Chapter 5 Stochastic Process

Stochastic process is a mathematical concept for expressing or describing random phenomena that vary with time, and it is defined as a system of random variables

$$\{X_t(\omega), t \in T\} \qquad \big(\omega \in (\Omega, P)\big).$$

Here, T is called the **time domain**, and it represents the set of observation times. Usually T is taken to be an interval of real numbers, though it can also be a discrete set of real numbers (for example $\boldsymbol{N}$). The former is a stochastic process of **continuous-time** variable, while the latter is a stochastic process of **discrete-time** variable. When $X_t(\omega)$ is a realization of a sample point ω, it is the observed value at time t, and $X_t(\omega)$ manifests itself as a point in a topological space E, which is known as the **state space**. Usually E is taken to be $\boldsymbol{R}^n$ $(n = 1, 2, \cdots)$, but it can be any topological space. When $\{X_t(\omega), t \in T\}$ is called the stochastic process, it is assumed that

> "for all t, $X_t(\omega)$ is an E-valued topological random variable $\big($ i.e., $X_t : \Omega \to E$ is measurable $\mathcal{D}(P)/\mathcal{B}(E)\big)$".

Henceforth, unless otherwise stated,

$$T \text{ is an interval,} \qquad E \text{ is } \boldsymbol{R}^1.$$

Oftentimes, stochastic process is called simply as a **process**.

Stochastic process is analogous to function in mathematical analysis; it is the most important concept in probability theory. While functions such as measurable function, continuous function, and analytic function are generic concepts in mathematical analysis, concepts such as martingale, additive process, Markov process, and stationary process are analogously generic in probability theory. Moreover, similar to index function and Bessel function being the examples of special functions in analysis, there are also special processes in probability theory such as Wiener process and Poisson process. But if one were to do a direct translation of the classification of functions or the form of differential equation from mathematical analysis, then one would not obtain something interesting in probability theory. The intuitive background of stochastic process is the temporal variation of random phenomena. By having the intuitive background in mind while developing a theory, the resulting theory will generate something not only uniquely interesting in probability theory but also useful.

§ 5.1 Function Space C and D

To begin and as a preparation, I will explain Polish space and Kolmogorov σ additive family.

Let ρ be the metric (function) on S. When the metric space (S, ρ) is complete[1], ρ is said to be a **complete metric** on S, and when (S, ρ) is separable[2], ρ is said to be a **separable metric**. The topology on S that is determined by a complete and separable metric on S is named the **Polish topology**. When τ is the metric on S that gives rise to the Polish topology on S, (S, τ) is known as the **Polish space**. Though Polish space is a topological space, one can treat it as a complete separable metric space that is equipped with a complete and separable metric, which determines its topology. A countable set with a discrete topology and $\boldsymbol{R}^n$ $(n = 1, 2, \cdots, \infty)$ with an ordinary topology are Polish spaces. The topological product of countable Polish spaces is itself a Polish space.

As mentioned previously, the family of Borel sets of topological space $S = (S, \tau)$ is denoted by $\mathcal{B}(S)$. When the topology τ is to be explicitly shown, it is denoted by $\mathcal{B}_\tau(S)$.

Theorem 5.1 The Polish topology τ on S and the Hausdorff topology τ_1 are given. If τ_1 is weaker than τ, then

$$\mathcal{B}_\tau(S) = \mathcal{B}_{\tau_1}(S).$$

Because the proof of this theorem requires the knowledge of analytic set theory, it is omitted.

The totality of all real-valued functions on set T is denoted as $\boldsymbol{R}^T$. An arbitrary subset of $\boldsymbol{R}^T$ is called the **function space** on T. With respect to the function space F on T, the map

$$\pi_t : \; F \longrightarrow \boldsymbol{R}^1 \,, \quad f \longmapsto f(t)$$

is called the image at t.

When F is the function space on T, $\pi_t^{-1}(\mathcal{B}^1)$ is the σ additive family on F. Write the topological product space as

$$\mathcal{B}_K(F) = \bigvee_{t \in T} \pi_t^{-1}(\mathcal{B}^1) \equiv \sup_{t \in T} \pi_t^{-1}(\mathcal{B}^1) \,,$$

and $\mathcal{B}_K(F)$ is, for all t, the smallest in the σ additive family $\mathcal{B}$ for which it is π_t measurable. $\mathcal{B}_K(F)$ is called the **Kolmogorov $\boldsymbol{\sigma}$ additive family** on F.

Henceforth through this section, T is taken to be an interval of real numbers. Of the function spaces on T, those of particularly important to the theory of stochastic processes are

$$\begin{aligned} C = C(T) \;&=\; \text{the totality of all continuous functions on } T \,, \\ D = D(T) \;&=\; \text{the totality of all discontinuous functions of the first kind on } T \,. \end{aligned}$$

[1]Translator's note: A Cauchy sequence in the metric space is a sequence whose members are arbitrarily close to each other as the sequence progresses. More precisely, if for every positive real number $\epsilon > 0$, there is a positive integer N such that for all natural numbers $m, n > N$, the distance between the m-th member and the n-th member in the sequence is less than ϵ. The metric space S for which every Cauchy sequence has a limit in S is called complete.

[2]Translator's note: A topological space is said to be separable if it contains a countable dense subset; that is, there exists a sequence $\{x_i\}_{i=1}^{\infty}$ of elements of the space such that every nonempty open subset of the space contains at least one element of the sequence.

If all the points of T (less the rightmost edge) are right continuous and if left limits exist for all the points of T (less the leftmost edge), then $f :\to \boldsymbol{R}^1$ is **discontinuous of the first kind**.

$C = C(T)$ is a well known space in functional analysis. When T is compact[3], C is topologicalized by the metric

$$\rho_u(f, g) = \sup_{t \in T} |f(t) - g(t)| \, .$$

We denote this topology by τ_u. Since

$$\rho_u(f_n, f) \longrightarrow 0 \Longleftrightarrow f_n \text{ is uniformly convergent to } f \, ,$$

τ_u is said to be a **uniformly convergent topology**. As ρ_u is a complete separable metric on C, τ_u is a Polish topology. When T is not compact, it may be casted as

$$T = \bigcup_n T_n \, , \qquad T_n \text{ is a compact space, and } T_1 \subset T_2 \subset \cdots$$

If we let

$$\rho_u(f, g) = \sum_{n=1}^{\infty} 2^{-n} \left(\sup_{t \in T_n} |f(t) - g(t)| \wedge 1 \right) \, ,$$

then ρ_u is a complete separable metric on C, and it determines the Polish topology τ_u. Since

$$\rho_u(f_n, f) \longrightarrow 0 \Longleftrightarrow f_n \text{ is uniformly convergent to } f \text{ in the generalized sense} \, ,$$

τ_u is said to be a **generalized uniformly convergent topology**. Henceforth, unless otherwise stated, C is taken to be a Polish space topologicalized by τ_u.

Though $D = D(T)$ might not be the object of study in functional analysis, it is as important as C in probability theory. On D, ρ_u and τ_u could be defined in the same way as on C, but this definition is not appropriate. First, ρ_u is a complete metric but not separable. Accordingly, τ_u does not possess countable open support. As Polish topology has countable open support, τ_u is not a Polish topology. The following case shows that ρ_u is not separable. For example, if we let

$$T = [0, 1] \, , \qquad f_\alpha = 1_{[0,\alpha)} \quad (\alpha \in T) \, ,$$

then $\{f_\alpha, \alpha \in T\}$ is an uncountable subset in D,

$$\rho_u(f_\alpha, f_\beta) = 1 \qquad (\alpha \neq \beta)$$

Consequently, since "α and β are close to each other but f_α and f_β are not close with respect to τ_u," it is not possible to say that τ_u as a topology is a natural object on D.

[3]Translator's note: An open cover of a space X is a collection of open subsets such that X is the union of all open subsets in the collection. A space T is said to be compact if for every open cover of T there exists a finite subcover of T. Intuitively, if one takes an infinite number of "steps" in the space, eventually one must get arbitrarily close to some other point of the space.

It would be ideal to introduce a topology that is more natural than τ_u, and with respect to this topology, D becomes a Polish space. Such topology was proposed by A. V. Skorohod, and thus it is named as the Skorohod topology. Roughly speaking, "f and g are close under Skorohod topology," means "by changing slightly the way the t of $f(t)$ is measured, $f(t)$ becomes uniformly close to $g(t)$ in the general sense." The more precise definition is as follows. Let Φ be the totality of all order-preserving topological mappings that map T onto T. Under the operation of mapping convolution, Φ is a group. The identity map i is the unit element of Φ. The inverse map φ^{-1} is the inverse element of Φ. Clearly,

$$\Phi \subset C \subset D\,.$$

Moreover, for $f, g \in D$,

$$\rho_u(f,g) = \rho_u(f \circ \varphi, g \circ \varphi) \qquad (\varphi \in \Phi)$$

is valid. Now, let

$$\rho_S(f,g) = \inf_{\varphi \in \Phi} \Big(\rho_u(f \circ g) + \rho_u(\varphi, i) \Big)\,,$$

and ρ_S becomes the metric on D. The topology τ_S determined by ρ_S is called the **Skorohod topology**.

In the defining formula of $\rho_S(f,g)$, let $\varphi = i$ to obtain

$$\rho_S(f,g) \le \rho_u(f,g)\,.$$

Hence, τ_S is a topology weaker than τ_u. Furthermore, if $\varphi_{\alpha\beta} \in \Phi$ is such that

$$\varphi_{\alpha\beta}(\beta) = \text{respectively linear on } [0,\alpha], [\alpha,\beta], [\beta,1]\,,$$

then in the above example, since

$$\rho_S(f_\alpha, f_\beta) \le \rho_u(\varphi_{\alpha\beta}, i) \le |\alpha - \beta|$$

is obtained, when α is close to β, f_α and f_β are close under τ_S.

Though ρ_S is separable, it is not complete (see Example 5.1 below). But τ_S is a Polish topology. To show this, it is sufficient to define a complete separable metric for which ρ_S is topologically equivalent. The following Billingsley's metric ρ_B has this property. To make the essential point clear, we explain with compact T being compact. For $\varphi \in \Phi$, let

$$\lambda(\phi) = \sup_{t \ne s} \left| \log \frac{\varphi(t) - \varphi(s)}{t - s} \right|\,.$$

To say that $\lambda(\varphi)$ is small, we mean that the gradient of $\phi(t)$ is uniformly close to 1. This is a condition stronger than "uniformly close to t."

$$\Psi = \big\{ \varphi \in \Phi \big| \lambda(\varphi) < \infty \big\}$$

is a subgroup of Φ, and

$$\lambda(\varphi^{-1}) = \lambda(\varphi)\,, \qquad \lambda(\varphi \circ \psi) \le \lambda(\varphi) + \lambda(\psi)$$

Now, let

$$\rho_B(f,g) = \inf_{\psi \in \Psi} \big(\rho_u(f \circ \psi, g) + \lambda(\psi)\big)$$

and ρ_B becomes a complete metric on D, and gives a topology τ_S same as ρ_S does. The proof of this statement is technically complicated and it is omitted.

Henceforth, $D = D(T)$ is topologicalized as τ_S, and is deemed as a Polish space.

Theorem 5.2

(i) $\mathcal{B}(C) = \mathcal{B}_K(C)$

(ii) $\mathcal{B}(D) = \mathcal{B}_K(D)$.

Proof Though T is assumed to be compact, the proof is essentially the same when T is generalized.

(i) The map $\pi_t : C \to \boldsymbol{R}^1$ is continuous and of course measurable $\mathcal{B}(C)$. Accordingly

$$\mathcal{B}_K(C) \subset \mathcal{B}(C)\,.$$

It is said to be the reverse containing relationship. Fix an arbitrary $g \in C$ and let

$$F_g(f) = \rho_u(f,g) \equiv \sup_{t \in T} |f(t) - g(t)|\,.$$

From the continuity of f, g,

$$F_g(f) = \sup_{t \in T \cap \boldsymbol{Q}} |f(t) - g(t)| = \sup_{t \in T \cap \boldsymbol{Q}} |\pi_t(f) - \pi_t(g)|\,.$$

As π_t is measurable $B_K(C)$, so is $F_g : C \to R^1$. Hence

$$U(g,r) = \{f \in C | \rho_u(f,g) < r\} = \{f \in C | F_g(f) < r\} \in \mathcal{B}_K(C)$$

As ρ_u is separable, $\mathcal{B}(C)$ is generated by $U(g,r)$, $g \in C$, $r > 0$. In this way, we obtain $\mathcal{B}(C) \subset \mathcal{B}_K(C)$, and in the end we know that $\mathcal{B}(C)$ and $\mathcal{B}_K(C)$ coincide.

(ii) As T is compact, the elements of $D = D(T)$ are bounded and Borel measurable (hence integrable) functions.

$$\rho_\lambda(f,g) = \int_T |f(t) - g(t)|\, dt$$

determines the Hausdorff topology[4] on D. We denote this topology by τ_λ, and the family of Borel sets on D with respect to this topology τ_λ by $B_\lambda(D)$. If we could have

$$\mathcal{B}(D) = \mathcal{B}_\lambda(D)\,, \quad \text{and} \quad \mathcal{B}_\lambda(D) = \mathcal{B}_K(D)\,,$$

[4]Translator's note: Two points x and y in a topological space X can be separated by neighborhoods if there exists a neighborhood U of x and a neighborhood V of y such that U and V are disjoint, i.e., $U \cap V = \emptyset$. All metric spaces are Hausdorff.

then we would have $\mathcal{B}(D) = \mathcal{B}_K(D)$ and end of proof.

First, we show that $\mathcal{B}(D) = \mathcal{B}_\lambda(D)$. Since τ_S is Polish topology, from Theorem 5.1, if we could say that τ_λ is weaker than τ_S, i.e., if we could say that

$$\rho_S(f_n\,, f) \longrightarrow 0 \Longrightarrow \rho_\lambda(f_n\,, f) \longrightarrow 0\,,$$

then it would be sufficient. From the assumption, there exists $\varphi_n \in \Phi$, and we have

$$\rho_u(f_n\,, f \circ \varphi_n) \longrightarrow 0\,, \quad \rho_n(\varphi_n\,, i) \longrightarrow 0\,.$$

Therefore, with regard to t in a uniform fashion

$$f_n(t) - f\big(\varphi_n(t)\big) \longrightarrow 0\,, \quad \varphi_n(t) \longrightarrow t\,.$$

Since $f(\in D)$ is bounded function, from the first equation above, $\{f_n\}$ is a uniformly bounded function series. Moreover, from the second equation, at the continuous point t of f, we obtain

$$f_n(t) \longrightarrow f(t)\,.$$

The discontinuous points of $f(\in D)$ are countable, by the dominated convergence theorem

$$\rho_\lambda(f_n, f) = \int_T |f_n(t) - f(t)|\, dt \longrightarrow 0\,.$$

Thus, $\mathcal{B}(D) = \mathcal{B}_\lambda(D)$ is proven.

What remains is the proof of $\mathcal{B}_\lambda(D) = \mathcal{B}_K(D)$. With respect to $[a, b] \subset T$, let

$$I_{a,b}(f) = \int_a^b f(t)\, dt.$$

Since

$$|I_{a,b}(f) - I_{a,b}(g)| \leqq \int_a^b |f(t) - g(t)| dt \leqq \rho_\lambda(f, g),$$

it follows that $I_{a,b} : D \to \boldsymbol{R}^1$ is continuous with respect to τ_λ. From the right continuous property of $f(\in D)$,

$$\pi_t(f) = f(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} I_{a,a+\epsilon}(f).$$

What this means is that $\pi_t : D \to \boldsymbol{R}^1$ is $\mathcal{B}_\lambda(D)$ measurable, and we obtain

$$\mathcal{B}_K(D) \subset \mathcal{B}_\lambda(D).$$

Since ρ_λ is a decomposable metric on D, we can think about the proof of $\mathcal{B}(C) \subset \mathcal{B}_K(C)$ in the same fashion. If we can show that for any arbitrary $g \in D$,

$$F(f) = \rho_B \lambda(f, g) = \int_T |f(t) - g(t)|\, dt$$

is $\mathcal{B}_K(D)$, then we obtain

$$\mathcal{B}_\lambda(D) \subset \mathcal{B}_K(D),$$

and the proof is completed.

If we let

$$F_n(f) = \sum_{i=1}^{n} \big|f(t_{n,i}) - g(t_{n,i})\big|\big(t_{n,i} - t_{n,i-1}\big),$$

since $f(t_{n,i}) = \pi_{t_{n,i}}(f)$, $F_n : D \to \boldsymbol{R}^1$ is $\mathcal{B}_K(D)$ measurable. Because $f, g(\in D)$ is Riemann-integrable, we have

$$F(f) = \lim_{n\to\infty} F_n(f).$$

What this means is that F is $\mathcal{B}_K(D)$ measurable.

Exercise 5.1 In relation to $D = D(T) \quad (T = [0,1])$, prove the following propositions:

(i) $f \in D$ is the uniformly convergent limit of a right-continuous sequence of step functions. The reverse is also true.

(ii) $f_n = 1_{[0,1/n)} \quad (n = 1, 2, \cdots)$ is a Cauchy sequence with respect to the metric ρ_S in D. It is not convergent with respect to ρ_S.

(iii) The above-mentioned $\{f_n\}$ is not a Cauchy sequence with respect to ρ_B.

(iv) ρ_B, ρ_S, and ρ_λ satisfy the conditions of metrics (in D).

(v) τ_u on C is the relative form of τ_S on D to C.

[Hint] When $f_1, f_2, \ldots, f \in C$, show that $\rho_S(f_n, f) \to 0$ and $\rho_u(f_n, f) \to 0$ are equal.

Christopher Ting
Translation Version 1.0
Date: 10 May, 2017