Biometrika (2010), **97**, 2, *pp.* 497–504
doi: 10.1093/biomet/asq002

Advance Access publication 15 April 2010
Printed in Great Britain

Miscellanea

Objective Bayes and conditional inference in exponential families

BY THOMAS J. DICICCIO

Department of Social Statistics, Cornell University, Ithaca, New York 14853, U.S.A.
tjd9@cornell.edu

AND G. ALASTAIR YOUNG

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
alastair.young@imperial.ac.uk

SUMMARY

Objective Bayes methodology is considered for conditional frequentist inference about a canonical parameter in a multi-parameter exponential family. A condition is derived under which posterior Bayes quantiles match the conditional frequentist coverage to a higher-order approximation in terms of the sample size. This condition is on the model, not on the prior, and it ensures that any first-order probability matching prior in the unconditional sense automatically yields higher-order conditional probability matching. Objective Bayes methods are compared to parametric bootstrap and analytic methods for higher-order conditional frequentist inference.

Some key words: Bootstrap method; Conditional inference; Full exponential family; Nuisance parameter; Objective Bayes inference; Probability matching; Signed root likelihood ratio statistic.

1. INTRODUCTION

In Bayesian parametric inference, it is natural to consider the use of an objective prior in the absence of subjective prior information about the parameter of interest. This leads to posterior probability quantiles that have the correct frequentist interpretation, at least to some higher-order approximation in terms of the sample size. Such priors are termed probability matching priors. In the context of inference for a canonical parameter in a multi-dimensional exponential family model, however, the appropriate frequentist inference is a conditional one. In this paper, we consider a condition under which posterior quantiles have the correct conditional frequentist interpretation to higher order. It turns out that the condition for higher-order conditional frequentist accuracy reduces to a condition on the model, not on the prior: when the condition is satisfied, any prior that is first-order probability matching in the unconditional sense automatically yields higher-order conditional probability matching. A key motivation is that the conceptually simple objective Bayes route may provide accurate approximation to more complicated frequentist procedures. We provide numerical illustrations involving the inverse Gaussian and gamma distributions, and discuss the relationship between the objective Bayes inference and both analytic and parametric bootstrap methods for approximate conditional frequentist inference.

Let $Y = (Y_1, \ldots, Y_n)$ be a random sample from an underlying continuous distribution having density $f(y;\theta)$, indexed by a d-dimensional parameter θ, and let $\psi = g(\theta)$ be a scalar parameter of interest. Without loss, we may suppose that $\theta = (\psi, \lambda)$, where λ is a nuisance parameter.

Let $\ell(\theta) = \ell(\theta; Y)$ be the loglikelihood for θ, and denote by $\hat\theta = (\hat\psi, \hat\lambda)$ the overall maximum likelihood estimator of θ. The likelihood ratio statistic is $w(\psi) = 2\{\ell(\hat\theta) - \ell(\psi, \hat\lambda_\psi)\}$, with $\hat\lambda_\psi$ the constrained maximum likelihood estimator of λ for fixed value of ψ. We will be concerned with the signed square root likelihood ratio statistic $r(\psi) = \text{sgn}(\hat{\psi} - \psi)w(\psi)^{1/2}$.

In the absence of subjective prior information about θ, it is natural to use a prior that leads to posterior probability limits for ψ that are also frequentist confidence limits, in the sense that

$$\text{pr}_\theta\left\{\psi \leqslant \psi^{(1-\alpha)}(\pi, Y)\right\} = 1 - \alpha + O\left(n^{-m/2}\right), \tag{1}$$

for each $0 < \alpha < 1$, where $m = 2$ or 3. Here, $\psi^{(1-\alpha)}(\pi, Y)$ is the $1 - \alpha$ quantile of the marginal posterior of ψ given data Y under prior $\pi(\psi, \lambda)$, and pr_θ denotes frequentist probability, under repeated sampling of Y, under parameter θ. If the condition (1) holds with $m = 2$, we speak of $\pi(\psi, \lambda)$ as a first-order probability matching prior, while if it holds with $m = 3$, we speak of $\pi(\psi, \lambda)$ as being a second-order probability matching prior. A second-order probability matching prior yields posterior limits that are frequentist limits of coverage error $O(n^{-3/2})$; this third-order frequentist accuracy is typically not straightforward to be achieved directly. For general scalar parameter models, Welch & Peers (1963) showed that Jeffreys' prior is first-order probability matching and gave conditions on the model under which second-order probability matching is attained. Peers (1965) considered the nuisance parameter case and gave conditions on the prior density under which first-order probability matching occurs; see also Tibshirani (1989). Mukerjee & Dey (1993) and Mukerjee & Ghosh (1997) gave conditions that ensure second-order probability matching in the nuisance parameter case; see Datta & Mukerjee (2004).

Suppose specifically that the loglikelihood is of the form

$$\ell(\theta) = \psi t(Y) + \lambda' c(Y) - k(\psi, \lambda) + b(Y), \tag{2}$$

so that the interest parameter ψ is a canonical parameter of a multi-parameter exponential family. Then the conditional sampling distribution of $T = t(Y)$ given observed value c of $C = c(Y)$ depends only on ψ, so that from a frequentist perspective, conditioning on the observed data value of C eliminates the nuisance parameter. The appropriate frequentist inference on ψ is based on the sampling distribution of T, given the observed value c; see, for example, Young & Smith (2005, Ch. 5). This sampling distribution is completely specified once ψ is fixed.

This conditional inference has the optimality property under unconditional repeated sampling of being uniformly most powerful unbiased; see Young & Smith (2005, § 7.2). In practice, however, the exact inference may be difficult, even impossible, to construct: the relevant conditional distribution typically requires awkward analytic or numerical calculations.

In the exponential family context, the appropriate frequentist inference to match is the conditional one. The requirement (1) should be replaced by that of conditional probability matching:

$$\text{pr}_\theta\left\{\psi \leqslant \psi^{(1-\alpha)}(\pi, Y) \mid C = c\right\} = 1 - \alpha + O\left(n^{-m/2}\right). \tag{3}$$

2. THEORETICAL DEVELOPMENT

Consider a continuous random variable Y with distribution belonging to a full exponential model of dimension d. Let the canonical parameter of the exponential model be denoted by $\theta = (\theta^1, \ldots, \theta^d)$ and suppose that $\psi = \theta^1$ is the scalar parameter of interest, with $\lambda = (\theta^2, \ldots, \theta^d)$ being a vector nuisance parameter. Write the density of Y as

$$f_Y(y; \theta) = \exp\{\theta^r t_r - k(\theta) + b(y)\}, \quad t_1 = t_1(y), \ldots, t_d = t_d(y). \tag{4}$$

Here, the summation convention is adopted, under which we automatically sum with respect to a letter represented both as a subscript and as a superscript. Based on a random sample $Y_1, \ldots, Y_n$ of size n, the loglikelihood function is, apart from a constant, $\ell(\theta) = n\{\theta^r \bar{t}_r - k(\theta)\}$, with $\bar{t}_r = \sum_{i=1}^n t_r(Y_i)/n$, which is of the form (2) assumed in our discussion above.

In the analysis that follows, derivatives will be denoted by subscripts: $\ell_r(\theta) = \partial\ell(\theta)/\partial\theta^r = n\{\bar{t}_r - k_r(\theta)\}$, where $k_r(\theta) = \partial k(\theta)/\partial\theta^r (r = 1, \ldots, d)$. We have $\ell_{rs}(\theta) = \partial^2\ell(\theta)/\partial\theta^r\theta^s = -nk_{rs}(\theta)$, where $k_{rs}(\theta) = \partial^2 k(\theta)/\partial\theta^r\theta^s$ $(r, s = 1, \ldots, d)$. In the formulae that follow, indices $r, s, t, \ldots$ are understood to run through $1, \ldots, d$. Let $\ell_{\theta\theta} = (\ell_{rs})$ be the $d \times d$ matrix with components $\ell_{rs}(\theta)$, and let $\ell_{\lambda\lambda}$ be the $(d-1) \times (d-1)$ submatrix corresponding to the nuisance parameter λ. Denote the inverse of $\ell_{\theta\theta}$ by $\ell^{\theta\theta} = (\ell^{rs})$. For ease of notation, write $\ell^{\psi\psi} = \ell^{11}$.

As before, let $\hat\theta = (\hat\theta^1, \ldots, \hat\theta^d)$ be the global maximum likelihood estimator of θ and let $\tilde\theta = \tilde\theta(\psi) = (\tilde\theta^1, \ldots, \tilde\theta^d)$ be the constrained maximum likelihood estimator of θ for a given value of $\psi = \theta^1$. Evaluation of functions of θ at $\hat\theta$ and $\tilde\theta$ will be denoted by $\hat{}$ and $\tilde{}$, respectively. We have $\tilde\ell_\lambda = 0$, and $-\hat\ell_{\theta\theta}$ is the observed information.

Our analysis follows the approach of Casella et al. (1995). For values ψ_0 such that $\hat\psi - \psi_0$ is of order $O(n^{-1/2})$, we have

$$\text{pr}(\psi \geqslant \psi_0 \mid Y) = \Phi\{r(\psi_0)\} + \varphi\{r(\psi_0)\}\{r(\psi_0)^{-1} - u_B(\psi_0)^{-1}\} + O\left(n^{-3/2}\right), \tag{5}$$

where Φ and φ are the standard normal distribution and density functions, respectively; r is the signed square root likelihood ratio statistic and $u_B(\psi) = \tilde\ell_\psi(|-\tilde\ell_{\lambda\lambda}|/|-\hat\ell_{\theta\theta}|)^{1/2}(\hat\pi/\tilde\pi)$.

From the frequentist perspective (Barndorff-Nielsen, 1986), a standard normal approximation to the relevant conditional distribution of

$$r^*(\psi) = r(\psi) + r(\psi)^{-1}\log\{u_F(\psi)/r(\psi)\} \tag{6}$$

has error of order $O(n^{-3/2})$, where $u_F(\psi) = (\hat\psi - \psi)(|-\hat\ell_{\theta\theta}|/|-\tilde\ell_{\lambda\lambda}|)^{1/2}$ in the exponential family context. It is easily seen that

$$\Phi\{r^*(\psi)\} = \Phi\{r(\psi)\} + \varphi\{r(\psi)\}\{r(\psi)^{-1} - u_F(\psi)^{-1}\} + O\left(n^{-3/2}\right). \tag{7}$$

Suppose the prior is such that $u_B = u_F + O_p(n^{-3/2})$, and fix ψ. By combining (5) and (7), under this condition, for given Y, the event $\psi \leqslant \psi^{(1-\alpha)}(\pi, Y)$ is equivalent to $\Phi\{r^*(\psi)\} + O(n^{-3/2}) \geqslant \alpha$. Therefore, by the delta method, we have from a repeated sampling perspective,

$$\begin{aligned}\text{pr}_\theta\left\{\psi \leqslant \psi^{(1-\alpha)}(\pi, Y) \mid C = c\right\} &= \text{pr}_\theta\left[\Phi\{r^*(\psi)\} + O_p\left(n^{-3/2}\right) \geqslant \alpha \mid C = c\right] \\ &= \text{pr}_\theta\left\{r^*(\psi) + O_p\left(n^{-3/2}\right) \geqslant z_\alpha \mid C = c\right\} = 1 - \alpha + O\left(n^{-3/2}\right),\end{aligned}$$

where $\Phi(z_\alpha) = \alpha$. Thus, the Bayesian confidence limits have conditional frequentist coverage error of order $O(n^{-3/2})$, the requirement of second-order conditional probability matching.

By using the standard result from linear algebra that $-\hat\ell^{\psi\psi} = |-\hat\ell_{\lambda\lambda}|/|-\hat\ell_{\theta\theta}|$, the condition on the prior that $u_B = u_F + O_p(n^{-3/2})$ may be written as

$$\frac{\tilde\pi}{\hat\pi}\frac{|-\tilde\ell_{\lambda\lambda}|^{-1}}{|-\hat\ell_{\lambda\lambda}|^{-1}}\frac{(-\tilde\ell^{\psi\psi})^{1/2}}{(-\hat\ell^{\psi\psi})^{1/2}} = \frac{z_S}{z_W} + O_p\left(n^{-3/2}\right),$$

where $z_S = \tilde\ell_\psi(-\tilde\ell^{\psi\psi})^{1/2}$ and $z_W = (\hat\psi - \psi)(-\hat\ell^{\psi\psi})^{-1/2}$ are both asymptotically standard normal pivots. We derive a simple condition under which $z_S/z_W = 1 + O_p(n^{-3/2})$. Under this condition, any prior π satisfying

$$\frac{\tilde\pi}{\hat\pi} = \frac{|-\tilde\ell_{\lambda\lambda}|}{|-\hat\ell_{\lambda\lambda}|}\frac{(-\tilde\ell^{\psi\psi})^{-1/2}}{(-\hat\ell^{\psi\psi})^{-1/2}} + O_p\left(n^{-3/2}\right), \tag{8}$$

is second-order conditional probability matching: (3) is satisfied with $m = 3$.

To derive a Taylor expansion of z_S around $\psi = \hat\psi$ having error of order $O_p(n^{-1})$, write $z_S = sv^{1/2}$, where $s = s(\psi) = \tilde\ell_\psi$ and $v = v(\psi) = -\tilde\ell^{\psi\psi}$. Since $\hat s = \hat\ell_\psi = 0$, we have $\hat z_S = 0$. Now, $(z_S)_1 = s_1 v^{1/2} + (sv^{-1/2}v_1)/2$; differentiating $s = \ell_1(\tilde\theta)$ yields $s_1 = \tilde\ell_{1r}\tilde\theta_1^r$, where $\tilde\theta_1^r = \partial\tilde\theta^r/\partial\theta^1$, and differentiating $\ell_\lambda(\tilde\theta) = 0$ yields $\tilde\theta_1^r = \tilde\ell^{r1}/\tilde\ell^{11}$. Thus, $s_1 = -1/\tilde\ell^{11} = -v^{-1}$, so $(z_S)_1 = v^{-1/2} + (sv^{-1/2}v_1)/2$. It follows that $(\hat z_S)_1 = -(-\hat\ell^{\psi\psi})^{-1/2}$ and $(z_S)_{11} = \{s(v^{-1/2}v_1)_1\}/2$, whence $(\hat z_S)_{11} = 0$. The desired expansion is $z_S = (\hat\psi - \psi)(-\hat\ell^{\psi\psi})^{-1/2} + O_p(n^{-1})$, which shows that $z_S/z_W = 1 + O_p(n^{-1})$. This result holds generally, not just for exponential families. Although z_S is invariant under monotonically increasing transformations of ψ, it is different from the usual standardized score statistic $s\hat v^{1/2}$; note that v can be negative in finite samples.

Since $(z_S)_{111} = -\{v^{-1}(v^{-1/2}v_1)_1\}/2 + \{s(v^{-1/2}v_1)_{11}\}/2$, we have $(\hat{z}_S)_{111} = -\{\hat{v}^{-1}(v^{-1/2}\hat{v}_1)_1\}/2$, where $\hat{v}^{-1} = (-\hat{\ell}^{\psi\psi})^{-1}$ cannot be 0. Thus, a sufficient condition for $(\hat{z}_S)_{111} = 0$, and hence, for $z_S/z_W = 1 + O_p(n^{-3/2})$, is that $v^{-1/2}v_1$ be a constant in ψ. Applying the identity $\partial\ell^{rs}/\partial\theta^t = -\ell^{ru}\ell^{sv}\ell_{uvt}$ to $v = -\tilde{\ell}^{\psi\psi} = -\tilde{\ell}^{11}$ yields $v_1 = \tilde{\ell}_{rst}\tilde{\ell}^{r1}\tilde{\ell}^{s1}\tilde{\theta}^t_1 = \tilde{\ell}_{rst}\tilde{\ell}^{r1}\tilde{\ell}^{s1}\tilde{\ell}^{t1}(\tilde{\ell}^{11})^{-1}$. It follows that $v^{-1/2}v_1 = -\tilde{\ell}_{rst}\tilde{\ell}^{r1}\tilde{\ell}^{s1}\tilde{\ell}^{t1}(-\tilde{\ell}^{11})^{-3/2}$, so $v^{-1/2}v_1$ is constant in ψ if $\ell_{rst}\ell^{r1}\ell^{s1}\ell^{t1}(-\ell^{11})^{-3/2}$ is constant in θ. For the exponential density (4), $\ell_{rst}\ell^{r1}\ell^{s1}\ell^{t1}(-\ell^{11})^{-3/2} = k_{rst}k^{r1}k^{s1}k^{t1}(k^{11})^{-3/2}$.

Hence, in summary, constancy of $g(\theta) = k_{rst}k^{r1}k^{s1}k^{t1}(k^{11})^{-3/2}$ is a sufficient condition for second-order conditional probability matching by any prior satisfying (8). This condition is on the model, not on the prior, and it is the same condition as that derived by Datta & Mukerjee (2004, § 2.5.4), in an analysis for the exponential family context that makes no regard to the conditioning appropriate in this context. Clearly, the prior $\pi \propto |-\ell_{\lambda\lambda}|(-\ell^{\psi\psi})^{-1/2}$ satisfies (8).

In the exponential family model (4), we have $\tilde{k}_a = \hat{k}_a = \bar{t}_a$, so any prior of the form

$$\pi \propto |-\ell_{\lambda\lambda}|(-\ell^{\psi\psi})^{-1/2}h(k_2, \ldots, k_d), \tag{9}$$

for arbitrary function h, also satisfies (8). Tibshirani (1989) considered a class of first-order probability matching priors, for the situation where the interest parameter and the nuisance parameter are orthogonal. In the current exponential family context this class corresponds exactly, under reparameterization, to the class of priors (9) for the canonical parameterization assumed in (4). If the model condition is satisfied, any member of this class of unconditional first-order matching priors automatically yields second-order conditional probability matching.

3. Numerical examples

3·1. *Inverse Gaussian distribution*

Consider a random sample $Y_1, \ldots, Y_n$ from the inverse Gaussian density

$$f(y; \psi, \lambda) = \{\psi/(2\pi)\}^{1/2}y^{-3/2}\exp\left\{-\tfrac{1}{2}(\psi y^{-1} + \lambda y) + (\psi\lambda)^{1/2}\right\}, \quad y > 0,\ \psi, \lambda > 0,$$

where the interest parameter ψ is the shape parameter of the distribution and λ is nuisance.

The appropriate conditional frequentist inference is based on the conditional distribution of $S = n^{-1}\sum_{i=1}^n Y_i^{-1}$ given $C = \bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, or, equivalently, on the marginal distribution of $V = \sum_{i=1}^n (Y_i^{-1} - \bar{Y}^{-1})$, since V is independent of C. We have $\psi V \sim \chi^2_{n-1}$.

We consider objective Bayes inference on ψ, based on four different priors specified as follows. Prior I: $\pi(\psi, \lambda) \propto \psi^{-3/2}\lambda^{-1/2}$; Prior II: $\pi(\psi, \lambda) \propto \psi^{-3/4}\lambda^{-3/4}$; Prior III: $\pi(\psi, \lambda) \propto \psi^{-1}\lambda^{-3/4}$; and Prior IV: $\pi(\psi, \lambda) \propto \psi^{-5/4}\lambda^{-3/4}$. The model condition that $g(\theta)$ be constant is satisfied in this example, with $g(\theta) \equiv -2^{3/2}$. Further, the condition (8) is satisfied by any prior of the form $\pi(\psi, \lambda) \propto \psi^{-1/2-a}\lambda^{a-3/2}$, so that both Priors I and IV should yield confidence limits of conditional frequentist coverage error of order $O(n^{-3/2})$; such nonuniqueness of a second-order conditional probability matching prior is, of course, typical. In this inference problem, Prior II is the Jeffreys prior, while Prior III is the reference prior (Liseo, 1993); neither of these priors yields second-order conditional probability matching.

Example 1. We consider the conditional frequentist confidence levels of posterior 5% and 95% quantiles, for a data configuration with observed values $s = 2{\cdot}0$, $c = 3{\cdot}0$ of the sufficient statistics, and for varying sample size n. Table 1 shows the 5% and 95% confidence limits obtained under the four priors in parentheses, with the corresponding exact conditional frequentist coverage levels.

To understand the table entries as conditional repeated sampling coverages, note that $\psi^{(1-\alpha)}(\pi, Y) \equiv \psi^{(1-\alpha)}(\pi, S, C)$ is a monotonically decreasing function of S for a fixed value of C. For fixed ψ_0, we have that $\{\psi_0 \leqslant \psi^{(1-\alpha)}(\pi, Y)\} \equiv \{S \leqslant s_0\}$, where s_0 has $\psi^{(1-\alpha)}(\pi, s_0, c) = \psi_0$. Hence, the conditional frequentist coverage under parameter value ψ_0 is $\mathrm{pr}\{\psi_0 \leqslant \psi^{(1-\alpha)}(\pi, Y) \mid C = c; \psi_0\} = \mathrm{pr}\{S \leqslant s_0 \mid C = c; \psi_0\}$, and this latter probability is $\mathrm{pr}\{V \leqslant n(s_0 - 1/c) \mid C = c; \psi_0\} = \mathrm{pr}\{V \leqslant n(s_0 - 1/c); \psi_0\} = \mathrm{pr}\{\chi^2_{n-1} \leqslant \psi_0 n(s_0 - 1/c)\}$, since V is independent of C. So, for example, conditional on $C = 3{\cdot}0$, if $n = 10$ and $\psi_0 = 0{\cdot}210$, the posterior 5% quantile using Prior I has conditional frequentist coverage

Table 1. *Conditional frequentist coverage levels of 5% and 95% confidence limits obtained from different Bayes priors, inverse Gaussian shape example. Actual limits are shown in parentheses beneath the coverage figures. Priors are detailed in the text*

	Prior I		Prior II		Prior III		Prior IV	
n	5%	95%	5%	95%	5%	95%	5%	95%
10	5·89	95·46	10·34	97·10	8·32	96·39	6·52	95·51
	(0·21)	(1·03)	(0·25)	(1·12)	(0·24)	(1·08)	(0·22)	(1·04)
20	5·39	95·27	7·62	96·35	6·44	95·75	5·39	95·03
	(0·31)	(0·91)	(0·33)	(0·94)	(0·32)	(0·92)	(0·31)	(0·91)
50	5·21	95·00	6·05	95·85	5·56	95·41	5·09	94·92
	(0·41)	(0·80)	(0·42)	(0·81)	(0·41)	(0·80)	(0·41)	(0·79)

equal to $\text{pr}(\chi^2_9 \leqslant 3{\cdot}5) = 5{\cdot}89\%$. As expected, Priors I and IV yield greater conditional frequentist accuracy, while it is clear also that different second-order conditional probability matching priors can yield noticeably different conditional frequentist properties.

3·2. *Gamma distribution*

Now suppose that $Y_1, \ldots, Y_n$ constitute a random sample from the gamma density

$$f(y; \psi, \lambda) = \lambda^{\psi} y^{\psi-1} \exp(-\lambda y)\Gamma(\psi)^{-1}, \quad y > 0,\ \psi, \lambda > 0. \tag{10}$$

We take the parameter of interest as the shape parameter ψ, with the scale parameter λ as nuisance. Conditional frequentist inference is based on the sampling distribution of $Q = \prod_{i=1}^n Y_i$, given the observed value c of $C = \sum_{i=1}^n Y_i$. This is equivalent (Pace & Salvan, 1997, Ex. 5.14) to basing inference on the marginal distribution of the statistic that is the ratio of the geometric and arithmetic means of Y_i. This distribution is complicated (Keating et al., 1990), but easily simulated: the distribution depends only on the interest parameter ψ, so the distribution can be simulated under an arbitrary setting of the nuisance parameter λ.

Let $\rho(\psi) = \{\psi\xi(\psi) - 1\}^{1/2}$, where $\xi(\psi) = (\partial^2/\partial\psi^2)\log\Gamma(\psi)$. We consider Bayesian inference under three priors specified as follows. Prior I: $\pi(\psi, \lambda) \propto \rho(\psi)\psi^{-1/2}\lambda^{-1}$; Prior II: $\pi(\psi, \lambda) \propto \rho(\psi)\lambda^{-1}$; and Prior III: $\pi(\psi, \lambda) \propto \{\rho(\psi)\}^2\psi^{-1/2}\lambda^{-1}$. Of these, Prior II is first-order probability matching, but the model condition that $g(\theta)$ be constant is not satisfied: indeed, $g(\theta) = \psi^{-1/2}\{\rho(\psi)\}^{-3/2}\{\psi^2\xi'(\psi) + 1\}$, which depends on ψ. Therefore, second-order conditional probability matching is not achievable by a first-order matching prior in this inference problem. Prior I is the Jeffreys prior, and Prior III is the reference prior; see Liseo (1993).

Example 2. As in our first example, we examine conditional frequentist confidence levels of posterior 5% and 95% quantiles, now for observed values $q = 1{\cdot}0$, $c = 20{\cdot}0$ of the sufficient statistics, again for varying n. Table 2 shows the 5% and 95% confidence limits obtained under the three priors in parentheses, with the corresponding exact conditional frequentist coverage levels. The first-order probability matching Prior II is seen to yield good conditional frequentist coverage properties.

4. Relationship with parametric bootstrap

In the current exponential family context, a simple unconditional parametric bootstrap, as considered by DiCiccio et al. (2001) and Lee & Young (2005), approximates the exact conditional frequentist inference to the same order of error, $O(n^{-3/2})$, as second-order conditional probability matching objective Bayes; see DiCiccio & Young (2008). This parametric bootstrap is based on simulating the unconditional distribution of the statistic $r(\psi)$, under the model with parameter value set as $(\psi, \hat{\lambda}_\psi)$, that is, with the nuisance parameter set as its constrained maximum likelihood value for the given data. A further, direct and analytic frequentist approach to approximation to the exact conditional inference to order $O(n^{-3/2})$ is

Table 2. *Conditional frequentist coverage levels of 5% and 95% confidence limits obtained from objective Bayes, bootstrap and analytic procedures; gamma shape example. Actual limits are shown in parentheses beneath the coverage figures. Priors are detailed in the text*

	Prior I		Prior II		Prior III		Bootstrap		r^*	
n	5%	95%	5%	95%	5%	95%	5%	95%	5%	95%
5	9·42	97·09	5·18	95·03	2·21	90·90	5·07	95·01	5·67	95·35
	(0·15)	(0·91)	(0·12)	(0·82)	(0·09)	(0·71)	(0·12)	(0·82)	(0·13)	(0·83)
10	7·84	96·64	5·11	95·01	2·88	92·25	5·00	95·00	5·19	95·11
	(0·40)	(1·46)	(0·36)	(1·37)	(0·31)	(1·27)	(0·36)	(1·37)	(0·36)	(1·37)
15	7·32	96·45	5·05	95·01	3·10	92·78	4·98	95·00	5·06	95·06
	(0·98)	(3·05)	(0·91)	(2·91)	(0·83)	(2·75)	(0·91)	(2·91)	(0·91)	(2·91)

(Barndorff-Nielsen, 1986) based on standard normal approximation to the distribution of the statistic $r^*(\psi)$, given by (6). Book length treatments of the analytic approach were given by Barnforff-Nielsen & Cox (1994), Severini (2000) and Brazzale et al. (2007). A comparative evaluation of the three routes given by Young (2009) suggests strongly that the bootstrap approach should be judged as providing the most accurate approximation to exact conditional frequentist inference in exponential families.

Though conceptually simple, a computational drawback to the objective Bayes procedure is the need for integration of the joint posterior of (ψ, λ) over the nuisance parameter λ, to obtain the marginal posterior for the interest parameter ψ. The bootstrap may be used to approximate a second-order conditional probability matching objective Bayes inference, obtaining approximations to objective Bayesian posterior quantiles by simulation, thus avoiding any awkward analytic calculation. Further, given the remarkable conditional accuracy properties of the parametric bootstrap noted by DiCiccio & Young (2008), we may expect, at least for moderate n, to be able to use the bootstrap to evaluate the conditional frequentist properties of objective Bayes procedures in cases where the exact conditional frequentist inference is intractable or impossible.

Example 2 (continued). We now approximate the exact conditional frequentist confidence limits by the parametric bootstrap procedure and by the analytic procedure based on the r^* statistic, for the same data configuration, $q = 1{\cdot}0$, $c = 20{\cdot}0$, as considered before. The bootstrap limits are based on a simulation that approximates the sampling distribution of the signed root statistic r by the drawing of 5 million samples from the relevant fitted distribution. Table 2 reports the conditional frequentist confidence levels of the bootstrap and analytic limits. The unconditional bootstrap yields more accurate approximation to exact conditional frequentist inference than does the objective Bayes procedure, though, of course, the latter does not give second-order conditional probability matching in this example. The bootstrap produces better approximation than r^*, which, surprisingly, is less accurate than the objective Bayes method.

5. Further remarks and comments

Conditioning also eliminates nuisance parameters in other settings, notably when the interest parameter ψ is a ratio of canonical parameters in an exponential family. A theoretical analysis of objective Bayes procedures for this case, along the lines of that given in § 2, does not yield any simple, transparent conditions under which second-order conditional frequentist matching is achievable by a first-order probability matching prior. However, identification of second-order conditional probability matching priors, in particular key examples, is straightforward.

Example 3. Let X and Y be independent exponential random variables, with rate parameters λ and $\psi\lambda$, respectively, and assume that the interest parameter is the ratio of the two rates, ψ. Under a joint prior of the form $\pi(\psi, \lambda) \propto 1/(\psi\lambda)$, the marginal posterior density for ψ, given $X = x$, $Y = y$, is of the form $\pi(\psi \mid x, y) = xy/(x + \psi y)^2$. Then the posterior $1 - \alpha$ quantile is $\psi^{(1-\alpha)}(\pi, X, Y) = (1/\alpha - 1)X/Y$. The appropriate conditional frequentist inference involves (Reid, 1995, Ex. 5.1) conditioning on the linear pivotal $s(\psi) = X + \psi Y$, and, conditional on $s(\psi) = s$, X is uniformly distributed on $[0, s]$. Hence, the

Table 3. *Conditional frequentist coverage levels, estimated by parametric bootstrap, of 5% and 95% confidence limits obtained from two Bayes priors, gamma mean example. Actual limits are shown in parentheses beneath the coverage figures. Priors are detailed in the text*

	Prior IV		Prior I	
n	5%	95%	5%	95%
5	5·13	85·58	6·23	83·31
	(1·52)	(14·52)	(1·62)	(12·77)
10	5·15	94·19	5·84	93·46
	(1·17)	(4·21)	(1·20)	(4·07)
15	5·05	95·05	5·59	94·39
	(0·98)	(1·94)	(0·99)	(1·91)

assumed prior is exactly conditional frequentist probability matching, with $\theta = (\psi, \lambda)$, for

$$\text{pr}_\theta\{\psi \leqslant \psi^{(1-\alpha)}(\pi, X, Y) \mid s(\psi) = s\} = \text{pr}\{X \geqslant \alpha s(\psi) \mid s(\psi) = s\} = 1 - \alpha.$$

Example 4. Consider inference for the mean $\mu \equiv \psi/\lambda$ of the gamma density (10) with the shape parameter ψ as nuisance. Here, Prior IV: $\pi(\mu, \psi) \propto \{\rho(\psi)\}^2\mu^{-1}$ is the unique second-order unconditional probability matching prior, and it is easily shown to be second-order conditional probability matching (Casella et al., 1995). Table 3 shows, for the data configuration and sample sizes used in Table 2, conditional frequentist confidence levels of posterior 5% and 95% limits, from Prior I, considered in § 3·2, and Prior IV. Jensen (1986) argued for this problem that the exact conditional frequentist inference is intractable unless $n \leqslant 3$; for larger values of n, approximation is necessary. DiCiccio & Young (2008) provided evidence that the parametric bootstrap yields essentially exact approximation in this problem, and the entries of Table 3 were obtained this way by using 5 million bootstrap samples. The higher-order conditional matching property of Prior IV is evident.

In this paper we have been concerned with identification of priors on the full parameter, which, when combined with the full data likelihood, yield a marginal posterior for the interest parameter having quantiles with the correct conditional frequentist interpretation. Chang & Mukerjee (2006) considered an alternative approach to probability matching, in which a prior is specified only for the interest parameter, the objective being that when this is combined with an adjusted profile likelihood for that interest parameter, posterior quantiles have the correct unconditional frequentist coverage. The family of likelihoods allowed in their analysis includes the conditional likelihood appropriate in the multi-parameter exponential family setting, and conditional probability matching within that framework is worthy of investigation. Further, since probability matching is typically achieved by improper priors and is with respect to a conditional frequentist approach, it is of some interest to examine whether the approach yields, or avoids, marginalization paradoxes (Dawid et al., 1973) typical of objective Bayes methodology.

Acknowledgement

The authors thank the editor, associate editor and two referees for particularly helpful and insightful comments that led to an improved version of the manuscript.

References

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–22.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

Brazzale, A. R., Davison, A. C. & Reid, N (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge: Cambridge University Press.

Casella, G., DiCiccio, T. J. & Wells, M. T. (1995). [Inference based on estimation functions in the presence of nuisance parameters]: Comment: Alternative aspects of conditional inference. *Statist. Sci.* **10**, 179–85.

CHANG, I. H. & MUKERJEE, R. (2006). Probability matching property of adjusted likelihoods. *Statist. Prob. Lett.* **76**, 838–42.

DATTA, G. S. & MUKERJEE, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. New York: Springer.

DAWID, A. P., STONE, M. & ZIDEK, J. V. (1973). Marginalisation paradoxes in Bayesian and structural inference (with Discussion). *J. R. Statist. Soc.* B **35**, 189–233.

DICICCIO, T. J., MARTIN, M. A. & STERN, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Statist*. **29**, 67–76.

DICICCIO, T. J. & YOUNG, G. A. (2008). Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families. *Biometrika* **95**, 747–58.

JENSEN, J. L. (1986). Inference for the mean of a gamma distribution with unknown shape parameter. *Scand. J. Statist.* **13**, 135–51.

KEATING, J. P., GLASER, R. E. & KETCHUM, N. S. (1990). Testing hypotheses about the shape parameter of a gamma distribution. *Technometrics* **32**, 67–82.

LEE, S. M. S. & YOUNG, G. A. (2005). Parametric bootstrapping with nuisance parameters. *Statist. Prob. Lett.* **71**, 143–53.

LISEO, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295–304.

MUKERJEE, R. & DEY, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* **80**, 499–505.

MUKERJEE, R. & GHOSH, M. (1997). Second-order probability matching priors. *Biometrika* **84**, 970–5.

PACE, L. & SALVAN, A. (1997). *Principles of Statistical Inference: From a Neo-Fisherian Perspective*. Singapore: World Scientific.

PEERS, H. W. (1965). On confidence sets and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc.* B **27**, 9–16.

REID, N. (1995). The roles of conditioning in inference. *Statist. Sci.* **10**, 138–57.

SEVERINI, T. A. (2000). *Likelihood Methods in Statistics.* New York: Oxford University Press.

TIBSHIRANI, T. J. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–8.

WELCH, B. L. & PEERS, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Statist. Soc.* B **62**, 159–80.

YOUNG, G. A. (2009). Routes to higher-order accuracy in parametric inference. *Aust. New Zeal. J. Statist.* **51**, 115–26.

YOUNG, G. A. & SMITH, R. L. (2005). *Essentials of Statistical Inference*. Cambridge: Cambridge University Press.

[*Received October* 2008. *Revised September* 2009]