MEASURE AND PROBABILITY

H. Watanabe

Professor emeritus, Kyushu University, Fukuoka, Japan

Keywords: Measure, algebra (field), σ-algebra (field), measurable function, Lebesgue integral, bounded convergence, Fatou's lemma, product measure, Radon-Nikodym, signed measure, Radon measure, Haar measure, probability, probability space, Brownian motion, convergence in law, convergence in probability, almost sure convergence, Borel-Cantelli lemma, Tchebychev inequality, characteristic function, moment generating function, normal (Gaussian) distribution, Poisson distribution, infinitely divisible distribution, stable law, law of large numbers, central limit theorem, random walk, transient, recurrent, conditional probability, conditional expectation, martingale, increasing process, stopping time, stochastic integral, Lévy process, Lévy measure, stationary process, ergodic theorem, measure-preserving transformation, Gaussian process, Markov process, transition probability, Feller process, Itô process, stochastic differential equation, Feynman-Kac formula, Markov chain, queuing process, Black-Scholes formula.

Contents

Summary

The Riemann integral confronts several inconveniencies, for example, the interchange of the limiting procedure and the integration sign, and the integral of singular functions. The Lebesgue integral overcomes these difficulties. The Lebesgue integral is extended on abstract spaces and provides a sound and powerful tool for probability theory. Probability theory is an integrated system associated with several uncertain phenomenological events. Markov processes and stationary processes are well established fields of probability theory. Stochastic dynamical theory finds many useful applications.

1. Introduction

Measure is a generalized notion such as length, area and volume. Let $I = (a,b]$ be the set of real numbers x satisfying $a < x \le b$. It is called a left open right closed interval. The length of I is defined to be ($b - a$). Also the length of intervals $[a,b), (a,b)$, and $[a,b]$ is given by $(b-a)$. If $I_i = (a_i, b_i]$ $(i = 1, 2, ..., n)$ are disjoint intervals in the sense that $I_i \cap I_j = \emptyset$ whenever $i \ne j$, then the length of the set $\cup_{i=1}^{n} I_i$ is defined by $\sum_{i=1}^{n} (b_i - a_i)$. By such procedure, one can define the length of a set which is the union of finite intervals. The extension of the notion of length for very general subsets of the real line has been achieved by Émile Borel and Henri Lebesgue around 1900.

Measures are defined on a σ-algebra which is a family of subsets of an abstract set, closed under complementation and countable union. A measure is a countably additive non-negative set function on the σ-algebra. An integral with respect to a measure has a notable property: passage to the limit can be performed under the integral sign. Also a product of two different measures is defined and an interchange of integrations with respect to the two measures is allowed under mild conditions. A relation exists between two different measures: absolute continuity and singularity. A measure is introduced on locally compact topological group which is called a Haar measure.

Theory of probability is concerned with the analysis of mathematical models which arise in the description of random phenomena, such as random coin tossing, the random movement of particles and so on. Despite its simple structure, one can find diverse applications in physics, biology, engineering, finance and so on. In order to build stochastic models, we are required to construct a probability space. A mathematically sound formalism for this purpose is founded by Kolmogorov (1933). A probability is nothing but a measure with a mass of 1. However, it becomes necessary to consider a family of measures, finite or countable or even uncountable. First, important concepts are independence and dependence which determine the relations of the family of probability measures. There are several mode of dependence, Markov property, Martingale, Stationarity, etc. Law of large numbers and central limit theorem are central topics in probability theory which showed up in a very great number of quite different situations. The study of sums of independent random variables is accomplished by Kolmogorov, Lévy, and Khintchine and summarized in the book of Gnedenko-Kolmogorov (1949). Lévy (1937) initiated the study of martingale and Ville (1939)

used the word martingale for a sequence of random variables. Doob (1940) established a rigorous and deep theory of martingale. Analytical study of Markov processes was first conducted systematically by Kolmogorov (1931) who established that the transition probability satisfies partial differential equations of parabolic type. Feller (1936, 1952, 1954, 1957) developed such analytical study. Itô showed that Markov processes could be obtained constructively by solving stochastic differential equations and established the theory of stochastic integration. Doob (1953) recognized that stochastic integrals are martingales. Kunita and Watanabe (1967) pushed such ideas further. However, some of their results were contained implicitly in the work of Itô (1951) on multiple Wiener integral. The most important process is the Brownian motion. Brownian movement is the name given to the irregular movement of pollen suspended in water, observed by the botanist Robert Brown in 1828. Bachelier (1900) wrote a paper "Théorie de la spéculation" in which he proposed the Brownian motion and considered it as the price of stock. Einstein also proposed another formulation. A mathematically rigorous model was established by Wiener (1923). Precise properties of sample functions of the Brownian motion were studied by Kakutani, Erdös, Chung, Sirao, Taylor and others

There is another interpretation of Itô processes. They can be considered as dynamical systems perturbed by noises. Such a point of view makes them applicable to stochastic control theory, stochastic finance, stochastic quantum field theory (Euclidean quantum field theory by E. Nelson), stochastic Navier-Stokes equation, and stochastic KDV equation. To discuss Itô stochastic differential equations in such a variety of situations, it becomes necessary to establish infinite-dimensional stochastic analysis on Banach spaces. These are the future directions to be developed.

In stationary processes, the main topic is the ergodic theorem. It originates from "ergodic hypotheses" introduced by L. Boltzmann (1887) and related statistical mechanics. On the mathematical side, there are contributions by Birkhoff (1932), E. Hopf (1937), Yoshida-Kakutani (1939) etc. Markov chain is initiated by Markov (1906) and developed by Hostinsky, Fréchet, Romanovski, Feller, Kolmogorov, Kakutani, Doob and others. Markov chains can be found in all applied fields, physics, chemistry, population genetics, queuing theory etc.

2. Measure

2.1. Fields of Sets

To state results it is convenient to introduce several definitions. Let. Ω be a set and $\mathcal{C}$ be a collection of subsets of Ω with the empty set $\emptyset \in \mathcal{C}$. $\mathcal{C}$ is called an algebra (or field) if $A, B \in \mathcal{C}$ implies that A^c and $A \cup B$ are in $\mathcal{C}$. Then, since $A \cap B = (A^c \cap B^c)^c$, it follows that $A \cap B \in \mathcal{C}$. An algebra $\mathcal{F}$ is called a σ-algebra if for any sequence A_n of sets in $\mathcal{F}$ their intersection $\bigcap_{n \geq 1} A_n$ is in $\mathcal{F}$. If $\Omega = R$ (1-dimensional Euclidean space) and $\mathcal{C}$ = the collection of sets of the form $\bigcup_{i=1}^{n} (a_i, b_i]$ where $-\infty \leq a_i < b_i \leq \infty$, then $\mathcal{C}$ is an algebra.

2.2. Lebesgue Measure

A function μ from $\mathcal{C}$ into $[0,\infty]$ is said to be finitely additive if and only if (iff) $\mu(\emptyset) = 0$ and $\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$ whenever A_i are disjoint, $A_i \in \mathcal{C}$ for $i = 1, 2, ..., n$ and $\bigcup_{i=1}^{n} A_i \in \mathcal{C}$. When $\mathcal{C}$ consists of all finite unions of intervals, the length of a set in $\mathcal{C}$ satisfies this requirement. The measure of a countable union of non-overlapping intervals is defined as the sum of their lengths, finite or not. The measures of a single point and any countable family of points are zero. This definition of measure can be extended to the class of Borel measurable sets. This is the smallest collection of subsets of R which contains all subintervals of R and is closed under countable union, countable intersection and complementation. Furthermore, the measure of a countable union of disjoint Borel measurable sets is the sum of their individual measures. This extension of measure to Borel measurable sets is uniquely done. Finally, the class of Lebesgue measurable sets consists of all sets obtained as the countable union of Borel measurable sets and any subsets of a Borel measurable set of measure 0, where the measures of the latter sets are defined as 0. The Lebesgue measure can be defined as follows; for any measurable set E,

$$\mu(E) = \inf \sum_{n=1}^{\infty} \mu(I_n).$$

Where infimum is taken over the class of countable coverings of E by intervals I_n such that $\bigcup_{n=1}^{\infty} I_n \supset E$. There exist uncountable measurable sets with measure 0. Let

$$G = (\frac{1}{3}, \frac{2}{3})$$
$$\cup (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$$
$$\cup (\frac{1}{27}, \frac{2}{27}) \cup (\frac{7}{27}, \frac{8}{27}) \cup (\frac{19}{27}, \frac{20}{27}) \cup (\frac{25}{27}, \frac{26}{27})...$$

Then, the measure of G is equal to $\sum_{n=1}^{\infty} 2^{n-1} 3^{-n} = \frac{1}{3}(1 - \frac{2}{3})^{-1} = 1$. Therefore, the measure of the set $C = [0,1] \cap G^c$ is 0. The set C is called Cantor set.

The Lebesgue measure is a function l of Lebesgue measurable sets with values in $[0,\infty]$ and has the following useful properties. $A, B_1, B_2, ...$ are Lebesgue measurable sets;

(1.a) $0 \le l(A)$
(1.b) $l(A) \le l(B)$ if $A \subset B$.
(1.c) $l(\bigcup_{n=1}^{\infty} B_n) \le \sum_{n=1}^{\infty} l(B_n)$.
(1.d) $l(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} l(B_n)$ if $B_i \cap B_j$ is empty for $i \ne j$.
(1.e) $l(B_n) \uparrow l(A)$ if $B_1 \subset B_2 \subset ....$ and $\bigcup_{n=1}^{\infty} B_n = A$.

(1.f) $l(B_n) \downarrow l(A)$ if $B_1 \supset B_2 \supset ....,\bigcap B_{n=1}^{\infty} = A$ and $l(B_1) < \infty$.

(1.g) $l(A) = l(x + A)$, where $x + A = \{x + y; y \in A\}$.

Let $\mathcal{A}$ be an algebra. If μ is a measure on $\mathcal{A}$, μ can be extended uniquely to the σ-algebra $\mathcal{F}(\mathcal{A})$ generated by $\mathcal{A}$.

2.3. Measures

In general setting, Ω can be an abstract space. In analysis and probability theory, the measure works mostly on the complete, separable metric space. The domain of the measure is the smallest σ-algebra containing all open sets in the metric topology. Its member is called Borel set. It is called a topological σ-algebra and will be denoted by $\mathcal{B}$. The measure is a function μ of the Borel sets with the properties (1.a) ~ (1.f). A measure space $(\Omega, \mathcal{F}, \mu)$ is said to be *complete* if $\mathcal{F}$ contains all subsets of any set of measure 0. If $\mu(\Omega) < \infty$, then for any Borel set A,

$$\mu(A) = \sup[\mu(K); K \text{ compact } K \subset A, K \in \mathcal{B}]$$

holds. A measure with such a property is called regular. Also it is tight, namely for any $\epsilon > 0$, there is a compact set K_ϵ such that $\mu(K_\epsilon^c) < \epsilon$. Every finite measure on a complete separable metric space is tight and regular.

A measure μ will be called a σ-finite measure on $(\Omega, \mathcal{F})$. If there is a sequence of sets $A_n \in \mathcal{F}$ such that $\mu(A_n) < \infty$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$.

Assuming the axiom of choice (also known as Zermelo's axiom), the existence of non-measurable sets can be proved. On the other hand, the existence of a non-Lebesgue measurable set cannot be proved in Zermelo-Frankel set theory if the use of the axiom of choice is forbidden.

2.4. Measurable Functions

A real function f on R is said to be measurable if for every a and b, the set $\{x; a \le f(x) < b\}$ is a measurable set. In general setting, let $(\Omega_i, \mathcal{B}_i), i = 1, 2$ be measurable spaces, where $\mathcal{B}_i$ is a σ-algebra of subsets of Ω_i. If f is a function from Ω_1 into Ω_2, then f is called measurable iff $f^{-1}(B) = \{\omega : f(\omega) \in B\} \in \mathcal{B}_1$ for all $B \in \mathcal{B}_2$.

The class of real measurable functions has the following properties.

(2.a) If f is measurable and c is a scalar, then cf is measurable.

(2.b) If f and g are measurable, then $f \pm g, fg$ and $f / g \ (g \ne 0)$ are measurable.

(2.c) Let (f_n) be a sequence of real functions. Then the functions $\sup_n f_n(\omega), \inf_n f_n(\omega), \ \limsup_{n\to\infty} f_n(\omega), \liminf_{n\to\infty} f_n(\omega)$ are all measurable.

A function f on Ω is called a simple function if it is expressed as

$$f(\omega) = \sum_{i=1}^n a_i I_{A_i}(\omega), \tag{1.1}$$

where A_i are disjoint sets and $I_A(\omega)$ denotes the indicator function of the set A, being 1 for $\omega \in A$ and 0 for $\omega \notin A$.

For any measurable function $f \geq 0$, let f_n be defined as follows. If $k2^{-n} \leq f(\omega) < (k+1)2^{-n}$ for an integer k satisfying $k < n2^n$, $f_n(\omega) = k2^{-n}$ and $f_n(\omega) = n$ for $f(\omega) \geq n$. Then it follows that f_n is measurable, simple and $0 \leq f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$. Namely, any measurable function can be approximated by simple measurable functions.

2.5. Integral

If $(\Omega, \mathcal{F}, \mu)$ is any measure space and f any simple function on Ω, as in (1.1) with $a_i \geq 0$ for all i, then, the integral of f with respect to μ is defined by

$$\sum_i a_i \mu(A_i) \in [0, \infty] (0 \cdot \infty \text{ is taken to be 0}).$$

which is denoted by $\int f \, d\mu$ or $\int_\Omega f(\omega) d\mu(\omega)$. The integral of $f \geq 0$ is defined by the limit of the integral of simple functions f_n above as $n \to \infty$,

$$\int f_n d\mu = \sum_{k=1}^{n2^{-n}} k2^{-n}\mu(\omega : k2^{-n} \leq f(\omega) < (k+1)2^{-n}) + n \times \mu(\omega : f(\omega) \geq n) \uparrow \int f d\mu$$

A measurable function f is integrable if $\int |f| d\mu < \infty$.

Let $f^+(\omega) = f(\omega) \vee 0$ and $f^-(\omega) = (-f(\omega)) \vee 0,$

where $a \vee b = \max(a, b)$.

Then the integral of f is defined by

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

It is said to be summable if the integral of f^+ or f^- is finite.

$f \le g$ $\mu \sim$ *almost everywhere* (a.e.) ($\mu \sim$ a.e.) means that $f \le g$ holds outside of a μ-negligible set (i.e. a set of μ-measure 0), namely $\mu(\{\omega : f(\omega) \ge g(\omega)\}) = 0$.

The most important properties of the integral are listed below.

(3.a) If $f \ge 0$ $\mu \sim$ a.e., then $\int f d\mu \ge 0$

(3.b) If f is integrable and a is real, then $\int af d\mu = a\int f d\mu$.

(3.c) If f and g are integrable, then $\int(f+g)d\mu = \int f d\mu + \int g d\mu$

(3.d) If f and g are integrable such that $f \ge g$ $\mu \sim$ a.e., then $\int f d\mu \ge \int g d\mu$.

(3.e) If f, g are integrable and $f = g$ $\mu \sim$ a.e., then $\int f d\mu = \int g d\mu$.

(3.f) If f is summable, then $\left| \int f d\mu \right| \le \int |f| d\mu$

(3.g) (*Bounded convergence theorem*) Let E be a set with $\mu(E) < \infty$. If f_n vanishes on E^c, $|f_n| \le M$, and $f_n \to f$ $\mu \sim$ a.e., then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$$

(3.h) (*Fatou's lemma*) If $f_n \ge 0$, , then

$$\liminf_{n\to\infty} \int f_n d\mu \ge \int (\liminf_{n\to\infty} f_n)d\mu.$$

(3.i) (*Monotone convergence theorem*)

If $f_n \ge 0$ and $f_n \uparrow f$, then $\int f_n d\mu \uparrow \int f d\mu$.

(3.j) (*Dominated convergence theorem*) If $f_n \to f$ $a.e., |f_n| \le g$ for all n and g is integrable, then $\int f_n d\mu \to \int f d\mu$

Even if $f_n \to f$, the convergence of the integral $\int f_n d\mu \to \int f d\mu$ is not guaranteed in general. (3.g), (3.j) give very useful sufficient simple conditions for the convergence. The inequality (3.h) which holds for any positive functions is also useful in more sophisticated proof of convergence in some cases.

2.6. Product Measures

If $(\Omega_i, \mathcal{F}_i, \mu_i), i = 1, 2, ..., n$ are σ-finite measure spaces, and

$\Omega = \Omega_1 \times \Omega_2 \times ..... \Omega_n = \{(\omega_1, \omega_2, ..., \omega_n) : \omega_i \in \Omega_i, i = 1, 2, ..., n\}$, there is a unique measure μ on the σ-algebra $\mathcal{F}$ generated by sets of the form $A_1 \times A_2 \times .... \times A_n, A_i \in \mathcal{F}_i$, satisfying

the product formula.

$$\mu\left(A_1 \times .... \times A_n\right) = \prod_{i=1}^{n} \mu_i\left(A_i\right)$$

(3.k) (*Fubini's theorem*) if $f \geq 0$ or $\int |f| d\mu < \infty$, then

$$\int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2)\right)\mu_1(d\omega_1) = \int_{\Omega_1 \times \Omega_2} f\, d\mu$$

$$= \int_{\Omega 2}\left(\int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)\right)\mu_2(d\omega_2)$$

The change of the order of integration, guaranteed by this formula under the simple condition of either the positivity or the integrability, is often useful in a concrete computation of integrals.

2.7. Relations between two Measures

Let v and μ be two measures on the same measurable space $(\Omega, \mathcal{F})$. The measure v is said to be absolutely continuous with respect to the measure μ, in symbols $v \prec \mu$, if $v(A) = 0$, whenever $\mu(A) = 0$. μ and v are *singular*, denoted $v \perp \mu$, if there is a measurable set A with $\mu(A) = v(\Omega - A) = 0$

(4.a) (*Lebesgue decomposition*) Let $(\Omega, \mathcal{F})$ be a measurable space and μ and v two σ-finite measures on it. Then there are unique measures v_{ac} and v_s such that $v = v_{ac} + v_c, v_{ac} \prec \mu$ and $v_s \perp \mu$.

(4.b) (*Radon-Nikodym derivative*) If μ, v are σ-finite measures and v is absolutely continuous with respect to μ, then there is a measurable function $h \geq 0$ for which $v(E) = \int_E h d\mu$ holds for all $E \in \mathcal{F}$. The function h is μ- almost unique in the sense that any two such functions h are equal $\mu \sim a.e.$. This function h is called the Radon-Nikodym derivative and denoted as $\frac{dv}{d\mu}$.

-
-
-

Bibliography

Dudley R.M.(1989). *Real Analysis and Probability*, 427 pp. Belmont, CA,USA: Wadsworth. [This presents a comprehensive description of set theory, topology, measure and foundations of probability]

Durret R. (1991) *Probability: Theory and Examples*, 453 pp. Belmont, CA, USA: Wadsworth. [This book is a text book for a one-year graduate course in probability with short resumé about measure theory]

Feller W. (1968). *An Introduction to Probability Theory and Its Applications* vol I, 509 pp. New York, USA: Wiley [This is a famous source book of classical probability.]

Feller W. (1966). *An Introduction to Probability Theory and Its Applications* vol. II, 626 pp. New York, USA: Wiley [This is mainly concerned with independent random variables.]

Glimm J. and Sharp D. (1999). *Stochastic Partial Differential Equations: Selected Applications in Continuum Physics*, in Mathematical Surveys and Monographs vol 44, 3-44 pp. Providence , Rhode Island, USA : American Mathematical Society . [Authors describe the role of probability in applied fields.]

Ikeda N. and Watanabe *S*. (1989) *Stochastic Differential Equations and Diffusion Processes*. 555 pp. New York, NY, USA : Elsevier . [This book is concentrated on stochastic differential calculus.]

Itô K. Mckean H.P. (1965) *Diffusion Processes and Their Sample Paths* 321 pp. New York, NY, USA: Springer-Verlag. [This book is a complete description about 1-dimensional diffusion processes.]

Ledoux M. and Talagrand M. *Probability in Banach Spaces: Isoperimetry and Processes*, 480 pp. New York, NY, USA: Spriger-Verlag. [This book provides a profound knowledge of isometric inequalities and regularity of random processes].

Revuz D. and Yor M. *Continuous Martingale and Brownian Motion* 557 pp. Berlin, Germany: Springer-Verlag. [This book focuses on the probabilistic theory of Brownian motion.]

Zabczyk J. *Parabolic equations* on Hilbert spaces, Lecture Notes in Mathematics, 1715, 117-213 pp. New York, NY, USA: Springer-Verlag. [This lecture deals with second order parabolic equations on a separable Hilbert space.]

Biographical Sketch

H. Watanabe was born on February 3, 1928, in Shizuoka, Japan. He received B.S. in 1951, from the Tokyo university of Literature and Science and D.S. in 1964 from Kyushu university. He was research assistant, The Institute of Mathematical Statistics(1951-1954), Defense Academy (1954-1963), as assistant, lecturer and assistant professor, associate professor, Nagoya Institute of technology (1963-1966), Professor(1966-1968), Shizuoka university, and Professor (1968-1991) Kyushu university.