Differential Equations Determining a Markoff Process*

Kiyosi Itô

Introduction

(I) For a simple Markoff process $x_1, x_2, \cdots$ parametrized by the set of natural numbers and having a finite number of possible states $a_1, a_2, \ldots, a_m$, one can consider many kinds of transition probabilities. For example, the probability of the event $x_l = a_j$ given the condition that $x_k = a_i$, or the probability of $x_{n+1} = a_{n+1}$ given that $x_1 = a_{i_1}$, $x_2 = a_{i_2}, \ldots, x_n = a_{i_n}$, and so forth. However, the computation of these probabilities can be reduced eventually to the consideration of the probabilities $p_{ij}^{(k)}$ $(k = 1, 2, \cdots; i, j = 1, 2, \cdots m)$ for the event $x_{k+1} = a_j$ given that $x_k = a_i$. This fact is explained, for example, in the book by Kolmogoroff[1]. Let us call $p_{ij}^{(k)}$'s the basic transition probabilities from now on.

Even when the number of possible states is not finite, the situation remains the same if, for instance, the states are represented by the set of real numbers.

However, if the Markoff process is parametrized not by the natural numbers but by the reals, namely, if the process has continuous parameter, then the situation becomes more complicated[2].

More generally, for a simple Markoff process with its states being represented by the real numbers and having continuous parameter, the problem of determining quantities corresponding to $p_{ij}^{(k)}$ mentioned above and of constructing the corresponding Markoff process once these quantities are given has been investigated systematically by Kolmogoroff[3], who reduced the problem to the study of differential equations or integro-differential equations satisfied by the transition probability function.

W. Feller[4] has proved under fairly strong assumptions that these equations possess a unique solution and furthermore that the solution exhibits the properties of transition probability function.

However, if we adopt more strict point of view such as the one J. Doob[5] has applied toward his investigation of stochastic processes, it seems to us that the aforementioned work done by Feller is not quite adequate. For example, even though the differential equation determining the transition probability function of a continuous stochastic process was solved in §3 of that paper, no proof was given of the fact that it is possible to introduce by means of this solution a probability measure on some "continuous" function space.

The objective of this article, then, is:

1) to formulate the problem precisely, and
2) to give a rigorous proof, à la Doob, for the existence of continuous parameter stochastic processes.

(II) **Remark:** When x is a real valued random variable, a proposition concerning x can be represented in the form $x \in E$, where E is a subset of R^1. Therefore, if $x^{-1}(E)$ is P-measurable, it makes sense to consider a proposition of the form $x \in E$. Furthermore if $x \sim x'$, namely, if $P(x \neq x') = 0$, then the P-measure of the symmetric difference $(x \in E)\ \Delta\ (x' \in E)$ is 0, and therefore, the proposition $x \in E$ is a "permissible concept."[6]

Next, when x and y are a pair of real-valued random variables, then propositions concerning x and y can be represented in the form "the pair (x,y) belongs to some subset of R^2." For instance the proposition $x < y$ can be represented as $(x,y) \in E\ \{(\lambda,\mu);\ \lambda < \mu\}$. Consequently, it is clear that a proposition concerning x and y can be discussed if the corresponding subset of R^2 is a Borel set.

The situation remains the same for propositions concerning countable number of random variables, but it is different for the case where uncountably many random variables are involved. This is because the notion of the joint random variable for the case of uncountably many random variables is, in general, a "non-permissible concept"[7].

For example, when x_t is a real-valued stochastic process with continuous parameter t, the consideration of propositions such as "x_t is continuous in t" or "the least upper bound of x_t is M" necessitates the consideration of x_t jointly for all t. This is precisely where the problem arises. (Of course, if one is concerned with the proposition that x_t is continuous at $t = t_0$ with respect to the topology of convergence in probability, one can express it as $\lim_{t\to t_0} P(\ |x_t - x_{t_0}| > \epsilon) = 0$. Namely, what is involved here is a proposition $|x_t - x_{t_0}| > \epsilon$ which concerns only with a pair of random variables x_t and x_{t_0}, and accordingly no difficulty of the type mentioned above occurs).

However, it is not correct to say that it is impossible to consider propositions such as "x_t is continuous in t."

Definition. We say that x_t is continuous in t $(\in [a,b])$ if there exists a random variable y taking values in the space of continuous functions on $[a,b]$ such that for each t $P(x_t = y_t) = 1$, where y_t denotes the value of y at t.

Remark 1. y as above is determined uniquely by x_t. Hence, we shall call this y the joint random variable for x_t and denote it by $(x_\tau\ ; a \leqslant \tau \leqslant b)$ or more simply by x_{ab}. Furthermore, if $\alpha < \beta$ are real numbers in $[a,b]$, a random variable taking its values in the space of continuous functions on the interval $[\alpha,\beta]$ is obtained if we restrict y to $[\alpha,\beta]$. We denote this random variable by $x_{\alpha\beta}$.

* Translated from the original Japanese. First published in *Journ. Pan-Japan Math. Coll.* No. 1077 (1942).

Remark 2. $(y_\tau;\ a \leqslant \tau \leqslant b)$ equals y itself. This fact is obvious, but this is the reason which justifies the statement "y is x_{ab}."

Remark 3. The proposition "x_t is continuous at each fixed t with respect to the topology of convergence in probability" and the proposition "x_t is continuous in the sense of definition above" do not coincide. P. Lévy has distinguished the two concepts by saying "x_t has no fixed discontinuity points" in the case of the former and "x_t has no moving discontinuity points" for the latter[8]. (Of course, Lévy was concerned only with the case of differential processes.) Therefore, if $\{y_t\}$ has no moving discontinuity points, then with $(y_\tau;\ a \leqslant \tau \leqslant b)$ interpreted as above, it is possible to consider the proposition "$|y_\tau|$ is smaller than some finite number M".

I. Differentiation

§1. Definition of Differentiation of a Markoff Process

Let $\{y_t\}$ be a (simple) Markoff process and denote by F_{t_0t} the conditional probability distribution[1] of $y_t - y_{t_0}$ given that "y_{t_0} is determined". F_{t_0t} is clearly a P_{y_t}-measurable (ρ) function[2] of y_t, where ρ denotes the Lévy distance among probability distributions.

Definition 1.1.[3]
If

$$(1) \qquad F_{t_0t}^{*[1/t-t_0]}$$

(here $[a]$ is the integer part of the number a, and "$*k$" denotes the k-fold convolution) converges in probability with respect to the Lévy distance ρ as $t \to t_0+0$, then we call the limit random variable (taking values in the space of probability distributions) the derivative of $\{y_t\}$ at t_0 and denote it by

$$(2) \qquad D_{t_0}\{y_t\} \text{ or } Dy_{t_0}.$$

Corollary 1.1. *Dy_{t_0} is an infinitely divisible probability distribution.*[4]

Proof. It is possible to choose a sequence $t_1 \geqslant t_2 \geqslant \cdots \to t_0$ suitably so that

[1] This journal, vol **234**, Article #1033.

[2] This journal, vol **234**, Article #1033, *Ibid.* vol **235** Article #1043.

[3] This definition differs somewhat from the definition given by the author in Article #1033, volume **234** of this journal. However, the remarks made there are, of course, relevant here also.

[4] To be more precise, Corollary 1.1 is valid "with probability 1." However, as we remarked earlier, we shall omit the expression "with probability 1" unless it has to be emphasized.

$$(3) \qquad F_{t_0t_n}^{*[1/t_n-t_0]} \qquad (n = 1, 2, \cdots)$$

converges with probability 1. When $F_{t_0t_n}^{*[1/t_n-t_0]}$ converges, its limit would be a so-called "limit law" in the sense of Khintchine and, therefore, it is infinitely divisible. Consequently, with probability 1, Dy_{t_0} is an infinitely divisible probability distribution.

Dy_{t_0} obtained above is a function of t_0 as well as of y_{t_0}, and so, we denote it by $L(t_0, y_{t_0})$ corresponds precisely to the "basic transition probability" discussed in the Introduction.

The precise formulation of the problem of Kolmogoroff, then, is to solve the equation

$$(4) \qquad Dy_t = L(t, y_t)$$

when the quantity $L(t,y)$ is given.

§2. A Comparison Theorem

Let us prove a comparison theorem for Dy_{t_0} which we shall make use of later.

Theorem 2.1. *Let $\{y_t\}$, $\{z_t\}$ be simple Markoff processes satisfying the following conditions:*

(1) $y_{t_0} = z_{t_0}$.
(2) $E(y_t - z_t \mid y_{t_0}) = o(t - t_0)$, *where o is the Landau symbol.*
(3) $\sigma(y_t - z_t \mid y_{t_0}) = o(\sqrt{t - t_0})$.

(Here $E(x|y)$ denotes the conditional expectation of x given y and $\sigma(x|y)$ denotes the conditional standard deviation of x given y. Also, the quantity o may depend on t_0 or y_{t_0}). Then, whenever Dz_{t_0} exists, Dy_{t_0} exists also, and $Dy_{t_0} = Dz_{t_0}$ holds.

Proof. For given ϵ and η, we can choose $\delta(\epsilon,\eta)$ sufficiently small so that if $|t - t_0| < \delta(\epsilon,\eta)$ then with probability bigger than $1 - \eta$ the following are satisfied:

$$(4) \qquad \rho(F_{z_t-z_{t_0}|z_{t_0}}^{*[1/t-t_0]}, Dz_{t_0}) < \epsilon \text{ [5]}$$

(here by $F_{x|y}$ we mean the conditional probability distribution of x given y)

$$(5) \qquad |E(y_t - z_t \mid y_{t_0})| < \epsilon(t - t_0)$$

$$(6) \qquad |\sigma(y_t - z_t \mid y_{t_0})|^2 < \epsilon(t - t_0).$$

[5] $F_{x|y}$ denotes the conditional probability distribution of x given y.

Since $y_{t_0} = z_{t_0}$, we have

$$y_t - y_{t_0} = (z_t - z_{t_0}) + (y_t - z_t). \tag{7}$$

Let us denote points of the $2n$ dimensional Euclidean space R^{2n} by $(\zeta_1, \xi_1, \zeta_2, \xi_2, \cdots \zeta_n, \xi_n)$, and introduce a probability measure of R^{2n} by

$$P'(d\zeta_1 d\xi_1 d\zeta_2 d\xi_2 \cdots d\zeta_n d\xi_n) = \prod_{i=1}^{n} F(d\zeta_i d\xi_i), \tag{8}$$

where F denotes the conditional probability distribution

$$F_{(z_t - z_{t_0},\, y_t - z_t) | z_{t_0}}. \tag{9}$$

The mapping

$$(\zeta_1, \xi_1, \zeta_2, \xi_2 \cdots \zeta_n, \xi_n) \to \zeta_i \quad (\text{or } \xi_i) \tag{10}$$

may be regarded as a random variable defined on the probability space (R^{2n}, P'). We denote this random variable again by ζ_i (or ξ_i). If we define

$$\eta_i = \zeta_i + \xi_i, \qquad \eta = \sum \eta_i, \qquad \zeta = \sum \zeta_i, \qquad \xi = \sum \xi_i \tag{11}$$

then $\{\eta_i\}$ gives a set of mutually independent random variables, and so do $\{\zeta_i\}$ and $\{\xi_i\}$. Furthermore, the probability distribution functions of ζ_i, ξ_i, η_i are given, respectively, by

$$F_{z_t - z_{t_0} | y_{t_0}} \qquad (\text{that is, } F_{z_t - z_{t_0} | z_{t_0}})$$

$$F_{y_t - z_t | y_{t_0}} \quad \text{and} \quad F_{y_t - y_{t_0} | z_{t_0}}.$$

Therefore, if we take $n = [1/t - t_0]$, then

$$F^{*[1/t - t_0]}_{z_t - z_{t_0} | z_{t_0}} = F'_\zeta \tag{12}$$

$$F^{*[1/t - t_0]}_{y_t - y_{t_0} | y_{t_0}} = F'_\eta$$

where F'_ζ and F'_η are probability distribution functions for ζ and η, respectively. Therefore, by (4)

$$\rho(F'_\zeta, Dz_{t_0}) < \epsilon, \tag{13}$$

and since $\xi_1, \xi_2, \cdots \xi_n$ are independent, we obtain

$$(\sigma(\xi))^2 = \sum_{i=1}^{n} (\sigma(\xi_1))^2 < \left[\frac{1}{t - t_0}\right] (t - t_0)\, \epsilon \leqslant \epsilon \tag{14}$$

and

$$|E(\xi)| \leqslant \sum_{i=1}^{n} |E(\xi_i)| < \left[\frac{1}{t - t_0}\right] (t - t_0)\epsilon \leqslant \epsilon. \tag{15}$$

Consequently, $E(\xi^2) = (E(\xi))^2 + (\sigma(\xi))^2 < \epsilon + \epsilon^2 < 2\epsilon$ $(\epsilon < 1)$, and

$$P(|\xi| > \epsilon^{1/4}) < 2\epsilon^{1/2},$$

that is, $P(|\eta - \zeta| > \epsilon^{1/4}) < 2\epsilon^{1/2}$, which in turn implies that

$$d(\eta, \zeta) < \epsilon^{1/4} + 2\epsilon^{1/2} < 3\epsilon^{1/4}. \text{ [6]}$$

According to P. Lévy[7], we also have

$$\rho(F_\eta', F_\zeta') < 3\sqrt{2}\epsilon^{1/4}. \tag{16}$$

From (13) and (16) it follows that

$$\rho(F_\eta', Dz_{t_0}) < \epsilon + 3\sqrt{2}\epsilon^{1/4} < (3\sqrt{2} + 1)\epsilon^{1/4}, \tag{17}$$

and from (12) and (17) it follows further that

$$\rho(F^{*[1/t - t_0]}_{y_t - y_{t_0} | y_{t_0}}, Dz_{t_0}) < (3\sqrt{2} + 1)\epsilon^{1/4}. \tag{18}$$

Thus, we see that the inequality (18) is valid for an arbitrary pair ϵ, η with probability greater than $1 - \eta$ as long as $|t - t_0| < \delta(\epsilon, \eta)$ is satisfied. We therefore conclude that $Dy_{t_0} = Dz_{t_0}$.

Theorem 2.2. *Let $\{y_t\}$ and $\{z_t\}$ be simple Markoff processes satisfying the following:*

$$y_{t_0} = z_{t_0} \tag{19}$$

$$d(y_t, z_t) = o(t - t_0) \tag{20}$$

(here the quantity o may depend on t_0 and y_{t_0}).

Then, if Dz_{t_0} exists, so does Dy_{t_0} and, furthermore, $Dy_{t_0} = Dz_{t_0}$ is satisfied.

Remark. $d(y_t, z_t) = \inf_{\alpha > 0}\{P(|y_t - y_t| > \alpha) + \alpha\}$.

Proof. In view of hypothesis (20), we have, if $\delta(\epsilon)$ is chosen sufficiently small,

$$P\{|y_t - z_t| > \epsilon(t - t_0)\} < \epsilon(t - t_0)$$

[6,7] Theorie de l'addition des variables aléatorires, p. 51.
$d(x, y) = \inf_{\eta > 0}\{P\{|x - y| > \eta\} + \eta\}$.

whenever $|t - t_0| < \delta(\epsilon)$.

With the same notations as in the proof of the preceeding theorem, we obtain

$$P(|\xi_i| > \epsilon(t - t_0)) < \epsilon(t - t_0) \qquad (i = 1, 2, \cdots n)$$

$$P(|\xi| > \epsilon \left[\frac{1}{t-t_0}\right](t-t_0)) < \sum_{i=1}^{n} P(|\xi_i| > \epsilon(t-t_0)) \quad (\text{where } n = \left[\frac{1}{t-t_0}\right])$$

$$< n\epsilon(t - t_0) = \epsilon(t - t_0)\left[\frac{1}{t-t_0}\right].$$

Therefore, $P(|\xi| > 2\epsilon) < 2\epsilon$.

We can then conclude the proof of this theorem just as in the proof of the preceding theorem.

§3. Examples

Example 1. When $\{x_t\}$ is a (temporally) homogeneous differential process, $Dx_t = F_{x_t - x_0}$ (independent of (t, x_t)).

Example 2. It happens quite frequently that $\{x_t\}$ is a differential process and the characteristic function $\varphi_{t,s}(z)$ of the difference $x_s - x_t$ is given in the following form:

$$\log \varphi_{t,s}(z) = \int_s^t \psi_\tau(z) d\tau,$$

where

$$\psi_\tau(z) = im_\tau z - \frac{\rho_\tau^2}{2} z^2 + \left(\int_{-\infty}^{-0} + \int_{+0}^{+\infty}\right)(e^{izu} - 1 - \frac{izu}{1+u^2}) n_\tau(du).$$

Let us assume that $\psi_\tau(z)$ is continuous in τ when z is fixed and is equicontinuous in some neighborhood of $z = 0$. Then, we have $\log\varphi_{Dx_t}(z) = \psi_t(z)$ (independent of x_t). Here $\varphi_{Dx_t}(z)$ denotes the characteristic function of the probability distribution Dx_t.

Conversely, one can construct a corresponding differential process when $\psi_t(z)$ is given.

Example 3. Let $\{x_t\}$ be a Brownian motion. Namely, $\{x_t\}$ is a temporally homogeneous differential process without moving discontinuity points and $x_1 - x_0$ has the normal distribution. Then, by example 1, we see that

(1) $\qquad Dx_t =$ normal distribution.

In the sequel, we shall denote for the sake of simplicity, by $G(a,b)$ the Gaussian distribution with mean a and standard deviation b. Then, the normal distribution is written as $G(0,1)$. If we consider $\{y_t\}$ given by

(2) $$y_t = (x_t - x_0)^2,$$

then since

(3) $$y_t = (x_{t_0} - x_0)^2 + (x_t - x_{t_0})^2 + 2(x_{t_0} - x_0)(x_t - x_{t_0}),$$

we see that y_t is a simple Markoff process. From

(4) $$\begin{aligned} y_t &= y_{t_0} + (x_t - x_{t_0})^2 + 2(x_{t_0} - x_0)(x_t - x_{t_0}) \\ &= \{y_{t_0} + (t - t_0) + 2(x_{t_0} - x_0)(x_t - x_{t_0})\} \\ &\quad + \{(x_t - x_{t_0})^2 - (t - t_0)\} \end{aligned}$$

it follows that if we denote by z_t the quantity inside of the first { } on the right-hand side above, then

(5) $$y_{t_0} = z_{t_0}$$

(6) $$\begin{aligned} E(y_t - z_t) &= E\{(x_t - x_{t_0})^2 - (t - t_0)\} \\ &= t - t_0 - (t - t_0) = 0 \end{aligned}$$

(7) $$\begin{aligned} (\sigma\{y_t - z_t\})^2 &= E\{(y_t - z_t)^2\} \\ &= E\{(x_t - x_{t_0})^4\} - (t - t_0)^2 \end{aligned}$$

$$\begin{aligned} E\{(x_t - x_{t_0})^4\} &= \int_{-\infty}^{\infty} \frac{\xi^4}{\sqrt{2\pi(t-t_0)}} \exp\left\{-\frac{\xi^2}{2(t-t_0)}\right\} d\xi \\ &= \left(\int_{-\infty}^{\infty} \frac{\lambda^4}{\sqrt{2\pi}} \exp\left\{-\frac{\lambda^2}{2}\right\} d\lambda\right)(t-t_0)^2 \\ &= 3(t - t_0)^2. \end{aligned}$$

Therefore, we have

(8) $$(\sigma\{y_t - z_t\})^2 = 2(t - t_0)^2 = o(t - t_0).$$

On the other hand, from the definition of z_t it follows that

$$F_{z_t - z_{t_0} | x_{t_0} - x_0} = G(t - t_0, 2|x_{t_0} - x_0|\sqrt{t - t_0}).$$

Since the right-hand side above depends only on $|x_{t_0} - x_0|$, we conclude that

$$F_{z_t - z_{t_0} | (x_{t_0} - x_0)^2} = G(t - t_0, 2|x_{t_0} - x_0|\sqrt{t - t_0}),$$

which implies that

$$F_{z_t - z_{t_0} | z_{t_0}} = G(t - t_0, 2 \mid x_{t_0} - x_0 \mid \sqrt{t - t_0}).$$

Consequently, we have $Dz_{t_0} = G(1, 2\sqrt{z_{t_0}})$. Therefore, if we use Theorem 2.1 we can conclude from the equation above and from (5), (6), and (8) that

$$Dy_{t_0} = G(1, 2\sqrt{z_{t_0}}).$$

II. Integration

4. Definite Integral

Let x_t be a Brownian motion satisfying $x_0 = 0$.[8] Assume that b_t is a function of x_{0t} (that is, of $(x_\tau: 0 \leqslant \tau \leqslant t)$)[9] having no moving discontinuity points. Denote by Δ a $2n+1$-tuple of real numbers $t_0, t_1, \ldots, t_n, \tau_0, \tau_1, \cdots \tau_{n-1}$ satisfying the following conditions:

(1) $$0 = t_0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n = 1$$

(2) $$0 \leqslant \tau_0, \tau_1, \tau_2, \cdots \tau_{n-1} \leqslant 1$$

(3) $$\tau_i \leqslant t_i \ (i = 0, 1, 2, \cdots n-1).$$

Define $d(\Delta) = \max_{1 \leqslant i \leqslant n}(t_i - \tau_{i-1})$. Then, clearly, we have

(4) $$t_i - \tau_{i-1} \leqslant d(\Delta) \qquad (i = 1, 2, \cdots n).$$

Let us call

(5) $$y_\Delta = \sum_{i=1}^{n} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})$$

a δ-sum over the partition $(t_0, t_1, \cdots t_n)$ when $d(\Delta) < \delta$.

Theorem 4.1. *y_Δ converges in probability as $d(\Delta) \to 0$.*

Proof. Let $s_1, s_2, \ldots, s_m$ be a refinement of $t_1, t_2, \cdots t_n$ and let $\{\sigma_i\}$ be given by

(6) $$\sigma_{i-1} = \tau_{k-1} \quad \text{if } (s_{i-1}, s_i) \subset (t_{k-1}, t_k).$$

Then

[8] Cf. Example 3 of § 3.

[9] Cf. Introduction (II) of this article.

(7) $$\sum_{i=1}^{n} b_{\sigma_{i-1}}(x_{s_i} - x_{s_{i-1}})$$

is also a δ-sum, and it equals y_Δ. We call (7) the refined representation of y_Δ over the partitions $s_1, s_2, \cdots s_m$. When two δ-sums, y_Δ and $y_{\Delta'}$ are given, we can consider the common refinement of partitions for Δ and Δ' by taking the union of sub-division points for each, and by considering the corresponding refined representations of y_Δ and $y_{\Delta'}$, we can represent y_Δ and $y_{\Delta'}$ as δ-sums over a single partition.

(8) $$y_\Delta = \sum_{i=1}^{n} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})$$

(9) $$y_{\Delta'} = \sum_{i=1}^{n} b_{\tau'_{i-1}}(x_{t_i} - x_{t_{i-1}})$$

Therefore,

(10) $$y_\Delta - y_{\Delta'} = \sum (b_{\tau_{i-1}} - b_{\tau'_{i-1}})(x_{t_i} - x_{t_{i-1}}).$$

Since b_t has no moving discontinuity points, we can, for any given ϵ, η, choose $\delta(\epsilon,\eta)$ sufficiently small so that

(11) $$P\left[\bigcap_{|t-s| < \delta(\epsilon,\eta)} (|b_t - b_s| < \epsilon)\right] > 1 - \eta.$$

If we now choose both $d(\Delta)$ and $d(\Delta')$ to be smaller than $\delta(\epsilon,\eta)$, and define c_i $(i=1,2, \cdots n)$ by

(12) $$c_i = b_{\tau_{i-1}} - b_{\tau'_{i-1}} \quad \text{if } |b_{\tau_{i-1}} - b_{\tau'_{i-1}}| < \epsilon$$

and

$$= 0 \quad \text{if } |b_{\tau_{i-1}} - b_{\tau'}| \geqslant \epsilon,$$

then by (11) we obtain

(13) $$P(y_\Delta - y_{\Delta'} \neq \sum_i c_i(x_{t_i} - x_{t_{i-1}})) < \eta,$$

and since c_i is clearly a function of $x_{0t_{i-1}}$ we see that c_i and $x_{t_i} - x_{t_{i-1}}$ are mutually independent.

Furthermore,

(14) $$E\{(\sum_i c_i(x_{t_i} - x_{t_{i-1}}))^2\}$$

$$= \sum_i E\{c_i^2(x_{t_i} - x_{t_{i-1}})^2\} + 2\sum_{i<j} E\{c_i c_j (x_{t_i} - x_{t_{i-1}})(x_{t_j} - x_{t_{j-1}})\}$$

$$= \sum_i E(c_i^2)\, E\{(x_{t_i}-x_{t_{i-1}})^2\} + 2\sum_{i<j} E\{c_ic_j(x_{t_i}-x_{t_{i-1}})\}E\{x_{t_j}-x_{t_{j-1}}\}$$ [10]

$$\leqslant \sum_i \epsilon^2(t_i - t_{i-1}) = \epsilon^2.$$

Consequently,

(15) $$P\{|\sum_i c_i(x_{t_i}-x_{t_{i-1}})| > \sqrt{\epsilon}\} < \epsilon.$$

From (13) and (14) we obtain

$$P\{|y_\Delta - y_{\Delta'}| > \sqrt{\epsilon}\} < \epsilon + \eta \qquad \text{q.e.d}$$

We note that in the argument above, the interval [0, 1] may be replaced by any interval $[t, s]$.

Definition 4.1. By $\int_0^1 b_\tau dx_\tau$ we shall mean the limit of y_Δ whose existence was proved in Theorem 4.1. We interpret $\int_t^s b_\tau dx_\tau$ similarly.

5. Theorems on Definite Integrals

Throughout this section, we assume that $\{x_t\}$ is a Brownian motion and that integrands (b_t, c_t, etc.) considered are functions of x_{0t} having no moving discontinuity points.

The integral defined in the preceding section satisfies the following properties familiar for the ordinary integral:

Theorem 5.1. $\int_t^s dx_\tau = x_s - x_t$.

Theorem 5.2. $\int_t^s(\lambda b_\tau + \mu c_\tau)dx_\tau = \lambda\int_t^s b_\tau dx_\tau + \mu\int_t^s c_\tau dx_\tau$.

Theorem 5.3. *If* $t < s < u$,

$$\int_t^s b_\tau dx_\tau + \int_s^u b_\tau dx_\tau = \int_t^u b_\tau dx_\tau.$$

Theorem 5.4. *Let* $y = \int_t^s b_\tau dx_\tau$. *If there exists a continuous function* $M(\tau)$ *such that*

(1) $$E(b_\tau^2) \leqslant M(\tau),$$

[10] Since $|c_i| \leqslant \epsilon$, $E(c_i^2)$ exists and since c_i and $x_{t_i} - x_{t_{i-1}}$ are independent,

$$E\{c_i^2(x_{t_i} - x_{t_{i-1}})^2\} = E(c_i^2)E\{(x_{t_i} - x_{t_{i-1}})^2\}$$

$c_ic_j(x_{t_i} - x_{t_{i-1}})$ is a function of $x_{0t_{j-1}}$, and is independent of $x_{t_j} - x_{t_{j-1}}$ since $i < j$. Furthermore, $E(c_ic_j(x_{t_i} - x_{t_{i-1}}))$ exists because $|c_ic_j(x_{t_i} - x_{t_{i-1}})| \leqslant \epsilon^2|x_{t_i} - x_{t_{i-1}}|$. Consequently,

$$E\{c_ic_j(x_{t_i} - x_{t_{i-1}})(x_{t_j} - x_{t_{j-1}})\} = E\{c_ic_j(x_{t_i} - x_{t_{i-1}})\}\, E\{x_{t_j} - x_{t_{j-1}}\}.$$

then

(2) $$E(y^2) \leqslant \int_t^s M(\tau)d\tau.$$

Theorem 5.5. *If a sequence* $\{b_\tau^{(n)}\}$ $(n=1,2,3,\cdots)$ *of stochastic processes converges to* $\{b_\tau\}$ *in the sense of strong topology, namely, if*

(3) $$P\{\sup_{s\leqslant\tau\leqslant t} |b_\tau^{(n)} - b_\tau| > \epsilon\} \to 0 \quad \text{as } n\to\infty,$$

then $\int_t^s b_\tau^{(n)} dx_\tau$ *converges in probability to* $\int_t^s b_\tau dx_\tau$.

Theorems 5.1–5.3 are obvious. In order to prove Theorems 5.4 and 5.5, we shall first prove the following lemma.

Lemma 5.1. *Let* $x, x_1, x_2, \cdots$ *be real-valued random variables satisfying the following properties:*

(4) $$x_1, x_2, \cdots \text{ converges with probability 1 to } x,$$

(5) $$x_1, x_2 \cdots \geqslant 0,$$

(6) $$E(x_n) \leqslant e_n \qquad (n = 1, 2, \cdots),$$

(7) $$e_n \to e.$$

Then, we have

(8) $$E(x) \leqslant e.$$

Proof. Let $y_n = \inf\{x_n, x_{n+1}, \cdots\}$. Then,

(9) $$0 = y_1 \leqslant y_2 \leqslant \cdots \to x$$

(10) $$0 \leqslant E(y_n) \leqslant E(x_n) \leqslant e_n \qquad (n = 1, 2, \cdots).$$

Since $\{y_n\}$ is monotone increasing by (9),

$$E(x) = E(\lim_{n\to\infty} y_n) = \lim E(y_n).$$

On the other hand, from (10) it follows that

$$\lim_{n\to\infty} E(y_n) \leqslant \overline{\lim}\, e_n = \lim e_n = e.$$

Therefore, we obtain $E(x) \leqslant e$.

Proof of Theorem 5.4. We can choose a sequence $\{\Delta_n\}$ suitably with $d(\Delta_n) \to 0$ so that

$$y = \int_t^s b_\tau dx_\tau = \lim_{n\to\infty} \sum_{\Delta_n} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}}) \tag{11}$$

holds with probability 1.[11] Therefore

$$y_2 = \left(\int_t^s b_\tau dx_\tau\right)^2 = \lim_{n\to\infty}\left(\sum_{\Delta_n} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})\right)^2.$$

If we let $y_n = \sum_{\Delta_n} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})$, then

$$E(y_n^2) = \sum_i E(b_{\tau_{i-1}}^2)\, E\{(x_{t_i} - x_{t_{i-1}})^2\}$$

$$+\, 2\sum_{i<j} E\{b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})b_{\tau_{j-1}}\}E(x_{t_j} - x_{t_{j-1}}).$$ [12]

Since $E(x_{t_i} - x_{t_{i-1}}) = 0$, we obtain $E(y_n^2) \leqslant \sum_{\Delta_n} M(\tau_{i-1})(t_i - t_{i-1})$.
Since the right-hand side above tends to $\int_t^s M(\tau)d\tau$ as $n\to\infty$, we obtain

$$E(y^2) \leqslant \int_t^s M(\tau)d\tau \quad \text{in view of Lemma 5.1.}$$

Proof of Theorem 5.5. Because of the assumption of convergence in probability, we can choose, for given ϵ and η, n sufficiently large so that

$$P\{\sup_{t\leqslant\tau\leqslant s} |b_\tau^{(n)} - b_\tau| \geqslant \epsilon\} < \eta. \tag{12}$$

Let us now define c_τ as follows:

$$\begin{aligned} c_\tau &= \epsilon && \text{if } b_\tau^{(n)} - b_\tau > \epsilon \\ &= b_\tau^{(n)} - b_\tau && \text{if } -\epsilon \leqslant b_\tau^{(n)} - b_\tau \leqslant \epsilon \\ &= -\epsilon && \text{if } b_\tau^{(n)} - b_\tau < -\epsilon. \end{aligned}$$

Then, c_τ is a function of $x_{0\tau}$ and the stochastic process $\{c_\tau\}$ has no moving discontinuity points. Therefore, one can consider the integral $\int_t^s c_\tau dx_\tau$. According to (12) we have

[11] P. Lévy: *Ibid.*, pp. 55 and 56.

[12]
$$\begin{aligned} &E(|b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})b_{\tau_{j-1}}|) \\ &\leqslant \sqrt{E\{b_{\tau_{i-1}}^2(x_{t_i} - x_{t_{i-1}})^2\}\, E\{b_{\tau_{j-1}}^2\}} \\ &= \sqrt{E(b_{\tau_{i-1}}^2)E\{(x_{t_i} - x_{t_{i-1}})^2\}\, E(b_{\tau_{j-1}}^2)} \quad (\text{since } b_{\tau_{i-1}} \text{ is independent of } x_{t_i} - x_{t_{i-1}}) \\ &\leqslant \sqrt{M(\tau_{i-1})M(\tau_{j-1})(t_i - t_{i-1})}. \end{aligned}$$

Therefore, $b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})b_{\tau_{j-1}}$ is integrable.

$$P\left\{\int_t^s c_\tau dx_\tau \neq \int_t^s (b_\tau^{(n)} - b_\tau)dx_\tau\right\} < \eta. \tag{13}$$

and furthermore, $|c_\tau| \leqslant \epsilon$ and hence $E(c_\tau^2) \leqslant \epsilon^2$. Therefore we have

$$E\left\{\left(\int_t^s c_\tau dx_\tau\right)^2\right\} \leqslant \epsilon^2(s - t),$$

from which it follows that

$$P\left\{\left|\int_t^s c_\tau dx_\tau\right| > \sqrt{\epsilon}\right\} \leqslant \epsilon(s - t). \tag{14}$$

From (13) and (14) we obtain that

$$P\left\{\left|\int_t^s b_\tau^{(n)}\, dx_\tau - \int_t^s b_\tau dx_\tau\right| > \sqrt{\epsilon}\right\} \leqslant \epsilon(s-t) + \eta,$$

from which we conclude that $\int_t^s b_\tau^{(n)}\, dx_\tau$ converges in probability to $\int_t^s b_\tau dx_\tau$.

6. Indefinite Integrals

Theorem 6.1. $\{\int_0^t b_\tau dx_\tau\}$ $(0 \leqslant t \leqslant 1)$ *has no moving discontinuity points.*

More precisely, there exists a random variable y taking values in the space of continuous functions on [0,1] such that for arbitrary t $(0 \leqslant t \leqslant 1)$

$$P\left\{\int_0^t b_\tau dx_\tau = y_t\right\} = 1 \quad \text{is satisfied,}$$

where y_t denotes the real-valued random variable taking the value of y at t. (The fact that such y is determined uniquely up to equivalence was explained in Introduction (II).)

Definition 6.1. We call y in Theorem 6.1 the indefinite integral of b_τ with respect to x_τ.

Proof of Theorem 6.1. $\{y_\Delta(t)\}$ defined by

$$y_\Delta(t) = \sum_{i=1}^{k-1} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}}) + b_{\tau_{k-1}}(x_t - x_{t_{k-1}}) \quad \text{for } t \in [t_{k-1}, t_k] \tag{1}$$

obviously has no moving discontinuity points. Therefore, we may regard $y_\Delta(t)$ as the value at t of some random variable—which we denote by y_Δ also—taking values in the space of continuous functions. It suffices, therefore, to show that, as $d(\Delta)\to 0$, $y_\Delta = (y_\Delta(t); 0 \leqslant t \leqslant 1)$ converges in probability with respect to the strong topology. This is because it would then follow that we can choose a sequence $\{\Delta_n\}$ so that the sequence $\{y_{\Delta_n}\}$ ($n=1, 2, \cdots$) converges with probability 1, and if we donate the limit by y, then y would be a continuous function with probability 1 since it is with probability 1 the limit of the uniformly convergent sequence of continuous functions $\{y_{\Delta_n}\}$.

Let us first state a lemma.

Lemma 6.1. *Let $x_1, x_2, \cdots, x_n$ be mutually independent real-valued random variables and let for each $i = 1, 2, \cdots, n$ y_i be a real-valued random variable independent of $(x_i, x_{i+1}, \cdots, x_n)$. If*

(2) $$E(x_i) = 0, \quad E(y_i^2) < \infty \qquad (i = 1, 2, \ldots, n)$$

then

(3) $$P\left\{\max_{1\leqslant p\leqslant n} |y_1x_1 + y_2x_2 + \cdots + y_kx_k| \geqslant l\right\} \leqslant \frac{E\{(y_1x_1 + y_2x_2 + \cdots + y_nx_n)^2\}}{l^2}.$$

This lemma is an extension of the Kolmogoroff inequality (which corresponds to the case where $y_1 = y_2 = \cdots = y_n = 1$), and its proof is identical with that for the latter. Hence, we shall omit its proof.

Going back to the proof of Theorem 6.1, let us suppose that $s_1, s_2, \cdots$ is a sequence of points dense in (0,1), and represent by $0 = t_0 < t_1 < \cdots < t_n = 1$ a new partition of (0,1) obtained by joining $s_1, s_2, \cdots s_m$ to the points of partitions Δ' and Δ. Then we have

(4) $$y_\Delta(t_k) = \sum_{i=1}^{k} b_{\tau_{i-1}}(x_{t_i} - x_{t_{i-1}})$$

(5) $$y_{\Delta'}(t_k) = \sum_{i=1}^{k} b_{\tau'_{i-1}}(x_{t_i} - x_{t_{i-1}}).$$

Here we choose $\delta(\epsilon,\eta)$ as in (11) of §4, and $d(\Delta)$, $d(\Delta')$ are $< \delta(\epsilon,\eta)$. Then, the quantities on the right hand side of (4) and (5) both represent a $\delta(\epsilon,\eta)$-sum. Now,

(6) $$y_\Delta(t_k) - y_\Delta'(t_k) = \sum_{i=1}^{k} (b_{\tau_{i-1}} - b_{\tau'_{i-1}})(x_{t_i} - x_{t_{i-1}})$$

and by defining c_i's as in (12) of §4 we obtain

(7) $$P\left\{\bigcup_{k=1}^{n}(y_\Delta(t_k) - y_\Delta'(t_k) \neq \sum_{i=1}^{k} c_i(x_{t_i} - x_{t_{i-1}}))\right\} < \eta.$$

We also have

(8) $$E\left\{\left[\sum_{i=1}^{n} c_i(x_{t_i} - x_{t_{i-1}})\right]^2\right\} < \epsilon^2.$$

Applying Lemma 6.1 by taking $c_i\,(x_{t_i} - x_{t_{i-1}})$ and $\sqrt{\epsilon}$ to be y_i, x_i and l, respectively, of that lemma, we obtain

(9) $$P\left\{\max_{1\leqslant k\leqslant n} \left| \sum_{i=1}^{k} c_i\,(x_{t_i} - x_{t_{i-1}}) \right| \geqslant \sqrt{\epsilon}\right\} \leqslant \epsilon.$$

From (7) and (9) it then follows that

(10) $$P\left\{\max_{1\leqslant k\leqslant n} | y_\Delta(t_k) - y_{\Delta'}(t_k) | \geqslant \sqrt{\epsilon}\right\} \leqslant \epsilon + \eta,$$

and therefore that

(11) $$P\left\{\max_{1\leqslant i\leqslant m} | y_\Delta(s_i) - y_{\Delta'}(s_i) | \geqslant \sqrt{\epsilon}\right\} \leqslant \epsilon + \eta.$$

As $m \to \infty$, the set within { } of the left-hand side above increases to the set

$$\left[\sup_{k=1}^{\infty} | y_\Delta(s_k) - y_{\Delta'}(s_k) | \geqslant \sqrt{\epsilon}\right].$$

Since $y_\Delta(\tau)$ and $y_{\Delta'}(\tau)$ are both continuous functions of τ, and since $\{s_i\}$ is dense in (0,1), we conclude that

$$\sup_{k=1}^{\infty} | y_\Delta(s_k) - y_{\Delta'}(s_k) | = \sup_{0\leqslant\tau\leqslant 1} | y_\Delta(\tau) - y_\Delta'(\tau) |,$$

and therefore we obtain

$$P\left\{\sup_{0\leqslant\tau\leqslant 1} | y_\Delta(\tau) - y_{\Delta'}(\tau) | \geqslant \sqrt{\epsilon}\right\} \leqslant \epsilon + \eta. \qquad \text{q.e.d.}$$

§7. Examples of Indefinite Integrals

Example 1.

$$\int_0^t x_\tau dx_\tau = \frac{1}{2}x_t^2 - \frac{1}{2}t.$$

Example 2.

$$\int_0^t x_\tau^2 dx_\tau = \frac{1}{3}x_t^3 - \int_0^t x_\tau d\tau.$$

Example 3.

$$\int_0^t a(x_\tau)dx_\tau = \int_0^{x_t} a(\lambda)d\lambda - \int_0^t \frac{a'(x_\tau)}{2}d\tau,$$

where we assume that $a'(\lambda)$ is continuous in λ.

Since Examples 1 and 2 are special cases of Example 3, we shall prove Example 3 only. Let us first of all define

(1) $$b(\xi) = \int_0^\xi a(\lambda)\,d\lambda.$$

Since $\{x_t\}$ is a process without moving discontinuity points, x_t is bounded in [0,1] with probability 1 (see the remark in Introduction (II)). Therefore, if we choose, for a given η, M sufficiently large, we have

(2) $$P\left\{\sup_{0\leqslant\tau\leqslant 1}|x_t| \leqslant M\right\} > 1 - \eta.$$

Since $a'(\xi)$ is continuous and hence is uniformly continuous on $|\xi| \leqslant M$, we have, if we choose δ sufficiently small for a given ϵ,

(3) $$|a'(\xi') - a'(\xi)| < \epsilon \quad \text{whenever } |\xi' - \xi| < \delta.$$

The fact that with probability 1 x_t is continuous implies also that for the δ as above, we have, if we choose γ sufficiently small,

(4) $$P\left[\bigcap_{\substack{|t-s|<\gamma \\ 0\leqslant t,s\leqslant 1}} (|x_t - x_s| < \delta)\right] > 1 - \eta.$$

Now, let $\Omega' = (\bigcap_{|t-s|<\gamma}(|x_t - x_s| < \delta))\cdot(\sup_{0\leqslant\tau\leqslant 1}|x_\tau| < M)$, then we see from (2) and (4) that

(4′) $$P(\Omega') > 1 - 2\eta.$$

The continuity of $a'(\lambda)$ implies that the function $b(\xi)$ has continuous derivatives up to the second order. Therefore,

(5) $$b(\xi') - b(\xi)$$
$$= b'(\xi)(\xi'-\xi) + \frac{b''(x)}{2}(\xi'-\xi)^2 + \frac{b''(\xi + \theta(\xi'-\xi)) - b''(\xi)}{2}(\xi'-\xi)^2$$
$$= a(\xi)(\xi'-\xi) + \frac{a'(\xi)}{2}(\xi'-\xi)^2 + \frac{a'(\xi+\theta(\xi'-\xi)) - a'(\xi)}{2}(\xi' - \xi)^2,$$

where θ lies in the interval [0,1]. Therefore,

(6) $$b(x_s) - b(x_t)$$
$$= a(x_t)(x_s - x_t) + \frac{a'(x_t)}{2}(x_s - x_t)^2 + \frac{a'(x_t + \theta(x_s - x_t)) - a'(x_t)}{2}(x_s - x_t)^2.$$

Whenever $|t-s| < \gamma$, we have on Ω'

$$|x_t + \theta(x_s - x_t)| \leqslant \max(|x_t|, |x_s|) < M,$$

and

$$|x_t + \theta(x_s - x_t) - x_t| \leqslant |x_s - x_t| < \delta.$$

Therefore, by (5) we get

(7) $$|a'(x_t + \theta(x_s - x_t)) - a'(x_t)| \leqslant \epsilon.$$

Let us now define $c_{t,s}$ by

(8) $$c_{t,s} \equiv \tfrac{1}{2}(a'(x_t + \theta(x_s - x_t)) - a'(x_t))$$

if the inequality (7) is satisfied, and

(9) $$c_{t,s} \equiv 0$$

otherwise. $a'(\xi)$, being continuous in ξ, is bounded on $|\xi| \leqslant M$; so, let $\sup_{|\xi|\leqslant M}|a'(\xi)| = R$.

Define e_t by

$$e_t = \begin{cases} a'(x_t) & \text{if } |a'(x_t)| \leqslant R \\ 0 & \text{otherwise.} \end{cases}$$

On the set Ω', $c_{t,s}$ is given by (8) and $e_t = a'(x_t)$; consequently, on Ω' we have, as long as $|t-s| < \gamma$,

(10) $$b(x_s) - b(x_t)$$
$$= a(x_t)(x_s - x_t) + \frac{a'(x_t)}{2}(s - t) + c_{t,s}(x_s - x_t)^2$$
$$+ \frac{1}{2}e_t((x_s - x_t)^2 - (s - t)).$$

Let us choose points $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$ such that $\max_{1\leqslant i\leqslant n}|t_i - t_{i-1}| < \gamma$, then we have on Ω'

(11) $$b(x_t) - b(x_0)$$
$$= \sum_{i=1}^{n}(b(x_{t_i}) - b(x_{t_{i-1}}))$$

$$= \sum_{i=1}^{n} a(x_{t_{i-1}})(x_{t_i} - x_{t_{i-1}}) + \sum_{i=1}^{n} \frac{a'(x_{t_{i-1}})}{2}(t_i - t_{i-1})$$

$$+ \sum_{i=1}^{n} c_{t_{i-1}t_i}(x_{t_i} - x_{t_{i-1}})^2 + \sum_{i=1}^{n} \frac{1}{2} e_{t_{i-1}}((x_{t_i} - x_{t_{i-1}})^2 - (t_i - t_{i-1})).$$

If we choose each $|t_i - t_{i-1}|$ sufficiently small (less than β, for example), we obtain

$$P\left\{ \left| \sum_{i=1}^{n} a(x_{t_{i-1}})(x_{t_i} - x_{t_{i-1}}) - \int_0^t a(x_\tau)dx_\tau \right| > \epsilon \right\} < \eta \tag{12}$$

$$P\left\{ \left| \sum_{i=1}^{n} \frac{a'(x_{t_{i-1}})}{2}(t_i - t_{i-1}) - \int_0^t \frac{a'(x_\tau)}{2} d\tau \right| > \epsilon \right\} < \eta \tag{13}$$

$$E\left(\left| \sum_{i=1}^{n} c_{t_{i-1}t_i}(x_{t_i} - x_{t_{i-1}})^2 \right| \right) \leqslant E\left(\sum_{i=1}^{n} \epsilon(x_{t_i} - x_{t_{i-1}})^2 \right) \leqslant \epsilon t \leqslant \epsilon$$

$$P\left\{ \left| \sum_{i=1}^{n} c_{t_{i-1}t_i}(x_{t_i} - x_{t_{i-1}})^2 \right| > \sqrt{\epsilon} \right\} < \sqrt{\epsilon}. \tag{14}$$

If we take each $|t_i - t_{i-1}|$ to be smaller than ϵ/R^2, we also obtain

$$E\left\{ \left(\sum_{i=1}^{n} \frac{e_{t_{i-1}}}{2}((x_{t_i} - x_{t_{i-1}})^2 - (t_i - t_{i-1}) \right)^2 \right\} \leqslant \sum_{i=1}^{n} \frac{R^2}{4} \left(2(t_i - t_{i-1})^2 \right)$$

$$\leqslant \frac{R^2}{4} \frac{\epsilon}{R^2} 2 \sum_{i=1}^{n} (t_i - t_{i-1}) \leqslant \frac{\epsilon}{2} t \leqslant \frac{\epsilon}{2}$$

and so,

$$P\left\{ \left| \sum_{i=1}^{n} \frac{e_{t_{i-1}}}{2}((x_{t_i} - x_{t_{i-1}})^2 - (t_i - t_{i-1})) \right| > \sqrt[4]{\epsilon} \right\} \leqslant \frac{\sqrt{\epsilon}}{2}. \tag{15}$$

From (4′) and (11)-(15), it follows that

$$P\left\{ \left| \int_0^{x_t} a(\lambda)d\lambda - \int_0^t a(x_\tau)dx_\tau - \int_0^t \frac{a'(x_\tau)}{2} d\tau \right| > 2\epsilon + \sqrt{\epsilon} + \sqrt[4]{\epsilon} \right\}$$

$$< 2\eta + 2\eta + \epsilon + \frac{\sqrt{\epsilon}}{2}.$$

Since ϵ and η are arbitrary, we have, for each fixed t,

$$\int_0^{x_t} a(\lambda)d\lambda = \int_0^t a(x_\tau)dx_\tau + \int_0^t \frac{a'(x_\tau)}{2} d\tau$$

and hence

$$\int_0^t a(x_\tau)dx_\tau = \int_0^{x_t} a(\lambda)d\lambda - \int_0^t \frac{a'(x_\tau)}{2} d\tau \tag{16}$$

holds with probability 1. The fact that both sides of (16) are stochastic processes having no moving discontinuity points implies that (16) is in fact valid for every t.

Example 4. When $a(\tau)$ and $b(\tau)$ are continuous functions of τ and x_τ is a Brownian motion,

$$y_t = \int_0^t a(\tau)d\tau + \int_0^t b(\tau)dx_\tau$$

is a differential process and $y_s - y_t$ has

$$G\left[\int_t^s a(\tau)d\tau, \sqrt{\int_t^s (b(\tau))^2 d\tau} \right]$$

distribution. Here $G(\alpha,\beta)$ denotes the Gaussian distribution with mean α and standard deviation β.

Proof. The fact that y_t is a differential process is clear. If we choose $\{\Delta_n\}$ suitably, we have

$$y_s - y_t = \lim_{n\to\infty} \left\{ \int_s^t a(\tau)d\tau + \sum_{\Delta_n} b(\tau_{i-1})(x_{t_i} - x_{t_{i-1}}) \right\}.$$

But the quantity within { } has a Gaussian distribution with its mean equal to $\int_s^t a(\tau)d\tau$ and its standard deviation equal to

$$\sqrt{\sum_i (b(\tau_{i-1}))^2 (t_i - t_{i-1})} \to \sqrt{\int_s^t (b(\tau))^2 d\tau}. \qquad \text{q.e.d.}$$

§8. An Inequality Concerning the Absolute Value of an Indefinite Integral

Theorem 8.1. *Let $\{x_t\}$ and $\{b_t\}$ be as in §4. We assume further that $E(b_t^2)$ is continuous in t. Then we have*

$$P\left(\sup_{0\leqslant t\leqslant 1} \left| \int_0^t b_\tau dx_\tau \right| \geqslant l \right) \leqslant \frac{\int_0^1 E(b_\tau^2)d\tau}{l^2}. \tag{1}$$

Remark. This theorem is nothing but Lemma 6.1 with the sum being replaced by the integral.

Proof of Theorem 8.1. Let us consider $y_\Delta(t)$ which was defined by (1) in the proof of Theorem 6.1. If we choose suitably $\{\Delta_n\}$ with $d(\Delta_n)\to 0$, $y_{\Delta_n}(t)$, which we denote simply by $y_n(t)$ in the sequel, converges uniformly to $\int_0^t b_\tau dx_\tau$ with probability 1.

Let α_1, a_2, $\cdots$ be a sequence of points dense in (0,1) and let us consider the partition $0 = s_0 < s_1 < s_2 \cdots < s_n = 1$ obtained by joining the points $\alpha_1, \alpha_2, \cdots, \alpha_k$ to the points of partition Δ_m. Then the refined representation of $y_m = y_{\Delta_m}$ with respect to $s_0 < s_1 < \cdots < s_n$ is given by

$$y_m(s_j) = \sum_{i=1}^{j} b_{\sigma_{i-1}}(x_{s_i} - x_{s_{i-1}}).$$

Now,

$$E\left\{\left[\sum_{i=1}^{n} b_{\sigma_{i-1}}(x_{s_i} - x_{s_{i-1}})\right]^2\right\} \leqslant \sum_{i=1}^{n} E(b^2_{\sigma_{i-1}})(s_i - s_{i-1})$$

$$= \sum_{\Delta_m} E(b^2_{\tau_{i-1}})(t_i - t_{i-1}).$$

By Lemma 6.1

$$P\left\{\max_{1 \leqslant j \leqslant n} |y_m(s_j)| \geqslant l\right\} \leqslant \frac{1}{l^2} \sum_{\Delta_m} E(b^2_{\tau_{i-1}})(t_i - t_{i-1}).$$

Consequently,

$$P\left\{\max_{1 \leqslant i \leqslant k} |y_m(\alpha_i)| \geqslant l\right\} \leqslant \frac{1}{l^2} \sum_{\Delta_m} E(b^2_{\tau_{i-1}})(t_i - t_{i-1}).$$

Noting that $y_m(\tau)$ has no moving discontinuity points, we obtain, by letting $k \to \infty$,

$$P\left\{\sup_{0 \leqslant \tau \leqslant 1} |y_m(\tau)| \geqslant l\right\} \leqslant \frac{1}{l^2} \sum_{\Delta_m} E(b^2_{\tau_{i-1}})(t_i - t_{i-1}).$$

Finally, by letting $m \to \infty$, we conclude that

$$P\left\{\sup_{0 \leqslant t \leqslant 1} \left|\int_0^t b_\tau dx_\tau\right| \geqslant l\right\} \leqslant \frac{1}{l^2} \int_0^t E(b_\tau^2) d\tau.$$

III. Differential Equations and Integral Equations

§9. We intend in this chapter to solve the differential equation

(1) $$Dy_t = G(a(t,y_t),\ b(t,y_t))$$

under the initial condition

(2) $$y_0 = c.$$

Here $G(\alpha,\beta)$ represents the Gaussian distribution with mean α and standard deviation β.

Let us first state the theorem.

Theorem 9.1. *Suppose $a(t,y)$ and $b(t,y)$ are both continuous in (t,y) and suppose further that there exist constants A and B such that*

(3) $$|a(t,y) - a(t,y')| \leqslant A\,|y - y'|$$

$$|b(t,y) - b(t,y')| \leqslant B\,|y - y'|$$

$$\text{for } 0 \leqslant t \leqslant 1,\ -\infty < y,\ y' < \infty.$$

Then, for a Brownian motion x_t, the equation

(4) $$y_t = c + \int_0^t a(\tau, y_\tau) d\tau + \int_0^t b(\tau, y_\tau) dx_\tau$$

has one and only one solution y_t and it satisfies the equation (1).

§10. Proof of the Existence of a Solution to the Integral Equation Given Above (the Method of Successive Approximations!)

Lemma 10.1. *Let $\{a_t\}$ be a stochastic process having no moving discontinuity points, and suppose that there exists a continuous function $M(t)$ satisfying $E(a_t^2) \leqslant M(t)$. Then, we have*

(1) $$E\left\{\left(\int_t^s a_\tau d\tau\right)^2\right\} \leqslant (s - t)\int_t^s M(\tau) d\tau.$$

Proof. We have

(2) $$\left(\int_t^s a_\tau d\tau\right)^2 \leqslant (s - t) \int_t^s a_\tau^2 d\tau$$

and

$$E\left\{\left(\int_t^s a_\tau d\tau\right)^2\right\} \leqslant (s - t)\, E\left\{\int_t^s a_\tau^2 d\tau\right\} \leqslant (s - t)\int_t^s M(\tau) d\tau,$$

where the last inequality can be proved by using Lemma 5.1.

Let us define $y_t^{(k)}$ $(k=1,2,\cdots)$ successively as follows:

(3) $$y_t^{(0)} = c$$

(4) $y_t^{(k)} = c + \int_0^t a(\tau, y_\tau^{(k-1)})d\tau + \int_0^t b(\tau, y_\tau^{(k-1)})dx_\tau \qquad (k=1,2,\cdots).$

First, we shall show that for each fixed t $(0 \leqslant t \leqslant 1)$ $y_t^{(k)}(k=1,2,\cdots)$ converges in the mean square.[13] We note that

$$y_t^{(1)} - y_t^{(0)} = \int_0^t a(\tau,c)d\tau + \int_0^t b(\tau,c)dx_\tau. \tag{5}$$

Since $a(\tau,c)$ and $b(\tau,c)$ are continuous in τ, their absolute values are dominated over the closed interval $0 \leqslant \tau \leqslant 1$ by some finite constant M, namely,

$$|a(\tau,c)| \leqslant M, \quad |b(\tau,c)| \leqslant M \qquad (0 \leqslant \tau \leqslant 1). \tag{6}$$

Consequently, by Lemma 10.1, we obtain from the first inequality of (6)

$$E\left\{ \left(\int_0^t a(\tau,c)d\tau\right)^2 \right\} \leqslant t\int_0^t M^2 d\tau \leqslant M^2t^2 \leqslant M^2 t \qquad (0 \leqslant t \leqslant 1). \tag{7}$$

By using Theorem 5.4, we obtain from the second inequality of (6)

$$E\left\{ \left(\int_0^t b(\tau,c)dx_\tau\right)^2 \right\} \leqslant \int_0^t M^2 d\tau = M^2 t. \tag{8}$$

From (5), (7), and (8), and utilizing

$$E\{(x+y)^2\} \leqslant \left\{\sqrt{E(x^2)} + \sqrt{E(y^2)}\right\}^2, \tag{9}$$ [14]

we conclude that

$$E\{(y_t^{(1)} - y_t^{(0)})^2\} \leqslant 4M^2 t. \tag{10}$$

Next, let us prove that for each n the following inequalities are valid:

$$E\left\{ \left(\int_0^t (a(\tau,y_\tau^{(n)}) - a(\tau,y_\tau^{(n-1)}))d\tau\right)^2 \right\} \leqslant 4M^2R^{2(n-1)}\frac{t^{n+1}}{(n+1)!}A^2, \tag{7'}$$

$$E\left\{ \left(\int_0^t (b(\tau,y_\tau^{(n)}) - b(\tau,y_\tau^{n-1})dx_\tau\right)^2 \right\} \leqslant 4M^2R^{2(n-1)}\frac{t^{n+1}}{(n+1)!}B^2, \tag{8'}$$

$$E\left\{ (y_t^{(n+1)} - y_t^{(n)})^2 \right\} \leqslant 4M^2R^{2n}\frac{t^{n+1}}{(n+1)!}, \quad \text{where } R \equiv A + B. \tag{10'}$$

[13] Cf. P. Lévy: *Ibid.* p. 52. 3° La convergence en moyenne. We consider the case of $d=2$ here. Namely, the set of all real-valued random variables x with $E(x^2) < \infty$ comprises a complete metric space with respect to the metric $\rho_m(x,y) = \sqrt{E((x-y)^2)}$. The convergence with respect to this metric is called the convergence in the mean square.

[14] This inequality means that the distance function ρ_m described in footnote (13) satisfies the triangle inequality.

If we define $a(\tau, y_\tau^{(-1)}) = 0$, $b(\tau, y_\tau^{(-1)}) = 0$, then (7′), (8′), and (10′) are all valid for $n = 0$ in view of (7), (8), and (10). The proof for the case of general n is done by induction.

So, let us suppose (7′), (8′), and (10′) are valid for $n-1$. Then, from (3) of §9 we obtain

$$|a(\tau,y_\tau^{(n)}) - a(\tau,y_\tau^{(n-1)})| \leqslant A\,|y_\tau^{(n)} - y_\tau^{(n-1)}|$$

$$E\left\{ \left(a(\tau,y_\tau^{(n)}) - a(\tau,y_\tau^{(n-1)})\right)^2 \right\} \leqslant A^2 E\left\{ \left(y_\tau^{(n)} - y_\tau^{(n-1)}\right)^2 \right\}$$
$$\leqslant 4M^2\,R^{2(n-1)}\frac{t^n}{n!}A^2.$$

By Lemma 10.1 we have (noting that $0 \leqslant t \leqslant 1$)

$$E\left\{ \left(\int_0^t (a(\tau,y_\tau^{(n)}) - a(\tau,y_\tau^{(n-1)}))d\tau\right)^2 \right\}$$
$$\leqslant \int_0^t 4M^2R^{2(n-1)}\frac{\tau^n}{n!}A^2 d\tau = 4M^2R^{2(n-1)}\frac{t^{n+1}}{(n+1)!}A^2,$$

which shows that (7′) is valid for n.

In the same way (8′) can be proved for n by using Theorem 5.4 in place of Lemma 10.1.

From (7′), (8′), and the following identity (*) we can deduce (10′) by using the inequality (9):

$$(*) \quad y_t^{(n+1)} - y_t^{(n)}$$
$$= \int_0^t (a(\tau,y_\tau^{(n)}) - a(\tau,y_\tau^{(n-1)}))d\tau + \int_0^t (b(\tau,y_\tau^{(n)}) - b(\tau,y_\tau^{(n-1)}))dx_\tau.$$

Now, since it follows from (10′) that

$$\rho_m(y_t^{(n+1)}, y_t^{(n)}) \leqslant \sqrt{4M^2R^{2n}\frac{t^{n+1}}{(n+1)!}}, \quad \text{where } \rho_m(x,y) = \sqrt{E((x-y)^2)},$$ [15]

[15] Cf. footnote (13).

and since

$$\sum_{n=0}^{\infty}\sqrt{4M^2R^{2n}\frac{t^{n+1}}{(n+1)!}} < \infty,$$

the sequence

(11) $$y_t^{(n)} = y_t^{(0)} + (y_t^{(1)} - y_t^{(0)}) + \cdots + (y_t^{(n)} - y_t^{(n-1)})$$

converges in the mean square. If we denote the limit by y_t, then $E(y_t^2) < \infty$ and y_t has a well-defined mean value and standard deviation.

Next we shall show that $\{y_t\}$ is a stochastic process without moving discontinuity points. Since $\{y_t^{(n)}\}$ $(n = 1,2, \cdots)$ is, by Lemma 6.1, a sequence of stochastic processes without moving discontinuity points, it suffices to show that the sequence $\{y_t^{(n)}\}$ converges in probability with respect to the strong topology (cf. the proof of Theorem 6.1 where similar arguments were used). First, letting $t = 1$ in (8′) and using Theorem 8.1, we obtain

(12) $$P\left\{\sup_{0\leqslant t\leqslant 1}\left|\int_0^t b(\tau,y_\tau^{(n)}) - b(\tau,y_\tau^{(n-1)})dx_\tau\right| > \frac{1}{2^n}\right\}$$
$$\leqslant 4M^2R^{2(n-1)}\frac{1}{(n+1)!}B^2 2^{2n} \qquad (n = 1, 2, \cdots).$$

Since the infinite series whose general term is given by the right hand side of (12) converges, the following statement holds with probability 1 because of the Borel-Cantelli Lemma: For all sufficiently large n

(13) $$\sup_{0\leqslant t\leqslant 1}\left|\int_0^t (b(\tau,y_\tau^{(n)}) - b(\tau,y_\tau^{(n-1)}))dx_\tau\right| \leqslant \frac{1}{2^n}.$$

Consequently, for any $\eta > 0$, we can choose $n_0(\eta)$ sufficiently large so that

(14) the inequality (13) is valid for all $n \geqslant n_0(\eta)$

everywhere except on a set of probability η. Furthermore, since $a(\tau,y_\tau^{(n_0)}) - a(\tau,y_\tau^{(n_0-1)})$ has no moving discontinuity points, we have for a sufficiently large $K = K(\eta)$,

(15) $$|a(\tau,y_\tau^{(n_0)}) - a(\tau,y_\tau^{(n_0-1)})| \leqslant K \qquad (0 \leqslant t \leqslant 1)$$

everywhere except on a set of probability η. If we denote by Ω' the set where (14) and (15) are valid simultaneously, we have

(16) $$P(\Omega') > 1 - 2\eta.$$

If we let $n = n_0$ in the identity (*) above, and use (14) and (15), we can conclude that

(17) $$|y_t^{(n_0+1)} - y_t^{(n_0)}| \leqslant Kt + \left(\frac{1}{2}\right)^{n_0} \qquad (0 \leqslant t \leqslant 1)$$

is valid on Ω'.

Next, we let $n = n_0 + 1$ in the identity (*) and use (14), (17), and (3) of §9 to obtain

$$|y_t^{(n_0+2)} - y_t^{(n_0+1)}| \leqslant \int_0^t A|y_t^{(n_0+1)} - y_t^{(n_0)}|\,dt + (1/2)^{n_0+1}$$
$$\leqslant AK\frac{t^2}{2} + (1/2)^{n_0}(tA) + (1/2)^{n_0+1}.$$

If we repeat the same procedure, we obtain

$$|y_t^{(n_0+m)} - y_t^{(n_0+m-1)}| \leqslant A^{m-1}K\frac{t^m}{m!} + \sum_{r=0}^{m-1}\left(\frac{1}{2}\right)^{n_0+r}\frac{(tA)^{m-1-r}}{(m-1-r)!}$$
$$= A^{m-1}K\frac{t^m}{m!} + \sum_{r=0}^{m-1}\left(\frac{1}{2}\right)^{n_0+m-1-r}\frac{(tA)^r}{r!}$$
$$= A^{m-1}K\frac{t^m}{m!} + \left(\frac{1}{2}\right)^{n_0+m-1}\sum_{r=0}^{m-1}\frac{(2tA)^r}{r!}$$
$$\leqslant A^{m-1}K\frac{1}{m!} + \left(\frac{1}{2}\right)^{n_0+m-1}e^{2A} \quad (m=1,2,3, \cdots).$$

Since the infinite series whose general term is given by the right-hand side above converges, we conclude that

$$\sum_{m=1}^{\infty}(y_t^{(n_0+m)} - y_t^{(n_0+m-1)})$$

converges uniformly in t on the set Ω'. Therefore, the sequence $\{y_t^{(n)}\}$ of stochastic processes converges in probability with respect to the strong topology, and thus its limit $\{y_t\}$ has no moving discontinuity points.

From the definition of $y_t^{(n)}$ it follows also that $y_t^{(n)}$ is a function of x_{0t}, and consequently, so are y_t, $a(t, y_t)$, and $b(t, y_t)$.

Equation (3) of §9 implies also that since y_t does not have any moving discontinuity points, $a(t,y_t)$ and $b(t,y_t)$ do not have them either.

Furthermore, from the fact that $\{y_t^{(n)}\}$ converges to $\{y_t\}$ in probability with respect to the strong topology it follows that $\int_0^t a(\tau, y_\tau^{(n)})d\tau$ converges in probability to $\int_0^t a(\tau, y_\tau)d\tau$, and by virtue of Theorem 5.5, $\int_0^t b(\tau, y_\tau^{(n)})dx_\tau$ converges in probability to $\int_0^t (b\tau, y_\tau)dx_\tau$. Consequently, y_t satisfies the integral equation (4) of §9 in view of equation (4).

§11. Proof of the Uniqueness of the Solution of the Integral Equation of §9

1^0. Let us suppose that there are two solutions y'_t and y''_t for the case when $|a(\tau,y)|$ and $|b(\tau,y)|$ are bounded (say by M) in $0 \leqslant \tau \leqslant 1$, $-\infty < y < \infty$. Then,

$$(1)\quad y'_t - y''_t = \int_0^t (a(\tau,y'_\tau) - a(\tau,y''_\tau))d\tau + \int_0^t (b(\tau,y'_\tau) - b(\tau,y''_\tau))dx_\tau$$

$$E\left\{ \left(\int_0^t (a(\tau,y'_\tau) - a(\tau,y''_\tau))d\tau\right)^2 \right\} \leqslant \left(\int_0^t M d\tau\right)^2 \leqslant M^2t^2 < \infty.$$

By Theorem 5.4, we also have

$$E\left\{ \left(\int_0^t (b(\tau,y'_\tau) - b(\tau,y''_\tau))dx_\tau\right)^2 \right\} \leqslant \int_0^t M^2 d\tau = M^2 t < \infty.$$

From (1) and the two preceding inequalities we obtain, in view of $E((x+y)^2) \leqslant (\sqrt{E(x^2)} + \sqrt{E(y^2)})^2$,

$$(2)\quad E\{(y'_t - y''_t)^2\} \leqslant (Mt + M\sqrt{t})^2 \leqslant 4M^2 \qquad (0 \leqslant t \leqslant 1)$$

Now, if $E\{(y'_t - y''_t)^2\} \leqslant K(t)$, it then follows from (3) of §9, Lemma 10.1, and Theorem 5.4 that

$$E\left\{ \left(\int_0^t (a(\tau,y'_\tau) - a(\tau,y''_\tau))d\tau\right)^2 \right\} \leqslant t\int_0^t A^2K(\tau)d\tau \leqslant A^2\int_0^t K(\tau)d\tau,$$

$$E\left\{ \left(\int_0^t (b(\tau,y'_\tau) - b(\tau,y''_\tau))dx_\tau\right)^2 \right\} \leqslant \int_0^t B^2K(\tau)d\tau = B^2\int_0^t K(\tau)d\tau.$$

From (1) and the preceding two inequalities we obtain

$$(3)\quad E\{(y'_t - y''_t)^2\} \leqslant (A + B)^2\int_0^t K(\tau)d\tau,$$

and from (2) and (3) we conclude that

$$(4)\quad E\{(y'_t - y''_t)^2\} \leqslant (A + B)^2 4M^2 t.$$

By using (4) and (5) again, we obtain

$$E\{(y'_t - y''_t)^2\} \leqslant (A + B)^4 4M^2 \frac{t^2}{2},$$

and repeating, we deduce that

$$(5)\quad E\{(y'_t - y''_t)^2\} \leqslant 4M^2\frac{((A+B)^2t)^n}{n!} \to 0 \quad \text{as } n \to \infty.$$

Therefore, $E\{(y'_t - y''_t)^2\} = 0$ and hence

$$P(y'_t = y''_t) = 1.$$

2^0. **The General Case.** Since both $|a(t,c)|$ and $|b(t,c)|$ are continuous functions of t they are bounded by some finite number α on $0 \leqslant t \leqslant 1$. Therefore,

$$(6)\quad |a(t,y)| \leqslant |a(t,c)| + |a(t,y) - a(t,c)| \leqslant \alpha + A|y - c|$$

$$(7)\quad |b(t,y)| \leqslant \alpha + B|y - c|.$$

Let us define Ω_K by

$$(8)\quad \Omega_K = \left(\sup_{0\leqslant\tau\leqslant 1}|y'_\tau - c| < K\right)\cdot\left(\sup_{0\leqslant\tau\leqslant 1}|y''_\tau - c| < K\right).$$

Then, Ω_K increases as K increases, and

$$(9)\quad P(\Omega_K) \to 1 \quad \text{as } K \to \infty.$$

Furthermore, on Ω_K we have

$$(10)\quad |a(t,y_t)|,\ |b(t,y_t)| \leqslant 2\alpha + (A+B)K \qquad (0 \leqslant t \leqslant 1).$$

Denote the quantity on the right-hand side above by M, and define $b_M(t,y)$ as follows:

$$(11)\quad \begin{aligned} b_M(t,y) &= M && \text{if } b(t,y) > M \\ &= b(t,y) && \text{if } -M \leqslant b(t,y) \leqslant M \\ &= -M && \text{if } b(t,y) < -M. \end{aligned}$$

Let us now suppose that there are two distinct solutions, and call them y'_t and y''_t. Then,

$$(12)\quad P(y'_t \neq y''_t) > 0.$$

If we can show that (12) leads to a contradiction, then we shall have the uniqueness of the solution. Now,

$$P(\Omega_K \cdot (y'_t \neq y''_t)) \geqslant P(y'_t \neq y''_t) - (1 - P(\Omega_K)). \tag{13}$$

Hence, from (9), (12), and (13) it follows that

$$P(\Omega_K \cdot (y'_t \neq y''_t)) > 0 \tag{14}$$

holds for sufficiently large K. But on the set Ω_K we have

$$b_M(t,y_t) = b(t,y_t),\ a_M(t,y_t) = a(t,y_t) \qquad (0 \leqslant t \leqslant 1),$$

and therefore, both y'_t and y''_t satisfy on Ω_K the equation

$$y_t = c + \int_0^t a_M(\tau,y_\tau)d\tau + \int_0^t b_M(\tau,y_\tau)dx_\tau. \tag{15}$$

By definition, $a_M(\tau,y_\tau)$ and $b_M(\tau,y_\tau)$ satisfy condition (3) of §9 and their absolute values are dominated by M. Therefore, by what was already proved in 1^0, we see that the solution of (15) is unique. If we call this solution $y_t^{(M)}$, then on Ω_K we have $y_t^{(M)} = y'_t = y''_t$ except possibly on a subset of probability 0. This says that $P\{\Omega_K \cdot (y'_t \neq y''_t)\} = 0$, which contradicts (14).

§12. Proof That the Solution to the Integral Equation of § 9 Is a Markoff Process

Denote by $z_s = f(t, s, \eta, \tilde{x}_{ts})$ the solution of

$$z_s = \eta + \int_t^s a(\tau,z_\tau)d\tau + \int_t^s b(\tau,z_\tau)dx_\tau. \tag{1}$$

Here, $\tilde{x}_{ts} = (x_\tau - x_t;\ t \leqslant \tau \leqslant s)$.

Denote by $\tilde{N}_{ts}(\eta)$ the set of ω for which f above is not well-defined. This set is described by some conditions on $\tilde{x}_{ts}$. Let y_s be the solution obtained in the preceding section, and let

$$\begin{aligned} y'_s &= y_s && \text{for } 0 \leqslant s \leqslant t \\ &= f(t, s, y_t, \tilde{x}_{ts}) && \text{for } s > t\,. \end{aligned} \tag{2}$$

Now, $f(t, s, y_t, \tilde{x}_{ts})$ is not well-defined either when

$$y_t \text{ is not well-defined} \tag{3}$$

or when

$$\omega \text{ belongs to } \bigcup_\eta (y_t = \eta)(\tilde{x}_{ts} \in \tilde{N}_{ts}(\eta)). \tag{4}$$

But the probability of the set described in (3) is obviously 0, while the probability of the set in (4) is given by

$$\int_{-\infty}^{\infty} F_{y_t}(d\eta)P(\tilde{N}_{ts}(y_t) \mid y_t = \eta), \tag{5}$$

where F_y denotes the probability distribution function of y. Since y_t is a function of x_{0t}, $\tilde{x}_{ts}$ and y_t are independent. Consequently, $P(\tilde{N}_{ts}(y_t) \mid y_t = \eta) = P(\tilde{N}_{ts}(\eta)) = 0$. This implies that the integral (5) is also equal to 0, and hence the probability of the set in (4) is 0.

If $s \leqslant t$, we have

$$\begin{aligned} y'_t = y_s &= c + \int_0^s a(\tau,y_\tau)d\tau + \int_0^s b(\tau,y_\tau)dx_\tau \\ &= c + \int_0^s a(\tau,y'_\tau)d\tau + \int_0^s b(\tau,y'_\tau)dx_\tau, \end{aligned} \tag{6}$$

while if $s > t$, we have

$$\begin{aligned} y'_s &= f(t, s, y_t, \tilde{x}_{ts}) \\ &= y_t + \int_t^s a(\tau,y'_\tau)d\tau + \int_t^s b(\tau, y'_\tau)dx_\tau \quad \text{(Definition of } f\text{!).} \end{aligned} \tag{7}$$

Substituting into (7) the equation (6) with $s = t$, we obtain

$$y'_s = c + \int_0^s a(\tau,y'_\tau)d\tau + \int_0^s b(\tau,y'_\tau)dx_\tau. \tag{8}$$

Therefore, $\{y'_s\}$ coincides with the solution $\{y_s\}$ of the integral equation of §9, and this implies that $y_s = f(t, s, y_t, \tilde{x}_{ts})$, so

$$F_{y_s|x_{0t}=\xi_{0t}} = F_{f(t,s,y_t,\tilde{x}_{ts})|x_{0t}=\xi_{0t}} = F_{f(t,s,y_t(\xi_{0t}),\tilde{x}_{ts})}$$

(we obtain the last identity since $f(t,s,y_t(\xi_{0t}), \tilde{x}_{ts})$ is a function of $\tilde{x}_{ts}$ and hence is independent of x_{0t}). This says that $F_{y_s|x_{0t}=\xi_{0t}}$ depends only on $y_t(\xi_{0t})$. Since y_{0t} is a function of x_{0t}, $F_{y_s|y_{0t}=\eta_{0t}}$ also depends only on η_t.

§13. Proof That the Solution of the Integral Equation of §9 Satisfies the Differential Equation (1) of §9

Let

$$y_t = c + \int_0^t a(\tau,y_\tau)d\tau + \int_0^t b(\tau,y_\tau)dx_\tau. \tag{1}$$

If we consider the conditional probability distribution of the solution y_t of equation (1) for $t \geqslant t_0$ given that $y_{t_0} = \eta$, then from what was discussed in the beginning of the preceding section it follows that this conditional probability distribution coincides with the (unconditional) probability distribution of the solution for the equation

$$y_t = \eta + \int_{t_0}^t a(\tau,y_\tau)d\tau + \int_{t_0}^t b(\tau,y_\tau)dx_\tau \qquad (t \geqslant t_0). \tag{2}$$

Now,

$$\begin{aligned} y_t - y_{t_0} &= y_t - \eta \\ &= \int_{t_0}^{t} a(\tau, y_\tau)d\tau + \int_{t_0}^{t} b(\tau, y_\tau)dx_\tau \\ &= \int_{t_0}^{t} a(\tau, y_{t_0})d\tau + \int_{t_0}^{t} b(\tau, y_{t_0})dx_\tau \\ &\quad + \int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau \\ &\quad + \int_{t_0}^{t}(b(\tau, y_\tau) - b(\tau, y_{t_0}))dx_\tau \end{aligned}$$

and hence

$$(2') \qquad y_t - y_{t_0} = \int_{t_0}^{t} a(\tau, y_{t_0})d\tau + \int_{t_0}^{t} b(\tau, y_{t_0})dx_\tau$$

$$+ \int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau + \sum_{\Delta}^{t}(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))(x_{t_i} - x_{t_{i-1}})$$

$$+ \left\{ \int_{t_0}^{t}(b(\tau, y_\tau) - b(\tau, y_{t_0}))dx_\tau - \sum_{\Delta}^{t}(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))(x_{t_i} - x_{t_{i-1}}) \right\}$$

where the sum $\sum_{\Delta}^{t}$ is defined just as $y_\Delta(t)$ was defined in (1) of §6. If we denote by γ_t the quantity appearing inside of { } in (2)′, and if we choose $d(\Delta)$ sufficiently small, then

$$(3) \qquad P\{|\gamma_t| \geqslant \epsilon(t - t_0)\} < \epsilon(t - t_0).$$

Now, since the solution of (1) should be obtained by the method of successive approximations of §10, letting M be the maximum value for $s_0 \leqslant \tau \leqslant 1$ of the quantities $|a(\tau, \eta)| \cdot |b(\tau, \eta)|$, we see that

$$E\{(y_t - \eta)^2\} \leqslant \left[\sum_{n=0}^{\infty} \sqrt{4M^2R^{2n}\frac{(t - t_0)^{n+1}}{(n+1)!}} \right]^2, \quad \text{where } R = A + B$$

$$\leqslant K(t - t_0), \quad \text{where } K \equiv \sum_{n=0}^{\infty} \sqrt{4M^2R^{2n}\frac{1}{(n+1)!}} < \infty,$$

namely, $E\{(y_t - y_{t_0})^2\} \leqslant K(t - t_0)$. Therefore

$$\begin{aligned} E\{(a(\tau, y_\tau) - a(\tau, y_{t_0}))^2\} &\leqslant A^2E\{(y_t - y_{t_0})^2\} \\ &\leqslant A^2K\,(t - t_0), \end{aligned}$$

from which it follows that

$$(4) \quad E\left\{ \left[\int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau \right]^2 \right\}$$

$$\leqslant (t - t_0) \int_{t_0}^{t} KA^2(t - t_0)dt = \frac{A^2K}{2}(t - t_0)^3 = o(t - t_0).$$

We also have

$$(5) \quad \left| E\left\{ \int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau \right\} \right|$$

$$\leqslant \sqrt{E\left\{ \int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau)^2 \right\}}$$

$$\leqslant \frac{\sqrt{A^2K}}{2}(t - t_0)^{3/2} = o(t - t_0).$$

From (4) and (5) we obtain

$$(5') \qquad \sigma\left\{ \int_{t_0}^{t}(a(\tau, y_\tau) - a(\tau, y_{t_0}))d\tau \right\} = o(\sqrt{t - t_0}).$$

Next, from $E\{(y_t - y_{t_0})^2\} \leqslant K(t - t_0)$ we obtain

$$E\{(b(t, y_t) - b(t, y_{t_0}))^2\} \leqslant B^2K(t - t_0).$$

Therefore, the mean of $b(t, y_t) - b(t, y_{t_0})$ exists also, and

$$(6) \quad E\left\{ \sum_{\Delta}^{t}(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))(x_{t_i} - x_{t_{i-1}}) \right\}$$

$$= \sum_{\Delta}^{t} E(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))E(x_{t_i} - x_{t_{i-1}})$$

$$= 0;$$

$$(7) \quad E\left\{ \left[\sum_{\Delta}^{t}(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))(x_{t_i} - x_{t_{i-1}}) \right]^2 \right\}$$

$$\leqslant \sum_{\Delta}^{t} B^2K(\tau_i - t_0)(t_i - t_{i-1})$$

$$\leqslant \int_{t_0}^{t} B^2K(\tau - t_0)d\tau \quad (\text{note that } t_0 \leqslant \tau_i \leqslant t_{i-1} \leqslant t_i)$$

$$= \frac{B^2K}{2}(t - t_0)^2 = o(t - t_0).$$

From (6) and (7) we obtain that

$$(7') \qquad \sigma\left(\sum_{\Delta}^{t}(b(\tau_{i-1}, y_{\tau_{i-1}}) - b(\tau_{i-1}, y_{t_0}))(x_{t_i} - x_{t_{i-1}})\right) = o(\sqrt{t-t_0}).$$

From (2′), (3), (4), (5′), (6), and (7′) it follows that Dy_{t_0} equals the derivative Dz_{t_0} of

$$(8) \qquad z_t = y_{t_0} + \int_{t_0}^{t} a(\tau, y_{t_0})d\tau + \int_{t_0}^{t} b(\tau, y_{t_0})dx_\tau.$$

But

$$F_{z_t - z_{t_0}} = G\left(\int_{t_0}^{t} a(\tau, y_{t_0})d\tau, \sqrt{\int_{t_0}^{t}(b(\tau, y_{t_0}))^2 d\tau}\right).$$

Therefore,

$$F_{z_t - z_{t_0}}^{*[1/t-t_0]} = G\left(\left[\frac{1}{t-t_0}\right]\int_{t_0}^{t} a(\tau, y_{t_0})d\tau, \sqrt{\left[\frac{1}{t-t_0}\right]\int_{t_0}^{t}(b(\tau, y_{t_0}))^2 d\tau}\right)$$

Now,

$$\lim_{t \to t_0}\left[\frac{1}{t-t_0}\right]\int_{t_0}^{t} a(\tau, y_{t_0})d\tau$$

$$= \lim_{t \to t_0}\frac{1}{t-t_0}\int_{t_0}^{t} a(\tau, y_{t_0})d\tau = a(t_0, y_{t_0}).$$

Similarly,

$$\lim_{t \to t_0}\left[\frac{1}{t-t_0}\right]\int_{t_0}^{t}(b(\tau, y_{t_0}))^2 d\tau = (b(t_0, y_{t_0}))^2.$$

Therefore,

$$Dy_{t_0} = Dz_{t_0} = G(a(t_0, y_{t_0}), b(t_0, y_{t_0})). \qquad \text{q.e.d.}$$

References

[1] A. Kolmogoroff: Grundbegriffe der Wahrscheinlichkeitsrechnung, p. 12.

[2] M. Fréchet: Recherches théoriques modernes sur la théorie des probabilités, tom 2, p. 215.

[3] A. Kolmogoroff: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung (*Math. Ann.* **104**, p. 415).

[4] W. Feller: Zur Theorie der stochastischen Prozesse (Existenz und Eindeutigkeitssätze) (*Math. Ann.* **113**, p. 113).

[5] J. L. Doob: Stochastic processes depending on a continuous parameter (*Trans. Am. Math. Soc.* **42** (1937)); Stochastic processes with an integral-valued parameter (*ibid* **44** (1938)).

[6] Article #1033 (this journal, vol **234**) §2.

[7] Article #1033 (this journal, vol **234**) §4.