SOME NEW RESULTS IN SEQUENTIAL ESTIMATION THEORY

YU. V. LINNIK and I. V. ROMANOVSKY
MATHEMATICAL INSTITUTE OF THE ACADEMY OF SCIENCES, LENINGRAD

1. The setup of the problem; some general remarks

The problem of sequential estimation attracted some fresh interest recently. We can note, basically, two directions in the corresponding recent literature: asymptotic investigations (see for instance [1], [2], [3], [4]) and exact formulas for "small samples" (see [5], [6], [7], [8]). We shall give here an account of some recent results in both these directions.

In articles [1] to [4] a Bayesian approach to the sequential estimation of parameters is considered. Here we use another, non-Bayesian approach.

We can consider sufficiently large families of stochastic processes with independent increments (and discrete or continuous time). We shall always study scalar processes unless stipulated otherwise.

In the case of discrete time, we shall suppose it integer valued so that our process will be reduced to the repeated sample of a certain population with a distribution in a family $\mathscr{P}_\theta$ characterized by a density $f(x, \theta)$ with respect to the Lebesgue or the counting measure. The parameter θ will be always scalar, unless stipulated otherwise.

We shall consider here mostly the processes with discrete time, addressing ourselves, say, to the standard Poisson process only in Section 3.

However, many of the results expounded here can be transferred to the continuous time case.

In both cases we shall consider Markov stopping times τ (see [9] for the definition) and scalar statistics T_τ that are unbiased estimates of a scalar function $g(\theta)$ of the parameter θ

$$E_\theta(T_\tau) = g(\theta). \tag{1.1}$$

Moreover, we must choose among such unbiased estimates a statistic $\tilde{T}_\tau$ with the minimal variance $D_\theta(\tilde{T}_\tau)$ under the condition

$$E_\theta(\tau) \leqq n, \tag{1.2}$$

where n is a given number. From (1.2) it follows that $P_\theta(\tau < \infty) = 1$.

Note that if such a statistic exists, this statistic and the corresponding stopping rule may depend upon θ. Thus a statistic which is optimal in the above sense uniformly with respect to θ does not exist in general.

However, if instead of exact optimality we consider asymptotic optimality, for $n \to \infty$ (which, of course, corresponds to the individual cost of an experiment converging to zero), the situation changes: the existence of the asymptotically optimal stopping rule and statistic T_τ is proved for the Bayes setup in the works of Bickel and Yahav [1] to [4].

In the present article we shall indicate asymptotically optimal estimates T_τ and stopping rules in the above mentioned setup. The presence or absence of the discontinuities in the corresponding information quantities will be very essential for the subject. An exact quantitative formulation will be given later.

We shall give some results for "small samples" about existence conditions for sampling plans which are efficient in the Rao-Cramér sense, about the types of such plans, and about the characterization of first hit plans by the simplest properties of Markov stopping times.

2. Asymptotic results in the case of the absence of discontinuities in the information quantities

In the present section we want to propound the point of view which indicates that, roughly speaking, if there are no discontinuities in the information quantities and the cost of the experiment is small, the sequential estimation in the setup described above can give only an infinitely small gain in comparison with the fixed sample method of estimation. We shall consider the family $\mathscr{P}_\theta$ of distributions with the density $f(x, \theta)$ with respect to the Lebesgue measure (some other requirements will be imposed later on).

Here we shall restrict our attention to the processes with independent increments and discrete time, that is, the simplest case of a repeated sample $x_1, x_2, \cdots$. For the continuous time case the corresponding theory is not yet worked out; to all appearances it can be made by analogy.

We shall look for an asymptotically unbiased estimate $\tilde{T}_\tau$ of the parameter θ which minimizes the variance $D_\theta(T_\tau)$ up to the infinitely small quantities of a given order under conditions $E_\theta \tau \leqq n$ and $n \to \infty$.

We shall show that, roughly speaking, in the absence of discontinuities in the information quantities such a statistic T_t can be constructed in the class of trivial stopping rules (τ = constant).

In the proof, we shall rely upon some theorems of Ibragimov and Hasminski on the asymptotic behavior of generalized Bayes estimates for constant sample sizes [12].

By the absence of discontinuities in the information quantities we mean the fulfillment of the following conditions of Ibragimov and Hasminski which we shall strengthen somewhat for the sake of convenience in the exposition.

CONDITION 1. *The density $f(x, \theta)$ with respect to the Lebesgue measure is measurable in both arguments, and $\int\int |\theta| f(x, \theta) f(x, \theta_0)\, dx d\theta < \infty$ for each point θ_0 in the interval of the values of the argument Θ.*

CONDITION 2. *As $|\theta| \to \infty$, the integral $\int f(x, \theta) f(x, \theta_0)\, dx \to 0$.*

CONDITION 3. *As* $\varepsilon \to 0$, *the integral* $\int f^{1-\varepsilon}(x, \theta_0)\, dx \to 1$.

CONDITION 4. *For each set of real numbers* $\theta_1, \cdots, \theta_s$ *and for appropriately chosen intervals* $[0, T_1], \cdots, [0, T_s]$, *we shall have for all* $t_j \in [0, T_j]$ *and* $\xi \downarrow 0$

$$(2.1) \qquad \int \prod_{j=1}^{s} \left(\frac{f(x, \theta_0 + \xi\theta_j)}{f(x, \theta_0)}\right)^{t_j} f(x, \theta_0)\, dx = 1 + \xi^\alpha a(t_1, \cdots, t_s; \theta_1, \cdots, \theta_s) + o(\xi^\alpha),$$

where the number α *does not depend upon* $\theta_1, \cdots, \theta_s$ *and the function* $a(\cdot) \not\equiv 0$.

CONDITION 5. *For all* θ_1, θ_2 *with* $|\theta_j| \leqq H < \infty$, $i = 1, 2$, *we have for* $\xi \to 0$

$$(2.2) \qquad \left|\int\left\{\log \frac{f(x, \theta_0 + \xi\theta_1)}{[f(x, \theta_0 + \xi\theta_2)]}\right\} f(x, \theta_0 + \xi\theta_2)\, dx\right| \leqq |\xi|^\alpha c_1(|\theta_2 - \theta_1|),$$

$$(2.3) \qquad \left|\int\left\{\log^2 \frac{f(x, \theta_0 + \xi\theta_1)}{[f(x, \theta_0 + \xi\theta_2)]}\right\} f(x, \theta_0 + \xi\theta_2)\, dx\right| \leqq |\xi|^\alpha c_2(|\theta_2 - \theta_1|),$$

where $[f(x, \theta_0 + \xi\theta_2)]$ *coincides with* $f(x, \theta_0 + \xi\theta_2)$ *if* $f(\cdot, \cdot) \neq 0$ *and equals* 1 *if* $f(\cdot, \cdot) = 0$. *The functions* $c_i(h)$, $i = 1, 2$, *depend upon* H, *and* $c_i(h) \to 0$ *for* $h \to 0$.

Under Conditions 1 to 5 the theorems of Ibragimov and Hasminski [12] can be applied to the study of the asymptotic behavior of the variance of the Pitman estimate for a repeated sample.

We form the Pitman estimates $\tilde{\theta}_n$ for the parameter θ. as

$$(2.4) \qquad \tilde{\theta}_n = \int \theta p_n(\theta)\, d\theta,$$

where

$$(2.5) \qquad p_n(\theta) = \frac{\prod_{i=1}^{n} f(x_i, \theta)}{\int \prod_{i=1}^{n} f(x_i, \theta)\, d\theta}.$$

First consider the case of the location parameter, that is, $f(x, \theta) = f(x - \theta)$. Apart from Conditions 1 to 5, we also assume the finiteness of the following quantities (a) $\int x^2 f(x)\, dx$, (b) the information type quantities $E_0|\ell^{(n)}(x_i)|^s$ with $n \leqq 5$ and $s \leqq 20$, and (c) $E_0 \max_{|\theta| \leqq \varepsilon} |\ell^{(s)}(x_i - \theta)|^s$ with $s \leqq 20$ and $\varepsilon > 0$, where ε is a small given number. Here $\ell(x) = \log f(x)$, and the above include the Fisher information quantity $I = E_0(\partial \log f/\partial x)^2$.

Then, according to [12] we can assert that the unbiased Pitman estimate (2.4) for the location parameter θ_0 has a finite variance; moreover

$$(2.6) \qquad E_\theta(\tilde{\theta}_n - \theta)^2 = \frac{1}{I_n} + \frac{c_1}{n^2} + o\left(\frac{1}{n^2}\right)$$

where c_1 is a constant depending upon $f(x)$ only.

In the general case, let $\ell(x, \theta) = \log f(x, \theta)$ and let $I_\theta = E(\partial\ell/\partial\theta)^2$ be the Fisher information number. If there is an $\varepsilon > 0$ such that the moments $E_\theta|\ell_\theta^{(n)}(x, \theta_0)|^s$, with $n \leqq 5$ and $s \leqq 20$, and $E_\theta \max_{|\theta| \leqq \varepsilon} |\ell_\theta^s(x_1, \theta_0 - \theta)|^s$, with

$s \leqq 20$, are finite and if $\max_\theta E_\theta \tilde{\theta}_n^2 < \infty$, we have

$$E_\theta(\tilde{\theta}_n - \theta) = \frac{c_0(\theta)}{n} + o\left(\frac{1}{n}\right), \tag{2.7}$$

$$E_\theta(\tilde{\theta}_n - \theta)^2 = \frac{1}{I_\theta n} + \frac{c_1(\theta)}{n^2} + o\left(\frac{1}{n^2}\right), \tag{2.8}$$

where the $c_i(\theta)$ for $i = 0, 1$, are certain constructively given functions.

We return to the location parameter and formula (2.6). Here $\tilde{\theta}_n$ is an unbiased estimate of the parameter θ which possesses a variance. The relation (2.6), by comparison with the well known Rao-Cramér inequality shows that asymptotically, for large values of n, the estimate $\tilde{\theta}_{[n]}$ is very good.

Consider a Markov stopping rule τ with $E_\theta \tau < \infty$, and an unbiased estimate T_τ of the location parameter θ.

According to the well known Wolfowitz inequality (see [13]), we have

$$D(T_\tau) \geqq \frac{1}{IE_\theta(\tau)}. \tag{2.9}$$

If the condition

$$E_\theta \tau \leqq n \tag{2.10}$$

holds. Then (2.6) implies directly that

$$D(T_\tau) \geqq D(\tilde{\theta}_{[n]}) \left[1 + o\left(\frac{1}{n}\right)\right]. \tag{2.11}$$

We can formulate (2.11) as the following theorem.

THEOREM 1. *Assume that the Ibragimov-Hasminski conditions for the absence of discontinuities in the information quantities are satisfied. Then for any unbiased sequential estimate of the location parameter θ subject to the restriction $E_\theta(\tau) \leqq n$, the relative improvement of the variance over that of the constant sample size method is at most of the order $o(1/n)$.*

For the general case, taking into account the bias (see Equations (2.7) and (2.8)), we get a similar result.

In that case we must take an estimate $\tilde{\theta}_n$ having bias $o(1/n)$. We obtain Theorem 2.

THEOREM 2. *Assume that the Ibragimov-Hasminski conditions for the absence of discontinuities in the information quantities are satisfied. Then for any unbiased sequential estimate of the location parameter θ subject to the restriction $E_\theta(\tau) \leqq n$, the relative improvement of the mean quadratic deviation is at most $o(1/n)$.*

For the proof of Theorem 2, we use (2.7) and (2.8). As an asymptotically optimal estimate for the sample size n, we take $\tilde{\theta}_{[n]}$. In forming the Wolfowitz inequality, we use (2.8). In other respects we argue as before.

Note that in the Theorems 1 and 2 we consider the mean quadratic deviation from the parameter value of the estimate $\tilde{\theta}_{[n]}$ itself, not that of the normed limit distribution. The computation of such a deviation often presents considerable difficulties (for analogous computations relating to the maximum likelihood estimates, see, for instance, [14]).

3. Sequential estimation in the case of discontinuities in the information quantities

In this section we shall give certain examples where the presence of discontinuities in the information quantities leads to the existence of sequential estimation procedures giving a considerable gain in the variance of the estimate over that of the fixed sample size method. Our examples relate to the location parameter θ and the density $f(x - \theta)$ which is continuous for $|x - \theta| < \frac{1}{2}$ and has the carrier $|x - \theta| \leqq \frac{1}{2}$.

As $f(x - \theta) = 0$ for $|x - \theta| > \frac{1}{2}$, we have discontinuity in the expressions for the Fisher information quantity and cannot use the Rao-Cramér and Wolfowitz inequalities.

We assume that f satisfies conditions insuring, for sample sizes $n \geqq 2$, the existence of the Pitman estimate

$$\tilde{\theta}_n = \int \theta p_n(\theta)\, d\theta, \tag{3.1}$$

where

$$p_n(\theta) = \prod_{i=1}^{n} f(x_i - \theta) \left[\int \prod_{i=1}^{n} f(x_i - \theta)\, d\theta \right]^{-1}. \tag{3.2}$$

As is well known (see [10] where the literature is indicated), the Pitman estimate $\tilde{\theta}_n$ will be unbiased and optimal with respect to the variance estimate of θ in the class of all "regular" estimates T_n (that is, such that $T_n(x_1 + c, \cdots, x_n + c) = T(x_1, \cdots, x_n) + c$). Therefore we can let this estimate exemplify the constant sample size n and compare its variance with that of estimates given by sequential methods for $E_\theta \tau \leqq n$.

For the simplest case of the uniform distribution $f(x) = 1$ with $|x| \leqq \frac{1}{2}$, we have (see [10]) $\tilde{\theta}_n = \frac{1}{2}(x_{\max} - x_{\min})$ and $D(\tilde{\theta}_n) = [2(n + 1)(n + 2)]^{-1}$. Choose the stopping rule $\tau = \min_n\{n; x_{\max} - x_{\min} > 1 - \varepsilon(n)\}$ and use the estimate $\tilde{\theta}_\tau$. For a suitable $\varepsilon(n)$ it will be unbiased such that $E_\theta(\tau) = n$ and $D(\tilde{\theta}_\tau) = \frac{1}{6}n^2$. It follows that

$$D(\tilde{\theta}_\tau)/D(\theta_n) \to \tfrac{1}{3} \tag{3.3}$$

for $n \to \infty$ and that the sequential estimation improves the asymptotic variance by a factor of three.

Note that in this case for any strictly monotonic function W the stopping times $\tau = \min\{n; |x_{\max} - x_{\min}| > 1 - \alpha\}$ minimize the risk function $W(|x - \theta|) + c\tau$ for an appropriate choice of α.

It is interesting that a similar result holds for much more general distributions. I. I. Iaura and A. N. Shalyt were kind enough to make a calculation requested by the present authors. It was based upon the article of Ibragimov and Hasminski and led to the following theorem.

THEOREM 3. *Let the density $f(x)$ having the carrier $|x| < \frac{1}{2}$, be symmetric and continuous together with its first derivative for $|x| < \frac{1}{2}$. Assume $f(x) = 1$ for $0 < \frac{1}{2} - |x| \leqq \varepsilon$ where $\varepsilon > 0$ is a constant. Then the stopping rule $\tau = \min\{\nu : x^{\nu}_{\max} - x^{\nu}_{\min} > 1 - 2/n\}$ leads to the unbiased estimate $\theta_{\tau} = \frac{1}{2}(x^{\tau}_{\max} - x^{\tau}_{\min})$ and gives a variance asymptotically three times smaller than the fixed sample size method.*

Thus we see that the presence of discontinuities in the information quantities may lead to a considerable gain in variance by use of sequential instead of constant sample size estimation. The study of such a connection between discontinuities in the information quantities and the improvement of the estimation quality in the sequential case seems to be interesting.

We must remark that these discontinuities do not always lead to an asymptotic gain in the variance. Thus, Shalyt informed the authors that for the density $f(x - \theta) = \exp\{-(x - \theta)\}$ for $x \geqq \theta$ and $f(x - \theta) = 0$ for $x < \theta$, there is no such gain.

4. Binomial and multinomial processes: the Poisson process

In this section we shall first consider estimation plans for processes with discrete time and a finite number of states. Such processes, inasmuch as the structure of the set of values of the corresponding random variable is not essential, are described by the multinomial scheme. The frequency vector describing the appearance of different states is a sufficient statistic in this case.

The estimation plans for the binomial case were studied in [5] and [6]. In particular, in [5] a statistic is given which is an unbiased estimate of the vector of probabilities of the states or of a given polynomial function of these probabilities. It is natural to indicate the plans for which such an estimate is unique.

A Markov stopping rule τ is called complete if the only unbiased estimate of zero is the trivial statistic $T_{\tau} \equiv 0$. In the case of bounded binomial schemes a necessary and sufficient condition for completeness of a Markov stopping rule is indicated in the work of De Groot [5]. In the multinomial case the following theorem of Zaidman holds.

THEOREM 4A. *For the completeness of a Markov stopping rule it is necessary that it be nonrandomized.*

This means that all the points of the phase space are subdivided into three groups: the set of boundary points B, the set of transition points, and the set of unattainable points. A plan with such a Markov stopping rule and with the statistic obtained by the Rao-Blackwell process will be called a first hit plan in what follows.

A plan called "finite for θ" if $E_{\theta}\tau < \infty$, and "finite" if it is finite for all $\theta \in \Theta$. If no proper subset $B' \subset B$ is a set of boundary points of a finite plan, the plan is called a minimal plan.

THEOREM 4B. *For the completeness of a first hit plan minimality is necessary.*

The proof of this theorem as well as Theorem 4A is based upon the construction of a nontrivial unbiased estimate of zero when the conditions of the theorem are violated.

Necessary and sufficient conditions of a geometrical nature for the completeness of multinomial first hit plans have not yet been found. However there are sufficient conditions valid for a large class of plans. In particular, the following theorem holds.

THEOREM 5. *Suppose that for an n-normal bounded plan, for each $x \in B$ there is a number i such that all the points $y = (y_1, \cdots, y_n)$ satisfying the conditions*

$$(4.1) \qquad y_i > x_i; \qquad y_j \leqq x_j, \qquad j \neq i; \qquad \sum_j y_j = \sum x_j + 1,$$

are unattainable. Then the plan is complete.

The proof of this theorem is effected by induction. All the points of the boundary B are subdivided into the sets $B_k, B_{k+1}, \cdots, B_\ell$ on which the statistic τ has constant values (k being the minimal value). Further construct a new plan S' in which the set B_k is included in the set of transition points, whereas B_{k+1} is extended to $B'_{k+1} = B_k + B_\Delta$ so as to make the new plan S' closed and minimal. The induction hypothesis is that S' is a complete plan (the induction basis is the constant sample size plan $\tau = \ell$). Under these conditions we must prove that if $\phi(x)$ is a function on B and

$$(4.2) \qquad E_p\phi \equiv 0$$

p being the vector of the state probabilities, then $\phi \equiv 0$. Now the process of induction requires proving that $\phi = 0$ on B_k. The completeness of S' requires proving that for each $y \in S'$ we must have the equality

$$(4.3) \qquad EK(0, x)\phi(x) = 0,$$

where the summation extends over all the points $x \in S_k$ which can be reached starting from the point y, and where $K(0, x)$ is the number of trajectories reaching x and starting from the origin. Under the conditions of the theorem, the system (4.3) proves to have as many equations as variables and its matrix is triangular. This follows from the possibility of indexing the points of B_k and B_Δ in such a way that passage from the ith point of B_k to the jth point of B_Δ is impossible for $j > i$ (this indexing will be needed later also). Moreover as the matrix of the system (4.3) is nonsingular, we have $\phi \equiv 0$ on B_k which terminates the induction.

Note that for the binomial case condition, (4.1) is necessary because it is fulfilled only for the plans described by De Groot [5]. The plans described in Theorem 5 have another interesting property, which is in a sense the inverse of the completeness property.

THEOREM 6. *If S and S' are two different first hit plans under the conditions of Theorem 5, then*

$$(4.4) \qquad E_p\tau \neq E_p\tau'.$$

This theorem is proved by the same process of induction as the previous one.

We prove step by step the coincidence of the sets B and B'. Having proved for an appropriate set R, $R \subset B$ and $R \subset B'$, we study $E_p(\tau, B\backslash R)$ and $E_p(\tau, B'\backslash R)$ where

$$E_p(\tau, M) = P_p(M)E_p(\tau|x \in M), \tag{4.5}$$

where $P_p(M)$ is the probability of reaching M from the origin, and where $E_p(\tau|x \in M)$ is the conditional expectation of τ given that we start from $x \in M$. The functions $E_p(\tau, B\backslash R)$ and $E_p(\tau, B'\backslash R)$ are polynomials in P_i.

Let plan S have the corresponding boundary set $B = \{B_k, \cdots, B_\ell\}$ and let plan S' have the set $B' = \{B'_{k'}, \cdots, B'_{\ell'}\}$ (see the proof of Theorem 5). Without loss of generality suppose that $k \leqq k'$. We shall prove that $k = k'$ and $B_k = B_{k'}$. The vectors of B_k can be subdivided into several subsets, each indexed as in the proof of Theorem 5, by writing the components of vectors in a sequence fixed for each subset separately and indexing in decreasing lexicographical order, beginning with the vector $(k, 0, \cdots, 0)$ in the sequence selected for this set.

We take the first of these subsets of B_k and the first of its vectors x^0 and make the first component of the vector of the state probabilities tend to 1. We get $\lim E_p\tau' = K(0, x^0)k$ and so the vector x^0 must belong to B'. Then we continue the process in the same way with the stipulation that before making the first component of the probability vector tend to 1, we must divide $E_p(\tau, B\backslash R)$ and $E_p(t, B'\backslash R)$, which are equal, by $\Pi P x_i^0$.

The proof does not hold for an arbitrary complete plan (except for the binomial case) but one can conjecture that not only complete, but even simply minimal plans have this property. However, this question remains open. For nonminimal plans there are examples of families of plans with the same $E_p\tau$.

For sampling without repetition there is no separate completeness problem because the following theorem holds.

THEOREM 7. *For a plan S to be complete for sampling without repetition it is necessary and sufficient that it be complete for the repeated sample method.*

We pass now to the Poisson process. This process can be treated as the limit case of the binomial process and the results relating to it are limit cases of the analogous facts for the Bernoulli scheme. For instance, the analogues of Theorems 5 and 6 are as follows.

THEOREM 8. *For the completeness of a bounded first hit plan S given by a set of boundary points, consisting of a finite number of segments parallel to the time axis, in the space of the sufficient statistics $(t, \eta(t))$, it is necessary and sufficient that the common length of these segments be equal to the essential value of the maximum stopping time corresponding to this plan.*

THEOREM 9. *A complete bounded first hit plan with a finite number of boundary segments is determined by the values of the mean stopping time of this process as a function of the intensity λ.*

The proofs of these theorems are analogous to the proofs for the binomial case and use induction on the length of the maximum stopping time. (See also [7], [8].)

5. Description of efficient sequential plans for the renewal process

In the present section we shall consider plans of sequential estimation for parameters of stochastic processes with independent increments which are renewal processes.

We shall consider sequential plans which are efficient in the sense of Wolfowitz's identity on a certain interval of parameter values. Investigations of this kind were started by De Groot [5] in 1959 and developed considerably by Trybula [6] in 1968. Trybula considered the Poisson process, Brownian motion, and so forth. In Section 7 of his interesting work he makes some general remarks on efficient plans of sequential estimation for homogeneous processes with independent increments.

One can consider homogeneous processes with independent increments and with discrete or continuous time.

Consider the case of discrete time. This will correspond to the repeated sample $x_1, x_2, \cdots$.

We shall suppose that the quantities x_i have a density $f(x, \theta) > 0$ with respect to the Lebesgue measure or the counting measure with carrier consisting of integer numbers. The parameter θ will lie in a certain interval Θ of the real axis. Let τ be a Markov stopping time and let $L(x, \tau, \theta)$ be the likelihood function. We shall suppose that requirements sufficient for fulfillment of the Rao-Cramér information inequalities hold (regarding these requirements, see for instance [15], [16]). Let T_τ be an estimate which is unbiased and efficient in the sense of the variance for a given function $g(\theta)$ of the parameter in a certain interval of parameter values. In this case

$$T_\tau - g(\theta) = h(\theta) \frac{\partial \log L}{\partial \theta}. \tag{5.1}$$

Suppose now that the function $g(\theta)$ has a derivative $g'(\theta)$ having no zeros on a certain interval $I_1 \subset \Theta$. Then we can replace the parameter θ by the parameter $\theta_1 = g(\theta)$ and use the following theorem proved by Kagan in [17].

THEOREM 10. (Kagan). *Assume that the following conditions are satisfied:*

(1) *the density $f(x, \theta)$ satisfies the standard regularity conditions (see, for instance,* [17]);

(2) *the Markov stopping time τ is such that for a certain $n > 2$ the set $\{\tau = n\} = M_n$ contains $\overline{\Delta} = \Delta_1 \times \Delta_2 \times \cdots \times \Delta_n$, where the Δ_i are the intervals of R^1;*

(3) *the statistic $[T_\tau, \tau]$ is sufficient for θ in a certain interval;*

(4) *the mapping $T_n \to R^1$ is nontrivial on $\overline{\Delta}$ and for each pair of values θ, θ' in the interval mentioned above, there exists an i, $1 \leqq i \leqq n$, such that for $x \in \Delta_i$, $f(x, \theta)$ is not a multiple of $f(x, \theta')$.*

Then $f(x, \theta)$ is of exponential type.

Thus, under the conditions of the theorem the density $f(x, \theta)$ corresponds to the exponential family of distributions

$$f(x, \theta) = \exp\{\theta T(x) - \rho(\theta)\} \tag{5.2}$$

for a certain interval of values of θ. Here $\rho(\theta)$ is the norming function and $T(x)$ is a statistic such that for each value $\tau = n$, $T_\tau = T_n = \Sigma_{i=1}^n T(x_i)$ gives the value of the sufficient statistic. An analogous result will hold also for densities with respect to the counting measure with carrier consisting of integers.

Without restricting the generality, we can suppose that relations (5.1) and (5.2) hold in an interval of values of θ where we can choose the values θ_1, and θ_2, and we can write equation (5.1) for them. Following articles [5] and [6] we can subtract one equality from the other. Then we shall get the linear relation

$$aT_\tau + b\tau + c = 0 \tag{5.3}$$

so that a sequential estimation plan which is efficient in the sense of the variance and corresponds to τ must be a first hit plan determined by the linear relation (5.3).

Let us suppose now that the sufficient statistic T_τ is integer valued and nonnegative so that the process $\{T_t, t = 1, 2, \cdots\}$ is a renewal process. Let us mark the integer values time on the x axis and the values of the statistic T_t on the y axis. Then our process will be represented in the first quadrant.

In order that the plan with the boundary (5.3) be closed for a certain interval of values of the parameter θ the fulfillment of the following relation is necessary:

$$\tau = AT_\tau + B, \tag{5.4}$$

A and B being positive integers.

Denote by $Cl(S)$ the set on which the plan S is closed (see [9], [10]), that is, the set of values θ for which $E_\theta \tau < \infty$.

Suppose that on an interval $I \subset \Theta$, $E_\theta \tau$ and $E_\theta \tau^2$ exist. Let $K(\tau, T_\tau)$ be the number of "trajectories" starting from the point (0, 0) and reaching the point (τ, T_τ) on the plan boundary (it is finite because we consider only renewal processes). Then we have

$$\sum_{\partial s} K(\tau, T_\tau) \exp\{\theta T_\tau - \tau\rho(\theta)\} = 1. \tag{5.5}$$

Differentiating this identity with respect to θ for $\theta \in I$, we get under certain natural conditions the Wald identity $E_\theta T_\tau = \rho'(\theta) E_\theta \tau$. From this and from (5.4), it follows that

$$E_\theta \tau = B[1 - A\rho'(\theta)]^{-1}. \tag{5.6}$$

Then (5.4) implies

$$\begin{aligned} E_\theta \tau &= A_\tau + B, \\ E_\theta t^2 - AE_\theta T_\tau + BE_\theta \tau &= 0, \\ E_\theta T_\tau \tau - AE(T_\tau^2) - BE_\theta T_\tau &= 0 \end{aligned} \tag{5.7}$$

and a straightforward calculation gives

$$D_\theta \tau = A^2 B\rho''(\theta)[1 - A\rho'(\theta)]^3. \tag{5.8}$$

(The calculation was made by Nz. M. Halfina to whom the authors wish to express their gratitude.)

This expression coincides with I_θ^{-1}. By dint of Wolfowitz's inequality this implies the efficiency of the unbiased estimate of the function (5.6) obtained from plan S, in the interval $I \subset Cl(S)$. All other efficiently estimable functions can be obtained from (5.6) with the help of linear transformations with constant coefficients (see [16]).

If $f(x, \theta)$ is the density with respect to the Lebesgue measure we shall again suppose that the statistic $T(x)$ is nonnegative. It is assumed to have a continuous density. In that case only constant sample size plans can be efficient, as a rule. Some particular cases of the processes with independent increments were considered by Trybula [6]. To study them systematically as we did for the discrete time case a generalization of Kagan's theorem from [17] for the continuous time case is needed. It has not yet been obtained.

6. Unsolved problems

We shall indicate here some unsolved problems of the theory of sequential estimation which seem to present a certain interest:

(1) generalize the results of Ibragimov and Hasminski to the case of multivariate distributions and formulate corresponding consequences for sequential analysis;

(2) compare the asymptotic properties of sequential estimation and fixed sample size method for nonquadratic losses and in the absence of discontinuities in the analogue of the information quantities;

(3) formulate and extend the above problems in the continuous time case;

(4) study the appearance of discontinuities in the information quantities and the possibilities of applying sequential analysis to improve the quality of the estimate in the case of discrete time and scalar or vector parameters for experiments which have a small cost;

(5) study the same question for the case of continuous time and scalar or vector parameter;

(6) find the geometric conditions for completeness of unbounded binomial plans which are closed only on a proper subset of the set (0, 1);

(7) find the geometric conditions for completeness of multinomial plans;

(8) in the multinomial case, investigate the possibility that a bounded minimal plan be determined by the values of $E_p\tau$;

(9) investigate the most general conditions for the existence of linear optimal estimation plans for homogeneous processes with independent increments and discrete or continuous time.

REFERENCES

[1] P. J. Bickel and J. A. Yahav, "Asymptotically pointwise optimal procedures in sequential analysis," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 1, pp. 401–413.

[2] ———, "Asymptotically optimal Bayes and minimax procedures in sequential estimation," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 442–456.

[3] ———, "Some contributions to the asymptotic theory of Bayes solutions," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*, Vol. 2 (1969), pp. 257–276.

[4] ———, "On an A. P. O. rule in sequential estimation with quadratic loss," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 417–427.

[5] M. H. DeGroot, "Unbiased sequential estimation for binomial population," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 80–101.

[6] S. Trybula, "Sequential estimation in processes with independent increments," *Dissertationes Math. Rozprawy Mat.*, Vol. 60 (1968), pp. 1–50.

[7] R. A. Zaidman, Yu. V. Linnik, and S. V. Romanovsky, "Sequential estimation plans and Markov stopping times," *Dokl. Akad. Nauk SSSR*, Vol. 185 (1969), pp. 1222–1225. (In Russian.)

[8] R. A. Zaidman, Yu. V. Linnik, and V. N. Sudakov, "On sequential estimation and Markov stopping times for processes with independent increments," *USSR–Japan Symposium on Probability, Habarovsk, August* 1969, Novosibirsk, "Nauke," 1969, pp. 127–143. (In Russian.)

[9] A. N. Shiryayev, *Statistical Sequential Analysis*, "Nauke," 1969.

[10] A. N. Shalyt, "Some results on the sequential estimation of the location parameter," *Dokl. Akad. Nauk SSSR*, Vol. 189 (1969), pp. 57–58.

[11] ———, "On the optimal behavior of the location parameter in the class of invariant procedures," *Dokl. Akad. Nauk SSSR* (1970). (In Russian.)

[12] I. A. Ibragimov and R. Z. Hasminski, "On the asymptotic behavior of generalized Bayes estimates," *Dokl. Akad. Nauk SSSR*, Vol. 194 (1970), pp. 257–260. (In Russian.)

[13] J. Wolfowitz, "The efficiency of sequential estimation and Wald's equation for sequential processes," *Ann. Math. Statist.*, Vol. 18 (1947), pp. 215–230.

[14] Yu. V. Linnik and N. M. Mitrofanova, "Some asymptotic expansions for the distribution of the maximum likelihood estimate," *Contributions to Statistics Presented to Professor P. C. Mahalanobis on the Occasion of his Seventieth Birthday* (edited by C. R. Rao), Oxford and New York, Pergamon Press, 1965, pp. 229–238.

[15] L. N. Bolshev, "A refinement of the Cramér-Rao inequality," *Teor. Verojatnost. i Primenen.*, Vol. 6 (1961), pp. 319–326.

[16] Yu. V. Linnik, "Some remarks on the inequalities of Rao-Cramér, and Bhattacharya," *Matem. Zametki* (1970).

[17] A. M. Kagan, "Remarks on the sufficiency in a sequential process," *Matem. Zametki* (1970). (In Russian.)

[18] R. A. Zaidman, "Completeness of Markov stopping rules," *Zap. Naučn. Sem. Leningrad Otdel. Mat. Inst. Steklov* (1970). (In Russian.)

[19] Yu. V. Linnik, *Statistical Problems with Nuisance Parameters*, Moscow, Nauke, 1966. (English translation, Providence, American Mathematical Society, 1968.)