KYBERNETIKA — VOLUME *46* (2010), NUMBER 4, PAGES 754–770

OPTIMAL SEQUENTIAL PROCEDURES WITH BAYES DECISION RULES

Andrey Novikov

In this article, a general problem of sequential statistical inference for general discrete-time stochastic processes is considered. The problem is to minimize an average sample number given that Bayesian risk due to incorrect decision does not exceed some given bound. We characterize the form of optimal sequential stopping rules in this problem. In particular, we have a characterization of the form of optimal sequential decision procedures when the Bayesian risk includes both the loss due to incorrect decision and the cost of observations.

Keywords: sequential analysis, discrete-time stochastic process, dependent observations, statistical decision problem, Bayes decision, randomized stopping time, optimal stopping rule, existence and uniqueness of optimal sequential decision procedure

Classification: 62L10, 62L15, 62C10, 60G40

1. INTRODUCTION

Let $X_1, X_2, \dots, X_n, \dots$ be a discrete-time stochastic process, whose distribution depends on an unknown parameter θ, $\theta \in \Theta$. In this article, we consider a general problem of sequential statistical decision making based on the observations of this process.

Let us suppose that for any $n = 1, 2, \dots$, the vector $(X_1, X_2, \dots, X_n)$ has a probability "density" function

$$f_\theta^n = f_\theta^n(x_1, x_2, \dots, x_n) \tag{1}$$

(Radon–Nikodym derivative of its distribution) with respect to a product-measure

$$\mu^n = \mu \otimes \mu \otimes \cdots \otimes \mu,$$

with some σ-finite measure μ on the respective space. As usual in the Bayesian context, we suppose that $f_\theta^n(x_1, x_2, \dots, x_n)$ is measurable with respect to $(\theta, x_1, \dots, x_n)$, for any $n = 1, 2, \dots$.

Let us define a *sequential statistical procedure* as a pair (ψ, δ), being ψ a (randomized) *stopping rule*,

$$\psi = (\psi_1, \psi_2, \dots, \psi_n, \dots),$$

and δ a *decision rule*

$$\delta = (\delta_1, \delta_2, \dots, \delta_n, \dots)\,,$$

supposing that

$$\psi_n = \psi_n(x_1, x_2, \dots, x_n)$$

and

$$\delta_n = \delta_n(x_1, x_2, \dots, x_n)$$

are measurable functions, $\psi_n(x_1, \dots, x_n) \in [0,1]$, $\delta_n(x_1, \dots, x_n) \in \mathscr{D}$ (a decision space), for any observation vector $(x_1, \dots, x_n)$, for any $n = 1, 2, \dots$ (see, for example, [1, 7, 8, 9, 21]).

The interpretation of these elements is as follows.

The value of $\psi_n(x_1, \dots, x_n)$ is interpreted as the conditional probability *to stop and proceed to decision making*, given that that we came to stage n of the experiment and that the observations up to stage n were $(x_1, x_2, \dots, x_n)$. If there is no stop, the experiment continues to the next stage and an additional observation x_{n+1} is taken. Then the rule ψ_{n+1} is applied to $x_1, \dots, x_n, x_{n+1}$ in the same way as above, etc., until the experiment eventually stops.

When the experiments stops at stage n, being $(x_1, \dots, x_n)$ the data vector observed, the decision specified by $\delta_n(x_1, \dots, x_n)$ is taken, and the sequential statistical experiment stops.

The stopping rule ψ generates, by the above process, a random variable τ_ψ (randomized *stopping time*), which may be defined as follows. Let $U_1, U_2 \dots, U_n, \dots$ be a sequence of independent and identically distributed (i.i.d.) random variables uniformly distributed on $[0,1]$ (randomization variables), such that the process $(U_1, U_2, \dots)$ is independent of the process of observations $(X_1, X_2, \dots)$. Then let us say that $\tau_\psi = n$ if, and only if,

$$U_1 > \psi_1(X_1),\ \dots,\ U_{n-1} > \psi_{n-1}(X_1, \dots, X_{n-1}),\ \text{and}\ U_n \le \psi_n(X_1, \dots X_n),$$

$n = 1, 2, \dots$.

It is easy to see that the distribution of τ_ψ is given by

$$P_\theta(\tau_\psi = n) = E_\theta(1 - \psi_1)(1 - \psi_2) \dots (1 - \psi_{n-1})\psi_n, \quad n = 1, 2, \dots. \tag{2}$$

In (2), ψ_n stands for $\psi_n(X_1, \dots, X_n)$, unlike its previous definition as $\psi_n = \psi_n(x_1, \dots, x_n)$. We use this "duality" throughout the paper, applying, for any $F_n = F_n(x_1, \dots, x_n)$ or $F_n = F_n(X_1, \dots X_n)$ the following general rule: when F_n is under the probability or expectation sign, it is $F_n(X_1, \dots, X_n)$, otherwise it is $F_n(x_1, \dots, x_n)$.

Let $w(\theta, d)$ be a non-negative loss function (measurable with respect to (θ, d), $\theta \in \Theta$, $d \in \mathscr{D}$) and π_1 any probability measure on Θ. We define the average loss of the sequential statistical procedure (ψ, δ) as

$$W(\psi, \delta) = \sum_{n=1}^{\infty} \int [E_\theta (1 - \psi_1) \dots (1 - \psi_{n-1}) \psi_n w(\theta, \delta_n)]\, \mathrm{d}\pi_1(\theta). \tag{3}$$

and its *average sample number*, given θ, as

$$N(\theta; \psi) = E_\theta \tau_\psi \tag{4}$$

(we suppose that $N(\theta;\psi)=\infty$ if $\sum_{n=1}^{\infty}P_\theta(\tau_\psi=n)<1$ in (2)).

Let us also define its "weighted" value

$$N(\psi)=\int N(\theta;\psi)\,\mathrm{d}\pi_2(\theta), \tag{5}$$

where π_2 is some probability measure on Θ, giving "weights" to the particular values of θ.

Our main goal is minimizing $N(\psi)$ over all sequential decision procedures (ψ,δ) subject to

$$W(\psi,\delta)\le w, \tag{6}$$

where w is some positive constant, supposing that π_1 in (3) and π_2 in (5) are, generally speaking, two *different* probability measures. We only consider the cases when there exist procedures (ψ,δ) satisfying (6).

Sometimes it is necessary to put the risk under control in a more detailed way. Let $\Theta_1,\dots,\Theta_k$ be some subsets of the parametric space such that $\Theta_i\bigcap\Theta_j=\emptyset$ if $i\neq j$, $i,j=1,\dots,k$. Then, instead of (6), we may want to guarantee that

$$W_i(\psi,\delta)=\sum_{n=1}^{\infty}\int_{\Theta_i}E_\theta(1-\psi_1)\dots(1-\psi_{n-1})\psi_n w(\theta,\delta_n)\,\mathrm{d}\pi_1(\theta)\le w_i, \tag{7}$$

with some $w_i>0$, for any $i=1,\dots,k$, when minimizing $N(\psi)$.

To advocate restricting the sequential procedures by (7), let us see a particular case of hypothesis testing.

Let $H_1:\theta=\theta_1$ and $H_2:\theta=\theta_2$ be two simple hypotheses about the parameter value, and let

$$w(\theta,d)=\begin{cases}1 & \text{if}\quad \theta=\theta_1 \text{ and } d=2,\\ 1 & \text{if}\quad \theta=\theta_2\ \text{ and } d=1,\\ 0 & \text{otherwise,}\end{cases}$$

and $\pi_1(\{\theta_1\})=\pi$, $\pi_1(\{\theta_2\})=1-\pi$, with some $0<\pi<1$. Then, letting $\Theta_i=\{\theta_i\}$, $i=1,2$, in (7), we have that

$$W_1(\psi,\delta)=\pi P_{\theta_1}(\text{ reject } H_1)=\pi\alpha(\psi,\delta)$$

and

$$W_2(\psi,\delta)=(1-\pi)P_{\theta_2}(\text{accept } H_1)=(1-\pi)\beta(\psi,\delta),$$

where $\alpha(\psi,\delta)$ and $\beta(\psi,\delta)$ are the type I and type II error probabilities. Thus, taking in (7) $w_1=\pi\alpha$, $w_2=(1-\pi)\beta$, with some $\alpha,\beta\in(0,1)$, we see that (7) is equivalent to

$$\alpha(\psi,\delta)\le\alpha,\quad \text{and}\quad \beta(\psi,\delta)\le\beta. \tag{8}$$

Let now $\pi_2(\{\theta_0\})=1$ and suppose that the observations are i.i.d. Then our problem of minimizing $N(\psi)=N(\theta_0;\psi)$ under restrictions (8) is the classical Wald and Wolfowitz problem of minimizing the expected sample size (see [22]). It is well known that its solution is given by the sequential probability ratio test (SPRT), and that it minimizes the expected sample size under the alternative hypothesis as well (see [12, 22]).

On the other hand, if $\pi_2(\{\theta\}) = 1$ with $\theta \neq \theta_0$ and $\theta \neq \theta_1$, we have the problem known as the modified Kiefer–Weiss problem, the problem of minimizing the expected sample size, under θ, among all sequential tests subject to (8) (see [10, 23]). The general structure of the optimal sequential test in this problem is given by Lorden [12] for i.i.d. observations.

So, we see that considering natural particular cases of sequential procedures subject to (7) and using different choices of π_1 in (3) and π_2 in (5) we extend known problems for i.i.d. observations to the case of general discrete-time stochastic processes.

The method we use in this article was originally developed for testing of two hypotheses [17], then extended for multiple hypothesis testing problems [15], and to composite hypothesis testing [18]. An extension of the same method for hypothesis testing problems when control variables are present can be found in [14].

A more general, than used in this article, setting for Bayes-type decision problems, where both the cost of observations and the loss functions depend on the true value of the parameter and on the observations, is considered in [16].

From this time on, our aim will be minimizing $N(\psi)$, defined by (5), in the class of sequential statistical procedures subject to (7).

In Section 2, we reduce the problem to an optimal stopping problem. In Section 3, we give a solution to the optimal stopping problems in the class of truncated stopping rules, and in Section 4 in some natural class of non-truncated stopping rules. In particular, in Section 4 we give a solution to the problem of minimizing $N(\psi)$ in the class of all statistical procedures satisfying $W_i(\psi, \delta) \leq w_i$, $i = 1, \dots, k$ (see Remark 4.10).

2. REDUCTION TO AN OPTIMAL STOPPING PROBLEM

In this section, the problem of minimizing the average sample number (5) over all sequential procedures subject to (7) will be reduced to an optimal stopping problem. This is a usual treatment of conditional problems in sequential hypothesis testing (see, for example, [2, 3, 12, 13, 19]). We will use the same ideas to treat the general statistical decision problem described above.

Let us define the following Lagrange-multiplier function:

$$L(\psi, \delta) = L(\psi, \delta; \lambda_1, \dots, \lambda_k) = N(\psi) + \sum_{i=1}^{k} \lambda_i W_i(\psi, \delta) \tag{9}$$

where $\lambda_i \geq 0$, $i = 1, \dots, k$ are some constant multipliers.

Let Δ be a class of sequential statistical procedures.

The following Theorem is a direct application of the method of Lagrange multipliers to the above optimization problem.

Theorem 2.1. Let there exist $\lambda_i > 0$, $i = 1, \dots, k$, and a procedure $(\psi^*, \delta^*) \in \Delta$ such that for any procedure $(\psi, \delta) \in \Delta$

$$L(\psi^*, \delta^*; \lambda_1, \dots, \lambda_k) \leq L(\psi, \delta; \lambda_1, \dots, \lambda_k) \tag{10}$$

holds and such that

$$W_i(\psi^*,\delta^*) = w_i, \quad i = 1,\dots k. \tag{11}$$

Then for any test $(\psi,\delta) \in \Delta$ satisfying

$$W_i(\psi,\delta) \leq w_i, \quad i = 1,2,\dots,k, \tag{12}$$

it holds

$$N(\psi^*) \leq N(\psi). \tag{13}$$

The inequality in (13) is strict if at least one of the inequalities (12) is strict.

Proof. Let $(\psi,\delta) \in \Delta$ be any procedure satisfying (12). Because of (10),

$$L(\psi^*,\delta^*;\lambda_1,\dots,\lambda_k) = N(\psi^*) + \sum_{i=1}^{k} \lambda_i W_i(\psi^*,\delta^*) \leq L(\psi,\delta;\lambda_1,\dots,\lambda_k) \tag{14}$$

$$= N(\psi) + \sum_{i=1}^{k} \lambda_i W_i(\psi,\delta) \leq N(\psi) + \sum_{i=1}^{k} \lambda_i w_i, \tag{15}$$

where to get the last inequality we used (12). Taking into account conditions (11) we get from this that

$$N(\psi^*) \leq N(\psi).$$

To get the last statement of the theorem we note that if $N(\psi^*) = N(\psi)$ then there are equalities in (14) – (15) instead of the inequalities, which is only possible if $W_i(\psi,\phi) = w_i$ for any $i = 1,\dots,k$. □

Remark 2.2. It is easy to see that defining a new loss function $w'(\theta,d)$ which is equal to $\lambda_i w(\theta,d)$ whenever $\theta \in \Theta_i, \quad i = 1,\dots,k$, we have that the weighted average loss $W(\psi,\delta)$ defined by (3) with $w(\theta,d) = w'(\theta,d)$ coincides with the second summand in (9).

Because of this, we treat in what follows only the case of one summand ($k = 1$) in (9), being the Lagrange-multiplier function defined as

$$L(\psi,\delta;\lambda) = N(\psi) + \lambda W(\psi,\delta). \tag{16}$$

It is obvious that the problem of minimization of (16) is equivalent to that of minimization of

$$R(\psi,\delta;c) = cN(\psi) + W(\psi,\delta), \tag{17}$$

where $c > 0$ is any constant, and, in the rest of the article, we will solve the problem of minimizing (17), instead of (16). This is because the problem of minimization of (17) is interesting by itself, without its relation to the conditional problem above. For example, if $\pi_2 = \pi_1 = \pi$, it is easy to see that it is equivalent to the problem of Bayesian sequential decision-making, with the prior distribution π and a fixed cost c per observation. The latter set-up is fundamental in the sequential analysis (see [7, 8, 9, 21, 24], among many others).

Because of Theorem 2.1, from this time on, our main focus will be on the unrestricted minimization of $R(\psi,\delta;c)$, over all sequential decision procedures.

Let us suppose, additionally to the assumptions of Introduction, that for any $n=1,2\dots$ there exists a decision function $\delta_n^B=\delta_n^B(x_1,\dots,x_n)$ such that for any $d\in\mathscr{D}$

$$\int w(\theta,d)f_\theta^n(x_1,\dots,x_n)\,\mathrm{d}\pi_1(\theta)\geq\int w(\theta,\delta_n^B(x_1,\dots,x_n))f_\theta^n(x_1,\dots,x_n)\,\mathrm{d}\pi_1(\theta) \tag{18}$$

for μ^n-almost all $(x_1,\dots,x_n)$. Then δ_n^B is called the Bayesian decision function based on n observations. We do not discuss in this article the questions of the existence of Bayesian decision functions, we just suppose that they exist for any $n=1,2,\dots$ referring, e. g., to [21] for an extensive underlying theory.

Let us denote by $l_n=l_n(x_1,\dots,x_n)$ the right-hand side of (18). It easily follows from (18) that

$$\int l_n d\mu^n=\inf_{\delta_n}\int E_\theta w(\theta,\delta_n)\,\mathrm{d}\pi_1(\theta), \tag{19}$$

thus

$$\int l_1\,\mathrm{d}\mu^1\geq\int l_2\,\mathrm{d}\mu^2\geq\dots.$$

Because of that, we suppose that

$$\int l_1(x)\,\mathrm{d}\mu(x)<\infty$$

which makes all the Bayesian risks (19) finite, for any $n=1,2,\dots$.

Let $\delta^B=(\delta_1^B,\delta_2^B,\dots)$. The following Theorem shows that the only decision rules worth our attention are the Bayesian ones. Its “if”-part is, in essence, Theorem 5.2.1 [9].

Let for any $n=1,2,\dots$ and for any stopping rule ψ

$$s_n^\psi=(1-\psi_1)\dots(1-\psi_{n-1})\psi_n,$$

and let

$$S_n^\psi=\{(x_1,\dots,x_n):s_n^\psi(x_1,\dots,x_n)>0\}$$

for all $n=1,2,\dots$.

Theorem 2.3. For any sequential procedure (ψ,δ)

$$W(\psi,\delta)\geq W(\psi,\delta^B)=\sum_{n=1}^\infty\int s_n^\psi l_n\,\mathrm{d}\mu^n. \tag{20}$$

Supposing that the right-hand side of (20) is finite, the equality in (20) is only possible if

$$\int w(\theta,\delta_n)f_\theta^n\,\mathrm{d}\pi_1(\theta)=\int w(\theta,\delta_n^B)f_\theta^n\,\mathrm{d}\pi_1(\theta)$$

μ^n-almost everywhere on S_n^ψ for all $n=1,2,\dots$.

P r o o f. It is easy to see that $W(\psi,\delta)$ on the left-hand side of (20) has the following equivalent form:

$$W(\psi,\delta)=\sum_{n=1}^{\infty}\int s_n^{\psi}\int w(\theta,\delta_n)f_\theta^n\,\mathrm{d}\pi_1(\theta)\,\mathrm{d}\mu^n. \tag{21}$$

Applying (18) under the integral sign in each summand in (21) we immediately have:

$$W(\psi,\delta)\geq\sum_{n=1}^{\infty}\int s_n^{\psi}\int w(\theta,\delta_n^B)f_\theta^n\,\mathrm{d}\pi_1(\theta)\,\mathrm{d}\mu^n=W(\psi,\delta^B). \tag{22}$$

If $W(\psi,\delta^B)<\infty$, then (22) is equivalent to

$$\sum_{n=1}^{\infty}\int s_n^{\psi}\Delta_n\,\mathrm{d}\mu^n\geq 0,$$

where

$$\Delta_n=\int w(\theta,\delta_n)f_\theta^n\,\mathrm{d}\pi_1(\theta)-\int w(\theta,\delta_n^B)f_\theta^n\,\mathrm{d}\pi_1(\theta),$$

which is, due to (18), non-negative μ^n-almost everywhere for all $n=1,2,\ldots$. Thus, there is an equality in (22) if and only if $\Delta_n=0$ μ^n-almost everywhere on $S_n^{\psi}=\{s_n^{\psi}>0\}$ for all $n=1,2,\ldots$. □

Because of (17), it follows from Theorem 2.3 that for any sequential decision procedure (ψ,δ)

$$R(\psi,\delta;c)\geq R(\psi,\delta^B;c). \tag{23}$$

The following lemma gives the right-hand side of (23) a more convenient form.

For any probability measure π on Θ let us denote

$$P^{\pi}(\tau_\psi=n)\equiv\int P_\theta(\tau_\psi=n)\,\mathrm{d}\pi(\theta)=\int E_\theta s_n^{\psi}\,\mathrm{d}\pi(\theta),$$

for $n=1,2,\ldots$ Respectively, $P^{\pi}(\tau_\psi<\infty)=\sum_{n=1}^{\infty}P^{\pi}(\tau_\psi=n)$, and

$$E^{\pi}\tau_\psi=\int E_\theta\tau_\psi\,\mathrm{d}\pi(\theta).$$

Lemma 2.4. If

$$P^{\pi_2}(\tau_\psi<\infty)=1 \tag{24}$$

then

$$R(\psi,\delta^B;c)=\sum_{n=1}^{\infty}\int s_n^{\psi}\,(cnf^n+l_n)\,\mathrm{d}\mu^n, \tag{25}$$

where, by definition,

$$f^n=f^n(x_1,\ldots,x_n)=\int f_\theta^n(x_1,\ldots,x_n)\,\mathrm{d}\pi_2(\theta). \tag{26}$$

P r o o f. By Theorem 2.3,

$$R(\psi, \delta^B; c) = cN(\psi) + W(\psi, \delta^B) = cN(\psi) + \sum_{n=1}^{\infty} \int s_n^{\psi} l_n \, \mathrm{d}\mu^n. \tag{27}$$

If now (24) is fulfilled, then, by the Fubini theorem,

$$N(\psi) = \int \sum_{n=1}^{\infty} nE_{\theta} s_n^{\psi} \, \mathrm{d}\pi_2(\theta) = \sum_{n=1}^{\infty} \int E_{\theta} n s_n^{\psi} \, \mathrm{d}\pi_2(\theta)$$

$$= \sum_{n=1}^{\infty} \int s_n^{\psi} \left(n \int f_{\theta}^n \, \mathrm{d}\pi_2(\theta) \right) \mathrm{d}\mu^n = \sum_{n=1}^{\infty} \int s_n^{\psi} n f^n \, \mathrm{d}\mu^n,$$

so, combining this with (27), we get (25). □

Let us denote

$$R(\psi) = R(\psi; c) = R(\psi, \delta^B; c). \tag{28}$$

By Lemma 2.4,

$$R(\psi) = \begin{cases} \sum_{n=1}^{\infty} \int s_n^{\psi} \left(cnf^n + l_n \right) \mathrm{d}\mu^n, \ \text{if } P^{\pi_2}(\tau_{\psi} < \infty) = 1, \\ \infty, \ \text{otherwise.} \end{cases} \tag{29}$$

The aim of what follows is to minimize $R(\psi)$ over all stopping rules. In this way, our problem of minimization of $R(\psi, \delta)$ is reduced to an optimal stopping problem.

3. OPTIMAL TRUNCATED STOPPING RULES

In this section, as a first step, we characterize the structure of optimal stopping rules in the class $\mathscr{F}^N$, $N \geq 2$, of all truncated stopping rules, i. e., such that

$$\psi = (\psi_1, \psi_2, \ldots, \psi_{N-1}, 1, \ldots) \tag{30}$$

(if $(1-\psi_1)\ldots(1-\psi_n) = 0$ μ^n-almost everywhere for some $n < N$, we suppose that $\psi_k \equiv 1$ for any $k > n$, so $\mathscr{F}^N \subset \mathscr{F}^{N+1}$, $N = 1, 2, \ldots$).

Obviously, for any $\psi \in \mathscr{F}^N$

$$R(\psi) = R_N(\psi) = \sum_{n=1}^{N-1} \int s_n^{\psi}(cnf^n + l_n) \, \mathrm{d}\mu^n + \int t_N^{\psi} \left(cNf^N + l_N \right) \mathrm{d}\mu^N,$$

where for any $n = 1, 2, \ldots$

$$t_n^{\psi} = t_n^{\psi}(x_1, \ldots, x_n) = (1 - \psi_1(x_1))(1 - \psi_2(x_1, x_2)) \ldots (1 - \psi_{n-1}(x_1, \ldots, x_{n-1}))$$

(we suppose, by definition, that $t_1^{\psi} \equiv 1$).

Let us introduce a sequence of functions V_n^N, $n=1,\dots,N$, which will define optimal stopping rules. Let $V_N^N \equiv l_N$, and recursively for $n=N-1,N-2,\dots 1$

$$V_n^N = \min\{l_n, Q_n^N\}, \tag{31}$$

where

$$Q_n^N = Q_n^N(x_1,\dots,x_n) = cf^n(x_1,\dots,x_n) + \int V_{n+1}^N(x_1,\dots,x_{n+1})\,\mathrm{d}\mu(x_{n+1}), \tag{32}$$

$n=0,1,\dots,N-1$ (we assume that $f^0\equiv 1$). Please, remember that all V_n^N and Q_n^N implicitly depend on the "unitary observation cost" c.

The following theorem characterizes the structure of optimal stopping rules in $\mathscr{F}^N$.

Theorem 3.1. For all $\psi \in \mathscr{F}^N$

$$R_N(\psi) \geq Q_0^N. \tag{33}$$

The lower bound in (33) is attained by a $\psi \in \mathscr{F}^N$ if and only if

$$I_{\{l_n<Q_n^N\}} \leq \psi_n \leq I_{\{l_n\leq Q_n^N\}} \tag{34}$$

μ^n-almost everywhere on

$$T_n^\psi = \{(x_1,\dots,x_n) : t_n^\psi(x_1,\dots,x_n) > 0\},$$

for all $n=1,2,\dots,N-1$.

The proof of Theorem 3.1 can be conducted following the lines of the proof of Theorem 3.1 in [17] (in a less formal way, the same routine is used to obtain Theorem 4 in [15]). In fact, both of these theorems are particular cases of Theorem 3.1.

Remark 3.2. Despite that ψ satisfying (34) is optimal among all truncated stopping rules in $\mathscr{F}^N$, it only makes practical sense if

$$l_0 = \inf_d \int w(\theta,d)\,\mathrm{d}\pi_1(\theta) \geq Q_0^N. \tag{35}$$

Indeed, if (35) does not hold, we can, without taking any observation, make any decision d_0 such that $\int w(\theta,d_0)\,\mathrm{d}\pi_1(\theta) < Q_0^N$, and this guarantees that this trivial procedure (something like "(ψ_0,d_0)" with $R(\psi_0,d_0) = \int w(\theta,d_0)\,\mathrm{d}\pi_1(\theta) < Q_0^N$) performs better than the best procedure with the optimal stopping time in $\mathscr{F}^N$.

Because of this, V_0^N, defined by (31) for $n=0$, may be considered the "minimum value of $R(\psi)$", when taking no observations is allowed.

Remark 3.3. When π_2 in (5) coincides with π_1 in (3) (Bayesian setting), an optimal truncated (non-randomized) stopping rule for minimizing (17) is provided by Theorem 5.2.2 in [9]. Theorem 3.1 describes the class of *all randomized* optimal stopping rules for the same problem in this particular case. This may be irrelevant if one is interested in the purely Bayesian problem, because any of these stopping rules provides the same minimum value of the risk.

Nevertheless, this extension of the class of optimal procedures may be useful for complying with (11) in Theorem 2.1 when seeking for optimal sequential procedures for the original conditional problem (minimization of $N(\psi)$ given that $W_i(\psi,\delta)\leq w_i$, $i=1,\dots,k$, see Introduction and the discussion therein). This is very much like in non-sequential hypothesis testing, where the randomization is crucial for finding the optimal level-α test in the Neyman-Pearson problem (see, for example, [11]).

4. OPTIMAL NON-TRUNCATED STOPPING RULES

In this section, we solve the problem of minimization of $R(\psi)$ in natural classes of non-truncated stopping rules ψ.

Let ψ be any stopping rule. Define

$$R_N(\psi)=R_N(\psi;c)=\sum_{n=1}^{N-1}\int s_n^{\psi}(cnf^n+l_n)\,\mathrm{d}\mu^n+\int t_N^{\psi}\left(cNf^N+l_N\right)\,\mathrm{d}\mu^N. \quad (36)$$

This is the "risk" (17) for ψ truncated at N, i.e. the rule with the components $\psi^N=(\psi_1,\psi_2,\dots,\psi_{N-1},1,\dots)$: $R_N(\psi)=R(\psi^N)$.

Because ψ^N is truncated, the results of the preceding section apply, in particular, the lower bound of (33). Very much like in [17] and in [15], our aim is to pass to the limit, as $N\to\infty$, in order to obtain a lower bound for $R(\psi)$, and conditions for attaining this bound.

It is easy to see that $V_n^N(x_1,\dots,x_n)\geq V_n^{N+1}(x_1,\dots,x_n)$ for all $N\geq n$, and for all $(x_1,\dots,x_n)$, $n\geq 1$ (see, for example, Lemma 3.3 in [17]). Thus, for any $n\geq 1$ there exists

$$V_n=V_n(x_1,\dots,x_n)=\lim_{N\to\infty}V_n^N(x_1,\dots,x_n),$$

(V_n implicitly depend on c, as V_n^N do). It immediately follows from the dominated convergence theorem that for all $n\geq 1$

$$\lim_{N\to\infty}Q_n^N(x_1,\dots,x_n)=cf^n(x_1,\dots,x_n)+\int V_{n+1}(x_1,\dots,x_{n+1})\,\mathrm{d}\mu(x_{n+1}) \quad (37)$$

(see (32)). Let $Q_n=Q_n(x_1,\dots,x_n)=\lim_{N\to\infty}Q_n^N(x_1,\dots,x_n)$.

In addition, passing to the limit, as $N\to\infty$, in (31) we obtain

$$V_n=\min\{l_n,Q_n\},\quad n=1,2,\dots.$$

Let now $\mathscr{F}$ be any class of stopping rules such that $\psi\in\mathscr{F}$ entails $R_N(\psi)\to R(\psi)$, as $N\to\infty$ (let us call such stopping rules *truncatable*). It is easy to see that such classes exist, for example, any $\mathscr{F}^N$ has this property. Moreover, we will assume that all truncated stopping rules are included in $\mathscr{F}$, i.e. that $\bigcup_{N\geq 1}\mathscr{F}^N\subset\mathscr{F}$.

It follows from Theorem 3.1 now that for all $\psi\in\mathscr{F}$

$$R(\psi)\geq Q_0. \tag{38}$$

The following lemma states that, in fact, the lower bound in (38) is the infimum of the risk $R(\psi)$ over $\psi\in\mathscr{F}$.

Lemma 4.1.

$$Q_0=\inf_{\psi\in\mathscr{F}}R(\psi).$$

The proof of Lemma 4.1 is very close to that of Lemma 3.5 in [17] (see also Lemma 6 in [15]) and is omitted here.

Remark 4.2. Again (see Remark 3.3), if $\pi_1=\pi_2$, Lemma 4.1 is essentially Theorem 5.2.3 in [9] (see also Section 7.2 of [8]) .

The following Theorem gives the structure of optimal stopping rules in $\mathscr{F}$.

Theorem 4.3. If there exists $\psi\in\mathscr{F}$ such that

$$R(\psi)=\inf_{\psi'\in\mathscr{F}}R(\psi'), \tag{39}$$

then

$$I_{\{l_n<Q_n\}}\leq\psi_n\leq I_{\{l_n\leq Q_n\}} \tag{40}$$

μ^n-almost everywhere on T_n^ψ for all $n=1,2,\ldots$.

On the other hand, if a stopping rule ψ satisfies (40) μ^n-almost everywhere on T_n^ψ for all $n=1,2,\ldots$, and $\psi\in\mathscr{F}$, then ψ satisfies (39) as well.

The proof of Theorem 4.3 is very close to the proof of Theorem 3.2 in [17] or Theorem 6 in [15] and is omitted here.

It follows from Theorem 4.3 that "$\psi\in\mathscr{F}$" is a sufficient condition for the optimality of a stopping rule ψ satisfying (40). In the hypothesis testing problems considered in [17] and in [15], there are large classes of problems (called truncatable) for which $R_N(\psi)\to R(\psi)$, as $N\to\infty$, for *all* stopping times ψ. In this article, we also identify the problems where this is the case.

The following Lemma gives a necessary and sufficient condition for truncatablity of a stopping rule.

Lemma 4.4. A stopping rule ψ with $R(\psi)<\infty$ is truncatable if and only if

$$\int t_N^\psi l_N\,\mathrm{d}\mu^N\to 0,\quad\text{as}\quad N\to\infty. \tag{41}$$

If $R(\psi)=\infty$, then $R_N(\psi)\to\infty$, $N\to\infty$.

P r o o f. Let ψ be such that $R(\psi) < \infty$.

Suppose that (41) is fulfilled. Then, by (36)

$$R(\psi) - R_N(\psi) = \sum_{n=N}^{\infty} \int s_n^{\psi}(cnf^n + l_n)\,\mathrm{d}\mu^n - c\int t_N^{\psi} N f^N d\mu^N + \int t_N^{\psi} l_N\,\mathrm{d}\mu^N. \quad (42)$$

The first summand converges to zero, as $N \to \infty$, being the tail of a convergent series (this is because $R(\psi) < \infty$).

The third summand in (42) goes to 0 as $N \to \infty$, because of (41).

The integral in the second summand in (42) is equal to

$$NP^{\pi_2}(\tau_\psi \geq N) \leq E^{\pi_2}\tau_\psi I_{\{\tau_\psi \geq N\}} \to 0,$$

as $N \to \infty$, because $E^{\pi_2}\tau_\psi < \infty$ (this is due to $R(\psi) < \infty$ again).

It follows from (42) now that $R_N(\psi) \to R(\psi) < \infty$ as $N \to \infty$.

Let us suppose now that $R_N(\psi) \to R(\psi) < \infty$ as $N \to \infty$. For the same reasons as above, the first two summands on the right-hand side of (42) tend to 0 as $N \to \infty$, therefore so does the third, i.e. (41) follows. The first assertion of Lemma 4.4 is proved.

If $R(\psi) = \infty$, this may be because $P^{\pi_2}(\tau_\psi < \infty) < 1$, or, if not, because

$$\sum_{n=1}^{\infty} \int s_n^{\psi}(cnf^n + l_n)\,\mathrm{d}\mu^n = \infty.$$

In the latter case, obviously,

$$R_N(\psi) \geq \sum_{n=1}^{N-1} \int s_n^{\psi}(cnf^n + l_n)\,\mathrm{d}\mu^n \to \infty, \quad \text{as} \quad N \to \infty.$$

In the former case,

$$R_N(\psi) \geq c\int t_N^{\psi} N f^N\,\mathrm{d}\mu^N = cNP^{\pi_2}(\tau_\psi \geq N) \to \infty,$$

as $N \to \infty$, as well. □

Let us say that the problem (of minimization of $R(\psi)$) is truncatable if all stopping rule ψ are truncatable.

Corollary 4.5 below gives some practical sufficient conditions for truncatability of a problem.

Corollary 4.5. The problem of minimization of $R(\psi)$ is truncatable if
i) the loss function w is bounded, and

$$R(\psi) < \infty \quad \text{implies that} \quad P^{\pi_1}(\tau_\psi < \infty) = 1, \quad (43)$$

or
ii)

$$\int l_N\,\mathrm{d}\mu^N \to 0, \quad (44)$$

as $N \to \infty$.

P r o o f. If $w(\theta,d)<M<\infty$ for any θ and d, then, by the definition of l_N,

$$\int t_N^\psi l_N \,\mathrm{d}\mu^N \le M\int t_N^\psi\left(\int f_\theta^N\,\mathrm{d}\pi_1(\theta)\right)\mathrm{d}\mu^N = MP^{\pi_1}(\tau_\psi\ge N). \tag{45}$$

If now $R(\psi)<\infty$, then by (43) the right-hand side of (45) tends to 0, as $N\to\infty$, i.e. (41) is fulfilled for any ψ such that $R(\psi)<\infty$. Thus, by Lemma 4.4 any ψ is truncatable.

If (44) is fulfilled, then (41) is satisfied for any ψ. Again, by Lemma 4.4 any ψ is truncatable. □

Remark 4.6. Condition i) of Corollary 4.5 is fulfilled for any Bayesian hypothesis testing problem (i.e. when $\pi_1=\pi_2=\pi$) with bounded loss function (see, for example, [17] and [15]). Indeed, in this case $R(\psi)<\infty$ implies $E^\pi\tau_\psi<\infty$, so, in particular, $P^\pi(\tau_\psi<\infty)=1$.

Remark 4.7. It is easy to see that Condition ii) of Corollary 4.5 is equivalent to

$$\int_\Theta E_\theta w(\theta,\delta_N^B)\,\mathrm{d}\pi_1(\theta)\to 0,\ N\to\infty,$$

i.e. that the Bayesian risk, with respect to the prior distribution π_1, of an optimal procedure based on sample of a fixed size N, vanishes as $N\to\infty$. This is a very typical behavior of statistical risks.

The following Theorem is an immediate consequence of Theorem 4.3.

Theorem 4.8. Let the problem of minimization of $R(\psi)$ be truncatable, and let $\mathscr{F}$ be the set of all stopping rules. Then

$$R(\psi)=\inf_{\psi'\in\mathscr{F}} R(\psi') \tag{46}$$

if and only if

$$I_{\{l_n<Q_n\}}\le\psi_n\le I_{\{l_n\le Q_n\}} \tag{47}$$

μ^n-almost everywhere on T_n^ψ for all $n=1,2,\dots$.

Remark 4.9. Once again (see Remark 3.2), the optimal stopping rule ψ from Theorem 4.8 (and Theorem 4.3) only makes practical sense if $l_0\ge Q_0=\inf_{\psi\in\mathscr{F}}R(\psi)$, because otherwise the trivial rule, which does not take any observation, performs better than ψ, from the point of view of minimization of $R(\psi)$.

Remark 4.10. Combining Theorems 2.1, 2.3, and 4.8 we immediately have the following solution to the conditional problem posed in Introduction.

Let $\lambda_1,\dots,\lambda_k$ be arbitrary positive constants. Let δ_n^B, $n=1,2,\dots$ be Bayesian, with respect to π_1, decision rules for the "loss function"

$$w'(\theta,d)=\sum_{i=1}^k\lambda_i w(\theta,d)I_{\Theta_i}(\theta),$$

i. e. such that for all $d \in \mathscr{D}$

$$\sum_{i=1}^{k} \lambda_i \int_{\Theta_i} w(\theta, d) f_\theta^n \,\mathrm{d}\pi_1(\theta) \geq l_n = \sum_{i=1}^{k} \lambda_i \int_{\Theta_i} w(\theta, \delta_n^B) f_\theta^n \,\mathrm{d}\pi_1(\theta) \tag{48}$$

μ^n-almost everywhere (remember that $\delta_n^B = \delta_n^B(x_1, \dots, x_n)$ and $f_\theta^n = f_\theta^n(x_1, \dots, x_n)$).

For any $N \geq 1$ define $V_N^N = l_N$, and $V_n^N = \min\{l_n, Q_n^N\}$ for $n = N-1, N-2, \dots, 1$, where $Q_n^N = f^n + \int l_{n+1} \,\mathrm{d}\mu(x_{n+1})$, with $f^n = \int_\Theta f_\theta^n \,\mathrm{d}\pi_2(\theta)$ (cf. (31) and (32)).

Let also $V_n = \lim_{N\to\infty} V_n^N$ and $Q_n = \lim_{N\to\infty} Q_n^N$, $n = 1, 2, \dots$.

Suppose, finally, that the problem is truncatable (see Corollary 4.5 for sufficient conditions for that).

Let $\psi = (\psi_1, \psi_2, \dots)$ be any stopping rule satisfying

$$I_{\{l_n < Q_n\}} \leq \psi_n \leq I_{\{l_n \leq Q_n\}} \tag{49}$$

μ^n-almost everywhere on T_n^ψ for all $n = 1, 2, \dots$.

Then for any sequential decision procedure (ψ', δ) such that

$$W_i(\psi', \delta) \leq W_i(\psi, \delta^B), \quad i = 1, \dots, k, \tag{50}$$

it holds

$$N(\psi) \leq N(\psi'). \tag{51}$$

The inequality in (51) is strict if at least one of the inequalities in (50) is strict.

If there are equalities in all of the inequalities in (50) and (51), then

$$I_{\{l_n < Q_n\}} \leq \psi_n' \leq I_{\{l_n \leq Q_n\}} \tag{52}$$

μ^n-almost everywhere on $T_n^{\psi'}$ for all $n = 1, 2, \dots$, and

$$\sum_{i=1}^{k} \lambda_i \int_{\Theta_i} w(\theta, \delta_n) f_\theta^n \,\mathrm{d}\pi_1(\theta) = \sum_{i=1}^{k} \lambda_i \int_{\Theta_i} w(\theta, \delta_n^B) f_\theta^n \,\mathrm{d}\pi_1(\theta)$$

μ^n-almost everywhere on $S_n^{\psi'}$ for all $n = 1, 2, \dots$.

For Bayesian problems (when $\pi_1 = \pi_2 = \pi$) Theorem 4.8 can be reformulated in the following equivalent way.

Let

$$R_n = \frac{l_n}{f^n} = \frac{\int_\Theta f_\theta^n w(\theta, \delta_n^B) \,\mathrm{d}\pi(\theta)}{\int_\Theta f_\theta^n \,\mathrm{d}\pi(\theta)}$$

be the posterior risk (see, e. g., [1]). Let $v_N^N \equiv R_N(X_1, \dots, X_n)$, and recursively for $n = N-1, N-2, \dots, 1$

$$v_n^N(X_1, \dots, X_n) = \min\{R_n(X_1, \dots, X_n), q_n^N(X_1, \dots, X_n)\},$$

where

$$q_n^N(X_1, \dots, X_n) = c + E^\pi\{v_{n+1}^N | X_1, \dots, X_n\}$$

(E^π stands for the expectation with respect to the family of finite-dimensional densities $f^n = \int_\Theta f_\theta^n \, \mathrm{d}\pi(\theta)$, $n = 1, 2, \ldots$, meaning, in particular, that

$$E^\pi\{v_{n+1}^N | x_1, \ldots, x_n\} = \int \frac{v_{n+1}^N(x_1, \ldots, x_{n+1}) f^{n+1}(x_1, \ldots, x_{n+1})}{f^n(x_1, \ldots, x_n)} \, \mathrm{d}\mu(x_{n+1})).$$

Let, finally, $v_n = v_n(X_1, \ldots, X_n) = \lim_{N\to\infty} v_n^N(X_1, \ldots, X_n)$, and $q_n = q_n(X_1, \ldots, X_n) = \lim_{N\to\infty} q_n^N(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$.

Then, the following reformulation of Theorem 4.8 gives, for a truncatable Bayesian problem, the structure of all Bayesian randomized tests (cf. Theorem 7, Ch. 7, in [1]).

Theorem 4.11. Let the problem of minimization of $R(\psi)$ be truncatable, and let $\mathscr{F}$ be the set of all stopping rules. Then

$$R(\psi) = \inf_{\psi' \in \mathscr{F}} R(\psi') \tag{53}$$

if and only if

$$I_{\{R_n < q_n\}} \le \psi_n \le I_{\{R_n \le q_n\}} \tag{54}$$

P^π-almost surely on T_n^ψ for all $n = 1, 2, \ldots$.

Remark 4.12. More general variants of Theorem 4.11, for cases when the loss function due to incorrect decision is of the form $w(\theta, d) = w_n(\theta, d; x_1, \ldots, x_n)$ and/or the cost of the observations $(x_1, \ldots, x_n)$ is of type $K_\theta^n(x_1, \ldots, x_n)$, can easily be deduced from Theorem 4 [16]. In particular, this gives the structure of optimal sequential multiple hypotheses tests for the problem considered in Section 9.4 of [24].

Remark 4.13. Theorem 4.11, in particular, gives a solution to optimal sequential hypothesis testing problems considered in [6] and [5] (where the general theory of optimal stopping is used, see [4] or [20]). See [17] and [15] for a more detailed description of the respective Bayesian sequential procedures.

ACKNOWLEDGEMENT

The author thanks the anonymous referees for their valuable comments and suggestions on the improvement of an earlier version of this paper.

The author greatly appreciates the support of the Autonomous Metropolitan University, Mexico City, Mexico, where this work was done, and the support of the National System of Investigators (SNI) of CONACyT, Mexico.

This work is also partially supported by Mexico's CONACyT Grant no. CB-2005-C01-49854-F.

(Received June 8, 2008)

REFERENCES

[1] J. O. Berger: Statistical Decision Theory and Sequential Analysis. Second edition. Springer-Verlag, New York, Berlin, Heidelberg, Tokyo 1985.

[2] R. H. Berk: Locally most powerful sequential tests. Ann. Statist. *3* (1975), 373–381.

[3] E. Castillo and J. García: Necessary conditions for optimal truncated sequential tests. Simple hypotheses (in Spanish). Stochastica *7* (1983), 1, 63–81.

[4] Y. S. Chow, H. Robbins, and D. Siegmund: Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin Company, Boston 1971.

[5] J. Cochlar: The optimum sequential test of a finite number of hypotheses for statistically dependent observations. Kybernetika *16* (1980), 36–47.

[6] J. Cochlar and I. Vrana: On the optimum sequential test of two hypotheses for statistically dependent observations. Kybernetika *14* (1978), 57–69.

[7] M. H. DeGroot: Optimal Statistical Decisions. McGraw-Hill Book Co., New York, London, Sydney 1970.

[8] T. Ferguson: Mathematical Statistics: A Decision Theoretic Approach. Probability and Mathematical Statistics, Vol. 1. Academic Press, New York, London 1967.

[9] M. Ghosh, N. Mukhopadhyay, and P. K. Sen: Sequential Estimation. John Wiley & Sons, New York, Chichester, Weinheim, Brisbane, Singapore, Toronto 1997.

[10] J. Kiefer and L. Weiss: Some properties of generalized sequential probability ratio tests. Ann. Math. Statist. *28* (1957), 57–75.

[11] E. L. Lehmann: Testing Statistical Hypotheses. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London 1959.

[12] G. Lorden: Structure of sequential tests minimizing an expected sample size. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete *51* (1980), 291–302.

[13] U. Müller-Funk, F. Pukelsheim, and H. Witting: Locally most powerful tests for two-sided hypotheses. Probability and statistical decision theory, Vol. A (Bad Tatzmannsdorf 1983), pp. 31–56, Reidel, Dordrecht 1985.

[14] A. Novikov: Optimal sequential multiple hypothesis testing in presence of control variables. Kybernetika *45* (2009), 3, 507–528.

[15] A. Novikov: Optimal sequential multiple hypothesis tests, Kybernetika *45* (2009), 2, 309–330.

[16] A. Novikov: Optimal sequential procedures with Bayes decision rules. Internat. Math. Forum *5* (2010), 43, 2137–2147.

[17] A. Novikov: Optimal sequential tests for two simple hypotheses. Sequential Analysis *28* (2009), 2, 188–217.

[18] A. Novikov and P. Novikov: Locally most powerful sequential tests of a simple hypothesis vs. one-sided alternatives. Journal of Statistical Planning and Inference *140* (2010), 3, 750-765.

[19] N. Schmitz: Optimal Sequentially Planned Decision Procedures. Lecture Notes in Statistics *79* (1993), New York: Springer-Verlag.

[20] A. N. Shiryayev: Optimal Stopping Rules. Springer-Verlag, Berlin, Heidelberg, New York 1978.

[21] A. Wald: Statistical Decision Functions. John Wiley & Sons, Inc., New York, London, Sydney 1971.

[22] A. Wald and J. Wolfowitz: Optimum character of the sequential probability ratio test. Ann. Math. Statist. *19* (1948), 326–339.

[23] L. Weiss: On sequential tests which minimize the maximum expected sample size. J. Amer. Statist. Assoc. *57* (1962), 551–566.

[24] S. Zacks: The Theory of Statistical Inference. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, London, Sydney 1971.

Andrey Novikov, Departamento de Matemáticas, Universidad Autónoma Metropolitana – Unidad Iztapalapa, San Rafael Atlixco 186, col. Vicentina, C.P. 09340, México D.F.. México

e-mail: an@xanum.uam.mx