Bayes factors for research workers

Ly, A.

Chapter 10

Analytic Posteriors for Pearson's Correlation Coefficient

Abstract

Pearson's correlation is one of the most common measures of linear dependence. Recently, Bernardo (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: Bivariate normal distribution, hypergeometric functions, reference priors.

10.1 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-stabilising 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,

This chapter is published online as: Ly, A., Marsman, M., and Wagenmakers, E.–J. (2017). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica.* doi: `http://dx.doi.org/10.1111/stan.12111`.

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.

10.2 Notation and result

Let $(X_1, X_2)'$ have a bivariate normal distribution with mean $\mu = (\mu_1, \mu_2)'$ 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, σ_2^2 are the population variances of X_1 and X_2, and where ρ is

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sigma_1\sigma_2} = \frac{E(X_1X_2) - \mu_1\mu_2}{\sigma_1\sigma_2}. \tag{10.2.1}$$

Pearson's correlation coefficient ρ measures the linear association between X_1 and X_2. In brief, the model is parameterised 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 summarised as $d = (n, \bar{x}_1, \bar{x}_2, s_1, s_2, r)$, where

$$r = \frac{1}{n}\sum_{j=1}^{n}\Big(\frac{x_{1j} - \bar{x}_1}{s_1}\Big)\Big(\frac{x_{2j} - \bar{x}_2}{s_2}\Big)$$

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 d are functionally related to the parameters by the following likelihood function

$$\begin{aligned} f(d \,|\, \theta) = {} & \Big(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\Big)^{-n} \\ & \times \exp\Big(-\frac{n}{2(1-\rho^2)}\Big[\Big(\frac{\bar{x}_1 - \mu_1}{\sigma_1^2}\Big)^2 - 2\rho\frac{(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)}{\sigma_1\sigma_2} + \Big(\frac{\bar{x}_2 - \mu_2}{\sigma_2^2}\Big)^2\Big]\Big) \\ & \times \exp\Big(-\frac{n}{2(1-\rho^2)}\Big[\Big(\frac{s_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{rs_1s_2}{\sigma_1\sigma_2}\Big) + \Big(\frac{s_2}{\sigma_2}\Big)^2\Big]\Big). \end{aligned} \tag{10.2.2}$$

For inference we use the following class of priors

$$\pi_\eta(\theta) \propto \underbrace{(1-\rho^2)^{\alpha-1}(1+\rho^2)^{\frac{\beta}{2}}}_{\pi_{\alpha,\beta}(\rho)}\underbrace{\sigma_1^{\gamma-1}}_{\pi_\gamma(\sigma_1)}\underbrace{\sigma_2^{\delta-1}}_{\pi_\delta(\sigma_2)}, \tag{10.2.3}$$

where η denotes the hyperparameters, that is, $\eta = (\alpha, \beta, \gamma, \delta)$. This class of priors is inspired by the one José Bernardo 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 Eq. (10.2.3), 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., 2017c), 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 Eq. (10.2.2) 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.[1]

If we set $\alpha = \beta = \gamma = \delta = 0$ in Eq. (10.2.3), 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 marginalisation 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 Eq. (10.2.3) correspond to a subclass of the generalised 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 Eq. (10.2.3). 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 10.2.1 (The reduced likelihood $h_{\gamma,\delta}(n, r \mid \rho)$)**.** *If* $|r| < 1$, $n > \gamma + 1$ *and* $n > \delta + 1$, *then the likelihood* $f(d \mid \theta)$ *times the prior Eq.* (10.2.3) *with the common parameters* $\theta_0 = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ *integrated out is a function* $f_{\gamma,\delta}$ *that factors as*

$$f_{\gamma,\delta}(d \mid \rho) = p_{\gamma,\delta}(d_0) h_{\gamma,\delta}(n, r \mid \rho). \tag{10.2.4}$$

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

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

$$\begin{aligned} p_{\gamma,\delta}(d_0) &= \int\int\int\int f(d \,|\, \theta_0, \rho = 0)\pi_\gamma(\sigma_1)\pi_\delta(\sigma_2)\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\sigma_1\mathrm{d}\sigma_2 \qquad (10.2.5)\\ &= 2^{\frac{-\gamma-\delta-4}{2}}\frac{\pi^{1-n}}{n}(ns_1^2)^{\frac{1+\gamma-n}{2}}(ns_2^2)^{\frac{1+\delta-n}{2}}\Gamma\Big(\frac{n-\gamma-1}{2}\Big)\Gamma\Big(\frac{n-\delta-1}{2}\Big), \end{aligned}$$

where $d_0 = (n, \bar{x}_1, \bar{x}_2, s_1, s_2)$. We refer to the second factor as the reduced likelihood, a function of ρ which is given by a sum of an even and an odd function, that is, $h_{\gamma,\delta} = A_{\gamma,\delta} + B_{\gamma,\delta}$ where

$$A_{\gamma,\delta}(n, r \,|\, \rho) = (1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}}\ {}_2F_1\Big(\frac{n-\gamma-1}{2}, \frac{n-\delta-1}{2}\ ;\ \frac{1}{2}\ ;\ r^2\rho^2\Big), \qquad (10.2.6)$$

$$B_{\gamma,\delta}(n, r \,|\, \rho) = 2r\rho(1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}}W_{\gamma,\delta}(n)\ {}_2F_1\Big(\frac{n-\gamma}{2}, \frac{n-\delta}{2}\ ;\ \frac{3}{2}\ ;\ r^2\rho^2\Big), \qquad (10.2.7)$$

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

Proof. To derive $f_{\gamma,\delta}(d \,|\, \rho)$ 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

$$\begin{aligned} f(d \,|\, \sigma_1, \sigma_2, \rho) &= \frac{\Big(2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2\Big)^{1-n}}{n} \qquad (10.2.8)\\ &\quad\times \exp\Big(\frac{-n}{2(1-\rho^2)}\Big[\frac{s_1^2}{\sigma_1^2} - 2\rho\frac{rs_1s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2}\Big]\Big), \end{aligned}$$

where we abbreviated $f(d \,|\, \sigma_1, \sigma_2, \rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(d \,|\, \theta_0, \rho)\mathrm{d}\mu_1\mathrm{d}\mu_2$. The factor $p_{\gamma,\delta}(d_0)$ follows directly by setting ρ to zero in Eq. (10.2.8) and two independent gamma integrals with respect to σ_1 and σ_2 resulting in Eq. (10.2.5). These gamma integrals cannot be used when ρ is not zero. For $f_{\gamma,\delta}(d \,|\, \rho)$ which is a function of ρ, we use results from special functions theory.
(ii) For the second integral, we collect only that part of Eq. (10.2.8) that involves σ_1 into a function g, that is,

$$\int_0^\infty g(d \,|\, \sigma_1)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1 = \int_0^\infty \sigma_1^{\gamma-n}\exp\Big(-\frac{ns_1^2}{2(1-\rho^2)}\frac{1}{\sigma_1^2} + \frac{ns_1s_2}{\sigma_2(1-\rho^2)}r\rho\frac{1}{\sigma_1}\Big)\mathrm{d}\sigma_1.$$

The assumption $n > \gamma + 1$ and the substitution $u = \sigma_1^{-1}$ allow us to solve this integral using Lemma 10.A.1, which we distilled from the Bateman manuscript project (Bateman 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

$$\int_0^\infty g(d \,|\, \sigma_1)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1 = 2^{\frac{n-\gamma-3}{2}}\Big(\frac{1-\rho^2}{ns_1^2}\Big)^{\frac{n-\gamma-1}{2}}\Big[\mathring{A}_\gamma + \mathring{B}_\gamma\Big], \qquad (10.2.9)$$

where

$$\mathring{A}_\gamma = \Gamma\Big(\frac{n-\gamma-1}{2}\Big)\ {}_1F_1\Big(\frac{n-\gamma-1}{2}\,;\,\frac{1}{2}\,;\,\frac{ns_2^2(r\rho)^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}\Big), \tag{10.2.10}$$

$$\mathring{B}_\gamma = \sqrt{\frac{2ns_2^2(r\rho)^2}{(1-\rho^2)}}\sigma_2^{-1}\Gamma\Big(\frac{n-\gamma}{2}\Big)\ {}_1F_1\Big(\frac{n-\gamma}{2}\,;\,\frac{3}{2}\,;\,\frac{ns_2^2(r\rho)^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}\Big), \tag{10.2.11}$$

and where ${}_1F_1$ denotes the confluent hypergeometric function. The functions $\mathring{A}_\gamma$ and $\mathring{B}_\gamma$ are the even and odd solution 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 $f_\gamma(d\,|\,\sigma_2,\rho) = \int_0^\infty f(d\,|\,\sigma_1,\sigma_2,\rho)\pi_\gamma(\sigma_1)\mathrm{d}\sigma_1$, we see that $f_{\gamma,\delta}(d\,|\,\rho)$ follows from integrating σ_2 out of the following expression

$$\begin{aligned} f_\gamma(d\,|\,\sigma_2,\rho)\pi_\delta(\sigma_2) = {} & 2^{\frac{-n-\gamma-1}{2}}\frac{\pi^{1-n}}{n}(ns_1^2)^{\frac{1+\gamma-n}{2}}(1-\rho^2)^{\frac{-\gamma}{2}} \\ & \times\Big[\breve{A}_\gamma(d\,|\,\sigma_2,\rho) + \breve{B}_\gamma(d\,|\,\sigma_2,\rho)\Big], \end{aligned} \tag{10.2.12}$$

where

$$\begin{aligned} \breve{A}_\gamma &= \Gamma\Big(\frac{n-\gamma-1}{2}\Big)k(d\,|\,\rho,\sigma_2) \\ \breve{B}_\gamma &= \Big(\frac{2ns_2^2}{1-\rho^2}\Big)^{\frac{1}{2}} r\rho\Gamma\Big(\frac{n-\gamma}{2}\Big)l(d\,|\,\rho,\sigma_2), \end{aligned} \tag{10.2.13}$$

and where

$$k(d\,|\,\rho,\sigma_2) = \sigma_2^{\delta-n}e^{-\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}}\ {}_1F_1\Big(\frac{n-\gamma-1}{2}\,;\,\frac{1}{2}\,;\,(r\rho)^2\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}\Big) \tag{10.2.14}$$

$$l(d\,|\,\rho,\sigma_2) = \sigma_2^{\delta-n-1}e^{-\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}}\ {}_1F_1\Big(\frac{n-\gamma}{2}\,;\,\frac{3}{2}\,;\,(r\rho)^2\frac{ns_2^2}{2(1-\rho^2)}\frac{1}{\sigma_2^2}\Big) \tag{10.2.15}$$

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

$$\begin{aligned} \int_0^\infty k(d\,|\,\rho,\sigma_2)\mathrm{d}\sigma_2 = {} & 2^{\frac{n-\delta-3}{2}}\Big(\frac{1-\rho^2}{ns_2^2}\Big)^{\frac{n-\delta-1}{2}}\Gamma\Big(\frac{n-\delta-1}{2}\Big) \\ & \times {}_2F_1\Big(\frac{n-\gamma-1}{2},\frac{n-\delta-1}{2}\,;\,\frac{1}{2}\,;\,r^2\rho^2\Big), \end{aligned} \tag{10.2.16}$$

$$\begin{aligned} \int_0^\infty l(d\,|\,\rho,\sigma_2)\mathrm{d}\sigma_2 = {} & 2^{\frac{n-\delta-2}{2}}\Big(\frac{1-\rho^2}{ns_2^2}\Big)^{\frac{n-\delta}{2}}\Gamma\Big(\frac{n-\delta}{2}\Big) \\ & \times {}_2F_1\Big(\frac{n-\gamma}{2},\frac{n-\delta}{2}\,;\,\frac{3}{2}\,;\,r^2\rho^2\Big). \end{aligned} \tag{10.2.17}$$

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

$$\frac{\tilde{A}_{\gamma,\delta}(d \,|\, \rho)}{p_{\gamma,\delta}(d_0)} = (1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} \, {}_2F_1\Big(\frac{n-\gamma-1}{2}, \frac{n-\delta-1}{2} \,;\, \frac{1}{2} \,;\, r^2\rho^2\Big),$$

$$\frac{\tilde{B}_{\gamma,\delta}(d \,|\, \rho)}{p_{\gamma,\delta}(d_0)} = 2r\rho(1-\rho^2)^{\frac{n-\gamma-\delta-1}{2}} W_{\gamma,\delta}(n) \, {}_2F_1\Big(\frac{n-\gamma}{2}, \frac{n-\delta}{2} \,;\, \frac{3}{2} \,;\, r^2\rho^2\Big).$$

Hence, $f_{\gamma,\delta}(d \,|\, \rho)$ is of the asserted form. Note that $A_{\gamma,\delta} = \frac{\tilde{A}_{\gamma,\delta}(d \,|\, \rho)}{p_{\gamma,\delta}(d_0)}$ is even, while $\frac{\tilde{B}_{\gamma,\delta}(d \,|\, \rho)}{p_{\gamma,\delta}(d_0)}$ is an odd function of ρ. □

This main theorem confirms Lindley's insights; $h_{\gamma,\delta}(n, r \,|\, \rho)$ 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 d then factors into $p_\eta(d) = p_{\gamma,\delta}(d_0) p_{\alpha,\beta}(n, r \,;\, \gamma, \delta)$, where $p_{\alpha,\beta}(n, r \,;\, \gamma, \delta) = \int h_{\gamma,\delta}(n, r \,|\, \rho) \pi_{\alpha,\beta}(\rho) \mathrm{d}\rho$ is the normalising constant of the marginal posterior of ρ. More importantly, the fact that the reduced likelihood is the sum of an even and an odd function allows us to fully characterise the posterior distribution of ρ for the priors Eq. (10.2.3) 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 normalised as

$$\pi_{\alpha,\beta}(\rho) = \frac{(1-\rho^2)^{\alpha-1}(1+\rho^2)^{\frac{\beta}{2}}}{\mathcal{B}(\frac{1}{2}, \alpha) {}_2F_1(-\frac{\beta}{2}, \frac{1}{2} \,;\, \frac{1}{2} + \alpha \,;\, -1)}, \tag{10.2.18}$$

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 normalisation constant, i.e., $\mathcal{B}(\frac{1}{2}, \alpha)$, and, subsequently, let $\alpha \to 0$.

Corollary 10.2.1 (Characterisation of the marginal posteriors of ρ)**.** *If $n > \gamma + \delta - 2\alpha + 1$, then the main theorem implies that whenever the marginal likelihood with all the parameters integrated out factors as $p_\eta(d) = p_{\gamma,\delta}(d_0) p_{\alpha,\beta}(n, r \,;\, \gamma, \delta)$, where*

$$p_{\alpha,\beta}(n, r \,;\, \gamma, \delta) = \int_{-1}^{1} h_{\gamma,\delta}(n, r \,|\, \rho) \pi_{\alpha,\beta}(\rho) \, d\rho = \int_{-1}^{1} A_{\gamma,\delta}(n, r \,|\, \rho) \pi_{\alpha,\beta}(\rho) \, d\rho, \tag{10.2.19}$$

defines the normalising 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

$$E_{\alpha,\beta}(\rho^k \,|\, n, r \,;\, \gamma, \delta) = \begin{cases} \frac{1}{p_{\alpha,\beta}(n, r \,;\, \gamma, \delta)} \int_{-1}^{1} \rho^k A_{\gamma,\delta}(n, r \,|\, \rho) \pi_{\alpha,\beta}(\rho) \, d\rho & \textit{if } k \textit{ is even,} \\ \frac{1}{p_{\alpha,\beta}(n, r \,;\, \gamma, \delta)} \int_{-1}^{1} \rho^k B_{\gamma,\delta}(n, r \,|\, \rho) \pi_{\alpha,\beta}(\rho) \, d\rho & \textit{if } k \textit{ is odd.} \end{cases} \tag{10.2.20}$$

These posterior moments define the series

$$E_{\alpha,\beta}(\rho^k \,|\, n, r\,;\, \gamma, \delta) = \begin{cases} \frac{1}{C_{\alpha,\beta}} \sum_{m=0}^{\infty} \frac{(\frac{n-\gamma-1}{2})_m (\frac{n-\delta-1}{2})_m}{(\frac{1}{2})_m m!} a_{k,m} r^{2m} & \textit{if } k \textit{ is even,} \\ \frac{2W_{\gamma,\delta}(n)}{C_{\alpha,\beta}} \sum_{m=0}^{\infty} \frac{(\frac{n-\gamma}{2})_m (\frac{n-\delta}{2})_m}{(\frac{3}{2})_m m!} b_{k,m} r^{2m+1} & \textit{if } k \textit{ is odd,} \end{cases} \tag{10.2.21}$$

where $C_{\alpha,\beta} = \mathcal{B}(\frac{1}{2}, \alpha) {}_2F_1(\frac{-\beta}{2}, \frac{1}{2}\,;\, \alpha + \frac{1}{2}\,;\, -1)$ *is the normalisation constant of the prior Eq.* (10.2.18), $W_{\gamma,\delta}(n)$ *is the ratios of gamma functions as defined under Eq.* (10.2.7) *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*

$$a_{k,m} = \mathcal{B}\Big(\frac{1}{2} + \frac{k+2m}{2}, \alpha + \frac{n-\gamma-\delta-1}{2}\Big) \tag{10.2.22}$$
$$\times {}_2F_1\Big(\frac{-\beta}{2}, \frac{k+2m+1}{2}\,;\, \frac{k+2m+2\alpha+n-\gamma-\delta}{2}\,;\, -1\Big),$$
$$b_{k,m} = \mathcal{B}\Big(\frac{1}{2} + \frac{k+2m+1}{2}, \alpha + \frac{n-\gamma-\delta-1}{2}\Big) \tag{10.2.23}$$
$$\times {}_2F_1\Big(\frac{-\beta}{2}, \frac{k+2m+2}{2}\,;\, \frac{k+2m+2\alpha+n-\gamma-\delta+1}{2}\,;\, -1\Big).$$

The series defined in Eq. (10.2.21) *are hypergeometric when* β *is a non-negative integer.* ◇

Proof. The series $E_{\alpha,\beta}(\rho^k \,|\, n, r\,;\, \gamma, \delta)$ 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 $x = \rho^2$ allows us to solve these integrals using Eq. (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 $|r| < 1$. □

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

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

Corollary 10.3.1 (Characterisation 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*

$$\pi_\alpha(\rho \,|\, n, r\,;\, \gamma, \delta) = \frac{(1-\rho^2)^{\frac{2\alpha+n-\gamma-\delta-3}{2}}}{p_\alpha(n, r\,;\, \gamma, \delta)\mathcal{B}(\frac{1}{2}, \alpha)} \tag{10.3.1}$$
$$\times \Big[{}_2F_1\Big(\frac{n-\gamma-1}{2}, \frac{n-\delta-1}{2}\,;\, \frac{1}{2}\,;\, r^2\rho^2\Big)$$
$$+ 2r\rho W_{\gamma,\delta}(n) {}_2F_1\Big(\frac{n-\gamma}{2}, \frac{n-\delta}{2}\,;\, \frac{3}{2}\,;\, r^2\rho^2\Big)\Big],$$

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

$$p_\alpha(n, r\,;\,\gamma, \delta) = \mathcal{B}\Big(\frac{1}{2}, \alpha + \frac{n-\gamma-\delta-1}{2}\Big) \Big/ \mathcal{B}\Big(\frac{1}{2}, \alpha\Big) \tag{10.3.2}$$
$$\times {}_2F_1\Big(\frac{n-\gamma-1}{2}, \frac{n-\delta-1}{2}\,;\, \alpha + \frac{n-\gamma-\delta}{2}\,;\, r^2\Big).$$

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

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

when k is even, and

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

when k is odd. ◇

Proof. The assumption $n > \gamma + \delta - 2\alpha + 1$ and the substitution $x = \rho^2$ allows us to use Eq. (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 normalising 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 generalised 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 Eq. (10.2.3). Lindley's reference posterior for ρ is given by

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

which follows from Eq. (10.3.1) 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 (IMH; Tierney, 1994) with 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 IMH 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`.

10.A A lemma distilled from the Bateman Project

Lemma 10.A.1. *For $a, c > 0$ the following equality holds*

$$\int_0^\infty u^{c-1} \exp\Big(- au^2 - bu\Big) du = 2^{-1} a^{-\frac{c}{2}} \Big[\mathring{A}(a,b,c) + \mathring{B}(a,b,c) \Big], \qquad (10.A.1)$$

that is, the integral is solved by the functions

$$\begin{aligned} \mathring{A}(a,b,c) &= \Gamma\Big(\frac{c}{2}\Big)\, {}_1F_1\Big(\frac{c}{2}\,;\, \frac{1}{2}\,;\, \frac{b^2}{4a}\Big), \\ \mathring{B}(a,b,c) &= -\frac{b}{\sqrt{a}} \Gamma\Big(\frac{c+1}{2}\Big)\, {}_1F_1\Big(\frac{c+1}{2}\,;\, \frac{3}{2}\,;\, \frac{b^2}{4a}\Big), \end{aligned} \qquad (10.A.2)$$

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

Proof. By Bateman et al. (1954, p. 313, Eq. 13) we note that,

$$\int_0^\infty u^{c-1} \exp\Big(- au^2 - bu\Big) \mathrm{d}u = (2a)^{\frac{-c}{2}} \Gamma(c) \exp\Big(\frac{b^2}{8a}\Big) D_{-c}\Big(\frac{b}{\sqrt{2a}}\Big), \qquad (10.A.3)$$

where $D_\lambda(z)$ is Whittaker's (1902) parabolic cylinder function (Abramowitz and Stegun, 1964). By virtue of Eq. (4) on p. 117 of Bateman et al. (1953), we can decompose $D_\lambda(z)$ into a sum of an even and odd function. Replacing this decomposition for $D_\lambda(z)$ in Eq. (10.A.3) and an application of the duplication formula of the gamma function yields the statement. □