Review

January 2, 2017 Debdeep Pati

1 Measure and Integral

Definition 1. (Measurable space and measurable sets). Let Ω be the universal set (sample space) with σ-field $\mathcal{A}$. Then $(\Omega, \mathcal{A})$ is called measurable space and the subsets of $\mathcal{A}$ are called measurable sets.

Definition 2. (Measure or probability measure). A non-negative σ-additive set function μ on a σ-algebra is called measure. It is called probability measure if $\mu(\Omega)(= P(\Omega)) = 1$.

Definition 3. A Lebesgue-Stieljes measure on $\mathbb{R}$ is a measure μ on $\mathcal{B}(\mathbb{R})$ such that $\mu(I) < \infty$ for each bounded interval I. A distribution function on $\mathbb{R}$ is a map $F : \mathbb{R} \to \mathbb{R}$ that is increasing ($a < b$ implies $F(a) \leq F(b)$) and right continuous $\lim_{h \downarrow 0} F(x + h) = F(x)$).

Theorem 1. Let μ be a Lebesgue-Stieljes (LS) measure on $\mathbb{R}$. Let $F : \mathbb{R} \to \mathbb{R}$ be defined, up to an additive constant, $F(b) - F(a) = \mu(a, b]$. Then F is a distribution function.

Theorem 2. Let F be a distribution function on $\mathbb{R}$, and let $\mu(a, b] = F(b) - F(a), a < b$. There is a unique extension of μ to a LS measure on $\mathbb{R}$.

Definition 4. (Measurable function). The function $f : \Omega_1 \to \Omega_2$ is measurable relative to the σ-algebras $\mathcal{A}_i, i = 1, 2$ iff $f^{-1}(A) \in \mathcal{A}_1$ for all $A \in \mathcal{A}_2$, i.e., $f^{-1}(\mathcal{A}_2) = \mathcal{A}_1$.

Definition 5. A measurable function $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ is called Borel measurable.

Theorem 3. Let $f_1, f_2, \ldots,$ be Borel measurable and $f_n \to f$. Then $f(= \lim f_n)$ is Borel measurable. (The same applies to lim sup and lim inf.)

Theorem 4. Any Borel-measurable function $f \geq 0$ is the limit of an increasing sequence of simple functions.

Definition 6. (Integral). Let μ be a measure on a σ-algebra $\mathcal{A}$.

a. For $f = 1_A, \mu f = \int f d\mu = \int 1_A d\mu = \mu A$.

b. For $f = \sum_{k=1}^{n} \alpha_k 1_{A_k}$, set $\mu f = \int \sum_{k=1}^{n} \alpha_k 1_{A_k} d\mu = \sum_{k=1}^{n} \alpha_k \mu A_k$ provided $+\infty$ and $-\infty$ do not occur in the sum together.

c. $f \geq 0$ is Borel-measurable, set $\mu f = \sup\{\mu s : s \,\text{is simple}, 0 \leq s \leq f\}$.

d. For Borel-measurable f, set $\mu f = \mu f^+ - \mu f^-$ provided $\infty - \infty$ can be excluded. f is called integrable if μf is finite.

Theorem 5. (Radon-Nikodym theorem). Let μ be a σ-finite measure and ν be a (σ-finite) signed measure on $\mathcal{A}$ with $\nu \ll \mu$. Then there is a measurable function $f : \Omega \to \mathbb{R}$ with

$$\nu A = \mu 1_A f = \int_A f d\mu$$

for all $A \in \mathcal{A}$. If g is another function with $\nu A = \mu 1_A g$, then, $f = g$ everywhere. f is called μ-density or Radon-Nikodym density.

Definition 7. (Measurable rectangles and product-σ-algebra). Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ be the cartesian product of $\Omega_k, k = 1, 2, \ldots, n$, and $\mathcal{A}_k$ the associated σ-algebras. A measurable rectangle in Ω is a set

$$\times_{i=1}^n A_i = A_1 \times \cdots \times A_n, A_k \in \mathcal{A}_k.$$

The σ-algebra generated by the measurable rectangles is called product σ-algebra, $\mathcal{A} = \otimes \mathcal{A}_k$, and $(\Omega, \mathcal{A})$ is called product measurable space.

Theorem 6. (Fubini's theorem). Let $(\Omega_k, \mathcal{A}_k, \mu_k), k = 1, 2$ be σ-finite measure spaces and let $f \in L_1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu)$, where $\mu = \mu_1 \otimes \mu_2$ denotes the product measure (with $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$). Then there are sets B_1 and B_2 such that $\mu_k(\Omega_k \backslash B_k) = 0$, for $k = 1, 2$, and (a) for $\omega_1 \in B_1$, $f(\omega_1, \cdot) \in L_1(\Omega_2, \mathcal{A}_2, \mu_2)$ and $g_1(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2) 1_{B_1}(\omega_1)$ is $\mathcal{A}_1$ measurable; (b) for $\omega_2 \in B_2, f(\cdot, \omega_2) \in L_1(\Omega_1, \mathcal{A}_1, \mu_1)$ and $g_2(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1) 1_{B_2}(\omega_2)$ is $\mathcal{A}_2$ measurable.

In particular,

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2)\mu_1(d\omega_1) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)\mu_2(d\omega_2) = \int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2).$$

2 Convergence of random variables and strong laws of large numbers

In the following, we consider the probability space $(\Omega, \mathcal{A}, P)$.

Definition 8. A Borel measurable function $X : \Omega \to \mathbb{R}^n$ is caled a random vector (or random variable if $n = 1$). The probability measure P^X on $\mathcal{B}^n$ induced by X is defined by

$$P^X B = P(X \in B) = P\{\omega : X(\omega) \in B\} = P X^{-1} B, B \in \mathcal{B}^n.$$

Definition 9. Consider random variables $X, X_1, X_2, \ldots \in L_p, p > 0$.

a. X_n convergence to X in L_p (in p-th norm or "in pth mean"), $X_n \to X$ in L_p, if $\|X_n - X\|_p = E(|X_n - X|^p)^{1/p} \to 0$

b. X_n converges to X in probability, $X_n \overset{p}{\to} X$, if $\forall \epsilon > 0, P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$.

c. X_n converges to X almost surely if there is a $N \subset \Omega$ with $PN = 0$ such that for all $\omega \notin N$, $X_n(\omega) \to X(\omega)$ (or $P\{(\lim_{n\to\infty} X_n = X)\} = 1$).

Remark 1. $X_n \overset{L_p}{\to} X \Rightarrow X_n \overset{p}{\to} X$, $X_n \overset{a.s.}{\to} X \Rightarrow X_n \overset{p}{\to} X$. The reverse is not always true.

Theorem 7. (SLLN). Suppose that $X_1, X_2, \ldots \in L_2$ are independent and $(b_n)_{n\in\mathbb{N}}$ is a sequence with $0 < b_n \uparrow \infty$. If $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)/b_n^2 < \infty$, then, for $S_n = \sum_{i=1}^{n} X_i$,

$$\frac{S_n - ES_n}{b_n} \overset{n\to\infty}{\to} 0$$

almost surely. A special case is $X_1, X_2, \ldots$ are i.i.d and $b_n = n$. Then $S_n/b_n = (1/n)\sum_{i=1}^{n} X_i \overset{a.s.}{\to} EX_1$.

Remark 2. Marcinkiewics-Zygmund SLLNs. Suppose $X_1, X_2, \ldots$ are identically distributed random variables and $p \in (0, 2)$. Then

a. If $X_1, X_2, \ldots,$ are pairwise independent and $(S_n - nc)/n^{1/p}$ converges a.s. for some $c \in \mathbb{R}$, then $E|X_1|^p < \infty$.

b. If $E|X_1|^p < \infty$ and $X_1, X_2, \ldots$ are independent, then $(S_n - nc)/n^{1/p}$ converges a.s. with any $c \in \mathbb{R}$ if $p \in (0, 1)$ and $c = EX_1$ if $p \in [1, 2)$.

Corollary 1. (Kolmogorov's SLLN). Suppose $X_1, X_2, \ldots,$ are i.i.d. random variables. Then $(S_n - nc)/n$ converges a.s. for some $c \in \mathbb{R}$ if and only if $E|X_1| < \infty$, in which case, $c = EX_1$.

Definition 10. *Consider probability measures $P, P_1, P_2, \ldots$ on $\mathcal{B}$. Then P_n converges to P, written as $P_n \Rightarrow P$, if for all bounded continuous functions f, $P_n f \to Pf$. If $X, X_1, X_2, \ldots$ are random variables with $P^{X_n} \Rightarrow P^X$, then we say X_n converges to X in distribution and write $X_n \Rightarrow X$ or $X_n \overset{d}{\to} X$.*

Remark 3. $X_n \overset{p}{\to} X \Rightarrow X_n \overset{d}{\to} X$. *The reverse is in general not true.*

3 Convergence of integrals and expectations

We begin with $(\Omega, \mathcal{A}, \mu)$ (which contains $\mu = P$ as a special case).

Theorem 8. (Beppo Levi's monotone convergence theorem). Let $0 \le f_1 \le f_2 \le \cdots$ and $f = \lim_{n\to\infty} f_n$. Then, $\mu f_n \to \mu_f$, *i.e.*, $\lim_{n\to\infty} \mu f_n = \mu(\lim_{n\to\infty} f_n)$. In the special case $\mu = P$ and $f_n = X_n$, we have $\lim_{n\to\infty} EX_n = E(\lim_{n\to\infty} X_n)$.

Lemma 1. (Fatou's Lemma). Consider $f_n \ge 0, n \in \mathbb{N}$ are measurable. Then,

a. $\lim_{n\to\infty} \inf_{k\ge n} \mu f_k \ge \mu(\lim_{n\to\infty} \inf_{k\ge n} f_k)$.

b. $\lim_{n\to\infty} \sup_{k\ge n} \mu f_k \le \mu(\lim_{n\to\infty} \sup_{k\ge n} f_k)$.

Theorem 9. (Lebesgue dominated convergence theorem). Consider $f_1, f_2, \dots,$ are measurable, $f_n \to f$, and $|f_n| \le g$ where g is integrable. Then, f is integrable and $\mu f_n \to \mu f$, i.e., $\lim_{n\to\infty} \mu f_n = \mu(\lim_{n\to\infty} f_n)$.

Theorem 10. If $\{T_\lambda : \lambda \in \Lambda\}$ is UI and $T_n \xrightarrow{d} T$, then $ET_n \to ET$.

4 Important Asymptotic Theorems

Theorem 11. *(Scheffé Lemma). Consider probability measures $P, P_1, P_2, \dots$ with Lebesgue-densities $f, f_1, f_2, \dots$. If $f_n \overset{n\to\infty}{\to} f$ pointwise, then*

$$\sup_{A\in\mathcal{B}} |P_n A - PA| \to 0.$$

Theorem 12. Suppose $X, X_1, X_2, \dots$ are random variables with distribution functions $F, F_1, F_2, \dots,$ i.e., $F_n(t) = P(X_n \le t)$. Then the following two statements are equivalent

a. $P^{X_n} \Rightarrow P^X$

b. $F_n(t) \to F(t)$ if F is continuous at t.

Remark 4. *a. A sequence of random vectors $X_1, X_2, \dots,$ (taking values in $\mathbb{R}^d$) is said to converge in distribution to a random vector X, "$X_n \overset{d}{=} X$" or "$P^{X_n} \Rightarrow P^X$", if*

$$P(X_n \le x) \to P(X \le x), \quad \textit{for all} \quad x \in C(n \to \infty),$$

where "$\le$" refers to the components and where $C = \{x \in \mathbb{R}^d : P(X_i = x_i) = 0 \, for \, 1, \dots, d\}$.

b. Cramer-Wold device:

$$X_n \overset{d}{=} X \Leftrightarrow a^{\mathrm{T}} X_n \overset{d}{=} a^{\mathrm{T}} X, \textit{ for all} \quad a \in \mathbb{R}^d.$$

Theorem 13. (Portmanteau theorem). Consider probability measures $P, P_1, P_2, \dots,$ on $\mathcal{B}$. Then the following statements are equivalent.

a. $P_n \Rightarrow P$.

b. $\text{liminf} P_n A \geq PA$, for all open sets $A \subset \mathbb{R}$.

c. $\text{limsup} P_n A \leq PA$, for all closed sets $A \subset \mathbb{R}$.

d. $P_n A \to PA$, for all A with $P(\delta A) = 0$, where δA denotes the boundary of A.

Remark 5. *a. Portmanteau's theorem can also be phrased for random variables: replace P_n and P by the induced measures P^{X_n} and P^X.*

b. Another statement refers to te characteristic function: if $X_n \overset{d}{=} X$ as $n \to \infty$, then $E(e^{itX_n}) \to E(e^{itX})$. The reverse holds if $E(e^{itX})$ is continuous at 0.

c. Further, if $E|X_n|^k < \infty$ for all $k \in \mathbb{N}$ and $Ee^{tX} < \infty$ for all $|t| < \epsilon$, and

$$EX_n^k \to EX^k \, as \, n \to \infty, for \; all \, k \in \mathbb{N},$$

then $X_n \overset{d}{=} X$ as $n \to \infty$.

Theorem 14. (Polya's theorem). If $T_n \overset{d}{\to} T$ and if, additionally, the distribution function of T is continuous on $\mathbb{R}$, then $P(T_n \leq x)$ converges uniformly,

$$\sup_{t \in \mathbb{R}} |P(T_n \leq t) - P(T \leq t)| \to 0 \quad (n \to \infty).$$

Theorem 15. (Continuous mapping theorem). Let h be a measurable function and $X, X_1, X_2, \ldots, X_n$ are random variables. Then,

a. If $X_n \overset{d}{\to} X$ and h is continuous, then $h(X_n) \overset{d}{\to} h(X)$.

b. If $X_n \overset{p}{\to} c$ and h is continuous, then $h(X_n) \overset{p}{\to} h(c)$.

Theorem 16. (Classical Slutsky's theorem). Consider random variables $(X_n, Y_n)_{n \in \mathbb{N}}$ defined on $(\Omega_n \mathcal{A}_n, P_n)$. Suppose $X_n \overset{d}{\to} X$ and $Y_n \overset{p}{\to} c$ for some $c \in \mathbb{R}$. Then

a. $X_n + Y_n \overset{d}{\to} X + c$.

b. $X_n Y_n \overset{d}{\to} cX$.

c. $X_n / Y_n \overset{d}{\to} X/c, \quad \text{provided} \, c \neq 0$.

Theorem 17. (Generalized Slutsky's theorem). Consider (X_n, Y_n) with $X_n \overset{d}{\to} X$ and $Y_n \overset{p}{\to} c$. Suppose h is continuous, then $h(X_n, Y_n) \overset{d}{\to} h(X, c)$.

Theorem 18. (Central limit theorem (i.i.d. version)). If $X_1, X_2, \ldots$ are i.i.d random variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$, then

$$n^{-1/2} \sum_{i=1}^{n} (X_i - \mu) \Rightarrow \text{N}(0, \sigma^2)$$

or

$$\frac{\sum X_i - E(\sum X_i)}{\sqrt{\text{Var}(X_i)}} \Rightarrow X \sim \text{N}(0, 1).$$

Remark 6. There are many generalizations.

a. Example: the Lindeberg- Feller central limit theorem for independent but not identically distributed random variables $X_{n1}, \ldots, X_{nn}$ with $E(X_{ni}) = \mu_{ni}$ and $\text{Var}(X_i) = \sigma^2_{ni}$ (which contains the i.i.d version as a special case). Then

$$n^{-1/2} \sum (X_i - \mu_{ni}) \Rightarrow \text{N}(0, \sigma^2).$$

follows from the Feller condition $n^{-1} \sum_{i=1}^{n} \sigma^2_{ni} \to \sigma^2 < \infty$ and the Lindeberg condition

$$\frac{1}{n} \sum_{i=1}^{n} E\big(1_{\{|X_{ni} - \mu_{ni}| > \sqrt{n}\epsilon\}} \, |X_{ni} - \mu_{ni}|^2\big) \overset{n \to \infty}{\to} X.$$

b. A multivariate version of the Lindeberg central limit theorem: For each $n \geq 1$, let $\{X_{in}, i = 1, \ldots, r_n\}$ be a collection of independent mean zero variables satisfying $\sum_{i=1}^{r_n} E(X_{in} X_{in}^{\mathrm{T}}) = \mathbb{I}$. Suppose that

$$\lim_{n \to \infty} \sum_{i=1}^{r_n} E[\|X_{in}\|^2 \, 1(\|X_{in}\| > \epsilon)] = 0 \quad \text{for all} \, \epsilon > 0,$$

Then $\sum_{i=1}^{r_n} X_{in} \overset{d}{\to} \text{N}(0, \mathbb{I})$.

5 Conditional expectation

Definition 11. (Conditional Expectation). Let $B \in \mathcal{A}$ and $PB > 0$. Suppose X is $\mathcal{A}$-measurable and integrable. The conditional expectation X given B is

$$E(X \mid B) = \frac{1}{PB} E 1_B X = \int X(\omega) P(d\omega \mid B).$$

Theorem 19. (Alternative definition). $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{A}$. Suppose X is $\mathcal{A}$-measurable and integrable. The conditional expectation $E(X \mid \mathcal{G})$ of X given $\mathcal{G}$ is defined as

a. $E(X \mid \mathcal{G})$ is $\mathcal{G}$ measurable

b. $E1_C X = E1_C E(X \mid \mathcal{G})$ for all $C \in \mathcal{G}$.

$E(X \mid \mathcal{G})$ exists and is almost unique.

Theorem 20. The conditional expectation $E(X \mid \mathcal{A}(Y))$ can be written as $E(X \mid Y) \circ Y$. It is characterized by

$$E1_{Y \in B} X = \int_B E(X \mid Y = y) P^Y(dy).$$

Theorem 21. (Monotone convergence theorem for conditional expectations). Suppose $0 \leq X_n \uparrow X$. Then $E(X_n \mid \mathcal{G}) \uparrow E(X \mid \mathcal{G})$.

Theorem 22. (Dominated convergence theorem for conditional expectations). Suppose $X_n \to X$ Then $E(X_n \mid \mathcal{G}) \uparrow E(X \mid \mathcal{G})$.

Theorem 23. Assume Y is $\mathcal{G}$-measurable, X and XY are integrable. Then $E(XY \mid \mathcal{G} = YE(X \mid \mathcal{G})$. Hence $E(XY \mid Y) = YE(X \mid Y))$.

Theorem 24. (Jensen's inequality). Let I be an open interval and $f : I \mapsto \mathbb{R}$ be convex. $X : \Omega \mapsto I$ is $\mathcal{A}$-measurable and integrable. Then

$$E(f \circ X \mid \mathcal{G}) \geq f \circ E(X \mid \mathcal{G}).$$