Statistics & Decisions 13, 131 – 139 (1995)

FREQUENTIST VALIDITY OF HIGHEST POSTERIOR DENSITY REGIONS IN THE PRESENCE OF NUISANCE PARAMETERS

Jayanta K. Ghosh and Rahul Mukerjee

Abstract

Priors ensuring frequentist validity, up to $o(n^{-1})$, of credible regions based on the highest posterior density have been characterized in the presence of nuisance parameters. In this connection, the consequences of an orthogonal parametrization have also been discussed.

1. Introduction

In recent years, there has been a revival of interest in problems relating to approximate frequentist validity of posterior credible regions. As noted in Tibshirani (1989), apart from providing a method for constructing accurate frequentist confidence regions, such studies are also helpful in defining noninformative priors which could be potentially useful for comparative purposes in a Bayesian analysis. The results available in the literature in this general area include those related to the approximate frequentist validity of one-sided posterior regions based on posterior quantiles ([18], [13], [16], [17], [12]), posterior regions based on the inversion of likelihood ratio and related statistics ([7], [8]) and highest posterior density (HPD) regions ([14], [9]). We refer to [11] and [15] for further interesting results and references.

As for the problem of characterizing priors ensuring approximate frequentist validity of HPD regions, it appears that not much work has as yet been done on models involving nuisance parameters. In consideration of the current interest on such models, in this work we propose to fill up this gap to some extent. As a special case, models with an orthogonal parametrization ([5]) have been considered. An advantage of our approach is that it does not require an explicit evaluation of all the coefficients involved in the (approximate) posterior density of the interest parameter and this keeps the algebra relatively simple (see Section 3).

AMS subject classification: 62F15, 62F25, 62E20

Key words and phrases: highest posterior density, non-informative prior, parametric orthogonality.

2. The HPD Region

Let $\{X_i\}$, $i \geq 1$, be a sequence of independent and indentically distributed, possibly vector-valued, random variables with common density $f(x;\theta)$, where $\theta = (\theta_1, \theta_2)'$, θ_1 is the interest parameter and θ_2 is the nuisance parameter. For notational simplicity, in this and the next section, we assume that θ_1 and θ_2 are both one-dimensional, i.e., $\theta \in R^2$ or some open subset thereof. The case where θ_1 and θ_2 are both possibly multi-dimensional will be briefly discussed in Section 4. We make the assumptions of [10]. Let θ have a prior density $\pi(.)$ which is positive and thrice continuously differentiable for all θ. In case $\pi(.)$ is not proper, as assumed in [10], we shall require that there exists an $n_0(>0)$ such that for all $X_1, \ldots, X_{n_0}$, the posterior of θ given $X_1, \ldots, X_{n_0}$ is proper. Let $X = (X_1, \ldots, X_n)'$ where n is the sample size. All formal expansions for the posterior, as used here, are valid for sample points in a set S, which may be defined along the line of Bickel and Ghosh (1990; Section 2 with $m = 3$), with P_θ-probability $1+0(n^{-2})$ uniformly on compact sets of θ. Let $\ell(\theta) = n^{-1} \sum_{i=1}^{n} \log f(X_i;\theta)$. Denoting the maximum likelihood estimator of θ by $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$, for $i,j = 0,1,2,\ldots$, let $\ell_{ij}(\theta) = D_1^i D_2^j \ell(\theta)$, $b_{ij} = \ell_{ij}(\hat{\theta})$, $c_{ij} = -b_{ij}$,

$$C = \begin{pmatrix} c_{20} & c_{11} \\ c_{11} & c_{02} \end{pmatrix},$$

where D_i is the operator of partial differentiation with respect to θ_i. The matrix C is positive definite over the set S mentioned above. Also, we write $\hat{\pi} = \pi(\hat{\theta})$, and for $i,j = 0,1,2,\ldots$, $\pi_{ij}(\theta) = D_1^i D_2^j \pi(\theta)$, $\hat{\pi}_{ij} = \pi_{ij}(\hat{\theta})$.

Let $h = n^{\frac{1}{2}}(\theta_1 - \hat{\theta}_1)$. Then starting from equation (2.2) in [7] or equation (3.1) in [12], it can be seen that the posterior density of h, under the prior $\pi(.)$, is given by

$$\begin{aligned} \tilde{\pi}(h|X) &= \phi(h;Q^{-1})[1 + n^{-\frac{1}{2}}\{B_1(\pi,X)h + B_3(X)h^3\} + n^{-1}\{B_2(\pi,X)(h^2 - Q^{-1}) \\ &\quad + B_4(\pi,X)(h^4 - 3Q^{-2}) + B_6(X)(h^6 - 15Q^{-3})\}] + o(n^{-1}), \quad (2.1) \end{aligned}$$

where $\phi(.;Q^{-1})$ is the univariate normal density with mean zero and variance Q^{-1}, $Q = c_{20} - c_{02}^{-1}c_{11}^2$,

$$\begin{aligned} B_1(\pi,X) &= B_{11}(\pi,X) + B_{12}(X),\ B_2(\pi,X) = B_{21}(\pi,X) + B_{22}(X), && (2.2a) \\ B_4(\pi,X) &= B_{41}(\pi,X) + B_{42}(X),\ B_{41}(\pi,X) = B_{11}(\pi,X)B_3(X), && (2.2b) \\ B_{11}(\pi,X) &= \hat{\pi}^{-1}(\hat{\pi}_{10} - c_{11}c_{02}^{-1}\hat{\pi}_{01}),\ B_{12}(X) = \frac{1}{2}(b_{12}c_{02}^{-1} - b_{03}c_{11}c_{02}^{-2}), && (2.2c) \\ B_3(X) &= \frac{1}{6}(b_{30} - 3b_{21}c_{11}c_{02}^{-1} + 3b_{12}c_{11}^2c_{02}^{-2} - b_{03}c_{11}^3c_{02}^{-3}), && (2.2d) \\ B_{21}(\pi,X) &= (2\hat{\pi})^{-1}\{\hat{\pi}_{20} - 2c_{11}c_{02}^{-1}\hat{\pi}_{11} + c_{11}^2c_{02}^{-2}\hat{\pi}_{02} + c_{02}^{-1}\hat{\pi}_{10}(b_{12} - b_{03}c_{11}c_{02}^{-1}) \\ &\quad + c_{02}^{-1}\hat{\pi}_{01}(b_{21} - 3b_{12}c_{11}c_{02}^{-1} + 2b_{03}c_{11}^2c_{02}^{-2})\}, && (2.2e) \end{aligned}$$

and $B_{22}(X)$, $B_{42}(X)$, $B_6(X)$, like $B_{12}(X)$, $B_3(X)$, are functions of X which are at most of order $0(1)$ and do not involve π or its derivatives. The detailed expressions for $B_{22}(X)$,

$B_{42}(X)$, $B_6(X)$ are not required in the sequel. From (2.1), it is not hard to see that $\tilde{\pi}(h|X)$ can as well be expressed as

$$\tilde{\pi}(h|X) = \phi(0;Q^{-1})\exp\{\frac{1}{2}n^{-1}B_7(\pi,X) - \frac{1}{2}U(\pi,X,h)\}/\{1+n^{-1}B_8(\pi,X)\} + o(n^{-1}), \quad (2.3)$$

where $B_7(\pi,X) = Q^{-1}\{B_1(\pi,X)\}^2$, $B_8(\pi,X) = Q^{-1}B_2(\pi,X)+3Q^{-2}B_4(\pi,X)+15Q^{-3}B_6(X)$, and

$$\begin{aligned} U(\pi,X,h) &= h^2Q - 2n^{-\frac{1}{2}}\{B_1(\pi,X)h + B_3(X)h^3\} \\ &\quad + n^{-1}[B_7(\pi,X) + \{B_1(\pi,X)h + B_3(X)h^3\}^2 \\ &\quad - 2\{B_2(\pi,X)h^2 + B_4(\pi,X)h^4 + B_6(X)h^6\}] \end{aligned} \quad (2.4)$$

The inclusion of $B_7(\pi,X)$ in $U(\pi,X,h)$ helps in expressing the approximate posterior characteristic function (c.f.) of $U(\pi,X,h)$ in a neat form.

From (2.1), (2.4), it can be seen after some algebra that the approximate posterior c.f. of $U(\pi,X,h)$ under the prior $\pi(.)$ is given by

$$(1-2\xi)^{-\frac{1}{2}}[1+n^{-1}\{F_0(\pi,X) + (1-2\xi)^{-1}F_1(\pi,X) + (1-2\xi)^{-2}F_2(X)\}] + o(n^{-1}),$$

where $\xi = (-1)^{\frac{1}{2}}t$, and

$$F_0(\pi,X) = 3Q^{-2}\{B_1(\pi,X)B_3(X) - B_4(\pi,X)\} - 15B_6(X)Q^{-3}, \quad (2.5a)$$

$$F_1(\pi,X) = 3Q^{-2}\{B_4(\pi,X) - B_1(\pi,X)B_3(X)\} + \frac{15}{2}Q^{-3}\{B_3(X)\}^2, \quad (2.5b)$$

$$F_2(X) = 15Q^{-3}[B_6(X) - \frac{1}{2}\{B_3(X)\}^2]. \quad (2.5c)$$

For positive integral ν, let $\Omega_\nu(.)$ and $\omega_\nu(.)$ denote respectively the cumulative distribution function and the probability density function of a central chi-square variate with ν degrees of freedom (d.f.). Also, let z^2 be the upper α-point $(0<\alpha<1)$ of a central chi-square variate with 1 d.f. and

$$U_{(1-\alpha)}(\pi,X) = z^2 - (n\omega_1(z^2))^{-1}\{F_0(\pi,X)\Omega_1(z^2) + F_1(\pi,X)\Omega_3(z^2) + F_2(X)\Omega_5(z^2)\}. \quad (2.6)$$

Then inverting the above approximate posterior c.f. of $U(\pi,X,h)$, which can be justified as in [3], [4], [1], we get

$$P^{\pi}[U(\pi,X,h) \le U_{(1-\alpha)}(\pi,X)|X] = 1-\alpha+o(n^{-1}), \quad (2.7)$$

where $p^{\pi}[.|X]$ is the posterior probability measure under the prior $\pi(.)$. Writing $h \equiv h(\theta_1)$, by (2.3), (2.7), the HPD region for θ_1 with posterior coverage probability $1-\alpha+o(n^{-1})$ can be expressed as

$$\{\theta_1 : \tilde{\pi}(h(\theta_1)|X) \ge \phi(0;Q^{-1})\exp(\frac{1}{2}n^{-1}B_7(\pi,X) - \frac{1}{2}U_{(1-\alpha)}(\pi,X))/(1+n^{-1}B_8(\pi,X))\},$$

which is approximable, up to $o(n^{-1})$, by $\{\theta_1 : U(\pi,X,h(\theta_1)) \le U_{(1-\alpha)}(\pi,X)\}$.

3. Frequentist Calculations

We now proceed along the line of Ghosh and Mukerjee (1991, 1993) to calculate $P_\theta\{U(\pi,X,h) \leq U_{(1-\alpha)}(\pi,X)\}$. To that effect, we consider a prior $\pi^*(.)$ satisfying the regularity conditions in [1, Section 2 with $m=3$] which are slightly stronger than those in [10], and make Edgeworth assumptions as in [1, p. 1078]. Then by (2.2), (2.5), (2.6), after some simplification,

$$P^{\pi^*}[U(\pi,X,h) \leq U_{(1-\alpha)}(\pi,X)|X] = 1-\alpha+n^{-1}\{\Omega_3(z^2)-\Omega_1(z^2)\}g(\pi,\pi^*,X) + o(n^{-1}), \tag{3.1a}$$

where

$$g(\pi,\pi^*,X) = Q^{-1}[B_{21}(\pi^*,X)-B_{21}(\pi,X)+B_1(\pi,X)\{B_{11}(\pi,X)-B_{11}(\pi^*,X)\}] + 3Q^{-2}B_3(X)\{B_{11}(\pi^*,X)-B_{11}(\pi,X)\}. \tag{3.1b}$$

The derivation of (3.1) is again based on the inversion of the approximate posterior c.f. of $U(\pi,X,h)$ under the prior $\pi^*(.)$.

Let $I=((I_{ij}))$ be the 2×2 Fisher information matrix (per observation) at θ which is assumed to be positive definite at each θ. For $i,j,i',j'=0,1,2,\ldots$, let

$$K_{ij}=E_\theta\{D_1^iD_2^j\log f(X_1;\theta)\},\ K_{ij.i'j'}=E_\theta[\{D_1^iD_2^j\log f(X_1,\theta)\}\{D_1^{i'}D_2^{j'}\log f(X_1,\theta)\}].$$

Also, let

$$A=I_{11}-I_{22}^{-1}I_{12}^2,\ \lambda=-I_{12}/I_{22},\ \mu=K_{21}+2\lambda K_{12}+\lambda^2K_{03},\ \psi=K_{30}+3\lambda K_{21}+3\lambda^2K_{12}+\lambda^3K_{03}.$$

Note that A, λ, μ, ψ, I_{ij}, K_{ij}, $K_{ij.i'j'}$, etc. are all functions of θ.

Let $\pi\equiv\pi(\theta)$, $\pi^*\equiv\pi^*(\theta)$. By (2.2), (3.1b), under θ,

$$g(\pi,\pi^*,X)=\bar{g}(\pi,\pi^*,\theta)+o(1), \tag{3.2a}$$

where

$$\begin{aligned}\bar{g}(\pi,\pi^*,\theta) = {} & A^{-1}[(2\pi^*)^{-1}(D_1^2\pi^*+2\lambda D_1D_2\pi^*+\lambda^2D_2^2\pi^*)-(2\pi)^{-1}(D_1^2\pi+2\lambda D_1D_2\pi+\lambda^2D_2^2\pi)\\ & +(2I_{22}\pi^*)^{-1}\mu D_2\pi^*-(2I_{22}\pi)^{-1}\mu D_2\pi\\ & +\pi^{-1}(D_1\pi+\lambda D_2\pi)\{\pi^{-1}(D_1\pi+\lambda D_2\pi)-(\pi^*)^{-1}(D_1\pi^*+\lambda D_2\pi^*)\}]\\ & +\frac{1}{2}A^{-2}\psi\{(\pi^*)^{-1}(D_1\pi^*+\lambda D_2\pi^*)-\pi^{-1}(D_1\pi+\lambda D_2\pi)\}.\end{aligned} \tag{3.2b}$$

In consideration of (3.1a), (3.2a), for a fixed $\pi(.)$,

$$P_\theta\{U(\pi,X,h)\leq U_{(1-\alpha)}(\pi,X)\} = 1-\alpha+n^{-1}\{\Omega_3(z^2)-\Omega_1(z^2)\}\{(2\pi)^{-1}H_\pi(\theta)\} + o(n^{-1}), \tag{3.3}$$

where the factor $\{(2\pi)^{-1}H_\pi(\theta)\}$ in the right-hand side is obtained by integrating $\bar{g}(\pi,\pi^*,\theta)$ by parts with respect to a $\pi^*(\theta)$ such that $\pi^*(.)$ and its first partial derivatives vanish on the boundary of a rectangle containing θ and then allowing $\pi^*(.)$ to converge weakly to the degenerate measure at θ. This approach, reminiscent of that in [6], was used earlier in [7] and [9]. Explicit calculation, based on (3.2b), shows that

$$\begin{aligned} H_\pi(\theta) &= D_1^2(A^{-1}\pi) + 2D_1D_2(\lambda A^{-1}\pi) + D_2^2(\lambda^2A^{-1}\pi) - D_2\{\mu(I_{22}A)^{-1}\pi\} \\ &\quad - D_2(\lambda\psi A^{-2}\pi) - D_1(\psi A^{-2}\pi). \end{aligned} \tag{3.4}$$

By (3.3), for each α, frequentist validity, up to $o(n^{-1})$, holds for the HPD region for θ_1 if and only if $\pi(.)$ satisfies the partial differential equation

$$H_\pi(\theta) = 0, \tag{3.5}$$

where $H_\pi(\theta)$ is given by (3.4). Note that the equation (3.5) does not involve α. *This equation represents the main result of this section.*

EXAMPLE 1. Consider the location-scale model with $f(x;\theta) = \theta_2^{-1}f^*\{\theta_2^{-1}(x-\theta_1)\}$, where the location parameter $\theta_1(-\infty < \theta_1 < \infty)$ is of interest and the scale parameter $\theta_2(>0)$ is the nuisance parameter. Then λ is a constant, each of I_{22} and A is proportional to θ_2^{-2} and each of μ and ψ is proportional to θ_2^{-3} (provided they exist). Hence, $\pi(\theta) \propto \theta_2^{-1}$ satisfies (3.5) (cf. [2]). In a location-scale model if instead the scale parameter is of interest then denoting the scale and location parameters by θ_1 and θ_2 respectively it can be similarly seen that $\pi(\theta) \propto \theta_1^{-1}$ satisfies (3.5).

In the rest of this section, we consider models where global parametric orthogonality ([5]) holds, i.e., I_{12} equals zero indentically in θ. Then $\lambda = 0$, $A = I_{11}$, $\mu = K_{21}$, $\psi = K_{30}$, and using the regularity condition $D_1I_{11} = -(K_{30} + K_{10.20})$, from (3.4) it can be seen that (3.5) reduces to

$$D_1\{I_{11}^{-1}(D_1\pi)\} + D_1(I_{11}^{-2}K_{10.20}\pi) - D_2\{(I_{11}I_{22})^{-1}K_{21}\pi\} = 0. \tag{3.6}$$

Under global parametric orthogonality, priors of the form

$$\pi(\theta) = d(\theta_2)I_{11}^{\frac{1}{2}}, \tag{3.7}$$

where $d(\theta_2)(>0)$ is a function of θ_2 alone, are of special interest. Tibshirani (1989) showed that such priors ensure frequentist validity, up to $o(n^{-\frac{1}{2}})$, of the posterior quantiles of θ_1. It is easy to see that a prior of the form (3.7) satisfies (3.6) if and only if $d(\theta_2)$ satisfies

$$\frac{1}{2}D_1\{I_{11}^{-\frac{3}{2}}(K_{10.20} - K_{30})\} - \{d(\theta_2)\}^{-1}D_2\{d(\theta_2)(I_{11}^{\frac{1}{2}}I_{22})^{-1}K_{21}\} = 0. \tag{3.8}$$

The above is similar to but not identical with a condition in [12] who studied the problem of ensuring frequentist validity, up to $o(n^{-1})$, of the posterior quantiles of θ_1.

EXAMPLE 2. We consider a version of the exponential regression model of Cox and Reid (1987). This is given by

$$f(x;\theta) = \Pi_{i=1}^r[\theta_2^{-1}\exp(-\theta_1y_i)\exp\{-\theta_2^{-1}x^{(i)}e^{-\theta_1y_i}\}],\ x^{(1)},\ \ldots,\ x^{(r)} > 0,$$

where $x = (x^{(1)}, \ldots, x^{(r)})'$, $-\infty < \theta_1 < \infty$, $\theta_2 > 0$, $r \geq 2$, and $y_1, \ldots, y_r$ are constants, not all equal, satisfying $y_1 + \ldots + y_r = 0$. Then global parametric orthogonality holds and one can check that $I_{11} = y_1^2 + \ldots + y_r^2$, $I_{22} = r\theta_2^{-2}$, $K_{21} = \theta_2^{-1} I_{11}$, $K_{30} = -K_{10.20} = y_1^3 + \cdots + y_r^3$. Hence (3.8) is uniquely satisfied by $d(\theta_2) \propto \theta_2^{-1}$. The same solution was reported in [12] in their context.

EXAMPLE 3. This relates to the ratio of independent normal means ([5]). Let $f(x;\theta) = \phi(x^{(1)} - s_1(\theta))\phi(x^{(2)} - s_2(\theta))$, where $\phi(.)$ is the standard univariate normal density, $x = (x^{(1)}, x^{(2)})'$, $s_1(\theta) = \theta_1\theta_2/(\theta_1^2+1)^{\frac{1}{2}}$, $s_2(\theta) = \theta_2/(\theta_1^2+1)^{\frac{1}{2}}$, and $\theta_1, \theta_2 > 0$. Note that $\theta_1 = s_1(\theta)/s_2(\theta)$ and that under this parametrization global parametric orthogonality holds. Here $I_{11} = \theta_2^2/(\theta_1^2+1)^2$, $I_{22} = 1$, $K_{21} = -\theta_2/(\theta_1^2+1)^2$, $K_{30} = -3K_{10.20} = 6\theta_1\theta_2^2/(\theta_1^2+1)^3$. Hence it can be seen that no solution to (3.8) is available, i.e., no prior of the form (3.7) satisfies (3.6). Nevertheless, (3.6) does have a solution, e.g., $\pi(\theta) = \theta_2(\theta_1^2+1)$ solves (3.6). Unfortunately, however, under this solution of (3.6), the posterior of θ given X may not be proper even for large n unless the parameter space is so restricted that $s_2(\theta)$ is bounded away from zero.

EXAMPLE 4. This relates to the ratio of independent exponential means. Let

$$f(x;\theta) = \{s_1(\theta)s_2(\theta)\}^{-1}\exp[-\{(s_1(\theta))^{-1}x^{(1)} + (s_2(\theta))^{-1}x^{(2)}\}],\ x^{(1)},\ x^{(2)} > 0,$$

where $x = (x^{(1)}, x^{(2)})'$, $s_1(\theta) = \theta_2\theta_1^{\frac{1}{2}}$, $s_2(\theta) = \theta_2\theta_1^{-\frac{1}{2}}$, and $\theta_1, \theta_2 > 0$. Note that $\theta_1 = s_1(\theta)/s_2(\theta)$ and that under this parametrization global parametric orthogonality holds. Here $I_{11} = \frac{1}{2}\theta_1^{-2}$, $I_{22} = 2\theta_2^{-2}$, $K_{21} = \frac{1}{2}(\theta_2\theta_1^2)^{-1}$, $K_{30} = -3K_{10.20} = \frac{3}{2}\theta_1^{-3}$. Hence (3.8) is uniquely solved by $d(\theta_2) \propto \theta_2^{-1}$, which is the same as the solution obtained in [12] in their context.

4. Extension to the General Multiparameter Case

Before concluding, we briefly indicate the result corresponding to (3.4) and (3.5) in the general multiparameter case. The set-up is as in Section 2 with the change that now $\theta = (\theta_1, \ldots, \theta_p)'$ is p-dimensional, $\theta^{(1)} = (\theta_1, \ldots, \theta_q)'$ represents the q-dimensional interest parameter and $\theta^{(2)} = (\theta_{q+1}, \ldots, \theta_p)'$ represents the $(p\text{-}q)$-dimensional nuisance parameter, where $1 \leq q < p$. Let $I = ((I_{ij}))$ be the $p \times p$ per observation Fisher information matrix which is assumed to be positive definite at each θ. Let $I^{-1} = ((I^{ij}))$ and for $1 \leq i, j, k, u \leq p$, let

$$I^{(1)}_{ijku} = I^{ij}I^{ku} + I^{ik}I^{ju} + I^{iu}I^{jk}.$$

Partition I as

$$I = \begin{pmatrix} I_{(11)} & I_{(12)} \\ I_{(21)} & I_{(22)} \end{pmatrix},$$

where $I_{(11)}$ is $q \times q$ and $I_{(12)}$, $I_{(21)}$, $I_{(22)}$ are of appropriate orders. For $1 \leq i,\ j \leq q$, let τ_{ij} denote the (i,j)th element of $I_{(11)} - I_{(12)}I_{(22)}^{-1}I_{(21)}$, and for $1 \leq i \leq q$, $q+1 \leq v \leq p$, define $\lambda_{vi} = \sum_{u=1}^{q} I^{vu}\tau_{ui}$. Also, with the rows and columns of $I_{(22)}^{-1}$ labelled by $q+1, \ldots, p$, for $q+1 \le v,\ v' \le p$, let $\sigma_{vv'}$, denote the (v,v')th element of $I^{-1}_{(22)}$. For $1 \le i,j,k \le p$, let $L_{ijk} = E_\theta\{D_iD_jD_k \log f(X_1;\theta)\}$, where D_i is the operator of partial differentiation with respect to θ_i and for $1 \le i,\ j,\ k \le q$, let

$$\begin{aligned}\psi_{ijk} &= L_{ijk} + 3\Sigma_v L_{ijv}\lambda_{vk} + 3\Sigma_v\Sigma_{v'} L_{ivv'}\lambda_{vj}\lambda_{v'k} \\ &\quad + \Sigma_v\Sigma_{v'}\Sigma_{v''} L_{vv'v''}\lambda_{vi}\lambda_{v'j}\lambda_{v''k}.\end{aligned}$$

where the summation over each of $v,\ v',\ v''$, is on the range $q+1$ to p. It may be noted that the quantities defined above are functions of θ.

Under this set-up, it can be shown that frequentist validity, up to $o(n^{-1})$, holds for the HPD regions for $\theta^{(1)}$ if and only if $\pi \equiv \pi(.)$ satisfies

$$\begin{aligned}\Sigma_i\Sigma_j\ [D_iD_j(I^{ij}\pi) &+ 2\Sigma_v D_iD_v(\lambda_{vj}I^{ij}\pi) + \Sigma_v\Sigma_{v'}D_vD_{v'}(\lambda_{vi}\lambda_{v'j}I^{ij}\pi) \\ &- \Sigma_w D_w\{(\Sigma_v L_{ijv}\sigma_{wv} + 2\Sigma_v\Sigma_{v'}L_{ivv'}\lambda_{vj}\sigma_{wv'} \\ &\qquad + \Sigma_v\Sigma_{v'}\Sigma_{v''}L_{vv'v''}\lambda_{vi}\lambda_{v'j}\sigma_{wv''})I^{ij}\pi\}] \\ &- \frac{1}{3}\Sigma_i\Sigma_j\Sigma_k\Sigma_u\{D_u(\psi_{ijk}I^{(1)}_{ijku}\pi) + \Sigma_v D_v(\psi_{ijk}\lambda_{vu}I^{(1)}_{ijku}\pi)\} = 0, \quad (4.1)\end{aligned}$$

where the summations on i,j,k,u are over the range 1 to q and the summations on w,v,v',v'' are on the range $q+1$ to p. Equation (4.1) generalizes (3.4), (3.5) to the multiparameter case. It can be proved proceeding as in Sections 2 and 3 but with heavier notation and algebra. It may be checked that (3.4), (3.5) and (4.1) are in agreement with the findings in [14] and [9] who considered the same problem in the absence of nuisance parameters.

Acknowlegement. The work of RM was supported by a grant from the Centre for Management and Development Studies, Indian Institute of Management, Calcutta. Thanks are due to a referee for helpful comments.

References

[1] Bickel, P.J. and Ghosh, J.K.: A decomposition for the likelihood ratio statistic and the Bartlett correction - A Bayesian argument. Ann. Statist. *18*, *1070–1090* *(1990)*.

[2] Box, G.E.P. and Tiao, G.C.: Bayesian Inference in Statistical Analysis. Reading, Massachusetts, Addison-Wesley (1973).

[3] Chandra, T.K.: Asymptotic expansions and deficiency. Ph.D. thesis, Indian Statistical Institute.

[4] Chandra, T.K. and Ghosh, J.K.: Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. Sankhyā A 41, 22–47 (1979).

[5] Cox, D.R. and Reid, N.: Parameter orthogonality and approximate conditional inference (with discussion). J. Roy. Statist. Soc. B 49, 1–39 (1987).

[6] Dawid, A.P.: Fisherian inference in likelihood and prequential frames of reference (with discussion). J. Roy. Statist. Soc. B 53, 79–109 (1991).

[7] Ghosh, J.K. and Mukerjee, R.: Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. J. Multivariate Anal. 38, 385–393 (1991).

[8] Ghosh, J.K. and Mukerjee, R.: Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. J. Roy. Statist. Soc. B 54, 867–875 (1992).

[9] Ghosh, J.K. and Mukerjee, R.: Frequentist validity of highest posterior density regions in the multiparameter case. Ann. Inst. Statist. Math. 45, 293–302 (1993).

[10] Johnson, R.A.: Asymptotic expansions associated with posterior distributions. Ann. Math. Statist. 41, 851–864 (1970).

[11] Lee, C.B.: Comparison of frequentist coverage probability and Bayesian posterior coverage probability, and applications. Ph.D. thesis, Purdue University (1989).

[12] Mukerjee, R. and Dey, D.: Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. Biometrika 80, 499–505 (1993).

[13] Peers, H.W.: On confidence sets and Bayesian probability points in the case of several parameters. J. Roy. Statist. Soc. B 27, 9–16 (1965).

[14] Peers, H.W.: Confidence properties of Bayesian interval estimates. J. Roy. Statist. Soc. B 30, 535–544 (1968).

[15] Severini, T.A.: On the relationship between Bayesian and non-Bayesian interval estimates. J. Roy. Statist. Soc. B 53, 611–618 (1991).

[16] Stein, C.: On coverage probability of confidence sets based on a prior distribution. In: Sequential Methods in Statistics, Banach Center Publications 16, 485–514. Warsaw, Polish Scientific Publishers (1985).

[17] Tibshirani, R.: Noninformative priors for one parameter of many. Biometrika 76, 604–608 (1989).

[18] Welch, B. and Peers, H.W.: On formulae for confidence points based on integrals of weighted likelihoods. J. Roy. Statist. Soc. B 25, 318–329 (1963).

Jayanta K. Ghosh
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399
U.S.A.

Rahul Mukerjee
Indian Institute of Management
P.O. Box 16757
Calcutta 700 027
INDIA