Estimateurs bayésiens et intervalles de crédibilité pour des moyennes normales ordonnées en présence d'une contrainte paramétrique incertaine
Bayes estimators and credible sets for ordered normal means with an uncertain constraint

par

Courtney Drew

mémoire présenté au Département de mathématiques en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, septembre 2021

Le 2 septembre 2021

Le jury a accepté le mémoire de Madame Courtney Drew dans sa version finale.

Membres du jury

Professeur Éric Marchand

Directeur de recherche

Département de mathématiques

Professeur William E. Strawderman

Membre externe

Department of Statistics - Rutgers University

Professeur Klaus Herrmann

Président-rapporteur

Département de mathématiques

ABSTRACT

This work concerns parameter estimation for a bivariate normal model where the difference of the two normal means is bounded. A hierarchical prior is considered which allows uncertainty to be introduced in the parametric constraint. We begin by studying the case of a uniform prior on $\theta_1 - \theta_2 \geq m$ where $m \sim N(0, \sigma_m^2)$. We show that the Bayes estimator of θ_1 dominates X_1 under squared error loss, which extends the known result for $\sigma_m^2 = 0$; that is, without uncertainty in the constraint. An ad hoc credible set for θ_1 is then presented along with an analysis of its frequentist properties including coverage probability. Finally, the last chapter considers two modifications to the previous model : a parametric restriction of the form $|\theta_1 - \theta_2| \leq m$ and the case of unknown variances.

SOMMAIRE

Ce mémoire porte sur l'estimation paramétrique d'un modèle normal bivarié où la différence des deux moyennes normales est bornée. Une loi a priori hiérarchique est considérée, ce qui permet d'introduire de l'incertitude dans la restriction paramétrique. En premier lieu, le cas d'une loi a priori uniforme sur $\theta_1 - \theta_2 \geq m$ est étudié où $m \sim N(0, \sigma_m^2)$. On montre que l'estimateur de Bayes de θ_1 domine X_1 en termes de perte quadratique lorsque $\theta_1 - \theta_2 \geq 0$. Ceci étend le résultat connu pour le cas $\sigma_m^2 = 0$; c'est-à-dire, en absence d'incertitude dans la contrainte. On présente ensuite un intervalle de crédibilité ad hoc pour θ_1 ainsi qu'une analyse de ses propriétés fréquentistes dont la probabilité de recouvrement. Finalement, le dernier chapitre considère deux modifications au modèle précédent : une restriction paramétrique de la forme $|\theta_1 - \theta_2| \leq m$ et le cas de variances inconnues.

ACKNOWLEDGEMENTS

My gratitude goes first and foremost to my research supervisor, Professor Éric Marchand. Thank you for your guidance, availability, advice and support. Your interest in your students' success, both academic and otherwise, has not gone unnoticed. I am also grateful for my family, my mother and brother, for their continued presence, and for numerous discussions with friends and colleagues. Finally, I acknowledge funding from the Département de mathématiques and fellowship support from the Natural Sciences and Engineering Research Council of Canada, as well as from the Fonds de Recherche du Québec.

Courtney Drew

Sherbrooke, July 2021

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

This work concerns parameter estimation for a bivariate normal model in the presence of a constraint on the difference of the two normal means. The larger topic at hand is thus estimation in restricted parameter spaces, a field which has gained attention since the 1950s. Parametric constraints arise naturally in many situations and in various forms. Perhaps the most fundamental example of such a problem was studied by Katz [Kat61] in 1961 and concerns point estimation of a positive normal mean under squared error loss. While there is no guarantee that estimators retain their favourable properties when a parametric restriction is introduced, some do carry over and it is of interest to study such problems (see van Eeden [vE06]).

The problem of estimating the larger of two normal means has been well studied, dating back to the late 1960s with a series of papers by Blumenthal & Cohen [BC68a, BC68b] followed by Cohen & Sackrowitz [CS70]. Questions of admissibility and minimaxity are discussed for both the simultaneous estimation of the two means as well as the estimation of the larger mean. Several estimators are studied including restricted and unrestricted maximum likelihood estimators as well as Bayes estimators with respect to a uniform prior on the restricted parameter space. In particular, the Bayes estimator, for both (θ_1, θ_2) in the simultaneous estimation of the two means and θ_1 in estimating the larger normal mean, is shown to be minimax and admissible on $\mathbb{R}^2$ under squared error loss.

A recent motivation for this work comes from a paper by Marchand & Nicoleris [MN19]. The authors consider estimating a normal mean which is suspected to be positive. They introduce uncertainty in the lower bound by using a hierarchical prior (see O'Hagan & Leonard [OL76] for an early use of this idea), yielding the following model and prior :

$$X|\theta, m \sim N(\theta, \sigma^2), \quad \theta|m \sim N(\mu, \tau^2)\mathbb{1}_{[m,\infty)}(\theta), \quad m \sim N(0, \sigma_m^2),$$

where μ and $\sigma^2, \tau^2, \sigma_m^2 \geq 0$ are known. The case $\sigma_m^2 = 0$ recovers the deterministic constraint $\theta \geq 0$ which corresponds to the problem considered by Katz [Kat61]. Marchand & Nicoleris proceed to show that a class of generalized Bayes estimators of θ_1 dominate the benchmark estimator X_1 and are thus minimax. The results obtained under this uncertain constraint framework, along with the known two-sample point estimation results when $\theta_1 \geq \theta_2$ (without uncertainty), suggest the idea of adding uncertainty to the bound in the two-sample problem and serve as motivation for this work. The performance of point estimators in such problems also leads to interest in Bayes credible sets. Bayesian credible sets typically do not have matching coverage probability, and even in rare occasions where they do, adding a parametric restriction further perturbs the probability matching (e.g., Mandelkern [Man02] ; Marchand & Strawderman [MS06]). Investigating the behaviour of the frequentist coverage of Bayes credible sets in restricted parameter spaces remains an interesting topic, especially for multiparameter problems. For example, there are a number of situations where Bayes credible sets associated with a non-informative or diffuse prior provide coverage probability close to the nominal credibility (e.g., Zhang & Woodfoofe [ZW03] ; Roe & Woodroofe [RW00] ; Marchand & Strawderman [MS06] ; Marchand et al. [MSBL08] ; Marchand & Strawderman [MS13] ; Ghashim et al. [GMS16]).

This document is organized as follows. Chapter 1 is dedicated to preliminary theory, presenting both definitions and results which are used later. The chapter begins by presenting relevant distributions along with related properties. The focus then shifts to parameter estimation where both point and interval estimation are discussed. Measures of goodness of estimates such as frequentist coverage probability, credibility, minimaxity and admissibility are defined with an emphasis placed on the differences between frequentist and Bayesian approaches.

The main portion of this work is presented in Chapter 2 and concerns estimation of the suspected larger of two normal means. This chapter forms the basis of a recently submitted paper. We consider a model of the form $X_i \sim N(\theta_i, \sigma_i^2)$, X_i independent for $i = 1, 2$ and σ_i^2 known, with a hierarchical prior $\pi(\theta_1, \theta_2|m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2)$, $m \sim N(0, \sigma_m^2)$ where σ_m^2 is known. The degenerate case without uncertainty in the parametric restriction ; that is, for a prior $\pi(\theta_1, \theta_2) = \mathbb{1}_{[0,\infty)}(\theta_1 - \theta_2)$ (or equivalently taking $\sigma_m^2 \to 0$ in the hierarchical prior), has been well studied in the literature, notably by Blumenthal & Cohen [BC68b] and Cohen & Sackrowitz [CS70]. We obtain a class of minimax Bayes estimators of θ_1 that dominate the benchmark estimator X_1 under squared error loss, thus extending the known minimaxity result for $\sigma_m^2 = 0$. We also address questions of admissibility. We then turn to the construction of Bayesian credible sets for θ_1 and introduce an ad hoc credible interval of a standard form $\mathbb{E}[\theta|x] \pm z_{\alpha/2}\sqrt{\mathbb{V}[\theta|x]}$ with approximate $1-\alpha$ credibility (see Denis [Den10]). An ad hoc procedure of this type has been advocated as a reasonable approach by several authors (e.g., Berger [Ber85]). We study the ad hoc interval's frequentist coverage probability and provide numerical evidence of it closely matching the nominal credibility. Finally, we incorporate a spending function and study its impact on the frequentist coverage probability and credibility.

The last chapter complements Chapter 2 by presenting two modifications to the model and parametric constraint and studying the resulting ad hoc credible interval defined in Chapter 2. Section 3.1 considers the same model for X_1 and X_2 as in Chapter 2, but now doubly-bounds the difference of means $\theta_1 - \theta_2$. Section 3.2 deals with the same type of constraint (i.e., ordered means), but with the variances of X_1 and X_2 unknown (but equal).

CHAPTER 1

Preliminary theory

In this section, we lay a (non-exhaustive) theoretical foundation for the results to come by presenting some preliminary theory pertaining to normal and skewed distributions as well as Bayesian inference, in particular to point and interval estimation. We also introduce other useful functions and results.

1.1 Multivariate normal distribution

The following well-known theory pertaining to the multivariate normal distribution can be found in Muirhead [Mui82].

Definition 1.1. *X has a p-variate normal distribution with mean μ and positive definite covariance matrix Σ when it has density on $\mathbb{R}^p$ given by*

$$f_X(x) = \frac{1}{(2\pi)^{p/2}(\det\Sigma)^{1/2}} \exp\left\{ -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}.$$

We denote such a distribution $X \sim N_p(\mu, \Sigma)$.

Lemma 1.2. *Let $X \sim N_p(\mu, \Sigma)$. Then the moment generating function of X is given by*

$$M_X(t) = \mathbb{E}\left[e^{t^\top X}\right] = e^{t^\top \mu + \frac{1}{2} t^\top \Sigma t}.$$

For $p = 2$, let $\mathbb{V}[X_i] = \sigma_i^2$ for $i = 1, 2$ and let ρ denote the Pearson correlation coefficient between X_1 and X_2. Such a situation corresponds to the bivariate normal distribution, which can be standardized by setting $Z_i = \frac{X_i - \mu_i}{\sigma_i}$ for $i = 1, 2$. The joint density of Z_1 and Z_2 can easily be shown to be

$$f_Z(z_1, z_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}(z_1^2 + z_2^2 - 2\rho z_1 z_2)\right\}.$$

Lemma 1.3. *Let $(X, Y) \sim N_2\left(\mu = \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \Sigma = \begin{bmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right)$ with $|\rho| \leq 1$. Then the conditional distribution of Y given $X = x$ is given by*

$$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right).$$

Lemma 1.4. *Let $X \sim N_p(\mu, \Sigma)$ and $Y = AX + b$, where A is $q \times p$ and b is $q \times 1$. Then*

$$Y \sim N_q\left(A\mu + b, A\Sigma A^\top\right).$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[e^{t^\top Y}\right] &= \mathbb{E}\left[e^{t^\top (AX+b)}\right] = e^{t^\top b}\,\mathbb{E}\left[e^{t^\top AX}\right] = e^{t^\top b} M_X(A^\top t) \\
&= e^{t^\top b} e^{t^\top A\mu + \frac{1}{2}t^\top (A\Sigma A^\top)t} = e^{t^\top (A\mu + b) + \frac{1}{2}t^\top (A\Sigma A^\top)t},
\end{aligned}$$

which one recognizes as the moment generating function of a $N_q(A\mu + b, A\Sigma A^\top)$ distribution. $\square$

Corollary 1.5. *Let $(X_1, X_2)^\top \sim N_2\left(\mu = \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \Sigma = \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right)$. Then setting $Y_1 = X_1$ and $Y_2 = \frac{1}{\sqrt{1-\rho^2}}\left(X_2 - \frac{\rho\sigma_2}{\sigma_1}X_1\right)$, we have*

$$(Y_1, Y_2)^\top \sim N_2\left(\mu' = \begin{bmatrix}\mu_1 \\ \frac{1}{\sqrt{1-\rho^2}}\left(\mu_2 - \frac{\rho\sigma_2}{\sigma_1}\mu_1\right)\end{bmatrix}, \Sigma' = \begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}\right).$$

In particular, Y_1 and Y_2 are independent.

Proof. We use Lemma 1.4 with $p = 2$, $A = \begin{bmatrix} 1 & 0 \\ -\frac{\rho\sigma_2}{\sigma_1\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}$ and $b = 0$. From this, we immediately obtain μ' and the covariance matrix is evaluated as follows

$$A\Sigma A^\top = \begin{bmatrix} 1 & 0 \\ -\frac{\rho\sigma_2}{\sigma_1\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & -\frac{\rho\sigma_2}{\sigma_1\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

□

1.2 Skewed distributions

Skewed distributions arise from various situations, one of which is hidden truncation. Arnold & Beaver [AB02] use the example of waist sizes for uniforms of elite troops, of which members are only selected if they are above a certain height, but it is easy to think of other examples of such a truncation where the units are selected only if they meet a certain threshold requirement. In the troop uniform example, a normal distribution wouldn't be appropriate to model the waist sizes since the latter are positively skewed. It would be better to use a skewed distribution. In this example, the truncation is not so hidden, but it is possible to have instances where there was some subliminal truncation or selection (see Arnold & Beaver [AB02] for more examples) which render the choice of a skewed distribution less obvious, yet more appropriate than that of a symmetric distribution. In a more formal estimation setting, skewed distributions can arise in posterior distributions where the prior is truncated to some restricted parameter space (see Chapter 2). We refer the reader to Azzalini [Azz14] for a detailed account of skewed distributions, with an emphasis on the construction of such distributions from the starting point of a continuous symmetric density. We denote throughout $\phi(\cdot)$ and $\Phi(\cdot)$ as the standard normal density and cdf respectively.

Definition 1.6. *ν has a truncated normal distribution with parameters $\alpha \in \mathbb{R}, \tau^2 > 0$*

and (a,b)*,* $-\infty \le a < b \le \infty$*, when it has density on* (a,b) *given by*

$$f(\nu) = \frac{1}{\tau} \frac{\phi\left(\frac{\nu-\alpha}{\tau}\right)}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)} \mathbb{1}_{(a,b)}(\nu).$$

We denote such a distribution $\nu \sim TN(\alpha, \tau^2, (a,b))$*.*

Lemma 1.7. *Let* $\nu \sim TN(\alpha, \tau^2, (a,b))$*. Then the expectation and variance of* ν *are given by*

$$\mathbb{E}[\nu] = \alpha + \tau \frac{\phi\left(\frac{a-\alpha}{\tau}\right) - \phi\left(\frac{b-\alpha}{\tau}\right)}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)}; \tag{1.1}$$

$$\mathbb{V}[\nu] = \tau^2 \left[1 + \frac{\left(\frac{a-\alpha}{\tau}\right)\phi\left(\frac{a-\alpha}{\tau}\right) - \left(\frac{b-\alpha}{\tau}\right)\phi\left(\frac{b-\alpha}{\tau}\right)}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)} - \left(\frac{\phi\left(\frac{a-\alpha}{\tau}\right) - \phi\left(\frac{b-\alpha}{\tau}\right)}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)}\right)^2\right]. \tag{1.2}$$

Proof. The moment generating function of ν is given by

$$\begin{aligned}
M_\nu(t) &= \mathbb{E}[e^{t\,\nu}] \\
&= \int_a^b e^{t\,\nu} \frac{1}{\tau} \frac{\phi\left(\frac{\nu-\alpha}{\tau}\right)}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)} d\nu \\
&= \frac{e^{-\frac{1}{2\tau^2}(\alpha^2 - (\tau^2 t + \alpha)^2)}}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)} \int_a^b \frac{1}{\tau} \phi\left(\frac{\nu - (\tau^2 t + \alpha)}{\tau}\right) d\nu \\
&= \frac{e^{\alpha\,t + \frac{\tau^2 t^2}{2}}}{\Phi\left(\frac{b-\alpha}{\tau}\right) - \Phi\left(\frac{a-\alpha}{\tau}\right)} \left[\Phi\left(\frac{b - (\tau^2 t + \alpha)}{\tau}\right) - \Phi\left(\frac{a - (\tau^2 t + \alpha)}{\tau}\right)\right],
\end{aligned}$$

from which we obtain $\mathbb{E}[\nu] = M_\nu'(0)$ and $\mathbb{V}[\nu] = \mathbb{E}[\nu^2] - (\mathbb{E}[\nu])^2 = M_\nu''(0) - (M_\nu'(0))^2$. □

Definition 1.8. *U has an extended skew-normal distribution with parameters* $c, d \in \mathbb{R}$ *when it has density on* $\mathbb{R}$ *given by*

$$f(u) = \frac{\phi(u)\Phi\left(cu + d\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)},$$

and we write $U \sim SN(c,d)$*.*

This density was proposed by Azzalini [Azz85] in 1985 and was further studied by Arnold & Beaver [AB02] among others. One recovers Azzalini's [Azz85] original skew-normal distribution, with density $f(u) = 2\phi(u)\Phi(c\,u)$, for $c \in \mathbb{R}$, with $d = 0$. The choice $c = 0$ of the shape parameter recovers the standard normal density. The cases $c > 0$ yield positively skewed distributions whereas negatively skewed distributions arise when $c < 0$.

Remark 1.9. *We can introduce location and scale parameters, μ and β respectively, in Definition 1.8 to obtain the following augmented class of densities*

$$f(u) = \frac{1}{\beta} \frac{\phi\left(\frac{u-\mu}{\beta}\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \Phi\left(\frac{c(u-\mu)}{\beta} + d\right).$$

We denote such distributions as $U \sim SN(c, d, \mu, \beta)$.

The following well-known result is useful for several calculations in this work, notably in computing posterior distributions in the following chapters.

Lemma 1.10. *Let $Z \sim N(0,1)$ and $\lambda, \varepsilon \in \mathbb{R}$. Then $\mathbb{E}\big[\Phi(\lambda(Z+\varepsilon))\big] = \Phi\left(\frac{\lambda\varepsilon}{\sqrt{1+\lambda^2}}\right)$.*

Proof. We have

$$\mathbb{E}\big[\Phi(\lambda(Z+\varepsilon))\big] = \int_{-\infty}^{\infty}\int_{-\infty}^{\lambda(z+\varepsilon)} \phi(y)\phi(z)dydz = \mathbb{P}\left[Y - \lambda Z \le \lambda\varepsilon\right] = \Phi\left(\frac{\lambda\,\varepsilon}{\sqrt{1+\lambda^2}}\right),$$

where $Y \sim N(0,1)$ is independent of Z. □

Lemma 1.11. *For $U \sim SN(c, d)$, the moment generating function, expectation and variance of U are given respectively by*

$$M_U(t) = \frac{e^{\frac{t^2}{2}}}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \Phi\left(\frac{c\,t + d}{\sqrt{1+c^2}}\right),$$

$$\mathbb{E}[U] = \frac{c}{\sqrt{1+c^2}} \frac{\phi\left(\frac{d}{\sqrt{1+c^2}}\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)}, \text{ and}$$

$$\mathbb{V}[U] = 1 - \frac{d\,c^2}{(1+c^2)^{3/2}} \frac{\phi\left(\frac{d}{\sqrt{1+c^2}}\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} - \frac{c^2}{1+c^2}\left[\frac{\phi\left(\frac{d}{\sqrt{1+c^2}}\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)}\right]^2.$$

Proof. With the change of variables $y = u - t$ ($dy = du$), we have

$$\begin{aligned}
M_U(t) &= \mathbb{E}\left[e^{t\,U}\right] \\
&= \int_{-\infty}^{\infty} e^{tu} \frac{\phi(u)\Phi\left(c\,u + d\right)}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} du \\
&= \frac{1}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{t(y+t)} e^{-\frac{(y+t)^2}{2}} \Phi\left(c\,(y+t) + d\right) dy \\
&= \frac{e^{\frac{t^2}{2}}}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \int_{-\infty}^{\infty} \phi(y)\Phi\left(c\left(y + t + \frac{d}{c}\right)\right) dy \\
&= \frac{e^{\frac{t^2}{2}}}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \mathbb{E}\left[\Phi\left(c\left(Y + t + \frac{d}{c}\right)\right)\right] \\
&= \frac{e^{\frac{t^2}{2}}}{\Phi\left(\frac{d}{\sqrt{1+c^2}}\right)} \Phi\left(\frac{c\,t + d}{\sqrt{1+c^2}}\right),
\end{aligned}$$

where $Y \sim N(0,1)$ and making use of Lemma 1.10. The expectation and variance follow by straightforward calculations. □

Skew-normal distributions relate to mean mixtures of normal distributions (e.g., Negarestani et al. [NJSB19]), as given by the following results.

Definition 1.12. *Let $X|\nu \sim N(\theta + a\nu, \sigma^2), \nu \sim G$ with density g, where $a \neq 0$ and σ^2 is known. Then X is said to be a mean mixture of normals and we write $X \sim MMN(\theta, a, \sigma^2, g)$.*

Lemma 1.13. *Let $X \sim MMN(\theta, a, \sigma^2, g)$, $a \neq 0$, where $g \sim TN(\alpha, \tau^2, (0, \infty))$. Then*

$$X \sim SN\left(\frac{a\tau}{\sigma}, \frac{\alpha\sqrt{a^2\tau^2 + \sigma^2}}{\sigma\tau}, \theta + a\alpha, \sqrt{a^2\tau^2 + \sigma^2}\right).$$

Proof. For $X|\nu \sim N(\theta + a\nu, \sigma^2)$, $a \neq 0$, with $\nu \sim TN(\alpha, \tau^2, (0,\infty))$, the marginal density of X is given by

$$\begin{aligned}
f_X(x) &= \int_{\mathbb{R}} f_{X|\nu=y}(x|y) f_\nu(y) dy \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-(\theta+ay))^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\,\tau} \frac{1}{\Phi\left(\frac{\alpha}{\tau}\right)} e^{-\frac{(y-\alpha)^2}{2\tau^2}} dy \\
&= \frac{1}{2\pi\sigma\tau\Phi\left(\frac{\alpha}{\tau}\right)} \int_0^\infty e^{-\frac{x^2}{2\sigma^2} + \frac{x(\theta+ay)}{\sigma^2} - \frac{(\theta+ay)^2}{2\sigma^2}} \cdot e^{-\frac{y^2}{2\tau^2} + \frac{\alpha y}{\tau^2} - \frac{\alpha^2}{2\tau^2}} dy \\
&= \frac{e^{-\frac{1}{2\sigma^2}(x-\theta)^2} e^{-\frac{\alpha^2}{2\tau^2}}}{2\pi\sigma\tau\Phi\left(\frac{\alpha}{\tau}\right)} \int_0^\infty e^{-\frac{1}{2}\left(\frac{a^2\tau^2+\sigma^2}{\sigma^2\tau^2}\right)y^2 + \left(\frac{a}{\sigma^2}(x-\theta) + \frac{\alpha}{\tau^2}\right)y} dy \\
&= \frac{e^{-\frac{1}{2\sigma^2}(x-\theta)^2} e^{-\frac{\alpha^2}{2\tau^2}}}{2\pi\sigma\tau\Phi\left(\frac{\alpha}{\tau}\right)} \cdot e^{\frac{1}{2}\frac{\left(a(x-\theta)\tau^2 + \alpha\sigma^2\right)^2}{\sigma^2\tau^2(a^2\tau^2+\sigma^2)}} \int_0^\infty e^{-\frac{1}{2}\left(\frac{a^2\tau^2+\sigma^2}{\sigma^2\tau^2}\right)\left(y - \frac{a\tau^2(x-\theta)+\alpha\sigma^2}{a^2\tau^2+\sigma^2}\right)^2} dy.
\end{aligned}$$

One recognizes the last integrand as the (unnormalized) density of a truncated normal on $(0,\infty)$ with location and scale parameters given respectively by

$$\tilde{\mu} = \frac{a\tau^2(x-\theta) + \alpha\sigma^2}{a^2\tau^2 + \sigma^2} \text{ and } \tilde{\sigma} = \sqrt{\frac{\sigma^2\tau^2}{a^2\tau^2+\sigma^2}}.$$

We therefore have

$$\begin{aligned}
f_X(x) &= \frac{e^{-\frac{1}{2\sigma^2}(x-\theta)^2} e^{-\frac{\alpha^2}{2\tau^2}}}{\sqrt{2\pi}\,\sigma\tau\Phi\left(\frac{\alpha}{\tau}\right)} \cdot e^{\frac{1}{2}\frac{\left(a(x-\theta)\tau^2+\alpha\sigma^2\right)^2}{\sigma^2\tau^2(a^2\tau^2+\sigma^2)}} \left(\frac{\sigma\tau}{\sqrt{a^2\tau^2+\sigma^2}}\right) \Phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{a^2\tau^2+\sigma^2}} \frac{1}{\Phi\left(\frac{\alpha}{\tau}\right)} \Phi\left(\frac{a\tau^2(x-\theta)+\alpha\sigma^2}{\sigma\tau\sqrt{a^2\tau^2+\sigma^2}}\right) e^{-\frac{1}{2\sigma^2}(x-\theta)^2 - \frac{\alpha^2}{2\tau^2} + \frac{1}{2}\frac{\left(a(x-\theta)\tau^2+\alpha\sigma^2\right)^2}{\sigma^2\tau^2(a^2\tau^2+\sigma^2)}} \\
&= \frac{1}{\sqrt{a^2\tau^2+\sigma^2}} \phi\left(\frac{x-\theta-a\alpha}{\sqrt{a^2\tau^2+\sigma^2}}\right) \frac{1}{\Phi\left(\frac{\alpha}{\tau}\right)} \Phi\left(\frac{a\tau^2(x-\theta)+\alpha\sigma^2}{\sigma\tau\sqrt{a^2\tau^2+\sigma^2}}\right),
\end{aligned}$$

from which we can conclude that

$$X \sim SN\left(c = \frac{a\tau}{\sigma}, d = \frac{\alpha\sqrt{a^2\tau^2+\sigma^2}}{\sigma\tau}, \mu = \theta + a\alpha, \beta = \sqrt{a^2\tau^2+\sigma^2}\right).$$

□

Note that without loss of generality, one can set $\theta = 0$ and $\tau = 1$. Lemma 1.13 indicates how a choice of (a, σ^2, α) for a MMN distribution leads to a $SN(c, d, \mu, \beta)$ distribution. Conversely, from parameters (c, d, μ, β), there is one redundant parameter. For instance, for $\alpha = \frac{d}{\sqrt{1+c^2}}$, $a = \frac{\mu\sqrt{1+c^2}}{d}$ and $\sigma^2 = \frac{\mu^2(1+c^2)}{c^2 d^2}$, we have $\beta = \frac{\mu(1+c^2)}{d}$ as being redundant. The next result summarizes how to generate a $SN(c, d)$ distribution.

Corollary 1.14. *Let $Z|\nu' \sim N(\sqrt{1-\rho}\,\nu', \rho)$ with $\nu' \sim TN(0, 1, (-\alpha, \infty))$ and $0 < \rho < 1$. Then $Z \sim SN\left(c = \sqrt{\frac{1-\rho}{\rho}}, d = \frac{\alpha}{\sqrt{\rho}}\right)$. Equivalently, if $U \sim SN(c, d)$, then U admits the representation $U|\nu' \sim N(\sqrt{1-\rho}\,\nu', \rho)$ with $\nu' \sim TN(0, 1, (-\alpha, \infty))$ and where $\rho = \frac{1}{1+c^2}$ and $\alpha = \frac{d}{\sqrt{1+c^2}}$.*

Proof. From Lemma 1.13, the representation

$$Z = \frac{X - a\alpha}{\sqrt{a^2+\sigma^2}} \,\Big|\, \nu \sim N\left(\frac{a(\nu-\alpha)}{\sqrt{a^2+\sigma^2}}, \frac{\sigma^2}{a^2+\sigma^2}\right)$$

with $\nu \sim TN(\alpha, 1, (0, \infty))$ implies that

$$Z \sim SN\left(c = \frac{a}{\sigma}, d = \frac{\alpha\sqrt{a^2+\sigma^2}}{\sigma}\right).$$

Conversely, for $Z|\nu' \sim N(\sqrt{1-\rho}\,\nu', \rho)$ and setting $\nu' \overset{d}{=} \nu - \alpha \sim TN(0, 1, (-\alpha, \infty))$, we obtain from Lemma 1.13 that

$$Z \sim SN\left(c = \sqrt{\frac{1-\rho}{\rho}}, d = \frac{\alpha}{\sqrt{\rho}}\right),$$

with $\rho = \frac{\sigma^2}{a^2+\sigma^2}$. □

1.3 Reverse Mill's ratio

An important function which appears consistently throughout this work is the reverse Mill's ratio, also sometimes referred to as the reversed hazard rate and which relates to the Mill's ratio. We introduce this function and provide some of its properties here. We refer the reader to Baricz [Bar08] for a wealth of other properties.

Definition 1.15. *The reverse Mill's ratio is given by*

$$R(t) = \frac{\phi(t)}{\Phi(t)}, \; t \in \mathbb{R}.$$

Lemma 1.16. *(Properties of the reverse Mill's ratio)*

(a) $R(t) + t > 0 \quad \forall t \in \mathbb{R}$ *;*

(b) $R(t)$ *is a decreasing function of* t *with* $\lim_{t \to -\infty} R(t) = +\infty$ *and* $\lim_{t \to \infty} R(t) = 0$ *;*

(c) $R'(t) = -R(t)\big(t + R(t)\big)$ *is an increasing function of* t *(i.e.,* $R(t)$ *is convex) with bounds* $-1 < R'(t) < 0 \quad \forall t \in \mathbb{R}$ *;*

(d) $\lim_{t \to -\infty} R(t)(t + R(t)) = 1$ *and* $\lim_{t \to -\infty} (t + R(t)) = 0$ *;*

(e) $\lim_{t \to \infty} (tR(t)) = 0$ *;*

(f) $\lim_{t \to -\infty} \frac{R(t)}{t} = -1.$

Proof. These properties are well-known and can be found in the literature (e.g., Gordon [Gor41] ; Baricz [Bar08] ; Sampford [Sam53]). We provide an instructional proof for part (a).

(a) This property can be obtained from a probabilistic point of view. For $X \sim N(\mu, 1)$ and prior $\pi(\mu) = \mathbb{1}_{(0,\infty)}(\mu)$, we obtain the posterior $\mu|x \sim TN(x, 1, (0, \infty))$, a density which is defined on $(0, \infty)$. From (1.1) of Lemma 1.7, we obtain that $\mathbb{E}[\mu|x] = x + R(x)$ from which we can conclude that $x + R(x) > 0$ since $\mathbb{P}[\mu > 0|x] = 1$.

□

1.4 Monotone likelihood ratio

The following general theory concerning monotone likelihood ratios can be found in Lehmann & Romano [LR05] and Casella & Berger [CB02] among others.

Definition 1.17. *The family of densities* $\{p_\theta(x) : \theta \in \Theta\}$ *has monotone likelihood ratio in* x *(or in a statistic* $T(x)$*) if for all* $\theta_1 > \theta_0$, $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}$ *is monotone in* x *(or* $T(x)$*).*

Example 1.18. *An example which relates to our work concerns a truncated normal distribution.* $X \sim TN(\theta, \sigma^2, (a, b))$ *with* a, b *and* $\sigma^2 > 0$ *fixed has increasing likelihood ratio in* x*. Indeed, for* $\theta_1 > \theta_0$*, we have*

$$\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \frac{\Phi\left(\frac{b-\theta_0}{\sigma}\right) - \Phi\left(\frac{a-\theta_0}{\sigma}\right)}{\Phi\left(\frac{b-\theta_1}{\sigma}\right) - \Phi\left(\frac{a-\theta_1}{\sigma}\right)} \cdot \frac{\phi\left(\frac{x-\theta_1}{\sigma}\right)}{\phi\left(\frac{x-\theta_0}{\sigma}\right)} = A \cdot e^{\frac{x}{\sigma^2}(\theta_1-\theta_0)}, \quad A > 0,$$

which is increasing in x *since* $f(t) = e^t$ *is an increasing function of* t *and* $\theta_1 > \theta_0$*. Setting* $a = -\infty$ *and* $b = \infty$ *recovers the result for the normal distribution. This can also be generalized to all one-parameter densities in the exponential family, admitting density* $f_\theta(x) = h(x)c(\theta)e^{\eta(\theta)t(x)}$*, provided that* $\eta(\theta)$ *is non-decreasing.*

Lemma 1.19. *Let* $X \in \mathbb{R}^1$ *be a random variable and let* $f(x)$ *and* $g(x)$ *be increasing functions of* x*. Then* $\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)]$*, assuming all expectations exist. In other words,* $\mathrm{Cov}(f(X), g(X)) \geq 0$.

Proof. Let X and Y be i.i.d. Since f and g are increasing, for all $x, y \in \mathbb{R}$, we have $(f(x) - f(y))(g(x) - g(y)) \geq 0$. Thus, by the monotonicity of expectation, we have

$$\mathbb{E}[(f(X) - f(Y))(g(X) - g(Y))] \geq 0.$$

Expanding and using the fact that X and Y are i.i.d., we have

$$
\begin{aligned}
&\mathbb{E}[f(X)g(X) - f(X)g(Y) - f(Y)g(X) + f(Y)g(Y)] \geq 0 \\
\implies &\mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(Y)] - \mathbb{E}[f(Y)]\mathbb{E}[g(X)] + \mathbb{E}[f(Y)g(Y)] \geq 0 \\
\implies &2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)]\mathbb{E}[g(X)] \geq 0 \\
\implies &\mathrm{Cov}(f(X), g(X)) \geq 0.
\end{aligned}
$$

$\square$

Theorem 1.20. *Let $\{p_\theta : \theta \in \Theta\}$ have increasing likelihood ratio in x, with $p_{\theta_0} > 0$ whenever $p_{\theta_1} > 0$, and for $g : \mathbb{R}^1 \to \mathbb{R}^1$, let $g(x)$ be increasing in x. Then $\mathbb{E}_\theta[g(X)]$ is increasing in θ, given the expectations exist. In other words, the ordering of the family transfers to the ordering of the function.*

Proof. Let $\{p_\theta : \theta \in \Theta\}$ have increasing likelihood ratio in x, with $p_{\theta_0} > 0$ whenever $p_{\theta_1} > 0$, and let $g(x)$ be increasing in x. Let $\theta_1 > \theta_0$. Then,

$$
\begin{aligned}
\mathbb{E}_{\theta_1}[g(X)] &= \int g(x) p_{\theta_1}(x) d\mu(x) \\
&= \int g(x) \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) d\mu(x) \\
&= \mathbb{E}_{\theta_0}\left[g(X) \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}\right] \\
&\geq \mathbb{E}_{\theta_0}[g(X)] \mathbb{E}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}\right] && (1.3) \\
&= \mathbb{E}_{\theta_0}[g(X)], && (1.4)
\end{aligned}
$$

where (1.3) holds by Lemma 1.19 since $g(x)$ and $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}$ are both increasing functions of x, and (1.4) holds since p_{θ_1} is a density, so $\mathbb{E}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}\right] = 1$. $\square$

1.5 Estimation

1.5.1 Interval estimation

Both point estimation and interval estimation have the common objective of estimating some unknown parameter. As their names indicate, point estimation gives a single value estimate whereas interval estimation gives an estimate within some margin of error. Of course, one wishes to obtain the most precise estimate possible, which would lead to point estimation. The trade-off however is in the confidence that one has in this estimate. Interval estimation becomes a question of finding a balance between the preciseness of the estimate (the length of the interval) and the level of confidence that this interval contains the true parameter. Such confidence can be expressed in terms of coverage probability, which is defined in Definition 1.21.

Definition 1.21 (Casella & Berger [CB02]). *Let $I(X) = [l(X), u(X)]$ be an interval estimator for some function $\tau(\theta) \in \mathbb{R}^1$ of a parameter $\theta \in \mathbb{R}^k$, where $X \sim F_\theta(\cdot)$. Then the coverage probability of $I(X)$; that is, the probability that $I(X)$ covers the true value $\tau(\theta)$, is given by $\mathbb{P}_\theta[\tau(\theta) \in [l(X), u(X)]]$.*

There exist several methods to construct interval estimators. From a frequentist perspective, one thinks of inverting test statistics or using pivotal quantities for example (see Casella & Berger [CB02]). Here we limit ourselves to a Bayesian setting and concern ourselves with constructing Bayesian credible sets (the counterpart of confidence intervals in a frequentist setting). If the credible set takes the form of an interval, it is common to refer to it as a credible interval. It is important to realize that from a frequentist perspective, it is the confidence interval itself which is random, not the parameter. The parameter is considered fixed. On the other hand, under a Bayesian model, the unknown parameter is a random variable which follows some prior distribution π. This prior can be proper or improper (informative or non-informative) and is subjectively determined. It represents experts' prior beliefs or understanding on the matter, or lack of prior knowledge (in the case of a non-informative prior). Given a model for X and a prior π, one can use Bayes' rule to determine the posterior density.

Definition 1.22. *For $X|\theta$ having density $f(x|\theta)$ and prior density $\pi(\theta)$ (with respect to the Lebesgue measure on $\mathbb{R}^k$), the posterior density of θ is given by $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$, where $m(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$ is the marginal density of X.*

Example 1.23. *Let $X_1, \dots, X_n \sim N(\theta, \sigma^2)$ i.i.d. and $\theta \sim N(\mu, \tau^2)$, where μ, σ^2 and τ^2 are known. Then by Definition 1.22, with*

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} \text{ and } \pi(\theta) = (2\pi\tau^2)^{-\frac{1}{2}} e^{-\frac{1}{2\tau^2}(\theta-\mu)^2},$$

and by completing the square, one obtains that

$$\theta|x_1, \dots, x_n \sim N(\tilde{\mu}(x), \tilde{\tau}^2(x)),$$

where

$$\tilde{\mu}(x) = \frac{\sigma^2\mu + n\tau^2\bar{x}}{\sigma^2 + n\tau^2} \text{ and } \tilde{\tau}^2(x) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

In a Bayesian setting, all credible sets will be based on the posterior distribution. As was the case in the frequentist setting, there are several ways to construct Bayesian credible sets. A few common credible sets are highest posterior density and equal-tails.

Definition 1.24 (Casella & Berger [CB02])**.** *Let $X|\theta$ have density $f(x|\theta)$, prior $\pi(\theta)$ and posterior density $\pi(\theta|x)$. A $1-\alpha$ level highest posterior density (HPD) credible set for θ is given by*

$$A = \{\theta : \pi(\theta|x) \geq k\} \text{ such that } \int_A \pi(\theta|x)d\theta = 1 - \alpha.$$

For $k = 1$, a unimodal posterior density guarantees that the HPD set will be an interval. Otherwise, it is possible that the region does not take the form of an interval. Moreover, by construction, the HPD interval is the shortest $1 - \alpha$ level credible interval (though it can coincide with the equal-tails interval for example if the posterior distribution is symmetric).

Definition 1.25. *Let $X|\theta$ have density $f(x|\theta)$, prior $\pi(\theta)$ and posterior density $\pi(\theta|x)$. A $1 - \alpha$ level equal-tails credible interval is given by $[a(x), b(x)]$, where*

$$\int_{-\infty}^{a(x)} \pi(\theta|x)d\theta = \frac{\alpha}{2} \quad \textit{and} \quad \int_{b(x)}^{\infty} \pi(\theta|x)d\theta = \frac{\alpha}{2}.$$

While the HPD and equal-tails credible sets are conceptually easy to devise, calculating the bounds of the intervals must sometimes rely on numerical evaluations. Indeed, obtaining analytic bounds requires a closed form for the quantiles of the posterior distribution.

A measure of the performance of a credible set is its credibility; that is, the posterior probability that the random parameter is in the credible set.

Definition 1.26 (Casella & Berger [CB02])**.** *Let $X|\theta$ have density $f(x|\theta)$, prior $\pi(\theta)$ and posterior density $\pi(\theta|x)$, and let A be some set of possible values of $\tau(\theta) \in \mathbb{R}^1$, where $\theta \in \mathbb{R}^k$. The credibility of A, given some observation x, is given by*

$$\mathbb{P}[\tau(\theta) \in A|x] = \int_A \pi(\theta|x)d\theta.$$

Some sets have exact credibility by construction (for example the equal-tails credible set) whereas others do not. It is important to recognize the distinction between credibility and frequentist coverage probability. Much work has been done on the study of frequentist coverage of Bayesian credible sets, and on determining lower bounds for the coverage. See Section 2.4 for more on this.

Although there is typically a discrepancy between frequentist coverage and credibility, there are also rare occasions where the frequentist and Bayesian approaches are in agreement, as can be seen in Example 1.27 (e.g., Evans & Rosenthal [ER06]).

Example 1.27. *Let $(x_1, \ldots, x_n)$ be a sample from a $N(\mu, \sigma^2)$ model with σ^2 known (the case where σ^2 is unknown also yields a similar outcome). Then an exact $1-\alpha$ confidence interval for μ is given by $I(x) = \left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$. The interval $I(x)$ also corresponds to the Bayes HPD credible set under a uniform prior for μ on $\mathbb{R}$, which has exact $1-\alpha$ credibility. There is hence a convergence of the two approaches for this example.*

1.5.2 Point estimation

For the following, we denote Θ to be the parameter space. We also refer to a loss function $L(\theta, \delta)$. While there exist numerous loss functions, we focus on squared error loss, given by $L(\theta, \delta) = (\delta - \theta)^2$. The following theory on point estimation can be found in Lehmann & Casella [LC98] among others.

Definition 1.28. *Let δ be an estimator of θ and let $L(\theta, \delta)$ be a loss function. Then the associated risk function is given by*

$$R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))].$$

Definition 1.29. *An estimator δ^* of $\tau(\theta) \in \mathbb{R}^1$ ($\theta \in \mathbb{R}^k$) is minimax with respect to a risk function $R(\theta, \delta)$ if*

$$\sup_{\theta\in\Theta} R(\theta, \delta^*) = \inf_{\delta} \sup_{\theta\in\Theta} R(\theta, \delta).$$

Definition 1.30.

a) As estimators of $\tau(\theta)$, an estimator δ^* dominates δ if $R(\theta, \delta^*) \leq R(\theta, \delta)$ $\forall \theta \in \Theta^k$, and with strict inequality for at least one θ.

b) An estimator δ is inadmissible if there exists another estimator δ^* which dominates δ. Otherwise, δ is said to be admissible.

Definition 1.31. *For $X|\theta \sim f(x|\theta)$, prior density $\pi(\theta)$ and loss function $L(\theta, \delta)$ in estimating $\tau(\theta)$, we call $\delta_\pi(x)$ a Bayes estimator of $\tau(\theta)$ if it minimizes $\mathbb{E}[L(\theta, \delta)|x]$ in δ.*

Lemma 1.32. *Let δ^* be a minimax estimator and let δ' dominate δ^*. Then δ' is also minimax.*

Proof. Let δ^* be a minimax estimator of θ. Then

$$\sup_{\theta \in \Theta} R(\theta, \delta^*) = \inf_{\delta} \sup_{\theta \in \Theta} R(\theta, \delta).$$

Moreover, suppose that δ' dominates δ^*; that is, $R(\theta, \delta') \leq R(\theta, \delta^*)$ for all $\theta \in \Theta$, and with strict inequality for at least one $\theta \in \Theta$. This implies that

$$\sup_{\theta \in \Theta} R(\theta, \delta') \leq \sup_{\theta \in \Theta} R(\theta, \delta^*) = \inf_{\delta} \sup_{\theta \in \Theta} R(\theta, \delta),$$

with equality since δ^* is minimax. Thus, δ' is also minimax. $\square$

An illustration of the result above can be found in Example 1.36.

Lemma 1.33. *Under squared error loss, $L(\theta, \delta) = (\delta - \theta)^2$, the Bayes estimator $\delta_\pi(x)$ is the posterior expectation of θ; that is, $\delta_\pi(x) = \mathbb{E}[\theta|x]$, provided that $\mathbb{E}[\theta^2|x] < \infty$.*

Proof. We have

$$\mathbb{E}\left[(\delta - \theta)^2|x\right] = \delta^2 - 2\,\delta\,\mathbb{E}[\theta|x] + \mathbb{E}[\theta^2|x],$$

which can easily be differentiated and found to be minimized for $\delta = \mathbb{E}[\theta|x]$. $\square$

Remark 1.34. *Note that for $X \sim N(\theta, I)$, $L(\theta, \delta) = |\theta - \delta|^2$ and prior π, the Bayes estimator for θ can be written as $\delta_\pi(x) = x + \nabla \log m_\pi(x)$, where $m_\pi(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$ is the marginal density of X and ∇ represents the gradient operator.*

1.5.3 Estimation in restricted parameter spaces

Literature on estimation in restricted parameter spaces is plentiful and dates back to the 1950s. We refer the reader to van Eeden [vE06] and Marchand & Strawderman [MS04] for an overview of past results. One important element to consider when looking at estimation in restricted parameter spaces is that good properties of estimators in the unrestricted space (one thinks of admissibility and minimaxity for example) do not necessarily carry over to the restricted space. Perhaps the most foundational example of estimation in restricted parameter spaces concerns a positive normal mean.

Example 1.35 (van Eeden [vE06])**.** *For $X \sim N(\theta, 1)$ and squared error loss, the MLE of θ in the unrestricted parameter space ($\theta \in \mathbb{R}$), namely X, is unbiased, admissible and minimax. Moreover, it is normally distributed. However, for the restricted space $\theta \geq 0$, the (restricted) MLE is $\max\{0, X\}$, and it clearly dominates X since the latter takes values outside of the parameter space and $L(\theta, \hat{\theta}_{MLE}(X)) \leq L(\theta, X)$ for all $x < 0, \theta \geq 0$, with equality if and only if $x \geq 0, \theta \geq 0$. Katz [Kat61] showed that the MLE is minimax, but it is not admissible (Sacks [Sac63]). It is also a biased estimator of θ and certainly does not follow a normal distribution. (See Lehmann & Casella [LC98] Example 2.8 and Example 2.9 where the information inequality is used for such proofs. A more general method of showing admissiblity is Blyth's mehtod, which uses sequences of priors (Lehmann & Casella [LC98], Theorem 7.13).)*

Another significant estimator concerning a positive normal mean which is discussed by Katz [Kat61] is the Bayes estimator under squared error loss with respect to a uniform prior on the restricted parameter space.

Example 1.36. *Let $X \sim N(\theta, \sigma^2)$ with improper prior $\pi(\theta) = \mathbb{1}_{[0,\infty)}(\theta)$. The posterior* *distribution of θ is given by*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} = \frac{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\theta)^2}{2\sigma^2}}\mathbb{1}_{[0,\infty)}(\theta)}{\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\theta)^2}{2\sigma^2}}d\theta} = \frac{\frac{1}{\sigma}\phi\left(\frac{\theta-x}{\sigma}\right)}{\Phi\left(\frac{x}{\sigma}\right)}\mathbb{1}_{[0,\infty)}(\theta),$$

which one recognizes as a TN$(x, \sigma^2, (0,\infty))$ distribution. The Bayes estimator under squared error loss is therefore given by Lemma 1.7 to be $\delta_\pi(x) = \mathbb{E}[\theta|x] = x + \sigma R\left(\frac{x}{\sigma}\right)$, and can be shown to dominate the unrestricted MLE X, and is thus minimax, and admissible.

The previous examples have dealt with one-sample problems, where a normal mean has been bounded below. The focus of this work however concerns a difference of normal means being bounded. It is therefore of interest to present some theory and examples of such situations. In a one-dimensional setting, the different plausible possibilities of parametric bounds are quite limited. The parameter can be lower-bounded, $\theta \geq m$, or doubly-bounded, $-m \leq \theta \leq m$. For an upper-bounded parameter, a simple transformation of variables (change of signs) easily turns this case into that of a lower-bounded parameter. In higher dimensions, the possibilities are much more plentiful. In a k-dimensional setting, there can be a complete ordering of parameters, $\theta_1 \leq \ldots \leq \theta_k$, or an incomplete ordering such as a tree-ordering $\theta_1 \leq \theta_i$ for all $i \in \{2, \ldots, k\}$ or an umbrella-ordering $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_i \geq \theta_{i+1} \geq \ldots \geq \theta_k$ to name just a few (see van Eeden [vE06]). We will later focus on a two-dimensional problem and consider bounds of the form $\theta_1 - \theta_2 \geq m$ (Chapter 2) and $|\theta_1 - \theta_2| \leq m$ (Chapter 3).

Example 1.37. *Let $X_i \sim N(\theta_i, \sigma_i^2)$, $i = 1, 2$, independent with σ_i^2 known. For the unrestricted parameter space $(\theta_1, \theta_2) \in \mathbb{R}^2$, the MLE is simply (X_1, X_2). For the restricted parameter space $\theta_1 \geq \theta_2$, it is convenient to consider the problem of observing $\left(Y_1 = \frac{X_1 - X_2}{2}, Y_2 = \frac{X_1 + X_2}{2}\right)$ instead of (X_1, X_2), and thus estimating $\left(\mu_1 = \frac{\theta_1 - \theta_2}{2}, \mu_2 = \frac{\theta_1 + \theta_2}{2}\right)$ (where $\mu_1 \geq 0$ and $\mu_2 \in \mathbb{R}$) instead of (θ_1, θ_2). We have $\hat{\mu}_{1,MLE}(X) = \max\left\{0, \frac{X_1 - X_2}{2}\right\}$ and $\hat{\mu}_{2,MLE}(X) = \frac{X_1 + X_2}{2}$. Since $(\theta_1, \theta_2) = (\mu_1 + \mu_2, \mu_2 - \mu_1)$,* *we obtain*

$$\begin{aligned}\hat{\theta}_{MLE}(X) &= \big(\hat{\mu}_{1,MLE}(X) + \hat{\mu}_{2,MLE}(X), \hat{\mu}_{2,MLE}(X) - \hat{\mu}_{1,MLE}(X)\big) \\ &= \begin{cases} (X_1, X_2) & \textit{if } X_1 \geq X_2 \\ \left(\frac{X_1+X_2}{2}, \frac{X_1+X_2}{2}\right) & \textit{else.} \end{cases}\end{aligned}$$

Estimation of the larger of two normal means is a problem that has been well studied, and results pertaining to it can be readily found. Two situations can be considered : 1) the simultaneous estimation of the two means, and 2) the estimation of the larger (or smaller) of the two means. We proceed with a brief account of related results and refer the reader to van Eeden [vE06], van Eeden & Zidek [vEZ02] and Marchand & Strawderman [MS04] for a more complete picture. While there exist many estimators in the literature, we will focus on three of them, namely the unrestricted MLE (δ_{ULE}), the MLE (δ_{MLE}) and the Bayes estimator with respect to a uniform prior on the restricted parameter space, often called the Pitman estimator (δ_P).

- Blumenthal & Cohen [BC68b] consider the simultaneous estimation of (θ_1, θ_2) under squared error loss and show that δ_P is minimax and admissible on $\mathbb{R}^2$.
- Patra & Kumar [PK17] extend the minimaxity and admissibility properties of δ_P to cover a correlated bivariate normal model.
- Cohen & Sackrowitz [CS70] consider the estimation of the larger of two normal means and show that δ_P is minimax and admissible on $\mathbb{R}^2$. Note that their estimator is incorrect, as is mentioned by van Eeden & Zidek [vEZ02]. These same authors provide a simpler proof of the admissibility result and refer to Kumar & Sharma [KS88] for a more straightforward proof of the minimaxity result.
- Van Eeden & Zidek [vEZ02] also state the following results in estimating the smaller of two normal means : δ_{MLE} and δ_{ULE} are minimax, δ_P and δ_{MLE} dominate δ_{ULE}, and δ_{MLE} is inadmissible.

A very useful result comes from the well-established rotation technique (see Blumenthal & Cohen [BC68a] for an early use of this technique for ordered translation parameters). Simply put, this technique reduces a two-dimensional problem to one dimension. Consider the problem of estimating θ_1 under squared error loss $L(\theta, d) = (d - \theta_1)^2$ with $X_1 \sim N(\theta_1, \sigma_1^2)$, $X_2 \sim N(\theta_2, \sigma_2^2)$ independent, σ_1^2, σ_2^2 known, and with the additional prior information $\theta_1 - \theta_2 \in A \subset \mathbb{R}$. As reviewed by Marchand & Strawderman [MS04], one can define the following class of estimators :

$$\mathcal{C}_1 = \left\{ \delta_\phi(X) = Y_2 + \phi(Y_1) \text{ where } Y_1 = \frac{X_1 - X_2}{1+\tau}, Y_2 = \frac{\tau X_1 + X_2}{1+\tau}, \text{ and } \tau = \frac{\sigma_2^2}{\sigma_1^2} \right\}, \quad (1.5)$$

where $\phi(\cdot)$ is some function. Under such a model for X, it is easy to verify that Y_1 and Y_2 are independently distributed (this follows from $\mathrm{Cov}(Y_1, Y_2) = 0$ since $(Y_1, Y_2)^\top$ is a bivariate normal distribution) 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

$$\begin{aligned}
R(\theta, \delta_\phi(X)) &= \mathbb{E}_\theta\left[(Y_2 + \phi(Y_1) - \theta_1)^2\right] \\
&= \mathbb{E}_\theta\left[\left(Y_2 + \phi(Y_1) - \left(\frac{\tau\theta_1 + \theta_2}{1+\tau} + \frac{\theta_1 - \theta_2}{1+\tau}\right)\right)^2\right] \\
&= \mathbb{E}_\theta\left[\left(\left(Y_2 - \frac{\tau\theta_1 + \theta_2}{1+\tau}\right) + \left(\phi(Y_1) - \frac{\theta_1 - \theta_2}{1+\tau}\right)\right)^2\right] \\
&= \mathbb{E}_\theta\left[(Y_2 - \mu_2)^2 + (\phi(Y_1) - \mu_1)^2 + 2(Y_2 - \mu_2)(\phi(Y_1) - \mu_1)\right] \\
&= \mathbb{E}_\theta\left[(Y_2 - \mu_2)^2\right] + \mathbb{E}_\theta\left[(\phi(Y_1) - \mu_1)^2\right], \qquad (1.6)
\end{aligned}$$

since Y_1 and Y_2 are independent. 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 (since the first term of (1.6) is not affected by the choice of $\delta_\phi \in \mathcal{C}_1$). This leads immediately to the following result which is also reviewed by Marchand & Strawderman [MS04].

Lemma 1.38. *For estimating θ_1 under the model $X_1 \sim N(\theta_1, \sigma_1^2)$, $X_2 \sim N(\theta_2, \sigma_2^2)$ independent, σ_1^2, σ_2^2 known, squared error loss $L(\theta_1, d) = (d - \theta_1)^2$, and 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 $\mu_1 \in \mathcal{C} = \{y : (1+\tau)y \in A\}$.*

Remark 1.39. *Of particular interest is the choice $\delta_{\phi_0}(X) = X_1$, the unrestricted MLE of θ_1, which is obtained by taking $\phi(Y_1) = Y_1$.*

Under the setting of the rotation technique presented here and by virtue of Lemma 1.38, any estimator which dominates in one dimension will also dominate in the two-sample problem. Much work has been done on such one-sample problems, one of which is Kubokawa's method, reviewed by Marchand & Strawderman [MS04] (also see Marchand & Strawderman [MS05]) and given below for the case of a normal model.

Theorem 1.40. *Let $W \sim N(\varepsilon, \sigma_W^2)$ with $\varepsilon \geq 0$. Then estimators $\delta_h(W) = W + h(W)$ dominate $\delta_0(W) = W$ under squared error loss whenever h is absolutely continuous, decreasing on $\mathbb{R}$ and $0 \leq h(w) \leq h_U(w)$ for all $w \in \mathbb{R}$ ($\delta_h \neq \delta_0$), and where $h_U(w) = \sigma_W R\left(\frac{w}{\sigma_W}\right)$.*

Proof. This is a particular case of a more general result which is stated and proved by Marchand & Strawderman [MS04]. The expression for h_U can be found in Example 1.36 as part of the Bayes estimator for W under a uniform prior on the restricted parameter space. $\square$

This generates many improvements on X_1 for estimating θ_1 when $\theta_1 \geq \theta_2$ which is summarized in the following corollary based on Lemma 1.38 and Theorem 1.40.

Corollary 1.41. *Estimators $\delta_h(Y_1) = Y_1 + h(Y_1)$ dominate $\delta_0(Y_1) = Y_1$ in estimating $\mu_1 = \frac{\theta_1 - \theta_2}{1+\tau}$ whenever h is absolutely continuous, decreasing on $\mathbb{R}$ and $0 \leq h(y_1) \leq h_U(y_1)$ for all $y_1 \in \mathbb{R}$ ($\delta_h \neq \delta_0$), and where $h_U(y_1) = \sigma_{Y_1} R\left(\frac{y_1}{\sigma_{Y_1}}\right)$. Consequently, estimators $\delta_\phi(X) = Y_2 + \delta_h(Y_1)$ dominate X_1 in estimating θ_1.*

Maruyama & Iwasaki [MI05] provide such estimators of μ_1, including Bayes and admissible estimators.

The results above pertain to point estimation in restricted parameter spaces; however, much work has also been carried out for interval estimation (e.g., Gelfand [GSL92]). It has already been mentioned that Bayesian credible sets do not guarantee matching frequentist coverage probability (they are not designed to do so), a class of exceptions being given by Example 1.27. This exception holds more generally for location and scale models with no 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 and lower coverage than credibility occurring (e.g., Mandelkern [Man02]; Marchand & Strawderman [MS06]). It is nonetheless of interest to study the frequentist coverage probability of Bayesian credible sets (e.g., Fraser [Fra11]; Wasserman [Was11]). Example 1.42 presents a situation where a lower bound for the frequentist coverage can be obtained.

Example 1.42 (Zhang & Woodroofe [ZW03]; Marchand & Strawderman [MS06]; Marchand et al. [MSBL08]). *For* $X \sim N(\theta, \sigma^2)$, $W \sim Ga(r/2, 2\sigma^2)$ *and prior* $\pi(\theta, \sigma) = \frac{1}{\sigma}\mathbb{1}_{[a,\infty)}(\theta)\mathbb{1}_{(0,\infty)}(\sigma)$, *a lower bound for the frequentist coverage probability of the* $100 \times (1-\alpha)\%$ *HPD credible set* $I_\pi(X, W)$ *is given by (Zhang & Woodroofe [ZW03])*

$$\mathbb{P}_{\theta,\sigma}\left[\theta \in I_\pi(X, W)\right] \geq \frac{1-\alpha}{1+\alpha},$$

for all $\theta \geq a$ *and* $\sigma > 0$. *Marchand & Strawderman [MS06] extend such a result to a more general setting, and Marchand et al. [MSBL08] improve on this lower bound (to* $1 - \frac{3\alpha}{2}$ *for* $\alpha < \frac{1}{3}$*) for logconcave pivotal densities such as the one related to a* $N(\theta, \sigma^2)$ *model with known* σ^2 *and prior density* $\pi(\theta) = \mathbb{1}_{(0,\infty)}(\theta)$.

1.5.4 Uncertain parametric restrictions

As mentioned earlier, when additional information is known about the parameters, one wishes to profit from this information. Some parametric constraints are known with certainty and all data that could possibly be collected would never contradict them. Such certainty in the additional information is not necessarily at hand though. For example, experts might compare crops with two different fertilizers and have a strong prior belief that one fertilizer outperforms the other, but it remains possible to obtain data which contradicts this belief. In this case, does one wish to allow the estimates to go against the parametric constraint ?

This question arose in the early 2000s under the context of two high-energy physics problems (see Mandelkern [Man02]). The first involved estimating the mass of a neutrino. It had long been believed that neutrinos were massless, but calibration was yielding data that indicated otherwise. More specifically, this setting involved estimating θ where $X \sim N(\theta, \sigma^2)$, σ^2 is known and $\theta \geq 0$. The second setting concerned the estimation of θ, a signal mean, under a model where $X \sim \text{Poisson}(\theta + b)$, $b > 0$ is a known background mean and $\theta \geq 0$.

O'Hagan & Leonard [OL76] address this issue of data conflicting the model and suggest a hierarchical prior to account for the uncertainty in the constraint. Doing so 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. Liseo & Loperfido [LL03] extend results from O'Hagan & Leonard [OL76] and consider uncertain linear restrictions.

Example 1.43. *Marchand & Nicoleris [MN19] consider such a hierarchical prior for a* *lower-bounded normal mean, where the model and prior are given by*

$$X|\theta, m \sim N(\theta, \sigma^2), \quad \theta|m \sim N(\mu, \tau^2)\mathbb{1}_{[m,\infty)}(\theta), \quad m \sim N(0, \sigma_m^2),$$

with μ and $\sigma^2, \tau^2, \sigma_m^2 \geq 0$ known. The case $\sigma_m^2 = 0$ recovers the deterministic constraint $\theta \geq 0$ (which corresponds to the problem considered by Katz [Kat61]), while $\tau^2 \to \infty$ yields a flat prior on the restricted parameter space.

Marchand & Nicoleris [MN19] proceed to present a class of minimax Bayes estimators, the result of which is given below.

Theorem 1.44 (Marchand & Nicoleris [MN19])**.** *For $X \sim N(\theta, \sigma^2)$, squared error loss $L(\theta, d) = (d - \theta)^2$ and parametric restriction $\theta \geq 0$, the class of estimators $\delta_c(X) = X + c\sigma R\left(\frac{cX}{\sigma}\right)$, $c \in (0, 1]$, dominates $\delta_0(X) = X$. Moreover, this class of estimators contains Bayes point estimators of θ under the hierarchical prior $\pi(\theta|m) = \mathbb{1}_{[m,\infty)}(\theta)$ with $m \sim N(0, \sigma_m^2)$, namely δ_c, with $c = \frac{\sigma}{\sqrt{\sigma^2 + \sigma_m^2}}$.*

Proof. Their proof relies on the application of Stein's integration by parts identity and then change of sign arguments, and can be found in [MN19]. □

Remark 1.45. *Note that this class of estimators provided by Marchand & Nicoleris [MN19] does not fall in the class of dominating estimators given in Theorem 1.40 since the condition $h(x) \leq h_U(x)$ for all $x \in \mathbb{R}$ is not satisfied.*

In terms of interval estimation under uncertain parametric restrictions, we refer the reader to Madi et al. [MLT00] among others.

CHAPTER 2

Estimating the suspected larger of two normal means

This chapter forms the basis of a recently submitted paper co-authored by Courtney Drew and Éric Marchand.

Abstract

For X_1, X_2 independently 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 $1-\alpha$ credibility 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.

2.1 Introduction

It has long been known, for a bivariate normal model with X_1 and 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 a 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 & Sackrowitz [CS70]). However, there are situations where one would not expect this bound to hold exactly, and one could envisage introducing uncertainty in the parametric constraint. This has been previously proposed (see O'Hagan & Leonard [OL76] for an early reference where uncertainty is introduced through a hierarchical prior, as well as Liseo & Loperfido [LL03] for a later reference with 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.

In this work, 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 & Cohen [BC68a]) and a one-sample minimax finding by Marchand & Nicoleris [MN19] 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 case 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 the known analyses for $\sigma_m^2 = 0$ carried out previously by Cohen & Sackrowitz [CS70], van Eeden & Zidek [vEZ02], and Kumar & Sharma [KS88], 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 with no 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 (Mandelkern [Man02] ; Marchand & Strawderman [MS06]). We point out that there has been much work on evaluating and computing Bayesian posterior densities and estimates in the presence of parametric restrictions, notably for ordered parameters with or without nuisance parameters (e.g., Gelfand et al. [GSL92] ; Madi et al. [MLT00]).

We introduce below an *ad hoc* Bayes credible set with approximate $1 - \alpha$ credibility (based again on the prior $\pi(\theta_1, \theta_2 \,|\, m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2)$ 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 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 probability. Namely, as illustrated by Marchand & Strawderman [MS13], as well as Ghashim et al. [GMS16], we believe that the characterization or determination of such Bayes credible sets through a spending function is impactful and merits to be considered. Thus, the analysis and illustrations presented here also make use of the concept of a spending function, the choice of which is guided.

This chapter is organized as follows. After having extracted and interpreted some useful properties of the posterior distributions in Section 2.2, which relate to extended skew-normal densities, the dominance and minimax results are presented and commented on in Section 2.3. Related questions of admissibility are also addressed. Section 2.4 deals with proposed Bayes credible sets for θ_1, focusing mostly on their frequentist coverage probability, but also discussing their expected length and exact credibility. The findings are commented on at length and illustrated with several figures. Section 2.4.1 defines and explores an ad hoc credible interval and Section 2.4.2 expands on modifications which make use of the concept of a spending function. Concluding remarks in Section 2.5 summarize the findings and implications.

2.2 Posterior analysis

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

$$\begin{aligned} &X_1 \sim N(\theta_1, \sigma_1^2), X_2 \sim N(\theta_2, \sigma_2^2); \\ &\pi(\theta_1, \theta_2 | m) = \mathbb{1}_{[m,\infty)}(\theta_1 - \theta_2), m \sim N(0, \sigma_m^2), \end{aligned} \tag{2.1}$$

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, but where the bound is not deterministic ; that is, there is some uncertainty on m. That being said, on average, $m = 0$. For $\sigma_m > 0$, an alternative and equivalent representation of the prior in (2.1) is readily obtained by integrating out m and yielding the improper density $\pi(\theta_1, \theta_2) = \Phi\left(\frac{\theta_1 - \theta_2}{\sigma_m}\right)$.

Remark 2.1.

(a) *The situation given by* (2.1) *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* $X_2' \sim N(\mu_2 = c\,\theta_2, \sigma_2^{2'} = c^2\sigma_2^2)$.

(b) *Similarly, analysis for model* (2.1) *will yield applications for correlated normally distributed variables, specifically for*

$$W = (W_1, W_2)^\top \sim N_2\left(\xi = \begin{bmatrix}\xi_1 \\ \xi_2\end{bmatrix}, \Sigma = \begin{bmatrix}\tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2\end{bmatrix}\right),$$

with $\xi_1 - \xi_2 \geq m$, *correlation coefficient* $\rho = \rho(W_1, W_2) \in (-1, 1)$, *such that* $\lambda = \rho\frac{\tau_1}{\tau_2} \neq 1$. *This is achieved by setting* $X_1 = W_1 - \lambda W_2$ *and* $X_2 = W_2$, *which implies that* $X_1 \sim N(\xi_1 - \lambda\xi_2, \tau_1^2(1-\rho^2))$ *and* $X_2 \sim N(\xi_2, \tau_2^2)$, *and whereupon part (a) applies with* $\theta_1 = \xi_1 - \lambda\xi_2$, $\theta_2 = \xi_2$, $c = (1-\lambda)$, $\sigma_1^2 = \tau_1^2(1-\rho^2)$ *and* $\sigma_2^2 = \tau_2^2$.

Remark 2.2. *There exist many instances where summary statistics are well modeled by normal observables such as in* (2.1) *or the variants of Remark 2.1. Common occurrences* *arise through sufficiency or asymptotically justified approximations. An example emerges in a basic linear model $W \sim N_n(Z^\top \beta, \sigma^2 I_n)$ with $Z(n \times p)$ of full rank p, the least squares estimator $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^\top = (Z^\top Z)^{-1} Z^\top W$, and the summary statistics $X_1 = \hat{\beta}_1$ 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* (2.1) *applies whether or not $\hat{\beta}_1$ and $\hat{\beta}_2$ are correlated.*

Theorem 2.3. *Under the model and prior given by* (2.1)*, the joint posterior density of (θ_1, θ_2) is given by*

$$\pi(\theta_1, \theta_2 | x) = \frac{\frac{1}{\sigma_1 \sigma_2} \phi\left(\frac{\theta_1 - x_1}{\sigma_1}\right) \phi\left(\frac{\theta_2 - x_2}{\sigma_2}\right) \Phi\left(\frac{\theta_1 - \theta_2}{\sigma_m}\right)}{\Phi\left(\frac{x_1 - x_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_m^2}}\right)}, \quad \textit{for } (\theta_1, \theta_2)^\top \in \mathbb{R}^2.$$

Proof. This follows from writing the joint posterior density of (θ_1, θ_2) as

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

where

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

and using Lemma 1.10 to evaluate the integrals. □

Theorem 2.4. *Under the model and prior given by* (2.1)*, the marginal posterior density of $U = \frac{\theta_1 - x_1}{\sigma_1}$ is given by*

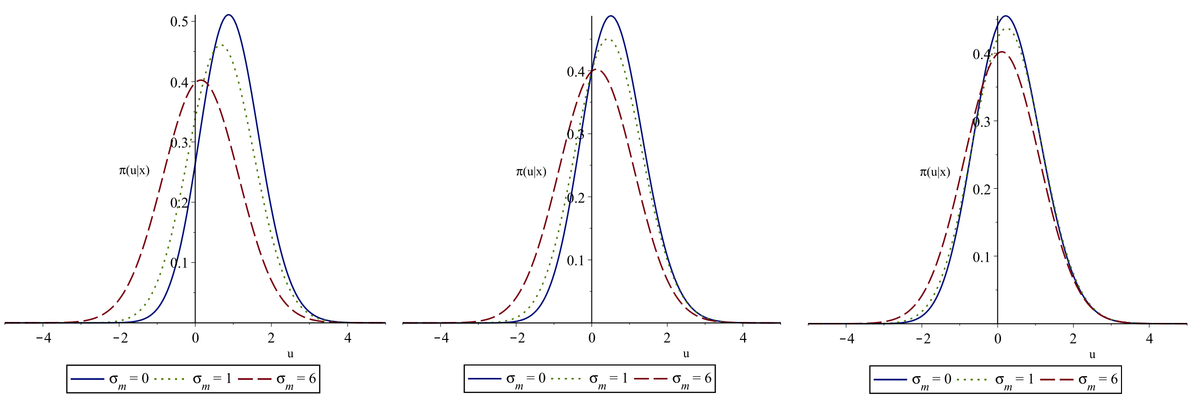
$$\pi(u|x) = \frac{\phi(u) \cdot \Phi\left(\frac{\sigma_1 u + d}{\sqrt{\sigma_2^2 + \sigma_m^2}}\right)}{\Phi\left(\frac{d}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_m^2}}\right)}, \tag{2.2}$$

where $d = x_1 - x_2$ and for $u \in \mathbb{R}$.

Proof. This follows from Theorem 2.3 and again using Lemma 1.10 to evaluate the integral over θ_2. □

One recognizes the posterior density in (2.2) as a $SN\left(\frac{\sigma_1}{\sqrt{\sigma_2^2+\sigma_m^2}}, \frac{d}{\sqrt{\sigma_2^2+\sigma_m^2}}\right)$, as given by Definition 1.8. The first parameter is always positive, which implies that the posterior density is always positively skewed. Note that the posterior density of U depends on the observations x_1 and x_2, but solely through their difference $d = x_1 - x_2$. Also note that this density holds for all $\sigma_m^2 \geq 0$; that is, whether or not there is uncertainty in the bound.

Remark 2.5. *As σ_m^2 increases, Eq. (2.2) tends towards a standard normal density; that is, $\lim_{\sigma_m^2 \to \infty} \pi(u|x) = \phi(u)$. Heuristically, the case $\sigma_m^2 \to \infty$ corresponds to an absence of additional information provided by X_2, and in such a case, a $N(x_1, \sigma_1^2)$ posterior would be expected for θ_1, which translates to a standard normal posterior for $U = \frac{\theta_1 - x_1}{\sigma_1}$.*



(a) $d = -1$ (b) $d = 0$ (c) $d = 1$

Figure 2.1 – Posterior density of U for $d = -1, 0, 1$, $\sigma_1^2 = \sigma_2^2 = 1$ and varying σ_m.

Figure 2.1 presents the posterior density $\pi(u|x)$ for varying σ_m and for $d = -1, 0, 1$. The behaviour described in Remark 2.5 as σ_m^2 increases is apparent in these graphs. Moreover, a similar observation can be made as d increases (i.e., the posterior density becomes less skewed) since $\lim_{d \to \infty} \pi(u|x) = \phi(u)$.

One readily obtains the posterior moment generating function, expectation and variance of U using Lemma 1.11.

Lemma 2.6. *Under the context of Theorem 2.4, the posterior moment generating function, expectation and variance of* U *are given respectively by*

$$M_{U|x}(t) = \frac{e^{\frac{t^2}{2}}}{\Phi\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)}\Phi\left(\frac{t\sigma_1 + d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right),$$

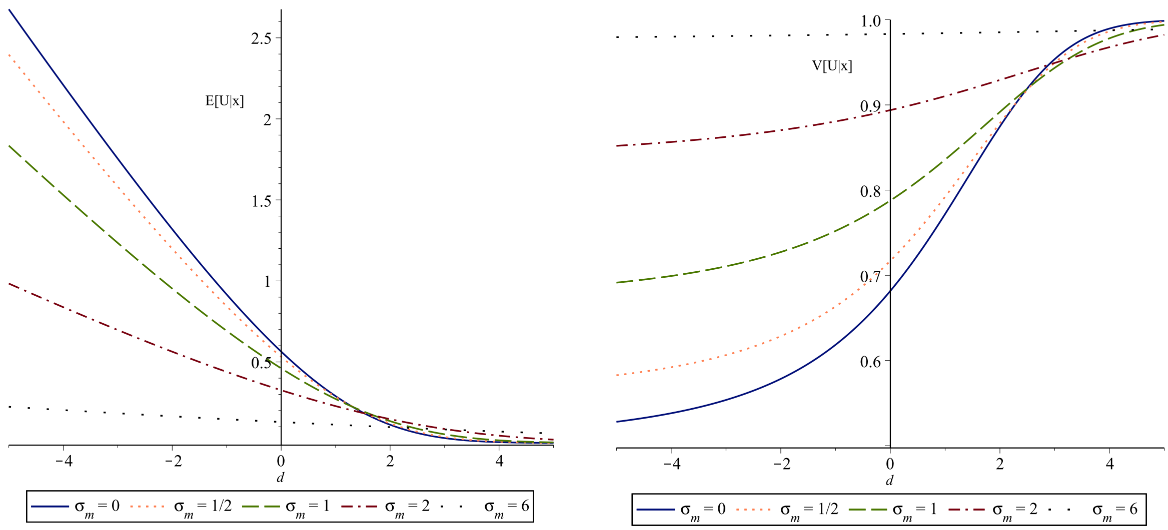
$$\mathbb{E}[U|x] = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}R\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right), \text{ and}$$

$$\mathbb{V}[U|x] = 1-\frac{\sigma_1^2 d}{(\sigma_1^2+\sigma_2^2+\sigma_m^2)^{3/2}}R\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)-\frac{\sigma_1^2}{(\sigma_1^2+\sigma_2^2+\sigma_m^2)}R^2\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right),$$

where $R(t) = \frac{\phi(t)}{\Phi(t)}$ *is the reverse Mill's ratio given in Definition 1.15.*

Proof. These can be obtained directly from Lemma 1.11. □

In Section 2.4, 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. Figure 2.2 presents the posterior expectation and variance of U as a function of $d = x_1 - x_2$ for fixed $\sigma_1^2 = \sigma_2^2 = 1$ and varying σ_m.



(a) Posterior expectation

(b) Posterior variance

Figure 2.2 – Posterior expectation and variance of U as a function of $d = x_1 - x_2$ for $\sigma_1^2 = \sigma_2^2 = 1$ and varying σ_m.

Several properties of the posterior expectation and variance become apparent in Figure 2.2, and we summarize these features in the following result.

Lemma 2.7. *In the setting of Lemma 2.6, 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 given in Lemma 1.16.

(a) This follows from part (b) of Lemma 1.16 and writing the limit as

$$\lim_{d\to-\infty}\frac{\mathbb{E}[U|x]}{d}=\lim_{d\to-\infty}\frac{\sigma_1}{\sigma_1^2+\sigma_2^2}\frac{R\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}{\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2}}}=-\frac{\sigma_1}{\sigma_1^2+\sigma_2^2},$$

where we use part (f) of Lemma 1.16.

(b) We have

$$\mathbb{V}[U|x]=1-\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2+\sigma_m^2}R\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)$$
$$\cdot\left[\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}+R\left(\frac{d}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)\right],$$

from which we can conclude using parts (b), (c), (d) and (e) of Lemma 1.16.

(c) This follows from parts (b) and (c) of Lemma 1.16.

□

Remark 2.8. *The case $\sigma_m = 0$ warrants particular attention since this corresponds to the case where there is no uncertainty in the bound and the parametric restriction is reduced to $\theta_1 \geq \theta_2$. One recovers results for this degenerate case in the literature, notably in Cohen & Sackrowitz [CS70] and Blumenthal & Cohen [BC68b]. Moreover, the case $\sigma_m \to \infty$ corresponds to an absence of additional information. As carried out in previous work (e.g., van Eeden & Zidek [vEZ02]), it is useful to consider heuristics related to these limiting cases in order to gain an understanding of the contribution of the additional information.*

(a) If $\sigma_m = 0$ and $x_1 \gg x_2$, then $d = x_1 - x_2$ is large and under the assumption of our model that $\theta_1 \geq \theta_2$, x_2 provides very little additional information. We would therefore expect to obtain results similar to, and equal to in the limiting case, those that are obtained if we only had information on x_1. 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 (2.2) 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

$$X_i|\theta_1 \sim N(\theta_1, \sigma_i^2),$$

with X_i 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 implies that for very small d, we have

$$\mathbb{E}[\theta_1|x] \approx \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \quad \textit{and} \quad \mathbb{V}[\theta_1|x] \approx \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.3}$$

From this, since $U = \frac{\theta_1 - x_1}{\sigma_1}$, we have for very small d that

$$\mathbb{V}[U|x] = \frac{1}{\sigma_1^2}\mathbb{V}[\theta_1|x] \approx \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

which matches the limiting value as $d \to -\infty$ given in part (b) of Lemma 2.7 (taking $\sigma_m = 0$). Additionally, this relates to part (a) of Lemma 2.7 since for very small d and $\sigma_m = 0$, (2.3) gives

$$\mathbb{E}\left[\frac{U}{d}|x\right] = \mathbb{E}\left[\frac{\theta_1 - x_1}{d\sigma_1}|x\right] \approx \frac{1}{d\sigma_1}\left(\frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} - x_1\right) = -\frac{\sigma_1}{\sigma_1^2 + \sigma_2^2}.$$

Finally, throughout this work we focus on the impact of σ_m^2, but one can also explore the effect of σ_1^2 and σ_2^2. If $\sigma_m^2 = 0$, and $\sigma_2^2 \gg \sigma_1^2$, then X_2 has much more variability than X_1, which implies that the additional information provided by x_2 is less valuable. Similar heuristic arguments can be made to those in part (a) of Remark 2.8 for limiting cases.

2.3 Point estimation

This section concerns the efficiency of point estimators of θ_1 for model (2.1). We obtain a class of Bayesian estimators that dominate X_1. From Cohen & Sackrowitz [CS70], it is known that X_1 is minimax and admissible for $\theta_1 \geq \theta_2$, which renders our class of estimators also minimax. The result stems from the rotation technique and is obtained by combining Lemma 1.38 with the one-sample result (in the presence of uncertainty in the parametric constraint) by Marchand & Nicoleris [MN19], given by Theorem 1.44.

Theorem 2.9. *Let X be distributed according to model* (2.1), $\tau = \frac{\sigma_2^2}{\sigma_1^2}$, *with squared error loss for estimating θ_1, $L(\theta_1, d) = (d - \theta_1)^2$. Then under the additional information $\theta_1 - \theta_2 \geq 0$, estimators of the form*

$$\delta_{\phi_c}(X) = X_1 + \frac{c\,\sigma_1}{\sqrt{1+\tau}} R\left(\frac{c\,(X_1 - X_2)}{\sigma_1\sqrt{1+\tau}}\right) \tag{2.4}$$

dominate X_1 for $c \in (0,1]$ and are hence minimax. 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 (2.1) *; that is,*

$$\delta_{\pi_{\sigma_m}}(X) = \mathbb{E}[\theta_1|X] = X_1 + \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_m^2}} R\left(\frac{X_1 - X_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_m^2}}\right). \tag{2.5}$$

Proof. Under the setting of (1.5), Theorem 1.44 asserts that estimators of the form

$$\delta_c(Y_1) = Y_1 + c\sigma_{Y_1} R\left(\frac{cY_1}{\sigma_{Y_1}}\right) = \frac{X_1 - X_2}{1+\tau} + \frac{c\,\sigma_1}{\sqrt{1+\tau}} R\left(\frac{c\,(X_1 - X_2)}{\sigma_1\,\sqrt{1+\tau}}\right) \tag{2.6}$$

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 1.38 yields (2.4) as a class of estimators which dominate X_1 for $c \in (0,1]$. □

Theorem 2.9 provides a class of (generalized) Bayes 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, as opposed to the "certain constraint" $\theta_1 \geq \theta_2$ associated with $\sigma_m = 0$, the prior 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. 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 $\theta_1 < \theta_2$ (see Figure 2.3). 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 (see Marchand & Nicoleris [MN19]). The behaviour of Bayes credible sets for the same priors in terms of frequentist coverage probability, which will be the object of study in Section 2.4, will be analogous.

The rotation technique (more specifically Eq. (1.6)) allows one to express the risk as a function of $\mu_1 = \frac{\theta_1 - \theta_2}{1+\tau}$. Figure 2.3 does so for $\sigma_1^2 = \sigma_2^2 = 1$ and varying σ_m.

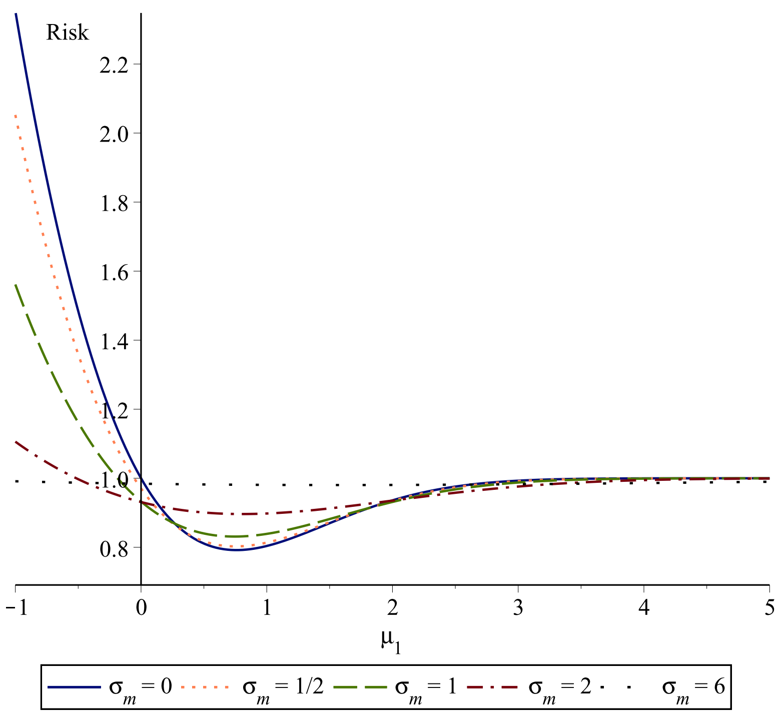


Figure 2.3 – Squared error risk as a function of $\mu_1 = \frac{\theta_1 - \theta_2}{1+\tau}$ for $\sigma_1^2 = \sigma_2^2 = 1$ and varying σ_m.

One observes in Figure 2.3 the improved performance of the Bayes estimator for $\mu_1 \leq 0$ as σ_m increases. This is due to the robustness of the estimator which follows from the increased flexibility of the model with uncertainty in the constraint. The minimax risk for these choices of parameters is 1 and is attained at 0 and $+\infty$ for $\sigma_m = 0$, but only at $+\infty$ on the restricted parameter space $\mu_1 > 0$ for other choices of σ_m (see Marchand & Nicoleris [MN19] ; Marchand & Strawderman [MS05]). The minimaxity for all choices of σ_m for $\mu_1 \geq 0$, on account of Theorem 2.9, is also visible. In fact, for $\sigma_m > 0$, the estimator remains minimax for some negative values of μ_1. For instance, the Bayes estimator with $\sigma_m = 2$ improves on the benchmark estimator X_1 for $\theta \in (\theta_0, 0)$ with $\theta_0 \approx -0.515$.

We complete this section with observations relative to the issues of admissibility and minimaxity for model (2.1), for both the unrestricted parameter space $\theta \in \mathbb{R}^2$ and the restricted parameter space $\Theta = \{\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2 : \theta_1 \geq \theta_2\}$. Moreover, we not only address the performance of estimators $\hat{\theta}_1(X)$ of θ_1 for risk $R_1(\theta, \hat{\theta}_1) = \mathbb{E}[(\hat{\theta}_1(X) - \theta_1)^2]$, but also the dual performance of estimators $\hat{\theta}_2(X)$ of θ_2 for risk $R_2(\theta, \hat{\theta}_2) = \mathbb{E}[(\hat{\theta}_2(X) - \theta_2)^2]$, and the performance of estimators $\hat{\theta}(X)$ of θ for risk $R(\theta, \hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta}(X) - \theta\right)^2\right] = R_1(\theta, \hat{\theta}_1) + R_2(\theta, \hat{\theta}_2)$.

(a) Consider the joint estimator of $(\theta_1, \theta_2)^\top$ given by

$$\delta^\pi(X) = E[\theta|X] = \left(X_1 + h(X_1 - X_2), X_2 - h(X_1 - X_2)\right)^\top, \tag{2.7}$$

where $h(t) = \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}} R\left(\frac{t}{\sqrt{\sigma_1^2+\sigma_2^2+\sigma_m^2}}\right)$.

Theorem 2.9 gives the dominance of the first component of Eq. (2.7) over X_1. In estimating θ_2, the second component of $\delta^\pi(X)$ also dominates X_2 on Θ. This can be seen by setting $Z = (Z_1, Z_2)^\top = -X$ and observing that Theorem 2.9 applies for $Z \sim N_2 \left(\mu \, , \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right)$ with $\mu = (\mu_1, \mu_2)^\top = -\theta$ and implies that

$$\begin{aligned} \mathbb{E}\left[(Z_2 + h(Z_2 - Z_1) - \mu_2)^2 \right] &\leq \mathbb{E}\left[(Z_2 - \mu_2)^2 \right] \text{ for all } \mu_1 \leq \mu_2 \\ \Longleftrightarrow R_2(\theta, X_2 - h(X_1 - X_2)) &\leq R_2(\theta, X_2) \text{ for } \theta \in \Theta \, . \end{aligned}$$

In other words, for estimating the larger of two normal means, we have that $Z_2 + h(Z_2 - Z_1) = -X_2 + h(X_1 - X_2)$ dominates Z_2 in estimating μ_2, where $\mu_1 \leq \mu_2$. Thus, $X_2 - h(X_1 - X_2)$ dominates X_2 in estimating θ_2 when $\theta_1 \geq \theta_2$.

(b) From above, it follows immediately that $\delta^\pi(X)$ dominates X as an estimator of θ for risk $R_1 + R_2$ and $\theta \in \Theta$. For $\sigma_m^2 = 0$, Blumenthal & Cohen [BC68b] showed that $\delta^\pi(X)$, which is the Bayes estimator for the uniform prior on Θ, is both minimax and admissible. The minimax property also holds for $\sigma_m^2 > 0$, and this is a novel finding. It follows since $\delta^\pi(X)$ dominates X, and X is itself minimax for $\theta \in \Theta$. Such a minimax result has been known for a long time (e.g., Blumenthal & Cohen [BC68b]) and holds in many situations (see van Eeden [vE06]; Marchand & Strawderman [MS12] for elements of review) where a minimum risk equivariant estimator designed for an untruncated parameter space remains minimax when evaluated on a parametric restriction.

Although the estimator $\delta^\pi(X)$ is designed for situations where it is only suspected that $\theta_1 \geq \theta_2$ and not certain, its minimaxity for $\theta_1 \geq \theta_2$ remains an attractive feature. For the constrained parameter space $\theta_1 \geq \theta_2$, only the estimator $\delta^\pi(X)$ corresponding to $\sigma_m^2 = 0$ remains admissible as otherwise, $\mathbb{P}_\theta[\delta^\pi(X) \notin \Theta] > 0$ for all $\theta \in \Theta$, and $\delta^\pi(X)$ can be improved uniformly in risk by taking its projection onto the constraint boundary $\theta_1 = \theta_2$. However, $\delta^\pi(X)$ could well be admissible for $\theta \in \mathbb{R}^2$ and this merits further investigation.

2.4 Bayes credible sets

Having evaluated the posterior distribution of θ_1 under model and prior (2.1) and determined the attractive performance of a class of corresponding Bayes estimators, 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. A way to observe this is through the length of the intervals, as well as by considering the 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 HPD or equal-tails for example, we focus on an ad hoc interval with approximate $1-\alpha$ credibility, mainly 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 Section 2.4.1, the ad hoc credible set studied is of a standard form $\mathbb{E}[\theta|x] \pm z_{\alpha/2}\sqrt{\mathbb{V}[\theta|x]}$ (e.g., Berger [Ber85]). In Section 2.4.2, we propose and study a modification based on the idea of a *spending function* (e.g., Marchand & Strawderman [MS13]) that shifts the above credible set towards lower values.

2.4.1 An ad hoc credible set

The Bayes credible set studied here is given by Definition 2.10.

Definition 2.10. *Let $\mathbb{E}[U|x]$ and $\mathbb{V}[U|x]$ denote respectively the posterior expectation and variance of U given by Lemma 2.6. The ad hoc Bayes credible interval for θ_1 (i.e., for $\sigma_1 U + X_1$) is defined as*

$$I_{\mathrm{ah}}(X) = [X_1 + l(X_1 - X_2), X_1 + u(X_1 - X_2)], \tag{2.8}$$

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* $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

One observes that the ad hoc interval is centered at the posterior mean of θ_1 ; that is, at $\mathbb{E}[\theta_1|x] = \sigma_1\mathbb{E}[U|x] + x_1$, and extends on either side of the posterior mean by equal amounts, namely $z_{\alpha/2}\sigma_1\sqrt{\mathbb{V}[U|x]}$. This construction is generalized in Section 2.4.2 to allow the interval to extend by different amounts on either side of the posterior mean. Also note that explicit endpoints are obtained for the ad hoc interval, up to the definition of the reverse Mill's ratio which involves a standard normal cdf.

Remark 2.11. *Limiting values of the functions* $l(d)$ *and* $u(d)$ *in Eq.* (2.8) *can easily be obtained from the properties of the posterior expectation and variance of* U *given in Lemma 2.7. Indeed, from parts (a) and (b) of Lemma 2.7, we have* $\lim_{d\to\infty} \mathbb{E}[U|x] = 0$ *and* $\lim_{d\to\infty} \mathbb{V}[U|x] = 1$, *which imply that*

$$\lim_{d\to\infty} u(d) = -\lim_{d\to\infty} l(d) = \sigma_1 z_{\alpha/2}. \tag{2.9}$$

Similarly, we have $\lim_{\sigma_m^2\to\infty} \mathbb{E}[U|x] = 0$ *and* $\lim_{\sigma_m^2\to\infty} \mathbb{V}[U|x] = 1$, *which yield*

$$\lim_{\sigma_m^2\to\infty} u(d) = -\lim_{\sigma_m^2\to\infty} l(d) = \sigma_1 z_{\alpha/2}. \tag{2.10}$$

Moreover, it can also be determined from Lemma 2.7 that $l(d)$ *is decreasing in* d *and is always greater than* $-\sigma_1 z_{\alpha/2}$. *An equivalent analysis for* $u(d)$ *is not readily available.*

Theorem 2.12 (also see Denis [Den10]) gives an expression for the frequentist coverage probability (see Definition 1.21) of a more general interval for θ_1, of which $I_{\mathrm{ah}}(X)$ is a particular case. The following result is useful for numerical evaluations.

Theorem 2.12. *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 of $I(X)$, i.e., $\mathbb{P}[\theta_1 \in I(X)]$, is given by*

$$C(\theta_1,\theta_2) = \mathbb{E}^Z\left[\Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}\,u\left\{\sqrt{\sigma_1^2+\sigma_2^2}\,Z+\beta\right\}+\frac{\sigma_1}{\sigma_2}Z\right)\right.$$
$$\left.-\Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}\,l\left\{\sqrt{\sigma_1^2+\sigma_2^2}\,Z+\beta\right\}+\frac{\sigma_1}{\sigma_2}Z\right)\right], \tag{2.11}$$

where $\beta = \theta_1 - \theta_2$ and $Z \sim N(0,1)$.

Proof. We have

$$\begin{aligned}
C(\theta_1,\theta_2) &= \mathbb{P}_\theta\left[\theta_1 \in I(X)\right] \\
&= \mathbb{P}_\theta\left[X_1 + l\{X_1 - X_2\} \le \theta_1 \le X_1 + u\{X_1 - X_2\}\right] \\
&= \mathbb{P}_\theta\left[-u\{(X_1-\theta_1)-(X_2-\theta_2)+(\theta_1-\theta_2)\} \le X_1 - \theta_1\right. \\
&\qquad \left.\le -l\{(X_1-\theta_1)-(X_2-\theta_2)+(\theta_1-\theta_2)\}\right] \\
&= \mathbb{P}_\theta\left[-u\{Y_1 - Y_2 + \beta\} \le Y_1 \le -l\{Y_1 - Y_2 + \beta\}\right],
\end{aligned}$$

where $Y_i = X_i - \theta_i \sim N(0,\sigma_i^2)$, $i = 1,2$, are independent and $\beta = \theta_1 - \theta_2$. Now, we have bounds for Y_1 which depend on $Y_1 - Y_2$, where $Y = Y_1 - Y_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$. From Lemma 1.3, we have

$$Y_1|Y \sim N\left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}Y, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right),$$

since

$$\rho_{Y_1,Y} = \frac{\mathrm{Cov}(Y_1,Y)}{\sigma_{Y_1}\sigma_Y} = \frac{\sigma_{Y_1}}{\sigma_Y} = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$$

due to the independence of Y_1 and Y_2. Now, by conditioning, we have

$$\begin{aligned}
C(\theta_1,\theta_2) &= \mathbb{P}\big[Y_1 \in [-u\{Y+\beta\}, -l\{Y+\beta\}]\big] \\
&= \mathbb{E}\big[\mathbb{1}_{[-u\{Y+\beta\},-l\{Y+\beta\}]}(Y_1)\big] \\
&= \mathbb{E}^Y\Big[\mathbb{E}\big[\mathbb{1}_{[-u\{Y+\beta\},-l\{Y+\beta\}]}(Y_1)|Y\big]\Big] \\
&= \mathbb{E}^Y\Big[\mathbb{P}\big[Y_1 \in [-u\{Y+\beta\},-l\{Y+\beta\}]|Y\big]\Big].
\end{aligned}$$

Setting $Z = \frac{Y_1 - Y_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$ and $Z' = \frac{Y_1 - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z}{\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}}$, we obtain $(Z, Z')^T \sim N_2(0, I_2)$, and the coverage can be expressed as

$$\begin{aligned}
C(\theta_1, \theta_2) &= \mathbb{P}\left[\sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}} \left(-u\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z\right) \leq Z' \right. \\
&\qquad \left. \leq \sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}} \left(-l\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z\right)\right] \\
&= \mathbb{E}^Z\left[\mathbb{P}\left[\sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}} \left(-u\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z\right) \leq Z' \right.\right. \\
&\qquad \left.\left. \leq \sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}} \left(-l\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} - \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} Z\right) \,\middle|\, Z\right]\right] \\
&= \mathbb{E}^Z\left[\Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} u\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} + \frac{\sigma_1}{\sigma_2} Z\right)\right] \\
&- \mathbb{E}^Z\left[\Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} l\left\{\sqrt{\sigma_1^2+\sigma_2^2}\, Z + \beta\right\} + \frac{\sigma_1}{\sigma_2} Z\right)\right].
\end{aligned}$$

$\square$

As a first illustration, Figure 2.4 presents the frequentist coverage probability of the ad hoc interval as a function of $\beta = \theta_1 - \theta_2$ for $\sigma_1^2 = \sigma_2^2 = 1$, a 0.95 nominal level and varying σ_m.

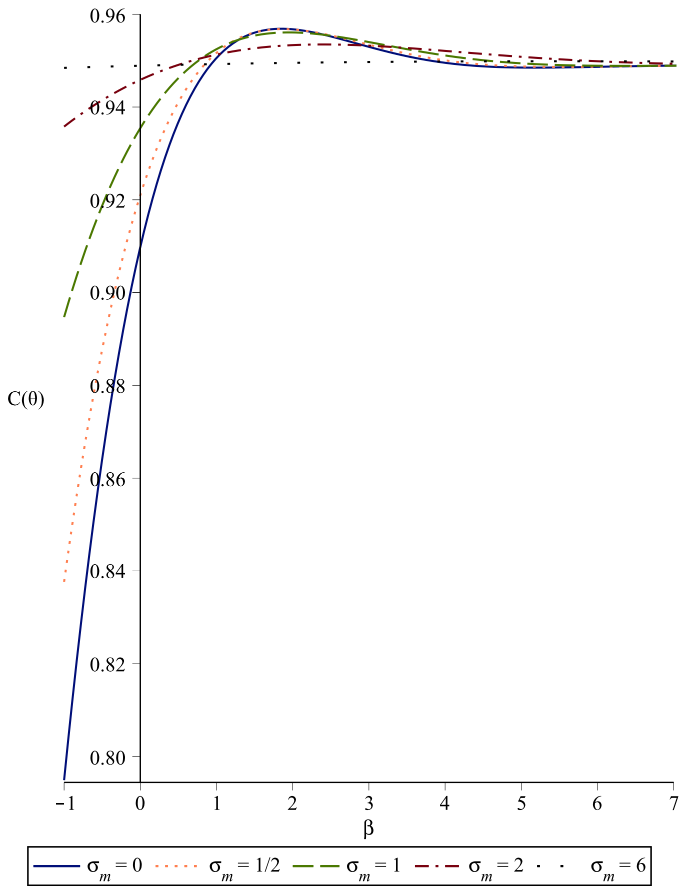


Figure 2.4 – Frequentist coverage probability of the ad hoc interval ($1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $\beta = \theta_1 - \theta_2$ for varying σ_m.

While the maximum coverage appears to decrease in σ_m^2, the overall discrepancy between frequentist coverage and nominal credibility for $\beta \geq 0$ (as well as for small negative values of β) tends to diminish as σ_m increases. The coverage of $I_{ah}(X)$ at $\beta = 0$ also appears to increase as σ_m^2 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^2 correlate with more uncertainty in the bound $\beta \geq 0$, which in turn becomes reflected in the posterior distribution. Thus, even though we suspect that $\beta \geq 0$, the ad hoc interval performs reasonably well for some negative values of β as well. Moreover, we have $\lim_{\beta\to\infty} C(\theta) = 1-\alpha$. This can be shown in the same way as in Remark 2.13 below for $\sigma_m^2 \to \infty$ by the limits in Eq. (2.9). We noted similar overall behaviour of $I_{ah}(X)$ for other nominal levels such as 0.80, 0.90 and 0.99.

Remark 2.13. *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* $\sigma_m^2 \to \infty$ *in* (2.2)*, yielding* $\lim_{\sigma_m^2\to\infty} \pi(u|x) = \phi(u)$*,* $\forall u \in \mathbb{R}$*. Accordingly, one infers that* $\lim_{\sigma_m^2\to\infty} C(\theta) = 1-\alpha$*,* $\forall\theta \in \mathbb{R}^2$*, and this is illustrated in Figure 2.4 (for* $\theta_1 \geq \theta_2$ *mostly) with the flattening out around the nominal level observed as* σ_m^2 *increases. It is also easy to see this by using the limits in Eq.* (2.10)*, which yield*

$$\begin{aligned}
\lim_{\sigma_m^2\to\infty} C(\theta) &= \mathbb{E}^Z\left[\Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(\sigma_1 z_{\alpha/2}) + \frac{\sigma_1}{\sigma_2}Z\right) - \Phi\left(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(-\sigma_1 z_{\alpha/2}) + \frac{\sigma_1}{\sigma_2}Z\right)\right] \\
&= \mathbb{E}^Z\left[\Phi\left(\frac{\sigma_1}{\sigma_2}\left(Z + \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1}z_{\alpha/2}\right)\right) - \Phi\left(\frac{\sigma_1}{\sigma_2}\left(Z - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1}z_{\alpha/2}\right)\right)\right] \\
&= \Phi\left(z_{\alpha/2}\right) - \Phi\left(-z_{\alpha/2}\right) \\
&= \left(1-\frac{\alpha}{2}\right) - \frac{\alpha}{2} \\
&= 1-\alpha,
\end{aligned}$$

making use of the dominated convergence theorem and Lemma 1.10 to evaluate the expectations.

For X distributed according to model (2.1), we have $d = X_1 - X_2 \sim N(\theta_1-\theta_2, \sigma_1^2+\sigma_2^2)$. The expected length of the ad hoc interval is therefore given by

$$\mathbb{E}_\theta[L_{ah}(X)] = \mathbb{E}_\theta\left[2\sigma_1 z_{\alpha/2}\sqrt{\mathbb{V}[U|X]}\right]. \tag{2.12}$$

Figure 2.5 presents the expected length of the ad hoc interval as a function of $\beta = \theta_1 - \theta_2$ for $\sigma_1^2 = \sigma_2^2 = 1$, a 0.95 nominal level and varying σ_m.

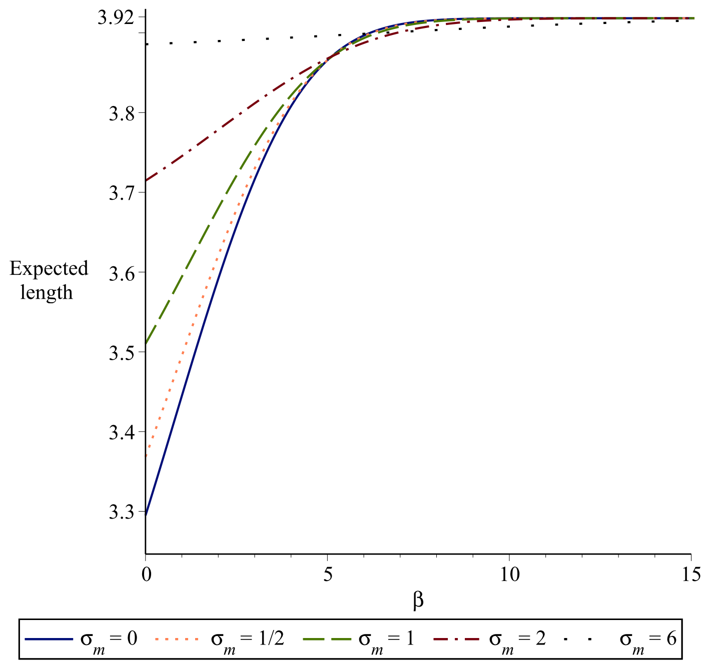


Figure 2.5 – Expected length of the ad hoc interval ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $\beta = \theta_1 - \theta_2$ for varying σ_m.

Since $\lim_{\sigma_m^2 \to \infty} \mathbb{V}[U|x] = 1$, we have $\lim_{\sigma_m^2 \to \infty} \mathbb{E}_\theta[L_{ah}(X)] = 2\sigma_1 z_{\alpha/2}$. For $\sigma_1^2 = 1$ and $\alpha = 0.05$ as in Figure 2.5, there is a flattening out of the expected length around the benchmark length $2z_{0.025} \approx 3.92$ as σ_m^2 increases. Moreover, we do not observe an ordering of the curves in Figure 2.5 since $\mathbb{V}[U|x]$ is not monotone increasing in σ_m^2 for all d. It is the case though that $\mathbb{V}[U|x]$ is an increasing function of d by Lemma 2.7. Since the family of distributions for d, which are $N(\beta, \sigma_1^2 + \sigma_2^2)$, has monotone increasing likelihood ratio in d with parameter β, it follows from Theorem 1.20 that $\mathbb{E}_\theta[\mathbb{V}[U|X]]$ is increasing in β. This in turn implies that the expected length of the ad hoc interval is also increasing in β for all $\sigma_m \geq 0$. Finally, one can also prove that $\lim_{\beta \to \infty} \mathbb{E}_\theta[L_{ah}(X)] = 2\sigma_1 z_{\alpha/2}$ which follows from the property $\mathbb{V}[U|X] \to 1$ as $d \to \infty$. This can be understood intuitively since d follows a normal distribution with mean β, so as β gets large, mass is being shifted towards larger values of d.

We also consider the exact credibility (see Definition 1.26) of the ad hoc interval, defined as $\mathbb{P}[\theta_1 \in I_{ah}(X)|x]$, or equivalently

$$\mathbb{P}[U \in [l(d), u(d)]|x] = \int_{l(d)}^{u(d)} \pi(u|x)du, \tag{2.13}$$

where $l(d) = \mathbb{E}[U|x] - z_{\alpha/2}\sqrt{\mathbb{V}[U|x]}$ and $u(d) = \mathbb{E}[U|x] + z_{\alpha/2}\sqrt{\mathbb{V}[U|x]}$, and $\pi(u|x)$ is the posterior density given by Eq. (2.2).

Figure 2.6 presents the credibility of the ad hoc interval as a function of $d = x_1 - x_2$ with $1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$ and varying values of σ_m.

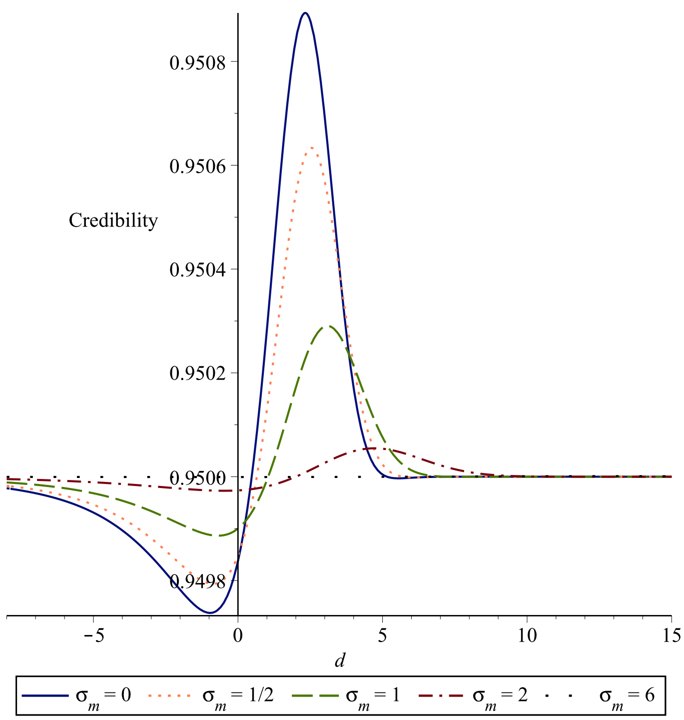


Figure 2.6 – Credibility of the ad hoc interval $(1 - \alpha = 0.95,\ \sigma_1^2 = \sigma_2^2 = 1)$ as a function of $d = x_1 - x_2$ for varying σ_m.

Examining Figure 2.6, we observe 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 $\sigma_m^2 \to \infty$. Indeed, if the posterior density was that of a standard normal, exact credibility would be attained. 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 (i.e., 0.80, 0.90 and 0.99) not shown here. A theoretical justification of this in terms of a bound on the discrepancy would be most interesting to establish.

To summarize this subsection, we construct an ad hoc Bayes credible interval for θ_1 in a standard way (centered at the posterior mean and extending on either side by equal amounts) and study its frequentist coverage probability, among other features. We provide numerical evidence of this interval's approximate $1 - \alpha$ credibility, as well as quite remarkable frequentist coverage close to the nominal credibility.

2.4.2 An ad hoc credible set 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. If the posterior distribution was normal, then a $(1 - \alpha)$ level interval would discard $\frac{\alpha}{2}$ in both tails due to the symmetric nature of the normal distribution. However as its name indicates, the skew-normal distribution is not symmetric about the mean. It would therefore be 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 the nominal credibility (see Figure 2.10).

This idea of discarding unequal amounts in the tails is referred to as a spending function in Ghashim et al. [GMS16], and previously in Marchand & Strawderman [MS13]. We consider the situation where we discard $k\alpha$ in the left tail and $(1-k)\alpha$ in the right tail, where $k \leq \frac{1}{2}$ ($k = \frac{1}{2}$ corresponds to the ad hoc interval in Definition 2.10). The adjustment in this direction with $k < 1/2$ is motivated by a relatively smaller coverage for $\beta = \theta_1 - \theta_2$ closer to 0 (see Figure 2.4).

Definition 2.14. *Let $\mathbb{E}[U|x]$ and $\mathbb{V}[U|x]$ denote respectively the posterior expectation and variance of U given by Lemma 2.6. The ad hoc Bayes credible interval for θ_1 (i.e., for $\sigma_1 U + X_1$) defined in terms of a spending function is given by*

$$I'_{\mathrm{ah}}(X) = [X_1 + l'(X_1 - X_2), X_1 + u'(X_1 - X_2)], \tag{2.14}$$

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 2.12 holds for general $u(d)$ and $l(d)$, so Eq. (2.11) holds here for all values of k. Figure 2.7 presents the frequentist coverage probability of the ad hoc interval for $\sigma_1^2 = \sigma_2^2 = 1, \sigma_m = 0$, a 0.95 nominal level and varying values of k in the spending function.

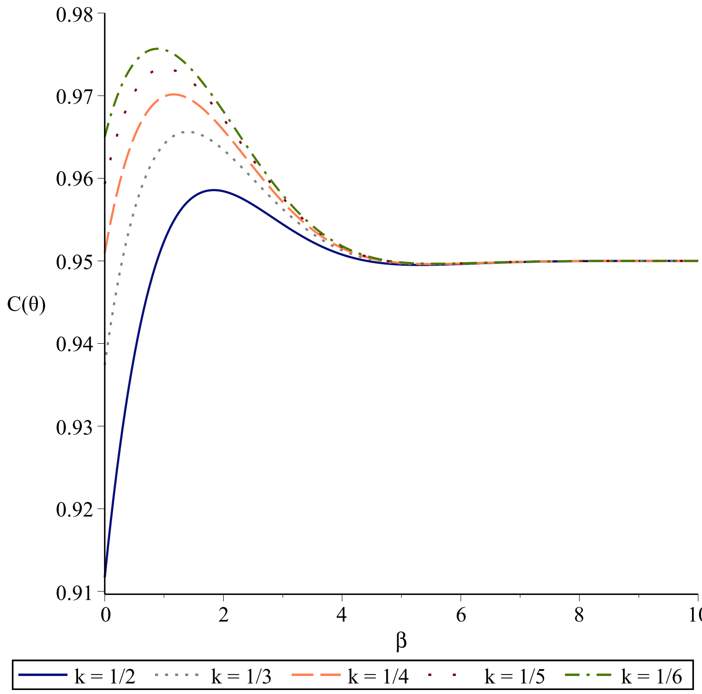


Figure 2.7 – Frequentist coverage probability of the ad hoc interval ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_m^2 = 0$) as a function of $\beta = \theta_1 - \theta_2$ for varying values of k in the spending function.

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. As shown in Figure 2.8, it also appears that $C(0) \geq 1-\alpha$ for $k \leq 1/4$ (the same behaviour was noticed for some other values of $1-\alpha$ not shown here). 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 is true, then the coverage would always remain above the nominal level if it starts above the nominal values ; that is, if $C(\theta) > 1-\alpha$ for $\theta_1 - \theta_2 = 0$, which would provide a lower bound for the coverage. Furthermore, there appears to be an ordering of the curves in terms of k, and it would be interesting to validate whether these curves intersect or not.

While the minimum coverage is not always attained at $\beta = 0$, it can still be interesting to examine the coverage at this particular value, where θ_1 coincides with θ_2. Recall that one of the objectives with the spending function was to increase the coverage for small values of $\beta \geq 0$. Figure 2.8 presents the coverage at $\beta = 0$ of the ad hoc interval as a function of $1 - \alpha$, for $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_m^2 = 0$ and varying values of k in the spending function.

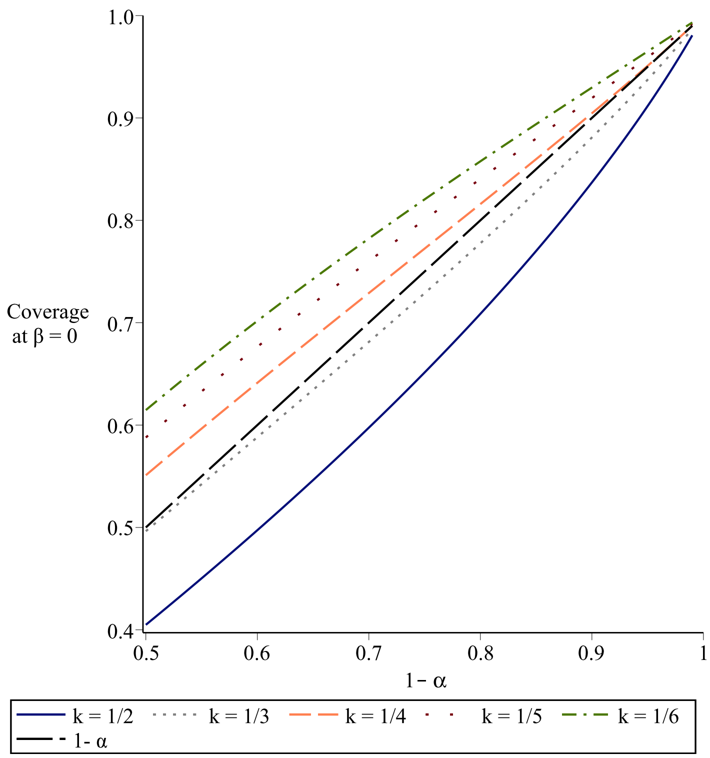


Figure 2.8 – Coverage at $\beta = 0$ of the ad hoc interval ($\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_m^2 = 0$) as a function of $1 - \alpha$ with a spending function.

One observes an ordering of the curves, with higher coverage at $\beta = 0$ for smaller values of k. It would be interesting to obtain a theoretical result relating the coverage at $\beta = 0$ to the value of k in the spending function.

Remark 2.15. *The length of the credible interval* (2.14) *is given by*

$$L'_{ah}(x) = \sigma_1\sqrt{\mathbb{V}[U|x]}(z_{(1-k)\alpha} + z_{k\alpha}).$$

Minimizing $L'_{ah}(x)$ is equivalent to minimizing $f(k) = z_{(1-k)\alpha} + z_{k\alpha}$. We observe that f is monotone decreasing in k for $k \leq 1/2$, which is minimized for $k = 1/2$. The length of the credible interval (2.14) therefore increases as k strays from $1/2$. However, for the sake of comparison, for $\alpha = 0.05$, we have $f\left(\frac{1}{2}\right) \approx 3.92$, $f\left(\frac{1}{3}\right) \approx 3.96$, $f\left(\frac{1}{4}\right) \approx 4.02$, $f\left(\frac{1}{5}\right) \approx 4.08$ and $f\left(\frac{1}{6}\right) \approx 4.13$, the latter value representing about a 5% increase in length only in comparison to $k = 1/2$.

With $d = X_1 - X_2$ distributed as before, the expected length of the credible interval given by Eq. (2.14) can be expressed as $\mathbb{E}_\theta[L'_{ah}(X)] = \mathbb{E}_\theta\left[\sigma_1\sqrt{\mathbb{V}[U|X]}(z_{(1-k)\alpha} + z_{k\alpha})\right]$. Figure 2.9 presents the expected length of the ad hoc interval as a function of $\beta = \theta_1 - \theta_2$ for $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_m^2 = 0$, a 0.95 nominal level and varying values of k.

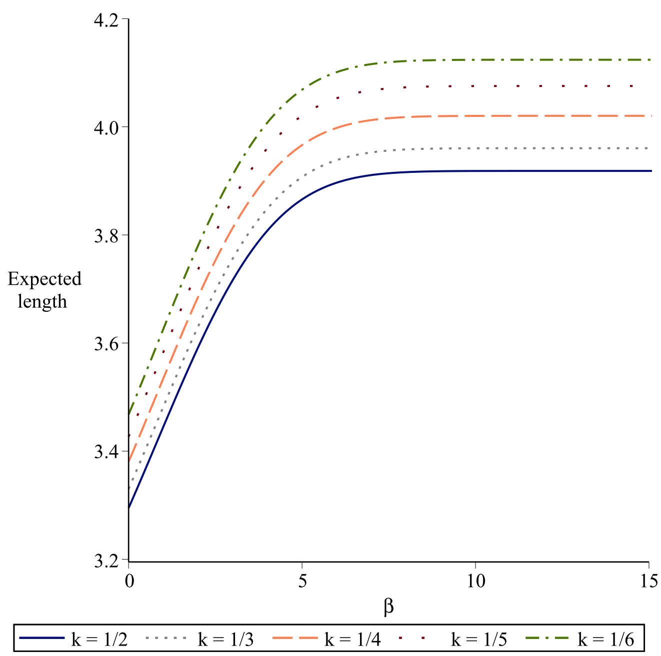


Figure 2.9 – Expected length of the ad hoc interval ($1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_m^2 = 0$) as a function of $\beta = \theta_1 - \theta_2$ for varying values of k in the spending function.

As $\beta \to \infty$, we observe that the expected length goes to the limiting values $f(k)$ given in Remark 2.15 ($\sigma_1^2 = 1$ here). We also notice an ordering of the curves as a function of k, with the expected length increasing for smaller k. Moreover, the expected length increases in β since $\mathbb{E}_\theta[\mathbb{V}[U|x]]$ is increasing in β. In comparison, the benchmark credible set $X_1 \pm \sigma_1 z_{\alpha/2}$ which does not incorporate the additional information has length $2\sigma_1 z_{\alpha/2} \approx 3.92$. The expected length of the credible interval (2.14) is therefore quite good, especially for smaller values of β.

Figure 2.10 presents the credibility of the ad hoc interval as a function of $d = x_1 - x_2$ for $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_m^2 = 0$, a 0.95 nominal level and varying values of k.

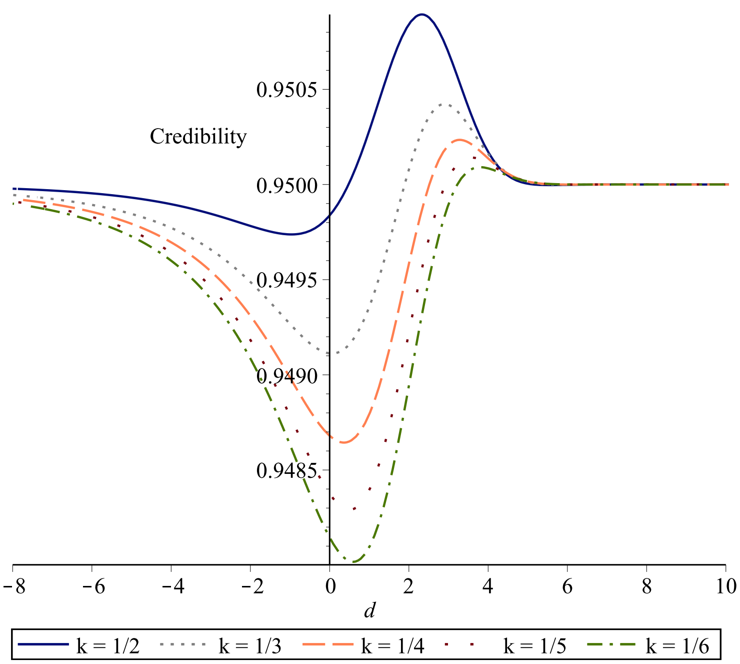


Figure 2.10 – Credibility of the ad hoc interval ($1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_m^2 = 0$) as a function of $d = x_1 - x_2$ for varying values of k in the spending function.

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. While there appears to be an ordering of the curves around $d = 0$, it does seem that they intersect around $d \approx 4.5$ for this particular nominal level. Moreover, such an ordering appears to depend on α (for $\alpha = 0.1$ for example, there is only an ordering for more extreme values of d). For the sake of comparison, Table 2.1 gives an approximate maximum discrepancy of the credibility for $k = 1/4$ and varying values of $1 - \alpha$.

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

TABLE 2.1 – Approximate maximum credibility discrepancy for the ad hoc interval with $k = 1/4$ in the spending function, $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_m^2 = 0$.

While the graphs of coverage appear to maintain a similar appearance for different nominal levels, the same cannot be said for the credibility. For instance, while higher than nominal credibility is achieved for $d > 0$ in Figure 2.10, the behaviour of the curves appears to flip for a 0.80 nominal level (see Figure 2.11).

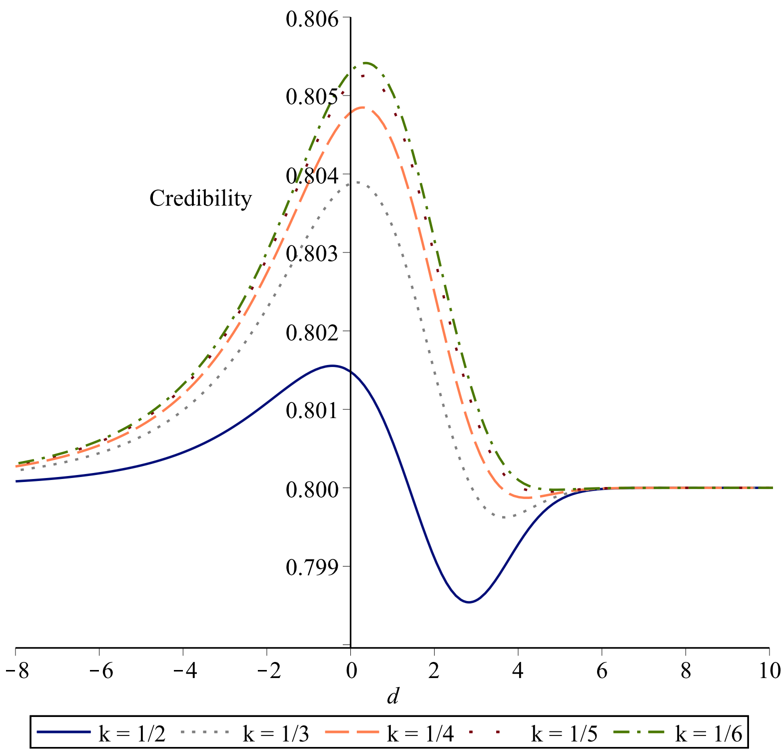


Figure 2.11 – Credibility of the ad hoc interval ($1 - \alpha = 0.80$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_m^2 = 0$) as a function of $d = x_1 - x_2$ for varying values of k in the spending function.

2.5 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 $\sigma_m^2 > 0$ 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 $\sigma_m^2 = 0$. 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^2 ranging from the "no-useful additional information case" ($\sigma_m^2 \to \infty$) to the certain constraint $\theta_1 \geq \theta_2$ ($\sigma_m^2 = 0$). While the focus is for the parameter space $\theta_1 \geq \theta_2$, the Bayes estimator nonetheless performs for $\theta_1 < \theta_2$ due to the shifting of the posterior distribution when there is uncertainty in the constraint. 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 choice of the value k in the spending function leads to a compromise between the frequentist coverage and the length of the ad hoc interval, which renders the question of an optimal value of k legitimate and not so straightforward to answer.

The presentation of these findings leaves open several interesting questions about analytically derived lower bounds on coverage probabilities which bring into play the model variances, the choice of σ_m^2, as well as the spending function setting. It would be particularly interesting to proceed with an analysis for an unknown variances extension of model (2.1). 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.

CHAPTER 3

Other models

The results in Chapter 2 raise the idea of considering similar problems with other models or parametric constraints. This chapter presents two such modifications. Section 3.1 considers the same model for X_1 and X_2 as in (2.1), but with the differences of means being doubly-bounded (i.e., bounded above and below). Section 3.2 deals with the same type of constraint (i.e., $\mu_1 \geq \mu_2$) as Chapter 2, but the variances of the normally distributed X_1 and X_2 are now unknown (but equal). Note that there is no uncertainty in the parametric bound in this chapter. The focus here is on the ad hoc procedure presented in Definition 2.10 and on properties of resulting credible intervals. This naturally entails a posterior analysis of the given model and prior. While we do not concern ourselves with point estimation for the models in this chapter, it is important to mention that much work has been done on such problems. We refer the reader to Marchand & Strawderman [MS04] and van Eeden [vE06] for elements of review.

3.1 Doubly-bounded constraint

Consider the following model for $X = (X_1, X_2)^\top$ and prior :

$$\begin{aligned}
&X_1 \sim N(\theta_1, \sigma_1^2),\ X_2 \sim N(\theta_2, \sigma_2^2);\\
&\pi(\theta_1, \theta_2) = \mathbb{1}_{[-m,m]}(\theta_1 - \theta_2),
\end{aligned} \tag{3.1}$$

where X_1 and X_2 are independently distributed and $\sigma_1, \sigma_2, m > 0$ are known. The parameter space is therefore restricted to the region $|\theta_1 - \theta_2| \leq m$, and a flat prior is placed on this restricted parameter space. This parameter space models a situation where it is believed that the two means θ_1 and θ_2 are not too different from one another, and the preciseness of how close they are is determined by the value m.

Lemma 3.1. *Under the context of* (3.1), *the joint posterior density of* (θ_1, θ_2) *is given by*

$$\pi(\theta_1, \theta_2 | x) = \frac{\frac{1}{\sigma_1 \sigma_2} \phi\left(\frac{\theta_1 - x_1}{\sigma_1}\right) \phi\left(\frac{\theta_2 - x_2}{\sigma_2}\right)}{\Phi\left(\frac{m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{-m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)} \mathbb{1}_{[-m,m]}(\theta_1 - \theta_2).$$

Proof. This follows from writing the joint posterior density of $(\theta_1, \theta_2)^\top$ as

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

where

$$f(x|\theta)\pi(\theta) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2\sigma_1^2}(\theta_1 - x_1)^2} e^{-\frac{1}{2\sigma_2^2}(\theta_2 - x_2)^2} \mathbb{1}_{[-m,m]}(\theta_1 - \theta_2),$$

and observing that the normalizing constant can be expressed as follows

$$\mathbb{P}[-m \leq W \leq m] = \Phi\left(\frac{m - (x_1 - x_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) - \Phi\left(\frac{-m - (x_1 - x_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right),$$

where $Y_1 \sim N(x_1, \sigma_1^2)$ and $Y_2 \sim N(x_2, \sigma_2^2)$ with Y_1 and Y_2 independent, and setting $W = Y_1 - Y_2 \sim N\big(x_1 - x_2, \sigma_1^2 + \sigma_2^2\big)$. $\square$

Lemma 3.2. *Under the model and prior given by* (3.1), *the marginal posterior density of* θ_1 *is given by*

$$\pi(\theta_1|x) = \frac{\frac{1}{\sigma_1}\phi\left(\frac{\theta_1-x_1}{\sigma_1}\right)\left[\Phi\left(\frac{\theta_1+m-x_2}{\sigma_2}\right)-\Phi\left(\frac{\theta_1-m-x_2}{\sigma_2}\right)\right]}{\Phi\left(\frac{m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)-\Phi\left(\frac{-m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}, \quad \theta_1 \in \mathbb{R}.$$

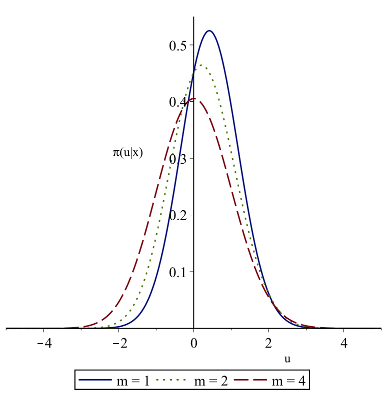
Proof. This follows from Lemma 3.1 by integrating out θ_2 over the region $\theta_1 - m \leq \theta_2 \leq \theta_1 + m$. □

Remark 3.3. *As in Chapter 2, it is useful to consider* $U = \frac{\theta_1 - x_1}{\sigma_1}$ *and* $d = x_1 - x_2$. *Lemma 3.2 yields the following representation of the posterior density*

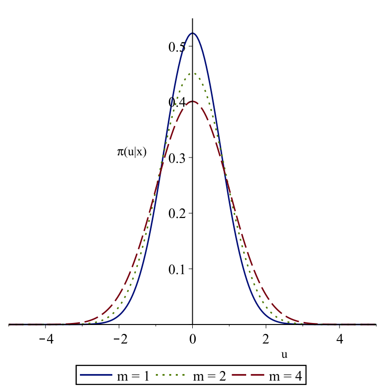
$$\pi(u|x) = \frac{\phi(u)\left[\Phi\left(\frac{\sigma_1 u+d+m}{\sigma_2}\right)-\Phi\left(\frac{\sigma_1 u+d-m}{\sigma_2}\right)\right]}{\Phi\left(\frac{m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)-\Phi\left(\frac{-m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}, \; u \in \mathbb{R}, \tag{3.2}$$

which one recognizes as the difference of two extended skew-normal densities (such densities appear in van Eeden & Zidek [vEZ04] for instance).

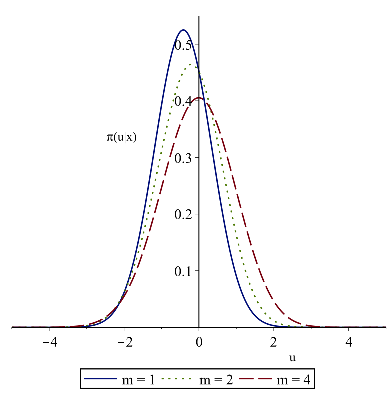
Figure 3.1 presents the posterior density given by (3.2) for $d = -1, 0, 1$, $\sigma_1^2 = \sigma_2^2 = 1$ and varying values of the constraint m.



(a) $d = -1$



(b) $d = 0$



(c) $d = 1$

Figure 3.1 – Posterior density of U for $d = -1, 0, 1$, $\sigma_1^2 = \sigma_2^2 = 1$ and varying m.

Remark 3.4. *As m increases, the value of the additional information on the bound of the difference $\theta_1 - \theta_2$ decreases, and heuristically, as $m \to \infty$, we would expect to get the same posterior density for θ_1 as if we only had information on X_1. This is indeed the case since from* (3.2)*, we have $\lim_{m\to\infty} \pi(u|x) = \phi(u)$, which translates to a $N(x_1, \sigma_1^2)$ posterior for θ_1. This behaviour is apparent in Figure 3.1.*

One can readily obtain the posterior moment generating function, expectation and variance of U.

Lemma 3.5. *Under the context of* (3.1) *and Remark 3.3, the posterior moment generating function, expectation and variance of U are given respectively by*

$$M_{U|x}(t) = \frac{e^{\frac{t^2}{2}}\left[\Phi\left(\frac{\sigma_1 t+d+m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{\sigma_1 t+d-m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\right]}{\Phi\left(\frac{m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{-m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}, \tag{3.3}$$

$$\mathbb{E}[U|x] = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\;\frac{\phi\left(\frac{d+m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \phi\left(\frac{d-m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}{\Phi\left(\frac{m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{-m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}, \text{ and} \tag{3.4}$$

$$\begin{aligned}\mathrm{Var}[U|x] = {}& \frac{1}{\Phi\left(\frac{m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{-m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}\left[\Phi\left(\frac{d+m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{d-m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\right. \\ & \left. + \frac{\sigma_1^2}{(\sigma_1^2+\sigma_2^2)^{3/2}}\left\{(d-m)\phi\left(\frac{d-m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - (d+m)\phi\left(\frac{d+m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\right\}\right] \\ & - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\left[\frac{\phi\left(\frac{d+m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \phi\left(\frac{d-m}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}{\Phi\left(\frac{m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - \Phi\left(\frac{-m-d}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)}\right]^2.\end{aligned} \tag{3.5}$$

Proof. The posterior moment generating function can be obtained in a similar way as in Lemma 2.6, making use of Lemma 1.10 to evaluate the integral. The posterior expectation and variance follow immediately. □

The following result relates to Lemma 3.5.

Lemma 3.6. *Let f_d be a family of densities on $\mathbb{R}$ such that $f_d(u) = f_{-d}(-u)$ for all $u, d \in \mathbb{R}$. Then, provided expectations exist, we have*

$$\mathbb{E}_d\left[U^k\right] = (-1)^k \mathbb{E}_{-d}\left[U^k\right], \text{ for } k \in \mathbb{N}_+.$$

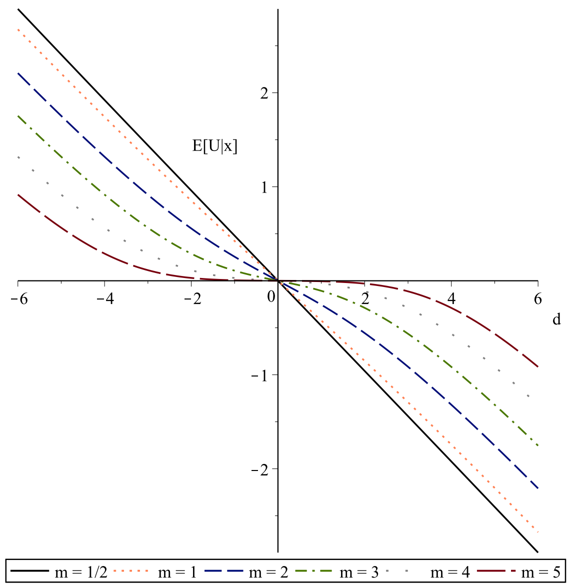
Consequently, $\mathbb{E}_d[U]$ is an odd function of d and $\mathbb{V}_d[U]$ is an even function of d.

Proof. We have

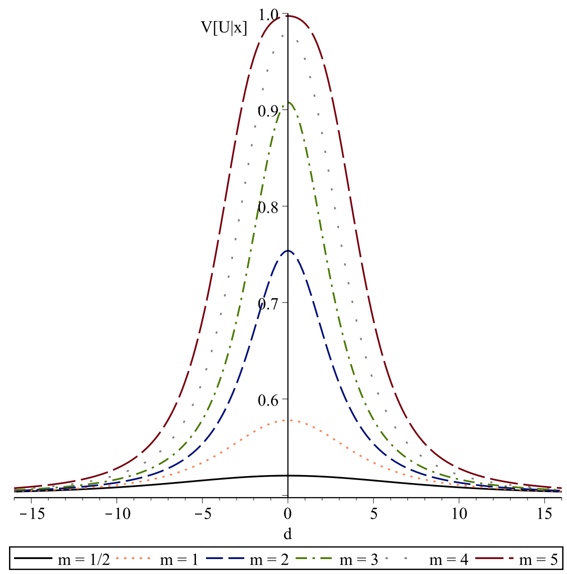
$$\mathbb{E}_d\left[U^k\right] = \int_{\mathbb{R}} u^k f_d(u) du = \int_{\mathbb{R}} u^k f_{-d}(-u) du = \int_{\mathbb{R}} (-1)^k u^k f_{-d}(u) du = (-1)^k \mathbb{E}_{-d}\left[U^k\right].$$

□

One can easily verify that (3.2) belongs to the family of densities defined in Lemma 3.6, and thus $\mathbb{E}[U|x]$ and $\mathbb{V}[U|x]$ given in Lemma 3.5 are respectively odd and even functions of $d = x_1 - x_2$.



(a) Posterior expectation



(b) Posterior variance

Figure 3.2 – Posterior expectation and variance of U as a function of $d = x_1 - x_2$ for $\sigma_1^2 = \sigma_2^2 = 1$ and varying m.

Figure 3.2 presents the posterior expectation and variance of U as a function of $d = x_1 - x_2$ for $\sigma_1^2 = \sigma_2^2 = 1$ and varying values of the constraint m. Several observations can be made from Figure 3.2, perhaps most noticeably the presence of symmetry due to the even and odd nature of these functions. Further properties are given in Remark 3.8 which make use of the following lemma.

Lemma 3.7 (Marchand & Sadeghkhani [MS18])**.** *Let* $X_1 \sim N(\theta_1, \sigma_1^2)$ *and* $X_2 \sim N(\theta_2, \sigma_2^2)$ *be independently distributed with prior* $\pi(\theta) = \mathbb{1}_{[-m,m]}(\theta_1 - \theta_2)$. *Then, conditional on* $x = (x_1, x_2)$, $\omega_1 = \theta_1 - \theta_2$ *and* $\omega_2 = r\theta_1 + \theta_2$, *where* $r = \frac{\sigma_2^2}{\sigma_1^2}$, *are independently distributed with*

$$\omega_1 \sim TN(x_1 - x_2, \sigma_1^2 + \sigma_2^2, (-m, m)) \quad \text{and} \quad \omega_2 \sim N(rx_1 + x_2, 2\sigma_2^2).$$

Remark 3.8. *Model* (3.1) *corresponds to the context of Lemma 3.7, which allows one to* *express the posterior expectation of θ_1 as follows, which is helpful for heuristic arguments :*

$$\mathbb{E}[\theta_1|x] = \frac{1}{1+r}\mathbb{E}[\omega_1 + \omega_2|x] = \frac{1}{1+r}\mathbb{E}[\omega_1|x] + \frac{rx_1 + x_2}{1+r},$$

which in turn implies that

$$\mathbb{E}\left[\frac{\theta_1 - x_1}{\sigma_1} \,\middle|\, x\right] = \frac{1}{\sigma_1(1+r)}\mathbb{E}[\omega_1|x] + \frac{x_2 - x_1}{\sigma_1(1+r)},$$

and thus

$$\frac{\mathbb{E}[U|x]}{d} = \frac{\mathbb{E}[\omega_1|x]}{\sigma_1 d(1+r)} - \frac{1}{\sigma_1(1+r)}.$$

$\mathbb{E}[\omega_1|x]$ *can be obtained from* (1.1)*, and since* $\mathbb{E}[\omega_1|x] \in [-m, m]$*, it follows that*

$$\lim_{d\to\pm\infty} \frac{\mathbb{E}[U|x]}{d} = -\frac{1}{\sigma_1(1+r)}.$$

From Figure 3.2, one also notices that for all m, $\mathbb{E}[U|x]$ appears to be a monotone decreasing function of d. Moreover, the magnitude of $\mathbb{E}[U|x]$ appears to increase for all d as m decreases.

Heuristically, if $x_1 \gg x_2$ or $x_2 \gg x_1$, then $|d|$ is large and under the assumption of our model that $|\theta_1 - \theta_2| \leq m$, as long as m is not too large, the data appears to contradict the model. We would therefore center all posterior belief on the boundary $|\theta_1 - \theta_2| = m$. If $x_1 \gg x_2$, we would believe that $\theta_1 - \theta_2 = m$ and conversely if $x_2 \gg x_1$, we would center our belief on the boundary $\theta_1 - \theta_2 = -m$. Moreover, as $d \to \infty$, ω_1 converges to a point mass at m, and thus for large d, we have

$$\mathbb{E}[U|x] \approx \frac{m}{\sigma_1(1+r)} - \frac{d}{\sigma_1(1+r)}.$$

Similarly, as $d \to -\infty$, ω_1 converges to $-m$, and hence for small d, we have

$$\mathbb{E}[U|x] \approx -\frac{m}{\sigma_1(1+r)} - \frac{d}{\sigma_1(1+r)}.$$

Such a behaviour can be noticed in Figure 3.2. For instance, for $\sigma_1^2 = \sigma_2^2 = 1$ and large d, we have $\mathbb{E}[U|x] \approx \frac{m}{2} - \frac{d}{2}$, which matches the slopes of the curves appearing to approach $-\frac{1}{2}$.

The same type of reasoning can be made for the posterior variance. By the independence of ω_1 and ω_2 in Lemma 3.7, we have

$$\mathbb{V}[U|x] = \frac{1}{\sigma_1^2}\mathbb{V}[\theta_1|x] = \frac{1}{\sigma_1^2}\mathbb{V}\left[\frac{\omega_1+\omega_2}{1+r}\,\middle|\, x\right] = \frac{1}{\sigma_1^2(1+r)^2}\left(\mathbb{V}[\omega_1|x] + 2\sigma_2^2\right),$$

where $\mathbb{V}[\omega_1|x]$ can be obtained from (1.2).

Further inference on the behaviour of the posterior variance can be derived from the following lemma.

Lemma 3.9 (Chen [Che13]; also Chen et al. [CvEZ10])**.** *Let X be a continuous random variable with log-concave cdf $F(x)$ on the interval C. Then $\mathbb{V}[X|X \in A] \leq \mathbb{V}[X|X \in B]$ for any intervals $A \subset B \subset C$.*

It is well known that normal distributions have log-concave cdfs, and thus from Lemma 3.9, we have that $\mathbb{V}[\omega_1|x]$ (and consequently $\mathbb{V}[U|x]$) is increasing in m. This is apparent in Figure 3.2. Moreover, due to the posterior distribution of ω_1 converging to a point mass as $d \to \pm\infty$, we also obtain that $\lim_{d\to\pm\infty} \mathbb{V}[U|x] = \frac{2\sigma_2^2}{\sigma_1^2(1+r)^2} = \frac{2r}{(1+r)^2}$. This coincides with the limiting behaviour observed in Figure 3.2 where the variance tends towards $\frac{1}{2}$ $(r = 1)$. One also observes that $\mathbb{V}[\omega_1|x]$ is decreasing in $|d|$.

With expressions for $\mathbb{E}[U|x]$ and $\mathbb{V}[U|x]$, we can construct the ad hoc credible set presented in Chapter 2 and given by Definition 2.10. Figure 3.3 displays the bounds of the ad hoc interval for U as a function of $d = x_1 - x_2$ for $1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$ and varying values of m. With the symmetry of the posterior expectation and variance of U, one is not surprised by the symmetry in Figure 3.3. There appears to be critical values of d where the upper, and lower, bounds intersect for all m. Graphically, those values appear to be $d_1^* \approx -1.15$ and $d_2^* \approx 1.15$ respectively for $1-\alpha = 0.95$.

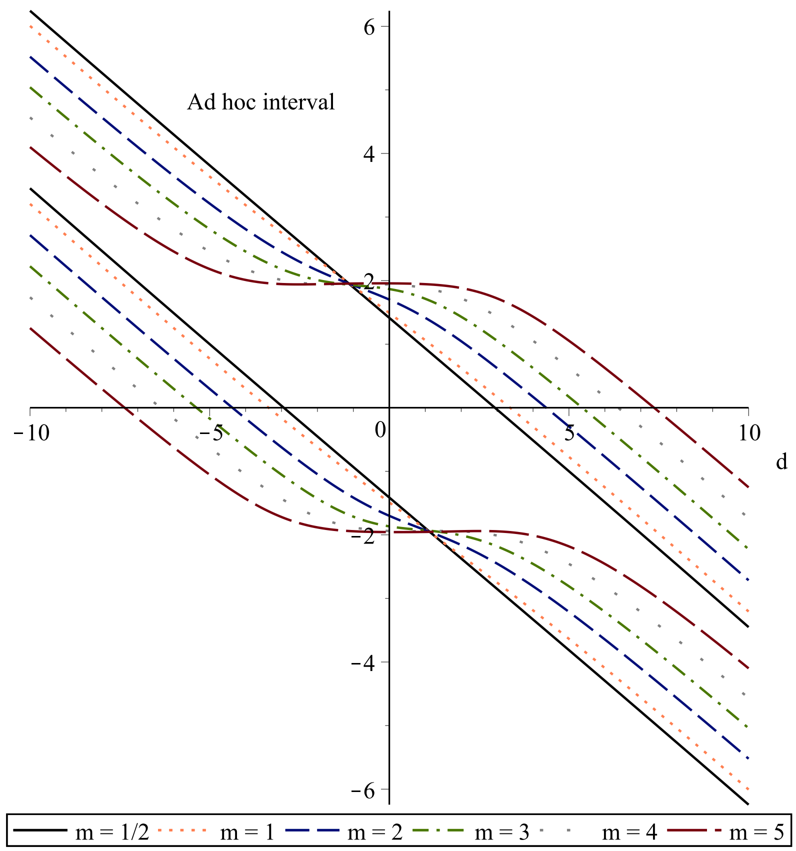


Figure 3.3 – Ad hoc interval for U ($1-\alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $d = x_1 - x_2$ for varying values of m.

Figure 3.4 presents the credibility of the ad hoc interval (again given by (2.13) but with the posterior density, expectation and variance being those corresponding to the doubly-bounded constraint) as a function of $d = x_1 - x_2$ for $1-\alpha = 0.95$ and $\sigma_1^2 = \sigma_2^2 = 1$.

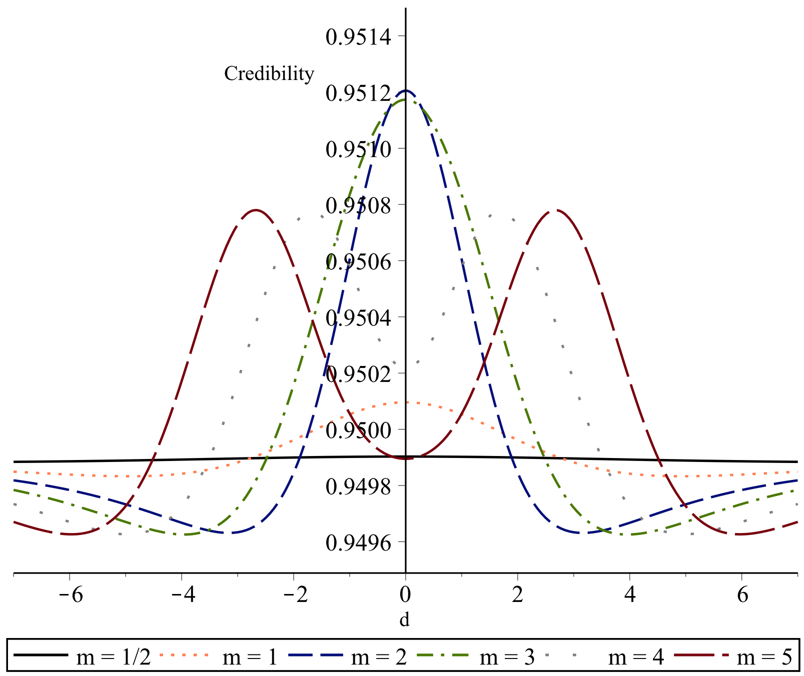


Figure 3.4 – Credibility of the ad hoc interval ($1 - \alpha = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$) as a function of $d = x_1 - x_2$ for varying values of m.

From Figure 3.4, we observe that the exact credibility fluctuates around the nominal 0.95 level, attaining both lower and higher than nominal values. The credibility is nonetheless quite close to the nominal level. An interesting feature is that $d = 0$ appears to correspond to either a local mininum or maximum, depending on the value of m. The credibility is also symmetric about $d = 0$, which is again not surprising due to the symmetric nature of the constraint.

While Eq. (2.11) for the coverage probability given in Theorem 2.12 holds for the model (3.1), we were not successful in obtaining graphs of such coverage. It would be interesting to study the frequentist properties of the ad hoc interval under this doubly-bounded constraint, in a manner similar to what was done in Chapter 2 for the lower-bounded constraint. It would also be of interest to extend this section to include uncertainty on the value m of the constraint.

Finally, we draw attention to a few point estimation results which relate to this problem and can be found in the literature. For instance, the rotation technique described in Chapter 1 (and the corresponding Lemma 1.38) apply to this doubly-bounded contraint model, which implies that finding estimators that dominate a benchmark estimator under squared error loss is reduced to a univariate problem. Such a univariate result can be found in van Eeden & Zidek [vEZ01] among others, where a generalized Bayes estimator of θ_1 with respect to a flat prior on the restricted parameter space is shown to dominate the unrestricted MLE X_1. Marchand & Perron [MP01] also provide estimators that dominate the MLE of θ_1 under squared error loss and in this regard permit extensions to the multivariate case with $X_1 \sim N_d(\theta_1, \sigma_1^2 I_d)$, $X_2 \sim N_d(\theta_2, \sigma_2^2 I_d)$ and $\|\theta_1 - \theta_2\| \leq m$.

3.2 Unknown but equal variances

This section considers the following model for $X = (X_1, X_2)^\top$ and prior :

$$\begin{aligned} &X_{11}, ..., X_{1n_1} \text{ i.i.d. } N(\mu_1, \sigma^2), \;\; X_{21}, ..., X_{2n_2} \text{ i.i.d. } N(\mu_2, \sigma^2); \\ &\pi(\mu_1, \mu_2, \sigma) = \frac{1}{\sigma} \mathbb{1}_{\{\mu_1 \geq \mu_2\}}(\mu_1, \mu_2) \mathbb{1}_{(0,\infty)}(\sigma), \end{aligned} \tag{3.6}$$

where X_1 and X_2 are independently distributed and $\sigma > 0$ is now unknown. This corresponds to a situation where we have two independent samples from normal distributions with means μ_1 and μ_2, and it is known that there is an ordering of the means (i.e., $\mu_1 \geq \mu_2$). Note that we can reduce this problem to the sufficient statistics of sample means and variances, namely

$$\bar{x_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i},\ \bar{x_2} = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i},\ s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x_1})^2,\ s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_{2i}-\bar{x_2})^2.$$

Definition 3.10. *T has a Student t-distribution with shape parameter $\alpha > 0$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ when it has density on $\mathbb{R}$ given by*

$$f_{\alpha,\mu,\sigma}(t) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sigma(\alpha\pi)^{1/2}\Gamma\left(\frac{\alpha}{2}\right)}\left(1+\frac{(t-\mu)^2}{\alpha\sigma^2}\right)^{-\left(\frac{\alpha+1}{2}\right)}.$$

We denote such a distribution $T \sim t(\alpha,\mu,\sigma^2)$. We also denote the cdf of such a distribution as $F_{\alpha,\mu,\sigma}(\cdot)$.

Theorem 3.11. *Under the context of* (3.6)*, the marginal posterior distribution of μ_1 is given by*

$$\mu_1|x \propto f_{\alpha_1,\bar{x}_1,\sigma_1}(\mu_1)\cdot F_{\alpha_2,\bar{x}_2,\sigma_2}(\mu_1), \tag{3.7}$$

where $\alpha_1 = n_1+n_2-3$, $\sigma_1^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1(n_1+n_2-3)}$, $\alpha_2 = n_1+n_2-2$ and $\sigma_2^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+n_1(\bar{x_1}-\mu_1)^2}{n_2(n_1+n_2-2)}$.

Proof. The joint posterior density of (μ_1,μ_2,σ^2) is given by

$$\pi(\mu_1,\mu_2,\sigma^2|x) = \frac{f(x_1|\mu_1,\sigma^2)f(x_2|\mu_2,\sigma^2)\pi(\mu_1,\mu_2,\sigma^2)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\mu_1}\int_0^{\infty} f(x_1|\mu_1,\sigma^2)f(x_2|\mu_2,\sigma^2)\pi(\mu_1,\mu_2,\sigma^2)d\sigma^2 d\mu_2 d\mu_1},$$

where

$$f(x_j|\mu_j,\sigma^2) = \prod_{i=1}^{n_j}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_{ji}-\mu_j)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n_j/2}}e^{-\frac{(n_j-1)s_j^2}{2\sigma^2}}e^{-\frac{n_j}{2\sigma^2}(\bar{x_j}-\mu_j)^2}.$$

With the transformation $(\mu_1,\mu_2,\sigma^2) \xrightarrow{T} \left(\mu_1,\mu_2,\phi=\frac{1}{\sigma^2}\right)$, we can write

$$\pi(\mu_1,\mu_2,\phi|x) \propto \frac{1}{(2\pi)^{\frac{n_1+n_2}{2}}}\cdot\phi^{\frac{n_1+n_2-1}{2}-1}\cdot e^{-\frac{\phi}{2}\left[(n_1-1)s_1^2+(n_2-1)s_2^2+n_1(\bar{x_1}-\mu_1)^2+n_2(\bar{x_2}-\mu_2)^2\right]}$$
$$\cdot\,\mathbb{1}_{\{\mu_1\geq\mu_2\}}(\mu_1,\mu_2)\mathbb{1}_{(0,\infty)}(\phi).$$

The joint posterior density of (μ_1, μ_2) is then obtained as follows by integrating out ϕ :

$$\pi(\mu_1, \mu_2|x) \propto \int_0^{\infty} \phi^{\tilde{\alpha}-1} \cdot e^{-\tilde{\beta}\phi} \mathbb{1}_{\{\mu_1 \geq \mu_2\}}(\mu_1, \mu_2) d\phi,$$

where

$$\tilde{\alpha} = \frac{n_1 + n_2 - 1}{2}, \text{ and}$$

$$\tilde{\beta} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + n_1(\bar{x_1} - \mu_1)^2 + n_2(\bar{x_2} - \mu_2)^2}{2}.$$

Making use of the property $\int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^{\alpha}}$, we obtain the following expression for the marginal posterior density of μ_1 :

$$\begin{aligned}
\pi(\mu_1|x) &\propto \int_{-\infty}^{\mu_1} \tilde{\beta}^{-\tilde{\alpha}} d\mu_2 \\
&\propto \left(\frac{D}{2}\right)^{-\left(\frac{n_1+n_2-1}{2}\right)} \int_{-\infty}^{\mu_1} \left(1 + \frac{(\mu_2 - \bar{x_2})^2}{(n_1 + n_2 - 2)\frac{D}{n_2(n_1+n_2-2)}}\right)^{-\left(\frac{(n_1+n_2-2)+1}{2}\right)} d\mu_2,
\end{aligned} \tag{3.8}$$

where $D = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + n_1(\bar{x}_1 - \mu_1)^2$. We recognize in (3.8) the cdf of a Student t-distribution with parameters $\alpha_2 = n_1 + n_2 - 2$, $\mu_2 = \bar{x_2}$ and $\sigma_2^2 = \frac{D}{n_2(n_1+n_2-2)}$. Setting $E = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2$, (3.8) becomes

$$\begin{aligned}
\pi(\mu_1|x) &\propto \left(E + n_1(\bar{x_1} - \mu_1)^2\right)^{-\left(\frac{n_1+n_2-2}{2}\right)} \cdot \mathbb{P}\left[\mathrm{t}(\alpha_2, \bar{x_2}, \sigma_2^2) \leq \mu_1\right] \\
&\propto \left(1 + \frac{(\mu_1 - \bar{x_1})^2}{E/n_1}\right)^{-\left(\frac{(n_1+n_2-3)+1}{2}\right)} \cdot \mathbb{P}\left[\mathrm{t}(\alpha_2, \bar{x_2}, \sigma_2^2) \leq \mu_1\right].
\end{aligned} \tag{3.9}$$

Finally, we recognize the first part of (3.9) as the density of a Student t-distribution with parameters $\alpha_1 = n_1 + n_2 - 3$, $\mu_1 = \bar{x_1}$ and $\sigma_1^2 = \frac{E}{n_1(n_1+n_2-3)}$. □

The density given by (3.7) can be called a skew-Student distribution, similarly to the skew-normal distribution in Definition 1.8, but where all distributions involved are Student t-distributions instead of normal. This is mentioned by Azzalini [Azz14] (also see Sadeghkani & Ahmed [SA20]).

One difficulty arising with the skew-Student distribution is the calculation of the normalization constant. This requires computing an integral of the form

$$\int_{-\infty}^{\infty} f_{\alpha_1,\bar{x}_1,\sigma_1}(\mu_1) \cdot F_{\alpha_2,\bar{x}_2,\sigma_2}(\mu_1) d\mu_1, \tag{3.10}$$

where the parameters of the Student t-distributions are as previously defined. Defining random variables $U \sim t(\alpha_1, \bar{x_1}, \sigma_1^2)$ and $V \sim t(\alpha_2, \bar{x_2}, \sigma_2^2)$, (3.10) is equivalent to $\mathbb{P}[U \leq V] = \mathbb{P}[U - V \leq 0]$, which is the cdf at zero of $U - V$. This brings into play the distribution of $U - V$. In general, the distribution of sums or differences of random variables can be obtained through convolutions, but this will require numerical evaluation. As opposed to Chapter 2 and Section 3.1, we do not obtain explicit expressions for the posterior expectation and variance of μ_1, quantities which are used in the construction of the ad hoc credible set. We therefore have only obtained numerical results, an example of which is given in Figure 3.5.

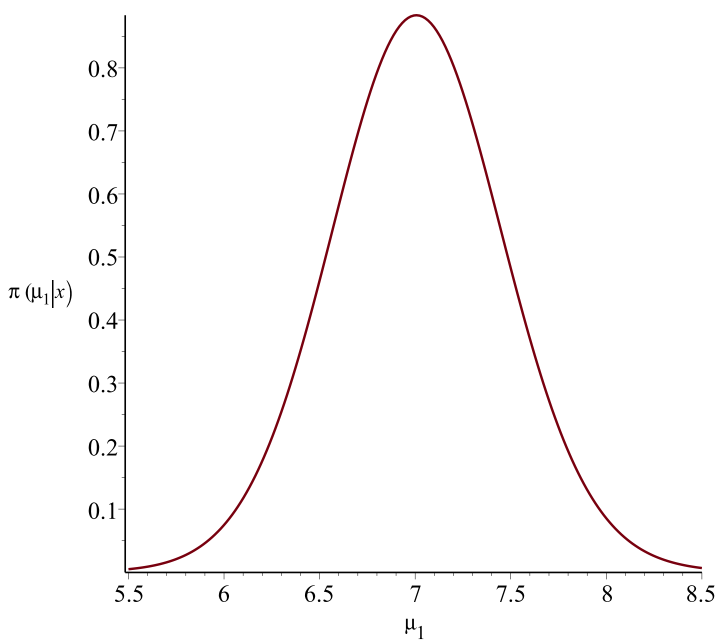


Figure 3.5 – Posterior density of μ_1 for $n_1 = n_2 = 20$, $\overline{x_1} = 7$, $\overline{x_2} = 5$ and $s_1 = s_2 = 2$.

For the choice of parameters in Figure 3.5, we have $\mathbb{E}[\mu_1|x] \approx 7.012$ and $\mathbb{V}[\mu_1|x] \approx 0.212$, which yield an ad hoc interval (of the same form as Definition 2.10) for μ_1 of approximately $[6.110, 7.913]$ for $1-\alpha = 0.95$.

While the analysis in this section has been limited, it has nonetheless illustrated another situation where the ad hoc procedure can be used, although relying on numerical evaluations of the posterior expectation and variance. It also demonstrates that even though this procedure is easy to implement in theory (and was seemingly easy to compute in Chapter 2 and Section 3.1), it can have its challenges depending on the model.

CONCLUSION

This work concerns the estimation of one of two normal means when there exists a constraint bounding and relating the difference of the two means. Chapter 1 presents some preliminary definitions and theory which relate to results in Chapters 2 and 3. The main part of this work is found in Chapter 2 where interest lies in estimating the suspected larger of two normal means. In this chapter, uncertainty is introduced in the parametric constraint in the form of a hierarchical prior. Point estimation results are obtained, yielding a class of minimax Bayes estimators which dominate the unrestricted MLE. The focus then shifts to the construction of an ad hoc credible interval for θ_1 and frequentist properties of this interval are studied, notably the closeness of its frequentist coverage probability to the nominal credibility. In Section 2.4.2, a spending function is incorporated and its impact is studied, in particular regarding improved frequentist coverage probability. Finally, Chapter 3 presents two modifications to the model in Chapter 2. Section 3.1 modifies the parametric constraint and doubly bounds the difference of the normal means. Section 3.2 considers the same type of constraint as in Chapter 2, but now for a model with unknown (but equal) variances. In both sections, we proceed with a posterior analysis and discuss the ad hoc credible interval.

There are numerous avenues that would be of interest to pursue related to this work. First of all, the uncertainty in the parametric constraint, as well as the idea of the spending function, can be included in other models and for different types of constraints. This work considers an ad hoc credible interval, but one could also consider other types of credible sets, such as an HPD or equal-tailed interval for instance, and conduct comparisons between them. Furthermore, it would be of interest to obtain further analytic results relating to some of the topics presented, notably a lower bound for the frequentist coverage probability of the ad hoc interval. Finally, the point estimation results relating to risk under squared error loss in Chapter 2 can seemingly relate to risks of predictive density estimation under Kullback-Leibler loss.

Bibliography

[AB02] B.C. Arnold and R.J. Beaver. Skewed multivariate models related to hidden truncation and/or selective reporting. *Sociedad de Estadística e Investigación Operativa Test*, 11(1) :7–54, 2002.

[Azz85] A. Azzalini. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, 12(2) :171–178, 1985.

[Azz14] A. Azzalini. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2014.

[Bar08] A. Baricz. Mill's ratio : Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 340 :1362–1370, 2008.

[BC68a] S. Blumenthal and A. Cohen. Estimation of the larger translation parameter. *The Annals of Mathematical Statistics*, 39(2) :502–516, 1968.

[BC68b] S. Blumenthal and A. Cohen. Estimation of two ordered translation parameters. *The Annals of Mathematical Statistics*, 39(2) :517–530, 1968.

[Ber85] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer, 2 edition, 1985.

[CB02] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Advanced Series. Duxbury Thomson Learning, 2 edition, 2002.

[Che13] J. Chen. A Partial Order on Uncertainty and Information. *Journal of Theoretical Probability*, 26 :349–359, 2013.

[CS70] A. Cohen and H.B. Sackrowitz. Estimation of the last mean of a monotone sequence. *The Annals of Mathematical Statistics*, 41(6) :2021–2034, 1970.

[CvEZ10] J. Chen, C. van Eeden, and J. Zidek. Uncertainty and the conditional variance. *Statistics and Probability Letters*, 80 :1764–1770, 2010.

[Den10] M. Denis. *Méthodes de modélisation bayésienne et applications en recherche clinique*. PhD thesis, Université Montpellier 1, 2010.

[ER06] M.J. Evans and J.S. Rosenthal. *Probability and Statistics*. University of Toronto, 2 edition, 2006.

[Fra11] D.A.S. Fraser. Is Bayes Posterior just Quick and Dirty Confidence ? *Statistical Science*, 26(3) :299–316, 2011.

[GMS16] E. Ghashim, É. Marchand, and W.E. Strawderman. On a better lower bound for the frequentist probability of coverage of bayesian credible intervals in restricted parameter spaces. *Statistical Methodology*, 31 :43–57, 2016.

[Gor41] R.D. Gordon. Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. *The Annals of Mathematical Statistics*, 12(3) :364–366, 1941.

[GSL92] A.E. Gelfand, A.F.M Smith, and T.M. Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87 :523–532, 1992.

[Kat61] M. Katz. Admissible and Minimax Estimates of Parameters in Truncated Spaces. *The Annals of Mathematical Statistics*, 32 :136–142, 1961.

[KS88] S. Kumar and D. Sharma. Simultaneous estimation of ordered parameters. *Communications in Statistics : theory and methods*, 17 :4315–4336, 1988.

[LC98] E.L. Lehmann and G. Casella. *Theory of Point Estimation.* Springer Texts in Statistics. Springer, 2 edition, 1998.

[LL03] B. Liseo and N. Loperfido. A Bayesian interpretation of the multivariate skew-normal distribution. *Statistics & Probability Letters*, 61 :395–401, 2003.

[LR05] E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses.* Springer Texts in Statistics. Springer, 3 edition, 2005.

[Man02] M. Mandelkern. Setting Confidence Intervals for Bounded Parameters. *Statistical Science*, 17(2) :149–172, 2002.

[MI05] Y. Maruyama and K. Iwasaki. Sensitivity of minimaxity and admissibility in the estimation of a positive normal mean. *Annals of the Institute of Statistical Mathematics*, 57(1) :145–156, 2005.

[MLT00] M.T. Madi, T. Leonard, and K.-W. Tsui. Bayes inference for treatment effects with uncertain order constraints. *Statistics & Probability Letters*, 49 :277–283, 2000.

[MN19] É. Marchand and T. Nicoleris. On Bayes minimax estimators for a normal mean with an uncertain constraint. *Communications in Statistics - Theory and Methods*, 5 :1873–1883, 2019.

[MP01] É. Marchand and F. Perron. Improving on the mle of a bounded normal mean. *The Annals of Statistics*, 29(4) :1078–1093, 2001.

[MS04] É. Marchand and W.E. Strawderman. Estimation in restricted parameter spaces : A review. *IMS Lecture Notes Monograph*, 45 :1–24, 2004.

[MS05] É. Marchand and W.E. Strawderman. Improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or half-interval. *Annals of the Institute of Statistical Mathematics*, 57(1) :129–143, 2005.

[MS06] É. Marchand and W.E. Strawderman. On the behavior of Bayesian credible intervals for some restricted parameter space problems. *IMS Lecture Notes – Monograph Series, Recent Developments in Nonparametric Inference and Probability*, 50 :112–126, 2006.

[MS12] É. Marchand and W.E. Strawderman. A unified minimax result for restricted parameter spaces. *Bernoulli*, 18 :635–643, 2012.

[MS13] É. Marchand and W.E. Strawderman. On Bayesian credible sets, restricted parameter spaces and frequentist coverage. *Electronic Journal of Statistics*, 7 :1419–1431, 2013.

[MS18] É. Marchand and A. Sadeghkhani. On predictive density estimation with additional information. *Electronic Journal of Statistics*, 12 :4209–4238, 2018.

[MSBL08] É. Marchand, W.E. Strawderman, K. Bosa, and A. Lmoudden. On the frequentist coverage of Bayesian credible intervals for lower bounded means. *Electronic Journal of Statistics*, 2 :1028–1042, 2008.

[Mui82] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*, volume 32 of *Wiley Series in Probability and Statistics*. Wiley, 1982.

[NJSB19] H. Negarestani, A. Jamalizadeh, S. Shafiei, and N. Balakrishnan. Mean mixtures of normal distributions : properties, inference and application. *Metrika*, 82 :501–528, 2019.

[OL76] A. O'Hagan and T. Leonard. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1) :201–203, 1976.

[PK17] L.K. Patra and S. Kumar. Estimating ordered means of a bivariate normal distribution. *American Journal of Mathematical and Management Sciences*, 36(2) :118–136, 2017.

[RW00] B.P. Roe and M.B. Woodroofe. Setting confidence belts. *Physical Review D*, 63, 2000.

[SA20] A. Sadeghkhani and S.E. Ahmed. Bayesian Predictive Densities as an Interpretation of a Class of Skew-Student t Distributions with Application to Medical Data. *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management*, pages 416–428, 2020.

[Sac63] J. Sacks. Generalized Bayes solutions in estimation problems. *The Annals of Mathematical Statistics*, 34 :751–768, 1963.

[Sam53] M.R. Sampford. Some Inequalities on Mill's Ratio and Related Functions. *The Annals of Mathematical Statistics*, 24(1) :130–132, 1953.

[vE06] C. van Eeden. *Restricted Parameter Space Estimation Problems.* Lecture Notes in Statistics. Springer, 2006.

[vEZ01] C. van Eeden and J.V. Zidek. Estimating one of two normal means when their difference is bounded. *Statistics & Probability Letters*, 51 :277–284, 2001.

[vEZ02] C. van Eeden and J.V. Zidek. Combining sample information in estimating ordered normal means. *Sankhyā : The Indian Journal of Statistics*, 64 :588–610, 2002.

[vEZ04] C. van Eeden and J.V. Zidek. Combining the data from two normal populations to estimate the mean of one when their means difference is bounded. *Journal of Multivariate Analysis*, 88 :19–46, 2004.

[Was11] L. Wasserman. Frasian Inference. *Statistical Science*, 26(3) :322–325, 2011.

[ZW03] T. Zhang and M. Woodroofe. Credible and confidence sets for restricted parameter spaces. *Journal of Statistical Planning and Inference*, 115 :479–490, 2003.