Consistency of Bayesian Estimates for the Sum of Squared Normal Means with a Normal Prior

by

Michael Evans
Department of Statistics
University of Toronto

and

Mohammed Shakhatreh
Department of Statistics
University of Toronto

Technical Report No. 0706 February 21, 2007

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Consistency of Bayesian Estimates for the Sum of Squared Normal Means with a Normal Prior

Michael Evans and Mohammed Shakhatreh
Department of Statistics
University of Toronto

Abstract: We consider the problem of estimating the sum of squared means when the data are independent values from normal distributions with known variance and unknown, possibly different, means. This example has posed difficulties for many approaches to inference. We consider a Bayesian formulation with a diffuse proper prior and examine the consistency properties of several estimates derived from Bayesian considerations. We prove that a particular Bayesian estimate (LRSE) is consistent in a much wider set of circumstances than other Bayesian estimates like the posterior mean and mode. We show that the LRSE is either equal to the positive part of the UMVUE or differs from it with a relative error no greater than $2/n$. We also prove a consistency result for interval estimation and discuss checking for prior-data conflict. We conclude that a general inference principle applied to this example leads to reasonable inferences and that there is not a problem with the Bayesian formulation or prior chosen.

Key words and phrases: consistency, Bayesian estimates, least relative surprise estimate, interval estimates, prior-data conflict.

1 Introduction

Let $x^n = (x_1, \ldots, x_n)$ be n independent random variables where $x_i \sim N(\theta_i, 1)$. Suppose that our interest is in making inferences about $\tau_n^2 = \sum_{i=1}^n \theta_i^2$. This problem is well-known to lead to difficulties for various approaches to deriving inferences. The behavior of inference rules in contexts like this provide insights into their relative strengths and weaknesses, e.g., see Stein (1959) for perhaps the first such use of this example.

For example, in frequentist contexts the plug-in MLE is given by $||x^n||^2 = \sum_{i=1}^n x_i^2$ and this has expectation equal to $\tau_n^2 + n$ and so is seriously biased for large n. By contrast, the UMVUE is $||x^n||^2 - n$, but this may take negative values and so it seems more appropriate to use the biased estimator $(||x^n||^2 - n)_+$. An easy argument shows that $(||x^n||^2 - n)_+$ has smaller MSE than the UMVUE. Chow (1987) has proven that this estimator is not admissible with respect to mean squared error. Perlman and Rasmussen(1975), Neff and Stawderman

(1976), and Kubokawa, Robert and Saleh (1993) have considered various classes of estimators that have smaller MSE than the UMVUE.

If we just consider the observed data to be $||x^n||^2$, i.e., ignore the fact that we observe the individual components of x^n, then we can derive an MLE of τ^2 based on the fact that $||x^n||^2 \sim$ Chi-squared(n, τ_n^2), where Chi-squared(n, δ) denotes the Chi-squared distribution with n degrees of freedom and noncentrality δ. Strictly speaking, however, this estimator is not an MLE for the problem we are considering. Some further principle beyond likelihood is then required to justify this estimator. Saxena and Alam (1982), have shown that $(||x^n||^2 - n)_+$ has smaller MSE than this "MLE" estimator when using squared error loss.

In this paper we examine inferences for τ_n^2 in Bayesian contexts. For the prior we suppose that $\theta_1, \theta_2 \ldots$ are i.i.d. with a $N(0, \sigma^2)$ distribution. Generally we will think of σ^2 as large and known so that our choice of prior represents diffuse beliefs about the θ_i. Our purpose here is to consider the consistency behavior of various Bayesian estimators derived according to some principle applied to the specified sampling model and prior. In particular, we do not allow the prior to be changed.

The prior distribution of τ_n^2/σ^2 is Chi-squared$(n, 0)$. From this we can deduce that the posterior distribution of $(1 + 1/\sigma^2)\tau_n^2$ is Chi-squared$(n, (1 + 1/\sigma^2)^{-1}||x||^2)$. Using squared error loss, the Bayes estimate of τ^2 is given by the mean of the posterior distribution of τ_n^2 and this is $m_n = (1 + 1/\sigma^2)^{-2}||x^n||^2 + (1+1/\sigma^2)^{-1}n$. As with the plug-in MLE, this would appear to be a rather poor estimate when n is large, see, however, the discussion of the consistency of this estimator in section 2. Note that when $\sigma^2 \to \infty$ the Bayes estimate converges to $||x^n||^2 + n$ which is the formal Bayes estimate when we use an improper flat prior on each θ_i, see also de Waal (1974). Another commonly used Bayesian estimator is the mode of the posterior density of τ_n^2 which, in this case, cannot be obtained in closed form. We discuss the consistency of this estimator in section 4 and will show that it is essentially equivalent to the posterior mean.

Various authors have commented on this example and point out the relatively poor performance of Bayesian inferences, particularly when σ^2 is large. As such, this example raises several questions that have broader significance. Is there something intrinsically wrong with putting a diffuse proper prior on a high-dimensional parameter such that we cannot trust our inferences about marginal parameters? There doesn't appear to be anything wrong with saying we have little information about a normal mean θ and using a $N(0, \sigma^2)$ prior distribution with large fixed σ^2 to reflect this. Why should doing this with n independent normal means matter? Are we forced here, as some advocate, to abandon the basic Bayesian paradigm of prescribing a sampling model and prior and deriving our inferences from these ingredients, to require that our choice of prior depend on the parameter of interest? We argue here that the problem in this example is not associated with the prior chosen for the full parameter and that suitable inferences can be obtained. Furthermore, even Bayesian inferences that appear to be poor, such as the posterior mean, can be seen to be behaving appropriately when we take into account other aspects of a Bayesian analysis.

We note first that Bayes theorem only tells us that probability statements about the unknown parameter must be based on the posterior. In essence this is not a theorem when we use it in this way, but a principle or axiom that most believe makes sense. To determine an inference, such as an estimate, we need to specify further ingredients to the basic Bayesian formulation.

One possibility is to add a loss function. This leads to choosing a Bayes rule as the estimate of τ_n^2. The inference is dependent on the choice of loss and, unless we had very specific reason to use some particular loss function, we might look for an approach to deriving inferences that does not require such a choice.

An alternative path arises from specifying a further principle or axiom as follows. Suppose we want to form a set $C_\gamma(x^n)$ that contains the unknown value of τ_n^2 with posterior probability γ. There will be many such sets and Bayes theorem does not tell us how to choose among them. It is clear that we cannot be arbitrary about this else one could choose a set containing any particular value of interest. We need a further principle or axiom to tell us how to choose the set. If we have such a principle and this leads, as it should, to sets that nest as we change γ, then we can take the central value (i.e., the limiting set as $\gamma \to 0$) as our point estimate. Various principles have been used with perhaps the most prominent being the hpd-principle where we select the set $\{\tau^2 : \pi\left(\tau^2 \,|\, x^n\right) > k_\gamma\}$, with $\pi(\cdot \,|\, x^n)$ the posterior density of τ_n^2 and k_γ chosen so that the set has posterior probability content equal to γ. This can also be shown to be equal to the set having smallest length measure among all sets having posterior probability content greater than or equal to γ. The limiting estimator is easily seen to be a value that maximizes $\pi(\cdot \,|\, x^n)$, i.e., the mode of the posterior density of τ^2 with respect to length measure.

More generally we can form hpd-like regions by taking an arbitrary measure Λ on the set of possibilities for τ^2 and find the set that has minimal Λ measure among all sets having posterior probability content equal to γ, see Evans, Guttman and Swartz (2006) for details. In many ways the most natural choice is $\Lambda = \Pi$, where Π is the marginal prior measure for τ_n^2, as this leads to regions that are invariant under arbitrary, smooth transformations of τ_n^2. Such regions were derived, from an alternative point-of-view, in Evans (1997) and referred to there as relative surprise regions. A γ-relative surprise region then takes the form $C_\gamma(x^n) = \{\tau^2 : \pi(\tau^2 \,|\, x^n)/\pi(\tau^2) > k_\gamma\}$ where π is the prior density of τ_n^2 and k_γ is chosen so that this set contains γ of the posterior probability. This leads to the estimate being a value that maximizes $\pi(\tau^2 \,|\, x^n)/\pi(\tau^2)$ and is referred to as a least relative surprise estimate (LRSE). We refer to this approach to determining inferences as the relative surprise principle.

The function $\pi(\tau^2 \,|\, x^n)/\pi(\tau^2)$ is also referred to in the literature as an integrated likelihood as it arises as the expectation of the likelihood under the conditional prior of $(\theta_1, \ldots, \theta_n)$ given τ_n^2. Berger, Liseo and Wolpert (1999) describe various advantages of using integrated likelihoods as opposed to other kinds of likelihoods. In this case, it is clear that we obtain $\pi(\tau^2 \,|\, x^n)/\pi(\tau^2)$ by taking the conditional prior of $(\theta_1, \ldots, \theta_n)$ given τ_n^2 to be the uniform prior on the sphere of radius τ_n. That the integrated likelihood also arises via the relative surprise principle provides further support for this form of the likelihood. Also, the LRSE for τ_n^2 is equal to the "MLE" discussed above. We prefer to refer to the estimator as the LRSE, however, as it reflects the fact that the estimate is derived from a general principle that arises quite naturally in Bayesian contexts.

In this example the LRSE does not depend on the hyperparameter σ^2 and so could be thought of as the estimate when $\sigma^2 \to \infty$ as well. This is not a general characteristic of relative surprise inferences for marginal parameters. In fact, there really is a dependence here, because we would naturally quote $C_\gamma(x^n)$ together with γ, perhaps for several γ, as our quantification of the accuracy of the LRSE and, for a prescribed γ, the interval $C_\gamma(x^n)$ does depend on σ^2 as we show in section 5. In section 3 we establish the consistency of the LRSE and show that is it essentially equivalent to $(||x^n||^2 - n)_+$. Note that, while it is known that the LRSE in this example is not admissible when the loss function is squared error, it possesses what might be considered a much more important property. In particular, the LRSE is invariant under all smooth reparameterizations, i.e., if we have the LRSE in one parameterization, then the LRSE in another parameterization is obtained by simply transforming the LRSE by the relevant transformation. This invariance property could be viewed as a necessary coherency property of a theory of inference. Of course, in a particular application we may introduce other ingredients into a problem, such as a loss function, that render such invariance impossible.

An alternative class of estimates arises by using intervals obtained by discarding different amounts of posterior probability in each tail of the posterior. For example, we could discard $\alpha(1-\gamma)$ in the left tail and $(1-\alpha)(1-\gamma)$ in the right tail. This leads to the limiting estimator being an α-th posterior quantile. One drawback of this approach is that it is unclear how to generalize this to a general principle for obtaining inferences about a higher dimensional parameter, or a non-Euclidean parameter, or even providing a reasonable general justification for this when we have a 1-dimensional parameter and the posterior is multimodal. In section 4 we make some comments about the posterior median, i.e., when $\alpha = 1/2$.

In section 5 we consider interval estimates for τ_n^2/n, establish the consistency of relative surprise intervals under certain conditions and discuss the precision of inferences. In section 6 we consider checking for prior-data conflict and in section 7 draw some conclusions.

2 The Posterior Mean

The parameter τ_n^2 is changing with n, so we say that a sequence of estimators $t_n(x^n)$ is consistent for τ_n^2 if $t(x^n)/n - \tau_n^2/n \xrightarrow{P} 0$. Note that $E(x_i^2) = 1 + \theta_i^2$ and $Var(x_i^2) = 2 + 4\theta_i^2$. Now let $P = \prod_{i=1}^{\infty} P_{\theta_i}$ for some sequence $\theta_1, \theta_2, \ldots$. Then $P\left(\left|||x^n||^2/n - 1 - \tau_n^2/n\right| > \epsilon\right) \leq \epsilon^{-2} E\left(||x^n||^2/n - 1 - \tau_n^2/n\right)^2 = 2/n\epsilon^2 + \left(4/n\epsilon^2\right)\left(\tau_n^2/n\right)$ by Markov's inequality and so convergence of $||x^n||^2/n$ is guaranteed when $\tau_n^2 = o(n^2)$. This is a fairly weak restriction on the sequence $\theta_1, \theta_2, \ldots$, e.g., it is satisfied whenever the sequence is bounded, and seems necessary to talk meaningfully about consistency.

The following proposition summarizes the behavior of several estimators.

Proposition 1. When $\tau_n^2 = o(n^2)$ we have that
(i) the plug-in MLE $||x^n||^2$ is inconsistent,
(ii) the UMVUE $||x^n||^2 - n$ and $(||x^n||^2 - n)_+$ are consistent,
(iii) the posterior mean m_n of τ_n^2 is consistent if and only if $\tau_n^2/n \to \sigma^2$ and
(iv) the limiting posterior mean as $\sigma^2 \to \infty$ is inconsistent.
Proof: By the preceding argument we have that

$$||x^n||^2/n - \tau_n^2/n = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - \theta_i^2\right) \xrightarrow{P} 1 \tag{1}$$

and so the plug-in MLE is inconsistent. It is then immediate that $||x||^2 - n$ is consistent and since $\left|(||x^n||^2 - n)_+/n - \tau_n^2/n\right| \le \left|(||x^n||^2 - n)/n - \tau_n^2/n\right|$ this implies that $(||x^n||^2 - n)_+$ is consistent. Further,

$$\begin{aligned} m_n/n - \tau_n^2/n &= (1 + 1/\sigma^2)^{-2}(||x^n||^2/n) + (1 + 1/\sigma^2)^{-1} - \tau_n^2/n \\ &= (1 + 1/\sigma^2)^{-2}(||x^n||^2/n - 1 - \tau_n^2/n) + \\ &\quad (1 + 1/\sigma^2)^{-1} + (1 + 1/\sigma^2)^{-2} - (1 - (1 + 1/\sigma^2)^{-2})(\tau_n^2/n) \end{aligned}$$

and so, from (1), m_n is consistent if and only if

$$\lim_{n\to\infty} \tau_n^2/n = \frac{(1 + 1/\sigma^2)^{-2} + (1 + 1/\sigma^2)^{-1}}{\{1 - (1 + 1/\sigma^2)^{-2}\}} = \sigma^2.$$

As $\sigma^2 \to \infty$ the limiting posterior mean is $||x||^2/n + 1$ and this is inconsistent. The consistency of the posterior mean when $\lim_{n\to\infty} \tau_n^2/n = \sigma^2$ is in some ways very natural. For the prior we are using forces this convergence on the sequence $\theta_1, \theta_2, \ldots$ by the strong law of large numbers.

Since the consistency of the posterior mean requires that $\tau_n^2/n \to \sigma^2$, and σ^2 is known, we might wonder if this result has any practical relevance. In the following section we show that applying the relative surprise principle to the specified model and prior yields an estimator that is arguably better even when $\tau_n^2/n \to \sigma^2$ and is applicable in a much wider class of situations.

3 The LRSE

The modified Bessel function of order p is defined as

$$I_p(x) = (x/2)^p \sum_{k=0}^{\infty} \left(x^2/4\right)^k / \left(k!\Gamma\left(p + k + 1\right)\right).$$

We need the following properties of I_p.

Lemma 2. We have that for $x \ge 0$
(i) $xI_p(x)/I_{p-1}(x)$ is strictly increasing in x,

(ii) $I_p(x)/xI_{p-1}(x)$ is strictly decreasing in x,
(iii) $I_p(x)/I_{p-1}(x) \geq I_{p+1}(x)/I_p(x)$,
(iv) $I_{p+1}(x)/I_{p-1}(x) \leq I_p^2(x)/I_{p-1}^2(x)$,
(v) $I_p(x) \sim e^x/\sqrt{2\pi\, x}$ as $x \to \infty$,
(vi) $I_{p-1}(x) = (2p/x)I_p(x) + I_{p+1}(x)$.
Proof: See the Appendix.

To obtain the LRSE we need to find τ^2 maximizing the ratio $\pi(\tau^2|\, x)/\pi(\tau^2)$. Note that it is immediate that the LRSE is always nonnegative. The prior density of τ_n^2 is $\pi(\tau^2) \propto (\tau^2)^{(n/2)-1}\exp\{-\tau^2/2\sigma^2\}$, and the posterior density is

$$\pi(\tau^2 \,|\, x^n) \propto (\tau^2)^{(n/2)-1} e^{\left\{-(1+1/\sigma^2)\tau^2/2\right\}} \sum_{k=0}^{\infty} \left(||x^n||^2\tau^2/4\right)^k / \left(k!\Gamma\left((n/2)+k\right)\right).$$

The ratio of the posterior to prior density is then

$$\pi(\tau^2 \,|\, x^n)/\pi(\tau^2) \propto e^{-\tau^2/2} \sum_{k=0}^{\infty} \left(||x^n||^2\tau^2/4\right)^k / \left(k!\Gamma\left((n/2)+k\right)\right)$$

and, since this is a smooth function of $\tau^2 \geq 0$, the maximum must be either at 0 or is a critical point of $g(\tau^2) = \ln \pi(\tau^2|\, x^n)/\pi(\tau^2)$ where

$$\begin{aligned} \frac{dg(\tau^2)}{d\tau^2} &= -\frac{1}{2} + \left(\sum_{k=0}^{\infty} k \frac{\left(||x^n||^2/4\right)^k (\tau^2)^{k-1}}{k!\Gamma\left((n/2)+k\right)}\right) / \sum_{k=0}^{\infty} \frac{\left(||x^n||^2\tau^2/4\right)^k}{k!\Gamma\left((n/2)+k\right)} \\ &= -(1/2) + (1/2)(||x^n||/\tau) I_{n/2}\left(||x^n||\tau\right) I_{(n/2)-1}^{-1}\left(||x^n||\tau\right). \qquad (2) \end{aligned}$$

Setting (2) equal to 0, we have that, when the LRSE is not 0, then it is a solution of $H_n(\tau^2) = (||x^n||/\tau)I_{n/2}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right) = 1$.

We need the following result.

Lemma 3. For $H_n(\tau^2) = (||x^n||/\tau)I_{n/2}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right)$ we have that
(i) $H_n(\tau^2)$ is a strictly decreasing function for $\tau^2 \geq 0$,
(ii) $\lim_{\tau^2 \to 0+} H_n(\tau^2) = ||x^n||^2/n$, and
(iii) $\lim_{\tau^2 \longrightarrow \infty} H_n(\tau^2) = 0$.
Proof: Part (i) follows from Lemma 2(ii). Also we have that $\lim_{\tau^2 \longrightarrow 0^+} H_n(\tau^2) = (1/2)\,||x^n||^2 \left(\Gamma\left(n/2\right)/\Gamma\left(n/2+1\right)\right) = ||x^n||^2/n$ and (ii) is established. From Lemma 2(v), we have that $I_{n/2}\left(||x^n||\tau\right) \sim e^{||x||\,\tau}/(2\pi||x^n||\tau)^{1/2}$ as $\tau^2 \longrightarrow \infty$ and so $\lim_{\tau^2 \longrightarrow \infty} H_n(\tau^2) = \lim_{\tau^2 \longrightarrow \infty}(||x^n||/\tau) = 0$.

We now establish the existence and uniqueness of the LRSE.

Proposition 4. If $||x^n||^2/n \geq 1$ then the LRSE is the unique solution of $H_n(\tau^2) = 1$ and if $||x^n||^2/n < 1$ then the LRSE equals 0.
Proof: If $||x^n||^2/n < 1$ then, by Lemma 3, $H_n(\tau^2) < 1$. Therefore by (2), $dg(\tau^2)/d\tau^2 < 0$ for all $\tau^2 \geq 0$ and this implies that g is decreasing on $[0, \infty)$ and so the LRSE is 0. If $||x^n||^2/n = 1$ then, $dg(0)/d\tau^2 = 0$ and $dg(\tau^2)/d\tau^2 < 0$ on $(0, \infty)$ and so the LRSE equals 0 and is the unique solution to $H_n(\tau^2) = 1$. If $||x^n||^2/n > 1$ then from Lemma 3 there is a unique solution to $H_n(\tau^2) = 1$ and

(2) establishes that $dg(\tau^2)/d\tau^2 > 0$ to the left of this value and $dg(\tau^2)/d\tau^2 < 0$ to the right of this value which proves that it is the LRSE.

Now suppose that τ_n^2/n is bounded away from 0 in the sense that, for given $\epsilon > 0$ there exists n_ϵ such that for all $n \geq n_\epsilon$ we have that $\tau_n^2/n \geq \epsilon$. Then, for all $n \geq n_\epsilon$, we have that $P(||x^n||^2/n \geq 1) \geq P(||x^n||^2/n \geq 1 + \tau_n^2/n - \epsilon) \geq P(\left| ||x^n||^2/n - 1 - \tau_n^2/n \right| \leq \epsilon)$ and, when $\tau_n^2 = o(n^2)$, the probability on the right goes to 1 by Proposition 1. So in this case we have that with high probability the LRSE is given by the unique solution to $H_n(\tau^2) = 1$.

We now establish the consistency of the LRSE.

Proposition 5. When $\tau_n^2 = o(n^2)$ and τ_n^2/n is bounded away from 0, then the LRSE is consistent.

Proof: Let $C_n = \left\{ ||x^n||^2/n \geq 1 \right\}$ and suppose $x^n \in C_n$ for each n. Let $\hat{\tau}_n^2$ denote the unique solution to $H_n(\tau^2) = 1$ for each such x^n. We prove that, for $\epsilon > 0$, then $\lim_{n\to\infty} P\left(\left| \hat{\tau}_n^2/n - \tau_n^2/n \right| > \epsilon \,|\, C_n \right) = 0$. Then since $P(C_n) \to 1$, we have that $P\left(\left| \hat{\tau}_n^2/n - \tau_n^2/n \right| \leq \epsilon \right) \geq P\left(\left\{ \left| \hat{\tau}_n^2/n - \tau_n^2/n \right| \leq \epsilon \right\} \cap C_n \right) = P\left(\left| \hat{\tau}_n^2/n - \tau_n^2/n \right| \leq \epsilon \,|\, C_n \right) P(C_n) \to 1$ and so the consistency of the LRSE will be established.

Assuming that $x^n \in C_n$ then the LRSE satisfies $H_n(\tau^2) = 1$. Multiplying both sides of this equation by τ^2 gives the equivalent equation

$$\tau^2 = ||x^n||\tau I_{n/2}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right). \tag{3}$$

Now $I_{n/2}\left(||x^n||\tau\right) = (||x^n||\tau/n)\{I_{(n/2)-1}\left(||x^n||\tau\right) - I_{(n/2)+1}\left(||x^n||\tau\right)\}$ by Lemma 2(vi) and therefore, $\tau^2 = (||x^n||^2\tau^2/n)\left\{1 - I_{(n/2)+1}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right)\right\}$. Thus from Lemma 2(iv), and using (3), we have

$$\tau^2 \geq \frac{||x^n||^2\tau^2}{n}\left\{1 - \frac{I_{n/2}^2\left(||x^n||\tau\right)}{I_{(n/2)-1}^2\left(||x^n||^2\tau^2\right)}\right\} = \frac{||x^n||^2\tau^2}{n}\left\{1 - \frac{\tau^2}{||x^n||^2}\right\}$$

and so

$$\hat{\tau}_n^2 \geq ||x^n||^2 - n. \tag{4}$$

Applying Lemma 2(vi) to the denominator in (3) we have

$$\begin{aligned}
\tau^2 &= ||x^n||\tau\, I_{n/2}\left(||x^n||\tau\right)/\left\{(n/||x^n||\tau)\, I_{n/2}\left(||x^n||\tau\right) + I_{(n/2)+1}\left(||x^n||\tau\right)\right\} \\
&= ||x^n||^2\tau^2/\left\{n + ||x^n||\tau I_{(n/2)+1}\left(||x^n||\tau\right)/I_{n/2}\left(||x^n||\tau\right)\right\}
\end{aligned}$$

and rearranging this gives

$$||x^n||\tau I_{(n/2)+1}\left(||x^n||\tau\right)/I_{n/2}\left(||x^n||\tau\right) = ||x^n||^2 - n. \tag{5}$$

Now applying Lemma 2(vi) to the numerator in (3), apply Lemma 2(iii) and Lemma 2(iv) again to obtain

$$\begin{aligned}
\tau^2 &= ||x^n||\tau\left\{\frac{n+2}{||x^n||\tau}\frac{I_{(n/2)+1}}{I_{(n/2)-1}}+\frac{I_{(n/2)+2}}{I_{(n/2)-1}}\right\}\\
&= ||x^n||\tau\left\{\frac{n+2}{||x^n||\tau}\frac{I_{(n/2)+1}/I_{n/2}}{I_{(n/2)-1}/I_{n/2}}+\frac{I_{(n/2)+2}/I_{(n/2)+1}}{I_{(n/2)-1}/I_{(n/2)+1}}\right\}\\
&\le ||x^n||\tau\left\{\frac{n+2}{||x^n||\tau}\frac{I_{(n/2)+1}/I_{n/2}}{I_{(n/2)-1}/I_{n/2}}+\frac{I_{(n/2)+1}/I_{n/2}}{I_{(n/2)-1}/I_{(n/2)+1}}\right\}\\
&= ||x^n||\tau\frac{I_{(n/2)+1}}{I_{n/2}}\left\{\frac{n+2}{||x^n||\tau}\frac{I_{n/2}}{I_{(n/2)-1}}+\frac{I_{(n/2)+1}}{I_{(n/2)-1}}\right\}\\
&\le ||x^n||\tau\frac{I_{(n/2)+1}}{I_{n/2}}\left\{\frac{n+2}{||x^n||\tau}\frac{I_{n/2}}{I_{(n/2)-1}}+\frac{I^2_{n/2}}{I^2_{(n/2)-1}}\right\}.
\end{aligned}$$

On the right side of this apply (5) to $||x^n||\tau I_{(n/2)+1}/I_{n/2}$ and finally (3) to $I_{n/2}/I_{(n/2)-1}$ to obtain $\tau^2 \le \left(||x^n||^2 - n\right)\left\{(n+2)/||x^n||^2 + \tau^2/||x^n||^2\right\}$. Rearranging this inequality we conclude that $\hat{\tau}^2_n \le \{||x^n||^2 - n\}\,(n+2)\,/n$. Combining this with (4) we have that, whenever C_n is true, then

$$\left(||x^n||^2 - n\right) \le \hat{\tau}^2_n \le \left(||x^n||^2 - n\right)(n+2)\,/n. \tag{6}$$

As $P\left(\left|||x^n||^2/n - 1 - \tau^2_n/n\right| \le \epsilon\right) = P\left(\left|||x^n||^2/n - 1 - \tau^2_n/n\right| \le \epsilon \,|\, C_n\right) P(C_n)$, and $P(C_n) \to 1$, then (1) implies that $P\left(\left|||x^n||^2/n - 1 - \tau^2_n/n\right| \le \epsilon \,|\, C_n\right) \to 1$. Similarly, $P\left(\left|\left(||x^n||^2/n\right)(n+2)\,/n - (n+2)\,/n - \tau^2_n/n\right| \le \epsilon \,|\, C_n\right) \to 1$ and from (6) we conclude that $P\left(\left|\hat{\tau}^2_n/n - \tau^2_n/n\right| \le \epsilon \,|\, C_n\right) \to 1$. This implies that the LRSE is consistent.

From the proof of Proposition 5 we obtain the result that the LRSE is effectively the same as the estimator $(||x^n||^2 - n)_+$ when n is large.

Corollary 6. For the LRSE $\hat{\tau}^2_n$ we have that, whenever $\hat{\tau}^2_n = 0$, then $\hat{\tau}^2_n = (||x^n||^2 - n)_+$ and, whenever $\hat{\tau}^2_n > 0$, then $1 \le \hat{\tau}^2_n/(||x^n||^2 - n)_+ \le 1 + 2/n$.

Proof: When $\hat{\tau}^2_n = 0$ then either $x^n \notin C_n$ or $||x^n||^2 - n = 0$. In either case $(||x^n||^2 - n)_+ = 0$. When $x^n \in C_n$, and $\hat{\tau}^2_n > 0$, then (6) implies $||x^n||^2 - n > 0$, and so $(||x^n||^2 - n)_+ = ||x^n||^2 - n$. Then (6) implies that $1 \le \hat{\tau}^2_n/(||x^n||^2 - n)_+ \le 1 + 2/n$.

Therefore, when $\hat{\tau}^2_n = 0$ the absolute difference between the LRSE and $(||x^n||^2 - n)_+$ is 0, and when $\hat{\tau}^2_n \ne 0$ the absolute relative difference between the LRSE and $(||x^n||^2 - n)_+$ is bounded above by $2/n$.

We note that when τ^2_n/n converges to a nonzero value, then $\tau^2_n = o(n^2)$ and τ^2_n/n is bounded away from 0, so $\hat{\tau}^2_n$ is consistent. In particular, if $\sigma^2 > 0$ and $\tau^2_n/n \to \sigma^2$, then the LRSE is consistent. Therefore the LRSE is consistent in much wider generality than the posterior mean which is only consistent when $\tau^2_n/n \to \sigma^2$.

It is also of interest to see what happens when $\tau^2_n/n \to 0$. This drops the "bounded away from 0" requirement in Proposition 3, but requires the convergence of τ^2_n/n.

Proposition 7. When $\tau_n^2/n \to 0$, then the LRSE is consistent.

Proof: In this case we have that $||x^n||^2/n \xrightarrow{P} 1$ by (1). If $||x^n||^2/n \le 1$, then $\hat{\tau}_n^2 = 0$. Therefore, if $\epsilon > 0$ and $\hat{\tau}_n^2/n > \epsilon$ this entails that $||x^n||^2/n > 1$ and arguing, just as in the proof of Proposition 5, we must have that (6) holds. Then, $P\left(\hat{\tau}_n^2/n > \epsilon\right) \le P(\left(||x^n||^2 - n\right)(n+2)/n^2 > \epsilon) = P(||x^n||^2/n > 1 + \epsilon/(1+2/n)) \le P(\left(||x^n||^2/n > 1 + \epsilon/3)\right) \to 0$ as $n \to \infty$ and giving the result.

4 The Posterior Mode

Letting $f(\tau^2) = \ln \pi(\tau^2 \,|\, x^n)$ we have that

$$f(\tau^2) = C + (n/2 - 1)\ln \tau^2 - \left(1 + 1/\sigma^2\right)(\tau^2/2) + \ln \sum_{k=0}^{\infty} \frac{\left(||x^n||^2\tau^2/4\right)^k}{k!\Gamma\left(n/2 + k\right)}.$$

If $n > 2$, then $\pi(0 \,|\, x^n) = 0$ and so the mode is a critical point of f. Using the definition of $H_n(\tau^2)$ from (2), we have $2d\,f(\tau^2)/d\tau^2 = (n-2)/\tau^2 - \left(1 + 1/\sigma^2\right) + H_n(\tau^2)$, and so the mode is a solution to $(n-2)/\tau^2 + H_n(\tau^2) = \left(1 + 1/\sigma^2\right)$.

The following properties of $G_n(\tau^2) = (n-2)/\tau^2 + H_n(\tau^2)$ follow immediately from Lemma 3.

Lemma 8. The function G_n satisfies
(i) $G_n(\tau^2)$ is a strictly decreasing function,
(ii) $\lim_{\tau^2 \longrightarrow 0^+} G_n(\tau^2) = \infty$, and
(iii) $\lim_{\tau^2 \longrightarrow \infty} G_n(\tau^2) = 0$.

Note that Lemma 8 establishes that there is always a solution to $\left(1 + 1/\sigma^2\right) = G_n(\tau^2)$ for $\tau^2 \in [0, \infty)$ and it is unique. This solution is the posterior mode $\tilde{\tau}_n^2$.

We now consider the consistency of the mode.

Proposition 9. When $\tau_n^2 = o(n^2)$ we have that $\tilde{\tau}_n^2$ is consistent if and only if the posterior mean is consistent. Further $m(n)/n - \tilde{\tau}_n^2/n \to 0$ in probability.

Proof: We assume $n > 2$ hereafter. Putting $l(\tau^2) = \left(1 + 1/\sigma^2\right)\tau^2 - (n-2)$, we have that the mode satisfies

$$l(\tau^2) = ||x^n||\tau I_{n/2}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right). \tag{7}$$

Applying Lemma 2(vi) to the numerator in (7), we obtain $l(\tau^2) = (||x^n||^2\tau^2/n) \times \left\{1 - I_{(n/2)+1}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right)\right\}$. Then using Lemma 2(iv) and (8) we have that

$$l(\tau^2) \ge \frac{||x^n||^2\tau^2}{n}\left\{1 - \frac{I_{n/2}^2\left(||x^n||\tau\right)}{I_{(n/2)-1}^2\left(||x^n||^2\tau^2\right)}\right\} = \frac{||x^n||^2\tau^2}{n}\left\{1 - \frac{l(\tau^2)^2}{||x^n||^2\tau^2}\right\}.$$

and rearranging this conclude that the mode satisfies

$$\left(1 + 1/\sigma^2\right)^2\tau^4 - \{||x^n||^2 + \left(1 + 1/\sigma^2\right)(n-4)\}\tau^2 - 2(n-2) \ge 0. \tag{8}$$

The roots of the quadratic in τ^2 in (8) are given by $r_1(n) \pm r_2(n)$ where

$$\begin{aligned} r_1(n) &= \left(||x^n||^2 + \left(1+1/\sigma^2\right)(n-4)\right)/2\left(1+1/\sigma^2\right)^2, \\ r_2(n) &= (r_1^2(n) + 2(n-2)/\left(1+1/\sigma^2\right)^2)^{1/2}. \end{aligned}$$

Since it is clear that $r_1(n) - r_2(n) < 0$ we have shown that the mode satisfies

$$r_1(n) + r_2(n) \leq \tilde{\tau}_n^2. \tag{9}$$

Now observe that

$$\begin{aligned} |r_1(n)/n - r_2(n)/n| &= \left|\sqrt{r_1^2(n)/n^2} - \sqrt{r_1^2(n)/n^2 + 2(n-2)/n^2\left(1+1/\sigma^2\right)^2}\right| \\ &= \frac{2(n-2)/n^2\left(1+1/\sigma^2\right)^2}{\sqrt{r_1^2(n)/n^2} + \sqrt{r_1^2(n)/n^2 + 2(n-2)/n^2\left(1+1/\sigma^2\right)^2}} \\ &\leq \sqrt{2(n-2)/n^2\left(1+1/\sigma^2\right)^2} \to 0 \end{aligned}$$

and so $r_2(n)/n \xrightarrow{P} r_1(n)/n$, which gives that $r_1(n) + r_2(n) \xrightarrow{P} 2r_1(n)/n$.

Now the posterior mean is given by $m(n) = (1+1/\sigma^2)^{-2}||x^n||^2 + n(1+1/\sigma^2)^{-1}$ and $m(n)/n - 2r_1(n)/n = (4/n)\left(1+1/\sigma^2\right)^{-1} \to 0$ which implies that $m(n)/n - (r_1(n) + r_2(n)) \xrightarrow{P} 0$.

Now apply Lemma 2(vi) to the denominator on the right-hand side of (7) to get $l(\tau^2) = ||x^n||^2\tau^2/\left\{n + ||x^n||\tau I_{(n/2)+1}\left(||x^n||\tau\right)/I_{n/2}\left(||x^n||\tau\right)\right\}$. Rearranging this gives

$$(||x^n||\tau) I_{(n/2)+1}\left(||x^n||\tau\right)/I_{n/2}\left(||x^n||\tau\right) = (||x^n||^2\tau^2/l(\tau^2)) - n. \tag{10}$$

Just as in the proof of Proposition 5 we have that

$$l(\tau^2) \leq ||x^n||\tau \frac{I_{(n/2)+1}}{I_{n/2}} \left\{ \frac{n+2}{||x^n||\tau} \frac{I_{n/2}}{I_{(n/2)-1}} + \frac{I_{n/2}^2}{I_{(n/2)-1}^2} \right\}$$

and then using (7)

$$l(\tau^2) \leq ||x^n||\tau \frac{I_{(n/2)+1}}{I_{n/2}} \left\{ \frac{(n+2)l(\tau^2)}{||x^n||^2\tau^2} + \frac{l^2(\tau^2)}{||x^n||^2\tau^2} \right\}.$$

Now use (10) to obtain $l(\tau^2) \leq \{||x||^2\tau^2 - nl(\tau^2)\}\{(n+2) + l(\tau^2)\}/(||x^n||^2\tau^2)$ or

$$(||x^n||^2\tau^2)l(\tau^2) - \{||x||^2\tau^2 - nl(\tau^2)\}\{(n+2) + l(\tau^2)\} \leq 0 \tag{11}$$

and note that the expression of the left is a quadratic in τ^2. Collecting coefficients this quadratic is given by $n\left(1+1/\sigma^2\right)^2\tau^4 - \{(n+2)||x||^2 - n(n-6)\left(1+1/\sigma^2\right))\}\tau^2 - 4n(n-2)$. As the coefficient of τ^4 is positive, then (11) implies that

$$\tilde{\tau}_n^2 \leq s_1(n) + s_2(n) \tag{12}$$

where $s_1(n)+s_2(n)$ is the largest root of the quadratic and so

$$\begin{aligned} s_1(n) &= \left(1+1/\sigma^2\right)^{-2}\{(n+2)||x^n||^2 - n(n-6)\left(1+1/\sigma^2\right)\}/2n \\ s_2(n) &= \sqrt{s_1^2(n)+4(1-2/n)\left(1+1/\sigma^2\right)^{-4}}. \end{aligned}$$

Observe that $|s_1(n)/n - s_2(n)/n| \to 0$, so $s_1(n)+s_2(n) \xrightarrow{P} 2s_1(n)$. Now

$$\begin{aligned} & m(n)/n - 2s_1(n)/n = (1+1/\sigma^2)^{-2}||x^n||^2/n + (1+1/\sigma^2)^{-1} - 2s_1(n) \\ = \; & (1+1/\sigma^2)^{-2}(-2||x^n||^2/n^2) + (1+1/\sigma^2)^{-1} - (1-6/n)\left(1+1/\sigma^2\right) \end{aligned}$$

and so $m(n)/n - 2s_1(n)/n \xrightarrow{P} 0$, since $\tau_n^2 = o(n^2)$ and (1) imply $||x^n||^2/n^2 \xrightarrow{P} 0$.

By (9) and (12)

$$\begin{aligned} & (r_1(n)/n + r_2(n)/n - m(n)/n) + \left(m(n)/n - \tau_n^2/n\right) \\ = \; & r_1(n)/n + r_2(n)/n - \tau_n^2/n \le \tilde{\tau}_n^2/n - \tau_n^2/n \le s_1(n)/n + s_2(n)/n - \tau_n^2/n \\ = \; & (s_1(n)/n + s_2(n)/n - m(n)/n) + \left(m(n)/n - \tau_n^2/n\right) \end{aligned}$$

and this establishes the result.

Proposition 1 establishes the inconsistency of the posterior mean except in very limited circumstances and so, by Proposition 9, this comment applies to the posterior mode as well. As the posterior is a right-skewed distribution it seems likely that the posterior median lies between the mode and mean and, if this is the case, then Proposition 9 applies to this estimator as well.

5 Interval Estimates

For the consistency of interval estimates, we say two intervals are asymptotically equivalent if the differences in the respective endpoints go to 0 and the ratio of their lengths goes to 1 in probability, as $n \to \infty$. Then, if an interval is asymptotically equivalent to an interval that always contains the true value of the parameter, we say it is consistent. For example, in a sample of n from a $N(\mu,1)$ distribution with μ unknown, an interval of the form $\bar{x} \pm z_*/\sqrt{n}$ with z_* a constant, is asymptotically equivalent to $\mu \pm z_*/\sqrt{n}$, and so $\bar{x} \pm z_*/\sqrt{n}$ is consistent. Such consistency seems like a natural requirement of any interval estimator. Note that consistency results for interval estimates don't say anything about their long-run relative frequency of containing the true value of the parameter, although we would expect inconsistent intervals to do rather poorly in this regard.

If $X_n \,|\, \delta_n^2 \sim$ Chi-squared(n, δ_n^2) and $\delta_n^2/n \xrightarrow{P} \delta_*^2$, then it is easy to show that $(X_n - E(X_n))/(Var(X_n))^{1/2} \xrightarrow{D} N(0,1)$. So if $\tau_n^2/n \to \tau_*^2$ then $||x^n||^2/n \xrightarrow{P} 1+\tau_*^2$ and so $(\tau^2 - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \,|\, x^n \xrightarrow{D} N(0,1)$ where $E(\tau^2 \,|\, x^n) = (1+1/\sigma^2)^{-1}\left\{n + (1+1/\sigma^2)^{-1}||x^n||^2\right\}$, and $Var(\tau^2 \,|\, x^n) = 2(1+1/\sigma^2)^{-2}\{n + 2(1+1/\sigma^2)^{-1}||x^n||^2\}$. From this we get an approximate γ-credible interval for

τ_n^2/n, obtained by discarding $(1-\gamma)/2$ of the probability in each tail of the posterior, given by

$$\begin{aligned}&\left\{(1+1/\sigma^2)^{-1}+(1+1/\sigma^2)^{-2}(||x^n||^2/n)\right\}\\&\pm\sqrt{2(1+1/\sigma^2)^{-2}+4(1+1/\sigma^2)^{-3}(||x^n||^2/n)}(z_{(1+\gamma)/2}/\sqrt{n})\end{aligned}\qquad(13)$$

where $z_{(1+\gamma)/2}$ is the $(1+\gamma)/2$-quantile of the $N(0,1)$ distribution.

Now consider the interval given by

$$\begin{aligned}&\left\{(1+1/\sigma^2)^{-1}+(1+1/\sigma^2)^{-2}(1+\tau_*^2)\right\}\\&\pm\sqrt{2(1+1/\sigma^2)^{-2}+4(1+1/\sigma^2)^{-3}(1+\tau_*^2)}(z_{(1+\gamma)/2}/\sqrt{n}).\end{aligned}\qquad(14)$$

Comparing (13) and (14) we see that the differences in the respective endpoints converge to 0, and the ratio of their lengths goes to 1, in probability as $n\to\infty$. So (13) and (14) are asymptotically equivalent. Now consider whether or not (14) contains $\tau_n^2/n\approx\tau_*^2$. The asymptotic error in the posterior mean is given by

$$\begin{aligned}&\tau_*^2-\left\{(1+1/\sigma^2)^{-1}+(1+1/\sigma^2)^{-2}(1+\tau_*^2)\right\}\\=\;&\left(1+1/\sigma^2\right)^{-2}\left(2+1/\sigma^2\right)\left(\tau_*^2/\sigma^2-1\right).\end{aligned}\qquad(15)$$

Note that τ_*^2 is in (14) if and only if 0 is in the interval obtained by adding (15) to each point in (14). From this we see that the true value τ_*^2 is always in (14) when $\tau_*^2=\sigma^2$. Now suppose that $\tau_*^2\neq\sigma^2$. If $\tau_*^2/\sigma^2>1$ then the sum of (15) and the left-hand endpoint of (14) is greater than 0 for all n large enough. If $\tau_*^2/\sigma^2<1$ then the sum of (15) and the right-hand endpoint of (14) is less than 0 for all n large enough. Therefore, when $\tau_*^2\neq\sigma^2$, (14) will never contain τ_*^2 for all n large enough. These conclusions are independent of σ^2. Also τ_*^2 is never in the interval for all n when $\sigma^2\to\infty$ and so (13) is inconsistent.

Due to difficulties in approximating the noncentral Chi-squared density function we have not been able to obtain useful approximate forms for hpd and relative surprise intervals in general. We note, however, that the results of section 4 suggest that hpd intervals will have the very poor coverage properties of the credible intervals constructed above. Further, the consistency of the LRSE, and the fact that the LRSE is always in any relative surprise interval, suggest the coverage properties of relative surprise intervals will be much improved. This is confirmed by Proposition 10 and the simulation results reported in Evans (1997).

We can establish the consistency of relative surprise intervals under the additional condition that $\tau_n^2/n\to\tau_*^2=0$. For example, this situation obtains when $\tau_n^2=\sum_{i=1}^n\theta_i^2$ is convergent.

Proposition 10. When $\tau_n^2=o(n)$ the γ-relative surprise interval for τ_n^2/n is asymptotically equivalent to the interval with left-hand endpoint equal to 0 and right-hand endpoint equal to $r_n^*=\left\{(1+1/\sigma^2)^{-1}+(1+1/\sigma^2)^{-2})\right\}+(2(1+1/\sigma^2)^{-2}+4(1+1/\sigma^2)^{-3})^{1/2}(z_\gamma/\sqrt{n})$ and so is consistent.

Proof: Note that the developments in section 3 imply that the γ-relative surprise interval $(l_n(x^n), r_n(x^n)) \subset [0, \infty)$ for τ_n^2 contains the LRSE $\hat{\tau}_n^2$. By Proposition 7, $\hat{\tau}_n^2/n$ converges in probability to 0 and so $l_n(x^n)/n$ must also converge in probability to 0.

The proof that $r_n(x^n)/n \xrightarrow{P} r_n^*$ is more difficult. First observe that $(nr_n^* - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \xrightarrow{P} z_\gamma$ and so it suffices to prove that $(r_n(x^n) - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \xrightarrow{P} z_\gamma$. These results also imply that $(r_n(x^n) - l_n(x^n))/nr_n^* \xrightarrow{P} 1$.

If $A_n(z) = \{x^n : (l_n(x^n) - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \leq z\}$, then $P(A_n(z)) \to 1$ for all z. Denote the posterior of τ_n^2 by $\Pi(\cdot \,|\, x^n)$. Let $\epsilon > 0$ and write

$$\begin{aligned} P(\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon) \quad &= \quad P(A_n(z) \cap \{\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon\}) + \\ & \quad\quad P(A_n^c(z) \cap \{\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon\}). \end{aligned} \tag{16}$$

Note that $x^n \in A_n(z)$ implies

$$\begin{aligned} & \Pi\left(\frac{\tau^2 - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \leq \frac{l_n(x^n) - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \,\middle|\, x^n \right) \\ \leq \quad & \Pi\left(\frac{\tau^2 - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \leq z \,\middle|\, x^n \right) \end{aligned}$$

and so the first term in (16) satisfies

$$\begin{aligned} & P(A_n(z) \cap \{\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon\}) \\ \leq \quad & P\left(A_n(z) \cap \left\{ \Pi\left(\frac{\tau^2 - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \leq z \,\middle|\, x^n \right) > \epsilon \right\} \right) \\ \leq \quad & P\left(\left\{ \Pi\left(\frac{\tau^2 - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \leq z \,\middle|\, x^n \right) > \epsilon \right\} \right). \end{aligned} \tag{17}$$

Let Φ denote the $N(0,1)$ cdf and choose z so that $\Phi(z) < \epsilon/2$. Since $(\tau^2 - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \xrightarrow{D} N(0,1)$, we have that

$$\Pi\left(\frac{\tau^2 - E(\tau^2 \,|\, x^n)}{(Var(\tau^2 \,|\, x^n))^{1/2}} \leq z \,\middle|\, x^n \right) \to \Phi(z)$$

for every data sequence x^n and so this convergence is also in probability with respect to P. This implies that (17) converges to 0. Further, the second term in (16) is bounded above by $P(A_n^c(z))$ which converges to 0. Accordingly we have proved that $\Pi([0, l_n(x^n)] \,|\, x^n) \xrightarrow{P} 0$. Also, we always have that $\gamma = \Pi([l_n(x^n), r_n(x^n)] \,|\, x^n) = \Pi([0, r_n(x^n)] \,|\, x^n) - \Pi([0, l_n(x^n)] \,|\, x^n)$ and so $\Pi([0, r_n(x^n)] \,|\, x^n) \xrightarrow{P} \gamma$.

Now let $\epsilon > 0$ and $B_n = \{x^n : (r_n(x^n) - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} > z_\gamma + \epsilon\}$. Put $\epsilon' = \Phi(z_\gamma + \epsilon) - \gamma$ and write

$$\begin{aligned} P(B_n) \quad &= \quad P(B_n \cap \{\Pi([0, r_n(x^n)] \,|\, x^n) > \gamma + \epsilon'/2\}) + \\ & \quad\quad P(B_n \cap \{\Pi([0, r_n(x^n)] \,|\, x^n) \leq \gamma + \epsilon'/2\}). \end{aligned} \tag{18}$$

The first term in (18) is bounded above by $P(\Pi([0,r_n(x^n)]\,|\,x^n) > \gamma+\epsilon'/2)$ and this converges to 0. For the second term in (18) we have that

$$\begin{aligned}
& P(B_n \cap \{\Pi([0,r_n(x^n)]\,|\,x^n) \le \gamma+\epsilon'/2\}) \\
= \;& P\left(B_n \cap \left\{ \begin{array}{c} \Pi\left(\frac{\tau^2 - E(\tau^2\,|\,x^n)}{(Var(\tau^2\,|\,x^n))^{1/2}} \le \frac{r_n(x^n)-E(\tau^2\,|\,x^n)}{(Var(\tau^2\,|\,x^n))^{1/2}} \,\middle|\, x^n\right) \\ \le \gamma+\epsilon'/2 \end{array} \right\}\right) \\
\le \;& P\left(B_n \cap \left\{\Pi\left(\frac{\tau^2 - E(\tau^2\,|\,x^n)}{(Var(\tau^2\,|\,x^n))^{1/2}} \le z_\gamma + \epsilon \,|\, x^n\right) \le \gamma + \epsilon'/2\right\}\right) \qquad (19)
\end{aligned}$$

and

$$\Pi\left(\frac{\tau^2 - E(\tau^2\,|\,x^n)}{(Var(\tau^2\,|\,x^n))^{1/2}}) \le z_\gamma + \epsilon \,|\, x^n\right) \xrightarrow{P} \Phi(z_\gamma+\epsilon) = \gamma + \epsilon'$$

implies that (19) converges to 0. Similarly if we set $C_n = \{x^n : (r_n(x^n) - E(\tau^2\,|\,x^n))/(Var(\tau^2\,|\,x^n))^{1/2} < z_\gamma - \epsilon\}$ we get that $P(C_n) \to 0$ and so $(r_n(x^n) - E(\tau^2\,|\,x^n))/(Var(\tau^2\,|\,x^n))^{1/2} \xrightarrow{P} z_\gamma$. This completes the proof.

Notice that the interval $[0, r_n^*] \to [0, (1+1/\sigma^2)^{-1} + (1+1/\sigma^2)^{-2}]$ as $n \to \infty$ and so the γ-relative surprise interval for τ_n^2/n does not shrink to $\{0\}$ as we increase n. When $\sigma^2 \to \infty$ this interval is $[0,2]$. So under a diffuse prior there is a level of uncertainty about the true value of τ_n^2/n that cannot be avoided no matter how large n is. To see that this makes sense, suppose that τ_n^2 converges. Then, even if we were to exactly observe n values θ_i, we would still have no idea as to whether or not τ_n^2 is close to $\sum_{i=1}^\infty \theta_i^2$. So even with no error in the observations, there is a fundamental uncertainty that cannot be decreased by increasing n. When $\sigma^2 \to 0$, then $[0, r_n^*] \to \{0\}$ for all n. So if we have very precise information that the the θ_i are close to 0, then this uncertainty is largely avoided. Effectively the prior controls the precision of inferences more than n.

6 Checking for Prior-Data Conflict

We have shown that the $N(0,\sigma^2)$ prior can lead to sensible Bayesian estimates of τ_n^2. Still we might ask if this prior makes sense in a particular problem. It is argued in Evans and Moshonov (2006) that an important aspect of a Bayesian analysis is to check for prior-data conflict and, that this is something we check for after checking that the sampling model is consistent with the data. The sampling model is consistent with the data provided there is at least one distribution in the sampling model for which the observed data is not surprising. Given that the θ_i are arbitrary, it is clear that the sampling model is always consistent with the observed data and so the only check here is for prior-data conflict.

A prior-data conflict exists when the prior places its mass primarily on parameter values for which the observed data is surprising. In this case checking for prior-data conflict entails comparing the observed value of the minimal sufficient statistic $T(x^n) = x^n$ with its prior predictive distribution M_T, the $N_n(0, (1+\sigma^2)I)$ distribution, to see if it is reasonable. So we need to check if x^n is out in the tails of this distribution and clearly this can be measured by how far $||x^n||^2$ is from the origin. Note that the prior predictive distribution of $(1+\sigma^2)^{-1}||x^n||^2$ is Chi-squared$(n,0)$ and so we can make this comparison by computing $P(X^2 > (1+\sigma^2)^{-1}||x^n||^2)$ when $X^2 \sim$ Chi-squared$(n,0)$. From section 5 we see that the effect of the prior remains even as $n \to \infty$ so prior-data conflict cannot be ignored.

Now suppose that we have evidence of a prior-data conflict. Clearly this is caused by the selected value of σ^2 being too small. As $\sigma^2 \to \infty$ the P-value converges to 1. As discussed in Evans and Moshonov (2006), such a sequence of priors satisfies at least a necessary requirement for a sequence of priors to be noninformative, namely, that we never find evidence of prior-data conflict no matter what data is obtained. A reasonable approach then, when we have evidence of a prior-data conflict existing, is to choose σ^2 much larger so that the conflict is avoided.

7 Conclusions

Our overall concern here has been to show that for the problem stated, a reasonable Bayesian solution is available that does not require that we choose a different prior, or somehow modify the prior, to suit the parameter of interest. We do not claim that the modification of priors to suit marginal parameters is something that can be always avoided, only that in this problem it does not seem to be necessary. For large σ^2 the chosen prior behaves like a noninformative prior should, at least when considering estimation of τ_n^2.

8 Appendix

Proof of Lemma 2. Parts (i), (ii) and (iii) are stated in Saxena and Alam (1982) without proof. Part (i) follows since $I_{p-1}(x) = (x/2)^{p-1}\sum_{k=0}^{\infty} a_k(x^2/4)^k$ so $(xI_p(x)/I_{p-1}(x))' = 2(\sum_{k=0}^{\infty} ka_k(x^2/4)^k/\sum_{k=0}^{\infty} a_k(x^2/4)^k)' = (4/x)n(x)/d(x)$ with $n(x) = \{\sum_{k=0}^{\infty} k^2 a_k(x^2/4)^k \sum_{k=0}^{\infty} a_k(x^2/4)^k - (\sum_{k=0}^{\infty} ka_k(x^2/4)^k)^2\}$ and $d(x) = (\sum_{k=0}^{\infty} .a_k(x^2/4)^k)^2$. Then, $c_m = \sum_{k=0}^{m} k(2k-m)a_k a_{m-k}$ is the coefficient of $\left(x^2/4\right)^m$ in $n(x)$. Using results in Kemp and Kemp (1956) we see that the coefficients $a_k a_{m-k}$ define a Type I A(i) generalized hypergeometric distribution(a,b,n) where $a = m, b = 2(p-1)+m$ and $n = p-1+m$. If we denote such a random variable by X_m we see that $c_0 = 0$ and, for $m \geq 1$, c_m is a positive constant times $2E(X_m^2) - mE(X_m) = 2Var(X_m) + E(X_m)(2E(X_m) - m) = 2Var(X_m) > 0$ since, by Kemp and Kemp (1956), $E(X_m) = na/(a+b) = (p-1+m)m/2(p-1+m) = m/2$. This implies the result.

Note that $I_p(x)/xI_{p-1}(x) = (1/2)\sum_{k=0}^{\infty}(p+k)^{-1}a_k(x^2/4)^k/\sum_{k=0}^{\infty} a_k(x^2/4)^k)$. The derivative of this is $(2/x)n_*(x)/d(x)$ where, after collecting terms in $n_*(x)$, the coefficient of $\left(x^2/4\right)^m$ is $-m\sum_{k=0}^{m}(p+k)^{-1}a_k a_{m-k}$ which implies part (ii).

We have that $I_p(x)/I_{p-1}(x) = (2/x)\sum_{k=0}^{\infty} ka_k(x^2/4)^k/\sum_{k=0}^{\infty} a_k(x^2/4)^k$ so

$$\frac{I_{p+1}(x)/I_p(x)}{I_p(x)/I_{p-1}(x)} = \frac{\sum_{m=0}^{\infty}(\sum_{k=0}^{m} k(k-1)a_k a_{m-k})(x^2/4)^m}{\sum_{m=0}^{\infty}(\sum_{k=0}^{m} k(m-k)a_k a_{m-k})(x^2/4)^m}.$$

Now compare the coefficients of $(x^2/4)^m$ in the numerator and denominator. When $m = 0$ they are the same. When $m \geq 1$ then the results in Kemp and Kemp (1956) indicate that the numerator coefficient is a positive constant c times $Var(X_m) + E(X_m)(E(X_m) - 1)$ and the denominator coefficient is $E(X_m)(m - E(X_m)) - Var(X_m)$ where $E(X_m) = m/2$ and $Var(X_m) = (m/4)(2(p-1)+m)/(2(p-1+m)-1)$. Then comparing these coefficients we see that they are equal when $m = 1$ and otherwise the numerator coefficient is strictly smaller than the denominator coefficient. This implies the result.

Part (iv) follows from part (iii) since $I_{p+1}/I_{p-1} = (I_{p+1}/I_p) \,/\, (I_{p-1}/I_p) \leq (I_p/I_{p-1}) \,/\, (I_{p-1}/I_p)$. Parts (v) and (vi) are standard results that are found in many references on Bessel functions, e.g., Abramowitz and Stegun (1970).

References

Abramowitz, M. and Stegun, I.A. eds. (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York.

Berger, J.O., Liseo, B. and Wolpert, R.L. (1999). Integrated likelihood methods for eliminating nuisance parameters (with discussion). Stat. Sci., Vol. 14, No. 1, 1-28.

Chow, M.S. (1987). A complete class theorem for estimating a noncentrality parameter. Ann. Statist., Vol. 15, No. 2, 800-804.

De Waal, D. J. (1974). Bayes estimate of the noncentrality parameter in multivariate analysis. Comm. Statist. ,**3**, 73-79.

Evans, M.(1997). Bayesian procedures derived via the concept of relative surprise. Comm. Statist., **26**, 1125-1143.

Evans, M., Guttman, I. and Swartz, T. (2006). Optimality and computations for relative surprise inferences. Can. J. of Statist., Vol. 34, No. 1, 113-129.

Evans, M. and Moshonov, H. (2006). Checking for prior-data conflict. Bayesian Analysis, Volume 1, Number 4, pp. 893-914.

Kemp, C.D. and Kemp, A.W. (1956). Generalized hypergeometric distributions. J. of the Royal Statist. Soc., 2, 202-211.

Kubokawa, T., Robert, C.P. and Saleh, A.K.Md.E. (1993). Estimation of noncentrality parameters. Can. J. of Statist., Vol. 21, No. 1, 45-57.

Perlman, M.D., and Rasmussen, V. A. (1975). Some remarks on estimating a noncentrality parameter. Comm. Statist. ,**4**, 455-468.

Saxena, K.M.L., and Alam, K.(1982). Estimation of the non-centrality parameter of a chi squared distribution. Ann. Statist., **10**, 1012-1016.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals, Ann. Math. Statist, 30, 877-880.