1 Random variables, independence, integration and conditioning

1.1 Measurable functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Under measurable function $f : \Omega \to \mathbb{R}$ we understand a function that satisfies $f^{-1}(B) \in \mathcal{F}$ for each Borel set $B \in \mathcal{B}(\mathbb{R})$, where

$$f^{-1}(B) := \{x \in \Omega : f(x) \in B\}.$$

Resorting to generators, it is enough to require this measurability condition to hold for $B = (-\infty, x]$ with x running over the set of rational numbers. In the context of a probability space we speak of random variables and use notation X, Y, ξ etc in place of f.

A function obtained by algebraic and analytic manipulations with a countable set (f_n) of measurable functions is again a measurable function. For instance $\limsup f_n$ is measurable (in general, as function into extended real line $\mathbb{R} \cup \{\infty\}$).

The indicator function of $A \in \mathcal{F}$

$$1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases} .$$

is measurable, and so are the *simple* functions of the form

$$f(x) = \sum_{j=1}^{n} y_j 1_{A_j}(x), \quad y_j \in \mathbb{R}\,.$$

With any family of measurable functions $\{f_t,\ t \in T\}$ we associate σ-algebra $\sigma(f_t,\ t \in T)$, generated by sets $\{f_t^{-1}(B) :\ t \in T, B \in \mathcal{B}(\mathbb{R})\}$. This is the smallest sub-σ-algebra of $\mathcal{F}$ which makes all f_t's measurable.

Let $(\Omega_t, \mathcal{F}_t),\ t \in T$, be measurable spaces. Their product is the measurable space

$$\prod_{t\in T} \Omega_t = \{(\omega_t,\ t \in T) : \omega_t \in \Omega_t\}, \quad \bigotimes_{t\in T} \mathcal{F}_t,$$

where the product σ-algebra is generated by cylinder sets

$$A_t \times \prod_{s\neq t} \Omega_s, \ \ A_t \in \mathcal{F}_t.$$

A measurable function can be used to transfer measure from Ω to $\mathbb{R}$ as $\mu \mapsto \mu_f$, where

$$\mu_f(B) := \mu(f^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

In the case of probability space, the measure on $\mathbb{R}$, induced by random variable X, is called *probability distribution* of X. The measure of halfline,

$$F_X(x) = \mathbb{P}(X \le x), \quad x \in \mathbb{R}$$

is known as the cumulative distribution function of X.

Example For simple function

$$f = \sum_{j=1}^{n} y_j 1_{A_j}$$

the induced measure on $\mathbb{R}$ is discrete,

$$\mu_f = \sum_{j=1}^{n} \mu(A_j)\delta_{y_j},$$

charging point y_j with mass $\mu(A_j)$.

1.2 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be probability space. Events $(A_i, i \in I) \subset \mathcal{F}$ are called independent if for every selection of distinct $i_1, \dots, i_k \in I$

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Let $(\mathcal{F}_t, t \in T)$ be sub-σ-algebras of $\mathcal{F}$. They are called independent if for any choice of distinct indices $t_1, \dots, t_k$ any events $A_{t_1} \in \mathcal{F}_{t_1}, \dots, A_{t_k} \in \mathcal{F}_{t_k}$ are independent.

Random variables X_i's are independent if their induced σ-algebras $\sigma(X_i)$ are independent.

1.3 Tail events

Let $A_i \in \mathcal{F}$ be events, $i \in \mathbb{N}$. Consider the event 'A_n occurs inifnitely often' (more precisely, 'infinitely many of A_n's occur')

$$\{A_n \text{ i.o.}\} := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Theorem. Borel-Cantelli Lemma)

(a) *If* $\sum_n \mathbb{P}(A_n) < \infty$ *then* $\mathbb{P}(A_n \text{ i.o.}) = 0$,

(b) *If* $A_1, A_2, \dots$ *are independent and* $\sum_n \mathbb{P}(A_n) = \infty$ *then* $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. (a) is exercise in Lecture 1. We focus on (b). We have

$$\{A_n \text{ i.o.}\}^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Clearly

$$\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \subset \bigcup_{n=2}^{\infty} \bigcap_{k=n}^{\infty} A_k \subset \cdots,$$

hence

$$\{A_n \text{ i.o.}\}^c = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty} \lim_{m\to\infty} \left(\bigcap_{k=n}^{m} A_k^c\right) =$$

using independence and that $\sum_n \mathbb{P}(A_n) = \infty$

$$= \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} (1 - \mathbb{P}(A_k)) = 0.$$

□

Example Let $X_1, X_2, \dots$ be independent $\mathcal{N}(0,1)$-distributed random variables (any other continuous distribution would also work). We say that there is a record at index n if $X_n = \max(X_1, \dots, X_n)$, call this event A_n. One can check that $\mathbb{P}(A_n) = 1/n$ and that the events are independent. Since $\sum_n 1/n = \infty$ the number of records is infinite with probability 1.

The Borel-Cantelli Lemma exemplifies situation where probability of some 'distant' event may assume only values 0 and 1.

Let $\mathcal{F}_j, j \in \mathbb{N}$, be σ-algebras (sub-σ-algebras of $\mathcal{F}$). We define *tail* σ-algebra as

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \mathcal{F}_k\right).$$

Each $A \in \mathcal{T}$ is called tail event.

Example In the coin-tossing space, let $\mathcal{F}_n$ be the σ-algebra generated by outcomes in n first trials. The event 'the pattern 1011101 occurs infinitely many times in the sequence' is a tail event.

Theorem. (Kolmogorov's $0-1$ law) *If $\mathcal{F}_1, \mathcal{F}_2, \cdots$ are independent, then $\mathcal{T}$ is trivial in the sense that $\mathbb{P}(A) = 0$ or 1 for each $A \in \mathcal{T}$.*

Proof. Suppose A is a tail event, since $A \in \bigcup_{k=n}^{\infty} \mathcal{F}_k$, we have that A is independent of $\mathcal{F}_1, \dots, \mathcal{F}_{n-1}$. Since this holds for every n, A is independent of $\mathcal{T}$. In particular, A is independent of itself, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\,\mathbb{P}(A)$, which is only possible when $\mathbb{P}(A)$ is 0 or 1. □

Example Let $X_1, X_2, \dots$ be independent random variables, generating σ-algebras $\sigma(X_j)$ $j \in \mathbb{N}$. The event

$$A = \{\omega \in \Omega : \sum_{n=1}^{\infty} X_n < \infty$$

is a tail event, therefore can only have probability 0 or 1.

Theorem. (Kolmogorov's Three Series Theorem) *Series $\sum_{n=1}^{\infty} X_n$ of independent random variable converges alsmost surely if and only if the following conditions hold with some constant $c > 0$*

(i) $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c) < \infty$,

(ii) $\sum_{n=1}^{\infty} \mathbb{E}(X_n 1_{\{|X_n| \le c\}}) < \infty$,

(iii) $\sum_{n=1}^{\infty} \mathrm{Var}(X_n 1_{\{|X_n| \le c\}}) < \infty$.

Example For normal random variables $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ convergence of the series $\sum_n X_n$ holds if and only if $\sum_n m_n < \infty$ and $\sum_n \sigma_n^2 < \infty$.

1.4 Lebesgue integral and expectation

In your courses you defined the expectation of discrete random variable X with values $x_1, x_2, \dots$ as

$$\mathbb{E}\, X = \sum_j x_j\, \mathbb{P}(X = x_j),$$

and if X has a density φ as

$$\mathbb{E}\, X = \int_{-\infty}^{\infty} x\varphi(x) dx.$$

These are unified by the general concept of Lebesgue integral.

For measurable function f on $(\Omega, \mathcal{F}, \mathbb{P})$ we wish to define

$$\int_\Omega f(x) d\mu(x).$$

Suppose first that f is nonnegative. For simple

$$f(x) = \sum_{j=1}^{k} y_j 1_{A_j}(x),$$

we set

$$\int_\Omega f(x) d\mu(x) = \sum_{j=1}^{k} y_j \mu(A_j).$$

For the general $f \geq 0$, consider sets

$$A_{jk} = \begin{cases} \{x : \frac{k}{2^j} \leq f(x) < \frac{k+1}{2^j}\}, & k = 0, 1, \ldots, j2^j - 1, \\ \{x : f(x) \geq j\}, \ k = j2^j, \end{cases}$$

and simple functions

$$f_j(x) = \sum_{k=0}^{j2^j} \frac{k}{2^j} \, 1_{A_{jk}}(x),$$

so

$$\int_\Omega f_j(x) d\mu(x) = \sum_{k=0}^{j2^j} \frac{k}{2^j} \, \mu(A_{jk}),$$

which we consider as a lower approximation for Lebesgue integral. The *Lebesgue integral* of f is defined as the limit

$$\int_\Omega f(x) d\mu(x) := \lim_{j \to \infty} \int_\Omega f_j(x) d\mu(x).$$

Example Let $f(x) = 1_{[0,1] \setminus \mathbb{Q}}$ be the indicator function of irrational numbers on $[0, 1]$. The Riemann integral over $[0, 1]$ does not exist, because every upper integral sum is 1, and every lower is 0. The Lebesgue integral is

$$\int_{[0,1]} f(x) dx = 1 \cdot \lambda([0, 1] \setminus \mathbb{Q}) + 0 \cdot \lambda([0, 1] \cap \mathbb{Q}) = 1.$$

Note that here dx means the same as $d\lambda(x)$.

For the general $f : \Omega \to \mathbb{R}$ let $f_+(x) = \max(f(x), 0), f_-(x) = \max(-f(x), 0)$ be positive and negative parts, then $f(x) = f_+(x) - f_-(x)$. If

$$\int_\Omega |f(x)| d\mu(x) < \infty$$

we say that f is integrable and we define the Lebesgue integral of f as

$$\int_\Omega f(x) d\mu(x) = \int_\Omega f_+(x) d\mu(x) - \int_\Omega f_-(x) d\mu(x).$$

Note that $\int_0^\infty (\sin x)/x dx = \pi/2$ exists as improper Riemann integral over $\mathbb{R}_+$, but not as Lebesgue integral because $\int_0^\infty |(\sin x)/x| dx = \infty$.

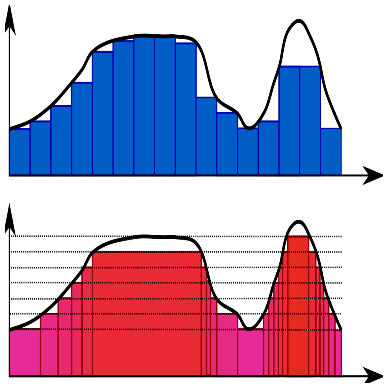

Figure 1: Lower Riemann integral sum and integral sum for Lebesgue integral.

1.5 Absolute continuity of measures

Let $p_j, j \in \mathbb{N}$ be positive numbers, with sum 1. We may treat the identity function on $\mathbb{N}$ as a random variable $X : \mathbb{N} \to \mathbb{N}$ with the probability mass function $P = (p_j, j \in \mathbb{N})$. For any function g we have the expectation calculated as

$$\mathbb{E}_P\, g(X) = \sum_{j\in\mathbb{N}} g(j)p_j$$

If $Q = (Q_j, j \in \mathbb{N})$ is some other probability mass function, the corresponding expectation is

$$\mathbb{E}_Q\, g(X) = \sum_{j\in\mathbb{N}} g(j)q_j.$$

To write the Q-expectation in terms of P, let $\xi(j) = p_j/q_j$, then

$$\mathbb{E}_Q\, g(X) = \sum_{j\in\mathbb{N}} g(j)\xi_j p_j = \mathbb{E}_P\, [\xi g(X)].$$

The random variable ξ is an instance of the Radon-Nikodym derivative/density.

In full generality, let μ, ν be two measures on $(\Omega, \mathcal{F})$. Call ν *absolutely continuous* with respect to μ, written as $\mu \gg \nu$ if

$$A \in \mathcal{F},\ \mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0.$$

The measures are called *equivalent*, denoted $\mu \sim \nu$, if

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0,$$

which means that the measures have the same null-sets.

Theorem. (Radon-Nikodym theorem.) *If $\mu \gg \nu$ then there exists a nonnegative measurable function ξ on Ω such that for any measurable f*

$$\int_\Omega f(x)d\nu(x) = \int_\Omega f(x)\xi(x)d\nu(x),$$

provided one of the integrals exists.

In particular, $\nu(A) = \int_A \xi(x)d\mu(x)$. We write $\xi = \frac{d\nu}{d\mu}$ and call ξ the Radon-Nikodym deritative of ν with respect to μ.

Example For λ the lebesgue measure, ν the normal $\mathcal{N}(0,1)$ distribution, the Radon-Nikodym derivative is the normal density

$$\xi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi.}}$$

1.6 Conditional expectation

For two random variables, X, Y, recall that the conditional expectation $\mathbb{E}[X|Y]$ is defined as follows. Calculate the function $h(y) = \mathbb{E}[X|Y = y]$, in case of discrete random variables as

$$\mathbb{E}[X|Y = y_j] = \sum_j x_i \, \mathbb{P}(X = x_i|Y = y_j),$$

or when (X, Y) have joint density as

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)dx,$$

where

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

is the conditional density. Then define $\mathbb{E}[X|Y] = h(Y)$ by subsituting random variable Y for dummy variable y.

Intuitively, $\mathbb{E}[X|X] = X$, and in the discrete case this is easily checked. When X has density, this is still true but we cannot use the above formula with $Y = X$, because (X, X) has no *joint* density function.

We wish to introduce more general conditional expectation $\mathbb{E}[X|\mathcal{G}]$ given sigma-algebra $\mathcal{G} \subset \mathcal{F}$. Suppose first $X \geq 0$. Let

$$\mathbb{Q}(A) := \mathbb{E}[X \cdot 1_A] = \int_A X d\,\mathbb{P}.$$

For disjoint sets $A_n \in \mathcal{G}$

$$\int_{\cup_n A_n} X d\,\mathbb{P} = \sum_n \int_{A_n} X d\,\mathbb{P},$$

which entails that $\mathbb{Q}$ is a measure, and $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$. By the Radon-Nikodym theorem there exists a $\mathcal{G}$-measurable random variable ξ such that

$$\mathbb{Q}(A) = \int_A \xi d\,\mathbb{P}.$$

We denote this variable as

$$\xi = \mathbb{E}[X|\mathcal{G}],$$

and call it the conditional expectattion of X given $\mathcal{G}$. The defining property is

$$\int_A X d\,\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}] d\,\mathbb{P}, \qquad A \in \mathcal{G}.$$

For any X, we write $X = X_+ - X_-$ and define the conditional expectation by

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X_+|\mathcal{G}] - \mathbb{E}[X_-|\mathcal{G}],$$

which exists if X is integrable.

The following rules will be used in the sequel:

(i) $\mathbb{E}[X|\{\varnothing, \Omega\}) = \mathbb{E}\,X$,

(ii) $\mathbb{E}[aX + bY|\mathcal{G}] = a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]$,

(iii) $\mathbb{E}[1|\mathcal{G}] = 1$,

(iv) taking out what is known: if Y is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = Y \cdot \mathbb{E}[X|\mathcal{G}],$$

(v) tower property: for $\mathcal{G}_1 \subset \mathcal{G}_2$

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[\mathcal{G}_\in],$$

in particular $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}\, X$.

Exercises

1. Let U be uniformly distributed over $[0,1]$, let X_n be the nth digit of U. Show that $X_1, X_2, \ldots$ are independent. For each q find the probability $\mathbb{P}(\lim_{n\to\infty} n^{-1} \sum_{j=1}^n X_j = q)$.
2. Let $Y_1, Y_2, \ldots$ be independent, exponentially distributed random variables with $\mathbb{E}\, Y_i = 1$. Show that $\mathbb{P}(Y_n > \log n \text{ i.o.}) = 1$.
3. Show that condition (i) in the three series theorem is necessary for convergence of the series.
4. Let $X_1, X_2, \ldots$ be arbitrary random variables. Prove that if $\sum_{j=1}^\infty |X_j| < \infty$ then the series $\sum_{j=1}^\infty X_j$ converges absolutely with probability one. Hint: use Chebyshev's inequality to estimate probabilities.
5. Let μ be a normal distribution $\mathcal{N}(m, \sigma^2)$, and ν the exponential distribution with parameter β. Argue that $\mu \gg \nu$ and find the Radon-Nikodym derivative $d\nu/d\mu$.
6. Let $A_{i,j}$ be a system of disjoint events, with $\cup_{i,j} A_{i,j} = \Omega$. Let $A_i = \cup_j A_{i,j}$. Let $\mathcal{G}_2$ be generated by all $A_{i,j}$'s, and let $\mathcal{G}_1$ be generated by A_i's. Describe as precise as you can the random variables $\mathbb{E}[X|\mathcal{G}_1], \mathbb{E}[X|\mathcal{G}_2]$. Assuming $\mathbb{P}(A_{i,j}) > 0$, prove the tower property in this example.

Literature

1. S. Resnick, A probability path, Springer 2003.

2. R. Schilling, Measures, integrals and martingales, CUP 2005.

3. A. Shiryaev, Probability, Springer, 1996.