Minimax estimation of constrained parametric functions for discrete families of distributions

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

a Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, IRAN (e-mail: m-jafari@sbu.ac.ir)

b Université de Sherbrooke, Département de mathématiques, Sherbrooke, QC, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)

ABSTRACT

For a vast class of discrete model families where the natural parameter is constrained to an interval, we give conditions for which the Bayes estimator with respect to a boundary supported prior is minimax under squared error loss type functions. Building on a general development of Éric Marchand and Ahmad Parsian, applicable to squared error loss, we obtain extensions to various parametric functions and squared error loss type functions. We provide illustrations for various distributions and parametric functions, and these include examples for many common discrete distributions, as well as when the parametric function is a zero-count probability, an odds-ratio, a Binomial variance, and a Negative Binomial variance, among others.

Keywords: Minimax estimation, restricted parameter space, discrete distributions, squared error loss, zero count probability, odds-ratio, Binomial variance, Negative Binomial variance.

MSC: primary 62C10; 62F10; 62F30; 62F99

[1]Research supported by a grant of the Institute for Research and Planning in Higher Education, Ministry of Science, Research and Technology, Iran

[2]Research supported by NSERC of Canada

1 Introduction

With the criteria of minimaxity playing a vital role in the development of statistical theory and practice (e.g., Brown, 1994 or Strawderman 2000), and the emergence of new statistical techniques and problems with constraints on the parameters to be estimated (e.g., as reviewed by Marchand and Strawderman, 2004; or Mandelkern, 2002), there has been much interest and work in the problem of minimax estimation in restricted parameter spaces.

Here, for certain types of discrete models, represented by probability functions $p_\theta(\cdot)$ with θ restricted to a known interval $[a, b]$; we search for explicit conditions under which the minimax estimator is Bayes with respect to a boundary supported prior. More precisely, we are concerned with:

(A) minimax estimation of $\gamma(\theta)$ under squared error loss

$$L_1(\theta, \delta) = (\delta - \gamma(\theta))^2;$$

(B) minimax estimation of θ under "γ-loss"

$$L_2(\theta, \delta) = (\gamma(\delta) - \gamma(\theta))^2;$$

where $\gamma(\cdot)$ is a monotone function. Loss functions L_1 and L_2 are mathematically equivalent (see Lemma 2), but stem from separate practical perspectives. The use of loss function L_2 is interesting for its own sake, but it also has been motivated recently in the work of Ganjali and Shafie (2006), who studied Lehmann's (1951) criteria of L-unbiasedness associated with γ-loss L_2, and illustrated how in some cases plausible estimators arise, in opposition to implausible usual unbiased estimators. Their motivating example concerns a $X \sim$ Poisson(λ) model and the zero count probability with $\theta = e^{-\lambda}$, where the L-unbiased estimator associated with L_2 (with $\gamma(z) = \log z$) is the plausible $\delta(X) = e^{-X}$ in comparison to the usual unbiased (UMVU) estimator of θ, which takes negative values. The use of loss L_2 is also illustrated in Schäbe (1991).

On the other hand, as in (A) with squared error loss L_1, interest could center around the estimation of a parametric function $\gamma(\theta)$, such as a zero count probability $\gamma(\theta) = P_\theta(X = 0)$, an odds-ratio $\gamma(\theta) = \frac{P_\theta(X \in E)}{1 - P_\theta(X \in E)}$ for an event E, or a variance $\gamma(\theta) = Var_\theta(X)$, among others, under the constraint $\theta \in [a, b]$.

For the particular problem of investigating conditions under the least favorable prior is supported on the boundary of a compact parameter space, the results of Casella and Strawderman (1981), as well as those of Zinzius (1981) have inspired a lot of further work in this area. Techniques used can be roughly divided into either: **(i)** various variational considerations such as sign change arguments (e.g., Casella and Strawderman, 1981; Berry, 1990; Marchand and Strawderman, 2004; Marchand and Parsian, 2006), and **(ii)** existence proofs of such least favorable priors using the "convexity" (of the risk) technique (e.g., Zinzius, 1981; DasGupta, 1985; Bader and Bischoff, 2003; Boratyńska, 2005).

Marchand and Parsian (2006) obtained, for a large class of discrete models a unified approach in analyzing under squared error loss, the minimaxity of a Bayes estimator associated with a boundary supported prior; obtaining either sufficient or necessary and sufficient conditions. This paper extends these results to loss functions L_1 and L_2 with $\gamma(t) \neq t$. As in the work of Marchand and Parsian (2006), and Marchand and MacGibbon (2000), a key variational element (see proof of Theorem 1) involves a third derivative argument. Various examples relative to both loss types L_1 and L_2 are given in Section 3. These include the interesting cases of zero count probabilities, odd ratios, and Binomial or Negative Binomial variances, among others.

2 Preliminary results and definitions

Our framework closely resembles the one introduced by Marchand and Parsian (2006), and several of the following preliminary results and definitions are reproduced here for sake of completeness. Our results apply to observable random vectors $\mathbf{X} =$

$(X_1,\ldots,X_n)$; $n \geq 1$, where the X_i's are identically but not necessarily independently distributed, discrete random variables with joint probability function $p_\theta(\mathbf{x}) = P_\theta(\mathbf{X}=\mathbf{x}); \theta \in [a,b] \subseteq \Theta$; such that the distribution of $\mathbf{X}$ at $\theta = a$ is degenerate at $(\underline{s},\ldots,\underline{s})$; Θ being the natural parameter space, i.e., $\Theta = \{\theta : p_\theta(\cdot)$ is a probability function $\}$. Since we can translate (i.e., X_i to $X_i - \underline{s}$), we assume hereafter that $\underline{s} = 0$. Let $A = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$, denote $G(n,\theta) = P_\theta(\mathbf{X} \in A)$, and observe that $G(n,a) = 1$ by assumption. We work throughout the paper with bounded and strictly monotone $\gamma(\theta)$'s such that $G(n,\gamma^{-1}(z))$, $z \in \gamma(\Theta)$, is three times continuously differentiable with respect to z with a bounded third derivative. Without loss of generality, we handle the cases where $\gamma(\theta)$ is strictly increasing with $\gamma(a)=0$, since we can always, in failing cases, set $\epsilon(\theta) = \gamma(a) - \gamma(\theta)$ and exploit the equivalence between the risks $E_\theta[(\delta'(X)-\epsilon(\theta))^2]$ and $E_\theta[(\delta(X)-\gamma(\theta))^2]$ with $\delta'(X) = \gamma(a) - \delta(X)$; as well as the equivalence between the minimaxity of $\delta^*(X)$ as an estimator of $\gamma(\theta)$, with the minimaxity of $\gamma(a) - \delta^*(X)$ as an estimator of $\epsilon(\theta)$. Finally, for sake of simplicity but mostly in view of the several applications that follow, we further assume that $a = 0, b = m$, even though cases where $a \neq 0$ can be handled along similar lines.

Definition 1 *For a given $\gamma(\cdot)$, define $\mathcal{C}_\gamma$ as the class of all families $p_\theta(\cdot)$ for $\mathbf{X}$ such that $G(n,0)=1$ and*

$$(-1)^k \frac{\partial^k}{\partial z^k} G(n,\gamma^{-1}(z)) \geq 0, \;\; z \in \gamma(\Theta), \;\; for\ k = 1,2,3;$$

with strict inequality for $k=1$ and the derivative evaluated at 0^+. As well, for the identity case $\gamma(\theta)=\theta$, denote $\mathcal{C}_\gamma$ as $\mathcal{C}$.

Completely monotone $G(n,\theta)$'s in $\gamma(\theta)$, which possess derivatives of alternating signs for all positive order, generate families in $\mathcal{C}_\gamma$; and their properties play an important role in Marchand and Parsian (2006), as well as in our work below. We also refer to the following lemma, which is very similar to a lemma on completely monotone functions by Feller (1996, page 417), in order to identify $\gamma(\cdot)$'s and $p_\theta(\cdot)$'s, such that $p_\theta(\cdot) \in \mathcal{C}$ implies $p_\theta(\cdot) \in \mathcal{C}_\gamma$.

Lemma 1 *Let $p_\theta(\cdot) \in \mathcal{C}$, and let $\gamma(\cdot)$ be an increasing and bounded function with inverse $\gamma^{-1}(\cdot)$. Then $p_\theta(\cdot) \in \mathcal{C}_\gamma$ whenever $(-1)^k \frac{d^k}{dz^k}\gamma^{-1}(z) \leq 0$ for $k = 2, 3$.*

Proof. Setting $z = \gamma(\theta)$, the result follows easily by examining directly the derivatives of the composition $G(n, \gamma^{-1}(z))$.

We pursue with various results relative to risk functions and the minimax criterion, in particular as associated with two-point boundary priors. As mentioned above and as now stated, loss functions L_1 and L_2 will necessarily lead to equivalent analyses. This will permit us to focus with developments relative to loss L_1 only for awhile, and relegate L_2 loss applications to the latter part of Section 3.

Lemma 2 *The estimator $\gamma(\delta(X))$ is minimax for estimating $\gamma(\theta)$ under loss L_1 if and only if the estimator $\delta(X)$ is minimax for estimating θ under loss L_2.*

Proof. The result follows easily given that the risks of the estimators $\gamma(\delta(X))$ under L_1 and $\delta(X)$ under L_2 are, as functions of θ, equal.

Our results rely on a following well known criteria for minimaxity applied to boundary two point priors. As well from now on, we set $z = \gamma(\theta)$ and express risks as functions of z.

Lemma 3 *A two point boundary prior π (for $\gamma(\theta)$) on $\{0, \gamma(m)\}$ is least favourable, and the corresponding Bayes estimator $\delta_\pi(X)$ is minimax if and only if*

$$R(0, \delta_\pi(X)) = R(\gamma(m), \delta_\pi(X)) = sup\{R(z, \delta_\pi); 0 \leq z \leq \gamma(m)\} \tag{1}$$

For our model, we consider two-point (boundary) priors on $\{0, \gamma(m)\}$. The following result, which is easily verified, identifies the unique "equalizer" rule among two-point boundary priors as well as its risk function, which we use for further analysis. It is followed by a necessary, and useful, condition for this equalizer rule to be minimax.

Lemma 4 *(a) Among two point boundary priors, the unique equalizer Bayes rule under loss L_1 is given by*

$$\delta^*(\mathbf{x}) = y^* 1_A(\mathbf{x}) + \gamma(m) 1_{A^C}(\mathbf{x}), \tag{2}$$

with

$$y^* = \frac{\gamma(m)\sqrt{G(n,m)}}{1+\sqrt{G(n,m)}}; \tag{3}$$

(b) the risk function of $\delta^*(X)$ under loss L_1, expressed as a function of $z = \gamma(\theta)$, may be written as

$$R(z, \delta^*(X)) = (\gamma(m) - z)^2 + \big\{(y^* - z))^2 - (\gamma(m) - z)^2\big\} G(n, \gamma^{-1}(z)). \tag{4}$$

Lemma 5 *For a family $p_\theta(\cdot) \in \mathcal{C}_\gamma$, and for estimating $\gamma(\theta)$ under loss L_1, a necessary condition for (1) to be satisfied with $\delta_\pi(\mathbf{X}) = \delta^*(\mathbf{X})$ is $\gamma(m) \le c_0(n,\gamma)$, where $c_0(n,\gamma)$ is a unique (positive) solution in c of the equation $T_\gamma(c) = 0$ with*

$$T_\gamma(c) = c + \Big(\frac{2}{\frac{\partial}{\partial z} G(n, \gamma^{-1}(z))|_{z=0}}\Big) \frac{\sqrt{G(n,\gamma^{-1}(c))}(1+\sqrt{G(n,\gamma^{-1}(c)\ )}}{1+2\sqrt{G(n,\gamma^{-1}(c))}}. \tag{5}$$

Proof. The result follows, as in Marchand and Parsian (2006, Lemma 4), by working with the necessary condition $\frac{\partial}{\partial z} R(z, \delta^*(X))|_{z=0} \le 0$ for (1) to hold. Note that the monotonicity of $T_\gamma(c)$ is used.

Under squared error loss (i.e., $\gamma(t) = t$ in L_1), Marchand and Parsian (2006) obtained simple conditions for the equalizer rule $\delta^*(X)$ to be minimax. The following theorem provides at the same time an extension to estimating $\gamma(\theta)$ under loss L_1, as well as a slight improvement in the conditions. This improvement is illustrated with the example that follows and concerns an application to a Negative Binomial model. Hereafter, we denote $c_0(n)$ and $T(m)$ as Lemma 5's $c_0(n,\gamma)$ and $T_\gamma(m)$ respectively with $\gamma(t) = t$. Also, define $k(z) = z\frac{\partial^2}{\partial z^2} G(n, \gamma^{-1}(z)) + 2\frac{\partial}{\partial z} G(n, \gamma^{-1}(z))$, and let $c_1(n,\gamma) := \sup\{z \in \gamma(\Theta) |\ k'(s) \ge 0 \ \textit{for all}\ s \in [0, z]\}$. Observe that for $p_\theta \in \mathcal{C}_\gamma$, we must have $c_1(n,\gamma) > 0$.

Theorem 1 *For estimating $z = \gamma(\theta)$ under squared error loss (i.e., L_1) with $p_\theta(\cdot) \in \mathcal{C}_\gamma$,*

(a) $\delta^*(\mathbf{X})$ is minimax whenever $\gamma(m) \leq c_0(n,\gamma) \wedge c_1(n,\gamma)$;

(b) Whenever $c_0(n,\gamma) \leq c_1(n,\gamma)$; or equivalently $T_\gamma(c_1(n,\gamma)) \geq 0$; $\delta^*(X)$ is minimax iff $\gamma(m) \leq c_0(n,\gamma)$.[3]

Proof. The necessity in part (b) is a consequence of Lemma 5, while the sufficiency follows from part (a). In part (a), as in Marchand and Parsian (2006), it suffices for values of m such that $\gamma(m) \leq c_0(n,\gamma)$ that $\frac{\partial^2}{\partial z^2} R(z, \delta^*(X))$ be nondecreasing in z, $z \in (0, c_1(n,\gamma))$, for $\delta^*(X)$ be minimax. From (4), observe that

$$\frac{\partial^2}{\partial z^2} R(z, \delta^*(X)) = 2 + (\gamma(m)^2 - (y^*)^2)(-\frac{\partial^2}{\partial z^2} G(n, \gamma^{-1}(z))) + 2(\gamma(m) - y^*)k(z),$$

which is indeed nondecreasing for $z \in (0, c_1(n,\gamma))$, since $p_\theta(\cdot) \in \mathcal{C}_\gamma$, $y^* \leq \gamma(m)$, and $k(z)$ increases by assumption for this range of z values.

Example 1 *Consider $X_1, \ldots, X_n$ independently and identically distributed as Negative Binomial with parameters (α, p) (α known, $0 < p \leq 1$), $\theta = E_\theta(X_i) = \alpha(\frac{1}{p} - 1)$, and $G(n,\theta) = (\frac{\alpha}{\alpha+\theta})^{n\alpha}$. Marchand and Parsian (2006) showed that, whenever $n\alpha \geq s_0 \approx 11.876904$, $\delta^*(X)$ given in (2) is minimax for estimating $\gamma(\theta) = \theta$ iff $m \leq c_0(n)$. However, for $n\alpha < s_0 \approx 11.876904$, they only obtain a sufficient condition. Using Theorem 1, we can show indeed that the condition $m \leq c_0(n)$ is always necessary and sufficient. To see this, observe that $k(z) = z\frac{\partial^2}{\partial z^2} G(n, \gamma^{-1}(z)) + 2\frac{\partial}{\partial z} G(n, \gamma^{-1}(z)) \propto (\alpha + z)^{-(n\alpha+2)}\{(n\alpha - 1)z - 2\alpha\}$; $k'(z) \propto (n\alpha + 1)(\alpha + z)^{-(n\alpha+3)}\{3\alpha - (n\alpha - 1)z\}$, which tells us that $c_1(n) =$"∞" for $n\alpha \leq 1$, and $c_1(n) = \frac{3\alpha}{n\alpha - 1}$ otherwise. Finally, to establish that part (b) of Theorem 1 can be applied, evaluate for $n\alpha > 1$*

$$T(c_1(n)) = \frac{1}{n}\{\frac{3n\alpha}{n\alpha - 1} - 2\nu\{(\frac{n\alpha - 1}{n\alpha + 2})^{n\alpha/2}\}\},$$

with $\nu(y) = \frac{y(1+y)}{1+2y}$, checking directly that $T(c_1(n)) \geq 0$.

[3] In the identity case, the above c_0 matches Marchand and Parsian's m_0, but the c_1 is larger than their m_1, thus providing an improvement.

3 Applications

In this section, we describe various applications for both loss functions L_1 and L_2. In view of Theorem 1, the critical condition to consider is that of $p_\theta(\cdot) \in \mathcal{C}_\gamma$, which involves checking that both $\frac{\partial^2}{\partial z^2} G(n, \gamma^{-1}(z))$ and $-\frac{\partial^3}{\partial z^2} G(n, \gamma^{-1}(z))$ are nonnegative.

Example 2 *(non-zero count or zero count probabilities)*
Suppose the X_i's are independent and that we wish to estimate, for $\theta \in [0, m]$, $z = \gamma(\theta) = P_\theta(X_1 > 0) = 1 - G(1, \theta)$, where $P_\theta(X_1 > 0)$ is increasing in θ with $P_0(X_1 > 0) = 0$. By re-parametrization, we obtain a Binomial model $G(n, \gamma^{-1}(z)) = (G(1, \gamma^{-1}(z)))^n = (1 - z)^n$; permitting us to apply either Theorem 1 or known Binomial model results of Marchand and MacGibbon (2000). Namely, we infer that $\delta^*(X)$, given by (2), is minimax iff $P_m(X_1 > 0) \leq c_0(n, \gamma)$, with $c_0(n, \gamma) \approx 0,912955/n$ as given by Marchand and MacGibbon. Notice the condition also applies to the minimaxity of $1 - \delta^*(X)$ as an estimator of the zero count probability $P_\theta(X_1 = 0)$ (see paragraph preceding Definition 1). As well, observe that we do not require $p_\theta(\cdot) \in \mathcal{C}_\gamma$, but rather only decreasing $P_\theta(X_1 > 0)$ with $P_0(X_1 > 0) = 0$.

Example 3 (*odds ratios of the type $P_\theta(X_1 > 0)/P_\theta(X_1 = 0)$*)
Suppose the X_i's are independent and that we wish to estimate, for $\theta \in [0, m]$, an odds ratio of the type $z = \gamma(\theta) = P_\theta(X_1 > 0)/P_\theta(X_1 = 0)$, where $P_\theta(X_1 > 0)$ is increasing in θ with $P_0(X_1 > 0) = 0$. Again by re-parametrization, we obtain a Negative Binomial model $G(n, \gamma^{-1}(z)) = (1 + z)^{-n}$; permitting us to apply either Theorem 1, or even Example 1 directly. As above, we do not require $p_\theta(\cdot) \in \mathcal{C}_\gamma$, but rather only decreasing $P_\theta(X_1 > 0)$ with $P_0(X_1 > 0) = 0$.

Example 4 *(cases where $(-1)^k \frac{\partial^k}{\partial z^k} \gamma^{-1}(z) \leq 0, k = 2, 3$)*
Lemma 1 provides a mechanism to generate many examples where Theorem 1 may be applied by simple selections of $p_\theta(\cdot) \in \mathcal{C}$ and $\gamma(\cdot)$ such that $(-1)^k \frac{\partial^k}{\partial z^k} \gamma^{-1}(z) \leq 0, k = 2, 3$. Two examples of such γ's are given by: **(i)** *$\gamma(\theta) = \frac{\theta}{c - \theta}$ in cases where Θ is*

bounded above by c, and **(ii)** *cases where* $\gamma(\theta) = A\theta + B\theta^2$ *with* $A > 0, B > 0$. *A particular interesting application in* **(ii)** *is given by a Negative Binomial model (see Example 1) variance, which is quadratic in its mean* θ *with* $A = 1$ *and* $B = 1/\alpha$, *and where* $G(n, \gamma^{-1}(z)) = (\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{z}{\alpha^2}}\ )^{-n\alpha}$. *Finally, we mention that further applications are given below in Examples 6 and 7.*

The next example is one where $\gamma(\cdot)$ does not satisfy the conditions of Lemma 1, but where Theorem 1 can be applied nevertheless.

Example 5 *(Binomial variance)*
Let $X_1, \ldots, X_n$ *be independent Bernouilli*(θ) *and consider estimating* $\gamma(\theta) = Var_\theta(X_1) = \theta(1-\theta)$, *with the constraint* $\theta \in [0, m]$; $m \le \frac{1}{2}$. *Here* $\gamma^{-1}(z) = \frac{1}{2} - \sqrt{\frac{1}{4} - z}$ *does not satisfy the conditions of Lemma 1, but we can work directly with* $G(n, \gamma^{-1}(z)) = (1 - \gamma^{-1}(z))^n = (\frac{1}{2} + \sqrt{\frac{1}{4} - z})^n$, *showing directly in the Appendix that*

$$(-1)^k \frac{\partial^k}{\partial z^k} (\frac{1}{2} + \sqrt{\frac{1}{4} - z}\ )^n \ge 0; k = 1, 2, 3; \text{ whenever } n \ge 4 \text{ and } m \le \frac{1}{2}\frac{n-3}{n-2}. \quad (6)$$

Hence, we may apply part (a) of Theorem 1 for estimating an upper bounded $Var_\theta(X_1)$, *whenever* $n \ge 4$ *and* $m \le \frac{1}{2}\frac{n-3}{n-2}$. *Furthermore, it is shown as well in the Appendix that*

$$c_0(n, \gamma) \le c_1(n, \gamma) \text{ for all } n \ge 5, \quad (7)$$

which tells us that $\delta^*(X)$, *given in (2) with* $y^* = m\frac{(1-m)^{n/2+1}}{1+(1-m)^{n/2}}$, *is minimax for estimating* $\gamma(\theta) = Var_\theta(X_1)$, $\theta \in [0, m]$, *as long as* $m(1-m) \le c_0(n, \gamma)$; $c_0(n, \gamma)$ *derived from (5) as the solution in c of the equation*

$$c = \frac{2}{n} \frac{(\frac{1}{2} + \sqrt{\frac{1}{4} - c}\ )^{n/2} + (\frac{1}{2} + \sqrt{\frac{1}{4} - c}\ )^n}{1 + 2(\frac{1}{2} + \sqrt{\frac{1}{4} - c}\ )^{n/2}}.$$

We conclude with two examples relative to loss L_2 (but also L_1), and where we make use of Lemma 2.

Example 6 *(Loss L_2 with $\gamma(\theta) = -\log(1-\theta)$)*

Suppose we are concerned with estimating θ; $\theta \leq m$; under loss L_2 with $\gamma(t) = -\log(1-t)$, in cases where $\sup \Theta < 1$. For the corresponding L_1 problem, we have $\gamma^{-1}(z) = 1-e^{-z}$, which satisfies the conditions of Lemma 1. Therefore, the conditions of Theorem 1 are necessarily satisfied with $G(n, \gamma^{-1}(z)) = G(n, 1-e^{-z})$, and applicable to the estimator $\delta^*(X)$ given in (2), as soon as $p_\theta(\cdot) \in \mathcal{C}$. In such cases, by virtue of Lemma 2, we have that $1 - e^{-\delta^*(X)}$ is minimax for estimating θ; $\theta \in [0, m]$; under loss L_2 with $\gamma(t) = -\log(1-t)$ for sufficiently small m (or $\gamma(m)$) as prescribed by Theorem 1. As a specific example, consider a Binomial model $G(n,\theta) = (1-\theta)^n$, for which we obtain $G(n, 1-e^{-z}) = e^{-nz}$, in other words a Poisson model $G(n, \cdot)$. So, either applying Theorem 1, or directly referring to Marchand and Parsian's (2006) Example 1 for Poisson models, we obtain a minimax result for estimating an upper bounded Binomial(n,θ) proportion θ under loss L_2 with $\gamma(t) = -\log(1-t)$. Namely, the estimator $1 - e^{-\delta^*(X)}$, given by (2) with $\gamma(t) = -\log(1-t)$, is minimax iff $\log(1-m) \leq \frac{c_0}{n}$, or equivalently $m \leq 1 - e^{-c_0/n}$, with $c_0 \approx 0.912955$ (for large n).

Example 7 *(Loss L_2 with $\gamma(\theta) = e^{a\theta} - 1; a > 0$)*

Suppose now that we are concerned with estimating θ under loss L_2 with $\gamma(t) = e^{at} - 1, a > 0$. Here $\gamma^{-1}(z) = \frac{1}{a}\log(1+z)$ satisfies the conditions of Lemma 1, implying that Theorem 1 is applicable with $G(n, \gamma^{-1}(z)) = G(n, \frac{1}{a}\log(1+z))$ as long as $p_\theta(\cdot) \in \mathcal{C}$. In such cases, the minimaxity for sufficiently small m of Theorem 1's $\delta^*(X)$, as an estimator of $\gamma(\theta) = e^{a\theta} - 1$, under squared error loss L_1 implies the minimaxity of $\frac{1}{a}\log(1+\delta^*(X))$ for the same sufficiently small m values, as an estimator of θ under loss L_2 with $\gamma(t) = e^{at} - 1$. As a specific example, consider a Poisson(θ) model $G(n,\theta) = e^{-n\theta}$, for which we obtain $G(n, \frac{1}{a}\log(1+z)) = (1+z)^{-n/a}$; i.e., a Negative Binomial $G(n,z)$. Therefore squared error results pertaining to Example 1 yield minimaxity results applied to $\frac{1}{a}\log(1+\delta^*(X))$ for estimating a Poisson mean parameter θ under loss L_2 with $\gamma(t) = e^{at} - 1$. In both cases, the upper bound on m

becomes $\frac{1}{a}\log(1+c_0(n))$*, with* $c_0(n)$ *given in Example 1.*

Acknowledgements

The authors are thankful to two referees whose comments and suggestions improved the presentation of our results.

References

Bader, G. and Bischoff, W. (2003). Old and new aspects of minimax estimation of a bounded parameter, *Mathematical Statistics and Applications, Festschrift for Constance van Eeden.* IMS Lecture Notes-Monograph Series, **42**, pp. 15-30.

Berry, C. (1990). Minimax estimation of a bounded normal mean vector, *Journal of Multivariate Analysis*, **35**, 130-139.

Boratyńska, A. (2005). Two point priors and minimax estimation of a bounded parameter under convex loss, *Applicationes Mathematicae*, **32**, 145-153.

Brown, L.D. (1994). Minimaxity, more or less, in *Statistical Decision Theory and Related Topics, V*, editors S.S. Gupta and J.O Berger. Springer, New York, pp. 1-18.

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

DasGupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set, *Sankhyā: Series A*, **47**, 326-332.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*: Volume II, John Wiley & Sons.

Ganjali, M. and Shafie, K. (2006). Estimation based on an appropriate choice of loss function, *Journal of the Iranian Statistical Society*, **4**, to appear.

Lehmann, E.L. (1951). A general concept of unbiasedness, *Annals of Mathematical Statistics*, **22**, 587-592.

Mandelkern, M. (2002). Setting Confidence Intervals for Bounded Parameters with discussion, *Statistical Science*, **17**, 149-172.

Marchand, É. and MacGibbon, B. (2000). Minimax estimation of a constrained binomial proportion, *Statistics and Decisions*, **18**, 129-167.

Marchand, É. and Parsian, A. (2006). Minimax estimation of a bounded discrete parameter, *Statistics & Probability Letters*, **76** , 547-554.

Marchand, É. and Strawderman, W. E. (2004). Estimation in restricted parameter spaces: A review, *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, pp. 1-24.

Schäbe, H. (1991). Bayes estimates under asymmetric loss, *IEEE Transactions on Reliability*, **40**, 63-67.

Strawderman, W. E. (2000). Minimaxity, *Journal of the American Statistical Assocation*, **95**, 1364-1368.

Zinzius, E. (1981). Minimaxschätzer für den Mittelwert θ einer normalverteilten Zufallsgröβe mit bekannter Vaianz bei vorgegebener oberer und unterer Schrannke für θ. *Math. Operationsforsch. Statist. Ser. Statist.*, **12**, 551-557.

4 Appendix

4.1 Proof of (6)

Working directly with derivatives, we have

(a) $\frac{\partial}{\partial z}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^n = -\frac{n}{2}(\frac{1}{4}-z)^{-1/2}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^{n-1} < 0,$

(b) $\frac{\partial^2}{\partial z^2}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^n = \frac{n}{4}(\frac{1}{4}-z)^{-3/2}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^{n-2}((n-2)\sqrt{\frac{1}{4}-z}-\frac{1}{2}),$

(c) $\frac{\partial^3}{\partial z^3}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^n = -\frac{n}{8}(\frac{1}{4}-z)^{-5/2}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^{n-3}\{3(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^2 - 3(n-1)(\frac{1}{4}-z)^{1/2}(\frac{1}{2}+\sqrt{\frac{1}{4}-z}) + (n-1)(n-2)(\frac{1}{4}-z)\}$.

Using (b), we have indeed $\frac{\partial^2}{\partial z^2}G(n,\gamma^{-1}(z)) \geq 0; z \leq m(1-m)$; whenever $(n-2)\sqrt{\frac{1}{4}-z} \leq \frac{1}{2}$, or equivalently $m \leq \frac{1}{2}\frac{n-3}{n-2}$. Finally, set $t = \sqrt{\frac{1}{4}-z}$, and observe that $\frac{\partial^3}{\partial z^3}(\frac{1}{2}+\sqrt{\frac{1}{4}-z})^n \leq 0$, as long as

$$t^2(n^2-6n+8) - \frac{3t}{2}(n-3) + \frac{3}{4} \geq 0$$

which is always true for $n \geq 4, t \leq \frac{1}{2}$, thus completing the proof.

4.2 Proof of (7)

We will make use of the following lemma.

Lemma 6 *For $n \geq 4$, we have $(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{n}}\,)^n \leq \frac{1}{2}$.*

Proof. Given that $(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{n}}\,)^n = (1-\frac{a_n}{n})^n$ with $a_n = (\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{n}}\,)^{-1}$, it follows as $a_n \geq 1$ that $(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{n}})^n \leq (1-\frac{1}{n})^n \leq e^{-1} \leq \frac{1}{2}$.

To establish the result, it will suffice to show that: **(i)** $T_\gamma(\frac{1}{n}) \geq 0$ for all $n \geq 4$ and **(ii)** $c_1(n,\gamma) \geq \frac{1}{n}$ for all $n \geq 5$, which will imply $T_\gamma(c_1(n,\gamma)) \geq 0$; $n \geq 5$ as specified in Theorem 1. For **(i)**, evaluating (5), setting $w_n = (\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{n}})^{n/2}$ and using Lemma 6, we obtain

$$T_\gamma(\frac{1}{n}) = \frac{1}{n} - \frac{2}{n}\frac{w_n+w_n^2}{1+2w_n} \geq \frac{1}{n} - \frac{2}{n}\frac{w_n+1/2}{1+2w_n} = 0,$$

establishing the result.

To establish **(ii)**, it suffices to show that $k'(\frac{1}{n}) \geq 0; n \geq 5$; with $k'(z)$ given prior to Theorem 1. Setting $t = \sqrt{\frac{1}{4}-z}$, we have

$$\begin{aligned} k'(z) &= 3\frac{\partial^2}{\partial z^2}G(n,\gamma^{-1}(z)) + z\frac{\partial^3}{\partial z^3}G(n,\gamma^{-1}(z)) \\ &= \frac{n}{8}t^{-5}(\frac{1}{2}+t)^{n-2}\psi_n(t) \end{aligned}$$

where

$$\psi_n(t) = (n^2-4)t^3 + (\frac{n^2}{2}+\frac{3}{2}n-\frac{5}{2})t^2 + (\frac{3n}{4}-\frac{3}{2})t - \frac{3}{8}.$$

Now, since $\psi_n(t)$ increases in t, the sought after property $k'(\frac{1}{n}) > 0$; $n \geq 5$; is equivalent to $\psi_n(\sqrt{\frac{1}{4} - \frac{1}{n}}) \geq 0$; $n \geq 5$. Finally, by identifying $A(n) = (\frac{n^2+3n-5}{2})(\frac{1}{4} - \frac{1}{n}) - \frac{3}{8}$ and exploiting the fact that $A(n)$ increases in n; $n \geq 4$, we infer with the deletion of the terms in t^3 and t in $\psi_n(t)$ that $\psi_n(\sqrt{\frac{1}{4} - \frac{1}{n}}) \geq A(n) \geq A(5) = \frac{1}{2} \geq 0$, establishing result **(ii)** and completing the proof.