Stat Comput (2007) 17: 281–291
DOI 10.1007/s11222-007-9032-0

Minimax estimation of a probability of success under LINEX loss

Alicja Jokiel-Rokita · Ryszard Magiera

Received: 17 July 2006 / Accepted: 6 March 2007 / Published online: 28 July 2007

Abstract Minimax estimation of a binomial probability under LINEX loss function is considered. It is shown that no equalizer estimator is available in the statistical decision problem under consideration. It is pointed out that the problem can be solved by determining the Bayes estimator with respect to a least favorable distribution having finite support. In this situation, the optimal estimator and the least favorable distribution can be determined only by using numerical methods. Some properties of the minimax estimators and the corresponding least favorable prior distributions are provided depending on the parameters of the loss function. The properties presented are exploited in computing the minimax estimators and the least favorable distributions. The results obtained can be applied to determine minimax estimators of a cumulative distribution function and minimax estimators of a survival function.

Keywords Minimax estimation · Binomial probability · LINEX loss function · Finitely supported least favorable priors

1 Introduction

The problem of estimation of a probability of success in Bernoulli trials is engaging a good deal of attention in the literature under various optimality criteria. Searching for optimal procedures in estimating the probability of success has an important significance from a theoretical and practical point of view as well.

A. Jokiel-Rokita · R. Magiera (✉)
Institute of Mathematics and Computer Science, Wrocław University of Technology, 50-370 Wrocław, Poland
e-mail: ryszard.magiera@pwr.wroc.pl

In this paper we focus our attention upon the estimation of the probability of success under minimax criterion and in the case when the loss associated with the error of estimation is determined by a linear-exponential loss function, the so called LINEX loss function (LLF) defined by (2).

Let X denote a random variable having the binomial distribution $\mathcal{B}(n, \vartheta)$. It is well known that the minimax estimator under the quadratic loss function in estimating the probability of success ϑ can be easily calculated as a Bayes estimator having constant risk function. This result is due to Hodges and Lehmann (1950) who proved that under quadratic loss function the minimax estimator of ϑ is

$$d^*(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}$$

having the constant risk $\mathcal{R}(\vartheta, d^*) = 1/4(1 + \sqrt{n})^2$, and the least favorable distribution is the beta distribution $\mathcal{B}e(\sqrt{n}/2, \sqrt{n}/2)$. However, under some other important and widely used loss functions no equalizer estimators of the binomial probability ϑ are available.

In the case of a weighted absolute error loss, minimax estimators of the probability of success were considered by Eichenauer-Herrmann et al. (1990). They calculated the minimax estimators numerically by performing an iterative method.

Minimax estimators of a binomial probability under entropy loss function were considered by Wieczorkowski and Zieliński (1992) and Rukhin (1993). The minimax estimators were calculated numerically by Wieczorkowski (1998) by using the Box complex method of constrained optimization.

In this paper minimax estimation of the probability of success under the LLF is considered. It is shown that no equalizer estimator is available in the statistical decision problem under consideration. This leads to the problem of searching for a minimax estimator that is Bayes with respect to a least favorable distribution having finite support. In this situation, the optimal estimator and the least favorable distribution can be determined only by using numerical methods. Some properties of the minimax estimators and the corresponding least favorable prior distributions are provided depending on the parameters of the loss function. The properties presented are exploited in computing the minimax estimators and the least favorable distributions. In the final section of the paper some numerical results are presented. The results obtained in the paper can be applied to determine minimax estimators of a cumulative distribution function and minimax estimators of a survival function (see Jokiel-Rokita and Magiera 2007).

2 Preliminaries

2.1 The statistical estimation problem and basic facts

Let $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\})$ denote a statistical space and let $(\Theta, \mathcal{D}, \mathcal{R})$ be the associated statistical decision problem, where $\mathcal{X}$ is the observation space of a random variable X having density $dP_\vartheta/d\mu = f(x|\vartheta)$ with respect to a σ-finite measure μ on $(\mathcal{X}, \mathcal{B})$; Θ is the parameter space; $\mathcal{D}$ is a decision space consisting of decision functions (decision procedures) $d(x)$, $x \in \mathcal{X}$, having finite risk function $\mathcal{R}$. Statistical inference on the unknown parameter $\vartheta \in \Theta$ or its function $g(\vartheta)$ then consists of choosing a decision procedure $d(x) \in \mathcal{D}$ possessing some optimal properties. We will consider a problem of estimation using a Bayesian approach and use the minimax criterion as the criterion of optimality.

Let $\mathcal{L}(\vartheta, d(X))$ be the loss function in estimating a function $g(\vartheta)$. The function $\mathcal{L}(\vartheta, d(X))$ represents the loss due to the statistician when ϑ is the true value of the parameter and $d(X)$ is the chosen estimator of $g(\vartheta)$.

The risk function associated with an estimator $d(X)$ in estimating the parameter ϑ is defined by

$$\mathcal{R}(\vartheta, d) = E_\vartheta[\mathcal{L}(\vartheta, d(X))] = \int_{\mathcal{X}} \mathcal{L}(\vartheta, d(x)) f(x|\vartheta)\mu(dx).$$

An estimator d^* is said to be minimax if

$$\sup_{\vartheta\in\Theta} \mathcal{R}(\vartheta, d^*) = \inf_{d\in\mathcal{D}} \sup_{\vartheta\in\Theta} \mathcal{R}(\vartheta, d). \tag{1}$$

Let π be a prior distribution of the parameter ϑ on the parameter space Θ. The distribution π will be also identified with the density $\pi(\vartheta)$ with respect to a σ-finite measure ν on Θ. Given a prior distribution π, the expected risk is

$$r(\pi, d) = E^\pi[\mathcal{R}(\vartheta, d)] = \int_\Theta \mathcal{R}(\vartheta, d) d\pi(\vartheta)$$
$$= \int_\Theta \int_{\mathcal{X}} \mathcal{L}(\vartheta, d(x)) f(x|\vartheta)\pi(\vartheta)\mu(dx)\nu(d\vartheta).$$

The posterior expected loss when $X = x$ is defined by

$$\varrho(\pi, d|x) = E^\pi[\mathcal{L}(\vartheta, d)|x] = \int_\Theta \mathcal{L}(\vartheta, d)\pi(\vartheta|x)\nu(d\vartheta),$$

where $\pi(\vartheta|x)$ denotes the posterior distribution of the parameter ϑ given $X = x$.

An estimator d^π satisfying

$$r(\pi, d^\pi) = \inf_{d\in\mathcal{D}} r(\pi, d)$$

is said to be Bayes with respect to the prior π. The value $r(\pi) = r(\pi, d^\pi) = \int_\Theta \mathcal{R}(\vartheta, d^\pi) d\pi(\vartheta)$ is then called the Bayes risk.

A prior π^* satisfying

$$\inf_{d\in\mathcal{D}} r(\pi^*, d) = \sup_{\pi\in\Pi} \inf_{d\in\mathcal{D}} r(\pi, d)$$

is said to be least favorable.

Let d^π be a Bayes estimator with respect to a prior distribution π. If the prior distribution π is such that

$$r(\pi) = r(\pi, d^\pi) = \sup_{\vartheta\in\Theta} \mathcal{R}(\vartheta, d^\pi),$$

then d^π is minimax; if d^π is the unique Bayes with respect to π, it is the unique minimax estimator; π is least favorable.

2.2 A method of finding minimax estimation procedures

If d_0 is a Bayes estimator with respect to a prior distribution π_0 and if $\mathcal{R}(\vartheta, d_0) \le r(\pi_0)$ for every ϑ in the support of π_0, d_0 is minimax and π_0 is the least favorable distribution. An estimator d_0 is said to be an equalizer estimator if $\mathcal{R}(\vartheta, d_0) = const$ for every $\vartheta \in \Theta$. If d_0 is an equalizer estimator and Bayes with respect to a prior distribution π, it is minimax.

Minimax estimators as well as least favorable distributions do not necessarily exist. If $\mathcal{D} \subset \mathbb{R}^k$ is a convex compact set and if $\mathcal{L}(\vartheta, d)$ is continuous and convex as a function of d for every $\vartheta \in \Theta$, there exists a nonrandomized minimax estimator.

Sufficient conditions for the existence of least favorable distributions are given by Kempthorne (1987). Suppose that the statistical decision problem satisfies the following assumptions:

(A1) For every distribution π on Θ, the Bayes procedure d^π with respect to π is unique a.s. for all ϑ;
(A2) For any sequence $(\pi_n), n = 1, 2, \ldots,$ of distributions on Θ converging weakly to π, the sequence $(\mathcal{R}(\cdot, d^{\pi_n})), n = 1, 2, \ldots,$ of the corresponding Bayes procedures converges uniformly on compacts to the risk function $R(\cdot, d^\pi)$ of the Bayes procedure d^π;
(A3) The parameter space Θ is a compact separable metric space;

(A4) The risk function $\mathcal{R}(\vartheta, d)$ for any decision procedure d is an upper-semicontinuous function of ϑ.

Kempthorne has proved the following theorems.

Theorem 1 *Under assumptions* (A1) *and* (A2), *if distribution* π^* *is least favorable*, *then*

$$r(\pi^*) = \max_{\vartheta\in\Theta} \mathcal{R}(\vartheta, d^{\pi^*}).$$

Theorem 2 *Under assumptions* (A1)–(A4), *a least favorable distribution* π^* *exists*.

Under the additional assumption

(A5) The parameter ϑ is real-valued and for any procedure d, the risk function $\mathcal{R}(\vartheta, d)$ is an analytic function of ϑ

the following theorem holds:

Theorem 3 *Suppose that the decision problem satisfies conditions* (A1)–(A5). *Then a least favorable distribution* π^* *has one of the following properties*:

(a) *The risk of the Bayes procedure* d^{π^*} *is constant on* Θ

or

(b) *The support of* π^* *is discrete and finite*.

3 Minimax estimators of the binomial probability under the LLF and the least favorable distributions

Let X be a random variable having the binomial distribution $\mathcal{B}(n, \vartheta)$. Consider the problem of estimation of the probability of success ϑ under the LLF

$$\mathcal{L}(\vartheta, d) = \beta\{\exp[\alpha(\vartheta - d)] - \alpha(\vartheta - d) - 1\}, \tag{2}$$

where $\alpha \neq 0$ and $\beta > 0$. In this case the sample space is $\mathcal{X} = \{0, 1, \dots, n\}$. Any function $d = d(x) : \mathcal{X} \mapsto [0, 1]$ is an estimate of the parameter ϑ.

The LLF is a convex and asymmetric function with respect to the error $\vartheta - d$. It is useful in the estimation problems when overestimation is considered more serious than underestimation or vice versa. The parameter β serves to scale the loss function and the parameter α serves to determine its shape. In the case $\alpha < 0$ the loss function indicates that overestimation is more costly than underestimation. The opposite is true, when $\alpha > 0$.

The risk function of an estimator $d(X)$ of ϑ is of the form

$$\mathcal{R}(\vartheta, d(X)) = \beta \sum_{x=0}^{n} \Big\{\exp[\alpha(\vartheta - d(x))] - \alpha(\vartheta - d(x)) - 1\Big\} \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x}. \tag{3}$$

Theorem 4 *There is no equalizer estimator available in the problem of estimating* ϑ *under the LLF.*

Proof Without loss of generality we can assume that $\beta = 1$. Suppose that d^* is an equalizer estimator, i.e., $\mathcal{R}(\vartheta, d^*(X)) = const$ for every $\vartheta \in [0, 1]$. It then holds in particular that

$$\mathcal{R}(0, d^*(X)) = \mathcal{R}(1, d^*(X)), \tag{4}$$

where

$$\mathcal{R}(0, d^*(X)) = \exp(-\alpha d^*(0)) + \alpha d^*(0) - 1,$$

$$\mathcal{R}(1, d^*(X)) = \exp[\alpha(1 - d^*(n))] - \alpha(1 - d^*(n)) - 1.$$

Thus, equality (4) holds if and only if

$$\exp(-\alpha d^*(0)) + \alpha d^*(0) = \exp[\alpha(1 - d^*(n))] - \alpha(1 - d^*(n)). \tag{5}$$

Suppose tentatively that $d^*(n) = d^*(0) + h$, where $h \in [0, 1]$, since $d^*(0) \in [0, 1]$ and $d^*(n) \in [0, 1]$. We will show later in Theorem 8 that it suffices to consider such estimators for which $d(x - 1) \le d(x)$, $x = 1, 2, \dots, n$. Equality (5) can be rewritten in the form

$$\exp(-\alpha d^*(0)) + \alpha d^*(0) = \exp[\alpha(1 - d^*(0) - h)] - \alpha(1 - d^*(0) - h)$$

or

$$\exp(-\alpha d^*(0))\{1 - \exp[\alpha(1 - h)]\} + \alpha(1 - h) = 0.$$

Denote

$$f(h) = \exp(-\alpha d^*(0))\{1 - \exp[\alpha(1 - h)]\} + \alpha(1 - h).$$

We show that $f(h)$ has a unique zero for $h = 1$. Note that

$$f'(h) = \alpha\{\exp[\alpha(1 - d^*(n))] - 1\}.$$

Since $d^*(n) \in [0, 1]$ we have the inequality $\exp[\alpha(1 - d^*(n))] \ge 1$ when $\alpha > 0$, or the inequality $\exp[\alpha(1 - d^*(n))] \le 1$ when $\alpha < 0$. Thus $f'(h) \ge 0$ for every $h \in [0, 1]$ (and $\alpha \neq 0$). The function $f(h)$ is strictly increasing on the interval $[0, 1]$ and $f(1) = 0$. But, for $h = 1$, it holds $d^*(0) = 0$, $d^*(n) = 1$ and then $\mathcal{R}(0, d^*) = 0$ and $\mathcal{R}(1, d^*) = 0$. It leads to the contradiction that d^* has constant risk, because there does not exist such an estimator for which the risk function could be equal to zero for every $\vartheta \in [0, 1]$. □

Lemma 1 *The Bayes estimator* d^π *associated with a prior distribution* π *in estimating* ϑ *under the LLF defined by* (2) *is of the form*

$$d^\pi(x) = \frac{1}{\alpha} E^\pi[\exp(\alpha\vartheta)|x]. \tag{6}$$

Proof It is well known that an estimator minimizing the expected risk $r(\pi, d)$ can be obtained by selecting, for every $x \in \mathcal{X}$, the value $d(x)$ which minimizes the posterior expected loss $\varrho(\pi, d|x) = E^\pi[\mathcal{L}(\vartheta, d(x))|x]$. The posterior expected loss has the following form

$$\varrho(\pi, d|x) = \beta \exp[-\alpha d(x)]E^\pi[\exp(\alpha\vartheta)|x] - \alpha\beta E^\pi(\vartheta|x) + \alpha\beta d(x) - \beta$$

and it attains its minimum for d^π defined by (6). □

It is easy to verify that the statistical decision problem considered satisfies conditions (A1)–(A5). It follows from Theorem 3 that in the case when no equalizer Bayes decision is available, then the least favorable distribution is discrete on a finite support. Thus, let the prior distribution π be represented by a finite sequence $\pi = (\vartheta_i, p_i)$, $i = 1, \dots, m$, where ϑ_i are the supporting points and $p_i = \pi(\vartheta_i)$. For a function $h(\vartheta)$ the posterior expectation $E^\pi\{h(\vartheta)|x\}$ is determined by

$$E^\pi\{h(\vartheta)|x\} = \frac{\int_\Theta h(\vartheta) f(x|\vartheta)\pi(\vartheta)\nu(d\vartheta)}{\int_\Theta f(x|\vartheta)\pi(\vartheta)\nu(d\vartheta)}.$$

Hence, by Lemma 1 the Bayes estimator with respect to the finite supported prior π can be evaluated by the formula

$$d^\pi(x) = \frac{1}{\alpha}\log\left[\frac{\sum_{i=1}^m p_i \exp(\alpha\vartheta_i)\vartheta_i^x(1-\vartheta_i)^{n-x}}{\sum_{i=1}^m p_i \vartheta_i^x(1-\vartheta_i)^{n-x}}\right],$$

$x = 0, 1, \dots, n$. The expected risk is

$$r(\pi) = r(\pi, d^\pi) = \sum_{i=1}^m p_i \mathcal{R}(\vartheta_i, d^\pi),$$

where the risk function $\mathcal{R}(\vartheta, d)$ is defined by (3).

Theorem 5 *Let $d_\alpha(X)$ be the minimax estimator of ϑ indexed by the parameter α of the LLF defined by (2). Then $1 - d_\alpha(n - X)$ is the minimax estimator of ϑ under the LLF with the parameter $-\alpha$.*

Proof Let $\mathcal{R}_\alpha(\vartheta, d)$ denote the risk function associated with an estimator d under the LLF defined by (2). Then

$$\begin{aligned}\mathcal{R}_\alpha(\vartheta, d) &= \sum_{x=0}^n \mathcal{L}(\vartheta, d(x))\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} \\ &= \sum_{x=0}^n \big\{\exp[\alpha(\vartheta - d(x))] - \alpha(\vartheta - d(x)) - 1\big\} \\ &\quad \times \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} \\ &= \sum_{x=0}^n \big\{\exp[(-\alpha)((1-\vartheta) - (1-d(x)))] \\ &\quad - (-\alpha)((1-\vartheta) - (1-d(x))) - 1\big\} \\ &\quad \times \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}.\end{aligned}$$

Denote $1 - \vartheta = \eta$ and $x = n - y$. Then, taking into account the relation $\binom{n}{n-y} = \binom{n}{y}$ we have

$$\begin{aligned}\mathcal{R}_\alpha(\vartheta, d) &= \sum_{y=0}^n \big\{\exp[(-\alpha)(\eta - (1 - d(n-y)))] \\ &\quad - (-\alpha)(\eta - (1 - d(n-y))) - 1\big\} \\ &\quad \times \binom{n}{n-y}(1-\eta)^{n-y}\eta^y \\ &= \sum_{y=0}^n \big\{\exp[(-\alpha)(\eta - (1 - d(n-y)))] \\ &\quad - (-\alpha)(\eta - (1 - d(n-y))) - 1\big\} \\ &\quad \times \binom{n}{y}\eta^y(1-\eta)^{n-y}.\end{aligned}$$

Thus we have

$$\mathcal{R}_\alpha(\vartheta, d(X)) = \mathcal{R}_{-\alpha}(\eta, 1 - d(n - X)).$$

Therefore, if $d_\alpha(X)$ is the estimator satisfying

$$\sup_{\vartheta\in[0,1]} \mathcal{R}_\alpha(\vartheta, d_\alpha) = \inf_d \sup_{\vartheta\in[0,1]} \mathcal{R}_\alpha(\vartheta, d),$$

then $1 - d_\alpha(n - X)$ is the estimator satisfying

$$\sup_{\eta\in[0,1]} \mathcal{R}_{-\alpha}(\eta, 1 - d_\alpha(n - X)) = \inf_d \sup_{\eta\in[0,1]} \mathcal{R}_{-\alpha}(\eta, d).$$ □

Theorem 6 *Let $\pi_\alpha^* = (\vartheta_i^*, p_i^*)$, $i = 1, \dots, m$ be the least favorable distribution in the problem of estimating the binomial parameter ϑ under the LLF defined by (2). Then $\pi_{-\alpha}^* = (1 - \vartheta_{m-i+1}^*, p_{m-i+1}^*)$, $i = 1, \dots, m$, is the least favorable distribution under the LLF with the parameter $-\alpha$.*

The proof follows from the arguments of Theorem 5.

Theorem 7 *The support of the least favorable prior distribution consists of at most $2(n + 1)$ points.*

The proof follows from a standard analysis of the behavior of the risk function.

Theorem 8 *In the problem of estimating the binomial parameter ϑ under the LLF the class of nondecreasing estimators is essentially complete.*

Table 1 The values of the supporting points and the corresponding probability mass of the least favorable distributions

$\alpha = 1, \quad \beta = 1$

$n = 4$		$n = 5$		$n = 6$		$n = 7$	
ϑ	p	ϑ	p	ϑ	p	ϑ	p
0.115045	0.304812	0.098213	0.231670	0.083406	0.171014	0.061590	0.103670
0.511592	0.443151	0.421193	0.408873	0.345288	0.356439	0.250156	0.262033
0.890022	0.252037	0.785787	0.309668	0.670383	0.332376	0.509354	0.317859
		0.994886	0.049789	0.920159	0.140171	0.763033	0.233119
						0.941470	0.083318

$n = 8$		$n = 9$		$n = 10$		$n = 12$	
ϑ	p	ϑ	p	ϑ	p	ϑ	p
0.064970	0.096789	0.052997	0.063641	0.052959	0.055541	0.045290	0.033235
0.252221	0.262292	0.200650	0.195848	0.194727	0.184478	0.158738	0.129360
0.507457	0.325878	0.409087	0.283111	0.395938	0.282031	0.322034	0.230368
0.758418	0.235629	0.632581	0.260012	0.616240	0.269745	0.510201	0.264028
0.937683	0.079412	0.825944	0.153410	0.812880	0.162421	0.695482	0.209144
		0.954818	0.043979	0.949112	0.045785	0.851726	0.107859
						0.957932	0.026005

$\alpha = 10, \quad \beta = 1$

$n = 4$		$n = 5$		$n = 6$		$n = 7$	
ϑ	p	ϑ	p	ϑ	p	ϑ	p
0.119954	0.591660	0.117328	0.546959	0.131185	0.503541	0.000001	0.068686
0.668179	0.338173	0.602829	0.362760	0.567077	0.383616	0.180326	0.425193
0.953222	0.070167	0.913976	0.090281	0.887276	0.112843	0.542162	0.354303
						0.831275	0.133338
						0.972870	0.018479

$n = 8$		$n = 9$		$n = 10$		$n = 12$	
ϑ	p	ϑ	p	ϑ	p	ϑ	p
0.008478	0.066235	0.057922	0.142975	0.062288	0.131341	0.000001	0.009760
0.182709	0.385401	0.243616	0.348784	0.239450	0.324399	0.097136	0.146125
0.504388	0.355383	0.528081	0.321541	0.497045	0.317057	0.266846	0.290228
0.784056	0.158636	0.782600	0.151913	0.738122	0.170756	0.484734	0.290556
0.947547	0.034345	0.942909	0.034786	0.906738	0.052858	0.695180	0.178624
				0.994385	0.003589	0.858230	0.070799
						0.960422	0.013908

Proof The binomial distribution belongs to the family of distributions having monotone likelihood ratio. The action space $[0, 1]$ is a closed subset of $\mathbb{R}$ and $\Theta = [0, 1]$ is an interval in $\mathbb{R}$. Moreover it is easy to see that the LLF $\mathcal{L}(\vartheta, a)$ attains its minimum as a function of a at a point $a = g(\vartheta)$, where g is an increasing function of ϑ and that $\mathcal{L}(\vartheta, a)$, considered as a function of a, increases as a moves away from $g(\vartheta)$. Thus the conditions of a statistical decision problem to be monotone are satisfied for $g(\vartheta) = \vartheta$ (for the notion of monotone decision problem see Berger 1985). From the result of Brown et al. (1976) it follows that for a monotone decision problem, the class of monotone decision procedures is essentially complete. □

4 Numerical determination of minimax estimators and least favorable prior distributions

The values of minimax estimators d^* under the LLF can only be determined by using numerical methods. This can

Table 2 The values of minimax estimators

$\alpha = 1,\quad \beta = 1$								
x	n							
	4	5	6	7	8	9	10	12
0	0.171461	0.158612	0.148550	0.140363	0.133515	0.127665	0.122585	0.114138
1	0.345228	0.301754	0.270569	0.246878	0.228138	0.212864	0.200120	0.179947
2	0.513890	0.442269	0.391043	0.352402	0.322086	0.297579	0.277298	0.245545
3	0.677915	0.580278	0.510017	0.456955	0.415368	0.381818	0.354121	0.310925
4	0.837808	0.715927	0.627542	0.560560	0.507995	0.465584	0.430592	0.376105
5		0.849352	0.743672	0.663238	0.599977	0.548884	0.506714	0.441056
6			0.858458	0.765014	0.691326	0.631723	0.582492	0.505826
7				0.865910	0.782055	0.714109	0.657928	0.570358
8					0.872174	0.796048	0.733027	0.634712
9						0.877546	0.807792	0.698841
10							0.882227	0.762780
11								0.826511
12								0.890045
r^*	0.013894	0.011939	0.010507	0.009406	0.008529	0.007813	0.007216	0.006273
$\alpha = 10,\quad \beta = 1$								
x	n							
	4	5	6	7	8	9	10	12
0	0.251802	0.210688	0.194929	0.178708	0.167488	0.158218	0.150420	0.137821
1	0.479403	0.400088	0.341438	0.309334	0.281847	0.259704	0.241524	0.213533
2	0.647871	0.566501	0.500029	0.434416	0.390167	0.356434	0.329251	0.287277
3	0.747961	0.658622	0.592912	0.541108	0.491680	0.448773	0.413229	0.358702
4	0.882458	0.777481	0.691517	0.630874	0.577427	0.532180	0.492317	0.427683
5		0.875521	0.805459	0.725189	0.663542	0.610664	0.566337	0.493999
6			0.867371	0.807059	0.743562	0.688231	0.638596	0.557896
7				0.887669	0.818890	0.758652	0.706944	0.620070
8					0.892347	0.829508	0.773342	0.680059
9						0.895786	0.837392	0.738625
10							0.899525	0.795521
11								0.851089
12								0.905344
r^*	1.48	1.23896	1.07	0.954526	0.862216	0.788	0.726481	0.630242

be done by one of the methods of constrained optimization, for example by the Nelder and Mead (1965) simplex method or by the Box (1965) complex method. The Kempthorne (1987) algorithm determines the least favorable priors and the minimax estimators.

To evaluate the minimax estimators d^* we have implemented the Box complex method and the Kempthorne algorithm. The algorithm constructed and based on the Box complex method relies on the condition for minimaxity defined by (1) and uses the monotone property of estimators given by Theorem 8. The other algorithm used is a slight modification of the Kempthorne algorithm. The only difference relies on using at each step a procedure searching for a global maximum instead of searching for local maxima. The Kempthorne algorithm relies on the determination of the least favorable distribution and is based on Theorems 1 and 2. To illustrate the method we describe the algorithm. The algorithm we have used has four basic steps which iterate if necessary until the solution is achieved with a given accuracy.

Step 1: Specify π_0 as an initial guess for the least favorable distribution. Let $\{\vartheta_{i,0}, i = 1, 2, \ldots, m_0\}$ be the points of the support of the distribution π_0 and let $\{p_{i,0}, i = 1, 2, \ldots, m_0\}$ be the corresponding probabilities, such

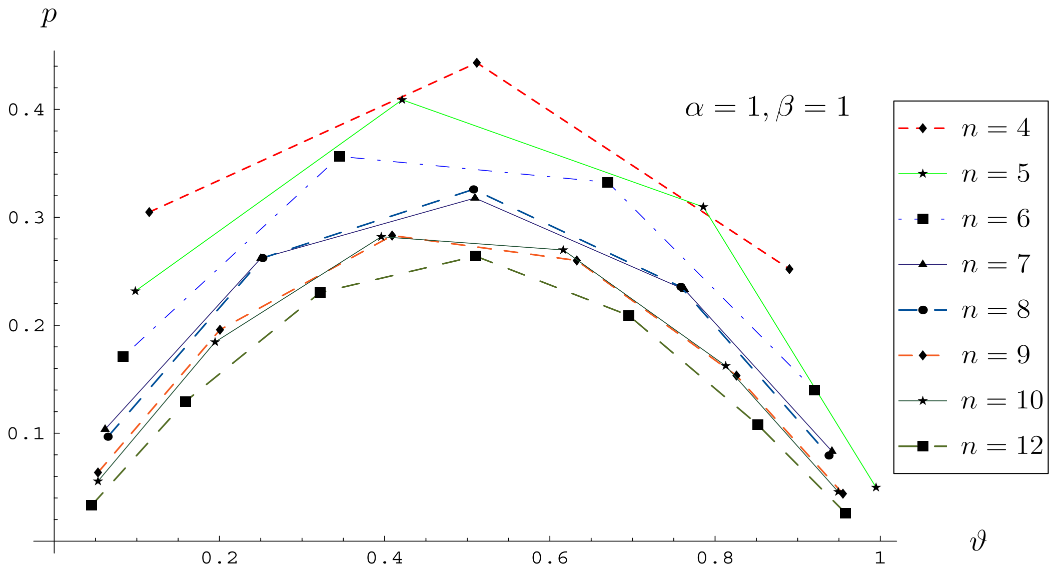

Fig. 1 The probability mass points of the support of the least favorable distribution for various n and $\alpha = 1$, $\beta = 1$

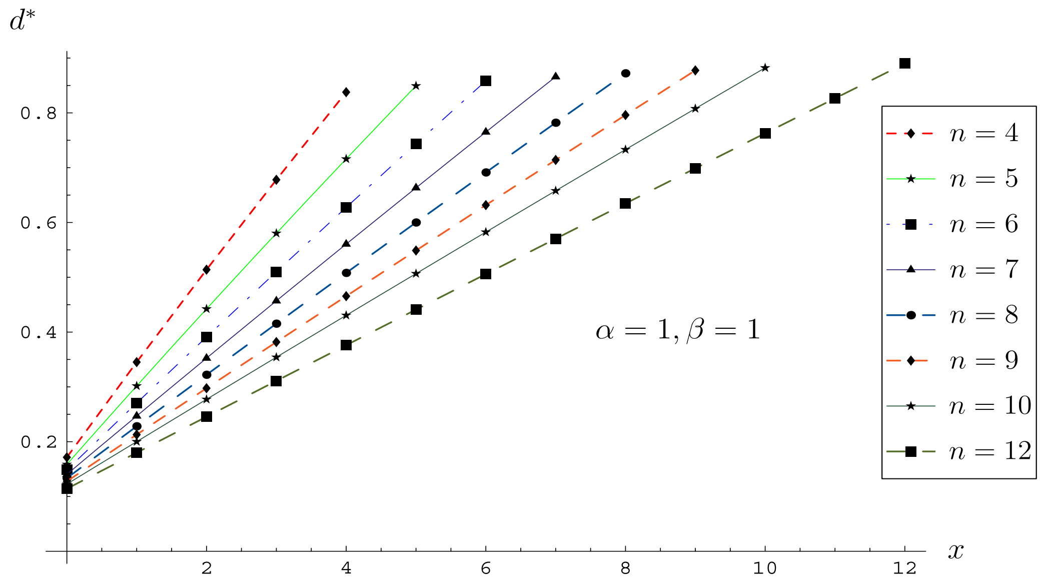

Fig. 2 The values of minimax estimators for various n and $\alpha = 1$, $\beta = 1$

that

$$\pi_0(\vartheta_{i,0}) = p_{i,0}, \quad p_{i,0} > 0 \text{ and } \sum_{i=1}^{m_0} p_{i,0} = 1.$$

The choice of the initial least favorable distribution does not affect the result of the algorithm. As the initial distribution π_0 we have chosen the two-point distribution: $\pi_0(0) = \pi_0(1) = 0.5$.

Step 2: Determine π_1 as a global maximum of the Bayes risk $r(\pi)$, where π is a prior distribution. The Bayes risk with respect to the prior distribution π with the support size m_0 is a function of $2m_0$ arguments $(\vartheta_1, \vartheta_2, \ldots, \vartheta_{m_0}, p_1, p_2, \ldots, p_{m_0})$ for which $\pi(\vartheta_i) = p_i$, $i = 1, 2, \ldots, m_0$. Starting from the values given by the distribution π_0, find the distribution π_1, supported on m_0 points, which maximizes (globally) the Bayes risk. This is the constrained maximization of a function on $\mathbb{R}^{2m_0}$, subject to the constraints:

$$p_i \geq 0, \quad i = 1, 2, \ldots, m_0, \ \sum_{i=1}^{m_0} p_i = 1,$$

$$\vartheta_i \in \Theta = [0, 1], \quad i = 1, 2, \ldots, m_0.$$

Step 3: Find the parameter value ϑ_1^* maximizing the risk function of the Bayes estimator d^{π_1}. If $\mathcal{R}(\vartheta_1^*, d^{\pi_1}) =$

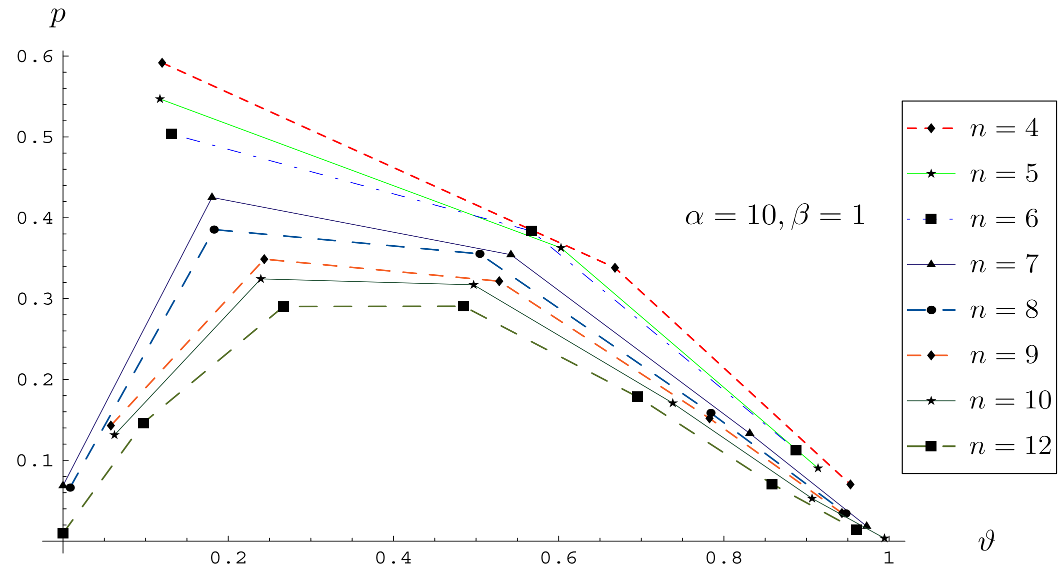

Fig. 3 The probability mass points of the support of the least favorable distribution for various n and $\alpha = 10$, $\beta = 1$

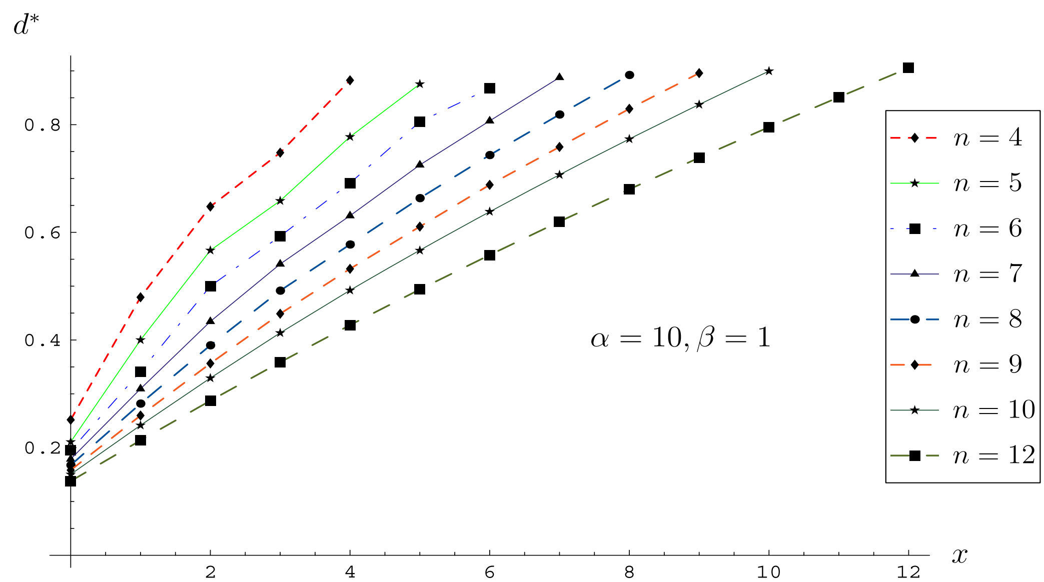

Fig. 4 The values of minimax estimators for various n and $\alpha = 10$, $\beta = 1$

$r(\pi_1)$ (we have assumed the relative accuracy to be less than 10^{-3}), then stop; the distribution π_1 is least favorable and d^{π_1} is minimax. Otherwise, proceed to the next step.

Step 4: Specify a new guess $\pi_{0,\text{new}}$ for the least favorable distribution and return to Step 2. The distribution $\pi_{0,\text{new}}$ is supposed to be a mixture of the distributions π_1 and π_1^*, the point distribution at ϑ_1^*. Specifically, for some $a > 0$ (e.g. $a = 0.1$), put

$$\pi_{0,\text{new}}(\vartheta) = \begin{cases} (1-a)p_{i,1}, & \vartheta = \vartheta_{i,1}, i = 1, 2, \ldots, m_0, \\ a, & \vartheta = \vartheta_1^*. \end{cases}$$

The support size of the distribution $\pi_{0,\text{new}}$ is now $m_{0,\text{new}} = m_0 + 1$.

In determining the minimax estimators we wanted simultaneously to examine the procedures available in Mathematica 5.2 (Licence L4615-2807, Premier Service) in respect to their satisfactory functionality, under the algorithm assumed, for solving the advanced numerical problem of mathematical statistics. In constructing a computer program, we implemented the NMaximize procedure for finding constrained global optima.

We present some numerical results obtained by using the computer program based on the Kempthorne method of find-

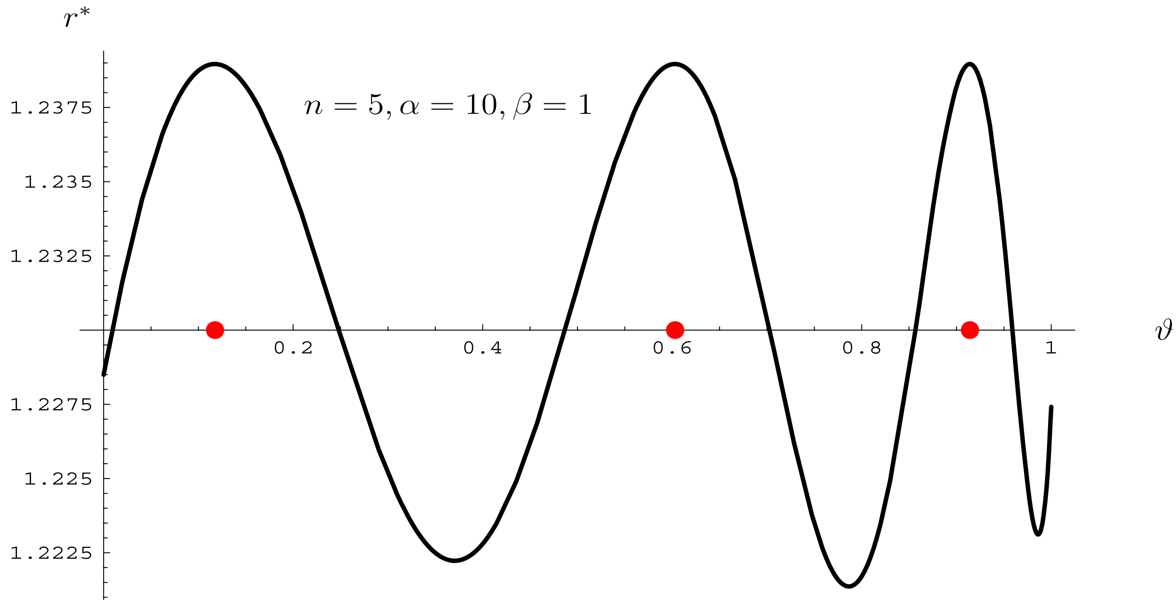

Fig. 5 The plot of the risk function associated with the minimax estimator for $n = 5$, $\alpha = 10$, $\beta = 1$, and the supporting points of the least favorable distribution

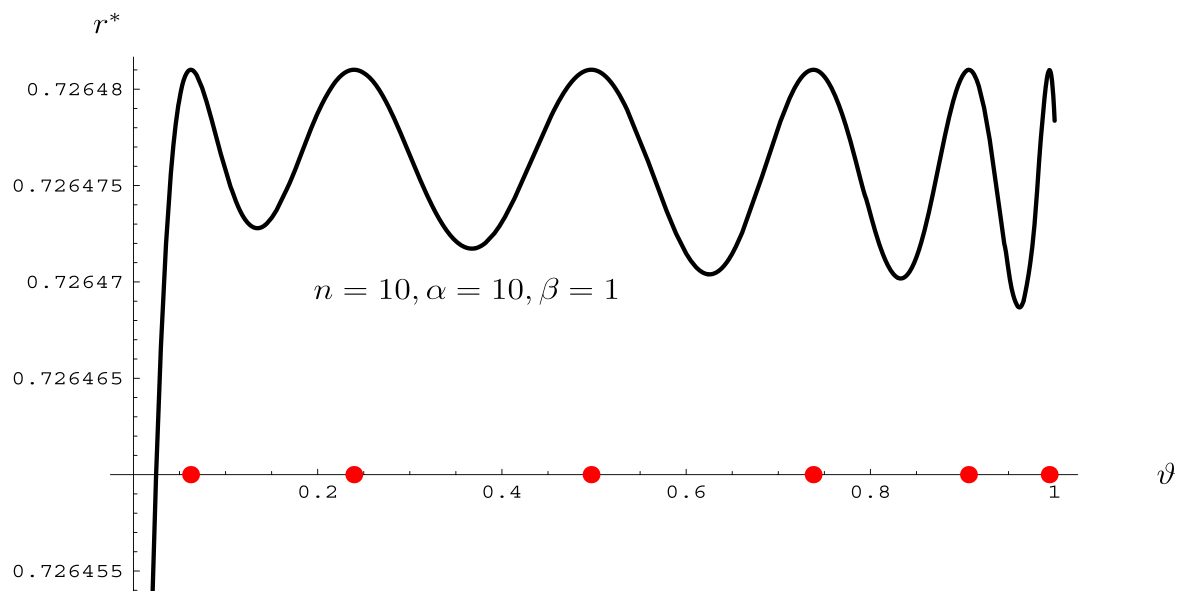

Fig. 6 The plot of the risk function associated with the minimax estimator for $n = 10$, $\alpha = 10$, $\beta = 1$, and the supporting points of the least favorable distribution

ing the least favorable prior distributions. Table 1 contains the values of the supporting points and the corresponding probability mass of the least favorable distributions for some numbers of n and for the LLF parameters $\alpha = 1$, 10 and $\beta = 1$. The values of the minimax estimators of the probability of success for the corresponding values of n, α, β are given in Table 2. The probability mass points of the support of the least favorable distributions for various n and $\alpha = 1$, 10, $\beta = 1$ are then plotted on Figs. 1 and 3. The values of the corresponding minimax estimators are presented on Figs. 2 and 4. The plots of the risk functions associated with the minimax estimators for $n = 15$, 10, 12 and $\alpha = 10$, $\beta = 1$ are shown on Figs. 5–7.

It is evidently seen that the minimax estimators and the least favorable prior distributions remain the same if only the parameter β of the LLF is changed. The change of the parameter β affects only the value of the risk $r^* = r(\pi^*)$. Therefore the sample of the results presents the minimax es-

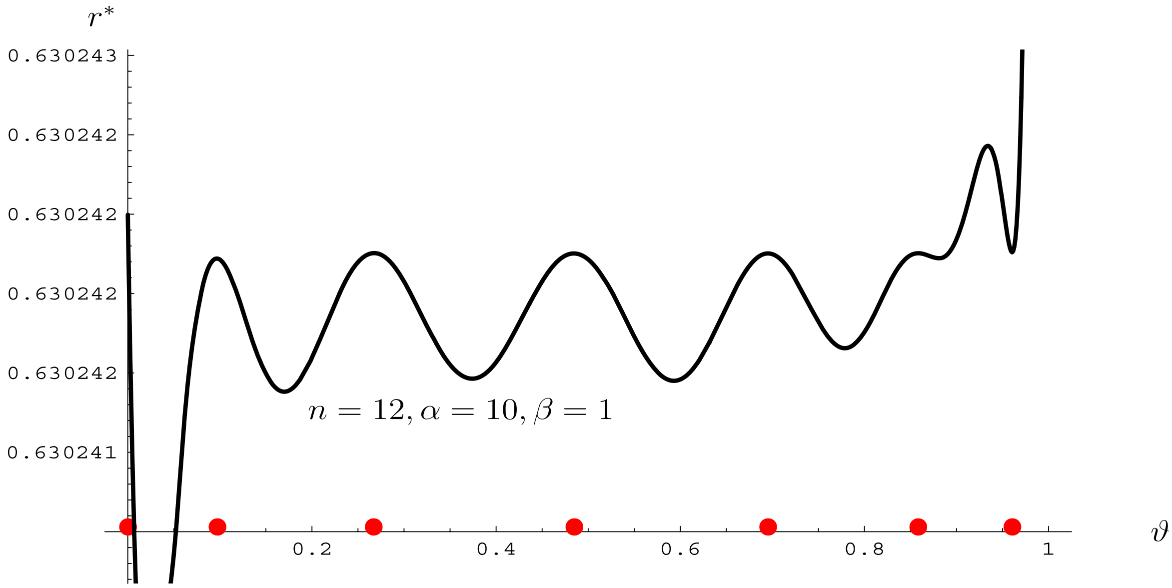

Fig. 7 The plot of the risk function associated with the minimax estimator for $n = 12$, $\alpha = 10$, $\beta = 1$, and the supporting points of the least favorable distribution

timators and the corresponding least favorable distributions in the case when $\beta = 1$. The values of minimax estimators under the LLF with opposite sign of α can be obtained by Theorem 5.

5 Applications

The problem of finding optimal, under various criteria, estimators of a probability of success is of significant importance in reliability systems.

An important application of the results obtained in the paper is the use of them in a nonparametric problem. In the paper of Jokiel-Rokita and Magiera (2007) it is shown that some problems of finding minimax estimators of a cumulative distribution function under convex loss function can be solved by converting the nonparametric problem to the parametric one. Let $\mathbf{X}$ be a sample from a cumulative distribution function $F(t)$ defined on $\mathbb{R}$. No assumptions are imposed on the unknown function F. Denote by $\delta(t;\mathbf{X})$ an estimator of F and take into account the following loss function

$$\mathcal{L}_{NP}(F,\delta(t;\mathbf{X})) = \int_{\mathbb{R}} G(F(t) - \delta(t;\mathbf{X}))H(F(t))dW(t), \qquad (7)$$

where $G(\Delta)$, $\Delta = F - \delta$, is a convex function such that $G(\Delta)$ is decreasing for $\Delta \in (-\infty, 0)$, increasing for $\Delta \in (0,\infty)$ and $G(0) = 0$; $H(\cdot)$ is a continuous and positive function. The function W is a given non-null finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The measure W is treated as a weight function. The risk function of any estimator δ of F is defined by $\mathcal{R}(F,\delta) = E_F[\mathcal{L}_{NP}(F,\delta)]$. An estimator δ^0 of F satisfying the condition $\sup_F \mathcal{R}(F,\delta^0) = \inf_\delta \sup_F \mathcal{R}(F,\delta)$ is said to be minimax. Denote by $(X_1^*,\ldots,X_n^*)$ the ordered sample of $\mathbf{X}$. It follows from Theorem 1 of Jokiel-Rokita and Magiera (2007) that the minimax estimator of F under the loss function defined by (7) is of the form

$$\delta^0(t;\mathbf{X}) = \sum_{i=0}^{n} d(i)\mathbf{1}_{[X_i^*,X_{i+1}^*)}(t),$$

where $d(i)$, $i = 0,1,\ldots,n$, is the minimax estimator under the loss function

$$\mathcal{L}_B(\vartheta,d) = G(\vartheta - d)H(\vartheta)$$

in the problem of estimating the probability of success ϑ of the binomial distribution $\mathcal{B}(n,\vartheta)$. In particular, for $H(\vartheta) \equiv 1$ and $G(\vartheta - d) = \beta\{\exp[\alpha(\vartheta - d)] - \alpha(\vartheta - d) - 1\}$, the minimax estimates $d(i)$, $i = 0,1,\ldots,n$, evaluated under the LLF given by (2) in the parametric problem, determine the minimax estimator $\delta^0(t;\mathbf{X})$ of an unknown cumulative distribution function $F(t)$ under the loss function

$$\mathcal{L}_{NP}(F,\delta(t;\mathbf{X})) = \beta\int_{\mathbb{R}}\big\{\exp[\alpha(F(t) - \delta(t;\mathbf{X}))] - \alpha(F(t) - \delta(t;\mathbf{X})) - 1\big\}dW(t).$$

The result can also be used in the case when the cumulative distribution function F has a nonnegative support as well as to determine the estimators of unknown survival functions.

References

Berger, J.: Statistical Decision Theory and Bayesian Analysis, 2nd edn. Springer, New York (1985)

Box, M.J.: A new method of constrained optimization and a comparison with other methods. Comput. J. **8**, 47–52 (1965)

Brown, L., Cohen, A., Strawderman, W.: A complete class theorem for strict monotone likelihood ratio with applications. Ann. Math. Stat. **4**(4), 712–722 (1976)

Eichenauer-Herrmann, J., Gohout, W., Lehn, J.: Minimax estimation of a binomial probability under weighted absolute error loss. Stat. Decis. **8**, 37–45 (1990)

Hodges, J., Lehmann, E.: Some problems in minimax point estimation. Ann. Math. Stat. **21**, 182–197 (1950)

Jokiel-Rokita, A., Magiera, R.: Minimax estimation of a cumulative distribution function by converting to a parametric problem. *Metrika* (2007, to appear)

Kempthorne, P.J.: Numerical specification of discrete least favorable prior distributions. SIAM J. Sci. Stat. Comput. **8**(2), 171–184 (1987)

Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**, 308–313 (1965)

Rukhin, A.: Minimax estimation of the binomial parameter under entropy loss. Stat. Decis. **13**(suppl.), 69–81 (1993)

Wieczorkowski, R.: Calculating the minimax estimation of a binomial probability with entropy loss function and its comparison with other estimators of a binomial probability. Stat. Decis. **16**, 289–298 (1998)

Wieczorkowski, R., Zieliński, R.: Minimax estimation of binomial probability with entropy loss function. Stat. Decis. **10**, 39–44 (1992)