Bayesian and minimax estimators of loss

Christine Allard[1] · Éric Marchand[1]

Abstract

We study the problem of loss estimation that involves for an observable $X \sim f_\theta$ the choice of a first-stage estimator $\hat{\gamma}$ of $\gamma(\theta)$, incurred loss $L = L(\theta, \hat{\gamma})$, and the choice of a second-stage estimator $\hat{L}$ of L. We consider both: (i) a sequential version where the first-stage estimate and loss are fixed and optimization is performed at the second-stage level, and (ii) a simultaneous version with a Rukhin-type loss function designed for the evaluation of $(\hat{\gamma}, \hat{L})$ as an estimator of (γ, L). We explore various Bayesian solutions and provide minimax estimators for both situations (i) and (ii). The analysis is carried out for several probability models, including multivariate normal models $N_d(\theta, \sigma^2 I_d)$ with both known and unknown σ^2, Gamma, univariate and multivariate Poisson, and negative binomial models, and relates to different choices of the first-stage and second-stage losses. The minimax findings are achieved by identifying a least favourable sequence of priors and depend critically on particular Bayesian solution properties, namely situations where the second-stage estimator $\hat{L}(x)$ is constant as a function of x.

Keywords Bayes estimation · Loss estimator · Minimax · Posterior distribution · Rukhin-type loss · Unbiased estimator

1 Introduction

Reporting on the precision of statistical decisions, whether it relates to a standard error of an estimate in multiple regression or survey sampling, the power of a test, or the coverage probability of an interval estimate etc., are central to the practice of statistics. Whereas a frequentist approach typically prescribes the level of risk

C. Allard and E. Marchand contributed equally to this work.

✉ Christine Allard
christine.allard@usherbrooke.ca

✉ Éric Marchand
eric.marchand@usherbrooke.ca

[1] Département de mathématiques, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, QC J1K 2R1, Canada

 Springer

prior to the collection of data, while a Bayesian approach typically relates to post-data inference to assess precision, such correspondences are not exclusively the case and various approaches have been presented in the literature, for instance by Berger (1985); Goutis and Casella (1995), as well as the references therein. The setting or search for efficient accuracy reports thus matters and we investigate here various issues and optimality properties cast in a loss estimation framework. The seminal work of Johnstone (1988) put forth a framework for loss estimation with an emphasis on the multivariate normal model and the discovery of improvements in terms of frequentist risk on usual unbiased estimators. Earlier work in this direction includes that of Sandved (1968), but many researchers engaged after Johnstone's work with further investigation towards various extensions in terms of different contexts and other models (e.g., multivariable regression, model selection, spherical symmetry), optimality properties (e.g., admissibility and dominance), and further related issues (e.g., (Boisbunon et al., 2014); (Fourdrinier & Strawderman, 2003); (Fourdrinier & Wells, 1995a, b, 2012); (Fourdrinier & Lepelletier, 2008); (Lele, 1992, 1993); (Lu & Berger, 1989); (Maruyama, 1997); (Matsuda & Strawderman, 2019); (Matsuda, 2024); (Narayanan & Wells, 2015); (Wan & Zou, 2004)).

The approach which we adopt and study is that of accompanying an estimator $\hat{\gamma}$ of an unknown parameter $\gamma(\theta)$ by an estimator $\hat{L}$ of a first-stage loss $L = L(\theta, \hat{\gamma}(x))$. To assess the accuracy of such a $\hat{L}$, we work with a second-stage loss typically of the form $W(L, \hat{L})$. We explore Bayesian estimators $\hat{L}_\pi$ of L and their properties for a given prior density π. We also address minimax optimality for estimating L and provide minimax solutions, which capitalize on the behaviour of Bayesian solutions. Minimax solutions serve as a benchmark for evaluating competing estimators, and have been relatively unexplored in the context of loss estimation.

A particular interesting and different approach, which combines both the first-stage estimation process and the second-stage loss estimation component was proposed and analyzed by Rukhin (1988a, b). Here again, we provide minimax solutions $(\hat{\gamma}, \hat{L})$ which depend critically on the existence of prior densities and associated Bayesian estimators $\hat{L}_\pi$ that do not depend on the observed data.

The paper is organized as follows. Section 2 describes the adopted language of loss estimation namely with the aid of definitions and notations. Section 3 explores Bayesian solutions in cases where both the first-stage and second-stage estimators $\hat{\gamma}_\pi$ and $\hat{L}_\pi$ are derived with respect to the same prior, with an emphasis on varied choices of the first-stage and second-stage losses, and namely departures from the ubiquitous squared error second-stage loss in the literature. Examples of models include normal with or without a known covariance structure, Gamma, univariate and multivariate Poisson, Negative Binomial, and location exponential. We also come across a surprising large number of cases where the Bayes estimator $\hat{L}$ of loss is constant as a function of the observable data x, including cases where even the posterior distribution $L|x$ is free of x. We further explore links between Bayesian and unbiased estimators of loss.

Section 4 provides minimax estimators $\hat{L}$ of loss $L(\theta, \hat{\gamma}(x))$ for different combinations of probability models and losses, and when the first-stage estimate has previously been obtained. To the best of our knowledge, the only known previously analyzed case involves normal models and first and second-stage squared error losses (Johnstone, 1988). We extend the finding to a wider class of first and second-stage losses, to an

unknown covariance structure, and proceed with minimax results for Gamma models. The results are obtained through the determination of a least favourable sequence $\{\pi_n\}$ of prior densities and an extended Bayes estimator with constant risk.

We point out that, unlike usual estimation problems, estimands of the form $L(\theta, \hat{\gamma}(x))$ depend not only on θ but also on x. Nevertheless, approaches to establishing minimaxity remain familiar ones, namely through the determination of a sequence π_n of least favourable prior densities.

Finally, we obtain minimax solutions $(\hat{\gamma}_0, \hat{L}_0)$ for estimating (γ, L) simultaneously under Rukhin-type losses, by using a sequence of priors approach to obtain an extended Bayes estimator with constant frequentist risk. Properties established or observed in Sect. 3, namely the constancy of Bayesian solutions $\hat{L}_\pi(x)$ as functions of x play a key role for the analysis.

2 Definitions and other preliminaries

Throughout, we consider a model density $f_\theta, \theta \in \Theta$, for an observable X, a parametric function $\gamma(\theta)$ of interest; mostly taken as identity; and the loss $L = L(\theta, \hat{\gamma})$ measuring the level of accuracy (or inaccuracy) of $\hat{\gamma}(X)$ as an estimator of $\gamma(\theta)$. Such a loss L is referred to as a first-stage loss, while loss $W(L, \hat{L}) \in [0, \infty)$ used for estimating L by $\hat{L}(X)$ is referred to as the second-stage loss.

Remark 1 A common and default choice in the literature for the second-stage loss W has been squared error $(\hat{L} - L)^2$ loss. This has been described as a matter of convenience, but it is also the case that the developments for multivariate normal models, as well as spherically symmetric models, bring into play Stein's lemma and convenient loss estimation representations arising for squared error loss (e.g. (Fourdrinier et al., 2018)). However, since a loss L is more synonymous with a positive scale parameter than a location parameter, it seems desirable to consider a scale invariant second-stage loss of the form $\rho(\frac{\hat{L}}{L})$, plausible choices satisfying the bowl-shaped property $\rho(t)$ decreasing for $t \in (0, 1)$ and increasing for $t > 1$. The developments in this manuscript thus address such losses amongst a wider choice of second-stage losses. Such examples include weighted squared error loss with $\rho(t) = (t - 1)^2$, squared log error loss with $\rho(t) = (\log t)^2$, symmetric versions with $\rho(t) = \rho(\frac{1}{t})$ like $\rho(t) = t + \frac{1}{t} - 2$ (e.g., (Mozgunov et al., 2019)), entropy loss with $\rho(t) = t - \log t - 1$, and variants of the above (except log error) with $\rho_m(t) = \rho(t^m)$, $m \neq 0$, the case $m = -1$ being most prominent for squared error and entropy.

For a given prior density π for θ and estimate $\hat{\gamma}(x)$, a Bayes estimator $\hat{L}_\pi(X)$ of $L = L(\theta, \hat{\gamma}(x))$ minimizes in $\hat{L}$ the expected posterior loss $\mathbb{E}(W(L, \hat{L})|x)$. As addressed in Sect. 3, it is particularly interesting to study cases where $\hat{\gamma} \equiv \hat{\gamma}_\pi$, i.e., the Bayes estimator of $\gamma(\theta)$ under the same prior. However, it is still useful to consider the more general context, for instance because the first-stage estimator may be imposed and not be Bayesian, or a theoretical assessment of a least favourable sequence of priors such as the one pursued in Sect. 4 requires it. In such a framework, the stated goal is how to report on sensible estimates of L for a given $\hat{\gamma}(x)$.

The frequentist risk performance of a given estimator $\hat{L}$ for estimating a loss $L = L(\theta, \hat{\gamma}(x))$ is given by

$$R(\theta, \hat{L}) = \mathbb{E}_\theta\big(W(L, \hat{L}(X))\big), \theta \in \Theta, \tag{1}$$

and different choices of $\hat{L}$ can be compared leading to the usual definitions of dominance, admissibility, and inadmissibility. Our findings relate also to the minimax criterion with estimator $\hat{L}_m(X)$ being a minimax estimator of L whenever $\sup_{\theta \in \Theta}\{R(\theta, \hat{L}_m)\} = \inf_{\hat{L}} \sup_{\theta \in \Theta}\{R(\theta, \hat{L})\}$.

Another criterion present in the literature, that sometimes interacts with Bayesianity, is that of unbiasedness. For a given estimator $\hat{\gamma}(X)$ of $\gamma(\theta)$, an estimator $\hat{L}(X)$ of loss $L(\theta, \hat{\gamma})$ is said to be unbiased if

$$\mathbb{E}_\theta \hat{L}(X) = \mathbb{E}_\theta L\big(\theta, \hat{\gamma}(X)\big), \text{ for all } \theta, \tag{2}$$

i.e., $\hat{L}(X)$ is an unbiased estimator of the frequentist risk of $\hat{\gamma}(X)$ as an estimator of $\gamma(\theta)$.

Example 1 As an illustration, consider the normal model $X \sim N_d(\theta, \sigma^2 I_d)$ with known σ^2, the benchmark estimator $\hat{\theta}_0(X) = X$ of θ, and the incurred squared error loss $L = L(\theta, \hat{\theta}(X)) = \|X - \theta\|^2$. Johnstone (1988) showed that the estimator $\hat{L}_0(X) = d\sigma^2$; equal here to the mean squared error risk of $\hat{\theta}_0$ making it an unbiased estimator of loss; matches the (generalized) Bayesian estimator $\hat{L}_{\pi_0}(X)$ of L for second-stage squared error loss $W(L, \hat{L}) = (\hat{L} - L)^2$ and the uniform prior density $\pi_0(\theta) = 1$. He also established that $\hat{L}_{\pi_0}$ is minimax for all $d \geq 1$, admissible for $d \leq 4$, and inadmissible for $d \geq 5$ providing dominating estimators $\hat{L}$ in such a latter case. Such dominating procedures are necessarily minimax. Our findings (Sect. 4.1) for normal models relate to minimaxity for different choices of first-stage (L) and second-stage (W) losses, and address the case of unknown σ^2 (Sect. 4.1.1).

Rukhin (1988a, b) studied the efficiency of estimators, namely in terms of admissibility, with his proposal to combine the two stages of estimation and to measure the performance of the pair $(\hat{\gamma}, \hat{L})$ for estimating $\big(\gamma(\theta), L\big)$, with $L = L(\theta, \hat{\gamma})$ by the loss

$$\hat{L}^{-1/2}L(\theta, \hat{\gamma}) + \hat{L}^{1/2}, \tag{3}$$

as well as extensions

$$W(\theta, \hat{\gamma}, \hat{L}) = h'(\hat{L})L(\theta, \hat{\gamma}) - h'(\hat{L})\hat{L} + h(\hat{L}), \tag{4}$$

h being an increasing and concave function on $(0, \infty)$, and the former being a particular case of the latter for $h(\hat{L}) = 2\hat{L}^{1/2}$. The two components are referred to as error of estimation and precision of estimation, and such a loss is appealing namely since $\hat{L} = L$ minimizes the loss for fixed L and since the Bayesian estimator of L for a given prior π is equal to the posterior expectation $\hat{L}_\pi(x) = \mathbb{E}(L|x)$.

3 Bayesian estimators

In this section, we record various interesting scenarios concerning Bayesian inference about a given loss $L = L(\theta, \hat{\gamma})$ incurred by estimator $\hat{\gamma}(X)$ for estimating $\gamma(\theta)$. Different choices of L and the second-stage loss $W(L, \hat{L})$ are considered for the determination of a point estimator $\hat{L}_\pi$, and we also describe directly the posterior distribution $L|x$ in some instances.

Section 3.1 deals with situations where a same prior π is used to determine both the choices of $\hat{\gamma}_\pi$ and $\hat{L}_\pi$, with a particular focus on Poisson and negative binomial models. While these last two examples involve cases where point estimates $\hat{L}_\pi(x)$ do not depend on x, Section 3.2 deals with specific situations where the posterior distribution $L|x$ does not depend on x. The features and representations provided in the Section will prove useful for the minimax findings of Sections 4 and 5.

3.1 Examples of $(\hat{\gamma}_\pi, \hat{L}_\pi)$ pairs

From a Bayesian perspective with the same prior π as for the first-stage, one could naturally consider the posterior distribution of $L|x$ for inference about L. Minimizing the expected posterior loss $W(L, \hat{L})$ in $\hat{L}$ produces Bayes estimate $\hat{L}_\pi$ and the ensemble produces pairs $(\hat{\gamma}_\pi, \hat{L}_\pi)$ which we investigate and illustrate in this section.

We begin with the familiar case of squared error loss $L(\theta, \hat{\theta}) = \|\hat{\theta} - \theta\|^2$ for estimating $\gamma(\theta) = \theta \in \mathbb{R}^d$ based on $X \sim f_\theta(\cdot)$. Assuming that the posterior covariance matrix $Cov(\theta|x)$ of θ exists, we have

$$\hat{\gamma}_\pi(X) = \mathbb{E}(\theta|x) \text{ with incurred loss } L = \|\theta - \mathbb{E}(\theta|x)\|^2.$$

Now, if the second-stage loss is again squared error loss, i.e., $W(L, \hat{L}) = (\hat{L} - L)^2$, then we obtain

$$\hat{L}_\pi(x) = \mathbb{E}(L|x) = \text{tr } Cov(\theta|x). \tag{5}$$

Example 2 For location family densities $f_\theta(x) = f_0(x - \theta)$, $x, \theta \in \mathbb{R}^d$, such that $\mathbb{E}_\theta(X) = \theta$, which include spherically symmetric densities $g(\|x - \theta\|^2)$, and non-informative prior $\pi(\theta) = 1$, we obtain $\hat{\gamma}_\pi(x) = x$. Since $x - \theta|x =^d X - \theta|\theta$ for all x, θ, it follows that the posterior distribution of $L|x$, which is that of $\|\theta - x\|^2|\theta$, is independent of x. Consequently, $\hat{L}_\pi(x)$ is a constant in terms of x. This observation includes the familiar multivariate normal case with $X|\theta \sim N_d(\theta, \sigma^2 I_d)$ where $L|x \sim \sigma^2 \chi_d^2$ and $\hat{L}_\pi(x) = d\sigma^2$. With the posterior distribution of L independent of x, Bayes estimators $\hat{L}_\pi(X)$ associated with other second-stage losses or even credibility intervals for L will necessarily be independent of x. These last features fit into a more general structure expanded upon with Theorem 1 and include proper priors as with the following example.

Example 3 The general normal model with conjugate normal priors is also quite tractable. Indeed, for model $X|\theta \sim N_d(\theta, \sigma^2 I_d)$ and prior $\theta \sim N_d(\mu, \tau^2 I_d)$, we have $\theta|x \sim N_d(\mu(x), \tau^2(x)I_d)$ with $\hat{\theta}_\pi(x) = \mathbb{E}(\theta|x) = \frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\mu$ and

Table 1 Bayesian loss estimators

$\rho(t)$	$\hat{L}_\pi$ (general)	$\hat{L}_\pi$ (conditions)
$\rho_A(t) = (t^m - 1)^2$	$\left(\frac{\mathbb{E}(L^{-m}\|x)}{\mathbb{E}(L^{-2m}\|x)}\right)^{1/m}$	$2\tau_0^2 \left(\frac{\Gamma(\frac{d}{2}-m)}{\Gamma(\frac{d}{2}-2m)}\right)^{1/m}$ for $d > 4m$
$\rho_m(t) = t^m - m\log t - 1$	$(\mathbb{E}(L^{-m}\|x))^{-1/m}$	$2\tau_0^2 \left(\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}-m)}\right)^{1/m}$ for $d > 2m$
$\rho_B(t) = t + \frac{1}{t} - 2$	$\sqrt{\frac{\mathbb{E}(L\|x)}{\mathbb{E}(\frac{1}{L}\|x)}}$	$\tau_0^2\sqrt{d(d-2)}$ for $d \geq 3$
$\rho_C(t) = (\log t)^2$	$e^{\mathbb{E}(\log L\|x)}$	$2\tau_0^2 e^{\Psi(\frac{d}{2})}$ for $d \geq 1$

$\tau^2(x) = \tau_0^2 = \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}$. From this, one obtains

$$L|x = \|\theta - \mathbb{E}(\theta|x)\|^2 \big| x \sim \tau_0^2 \chi_d^2, \tag{6}$$

again independent of x. For second-stage squared error loss, one obtains $\hat{L}_\pi(x) = \tau_0^2 d$.

As addressed in Remark 1, second-stage losses of the form $\rho(\frac{\hat{L}}{L})$ are desirable alternatives given their scale invariance. Table 1 provides Bayesian estimators $\hat{L}_\pi$ of L for some choices of ρ. The third column is specific to the normal model context here, while the second column expressions apply more broadly. For the loss $\rho_C(t)$, $\Psi(t)$ is the Digamma function given by $\Psi(t) = \frac{d}{dt}\log\Gamma(t)$. Various degrees of shrinkage or expansion in comparison to second-stage squared error loss; for which $\hat{L}_\pi(X) = d\tau_0^2$; are observable. For instance with loss ρ_A and $d \geq 5$, we have $\hat{L}_\pi(X) = \tau_0^2(d+2)$ versus $\hat{L}_\pi(X) = \tau_0^2(d-4)$ according to the selections $m = -1$ or $m = 1$, respectively. Shrinkage occurs for ρ_B and ρ_C (see Remark 5), while $\hat{L}_\pi$ is decreasing as a function of m for both ρ_A and ρ_m, with shrinkage iff $m > -1$ for ρ_m, and iff $m > m_0(d)$ for ρ_A with $m_0(d) \in (-1, 0)$ such that $\hat{L}_\pi = d\tau_0^2$ at $m = m_0(d)$ (see Appendix).

The next examples involve weighted squared error loss as the first-stage which is a typical choice when the model variance varies with θ, such as Poisson and negative binomial. To this end, consider first-stage loss as weighted squared error loss $L(\theta, \hat{\gamma}) = \omega(\theta)(\hat{\gamma} - \gamma(\theta))^2$ for $X \sim f_\theta$, $\gamma(\theta) \in \mathbb{R}$. Given a posterior density for θ, the Bayes estimator of $\gamma(\theta)$ is given, whenever it exists, by the familiar

$$\hat{\gamma}_\pi(x) = \frac{\mathbb{E}(\gamma(\theta)\omega(\theta)|x)}{\mathbb{E}(\omega(\theta)|x)}, \tag{7}$$

with incurred loss given by $L = \omega(\theta)(\hat{\gamma}_\pi(x) - \gamma(\theta))^2$. For second-stage squared error loss $(\hat{L} - L)^2$, we obtain the Bayes estimator

$$\hat{L}_\pi(x) = \mathbb{E}(L|x) = \mathbb{E}(\gamma(\theta)^2\omega(\theta)|x) - \frac{\mathbb{E}^2(\gamma(\theta)\omega(\theta)|x)}{\mathbb{E}(\omega(\theta)|x)}, \tag{8}$$

as long as $\mathbb{E}(L^2|x)$ exists.

3.1.1 Poisson distribution

Consider a Poisson model $X|\theta \sim$ Poisson(θ) with a Gamma prior $\theta \sim$ G(a, b) (density proportional to $\theta^{a-1}e^{-\theta b}\mathbb{I}_{(0,\infty)}(\theta)$ throughout the manuscript), and the estimation of $\gamma(\theta) = \theta$ with normalized squared error loss $\frac{(\hat{\theta}-\theta)^2}{\theta}$. The set-up leads to $\theta|x \sim$ Ga$(a + x, 1 + b)$, and Bayes estimator

$$\hat{\gamma}_\pi(x) = \frac{1}{\mathbb{E}(1/\theta|x)} = \frac{a + x - 1}{1 + b},$$ (9)

for $a > 1, b > 0$. For the case $(a, b) = (1, 0)$, i.e., the uniform prior density on $(0, \infty)$, the (generalized) Bayes estimator is also given by (9), i.e., $\hat{\gamma}_\pi(X) = X$, and, moreover, is the unique minimax estimator of θ. The incurred loss by the Bayes estimator (9) becomes

$$L = \frac{(\frac{a+x-1}{1+b} - \theta)^2}{\theta},$$ (10)

and the Bayes estimator $\hat{L}_\pi$ in (8) becomes

$$\hat{L}_\pi(x) = \mathbb{E}(\theta|x) - \frac{1}{\mathbb{E}(\frac{1}{\theta}|x)} = \frac{a + x}{1 + b} - \frac{a + x - 1}{1 + b} = \frac{1}{1 + b},$$ (11)

provided $a > 2$ as the existence of $\mathbb{E}(L^2|x)$ requires finite $\mathbb{E}(\theta^{-2}|x)$ which in turn necessitates $a > 2$. Observe that estimate (11) is independent of x and of the hyperparameter a.

Remark 2 A few remarks:

(I) Under the above set-up, Lele (1993) established the admissibility of the posterior expectation $\hat{L}_0(X) = \mathbb{E}(L|X) = 1$ as an estimator of $L = \frac{(X-\theta)^2}{\theta}$ under squared error loss $(\hat{L} - L)^2$. Another interesting property of $\hat{L}_0(X)$ is that of unbiasedness, as can be seen by the risk calculation $\mathbb{E}(L|\theta) = 1$.

(II) The frequentist risk of $\hat{L}_0(X)$ is given by

$$R(\theta, \hat{L}_0) = \mathbb{E}(\hat{L}_0(X) - L)^2 = \mathbb{V}(L|\theta) = \frac{\mathbb{E}(X - \theta)^4}{\theta^2} - 1 = 2 + \frac{1}{\theta},$$

using the fact the fourth central moment of a Poisson distribution with mean θ is given by $\mathbb{E}(X - \theta)^4 = \theta(1 + 3\theta)$. Observe that the supremum risk is equal to $+\infty$ which is not conducive to the property of minimaxity. However, $\hat{L}_0(X)$ is (unique) minimax for estimating L under weighted second-stage loss $\frac{\theta}{2\theta+1}(\hat{L} - L)^2$ because it remains admissible, it has constant risk, and such estimators are necessarily minimax.

(III) For $a > 2$ and $b > 0$, the estimators $\hat{L}_\pi(X)$ of L given in (11) are proper Bayes and admissible since the corresponding integrated Bayes risks are finite. The finiteness can be justified by the fact that $\int_0^\infty R(\theta, \hat{L}_\pi)\pi(\theta)d\theta \leq \int_0^\infty R(\theta, \hat{L}_0)\pi(\theta)d\theta = \int_0^\infty (2 + \frac{1}{\theta})\pi(\theta)d\theta = 2 + \frac{b}{a-1} < \infty$

3.1.2 Poisson distribution (multivariate case)

As a multivariate extension, consider $X = (X_1, \ldots, X_d)$ with $X_i \sim \text{Poisson}(\theta_i)$ independent, the first-stage loss $L = L(\theta, \hat{\theta}) = \sum_{i=1}^d \frac{(\hat{\theta}_i - \theta_i)^2}{\theta_i}$ for estimating $\theta = (\theta_1, \ldots, \theta_d)$ based on $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_d)$, and second-stage squared error loss. Proceeding as for the case $d = 1$, we have for a given prior π the Bayes estimators (whenever well defined):

$$\hat{\theta}_{\pi,i}(x) = \left(\mathbb{E}(\theta_i^{-1}|x)\right)^{-1}, i = 1, \ldots, d, \tag{12}$$

and

$$\hat{L}_\pi(x) = \mathbb{E}(L|x) = \sum_{i=1}^d \mathbb{E}(\theta_i|x) - \sum_{i=1}^d \left(\mathbb{E}(\theta_i^{-1}|x)\right)^{-1}. \tag{13}$$

As an example, a familiar prior specification choice for π (e.g., (Clevenson & Zidek, 1975)) brings into play $S = \sum_{i=1}^d \theta_i$ and $U_i = \frac{\theta_i}{S}$, $i = 1, \ldots, d$, and density $(S, U) \sim h(s)\mathbb{I}_{\{1\}}(\sum_i u_i)$, where $h(\cdot)$ is a density on $\mathbb{R}_+$. With such a choice and setting $Z = \sum_{i=1}^d x_i$ hereafter, one obtains the posterior density representation $U|s, x \sim \text{Dirichlet}(x_1 + 1, \ldots, x_d + 1)$ and $h(s|x) \propto s^Z e^{-s} h(s)$. With U and S independently distributed under the posterior and $\text{Beta}(x_i + 1, Z - x_i + d - 1)$ marginals for the U_i's, the evaluation of (12) and (13) is facilitated and yields a Clevenson–Zidek type estimator of θ and accompanying loss estimator

$$\hat{\theta}_\pi(X) = \frac{X}{Z + d - 1}\left(\mathbb{E}(S^{-1}|X)\right)^{-1}, \text{ and } \hat{L}_\pi(X) = \mathbb{E}(S|X) - \frac{Z}{Z + d - 1}\left(\mathbb{E}(S^{-1}|X)\right)^{-1}.$$

For a gamma prior $S \sim \text{G}(a, b)$ with $a \geq 1$, we have $S|x \sim \text{G}(a + Z, b + 1)$ and the above reduces to

$$\hat{\theta}_{\pi,a,b}(X) = \frac{X}{Z + d - 1}\frac{a + Z - 1}{b + 1}, \text{ and } \hat{L}_{\pi,a,b}(X) = \frac{dZ + a(d - 1)}{(b + 1)(Z + d - 1)}.$$

Notice that the univariate $\hat{L}_\pi$ given previously in (11) is recovered from the above for $d = 1$, while the case $a = d, b = 0$ yields the unbiased estimators $\hat{\theta}_{\pi,d,0}(X) = X$ and $\hat{L}_{\pi,d,0}(X) = d$. We point out that Lele (1992, 1993) established: **(i)** in the latter case, the admissibility of $\hat{L}_\pi(X) = d$ as an estimator of loss $L(\theta, X)$ for $d = 1, 2$, and inadmissibility for $d \geq 3$; and **(ii)** the admissibility of $\hat{L}_{\pi,1,0}$ as an estimator of $L(\theta, \hat{\theta}_{\pi,1,0}(x))$ for all $d \geq 1$. As in Remark 2 for the bivariate case, we point out that

$\hat{L}_{\pi,d,0}(X) = 2$ has frequentist risk equal to $4 + \frac{1}{\theta_1} + \frac{1}{\theta_2}$, infinite supremum risk, and that it is minimax for weighted squared error second-stage loss $\frac{(\hat{L}-L)^2}{4+\frac{1}{\theta_1}+\frac{1}{\theta_2}}$.

Remark 3 In the specific situation where $S \sim G(d, b)$, one verifies that the above prior reduces to independently distributed $\theta_i \sim G(1, b)$ for $i = 1, \ldots, d$. Since the X_i's are also independently distributed given the θ_i's, the multivariate Bayesian estimation problem reduces to the juxtaposition of d independent univariate problems as analyzed in the previous section. For instance, expressions (9) and (11) applied to the components θ_i lead to the above $\hat{\theta}_{\pi,d,b}(X)$ and $\hat{L}_{\pi,d,b}(X)$, and the same remains true for the improper proper choice with $b = 0$.

3.1.3 Negative binomial distribution

Consider a negative binomial model $X \sim NB(r, \theta)$ such that

$$\mathbb{P}(X = x|\theta) = \frac{(r)_x}{x!} \left(\frac{r}{\theta + r}\right)^r \left(\frac{\theta}{\theta + r}\right)^x \mathbb{I}_{\mathbb{N}}(x), \qquad (14)$$

with $r > 0$, $\mathbb{E}(X|\theta) = \theta > 0$, $(r)_x$ the Pochhammer symbol representing the quantity $(r)_x = \frac{\Gamma(r+x)}{\Gamma(r)}$ and where we study pairs $(\hat{\theta}_\pi, \hat{L}_\pi)$ for a class of Beta type II priors for θ which are conjugate and defined more generally as follows.

Definition 1 A Beta type II distribution, denoted as $Y \sim B2(a, b, \sigma)$ with $a, b, \sigma > 0$ has density of the form

$$f(y) = \sigma^b \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \frac{y^{a-1}}{(\sigma + y)^{a+b}} \mathbb{I}_{(0,\infty)}(y).$$

The following identity, which is readily verified, will be particularly useful.

Lemma 1 For $Y \sim B2(a, b, \sigma)$, $\gamma_1 > -a$, and $\gamma_2 > \gamma_1 - b$, we have

$$\mathbb{E}\left(\frac{Y^{\gamma_1}}{(\sigma + Y)^{\gamma_2}}\right) = \frac{(a)_{\gamma_1}}{(a+b)_{\gamma_2}} \frac{(b)_{\gamma_2 - \gamma_1}}{\sigma^{\gamma_2 - \gamma_1}} \qquad (15)$$

where, for $\alpha > 0$ and $\alpha + m > 0$, $(\alpha)_m$ is the Pochhammer symbol as defined above.

It is simple to verify the following (e.g., (Ferguson, 1968), page 96).

Lemma 2 For $X|\theta \sim NB(r, \theta)$ with prior $\theta \sim B2(a, b, r)$, the posterior distribution is $\theta|x \sim B2(a + x, b + r, r)$.

Now with such a prior, for estimating θ, since $\mathbb{V}(X|\theta) = \theta(\theta + r)/r$, under normalized squared error loss

$$L(\theta, \hat{\theta}) = \frac{(\hat{\theta} - \theta)^2}{\theta(\theta + r)}, \qquad (16)$$

the Bayes estimate of θ may be derived from (7) and (15) as

$$\hat{\theta}_\pi(x) = \frac{\mathbb{E}\big(\frac{1}{(\theta+r)}|x\big)}{\mathbb{E}\big(\frac{1}{\theta(\theta+r)}|x\big)} = r\frac{a+x-1}{b+r+1},$$

for $a + x > 1$. For $a = 1$ and $x = 0$, a direct evaluation yields $\hat{\theta}_\pi(0) = 0$, which matches the above extended to $x = 0$. The associated loss $L(\theta, \hat{\theta}_\pi(x))$ has posterior expectation for $a > 1$ equal to

$$\hat{L}_\pi(x) = \mathbb{E}\big(\frac{\theta}{\theta+r}|x\big) - \frac{\mathbb{E}^2\big(\frac{1}{\theta+r}|x\big)}{\mathbb{E}\big(\frac{1}{\theta(\theta+r)}|x\big)}$$

$$= \frac{a+x}{a+b+x+r} - \big(\frac{b+r}{b+r+1}\big)\big(\frac{a+x-1}{a+b+x+r}\big) = \frac{1}{b+r+1},$$

making use of (8) and (15). Interestingly, the estimator does not depend on a and is constant as a function of x and this property will play a key role for the minimax findings of Sect. 5. We conclude by pointing out that the above applies to the improper prior $\theta \sim B2(1, 0, r)$ yielding the generalized Bayes estimator $\hat{\theta}_0(x) = rx/(r+1)$. It is known (e.g., Ferguson, 1968) that the estimator $\hat{\theta}_0$ is minimax with minimax risk equal to $1/(r+1)$.

3.2 Posterior distributions for loss that do not depend on *x*

There are a good number of instances; some of which have appeared in the literature; where both the posterior distribution of loss $L(\theta, \hat{\gamma})$ and (consequently) the Bayes estimate with respect to loss $W(L, \hat{L})$ are free of the observed x. Such a property is particularly interesting and will play a critical role for the minimax implications in Sect. 4. We describe situations where such a property arises and collect some examples. The situations correspond to similar scenarios mathematically and relate specifically to cases where the posterior density admits: (I) a location invariant, (II) a scale invariant, or (III) a location-scale invariant structure.

Theorem 1 *Suppose that the posterior density for $\gamma(\theta) = \theta$ is, for all x, location invariant of the form $\theta|x \sim f(\theta - \mu(x))$ and that the first-stage loss for estimating θ is location invariant, i.e., of the form $L(\theta, \hat{\theta}) = \beta(\hat{\theta} - \theta); \beta : \mathbb{R}^d \to \mathbb{R}_+$. Then,*

(a) *the Bayes estimator $\hat{\theta}_\pi(X)$, whenever it exists, is of the form $\hat{\theta}_\pi(x) = \mu(x) + k$, k being a constant;*
(b) *the posterior distribution of the loss $\beta(\hat{\theta}_\pi(x) - \theta)$ is free of x;*
(c) *the Bayes estimator $\hat{L}_\pi(x)$ of the loss $\beta(\hat{\theta}_\pi(x) - \theta)$ with respect to second-stage loss $W(L, \hat{L})$ is, whenever it exists, free of x.*

Proof Part **(c)** follows from part **(b)**. Now, observe that

$$\inf_{\hat{\theta}\in\mathbb{R}^d} \mathbb{E}\Big\{\beta(\hat{\theta} - \theta)|x\Big\} = \inf_{\hat{\theta}\in\mathbb{R}^d} \mathbb{E}\Big\{\beta\Big(\hat{\theta} - \mu(x) - \big(\theta - \mu(x)\big)\Big)|x\Big\}$$
$$= \inf_{\hat{\alpha}\in\mathbb{R}^d} \mathbb{E}\big\{\beta(\hat{\alpha} - Z)|x\big\}$$
$$= \mathbb{E}\big\{\beta(\hat{\alpha}_\pi(x) - Z)|x\big\},$$

for all x, with $\hat{\alpha} = \hat{\theta} - \mu(x)$, $\hat{\alpha}_\pi(x) = \hat{\theta}_\pi(x) - \mu(x)$, and $Z|x =^d \theta - \mu(x)|x$. Part **(a)** follows since the distribution of $Z|x$ does not depend on x and hence the minimizing $\hat{\alpha}$ (i.e., $\hat{\alpha}_\pi(x)$) does not depend on x. Finally, part **(b)** follows since

$$\beta\big(\hat{\theta}_\pi(x) - \theta\big)|x =^d \beta(k - Z)|x$$

is free of x. □

We pursue with similar findings as described in (II) and (III) above

Theorem 2 *Suppose that the posterior density for θ is, for all x, scale invariant of the form $\theta|x \sim \frac{1}{\sigma(x)} f(\frac{\theta}{\sigma(x)})$, $\theta \in \mathbb{R}_+$, and that the first-stage loss for estimating θ is scale invariant, i.e., of the form $L(\theta, \hat{\theta}) = \rho(\frac{\hat{\theta}}{\theta})$; $\rho : \mathbb{R} \to \mathbb{R}_+$. Then,*

(a) *the Bayes estimator $\hat{\theta}_\pi(X)$, whenever it exists, is of the form $\hat{\theta}_\pi(x) = k\sigma(x)$, k being a constant;*

(b) *the posterior distribution of the loss $\rho(\frac{\hat{\theta}_\pi(x)}{\theta})$ is free of x;*

(c) *the Bayes estimator $\hat{L}_\pi(x)$ of the loss $\rho(\frac{\hat{\theta}_\pi(x)}{\theta})$ with respect to second-stage loss $W(L, \hat{L})$ is, whenever it exists, free of x.*

Proof A similar development to the proof of Theorem 1 establishes the results. □

The next result inspired initially by the context of estimation of a multivariate normal mean with unknown covariance matrix (see Example 7) is presented in a more general setting.

Theorem 3 *Suppose that the posterior density for $\theta = (\theta_1, \theta_2)$ is, for all x, location-scale invariant of the form*

$$\theta|x \sim \frac{1}{\sigma(x)\theta_2^d} f\left(\frac{\theta_1 - \mu(x)}{\theta_2}, \frac{\theta_2}{\sigma(x)}\right), \tag{17}$$

with $\theta_1 \in \mathbb{R}^d$, $\theta_2 \in \mathbb{R}_+$, and that the first-stage loss for estimating θ_1 is location-scale invariant, i.e., of the form $L(\theta, \hat{\theta}_1) = \rho(\frac{\hat{\theta}_1 - \theta_1}{\theta_2})$. Then,

(a) *the Bayes estimator $\hat{\theta}_{1,\pi}(X)$, whenever it exists, is of the form $\hat{\theta}_{1,\pi}(x) = \mu(x) + k\sigma(x)$, k being a constant;*

(b) *the posterior distribution of the loss $L\big(\theta, \hat{\theta}_{1,\pi}(x)\big)$ is free of x;*

(c) *the Bayes estimator $\hat{L}_\pi(x)$ of the loss $L\big(\theta, \hat{\theta}_{1,\pi}(x)\big)$ with respect to second-stage loss $W(L, \hat{L})$ is, whenever it exists, free of x.*

Proof Part **(c)** follows from part **(b)**. Observe that

$$\inf_{\hat{\theta}_1 \in \mathbb{R}^d} \mathbb{E}\left\{\rho\left(\frac{\hat{\theta}_1 - \theta_1}{\theta_2}\right)\Big|x\right\} = \inf_{\alpha \in \mathbb{R}^d} \mathbb{E}\left\{\rho\left(\frac{\alpha - Z}{V}\right)\Big|x\right\},$$

with $\alpha = \frac{\hat{\theta}_1 - \mu(x)}{\sigma(x)}$, $Z|x \stackrel{d}{=} \frac{\theta_1 - \mu(x)}{\sigma(x)}|x$, and $V|x \stackrel{d}{=} \frac{\theta_2}{\sigma(x)}|x$. Since the pair $(Z, V)|x$ has joint density $\frac{1}{v^d} f(\frac{z}{v}, v)$ which is free of x, the minimizing α is free of x which yields part **(a)**. Finally for part **(b)**, observe that $\rho(\frac{\hat{\theta}_{1,\pi}(x) - \theta_1}{\theta_2})|x \stackrel{d}{=} \rho(\frac{k-Z}{V})|x$, which is indeed free of x given the above. $\qquad \square$

3.2.1 Examples

First examples that come to mind are given by the non-informative prior density choices:

(i) $\pi(\theta) = 1$ for the location model density $X|\theta \sim f_0(x - \theta)$, $x, \theta \in \mathbb{R}^d$, with $f(t) = f_0(-t)$ and $\mu(x) = x$ (Theorem 1);

(ii) $\pi(\theta) = \frac{1}{\theta}$ for the scale model density $X|\theta \sim \frac{1}{\theta} f_1(\frac{x}{\theta})$, $x, \theta \in \mathbb{R}_+$, with $f(u) = \frac{1}{u^2} f_1(\frac{1}{u})$ and $\sigma(x) = x$ (Theorem 2);

(iii) $\pi(\theta) = \frac{1}{\theta_2} \mathbb{I}_{(0,\infty)}(\theta_2) \mathbb{I}_{\mathbb{R}^d}(\theta_1)$ for $X = (X_1, X_2)|\theta \sim \frac{1}{\theta_2^{d+1}} f_{0,1}(\frac{x_1 - \theta_1}{\theta_2}, \frac{x_2}{\theta_2})$ with $f(u, v) = \frac{1}{v^2} f_{0,1}(-u, \frac{1}{v})$, $\mu(x) = x_1$, $\sigma(x) = x_2$ (Theorem 3).

Applications of the above theorems are however not limited to such improper priors and we pursue with further proper prior examples.

Example 4 (Multivariate normal model with known covariance) Theorem 1 applies for the normal model set-up of Example 3 since the posterior distribution is of the form $f(\theta - \mu(x))$. For instance, under second-stage squared error loss, identity (6) and $\hat{L}_\pi(X) = \tau_0^2 d$ are illustrative of parts **(b)** and **(c)** of the theorem.

Theorem 1 applies as well to many other first-stage and second-stage losses, such as L^q and reflected normal first-stage losses $\beta(t) = \|t\|^q$ and $1 - e^{-c\|t\|^2}$, with $c > 0$; and second-stage losses of the form $\rho(\frac{\hat{L}}{L})$ such as those of Remark 1. Finally, as previously mentioned, the above observations apply for the improper prior density $\pi(\theta) = 1$ with corresponding expressions obtained by taking $\tau^2 = +\infty$.

Example 5 (A Gamma model) Gamma distributed sufficient statistics appear in many contexts and we consider here $X|\theta \sim G(\alpha, \theta)$ with a Gamma distributed prior $\theta \sim G(a, b)$, which results in the scale invariant form of Theorem 2 with f a $G(\alpha + a, 1)$ density and $\sigma(x) = (x + b)^{-1}$. Therefore, Theorem 2 applies for scale invariant losses $\rho(\frac{\hat{\theta}}{\theta})$ as those referred to in Remark 1. As an illustration, take entropy-type losses of the form with $\rho_m(t) = t^m - m \log(t) - 1$, $m \neq 0$, for which one obtains for $m < a + \alpha$ the Bayes estimator

$$\hat{\theta}_\pi(x) = \left\{ \mathbb{E}(\theta^{-m}|x) \right\}^{-1/m} = k\sigma(x),$$

with $k = \left\{ \frac{\Gamma(a+\alpha)}{\Gamma(a+\alpha-m)} \right\}^{1/m}$, and the posterior distribution of the loss $\rho_m(\frac{\hat{\theta}_\pi(x)}{\theta})$ free of x and matching that of $\rho_m(Z^{-1})$ (or equivalently $\rho_{-m}(Z)$) with $Z \sim G(a + \alpha, k)$. Finally, the Bayes estimate $\hat{L}_\pi(x)$ will also be free of x for any second-stage loss. For

the case of squared error loss $W(L, \hat{L}) = (\hat{L} - L)^2$, one obtains

$$
\begin{aligned}
\hat{L}_\pi(x) &= \mathbb{E}\left\{ \rho_m\left(\frac{\hat{\theta}_\pi(x)}{\theta} \right) \Big| x \right\} \\
&= \mathbb{E}\left\{ \rho_m(Z^{-1}) \right\} \\
&= \mathbb{E}\left(Z^{-m} + m\mathbb{E}\log Z - 1 \right) \\
&= m\Psi(a+\alpha) + \log\left(\frac{\Gamma(a+\alpha-m)}{\Gamma(a+\alpha)} \right),
\end{aligned}
\tag{18}
$$

for $m < a + \alpha$, where Ψ is the Digamma function.

To conclude, as previously mentioned, we point out that the above expressions are applicable for the improper density $\pi_0(\theta) = \frac{1}{\theta}$ on $(0, \infty)$ by setting $a = b = 0$. Related minimax properties are investigated in Sect. 4.2

Example 6 (An exponential location model) Consider $X_1, \ldots, X_n$ i.i.d. from an exponential distribution with location parameter θ and density $e^{-(t-\theta)}\mathbb{I}_{(\theta,\infty)}(t)$ (fixing the scale without loss of generality) with a Gamma prior $\theta \sim \mathrm{G}(a, b)$, $a > 0$ and $b = n$. This yields a posterior density of the form $\theta|x \sim \frac{1}{\sigma(x)}f(\frac{\theta}{\sigma(x)})$ with $\sigma(x) = x_{(1)} = \min\{x_1, \ldots, x_n\}$, and $f(u) = au^{a-1}\mathbb{I}_{(0,1)}(u)$, i.e., with the density of $U \sim \mathrm{Beta}(a, 1)$. Theorem 2 thus applies for any first-stage scale invariant and second-stage losses.

As an illustration, consider the entropy-type loss $L(\theta, \hat{\theta}) = \frac{\theta}{\hat{\theta}} - \log(\frac{\theta}{\hat{\theta}}) - 1$, yielding $\hat{\theta}_\pi(x) = \frac{a}{a+1}x_{(1)}$ and loss $L = L(\theta, \hat{\theta}_\pi(x))$ distributed under the posterior as $\frac{a+1}{a}U - \log(U) - \log(1 + \frac{1}{a}) - 1$, for $U \sim \mathrm{Beta}(a, 1)$ which is indeed free of x. Finally, for squared error second-stage loss, the Bayes estimator of L is given by

$$
\hat{L}_\pi(x) = \mathbb{E}(L|x) = \frac{1}{a} - \log\left(1 + \frac{1}{a} \right),
$$

since $\mathbb{E}(U) = \frac{a}{a+1}$ and $\mathbb{E}(\log U) = -\frac{1}{a}$.

Example 7 (Multivariate normal model with unknown covariance) Based on $X = (X_1, \ldots, X_n)^\top$ with $X_i \sim N_d(\mu, \sigma^2 I_d)$ independently distributed components, setting $\theta = (\theta_1, \theta_2) = (\mu, \sigma)$, consider estimating θ_1 under location scale invariant loss $L(\theta, \hat{\theta}_1) = \rho(\frac{\hat{\theta}_1 - \theta_1}{\theta_2})$, such as the typical case $\rho(t) = \|t\|^2$. For this set-up, a sufficient statistic is given by $(\bar{X}, S)$ with $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S = \sum_{i=1}^n \|X_i - \bar{X}\|^2$. Furthermore, $\bar{X}$ and S are independently distributed as $\bar{X}|\theta \sim N_d(\theta_1, (\sigma^2/n)I_d)$ and $S|\theta \sim \mathrm{G}(k/2, 1/2\sigma^2)$ with $k = (n-1)d$.

Now consider a normal-gamma conjugate prior distribution for θ such that

$$
\theta_1|\theta_2 \sim N_d\left(\xi, \frac{\theta_2^2}{c}I_d \right) \text{ with } \frac{1}{\theta_2^2} \sim \mathrm{G}(a, b),
$$

denoted $\theta \sim \mathrm{NG}(\xi, c, a, b)$, with hyperparameters $\xi \in \mathbb{R}^d, a, b, c > 0$. Calculations lead to the posterior density

$$\theta | x \sim NG\big(\xi(x), c + n, a(x), b(x)\big),$$

with $\xi(x) = \frac{n\bar{x}+c\xi}{n+c}, a(x) = a + \frac{d+k}{2}$, and $b(x) = \frac{s+2b+\frac{nc}{n+c}\|\bar{x}-\xi\|^2}{2}$. The corresponding posterior density of θ can be seen to match form (17) with $\xi(x)$ as given, $\sigma(x) = \sqrt{b(x)}$, and f the joint density of (V_1, V_2) where V_1 and V_2 are independently distributed as $V_1 \sim N_d(0, \frac{1}{c+n}I_d)$ and $V_2^{-2} \sim \mathrm{G}(a + \frac{d+k}{2}, 1)$. The last representation is obtained by the transformation $(\theta_1, \theta_2) \rightarrow (V_1 = \frac{\theta_1 - \xi(x)}{\theta_2}, V_2 = \frac{\theta_2}{\sigma(x)})$ under the posterior distribution.

Theorem 3 thus applies for any first-stage location-scale invariant and second-stage losses. For instance, the familiar weighted squared error loss $L(\theta, \hat{\theta}_1) = \frac{\|\hat{\theta}_1 - \theta_1\|^2}{\theta_2^2}$ leads to $\hat{\theta}_{1,\pi}(x) = \xi(x)$, [1] and loss $L = \frac{\|\theta_1 - \xi(x)\|^2}{\theta_2^2}$ whose posterior distribution (i.e., $\|V_1\|^2$ with V_1 as above) is that of a $\frac{1}{n+c}\chi_d^2(0)$ distribution.

Remark 4 Further potential applications of Theorems 1 and 3 may arise when the posterior distribution can be well approximated by a multivariate normal distribution. Such a situation occurs with the Bernstein-von Mises theorem and the convergence (under regularity conditions; e.g., (DasGupta, 2008) for an exposition) of $\sqrt{n}(\theta - \hat{\theta}_{mle})|x$ to a $N_d(0, I_{\theta_0})$, I_{θ_0} being the Fisher information matrix at the true parameter θ_0, and $\hat{\theta}_{mle}$ the maximum likelihood estimator.

4 Minimax findings for a given loss

In this section, we present different scenarios with loss estimators that are minimax and therefore a benchmark against which we can assess other loss estimators. The results are subdivided into two parts: (i) multivariate normal models with independent components and common variance, with or without a known variance; and (ii) univariate gamma models. Throughout, the first-stage estimator and associated loss are given, and the task consists in estimating the loss. For normal models, a quite general class of first-stage losses which are functions of squared error is considered with various second-stage losses of the form $\rho(\frac{\hat{L}}{L})$, while the analysis for the gamma distribution involves first-stage entropy loss and second-stage squared error loss. The theoretical results are accompanied by observations and illustrations.

---

[1] This can be seen as follows.

$$\hat{\theta}_{1,\pi}(x) = \frac{\mathbb{E}\left(\frac{\theta_1}{\theta_2^2}|x\right)}{\mathbb{E}\left(\frac{1}{\theta_2^2}|x\right)} = \frac{\mathbb{E}^{\theta_2|x}\left(\mathbb{E}\left(\frac{\theta_1}{\theta_2^2}|x, \theta_2\right)\right)}{\mathbb{E}\left(\frac{1}{\theta_2^2}|x\right)} = \frac{\mathbb{E}\left(\frac{\xi(x)}{\theta_2^2}|x\right)}{\mathbb{E}\left(\frac{1}{\theta_2^2}|x\right)} = \xi(x). \tag{19}$$

4.1 Normal models

We first study models $X \sim N_d(\theta, \sigma^2 I_d)$ with known σ^2 before addressing the unknown σ^2 case with an i.i.d. sample. We consider first-stage losses of the form $\beta\left(\frac{\|\hat{\theta}-\theta\|^2}{\sigma^2}\right)$ for estimating θ, with $\beta(\cdot)$ absolutely continuous and strictly increasing on $[0, \infty)$, and the loss

$$L = \beta\left(\frac{\|X - \theta\|^2}{\sigma^2}\right) \tag{20}$$

incurred by estimator $\hat{\theta}_0(X) = X$. In this set-up, the first-stage frequentist risk is given by $R(\theta, \hat{\theta}_0) = \mathbb{E}_\theta\left\{\beta\left(\frac{\|X-\theta\|^2}{\sigma^2}\right)\right\} = \mathbb{E}\{\beta(Z)\}$ with $Z \sim \chi_d^2$, and we assume that it is finite. In the identity case $\beta(t) = t$, the estimator $\hat{\theta}_0(X)$ is best equivariant, minimax and (generalized) Bayes with respect to the uniform prior density $\pi_0(\theta) = 1$ (e.g., (Fourdrinier et al., 2018)). These properties also hold more general β, and even for $X \sim f(\|x - \theta\|^2)$ with decreasing f (e.g., (Kubokawa et al., 2015)).

As a second-stage loss, consider the entropy-type loss

$$W(L, \hat{L}) = \rho_m\left(\frac{\hat{L}}{L}\right) \text{ with } \rho_m(t) = t^m - m\log(t) - 1, \ m \neq 0, \tag{21}$$

and the Bayes estimator $\hat{L}_{\pi_0}(X)$ of loss L with respect to prior density π_0. Since the posterior distribution $\frac{\|x-\theta\|^2}{\sigma^2}|x$ is χ_d^2 (see Example 3) independently of x, a direct minimization of the expected posterior loss tells us that

$$
\begin{aligned}
\hat{L}_{\pi_0}(x) &= \operatorname{argmin}_{\hat{L}}\left\{\mathbb{E}\left(\rho_m\left(\frac{\hat{L}}{L}\right)|x\right)\right\} \\
&= \operatorname{argmin}_{\hat{L}}\left\{\hat{L}^m \mathbb{E}(L^{-m}|x) - m\log\hat{L} + m\mathbb{E}(\log L|x) - 1\right\} \\
&= \left\{\mathbb{E}(L^{-m}|x)\right\}^{-1/m} \\
&= \left\{\mathbb{E}\left((\beta(Z))^{-m}\right)\right\}^{-1/m},
\end{aligned} \tag{22}
$$

for all $m \neq 0$, and that

$$
\begin{aligned}
\inf_{\hat{L}}\left\{\mathbb{E}\left(\rho_m\left(\frac{\hat{L}}{L}\right)|x\right)\right\} &= \mathbb{E}\left(\rho_m\left(\frac{\hat{L}_{\pi_0}}{L}\right)|x\right) \\
&= m\mathbb{E}\left(\log\beta(Z)\right) + \log\mathbb{E}\left((\beta(Z))^{-m}\right) \\
&= \bar{R} \ (say), \tag{23}
\end{aligned}
$$

as long as these expectations exist. $\hat{L}_{\pi_0}$ takes values on $(0, \infty)$ and a specific example is provided after the proof. Also observe that $\hat{L}_{\pi_0}$ has constant risk $R(\theta, \hat{L}_{\pi_0}) = \bar{R}$. This also can be seen directly by the equivalence of the frequentist and posterior

distributions of $\frac{\|X-\theta\|^2}{\sigma^2}$ which tells us that $\frac{\hat{L}_{\pi_0}}{L}|\theta =^d \frac{\hat{L}_{\pi_0}}{L}|x$ for all θ, x and therefore $R(\theta, \hat{L}_{\pi_0}) = \mathbb{E}\left(\rho_m(\frac{\hat{L}_{\pi_0}}{L})|\theta\right) = \mathbb{E}\left(\rho_m(\frac{\hat{L}_{\pi_0}}{L})|x\right) = \bar{R}.$

Theorem 4 *Under the above set-up and assumptions, for estimating the first-stage loss L in (20) under second-stage loss (21), the minimax estimator and risk are given by $\hat{L}_{\pi_0}(X)$ and $\bar{R}$, respectively.*

Proof Since $\hat{L}_{\pi_0}(X)$ has constant risk $\bar{R} = R(\theta, \hat{L}_{\pi_0})$, it suffices to show that $\hat{L}_{\pi_0}(X)$ is an extended Bayes estimator with respect to the sequence of priors $\pi_n \sim N_d(0, n\sigma^2 I_d), n \geq 1$, which is to show that

$$\lim_{n\to\infty} r_n = \bar{R}, \tag{24}$$

with r_n the integrated Bayes risk associated with π_n. For prior π_n, we have $\theta|x \sim N_d(\frac{n}{n+1}x, \frac{n\sigma^2}{n+1}I_d)$ implying that $\frac{\|X-\theta\|^2}{\sigma^2}|x \sim \frac{n}{n+1}\chi_d^2(\frac{y}{n})$, with $y = \frac{\|x\|^2}{(n+1)\sigma^2}$. Now, set Z_n such that $Z_n|y \sim \frac{n}{n+1}\chi_d^2(\frac{y}{n})$ and y as above. Then as earlier, the Bayesian optimization problem for π_n results in

$$\hat{L}_{\pi_n}(x) = \left\{\mathbb{E}\left((\beta(\frac{\|x-\theta\|^2}{\sigma^2}))^{-m}|x\right)\right\}^{-1/m} = \left\{\mathbb{E}\left((\beta(Z_n))^{-m}|y\right)\right\}^{-1/m}$$

and

$$\inf_{\hat{L}} \left\{\mathbb{E}_n\left(\rho_m(\frac{\hat{L}}{L})|x\right)\right\} = \mathbb{E}_n\left(\rho_m(\frac{\hat{L}_{\pi_n}}{L})|x\right)$$
$$= m\mathbb{E}\left(\log \beta(Z_n)|y\right) + \log \mathbb{E}\left((\beta(Z_n))^{-m}|y\right)$$
$$= u_n(y) \text{ say.}$$

From the above, we have

$$r_n = \mathbb{E}(u_n(Y)) \text{ with } Y = \frac{\|X\|^2}{(n+1)\sigma^2}. \tag{25}$$

Now observe that the marginal distribution of X is $N_d(0, (n+1)\sigma^2 I_d)$ implying that the marginal distribution of Y is χ_d^2 free of n. Finally, by the dominated convergence theorem with $|u_n(y)| \leq v_n(y)$, $v_n(y) = \mathbb{E}_n\left(\rho_m(\frac{\hat{L}_{\pi_0}}{L})|x\right)$ for all $n \geq 1$ and $y > 0$, $\mathbb{E}(v_n(Y)) = \mathbb{E}_n\mathbb{E}\left(\rho_m(\frac{\hat{L}_{\pi_0}}{L})|\theta\right) = \bar{R}$ (independently of n), and an application of Lemma 6, which is stated and proven in the Appendix, we infer that

$$\lim_{n\to\infty} r_n = \mathbb{E}^Y \left\{\lim_{n\to\infty} m\mathbb{E}\left(\log \beta(Z_n)|Y\right) + \lim_{n\to\infty} \log \mathbb{E}\left((\beta(Z_n))^{-m}|Y\right)\right\}$$
$$= \mathbb{E}^Y \left\{m\mathbb{E}\left(\log \beta(Z)\right) + \log \mathbb{E}\left((\beta(Z))^{-m}\right)\right\}$$
$$= m\mathbb{E}\left(\log \beta(Z)\right) + \log \mathbb{E}\left((\beta(Z))^{-m}\right)$$

which is (24) and completes the proof. $\qquad\square$

Table 2 Bayesian Estimators, Bayes Risks and Minimax Risks

$\rho_j(\frac{\hat{L}}{L})$	$\hat{L}_{\pi_0,j}(X)$	$r_{n,j} = \mathbb{E}(g_j(Y)), Y \sim \chi_d^2$	Minimax risk
$\left((\frac{\hat{L}}{L})^m - 1\right)^2$	$\left\{\frac{\mathbb{E}(\beta^{-m}(Z))}{\mathbb{E}(\beta^{-2m}(Z))}\right\}^{1/m}$	$1 - \frac{\mathbb{E}^2(\beta^{-m}(Z_n)|Y)}{\mathbb{E}(\beta^{-2m}(Z_n)|Y)}$	$1 - \frac{\mathbb{E}^2(\beta^{-m}(Z))}{\mathbb{E}(\beta^{-2m}(Z))}$
$\frac{1}{2}(\frac{\hat{L}}{L} + \frac{L}{\hat{L}} - 2)$	$\sqrt{\frac{\mathbb{E}(\beta(Z))}{\mathbb{E}(\frac{1}{\beta(Z)})}}$	$\sqrt{\mathbb{E}(\beta(Z_n)|Y)\mathbb{E}(\frac{1}{\beta(Z_n)}|Y)} - 1$	$\sqrt{\mathbb{E}(\beta(Z))\mathbb{E}(\frac{1}{\beta(Z)})} - 1$
$\left(\log(\frac{\hat{L}}{L})\right)^2$	$e^{\mathbb{E}(\log\beta(Z))}$	$\mathbb{V}(\log\beta(Z_n)|Y)$	$\mathbb{V}(\log\beta(Z))$

Theorem 4 applies quite generally for various choices of β and loss ρ_m. and the approach is unified. A particular interesting case is given by first-stage L^q losses with $\beta(t) = t^q, q > 0$. Calculations are easily carried out with the moments of $Z \sim \chi_d^2$ yielding the minimax loss estimator $\hat{L}_{\pi_0}(X) = 2^q \left(\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}-mq)}\right)^{1/m}$ of $L = \frac{\|x-\theta\|^{2q}}{\sigma^{2q}}$ for $m < d/2q$.

An analogous approach establishes the minimaxity of the generalized Bayes loss estimator $\hat{L}_{\pi_0}$ for various other interesting loss functions of the form $\rho(\frac{\hat{L}}{L})$. We summarize such findings as follows.

Theorem 5 *Consider the set-up of Theorem 4 with its corresponding assumptions, and the problem of estimating the first-stage loss (20) under second-stage loss $\rho_j(\frac{\hat{L}}{L})$, $j = A, B, C$ where $\rho_A(t) = (t^m - 1)^2, m \neq 0, \rho_B(t) = \frac{1}{2}(t + \frac{1}{t} - 2)$, and $\rho_C(t) = (\log t)^2$, then assuming existence and finite risk, the generalized Bayes estimator $\hat{L}_{\pi_0,j}(X)$, $j = A, B, C$, with respect to the uniform prior density is minimax, its frequentist risk is constant and matches the minimax risk.*

Proof In each of the three cases, a proof is quite analogous to that of Theorem 4 with estimators $\hat{L}_{\pi_0,j}(X)$, the integrated Bayes risk $r_{n,j} = \mathbb{E}(g_j(Y)), n \geq 1$ and the minimax risk varying for $j = A, B, C$ as presented in Table 2 with $Z_n|Y \sim \frac{n}{n+1}\chi_d^2(\frac{Y}{n})$, $Y \sim \chi_d^2, Z \sim \chi_d^2$. □

Remark 5 Some remarks:

(I) For squared error second-stage loss $(\hat{L} - L)^2$, the generalized Bayes estimator $\hat{L}_{\pi_0}$ of $L = \beta\left(\frac{\|X-\theta\|^2}{\sigma^2}\right)$ with respect to the uniform prior density π_0 is given by $\mathbb{E}(L|x) = \mathbb{E}(\beta(Z))$ with $Z \sim \chi_d^2$ (as for entropy loss ρ_{-1}), assuming finite $\mathbb{E}(\beta^2(Z))$. The methodology of Theorems 4 and 5 can be applied to establish the minimaxity of $\hat{L}_{\pi_0}$. This represents an extension of the identity case $\beta(t) = t$ proven by Johnstone (1988).

(II) Johnstone established in the identity case and for the squared error second-stage loss the inadmissibility of $\hat{L}_{\pi_0}(X) = d$ for $d \geq 5$ by producing estimators $\hat{L}$ that dominate $\hat{L}_{\pi_0}$. These estimators $\hat{L}$, and others appearing later in the literature, are "shrinkers" exploiting a potential defect of $\hat{L}_{\pi_0}$. In comparison, it can be shown using various applications of Jensen's inequality and the covariance inequality

that the Bayes estimators $\hat{L}_{\pi_0}$ in Theorem 5 for ρ_B, ρ_C, and for ρ_A with $m > 0$, as well as those of Theorem 4 for $m > -1$ are shrinkers in the sense that $\hat{L}_{\pi_0}(X) < \mathbb{E}\big(\beta(Z)\big)$. In contrast, they are expanders for $m < -1$ for both ρ_m and ρ_A. Such comparisons apply as well beyond the set-up here, for other models and priors, namely for Example 3's general $\hat{L}_\pi$ expressions, and in comparison to the benchmark posterior expectation estimator $\mathbb{E}(L|X)$ (see Appendix). Stronger properties can undoubtedly be established in specific cases, such as the identity case seen in Example 3. Finally, we point out for a given loss of Theorem 4, or Theorem 5, that the benchmark unbiased procedure $\hat{L}_0(X) = \mathbb{E}\big(\beta(Z)\big)$ is dominated in terms of frequentist risk by $\hat{L}_{\pi_0}(X)$ unless these two estimators coincide (e.g., ρ_{-1}).

(III) The considerations above also informs us on a "conservativeness" criterion for selecting a loss estimator which stipulates that

$$\mathbb{E}_\theta \hat{L}(X) \geq \mathbb{E}_\theta L(\theta, \hat{\gamma}(X)) \text{ for all } \theta, \tag{26}$$

(equality being (2)), put forth by Brown (1978), Lu and Berger (1989), and others. In our context, such an "expander" property does not follow in general, and rather is inherited (or disinherited) by the choice of the second-stage loss. The adherence to (26) is rather involved in general, but it thus can be controlled in this Example by the choice of the second-stage loss.

4.1.1 Unknown σ^2

For the unknown σ^2 case, a familiar argument (e.g., (Lehmann & Casella, 1998)) coupled with properties of $\hat{L}_{\pi_0}(X)$ for the known σ^2 case, leads to minimax findings via the following lemma.

Lemma 3 *For $X = (X_1, \ldots, X_n)^\top$ with independently distributed components $X_i \sim N_d(\mu, \sigma^2 I_d)$, $\theta = (\mu, \sigma^2)$, consider estimating the loss $L = \beta(\frac{\|\bar{x}-\mu\|^2}{\sigma^2/n})$ incurred by $\hat{\gamma}(X) = \bar{X}$ for estimating $\gamma(\theta) = \mu$, with $\beta(\cdot)$ absolutely continuous and strictly increasing as in Sect. 4.1. Suppose that $\hat{L}_0(X)$ is under second-stage loss $W(L, \hat{L})$ free of σ^2, minimax, and with constant risk $\bar{R} = \mathbb{E}_\theta\big\{W(L, \hat{L}_0(X))\big\}$ for estimating L regardless of σ^2. Then, $\hat{L}_0(X)$ remains minimax for unknown σ^2 with minimax risk $\bar{R}$.*

Proof Suppose, in order to establish a contradiction, that there exists another estimator $\hat{L}(X)$ such that

$$\sup_\theta \mathbb{E}_\theta\big\{W(L, \hat{L}(X))\big\} < \sup_\theta \mathbb{E}_\theta\big\{W(L, \hat{L}_0(X))\big\} = \bar{R}.$$

Then, for fixed $\sigma^2 = \sigma_0^2$, we would have

$$\sup_{\theta=(\mu,\sigma_0^2)} \mathbb{E}_\theta\big\{W(L, \hat{L}(X))\big\} < \sup_\theta \mathbb{E}_\theta\big\{W(L, \hat{L}_0(X))\big\} < \bar{R},$$

which would be not possible given the assumed minimax property of $\hat{L}_0(X)$ for $\sigma^2 = \sigma_0^2$. $\qquad\square$

The above coupled with the results of the previous section leads to the following.

Corollary 1 *In the set-up of Lemma 3 with second-stage loss (21), the loss estimator $\hat{L}_{\pi_0}$ given in () is minimax for estimating $L = \beta(\frac{\|\bar{x}-\mu\|^2}{\sigma^2/n})$. Furthermore, minimaxity is also achieved by the generalized Bayes estimators $\hat{L}_{\pi_0,j}(X)$ of Theorem 5 for the corresponding losses $\rho_j(\frac{\hat{L}}{L})$, $j = A, B, C$.*

Proof The results are deduced immediately as consequences of Lemma 3, Theorems 4, and 5. $\qquad\square$

4.2 Gamma models

We revisit here the Gamma model $X|\theta \sim G(\alpha, \theta)$ of Example 5 with first-stage entropy-type loss $\rho_m(\frac{\hat{\theta}}{\theta})$ for estimating θ and with $m < \alpha/2$. We consider the loss associated with $\hat{\theta}_{\pi_0}(X) = \frac{k}{X}$, $k = \{\frac{\Gamma(\alpha)}{\Gamma(\alpha-m)}\}^{1/m}$, which as an estimator of θ, is generalized Bayes for the improper density $\pi_0(\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\infty)}(\theta)$, as well as minimax with constant risk.

With the above first-stage estimator given, the task becomes to estimate

$$L = \left(\frac{k}{\theta x}\right)^m - m\log\left(\frac{k}{\theta x}\right) - 1, \qquad (27)$$

and we investigate second-stage squared error loss $(\hat{L} - L)^2$. We establish below the minimaxity of the Bayes estimator

$$\hat{L}_{\pi_0}(X) = m\Psi(\alpha) + \log\left(\frac{\Gamma(\alpha - m)}{\Gamma(\alpha)}\right), \qquad (28)$$

given in Example 5 for $a = b = 0$. We will require the following Gamma distribution properties, which are derivable in a straightforward manner, and related frequentist risk expression for $\hat{L}_{\pi_0}$.

Lemma 4 *Let $W \sim G(\xi, \beta)$ and $h > -\xi/2$. We have:*
(a) $\mathbb{V}(W^h) = \beta^{-2h}\left\{\frac{\Gamma(\xi+2h)}{\Gamma(\xi)} - \left(\frac{\Gamma(\xi+h)}{\Gamma(\xi)}\right)^2\right\}$,
(b) $\mathbb{V}(\log W) = \Psi'(\xi)$,
(c) $Cov(W^h, \log W) = \frac{1}{\beta^h}\frac{\Gamma(h+\xi)}{\Gamma(\xi)}\{\Psi(\xi + h) - \Psi(\xi)\}$,

Γ and Ψ being the Gamma and Digamma functions.

Lemma 5 *The estimator $\hat{L}_{\pi_0}(X)$ of L has constant in θ frequentist risk given by*

$$R(\theta, \hat{L}_{\pi_0}) = k^{2m}\left\{\frac{\Gamma(\alpha - 2m)}{\Gamma(\alpha)} - \left(\frac{\Gamma(\alpha - m)}{\Gamma(\alpha)}\right)^2\right\} + m^2\Psi'(\alpha)+$$

$$+ 2mk^m \frac{\Gamma(\alpha - m)}{\Gamma(\alpha)} \{\Psi(\alpha - m) - \Psi(\alpha)\}. \tag{29}$$

Proof Let $Z \sim G(\alpha, k)$. For the non-informative prior distribution here, one can verify as in Theorem 3 that $\frac{X\theta}{k}|\theta =^d \frac{X\theta}{k}|x =^d Z$ for all θ, x, i.e., the frequentist and posterior distributions coincide. This implies that $L|\theta =^d L|x$ for all θ, x, so that $\mathbb{E}_\theta(L) = \mathbb{E}(L|x) = \hat{L}_{\pi_0}(x)$. Now, since the second-stage loss is squared error, the frequentist risk is equal to

$$
\begin{aligned}
R(L, \hat{L}) = \mathbb{E}_\theta\big((L - \mathbb{E}(L|X))^2\big) &= \mathbb{V}(L|\theta) \\
&= \mathbb{V}\big(Z^{-m} + m \log Z\big) \\
&= \mathbb{V}(Z^{-m}) + m^2 \mathbb{V}(\log Z) + 2m\mathrm{Cov}(Z^{-m}, \log Z).
\end{aligned}
$$

The result then follows by applying Lemma 4 to the above for $\xi = \alpha, h = -m$. □

We now proceed with the main result of this section.

Theorem 6 *For $X \sim G(\alpha, \theta)$ with known α ($\alpha > 2m$) and unknown $\theta \in \mathbb{R}_+$, first-stage loss $\rho_m(\frac{\hat{\theta}}{\theta})$, and second-stage squared error loss, the generalized Bayes estimator $\hat{L}_{\pi_0}(X)$ given in (28) of L is minimax with minimax risk given by (29).*

Proof We show that $\hat{L}_{\pi_0}(X)$ is an extended Bayes estimator of L with respect to the sequence of prior densities $\pi_n := G(a_n, b_n)$, with $a_n = b_n = \frac{1}{n}, n \geq 1$. Since the risk $R(\theta, \hat{L}_{\pi_0}) = \bar{R}$ is constant in θ, establishing (24) with r_n the integrated Bayes risk with respect to π_n will suffice to prove the above.

We have for a given n and π_n,

$$r_n = \mathbb{E}_n^X \left\{ \mathbb{E}_n\big(L - \hat{L}_{\pi_n}(X)\big)^2 \big| X \right\} = \mathbb{E}_n^X\big(\mathbb{V}_n(L|X)\big),$$

where $\hat{L}_{\pi_n}(X) = \mathbb{E}_n(L|X)$ is the Bayes estimator of L, $\mathbb{V}_n(L|X)$ is the posterior variance of L, and the expectation $\mathbb{E}_n^X$ is taken with respect to the marginal distribution of X.

Under π_n, we have $\theta|x \sim G\big(\alpha + a_n, (x + b_n)\big)$ so that $\frac{\theta x}{k}|x \sim G(\alpha + a_n, \frac{k(x + b_n)}{x})$. Setting $Y = Y(X) = \frac{k(X + b_n)}{X}$, we can write

$$
\begin{aligned}
\mathbb{V}_n(L|X) &= \mathbb{V}_n\big((\tfrac{k}{\theta X})^m - m \log(\tfrac{k}{\theta X})\big| X\big) \\
&= \mathbb{V}_n\big(W^{-m} + m \log W \big| Y\big),
\end{aligned}
$$

with $W|Y \sim G(\alpha + a_n, Y)$. Expanding the above variance as in Lemma 5 and again making use of Lemma 4, one obtains

$$\mathbb{V}_n(L|X) = C_n\{Y(X)\}^{2m} + m^2\Psi'(a_n + \alpha) + 2m\{Y(X)\}^m D_n, \tag{30}$$

with

$$C_n = \frac{\Gamma(a_n + \alpha - 2m)}{\Gamma(a_n + \alpha)} - \frac{\Gamma^2(a_n + \alpha - m)}{\Gamma^2(a_n + \alpha)},$$

and

$$D_n = \frac{\Gamma(a_n + \alpha - m)}{\Gamma(a_n + \alpha)}\{\Psi(a_n + \alpha - m) - \Psi(a_n + \alpha)\}.$$

Now, since $X \sim B2(\alpha, a_n, b_n)$ under π_n, i.e., a Beta type II distribution as in Definition 1, we have by virtue of Lemma 1

$$\mathbb{E}_n(Y(X)^h) = \mathbb{E}\left(\left(\frac{X}{k(X + b_n)}\right)^{-h}\right) = k^h \frac{(\alpha)_{-h}}{(\alpha + a_n)_{-h}} \text{ for } h < \alpha,$$

so that $\lim_{n \to \infty} \mathbb{E}_n((Y(X))^h) = k^h$ for $h = m$ and $h = 2m$. Finally with the above and (30), we obtain directly

$$\lim_{n \to \infty} r_n = \lim_{n \to \infty} \mathbb{E}_n^X\{\mathbb{V}_n(L|X)\} = R(\theta, \hat{L}_{\pi_0}) = \bar{R},$$

as given in (29), completing the proof. □

5 Minimaxity for Rukhin-type losses

Whereas the minimax findings of the previous sections apply to decisions problems that are sequential in nature, i.e., the decision of interest which is that of estimating a loss $L = L(\theta, \hat{\gamma}(x))$ is assessed for optimality after having observed the data x, Rukhin's loss in (3) or (4) applies to the problem of estimating $(\gamma(\theta), L)$ simultaneously. Whereas (Rukhin, 1988a, b) investigated questions of admissibility of pairs $(\hat{\gamma}, \hat{L})$, our investigation here pertains to minimaxity. An estimator $(\hat{\gamma}_m, \hat{L}_m)$ of $(\gamma(\theta), L)$ is defined to be minimax for loss $W(\theta, \hat{\gamma}, \hat{L})$ if $\sup_\theta \{W(\theta, \hat{\gamma}_m, \hat{L}_m)\} \leq \sup_\theta \{W(\theta, \hat{\gamma}, \hat{L})\}$ for all $(\hat{\gamma}, \hat{L})$. Since we will investigate the behaviour of a sequence of Bayes estimators, we point out that a Bayes estimator $(\hat{\gamma}_\pi, \hat{L}_\pi)$ of (γ, L) under loss $W(\theta, \hat{\gamma}, \hat{L})$ and prior π is, whenever it exists, given by $\hat{L} = \mathbb{E}(L|x)$ and $\hat{\gamma}_\pi$ the Bayes point estimator of $\hat{\gamma}(\theta)$ under first-stage loss $L = L(\theta, \hat{\gamma}(x))$ (independently of the choice of h).

We capitalize on a combination of properties and findings of the previous sections to establish a minimax result which we frame as follows.

Theorem 7 *Consider a given model $X \sim f_\theta$ and loss $W = W(\theta, \hat{\gamma}, \hat{L})$ as in (4) for estimating (γ, L) with $\gamma = \gamma(\theta)$ and $L = L(\theta, \hat{\gamma})$. Suppose there exist an estimator $(\hat{\gamma}_0, \hat{L}_0)$ and a sequence of proper densities $\{\pi_n; n \geq 1\}$ such that: (i) $\hat{\gamma}_0(X)$ is for first-stage loss L an extended Bayes estimator of γ with constant risk in θ; (ii) the*

Bayes estimator $\hat{L}_{\pi_n}(x)$ is for $n \geq 1$ constant as a function of x; and **(iii)** *the estimator $\hat{L}_0(x)$ is the limit of $\hat{L}_{\pi_n}(x)$ as $n \to \infty$. Then $(\hat{\gamma}_0, \hat{L}_0)$ is minimax.*

Proof Denote $R_W(\theta, (\hat{\gamma}, \hat{L}))$ as the frequentist risk of estimator $(\hat{\gamma}, \hat{L})$ under loss $W(\theta, \hat{\gamma}, \hat{L})$; $R(\theta, \hat{\gamma})$ as the first-stage risk of $\hat{\gamma}$; $\bar{R}$ as the constant first-stage risk of $\hat{\gamma}_0$; r_n and r_n^W as the integrated Bayes risks with respect to π_n associated with the first-stage loss L and global loss W, respectively. As well, denote the constant values of $\hat{L}_{\pi_n}(x)$ and $\hat{L}_0(x)$ as c_n and $c = \lim_{n \to \infty} c_n$. We have

$$
\begin{aligned}
R_W(\theta, (\hat{\gamma}_0, \hat{L}_0)) &= \mathbb{E}_\theta(W(\theta, \hat{\gamma}_0, \hat{L}_0)) \\
&= h'(c)\bar{R} - ch'(c) + h(c) \\
&= \bar{R}_W,
\end{aligned}
$$

which is constant as a function of θ. To establish the result, it will suffice to show that

$$
\lim_{n \to \infty} r_n^W = \bar{R}_W \tag{31}
$$

which implies that the pair $(\hat{\gamma}_0, \hat{L}_0)$ is an extended Bayes equalizer rule with respect to the loss $W(\theta, \hat{\gamma}, \hat{L})$ which hence results in a minimax solution. As above, it is the case that

$$
R_W(\theta, (\hat{\gamma}_{\pi_n}, \hat{L}_{\pi_n})) = h'(c_n)R(\theta, \hat{\gamma}_{\pi_n}) - c_n h'(c_n) + h(c_n),
$$

which implies that

$$
r_n^W = h'(c_n)r_n - c_n h'(c_n) + h(c_n).
$$

Finally, condition (31) is verified with the above expressions since, by assumptions, $\hat{\gamma}_0$ is extended Bayes with $\lim_{n \to \infty} r_n = \bar{R}$ and $\lim_{n \to \infty} c_n = c$. $\qquad \square$

Observe that the result is quite general and the minimaxity holds irrespectively of the choice of h in loss function (4), as is the case for the determination of a Bayes estimator of $(\gamma(\theta), L)$. The above theorem paves the way for various applications which build on the results contained in the previous sections and we present as a series of examples. A critical property is the one where the Bayes estimators $\hat{L}_{\pi_n}(x)$ are free of x, situations that were expanded on in Sect. 3.

Example 8 (Normal model) Theorem 7 applies for $X \sim N_d(\theta, \sigma^2 I_d)$, $\gamma(\theta) = \theta$, simultaneous loss W as in (4) with first-stage squared error loss $L = \frac{\|\hat{\theta} - \theta\|^2}{\sigma^2}$, with the estimator $(\hat{\gamma}_0(X) = X, \hat{L}_0(X) = d)$ which is generalized Bayes for (θ, L) and the uniform prior density $\pi(\theta) = 1$. Indeed, with the prior sequence of densities $\theta \sim^{\pi_n} N_d(0, n\sigma^2)$, the minimaxity follows from Theorem 7 since: **(i)** $\hat{\gamma}_0(X)$ is extended Bayes relative to $\{\pi_n; n \geq 1\}$ with constant risk equal to d, **(ii)** $\hat{L}_{\pi_n}(x) = \frac{nd}{n+1}$ is constant as a function of x, and **(iii)** converges to $\hat{L}_0(x) = d$.

Example 9 (Normal model continued) The above example can be extended to the choice of $L = \beta\left(\frac{\|\hat{\theta}-\theta\|^2}{\sigma^2}\right)$ with β a continuous and strictly increasing function on $\mathbb{R}_+$. Let $Z \sim \chi_d^2(0)$. Then, the estimator $(\hat{\gamma}, \hat{L})$ with $\hat{\gamma}_0(x) = x$ and $\hat{L}_0(x) = \mathbb{E}(\beta(Z))$ (as long as the latter is finite) can be shown to be minimax for estimating (θ, L) under loss W. It is also generalized Bayes for the uniform prior density.

A justification of condition **(i)** of Theorem 7 is as follows. A result in Yadegari (2017) (Theorem 2.2, page 24), that applies when the posterior distribution of θ is normal, tells us that the first-stage Bayes estimator of θ under loss L and prior π_n is given by the posterior mean $\frac{nx}{n+1}$, independently of β, the posterior being $\theta|x \sim N_d\left(\frac{nx}{n+1}, \left(\frac{n\sigma^2}{n+1}\right)I_d\right)$ under prior π_n. It follows from this that the minimum expected posterior loss, equivalently $\hat{L}_{\pi_n}(x)$, is equal to $\mathbb{E}\left(\frac{n}{n+1}\beta(Z)\right)$ since $\left(\frac{\|\hat{\theta}-\theta\|^2}{\sigma^2}\right)|x \sim \frac{n}{n+1}\chi_d^2(0)$. Now, since this is free of x, one infers that the integrated Bayes risk r_n is equal to the minimum expected posterior loss and thus converges to $\mathbb{E}(\beta(Z))$ which can be seen as an application of Lemma 6 (for $y = 0$). Since this matches the constant risk of $\hat{\gamma}_0(X)$, we infer that the latter is also extended Bayes, whence condition **(i)** of Theorem 7. From the above, we infer have that **(ii)** $\hat{L}_{\pi_n}(x) = \mathbb{E}\left(\frac{n}{n+1}\beta(Z)\right)$ is free of x, and which **(iii)** converges to $\hat{L}_0(x)$, establishing the minimaxity.

Example 10 (Gamma model) For a Gamma model $X \sim G(\alpha, \theta)$ (i.e., Example 5, we apply Theorem 7 for estimating $\gamma(\theta) = \theta$ and $L(\theta, \hat{\theta}) = \left(\frac{\hat{\theta}}{\theta}\right)^m - m\log(\frac{\hat{\theta}}{\theta}) - 1$ simultaneously under loss W in (3), with $m < \alpha$. We show that the Bayes estimator $(\hat{\theta}_0, \hat{L}_0)$ of (θ, L) with respect to the improper prior density $\pi(\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\infty)}(\theta)$, given by $\hat{\theta}_0(X) = \left\{\frac{\Gamma(\alpha)}{\Gamma(\alpha-m)}\right\}^{1/m}\frac{1}{X}$ and $\hat{L}_0(X) = m\Psi(\alpha) + \log\left(\frac{\Gamma(\alpha-m)}{\Gamma(\alpha)}\right)$ is minimax. Indeed, with the sequence of prior $G(\frac{1}{n}, \frac{1}{n})$ densities π_n, the minimaxity follows since: **(i)** $\hat{\theta}_0(X)$ can be shown to be extended Bayes with constant risk given by $\hat{L}_0(X)$, **(ii)** the Bayes estimator $\hat{L}_{\pi_n}(x)$ of L is a constant given by (18) with $a = 1/n$, and **(iii)** converges to $\hat{L}_0$ as $n \to \infty$.

Theorem 7 also applies for other first-stage losses, such as the familiar scale invariant squared error loss $L(\theta, \hat{\theta}) = (\frac{\hat{\theta}}{\theta} - 1)^2$ with $\alpha > 2$. In this case, a minimax solution is $\hat{\theta}_0(X) = \frac{\alpha-2}{X}$ and $\hat{L}_0(X) = \frac{1}{\alpha-1}$, and the conditions of the theorem can be verified with the same prior sequence $\{\pi_n\}$ as above with $\hat{L}_{\pi_n}(X) = \frac{1}{\alpha-1+n^{-1}}$ computable from (8).

Example 11 (Poisson model) Theorem 7 leads to the following application for the Poisson models of Sect. 3.1.1 and 3.1.2. With the set-up of Sect. 3.1.2, for estimating (θ, L) under loss (3) with loss $L(\theta, \hat{\theta}) = \sum_{i=1}^d \frac{(\hat{\theta}_i - \theta_i)^2}{\theta_i}$, we infer that $(\hat{\theta}_0(X), \hat{L}_\pi(X)) = (X, d)$ is minimax by considering the sequence of priors π_n such that $S \sim G(a_n = d, b_n = \frac{1}{n})$. Indeed for such a sequence, we may show, namely by using Remark 3, that: **(i)** $\hat{\theta}_0(X)$ is extended Bayes with constant risk given by d, **(ii)** the Bayes estimator $\hat{L}_{\pi_n}(x)$ of L is a constant as a function of x given by $\frac{d}{1+\frac{1}{n}}$, and which **(iii)** converges to $\hat{L}_0$ as $n \to \infty$.

Example 12 (Negative binomial model) We consider the set-up of Sect. 3.1.3 with $X \sim \text{NB}(r, \theta)$ as in (14), and the problem of estimating (θ, L) for simultaneous loss W as in (3), with $L = L(\theta, \hat{\theta})$ the weighted squared error loss given in (16). Theorem 7 applies in establishing the minimaxity of $(\hat{\theta}_0, \hat{L}_0)$ with $\hat{\theta}_0(X) = \frac{rX}{r+1}$ and $\hat{L}_0(X) = \frac{1}{r+1}$. Indeed with the sequence of $B2(a_n, b_n, r)$ prior densities π_n for θ with $a_n = 1$ and $b_n = \frac{1}{n}$, it is the case that: **(i)** $\hat{\theta}_0(X)$ is extended Bayes with constant risk $\bar{R} = \frac{1}{r+1}$, **(ii)** $\hat{L}_{\pi_n}(x) = \frac{1}{r+1+n^{-1}}$ is free of x, and **(iii)** converges to $\hat{L}_0(x)$ as $n \to \infty$.

The results above paired with the particular features of the $(\hat{\theta}_0, \hat{L}_0)$ minimax solutions lead to further minimax estimators with the simple observation that $(\hat{\theta}_1, \hat{L}_0)$ dominates $(\hat{\theta}_0, \hat{L}_0)$ under loss (3) whenever $\hat{\theta}_1$ dominates $\hat{\theta}_0$ under first-stage loss $L(\theta, \hat{\theta})$, given that $\hat{L}_0(X)$ is a constant. We thus have the following implications for the multivariate normal and Poisson models, for $d \geq 3$ and $d \geq 2$ respectively, since there exist (many) dominating estimators $\hat{\theta}_1(X)$ of $\hat{\theta}_0(X) = X$. The same applies for the multivariate normal model with a loss function which is a concave function of squared error loss and $d \geq 4$ (see for instance, (Fourdrinier et al., 2018)).

Corollary 2 (a) *For the normal model context of Example 8 with $d \geq 3$, an estimator $(\hat{\theta}, \hat{L})$ is minimax for estimating (θ, L) whenever $\hat{\theta}(X)$ dominates $\hat{\theta}_0(X) = X$ under first-stage loss $\frac{\|\hat{\theta}-\theta\|^2}{\sigma^2}$;*

(b) *For the normal model context of Example 9 with $d \geq 4$ and concave β, an estimator $(\hat{\theta}, \hat{L})$ is minimax for estimating (θ, L) whenever $\hat{\theta}(X)$ dominates $\hat{\theta}_0(X) = X$ under first-stage loss $\beta\left(\frac{\|\hat{\theta}-\theta\|^2}{\sigma^2}\right)$;*

(c) *For the Poisson model context of Example 11 with $d \geq 2$, an estimator $(\hat{\theta}, \hat{L})$ is minimax for estimating (θ, L) whenever $\hat{\theta}(X)$ dominates $\hat{\theta}_0(X) = X$ under first-stage loss $\sum_{i=1}^{d} \frac{(\hat{\theta}_i - \theta_i)^2}{\theta_i}$.*

6 Concluding remarks

This paper brings into play original contributions and analyses for loss estimation that culminate with minimax findings for: **(i)** estimating a first-stage loss $L = L(\theta, \hat{\gamma})$, and for **(ii)** estimating jointly $(\gamma(\theta), L)$ under a Rukhin-type loss. Various models and choices of the first and second-stage losses were analysed. Our work also clarifies the structure of various Bayesian solutions, properties of which become critical for the minimax analyses.

All in all, the optimality properties obtained here serve as a guide on how one can sensibly report on an incurred loss in both situations **(i)** and **(ii)**. Notwithstanding existing results, related questions of admissibility questions remain unanswered, namely in the context of Example 1 for different second-stage losses where it would be interesting to revisit the effect of the dimension d in the d−variate normal case.

Appendix A

A.1

The following result was used in Sect. 4.1.

Lemma 6 *Let $Z_n \sim \frac{n}{n+1} \chi_d^2(\frac{y}{n})$ for $n \geq 1$, $d \geq 1$, and fixed $y \geq 0$, and let g be a positive valued function such that $\mathbb{E}(g(Z_1)) < \infty$. Then, we have $\lim_{n\to\infty} \mathbb{E}(g(Z_n)) = \mathbb{E}(g(Z))$ with $Z \sim \chi_d^2(0)$.*

Proof Denote h_n and f_n as the density of Z_n and $\frac{n+1}{n} Z_n \sim \chi_d^2(\frac{y}{n})$ respectively. We seek to apply the dominated convergence theorem and we make use of the $W \sim \chi_\nu^2(\lambda)$ density representation

$$\left(\frac{1}{2}\right)^{\nu/2} \frac{w^{\frac{\nu-2}{2}}}{\Gamma(\frac{\nu}{2})} \, {}_0F_1\left(-; \frac{\nu}{2}; \frac{\lambda w}{4}\right) e^{-(\lambda+w)/2},$$

for $w > 0$, $\nu > 0$, $\lambda \geq 0$ and where ${}_0F_1(-; b; t) = \sum_{k\geq 0} \frac{t^k}{(b)_k k!}$. From this, we have for all $n \geq 2$ and $z > 0$:

$$\begin{aligned}
g(z)h_n(z) &= g(z)f_n(\tfrac{(n+1)z}{n})\tfrac{n+1}{n} \\
&\leq \tfrac{3}{2} g(z) f_n(\tfrac{n+1}{n}z) \\
&= \tfrac{3}{2} g(z)\left(\tfrac{1}{2}\right)^{d/2} \frac{(\tfrac{(n+1)z}{n})^{\frac{d-2}{2}}}{\Gamma(\tfrac{d}{2})} \, {}_0F_1(-; \tfrac{d}{2}; \tfrac{n+1}{4n^2}zy)\, e^{-\frac{(y+(n+1)z)}{2n}} \\
&\leq \tfrac{3}{2} K e^{\frac{y}{2}} g(z) f_1(z),
\end{aligned}$$

with $K = \max\{1, (\frac{3}{2})^{\frac{d-2}{2}}\}$, where we have exploited the fact that ${}_0F_1(-; b; t)$ is for fixed $b > 0$ increasing in $t > 0$, as well as make use of the inequality $(\frac{n+1}{n})^{\frac{d-2}{2}} \leq K$ for all $d \geq 1$ and $n \geq 2$. The result then follows by dominated convergence. □

A.2

As a complement to Example 3, here are justifications to the effect that the constant loss estimate $\hat{L}_\pi$ decreases as a function of m for losses ρ_A and ρ_m and that the defined cut-off point $m_0(d)$ takes values between -1 and 0. We have

$$\log\left(\frac{\hat{L}_\pi(x)}{2\tau_0^2}\right) = f(d, m) = \frac{1}{m}\left\{\log\Gamma\left(\frac{d}{2}\right) - \log\Gamma\left(\frac{d}{2} - m\right)\right\},$$

for ρ_m, and $\log\left(\frac{\hat{L}_\pi(x)}{2\tau_0^2}\right) = f(d - 2m, m)$ for ρ_A. Now, observe that $f(d, m)$ increases in d, and decreases in m, the former being a consequence of the strict logconvexity of

the gamma function, and the latter since

$$\frac{\partial}{\partial m} f(d, m) = -\frac{1}{m^2} \left\{ \log \Gamma\left(\frac{d}{2}\right) - \log \Gamma\left(\frac{d}{2} - m\right) - m\Psi\left(\frac{d}{2} - m\right) \right\} < 0,$$

with the inequality due to the ordering $u'(a) < \frac{u(a+b)-u(a)}{b} < u'(a + b)$ for $a, b \in \mathbb{R}_+$ and strictly convex and differentiable functions $u(\cdot)$ on $\mathbb{R}_+$. The above tells us directly that $\hat{L}_\pi$ decreases in m for loss ρ_m, but it is also the case for loss ρ_A since $f(d - 2m_1, m_1) > f(d - 2m_2, m_1) > f(d - 2m_2, m_2)$ for $m_1 < m_2 < \frac{d}{4}$.

For the bounds on $m_0(d)$ which apply to loss ρ_A, it suffices to observe that $\hat{L}_\pi = \tau_0^2(d + 2)$ for $m = -1$, calculate the limiting value $\hat{L}_\pi = 2\tau_0^2 e^{\Psi(\frac{d}{2})}$ as $m \to 0$, and then infer that $\lim_{m\to 0} \hat{L}_\pi \leq \tau_0^2 d$ by virtue of the Digamma function inequality $\Psi(\alpha) < \log(\alpha)$ for $\alpha > 0$. More generally, one shows that $\lim_{m\to 0} \left(\frac{\mathbb{E}(L^{-m}|x)}{\mathbb{E}(L^{-2m}|x)}\right)^{1/m} = e^{\mathbb{E}(\log L|x)}$, so that the squared log error loss arises as the limiting loss $\left((\frac{\hat{L}}{L})^m - 1\right)^2$ when $m \to 0$.

A.3

Here are elements of justification for the stated properties of Remark 5. We make use of two inequalities, first Jensen's inequality for concave h, $\mathbb{E}(h(\beta(Z))) \leq h(\mathbb{E}(\beta(Z)))$ and a Covariance inequality for increasing f and decreasing g, $\mathbb{E}(f(\beta(Z))g(\beta(Z))) \leq \mathbb{E}(f(\beta(Z)))\mathbb{E}(g(\beta(Z)))$. The implications for losses ρ_m and ρ_C follow with Jensen's inequality using $h(t) = t^{-m}$ or $-t^{-m}$ depending on the value of m, and $h(t) = \log(t)$, respectively. The shrinkage for ρ_A with $m \in (0, 1)$ follows with the covariance inequality applied for $f(t) = t^m$ and $g(t) = t^{-2m}$, telling us that $\left(\hat{L}_{\pi_0}(x)\right)^m < \mathbb{E}(L^m|x)$, followed by Jensen's inequality applied to $h(t) = t^m$. The shrinkage that occurs for ρ_B follows from the use of the covariance inequality with $f(t) = t$ and $g(t) = t^{-1}$. There remains loss ρ_A, the Bayes estimator $\left\{\frac{\mathbb{E}(L^{-m}|x)}{\mathbb{E}(L^{-2m}|x)}\right\}^{1/m}$ and its properties of shrinkage for $m > 0$, and expansion for $m < -1$, in comparison to the benchmark estimator $\hat{L}_0(X) = \mathbb{E}(L|X)$. These properties follow directly from the following inequality, which is also of independent interest.

Lemma 7 *The following inequality holds for a positive random variable T:*

$$\frac{\mathbb{E}(T^{-m})}{\mathbb{E}(T^{-2m})} \leq (\mathbb{E}(T))^m \text{ for } m > 0 \text{ and } m \leq -1$$

assuming existence of the above expectations.

Proof For a positive real number N, we set $\lfloor N \rfloor$ and $\{N\}$ as integer and fractional parts defined here as $\lfloor N \rfloor = \sup\{j \in \mathbb{N} : j < N\}$ and $\{N\} = N - \lfloor N \rfloor$. The result has been previously shown for $m \in (0, 1)$. For $m \geq 1$, the result follows by applying

the covariance inequality $\lfloor m \rfloor + 1$ times as follows

$$\mathbb{E}(T^{-m}) \leq \mathbb{E}(T^{-m-1})\mathbb{E}(T) \leq \cdots \leq \mathbb{E}(T^{-m-\lfloor m \rfloor})\big(\mathbb{E}(T)\big)^{\lfloor m \rfloor}$$
$$\leq \mathbb{E}(T^{-m-\lfloor m \rfloor - \{m\}})\big(\mathbb{E}(T)\big)^{\lfloor m \rfloor + \{m\}}$$
$$= \mathbb{E}(T^{-2m})\big(\mathbb{E}(T)\big)^{m}.$$

Similarly for $m < -1$, the inequality follows as

$$\mathbb{E}(T^{-2m}) \geq \mathbb{E}(T^{-2m-1})\mathbb{E}(T) \geq \cdots \geq \mathbb{E}(T^{-2m-\lfloor -m \rfloor})\big(\mathbb{E}(T)\big)^{\lfloor -m \rfloor}$$
$$\geq \mathbb{E}(T^{-2m-\lfloor -m \rfloor - \{-m\}})\big(\mathbb{E}(T)\big)^{\lfloor -m \rfloor + \{-m\}}$$
$$= \mathbb{E}(T^{-m})\big(\mathbb{E}(T)\big)^{-m}.$$

$\square$

Declarations

References

Boisbunon, A., Canu, S., Fourdrinier, D., Strawderman, W. E., & Wells, M. T. (2014). Akaike's information criterion, C_p and estimators of loss for elliptically symmetric distributions. *International Statistical Review, 82*(3), 422–439.

Berger, J.O. (1985). The frequentist viewpoint and conditioning. In: LeCam, L.M., Olshen, R.A. (eds.) Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Berkeley, Calif., 1983), Wadsworth Statist./Probab. Ser., vol. 1, pp. 15–44 Wadsworth, Belmont, California.

Brown, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Annals of Statistics, 6*(1), 59–71.

Clevenson, M. L., & Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *Jour. Amer. Statist. Assoc., 70*, 698–705.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer, New York: Springer Texts. in Statistics.

Ferguson, T. (1968). *Mathematical Statistics: A Decision Theoretic Approach*. New York and London: Academic Press.

Fourdrinier, D., & Lepelletier, P. (2008). Estimating a general function of a quadratic function. *Annals of the Institute of Statistical Mathematics, 60*(1), 85–119.

Fourdrinier, D., & Strawderman, W. E. (2003). On Bayes and unbiased estimators of loss. *Annals of the Institute of Statistical Mathematics, 55*(4), 803–816.

Fourdrinier, D., Strawderman, W. E., & Wells, M. T. (2018). *Shrinkage Estimation*. New York: Springer.

Fourdrinier, D., & Wells, M. T. (1995). Estimation of a loss function for spherically symmetric distributions in the general linear model. *Annals of Statistics, 23*(2), 571–592.

Fourdrinier, D., & Wells, M. T. (1995). Loss estimation for spherically symmetric distributions. *Journal of Multivariate Analysis, 53*(2), 311–331.

Fourdrinier, D., & Wells, M. T. (2012). On improved loss estimation for shrinkage estimators. *Statistical Science, 27*(1), 61–81.

Goutis, C., & Casella, G. (1995). Frequentist post-data inference. *International Statistical Review, 63*(3), 325–344.

Johnstone, I. (1988). On inadmissibility of some unbiased estimates of loss. In: Gupta, S.S., Berger, J.O. (eds.) Statistical Decision Theory and Related Topics, IV, (West Lafayette, Ind., 1986). vol. 1, pp. 361–379. Springer, New York

Kubokawa, T., Marchand, É., & Strawderman, W. E. (2015). On improved shrinkage estimators under concave loss. *Statistics and Probability Letters, 96*, 241–246.

Lu, K. L., & Berger, J. O. (1989). Estimation of normal means: frequentist estimation of loss. *Annals of Statistics, 17*(2), 890–906.

Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation* (2nd ed.). New York: Springer.

Lele, C. (1992). Inadmissibility of loss estimators. *Statistics and Decisions, 10*(4), 309–322.

Lele, C. (1993). Admissibility results in loss estimation. *Annals of Statistics, 21*(1), 378–390.

Maruyama, Y. (1997). A new positive estimator of loss function. *Statistics and Probability Letters, 36*(3), 269–274.

Matsuda, T. (2024). Inadmissibility of the corrected Akaike information criterion. *Bernoulli, 30*(2), 1416–1440.

Mozgunov, P., Jaki, T., & Gasparini, M. (2019). Loss functions in restricted parameter spaces and their Bayesian applications. *Journal of Applied Statistics, 46*(13), 2314–2337.

Matsuda, T., & Strawderman, W. E. (2019). Improved loss estimation for a normal mean matrix. *Journal of Multivariate Analysis, 169*, 300–311.

Narayanan, R., & Wells, M. T. (2015). Improved loss estimation for the LASSO: a variable selection tool. *Sankhya B, 77*(1), 45–74.

Rukhin AL. (1988). Estimated loss and admissible loss estimators. In: Gupta, S.S., Berger, J.O. (eds.) Statistical Decision Theory and Related Topics, IV (West Lafayette, Ind., 1986). vol. 1, pp. 409–418. Springer, New York

Rukhin, A. L. (1988). Loss functions for loss estimation. *Annals of Statistics, 16*(3), 1262–1269.

Sandved, E. (1968). Ancillary statistics and estimation of the loss in estimation problems. *Annals of Mathematical Statistics, 39*(5), 1756–1758.

Wan, A. T. K., & Zou, G. (2004). On unbiased and improved loss estimation for the mean of a multivariate normal distribution with unknown variance. *Journal of Statistical Planning and Inference, 119*(1), 17–22.

Yadegari I. (2017). Prédiction, inférence sélective et quelques problèmes connexes. phdthesis, Faculté des sciences, Université de Sherbrooke, Canada. https://savoirs.usherbrooke.ca/handle/11143/10167