Sankhyā : The Indian Journal of Statistics
2014, Volume 76-A, Part 1, pp. 25-47

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

Michael Evans
University of Toronto, Toronto, Canada
Mohammed Shakhatreh
Jordan University of Science and Technology, Irbid, Jordan

Abstract

We consider the problem of estimating the sum of squared means when the data $(x_1, \ldots, x_n)$ are independent values with $x_i \sim N(\theta_i, 1)$ and $\theta_1, \theta_2 \ldots$ are *a priori* i.i.d. $N(0, \sigma^2)$ with σ^2 known. This example has posed difficulties for many approaches to inference. We examine the consistency properties of several estimators derived from Bayesian considerations. We prove that a particular Bayesian estimate (LRSE) is consistent in a 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. While it can be argued that the choice of the $N(0, \sigma^2)$ prior is inappropriate when σ^2 is chosen large to reflect noninformativity, this argument is not applicable when σ^2 is chosen to reflect knowledge about the unknowns. As such it is important to show that there are consistent Bayesian estimation procedures using this prior.

AMS (2000) *subject classification.* Primary 62F15; Secondary 62F10.
Keywords 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 Strawderman (1976), Gelfand (1983), 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. The hyperparameter σ^2 is chosen to reflect prior beliefs about the θ_i. We then consider the consistency behavior of various Bayesian estimators, derived according to some principle applied to the specified sampling model and prior.

The prior distribution of τ_n^2/σ^2 is then Chi-squared$(n, 0)$. From this we can deduce that the posterior distribution of $(1 + 1/\sigma^2)\tau_n^2$ is Chi-squared $(n, (1 + 1/\sigma^2)^{-1}||x^n||^2)$. Using squared error loss, the Bayes estimate of τ_n^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 shown in Section 3 this estimator is inconsistent (see Section 3 for the precise definition of consistency used here) except in very limited circumstances. Another commonly used Bayesian estimator is the mode of the posterior density of τ_n^2 which cannot be obtained in closed form. We discuss the consistency of this estimator in Section 5 and show that it is essentially equivalent to the posterior mean. Note that the inconsistency of these estimators holds for every value of σ^2, i.e. it does not depend on choosing σ^2 to be large to reflect diffuse beliefs about the θ_i. The relatively poor performance of these Bayesian inferences has been pointed out by others such as Efron (1973).

In Section 2 we discuss a general approach to deriving Bayesian inferences that we refer to as *relative surprise* or *relative belief inferences*. Suppose we have a statistical model where θ is the parameter indexing the distributions and we are interested in making inferences about a marginal parameter $\psi = \Psi(\theta)$. Relative surprise inferences are based on using the relative belief ratio $RB(\psi)$ as a characterization of the statistical evidence that the true value of $\Psi(\theta)$ is given by ψ. The quantity $RB(\psi)$ measures how beliefs have changed from *a priori* to *a posteriori* that $\Psi(\theta) = \psi$. This leads immediately to estimating the true value of $\Psi(\theta)$ by the value of ψ which maximizes $RB(\psi)$ and to a measure of the accuracy of this estimate through credible regions of a necessary form. Furthermore, $RB(\psi_0)$ is a direct assessment of the evidence for the hypothesis $H_0 : \Psi(\theta) = \psi_0$. We have that $RB(\psi_0) > 1$ is evidence in favor of H_0 and $RB(\psi_0) < 1$ is evidence against H_0 and the strength of this evidence is assessed via the posterior probability that the true value of $\Psi(\theta)$ has a relative belief ratio no greater than $RB(\psi_0)$. So, for relative surprise inferences, both estimation and hypothesis assessment proceed from a common characterization of statistical evidence. In Section 2 we discuss various general results that have been established concerning the properties of this approach to inference and that demonstrate its virtues when compared to more commonly used Bayesian inferences.

The consistency properties of the posterior mean of τ_n^2, the relative surprise estimator of τ_n^2, and the posterior mode of τ_n^2 are discussed in Sections 3, 4 and 5, respectively. It is shown that the relative surprise estimator of τ_n^2 has superior consistency properties when compared to the other Bayesian estimators. In Section 6 we consider the much harder problem of credible intervals for τ_n^2, and establish the consistency of relative surprise credible intervals under some conditions.

It has been argued that the $N(0, \sigma^2)$ prior is a poor choice from various points of view. For example, as just discussed, some common Bayesian estimators are inconsistent except in very isolated circumstances. It is contrary to the essential coherency of Bayesian inference, however, to rule out a proper prior because one does not like the inferences it produces. For example, suppose σ^2 is chosen to represent knowledge concerning effects $\theta_1, \theta_2 \ldots$, i.e., σ^2 is not just chosen large to reflect diffuse knowledge about these quantities. Given that we know the scale on which the x_i values are being measured, it is certainly reasonable to suppose that we have such information. In this situation, we do not believe that it is logical to require the prior be changed, although we acknowledge that others may have a different view on this. As opposed to changing the prior, we can look for Bayesian inferences that possess appropriate properties. Actually, it is not our purpose to defend this prior, but to show that it is possible with this choice to obtain a Bayesian estimator that avoids the difficulties encountered by more standard choices.

Perhaps the greatest objection to the use of this prior arises when we think of choosing σ^2 to be large to reflect diffuse beliefs about the means. Intuitively it might seem that choosing σ^2 very large is a sensible approach in such a context. When $\sigma^2 \to \infty$, however, the posterior mean converges to $||x^n||^2 + n$ and this estimator is also inconsistent. This is also the formal Bayes estimate when we use an improper flat prior on each θ_i, see de Waal (1974). Accordingly, different noninformative priors have been considered for this problem. In particular, Berger, Philippe and Robert (1998) use the reference prior approach of Berger and Bernardo (1992) and derive the Bayes estimate under squared error loss in terms of the confluent hypergeometric function. Although no proof is provided, it seems likely that this estimator is consistent. Also, one could place the noninformative improper prior $(1 + \sigma^2)^{-1}$ on σ^2 which leads to the Bayes estimator, under squared error loss, given by $||x^n||^2 - 2n + 2n/||x^n||$ and this is consistent under weak conditions. These solutions are definitely appealing, but require noninformative priors and do not resolve the issue of obtaining an appropriate Bayesian estimator when σ^2 is chosen to reflect knowledge about the θ_i values so that we have an informative, proper prior.

It is also possible to compare the frequency properties of inferences obtained by different priors. The inferences obtained by using various improper priors can be shown to have superior frequency properties when compared with those inferences obtained via the $N(0, \sigma^2)$ prior with σ^2 large. So the $N(0, \sigma^2)$ prior is a poor choice when our goal is good frequentist properties. We emphasize that we are not advocating the $N(0, \sigma^2)$ prior with σ^2 large as a representative of noninformativity, but we believe it is a perfectly sensible choice when σ^2 is chosen informatively as then uniform frequency properties are not relevant.

From a practical viewpoint, the problem discussed here is somewhat artificial. In spite of this, it has been considered by many authors as a kind of test case for various approaches to inference. Our results for this problem based on relative beliefs, together with the general results for relative beliefs discussed in Section 2, demonstrate that these inferences have distinct advantages for proper Bayesian inference.

2 Relative Surprise Inferences

In Evans (1997) the relative surprise approach to deriving inferences was introduced. Suppose the full model has parameter space Θ and we are interested in making inferences about a marginal parameter $\psi = \Psi(\theta)$ where $\Psi : \Theta \to \Upsilon$. Denote the prior measure on Θ by Π and the posterior measure after observing data x, by $\Pi(\cdot \mid x)$. Further, let π_Ψ and $\pi_\Psi(\cdot \mid x)$ denote the prior and posterior densities of ψ with respect to some support measure. Then the *relative belief ratio* of a value ψ is given by $RB(\psi) = \pi_\Psi(\psi \mid x)/\pi_\Psi(\psi)$. This ratio measures the change in belief in the value ψ being the true value of $\Psi(\theta)$ from *a priori* to *a posteriori* and so we take $RB(\psi)$ as the evidence that ψ is the true value of $\Psi(\theta)$. As discussed in Evans, Guttman and Swartz (2006), this ratio can be seen to be the limiting value of the relative belief ratio $\Pi_\Psi(B \mid x)/\Pi_\Psi(B)$ for an appropriate sequence of sets $B \downarrow \{\psi\}$ as $\Pi_\Psi(\cdot \mid x)$ is absolutely continuous with respect to the prior Π_Ψ. Also, whenever $\Pi_\Psi(\{\psi\}) = 0$, as in the problem considered in this paper, then $RB(\psi)$ is the limiting value of the sequence of Bayes factors in favor of B as $B \downarrow \{\psi\}$.

Notice that the interpretation of $RB(\psi)$ as the evidence that ψ is the true value, imposes a necessary total ordering on the possible values for ψ. For ψ_1 is preferred to ψ_2 whenever $RB(\psi_1) \geq RB(\psi_2)$ as the observed data have lead to an increase in belief for ψ_1 at least as large as that for ψ_2. As we now discuss, this total ordering determines the inferences.

The best estimate of $\Psi(\theta)$ is the value of ψ for which the evidence is greatest, called the *least relative surprise estimator* (LRSE), namely, $\psi_{\mathrm{LRSE}}(x) = \arg\sup RB(\psi)$. A γ-credible region for $\Psi(\theta)$ must take the form $C_\gamma(x) = \{\psi : RB(\psi) \geq c_\gamma(x)\}$ where $c_\gamma(x) = \inf\{k : \Pi_\Psi(RB(\psi) > k \mid T(x)) \leq \gamma\}$ as, if $RB(\psi_1) \geq RB(\psi_2)$ and $\psi_2 \in C_\gamma(x)$, then we must have $\psi_1 \in C_\gamma(x)$. Note that $C_{\gamma_1}(x) \subset C_{\gamma_2}(x)$ when $\gamma_1 \leq \gamma_2$ and $\psi_{\mathrm{LRSE}}(x) \in C_\gamma(x)$ for each γ that leads to a nonempty set. The size of $C_\gamma(x)$, for suitably chosen γ, is then a measure of the accuracy of $\psi_{\mathrm{LRSE}}(x)$ where the interpretation of size is application dependent.

To assess a hypothesis $H_0 : \Psi(\theta) = \psi_0$, the value $RB(\psi_0)$ gives the evidence as to whether H_0 is true or false. For $RB(\psi_0) > 1$ means that the probability of ψ_0 has increased by the factor $RB(\psi_0)$ from *a priori* to *a posteriori* so we have evidence in favor of H_0 and the larger $RB(\psi_0)$ is, the more evidence we have in favor of H_0. If $RB(\psi_0) < 1$, then the probability of ψ_0 has decreased by the factor $RB(\psi_0)$ from *a priori* to *a posteriori*, we have evidence against H_0 and the smaller $RB(\psi_0)$ is, the more evidence we have against H_0. Just stating $RB(\psi_0)$ as evidence, however, is only part of the story as it is not immediately clear, for example, exactly how much evidence in favor of H_0 a value such as $RB(\psi_0) = 20$ is. To assess the strength of the evidence we compute the *observed relative surprise* (*ORS*) *at* ψ_0, namely, $\Pi_\Psi(RB(\psi) \leq RB(\psi_0) \mid x)$ which is the posterior probability that the true value of $\Psi(\theta)$ has a relative belief ratio no larger than the hypothesized value. If $RB(\psi_0) < 1$, so we have evidence against H_0, and the ORS is small, then we have strong evidence against H_0 because there is a large posterior probability that the true value has a larger relative belief ratio, while a large value of the ORS would indicate only weak evidence against H_0. If $RB(\psi_0) > 1$, so we have evidence in favor of H_0, and the ORS is small, then we have only weak evidence in favor H_0 because there is a large posterior probability that the true value has a larger relative belief ratio, while a large value of the ORS would indicate strong evidence in favor of H_0. Notice that the interpretation of the ORS is dependent on whether $RB(\psi_0)$ is smaller or greater than 1 and it is not to be interpreted like a p-value. If we consider the ORS as a measure of the accuracy of the evidence, then it is interesting to note that $C_\gamma(x) = \{\psi_0 : \Pi_\Psi\left(RB(\psi) \leq RB(\psi_0) \,|\, x\right) \geq 1-\gamma\}$ and $\Pi_\Psi\left(RB(\psi) \leq RB(\psi_0) \,|\, x\right) = 1 - \inf\{\gamma : \psi_0 \in C_\gamma(x)\}$ so the measures of accuracy in estimation and hypothesis assessment are intimately related. Note also that $\psi_{\text{LRSE}}(x)$ minimizes the ORS which leads to the name of this estimator.

It is the properties of $RB(\psi_0)$ that determine whether or not it is suitable as a measure of statistical evidence. For example, in contrast to Bayesian inferences like the mean, mode or hpd regions, relative surprise inferences are invariant under smooth transformations. This follows from the fact that, if $\lambda = \Lambda(\psi)$ for some 1–1, smooth function Λ, then $RB(\lambda) = RB(\psi)$ as Jacobians cancel in the numerator and denominator. So in particular, $\lambda_{\text{LRSE}}(x) = \Lambda(\psi_{\text{LRSE}}(x))$. It is proven in Evans et al. (2006) that $C_\gamma(x)$ has an optimality property among all γ-credible regions, namely, $\Pi_\Psi(C)$ is minimized among all sets having $\Pi_\Psi\left(C \,|\, x\right) \geq \gamma$ by taking $C = C_\gamma(x)$. In Evans and Shakhatreh (2008) $C_\gamma(x)$ is shown to minimize the prior probability of covering a false value and is unbiased in the sense that this probability is bounded above by the prior probability of containing the true value. Furthermore, among all sets B satisfying $\Pi_\Psi\left(B \,|\, x\right) = \Pi_\Psi\left(C_\gamma(x) \,|\, x\right)$, the set $C_\gamma(x)$ maximizes both the relative belief ratio and the Bayes factor and this maximized value is greater than 1. In Baskurt and Evans (2013) a full discussion of the relationship of relative belief ratios to Bayes factors is provided, and various *a priori* and *a posteriori* inequalities involving $RB(\psi_0)$ are established. For example, it is always the case that $\Pi\left(RB(\Psi(\theta)) \leq RB(\psi_0) \,|\, x\right) \leq RB(\psi_0)$ so that a small value of $RB(\psi_0)$ is always strong evidence against H_0.

In Evans and Jang (2011c) it is proved that the LRSE has optimal decision-theoretic properties. When the number of possible values for ψ is finite and the loss function is given by $L(\theta, \psi) = I(\Psi(\theta) \neq \psi)/\pi_\Psi\left(\Psi(\theta)\right)$, where $I(\Psi(\theta) \neq \psi)$ is the indicator function for the event $\Psi(\theta) \neq \psi$, then $\psi_{\text{LRSE}}(x)$ is a Bayes rule. This loss function penalizes incorrect values of

ψ more severely when the true value of $\Psi(\theta)$ is in the tails of the prior. When the number of possible values for ψ is infinite, the loss function needs to be truncated to ensure a bounded loss and then $\psi_{\text{LRSE}}(x)$ is a limit of Bayes rules as the truncation parameter goes to infinity; see Evans and Jang (2011c) for details. So the LRSE also has a decision-theoretic justification and thus is well-supported as an estimator both intuitively and theoretically. Given this, it is perhaps not surprising that the LRSE performs well in the problem under discussion in this paper.

The function $\pi_\Psi(\psi \,|\, x)/\pi_\Psi(\psi)$ is sometimes referred to in the literature as an integrated likelihood as it arises as the expectation of the likelihood under the conditional prior of θ given $\Psi(\theta) = \psi$. It is worth noting, however, that, in contrast to an integrated likelihood, $RB(\psi)$ cannot be multiplied by a positive constant and retain its interpretation as a relative belief ratio. Berger, Liseo and Wolpert (1999) describe various advantages of using integrated likelihoods as opposed to other kinds of likelihoods in the context of improper priors. That the integrated likelihood also arises via the relative surprise principle provides further support for this form of the likelihood.

It is also worth noting that the estimator $\psi_{\text{LRSE}}(x)$ is completely robust to the choice of the marginal prior π_Ψ. This robustness can be viewed as an additional strength of this estimator. The measure of accuracy of this estimate, as expressed by the size of the region $C_\gamma(x)$ for some γ, however, is dependent on π_Ψ. It is possible that there is a conflict between π_Ψ and the likelihood. This is referred to as prior-data conflict and determining whether or not this exists can be assessed using the methods discussed in Evans and Moshonov (2006), see Section 7. In situations where prior-data conflict exists, we do not advocate the use of inferences based on the prior. Methods for modifying a prior to avoid prior-data conflict, when it has been detected, are presented in Evans and Jang (2011b). When no prior-data conflict exists then, although the likelihood and prior may be saying somewhat different things about the true value of ψ, these assertions are at least not in conflict.

3 The Posterior Mean

The parameter τ_n^2 is changing with n so we use the following definition of consistency here.

Definition 3.1. *A sequence of estimators $t_n(x^n)$ is consistent for τ_n^2 if $t(x^n)/n - \tau_n^2/n \xrightarrow{P} 0$ as $n \to \infty$.*

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 $P\left(\left| ||x^n||^2/n - 1 - \tau_n^2/n \right| > \epsilon\right) \le \epsilon^{-2} E(||x^n||^2/n - 1 - \tau_n^2/n)^2 = 2/n\epsilon^2 + \left(4/n\epsilon^2\right)\left(\tau_n^2/n\right)$ 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 and is proved in the Appendix.

PROPOSITION 3.1. *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.*

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 SLLN. Still it seems like a very special situation.

4 The LRSE

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

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

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

$$\begin{aligned}\frac{dg(\tau^2)}{d\tau^2} &= -1/2 + \left(\sum_{k=0}^{\infty} k \frac{\left(||x^n||^2/4\right)^k (\tau^2)^{k-1}}{k!\Gamma\left((n/2)+k\right)}\right) / \sum_{k=0}^{\infty} \frac{\left(||x^n||^2\tau^2/4\right)^k}{k!\Gamma\left((n/2)+k\right)} \\ &= -(1/2) + (1/2)(||x^n||/\tau) I_{n/2}\left(||x^n||\tau\right) I^{-1}_{(n/2)-1}\left(||x^n||\tau\right) \\ &= -(1/2) + (1/2)H_n(\tau^2) \qquad (4.1)\end{aligned}$$

and $I_p(x) = (x/2)^p \sum_{k=0}^{\infty} \left(x^2/4\right)^k / \left(k!\Gamma\left(p+k+1\right)\right)$ is the modified Bessel function of order p. Setting (4.1) equal to 0, we have that, when the LRSE is not 0, then it is a solution of $H_n(\tau^2) = 1$. The following result establishes the existence and uniqueness of the LRSE and is proved in the Appendix.

PROPOSITION 4.1. *If* $||x^n||^2/n \geq 1$*, the LRSE is the unique solution of* $H_n(\tau^2) = 1$ *and, if* $||x^n||^2/n < 1$*, the LRSE equals* 0*.*

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 3.1. So in this case we have that with high probability the LRSE is given by the unique solution to $H_n(\tau^2) = 1$.

The following results are proved in the Appendix.

PROPOSITION 4.2. *When* $\tau_n^2 = o(n^2)$ *and* τ_n^2/n *is bounded away from 0, the LRSE is consistent.*

So the LRSE is consistent under very general circumstances.

COROLLARY 4.1. *We have that, whenever* $||x^n||^2 - n < 0$*, then* $\hat{\tau}_n^2 = (||x^n||^2 - n)_+ = 0$ *and, whenever* $\hat{\tau}_n^2 > 0$*, then* $1 \leq \hat{\tau}_n^2/(||x^n||^2 - n)_+ \leq 1 + 2/n$*.*

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

We note that when τ_n^2/n converges to any nonzero value, then $\tau_n^2 = o(n^2)$ and τ_n^2/n is bounded away from 0, so $\hat{\tau}_n^2$ is consistent by Proposition 4.2. Therefore the LRSE is consistent in much greater generality than the posterior mean which is only consistent when $\tau_n^2/n \to \sigma^2$. It is also of interest to see what happens when $\tau_n^2/n \to 0$. The following result, proved in the Appendix, drops the "bounded away from 0" requirement in Proposition 4.2.

PROPOSITION 4.3. *When* $\tau_n^2/n \to 0$*, the LRSE is consistent.*

The condition $\tau_n^2/n \to 0$ means that the effects θ_i become vanishingly small as i grows, e.g., whenever τ_n^2 converges. This situation is relevant when we can conceive of only a finite number of effects being material, although we do not know which these are.

In this example the LRSE does not depend on the hyperparameter σ^2. This is not a general characteristic of relative surprise inferences for marginal parameters, namely, there is generally a dependence on hyperparameters. 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 6.

5 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 (4.1), we have $2df(\tau^2)/d\tau^2 = (n-2)/\tau^2 - \left(1+1/\sigma^2\right) + H_n(\tau^2)$. Putting $G_n(\tau^2) = (n-2)/\tau^2 + H_n(\tau^2)$ we have that the mode is a solution to $G_n(\tau^2) = \left(1+1/\sigma^2\right)$. In the Appendix we show that there is always a unique solution to $G_n(\tau^2) = \left(1+1/\sigma^2\right)$ for $\tau^2 \in [0,\infty)$. This solution is the posterior mode $\tilde{\tau}_n^2$.

In the Appendix we prove the following concerning the consistency of $\tilde{\tau}_n^2$.

PROPOSITION 5.1. *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.*

Proposition 3.1 establishes the inconsistency of the posterior mean except in very limited circumstances and so, by Proposition 5.1, 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 5.1 applies to this estimator as well.

6 Credible Intervals

For the consistency of interval estimates we use the following definition.

DEFINITION 6.1. *Intervals (a_n, b_n) and (c_n, d_n) are asymptotically equivalent if $\lim_{n\to\infty}(a_n - c_n) = 0, \lim_{n\to\infty}(b_n - d_n) = 0$ and $\lim_{n\to\infty}(b_n - a_n)/(d_n - c_n) = 1$. A credible interval for a parameter is said to be a consistent interval for the parameter if it is asymptotically equivalent to an interval that always contains the true value of the parameter.*

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 $X_n \mid \delta_n^2 \sim$ 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 $(\tau^2 - E(\tau^2 \mid x^n))/(Var(\tau^2 \mid x^n))^{1/2} \mid x^n \xrightarrow{D} N(0,1)$ where $E(\tau^2 \mid x^n) = (1+1/\sigma^2)^{-1}\left\{n + (1+1/\sigma^2)^{-1}||x^n||^2\right\}$, and $Var(\tau^2 \mid x^n) = 2(1+1/\sigma^2)^{-2}\{n + 2(1+1/\sigma^2)^{-1}||x^n||^2\}$. 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

$$\begin{aligned}&\left\{(1+1/\sigma^2)^{-1} + (1+1/\sigma^2)^{-2}(||x^n||^2/n)\right\}\\ &\pm \sqrt{2(1+1/\sigma^2)^{-2} + 4(1+1/\sigma^2)^{-3}(||x^n||^2/n)}(z_{(1+\gamma)/2}/\sqrt{n})\end{aligned} \tag{6.1}$$

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

Now consider the interval given by

$$\begin{aligned}&\left\{(1+1/\sigma^2)^{-1} + (1+1/\sigma^2)^{-2}(1+\tau_*^2)\right\}\\ &\pm \sqrt{2(1+1/\sigma^2)^{-2} + 4(1+1/\sigma^2)^{-3}(1+\tau_*^2)}(z_{(1+\gamma)/2}/\sqrt{n}).\end{aligned} \tag{6.2}$$

Comparing (6.1) and (6.2) 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 (6.1) and (6.2) are asymptotically equivalent. Now consider whether or not (6.2) contains $\tau_n^2/n \approx \tau_*^2$. The asymptotic error in the posterior mean is given by

$$\begin{aligned}&\tau_*^2 - \left\{(1+1/\sigma^2)^{-1} + (1+1/\sigma^2)^{-2}(1+\tau_*^2)\right\}\\ &= \left(1+1/\sigma^2\right)^{-2}\left(2+1/\sigma^2\right)\left(\tau_*^2/\sigma^2 - 1\right).\end{aligned} \tag{6.3}$$

Note that τ_*^2 is in (6.2) if and only if 0 is in the interval obtained by adding (6.3) to each point in (6.2). From this we see that the true value τ_*^2 is always in (6.2) when $\tau_*^2 = \sigma^2$. Now suppose that $\tau_*^2 \neq \sigma^2$. If $\tau_*^2/\sigma^2 > 1$ then the sum of (6.3) and the left-hand endpoint of (6.2) is greater than 0 for all n large enough. If $\tau_*^2/\sigma^2 < 1$ then the sum of (6.3) and the right-hand endpoint of (6.2) is less than 0 for all n large enough. Therefore, when $\tau_*^2 \neq \sigma^2$, (6.2) 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 (6.1) 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 5 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 6.1 and the simulation results reported in Evans (1997).

The following result establishes the consistency of relative surprise intervals under the condition that $\tau_n^2/n \to 0$ and is proved in the Appendix.

PROPOSITION 6.1. *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.*

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 θ_i are close to 0, then this uncertainty is largely avoided. Effectively the prior controls the precision of inferences more than n.

7 Checking for Prior-Data Conflict

We have shown that a reasonable Bayesian estimator of τ_n^2 can be obtained when using a $N(0, \sigma^2)$ prior. Still we might ask if this prior, with a specific value chosen for σ^2, 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 do 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, in this case, the sampling model is always consistent with the data and so we only check 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. As discussed in Evans and Moshonov (2006), checking for prior-data conflict then entails comparing the observed value of a minimal sufficient statistic $T(x^n)$ with its prior predictive distribution M_T to see if it is a reasonable value. This comparison leads to computing the tail probability $M_T(m_T(t) \leq m_T(T(x^n)))$ where m_T is the prior predictive density and $t \sim M_T$. It is proved in Evans and Jang (2011a) that this tail probability is assessing whether or not the true value of the parameter is in the tails of the prior. In this case, $T(x^n) = x^n$ and M_T is the $N_n(0, (1+\sigma^2)I)$ distribution. So $M_T(m_T(t) \leq m_T(T(x^n))) = P(X^2 > (1+\sigma^2)^{-1}||x^n||^2)$ where $X^2 \sim$ Chi-squared$(n, 0)$ as the prior predictive distribution of $(1+\sigma^2)^{-1}||x^n||^2$ is Chi-squared$(n, 0)$.

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 $\sigma^2 \to \infty$ then $P(X^2 > (1+\sigma^2)^{-1}||x^n||^2) \to 1$. 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 any 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 larger so that the conflict is avoided. This process of modifying the prior when a conflict is encountered is discussed further in Evans and Jang (2011b). In general, there is no reason to rule out using a $N(0, \sigma^2)$ prior in informative settings.

8 Conclusions

Discussion of the estimation problem considered in this paper has focused on the informative case. Various arguments can be advanced for noninformative priors that avoid the poor behavior, whether frequentist or Bayesian, of the standard Bayesian estimators when using a diffuse $N(0, \sigma^2)$ prior. The problem of inconsistency still remains, however, when σ^2 is chosen informatively. So we need an approach that leads to Bayesian inference procedures that have appropriate properties without requiring the incoherent behavior of modifying a proper prior. This paper has shown that the LRSE has better consistency properties than other Bayesian estimators in this challenging problem. Of course, one could choose to ignore this and require instead that the estimator have optimal properties with respect to some specific loss function such as quadratic loss. As described in Evans and Jang (2011c), however, the LRSE also has optimal decision-theoretic properties.

Acknowledgement. The authors thank the referees for a number of helpful comments.

References

ABRAMOWITZ, M. and STEGUN, I.A. (eds.) (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover, New York.

BASKURT, Z. and EVANS, M. (2013). Hypothesis assessment and inequalities for Bayes factors and relative belief ratios. *Bayesian Anal.*, **8**, 569–590.

BERGER, J.O. and BERNARDO, J. (1992). On the development of reference priors. In *Bayesian Statistics*, vol. 4 (J. Berger, J. Bernardo, P. Dawid and A.F.M. Smith, eds.). Oxford University Press, London, pp. 35–49.

BERGER, J.O., PHILIPPE, A. and ROBERT, C.P. (1998). Estimation of quadratic functions: Noninformative priors for noncentrality parameters. *Stat. Sin.*, **8**, 359–378.

BERGER, J.O., LISEO, B. and WOLPERT, R.L. (1999). Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Stat. Sci.*, **14**, 1–28.

CHOW, M.S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.*, **15**, 800–804.

DE WAAL, D.J. (1974). Bayes estimate of the noncentrality parameter in multivariate analysis. *Comm. Statist.*, **3**, 73–79.

EFRON, B. (1973). Discussion of Dawid, A.P., Stone, M. and Zidek, J.V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. R. Stat. Soc. B*, **2**, 189–233.

EVANS, M. (1997). Bayesian inference 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. *Canad. J. Statist.*, **34**, 113–129.

EVANS, M. and JANG, G.H. (2011a). A limit result for the prior predictive applied to checking for prior-data conflict. *Statist. Probab. Lett.*, **81**, 1034–1038.

EVANS, M. and JANG, G.H. (2011b). Weak informativity and the information in one prior relative to another. *Statist. Sci.*, **26**, 423–439.

EVANS, M. and JANG, G.H. (2011c). *Inferences from prior-based loss functions.* Tech. Rep. No. 1104, Dept. of Statistics, U. of Toronto.

EVANS, M. and MOSHONOV, H. (2006). Checking for prior-data conflict. *Bayesian Anal.*, **1**, 893–914.

EVANS, M. and SHAKHATREH, M. (2008). Optimal properties of some Bayesian inferences. *Electron. J. Stat.*, **2**, 1268–1280.

GELFAND, A. (1983). Estimation in noncentral distributions. *Comm. Statist. A*, **12**, 463–475.

KEMP, C.D. and KEMP, A.W. (1956). Generalized hypergeometric distributions. *J. R. Stat. Soc.*, **2**, 202–211.

KUBOKAWA, T., ROBERT, C.P. and SALEH, A.K.MD.E. (1993). Estimation of noncentrality parameters. *Canad. J. Statist.*, **21**, 45–57.

NEFF, N. and STRAWDERMAN, W.E. (1976). Estimation of the parameter of a non-central chi square distribution. *Comm. Statist.*, **A5**, 65–76.

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.

Appendix

Proof of Proposition 3.1. By the argument preceding the statement of Proposition 3.1 we have that

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

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

Proof of Proposition 4.1. We need the following properties of the modified Bessel function I_p.

Lemma A.1. *We have that for $x \geq 0$*

(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. Parts (i), (ii) and (iii) are stated in Saxena and Alam (1982) without proof. Part (i) follows as we have that $I_{p-1}(x) = (x/2)^{p-1} \sum_{k=0}^{\infty} a_k(x^2/4)^k$ 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 $n(x) = \{\sum_{k=0}^{\infty} k^2a_k(x^2/4)^k \sum_{k=0}^{\infty} a_k(x^2/4)^k - (\sum_{k=0}^{\infty} ka_k(x^2/4)^k)^2\}$ 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 (1972).

We also need the following result.

Lemma A.2. *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) $\lim_{\tau^2\to\infty}H_n(\tau^2)=0$.

Proof. Part (i) follows from Lemma A.1(ii). Also we have that $\lim_{\tau^2\to 0^+} H_n(\tau^2) = (1/2)\,||x^n||^2\left(\Gamma\left(n/2\right)/\Gamma\left(n/2+1\right)\right) = ||x^n||^2/n$ and (ii) is established. From Lemma A.1(v), we have that $I_{n/2}(||x^n||\tau) \sim e^{||x||\tau}/(2\pi||x^n||\tau)^{1/2}$ as $\tau^2 \to \infty$ and so $\lim_{\tau^2\to\infty} H_n(\tau^2) = \lim_{\tau^2\to\infty}(||x^n||/\tau) = 0$.

The proof of Proposition 4.1 then proceeds as follows. If $||x^n||^2/n < 1$ then, by Lemma A.2, $H_n(\tau^2) < 1$. Therefore by (4.1), $dg(\tau^2)/d\tau^2 < 0$ 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 A.2 there is a unique solution to $H_n(\tau^2) = 1$ and (4.1) 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.

PROOF OF PROPOSITION 4.2, COROLLARY 4.1 AND PROPOSITION 4.3. 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 $P\left(\left|\hat{\tau}_n^2/n - \tau_n^2/n\right| \leq \epsilon\right) \geq P\left(\left\{\left|\hat{\tau}_n^2/n - \tau_n^2/n\right| \leq \epsilon\right\} \cap C_n\right) = P(\left|\hat{\tau}_n^2/n - \tau_n^2/n\right| \leq \epsilon \,|\, C_n)P(C_n) \to 1$ 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

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

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

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

and so

$$\hat{\tau}_n^2 \geq ||x^n||^2 - n. \tag{A.3}$$

Applying Lemma A.1(vi) to the denominator in (A.2) we have

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

and rearranging this gives

$$||x^n||\tau I_{(n/2)+1}(||x^n||\tau)/I_{n/2}(||x^n||\tau) = ||x^n||^2 - n. \tag{A.4}$$

Now applying Lemma A.1(vi) to the numerator in (A.2), apply Lemma A.1(iii) and Lemma A.1(iv) again to obtain

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

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

$$\left(||x^n||^2 - n\right) \leq \hat{\tau}_n^2 \leq \left(||x^n||^2 - n\right)(n+2)/n. \tag{A.5}$$

As $P\left(\left|||x^n||^2/n - 1 - \tau_n^2/n\right| \leq \epsilon\right) = P\left(\left|||x^n||^2/n - 1 - \tau_n^2/n\right| \leq \epsilon \,|\, C_n\right) P(C_n)$, and $P(C_n) \to 1$, then (A.1) implies that $P\left(\left|||x^n||^2/n - 1 - \tau_n^2/n\right| \leq \epsilon \,|\, C_n\right) \to 1$. Similarly, $P\left(\left|\left(||x^n||^2/n\right)(n+2)/n - (n+2)/n - \tau_n^2/n\right| \leq \epsilon \,|\, C_n\right) \to 1$ and from (A.5) we conclude that $P\left(\left|\hat{\tau}_n^2/n - \tau_n^2/n\right| \leq \epsilon \,|\, C_n\right) \to 1$. This proves Proposition 4.2.

The proof of Corollary 1 then proceeds as follows. When $\hat{\tau}_n^2 = 0$ 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 $\hat{\tau}_n^2 > 0$, then (A.5) implies $||x^n||^2 - n > 0$, and so $(||x^n||^2 - n)_+ = ||x^n||^2 - n$. Then (A.5) implies that $1 \leq \hat{\tau}_n^2/(||x^n||^2 - n)_+ \leq 1 + 2/n$.

The proof of Proposition 4.3 proceeds as follows. We have that $||x^n||^2/n \xrightarrow{P} 1$ by (A.1). If $||x^n||^2/n \leq 1$, 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 4.3, we must have that (A.5) holds. Then, $P\left(\hat{\tau}_n^2/n > \epsilon\right) \leq P(\left(||x^n||^2 - n\right)(n+2)/n^2 > \epsilon) = P(||x^n||^2/n > 1 + \epsilon/(1+2/n)) \leq P((||x^n||^2/n > 1 + \epsilon/3)) \to 0$ as $n \to \infty$ establishing the result.

Proof of Proposition 5.1. The following properties of $G_n(\tau^2) = (n-2)/\tau^2 + H_n(\tau^2)$ follow immediately from Lemma A.2.

LEMMA A.3. *The function G_n satisfies*

(i) $G_n(\tau^2)$ *is a strictly decreasing function,*

(ii) $\lim_{\tau^2 \to 0^+} G_n(\tau^2) = \infty$, *and*

(iii) $\lim_{\tau^2 \to \infty} G_n(\tau^2) = 0$.

Note that Lemma A.3 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 proceed to the proof of Proposition 5.1. 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

$$l(\tau^2) = ||x^n||\tau I_{n/2}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right). \tag{A.6}$$

Applying Lemma A.1(vi) to the numerator in (A.6), we obtain $l(\tau^2) = \left(||x^n||^2\tau^2/n\right) \times \left\{1 - I_{(n/2)+1}\left(||x^n||\tau\right)/I_{(n/2)-1}\left(||x^n||\tau\right)\right\}$. Then using Lemma A.1(iv) and (A.6) we have that

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

and rearranging this conclude that the mode satisfies

$$\left(1 + 1/\sigma^2\right)^2\tau^4 - \{||x^n||^2 + \left(1 + 1/\sigma^2\right)(n-4)\}\tau^2 - 2(n-2) \geq 0. \tag{A.7}$$

The roots of the quadratic in τ^2 in (A.7) are given by $r_1(n) \pm r_2(n)$ where

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

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

$$r_1(n) + r_2(n) \leq \tilde{\tau}_n^2. \tag{A.8}$$

Now observe that

$$|r_1(n)/n - r_2(n)/n| = \left|\sqrt{r_1^2(n)/n^2} - \sqrt{r_1^2(n)/n^2 + 2(n-2)/n^2\left(1 + 1/\sigma^2\right)^2}\right|$$
$$= \frac{2(n-2)/n^2\left(1 + 1/\sigma^2\right)^2}{\sqrt{r_1^2(n)/n^2} + \sqrt{r_1^2(n)/n^2 + 2(n-2)/n^2\left(1 + 1/\sigma^2\right)^2}}$$

$$\leq \sqrt{2(n-2)/n^2 \left(1+1/\sigma^2\right)^2} \to 0$$

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 $m(n)/n - 2r_1(n)/n = (4/n)\,(1+1/\sigma^2)^{-1} \to 0$ which implies that $m(n)/n - (r_1(n)+r_2(n)) \xrightarrow{P} 0$.

Now apply Lemma A.1(vi) to the denominator on the right-hand side of (A.6) 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

$$(||x^n||\tau)I_{(n/2)+1}\left(||x^n||\tau\right)/I_{n/2}\left(||x^n||\tau\right) = (||x^n||^2\tau^2/l(\tau^2)) - n. \tag{A.9}$$

Just as in the proof of Proposition 4.2 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 (A.6)

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

Now use (A.9) to obtain $l(\tau^2) \leq \{||x||^2\tau^2 - nl(\tau^2)\}\{(n+2)+l(\tau^2)\}/(||x^n||^2\tau^2)$ or

$$(||x^n||^2\tau^2)l(\tau^2) - \{||x||^2\tau^2 - nl(\tau^2)\}\{(n+2)+l(\tau^2)\} \leq 0 \tag{A.10}$$

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 (A.10) implies that

$$\tilde{\tau}_n^2 \leq s_1(n) + s_2(n) \tag{A.11}$$

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

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

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 (A.1) imply $||x^n||^2/n^2 \xrightarrow{P} 0$.

By (A.8) and (A.11)

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

and this establishes the result.

PROOF OF PROPOSITION 6.1. Note that the developments in Section 2 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 4.3, $\hat{\tau}_n^2/n$ 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 $A_n(z) = \{x^n : (l_n(x^n) - E(\tau^2 \,|\, x^n))/(Var(\tau^2 \,|\, x^n))^{1/2} \le z\}$, 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

$$\begin{aligned} P(\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon) \;=\; & P(A_n(z) \cap \{\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon\}) \\ & + P(A_n^c(z) \cap \{\Pi([0, l_n(x^n)] \,|\, x^n) > \epsilon\}). \end{aligned} \tag{A.12}$$

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

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

and so the first term in (A.12) satisfies

$$
\begin{aligned}
& P(A_n(z) \cap \{\Pi([0, l_n(x^n)] \mid x^n) > \epsilon\}) \\
& \leq P\left(A_n(z) \cap \left\{\Pi\left(\frac{\tau^2 - E(\tau^2 \mid x^n)}{(Var(\tau^2 \mid x^n))^{1/2}} \leq z \,\middle|\, x^n\right) > \epsilon\right\}\right) \\
& \leq P\left(\left\{\Pi\left(\frac{\tau^2 - E(\tau^2 \mid x^n)}{(Var(\tau^2 \mid x^n))^{1/2}} \leq z \,\middle|\, x^n\right) > \epsilon\right\}\right). \qquad \text{(A.13)}
\end{aligned}
$$

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

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

for every data sequence x^n and so this convergence is also in probability with respect to P. This implies that (A.13) converges to 0. Further, the second term in (A.12) is bounded above by $P(A_n^c(z))$ which converges to 0. Accordingly, we have proved that $\Pi([0, l_n(x^n)] \mid x^n) \xrightarrow{P} 0$. Also, we always have that $\gamma = \Pi([l_n(x^n), r_n(x^n)] \mid x^n) = \Pi([0, r_n(x^n)] \mid x^n) - \Pi([0, l_n(x^n)] \mid x^n)$ and so $\Pi([0, r_n(x^n)] \mid x^n) \xrightarrow{P} \gamma$.

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

$$
\begin{aligned}
P(B_n) \;=\; & P(B_n \cap \{\Pi([0, r_n(x^n)] \mid x^n) > \gamma + \epsilon'/2\}) \\
& + P(B_n \cap \{\Pi([0, r_n(x^n)] \mid x^n) \leq \gamma + \epsilon'/2\}). \qquad \text{(A.14)}
\end{aligned}
$$

The first term in (A.14) is bounded above by $P(\Pi([0, r_n(x^n)] \mid x^n) > \gamma + \epsilon'/2)$ and this converges to 0. For the second term in (A.14) we have that

$$
\begin{aligned}
& P(B_n \cap \{\Pi([0, r_n(x^n)] \mid x^n) \leq \gamma + \epsilon'/2\}) \\
& = P\left(B_n \cap \left\{\begin{array}{c} \Pi\left(\frac{\tau^2 - E(\tau^2 \mid x^n)}{(Var(\tau^2 \mid x^n))^{1/2}} \leq \frac{r_n(x^n) - E(\tau^2 \mid x^n)}{(Var(\tau^2 \mid x^n))^{1/2}} \,\middle|\, x^n\right) \\ \leq \gamma + \epsilon'/2 \end{array}\right\}\right) \\
& \leq P\left(B_n \cap \left\{\Pi\left(\frac{\tau^2 - E(\tau^2 \mid x^n)}{(Var(\tau^2 \mid x^n))^{1/2}} \leq z_\gamma + \epsilon \mid x^n\right) \leq \gamma + \epsilon'/2\right\}\right) \qquad \text{(A.15)}
\end{aligned}
$$

and

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

implies that (A.15) converges to 0. Similarly, if we set $C_n = \{x^n : (r_n(x^n) - E(\tau^2 \mid x^n))/(Var(\tau^2 \mid x^n))^{1/2} < z_\gamma - \epsilon\}$ we get that $P(C_n) \to 0$ and so $(r_n(x^n) - E(\tau^2 \mid x^n))/(Var(\tau^2 \mid x^n))^{1/2} \xrightarrow{P} z_\gamma$. This completes the proof.

Michael Evans
Department of Statistics
University of Toronto
Toronto, ON M5S 3G3, Canada
E-mail: mevans@utstat.utoronto.ca

Mohammed Shakhatreh
Department of Mathematics
and Statistics, Jordan University
of Science and Technology
P.O. Box 3030, Irbid 22110, Jordan
E-mail: mkshakhatreh6@just.edu.jo

Paper received: 21 May 2012; revised: 20 August 2013.