MATHEMATICS

R. L. STRATONOVICH

1961

SovietRxiv

Abstract

Full Text

MATHEMATICS

R. L. STRATONOVICH

CONDITIONAL MARKOV PROCESSES IN PROBLEMS OF MATHEMATICAL STATISTICS AND DYNAMIC PROGRAMMING

(Presented by Academician A. N. Kolmogorov on 19 V 1961)

The theory of conditional Markov processes ([1]) makes it possible to obtain exact or approximate solutions of a large number of problems in mathematical statistics and dynamic programming. Problems of this class are distinguished by the fact that they contain a sequential quality criterion and a sequentially proceeding Markov process (independent trials are a special case).

Let a parameter $t \in T$ be given, having the meaning of time, discrete or continuous. For simplicity of exposition, restricting ourselves to the first case, we shall assume $T = \{t_k : k = 0, 1, 2, ...\}$, $t_k - t_{k-1} = \Delta$, $t_0 = 0$. At each time $t \in T$ there is the possibility of making a controlling decision $u_t \in U_t$. In the general case the possibilities of making one or another decision are limited by the preceding decisions $u_0^{t-\Delta}$ (here and below $u_b^a = \{u_\tau : a \le \tau \le b,\ \tau \in T\}$ and, moreover, $u = \{u_\tau : \tau \in T\}$). We shall assume the following Markov property of the restrictions: the restriction of the choice at time t depends only on the decision $u_{t-\Delta}$ made at the preceding time. Therefore $u_t \in U_t(u_0^{t-\Delta}) = U_t(u_{t-\Delta})$.

Next, let to each control u_t there correspond some phase space $E_t(u_t)$ with points ξ_t. The pair (u_t, ξ_t) will be denoted by ξ_t. For a given control u the process $\{\xi_t\}$ is random and is described by a probability measure $P_u(d\xi)$, which has the obvious property: $P_u(d\xi_a^b)$ also does not depend on $u_{b+\Delta}^\infty$. The theory is effective under the assumption that ζ (for any admissible functions u) forms a Markov process with given transition probabilities

$$P_u(d\xi_{t+\Delta} \mid \xi_t) = P_{u_t, t+\Delta}(d\xi_{t+\Delta} \mid \xi_t) \equiv P(d\xi_{t+\Delta} \mid \xi_t);$$

$$P_u(d\xi_{t+\Delta}^T \mid \xi_0^t) = P_{u_t, T}(d\xi_{t+\Delta}^T \mid \xi_t) \equiv P(d\xi_{t+\Delta}^T \mid \xi_t). \tag{1}$$

It is required to find a decision rule

$$u_t = \delta_t(x) = \delta_t(x_0^{t-\Delta}) \in U_t(u_0^{t-\Delta}), \tag{2}$$

adopted on the basis of the observed values $x_0^{t-\Delta}$, each of which is a known function

$$x_t = f_t(\xi_t, u_t) \tag{3}$$

of the point ξ_t corresponding to the same time instant.

A loss function $F_t(\xi_t)$ is specified, and the desired decision (2) must be optimal in the sense that it must correspond to the minimal inte-

mean loss. The measure $P_u(d\xi)$ and the decisions (2) generate the measure

$$\begin{aligned} P_{\delta(x)}(dx, d\xi) = P_{\delta_0}(dx_0)\, P_{\delta(x_0)}(dx_\Delta \mid x_0)\, P_{\delta(x_\Delta)}(dx_{2\Delta} \mid x_0^\Delta) \ldots \\ \ldots P_{\delta(x_0^t)}(dx_{t+\Delta} \mid x_0^t) \ldots P_{\delta(x_0^{T-\Delta})}(dx_T, d\xi \mid x_0^{T-\Delta}), \end{aligned} \tag{4}$$

where $\delta(x_0^t) = \{\delta_{\tau+\Delta}(x_0^\tau) : 0 \le \tau \le t\}$. Therefore the mean loss can be written as

$$R = \min_\delta \int P_{\delta(x)}(dx, d\xi) \int_0^T d\tau\, F_\tau(\xi_\tau, \delta_\tau(x_0^\tau)), \tag{5}$$

where the minimum is sought over the class of all admissible decisions.

Using the fact that

$$P_{\delta(x_0^t)}(dx_{t+\Delta} \mid x_0^t) = P_{\delta_0^{t+\Delta}}(dx_{t+\Delta} \mid x_0^t)$$

does not depend on $\delta_{t+2\Delta}^T$, we write (5) in the form

$$\begin{aligned} R = \min_{\delta_0} \Bigg\{ \int P_{\delta_0}(dx_0) \min_{\delta_\Delta} P_{\delta_0^\Delta}(dx_\Delta \mid x_0) \ldots \min_{\delta_{t+\Delta}} P_{\delta_0^{t+\Delta}}(dx_{t+\Delta} \mid x_0^t) \ldots \\ \ldots \min_{\delta_T} P_{\delta_0^T}(dx_T, d\xi \mid x_0^{T-\Delta}) \int_0^T d\tau\, F_\tau(\xi_\tau, \delta_\tau(x_0^\tau)) \Bigg\} \end{aligned} \tag{6}$$

$$(\delta_{t+\Delta} = U_t(\delta_0^t)).$$

We shall carry out the minimization successively, beginning from the right. Introduce the function

$$\begin{aligned} S(t \mid x_0^t, \delta_0^t) = \min_{\delta_{t+\Delta}} \Bigg\{ \int P_{\delta_0^{t+\Delta}}(dx_{t+\Delta} \mid x_0^t) \ldots \\ \ldots \min_{\delta_T} P_{\delta_0^T}(dx_T, d\xi \mid x_0^{T-\Delta}) \int_{t+\Delta}^T d\tau\, F_\tau(\xi_\tau, \delta_\tau) \Bigg\} \end{aligned}$$

$$= \min_{\delta_{t+\Delta}^T} \int P_\delta(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid x_0^t) \int_{t+\Delta}^T d\tau\, F_\tau(\xi_\tau, \delta_\tau) \tag{7}$$

(the decisions $\delta_{t+\Delta}^T$ being compared are compatible with δ_0^t).

It is easy to see that the minimal mean loss (6) coincides with

$$S(-\Delta) = \min_{\delta_0} \int S(0 \mid x_0, \delta_0)\, P(dx_0),$$

and therefore the problem of minimizing the loss is reduced to the problem of finding the indicated function. Let us derive recurrence relations for it. Splitting the integral over τ in (7) into the sum of the terms

$$F_{t+\Delta}(\xi_{t+\Delta}, \delta_{t+\Delta})\, \Delta + \int_{t+2\Delta}^T d\tau\, F_\tau(\xi_\tau, \delta_\tau)$$

and taking into account that the first term does not depend on $\xi_{t+2\Delta}^T, \delta_{t+2\Delta}^T$ (and therefore is easily averaged and minimized), while the second term, under minimization with respect to $\delta_{t+2\Delta}^T$ and averaging, leads to $S(t+\Delta \mid x_0^{t+\Delta}, \delta_0^{t+\Delta})$, we obtain

$$\begin{aligned} S(t \mid x_0^t, \delta_0^t) = \min_{\delta_{t+\Delta}} \Bigg[\int P_{\delta_0^{t+\Delta}}(d\xi_{t+\Delta} \mid x_0^t) F_{t+\Delta}(\xi_{t+\Delta}, \delta_{t+\Delta})\, \Delta + \\ + \int P_{\delta_0^{t+\Delta}}(dx_{t+\Delta} \mid x_0^t)\, S(t+\Delta \mid x_0^{t+\Delta}, \delta_0^{t+\Delta}) \Bigg]. \end{aligned} \tag{8}$$

The solution $u_{t+\Delta} = \delta_{t+\Delta} \in U(\delta_0^t)$, corresponding to the minimum of this expression, depends only on x_0^t, δ_0^t and is the desired optimal solution (2).

Until now we have not used the Markov properties of the processes ξ, u. Let us consider the simplifications following from them. Transforming the measure occurring in (7) to the form

$$\begin{aligned} P_u\left(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid x_0^t\right) = \int_{E_t} P_u\left(dx_{t+\Delta}^T, d\xi_t^T \mid x_0^t\right) = \\ = \int_{E_t} P_u\left(d\xi_t \mid x_0^t\right) P_u\left(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid \xi_t, x_0^t\right), \end{aligned} \tag{9}$$

we note that

$$P_u\left(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid \xi_t, x_0^t\right) = P_{u_t^T}\left(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid \xi_t\right)$$

does not depend on $x_0^t, u_0^{t-\Delta}$ by virtue of (1). Therefore the function (7) can be written as

$$S\left(t \mid x_0^t, \delta_0^t\right) = \int_{E_t} P_{\delta_0^t}\left(d\xi_t \mid x_0^t\right) S\left(t \mid \xi_t, \delta_t\right), \tag{10}$$

where

$$S\left(t \mid \xi_t, u_t\right) = \min_{u_{t+\Delta}^T} \int P_{u_t^T}\left(dx_{t+\Delta}^T, d\xi_{t+\Delta}^T \mid \xi_t\right) \int_{t+\Delta}^T d\tau\, F_\tau(\xi_\tau, u_\tau).$$

Hence it is seen that the function

$$S\left(t \mid x_0^t, u_0^t\right) = S\left(t, u_t, W_t(d\xi_t)\right) \tag{11}$$

depends on $x_0^t, u_0^{t-\Delta}$ only through the posterior distribution $W_t(d\xi_t) = P_{u_0^t}(d\xi_t \mid x_0^t)$.

Equation (8) takes the form

$$S\left(t, u_t, W_t(d\xi_t)\right) = \tag{12}$$

$$= \min_{u_{t+\Delta} \in U_t(u_t)} \left[M_{ps} F_{t+\Delta}(\xi_{t+\Delta}, u_{t+\Delta})\Delta + M_{ps} S\left(t+\Delta, u_{t+\Delta}, W_{t+\Delta}(d\xi_{t+\Delta})\right)\right].$$

Here M_{ps} is the symbol of posterior averaging corresponding to the distribution $W_t(d\xi_t)$. In the present case it denotes the operation of integration with weights

$$P_{\delta_0^{t+\Delta}}\left(d\xi_{t+\Delta} \mid x_0^t\right) = \int_{E_t} P_{\delta_t^{t+\Delta}}\left(d\xi_{t+\Delta} \mid \xi_t\right) W_t(d\xi_t);$$

$$P_{\delta_0^{t+\Delta}}\left(dx_{t+\Delta} \mid x_0^t\right) = \int_{E_t} P_{\delta_t^{t+\Delta}}\left(dx_{t+\Delta} \mid \xi_t\right) W_t(d\xi_t). \tag{13}$$

The last equalities are derived analogously to (9). If there are no constraints on u ($U_t(u_{t-\Delta})$ does not depend on $u_{t-\Delta}$), then the dependence in (11) on u_t disappears.

The behavior of the posterior probabilities $W_t(d\xi_t)$ has been studied in the theory of conditional Markov processes. The dependence of $P_u(d\xi)$ on y which occurs in the present case introduces nothing new and does not require a generalization of the theory. The results of the theory of conditional Markov processes make it possible to solve equation (12) and find the function $S(t, u_t, W_t)$, and consequently also the optimal solutions $u_t = d_t(x)$.

Having set out the general formulation of the problem and the method of its solution, we proceed to a survey of various special cases pertaining to mathematical statistics and dynamic programming.

1. **Optimal filtering.** In this case, both the Markov transition probability (1) and the observed data (3) do not depend on the decision $u_t = d_t$. The Markov process $\zeta = (x, y)$ consists of observed (x) and unobserved (y) components. The quality function $F_t(x_t, y_t, u_t)$ most often depends only on y_t and u_t. Constraints $u_t \in U_t(u_{t-\Delta})$ are usually absent. An important special case is the estimation of parameters, when the function $y_t = y_0$ is constant.

2. **Sequential Wald analysis**(2). The course of the process $\{\zeta_t\}$ does not depend on u. The decision $u_t = (D_t, z_t)$ splits into the decision to continue observation $(D_t = 1)$ or to terminate it $(D_t = 0)$, and into the estimate decision z_t. The loss function depends on u_t: $F_t(\xi_t) = F_t(y_t, z_t, D_t)$. The observed data (3) depend on D_t, but do not depend on z_t. There are certain Markov constraints on u_t: observation cannot be resumed if it has been terminated.

3. **Controlled observation in the sense of Kolmogorov.** The process $\{\xi_t\}$ does not depend on u. The information obtained $f_t(\zeta_t, u_t)$ depends essentially on u_t. In other respects various variants are possible.

4. **Dynamic programming**(3). The observed process $x_t = f_t(\zeta_t)$ explicitly does not depend on u. The basic dynamic process z_t depends essentially on u; for example, it is described by the equation

$$\frac{dz_t}{dt} = \varphi(z_t, \eta_t, u_t)$$

(η_t is a random disturbance). The quality function of the control $F_t(\zeta_t, u_t)$ often does not explicitly depend on u_t. Consider two principal special cases.

a) Suppose it is required to track a known function y_t with quality function $F_t(z_t) = \Phi_t(y_t, z_t)$. The process $x_t = \tilde{z}_t$, which is in some way connected with z_t, is observed. The theory set out above is applicable if the processes $\zeta = (z, \tilde{z}, \eta)$ (possibly augmented by some other processes) form, in their totality, a Markov process.

b) It is required to track a random process y_t, so that $F_t(y_t, z_t) = \Phi_t(y_t, z_t)$. In addition to $\tilde{z}$, the operator has at his disposal the process y or a process $\tilde{y}$ connected with it. In this case $x = (\tilde{y}, \tilde{z})$, and one may set $\zeta = (z, \tilde{z}, \eta, y, \tilde{y})$. Of course, in some cases η_t may be absent, y and $\tilde{y}$ may coincide, etc.

5. Less studied are those cases in which both $P_u(d\xi_{t+\Delta} \mid \xi_t)$ and $f_t(\zeta_t, u_t)$ depend essentially on u, in which constraints on u are important, and so on.

Moscow State University
named after M. V. Lomonosov

Received
12 V 1961

REFERENCES CITED

1. R. L. Stratonovich, *Theory of Probability and Its Applications*, **5**, 172 (1960).

2. D. Blackwell, M. A. Girshick, *Theory of Games and Statistical Decisions*, IL, 1958.

3. R. Bellman, *Dynamic Programming*, IL, 1960.

Note: Figure translations are in progress. See original paper for figures.