On estimation with weighted balanced type loss function

Mohammad Jafari Jozani,[a] Éric Marchand,[b,1] Ahmad Parsian,[c,2]

a Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, IRAN (e-mail: mjafari2002@yahoo.com)

b University of New Brunswick and Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)

c School of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156, IRAN (e-mail: ahmad_p@cc.iut.ac.ir)

ABSTRACT

For estimating an unknown parameter θ, we introduce and motivate the use of the balanced type loss function: $\mathrm{L}_{w,\delta_0}(\theta,\delta) = \omega\, q(\theta)(\delta-\delta_0)^2 + (1-\omega)q(\theta)(\delta-\theta)^2$, where $0 \le \omega \le 1$, $q(\theta)$ is a positive weight function, and δ_0 is a general "target" estimator. Developments and various examples are given with regards to the issues of admissibility, dominance, Bayesianity, and minimaxity. In many cases, as in Dey, Ghosh, and Strawderman (1999), we show that results for loss L_{w,δ_0} may be inferred directly from corresponding results for weighted squared error loss (i.e, $\omega = 0$). Specific issues related to constrained parameter spaces, which include the choice of the target estimator, are addressed. Finally, we derive minimax estimators of a bounded normal mean θ under loss L_{w,δ_0} with δ_0 being the maximum likelihood estimator of θ.

Keywords: Balanced loss function; Admissibility; Bayes estimator; Minimax estimation; Constrained parameter spaces

MSC: 62C10; 62C15; 62C20; 62F10; 62F30

[1]research supported by NSERC of Canada

[2]research supported by a grant of the Research Council of the Isfahan University of Technology

1 Introduction

This paper deals with the estimation of an unknown parameter θ under the model $X = (X_1, \ldots, X_n) \sim F_\theta$, and with loss

$$\mathrm{L}_{w,\delta_0}(\theta, \delta) = \omega\, q(\theta)(\delta - \delta_0)^2 + (1-\omega) q(\theta)(\delta - \theta)^2; \tag{1}$$

where $0 \leq \omega \leq 1$, $q(\theta)$ is a positive weight function, and δ_0 is a "target" estimator. This loss, which depends on the observed value of $\delta_0(X)$, reflects a desire of closeness of δ both to: (i) θ in terms of weighted squared error loss, and (ii) the target estimator δ_0 in terms of weighted squared distance. The $\omega q(\theta)\,(\delta - \delta_0)^2$ part of the loss is analogous to a penalty term for lack of smoothness in non-parametric regression. The weight ω in (1) calibrates the relative importance of these two criteria. The choice of loss (1) is inspired by the Zellner's (1994) balanced loss function, which has been studied by several others and which corresponds to the case $q(\theta) = 1$, and a least squares δ_0. Indeed, for the particular case where $\delta_0(X) = \bar{X}$, loss (1) becomes equivalent to loss:

$$\frac{\omega}{n} \sum_{i=1}^{n} (X_i - \delta)^2 + (1-\omega)(\delta - \theta)^2;$$

highlighting a goodness of fit term (i.e, $\frac{\omega}{n}\sum_{i=1}^{n}(X_i - \delta)^2$) with the least squares estimates. However, it is useful to view δ_0 as a more general target estimator than least squares as a unifying scheme for establishing decision theoretic properties such as admissibility, Bayesianity, and minimaxity. Moreover, we study the application of (1) in the context of restricted parameter spaces Θ, where it will be desirable for the target estimator to take values solely in Θ.

As shown by Dey, Ghosh and Strawderman (1999), the issues of admissibility and dominance, as well as Bayesianity for BLF essentially does not depend on ω, and can be inferred by previously or to be established results under squared error loss (i.e, $\omega = 0$). As one may anticipate, similar connections can be drawn for losses in (1) for more general $(q(\theta), \delta_0)$. Section 2 deals with such connections, and expands on various illustrations. Moreover, we also develop new connections with regards to minimaxity in Section 3, and present various applications, which include the problem of estimating a bounded normal mean θ under loss $\mathrm{L}_{\omega,\delta_0}$, with $q \equiv 1$ and δ_0 the maximum likelihood estimator.

2 Admissibility and Dominance

For an estimator $\delta(X)$, we will denote the risk under loss (1) as $\mathrm{R}_{w,\delta_0}(\theta, \delta(X))$. We assume throughout risk finiteness of $\mathrm{R}_{w,\delta_0}(\theta, \delta(X))$. For the case $\omega = 0$, we will simply write $\mathrm{R}_0(\theta, \delta(X))$ as the risk becomes independent of δ_0. For similar reasons, we will write L_{0,δ_0} as L_0, and the difference in losses below Δ_{0,δ_0} as Δ_0. As in Dey, Ghosh, and Strawderman (1999), the connections between losses $\mathrm{L}_{\omega,\delta_0}$; $0 \leq \omega < 1$; are driven by Lemma 1's decompositions of the risk $\mathrm{R}_{w,\delta_0}(\theta, \delta(X))$ and of the difference in losses:

$$\Delta_{\omega,\delta_0}(\theta, \delta_1, \delta_2) = \mathrm{L}_{\omega,\delta_0}(\theta, \delta_1) - \mathrm{L}_{\omega,\delta_0}(\theta, \delta_2).$$

Lemma 1 *We have*

(a) $\Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g, \delta_0) = (1-\omega)^2 \Delta_0(\theta, \delta_0 + g, \delta_0)$*;*

(b) $\Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g_1, \delta_0 + (1-\omega)g_2) = (1-\omega)^2 \Delta_0(\theta, \delta_0 + g_1, \delta_0 + g_2)$*;*

(c) $R_{\omega,\delta_0}(\theta, \delta_0(X) + (1-\omega)g(X)) = \omega(1-\omega)R_0(\theta, \delta_0(X)) + (1-\omega)^2 R_0(\theta, \delta_0(X) + g(X))$*.*

Proof. In order to establish part (c), first take expectations on both sides of the equality in part (a) to obtain

$$\mathrm{R}_{\omega,\delta_0}(\theta, \delta_0(X)+(1-\omega)g(X)) = \mathrm{R}_{\omega,\delta_0}(\theta, \delta_0(X))+(1-\omega)^2\{\mathrm{R}_0(\theta, \delta_0(X)+g(X))-\mathrm{R}_0(\theta, \delta_0(X))\}\,.$$

Now, part (c) follows given that $\mathrm{R}_{\omega,\delta_0}(\theta, \delta_0(X)) = (1-\omega)\mathrm{R}_0(\theta, \delta_0(X))$. Part (b) follows directly from (a) as $\Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g_1, \delta_0 + (1-\omega)g_2) = \Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g_1, \delta_0) - \Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g_2, \delta_0)$. Finally, to establish part (a), a straightforward expansion leads to

$$\begin{aligned} \mathrm{L}_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g) &= q(\theta)\{g^2(1-\omega)^2 + 2g(1-\omega)^2(\delta_0 - \theta) + (1-\omega)(\delta_0-\theta)^2\} \\ &= q(\theta)(1-\omega)^2\{g^2 + 2g(\delta_0-\theta)\} + \mathrm{L}_{\omega,\delta_0}(\theta, \delta_0), \end{aligned}$$

which implies $\Delta_{\omega,\delta_0}(\theta, \delta_0 + (1-\omega)g, \delta_0) = q(\theta)(1-\omega)^2\{g^2 + 2g(\delta_0-\theta)\}$, and the desired result.

Now, the following is immediate.

Corollary 1 *(a) The estimator* $\delta_0(X) + (1-\omega)g(X)$ *is admissible under loss* L_{ω,δ_0} *if and only if the estimator* $\delta_0(X) + g(X)$ *is admissible under loss* L_0*;*

(b) $\delta_0(X) + (1-\omega)g_1(X)$ *dominates* $\delta_0(X) + (1-\omega)g_2(X)$ *under loss* L_{ω,δ_0} *if and only if* $\delta_0(X) + g_1(X)$ *dominates* $\delta_0(X) + g_2(X)$ *under loss* L_0*.*

To complement the above with regards to admissibility and dominance, similar connections between Bayes estimators for varying ω are easily derived and given in the next lemma. In particular, the Bayes (or Generalized Bayes) estimator under loss $\mathrm{L}_{\omega,\delta_0}$ may be expressed simply as a convex linear combination of the Bayes (or Generalized Bayes) estimator under loss L_0 (weight $1-\omega$) and δ_0 (weight ω).

Lemma 2 *Under loss* L_{ω,δ_0} *(as in (1)), the Bayes estimator of* θ *under prior* π *admits representation:*

$$\delta_\pi(X) = \omega\delta_0(X) + (1-\omega)\frac{E_\pi(\theta q(\theta)|X)}{E_\pi(q(\theta)|X)};$$

or, equivalently,

$$\delta_\pi(X) = \delta_0(X) + (1-\omega)\{\frac{E_\pi(\theta q(\theta)|X)}{E_\pi(q(\theta)|X)} - \delta_0(X)\}; \qquad (2)$$

as long as $E_\pi(\theta^i q(\theta)|X) < \infty\,; i = 0, 1$*; with probability one.*

Proof. The proof, which is straightforward and similar to the one given by Dey, Ghosh and Strawderman (1999) for $q(\cdot) = 1$, is omitted. (It even follows from Dey, Ghosh and Strawderman's result for $q(\cdot) = 1$ given that Bayesian estimators for (q, π) are necessarily equivalent to Bayesian estimators for $(q \equiv 1, \pi' \equiv q \times \pi)$.)

Results concerning dominance, admissibility, and Bayesianity of a scalar parameter under squared error or weighted squared losses are, of course, plentiful in the literature. For instance, many such results may be found in the textbook of Lehmann and Casella (1998), as well in various papers dating back at least 50 years. Corollary 1 and Lemma 2 imply that, necessarily, such results have easily stated analogs under loss (1), and for arbitrary δ_0. The reverse is also true. Hence, problems that address the issues of dominance, admissibility, and/or Bayesianity, with respect to a loss function $\mathrm{L}_{\omega,\delta_0}$ can be handled, and perhaps have already been handled, by an analysis for loss L_0.

A lengthy enumeration of examples is not pertinent here. Rather, we illustrate the above connections with a few immediate examples, and we emphasize with later examples in Section 3 other related issues such as the minimax criterion and the choice of the target estimator in restricted parameter spaces.

Example 1 *(Affine linear estimators) A sampling of familiar results concerning the performance of affine linear estimators under the assumptions $E_\theta(X) = \theta$, $Var_\theta(X) = \sigma^2(\theta) = \sigma^2$ (e.g., Lehmann and Casella, 1998; Gupta, 1966) indicate that, for estimating θ under squared error loss:*

(i) $cX + d$ is inadmissible whenever $c > 1$, $c < 0$, or $c = 1, d \neq 0$;

(ii) $cX + d$ is, for $X \sim N(\theta, 1)$, admissible iff $c = 1, d = 0$ or $c \in [0, 1), d \in \Re$;

(iii) $cX + d$ is, for $X \sim Gamma(\alpha, \theta/\alpha)$, admissible iff $c = \alpha/(\alpha + 1), d = 0$ or $c \in (0, \alpha/(\alpha + 1)], d > 0$.

The inadmissibility results in (i) translates, for an arbitrary target estimator δ_0, to the inadmissibility under loss L_{ω,δ_0} of estimators of the form $\delta_0(X) + (1 - \omega)(cX + d - \delta_0(X)) = \omega\delta_0(X) + (1-\omega)(cX+d)$ for $c > 1$, $c < 0$, or $c = 1, d \neq 0$. In contrast, given (ii), estimators of the form $\omega\delta_0(X) + (1 - \omega)(cX + d)$ are, under loss L_{ω,δ_0} with $X \sim N(\theta, 1)$, admissible iff $c = 1, d = 0$; or $c \in [0, 1), d \in \Re$. In particular, if $\delta_0(X) = X$, we obtain the inadmissibility of $c'X + d'$ (under loss L_{ω,δ_0} with $\delta_0(X) = X$) for $c' > 1$, $c' < \omega$, or $c' = \omega, d' \neq 0$; and the admissibility for $X \sim N(\theta, 1)$ of $c'X + d'$ iff $c' = 1, d' = 0$; or $c' \in [\omega, 1), d' \in \Re$. Similar inferences can be drawn from (iii). Namely for $X \sim Gamma(\alpha, \theta/\alpha)$, $c'X + d'$ is under loss L_{ω,δ_0}, with $\delta_0(X) = X$, admissible for estimating θ iff $c' = (\alpha + \omega)/(\alpha + 1), d' = 0$ or $c' \in (\omega, (\alpha + \omega)/(\alpha + 1)], d > 0$.

Example 2 *(Minimum risk equivariant estimators) Consider the location model $X \sim f_0(x-\theta)$; with $E_0(X^2) < \infty$; and take L_{ω,δ_0} as in (1) with $q \equiv 1$. It is well known that, under loss L_0, location invariant estimators of the form $X + c$ have constant risk, and that an optimal choice (MRE) is given by $\delta^*(X) = X + c^*$ (where $c^* = -E_0(X)$). Furthermore, the estimator $\delta^*(X)$ is, under loss L_0, minimax, admissible, and Bayes with respect to the right-invariant*

Haar prior $\pi^*(\theta) = 1$. The results of this section tell us that, under loss L_{ω,δ_0}, the estimator $\delta^{**}(X) = \omega\delta_0(X) + (1-\omega)\delta^*(X)$ is admissible, and Bayes with respect to the right-invariant Haar prior $\pi^*(\theta) = 1$. In particular, if the target estimator δ_0 is location invariant of the form $X + c_0$, then $\delta^{**}(X)$ (given by $X + \omega c_0 + (1-\omega)c^*$; and equal, of course, to $\delta^{**}(X)$ if $c_0 = c^*$) also: (i) has constant risk (given part (c) of Lemma 1); (ii) is optimal among location-invariant estimators $X + c$; (iii) is, by virtue of Theorem 1 below, minimax. Finally, we point out that similar developments for observables $X_1, \ldots, X_n$ and for scale parameter models follow along the same lines (also see Example 6).

3 Minimaxity

Here, we capitalize on the results of Lemma 1 to present general results and various examples pertaining to minimax estimators under loss $\mathrm{L}_{\omega,\delta_0}$. The first part of the section deals with cases where the target estimator δ_0 has constant risk under loss L_0, while the second part addresses more general cases and concludes with an application concerning the estimation of a bounded normal mean.

Theorem 1 *Suppose $\delta_0(X)$ has constant risk γ under loss L_0; then $\delta_0(X) + (1-\omega)g(X)$ is minimax under loss L_{ω,δ_0}, with minimax risk $(1-\omega)(\omega\gamma + (1-\omega)r)$, if and only if $\delta_0(X) + g(X)$ is minimax under loss L_0, with minimax risk r.*

Proof. Since $\mathrm{R}_0(\theta, \delta_0(X)) = \gamma$ by assumption, we obtain from part (c) of Lemma 1

$$\mathrm{R}_{\omega,\delta_0}(\theta, \delta_0(X) + (1-\omega)g(X)) = \omega(1-\omega)\gamma + (1-\omega)^2 \mathrm{R}_0(\theta, \delta_0(X) + g(X))\,; \qquad (3)$$

telling us that

$$\sup_\theta \{\mathrm{R}_{\omega,\delta_0}(\theta, \delta_0(X) + (1-\omega)g(X))\} = \omega(1-\omega)\gamma + (1-\omega)^2 \sup_\theta \{\mathrm{R}_0(\theta, \delta_0(X) + g(X))\}\,,$$

and yielding the result.

Corollary 2 *Suppose the estimator $\delta_0(X)$ has constant risk γ under loss L_0. Then $\delta_0(X)$ is minimax under loss L_{ω,δ_0} with constant (and minimax) risk $(1-\omega)\gamma$ if and if only $\delta_0(X)$ is minimax under loss L_0 with constant (and minimax) risk γ.*

Proof. Apply Theorem 1 with $g(X) = 0$.

We pursue with a selection of applications of Theorem 1 and Corollary 2.

Example 3 *(Poisson) Let $X \sim$ Poisson(θ); $\theta \in (0, \infty)$; with $L_0(\theta, \delta) = \frac{(\delta-\theta)^2}{\theta}$ (i.e., information-normalized loss). Since $\delta_0(X) = X$ is minimax under loss L_0 with constant (and minimax) risk 1, Corollary 2 tells us that $\delta_0(X) = X$ is also minimax under any loss L_{ω,δ_0}; ($\omega \in (0,1]$, weight $q(\theta) = \frac{1}{\theta}$); with constant (and minimax) risk $1-\omega$.*

Example 4 *(Binomial) Let $X \sim Bi(n,\theta)$; $\theta \in (0,1)$; with $L_0(\theta,\delta) = \frac{(\delta-\theta)^2}{\theta(1-\theta)}$ (i.e., information-normalized loss). Since $\delta_0(X) = \frac{X}{n}$ is minimax under loss L_0 with constant (and minimax) risk $\frac{1}{n}$, Corollary 2 tells us that $\delta_0(X) = \frac{X}{n}$ is also minimax under any loss L_{ω,δ_0}; ($\omega \in (0,1]$, weight $q(\theta) = \frac{1}{\theta(1-\theta)}$); with constant (and minimax) risk $\frac{1-\omega}{n}$. In contrast, if $L_0(\theta,\delta) = (\delta-\theta)^2$, Corollary 2 applies for the constant risk and minimax under L_0 estimator $\delta_0(X) = \frac{X+(\sqrt{n}/2)}{n+\sqrt{n}}$, and implies the minimaxity of this $\delta_0(X)$ for all losses L_{ω,δ_0}; $\omega \in (0,1]$. Finally, if $L_0(\theta,\delta) = (\delta-\theta)^2$ but $\delta_0(X) = \frac{X}{n}$, Theorem 1 does not apply, and one can proceed directly in finding a constant risk proper Bayes (and hence) minimax estimator (see Sanjari Farsipour and Asgharzadeh, 2003).*

In the next example, the target estimator δ_0 is not minimax, but we can still make use of Theorem 1.

Example 5 *(Negative Binomial) Consider $X \sim NBi(\alpha,p)$; $\alpha > 0$ known, $p \in (0,1)$ unknown, with $\theta = E(X) = \alpha(\frac{1}{p}-1)$; $Var_\theta(X) = \alpha\theta(1+\theta)$; and $P_\theta(X=x) = \frac{\Gamma(\alpha+x)}{x!\Gamma(\alpha)}(\frac{\alpha}{\alpha+\theta})^\alpha(\frac{\theta}{\alpha+\theta})^x I_{\{0,1,\ldots\}}(x)$ Now, consider $L_0(\theta,\delta) = \frac{(\delta-\theta)^2}{Var_\theta(X)}$ for which the maximum likelihood estimator $\delta_0(X) = X$ has constant risk equal to $\gamma = 1$. Given that, under loss L_0, $\frac{X}{1+\alpha}$ $(= X - \frac{\alpha}{1+\alpha}X)$ is minimax, with minimax risk $r = \frac{\alpha}{1+\alpha}$, (e.g., Ferguson, 1967), Theorem 1 tells us that $X - (1-\omega)\frac{\alpha}{1+\alpha}X = (\frac{1+\omega\alpha}{1+\alpha})X$ is minimax under loss L_{ω,δ_0}; with minimax risk $(1-\omega)((\omega \times 1) + (1-\omega)(\frac{\alpha}{1+\alpha})) = \frac{(1-\omega)(\alpha+\omega)}{1+\alpha}$. Moreover, since $\frac{X}{1+\alpha}$ is admissible and dominates $\delta_0(X) = X$ under loss L_0, it follows from Corollary 1 that $(\frac{1+\omega\alpha}{1+\alpha})X$ is admissible and dominates $\delta_0(X) = X$ for any loss $L_{\omega,\delta}; \omega \in (0,1)$. Finally, we point out that other admissibility and Bayesianity results under loss L_0, such as those given by Chou (1995), can be translated via Corollary 1 and Lemma 2 to admissibility and Bayesianity results for loss L_{ω,δ_0}.*

In the next example, there exists many minimax estimators.

Example 6 *(Lower bounded scale parameter) For the model $X \sim Gamma(\alpha,\theta)$; (density proportional to $x^{\alpha-1}e^{-x/\theta}$); consider estimating a lower bounded θ, $\theta \geq a > 0$ (a known), with loss L_{ω,δ_0} as in (1) with $q(\theta) = \theta^{-2}$. Set $h_\alpha(z) = \frac{z^{\alpha+1}e^{-z}}{\int_0^z t^{\alpha+1}e^{-t}dt}$. Van Eeden (1995) showed that, under loss L_0, the Bayes estimator $\delta_a^*(X)$ with respect to the prior $\pi_a^*(\theta) = \theta^{-1}1_{(a,\infty)}(\theta)$, given by $\delta_a^*(x) = \frac{x}{\alpha+1}(1+h_{\alpha+2}(\frac{x}{a}))$, is an admissible and minimax estimator of θ, with minimax risk $r = \frac{1}{\alpha+1}$ (also see Jafari Jozani, Nematollahi, and Shafie 2002). For general target estimator δ_0, the results of Section 2 tell us that the estimator $\delta_a^{**}(X) = \omega\delta_0(X) + (1-\omega)\delta_a^*(X)$ is an admissible Generalized Bayes estimator with respect to π_a^*. If the target estimator δ_0 is chosen to be the MRE estimator $\frac{X}{\alpha+1}$, then it follows from Theorem 1 that $\delta_a^{**}(X)$; given now by $\frac{X}{\alpha+1} + (1-\omega)\delta_a^*(X) = \frac{X}{\alpha+1}(1+(1-\omega)h_{\alpha+2}(\frac{X}{a}))$; is not only admissible, but also minimax, with minimax risk $\frac{1-\omega}{\alpha+1}$ (since $\gamma = R_0(\theta,\frac{X}{\alpha+1}) = \frac{1}{\alpha+1}$)). Observe as well that there exists many minimax estimators under loss L_0, and hence many under loss L_{ω,δ_0}, when δ_0 is the MRE estimator. Similar developments for lower bounded scale or location parameters under losses L_0, which can be applied to losses L_{ω,δ_0}, are given by Marchand and Strawderman (2005a, 2005b).*

Remark 1 *With respect to the previous example, if the target estimator δ_0 takes values outside the parameter space $[a,\infty)$, there is no guarantee for instance that $P_\theta(\delta_a^{**}(X) \in [a,\infty)) = 1$. In fact, if $\delta_0(X) = \frac{X}{\alpha+1}$, then*

$$\lim_{x\to 0} \delta_a^{**}(x) = (1-\omega) \lim_{x \to 0} \delta_a^{*}(x) = (1-\omega)\frac{\alpha+2}{\alpha+1}a,$$

since $\lim_{x\to 0}\delta_a^*(x) = \frac{\alpha+2}{\alpha+1}a$ (van Eeden, 1995). This tells us that $\delta_a^{**}(x)$ would take values within the parameter space if and only if $\omega \le \frac{1}{\alpha+2}$. This phenomenon is not restricted to this example, and occurs since the loss function L_{ω,δ_0} reflects the closeness of δ to the target δ_0, rather than the criterion $\delta \in \Theta$. It was our motivation to study, present examples, and have results available for general target estimators δ_0, which are not necessarily least-squares, and which take values solely in the parameter space. For instance, the general admissibility and Bayesian properties of $\delta_a^{**}(X)$ in Example 6 hold for the truncated version of the MRE estimator $\delta_0(X) = \max(a, \frac{X}{\alpha+1})$; or the maximum likelihood estimator $\delta_0(X) = max(a, X/\alpha)$.

Notwithstanding the various minimax estimation applications of Theorem 1 and Corollary 2 given in Examples 2 to 6, they are limited to cases where the target estimator $\delta_0(X)$ has constant risk. In contrast, the following development relates to situations where $\delta_0(X)$ does not have constant risk, but rather possesses risk properties permitting us to exploit the following well-known criterion for minimaxity (e.g., Lehmann and Casella, 1998, section 5.1).

Lemma 3 *If δ_π is a Bayes estimator with respect to a proper prior π, and $S_\pi = \{\theta \in \Theta : \sup_\theta\{R(\theta,\delta_\pi);\ \theta\in\Theta\} = R(\theta,\delta_\pi)\}$, then δ_π is minimax whenever $P_\pi(\theta \in S_\pi) = 1$.*

To pursue, we define for loss $\mathrm{L}_{\omega,\delta_0}$ and for an estimator $\delta(X)$:

$$\mathrm{S}_{\omega,\delta_0}(\delta) = \{\theta \in \Theta : \sup_\theta\{\mathrm{R}_{\omega,\delta_0}(\theta,\delta(X));\ \theta \in \Theta\} = \mathrm{R}_{\omega,\delta_0}(\theta,\delta(X))\};$$

(i.e., the set of θ's such that $\delta(X)$ attains its maximum risk under loss $\mathrm{L}_{\omega,\delta_0}$); and write $\mathrm{S}_0(\delta)$ for $\mathrm{S}_{0,\delta_0}(\delta)$.

Lemma 4 *Suppose $\delta_\pi(X) = \delta_0(X) + g_\pi(X)$ is unique Bayes under loss L_0 for proper prior π. Suppose further that $P_\pi(\theta \in S_0(\delta_\pi)) = P_\pi(\theta \in S_0(\delta_0)) = 1$, then the estimator $\delta_\pi(X) = \delta_0(X) + (1-\omega)g_\pi(X)$ is unique minimax under loss $L_{\omega,\delta}$.*[3]

Proof. Given that under loss $\mathrm{L}_{\omega,\delta_0}$, $\delta'_\pi(X) = \delta_0(X) + (1-\omega)g_\pi(X)$ is unique Bayes with respect to π by virtue of (2), it suffices to show that

$$\mathrm{P}_\pi(\theta \in \mathrm{S}_{\omega,\delta_0}(\delta'_\pi)) = 1, \tag{4}$$

since we could then apply Lemma 3 to infer the desired result. But (4) follows by Lemma 1 part (c)'s representation of the risk $\mathrm{R}_{\omega,\delta_0}(\theta,\delta'_\pi(X))$ given that, by assumption, both $\mathrm{R}_0(\theta,\delta_0(X))$ and $\mathrm{R}_0(\theta,\delta_\pi(X))$ attain their supremum on the same probability one set.

[3]we note that these assumptions imply that $\delta_0(X) + g_\pi(X)$ is minimax under loss L_0

We now give an application of Lemma 4 to the normal distribution case $X \sim \mathrm{N}(\theta, \sigma^2)$, where θ is bounded to an interval $[-m, m]$, and where the target estimator δ_0 is chosen as the maximum likelihood estimator $\delta_{\mathrm{mle}}(X) = (m \wedge X)\mathrm{sgn}(X)$. Incidentally, Marchand and Perron (2001) provide various dominating estimators $(\delta_0(X)+g(X))$ of $\delta_{\mathrm{mle}}(X)$ under squared error loss L_0, which translate to various dominating estimators $(\delta_0(X)+(1-\omega)g(X))$ of $\delta_{\mathrm{mle}}(X)$ under loss $\mathrm{L}_{\omega,\delta_0}$ with $\delta_0 = \delta_{\mathrm{mle}}$. An example is given by the minimax estimator of Theorem 2 when $m \leq \sigma$. Returning to the minimax estimation of θ, we capitalize on Casella and Strawderman's (1981) work giving conditions for which the Bayes estimator, $\delta_{BU}(X)$, with respect to the (boundary) uniform prior π_{BU} on $\{-m, m\}$ is minimax.

Theorem 2 *For estimating $\theta \in [-m, m]$ with loss L_{ω,δ_0}, as in (1) with $q(\theta) = 1$ and $\delta_0 = \delta_{mle}$, under the model: $X \sim N(\theta, \sigma^2)$ (with pdf $\frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})$; cdf $\Phi(\frac{x-\theta}{\sigma})$) with known σ, the estimator $(1-\omega)\delta_{BU}(X) + \omega\delta_{mle}(X)$ is unique minimax as long as $m \leq m_0\sigma$, with $m_0 \approx 1.0567$ and $\delta_{BU}(X) = \frac{m}{\sigma}\tanh(mX/\sigma^2)$.*

Proof. We use Lemma 4 above, and Lemma 5 which is stated and proven in the Appendix. From Casella and Strawderman (1981), we have $\mathrm{P}_\pi(\theta \in \mathrm{S}_0(\delta_\pi)) = 1$, for $\pi = \pi_{BU}$ and $m \leq m_0\sigma$. Furthermore, we show in Lemma 5 below that the maximum value of $\mathrm{R}(\theta, \delta_{\mathrm{mle}}(X))$ is necessarily attained at $\theta = \pm m$ whenever $m \leq \sqrt{2}\sigma$. Since $\sqrt{2} \geq m_0$, in other words $\mathrm{P}_\pi(\theta \in \mathrm{S}_0(\delta_0)) = 1$ for $\pi = \pi_{BU}$ and $m \leq m_0\sigma$, the result follows directly as a consequence of Lemma 4.

References

Casella, G., Strawderman, W.E. (1981). *Estimating a bounded normal mean.* Annals of Statistics, **9**, 870-878.

Chou, J.P. (1995). *Admissibility of conjugate Bayes estimators for the mean of a negative binomial distribution.* Statistics & Decisions, **13**, 301-306.

Dey, D., Ghosh, M., Strawderman, W.E. (1999). *On estimation with balanced loss functions.* Statistics & Probability Letters, **45**, 97-101.

Ferguson, T. S. (1967). *Mathematical statistics: A decision theoretic approach.* Academic Press, New York-London.

Gupta, M.K. (1966). *On the admissibility of linear estimates for estimating the mean of distributions of the one parameter exponential family*, Calcutta Statistical Association Bulletin, **15**, 14-19.

Jafari Jozani, M., Nematollahi, N., Shafie, K. (2002). *An admissible minimax estimator of a bounded scale-parameter in a subclass of the exponential family under scale-invariant squared error loss.* Statistics & Probability Letters, **60**, 437-444.

Lehmann, E. L., Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag, New York, 2nd edition.

Marchand, Éric, Perron, François (2001). *Improving on the MLE of a Bounded Normal Mean.* Annals of Statistics, **29**, 1078-1093.

Marchand, É., Strawderman, W. E. (2005a). *On Improving on the Minimum Risk Equivariant Estimator of a Location Parameter which is Constrained to an Interval or a Half-Interval.* Annals of the Institute of Statistical Mathematics, **57**, 129-143.

Marchand, É., Strawderman, W. E. (2005b). *On Improving on the Minimum Risk Equivariant Estimator of a Scale Parameter under a Lower-Bound Constraint.* Journal of Statistical Planning and Inference, **134**, 90-101.

Sanjari Farsipour, N. and Asgharzadeh, A. (2004). *Estimation of the parameter of a Bernouilli distribution using a balanced loss function.* Personal communication.

van Eeden, C. (1995) *Minimax estimation of a lower-bounded scale parameter of a gamma distribution for scale-invariant squared error loss.* Canadian Journal of Statistics, **23**, 245-256.

Zellner, A. (1994). *Bayesian and Non-Bayesian estimation using balanced loss functions.* Statistical Decision Theory and Methods V, (J.O. Berger and S.S. Gupta Eds). NewYork: Springer-Verlag, 337-390.

4 Appendix

Lemma 5 *For $X \sim N(\theta, \sigma^2)$ with $|\theta| \leq m$, the maximum risk under squared error loss of the estimator $\delta_{mle}(X)$ is attained on the boundary $\{-m, m\}$, whenever $m \leq \sqrt{2}\sigma$.*

Proof. Without loss of generality, set $\sigma = 1$. We first show that, for all $m \in (0, \sqrt{2})$,

$$(A) \qquad \sup_{\theta \in [-m,m]} \{ \mathrm{R}_0(\theta, \delta_{\mathrm{mle}}(X)) \} = \max\{ \mathrm{R}_0(0, \delta_{\mathrm{mle}}(X)), \mathrm{R}_0(m, \delta_{\mathrm{mle}}(X)) \}.$$

Secondly, we show

$$(B) \qquad \mathrm{R}_0(0, \delta_{\mathrm{mle}}(X)) \leq \mathrm{R}_0(m, \delta_{\mathrm{mle}}(X)), \text{ iff } m \leq m_1 \approx 1.96422.$$

Clearly, (A) and (B) (which are both of independent interest), will then suffice to establish the lemma. To establish (A), we proceed directly by showing that, for $m \leq \sqrt{2}$,

$$\frac{\partial}{\partial^3 \theta} \mathrm{R}_0(\theta, \delta_{\mathrm{mle}}(X)) \geq 0,$$

for $\theta \in [0, m]$. Indeed, observe that either: **(i)** the convexity of $\mathrm{R}_0(\theta, \delta_{\mathrm{mle}}(X))$ for $\theta \in [0, m]$; or **(ii)** a change from concavity to convexity of $\mathrm{R}_0(\theta, \delta_{\mathrm{mle}}(X))$ on $[0, m]$ both imply (A). Note that the sufficiency of **(ii)** exploits the fact that the risk of δ_{mle} is an even function of θ, which implies that its derivative at $\theta = 0^+$ is 0; and that the risk will be decreasing in θ for positive θ in a neighbourhood of 0.

Expanding directly the risk $\mathrm{E}_\theta[(\delta_{\text{mle}}(X)-\theta)^2]$, we obtain

$$\mathrm{R}_0(\theta,\delta_{\text{mle}}(X)) = \int_{-m}^{m}(x-\theta)^2\phi(x-\theta)dx + (m-\theta)^2\Phi(\theta-m) + (m+\theta)^2\Phi(-(\theta+m)). \quad (5)$$

Direct computations now yield

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial\theta}\mathrm{R}_0(\theta,\delta_{\text{mle}}(X)) &= (m+\theta)\Phi(-(\theta+m)) + (\theta-m)\Phi(\theta-m)\,;\\
\frac{1}{2}\frac{\partial}{\partial^2\theta}\mathrm{R}_0(\theta,\delta_{\text{mle}}(X)) &= \Phi(-\theta-m)+\Phi(\theta-m)-(m+\theta)\phi(m+\theta)+(\theta-m)\phi(\theta-m)\,;\\
\frac{1}{2}\frac{\partial}{\partial^3\theta}\mathrm{R}_0(\theta,\delta_{\text{mle}}(X)) &= \phi(\theta-m)(2-(m-\theta)^2))-\phi(\theta+m)(2-(m+\theta)^2))\,.
\end{aligned}$$

Finally, result (A) follows as $\phi(\theta-m)\geq\phi(\theta+m)\geq 0$; $2-(m-\theta)^2 \geq 2-(m+\theta)^2$; and $2-(m-\theta)^2\geq 0$; for $\theta\in[0,m]$ with $m\leq\sqrt{2}$.

To establish (B), use (5) to write

$$\begin{aligned}
\mathrm{R}_0(0,\delta_{\text{mle}}(X)) &= 2m^2\Phi(-m) + 2\textstyle\int_0^m x^2\phi(x)dx\,;\\
\textit{and}\quad \mathrm{R}_0(m,\delta_{\text{mle}}(X)) &= 4m^2\Phi(-2m) + \textstyle\int_0^{2m} x^2\phi(x)dx.
\end{aligned}$$

From this, work with the difference in risks $D(m) = \mathrm{R}_0(m,\delta_{\text{mle}}(X)) - \mathrm{R}_0(0,\delta_{\text{mle}}(X))$ to obtain

$$\frac{\partial}{\partial m}D(m) = 4m\{2\Phi(-2m) - m\phi(2m) - \Phi(-m)\} = 4mN(m) \text{ (say)};$$

$$\text{and } \frac{\partial}{\partial m}N(m) = \phi(2m)\,\{e^{3m^2}+4m^2-5\}.$$

Now, observe that $N(0)=\frac{1}{2}>0$, $\lim_{m\to\infty}N(m)=0$, and that $\frac{\partial}{\partial m}N(m)$ changes signs from - to + on $(0,\infty)$. These conditions imply that $N(m)$ changes signs once from + to - for $m\in(0,\infty)$, which in turn tells us that $D(m)$ has at most one signs change from + to - for $m\in(0,\infty)$ (as $D(0)=0$). Finally, a numerical evaluation yields $D(m)=0$ for $m=m_1\approx 1.96422$.