APPLICATIONES MATHEMATICAE
25,1 (1998), pp. 85–100

L. GAJEK and B. MIZERA-FLORCZAK (Łódź)

INFORMATION INEQUALITIES FOR THE MINIMAX RISK OF SEQUENTIAL ESTIMATORS (WITH APPLICATIONS)

Abstract. Information inequalities for the minimax risk of sequential estimators are derived in the case where the loss is measured by the squared error of estimation plus a linear functional of the number of observations. The results are applied to construct minimax sequential estimators of: the failure rate in an exponential model with censored data, the expected proportion of uncensored observations in the proportional hazards model, the odds ratio in a binomial distribution and the expectation of exponential type random variables.

1. Introduction. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random vectors (i.i.d. r.v.'s) in $\mathbb{R}^l$ each with probability distribution P_θ with $\theta \in \Theta$, where Θ is an open interval of reals. Assume that the family $\{P_\theta : \theta \in \Theta\}$ is dominated by some σ-finite measure μ on $\mathbb{R}^l$ and let $p_\theta(x)$ denote $dP_\theta/d\mu$ at the point $x \in \mathbb{R}^l$.

In this paper we consider minimax estimation of the parameter θ^s, where $s \neq 0$ is a given real number, under squared error loss L with a weight $h(\theta)$:

$$L(\tau, \theta) = (\tau - \theta^s)^2 h(\theta).$$

We shall investigate estimators T_N of θ^s under a sequential sampling scheme with the random variable N denoting the number of observations. Let $c(\theta)$ denote the average cost of a single observation. Usually it is assumed that $c(\theta)$ does not depend on θ, but it is more natural to assume that the cost of observing each X_i is a function of X_i, say $\xi(X_i)$. Then the average cost of observing the whole sample $X_1, \ldots, X_N$, where N is a stopping time, is

1991 *Mathematics Subject Classification*: 62C20, 62F10, 62L12.
Key words and phrases: minimax estimation, information inequalities, censored data, proportional hazard model, odds ratio.

equal to

$$\mathrm{E}_\theta\Big[\sum_{i=1}^{N}\xi(X_i)\Big] = \mathrm{E}_\theta N \cdot \mathrm{E}_\theta \xi(X_1)$$

by the Wald lemma. Hence the total risk of the sequential procedure T_N is

$$R(T_N,\theta) = \mathrm{E}_\theta[L(T_N,\theta)] + c(\theta)\mathrm{E}_\theta N, \tag{1}$$

where $c(\theta) = \mathrm{E}_\theta\xi(X_1)$.

The main aim of this article is to provide general information inequalities for the minimax value when the risk of sequential estimators is of the form (1) and when the parameter space is $\Theta = (0,\theta_1)$ for some $0 < \theta_1 \le \infty$ or $\Theta = (\theta_2,\infty)$ for some $0 \le \theta_2 < \infty$ (see Theorem 2.1 in Section 2). These bounds are applicable to a wide range of estimation problems.

First we use the results obtained to estimate the exponential mean lifetime in the model with censored data under the risk defined by (1). There are several proposals of sequential procedures in the above model. Gardiner and Susarla (1984) and Gardiner, Susarla, and van Ryzin (1986) proposed sequential asymptotically risk efficient procedures. Some asymptotic distribution results for the procedures introduced in Gardiner, Susarla and van Ryzin (1986) can be found in Gardiner and Susarla (1991). Bayesian sequential estimation with censored data was investigated in Tahir (1988). However, very few papers concern minimax estimation from censored data. In particular, minimax estimation in the exponential failure time model under the presence of a censoring mechanism was considered by Gajek and Gather (1991) for the case of fixed sample size. They gave a lower bound on the minimax risk but a minimax estimator was not found. The problem of minimax sequential estimation in the model considered by Gajek and Gather (1991) was investigated in Mizera (1996) under some additional restrictions on the expected number of observations. One of the motivations for the present paper is to construct a minimax estimator in the same model provided that the sample size is randomly chosen and that the risk function incorporates the cost of observations (see Section 3).

In Section 4 we investigate the problem of estimating the expected proportion of uncensored observations in the proportional hazards model. Recall that subject to this model the distribution function G of the censoring random variable Y satisfies the equation

$$1 - G(y) = [1 - F(y)]^d \quad \text{for all } y,$$

where F denotes the distribution function of the censored r.v. X and $d > 0$ is the censoring parameter. The expected proportion of the uncensored observations, θ, is then equal to $1/(d+1)$. In Section 4 we propose a simple sequential estimator of θ and prove its minimaxity by applying the bounds proven in Section 2.

In Section 5 we apply Theorem 2.1 to the problem of minimax estimation of the expectation of a one-parameter exponential type family of probability distributions. Bayesian sequential estimation in an exponential family was considered e.g. by Mayer Alvo (1977). Using Bayesian methods Magiera (1977) investigated minimax estimation with the cost function depending only on time for an exponential class of processes with continuous time. He showed that under some additional assumptions the fixed-time plan is minimax. We consider discrete time exponential processes under the condition that the cost is some real function possibly depending on an unknown parameter.

Finally, we consider the problem of estimating the odds ratio θ in a binomial distribution. In Section 5 we propose a simple sequential estimator of θ and derive its minimaxity from the information inequality of Theorem 2.1.

2. Lower bounds on the minimax value. Let N be a random variable defined on the same probability space $(\Omega, \mathcal{S}, \mathbb{P}_\theta)$ as the sequence $X_1, X_2, \ldots$, where $\mathbb{P}_\theta$ is the product measure generated by P_θ. Let $\sigma(X_1, \ldots, X_n)$ denote the σ-algebra generated by the finite sequence $X_1, \ldots, X_n$. If the r.v. N is integer-valued and

(i) $\{\omega \in \Omega : N(\omega) \le n\} \in \sigma(X_1, \ldots, X_n)$ for $n = 1, 2, \ldots$,

(ii) $\mathbb{P}_\theta(\{\omega \in \Omega : N(\omega) < \infty\}) = 1$,

then N is called a stopping time (see e.g. Chow, Robbins and Siegmund (1971)) or a *proper stopping time* (see Woodroofe (1982)). Let $T_n = T_n(X_1, \ldots, X_n)$ be an estimator of θ^s, $s \ne 0$, based on n observations $X_1, \ldots, X_n$. Having a sequence $(T_n,\ n = 1, 2, \ldots)$ of statistics and the stopping time N, we construct a sequential estimator T_N. Throughout the paper we assume that the following Cramér–Rao–Wolfowitz inequality holds (see e.g. Wolfowitz (1947)):

$$\mathrm{E}_\theta[T_N - \theta^s]^2 \ge b^2(\theta) + \frac{[s\theta^{s-1} + b'(\theta)]^2}{I(\theta)\mathrm{E}_\theta N} \quad \text{for all } \theta \in \Theta, \tag{2}$$

where $b(\theta) = \mathrm{E}_\theta T_N - \theta^s$ and $I(\theta) = \mathrm{Var}_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(X_1)\right]$.

Let $h(\cdot)$ be a positive weight function and $c(\cdot)$ be a positive cost function. Then we have the following

THEOREM 2.1. *Assume that* (2) *holds and* $\Theta = (0, \theta_1)$ *for some* $0 < \theta_1 \le \infty$ (*resp.* $\Theta = (\theta_2, \infty)$ *for some* $0 \le \theta_2 < \infty$).

(i) *Assume that* $0 < c(\theta) \le s^{-2}\theta^{2s+2} I(\theta) h(\theta)$ *for all* $\theta \in \Theta$ *such that* $\theta < \delta$, *for some* $\delta > 0$ (*resp.* $\theta > \kappa$, *for some* $\kappa > 0$). *If the limits as* $\theta \to 0$

(*resp.* $\theta \to \infty$) *of* $h(\theta)\theta^{2s}$ *and* $\theta^{-2}c(\theta)/I(\theta)$ *exist and are finite*, *then*

$$\limsup_{\substack{\theta\to 0\\(\theta\to\infty)}}\{\mathrm{E}_\theta[T_N-\theta^s]^2h(\theta)+c(\theta)\mathrm{E}_\theta N\}$$
$$\geq \lim_{\substack{\theta\to 0\\(\theta\to\infty)}} |s|\theta^{s-1}\sqrt{\frac{c(\theta)h(\theta)}{I(\theta)}}\left(2-\frac{|s|}{\theta^{s+1}}\sqrt{\frac{c(\theta)}{I(\theta)h(\theta)}}\right).$$

(ii) *Assume that* $c(\theta)\geq s^{-2}\theta^{2s+2}I(\theta)h(\theta)$ *for all* $\theta\in\Theta$ *such that* $\theta<\delta$, *for some* $\delta>0$ (*resp.* $\theta>\kappa$, *for some* $\kappa>0$). *If the limit as* $\theta\to 0$ (*resp.* $\theta\to\infty$) *of* $h(\theta)\theta^{2s}$ *exists and is finite*, *then*

$$\limsup_{\substack{\theta\to 0\\(\theta\to\infty)}}\{\mathrm{E}_\theta[T_N-\theta^s]^2h(\theta)+c(\theta)\mathrm{E}_\theta N\}\geq \lim_{\substack{\theta\to 0\\(\theta\to\infty)}} h(\theta)\theta^{2s}.$$

In the proof we shall need the following three lemmas.

LEMMA 2.2. *For all* $A,B,z>0$, *we have*

$$\frac{A}{z}+Bz\geq 2\sqrt{AB}.$$

LEMMA 2.3. *For all* $D>0$ *and* $z\in\mathbb{R}$, *we have*

$$z^2+D|1+z|\geq\begin{cases} D-D^2/4 & \text{for } 0<D\leq 2,\\ 1 & \text{for } D\geq 2.\end{cases}$$

LEMMA 2.4 (extended L'Hospital rule). *Let* $x_0\in[a,b]$ *and* $\mathcal{D}=(a,b)\setminus\{x_0\}$. *Assume that* $f,g:\mathcal{D}\to\mathbb{R}$ *are differentiable. If* $g'(x)\neq 0$ *for every* $x\in\mathcal{D}$ *and*

$$\lim_{x\to x_0} f(x)=\lim_{x\to x_0} g(x)=0,\ +\infty \text{ or } -\infty,$$

then

$$\liminf_{x\to x_0}\frac{f'(x)}{g'(x)}\leq\liminf_{x\to x_0}\frac{f(x)}{g(x)}\leq\limsup_{x\to x_0}\frac{f(x)}{g(x)}\leq\limsup_{x\to x_0}\frac{f'(x)}{g'(x)}.$$

We omit the proofs of Lemmas 2.2 and 2.3 as they are elementary. The proof of Lemma 2.4 can be found in Gajek (1987).

Proof of Theorem 2.1. (i) The proof is somewhat similar to the proofs of Theorem 2 in Gajek (1987) and Theorem 2.7 in Gajek (1988). First, we prove the bound as $\theta\to 0$. Observe that the lower bound given in (i) is equal to 0 if $\lim_{\theta\to 0}\theta^{2s}h(\theta)=0$, so without loss of generality we can assume that this limit is positive. Applying (2) and Lemma 2.2 we obtain for all $\theta\in\Theta$,

$$
\begin{aligned}
(3)\quad & \mathrm{E}_\theta[T_N-\theta^s]^2h(\theta)+c(\theta)\mathrm{E}_\theta N \\
&\geq \left[\left(\frac{b(\theta)}{\theta^s}\right)^2+\frac{s^2\left(1+\frac{b'(\theta)}{s\theta^{s-1}}\right)^2}{\theta^2 I(\theta)\mathrm{E}_\theta N}\right]\theta^{2s}h(\theta)+c(\theta)\mathrm{E}_\theta N \\
&\geq \left[\left(\frac{b'(\theta)}{s\theta^{s-1}}\right)^2+\frac{2|s|}{\theta^{s+1}}\sqrt{\frac{c(\theta)}{I(\theta)h(\theta)}}\left|1+\frac{b'(\theta)}{s\theta^{s-1}}\right|\right]\theta^{2s}h(\theta) \\
&\quad+\left[\left(\frac{b(\theta)}{\theta^s}\right)^2-\left(\frac{b'(\theta)}{s\theta^{s-1}}\right)^2\right]\theta^{2s}h(\theta).
\end{aligned}
$$

Since for some $\delta>0$ we have $0<c(\theta)\leq s^{-2}\theta^{2s+2}I(\theta)h(\theta)$ for all $\theta<\delta$, therefore

$$
0<\frac{2|s|}{\theta^{s+1}}\sqrt{\frac{c(\theta)}{I(\theta)h(\theta)}}\leq 2\qquad\text{for all } \theta<\delta,
$$

and from Lemma 2.3 and (3) we have

$$
\begin{aligned}
(4)\quad & \mathrm{E}_\theta[T_N-\theta^s]^2h(\theta)+c(\theta)\mathrm{E}_\theta N \\
&\geq |s|\theta^{s-1}\sqrt{\frac{c(\theta)h(\theta)}{I(\theta)}}\left(2-\frac{|s|}{\theta^{s+1}}\sqrt{\frac{c(\theta)}{I(\theta)h(\theta)}}\right) \\
&\quad+\left[\left(\frac{b(\theta)}{\theta^s}\right)^2-\left(\frac{b'(\theta)}{s\theta^{s-1}}\right)^2\right]\theta^{2s}h(\theta)\qquad\text{for all } \theta<\delta.
\end{aligned}
$$

Observe that (i) holds if the left hand side of (3) is unbounded on each interval $(0,\delta_1)$. So assume the opposite. Then the condition $\lim_{\theta\to 0}\theta^{2s}h(\theta)<\infty$ and the first inequality in (3) imply together that $\lim_{\theta\to 0} b(\theta)=0$. In order to prove (i) it is enough to show that

$$
(5)\qquad \limsup_{\theta\to 0}\left[\left(\frac{b(\theta)}{\theta^s}\right)^2-\left(\frac{b'(\theta)}{s\theta^{s-1}}\right)^2\right]\geq 0
$$

and next to combine it with (4). Suppose that (5) is not satisfied. Then, for some $\varepsilon>0$, $(b(\theta)/\theta^s)^2-(b'(\theta)/(s\theta^{s-1}))^2<0$ for $\theta<\varepsilon$. Hence $b'(\theta)\neq 0$ for $\theta<\varepsilon$ and by Theorem 5.12 of Rudin (1976), either $b'(\theta)>0$ for all $\theta<\varepsilon$ or the reverse inequality holds on $(0,\varepsilon)$. In the first case we have for $s>0$ and $\theta\in(0,\varepsilon)$,

$$
(6)\qquad -\frac{b'(\theta)}{s\theta^{s-1}}<\frac{b(\theta)}{\theta^s}<\frac{b'(\theta)}{s\theta^{s-1}}.
$$

Since

$$
\frac{d}{d\theta}\left[\frac{b(\theta)}{\theta^s}\right]=s\theta^{-1}\left[\frac{b'(\theta)}{s\theta^{s-1}}-\frac{b(\theta)}{\theta^s}\right],
$$

from (6) it follows that $b(\theta)/\theta^s$ is increasing on $(0,\varepsilon)$ and so the limit of $b(\theta)/\theta^s$ exists as $\theta\to 0$. Hence, from Lemma 2.4 and (6), we obtain

$$\liminf_{\theta\to 0}\frac{b'(\theta)}{s\theta^{s-1}}=\liminf_{\theta\to 0}\frac{b(\theta)}{\theta^s}=\limsup_{\theta\to 0}\frac{b(\theta)}{\theta^s}\le\limsup_{\theta\to 0}\frac{b'(\theta)}{s\theta^{s-1}}.$$

When $s<0$ the opposite inequalities in (6) hold but $b(\theta)/\theta^s$ is still increasing on $(0,\varepsilon)$. Hence the limit of $b(\theta)/\theta^s$ exists and $b(\theta)\to\infty$ as $\theta\to 0$, and by Lemma 2.4,

$$\liminf_{\theta\to 0}\frac{b'(\theta)}{s\theta^{s-1}}\le\liminf_{\theta\to 0}\frac{b(\theta)}{\theta^s}=\limsup_{\theta\to 0}\frac{b(\theta)}{\theta^s}=\limsup_{\theta\to 0}\frac{b'(\theta)}{s\theta^{s-1}}.$$

In each case if $\lim_{\theta\to 0}b(\theta)/\theta^s$ is finite, then (5) is satisfied, a contradiction; if not, then (i) follows directly from (3). The case $b'(\theta)<0$ for $\theta\in(0,\varepsilon)$ can be treated in the same way. The proof of the theorem for $\theta\to\infty$ is a bit more complex though quite similar. ∎

Theorem 2.1 shows that the minimax risk of each estimator which satisfies inequality (2) depends neither on the estimator nor on the stopping time. Now consider the scale family of Lebesgue densities

$$\mathcal{F}=\{f_\theta: f_\theta(x)=(1/\theta)f_1(x/\theta),\ x\ge 0,\ \theta\in\Theta\}, \tag{7}$$

where f_1 is a given Lebesgue density. Assume that f_1 is differentiable. Let $I(\theta)$ denote the Fisher information of a single observation X, which has Lebesgue density $f_\theta\in\mathcal{F}$. It is easy to show that

$$I(\theta)=\theta^{-2}A_1, \tag{8}$$

where

$$A_1=\int_0^\infty\left[1+\frac{uf_1'(u)}{f_1(u)}\right]^2 f_1(u)\,du. \tag{9}$$

Let $X_1,X_2,\ldots$ be i.i.d. r.v.'s each with Lebesgue density f_θ, where $f_\theta\in\mathcal{F}$. Let V denote the minimax value in estimating θ^s, $s\ne 0$, under the risk given by (1) and the weight function $h(\theta)=\theta^{-2s}$, i.e.

$$V=\inf_{T_N}\sup_{\theta\in\Theta}\{\mathrm{E}_\theta[T_N-\theta^s]^2\theta^{-2s}+c(\theta)\mathrm{E}_\theta N\}.$$

Then the following result holds.

Proposition 2.5. *Assume that* $\Theta=(0,\theta_1)$ *for some* $0<\theta_1\le\infty$ (*resp.* $\Theta=(\theta_2,\infty)$ *for some* $0\le\theta_2<\infty$). *Let* A_1 *be defined by* (9).

(i) *If* $0<c(\theta)\le s^{-2}A_1$ *for all* $\theta\in\Theta$ *such that* $\theta<\delta$, *for some* $\delta>0$ (*resp.* $\theta>\kappa$, *for some* $\kappa>0$), *then*

$$V\ge\lim_{\substack{\theta\to 0\\(\theta\to\infty)}}|s|\sqrt{\frac{c(\theta)}{A_1}}\left(2-|s|\sqrt{\frac{c(\theta)}{A_1}}\right)$$

provided that the right-hand side exists.

(ii) *If* $c(\theta) \geq s^{-2}A_1$ *for all* $\theta \in \Theta$ *such that* $\theta < \delta$, *for some* $\delta > 0$ (*resp.* $\theta > \kappa$, *for some* $\kappa > 0$), *then* $V \geq 1$.

Proof. The result follows from Theorem 2.1 for $h(\theta) = \theta^{-2s}$ and $I(\theta)$ defined by (8). ■

Proposition 2.5(ii) implies that if the cost of collecting observations is too large then trivial estimators may be both minimax and admissible. To be more precise, define T_0 to be a given constant $C^s \in \mathbb{R}$ and assume that $c(\theta) \geq s^{-2}A_1$. Then the lower bound on V is equal to 1 provided that $\mathrm{E}_\theta N > 0$. If $\mathrm{E}_\theta N = 0$ for some $\theta \in \Theta$, then $T_N = T_0$ a.s. $[P_\theta]$. Moreover, if $s > 0$ and $\Theta = [\theta_2, \infty)$, where $0 < \theta_2 < \infty$, then

$$\inf_{C \in \mathbb{R}} \sup_{\theta \in \Theta} R(T_0, \theta) = 1.$$

If $s < 0$ and $\Theta = (0, \theta_1]$ for some $\theta_1 > 0$, then the same holds. So $V \geq 1$ no matter what $\mathrm{E}_\theta N$ is, provided Θ is a properly truncated parameter space. On the other hand, the estimator $T_N \equiv C^s$ makes the supremum of the risk equal 1, for every $C \in [2^{1/s}\theta_1, \theta_1]$ when $\Theta = (0, \theta_1]$ and $s < 0$. Hence it is a minimax (and admissible) estimator of θ^s in this case. If $s > 0$ and $\Theta = [\theta_2, \infty)$ for some $\theta_2 > 0$ the estimator $T_N \equiv C^s$ with $C \in [\theta_2, 2^{1/s}\theta_2]$ is minimax and admissible (to prove admissibility, it is sufficient to notice that $T_N = C^s$ is the unique locally optimal estimator at the point $\theta = C$).

3. Application to censored data from an exponential distribution. Assume now that each random variable $X_1, X_2, \ldots$ has an exponential distribution with Lebesgue density

$$f_\theta(x) = (1/\theta)\exp(-x/\theta), \qquad x \geq 0,\ \theta \in \Theta,$$

where $\Theta = (0, \theta_1)$ for some $0 < \theta_1 \leq \infty$ or $\Theta = (\theta_2, \infty)$ for some $0 \leq \theta_2 < \infty$. Assume that the data consist of the sequence $(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots$ defined by

$$Z_i = \min(X_i, y_0), \qquad \Delta_i = \begin{cases} 1 & \text{for } X_i \leq y_0, \\ 0 & \text{for } X_i > y_0, \end{cases} \qquad i = 1, 2, \ldots,$$

where $y_0 > 0$ is a given censoring time. The average cost of each observation is denoted by $c(\theta)$. Let a random variable N_r, denoting the number of observations, be described in the following way:

$$N_r = \min\Big\{m : \sum_{i=1}^{m} \Delta_i \geq r\Big\} \qquad \text{for } r = 1, 2, \ldots,$$

where r is a given integer. Define

$$p := P(\Delta = 1) = P(X \leq y_0) = 1 - \exp(-y_0/\theta).$$

THEOREM 3.1. *Assume that* $\Theta = (0, \theta_1)$ *for some* $0 < \theta_1 \le \infty$ *and* $\sup_{\theta\in\Theta}[c(\theta)/p] = \lim_{\theta\to 0}[c(\theta)/p] = 1/(k+1)^2$, *where* k *is a positive integer. The estimator* $\tilde{\theta}_{N_k}$ *defined by*

$$\tilde{\theta}_{N_k} = \frac{1}{k+1}\Big[\sum_{i=1}^{k} X_{i,N_k} + (N_k - k)y_0\Big],$$

where X_{i,N_l} *is the ith order statistic from* $X_1, \ldots, X_{N_l}$, *is a minimax estimator of* $\theta \in \Theta$ *under the loss weighted by* θ^{-2}.

Proof. Observe that

$$P_\theta(N_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = k, k+1, \ldots$$

It can be shown that, under the above conditions, the order statistics $X_{1,N_k}, \ldots, X_{k,N_k}$ have a joint conditional density given $N_k = n$, which is equal to the joint density of the order statistics from a random sample consisting of k i.i.d. observations from a truncated distribution with density

$$h(u) = (\theta p)^{-1}\exp(-u/\theta), \quad 0 < u \le y_0. \tag{10}$$

Note that

$$\begin{aligned} \mathrm{E}\widetilde{\theta}_{N_k} &= \mathrm{E}[\mathrm{E}(\widetilde{\theta}_{N_k} \mid N_k)] = \mathrm{E}\left[\frac{1}{k+1}\mathrm{E}\Big(\sum_{i=1}^{k} X_{i,N_k} + (N_k-k)y_0 \mid N_k\Big)\right] \\ &= \mathrm{E}\left[\frac{k}{k+1}\mathrm{E}U + \frac{1}{k+1}(N_k-k)y_0\right], \end{aligned}$$

where U has density (10). Since $\mathrm{E}U = \theta - y_0 q/p$, $\mathrm{E}N_k = k/p$, where $p = 1-\exp(-y_0/\theta)$, $q = 1-p$, therefore

$$\mathrm{E}\widetilde{\theta}_{N_k} = \frac{k}{k+1}\theta. \tag{11}$$

Further, observe that

$$\mathrm{E}(\widetilde{\theta}_{N_k} - \theta)^2 = \mathrm{Var}\,\widetilde{\theta}_{N_k} + (\mathrm{E}\widetilde{\theta}_{N_k} - \theta)^2, \tag{12}$$

where $\mathrm{Var}\,\widetilde{\theta}_{N_k} = \mathrm{E}[\mathrm{Var}(\widetilde{\theta}_{N_k} \mid N_k)] + \mathrm{Var}[\mathrm{E}(\widetilde{\theta}_{N_k} \mid N_k)]$. As $\mathrm{E}[\mathrm{Var}(\widetilde{\theta}_{N_k} \mid N_k)] = (k/(k+1)^2)\,\mathrm{Var}\,U = (k/(k+1)^2)(\theta^2 - y_0^2 q/p^2)$ and $\mathrm{Var}[\mathrm{E}(\widetilde{\theta}_{N_k} \mid N_k)] = y_0^2 k(1-p)/[p^2(k+1)^2]$, therefore

$$\mathrm{Var}\,\widetilde{\theta}_{N_k} = \frac{k}{(k+1)^2}\theta^2. \tag{13}$$

From (11)–(13) we obtain $\mathrm{E}(\widetilde{\theta}_{N_k} - \theta)^2 = \theta^2/(k+1)$. Hence and from the fact that $\sup_{\theta\in\Theta}[c(\theta)/p] = 1/(k+1)^2$ we have

$$\sup_{\theta\in\Theta}[\mathrm{E}(\widetilde{\theta}_{N_k} - \theta)^2\theta^{-2} + c(\theta)\mathrm{E}N_k] = \frac{2k+1}{(k+1)^2}. \tag{14}$$

Since the Fisher information $I_1(\theta)$ of a single observation (Z, Δ) is equal to $I_1(\theta) = \theta^{-2}p$ and $\lim_{\theta\to 0}[c(\theta)/p] = 1/(k+1)^2$, from Theorem 2.1(i) for $\theta \to 0$ we have the following inequality for a sequential estimator θ^*_M of θ with stopping time M:

$$\sup_{\theta\in\Theta}[\mathrm{E}_\theta(\theta^*_M - \theta)^2\theta^{-2} + c(\theta)\mathrm{E}M] \geq \frac{2k+1}{(k+1)^2}. \tag{15}$$

Now (14) and (15) imply the assertion. ■

THEOREM 3.2. *Assume that* $\Theta = (\theta_2, \infty)$ *for some* $0 \leq \theta_2 < \infty$ *and* $\sup_{\theta\in\Theta}[c(\theta)/p] = \lim_{\theta\to\infty}[c(\theta)/p] = 1/(k+1)^2$, *where k is a positive integer. The estimator* $\widetilde{\theta}_{N_k}$ *defined in Theorem* 3.1 *is a minimax estimator of* $\theta \in \Theta$.

Proof. As in Theorem 3.1, from Theorem 2.1(i) for $\theta \to \infty$ we obtain for a sequential estimator $\check{\theta}_M$ of θ,

$$\sup_{\theta\in\Theta}[\mathrm{E}_\theta(\check{\theta}_M - \theta)^2\theta^{-2} + c(\theta)\mathrm{E}_\theta M] \geq \frac{2k+1}{(k+1)^2}.$$

Now the result follows from (14). ■

REMARK 3.3. The assumption that $c(\theta)/p = \text{const.}$ seems to be quite natural in life time experiments. Indeed, suppose that the items on test are observed only for a certain period of time y_0 and classified afterwards. If the life time X is greater than y_0, the item is classified for sale, otherwise the loss is c. Then $c(\theta) = cp$, where $p = P(X \leq y_0)$.

4. Minimax estimation in the proportional hazards model. Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s with absolutely continuous distribution function F. Assume that $X_1, X_2, \ldots$ are censored on the right by i.i.d. r.v.'s $Y_1, Y_2, \ldots$ which have a common distribution function G, so that the observations available are the pairs (Z_i, Δ_i), where

$$Z_i = \min(X_i, Y_i), \qquad \Delta_i = \begin{cases} 1, & X_i \leq Y_i, \\ 0, & X_i > Y_i, \end{cases} \qquad i = 1, 2, \ldots$$

Assume that the sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are independent. Write $P(Z > t) = 1 - H(t)$, $t \in \mathbb{R}$.

The *proportional hazards model* is a parametric-nonparametric model in which there exists a positive constant d, the so-called *censoring parameter*, such that

$$1 - G(x) = (1 - F(x))^d, \qquad x \in \mathbb{R}.$$

In this model, the expected proportion $p = P(\delta = 1)$ of uncensored observations satisfies the equation $p = (1+d)^{-1}$ and

$$1 - F(x) = (1 - H(x))^p, \qquad x \in \mathbb{R}.$$

Note that the special case $d = 0$ (or $p = 1$) may be identified with the lack of censoring.

This model was considered by Koziol and Green (1976), Csörgő (1988), Csörgő and Mielniczuk (1988) and others (for a more complete list of references and a survey of results, see Csörgő (1988)). In this section we consider minimax estimation of p provided that the form of F is known. Let $r(x)$ be the *hazard function* of X, i.e.

$$r(x) = f(x)/[1 - F(x)], \qquad x \in S,$$

where f is the density of X and $S = \{x \in \mathbb{R} : 0 < f(x)\}$. Let $T(x) = -\log(1 - F(x))$, $x \in S$, be the *cumulative hazard function* of X. Obviously $F(x) = 1 - \exp(-T(x))$ and $f(x) = r(x)\exp(-T(x))$ a.e. on S.

The aim of this section is to find a minimax sequential estimator of p under the normalized square error loss function

$$L(\widetilde{p}, p) = (\widetilde{p} - p)^2/p^2,$$

with constant cost function, $c(p) = c_0$, per observation. A natural estimator of p, and to the best of our knowledge the only proposal in the literature, is based on the statistic

$$\widehat{p}_n = \frac{1}{n}\sum_{i=1}^{n} \Delta_i. \tag{16}$$

Clearly, $\widehat{p}_n$ is unbiased and $\operatorname{Var}\widehat{p}_n = p(1-p)/n$.

However, using the statistics (16) does not lead to constructing a minimax sequential estimator. In fact, such an estimator is based on the statistic

$$p_n^* = \frac{1}{n+1}\sum_{i=1}^{n} T(Z_i)$$

and, what is quite surprising, does not use any information from the sequence $(\Delta_1, \Delta_2, \ldots)$!

THEOREM 4.1. *Assume that $p \in (0, \overline{p})$ with $0 < \overline{p} \le 1$. Let $0 < c_0 < 1$ and define the stopping time $N_0 = 1/\sqrt{c_0} - 1$ with probability 1. Then the estimator*

$$p_{N_0}^* = \frac{1}{N_0+1}\sum_{i=1}^{N_0} T(Z_i)$$

is a minimax sequential estimator in the proportional hazards model considered above.

Proof. It is well known that Δ and Z are independent in the proportional hazards model and hence it is easy to show that the Fisher information $I_{(Z_1,\Delta_1)}(p)$ of a single observation (Z_1, Δ_1) is equal to $1/[p^2(1-p)]$.

Further, applying Theorem 2.1(i) with $h(p) = p^{-2}$, $c(p) = c_0$, $0 < c_0 \le 1$ and $p \to 0$, we get the following lower bound for the minimax value in the problem of estimating p by a sequential estimator $\widetilde{p}_M$:

$$\sup_{p\in(0,1)} \{\mathrm{E}(\widetilde{p}_M - p)^2 p^{-2} + c_0 \mathrm{E} M\} \ge 2\sqrt{c_0} - c_0. \tag{17}$$

Since Z_1 has distribution function $H(t) = 1 - [1 - F(t)]^{1/p}$, therefore

$$\begin{aligned} P(T(Z_1) \le t) &= P(-\log[1 - F(Z_1)] \le t) = P(1 - F(Z_1) \ge e^{-t}) \\ &= P([1 - F(Z_1)]^{1/p} \ge e^{-t/p}) = P(1 - H(Z_1) \ge e^{-t/p}) \\ &= P(H(Z_1) \le 1 - e^{-t/p}) = 1 - e^{-t/p}, \quad t > 0, \end{aligned}$$

because $H(Z_1)$ is uniformly distributed. Hence $T(Z_1)$ is exponentially distributed with scale parameter p, so $\sum_{i=1}^n T(Z_i)$ has a gamma distribution. Therefore

$$\mathrm{E}\left[\frac{1}{n+1}\sum_{i=1}^n T(Z_i)\right] = \frac{n}{n+1}p \quad \text{and} \quad \mathrm{Var}\left[\frac{1}{n+1}\sum_{i=1}^n T(Z_i)\right] = \frac{n}{(n+1)^2}p^2.$$

Since $N_0 = n_0$ with probability 1 and $p_n^* = (n+1)^{-1}\sum_{i=1}^n T(Z_i)$, therefore

$$\mathrm{E} p_{N_0}^* = \frac{n_0}{n_0+1}p, \quad \mathrm{Var}\, p_{N_0}^* = \frac{n_0}{(n_0+1)^2}p^2 \quad \text{and} \quad \mathrm{E}\,(p_{N_0}^* - p)^2 p^{-2} = \frac{1}{n_0+1}.$$

Hence

$$R(p_{N_0}^*, p) = \mathrm{E}\,(p_{N_0}^* - p)^2 p^{-2} + c_0 \mathrm{E} N_0 = \frac{1}{n_0+1} + c_0 n_0. \tag{18}$$

From (17), (18) and the fact that $n_0 = 1/\sqrt{c_0} - 1$, $0 < c_0 \le 1$, we obtain the assertion. ∎

5. Some other applications. In this section we show that Theorem 2.1 is applicable to a variety of problems. First of all we get immediately a result analogous to that of Magiera (1977), who considered minimax sequential estimation of continuous time exponential type stochastic processes. What is more, our method proves minimaxity of the estimator (20) below also in the case $\beta = 0$, which was not covered by Magiera (1977). It is also worth noting that our result concerns discrete time processes and that the cost function c may depend on θ.

The second example of applications of Theorem 2.1 concerns estimation of the odds ratio and the inverse of the success probability in a sequence of Bernoulli trials. It should be stressed that the result does not follow from the first part of this section.

Minimax estimation for an exponential type family of distributions. Consider a one-parameter exponential family of probability distributions P_θ,

$\theta \in \Theta \subset (0, \infty)$, which are absolutely continuous with respect to a σ-finite measure μ on $\mathbb{R}$, with Radon–Nikodym derivatives

$$\frac{dP_\theta(x)}{d\mu(x)} \equiv p_\theta(x) = w(x)\exp[\eta(\theta)t(x) + a(\theta)], \qquad x \in \mathbb{R}, \tag{19}$$

where $t(x)$, $w(x)$ denote measurable functions and $a(\theta)$, $\eta(\theta)$ are some real functions defined on Θ.

Let X be a random variable with Radon–Nikodym density $p_\theta(x)$ defined by (19) and let $I(\theta)$ denote the Fisher information of X. Reparametrizing family (19) if necessary, we may also assume that $\theta = \mathrm{E}t(X)$. Then $I(\theta) = 1/\operatorname{Var} t(X)$ (see e.g. Lehmann (1983), Theorem 6.2). Assume that $X, X_1, X_2, \ldots$ are i.i.d. random variables. We are interested in minimax estimation of the parameter $\theta = \mathrm{E}t(X)$ under the risk defined by (1) with weight function $h(\theta) = I(\theta)$. For the time being assume that $c(\theta) \equiv c$, where $c \in \mathbb{R}_+$.

Consider the following sequential estimator of θ:

$$\check{\theta}_{N,\beta} = \sum_{i=1}^{N} t(X_i)/(N + \beta) \tag{20}$$

where the r.v. N is a stopping time and $\beta \in \mathbb{R}$. If N is equal, with probability 1, to a constant n_0, the pair $(N, \check{\theta}_{N,\beta})$ is called a *fixed-time plan* (see e.g. Magiera (1977)). Then we have

$$\begin{aligned}\mathrm{E}\,(\check{\theta}_{N,\beta} - \theta)^2 &= \operatorname{Var}\check{\theta}_{N,\beta} + (\mathrm{E}\check{\theta}_{N,\beta} - \theta)^2\\ &= \frac{1}{(n_0+\beta)^2} n_0 \operatorname{Var} t(X) + \left(\frac{n_0}{n_0+\beta}\theta - \theta\right)^2\\ &= \frac{n_0 + \beta^2\theta^2 I(\theta)}{(n_0+\beta)^2 I(\theta)}.\end{aligned}$$

Hence

$$R(\check{\theta}_{N,\beta}, N) = \mathrm{E}\,(\check{\theta}_{N,\beta} - \theta)^2 I(\theta) + c\mathrm{E}N = \frac{n_0 + \beta^2\theta^2 I(\theta)}{(n_0+\beta)^2} + cn_0. \tag{21}$$

Applying a simple extension of Theorem 2.1(i) and (21) it is easy to check that under the above assumptions the following result holds:

PROPOSITION 5.1. *Assume that* $\Theta = (0, \theta_1)$ *for some* $0 < \theta_1 \le \infty$ (*resp.* $\Theta = (\theta_2, \infty)$ *for some* $0 \le \theta_2 < \infty$).

(i) *Assume that* $\lim_{\theta\to 0} \theta^2 I(\theta) = \infty$ (*resp.* $\lim_{\theta\to\infty} \theta^2 I(\theta) = \infty$) *and define the stopping time* N_1 *to be equal to* $1/\sqrt{c}$ *with probability* 1. *Then* $\check{\theta}_{N_1,0}$, *defined by* (20), *is a minimax estimator of* $\theta \in \Theta$.

(ii) *Assume that* $\sup_{\theta\in\Theta} \theta^2 I(\theta) = \lim_{\theta\to 0} \theta^2 I(\theta) = 1/\beta$ (*resp.* $\sup_{\theta\in\Theta} \theta^2 I(\theta) = \lim_{\theta\to\infty} \theta^2 I(\theta) = 1/\beta$) *for some* $\beta > 0$. *Let* c *be such* *that* $0 < c < 1/\beta^2$ *and define the stopping time* N_2 *to be equal to* $1/\sqrt{c} - \beta$ *with probability* 1. *Then* $\check{\theta}_{N_2,\beta}$ *is a minimax estimator of* $\theta \in \Theta$.

Similar results to Proposition 5.1 can be easily deduced from Theorem 2.1 in the general case when the cost function c depends on the unknown parameter θ. For example, we have

REMARK 5.2. Assume that $\Theta = (0, \theta_1)$ for some $0 < \theta_1 \le \infty$ (resp. $\Theta = (\theta_2, \infty)$ for some $0 \le \theta_2 < \infty$) and $\lim_{\theta\to 0} \theta^2 I(\theta) = \infty$ (resp. $\lim_{\theta\to\infty} \theta^2 I(\theta) = \infty$). Let $c(\theta) > 0$ be such that $\sup_{\theta\in\Theta} c(\theta) = \lim_{\theta\to 0} c(\theta) = c$ (resp. $\sup_{\theta\in\Theta} c(\theta) = \lim_{\theta\to\infty} c(\theta) = c$) for some $c > 0$ and consider the stopping time $N_3 = 1/\sqrt{c}$ with probability 1. Then $\check{\theta}_{N_3,0}$ is a minimax estimator of $\theta \in \Theta$.

Now we present some examples of probability distributions that satisfy the conditions of Proposition 5.1.

EXAMPLES. (a) Consider the r.v. X with gamma distribution of density

$$f_\theta(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

with respect to the Lebesgue measure, where $x > 0$, α is a given positive real and $\theta \in \Theta$. The parameter θ is estimated under the risk (1) with weight function $h(\theta) = I(\theta)$. Suppose that $\Theta = (0, \theta_1)$ for some $0 < \theta_1 \le \infty$, or $\Theta = (\theta_2, \infty)$ for $0 \le \theta_2 < \infty$. According to the notation introduced earlier $t(X) = X/\alpha$, $\mathrm{E}t(X) = \theta$, $I(\theta) = \alpha/\theta^2$. Then

$$\sup_{\theta\in\Theta} \theta^2 I(\theta) = \alpha \quad \text{and} \quad \lim_{\theta\to 0\,(\theta\to\infty)} \theta^2 I(\theta) = \alpha$$

and the conditions of Proposition 5.1(ii) are satisfied for $\beta = 1/\alpha$.

(b) Suppose that the r.v. X has probability function

$$P_\theta(X = x) = \frac{e^{-\theta}\theta^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots,\ \theta \in \Theta$$

where $\Theta = (\theta_2, \infty)$ for some $0 \le \theta_2 < \infty$. Then $t(X) = X$, $\mathrm{E}X = \theta$, $I(\theta) = 1/\theta$ and $\lim_{\theta\to\infty} \theta^2 I(\theta) = \infty$. By Proposition 5.1(i) the estimator $\check{\theta}_{N_1,0}$ is a minimax estimator of $\theta \in \Theta$.

(c) Suppose that the r.v. X is normally distributed with Lebesgue density

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp[-x^2/(2\theta)]$$

for $x \in \mathbb{R}$, $\theta \in \Theta$, where $\Theta = (0, \theta_1)$ for some $0 < \theta_1 \le \infty$ or $\Theta = (\theta_2, \infty)$ for some $0 \le \theta_2 \le \infty$. Let the parameter θ be estimated under the risk (1) with the weight function $h(\theta) = I(\theta)$. We obtain $t(X) = X^2$, $\mathrm{E}t(X) = \theta$, $I(\theta) = 1/(2\theta^2)$ and Proposition 5.1(ii) holds for $\beta = 2$.

Minimax estimation of the inverse of the success probability in a sequence of Bernoulli trials. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s with $P(X_i = 1) = p_1$, $P(X_i = 0) = q_1 = 1 - p_1$, where $p_1 \in (0, \overline{p}_1)$ for some $0 < \overline{p}_1 < 1$, $i = 1, 2, \ldots$ The minimax estimation of the function $\theta = p_1^{-1}$, $\theta \in (1/\overline{p}_1, \infty)$, is considered under the normalized squared error loss of estimation with weight $h(\theta) = \theta^{-2}$. Assume also that the risk incorporates some average cost $K(\theta)$ of collecting observations. Suppose that the cost $\xi(X_i)$ of the observation X_i is defined by

$$\xi(X_i) = \begin{cases} 0 & \text{for } X_i = 0, \\ c_1 & \text{for } X_i = 1, \ i = 1, 2, \ldots, \end{cases}$$

with fixed constant $c_1 \in (0, 1)$. Observe that $K(\theta) = \mathrm{E}[\sum_{i=1}^M \xi(X_i)]$, where M denotes a stopping rule. By the Wald lemma, $K(\theta) = c_1 \mathrm{E} M/\theta$ and the total risk of an estimator $\widehat{\theta}_M$ of θ is

$$R(\widehat{\theta}_M, \theta) = \frac{\mathrm{E}(\widehat{\theta}_M - \theta)^2}{\theta^2} + \frac{c_1 \mathrm{E} M}{\theta}.$$

Consider the stopping time

$$N_r = \min\Big\{m : \sum_{i=1}^m X_i = r\Big\} \quad \text{for } r = 1, 2, \ldots,$$

with fixed r (see also Section 3) and the estimator

$$\theta^*_{N_r} = \frac{N_r}{r+1}. \tag{22}$$

It is easy to check that

$$R(\theta^*_{N_r}, \theta) = \frac{r(1 - \theta^{-1}) + 1}{(r+1)^2} + c_1 r$$

and

$$\sup_{\theta > 1/\overline{p}_1} R(\theta^*_{N_r}, \theta) = \frac{1}{r+1} + c_1 r. \tag{23}$$

Since the Fisher information $I^*(\theta)$ is $1/[\theta^2(\theta - 1)]$, from Theorem 2.1(i) for $\theta \to \infty$ we have

$$\sup_{\theta > 1/\overline{p}_1} R(\widehat{\theta}_M, \theta) \ge \sqrt{c_1}(2 - \sqrt{c_1}).$$

Now (23) shows that for $r = (1 - \sqrt{c_1})/\sqrt{c_1}$, $\theta^*_{N_r}$ given by (22) is a minimax estimator of $\theta = p_1^{-1}$, whenever $\theta \in (1/\overline{p}_1, \infty)$.

Minimax estimation of the odds ratio. Now consider the problem of estimating the odds ratio $\theta = p_1/q_1$ from a sequence $X_1, X_2, \ldots$ of i.i.d. r.v.'s with $P(X_i = 1) = p_1$ and $P(X_i = 0) = q_1 = 1 - p_1$, where $q_1 \in (0, \overline{q})$ for some $0 < \overline{q} < 1$. Since $\theta = q_1^{-1} - 1$ and the r.v.'s $X_i' = 1 - X_i$ are two-point distributed with success probability $q_1 \equiv P(X'_i = 1)$, estimating θ is formally equivalent to estimating the inverse of the success probability q_1^{-1} (shifted by a constant) from the sequence $X'_1, X'_2, \ldots$ From the previous section it follows that the statistic

$$\theta^*_{N'_r} = \frac{N'_r}{r+1} - 1,$$

where the stopping time is

$$N'_r = \min\Big\{m : \sum_{i=1}^{m}(1 - X_i) = r\Big\},$$

is a minimax estimator of the odds ratio $\theta = p_1/q_1$ under the square error loss with weight function $h(\theta) = q_1^2 = (1+\theta)^{-2}$ provided that the cost function is $c(\theta) = c_1 q_1 = c_1(1+\theta)^{-1}$ and $r = (1-\sqrt{c_1})/\sqrt{c_1}$. The average cost of observation, $c(\theta)$, is equal to $c_1 q_1$ whenever the cost $\xi(X_i)$ of the observation X_i satisfies

$$\xi(X_i) = \begin{cases} c_1 & \text{for } X_i = 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \ldots$

References

M. Alvo (1977), *Bayesian sequential estimation*, Ann. Statist. 5, 955–968.

Y. S. Chow, H. Robbins and D. Siegmund (1971), *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston.

S. Csörgő (1988), *Estimation in the proportional hazards model of random censorship*, Statistics 19, 437–463.

S. Csörgő and J. Mielniczuk (1988), *Density estimation in the simple proportional hazards model*, Statist. Probab. Letters 6, 419–426.

L. Gajek (1987), *An improper Cramér–Rao lower bound*, Zastos. Mat. 19, 241–256.

L. Gajek (1988), *On minimax value in the scale model with truncated data*, Ann. Statist. 16, 669–677.

L. Gajek and U. Gather (1991), *Estimating a scale parameter under censorship*, Statistics 22, 529–549.

J. C. Gardiner and V. Susarla (1984), *Risk-efficient estimation of the mean exponential survival time under random censoring*, Proc. Nat. Acad. Sci. U.S.A. 81, 5906–5909.

J. C. Gardiner and V. Susarla (1991), *Some asymptotic distribution results in time-sequential estimation of the mean exponential survival time*, Canad. J. Statist. 19, 425–436.

J. C. Gardiner, V. Susarla and J. van Ryzin (1986), *Time sequential estimation of the exponential mean under random withdrawals*, Ann. Statist. 14, 607–618.

J. A. Koziol and S. B. Green (1976), *A Cramér–von Mises statistic for randomly censored data*, Biometrika 63, 465–474.

E. L. Lehmann (1983), *Theory of Point Estimation*, Wiley, New York.

R. Magiera (1977), *On sequential minimax estimation for the exponential class of processes*, Zastos. Mat. 15, 445–454.

B. Mizera (1996), *Lower bounds on the minimax risk of sequential estimators*, Statistics 28, 123–129.

W. Rudin (1976), *Principles of Mathematical Analysis*, McGraw-Hill, New York.

M. Tahir (1988), *Asymptotically optimal Bayesian sequential point estimation with censored data*, Sequential Anal. 7, 227–237.

J. Wolfowitz (1947), *The efficiency of sequential estimates and Wald's equation for sequential processes*, Ann. Math. Statist. 19, 215–230.

M. Woodroofe (1982), *Nonlinear Renewal Theory in Sequential Analysis*, CBMS-NSF Regional Conf. Ser. Appl. Math. 39, SIAM, Philadelphia.

Lesław Gajek and B. Mizera-Florczak
Institute of Mathematics
Technical University of Łódź
ul. Żwirki 36
90-924 Łódź, Poland

Received on April 1, 1997