DEMONSTRATIO MATHEMATICA
Vol. XV No 3 1982

Elżbieta Ferenstein

OPTIMAL OBSERVATION STRATEGY FOR A CLASS OF RANDOM EXPERIMENTS

1. Introduction

The objective of most sequential decision problems is to find a sequential decision procedure, consisting of a stopping time and a decision rule, which has some optimal properties. In some cases such a procedure ought to minimize (in a certain sense) a risk function defined as a sum of two risks: the mean cost of observations and the mean loss connected with the chosen decision rule. In other cases the goal is to find procedures which simultaneously minimize one of the above risks and ensure a sufficiently small value of the other one. Thus the control of a sample (observations of a random process under investigation) consists in the choice of a stopping time.

This paper refers to more general sequential decision problems which arise in practical situations with partially observed processes. The idea of partial observability of random processes was discussed by Pleszczyńska and Dąbrowska in [6].

In this case the problem is aimed at finding an observation strategy providing a stopping time and characterizing the form of observations of the process up to that time, which satisfies some requirements.

The aim of the paper is to describe optimal observation strategies which minimize the mean cost of observations and ensure the existence of a sufficiently good decision rule. Exact solutions in some simple cases will be presented.

2. The statement of the problem

Assume that $\{(X_n,Y_n)\}$ is a sequence of independent identically distributed bivariate random variables. The distribution of (X_1,Y_1) is known to belong to a given family $\{P_\vartheta : \vartheta \in \Theta\}$ of distributions on $(R^2, \mathcal{B}(R^2))$, where Θ is a given set of parameters. Accordingly the statistical space $(\Omega, \mathcal{A}, \mathcal{P})$ can be defined so that: $\Omega = R^2 \times R^2 \times \ldots$, $\mathcal{A} = \mathcal{B}(R^2) \otimes \mathcal{B}(R^2) \otimes \ldots$, $\mathcal{P} = \{P_\vartheta^\infty = P_\vartheta \otimes P_\vartheta \otimes \ldots, \ \vartheta \in \Theta\}$. Any sample $\omega = \{(x_n,y_n)\}$ presents current realization of the random phenomenon under investigation.

Let us suppose that there are given:

- a set of decisions D and a σ-field $\mathcal{D}$ of subsets of D,
- a loss function $L : \Theta \times D \to R^+$,
- cost functions $L_1 = R^2 \to R^+$, $L_2 : R^2 \to R^+$.

Interpretation of the functions L_1, L_2 will be given in the sequel.

Let us assume that the observation strategy is defined by a stopping time ν from some given set $\mathcal{T}$ and a sequence $(\{B_n\})$ of events from $\mathcal{A}$ so that for any realization $\omega \in \Omega$ we observe $X_1(\omega),\ldots,X_{\nu(\omega)}(\omega)$ and

$$Y_i(\omega) \quad \text{iff} \quad \omega \in B_i, \quad i = 1,\ldots,\nu(\omega).$$

Besides if at the i-th step $(X_i(\omega),Y_i(\omega)) = (x,y)$ is observed (i.e. $\omega \in B_i$) we pay $L_1(x,y)$, otherwise (i.e. $\omega \notin B_i$) we pay $L_2(x,y)$. Such a way of observation corresponds to situations in which a feature of independently investigated elements of population constitutes a bivariate random variable and the cost of observations of the second coordinate is relatively high in comparison with the first one.

Roughly speaking the problem is aimed at finding an observation strategy, which ensures the existence of a decision rule $\delta : \Omega \to D$ for which the mean loss $R(\vartheta,\delta) = E_\vartheta L(\vartheta,\delta)$ is sufficiently small and simultaneously the mean value of global cost of observations is minimal.

It is obvious that for any $n \leq \nu(\omega), \omega \in \Omega$, the observation at the n-th step is completely described by a random variable ξ_n, such that $\xi_n(\omega) = (X_n(\omega), \varepsilon_n(\omega), \varepsilon_n(\omega), Y_n(\omega))$, $\varepsilon_n(\omega) = \chi_{B_n}(\omega)$. After n steps the observation can be written as the realization of $T_n = (\xi_1, \dots, \xi_n)$. Therefore there is a natural restriction on observation strategies such that $B_1 \in \sigma(X_1)$, $B_n \in \sigma(T_{n-1}, X_n)$, $n = 2,3,\dots$, and $\mathcal{T}$ is a subset of the set of stopping times with respect to the sequence of σ-fields $\{\sigma(T_n)\}$. Thus any observation strategy is a pair $\lambda = (\nu, \{B_n\})$, where ν and $\{B_n\}$ satisfy the assumptions mentioned above. The cost function corresponding to λ has the following form

$$H(\lambda,\omega) = \sum_{n=1}^{\nu(\omega)} L_1(X_n(\omega), Y_n(\omega))\, \chi_{B_n}(\omega) + L_2(X_n(\omega), Y_n(\omega))\, \chi_{\bar{B}_n}(\omega),$$

where $\bar{B}_n = \Omega \setminus B_n$.

Assume that Δ_λ is a given set of decision rules such that $\Delta_\lambda \subset \{\delta : \Omega \to D / \delta$ is $\mathcal{A}_\nu$-measurable$\}$, where $\mathcal{A}_\nu$ denotes the σ-field of observable events, i.e.

$$\mathcal{A}_\nu = \{A \in \mathcal{A} / \forall n \; \{\omega : \nu(\omega) = n\} \cap A \in \mathcal{A}_n\},$$

where $\mathcal{A}_n = \sigma(T_n)$, $n = 1,2,\dots$.

Let us introduce denotations:

$$\Delta_\lambda(\varepsilon) = \{\delta \in \Delta_\lambda / \mathcal{R}(\vartheta, \delta) \leq \varepsilon \,, \; \vartheta \in \Theta\}$$

$$\tilde{R}(\vartheta, \lambda) = E_\vartheta H(\lambda, \cdot),$$

$$\Lambda_\varepsilon = \{\lambda = (\nu, \{B_n\}) / \exists \delta \in \Delta_\lambda(\varepsilon) \; \tilde{\mathcal{R}}(\vartheta, \lambda) < \infty \;, \; \vartheta \in \Theta\}$$

where ε is a positive real number.

For any given $\varepsilon > 0$ we shall define an optimal observation strategy $\lambda_\varepsilon^0 \in \Lambda_\varepsilon$ so that

$$\sup_{\vartheta \in \Theta} \tilde{R}(\vartheta, \lambda_\varepsilon^0) = \inf_{\lambda \in \Lambda_\varepsilon} \sup_{\vartheta \in \Theta} \tilde{R}(\vartheta, \lambda).$$

3. Examples of optimal observation strategies

Examples of optimal observation strategies given below refer to trivial stopping times, i.e. constant ones. In this case any observation strategy has the form $\lambda = (N, \{B_n\})$, $N \in \mathcal{T} = \{1,2,\dots\}$ and $\Delta_N = \{\delta = f(T_N)/f \in \mathcal{F}_{|N}\}$, where $\mathcal{F}_N$ is a given subset of the set of Borel functions from R^{3N} into D.

To obtain explicit results some simple calculations are needed. Let $\{A_n\}$ denote a sequence of Borel sets such that for the given observation strategy $\lambda = (N, \{B_n\})$ the following conditions hold:

$$B_1 = \{\omega / X_1(\omega) \in A_1\}, \quad B_n = \{\omega / (T_{n-1}(\omega), X_n(\omega)) \in A_n\}, \; n=2,3,\dots .$$

Under these conditions the cost function H takes the form

$$H(\lambda, \omega) = L_1(X_1(\omega), Y_1(\omega))\, \chi_{A_1}(X_1(\omega)) + L_2(X_1(\omega), Y_1(\omega))\, \chi_{\bar{A}_1}(X_1(\omega)) +$$
$$+ \sum_{n=2}^{N} L_1(X_n(\omega), Y_n(\omega))\, \chi_{A_n}((T_{n-1}(\omega), X_n(\omega))) +$$
$$+ L_2(X_n(\omega), Y_n(\omega))\, \chi_{\bar{A}_n}((T_{n-1}(\omega), X_n(\omega))).$$

For any $\vartheta \in \Theta$, $n = 1,2,\dots$, let us denote by $\mu_{\vartheta,n}$ the probability distribution on $(R^{3n}, \mathcal{B}(R^{3n}))$ generated by T_n. Taking into account the independency of T_n and (X_{n+1}, Y_{n+1}) we obtain that $\mu_{\vartheta,n} \bullet P_\vartheta$ is their joint probability distribution. Now the risk function $\tilde{R}$ can be written in the following form

$$\tilde{R}(\vartheta,\lambda) = \int_{A_1\times R^1} L_1(x,y)dP_\vartheta + \int_{\bar{A}_1\times R^1} L_2(x,y)dP_\vartheta +$$

$$+ \sum_{n=2}^{N}\left(\int_{A_n\times R^1} L_1(x,y)d\mu_{\vartheta,n-1}\otimes P_\vartheta + \int_{\bar{A}_n\times R^1} L_2(x,y)d\mu_{\vartheta,n-1}\otimes P_\vartheta\right).$$

Let us introduce denotations

$$r(\vartheta,A) = \int_{A\times R^1} L_1(x,y)dP_\vartheta + \int_{\bar{A}\times R^1} L_2(x,y)dP_\vartheta, \quad A\in\mathcal{B}(R^1),$$

$$A_n(t) = \left\{x\in R^1 \,\middle|\, (t,x)\in A_n\right\}, \quad t\in R^{3(n-1)}, \quad n = 2,3,\dots .$$

Due to Fubini's theorem the function R has the form

$$(3.1) \quad \tilde{R}(\vartheta,\lambda) = r(\vartheta,A_1) + \sum_{n=2}^{N}\int_{R^{3(n-1)}} r(\vartheta,A_n(t))d\mu_{\vartheta,n-1}(t).$$

P r o p o s i t i o n . Let us assume that

P1. $\left(\exists\, \vartheta_0\in\Theta,\ A^0\in\mathcal{B}(R^1)\right)\ (\forall\,\vartheta\in\Theta,\ A\in\mathcal{B}(R^1))$

$$r(\vartheta,A^0) \leqslant r(\vartheta_0,A^0) \leqslant r(\vartheta_0,A).$$

P2. $(\forall n)\ (\forall\lambda_n = (n,\{B_k\}))\ (\exists\,\delta_n\in\Delta_{\lambda_n})$

$$\inf_{\delta\in\Delta_{\lambda_n}}\sup_{\vartheta\in\Theta} R(\vartheta,\delta) = \sup_{\vartheta\in\Theta} R(\vartheta,\delta_n) = \mu_n, \quad \lim_{n\to\infty}\mu_n = 0.$$

Then for any $\varepsilon > 0$ there exists an optimal observation strategy $\lambda_{\varepsilon}^{o} = (N_{\varepsilon}^{o}, \{B_k^o\})$ such that

$$B_k^o = \{\omega \in \Omega / X_k(\omega) \in A^o\}, \quad k=1,2,\dots ,$$

$$N^o = \inf \{n \geq 1/\mu_n \leqslant \varepsilon\}.$$

P r o o f . Let $\lambda_n = (n, \{B_k\})$ be any fixed observation strategy. In view of P1., for $k = 2,3,\dots$, $t \in R^{3(k-1)}$, $A_k(t) \in \mathcal{B}(R^1)$

$$r(\vartheta, A^o) \leqslant r(\vartheta_o, A^o) \leqslant r(\vartheta, A_k(t)), \quad \vartheta \in \Theta .$$

Then the formula (3.1) and the condition P1. imply

$$\tilde{R}(\vartheta, \lambda_n^o) \leqslant \tilde{R}(\vartheta_o, \lambda_n^o) \leqslant \tilde{R}(\vartheta_o, \lambda_n), \quad \lambda_n^o = (n, \{B_o^k\}).$$

Therefore in view of P2. it suffices to note that for any $\varepsilon > 0$ and any $\lambda_n \in \Lambda_{\varepsilon}$

$$\tilde{R}(\vartheta, \lambda_{\varepsilon}^o) \leqslant \tilde{R}(\vartheta_o, \lambda_n).$$

E x a m p l e 1. Let us suppose that $\Theta = (\alpha, \beta) \subset R^1$ and for any $\vartheta \in \Theta$ the following conditions hold:

(i) there exist a density function $p_{\vartheta}(x,y)$, $(x,y) \in R^2$ of the distribution P_{ϑ} with respect to some product measure $\mu = \mu_X \otimes \mu_Y$.

(ii) the functions $\Phi_i(x,\vartheta) = \int_{R^1} L_i(x,y) p_{\vartheta}(x,y) d\mu_Y(y)$, $i=1,2$, are integrable with respect to μ_X.

(iii) there exist integrable partial derivatives $\frac{\partial \Phi_i(x,\vartheta)}{\partial \vartheta}$, $x \in R^1$, for $i = 1,2$, and the following equalities are fulfilled:

$$\frac{d}{d\vartheta}\int_A \Phi_i(x,\vartheta)\,d\mu_X(x) = \int_A \frac{\partial\Phi_i(x,\vartheta)}{\partial\vartheta}\,d\mu_X(x), \quad A\in\mathfrak{B}(R^1).$$

(iv) there exists $\vartheta_0 \in \Theta$ such that

$$\int_{A^0} \frac{\partial\Phi_1(x,\vartheta)}{\partial\vartheta}\,d\mu_X(x) = -\int_{\bar{A}^0} \frac{\partial\Phi_2(x,\vartheta)}{\partial\vartheta}\,d\mu_X(x),$$

where $A^0 = \left\{x \in R^1/\Phi_1(x,\vartheta_0) \leqslant \Phi_2(x,\vartheta_0)\right\}$.

Then the condition P1. of the proposition is satisfied.

P r o o f . According to (ii) the function r has the form

$$(3.2) \quad r(\vartheta,A) = \int_A \Phi_1(x,\vartheta)\,d\mu_X(x) + \int_{\bar{A}} \Phi_2(x,\vartheta)\,d\mu_X(x), \quad A\in\mathfrak{B}(R^1)$$

Due to the definition of A^0

$$(3.3) \qquad r(\vartheta_0,A^0) = \min_{A\in\mathfrak{B}(R^1)} r(\vartheta_0,A).$$

Besides in view of the conditions (iii), (iv)

$$(3.4) \qquad \frac{dr(\vartheta,A^0)}{d\vartheta} = 0,$$

for any $\vartheta\in\Theta$. It follows that $r(\vartheta,A^0)$ does not depend on ϑ. Taking into account (3.3) and (3.4) and applying the usual reasoning in classical discrimination problems we state that the condition P1. is fulfilled.

E x a m p l e 2. Assume that $\Theta = (0,1)$ and for any $\vartheta\in\Theta$ the distribution P_ϑ is defined as

$$P_{\vartheta}(A\times\{1\}) = \int_A \vartheta f_1(x)dx, \quad P_{\vartheta}(A\times\{0\}) = \int_A (1-\vartheta)f_2(x)dx,$$

$$P_{\vartheta}(R^1\times\{0,1\}) = 1, \quad A\in\mathcal{B}(R^1),$$

where f_1, f_2 are fixed density functions with respect to the Lebesgue measure, not proportional over a set of a positive measure.

Besides suppose that

$$L_1(x,y) = \begin{cases} 0, & y=0 \\ L_1, & y=1 \end{cases}, \quad L_2(x,y) = \begin{cases} L_2, & y=0 \\ 0, & y=1 \end{cases}.$$

Let us restrict ourselves to the strategies $\lambda_n = (n, \{B_k\})$, where $B_k = \{\omega\in\Omega / X_k(\omega)\in A\}$, $A\in\mathcal{B}(R^1)$, $A\neq\phi$, $k = 1,2,\dots$, $n = 1,2,\dots$. Let us assume that the loss function L depends additionally on λ_n in the following way:

$$L(\vartheta,d,\lambda_n) = \frac{\gamma_A}{\vartheta(1-\vartheta\gamma_A)}(\vartheta-d)^2, \quad \gamma_A = \int_A f_1(x)dx, \quad \vartheta\in\Theta, \ d\in\Theta \quad ,$$

and let

$$\Delta_{\lambda_n} = \left\{\delta \,|\, \delta = \sum_{i=1}^{n} \alpha_i\varepsilon_i y_i, \quad \varepsilon_i = \chi_A(x_i), \quad \alpha_i\in R^1, \quad E_{\vartheta}\delta = \vartheta\right\},$$

$$R(\vartheta,\delta,\lambda_n) = E_{\vartheta}L(\vartheta,\delta,\lambda_n), \quad \vartheta\in\Theta, \quad \delta\in\Delta_{\lambda_n}.$$

Under the above assumptions the conditions P1., P2. of the Proposition are fulfilled.

P r o o f : The form of P_{ϑ} implies that

$$r(\vartheta,A) = \int_A \vartheta L_1 f_1(x)dx + \int_{\bar{A}} (1-\vartheta)\, L_2 f_2(x)dx.$$

Now following the discrimination problem we state that there exist a positive constant C such that $A^o = \{x/C\, L_1 f_1(x) \leqslant \leqslant L_2 f_2(x)\}$ and $\vartheta_0 = C(1+C)^{-1}$ satisfy P1.

It is obvious that for any λ_n $\{\varepsilon_k Y_k\}$ is the sequence of i.i.d. random variables with $E\,(\varepsilon_1 Y_1) = \vartheta\,\gamma_A$, $D^2_\vartheta(\varepsilon_1 Y_1) = = \vartheta\,\gamma_A(1-\vartheta\gamma_A)$. The Gauss Markov theorem implies that

$$\inf_{\delta \in \Delta_{\lambda_n}} E_\vartheta(\vartheta-\delta)^2 = E_\vartheta(\vartheta-\hat{\vartheta}_n)^2, \quad \vartheta \in \Theta$$

where

$$\hat{\vartheta}_n = (n\,\gamma_A)^{-1} \sum_{i=1}^{n} \varepsilon_i Y_i .$$

Hence

$$\inf_{\delta \in \Delta_\lambda} \sup_{\vartheta\in\Theta} R(\vartheta,\delta,\lambda_n) = \sup_{\vartheta\in\Theta} R(\vartheta,\hat{\vartheta}_n,\lambda_n) = n^{-1}.$$

The above equalities imply the condition P2.

Acknowledgement. I wish to thank E. Pleszczyńska and D. Dąbrowska for useful comments and discussions.

REFERENCES

[1] T.W. Anderson: Introduction to multivariate statistical analysis, New York 1958.

[2] M.H. DeGroot: Optimal statistical decisions, New York 1970.

[3] T.S. Ferguson: Mathematical statistics. A decision theoretic approach. New York 1967.

[4] A.W. Marshall, I. Olkin: A general approach to some screening and classification problems, J. Roy. Stat. Soc. Ser. B30 (1968) 407-444.

[5] M. M u s i e l a : On sequential estimation of parameters of continuous Gaussian Markov processes, Institute of Math., Polish Academy of Sciences, Preprint No. 188 (1979).

[6] E. P l e s z c z y ń s k a , D. D ą b r ó w s k a : On partial observability in statistical models, Math. Operationsforsch, Statist. 11 (1980) 49-59.

INSTITUTE OF MATHEMATICS, TECHNICAL UNIVERSITY OF WARSAW
Received May 6, 1981.