On the Estimation of an Eigenvalue by an Additive Functional of a Stochastic Process, With Special Reference to the Kac-Donsker Method[1]

R. Fortet

A "Monte Carlo" method is described for the determination of the eigenvalues and the Fredholm determinant of certain Fredholm integral equations with positive kernel $\Gamma(t,\tau)$. The method is based on a theorem by Kac and Siegert. An appropriate stochastic process is constructed from a Poisson process, for the case that $\Gamma(t,\tau)$ depends on $t-\tau$ only.

The second part of the paper contains a discussion of the various errors inherent in the method of Donsker and Kac for the determination of the lowest eigenvalue of Schrodinger's equation.

1. Introduction

Kac and Donsker [1, 2][2] have given a "Monte Carlo" method for estimating the smallest eigenvalue of a linear operator, when this operator is of a certain type. The starting point of their method is to consider an additive functional of a Wiener-Levy process. In what follows we intend to give: 1°) a different method (but which also consists of considering an additive functional of a random process) of estimating the smallest eigenvalues of some integral equations with kernals of nonnegative type; 2°) some remarks on the Kac-Donsker method.

2. Integral Equations With Positive Definite Kernel

For this first part, the following theorem will be fundamental. *Theorem:* Let $X(t)$ be a real, Laplacian[3] rf,[4] defined for a $<t<b$ (a and b finite) and the covariance $\Gamma(t,\tau)$ of which is a continuous function of (t,τ) on the domain $(a\leqq t\leqq b,\ a\leqq\tau\leqq b)$. Let us consider the rv:[4]

$$Y=\int_a^b X^2(t)\,dt \tag{1}$$

and the integral equation:

$$f(t)-\lambda\int_a^b \Gamma(t,\tau)f(\tau)\,d\tau=g(t). \tag{2}$$

If $D(\lambda)$ is the Fredholm's determinant of the equation (2), the cf[5] $\phi(v)$ of Y is equal to $D(2iv)^{-\frac{1}{2}}$.

This theorem was stated by Kac and Siegert [3,4] (Kac gave only some weaker results, but the generalization is obvious; we gave a proof of the general theorem in [5]). It is easy to give assumptions under which the theorem is valid if $a=-\infty$, or $b=+\infty$, or $a=-\infty$ and $b=+\infty$; also it follows from a paper by Kac [4] that the theorem remains valid if, instead of (1) and (2), we consider the rv:

$$\int_a^b h(t)X^2(t)\,dt, \qquad h(t)\geqq 0$$

and the integral equation:

$$f(t)-\lambda\int_a^b \frac{\Gamma(t,\tau)}{\sqrt{h(t)h(\tau)}}f(\tau)\,d\tau=g(t).$$

But for the principle of the method, it will be sufficient to restrict ourselves to the above statement.

Principle of the method: We consider an integral equation (2), with a continuous kernel of nonnegative type $\Gamma(t,\tau)$, and we would like to estimate its smallest eigenvalues, and more generally its Fredholm's determinant $D(\lambda)$. Now $\Gamma(t,\tau)$, being of nonnegative type, may be considered as a covariance of a Laplacian process $X(t)$, which is entirely determined (see [8]) by $\Gamma(t,\tau)$. We assume that some random game has been set up that implies a realization of $X(t)$ and, consequently, of Y, as defined by (1). We make n independent trials, obtaining n values $y_1, y_2, \cdots, y_n$ of Y; from these y_j's, we can deduce the following fr[5] $G_n(y)$:

$$G_n(y)=\frac{1}{n}\times[\text{number of those } y_j\text{'s which are}<y];$$

and it is well known that $G_n(y)$ is an estimate of the fr $G(y)$ of Y. Hence, we have an estimate $\phi_n(v)$ of c f $\phi(v)$ of Y by:

$$\phi_n(v)=\int_0^{+\infty} e^{ivy}dG_n(y) \tag{3}$$

(the integral is extended from 0 only to $+\infty$ because Y is $\geqq 0$); but (3) is equivalent to:

$$\phi_n(v)=\frac{1}{n}\sum_j e^{iv y_j} \tag{3}'$$

[1] The preparation of this paper was sponsored (in part) by the Office of Naval Research.
[2] Figures in brackets indicate the literature references at the end of this paper.
[3] This is, Gaussian.
[4] rf, random function; rv, random variable.
[5] cf, characteristic function; fr, function of repartition (i. e., cumulative distribution function).

Now $\phi(v)$ is the me[6] of the rv $Z(v)=e^{ivY}$; hence we obtain an estimate $D_n(\lambda)$ of $D(\lambda)$ by the preceding theorem, by putting:

$$D_n(2iv)=\frac{1}{\phi_n^2(v)}, \tag{4}$$

However, v being real in (3), (4) gives an approximation of $D(\lambda)$ only for the values of λ that are purely imaginary; and the roots of $D(\lambda)$, which are real and positive, are not obtained by this procedure. But we can operate in the two following ways:

(A) Under the preceding assumptions, $D(\lambda)$ is an entire function of genus at most 1; hence $D(\lambda)$ is representable by an entire series:

$$D(\lambda)=a_0+\frac{a_1}{1!}\lambda+\ldots+\frac{a_k}{k!}\lambda^k+\ldots, \tag{5}$$

where the a_k's are the derivatives of $D(\lambda)$ for $\lambda=0$. The formula:

$$\phi(v)=D(2iv)^{-1}$$

shows that $\phi(v)$ is indefinitely differentiable for v close to 0 and that the a_k's can be deduced from the derivatives of $\phi(v)$ for $v=0$. These derivatives are equal to $i^1M_1, i^2M_2, \ldots, i^kM_k, \ldots$, where the M_k are the moments of Y, and these moments may be estimated, by some well-known statistical procedures, from the y_j's; hence we can obtain estimates a_k^* of the a_k's and an approximate representation $D_n(\lambda)$ of $D(\lambda)$ by:

$$D_n(\lambda)=\sum_k \frac{a_k^*}{k!}\lambda^k.$$

But it is well known by statisticians that, if n is not very large, it is difficult to obtain good estimates of M_k for $k>8$; it will be necessary, in general, to adopt an approximate representation of $D(\lambda)$ by a polynomial of the following type:

$$D_n^*(\lambda)=\sum_{k=1}^{s}\frac{a_k^*}{k!}\lambda^k \tag{6}$$

with $s\leqq 8$. But this seems to be sufficient in some cases.[7] For instance, in order to estimate the 2 or 3 lowest roots $\lambda_1, \lambda_2, \lambda_3, \ldots$ of $D(\lambda)$, we can obtain numerically the lowest roots $\lambda_1^*, \lambda_2^*, \ldots$ of (6), and these λ_j^* may be considered as good approximations for $\lambda_1, \lambda_2, \ldots$

(B) We can also employ the following procedure: $D(\lambda)$ being an entire function of genus at most 1, there are two positive numbers A and ρ such that:

$$|D(\lambda)|\leqq Ae^{\rho'|\lambda|}$$

for every λ and every $\rho'>\rho$; hence, the function

$$\Delta(s)=\sum_k \frac{a_k}{s^{k+1}}$$

considered as a series in $\frac{1}{s}$, has a radius of convergence $\frac{1}{\rho}>0$, and

$$\Delta(s)=\int_{-\infty}^{+\infty}e^{s\lambda}D(\lambda)d\lambda$$

if $\mathscr{R}(s)>\rho>0$; hence for every λ (and particularly for λ real and >0), we have:

$$D(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-\lambda(\alpha+i\beta)}\Delta(\alpha+i\beta)d\beta \tag{7}$$

for any fixed real $\alpha>\rho$; if λ is real and <0, the integral:

$$\phi\left(-i\frac{\lambda}{2}\right)=\int_0^{+\infty}e^{\frac{\lambda y}{2}}dG(y) \tag{8}$$

has a meaning, and, as a consequence of the preceding theorem, we have:

$$D(\lambda)=\phi\left(-i\frac{\lambda}{2}\right)^{-2} \qquad \text{for } \lambda \text{ real and } <0 \tag{9}$$

Hence we can obtain, by a statistical procedure, an estimate of $\phi\left(-i\frac{\lambda}{2}\right)$ by (8), then an estimate of $D(\lambda)$ ($\lambda<0$) by (9), then an estimate of $\Delta(\alpha+i\beta)$ by (6), and finally an estimate of $D(\lambda)$ for $\lambda>0$ (or for any λ) by (7). There are two numerical integrations [(6) and (7)] to be performed, and this procedure does not seem to be of practical interest.

Realization of the game: Another difficulty lies in the practical realization of $X(t)$. This question is also interesting from a theoretical point of view. It may happen that there is an obvious procedure for this realization, with a sufficiently close approximation. This happens for instance if $a=0$, $b>0$ and if $\Gamma(t,\tau)=\min(t,\tau)$; in this case, $X(t)$ is a Wiener-Levy process (with $X(0)=0$, $0\leqq t<b$) and one can see in [2] how it is possible to realize (approximately) $X(t)$.[8]

In many cases it is possible to reduce $X(t)$ to a Wiener-Levy process, as for instance if $X(t)$ is a Markoff process (see [5, p. 198]); that happens if $\Gamma(t,\tau)=e^{-k|\tau-t|}$, where k is any constant. But in general, for a given $\Gamma(t,\tau)$, we do not know if $X(t)$ is or is not a Markoff process (to date, there is no general theorem about this). On the other hand, the reduction of $X(t)$ to a Wiener-Levy process needs some computation which, although easy to perform, may be lengthy.

We can look for a realization of $X(t)$ in another direction. First, we mention that, $X(t)$ being a permanent process, it cannot be realized rigorously: we can only obtain a process $X^*(t)$ that is an approximation of $X(t)$. Then too, the game concerns, not Y, but

$$Y^*=\int_a^b X^*(t)\,dt.$$

[6] m. e. mathematical expectation; the mathematical expectation of a rv X is represented by $E(X)$.

[7] There is a good method for obtaining estimates of the lowest roots of a Fredholm determinant when its first coefficients are known.

[8] $X(t)$ is a Wiener-Levy process if the r. v. $X(\tau)-X(t)$ [with $\tau>t$] is independent of the r. v. $X(u)$ for any $u\leqq t$, and if it is a Laplacian r. v. with m. e. equal to 0 and a standard deviation equal to $\sqrt{\tau-t}$. By definition, $\Gamma(t,\tau)$ is the m. e. of the product $X(t)\cdot X(\tau)=X(t)[X(t)+(X(\tau)-X(t))]$, and if $X(0)=0$, $0\leqq t<a$, the m. e. of this is equal to t, that is to say: $\min(t,\tau)$, since $\tau>t$.

This substitution is valid only if we can prove that the fr of Y^* is an approximation of that of Y. But, because this is an intuitive feature (at least under some assumptions), we shall admit it.

Let $N(t)$ be a Poisson's process, homogeneous and with density m; let $R(t,\tau)$ be an ordinary function defined over the domain $\mathfrak{d}_1$: $-\infty<t,\ \tau<+\infty$, and such that, for every t,

$$\int_{-\infty}^{+\infty} R^2(t,\tau)d\tau<+\infty \quad \text{(Lebesgue integral.)}$$

We put:

$$N^*(t)=N(t)-E[N(t)]=N(t)-mt$$

and let $X^*(t)$ be the process defined by:

$$X^*(t)=\lim_{\alpha\to-\infty,\ \beta\to+\infty} \text{mq}\left\{\text{inc}\frac{l}{\sqrt{m}}\int_\alpha^\beta R(t,\tau)dN^*(\tau)\right\}^0 \quad (10)$$

(for the definition of a Poisson process, see for instance [7, p. 212]: for the precise meaning of (10), see [5]). In what follows, we shall call such an $X^*(t)$ a "Poisson's rf".

In general $X^*(t)$ may be simply represented by:

$$X^*(t)=\frac{l}{\sqrt{m}}\left(\sum_j R(t,\tau_j)-m\int_{-\infty}^{+\infty} R(t,\tau)d\tau\right) \quad (10)'$$

where the τ_j's are the jumps of $N(t)$. It is possible, from a collection of random digits, to realize correctly a Poisson's process: hence it is possible to realize a Poisson's rf; in fact, it is possible to think of a device (employing electrical noise, or emission of α- particles, etc. . . .) giving $X^*(t)$ in a physical way.

It has been proved (see, for instance [5]), that, if $m\to+\infty$, $X^*(t)$ tends toward the Laplacian process, the covariance $\Gamma(t,\tau)$ of which is given by:

$$\Gamma(t,\tau)=\int_{-\infty}^{+\infty} R(t,u)\,R(\tau,u)du. \quad (11)$$

Hence, for a given $\Gamma(t,\tau)$, the problem of realizing approximately a Laplacian process $X(t)$ with covariance $\Gamma(t,\tau)$ is solved if we can determine an $R(t,\tau)$ defined over $\mathfrak{d}_1$, with

$$\int_{-\infty}^{+\infty} R^2(t,\tau)\,d\tau<+\infty \qquad \text{for every } t$$

and such that (11) would be satisfied, at least over the following domain $\mathfrak{d}$:

$$a\leqq t,\tau\leqq b.$$

Hence, the first step is the theoretical study of the existence of solutions $R(t,\tau)$ for (11); but our practical aim will be reached only if there is a solution which is easy to determine numerically. We shall consider: first a particular case, and second, the general case.

* mq, in quadratic mean; sc, almost certain (with probability 1); inc means: stochastic integral with probability 1. For definition of these terms, see [7].

(1) Let us assume that there exists a function $\Gamma_1(\tau-t)$ of $(\tau-t)$ only, defined over $\mathfrak{d}_1$, symmetric and of the nonnegative type (over $\mathfrak{d}_1$), and such that $\Gamma(t,\tau)=\Gamma_1(\tau-t)$ over $\mathfrak{d}$. We shall put $h=\tau-t$, $\Gamma_1(\tau-t)=r(h)$; in this case, $X(t)$ is, at least over the interval $(a,\ b)$, a stationary process, and $r(h)$ is a positive definite function [see (8)]. It is sufficient to have a solution of

$$\Gamma_1(\tau-t)=\int_{-\infty}^{+\infty} R(t,u)\,R(\tau,u)\,du \qquad (11)_1$$

over $\mathfrak{d}_1$. It is possible that every solution $R(t,\tau)$ of $(11)_1$ depends on $(\tau-t)$ only, but that is not sure. But it is sufficient to look for this kind of solution; that is to say, to look for (real) functions $R(u)$ such that:

$$\int_{-\infty}^{+\infty} R^2(u)\,du<+\infty \qquad r(h)=\int_{-\infty}^{+\infty} R(u)\,R(u+h)\,du. \quad (11)'$$

The corresponding $X^*(t)$ will be stationary itself. We proved in [9] that $(11)'$ has solutions only if $r(h)$ is continuous [hence, $r(h)$ is a cf] and if the spectral function $F(\omega)$ of $r(h)$ is absolutely continuous, that is to say admits a derivative $f(\omega)$; in this case, $R(u)$ is a solution of $(11)'$ if, and only if,

$$R(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sqrt{f(\omega)}\;e^{i\psi(\omega)-i\omega u}\,d\omega$$

[Fourier-Plancherel transform] (12)

where $\psi(\omega)$ is any odd function.[10] We can consider that this result and (12) give a convenient answer to our problem.

(2) The general case is much more difficult, and it seems that the only result is the following theorem, that we proved in [9]. If $\Gamma(t,\tau)$, supposed to be defined over $\mathfrak{d}_1$ for instance, is continuous (as a function of the two variables t,τ) over any bounded domain, there is at least one solution for (11), valid over $\mathfrak{d}_1$; but we do not know any easy way to compute numerically this solution, or any other solution (it is easy to see that, in general, (11) has many, and even infinitely many, solutions). *Conclusion:* The interest of the Monte-Carlo method under consideration here would be that it can give simultaneously several eigenvalues of (2); but it seems possible to perform it only in the case where $\Gamma(t,\tau)$ depends on $(\tau-t)$ only; even in this case, the method is complicated, but it might be interesting to try it.

3. The Kac-Donsker Method

Let us consider the equation:

$$\tfrac{1}{2}\frac{d^2\Psi}{dx^2}-V(x)\,\Psi(x)=-\lambda\Psi(x), \qquad (13)$$

[10] In this paper, $\Gamma(t,\tau)$ and $R(t,\tau)$ are always supposed to be real. On the other hand, it is well known that, if $r(h)$ is a c. f., there is a real nondecreasing function $F(\omega)(-\infty<\omega<+\infty)$, with: $F(-\infty)=0$, $F(+\infty)=1$, and such that:

$$r(h)=\int_{-\infty}^{+\infty} e^{i\omega h}dF(\omega)$$

(Theorem of Bochner).

where λ is a constant; $\Psi(x)$ and $V(x)$ are functions defined over $(-\infty, +\infty)$; $V(x)$ is given and $\geqq 0$:

$$V(x) \geqq 0.^{11} \qquad (14)$$

Under some general assumptions on V, (13) has nonnull solutions only for some positive values λ_1, λ_2, . . ., λ_j, . . . of λ (these λ_j's are the eigenvalues of (13); we assume the λ_j's ordered by increasing values). Kac and Donsker (see [2]; we adhere to the notation of [2]) try to estimate λ_1 (their method can be extended to λ_2, λ_3, . . ., and also to the computation of the corresponding eigenfunctions; but for the discussion of the method, we shall restrict ourselves to the estimation of λ_1), in the following way:

Let $\Psi_j(x)$ be the eigenfunction corresponding to λ_j; that is to say the nonnull solution of (13) for $\lambda=\lambda_j$; we suppose that the Ψ_j's are normalized. We put

$$L(t)=\int_0^t V[X(u)]\,du, \qquad (t \geqq 0)$$

where $X(t)$ is a Wiener-Levy process with $X(0)=0$; also we put

$$Z(s;t)=\exp.[-sL(t)] \qquad (s \geqq 0).$$

Kac proved that:

$$E[Z(1;t)]=\sum_j e^{-\lambda_j t}\Psi_j(0)\int_{-\infty}^{+\infty}\Psi_j(x)\,dx. \qquad (15)$$

If t is large, we have:

$$E[Z(1;t)] \sim e^{-\lambda_1 t}\Psi_1(0)\int_{-\infty}^{+\infty}\Psi_1(x)\,dx.$$

Hence:

$$\lambda_1=\lim_{t\to+\infty}-\frac{1}{t}\log E[Z(1;t)]$$

or:

$$\lambda_1 \sim -\frac{1}{t}\log E[Z(1;t)] \qquad \text{if } t \text{ is large.} \qquad (16)_1$$

We can estimate λ_1 by $(16)_1$, but Kac and Donsker showed that, in order to avoid the use of too large values of t, it is better to consider two different and sufficiently large values of t, t_1 and t_2, and to estimate λ_1 by:

$$\lambda_1 \sim \frac{1}{t_2-t_1}\log\frac{E[Z(1;t_1)]}{E[Z(1;t_2)]}. \qquad (16)$$

Now, if $X_1, X_2, \ldots, X_k, \ldots$ are mutually independent rv with the same fr, each with me equal to 0 and standard deviation equal to 1, we put:

$$S_k=X_1+X_2+\ldots+X_k,$$

$$L_n(t)=\frac{1}{n}\sum_{k<nt}V\left(\frac{S_k}{\sqrt{n}}\right),$$

$$Y_n(t)=\exp.[-L_n(t)]. \qquad (17)$$

[11] It is possible to replace (14) by $V(x) \geqq k$, where k is any constant. In comparison with some other papers on Monte-Carlo methods, for instance by Wasow [1] (see also [12]), it seems that, in order for the method to be applicable, some assumption on V is necessary, but a weaker one than (14) ought to be sufficient. It would be worth while to study this question. We mention that (13) is a Schrödinger's equation.

Kac proved [1] that, under general assumptions on V, if $n \to +\infty$, the fr of $L_n(t)$ tends toward the fr of $L(t)$. It follows that $E[Y_n(t)]$ tends toward $E[Z(1;t)]$ [here, (14) is essential]; hence, we can estimate λ_1 by:

$$\lambda_1 \sim \frac{1}{t_2-t_1}\log\frac{E[Y_n(t_1)]}{E[Y_n(t_2)]} \qquad (16)'$$

if n is sufficiently large. Hence the procedure is the following: t_1, t_2, n being properly chosen, we perform a large number m of independent realizations $Y_n^1(t_1)$, $Y_n^2(t_1)$, . . ., $Y_n^m(t_1)$ and $Y_n^1(t_2)$, $Y_n^2(t_2)$, . . ., $Y_n^m(t_2)$ of $Y_n(t_1)$ and $Y_n(t_2)$; and we estimate $E[Y_n(t_1)]$ and $E[Y_n(t_2)]$ by the experimental values:

$$U^m(t_1)=\frac{1}{m}\sum_{k=1}^{m}Y_n^k(t_1) \qquad U^m(t_2)=\frac{1}{m}\sum_{k=1}^{m}Y_n^k(t_2).$$

For further details on the procedure, see [2].

Hence, we have to consider three errors:

A. A statistical error, arising from the fact that $U^m(t_1)$ and $U^m(t_2)$ are not rigorously equal to $E[Y_n(t_1)]$ and $E[Y_n(t_2)]$;

B. An error caused by the fact that $E[Y_n(t_1)]$ and $E[Y_n(t_2)]$ are not rigorously equal to $E[Z(1;t_1)]$ and $E[Z(1;t_2)]$ [replacement of (16) by (16)']

C. An error caused by the fact that (16) is only an approximation.

There is a fourth error, because the random digits, which we are ultimately obliged to use in the computational procedure, are never perfect random digits; but this error seems to us very small in comparison to A, B, C. In fact, in all the experiments performed to date, of which the author is aware, the results are in good agreement with a hypothesis of perfect randomness of the random digits; consequently, we shall not consider this error in what follows.

Discussion of the errors A, B, C: It will be convenient for the discussion to take a definite example, so we shall take $V(x)=x^2$, because in this case the λ_j's and the ψ_j's are known; but we shall see that some of the conclusions may depend on V. We assume $t_2<t_1$.

C. Error C is the easiest to discuss. It is not connected with probability theory. We can readily estimate the proper order of magnitude for t_1 and t_2: if $V(x)=x^2$, $\lambda_1=0{,}707$, . . ., $\lambda_2=2{,}121$ . . ., $\lambda_j=\frac{2j-1}{\sqrt{2}}$, . . .; if t_2 is about 3 or 4, and (t_1-t_2) about 1 or 2, the *absolute* error is about 1/200; we need relatively large values, as: $t_2=5$, $t_1=8$, to have the error about 1/1000. For further details, see [2]. In what follows, we assume that t_1 and t_2 are definitely chosen.

B. We know almost nothing about error B; when t_1 and t_2 are fixed, it depends on two elements: the fr of the X_k's, and the value of n. Let us assume that the X_k's have the following fr:

$$Pr(X_k=1)=Pr(X_k=-1)=\tfrac{1}{2}.$$

The function $V(x)=x^2$ increases relatively quickly with x; and the expected order of magnitude of $|S_k|$, which is $\sqrt{k}$, also increases; hence we can admit that, in (17) only the S_k's with $k \geqq nt/d$, where d is something like 3, are important; hence it will be necessary, in order that B be small, that the S_k's, for $k \geqq nt/3$ have a fr close to Laplace's fr. From known results, see [7], p. 153—, it follows that we must take $nt \geqq 1000$ for a fair approximation, and $nt \geqq 2000$ for a good approximation. We do not get a precise estimate of the error B by this argument but we see that, in order to be able to take an n which is not too large, it is better to take for the X_k's a symmetric fr; because in this case there is a faster approach to Laplace's law. It is clearly best to take the X_k's (and hence the S_k's) directly with a Laplace's law: this is more complicated from a practical point of view; however it is possible to realize a Laplacian rv with a good approximation. In this case, it is possible to obtain an estimate of the error. For we can suppose that $S_k/\sqrt{n}$ is $X(k/n)$, and if $X_n(t)$ is the rf defined by:

$$X_n(t)=X\left(\frac{k}{n}\right) \text{ for: } \frac{k}{n} \leqq t < \frac{k+1}{n},$$

we have

$$\mathrm{L}_n(t)=\frac{1}{n}\sum_k V\left[X\left(\frac{k}{n}\right)\right]=\int_0^t V[X_n(u)]\,du.$$

If $V(x)=x^2$, we can write:

$$\mathrm{L}(t)-\mathrm{L}_n(t)=\int_0^t [X^2(u)-X_n^2(u)]\,du$$

$$=\int_0^t [X(u)+X_n(u)][X(u)-X_n(u)]\,du;$$

$$E(|\mathrm{L}(t)-\mathrm{L}_n(t)|) \leqq \int_0^t E(|X(u)+X_n(u)||X(u)-X_n(u)|)\,du.$$

By Schwarz's inequality, we have

$$E(|X(u)+X_n(u)||X(u)-X_n(u)|)$$

$$\leqq \sqrt{E\{[X(u)+X_n(u)]^2\}\times E\{[X(u)-X_n(u)]^2\}}$$

$$\leqq \sqrt{\left(u+3\,\frac{k_n}{n}\right)\left(u-\frac{k_n}{n}\right)}$$

where k_n is the largest integer such that: $k_n/n \leqq u$. It follows that:

$$E(|\mathrm{L}(t)-\mathrm{L}_n(t)|) \leqq 2/3\,\frac{t^{3/2}}{\sqrt{n}}$$

Hence;

$$Z(1;t)-Y_n(t)=e^{-\mathrm{L}(t)}-e^{-\mathrm{L}_n(t)}=e^{-\mathrm{L}(t)}[1-e^{\mathrm{L}(t)-\mathrm{L}_n(t)}]$$

$$e^{\mathrm{L}(t)-\mathrm{L}_n(t)}=1+e^{\theta[\mathrm{L}(t)-\mathrm{L}_n(t)]}[\mathrm{L}(t)-\mathrm{L}_n(t)] \qquad 0<\theta<1$$

$$Z(1;t)-Y_n(t)=-[\mathrm{L}(t)-\mathrm{L}_n(t)]\cdot e^{-\mathrm{L}(t)+\theta[\mathrm{L}(t)-\mathrm{L}_n(t)]}$$

$\mathrm{L}(t)$ and $\mathrm{L}_n(t)$ being $\geqq 0$, we have:

$$0 \leqq e^{-\mathrm{L}(t)+\theta[\mathrm{L}(t)-\mathrm{L}_n(t)]} \leqq 1$$

Hence

$$|Z(1;t)-Y_n(t)| \leqq |\mathrm{L}(t)-\mathrm{L}_n(t)|$$

$$|E[Z(1;t)]-E[Y_n(t)]| \leqq 2/3\,\frac{t^{3/2}}{\sqrt{n}}. \qquad (18)$$

An analogous result may be obtained for $V(x)=|x|^\alpha$ with α equals to any nonnegative number; and more generally for a large class of nonnegative functions $V(x)$. For $V(x)=|x|^\alpha$ with $-1<\alpha<0$ (see appendix) it seems more difficult to obtain a limitation like (18).

But (18) gives an upper bound for an absolute error, and we need rather a bound for a relative error; but it seems more difficult to obtain this.

On the other hand, we are not sure that (18) is the least upper bound for the absolute error; about this, we can say two things:

(1) In (18) the orders of magnitude with respect to n and with respect to t seem to be the right orders; hence the absolute error is increasing when t is increasing for a given fixed n. We know that, for error C, we have to take t sufficiently large. With $V(x)=x^2$ the following bad feature appears, which will be called the feature "F_1" in what follows. It is that

$$E[Z(1;t)]=[ch(t\sqrt{2})]^{-1} \text{ [see appendix (23)]}$$

is exponentially decreasing when $t \to +\infty$; hence the relative error is quickly increasing. For instance, if we choose $t=5$ (which is not a very large value), we have

$$E[Z(1;5)] \approx 0{,}043\ \ldots,$$

and if we use (18), we find that we have to take $n \geqq 3000$ in order to have a relative error about 1/100.

F_1 seems to be related to the fact that $V(x)=x^2$ is not bounded as $x \to +\infty$.

(2) From the experiments performed to date, the error B seems smaller than indicated by (18); probably, the coefficient ⅔ in (18) may be replaced by a smaller one; this does not eliminate F_1, but it does perhaps indicate that F_1 is not very important practically.

A. We shall now discuss the probable order of magnitude of the error A, as a function of m; this order, for the relative error, is $\sigma/\mu\sqrt{m}$, where μ is the me of $Y_n(t)$ and σ its standard deviation. We know neither μ nor σ; but (18) shows that μ is close to $E[Z(1;t)]$, if n is large (but that is necessary for B). It is easy to obtain an analogous inequality which shows that σ is close to the standard deviation of $Z(1;t)$; it is possible to avoid this interference of two unknown quantities μ and σ, with a slight modification of our procedure, but it seems sufficient for our purpose to identify μ with $E[Z(1;t)]$ and σ with the standard deviation of $Z(1;t)$.

With $V(x)=x^2$, we know that $\mu=[ch(t\sqrt{2})]^{-1/2}$, that is to say: $\mu\approx 0{,}043$ for $t=5$; we have also [see appendix (21)]:

$$E[Z(1;t)^2]=E[e^{-2L(t)}]=E[Z(2;t)]=(ch\,2t)^{-1/2}$$

that is to say, if $t=5$:

$$E[Z(1;5)^2]=0{,}000953.$$

Hence

$$\sigma^2=E[Z(1;5)^2]-[E[Z(1;5)]]^2\sim 10^{-4}\cdot 75.8$$

$$\frac{\sigma}{\mu}\approx 2.$$

To have a negligible probability of a relative error of more than one percent, we have to take $m=4.10^4$, which is a very large number. The reason is that we encounter a second bad feature, the following feature "F_2":

We have:

$$\mu=E[Z(1;t)]=[ch(t\sqrt{2}]^{-1/2}$$

$$\sigma^2=E[Z(1;t)^2]-[E[Z(1;t)]]^2=(ch\,2t)^{-1/2}-(ch\sqrt{2}\,t)^{-1}.$$

Hence if t is large (in fact, for $t>2$):

$$\mu\sim\sqrt{2}\,e^{-\frac{t}{\sqrt{2}}}\qquad \sigma\sim 2^{\frac{1}{4}}e^{-\frac{t}{2}}$$

$$\frac{\sigma}{\mu}\sim 2^{-\frac{1}{4}}e^{\frac{t}{2}(\sqrt{2}-1)}$$

which tends toward $+\infty$ when $t\to+\infty$.

Another aspect of the same fact is the following: more generally, we take $V(x)=|x|^\alpha$, with $\alpha>-1$ (see appendix); let $G(t;\alpha)$ and $G(\alpha)$ be the fr's of $L(t)$ and $L(1)$; we can write:

$$E[Z(1;t)]=\int_0^{+\infty}e^{-\alpha}\,dG(t;\alpha).$$

But, if $\nu=E[L(1)]$ and if δ is the standard deviation of $L(1)$, we can deduce from a remark in the appendix, and an integration by parts, that:

$$E[Z(1;t)=-G\left(-\frac{\nu}{\delta}\right)+\int_0^{+\infty}G(\beta)e^{-\alpha}d\alpha.$$

With a being any fixed positive number, we put:

$$A(a;t)=\int_0^{at^{1+\frac{\alpha}{2}}}G(\beta)e^{-\alpha}d\alpha$$

$$B(a;t)=\int_{at^{1+\frac{\alpha}{2}}}^{+\infty}G(\beta)e^{-\alpha}d\alpha\leqq e^{-at^{\alpha}}t^{1+\frac{\alpha}{2}}$$

Hence

$$E[Z(1;t)]=-G\left(-\frac{\nu}{\delta}\right)+A(a;t)+B(a;t).$$

We known by [2] that:

$$E[Z(1;t)]=e^{(-\lambda_1+\epsilon_1)t}$$

where $\epsilon_1\to 0$ if $t\to+\infty$. Hence, if $\alpha>0$, we have

$$\lambda_1=-\lim_{t\to+\infty}\frac{1}{t}\log\left[-G\left(-\frac{\nu}{\delta}\right)+A(a;t)\right]$$

and this is valid for any $a>0$; $-G(-\nu/\delta)+A(a;t)$ is depending only on the values of $G(\mu)$ for: $-\nu/\delta\leqq\mu\leqq-\nu/\delta+a$. Hence λ_1 is a local characteristic of $G(\alpha)$, in the neighborhood (and to the right) of the value $\alpha=-\nu/\delta$; hence a good estimation of λ_1 is equivalent is a good *local* statistical estimation of $G(\alpha)$ [implying, for instance, a good estimation of several derivatives of $G(\alpha)$ for $\alpha=-\nu/\delta$]. It is obvious and well known that such an estimation is very difficult.

But this conclusion may become wrong if $\alpha\leqq 0$; that is to say, if $\alpha\leqq 0$, the feature F_2 may disappear.

Conclusion: We can conclude that the Kac-Donsker method gives an *asymptotic* estimation of λ_1; that is to say that we must take t sufficiently large [we saw that values like 4 or 5 are scarcely sufficient]; but when t is large, features F_1 and F_2 imply that n and m have to be very large. The computation will therefore be a lengthy one if even only very nominal accuracy is to be achieved. This is valid for $V(x)=x^2$ and for a large class of some increasing V of the same kind.

But we saw that F_2 may disappear for $V(x)=|x|^\alpha$ with $\alpha<0$; perhaps F_1 may also disappear in such a case, and consequently the method may be much better. Hence, it seems that we have two problems:

(a) To examine if there is a class of V's such as features F_1 and F_2 disappear for V belonging to this class.

(b) To examine if the method can be improved, even when F_1 and F_2 are present, eventually by some change in the method or in the procedure.

Concerning problem (b), we report the following remark by M. Kac: if we consider, instead of $E[Z(1;t)]$,

$$E\{Z(1;t)\,\psi_1[X(t)]\},$$

then he has proved that: [12]

$$E\{Z(1;t)\cdot\psi_1[X(t)]\}=e^{-\lambda_1 t}\psi_1(0). \tag{19}$$

Hence

$$\lambda_1=-\frac{1}{t_1-t_2}\log\frac{Z(1;t_1)\psi_1[X(t_1)]}{Z(1;t_2)\psi_1[X(t_2)]}. \tag{20}$$

Now (20) is no longer an asymptotic result and we can choose t_1 and t_2 as we like. Hence we have the following method: we use $Z(1;t)\,\psi_1[X(t)]$ instead of $Z(1;t)$; and with t_1 and t_2 sufficiently small, F_1 disappears, and also F_2, at least in the case $V(x)=x^2$.

The difficulty is that we have to know ψ_1 in advance. But in practice we need only a rough approximation of ψ_1; it may even be sufficient,

[12] M. Kac will soon publish the proof and some complementary explanations.

practically, to operate, instead of ψ_1, with any function ψ such that

$$\int_{-\infty}^{+\infty} \psi(x)\psi_2(x)\,dx$$

is small. (In this case, (19) is not rigorous, but may be a sufficient approximation). Under these conditions, it seems possible to determine such a ψ by a preliminary rough experiment; it would be interesting to try it, but in any case the Kac-Donsker method became more complicated.

It appears that in general [even if $V(x) \neq x^2$], this procedure will avoid F_1; but in some interesting cases, F_2 still remains. For instance, Kac has studied a three-dimensional case, where (13) is replaced by:

$$\frac{1}{2}\Delta\psi - \frac{1}{r}\psi = -\lambda\psi \qquad (13a)$$

where

$$r = \sqrt{x^2+y^2+z^2}, \qquad V = \frac{1}{r}.$$

It is a complicated case, because (13a) has not only a discrete spectrum, but also a continuous one. However the method can be applied, with a 3-dimensional Wiener-Levy process $[X(t), Y(t), Z(t)]$, and putting

$$L(t) = \int_0^t \frac{1}{\sqrt{X^2(u)+Y^2(u)+Z^2(u)}}\,du.$$

It happens that, with the introduction of ψ_1 as above, we can avoid F_1, but not F_2, in the sense that the ratio σ/μ remains large even for small t. The reason is that for small t, $[X^2(t)+Y^2(t)+Z^2(t)]^{-1/2}$ is very large.

A useful device in many Monte Carlo methods is "importance sampling" which consists in playing the game not with the natural distribution functions, but with some other distribution functions conveniently chosen. But here the game is played with the distribution function of the X_t's, and Kac has proved that this distribution function is practically irrelevant.

The greatest hope seems in the following direction. Considering the case $V(x)=x^2$, for instance, we saw that the problem reduces to a good statistical determination of $G(\alpha)$ for α close to ν/δ. Let m be the total number of samples; let N_m be the number of the samples for which

$$-\frac{\nu}{\delta} \leqq L(1) \leqq -\frac{\nu}{\delta}+a$$

where a is a given positive small number. The determination of $G(\alpha)$, in the neighborhood of $-\nu/\delta$, may be considered good if N_m is greater than a definite number N; we can stop the game for the first m such that

$$N_m > N$$

and, by chance, this may happen for a relatively small m: in other words, we can follow a *sequential procedure*.

On the other hand, the fact that λ_1 is a local characteristic of $G(\alpha)$ in the neighborhood of $-\nu/\delta$ does mean that the knowledge of $G(\alpha)$ for some other values of α cannot give information about λ_1. We can consider the general problem of the statistical analysis of the results with respect to the spectrum of (13); but in the present state of the statistics, there seems to be little hope in this direction.

4. References

[1] M. Kac, On the distribution of certain Wiener functionals, Trans. Am. Math. Soc. **65**, 1 (1949).

[2] M. Kac and M. D. Donsker, A sampling method for determining the lowest eigenvalue and the principal eigenfunction of Schördinger's equation, J. Research NBS **44**, 551 (1950) RP2102.

[3] M. Kac and A. J. F. Siegert, On the theory of noise in radio receivers with square law detectors, J. Applied Phys. **18**, 383 (1947).

[4] M. Kac, On some connections between probability theory and differential and integral equations, Berkeley Second Symposium on Mathematical Statistics and Probability (1950).

[5] Fortet, Random functions derived from a Poisson process, Berkeley Second Symposium on Mathematical Statistics and Probability (1950).

[6] R. Fortet, Les fonctions aléatoires du type de Markoff, J. Math. **22**, 177 (1943).

[7] R. Fortet, Elements de calcul des probabilitées, CNRS ed. (Paris 1950).

[8] M. Loève, Fonctions aléatoires du second ordre, dans P. Levy, Prodessus Stochastiques (Gauthier-Villars, Paris, 1948).

[9] R. Fortet, On some functionals of a Laplacian process, J. Research NBS **48**, 32 (1951) RP 2280.

[10] R. Fortet, Additive functionals of a Markoff process, Ann. Math. (publication pending).

[11] W. Wasow, Random walks and the eigenvalues of elliptic difference equations, J. Research NBS **44**, 65 (1951) RP2176.

[12] G. E. Forsythe and R. A. Leibler, Matrix inversion by a Monte Carlo method, MTAC **IV**, No. 31, 127 (1950).

5. Appendix

$X(t)$ being a Wiener-Levy process, with $X(0)=0$, we consider the following functional:

$$L(t) = \int_0^t |X(u)|^\alpha\,du \qquad (t>0).$$

This stochastic integral has a meaning when $\alpha > -1$, in the following sense: $X(t)$ being ac a continuous function, $|X(t)|^\alpha$ is ac a continuous function, and the integral

$$\int_0^t |X(u)|^\alpha\,du$$

exists, but may be infinite. But if $\alpha > -1$, it is ac finite, because in the first place,

$$E(|X(u)|^\alpha) = \frac{1}{\sqrt{2\pi u}}\int_{-\infty}^{+\infty} |x|^\alpha e^{-\frac{x^2}{2u}}\,dx = K u^{\frac{\alpha}{2}},$$

where

$$K = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} |y|^\alpha e^{-\frac{y^2}{2}}\,dy,$$

and K is $< +\infty$ if $\alpha > -1$. Therefore

$$E[\mathrm{L}(t)]=\int_0^t E|X(u)|^\alpha\,du=K\int_0^t u^{\frac{\alpha}{2}}\,du<+\infty,$$

the first member of the equality following from Fubini's Theorem. In what follows we suppose $\alpha>-1$. Now $\mathrm{L}(t)$ has the same law as $t^{1+\frac{\alpha}{2}}\mathrm{L}(1)$ does, because, if we put $u=tv$, we have

$$\mathrm{L}(t)=t\int_0^1|X(tv)|^\alpha dv$$

and if we consider $X(tv)/\sqrt{t}$, it may be considered as a Wiener-Levy process in respect to v only; hence we have

$$\mathrm{L}(t)=t^{1+\frac{\alpha}{2}}\int_0^1|Y(v)|^\alpha\,dv$$

where $\int_0^1|Y(v)|^\alpha\,dv$ has the same law as $\mathrm{L}(1)$ does.

As an application of the preceding formula, the value of $E[Z(s;t)]$ for $\alpha=2$ will now be computed.

We know, by the theorem of Part I, that the cf $\phi(v)$ of $\mathrm{L}(1)$ is equal to $D(2iv)^{-\frac{1}{2}}$, where $D(\lambda)$ is the Fredholm's determinant of the integral equation

$$f(t)=\lambda\int_0^1\min(t,\tau)\,f(\tau)\,d\tau, \tag{21}$$

since $\Gamma(t,\tau)=\min(t,\tau)$ for a Wiener-Levy process; from (21) we deduce

$$f(t)=\lambda\int_0^t\tau f(\tau)\,d\tau+\lambda t\int_t^1 f(\tau)\,d\tau.$$

Hence $f(0)=0$; then

$$f'(t)=\lambda\int_t^1 f(\tau)\,d\tau$$

hence $f'(1)=0$; then

$$f''(t)=-\lambda f(t). \tag{22}$$

Hence (21) is equivalent to (22), with the boundary conditions $f(0)=f'(1)=0$. The solutions of (22) are

$$f(t)=A\cos\sqrt{\lambda}\,t+B\sin\sqrt{\lambda}t.$$

In order to have $f(0)=f'(1)=0$, we must have $A=0$ and $\lambda=\pi^2/4\ (1+2k)^2$ $(k=0,\ 1,\ 2,\ \ldots)$; hence, putting $\lambda=\mu^2$, we have

$$D(\lambda)=\prod_k\left[1-\frac{\lambda}{\frac{\pi^2}{4}(1+2k)^2}\right]=\prod_k\left[1-\frac{\mu}{\frac{\pi}{2}(1+2k)}\right]=$$

$$\cos\mu=\cos(\sqrt{\lambda}).$$

Therefore

$$\phi(v)=\cos(\sqrt{2iv})^{-\frac{1}{2}}.$$

Now $\phi(v)$ is equal to $E[e^{ivL(1)}]$; hence, in the notation of the preceding pages

$$E[Z(s;1)]=E[e^{-sL(1)}]=\phi(is)=(\cos i\sqrt{2s})^{-\frac{1}{2}}=(\mathrm{ch}\sqrt{2s})^{-\frac{1}{2}}.$$

Now, from an earlier remark, we have

$$E[Z(s;t)]=E[e^{-sL(t)}]=E[e^{-st^2L(1)}]=E[Z(st^2;1)].$$

Hence

$$E[Z(s;t)]=[\mathrm{ch}(t\sqrt{2s})]^{-\frac{1}{2}}. \tag{23}$$

Los Angeles, January 26, 1951.