A simple analysis of the exact probability matching prior in the location-scale model

Thomas J. DiCiccio, Todd A. Kuffner & G. Alastair Young

Submit your article to this journal

View related articles

View Crossmark data

A simple analysis of the exact probability matching prior in the location-scale model

Thomas J. DiCiccio

Department of Social Statistics, Cornell University

e-mail: tjd9@cornell.edu

Todd A. Kuffner

Department of Mathematics, Washington University in St. Louis

e-mail: kuffner@wustl.edu

and

G. Alastair Young

Department of Mathematics, Imperial College London

London SW7 2AZ, U.K.

e-mail: alastair.young@imperial.ac.uk

Abstract

It has long been asserted that in univariate location-scale models, when concerned with inference for either the location or scale parameter, the use of the inverse of the scale parameter as a Bayesian prior yields posterior credible sets which have exactly the correct frequentist confidence set interpretation. This claim dates to at least Peers (1965), and has subsequently been noted by various authors, with varying degrees of justification. We present a simple, direct demonstration of the exact matching property of the posterior credible sets derived under use of this prior in the univariate location-scale model. This is done by establishing an equivalence between the conditional frequentist and posterior densities of the pivotal quantities on which conditional frequentist inferences are based.

1   A Brief History

The purpose of this note is to give a detailed and direct account of an exact probability matching result for the univariate location-scale model, i.e. a result in which a particular choice of prior for Bayesian inference results in posterior credible sets which have exactly the correct frequentist confidence set interpretation. The setting we consider is that of an independent and identically distributed sample $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ from the family $\sigma^{-1} f\{(y - \mu)/\sigma\}$ where $f(\cdot)$ is a known probability density function defined on $\mathbb{R}$, $-\infty < \mu < \infty$ and $\sigma > 0$. We will consider inference for a scalar interest parameter, regarding the other parameter as nuisance. Thus we will investigate the relevant frequentist coverage of the marginal posterior quantiles of μ and σ, separately.

Various authors, including Fisher (1934) and Fraser (1979), have argued that in location-scale models, from a frequentist perspective there are strong reasons to draw inferences conditionally on the observed value of an ancillary statistic. Consequently, as noted by DiCiccio et al. (2012), when considering probability matching priors, the correct frequentist inference to match is a conditional one. Suppose that σ is the interest parameter and denote by $\sigma_{B,1-\alpha} \equiv \sigma_{B,1-\alpha}(\pi(\mu, \sigma), \mathbf{Y})$ the $1 - \alpha$ marginal posterior quantile for σ under the prior $\pi(\mu, \sigma)$. A conditional probability matching prior, $\pi(\mu, \sigma)$, is one which satisfies

$$\mathrm{Pr}_{\mu,\sigma|\mathbf{A}=\mathbf{a}}\{\sigma \le \sigma_{B,1-\alpha}|\mathbf{A} = \mathbf{a}\} = 1 - \alpha + O(n^{-m/2})$$

for all $\alpha \in (0, 1)$ for $m = 2$ or 3 which correspond to first- or second-order matching, n is the sample size and $Pr_{\mu,\sigma|\mathbf{A}=\mathbf{a}}$ is the conditional frequentist probability under repeated sampling of $\mathbf{Y}$, conditioning on the observed value of an ancillary statistic $\mathbf{A}$. This states that the $1 - \alpha$ quantile of the marginal posterior density of σ under prior $\pi(\mu, \sigma)$ has conditional frequentist coverage probability $1 - \alpha$, to error of order $O(n^{-m/2})$. Simple application of the law of iterated expectations shows that a conditional matching prior is also an unconditional matching prior to the same order. An identical definition of a conditional probability matching prior when μ is the interest parameter results from reversing the roles of σ and μ in the above.

All smooth priors are probability matching in a weak sense ($m = 1$ in the above); this is a consequence of the equivalence, to $O(n^{-1/2})$, of frequentist and Bayesian normal approximation. Datta and Mukerjee (2004) and Datta and Sweeting (2005) provide thorough reviews of known results,

including both approximate and exact matching results in the single parameter and multiparameter settings.

The result that there is exact probability matching for inference about a scalar parameter in the location-scale model for the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ has been stated by various authors. The earliest reference is Peers (1965); others include Lawless (1972, 1982) and DiCiccio and Martin (1993). However, to the best of our knowledge, a direct, general demonstration of this result is missing from the literature.

Datta and Mukerjee (2004)[p. 26] note that in the univariate normal location-scale model, the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ is exact unconditional frequentist probability matching, regardless of whether μ or σ is the interest parameter. This is because, under this prior, the unconditional frequentist and posterior distributions of certain pivots coincide. Earlier references for this observation include Guttman (1970, Ch. 7), Box and Tiao (1973, Ch. 2), and Sun and Ye (1996). In the present note, we show that such a result is actually true quite generally.

Severini et al. (2002) proved a related result about exact matching for predictive highest posterior density regions in group transformation models, of which the multivariate location-scale model is a particular example considered. This result is due to an invariance property of the highest predictive density region, and is essentially an extension of the invariance results derived in Hora and Buehler (1966, 1967).

Here we are concerned with the conditional frequentist matching property of posterior credible sets for a scalar interest parameter, and present a detailed argument confirming the exact matching property of the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$. Note also that in the location-scale model, Jeffreys (1961) recommended use of this prior instead of the Jeffreys prior, $\pi(\mu, \sigma) \propto \sigma^{-2}$. Thus the results described here may be interpreted as support for his recommendation.

2 Demonstrating exactness

Property: *The prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ yields exact conditional probability matching in the univariate location-scale model whether μ or σ is the interest parameter.*

We verify the property by establishing an equivalence between, respectively, the marginal con-

ditional frequentist confidence limits and marginal posterior credible limits for the parameter of interest. These limits are derived from the joint conditional frequentist and joint posterior densities of suitable pivotal quantities. The motivation for this approach is that in the conditional frequentist framework, confidence sets are constructed using the conditional distributions of pivotals. In the location-scale model, a particular choice of pivotal quantities yields straightforward construction of confidence sets for either parameter directly from the marginal distribution of the corresponding pivotal. Bayesian and frequentist procedures for constructing credible and confidence sets using, respectively, the joint posterior and joint conditional frequentist densities of suitable pivotal quantities, are exactly the same. Thus to establish the result, it is sufficient to demonstrate that the Bayesian and frequentist frameworks base inference on the same joint density.

We first summarize the procedure for exact conditional frequentist inference as suggested by Fisher (1934) and more thoroughly examined by Fraser (1979, Ch. 6); additional details and references may be found in Lawless (1982, Appendix E) and Pace and Salvan (1997, Ch. 7). The joint density of the sample for given values of (μ, σ) is defined as

$$p_{\mathbf{y}}(\mathbf{y}; \mu, \sigma) = \sigma^{-n} \prod_{i=1}^{n} f\{(y_i - \mu)/\sigma\}.$$

It is assumed that the maximum likelihood estimators for (μ, σ), denoted by $(\hat{\mu}, \hat{\sigma})$, are unique and exist with probability one. The configuration statistic,

$$\mathbf{A} = (A_1, \ldots, A_n) = (\frac{Y_1 - \hat{\mu}}{\hat{\sigma}}, \ldots, \frac{Y_n - \hat{\mu}}{\hat{\sigma}}).$$

is an exact ancillary. This statistic is distribution constant, in the sense that its distribution does not depend on any unknown parameters, and only $n - 2$ elements of this random vector are functionally independent. To appreciate this last property, simply write the likelihood in the form $L(\mu, \sigma; \hat{\mu}, \hat{\sigma}, \mathbf{a})$, that is, in terms of the minimal sufficient statistic $(\hat{\mu}, \hat{\sigma}, \mathbf{a})$, and observe that the likelihood equations give two constraints involving the ancillary. In particular, A_{n-1} and A_n may be expressed in terms of $(A_1, \ldots, A_{n-2})$. Moreover, the quantity

$$(Q_1, Q_2) = (\frac{\hat{\mu} - \mu}{\hat{\sigma}}, \frac{\hat{\sigma}}{\sigma})$$

is pivotal with respect to the parameters (μ, σ) in the sense that the joint distribution of (Q_1, Q_2), conditional on the ancillary statistic, does not depend on (μ, σ).

Standard joint density manipulations involving transformations from the joint density of $(Y_1, \ldots, Y_n)$ to the joint density of $(\hat{\mu}, \hat{\sigma}, A_1, \ldots, A_{n-2})$ shows that the exact joint conditional frequentist density of (Q_1, Q_2), given $\mathbf{A} = \mathbf{a}$, is of the form

$$p_{Q_1, Q_2 | \mathbf{A}=\mathbf{a}}(q_1, q_2 | \mathbf{A} = \mathbf{a}) = c(\mathbf{a}) q_2^{n-2} p_{Q_1, Q_2}(q_1, q_2)$$

$$= c(\mathbf{a}) q_2^{n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i), \tag{1}$$

where the normalizing constant $c(\mathbf{a})$ depends on $\mathbf{a}$ only and is defined by

$$c(\mathbf{a}) \int_0^\infty \int_{-\infty}^\infty q_2^{n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i) dq_1 dq_2 = 1. \tag{2}$$

Exact conditional frequentist inference makes use of (1) to construct a confidence set for, respectively, μ or σ. Let $q_{1, F, \alpha}$ denote the α quantile of the conditional frequentist marginal distribution of q_1. That is,

$$\int_{-\infty}^{q_{1, F, \alpha}} \int_0^\infty p_{Q_1, Q_2 | \mathbf{A}=\mathbf{a}}(q_1, q_2 | \mathbf{A} = \mathbf{a}) dq_2 dq_1 = \alpha. \tag{3}$$

Similarly let $q_{2, F, \alpha}$ denote the α quantile of the conditional frequentist marginal distribution of q_2,

$$\int_0^{q_{2, F, \alpha}} \int_{-\infty}^\infty p_{Q_1, Q_2 | \mathbf{A}=\mathbf{a}}(q_1, q_2 | \mathbf{A} = \mathbf{a}) dq_1 dq_2 = \alpha. \tag{4}$$

Fix (μ, σ). Conditioning on $\mathbf{A} = \mathbf{a}$, the event $\{q_1 \geq q_{1, F, \alpha}\}$ is equivalent to the event $\{\hat{\mu} - \hat{\sigma} q_{1, F, \alpha} \geq \mu\}$. Also, the event $\{q_2 \geq q_{2, F, \alpha}\} \equiv \{\hat{\sigma} / q_{2, F, \alpha} \geq \sigma\}$. Thus, an upper $1 - \alpha$ one-sided conditional frequentist confidence limit for μ, say $\mu_{F, 1-\alpha}$ may be found directly from the corresponding limit for q_1 and similarly the limit for σ, say $\sigma_{F, 1-\alpha}$, may be obtained from the limit for q_2. Formally, under repeated sampling of $\mathbf{Y}$, $\Pr_{\mu, \sigma | \mathbf{A}=\mathbf{a}}\{\mu \leq \hat{\mu} - \hat{\sigma} q_{1, F, \alpha} | \mathbf{A} = \mathbf{a}\} = 1 - \alpha$ and $\Pr_{\mu, \sigma | \mathbf{A}=\mathbf{a}}\{\sigma \leq \hat{\sigma} / q_{2, F, \alpha} | \mathbf{A} = \mathbf{a}\} = 1 - \alpha$.

Turning to the Bayesian perspective, inference is conditioned on the full data $\mathbf{y}$. The joint posterior density $\pi(\mu, \sigma | \mathbf{Y} = \mathbf{y})$ is defined by

$$\pi(\mu, \sigma | \mathbf{Y} = \mathbf{y}) = \frac{\pi(\mu, \sigma) p(\mathbf{y}; \mu, \sigma)}{\int_0^\infty \int_{-\infty}^\infty \pi(\mu', \sigma') p(\mathbf{y}; \mu' \sigma') d\mu' d\sigma'}.$$

Expressing the likelihood in the form $L(\mu, \sigma; \hat{\mu}, \hat{\sigma}, \mathbf{a})$ yields

$$\pi(\mu, \sigma | \mathbf{Y} = \mathbf{y}) \propto \pi(\mu, \sigma) \sigma^{-n} \prod_{i=1}^{n} f\{\frac{\hat{\sigma}}{\sigma}(a_i + \frac{\hat{\mu} - \mu}{\hat{\sigma}})\},$$

and using the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$, we have the joint posterior density of the parameters

$$\pi(\mu, \sigma | \mathbf{Y} = \mathbf{y}) = s\sigma^{-n-1} \prod_{i=1}^{n} f\{\frac{\hat{\sigma}}{\sigma}(a_i + \frac{\hat{\mu} - \mu}{\hat{\sigma}})\}, \tag{5}$$

where the normalizing constant s is determined by

$$s \int_0^\infty \int_{-\infty}^\infty \sigma^{-n-1} \prod_{i=1}^{n} f\{\frac{\hat{\sigma}}{\sigma}(a_i + \frac{\hat{\mu} - \mu}{\hat{\sigma}})\}d\mu d\sigma = 1. \tag{6}$$

We now find the joint posterior density of (Q_1, Q_2) which, conditional on the data, is a one-to-one transformation of the parameters (μ, σ), treated as random quantities in the Bayesian framework. In order to show that the posterior density of (Q_1, Q_2) is exactly equal to the conditional frequentist density given by (1), the relationship between the normalizing constants must be discovered. We could start with the posterior density in (5), find the posterior density of (Q_1, Q_2) via the usual route and finally solve for the relationship between normalizing constants, but we instead choose to do everything at once.

We integrate (6) using the substitution $(Q_1, Q_2) = \boldsymbol{\varphi}(\mu, \sigma) = ((\hat{\mu} - \mu)/\hat{\sigma}, \hat{\sigma}/\sigma)$ and, by setting this equal to (2), establish a relationship between s and $c(\mathbf{a})$. Explicitly,

$$\begin{aligned}
1 = \int_0^\infty \int_{-\infty}^\infty \pi(\mu, \sigma | \mathbf{Y} = \mathbf{y})d\mu d\sigma &= s \int_0^\infty \int_{-\infty}^\infty \sigma^{-n-1} \prod_{i=1}^{n} f\{\frac{\hat{\sigma}}{\sigma}(a_i + \frac{\hat{\mu} - \mu}{\hat{\sigma}})\}d\mu d\sigma \\
&= s \int_0^\infty \int_{-\infty}^\infty (\hat{\sigma}/q_2)^{-n-1} f(q_1 q_2 + q_2 a_i) |\det J| dq_1 dq_2 \\
&= s \int_0^\infty \int_{-\infty}^\infty \hat{\sigma}^{-n+1} q_2^{n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i) dq_1 dq_2,
\end{aligned}$$

where $|\det J| = (\hat{\sigma}/q_2)^2$ is the absolute value of the Jacobian determinant of $(\mu, \sigma)(q_1, q_2) = \boldsymbol{\varphi}^{-1}(q_1, q_2) = (\hat{\mu} - \hat{\sigma}q_1, \hat{\sigma}/q_2)$. Comparison with (2) yields the relationship between the normalizing constants,

$$s \equiv \hat{\sigma}^{n-1} c(\mathbf{a}). \tag{7}$$

Thus the joint posterior density of (Q_1, Q_2) is given by

$$\begin{aligned}
\pi(q_1, q_2 | \mathbf{Y} = \mathbf{y}) &= s q_2^{n+1} \hat{\sigma}^{-n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i) |\det J| \\
&= c(\mathbf{a}) q_2^{n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i). \tag{8}
\end{aligned}$$

Note that this is exactly equal to the joint conditional frequentist density given in (1).

The α quantiles of the marginal posterior distributions, denoted by $q_{1,B,\alpha}$ and $q_{2,B,\alpha}$, are respectively defined by

$$\int_{-\infty}^{q_{1,B,\alpha}} \int_{0}^{\infty} \pi(q_1, q_2 | \mathbf{Y} = \mathbf{y}) dq_2 dq_1 = \alpha, \tag{9}$$

and

$$\int_{0}^{q_{2,B,\alpha}} \int_{-\infty}^{\infty} \pi(q_1, q_2 | \mathbf{Y} = \mathbf{y}) dq_1 dq_2 = \alpha. \tag{10}$$

Comparison with (3) and (4) confirms that $q_{1,F,\alpha} = q_{1,B,\alpha}$ and $q_{2,F,\alpha} = q_{2,B,\alpha}$. The construction of credible sets when either μ or σ is the interest parameter exactly parallels the procedure in the conditional frequentist setting. In particular, the $1 - \alpha$ upper credible limits for μ and σ, denoted by $\mu_{B,1-\alpha}$ and $\sigma_{B,1-\alpha}$, satisfy

$$\text{Pr}_{\mu,\sigma|\mathbf{A}=\mathbf{a}}(\mu \leq \mu_{B,1-\alpha} | \mathbf{A} = \mathbf{a}) = 1 - \alpha$$

and

$$\text{Pr}_{\mu,\sigma|\mathbf{A}=\mathbf{a}}(\sigma \leq \sigma_{B,1-\alpha} | \mathbf{A} = \mathbf{a}) = 1 - \alpha,$$

i.e. the conditional frequentist coverage of the posterior credible set under prior $\pi(\mu, \sigma) \propto \sigma^{-1}$, is exactly $1 - \alpha$, whether μ or σ is the parameter of interest.

In summary, we have demonstrated that the formal equivalence between the frequentist conditional density of (Q_1, Q_2) and the Bayesian posterior density of (Q_1, Q_2) under the prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ allows us to deduce directly that this prior is exact conditional probability matching, when either μ or σ is the interest parameter. It follows immediately that this prior is an exact probability matching prior in the usual, unconditional sense. That Bayesian posterior quantiles under this prior act as frequentist confidence limits both conditionally and unconditionally is a result that holds quite generally, even when the marginal posterior of interest is analytically intractable.

References

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, New York: John Wiley.

Datta, G.S., and Mukerjee, R. (2004), *Probability Matching Priors: Higher Order Asymptotics*, New York, NY: Springer.

Datta, G.S., and Sweeting, T.J. (2005), "Probability matching priors," in *Handbook of Statistics* **25** *Bayesian Thinking: Modeling and Computation*, eds. D.K. Dey and C.R. Rao,, 91–114. North-Holland, Amsterdam.

DiCiccio, T. J., and Martin, M. A. (1993), "Simple modifications for signed roots of likelihood ratio statistics," Journal of the Royal Statistical Society, *Series B*, 55, 305–316.

DiCiccio, T. J., Kuffner, T. A., and Young, G. A. (2012), "Objective Bayes, conditional inference and the signed root likelihood ratio statistic," *Biometrika*, 99, 675–686.

Fisher, R.A. (1934), "Two new properties of mathematical likelihood," *Proceedings of the Royal Society of London A*, 134, 285–307.

Fraser, D. A. S. (1979), *Inference and Linear Models*, New York, NY: McGraw-Hill.

Guttman, I. (1970), *Statistical Tolerance Regions: Classical and Bayesian*, London: Charles Griffin.

Hora, R.B., and Buehler, R.J. (1966), "Fiducial theory and invariant estimation," *Annals of Mathematical Statistics*, 37, 643–656.

Hora, R.B., and Buehler, R.J. (1967), "Fiducial theory and invariant prediction," *Annals of Mathematical Statistics*, 38, 785–801.

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: Oxford University Press.

Lawless, J.F. (1972), "Conditional confidence interval procedures for the location and scale parameters of the Cauchy and logistic distributions," *Biometrika*, 59, 377–386.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York, NY: John Wiley & Sons.

Pace, L., and Salvan, A. (1997), *Principles of Statistical Inference from a Neo-Fisherian Perspective*, Singapore: World Scientific.

Peers, H.W. (1965), "On confidence points and Bayesian probability points in the case of several parameters," Journal of the Royal Statistical Society, *Series B*, 27, 9–16.

Severini, T.A., Mukerjee, R., and Ghosh, M. (2002), "On an exact probability matching property of right-invariant priors," *Biometrika*, 89, 952–957.

Sun, D., and Ye, K. (1996). "Frequentist validity of posterior quantiles for a two-parameter exponential family," *Biometrika*, 83, 55–65.