IMPROVING ON THE MLE OF p FOR A BINOMIAL(n,p) WHEN p IS AROUND $\frac{1}{2}$

François Perron
Université de Montréal

We consider the problem of estimating the parameter p of a binomial (n,p) distribution when p lies in a symmetric interval of length $m/\sqrt{n}$, $m < \sqrt{n}$. We establish sufficient conditions for the domination of the maximum likelihood estimator with quadratic loss. We suggest three other estimators for the estimation of p. The first two dominate the maximum likelihood estimator. The first one comes from Moors (1985) and corresponds to the bayesian estimator with respect to the symmetric prior concentrated on the end points if and only if $m \leq 1$ when n is odd or $m \leq \sqrt{n/(n-1)}$ when n is even. The second estimator comes from Charras and van Eeden (1991); it is in fact the maximum likelihood estimator for the problem where m is replaced by m_0, $0 < m_0 < m$. We give an algorithm for the selection of m_0. The third is the Bayes estimator with respect to the prior having a density proportional to $(p(1-p))^{-1}$. This estimator dominates the maximum likelihood estimator for some values of (n,m) but not for all of them. We give simple sufficient conditions for the domination of the Bayes estimator over the maximum likelihood estimator. It is clear that the maximum likelihood estimator is inappropriate when either n or m is small. When n is large, all of the estimators have approximately the same behaviour except for the last. Numerical evaluations illustrate our comments.

1. Introduction

Assume that the statistician observes x, the realisation of X, a binomial(n,p) random variable. Consider the problem of estimating the proportion p with quadratic loss when p lies in a symmetric interval around $\frac{1}{2}$. In many situations, prior knowledge tells us that this symmetric interval has length less than 1 simply because successes and failures are not rare events. For example, in an effort to protect the privacy of the respondant, Warner (1965) has developed a method where one is interested in the estimation of π and $p = \pi P + (1-\pi)(1-P)$. In his setup, P is known, $\frac{1}{2} < P < 1$ and the distribution of X is a binomial(n,p), therefore, $1 - P \leq p \leq P$. In the following, m will be fixed, $0 < m < \sqrt{n}$, and we shall set $p = (1 + \theta/\sqrt{n})/2$, $\Theta(m) = \{\theta \in \mathbb{R} : \|\theta\| \leq m\}$.

In this problem, the maximum likelihood estimator δ_{mle} is the truncation of the empirical proportion (x/n) on the parameter space. It is given by $\delta_{\text{mle}}(x) = \{1 + [|2x/n - 1| \wedge (m/\sqrt{n})] \operatorname{sgn}(2x/n - 1)\}/2$. This estimator is inadmissible because it takes values on the boundary of the parameter space, (see, Sacks, 1963, DasGupta 1985 or Charras and van Eeden, 1991). Actually, Charras and van Eeden (1991) specifically treat our problem in their Example 5.2. They propose that we modify the maximum likelihood

AMS subject classifications: 62F10, 62F15, 62F30.

estimator when it touches a boundary point. They mention that for some constant m_0, $0 < m_0 < m$, the maximum likelihood estimator in the problem $|\theta| \leq m_0$ dominates δ_{mle}. In this paper, we shall provide an algorithm for the selection of m_0 and the corresponding estimator will be called δ_{cve}. In his thesis, Moors (1985) studied this problem in great detail. He found that, for each value of x, $\delta(x)$ must belong to a certain closed interval. If $\delta(x_0)$ is not in the interval then it is preferable to replace $\delta(x_0)$ by the closest end point in the corresponding interval. This modified version generates a better estimator. We apply this technique to δ_{mle}, that gives us the estimator δ_{mrs}. Finally, Marchand and MacGibbon (2000) have also worked on the present problem in the very special cases $n = 1$ and $n = 2$.

The strategy behind the works of Charras and van Eeden (1991) and Moors (1985) consists in having a criterion such that when this criterion is not met a correction is proposed. These approaches are oriented towards finding a complete class. They are not helpful in verifying if another estimator dominates the maximum likelihood estimator. For instance, it is easy to show that a Bayes estimator with respect to a symmetric prior distribution will always satisfy their criterion but we still do not know if this Bayes estimator dominates δ_{mle}. A series of sufficient conditions for the domination of δ_{mle} will be established in Section 3. These conditions depend strongly on the binomial distribution. In Section 2, we shall analyse some properties for this distribution. Many of the results in Setions 2 and 3 are inspired by the methodology used in Marchand and Perron's (2001) paper. Our Corollary 3.3 of Section 3 says that if m is small enough then any symmetric estimator (i.e., $\delta(n-x) = 1 - \delta(x)$) such that δ shrinks δ_{mle} towards $\frac{1}{2}$ will dominate δ_{mle}. Our Corollary 3.2 of Section 3 says that if δ satisfies some regularity conditions and δ is not far from δ_{mle} then δ dominates δ_{mle}. Many Bayes estimator will satisfy these regularity conditions. We study the one corresponding to a prior density proportional to $(p(1-p))^{-1}$ called δ_{bys}. Numerical results are given in Section 4. We provide some evaluations for m_0. We compare graphically the risk function of δ_{mle}, δ_{mrs}, δ_{cve} and δ_{bys}.

2. Definitions and Preliminaries

Let

$$\bar{X} = X/n, \quad S = \text{sgn}(2\bar{X} - 1), \quad R = \sqrt{n}|2\bar{X} - 1|,$$
$$p = (1 + \theta/\sqrt{n})/2, \quad \Theta(m) = \{\theta : |\theta| \leq m\}$$

and

$$\lambda = |\theta|, \quad c(n, \lambda) = \log(1 + \lambda/\sqrt{n}) - \log(1 - \lambda/\sqrt{n}).$$

All the estimators considered below are symmetric, i.e., $\delta(n - x) = 1 - \delta(x)$ for $x = 0, 1, \ldots, n$. Any symmetric estimator can be parametrized with a function g via the following formula:

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

For instance, the maximum likelihood estimator corresponds to $g(r) = 1 \wedge r/m$. The quadratic loss is used to obtain the risk, i.e., $\mathrm{R}(\theta, \delta_g) = \mathrm{E}_\theta[(\delta_g(X) - p)^2]$. After simplifications we obtain $\mathrm{R}(\theta, \delta_g) = \mathrm{E}_\theta[(g(R)mS - \theta)^2]/4n$. Our dominance results are based on conditional risk decompositions. In short, if we partition the sample space into $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ say, and we show that $\mathrm{E}_\theta[(\delta_{\mathrm{mle}}(X) - \theta)^2 - (\delta(X) - \theta)^2 | \mathcal{A}_i] \geq 0$ for all $\theta \in \Theta(m)$, $i = 1, 2, \ldots, k$ then $\mathrm{R}(\theta, \delta_{\mathrm{mle}}) \geq \mathrm{R}(\theta, \delta)$ for all $\theta \in \Theta(m)$. In the following, the partitions will depend on the statistic R. Therefore, it is important to analyse the mass function of R and the conditional expectation of S given that R is fixed.

Lemma 2.1. *Let* $0 \leq \lambda < \sqrt{n}$. *The probability mass function* $f_n(\lambda, \cdot)$ *of* R *is defined on* $\mathcal{R}_n$ *with* $\mathcal{R}_n = \{\sqrt{n}(2k/n - 1) : n/2 \leq k \leq n\}$. *For* $r \in \mathcal{R}_n$, *it is given by:*

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

and the family of distributions has monotone likelihood ratio. Moreover, if $r_1, r_2 \in \mathcal{R}_n$, $r_1 < r_2$ *then* $f_n(\lambda, r_2)/f_n(\lambda, r_1)$ *is nondecreasing in* λ.

Proof. The derivation of $f_n(\lambda, \cdot)$ is direct. It is easy to see that the nondecreasing property in t of the expression $\cosh(at)/\cosh(bt)$ for all $0 \leq b < a$ implies that the ratio $f_n(\lambda_2, r)/f_n(\lambda_1, r)$ is nondecreasing in r for all $0 \leq \lambda_1 < \lambda_2 < \sqrt{n}$, i.e., the family of distributions has monotone likelihood ratio. In fact,

$$\frac{\partial}{\partial t}\left\{\frac{\cosh(at)}{\cosh(bt)}\right\} = \frac{(a-b)}{2}\frac{\sinh((a+b)t)}{\cosh^2(bt)} + \frac{(a+b)}{2}\frac{\sinh((a-b)t)}{\cosh^2(bt)} \geq 0$$

for all t, $0 \leq b < a$. If $\{r, r + 2/\sqrt{n}\} \subset \mathcal{R}_n$ then

$$\frac{\partial}{\partial \lambda}\left\{\frac{f_n(\lambda, r + 2/\sqrt{n})}{f_n(\lambda, r)}\right\}$$

$$= \begin{cases} [2\sqrt{n}/(\sqrt{n} - \lambda)^2](1 - \exp\{-2c(n, \lambda)\}) \geq 0, & \text{if } r = 0 \\ [2\sqrt{n}/(\sqrt{n} - \lambda)^2][r(1 - e^{-2c(n,\lambda)})/(e^{rc(n,\lambda)} + e^{-rc(n,\lambda)})^2 & \\ \quad + (e^{rc(n,\lambda)} - e^{-(r+2)c(n,\lambda)})/(e^{rc(n,\lambda)} + e^{-rc(n,\lambda)})] \geq 0, & \text{if } r > 0 \end{cases}$$

for all $0 \leq \lambda < \sqrt{n}$. This implies that $f_n(\lambda, r_2)/f_n(\lambda, r_1)$ is nondecreasing in λ on $[0, \sqrt{n})$ for any $r_1, r_2 \in \mathcal{R}_n$, $r_1 < r_2$. □

We introduce a function called ρ_n. This function plays a key role in the derivation of our estimators. The domain of the function is $[0, \sqrt{n}) \times [0, \infty)$, the image is $[0, 1)$ and the function is given by

$$\rho_n(\lambda, r) = \tanh\left(c(n, \lambda)\frac{\sqrt{n}r}{2}\right).$$

Lemma 2.2. *If $0 \le \lambda < \sqrt{n}$ and $r \in \mathcal{R}_n$ then $\mathrm{E}_\lambda[S|R = r] = \rho_n(\lambda, r)$. The Bayes estimator δ_{BU}, with respect to the uniform prior on $\{-m/\sqrt{n}, m/\sqrt{n}\}$, is given by*

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

Moreover,

(a) *If $0 < \lambda < \sqrt{n}$ then $\rho_n(\lambda, \cdot)$ is increasing with $\rho_n(\lambda, 0) = 0$ and $\rho_n(\lambda, r) \to 1$ as $r \to \infty$.*

(b) *The function $\rho_n(\lambda, \cdot)$ is concave, $\rho_n(\lambda, r) \le \lambda r$ for all $r \in \mathcal{R}_n$, n odd and $\rho_n(\lambda, r) \le \lambda r/(1 + \lambda^2/n)$ for all $r \in \mathcal{R}_n$, n even.*

(c) *If $r > 0$ then $\rho_n(\cdot, r)$ is increasing on $[0, \sqrt{n})$.*

(d) *If $r \ge 0$ then the expression*

$$\frac{(1 + \lambda/\sqrt{n})^{\sqrt{n}r+1} + (1 - \lambda/\sqrt{n})^{\sqrt{n}r+1}}{(1 + \lambda/\sqrt{n})^{\sqrt{n}r} + (1 - \lambda/\sqrt{n})^{\sqrt{n}r}}$$

is nondecreasing in λ on $[0, \sqrt{n})$.

Proof. If $0 \le \lambda < \sqrt{n}$ and $r \in \mathcal{R}_n$ then $\mathrm{E}_\lambda[S \mid R = r] = \rho_n(\lambda, r)$ and $\delta_{\mathrm{BU}}(x) = (1 + \rho_n(m, r)ms/\sqrt{n})/2$. Using the properties of the tanh function we obtain the results of part (a), part (c) and the concavity of $\rho_n(\lambda, \cdot)$ for $0 \le \lambda < \sqrt{n}$. Since $\rho_n(\lambda, r) = \lambda r$ for $r = 0, 1/\sqrt{n}$ and $\rho_n(\lambda, r) = \lambda r/(1+\lambda^2/n)$ for $r = 0, 2/\sqrt{n}$, $0 \le \lambda < \sqrt{n}$, the concavity of $\rho_n(\lambda, \cdot)$ implies that $\rho_n(\lambda, r) \le \lambda r$ for all $r \in \mathcal{R}_n$, n odd, $0 \le \lambda < \sqrt{n}$ and $\rho_n(\lambda, r) \le \lambda r/(1 + \lambda^2/n)$ for all $r \in \mathcal{R}_n$, n even, $0 \le \lambda < \sqrt{n}$. Finally, if $0 \le \lambda < \sqrt{n}$ and $r \ge 0$ then

$$\frac{\partial}{\partial \lambda}\left\{\frac{(1 + \lambda/\sqrt{n})^{\sqrt{n}r+1} + (1 - \lambda/\sqrt{n})^{\sqrt{n}r+1}}{(1 + \lambda/\sqrt{n})^{\sqrt{n}r} + (1 - \lambda/\sqrt{n})^{\sqrt{n}r}}\right\} = \frac{1}{\sqrt{n}}\rho_n(\lambda, r) \ge 0$$

which gives the proof of part (d). □

Some of the risk function decompositions below will bring into play the conditional expectations

$$\alpha_n(\lambda, m) = \mathrm{E}_\theta[\rho_n(\lambda, R) \mid R > m],$$

and

$$\beta_n(\lambda, m) = \mathrm{E}_\theta[\lambda \rho_n(\lambda, R)/R \mid 0 < R \leq m],$$

where $0 < \lambda \leq m$, as well as the functions

$$\mathrm{A}_n(m) = \sup_{0<\lambda\leq m} \alpha_n(\lambda, m), \quad \text{and} \quad \mathrm{B}_n(m) = \sup_{0<\lambda\leq m} \beta_n(\lambda, m).$$

Notice that the definition of B_n applies only in the case where $\mathrm{P}_\lambda[0 < R \leq m] > 0$. The following properties, which are proved in the appendix, will be required.

Lemma 2.3. (a) *The function $\alpha_n(\cdot,\cdot)$ is increasing in both arguments and, consequently, $\mathrm{A}_n(m) = \alpha_n(m, m)$. Furthermore, $\mathrm{A}_n(m) \to 0$ as $m \to 0$ and $\mathrm{A}_n(m) \to 1$ as $m \to \infty$.*

(b) *If n is odd then $\beta_n(\cdot, m)$ is an increasing function on $(0, \sqrt{2n/(n+1)}]$ and, consequently, $\mathrm{B}_n(m) = \beta_n(m, m)$ whenever $m \leq \sqrt{2n/(n+1)}$. More generaly, $0 \leq \mathrm{B}_n(m) < m^2$ and $\lim_{m\to\sqrt{n}} \mathrm{B}_n(m) \geq 1$.*

Finally, since A_n is increasing from $(0, \sqrt{n})$ onto $(0, 1)$ we shall denote by A_n^{-1} the inverse function of A_n. This inverse function is defined on $(0, 1)$ onto $(0, \sqrt{n})$.

3. Dominance results

In this section, the space $\mathcal{R}_n$ is partitioned and the conditional risks are based on these partitions. All symmetric estimators satisfy $g(0) = 0$ if $0 \in \mathcal{R}_n$. The maximum likelihood estimator corresponds to a function g which is linear on $(0, m] \cap \mathcal{R}_n$ and constant on $(m, \sqrt{n}] \cap \mathcal{R}_n$. Therefore, it seems natural to consider the partition $\mathcal{R}_n = (\{0\} \cap \mathcal{R}_n) \cup ((0, m] \cap \mathcal{R}_n) \cup ((m, \sqrt{n}] \cap \mathcal{R}_n)$. Theorems 3.1 and 3.2 give some properties of conditional risks given that R belongs to $(0, m] \cap \mathcal{R}_n$ and $(m, \sqrt{n}] \cap \mathcal{R}_n$ respectively. Corollaries 3.1 and 3.2 are based on Theorems 3.1 and 3.2 and they provide general results. In both Theorems 3.1 and 3.2 the function g has to be nondecreasing. In Theorem 3.3 we show that if a prior π is symmetric then the corresponding estimator δ_π is also symmetric so

$$\delta_\pi(x) = \frac{1}{2}\left(1 + g_\pi(r)\frac{ms}{\sqrt{n}}\right).$$

Moreover, we show that if π does not assign probability one to the event $\{\theta = 0\}$ then g_π has to be increasing. In Theorem 3.4 the set $\{r : \rho_n(m, r) \geq r/m, r \in \mathcal{R}_n\}$ belongs to the partition and all other elements are singletons. The idea is that the maximum likelihood estimator and the other estimator have the same values on $\{r : \rho_n(m, r) \geq r/m, r \in \mathcal{R}_n\}$. In Corollaries 3.3

and 3.4 we use the results of Theorems 3.1 and 3.4 where two elements of the partition are $\{r : \rho_n(m,r) \geq r/m, r \in \mathcal{R}_n\}$, $(m, \sqrt{n}] \cap \mathcal{R}_n$ while all others are singletons of $\mathcal{R}_n$.

Theorem 3.1. *Let g be a nondecreasing function on $(m, \sqrt{n}] \cap \mathcal{R}_n$. If*

$$(2\,\mathrm{A}_n(m) - 1) \leq g(r) \leq 1 \quad \text{for all } r \in (m, \sqrt{n}] \cap \mathcal{R}_n$$

then

$$\mathrm{E}_\theta\big[\mathrm{L}\big(\theta, \delta_{\mathrm{mle}}(X)\big) - \mathrm{L}\big(\theta, \delta_g(X)\big) \mid R > m\big] \geq 0 \quad \text{for all } \theta \in \Theta(m).$$

Proof. First, Lemma 2.3 shows that $(2\,\mathrm{A}_n(m) - 1) \leq 1$, for all $m > 0$, which implies that the given condition on g is not vacuous. Decomposing the difference into conditional risks, we obtain

$$\begin{aligned}
&\mathrm{E}_\theta\big[\mathrm{L}\big(\theta, \delta_{\mathrm{mle}}(X)\big) - \mathrm{L}\big(\theta, \delta_g(X)\big) \mid R > m\big] \\
&\quad = \frac{m^2}{4n}\,\mathrm{E}_\theta\left[\big(1 - g(R)\big)\left(g(R) - \left\{2\frac{\lambda}{m}\rho_n(\lambda, R) - 1\right\}\right) \,\middle|\, R > m\right] \\
&\quad \geq \frac{m^2}{4n}\,\mathrm{E}_\theta\left[\big(1 - g(R)\big)\left(g(R) - \left\{2\frac{\lambda}{m}\alpha_n(\lambda, m) - 1\right\}\right) \,\middle|\, R > m\right] \\
&\quad \geq \frac{m^2}{4n}\,\mathrm{E}_\theta\big[\big(1 - g(R)\big)\big(g(R) - \{2\,\mathrm{A}_n(m) - 1\}\big) \mid R > m\big] \geq 0
\end{aligned}$$

where the equality comes from the conditional expectation given R and Lemma 2.2; the first inequality holds by virtue of the inequality

$$\mathrm{Cov}_\theta[\rho_n(\lambda, R), g(R) \mid R > m] \geq 0,$$

which in turn is valid since both ρ_n and g are nondecreasing on $(m, \sqrt{n}] \cap \mathcal{R}_n$, and the second inequality follows from Lemma 2.3. □

Corollary 3.1. *Consider that $m < 1/\sqrt{n}$ when n is odd or $m < 2/\sqrt{n}$ when n is even. If the conditions of Theorem* 3.1 *are satisfied then* $\mathrm{R}(\theta, \delta_{\mathrm{mle}}) - \mathrm{R}(\theta, \delta_g) \geq 0$ *for all $\theta \in \Theta(m)$.*

Proof. We obtain that

$$\mathrm{R}(\theta, \delta_{\mathrm{mle}}) - \mathrm{R}(\theta, \delta_g) = \mathrm{E}_\theta\big[\mathrm{L}\big(\theta, \delta_{\mathrm{mle}}(X)\big) - \mathrm{L}\big(\theta, \delta_g(X)\big) \mid R > m\big]\,\mathrm{P}_\theta[R > 0]$$

for all $\theta \in \Theta(m)$. □

Remark 3.1. If $n = 1$ then $\mathrm{A}_n(m) = m$ and the conditions of Corollary 3.1 are satisfied if and only if $2m - 1 \leq g(1) \leq 1$. If $n = 2$ then $\mathrm{A}_n(m) = \sqrt{2}m/(1 + m^2/2)$ and the conditions of Corollary 3.1 are satisfied if and only if $2\sqrt{2}m/(1 + m^2/2) - 1 \leq g(\sqrt{2}) \leq 1$. These two results are identical to the ones in Marchand and MacGibbon (2000, Theorem 4.2 p. 145).

Theorem 3.2. *Assume that $m \geq 1/\sqrt{n}$ when n is odd and $m \geq 2/\sqrt{n}$ when n is even. Let $g(r)$ and $r(r-mg(r))$ be nondecreasing in r on $(1/\sqrt{n}, m] \cap \mathcal{R}_n$. If*

$$\mathrm{B}_n(m) \leq 1$$

and

$$(2\,\mathrm{B}_n(m) - 1)r/m \leq g(r) \leq r/m \quad \text{for all } r \in (1/\sqrt{n}, m] \cap \mathcal{R}_n$$

then

$$\mathrm{E}_\theta\big[\mathrm{L}(\theta, \delta_{\mathrm{mle}}(X)) - \mathrm{L}(\theta, \delta_g(X)) \;\big|\; 0 < R \leq m\big] \geq 0 \quad \text{for all } \theta \in \Theta(m).$$

Proof. First, note that $\{m : \mathrm{B}_n(m) < 1\} \neq \varnothing$ since the properties of B_n in Lemma 2.3 imply that $(0, 1] \subset \{m : \mathrm{B}_n(m) < 1\}$. Let $\mathcal{A} = \{0 < R \leq m\}$. We obtain

$$\begin{aligned}
&\mathrm{E}_\theta\big[\mathrm{L}(\theta, \delta_{\mathrm{mle}}(X)) - \mathrm{L}(\theta, \delta_g(X)) \mid \mathcal{A}\big] \\
&\quad = \frac{m^2}{4n}\,\mathrm{E}_\theta\left[\left(\frac{R}{m} - g(R)\right)\left(g(R) - \left\{2\lambda\frac{\rho_n(\lambda, R)}{R} - 1\right\}\frac{R}{m}\right) \,\middle|\, \mathcal{A}\right] \\
&\quad \geq \frac{m^2}{4n}\,\mathrm{E}_\theta\left[\left(\frac{R}{m} - g(R)\right)\left(g(R) - \{2\beta_n(\lambda, m) - 1\}\frac{R}{m}\right) \,\middle|\, \mathcal{A}\right] \\
&\quad \geq \frac{m^2}{4n}\,\mathrm{E}_\theta\left[\left(\frac{R}{m} - g(R)\right)\left(g(R) - \{2\,\mathrm{B}_n(m) - 1\}\frac{R}{m}\right) \,\middle|\, \mathcal{A}\right] \geq 0
\end{aligned}$$

where the first inequality holds because $r(r - mg(r))$ is nondecreasing in r on $(1/\sqrt{n}, m] \cap \mathcal{R}_n$ and $\rho_n(\lambda, r)/r$ is nonincreasing in r for $r \in (1/\sqrt{n}, m] \cap \mathcal{R}_n$ which implies that $\mathrm{Cov}_\theta\big[R(R - mg(R)), \rho_n(\lambda, R)/R \;\big|\; 0 < R \leq m\big] \leq 0$. The second inequality comes from Lemma 2.3. □

Corollary 3.2. *Let $m_- = \sup\{r : r \leq m, r \in \mathcal{R}_n \cup \{0\}\}$ and $m_+ = \inf\{r : r > m, r \in \mathcal{R}_n\}$. Let $0 \leq g(r) \leq 1 \wedge r/m$ for all $r \in \mathcal{R}_n$ with $g(0) = 0$. If*

- *g is nondecreasing on $\mathcal{R}_n$,*
- *$g(r)/r$ is nonincreasing in r for all $r \in \mathcal{R}_n \setminus \{0\}$,*
- *$g(m_-) \geq (2\,\mathrm{B}_n(m) - 1)m_-/m$ and $g(m_+) \geq 2\,\mathrm{A}_n(m) - 1$*

then

$$\mathrm{R}(\theta, \delta_g) \leq \mathrm{R}(\theta, \delta_{mle}) \quad \text{for all } \theta \in \Theta(m).$$

Proof. Since g is nondecreasing and bounded by 1 on $\mathcal{R}_n$, $g(m_+) \geq 2\,\mathrm{A}_n(m) - 1$ implies that $2\,\mathrm{A}_n(m) - 1 \leq g(r) \leq 1$ for all $r \in (m, \sqrt{n}] \cap \mathcal{R}_n$ and the conditions of Theorem 3.1 are satisfied. If $(0, m] \cap \mathcal{R}_n = \emptyset$ then the proof is complete. Otherwise, since $g(r)/r$ is nonincreasing in r for all

$r \in \mathcal{R}_n \backslash \{0\}$ we obtain that $r(r - mg(r))$ is nondecreasing in r on $\mathcal{R}_n$. Finally, $g(m_-) \geq (2\,\mathrm{B}_n(m) - 1)m_-/m$ implies that $(2\,\mathrm{B}_n(m) - 1)r \leq mg(r) \leq r$ for all $0 < r \leq m$, $r \in \mathcal{R}_n$ and the conditions of Theorem 3.2 are satisfied. □

Theorem 3.3. *Let $T = \lambda$. For a given symmetric prior π on $\Theta(m)$ with $\pi(\{T = 0\}) < 1$ the Bayes estimator δ_π is given by $\delta_\pi(x) = (1 + g_\pi(r)ms/\sqrt{n})/2$ where $g_\pi(r) = \mathrm{E}[T\rho_n(T, r)/m \mid R = r]$. In other words, $g_\pi(r)$ is the expectation of $T\rho_n(T, r)/m$ with respect to the posterior distribution of T. Moreover, g_π is increasing with $g_\pi(0) = 0$ and $0 \leq g_\pi \leq \rho_n(m, \cdot)$.*

Proof. Assume, without loss of generality, that $r \neq 0$, $r \in \mathcal{R}_n$. We use the representation $\theta = TU$ with $T = |\theta|$ and $U = \mathrm{sgn}(\theta)$. T is distributed according to a probability measure σ on $[0, m]$ and, conditionally on the event $T = t$, $t \neq 0$, U is uniformly distributed on $\{-1, 1\}$ while $\mathrm{P}[U = 0 \mid T = 0] = 1$. This representation now implies that the posterior distribution of T has a density, with respect to the measure σ, proportional to $f_n(t, r)$, that is,

$$\sigma(\mathrm{d}t \mid x) = \frac{f_n(t, r)}{\int_0^m f_n(\tau, r)\sigma(\mathrm{d}\tau)}\sigma(\mathrm{d}t) \quad \text{for } t \in [0, m],$$

and that, conditioned on the event $T = t$, the posterior distribution of U is given by

$$\mathrm{P}[U = u \mid T = t, X = x] = \begin{cases} (1 + \rho_n(t, r)su)/2 & \text{if } t > 0,\ u \in \{-1, 1\} \\ 1 & \text{if } t = 0,\ u = 0. \end{cases}$$

The Bayes estimator of θ is given by $\mathrm{E}[\theta \mid X = x]$ and

$$\begin{aligned} \mathrm{E}[\theta \mid X = x] &= \mathrm{E}[TU \mid X = x] \\ &= \mathrm{E}[T\,\mathrm{E}[U \mid T, X = x] \mid X = x] \\ &= \mathrm{E}[T\rho_n(T, r)s \mid X = x] \\ &= \mathrm{E}[T\rho_n(T, r) \mid R = r]s \\ &= g_\pi(r)ms \end{aligned}$$

where the fourth equality holds because the posterior distribution of T depends on x through r only. Since $0 \leq \rho_n(t, r) \leq \rho_n(m, r)$ for all $t \in \Theta(m)$, we obtain $0 \leq g_\pi(r) \leq \rho_n(m, r)$. Finally, for $0 < r_1 < r_2$, $r_1, r_2 \in \mathcal{R}_n$ we obtain

$$\begin{aligned} g_\pi(r_2) &= \mathrm{E}[T\rho_n(T, r_2)/m \mid R = r_2] \\ &> \mathrm{E}[T\rho_n(T, r_1)/m \mid R = r_2] \\ &\geq \mathrm{E}[T\rho_n(T, r_1)/m \mid R = r_1] \\ &= g_\pi(r_1) \end{aligned}$$

where the inequalities come from the monotonicity property of $\rho_n(\cdot, r)$ and the fact that the conditional distribution of T given that $R = r$ has monotone likelihood ratio when r is viewed as the parameter (this is a direct consequence of Lemma 2.2 part (d)). □

Theorem 3.4. *Assume that* $r \in \mathcal{R}_n$ *and* $\rho_n(m, r) < r/m$. *If*

$$2\rho_n(m, r) - (1 \wedge r/m) \leq g(r) \leq 1 \wedge r/m$$

then

$$\mathrm{E}_\theta\big[\mathrm{L}(\theta, \delta_{\mathrm{mle}}(X)) - \mathrm{L}(\theta, \delta_g(X)) \mid R = r\big] \geq 0 \quad \textit{for all } \theta \in \Theta(m).$$

Proof. We have

$$\begin{aligned}\mathrm{E}_\theta\big[\mathrm{L}(\theta, \delta_{\mathrm{mle}}(X)) &- \mathrm{L}(\theta, \delta_g(X)) \mid R = r\big] \\ &= \frac{m^2}{4n}(1 \wedge r/m - g(r))(g(r) - \{2\lambda\rho_n(\lambda, r)/m - 1 \wedge r/m\}) \\ &\geq \frac{m^2}{4n}(1 \wedge r/m - g(r))(g(r) - \{2\rho_n(m, r) - 1 \wedge r/m\}) \geq 0\end{aligned}$$

where the first inequality comes from the monotonicity of the function $\rho_n(\cdot, r)$. □

Remark 3.2. It is possible to improve on the class of estimators in Theorem 3.4 by adding the condition $g(r) \leq \rho_n(m, r) \wedge r/m$. In fact, Moors (1985) shows that if g does not satisfy the condition $0 \leq g(r) \leq \rho_n(m, r)$ for all $r \in \mathcal{R}_n$ then we can replace g by g^* where $g^*(r) = g(r)$ when $0 \leq g(r) \leq \rho_n(m, r)$ and $0 \leq g^*(r) \leq \rho_n(m, r)$ for all $r \in \mathcal{R}_n$.

Corollary 3.3. *Let* $0 \leq g(r) \leq 1 \wedge r/m$ *for* $r \in \mathcal{R}_n$. *If*

$$g \textit{ is nondecreasing on } \mathcal{R}_n \quad \textit{and} \quad m \leq \mathrm{A}_n^{-1}(1/2) \wedge \sqrt{\tfrac{1}{2}},$$

then

$$\mathrm{R}(\theta, \delta_g) \leq \mathrm{R}(\theta, \delta_{\mathrm{mle}}) \quad \textit{for all } \theta \in \Theta(m).$$

Proof. This proof is based on verifying the conditions of Theorem 3.1 on $r \in (m, \sqrt{n}] \cap \mathcal{R}_n$ and the conditions of Theorem 3.4 on $r \in (0, m] \cap \mathcal{R}_n$. To apply Theorem 3.1 we need to verify that $(2\,\mathrm{A}_n(m) - 1) \leq g(r) \leq 1$ for all $r > m$, $r \in \mathcal{R}_n$. The conditions of Corollary 3.3 imply that $2\,\mathrm{A}_n(m) - 1 \leq 0$. Therefore, if g is nondecreasing on $\mathcal{R}_n$ and $0 \leq g(r) \leq 1$ for $r > m$, $r \in \mathcal{R}_n$ then the conditions of Theorem 3.1 will be satisfied. To apply Theorem 3.4, we need to verify that $(2\rho_n(m, r) - r/m) \leq g(r) \leq r/m$ for all $r \in (0, m] \cap \mathcal{R}_n$. Theorem 3.4 applies whenever $\rho_n(r, m) \leq r/m$. Part (b) of Lemma 2.2 tells us that $\rho_n(m, r) \leq mr$ for all $r \in \mathcal{R}_n$. Since $m \leq \sqrt{\frac{1}{2}}$, we obtain $\rho_n(m, r) \leq r/m$ for all $r \in \mathcal{R}_n$ and $2\rho_n(m, r) - 1 \leq 0$ for all $r \in (0, m] \cap \mathcal{R}_n$. Therefore, g satisfies the conditions of Theorem 3.4 for $r \in (0, m] \cap \mathcal{R}_n$. □

Corollary 3.4. *Let* $m_0 = m[(2\,\mathrm{A}_n(m) - 1) \vee \sup\{2(\rho_n(m,r) \wedge r/m) - r/m : 0 \le r \le m, r \in \mathcal{R}_n \cup \{0\}\}]$. *Assume that* $m_0 \le m_1 \le m$. *If* δ_g *is the maximum likelihood estimator for the new problem where* $|\theta| \le m_1$, *then* $\mathrm{R}(\theta, \delta_g) \le \mathrm{R}(\theta, \delta_{\mathrm{mle}})$ *for all* $\theta \in \Theta(m)$.

Proof. We have $g(r) = (m_1 \wedge r)/m$. Notice that if $r \ge m_0$ then $\rho_n(m,r) \le r/m$ so $\delta_g(r) = \delta_{\mathrm{mle}}(r)$ on the set $\{r : \rho_n(m,r) > r/m\}$. The rest of the proof is based on verifying the conditions of Theorem 3.1 on $r \in (m, \sqrt{n}] \cap \mathcal{R}_n$ and the conditions of Theorem 3.4 on $r \in (0,m] \cap \mathcal{R}_n \cap \{r : \rho_n(m,r) \le r/m\}$. □

4. Examples and numerical evaluations

In this section we shall consider three competitors to the maximum likelihood estimator. Two of them will always dominate the maximum likelihood estimator while the third will sometimes dominate the maximum likelihood estimator. We have chosen these estimators for their simplicity but we could have provided more complicated ones using a bayesian approach.

Example 4.1. We set $g(r) = \rho_n(m,r) \wedge r/m$ and denote the estimator by δ_{mrs}.

This example comes from Moors (1985). Since δ_{mrs} satisfies the conditions of Theorem 3.4 it dominates δ_{mle}. We obtain that $\delta_{\mathrm{mrs}} = \delta_{\mathrm{BU}}$ if and only if $m \le 1$ when n is odd or $m \le \sqrt{n/(n-1)}$ when n is even. Notice that δ_{BU} is a Bayes rule, so it is admissible.

Example 4.2. We set $g(r) = (m_0 \wedge r)/m$ and denote the estimator by δ_{cve}.

In Charras and van Eeden (1991) it is shown that there exists m_0 such that δ_{cve} dominates δ_{mle}. We give an explicit value to m_0 in Corollary 3.4 and use it in this example. In Charras and van Eeden (1991) it is also shown that the class of Bayes estimators is complete. In our case, it is not difficult to see that the class of Bayes estimators with respect to symmetric priors is complete for the class of symmetric estimators. Notice that if $m \le (n-2)/\sqrt{n}$ with $n > 2$ then any estimator satisfying the conditions of Corollary 3.4 will correspond to a function g which is not strictly increasing. However, Bayes estimators are associated with a strictly increasing function g (see Theorem 3.3). Therefore, any estimator satisfying the conditions of Corollary 3.4 is inadmissible. This gives a partial answer to the open question raised in a remark on page 127 in Charras and van Eeden (1991).

Since we know that the class of Bayes estimators with respect to symmetric priors is complete it is then important to study at least one estimator in this class.

Example 4.3. We set $g(r) = \mathrm{E}[T\rho_n(T,r)/m \mid R = r]$ where the distribution of T is the conditional distribution of λ given that $R = r$ and the prior density of p is proportional to $(p(1-p))^{-1}$. The estimator is called δ_{bys}.

In other words, δ_{bys} is the Bayes estimator of p based on the prior density proportional to $\big(p(1-p)\big)^{-1}$. This estimator has been inspired by the methodology developed in Marchand and Perron (2001). As m tends to $\sqrt{n}$, n being fixed, δ_{bys} tends to δ_{mle}. It can be shown that the function g will satisfy the first conditions of Corollary 3.2 but we still have to verify numerically that $g(m_+) \geq 2\,\mathrm{A}_n(m) - 1$ and $g(m_-) \geq (2\,\mathrm{B}_n(m) - 1)m_-/m$. If g satisfies these two conditions then δ_{bys} dominates δ_{mle}. Numerical evaluations show that the two conditions are satisfied for small values of m as it can be seen in Table 1.

For example, set $n = 10$. Table 1 tells us that if $m \leq .601$ then our Bayes estimator dominates the maximum likelihood estimator. Notice that $m = .601$ corresponds to $p \in [0.405; 0,595]$ with $0.595 = (1 + 0.601/\sqrt{10})/2$. In Fig. 2, $[0.4, 0.6] \not\subset [0.405; 0,595]$ so we do not know if δ_{bys} dominates δ_{mle} because our approach uses only sufficient conditions. However, we can see graphically that δ_{bys} dominates δ_{mle}. For $p \in [0.4, 0.6]$, we can verify from Table 1 that δ_{bys} dominates δ_{mle} whenever $n = 1, 2, 3, 4, 5, 6$ and 8. In general, the risk of δ_{bys} is smaller than the one of δ_{mle} when λ is small. In fact, a minor modification to Theorem 3.4 will give that $\mathrm{R}(\theta, \delta_{\mathrm{mle}}) > \mathrm{R}(\theta, \delta_{\mathrm{bys}})$ for $\lambda \in [0, a]$ if $2a\rho_n(a, r)/m \leq g_\pi(r) + (1 \wedge r/m)$ for all $r \in \mathcal{R}_n$ but $a\rho_n(a, r)$ tends to 0 as a tends to 0.

We know from theoretical considerations that if n is small or m is small then δ_{mle} is a bad estimator. Suppose that the nature of the problem tells us that p must be between 0.4 and 0.6. In this case, $m = 0.2\sqrt{n}$ and $\delta_{\mathrm{mrs}} = \delta_{\mathrm{BU}}$ if and only if $n \leq 26$. The estimator δ_{cve} corresponds to the maximum likelihood estimator for the problem $p \in [(1 - m_0/\sqrt{n})/2, (1 + m_0/\sqrt{n})/2]$. Numerical evaluations are given in Table 2.

From Table 2 we shall expect a very slight improvement of δ_{cve} over δ_{mle} when n is large and $m\sqrt{n}$ is fixed. We evaluate the risk functions in Figs. 1, 2 and 3 numerically. We have used $n = 10$, 25, 1000. When $n = 10$, δ_{mle} takes only three values, 0.4, 0.5 and 0.6. The estimator δ_{cve} takes the three values 0.45, 0.5 and 0.55. We can see from Fig. 1 that all of estimators given in the three examples dominate δ_{mle}. We would have observed the same phenomenon for $n < 10$. In these cases, the risk function of δ_{cve} stays between those of δ_{mrs} and δ_{bys}. Even if δ_{mle} takes the value 0.6 for $x > 5$,

Table 1.

n	1	2	3	4	5	6	7	8	9	10	15	20	25	30
$m \leq$	.615	.485	.577	.547	.456	.574	.487	.590	.507	.601	.544	.504	.573	.532

Table 2.

n	1–4	5	6	7	8	9	10	15	20	25	30	50
$(1 + m_0/\sqrt{n})/2$	.500	.518	.501	.525	.510	.532	.549	.569	.581	.588	.592	.599

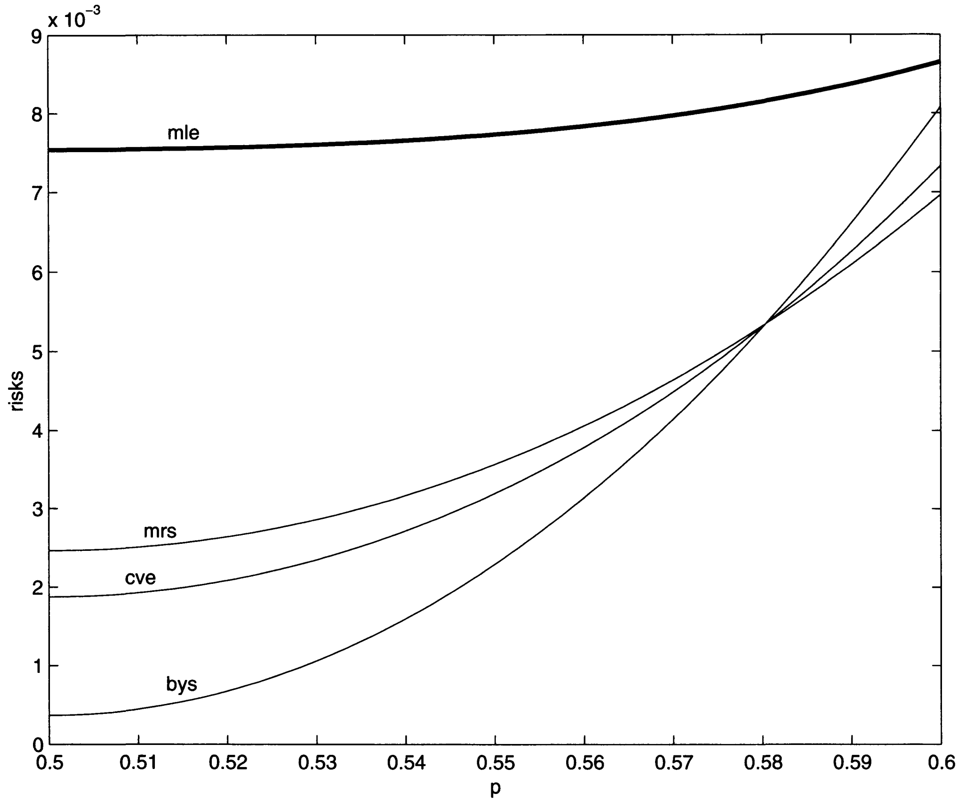

Figure 1. Risk functions for the estimation of p in a binomial (n, p) with $n = 10$ and $p \in [0.4, 0.6]$.

mle maximum likelihood estimator

cve mle for the problem $p \in [0.450, 0.550]$

mrs Moors' estimator

bys Bayes, prior proportional to $\big(p(1-p)\big)^{-1}$

its risk function is the worst at 0.6. When $n = 25$, δ_{mrs} is still a Bayes estimator. However, Fig. 2 shows that the improvements over δ_{mle} are less impressive for δ_{mrs} and δ_{cve}. From this figure, we see that δ_{mrs} dominates δ_{cve}. Here, the risk function of δ_{mle} becomes better as p approaches a boundary point. The risk function of the estimator δ_{bys} behaves differently from the other risk functions. The Bayes estimator no longer dominates δ_{mle} but the improvement on the risk function is quite big when p is near $\frac{1}{2}$ and holds until p approaches a boundary point. In a real situation, one might expect that n is large, something of the order of 1000. When n is that large, using δ_{mle}, δ_{cve} or δ_{mrs} makes no perceptible difference. The differences between δ_{mle} and δ_{bys} will be more significant when r is close to m. Therefore, the risk functions will be approximately the same when p is near $\frac{1}{2}$, something in

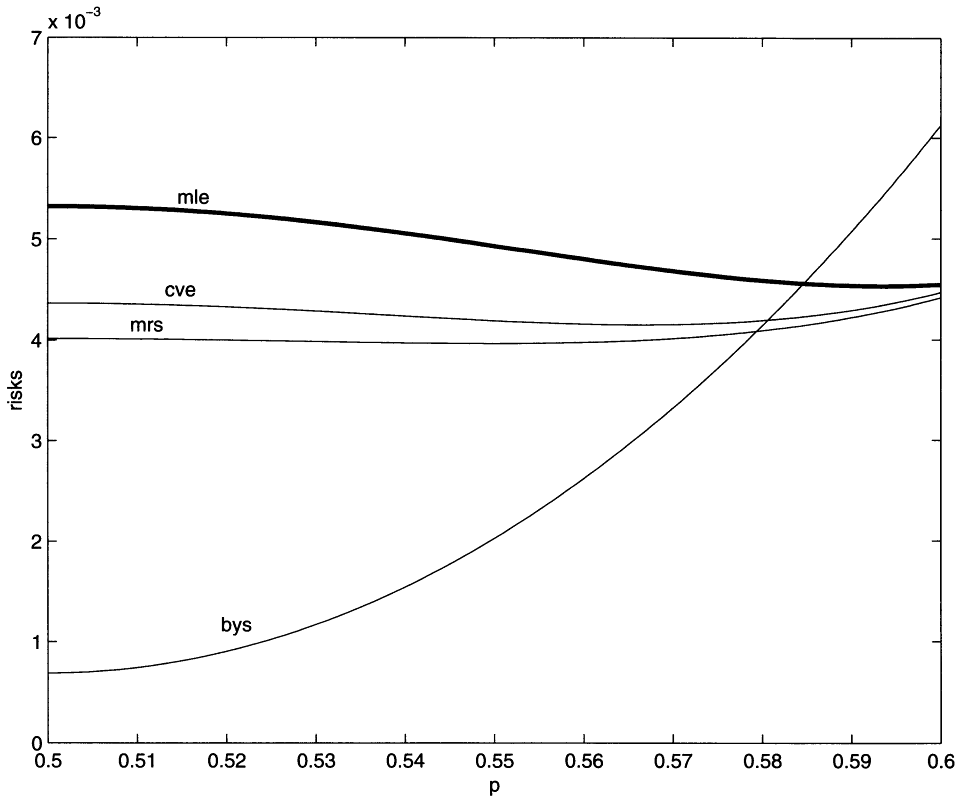

Figure 2. Risk functions for the estimation of p in a binomial (n,p) with $n = 25$ and $p \in [0.4, 0.6]$.

mle maximum likelihood estimator
cve mle for the problem $p \in [0.412, 0.588]$
mrs Moors' estimator
bys Bayes, prior proportional to $\big(p(1-p)\big)^{-1}$

the neighbourhood of $1/4n$. One would then have the same result as if there were no constraints on p. Similarly, when p is on a boundary point, δ_{mle} will be on target with a probability almost equal to $\frac{1}{2}$ so we should expect a risk of $0.4 \cdot 0.6/2n$ for δ_{mle}. The Bayes estimator still shrinks δ_{mle} towards $\frac{1}{2}$ even if r is close to m and this may explain why there is a sudden growth in its risk function when p is close to the boundary. These phenomena were also observed for large values of n in numerous numerical evaluations and Fig. 3 shows the case $n = 1000$.

In conclusion, if the length of the parameter space for p is of the order of $O(1/\sqrt{n})$ then any of the suggested estimators will outperform the maximum likelihood estimator. When n is small, one should never use the maximum likelihood estimator. If the length of the parameter space for p is not of

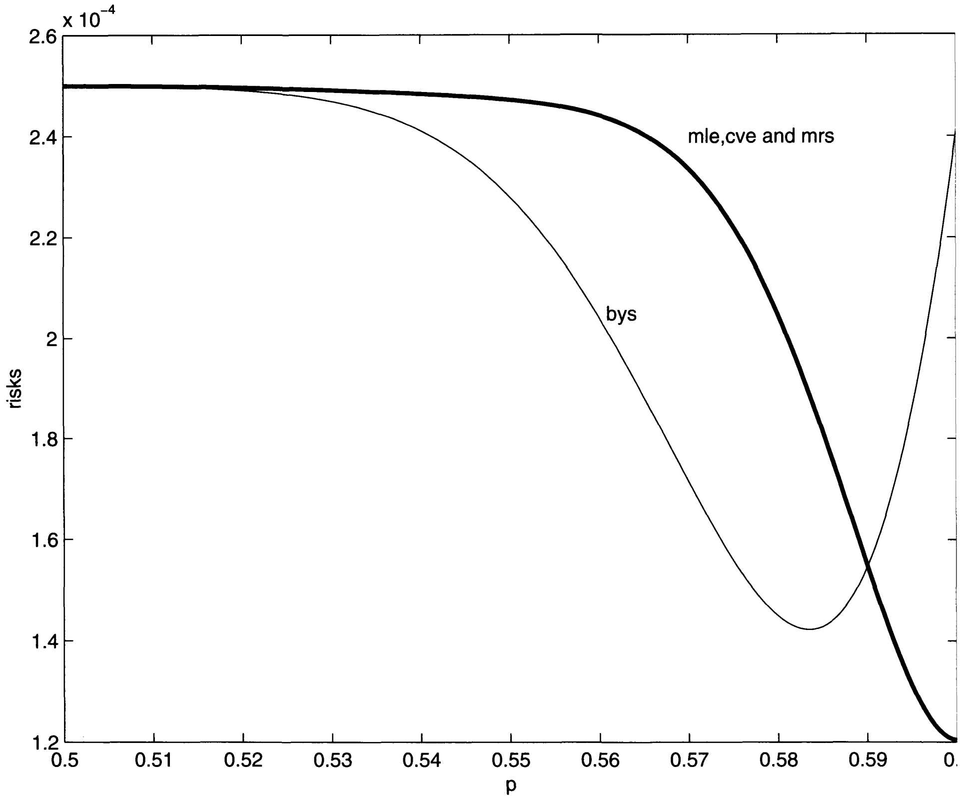

Figure 3. Risk functions for the estimation of p in a binomial (n, p) with $n = 10$ and $p \in [0.4, 0.6]$.

mle maximum likelihood estimator

cve mle for the problem $p \in [0.400, 0.500]$

mrs Moors' estimator

bys Bayes, prior proportional to $\big(p(1-p)\big)^{-1}$

the order of $O(1/\sqrt{n})$ but the choice of m is conservative (that is: $\Theta(m)$ is too large), then it seems preferable to select δ_{bys} even though δ_{bys} does not dominate δ_{mle}. As we mentioned in the beginning of this section, it is possible to develop Bayes estimators which perform better in comparison to δ_{mle} but this goes beyond the scope of this paper.

APPENDIX

A. Proof of Lemma 2.3

(a) Let $0 \le \lambda_1 < \lambda_2 < m$. From Lemma 2.2 we obtain

$$\begin{aligned}\alpha_n(\lambda_1, m) &= \mathrm{E}_{\lambda_1}[\rho_n(\lambda_1, R) \mid R > m] \\ &\le \mathrm{E}_{\lambda_1}[\rho_n(\lambda_2, R) \mid R > m] \\ &\le \mathrm{E}_{\lambda_2}[\rho_n(\lambda_2, R) \mid R > m] \\ &= \alpha_n(\lambda_2, m)\end{aligned}$$

where the first inequality follows from the monotone increasing property of ρ_n, and the second inequality follows from the fact that the probability distribution function of R has a monotone likelihood ratio. Assume that $0 \le \lambda \le m_1 < m_2 < \sqrt{n}$. If $\mathrm{P}_\lambda[m_1 < R \le m_2] = 0$ then $\alpha_n(\lambda, m_1) = \alpha_n(\lambda, m_2)$. Otherwise,

$$\begin{aligned}\alpha_n(\lambda, m_1) = {} & \frac{\mathrm{P}_\lambda[R > m_2]}{\mathrm{P}_\lambda[R > m_1]}\alpha_n(\lambda, m_2) \\ & + \left(1 - \frac{\mathrm{P}_\lambda[R > m_2]}{\mathrm{P}_\lambda[R > m_1]}\right)\mathrm{E}_\lambda[\rho_n(\lambda, R) \mid m_1 < R \le m_2] < \alpha_n(m_2, \lambda),\end{aligned}$$

since the increasing property of ρ_n leads to $\rho_n(\lambda, r) \le \alpha(\lambda, m_2)$ on $\{r : m_1 < r \le m_2\}$. Finally, from Lemma 2.2 with $|\theta| = m$, we obtain

$$0 \le \mathrm{A}_n(m) = \mathrm{E}_\lambda\left[R\frac{\rho_n(m, R)}{R} \,\middle|\, R > m\right] \le m\,\mathrm{E}_\lambda[R \mid R > m] \to 0, \quad \text{as } m \to 0$$

and

$$\lim_{m \to \sqrt{n}} \mathrm{A}_n(m) = \lim_{m \to \sqrt{n}} \rho_n(m, \sqrt{n}) = 1.$$

(b) Notice that in this proof, the condition saying that n is odd is used only to state that $\mathrm{E}_\lambda[\rho_n(\lambda, R)R | 0 < R \le m]) = \mathrm{E}_\lambda[\rho_n(\lambda, R)R | R \le m])$. Now, straightforward computations give

$$\begin{aligned}&\frac{\partial \beta_n(\lambda, m)}{\partial \lambda} \\ &= \frac{1}{\lambda}\left\{\frac{\lambda^2}{1 - \lambda^2/n} + \beta_n(\lambda, m)\left(1 - \lambda 1 - \lambda^2/n\,\mathrm{E}_\lambda[\rho_n(\lambda, R)R \mid 0 < R \le m]\right)\right\}.\end{aligned}$$

Moreover, since $r\rho_n(\lambda, r)$ is nondecreasing in r, when n is odd $\mathrm{P}_\lambda[R=0]=0$. Thus

$$\begin{aligned}\mathrm{E}_\lambda[\rho_n(\lambda,R)R \mid 0<R\le m] &\le \mathrm{E}_\lambda[\rho_n(\lambda,R)R]\\ &= \mathrm{E}_\lambda[\sqrt{n}(2\bar{X}-1)] = \lambda\end{aligned}$$

and $\beta_n(\lambda,m)\le\lambda^2$ because $\mathrm{P}_\lambda[\rho_n(\lambda R)\le\lambda R]=1$. If $\lambda\le 1$ then $\partial\beta_n(\lambda,m)/\partial\lambda$ is the sum of two positive elements so $\beta_n(\lambda,m)$ is increasing in λ. If $1<\lambda\le\sqrt{2n/(n+1)}$ then $1-\lambda^2/(1-\lambda^2/n)<0$ and

$$\begin{aligned}&\frac{\partial\beta_n(\lambda,m)}{\partial\lambda}\\ &=\frac{1}{\lambda}\left\{\frac{\lambda^2}{1-\lambda^2/n}+\beta_n(\lambda,m)\left(1-\frac{\lambda}{1-\lambda^2/n}\mathrm{E}_\lambda[\rho_n(\lambda,R)R\mid 0<R\le m]\right)\right\}\\ &\ge\frac{1}{\lambda}\left\{\frac{\lambda^2}{1-\lambda^2/n}+\beta_n(\lambda,m)\left(1-\frac{\lambda^2}{1-\lambda^2/n}\right)\right\}\\ &\ge\frac{\lambda}{(1-\lambda^2/n)}[2-(n+1)\lambda^2/n]\ge 0.\end{aligned}$$

Finally,

$$\beta_n(\lambda,m)\le\lambda^2\le m^2 \quad \text{for all } \lambda\le m$$

so $\mathrm{B}_n(m)\le m^2$ and, for $m\ge 1/\sqrt{n}$,

$$\mathrm{B}_n(m)\ge\beta_n(m,m)\ge\mathrm{E}_m[\rho_n(m,R)\mid 0<R\le m]\to_{m\to\sqrt{n}} 1. \qquad \square$$

Acknowledgements. This work has been supported by NSERC of Canada. The author is grateful to Constance van Eeden for all the preprints she sent to him, and, especially, for providing a copy of Moors' thesis. Many thanks to the editorial commitee and, especially, to the referees for their help. The author is also thankful to Professor Martin Goldstein, who helped with the English and to Claude Beauchamp, who helped with the graphs.

REFERENCES

Charras, A. and van Eeden, C. (1991). Bayes and admissibility properties of estimators in truncated parameter spaces. *Canad. J. Statist.* 19, 121–134.

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.

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. *Ann. Statist.* 29, 1066–1081.

Moors, J.J.A. (1985). *Estimation in Truncated Parameter Spaces.* Ph.D. thesis, Tilburg University.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* 34, 751–768.

Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* 62, 63–69.

François Perron
Département de mathématiques et statistique
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal, QC H3C 3J7
Canada
perronf@dms.umontreal.ca