NATIONAL BUREAU OF STANDARDS REPORT

1293

INTRODUCTION TO THE THEORY OF STOCHASTIC PROCESSES
DEPENDING ON A CONTINUOUS PARAMETER

By

Henry B. Mann
Ohio State University

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, Secretary

NATIONAL BUREAU OF STANDARDS
A. V. Astin, Acting Director

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

1. **ELECTRICITY.** Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Electrochemistry.

2. **OPTICS AND METROLOGY.** Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

3. **HEAT AND POWER.** Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels.

4. **ATOMIC AND RADIATION PHYSICS.** Spectroscopy. Radiometry. Mass Spectrometry. Physical Electronics. Electron Physics. Atomic Physics. Neutron Measurements. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

5. **CHEMISTRY.** Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

6. **MECHANICS.** Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

7. **ORGANIC AND FIBROUS MATERIALS.** Rubber. Textiles. Paper. Leather. Testing and Specifications. Organic Plastics. Dental Research.

8. **METALLURGY.** Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

9. **MINERAL PRODUCTS.** Porcelain and Pottery. Glass. Refractories. Enameled Metals. Building Stone. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

10. **BUILDING TECHNOLOGY.** Structural Engineering. Fire Protection. Heating and Air Conditioning. Exterior and Interior Coverings. Codes and Specifications.

11. **APPLIED MATHEMATICS.** Numerical Analysis. Computation. Statistical Engineering. Machine Development.

12. **ELECTRONICS.** Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

13. **ORDNANCE DEVELOPMENT.** Mechanical Research and Development. Electromechanical Fuzes. Technical Services. Missile Fuzing Research. Missile Fuzing Development. Projectile Fuzes. Ordnance Components. Ordnance Tests. Ordnance Research.

14. **RADIO PROPAGATION.** Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

15. **MISSILE DEVELOPMENT.** Missile Engineering. Missile Dynamics. Missile Intelligence. Missile Instrumentation. Technical Services. Combustion.

NBS PROJECT **NBS REPORT**

May 1951

1103-11-1107 1293

INTRODUCTION TO THE THEORY OF STOCHASTIC PROCESSES DEPENDING ON A CONTINUOUS PARAMETER

By

Henry B. Mann

Ohio State University

NBS

This monograph to be published in the NBS Applied Mathematics Series is the outgrowth of a series of lectures given at the National Bureau of Standards in June, 1949, under the sponsorship of the Statistical Engineering Laboratory.

Contemplator enim, cum solis lumina cumque
Inserti fundunt radii per opaca domorum,
Multi minuta modis multis per inane videbis
Corpora misceri, radiorum lumine in ipso.
Et velut eterno certamine proelia, pugnas
Edere turmatim certantia nec dare pausam,
Conciliis et discidiis exercita crebris.
Conicere ut possis ex hoc, primordia rerum
Quale sit in magno iactari semper inani.
Dumtaxat rerum magnarum parva potest res
Exemplare dare et vestigia notitiai.
Hoc etiam magis haec animum te advertere par est
Corpora, quae in solis radiis turbare videntur
Quod tales turbae motus quoque materiai
Significant clandestinos caecosque subesse.
Multa videbis enim plagis ibi percita caecis
Commutare viam retroque pulsa reverti
Nunc huc nunc illuc in cunctas undique partes,

Titus Lucretius Carus
De Rerum Natura: Vol. II. Vers 113-130.

Let us observe as brightly the rays of the sun
Penetrate in streams the darkness of our houses
Thousands of tiny bodies dancing in space
Approaching each other and parting in the bright light of the sun.
As if fighting a battle without pause through the ages,
Like an army of soldiers restlessly warring,
They advance and retreat in motion never to cease.
May you conjecture from this the very nature of matter,
How it is ceaselessly tossed through the vastness of space.
Thus a phenomenon small as it seems and of little importance
Often does indicate things highly important and great.
Hence it is well worthwhile to observe these bodies
Whirling and dancing without rest in the sunlight,
Since such irregular motion of visible bodies
Is a sure indication of the invisible motion of matter.
For you can see these bodies constantly changing direction,
Often reversing their motion all of a sudden
And propelled by invisible impacts moving this way and that way.

TABLE OF CONTENTS

FOREWORD

The theory of stochastic processes is steadily gaining in importance and the applications are ever widening. Nevertheless, it is at present not easy to study this subject, since the literature, although extensive, is widely scattered.

This situation motivated the National Bureau of Standards to invite Professor Henry B. Mann to give a series of lectures on stochastic processes and to write a monograph on the subject. The lectures were given in the period from March 1949 to June 1949, during which Dr. Mann was a member of the staff of the Bureau's Statistical Engineering Laboratory.

It is well known that the theory of stochastic processes depending on a continuous parameter can be developed in a satisfactory way by studying random functions or by considering probability measures in function space. The author of the present monograph has however adopted a different approach which is similar to the definition of a stochastic process given by E. Slutsky. A random variable is considered to be a symbol with which a distribution function is associated, and a stochastic process is then defined as a set of random variables. This approach leads to a theory which for many practical purposes is equivalent to the direct measure-theoretical approach. It has the advantage that the technicalities of measure theory seem less obstrusive at the outset, although for logical completness they must enter sooner or later if the theory is to be developed in a well-rounded way.

It is hoped that this modest little volume, written by a distinguished contemporary mathematician, will be useful and interesting in various ways. The argument is addressed uncompromisingly to educated mathematicians, and they will not fail to be impressed by the skillful way in which the author develops the theory from his chosen starting point. The user of time-continuous processes in the applied fields who is not interested in the methods of proof may still appreciate having a number of important definitions and results conveniently gathered here between two covers.

Finally, it is hoped that the publication of the monograph will stimulate further expository efforts in the important field of time-continuous stochastic processes, and that in particular the day will come a little sooner than it otherwise might have, when a comprehensive but readable textbook on this subject, using the measure-theoretical approach, appears in the English language.

J. H. Curtiss

National Applied Mathematics Laboratories
National Bureau of Standards
Washington 25, D. C.
October 1951

INTRODUCTION

The study of stochastic processes is becoming increasingly important in many branches of science and accordingly the mathematical theory of stochastic processes has progressed rapidly during the last two decades. This rapid progress has resulted in a large diversification of notation and terminology which makes it difficult even for a mathematician to inform himself on the subject. It seemed, therefore, advisable to bring together under a unified terminology and notation some of the basic definitions and results of this theory. The viewpoint taken was that of the mathematical statistician, and the stochastic process was accordingly defined as a family of distribution functions satisfying certain consistency relations. It was one of the goals of the present monograph to develop the theory of stochastic processes from this viewpoint with as little appeal to abstract measure theory as possible. In most practical problems information about random variables can be obtained only in terms of their joint distribution function, and it is the opinion of the author that a treatise on stochastic processes will be most useful to the statistician if the definitions, theorems, and proofs are given in these terms. It is in many cases almost impossible to trace a result to one particular author, and it was therefore decided to omit references altogether. This does not mean that the author claims credit for any particular result. To the author's knowledge only theorem 7 of chapter 1 and most of chapter 3 are new. (After completion of chapter 3 the author was informed by H. Rubin that some of the

results of this chapter had previously been obtained by him and L. Savage, but their results were never published.) In his presentation of the theory of stochastic processes, as well as in chapter 4, the author has followed the presentations of M. Loeve given in Paul Levy's book on stochastic processes and in M. Loeve's paper "On sets of probability laws and their limit elements" (University of California Press, 1950), respectively. In the treatment of counter data in chapter 5 the author has used W. Feller's approach and his masterful presentation in the Courant Anniversary Volume. The treatment of the Ornstein Uhlenbeck process in chapter 2 follows a presentation given by J. L. Doob (Ann. of Math., Vol. 43, No. 2).

My thanks are due to Dr. Eugene Lukacs for his valuable help in preparing the final form of the manuscript and to Mr. P. Moranda who read the proofs and prepared the index. I also wish to thank Professor M. Loeve for many helpful discussions on the subject.

H. B. MANN

Ohio State University
May 1951

Chapter 1

FUNDAMENTAL CONCEPTS

1. Random variables. We consider a finite or infinite set of symbols $(x, y, \ldots)$ such that to every finite set of symbols $x_1, \ldots, x_n$ there is defined a right continuous distribution function

(1.1) $$F_{x_1, x_2, \ldots, x_n}(a_1, a_2, \ldots, a_n) = P(x_1 \leq a_1, \ldots, x_n \leq a_n)$$

called the probability of the event $x_1 \leq a_1, \ldots, x_n \leq a_n$.

The distribution functions of the family given by (1.1) satisfy the following equations:

(1.2) $$F_{x_{i_1}, \ldots, x_{i_n}}(a_{i_1}, \ldots, a_{i_n}) = F_{x_1, \ldots, x_n}(a_1, \ldots, a_n)$$

where $i_1, \ldots, i_n$ is any permutation of the numbers $1, 2, \ldots, n$;

(1.3) $$F_{x_1, \ldots, x_n}(a_1, \ldots, a_{n-1}, \infty) = F_{x_1, \ldots, x_{n-1}}(a_1, \ldots, a_{n-1})$$

The symbol x_i is called a random variable. For every Borel set A in the n-dimensional Euclidean space we define the symbol $P[(x_1, \ldots, x_n) \subset A]$, called the probability that the "point" $(x_1, \ldots, x_n)$ lies in A by the equation

$$P[(x_1, \ldots, x_n) \subset A] = \int_A dF_{x_1, \ldots, x_n}(a_1, \ldots, a_n)$$

If $g(a_1,\ldots,a_m)$ is a Borel measurable function, then we can define a new random variable $g(x_1,\ldots,x_m)$ by the equations

$$F_{g,y_1,\ldots,y_m}(a,b_1,\ldots,b_m) = P(g \leq a,\ y_1 \leq b_1,\ldots,\ y_m \leq b_m)$$

$$= F_{y_1\ldots g\ldots y_m}(b_1,\ldots,a,\ldots,b_m)$$

We now consider sequences $\{x_j\}$ of random variables. The notion of convergence of such a sequence can be defined in various ways. In our representation of the theory of stochastic processes we shall however use mainly the following definition.

2. Convergence. A sequence $\{x_j\}$ will be called convergent if for every $\varepsilon > 0$, $\eta > 0$ there exists an $N(\varepsilon,\eta)$ such that

(1.4) $$P(|x_{n+h} - x_n| \geq \varepsilon) \leq \eta$$

for $n > N(\varepsilon,\eta)$ and all h. If there exists a random variable x such that $\lim\limits_{n\to\infty} P(|x_n - x| \geq \varepsilon) = 0$ for all ε then we shall write $\operatorname{plim}\limits_{n\to\infty} x_n = x$ and say that $\{x_n\}$ converges to x or that x is the probability limit of the sequence $\{x_n\}$. The convergence defined above is usually termed convergence in probability. This definition can be extended in an obvious manner to random vectors.

We proceed to formulate an important property of convergent sequences.

<u>Theorem 1.1</u> Let $\{x_n\}$ be a sequence of random variables. There exists a random variable x such that $\underset{n \to \infty}{\text{plim}}\, x_n = x$ if and only if the sequence $\{x_n\}$ converges. Moreover, if $F_{x_n y_1 \ldots y_m}$ are the distribution functions of x_n $(n = 1, 2, \ldots)$ and $y_1 \ldots y_m$ then $\lim_{n \to \infty} F_{x_n y_1 \ldots y_m} = F_{x y_1 \ldots y_m}$ for all points $(t, b_1, \ldots, b_m)$ for which the function $F_{x y_1 \ldots y_m}(t, b_1, \ldots, b_m)$ is continuous in t.

Theorem 1.1 gives a condition for convergence in probability similar to Cauchy's criterion. This condition was first established by E. Slutsky[1]. We proceed to prove[2] theorem 1.1. As a first step we assume that the sequence $\{x_n\}$ is convergent and show the existence of a random variable $x = \underset{n \to \infty}{\text{plim}}\, x_n$. In the following we write for abbreviation

$$g_n(c) = F_{x_n y_1 \ldots y_m}(c, b_1, \ldots, b_m)$$

We first prove the following lemma.

---

[1] Metron 5, 3-89 (1925); C. R. Acad. Sci. Paris, <u>187</u>, 370-372(1928).

[2] The proof of theorem 1.1 may be skipped in a first reading without affecting the understanding of the rest of the monograph.

<u>Lemma 1.1</u> If δ and η are any positive numbers, then for sufficiently large n and all $h \geq 0$

$$(1.5) \qquad g_n(c+\delta) + \eta \geq g_{n+h}(c) \geq g_n(c-\delta) - \eta$$

for all c.

For abbreviation we write for any event $\mathcal{E}$

$$P_b(\mathcal{E}) = P(\mathcal{E},\ y_1 \leq b_1, \ldots, y_m \leq b_m),$$

so that in particular

$$P_b(x_n \leq c) = g_n(c).$$

Then

$$P_b(x_n \leq c+\delta) \geq P_b(x_n \leq c+\delta,\ x_{n+h} \leq c) \geq P_b(x_{n+h} \leq c,\ |x_{n+h} - x_n| \leq \delta).$$

Since the set of points (x_{n+h}, x_n) for which $|x_{n+h} - x_n| > \delta$ includes the points for which $|x_{n+h} - x_n| > \delta$ and $x_{n+h} \leq c$, we have

$$P_b(x_{n+h} \leq c,\ |x_{n+h} - x_n| \leq \delta) \geq P_b(x_{n+h} \leq c) - P(|x_{n+h} - x_n| > \delta)$$

Hence for sufficiently large n and all $h \geq 0$

$$g_n(c+\delta) \geq g_{n+h}(c) - \eta .$$

Similarly

$$g_{n+h}(c) \geq g_n(c-\delta) - \eta$$

and (1.5) follows.

For a sequence of functions $\{f_n(t)\}$ we shall write $\operatorname*{Lim}_{n\to\infty} f_n(t) = f(t)$ if $\lim_{n\to\infty} f_n(t) = f(t)$ for every continuity point of $f(t)$.

Lemma 1.2 There exists a non-decreasing function $g(c)$ such that

$$\lim_{n\to\infty} g_n(c) = g(c).$$

The functions $g_n(c)$ are non decreasing and bounded hence by Helly's theorem[8] there exists a subsequence $g_{n_i}(c)$ such that

$$(1.6) \qquad \lim_{i\to\infty} g_{n_i}(c) = g(c)$$

where $g(c)$ is non-decreasing.

Let t be a continuity point of $g(c)$. Fix $\eta>0$ and choose δ positive and arbitrarily small and so that $t+\delta$ and $t-\delta$ are continuity points of $g(c)$ and

$$g(t+\delta) - g(t) \leq \eta, \quad g(t) - g(t-\delta) \leq \eta .$$

For sufficiently large n_i and n we then have by (1.5)

$$g_{n_i}(t+\delta) + \eta \geq g_n(t) \geq g_{n_i}(t-\delta) - \eta ,$$

and therefore

$$g(t+\delta) + \eta \geq g_n(t) \geq g(t-\delta) - \eta .$$

By choice of δ we have

$$(1.7) \qquad g(t) + 2\eta \geq g_n(t) \geq g(t) - 2\eta$$

where η can be made arbitrarily small for sufficiently large n.

---

[8] Helly's theorem (see, for instance, D. V. Widder, The Laplace Transform 27) states. If the real non-decreasing functions $\alpha_n(x)$ and the positive constant A are such that $|\alpha_n(x)| < A$ $(n = 0, 1, 2, \ldots)$, $a \leq x \leq b$; then there exists a subsequence $\{\alpha_{n_i}(x)\}$ of $\{\alpha_n(x)\}$ and a non-decreasing bounded function $\alpha(x)$ such that

$$\lim_{i\to\infty} \alpha_{n_i}(x) = \alpha(x) \qquad (a \leq x \leq b)$$

It then follows that

$$\lim_{n \to \infty} g_n(t) = g(t)$$

and lemma 1.2 is proved.

We now define a symbol x by the equations

$$F_{xy_1 \ldots y_m}(t, b_1, \ldots, b_m) = F_{y_1 \ldots y_{i-1} x y_i \ldots y_m}(b_1, \ldots, b_{i-1}, t, b_i, \ldots, b_m)$$

$$= g(t + 0)$$

To prove that x is a random variable we have to show that $F_{xy_1 \ldots y_m}$ is a distribution function and that

$$(1.8) \qquad F_{xy_1 \ldots y_m}(t, b_1, \ldots, b_{m-1}, \infty) = F_{xy_1 \ldots y_{m-1}}(t, b_1, \ldots, b_{m-1})$$

and

$$(1.9) \qquad \lim_{t \to \infty} F_{xy_1 \ldots y_m}(t, b_1, \ldots, b_m) = F_{y_1 \ldots y_m}(b_1, b_2, \ldots, b_m) .$$

That the interval function corresponding to $F_{xy_1 \ldots y_m}$ is non-negative is obvious since it is a limit of functions $F_{x_n y_1 \ldots y_m}$ with this property. We therefore have merely to show that (1.8) and (1.9) hold and that $F_{xy_1 \ldots y_m}$ tends to zero if any one of its arguments tends to $-\infty$.

We have

$$0 \leq F_{x_n y_1 \ldots y_m}(t, b_1, \ldots, b_m) \leq F_{y_j}(b_j)$$

and hence

$$0 \leq F_{x y_1 \ldots y_m}(t, b_1, \ldots, b_m) \leq F_{y_j}(b_j)$$

so that

$$\lim_{b_j \to -\infty} F_{x y_1 \ldots y_m}(t, b_1, \ldots, b_m) = 0$$

Furthermore from (1.5) for arbitrarily small η and all t

$$F_{x_n y_1 \ldots y_m}(t+\delta, b_1, \ldots, b_m) + \eta \geq F_{x y_1 \ldots y_m}(t, b_1, \ldots, b_m) \geq 0$$

for sufficiently large n uniformly in t. We therefore have

$$\lim_{t \to -\infty} F_{x y_1 \ldots y_m}(t, b_1, \ldots, b_m) = 0 .$$

Furthermore

$$0 \leq F_{x_n y_1 \ldots y_{m-1}}(t, b_1, \ldots, b_{m-1}) - F_{x_n y_1 \ldots y_m}(t, b_1, \ldots, b_m)$$

$$\leq 1 - F_{y_m}(b_m) ;$$

if we let first $n \to \infty$ and then $b_m \to \infty$ we obtain (1.8).

To prove (1.9) we choose a continuity point c of $F_{x y_1 \ldots y_m}(t, b_1 \ldots b_m)$ so large that for fixed δ and η

$$F_{x_n y_1 \ldots y_m}(c-\delta, b_1, \ldots, b_m) = F_{y_1 \ldots y_m}(b_1, \ldots, b_m) - \eta\theta$$

and also

$$F_{x_n y_1 \ldots y_m}(c+\delta, b_1, \ldots, b_m) = F_{y_1 \ldots y_m}(b_1, \ldots, b_m) - \eta\theta'$$

where $0 < \theta < 1$ and $0 < \theta' < 1$.

[illegible] this term (1.5) we get:

$$F_{y_1 \ldots y_m}(b_1, \ldots, b_m) + \eta \geq F_{x_{n+h} y_1 \ldots y_m}(a, b_1, \ldots, b_m)$$

$$\geq F_{y_1 \ldots y_m}(b_1, \ldots, b_m) - 2\eta .$$

By letting $h \to \infty$ and considering that η may be chosen arbitrarily small we obtain (1.9).

We proceed to prove that $\operatorname{plim}_{n \to \infty} x_n = x$. We represent the domain $|y - x| > \varepsilon$ by the sum of a denumerable number of intervals whose corner points are continuity points of $F_{x_n x}$ for all n. To do this we remember that there can be at most a denumerable number of points with a positive probability in the plane, so that we can construct an interval netting[4] which avoids these points. Since the set $|y - x| > \varepsilon$ is open, every point of this set is contained in one interval of this netting which is entirely contained in the set. Therefore we have merely to take all the intervals of our netting which lie entirely in the set $|y - x| > \varepsilon$. Thereby we proceed in such a way that if I_1, I_2 are two such intervals and

---

[4] A set of lines parallel to the x and y axis which divides the plane into rectangles is called a mesh. The maximum area of any of these rectangles is called the modulus of the mesh. A sequence (M_i) of meshes such that M_{i+1} is a refinement of M_i and such that the moduli m_i converge to zero is called a netting.

$I_1 \supset I_2$ then only I_1 is chosen for our interval covering.

Let $I_1, I_2, \ldots$ be these intervals and denote by $P_{x_n y}(I_k)$ the probability that the point (x_n, y) will fall into the interior of the interval I_k or on either its right ~~and~~ or upper boundary. Then for sufficiently large n and arbitrary η

$$(1.11) \quad P(|x_{n+h} - x_n| > \varepsilon) = \sum_k P_{x_n x_{n+h}}(I_k) \leq \eta$$

for all h.

Furthermore

$$P(|x_n - x| > \varepsilon) = \sum_k P_{x_n x}(I_k)$$

Both sums converge. From now on consider n as fixed. Choose N so that for some $\eta > 0$

$$\sum_{k=N+1}^{\infty} P_{x_n x}(I_k) < \eta .$$

Next choose h so that for the first N intervals

$$|P_{x_n x_{n+h}}(I_k) - P_{x_n x}(I_k)| \leq \frac{\eta}{N} .$$

Then

$$P(|x_n - x| > \varepsilon) = \sum_1^{\infty} P_{x_n x}(I_k) \leq \sum_1^{\infty} P_{x_n x_{n+h}}(I_k) + 2\eta \leq 3\eta .$$

Since η was arbitrary $\lim_{n\to\infty} P(|x_n - x| > \varepsilon) = 0$ or $\operatorname{plim}_{n\to\infty} x_n = x$.

[The relation $\operatorname{plim}_{n\to\infty} x_n = x$ also follows from the fact that the characteristic function of $x_n - x_{n+h}$ converges to the characteristic function of $x_n - x$].

On the other hand if $\operatorname{plim}_{n\to\infty} x_n = x$ then for sufficiently large n and arbitrary η

$$P(|x_{n+h}-x_n| \leq \varepsilon) \geq P(|x_n-x| \leq \frac{\varepsilon}{2} \text{ and } |x_{n+h}-x| \leq \frac{\varepsilon}{2})$$

$$\geq P(|x_n-x| \leq \frac{\varepsilon}{2}) - P(|x_{n+h}-x| > \frac{\varepsilon}{2}) \geq 1-\eta \, .$$

Hence the sequence $\{x_n\}$ converges and theorem 1.1 is proved.

3. Stochastic processes. A set of random variables x_t where t is chosen out of some set of real numbers is called a stochastic process. If the set of indices t is an interval then the stochastic process is said to depend on a continuous parameter. Such a process is called continuous in $[a,b]$ if for every sequence $\{h_i\}$ with $\lim_{i\to\infty} h_i = 0$, $\operatorname{plim}_{i\to\infty} x_{t+h_i} = x_t$ for $a \leq t \leq b$.

The expression

$$\int_{-\infty}^{+\infty} t \, dF_y(t) = E(y)$$

is called the mathematical expectation of y. The expression

$$E\{[x - E(x)][y - E(y)]\} = \sigma_{xy}$$

is called the covariance between x and y. The covariance between x_{t_1} and x_{t_2} will be denoted by $\sigma_{t_1 t_2}$ and called the covariance function of the process.

4. Convergence in the mean. A sequence of random variables $\{x_n\}$ is said to converge in the mean to a random variable x in symbols $\underset{n\to\infty}{\mathrm{l.i.m.}}\, x_n = x$ if

$$\lim_{n\to\infty} E(x_n - x)^2 = 0 . \tag{1.12}$$

We shall now prove several very useful lemmas on convergence in the mean and convergence in probability.

Lemma 1.3. If $\underset{n\to\infty}{\mathrm{l.i.m.}}\, x_n = x$ then $\underset{n\to\infty}{\mathrm{plim}}\, x_n = x$.

This follows immediately from Tchebicheff's inequality.

Lemma 1.4. If $\underset{n\to\infty}{\mathrm{l.i.m.}}\, x_n = x$ then $\lim\limits_{n\to\infty} E(x_n) = E(x)$.

Since $E(x_n - x)^2 \leq \varepsilon$ for sufficiently large n we also have

$$\varepsilon \geq E(x_n - x)^2 = \sigma^2_{x_n - x} + [E(x_n - x)]^2 \geq [E(x_n) - E(x)]^2 .$$

Lemma 1.5. If $\{y_h\}$ is a sequence of non-negative[5] random variables and if $\underset{h\to\infty}{\mathrm{plim}}\, y_h = y$ and $E(y_h) \leq M$ then $E(y) \leq M$.

Under the conditions of the lemma and in view of theorem 1.1 we have $\lim\limits_{h\to\infty} F_h(t) = F(t)$ where F_h and F are the cumulative distribution functions of y_h and y respectively. Suppose $E(y) > M$ then there exists a continuity point A of $F(t)$ such that $\int_0^A t\, dF(t) > M$. However $\int_0^A t\, dF_h(t) \leq M$ and $\lim\limits_{h\to\infty} \int_0^A t\, dF_h(t) = \int_0^A t\, dF(t)$, a contradiction.

---

[5] I.e., if $P(y_h < 0) = 0$.

Lemma 1.6. The sequence $\{x_n\}$ converges in the mean to a random variable x if and only if to every $\varepsilon > 0$ there exists an N such that

(1.13) $\quad E(x_m - x_n)^2 \leq \varepsilon \quad$ for all $m, n \geq N$.

Suppose first that there exists a random variable x such that $\underset{n\to\infty}{\text{l.i.m.}}\ x_n = x$. Then for sufficiently large n and m and arbitrary ε

$$E(x_n - x)^2 \leq \varepsilon \quad , \quad E(x_m - x)^2 \leq \varepsilon \ .$$

But

$$E(x_m - x_n)^2 = E(x_m - x)^2 + E(x_n - x)^2 - 2E[(x_m - x)(x_n - x)]$$

and by Schwartz's inequality

$$|E(x_m - x)(x_n - x)| \leq \sqrt{E(x_m - x)^2\ E(x_n - x)^2} \leq \varepsilon \ ;$$

hence

$$E(x_m - x_n)^2 \leq 4\varepsilon \ .$$

On the other hand from $E(x_m - x_n)^2 \leq \varepsilon$ it follows by Tchebicheff's inequality that

$$P(|x_m - x_n| \geq t\sqrt{\varepsilon}\ ) \leq \frac{1}{t^2} \ .$$

Thus $\underset{n\to\infty}{\text{plim}}\ x_n = x$ exists by theorem 1.1. It follows also that $\underset{n\to\infty}{\text{plim}}\ (x_m - x_n)^2 = (x_m - x)^2$ and thus by lemma 1.5

$$E(x_m - x)^2 \leq \varepsilon \quad \text{and} \quad \underset{m\to\infty}{\text{l.i.m.}}\ x_m = x \ .$$

Lemma 1.7. If $\underset{n\to\infty}{\text{l.i.m.}}\ x_n = x$, $\underset{n\to\infty}{\text{l.i.m.}}\ y_n = y$ and if $E(x_n^2)$, $E(y_n^2)$ exist then $\lim\limits_{n\to\infty} E(x_n y_n) = E(xy)$.

We show first that $E(x_n^2)$, $E(y_n^2)$ are bounded with respect to n. By virtue of lemma 1.6 there exists an m and an M such that $E(x_m-x_n)^2 \leq M$ for all n.[6] From the almost trivial inequality $a^2 \leq 2[(a-b)^2+b^2]$ we see that for all m

$$E(x_n^2) \leq 2[E(x_m-x_n)^2 + E(x_m^2)] \leq 2[M + E(x_m^2)]$$

From this inequality it follows that $E(x_m^2)$ is bounded for all m. Since $|E(x_n y_n)| \leq \sqrt{E(x_n^2)E(y_n^2)}$ it follows moreover that $E(xy)$ exists and furthermore

$$(1.14) \quad |E(x_n y_n - xy)| = |E[x_n(y_n-y) + y(x_n-x)]|$$
$$\leq \sqrt{E(x_n^2)E(y_n-y)^2} + \sqrt{E(y^2)E(x_n-x)^2}$$

$E(y^2)$ exists by lemma 1.5 and $E(x_n^2)$ is bounded and since $E(y_n-y)^2$, $E(x_n-x)^2$ converge to zero, the right-hand side of (1.14) converges to zero, which proves lemma 1.7. Lemma 1.7 may also be written in the form

$$(1.15) \quad E\{(\underset{n\to\infty}{\mathrm{l.i.m.}}\, x_n)(\underset{n\to\infty}{\mathrm{l.i.m.}}\, y_n)\} = \lim_{n\to\infty} E(x_n y_n)$$

or

$$(1.15a) \quad \sigma\{(\underset{n\to\infty}{\mathrm{l.i.m.}}\, x_n)(\underset{n\to\infty}{\mathrm{l.i.m.}}\, y_n)\} = \lim_{n\to\infty} \sigma_{x_n y_n}$$

<u>Corollary to lemma 1.7.</u> The sequence $\{x_n\}$ converges in the mean if and only if $\lim_{\substack{n\to\infty\\ m\to\infty}} E(x_n x_m)$ exists irrespective of the manner in which n and m tend to infinity.

- -

[6] This is seen if we determine N -- according to lemma 1.6 --, so that $E(x_m-x_n)^2 \leq \varepsilon$ for $m,n \geq N$ and then take, for a fixed $m \geq N$, $M = \max\{E(x_m-x_1)^2, E(x_m-x_2)^2, \ldots, E(x_m-x_N)^2; \varepsilon\}$

From lemma 1.7 it follows immediately that the condition is necessary. To show that the condition is also sufficient we assume that $\lim_{\substack{n\to\infty \\ m\to\infty}} E(x_n x_m)$ exists and is independent of the manner in which n and m go to infinity. Then $\lim_{\substack{n\to\infty \\ m\to\infty}} E(x_n - x_m)^2 = 0$ hence it is possible to find for every $\varepsilon > 0$ an $N = N(\varepsilon)$ such that $E(x_n - x_m)^2 \leq \varepsilon$ for $n, m \geq N$. We see therefore from lemma 1.6 that $\underset{n\to\infty}{\text{l.i.m.}}\, x_n$ exists. This corollary is due to M. Loève.

5. Differentiation. In order to be able to define derivatives of stochastic processes we have to extend the concepts of limits in probability and limits in the mean. Suppose that for every h in an interval (a, b) a random variable x_h is defined. If for every sequence $\{h_i\}$ with $\lim_{i\to\infty} h_i = a$, $\operatorname{plim}_{i\to\infty} x_{h_i} = x$ exists then we write $\operatorname{plim}_{h\to a} x_h = x$. In a similar manner we define $\underset{h\to a}{\text{l.i.m.}}\, x_h = \underset{i\to a}{\text{l.i.m.}}\, x_{h_i}$.

The process x_t is called differentiable at the point t if

$$\operatorname{plim}_{h\to 0} \frac{x_{t+h} - x_t}{h} = x_t'$$

exists. The stochastic process x_t' is called the derivative of x_t.

In the following we assume $E(x_t) = 0$. The modifications of our statements for the case $E(x_t) \neq 0$ will be obvious.

6. Stochastic processes of second order. A stochastic process x_t is called of second order if for any values t_1, t_2 the covariance $\sigma_{t_1 t_2}$ exists. The process x_t is called differentiable l.i.m. if $\underset{h \to 0}{\text{l.i.m.}} \frac{x_{t+h} - x_t}{h} = x'_t$ exists.

Theorem 1.2. Necessary and sufficient that the process x_t is differentiable l.i.m. is that the limit

$$(1.16) \quad \lim_{\substack{h \to 0 \\ k \to 0}} \frac{\sigma_{t+h,\,t+k} - \sigma_{t+h,\,t} - \sigma_{t,\,t+k} + \sigma_{t,\,t}}{hk} = \Sigma_{t,t}$$

exist. The covariance function $\sigma_{t_1 t_2}$ is then twice differentiable and $\frac{\partial^2 \sigma_{t_1 t_2}}{\partial t_1 \partial t_2} = \frac{\partial^2 \sigma_{t_1 t_2}}{\partial t_2 \partial t_1}$. Moreover x'_t is a stochastic process of second order and its covariance function is $\frac{\partial^2 \sigma_{t_1 t_2}}{\partial t_1 \partial t_2}$.

The covariance between x_{t^*} and x'_t is given by $\frac{\partial \sigma_{t^* t}}{\partial t}$.

Proof: Consider a sequence of difference quotients $\frac{x_{t+h} - x_t}{h}$.

We have

$$E\left[\frac{x_{t+h} - x_t}{h} \cdot \frac{x_{t+k} - x_t}{k}\right] = \frac{\sigma_{t+h,\,t+k} - \sigma_{t+h,\,t} - \sigma_{t,\,t+k} + \sigma_{t,\,t}}{hk}$$

and by the corollary to lemma 1.7 the relation (1.16) is necessary and sufficient for $\underset{h \to 0}{\text{l.i.m.}} \frac{x_{t+h} - x_t}{h} = x'_t$ to exist. The expression $\Sigma_{t,t}$ in (1.16) is called the generalized second derivative.

We moreover have by lemma 1.4 $E(x_t') = 0$ since $E(x_t) = E(x_{t+h}) = 0$. Furthermore by lemma 1.7 $\sigma_{x_{t^*} x_t'}$ exists and

$$(1.17) \quad \sigma_{x_{t^*} x_t'} = \lim_{h \to 0} E\left[x_{t^*} \frac{x_{t+h} - x_t}{h}\right] = \lim_{h \to 0} \frac{\sigma_{t+h,\, t^*} - \sigma_{tt^*}}{h} = \frac{\partial \sigma_{tt^*}}{\partial t} .$$

Thus $\frac{\partial \sigma_{tt^*}}{\partial t}$ exists. It also follows from lemma 1.7 that $\sigma_{x_t' x_{t^*}'}$ exists and

$$\sigma_{x_t' x_{t^*}'} = \lim_{\substack{h \to 0 \\ k \to 0}} E\left[\frac{x_{t+h} - x_t}{h} \cdot \frac{x_{t^*+k} - x_{t^*}}{k}\right]$$

$$= \lim_{\substack{h \to 0 \\ k \to 0}} \frac{\sigma_{t+h,\, t^*+k} - \sigma_{t,\, t^*+k} - \sigma_{t+h,\, t^*} + \sigma_{t,\, t^*}}{hk} = \Sigma_{t,\, t^*} .$$

It easily follows that σ_{tt^*} is twice differentiable and that

$$\frac{\partial^2 \sigma_{tt^*}}{\partial t\, \partial t^*} = \frac{\partial^2 \sigma_{tt^*}}{\partial t^*\, \partial t} = \Sigma_{t,\, t^*} .$$

It is well known that the generalized second derivative of any function $f(x,y)$ exists if $\frac{\partial^2 f}{\partial x\, \partial y}$ exists and is continuous.

Thus we have

<u>Corollary to theorem 1.2.</u> If x_t is a stochastic process of second order with covariance function σ_{tt^*} and if $\frac{\partial^2 \sigma_{tt^*}}{\partial t\, \partial t^*}$ exists and is continuous, ~~at $t = t^*$~~ then x_t' exists l.i.m. and its covariance function is $\frac{\partial^2 \sigma_{tt^*}}{\partial t\, \partial t^*} = \Sigma_{t,\, t^*}$.

The corollary to theorem 1.2 gives a convenient method to find the covariance function of x_t' if x_t' exists l.i.m.

4. Integration. Let x_t be a stochastic process defined for $a \le t \le b$. We subdivide the interval from a to b into n parts by means of the points $a = t_0, t_1, \ldots, t_n = b$ and put $\max(t_i - t_{i-1}) = \delta$. The number δ is called the modulus of the subdivision. Within every interval $t_{i-1} \le t \le t_i$ we choose a value t_i^* and form the sum

$$Y = \sum x_{t_i^*}(t_i - t_{i-1}) \tag{1.17}$$

Y is a random variable. Now consider a sequence $\{S_m\}$ of subdivisions S_m with moduli δ_m such that $\lim_{m \to \infty} \delta_m = 0$. Let $X(m)$ be the random variable corresponding by (1.17) to the subdivision S_m and some choice of the t_m^*. If then $\lim_{m \to \infty} X(m) = X$ exists and is equal for all sequences $\{S_m\}$ with modulus converging to zero and all choices of t_i^* then X is called the integral of x_t and we write

$$\int_a^b x_t \, dt = X \tag{1.18}$$

Strong continuity.[7] In the following we denote by $\mathcal{E}(\delta, \varepsilon, S)$ the event that the relations $|x_{t_i} - x_{t_k}| \leq \varepsilon$ are simultaneously satisfied for all pairs (t_i, t_k) with $|t_i - t_k| < \delta$ and belonging to a finite set S of points contained in $[a, b]$. $P[\mathcal{E}(\delta, \varepsilon, S)]$ is then the probability that the inequalities $|x_{t_i} - x_{t_k}| \leq \varepsilon$ are simultaneously fulfilled for all pairs (t_i, t_k) of a finite set S of points for which $|t_i - t_k| \leq \delta$.

The process x_t is called strongly continuous in an interval $[a, b]$ if to every ε and η there exists a $\delta = \delta(\varepsilon, \eta)$ such that for every finite set S of points contained in $[a, b]$

$$P\left[\mathcal{E}(\delta, \varepsilon, S)\right] \geq 1 - \eta. \tag{1.20}$$

For any stochastic process x_t consider a set $S = (t_1, \ldots, t_n)$ where $a \leq t_i \leq b$ $(i = 1, \ldots, n)$. We denote by M_{abS} the largest of the values $x_{t_1}, \ldots, x_{t_n}$. Let $\{S_i\}$ be a sequence of subdivisions of the interval $[a, b]$ whose moduli converge to zero. If $\operatorname{plim}_{i \to \infty} M_{abS_i} = M_{ab}$ exists and is the same for all sequences $\{S_i\}$ whose moduli converge to zero then we shall call M_{ab} the maximum of x_t in $[a, b]$. The minimum m_{ab} is similarly defined.

---

[7] The ~~definition~~ concept is due to P. Lévy.

To simplify the notation we also introduce $V_{ab} = M_{ab} - m_{ab}$ and $V_{abS} = M_{abS} - m_{abS}$. We next derive a criterion for the strong continuity of a process:

Theorem 1.3. A process x_t is strongly continuous in $[a,b]$ if and only if

(i) it possesses a maximum $M_{tt'}$ and a minimum $m_{tt'}$ in every subinterval $[t, t']$ of $[a,b]$;

(ii) for every $\varepsilon > 0$, $\eta > 0$ there exists a δ such that for every subdivision $S = \{a = t_0, t_1, \ldots, t_n = b\}$ with modulus less than δ, it is true that

$$P\left(V_{t_{i-1}t_i} \leq \varepsilon; i = 1, \ldots, n\right) \geq 1 - \eta. \tag{1.21}$$

We emphasize that (1.21) means that the probability of the simultaneous fulfillment of all the inequalities $V_{t_{i-1}t_i} \leq \varepsilon$ $(i=1,\ldots,n)$ must exceed $1-\eta$.[8]

---

[8] If R_i $(i=1,2,\ldots,n)$ are n events then $P(R_i; i\ 1,\ldots,n)$ means the probability that all n events occur simultaneously. Thus $P(R_i; i=1,\ldots,n) > k$ means that the probability of the simultaneous occurrence of all n events exceeds k, this should be carefully distinguished from the statement $P(R_i) > k$, $(i=1,\ldots,n)$, which means that the probability of the occurrence of each single event R_i exceeds k which does not imply any thing about their joint occurrence. We further emphasize that in conditional probabilities the condition is separated not by a semicolon but by a vertical bar.

We shall use the following:

Lemma 1.8. If $P(x_n \geq y) = 1$ and $\operatorname*{plim}_{n \to \infty} x_n = x$ then $P(x \geq y) = 1$.

The proof of lemma 1.8 is left to the reader.

Proof of theorem 1.3. Suppose first that conditions (i) and (ii) are fulfilled. Let S be a finite set of points. We can consider sequences of subdivisions $\{S_n\}$ such that each S_n contains S. It follows then from lemma 1.8 that

$$P(V_{ab} \geq V_{abS}) = 1 . \tag{1.22}$$

Now let $t_1', \ldots, t_m'$ be the points of S and consider the sequence $\{S_n\}$ of subdivisions $(t_0, \ldots, t_n)$ with $t_k = a + \frac{k(b-a)}{n}$. For sufficiently large n and arbitrary ε, η we have by (ii)

$$P(V_{t_{i-1}t_i} \leq \varepsilon; i=1, \ldots, n) \geq 1-\eta . \tag{1.23}$$

The relations $V_{t_{i-1}t_i} \leq \varepsilon$, $i=1, 2, \ldots, n$ imply the relations $V_{t_{i-1}t_{i+1}} \leq 2\varepsilon$, $i = 1, 2, \ldots, n-1$. Hence from (1.23) and lemma 1.8 it follows that

$$P(V_{t_{i-1}t_{i+1}} \leq 2\varepsilon; i=1, 2, \ldots, n-1) \geq 1-\eta .$$

Any two points t_i', t_k' with $|t_i' - t_k'| \leq \frac{1}{n}$ lie together in one of the intervals (t_{i-1}, t_{i+1}). Hence the above inequality and (1.22) imply ~~implies~~

$$(1.24) \qquad P\left[\mathcal{E}\left(\tfrac{1}{n}, \varepsilon, S\right)\right] \geq 1-\eta .$$

Thus (i) and (ii) imply strong continuity.

We next show that the condition is necessary. We assume that x_t is strongly continuous and let $a \leq \bar{a} < \bar{b} \leq b$. We consider two subdivisions S_n and S_m of $[\bar{a}, \bar{b}]$ both of modulus less than δ and the maxima $M_{\bar{a}\bar{b}S_n}$ and $M_{\bar{a}\bar{b}S_m}$. Let S consist of the points S_n and S_m. Then the relation $|M_{\bar{a}\bar{b}S_n} - M_{\bar{a}\bar{b}S_m}| > \varepsilon$ implies that for two points t, t' of S we must have $|x_t - x_{t'}| > \varepsilon$, $|t-t'| \leq \delta$ hence by (1.20) for sufficiently small δ and arbitrary ε, η

$$(1.25) \qquad P(|M_{\bar{a}\bar{b}S_n} - M_{\bar{a}\bar{b}S_m}| > \varepsilon) \leq \eta .$$

On account of theorem 1.1 the relation (1.25) implies that $\operatorname*{plim}_{n\to\infty} M_{\bar{a}\bar{b}S_n} = M_{\bar{a}\bar{b}}$ exists. In a similar manner it is shown that $m_{\bar{a}\bar{b}}$ exists so that condition (i) of theorem 1.3 is satisfied.

Now choose δ so that for every finite set S of points $t_1, \ldots, t_n$ and arbitrary ε, η

$$(1.26) \qquad P\left[\mathcal{E}\left(\delta, \tfrac{\varepsilon}{2}, S\right)\right] \geq 1-\eta .$$

Now let $S = \{a = t_0, t_1, \ldots, t_n = b\}$ be any subdivision of modulus less than δ, $\{S_n\}$ a sequence of subdivisions with moduli converging to zero and containing the points of S. The relation (1.26) implies

$$(1.27) \qquad P(V_{t_{i-1}t_iS_n} \leq \frac{\varepsilon}{2}\,;\ i = 1, \ldots, n) \geq 1 - \eta.$$

Now choose ε^* so that $\frac{\varepsilon}{2} \leq \varepsilon^* \leq \varepsilon$ and so that ε^* is a continuity point of the distributions of $V_{t_{i-1}t_i}$; $i = 1, 2, \ldots, n$. It then follows from (1.27)

$$P(V_{t_{i-1}t_i} \leq \varepsilon^*;\ i = 1, 2, \ldots, n) \geq 1 - \eta\ .$$

This completes the proof of theorem 1.3.

<u>Theorem 1.4.</u> Let x_t be a strongly continuous process, then

1) $X_t = \int_a^t x_\tau d\tau$ exists for every t

2) $x_t = \frac{dX_t}{dt}$

Proof: By theorem 1.3 $m_{\tau\tau'}, M_{\tau\tau'}$ exist for all pairs τ, τ' and we have for every choice of points $a = t_0, t_1, t_2, \ldots t_n = t$ and $t_{i-1} \leq t_i^* \leq t_i\ (i = 1, \ldots, n)$

$$(1.28) \quad \Sigma m_{t_{i-1}t_i}(t_i - t_{i-1}) \leq \Sigma x_{t_i^*}(t_i - t_{i-1}) \leq \Sigma M_{t_{i-1}t_i}(t_i - t_{i-1})\ .$$

To understand this inequality correctly we must remember that $m_{t_{i-1}t_i}$, $M_{t_{i-1}t_i}$, $x_{t_i^*}$ $(i = 1, 2, \ldots, n)$ are random variables and that their joint distribution is such that the inequality (1.28) holds with probability one.

Since the process is strongly continuous we have for any subdivision with sufficiently small modulus δ

$$P\left[\Sigma M_{t_{i-1}t_i}(t_i - t_{i-1}) - \Sigma m_{t_{i-1}t_i}(t_i - t_{i-1}) \leq \varepsilon(b-a)\right] \geq 1-\eta \, .$$

If S is a subdivision then we call $Y(S) = \Sigma M_{t_{i-1}t_i}(t_i - t_{i-1})$ the upper sum and $y(S) = \Sigma m_{t_{i-1}t_i}(t_i - t_{i-1})$ the lower sum corresponding to the subdivision S. We consider now a sequence of subdivisions $\{S_j\}$ with moduli $\{\delta_j\}$ such that

$$\lim_{j\to\infty} \delta_j = 0 \quad \text{and} \quad S_m \subset S_n \quad \text{if} \quad m < n \, .$$

If $Y_j = Y(S_j)$ and $y_j = y(S_j)$ are the corresponding upper and lower sums then

$$Y_j - y_j \geq y_{j+k} - y_j \geq 0$$

and hence for sufficiently large n

$$P\left[0 \leq y_{n+k} - y_n \leq \varepsilon(b-a)\right] \geq 1-\eta \, .$$

Hence the sequences of random variables $\{y_n\}$ and $\{Y_n\}$ converge and $\operatorname{plim}_{n\to\infty} y_n = \operatorname{plim}_{n\to\infty} Y_n$. From here on the proof of the existence of $X_t = \int_a^t x_\tau d\tau$ is precisely the same as that of the existence of the ordinary Riemann integral of a continuous function.

It follows also from (1.25) that

$$M_{tt'}(t'-t) \geq \int_t^{t'} x_t dt \geq m_{tt'}(t'-t) .$$

Consider now the quotient

$$\frac{X_{t_2}-X_{t_1}}{t_2-t_1} = \int_{t_1}^{t_2} \frac{x_t dt}{t_2-t_1} .$$

We have

$$m_{t_1t_2} \leq \frac{X_{t_2}-X_{t_1}}{t_2-t_1} \leq M_{t_1t_2} ,$$

and

$m_{t_1t_2} \leq x_{t_1} \leq M_{t_1t_2}$ and since the process is strongly continuous

$\underset{t_2 \to t_1}{\text{plim}} (M_{t_1t_2} - m_{t_1t_2}) = 0$. Hence $\underset{t_2 \to t_1}{\text{plim}}\, m_{t_1t_2} = \underset{t_2 \to t_1}{\text{plim}}\, M_{t_1t_2} = x_{t_1}$

and thus

$$\underset{t_2 \to t_1}{\text{plim}} \frac{X_{t_2}-X_{t_1}}{t_2-t_1} = x_{t_1} .$$

This completes the proof of theorem 1.4 .

We shall now consider stochastic processes of second order. We shall say that

$$X_t = \int_s^t x_t dt \quad \text{exists l.i.m.}$$

if the Riemann sums $\Sigma x_{t_i^*}(t_i - t_{i-1})$ converge in the mean.

<u>Theorem 1.5.</u> Let x_t be a process with covariance function $\sigma_{t_1t_2}$. The process is integrable l.i.m. in $[a,b]$ if and only if for any t in $[a,b]$, ${}_a\int^t {}_a\int^t \sigma_{t_1t_2}dt_1dt_2$ exists. The covariance function $\Sigma_{t_1t_2}$ of $X_t = {}_a\int^t x_t dt$ is given by

$$(1.29) \qquad \Sigma_{t_1t_2} = {}_a\int^{t_1} {}_a\int^{t_2} \sigma_{\tau_1\tau_2} d\tau_1 d\tau_2 .$$

The process X_t is differentiable l.i.m. and $X_t' = x_t$ if $\sigma_{t_1t_2}$ is continuous.

To prove theorem 1.5 we apply the corollary to lemma 1.7 to a sequence of Riemann sums $\{\Sigma_n\}$ with moduli going to zero. We have

$$E(\Sigma_n\Sigma_m) = E\left[\Sigma x_{t_i^*}(t_i-t_{i-1})\Sigma x_{t_j^*}(t_j-t_{j-1})\right] = \sum_{ij}\Sigma\sigma_{t_i^*t_j^*}(t_i-t_{i-1})(t_j-t_{j-1})$$

If n and m go to infinity in any manner we have

$$\lim_{\substack{n\to\infty \\ m\to\infty}} E(\Sigma_n\Sigma_m) = {}_a\int^t {}_a\int^t \sigma_{t_1t_2}dt_1dt_2 .$$

Thus $\underset{n\to\infty}{\text{l.i.m.}}\ \Sigma_n = \Sigma$ exists.

Moreover if $\{\Sigma_n'\}$ is any other sequence of Riemann sums we put $\Sigma_{2n}'' = \Sigma_n$, $\Sigma_{2n+1}'' = \Sigma_n'$. Since l.i.m. Σ_n'' exists we must have

$$\Sigma = \underset{n\to\infty}{\text{l.i.m.}}\ \Sigma_n = \underset{n\to\infty}{\text{l.i.m.}}\ \Sigma_n' = {}_a\int^t x_t dt = X_t .$$

It follows also easily from lemma 1.7 that $E(X_t X_{t'})$ exists and that

$$E(X_t X_{t'}) = \int_a^{t}\int_a^{t'} \sigma_{t_1 t_2}\, dt_1\, dt_2 . \tag{1.29a}$$

We further have

$$Y_{t'} = \frac{X_{t'} - X_t}{t' - t} - x_t = \int_t^{t'} \frac{x_\tau - x_t}{t' - t}\, d\tau .$$

From (1.29a) it follows that

$$\sigma_{Y_{t'} Y_{t'}} = \frac{1}{(t'-t)^2}\int_t^{t'}\int_t^{t'} (\sigma_{\tau\tau'} - \sigma_{t\tau'} - \sigma_{t\tau} + \sigma_{tt})\, d\tau\, d\tau' .$$

If $\sigma_{\tau\tau'}$ is continuous then by the mean value theorem of integral calculus $\sigma_{Y_{t'} Y_{t'}}$ becomes arbitrarily small if t' approaches t. Thus $x_t = X'_t$ l.i.m.

Let x_t, y_t be two ~~random variables~~ stochastic processes, then to every subdivision $S = (a = t_0, t_1, t_2, \ldots, t_n = b)$ we can form Riemann-Stieltjes sums

$$X(S) = \sum_{i=1}^{n} x_{t_i^*}(y_{t_i} - y_{t_{i-1}}) \tag{1.30}$$

If now for every sequence $\{S_n\}$ of subdivisions with modulus converging to zero $\operatorname{plim}_{n \to \infty} X(S_n)$ exists and is independent of the particular sequence $\{S_n\}$ and of the choice of points t_i^* $(t_{i-1} \leq t_i^* \leq t_i)$ then we shall write

$$X = \operatorname{plim}_{n \to \infty} X(S_n) = \int_a^b x_t dy_t$$

We shall call X the integral of x_t with respect to y_t. If the random variables $X(S_n)$ converge in the mean to X we shall say that

$$\int_a^b x_t dy_t \quad \text{exists l.i.m.}$$

<u>Theorem 1.6</u>. Let x_t, y_t be two independent stochastic processes of second order (that is to say x_t is independent of $y_{t'}$ for any t and any t') with covariance functions $\sigma_{tt'}$, $\rho_{tt'}$ respectively. The integral of x_t with respect to y_t exists l.i.m. for every interval $[a, t]$ contained in $[a, b]$ if and only if

(1.31)
$$\int_a^t \int_a^t \sigma_{t_1 t_2} d^2\rho_{t_1 t_2}$$

exists. The covariance function of

$$X_t = \int_a^t x_t dy_t$$

is moreover given by

$$\Sigma_{t_1 t_2} = \int_a^{t_1} \int_a^{t_2} \sigma_{\tau_1 \tau_2} d^2\rho_{\tau_1 \tau_2} .$$

The proof of theorem 1.6 is analogous to that of theorem 1.5 and is left to the reader.

By $P(E|E')$ we denote the conditional probability that E will happen provided E' has happened.[9] We have called a process continuous if $\operatorname{plim}_{\tau \to 0} x_{t+\tau} = x_t$. A process will be called uniformly continuous in $[a, b]$ if to every $\varepsilon > 0$, $\eta > 0$ there exists a $\delta(\varepsilon, \eta)$ independent of t such that

$$P\{|x_{t+\tau} - x_t| \leq \varepsilon\} \geq 1-\eta \quad \text{for every} \quad |\tau| \leq \delta(\varepsilon, \eta)$$

Lemma 1.9. If a process is continuous in a closed interval $[a, b]$ then it is uniformly continuous in $[a, b]$.

Proof: Consider a monotone decreasing sequence $\{\tau_i\}$ with $\lim_{i \to \infty} \tau_i = 0$. For every i consider the set of points t for which

$$P\{|x_{t+\tau_i} - x_t| > \varepsilon\} > \eta$$

Assume the lemma to be false, then we can construct a sequence $t_1, t_2, \ldots$ such that for some $\varepsilon > 0$, $\eta > 0$

$$P\{|x_{t_i+\tau} - x_{t_i}| > 2\varepsilon\} > 2\eta$$

for some $\tau < \tau_i$. Let t be an accumulation point of the sequence $\{t_i\}$. Then for every ~~ε, η and δ~~ $\delta > 0$ we can find a t_i arbitrarily close to t such that $P\{|x_{t_i+\tau} - x_{t_i}| > 2\varepsilon\} > 2\eta$ for some $|\tau| < \frac{\delta}{2}$.

---

[9] For the concept of conditional probability the reader is referred to Kolmogoroff, Grundbegriffe der Wahrscheinlichkeitsrechnung, Chapter 5, par. 1 and 3.

Choose now $|t_1 - t| < \frac{\delta}{2}$ and t_1 so close to t that

$$P(|x_{t_1} - x_t| > \varepsilon) < \eta .$$

Then

$$P(|x_{t_1+\tau} - x_t| > \varepsilon) \geq P(|x_{t_1+\tau} - x_{t_1}| > 2\varepsilon \, , \, |x_{t_1} - x_t| \leq \varepsilon)$$

$$\geq P(|x_{t_1+\tau} - x_{t_1}| > 2\varepsilon) - P(|x_{t_1} - x_t| > \varepsilon) > \eta \, .$$

Hence for arbitrary $\delta > 0$ and some $\varepsilon > 0$, $\eta > 0$ there exist values $\tau < \delta$ such that

$$P(|x_{t+\tau} - x_t| > \varepsilon) > \eta$$

in contradiction with our assumption of continuity of the process x_t.

We derive next a sufficient condition for the existence of M_{ab} and m_{ab}. We have to consider in this connection the event $\{x_k < x_i$ for $k < i$, $x_k \leq x_i$ for $k > i\}$ and we denote this event by A_i and state the following

<u>Theorem 1.7</u>. Let $S = \{a \leq t_1 < t_2, \ldots < t_n \leq b\}$ be a set of points in $[a, b]$ and let x_t be a stochastic process which

(i) is continuous in $[a, b]$,

(ii) is such that for sufficiently small τ — which is independent of S — .

$$(1.22) \qquad P(|x_{t_i+\tau} - x_{t_i}| > \varepsilon \,|\, A_i) \leq K \, P(|x_{t_i+\tau} - x_{t_i}| > \varepsilon)$$

where K is a constant independent of the choice of S. Then M_{ab} and m_{ab} exist.

Proof: Let $S_i = (t_1, \ldots, t_n)$ and $S_j = (t'_1, \ldots, t'_n)$ be two subdivisions of moduli less than δ and put for short $x_{t_\alpha} = x_\alpha$, $x_{t'_\alpha} = x'_\alpha$.

To every x_k we can find an $x'_{j_k} = y_k$ such that $t_{j_k} - t_k < \delta$.

For $P(A_k) \neq 0$ we thus have on account of Lemma 1.9 for arbitrary η and sufficiently small δ

$$(1.33) \quad P(A_k, x_k - y_k > \varepsilon) = P(A_k) P(x_k - y_k > \varepsilon \mid A_k) \leq K\eta P(A_k).$$

If $P(A_k) = 0$ the inequality (1.33) is also valid.

Denote by B_k the joint occurrence of the events A_k and $x_k - y_k \leq \varepsilon$ and let B be the event that at least one of the events B_k occurs. By definition A_k implies $x_k = M_{abS_i}$ so that B implies the existence of some y_k such that $y_k \geq M_{abS_i} - \varepsilon$. From this it follows in turn that $M_{abS_j} \geq M_{abS_i} - \varepsilon$. Therefore we see that

$$P(M_{abS_j} \geq M_{abS_i} - \varepsilon) \geq P(B) \geq \sum_k P(B_k).$$

From (1.33) we obtain easily

$$P(B_k) = P(A_k) P(x_k - y_k \leq \varepsilon \mid A_k) \geq (1 - K\eta) P(A_k).$$

The events A_k exclude each other and exhaust all the possibilities so that by adding these inequalities we obtain

$$\sum_k P(B_k) \geq 1-K\eta$$

Therefore,

(1.34) $$P(M_{abS_j} \geq M_{abS_i} - \varepsilon) \geq 1-K\eta.$$

Similarly we obtain

(1.34a) $$P(M_{abS_i} \geq M_{abS_j} - \varepsilon) \geq 1-K\eta.$$

Hence

(1.35) $$P(|M_{abS_i} - M_{abS_j}| \leq \varepsilon) \geq 1-2K\eta.$$

The existence of M_{ab} follows easily from (1.35) using theorem 1.1 and the existence of m_{ab} is proved similarly.

CHAPTER 2

SPECIAL PROCESSES

1. The fundamental random process.

A small particle suspended in a gas is subjected to a continual bombardment by the molecules of this gas. The individual impacts imparted by these molecules are small compared to the mass of the ~~gas~~ particle and the number of impacts per second is very large. The impacts are received from all directions and are randomly distributed. Moreover, if we neglect the velocity of the particle itself, which is small compared to the velocity of the molecules, the distribution of these impacts at time t will be independent of the momentum of the particle at time $t' \leq t$. If we denote by x_t the momentum of the particle, it will therefore be reasonable to assume that $x_{t+\tau}-x_t$ is independent of x_{t^*} for $t^* \leq t$. The motion of the particle is called the Brownian motion.

The momentum of the particle is a special example of a more general type of stochastic processes, called Markoff processes, which satisfy for $t_1 < t_2 < \ldots < t_n < t$ and $\tau > 0$ the equation

$$P(x_{t+\tau} \leq A | x_{t_1}, \ldots, x_{t_n}, x_t) = P(x_{t+\tau} \leq A | x_t) .$$

In words, the conditional distribution of $x_{t+\tau}$ (where $\tau > 0$), given the values of $x_{t_1}, \ldots, x_{t_n}, x_t$ (where $t_1 < t_2 < \ldots < t_n < t$) is the same as the conditional distribution of $x_{t+\tau}$ given x_t.

More conversationally speaking, if the present value of x_t is known the distribution of any future values is independent of the way in which the present value was reached.

In our special case of the motion of a particle suspended in a gas we shall make the following assumptions about its momentum x_t:

Assumption 1.

(2.1)
$$x_{t+\tau} = x_t + \varepsilon_{t,\tau}$$

where $\varepsilon_{t,\tau}$ is a random variable with mean zero and is independent of x_t and also of $\varepsilon_{t',\tau'}$ if the intervals $(t, t+\tau)$, $(t', t'+\tau')$ do not overlap.

Assumption 2.

The distribution of $\varepsilon_{t,\tau}$ depends only on τ.

Assumption 3.

The variance of $\varepsilon_{t,\tau}$ exists and is a measurable function of τ.

We have for $\tau = \tau_1 + \tau_2$, $\tau_1 \geq 0$, $\tau_2 \geq 0$

$$\varepsilon_{t,\tau} = \varepsilon_{t,\tau_1} + \varepsilon_{t+\tau_1,\tau_2}$$

and hence

(2.2)
$$\sigma^2_{\tau_1} + \sigma^2_{\tau_2} = \sigma^2_{\tau_1+\tau_2}$$

where $\sigma^2_\tau = \sigma^2_{\varepsilon_{t,\tau}}$.

From (2.2) and assumption 3 it follows by a well known theorem[10] that

(2.3) $$\sigma^2_\tau = c\tau$$

where c is some positive constant. Thus $x_{t+\tau}$ converges to x_t in the mean with decreasing τ and the process is continuous l.i.m. ¶ Suppose further that $x_0 = 0$. It follows then from (2.3) and assumption 1 that

(2.4) $$\begin{cases} \sigma^2_{x_t} = ct \\ \sigma_{x_t x_{t+\tau}} = \sigma\{x_t(x_{t+\tau} - x_t + x_t)\} = \sigma^2_{x_t} = ct \end{cases}.$$

¶ The assumptions 1 to 3 define an important class of Markoff processes, sometimes called differential processes.[11] ¶ In order to define completely our mathematical model for the Brownian motion we must also take account of the fact that we regard the impacts from the molecules as coming in a continuous stream so that large changes of the momentum in a short time interval become much less likely than small ones.

---

[10] If a measurable function f(x) satisfies the functional equation $f(x+y) = f(x) + f(y)$ then $f(x) = cx$. Proof of this theorem may be found in H. Hahn, Theorie der reellen Funktionen, Erster Band, pp. 581-3, J. Springer, Berlin (1921).

[11] We distinguish differential process from "general differential processes" (Chapter 4).

We therefore impose the following additional condition, called the Lindeberg condition, on the distribution functions $F_\tau(a)$ of $\varepsilon_{t,\tau}$.

<u>Assumption 4 (Lindeberg condition)</u>:

For sufficiently small τ and arbitrary $\rho > 0$, $\eta > 0$

$$\int_{|a| > \rho} a^2 \, dF_\tau(a) < \eta\sigma_\tau^2 . \tag{2.5}$$

Condition (2.5) may perhaps best be understood if we discuss an important case where it is fulfilled.

<u>Theorem 2.1.</u> If $\sigma_\tau^2 = \int_{-\infty}^{+\infty} a^2 \, dF_\tau(a)$ exists and if $P(\frac{\varepsilon_{t,\tau}}{\sigma_\tau} < a) = F(a)$ is independent of τ then $\varepsilon_{t,\tau}$ fulfills the Lindeberg condition (2.5).

Proof: From the conditions of the theorem we have

$$F_\tau(a\sigma_\tau) = F(a) , \quad F_\tau(a) = F(a/\sigma_\tau) .$$

Hence for arbitrarily small $\rho > 0$ and $\eta > 0$

$$\int_{|a|>\rho} a^2 \, dF_\tau(a) = \int_{|a|>\rho} a^2 \, dF(a/\sigma_\tau) = \sigma_\tau^2 \int_{|a|>\rho} \frac{a^2}{\sigma_\tau^2} \, dF(\frac{a}{\sigma_\tau})$$

$$= \sigma_\tau^2 \int_{|y|>\rho/\sigma_\tau} y^2 \, dF(y) \leq \eta\sigma_\tau^2$$

the inequality holding for sufficiently small τ since $\sigma_\tau^2 = c\tau$.

We next prove

Theorem 2.2. If $\varepsilon_{t,\tau}$ fulfills the Lindeberg condition then $\varepsilon_{t,\tau}$ is normally distributed with variance $c\tau$.

As long as we consider only the distribution of $\varepsilon_{t,\tau}$ there will not be any danger of confusion if we write ε_τ for $\varepsilon_{t,\tau}$. We shall use the following

Lemma 2.1: Let $x_1, x_2, \ldots, x_k$ be independent random variables with mean zero and with distribution functions $F_1, \ldots, F_k$ respectively and $\sigma^2_{(x_1+\ldots+x_k)} = 1$ then to every δ there exists a ρ and an η such that

$$\left| P(x_1+\ldots+x_k < a) - \int_{-\infty}^{a} (1/\sqrt{2\pi}) e^{-x^2/2} \, dx \right| < \delta$$

whenever for all k

$$\int_{|x|>\rho} x^2 \, dF_k(x) < \eta \sigma^2_{x_k}$$

A proof of this lemma can be found for instance in Khintchine's "Asymptotische Gesetze der Wahrscheinlichkeitsrechnung", p.3 (Ergebnisse der Mathematik, J. Springer, Berlin 1933).

To prove theorem 2.2 we put $\varepsilon'_\tau = \varepsilon_\tau/\sqrt{c\tau}$, then ε'_τ has variance 1. We divide the interval $0 \leq t \leq \tau$ into n equal parts and put

$$\varepsilon_i = \{(x_{t+(i\tau/n)} - x_{t+[(i-1)\tau/n]}\}/\sqrt{c\tau} \quad \text{for } i=1,2,\ldots,n .$$

The ε_i are independently distributed and $\varepsilon'_\tau = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n$.
Since the ε_i fulfill the Lindeberg condition we see from lemma 2.1 that the distribution of ε'_τ differs arbitrarily little from the normal distribution with unit variance. Hence $\varepsilon_\tau = \sigma_\tau \varepsilon'_\tau$ is normally distributed with variance $\sigma_\tau^2 = c\tau$.

The processes defined by assumptions 1, 2, 3, and 4 will be called fundamental random processes (abbreviated F.R.P.).

In the following we shall repeatedly use the fact that the distribution of the limit of a sequence of random variables equals the limit of the distribution functions in all its continuity points.

2. Further properties of the F.R.P.

Theorem 2.3. Every F.R.P. is strongly continuous.

Without loss of generality we shall assume $c = 1$; that is, $E[(x_{t+\tau} - x_t)^2] = \tau$. We derive next several lemmas needed for the proof of theorem 2.3.

Lemma 2.2. For $a > 0$ we have

$$\int_a^\infty e^{-x^2/2}\,dx \leq e^{-a^2/2}/a \; . \tag{2.6}$$

Proof:

$$a\int_a^\infty e^{-x^2/2}\,dx < \int_a^\infty x e^{-x^2/2}\,dx = e^{-a^2/2} \; .$$

Lemma 2.3. In an interval (t, t') of length δ and for every finite set S of points in this interval

$$P(V_{tt'S} \geq M) \leq (8e^{-M^2/8\delta}/M\sqrt{2\pi})\sqrt{\delta} \; . \tag{2.7}$$

We add the points t and t' to S. This can only increase the value of $\delta_{tt'S} = M_{tt'S} - m_{tt'S}$. Without loss of generality we may further assume $x_t = 0$, since we could otherwise consider the process $x'_s = x_s - x_t$. Let $t_0 = t, t_1, t_2, \ldots, t_n = t'$ be the subdivision points of S and put $x_i = x_{t_i}$. Let A_i $(i=1,2,\ldots,n)$ be the event $\{x_1 < M, \ldots, x_{i-1} < M, x_i \geq M\}$. The events A_i are exclusive and exhaust all cases for which $M_{tt'S} \geq M$.

We have with $x_0 = 0$, $M \geq 0$

$$P(A_i, x_n \geq M) = P(A_i)\, P(x_n \geq M | A_i) = P(A_i)\, P(x_n - x_i \geq M - x_i | A_i)$$

$$\geq P(A_i)\, P(x_n - x_i \geq 0 | x_1 < M, \ldots, x_{i-1} < M, x_i \geq M) \; ;$$

on account of our assumptions we know that $x_n - x_i$ is independently distributed of $x_1, \ldots, x_i$ and has a normal distribution with zero mean and variance $t_n - t_i$ so that

$$(2.8) \quad P(A_i, x_n \geq M) \geq P(A_i) \int_0^\infty \frac{1}{\sqrt{2\pi(t_n - t_i)}} \exp\left[-\frac{1}{2}\frac{(x_n - x_i)^2}{(t_n - t_i)}\right] d(x_n - x_i)$$

$$= \frac{1}{2} P(A_i) \, .$$

The events $\{A_i, x_n \geq M\}$ comprise all cases for which $x_n \geq M$ and are mutually exclusive. Adding (2.8) over all i we therefore obtain

$$(2.9) \qquad 2P(x_n \geq M) \geq P(M_{tt'S} \geq M) \, .$$

The left side is by (2,6) smaller than

$$(2.10) \qquad (2\sqrt{\delta}/M\sqrt{2\pi})e^{-M^2/2\delta}$$

The same estimate is obtained also for $P(m_{tt'S} \leq -M)$.

Furthermore

$$P(V_{tt'S} \geq M) \leq P(\text{either } M_{tt'S} \geq \tfrac{M}{2} \text{ or } m_{tt'S} \leq \tfrac{-M}{2})$$

$$\leq P(M_{tt'S} \geq \tfrac{M}{2}) + P(m_{tt'S} \leq \tfrac{-M}{2}) \leq (8\sqrt{\delta}/M\sqrt{2\pi})e^{-M^2/8\delta},$$

which establishes (2,7).

Lemma 2,4. Let $t'-t=\ell$ and consider subdivisions $S_n=\{t=t_0, t_1, \ldots, t_n=t'\}$ such that $t_i-t_{i-1}=\ell/n=\delta_n$. Then there exists for every positive ε, η an N such that for an arbitrary ~~choice~~ finite set S of points

$$(2.11) \qquad P(V_{t_{i-1}t_iS} \leq \varepsilon;\ i=1,\ldots,n) \geq 1-\eta \quad \text{for } n>N.$$

To prove (2,11) we add the points of S_n to S. This will at most decrease the probability in (2,11). Since the distribution of $V_{t_{i-1}t_iS}$ is independent of the distribution of that of $V_{t_{j-1}t_jS}$ for $i \neq j$ we have by lemma 2,3

$$(2.12) \quad P(V_{t_{i-1}t_iS} \leq \varepsilon;\ i=1,\ldots,n) \geq \left[1-\frac{8\exp(-n\varepsilon^2/8\ell)\sqrt{\ell}}{\sqrt{2\pi}\,\varepsilon\sqrt{n}}\right]^n$$

$$\geq (1-ke^{-nk'})^n$$

where k and k' are positive constants independent of n. Since it is easily seen that $\lim_{n\to\infty}(1 - k e^{-nk'})^n = 1$, lemma 2.4 follows.

<u>Lemma 2.5.</u> $M_{tt'}$ and $m_{tt'}$ exist for every interval $[t,t']$.

Consider a sequence of subdivisions $\{S_i\}$ of the interval $[t,t']$ of moduli $\{\delta_i\}$ with $\lim_{i\to\infty}\delta_i = 0$. If S_i and S_j have both sufficiently small moduli then in every interval of length δ_n of lemma 2.4 there will be at least one point of S_i and one of S_j. Hence applying lemma 2.4 to the union of S_i and S_j we obtain

$$P(|M_{tt'}S_i - M_{tt'}S_j| \geq \varepsilon) \leq \eta .$$

Thus $\operatorname{plim}_{i\to\infty} M_{tt'}S_i = M_{tt'}$ exists for every sequence $\{S_n\}$ whose moduli converge to zero and is the same for every such sequence. In a similar manner it is possible to prove that $\operatorname{plim}_{i\to\infty} m_{tt'}S_i = m_{tt'}$ exists.

To prove theorem 2.3 let S be any subdivision of modulus $\leq \delta/2$ and consider

$$(2.13) \qquad P(V_{t_{i-1}t_i} \leq \varepsilon;\ i=1,2,\ldots,n) .$$

We form a new subdivision with $\delta/2 \leq t_i - t_{i-1} \leq \delta$ by deleting points of S. The probability (2.13) for this new subdivision is smaller than that for S.

By lemma 2.3 we have, since the distribution of $V_{t_{i-1}t_iS}$ converges to the distribution of $V_{t_{i-1}t_i}$,

$$(2.14) \qquad P(V_{t_{i-1}t_i} \leq \varepsilon; i=1,\ldots,n) \geq \left[1-\frac{8\exp(-\varepsilon^2/8\delta)}{\varepsilon\sqrt{2\pi}}\right]^{2\ell/\delta}$$

The limit on the right of (2.14) is 1 for $\delta \to 0$, which proves theorem 2.3.

<u>Theorem 2.4.</u> For the F.R.P.

$$(2.15) \qquad P(M_{ab}-x_a \geq M) = 2P(x_b-x_a \geq M)$$

Proof: We consider the proof of lemma 2.3 and write $t'=b$, $t=a$ we see from (2.9) that $2P(x_b \geq M) \geq P(M_{abS_i} \geq M)$. Here S_i may be any set of points in the interval $[a,b]$. Let S_i be an element of a sequence of subdivisions $\{S_i\}$ whose moduli go to zero. According to lemma 2.5 $\operatorname{plim}_{i\to\infty} M_{abS_i} = M_{ab}$ exists, hence

$$(2.16) \qquad 2P(x_b \geq M) \geq P(M_{ab} \geq M)$$

We consider again the events A_i of lemma 2.3 for a subdivision of $[a,b]$ into equal parts. We then have for ~~a point~~ x_c with $c > b$ and with $x_a = 0$

$$(2.17)\quad P(A_i, x_c \geq M) = P(A_i, x_c - x_i \geq 0) + P(A_i, M \leq x_c < x_i)$$

$$= \tfrac{1}{2}P(A_i) + P(A_i, M \leq x_c < x_i)$$

$$\leq \tfrac{1}{2}P(A_i) + P(A_i, x_{i-1} \leq x_c < x_i) .$$

Let ε be an arbitrary positive number; then

$$(2.18)\quad P(A_i, x_{i-1} \leq x_c < x_i) = P(A_i, x_i - x_{i-1} \leq \varepsilon, x_{i-1} \leq x_c < x_i)$$

$$+ P(A_i, x_i - x_{i-1} > \varepsilon, x_{i-1} \leq x_c < x_i)$$

$$\leq P(A_i, 0 < x_i - x_c \leq \varepsilon) + P(x_i - x_{i-1} > \varepsilon)$$

$$\leq P(A_i, 0 \leq x_b - x_c \leq \varepsilon) + P(x_i - x_{i-1} > \varepsilon)$$

since $x_i - x_c$ has larger variance then $x_b - x_c$ and both are normally distributed with mean zero. Adding the inequalities (2.17) and using (2.18) we obtain

$$(2.19)\quad P(x_c \geq M, M_{ab} \geq M) \leq \tfrac{1}{2}P(M_{ab} \geq M) + P(0 \leq x_b - x_c \leq \varepsilon)$$

$$+ \Sigma P(x_i - x_{i-1} > \varepsilon) .$$

From (2.7) we see easily that $\Sigma P(x_i - x_{i-1} > \varepsilon)$ converges to zero for every positive ε and every sequence $\{S_n\}$ whose moduli converges to zero. Since ε was arbitrary we have

$$(2.20)\quad P(x_c \geq M, M_{ab} \geq M) \leq \tfrac{1}{2}P(M_{ab} \geq M) .$$

Since c may be chosen arbitrarily close to b, it follows from (2.20) that

(2.21) $$2P(x_b \geq M) \leq P(M_{ab} \geq M) .$$

The inequalities (2.21) and (2.16) together imply theorem 2.4.

<u>Corollary to theorem 2.4.</u> Let $S_1, S_2, \ldots$ be a sequence of subdivisions of the interval $(t, t+\tau)$ with moduli converging to zero and consider for each $S_n = \{t=t_1, \ldots, t_n = t+\tau\}$ the probability

$$P_n = P(x_{t_2} - x_t \leq 0, \ldots, x_{t_n} - x_t \leq 0) ;$$

then $\lim_{n\to\infty} P_n = 0$.

Proof: We have

$$\lim_{n\to\infty} P_n = P(M_{t,t+\tau} - x_t = 0) = 1 - P(M_{t,t+\tau} - x_t > 0) = 1 - 2P(x_{t+\tau} - x_t > 0) = 0 .$$

<u>Theorem 2.5.</u> If $\rho \geq 1/2$, $\operatorname{plim}_{\tau\to 0} (x_{t+\tau} - x_t)/\tau^\rho$ does not exist.

It is zero if $\rho < 1/2$.

The proof of this theorem is left to the reader.

<u>Theorem 2.6.</u> Let $[a,b]$ be any interval and ε, η arbitrary positive numbers and $\rho < 1/2$. Then there exists a δ such that for every subdivision $S = \{a = t_0, t_1, \ldots, t_n = b\}$ of modulus not exceeding δ

$$P[V_{t_{i-1}t_i}/(t_i - t_{i-1})^\rho \leq \varepsilon;\ i=1, \ldots, n] \geq 1-\eta$$

For the proof we need the following

Lemma 2.6. For $\delta_1 > 0$, $\delta_2 > 0$, $k \geq 0$, $k' > 0$, $\nu > 0$, and $\delta_1 + \delta_2$ sufficiently small

(2.22) $[1-k\exp(-k'/\delta_1^\nu)][1-k\exp(-k'/\delta_2^\nu)] \geq 1-k\exp[-k'/(\delta_1+\delta_2)^\nu]$

Proof of lemma 2.6: The left side of (2.22) is not smaller than

$$1-k\{\exp(-k'/\delta_1^\nu) + \exp(-k'/\delta_2^\nu)\} .$$

Hence (2.22) is proved if we prove for sufficiently small δ_1 and δ_2

$$F(\delta_1, \delta_2) = \exp[-k'/(\delta_1+\delta_2)^\nu] - \exp[-k'/\delta_1^\nu] - \exp[-k'/\delta_2^\nu] \geq 0$$

We have

$$\lim_{\substack{\delta_1 \to 0 \\ \delta_2 \to 0}} F(\delta_1, \delta_2) = 0 \qquad \text{and}$$

$$\frac{\partial F}{\partial \delta_1} = \nu k' \left\{ \frac{\exp[-k'/(\delta_1+\delta_2)^\nu]}{(\delta_1+\delta_2)^{\nu+1}} - \frac{\exp[-k'/\delta_1^\nu]}{\delta_1^{\nu+1}} \right\}$$

$$\frac{\partial F}{\partial \delta_2} = \nu k' \left\{ \frac{\exp[-k'/(\delta_1+\delta_2)^\nu}{(\delta_1+\delta_2)^{\nu+1}} - \frac{\exp[-k'/\delta_2^\nu]}{\delta_2^{\nu+1}} \right\}$$

The function $x^{-v-1} \exp(-k'/x^v)$ is monotonically increasing for sufficiently small x. We have therefore $\partial F/\partial\delta_1 > 0$, $\partial F/\partial\delta_2 > 0$ for sufficiently small $\delta_1+\delta_2$. Hence $F(\delta_1, \delta_2)$ is positive for sufficiently small $\delta_1+\delta_2$.

We proceed to prove theorem 2.6. We have by (2.7) with $1-2\rho = v$, $\delta_i = t_i - t_{i-1} \leq 1$

$$P\left[(V_{t_{i-1}t_i})/\delta_i^{\rho} \leq \varepsilon\right] \geq 1 - (8\sqrt{\delta_i^v}/\varepsilon\sqrt{2\pi}) \exp(-\varepsilon^2/8\delta_i^v)$$

$$\geq 1 - k \exp(-k'/\delta_i^v)$$

where k and k' are independent of the subdivision.

Thus

$$(2.23) \quad P = P\left[\frac{V_{t_{i-1}t_i}}{(t_i - t_{i-1})^{\rho}} \leq \varepsilon;\ i = 1,\ldots,n\right] \geq \prod_{i=1}^{i=n}\left[1 - k\exp\left(-\frac{k'}{\delta_i^v}\right)\right].$$

Now let S have modulus $\frac{\delta}{2}$ then by lemma 2.6 we may combine the intervals to the right of (2.23) in such a way that all intervals are at least of length $\frac{\delta}{2}$ and at most of length δ. Hence

$$(2.24) \qquad P \geq \left[1 - k\exp(-k'/\delta^v)\right]^{2(b-a)/\delta}$$

and the right hand side of (2.24) is arbitrarily close to one if δ is sufficiently small. This completes the proof of theorem 2.6.

A process is called Gaussian if the joint distribution of $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ is normal for every choice of $t_1, t_2, \ldots, t_n$.

We now consider the integral of a F.R.P. and prove

Theorem 2.7. Let x_t be a F.R.P. with variance ct. The process $X_t = \int_0^t x_\tau \, d\tau$ is a Gaussian process with mean zero and covariance ~~correlation~~ function $\sigma_{tt'} = \frac{c}{2} \max(t, t')[\min(t, t')]^2 - \frac{c}{6}[\min(t, t')]^3$.

The integral X_t exists l.i.m. by theorem 1.5. By lemma 1.4 we have $E(X_t) = 0$ and by theorem 1.5 for $t' \geq t$

$$\sigma_{tt'} = c \int_0^t \int_0^{t'} \min(\tau, \tau') d\tau d\tau' = c \int_0^t d\tau \int_0^\tau \tau' d\tau' + c \int_0^t \tau \, d\tau \int_\tau^{t'} d\tau' = c \, \frac{t' t^2}{2} - c \, \frac{t^3}{6} .$$

Each of the approximating Riemann sums is normally distributed and that X_t itself is Gaussian follows from the following lemma:

Lemma 2.7. Let $x_n = (x_n^1, x_n^2, \ldots, x_n^s)$ be a sequence of normally distributed vectors with mean 0 and assume that $\lim_{n\to\infty} \sigma_{x_n^i x_n^j} = \sigma_{ij}$ exists. If $\operatorname{plim}_{n\to\infty} x_n = x$ then x is normally distributed with mean zero and covariance matrix σ_{ij}.

Proof: The inequality (1.5) may be derived also for vectors if we interpret $x \leq a$ to mean that the vector a-x has non-negative components. Lemma 2.7 then follows easily from the fact that for arbitrarily small δ and sufficiently large n

$$F_n(a+\delta) + \delta \geq F(a) \geq F_n(a-\delta) - \delta \ .$$

where $F_n(a)$, $F(a)$ are the distribution functions of x_n and x respectively.

3. Frictional effects. The Ornstein-Uhlenbeck process.

We have so far in the Brownian motion neglected the effect of the motion of the particle itself on $\varepsilon_{t\tau}$. If the particle has the momentum x_t then the random impulses will have a mean value proportional to x_t itself. This leads to the equation

$$(2.25) \quad \begin{cases} x_{t+\tau} = a_\tau x_t + \varepsilon_{t,\tau} \\ a_0 = 1 = \lim\limits_{\tau \to 0} a_\tau \ , \ a_\tau < 1 \text{ for } \tau > 0 \ , \ E(x_0) = 0 \end{cases}$$

where again $\varepsilon_{t,\tau}$ is normally distributed with mean value zero and variance σ_τ^2 and is independent of x_t and of $\varepsilon_{t',\tau'}$ if the intervals $(t, t+\tau)$, $(t', t'+\tau')$ do not overlap. We shall further assume that x_t is normally distributed and that a_τ is a measurable function of τ.

It follows from (2.25) that

$$(2.26)\quad \begin{cases} x_{t+\tau_1+\tau_2} = a_{\tau_2} a_{\tau_1} x_t + a_{\tau_2}\varepsilon_{t,\tau_1} + \varepsilon_{t+\tau_1,\tau_2} \\ E(x_{t+\tau_1+\tau_2} \mid x_t) = a_{\tau_1+\tau_2} x_t = a_{\tau_1} a_{\tau_2} x_t \; . \end{cases}$$

Hence $a_{\tau_1+\tau_2} = a_{\tau_1} a_{\tau_2}$ from which it follows that $a_\tau = e^{\alpha\tau}$ and since $a_\tau < 1$, $a_\tau = e^{-\beta\tau}$, $\beta > 0$. We further have from (2.25) and (2.26)

$$a_{\tau_2}\varepsilon_{t,\tau_1} + \varepsilon_{t+\tau_1,\tau_2} = \varepsilon_{t,\tau_1+\tau_2} \; .$$

Thus

$$a_{\tau_2}^2 \sigma_{\tau_1}^2 + \sigma_{\tau_2}^2 = \sigma_{\tau_1+\tau_2}^2 = a_{\tau_1}^2 \sigma_{\tau_2}^2 + \sigma_{\tau_1}^2$$

or

$$(\sigma_{\tau_1}^2 / \sigma_{\tau_2}^2) = (1 - a_{\tau_1}^2)/(1 - a_{\tau_2}^2) = (1 - e^{-2\beta\tau_1})/(1 - e^{-2\beta\tau_2}) \; .$$

Therefore

$$(2.27)\qquad \sigma_\tau^2 = \sigma^2(1 - e^{-2\beta\tau}) \; , \quad \beta > 0 \; .$$

Moreover from (2.25) and our assumption about $\varepsilon_{t,\tau}$ we have

$$(2.28)\qquad \sigma_{x_{t+\tau}}^2 = a_\tau^2 \sigma_{x_t}^2 + \sigma^2(1 - e^{-2\beta\tau})$$

⟵ If τ approaches infinity then $\sigma_{x_{t+\tau}}^2$ approaches σ^2. That is to say, if the particle has been subjected to these random impacts for a long time then the distribution of its momentum approaches a steady state.

We shall therefore assume that the process is stationary, that is to say, that the joint distribution of $x_{t_1}, \ldots, x_{t_n}$ is the same as that of $x_{t_1+h}, \ldots, x_{t_n+h}$. Under this assumption $\sigma^2_{x_{t+\tau}} = \sigma^2_{x_t} = \sigma^2$ and it follows from (2.27) that

(2.29) $$\sigma_{x_t x_{t+\tau}} = a_\tau \sigma^2 = \sigma^2 \exp(-\beta\tau) \text{ for } \tau \geq 0 .$$

The x_t process, satisfying the assumptions listed above, was first considered by L. S. Ornstein and G. E. Uhlenbeck. We will call it the Ornstein-Uhlenbeck process (abbreviated O. U. P.).

We next consider the process given by (2.30).

(2.30) $$\begin{cases} v_t = t^{\frac{1}{2}} x_{(\log t)/2\beta} & \text{for } t > 0 \\ v_t = 0 & \text{for } t \leq 0 \end{cases}$$

we have for $t' \geq t$

(2.31) $$\begin{cases} E(v_t v_{t'}) = \sigma^2 t \\ E[(v_{t'} - v_t) v_s] = 0 & \text{for } 0 \leq s < t < t' . \end{cases}$$

Moreover the $v_{t_1}, \ldots, v_{t_n}$ are jointly normally distributed. Thus the v_t process is a F.R.P. From the properties of the v_t process we obtain

Theorem 2.8. Let x_t be an O.U.P. Then

(I) x_t is continuous;

(II) x_t is not differentiable;

(III) M_{ab} and m_{ab} exist for the x_t process and it is strongly continuous;

(IV) the following equation — derived from theorem 2.4 — holds for $M \geq 0$

$$(2.32) \qquad \underset{-\infty \leq \tau \leq (1/2\beta)\log t}{P}(e^{\beta\tau}x_\tau < M) = 1 - 2\int_{\frac{M}{\sigma\sqrt{t}}}^{\infty}(1/\sqrt{2\pi})\exp(-x^2/2)\,dx$$

In equation (2.32) we have written

$$\underset{a \leq t \leq b}{P}(y_t \leq M) \quad \text{for} \quad P(M_{ab} \leq M)$$

where M_{ab} is the maximum of the process y_t in $a \leq t \leq b$. This notation will also be used in what follows.

Equation (2.32) does not seem of great use as it stands, but we can obtain from it a bound for $\underset{\tau_1 \leq \tau \leq \tau_2}{P}(x_\tau \leq M)$ as follows

$$\underset{-\infty \leq \tau \leq \tau_2}{P}(e^{\beta\tau}x_\tau \leq M) \leq \underset{\tau_1 \leq \tau \leq \tau_2}{P}(e^{\beta\tau}x_\tau \leq M) \leq \underset{\tau_1 \leq \tau \leq \tau_2}{P}(x_\tau \leq Me^{-\beta\tau_1});$$

and thus

$$(2.33)\qquad \underset{\tau_1\le\tau\le\tau_2}{P}(x_\tau \le M) \ge 1 - 2\int_{Me^{-\beta(\tau_2-\tau_1)}/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \quad .$$

<u>Theorem 2.9.</u> Let x_t be an O.U.P. The integral $X_t = \int_0^t x_\tau d\tau$ of this process exists then l.i.m. and for $t_2 \ge t_1$ its covariance function is given by

$$(2.34)\qquad R_{t_1t_2} = \frac{\sigma^2}{\beta^2}\left[e^{-\beta t_1} + e^{-\beta t_2} + 2\beta t_1 - 1 - e^{-\beta(t_2-t_1)}\right] .$$

Moreover

$$(2.35)\qquad B_t = \beta(X_t - X_0) + x_t - x_0$$

is a F.R.P.

Proof: For $t_2 \ge t_1$ we have by theorem 1.5

$$R_{t_1t_2} \cancel{= \sigma^2\int_0^{t}\int_0^{t} e^{-\beta(\tau_1-\tau_2)}\, d\tau_1 d\tau_2} = \sigma^2\int_0^{t_1}\int_0^{\tau_1} e^{-\beta(\tau_1-\tau_2)}\, d\tau_1 d\tau_2$$

$$+\sigma^2\int_0^{t_1}\int_{\tau_1}^{t_2} e^{-\beta(\tau_2-\tau_1)}\, d\tau_2 d\tau_1$$

$$= \frac{\sigma^2}{\beta^2}\left[e^{-\beta t_1} + e^{-\beta t_2} + 2\beta t_1 - 1 - e^{-\beta(t_2-t_1)}\right];$$

and for $t_1 = t_2$

$$E(X_t - X_0)^2 = \frac{2\sigma^2}{\beta^2}(e^{-\beta t} + \beta t - 1)$$

A straightforward calculation gives for $t_2 \geq t_1 \geq s_2 \geq s_1$:

$$(2.36) \qquad E[(X_{t_2} - X_{t_1})(X_{s_2} - X_{s_1})]$$

$$= \frac{\sigma^2}{\beta^3}(e^{\beta s_2} - e^{\beta s_1})(e^{-\beta t_1} - e^{-\beta t_2})$$

$$= -\frac{1}{\beta^2} E[(x_{t_2} - x_{t_1})(x_{s_2} - x_{s_1})]$$

Using $F(X_t x_s) = E(X_t \lim_{h \to 0} \frac{X_{s+h} - X_s}{h})$

we obtain from lemma 1.7 and (2.34)

$$(2.37) \quad E(X_t x_s) = \begin{cases} \frac{\sigma^2}{\beta}\left[2 - e^{-\beta s} - e^{-\beta(t-s)}\right] & \text{if } t > s \\ \frac{\sigma^2}{\beta}\left[e^{-\beta(s-t)} - e^{-\beta s}\right] & \text{if } t \leq s \end{cases} .$$

From (2.37) we see easily

$$(2.38) \qquad E[(X_{t_2} - X_{t_1})(x_{s_2} - x_{s_1})]$$

$$= \frac{\sigma^2}{\beta}(e^{\beta s_2} - e^{\beta s_1})(e^{-\beta t_1} - e^{-\beta t_2})$$

$$= -E[(X_{s_2} - X_{s_1})(x_{t_2} - x_{t_1})]$$

It is easily seen from the relations (2.36) and (2.38) that

(2.35) $$B_t = \beta(X_t - X_0) + x_t - x_0$$

has the property that for $t_2 \geq t_1 \geq s_2 \geq s_1$ the difference $B_{t_2} - B_{t_1}$ is independent of $B_{s_2} - B_{s_1}$ and since B_t is normally distributed it is a F.R.P.

From (2.35) we have

$$B_t - B_{t'} = \beta(X_t - X_t') + x_t - x_{t'} .$$

Thus for every function $f(t)$ for which the operations indicated below have meaning, we have

(2.39) $$\int_a^t f(t)\, dx_t = -\beta\int_a^t f(t) x_t\, dt + \int_a^t f(t)\, dB_t .$$

We may write (2.39) as a stochastic differential equation

(2.40) $$dx_t = -\beta x_t\, dt + dB_t .$$

In the form (2.40) a stochastic differential equation has meaning even if the processes are not differentiable. This interpretation of a stochastic differential equation is due to J.L.Doob. The equation (2.40) may be interpreted as the equation of the motion of a particle (x_t being its velocity at time t) subject to random impacts when the frictional force is proportional to its velocity. One could also interpret x_t as an electric potential subject to random changes when the decrease in potential is proportional to the potential itself.

(Thus we may consider a condenser which is charged by a randomly fluctuating current and at the same time grounded through a resistance). In short, equation (2.40) describes any situation in which a quantity x_t is subject to random changes and to a systematic decrease proportional to x_t itself.

CHAPTER 3

ESTIMATION OF PARAMETERS

In the preceding chapter we discussed Markoff processes; we shall now apply our results to obtain estimates of the parameters determining these processes from observations. In our estimating procedures we shall assume that we have at least one curve at our disposal registering the values x_t for all values $0 \leq t \leq T$. Actually it would be sufficient to know x_t for any dense set in this interval. This procedure may not seem realistic since we never observe the process for every time point. Every method of registering the curve described by x_t will itself affect x_t and in particular smooth the path curve of x_t. Thus what we observe is really a modified process.

However the methods of observation may be so refined as to give us the value of x_t in a large number of points and at any rate the variances of our estimates if obtained from discrete points may also be computed.

1. Estimation of the parameter of the F.R.P.

The F.R.P. is completely known if the constant $\frac{E(x_{t+\tau} - x_t)^2}{\tau}$ = c is known. We first discuss the estimation of the parameter c of a F.R.P.

Theorem 3.1 If x_t is a F.R.P. and if it is known in a dense set in an arbitrary small interval, then it is possible to estimate the parameter c with arbitrarily high precision.

Proof: Assume that N observations are taken in the interval $0 \leq t \leq T$, let $\tau = T/N$ and $x_{n\tau}$ $(n = 0,1,2,\ldots,N)$ be the sample value at the time $n\tau$. Since x_t is a F.R.P. the variates

$$y_n = \frac{x_{n\tau} - x_{(n-1)\tau}}{\sqrt{\tau}} = \frac{\varepsilon_{(n-1)\tau,\tau}}{\sqrt{\tau}} \quad (n=1,2,\ldots,N)$$

are normally and independently distributed with mean zero and variance c. The maximum likelihood estimate of the variance of y_n is therefore given by

$$(3.1) \qquad \hat{c} = \frac{1}{N}\frac{1}{\tau}\sum_{n=1}^{N}(x_{n\tau} - x_{(n-1)\tau})^2 = \frac{1}{T}\sum_{n=1}^{N}(x_{n\tau} - x_{(n-1)\tau})^2$$

We have $E\hat{c} = c$. Moreover $N\hat{c}/c$ has the chi-square distribution with N degrees of freedom. Its variance is therefore $2N$. Hence $\hat{c}$ has the variance $2c^2/N$ and this can be made arbitrarily small by taking N large enough.

Thus if it were possible to observe the process completely in any interval, however small, we could determine c accurately. Actually however every registering instrument will introduce a time lag and will thus smooth the process. We may infer however from our result that the points for which we read the value of x_t should be spaced as closely together as possible. That is to say, as close as is consistent with the assumption that the values of $x_{t+\tau} - x_t$ obtained still represent the actual values supplied by the F.R.P.

2. F.R.P. with mean value function.

More generally we prove

Theorem 3.2 If y_t is a stochastic process such that $y_t = x_t + f(t)$ where x_t is a F.R.P. and $f(t)$ a function of bounded variation satisfying in $(0,T)$ a Lipschitz condition $\left|\frac{f(t+\tau)-f(t)}{\tau}\right| \leq M$ for some M, and if y_t is known in a dense set of an arbitrarily small interval, then it is possible to estimate the parameter c of the F.R.P. x_t with arbitrarily high precision.

Proof: Let again $\tau = T/N$ and consider the sample points $x_{n\tau}$, $(n = 0,1,2,\ldots,N)$. Denote by $\hat{c} = \frac{1}{N}\sum_{n=1}^{N} \frac{(y_{n\tau} - y_{(n-1)\tau})^2}{\tau}$. Then

$$E\left(\frac{\hat{c}}{c}\right)^2 = \frac{1}{N^2} E\left\{\sum \frac{[x_{n\tau} - x_{\overline{n-1}\tau} + f(n\tau) - f(\overline{n-1}\tau)]^2}{c\tau}\right\}^2$$

Here and in the immediately following formulae the summation is to be extended from $n=1$ to $n=N$. We write also $\overline{n-1}$ for $(n-1)$ to simplify such expressions as $x_{(n-1)\tau} = x_{\overline{n-1}\tau}$, $y_{(n-1)\tau} = y_{\overline{n-1}\tau}$ or $f[(n-1)\tau] = f(\overline{n-1}\tau)$. Then

$$E\left(\frac{\hat{c}}{c}\right)^2 = \frac{1}{N^2}. E\left\{\frac{1}{c\tau}\sum(x_{n\tau} - x_{\overline{n-1}\tau})^2 + \frac{2}{c\tau}\sum[f(n\tau) - f(\overline{n-1}\tau)](x_{n\tau} - x_{\overline{n-1}\tau}) + \frac{1}{c\tau}\sum[f(n\tau) - f(\overline{n-1}\tau)]^2\right\}^2$$

We expand the right member of this expression; a considerable simplification follows from the assumption that x_t is a F.R.P., we use in particular the fact that $\frac{x_{n\tau} - x_{\overline{n-1}\tau}}{\sqrt{c\tau}}$ is normally distribut

with zero mean and unit variance and is independent of $\frac{x_{m\tau} - x_{\overline{m-1}\tau}}{\sqrt{c\tau}}$ for $m \neq n$. Thus we obtain

$$E\left(\frac{\hat{c}}{c}\right)^2 = 1 + \frac{2}{N} + \frac{2}{N}\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{c\tau} + \frac{4}{N^2}\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{c\tau}$$

$$+ \frac{1}{N^2}\left\{\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{c\tau}\right\}^2,$$

and

$$E\left(\frac{\hat{c}}{c}\right) = 1 + \frac{1}{N}\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{c\tau}$$

Hence

$$E(\hat{c} - c)^2 = \frac{2c^2}{N} + \frac{4c}{N2}\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{\tau} + \frac{1}{N^2}\left\{\sum \frac{[f(n\tau) - f(\overline{n-1}\tau)]^2}{\tau}\right\}^2.$$

Thus $\hat{c}$ converges in the mean to c and $\hat{c} - c$ is stochastically of the order $1/\sqrt{N}$. In fact $E(\hat{c} - c)^2 \leq \frac{2c^2}{N} + \frac{4c}{N^2} MV + \frac{1}{N^2}(MV)^2$.

Here $\left|\frac{f(t+\tau) - f(t)}{\tau}\right| \leq M$ and V is the variation of $f(t)$ so that also $V \leq MT$ where T is the length of the interval.

Thus in estimating the function $f(t)$ we may assume c to be known, if we know y_t in any interval completely.

We shall discuss two examples of the function $f(t)$. In the first we assume that

$$f(t) = at$$

(Since we can always consider the process $y_t - y_0$, this assumption is identical with the assumption $f(t) = at + b$.) We then know that the $y_{t+\tau} - y_t$ are normally distributed and independent in non-overlapping intervals with mean $a\tau$ and variance $c\tau$. Hence the maximum likelihood estimate of a, given the values at time $0, \tau, \ldots, N\tau$ becomes

$$\hat{a} = \frac{1}{T}\Sigma(y_{n\tau} - y_{\overline{n-1}\,\tau}) = \frac{y_T - y_0}{T}$$

Its variance is

$$\sigma_{\hat{a}}^2 = \frac{c}{T}$$

For the second example we assume that f(t) is given by

$$f(t) = \sum_{j=1}^{m} (\alpha_j \cos jt + \beta_j \sin jt)$$

and that the values of y_t are known in the interval $(0, 2\pi)$ and that $y_0 = 0$. If we just choose the values $y_0, y_\tau, y_{2\tau}, \ldots, y_{n\tau}$ where $n\tau = 2\pi$, the maximum likelihood estimates of the α_j and β_j will be given by those values, which minimize the expression

$$\sum_{i=1}^{n} \{y_{i\tau} - y_{\overline{i-1}\,\tau} - \sum_{j=1}^{m} \alpha_j[\cos ij\tau - \cos(i-1)j\tau]$$
$$- \sum_{j=1}^{m} \beta_j[\sin ij\tau - \sin(i-1)j\tau]\}^2$$

Hence the maximum likelihood equations are

$$(3.2.1)\quad \sum_{i=1}^{n} (y_{i\tau} - y_{\overline{i-1}\,\tau})[\cos ik\tau - \cos(i-1)k\tau]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\alpha}_j [\cos ij\tau - \cos(i-1)j\tau][\cos ik\tau - \cos(i-1)k\tau]$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\beta}_j [\sin ij\tau - \sin(i-1)j\tau][\cos ik\tau - \cos(i-1)k\tau]$$

$$k = 1, \ldots, m$$

and

$$(3.2.2)\quad \sum_{i=1}^{n} (y_{i\tau} - y_{\overline{i-1}\,\tau})[\sin ik\tau - \sin(i-1)k\tau]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\alpha}_j [\cos ij\tau - \cos(i-1)j\tau][\sin ik\tau - \sin(i-1)k\tau]$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\beta}_j [\sin ij\tau - \sin(i-1)j\tau][\sin ik\tau - \sin(i-1)k\tau]$$

$$k = 1, 2, \ldots, m$$

If we divide the first of these equations by τ and let τ go to zero, we obtain because of the orthogonality of the sine and cosine functions

$$-k \int_0^{2\pi} \sin kt \, dy_t = \hat{\alpha}_k k^2 \int_0^{2\pi} \sin^2 kt \, dt = \hat{\alpha}_k \pi k^2$$

The rules of calculation for the integral in the left member of this equation are completely analogous to those for the ordinary Riemann-Stieltjes integral. Integration by parts on the left gives therefore

$$\int_0^{2\pi} y_t \cos kt \, dt = \hat{\alpha}_k \pi$$

and thus

$$\hat{\alpha}_k = \frac{1}{\pi} \int_0^{2\pi} y_t \cos kt \, dt = -\frac{1}{\pi k} \int_0^{2\pi} \sin kt \, dy_t \tag{3.3.1}$$

Similarly, we obtain from (3.2.2)

$$k \int_0^{2\pi} \cos kt \, dy_t = \pi k^2 \hat{\beta}_k$$

Integration by parts gives again

$$\hat{\beta}_k = \frac{1}{\pi k}(y_{2\pi} - y_0) + \frac{1}{\pi} \int_0^{2\pi} y_t \sin kt \, dt = \frac{1}{\pi k} \int_0^{2\pi} \cos kt \, dy_t \tag{3.3.2}$$

The integrals in (3.3.1) and (3.3.2) are to be understood as stochastic limits of Riemann sums, and it is easy to see from the corollary to lemma 1.7 that these Riemann sums converge in the mean. From lemma 1.4 we see therefore

$$E(\hat{\alpha}_k) = \alpha_k \quad , \qquad E(\hat{\beta}_k) = \beta_k$$

and similarly from lemma 1.7

$$\sigma^2_{\hat{\alpha}_k} = \frac{1}{\pi^2 k^2} \lim_{\substack{\delta_i \to 0 \\ \delta_j \to 0}} \sum_i \sum_j \sin kt_i \sin kt_j \, \sigma\{(y_{t_i} - y_{t_{i-1}})(y_{t_j} - y_{t_{j-1}})\}$$

where $\delta_i = \max|t_i - t_{i-1}|$ and $\delta_j = \max|t_j - t_{j-1}|$.

Since the increments of y_t in non-overlapping intervals are independent of each other we have

$$\sigma^2_{\hat{\alpha}_k} = \frac{1}{\pi^2 k^2} \lim_{\delta_i \to 0} \sum_i c(t_i - t_{i-1}) \sin^2 kt_i = \frac{c}{\pi^2 k^2} \int_0^{2\pi} \sin^2 kt \, dt$$

so that

(3.8) $$\sigma^2_{\hat{\alpha}_k} = \frac{c}{\pi k^2}$$

and similarly

(3.9) $$\sigma^2_{\hat{\beta}_k} = \frac{c}{\pi k^2}$$

Further

(3.10) $$\sigma_{\hat{\alpha}_k \hat{\alpha}_\ell} = \sigma_{\hat{\beta}_k \hat{\beta}_\ell} = 0 \quad \text{for } k \neq \ell$$

$$\sigma_{\hat{\alpha}_k \hat{\beta}_\ell} = 0$$

From lemma 2.6 we see moreover that $\hat{\alpha}_k$ and $\hat{\beta}_k$ are normally distributed.

Suppose now that we take observations in the interval $(0, T)$ so that we have

(3.11) $$f(t) = \sum [\alpha_n \cos 2\pi n \tfrac{t}{T} + \beta_n \sin 2\pi n \tfrac{t}{T}]$$

We put $\tau = \frac{2\pi t}{T}$ and $y^*_\tau = y_t$. Then

$$\mathrm{Var}\left\{\frac{(y^*_{\tau+k} - y^*_\tau)}{\sqrt{k}}\right\} = c^* = \mathrm{Var}\left\{\frac{(y_{t+(kT/2\pi)} - y_t)}{\sqrt{k}}\right\} = \frac{Tc}{2\pi}$$

The maximum likelihood estimates $\hat{\alpha}_k$ and $\hat{\beta}_k$ then become

$$\hat{\alpha}_k = -\frac{1}{\pi k}\int_0^T \sin(2\pi k \tfrac{t}{T})dy_t$$

$$\hat{\beta}_k = \frac{1}{\pi k}\int_0^T \cos(2\pi k \tfrac{t}{T})dy_t$$

and

$$\sigma^2_{\hat{\alpha}_k} = \sigma^2_{\hat{\beta}_k} = \frac{Tc}{2\pi^2 k^2}, \quad \sigma_{\hat{\alpha}_k \hat{\beta}_\ell} = 0, \quad \sigma_{\hat{\alpha}_k \hat{\alpha}_\ell} = \sigma_{\hat{\beta}_k \hat{\beta}_\ell} = 0 \text{ for } k \neq \ell.$$

Thus a confidence region for the α_k, β_ℓ is given by

$$\chi^2 = \sum \frac{2\pi^2 k^2}{cT}\left[(\hat{\alpha}_k - \alpha_k)^2 + (\hat{\beta}_k - \beta_k)^2\right] \leq M,$$

where the sum runs over those terms $(\hat{\alpha}_k - \alpha_k)^2$, $(\hat{\beta}_k - \beta_k)^2$ which are not zero by assumption, and χ^2 has the χ^2-distribution with the number of degrees of freedom equal to the number of terms in the sums on the right. The estimates $\hat{\alpha}_k$, $\hat{\beta}_k$ are consistent in the following sense. Suppose f(t) is given by (3.11) and we observe y_t in the interval VT where V is an integer. We then have

$$f(t)=\sum_{n=1}^{m}\left[\alpha_n \cos 2\pi\frac{(nV)t}{VT} + \beta_n \sin 2\pi\frac{(nV)t}{VT}\right]$$

so that

$$\sigma^2_{\hat{\alpha}_k} = \sigma^2_{\hat{\beta}_k} = \frac{VTc}{2\pi^2V^2k^2} = \frac{Tc}{2\pi^2Vk^2}$$

and thus $\underset{V\to\infty}{\text{plim}}\ \hat{\alpha}_k = \alpha_k$. Similarly, $\underset{V\to\infty}{\text{plim}}\ \hat{\beta}_k = \beta_k$.

3. *Estimation of parameters for the O.U.P.*

We now turn to the discussion of the O.U.P. given by (2.25) and prove

Theorem 3.3 If x_t is an O.U.P. determined by the two parameters β and σ^2 and if the values of x_t are known in a dense set in any interval $0 \leq t \leq T$, then it is possible to determine $\sigma^2\beta$ with arbitrarily high precision.

Proof: We form with $N\tau = T$

$$(3.12) \qquad D = \frac{1}{N}\sum \frac{(x_{n\tau} - x_{\overline{n-1}\,\tau})^2}{\tau}$$

We have $E(D) = \frac{2\sigma^2}{\tau}(1-a_\tau) = 2\sigma^2\frac{(1-e^{-\beta\tau})}{\tau}$.

For $\tau \to 0$ this converges to $2\sigma^2\beta$.

We now compute the variance of D. For this purpose we shall need the value of $E(x_t x_{t'} x_{t''} x_{t'''})$ with $t \leq t' \leq t'' \leq t'''$. We have for $t \leq t' \leq t'' \leq t'''$ on replacing $x_{t'''}$ by $a_{t'''-t''} x_{t''} + \varepsilon_{t'',t'''-t''}$ and analogous transformations

$$E(x_t x_{t'} x_{t''} x_{t'''}) = a_{t'''-t''} E(x_t x_{t'} x^2_{t''})$$

$$E(x_t x_{t'} x^2_{t''}) = E\left[x_t x_{t'} (a_{t''-t'} x_{t'} + \varepsilon_{t',t''-t'})^2\right]$$

$$= a^2_{t''-t'} E(x_t x^3_{t'}) + a_{t'-t}(1 - a^2_{t''-t'})\sigma^4$$

and

$$E(x_t x^3_{t'}) = E\left[x_t (a_{t'-t} x_t + \varepsilon_{t,t'-t})^3\right]$$

$$= a^3_{t'-t} E(x^4_t) + 3a_{t'-t} E(x^2_t)\, \sigma^2 (1 - a^2_{t'-t})$$

$$= \sigma^4\left[3a^3_{t'-t} + 3a_{t'-t} - 3a^3_{t'-t}\right] = 3a_{t'-t}\sigma^4$$

Thus

$$E(x_t x_{t'} x^2_{t''}) = \sigma^4\left[3a^2_{t''-t'} a_{t'-t} + a_{t'-t}(1 - a^2_{t''-t'})\right]$$

$$= \sigma^4(a_{t'-t} + 2a^2_{t''-t'} a_{t'-t})$$

and

$$E(x_t x_{t'} x_{t''} x_{t'''}) = \sigma^4(a_{t'''-t''} a_{t'-t} + 2a_{t'''-t''} a^2_{t''-t'} a_{t'-t})$$

$$= \sigma^4(a_{t'''-t''} a_{t'-t} + 2a_{t'''-t'} a_{t''-t})$$

or

$$E(x_t x_{t'} x_{t''} x_{t'''}) = \sigma^4 \left\{ \exp\left[-\beta(t''' - t'' + t' - t)\right] + 2 \exp\left[-\beta(t''' + t'' - t' - t)\right] \right\} \tag{13}$$

For $n < m$ we find easily

$$E\left[(x_{n\tau} - x_{\overline{n-1}\,\tau})^2 (x_{m\tau} - x_{\overline{m-1}\,\tau})^2\right]$$

$$= E\left\{(x_{n\tau} - x_{\overline{n-1}\,\tau})^2 \left[(a_\tau - 1)x_{\overline{m-1}\,\tau} + \varepsilon_{\overline{m-1}\tau,\,\tau}\right]^2\right\}$$

$$= (a_\tau - 1)^2 \left[E(x^2_{n\tau}\, x^2_{\overline{m-1}\,\tau}) - 2E(x_{\overline{n-1}\,\tau}\, x_{n\tau}\, x^2_{\overline{m-1}\,\tau}) + E(x^2_{\overline{n-1}\tau} x^2_{\overline{m-1}\,\tau})\right.$$

$$\left. + 2\sigma^4(1 + a_\tau)\right].$$

From (3.13) we see then

$$E\left[(x_{n\tau} - x_{\overline{n-1}\,\tau})^2 (x_{m\tau} - x_{\overline{m-1}\,\tau})^2\right]$$

$$= (a_\tau - 1)^2\ \sigma^4 \left\{1 + 2e^{-2\beta(m-n-1)\tau} - 2e^{-\beta\tau} - 4e^{-\beta\tau[2(m-n)-1]}\right.$$

$$\left. + 1 + 2e^{-2\beta(m-n)\tau} + 2 + 2a_\tau\right\}.$$

Since $a_\tau = e^{-\beta\tau}$ we finally have for $m > n$

$$(3.14)\qquad E\left[(x_{n\tau} - x_{\overline{n-1}\,\tau})^2 (x_{m\tau} - x_{\overline{m-1}\,\tau})^2\right]$$

$$= 2\sigma^4 (1 - a_\tau)^2 \left[2 + (1 - a_\tau)^2\, e^{-2\beta(m-n-1)\tau}\right].$$

Further

$$E(x_{n\tau} - x_{\overline{n-1}\,\tau})^4 = E[(a_\tau - 1)x_{\overline{n-1}\,\tau} + \varepsilon_{\overline{n-1}\,\tau,\,\tau}]^4$$

$$= (a_\tau - 1)^4\, E(x^4_{\overline{n-1}\,\tau}) + 6(a_\tau - 1)^2\, E(x^2_{\overline{n-1}\,\tau})(1 - a_\tau^2)\sigma^2 + 3\sigma^4(1 - a_\tau^2)^2$$

$$= 3\sigma^4[(a_\tau - 1)^4 + 2(a_\tau - 1)^2(1 - a_\tau^2) + (1 - a_\tau^2)^2]$$

$$= 3\sigma^4(1 - a_\tau)^2\, [(1 - a_\tau)^2 + 2(1 - a_\tau^2) + (1 + a_\tau)^2],$$

that is

$$(3.15) \qquad E(x_{n\tau} - x_{\overline{n-1}\,\tau})^4 = 12\,\sigma^4(1 - a_\tau)^2.$$

From (3.12), (3.14), and (3.15) we have

$$E(D^2) = \frac{\sigma^4}{N^2\tau^2}(1 - a_\tau)^2\left[12N + 4N(N-1) + 4(1 - a_\tau)^2 \sum_{n=1}^{N-1}\sum_{m=n+1}^{N} e^{-2\beta(m-n-1)\tau}\right]$$

so that

$$\sigma_D^2 = \frac{4\sigma^4(1 - a_\tau)^2}{N^2\tau^2}\left[2N + (1 - a_\tau)^2 \sum_{n=1}^{N-1}\sum_{m=n+1}^{N} e^{-2\beta(m-n-1)\tau}\right].$$

We now compute

$$A = \sum_{n=1}^{N-1}\sum_{m=n+1}^{N}(a_\tau^2)^{m-n-1} = \sum_{n=1}^{N-1}\sum_{m=n}^{N-1}(a_\tau^2)^{m-n}.$$

By using the formula for the sum of a geometric series it is easily seen that

$$\sum_{n=1}^{N-1} \sum_{m=n}^{N-1} b^{m-n} = \frac{N}{1-b} - \frac{1-b^N}{(1-b)^2}$$

hence

$$A = \frac{N}{1-a_\tau^2} - \frac{1-a_\tau^{2N}}{(1-a_\tau^2)^2} \quad .$$

We substitute this in the expression for σ_D^2 and obtain

$$\sigma_D^2 = \frac{4\sigma^4(1-a_\tau)^2}{N^2\tau^2}\left[2N + \frac{N(1-a_\tau)}{1+a_\tau} - \frac{1-a_\tau^{2N}}{(1+a_\tau)^2}\right]$$

or

$$(3.16) \qquad \sigma_D^2 = \frac{4\sigma^4(1-a_\tau)^2}{N^2\tau^2}\left[2N + \frac{N-1-Na_\tau^2+a_\tau^{2N}}{(1+a_\tau)^2}\right] .$$

Equation (3.16) shows that σ_D^2 can be made arbitrarily small by making N large enough. In fact

$$\lim_{N\to\infty} \sigma_{\sqrt{N}D}^2 = 8\sigma^4\beta^2 \quad \text{and} \quad \lim_{N\to\infty} E(D) = 2\beta\sigma^2;$$

also

$$E(D-2\beta\sigma^2)^2 = \sigma_D^2 + 4\sigma^4\left(\frac{1-e^{-\beta\tau}}{\tau} - \beta\right)^2$$

and

$$E[\sqrt{N}(D - 2\beta\sigma^2)] = 2\sigma^2\sqrt{N}\left[\frac{1 - e^{-\beta\tau}}{\tau} - \beta\right]$$

and therefore since $\tau = T/N$

$$\lim_{N \to \infty} E[\sqrt{N}(D - 2\beta\sigma^2)] = 0$$

We proceed to prove that the limit distribution of $\sqrt{N}(D - 2\beta\sigma^2)$ is normal. This may be seen as follows:

$$\sqrt{N}\, D = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{(x_{n\tau} - x_{\overline{n-1}\,\tau})^2}{\tau} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \frac{[(a_\tau - 1)x_{n\tau} + \varepsilon_{n\tau,\tau}]^2}{\tau}$$

or

$$\text{(3.17)} \quad \sqrt{N}\, D = \frac{1}{\sqrt{N}} \frac{(a_\tau - 1)^2}{\tau} \sum_{0}^{N-1} x_{n\tau}^2 + \frac{2}{\sqrt{N}} \frac{(a_\tau - 1)}{\tau} \sum_{0}^{N-1} x_{n\tau}\varepsilon_{n\tau,\tau} + \frac{1}{\sqrt{N}} \sum_{0}^{N-1} \frac{\varepsilon_{n\tau,\tau}^2}{\tau}.$$

The last sum is a sum of independently distributed variables all with the same distribution and converges to the normal distribution by lemma 2.1. Thus the normality of the limit distribution will be proved if we can prove that the first two sums in (3.17) converge to zero.

We therefore put

(3.18) $\qquad \Sigma_1 = \frac{1}{\sqrt{N}} \frac{(a_\tau - 1)^2}{\tau} \sum_0^{N-1} x_{n\tau}^2$ and $\Sigma_2 = \frac{2}{\sqrt{N}} \frac{a_\tau - 1}{\tau} \sum_0^{N-1} x_{n\tau}\, \varepsilon_{n\tau,\tau}.$

We have by (3.13)

$$E(\Sigma_1^2) = \frac{\sigma^4}{N} \frac{(a_\tau - 1)^4}{\tau^2} \left[N^2 + 2N + 4 \sum_0^{N-2} \sum_{n+1}^{N-1} e^{-2\beta(m-n)\tau}\right].$$

The double sum in the bracket can be easily determined and we have

$$E(\Sigma_1^2) = \frac{\sigma^4 (a_\tau - 1)^4}{\tau^2} \left[N + \frac{2(1+a_\tau^2)}{1-a_\tau^2} - \frac{4a_\tau^2}{N} \frac{1-e^{-2\beta\tau N}}{(1-a_\tau^2)^2}\right].$$

Clearly $\lim\limits_{N\to\infty} E(\Sigma_1^2) = 0$.

Further

$$E(\Sigma_2^2) = \frac{4}{N} \frac{(a_\tau - 1)^2}{\tau^2} \sum_0^{N-1} E(x_{n\tau}^2\, \varepsilon_{n\tau,\tau}^2) = \frac{4(a_\tau - 1)^2}{\tau^2} \sigma^4 (1-a_\tau^2)$$

and therefore $\lim\limits_{N\to\infty} E(\Sigma_2^2) = 0$.

Therefore $\operatorname*{l.i.m.}\limits_{N\to\infty} \Sigma_1 = \operatorname*{l.i.m.}\limits_{N\to\infty} \Sigma_2 = 0$ so that

$$\operatorname*{l.i.m.}_{N\to\infty} \left(\sqrt{N}\, D - \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \frac{\varepsilon_{n\tau,\tau}^2}{\tau}\right) = 0,$$

and since the second term on the left is normally distributed with mean $2\beta\sigma^2\sqrt{N}$ it follows that $\sqrt{N}(D - 2\beta\sigma^2)$ is in the limit normally distributed with mean zero and variance $8\sigma^4\beta^2$.

To estimate σ^2 separately one might use the estimate

$$\hat{\sigma}^2 = \frac{1}{T}\int_0^T x_t^2 \, dt$$

Its variance is given by

$$\frac{1}{T^2}\int_0^T\int_0^T E(x_t^2 x_{t'}^2)\,dt\,dt' - \sigma^4$$

$$= \frac{2\sigma^4}{T^2}\left\{\int_0^T\int_0^t e^{-2\beta(t-t')}\,dt'dt + \int_0^T\int_t^T e^{-2\beta(t'-t)}\,dt'dt\right\}$$

$$= \frac{2\sigma^4}{\beta T} + \frac{\sigma^4(e^{-2\beta T} - 1)}{\beta^2 T^2}$$

If βT is large enough compared with σ^4 this comparatively simple estimate may be quite satisfactory.

CHAPTER 4

THE GENERAL DIFFERENTIAL PROCESS

We shall consider processes x_t with the following properties:

(1) x_t is a continuous process (not necessarily strongly continuous);

(2) Let $x_{t+\tau}-x_t = \varepsilon_{t,\tau}$. The random variables $\varepsilon_{t_1,\tau_1}, \varepsilon_{t_2,\tau_2}, \ldots, \varepsilon_{t_n,\tau_n}$ are completely independent of each other if the intervals $(t_1, t_1+\tau_1), (t_2, t_2+\tau_2), \ldots, (t_n, t_n+\tau_n)$ do not overlap;

(3) The distribution of $\varepsilon_{t,\tau}$ is independent of t.

Processes satisfying these three conditions will be called general differential processes. In this chapter we shall find a general expression giving the distributions of $x_{t+\tau}-x_t$ for all possible general differential processes. Processes satisfying equation (2.1) and assumptions 1, 2, 3, of chapter 2 are differential processes of second order and we discussed in chapter 2 the special case where $x_{t+\tau}-x_t$ is normally distributed.

We shall first discuss another special case in which the increments $x_{t+\tau}-x_t$ have a discontinuous distribution. A practical example for such a process is, for instance, the total of insurance claims raised against an insurance company as a result of randomly distributed accidents. An important special case, fundamental also for the understanding of the more general problem,

is that in which x_t increases a randomly distributed number of times within every time interval but each time by the same amount. We shall call such an increase a shot.

We make the following assumptions:

(1) The probability $p_\tau^{(k)}$ that k shots will occur during the interval $(t, t+\tau)$ is independent of t and of the number of shots that have occurred up to and including the time t.

(2) The probability $q_\tau^{(2)}$ that more than one shot will occur in a time interval of length τ is of smaller order than τ. In symbols,

$$q_\tau^{(2)} = o(\tau) \quad \text{or} \quad \lim_{\tau \to 0}(q_\tau^{(2)}/\tau) = 0.$$

(3) $p_\tau^{(k)}$ is a measurable function of τ.

We clearly have, if $\tau_1 + \tau_2 = \tau$

$$p_\tau^{(0)} = p_{\tau_1}^{(0)} \, p_{\tau_2}^{(0)} \quad . \tag{4.1}$$

From (4.1) and the measurability of $p_\tau^{(0)}$ it follows that $p_\tau^{(0)} = e^{\alpha\tau}$ and since $p_\tau^{(0)} \leq 1$ we have $\alpha \leq 0$. Moreover, if there are any shots to be expected we must have $\alpha < 0$, $\alpha = -\mu$ where $\mu > 0$. Thus

$$p_\tau^{(0)} = e^{-\mu\tau} \, , \; \mu > 0 \quad . \tag{4.2}$$

We now divide the interval $(t, t+\tau)$ into N parts. Then for sufficiently large N the probability that two shots will occur in any of the intervals can be made arbitrarily small so that if $p_\tau^{(k)}$ denotes the probability that k shots will occur during the time interval τ

$$(4.3) \qquad p_\tau^{(k)} = \binom{N}{k} [1 - \exp(-\frac{\mu\tau}{N})]^k \exp[-\frac{\mu\tau}{N}(N-k)] + o(1).$$

For $N \to \infty$ we then have

$$(4.4) \qquad p_\tau^{(k)} = \frac{e^{-\mu\tau}(\mu\tau)^k}{k!}$$

The distribution (4.4) is the Poisson distribution, its mean and variance are both equal to $\mu\tau$.

We next consider a situation in which the assumptions (1), (2), (3) regarding $p_\tau^{(k)}$ hold but where the increment of x_t at each shot varies and has itself a probability distribution $\phi(x) = \phi_1(x)$ and we shall also assume that the increases in different shots are independent of each other. If $\phi_k(x)$ is the distribution of the sum of k independent random variables, each with distribution $\phi(x)$, then the distribution of the total increase provided that k shots have occurred is $\phi_k(x)$. Thus the distribution of $x_{t+\tau} - x_t$ is given by

$$(4.5) \qquad P(x_{t+\tau} - x_t \leq A) = \sum_{k=0}^{\infty} \frac{e^{-\mu\tau}(\mu\tau)^k}{k!} \phi_k(A)$$

$$\text{with} \quad \phi_0(A) \begin{cases} 0 & \text{for } A < 0 \\ 1 & \text{for } A \geq 0 \end{cases}$$

In the following we shall need the characteristic function [abbreviated c.f.]

$$(4.6) \qquad f_\tau(s) = E\{\exp[is(x_{t+\tau} - x_t)]\}$$

of the distribution (4.5). An easy calculation gives, if $g(s)$ denotes the c.f. of $\phi(x)$

$$(4.7) \qquad f_\tau(s) = \sum_{k=0}^{\infty} \frac{e^{-(\mu\tau)}(\mu\tau)^k}{k!} [g(s)]^k = \exp\{\mu\tau[g(s) - 1]\}.$$

The distribution (4.5) is called the generalized Poisson distribution.

We return now to the general differential process. We have

$$\mathcal{E}_{t,\tau} = \mathcal{E}_{t,\frac{\tau}{n}} + \mathcal{E}_{t+\frac{\tau}{n},\frac{\tau}{n}} + \cdots + \mathcal{E}_{t+\frac{n-1}{n}\tau,\frac{\tau}{n}}$$

Hence, if $\phi_\tau(s)$ denotes the c.f. of $\mathcal{E}_{t,\tau}$ we have

$$(4.8) \qquad \phi_\tau(s) = [\phi_{\frac{\tau}{n}}(s)]^n.$$

From the continuity of x_t it follows that $\lim\limits_{\Delta\tau \to 0} \phi_{\tau+\Delta\tau}(s) = \phi_\tau(s)$.

Thus (4.8) implies

$$(4.9) \qquad \phi_\tau(s) = [\phi_1(s)]^\tau.$$

On the other hand every family of distribution functions $H_\tau(x)$ whose characteristic functions satisfy equation (4.9) is the distribution function of the increment of some differential process. Hence the general form of a differential process will be found if we find the general form of characteristic functions $\phi(s)$ that satisfy the condition that for every $\tau \geq 0$, $[\phi(s)]^\tau$ is a c.f. A distribution law whose c.f. satisfies this condition is called an infinitely divisible law (abbreviated, i.d.l.).

Our main result will be the following:

<u>Theorem 4.1.</u>[12] Let $\psi(s) = \log \phi(s)$. The function $\phi(s)$ is c.f. of an infinitely divisible law if and only if

$$(4.10) \qquad \psi(s) = isa + \int_{-\infty}^{+\infty} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right)\frac{1+x^2}{x^2}\, dG(x)$$

where a is real and $G(x)$ non-decreasing and bounded and the integrand is defined by continuity to be $-\frac{s^2}{2}$ for $x = 0$.

Fundamental for the proof of this theorem is the powerful continuity theorem of P. Lévy.

<u>Continuity Theorem:</u> Let $\{F_n(x)\}$ be a sequence of distribution functions, $\{f_n(s)\}$ the corresponding sequence of c.f.'s. The sequence $\{F_n(x)\}$ converges to a distribution function $F(x)$ if and only if $f_n(s)$ converges to a function $f(s)$ continuous for $s = 0$.

---

[12] Theorem 4.1 is due to P. Lévy (see, for instance, his "Theorie de l'addition des variables aléatoires", Gauthiers Villars Paris, 1937, p. 180). The following elegant proof is due to M. Loève, University of California Publ. in Stat., vol. 1, No. 5, 53-88 (1950).

For a proof the reader is referred to H. Cramér, Mathematical Methods of Statistics, 10.4.

We shall need this theorem in the following, slightly more general, form.

<u>Corollary to the Continuity Theorem</u>: Let $\{F_n(x)\}$ be a sequence of bounded monotone functions, $F_n(-\infty) = 0$ and let $\{f_n(s)\}$ be the sequence of their Fourier transforms

$$(4.11) \qquad f_n(s) = \int_{-\infty}^{+\infty} e^{ixs}\, dF_n(x) .$$

The sequence $\{F_n(x)\}$ converges to a bounded monotonic function $F(x)$ and $\lim_{n\to\infty} [F_n(\infty) - F_n(-\infty)] = F(\infty) - F(-\infty)$ if and only if the sequence $\{f_n(s)\}$ converges to a function $f(s)$ continuous at $s = 0$.

The corollary follows easily from the continuity theorem in observing that

$$f_n(0) = \int_{-\infty}^{+\infty} dF_n(x) = V_n .$$

Hence the variations V_n of $F_n(x)$ converge to $f(0)$. If $f(0) \neq 0$ then $\frac{F_n(x)}{V_n}$ converges by the continuity theorem to a distribution function $H(x)$ and $F_n(x)$ converges to $F(x) = H(x) f(0)$. If $f(0) = 0$ then $\lim_{n\to\infty} F_n(x) = 0$ and the theorem also holds.

We shall also need the

Helly-Bray theorem: [redacted] Let $\{F_n(x)\}$ be a sequence of distribution functions and $\lim\limits_{n\to\infty} F_n(x) = F(x)$. Let further $g(x)$ be everywhere continuous and assume that to every $\varepsilon > 0$ there exists an A such that $\int\limits_{|x| \geq A} |g(x)| \, dF_n(x) < \varepsilon$. Then

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} g(x) \, dF_n(x) = \int_{-\infty}^{+\infty} g(x) \, dF(x) .$$

For a proof of this theorem the reader is referred to H. Cramér, op. cit., p. 74.

It is easy to verify that the conditions of the Helly-Bray theorem are satisfied if

$$\lim_{n\to\infty} [F_n(\infty) - F_n(-\infty)] = F(\infty) - F(-\infty)$$

and if $g(x)$ is bounded.

We shall first show that every function $\psi(s)$ given by (4.10) is the logarithm of the c.f. of an i.d.l. For this it will be sufficient to show that every $\psi(s)$ given by (4.10) is the logarithm of a c.f. since it is obvious that with $\psi(s)$ also $\frac{\psi(s)}{n}$ satisfies (4.10).

---

[13] The theorem still holds if the functions $F_n(x)$ are uniformly bounded monotonic functions and if one or both of the limits of integration is infinite.

To see this, consider first

$$(4.12) \qquad I_\varepsilon(s) = \int\limits_{\varepsilon<|x|<\frac{1}{\varepsilon}} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right)\frac{1+x^2}{x^2}\, dG(x)$$

with $0 < \varepsilon < 1$.

$I_\varepsilon(s)$ may be written as the limit of Riemann-Stieltjes sums

$$(4.13) \qquad S_k = \sum_k \left[\lambda_k(e^{isx_k} - 1) + is\mu_k\right]$$

with

$$\lambda_k = \frac{1+x_k^2}{x_k^2}\,[G(x_{k+1}) - G(x_k)]$$

$$\mu_k = \frac{1}{x_k}[G(x_{k+1}) - G(x_k)].$$

It is easy to verify that e^{isu} is the c.f. of the distribution of a random variable which equals u with probability one. Hence on account of (4.7) we see that

$$\lambda_k(e^{isx_k} - 1) + is\mu_k$$

is logarithm of a c.f. and so consequently is S_k. Thus $I_\varepsilon(s)$ is a limit of logarithms of c.f.'s. Also $I_\varepsilon(s)$ is obviously continuous at $s=0$. By Lévy's continuity theorem $I_\varepsilon(s)$ is thus the logarithm of the c.f. of a distribution function. By the same theorem also

$$I_0(s) = \lim_{\varepsilon \to 0} I_\varepsilon(s) = \int_{|x|>0} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right) \frac{1+x^2}{x^2}\, dG(x)$$

is the logarithm of a c.f.

But $\psi(s)$

$$= isa + \int_{-\infty}^{+\infty} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right) \frac{1+x^2}{x^2}\, dG(x) = I_0(s) + isa - \frac{s^2}{2}[G(0+) - G(0-)]$$

is obtained by adding to $I_0(s) + isa$ the logarithm of the c.f. of a normal distribution. Thus $\psi(s)$ is the logarithm of a c.f. Thus the equation (4.10) is sufficient for $\psi(s)$ to be the logarithm of a c.f. of an i.d.l.

To prove also the necessity of (4.10) we need several lemmas.

<u>Lemma 4.1.</u> $G(x)$ and a in (4.10) are uniquely determined by $\psi(s)$.

We consider

$$(4.14) \quad e(s) = \int_0^1 \left[\psi(s) - \frac{\psi(s+h) - \psi(s-h)}{2}\right] dh =$$

$$\int_{-\infty}^{+\infty} e^{isx}\left(1 - \frac{\sin x}{x}\right) \frac{1+x^2}{x^2}\, dG(x) = \int_{-\infty}^{+\infty} e^{isx}\, d\phi(x)$$

where

$$(4.15) \qquad \phi(x) = \int_{-\infty}^{x} \left(1 - \frac{\sin y}{y}\right) \frac{1+y^2}{y^2}\, dG(y)$$

It is easy to verify that $(1-\frac{\sin y}{y})\frac{1+y^2}{y^2}$ is bounded above and below by positive constants. Thus $\phi(x)$ is monotone and of bounded variation. The Fourier inversion formula determines uniquely $\phi(x)$ given $e(s)$ and thus also

$$(4.16) \qquad G(x) = \int_{-\infty}^{x} \frac{y^2}{1+y^2}(1-\frac{\sin y}{y})^{-1}\, d\phi(y)$$

Finally a is determined from (4.10).

Lemma 4.2. If $\psi_n(s)$ determined by $G_n(x)$ and by a_n converges uniformly in every finite interval to a function b(s) continuous at the origin then $\lim_{n\to\infty} G_n(x) = G(x)$ and $\lim_{n\to\infty} a_n = a$ exist and $b(s) = \psi(s)$ is determined by a and G(x) by equation (4.10).

Proof: From P. Lévy's continuity theorem it follows that $e^{b(s)}$ is a c.f. hence everywhere continuous. Thus also b(s), its logarithm. Therefore, $e_n(s)$, defined by $\psi_n(s)$ by means of (4.14) converges to a continuous function. By the corollary to the Continuity Theorem it then follows that $\lim_{n\to\infty} \phi_n(x) = \phi(x)$ exists. Moreover $\phi(x)$ is non-decreasing and bounded. It follows from the Helly-Bray theorem that

$$\lim_{n\to\infty} G_n(x) = \lim_{n\to\infty}\int_{-\infty}^{x}(1-\frac{\sin y}{y})^{-1}\frac{y^2}{1+y^2}\, d\phi_n(y)$$

$$= \int_{-\infty}^{x}(1-\frac{\sin y}{y})^{-1}\frac{y^2}{1+y^2}\, d\phi(y) = G(x)$$

since the integrand is bounded. It further follows, also from the Helly-Bray theorem, that

$$\lim_{n\to\infty} I_n(s) = \lim_{n\to\infty} \int_{-\infty}^{+\infty} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right) \frac{1+x^2}{x^2}\, dG_n(x)$$

$$= \int_{-\infty}^{+\infty} \left(e^{isx} - 1 - \frac{isx}{1+x^2}\right) \frac{1+x^2}{x^2}\, dG(x) = I(s) .$$

Finally it follows from the convergence of $\psi_n(s)$ and $I_n(s)$ that also the sequence $\{a_n\}$ must converge and thus $b(s) = \psi(s)$.

The converse of lemma 4.2 follows immediately from the Helly-Bray theorem.

<u>Lemma 4.3.</u> The c.f. $\phi(s)$ of an i.d.l. is everywhere different from zero.

Consider $|\phi(s)|^{(1/n)}$ We have $\lim_{n\to\infty}[\phi(s)]^{(1/n)} = w(s)$ where $w(s) = 1$ for $\phi(s) \neq 0$ and $w(s) = 0$ for $\phi(s) = 0$. By Lévy's theorem $w(s)$ is a c.f. Since $w(s) = 1$ for $s = 0$ and $w(s)$ is continuous being a c.f. we must have $w(s) = 1$ everywhere. Thus $\phi(s) \neq 0$ everywhere,

<u>Lemma 4.4:</u> $\log x = \lim_{n\to\infty} n\{x^{(1/n)} - 1\}$, $x > 0$. Lemma 4.4 follows immediately from the rule of de l'Hospital.

<u>Lemma 4.5.</u> If $\phi(s)$ is the c.f. of an i.d.l, then there exists a sequence of functions $\psi_n(x)$ given by (4.10) such that

$$\log \phi(s) = \lim_{n\to\infty} \psi_n(s) .$$

Proof: We have by lemma 4.4

$$\log \phi(s) = \lim_{n\to\infty} n\{[\phi(s)]^{(1/n)} - 1\} = \lim_{n\to\infty} \psi_n(s)$$

uniformly in every finite interval of s since $\phi(s) \neq 0$ with

$$a_n = n \int_{-\infty}^{+\infty} \frac{y}{1+y^2}\, dF_n(y), \qquad G_n(x) = n \int_{-\infty}^{x} \frac{y^2}{1+y^2}\, dF_n(y) .$$

Here $F_n(x)$ is the d.f. whose c.f. is $[\phi(s)]^{(1/n)}$.

Proof of the necessity of (4.10): By lemma 4.3 $\phi(s) \neq 0$, hence $\log \phi(s)$ is defined everywhere and continuous. Moreover $\log \phi(s) = \lim_{n\to\infty} \psi_n(s)$ uniformly in s with $\psi_n(s)$ given by (4.10). But by lemma 4.2 $\lim_{n\to\infty} \psi_n(s) = \psi(s)$ where $\psi(s)$ is itself determined by (4.10). Thus theorem 4.1 is completely proved.

From our proof of the sufficiency of equation (4.10) follows the following corollary to theorem 4.1.

Corollary to theorem 4.1: If x is distributed according to an i.d.l. then $x = y + z$ where y is normally distributed and z is distributed as is the limit of a sequence of finite sums of independent random variables each of which is distributed according to (4.5) with

$$\Phi_1(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq a \end{cases}$$

Theorem 4.2. Let x_t be a differential process of second order, then $E(x_{t+\tau}-x_t) = \tau m$ and $Var(x_{t+\tau}-x_t) = \tau\sigma^2$ where m and σ^2 are constants independent of t and τ.

Proof: Let $\psi_\tau(s)$ be the logarithm of the c.f. of $x_{t+\tau}-x_t$. From (4.9) we see then that $\psi_\tau(s) = \tau\psi_1(s)$ where $\psi_1(s)$ is determined by (4.10). Therefore $\psi_\tau'(0) = \tau\psi_1'(0)$ and $\psi_\tau''(0) = \tau\psi_1''(0)$. From this and from $\psi_\tau'(0) = iE(x_{t+\tau}-x_t)$ and $\psi_\tau''(0) = -Var(x_{t+\tau}-x_t)$ we see that the ~~lemma~~ theorem holds.

The estimation procedures in the case of a process given by (4.4) are very simple. The parameter to be estimated is μ, its maximum likelihood estimate is x_T/T, the number of shots observed per unit of time, and the variance of this estimate is μ/T.

If the process is given by (4.5) then the increments observed are a sample from a population with the distribution $\phi(x)$. If $\phi(x)$ is given in parametric form then the proper estimation procedures are those appropriate for estimating the parameters of $\phi(x)$ from the observed sample of increments.

CHAPTER 5

DIFFERENTIAL PROCESSES MODIFIED BY MECHANICAL DEVICES

1. Filter effect.

In registering a stochastic process the registering device often spreads the effect over a certain period of time in such a way that an increase occurring at time $t = 0$ will produce an effect at time t and the observed value, or the output process, is obtained from the superposition of all the effects produced from increases that occurred in preceding time periods.

If we let y_t be the output process and x_t the input process and assume that the effect is proportional to the increase we then have

(5.1) $$y_t = \int_{-\infty}^{+\infty} f(t-\tau)\, dx_\tau$$

where $\int_{-\infty}^{+\infty} |f(t)|dt$ and $\int_{-\infty}^{+\infty} [f(t)]^2\, dt$ and the integral (5.1) are assumed to exist.

Taking the upper limit of the integral equal to infinity instead of zero leaves the possibility open that future changes will influence the present. If the present is not affected by the future, then $f(t)$ will be 0 for negative values of t.

In previous work we have often taken the point of view that the process x_t starts at some fixed time T. However we can also speak of the conditional distribution of x_τ given x_t for $\tau < t$ and thus $\int_A^B f(t-\tau)\, dx_\tau$ may be formed for every A and B and we proceed to prove

Theorem 5.1. If x_t is a differential process of second order with a weight function $f(t)$ then the integral $\int_{-\infty}^{+\infty} f(t-\tau)\, dx_\tau$ exists l.i.m.

In the following we use a simplified notation by writing x_i for x_{τ_i} and Δ_i for $\tau_i - \tau_{i-1}$.

Proof of theorem 5.1: We have

$$E[\int_A^B f(t-\tau)\, dx_\tau]^2 = \lim_{\Delta_i \to 0} E\{[\sum_i f(t-\tau_i^*)(x_i - x_{i-1})]^2\}$$

$$= \lim_{\Delta_i \to 0} E\{\sum_i f^2(t-\tau_i^*)(x_i - x_{i-1})^2$$

$$+ \sum_{i \neq j} f(t-\tau_i^*) f(t-\tau_j^*)(x_i - x_{i-1})(x_j - x_{j-1})\}$$

where $\tau_{i-1} \leq \tau_i^* \leq \tau_i$.

Since x_τ is a differential process of second order we have, writing σ_{ij} for the covariance of $(x_i - x_{i-1})$ and $(x_j - x_{j-1})$

$$\sigma_{ij} = \begin{cases} 0 & \text{if} \quad i \neq j \\ \sigma^2(\tau_i - \tau_{i-1}) & \text{if} \quad i = j \end{cases}$$

$$E(x_i - x_{i-1}) = m(\tau_i - \tau_{i-1}).$$

This follows from theorem 4.2.

Thus

$$(5.2)\quad E\Big[\int_A^B f(t-\tau)dx_\tau\Big]^2 = \sigma^2\int_A^B f^2(t-\tau)d\tau + m^2\Big[\int_A^B f(t-\tau)d\tau\Big]^2$$

$$= \sigma^2\int_{t-B}^{t-A} f^2(\tau)d\tau + m^2\Big[\int_{t-A}^{t-B} f(\tau)d\tau\Big]^2 .$$

Both integrals on the right of (5.2) converge to zero if A and B converge both to $+\infty$ or to $-\infty$. In fact, if $m = 0$, only the convergence of $\int_{-\infty}^{\infty}[f(t)]^2dt$ need be assumed; thus by lemma 1.6

$$\underset{\substack{A\to-\infty\\ B\to+\infty}}{\text{l.i.m.}}\int_A^B f(t-\tau)dx_\tau \quad \text{exists.}$$

We denote the characteristic function of the increment $x_{t+\tau} - x_t$ of a differential process by $\phi_\tau(s)$. From (4.9) we see that

$$(5.3)\qquad \phi_\tau(s) = \exp[\tau\log\phi_1(s)]$$

We compute next the characteristic function $\eta_t(s)$ of y_t, that is $\eta_t(s) = E(e^{isy_t})$. We have

$$(5.4)\qquad \eta_t(s) = E\{\exp[is\int_{-\infty}^{+\infty}f(t-\tau)dx_\tau]\}$$

$$= E\{\exp[is\ \underset{\Delta_j\to 0}{\text{plim}}\ \sum_j f(t-\tau_j^*)(x_j - x_{j-1})\}$$

The random variable in the exponent at the right of (5.4) converges in probability to $y_t = \int_{-\infty}^{+\infty} f(t-\tau)dx_\tau$. Hence its characteristic function converges to $\eta_t(s)$. We thus have

$$(5.5) \qquad \eta_t(s) = \lim_{\Delta_j \to 0} E\{\exp[is \sum_j f(t-\tau_j^*)(x_j - x_{j-1})]\}$$

Since the summands in the exponent are independent random variables we have

$$\eta_t(s) = \lim_{\Delta_j \to 0} \prod_j E\{\exp[isf(t-\tau_j^*)(x_j - x_{j-1})]\}$$

The characteristic function of $x_{t+\tau} - x_t$ is $\phi_\tau(s)$, therefore the characteristic function of $f(t-\tau_j^*)(x_j - x_{j-1})$ is

$\phi_{\Delta_j}[sf(t-\tau_j^*)] = \exp\{\Delta_j \log \phi_1[sf(t-\tau_j^*)]\}$ so that

$$(5.6) \qquad \log \eta_t(s) = \lim_{\Delta_j \to 0} \sum \Delta_j \log \phi_1[sf(t-\tau_j^*)]$$

$$= \int_{-\infty}^{+\infty} \log \phi_1[sf(t-\tau)]d\tau = \int_{-\infty}^{+\infty} \log \phi_1[sf(t)]dt .$$

The characteristic function of the joint distribution of $y_{t_1}, \ldots, y_{t_n}$, which completely determines the output process is formed in an analogous manner. We have

$$\eta_{t_1,\ldots,t_n}(s_1,\ldots,s_n) = E\{\exp i(s_1 y_{t_1} + \ldots + s_n y_{t_n})\}$$

$$= E\{\exp i \int_{-\infty}^{+\infty}[s_1 f(t_1-\tau)+\ldots+s_n f(t_n-\tau)]\}dx_\tau.$$

The argument employed in the case $n=1$ shows that

$$\text{(5.7)} \qquad \log \eta_{t_1,\ldots,t_n}(s_1,\ldots,s_n)$$

$$= \int_{-\infty}^{+\infty} \log\{\phi_1[s_1 f(t_1-\tau)+\ldots+ s_n f(t_n-\tau)]\}d\tau$$

It is seen from (5.7) that $\eta_{t_1,\ldots,t_n}(s_1,\ldots,s_n)$ and thus also the joint distribution of $y_{t_1},\ldots, y_{t_n}$ is invariant under translations in time. Hence the process is stationary.[14]

If the distribution of x_t is Gaussian (a differential process which is Gaussian is a F.R.P.) then also the distribution of y_t is Gaussian. The variance of y_t is given by $\sigma^2 \int_{\infty}^{+\infty}[f(t)]^2 dt$ and the covariance function is given by

$$\text{(5.8)} \quad R(t-t') = E\{[\int_{-\infty}^{+\infty} f(t-\tau)dx_\tau][\int_{-\infty}^{+\infty} f(t'-\tau)dx_\tau]\}$$

$$= \sigma^2 \int_{-\infty}^{+\infty} f(t-\tau)f(t'-\tau)d\tau = \sigma^2 \int_{-\infty}^{+\infty} f(\tau)f(t'-t+\tau)d\tau$$

- -

[14] A process is called stationary if the variables $x_{t_1},\ldots,x_{t_n}$ have the same distribution as the variables $x_{t_1+h},\ldots,x_{t_n+h}$.

This follows immediately by writing the integrals as limits of Riemann-Stieltjes sums and by then applying theorem 5.1 and lemma 1.4. Thus the resulting process is a stationary [redacted] Gaussian process with covariance function (5.8). If we put, for instance,

$$f(t) = \begin{cases} e^{-\beta t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

then we obtain the O.U.P. of Chapter 2 with $E(y_t^2) = \frac{\sigma^2}{2\beta}$

It may be seen from (5.7) and (5.8) that a large variety of output processes may be obtained from differential processes. If the differential process can be specified in parametric form and the function f(t) is also known at least in parametric form then (5.7) or in the most important special case (5.8) will give the output process in parametric form, so that the procedures of testing hypotheses about a finite number of parameters and of estimation in the parametric case become applicable although the difficulties of calculation may still be formidable.

In case nothing is known about either f(t) or $\phi_1(s)$ the only way known at present by which some inferences can be obtained is by the spectral analysis decribed in Chapter 6.

---

[14] A process is called stationary if the variables $x_{t_1}, \ldots, x_{t_n}$ have the same distribution as the variables $x_{t_1+h}, \ldots, x_{t_n+h}$.

2. Counter data

The modifying device may also operate in such a way that the modification of the input process is itself dependent on previous values of the input or output process. A frequently occurring example of this type of modification is provided by certain counter devices, which count random events. Due to the inertia of the counter device not all events will be counted. In particular we shall consider two types of such devices.

Type 1. After an event has been registered the counter remains locked during a certain time τ.

Type 2. After an event has happened the counter remains locked during a certain time τ.

A general and comprehensive treatment of probability problems in counter devices has been given by W. Feller (Courant Anniversary volume, 1948, pp. 105-115) and we shall here follow essentially Feller's representation. We shall assume that the input process is a Poisson process described by (4.4).

Let T_i, $i \geq 1$ be the time interval between the i-th and the (i+1)st registration, T_0 the time from the beginning, when the counter is locked, to the first registration. The T_i, $i \geq 1$ are independently distributed all with the same distribution. We denote the time up to the (k+1)st registration by

$$S_k = T_0 + T_1 + \ldots + T_k \tag{5.9}$$

Let N be the number of registrations during time t. We clearly have

$$p_k(t) = P(N = k) = P(S_{k-1} \leq t) - P(S_k \leq t)$$

Let T_k, $k \geq 1$ have the distribution function F, so that $P(T_k \leq t) = F(t)$. We write moreover $F_0(t)$ for the distribution function of T_0. Let $F_k(t) = P(S_k \leq t)$ then

$$p_k(t) = F_{k-1}(t) - F_k(t). \tag{5.10}$$

Since $S_{k+1} = S_k + T_{k+1}$ and since S_k and T_{k+1} are independent we have

$$F_{n+1}(t) = \int_0^t F_n(t-x)\, dF(x).$$

The characteristic function $\phi_t(s)$ of the random variable N is thus given by

$$p_0(t) + \sum_{k=1}^{\infty} e^{isk}[F_{k-1}(t) - F_k(t)] \tag{5.11}$$

$$= p_0(t) + e^{is}F_0(t) + \sum_{k=1}^{\infty} e^{is(k+1)}F_k(t) - \sum_{k=1}^{\infty} e^{isk}F_k(t).$$

Thus since $p_0(t) + F_0(t) = 1$, we obtain

$$\phi_t(s) = 1 + (e^{is} - 1)\sum_{k=0}^{\infty} e^{isk}F_k(t). \tag{5.12}$$

Hence

$$(5.13)\qquad \psi_t(s) = \frac{\phi_t(s)-1}{(e^{is}-1)} = \sum_{k=0}^{\infty} e^{isk}F_k(t)$$

$$= F_0(t) + \sum_{k=1}^{\infty}\int_0^t e^{isk}F_{k-1}(t-x)\,dF(x)$$

$$= F_0(t) + e^{is}\int_0^t \psi_{t-x}(s)\,dF(x).$$

In type 1 as well as in type 2 counters the value of N is bounded so that $\sum_{k=0}^{\infty}\frac{V_k}{k!}$ converges, where $V_k = V_k(t)$ is the k-th moment. Thus we may write

$$\phi_t(s) = \sum_{k=0}^{\infty}\frac{V_k}{k!}(is)^k$$

and for $s < 1$ also

$$\psi_t(s) = \left[\sum_{1}^{\infty}\frac{V_k}{k!}(is)^k\right]\left[is + \frac{(is)^2}{2!} + \dots\right]^{-1}$$

$$= \left(V_1 + \frac{isV_2}{2!} + \dots\right)\left(1 - \frac{is}{2} + \dots\right).$$

Hence the constant term in the expansion of $\psi_t(s)$ becomes V_1 and the coefficient of (is) becomes $(V_2 - V_1)/2$. From this and (5.13) we obtain equations for V_1 and V_2

$$(5.14)\begin{cases} V_1(t) = F_0(t) + \int_0^t V_1(t-x)dF(x) = E(N) \\ \\ V_2(t) = 2V_1(t) - F_0(t) + \int_0^t V_2(t-x)dF(x) = E(N^2) \end{cases}$$

We begin with the discussion of counters of type 1. From (4.2) we see that $F_o(t) = 1 - e^{-at}$ where $a > 0$ is the mean number of events per unit of time.

Further

$$F(t) = 1 - e^{-a(t-\tau)} \quad \text{for } t \geq \tau,$$

$$F(t) = 0 \quad \text{for } t < \tau$$

since the counter remains locked during the time τ after every registration. The first of the equations (5.14) then becomes

$$(5.15) \quad V_1(t) = \begin{cases} 1 - e^{-at} & \text{for } t \leq \tau \\ \\ 1 - e^{-at} + a\int_\tau^t V_1(t-x)e^{-a(x-\tau)}dx & \text{for } t > \tau. \end{cases}$$

We compare (5.15) with the more general equation

$$A(t) = \begin{cases} H(t) & \text{for } t \leq \tau \\ \\ H(t) + a\int_\tau^t A(t-x)e^{-a(x-\tau)}\,dx & \text{for } t > \tau. \end{cases}$$

If $H(t) \leq 1 - e^{-at}$ then $A(t) \leq V_1(t)$. If $H(t) > 1 - e^{-at}$ then $A(t) > V_1(t)$. This is certainly true for $0 \leq t \leq \tau$ and can easily be shown to hold for $t \leq (n+1)\tau$ if it holds for $t \leq n\tau$.

We put $A(t) = \frac{at}{1+a\tau} + c$. Then

$$H(t) = \begin{cases} \frac{at}{1+a\tau} + c & \text{for } t \leq \tau \\ 1 - \frac{e^{-a(t-\tau)}}{1+a\tau} + ce^{-a(t-\tau)} & \text{for } t \geq \tau \end{cases}$$

An elementary calculation shows that

$$H(t) \leq 1 - e^{-at} \quad \text{if } c = 0$$

$$H(t) > 1 - e^{-at} \quad \text{if } c = \frac{a^2\tau^2}{2(1+a\tau)}$$

Hence

$$(5.16) \qquad \frac{at}{1+a\tau} \leq V_1(t) \leq \frac{at}{1+a\tau} + \frac{a^2\tau^2}{2(1+a\tau)}$$

It is also possible to obtain from (5.14) an exact expression for $V_1(t)$. However, this does not seem to be of great interest since (5.16) shows that $\frac{at}{1+a\tau}$ approximates $V_1(t)$ very closely with a bounded error which is small compared to $V_1(t)$ unless $a\tau$ is very large. The exact expression for $V_1(t)$ is moreover very involved and hard to evaluate.

For the variance $B(t)$ of N given by $B(t) = V_2(t) - [V_1(t)]^2$, Feller found the asymptotic expression

$$(5.17) \qquad B(t) = \frac{at}{(1+a\tau)^3} + o(t) .$$

We now put

$$(5.18)\quad \begin{cases} f(s) = \int_0^\infty e^{-st}\, dF(t)\,; \quad f_k(s) = \int_0^\infty e^{-st}\, dF_k(t) \\ \mu(s) = \int_0^\infty V_1(t) e^{-st}\, dt \end{cases}$$

We have by (5.10)

$$V_1(t) = \sum_{k=1}^{\infty} k p_k(t) = \sum_{k=1}^{\infty} k[F_{k-1}(t) - F_k(t)] = \sum_{k=0}^{\infty} F_k(t)\,,$$

and by induction, using the well-known multiplicative property of the Laplace transform,

$$f_k(s) = f_0(s)[f(s)]^k$$

where

$$f_0(s) = \int_0^\infty e^{-st}\, dF_0(t)\,.$$

Thus

$$(5.19)\quad \mu(s) = \sum_0^\infty \int_0^\infty e^{-st} F_k(t) dt = \sum_0^\infty \frac{1}{s} \int_0^\infty e^{-st} dF_k(t) = \frac{1}{s} \frac{f_0(s)}{1-f(s)}.$$

We now proceed to discuss counters of type 2. The distribution function $F(t)$ of T_k must first be obtained. To this purpose we shall first obtain the distribution of the time T during which the counter is locked. The probability that once the counter is locked exactly ν events will prolong the locked time T is given by $q^\nu p$, where $p = e^{-a\tau}$, $q = 1 - e^{-a\tau}$, since p by (4.2) is the probability that no event will occur during time τ and q^ν the probability

that the time intervals between ν successive events will all be smaller than τ. Let now $T^{(i)}$ be the time elapsed between the $(i-1)$-st and the i-th event. The total locked time T, provided exactly ν events prolong the locked time, is then given by

$$T = T^{(1)} + T^{(2)} + \dots + T^{(\nu)} + \tau . \tag{5.20}$$

The conditional probability $U(t)$ that an event will occur during time t provided that $t \leq \tau$ is then given by

$$U(t) = \frac{1}{q}(1 - e^{-at}) . \tag{5.21}$$

Thus

$$u(s) = \int_0^{\tau} e^{-st}\, dU(t) = \frac{1}{q}\frac{a}{a+s}[1 - e^{-(a+s)\tau}] . \tag{5.22}$$

Let now ν events prolong the locked time and write $W_\nu(t) = P(T^{(1)} + \dots + T^{(\nu)} < t \mid \nu)$ for the probability that the locked time will be at most $t + \tau$, provided ν events prolong the locked time. From (5.22) and (5.18) we see that

$$\int_0^{\nu\tau} e^{-st}\, dW_\nu(t) = [u(s)]^\nu .$$

Consider now ν itself as a random variable. Then $W(t) = P(T^{(1)} + \dots T^{(\nu)} < t)$ is the probability that the locked time T is at most $t + \tau$. We thus have

$$\int_0^\infty e^{-st}\, dW(t) = p\sum[qu(s)]^\nu = \frac{p}{1-qu(s)} = p\{1-\frac{a}{a+s}[1-e^{-(a+s)\tau}]\}^{-1}$$

$$= \frac{(a+s)e^{-a\tau}}{s+ae^{-(a+s)\tau}} .$$

Let now $G(t) = P(T \leq t)$ then $G(t) = W(t-\tau)$ for $t \geq \tau$ and $G(t) = 0$ for $t < \tau$. Hence

$$(5.23) \quad \int_0^\infty e^{-st}\, dG(t) = \int_\tau^\infty e^{-st}\, dW(t-\tau) = e^{-s\tau}\int_0^\infty e^{-st}\, dW(t)$$

$$= \frac{(a+s)e^{-(a+s)\tau}}{s+ae^{-(a+s)\tau}} .$$

The time between two successive registrations is composed of the resolving time T and the time from the moment when the counter is free to the next event. The distribution of the latter is by (4.2) $1-e^{-at}$ and has the Laplace transform $a/(a+s)$ thus

$$f(s) = \int_0^\infty e^{-st}\, dF(t) = \frac{a\exp[-(a+s)\tau]}{s+a\exp[-(a+s)\tau]}$$

while $f_0(s) = \frac{a}{a+s}$. Substituting this into (5.19) yields

$$(5.24) \qquad \mu(s) = \frac{a[s+a\, e^{-(a+s)\tau}]}{s^2(a+s)} .$$

This is the Laplace transform (5.18) of the function

$$(5.25)\qquad V_1(t)=\begin{cases} 1-e^{-at} & \text{for } 0 \leq t \leq \tau \\ 1-e^{-a\tau}+(t-\tau)ae^{-a\tau} & \text{for } t \geq \tau \end{cases}$$

Since $V_1(t)$ is completely determined by its Laplace transform, formula (5.25) gives the expected number of registrations during time t in a counter of type 2.

A calculation similar to the one leading to (5.25) shows that the variance $B(t)$ of the number of counted events is given by

$$(5.26)\qquad B(t) = V_2(t) - [V_1(t)]^2$$

$$= ae^{-a\tau}(t-\tau)[1-2a\tau e^{-a\tau}] - e^{-a\tau} + (1+a\tau)^2 e^{-2a\tau}.$$

CHAPTER 6

The Fourier Analysis of Stochastic Processes.

1. General theory.

A function $f(t,t')$ in two variables is called monotonoid if $f(t,t') = g(t,t') - h(t,t')$ where g and h are two functions monotonic in t and t' in the same sense. We now prove:

Theorem 6.1. Let x_t be a stochastic process with a monotonoid and continuous covariance function $\sigma_{tt'}$ and $E(x_t) = 0$ then

(i) For $0 < t < T$ we have the expansion

$$(6.1)\quad x_t = \underset{m\to\infty}{\mathrm{l.i.m.}} \sum_{n=-m}^{n=+m} c_n \exp[2\pi i n t/T]$$

where

$$c_n = \frac{1}{T}\int_0^T x_t \exp[-2\pi i n t/T]\, dt$$

(ii) This limit is uniform in $0 < \varepsilon \le t < T - \varepsilon$

(iii) $$\sigma_{c_n c_m} = \frac{1}{T^2}\int_0^T\int_0^T \sigma_{tt'} \exp\left[-2\pi i\frac{nt+mt'}{T}\right]dt\, dt'$$

(iv) If the process is Gaussian, then any finite set of $c_n + \bar{c}_n$ and $c_n - \bar{c}_n$ are jointly normally distributed.

To simplify the proof we put $\tau = \frac{2\pi t}{T}$, $y_\tau = x_t$. Then τ goes from 0 to 2π as t goes from 0 to T and we have to prove the formula

$$(6.2)\quad y_\tau = \underset{m\to\infty}{\mathrm{l.i.m.}} \sum_{-m}^{m} c_n e^{in\tau},\quad c_n = \frac{1}{2\pi}\int_0^{2\pi} y_\tau e^{-in\tau} d\tau .$$

We thus have to prove

$$\lim_{m\to\infty} E[y_\tau^{(m)} - y_\tau]^2 = 0$$

uniformly in every interval $\varepsilon \leq \tau \leq 2\pi - \varepsilon$ where

$$y_\tau^{(m)} = \sum_{-m}^{+m} c_n e^{in\tau} \ .$$

We have

$$(6.3) \qquad y_\tau^{(m)} = \frac{1}{2\pi}\int_0^{2\pi} y_{\tau'}\left[\sum_{-m}^{+m} e^{in(\tau-\tau')}\right] d\tau'$$

Now

$$\sum_{n=-m}^{m} e^{in\alpha} = e^{-im\alpha}\,\frac{1-e^{i(2m+1)\alpha}}{1-e^{i\alpha}} = \frac{e^{-im\alpha}-e^{i(m+1)\alpha}}{1-e^{i\alpha}}$$

$$= \frac{e^{-im\alpha}-e^{i(m+1)\alpha}-e^{-i(m+1)\alpha}+e^{im\alpha}}{2(1-\cos\alpha)}$$

$$= \frac{\cos m\alpha - \cos(m+1)\alpha}{1-\cos\alpha} = \frac{\sin\frac{2m+1}{2}\alpha}{\sin\frac{\alpha}{2}}$$

Putting $\tau' = \tau + h$ we thus have

$$(6.4) \qquad y_\tau^{(m)} = \frac{1}{2\pi}\int_{-\tau}^{2\pi-\tau} y_{\tau+h}\,\frac{\sin\frac{2m+1}{2}h}{\sin\frac{h}{2}}\,dh$$

and

$$(6.5)\quad E[y_\tau^{(m)} - y_\tau]^2 = \frac{1}{4\pi^2} \int_{-\tau}^{2\pi-\tau} \int_{-\tau}^{2\pi-\tau} \sigma_{\tau+h,\,\tau+k} \frac{\sin\frac{2m+1}{2}h}{\sin\frac{h}{2}} \cdot \frac{\sin\frac{2m+1}{2}k}{\sin\frac{k}{2}}\, dh\, dk$$

$$- 2\,\frac{1}{2\pi} \int_{-\tau}^{2\pi-\tau} \sigma_{\tau,\,\tau+h} \frac{\sin\frac{2m+1}{2}h}{\sin\frac{h}{2}}\, dh + \sigma_{\tau\tau} \quad .$$

By well-known theorems on the Dirichlet integral[15] we have uniformly in $\varepsilon \leq \tau \leq 2\pi - \varepsilon$ $(\varepsilon > 0)$,

$$(6.6)\quad \lim_{m\to\infty} \frac{1}{4\pi^2} \int_{-\tau}^{2\pi-\tau} \int_{-\tau}^{2\pi-\tau} \sigma_{\tau+h,\,\tau+k} \frac{\sin\frac{2m+1}{2}h}{\sin\frac{h}{2}} \cdot \frac{\sin\frac{2m+1}{2}k}{\sin\frac{k}{2}}\, dh\, dk$$

$$= \lim_{m\to\infty} \frac{1}{2\pi} \int_{-\tau}^{2\pi-\tau} \sigma_{\tau,\,\tau+h} \frac{\sin\frac{2m+1}{2}h}{\sin\frac{h}{2}}\, dh = \sigma_{\tau\tau} \quad .$$

From (6.5) and (6.6) it follows that

$$(6.7)\quad \lim_{m\to\infty} E[y_\tau^{(m)} - y_\tau]^2 = 0 \quad \text{or} \quad \underset{m\to\infty}{\mathrm{l.i.m.}}\, y_\tau^{(m)} = y_\tau$$

uniformly in $\varepsilon \leq \tau \leq 2\pi - \varepsilon$ for every $\varepsilon > 0$.

---

[15] For the double Dirichlet integral see Hobson, "The theory of functions of a real variable and the theory of Fourier series", vol. II, pp. 705-9.

This completes the proof of the first two statements of theorem 6.1. (iii) is easily obtained by an elementary computation while (iv) follows from the representation of the Fourier coefficients as limits of Riemann sums.

2. Trigonometric expansion of the F. R. P.

As an example we shall represent the F.R.P. by a trigonometric series with random coefficients. The covariance function of the F.R.P. is $c \min(t,t')$ [see formula (2,4)], this is a monotonic and non-decreasing function of t and t' so that theorem 6.1 is applicable.

In this case c_n (that is the real and imaginary part of c_n) is normally distributed with mean zero and we have

$$(6.8)\quad E(y_\tau y_{\tau'}) = E(x_t x_{t'}) = c \min(t,t') = \frac{cT}{2\pi}\min(\tau,\tau') = c'\min(\tau,\tau')$$

Thus

$$E(c_n c_m) = c'\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\min(\tau,\tau')\exp[-in\tau-im\tau']\,d\tau\,d\tau'$$

$$= \frac{c'}{4\pi^2}\left\{\int_0^{2\pi}\left[\int_0^{\tau}\tau'\exp(-in\tau-im\tau')d\tau'\right]d\tau + \int_0^{2\pi}\left[\int_\tau^{2\pi}\tau\exp(-in\tau-im\tau')d\tau'\right]d\tau\right\}$$

For $n=m=0$ we obtain

$$(6.9.1)\quad E(c_0^2) = \frac{c'}{4\pi^2}\left\{\int_0^{2\pi}\frac{\tau^2}{2}d\tau + \int_0^{2}\tau(2\pi-\tau)d\tau\right\} = \frac{2\pi c'}{3} = \frac{cT}{3}$$

For $m\neq 0$ we have

(6.9.2) $$E(c_n c_m) = \frac{c'}{4\pi^2} \int_0^{2\pi} e^{-in\tau}\left(\frac{e^{-im\tau}}{m^2} - \frac{1}{m^2} - \frac{\tau}{im}\right) d\tau \; .$$

This gives

(6.9.3) $$E(c_0 c_m) = \frac{c'}{4\pi^2}\left[-\frac{2\pi^2}{im} - \frac{2\pi}{m^2}\right] = \frac{-cT}{4\pi m i} - \frac{cT}{4\pi^2 m^2} \quad \text{for } m \neq 0 ,$$

(6.9.4) $$E(c_m c_{-m}) = \frac{cT}{2\pi^2 m^2} \quad \text{for } m \neq 0 .$$

For $n \neq -m$, $n \neq 0$, $m \neq 0$ we obtain from (6.9.2)

(6.9.5) $$E(c_n c_m) = \frac{-cT}{4\pi^2 mn} .$$

If we write

$$x_t = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos 2\pi\frac{nt}{T} + b_n \sin 2\pi\frac{nt}{T}\right) \quad \text{l.i.m.}$$

we have

$$a_n = c_n + c_{-n}, \; b_n = i(c_n - c_{-n}) \quad \text{for } n > 0 , \; a_0 = c_0$$

and from this and the formulae (6.9.1) - (6.9.5) we find

$$
(6.10) \quad \begin{cases}
E(a_0^2) = cT/3 \\
E(a_0 a_n) = -cT/2\pi^2 n^2 \\
E(a_m a_n) = 0 \quad \text{for} \quad m \neq n,\ m \neq 0,\ n \neq 0 \\
E(a_n^2) = cT/2\pi^2 n^2 \\
E(a_n b_m) = 0 \quad \text{for } n \neq 0 \\
E(a_0 b_m) = -cT/2\pi m \\
E(b_n^2) = 3cT/2\pi^2 n^2 \\
E(b_n b_m) = cT/\pi^2 mn \quad \text{for} \quad m \neq n.
\end{cases}
$$

We shall now estimate $E\left[x_t - x_t^{(m)}\right]^2$ for a fixed m where

$$
x_t^{(m)} = a_0 + \sum_{n=1}^{m} a_n \cos 2\pi\frac{nt}{T} + \sum_{n=1}^{m} b_n \sin 2\pi\frac{nt}{T}.
$$

We have, using (6.10)

$$
E[x_t - x_t^{(m)}]^2 = \sum_{n=m+1}^{\infty} E(a_n^2)\cos^2 2\pi\frac{nt}{T} + \sum_{n=m+1}^{\infty} E(b_n^2)\sin^2 2\pi\frac{nt}{T}
$$

$$
\sum_{\substack{n=m+1 \\ n \neq k}}^{\infty} \sum_{k=m+1}^{\infty} E(b_n b_k)\sin 2\pi\frac{nt}{T} \sin 2\pi\frac{kt}{T}
$$

$$
= \frac{cT}{\pi^2}\left\{\frac{1}{2}\sum_{n=m+1}^{\infty}\frac{1}{n^2} + \left[\sum_{n=m+1}^{\infty}\frac{\sin\frac{2\pi t n}{T}}{n}\right]^2\right\}.
$$

Expanding the function $\pi - \alpha$ into a Fourier series we get

$$\pi - \alpha = 2\sum_{n=1}^{\infty} \frac{\sin n\alpha}{n},$$

hence

$$\sum_{n=m+1}^{\infty} \frac{\sin n\alpha}{n} = \frac{1}{2}(\pi - \alpha) - \sum_{n=1}^{m} \frac{\sin n\alpha}{n} = f(\alpha).$$

Differentiating this equation we have

$$f'(\alpha) = -\frac{1}{2} - \sum_{n=1}^{m} \cos n\alpha = -\frac{1}{2}\,\frac{\sin(m+\frac{1}{2})\alpha}{\sin\frac{\alpha}{2}},$$

and therefore

$$\sum_{n=m+1}^{\infty} \frac{\sin n\alpha}{n} = \frac{\pi}{2} - \frac{1}{2}\int_0^{\alpha} \frac{\sin(m+\frac{1}{2})t}{\sin\frac{t}{2}}\,dt = \frac{\pi}{2} - \int_0^{\frac{\alpha}{2}} \frac{\sin(2m+1)v}{\sin v}\,dv.$$

Hence

$$E[x_t - x_t^{(m)}]^2 = \frac{cT}{\pi^2}\left[\frac{1}{2}\sum_{m+1}^{\infty}\frac{1}{n^2}\cdot\left(\frac{\pi}{2} - \int_0^{\frac{\pi t}{T}} \frac{\sin(2m+1)v}{\sin v}\,dv\right)^2\right].$$

Thus for values of t not too close to 0 or T, $x_t^{(m)}$ is a good approximation to x_t.

Another and perhaps more useful formula may be obtained by deriving the following expansion

$$x_t - \frac{t}{T}x_T = \sum_{n=1}^{\infty}\left\{a_n\left[\cos\frac{2\pi nt}{T} - 1\right] + b_n \sin\frac{2\pi nt}{T}\right\}.$$

In this expansion the a_n and b_n are independently and normally distributed variables with mean zero and variances $\frac{cT}{2\pi^2 n^2}$, the a_n and b_n are also independent of x_T. The right side converges, moreover, uniformly in the mean to the left side. The proof can be obtained by first applying theorem 6.1 to the stochastic process $x_t - \frac{t}{T} x_T$ and determining the Fourier coefficients and their variances. It is then seen that the Fourier expansion thus obtained converges also l.i.m. for $t = T$ and thus $a_0 = -\sum_{n=1}^{\infty} a_n$. The proof is rather laborious but elementary and is therefore omitted.

Thus writing $x_T = a_0$ we have

$$(6.11) \qquad x_t = a_0 \frac{t}{T} + \sum_{n=1}^{\infty} \{a_n[\cos\frac{2\pi nt}{T} - 1] + b_n \sin\frac{2\pi nt}{T}\}$$

where

$$\sigma_{a_0}^2 = cT, \quad \sigma_{a_n}^2 = \sigma_{b_n}^2 = \frac{cT}{2\pi^2 n^2},$$

$$\sigma_{a_i a_j} = \sigma_{b_i b_j} = 0 \quad \text{for} \quad i \neq j, \quad \sigma_{a_i b_j} = 0$$

Except for the constant term, $-\sum_{n=1}^{\infty} a_n$, this is essentially the expansion discovered by Paley and Wiener.[16]

---

[16] Fourier transforms in the complex domain, p. 147.

3. Stationary processes.

We now return to the general theory and consider stationary processes.

We shall further assume that the covariance $E(x_t x_{t'}) = \sigma_{tt'}$ exists. We then have $\sigma_{tt'} = R(t-t')$ where $R(\tau)$ is an even function of τ.

We shall also consider a slightly more general class of processes, called quasistationary processes. A process x_t is said to be quasistationary if $E[x_t]$ is independent of t and if its covariance function exists and is given by $\sigma_{tt'} = R(t-t')$ where $R(\tau)$ is an even function of τ.

We assume now that $R(\tau)$ is continuous at the point $\tau = 0$ and show that $R(\tau)$ is then continuous everywhere. If $R(\tau)$ is continuous at $\tau = 0$ then we have $\lim_{\tau \to 0} E(x_{t+\tau} - x_t)^2 = 0$. From the definition of $R(\tau)$ we see that

$$\lim_{h \to 0} [R(\tau+h) - R(\tau)] = \lim_{h \to 0} E[(x_{\tau+h} - x_\tau)x_0] = 0$$

since

$$|E[(x_{\tau+h} - x_\tau)x_0]| \leq \sqrt{E(x_{\tau+h} - x_\tau)^2 \; E(x_0^2)} .$$

We next introduce the following definition: A function $f(t)$ is said to be positive definite if

(a) $f(t)$ is continuous and bounded on the real axis;

(b) $f(t)$ is Hermitian, that is $f(-t) = \overline{f(t)}$;

(c) for any positive integer m and any real numbers $z_1, z_2, \ldots, z_m$ and any complex numbers $u_1, u_2, \ldots, u_m$ we have

$$\sum_{h=1}^{m} \sum_{k=1}^{m} f(z_h - z_k) u_h \overline{u}_k \geq 0 .$$

From the preceding it is clear that $R(t)$ satisfies conditions (a) and (b) since $R(t)$ is real and even. We have only to prove that

$$S = \sum_1^m \sum_1^m R(t_h - t_k) u_h \overline{u}_k \geq 0 .$$

We have

$$S = \sum_1^m \sum_1^m u_h \overline{u}_k \, E(x_{\tau+t_h} x_{\tau+t_k}) = E\left\{\left(\sum_1^m u_h x_{\tau+t_h}\right)\left(\sum_1^m \overline{u}_k x_{\tau+t_k}\right)\right\}$$

$= E\left\{\left| \sum_1^m u_h x_{\tau+t_h} \right|^2\right\} \geq 0$. Therefore $R(t)$ is a positive definite function.

According to a theorem of S. Bochner[17] every positive definite function f(t) may be represented in the form

$$f(t) = \int_{-\infty}^{+\infty} e^{it\alpha}\, dV(\alpha)$$

where $V(\alpha)$ is a bounded non-decreasing function.

Thus we have

$$R(t) = \int_{-\infty}^{+\infty} e^{it\varpi}\, dg(\varpi) \tag{6.15}$$

where $g(\varpi)$ is a bounded and non-decreasing function. We may take $g(-\infty) = 0$. Then $g(\infty) = R(0)$ and $g(\alpha)/R(0)$ could therefore be defined as a distribution function. It will however simplify our formulae if we determine $g(\alpha)$ so that

$$g(\alpha) = \frac{g(\alpha+) + g(\alpha-)}{2} .$$

Since $R(t) = R(-t)$ we have

$$R(t) = \int_{-\infty}^{+\infty} e^{it\varpi}\, dg(\varpi) = \int_{-\infty}^{+\infty} e^{-it\varpi}\, dg(\varpi) = -\int_{-\infty}^{+\infty} e^{it\varpi}\, dg(-\varpi)$$

$$= \int_{-\infty}^{+\infty} e^{it\varpi}\, d[g(\infty)-g(-\varpi)] ,$$

---

[17] S. Bochner, Vorlesungen über Fouriersche Integrale, p. 76, Satz 23.

and since the function $g(\omega)$ is unique if $g(-\infty) = 0$ and

$g(\omega) = \frac{g(\omega+) + g(\omega-)}{2}$, we must have $g(\omega) = g(\infty) - g(-\omega)$

and for $\omega = 0$, $g(\infty) = 2g(0)$ and $g(\omega) - g(0) = g(0) - g(-\omega)$.

It is further well known that

$$g(\omega) - g(0) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{-T}^{T} R(t) \frac{e^{it\omega} - 1}{t}\, dt \; .$$

We may also write

$$R(t) = \frac{R(t)+R(-t)}{2} = \int_{-\infty}^{+\infty} \frac{e^{it\omega} + e^{-it\omega}}{2}\, dg(\omega)$$

$$= \int_{-\infty}^{\infty} \cos t\omega \, dg(\omega) = \int_{0}^{\infty} \cos t\omega \, d[g(\omega) - g(-\omega)]$$

$$= \int_{0}^{\infty} \cos t\omega \, dF(\omega) \; ,$$

where

$$(5.16) \quad \begin{cases} F(\omega) = g(\omega) - g(-\omega) = \dfrac{F(\omega+) + F(\omega-)}{2} \\[2ex] F(\infty) = g(\infty) = R(0) \\[2ex] F(0) = 0 \; . \end{cases}$$

Further

$$F(\omega) = g(\omega) - g(-\omega) = \lim_{T\to\infty} \frac{1}{2\pi i}\int_{-T}^{+T} R(t)\frac{e^{it\omega} - e^{-it\omega}}{t}\,dt$$

so that

$$F(\omega) = \frac{2}{\pi}\int_0^\infty R(t)\,\frac{\sin t\omega}{t}\,dt\,.$$

It may also be remarked that to every positive definite function $R(\tau)$ we may construct a Gaussian process with $R(\tau)$ as covariance function. This can be done by defining the distribution of $x_{t_1}, \dots, x_{t_n}$ to be a multivariate Gaussian distribution with covariance matrix $\| R(t_i - t_j) \|$. Since $R(t)$ is positive definite such a distribution always exists. It is then easy to verify that the family of distribution functions so defined satisfies the consistency conditions of chapter 1. Combining this with the result of Bochner we obtain

Theorem 6.2. The function $R(t)$ is the covariance function of a quasistationary process if and only if it is the Fourier transform of a bounded non-decreasing function.

4. The mean ergodic theorem.

We shall conclude this chapter with a proof of the mean ergodic theorem.

Theorem 6.3. (Mean ergodic theorem).[18] Let x_t be a quasistationary process with continuous covariance function $R(t)$ and mean value zero.

---

[18] This theorem is due to J. v. Neumann, Proc. Nat. Acad. Sci., vol. 18 (1932), pp. 70-82.

Then

$$(6.17) \qquad \underset{T\to\infty}{\mathrm{l.i.m.}}\ \frac{1}{T}\int_0^T e^{i\lambda t} x_t \, dt = a_\lambda$$

where a_λ is a random variable with variance $g(\lambda+) - g(\lambda-)$ and mean zero and where $E(a_\lambda a_\mu) = 0$ for $\lambda \neq -\mu$. The function $g(\lambda)$ is defined by (6.15).

We first prove the following lemma

Lemma 6.1: For any $0 \leq t \leq T$, $T \geq 1$ and for every ε

$$\left| \frac{1}{T}\int_0^T e^{i\lambda(t-\tau)} R(t-\tau)\, d\tau - \left[g(\lambda+\tfrac{\varepsilon}{T}) - g(\lambda-\tfrac{\varepsilon}{T})\right] \right|$$

$$\leq [g(\lambda+\varepsilon) - g(\lambda+) + g(-\lambda+\varepsilon) - g(-\lambda+)]$$

$$+ \frac{\varepsilon}{2}[g(\lambda+\varepsilon) - g(\lambda-\varepsilon)] + \frac{4}{\varepsilon T}[g(\infty) - g(-\infty)]$$

Proof: We put

$$\frac{1}{T}\int_0^T e^{i\lambda(t-\tau)} R(t-\tau)\, d\tau = \frac{1}{T}\int_0^T \int_{-\infty}^{\infty} \exp[i\lambda(t-\tau) + i\omega|t-\tau|]\, dg(\omega)\, d\tau$$

$$= I_1 + I_2 + I_3 + J_1 + J_2 + J_3$$

where

$$I_1 = \frac{1}{T}\int_0^t e^{i\lambda(t-\tau)} \int_{|\omega+\lambda|\geq\varepsilon} e^{i\omega(t-\tau)} dg(\omega)d\tau\,; \quad J_1 = \frac{1}{T}\int_t^T e^{i\lambda(t-\tau)} \int_{|\omega-\lambda|\geq\varepsilon} e^{-i\omega(t-\tau)} dg(\omega)d\tau\,;$$

$$I_2 = \frac{1}{T}\int_0^t e^{i\lambda(t-\tau)} \int_{\frac{\varepsilon}{T}\leq|\omega+\lambda|\leq\varepsilon} e^{i\omega(t-\tau)} dg(\omega)d\tau\,; \quad J_2 = \frac{1}{T}\int_t^T e^{i\lambda(t-\tau)} \int_{\frac{\varepsilon}{T}\leq|\omega-\lambda|\leq\varepsilon} e^{-i\omega(t-\tau)} dg(\omega)d\tau\,;$$

$$I_3 = \frac{1}{T}\int_0^t e^{i\lambda(t-\tau)} \int_{-\lambda-\frac{\varepsilon}{T}}^{-\lambda+\frac{\varepsilon}{T}} e^{i\omega(t-\tau)} dg(\omega)d\tau\,; \quad J_3 = \frac{1}{T}\int_t^T e^{i\lambda(t-\tau)} \int_{\lambda-\frac{\varepsilon}{T}}^{\lambda+\frac{\varepsilon}{T}} e^{-i\omega(t-\tau)} dg(\omega)d\tau\,.$$

These integrals converge absolutely. Hence we may interchange the order of integration whenever necessary. In this manner we obtain

$$|I_1| = \left|\frac{1}{T}\int_{|\omega+\lambda|\geq\varepsilon} dg(\omega) \int_0^t e^{i(\omega+\lambda)(t-\tau)}d\tau\right| = \left|\frac{1}{T}\int_{|\omega+\lambda|\geq\varepsilon} \frac{e^{i(\omega+\lambda)t}-1}{i(\omega+\lambda)}\, dg(\omega)\right|,$$

so that

(6.18) $$|I_1| \leq \frac{2}{\varepsilon T}\,[g(\infty)-g(-\infty)]$$

Similarly

(6.18a) $$|J_1| \leq \frac{2}{\varepsilon T}\,[g(\infty)-g(-\infty)]$$

We have, *for x real,*

(*) $$|e^{ix}-1| \leq |x|\,;$$

using this inequality we obtain from

$$|I_2| = \left|\frac{1}{T} \int\limits_{\frac{\varepsilon}{T} \le |\omega+\lambda| \le \varepsilon} \int_0^t \exp[i(\lambda+\omega)(t-\tau)]d\tau \, dg(\omega)\right|,$$

(6.19) $|I_2| \le \frac{t}{T}[g(\lambda+\varepsilon - g(\lambda+ \ \ +g(-\lambda+\varepsilon - g(-\lambda+ \ \ ]$

and similarly

(6.19a) $|J_2| \le \frac{T-t}{T}[g(\lambda+\varepsilon) - g(\lambda+) + g(-\lambda+\varepsilon) - g(-\lambda+)]$.

Since

$$\frac{t}{T}[g(\lambda+\frac{\varepsilon}{T}) - g(\lambda-\frac{\varepsilon}{T})] = \frac{1}{T}\int_{-\lambda-\frac{\varepsilon}{T}}^{-\lambda+\frac{\varepsilon}{T}} \int_0^t dg(\omega) \, d\tau$$

we have, using (*)

$$|I_3 - \frac{t}{T}[g(\lambda+\frac{\varepsilon}{T}) - g(\lambda-\frac{\varepsilon}{T})]| = \left|\frac{1}{T}\int_{-\lambda-\frac{\varepsilon}{T}}^{-\lambda+\frac{\varepsilon}{T}} \int_0^t [e^{i(\lambda+\omega)(t-\tau)} - 1] dg(\omega) \, d\tau\right|$$

$$\le \frac{1}{T}\int_{-\lambda-\frac{\varepsilon}{T}}^{-\lambda+\frac{\varepsilon}{T}} \int_0^t |\lambda+\omega|(t-\tau) \, dg(\omega) \, d\tau$$

$$\le \frac{\varepsilon t^2}{2T^2}\int_{-\lambda-\frac{\varepsilon}{T}}^{-\lambda+\frac{\varepsilon}{T}} dg(\omega) \le \frac{\varepsilon t}{2T}[g(\lambda+\frac{\varepsilon}{T}) - g(\lambda-\frac{\varepsilon}{T})]$$

Since $T \geq 1$ and since $g(x)$ is non-decreasing it is seen easily that

(6.20) $$\left| I_3 - \frac{t}{T}\left[g\left(\lambda+\frac{\varepsilon}{T}\right) - g\left(\lambda-\frac{\varepsilon}{T}\right)\right]\right| \leq \frac{\varepsilon t}{2T}[g(\lambda+\varepsilon) - g(\lambda-\varepsilon)] \;.$$

Similarly we obtain

(6.20a) $$\left| J_3 - \frac{T-t}{T}\left[g\left(\lambda+\frac{\varepsilon}{T}\right) - g\left(\lambda-\frac{\varepsilon}{T}\right)\right]\right| \leq \frac{\varepsilon(T-t)}{2T}[g(\lambda+\varepsilon) - g(\lambda-\varepsilon)] .$$

Lemma 6.1 then follows easily from (6.18), (6.18a), (6.19), (6.19a), (6.20), (6.20a).

Corollary 1 to lemma 6.1.

(6.21) $$\lim_{T\to\infty} \frac{1}{T}\int_0^T e^{i\lambda(t-\tau)}R(t-\tau)\,d\tau = g(\lambda+) - g(\lambda-) \,.$$

Corollary 2 to lemma 6.1.

(6.22) $$\lim_{\substack{T\to\infty \\ T'\to\infty}} \frac{1}{TT'}\int_0^T\int_0^{T'} e^{i\lambda(t-t')}R(t-t')\,dt\,dt' = g(\lambda+) - g(\lambda-) \;.$$

Proof. We may always write the double integral so that $T' > T$ so that lemma 6.1 is applicable and corollary 2 follows easily since ε is arbitrary.

In the proof of the mean ergodic theorem we shall operate with complex random variables. If $z = x+iy$ is a complex random variable with mean zero we shall define

$$\sigma_z^2 = E(z\bar{z}) = \sigma_x^2 + \sigma_y^2 \tag{6.23}$$

where $\bar{z} = x - iy$ is the complex conjugate to z. A sequence $\{z_n\} = \{x_n + iy_n\}$ of complex random variables converges if both $\{x_n\}$ and $\{y_n\}$ converge. From lemma 1.6 it follows that $\{z_n\}$ converges l.i.m. if and only if $E[(z_n - z_m)(\bar{z}_n - \bar{z}_m)]$ is arbitrarily small for sufficiently large n, m.

To show that $X_T = \frac{1}{T}\int_0^T x_t e^{i\lambda t}\,dt$ converges in the mean we consider

$$L_{TT'} = E[(X_T - X_{T'})(\bar{X}_T - \bar{X}_{T'})]$$

$$= \frac{1}{T^2}\int_0^T\int_0^T e^{i\lambda(t-t')}R(t-t')\,dt\,dt' + \frac{1}{T'^2}\int_0^{T'}\int_0^{T'} e^{i\lambda(t-t')}R(t-t')\,dt\,dt'$$

$$- \frac{2}{TT'}\int_0^T\int_0^{T'} e^{i\lambda(t-t')}R(t-t')\,dt\,dt'.$$

All three integrals converge to the same limit by (6.22). Thus $\underset{T\to\infty}{\text{l.i.m.}}\, X_T = a_\lambda$ exists. Moreover, by lemma 1.7 and (6.22)

$$\sigma^2_{a_\lambda} = \lim_{T\to\infty} \sigma^2_{X_T} = g(\lambda+) - g(\lambda-) .$$

For $\lambda \neq -\mu$ we further have

$$E(a_\lambda a_\mu) = \lim_{T\to\infty} \frac{1}{T^2}\int_0^T\int_0^T \exp(i\lambda t + i\mu t')R(t-t')dt\, dt'$$

$$= \lim_{T\to\infty} \frac{1}{T}\int_0^T e^{i(\lambda+\mu)t}\, \frac{1}{T}\int_0^T e^{-i\mu(t-t')}R(t-t')dt\, dt'$$

The second integral converges by corollary 1 of lemma 6.1 to $g(\mu+) - g(\mu-)$ uniformly in t. Thus

$$E(a_\lambda a_\mu) = \lim_{T\to\infty}\left\{\frac{1}{T}\int_0^T e^{i(\lambda+\mu)t}[g(\mu+) - g(\mu-)]dt + \eta(T)\right\}$$

where $\lim_{T\to\infty} \eta(T) = 0$. It easily follows that $E(a_\lambda a_\mu) = 0$.

<u>Theorem 6.4.</u> Let x_t be any quasistationary process with ~~correlation~~ covariance function $R(\tau)$ and let $g(\omega)$ be defined by (6.15). Further let λ_1, $\lambda_2, \ldots$ be the discontinuities of $g(\omega)$ and

$$a_\lambda = \underset{T\to\infty}{\text{l.i.m.}}\, \frac{1}{T}\int_0^T x_t e^{i\lambda t}\, dt$$

Then

$$x_t = \sum_{j=1}^{\infty} a_{\lambda_j} e^{-it\lambda_j} + y_t$$

where y_t is a quasistationary process such that

$$\underset{T\to\infty}{\text{l.i.m.}} \frac{1}{T}\int_0^T y_t e^{i\mu t}\, dt = 0$$

for all real numbers μ.

Proof: The sum $z_t = \sum_{j=1}^{\infty} a_{\lambda_j} e^{i\lambda_j t}$ converges in the mean

since

$$\sum_{j=m}^{n} \sigma^2_{a_{\lambda_j}} = \sum_{j=m}^{n} [g(\lambda_j+) - g(\lambda_j-)]$$

whilst

$$\sum_{j=1}^{n} [g(\lambda_j+) - g(\lambda_j-)] \leq g(\infty) - g(-\infty) \text{ for all } n .$$

Moreover

$$\underset{T\to\infty}{\text{l.i.m.}} \frac{1}{T}\int_0^T z_t e^{i\mu t}\, dt = \begin{cases} a_{\lambda_i} & \text{for } \mu = \lambda_i \, (i = 1, 2, \ldots) \\ 0 & \text{otherwise} \end{cases}$$

which proves theorem 6.4.

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The National Bureau of Standards is the principal agency of the Federal Government for fundamental and applied research in physics, mathematics, chemistry, and engineering. Its activities range from the determination of physical constants and properties of materials, the development and maintenance of the national standards of measurement in the physical sciences, and the development of methods and instruments of measurement, to the development of special devices for the military and civilian agencies of the Government. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various scientific and technical advisory services. A major portion of the NBS work is performed for other government agencies, particularly the Department of Defense and the Atomic Energy Commission. The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. The scope of activities is suggested in the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: the Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: the Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS