

# Coverage probability bias, objective Bayes and the likelihood principle

By TREVOR J. SWEETING

*Department of Mathematics and Statistics, University of Surrey, Guildford,  
GU2 7XH U.K.*

t.sweeting@surrey.ac.uk

## SUMMARY

We review objective Bayes procedures based on both parametric and predictive coverage probability bias and explore the extent to which such procedures contravene the likelihood principle in the case of a scalar parameter. The discussion encompasses choice of objective priors, objective posterior probability statements and objective predictive probability statements. We conclude with some remarks concerning the future development and implementation of objective priors based on small coverage probability bias.

*Some key words:* Approximate Bayesian inference; Coverage probability bias; Higher-order asymptotics; Likelihood principle; Objective Bayes.

## 1. INTRODUCTION

We discuss objective Bayesian methodology for parametric models from the point of view of coverage probability bias. There are many different approaches to deriving ‘objective’ Bayesian procedures; that is, procedures which are formally Bayesian, and for which there is some justification for using them in an automatic way. There is an extensive literature on the construction of objective, or reference, priors; see, for example, Kass & Wasserman (1996) and Bernardo & Ramón (1998). An objective Bayesian procedure may be regarded as a default method which can be applied in cases where prior information is sparse or not well understood, or differs between the stakeholders. Such procedures have some

appeal, especially in complex models where it may be very difficult to make an adequate subjective assessment of the prior distributions involved.

In the present paper we discuss objective Bayes methods which have some justification in terms of repeated sampling performance characteristics. More specifically, throughout this paper we shall use the term ‘objective Bayes’ to mean any Bayesian procedure, which can be justified on the basis of small coverage probability bias; that is, any Bayesian procedure that is well calibrated. Here we use the term ‘coverage probability’ to refer to the frequentist probability of some statement either about a model parameter or about a future observation. The rationale behind these ideas is that the resulting Bayesian statements are endowed with additional frequentist validity. A major aim of the paper is to elucidate some of the issues underlying objective Bayes construction via coverage probability bias with a view to the future development of this approach, especially in relation to objective prior construction for multiparameter models.

We interpret objective Bayes in a broad sense, encompassing objective choice of priors, objective posterior probability statements and objective predictive probability statements. There is an interplay between these ideas. For example, one might fix on an objective prior under which a wide range of posterior probability statements will have small coverage probability bias. In this case it could be argued that this would be an appropriate default prior to be used when prior information is sparse or not well understood. Alternatively, one might focus on a suitable proper prior to reflect one’s uncertainty, but then limit one’s probability statements to a restricted class for which the associated coverage probability bias is small. Such probability statements will be well calibrated, so this can be regarded as a frequency-based robustness property to alternative prior specifications. This might have some appeal when there is some concern about the specific prior adopted. We remark that, if a small coverage probability bias justification is used to obtain a default objective prior, then this in itself says nothing about the suitability of this prior for producing approximate subjective posterior statements. This would have to be argued by appeal to other considerations, such as

invariance or arguments based on limiting posteriors.

Since the concept of objective Bayes used here is based on coverage probability, alternative sampling rules may give rise to different objective Bayes methods. Thus the approach contravenes the strong likelihood principle, and is therefore incoherent from a Bayesian point of view. This is unavoidable, but we can at least ask to what extent coverage probability bias is affected by the sampling rule.

Following some preliminaries in §2, in §3 we review the development of objective Bayesian inference via assessment of the associated parametric asymptotic coverage probability bias. As is well known, the situation is quite different for one-sided and two-sided probability statements. The discussion covers choice of both objective priors and objective regions. In §4 there is a parallel review of recent work by Datta et al. (2000) on predictive coverage probability bias. In §5 we focus on the case of a single real parameter and explore the extent to which the methods of §§3 and 4 conform to the likelihood principle. In general, the construction of objective priors based on one-sided coverage probability bias will depend on the sampling rule. However, it may be possible to construct a data-dependent prior with the required matching property for a given class of sampling rules. In the case of likelihood and perturbed likelihood regions for the natural parameter of an exponential family model, it is shown that the objective classes of Bayes priors obtained in §3 are unchanged under certain classes of stopping rules. In the case of predictive coverage probability bias, it is shown that the construction of objective predictive intervals does not depend on the sampling rule at all and hence the likelihood principle is fully respected to the asymptotic order considered. Some concluding remarks are made in §6, including a discussion on the future development and implementation of objective priors based on small coverage probability bias. Finally, Appendices contain proofs of results stated in §§3 - 5. These proofs, which are only sketched here, use techniques in Sweeting (1995a, b) and Sweeting (1999).

## 2. PRELIMINARIES

We begin by describing the parametric framework and the main regularity conditions and definitions. Let  $X_n$  be an observation vector and assume that the family of density functions  $p_n(\cdot; \theta)$  of  $X_n$  depends on the unknown parameter  $\theta \in \Omega$ , an open subset of  $\mathcal{R}^d$ . Let  $\Pi_\Omega$  be the class of all probability densities on  $\Omega$  which are continuous and positive throughout  $\Omega$ . Then the posterior density of the unknown parameter value  $\Theta$ , regarded as a random quantity having prior density  $\pi \in \Pi_\Omega$ , is given by

$$\pi(\theta|X_n) \propto L_n(\theta)\pi(\theta), \tag{1}$$

where  $L_n(\theta) \propto p_n(X_n; \theta)$  is the likelihood function.

We assume throughout that there exists a unique local maximum likelihood estimator  $\hat{\theta}_n$  of  $\theta$ . Write  $l_n(\theta) = \log L_n(\theta)$ ,  $j_n(\theta) = -l_n''(\theta)$ , the negative Hessian matrix,  $J_n = j_n(\hat{\theta}_n)$ , the observed information matrix,  $u_n(\theta) = -J_n^{-1/2}l_n'(\theta)$ , a normalised score statistic,  $U_n = u_n(\Theta)$  and  $I_n(\theta) = E_\theta\{j_n(\theta)\}$ , Fisher's information matrix. The asymptotic results given in the present section and in §§3 and 4 are for the case of independent and identically distributed observations. However, under suitable regularity conditions these results will carry over to more general settings and in §5 we assume that the relevant results hold under specified alternative sampling rules which give rise to  $X_n$ .

Consider, then, the case where  $X_n$  comprises  $n$  independent observations from a common distribution  $F(\cdot; \theta)$  with density  $f(\cdot; \theta)$ . In that case we can define  $i(\theta) = I_n(\theta)/n$ , Fisher's information per observation. All approximations will be in the asymptotic sense as  $n \rightarrow \infty$  and we shall assume that the problem is sufficiently regular for the validity of the various asymptotic approximations. In particular, although stochastic  $O(\cdot)$  and  $o(\cdot)$  terms generally refer to probability convergence, we will assume that these can be strengthened to convergence in first mean whenever required in a proof.

The main development will be for the case  $d = 1$ , with an indication of multiparameter extensions. In this case, we write  $l_n^i$  for the  $i$ th derivative of  $l_n$

with respect to  $\theta$  and define the functions  $\rho_{ijk\dots} = \rho_{ijk\dots;n}$  and  $\bar{\rho}_{ijk\dots} = \bar{\rho}_{ijk\dots;n}$  by

$$\begin{aligned}\rho_{ijk\dots}(\theta) &= E_{\theta}\{l_n^i(\theta)l_n^j(\theta)l_n^k(\theta)\dots\} \\ \bar{\rho}_{ijk\dots}(\theta) &= \{I_n(\theta)\}^{-(i+j+k+\dots)/2}\rho_{ijk\dots}(\theta).\end{aligned}$$

The likelihood root, that is the signed loglikelihood ratio or directed likelihood, is defined by

$$R_n = \text{sgn}(\Theta - \hat{\theta}_n)[2\{l_n(\hat{\theta}_n) - l_n(\Theta)\}]^{1/2}.$$

Under a given prior  $\pi \in \Pi_{\Omega}$  the posterior distribution of the standardised likelihood root

$$\tilde{R}_n = R_n + R_n^{-1} \log(U_n/R_n) - R_n^{-1} \log\{\pi(\Theta)/\pi(\hat{\theta}_n)\}$$

is standard normal to  $O(n^{-3/2})$  (Sweeting, 1995a). Note that  $\tilde{R}_n$  may also be written as

$$\tilde{R}_n = R_n + R_n^{-1} \log(T_n/R_n) - R_n^{-1} \log\{\xi(\Theta)/\xi(\hat{\theta}_n)\}, \quad (2)$$

where  $T_n = \{i(\hat{\theta}_n)/i(\Theta)\}^{1/2}U_n$  and  $\xi(\theta) = \{i(\theta)\}^{-1/2}\pi(\theta)$ . The reason for the introduction of  $i(\theta)$  here is that each of the terms on the right-hand side of (2) is invariant under reparameterisation. Quantities related to (2) are also used as approximate pivots in frequentist inference. In particular, the sampling distribution of (2) without the final term is standard normal to third order for all location models. For a general model, again without the final term, a suitable modification of  $T_n$  yields a third-order standard normal sampling distribution, conditional on a suitable asymptotically ancillary statistic. See, for example, Barndorff-Nielsen (1986), Section 6.6 of Barndorff-Nielsen & Cox (1994) and Pierce & Peters (1994). The connection of (2) with frequentist inference in exponential families is discussed in Sweeting (1995b).

### 3. PARAMETRIC COVERAGE PROBABILITY BIAS

#### 3.1 *One-sided parametric intervals*

We begin by investigating the construction of objective priors on the basis of small coverage probability bias associated with posterior parametric statements in the case  $d = 1$ . Let  $0 < \alpha < 1$  and suppose that  $\pi \in \Pi_\Omega$ . Let  $t(\pi, \alpha)$  denote the upper  $\alpha$ -quantile of the posterior distribution of  $\theta$ , satisfying

$$\text{pr}\{\Theta > t(\pi, \alpha) | X_n\} = \alpha. \quad (3)$$

We wish to know under what conditions the coverage probability bias associated with (3) will be approximately zero. Formally, we ask when is it true that, to a given asymptotic order of approximation,

$$\text{pr}\{\theta > t(\pi, \alpha) | \theta\} = \alpha \quad (4)$$

very weakly, by which we shall mean that the relation

$$\text{pr}\{\Theta > t(\pi, \alpha) | \tau\} \equiv \int \text{pr}\{\theta > t(\pi, \alpha) | \theta\} \tau(\theta) d\theta = \alpha \quad (5)$$

holds to the asymptotic order considered for all smooth prior densities  $\tau \in \Pi_\Omega$ ; see, for example, Woodroffe (1986). For the present purpose of investigating frequency based robustness to alternative prior specifications, requiring relation (5) to hold only for smooth priors  $\tau$ , and not necessarily for point-mass priors, will be sufficient. It can also be argued that the weaker relation (5) makes sense from a frequentist standpoint if one is really concerned with performance in repeated use, as opposed to repeated sampling. We further note that this formulation avoids a technical issue in asymptotic analysis, where (5) may hold in an asymptotic sense for every smooth prior  $\tau \in \Pi_\Omega$ , but not for every point-mass prior.

The answer to the question posed above depends on the desired degree of approximation in (4). To the first order of approximation, the answer is simple and well known. Relation (4) holds to  $O(n^{-1/2})$  for every  $\pi \in \Pi_\Omega$  and  $0 < \alpha < 1$ . Thus, to the first order of approximation there is no coverage probability

bias associated with posterior probability statements. To the second order of approximation, the answer is again well known: relation(4) holds to  $O(n^{-1})$  for all  $\alpha$  if and only if  $\pi(\theta) \propto \{i(\theta)\}^{1/2}$  (Welch & Peers, 1963). Thus Jeffreys' invariant prior is second-order matching with respect to one-sided parametric regions and, on the basis of minimising the coverage probability bias associated with statements of the type (3), the objective prior is Jeffreys' prior.

The emergence of Jeffreys' prior as the required matching prior can also be seen by noting that when  $\pi(\theta)$  is Jeffreys' prior then (2) becomes

$$\tilde{R}_n^J \equiv R_n + R_n^{-1} \log(T_n/R_n). \quad (6)$$

It turns out that the sampling distribution of  $\tilde{R}_n^J$  is also approximately standard normal to  $O(n^{-1})$ ; see Barndorff-Nielsen & Chamberlin (1991), DiCiccio & Martin (1993) and Sweeting (1995a). In fact the conditional sampling distribution of  $\tilde{R}_n^J$  remains standard normal to this order given an arbitrary second-order locally ancillary statistic (Sweeting, 1995b). We will say that a prior distribution is ' $k$ th-order stably matching' if, conditional on any  $k$ th-order locally ancillary statistic, the relevant coverage probability bias is  $O(n^{-k/2})$ . Thus Jeffreys' prior is second-order stably matching with respect to one-sided parametric regions.

In general, the approximation in (4) is no better than  $O(n^{-1})$ , unless the skewness measure  $\bar{\rho}_{111}(\theta)$  happens to be independent of  $\theta$ . In that case the approximation is  $O(n^{-3/2})$  under Jeffreys' prior, as was shown by Welch & Peers (1963). The reason for this is that  $\bar{\rho}'_{111}(\theta)$  appears as a factor in the next term in the asymptotic expansion of the coverage probability.

In the multiparameter case the  $O(n^{-1/2})$  equivalence property continues to hold on account of the first-order equivalence of the Bayesian and frequentist normal approximations. To the second order of approximation, the Welch & Peers (1963) result was generalised by Tibshirani (1989) and Nicolaou (1993), following earlier work by Stein (1985). Suppose that  $(\psi, \lambda)$  is a one-to-one transformation of  $\theta$ , where  $\psi = \psi(\theta)$  is a scalar parametric function of interest, and  $\lambda = \lambda(\theta)$  is chosen to be orthogonal to  $\psi$  in the sense of Cox & Reid (1987). Then  $O(n^{-1})$  matching occurs with respect to one-sided parametric statements about  $\psi$  when

$\pi(\psi, \lambda) \propto g(\lambda)i_{\psi\psi}(\psi, \lambda)^{1/2}$ , where  $g(\lambda) > 0$  is arbitrary and  $i_{\psi\psi}(\psi, \lambda)$  is the information component associated with  $\psi$ . We note that in general it is not possible to obtain a single prior that will provide the desired  $O(n^{-1})$  matching property with respect to different scalar interest parameters  $\psi_1$  and  $\psi_2$ .

### 3.2 Two-sided parametric intervals

In this section we describe the construction of both objective priors and regions on the basis of small coverage probability bias associated with likelihood and related regions in the case  $d = 1$ . For such regions, coverage probability bias can be reduced to  $O(n^{-2})$  as a result of the cancellation of directional errors. However, the associated family of matching priors may not contain Jeffreys' prior, in which case the coverage probability bias for one-sided intervals will be  $O(n^{-1/2})$ . Matching priors for likelihood regions have been proposed in the literature as default priors for Bayesian inference, and the primary purpose of this section, along with §5.3, is to explore and clarify the issues surrounding probability matching for two-sided statements.

Let  $0 < \alpha < 1$  and suppose that  $\pi \in \Pi_\Omega$ . Let  $(t_1(\pi, \alpha), t_2(\pi, \alpha))$  be any interval having posterior probability  $\alpha$ ; that is,

$$\text{pr}\{t_1(\pi, \alpha) < \Theta < t_2(\pi, \alpha) | X_n\} = \alpha \quad (7)$$

As before, we ask when it is also true that, to a given degree of approximation,

$$\text{pr}\{t_1(\pi, \alpha) < \theta < t_2(\pi, \alpha) | \theta\} = \alpha \quad (8)$$

very weakly.

Although we trivially deduce from the discussion in §3.1 that (8) holds to  $O(n^{-1/2})$  for any smooth prior, and to  $O(n^{-1})$  under Jeffreys' prior, the order of approximation is usually better than this. In particular, suppose that the region in (7) is a likelihood region; that is,  $L_n\{t_1(\pi, \alpha)\} = L_n\{t_2(\pi, \alpha)\}$ . Then (8) holds to  $O(n^{-1})$  for every  $\pi \in \Pi_\Omega$  (Hartigan, 1966). Furthermore, it can be shown that (8) holds to  $O(n^{-2})$  for all priors of the form

$$\pi(\theta) = \{I_n(\theta)\}^{1/2} e^{-\tau_n(\theta)} \left[ k_1 + k_2 \int \{I_n(\theta)\}^{1/2} e^{\tau_n(\theta)} d\theta \right], \quad (9)$$



where  $\tau_n(\theta) = \frac{1}{2} \int \{I_n(\theta)\}^{1/2} \bar{\rho}_{111}(\theta) d\theta$  and  $k_1, k_2$  are arbitrary constants (Ghosh & Mukerjee, 1992a; Sweeting, 1995a). That is, every prior of the form (9) is fourth-order matching with respect to likelihood regions and, on the basis of minimising the coverage probability bias associated with statements of the type (7), we obtain the objective class of priors (9). Notice that this class contains Jeffreys' prior only in the special case where the skewness  $\bar{\rho}_{111}(\theta)$  is independent of  $\theta$ .

In the case of the exponential family model

$$l_n(\theta) = s\theta - n\kappa(\theta) \tag{10}$$

the above class of priors has a concise form when expressed in terms of the canonical parameter  $\theta$ . Since  $\rho_{12}(\theta) = 0$ , we have  $\rho_{111}(\theta) = I'_n(\theta) - 2\rho_{12}(\theta) = I'_n(\theta)$ , so that

$$\tau_n(\theta) = \frac{1}{2} \int \{I_n(\theta)\}^{-1} I'_n(\theta) d\theta = \frac{1}{2} \log I_n(\theta).$$

Therefore the class of matching priors (9) becomes simply

$$\pi(\theta) = k_1 + k_2\mu(\theta), \tag{11}$$

where  $\mu(\theta)$  is the expectation parameter. Note that this family includes the uniform prior on the canonical scale, which has some intuitive appeal as this is the flat parameterisation for the family. Note also that the family of priors (11) is equivalent to forming uniform priors in all parameterisations of the form

$$\phi = k_1\theta + k_2\kappa(\theta)$$

whenever these transformations are one-to-one. In §5 we establish a connection between this class of priors and a class of linear stopping rules. We note that the form (11) holds more generally for any model where  $\theta$  is the unique affine parameter for the model for which  $\rho_{12}(\theta) = 0$ . In this case  $\mu(\theta)$  is any indefinite integral of  $i(\theta)$ .

As previously remarked, the family (11) of matching priors for likelihood regions usually does not contain Jeffreys' prior, so the  $O(n^{-2})$  two-sided matching is

bought at the expense of  $O(n^{-1})$  one-sided matching. Clearly, this is an unattractive feature of objective priors obtained in this way. Furthermore, it can be argued that there is nothing special about likelihood regions for the construction of objective priors. Severini (1993) showed that, by a judicious choice of interval, it is possible to have agreement between posterior and coverage probabilities to third order, under any given smooth prior. This is an attractive property, especially where there may be some concern about the prior adopted, since it gives additional frequentist validity to subjective Bayesian probability statements. On the other hand, approximate frequentist confidence regions can be constructed which take some account of moderate prior information, having an approximate Bayesian interpretation under that prior. In Sweeting (1999) such regions are referred to as ‘Bayes-confidence regions’.

Consider any modification of the likelihood root  $R_n$  of the form

$$\bar{R}_n = R_n + R_n^{-1} \log g_n(R_n), \quad (12)$$

where  $(g_n)$  is a function sequence in the class  $\mathcal{A}^4$  defined in Sweeting (1999), and consider intervals of the form  $|\bar{R}_n| \leq c$ . We will refer to such intervals as ‘perturbed likelihood regions’ based on  $\bar{R}_n$ . Note that these are likelihood regions when  $g_n \equiv 1$  in (12). Then, for any prior  $\pi \in \Pi_\Omega$ , perturbed likelihood regions based on  $\bar{R}_n$  satisfying (7) will also satisfy (8) to  $O(n^{-1})$ . The reason for this is the cancellation of the  $O(n^{-1/2})$  directional errors in the sampling densities.

Now let  $\pi = \pi_0$ , a specific prior in  $\Pi_\Omega$ , and define  $\xi_0(\theta) = \{i(\theta)\}^{-1/2} \pi_0(\theta)$ . It is shown in Sweeting (1999) that perturbed likelihood regions based on the particular modification

$$\bar{R}_n = R_n + \frac{3}{2} R_n^{-1} \log(T_n/R_n) - \frac{1}{2} R_n^{-1} \log\{\xi_0(\Theta)/\xi_0(\hat{\theta}_n)\} \quad (13)$$

which satisfy (7) also satisfy (8) to  $O(n^{-2})$ . Such intervals are therefore invariant fourth-order Bayes-confidence regions. Furthermore, it is shown in Sweeting (1999) that (8) holds to  $O(n^{-3/2})$  conditionally on any second-order locally ancillary statistic. Thus all that is required for the construction of Bayes-confidence

intervals is a suitable change in the numerical coefficients of the two adjustment components of the standardised likelihood root (2). Numerical evaluations in Sweeting (1999) indicate that the coverage properties of these intervals are very good, even in quite small samples, although there may be some degradation for parameter values in regions of relatively low prior density. Thus, for a given prior, restriction to this class of intervals provides some robustness in the sense that the coverage properties under alternative neighbouring smooth priors would be approximately the same.

In view of the above results, there does not seem to be a compelling case for constructing default priors via probability matching for likelihood regions, since perturbed likelihood regions could serve the same purpose and would yield different default priors. One way to resolve the tension between objective Bayes procedures based on one- or two-sided intervals would be to use the matching prior, that is, Jeffreys' prior, for one-sided probabilities and, if desired, to use perturbed likelihood regions associated with Jeffreys' prior for two-sided probability statements. These regions will have the additional frequentist robustness property referred to above. Thus, one would take  $g_n(R_n) = (T_n/R_n)^{3/2}$  in (12), giving rise to  $O(n^{-1})$  coverage probability bias for all posterior statements coupled with  $O(n^{-2})$  coverage probability bias for two-sided posterior statements.

In the case of an exponential family model, an alternative simpler form of (13) is available, namely

$$\bar{R}_n = R_n - \frac{1}{2}R_n^{-1} \log\{\pi_0(\Theta)/\pi_0(\hat{\theta}_n)\}, \quad (14)$$

where  $\theta$  is the canonical parameter of the family. As is the case with likelihood regions, to each perturbed likelihood region based on (14) there is an associated class of matching priors. The following extension of (11) is proved in Appendix 1.

THEOREM 1. *Consider perturbed likelihood regions based on (14) for an exponential family model. Then the coverage probability bias of these regions is  $O(n^{-2})$  for all priors of the form*

$$\pi(\theta) = \pi_0(\theta) \{k_1 + k_2 \int \pi_0(\theta)^{-1} i(\theta) d\theta\}, \quad (15)$$

where  $k_1, k_2$  are arbitrary constants.

As with (11), in §5 we will relate this class of matching priors to a certain class of stopping rules. We note that a similar result to Theorem 1 can be obtained for a general model by considering the perturbed regions based on (13).

The extension to the multiparameter case of approximate Bayesian inference based on directed likelihood is described in Sweeting (1996). It should be possible to extend the two-sided results described above for a single parameter in a multiparameter setting by initial transformation to a directed likelihood vector; see also Ghosh & Mukerjee (1992b) for a review of matching priors for posterior and frequentist inference in the multiparameter case.

## 4. PREDICTIVE COVERAGE PROBABILITY BIAS

### 4.1. One-sided predictive intervals

In this section we consider the construction of objective priors on the basis of small coverage probability bias associated with Bayesian predictive statements. This question is discussed in Datta et al. (2000). Let  $0 < \alpha < 1$ ,  $\pi \in \Pi_\Omega$  and let  $Y$  be a future observation from  $F(\cdot; \theta)$ . Let  $y(\pi, \alpha)$  denote the upper  $\alpha$ -quantile of the predictive distribution of  $Y$ , satisfying

$$\text{pr}\{Y > y(\pi, \alpha) | X_n\} = \alpha. \quad (16)$$

We would like to know when is it also true that, to a given degree of approximation,

$$\text{pr}\{Y > y(\pi, \alpha) | \theta\} = \alpha \quad (17)$$

very weakly.

It turns out that (17) holds to  $O(n^{-1})$  for every  $\pi \in \Pi_\Omega$  and  $0 < \alpha < 1$ . This is because

$$\text{pr}\{Y > y(\pi, \alpha)|\theta\} = \text{pr}\{Y > y(\pi, \alpha)|\hat{\theta}_n\} + O(n^{-1})$$

and also

$$\begin{aligned} \text{pr}\{Y > y(\pi, \alpha)|X_n\} &= E[\text{pr}\{Y > y(\pi, \alpha)|\Theta\}|X_n] \\ &= \text{pr}\{Y > y(\pi, \alpha)|\hat{\theta}_n\} + O(n^{-1}). \end{aligned}$$

Thus, to second order, there is no coverage probability bias associated with Bayesian predictive probability statements. Note that this is one order higher than the corresponding property for parametric statements.

It is now natural to ask whether or not there exists a prior distribution for which (17) holds to a higher asymptotic order. Write  $D_j \equiv \partial/\partial\theta_j$ . Datta et al. (2000) show that, employing the summation convention,

$$\text{pr}\{Y > y(\pi, \alpha)|\theta\} = \alpha - \frac{1}{\pi(\theta)} D_s \{I_n^{st}(\theta)\mu_t(\theta, \alpha)\pi(\theta)\} + o(n^{-1}) \quad (18)$$

very weakly, where

$$\mu_t(\theta, \alpha) = \int_{q(\theta, \alpha)}^{\infty} D_t f(u; \theta) du$$

and  $q(\theta, \alpha)$  satisfies

$$\int_{q(\theta, \alpha)}^{\infty} f(u; \theta) du = \alpha.$$

It follows that (17) holds to  $o(n^{-1})$  if and only if  $\pi$  satisfies the partial differential equation

$$D_s \{i^{st}(\theta)\mu_t(\theta, \alpha)\pi(\theta)\} = 0. \quad (19)$$

In general solutions to (19) will depend on the level  $\alpha$ , in which case it is not possible to improve coverage probability bias beyond  $O(n^{-1})$ . On the other hand, in the case  $d = 1$  it is shown in Datta et al. (2000) that, if there does exist a prior satisfying (19) for all  $\alpha$ , then this prior must be Jeffreys' prior. In that case, it can further be shown that the coverage probability bias is actually  $O(n^{-2})$  under suitable regularity conditions. Examples include all location models. Recent work by the author indicates that Jeffreys' prior should produce good predictive

coverage properties even when there is no prior for which (19) is satisfied for all  $\alpha$ .

The solution to (19) in the multiparameter case is investigated by Datta et al. (2000). In particular, they show that, when there does exist a prior satisfying (19) which is free from  $\alpha$ , it is not necessarily Jeffreys' prior. Consideration of particular models indicates that the prior which does emerge has other attractive properties. For example, in location-scale models the predictive approach yields the commonly-used improper prior which is proportional to the inverse of the scale parameter, as opposed to the problematic Jeffreys' prior in this case. Overall, consideration of predictive coverage promises to be a valuable tool for the development of sensible objective priors.

#### 4.2. Two-sided predictive intervals

In this section we discuss the construction of objective predictive regions on the basis of small coverage probability bias. For  $0 < \alpha < 1$  and  $\pi \in \Pi_\Omega$ , let  $y_1(\pi, \alpha), y_2(\pi, \alpha)$  satisfy

$$\text{pr}\{y_1(\pi, \alpha) < Y < y_2(\pi, \alpha) | X_n\} = \alpha. \quad (20)$$

We would like to know when it is also true that, to a given level of approximation,

$$\text{pr}\{y_1(\pi, \alpha) < Y < y_2(\pi, \alpha) | \theta\} = \alpha \quad (21)$$

very weakly.

We trivially deduce from §2 that (21) holds to  $O(n^{-1})$  under any smooth prior, and possibly to  $o(n^{-1})$  under Jeffreys' prior in the case  $d = 1$ . Now let  $\pi_0$  be an arbitrary prior in  $\Pi_\Omega$ . Let  $\gamma_\alpha(\theta)$  be any function of  $\alpha$  and  $\theta$  satisfying  $0 < \alpha < \gamma_\alpha(\theta) < 1$  and write  $\gamma_\alpha = \gamma_\alpha(\hat{\theta}_n)$ . Then, from (16), (20) is satisfied if we take  $y_1(\pi_0, \alpha) = y(\pi_0, 1 - \gamma_\alpha + \alpha)$  and  $y_2(\pi_0, \alpha) = y(\pi_0, 1 - \gamma_\alpha)$ . It is shown in Datta et al. (2000) in the case  $d = 1$  that, under mild regularity conditions on  $F(\cdot; \theta)$ , the  $O(n^{-1})$  error term in (21) is then

$$-\frac{1}{\pi_0(\theta)} \frac{d}{d\theta} \left\{ \frac{\xi_\alpha(\theta) \pi_0(\theta)}{I_n(\theta)} \right\},$$

where  $\xi_\alpha(\theta) = \psi(\gamma_\alpha(\theta), \alpha, \theta)$  and

$$\psi(\gamma, \alpha, \theta) = \int_{q(\theta, 1-\gamma+\alpha)}^{q(\theta, 1-\gamma)} f_\theta(x; \theta) dx. \quad (22)$$

Equation (21) therefore holds to  $o(n^{-1})$  when  $\xi_\alpha(\theta) = 0$ . Datta et al. (2000) show that, under mild conditions, there is a unique choice of  $\gamma_\alpha(\theta)$  which achieves this. Furthermore, under additional conditions the error in (21) can be shown to be  $O(n^{-2})$ . Thus restriction to predictive probabilities for this class of intervals can be regarded as objective Bayes in the sense discussed in §3 in the case of Bayes-confidence intervals. The construction of analogous objective predictive regions in the multiparameter case is a topic for future investigation.

## 5. COVERAGE PROBABILITY BIAS AND THE LIKELIHOOD PRINCIPLE

### 5.1. *Preamble*

We have been concerned with Bayesian procedures possessing small coverage probability bias. However, since coverage probabilities involve the sampling rule, such procedures contravene the strong likelihood principle and are therefore incoherent from a Bayesian point of view. In this section we specialise to the case  $d = 1$  and explore the extent to which the procedures described in §§3 and 4 conform to the likelihood principle. Most of the technical details will be relegated to the Appendices.

It appears difficult to obtain useful general results which would apply to quite different models giving rise to the same likelihood. Instead, we focus on the extent to which objective Bayesian procedures are affected by the censoring model or the stopping rule. There are connections here with the work of Barndorff-Nielsen & Cox (1984) on the effect of the sampling rule on Bartlett corrections, and Pierce & Peters (1994) on its effect on general frequentist inference, based on Barndorff-Nielsen's  $r^*$ .

## 5.2. Data-dependent priors

We return to the construction of objective priors via coverage properties of one-sided posterior parametric statements. In general  $O(n^{-1})$  objective Bayesian statements of the form (3) based on Jeffreys' prior will not conform to the likelihood principle to this order. To see this, consider two alternative sampling rules which give rise to the same likelihood but different Fisher informations  $I_1(\theta), I_2(\theta)$ . Then, in order that the posterior distributions of  $\theta$  under the two forms of Jeffreys' prior agree to  $O(n^{-1})$ , we need

$$\frac{I_1(\theta)}{I_2(\theta)} \propto 1 + \eta_n(\theta),$$

where  $\eta_n(\theta)$  is  $O(n^{-1})$ . In general, however,  $\eta_n(\theta)$  will be  $O(n^{-1/2})$  and this relation will fail to hold. On the other hand, as we now show, it may be possible to construct a data-dependent prior which gives rise to  $O(n^{-1})$  coverage probability bias under certain alternative sampling mechanisms. Data-dependent priors have been used quite frequently in the literature as approximate priors.

Define the estimator  $\hat{I}_n(\theta)$  of  $I_n(\theta)$  by

$$\hat{I}_n(\theta) = -l_n''(\theta) + l_n'(\theta) \left\{ \frac{\sum_i l_{(i)}'(\theta) l_{(i)}''(\theta)}{\sum_i (l_{(i)}'(\theta))^2} \right\}, \quad (23)$$

where  $l_{(i)}(\theta)$  is the component of loglikelihood corresponding to the  $i$ th observation. It is straightforward to verify that the construction (23) is invariant under reparameterisation. It is shown in Appendix 2 that

$$\hat{I}_n(\theta) \propto I_n(\theta) \{1 + \eta_n(\theta)\}, \quad (24)$$

where  $\eta_n(\theta)$  is  $O(n^{-1})$ , which turns out to be sufficient to allow us to approximate Jeffreys' prior by the data-dependent prior

$$\hat{\pi}_n(\theta) = \{\hat{I}_n(\theta)\}^{1/2}. \quad (25)$$

Under this prior, the quantity  $T_n$  in (6) becomes

$$\hat{T}_n = -\{\hat{I}_n(\Theta)\}^{-1/2} l_n'(\Theta), \quad (26)$$



so that the posterior distribution of

$$\hat{R}_n^J = R_n + R_n \log(\hat{T}_n/R_n) \quad (27)$$

is standard normal to  $O(n^{-3/2})$ . The following result is proved in Appendix 2.

**THEOREM 2.**  *$\hat{\pi}_n(\theta)$  in (25) is a second-order stably matching data-dependent prior.*

It is important to note that the derivation of Theorem 2 requires independence of the observations. In particular, it can be shown that the result fails in the case of Type II censored data from an exponential distribution. However, the result does hold for Type I censoring. As an example, consider the case of  $n$  observations  $x_1, \dots, x_n$ , where  $x_i = \min(t_i, c_i)$ ,  $t_1, \dots, t_n$  are independent random variables from an exponential ( $\theta$ ) distribution and  $c_1, \dots, c_n$  are censoring times generated from a censoring distribution  $G$ . Then it is straightforward to check that

$$\theta^2 \hat{I}_n(\theta) = r - \frac{(r - s\theta)(r - u\theta)}{(r - 2u\theta + v\theta^2)},$$

where  $r$  is the number of uncensored observations,  $u$  is the sum of the uncensored observations,  $s = \sum_i x_i$  and  $v = \sum_i x_i^2$ . This estimator automatically adapts to the true Fisher information  $I_n(\theta) \propto \theta^{-2}\{1 - \phi(\theta)\}$ , where  $\phi(\cdot)$  is the Laplace transform associated with  $G$ . Note that if the censoring distribution is unknown then it will not be possible to identify  $I_n(\theta)$ .

Use of the data-dependent prior (25) is attractive in that it permits objective Bayesian inference to be performed independently of the underlying data censoring mechanism. However, as can be seen from the above example, (25) uses aspects of the data beyond the minimal sufficient statistic, and so the approach still contravenes the likelihood principle. Since (25) essentially estimates Jeffreys' prior under the true underlying censoring mechanism, such a procedure can only really begin to make sense from a purely Bayesian point of view if the censoring mechanism is regarded as being informative. In particular cases, it may be possible to justify a data-dependent prior as an approximation when the construction

of the experimental design is likely to have been based on prior beliefs about  $\theta$ . Such a possibility warrants further investigation. From a frequentist viewpoint, Theorem 2 indicates that, at least under non-sequential sampling, it is possible to perform conditional inference to  $O(n^{-1})$  independently of sampling mechanisms which give rise to the same likelihood. This observation agrees with the discussion in Pierce (1999), who observes that ‘ideal’ frequentist inference conforms to the second order with the likelihood principle in regard to censoring models.

We note that the objective priors in (11) or (15) do not in general respect the likelihood principle with respect to Type I censoring. This can be easily shown by evaluating the canonical parameter under the censoring model and we omit details. It is not known whether or not one can construct data-dependent priors in this case to achieve (8) to  $o(n^{-1})$  under Type I censoring.

### 5.3. Stopping rules and exponential families

In this section we reconsider the two-sided objective parametric intervals of the form (8). From §2, when these intervals are likelihood regions the class of objective priors for which (8) holds to  $O(n^{-2})$  is given by (9). We now specialise to the case of an exponential family model (10) for which  $s > 0$  and investigate the effect of alternative stopping rules on their associated coverage probability.

Define the class of linear stopping rules for which observation is stopped as soon as

$$c_1 s + c_2 n \geq m, \tag{28}$$

where  $c_1, c_2$  are nonnegative constants with  $c_1 + c_2 = 1$ , and consider asymptotics as  $m \rightarrow \infty$ . In order to gain some insight into the behaviour described later, consider the following heuristic argument. Ignoring any overshoot in the stopping rule (28), we can write

$$s\theta - n\kappa(\theta) = z\phi - m\tilde{\kappa}(\phi),$$

where  $z = s - n$ ,  $\phi = c_2\theta + c_1\kappa(\theta)$  and  $\tilde{\kappa}(\phi) = \kappa(\theta) - \theta$ . Thus, if we assume  $\phi$  is a one-to-one transformation of  $\theta$ , any stopping rule of the form (28) gives rise to

another exponential family model with canonical parameter  $\phi$ . Further assuming that the transformed family is sufficiently regular for (9) to hold, it follows from (11) that the class of priors for which (8) holds to  $O(m^{-2})$  is

$$\pi(\phi) = \tilde{k}_1 + \tilde{k}_2 \tilde{\kappa}'(\phi),$$

for arbitrary constants  $\tilde{k}_1$  and  $\tilde{k}_2$ . In terms of  $\theta$ , these priors become

$$\begin{aligned} \pi(\theta) &= [\tilde{k}_1 + \tilde{k}_2 \{\kappa'(\theta) - 1\} \{\phi'(\theta)\}^{-1}] \phi'(\theta) \\ &= \tilde{k}_1 \{c_2 + c_1 \kappa'(\theta)\} + \tilde{k}_2 \{\kappa'(\theta) - 1\} \\ &= k_1 + k_2 \kappa'(\theta), \end{aligned}$$

which gives precisely the class of objective priors in (11) under a fixed sample size.

Thus we have shown informally that the class of objective priors based on small coverage probability bias of likelihood regions conforms to the likelihood principle with respect to all linear stopping rules of the form (28). A formal derivation proceeds by considering all stopping rules of the form

$$nv(\hat{\theta}) \geq m. \tag{29}$$

Say that  $v(\theta)$  is asymptotically linear in  $\phi(\theta)$  if  $v$  can be written in the form  $v(\theta) = c_1 \phi + c_2 + o(1)$ . The following result is proved in Appendix 3.

**THEOREM 3.** *Consider stopping rules of the form (29) where  $v$  is twice differentiable. Then the coverage probability bias of likelihood regions under the class (11) of priors is  $o(m^{-1})$  if and only if  $v$  is asymptotically linear in the expectation parameter  $\mu$ ; that is, the stopping rule is of the form ‘stop when  $c_1 s + c_2 n + o(n) \geq m$ ’, where  $o(n)$  is arbitrary. Furthermore, the coverage probability bias under stopping rules of the form (28) under any prior in the class (11) is  $O(m^{-2})$ .*

Theorem 3 is related to the concluding remark in Barndorff-Nielsen & Cox (1984) that, although the Bartlett adjustment factor will in general depend on the sampling rule, ‘interesting families of sampling rules related in a simple way

may lead to the same value of the adjustment'. As emphasised by Pierce & Peters (1994), such results arise because of offsetting directional errors, and there is no corresponding result for one-sided probability statements.

We return to the discussion in §3.2 concerning the construction of objective priors on the basis of coverage properties of likelihood regions. From the point of view of coverage probability bias robustness under alternative stopping rules, from Theorem 3 and the preceding informal argument, likelihood regions have the distinctive property of conforming approximately to the likelihood principle under the class (28) of linear stopping rules. However, Theorem 4 below indicates that there is really nothing special about likelihood regions, since every perturbed likelihood region of the form (14) will conform approximately to the likelihood principle under stopping rules for which  $v$  is linear in some other parameterisation. Put another way, this implies that, for any given prior, the perturbed region based on (14) will have small coverage probability bias under the class of stopping rules defined in Theorem 4. The proof is given in Appendix 3.

**THEOREM 4.** *Consider stopping rules of the form (29) where  $v$  is twice differentiable. Then the coverage probability bias of perturbed likelihood regions based on (14) under the class (15) of priors is  $o(m^{-1})$  if and only if  $v$  is asymptotically linear in  $\phi \equiv \int \pi_0(\theta)^{-1} i(\theta) d\theta$ . Furthermore, the coverage probability bias under stopping rules for which  $v$  is linear in  $\phi$  under any prior in the class (15) is  $O(m^{-2})$ .*

It is of interest to note some particular cases. When  $\pi_0(\theta) \propto 1$ , we are back to likelihood regions and the class of matching priors is unchanged under stopping rules for which  $v$  is linear in the expectation parameter. When  $\pi_0$  is Jeffreys' prior, then the corresponding stopping rules have  $v$  linear in the constant-information parameterisation. Finally, stopping rules with  $v$  linear in the canonical parameter arise on taking perturbed regions corresponding to  $\pi_0(\theta) \propto i(\theta)$ .

It would be of interest to investigate whether or not there exist stopping rules for which the coverage property (8) holds to  $O(m^{-2})$  for non-exponential families. From the results for exponential families, one would expect that approximately

conformity to the likelihood principle under approximately linear stopping rules, provided that the statistical curvature of the model is not large.

#### 5.4. Predictive intervals

In this section we investigate the effect of alternative sampling rules on objective Bayesian inference based on predictive coverage probability bias in the case  $d = 1$ . In order to do this we need to specify the predictive distribution  $F(\cdot; \theta)$  of the future observation  $Y$ .

It follows from (18) that objective statements of the form (8) will not conform to the likelihood principle to  $o(n^{-1})$  because of the dependence of the functional form of Fisher's information on the sampling rule. In particular, consider the case where a predictive matching prior exists which is free from  $\alpha$ . Suppose initially that  $X_n$  consists of independent observations with common distribution  $F(\cdot; \theta)$ , the same as the specified predictive distribution of  $Y$ , and suppose further that there exists a predictive matching prior satisfying (19) for all  $\alpha$ . Then, from §4.1, this prior must be Jeffreys' prior and it follows from (19) that

$$D_1[\mu_1(\theta, \alpha)\{i(\theta)\}^{-1/2}] = 0 \quad (30)$$

for all  $\alpha$ . Now consider an alternative sampling rule giving rise to the same likelihood under which (18) holds with  $I_n = \bar{I}_n$ , Fisher's information associated with this sampling rule. Then, from (18),  $\bar{\pi}$  will be a predictive matching prior with respect to this sampling rule if it satisfies the equation

$$D_1[\{\bar{I}_n(\theta)\}^{-1}\mu_1(\theta, \alpha)\bar{\pi}(\theta)] = 0$$

for all  $\alpha$ . It now follows from (30) that

$$\bar{\pi}(\theta) \propto \{i(\theta)\}^{-1/2}\bar{I}_n(\theta).$$

Therefore a matching prior exists for every sampling rule for which (18) is valid and depends on both the predictive distribution  $F(\cdot; \theta)$  and the sampling rule. Furthermore, this prior only coincides with Jeffreys' prior when  $\bar{I}_n(\theta) \propto i(\theta)$ .

Finally, if it is possible to estimate  $\bar{I}_n(\theta)$  with accuracy given in (24), then a matching data-dependent prior may be used, as discussed in §5.2.

Consider next the objective predictive intervals described in §4.2. Again, we suppose that the predictive distribution  $F(\cdot; \theta)$  of  $Y$  is specified and that (18) holds for the sampling mechanism used to generate  $X_n$ . The key point here is that the construction of the function  $\gamma_\alpha(\theta)$  in §4.2 only depends on  $F$  and does not depend at all on the sampling rule. Furthermore, since  $\gamma_\alpha = \gamma_\alpha(\hat{\theta}_n)$ , the quantities  $y_1(\pi, \alpha)$  and  $y_2(\pi, \alpha)$  satisfying (20) and (21) are functions of likelihood-based quantities. Thus, remarkably, the objective predictive intervals constructed in Datta et al. (2000) conform to the likelihood principle to the order of approximation considered with respect to all sampling rules under which (18) is valid.

## 6. DISCUSSION

In this paper we have considered objective Bayesian procedures from the point of view of coverage probability bias. Objective Bayes has been interpreted in a broad sense, encompassing objective choice of priors, objective posterior probability statements and objective predictive probability statements. Since these methods involve sample space averaging, it is natural to enquire into their sensitivity to the sampling rule. The following issue arises in relation to the construction of objective priors. In the terminology of Dawid (1991), should we agree to use a default prior based on a specified ‘inferential model’, or should it be based on the appropriate ‘production model’? In the first case we might get poor coverage properties when the production model does not coincide with the inferential model. Furthermore, there will probably be some ambiguity in deciding on the appropriate inferential model. On the other hand, in the second case we will be violating the likelihood principle. Here we have taken the production model viewpoint and explored the extent to which the likelihood principle is violated. Data-dependent priors would seem to be especially useful when the production model is unknown.

Pierce & Peters (1994) conjecture that the effect due to the elimination of nuisance parameters in Barndorff-Nielsen's  $r^*$  should conform approximately to the likelihood principle. This suggests that conformity of objective procedures to the likelihood principle for a single parameter in the presence of nuisance parameters should be similar to the one-parameter case. Barndorff-Nielsen & Cox (1984) show that for the special case of an exponential failure time model with an exponential censoring mechanism, and treating the censoring rate parameter as a nuisance parameter, the sampling distribution of the likelihood ratio for the failure time parameter is unaffected by the censoring model to  $O(n^{-2})$ .

The difficulties associated with prior construction in Bayesian inference become more severe when the dimension of the parameter space is large. As model complexity increases it becomes more difficult to make sensible prior assignments, whether in a subjective or objective manner. At the same time, the effect of the prior specification on the final inference of interest will tend to become more pronounced. In the case of subjective assignment, the meaning of parameters will often be less clear and there will be a limit to the amount of elicitation that can be carried out in a finite time. On the other hand, naive objective assignments may lead to unappealing properties of the posterior, such as non propriety or paradoxical behaviour. Sensible ways of constructing and implementing priors based on coverage probability bias in the multiparameter case are therefore of real practical importance.

In the absence of nuisance parameters, the coverage requirement for posterior quantiles leads to Jeffreys' prior. However, in the multiparameter case different objective priors may emerge according to which parameter is viewed as the parameter of interest. Furthermore, in the multiparameter case there is a large class of priors giving rise to  $O(n^{-1})$  coverage probability bias. Mukerjee & Ghosh (1997) show that it may be possible to achieve  $o(n^{-1})$  coverage probability bias. The priors that emerge, however, differ depending on whether posterior quantiles or the posterior distribution function are considered.

In view of the ambiguities associated with the parametric approach, a pre-

dictive approach to the derivation of objective Bayes procedures appears to be particularly promising in the multiparameter case, especially when there is no reason to treat any parameter as the parameter of interest in preference to the others. Experience so far indicates that such an approach can yield sensible multiparameter priors. Furthermore, predictive coverage is less influenced by the sampling rule since, first, it is one asymptotic order higher than parametric coverage for one-sided intervals and, second, two-sided objective predictive intervals can be constructed independently of the sampling rule.

In addition to providing a sound basis for the construction of priors in general multiparameter problems, there is also some prospect of computer implementation of predictive matching priors via local solutions to equation (19). In some cases it may be possible to incorporate local predictive matching priors within Gibbs sampling schemes. It turns out that the derivation of (19) also lends itself to the construction of mixed subjective/objective priors. This would be an attractive option where there is real prior knowledge about some parametric functions, but otherwise considerably uncertainly. A related idea would be to achieve some mixed parametric/predictive matching. For example, the class of Tibshirani matching priors associated with a specified scalar parametric function could be obtained, and then a prior identified which achieved some optimal predictive matching within this class of priors. Finally, it would be of interest to explore the construction of predictive matching priors in various non-regular cases. The derivation of matching priors based on parametric coverage for a class of non-regular cases has recently been considered by Ghosal (1999). There is therefore a need for further investigation into the properties of priors constructed via predictive coverage and their practical implementation in the multiparameter case.

#### ACKNOWLEDGEMENT

I am very grateful to the editor, the associate editor and a referee for their many helpful comments and suggestions for improving the clarity of this paper.



## APPENDIX 1

### *Proof of Theorem 1*

Consider an arbitrary perturbed region of the form (12). The techniques here are similar to those in Sweeting (1999) and we only sketch the proof. Let  $s_B^\pi$  be the Bayesian Bartlett correction associated with (12) under the prior  $\pi$ , so that

$$\text{pr}\{|\bar{R}_n| \leq s_B^\pi z_{\alpha/2} | X_n\} = \alpha + O(n^{-2}).$$

We wish to determine the class of priors  $\pi$  for which

$$\text{pr}\{|\bar{R}_n| \leq s_B^\pi z_{\alpha/2} | \theta\} = \alpha + O(n^{-2}) \quad (\text{A1})$$

very weakly.

Let  $\psi$  be the constant-information parameterisation for which  $I_n(\psi) = n$ . From (14) in Sweeting (1999), the sampling density of  $\bar{R}_n/s_B^\pi$  has the very weak approximation

$$\bar{f}_n(r|\psi) \propto \phi(r) \bar{h}_n(r, \psi_0) (\partial\psi_0/\partial\psi) \{1 - \bar{b}_n(\psi_0)r^2\} \{1 + \bar{e}_n(\psi_0)r\} \quad (\text{A2})$$

to  $O(n^{-2})$ , where all terms are as defined in Sweeting (1999). Then (A1) is true if and only if the second coefficient of the modulating function in expression (A2) is zero. This coefficient is given by (15) in Sweeting (1999). Write  $\bar{q}_1(\psi) = \bar{a}_1(\psi) + \bar{p}_1(\psi)$  where  $\bar{p}_1(\psi)$  is the first coefficient in  $\bar{g}(r)$ . That is,  $\bar{p}_1(\psi)$  satisfies

$$p_1 = \bar{p}_1(\psi_0) + n^{-1}\xi + O(n^{-3/2})$$

under  $\psi_0$ , where  $p_1$  is the first coefficient of  $g$  and  $\xi$  is a random variable with  $E_0(\xi) = 0$ . Also, define  $\beta(\psi) = 3n^{1/2}\bar{a}_1(\psi) = \frac{1}{2}n^{1/2}\bar{\rho}_{111}(\psi)$  and  $\delta(\psi) = 2n^{1/2}\bar{p}_1(\psi)$ . Substituting for  $\bar{\gamma}_1, \bar{\gamma}_2, \bar{\alpha}_1, \bar{\alpha}_2, \beta$  and  $\delta$  in (15) of Sweeting (1999), after some algebra we obtain, in a similar fashion to §5.4 of Sweeting (1995a), the differential equation

$$\pi''(\psi) + [\{\beta(\psi) + \delta(\psi)\}\pi(\psi)]' = 0,$$

giving

$$\pi'(\psi) + \{\beta(\psi) + \delta(\psi)\}\pi(\psi) = C.$$

Finally, if we note the invariance of the quantities  $\bar{\rho}_{111}$  and  $\bar{p}_1$ , the solution to this equation expressed in the canonical parameterisation is found to be

$$\pi(\theta) = e^{-\omega_n(\theta)} \{k_1 + k_2 \int I_n(\theta) e^{\omega_n(\theta)} d\theta\},$$

where  $\omega_n(\theta) = 2 \int I_n(\theta)^{1/2} \bar{p}_1(\theta) d\theta$ .

Now let  $\pi_0$  be an arbitrary prior in  $\Pi_\Omega$  and take  $g(r) = \{\pi_0(\theta)/\pi_0(\hat{\theta}_n)\}^{-1/2}$ .

Then

$$I_n(\theta)^{1/2} \bar{p}_1(\theta) = -\frac{1}{2} \{\pi_0'(\theta_0)/\pi_0(\theta_0)\}$$

so that  $\exp\{-\omega_n(\theta)\} = \pi_0(\theta)$  and (15) follows.

## APPENDIX 2

### *Proof of Theorem 2*

Assume without loss of generality that  $I_n(\theta) = n$ . Applying one-term Taylor expansions about  $\theta_0$  and the law of large numbers, we see that

$$\begin{aligned} \hat{I}'_n(\hat{\theta}_n) &= -l_n'''(\hat{\theta}_n) + l_n''(\hat{\theta}_n) \left\{ \frac{\sum_i l'_{(i)}(\hat{\theta}_n) l''_{(i)}(\hat{\theta}_n)}{\sum_i (l'_{(i)}(\hat{\theta}_n))^2} \right\} \\ &= -l_n'''(\theta_0) + l_n''(\theta_0) \left\{ \frac{\sum_i l'_{(i)}(\theta_0) l''_{(i)}(\theta_0)}{\sum_i (l'_{(i)}(\theta_0))^2} \right\} + O(n^{1/2}) \\ &= -E_0\{l_n'''(\theta_0)\} + E_0\{l_n''(\theta_0)\} \left\{ \frac{\sum_i E_0\{l'_{(i)}(\theta_0) l''_{(i)}(\theta_0)\}}{\sum_i E_0\{(l'_{(i)}(\theta_0))^2\}} \right\} + O(n^{1/2}) \\ &= -\rho_3(\theta_0) - \rho_{12}(\theta_0) + O(n^{1/2}) \\ &= I'_n(\theta_0) + O(n^{1/2}) = O(n^{1/2}) \end{aligned}$$

since  $I'_n(\theta_0) = 0$ . A Taylor expansion of  $\hat{I}_n(\theta)$  about  $\hat{\theta}_n$  now gives

$$J_n^{-1} \hat{I}_n(\theta) = 1 + J_n^{-3/2} \hat{I}'_n(\hat{\theta}_n) r + O(n^{-1}) = 1 + O(n^{-1}). \quad (A3)$$

Relation (24) now follows from (A3) by converting to an arbitrary parameterisation.

We now trace through the arguments in §5.3 of Sweeting (1995b), with the following modifications. All the following statements hold to  $O(n^{-1})$ . First, the

posterior distribution of  $\hat{R}_n^J$  given by (27) under any prior  $\tau \in \Pi_\Omega$  is given by

$$\begin{aligned} \hat{f}_n^J(r|X_n) &\propto \phi(r)\{\hat{I}_n(\theta)/J_n\}^{-1/2}\{\tau(\theta)/\tau(\hat{\theta}_n)\} \\ &= \phi(r)\{\tau(\theta)/\tau(\hat{\theta}_n)\} \end{aligned} \tag{A4}$$

from (A3). Now let  $A_n$  be any second-order locally ancillary statistic. The argument in Sweeting (1995b) then gives

$$\hat{f}_n^J(r|A_n, \theta) = \phi(r)$$

very weakly. On comparison with (A4) we see that the conditional sampling distribution of  $\hat{R}_n^J$  is the same as the posterior distribution under the data-dependent prior (25), as required.

### APPENDIX 3

#### *Proofs of Theorems 3 and 4*

Since  $l''(\theta)$  is non-random, we have  $-l''(\theta) = ni(\theta)$  and  $-l'''(\theta) = ni'(\theta)$ . Let  $\alpha(\theta) = \tilde{E}(n)$ , where a  $\tilde{\cdot}$  indicates expectation under the stopping rule (29). We have the relations

$$\begin{aligned} \tilde{I}_n(\theta) &= (\alpha i)(\theta) \\ \tilde{I}'_n(\theta) &= (\alpha i)'(\theta) \\ \tilde{\rho}_3(\theta) &= -\alpha(\theta)i'(\theta), \end{aligned}$$

where  $(\alpha i)(\theta) = \alpha(\theta)i(\theta)$ . The final two relations along with the identity  $3\tilde{I}'_n(\theta) = \tilde{\rho}_{111}(\theta) - 2\tilde{\rho}_3(\theta)$  imply that

$$\tilde{\rho}_{111}(\theta) = 2\alpha'(\theta)i(\theta) + (\alpha i)'(\theta).$$

Since  $\tilde{E}(\hat{\theta}_n) = \theta + O(m^{-1})$ , a second-order Taylor expansion of  $v(\hat{\theta}_n)^{-1}$  about  $\theta$  gives

$$\alpha(\theta) = m/v(\theta) + O(1). \tag{A5}$$

Now transform to the constant-information parameterisation  $\psi$  under the stopping rule (29) for which  $\tilde{I}_n(\psi) = m$ . Then

$$\theta'(\psi) = \left\{ \frac{m}{(\alpha i)(\theta)} \right\}^{1/2}.$$

*Proof of Theorem 3.* Tracing through the development in Appendix 1 with  $g \equiv 1$ , we see that (8) is true under the stopping rule (29) to  $o(m^{-1})$  if and only if

$$\pi''(\psi) + \{\tilde{\beta}(\psi)\pi(\psi)\}' = o(1), \quad (\text{A6})$$

where  $\tilde{\beta}(\psi) = \frac{1}{2}m^{1/2}\tilde{\rho}_{111}(\psi)$ .

Now revert to the canonical parameterisation, noting that

$$\begin{aligned} \pi(\psi) &= \pi(\theta)\theta'(\psi) \\ \theta''(\psi) &= -\frac{1}{2}m^{1/2}(\alpha i)(\theta)^{-3/2}(\alpha i)'(\theta)\theta'(\psi) \\ \tilde{\beta}(\psi) &= \frac{1}{2}m^{1/2}\tilde{\rho}_{111}(\theta) = \frac{1}{2}m^{1/2}(\alpha i)(\theta)^{-3/2}\{2\alpha'(\theta)i(\theta) + (\alpha i)'(\theta)\}, \end{aligned} \quad (\text{A7})$$

to give

$$\begin{aligned} \pi'(\psi) + \tilde{\beta}(\psi)\pi(\psi) &= m \left\{ \frac{\pi'(\theta)}{(\alpha i)(\theta)} + \frac{\alpha'(\theta)\pi(\theta)}{\alpha(\theta)^2 i(\theta)} \right\} \\ &= \frac{\eta(\theta)\pi'(\theta)}{i(\theta)} - \frac{\eta'(\theta)\pi(\theta)}{i(\theta)}, \end{aligned} \quad (\text{A8})$$

where  $\eta(\theta) = m/\alpha(\theta)$ . Now, noting that the class (11) of matching priors under fixed sample size satisfies the equation

$$\{\pi'(\theta)/i(\theta)\}' = 0,$$

differentiate (A8) to obtain, from (A6),

$$\eta(\theta) \left\{ \frac{\pi'(\theta)}{i(\theta)} \right\}' - \left\{ \frac{\eta'(\theta)}{i(\theta)} \right\}' \pi(\theta) = o(1). \quad (\text{A9})$$

It follows that the class of matching priors under (29) is the same as that under fixed sample size sampling if and only if  $\{\eta'(\theta)/i(\theta)\}' = o(1)$ , which is equivalent to

$$\eta(\theta) = c_1\mu + c_2 + o(1).$$

From (A5) this is equivalent to the class of stopping rules which are asymptotically linear in  $\mu$ , as asserted.

Finally, for the class of stopping rules (28) we see that equation (A9) holds with  $o(1)$  replaced by  $O(m^{-1})$ . This implies that (8) holds to  $O(m^{-2})$ , as required.

*Proof of Theorem 4.* As in the proof of Theorem 3, tracing through the development in Appendix 2, we see that (8) is true under the stopping rule (29) to  $o(m^{-1})$  if and only if

$$\pi''(\psi) + [\{\tilde{\beta}(\psi) + \delta(\psi)\}\pi(\psi)]' = o(1). \quad (\text{A10})$$

Now revert to the canonical parameterisation, noting the relations (A7) and  $\bar{p}_1(\psi) = \bar{p}_1(\theta)$ , to give

$$\begin{aligned} \pi'(\psi) + \{\tilde{\beta}(\psi) + \tilde{\delta}(\psi)\}\pi(\psi) &= m \left[ \left\{ \frac{\pi_0(\theta)}{(\alpha i)(\theta)} \right\} \left\{ \frac{\pi(\theta)}{\pi_0(\theta)} \right\}' + \frac{\alpha'(\theta)\pi(\theta)}{\alpha(\theta)^2 i(\theta)} \right] \\ &= \left\{ \frac{\eta(\theta)\pi_0(\theta)}{i(\theta)} \right\} \left\{ \frac{\pi(\theta)}{\pi_0(\theta)} \right\}' - \frac{\eta'(\theta)\pi(\theta)}{i(\theta)}, \quad (\text{A11}) \end{aligned}$$

where  $\eta(\theta) = m/\alpha(\theta)$ . Now, noting that the class (15) of matching priors under fixed sample size satisfies the equation

$$\left[ \left\{ \frac{\pi_0(\theta)}{i(\theta)} \right\} \left\{ \frac{\pi(\theta)}{\pi_0(\theta)} \right\}' \right]' = 0,$$

differentiate (A11) to obtain, from (A10),

$$\eta(\theta) \left[ \left\{ \frac{\pi_0(\theta)}{i(\theta)} \right\} \left\{ \frac{\pi(\theta)}{\pi_0(\theta)} \right\}' \right]' - \left\{ \frac{\eta'(\theta)\pi_0(\theta)}{i(\theta)} \right\}' \left\{ \frac{\pi(\theta)}{\pi_0(\theta)} \right\} = o(1). \quad (\text{A12})$$

It follows that the class of matching priors under (29) is the same as that under fixed sample size sampling if and only if  $\{\eta'(\theta)\pi_0(\theta)/i(\theta)\}' = o(1)$ , which is equivalent to

$$\eta(\theta) = c_1\phi + c_2 + o(1),$$

where  $\phi = \int \pi_0(\theta)^{-1}i(\theta)d\theta$ . From (A5) this gives the asserted class of stopping rules.

Finally, for the class of stopping rules (29) which are linear in  $\phi$  we see that equation (A12) holds with  $o(1)$  replaced by  $O(m^{-1})$ . This implies that (8) holds to  $O(m^{-2})$ , as required.

## REFERENCES

- Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardised signed log likelihood ratio. *Biometrika* **73**, 307-22.
- Barndorff-Nielsen, O. E. & Chamberlin, S. R. (1991). An ancillary invariant modification of the signed log likelihood ratio. *Scand. J. Statist.* **18**, 341-52.
- Barndorff-Nielsen, O. E. & Cox, D. R. (1984). The effect of sampling rules on likelihood statistics. *Int. Statist. Rev.* **52**, 309-26.
- Barndorff-Nielsen, O. E. & Cox, D. R. (1994). *Inference and Asymptotics*. London: Chapman and Hall.
- Bernardo, J. M. & Ramón, J. M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* **47**, 101-35.
- Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *J. R. Statist. Soc. B* **49**, 1-39.
- Datta, G. S., Mukerjee, R., Ghosh, M. & Sweeting, T. J. (2000). Bayesian prediction with approximate frequentist validity. *Ann. Statist.* **28**, 1414-26.
- Dawid, A. P. (1991). Fisherian inference in likelihood and prequential frames of reference (with Discussion). *J. R. Statist. Soc. B* **53**, 79-109.
- DiCiccio, T. J. & Martin, A. M. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc. B* **55**, 305-16.
- Ghosh, J. K. & Mukerjee, R. (1992a). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. R. Statist. Soc. B* **54**, 867-75.

- Ghosh, J. K. & Mukerjee, R. (1992b). Non-informative priors (with Discussion). In *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 195-210. Oxford University Press.
- Ghosal, S. (1999). Probability matching priors for non-regular cases. *Biometrika* **86**, 956-64.
- Hartigan, J. A. (1966). Note on the confidence-prior of Welch and Peers. *J. R. Statist. Soc. B* **28**, 55-6.
- Kass & Wasserman (1996). The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.* **91**, 1343-70.
- Mukerjee, R. & Ghosh, J. K. (1997). Second-order probability matching priors. *Biometrika* **84**, 970-5.
- Nicolaou, A. (1993). Bayesian intervals with good frequency behaviour in the presence of nuisance parameters. *J. R. Statist. Soc. B* **55**, 377-90.
- Pierce, D. A. (1999). On the relation between frequency inference and likelihood. *Bull. Int. Statist. Inst.*, 52nd Session of the ISI.
- Pierce, D. A. & Peters, D. (1994). Higher-order asymptotics and the likelihood principle: one parameter models. *Biometrika* **81**, 1-10.
- Severini, T. A. (1993). Bayesian interval estimates which are also confidence intervals. *J. R. Statist. Soc. B* **55**, 533-40.
- Stein, C. M. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach Center Publications, **16**, Ed. Zieliński, pp. 485-514. Warsaw: PWN-Polish Scientific Publishers.
- Sweeting, T. J. (1995a). A framework for Bayesian and likelihood approximations in statistics. *Biometrika* **82**, 1-23.

- Sweeting, T. J. (1995b). A Bayesian approach to approximate conditional inference. *Biometrika* **82**, 25-36.
- Sweeting, T. J. (1996). Approximate Bayesian computation based on signed roots of log-density ratios (with Discussion). In *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 427-44. Oxford University Press.
- Sweeting, T. J. (1999). On the construction of Bayes-confidence regions. *J. R. Statist. Soc. B* **61**, 849-61.
- Tibshirani, R.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* **35**, 318-29.
- Woodroffe, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14**, 1049-67.