Analytic posteriors for Pearson's correlation coefficient

Ly, A.; Marsman, M.; Wagenmakers, E.-J.

DOI
10.1111/stan.12111

Publication date
2018

Document Version
Final published version

Published in
Statistica Neerlandica

License
CC BY-NC

Link to publication

Citation for published version (APA):
Ly, A., Marsman, M., & Wagenmakers, E.-J. (2018). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*, *72*(1), 4-13. https://doi.org/10.1111/stan.12111

Statistica Neerlandica (2018) Vol. 72, nr. 1, pp. 4–13
doi:10.1111/stan.12111

Analytic posteriors for Pearson's correlation coefficient

Alexander Ly, Maarten Marsman, and Eric-Jan Wagenmakers

Department of Psychological Methods, University of Amsterdam , PO Box 15906, 1001 NK Amsterdam, The Netherlands

Pearson's correlation is one of the most common measures of linear dependence. Recently, Bernardo (11th International Workshop on Objective Bayes Methodology, 2015) introduced a flexible class of priors to study this measure in a Bayesian setting. For this large class of priors, we show that the (marginal) posterior for Pearson's correlation coefficient and all of the posterior moments are analytic. Our results are available in the open-source software package JASP.

Keywords and Phrases: bivariate normal distribution, hypergeometric functions, reference priors.
MSC subject classifications: primary 62H20, 62E15, 62F15

1 Introduction

Pearson's product–moment correlation coefficient ρ is a measure of the linear dependency between two random variables. Its sampled version, commonly denoted by r, has been well studied by the founders of modern statistics such as Galton, Pearson, and Fisher. Based on geometrical insights, Fisher (1915, 1921) was able to derive the exact sampling distribution of r and established that this sampling distribution converges to a normal distribution as the sample size increases. Fisher's study of the correlation has led to the discovery of variance-stabilizing transformations, sufficiency (Fisher, 1920), and, arguably, the maximum likelihood estimator (Fisher, 1922; Stigler, 2007). Similar efforts were made in Bayesian statistics, which focus on inferring the unknown ρ from the data that were actually observed. This type of analysis requires the statistician to (i) choose a prior on the parameters, thus, also on ρ, and to (ii) calculate the posterior. Here we derive analytic posteriors for ρ given a large class of priors that include the recommendations of Jeffreys (1961), Lindley (1965), Bayarri (1981), and, more recently, Berger and Sun (2008) and Berger *et al.* (2015). Jeffreys's work on the correlation coefficient can also be found in the second edition of his book (Jeffreys, 1961), originally published in 1948; see Robert *et al.* (2009) for a modern re-read of Jeffreys's work. An earlier attempt at a Bayesian analysis of the correlation coefficient can be found in Jeffreys (1935). Before presenting the results, we first discuss some notations and recall the likelihood for the problem at hand.

*a.ly@uva.nl

Published by Wiley Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

2 Notation and result

Let $(X_1, X_2)'$ have a bivariate normal distribution with mean $\vec{\mu} = (\mu_1, \mu_2)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where σ_1^2 and σ_2^2 are the population variances of X_1 and X_2, and where ρ is

$$\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1\sigma_2} = \frac{E(X_1X_2) - \mu_1\mu_2}{\sigma_1\sigma_2}. \tag{1}$$

Pearson's correlation coefficient ρ measures the linear association between X_1 and X_2. In brief, the model is parametrized by the five unknowns $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Bivariate normal data consisting of n pairs of observations can be sufficiently summarized as $y = (n, \bar{x}_1, \bar{x}_2, s_1, s_2, r)$, where

$$r = \frac{\sum_{j=1}^{n}(x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2)}{ns_1s_2}$$

is the sample correlation coefficient, $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ the sample mean, and $s_i^2 = \frac{1}{n}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2$ the average sums of squares. The bivariate normal model implies that the observations y are functionally related to the parameters by the following likelihood function:

$$\begin{aligned} f(y \mid \theta) =& \left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-n} \\ & \times \exp\left(-\frac{n}{2(1-\rho^2)}\left[\frac{(\bar{x}_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(\bar{x}_1-\mu_1)(\bar{x}_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(\bar{x}_2-\mu_2)^2}{\sigma_2^2}\right]\right) \\ & \times \exp\left(-\frac{n}{2(1-\rho^2)}\left[\left(\frac{s_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{rs_1s_2}{\sigma_1\sigma_2}\right) + \left(\frac{s_2}{\sigma_2}\right)^2\right]\right). \end{aligned} \tag{2}$$

For inference we use the following class of priors:

$$\pi_\eta(\theta) \propto \underbrace{(1-\rho^2)^{\alpha-1}(1+\rho^2)^{\frac{\beta}{2}}}_{\pi_{\alpha,\beta}(\rho)} \underbrace{\sigma_1^{\gamma-1}}_{\pi_\gamma(\sigma_1)} \underbrace{\sigma_2^{\delta-1}}_{\pi_\delta(\sigma_2)}, \tag{3}$$

where η denotes the hyperparameters, that is, $\eta = (\alpha, \beta, \gamma, \delta)$. This class of priors is inspired by the one that José Bernardo (2015) used in his talk on reference priors for the bivariate normal distribution at the '11th International Workshop on Objective Bayes Methodology in honor of Susie Bayarri'. This class of priors contains certain recommended priors as special cases.

If we set $\alpha = 1, \beta = \gamma = \delta = 0$ in Equation (3), we retrieve the prior that Jeffreys recommended for both estimation and testing (Jeffreys, 1961, pp. 174–179 and 289–292). This recommendation is *not* the prior derived from Jeffreys's rule based on the Fisher information (e.g., Ly *et al.*, 2017), as discussed in Berger and Sun (2008). With $\alpha = 1, \beta = \gamma = \delta = 0$, thus, a uniform prior on ρ, Jeffreys showed that the marginal posterior for ρ is approximately proportional to $h_a(n, r \mid \rho)$, where

$$h_{\mathrm{a}}(n, r \mid \rho) = (1-\rho^2)^{\frac{n-1}{2}}(1-\rho r)^{\frac{3-2n}{2}},$$

represents the ρ-dependent part of the likelihood Equation (2) with $\theta_0 = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ integrated out. For n large enough, the function h_{a} is a good approximation to the true reduced likelihood $h_{\gamma,\delta}$ given below.[†]

If we set $\alpha = \beta = \gamma = \delta = 0$ in Equation (3), we retrieve Lindley's reference prior for ρ. Lindley (1965, pp. 214–221) established that the posterior of $\tanh^{-1}(\rho)$ is asymptotically normal with mean $\tanh^{-1}(r)$ and variance n^{-1}, which relates the Bayesian method of inference for ρ to that of Fisher. In Lindley's (1965, p. 216) derivation, it is explicitly stated that the likelihood with θ_0 integrated out cannot be expressed in terms of elementary functions. In his analysis, Lindley approximates the true reduced likelihood $h_{\gamma,\delta}$ with the same h_{a} that Jeffreys used before. Bayarri (1981) furthermore showed that with the choice $\gamma = \delta = 0$, the marginalization paradox (Dawid *et al.*, 1973) is avoided.

In their overview, Berger and Sun (2008) showed that for certain a, b with $\alpha = b/2 - 1$, $\beta = 0$, $\gamma = a - 2$, and $\delta = b - 1$, the priors in Equation (3) correspond to a subclass of the generalized Wishart distribution. Furthermore, a right-Haar prior (e.g., Sun and Berger, 2007) is retrieved when we set $\alpha = \beta = 0, \gamma = -1, \delta = 1$ in Equation (3). This right-Haar prior then has a posterior that can be constructed through simulations, that is, by simulating from a standard normal distribution and two chi-squared distributions (Berger and Sun, 2008, Table 1). This constructive posterior also corresponds to the fiducial distribution for ρ (e.g., Fraser, 1961; Hannig *et al.*, 2006). Another interesting case is given by $\alpha = 0, \beta = 1, \gamma = \delta = 0$, which corresponds to the one-at-a-time reference prior for σ_1 and σ_2; see also Jeffreys (1961 p. 187).

The analytic posteriors for ρ follow directly from exact knowledge of the reduced likelihood $h_{\gamma,\delta}(n, r \mid \rho)$, rather than its approximation used in previous work. We give full details, because we did not encounter this derivation in earlier work.

Theorem 1: **The reduced likelihood** $h_{\gamma,\delta}(n, r \mid \rho)$. If $|r| < 1$, $n > \gamma + 1$, and $n > \delta + 1$, then the likelihood $f(y \mid \theta)$ times the prior Equation (3) with $\theta_0 = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ integrated out is a function $f_{\gamma,\delta}$ that factors as

$$f_{\gamma,\delta}(y \mid \rho) = p_{\gamma,\delta}(y_0) h_{\gamma,\delta}(n, r \mid \rho). \tag{4}$$

The first factor is the marginal likelihood with ρ fixed at zero, which does not depend on r nor on ρ, that is,

$$\begin{aligned} p_{\gamma,\delta}(y_0) &= \int\int\int\int f(y \mid \theta_0, \rho = 0)\pi_\gamma(\sigma_1)\pi_\delta(\sigma_2)\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\sigma_1\mathrm{d}\sigma_2 \\ &= 2^{\frac{-\gamma-\delta-4}{2}} \frac{\pi^{1-n}}{n}(ns_1^2)^{\frac{1+\gamma-n}{2}}(ns_2^2)^{\frac{1+\delta-n}{2}}\Gamma\left(\tfrac{n-\gamma-1}{2}\right)\Gamma\left(\tfrac{n-\delta-1}{2}\right), \end{aligned} \tag{5}$$

[†]We thank an anonymous reviewer for clarifying how Jeffreys derived this approximation.

where $y_0 = (n, \bar{x}_1, \bar{x}_2, s_1, s_2)$. We refer to the second factor as the reduced likelihood, a function of ρ which is given by a sum of an even function and an odd function, that is, $h_{\gamma,\delta} = A_{\gamma,\delta} + B_{\gamma,\delta}$, where

$$A_{\gamma,\delta}(n, r \mid \rho) = (1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} {}_2F_1\left(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \tfrac{1}{2}\,;\, r^2\rho^2\right), \tag{6}$$

$$B_{\gamma,\delta}(n, r \mid \rho) = 2r\rho(1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} W_{\gamma,\delta}(n) {}_2F_1\left(\tfrac{n-\gamma}{2}, \tfrac{n-\delta}{2}\,;\, \tfrac{3}{2}\,;\, r^2\rho^2\right) \tag{7}$$

where $W_{\gamma,\delta}(n) = \left[\Gamma\left(\frac{n-\gamma}{2}\right)\Gamma\left(\frac{n-\delta}{2}\right)\right] \Big/ \left[\Gamma\left(\frac{n-\gamma-1}{2}\right)\Gamma\left(\frac{n-\delta-1}{2}\right)\right]$ and where ${}_2F_1$ denotes Gauss' hypergeometric function.

Proof. To derive $f_{\gamma,\delta}(y \mid \rho)$, we have to perform three integrals: (i) with respect to $\pi(\mu_1, \mu_2) \propto 1$, (ii) $\pi_\gamma(\sigma_1) \propto \sigma_1^{\gamma-1}$, and (iii) $\pi_\delta(\sigma_2) \propto \sigma_2^{\delta-1}$.

(i) The integral with respect to $\pi(\mu_1, \mu_2) \propto 1$ yields

$$f(y \mid \sigma_1, \sigma_2, \rho) = \frac{\left(2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2\right)^{1-n}}{n} \exp\left(\tfrac{-n}{2(1-\rho^2)}\left[\tfrac{s_1^2}{\sigma_1^2} - 2\rho\tfrac{rs_1s_2}{\sigma_1\sigma_2} + \tfrac{s_2^2}{\sigma_2}\right]\right), \tag{8}$$

where we abbreviated $f(y \mid \sigma_1, \sigma_2, \rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y \mid \theta_0, \rho)\mathrm{d}\mu_1\mathrm{d}\mu_2$. The factor $p_{\gamma,\delta}(y_0)$ follows directly by setting ρ to zero in Equation (8) and two independent gamma integrals with respect to σ_1 and σ_2 resulting in Equation (5). These gamma integrals cannot be used when ρ is not zero. For $f_{\gamma,\delta}(y \mid \rho)$, which is a function of ρ, we use results from special functions theory.

(ii) For the second integral, we collect only that part of Equation (8) that involves σ_1 into a function g, that is,

$$\int_0^\infty g(y \mid \sigma_1)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1 = \int_0^\infty \sigma_1^{\gamma-n}\exp\left(-\tfrac{ns_1^2}{2(1-\rho^2)}\tfrac{1}{\sigma_1^2} + \tfrac{ns_1s_2}{\sigma_2(1-\rho^2)}r\rho\tfrac{1}{\sigma_1}\right)\mathrm{d}\sigma_1.$$

The assumption $n > \gamma + 1$ and the substitution $u = \sigma_1^{-1}$ allow us to solve this integral using Lemma A.1, which we distilled from the Bateman manuscript project (Erdélyi *et al.*, 1954), with $a = \frac{ns_1^2}{2(1-\rho^2)}, b = -\frac{ns_1s_2}{(1-\rho^2)\sigma_2}r\rho$ and $c = n - \gamma - 1$. This yields

$$\int_0^\infty g(y \mid \sigma_1)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1 = 2^{\frac{n-\gamma-3}{2}}\left(\tfrac{1-\rho^2}{ns_1^2}\right)^{\frac{n-\gamma-1}{2}}\left[\mathring{A}_\gamma + \mathring{B}_\gamma\right], \tag{9}$$

where

$$\mathring{A}_\gamma = \Gamma\left(\tfrac{n-\gamma-1}{2}\right) {}_1F_1\left(\tfrac{n-\gamma-1}{2}\,;\, \tfrac{1}{2}\,;\, \tfrac{ns_2^2(r\rho)^2}{2(1-\rho^2)}\tfrac{1}{\sigma_2^2}\right), \tag{10}$$

$$\mathring{B}_\gamma = \sqrt{\tfrac{2ns_2^2(r\rho)^2}{(1-\rho^2)}}\sigma_2^{-1}\Gamma\left(\tfrac{n-\gamma}{2}\right) {}_1F_1\left(\tfrac{n-\gamma}{2}\,;\, \tfrac{3}{2}\,;\, \tfrac{ns_2^2(r\rho)^2}{2(1-\rho^2)}\tfrac{1}{\sigma_2^2}\right), \tag{11}$$

and where ${}_1F_1$ denotes the confluent hypergeometric function. The functions $\mathring{A}_\gamma$ and $\mathring{B}_\gamma$ are the even and odd solutions of Weber's differential equation in the variable $z = (r\rho)^2 \frac{ns_2^2}{2(1-\rho^2)\sigma_2^2}$, respectively.

(iii) With $f_\gamma(y \mid \sigma_2, \rho) = \int_0^\infty f(y \mid \sigma_1, \sigma_2, \rho)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1$, we see that $f_{\gamma,\delta}(y \mid \rho)$ follows from integrating σ_2 out of the following expression:

$$f_\gamma(y \mid \sigma_2, \rho)\pi_\delta(\sigma_2) = 2^{\frac{-n-\gamma-1}{2}} \frac{\pi^{1-n}}{n} (ns_1^2)^{\frac{1+\gamma-n}{2}} (1-\rho^2)^{\frac{-\gamma}{2}} \left[\breve{A}_\gamma(y \mid \sigma_2, \rho) + \breve{B}_\gamma(y \mid \sigma_2, \rho)\right],$$

where

$$\begin{aligned}\breve{A}_\gamma =& \Gamma\left(\tfrac{n-\gamma-1}{2}\right) \overbrace{\sigma_2^{\delta-n} e^{-\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}} {}_1F_1\left(\tfrac{n-\gamma-1}{2} ; \tfrac{1}{2} ; (r\rho)^2 \tfrac{ns_2^2}{2(1-\rho^2)} \tfrac{1}{\sigma_2^2}\right)}^{k(n,r \mid \rho, \sigma_2)}, \\ \breve{B}_\gamma =& \left(\tfrac{2ns_2^2}{1-\rho^2}\right)^{\frac{1}{2}} r\rho\Gamma\left(\tfrac{n-\gamma}{2}\right) \underbrace{\sigma_2^{\delta-n-1} e^{-\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}} {}_1F_1\left(\tfrac{n-\gamma}{2} ; \tfrac{3}{2} ; (r\rho)^2 \tfrac{ns_2^2}{2(1-\rho^2)} \tfrac{1}{\sigma_2^2}\right)}_{l(n,r \mid \rho, \sigma_2)}.\end{aligned} \tag{12}$$

Hence, the last integral with respect to σ_2 only involves the functions k and l in Equation (12). The assumption $n > \delta+1$ and the substitution $t = \frac{ns_2^2}{2(1-\rho^2)}\sigma_2^{-2}$, thus, $\mathrm{d}\sigma_2 = -\frac{1}{2}\sqrt{\frac{ns_2^2}{2(1-\rho^2)}} t^{-\frac{3}{2}}\mathrm{d}t$ allow us to solve this integral using Equation (7.621.4) from Gradshteyn and Ryzhik (2007, p. 822) with $s = 1$, $\tilde{k} = (r\rho)^2$. This yields

$$\begin{aligned}\int_0^\infty k(n, r \mid \rho, \sigma_2)\mathrm{d}\sigma_2 &= 2^{\frac{n-\delta-3}{2}} \left(\tfrac{1-\rho^2}{ns_2^2}\right)^{\frac{n-\delta-1}{2}} \Gamma\left(\tfrac{n-\delta-1}{2}\right) \\ &\quad \times {}_2F_1\left(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2} ; \tfrac{1}{2} ; r^2\rho^2\right), \\ \int_0^\infty l(n, r \mid \rho, \sigma_2)\mathrm{d}\sigma_2 &= 2^{\frac{n-\delta-2}{2}} \left(\tfrac{1-\rho^2}{ns_2^2}\right)^{\frac{n-\delta}{2}} \Gamma\left(\tfrac{n-\delta}{2}\right) {}_2F_1\left(\tfrac{n-\gamma}{2}, \tfrac{n-\delta}{2} ; \tfrac{3}{2} ; r^2\rho^2\right).\end{aligned}$$

After we combine the results, we see that $f_{\gamma,\delta}(y \mid \rho) = \tilde{A}_{\gamma,\delta}(y \mid \rho) + \tilde{B}_{\gamma,\delta}(y \mid \rho)$, where

$$\begin{aligned}\frac{\tilde{A}_{\gamma,\delta}(y \mid \rho)}{p_{\gamma,\delta}(y_0)} &= (1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} {}_2F_1\left(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \tfrac{1}{2}\,;\, r^2\rho^2\right),\\ \frac{\tilde{B}_{\gamma,\delta}(y \mid \rho)}{p_{\gamma,\delta}(y_0)} &= 2r\rho(1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} W_{\gamma,\delta}(n)\, {}_2F_1\left(\tfrac{n-\gamma}{2}, \tfrac{n-\delta}{2}\,;\, \tfrac{3}{2}\,;\, r^2\rho^2\right).\end{aligned}$$

Hence, $f_{\gamma,\delta}(y \mid \rho)$ is of the asserted form. Note that $A_{\gamma,\delta} = \frac{\tilde{A}_{\gamma,\delta}(y \mid \rho)}{p_{\gamma,\delta}(y_0)}$ is even, while $\frac{\tilde{B}_{\gamma,\delta}(y \mid \rho)}{p_{\gamma,\delta}(y_0)}$ is an odd function of ρ.

□

This main theorem confirms Lindley's insights; $h_{\gamma,\delta}(n, r \mid \rho)$ is indeed not expressible in terms of elementary functions, and the prior on ρ is updated by the data only through its sampled version r and the sample size n. As a result, the marginal likelihood for data y then factors into $p_\eta(y) = p_{\gamma,\delta}(y_0)p_{\alpha,\beta}(n, r\,;\, \gamma, \delta)$, where $p_{\alpha,\beta}(n, r\,;\, \gamma, \delta) = \int h_{\gamma,\delta}(n, r \mid \rho)\pi_{\alpha,\beta}(\rho) d\rho$ is the normalizing constant of the marginal posterior of ρ. More importantly, the fact that the reduced likelihood is the sum of an even function and an odd function allows us to fully characterize the posterior distribution of ρ for the priors Equation (3) in terms of its moments. These moments are easily computed, as the prior $\pi_{\alpha,\beta}(\rho)$ itself is symmetric around zero. Furthermore, the prior $\pi_{\alpha,\beta}(\rho)$ can be normalized as

$$\pi_{\alpha,\beta}(\rho) = \frac{(1-\rho^2)^{\alpha-1}(1+\rho^2)^{\frac{\beta}{2}}}{\mathcal{B}\left(\frac{1}{2}, \alpha\right) {}_2F_1\left(-\frac{\beta}{2}, \frac{1}{2}\,;\, \frac{1}{2}+\alpha\,;\, -1\right)}, \tag{13}$$

where $\mathcal{B}(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$ denotes the beta function. The case with $\beta = 0$ is also known as the (symmetric) stretched beta distribution on $(-1, 1)$ and leads to Lindley's reference prior when we ignore the normalization constant, that is, $\mathcal{B}\left(\frac{1}{2}, \alpha\right)$, and, subsequently, let $\alpha \to 0$.

Corollary 1: **Characterization of the marginal posteriors of** ρ. If $n > \gamma + \delta - 2\alpha + 1$, then the main theorem implies that the marginal likelihood with all the parameters integrated out factors as $p_\eta(y) = p_{\gamma,\delta}(y_0)p_{\alpha,\beta}(n, r\,;\, \gamma, \delta)$ where

$$p_{\alpha,\beta}(n, r\,;\, \gamma, \delta) = \int_{-1}^{1} h_{\gamma,\delta}(n, r \mid \rho)\pi_{\alpha,\beta}(\rho)\, \mathrm{d}\rho = \int_{-1}^{1} A_{\gamma,\delta}(n, r \mid \rho)\pi_{\alpha,\beta}(\rho) \mathrm{d}\rho, \tag{14}$$

defines the normalizing constant of the marginal posterior for ρ. Observe that the integral involving $B_{\gamma,\delta}$ is zero, because $B_{\gamma,\delta}$ is odd on $(-1, 1)$. More generally, the kth posterior moment of ρ is

$$E(\rho^k \mid n, r) = \begin{cases} \frac{1}{p_{\alpha,\beta}(n,r\,;\,\gamma,\delta)} \int_{-1}^{1} \rho^k A_{\gamma,\delta}(n, r \mid \rho)\pi_{\alpha,\beta}(\rho)\mathrm{d}\rho & \text{if } k \text{ is even,} \\ \frac{1}{p_{\alpha,\beta}(n,r\,;\,\gamma,\delta)} \int_{-1}^{1} \rho^k B_{\gamma,\delta}(n, r \mid \rho)\pi_{\alpha,\beta}(\rho)\mathrm{d}\rho & \text{if } k \text{ is odd.} \end{cases} \tag{15}$$

These posterior moments define the series

$$E(\rho^k \mid n, r) = \begin{cases} \frac{1}{C_{\alpha,\beta}} \sum_{m=0}^{\infty} \frac{\left(\frac{n-\gamma-1}{2}\right)_m \left(\frac{n-\delta-1}{2}\right)_m}{\left(\frac{1}{2}\right)_m m!} a_{k,m} r^{2m} & \text{if } k \text{ is even}, \\ \frac{2W_{\gamma,\delta}(n)}{C_{\alpha,\beta}} \sum_{m=0}^{\infty} \frac{\left(\frac{n-\gamma}{2}\right)_m \left(\frac{n-\delta}{2}\right)_m}{\left(\frac{3}{2}\right)_m m!} b_{k,m} r^{2m+1} & \text{if } k \text{ is odd}, \end{cases} \tag{16}$$

where $C_{\alpha,\beta} = \mathcal{B}\left(\frac{1}{2}, \alpha\right) {}_2F_1\left(\frac{-\beta}{2}, \frac{1}{2} ; \alpha + \frac{1}{2} ; -1\right)$ is the normalization constant of the prior Equation (13), $W_{\gamma,\delta}(n)$ is the ratios of gamma functions as defined under Equation (7), and $(x)_m = \frac{\Gamma(x+m)}{\Gamma(x)} = x(x+1)(x+2)\ldots(x+m-1)$ refers to the Pochhammer symbol for rising factorials. The terms $a_{k,m}$ and $b_{k,m}$ are

$$\begin{aligned} a_{k,m} &= \mathcal{B}\left(\tfrac{1}{2} + \tfrac{k+2m}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\right) {}_2F_1\left(\tfrac{-\beta}{2}, \tfrac{k+2m+1}{2} ; \tfrac{k+2m+2\alpha+n-\gamma-\delta}{2} ; -1\right), \\ b_{k,m} &= \mathcal{B}\left(\tfrac{1}{2} + \tfrac{k+2m+1}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\right) {}_2F_1\left(\tfrac{-\beta}{2}, \tfrac{k+2m+2}{2} ; \tfrac{k+2m+2\alpha+n-\gamma-\delta+1}{2} ; -1\right). \end{aligned}$$

The series defined in Equation (16) are hypergeometric when β is a non-negative integer.

PROOF. The series $E(\rho^k \mid n, r)$ result from term-wise integration of the hypergeometric functions in $A_{\gamma,\delta}$ and $B_{\gamma,\delta}$. The assumption $n > \gamma + \delta - 2\alpha + 1$ and the substitution $x = \rho^2$ allow us to solve these integrals using Equation (3.197.8) in GRADSHTEYN AND RYZHIK (2007, p. 317) with their $\tilde{\alpha} = 1$, $u = 1$, $\lambda = \frac{\beta}{2}$, $\mu = \alpha + \frac{n-\gamma-\delta-1}{2}$ and $\nu = \frac{1}{2} + \frac{k+2m}{2}$ when k is even, while we use $\nu = \frac{1}{2} + \frac{k+2m+1}{2}$ when k is odd. A direct application of the ratio test shows that the series converge when $|r| < 1$. □

3 Analytic posteriors for the case $\beta = 0$

For most of the priors discussed earlier, we have $\beta = 0$, which leads to the following simplification of the posterior.

COROLLARY 1: **Characterization of the marginal posteriors of ρ, when $\beta = 0$.** If $n > \gamma + \delta - 2\alpha + 1$ and $|r| < 1$, then the marginal posterior for ρ is

$$\begin{aligned} \pi(\rho \mid n, r) = & \frac{(1-\rho^2)^{\frac{2\alpha+n-\gamma-\delta-3}{2}}}{p_\alpha(n, r ; \gamma, \delta)\mathcal{B}\left(\frac{1}{2}, \alpha\right)} \\ & \times \left[{}_2F_1\left(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2} ; \tfrac{1}{2} ; r^2\rho^2\right) + 2r\rho W_{\gamma,\delta}(n)\, {}_2F_1\left(\tfrac{n-\gamma}{2}, \tfrac{n-\delta}{2} ; \tfrac{3}{2} ; r^2\rho^2\right)\right], \end{aligned} \tag{17}$$

where $p_\alpha(n, r\,;\, \gamma, \delta)$ refers to the normalizing constant of the (marginal) posterior of ρ, which is given by

$$p_\alpha(n, r\,;\, \gamma, \delta) = \mathcal{B}\big(\tfrac{1}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\big)\,{}_2F_1\big(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \alpha + \tfrac{n-\gamma-\delta}{2}\,;\, r^2\big)/\mathcal{B}\big(\tfrac{1}{2}, \alpha\big).$$

More generally, when $\beta = 0$, the kth posterior moment is

$$\frac{\mathcal{B}\big(\tfrac{1}{2} + \tfrac{k}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\big)\,{}_3F_2\big(\tfrac{k+1}{2}, \tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \tfrac{1}{2}, \tfrac{k+2\alpha+n-\gamma-\delta}{2}\,;\, r^2\big)}{\mathcal{B}\big(\tfrac{1}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\big)\,{}_2F_1\big(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \tfrac{2\alpha+n-\gamma-\delta}{2}\,;\, r^2\big)},$$

when k is even, and

$$2rW_{\gamma,\delta}(n)\frac{\mathcal{B}\big(\tfrac{1}{2} + \tfrac{k+1}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\big)\,{}_3F_2\big(\tfrac{k+2}{2}, \tfrac{n-\gamma}{2}, \tfrac{n-\delta}{2}\,;\, \tfrac{3}{2}, \tfrac{k+2\alpha+n-\gamma-\delta+1}{2}\,;\, r^2\big)}{\mathcal{B}\big(\tfrac{1}{2}, \alpha + \tfrac{n-\gamma-\delta-1}{2}\big)\,{}_2F_1\big(\tfrac{n-\gamma-1}{2}, \tfrac{n-\delta-1}{2}\,;\, \tfrac{2\alpha+n-\gamma-\delta}{2}\,;\, r^2\big)},$$

when k is odd.

Proof. The assumption $n > \gamma + \delta - 2\alpha + 1$ and the substitution $x = \rho^2$ allow us to use Equation (7.513.12) in Gradshteyn and Ryzhik (2007, p. 814) with $\mu = \alpha + \frac{n-\gamma-\delta-1}{2}$ and $\nu = \frac{1}{2} + \frac{k}{2}$ when k is even, while we use $\nu = \frac{1}{2} + \frac{k+1}{2}$ when k is odd. The normalizing constant of the posterior $p_\alpha(n, r\,;\, \gamma, \delta)$ is a special case with $k = 0$. □

The marginal posterior for ρ updated from the generalized Wishart prior, the right-Haar prior, and Jeffreys's recommendation then follow from a direct substitution of the values for α, γ, and δ as discussed under Equation (3). Lindley's reference posterior for ρ is given by

$$\frac{{}_2F_1\big(\tfrac{n-1}{2}, \tfrac{n-1}{2}\,;\, \tfrac{1}{2}\,;\, r^2\rho^2\big) + 2r\rho W_{0,0}(n)\,{}_2F_1\big(\tfrac{n}{2}, \tfrac{n}{2}\,;\, \tfrac{3}{2}\,;\, r^2\rho^2\big)}{\mathcal{B}\big(\tfrac{1}{2}, \tfrac{n-1}{2}\big)\,{}_2F_1\big(\tfrac{n-1}{2}, \tfrac{n-1}{2}\,;\, \tfrac{n}{2}\,;\, r^2\big)}(1-\rho^2)^{\frac{n-3}{2}},$$

which follows from Equation (17) by setting $\gamma = \delta = 0$ and, subsequently, letting $\alpha \to 0$.

Lastly, for those who wish to sample from the posterior distribution, we suggest the use of an independence-chain Metropolis algorithm (Tierney, 1994) using Lindley's normal approximation of the posterior of $\tanh^{-1}(\rho)$ as the proposal. This method could be used when Pearson's correlation is embedded within a hierarchical model, as the posterior for ρ will then be a full conditional distribution within a Gibbs sampler. For $\alpha = 1$, $\beta = \gamma = \delta = 0$, $n = 10$ observations and $r = 0.6$, the acceptance rate of the independence-chain Metropolis algorithm was already well above 75%, suggesting a fast convergence of the Markov chain. For n larger, the acceptance rate further increases. The R code for the independence-chain Metropolis algorithm can be found on the first author's home page. In addition, this analysis is also implemented in the open-source software package jasp (https://jasp-stats.org/).

Acknowledgements

This work was supported by the starting grant 'Bayes or Bust' awarded by the European Research Council (grant number 283876). The authors thank Christian Robert, Fabian Dablander, Tom Koornwinder, and an anonymous reviewer for helpful comments that improved an earlier version of this manuscript.

Appendix A: A Lemma distilled from the Bateman project

LEMMA A.1. For $a, c > 0$, the following equality holds:

$$\int_0^\infty u^{c-1} \exp\big(- au^2 - bu \big) \mathrm{d}u = 2^{-1} a^{-\frac{c}{2}} \Big[\mathring{A}(a,b,c) + \mathring{B}(a,b,c) \Big], \tag{A.1}$$

that is, the integral is solved by the functions

$$\begin{aligned} \mathring{A}(a,b,c) =& \Gamma\big(\tfrac{c}{2}\big) {}_1F_1\big(\tfrac{c}{2}\,;\,\tfrac{1}{2}\,;\,\tfrac{b^2}{4a}\big), \\ \mathring{B}(a,b,c) =& -\tfrac{b}{\sqrt{a}} \Gamma\big(\tfrac{c+1}{2}\big) {}_1F_1\big(\tfrac{c+1}{2}\,;\,\tfrac{3}{2}\,;\,\tfrac{b^2}{4a}\big), \end{aligned} \tag{A.2}$$

which define the even and odd solutions to Weber's differential equation in the variable $z = \frac{b}{\sqrt{2a}}$, respectively.

PROOF. By ERDÉLYI *et al.* (1954, p 313, Equation (13)), we note that

$$\int_0^\infty u^{c-1} \exp\big(- au^2 - bu \big) \mathrm{d}v = (2a)^{\frac{-c}{2}} \Gamma(c) \exp\big(\tfrac{b^2}{8a}\big) D_{-c}\big(\tfrac{b}{\sqrt{2a}}\big), \tag{A.3}$$

where $D_\lambda(z)$ is WHITTAKER's (1902) parabolic cylinder function (ABRAMOWITZ and STEGUN, 1992). By virtue of Equation (4) on p. 117 of ERDÉLYI *et al.* (1981), we can decompose $D_\lambda(z)$ into a sum of an even function and an odd function. Replacing this decomposition for $D_\lambda(z)$ in Equation (A.3) and an application of the duplication formula of the gamma function yields the statement. □

References

ABRAMOWITZ, M. and I. A. STEGUN (eds) (1992), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc., New York. Reprint of the 1972 edition. MR1225604 (94b:00012).

BAYARRI, M. J. (1981), Bayesian inference on the correlation coefficient of a bivariate normal population, *Trabajos de Estadística y de Investigación Operativa* **32**, 18–31, MR697200 (84i:62047).

BERGER, J. O., J. M. BERNARDO and D SUN (2015), Overall objective priors, *Bayesian Analysis* **10**, 189–221.

BERGER, J. O. and D. SUN (2008), Objective priors for the bivariate normal model, *The Annals of Statistics* **36**, 963–982, MR2396821 (2009b:62058).

BERNARDO, J. M. (2015), An overall prior for the five-parameter normal distribution, 11th International Workshop on Objective Bayes Methodology dedicated to Susie Bayarri; Valencia.

DAWID, A. P., M. STONE and J. V. ZIDEK (1973), Marginalization paradoxes in Bayesian and structural inference, *Journal of the Royal Statistical Society. Series B. Methodological* **35**, 189–233, With discussion by D. J. Bartholomew, A. D. McLaren, D. V. Lindley, Bradley Efron, J. Dickey, G. N. Wilkinson, A. P.Dempster, D. V. Hinkley, M. R. Novick, Seymour Geisser, D. A. S. Fraser and A. Zellner, and a reply by A. P. Dawid, M. Stone, and J. V. Zidek. MR 0365805 (51 #2057).

ERDÉLYI, A., W. MAGNUS, F. OBERHETTINGER and F. G. TRICOMI (1954), *Tables of Integral Transforms*, vol. I, McGraw-Hill Book Company, Inc., New York-Toronto-London. Based, in part, on notes left by Harry Bateman. MR0061695 (15,868a).

ERDÉLYI, A., W. MAGNUS, F. OBERHETTINGER and F. G. TRICOMI (1981), *Higher Transcendental Functions*, vol. I, Robert E. Krieger Publishing Co., Inc., Melbourne, Fla. Based on notes left by Harry Bateman, Reprint of the 1953 original. MR698779 (84h:33001a).

FISHER, R. A. (1915), Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika* **10**, 507–521.

FISHER, R. A. (1920), A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error, *Monthly Notices of the Royal Astronomical Society* **80**, 758–770.

FISHER, R. A. (1921), On the "probable error" of a coefficient of correlation deduced from a small sample., *Metron* **1**, 3–32.

FISHER, R. A. (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society A* **222**, 309–368.

FRASER, D. A. S. (1961), On fiducial inference, *Annals of Mathematical Statistics* **32**, 661–676, MR0130755 (24 ##A614).

GRADSHTEYN, I. S. and I. M. RYZHIK (2007), *Table of Integrals, Series, and Products* Seventh, Elsevier/Academic Press, Amsterdam. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX). MR2360010 (2008g:00005).

HANNIG, J., H. IYER and P. PATTERSON (2006), Fiducial generalized confidence intervals, *Journal of the American Statistical Association* **101**, 254–269.

JEFFREYS, H. (1935), Some tests of significance, treated by the theory of probability, *Proceedings of the Cambridge Philosophical Society* **31**, 203–222.

JEFFREYS, H. (1961), *Theory of Probability* Third, Oxford Science Publications, The Clarendon Press, Oxford University Press, New York. MR745621 (85f:60005).

LINDLEY, D. V. (1965), *Introduction to Probability and Statistics From a Bayesian Viewpoint. Part II: Inference*, Cambridge University Press, New York. MR0168084 (29 #5349).

LY, A., M. MARSMAN, A. J. VERHAGEN, R. P. P. P. GRASMAN and E. J. WA-GENMAKERS (2017), A tutorial on Fisher information, arXiv preprint arXiv:1705.01064.

ROBERT, C. P., N. CHOPIN and J. ROUSSEAU (2009), Harold Jeffreys's theory of probability revisited, *Statistical Science. A Review Journal of the Institute of Mathematical Statistics* **24**, 141–172, MR2655841 (2011c:62002).

STIGLER, S. M. (2007), The epic story of maximum likelihood, *Statistical Science. A Review Journal of the Institute of Mathematical Statistics* **22**, 598–620.MR2410255

SUN, D. and J. O. BERGER (2007), Objective Bayesian analysis for the multivariate normal model, *Oxford Sci. Publ.*, Bayesian Statistics 8, Oxford, Oxford Univ. Press, 525–562.MR2433206 (2009m:62019)

TIERNEY, L. (1994), Markov chains for exploring posterior distributions, *The Annals of Statistics* **22**, 1701–1762, With discussion and a rejoinder by the author. MR1329166 (96m:60150).

WHITTAKER, E. T. (1902), On the functions associated with the parabolic cylinder in harmonic analysis, *Proceedings of the London Mathematical Society* **1**, 417–427.

Received: 8 February 2016. Revised: 7 April 2017.