ALICJA JOKIEL-ROKITA (Wrocław)

THE BAYES SEQUENTIAL ESTIMATION OF
A NORMAL MEAN FROM DELAYED OBSERVATIONS

Abstract. The problem of estimating the mean of a normal distribution is considered in the special case when the data arrive at random times. Certain classes of Bayes sequential estimation procedures are derived under LINEX and reflected normal loss function and with the observation cost determined by a function of the stopping time and the number of observations up to this time.

1. Introduction. The paper deals with the problem of estimating the mean of a normal distribution in the case when the observations become available at random times.

The problem of estimating an unknown parameter of a distribution on the basis of randomly incoming data can appear in many practical situations. For example, in studying the effectiveness of experimental safety devices, relevant data may become available only as a result of accidents. Medical data can sometimes only be obtained when patients seek help or are somehow otherwise identified and examined at random times.

Consider the following model. Let Y_i, $i = 1, \ldots, n$, be independent identically distributed random variables having a normal distribution with unknown mean ϑ and known variance σ^2. It is assumed that Y_i is observed at time t_i, $i = 1, \ldots, n$, where $t_1, \ldots, t_n$ are the order statistics of positive exchangeable i.i.d. random variables $U_1, \ldots, U_n$ which are independent of $Y_1, \ldots, Y_n$.

Let

$$k(t) = \sum_{i=1}^{n} I_{[0,t]}(U_i)$$

---

denote the number of observations made up to time $t \geq 0$, and let $\mathcal{F}_t = \sigma\{k(s),\, s \leq t,\, Y_1, \ldots, Y_{k(t)}\}$ be the information which is available at time t. The problem is to estimate the parameter ϑ.

If observation is stopped at time t, the loss incurred is defined by

$$(1) \qquad L_t(\vartheta, d) = L(\vartheta, d) + c_A k(t) + c(t),$$

where $L(\vartheta, d)$ denotes the loss associated with estimation when ϑ is the true value of the parameter and d is the chosen estimate. The function $c(t)$ represents the cost of observing the process up to time t and it is supposed to be a differentiable and increasing convex function such that $c(0) = 0$; moreover, $c_A \geq 0$ is the cost of taking one observation.

Sequential estimation procedures of the form $\delta = (\tau, d(\tau))$ for estimating the parameter ϑ of the normal distribution will be considered, where τ is a stopping time with respect to $\mathcal{F}_t$, $t \geq 0$, and $d(\tau)$ is an $\mathcal{F}_\tau$-measurable function.

The estimation problem with delayed observations was investigated by Starr, Wardrop and Woodroofe (1976), who considered the case of estimating the mean of normally distributed observations with known variance under squared error loss. Some of their results were extended by Magiera (1982) to one-parameter exponential families of distributions (also under squared error loss).

In practice we are often faced with estimation problems in which over-estimation is considered more serious than underestimation or vice versa. Then using squared error loss (symmetric loss) is not appropriate, and we should use an asymmetric loss function. It has also been argued that unboundedness is undesired in many statistical problems. (For a full discussion of the objections to unbounded loss functions (especially quadratic loss) see Leon and Wu (1992) and Tribus and Szonyi (1989).)

In Section 2 of this paper a class of Bayes sequential procedures for estimating the distribution parameter ϑ of the observations $Y_1, \ldots, Y_n$ with delay will be derived under an asymmetric (LINEX) loss function given by (2) below. The result will be given in the case when the common distribution of $U_1, \ldots, U_n$ is known exactly as well as when it is unknown but required to be exponential. In Section 3 analogous results will be presented for the case of a bounded (reflected normal) loss function given by (11).

2. The Bayes sequential procedures under a LINEX loss function. Let us consider the problem of the Bayes sequential estimation of the parameter ϑ under loss (1) with the LINEX loss function $L(\vartheta, d)$ of the form

$$(2) \qquad L(\vartheta, d) = b\{\exp[a(\vartheta - d)] - a(\vartheta - d) - 1\},$$

where $a \neq 0$, $b > 0$. The LINEX loss function was introduced by Varian (1975). This convex and asymmetric loss function is useful when overestimation is considered more serious than underestimation or vice versa.

Sequential estimation procedures of the form $\delta = (\tau, d(\tau))$ for estimating the parameter ϑ of the normal distribution will be considered, where τ is a stopping time with respect to $\mathcal{F}_t$, $t \geq 0$, and $d(\tau)$ is an $\mathcal{F}_\tau$-measurable function. The risk function associated with a sequential procedure $\delta = (\tau, d(\tau))$ is defined by

$$R(\vartheta, \delta) = E_\vartheta[L_\tau(\vartheta, d(\tau))] = E_\vartheta[L(\vartheta, d(\tau)) + c_A k(\tau) + c(\tau)],$$

where E_ϑ means the expectation with respect to the conditional distribution given ϑ.

Assume that the prior information about the parameter ϑ is that it has the normal distribution (the natural conjugate distribution to the normal distribution in the case when the variance is known) with parameters μ and η^2.

The Bayes risk of the procedure δ connected with the prior distribution π is defined by

$$r(\pi, \delta) = E^\pi[R(\vartheta, \delta)] = \int_{\mathbb{R}} R(\vartheta, \delta) \, \pi(d\vartheta).$$

Define

$$X_t = \sum_{i=1}^{k(t)} Y_i.$$

The following lemma gives the form of the Bayes estimator of the parameter ϑ for any stopping time τ.

LEMMA 1. *For the loss function given by* (2) *and for any stopping time* τ, *the Bayes estimator of* ϑ *with respect to a prior distribution* π *given* $\mathcal{F}_\tau$ *is*

$$(3) \qquad d^*(\tau) = \frac{1}{k(\tau) + \varepsilon}\left(X_\tau + \varepsilon\mu + \frac{1}{2}\,a\sigma^2\right),$$

and the posterior expected loss is

$$E[L(\vartheta, d^*(\tau)) \,|\, \mathcal{F}_\tau] = \frac{1}{2}\,a^2\,\frac{\sigma^2}{k(\tau) + \varepsilon},$$

where $\varepsilon = \sigma^2/\eta^2$.

Proof. The posterior distribution π_t of the parameter ϑ, given $\mathcal{F}_t$, is the normal distribution with parameters μ_t and η_t^2, where

$$(4) \qquad \mu_t = \frac{1}{k(t) + \varepsilon}\,(X_t + \varepsilon\mu),$$

$$(5) \qquad \eta_t^2 = \frac{\sigma^2}{k(t) + \varepsilon}.$$

For a given time t, the form of the Bayes estimator $d^*(t)$ is obtained by using the general formula given in Zellner (1986), namely

$$d^*(t) = \frac{1}{a} \ln\{E^{\pi_t}[\exp(a\vartheta)]\} = \mu_t + \frac{1}{2}\, a\eta_t^2.$$

Straightforward calculations lead to the form of the posterior expected loss. For a random observation time τ the lemma follows from the strong Markov property. ∎

It follows from Lemma 1 that the sequential procedure can be identified with the stopping time. The Bayes estimation problem then reduces to the following optimal stopping problem. The total loss (cost) of observing the process up to time τ is defined to be

$$\mathcal{L}(k(\tau), \tau) = \frac{1}{2}\, a^2\, \frac{\sigma^2}{k(\tau) + \varepsilon} + c_A k(\tau) + c(\tau).$$

The problem is to find a stopping time τ^* which minimizes the expected total loss $E\mathcal{L}$ over all stopping times τ. Such a stopping time is called the *optimal stopping time* and the corresponding sequential procedure $(\tau^*, d^*(\tau^*))$ is called the *Bayes sequential procedure*. In the next subsection such procedures will be derived in the case when the common distribution of the random variables $U_1, \ldots, U_n$ is known exactly. In Subsection 2.2 the Bayes sequential procedures will be given when the distribution of $U_1, \ldots, U_n$ is unknown but required to be exponential.

2.1. *The Bayes sequential procedures: known F.* Assume that the random variables $U_1, \ldots, U_n$ are independent and have a common distribution function F. Suppose that $F(0) = 0$; $F(t) > 0$ for $t > 0$; F is absolutely continuous with density f; and f is the right hand derivative of F on $(0, \infty)$. Denote the class of such F by $\mathcal{G}$.

Let $\zeta = \sup\{t : F(t) < 1\}$, and let $\varrho(t) = f(t)[1 - F(t)]^{-1}$, $0 \le t < \zeta$, denote the failure rate. The process $k(t)$, $0 \le t \le \zeta$, is a nonstationary Markov chain with respect to $\mathcal{F}_t$, $0 \le t \le \zeta$, and its infinitesimal operator is

$$\mathcal{A}_t h(k) = (n - k)\varrho(t)[h(k + 1) - h(k)]$$

for $k \in E_n = \{0, 1, \ldots, n\}$ and all real-valued functions h on E_n (see Starr, Wardrop and Woodroofe (1976)).

Let h be a given real-valued function on E_n such that $0 \le h(k) < \infty$ for each $k \in E_n$, and let

(6) $$\mathcal{L}_h(t) = \mathcal{L}_h(k(t), t) = h(k(t)) + c(t),$$

$t \ge 0$, be the loss incurred if the process is stopped at time t. Suppose that $h(k) - h(k + 1)$ is nonincreasing for $k \le n - 1$ and that $F \in \mathcal{G}$ has a nonincreasing failure rate. Under the assumptions concerning the functions $\varrho(t), h(k)$ and $c(t)$ the so-called *monotone case* holds: once the infinitesimal

prospect of the loss $\mathcal{L}_h$ for the future becomes bad (greater than zero), it remains bad. If the loss $\mathcal{L}_h$ is in the monotone case, then using Dynkin's identity we find that the stopping time

$$\tau_h = \inf\{t \geq 0 : \mathcal{A}_t h(k(t)) + c'(t) \geq 0\}$$
$$= \inf\{t \geq 0 : [n - k(t)]\varrho(t)[h(k(t)) - h(k(t) + 1)] \leq c'(t)\}$$

is optimal.

This method has been used to derive explicitly optimal sequential procedures in other models (see, for example, Ross (1971), Chen and Wardrop (1980), Shapiro and Wardrop (1980)). We will also use this method to prove the following result.

THEOREM 1. *Suppose that $F \in \mathcal{G}$ has nonincreasing failure rate ϱ. Then the Bayes sequential procedure with respect to π is $\delta^* = (\tau^*, d^*(\tau^*))$, where*

$$(7) \quad \tau^* = \inf\left\{t \geq 0 : [n - k(t)]\varrho(t)\left[\frac{a^2\sigma^2}{2[k(t) + 1 + \varepsilon][k(t) + \varepsilon]} - c_A\right] \leq c'(t)\right\}$$

and

$$d^*(\tau^*) = \frac{1}{k(\tau^*) + \varepsilon}\left(X_{\tau^*} + \varepsilon\mu + \frac{1}{2}a\sigma^2\right).$$

Proof. Under the assumptions concerning the functions $\varrho(t)$ and $c(t)$ and taking in (6)

$$h(k(t)) = \frac{1}{2}a^2\frac{\sigma^2}{k(t) + \varepsilon} + c_A k(t),$$

which is associated with the model considered, we infer that the cost $\mathcal{L}_h(t)$ is in the monotone case. Thus, by Dynkin's identity, the stopping time given by (7) is optimal. The form of the Bayes estimator d^* follows from Lemma 1. ∎

2.2. *The Bayes sequential procedures: unknown F.* Let us now consider the problem of sequential estimation of the parameter ϑ in the case when $U_1, \ldots, U_n$ are conditionally independent and exponentially distributed with parameter w, given $W = w$, where W is a random variable having the gamma distribution $\mathcal{G}(\alpha, \beta)$, where $\alpha, \beta > 0$ are known. That is, the random variable W has the density function

$$(8) \qquad\qquad f(w) = \Gamma(\alpha)^{-1}\beta^\alpha w^{\alpha-1}e^{-\beta w}$$

for $w > 0$. The posterior distribution of W given $\mathcal{F}_t$ is $\mathcal{G}(\alpha_t, \beta_t)$ with

$$\alpha_t = \alpha + k(t), \quad \beta_t = \beta + \sum_{j=1}^{k(t)} t_j + [n - k(t)]t.$$

Denote the prior parameters of (8) by α_0 and set β_0, and $m = \alpha_0 + n$. It is easy to check that the process (α_t, β_t), $t \geq 0$, with values in the product

$\{\alpha_0, \alpha_0 + 1, \ldots, \alpha_0 + n\} \times (0, \infty)$, is a stationary Markov process with the infinitesimal operator

$$\mathcal{A}H(\alpha, \beta) = \alpha\beta^{-1}(m - \alpha)[H(\alpha + 1, \beta) - H(\alpha, \beta)] + (m - \alpha)H'(\alpha, \beta),$$

where $'$ denotes differentiation with respect to β (for the proof see Stadje (1990)). The domain of $\mathcal{A}$ includes all H which are continuously differentiable in β for each α.

Let the loss associated with the observation of the process (α_t, β_t), $t \geq 0$, up to time t be of the form

$$\mathcal{L}(t) = \mathcal{L}(\alpha_t, \beta_t, t) = H(\alpha_t, \beta_t) + c(t) = h(\alpha_t) + c(t),$$

where $h(\alpha)$ is a function on $\{\alpha_0, \alpha_0 + 1, \ldots, \alpha_0 + n\}$ such that $0 \leq h(\alpha) < \infty$ for $\alpha \geq \alpha_0$. If the function $\alpha(m - \alpha)[h(\alpha) - h(\alpha + 1)]$ is nonincreasing for $\alpha = \alpha_0, \alpha_0 + 1, \ldots, \alpha_0 + n - 1$, then using the same methods as in Subsection 2.1 one shows that the stopping time

$$(9) \qquad \tau^* = \inf\{t \geq 0 : \mathcal{A}H(\alpha_t, \beta_t) + c'(t) \geq 0\}$$
$$= \inf\{t \geq 0 : \alpha_t\beta_t^{-1}(m - \alpha_t)[h(\alpha_t) - h(\alpha_t + 1)] \leq c'(t)\}$$

is optimal. Notice that

$$\tau^* = \inf\{t \geq 0 : w_t(m - \alpha_t)[h(\alpha_t) - h(\alpha_t + 1)] \leq c'(t)\},$$

where

$$(10) \qquad\qquad\qquad w_t = \alpha_t\beta_t^{-1} = E(W \mid \mathcal{F}_t)$$

is the Bayes estimate of w at time t with respect to a prior distribution given by (8) under squared error loss.

In particular, for

$$h(\alpha_t) = \frac{1}{2}a^2\frac{\sigma^2}{\alpha_t - \alpha_0 + \varepsilon} + c_A(\alpha_t - \alpha_0) = \frac{1}{2}a^2\frac{\sigma^2}{k(t) + \varepsilon} + c_Ak(t),$$

which is associated with the model considered, the following theorem holds.

THEOREM 2. *Let the distribution of $U_1, \ldots, U_n$ be as described above and $\alpha \geq n - 1$. Then the sequential procedure $(\tau^*, d^*(\tau^*))$, where*

$$\tau^* = \inf\left\{t \geq 0 : w_t[n - k(t)]\left[\frac{a^2\sigma^2}{2[k(t) + 1 + \varepsilon][k(t) + \varepsilon]} - c_A\right] \leq c'(t)\right\}$$

and

$$d^*(\tau^*) = \frac{1}{k(\tau^*) + \varepsilon}\left(X_{\tau^*} + \varepsilon\mu + \frac{1}{2}a\sigma^2\right)$$

is Bayes with respect to π. ∎

3. The Bayes sequential procedures under a reflected normal loss function. Let us now consider the problem of the Bayes sequential estimation of the parameter ϑ under loss (1) with the reflected normal loss function $L(\vartheta, d)$ of the form

$$(11) \qquad L(\vartheta, d) = K\left\{1 - \exp\left[-\frac{(d - \vartheta)^2}{2\gamma^2}\right]\right\},$$

where $\gamma > 0$, $K > 0$, and K is the maximum loss parameter. The reflected normal (Spiring's) loss function is symmetric and bounded. It was employed by Spiring (1993) for location parameter estimation. This loss function is a monotone function of the squared error loss.

The following lemma gives the form of the Bayes estimator of the parameter ϑ with respect to a prior normal distribution with parameters μ and η^2.

LEMMA 2. *For the loss function given by* (11) *and for any stopping time τ, the Bayes estimator of ϑ with respect to a prior distribution π given $\mathcal{F}_\tau$ is*

$$d^*(\tau) = \frac{1}{k(\tau) + \varepsilon}\,(X_\tau + \varepsilon\mu),$$

and the posterior expected loss is

$$E[L(\vartheta, d^*(\tau)) \,|\, \mathcal{F}_\tau] = K\left[1 - \gamma\left(\frac{\sigma^2}{k(\tau) + \varepsilon} + \gamma^2\right)^{-1/2}\right].$$

Proof. For a given time t straightforward calculations lead to the form of the Bayes estimator d^* and the form of the posterior expected loss. For a random observation time τ the lemma follows from the strong Markov property. ∎

The following results can be obtained using the same arguments as in Section 2.

THEOREM 3. *Suppose that $F \in \mathcal{G}$ has nonincreasing failure rate ϱ. Then the sequential procedure $\delta^* = (\tau^*, d^*(\tau^*))$, where*

$$\tau^* = \inf\left\{t \geq 0 : \left[\frac{\sigma^2}{k(\tau) + 1 + \varepsilon} + \gamma^2\right]^{-1/2} - \left[\frac{\sigma^2}{k(\tau) + \varepsilon} + \gamma^2\right]^{-1/2} - c_A\right.$$
$$\left. \leq \frac{c'(t)}{K\gamma\varrho(t)[n - k(t)]}\right\}$$

and

$$d^*(\tau^*) = \frac{1}{k(\tau^*) + \varepsilon}\,(X_{\tau^*} + \varepsilon\mu),$$

is Bayes with respect to π. ∎

THEOREM 4. *Let the distribution of $U_1, \ldots, U_n$ be as described in Subsection 2.2 and $\alpha \geq n - 1$. Then the sequential procedure $(\tau^*, d^*(\tau^*))$, where*

$$\tau^* = \inf\left\{ t \geq 0 : \left[\frac{\sigma^2}{k(\tau) + 1 + \varepsilon} + \gamma^2\right]^{-1/2} - \left[\frac{\sigma^2}{k(\tau) + \varepsilon} + \gamma^2\right]^{-1/2} - c_A \right.$$
$$\left. \leq \frac{c'(t)}{K\gamma w_t[n - k(t)]} \right\}$$

with w_t given by (10) and

$$d^*(\tau^*) = \frac{1}{k(\tau^*) + \varepsilon}\left(X_{\tau^*} + \varepsilon\mu\right),$$

is Bayes with respect to π. ∎

References

E. E. Chen and R. L. Wardrop (1980), *Bayes sequential estimation in a life test and asymptotic properties*, Comm. Statist. Theory Methods A 9, 659–672.

R. W. Leon and C. F. G. Wu (1992), *A theory of performance measures in parametric design*, Statist. Sinica 2, 335–357.

R. Magiera (1982), *Estimation with delayed observations*, Zastos. Mat. 17, 249–258.

S. M. Ross (1971), *Infinitesimal look-ahead stopping rules*, Ann. Math. Statist. 42, 297–303.

C. P. Shapiro and R. L. Wardrop (1980), *Dynkin's identity applied to Bayes sequential estimation of a Poisson process rate*, Ann. Statist. 8, 171–182.

F. A. Spring (1993), *The reflected normal loss function*, Canad. J. Statist. 21 (3), 321–330.

W. Stadje (1990), *A sequential estimation procedure for the parameter of an exponential distribution*, Statistics 21, 239–250.

N. Starr, R. Wardrop and M. Woodroofe (1976), *Estimating a mean from delayed observations*, Z. Wahrsch. Verw. Gebiete 35, 103–113.

M. Tribus and G. Szonyi (1989), *An alternate view of the Taguchi approach*, Quality Progress, May, 46–52.

H. L. Varian (1975), *A Bayesian approach to real estate assessment*, in: Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, ed. by S. E. Fienberg and A. Zellner, North-Holland, Amsterdam, 195–208.

A. Zellner (1986), *Bayes estimation and prediction using asymmetric loss function*, J. Amer. Statist. Assoc. 81, 446–451.

Institute of Mathematics and Computer Science
Wrocław University of Technology
50-370 Wrocław, Poland
E-mail: arokita@im.pwr.wroc.pl