumulative tetthetsfunksjonen til en gammafordeling fritt tilgjengelig på internett.

Denne algoritmen klare å regne ut alfa for skjevhetsforhold opp til litt over 13 000 og takler i tillegg spesialtilfellet der  $n = s$  eller  $s = \emptyset$ . Algoritmen takler ikke tilfellet der  $n = s = \emptyset$  da dette tilfellet ikke gir noen mening for en gammafordeling. Algoritmen forutsetter en høyreskjev kurve. Hvis man har en venstreskjev kurve må tallene transformeres først. Se "Transformasjon av trippelanslag" på side 36.

## C.1 Algoritme for å finne Alfa

Inngangsparametre:  $n$ ,  $s$  og  $\emptyset$  (nedre, mest sannsynlig og øvre anslag)

1. Hvis  $n = s$  eller  $\emptyset = s$  returner alfa = 1,1563
2. Regn ut skjevhetsforholdet  $SF_{ns\emptyset}$  fra  $n$ ,  $s$  og  $\emptyset$ .
3. Hvis  $SF_{ns\emptyset} = 1$  returner alfa = 1 000 000 000 000. Se merknad 1
4. Hvis  $SF_{ns\emptyset} < 1$ ,  $SF_{ns\emptyset} = 1/SF_{ns\emptyset}$
5. Gjør et grovt anslag på alfa. Se merknad 2
6. Sett alfa = 10, beta = 10
7. Regn ut skjevhetsforholdet  $SF_{ab}$  fra alfa og beta. Se merknad 3
8. Regn ut den interne forholdet mellom skjevhetsforholdene.  $i = SF_{ab} / SF_{ns\emptyset}$
9. Hvis  $i = 1 \pm 0,0001$  avslutt og returner alfa
10. Hvis vi antall iterasjoner hittil er større enn 50 reduser  $i$ . Se merknad 4
11. Hvis  $i > 1$  alfa = alfa \*  $i$   
Gå tilbake til punkt 3
12. Hvis  $i < 1$  alfa = alfa / (2 -  $i$ )
13. Gå tilbake til punkt 3.

*Merkead 1*

Alfa bør i dette tilfellet settes så stor som mulig. Hvor stor den bør settes til vil være avhengig av datasystemet og hva verdien skal brukes til i senere bergninger. Det bunner ut i er at man skal sette den så stor som mulig slik at den ikke forårsaker overflow feil andre steder i programmet.

*Merkead 2*

Formel for å gjøre grovt anslag av alfa:

$$\text{alfa} = \frac{3978883,4}{f^{100}} - \frac{57,7483922}{e^f} + \frac{404,931744}{e^f} + 5,04650656$$

f i formelen er lik skjevhetsforholdet SFnsø i algoritmen.

*Merkead 3*

Se egen algoritme.

*Merkead 4*

Når vi har en veldig skjev kurve vil alfa være lav og små endringer i alfa representerer store endringer i skjevhetsforholdet. For å klare å få algoritmen til å konvergere må vi da redusere i for å få til mindre endringer av alfa.

i reduseres i første omgang etter hva den hittil beregnede alfa ligger på som følger:

$$\begin{aligned} \alpha < 2 & \quad i = \sqrt{i} \\ \alpha < 1,75 & \quad i = \sqrt[4]{i} \\ \alpha < 1,25 & \quad i = \sqrt[8]{i} \\ \alpha < 1,2 & \quad i = \sqrt[16]{i} \end{aligned}$$

I enkelte tilfeller vil heller ikke denne reduksjonen være tilstrekkelig til å få algoritmen til å konvergere, vi må i så fall redusere i enda mer. Hvis i fortsatt ikke har konverget når vi har nådd 5000 iterasjoner setter vi  $i = \sqrt[16]{i}$

Dette kommer i tillegg til den reduksjonen vi eventuelt har gjort tidligere som følge av lav alfa. Hvis algoritmen fortsatt ikke har konverget når vi har nådd 10 000 iterasjoner reduser vi igjen  $i = \sqrt[4]{i}$

Også denne gangen i tillegg til eventuelt andre reduksjoner som er gjort.

## C.2 Regne ut skjevhetsforholdet fra alfa og beta

Denne algoritmen kan brukes til å finne skjevhetsforhold for en gammafordelingen der man kjenner alfa og beta.

Inngangsparametre: alfa, beta, nkvan, økvan der nkvan og økvan er kvantilene vi benytter for nedre og øvre anslag.

1. Regn ut  $s$  ved  $s = \text{beta} * (\text{alfa} - 1)$
2. Regn ut  $n$  ved å ta den inverse kummulative tetthetsfunksjonen for gammafordelingen for nkvan, alfa og beta.
3. Regn ut  $\emptyset$  ved å ta den inverse kummulative tetthetsfunksjonen for gammafordelingen for økvan, alfa og beta.
4. Regn ut skjevhetsforholdet som normalt,  $\text{skjevhetsforhold} = (\emptyset - s) / (s - n)$

## C.3 Algoritme for å finne beta

Denne algoritmen baserer seg på at for en gitt alfa eller skjevhetsforhold, vil en beta-verdi representere en spesifikk avstand mellom to prosentkvantiler.

Denne algoritmen forutsetter en høyreskjev kurve. Hvis man har en venstreskjev kurve må tallene transformeres først. Se "Transformasjon av trippelanslag" på side 36. Hvis man benytter definisjonen av at venstreskjeve gammafordelinger har negativ beta må man i tillegg endre fortegn på beta etter å ha kjørt algoritmen som den står.

Inngangsparametere: alfa,  $n$ ,  $\emptyset$ ,  $n$ -kvantil,  $\emptyset$ -kvantil

1. Regn ut hva  $n$  og  $\emptyset$  vil være for den beregnede beta ved bruk av den inverse kummulative tetthetsfunksjonen for de aktuelle prosentkvantilene. Vi kaller disse verdiene  $x_n$  og  $x_\emptyset$ .
2. Regn ut  $i$  som følger:  $i = (\emptyset - n) / (x_\emptyset - x_n)$
3. Hvis  $i < 1,0001$  og  $i > 0,9999$  avslutt og returner beta
4. Hvis  $i > 1$   $\text{beta} = \text{beta} * i$ . Gå tilbake til 1.
5. Hvis  $\text{beta} = \text{beta} / (2 - i)$ . Gå tilbake til 1.

## C.4 Algoritme for å finne forskyvning

Forutsatt at man vet alfa og beta for en gammafordeling kan man regne ut hva forskyvningen vil være ved hjelp av et hvilket som helst punkt på kurven som man vet x-verdi og prosentkvantil for.

Inngangsparameter: alfa, beta, x-verdi, kvantil.

1. Regn ut x-verdien er for det oppgitte kvantilet for en uforskjøvet gammafordeling.
2. Returner resultatet fra 1 minus x-verdi





# Concept rapportserie

Papirtrykk: ISSN 0803-9763

Elektronisk utgave på internett: ISSN 0804-5585

Tilgjengelig på

[www.concept.ntnu.no/Publikasjoner/Rapportserie/concept\\_rapport.htm](http://www.concept.ntnu.no/Publikasjoner/Rapportserie/concept_rapport.htm)

Rapport	Tittel	Forfatter
Nr. 15	Kostnadsusikkerhet i store statlige investeringsprosjekter; Empiriske studier basert på KS2 Cost Uncertainty in large Public Investment Projects; Empirical studies based on QA2	Olav Torp (red.), Ole Morten Magnussen, Nils Olsson og Ole Jonny Klakegg
Nr. 14	Positiv usikkerhet og økt verdiskaping Positive uncertainty and increasing utility	Ingemund Jordanger
Nr. 13	Usikkerhetsanalyse – Feilkilder i metode og beregning Uncertainty analysis – sources of error in data and analysis	Kjell Austeng, Vibeke Binz og Frode Drevland
Nr. 12	Metoder for usikkerhetsanalyse Uncertainty analysis – Methodology	Kjell Austeng, Jon Terje Midtbø, Vidar Helland, Olav Torp og Ingemund Jordanger
Nr. 11	Usikkerhetsanalyse – Modellering, estimering og beregning Uncertainty analysis – modeling, estimation and calculation	Frode Drevland, Kjell Austeng og Olav Torp
Nr. 10	Usikkerhetsanalyse – Kontekst og grunnlag Uncertainty analysis – context and foundations	Kjell Austeng, Olav Torp, Jon Terje Midtbø, Ingemund Jordanger og Ole Morten Magnussen
Nr. 9	Bedre utforming av store offentlige investeringsprosjekter. Vurdering av behov, mål og effekt i tidligfasen Improved design of public investment projects. Making up-front appraisals of needs, objectives and effects	Petter Næss med bidrag fra Kjell Arne Brekke, Nils Olsson og Ole Jonny Klakegg
Nr. 8	Realopsjoner og fleksibilitet i store offentlige investeringsprosjekt Real options and flexibility in major public investment projects	Kjell Arne Brekke
Nr. 7	Hvordan trur vi at det blir? Effektvurderinger av store offentlige prosjekt An estimated guess. Up-front assessment of anticipated effects of major public investment projects	Nils Olsson
Nr. 6	Målformulering i store statlige investeringsprosjekt Formulation of objectives in major public investment projects	Ole Jonny Klakegg
Nr. 5	Bedre behovsanalyser. Erfaringer og anbefalinger om behovsanalyser i store offentlige investeringsprosjekt Needs analysis in major public investment projects. Lessons and recommendations	Petter Næss
Nr. 4	Konseptutvikling og evaluering i store statlige investeringsprosjekt Concept development and evaluation in major public investment projects	Hege Gry Solheim, Erik Dammen, Håvard O. Skaldebø, Eystein Myking, Elisabeth K. Svendsen og Paul Torgersen
Nr. 3	Beslutningsunderlag og beslutninger i store statlige investeringsprosjekt Decisions and basis for decisions in major public investment projects	Stein V. Larsen, Eilif Holte og Sverre Haanæs
Nr. 2	Statlig styring av prosjektledelse. Empiri og økonomiske prinsipper. Economic incentives in public project management	Dag Morten Dalen, Ola Lædre og Christian Riis
Nr. 1	Styring av prosjektporteføljer i staten. Usikkerhetsavsetning på porteføljenivå Public project portfolio management. Estimating provisions for uncertainty at portfolio level.	Stein Berntsen og Thorleif Sunde