holds, where the quantities with asterisks belong to the alternative distribution, the test based on $\chi^2$ is asymptotically consistent ($n \to \infty$). Concerning efficiency, the test is convenient for moderate $r$, the situation usually met in practice.

References


For further references see:


Presented to the seminar MATHMATICAL STATISTICS
September 15–December 18, 1976

Mathematical Statistics
Banach Center Publications, Volume 6
PWN—Polish Scientific Publishers
Warsaw 1980

SEQUENTIAL ESTIMATION IN PROCESSES WITH
INDEPENDENT INCREMENTS

WOLFGANG WINKLER
Sektor Mathematik, Technische Universität Dresden, Dresden, D.D.R.

1. Introduction

Sequential statistical estimation problems, which are characterized by a random sample size, have been considered about thirty years ago in connection with simple random processes with independent increments, especially for the Bernoulli process, Poisson process, and others. (See, for instance, [3], [4], [11] and [20].) The case of the Poisson process has been considered by Trybula [19] in 1968. In his paper Trybula solved the problem of describing sequential sample plans for the Poisson process, which are optimal in a certain sense and he studied the properties of such plans concerning efficiently estimable functions.

It is natural to ask for extensions of the statistical results to a wide class of processes with independent increments using arbitrary Markov times to stop the observation of the considered random process. In fact, this is possible for the so-called exponential class of random processes with independent increments.

In the following we give a short summary about some results in this direction, which are due to J. Franz, M. Magiera, S. Trybula, and W. Winkler. For details see [5]–[10], [13]–[16], [19].

2. The exponential class and stopping times

Let us denote by $X(t) = (X_1(t), ..., X_m(t))^T$ an $m$-dimensional random process defined on the probability space $(\Omega, \mathcal{F}, P)$, where $X(t)$ takes values in $E = \mathbb{R}^m$ and $t \in T$ runs over all non-negative integers or over all non-negative real numbers as usually. Moreover, let $\mathcal{F}_t$, $t \in T$, denote the $\sigma$-algebra generated by the random vectors $X(t)$, $t \in T$.

We assume that the probability measure $P$ depends on an unknown parameter $\Theta = (\theta_1, ..., \theta_r)^T \in \Theta \subseteq \mathbb{R}^r$ and we write $P = P_\Theta$.

DEFINITION. We shall say that $X(t)$, $t \in T$, belongs to the exponential class if the following conditions are satisfied:

\[ X(t) \in \text{Exp}(\lambda_t) \quad \text{for all } t \in T, \quad \lambda_t \in \mathbb{R}^+ . \]
(1) $X(t)$ is a right-continuous homogeneous random process with independent increments and $\mathbb{P}(X(0) = 0) = 1$. $\mathbb{E} X_i(t) < +\infty$ holds for all $i = 1, 2, \ldots, m$ and $t \in T$.

(2) The probability distributions at time $t$ are dominated by a $\sigma$-finite measure $\nu$ and the densities (Radon–Nikodym derivatives) may be represented in the form

$$f(x, t, \theta) = g(x, t) \exp \left( \sum_{i=1}^{m} a_i(\theta) x_i + b(\theta) t \right),$$

where $x = (x_1, \ldots, x_m)^T$ and $g$ is a non-negative function.

In the one-dimensional case ($m = 1$), for instance, the Bernoulli process, negative-binomial process, gamma process, Poisson process, and Wiener process belong to the exponential class. We remark that in the one-dimensional case the exponential class of processes with independent increments was treated by Magiera [13] and by Franz and Winkler [3], [8], [9], nearly at the same time.

Now, let us consider a stopping time $\tau$, that is, a random variable $\tau$ defined on $\Omega$ with values in $T \cup \{-\infty\}$ and the property: $\omega \in \Omega: \tau(\omega) \leq t \in \mathbb{B}_T$ for every $t \in T$. The $\sigma$-algebra of the $\tau$-past of the process $X(t)$ is denoted by

$$\mathbb{B}_T = \{ F \in \mathbb{B}: F \cap \{ \omega: \tau(\omega) \leq t \} \in \mathbb{B}, \text{ for every } t \in T \}.$$

In the following we assume $\mathbb{E} \tau < +\infty$; hence $\mathbb{P}(\tau < +\infty) = 1$ because we are interested in finite stopping times only.

We are now in position to formulate a representation theorem for the common distribution of the pair $(\tau, X(\tau))$, which will be very useful for statistical considerations.

**Theorem 1.** Let $X(t)$ be taken from the exponential class and let $\tau$ be any stopping time. Then for every $\theta_0$ there exists a probability measure $P_{\theta_0}$, which does not depend on the unknown parameter $\theta$, such that

$$P_{\theta_0}(\{ (\tau, X(\tau)) : \tau \in S \}) = \int_{\mathbb{R}_+} \exp \left( \sum_{i=1}^{m} a_i(\theta) x_i + b(\theta) \tau \right) d\nu_{\theta_0}$$

where $\Omega_\theta = \{ (\tau, X(\tau)) : \tau \in S \}, S \subseteq T \times \mathbb{R}$ and $a_i(\theta) = a_i(\theta_0) - a_i(\theta_0)$, $b(\theta) = b(\theta_0) - b(\theta_0)$.

Remark. Introducing $U = T \times \mathbb{R}, \mathcal{U} = \mathcal{T} \times \mathcal{C}$ and denoting by $Q$ the measure on $\mathcal{U}$ defined by $Q(S) = P(\{ (\tau, X(\tau)) : \tau \in S \}), S \in \mathcal{U}$, where $i(\omega), x_i(\omega) = (x_1(\omega), \ldots, x_m(\omega))^T$ are the coordinate functions of a point $u = (\tau, x) \in \mathcal{U}$, representation (2) may be written

$$Q_{\theta}(S) = \int_{\mathbb{R}_+} \exp \left( \sum_{i=1}^{m} a_i(\theta) x_i + b(\theta) i(\omega) \right) Q_{\theta_0}(d\omega).$$

Therefore in the case of exponential class we are able to get an explicit formula for the density occurring in the well-known Sudakov lemma (see [18]) which holds for a certain more general class of Markov processes.

The proof of Theorem 1 is given in [8] in the case of real processes; the proof of the multidimensional case needs only some obvious modifications.

3. **Sufficient statistics**

We now turn our attention to the problem of obtaining sufficient statistics. If we observe a process satisfying condition (1) of the definition of the exponential class during a fixed interval $[0, t]$, then the last observation $X(t)$ is a sufficient statistic for $\theta$. More precisely, we have that the process $X(t)$ belongs to the exponential class if and only if $X(t)$ is sufficient for every fixed observation interval $[0, t]$ and the following condition holds: For every $\theta_1, \theta_2$, the probability distributions of $X_{\theta_1}(t)$ and $X_{\theta_2}(t)$ are mutually absolutely continuous and the Radon–Nikodym derivatives $P_{\theta_1}(\tau, x)$ are continuous with respect to arguments $\tau$ and $x$ on $T \times \mathbb{R}$ (see [9], [10]).

If we observe a process with independent increments up to a random stopping time $\tau$, the problem of determining a sufficient statistics is more difficult. By my knowledge, a general solution is not known, but in the case of processes of the exponential class the following theorem gives a positive and expected answer (for the proof see [10]):

**Theorem 2.** Suppose that $X(t)$ belongs to the exponential class and let $\tau$ be any stopping time. Denote by $\mathbb{I}$ the smallest $\sigma$-algebra of $\mathbb{B}_T$ relative to which $(\tau, X(\tau))$ is measurable. Then $\mathbb{I}$ is a sufficient $\sigma$-algebra, i.e. the pair $(\tau, X(\tau))$ is a sufficient statistics to estimate $\theta$.

This theorem allows us to restrict our statistical considerations to observations of $\tau$ and $X(\tau)$. Knowing these quantities, we estimate the parameter vector $\theta$. The problem is to find stopping times $\tau$ and estimators $\phi(\tau, X(\tau))$ which will possess optimal properties.

4. **A general relation between the moments of $\tau$ and $X(\tau)$**

Let $X(t), t \in T$, be a random process of the exponential class. We suppose that the functions $a_1(\theta), \ldots, a_m(\theta), b(\theta)$ in formula (1) are differentiable with respect to the components of the parameter vector $\theta = (\theta_1, \ldots, \theta_m)^T$. We denote by $A = (a_i)$ the matrix with the elements $a_{ij} = \frac{\partial a_i(\theta)}{\partial \theta_j}, i, j = 1, 2, \ldots, m$ and let $b = \left( \frac{\partial b(\theta)}{\partial \theta_1}, \ldots, \frac{\partial b(\theta)}{\partial \theta_m} \right)^T$. Then for the expectation vector of $X(t)$ we have

$$A \cdot E_{\theta}(X(t)) + b = 0.$$

Let us now consider a random stopping time $\tau$. We ask for relations between the moments of $\tau$ and $X(\tau), \ldots, X_n(\tau)$. There are many results connected with some special cases, for instance, Bernoulli, negative-binomial, Poisson, Wiener process, and others. First of all we mention Wald's well-known identities; for further equations of this type see, for instance, [2], [12], [17]. In the following we shall generalize these equations to processes of the exponential class.
Let $\varphi = (\varphi_1, ..., \varphi_p)^T$ be a vector of functions $\varphi_l$, $l = 1, 2, ..., p$, defined on $\mathbb{T} \times \mathbb{E} \times \theta$, satisfying some regularity conditions which guarantee the differentiability under the sign of integration. Then it is possible to obtain the equation
\begin{equation}
E_\varphi(A \mathbb{X}(\tau) + b \tau) \varphi = \text{grad}_h(E_\varphi \varphi^T) - E_\varphi \text{grad}_\varphi \varphi^T.
\end{equation}
For proofs see [10].

We remark that in the special case $p = 1$, $\varphi = 1$ equation (5) turns to a generalization of the Wald first identity
\begin{equation}
A E_\varphi \mathbb{X}(\tau) + b E_\varphi \tau = 0
\end{equation}
which holds in general for $\mathbb{X}(\tau)$ belonging to the exponential class and any stopping time $\tau$.

5. Inequalities of the Cramér-Rao type

In this section we are interested in generalizations of inequalities of the Cramér–Rao type in sequential estimation theory. An interesting result in this direction has been stated by Magiera.

In [13] Magiera considered a class of (one-dimensional) random processes for which the Sudakov lemma is valid and it is assumed that $\theta$ is a real parameter. Let $P_{\theta}(t, x)$ be the density of the measure $Q_\theta$ with respect to $Q_0$, and let $\varphi(\tau, \mathbb{X}(\tau))$ be an unbiased estimator of a function $h(\theta)$ with $E_\varphi \varphi^2 < +\infty$, satisfying certain regularity conditions. Then we have
\begin{equation}
D_\varphi \varphi \geq \frac{\left( \frac{d}{d\theta} \varphi(\theta) \right)^2}{E_\varphi \left( \frac{d}{d\theta} \text{log} P_{\theta} \right)^2}.
\end{equation}
and equality holds at a particular value of the parameter $\theta$ if and only if
\begin{equation}
\varphi = \tilde{g}(\theta) \frac{d}{d\theta} \text{log} P_{\theta} + h(\theta),
\end{equation}
where $\tilde{g}(\theta) \neq 0$.

If in (7) the equality holds the estimator $\varphi$ is called efficient at $\theta$ and the function $h$ is efficiently estimable at the point $\theta$. If equality holds for every $\theta$ the estimator $\varphi$ is said to be efficient and $h$ is efficiently estimable.

In the case of the exponential class with
\[ f(x, t, \theta) = g(x, t) \exp \left( a(\theta) x + b(\theta) \right) \]
inequality (7) takes the form
\[ D_\varphi \varphi \geq \frac{\left( \frac{d}{d\theta} \varphi(\theta) \right)^2}{E_\varphi \left( \frac{d}{d\theta} \frac{d}{d\theta} \text{log} P_{\theta} \right)^2} \cdot \frac{1}{\left( \frac{d}{d\theta} \text{log} P_{\theta} \right)^2}. \]

Suppose that $\varphi$ is efficient. Then it can be shown that the probability distribution of $(\tau, \mathbb{X}(\tau))$ is accumulated on some straight line in $\mathbb{T} \times \mathbb{E}$. For further optimal sequential plan research see [19] for the Poisson process, [16] for the Bernoulli process, [5] for the negative-binomial process.

Let us now turn to the general case of the exponential class defined in Section 1. We consider $h(\theta) = (h_1(\theta), ..., h_p(\theta))^T$, where $h_i$ are functions of $\theta$, and let $\varphi = (\varphi_1, ..., \varphi_p)^T$ be an unbiased estimator of $h$. We define $H = \text{grad}_h h^T$ and $\Gamma = E_\varphi(A \mathbb{X}(\tau) + b \tau)^T(A \mathbb{X}(\tau) + b \tau)$. Since the $\varphi_l = \varphi_l(\tau, \mathbb{X}(\tau))$ are independent of $\theta$, it follows from (5) that
\begin{equation}
E_\varphi(A \mathbb{X}(\tau) + b \tau) \varphi = H.
\end{equation}
Under some regularity conditions, using the Schwartz inequality, it is possible to prove that
\begin{equation}
E_\varphi(\varphi - h)(\varphi - h)^T \geq \Gamma^{-1} H \Gamma^{-1}
\end{equation}
and the equality holds if and only if
\begin{equation}
\varphi = H^{-1} \Gamma^{-1} (A \mathbb{X}(\tau) + b \tau) + h.
\end{equation}
For details see [10]. If the estimator $\varphi$ is efficient in the sense that in (10) the equality holds for every $\theta$, it can be proved under proper conditions that the probability distribution of $(\tau, \mathbb{X}(\tau))$ is accumulated on some $m$-dimensional hyperplane in $\mathbb{T} \times \mathbb{E}$. Using this result, it is possible to discuss in detail the properties of efficiently estimable functions in the considered case.

6. Final remarks

An important problem in sequential estimation theory is to determine the distribution or at least the first moments of the stopping time $\tau$. This problem has been considered in special cases only. Let us consider the negative-binomial process $X(i)$, i.e.
\[ P(X(i) = k) = \frac{\Gamma(i + k)}{\Gamma(i) \Gamma(k + 1)} \cdot \frac{\theta^k}{(1 + \theta)^{i+k}}, \quad k = 0, 1, 2, ..., \]
to illustrate the method which was used at first by Trybula in the case of the Poisson process [19].

We assume that $\tau$ is the moment of the first attaining of the straight line, say $x = \frac{1}{r} \left( t - s, r > 0, s > 0 \right)$, on which the probability distribution of $(\tau, \mathbb{X}(\tau))$ is accumulated in the case of an efficient sequential plan. Let $p_k(s)$ denote the probability of the first attaining of this line for $x = k$. Then, moving this straight line to $x = \frac{1}{r} \left( t - t - t\Delta s \right)$, we obtain
\[ p_k(s + \Delta s) = \sum_{i=s}^{k} p_k(s + x) P(X(\Delta s) = v). \]
For $\Delta r \to 0$ this leads to a system of differential equations with the solution

$$p_{s}(x) = \frac{s^{(s+k(r+1))}}{I(s+k-r+1)/(k+1) - (1/\theta)^{s+k(r+1)}}.$$ 

Using the relations between the moments of $X(r)$ and $r$, we get

$$E r = \frac{s}{1 - \theta}, \quad D^2 r = \frac{\theta(1-\theta)r^2}{(1-\theta)^3}.$$ 

These results are due to Magiera [14]. For the binomial case see [16].

Another interesting method used by Franz [6] is connected with crossing problems. We assume that $X(t)$ is a process of the exponential class with the property that its realizations are either continuous with probability one or have only positive (or only negative) jumps. Then $Z(t) = \gamma t + 2X(t)$, $\delta < 0$, $\gamma + \delta > 0$, is also a process of the exponential class and stopping of $Z(t)$ at the level $s$ is equivalent to stopping of $X(t)$ at the straight line $X = \gamma t - s$. By $\tau_s = \inf\{t: Z(t) > s\}$ we denote the crossings of $Z(t)$ at the level $s$. By a result of Skorohod, $\tau_s$ has independent increments and under some further assumption it can be shown that $\tau_s$ is a process of the exponential class again with the density

$$f_{s}(t, s, \theta) = \frac{s}{\|s\|} \left( \gamma - \frac{\gamma^2}{\delta} - 1 \right) \exp \left( \frac{\theta(\gamma - s)}{\delta} - 1 \right).$$

and

$$E r = \frac{s}{\gamma + \delta}, \quad D^2 r = \frac{\gamma \gamma^2}{(\gamma + \delta)^3}.$$ 

Finally, we remark that it is possible to construct sequential confidence intervals of a fixed length for $\theta$ by using properties of the exponential class; see Franz [5]. These confidence intervals are asymptotically optimal in the sense of Chow and Robbins [1].

References


Sequential Estimation


Presented to the semester