MAT6082 Topics in Analysis

2^{nd} term, 2015-16

Teacher: Professor Ka-Sing Lau

Schedule: Wednesday, 2.30-5.00 pm

Venue: LSB 222

Topics: Introduction to Stochastic Calculus

In the past thirty years, there has been an increasing demand of stochastic calculus in mathematics as well as various disciplines such as mathematical finance, pde, physics and biology. The course is a rigorous introduction to this topic. The material include conditional expectation, Markov property, martingales, stochastic processes, Brownian motions, Ito's calculus, and stochastic differential equations.

Prerequisites

Students are expected to have good background in real analysis, probability theory and some basic knowledge of stochastic processes.

References:

1. A Course in Probability Theory, K.L. Chung, (1974).
2. Measure and Probability, P. Billingsley, (1986).
3. Introduction to Stochastic Integration, H.H. Kuo, (2006).
4. Intro. to Stochastic Calculus with Application, F. Klebaner, (2001).
5. Brownian Motion and Stoch. Cal., I. Karatzas and S. Shreve, (1998).
6. Stoch. Cal. for Finance II– Continuous time model, S. Shreve, (2004).

Everyone knows calculus deals with deterministic objects. On the other hand *stochastic calculus* deals with random phenomena. The theory was introduced by Kiyosi Ito in the 40's, and therefore stochastic calculus is also called *Ito calculus*. Besides its interest in mathematics, it has been used extensively in statistical mechanics in physics, the filter and control theory in engineering. Nowadays it is very popular in the option price and hedging in finance. For example the well-known Black-Scholes model is

$$dS(t) = rS(t)dt \ + \ \sigma S(t)dB(t)$$

where $S(t)$ is the stock price, σ is the volatility, and r is the interest rate, and $B(t)$ is the Brownian motion. The most important notion for us is the Brownian motion. As is known the botanist R. Brown (1828) discovered certain zigzag random movement of pollens suspended in liquid. A. Einstein (1915) argued that the movement is due to bombardment of particle by the molecules of the fluid. He set up some basic equations of Brownian motion and use them to study diffusion. It was N. Wiener (1923) who made a rigorous study of the Brownian motion using the then new theory of Lebesgue measure. Because of that a Brownian motion is also frequently called a Wiener process.

Just like calculus is based on the *fundamental theorem of calculus*, the Ito calculus is based on the *Ito Formula*: Let f be a twice differentiable function on $\mathbb{R}$, then

$$f(B(t)) - f(B(0)) = \int_0^T f'(B(t))dB(t) + \frac{1}{2}\int_0^T f''(B(t))dt$$

where $B(0) = 0$ to denote the motion starts at 0. There are formula for integration, for example, we have

$$\int_0^T B(t)dB(t) = \frac{1}{2}B(t)^2 - \frac{1}{2}T; \qquad \int_0^T tdB(t) = TB(T) - \int_0^T B(t)dt.$$

In this course, the prerequisite is real analysis and basic probability theory. In real analysis, one needs to know σ-fields, measurable functions, measures and integration theory, various convergence theorems, Fubini theorem and the Radon-Nikodym theorem. We will go through some of the probability theory on conditional expectation, optional r.v. (stopping time), Markov property, martingales ([1], [2]). Then we will go onto study the Brownian motion ([2], [3], [5]), the stochastic integration and the Ito calculus ([3], [4], [5]).

Chapter 1

Basic Probability Theory

1.1 Preliminaries

Let Ω be a set and let $\mathcal{F}$ be a family of subsets of Ω, $\mathcal{F}$ is called a *field* if it satisfies

(i) $\emptyset,\ \Omega \in \mathcal{F}$;

(ii) for any $A \in \mathcal{F},\ A^c \in \mathcal{F}$;

(iii) for any $A,\ B \in \mathcal{F},\ A \cup B \in \mathcal{F}$ (hence $A \cap B \in \mathcal{F}$).

It is called a *σ-field* if (iii) is replaced by

(iii)′ for any $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F},\ \cup_{n=1}^{\infty} A_n \in \mathcal{F}$ (hence $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$).

If $\Omega = \mathbb{R}$ and $\mathcal{F}$ is the smallest σ-field generated by the open sets, then we call it the Borel field and denote by $\mathcal{B}$.

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ such that $\mathcal{F}$ is a σ-field in Ω, and $P : \mathcal{F} \to [0, 1]$ satisfies

(i) $P(\Omega) = 1$

(ii) countable additivity : if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ is a disjoint family, then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

We call Ω a sample space, $A \in \mathcal{F}$ an event (or measurable set) and P a probability measure on Ω; an element $\omega \in \Omega$ is called an outcome.

Theorem 1.1.1. *(Caratheodory Extension Theorem) Let $\mathcal{F}_0$ be a field of subsets in Ω and let $\mathcal{F}$ be the $\sigma-$ field generated by $\mathcal{F}_0$. Let $P: \mathcal{F}_0 \to [0,1]$ satisfies (i) and (ii) (on $\mathcal{F}_0$). Then P can be extended uniquely to $\mathcal{F}$, and $(\Omega, \mathcal{F}, P)$ is a probability space.*

The proof of the theorem is to use the outer measure argument.

Example 1. Let $\Omega = [0,1]$, let $\mathcal{F}_0$ be the family of set consisting of finite disjoint unions of half open intervals $(a,b]$ and $[0,b]$, Let $P([a,b)) = |b-a|$. Then $\mathcal{F}$ is the Borel field and P is the Lebesgue measure on $[0,1]$.

Example 2. Let $\{(\Omega_n, \mathcal{F}_n, P_n)\}_n$ be a sequence of probability spaces. Let $\Omega = \prod_{n=1}^{\infty} \Omega_n$ be the product space and let $\mathcal{F}_0$ be the family of subsets of the form $E = \prod_{n=1}^{\infty} E_n$, where $E_n \in \mathcal{F}_n$, $E_n = \Omega_n$ except for finitely many n. Define

$$P(E) = \prod_{n=1}^{\infty} P(E_n)$$

Let $\mathcal{F}$ be the σ-field generated $\mathcal{F}_0$, then $(\Omega, \mathcal{F}, P)$ is the standard infinite product measure space.

Example 3. (Kolmogorov Extension Theorem) Let P_n be probability measures on $(\prod_{k=1}^{n} \Omega_k, \mathcal{F}_n)$ satisfying the following consistency condition: for $m \leq n$

$$P_n \circ {\pi_{nm}}^{-1} = P_m$$

where $\pi_{nm}(x_1 \cdots x_n) = (x_1 \cdots x_m)$. On $\Omega = \prod_{k=1}^{\infty} \Omega_k$, we let $\mathcal{F}_0$ be the field of sets $F = E \times \prod_{k=n+1}^{\infty} \Omega_k$, $E \in \mathcal{F}_n$ and let

$$P(F) = P_n(E).$$

Then this defines a probability spaces $(\Omega,\ \mathcal{F}, P)$, where $\mathcal{F}$ is the σ-field generated by $\mathcal{F}_0$.

Remark: The probability space in Example 2 is the underlying space for a sequence of independent random variables. Example 3 is for more general sequence of random variables (with the consistency condition).

A random variable (r.v.) X on $(\Omega, \mathcal{F})$ is an (extended) real valued function $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ such that for any Borel subset B of $\mathbb{R}$,

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

(i.e. X is $\mathcal{F}$-measurable). We denote this by $X \in \mathcal{F}$. It is well known that

– For $X \in \mathcal{F}$, X is either a simple function (i.e., $\sum_{k=1}^{n} a_k \chi_{A_k}(\omega)$ where $A_k \in \mathcal{F}$), or is the pointwise limit of a sequence of simple functions.

– Let $X \in \mathcal{F}$ and g is a Borel measurable function, then $g(X) \in \mathcal{F}$.

– If $\{X_n\} \subseteq \mathcal{F}$ and $\lim_{n\to\infty} X_n = X$, then $X \in \mathcal{F}$.

– Let $\mathcal{F}_X$ be the σ-field generated by X, i.e., the sub-σ-field $\{X^{-1}(B) : B \in \mathcal{B}\}$. Then for any $Y \in \mathcal{F}_X$, $Y = \varphi(X)$ for some extended-valued Borel function φ on $\mathbb{R}$.

Sketch of proof ([1, p.299]): First prove this for simple r.v. Y so that $Y = \phi(X)$ for some simple function ϕ. For a bounded r.v. $Y \geq 0$, we can find a sequence of increasing simple functions $\{Y_n\}$ such that $Y_n = \phi_n(X)$ and

$Y_n \nearrow Y$. Let $\phi(x) = \overline{\lim}_n \phi_n(x)$, hence $Y = \phi(X)$. Then prove Y for the general case.

A r.v. $X : (\Omega, \mathcal{F}) \to \mathbb{R}$ induces a *distribution (function)* on $\mathbb{R}$:

$$F(x) = F_X(x) = P(X \leq x).$$

It is a non-decreasing, right continuous function with $\lim_{n\to-\infty} F(x) = 0$, $\lim_{n\to\infty} F(x) = 1$. The distribution defines a measure μ

$$\mu((a,b]) = F(b) - F(a)$$

(use the Caratheodory Extension Theorem here). More directly, we can define μ by

$$\mu(B) = P(X^{-1}(B)) \ , \quad B \in \mathcal{B}.$$

The jump of F at x is $F(x) - F(x-) = P(X = x)$. A r.v. X is called a *discrete* if F is a jump function; X is called a *continuous* r.v. if F is continuous, i.e., $P(X = x) = 0$ for each $x \in \mathbb{R}$, and X is said to have a density function $f(x)$ if F is absolutely continuous with the Lebesgue measure and $f(x) = F'(x)$ a.e., equivalently $F(x) = \int_{-\infty}^{x} f(y)dy$.

For two random variables X, Y on $(\Omega, \mathcal{F})$, the random vector (X, Y) : $(\Omega, \mathcal{F}) \to \mathbb{R}^2$ induces a distribution F on $\mathbb{R}^2$

$$F(x,y) = P(X \leq x, \ Y \leq y)$$

and F is called the joint distribution of (X, Y), the corresponding measure μ is given by

$$\mu((a,b] \times (c,d]) = F(b,d) - F(a,d) - F(b,c) + F(a,c),$$

Similarly we can define the joint distribution $F(x_1 \cdots x_n)$ and the corresponding measure.

For a sequence of r.v., $\{X_n\}_{n=1}^{\infty}$, there are various notions of convergence.

(a) $X_n \to X$ a.e. (or a.s.) if $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ (pointwise) for $\omega \in \Omega \backslash E$ where $P(E) = 0$.

(b) $X_n \to X$ in probability if for any $\epsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.

(c) $X_n \to X$ in distribution if $F_n(x) \to F(x)$ at every point x of continuity. It is equivalent to $\mu_n \to \mu$ vaguely i.e., $\mu_n(f) \to \mu(f)$ for all $f \in C_0(\mathbb{R})$, the space of continuous functions vanish at ∞ (detail in [1]).

The following relationships are basic ([1] or Royden): $(a) \Rightarrow (b) \Rightarrow (c)$; $(b) \Rightarrow (a)$ on some subsequence. On the other hand we cannot expect (c) to imply (b) as the distribution does not determine X. For example consider the interval $[0, 1]$ with the Lebesgue measure, the r.v.'s $X_1 = \chi_{[0,\frac{1}{2}]}$, $X_2 = \chi_{[\frac{1}{2},1]}$, $X_3 = \chi_{[0,\frac{1}{4}]} + \chi_{[\frac{3}{4},1]}$ all have the same distribution.

The expectation of a random variable is defined as

$$E(X) = \int_{\Omega} X(\omega) dP(\omega) = \int_{-\infty}^{\infty} x dF(x) \ (= \int_{-\infty}^{\infty} x d\mu(x))$$

and for a Borel measurable h, we have

$$E(h(X)) = \int_{\Omega} h(X(\omega)) dP(\omega) = \int_{-\infty}^{\infty} h(x) dF(x).$$

The most basic convergence theorems are:

(a) Fatou lemma:

$$X_n \geq 0, \quad \text{then} \ \ E(\underline{\lim}_{n\to\infty} X_n) \ \leq \ \underline{\lim}_{n\to\infty} E(X_n).$$

(b) Monotone convergence theorem:

$$X_n \geq 0, \ X_n \nearrow X, \quad \text{then} \quad \lim_{n\to\infty} E(X_n) = E(X).$$

(c) Dominated convergence theorem:

$$|X_n| \leq Y,\ E(Y) < \infty \ \text{ and } \ X_n \to X \ a.e., \ \text{ then } \lim_{n\to\infty} E(X_n) = E(X).$$

We say that $X_n \to X$ in $L^p, p > 0$ if $E(|X|^p) < \infty$ and $E(|X_n - X|^p) \to 0$ as $n \to \infty$. It is known that L^p convergence implies convergence in probability. The converse also holds if we assume further $E(|X_n|^p) \to E(|X|^p) < \infty$ ([1], p.97).

Two events $A, B \in \mathcal{F}$ are said to be *independent* if

$$P(A \cap B) = P(A)P(B).$$

Similarly we say that the events $A_1, \cdots A_n \in \mathcal{F}$ are independent if for any subsets $A_{j_1}, \cdots, A_{j_k}$,

$$P(\bigcap_{i=1}^{k} A_{j_i}) = \prod_{i=1}^{k} P(A_{j_i}).$$

Two sub-σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ are said to be independent if any choice of sets of each of these σ-fields are independent. Two r.v.'s X, Y are independent if the σ-fields $\mathcal{F}_X$ and $\mathcal{F}_Y$ they generated are independent. Equivalently we have

$$P(X \leq x,\ Y \leq y) = P(X \leq x)\ P(Y \leq y),$$

(i.e., the joint distribution equals the product of their marginal distributions). We say that $X_1 \cdots X_n$ are independent if for any $X_{i_1} \cdots X_{i_k}$, their joint distribution is a product of their marginal distributions.

Proposition 1.1.2. *Let X, Y be independent, then $f(X)$ and $g(Y)$ are independent for any Borel measurable functions f and g.*

Exercises

1. Can you identify the interval $[0, 1]$ with the Lebesgue measure to the probability space for tossing a fair coin repeatedly?

2. Prove Proposition 1.1.2.

3. Suppose that $\sup_n |X_n| \le Y$ and $E(Y) < \infty$. Show that

$$E(\overline{\lim}_{n\to\infty} X_n) \ \ge \ \overline{\lim}_{n\to\infty} E(X_n)$$

4. If $p > 0$ and $E(|X|^p) < \infty$, then $x^p P(|X| > x) = o(1)$ as $x \to \infty$. The converse also holds for $E(|X|^{p-\epsilon}) < \infty$ for $0 < \epsilon < p$.

5. For any d.f. F and any $a \ge 0$, we have

$$\int_{-\infty}^{\infty} (F(x+a) - F(x))dx = a$$

6. Let X be a positive r.v. with a distribution F, then

$$\int_0^\infty (1 - F(x))\ dx = \int_0^\infty x\ dF(x).$$

and

$$E(X) = \int_0^\infty P(X > x)\ dx = \int_0^\infty P(X \ge x)\ dx$$

7. Let $\{X_n\}$ be a sequence of identically distributed r.v. with finite mean, then

$$\lim_n \frac{1}{n} E(\max_{1\le j\le n} |X_j|) = 0.$$

(Hint: use Ex.6 to express the mean of the maximum)

8. If X_1, X_2 are independent r.v.'s each takes values $+1$ and -1 with probability $\frac{1}{2}$, then the three r.v.'s $\{X_1, X_2, X_1X_2\}$ are pairwise independent but not independent.

9. A r.v. is independent of itself if and only if it is constant with probability one. Can X and $f(X)$ be independent when $f \in \mathcal{B}$?

10 . Let $\{X_j\}_{j=1}^n$ be independent with distributions $\{F_j\}_{j=1}^n$. Find the distribution for $\max_j X_j$ and $\min_j X_j$.

11. If X and Y are independent and $E(|X+Y|^p) < \infty$ for some $p > 0$, then $E(|X|^p) < \infty$ and $E(|Y|^p) < \infty$.

12. If X and Y are independent, $E(|X|^p) < \infty$ for some $p \geq 1$, and $E(Y) = 0$, then $E(|X+Y|^p) \geq E(|X|^p)$.

1.2 Conditional Expectation

Let $\Lambda \in \mathcal{F}$ with $P(\Lambda) > 0$, we define

$$P(E|\Lambda) = \frac{P(\Lambda \cap E)}{P(\Lambda)} \qquad \text{where} \quad P(\Lambda) > 0.$$

It follow that for a discrete random vector (X, Y),

$$P(Y = y|X = x) = \begin{cases} \dfrac{P(Y = y,\ X = x)}{P(X = x)}\ , & \text{if} \quad P(X = x) > 0\ , \\ 0\ , & \text{otherwise.} \end{cases}$$

Moreover if (X, Y) is a continuous random variable with joint density $f(x, y)$, the conditional density of Y given $X = x$ is

$$f(y|x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)}\ , & \text{if} \quad f_X(x) > 0\ , \\ 0\ , & \text{otherwise}\ . \end{cases}$$

where $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ is the marginal density. The conditional expectation of Y given $X = x$ is

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f(y|x) dy.$$

Note that

$$g(x) := E(Y|X = x) \quad \text{is a function on} \quad x\ ,$$

and hence

$$g(X(\cdot)) := E(Y|X(\cdot)) \quad \text{is a r.v. on} \quad \Omega\ . \tag{1.2.1}$$

In the following we have a more general consideration for the conditional expectation (and also the conditional probability): $E(Y|\mathcal{G})$ where $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$.

First let us look at a special case where $\mathcal{G}$ is generated by a measurable partition $\{\Lambda_n\}_n$ of Ω (each member in $\mathcal{G}$ is a union of $\{\Lambda_n\}_n$). Let Y be an integrable r.v., then

$$E(Y|\Lambda_n) = \int_\Omega Y(\omega)dP_{\Lambda_n}(\omega) = \frac{1}{P(\Lambda_n)}\int_{\Lambda_n} Y(\omega)dP(\omega). \tag{1.2.2}$$

(Here $P_{\Lambda_n}(\cdot) = \frac{P(\,\cdot\,\cap\,\Lambda_n)}{P(\Lambda_n)}$ is a probability measure for $P(\Lambda_n) > 0$). Consider the random variable (as in (1.2.1))

$$Z(\cdot) = E(Y|\mathcal{G})(\cdot) := \sum_n E(Y|\Lambda_n)\chi_{\Lambda_n}(\cdot) \in \mathcal{G}.$$

It is easy to see that if $\omega \in \Lambda_n$, then $Z(\omega) = E(Y|\Lambda_n)$, and moreover

$$\int_\Omega E(Y|\mathcal{G})dP = \sum_n \int_{\Lambda_n} E(Y|\mathcal{G})dP = \sum_n E(Y|\Lambda_n)P(\Lambda_n) = \int_\Omega YdP\ .$$

We can also replace Ω by $\Lambda \in \mathcal{G}$ and obtain

$$\int_\Lambda E(Y|\mathcal{G})dP = \int_\Lambda YdP \qquad \forall\ \Lambda \in \mathcal{G}.$$

Recall that for μ, ν two σ-finite measures on $(\Omega, \mathcal{F})$ and $\mu \geq 0$, ν is called *absolutely continuous* with respect to μ $(\nu \ll \mu)$ if for any $\Lambda \in \mathcal{F}$ and $\mu(\Lambda) = 0$, then $\nu(\Lambda) = 0$. The Radon-Nikodym theorem says that there exists $g = \dfrac{d\nu}{d\mu}$ such that

$$\nu(\Lambda) = \int_\Lambda g d\mu \qquad \forall\ \ \Lambda \in \mathcal{F}.$$

Theorem 1.2.1. *If $E(|Y|) < \infty$ and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$, t hen there exists a unique $\mathcal{G}$-measurable r.v., denote by $E(Y|\mathcal{G}) \in \mathcal{G}$, such that*

$$\int_\Lambda YdP = \int_\Lambda E(Y|\mathcal{G})\ dP \qquad \forall\ \Lambda \in \mathcal{G}.$$

Proof. Consider the set-valued function

$$\nu(\Lambda) = \int_\Lambda YdP \qquad \Lambda \in \mathcal{G}.$$

Then ν is a "signed measure" on $\mathcal{G}$. It satisfies

$$P(\Lambda) = 0 \implies \nu(\Lambda) = 0.$$

Hence ν is absolutely continuous with respect to P. By the Radon-Nikodym theorem, the derivative $g = \frac{d\nu}{dP} \in \mathcal{G}$ and

$$\int_\Lambda Y dP = v(\Lambda) = \int_\Lambda g dP \qquad \forall \ \Lambda \in \mathcal{G}.$$

This g is unique: for if we have $g_1 \in \mathcal{G}$ satisfies the same identity,

$$\int_\Lambda Y dP = v(\Lambda) = \int_\Lambda g_1 dP \qquad \forall \ \Lambda \in \mathcal{G}.$$

Let $\Lambda = \{g > g_1\} \in \mathcal{G}$, then $\int_\Lambda (g - g_1) dP = 0$ implies that $P(\Lambda) = 0$. We can reverse g and g_1 and hence we have $P(g \neq g_1) = 0$. It follows that $g = g_1$ $\mathcal{G}$-a.e.

Definition 1.2.2. *Given an integrable r.v. Y and a sub-σ-field $\mathcal{G}$, we say that $E(Y|\mathcal{G})$ is the* conditional expectation *of Y with respect to $\mathcal{G}$ (also denote by $E_{\mathcal{G}}(Y)$) if it satisfies*

(a) $E(Y|\mathcal{G}) \in \mathcal{G}$;

(b) $\int_\Lambda Y dP = \int_\Lambda E(Y|\mathcal{G}) dP \quad \forall \ \Lambda \in \mathcal{G}$.

If $Y = \chi_\Delta \in \mathcal{F}$, we define $P(\Delta|\mathcal{G}) = E(\chi_\Delta|\mathcal{G})$ and call this the conditional probability with respect to $\mathcal{G}$.

Note that the *conditional probability* can be put in the following way:

(a)′ $\quad P(\Delta|\mathcal{G}) \in \mathcal{G}$;

(b)′ $\quad P(\Delta \cap \Lambda) = \int_\Lambda P(\Delta|\mathcal{G}) dP \quad \forall \ \Lambda \in \mathcal{G}$.

It is a simple exercise to show that the original definition of $P(\Delta|\Lambda)$ agrees with this new definition by taking $\mathcal{G} = \{\emptyset, \ \Lambda, \ \Lambda^c, \ \Omega\}$.

Note that $E(Y|\mathcal{G})$ is "almost everywhere" defined, and we call one such function as a "version" of the conditional expectation. For brevity we will not mention the "a.e." in the conditional expectation unless necessary. If $\mathcal{G}$ is the sub-σ-field $\mathcal{F}_X$ generated by a r.v. X, we write $E(Y|X)$ instead of $E(Y|\mathcal{F}_X)$. Similarly we can define $E(Y|X_1, \cdots, X_n)$.

Proposition 1.2.3. *For $E(Y|X) \in \mathcal{F}_X$, there exists an extended-valued Borel measurable φ such that $E(Y|X) = \varphi(X)$, and φ is given by*

$$\varphi = \frac{d\lambda}{d\mu} \ ,$$

where $\lambda(B) = \int_{X^{-1}(B)} Y dP, \quad B \in \mathscr{B}$, and μ is the associated probability of the r.v. X on $\mathbb{R}$.

Proof. Since $E(Y|X) \in \mathcal{F}_X$, we can write $E(Y|X) = \varphi(X)$ for some Borel measurable φ (see §1). For $\Lambda \in \mathcal{F}$, there exists $B \in \mathcal{B}$ such that $\Lambda = X^{-1}(B)$. Hence

$$\int_\Lambda E(Y|X) dP = \int_\Omega \chi_B(X)\varphi(X) dP = \int_{\mathbb{R}} \chi_B(X)\varphi(X) d\mu = \int_B \varphi(x) d\mu$$

On the other hand by the definition of conditional probability,

$$\int_\Lambda E(Y|X) dP = \int_{X^{-1}(B)} Y dP = \lambda(B).$$

It follows that $\lambda(B) = \int_B \varphi(x) d\mu$ for all $B \in \mathcal{B}$. Hence $\varphi = \dfrac{d\lambda}{d\mu}$. $\square$

The following are some simple facts of the conditional expectation:

- If $\mathcal{G} = \{\phi, \Omega\}$, then $E(Y|\mathcal{G})$ is a constant function and equals $E(Y)$.

- If $\mathcal{G} = \{\phi, \Lambda, \Lambda^c, \Omega\}$, then $E(Y|\mathcal{G})$ is a simple function which equals $E(Y|\Lambda)$ on Λ, and equals $E(Y|\Lambda^c)$ on Λ^c,

– If $\mathcal{G} = \mathcal{F}$ or $Y \in \mathcal{G}$, then $E(Y|\mathcal{G}) = Y$.

– If (X, Y) has a joint density function, then $E(Y|X)$ coincides with the expression in (1.2.1).

Using the defining relationship of conditional expectation, we can show that the linearity, the basic inequalities and the convergence theorems for $E(\cdot)$ also hold for $E(\cdot\ |\mathcal{G})$. For example we have

Proposition 1.2.4. *(Jensen inequality) If φ is a convex function on $\mathbb{R}$, and Y and $\varphi(Y)$ are integrable r.v., then for each sub-σ-algebra $\mathcal{G}$,*

$$\varphi\big(E(Y|\mathcal{G})\big) \leq E\big(\varphi(Y)|\mathcal{G}\big)$$

Proof. If Y is a simple r.v., then $Y = \sum_{j=1}^{n} y_j \chi_{\Lambda_j}$ with $\Lambda \in \mathcal{F}$. It follows that

$$E(Y|\mathcal{G}) = \sum_{j=1}^{n} y_j E(\chi_{\Lambda_j}|\mathcal{G}) = \sum_{j=1}^{n} y_j P(Y_{\Lambda_j}|\mathcal{G})$$

and

$$E(\varphi(Y)|\mathcal{G}) = \sum_{j=1}^{n} \varphi(y_j) P(Y_{\Lambda_j}|\mathcal{G}).$$

Since $\sum_{j=1}^{n} P(\Lambda_j|\mathcal{G}) = 1$, the inequality holds by the convexity of φ.

In general we can find a sequence of simple r.v. $\{Y_m\}$ with $|Y_m| \leq |Y|$ and $Y_m \to Y$, then apply the above together with the dominated convergence theorem. $\Box$

Proposition 1.2.5. *Let Y and YZ be integrable r.v. and $Z \in \mathcal{G}$, then we have*

$$E(YZ|\mathcal{G}) = ZE(Y|\mathcal{G}).$$

Proof. It suffices to show that for $Y, Z \geq 0$

$$\int_\Lambda ZE(Y|\mathcal{G})dP = \int_\Lambda ZY dP \qquad \forall\, \Lambda \in \mathcal{G}.$$

Obviously, this is true for $Z = \chi_\Delta$, $\Delta \in \mathcal{G}$. We can pass it to the simple r.v. Then use the monotone convergence theorem to show that it hold for all $Z \geq 0$, and then the general integrable r.v. $\square$

Proposition 1.2.6. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be sub-σ-fields of $\mathcal{F}$ and $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Then for Y integrable r.v.*

$$E(E(Y|\mathcal{G}_2)|\mathcal{G}_1) = E(Y|\mathcal{G}_1) = E(E(Y|\mathcal{G}_1)|\mathcal{G}_2). \tag{1.2.3}$$

Moreover

$$E(Y|\mathcal{G}_1) = E(Y|\mathcal{G}_2) \quad \textit{iff} \quad E(Y|\mathcal{G}_2) \in \mathcal{G}_1. \tag{1.2.4}$$

Proof. Let $\Lambda \in \mathcal{G}_1$, then $\Lambda \in \mathcal{G}_2$. Hence

$$\int_\Lambda E(E(X|\mathcal{G}_2)|\mathcal{G}_1)dP = \int_\Lambda E(Y|\mathcal{G}_2)dP = \int_\Lambda Y dP = \int_\Lambda E(Y|\mathcal{G}_1)dP,$$

and the first identity in (1.2.3) follows. The second identity is by $E(Y|\mathcal{G}_1) \in \mathcal{G}_2$ (recall that $Z \in \mathcal{G}$ implies $E(Z|\mathcal{G}) = Z$).

For the last part, the necessity is trivial, and the sufficiency follows from the first identity. $\square$

As a simple consequence, we have

Corollary 1.2.7. $E(E(Y|X_1, X_2)|X_1) = E(Y|X_1) = E(E(Y|X_1)|X_1, X_2)$.

Exercises

1. (Bayes' rule) Let $\{\Lambda_n\}$ be a $\mathcal{F}$-measurable partition of Ω and let $E \in \mathcal{F}$ with $P(E) > 0$. Then

$$P(\Lambda_n|E) = \frac{P(\Lambda_n)\ P(E|\Lambda_n)}{\sum_n P(\Lambda_n)P(\Lambda_n|E)}\ .$$

2. If the random vector (X, Y) has probability density $p(x, y)$ and X is integrable, then one version of $E(X|X+Y=z)$ is given by

$$\int xp(x, z-x)dx \ \Big/ \int p(x, z-x)dx\ .$$

3. Let X be a r.v. such that $P(X > t) = e^{-t},\ t > 0$. Compute $E(X|X \vee t)$ and $E(X|X \wedge t)$ for $t > 0$. (Here $\vee$ and $\wedge$ mean maximum and minimum respectively.

4. If X is an integrable r.v., Y is a bounded r.v., and $\mathcal{G}$ is a sub-σ-field, then

$$E\big(E(X|\mathcal{G})Y\big) = E\big(XE(Y|\mathcal{G})\big).$$

5. Prove that $\mathrm{var}(E(Y|\mathcal{G})) \le \mathrm{var}(Y)$.

6. Let X, Y be two r.v., and let $\mathcal{G}$ be a sub-σ-field. Suppose

$$E(Y^2|\mathcal{G}) = X^2, \quad E(Y|\mathcal{G}) = X,$$

then $Y = X$ a.e.

7. Give an example that $E(E(Y|X_1)|X_2) \neq E(E(Y|X_2)|X_1)$. (*Hint: it suffices to find an example* $E(X|Y) \neq E(E(X|Y)|X)$ *for* Ω *to have three points*).