SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

ON COMPUTING THE OC AND ASN FOR SOME MARKOVIAN SPRT'S

GEORGICĂ OBREJA and GHEORGHIŢĂ ZBĂGANU

Centre of Mathematical Statistics, Bucharest, Romania

Let $(\Omega, \mathscr{K})$ be a measurable space and $(P_\theta)_{\theta\in\Theta}$ a family of unknown probabilities. Let $X_n\colon \Omega \to R$ be a sequence of random variables such that, for every $\theta\in\Theta$, $((X_n)_n, P_\theta)$ is a homogeneous Markov chain. Let $F_\theta^n = P_\theta \circ (X_1, \ldots, X_n)^{-1}$ and let $Q_\theta(x, dy)$ be the transition probability of the chain if P_θ is the true probability. Also, let Π_θ be the starting probability of the chain; $\Pi_\theta = P_\theta\circ(X_1)^{-1}$.

The intuitive signification is that we have a process X_n which should be Markovian by virtue of a previous model; the problem is to select the true transition kernel $Q_\theta(x, dy)$ from a family $(Q_\theta(z, \cdot))_\theta$ which is available to us from the theoretical model of the phenomenon and which is supposed to be large enough to contain the true one. It is not our ambition to test either if the phenomenon is Markovian or if the stock of available kernels is large enough.

We shall also suppose that there exists a measure λ on $(R, \mathscr{B}_R)$ which dominates all the $Q_\theta(x, \cdot)$ and Π_θ. Namely

$$Q_\theta(x, dy) = f_\theta(x, y)\lambda(dy), \qquad \Pi_\theta(dy) = \varrho_\theta(y)\,dy.$$

Then it is clear that

$$dF_\theta^n/d\lambda^n(x_1, \ldots, x_n) = \varrho_\theta(x_1) f_\theta(x_1, x_2) \ldots f_\theta(x_{n-1}, x_n) := L_n(x_1, \ldots, x_n; \theta) \quad (1)$$

If $(X_n)_n$ is a homogeneous random walk for every probability P_θ (i.e., X_n is a sum of i.i.d. random variables or, in other words, a homogeneous Markov chain with independent increments) then $f_\theta(x, y)$ is of the form $f_\theta(y-x)$ and (1) becomes

$$L_n(x_1, x_2, \ldots, x_n; \theta) = \varrho_\theta(x_1) f_\theta(x_1 - x_2) \ldots f_\theta(x_n - x_{n-1}). \qquad (2)$$

We shall construct the usual SPRT for testing the hypothesis H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$. Set

$$Z_n := \ln\left(\frac{L_n(X_1,\ldots,X_n;\theta_1)}{L_n(X_1,\ldots,X_n;\theta_0)}\right) = \ln\frac{\varrho_{\theta_1}(X_1)}{\varrho_{\theta_0}(X_1)} + \sum_{i=1}^{n-1} \ln\frac{f_{\theta_1}(X_i, X_{i+1})}{f_{\theta_0}(X_i, X_{i+1})}. \tag{3}$$

Let $b < 0 < a$. The test will be constructed in the usual manner, and it will be denoted by $T(b, a)$:

– if $b < Z_n < a$, continue sampling;

– if $Z_n \leqslant b$, stop and accept H_0 (reject H_1);

– if $Z_n \geqslant a$, stop and accept H_1 (reject H_0).

Denote

$$\tau(b, a) := \tau := \inf\{k \mid Z_k \notin (b, a)\}.$$

If $P_\theta(\tau < \infty) = 1$ for every $\theta \in \Theta$, we say that the test is *closed*.

Concerning the initial distributions Π_θ the statisticians agree with the following two alternatives:

– either suppose that the starting point of the chain is known; then the first term from (3) becomes equal to 0;

– or construct a different test for every starting point $X_1(\omega)$, substracting the first term from the barriers b, a.

In order to avoid more difficulties, we shall suppose that the starting point is known. In fact, the practician is interested in finding the transition kernel of the Markov chain more than in the starting probabilities. It seems to us that the second problem is not a consistent one unless we suppose that the chain is stationary; if we do that, the problem of the initial distribution becomes an analytical one (rather than a statistical one) because the stationary distribution, when existing, is given by the transition kernel.

In order to have no more trouble, we have to make another hypothesis: that the support of the functions $y \to f_\theta(x, y)$ does not depend on θ. It is only a technical supposition made in order not to deal with $\bar{R}$-valued processes. But it seems natural that if $L_n(X_1,\ldots,X_n;\theta_1) = 0$ and $L_n(X_1,\ldots,X_n;\theta_0) \neq 0$ one should reject H_1; and if $L_n(X_1,\ldots,X_n;\theta_1) \neq 0$ but $L_n(X_1,\ldots,X_n;\theta_0) = 0$, H_0 should be rejected; and if both quantities are equal to zero, one should reject both H_0 and H_1 and perhaps search for another model.

The problems we shall deal with are:

1) Find sufficient conditions to ensure the closure of the test.

2) Find sufficient conditions under which one can compute – at least approximately – the OC-function:

$$\mathrm{OC}(\theta) = P_\theta(Z_\tau \leqslant b) = \sum_{n=1}^{\infty} P_\theta\big(Z_i \in (b, a) \text{ for } i = 1,\ldots,n-1 \text{ but } Z_n \leqslant b\big).$$

3) Find sufficient conditions under which one can compute (approximately) the ASN-function:

$$\mathrm{ASN}(\theta) = E_\theta(\tau) = \sum_{n=1}^{\infty} nP_\theta(\tau = n).$$

Remark that if the test is closed we have $\mathrm{OC}(\theta_1) = 1-\mathrm{OC}(\theta_0)$, because $\mathrm{OC}(\theta)$ means the probability of accepting H_0 if the true parameter is θ. If θ_0 and θ_1 both fail to be true but $\mathrm{OC}(\theta)$ is small for $\theta \neq \theta_0$, the test can still be applied as a rejection tool.

In general, the class of the Markov chains for which the test $T(b, a)$ is closed for every $b < a$ is rather small. Here is a counter-example which points out that even in good cases the test may fail to be closed:

Let X_n be a random walk with an absorbing barrier in 0; that is, X_n: $\Omega \to \{0, 1, 2, \ldots\}$ and $P(X_{n+1} = i+1|X_n = i) = \theta$, $P(X_{n+1} = i-1|X_n = i) = 1-\theta$ if $i \neq 0$, but $P(X_{n+1} = 0|X_n = 0) = 1$ for every $\theta \in (0, 1)$. Here the set of parameters is $\Theta = (0, 1)$. Let $\theta_0 < \theta_1$ and construct $T(b, a)$ for testing H_0 against H_1. Then

$$L_{n+1}(X_1, \ldots, X_{n+1}|\theta) = \theta^{d_n}(1-\theta)^{s_n} = \theta^{d_n}(1-\theta)^{n-d_n-a_n}$$

where $d_n = \mathrm{card}\{i \leqslant n+1|X_{i+1}-X_i = 1\}$, $a_n = \mathrm{card}\{i \leqslant n+1|X_i = 0\}$ and $s_n = \mathrm{card}\{i \leqslant n+1|X_{i+1} = X_i-1\}$; of course $d_n+a_n+s_n = n$.

Therefore

$$Z_{n+1} = d_n \log\frac{\theta_1}{\theta_0}+(n-d_n-a_n)\log\frac{1-\theta_1}{1-\theta_0} = d_n \log\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}-(n-a_n)\log\frac{1-\theta_0}{1-\theta_1}.$$

Set $\tau(\omega) := \inf\{n|X_n(\omega) = 0\}$. Then it is known that $\theta < \frac{1}{2} \Rightarrow E_\theta(\tau) < \infty$; hence $P_\theta(\tau < \infty) = 1$. But $a_n = (n+1-\tau)_+$ implies that $n-a_n = \tau-1$ if $n+1 \geqslant \tau$ and $= n$ if $n \leqslant \tau$; then $n > \tau \Rightarrow d_n = d_\tau$. If $n \to \infty$ it follows that

$$Z_n \to \left(d_\tau \log\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}-(\tau-1)\log\frac{1-\theta_0}{1-\theta_1}\right) \quad P\text{-a.s.},$$

and this implies that, for some $b < a$, $P_\theta(Z_n \in (b, a)$ for every $n) > 0$.

The following proposition is a partial answer to the first question:

Proposition 1. *Each of the following two assumptions implies the closure of the test $T(b, a)$ for every $b < 0 < a$:*

I. *$((X_n)_n, P_\theta)$ is a homogeneous random walk for every θ.*

II. *The family $(P_\theta)_\theta$ is an exponential family in the following sense* (see [3]):

$$L(X_1, \ldots, X_n; \theta) = C_n(\theta)\cdot\exp(D_n(\theta)K_n(X_1, \ldots, X_n))$$

with $C_n(\theta) > 0$, $D_n(\theta) > 0$ *strictly monotonous and*

$$a.s.\text{-}\lim K_n(X_1,\ldots,X_n) = g(\theta) \pmod{P_\theta}$$

$$\lim\big(b-\log(C_n(\theta_1)/C_n(\theta_0))\big)/\big(D_n(\theta_1)-D_n(\theta_0)\big) = c(\theta)$$
$$= \lim\big(a-\log(C_n(\theta_1)/C(\theta_0))\big)/\big(D_n(\theta_1)-D_n(\theta_0)\big) \pmod{P_\theta} \qquad \textit{for every } \theta$$

and $P_\theta(g(\theta) \neq c(\theta)) = 1$.

Proof. I. If X_n is a random walk, $f(x, y) = f(y-x)$ and

$$Z_n = \sum_{i=1}^{n-1} \log\frac{f_{\theta_1}(X_{i+1}+X_i)}{f_{\theta_0}(X_{i+1}-X_i)} = \sum_{i=1}^{n-1} z_i.$$

Then z_i are i.i.d. random variables because $X_{i+1}-X_i$ are i.i.d.; this means that Z_n is a random walk itself. But for a random walk the following facts are well known (see for instance [2]):

– If $E_\theta(z_1) > 0$ then $\lim_n Z_n = \infty \pmod{P_\theta}$.

– If $E_\theta(z_1) < 0$ then $\lim_n Z_n = -\infty \pmod{P_\theta}$.

If $E_\theta(z_1) = 0$ but Z_1 is not identically $\pmod{P_\theta}$ equal to 0, then $\liminf_{n\to\infty} Z_n = -\infty$ and $\limsup_{n\to\infty} Z_n = \infty \pmod{P_\theta}$. In all these three cases $P_\theta(Z_n \in (b, a) \text{ for every } n) = 0$, i.e., the test is closed.

If $Z_1 \equiv 0$, then

$$f_{\theta_1}(X_2-X_1) = f_{\theta_0}(X_2-X_1) \pmod{P_\theta}.$$

This equality further implies that $\{f_{\theta_1} \neq f_{\theta_0}\} \subset (\operatorname{supp} f_\theta)^c$, which is absurd because we supposed that all the functions f_θ have the same support and of course that $\theta_0 \neq \theta_1 \Rightarrow f_{\theta_1} \neq f_{\theta_0}$.

II.

$$P_\theta(b < Z_n < a) = P_\theta\left(b < \log\frac{C_n(\theta_1)}{C_n(\theta_0)} + K_n\big(D_n(\theta_1)-D_n(\theta_0)\big) < a\right)$$
$$= P\left(\frac{b-\log(C_n(\theta_1)/C_n(\theta_0))}{D_n(\theta_1)-D_n(\theta_0)} < K_n < \frac{a-\log(C_n(\theta_1)/C_n(\theta_0))}{D_n(\theta_1)-D_n(\theta_0)}\right),$$

and the last expression tends to 0 due to our assumptions. ∎

Remark. The second condition is a highly restrictive one. In fact, the only examples we know to fulfil it are some random walks. For instance, if $X_n\colon \Omega \to Z$ is a random walk with

$$P_\theta(X_{n+1} = X_n+1) = \theta \qquad \text{and} \qquad P_\theta(X_{n+1} = X_n-1) = 1-\theta,$$

then

$$L_n(X_1,\ldots,X_n;\theta) = \theta^{d_n}(1-\theta)^{n-d_n}$$

where $d_n = \text{card}\{i \leqslant n+1 | X_{i+1} = X_i + 1\}$ and Assumptions II are fulfilled because

$$L(X_1, \ldots, X_n; \theta) = (1-\theta)^{n-1} \exp\big(d_n \log(\theta/(1-\theta))\big).$$

The choice may be $D_n(\theta) = n \log(\theta/(1-\theta))$, $K_n = d_n/n$, $C_n(\theta) = (1-\theta)^{n-1}$. In this case $K_n \underset{n\to\infty}{\longrightarrow} \theta \pmod{P_\theta}$ due to the strong law of large numbers and

$$P_\theta\text{-a.s. } \lim_n \frac{b - \log(C_n(\theta_1)/C_n(\theta_0))}{D_n(\theta_1) - D_n(\theta_0)}$$

$$= P_\theta\text{-a.s. } \lim_n \frac{a - \log(C_n(\theta_1)/C_n(\theta_0))}{D_n(\theta_1) - D_n(\theta_0)} = \log \frac{1-\theta_0}{1-\theta_1} \Big/ \log \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}.$$

We ignore the question whether there exist also other examples of exponential families if X_n are not necessarily i.i.d. or random walks.

Usually, when constructing a SPRT, one starts with two risks α, β, and sets $a = \log((1-\beta)/\alpha)$ and $b = \log(\beta/(1-\alpha))$. After executing the SPRT the two risks becomes α' and β'. One knows that $\alpha' + \beta' \leqslant \alpha + \beta$ provided that the test is closed.

The answer to the second question will be a rather partial one. Wald gave some approximative formulas to compute OC and ASN provided X_n are i.i.d. random variables. They are not at all easy to compute, Wald's approximations are rather difficult to check and, finally, the precision of his formulas is, from a pragmatical point of view, uncomputable. Ghosh conjectured an analogous formula for the general case ([3], p. 132). In the absence of any proof of his conjecture and for pragmatical reasons, the present authors have decided to try a completely different approach.

From now on we shall suppose that the process Z_n is a Markov chain itself. It is not clear at all what assumptions to make about the densities $f_\theta(x, y)$ to ensure that property, but at least it is clear that if $(X_n)_n$ is a random walk this is indeed the case.

Suppose that

$$P_\theta(Z_{n+1} \in dy | Z_n = x) = Q_\theta(x, dy) = q_\theta(x, y)\, dy$$

and that the density $q_\theta(x, y)$ is continuous in x for every fixed y.

Let $b < 0 < a$ and $\tau(\omega) = \inf\{m | Z_m(\omega) \notin (b, a)\}$. Add that $Z_0 \equiv 0$. The idea is to discretize the state space of the chain Z_n and to approximate it (in the sense of weak convergence) by some discrete chains $(Z_m(n))_m$, to compute the OC and ASN functions (OC_n, ASN_n) for the discrete state Markov chains and to give some conditions in which they converge to the OC and ASN functions of the chain.

We shall fix once for ever a parameter θ which will be omitted in notations. For instance, $Q(x, dy)$ means $Q_\theta(x, dy)$.

Let $L := a-b$ and let n be a fixed positive integer. Divide the set of real numbers in the intervals $\Delta_i^n := \left(a+\frac{i-1}{n}L,\ a+\frac{i}{n}L\right]$, $i\in Z$. Denote by ξ_i^n the point $a+\frac{i}{n}L$. We shall construct a sequence of transition probabilities by the relations

$$Q_n(x,\cdot) := \sum_{i\in Z} Q(x,\Delta_i^n)\varepsilon_{\xi_i^n}(\cdot) \tag{4}$$

where ε_α means the point probability measure concentrated on α. The intuitive signification of these new kernels should be obvious; namely, the Markov chains $(Z_m(n))_m$ given by the kernels $Q_n(x,\cdot)$ have the sets $(\xi_i^n)_i$ as state spaces and the transition matrices

$$P\big(Z_{j+1}(n)=\xi_m^n | Z_j(n)=\xi_i^n\big) = Q_n(\xi_i^n, \Delta_m^n).$$

These Markov chains are approximations of the chain $(Z_n)_n$ in the following sense:

PROPOSITION 2. *$Q_n^j(x,\cdot) \Rightarrow Q^j(x,\cdot)$ as n tends to infinity for every $x\in R$.*

Here "$\Rightarrow$" means weak convergence and $Q^j(x, A_1,\dots,A_j)$ means the product of Q by itself j times, that is,

$$Q^j(x,A) = \int\dots\int 1_A(x_1,\dots,x_j)Q(x,dx_1)Q(x_1,dx_2)\dots Q(x_{j-1},dx_j). \tag{5}$$

Proof. If $f\colon R^j\to R$ is a bounded continuous function, we shall write

$$Q^j(x,f) := \int\dots\int f(x_1,\dots,x_j)Q(x,dx_1)\dots Q(x_{j-1},dx_j). \tag{6}$$

First, let j be equal to 1. Remark that in this case

$$Q_n(x,f) = \int f(y)Q_n(x,dy) = \sum_i f(\xi_i^n)Q(x,\Delta_i^n) = Q(x,f_n),$$

where $f_n = \sum_i f(\xi_i^n)1_{\Delta_i^n}$. But if f is bounded and continuous, $\lim_n f_n = f$ and Lebesgue's dominated convergence theorem implies that

$$\lim_n Q_n(x,f) = Q(x,f) \quad \text{for every } x, \quad \text{i.e.,} \quad Q_n(x,\cdot)\Rightarrow Q(x,\cdot).$$

For an arbitrary j, it is enough to check that

$$\lim_n Q_n^j(x,f) = Q^j(x,f)$$

for uniformly continuous and bounded functions $f\colon R^j\to R$ (see [1]).

We have

$$Q^j(x,f) = \int\dots\int f(x_1,\dots,x_j)q(x,x_1),\dots,q(x_{j-1},x_j)dx_1\dots dx_j$$

and

$$Q_n^j(x,f) = \sum_{i_1}\cdots\sum_{i_j} f(\xi_{i_1}^n,\ldots,\xi_{i_j}^n)Q(x,\Delta_{i_1}^n)Q(\xi_{i_1}^n,\Delta_{i_2}^n)\ldots Q(\xi_{i_{j-1}}^n,\Delta_{i_j}^n)$$

$$= \sum_{i_1}\cdots\sum_{i_j}\int_{\Delta_{i_1}^n}\cdots\int_{\Delta_{i_j}^n} f(\xi_{i_1}^n,\ldots,\xi_{i_j}^n)q(x,x_1)q(\xi_{i_1}^n,x_2)\ldots q(\xi_{i_{j-1}}^n,x_j)\,dx_1\ldots dx_j$$

Therefore

$$\begin{aligned}
&|Q_n^j(x,f)-Q^j(x,f)| \\
&\quad\leqslant \sum_{i_1}\cdots\sum_{i_j}\int_{\Delta_{i_1}^n}\cdots\int_{\Delta_{i_1}^n} |f(x_1,\ldots,x_j)q(x,x_1),\ldots,q(x_{j-1},x_j)- \\
&\qquad -f(\xi_{i_1}^n,\ldots,\xi_{i_j}^n)q(x,x_1)q(\xi_{i_1}^n,x_2),\ldots,q(\xi_{i_{j-1}}^n,x_j)|\,dx_1\ldots dx_j \\
&\quad\leqslant \int\cdots\int |f(x_1,\ldots,x_j)|\,|\varrho(x_1,\ldots,x_j)-\varrho_n(x_1,\ldots,x_j)|\,dx_1\ldots dx_j + \\
&\qquad + \int\cdots\int |f(x_1,\ldots,x_j)-f_n(x_1,\ldots,x_j)|\,\varrho_n(x_1,\ldots,x_j)\,dx_1\ldots dx_j := \mathrm{I}+\mathrm{II}
\end{aligned}$$

where

$$\varrho(x_1,\ldots,x_j) = q(x,x_1)q(x_1,x_2),\ldots,q(x_{l-1},x_j),$$

$$\varrho_n(x_1,\ldots,x_j) = \sum_{i_1}\cdots\sum_{i_j} q(x,x_1)q(\xi_{i_1}^n,x_2)\ldots$$

$$\ldots q(\xi_{i_{j-1}}^n,x_j)\,\mathbf{1}_{\Delta_{i_1}^n}(x_1)\ldots\mathbf{1}_{\Delta_{i_j}^n}(x_j)$$

Remark that ϱ_n is indeed a probability density because

$$\int\cdots\int \varrho_n\,dx_1\ldots dx_j = \sum_{i_1}\cdots\sum_{i_j} Q(x,\Delta_{i_1}^n)Q(\xi_{i_1}^n,\Delta_{i_2}^n)\ldots Q(\xi_{i_{j-1}}^n,\Delta_{i_j}^n) = 1$$

and

$$f_n(x_1,\ldots,x_j) = \sum_{i_1}\cdots\sum_{i_j} f(\xi_{i_1}^n\ldots\xi_{i_j}^n)\,\mathbf{1}_{\Delta_{i_1}^n}(x_1)\ldots\mathbf{1}_{\Delta_{i_j}^n}(x_j).$$

Moreover, $f_n \to f$ as n tends to infinity, uniformly because f was supposed to be uniformly continuous. Hence $\mathrm{II} \leqslant \|f-f_n\| \to 0$ as n tends to infinity ($\|\cdot\|$ means the uniform norm).

As regards the first term, remark that $\lim_n \varrho_n = \varrho$ because $x \mapsto q(x,\cdot)$ is continuous, and

$$\mathrm{I} \leqslant \|f\|\int\cdots\int |\varrho-\varrho_n|\,dx_1\ldots dx_j$$

But

$$\begin{aligned}
&\int\cdots\int(\varrho-\varrho_n)\,dx_1\ldots dx_n \\
&\qquad = \int\cdots\int(\varrho-\varrho_n)_+\,dx_1\ldots dx_j - \int\cdots\int(\varrho-\varrho_n)_-\,dx_1\ldots dx_j = 0
\end{aligned}$$

implies that

$$\int \ldots \int |\varrho - \varrho_n| \, dx_1 \ldots dx_j = 2 \int \ldots \int (\varrho - \varrho_n)_+ \, dx_1 \ldots dx_j \to 0$$

due to Lebesgue's dominated convergence theorem as $n \to \infty$.

The proof is finished. ∎

Remark. $Q_n^j(x, \cdot)$ is the distribution of the selection $(Z_1(n), \ldots, Z_j(n))$ and $Q^j(x, \cdot)$ is one of the $(Z_1, \ldots, Z_j)$'s provided that $Z_0 = Z_0(n) = x$. Then Proposition 2 implies that, for every positive integer j and $b < a$,

$$\lim_n P(Z_i(n) \in (b, a) \text{ for every } i \leqslant j) = P(Z_i \in (b, a) \text{ for every } i \leqslant j).$$

Let

$$\tau_n(\omega) = \inf\{j | Z_j(n)(\omega) \notin (b, a)\}, \qquad \tau(\omega) = \inf\{j | Z_j(\omega) \notin (b, a)\}.$$

Also, let

$$\mathrm{OC}_n := P(Z_{\tau_n}(n) \leqslant b) \qquad \text{and} \qquad \mathrm{OC} := P(Z_\tau \leqslant b).$$

PROPOSITION 3. *If* $P(\tau_n < \infty) = P(\tau < \infty) = 1$, *then*

$$\lim_n \mathrm{OC}_n = \mathrm{OC}.$$

Proof. Let

$$p_j(n) = P(Z_i(n) \in (b, a) \text{ for } i < j \text{ but } Z_j(n) \leqslant b),$$

$$q_j(n) = P(Z_i(n) \in (b, a) \text{ for } i < j \text{ but } Z_j(n) \geqslant a),$$

and

$$p_j = P(Z_i \in (b, a) \text{ for } i < j,\ Z_j \leqslant b),\ q_j = P(Z_i \in (b, a) \text{ for } i < j,\ Z_j \geqslant a).$$

Then $p_j(n) \to p_j$, $q_j(n) \to q_j$ and the conditions of finiteness imposed on τ_n and τ imply that

$$\sum_{j=1}^{\infty} (p_j(n) + q_j(n)) = \sum_{j=1}^{\infty} (p_j + q_j) = 1.$$

Therefore

$$\sum_{j=1}^{\infty} (|p_j - p_j(n)| + |q_j - q_j(n)|) = 2 \sum_{j=1}^{\infty} ((p_j - p_j(n))_+ + (q_j - q_j(n))_+$$

and the last term converges to zero by dominated convergence.

The remark that $\mathrm{OC}_n = \sum_{j=1}^{\infty} p_j(n)$ and $\mathrm{OC} = \sum_{j=1}^{\infty} p_j$ ends the proof. ∎

Remark. If Z_j is a random walk, then $Z_j(n)$ are random walks too and the conditions from the above proposition are satisfied.

Let us now study the ASN function:

$$\mathrm{ASN} = E(\tau) = \sum_{j=1}^{\infty} jP(\tau = j).$$

But $(\tau = j) = (Z_i \in (b, a)$ for $i < j$ but $Z_j \notin (b, a))$. If we set $a_j := P(Z_i \in (b, a)$ for $i \leqslant j)$ and $a_0 = 1$, then $P(\tau = j) = a_{j-1} - a_j$; hence

$$\mathrm{ASN} = \sum_{j=1}^{\infty} j(a_{j-1} - a_j) = 1 + a_1 + a_2 + \ldots$$

Let

$$a_1(x) := Q(x, (b, a)), \quad \ldots, \quad a_{j+1}(x) = \int 1_{(b,a)}(y) a_j(y) Q(x, dy)$$

and

$$\mathrm{ASN}(x) := \sum_{j=0}^{\infty} a_j(x).$$

Of course that ASN becomes now ASN(0). Remark that

$$a_{j+1}(x) \leqslant \|a_j\| \, a_1(x)$$

($\|\cdot\|$ means again the uniform norm)

$$a_{j+2}(x) = \int_{(b,a)} a_{j+1}(y) Q(x, dy) = \int_{(b,a)} \int_{(b,a)} a(z) Q(y, dz) Q(x, dy) \leqslant \|a_j\| \, \|a_2\|$$

and, in general,

$$\|a_{j+m}\| \leqslant \|a_j\| \, \|a_m\|. \tag{7}$$

Moreover, $a_j(x)$ is a decreasing sequence and $0 \leqslant a_1(x) \leqslant 1$.

Lemma. 1° *If there exists a j such that $\|a_j\| < 1$, then* ASN$(x) < \infty$ *for every x and*

$$\mathrm{ASN}(x) \leqslant (1 + a_1(x) + \ldots + a_j(x))/(1 - \|a_j\|). \tag{8}$$

2° *If* $\sup_x Q(x, (b, a)) < 1$, *then* ASN$(x) < \infty$.

Proof. Let j be an integer number such that $\|a_j\| < 1$. Then inequalities (7) become

$$\|a_{jm+i}\| \leqslant \|a_j\|^m \, \|a_i\|$$

and everything becomes obvious. ∎

Proposition 4. 1° *Let $(Z_j)_j$ be a random walk such that $P(Z_{j+1} - Z_j)^{-1} = F$, $F \neq \varepsilon_0$. Let $Q(x, A) = F(A - x)$ be the kernel of Z. Then* ASN$(x) < \infty$ *for every x.*

2° *Let* $\mathrm{ASN}_n := E(\tau_n)$. *Suppose also that* $F = \varrho\cdot\lambda$ *where* λ *is the Lebesgue measure on the real line and* ϱ *is a continuous probability density. Then* $\mathrm{ASN}_n \to \mathrm{ASN}$ *as* $n \to \infty$.

Proof. 1° In this case $Q(x,f) = \int f(x+y)\,dF(y)$ and $a_1(x) = F\big((b-x,\ a-x)\big)$,

$$a_j(x) = P(Z_i \in (b-x,\ a-x) \text{ for } i \leqslant j) = \int_{(b,a)} Q(x, dx_1) \int_{(b,a)} Q(x_1, dx_2)\ldots \int_{(b,a)} Q(x_{j-1}, dx_j)$$

At the same time $a_1(x) = P\big(x+Z_1 \in (b,\ a)\big)$ and

$$a_{j+1}(x) = \int 1_{(b,a)}(x+y)\,a_j(x+y)\,dF(y). \tag{9}$$

As we have mentioned before, one knows that

$$\liminf_{j\to\infty} Z_j = -\infty \qquad \text{or} \qquad \limsup_{j\to\infty} Z_j = +\infty \qquad \text{a.s..} \tag{10}$$

This means that $a_j(x)$ tends to zero for every x as r tends to infinity. According to the previous lemma, the only thing one must check is that there exists a j such that $\|a_j\| < 1$.

Suppose ad absurdum that $\|a_j\| = 1$ for every integer j. Then for every integer j and, for every $\delta > 0$ except an arbitrarily small one there exists a real number $x_{j,\delta}$ such that $a_j(x_{j,\delta}) > 1-\delta$. In particular, for every n there exists an x_n such that

$$P\big(Z_j \in (b-x_n,\ a-x_n) \text{ for } j \leqslant n\big) > 1-1/2^n.$$

Let $A_n := \{x \mid a_n(x) > 1-1/2^n\}$; of course $A_n \downarrow \emptyset$. One must study two cases:

a) *The support of F is unbounded.* Then $\|a_1\| < 1$. Indeed, if $\|a_1\| = 1$, there exists a sequence x_n such that $a_1(x_n) > 1-1/2^n$. But $y > a-b \Rightarrow a_1(x_n \pm y) < 1/2^n$; hence it turns out that the set $\{x_n \mid n \geqslant n_0\}$ is bounded for some n_0 and the Borel–Cantelli lemma says that $F\big(\liminf(b-x_n,\ a-x_n)\big) = 1$, which contradicts the unboundedness of the support of F. Now the 1° is a simple consequence of the previous lemma 2°.

b) *The support of F is compact.* Then $A_n \subset A_1 \subset \mathrm{supp}(a_1)$, which is also compact, hence $\{x_n \mid n \geqslant 1\}$ is bounded and we can pick up a convergent subsequence $(x_{n_j})_j$ which converges to some x. For every fixed j and arbitrary m we have

$$P\big(Z_{m_i} \in (b-x_{n_{j+m}},\ a-x_{n_{j+m}}) \text{ for } i \leqslant j\big)$$
$$\geqslant P\big(Z_{n_i} \in (b-x_{n_{j+m}},\ a-x_{n_{j+m}}),\ i \leqslant j+m\big) = a_{n_{j+m}}(a_{n_{j+m}}) \geqslant 1-1/2^{n_{j+m}}.$$

An easy application of the Borel–Cantelli lemma gives (for m tending to infinity):

$$P\big(\omega \mid Z_{n_i} \in (b-x_{n_{j+m}},\ a-x_{n_{j+m}}),\ i \leqslant j \text{ for } m \geqslant \text{some } m_0(\omega)\big) = 1,$$

which further implies that

$$P(Z_{n_i} \in [b-x', a-x'] \text{ for every } i \leq j) = 1 \qquad \text{for every } j,$$

and this equality is impossible because it contradicts (10). We must admit that there exists a j such that $\|a_j\| < 1$.

2° If $F = \varrho \cdot \lambda$ and ϱ is a continuous density, then $Q(x, dy) = \varrho(y-x)\,dy$ and $x \to \varrho(y-x)$ is continuous; moreover, $Q_n(x, A) = F_n(A-x)$ with $F_n = \sum_i F(\Delta_i^n)\,\varepsilon_{\xi_i^n}$ and

$$a_{1n}(x) = \int_{(b,a)} Q_n(x, dy) = F_n((b-x, a-x)), \ldots, a_{j+1,n}(x) = \int_{(b,a)} a_{j,n}(y)\,Q_n(x, dy).$$

We see that $\lim_n a_{j,n}(x) = a_j(x)$ for every positive integer j according to Proposition 2. We claim now that the convergence is uniform in x for every j. Indeed, if $j = 1$,

$$\|a_1 - a_{1,n}\| = \sup_x |F - F_n|((b-x, a-x)) \leq 2 \sup_i F(\Delta_i^n)$$

and the last quantity converges to 0 because F is absolutely continuous with respect to the Lebesgue measure λ. For an arbitrary j observe that

$$\begin{aligned}\|a_{j+1} - a_{j+1,n}\| &= \Big| \int_{(b,a)} a_j\, dF*\varepsilon_x - \int_{(b,a)} a_{j,n}\, dF_n*\varepsilon_x \Big| \\ &\leq \int_{(b,a)} |a_j - a_{j,n}|\, dF*\varepsilon_x + \int_{(b,a)} a_{j,n}\, d|F*\varepsilon_x - F_n*\varepsilon_x| \\ &\leq \|a_j - a_{j,n}\| + \|a_1 - a_{1,n}\|.\end{aligned}$$

The last inequality shows, by recurrence, that $\lim_n \|a_{j+1} - a_{j+1,n}\| = 0$ provided that $\lim_n \|a_j - a_{j,n}\| = 0$.

Let $\varepsilon > 0$ be arbitrarily small and j such that $\|a_j\| < \varepsilon$.

Also, let n_0 be large enough to ensure that $n \geq n_0 \Rightarrow \|a_{j,n}\| < \varepsilon$. Then

$$\begin{aligned}|\mathrm{ASN}(x) - \mathrm{ASN}_n(x)| &\leq \sum_{i=0}^{j} |a_i(x) - a_{i,n}(x)| + \sum_{j+1}^{\infty} |a_i(x) - a_{i,n}(x)| \\ &\leq \sum_{i=0}^{j} |a_i(x) - a_{i,n}(x)| + \big(2\varepsilon/(1-\varepsilon)\big) \sum_{i=0}^{j} \big(a_i(x) + a_{i,n}(x)\big).\end{aligned}$$

Therefore

$$\limsup_n |\mathrm{ASN}(x) - \mathrm{ASN}_n(x)| \leq \frac{2\varepsilon}{1-\varepsilon} \sum_{i=0}^{j} \big(a_i(x) + a_{i,n}(x)\big) \leq \frac{2\varepsilon \cdot 2\,\mathrm{ASN}(x)}{1-\varepsilon}$$

and we are done because ε is arbitrary. ∎

It turns out that we can approximate ASN by ASN_n and OC by OC_n if the assumptions from the above proposition are fulfilled. The precision of the approximation procedure must be studied according to every particular case.

Now if $Z_0 = 0$, $Z_1, \ldots, Z_n$ is a Markov chain, $b < 0 < a$ and $\tau(\omega) = \inf\{j | Z_j \notin (b, a)\}$ we can construct the process

$$Y_i(\omega) = \begin{cases} X_i(\omega) & \text{if } i < \tau(\omega), \\ b & \text{if } i \geqslant \tau(\omega) \text{ and } Z_{\tau(\omega)}(\omega) \leqslant b, \\ a & \text{if } i \geqslant \tau(\omega) \text{ and } Z_{\tau(\omega)}(\omega) \geqslant a. \end{cases}$$

PROPOSITION 5. *$(Y_j)_j$ is a Markov chain and b, a are absorbing states.*

Proof. The proof relies on the following lemma. ∎

LEMMA. *Let $(X_n)_n$ be a Markov chain, $X_n\colon \Omega \to E$, and $(E, \mathscr{F})$ an arbitrary measurable space. Let $\Delta \in \mathscr{F}$, $\delta \notin E$ and $\tau(\omega) := \inf\{n | X_n \in \Delta\}$. Then the process $Y_n = X_n 1_{(n<\tau)} + \delta 1_{(n \geqslant \tau)}$ is a Markov chain having $(E \setminus \Delta) \cup \delta$ as a state space.*

Proof of the lemma. Observe that $\{\tau > n\} = (X_1 \notin \Delta_1, \ldots, X_n \notin \Delta_n) = (Y_n \neq \delta)$ belongs both to the σ-algebra generated by $X_1, \ldots, X_n$ and to the one generated by Y_n. On the other hand, if $X, Y\colon \Omega \to (E, \mathscr{F})$ are two arbitrary random variables and $A \subset (X = Y)$, $A \in \sigma(X) \cap \sigma(Y)$, it is easy to prove that for every integrable random variable $f\colon \Omega \to R$ we have the equality

$$E(f|X) 1_A = E(f|Y) 1_A. \tag{11}$$

Let B be a measurable set from $E \cup \delta$. Then two cases are possible: either $B \in \mathscr{F}$ and $\delta \notin B$, or $B = B' \cup \delta$, $B' \in \mathscr{F}$ and $\delta \notin B'$. Therefore it is sufficient to compute $P(Y_{n+1} \in B | Y_1, \ldots, Y_n)$ in the first case and $P(Y_{n+1} = \delta | Y_1, \ldots, Y_n)$ in the second. If $\delta \notin B$, we have:

$$\begin{aligned} P(Y_{n+1} \in B | Y_1, \ldots, Y_n) &= P(Y_{n+1} \in B,\ Y_n \neq \delta | Y_1, \ldots, Y_n) \\ &= P(Y_{n+1} \in B | Y_1, \ldots, Y_n) 1_{(Y_n \neq \delta)} = P(X_{n+1} \in B | Y_1, \ldots, Y_n) 1_{(Y_n \neq \delta)} \\ &= P(X_{n+1} \in B | X_1, \ldots, X_n) 1_{(Y_n \neq \delta)} = Q(X_n, B) 1_{(Y_n \neq \delta)}. \end{aligned}$$

In the last but one equality we used (11) with $A = (Y_n \neq \delta) = (\tau > n)$, $(X_1, \ldots, X_n) := X$, $Y := (Y_1, \ldots, Y_n)\colon \Omega \to E^n$. Of course $A \subset (X_1 = Y_1, \ldots, X_n = Y_n)$.

Moreover

$$\begin{aligned} &P(Y_{n+1} = \delta | Y_1, \ldots, Y_n) \\ &\quad = P(Y_{n+1} = \delta,\ Y_n = \delta | Y_1, \ldots, Y_n) + P(Y_{n+1} = \delta,\ Y_n \neq \delta | Y_1 \ldots Y_n). \end{aligned}$$

But

$$\begin{aligned} &P(Y_{n+1} = \delta,\ Y_n \neq \delta | Y_1, \ldots, Y_n) = P(X_{n+1} \in \Delta | Y_1, \ldots, Y_n) 1_{(Y_n \neq \delta)} \\ &\quad = P(X_{n+1} \in \Delta | X_1, \ldots, X_n) 1_{(Y_n \neq \delta)} = Q(X_n, \Delta) 1_{(Y_n \neq \delta)} = Q(Y_n, \delta) 1_{(Y_n \neq \delta)} \end{aligned}$$

and

$$P(Y_{n+1}=\delta|Y_1,\dots,Y_n)\,1_{(Y_n=\delta)}=1_{(Y_n=\delta)}.$$

In short

$$P(Y_{n+1}\in B|Y_1,\dots,Y_n)=\begin{cases}Q(Y_n,B)\,1_{(Y_n\neq\delta)} & \text{if } \delta\in B,\\ Q(Y_n,B'\cup\Delta)\,1_{(Y_n\neq\delta)}+1_{(Y_n=\delta)} & \text{if } B=B'\cup\delta,\end{cases}$$

that is, Y_n is a Markov chain. Its kernel is

$$\bar{Q}(y,B)=\begin{cases}Q(y,B)\,1_{\Delta^c}(y) & \text{if } \delta\notin B,\\ Q(y,B'\cup\Delta)\,1_{\Delta^c}(y)+1_{\{\delta\}}(y) & \text{if } B=B'\cup\delta.\end{cases}$$

To come back to our case, the kernel of Y is

$$\begin{aligned}\tilde{F}(x,y)&=:P(Y_{n+1}\leqslant y|Y_n=x)\\ &=\begin{cases}1_{[b,\infty)}(y) & \text{if } x=b,\\ F(x,y)\,1_{[b,a)}(y)+1_{[a,\infty)}(y) & \text{if } b<x<a,\\ 1_{[a,\infty)}(y) & \text{if } x=a\end{cases}\end{aligned}$$

where $F(x,y)=Q\big(x,(-\infty,y]\big)$.

The proof is finished. ∎

As a result of the discretization made in Proposition 2 we have obtained in fact a sequence of Markov chains $Z(n)$ having the set $E_n=:(\xi_i^n)_i$ as state spaces. Being discrete, their kernels are the matrices $q_{ij}(n)=Q(\xi_i^n,\Delta_j^n)$. After applying the transformations pointed out before Proposition 5, we obtain a sequence of Markov chains with two absorbing states b and a and $n-1$ transient states $\xi_1^n,\dots,\xi_{n-1}^n$; $b=\xi_0^n$, $a=\xi_n^n$. Let P_n be their transition matrices, $P_n=\big(p_{ij}(n)\big)_{i,j\leqslant n}$, their expression being:

$$p_{ij}(n)=\begin{cases}\delta_{ij} & \text{if } i=0 \text{ or } i=n,\\ q_{ij}(n) & \text{if } i,j\leqslant n-1,\ i,j\geqslant 1,\\ \sum\limits_{j\leqslant 0} q_{ij}(n) & \text{if } j=0,\ 1\leqslant i\leqslant n-1,\\ \sum\limits_{j\geqslant n} q_{ij}(n) & \text{if } j=n,\ 1\leqslant i\leqslant n-1.\end{cases}$$

The matrix P_n max be partitioned as follows:

$$\begin{array}{c|c|cccc|c}
 & \xi_0^n & \xi_1^n & \xi_2^n \cdots & & \xi_{n-1}^n & \xi_n^n\\
\xi_0^n & 1 & 0 & 0 \cdots & & 0 & 0\\
\hline
\xi_1^n & p_{1,0}(n) & p_{1,1}(n) & p_{1,2}(n) \cdots & & p_{1,n-1}(n) & p_{1,n}(n)\\
\xi_2^n & p_{2,0}(n) & p_{2,1}(n) & p_{2,2}(n) \cdots & & p_{2,n-1}(n) & p_{2,n}(n)\\
\vdots & \vdots & \vdots & T_n=\big(q_{ij}(n)\big)_{i,j\geqslant 1} & & \vdots & \\
\xi_{n-1}^n & p_{n-1,0}(n) & p_{n-1,1}(n) & \cdots & & p_{n-1,n-1}(n) & p_{n-1,n}(n)\\
\hline
\xi_n^n & 0 & 0 & \cdots & & 0 & 1
\end{array}$$

The functions OC_n may be interpreted as the probabilities that the chain is absorbed by the state b, and ASN_n as the expected time up to the absorption. Their expressions are known from the theory of finite Markov chains. Namely, since we supposed that $Y_0(n) = 0$ for every n, we have

$$P(Y_1(n) = \xi_i^n) = Q(0, \Delta_i^n) \qquad \text{for} \quad i = 1, \dots, n-1,$$
$$P(Y_1(n) = b) = Q(0, (-\infty, b]),$$
$$P(Y_1(n) = a) = Q(0, [a, \infty)).$$

Collect the first $n-1$ probabilities in a row vector Π_n. Also let V_n be the column vector $V_n(i) = p_{i,0}(n)$, $i = 1, \dots, n-1$, let I_n be the $n-1$ dimensional unity matrix and let S_n be an $(n-1)$-dimensional column vector filled with I's. Then.

PROPOSITION 6.

$$OC_n = \Pi_n(I_n - T_n)^{-1} V_n, \qquad ASN_n = \Pi_n(I_n - T_n)^{-1} S_n.$$

The proof is given for instance in [4]. Moreover, the moments of τ_n are computable by a recurrence formula ([4]) and, at least in the case of a random walk with good density of transition, they converge to the moments of τ.

References

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.
[2] K. L. Chung, *A Course in Probability Theory*, 2nd. ed., Academic Press, New York 1974.
[3] B. K. Ghosh, *Sequential Tests of Statistical Hypothesis*, Addison-Wesley, Reading, 1970.
[4] M. Iosifescu, *Markov Processes and Their Applications*, Wiley, New York 1981.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981