Minimax estimation of a Binomial proportion p when $|p - 1/2|$ is bounded

Éric Marchand[a], François Perron[b], and Rokhaya Gueye[b,1]

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

b Université de Montréal, Département de mathématiques et de statistique, C.P. 6128, succursale Centre-Ville, Montréal, Qc, CANADA, H3C 3J7 (e-mails: perronf@dms.umontreal.ca; gueye@dms.umontreal.ca)

Abstract

We consider the problem of estimating the parameter p of a Binomial(n, p) distribution when p lies in the symmetric interval about 1/2 of the form $[a, 1-a]$, with $a \in (0, 1/2)$. For a class of loss functions, which include the important cases of squared error and information-normalized losses, we investigate conditions for which the Bayes estimator, δ_{BU}, with respect to a symmetric prior concentrated on the end points of the parameter space is minimax. Our conditions are of the form $1 - 2a \leq c(n)$ with $c(n) = O(n^{-1/2})$, and various analytical evaluations, lower and upper bounds, and numerical evaluations are given for $c(n)$. For instance, the simple condition $1 - 2a \leq \frac{1}{\sqrt{2n}}$ guarantees, for all $n \geq 1$, the minimaxity of δ_{BU} under both squared error and information-normalized losses.

AMS 2000 subject classifications. 62F10, 62F15, 62F30 62F99

Key words and phrases: Minimax estimation, restricted parameter space, binomial distribution, squared error loss, information-normalized loss

[1] Part of this work appears in Rokhaya Gueye's Ph.D. thesis: *Problème d'estimation de paramètres avec restriction sur l'espace des paramètres*, Université de Montréal, 2004.

1. Introduction

In many statistical problems, there exists bounds on the values that unknown parameters can take. Accordingly, a large body of work concerned with estimation problems in restricted parameter spaces has emerged, as reviewed for instance by van Eeden (1996) or Marchand and Strawderman (2004). One influential, much studied, and useful criterion to select or to evaluate a procedure is minimaxity (e.g., Brown, 1994, Strawderman, 2000).

Inferential problems arising from binomial experiments is one of the classical problems in statistics offering many challenges. As an example indeed, the unexpected difficulties inherent to the choice of an interval estimate of a binomial proportion, and the relative inefficiency of the "standard" Wald confidence interval, has resurfaced recently with the work of Brown, Cai and DasGupta (2001), and Agresti and Coull (1998) among others.

We are concerned here with the problem of minimax estimation of a Binomial proportion p, where p is known to lie in a symmetric interval about $1/2$ of the form $[a, 1-a]$ with $0 < a < 1/2$. Not only does the constraint arise in Warner's (1965) random response model, but it also represents along with upper bound constraints of the form $p \leq b$, important test cases for the study of the estimation of a constrained binomial proportion. Marchand and MacGibbon's (2000) work is concerned with a general constraint for a Binomial proportion p and, for the particular type of constraint studied in this paper, they provide numerical minimax and linear minimax solutions estimators for both squared error and information-normalized loss functions. Recently, Braess and Dette (2004) show that these linear minimax solutions are also asymptotically minimax for squared error, information-normalized and entropy loss functions. Their results also apply to more general types of constraints on the proportion p. Earlier analytical and numerical solutions for squared error loss were given by Berry (1989) and Moors (1985). Related work from a Gamma-minimax perspective appears in Lehn and Rummel (1987).

It follows here from the work of DasGupta (1985) and invariance considerations that the least favourable prior is, for squared error loss, uniform on the boundary $\{a, 1-a\}$ for small enough interval length $1-2a$. This paper contributes to the analytical determination of simple sufficient conditions, of the form $1-2a \le c(n)$, for which the Bayes estimator δ_{BU} with respect to the 2-point uniform prior on $\{a, 1-a\}$ is minimax. The development below addresses both the important cases of squared error and information-normalized losses (among a quite vast class of weighted squared error losses) and is applicable and unified for all $n \ge 1$. This is of particular interest given the difficulties in working with the binomial probability function, as well as with various loss functions. For instance, we establish below the simple condition

$$1 - 2a \le \frac{1}{\sqrt{2n}}$$

for the minimaxity of δ_{BU} under both squared error (see part (b) of Corollary 2) and information-normalized (see Section 3.2) losses. Numerical evaluations of larger cutoff points $c(n)$ for minimaxity, as well as sufficient conditions for the non-minimaxity of δ_{BU} of the form $1-2a \ge c(n)$ are also given.

In related work, Perron (2003) investigates dominating estimators of the maximum likelihood estimator δ_{mle}. In particular, he establishes the sufficiency of the condition $1-2a \le \frac{1}{\sqrt{n}}$ for δ_{BU} to dominate δ_{mle} under squared error loss. Apart from addressing this particular estimation problem and providing various other dominating estimators, Perron (2003) also illustrates, because of the common symmetries, how methods introduced by Marchand and Perron (2001) for handling the estimation of a constrained multivariate normal mean can be used for this discrete Binomial model problem, yielding striking parallels as well in the conditions of dominance of the maximum likelihood procedure.

Section 2 contains preliminary results and definitions, some of which are borrowed from Perron (2003). Section 3 addresses conditions related to the minimaxity of δ_{BU} for both squared error and information-normalized losses.

2. Definitions and Preliminaries

Originally expressed in terms of (X, p), it is convenient, as in Perron (2003), to express our model in terms of (R, S, θ, λ) where $R = 2\sqrt{n}\,|\frac{X}{n} - \frac{1}{2}|$, $S = \mathrm{sgn}(\frac{X}{n} - \frac{1}{2})$, $\theta = 2\sqrt{n}\,(p - \frac{1}{2})$, and $\lambda = |\theta|$. Correspondingly, our constraint $p \in [a, 1-a]$ can be expressed as $\theta \in \Theta(m) = \{\theta : \lambda \leq m\}$; with $m = \sqrt{n}(1-2a)$. All the estimators considered below are symmetric; i.e. $\delta(n-x) = 1 - \delta(x)$ for $x = 0, 1, \ldots, n$; and any such estimator may be parametrized by a function g as:

$$\delta_g(x) = (1 + g(r)\frac{ms}{\sqrt{n}})/2\ .$$

We will study weighted and unweighted squared error loss functions of the form $L_w(\delta, p) = w(\lambda^2)(\delta - p)^2$; which will include the important cases of squared error ($w(t) = 1$) and information-normalized losses ($w(t) \propto (n-t)^{-1}$). The risk $\mathrm{R}(\theta, \delta_g) = w(\lambda^2)\mathrm{E}_\theta[\,(\delta_g(X) - p)^2]$ of a symmetric estimator δ_g becomes with some simplifications:

$$\mathrm{R}(\theta, \delta_g) = \frac{w(\lambda^2)}{4n}\mathrm{E}_\lambda\{m^2 g^2(R)\ + \lambda^2 - 2m\lambda g(R)\ E_\lambda[S|R]\}. \tag{1}$$

The next definitions and preparatory results relate to this above risk expression, namely to the distribution of R, and to the quantities $E_\lambda[S|R=r]$; $r > 0$.

Definition 1. *For $n \geq 1$, $\mathcal{R}_n = \{\frac{n-2k}{\sqrt{n}} : k \text{ integer}, 0 \leq k \leq n/2\}$.*

Lemma 1. *(Perron, 2003) The probability mass function $f_n(\lambda, \cdot)$ of R is supported on $\mathcal{R}_n$ and is given by:*

$$f_n(\lambda, r) = \begin{cases} 2^{-n}\binom{n}{n/2}(1-\lambda^2/n)^{n/2} & \text{if } r = 0 \\ 2^{-n}\binom{n}{(n+\sqrt{n}r)/2}(1-\lambda^2/n)^{n/2} 2\cosh(\tanh^{-1}(\frac{\lambda}{\sqrt{n}})\ \sqrt{n}\,r) & \text{if } r > 0. \end{cases}$$

Furthermore, the family of distributions of R has increasing monotone likelihood ratio (MLR) in R, with $\lambda \in [0, \sqrt{n})$ viewed as the parameter.

Definition 2. *For $(\lambda, r) \in [0, \sqrt{n}) \times \mathcal{R}_n$,*

$$\rho_n(\lambda, r) = \tanh(\tanh^{-1}(\frac{\lambda}{\sqrt{n}})\ \sqrt{n}\,r).$$

Lemma 2. *For $(\lambda, r) \in [0, \sqrt{n}) \times \mathcal{R}_n$, we have $E_\lambda[S|R = r] = \rho_n(\lambda, r)$; and the Bayes estimator δ_{BU}, with respect to the uniform prior (for p) on $\{a = \frac{1}{2}(1 - m/\sqrt{n}), 1 - a = \frac{1}{2}(1 + m/\sqrt{n})\}$, is given by*

$$\delta_{BU}(x) = (1 + \rho_n(m, r)\frac{ms}{\sqrt{n}})/2.$$

Moreover,

(a) For $r \in \mathcal{R}_n, r \neq 0$, $\rho_n(\cdot, r)$ is increasing, with $\rho_n(0, r) = 0$, and $\rho_n(\lambda, r) \to 1$ as $\lambda \to \sqrt{n}$.

(b) If $0 \leq \lambda < \sqrt{n}$, and $r \geq 0$, then $\rho_n(\lambda, 0) = 0$, $\rho_n(\lambda, \cdot)$ is increasing and concave, $\frac{\rho_n(\lambda,r)}{r}$ decreases in r, $\rho_n(\lambda, r) \leq \lambda r$ for all $r \in \mathcal{R}_n$, and $\rho_n(\lambda, r) \leq \lambda r/(1 + \lambda^2/n)$ for all $r \in \mathcal{R}_n$ with n even.

Lemma 3. *(a)* $\frac{\partial}{\partial\lambda}\rho_n(\lambda, r) = \frac{1}{(1-\lambda^2/n)} r[1 - \rho_n^2(\lambda, r)]$,

(b) $\frac{\partial}{\partial\lambda} \log(f_n(\lambda, r)) = \frac{1}{(1-\lambda^2/n)}[r\rho_n(\lambda, r) - \lambda]$,

(c) $E_\lambda[R\rho_n(\lambda, R)] = \lambda$,

(d) $E_\lambda[R^2] = \lambda^2 + (1 - \lambda^2/n)$,

(e) $E_\lambda[g(R)(R\rho_n(\lambda, R) - \lambda)] = Cov_\lambda[g(R), R\rho_n(\lambda, R)]$,

(f) $E_\lambda[Rh(R)(R - \lambda\rho_n(\lambda, R))] = Cov_\lambda[h(R), R(R - \lambda\rho_n(\lambda, R))] + (1 - \lambda^2/n)E_\lambda[h(R)]$,

Proof. Parts (a) and (b) are obtained by direct computations. Since $0 = \mathrm{E}_\lambda[\frac{\partial}{\partial\lambda} f_n(\lambda, R)]$, the proof of part (c) is immediate.[2] The proof of part (d) is also straightforward as $R^2 = n(2\bar{X} - 1)^2$ and $\mathrm{E}_\lambda[R^2] = 4n\mathrm{Var}_\lambda[\bar{X}] + n\mathrm{E}^2_\lambda[2\bar{X} - 1]$. Finally, parts (e) and (f) are direct applications of the results in parts (c) and (d). $\square$

Concerning Lemma 2, without giving details most of which are available in Perron (2003), it is nevertheless useful to point out that the given properties of $\rho_n(\lambda, r)$ are essentially governed by the properties of the

[2] Observe also that $\mathrm{E}_\lambda[R\rho_n(\lambda, R)] = \mathrm{E}_\lambda[RS] = \mathrm{E}_\lambda[\sqrt{n}\ (2\frac{X}{n} - 1)] = \lambda$ directly.

tanh function. Namely the upper bounds for $\rho_n(\lambda, r)$ are obtained by Definition 2's representation with the evaluations $\tanh(\tanh^{-1}(x)) = x$ and $\tanh(2\tanh^{-1}(x)) = \frac{2x}{1+x^2}$.

3. Minimaxity Results

Recalling a familiar criterion for minimaxity, the estimator δ_{BU} will be minimax if the supremum of its risk function $R(\theta, \delta_{BU})$ is attained on the support of the associated prior, i.e., $\{-m, m\}$. In particular, given the symmetry of $R(\theta, \delta_{BU})$ in θ about 0, it will suffice to show that $R(\theta, \delta_{BU})$ is increasing in $\lambda = |\theta|; \lambda \leq m$.

3.1. *Squared error loss*

We begin with squared error loss and we first establish conditions for which the risk $R(\theta, \delta_g)$ of an estimator δ_g increases in $|\theta|$; a property which is also of independent interest.

Theorem 1. *Consider g, a positive and nondecreasing function on $\mathcal{R}_n$. Define h as $h(r) = g(r)/r$ for $r \in \mathcal{R}_n$, $r \neq 0$ and let $h(0) = h(2/\sqrt{n})$ if $0 \in \mathcal{R}_n$. Assume that h is nonincreasing on $\mathcal{R}_n$. If $E_0[h(R)] \leq 1/2m$, then the risk $R(\theta, \delta_g)$, under squared error loss, is nondecreasing in λ, for $\lambda \in [0, m \wedge 1]$.*

Proof. First, risk expression (1) and Lemma 2 give us: $\mathrm{R}(\theta, \delta_g) = \frac{1}{4n}\,\mathrm{E}_\lambda[m^2 g^2(R) - 2m\lambda g(R)\rho_n(\lambda, R) + \lambda^2]$. Then, we have the following expansion:

$$\begin{aligned}
4n\frac{\partial}{\partial\lambda}\mathrm{R}(\theta, \delta_g) &= \frac{\partial}{\partial\lambda}\mathrm{E}_\lambda[m^2 g^2(R) - 2m\lambda g(R)\rho_n(\lambda, R) + \lambda^2] \\
&= \mathrm{E}_\lambda[\frac{\partial}{\partial\lambda}\{m^2 g^2(R) - 2m\lambda g(R)\rho_n(\lambda, R) + \lambda^2\}] \\
&+ \mathrm{E}_\lambda[\{m^2 g^2(R) - 2m\lambda g(R)\rho_n(\lambda, R) + \lambda^2\}\frac{\partial}{\partial\lambda}\log(f_n(\lambda, R))] \\
&= 2\mathrm{E}_\lambda[\lambda - m g(R)\rho_n(\lambda, R)] \\
&+ \frac{1}{(1-\lambda^2/n)}\mathrm{E}_\lambda[(m^2 g^2(R) + \lambda^2)(R\rho_n(\lambda, R) - \lambda) - 2m\lambda g(R)(R - \lambda\rho_n(\lambda, R))]
\end{aligned}$$

$$
\begin{aligned}
&= 2\lambda \mathrm{E}_\lambda[1-2mh(R)] - 2m\mathrm{Cov}_\lambda[h(R), R\rho_n(\lambda,R))] \\
&+ \frac{1}{(1-\lambda^2/n)}\{\mathrm{Cov}_\lambda[m^2g^2(R), R\rho_n(\lambda,R)] - 2m\lambda\mathrm{Cov}_\lambda[h(R), R^2-\lambda R\rho_n(\lambda,R))]\} \\
&\geq 2\lambda \mathrm{E}_\lambda[1-2mh(R)] \\
&\geq 2\lambda \mathrm{E}_0[1-2mh(R)] \\
&\geq 0.
\end{aligned}
$$

Here are some details regarding this series of equalities and inequalities. The third equality comes from Lemma 3 parts (a) and (b), while the fourth equality comes from Lemma 3 parts (e) and (f). The first inequality holds as each one of the expressions: $-h(r)$, $g^2(r)$, $\rho_n(\lambda,r)$, $r\rho_n(\lambda,r)$ and $r^2-\lambda r\rho_n(\lambda,r)$ are nondecreasing in r, $r\in\mathcal{R}_n$. To see this, observe first that Lemma 2 and the assumptions on g and h imply that the first four given expressions are nondecreasing in r for all $r\in\mathcal{R}_n$. For the last expression, write for $r\neq 0$, $r^2-\lambda r\rho_n(\lambda,r) = r^2(1-\frac{\lambda\rho_n(\lambda,r)}{r})$; and use Lemma 2 to infer that $r^2-\lambda r\rho_n(\lambda,r)$ is expressible as the product of two positive increasing functions given that $(1-\frac{\lambda\rho_n(\lambda,r)}{r}) \geq 1-\lambda^2 \geq 0$, for $\lambda\leq 1$. Finally, the next to last inequality comes from the monotone likelihood ratio property of Lemma 1 given that h is nonincreasing, while the last inequality holds by assumption. □

Corollary 1. *Consider g, a positive and nondecreasing function on $\mathcal{R}_n$. Define h as $h(r) = g(r)/r$ for $r\in\mathcal{R}_n$, $r\neq 0$ and let $h(0)=h(2/\sqrt{n})$ if $0\in\mathcal{R}_n$. Assume that h is nonincreasing on $\mathcal{R}_n$. Let $r_0=\min\{r: r\in\mathcal{R}_n\}$. If $m\leq 1/2h(r_0)$, then the risk $R(\theta,\delta_g)$,under squared error loss, is a nondecreasing expression in λ for $\lambda\in[0, m\wedge 1]$.*

Proof. The result follows from Theorem 1 as the nonincreasing property of h on $\mathcal{R}_n$ implies that $\mathrm{E}_0[h(R)]\leq h(r_0)$. □

We now apply the above in order to obtain sufficient conditions for the minimaxity of δ_{BU}. In doing so, we define hereafter h_{BU} such that $h_{BU}(m,r)=\rho_n(m,r)/r$ for all $r\in\mathcal{R}_n$, $r\neq 0$ and $h_{BU}(m,0) = h_{BU}(m,2/\sqrt{n})$.

Corollary 2. *Under squared error loss,*

(a) If $E_0[h_{BU}(m,R)] \leq 1/2m$ and $m \leq 1$, then $R(\theta, \delta_{BU})$ is nondecreasing in λ, and δ_{BU} is minimax;

(b) If $m \leq 1/\sqrt{2}$, δ_{BU} is minimax;

(c) If $m \leq 1/\sqrt{2-\frac{1}{n}}$ and n is even, then δ_{BU} is minimax.

Proof. Part (a) follows directly from Theorem 1, while parts (b) and (c) follow from Corollary 1 with the evaluations $h_{BU}(m, 1/\sqrt{n}) = m$ and $h_{BU}(m, 2/\sqrt{n}) = \frac{m}{1+m^2/n}$. $\square$

Consider the inequality $\mathrm{E}_0[h_{BU}(m,R)] \leq 1/2m$ of Corollary 2, part (a). Given Lemma 2's properties, we have that $\mathrm{E}_0[2mh_{BU}(m,R)]$ increases in m, with $\lim_{m\to\sqrt{n}} \mathrm{E}_0[2mh_{BU}(m,R)] = 2\sqrt{n} > 1$, which shows that sufficient condition (a) of Corollary 2 is, for cases where $m \leq 1$, equivalent to $m \leq m_0(n)$, with $2m_0(n)\mathrm{E}_0[h_{BU}(m_0(n),R)] = 1$. Observe also that Corollary 2's cutoff points $1/\sqrt{2}$ and $1/\sqrt{2-\frac{1}{n}}$ are lower bounds for $m_0(n)$. We pursue with a large sample analysis of $m_0(n)$, an upper bound, and numerical evaluations.

Lemma 4. *As $n \to \infty$, we have $m_0(n) \to \kappa_0 \approx 0.755233$, where κ_0 is the solution in m to the equation*

$$E[\tanh(mZ)/Z] = 1/2m,$$

for Z distributed according to a standard normal distribution.

Proof. Assume that m and r are fixed. Observe that

$$\tanh^{-1}(\frac{m}{\sqrt{n}})\,\sqrt{n}\,r \to mr \qquad \text{as } n \to \infty,$$

and that $\sqrt{n}(2\bar{X}-1)$ has a limiting standard normal distribution. Therefore,

$$\mathrm{E}_0[\frac{1}{R}\tanh(\tanh^{-1}(\frac{m}{\sqrt{n}})\,\sqrt{n}\,R)] \to \mathrm{E}[\tanh(mZ)/Z], \qquad \text{as } n \to \infty,$$

establishing the result. $\square$

Values of $m_0(n)$, which are seen to converge rapidly to their limiting value $\kappa_0 \approx 0.755233$, are presented in Figure 1 along with the lower envelopes $1/\sqrt{2}$ and $1/\sqrt{2-\frac{1}{n}}$, for n odd and n even respectively. An upper envelope $m_1(n)$, whose derivation is given below, also appears. It is clear from part (a) of Corollary 2 that values of $m \in \{m : m \leq m_0(n) \wedge 1\}$ are necessarily values of m for which $\mathrm{R}(0, \delta_{BU}) \leq \mathrm{R}(m, \delta_{BU})$. Now using (1), Lemma 2's properties of $\rho_n(m, r)$, and the monotone likelihood ratio property of Lemma 1, we obtain:

$$\mathrm{R}(0, \delta_{BU}) \leq \mathrm{R}(m, \delta_{BU}) \Leftrightarrow E_0[\rho^2(m, R)] + E_m[\rho^2(m, R)] \leq 1 \Leftrightarrow m \leq m_1(n);$$

where $E_0[\rho^2(m_1(n), R)] + E_{m_1(n)}[\rho^2(m_1(n), R)] \leq 1$. With $m_0(n)$ not fitting tightly to $m_1(n)$, the real interest in the values $m_1(n)$ however lies in our conjecture that $\delta_{BU}(X)$ is minimax if and only if $m \leq m_1(n)$, for all $n \geq 1$. The conjecture arises with numerical evidence that suggests that $\frac{\partial}{\partial\theta}\mathrm{R}(\theta, \delta_{BU})$ has, for $\theta \in [0, m]$, at most one sign change from $-$ to $+$. Such a behaviour would match as well the asymptotic behaviour (in n) of $\mathrm{R}(\theta, \delta_{BU})$, which is governed by properties relative to a normal model $N(\theta, 1)$ with $|\theta| \leq m$, as established by Casella and Strawderman (1981) or Berry (1990). Finally, as in Lemma 4 and independently of the conjecture, we have $m_1(n) \to \kappa_1 \approx 1.0567$ as $n \to \infty$, and the heuristic $\delta_{BU}(X)$ is minimax if and only if $m \leq \kappa_2$, with $\kappa_2 \approx \kappa_1$ for large n.

3.2. *Information-normalized and other weighted squared error losses*

Here, we consider loss functions of the form:

$$L_w(\delta, p) = w(\lambda^2)(\delta - p)^2, \tag{2}$$

with $w > 0$ and $w(\cdot)$ nondecreasing on $[0, m^2]$. These losses penalize deviations $|\delta - p|$ more heavily as p moves away from $1/2$. Let R_w be the risk function associated with L_w. A first key observation is that, if the prior is supported on the boundary of the parameter space, implying that $w(\lambda^2)$ is constant with probability one, then loss functions L_w produce the same Bayes estimators! Along with previous conditions for which $R(\theta, \delta_{BU})$ increases in λ, the following is then immediate for losses L_w given that w is nondecreasing.

Lemma 5. *If δ_{BU} is minimax under squared error loss with $\sup\{R(\theta, \delta_{BU}), |\theta| \le m\} = R(m, \delta_{BU})$, then δ_{BU} is minimax under any loss L_w as in (2).*

From the above, we can immediately affirm that Corollary 2 applies simultaneously for all losses L_w in (2). We now pursue with a somewhat finer analysis leading to analogues of Theorem 1 and part (a) of Corollary 2 for a general class of weights w in (2), who are further described in Theorem 2 (but also see Remark 1), and which arise as solutions of a certain differential inequation. An important subclass consists of weights of the form $w_c(t) = (c-t)^{-1}$; $t \in (0, m^2)$; with $c > m^2$. These weights lead, by expressing $w_c(\lambda^2)$ in terms of p, to losses $L_{w_c}(\delta, p) = (\delta - p)^2/\{(c-n) + 4np(1-p)\}$, and in particular the choice $c = n$ (i.e., w_n) corresponds to the case of information-normalized loss for our Binomial problem.

Theorem 2. *Consider g, a positive and nondecreasing function on $\mathcal{R}_n$. Define h as $h(r) = g(r)/r$ for $r \in \mathcal{R}_n$, $r \ne 0$ and let $h(0) = h(2/\sqrt{n})$ if $0 \in \mathcal{R}_n$. Assume that h is nonincreasing on $\mathcal{R}_n$, that $w > 0$, w is nondecreasing, and that $w(t) \ge (c-t)w'(t)$ for all $t \in [0, m^2]$, and for some $c > m^2$. If $E_0[1 - 2mh(R) + m^2 g^2(R)/c] \ge 0$, then the risk $R_w(\theta, \delta_g)$, under loss L_w, is a nondecreasing expression in λ for $\lambda \in [0, m \wedge 1]$.*

Proof. We have

$$4n \frac{\partial}{\partial \lambda} \mathrm{R}_w(\theta, \delta_g) = 4nw(\lambda^2) \frac{\partial}{\partial \lambda} \mathrm{R}(\theta, \delta_g) + 8n\lambda w'(\lambda^2) \mathrm{R}(\theta, \delta_g).$$

First, within the proof of Theorem 1 we learn that

$$4n\frac{\partial}{\partial\lambda}\mathrm{R}(\theta,\delta_g) \geq 2\lambda\mathrm{E}_0[1-2mh(R)].$$

Next, using (1) and similar arguments to those in the proof of Theorem 1, we obtain that

$$\begin{aligned}
8n\mathrm{R}(\theta,\delta_g) &= 2\mathrm{E}_\lambda[m^2g^2(R) - 2m\lambda g(R)\rho_n(\lambda,R) + \lambda^2] \\
&= 2\mathrm{E}_\lambda[m^2g^2(R) + \lambda^2(1-2mh(R))] - 4m\lambda\mathrm{Cov}_\lambda[h(R), R\rho_n(\lambda,R))] \\
&\geq 2\mathrm{E}_0[m^2g^2(R) + \lambda^2(1-2mh(R))].
\end{aligned}$$

Finally, since $w > 0$, $w' \geq 0$, and w is a solution of the differential inequation $w(t) + tw'(t) \geq cw'(t)$, we infer that

$$\begin{aligned}
4n\frac{\partial}{\partial\lambda}\mathrm{R}_w(\theta,\delta_g) &\geq 2\lambda w(\lambda^2)\mathrm{E}_0[1-2mh(R)] + 2\lambda w'(\lambda^2)\mathrm{E}_0[m^2g^2(R) + \lambda^2(1-2mh(R))] \\
&= 2\lambda\mathrm{E}_0[(w(\lambda^2) + \lambda^2 w'(\lambda^2))(1-2mh(R)) + w'(\lambda^2))m^2g^2(R)] \\
&\geq 2\lambda w'(\lambda^2)\mathrm{E}_0[c(1-2mh(R)) + m^2g^2(R)] \\
&\geq 0,
\end{aligned}$$

under the assumptions of the Theorem. □

Remark 1. *It is easily checked that weights of the form* $w(t) = \frac{k(c-t)}{c-t}$*, with increasing* $k(y)$*, decreasing* $\frac{k(y)}{y}$ *(or even concave* k *with* $k(0) = 0$*) satisfy the conditions of the Theorem. This provides a possibly more convenient representation of weights for which Theorem 2 applies. Now, the cases where* $k(.)$ *is constant are such weights, and the object of the next result.*

Corollary 3. *Under loss* $L_{w_c}(\delta,p) = \frac{(\delta-p)^2}{c-\theta^2}$ *with* $c > m^2$*,* δ_{BU} *is minimax whenever* $m \leq (1 \wedge m_0(c,n))$*, where* $m_0(c,n)$ *is the solution in* m *of the equation:*

$$E_0[1 - 2mh_{BU}(m,R) + \frac{m^2}{c}\rho_n^2(m,R)] = 0.$$

Proof. Theorem 2 tells us that δ_{BU} is minimax for

$$m \in \{m : \mathrm{E}_0[1 - 2mh_{BU}(m,R) + \frac{m^2}{c}\rho_n^2(m,R)] \geq 0, \text{ and } m \leq 1\}.$$

Hence, our result here will be established if we can show for instance that $\mathrm{E}_0[\frac{m^2}{c}\rho_n^2(m,R) - 2mh_{BU}(m,R)]$ decreases in $m; m \leq 1$. Now, by using part (a) of Lemma 3, we obtain after some simplifications (for $r > 0$):

$$\frac{\partial}{\partial m}\{\frac{m^2}{c}\rho_n^2(m,r) - 2m\frac{\rho_n(m,r)}{r}\} = 2\{\frac{m(1-\rho_n^2(m,r))}{1-\frac{m^2}{n}} + \frac{\rho_n(m,r)}{r}\}\,\{\frac{mr\rho_n(m,r)}{c} - 1\}.$$

Since the quantities $1 - \rho_n^2(m,r)$, $\frac{\rho_n(m,r)}{r}$, and $-r\rho_n(m,r)$ are decreasing in r (refer to Lemma 2), we have the inequality

$$\frac{\partial}{\partial m}\{E_0[\frac{m^2}{c}\rho_n^2(m,R) - 2mh_{BU}(m,R)]\} = 2\{E_0[\frac{m(1-\rho_n^2(m,R))}{1-\frac{m^2}{n}} + \frac{\rho_n(m,R)}{R}]\}\,\{E_0[\frac{mR\rho_n(m,R)}{c} - 1]\}.$$

Finally, the result follows with the MLR property of Lemma 1, the condition $c > m^2$, and part (c) of Lemma 3, which imply $E_0[\frac{mR\rho_n(m,R)}{c} - 1] \leq E_m[\frac{mR\rho_n(m,R)}{c} - 1] = E_m[\frac{m^2}{c} - 1] \leq 0$. □

We conclude by referring to the values of $m_2(n) = m_0(n,n)$ of Figure 1 and Corollary 3, which are cutoff points for which smaller values of m imply the minimaxity of δ_{BU} under information-normalized loss. As expected, values $m_2(n)$ lie above values $m_0(n)$, but the differences are hardly noticeable for large n as explained by the similarity of the losses for large n, i.e, $L_{w_n}(\delta,p) = \frac{(\delta-p)^2}{n-\lambda^2} \approx \frac{(\delta-p)^2}{n}$, for large n and $\lambda \leq m \leq 1$.

Acknowledgements

The work of Marchand and Perron is partially supported by NSERC of Canada. The authors gratefully acknowledge the assistance of Claude Beauchamp for some of the numerical work.

References

Agresti, A. and Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician*, **52**, 119-126.

Berry, C. (1989). Bayes minimax estimation of a Bernouilli p in a restricted parameter space. *Communications in Statistics: Theory and Methods*, **18**, 4607-4616.

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

Braess, D. and Dette, H. (2004). The asymptotic minimax risk for the estimation of constrained binomial and multinomial probabilities. To appear in *Sankhyā*.

Brown, L.D., Cai, T. and DasGupta, A. (2001). Interval estimation for a binomial proportion (with comments and rejoinder). *Statistical Science*, **16**, 101-133.

Casella, G. and Strawderman, W.E. (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 Ser. A*. **47**, 326-332.

Lehn, J. and Rummel, F. (1987). Gamma-minimax estimation of a binomial probability under squared error loss. *Statistics & Decisions*, **5**, 229-249.

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

Marchand, É. and Perron, F. (2001). Improving on the mle of a bounded normal mean. *The Annals of Statistics.* **29**, 1066-1081.

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. 21-44.

Moors, J.J.A. (1985). *Estimation in truncated parameter spaces.* Ph.D. thesis, Tilburg University.

Perron, F. (2003). Improving on the mle of p for a Binomial(n, p) when p is around $1/2$. *Festschrift for Constance van Eeden*, IMS Lecture Notes-Monograph Series, **42**, pp. 45-64.

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

van Eeden, C. (1996). Estimation in restricted parameter spaces- some history and recent developments. *CWI Quarterly*, **9**, 69-76.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association.* **62**, 63-69.