Objective Bayes, conditional inference and the signed root likelihood ratio statistic

BY THOMAS J. DICICCIO

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

TODD A. KUFFNER

Department of Mathematics, American University of Beirut, Beirut 1107 2020, Lebanon
tk30@aub.edu.lb

AND G. ALASTAIR YOUNG

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

SUMMARY

Bayesian properties of the signed root likelihood ratio statistic are analysed. Conditions for first-order probability matching are derived by the examination of the Bayesian posterior and frequentist means of this statistic. Second-order matching conditions are shown to arise from matching of the Bayesian posterior and frequentist variances of a mean-adjusted version of the signed root statistic. Conditions for conditional probability matching in ancillary statistic models are derived and discussed.

Some key words: Bayesian inference; Conditional inference; Nuisance parameter; Objective Bayes; Probability-matching prior; Signed root likelihood ratio statistic.

1. INTRODUCTION

In problems concerning inference on a scalar interest parameter in the presence of a nuisance parameter, the signed root likelihood ratio statistic is a fundamental object of statistical methodology. The focus of this paper is an analysis of the signed root statistic from an objective Bayes perspective, where the issue of identification of prior distributions that display the property of probability matching is central. Under probability matching, quantiles of the Bayesian posterior distribution have the property of being confidence limits in the conventional, frequentist sense.

Considering inference based on a sample of size n, we establish a number of results. If we choose the prior distribution in a Bayesian analysis so that the frequentist and Bayesian posterior means of the signed root statistic match to $O_p(n^{-1})$, then the prior achieves first-order probability matching: the Bayesian $1 - \alpha$ posterior quantile has frequentist coverage $1 - \alpha + O(n^{-1})$. It is observed that such matching of frequentist and Bayesian posterior means occurs when the prior satisfies the conditions noted by Peers (1965), who extended to the nuisance parameter case work of Welch & Peers (1963). We further obtain a simple condition, related to the Bayesian posterior and frequentist variances of a mean-adjusted version of the signed root statistic, under which the prior is second-order probability matching, so that the frequentist coverage of the Bayesian

$1 - \alpha$ quantile is $1 - \alpha + O(n^{-3/2})$. This condition is shown to be equivalent to that established by Mukerjee & Dey (1993) and Mukerjee & Ghosh (1997) by an analytically more elaborate shrinkage argument. Our analysis therefore provides a transparent and intuitive interpretation, in terms of the distributional properties of the signed root statistic, for existing conditions for probability-matching priors.

In particular statistical problems, specifically those involving inference on canonical parameters in multi-parameter exponential family models and in models admitting ancillary statistics, the appropriate frequentist inference is a conditional one, so that the relevant objective Bayesian notion is that of a conditional probability-matching prior. We further provide an analysis of the conditional frequentist behaviour of Bayesian posterior quantiles in ancillary statistic models, as was carried out for the exponential family context by DiCiccio & Young (2010). Their approach involves matching of higher-order conditional frequentist and Bayesian asymptotics, yielding simple conditions for probability matching. In the ancillary statistic context, this approach does not, however, yield any general, easily interpretable conditions, unlike the methodologies applied here. By considering the distributional properties of the signed root statistic, we note first that if the unconditional first-order probability-matching condition of Peers (1965), which we refer to as the Welch–Peers condition, is satisfied, then the prior automatically enjoys the property of first-order conditional probability matching. We then establish our key result, which identifies a simple condition involving the Bayesian posterior and conditional frequentist means of the signed root statistic, under which the conditions for second-order probability matching in an unconditional sense ensure also second-order conditional probability matching.

2. Notation

Consider a random vector $Y = (Y_1, \ldots, Y_n)$ having continuous probability distribution that depends on an unknown $(q + 1)$-dimensional parameter $\theta = (\theta^1, \ldots, \theta^{q+1})$, and denote the loglikelihood function for θ based on Y by $L(\theta)$. Suppose that θ is partitioned in the form $\theta = (\psi, \phi)$, where ψ is a scalar interest parameter and ϕ is a q-dimensional nuisance parameter. Let $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ be the overall maximum likelihood estimator of θ, and let $\tilde{\theta}(\psi) = \{\psi, \tilde{\phi}(\psi)\}$ be the constrained maximum likelihood estimator of θ for given ψ. The log profile likelihood function for ψ is $M(\psi) = L\{\tilde{\theta}(\psi)\}$.

In the asymptotic calculations that follow, standard conventions for denoting arrays and summation are used. For these conventions, it is understood that indices $i, j, k, \ldots$ range over $2, \ldots, q + 1$, and that indices $r, s, t, \ldots$ range over $1, \ldots, q + 1$. Summation over the relevant range is implied for any index appearing in an expression both as a subscript and as a superscript. Differentiation of the functions $L(\theta)$ and $M(\psi)$ is indicated by subscripts, so $L_r(\theta) = \partial L(\theta)/\partial \theta^r$, $L_{rs}(\theta) = \partial^2 L(\theta)/\partial \theta^r \partial \theta^s$, $M_1(\psi) = \partial M(\psi)/\partial \psi$, $M_{11}(\psi) = \partial^2 M(\psi)/\partial \psi^2$, etc. In this notation, $L_r(\hat{\theta}) = 0$ $(r = 1, \ldots, q + 1)$ and $M_1(\hat{\psi}) = 0$. Evaluation of the derivatives of $L(\theta)$ at $\hat{\theta}$ and the derivatives of $M(\psi)$ at $\hat{\psi}$ is indicated by placing a circumflex above the appropriate quantity; for example, $\hat{L}_r = L_r(\hat{\theta}) = 0$, $\hat{L}_{rs} = L_{rs}(\hat{\theta})$, $\hat{M}_1 = M_1(\hat{\psi}) = 0$, $\hat{M}_{11} = M_{11}(\hat{\psi})$, etc. Let $\lambda_r = E\{L_r(\theta)\} = 0$, $\lambda_{rs} = E\{L_{rs}(\theta)\}$, $\lambda_{rst} = E\{L_{rst}(\theta)\}$, etc., and define $l_r = L_r(\theta) - \lambda_r = L_r(\theta)$, $l_{rs} = L_{rs}(\theta) - \lambda_{rs}$, $l_{rst} = L_{rst}(\theta) - \lambda_{rst}$, etc. The constants λ_{rs}, λ_{rst}, etc. are assumed to be of order $O(n)$; the variables l_r, l_{rs}, l_{rst}, etc. have expectation zero and they are assumed to be of order $O_p(n^{1/2})$. The joint cumulants of l_r, l_{rs}, etc. are assumed to be of order $O(n)$. These assumptions are usually satisfied in regular situations involving independent observations.

In subsequent calculations, it is useful to extend the λ-notation: let

$$\lambda_{r,s} = E\{L_r(\theta)L_s(\theta)\}, \quad \lambda_{rs,t} = E\{L_{rs}(\theta)L_t(\theta)\}, \quad \lambda_{r,s,t} = E\{L_r(\theta)L_s(\theta)L_t(\theta)\},$$

etc. Identities involving the λs can be derived by repeated differentiation of the identity $\int \exp\{L(\theta)\}\mathrm{d}y = 1$; in particular,

$$\lambda_{rs} + \lambda_{r,s} = 0, \quad \lambda_{rst} + \lambda_{rs,t} + \lambda_{rt,s} + \lambda_{st,r} + \lambda_{r,s,t} = 0.$$

Differentiation of the definition $\lambda_{rs} = \int L_{rs}(\theta)\exp\{L(\theta)\}\,\mathrm{d}y$ yields $\lambda_{rs/t} = \lambda_{rst} + \lambda_{rs,t}$, where $\lambda_{rs/t} = \partial\lambda_{rs}/\partial\theta^t$.

Let (λ^{rs}), (L^{rs}) and $(\hat{L}^{rs})$ be the $(q+1) \times (q+1)$ matrix inverses of (λ_{rs}), (L_{rs}) and $(\hat{L}_{rs})$, respectively. Define $\tau^{rs} = \lambda^{r1}\lambda^{s1}/\lambda^{11}$, $\hat{T}^{rs} = \hat{L}^{r1}\hat{L}^{s1}/\hat{L}^{11}$, $\nu^{rs} = \lambda^{rs} - \tau^{rs}$ and $\hat{V}^{rs} = \hat{L}^{rs} - \hat{T}^{rs}$. Note that λ^{rs}, τ^{rs} and ν^{rs} are all of order $O(n^{-1})$, and $\hat{L}^{rs}$, $\hat{T}^{rs}$ and $\hat{V}^{rs}$ are all of order $O_p(n^{-1})$. Furthermore, $1/\lambda^{11} = \lambda_{r1}\lambda_{s1}\tau^{rs}$ is of order $O(n)$ and $1/\hat{L}^{11} = \hat{L}_{r1}\hat{L}_{s1}\hat{T}^{rs}$ is of order $O_p(n)$. Note that $\tau^{r1} = \lambda^{r1}$ and $\nu^{r1} = 0$; thus, the entries of $(q+1) \times (q+1)$ matrices (ν^{rs}) and $(\hat{V}^{rs})$ are all 0 except for the lower right-hand submatrices (ν^{ij}) and $(\hat{V}^{ij})$, which are the inverses of (λ_{ij}) and $(\hat{L}_{ij})$, respectively.

3. Signed root statistic and probability matching

The likelihood ratio statistic $W(\psi) = 2\{M(\hat{\psi}) - M(\psi)\}$ is useful for testing the null hypothesis $H_0 : \psi = \psi_0$ against the two-sided alternative $H_a : \psi \neq \psi_0$ or for constructing two-sided confidence intervals. However, for a scalar interest parameter, it is natural to conduct one-sided tests, where the alternative is either $H_a : \psi > \psi_0$ or $H_a : \psi < \psi_0$, or to construct one-sided, upper or lower, confidence limits. One-sided inferences can be achieved by considering the signed square root of the likelihood ratio statistic $R(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\{W(\psi)\}^{1/2}$, which has the standard normal distribution, $N(0, 1)$, to error of order $O(n^{-1/2})$, so that the $N(0, 1)$ approximation produces one-sided confidence limits having coverage error of that order. The order of the error in the $N(0, 1)$ approximation to the distribution of $R(\psi)$ can be reduced (DiCiccio & Stern, 1994a) to $O(n^{-1})$ by correcting for the mean of $R(\psi)$.

The signed root $R(\psi)$ behaves (DiCiccio & Stern, 1994b) identically from a Bayesian perspective: the posterior distribution of $R(\psi)$ is standard normal to error of order $O(n^{-1/2})$, and the order of the error in the $N(0, 1)$ approximation to the posterior distribution of $R(\psi)$ can be reduced to $O(n^{-1})$ by correcting for the posterior mean. These order statements are asserted conditionally, given the data, so it is appropriate to use $O(\cdot)$ rather than $O_p(\cdot)$ to describe the errors associated with the $N(0, 1)$ approximation. To distinguish frequentist and Bayesian inference, let $\mu_F = \mu_F(\theta)$ denote the frequentist mean of $R(\psi)$ and let $\mu_B = \mu_B(Y)$ denote the posterior mean. Applying the $N(0, 1)$ approximation to the posterior distribution of $R(\psi) - \mu_B$ yields approximate posterior percentage points for ψ having a posterior probability error of order $O(n^{-1})$. If the prior distribution is chosen so that $\mu_B = \mu_F + O_p(n^{-1})$, then the posterior distribution of $R(\psi) - \mu_B$ coincides with the frequentist one to an error of order $O(n^{-1})$. Thus, for such a prior distribution, the upper $1 - \alpha$ posterior percentage point is necessarily an approximate upper $1 - \alpha$ frequentist confidence limit having a coverage error of order $O(n^{-1})$. As noted, Peers (1965) derived a condition that the prior distribution must satisfy in order for it to have this first-order probability-matching property, although not by the method considered here of matching the Bayesian posterior and frequentist means of the signed root statistic.

DiCiccio & Stern (1994a) showed that μ_F is

$$\mu_F = -\tfrac{1}{2}\eta\lambda_{rst}\lambda^{r1}\lambda^{st} - \tfrac{1}{6}\eta^3\lambda_{rst}\lambda^{r1}\lambda^{s1}\lambda^{t1} + \eta\lambda_{rs/t}\lambda^{r1}\lambda^{st} + \tfrac{1}{2}\eta^3\lambda_{rs/t}\lambda^{r1}\lambda^{s1}\lambda^{t1} + O(n^{-3/2}),$$

where $\eta = (-\lambda^{11})^{-1/2}$. A general expression for μ_B is derived in the Appendix.

To compare μ_B and μ_F, note that

$$\mu_B = -\tfrac{1}{2}\eta\lambda_{rst}\lambda^{r1}\lambda^{rs} - \tfrac{1}{6}\eta^3\lambda_{rst}\lambda^{r1}\lambda^{s1}\lambda^{t1} + \eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1} + O_p(n^{-1})$$

in the frequentist sense; hence, $O_p(\cdot)$ is used for the error term in place of $O(\cdot)$. Thus, the condition $\mu_B = \mu_F + O_p(n^{-1})$ is met when the prior satisfies

$$\eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1} = \eta\lambda_{rs/t}\lambda^{r1}\lambda^{st} + \tfrac{1}{2}\eta^3\lambda_{rs/t}\lambda^{r1}\lambda^{s1}\lambda^{t1}.$$

A standard result of matrix algebra gives $(\lambda^{uv})_{/t} = -\lambda_{rs/t}\lambda^{ru}\lambda^{sv}$, so it follows that

$$\sum_r \frac{\partial \lambda^{r1}}{\partial \theta^r} = -\lambda_{rs/t}\lambda^{r1}\lambda^{st}, \quad \frac{\partial \eta}{\partial \theta^r} = -\tfrac{1}{2}\eta^3\lambda_{st/r}\lambda^{s1}\lambda^{t1};$$

consequently, the condition on the prior can be written as

$$\eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1} = -\sum_r \frac{\partial(\eta\lambda^{r1})}{\partial \theta^r}, \tag{1}$$

which is what we have termed the Welch–Peers condition.

In many cases, Welch–Peers priors are not unique; for example, under parameter orthogonality, the prior is (Tibshirani, 1989) essentially the Jeffreys prior on the interest parameter, multiplied by an arbitrary function of the nuisance parameter. As a means to choose among the Welch–Peers priors, it is natural to attempt to determine those that are second-order probability matching, i.e., those for which the percentage points are one-sided confidence limits having a coverage error of order $O(n^{-3/2})$. This problem has been well studied, and Mukerjee & Dey (1993) and Mukerjee & Ghosh (1997) have given a differential equation, auxiliary to the Welch–Peers condition, that the prior must satisfy for it to be second-order probability matching; see Datta & Mukerjee (2004, Theorem 2.4.1). Here we demonstrate that the auxiliary condition can be developed by considering a mean-adjusted version of the signed root statistic.

DiCiccio & Stern (1993) showed that the posterior expectation of $\{R(\psi)\}^2$ is $1 + a_B + O(n^{-3/2})$, where

$$a_B = \tfrac{1}{4}(\hat{L}^{rs}\hat{L}^{tu} - \hat{V}^{rs}\hat{V}^{tu})\hat{L}_{rstu} - \tfrac{1}{4}(\hat{L}^{ru}\hat{L}^{st}\hat{L}^{vw} - \hat{V}^{ru}\hat{V}^{st}\hat{V}^{vw})\hat{L}_{rst}\hat{L}_{uvw}$$
$$- \tfrac{1}{6}(\hat{L}^{ru}\hat{L}^{sw}\hat{L}^{tv} - \hat{V}^{ru}\hat{V}^{sw}\hat{V}^{tv})\hat{L}_{rst}\hat{L}_{uvw} + (\hat{L}^{rs}\hat{L}^{tu} - \hat{V}^{rs}\hat{V}^{tu})\hat{L}_{rst}\hat{\Pi}_u$$
$$- (\hat{L}^{rs} - \hat{V}^{rs})\hat{\Pi}_{rs},$$

and $\hat{\Pi}_r = \Pi_r(\hat{\theta})$, $\hat{\Pi}_{rs} = \Pi_{rs}(\hat{\theta})$, with $\Pi_r(\theta) = \pi_r(\theta)/\pi(\theta)$, $\Pi_{rs}(\theta) = \pi_{rs}(\theta)/\pi(\theta)$, $\pi_r(\theta) = \partial\pi(\theta)/\partial\theta^r$, $\pi_{rs}(\theta) = \partial^2\pi(\theta)/\partial\theta^r\partial\theta^s$. It follows that the posterior variance of $R(\psi) - \mu_B$ is

$$\sigma_B^2 = 1 + a_B - \mu_B^2 + O(n^{-3/2}).$$

When the prior satisfies condition (1), from a frequentist perspective, since μ_B is of order $O_\mathrm{p}(n^{-1/2})$ and $\mu_\mathrm{B} = \mu_\mathrm{F} + O_\mathrm{p}(n^{-1})$, where μ_F is of order $O(n^{-1/2})$, it follows that $\mu_\mathrm{B}^2 = \mu_\mathrm{F}^2 + O_\mathrm{p}(n^{-3/2})$, where μ_F^2 is of order $O(n^{-1})$; hence, the posterior variance satisfies $\sigma_\mathrm{B}^2 = 1 + a_\mathrm{B} - \mu_\mathrm{F}^2 + O_\mathrm{p}(n^{-3/2})$.

DiCiccio & Stern (1994b) showed that the frequentist expectation of $\{R(\psi)\}^2$ is $1 + a_\mathrm{F} + O(n^{-3/2})$, where

$$
\begin{aligned}
a_\mathrm{F} = {} & (\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})(\tfrac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}) \\
& - (\lambda^{ru}\lambda^{st}\lambda^{vw} - \nu^{ru}\nu^{st}\nu^{vw})(\tfrac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}) \\
& - (\lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})(\tfrac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}) + O(n^{-3/2}).
\end{aligned}
$$

It is shown in the Appendix that $E(\mu_\mathrm{B}) = \mu_\mathrm{F} + O(n^{-3/2})$, and so it follows as a particular case of the results of DiCiccio & Stern (1994a) that the frequentist variance of $R(\psi) - \mu_\mathrm{B}$ is

$$
\sigma_\mathrm{F}^2 = 1 + a_\mathrm{F} + 2\eta\mu_{\mathrm{F}/r}\lambda^{r1} - \mu_\mathrm{F}^2 + O(n^{-3/2}),
$$

where $\mu_{\mathrm{F}/r} = \partial\mu_\mathrm{F}/\partial\theta^r$.

From both the Bayesian and frequentist perspectives, the third- and higher-order cumulants of $R(\psi) - \mu_\mathrm{B}$ are (DiCiccio & Stern, 1994a) of order $O(n^{-3/2})$ or smaller. Thus, the marginal distribution of $\{R(\psi) - \mu_\mathrm{B}\}/\sigma_\mathrm{F}$ and the posterior distribution of $\{R(\psi) - \mu_\mathrm{B}\}/\sigma_\mathrm{B}$ are both standard normal to an error of order $O(n^{-3/2})$; moreover, if $\pi(\theta)$ is a prior density such that $\sigma_\mathrm{B}^2 = \sigma_\mathrm{F}^2 + O_\mathrm{p}(n^{-3/2})$, then $\pi(\theta)$ is a second-order probability-matching prior: a formal proof follows the same argument as that used in our main result in §4. The condition under which $\pi(\theta)$ will be a second-order probability-matching prior is therefore $a_\mathrm{B} = a_\mathrm{F} + 2\eta\mu_{\mathrm{F}/r}\lambda^{r1} + O_\mathrm{p}(n^{-3/2})$. Since

$$
\begin{aligned}
a_\mathrm{B} = {} & \tfrac{1}{4}(\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})\lambda_{rstu} - \tfrac{1}{4}(\lambda^{ru}\lambda^{st}\lambda^{vw} - \nu^{ru}\nu^{st}\nu^{vw})\lambda_{rst}\lambda_{uvw} \\
& - \tfrac{1}{6}(\lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})\lambda_{rst}\lambda_{uvw} + (\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})\lambda_{rst}\Pi_u \\
& - (\lambda^{rs} - \nu^{rs})\Pi_{rs} + O_\mathrm{p}(n^{-3/2}),
\end{aligned}
$$

this condition can be expressed as

$$
\begin{aligned}
a_\mathrm{F} + 2\eta\mu_{\mathrm{F}/r}\lambda^{r1} = {} & \tfrac{1}{4}(\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})\lambda_{rstu} - \tfrac{1}{4}(\lambda^{ru}\lambda^{st}\lambda^{vw} - \nu^{ru}\nu^{st}\nu^{vw})\lambda_{rst}\lambda_{uvw} \\
& - \tfrac{1}{6}(\lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})\lambda_{rst}\lambda_{uvw} + (\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})\lambda_{rst}\Pi_u \\
& - (\lambda^{rs} - \nu^{rs})\Pi_{rs}.
\end{aligned}
$$

By assuming that the prior density $\pi(\theta)$ satisfies the first-order probability-matching condition (1), the condition for second-order probability matching reduces to

$$
\begin{aligned}
\tau^{rs}\Pi_{rs} - \tau^{rs}\lambda^{tu}\lambda_{rst}\Pi_u = {} & (\tau^{rs}\nu^{tu} + \tfrac{1}{3}\tau^{rs}\tau^{tu})\lambda_{rst/u} - \lambda^{ru}\tau^{st}\lambda^{vw}\lambda_{rst}\lambda_{uv/w} \\
& - (\lambda^{ru}\nu^{sw}\tau^{tv} + \nu^{ru}\nu^{sw}\tau^{tv})\lambda_{rst}\lambda_{uv/w} \\
& + (\tau^{ru}\lambda^{st}\nu^{vw} + \tfrac{1}{2}\tau^{ru}\tau^{st}\tau^{vw})\lambda_{rs/t}\lambda_{uv/w} \\
& + \tau^{ru}\lambda^{sw}\nu^{tv}\lambda_{rs/t}\lambda_{uv/w},
\end{aligned}
$$

which may be written more succinctly as

$$\sum_r \frac{\partial\{\pi(\nu^{rs} + \frac{1}{3}\tau^{rs})\tau^{tu}\lambda_{stu}\}}{\partial\theta^r} + \sum_{r,s} \frac{\partial^2(\pi\,\tau^{rs})}{\partial\theta^r\partial\theta^s} = 0. \qquad (2)$$

This condition is that given by Mukerjee & Dey (1993), who considered a scalar nuisance parameter, and by Mukerjee & Ghosh (1997), who considered a vector nuisance parameter.

4. Conditional inference

The main motivation for our analysis is to investigate the conditional frequentist properties of approximate confidence limits obtained from Welch–Peers priors. Suppose that A is an ancillary statistic such that $(\hat\theta, A)$ is sufficient. To undertake calculations with respect to the conditional distribution of Y given A, or equivalently, with respect to the conditional distribution of $(\hat\theta, A)$ given A, it is useful to consider the versions of the λs obtained by applying their definitions to the conditional distribution. The resulting conditional quantities are distinguished notationally from the unconditional ones by the inclusion of the accent symbol, $\circ$. Since the conditional loglikelihood function $\mathring{L}(\theta)$ differs from the unconditional loglikelihood function $L(\theta)$ by a constant, i.e., a quantity that depends on A but not on θ, it follows that $\mathring{L}_r = L_r$, $\mathring{L}_{rs} = L_{rs}$, etc. Thus, $\mathring\lambda_r = E_{Y|A}(L_r)$, $\mathring\lambda_{rs} = E_{Y|A}(L_{rs})$, etc. Note that $\mathring\lambda_r = 0$. The quantities $\mathring\lambda_{rs}$, $\mathring\lambda_{rst}$, etc. are random variables depending on A, and they are assumed to be of order $O_p(n)$. The variables $\mathring{l}_r = L_r$, $\mathring{l}_{rs} = L_{rs} - \mathring\lambda_{rs}$, etc. all have zero conditional expectation and hence they have zero unconditional expectation, and they are assumed to be of order $O_p(n^{1/2})$. The joint conditional cumulants of $\mathring{l}_r$, $\mathring{l}_{rs}$, etc. depend on A, and they are assumed to be of order $O(n)$ given A and of order $O_p(n)$ unconditionally. The identities that hold for the λs immediately carry over to the $\mathring\lambda$s.

In the calculations that follow, it is necessary to take into account the differences between the λs and the $\mathring\lambda$s. To describe the difference between λ_{rs} and $\mathring\lambda_{rs}$, first note that

$$E_A(\mathring\lambda_{rs}) = E_A\{E_{Y|A}(L_{rs})\} = E_Y(L_{rs}) = \lambda_{rs};$$

moreover,

$$\mathrm{var}_A(\mathring\lambda_{rs}) = \mathrm{var}_A\{E_{Y|A}(L_{rs})\} = \mathrm{var}_Y(L_{rs}) - E_A\{\mathrm{var}_{Y|A}(L_{rs})\}$$
$$= \mathrm{var}_Y(l_{rs}) - E_A\{\mathrm{var}_{Y|A}(\mathring{l}_{rs})\} = O(n) - E_A\{O_p(n)\} = O(n).$$

With respect to the distribution of A, $\mathring\lambda_{rs}$ has mean λ_{rs} and variance of order $O(n)$; thus, $\mathring\lambda_{rs} - \lambda_{rs}$, $\mathring\lambda_{rst} - \lambda_{rst}$, etc. have zero expectation and variance of order $O(n)$. It is assumed that these quantities are of order $O_p(n^{1/2})$ and have joint cumulants of order $O(n)$ with respect to the distribution of A. In particular, $\mathring\lambda_{rs} = \lambda_{rs} + O_p(n^{1/2})$, $\mathring\lambda_{rst} = \lambda_{rst} + O_p(n^{1/2})$, etc.

Assume that the differentiation of the identity $\mathring\lambda_{rs} = \lambda_{rs} + O_p(n^{1/2})$ yields $\mathring\lambda_{rs/t} = \lambda_{rs/t} + O_p(n^{1/2})$. As a rule, differentiation of an asymptotic relation will preserve the asymptotic order, but care is necessary; see Barnforff-Nielsen & Cox (1994, Exercise 5.4). The difference between $\mathring\lambda_{rs/t}$ and $\lambda_{rs/t}$ indicated here constitutes an additional assumption of our calculations, though validity is immediate in particular cases, such as the example considered in § 5. Then

$$\mathring\lambda_{rs,t} = \mathring\lambda_{rs/t} - \mathring\lambda_{rst} = \lambda_{rs/t} - \lambda_{rst} + O_p(n^{1/2}) = \lambda_{rs,t} + O_p(n^{1/2}).$$

By working with the conditional density of Y given A in place of the marginal density of Y, it follows that $R(\psi)$ is conditionally $N(0, 1)$ to an error of order $O(n^{-1/2})$ and that the error in the $N(0, 1)$ approximation to the conditional distribution of $R(\psi)$ can be reduced to order $O(n^{-1})$ by adjusting for the conditional mean of $R(\psi)$. Denote the conditional mean by $\mathring{\mu}_F$; then $R(\psi) - \mathring{\mu}_F$ has conditionally the $N(0, 1)$ distribution to the error of order $O(n^{-1})$.

The calculations of DiCiccio & Stern (1994a) can be applied to the conditional distribution of $R(\psi)$ to show that

$$\mathring{\mu}_F = -\tfrac{1}{2}\mathring{\eta}\mathring{\lambda}_{rst}\mathring{\lambda}^{r1}\mathring{\lambda}^{st} - \tfrac{1}{6}\mathring{\eta}^3\mathring{\lambda}_{rst}\mathring{\lambda}^{r1}\mathring{\lambda}^{s1}\mathring{\lambda}^{t1} + \mathring{\eta}\mathring{\lambda}_{rs/t}\mathring{\lambda}^{r1}\mathring{\lambda}^{st} + \tfrac{1}{2}\mathring{\eta}^3\mathring{\lambda}_{rs/t}\mathring{\lambda}^{r1}\mathring{\lambda}^{s1}\mathring{\lambda}^{t1} + O(n^{-3/2}),$$

where $\mathring{\eta} = (-\mathring{\lambda}^{11})^{-1/2}$. Furthermore, the preceding comparisons of the λs and the corresponding $\mathring{\lambda}$s show that $\mathring{\mu}_F = \mu_F + O_p(n^{-1})$, and hence $\mu_B = \mathring{\mu}_F + O_p(n^{-1})$, provided the condition (1) holds. It follows that Welch–Peers priors satisfying (1) produce approximate confidence limits having a conditional coverage error of order $O(n^{-1})$. This result is implicit in DiCiccio & Martin (1993), who compare the Bayesian percentage points under an objective prior with the approximate, conditional confidence limits derived from Barndorff-Nielsen's r^* statistic (Barndorff-Nielsen, 1986); see also Nicolau (1993).

Since $\mu_B = \mathring{\mu}_F + O_p(n^{-1})$, it follows that $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-1})$. In resolving the conditional properties of second-order probability-matching priors, the crucial issue turns out to be to examine circumstances under which $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$ holds. Since $\mathring{\lambda}_r = 0$, an argument analogous to that given in the Appendix for $f(\theta)$ defined in (A1) shows that $E_{Y|A}(\mu_B) = E_{Y|A}\{f(\theta)\} + O(n^{-3/2})$, so the crucial criterion reduces to $E_{Y|A}\{f(\theta)\} = \mathring{\mu}_F + O(n^{-3/2})$, which holds provided

$$\mathring{\eta}\frac{\partial \log \pi(\theta)}{\partial \theta^r}\mathring{\lambda}^{r1} = \mathring{\eta}\mathring{\lambda}_{rs/t}\mathring{\lambda}^{r1}\mathring{\lambda}^{st} + \tfrac{1}{2}\mathring{\eta}^3\mathring{\lambda}_{rs/t}\mathring{\lambda}^{r1}\mathring{\lambda}^{s1}\mathring{\lambda}^{t1} + O(n^{-3/2}). \tag{3}$$

Suppose that π is a first-order probability-matching prior, so that (1) holds, that satisfies further the condition $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$. Let $\mathring{\sigma}_F^2$ denote the conditional frequentist variance of $R(\psi) - \mu_B$. From arguments similar to the ones given previously that showed $\mathring{\mu}_F = \mu_F + O_p(n^{-1})$, it follows that $\mathring{\sigma}_F^2 = \sigma_F^2 + O_p(n^{-3/2})$. To be specific, recall that

$$\sigma_F^2 = 1 + a_F + 2\eta\mu_{F/r}\lambda^{r1} - \mu_F^2 + O(n^{-3/2}),$$

where $a_F + 2\eta\mu_{F/r}\lambda^{r1} - \mu_F^2$ is of order $O(n^{-1})$ and can be expressed, to error of order $O(n^{-3/2})$, as a function of the λs. By applying identical calculations, which require the condition $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$, to the conditional distribution, it follows that

$$\mathring{\sigma}_F^2 = 1 + \mathring{a}_F + 2\mathring{\eta}\mathring{\mu}_{F/r}\mathring{\lambda}^{r1} - \mathring{\mu}_F^2 + O(n^{-3/2}),$$

where $\mathring{a}_F + 2\mathring{\eta}\mathring{\mu}_{F/r}\mathring{\lambda}^{r1} - \mathring{\mu}_F^2$ is of order $O_p(n^{-1})$ and can be expressed, to error of order $O_p(n^{-3/2})$, as the identical function as can its unconditional version, with each λ being replaced by its corresponding $\mathring{\lambda}$. Since by assumption each $\mathring{\lambda}$ differs from its corresponding λ by $O_p(n^{-1/2})$, it follows that $\mathring{\sigma}_F^2 = \sigma_F^2 + O_p(n^{-3/2})$. Hence, the condition that ensures that $\pi(\theta)$ is a second-order probability-matching prior in the marginal frequentist sense also ensures that it is a second-order probability-matching prior in the conditional frequentist sense.

Thus, if $\pi(\theta)$ is a second-order probability-matching prior in the marginal frequentist sense, then it is also a second-order probability-matching prior in the conditional frequentist

sense, provided $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$, i.e., provided (3) holds. This is satisfied if

$$\frac{\partial \log \pi(\theta)}{\partial \theta^r} \mathring{\lambda}^{r1} = \mathring{\lambda}_{rs/t} \mathring{\lambda}^{r1} \mathring{\lambda}^{st} + \tfrac{1}{2}\mathring{\eta}^2 \mathring{\lambda}_{rs/t} \mathring{\lambda}^{r1} \mathring{\lambda}^{s1} \mathring{\lambda}^{t1}. \tag{4}$$

We summarize our conclusions in the following theorem.

THEOREM 1. *Suppose that the prior π is such that* (4) *holds, so that* $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$. *If the prior also satisfies* (1) *and* (2), *so that* $\sigma_B^2 = \mathring{\sigma}_F^2 + O(n^{-3/2})$, *then the Bayesian quantile is second-order conditional probability matching.*

Proof. Given Y, let $\psi_l \equiv \psi^{1-\alpha}(\pi, Y)$ be the posterior $1 - \alpha$ quantile for the interest parameter, so that $\mathrm{pr}\{\psi \leqslant \psi^{1-\alpha}(\pi, Y) \mid Y\} = 1 - \alpha$. Also, given Y and provided that the log-likelihood is unimodal, $R(\psi)$ is a monotonic decreasing function of ψ. Therefore, $\mathrm{pr}\{R(\psi) \geqslant R(\psi_l) \mid Y\} = 1 - \alpha$, where $R(\psi)$ is the signed root statistic constructed from Y. That is,

$$\mathrm{pr}\left\{ \frac{R(\psi) - \mu_B}{\sigma_B} \geqslant \frac{R(\psi_l) - \mu_B}{\sigma_B} \mid Y \right\} = 1 - \alpha.$$

Since the posterior distribution of $\{R(\psi) - \mu_B\}/\sigma_B$ is $N(0, 1)$ to an error of order $O_p(n^{-3/2})$, it follows that, by the delta method from Hall (1992, §2.7), $\mathrm{pr}[N(0, 1) \geqslant \{R(\psi_l) - \mu_B\}/\sigma_B] = 1 - \alpha + O(n^{-3/2})$, so that $\{R(\psi_l) - \mu_B\}/\sigma_B = z_\alpha + O(n^{-3/2})$, in terms of the $N(0, 1)$ quantile z_α defined by $\Phi(z_\alpha) = \alpha$.

By the monotonicity of $R(\psi)$ given Y, the event $\psi \leqslant \psi_l$ is equivalent to the event $R(\psi) \geqslant R(\psi_l)$. Thus, from a conditional frequentist perspective, given an ancillary statistic $A = a$, we have, again using the delta method and by the frequentist distributional result for the mean-adjusted signed root statistic,

$$
\begin{aligned}
\mathrm{pr}(\psi \leqslant \psi_l \mid A = a) &= \mathrm{pr}\left\{ \frac{R(\psi) - \mu_B}{\sigma_B} \geqslant \frac{R(\psi_l) - \mu_B}{\sigma_B} \mid A = a \right\} \\
&= \mathrm{pr}\left\{ \frac{R(\psi) - \mu_B}{\mathring{\sigma}_F} + O_p(n^{-3/2}) \geqslant \frac{R(\psi_l) - \mu_B}{\sigma_B} \mid A = a \right\} \\
&= \mathrm{pr}\left\{ \frac{R(\psi) - \mu_B}{\mathring{\sigma}_F} \geqslant z_\alpha + O_p(n^{-3/2}) \mid A = a \right\} + O(n^{-3/2}) \\
&= \mathrm{pr}\{N(0, 1) \geqslant z_\alpha\} + O(n^{-3/2}) = 1 - \alpha + O(n^{-3/2}). \qquad \square
\end{aligned}
$$

5. DISCUSSION

Conditions under which a Bayesian prior on a scalar interest parameter in the presence of a nuisance parameter achieves probability matching have been shown in this paper to have direct interpretation in terms of the frequentist and Bayesian distributional properties of the signed root likelihood ratio statistic. A prior that is first-order probability matching in a marginal sense is necessarily first-order conditional probability matching. A prior that is second-order probability matching in the marginal sense yields second-order conditional probability matching, provided a simple condition (4), which may be interpreted as a conditional version of the marginal Welch–Peers condition (1), holds. Second-order unconditional matching is seen to correspond to matching of the frequentist and Bayesian variances of a specific mean-adjusted version of the signed root statistic, where the adjustment is by the Bayesian mean, under a prior that ensures

first-order matching. Similarly, conditional probability matching is typically achieved only under a very particular prior specification. These conclusions indicate, to our mind, that second-order probability matching is too stringent a criterion to be useful in practice.

For illustration, consider a location-scale model, with $Y_1, \ldots, Y_n$ an independent sample from a density of the form

$$f(y; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{y - \mu}{\sigma}\right).$$

The appropriate conditioning ancillary is the configuration statistic $A = (A_1, \ldots, A_n) = \{(Y_1 - \hat{\mu})/\hat{\sigma}, \ldots, (Y_n - \hat{\mu})/\hat{\sigma}\}$, and the loglikelihood is of the form

$$L(\mu, \sigma) = -n \log \sigma + \sum_{i=1}^{n} h\left(\frac{\hat{\mu} - \mu}{\sigma} + A_i \frac{\hat{\sigma}}{\sigma}\right),$$

with $h(\cdot) = \log g(\cdot)$. Then, since the conditional distribution of $\{(\hat{\mu} - \mu)/\sigma, \hat{\sigma}/\sigma\}$, given $A = a = (a_1, \ldots, a_n)$, does not depend on (μ, σ), simple calculations show $\mathring{\lambda}_{\mu\mu} = B/\sigma^2$, $\mathring{\lambda}_{\mu\sigma} = C/\sigma^2$, $\mathring{\lambda}_{\sigma\sigma} = D/\sigma^2$, where B, C, D are nonzero functions of a. The corresponding unconditional quantities are of the form $\lambda_{\mu\mu} = B_g/\sigma^2$, $\lambda_{\mu\sigma} = C_g/\sigma^2$, $\lambda_{\sigma\sigma} = D_g/\sigma^2$ for constants B_g, C_g and D_g depending only on g and n. In the notation of §4, it is then immediate that the differentiation of the relation $\mathring{\lambda}_{rs} = \lambda_{rs} + O_p(n^{1/2})$ yields the assumed relation $\mathring{\lambda}_{rs/t} = \lambda_{rs/t} + O_p(n^{1/2})$ in this case.

Consider the case where the location parameter μ is the interest parameter, with σ as the nuisance parameter. Then the right-hand side of the matching condition (4) is easily seen after some algebra to reduce to $\sigma C/E$, where $E = BD - C^2$. The left-hand side of (4) is, after some manipulation,

$$\frac{\sigma^2 D}{E} \frac{\partial \log \pi}{\partial \mu} - \frac{\sigma^2 C}{E} \frac{\partial \log \pi}{\partial \sigma},$$

so that the matching condition is satisfied if and only if

$$\frac{\partial \log \pi}{\partial \mu} = 0, \quad \frac{\partial \log \pi}{\partial \sigma} = -\frac{1}{\sigma},$$

which gives $\pi(\mu, \sigma) \propto 1/\sigma$. This prior is easily seen to satisfy (1) and (2), and is therefore second-order conditional probability matching. In fact, the prior is (Lawless, 1982, Appendix G) exact conditional probability matching, and is (Datta & Mukerjee, 2004, §2.5.2) the unique second-order unconditional matching prior.

Analogous calculations yield the same conclusion in the case where the scale parameter σ is the interest parameter, with μ the nuisance parameter: the unique solution to the conditional matching condition (4) is $\pi(\mu, \sigma) \propto 1/\sigma$, which is again exact conditional probability matching. Now, however, second-order marginal-matching priors are not necessarily unique. In the Cauchy location-scale model, for example, any prior of the form $\pi(\mu, \sigma) \propto d(\mu)/\sigma$, for any smooth positive function $d(\cdot)$, is second-order unconditional matching; see Datta & Mukerjee (2004, §2.5.2).

In general, therefore, second-order conditional probability matching is only achieved by the exact conditional probability-matching prior $\pi(\mu, \sigma) \propto 1/\sigma$. We have noted, however, that first-order conditional probability matching is achieved by any first-order unconditional probability-matching prior in the class satisfying the Welch–Peers condition (1).

Appendix

Derivation of μ_B

Repeated differentiation of the definition $M(\psi) = L\{\tilde{\theta}(\psi)\}$ yields $M_1(\hat{\psi}) = 0$, $M_{11}(\hat{\psi}) = 1/\hat{L}^{11}$, $M_{111}(\hat{\psi}) = \hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1}/(\hat{L}^{11})^3$. Taylor expansion about $\hat{\psi}$ yields

$$W(\psi) = -\hat{M}_{11}(\hat{\psi} - \psi)^2 + \tfrac{1}{3}\hat{M}_{111}(\hat{\psi} - \psi)^3 + O_p(n^{-1}).$$

Consequently,

$$W(\psi) = \{Z(\psi)\}^2 - \tfrac{1}{3}\hat{H}^3\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1}\{Z(\psi)\}^3 + O_p(n^{-1}),$$

where $Z(\psi) = (-\hat{M}_{11})^{1/2}(\hat{\psi} - \psi) = \hat{H}(\hat{\psi} - \psi)$ and $\hat{H} = (-\hat{M}_{11})^{1/2}$. Since $\hat{M}_{111} = -\hat{H}^6\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1}$, the signed root statistic $R(\psi)$ has the expansion

$$R(\psi) = Z(\psi) - \tfrac{1}{6}\hat{H}^3\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1}\{Z(\psi)\}^2 + O_p(n^{-1}).$$

The Laplace approximation to the marginal posterior density function of ψ given by Tierney & Kadane (1986) can be written as

$$\pi_{\psi|Y}(\psi) = c\exp\{B(\psi) + M(\psi) - M(\hat{\psi})\}\{1 + O(n^{-3/2})\}$$

for values of the argument ψ such that $\psi = \hat{\psi} + O(n^{-1/2})$, where c is a normalizing constant, $\pi(\theta) = \pi(\psi, \phi)$ is the prior density and

$$B(\psi) = -\tfrac{1}{2}\log\left\{\frac{|-L_{ij}(\psi, \tilde{\phi}_\psi)|}{|-L_{ij}(\hat{\psi}, \hat{\phi})|}\right\} + \log\left\{\frac{\pi(\psi, \tilde{\phi}_\psi)}{\pi(\hat{\psi}, \hat{\phi})}\right\}.$$

Differentiation of $B(\psi)$ yields

$$\hat{B}_1 = \tfrac{1}{2}\hat{H}^2\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{rs} + \tfrac{1}{2}\hat{H}^4\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1} - \hat{H}^2\left.\frac{\partial\log\pi(\theta)}{\partial\theta^r}\right|_{\theta=\hat{\theta}}\hat{L}^{r1},$$

which is of order $O(1)$. By Taylor expansion about $\hat{\psi}$,

$$\pi_{\psi|Y}(\psi) = (2\pi)^{-1/2}\hat{H}\exp\{-\tfrac{1}{2}\hat{H}^2(\hat{\psi} - \psi)^2\}\{1 - \hat{B}_1(\hat{\psi} - \psi) - \tfrac{1}{6}\hat{M}_{111}(\hat{\psi} - \psi)^3 + O(n^{-1})\}$$

so the marginal posterior density of $Z(\psi)$ has the expansion

$$\pi_{Z(\psi)|Y}(z) = (2\pi)^{-1/2}e^{-z^2/2}\{1 - \hat{H}^{-1}\hat{B}_1 z - \tfrac{1}{6}\hat{H}^{-3}\hat{M}_{111}z^3 + O(n^{-1})\},$$

from which it follows that

$$\mu_B = -\hat{H}^{-1}\hat{B}_1 - \tfrac{1}{2}\hat{H}^{-3}\hat{M}_{111} - \tfrac{1}{6}\hat{H}^3\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1} + O(n^{-1})$$

$$= -\tfrac{1}{2}\hat{H}\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{rs} - \tfrac{1}{6}\hat{H}^3\hat{L}_{rst}\hat{L}^{r1}\hat{L}^{s1}\hat{L}^{t1} + \hat{H}\left.\frac{\partial\log\pi(\theta)}{\partial\theta^r}\right|_{\theta=\hat{\theta}}\hat{L}^{r1} + O(n^{-1}).$$

A more careful analysis that takes higher-order terms into account, i.e., that includes terms of order $O(n^{-1})$, shows that the error term in the preceding formula is actually $O(n^{-3/2})$.

To be specific, a higher-order expansion of $R(\psi)$ would involve a term in $\{Z(\psi)\}^3$ with a coefficient that is of order $O(n^{-1})$. Moreover, a higher-order expansion of the marginal posterior density of $Z(\psi)$ would involve terms in z^4 and z^6, each having coefficients of order $O(n^{-1})$. Now, when calculating the contribution to the expectation from the $Z(\psi)$ term in $R(\psi)$, the z^4 and z^6 terms in the marginal posterior density of $Z(\psi)$ would yield 0. When calculating the contribution to the expectation from the $\{Z(\psi)\}^2$ term in $R(\psi)$, which has a coefficient of order $O(n^{-1/2})$, the z^4 and z^6 terms in the marginal posterior density of $Z(\psi)$, which have coefficients of order $O(n^{-1})$, would yield a term of order $O(n^{-3/2})$. When calculating the contribution to the expectation from the $\{Z(\psi)\}^3$ term in $R(\psi)$, which has a coefficient of order $O(n^{-1})$, the z and z^3 terms in the marginal posterior density of $Z(\psi)$, which have coefficients

of order $O(n^{-1/2})$, would yield terms of order $O(n^{-3/2})$, while the z^4 and z^6 terms in the marginal posterior density of $Z(\psi)$ would yield 0. Similar calculations, to an error of order $O(n^{-3/2})$, were given by DiCiccio & Stern (1993).

Higher-order analysis of the Bayesian mean

Since the focus here is the frequentist mean of μ_B, it is appropriate to describe error terms in expansions by using the $O_p(\cdot)$ notation instead of the $O(\cdot)$ notation. To simplify subsequent calculations, μ_B is conveniently written in the form $\mu_B = f(\hat{\theta}) + O_p(n^{-3/2})$, where

$$f(\theta) = -\tfrac{1}{2}HL_{rst}L^{r1}L^{rs} - \tfrac{1}{6}H^3 L_{rst}L^{r1}L^{s1}L^{t1} + H\frac{\partial \log \pi(\theta)}{\partial \theta^r}L^{r1}, \tag{A1}$$

and $H = (-L^{11})^{-1/2}$. Recall that $f(\theta)$ is $O_p(n^{-1/2})$. If the prior density $\pi(\theta)$ satisfies the Welch–Peers condition, then $\mu_B = \mu_F + O_p(n^{-1})$, and hence $E(\mu_B) = \mu_F + O(n^{-1})$. We establish here that $E(\mu_B) = \mu_F + O(n^{-3/2})$.

Note that, by Taylor expansion about θ,

$$f(\hat{\theta}) = f(\theta) + f_r(\theta)(\hat{\theta}^r - \theta^r) + O_p(n^{-3/2}) = f(\theta) - f_r(\theta)\lambda^{rs}l_s + O_p(n^{-3/2}),$$

since $\hat{\theta}^r - \theta^r = \lambda^{rs}l_s + O_p(n^{-1})$. Now $f_r(\theta)$ is $O_p(n^{-1/2})$, and it is a function of the likelihood derivatives. Let $f_r^\lambda(\theta)$ be the quantity obtained when each of the likelihood derivatives in $f_r(\theta)$ is replaced by its corresponding expectation, so $f_r^\lambda(\theta)$ is a nonrandom quantity depending on θ. Then $f_r(\theta) = f_r^\lambda(\theta) + O_p(n^{-1})$, and hence

$$E\{f_r(\theta)\lambda^{rs}l_s\} = E\{f_r^\lambda(\theta)\lambda^{rs}l_s + O_p(n^{-3/2})\} = O(n^{-3/2}).$$

It follows that $E(\mu_B) = E\{f(\theta)\} + O(n^{-3/2})$, so it is required to show that $E\{f(\theta)\} = \mu_F + O(n^{-3/2})$.

We have $L^{rs} = \lambda^{rs} - \lambda^{rt}\lambda^{su}l_{tu} + O_p(n^{-2})$, so that

$$H = (-L^{11})^{-1/2} = \eta - \tfrac{1}{2}\eta^3\lambda^{r1}\lambda^{s1}l_{rs} + O_p(n^{-1/2}), \quad H^3 = \eta^3 - \tfrac{3}{2}\eta^3\lambda^{r1}\lambda^{s1}l_{rs} + O_p(n^{-1/2}).$$

Consider first the final term of $f(\theta)$; the other terms can be handled similarly. It follows from the preceding equations that

$$H\frac{\partial \log \pi(\theta)}{\partial \theta^r}L^{r1} = \eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1} - \tfrac{1}{2}\eta^3\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1}\lambda^{s1}\lambda^{t1}l_{st}$$
$$- \eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{rs}\lambda^{t1}l_{st} + O_p(n^{-3/2}).$$

Hence,

$$E\left\{H\frac{\partial \log \pi(\theta)}{\partial \theta^r}L^{r1}\right\} = \eta\frac{\partial \log \pi(\theta)}{\partial \theta^r}\lambda^{r1} + O(n^{-3/2})$$
$$= \eta\lambda_{rs/t}\lambda^{r1}\lambda^{st} + \tfrac{1}{2}\eta^3\lambda_{rs/t}\lambda^{r1}\lambda^{s1}\lambda^{t1} + O(n^{-3/2}),$$

by virtue of the Welch–Peers condition (1). The other terms in $f(\theta)$ have

$$E(-\tfrac{1}{2}HL_{rst}L^{r1}L^{rs}) = -\tfrac{1}{2}\eta\lambda_{rst}\lambda^{r1}\lambda^{rs} + O(n^{-3/2}),$$
$$E(-\tfrac{1}{6}H^3 L_{rst}L^{r1}L^{s1}L^{t1}) = -\tfrac{1}{6}\eta^3\lambda_{rst}\lambda^{r1}\lambda^{s1}\lambda^{t1} + O(n^{-3/2}),$$

and combining these expressions yields the desired result, namely $E\{f(\theta)\} = \mu_F + O(n^{-3/2})$.

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.

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

DiCiccio, T. J. & Martin, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc.* B **55**, 305–16.

DiCiccio, T. J. & Stern, S. E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80**, 731–40.

DiCiccio, T. J. & Stern, S. E. (1994a). Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics. *Scand. J. Statist.* **21**, 447–60.

DiCiccio, T. J. & Stern, S. E. (1994b) Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. R. Statist. Soc.* B **56**, 397–408.

DiCiccio, T. J. & Young, G. A. (2010). Objective Bayes and conditional inference in exponential families. *Biometrika* **97**, 497–504.

Hall, P. G. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer.

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

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.

Nicolau, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. R. Statist. Soc.* B **55**, 377–90.

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.

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

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–6.

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.