Korrelationstheorie der stationären stochastischen Prozesse.

Von

A. Khintchine in Moskau.

§ 1.

Definition der stationären stochastischen Prozesse.

Ein *stochastischer Prozeß* ist eine einparametrige Schar x_t $(-\infty < t < +\infty)$ von zufälligen Variablen; zur Kennzeichnung des Prozesses gehört, daß für jeden endlichen Wertevorrat $t_1, t_2, \ldots, t_n$ das n-dimensionale Verteilungsgesetz der entsprechenden Variablen $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ vorgegeben sei; selbstverständlich müssen die so definierten Verteilungsgesetze in ihren Zusammenhängen allen Forderungen der Wahrscheinlichkeitsrechnung genügen.

Während sich die klassische Wahrscheinlichkeitsrechnung ausnahmslos mit diskreten Folgen von zufälligen Variablen beschäftigte, hat sich in der neuesten Zeit die Theorie der stochastischen Prozesse zu einem der wichtigsten Zweige dieser Wissenschaft entwickelt; die mit ihr verbundenen theoretischen Problemkreise eröffneten dem Mathematiker ein äußerst fruchtbares Untersuchungsfeld, während andererseits ihre Ergebnisse zahlreiche Anwendungen, hauptsächlich in verschiedenen Fragen der physikalischen und technischen Statistik, gefunden haben. Es hat sich unter anderem herausgestellt, daß die wichtigsten Verteilungsgesetze, die in der klassischen Wahrscheinlichkeitstheorie als asymptotische Formeln auftraten, in der Theorie der stochastischen Prozesse, wie auch zu erwarten war, die Rolle *exakter* Lösungen der entsprechenden Probleme übernehmen.

Die bisher erfolgten Untersuchungen beziehen sich fast ausnahmslos auf eine bestimmte Klasse von stochastischen Prozessen, der in theoretischer und praktischer Beziehung eine große Bedeutung zukommt. Wegen ihrer Analogie zu den „Markoffschen Ketten“, die unter den Folgen zufälliger Variablen eine hervorragende Rolle spielen, wollen wir sie als *Markoffsche Prozesse* bezeichnen. Das charakteristische Merkmal der Markoffschen Prozesse bildet der Umstand, daß für jeden Zeitpunkt die künftige Entwicklung des Vorganges nur von seinem gegenwärtigen Zustand, nicht aber von seiner Vorgeschichte abhängt. In präziser Weise hat das folgende Bedeutung: sind $t_1 < t_2 < t_3$ drei beliebige Zeitpunkte, a und b zwei beliebige reelle Zahlen, so ist das durch $x_{t_2} = b$ bedingte Verteilungsgesetz von x_{t_3} mit dem durch die vollständigere Angabe $x_{t_1} = a$, $x_{t_2} = b$ bedingten identisch; die durch die Gegenwart gelieferten Voraussagen über die Zukunft können also von einer eventuell hinzutretenden Kenntnis über die Vorgeschichte des betrachteten Vorgangs in keiner Weise beeinflußt werden[1]). Diese Art von Vorgängen, die eine in mathematischer Hinsicht verhältnismäßig leichte Behandlung zuläßt, ist in vielen Anwendungen auch der Wirklichkeit gut angepaßt (radioaktiver Zerfall, Fernsprechverkehr u. dgl. m.); viel zahlreicher sind aber wohl die physikalischen und technischen Fragestellungen, bei denen die Vorgeschichte des Vorgangs für das Urteil über seine künftige Entwicklung eine wesentliche Bedeutung hat und auch näherungsweise nicht vernachlässigt werden darf. So würde z. B. die Auffassung der Lagenänderung von beweglichen Teilchen in Diffusionsvorgängen oder Brownscher Bewegung als eines Markoffschen Prozesses offenbar bedeuten, daß der Trägheit keine Rechnung getragen wird. Allerdings könnte man in diesem Beispiel den Begriff der Gegenwart derart erweitern, daß er außer der Lage noch die Geschwindigkeit des Teilchens enthalten soll, was den erwähnten Übelstand zu beseitigen im Stande wäre. Demgegenüber gibt es aber zahlreiche Beispiele, wo die Einführung noch so vieler Parameter in die Kennzeichnung des gegenwärtigen Zustandes die Sachlage doch nicht verbessert; es ist hier vor allem die *statistische Mechanik* zu erwähnen. Betrachtet man die stationäre Bewegung des Phasenraums eines mechanischen Systems und kennzeichnet den gegenwärtigen Zustand eines Punktes durch die Angabe der Zelle, in welcher er augenblicklich enthalten ist, so lassen sich über seine künftige Bewegung verschiedene Wahrscheinlichkeitsaussagen machen. Jede Kenntnis über die Vorgeschichte des Punktes (d. h. jede Angabe über die Zelle, in welcher er zu einem früheren Zeitpunkt war) verändert aber dieses Wahrscheinlichkeitsurteil sehr wesentlich. Diese Tatsache und auch die durch sie bedingte Unmöglichkeit, die Probleme der statistischen Mechanik mit Hilfe der Markoffschen Prozesse in Angriff zu nehmen, hat Hadamard sehr klar in seinem Vortrag auf dem Kongreß in Bologna 1928 ausgesprochen[2]).

[1]) Die ziemlich verbreitete Ausdrucksweise „durch die Angabe von x_t wird die Verteilung von $x_{t'}$ $(t' > t)$ vollständig definiert" ist selbstverständlich ungenügend; eine richtige Fassung findet man z. B. bei G. Pólya (Sur quelques points de la théorie des probabilités, Ann. Inst. H. Poincaré, 1930) und A. Kolmogoroff (Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Math. Ann. 104 (1931), S. 415.

[2]) Atti del Congr. Intern. dei Matem. Bologna 1928, B. V, S. 133—139; insbesondere S. 138.

Unter den Vorgängen, in denen die Vorgeschichte ein wesentliches Bestimmungsstück der die Zukunft betreffenden Wahrscheinlichkeitsurteile ausmacht, muß vor allem eine Klasse ausgezeichnet werden, deren eingehende Untersuchung zweifellos den Ausgangspunkt aller allgemeineren Forschungen auf diesem Gebiete zu bilden hat. Es sind dies die zeitlich homogenen oder, wie ich sie kürzer nennen will, *stationären* Prozesse. Der durch die Variablenschar x_t gekennzeichnete stochastische Prozeß soll als *stationär* bezeichnet werden, wenn die Verteilungsgesetze der beiden endlichen Variablengruppen $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$ und $(x_{t_1+u}, x_{t_2+u}, \ldots, x_{t_n+u})$ einander identisch ausfallen; die Zahl n, die Zeitpunkte $t_1, t_2, \ldots, t_n$ und die Zeitstrecke u können dabei ganz beliebig gewählt werden. Es ist ohne weiteres klar, daß dieser Art von stochastischen Prozessen in verschiedenartigen Anwendungsgebieten eine bedeutende Rolle zukommen muß; insbesondere bildet die oben erwähnte Bewegung des Phasenraums eines mechanischen Systems wegen ihrer Stationarität und Maßinvarianz einen stationären Prozeß [3]), wodurch die Theorie dieser Prozesse eine fundamentale Bedeutung für die statistische Mechanik gewinnt. Aber auch in anderen Gebieten (so z. B. in der Vererbungslehre und Meteorologie) könnten vielleicht die Gesetzmäßigkeiten, die die stationären stochastischen Vorgänge beherrschen, praktische Wichtigkeit beanspruchen.

In dieser Abhandlung soll eine erste Skizze einer Theorie der stationären stochastischen Prozesse versucht werden. Die Methoden und Ergebnisse der vorliegenden Arbeit beschränken sich durchweg auf solche Eigenschaften der im Prozeß begriffenen zufälligen Variablen, die in den Momenten erster und zweiter Ordnung ihrer Verteilung ihren vollständigen Ausdruck finden; es schien mir deshalb angebracht, die ganze Skizze als eine *Korrelationstheorie* der in Frage kommenden Art von stochastischen Vorgängen zu kennzeichnen. Selbstverständlich bedeutet eine derartige Einschränkung einen bewußten Verzicht auf die Erkenntnis mancher tieferliegenden Gesetzmäßigkeit; ich hoffe jedoch, daß trotzdem die gewählte Behandlungsmethode die wichtigsten Grundzüge der Theorie genügend klar hervortreten läßt. Andererseits ist zu bemerken, daß diese Einschränkung des Problemkreises offenbar der Theorie ein wesentlich umfangreicheres Anwendungsgebiet verleiht: die oben definierte Voraussetzung der Stationarität kann nämlich, ohne daß an den Methoden und Ergebnissen etwas zu ändern wäre, durch die folgende weniger bindende ersetzt werden: *x_t soll einen Erwartungswert und eine Streuung haben, die von t unabhängig sind, und der Korrelationskoeffizient der Variablen x_t*

---

[3]) Als Wahrscheinlichkeit einer die Punkte des Phasenraums kennzeichnenden Eigenschaft ist dabei das relative Lebesguesche Maß der Menge derjenigen Punkte aufzufassen, die diese Eigenschaft aufweisen. Näheres darüber siehe in § 5.

und x_u soll eine Funktion von $|t-u|$ allein sein. In allem, was folgt, soll demnach die Stationarität des behandelten Prozesses diese allgemeinere Forderung bedeuten. Der einfacheren Schreibweise halber wollen wir darüber hinaus immer voraussetzen, daß $E(x_t) = 0$, $E(x_t^2) = 1$ ist, wo E hier und im folgenden ein Symbol der mathematischen Erwartung ist. $E(x_u x_v)$ ist dann der Korrelationskoeffizient der Größen x_u und x_v. Nach Voraussetzung dürfen wir

$$E(x_u x_v) = R(u-v)$$

setzen, wo $R(t)$ eine *gerade* Funktion ist, die wir die *Korrelationsfunktion* des betrachteten Prozesses nennen wollen. Ein stationärer Prozeß heiße *stetig*, wenn $R(+0) = 1$ ist; im Fall eines stetigen stationären Prozesses ist $R(t)$ eine stetige Funktion; denn für $\Delta t \to 0$ ist nach der Schwarzschen Ungleichung

$$|R(t+\Delta t) - R(t)| = |E(x_0 x_{t+\Delta t}) - E(x_0 x_t)| = |E[x_0(x_{t+\Delta t} - x_t)]|$$
$$\leqq \sqrt{E(x_0^2) E[(x_{t+\Delta t} - x_t)^2]} = \sqrt{E[(x_{\Delta t} - x_0)^2]} = \sqrt{2}\sqrt{1 - R(\Delta t)} \to 0.$$

Sind x_t und y_t zwei stochastisch unabhängige stationäre Prozesse (dies bedeutet, daß die Variablen x_t und $y_{t'}$ für alle Werte von t und t' stochastisch unabhängige Variable sind), so ist, wie man unmittelbar einsieht, auch der Prozeß $z_t = x_t + y_t$ stationär. Dasselbe gilt auch dann, wenn x_t und y_t *in stationärer Weise* abhängig sind; das soll bedeuten, daß $\frac{1}{2}\{E(x_t y_{t'}) + E(x_{t'} y_t)\}$ eine Funktion von $|t'-t|$ allein ist. Wir wollen diese Funktion mit $\varrho(t'-t)$ bezeichnen und im Fall $E(x_t) = E(y_t) = 0$, $E(x_t^2) = E(y_t^2) = 1$ die *gegenseitige Korrelationsfunktion* der beiden Prozesse x_t und y_t nennen. Sind $R_1(t)$, $R_2(t)$ bzw. die Korrelationsfunktionen der Prozesse x_t und y_t und $R(t)$ die Korrelationsfunktion von $x_t + y_t$, so ist wegen $E\{(x_t + y_t)^2\} = 2[1 + \varrho(0)]$

$$R(t-t') = \frac{E\{(x_t + y_t)(x_{t'} + y_{t'})\}}{2[1+\varrho(0)]} = \frac{R_1(t-t') + R_2(t-t') + 2\varrho(t-t')}{2[1+\varrho(0)]},$$

und folglich

$$\varrho(t) = [1 + \varrho(0)] R(t) - \tfrac{1}{2}[R_1(t) + R_2(t)]. \tag{1}$$

Sind die Prozesse x_t und y_t, also auch die Funktionen $R_1(t)$ und $R_2(t)$ stetig, so ist wegen

$$|\varrho(t') - \varrho(t)| = \tfrac{1}{2}|E\{x_0(y_{t'} - y_t) + y_0(x_{t'} - x_t)\}|$$
$$\leqq \tfrac{1}{2}\sqrt{E(x_0^2) E\{(y_{t'} - y_t)^2\}} + \tfrac{1}{2}\sqrt{E(y_0^2) E\{(x_{t'} - x_t)^2\}}$$
$$= \tfrac{1}{2}\sqrt{2\{1 - R_1(t'-t)\}} + \tfrac{1}{2}\sqrt{2\{1 - R_2(t'-t)\}}$$

auch $\varrho(t)$ eine stetige Funktion.

Im Gebiete der diskreten Reihen von zufälligen Variablen entsprechen den stationären stochastischen Prozessen *stationäre Reihen zufälliger Variablen*; man bezeichnet mit diesem Namen zweckmäßig solche Reihen, deren Glieder sämtlich gleichen Erwartungswert und gleiche Streuung besitzen, und in denen der Korrelationskoeffizient zweier Glieder eine Funktion ihres gegenseitigen Abstandes in der gegebenen Reihe ist. Über solche Reihen sind einige Untersuchungen von Slutsky[4]) und Romanovsky[5]) bekannt, die übrigens hauptsächlich einige Entartungsfälle behandeln. Ich habe[6]) vor kurzem gezeigt, daß solche Reihen ganz allgemein dem Gesetz der großen Zahlen unterliegen. Über den kontinuierlichen Fall ist, soviel ich weiß, bisher nichts veröffentlicht worden.

§ 2.

Allgemeine Form der Korrelationsfunktion im Fall eines stetigen stationären Prozesses.

Die Theorie, die im folgenden entwickelt werden soll, stützt sich in allem Wesentlichen auf eine Spektralzerlegung, deren Möglichkeit eine charakteristische Eigenschaft der Korrelationsfunktion eines stetigen stationären stochastischen Prozesses bildet. Es gilt nämlich folgender

Satz I. *Damit die Funktion $R(t)$ die Korrelationsfunktion eines stetigen stationären stochastischen Prozesses sei, ist notwendig und hinreichend, daß sie sich in der Form*

$$(2) \qquad R(t) = \int_{-\infty}^{+\infty} \cos t x \, d F(x)$$

darstellen läßt, wo $F(x)$ eine gewisse Verteilungsfunktion bedeutet.

Beweis: 1. Die Bedingung ist notwendig. Denn ist $R(t)$ die Korrelationsfunktion eines stetigen stationären stochastischen Prozesses, so ist $R(t)$ stetig und beschränkt. Ferner ist für beliebige reelle a, b und für eine beliebige in (a, b) stetige komplexwertige Funktion $\varphi(t)$

$$0 \leqq E\left\{\left|\int_a^b x_t \varphi(t) \, d t\right|^2\right\} = E\left\{\int_a^b\int_a^b x_u x_v \varphi(u) \overline{\varphi}(v) \, d u \, d v\right\}$$
$$= \int_a^b\int_a^b R(u-v) \varphi(u) \overline{\varphi}(v) \, d u \, d v.$$

4) C. R. Acad. Sci. Paris 185, 169 (1927).

5) Rend. Circ. Mat. Palermo 56 (1932).

6) Giorn. Ist. Ital. Attuari 3 (1932), S. 267; Rec. math. Soc. Math. Moscou 40 (1933), S. 124.

Nach einem Satz von S. Bochner[7]) folgt daraus, daß $R(t)$ in der Gestalt

$$R(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$$

dargestellt werden kann, wo $F(x)$ eine nichtabnehmende Funktion mit beschränkter Schwankung ist; wegen der Realität von $R(t)$ ist

$$R(t) = \int_{-\infty}^{+\infty} \cos tx \, dF(x);$$

endlich ist wegen $R(0) = 1$, $F(+\infty) - F(-\infty) = 1$, also $F(x)$ eine Verteilungsfunktion.

2. Die Bedingung ist hinreichend. Denn hat $R(t)$ die Form (2), so nehme man eine zufällige Variable z, die dem Verteilungsgesetz $F(z)$ unterliegt; ferner sei y eine zweite, von z unabhängige zufällige Variable, welche durch die Wahrscheinlichkeitsdichte

$$\psi(y) = \begin{cases} \frac{1}{2\pi} & \text{für } 0 < y < 2\pi, \\ 0 & \text{für } y < 0 \text{ und } y > 2\pi \end{cases}$$

gekennzeichnet ist. Setzt man

$$x_t = \sqrt{2} \cos(y + zt), \tag{3}$$

so ist, wie man leicht nachrechnet,

$$E(x_t) = 0, \quad E(x_t^2) = 1,$$

$$E(x_u x_v) = \int_{-\infty}^{+\infty} \cos[(u-v)z] \, dF(z) = R(u-v),$$

womit alles bewiesen ist.

Bemerkung: Die Formel (3) definiert einen stochastischen Prozeß, der nur im weiteren Sinne stationär ist. Man kann aber, wie ich einer mündlichen Mitteilung von Herrn Kolmogoroff entnehme, leicht einen Prozeß mit derselben Korrelationsfunktion angeben, der auch in dem ursprünglichen engeren Sinne stationär ist. Man wählt dazu für jedes endliche Variablensystem $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ ein n-dimensionales *normales* Verteilungsgesetz, das ja durch die vorgegebenen Korrelationskoeffizienten eindeutig festgelegt ist; die Gestalt (2) der Korrelationsfunktion garantiert nämlich, daß die im Exponenten auftretende quadratische Form definit ausfällt.

[7]) Vorlesungen über Fouriersche Integrale, Leipzig 1932, S. 76; Satz 23.

Haben wir es mit zwei stetigen stationären Prozessen zu tun, die stationär voneinander abhängen, so zeigt die Formel (1), daß die gegenseitige Korrelationsfunktion $\varrho(t)$ sich in der Form

$$\varrho(t) = \int_{-\infty}^{+\infty} \cos t x \, d\Phi(x) \tag{4}$$

darstellen läßt, wo $\Phi(x)$ eine Funktion mit beschränkter Schwankung im unendlichen Intervall $(-\infty, +\infty)$ bedeutet; es ist dabei offenbar

$$\Phi(+\infty) - \Phi(-\infty) = \varrho(0).$$

§ 3.

Eigenschaften der Korrelationsfunktion.

Aus der in § 2 gewonnenen Spektraldarstellung der Korrelationsfunktion lassen sich viele wichtige Eigenschaften derselben auf elementare Weise ableiten. Alle Anwendungen, die wir in diesem und im folgenden Paragraphen von dieser Fourierzerlegung zu machen haben, beruhen auf folgendem einfachen

Hilfssatz: *Es sei $\psi(x; a, b, \ldots)$ eine reelle stetige Funktion der reellen Variablen x, die außerdem noch von den ebenfalls reellen Parametern $a, b, \ldots$ abhängt und folgende Eigenschaften besitzt:*

1. *Es gibt eine positive Konstante C, so daß für alle $x, a, b, \ldots$*

$$|\psi(x; a, b, \ldots)| < C$$

ist;

2. *Es gibt eine reelle Zahl λ von der Beschaffenheit, daß für jedes feste $\delta > 0$ und für $a, b, \ldots \to \infty$*

$$\psi(x; a, b, \ldots) \to 0$$

gleichmäßig in $|x - \lambda| > \delta$ gilt;

3. *Es gibt eine reelle Zahl A von der Beschaffenheit, daß für beliebige feste positive Werte der Parameter $a, b, \ldots$ und für $x \to \lambda$*

$$\psi(x; a, b, \ldots) \to A$$

gilt.

Ist dann $\Phi(x)$ eine Funktion mit beschränkter Schwankung in $(-\infty, +\infty)$, so gilt für $a, b, \ldots \to \infty$

$$\chi(a, b, \ldots) = \int_{-\infty}^{+\infty} \psi(x; a, b, \ldots) \, d\Phi(x) \to A\,[\Phi(\lambda + 0) - \Phi(\lambda - 0)].$$

Beweis: Man setze $K = \int\limits_{-\infty}^{+\infty} |d\Phi(x)|$; $\varepsilon > 0$ sei beliebig vorgegeben; man wähle $\delta > 0$ so klein, daß

(5) $$\int\limits_{\lambda-\delta}^{\lambda-0} |dF(x)| < \frac{\varepsilon}{4C}, \quad \int\limits_{\lambda+0}^{\lambda+\delta} |dF(x)| < \frac{\varepsilon}{4C}$$

wird; für genügend große Werte von $a, b, \ldots$ ist dann wegen 2.

$$|\psi(x;\, a, b, \ldots)| < \frac{\varepsilon}{2K} \quad \text{für} \quad |x-\lambda| \geqq \delta,$$

und folglich

(6) $$\Bigl|\int\limits_{|x-\lambda|>\delta} \psi\, d\Phi(x)\Bigr| < \frac{\varepsilon}{2};$$

für jede positive Zahl $\eta < \delta$ hat man aber wegen (5)

(7) $$\Bigl|\int\limits_{\eta<|x-\lambda|<\delta} \psi\, d\Phi(x)\Bigr| < C \int\limits_{\eta<|x-\lambda|<\delta} |d\Phi(x)| < \frac{\varepsilon}{2};$$

aus (6) und (7) folgt

$$\Bigl|\chi(a, b, \ldots) - \int\limits_{\lambda-\eta}^{\lambda+\eta} \psi(x;\, a, b, \ldots)\, d\Phi(x)\Bigr| < \varepsilon;$$

und da hierin η beliebig klein ist, folgt nach der Eigenschaft 3. von ψ

$$|\chi(a, b, \ldots) - A\,\{\Phi(\lambda+0) - \Phi(\lambda-0)\}| < \varepsilon,$$

wenn nur $a, b, \ldots$ hinreichend groß sind; damit ist der Hilfssatz bewiesen.

Wir gehen nun dazu über, einige Eigenschaften der Korrelationsfunktion festzustellen; wir wollen uns dabei mit der gegenseitigen Korrelationsfunktion $\varrho(t)$ zweier stationärer Prozesse befassen, denn es genügt ja, diese Prozesse miteinander zu identifizieren, um den Spezialfall eines einzelnen Prozesses zu erhalten.

Satz 2. *Die Korrelationsfunktion $\varrho(t)$ besitzt einen bestimmten Mittelwert, d. h. es ist*

$$\lim_{T\to\infty} \frac{1}{T} \int\limits_0^T \varrho(t)\, dt$$

vorhanden[8]).

---

[8]) Den analogen Satz für diskrete Reihen zufälliger Variablen habe ich mittels einer ganz anderen Methode früher bewiesen [Giorn. Ist. Ital. Attuari 3 (1932), 267].

Beweis: Nach (3) ist

$$\frac{1}{T}\int_0^T \varrho(t)\,dt = \int_{-\infty}^{+\infty} \frac{\sin T x}{T x}\,d\Phi(x);$$

hierin genügt $\psi(x;\, T) = \frac{\sin T x}{T x}$ offenbar allen Voraussetzungen des soeben bewiesenen Hilfssatzes, wenn man $\lambda = 0$, $A = 1$ wählt; folglich ist

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varrho(t)\,dt = \Phi(+0) - \Phi(-0),$$

w. z. b. w.

Satz 3. *Die Korrelationsfunktion $\varrho(t)$ läßt sich in der Gestalt einer Summe*

$$\varrho(t) = \varrho_1(t) + \varrho_2(t)$$

darstellen, wo $\varrho_1(t)$ fastperiodisch ist und $|\varrho_2(t)|^2$ den Mittelwert Null hat.

Beweis: In (4) zerlege man $\Phi(x)$ in zwei Summanden

$$\Phi(x) = \Phi_1(x) + \Phi_2(x),$$

von denen $\Phi_1(x)$ ein Punktspektrum hat, während $\Phi_2(x)$ stetig ist, und setze

$$\varrho_1(t) = \int_{-\infty}^{+\infty} \cos t x\, d\Phi_1(x), \quad \varrho_2(t) = \int_{-\infty}^{+\infty} \cos t x\, d\Phi_2(x),$$

so daß $\varrho(t) = \varrho_1(t) + \varrho_2(t)$ wird. $\varrho_1(t)$ ist die Summe einer gleichmäßig konvergenten trigonometrischen Reihe und folglich fastperiodisch. Andererseits ist aber

$$\frac{1}{T}\int_0^T \{\varrho_2(t)\}^2\,dt = \frac{1}{T}\int_0^T \left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \cos t x \cos t y\, d\Phi_2(x)\, d\Phi_2(y)\right\}$$

$$= \frac{1}{2T}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\int_0^T [\cos t(x+y) + \cos t(x-y)]\,dt\right\} d\Phi_2(x)\,d\Phi_2(y)$$

$$= \frac{1}{2}\int_{-\infty}^{+\infty} d\Phi_2(y) \int_{-\infty}^{+\infty}\left\{\frac{\sin T(x+y)}{T(x+y)} + \frac{\sin T(x-y)}{T(x-y)}\right\} d\Phi_2(x).$$

Im Integral $\int_{-\infty}^{+\infty} \frac{\sin T(x-y)}{T(x-y)}\,d\Phi_2(x)$ ist aber $\psi(x;\, T) = \frac{\sin T(x-y)}{T(x-y)}$ wieder eine Funktion von der im oben bewiesenen Hilfssatz betrachteten Art, wenn man $\lambda = y$, $A = 1$ wählt; folglich konvergiert dieses Integral für $T \to \infty$ gegen $\Phi_2(y+0) - \Phi_2(y-0) = 0$, und zwar gleichmäßig in bezug auf y, wie man leicht daraus erschließt, daß die Funktion $\Phi_2(x)$ in $(-\infty, +\infty)$ von beschränkter Schwankung und folglich gleichmäßig stetig ist. Und da offenbar dasselbe auch für das andere Integral

$$\int_{-\infty}^{+\infty} \frac{\sin T(x+y)}{T(x+y)}\, d\Phi_2(x)$$

stattfindet, so konvergiert auch der ganze Ausdruck für $T \to \infty$ gegen Null, w. z. b. w.

§ 4.

Gesetz der großen Zahlen.

Es ist eine der wichtigsten Eigenschaften der stationären Reihen und Prozesse, daß sie eine Stabilität der arithmetischen Mittelbildungen aufweisen oder, wie man in der Wahrscheinlichkeitstheorie zu sagen pflegt, *dem Gesetz der großen Zahlen unterliegen.* Für einen stationären stochastischen Prozeß hat das folgende Bedeutung: man setze

$$\xi(T) = \frac{1}{T}\int_0^T x_t\, dt;$$

sind dann ε und δ beliebig kleine positive Zahlen, so ist für hinreichend großes T und für alle $U > 0$ die Wahrscheinlichkeit der Ungleichung

$$|\xi(T+U) - \xi(T)| > \varepsilon$$

kleiner als δ.

Um dies einzusehen, genügt es nach der Tschebycheffschen Ungleichung zu beweisen, daß für $T \to \infty$ gleichmäßig in bezug auf $U > 0$

$$\mu(T, U) = E\{[\xi(T+U) - \xi(T)]^2\} \to 0$$

gilt.

Beweis: Es ist

$$\begin{aligned}
[\xi(T+U) - \xi(T)]^2 &= \frac{1}{T^2(T+U)^2}\Big[T\int_0^{T+U} x_t\, dt - (T+U)\int_0^T x_t\, dt\Big]^2 \\
&= \frac{1}{T^2(T+U)^2}\Big[T\int_T^{T+U} x_t\, dt - U\int_0^T x_t\, dt\Big]^2 \\
&= \frac{1}{T^2(T+U)^2}\Big[T^2\int_T^{T+U}\int_T^{T+U} x_u x_v\, du\, dv + U^2\int_0^T\int_0^T x_u x_v\, du\, dv \\
&\qquad - 2UT\int_0^T\int_T^{T+U} x_u x_v\, du\, dv\Big],
\end{aligned}$$

und folglich wegen $E(x_u x_v) = R(u-v)$

$$\mu(T, U) = \frac{1}{T^2 (T+U)^2}\left[T^2 \int_0^U \int_0^U R(u-v)\,du\,dv + U^2 \int_0^T \int_0^T R(u-v)\,du\,dv \right.$$

$$\left. - 2UT \int_0^T \int_T^{T+U} R(u-v)\,du\,dv\right].$$

Setzt man hierin nach (2)

$$R(u-v) = \int_{-\infty}^{+\infty} \cos[(u-v)x]\,dF(x),$$

so ergibt eine elementare Rechnung

$$\mu(T, U) = \frac{U^2}{(T+U)^2} \int_{-\infty}^{+\infty} \left\{ \left(\frac{\sin \frac{1}{2} T x}{\frac{1}{2} T x}\right)^2 + \left(\frac{\sin \frac{1}{2} U x}{\frac{1}{2} U x}\right)^2 \right.$$

$$\left. - 2 \frac{\sin \frac{1}{2} T x}{\frac{1}{2} T x} \frac{\sin \frac{1}{2} U x}{\frac{1}{2} U x} \cos \frac{(T+U)x}{2} \right\} dF(x).$$

Wenn nun für $T \to \infty$ U beschränkt bleibt, wird der vor dem Integral stehende Faktor unendlich klein, und folglich konvergiert $\mu(T, U)$ gegen Null. Wird aber U unendlich groß, so genügt offenbar der Integrand $\psi(x; U, T)$ allen Voraussetzungen des im § 3 bewiesenen Hilfssatzes, wenn man daselbst $\lambda = 0$, $A = 0$ setzt. Es ist folglich auch in diesem Fall $\mu(T, U) \to 0$, und damit ist die Behauptung vollständig bewiesen.

Für stationäre *Reihen* habe ich den Beweis des analogen Satzes mittels ganz anderer Methoden schon früher gegeben[9]).

§ 5.

Mechanische Anwendungen.

Die in den vorstehenden Paragraphen geschilderte Methode steht in naher Beziehung zu einigen Untersuchungen über allgemeine dynamische Probleme, die in der letzten Zeit von Koopman, v. Neumann und E. Hopf veröffentlicht wurden[10]). Sie unterscheidet sich von der Koopman-v. Neumannschen Methode hauptsächlich dadurch, daß Operatorenbetrachtungen bei der hier gegebenen Anordnung vermieden werden können. Die entsprechenden mechanischen Sätze erscheinen dabei nicht

---

[9]) Vgl. [6]).

[10]) Vgl. insbesondere J. v. Neumann, Proof of the quasiergodic hypothesis, Proc. Nat. Acad. Sci. U. S. A. 18 (1932), 70; B. O. Koopman and J. v. Neumann, Dynamical systems of continuous spectra, ibid. 18 (1932), 255.

als Folgerungen, sondern bilden lediglich eine Umdeutung der hier entwickelten Theorie; ich habe diese Sätze vor einigen Monaten in einer Note[11]) zusammengefaßt.

Ist V der endlich vorausgesetzte Phasenraum eines Hamiltonschen Systems, E eine im Lebesgueschen Sinne meßbare Punktmenge in V von positivem Maß $m(E)$, F eine andere ebensolche Menge, und bedeuten E_t bzw. F_t die Mengen, in welche E bzw. F nach Abschluß einer Zeitstrecke t übergehen, so betrachte man die Funktion

$$\mu(t) = \tfrac{1}{2}[m(E_t F) + m(E F_t)].$$

Wie oben, kann leicht gezeigt werden, daß diese Funktion eine Spektralzerlegung der Form (4) zuläßt; denn ist P_t der Punkt von V, in welchen der Punkt P nach Ablauf der Zeitstrecke t übergeht, und bedeutet $\varphi(P)$ 1 oder 0, je nachdem $P \subset E$ oder $P \not\subset E$ ist, und $\psi(P)$ die entsprechende Funktion für die Menge F, so ist offenbar

$$\mu(t) = \tfrac{1}{2}\int [\varphi(P_t)\,\psi(P) + \varphi(P)\,\psi(P_t)]\,dV,$$

wo das Integral über den ganzen Phasenraum V zu erstrecken ist; diese Integration ist hier das Analogon der Bildung von mathematischen Erwartungen, und erlaubt infolgedessen alle Schlüsse, die wir in § 2 gezogen haben.

Aus der so gewonnenen Spektraldarstellung ergeben sich nun unmittelbar die Analoga der Sätze 2 und 3 von § 3: die Existenz von $\lim\limits_{T\to\infty} \frac{1}{T}\int\limits_0^T \mu(t)\,dt$ und die Zerlegung $\mu(t) = \mu_1(t) + \mu_2(t)$, wo $\mu_1(t)$ fastperiodisch und $\lim\limits_{T\to\infty} \frac{1}{T}\int\limits_0^T \{\mu_2(t)\}^2\,dt = 0$ ist. Das Gesetz der großen Zahlen in seiner in § 4 dargestellten Form läßt sich als der v. Neumannsche „Quasiergodensatz" deuten.

Moskau, Mathematisches Institut der Universität.

---

11) Proc. Nat. Acad. Sci. USA, 19 (1933), 567.

(Eingegangen am 2. 5. 1933.)