STATISTICAL ANALYSIS OF CERTAIN TYPES OF RANDOM FUNCTIONS

By H. Hurwitz, Jr. and M. Kac

Cornell University

1. Introduction. In solving certain physical problems (Brownian movements, shot effect) one is often led to the study of superpositions of random pulses. More precisely, one is led to sums of the type

$$F(t) = \sum_{j=1}^{N} f(t - t_j), \tag{1}$$

where N and the t_j's are random variables and a function $P(t)$ is given such that $\int_\Delta P(t)\, dt$ represents the average number of pulses occurring during the time interval Δ.

We propose to give a fairly detailed treatment of those statistical properties of $F(t)$ which may be of interest to a physicist and at the same time pay careful attention to the mathematical assumptions which underly the applications. It may also be pointed out that our results could be applied to the theory of time series.

2. Statistical assumptions and the distribution of N. The statistical assumptions can be formulated as follows:

1. The t_j's form an infinite sequence of independent identically distributed random variables each having $p(t)$ as its probability density.
2. N is capable of assuming the values $0, 1, 2, 3, \cdots$ only, and N is independent of the t_j's.
3. If $M(\Delta; N)$ denotes the number of those t_j's among the first N, which fall within the interval Δ, then for non-overlaping intervals Δ_1 and Δ_2 the random variables $M(\Delta_1; N)$ and $M(\Delta_2; N)$ are independent.

We now state our first theorem.[1]

Theorem 1. *Assumptions* 1, 2, 3 *imply that N is distributed according to Poisson's law, i.e.*

$$\text{Prob}\,\{N = r\} = e^{-h}\frac{h^r}{r!},$$

where $h = \int_{-\infty}^{+\infty} P(t)\, dt$.

[1] For a different approach to Poisson's distribution see W. Feller, *Math. Ann.* 113 (1937) in particular pp. 113–160.

Our proof is based on considerations of characteristic functions. Let $\psi_\Delta(x)$ be 1 if x belongs to the interval Δ and 0 otherwise. Thus

$$M(\Delta; N) = \sum_{k=1}^{N} \psi_\Delta(t_k).$$

From the independence of $M(\Delta_1; N)$ and $M(\Delta_2; N)$ it follows that for every pair of real numbers ξ and η we have[2]

$$E\left[\exp\left\{i\left(\xi \sum_{k=1}^{N} \psi_{\Delta_1}(t_k) + \eta \sum_{k=1}^{N} \psi_{\Delta_2}(t_k)\right)\right\}\right] = E\left[\exp\left\{i\xi \sum_{k=1}^{N} \psi_{\Delta_1}(t_k)\right\}\right] E\left[\exp\left\{i\eta \sum_{k=1}^{N} \psi_{\Delta_2}(t_k)\right\}\right],$$

where $E[x]$ denotes the mathematical expectation, or mean value, of x. Letting $q(r) = \text{Prob}\,\{N = r\}$ and using first the independence of N and the t_j's and then the fact that the t_j's are independent and identically distributed we obtain

$$\sum_{r=0}^{\infty} q(r)(E[\exp\{i(\xi\psi_{\Delta_1}(t) + \eta\psi_{\Delta_2}(t))\}])^r = \sum_{r=0}^{\infty} q(r)(E[\exp\{i\xi\psi_{\Delta_1}(t)\}])^r \sum_{r=0}^{\infty} q(r)(E[\exp\{i\eta\psi_{\Delta_2}(t)\}])^r. \tag{2}$$

An easy calculation gives

$$E[\exp\{i\xi\psi_{\Delta_1}(t)\}] = 1 + (e^{i\xi} - 1)\int_{\Delta_1} p(t)\,dt,$$

$$E[\exp\{i\eta\psi_{\Delta_2}(t)\}] = 1 + (e^{i\eta} - 1)\int_{\Delta_2} p(t)\,dt,$$

$$E[\exp\{i(\xi\psi_{\Delta_1}(t) + \eta\psi_{\Delta_2}(t))\}] = 1 + (e^{i\xi} - 1)\int_{\Delta_1} p(t)\,dt + (e^{i\eta} - 1)\int_{\Delta_2} p(t)\,dt.$$

The last equation follows from the fact that Δ_1 and Δ_2 do not overlap. Putting $\xi = \eta = \pi$, $x = 1 - 2\int_{\Delta_1} p(t)\,dt$, $y = 1 - 2\int_{\Delta_2} p(t)\,dt$, $\varphi(x) = \Sigma q(r)x^r$ we see that (2) yields the functional equation

$$\varphi(x + y - 1) = \varphi(x)\varphi(y). \tag{3}$$

One cannot ascertain that (3) holds for all real x and y. First of all the defining power series of $\varphi(x)$ is not known to converge outside the unit circle and secondly it is not obvious that each pair of real numbers x, y between -1 and 1 is such that non-overlapping intervals Δ_1, Δ_2 exist for which

$$x = 1 - 2\int_{\Delta_1} p(t)\,dt \quad \text{and} \quad y = 1 - 2\int_{\Delta_2} p(t)\,dt.$$

[2] We use the symbol $\bar{R}$ and $E[R]$ interchangeably to denote the average (mathematical expectation) of R.

However, if one restricts oneself to small Δ_1 and Δ_2 the functional equation (3) is seen to hold in a sufficiently small neighborhood of 1. This is sufficient (in view of the analyticity of φ in the unit circle) to determine $\varphi(x)$.

In fact, differentiating (3) first with respect to x and then with respect to y we get

$$\varphi'(x)\varphi'(y) = \varphi''(x + y - 1).$$

Letting $y = 1$ and putting $\varphi'(1) = h$ we have

$$\varphi''(x) = h\varphi'(x),$$

which yields immediately

$$\varphi(x) = Ae^{hx} + B.$$

An entirely elementary reasoning (which employs the fact that $Ae^{hx} + B$ must satisfy (3)) leads to the conclusion that $B = 0$, $A = e^{-h}$ which in turn implies at once that

$$q(r) = e^{-h} \frac{h^r}{r!}.$$

Finally,

$$\int_\Delta P(t)\, dt = E[M(N; \Delta)] = E\left[\sum_{k=1}^{N} \psi_\Delta(t_k)\right]$$
$$= \left(\int_\Delta p(t)\, dt\right) e^{-h} \sum_{r=0}^{\infty} \frac{h^{r+1}}{r!} = h \int_\Delta p(t)\, dt,$$

and therefore

$$\int_{-\infty}^{+\infty} P(t)\, dt = h, \qquad P(t) = hp(t).$$

Since h is the mean value of N (i.e. $\bar{N}$) we shall use $\bar{N}$ instead of h.

3. Fourier coefficients of $F(t)$ and their statistical properties. In physical applications it is often convenient to assume that the "pulse function" $f(t)$ is periodic with period $T (T$ large) and one might therefore restrict oneself to the interval $(0, T)$.

It is furthermore assumed that both $f(t)$ and $P(t)$ are sufficiently smooth[3] so as to justify the formal operations on Fourier series performed below. Since we work in the interval $(0, T)$ we assume that $P(t) = 0$ for $t < 0$ and $t > T$.

Expanding $f(t)$ in a Fourier series in $(0, T)$ we get

$$f(t) \sim \sum_{-\infty}^{\infty} a(\omega_k) \exp(i\omega_k t), \qquad \omega_k = \frac{2\pi k}{T},$$

[3] For instance $f(t)$ and $P(t)$ may be assumed to be of bounded variation. Actually, much less severe restrictions suffice but in investigations of this sort far reaching generality would only impair the exposition.

and thus

$$F(t) \sim \sum_{-\infty}^{\infty} a(\omega_k)b(\omega_k) \exp (i\omega_k t),$$

where

$$b(\omega_k) = \sum_{j=1}^{N} \exp (-i\omega_k t_j).$$

Note that

$$E[\exp (-i\omega t)] = \int_0^T \exp (-i\omega t)p(t)\, dt$$

$$= \frac{\frac{1}{T}\int_0^T \exp (-i\omega t)P(t)\, dt}{\frac{1}{T}\int_0^T P(t)\, dt} = \frac{\rho(\omega)}{\rho(0)} = c(\omega) - is(\omega)$$

and put

$$E[F(t)] = \frac{\bar{N}}{\rho(0)} \sum_{-\infty}^{+\infty} a(\omega_k)\rho(\omega_k) \exp (i\omega_k t) = T \sum_{-\infty}^{+\infty} a(\omega_k)\rho(\omega_k) \exp (i\omega_k t),$$

$$X_k^{(\bar{N})} = \frac{\sum_{j=1}^{N} \cos (\omega_k t_j) - \bar{N}c(\omega_k)}{\sqrt{\bar{N}}},$$

$$Y_k^{(\bar{N})} = \frac{\sum_{j=1}^{N} \sin (\omega_k t_j) - \bar{N}s(\omega_k)}{\sqrt{\bar{N}}}.$$

Thus remembering that $\bar{N} = \int_0^T P(t)\, dt$ we may write

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\bar{N}}} \sim \sum_{-\infty}^{+\infty} a(\omega_k)(X_k^{(\bar{N})} - iY_k^{(\bar{N})}) \exp (i\omega_k t)$$

or

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\rho(0)}} \sim \sqrt{T} \sum_{-\infty}^{\infty} a(\omega_k)(X_k^{(\bar{N})} - iY_k^{(\bar{N})}) \exp (i\omega_k t).$$

We can now state the following:

THEOREM 2. *In the limit as* $\bar{N} \to \infty$ *each* $X_k^{(\bar{N})}$ *(and* $Y_k^{(\bar{N})}$*) is normally distributed with mean* 0 *and variance* $\frac{1}{2} + \frac{1}{2}c(2\omega_k)$ $(\frac{1}{2} - \frac{1}{2}c(2\omega_k))$.

The proof, as usual, is based on the consideration of the characteristic function of $X_k^{(\bar{N})}$.

We have

$$E[\exp\{i\xi X_k^{(\bar N)}\}]$$

$$= \exp\{-i\xi\sqrt{\bar N}\, c(\omega_k)\} \exp(-\bar N) \sum_{r=0}^{\infty} \frac{(\bar N)^r}{r!} \left(E\left[\exp\left\{\frac{i\xi\cos\omega_k t}{\sqrt{\bar N}}\right\}\right]\right)^r$$

$$= \exp\{-i\xi\sqrt{\bar N}\, c(\omega_k)\} \exp(-\bar N) \exp\left\{\bar N E\left[\exp\left\{\frac{i\xi\cos\omega_k t}{\sqrt{\bar N}}\right\}\right]\right\}.$$

In deriving this formula use has been made of the facts that the t_j's are independent and identically distributed, that N is independent of the t_j's and that N is distributed according to Poisson's law. It is now easy to see that as $\bar N \to \infty$ the characteristic function of $X_k^{(\bar N)}$ approaches

$$\exp\{-(\tfrac{1}{4} + \tfrac{1}{4}c(2\omega_k))\xi^2\}$$

uniformly in every finite ξ-interval. This, in view of the continuity theorem for Fourier-Stieljes· transforms, implies our theorem. It should be mentioned that it is tacitly assumed that even though $\bar N = T\rho(0)$ approaches ∞ it does it in such a way that the ratio $\rho(\omega)/\rho(0)$ (and hence $c(\omega)$) remains constant (or more generally, approaches a limit).

By considering the characteristic function of the joint distribution of $X_k^{(\bar N)}$ and $X_l^{(\bar N)}(|k| \neq |l|)$ (or any other pair like, for instance, $X_k^{(\bar N)}$ and $Y_l^{(\bar N)}$, in which case no restriction on k, l is necessary) we are able to prove

THEOREM 3. *In the limit as $\bar N \to \infty$ the distinct Fourier coefficients of $(F(t) - E[F(t)])/\sqrt{\rho(0)}$ are normally correlated (i.e. their joint distribution function is the bivariate normal distribution).*

It is also clear that the higher correlations (i.e. between more than two coefficients) will lead to multivariate normal distributions with coefficients expressible in terms of Fourier coefficients of $P(t)$.

We do not state Theorem 3 in more definite terms because in the next section we shall give a more convenient and useful way of handling correlation properties of our Fourier coefficients.

4. Statistical structure of Fourier coefficients. Let us assume that $P(t) > \gamma > 0$ and that the Fourier series of $P(t)$ converges everywhere.

Expanding $\sqrt{P(t)}$ in a Fourier series in $(0, T)$ we have

$$\sqrt{P(t)} = \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t),$$

and in particular (since $p(t) = P(t)/\bar N$)

$$\sqrt{p(t_j)} = \frac{1}{\sqrt{\bar N}} \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t_j).$$

We can now write

$$
\begin{aligned}
b(\omega_k) &= \sum_{j=1}^{N} \exp(-i\omega_k t_j) = \sum_{j=1}^{N} \frac{\exp(-i\omega_k t_j)}{\sqrt{p(t_j)}} \sqrt{p(t_j)} \\
&= \frac{1}{\sqrt{\bar{N}}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l) \left\{ \sum_{j=1}^{N} \frac{\exp(i(\omega_l - \omega_k)t_j)}{\sqrt{p(t_j)}} \right\} \\
&= \frac{1}{\sqrt{\bar{N}}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} \right\} \\
&= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k)\sigma(-\omega_l) + \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{\bar{N}}} \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - T\sigma(-\omega_l) \right\} \\
&= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k)\sigma(-\omega_l) \\
&\qquad + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\sigma(-\omega_l) \right\}.
\end{aligned}
$$

Put $\sigma(\omega) = \alpha(\omega) + i\beta(\omega)$, note that by Parseval's relation

$$
\rho(\omega_k) = \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)\sigma(-\omega_l),
$$

and introduce random variables $U_l^{(\bar{N})}$ and $V_l^{(\bar{N})}$ by means of the formulas

$$
U_l^{(\bar{N})} = \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\,\alpha(\omega_l),
$$

$$
V_l^{(\bar{N})} = \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\sin(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\,\beta(\omega_l).
$$

Thus

$$
b(\omega_k) = T\rho(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)(U_l^{(\bar{N})} - iV_l^{(\bar{N})})l
$$

and we have the following theorem.

THEOREM 4. *In the limit as $\bar{N} \to \infty$ the random variables $U_0^{(\bar{N})}$, $U_1^{(\bar{N})}$, $V_1^{(\bar{N})}$, $U_2^{(\bar{N})}$, $V_2^{(\bar{N})}$, $\cdots$ are independent and normally distributed (each with mean 0 and variance $\frac{1}{2}$).*

This theorem can be proved in a manner exactly analogous to that of Theorem 2. We need only consider the characteristic functions of the joint distributions of U's and V's and treat them in the same way as we treated the characteristic function of the distribution of a single X in the proof of Theorem 2. One thing, however, should be strongly emphasized. The proof of independence (in the limit as $\bar{N} \to \infty$) of $U_l^{(\bar{N})}$ and $U_m^{(\bar{N})}$ ($|l| \neq |m|$), for instance, depends on proving that

$$
E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = 0.
$$

This in turn depends essentially on the fact that N is distributed according to Poisson's law.

In fact,

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = \frac{1}{\bar{N}T} E\left[\left(\sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^{N} \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right] - T\alpha(\omega_l)\alpha(\omega_m).$$

But

$$E\left[\left(\sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^{N} \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right]$$

$$= E\left[\sum_{j=1}^{N} \frac{\cos \omega_l t_j \cos \omega_m t_j}{p(t_j)}\right] + E\left[\sum_{j\neq k} \frac{\cos \omega_l t_j \cos \omega_m t_k}{\sqrt{p(t_j)}\sqrt{p(t_k)}}\right]$$

$$= E\left[\frac{N(N-1)}{(\bar{N})^2}\right] T\alpha(\omega_l)\alpha(\omega_m),$$

and finally

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = \left(\frac{\bar{N^2} - \bar{N}}{(\bar{N})^2} - 1\right) T\alpha(\omega_l)\alpha(\omega_m).$$

Since for Poisson's distribution $\bar{N^2} = \bar{N} + (\bar{N})^2$ we get

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = 0.$$

Also the proof that $E[\,|\,U_l^{(\bar{N})}\,|^2] = \frac{1}{2}$ employs essentially the fact that N is distributed according to Poisson's law.

In view of Theorem 4 we can restate Theorem 3 in a form which is both useful and illuminating inasmuch as it describes completely the statistical structure of the $b(\omega_k)$'s and hence of the Fourier coefficients of $F(t)$.

THEOREM 5. *For the purposes of finding correlations between the $b(\omega_k)$'s it suffices to replace each $b(\omega_k)$ (in the limit as $\bar{N} \to \infty$) by its "statistical representation"*

$$T\rho(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)A_l,$$

where A_{-l} is the complex conjugate of A_l, A_0, A_1, A_2, $\cdots$ a sequence of independent complex-valued random variables and each A_k is distributed in such a way that θ_k = arg A_k is uniformly distributed independent of A_k and the density of the probability distribution of $|\,A_k\,|$ is

$$2Ae^{-A^2}, \qquad (A \geq 0).$$

Theorem 5 was proved under the assumption $P(t) > \gamma > 0$. This assumption was needed to validate the convenient artifice of multiplying and dividing by $\sqrt{p(t_j)}$.

However, even in the case when $P(t)$ is not bounded from below by a positive number (it is always true that $P(t) \geq 0$) Theorem 5 remains true. It could be proved by direct but tedious considerations suggested in section 2.

Theorems 4 and 5 can be easily extended to the case when the pulses all have the same shape but may, at random, differ in magnitude. In other words, instead of sum (1) we may consider the sum

$$F(t) = \sum_{j=1}^{N} \epsilon_j f(t - t_j), \tag{4}$$

where the individual pulses are independent and a function $P(\epsilon, t)$ is given such that

$$\int_{\epsilon}^{\epsilon+\Delta\epsilon} \int_{t}^{t+\Delta t} P(\epsilon, t)\, dt\, d\epsilon$$

is the average number of pulses of "amplitude" between ϵ and $\epsilon + \Delta\epsilon$ occurring between t and $t + \Delta t$.

Theorems 4 and 5 still hold provided one replaces the Fourier coefficients of $P(t)$ by those of

$$\int_{-\infty}^{+\infty} \epsilon P(\epsilon, t)\, d\epsilon,$$

and the Fourier coefficients of $\sqrt{P(t)}$ by those of

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{+\infty} \epsilon^2 P(\epsilon, t)\, d\epsilon}.$$

5. Concluding remarks and summary. If one assumes that the number of pulses N in the time interval $(0, T)$ is constant instead of being a random variable obeying Poisson's law, then Theorems 4 and 5 fail. The failure is due to the fact that, for instance $E[U_l^{(\bar{N})} U_m^{(\bar{N})}]$ is no longer 0. However, as $T \to \infty$ the changes in correlation due to assuming N constant become negligible. On the other hand if one assumes that the number of pulses in each of the time intervals $(0, \tau)$, $(\tau, 2\tau)$, $\cdots$ is fixed, the changes in correlations become appreciable. This case can also be treated by the above methods.

The case in which $p(t)$ is independent of time has been considered in various connections by Schottky, Uhlenbeck and Goudsmidt and Rice[4]. Their investigations emphasized the importance and usefulness of the harmonic analysis of random functions.

In conclusion we summarize our results for the case of time-dependent $P(\epsilon, t)$

[4] W. Schottky, *Ann. d. Phys.* 57 (1919) pp. 541–567.

G. E. Uhlenbeck and S. Goudsmidt, *Phys. Rev.* 34 (1929) pp. 145–151.

S. O. Rice, mimeographed notes on mathematical analysis of random noise, as yet unpublished.

The authors are indebted to Mr. Rice for making his notes available to them.

by observing that in applications one may replace $F(t)$ by its "statistical representation"

$$E[F(t)] + \sqrt{T} \sum_{k=-\infty}^{\infty} a(\omega_k) \left\{ \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) A_l \right\} \exp(i\omega_k t), \tag{5}$$

where

$$\omega_k = \frac{2\pi k}{T},$$

$$E[F(t)] = T \sum_{-\infty}^{\infty} a(\omega_k)\rho(\omega_k) \exp(i\omega_k t),$$

$$\int_{-\infty}^{\infty} \epsilon P(\epsilon, t)\, d\epsilon = \sum_{k=-\infty}^{\infty} \rho(\omega_k) \exp(i\omega_k t),$$

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{\infty} \epsilon^2 P(\epsilon, t)\, d\epsilon} = \sum_{k=-\infty}^{\infty} \sigma(\omega_k) \exp(i\omega_k t),$$

and the A_l's are normally distributed complex-valued random variables for which

$$E[A_l] = 0, \qquad E[|A_l|^2] = 1, \qquad A_l^* = A_{-l}.$$

Furthermore, for $l \geq 0$ the A_l's are statistically independent.

Thus

$$F(t) - E[F(t)] \sim \sum_{k=-\infty}^{\infty} \lambda(\omega_k) \exp(i\omega_k t)$$

where the λ's are normally distributed complex-valued random variables obeying the relation

$$E[|\lambda(\omega)|^2] = |a(\omega)|^2 \int_0^T Q(t)\, dt.$$

If $Q(t)$ is periodic with frequency $\frac{\omega_{k_0}}{2\pi}$ then it follows that $\lambda(\omega')$ and $\lambda(\omega'')$ are independent unless $\omega' + \omega''$ or $\omega' - \omega''$ is an integral multiple of ω_{k_0}.
Finally, we mention that $F(t) - E[F(t)]$ is normally distributed with variance $s(t)$ given by the formula

$$s^2(t) = E[(F(t) - E(F(t)))^2] = T \sum_{k=-\infty}^{\infty} \gamma(\omega_k)\mu(\omega_k) \exp(i\omega_k t),$$

where $\gamma(\omega_k)$ is the Fourier coefficient of $Q(t)$ and $\mu(\omega_k)$ the Fourier coefficient of $f^2(t)$.