Metrika
https://doi.org/10.1007/s00184-024-00961-5

Estimating the suspected larger of two normal means

Courtney Drew[1] · Éric Marchand[2]

Received: 2 December 2022 / Accepted: 5 March 2024

Abstract
For X_1, X_2 independently and normally distributed with means θ_1 and θ_2, variances σ_1^2 and σ_2^2, we consider Bayesian inference about θ_1 with the difference $\theta_1 - \theta_2$ being lower-bounded by an uncertain m. We obtain a class of minimax Bayes estimators of θ_1, based on a posterior distribution for $(\theta_1, \theta_2)^\top$ taking values on $\mathbb{R}^2$, which dominate the unrestricted MLE under squared error loss for $\theta_1 - \theta_2 \geq 0$. We also construct and study an ad hoc credible set for θ_1 with approximate credibility $1 - \alpha$ and provide numerical evidence of its frequentist coverage probability closely matching the nominal credibility level. A spending function is incorporated which further increases the coverage.

Keywords Bayes estimator · Hierarchical prior · Point estimation · Interval estimation · Skew-normal · Additional information · Uncertain constraint

Mathematics Subject Classification 62F15 · 62F30 · 62F10 · 62C20

1 Introduction

It has long been known, for a bivariate normal model with X_1, X_2 independently distributed with means θ_1 and θ_2, and known variances σ_1^2 and σ_2^2, that the Bayes estimator of θ_1 with respect to the uniform prior on $\theta_1 \geq \theta_2$ dominates the benchmark minimax estimator X_1 when $\theta_1 \geq \theta_2$ under squared error loss (Cohen and Sackrowitz 1970). However, there are situations where one would not expect this bound to hold exactly, and one could envisage introducing uncertainty in the parametric bound. This has been previously proposed (see O'Hagan and Leonard 1976 where uncertainty

✉ Éric Marchand
eric.marchand@usherbrooke.ca

Courtney Drew
Courtney.Drew@statcan.gc.ca

1 Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, ON K1A 0T6, Canada

2 Département de Mathématiques, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada

is expressed through a hierarchical prior, as well as Liseo and Loperfido 2003 for uncertain linear restrictions) and allows for a more flexible and encompassing model, where the data is allowed to contradict the believed parametric constraint. Moreover, with such a model, one has the ability to take into account the degree of prior belief in the constraint. Despite the earlier work, little is known about the frequentist risk performance of associated Bayes point estimators or Bayes credible sets.

Here, we consider Bayesian inference about θ_1 for the two-sample normal problem with hierarchical prior density given by $\pi(\theta_1, \theta_2 \,|\, m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2)$ with $m \sim N(0, \sigma_m^2)$, and study the frequentist performance of (generalized) Bayesian point estimators and credible sets. We show that the Bayes estimator of θ_1 dominates X_1, and is hence minimax, under squared error loss for $\theta_1 - \theta_2 \geq 0$ and all choices of $\sigma_m^2 > 0$. We make use of the so-called rotation technique (e.g., Blumenthal and Cohen 1968a), and a one-sample minimax finding by Marchand and Nicoleris (2019) set in the context of a single normal mean with an uncertain lower bound. The proposed Bayesian estimators stem from posterior densities for $\theta = (\theta_1, \theta_2)^\top$ that take values on $\mathbb{R}^2$, but still pass the test of minimaxity for estimating θ_1 when evaluated on the restricted parameter space $\theta_1 \geq \theta_2$. In this sense, they are more flexible and desirable in the context of constraint uncertainty than their counterpart estimator when $\sigma_m^2 = 0$, for which the posterior density is concentrated on $\theta_1 \geq \theta_2$. The finding adds to known analyses for $\sigma_m^2 = 0$ carried out by Cohen and Sackrowitz (1970), van Eeden and Zidek (2002), and Kumar and Sharma (1988), among others.
The attractive performance of the proposed point estimators of the suspected larger of the two means, θ_1, leads to interest in Bayes credible sets, and to the investigation of the extent to which one can capitalize on this additional information. We namely focus on the performance of such credible sets as measured by frequentist coverage probability. Typically, Bayesian credible sets are far from guaranteeing matching coverage probability and are not designed to do so. Exceptions lie in location and scale models without parametric restrictions and non-informative priors. Even so, in such problems, in the face of a parametric restriction $\theta \in C$, the truncation of such non-informative priors on C perturbs probability matching, with both higher coverage and lower coverage than credibility occurring (e.g., Mandelkern 2002; Marchand and Strawderman 2006). We point out that there has been much work on evaluating Bayesian posterior densities and estimates with parametric restrictions, notably for ordered parameters with or without nuisance parameters (e.g., Gelfand et al. 1992; Madi et al. 2000).

We introduce below an *ad hoc* Bayes credible set with approximate $1-\alpha$ credibility (based again on the prior $\pi(\theta \,|\, m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2)$ with $m \sim N(0, \sigma_m^2)$), and study its frequentist coverage probability with evidence of very good matching to the nominal credibility $1 - \alpha$. Numerical evidence of the remarkable proximity between the actual and nomimal credibilities is also provided. We furthermore explore how the performance is affected by the choice of the hyperparameter σ_m, ranging from the case of a certain constraint, i.e., $\sigma_m = 0$, to the case of no useful information provided by X_2 when $\sigma_m \to \infty$.

For a given posterior distribution, there is no single definitive choice of a Bayes credible set and such a choice can be impactful in terms of frequentist coverage. Namely, as illustrated by Marchand and Strawderman (2013), as well as Ghashim et al. (2016), the characterization of Bayes credible sets through a spending function merits to be considered. Hence, the analysis and illustrations presented here involve a spending function, the choice of which is guided.

The paper is organized as follows. After having extracted and interpreted some useful properties of the posterior distributions in Sect. 2.1, which relate to extended skew-normal densities, the dominance and minimax results are presented and commented on in Sect. 2.2. Section 3 deals with proposed credible sets for θ_1, focussing mostly on their frequentist coverage probability. The findings are commented on at length and illustrated with several figures. Section 3.2 expands on modifications which make use of the concept of a spending function. A summary and further research questions are presented in Sect. 4. Finally, we mention that the developments in this paper also appear in the M.Sc. thesis (Drew 2021) of Courtney Drew.

2 Bayesian inference and minimax point estimators

2.1 Posterior analysis

We consider the following model for $X = (X_1, X_2)^T$ and hierarchical prior:

$$\begin{aligned} &X_1 \sim N(\theta_1, \sigma_1^2),\, X_2 \sim N(\theta_2, \sigma_2^2);\\ &\pi(\theta_1, \theta_2|m) \;=\; \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2)\,,\; m \sim N(0, \sigma_m^2), \end{aligned} \tag{1}$$

where X_1 and X_2 are independently distributed and $\sigma_1, \sigma_2, \sigma_m > 0$ are known. This corresponds to a situation where the difference of parameters $\theta_1 - \theta_2$ is bounded below by m, with uncertainty on m. We denote throughout ϕ and Φ as the standard normal pdf and cdf respectively. An alternative and equivalent representation of the prior in (1) is readily obtained by integrating out m yielding the improper density $\pi(\theta_1, \theta_2) = \Phi(\frac{\theta_1 - \theta_2}{\sigma_m})$.

Remark 2.1 **(a)** The situation given by (1) also covers the case of a parametric bound of the form $\theta_1 - c\,\theta_2 \geq m$, with $c \neq 0$. Setting $X_1' = X_1$, $X_2' = cX_2$, the constraint becomes re-expressible as $\mu_1 - \mu_2 \geq m$ with $X_1' \sim N(\mu_1, \sigma_1^2)$ and $X_2' \sim N(\mu_2 = c\theta_2, c^2\sigma_2^2)$.

(b) Analysis for (1) yields applications for correlated variables, specifically for $W = (W_1, W_2)^\top \sim N_2(\xi, \Sigma)$ with $\xi_1 - \xi_2 \geq m$, correlation coefficient $\rho = \rho(W_1, W_2) \in (-1, 1)$, such that $\lambda = \rho\sigma(W_1)/\sigma(W_2) \neq 1$. This is achieved by setting $X_1 = W_1 - \lambda W_2$, $X_2 = W_2$ whereupon part **(a)** applies with $\theta_1 = \xi_1 - \lambda\xi_2$, $\theta_2 = \xi_2$, $c = (1 - \lambda)$, $\sigma_1^2 = \mathbb{V}(W_1)(1 - \rho^2)$, and $\sigma_2^2 = \mathbb{V}(W_2)$.

Remark 2.2 There exist many instances with summary statistics well modelled by normal observables such as in (1). Common occurrences arise through sufficiency or asymptotically justified approximations. An example emerges in a basic linear model with $W \sim N_n(Z^\top\beta, \sigma^2 I_n)$ with $Z(n \times p)$ of full rank p, the least squares $\hat\beta = (\hat\beta_1, \ldots, \hat\beta_p)^\top = (Z^\top Z)^{-1} Z^\top W$, $X_1 = \hat\beta_1$ and $X_2 = \hat\beta_2$, where it is suspected that $\beta_1 \geq \beta_2$. In such cases, with the link presented in part **(b)** of Remark 2.1, analysis for (1) applies whether $\hat\beta_1$ and $\hat\beta_2$ are correlated or not.

The following known result is useful in analyzing the posterior density in (1).

Lemma 2.3 *Let* $Z \sim N(0,1)$ *and* $\nu, \varepsilon \in \mathbb{R}$. *Then* $\mathbb{E}\big[\Phi(\nu(Z+\varepsilon))\big] = \Phi\left(\frac{\nu\,\varepsilon}{\sqrt{1+\nu^2}}\right)$.

Proof Let $T \sim N(0,1)$ be independent of Z. Then, we can write $\mathbb{E}\big[\Phi(\nu(Z+\varepsilon))\big] = \mathbb{P}\big(T \leq \nu(Z+\varepsilon)\big) = \Phi\left(\frac{\nu\,\varepsilon}{\sqrt{1+\nu^2}}\right)$ since $T - \nu Z \sim N(0, 1+\nu^2)$. □

Theorem 2.4 *Under the model and prior given by (1), setting* $d = x_1 - x_2$, *the marginal posterior density of* $U = \frac{\theta_1 - x_1}{\sigma_1}$ *is given by*

$$\pi(u|x) = \frac{\phi(u)\,\Phi\left(\frac{\sigma_1 u + d}{\sqrt{\sigma_2^2+\sigma_m^2}}\right)}{\Phi\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)}. \tag{2}$$

Proof This follows from writing the marginal posterior density of θ_1 as

$$\pi(\theta_1|x) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_1-m} f(x|\theta)\,\pi(\theta|m)\,\pi(m)\,d\theta_2\,dm}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_1-m} f(x|\theta)\,\pi(\theta|m)\,\pi(m)\,d\theta_2\,dm\,d\theta_1},$$

where

$$\begin{aligned} & f(x|\theta)\,\pi(\theta|m)\,\pi(m) \\ & = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2\sigma_1^2}(x_1-\theta_1)^2} e^{-\frac{1}{2\sigma_2^2}(x_2-\theta_2)^2} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{m^2}{2\sigma_m^2}} \mathbb{1}_{[m,\infty)}(\theta_1-\theta_2), \end{aligned}$$

then using Lemma 2.3 to evaluate the integrals, and changing variables from θ_1 to U. □

One recognizes the posterior density in (2) as a skew-normal density of the form $\phi(u)\frac{\Phi(\alpha_1 u+\alpha_2)}{\Phi\left(\alpha_2/\sqrt{1+\alpha_1^2}\right)}$; $\alpha_1, \alpha_2 \in \mathbb{R}$ (e.g., Azzalini 1985; Arnold and Beaver 2002). Note that the density in (2) also holds for $\sigma_m = 0$. We next link properties of such extended skew-normal distributions to the posterior distribution (2).

Lemma 2.5 *Under the context of Theorem 2.4, the posterior moment generating function, expectation and variance of* U *are given respectively by*

$$\begin{aligned} M_{U|x}(t) &= \frac{e^{\frac{t^2}{2}}}{\Phi(d')}\Phi\left(t\sigma' + d'\right), \quad \mathbb{E}(U|x) = \sigma'\,R\left(d'\right), \\ \mathbb{V}(U|x) &= 1 - \sigma'^2\,d'\,R\left(d'\right) - \sigma'^2\,R^2\left(d'\right), \end{aligned}$$

with $\sigma' = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}$, $d' = \frac{x_1-x_2}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}$, *and where* $R(t) = \frac{\phi(t)}{\Phi(t)}$ *is the reverse Mill's ratio.*

Proof The moment generating function is readily computed by making a change of variables $u' = u - t$ and using Lemma 2.3. The posterior expectation and variance of U follow by straightforward calculations. $\square$

In Sect. 3, we construct an ad hoc credible set for θ_1 based on its posterior expectation and variance. It is therefore of interest to study the properties of these quantities, which in turn follow from well-known properties of the reverse Mill's ratio.

Lemma 2.6 *In the setting of Theorem 2.4, the following properties of* $\mathbb{E}(U|x)$ *and* $\mathbb{V}(U|x)$ *hold for* $d = x_1 - x_2$*:*

(a) $\mathbb{E}(U|x)$ *is a decreasing function of d with* $\lim_{d\to\infty} \mathbb{E}(U|x) = 0$, $\lim_{d\to-\infty} \mathbb{E}(U|x) = +\infty$ *and* $\lim_{d\to-\infty} \frac{\mathbb{E}(U|x)}{d} = -\frac{\sigma_1}{\sigma_1^2+\sigma_2^2+\sigma_m^2}$*;*

(b) $\mathbb{V}(U|x)$ *is an increasing function of d with* $\lim_{d\to\infty} \mathbb{V}(U|x) = 1$ *and* $\lim_{d\to-\infty} \mathbb{V}(U|x) = 1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2+\sigma_m^2}$*;*

(c) $\mathbb{E}(U|x)$ *is decreasing in* σ_m^2 *when* $d < 0$*, and* $\mathbb{V}(U|x)$ *is increasing in* σ_m^2 *when* $d < 0$.

Proof These results follow from properties of the reverse Mill's ratio, in particular $\lim_{t\to\infty} R(t) = 0$, $\lim_{t\to-\infty} R(t) = \infty$, $\lim_{t\to-\infty} R(t)\big(t + R(t)\big) = 1$, $\lim_{t\to\infty} tR(t) = 0$ and $\lim_{t\to-\infty} \frac{R(t)}{t} = -1$, as well as the fact that $R(t)$ is a decreasing function of t and $R'(t) = -R(t)\big(t + R(t)\big)$. $\square$

Remark 2.7 The case $\sigma_m = 0$, i.e., no uncertainty on the restriction $\theta_1 \geq \theta_2$, warrants particular attention. One recovers results for this degenerate case in literature, notably in Cohen and Sackrowitz (1970) and Blumenthal and Cohen (1968b). Moreover, the case $\sigma_m \to \infty$ corresponds to an absence of additional information. It is useful to consider heuristics related to these limiting cases in order to gain additional understanding.

(A) If $x_1 \gg x_2$, then $d = x_1 - x_2$ is large and, since $\theta_1 \geq \theta_2$ given that $\sigma_m = 0$, x_2 provides very little additional information. We would therefore expect to obtain results similar to those in the limiting case with information on x_1 only. This is indeed the case, since we would expect a $N(x_1, \sigma_1^2)$ posterior for θ_1, which matches the limiting density of U in (2) when $d \to \infty$.

(B) In the opposite situation where $\sigma_m = 0$ but $d \ll 0$, we have data which appears to contradict the model. Assuming the model is still correct, posterior belief would be concentrated on the boundary $\theta_1 = \theta_2$. This suggests the benchmark model

$$X_i|\theta_1 \sim N(\theta_1, \sigma_i^2) \text{ independent.}$$

For the flat prior $\pi(\theta_1) = 1$, the posterior distribution of θ_1 becomes

$$\theta_1|x \sim N\left(\frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right),$$

which for $U = \frac{\theta_1 - x_1}{\sigma_1}$ and very small d, yields the approximations:

$$\mathbb{E}\left(\frac{U}{d}|x\right) \approx -\frac{\sigma_1}{\sigma_1^2 + \sigma_2^2} \text{ and } \mathbb{V}(U|x) \approx \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

which match the limiting values as $d \to -\infty$ given in Lemma 2.6 (taking $\sigma_m = 0$).

2.2 Point estimation

This section concerns itself with the efficiency of point estimators of θ_1 for model (1). We obtain a class of Bayesian estimators that dominate X_1. From Cohen and Sackrowitz (1970), it is known that X_1 is minimax for $\theta_1 \geq \theta_2$, which renders our class of estimators also minimax. Consider the problem of estimating θ_1 under squared error loss $L(\theta, d) = (d - \theta_1)^2$ with X distributed according to model (1) and with the additional prior information $\theta_1 - \theta_2 \in A \subset \mathbb{R}$. As reviewed by Marchand and Strawderman (2004), it is pertinent to consider the class of estimators:

$$\mathcal{C}_1 = \left\{ \delta_\phi(X) = Y_2 + \phi(Y_1) \text{ where } Y_1 = \frac{X_1 - X_2}{1+\tau}, Y_2 = \frac{\tau X_1 + X_2}{1+\tau}, \text{ and } \tau = \frac{\sigma_2^2}{\sigma_1^2} \right\}. \tag{3}$$

Of particular interest is the choice $\delta_{\phi_0}(X) = X_1$, i.e., the MLE of θ_1 without parametric restrictions, obtained by taking $\phi(Y_1) = Y_1$. Under model (1), Y_1 and Y_2 are independently distributed with $Y_1 \sim N(\mu_1, \sigma_{Y_1}^2)$ and $Y_2 \sim N(\mu_2, \sigma_{Y_2}^2)$, where $\mu_1 = \frac{\theta_1 - \theta_2}{1+\tau}$, $\sigma_{Y_1}^2 = \frac{\sigma_1^2}{1+\tau}$, $\mu_2 = \frac{\tau\theta_1 + \theta_2}{1+\tau}$ and $\sigma_{Y_2}^2 = \frac{\tau\sigma_1^2}{1+\tau}$. Furthermore, the mean squared error of the estimator $\delta_\phi(X)$ reduces to

$$R(\theta, \delta_\phi(X)) = \mathbb{E}_\theta\left[(Y_2 + \phi(Y_1) - \theta_1)^2\right] = \mathbb{E}_\theta\left[(Y_2 - \mu_2)^2\right] + \mathbb{E}_\theta\left[(\phi(Y_1) - \mu_1)^2\right].$$

The efficiency of the estimator $\delta_\phi(X)$ in estimating θ_1 is therefore reliant on that of the estimator $\phi(Y_1)$ in estimating μ_1.

Lemma 2.8 *For estimating θ_1 in the context of model (1) under squared error loss $L(\theta, d) = (d - \theta_1)^2$, with prior additional information $\theta_1 - \theta_2 \in A \subset \mathbb{R}$, the estimator $\delta_{\phi_1}(X)$ dominates $\delta_{\phi_0}(X)$ if and only if $\phi_1(Y_1)$ dominates $\phi_0(Y_1)$ in the problem of estimating $\mu_1 \in \mathcal{C} = \{y : (1+\tau)y \in A\}$.*

We now use a recent result from Marchand and Nicoleris (2019) which gives a class of minimax Bayes estimators for a normal mean suspected to be positive.

Lemma 2.9 (Marchand and Nicoleris 2019) For $X \sim N(\epsilon, \sigma^2)$, squared error loss $L(\epsilon, d) = (d - \epsilon)^2$ and parametric restriction $\epsilon \geq 0$, estimators $\delta_c(X) =$

$X + c\sigma R\left(\frac{cX}{\sigma}\right)$, $c \in (0, 1]$, dominate $\delta_0(X) = X$. Moreover, this class of estimators contains Bayes point estimators of ϵ under the hierarchical prior density $\pi(\epsilon \mid m) = \mathbb{1}_{[m,\infty)}(\epsilon)$ with $m \sim N(0, \sigma_m^2)$, namely δ_c, with $c = \frac{\sigma}{\sqrt{\sigma^2+\sigma_m^2}}$.

Combining Lemma 2.8 and Lemma 2.9, one obtains the following result.

Theorem 2.10 *Let X be distributed according to model (1), $\tau = \frac{\sigma_2^2}{\sigma_1^2}$, with squared error loss for estimating θ_1, $L(\theta, d) = (d - \theta_1)^2$. Then under the additional information $\theta_1 - \theta_2 \geq 0$, estimators of the form*

$$\delta_{\phi_c}(X) = X_1 + \frac{c\,\sigma_1}{\sqrt{1+\tau}} R\left(\frac{c\,(X_1 - X_2)}{\sigma_1\sqrt{1+\tau}}\right) \tag{4}$$

dominate X_1, and are hence minimax, for $c \in (0, 1]$. Furthermore, the choice $c = \frac{\sqrt{1+\tau}}{\sqrt{1+\tau+\frac{\sigma_m^2}{\sigma_1^2}}}$ coincides with the Bayes estimator for θ_1 under the prior given in (1); that is,

$$\delta_{\pi_{\sigma_m}}(X) = \mathbb{E}[\theta_1|X] = X_1 + \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}} R\left(\frac{X_1 - X_2}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right). \tag{5}$$

Proof Under the setting of (3), Lemma 2.9 asserts that estimators of the form

$$\delta_c(Y_1) = Y_1 + c\sigma_{Y_1} R\left(\frac{cY_1}{\sigma_{Y_1}}\right) = \frac{X_1 - X_2}{1+\tau} + \frac{c\,\sigma_1}{\sqrt{1+\tau}} R\left(\frac{c\,(X_1 - X_2)}{\sigma_1\sqrt{1+\tau}}\right) \tag{6}$$

dominate $\delta_0(Y_1) = Y_1$ for $c \in (0, 1]$. Thus, with $\phi_0(Y_1) = Y_1$ and correspondingly $\delta_{\phi_0}(X) = Y_2 + Y_1 = X_1$, Lemma 2.8 yields (4) as a class of estimators which dominate X_1 for $c \in (0, 1]$. □

Theorem 2.10 provides a class of Bayesian estimators that dominate X_1 and are minimax for $\theta_1 \geq \theta_2$. As for the previously known result when $\sigma_m = 0$, the estimators $\delta_{\pi_{\sigma_m}}(X)$ incorporate the sample information X_2 but, in contrast, do not arise from a prior (or posterior) density for θ concentrated on $\theta_1 \geq \theta_2$. Expressed otherwise, choices with $\sigma_m > 0$ allow more flexibility for the data to contradict such a constraint and for it to be better reflected in the posterior distribution determination. Despite this accommodation, the estimators $\delta_{\pi_{\sigma_m}}(X)$ for $\sigma_m > 0$ still remain minimax for $\theta_1 \geq \theta_2$ and will have less inflated risk than $\delta_{\pi_0}(X)$ for parameter values of θ such that $\theta_1 < \theta_2$. The value of σ_m relates to the degree of confidence for which $\theta_1 - \theta_2 \geq m$ and impacts the corresponding risk accordingly. Several of the frequentist risk features above will be paralleled by the frequentist coverage analysis of Bayes credible sets, which is the object of study of Sect. 3. Finally, questions of minimaxity and admissibility, including simultaneous estimation of $\theta = (\theta_1, \theta_2)^\top$, are addressed in Drew (2021).

3 Bayes credible sets

Having evaluated the posterior distribution of θ_1 under model and prior (1), we now turn to the construction of a Bayesian credible set for θ_1 and the study of its frequentist coverage probability and length. One objective is to determine the effect of the additional information on the credible sets, notably by considering the length of the intervals, as well as their frequentist coverage probability and credibility. Naturally, one may strive to obtain a satisfactory compromise between a short interval and good coverage. While there exist several types of credible sets; one thinks of highest posterior density (HPD) or equal-tails for example; we focus on an ad hoc interval with approximate credibility $1-\alpha$ due to its ease of computation (i.e., explicit endpoints) and interpretation, which also presents the potential for further analytical determination of frequentist coverage probability. In Sect. 3.1, the ad hoc credible set studied is of a standard form $\mathbb{E}[\theta|x] \pm z_{\alpha/2}\,\sigma(\theta|x)$ (e.g., Berger 1985). In Sect. 3.2, we propose and study a modification based on the idea of a "spending function" (e.g., Marchand and Strawderman 2013) that shifts the above credible set towards lower values.

3.1 An ad hoc credible set

The Bayes credible set studied here is given by Definition 3.1.

Definition 3.1 Let $\mathbb{E}(U|x)$ and $\mathbb{V}(U|x)$ denote respectively the posterior expectation and variance of U given by Lemma 2.5. The ad hoc Bayes credible interval for θ_1 (i.e., for $\sigma_1 U + X_1$) is defined as

$$I_{\text{ah}}(X) = [X_1 + l(X_1 - X_2),\, X_1 + u(X_1 - X_2)], \tag{7}$$

where $l(d) = \sigma_1\mathbb{E}(U|x) - z_{\alpha/2}\sigma_1\sqrt{\mathbb{V}(U|x)}$ and $u(d) = \sigma_1\mathbb{E}(U|x) + z_{\alpha/2}\sigma_1\sqrt{\mathbb{V}(U|x)}$, and where $z_{\alpha/2} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.

Theorem 3.2 (also see Denis 2010) gives an expression for the frequentist coverage probability of a more general interval for θ_1, of which $I_{\text{ah}}(X)$ is a particular case.

Theorem 3.2 *Let $X_i \sim N(\theta_i, \sigma_i^2)$, $i = 1, 2$, independent, with $d = X_1 - X_2$, σ_i^2 known and consider an interval of the form $I(X) = [X_1 + l(d), X_1 + u(d)]$. Then the frequentist coverage probability, $\mathbb{P}[\theta_1 \in I(X)]$, is given by*

$$C(\theta) = \mathbb{E}^Z\left[\Phi\left(\gamma\; u\left\{\sqrt{\sigma_1^2+\sigma_2^2}\,Z+\beta\right\} + \frac{\sigma_1}{\sigma_2}Z\right)\right.$$
$$\left. -\Phi\left(\gamma\, l\left\{\sqrt{\sigma_1^2+\sigma_2^2}\,Z+\beta\right\} + \frac{\sigma_1}{\sigma_2}Z\right)\right], \tag{8}$$

where $\beta = \theta_1 - \theta_2$, $\gamma = \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}$, and $Z \sim N(0, 1)$.

Proof We have $C(\theta) = \mathbb{P}_\theta\left[\theta_1 \in I(X)\right] = \mathbb{P}_\theta\big[X_1 + l\{X_1 - X_2\} \le \theta_1 \le X_1 + u\{X_1 - X_2\}\big] = \mathbb{P}_\theta\left[-u\{Y_1 - Y_2 + \beta\} \le Y_1 \le -l\{Y_1 - Y_2 + \beta\}\right]$, where $Y_i = X_i -$

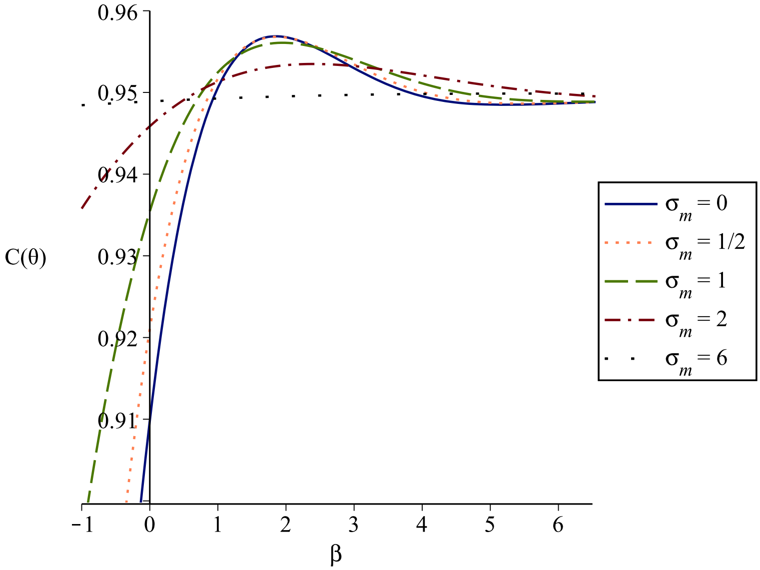

Fig. 1 Frequentist coverage probability of the ad hoc interval ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $\beta = \theta_1 - \theta_2$ for varying σ_m

$\theta_i \sim N(0, \sigma_i^2)$, $i = 1, 2$, are independent. Setting $Z = \frac{Y_1 - Y_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$ and $Z' = \gamma\left(Y_1 - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z\right)$, we obtain $(Z, Z')^T \sim N_2(0, I_2)$. Now, by conditioning, we have

$$\begin{aligned} C(\theta) &= \mathbb{P}\left[\gamma\left(-u\left\{\sqrt{\sigma_1^2+\sigma_2^2}Z+\beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}Z\right)\right. \\ &\left.\leq Z' \leq \gamma\left(-l\left\{\sqrt{\sigma_1^2+\sigma_2^2}Z+\beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}Z\right)\right] \\ &= \mathbb{E}^Z\left[\mathbb{P}\left[\gamma\left(-u\left\{\sqrt{\sigma_1^2+\sigma_2^2}Z+\beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}Z\right)\right.\right. \\ &\left.\left.\leq Z' \leq \gamma\left(-l\left\{\sqrt{\sigma_1^2+\sigma_2^2}Z+\beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}Z\right)\right]\right], \end{aligned}$$

which yields (8). □

As a first example, Fig. 1 presents the frequentist coverage probability of the ad hoc interval for $\sigma_1 = \sigma_2 = 1$, a 0.95 nominal level and varying σ_m.

While the maximum coverage appears to decrease in σ_m, the overall discrepancy between frequentist coverage and credibility for $\beta \geq 0$ tends to diminish as σ_m increases. The coverage of $I_{ah}(X)$ at $\beta = 0$ also appears to increase as σ_m increases (although it seems to remain below the nominal level $1-\alpha$). The same ordering occurs for negative values of β, which is understandable as larger values of σ_m correlate with more uncertainty on the bound $\beta \geq 0$, which in turn becomes reflected in the posterior distribution. Moreover, we have $\lim_{\beta \to \infty} C(\theta) = 1-\alpha$. This can be shown in the same way as in Remark 3.3 below for $\sigma_m \to \infty$ since we have $\lim_{d\to\infty} u(d) = -\lim_{d\to\infty} l(d) = \sigma_1 z_{\alpha/2}$. We noted similar overall behaviour of $I_{ah}(X)$ for other nominal levels such as 0.80, 0.90 and 0.99.

Remark 3.3 Without recourse to the additional information provided by X_2, a benchmark confidence interval for θ_1 is given by $X_1 \pm z_{\alpha/2}\sigma_1$. This interval arises from $I_{ah}(X)$ by taking $\sigma_m \to \infty$ in (2) and (7), yielding $\lim_{\sigma_m^2\to\infty} \pi(u|x) = \phi(u),\ \forall u \in \mathbb{R}$. Accordingly, one infers that $\lim_{\sigma_m\to\infty} C(\theta) = 1-\alpha,\ \forall\theta \in \mathbb{R}^2$, and this is illustrated in Fig. 1 (for $\theta_1 \geq \theta_2$ mostly) with the flattening out around the nominal level observed as σ_m increases.

We also consider the credibility $\mathbb{P}[\theta_1 \in I_{ah}(X)|x]$ of the ad hoc interval, also given by

$$\mathbb{P}[U \in [l(d), u(d)]|x] = \int_{l(d)}^{u(d)} \pi(u|x)du,$$

where $l(d) = \mathbb{E}[U|x] - z_{\alpha/2}\sqrt{\mathbb{V}(U|x)}$ and $u(d) = \mathbb{E}[U|x] + z_{\alpha/2}\sqrt{\mathbb{V}(U|x)}$.

Figure 2 presents the credibility as a function of $d = x_1 - x_2$ of the ad hoc interval with $1-\alpha = 0.95$, and $\sigma_1^2 = \sigma_2^2 = 1$ for varying values of σ_m. Examining Fig. 2, we notice that the credibility flattens out around the nominal level as σ_m increases, as was the case for the coverage probability, which is justified here by the fact that $\pi(u|x) \to \phi(u)$ as $\sigma_m^2 \to \infty$. For all values of σ_m, the exact credibility is remarkably close to the nominal level, with slightly higher credibility for positive d. Such closeness was equally observed for other nominal levels and other settings of σ_1^2 and σ_2^2.

3.2 Credible sets defined in terms of a spending function

The ad hoc procedure previously considered creates a credible set which is centered at the mean of the posterior distribution and which extends on either side of the mean by equal amounts. Given the asymmetry of the posterior density, it is justifiable to consider throwing out α_1 in one tail and α_2 in the other tail such that $\alpha_1 + \alpha_2 = \alpha$. As above, exact credibility will not be achieved for all x, but it turns out for practical purposes to be close to nominal credibility (see Fig. 4). This idea of discarding unequal amounts in the tails is referred to as a spending function in Ghashim et al. (2016), and previously in Marchand and Strawderman (2013). We consider the situation where we discard $k\alpha$ in the left tail and $(1-k)\alpha$ in the right tail. The adjustment in this direction

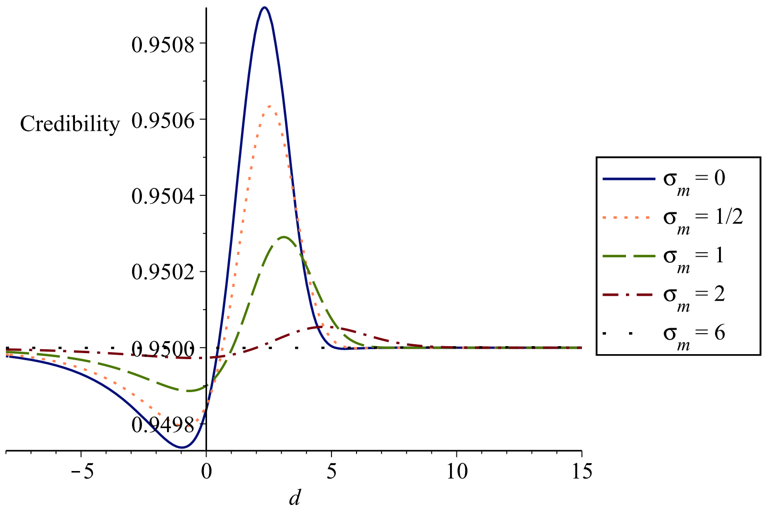

Fig. 2 Credibility of the ad hoc interval ($1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $d = x_1 - x_2$ for varying σ_m

with $k < 1/2$ is motivated by a relatively smaller coverage for $\beta = \theta_1 - \theta_2$ closer to 0 (see Fig. 1).

Definition 3.4 Let $\mathbb{E}(U|x)$ and $\mathbb{V}(U|x)$ denote respectively the posterior expectation and variance of U given by Lemma 2.5. The ad hoc Bayes credible interval for $\theta_1 = \sigma_1 U + X_1$ defined in terms of a spending function is given by

$$I'_{\text{ah}}(X) = [X_1 + l'(X_1 - X_2), X_1 + u'(X_1 - X_2)], \tag{9}$$

where $l'(d) = \sigma_1 \mathbb{E}(U|x) - z_{k\alpha}\, \sigma_1 \sqrt{\mathbb{V}(U|x)}$ and $u'(d) = \sigma_1 \mathbb{E}(U|x) + z_{(1-k)\alpha}\, \sigma_1 \sqrt{\mathbb{V}(U|x)}$, with $z_\alpha = \Phi^{-1}(1 - \alpha)$.

Theorem 3.2 holds for general $u(d)$ and $l(d)$, so Eq. (8) holds here for all values of k. Figure 3 presents the frequentist coverage probability of the ad hoc interval for $\sigma_1 = \sigma_2 = 1$, $\sigma_m = 0$, a 0.95 nominal level and varying values of k in the spending function.

Similarly to previous results, it is easy to show that $\lim_{\beta \to \infty} C(\theta) = 1 - \alpha$ for all k. The coverage at $\beta = 0$ appears to be a decreasing function of k. Further numerical exploration suggests that $C(0) \geq 1 - \alpha$ for $k \leq 1/4$, even for various other values of $1 - \alpha$. Moreover, for small values of k, the minimum coverage is no longer attained at $\beta = 0$. It would be interesting to investigate theoretically if the coverage has a local minimum after the initial peak or if it decreases monotonically towards the limiting value of $(1 - \alpha)$. If the latter were true, then the coverage would always be above the nominal value whenever $C(\theta) > 1 - \alpha$ for $\theta_1 - \theta_2 = 0$. Further illustration and observations about the coverage at $\beta = 0$ are provided by Drew (2021).

Figure 4 presents the credibility as a function of $d = x_1 - x_2$ of the ad hoc interval for $\sigma_1 = \sigma_2 = 1$, $\sigma_m = 0$, a 0.95 nominal level and varying values of k.

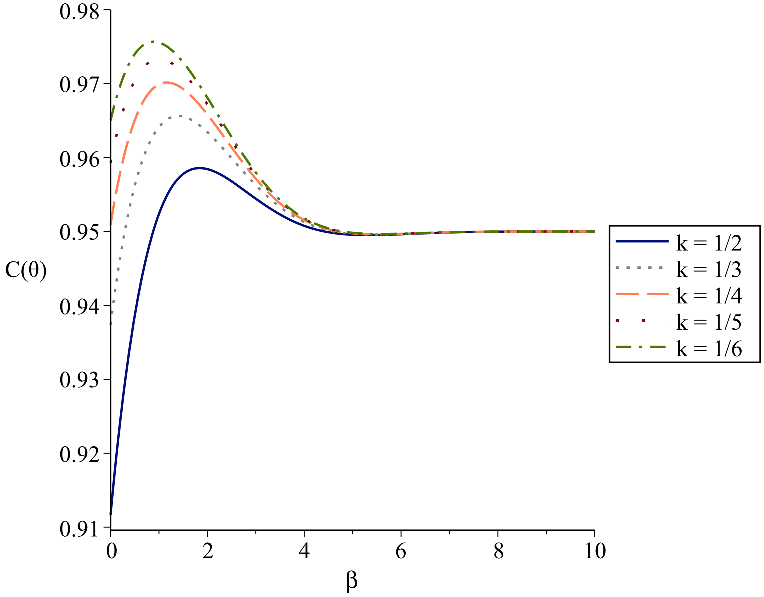

Fig. 3 Frequentist coverage probability of the ad hoc interval ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_m^2 = 0$) as a function of $\beta = \theta_1 - \theta_2$ for varying values of k in the spending function

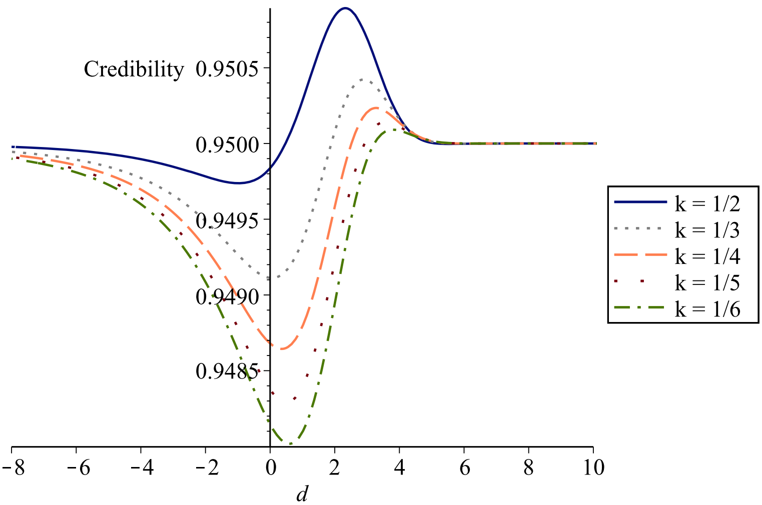

Fig. 4 Credibility of the ad hoc interval ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_m^2 = 0$) as a function of $d = x_1 - x_2$ for varying values of k in the spending function

The overall credibility appears to be the best when $k = 1/2$, and decrease as k decreases. That being said, for all values of k plotted here, the credibility remains extremely close to the nominal level. For the sake of further comparison, Table 1 gives

Table 1 Approximate maximum credibility discrepancy for the ad hoc interval with $k = 1/4$ in the spending function, $\sigma_1 = \sigma_2 = 1$ and $\sigma_m = 0$

$1 - \alpha$	0.80	0.90	0.95	0.99
Maximum discrepancy	0.0049	0.00065	0.0014	0.0017

an approximate maximum discrepancy of the credibility for $k = 1/4$ and varying values of $1 - \alpha$.

Remark 3.5 Unsurprisingly, the credible intervals $I_{ad}(X)$ and $I'_{ad}(X)$ typically lead to shorter intervals in comparison to the non-informative case $\sigma_m \to \infty$. The expected length of these credible intervals is further studied in Drew (2021) and illustrated for various settings of σ_m^2 and the spending function (i.e., k).

4 Concluding remarks

For estimating the suspected larger (θ_1) of two normal means (θ_1 and θ_2), we have studied the frequentist risk performance of Bayesian point and interval estimators associated with non-informative prior densities of the form:

$$\pi(\theta_1, \theta_2|m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2), \; m \sim N(0, \sigma_m^2).$$

Firstly, we establish for all $\sigma_m > 0$ the minimaxity of the Bayesian point estimator of θ_1 under squared error loss and when the supremum risk is taken on $\theta_1 \geq \theta_2$, thus extending the previously known result for $\sigma_m = 0$. Secondly, we provide ample evidence of satisfactory, or even excellent, frequentist performance of Bayesian credible sets for the same priors as measured on the set of parameter values $\theta_1 \geq \theta_2$, with such procedures capitalizing on the additional information available for θ_2. In doing so, we have elicited how the frequentist probability of coverage varies with the difference $\beta = \theta_1 - \theta_2$, as well as vary according to the choice of the hyperparameter σ_m ranging from the "no-useful additional information case" ($\sigma_m \to \infty$) to the certain constraint $\theta_1 \geq \theta_2$ ($\sigma_m = 0$). Moreover, we have further illustrated the role of a spending function in the construction of the Bayesian credible set and how its setting can give rise to even better frequentist coverage probability.

The findings of this paper also apply to situations where $m \sim N(\xi, \sigma_m^2)$ in (1) with $\xi \neq 0$. Indeed for such a case, we can set $X'_1 = X_1 - \xi$ and $\theta'_1 = \theta_1 - \xi$ so that point and interval estimates of θ'_1 based on (X'_1, X_2) with $\theta'_1 - \theta_2 \geq m'$, $m' =^d m - \xi \sim N(0, \sigma_m^2)$, translate to point and interval estimates of θ_1. For instance, he above strategy will generate point estimates $\hat{\theta}_1(x) = \hat{\theta}'_1(x'_1, x_2) + \xi$. Theorem 2.10's minimaxity result will then apply to the parametric restriction $\theta_1 - \theta_2 \geq \xi$, and Section 3's study of frequentist coverage probability which pertains to $\beta' = \theta'_1 - \theta_2$ will equate to $\beta = \theta_1 - \theta_2 \geq \xi$.

The results of this paper do leave open several interesting questions about analytically derived lower bounds on coverage probabilities which bring into play the model variances, the choice of σ_m, as well as the spending function setting. It would be particularly interesting to proceed with an analysis for an unknown variances extension of model (1). Finally, although we have focussed on a relatively simple two-parameter problem with normal observables, we do believe that the ideas or techniques put forth can be adapted to a wider range of settings, namely the incorporation of uncertainty on a parametric restriction and the use of a spending function in the construction of Bayesian credible sets.

Acknowledgements The authors are grateful to the Editor and a reviewer for thoughtful and helpful comments. Courtney Drew acknowledges fellowship support from the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as from the Fonds de recherche du Québec. Éric Marchand's research is supported in part by the NSERC of Canada. On behalf of both authors, the corresponding author states that there is no Conflict of interest.

References

Arnold BC, Beaver RJ (2002) Skewed multivariate models related to hidden truncation and/or selective reporting. Sociedad de Estadística e Investigación Operativa Test 11:7–54

Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12:171–178

Blumenthal S, Cohen A (1968) Estimation of the larger translation parameters. Ann Math Stat 39:502–516

Blumenthal S, Cohen A (1968) Estimation of two ordered translation parameters. Ann Math Stat 39:517–530

Berger JO (1985) Statistical Decision Theory and Bayesian Analysis. Springer Series in Statistics, Springer, second edition

Cohen A, Sackrowitz HB (1970) Estimation of the last mean of a monotone sequence. Ann Math Stat 41:2021–2034

Denis, M. Méthodes de modélisation bayésienne et applications en recherche clinique. PhD thesis. Université de Montpellier, 2010

Drew C (2021) Estimateurs bayésiens et intervalles de crédibilité pour des moyennes normales ordonnées en présence d'une contrainte paramétrique incertaine. M.Sc. thesis. Université de Sherbrooke. (http://hdl.handle.net/11143/18673)

Gelfand AE, Smith AFM, Lee TM (1992) Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. J Am Stat Assoc 87:523–532

Ghashim E, Marchand É, Strawderman WE (2016) On a better lower bound for the frequentist probability of coverage of Bayesian credible intervals in restricted parameter spaces. Stat Methodol 31:43–57

Kumar S, Sharma D (1988) Simultaneous estimation of ordered parameters. Commun Stat: Theory Methods 7:4315–4336

Liseo B, Loperfido N (2003) A Bayesian interpretation of the multivariate skew-normal distribution. Stat Prob Lett 61:395–401

Madi MT, Leonard T, Tsui K-W (2000) Bayes inference for treatment effects with uncertain order constraints. Stat Probab Lett 49:277–283

Mandelkern M (2002) Setting confidence intervals for bounded parameters with discussion. Stat Sci 17:149–172

Marchand É, Nicoleris T (2019) On Bayes minimax estimators for a normal mean with an uncertain constraint. Commun Stat: Theory Methods 50:1873–1883

Marchand É and Strawderman WE (2004) Estimation in restricted parameter spaces: a review. A Festschrift for Herman Rubin. IMS Lecture Notes-Monograph Series, pp 21–44

Marchand É, Strawderman WE (2006) On the behaviour of Bayesian credible intervals for some restricted parameter space problems. Recent Developments in Nonparametric Inference and Probability: A Festschrift for Michael Woodroofe, IMS Lecture Notes-Monograph Series 50:112–126

Marchand É, Strawderman WE (2013) On Bayesian credible sets, restricted parameter spaces and frequentist coverage. Electr J Stat 7:1419–1431

O'Hagan A, Leonard T (1976) Bayes estimation subject to uncertainty about parameter constraints. Biometrika 63:201–203 van Eeden C and Zidek JV (2002) Combining sample information in estimating ordered normal means. Sankhyā: The Indian J Stat 64:588–610