Agnieszka Stępień-Baran (Wrocław)

SEQUENTIAL ESTIMATION OF POWERS OF A SCALE PARAMETER FROM DELAYED OBSERVATIONS

Abstract. The problem of sequentially estimating powers of a scale parameter in a scale family and in a location-scale family is considered in the case when the observations become available at random times. Certain classes of sequential estimation procedures are derived under a scale invariant loss function and with the observation cost determined by a convex function of the stopping time and the number of observations up to that time.

1. Introduction. Estimation with delayed observations was investigated by Starr, Wardrop and Woodroofe (1976), who considered the case of Bayes estimation of a mean of normally distributed observations with known variance under a squared error loss and under the assumption that the observation cost involves only cost per unit time. Some of their results were generalized by Magiera (1996). He dealt with estimation of the mean value parameter of the exponential family of distributions under a weighted squared error loss and with the observation cost determined by a convex function of the stopping time. Jokiel-Rokita and Stępień (2007) studied the model with delayed observations for estimating a location parameter under a location invariant loss function and under the assumption that the observation cost is determined by a convex function of the stopping time and the number of observations up to that time.

We consider the following model. Let the sample $(X_1, \ldots, X_n)$ have a joint distribution P_σ with a Lebesgue p.d.f.

$$(1) \qquad \frac{1}{\sigma^n} f\left(\frac{x_1}{\sigma}, \ldots, \frac{x_n}{\sigma}\right),$$

where f is known and $\sigma > 0$ is an unknown scale parameter or a joint

distribution $P_{(\mu,\sigma)}$ with a Lebesgue p.d.f.

(2) $$\frac{1}{\sigma^n}\, f\left(\frac{x_1 - \mu}{\sigma}, \ldots, \frac{x_n - \mu}{\sigma}\right),$$

where f is known and both parameters, i.e., the location parameter $\mu \in \mathbb{R}$ and the scale parameter $\sigma > 0$ are unknown. It is assumed that X_i is observed at time t_i, $i = 1, \ldots, n$, where $t_1, \ldots, t_n$ are the values of the order statistics of positive i.i.d. random variables $U_1, \ldots, U_n$ which are independent of $X_1, \ldots, X_n$.

Let

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

denote the number of observations which have been made by time $t \geq 0$, and let $\mathcal{F}_t = \sigma\{k(s),\, s \leq t,\, X_1, \ldots, X_{k(t)}\}$ be the information which is available at time t.

The problem is to estimate the parameter $\eta = \sigma^r$, $r > 0$. If observation is stopped at time t, the loss incurred is defined by

(4) $$\mathcal{L}_t(\eta, d) := \mathcal{L}(\eta, d) + c_A k(t) + c(t),$$

where $\mathcal{L}(\eta, d)$ denotes the loss associated with estimation when η 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. It is supposed to be a differentiable and increasing convex function such that $c(0) = 0$. The constant $c_A \geq 0$ is the cost of taking one observation.

The family $\{P_\sigma : \sigma > 0\}$ with the p.d.f. given by (1) is called a *one-parameter scale family* and is invariant under the scale transformations $x \mapsto sx$ with $s > 0$. Consequently, the decision problem is invariant under scale transformation if and only if $\mathcal{L}(\eta, a) = \mathcal{L}(s^r\eta, s^r a)$ for all $s > 0$, which is equivalent to

(5) $$\mathcal{L}(\eta, a) = \gamma(a/\eta)$$

for a Borel function $\gamma(\cdot)$ on $[0, \infty)$. An estimator d of the parameter $\eta = \sigma^r$ is *scale equivariant* if and only if

$$d(sX_1, \ldots, sX_n) = s^r d(X_1, \ldots, X_n).$$

The family $\{P_{(\mu,\sigma)} : \mu \in \mathbb{R}, \sigma > 0\}$ with the p.d.f. given by (2) is called a *location-scale family* and is invariant under the location-scale transformations $x \mapsto sx + c$ with $s > 0$ and $c \in \mathbb{R}$. The decision problem is invariant under location-scale transformation provided that a loss function satisfies condition (5). An estimator d of $\eta = \sigma^r$ is *location-scale equivariant* if and only if

$$d(sX_1 + c, \ldots, sX_n + c) = s^r d(X_1, \ldots, X_n).$$

Suppose that we agree to take at least one observation. If we observe the process for $t \geq t_1$ units of time, then the conditional expected loss, given $k(t)$, associated with an equivariant estimator $d(\mathbf{X}_{k(t)})$ based on the random size sample $\mathbf{X}_{k(t)} = (X_1, \ldots, X_{k(t)})$ is of the form

$$(6) \qquad \mathcal{R}_t(\eta, d(\mathbf{X}_{k(t)})) := E_\eta[\mathcal{L}_t(\eta, d(\mathbf{X}_{k(t)})) \mid k(t)] = h(k(t)) + c(t),$$

where E_η means the expectation with respect to the conditional distribution given η. The form of the risk function $\mathcal{R}_t(\eta, d)$, given by (6), follows from the fact that the risk of any equivariant estimator of the parameter η in the invariant problem of estimation is independent of η (see e.g. Shao (2003), Theorem 4.7). The function h depends only on the loss function γ.

In Section 2 we present the method of finding a stopping time which minimizes the expected risk associated with a minimum risk equivariant (MRE) estimator of the parameter $\eta = \sigma^r$ over all stopping times. We consider a situation when the common distribution of the random variables $U_1, \ldots, U_n$ which can be interpreted as the lifetimes of n objects is known exactly. In Section 3 we apply the results of Section 2 to estimate powers of a scale parameter in some subclass of a one-parameter exponential family of distributions, and the normal variance and the scale parameter of a negative exponential distribution (in the case when the location parameter is unknown). We consider the following loss functions associated with estimation: a standardized squared error loss, a standardized LINEX loss function, Stein's loss function and a squared log error loss function. We construct optimal sequential estimation procedures under the aforementioned loss functions in the model with delayed observations.

2. The optimal stopping time. Suppose that in the estimation problem of the parameter $\eta = \sigma^r$ with the loss function $\mathcal{L}(\eta, d)$ there exists an MRE estimator, denoted by d^*. We look for a stopping time τ^* which minimizes the expected risk

$$(7) \qquad\qquad E[\mathcal{R}_\tau(\eta, d^*(\mathbf{X}_{k(\tau)}))] = E[h(k(\tau)) + c(\tau)]$$

over all stopping times $\tau \geq t_1$, $\tau \in \mathcal{T}$, where $\mathcal{T}$ denotes the class of $\mathcal{F}_t$-measurable functions. Such a stopping time will be called an *optimal stopping time*. Then we construct an *optimal sequential estimation procedure* of the form $(\tau^*, d^*(\mathbf{X}_{k(\tau^*)}))$.

Let the random variables $U_1, \ldots, U_n$ be independent and have a common known distribution function G. Suppose that $G(0) = 0$, $G(t) > 0$ for $t > 0$, G is absolutely continuous with density g, and g is the right hand derivative of G on $(0, \infty)$. Denote the class of such G by $\mathcal{G}$. Let $\zeta = \sup\{t : G(t) < 1\}$, and let $\varrho(t) = g(t)[1 - G(t)]^{-1}$, $0 \leq t < \zeta$, denote the failure rate. Under the above assumptions the process $k(t)$, given by (3), is a nonstationary Markov

chain with respect to $\mathcal{F}_t$, $0 \le t \le \zeta$ (see Starr et al. (1976)). The infinitesimal operator $\mathcal{A}_t$ of the process $k(t)$ at $\widetilde{h}$ is defined by

(8) $$\mathcal{A}_t \widetilde{h}(k) := \lim_{s \to 0^+} s^{-1} E[\widetilde{h}(k(t+s)) - \widetilde{h}(k(t)) \,|\, k(t) = k].$$

The domain $D_{\mathcal{A}_t}$ of $\mathcal{A}_t$ is the set of all bounded Borel measurable functions $\widetilde{h}$ on the set $\{0, 1, \ldots, n\}$ for which the limit in (8) exists boundedly pointwise for every $k \in \{0, 1, \ldots, n\}$.

To determine an optimal stopping time we use the following lemma which provides the form of the infinitesimal operator $\mathcal{A}_t$ of the process $k(t)$, given by (3).

LEMMA 1. *Let $\widetilde{h}$ be a given real-valued function on the set $\{0, 1, \ldots, n\}$. The infinitesimal operator $\mathcal{A}_t$ of the process $k(t)$, given by (3), at $\widetilde{h}$ is of the form*

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

Proof. Fix $k \in \{0, 1, \ldots, n\}$. It is clear that

$$E[\widetilde{h}(k(t+s)) - \widetilde{h}(k(t)) \,|\, k(t) = k]$$
$$= \sum_{i=k+1}^{n} [\widetilde{h}(i) - \widetilde{h}(k)]P(k(t+s) = i \,|\, k(t) = k)$$
$$= [\widetilde{h}(k+1) - \widetilde{h}(k)]P(k(t+s) = k+1 \,|\, k(t) = k)$$
$$\quad + \sum_{i=k+2}^{n} [\widetilde{h}(i) - \widetilde{h}(k)]P(k(t+s) = i \,|\, k(t) = k)$$
$$= [\widetilde{h}(k+1) - \widetilde{h}(k)](n-k)\frac{G(t+s) - G(t)}{1 - G(t)}\left[\frac{1 - G(t+s)}{1 - G(t)}\right]^{n-k-1}$$
$$\quad + \sum_{i=k+2}^{n} [\widetilde{h}(i) - \widetilde{h}(k)]P(k(t+s) = i \,|\, k(t) = k)$$
$$\le [\widetilde{h}(k+1) - \widetilde{h}(k)](n-k)\frac{G(t+s) - G(t)}{1 - G(t)}\left[\frac{1 - G(t+s)}{1 - G(t)}\right]^{n-k-1}$$
$$\quad + 2\sup_{i \le n}|\widetilde{h}(i)|P(k(t+s) \ge k+2 \,|\, k(t) = k)$$
$$= [\widetilde{h}(k+1) - \widetilde{h}(k)](n-k)\frac{G(t+s) - G(t)}{1 - G(t)}\left[\frac{1 - G(t+s)}{1 - G(t)}\right]^{n-k-1}$$
$$\quad + 2\sup_{i \le n}|\widetilde{h}(i)|\left\{1 - \left[\frac{1 - G(t+s)}{1 - G(t)}\right]^{n-k}\left[1 - (n-k)\frac{G(t+s) - G(t)}{[1 - G(t+s)]}\right]\right\}.$$

Now it is easy to see that

$$\lim_{s \to 0^+} \frac{E[\widetilde{h}(k(t+s)) - \widetilde{h}(k(t)) \mid k(t) = k]}{s} = (n-k)[\widetilde{h}(k+1) - \widetilde{h}(k)]\varrho(t)$$

and the lemma is proved. ∎

Let $\widetilde{h}(k) = h(k)$ for $k = 1, \ldots, n$, and $\widetilde{h}(0) = 0$. The following theorem determines the optimal stopping time τ^* for a large class of possible h.

THEOREM 1. *Suppose that $G \in \mathcal{G}$ has nonincreasing failure rate ϱ, and the function $h(k)$ in formula (6) is such that $h(k) - h(k+1)$ is nonincreasing for $k \in \{1, \ldots, n-1\}$. Then the stopping time*

$$(9) \qquad \tau^* = \inf\{t \geq t_1 : \mathcal{A}_t \widetilde{h}(k(t)) + c'(t) \geq 0\}$$
$$= \inf\{t \geq t_1 : (n - k(t))[h(k(t)) - h(k(t) + 1)]\varrho(t) \leq c'(t)\}$$

minimizes the expected risk given by (7) over all stopping times $\tau \geq t_1$, $\tau \in \mathcal{T}$.

Proof. See Jokiel-Rokita & Stępień (2007), proof of Theorem 1. ∎

3. Examples. In this section we use the solutions of Section 2 to estimate powers of a scale parameter in some subclass of a one-parameter exponential family of distributions, and the normal variance and the scale parameter of a negative exponential distribution (in the case when the location parameter is unknown), under the following loss functions associated with estimation: the standardized squared error loss

$$(10) \qquad \mathcal{L}(\eta, d) = \left(\frac{d}{\eta} - 1\right)^2,$$

the standardized LINEX loss function

$$(11) \qquad \mathcal{L}(\eta, d) = \exp\left\{a\left(\frac{d}{\eta} - 1\right)\right\} - a\left(\frac{d}{\eta} - 1\right) - 1 \quad \text{for } a \neq 0,$$

Stein's loss function (see James and Stein (1961))

$$(12) \qquad \mathcal{L}(\eta, d) = \frac{d}{\eta} - \ln\frac{d}{\eta} - 1,$$

and the squared log error loss function (see Brown (1968))

$$(13) \qquad \mathcal{L}(\eta, d) = (\ln d - \ln \eta)^2 = \left[\ln\frac{d}{\eta}\right]^2.$$

Taking MRE estimators as optimal estimators of powers of a scale parameter, we construct optimal sequential estimation procedures under the aforementioned loss functions in the model with observations which are available at random times.

3.1. *A subclass of a one-parameter exponential family of distributions with a scale parameter.* In many cases the p.d.f. given by (1) can be written in the following way:

$$(14) \qquad f_\eta(x_1,\ldots,x_n) = w(x_1,\ldots,x_n;n)\eta^{-\nu}\exp\{-T(x_1,\ldots,x_n)/\eta\},$$

where $w(x_1,\ldots,x_n;n)$ is a nonnegative function, $\eta = \sigma^r$ for some $r > 0$, $\nu = \lambda n$, $\lambda > 0$; for a known parameter λ, $T(X_1,\ldots,X_n)$ is a complete sufficient statistic for η; $T \sim \mathcal{G}(\nu,\eta)$. Examples of such models are: gamma distribution $\mathcal{G}(\mu,\sigma)$ with a known parameter μ and $\eta = \sigma$; Laplace distribution $\mathcal{L}a(0,\sigma)$ with $\eta = \sigma$; normal distribution $\mathcal{N}(0,\sigma^2)$ with $\eta = \sigma^2$; Maxwell distribution $\mathcal{M}a(\sigma)$ with $\eta = \sigma^2$.

Let $T(\mathbf{X}_{k(t)})$ be the statistic based on the random size sample $\mathbf{X}_{k(t)} = (X_1,\ldots,X_{k(t)})$, where $k(t)$, given by (3), denotes the number of observations made by time t. The following theorem provides the MRE estimators of the parameter $\eta = \sigma^r$ of distributions belonging to the subclass of a one-parameter exponential family of distributions defined by (14), and the corresponding risk functions under the loss functions given by (10), (11), (12) or (13).

THEOREM 2. *For any stopping time t,*

(a) *if the loss function is given by (10), then the MRE estimator of the parameter η is*

$$d_S^*(\mathbf{X}_{k(t)}) = \frac{T(\mathbf{X}_{k(t)})}{\lambda k(t) + 1},$$

and the risk function of the estimator d_S^* has the form

$$\mathcal{R}_t(\eta, d_S^*) = \frac{1}{\lambda k(t) + 1} + c_A k(t) + c(t);$$

(b) *if the loss function is given by (11), then the MRE estimator of η is*

$$d_L^*(\mathbf{X}_{k(t)}) = \frac{1 - \exp\left\{-\frac{a}{\lambda k(t)+1}\right\}}{a} T(\mathbf{X}_{k(t)}),$$

and the risk function of d_L^* has the form

$$\mathcal{R}_t(\eta, d_L^*) = (\lambda k(t) + 1)\left[\exp\left\{-\frac{a}{\lambda k(t) + 1}\right\} - 1\right] + a + c_A k(t) + c(t);$$

(c) *if the loss function is given by (12), then the MRE estimator of η is*

$$d_{ST}^*(\mathbf{X}_{k(t)}) = \frac{T(\mathbf{X}_{k(t)})}{\lambda k(t)},$$

and the risk function of d_{ST}^* has the form

$$\mathcal{R}_t(\eta, d_{ST}^*) = \ln \lambda k(t) - \psi(\lambda k(t)) + c_A k(t) + c(t),$$

where ψ denotes the digamma function;

(d) *if the loss function is given by (13), then the MRE estimator of η is*

$$d_{SL}^*(\mathbf{X}_{k(t)}) = \frac{T(\mathbf{X}_{k(t)})}{\exp\{\psi(\lambda k(t))\}},$$

and the risk function of d_{SL}^* has the form

$$\mathcal{R}_t(\eta, d_{SL}^*) = \psi'(\lambda k(t)) + c_A k(t) + c(t),$$

where ψ' denotes the derivative of the digamma function.

Proof. In this proof we use solutions for the classical model, i.e., for a fixed sample size. Let k be the fixed size of the sample $\mathbf{X}_k = (X_1, \ldots, X_k)$. The maximum likelihood estimator of the parameter η, $d_0(\mathbf{X}_k) = T(\mathbf{X}_k)/\lambda k$, is an equivariant estimator of η. The statistic $T(\mathbf{X}_k)$ is complete and sufficient for η (see e.g. Shao (2003), Proposition 2.1). The statistic $\mathbf{Z}_k = (Z_1, \ldots, Z_k)$, where $Z_i = X_i/X_k$ for $i = 1, \ldots, k-1$ and $Z_k = X_k/|X_k|$, is ancillary, i.e., its distribution does not depend on η. Hence, by Basu's theorem (see e.g. Shao (2003), Theorem 2.4) the statistics $T(\mathbf{X}_k)$ and $\mathbf{Z}_k$ are independent. Moreover, we have $T \sim \mathcal{G}(\lambda k, \eta)$, $\lambda > 0$.

According to e.g. Theorem 4.8 of Shao (2003), since $T(\mathbf{X}_k)$ and $\mathbf{Z}_k$ are independent, an MRE estimator of the parameter η is of the form

$$d^*(\mathbf{X}_k) = \frac{d_0(\mathbf{X}_k)}{w^*}$$

with a constant w^* which minimizes

$$E_1\left\{\gamma\left(\frac{d_0(\mathbf{X}_k)}{w}\right)\right\}$$

over w, where the expectation E_1 is calculated under the assumption that $\eta = 1$.

In this proof we also use the form of the moment generating function of a random variable which has gamma distribution with parameters $\nu, \eta > 0$, i.e., $M(t) = (1/(1-\eta t))^\nu$ for $t < 1/\eta$, and the following facts. If $X \sim \mathcal{G}(\nu, \eta)$, then $E[\ln X] = \psi(\nu) + \ln\eta$ and $E[\ln X]^2 = \psi'(\nu) + [\psi(\nu) + \ln\eta]^2$.

In the case (a), i.e., for the standardized squared error loss $\mathcal{L}(\eta, d)$ given by (10), on the basis of e.g. Corollary 4.1 of Shao (2003) the MRE estimator of the parameter η has the form

$$d_S^*(\mathbf{X}_k) = \frac{d_0(\mathbf{X}_k) E_1[d_0(\mathbf{X}_k)]}{E_1[d_0(\mathbf{X}_k)]^2} = \frac{\lambda k d_0(\mathbf{X}_k)}{\lambda k + 1} = \frac{T(\mathbf{X}_k)}{\lambda k + 1}.$$

Thus the risk associated with estimation is equal to

$$\mathcal{R}(\eta, d_S^*) = E_\eta[\mathcal{L}(\eta, d_S^*)] = \frac{E_\eta[T(\mathbf{X}_k)]^2}{(\lambda k + 1)^2 \eta^2} - \frac{2E_\eta[T(\mathbf{X}_k)]}{(\lambda k + 1)\eta} + 1 = \frac{1}{\lambda k + 1}.$$

Hence, we conclude that in this model with delayed observations the risk associated only with estimation errors reduces to

$$\mathcal{R}(\eta, d_S^*(\mathbf{X}_{k(t)})) := E_\eta[\mathcal{L}(\eta, d_S^*) \,|\, k(t)] = \frac{1}{\lambda k(t) + 1}.$$

Therefore the total risk, i.e., the total expected loss which consists of the risk function of the MRE estimator and the observation cost up to time t is of the form

$$\mathcal{R}_t(\eta, d_S^*) = E_\eta[\mathcal{L}_t(\eta, d_S^*) \,|\, k(t)] = \frac{1}{\lambda k(t) + 1} + c_A k(t) + c(t).$$

In the case (b), i.e., under the standardized LINEX loss function $\mathcal{L}(\eta, d)$ given by (11), to determine the MRE estimator of η we calculate the number w^* minimizing the expression

$$E_1\left[\exp\left\{a\left(\frac{T(\mathbf{X}_k)}{\lambda k w} - 1\right)\right\} - a\left(\frac{T(\mathbf{X}_k)}{\lambda k w} - 1\right) - 1\right]$$

$$= \exp\{-a\} M_{\eta=1}\left(\frac{a}{\lambda k w}\right) - \frac{a E_1[T(\mathbf{X}_k)]}{\lambda k w} + a - 1$$

$$= \frac{\exp\{-a\}}{\left(1 - \frac{a}{\lambda k w}\right)^{\lambda k}} - \frac{a}{w} + a - 1.$$

We obtain

$$w^* = \frac{a}{\lambda k\left[1 - \exp\left\{-\frac{a}{\lambda k + 1}\right\}\right]}.$$

Then the MRE estimator of the parameter η has the form

$$d_L^*(\mathbf{X}_k) = \frac{d_0(\mathbf{X}_k)}{\frac{a}{\lambda k[1 - \exp\{-\frac{a}{\lambda k + 1}\}]}} = \frac{1 - \exp\{-\frac{a}{\lambda k + 1}\}}{a} T(\mathbf{X}_k).$$

Thus the risk associated with estimation is equal to

$$\mathcal{R}(\eta, d_L^*) = E_\eta[\mathcal{L}(\eta, d_L^*)]$$

$$= \exp\{-a\} E_\eta\left[\exp\left\{\frac{1 - \exp\{-\frac{a}{\lambda k + 1}\}}{\eta} T(\mathbf{X}_k)\right\}\right]$$

$$- \frac{1 - \exp\{-\frac{a}{\lambda k + 1}\}}{\eta} E_\eta[T(\mathbf{X}_k)] + a - 1.$$

Using the form of the moment generating function of the statistic T (which

has gamma distribution), we obtain

$$\mathcal{R}(\eta, d_L^*) = \exp\{-a\} M\left(\frac{1 - \exp\{-\frac{a}{\lambda k+1}\}}{\eta}\right) - \lambda k\left[1 - \exp\left\{-\frac{a}{\lambda k + 1}\right\}\right]$$
$$+ a - 1$$
$$= (\lambda k + 1)\left[\exp\left\{-\frac{a}{\lambda k + 1}\right\} - 1\right] + a.$$

Hence, in this model with delayed observations the risk associated only with estimation errors reduces to

$$\mathcal{R}(\eta, d_L^*(\mathbf{X}_{k(t)})) := E_\eta[\mathcal{L}(\eta, d_L^*) \,|\, k(t)]$$
$$= (\lambda k(t) + 1)\left[\exp\left\{-\frac{a}{\lambda k(t) + 1}\right\} - 1\right] + a.$$

Therefore the total risk is

$$\mathcal{R}_t(\eta, d_L^*) = E_\eta[\mathcal{L}_t(\eta, d_L^*) \,|\, k(t)]$$
$$= (\lambda k(t) + 1)\left[\exp\left\{-\frac{a}{\lambda k(t) + 1}\right\} - 1\right] + a + c_A k(t) + c(t).$$

In the case (c), i.e., for Stein's loss function $\mathcal{L}(\eta, d)$ given by (12), to determine the MRE estimator of η we calculate the number w^* which minimizes the expression

$$E_1\left[\frac{T(\mathbf{X}_k)}{\lambda k w} - \ln\frac{T(\mathbf{X}_k)}{\lambda k w} - 1\right] = \frac{E_1[T(\mathbf{X}_k)]}{\lambda k w} + \ln \lambda k w - E_1[\ln T(\mathbf{X}_k)] - 1$$
$$= \frac{1}{w} + \ln w + \ln \lambda k - \psi(\lambda k) - 1.$$

We obtain $w^* = 1$. Then the MRE estimator of the parameter η is $d_{ST}^* = d_0$. Thus the risk associated with estimation is equal to

$$\mathcal{R}(\eta, d_{ST}^*) = E_\eta[\mathcal{L}(\eta, d_{ST}^*)] = \frac{E_\eta[T(\mathbf{X}_k)]}{\lambda k \eta} - E_\eta[\ln T(\mathbf{X}_k)] + \ln \lambda k \eta - 1$$
$$= \ln \lambda k - \psi(\lambda k).$$

Hence, in this model with delayed observations the risk associated only with estimation errors reduces to

$$\mathcal{R}(\eta, d_{ST}^*(\mathbf{X}_{k(t)})) := E_\eta[\mathcal{L}(\eta, d_{ST}^*) \,|\, k(t)]$$
$$= \ln \lambda k(t) - \psi(\lambda k(t)).$$

Therefore the total risk is

$$\mathcal{R}_t(\eta, d_{ST}^*) = E_\eta[\mathcal{L}_t(\eta, d_{ST}^*) \,|\, k(t)]$$
$$= \ln \lambda k(t) - \psi(\lambda k(t)) + c_A k(t) + c(t).$$

In the case (d), i.e., under the squared log error loss function $\mathcal{L}(\eta, d)$ given by (13), to determine the MRE estimator of η we calculate w^* minimizing

$$E_1\left[\ln^2 \frac{T(\mathbf{X}_k)}{\lambda k w}\right] = E_1[\ln^2 T(\mathbf{X}_k)] - 2(\ln \lambda k w)E_1[\ln T(\mathbf{X}_k)] + \ln^2 \lambda k w$$

$$= \psi'(\lambda k) + \psi^2(\lambda k) - 2(\ln \lambda k w)\psi(\lambda k) + \ln^2 \lambda k w.$$

We obtain

$$w^* = \frac{\exp\{\psi(\lambda k)\}}{\lambda k}.$$

Then the MRE estimator of η is

$$d_{SL}^*(\mathbf{X}_k) = \frac{d_0(\mathbf{X}_k)}{\frac{\exp\{\psi(\lambda k)\}}{\lambda k}} = \frac{T(\mathbf{X}_k)}{\exp\{\psi(\lambda k)\}}.$$

Thus the risk associated with estimation is equal to

$$\begin{aligned}
\mathcal{R}(\eta, d_{SL}^*) &= E_\eta[\mathcal{L}(\eta, d_{SL}^*)] \\
&= E_\eta[\ln^2 T(\mathbf{X}_k)] - 2[\psi(\lambda k) + \ln \eta]E_\eta[\ln T(\mathbf{X}_k)] + [\psi(\lambda k) + \ln \eta]^2 \\
&= \psi'(\lambda k).
\end{aligned}$$

Hence, in this model with delayed observations the risk associated only with estimation errors reduces to

$$\mathcal{R}(\eta, d_{SL}^*(\mathbf{X}_{k(t)})) := E_\eta[\mathcal{L}(\eta, d_{SL}^*) \,|\, k(t)] = \psi'(\lambda k(t)).$$

Therefore the total risk has the form

$$\mathcal{R}_t(\eta, d_{SL}^*) = E_\eta[\mathcal{L}_t(\eta, d_{SL}^*) \,|\, k(t)] = \psi'(\lambda k(t)) + c_A k(t) + c(t). \quad \blacksquare$$

On the basis of Theorems 1 and 2 we construct optimal sequential estimation procedures of the form $(\tau^*, d^*(\mathbf{X}_{k(\tau^*)}))$, where τ^* is defined by (9), and d^* is the corresponding sequential MRE estimator of $\eta = \sigma^r$ based on the random size sample $\mathbf{X}_{k(\tau^*)}$. The function $h(k)$ for the specified loss functions takes the respective forms

$$h_S(k) = \frac{1}{\lambda k + 1} + c_A k,$$

$$h_L(k) = (\lambda k + 1)\left[\exp\left\{-\frac{a}{\lambda k + 1}\right\} - 1\right] + a + c_A k,$$

$$h_{ST}(k) = \ln \lambda k - \psi(\lambda k) + c_A k,$$

$$h_{SL}(k) = \psi'(\lambda k) + c_A k.$$

The following theorem determines the optimal sequential estimation procedures under the loss function $\mathcal{L}(\eta, d)$ given by (10), (11), (12) or (13).

THEOREM 3. *Suppose that $G \in \mathcal{G}$ has nonincreasing failure rate ϱ.*

(a) *Under the loss function $\mathcal{L}_t(\eta, d)$ given by (4) with $\mathcal{L}(\eta, d)$ of the form (10), the sequential estimation procedure $(\tau_S^*, d_S^*(\mathbf{X}_{k(\tau_S^*)}))$, where*

$$\tau_S^* = \inf\left\{t \geq t_1 : [n-k(t)]\right.$$

$$\left.\times\left[\frac{\lambda}{(\lambda k(t)+1)(\lambda k(t)+\lambda+1)} - c_A\right]\varrho(t) \leq c'(t)\right\}$$

and

$$d_S^*(\mathbf{X}_{k(\tau_S^*)}) = \frac{T(\mathbf{X}_{k(\tau_S^*)})}{\lambda k(\tau_S^*)+1},$$

is optimal.

(b) *Under the loss function* $\mathcal{L}_t(\eta, d)$ *given by* (4) *with* $\mathcal{L}(\eta, d)$ *of the form* (11), *the sequential estimation procedure* $(\tau_L^*, d_L^*(\mathbf{X}_{k(\tau_L^*)}))$, *where*

$$\tau_L^* = \inf\left\{t \geq t_1 : [n-k(t)]\varrho(t)\left[(\lambda k(t)+1)\exp\left\{-\frac{a}{\lambda k(t)+1}\right\}\right.\right.$$

$$\left.\left. - (\lambda k(t)+\lambda+1)\exp\left\{-\frac{a}{\lambda k(t)+\lambda+1}\right\} + \lambda - c_A\right] \leq c'(t)\right\}$$

and

$$d_L^*(\mathbf{X}_{k(\tau_L^*)}) = \frac{1 - \exp\left\{-\frac{a}{\lambda k(\tau_L^*)+1}\right\}}{a} T(\mathbf{X}_{k(\tau_L^*)}),$$

is optimal.

(c) *Under the loss function* $\mathcal{L}_t(\eta, d)$ *given by* (4) *with* $\mathcal{L}(\eta, d)$ *of the form* (12), *the sequential estimation procedure* $(\tau_{ST}^*, d_{ST}^*(\mathbf{X}_{k(\tau_{ST}^*)}))$, *where*

$$\tau_{ST}^* = \inf\left\{t \geq t_1 : \left[\ln\frac{k(t)}{k(t)+1} + \psi(\lambda(k(t)+1)) - \psi(\lambda k(t)) - c_A\right]\right.$$

$$\left.\times [n-k(t)]\varrho(t) \leq c'(t)\right\}$$

and

$$d_{ST}^*(\mathbf{X}_{k(\tau_{ST}^*)}) = \frac{T(\mathbf{X}_{k(\tau_{ST}^*)})}{\lambda k(\tau_{ST}^*)},$$

is optimal.

(d) *Under the loss function* $\mathcal{L}_t(\eta, d)$ *given by* (4) *with* $\mathcal{L}(\eta, d)$ *of the form* (13), *the sequential estimation procedure* $(\tau_{SL}^*, d_{SL}^*(\mathbf{X}_{k(\tau_{SL}^*)}))$, *where*

$$\tau_{SL}^* = \inf\{t \geq t_1 : [n-k(t)][\psi'(\lambda k(t)) - \psi'(\lambda(k(t)+1)) - c_A]\varrho(t) \leq c'(t)\}$$

and

$$d_{SL}^*(\mathbf{X}_{k(\tau_{SL}^*)}) = \frac{T(\mathbf{X}_{k(\tau_{SL}^*)})}{\exp\{\psi(\lambda k(\tau_{SL}^*))\}},$$

is optimal.

Proof. We have to show that the assumption of Theorem 1 is satisfied, i.e., the function $v(k) := h(k) - h(k+1)$ is nonincreasing on the set $\{1, \ldots, n-1\}$. Hence, we need to verify the condition $v(k+1) - v(k) = 2h(k+1) - h(k) - h(k+2) \leq 0$, which is equivalent to $h(k+1) \leq (h(k) + h(k+2))/2$. This can be reduced to the verification that h is convex on the interval $[1, n-1]$.

(a) In the model with the standardized squared error loss we obtain

$$h_S''(k) = \frac{2\lambda^2}{(\lambda k + 1)^3} > 0 \quad \text{for } k > 0.$$

Hence, h_S is strictly convex on $(0, \infty)$.

(b) Under the standardized LINEX loss function, we again have

$$h_L''(k) = \exp\left\{-\frac{a}{\lambda k + 1}\right\} \frac{a^2 \lambda^2}{(\lambda k + 1)^3} > 0 \quad \text{for } k > 0.$$

(c) For Stein's loss function, the function h_{ST} can be written as

$$h_{ST}(k) = r(\lambda k) + \frac{1}{2\lambda k} + c_A k,$$

where $r(\lambda k) = \ln \lambda k - \frac{1}{2\lambda k} - \psi(\lambda k)$. Hence

$$h_{ST}''(k) = \lambda^2 r''(\lambda k) + \frac{1}{\lambda k^3} > 0 \quad \text{for } k > 0,$$

because $r''(x) > 0$ for any $x \in (0, \infty)$ by Theorem 1 of Qi et al. (2005).

(d) In the model with the squared log error loss function we have

$$h_{SL}''(k) = \lambda^2 \psi'''(\lambda k) > 0 \quad \text{for } k > 0,$$

since

$$\psi^m(x) = (-1)^{m+1} m! \sum_{i=0}^{\infty} \frac{1}{(x+i)^{m+1}} \text{ for } x \in \mathbb{R}_+ \text{ and } m \in \mathbb{N},$$

and in particular

$$\psi'''(\lambda k) = 6 \sum_{i=0}^{\infty} \frac{1}{(\lambda k + i)^4} > 0. \ \blacksquare$$

3.2. *The normal and the negative exponential distribution case.* Consider first the classical model. Let k be the fixed size of the sample $\mathbf{X}_k = (X_1, \ldots, X_k)$. Examples of location-scale distributions with the p.d.f. given by (2) are

(a) normal distribution $\mathcal{N}(\mu, \sigma^2)$; as an equivariant estimator of the parameter σ^2 we take

$$d_0(\mathbf{X}_k) = \frac{1}{k} V(\mathbf{X}_k) \quad \text{with} \quad V(\mathbf{X}_k) = \sum_{i=1}^{k} (X_i - \bar{X}_k)^2, V \sim \mathcal{G}\left(\frac{k-1}{2}, 2\sigma^2\right);$$

(b) negative exponential distribution $\mathcal{NE}(\mu, \sigma)$; as an equivariant estimator $d_0(\mathbf{X}_k)$ of the parameter σ we take

$$d_0(\mathbf{X}_k) = \frac{1}{k} V(\mathbf{X}_k) \quad \text{with} \quad V(\mathbf{X}_k) = \sum_{i=1}^{k}(X_i - X_{1:k}), V \sim \mathcal{G}(k-1, \sigma).$$

Notice that for both aforementioned distributions the statistic V has gamma distribution $\mathcal{G}(\lambda(n-1), r\sigma^r)$, where $r = 2$ and $\lambda = 1/2$ for the normal case, and $\lambda = r = 1$ for the negative exponential case.

Consider next the model with delayed observations. Let $V(\mathbf{X}_{k(t)})$ be the statistic based on the random size sample $\mathbf{X}_{k(t)} = (X_1, \ldots, X_{k(t)})$, where $k(t)$, given by (3), denotes the number of observations which have been made by time t. The following theorem provides the MRE estimators of the parameter $\eta = \sigma^r$ for the normal case ($r = 2$) and for the negative exponential case ($r = 1$), and the corresponding risk functions under the loss functions given by (10), (11), (12) or (13).

THEOREM 4. *For any stopping time* t,

(a) *if the loss function is given by* (10), *then the MRE estimator of the parameter* η *is*

$$d_S^*(\mathbf{X}_{k(t)}) = \frac{V(\mathbf{X}_{k(t)})}{r[\lambda(k(t)-1)+1]},$$

and the risk function of the estimator d_S^* *has the form*

$$\mathcal{R}_t(\eta, d_S^*) = \frac{1}{\lambda(k(t)-1)+1} + c_A k(t) + c(t);$$

(b) *if the loss function is given by* (11), *then the MRE estimator of* η *is*

$$d_L^*(\mathbf{X}_{k(t)}) = \frac{1 - \exp\left\{-\frac{a}{\lambda(k(t)-1)+1}\right\}}{ar} V(\mathbf{X}_{k(t)}),$$

and the risk function of d_L^* *has the form*

$$\mathcal{R}_t(\eta, d_L^*) = [\lambda(k(t)-1)+1]\left[\exp\left\{-\frac{a}{\lambda(k(t)-1)+1}\right\} - 1\right]$$
$$+ a + c_A k(t) + c(t);$$

(c) *if the loss function is given by* (12), *then the MRE estimator of* η *is*

$$d_{ST}^*(\mathbf{X}_{k(t)}) = \frac{V(\mathbf{X}_{k(t)})}{r\lambda(k(t)-1)},$$

and the risk function of d_{ST}^* *has the form*

$$\mathcal{R}_t(\eta, d_{ST}^*) = \ln \lambda(k(t)-1) - \psi(\lambda(k(t)-1)) + c_A k(t) + c(t),$$

where ψ *denotes the digamma function;*

(d) *if the loss function is given by* (13), *then the MRE estimator of η is*

$$d_{SL}^*(\mathbf{X}_{k(t)}) = \frac{V(\mathbf{X}_{k(t)})}{r \exp\{\psi(\lambda(k(t)-1))\}},$$

and the risk function of d_{SL}^* has the form

$$\mathcal{R}_t(\eta, d_{SL}^*) = \psi'(\lambda(k(t)-1)) + c_A k(t) + c(t),$$

where ψ' denotes the derivative of the digamma function.

Proof. The forms of the MRE estimators d_S^*, d_L^*, d_{ST}^* and d_{SL}^* were obtained from the general formula for MRE estimators of powers of a scale parameter in a location-scale family under loss function (5) (see e.g. Shao (2003), Section 4.2.3, General location-scale families). The formulas for the risk functions $\mathcal{R}_t(\eta, d_S^*)$, $\mathcal{R}_t(\eta, d_L^*)$, $\mathcal{R}_t(\eta, d_{ST}^*)$ and $\mathcal{R}_t(\eta, d_{SL}^*)$ follow from calculations which use the same technique as in the proof of Theorem 2, because the statistic V (just as T) has gamma distribution. ∎

The next theorem determines the optimal sequential estimation procedures under the loss function $\mathcal{L}(\eta, d)$ given by (10), (11), (12) and (13), respectively.

THEOREM 5. *Suppose that $G \in \mathcal{G}$ has nonincreasing failure rate ϱ.*

(a) *Under the loss function $\mathcal{L}_t(\eta, d)$ given by* (4) *with $\mathcal{L}(\eta, d)$ of the form* (10), *the sequential estimation procedure $(\tau_S^*, d_S^*(\mathbf{X}_{k(\tau_S^*)}))$, where*

$$\tau_S^* = \inf \left\{ t \geq t_1 : [n - k(t)] \right.$$
$$\left. \times \left[\frac{\lambda}{(\lambda k(t) + 1 - \lambda)(\lambda k(t) + 1)} - c_A \right] \varrho(t) \leq c'(t) \right\}$$

and

$$d_S^*(\mathbf{X}_{k(\tau_S^*)}) = \frac{V(\mathbf{X}_{k(\tau_S^*)})}{r[\lambda(k(\tau_S^*)-1)+1]},$$

is optimal.

(b) *Under the loss function $\mathcal{L}_t(\eta, d)$ given by* (4) *with $\mathcal{L}(\eta, d)$ of the form* (11), *the sequential estimation procedure $(\tau_L^*, d_L^*(\mathbf{X}_{k(\tau_L^*)}))$, where*

$$\tau_L^* = \inf \left\{ t \geq t_1 : [n-k(t)]\varrho(t) \left[(\lambda k(t)+1-\lambda) \exp\left\{ -\frac{a}{\lambda k(t) + 1 - \lambda} \right\} \right. \right.$$
$$\left. \left. - (\lambda k(t) + 1) \exp\left\{ -\frac{a}{\lambda k(t) + 1} \right\} + \lambda - c_A \right] \leq c'(t) \right\}$$

and

$$d_L^*(\mathbf{X}_{k(\tau_L^*)}) = \frac{1 - \exp\left\{-\frac{a}{\lambda(k(\tau_L^*)-1)+1}\right\}}{ar} V(\mathbf{X}_{k(\tau_L^*)}),$$

is optimal.

(c) *Under the loss function* $\mathcal{L}_t(\eta, d)$ *given by* (4) *with* $\mathcal{L}(\eta, d)$ *of the form* (12), *the sequential estimation procedure* $(\tau_{ST}^*, d_{ST}^*(\mathbf{X}_{k(\tau_{ST}^*)}))$, *where*

$$\tau_{ST}^* = \inf\left\{ t \ge t_2 : \left[\ln \frac{k(t)-1}{k(t)} + \psi(\lambda k(t)) - \psi(\lambda(k(t)-1)) - c_A\right] \right.$$
$$\left. \times [n - k(t)]\varrho(t) \le c'(t) \right\}$$

and

$$d_{ST}^*(\mathbf{X}_{k(\tau_{ST}^*)}) = \frac{V(\mathbf{X}_{k(\tau_{ST}^*)})}{r\lambda(k(\tau_{ST}^*)-1)},$$

is optimal.

(d) *Under the loss function* $\mathcal{L}_t(\eta, d)$ *given by* (4) *with* $\mathcal{L}(\eta, d)$ *of the form* (13), *the sequential estimation procedure* $(\tau_{SL}^*, d_{SL}^*(\mathbf{X}_{k(\tau_{SL}^*)}))$, *where*

$$\tau_{SL}^* = \inf\{t \ge t_1 : [n-k(t)][\psi'(\lambda(k(t)-1)) - \psi'(\lambda k(t)) - c_A]\varrho(t) \le c'(t)\}$$

and

$$d_{SL}^*(\mathbf{X}_{k(\tau_{SL}^*)}) = \frac{V(\mathbf{X}_{k(\tau_{SL}^*)})}{r \exp\{\psi(\lambda(k(\tau_{SL}^*)-1))\}},$$

is optimal.

Proof. The proof is analogous to the proof of Theorem 3. ∎

References

L. D. Brown (1968), *Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters*, Ann. Math. Statist. 39, 29–48.

W. James and C. Stein (1961), *Estimation with quadratic loss*, in: Proc. 4th Berkeley Sympos. Math. Statist. and Probab., Vol. I, Univ. California Press, Berkeley, CA, 361–379.

A. Jokiel-Rokita and A. Stępień, *Sequential estimation of a location parameter from delayed observations*, Statistical Papers, doi: 10.1007/s00362-007-0059-2 (2007).

R. Magiera (1996), *On a class of sequential estimation problems for one-parameter exponential families*, Sankhyā Ser. A 58, 160–170.

F. Qi, R.-Q. Cui, C.-P. Chen, and B.-N. Guo (2005), *Some completely monotonic functions involving polygamma functions and an application*, J. Math. Anal. Appl. 310, 303–308.

J. Shao (2003), *Mathematical Statistics*, Springer Texts in Statistics, 2nd ed., Springer, New York.

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

Institute of Mathematics and Computer Science
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: stepienbaran@interia.eu