



The Confidence Density for Correlation

Gunnar Taraldsen 

Norwegian University of Science and Technology, Trondheim, Norway

Abstract

Inference for correlation is central in statistics. From a Bayesian viewpoint, the final most complete outcome of inference for the correlation is the posterior distribution. An explicit formula for the posterior density for the correlation for the binormal is derived. This posterior is an optimal confidence distribution and corresponds to a standard objective prior. It coincides with the fiducial introduced by R.A. Fisher in 1930 in his first paper on fiducial inference. C.R. Rao derived an explicit elegant formula for this fiducial density, but the new formula using hypergeometric functions is better suited for numerical calculations. Several examples on real data are presented for illustration. A brief review of the connections between confidence distributions and Bayesian and fiducial inference is given in an [Appendix](#).

AMS (2000) subject classification. 62E15, 62F15, 62A01, 62H20, 62F25.

Keywords and phrases. Exact distribution, measures of association, Bayesian posterior, Fiducial, Fisher z-transform, Marginalization paradox.

1 Introduction

Fisher (1930) introduced the concept of a fiducial distribution. Fisher's first example is the fiducial density $\pi(\rho|r)$ for the correlation ρ of the binormal distribution. It is given by

$$\pi(\rho|r) = -\partial_\rho F(r|\rho), \quad (1.1)$$

where $F(r|\rho)$ is the cumulative distribution function for the empirical correlation r of a random sample of size n from the binormal distribution.

Earlier, Fisher (1915) had derived an explicit formula for the probability density $f(r|\rho) = \partial_r F(r|\rho)$ of the empirical correlation. Fisher's formula is

$$f(r|\rho) = \frac{(1 - \rho^2)^{\frac{\nu}{2}} \cdot (1 - r^2)^{\frac{\nu-3}{2}}}{\pi(\nu - 2)!} \partial_{\rho r}^{\nu-1} \left\{ \frac{\theta}{\sin \theta} \right\}, \quad (1.2)$$

where $\cos \theta = -\rho r$, $0 < \theta < \pi$, and $\nu = n - 1$ is the degrees of freedom. In principle, formula (1.1) and formula (1.2) give a method for deriving a

more explicit formula for $\pi(\rho|r)$. This is, however, not straightforward since a convenient closed form expression for $F(r|\rho)$ is missing.

The problem of deriving a more explicit formula for $\pi(\rho|r)$ was solved by C. R. Rao. It is stated as the very last formula in the classical book *Statistical Methods and Scientific Inference* by Fisher (1973, eq.(234)). Rao's formula is

$$\pi(\rho|r) = \frac{(1-r^2)^{\frac{\nu-1}{2}} \cdot (1-\rho^2)^{\frac{\nu-2}{2}}}{\pi(\nu-2)!} \partial_{\rho r}^{\nu-2} \left\{ \frac{\theta - \frac{1}{2} \sin 2\theta}{\sin^3 \theta} \right\}. \quad (1.3)$$

Unfortunately, this elegant formula by Rao is less known than the corresponding formula (1.2). The density $\pi(\rho|r)$ is more important than the density $f(r|\rho)$ in applications since it represents directly the resulting state of knowledge for the unknown correlation ρ based on the observed correlation r .

The Rao formula (1.3) for the density π of ρ is similar in form to the formula derived by Fisher (1915) for the density f of r defined in Eq. 1.2. Historically, it took quite some time to arrive at an alternative formula suitable for the numerical calculation of $f(r|\rho)$. Hotelling (1953) arrived at a formula using hypergeometric functions for calculation of $f(r|\rho)$, and this solution is now advocated by Stuart and Ord (1994, p.559–65) and Anderson (2003, p.122–26). Theorem 2.1 proved in Section 2 gives a similar explicit formula for $\pi(\rho|r)$ using hypergeometric functions.

The concept of a confidence distribution is possibly unfamiliar to many readers. A brief introduction is given in the next section. Then the main result in Theorem 2.1 is stated and proved. This gives the exact confidence density corresponding to the fiducial distribution defined by Fisher (1930). The final section contains a brief discussion of some consequences with some additional remarks on the relevance of the fiducial argument in current statistics. An Appendix explains in more detail the connection of the derived confidence distribution to the corresponding Bayesian posterior.

2 Theory

According to Cox (1958, p.363) a confidence distribution for a parameter can be defined directly or introduced as the set of all confidence intervals at different levels. A direct definition can be given starting with a random cumulative distribution function C depending on the data Y . For completeness the mathematical details are specified next.

It is assumed that (Ω, \mathcal{E}, P) is the underlying probability space following Kolmogorov (1933). The data Y is given by a measurable function $Y : \Omega \rightarrow \Omega_Y$ where Ω_Y is the sample space of the data. A statistical model

is defined by assuming that P is unknown, and depending on an unknown model parameter $\theta \in \Omega_\Theta$. A parameter $\gamma = \psi(\theta)$ is defined by a measurable function $\psi : \Omega_\Theta \rightarrow \Omega_\Gamma$. Let $U \sim U(0, 1)$ mean $P(U \leq u) = u$ for all $0 < u < 1$.

DEFINITION 2.1. A statistic C is an exact confidence distribution for a real parameter ρ if:

1. $\rho \mapsto C(\rho; y)$ is a cumulative distribution function for all y in the sample space of the data Y .
2. $C(\rho; Y) \sim U(0, 1)$

If the cumulative distribution function is differentiable with $\pi(\rho; y) = \partial_\rho C(\rho; y)$, then π is an exact confidence density for ρ .

It must be observed in Definition 2.1 that the both the law P_Y of the data Y and the parameter ρ depend on the model. Definition 2.1 is as given by Schweder and Hjort (2016, Definition 3.1, p.58). This is equivalent to demanding that the $W_p = C^{-1}(p)$ fractile of C defines an exact confidence interval $(-\infty, W_p]$ for ρ . The proof is given by

$$P(\rho \leq W_p) = P(C(\rho) \leq p) = p. \tag{2.4}$$

Again, it is important to notice that both probabilities depend on the law of the data Y , and ρ is a parameter of the law of the data.

In the particular case given by Eq. 1.1

$$C(\rho; y) = 1 - F(r|\rho) \tag{2.5}$$

where r is the empirical correlation of the random sample $y = ((y_1, y_2)_1, \dots, (y_1, y_2)_n)$ from the binormal law with correlation ρ . The unknown model is given by the model parameter $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ corresponding to the means μ_1, μ_2 , the standard deviations σ_1, σ_2 , and the correlation ρ .

For this case it follows then that $\pi(\rho|r) = \partial_\rho C(\rho; y)$ is an exact confidence density as explained originally by Fisher (1930, p.532–5). The proof follows by observing that the law of the empirical correlation r only depends on the correlation ρ , that $\rho \mapsto 1 - F(r|\rho)$ is a differentiable cumulative distribution function, and generally that $F(X) \sim U(0, 1)$ for a continuous random variable X with cumulative distribution function F .

An explicit formula for the density $\pi(\rho|r)$ does not follow from the above arguments, but is derived below using an alternative path.

THEOREM 2.1. *Let r be the empirical correlation of a random sample of size n from the binormal. The exact confidence density for the correlation ρ is*

$$\pi(\rho|r) = \frac{(1 - r^2)^{\frac{\nu-1}{2}} \cdot (1 - \rho^2)^{\frac{\nu-2}{2}} \cdot (1 - r\rho)^{\frac{1-2\nu}{2}}}{\sqrt{2}B(\nu + \frac{1}{2}, \frac{1}{2})} \cdot F\left(\frac{3}{2}, -\frac{1}{2}; \nu + \frac{1}{2}; \frac{1 + r\rho}{2}\right)$$

where B is the Beta function, F is the Gaussian hypergeometric function, and $\nu = n - 1 > 1$. The one-sided confidence intervals from $\pi(\rho|r)$ are uniformly most accurate invariant with respect to the location-scale groups on the two coordinates.

PROOF. The idea is to use the Elfving equation

$$\sqrt{u} \frac{\rho}{\sqrt{1 - \rho^2}} - \sqrt{v} \frac{r}{\sqrt{1 - r^2}} = z. \tag{2.6}$$

This relation was also obtained by Professor David Sprott as explained and proved by Fraser (1964, p.853, eq.(5.1)). Here $u \sim \chi^2(\nu), v \sim \chi^2(\nu - 1), z \sim N(0, 1)$ are independent. Equation 2.6 gives the law of ρ when r is known. The degrees of freedom ν equals $n - 1$ for sample size n . If the means are known, then $\nu = n$. Equation 2.6 is due to Elfving (1947) according to Lee (1971, p.117).

Equation 2.6 gives the conditional density of ρ given u, v . The marginal density of ρ follows then by integration over u, v . This integration is done by a change of variables resulting in a gamma integral and results in the density $\pi(\rho|r)$. The details are as follows.

The conditional density of $s = \rho/\sqrt{1 - \rho^2}$ given u, v is normal by Eq. 2.6 with $(s|u, v) \sim N(\sqrt{\frac{v}{u}}t, 1/u)$ where $t = r/\sqrt{1 - r^2}$. Using this, the law of u, v , and $ds = (1 - \rho^2)^{-3/2}d\rho$ give the joint density of ρ, u, v as

$$(1 - \rho^2)^{-3/2} \cdot \frac{u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \cdot \frac{v^{\frac{\nu-1}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{\nu-1}{2}} \Gamma(\frac{\nu-1}{2})} \cdot \sqrt{\frac{u}{2\pi}} e^{-\frac{u}{2}(s - \sqrt{\frac{v}{u}}t)^2}. \tag{2.7}$$

The terms in the exponential are

$$-\frac{1}{2} \left[\frac{u}{1 - \rho^2} - \frac{2\sqrt{uv}\rho r}{\sqrt{(1 - \rho^2)(1 - r^2)}} + \frac{v}{1 - r^2} \right] = -\frac{\nu(s_1^2 - 2s_1 s_2 r\rho + s_2^2)}{2(1 - r^2)} \tag{2.8}$$

using new coordinates (s_1, s_2) defined by $\nu s_1^2 = u(1 - r^2)/(1 - \rho^2)$ and $\nu s_2^2 = v$. Let $s_1 = \sqrt{\alpha} \exp(-\beta/2)$ and $s_2 = \sqrt{\alpha} \exp(\beta/2)$. The density for ρ, α, β from Eq. 2.7 is

$$\frac{2^{1-\nu} \nu^\nu}{\sqrt{\pi} \Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu-1}{2})} (1 - r^2)^{-\frac{\nu+1}{2}} (1 - \rho^2)^{\frac{\nu-2}{2}} e^{-\beta} \alpha^{\nu-1} e^{\frac{-\nu\alpha(\cosh(\beta) - \rho r)}{1 - r^2}}. \tag{2.9}$$

Integration over α gives $\pi(\rho|r)$ using the identity $\pi(\nu - 2)! = \sqrt{\pi}2^{\nu-2}\Gamma(\frac{\nu}{2})\Gamma(\frac{\nu-1}{2})$ and adjusting an integral representation of F (Olver et al. 2010, 14.3.9,14.12.4). The normalization factor from this is $[\nu(\nu-1)\Gamma(\nu-1)]/[\sqrt{2\pi}\Gamma(\nu + \frac{1}{2})]$ which simplifies into the stated $1/[\sqrt{2}B(\nu + \frac{1}{2}, \frac{1}{2})]$. This ends the proof of the formula. The optimality claim is a consequence of Problem 6.68 presented by Lehmann and Romano (2005, p.273).

The formula for $\pi(\rho|r)$ can be reformulated using Legendre functions similarly to the formula for $f(r|\rho)$ obtained by Fisher (1915, p.511). Interesting recursion relations can also be established. The details of this are not given here since Theorem 2.1 is of a form more suitable for numerical calculations.

The method of proof is also of independent interest since it gives an alternative and simpler derivation of the known exact formulas for $f(r|\rho)$ derived by Fisher (1915, p.507–11), Hotelling (1953, p.197–200), Stuart and Ord (1994, p.559–65), Anderson (2003, p.122–6). It gives also a possible path for derivation of exact confidence distributions for partial correlation functions.

The formula for $\pi(\rho|r)$ seems difficult to derive directly from the formula for $f(r|\rho)$ and Eq. 1.1. This is possibly the reason why an alternative explicit formula for $\pi(\rho|r)$ has been absent for so long. Another reason is given by the good approximation given by the Fisher (1921) z-transformation:

$$\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) - \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) \approx z / \sqrt{\nu - 2}. \tag{2.10}$$

Replacing Eq. 2.6 with Eq. 2.10 and solving with respect to ρ gives the z-transform confidence density

$$\tilde{\pi}(\rho|r) = \sqrt{\frac{\nu - 2}{2\pi}} (1 - \rho^2)^{-1} e^{\frac{2-\nu}{8} \left[\ln \left(\frac{(1+\rho)(1-r)}{(1-\rho)(1+r)} \right) \right]^2} \tag{2.11}$$

for $\nu = n - 1 > 2$. The density $\tilde{\pi}$ is a good approximation to π for large sample size n , and is surprisingly accurate also for moderate sample size.

3 Examples

The result of an experiment is given by four points with (x, y) coordinates (773, 727), (777, 735), (284, 286), and (519, 573). There are reasons a priori for assuming a linear relationship. This is further supported by Fig. 1, and a high value for the coefficient of determination $R^2 = 97.00\%$. The R^2 equals the square of the empirical correlation $r = 98.49\%$. This is an example of linear regression used extensively in applied sciences. A natural question is then: What about uncertainty? The focus in the following is the

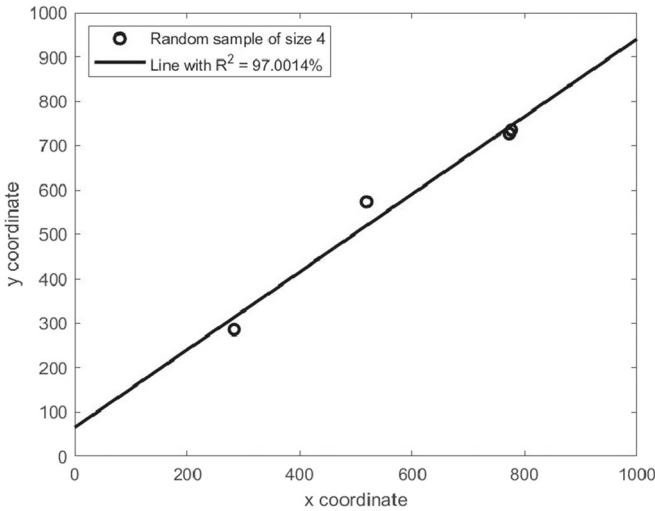


Figure 1: A sample of size 4 with a regression line corresponding to an example by Fisher (1930)

correlation, but other parameters are of course of possible interest depending on the concrete application.

An approximate 95% one sided confidence interval for the correlation ρ based on the Fisher (1921) z-transformation is [66.08, 100]%. Linear interpolation in the table presented by Fisher (1930) gives an exact 95% confidence interval [67.42, 100]%.¹ Our Theorem 2.1, without linear interpolation, gives the true exact 95% confidence interval [67.39, 100]%. This demonstrates that the z-transformation can be quite good also for small sample size.

More complete information is given by the confidence densities shown in Fig. 2. The exact confidence density in Fig. 2 illustrates the corresponding uncertainty corresponding to all possible confidence intervals with all possible confidence levels. The density is also the Bayesian posterior from a standard prior. It represents hence all information available for the correlation based on the observations. Figure 2 also shows the approximate confidence density $\tilde{\pi}$ from the z-transform. It is similar, but clearly $\tilde{\pi}$ is different from π .

Figure 3 shows the cd4 counts for 20 HIV-positive subjects (Efron 1998, p.101). The x-axis gives the baseline count and the y-axis gives the count

¹ $66.4037 + (71.6298 - 66.4037) * (98.4893 - 98.4298) / (98.7371 - 98.4298) = 67.4156$

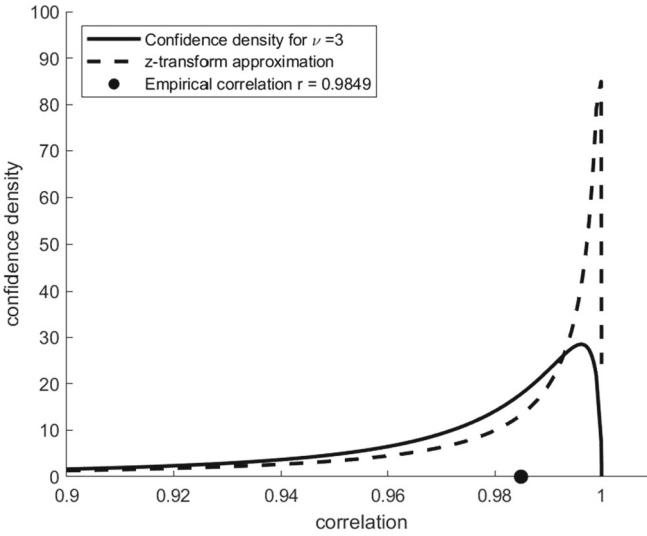


Figure 2: The confidence density and the z-transform density for the Fisher (1930, p.434) example

after one year of treatment with an experimental antiviral drug. The empirical correlation is $r = 0.7232$, and the equitail z-transform 90% approximate confidence interval is $[47.41, 86.51]\%$.

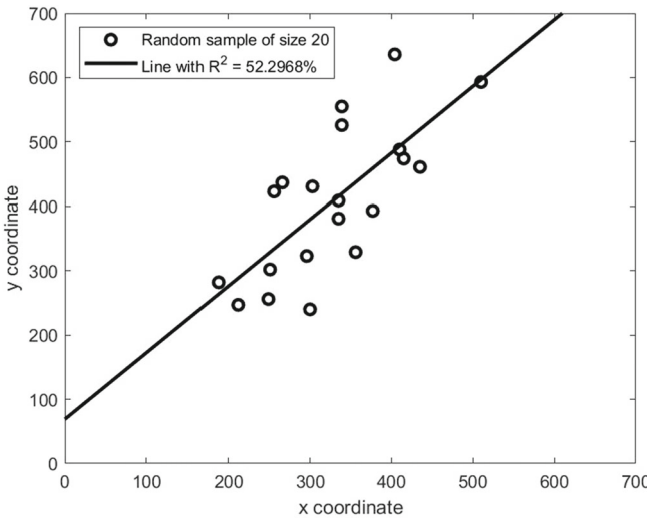


Figure 3: The HIV data of DiCiccio and Efron (1996, Table 1)

Figure 4 shows the closeness of the confidence density π and the z-transform density $\tilde{\pi}$. The exact equitail 90% confidence interval from Theorem 2.1 is [46.54, 85.74]%. It is shifted to the left as also can be inferred from Fig. 4. Efron (1998, p.101) discuss this example in more detail using bootstrap techniques.

As a final example, consider certain pairs of measurements with $r = 0.534$ taken on $n = 8$ children at age 24 months in connection with a study at a university hospital in Hong Kong. Figure 5 shows again the closeness of the confidence density and the z-transform density. Schweder and Hjort (2016, p.227, Figure 7.8) discusses this example in much more detail including different bootstrap approaches. Using the method of Fisher (1930) they arrive at the same plot of the confidence density using the exact distribution for the empirical correlation. This provides additional verification of the exact result in Theorem 2.1.

An alternative method for all examples is density estimation based on Markov Chain Monte Carlo methods from the standard prior for the binormal with five unknown parameters, but the explicit formula is preferable. The explicit formula can be used as a benchmark test case for a general MCMC posterior implementation.

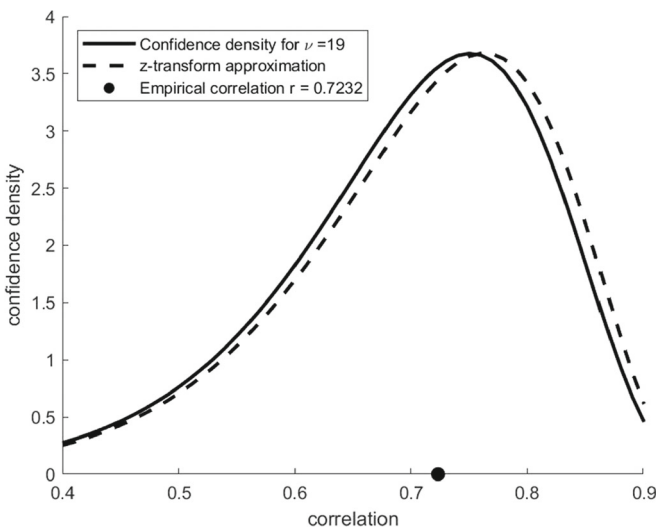


Figure 4: The confidence density and the z-transform density for HIV data DiCiccio and Efron (1996, Table 1)

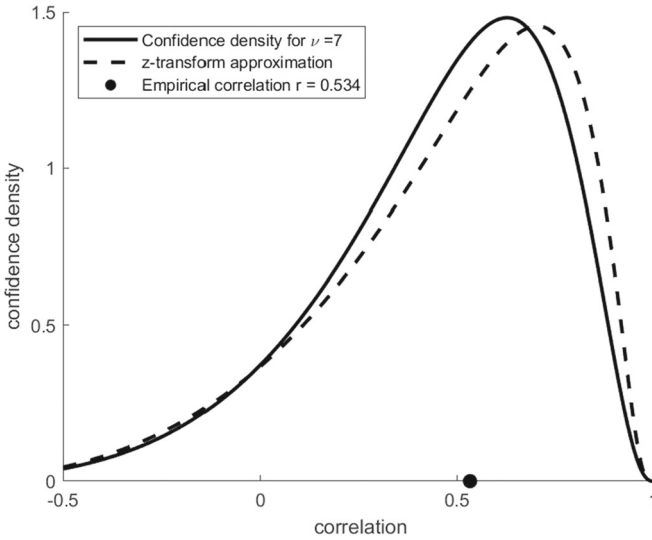


Figure 5: The confidence density and the z-transform density for data from 8 children Schweder and Hjort (2016, p.227, Figure 7.8)

4 Discussion

The fiducial density $\pi(\rho|r)$ coincides with the Bayesian posterior from the standard prior for the five unknown parameters in the binormal and gives optimal inference. It is an exact confidence density. This is explained by Taraldsen and Lindqvist (2013). The explicit formulas for $f(r|\rho)$ and $\pi(\rho|r)$ prove that the fiducial is not obtainable from a prior $\pi(\rho)$. This marginalization paradox is known, but the previously known proof is complicated Berger and Sun (2008, p.966-7).

In current mathematical statistics the problem of choice of a prior is central, and the current revival of the fiducial argument presents an alternative solution to this problem. In a non-parametric problem the choice of a Bayesian prior can be justified by establishing good asymptotic frequentist coverage properties (Castillo and Nickl, 2013; Ghosal and van der Vaart, 2017). Cui and Hannig (2019) demonstrate that this can also be done by a fiducial argument for a non-parametric problem without a Bayesian prior.

Schweder and Hjort (2016) present recent developments in the theory of confidence distributions and advocates this as an alternative to the calculation of Bayesian posteriors. Taraldsen and Lindqvist (2019) explain that the problem of choosing a prior, including non-parametric problems, can be solved by not choosing a prior, but rather using the information contained

in a data generating equation. Xie and Singh (2013) explain how the concept of a confidence distribution can be seen as the frequentist distribution estimator of a parameter.

In his initial work on the fiducial argument Fisher (1930, p.532–5) used a frequency interpretation for its justification. In later works, Fisher (1973, p.54–5) insisted on a more general interpretation: When there is no prior information, then the interpretation of the fiducial is exactly as for a Bayesian posterior. It is the state of knowledge of the parameter given the observations. The modern view is that the knowledge given by a prior is replaced by the knowledge inherent in a particular data generating equation. Taraldsen and Lindqvist (2013, p.331) demonstrate that this can give optimal inference in a non-parametric Hilbert space problem, and Cui and Hannig (2019) demonstrate superiority of a fiducial distribution in a non-parametric problem for survival functions under censoring.

Neyman (1937, eq.20) is usually credited for the invention of the theory of confidence intervals. Cox (1958, p.363–6) can likewise be credited for suggesting the use of confidence distributions in statistical inference. Actually, Fisher (1930, p.532–5) defines both concepts precisely, and uses the correlation coefficient and Eq. 1.1 as a concrete example with numerical calculations. Combining his initial results gives an algorithm for numerical calculation of exact confidence intervals for the correlation. Fisher (1930, Table, p.533) computed a table with exact 5% and 95% percentiles for sample size $n = 4$. His table, up to numerical rounding, is consistent with a more direct approach based on Theorem 2.1. This gives an independent check of the claim in Theorem 2.1, and also of the table calculated by Fisher.

Inference for correlation is, even if of a seemingly elementary kind, of central importance in applied statistics. It is almost impossible to find a linear regression plot without the accompanying R^2 . The exact solution by Fisher (1930) is rarely used. Standard software gives an approximate solution using the Fisher (1921) z-transform. An example using the z-transform is given by Efron (1998, p.101). The density in Theorem 2.1 gives the uncertainty associated with the estimated correlation, and hence also of the correlation squared. Approximate inference using the Fisher (1921) z-transform can, and should, be replaced by exact inference.

The fiducial density $\pi(\rho|r)$ corresponds to the very first example used by Fisher (1930) when he introduced his fiducial argument. Fisher justified this fiducial distribution by proving that the corresponding quantiles give

exact confidence intervals. This was a starting point for Neyman (1937) when formulating a general theory of confidence intervals. It can safely be concluded that the seminal paper by Fisher (1930) has been pivotal in the historical development of mathematical statistics, and it still is.

Acknowledgements. The author gratefully acknowledges the support and encouragement of Professor Bo Henry Lindqvist during many years of cooperation on this and similar problems. Constructive and positive remarks on a first version of the manuscript from a reviewer were also most helpful and are hereby acknowledged. This contributed to several improvements including the [Appendix](#).

Funding. Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Compliance with Ethical Standards. The author declares that he has no conflict of interest. I have no relevant financial or non-financial interests to disclose. No funds, grants, or other support was received.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- ANDERSON, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley-Interscience, Hoboken.
- BARNARD, G. A. (1995). Pivotal models and the Fiducial Argument. *Int. Stat. Rev./Revue Internationale de Statistique* **63**, 309–323.
- BERGER, J. O. and SUN, D. (2008). Objective priors for the bivariate normal model. *Ann. Stat.* **36**, 963–82.
- CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference (2nd edn)*. Thomson Learning, Duxbury.
- CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Stat.* **41**, 1999–2028.
- COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Stat.* **29**, 357–372.

- COX, D. R. and HINKLEY, D. V. (1974). *Theoretical statistics*. Chapman and Hall, London.
- CUI, Y. and HANNIG, J. (2019). Nonparametric generalized fiducial inference for survival functions under censoring. *Biometrika* **106**, 501–518.
- DAWID, P. (2020). Fiducial inference then and now. arXiv:2012.10689 [math, stat].
- DAWID, A. P. and WANG, J. (1993). Fiducial prediction and semi-Bayesian inference. *Ann. Stat.* **21**, 1119–1138.
- DICICCIO, T. J. and EFRON, B. (1996). Bootstrap confidence intervals. *Stat. Sci.* **11**, 189–212.
- EATON, M. L. (1989). *Group invariance applications in Statistics*. Institute of Mathematical Statistics.
- EATON, M. L. and SUDDERTH, W. D. (2012). Invariance, model matching and probability matching. *Sankhya A* **74**, 170–193.
- EFRON, B. (1998). R. A. Fisher in the 21st century (Invited paper presented at the 1996 R. A. Fisher Lecture). *Stat. Sci.* **13**, 95–122.
- ELFVING, G. (1947). A simple method of deducing certain distributions connected with multivariate sampling. *Scand. Actuar. J.* **1947**, 56–74.
- FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**, 507–21.
- FISHER, R. A. (1921). On the ‘probable error’ of a coefficient of correlation deduced from a small sample. *Metron* **1**, 1–32.
- FISHER, R. A. (1930). Inverse probability. *Proc. Camb. Phil. Soc.* **26**, 528–535.
- FISHER, R. A. (1973). *Statistical methods and scientific inference*. Hafner Press, New York.
- FRASER, D. A. S. (1964). On the definition of fiducial probability. *Bull. Int. Stat. Inst.* **40**, 842–856.
- FRASER, D. A. S. (1968). *The structure of inference*. Wiley, New York.
- FRASER, D. A. S. (1979). *Inference and linear models*. McGraw-Hill, New York.
- GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. *J. R. Stat. Soc. Ser. B (Methodol.)* **25**, 368–376.
- GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press, Cambridge.
- HANNIG, J., IYER, H., LAI, R. C. S. and LEE, T. C. M. (2016). Generalized fiducial inference: a review and new results. *J. Am. Stat. Assoc.* **111**, 1346–1361.
- HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *J. R. Stat. Soc. Ser. B (Methodol.)* **15**, 193–232.
- KENDALL, M. G. and STUART, A. (1961). *The advanced theory of statistics. Volume 2: inference and relationship*. Hafner Publishing Company, New York.
- KOLMOGOROV, A. (1933). *Foundations of the theory of probability* Chelsea (1956).
- LEE, Y. -S. (1971). Some results on the sampling distribution of the multiple correlation coefficient. *J. R. Stat. Soc. Ser. B (Methodol.)* **33**, 117–130.
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing statistical hypotheses*. Springer, New York.
- NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci.* **236**, 333–380.
- OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W. (EDS.) (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press, Cambridge.
- PEDERSEN, J. G. (1978). Fiducial inference. *Int. Stat. Rev./Revue Internationale de Statistique* **46**, 147–170.

- SCHWEDER, T. and HJORT, N. L. (2016). *Confidence, likelihood, probability: statistical inference with confidence distributions*. Cambridge University Press, Cambridge.
- SEIDENFELD, T. (1979). *Philosophical problems of statistical inference: learning from R.A. Fisher. Theory and Decision Library*. Springer, Netherlands.
- SPROTT, D. A. (2000). *Statistical inference in science. Springer Series in Statistics*. Springer, New York.
- STUART, A. and ORD, K. (1994). *Kendall's Advanced theory of statistics, distribution theory (volume 1 ed.)* Wiley, New York.
- STUART, A., ORD, K. and ARNOLD, S. (1999). *Kendall's advanced theory of statistics, classical inference and the linear model (6th edn), Volume 2A*. Wiley, New York.
- TARALDSEN, G. and LINDQVIST, B. H. (2013). Fiducial theory and optimal inference. *Ann. Stat.* **41**, 323–341.
- TARALDSEN, G. and LINDQVIST, B. H. (2015). Fiducial and posterior sampling. *Commun. Stat.: Theory Methods* **44**, 3754–3767.
- TARALDSEN, G. and LINDQVIST, B. H. (2018). Conditional fiducial models. *J. Stat. Plan. Inference* **195**, 141–152.
- TARALDSEN, G. and LINDQVIST, B. H. (2019). Discussion of ‘Nonparametric generalized fiducial inference for survival functions under censoring’. *Biometrika* **106**, 523–526.
- VERONESE, P. and MELILLI, E. (2015). Fiducial and confidence distributions for real exponential families. *Scand. J. Stat.* **42**, 471–484.
- XIE, M. -G. and SINGH, K. (2013). Confidence distribution, the Frequentist Distribution Estimator of a parameter: a review. *Int. Stat. Rev.* **81**, 3–39.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A: Confidence in Fiducial and Bayesian Inference?

The adjective “fiducial” comes from the Latin *fiducia* for faith and means something taken as standard of reference or something founded on faith or trust. The term “fiducial” is used in the following also as a noun instead of the longer “fiducial distribution”. Similarly, “posterior” is used as a noun instead of “posterior distribution”. The purpose of this Appendix is to briefly summarize mathematical theorems providing links between fiducial inference, Bayesian inference, and confidence distributions as needed for the correlation coefficient problem.

As noted by a reviewer, the procedures reinforced on the concept of fiducial probability are not always viewed undoubtedly. Fiducial inference has not been widely accepted. Different versions are presented in textbooks and reference works, and most often with critical remarks. Casella and Berger (2002, p.291), Kendall and Stuart (1961, p.134–), and Stuart et al. (1999, p.440–) present versions based directly on the likelihood, but Cox and Hinkley (1974, p.246), Sprott (2000, p.77), and Barnard (1995) present versions based on pivotal quantities. The version presented below is a generalization

of the version based on pivotal quantities. It is in particular not restricted to models defined by a likelihood. It is based on a data generating equation.

A more complete review of fiducial inference is beyond the scope of this Appendix, but Seidenfeld (1979), Dawid and Wang (1993), Lehmann and Romano (2005, p.175), Veronese and Melilli (2015), Hannig et al. (2016), Taraldsen and Lindqvist (2018), and Dawid (2020) can be consulted for discussion and further references. Pedersen (1978, p.147) concluded that the fiducial argument has had very limited success and that it was essentially dead. There was then hence little confidence in fiducial inference among most statisticians, and this is also true today.

In contrast, in the context of objective Bayesian inference, Efron (1998, p.107) suggests that maybe Fisher's biggest blunder will become a big hit in the 21st century. I agree. I have much confidence in fiducial inference. One reason for my optimism is given by the version of fiducial inference which is based on a data generating model. This is perfectly adapted to the available computational power in the 21st century. Fiducial inference is, by definition here in this [Appendix](#), understood as inference based on a data generating model.

DEFINITION A.2. A data generating model for a statistic r is given by a function $r = r(\rho, m)$ where ρ is a parameter and m has a known probability distribution. The model is simple if it can be uniquely solved to give a parameter generating function $\rho = \rho(r, m)$. The model is pivotal if it can be uniquely solved to give a pivotal $m = m(r, \rho)$. For a simple data generating model the fiducial is defined to be the distribution of $\rho(r, m)$ for fixed r .

In Definition A.2, and in the following, the term *function* is used as a synonym with *measurable function*. This means that all sets involved are assumed to be measurable spaces. The fundamental assumptions are hence as explained before Definition 2.1. A statistic is by definition a (measurable!) function of the data.

There are many different data generating models for a given statistical model. In general, as will be demonstrated, the corresponding fiducial depends on the data generating model. For the case of a real statistic, exemplified by r in Theorem 2.1, the fiducial is, however, uniquely determined by the statistical model under mild assumptions.

THEOREM A.2 (Uniqueness of the fiducial). *The fiducial distribution from of a real valued, strictly monotonic, and simple data generating model is uniquely determined by the sampling distribution of the statistic. If, additionally, the sampling distribution of the statistic is continuous, then the fiducial distribution is an exact confidence distribution.*

PROOF. This is Theorem 1 proved by Taraldsen and Lindqvist (2018, p.143). The idea is that inversion of the cumulative distribution function gives a data generating model which is both simple and pivotal and a calculation shows that the fiducial from this coincides with the initial model.

In the case of the correlation coefficient, the Elfving equation (2.6) can be solved to give a data generating model $r = r_E(\rho, m)$ with $m = (u, v, z)$. This is not a pivotal model, but Theorem A.2 gives that the fiducial is an exact confidence distribution as claimed in Theorem 2.1. It gives, furthermore, that the fiducial coincides with the fiducial from the data generating model $r = r_F(\rho, m) = F^{-1}(m, \rho)$ with $m \sim U(0, 1)$ corresponding to Eq. 1.1.

The fiducial in Theorem 2.1 also coincides with a Bayesian posterior from a random sample from the binormal as claimed in the Abstract.

THEOREM A.3. *The confidence density $\pi(\rho|r)$ in Theorem 2.1 is the posterior density from the prior density*

$$\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{\sigma_2^2(1 - \rho^2)} \quad (\text{A.12})$$

and also from the prior density

$$\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{\sigma_1^2(1 - \rho^2)}. \quad (\text{A.13})$$

PROOF. The Elfving equation (2.6) can be solved to give a parameter generating function $\rho = \rho(r, u, v, z)$. This parameter generating function equals the constructive posterior provided by Berger and Sun (2008, Table 1, p.66) for the posterior in Eq. A.13. It follows then from Theorem 2.1 that the confidence density $\pi(\rho|r)$ coincides with the posterior density as claimed in the abstract. Interchange of the coordinates shows that this also holds for the prior in Eq. A.12.

The details in the above proof requires several pages of computations. The connection of the derived confidence distribution to the corresponding Bayesian posterior is not straightforward. A direct proof by integrating out $\mu_1, \mu_2, \sigma_1, \sigma_2$ using the joint posterior density of $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ is also not a simple task. The explicit formula in Theorem 2.1 is hence important also from a Bayesian perspective.

Geisser and Cornfield (1963) consider a prior which is symmetric in the coordinates, and derive an explicit formula for the posterior density of the correlation by direct integration. It is of the same form as the density of the empirical correlation, but with r and ρ interchanged. They prove that it differs from the fiducial $\pi(\rho|r)$ using a difficult analytic argument. This

difference can be more easily seen now by direct comparison of the two explicit formulas. The posterior $\pi(\rho|r)$ is optimal as stated in Theorem 2.1, and is in this sense preferable compared to the posterior found by Geisser and Cornfield (1963, eq. 3.10).

An alternative proof of Theorem A.3 follows from arguments given originally by Fraser (1964, p.846). The argument starts with the data generating model

$$y_{1j} = \mu_1 + \sigma_1 m_{1j}, \quad y_{2j} = \mu_2 + \sigma_2(\rho m_{1j} + \sqrt{1 - \rho^2} m_{2j}) \quad (\text{A.14})$$

corresponding to Cholesky decomposition of the covariance matrix and independent $m_{ij} \sim N(0, 1)$. This is a group model. The fiducial can be computed by conditioning on a maximal invariant statistic, or alternatively by reduction to the minimal sufficient statistic as explained by Fraser (1968, 1979) and Taraldsen and Lindqvist (2013).

Fraser (1979, p.184) notes in particular that for an application the particular order for the coordinates is part of the specification of the system being investigated. It is assumed that the observed data have been generated by Eq. A.14. This corresponds to first drawing a random sample of x -coordinates, and then drawing a sample of y -coordinates as in linear regression. The choice of a particular Bayesian prior corresponds to the choice of a particular data generating model. This explains in this case the asymmetry and difference between the two priors in Theorem A.3. The asymmetry is part of the model assumptions both in the Bayesian analysis and the fiducial analysis.

A data generating equation $r = r(\rho, u, v, z)$ corresponding to the Elfving equation (2.6) can be computed directly from the data generating Eq. A.14 and the definition of r . Theorem A.2 ensures that the choice of the particular data generating Eq. A.14 is irrelevant for the resulting fiducial for ρ , but it is convenient for proving Theorem 2.1.

As explained by Taraldsen and Lindqvist (2013, p.329) it is known that the fiducial coincides with the posterior from a right Haar prior for a group model. A more general result was proven by Taraldsen and Lindqvist (2015, p.3756, Thm 2.1).

THEOREM A.4. *The fiducial from a simple data generating model $t = t(\theta, m)$ is the Bayes posterior from a σ -finite prior P_Θ if the distribution of $t(\theta, m)$ does not depend on m when $\theta \sim P_\Theta$.*

PROOF. The idea of the proof by Taraldsen and Lindqvist (2015, p.3765) is given by establishing a joint density for $(t(\Theta, M), M)$ using the Fubini theorem together with the general change-of-variables theorem from measure

theory. This is different from the more common change-of-variables theorem which requires differentiability. It should also be noted that there is no assumption of existence of a dominating measure. The proof holds in the generality as stated. It follows also that the posterior is proper.

The assumption of Theorem A.4 is fulfilled when $t = t(\theta, m) = \theta m$ is given by group multiplication and P_{Θ} is chosen as the right Haar prior. This gives in particular equality of the the joint posterior density of $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ and the joint fiducial. The priors in Theorem A.3 correspond to the right Haar priors corresponding to the choice of a Cholesky decomposition in lower- or upper-triangular matrices. Explicit expressions for right Haar priors are derived and discussed by Eaton (1989) in the more general context of group invariance in statistics.

The reader is possibly not completely happy with the asymmetry between the coordinates in the Haar priors in Theorem A.3. The conclusion for the correlation, however, holds equally well for the prior obtained by replacing σ_2 by σ_1 . It is the joint posterior that changes into the fiducial obtained from using upper triangular instead of lower triangular matrices in the Cholesky decomposition for the data generating equation. In both cases the posterior density for the correlation equals the fiducial in Theorem 2.1.

Fraser (1968, p.192-) gives a fiducial argument for the binormal based on a model which is symmetric in the coordinates. The resulting fiducial for ρ is as in Theorem 2.1, but it is unknown if the joint fiducial for $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ is a Bayes posterior. The problem of the binormal, and its natural generalizations, remains a very interesting problem from Bayesian, fiducial, and frequentist points of view. It can shed light on the more general problem of when a posterior or a fiducial is a confidence distribution. Eaton and Sudderth (2012) give more examples of Bayesian posteriors that are exact confidence distributions in the sense of giving exact confidence regions.

GUNNAR TARALDSEN
NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY, TRONDHEIM, NORWAY
E-mail: Gunnar.Taraldsen@ntnu.no

Paper received: 29 May 2021; accepted 13 September 2021.