Stat Papers (2010) 51:511–529
DOI 10.1007/s00362-008-0137-0

REGULAR ARTICLE

Optimal sequential estimation procedures of a function of a probability of success under LINEX loss

Jerzy Baran · Ryszard Magiera

Received: 14 June 2007 / Accepted: 7 March 2008 / Published online: 15 April 2008

Abstract In this paper, we investigate the problem of estimating a function $g(p)$, where p is the probability of success in a sequential sample of independent identically Bernoulli distributed random variables. As a loss associated with estimation we introduce a generalized LINEX loss function. We construct a sequential procedure possessing some asymptotically optimal properties in the case when p tends to zero. In this approach to the problem, the conditions are given, under which the stopping time is asymptotically efficient and normal, and the corresponding sequential estimator is asymptotically normal. The procedure constructed guarantees that its sequential risk is asymptotically equal to a prescribed constant.

Keywords Sequential procedure · Stopping rule · Asymptotic optimality of sequential procedures · Bernoulli trials · Probability of success · LINEX loss function

1 Introduction

In many situations, statistical inference problems cannot be solved on the basis of the fixed sample size, and the only way to resolve them is to construct sequential procedures. One of the first examples of such situation was considered by Stein (1945), who constructed a two-stage procedure for estimating the mean of a normal distribution with unknown variance.

In this paper, we focus our attention upon the Bernoulli case. The problem of sequential estimation of a probability of success in Bernoulli trials is engaging a good deal of attention in the literature—for example Haldane (1945) (one of the first papers),

J. Baran · R. Magiera (✉)
Institute of Mathematics and Computer Science, Wroclaw University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: Ryszard.Magiera@pwr.wroc.pl

Robbins and Siegmund (1976), Cabilio and Robbins (1975), Cabilio (1977), Alvo (1977). The papers of Girshick et al. (1946) and DeGroot (1959) contain the study of unbiased sequential procedures for the binomial process. DeGroot (1959) proved the efficiency of simple and inverse procedures. Magiera and Trybuła (1976) found a new efficient (in the Cramér–Rao–Wolfowitz inequality sense) sequential procedure for the binomial process, the so-called oblique plan. Braess and Dette (2004) considered minimax approach to the estimation of constrained binomial and multinomial probabilities.

In sequential point estimation we are interested in estimating a value of an unknown parameter (or function of an unknown parameter). Usually the loss incurred by the statistician and associated with the error of estimation is represented by a weighted function of the difference between the true value of an unknown parameter and its estimate (e.g. the loss given by formula (1) or (2)). The natural question arises how many observations we should take to guarantee that the decision based on these observations will have some desirable properties. If all parameters of the model were known, then we could calculate this optimal fixed sample size n^*. In reality we do not know these parameters, so we construct the so called sequential procedure (N,d_N) consisting of a random stopping time (a random sample size) N terminating the observations, and the decision (estimator) d_N taken when we stop sampling.

It is reasonable to demand that the sequential procedure (N, d_N) possesses some desirable optimal properties, such as: efficiency $N/n^* \to 1$, asymptotic normality and risk (expected loss) consistency. The first condition asserts that using the stopping rule N requires approximately the same number n^* of observations as in the case when all parameters of the model were known. The second concept is also a measure of efficiency. The latter term is understood in this paper in the following way. The risk for the sequential procedure (N, d_N) preserves approximately the same level as in the case of the optimal procedure with the fixed sample size n^*. We then say that the sequential procedure possesses the risk consistency property. All the asymptotic properties derived for the model considered are treated as asymptotic when the probability of success p tends to zero.

Hubert and Pyke (2000) considered the problem of sequentially estimating the probability of success p under the loss

$$\mathcal{L}(p, d_n) = p^b \left(d_n - p^a\right)^2, \tag{1}$$

where d_n denotes an estimator based on the sample mean, and the constants a and b satisfy the condition $2a + b < 0$. They required the risk to be close to a prescribed constant c and studied the asymptotic properties in the case when p approaches zero. They obtained the following efficiency and normality results: $N/n^*(p) \to 1$ as $p \to 0$ and

$$\frac{N - n^*(p)}{s\sqrt{p^{-1}n^*(p)}} \xrightarrow{\mathcal{D}} \mathcal{Z} \quad \text{as} \quad p \to 0,$$

respectively, where N is the stopping rule, $n^*(p)$ is the optimal fixed sample size minimizing the risk for a given p, s is a certain constant, $\mathcal{Z}$ is the standard normal distribution random variable and $\stackrel{\mathcal{D}}{\to}$ denotes the convergence in distribution. They also showed the sequential risk consistency:

$$\lim_{p\to 0} \text{Risk} = c,$$

where c is a prescribed constant.

The particular case $a = 1, b = -2$ in (1) for which $2a + b = 0$ is precisely the case examined by Robbins and Siegmund (1976). They required the risk to be equal to a constant c, like in Hubert and Pyke (2000), but their asymptotic study of the problem concerns the case when this constant is allowed to tend to zero for fixed p. In this case they obtained:

$$\frac{N}{n^*(p)} = cpqN \to 1 \text{ in probability as } c \to 0,$$
$$\frac{\text{Risk}}{c} \to 1 \text{ as } c \to 0,$$

where $q = 1 - p$. When $p \neq 1/2$,

$$\lim_{c\to 0} P\left\{\sqrt{\frac{1}{c}}\left(\frac{N - n^*(p)}{n^*(p)}\right) \leq t\right\} = \Phi\left(\frac{t}{q-p}\right),$$

where Φ is the Gaussian distribution function. For $p = 1/2$ we have

$$P\big(N - [4/c] \leq j\big) - G\big([4/c] - 4/c + j\big) \to 0,$$

where $j = 0, 1, \ldots, n = [4/c] + j \to \infty$ and G is the chi-square distribution function with one degree of freedom.

Cabilio and Robbins (1975) also examined this case adding a constant cost c per observation to the loss function. They constructed a sequential procedure (N, δ_N) and for fixed p obtained the following results as $c \to 0$:

$$N(cpq)^{1/2} = N(cp(1-p))^{1/2} \to 1 \text{ a.s.},$$
$$E_p\{N(cpq)^{1/2}\}^k \to 1, \quad k = 1, 2, \ldots,$$
$$\left(\frac{pq}{c}\right)^{1/2}\left(\frac{\delta_N - p}{pq}\right)^2 \stackrel{\mathcal{D}}{\to} \mathcal{Z}^2,$$
$$\left(\frac{pq}{c}\right)^{1/2} E_p\left(\frac{\delta_N - p}{pq}\right)^2 \to 1,$$

where δ_N is the decision when we stop at time N. This procedure is well defined for all values of p. However, for all fixed $c > 0$, when $p \to 0$ it does not perform well, because the convergence

$$\frac{E_p\left(\left(\frac{\delta_N - p}{pq}\right)^2 + cn\right)}{2\sqrt{c/pq}} \to 1,$$

is not uniform in p. They introduced a uniform prior distribution on p to avoid the latter inconvenience.

2 The underlying problem

In this paper, we consider the problem of sequential estimation of a function of the probability of success p under the generalized LINEX loss function defined by

$$\mathcal{L}(p, d_n) = h_1(p)\left(e^{h_2(p)\Delta_n} - h_2(p)\Delta_n - 1\right), \tag{2}$$

where $h_1(p)$, $h_2(p)$ are some functions of p and $\Delta_n = d_n - g(p)$, where d_n is an estimator of $g(p)$. In comparison with the standard LINEX loss it is a linear exponential weighted error loss function. As any LINEX function it is an asymmetric function of the error. A squared weighted loss function can be viewed as an approximation of the loss (2), when $h_2(p)$, possibly depending on some constants, is small.

Assuming $h_1(p) = b$, $h_2(p) = ap^l$ and $g(p) = p^s$, a stopping time N is constructed, which determines the sequential procedure $\delta_N = (N, d_N)$ possessing some asymptotically optimal properties in the case when p tends to zero. The conditions are given under which the stopping time N is asymptotically efficient, i.e. $N/n^*(p) \to 1$ as $p \to 0$, where $n^*(p)$ is the optimal fixed sample size for which the risk equals approximately a prescribed constant c. In this approach, when p tends to zero, the asymptotic normality of the stopping time and the corresponding sequential estimator of $g(p) = p^s$ is also obtained. The sequential procedure constructed guarantees that the sequential risk is also close to the predefined constat c when p is small.

3 The sequential procedure

Let $\{X_i : i \geq 1\}$ be a sequence of independent Bernoulli distributed, with parameter p, random variables. The general problem is to sequentially estimate the value $g(p)$ of a smooth function of p. As a measure of the error of estimation we assume the loss function defined by (2).

In the sequel the estimators will be assumed to have the form $d_n = g(\overline{X}_n)$, where $\overline{X}_n$ is the sample mean. From the well known inequality $e^x \geq 1 + x$ we know that $e^{h_2(p)\Delta_n} - h_2(p)\Delta_n - 1 \geq 0$, so we assume that $h_1(p) > 0$ to make the risk non-negative.

Let us note, that the risk under the LINEX loss given by (2) can be written in the following form

$$\mathcal{R}(p, d_n) = h_1(p)E\left(\exp(h_2(p)(g(\overline{X}_n) - g(p))) - h_2(p)(g(\overline{X}_n) - g(p)) - 1\right)$$
$$= h_1(p)E\left[\exp\left(\frac{h_2(p)\left|g'(p)\right|\sqrt{pq}}{\sqrt{n}}\frac{\sqrt{n}(g(\overline{X}_n) - g(p))}{|g'(p)|\sqrt{pq}}\right)\right.$$
$$\left.-\left(\frac{h_2(p)\left|g'(p)\right|\sqrt{pq}}{\sqrt{n}}\frac{\sqrt{n}(g(\overline{X}_n) - g(p))}{|g'(p)|\sqrt{pq}}\right) - 1\right].$$

If g is sufficiently smooth, then by delta method for in law approximations,

$$\frac{\sqrt{n}(g(\overline{X}_n) - g(p))}{|g'(p)|\sqrt{pq}} \xrightarrow{\mathcal{D}} \mathcal{Z} \quad \text{as } n \to \infty. \tag{3}$$

Using the approximation given by (3) and putting $q = 1 - p \cong 1$ (because we are interested in small p), we have

$$\mathcal{R}(p, d_n) \cong h_1(p)E\left(\exp\left(\frac{h_2(p)\left|g'(p)\right|\sqrt{p}}{\sqrt{n}}\mathcal{Z}\right) - 1\right).$$

Since the random variable

$$\exp\left(\frac{h_2(p)\left|g'(p)\right|\sqrt{p}}{\sqrt{n}}\mathcal{Z}\right)$$

has the $\mathcal{LN}(\mu, \sigma^2)$ with parameters $\mu = 0$ and $\sigma^2 = \left(h_2(p)g'(p)\right)^2 p/n$, we obtain

$$\mathcal{R}(p, d_n) \cong h_1(p)\ \exp\left(\frac{\left(h_2(p)g'(p)\right)^2 p}{2n}\right) - h_1(p).$$

In the sequel we shall construct the sequential procedure $\delta_N = (N, d_N)$ which guarantees that the sequential risk $\mathcal{R}(p, d_N)$ is close to the predefined constant c.

The optimal sample size n^*, when we know the true value of p, can be calculated from the equation

$$\mathcal{R}(p, d_n) \cong\ h_1(p)\ \exp\left(\frac{\left(h_2(p)g'(p)\right)^2 p}{2n}\right) - h_1(p) = c.$$

Hence, we obtain

$$n^* = \frac{\left(h_2(p)g'(p)\right)^2 p}{2\log\left(\frac{c}{h_1(p)} + 1\right)}. \tag{4}$$

Thus, if we know p, then the optimal sample size is given by (4). We denote this optimal n^* by $n^*(p)$. But we do not know the true value of p, so we can consider $n^*(\overline{X}_n)$ as an approximation of $n^*(p)$ (by taking the sample mean $\overline{X}_n$ instead of p). In this case it is natural to consider the stopping time

$$N = \inf\left\{n \geq 1 : n \geq n^*(\overline{X}_n)\right\}.$$

Let us note that

$$N = \inf\left\{n \geq 1 : n \geq \frac{\left(h_2(\overline{X}_n)g'(\overline{X}_n)\right)^2 \overline{X}_n}{2\log\left(\frac{c}{h_1(\overline{X}_n)} + 1\right)}\right\}. \tag{5}$$

We see that formula (5) has a quite general form. To make the problem more tractable, in the rest of this paper, we take into consideration the following assumptions:

$$\begin{aligned} h_1(p) &= b, \quad \text{where } b \in \mathbb{R}_+, \\ h_2(p) &= ap^l, \quad \text{where } a \in \mathbb{R}\setminus\{0\} \text{ and } l \in \mathbb{R}, \\ g(p) &= p^s, \quad \text{where } s \in \mathbb{R}\setminus\{0\}. \end{aligned}$$

Under these assumptions, the loss function and the risk function take the form

$$\begin{aligned} \mathcal{L}(p, d_n) &= b\left(e^{ap^l(d_n - p^s)} - ap^l\left(d_n - p^s\right) - 1\right) \\ &= b\left(e^{ap^l\left(\overline{X}_n^s - p^s\right)} - ap^l\left(\overline{X}_n^s - p^s\right) - 1\right) \end{aligned} \tag{6}$$

and

$$\mathcal{R}(p, d_n) \cong b\ \exp\left(\frac{(as)^2 p^{2(l+s)-1}}{2n}\right) - b,$$

respectively.

Let us remark that only the condition $l + s \leq 0$ is relevant to the problem. This remark is an easy consequence of the model assumed. In particular, if one takes d_n to be identically zero for every n, then by (6), $\mathcal{L}(p, d_n) = b\left(e^{-ap^{l+s}} + ap^{l+s} - 1\right)$. Thus, if $l + s > 0$, any stopping rule, including $N = 0$, satisfies

$$\mathcal{R}(p, d_N) = b\left(e^{-ap^{l+s}} + ap^{l+s} - 1\right) \overset{p\to 0}{\to} 0.$$

Then it means that if $l + s > 0$, one can estimate p^s with arbitrarily small risk as $p \to 0$ without taking any observations. Hence, we assume that $l + s \leq 0$ to avoid the above shortcoming, but the case $l + s = 0$ will be discussed separately in the further part of the paper (see Remark 1 in Appendix).

For $l = 0$ (and $s < 0$) the loss function takes the standard form of the LINEX loss for estimating p^s:

$$\mathcal{L}(p, d_n) = b\left(e^{a(d_n - p^s)} - a(d_n - p^s) - 1\right),$$

including the loss in estimating $1/p$. For $s = 1$ (and $l \le -1$) we have

$$\mathcal{L}(p, d_n) = b\left(e^{ap^l(d_n - p)} - ap^l\,(d_n - p) - 1\right).$$

This is the linear exponential weighted error loss function with the weight p^l of the error $\Delta_n = d_n - p$.

In the case considered, we give now a representation for the stopping time N. From formula (4) we have

$$n^*(p) = \left(\frac{as}{D}\right)^2 p^{2(l+s)-1}, \tag{7}$$

where

$$D^2 = 2\log\left(\frac{c}{b} + 1\right).$$

Thus

$$n^*(\overline{X}_n) = \left(\frac{as}{D}\right)^2 (\overline{X}_n)^{2(l+s)-1} = \left(\frac{1}{n}\right)^{2(l+s)-1} \left(\frac{as}{D}\right)^2 S_n^{2(l+s)-1},$$

and the stopping time N defined by (5) can be written in the following form

$$\begin{aligned} N &= \inf\left\{n \ge 1 : S_n^{2(l+s)-1} \le n^{2(l+s)} \left(\frac{as}{D}\right)^{-2}\right\} \\ &= \inf\left\{n \ge 1 : S_n \ge \left(\frac{as}{D}\right)^{-\frac{2}{2(l+s)-1}} n^{\frac{2(l+s)}{2(l+s)-1}}\right\}, \end{aligned} \tag{8}$$

provided that $l + s < 1/2$. In other case we can stop without any observations.

Assume that $l + s < 0$ and denote

$$\rho = \frac{2(l+s)}{2(l+s)-1}. \tag{9}$$

Observe that

$$0 < \rho = \frac{2(l+s)}{2(l+s)-1} = 1 + \frac{1}{2(l+s)-1} < 1,$$

because $2(l + s) - 1 < 0$. Therefore $\rho \in (0, 1)$. Moreover, $\frac{-1}{2(l+s)-1} = 1 - \rho$. Taking into account notation (9), formula (8) takes the form

$$N = \inf \left\{ n \geq 1 : S_n \geq \left(\frac{as}{D} \right)^{2(1-\rho)} n^{\rho} \right\}.$$

Thus the stopping rule N can be expressed in the form

$$N = \inf \{ n \geq 1 : S_n \geq C(n) \}, \tag{10}$$

where

$$C(n) = A^{1-\rho} n^{\rho} = A_1 n^{\rho}, \quad A_1 := A^{1-\rho} \tag{11}$$

with

$$A = \left(\frac{as}{D} \right)^2. \tag{12}$$

For further purposes, let us note that under the notations given above, the optimal fixed sample size $n^*(p)$ defined by (7) can be expressed in the form

$$n^*(p) = A \; p^{2(l+s)-1} = A \; p^{-\frac{1}{1-\rho}} = \left(\frac{A_1}{p} \right)^{\frac{1}{1-\rho}}. \tag{13}$$

Thus, we have shown that the stopping rule N for the problem of estimation under the LINEX loss function can be presented in a form analogous to that obtained by Hubert and Pyke (2000) for a squared error loss. We then can formulate the following theorems, obtained by using the arguments analogous to that of Hubert and Pyke (2000).

Theorem 1 (efficiency of the stopping time) *If* $l + s < 0$ *and* N, $n^*(p)$ *are defined by* (10), (13)*, respectively, then*

$$\frac{N}{n^*(p)} \xrightarrow{P} 1 \quad as \quad p \to 0.$$

Proof (sketch following the methods and arguments of Hubert and Pyke (2000)) The following formula holds

$$P\left(\left| \frac{N}{n^*(p)} - 1 \right| \leq \epsilon \right) = P\left((1-\epsilon) n^*(p) \leq N \leq (1+\epsilon) n^*(p) \right)$$
$$\geq 1 - P(\tilde{A}) - P(\tilde{B}),$$

where

$$\tilde{A} = \{N < n_m(p)\},$$
$$\tilde{B} = \left\{|S_i - ip| > Ap^{-h(\rho)}c_\epsilon, \quad \text{for some } 1 \le i \le L_\epsilon(p)\right\},$$

with $n_m(p)$ denoting the value of x where the function $C(x) - xp$ takes its maximum and

$$L_\epsilon(p) = (1+\epsilon)\, n^*(p), \quad h(\rho) = \rho/(1-\rho), \quad c_\epsilon = (1-\epsilon)^\rho - (1-\epsilon),$$

where ρ is defined by (9). Thus it suffices to show that $P(\tilde{A})$ and $P(\tilde{B})$ tend to zero when p approaches zero. To prove the first convergence, the Hájek–Rényi inequality is used to get

$$P(\tilde{A}) \le \int_0^\infty \frac{p}{(C(x) - xp)^2} 1_{(0, n_m(p)]}(x) dx,$$

and then the dominated convergence theorem is applied. The main tool for obtaining the second convergence is Kolmogorov's inequality. □

Theorem 2 (asymptotic normality of the stopping time) *Let $l + s < 0$ and let N, $n^*(p)$ be as in Theorem 1. Then*

$$\frac{N - n^*(p)}{s_p} \xrightarrow{\mathcal{D}} \mathcal{Z} \quad \text{as } p \to 0,$$

where $s_p^2 = (1-\rho)^{-2} p^{-1} n^*(p)$ and ρ is defined by (9).

Proof (sketch following the methods and arguments of Hubert and Pyke (2000)) Observe that

$$P\left(\frac{N - n^*(p)}{s_p} \le x\right) = P\left(N \le z_p(x)\right),$$

where $z_p(x) = n^*(p) + xs_p$, and it suffices to show that

$$\liminf_{p\to 0} P\left(N \le z_p(x)\right) \ge \Phi(x) \quad \text{and} \quad \limsup_{p\to 0} P\left(N \le z_p(x)\right) \le \Phi(x),$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

Since $s_p/n^*(p) = \text{const} \cdot p^{\rho/2(1-\rho)} \to 0$ as $p \to 0$, there exists a sequence $\left\{\epsilon_p, p > 0\right\}$ such that $\epsilon_p \to 0$ as $p \to 0$, $\epsilon_p > K s_p/n^*(p)$ for some known positive constant K and $(1-\epsilon_p)\, n^*(p) =: n_p$ is an integer.

In the first step one shows that

$$P\left(N \le z_p(x)\right) \ge P\left(Z(p) \ge \frac{\delta_{1,\epsilon_p}(x) - pn_p}{(pn_p)^{1/2}} + \epsilon\right) P\left(A_{x,\epsilon}\right),$$
$$P\left(N \le z_p(x)\right) \le P\left(Z(p) \ge \frac{\delta_{2,\epsilon_p}(x) - pn_p}{(pn_p)^{1/2}} - \epsilon\right) P\left(A_{x,\epsilon}\right) + P\left(A^c_{x,\epsilon}\right),$$

where

$$Z(p) = (S_{n_p} - pn_p)/\sqrt{pn_p},$$
$$\delta_{1,\epsilon_p}(x) - pn_p = -x\sqrt{pn_p}/\sqrt{1-\epsilon_p},$$
$$\delta_{2,\epsilon_p}(x) - pn_p = -pxs_p\left(1 - \rho + O\left(p^{\rho/2(1-\rho)}\right)\right)$$

and

$$A_{x,\epsilon} = \left\{|S_k - kp| \le \epsilon(pn_p)^{1/2}, k = 1, 2, \ldots, z_p(x) - n_p\right\}, \quad \epsilon > 0.$$

Using Kolmogorov's inequality, one obtains that $P\left(A^c_{x,\epsilon}\right) \to 0$ as $p \to 0$. Then, taking into account that $\epsilon_p > 0$ and $\epsilon_p \to 0$ as $p \to 0$, one shows that given $\eta > 0$ there exists $p_0 > 0$ such that

$$\frac{\delta_{1,\epsilon_p}(x) - pn_p}{\sqrt{pn_p}} = -x/\sqrt{1-\epsilon_p} < -x + \eta,$$
$$\frac{\delta_{2,\epsilon_p}(x) - pn_p}{\sqrt{pn_p}} = -x/\sqrt{1-\epsilon_p} - xO\left(p^{\rho/2(1-\rho)}\right) > -x - \eta$$

for all $p < p_0$ and all x. The proof of the theorem follows from the fact that $Z(p) \xrightarrow{\mathcal{D}} \mathcal{Z}$ as $p \to 0$. □

Let us remark that the convergence in the theorem given above is a version of the CLT for Bernoulli trials in which the asymptotics are driven by the parameter p approaching zero.

Corollary 1 (asymptotic normality of the sequential estimator) *Let the assumptions of Theorem 2 be satisfied. Then we have*

$$\frac{ap^l\left(\overline{X}^s_N - p^s\right)}{D} \xrightarrow{\mathcal{D}} \mathcal{Z} \quad as\ p \to 0.$$

Proof Since, by construction, $C(N) \leq S_N < C(N) + 1$ and by (11), $C(N) = A_1 N^{\rho}$, we obtain

$$A_1 N^{\rho} \leq S_N < A_1 N^{\rho} + 1,$$
$$A_1 N^{\rho-1} \leq \overline{X}_N < A_1 N^{\rho-1} + N^{-1}.$$

From Theorem 1 we know that for sufficiently small p, $\frac{1}{N}$ is so small as we want. Therefore

$$\overline{X}_N \cong A_1 N^{\rho-1} = \frac{C(N)}{N}.$$

Let

$$W_p = \frac{N - n^*(p)}{s_p}.$$

Note that from Theorem 2, $W_p \stackrel{\mathcal{D}}{\rightarrow} \mathcal{Z}$, when $p \to 0$. Hence

$$p^l \left\{ \left(\overline{X}_N\right)^s - p^s \right\} \cong p^l \left\{ \left(\frac{C(N)}{N}\right)^s - p^s \right\}$$

for respectively small p. Further

$$\begin{aligned}
p^l \left\{ \left(\overline{X}_N\right)^s - p^s \right\} &\cong p^l \left\{ \left(A_1 N^{\rho-1}\right)^s - p^s \right\} = p^l \left\{ A_1^s (N)^{s(\rho-1)} - p^s \right\} \\
&= p^{l+s} \left\{ A_1^s (N)^{s(\rho-1)} p^{-s} - 1 \right\} \\
&= p^{l+s} \left\{ A_1^s \left(\frac{N - n^*(p)}{s_p} s_p + n^*(p) \right)^{s(\rho-1)} p^{-s} - 1 \right\} \\
&= p^{l+s} \left\{ A_1^s \left(W_p s_p + n^*(p) \right)^{s(\rho-1)} p^{-s} - 1 \right\} \\
&= p^{l+s} \left\{ \left(\frac{A_1}{p}\right)^s n^*(p)^{-s(1-\rho)} \left(W_p \frac{s_p}{n^*(p)} + 1 \right)^{s(\rho-1)} - 1 \right\} \\
&= p^{l+s} \left\{ \left(W_p \frac{s_p}{n^*(p)} + 1 \right)^{s(\rho-1)} - 1 \right\},
\end{aligned}$$

where the last equality is a consequence of the definition of $n^*(p)$. From Taylor's theorem for $(x+1)^{s(\rho-1)}$ with $x = s_p W_p / n^*(p)$ we have

$$p^l \left\{ \left(\overline{X}_N\right)^s - p^s \right\} \cong p^{l+s} \left\{ 1 + s(\rho - 1) \frac{s_p}{n^*(p)} W_p + O\left(\left(W_p \frac{s_p}{n^*(p)} \right)^2 \right) - 1 \right\}.$$

Since

$$\frac{s_p}{n^*(p)} = \frac{1}{1-\rho} A_1^{-1/2(1-\rho)} p^{-(l+s)} \overset{p\to 0}{\to} 0,$$

we obtain

$$\begin{aligned} p^l \left\{ \left(\overline{X}_N\right)^s - p^s \right\} &\cong p^{l+s} \Bigg\{ 1 + s(\rho - 1) \frac{1}{1-\rho} A_1^{-1/2(1-\rho)} p^{-(l+s)} W_p \\ &\quad + O\left(\left(W_p \frac{1}{1-\rho} A_1^{-1/2(1-\rho)} p^{-(l+s)} \right)^2 \right) - 1 \Bigg\} \\ &= p^{l+s} \Bigg\{ -s(1-\rho) \frac{1}{1-\rho} A_1^{-1/2(1-\rho)} p^{-(l+s)} W_p \\ &\quad + O\left(\left(W_p \frac{1}{1-\rho} A_1^{-1/2(1-\rho)} p^{-(l+s)} \right)^2 \right) \Bigg\} \\ &= -s A_1^{-1/2(1-\rho)} W_p + O\left(W_p^2 \frac{1}{(1-\rho)^2} A_1^{-1/(1-\rho)} p^{-2(l+s)} \right) \\ &\overset{D}{\to} s A_1^{-1/2(1-\rho)} \mathcal{Z} = \frac{1}{a} D \mathcal{Z} \end{aligned}$$

as $p \to 0$. The convergence is a consequence of Theorem 2. Hence the proof is complete. □

4 The consistency of the sequential procedure

In order to show that the risk of the procedure with the stopping rule N is close to the pre-specified constant c the result on uniform integrability of the family of loss functions is needed. Theorems 3 and 4, connected with the uniform integrability, can be established by the arguments of Hubert and Pyke (2000). To establish these results, it is necessary to involve a minimum stopping time in order to avoid the effect of an initial sequence of successes. Therefore we introduce the lower bound

$$m_r = \inf\left\{k > 0 : C(k) > -r \min\{l, l+s\}\right\},$$

where $r \geq r_0$ and r_0 guarantees that we do not stop too fast, so that the approximation (3) is suitably exact.

Let

$$\begin{aligned} N_r &= \inf\{n \geq m_r : n \geq n^*(\overline{X}_n)\} \\ &= \inf\{n \geq m_r : S_n \geq C(n)\}. \end{aligned}$$

The rule N_r guarantees that we take minimum m_r observations.

Define

$$Z_{p,r} := p^l \left(\overline{X}^s_{N_r} - p^s\right), \quad Z^+_{p,r} := p^l \left|\overline{X}^s_{N_r} - p^s\right|. \tag{14}$$

Let

$$\tau_r = -r \min\{l, l+s\}, \tag{15}$$

$$\gamma = \begin{cases} 1 & \text{for } s > 0, \\ \min\left\{1, -\frac{1}{2s(1-\rho)}\right\} & \text{for } s < 0. \end{cases} \tag{16}$$

Theorem 3 *Let* τ_r *and* γ *be defined by* (15) *and* (16). *For* $\{Z^+_{p,r} : p \in (0,1)\}$ *as defined in* (14)*, there exist positive constants* $K, K_1, K_2, K_3, \tau > \tau_r$ *and* $0 < \epsilon < 1$ *such that for all* $u \geq 0$:

(i) $P\left(Z^+_{p,r} > u, N_r < n^*(p)\right) \leq K_1 p^\tau + K_2 e^{-K_3 u}$ *for p sufficiently small,*

(ii) $P\left(Z^+_{p,r} > u, N_r \geq n^*(p)\right) \leq e^{-K u^{2\gamma}}$ *for* $p \in (0,1)$.

Proof (sketch following the methods and arguments of Hubert and Pyke (2000)) If $s > 0$, then using the approximation $\overline{X}_{N_r} \cong C(N_r)/N_r$ one can show that $P\left(Z^+_{p,r} > u, N_r < n^*(p)\right) = P\left(A^+_{p-}\right)$, where

$$A^+_{p-} = \left\{N_r < n^*(p)\, t_{p-}(u)\right\}, \quad t_{p-}(u) = \left(1 + u p^{\rho/2(1-\rho)}\right)^{-1/s(1-\rho)}.$$

Let $\epsilon \in (0,1)$, set $v_p = \text{const} \cdot p^{(1-\epsilon)/2(1-\rho)}$, and let w_p be the value for which $C\left(n^*(p)v_p\right) - n^*(p)v_p p = C\left(n^*(p)w_p\right) - n^*(p)w_p p$. Moreover, define u_1 and u_2 by $t_{p-}(u_1) = v_p$, $t_{p-}(u_2) = w_p$, respectively.

Set $I_1 = (u_1, \infty)$, $I_2 = (u_2, u_1]$ and $I_3 = [0, u_2]$. Let $u \in I_1$. If u is such that $t_{p-}(u) < m_r$, then $P\left(A^+_{p-}\right) = 0$. Therefore, only those u are considered for which $t_{p-}(u) \geq m_r$. To find upper boundary for the probability $P\left(A^+_{p-}, u \in I_1\right)$, define u_3 by $n^*(p)\, t_{p-}(u_3) = m'$, where m' is a constant greater than m_r which will be given later. Then this probability is not greater than the probability that S_n crosses the line $C(m_r)$ for $n = m'$ ($u \in I_{1,1} = (u_3, \infty)$) plus the probability that S_n crosses the line $L_1(t) = \alpha_p t + \beta_p$ which passes through the points $(m', C(m'))$ and $(n^*(p)v_p, C(n^*(p)v_p))$ ($u \in I_{1,2} = (u_1, u_3)$). To estimate the latter one, Täcklind (1942) result is used to obtain: P (S_n crosses the line $L_1) \leq p^{k_0}$, where k_0 is a positive constant and p is sufficiently small. Choosing k_0, m' and ϵ such that $k_0 > -rl$, $C(m') > 2k_0$, $1/2 - \epsilon/2 - k_0/C(m') > 0$, gives

$$P\left(A^+_{p-}, u \in I_1\right) \leq \text{const} \cdot p^{\tau_r}$$

for p sufficiently small. If $u \in I_2$, $L_2(t) = pt + \delta$ is defined, where

$$\delta = C(n^*(p)v_p) - n^*(p)v_p p = C(n^*(p)w_p) - n^*(p)w_p p.$$

The probability of crossing the boundary consisting of the lines $\lceil C(m_r) \rceil$, L_1 and L_2 (defined on intervals $I_{1,1}$, $I_{1,2}$ and I_2, respectively) is not greater than the upper bound for the probability $P\left(A^+_{p-}, u \in I_1\right)$ plus the probability that S_n crosses the line L_2 between 0 and $n^*(p)$. For the latter one, using the Skorokhod and Bernstein inequalities yields

$$P\left(S_n \text{ crosses the line } L_2\right) \leq e^{-\text{const}\cdot p^{-\epsilon\rho/(1-\rho)}}.$$

Thus

$$P\left(A^+_{p-}, u \in I_2\right) \leq e^{-\text{const}\cdot p^{-\epsilon\rho/(1-\rho)}} + \text{const} \cdot p^{\tau_r}.$$

For $u \in I_3$ the method is analogous to that for $u \in I_2$. Thus, combining these probability bounds gives

$$P\left(A^+_{p-}\right) \leq \text{ const} \cdot p^{\tau_r} + e^{-\text{ const}\cdot u}$$

for $u \geq 0$ and p sufficiently small.

In the case $s < 0$ the proof is similar. Analogously, the intervals I_1, I_2 and I_3 are defined. The steps are the same as before but we choose k_0, ϵ and m' such that $k_0 > -r(s+l)$, $C(m') > 2k_0$, $1/2 - \epsilon/2 - k_0/C(m') > 0$.

The probability $P\left(Z^+_{p,r} > u, N_r \geq n^*(p)\right)$ can be expressed by

$$P\left(A^+_p\right) = P\left(N_r > n^*(p)\left(1 - up^{\rho/2(1-\rho)}\right)^{-1/s(1-\rho)}\right) \quad \text{for } s > 0,$$
$$P\left(A^-_p\right) = P\left(N_r > n^*(p)\left(1 + up^{\rho/2(1-\rho)}\right)^{-1/s(1-\rho)}\right) \quad \text{for } s < 0.$$

If $s > 0$, this probability is equal to zero unless $u < p^{-\rho/2(1-\rho)}$. By Bennet's inequality it can be shown that

$$P\left(A^+_p\right) \leq e^{-\text{const}\cdot u^2 J(x)} \quad \text{and} \quad P\left(A^-_p\right) \leq e^{-\text{const}\cdot u^{2\epsilon} J_\epsilon(x)},$$

where $x = up^{\rho/2(1-\rho)}$ and

$$J(x) = x^{-2}(1-x)^{-1/s(1-\rho)}\left(1 - (1-x)^{1/s}\right)^2,$$
$$J_\epsilon(x) = x^{-2\epsilon}(1+x)^{-1/s(1-\rho)}\left((1+x)^{1/s} - 1\right)^2.$$

Since the functions J and J_ϵ have a positive minimum (the latter one under $\epsilon = \gamma = \min\{1, -1/2s(1-\rho)\}$), one obtains

$$P\left(A^+_p\right) \leq e^{-\text{const}\cdot u^2} \quad \text{and} \quad P\left(A^-_p\right) \leq e^{-\text{const}\cdot u^{2\gamma}}$$

for $p \in (0, 1)$, $u > 0$ and an appropriate const. □

Theorem 4 *For $r > 0$, the family $\left\{(Z^{+}_{p,r})^r : 0 < p < 1\right\}$ is uniformly integrable.*

Proof (sketch following the methods and arguments of Hubert and Pyke (2000)) It suffices to show that for some $w > r$

$$E\left(Z^{+}_{p,r}\right)^{w} = \int_{0}^{\infty} P\left(Z^{+}_{p,r} > u\right) du^{w} = I_1 + I_2 < \infty,$$

where I_1 and I_2 are given by

$$\int_{0}^{\infty} P\left(Z^{+}_{p,r} > u, N_r < n^{*}(p)\right) du^{w} \quad \text{and} \quad \int_{0}^{\infty} P\left(Z^{+}_{p,r} > u, N_r \geq n^{*}(p)\right) du^{w},$$

respectively. Let us note that from definition of N_r and the condition $N_r < n^{*}(p)$ it follows that $\overline{X}_{N_r} > p$ (this will be discussed closer in the proof of Theorem 5), so I_1 is bounded by

$$\int_{0}^{p^{l}} P\left(Z^{+}_{p,r} > u, N_r < n^{*}(p)\right) du^{w} \quad \text{and} \quad \int_{0}^{p^{l+s}} P\left(Z^{+}_{p,r} > u, N_r < n^{*}(p)\right) du^{w}$$

for $s > 0$ and $s < 0$, respectively. Using Theorem 3 for p sufficiently small, say $p \leq p_0$, one obtains that $\sup_{p \leq p_0} I_1 < \infty$. Note that for $p > p_0 > 0$, the bounded function is integrated on a finite support, so $\sup_{0<p<1} I_1 < \infty$. In the case of the integral I_2, applying Theorem 3 yields $\sup_{0<p<1} I_2 < \infty$. □

Theorem 5 *Let $l + s < 0$, $\gamma > 1/2$ and let $Y_{p,r} = \exp\left\{aZ_{p,r}\right\}$, where $Z_{p,r}$ is given by (14). If $a > 0$, $s < 0$ or $a < 0$, $s > 0$, then the family $\left\{Y_{p,r} : 0 < p < 1\right\}$ is uniformly integrable.*

Proof It suffices to show that

$$E\left(Y^{w}_{p,r}\right) = \int_{0}^{\infty} P\left(Y_{p,r} > u\right) du^{w} < \infty,$$

where $w > 1$. Write

$$\int_{0}^{\infty} P\left(Y_{p,r} > u\right) du^{w} = \int_{0}^{\infty} P\left(Y_{p,r} > u, N_r < n^{*}(p)\right) du^{w} + \int_{0}^{\infty} P\left(Y_{p,r} > u, N_r \geq n^{*}(p)\right) du^{w}. \tag{17}$$

Let us note that in the case when $a > 0,\ s < 0$ or $a < 0,\ s > 0$, the random variable $Y_{p,r}$ is bounded by 1. This observation is an easy consequence of the fact that the condition $N_r < n^*(p)$ is equivalent to $\overline{X}_{N_r} > p$. This follows from the following argumentation. By definition, $N_r = \inf\{n \geq m_r : S_n \geq C(n) = A_1 n^{\rho}\}$, we have $\overline{X}_{N_r} = \frac{S_{N_r}}{N_r} \geq A_1 N_r^{\rho-1}$. If $N_r < n^*(p) = (A_1/p)^{1/(1-\rho)}$, then $A_1 N_r^{\rho-1} \geq p$. Thus $\overline{X}_{N_r} > p$, and consequently $Y_{p,r} \leq 1$. Hence, for all p we obtain

$$\int_0^{\infty} P\left(Y_{p,r} > u, N_r < n^*(p)\right) du^w = \int_0^1 P\left(Y_{p,r} > u, N_r < n^*(p)\right) du^w \leq \int_0^1 du^w < \infty.$$

The second integral in formula (17) can be written in the following form

$$\int_0^{\infty} P\left(Y_{p,r} > u, N_r \geq n^*(p)\right) du^w = \int_0^1 P\left(e^{aZ_{p,r}} > u, N_r \geq n^*(p)\right) du^w + \int_1^{\infty} P\left(e^{aZ_{p,r}} > u, N_r \geq n^*(p)\right) du^w. \tag{18}$$

Let us remark that

$$\int_0^1 P\left(e^{aZ_{p,r}} > u, N_r \geq n^*(p)\right) du^w \leq \int_0^1 du^w < \infty.$$

As concerns the second integral in formula (18), we observe that

$$\int_1^{\infty} P\left(e^{aZ_{p,r}} > u, N_r \geq n^*(p)\right) du^w \leq \int_1^{\infty} P\left(|a|\, Z_{p,r}^+ > \log u, N_r \geq n^*(p)\right) du^w.$$

By substituting $v = \log u / |a|$ into the above integral, we obtain

$$w \int_0^{\infty} P\left(Z_{p,r}^+ > v, N_r \geq n^*(p)\right) e^{|a|vw} dv \leq w \int_0^{\infty} e^{-Kv^{2\gamma}} e^{|a|vw} dv,$$

where K is a positive constant. The last inequality is a consequence of Theorem 3. If $\gamma > 1/2$, then

$$w \int_0^\infty e^{-Kv^{2\gamma}+|a|vw} dv < \infty,$$

which ends the proof. □

Let us note that

$$\mathcal{L}\left(p, d_{N_r}\right) = b\left(Y_{p,r} - aZ_{p,r} - 1\right) =: \mathcal{L}_r(p).$$

By Theorems 4 and 5 the pass with the limit under the expectation sign is allowed, and as a consequence the following theorem holds.

Theorem 6 (consistency of the sequential procedure) *By Theorems 4, 5 and Corollary 1, for any $r \geq r_0$ the sequential estimation procedures based on the stopping time N_r satisfy*

$$\lim_{p\to 0} \text{Risk} = \lim_{p\to 0} E\mathcal{L}_r(p) = c.$$

Proof By the uniform integrability of the $Y_{p,r}$, $Z^r_{p,r}$ and Corollary 1 we have

$$\lim_{p\to 0} E\ \mathcal{L}_r(p) = bE\ \lim_{p\to 0}\left(Y_{p,r} - a\ Z_{p,r} - 1\right) = bE\left(e^{a\ \frac{1}{a}\ D\ \mathcal{Z}} + a\ \frac{1}{a}\ D\ \mathcal{Z} - 1\right)$$
$$= bE\left(e^{D\ \mathcal{Z}}\right) - b.$$

Taking into account the fact that $\exp(D\ \mathcal{Z})$ has the log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$ with parameters $\mu = 0$ and $\sigma = D$, we get

$$\lim_{p\to 0} E\ \mathcal{L}_r(p) = b\ \exp\left(\frac{1}{2}\ D^2\right) - b = b\ \exp\left[\log\left(\frac{c}{b} + 1\right)\right] - b = c.$$

□

Appendix

Most approaches to the study of asymptotics of sequential procedures have usually fixed the distributional parameter and allowed some other parameter of the problem (e.g., accuracy or cost) to vary. There exist a very vast literature on sequential estimation under the loss being the sum of the loss associated with the error of estimation and the cost of observation. One constructs optimal procedures in the case when the cost per one observation tends to zero. The notions of optimal asymptotic properties (efficiency, normalities, consistency), when the observation unit cost tends to zero are analogous, but the methodology is different from this one used in this paper and for the first time presented by Hubert and Pyke (1997, 2000). Decreasing of the observation unit cost permits unboundedly increasing of the number n of observations. However, one is often interested what happens when the binomial parameter p is small for fixed confidence or fixed cost. It is then very reasonable to consider the approach to the asymptotic analysis of the problem by letting the parameter p tend to zero.

The optimal sequential procedures, known from the literature, derived under the assumption that the parameter p can take its value from the whole parameter space, are not usually a good proposals when this parameter takes its value close to the boundary. Therefore, finding sequential procedures which have optimal properties when $p \to 0$ is of some interest. From a practical point of view the problem of finding optimal sequential procedures for estimating a function of p when p is small is of great importance for stochastic systems demanding very high reliability.

Examples covered by the problem of estimation considered include some important special cases, in particular, estimation of $1/p$ under a standard LINEX loss function. The loss function of the form (2) is convex and asymmetric. It is useful in the estimation problems when overestimation is considered more serious than underestimation or vice versa (depending on the sign of h_2).

Remark 1 If we use the delta method to construct the stopping time we have to restrict ourselves to the case when $l + s < 0$.

This restriction is a natural consequence of the following discussion [analogous to that of Hubert and Pyke (2000)]. If $l + s = 0$, then from (7) we have

$$n^*(p) = A(s)p^{-1},$$

where the constant $A(s) = A$ is defined by (12), and consequently,

$$N = \inf\{n \geq 1 : S_n \geq A(s)\}. \tag{19}$$

Thus N is the waiting time until $[A(s)]$ successes. Note that in this case, for every $\epsilon > 0$,

$$P\left(\frac{N}{n^*(p)} < 1 - \epsilon\right) = P\left(N < (1-\epsilon)n^*(p)\right) = P\left(S_{[(1-\epsilon)n^*(p)]} \geq [A(s)]\right).$$

For $l + s = 0$, $(1-\epsilon)n^*(p)p = (1-\epsilon)A(s)$ is a constant, so that the distribution of $S_{[(1-\epsilon)n^*(p)]}$ converges to a Poisson distribution as $p \to 0$. Therefore, it cannot be true that $\frac{N}{n^*(p)} \xrightarrow{P} 1$ as $p \to 0$ when $l + s = 0$. Thus the stopping time defined by (19) does not determine a good sequential procedure for estimating p^s.

Remark 2 In this paper, we consider the LINEX loss with weighted error. The model considered can be somewhat expanded by taking

$$h_1(p) = bp^{l_1},$$
$$h_2(p) = ap^{l_2},$$

where l_1 and l_2 are chosen independently.

References

Alvo M (1977) Bayesian sequential estimates. Ann Stat 5(5):955–968

Anderson TW (1960) A modification of the sequential probability ratio test to reduce the sample size. Ann Math Stat 31:165–197

Braess D, Dette H (2004) The asymptotic minimax risk for the estimation of constrained binomial and multinomial probabilities. Sankhyā 66:707–732

Cabilio P (1977) Sequential estimation in Bernoulli trials. Ann Stat 5(2):342–356

Cabilio P, Robbins H (1975) Sequential estimation of p with squared relative error loss. Proc Nat Acad Sci USA 72:191–193

DeGroot MH (1959) Unbiased sequential estimation for binomial populations. Ann Math Stat 30:80–101

Girshick MA, Mosteller F, Savage LJ (1946) Unbiased estimates for certain binomial sampling problems with applications. Ann Math Stat 17:13–23

Haldane JBS (1945) On a method of estimating frequencies. Biometrika 33:222–225

Hubert SL, Pyke R (1997) A particular application of Brownian motion to sequential analysis. Stat Sin 7(1):109–126, empirical Bayes, sequential analysis and related topics in statistics and probability (New Brunswick, NJ, 1995)

Hubert SL, Pyke R (2000) Sequential estimation of functions of p for Bernoulli trials. In: Game theory, optimal stopping, probability and statistics. IMS lecture notes monograph series, vol 35. Inst Math Stat, Beachwood, OH, pp 263–294

Magiera R, Trybuła S (1976) Oblique plans for a binomial process (in Polish). Mat Stos 6(3):41–47

Robbins H, Siegmund D (1976) Sequential estimation of p in Bernoulli trials. In: Studies in probability and statistics (papers in honour of Edwin J. G. Pitman). North-Holland, Amsterdam, pp 103–107

Stein C (1945) A two-sample test for a linear hypothesis whose power is independent of the variance. Ann Math Stat 16:243–258

Täcklind S (1942) Sur le risque de ruine dans des jeux inéquitables. Skand Aktuarietidskr 25:1–42