Markov Processes and Martingales

Károly Simon

Department of Stochastics
Institute of Mathematics
Budapest University of Technology and Economics
www.math.bme.hu/~simonk

Measure Theory and Conditional Expectation (a review)

Introduction Basic Measure Theory

The most important reference material is

R. Durett Probability Theory and Examples 5. ed.

One can download it free of charge from:

https://sites.math.duke.edu/~rtd/PTE/pte.html

We do not follow this book closely but the material students of the course are supposed to know from previous semesters is available in this book.

1. When we want to emphasize that the set A is a disjoint union of $A_1, A_1, \dots A_n$ then we write

$$A = \bigsqcup_{i=1}^{n} A_i. \quad (1)$$

We remark that in Durett's book [5] the same disjoint union is denoted by $+_{i=1}^{n} A_i$. If A is not necessarily disjoint union of $A_1, A_1, \dots A_n$ then we write

$$A = \bigcup_{i=1}^{n} A_i.$$

2. Given the sequences a_n, b_n with $b_n > 0$ for every n. We write that
   - $a_n = o(b_n)$ if $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 0$.
   - $a_n = O(b_n)$ if $\limsup\limits_{n\to\infty} \frac{|a_n|}{b_n} < \infty$
   - $a_n = \Theta(b_n)$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$.
   - $a_n \sim b_n$ if $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 1$.
3. The Borel σ-algebra on $\mathbb{R}^d$ is denoted by $\mathfrak{R}^d$. When $d = 1$ we write simply $\mathfrak{R}$.
4. The d-dimensional Lebesgue measure is denoted by $\mathcal{L}^d$.
5. Given a set X.
   - The complement of an $A \subset X$ is denoted by A^c. If there is a topology on X then

   - $\overline{A}$ stands for the closure of A (if it makes sense),
   - A° is the interior of A.
6. Let $\mathcal{F}$ be a collection of subsets of a set X. Then we write $\sigma(\mathcal{F})$ for the generated σ-algebra.
7. When we say a set is *countable* we mean that it is either finite or countably infinite.

Definition 1.1 (semialgebra, algebra,σ-algebra)

Given a set Ω. A collection $\mathcal{S} \subset 2^\Omega$ is called

semialgebra: if
- $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$ and
- $\forall A \in \mathcal{F}$, A^c is a finite disjoint union of elements of $\mathcal{F}$.

algebra: if $\mathcal{S}$ is closed for all finite set operations.

σ-algebra: if $\mathcal{S}$ is closed for all countable set operations.

Definition 1.2

1. Assume that $\mathcal{S}$ is a semialgebra. Then we write $\overline{\mathcal{S}}$ for the generated algebra (which the collection of disjoint unions of sets from $\mathcal{S}$).
2. Let A be algebra. The σ-algebra $\sigma(A)$ generated by A is the smallest σ-algebra that contains A. This is the intersection of all σ-algebras that contain A.

Observe that any finitely additive set function defined on a semialgebra $\mathcal{S}$ extends to the generated algebra $\overline{\mathcal{S}}$ in an obvious way.

Basic Measure Theory

- Let $\mathcal{F}$ be a σ-algebra of a given set X. We say that a function $\mu : \mathcal{F} \to [0, \infty]$ is a measure if
  - (a) $\mu(\emptyset) = 0$ and
  - (b) $\mu\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$ for every disjoint sequence of sets $\{E_i\}_{i=1}^{\infty}$ in $\mathcal{F}$.
- We say that a set function ν is a pre-measure ν is defined on a algebra $\mathcal{A} \subset 2^X$ satisfying:
  - (i) $\nu(\emptyset) = 0$ and

Basic Measure Theory (cont.)

  - (ii) $\nu\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \nu(E_i)$ for every disjoint sequence of sets $\{E_i\}_{i=1}^{\infty}$ in $\mathcal{A}$.
- An outer measure ν on X is defined on all subsets of X takes values from $[0, \infty]$ such that
  - $\nu(\emptyset) = 0$,
  - $\nu(A) \leqslant \nu(B)$ if $A \subset B$,
  - $\nu(\cup_{i=1}^{\infty} E_i) \leqslant \sum_{i=1}^{\infty} \nu(E_i)$ for all sequence of sets $\{E_i\}_{i=1}^{\infty}$.

Basic Measure Theory (cont.)

- Now we assume that (X, d) is a metric space. We say that the outer measure ν is a metric outer measure if

$$\nu(A \cup B) = \nu(A) + \nu(B)$$

  holds for all $A, B \subset X$ with $\inf\{d(a, b) : a \in A, b \in B\} > 0$.
- The set function $\nu : \mathcal{A} \to [0, \infty]$ for an $\mathcal{A} \subset 2^X$ is called σ-finite if there exists $\{B_i\}_{i=1}^{\infty}$, $B_i \in \mathcal{A}$ such that $\Omega = \bigcup_{i=1}^{\infty} B_i$ and $\nu(B_i) < \infty$ for all i.

Carathéodory measurable sets

Let μ be a pre-measure on the algebra $\mathcal{A} \subset 2^X$. We define the corresponding outer measure

(2) $$\mu^*(B) := \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) : B \subset \bigcup A_i,\ A_i \in \mathcal{A}\right\}$$

and the family of Carathéodory-measurable sets:

(3) $$\mathcal{M} := \{E : \forall A \subset X : \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \backslash E)\}.$$

The proof of the following theorem is available in the Appendix of [5].

Carathéodory measurable sets (cont.)

Theorem 2.1

1. *$\mathcal{M}$ is a σ-algebra. We call it the σ-algebra of μ-measurable sets.*
2. *$\mathcal{M} \supset \sigma(\mathcal{A})$.*
3. *The restriction of μ^* to $\mathcal{M}$ is a measure.*
4. *$\forall B \in \mathcal{A}$ we have $\mu^*(B) = \mu(B)$.*
5. *If (X, d) is a metric space and μ^* is a metric outer measure. Then $\mathcal{M}$ contains the Borel sets. That is the restriction of μ^* to $\mathcal{M}$ is a Borel measure.*

Carathéodory's Extension Theorem

Let $\mathcal{S}$ be a semialgebra and let μ be a set function $\mu : \mathcal{S} \to [0, \infty]$ satisfying

1. If $S, S_i \in \mathcal{S}$ for $i = 1, 2, \dots, n$ s.t. $S = \bigsqcup_{i=1}^{n} S_i$ then $\mu(S) = \sum_{i=1}^{n} \mu(S_i)$. (That is μ is finitely additive.)
2. If $S, S_i \in \mathcal{S}$ for $i = 1, 2, \dots$ s.t. $S = \bigsqcup_{i=1}^{\infty} S_i$ then $\mu(S) \leqslant \sum_{i=1}^{\infty} \mu(S_i)$. (That is μ is sub-additive.)

Then

Carathéodory's Extension Theorem (cont.)

- (a) μ has a unique extension $\overline{\mu}$ that is a measure on the generated algebra $\overline{\mathcal{S}}$. That is $\overline{\mu}$ is a pre-measure on $\overline{\mathcal{S}}$
- (b) If $\overline{\mu}$ is σ-finite then there is a unique extension ν of μ that is a measure on $\sigma(\mathcal{S})$.
- (c) Assume that the measure ν in (b) is a probability measure. In this case, for every $\varepsilon > 0$ and for every $B \in \sigma(\mathcal{S})$ there exists an $A \in \overline{\mathcal{S}}$ such that $\mu(A \Delta B) < \varepsilon$. That is $\overline{\mathcal{S}}$ is dense in $\sigma(\mathcal{S})$ in the metric $\rho(A, B) := \nu(A \Delta B)$.

For the proof of all but the last assertion see [5, p.4].

Probability space

Definition 2.2 (Measurable space)

$(\Omega, \mathcal{F})$ is a measurable space if $\Omega \neq \emptyset$ is a set and $\mathcal{F} \subset 2^{\Omega}$ is a σ-algebra.

Definition 2.3 (Probability space)

$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space if $\mathbb{P}$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$.

π-λ systems

Definition 2.4

Let $\mathcal{P}, \mathcal{L} \subset 2^X$. We say that

1. $\mathcal{P}$ is a π-system if
$$A, B \in \mathcal{P} \Longrightarrow A \cap B \in \mathcal{P}$$
2. $\mathcal{L}$ is a λ-system if
   (i) $X \in \mathcal{L}$
   (ii) $A, B \in \mathcal{L}$ & $A \subset B \Longrightarrow B \setminus A \in \mathcal{L}$.
   (iii) $A_n \in \mathcal{L}$ & $A_n \uparrow A \Longrightarrow A \in \mathcal{L}$

Theorem 2.5

Assume that
(i) $\mathcal{P}$ *is a* π*-system,*
(ii) $\mathcal{L}$ *is a* λ*-system,*
(iii) $\mathcal{P} \subset \mathcal{L}$.
Then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

The proof is available in the Appendix A of Durrett's book [4] .

Theorem 2.6

Let $\mathcal{F}_1, \mathcal{F}_2 \subset 2^X$ *be* σ*-algebras and let* ν_1, ν_2 *be probability measures on* $\mathcal{F}_1, \mathcal{F}_2$ *respectively. Assume that*
(a) $\mathcal{P} \subset \mathcal{F}_1 \cap \mathcal{F}_2$ *is a* π*-system and*
(b) *The restriction of* ν_1 *to* $\mathcal{P}$ *agrees with the restriction of* ν_2 *to* $\mathcal{P}$.
Then the restrictions of ν_1 *and* ν_2 *to* $\sigma(\mathcal{P})$ *are the same.*

Properties of the integral

On the next three slides we use the following notation:

Notation 1

Let μ be a not necessarily finite measure on the measurable space $(\Omega, \mathcal{F})$. Let $\{f_n\}_{n=1}^{\infty}$ be sequence of real valued functions $f_n : \Omega \to \mathbb{R}$ which are measurable w.r.t. $\mathcal{F}$. When $\Omega = \mathbb{R}$ and we write $\int f(x)dx$ then we mean integration w.r.t. the Lebesgue measure. We write $L^p(\mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{R} | \int |f(x)|^p dx < \infty . \right\}$

About the definition and properties of the integral see [4, Section 1.4]. Here I mention only some of the most important theorems.

(a) Rieman-Lebesgue Lemma: Let $g \in L^1(\mathbb{R})$. Then $\lim\limits_{t \to \infty} \int g(x) \sin(tx) dx = 0$.

(b) Jensen's inequality: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and we assume that $f, \varphi \circ f \in L^1(\mu)$. Then $\varphi\left(\int f d\mu\right) \leqslant \int \varphi(f) d\mu$.

(c) Hölder's inequality Let $p, q \in (0, \infty)$ be conjugates, that is , $1/p + 1/q = 1$. Then

$$\int |fg| d\mu \leqslant \|f\|_p \cdot \|g\|_q. \quad (4)$$

When $p = q = 2$ then we obtain the Cauchy-Schwarz inequality:

$$\int |fg| d\mu \leqslant \|f\|_2 \cdot \|g\|_2. \quad (5)$$

Definition of a.s. convergence

Let $\{X_n\}_{n=1}^{\infty}$ be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is easy to see that the set $\Omega_0 := \left\{ \omega : \lim\limits_{n \to \infty} X_n \text{ exists} \right\}$ is measurable. If $\mathbb{P}(\Omega_0) = 1$ then we say that X_n converge almost surely. In this case we often write:

$$X_{\infty} := \limsup_{n \to \infty} X_n. \quad (6)$$

In this case $X_{\infty}(\omega) = \lim\limits_{n \to \infty} X_n(\omega)$ for $\mathbb{P}$-almost all ω. We express this is the following way: $X_{\infty} = \lim\limits_{n \to \infty} X_n$ a.s..

(e) Minkowski inequality Let $p \in [1, \infty]$ and $f, g \in L^p(\mu)$. Then $\|f + g\|_p \leqslant \|f\|_p + \|f\|_p$.

(f) Dominated Conv. Thm. Assume that there is a $g \in L^1(\mu)$ s.t. $|f_n| \leqslant g$ and $\lim_{n\to\infty} f_n = f$ a.e. (this means that for μ-almost all $\omega \in \Omega$ we have $\lim_{n\to\infty} f_n(\omega) = f(\omega)$.) Then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

(g) Monotone convergence thm. Let $f_n \geqslant 0$ and $f_n \uparrow f$. Then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

(h) Fatou Lemma If $f_n \geqslant 0$ then

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

Fubini Theorem

The basic reference is [4, Section 1.7] Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be two σ-finite measure spaces. Let $\Omega := X \times Y$ and let $\mathcal{F} := \mathcal{A} \times \mathcal{B}$ be the product σ-algebra which is generated by the semi-algebra

$$\mathcal{S} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}.$$

For all elements $A \times B \in \mathcal{S}$ we define $\nu(A \times B) := \mu_1(A) \cdot \mu_2(B)$. This measure ν can be extended uniquely $\mathcal{F}$. The resulted measure is called the product measure $\mu_1 \times \mu_2$.

Fubini Theorem (cont.)

Theorem 3.1 (Fubini)

Assume that $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be two σ-finite measure spaces. We assume that either $f \geqslant 0$ or $\int |f| d(\mu_1 \times \mu_2) < \infty$. Then

(8) $$\int_X \int_Y f(x,y) d\mu_2(y) d\mu_1(x) = \iint_{X \times Y} f(x,y) d\mu_1 \times \mu_2(x,y) = \int_Y \int_X f(x,y) d\mu_1(x) d\mu_2(y)$$

Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and let $(S, \mathcal{S})$ be a measurable space. A function $X : \Omega \to S$ is called measurable if

$$X^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{S}.$$

In this case we say that X is an S-valued random variable. If S is countable then X is a discrete random variable. In this case

$$p_X : S \to [0,1], \quad p_X(s) = \mathbb{P}(X = s)$$

Random variables (cont.)

is the (probability) mass function of X. The push forward measure

$$(\mathbb{P} \circ X^{-1})(A) := \mathbb{P}\left(X^{-1}(A)\right), \quad A \in \mathcal{S}.$$

is the distribution of X. Sometimes we denote the distribution of X by $\mathbb{P}_X$. Let $\mathbb{R}^*$ be the extended real line that is $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, \infty\}$ and $\mathfrak{R}^*$ is the σ-algebra generated by interval (a,b), $[-\infty, b)$, $(a, \infty]$ where $a, b \in \mathbb{R}$. If $(S, \mathcal{S}) = (\mathbb{R}^*, \mathfrak{R}^*)$ then we say that X is a random variable.

Independence

(a) Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

(b) Two r.v. X and Y are independent if for all $A, B \in \mathcal{R}$ we have $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

(c) Two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ are independent if for all $A \in \mathcal{F}, B \in \mathcal{G}$, the events A and B are independent.

(d) An infinite collection of objects (events, r.v., σ-algebras) is independent if every finite sub-collection is independent.

Independence (cont.)

(e) σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if

$$A_i \in \mathcal{F}_i \Longrightarrow \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

(f) R.v. $X_1, \ldots, X_n$ are independent if

$$B_i \in \mathcal{R} \Longrightarrow \mathbb{P}\left(\bigcap_{i=1}^{\infty} \{X_i \in B_i\}\right) = \prod_{i=1}^{\infty} \mathbb{P}(X_i \in B_i).$$

Independence (cont.)

(g) Events $A_1, \ldots A_n$ are independent if

$$I \subset \{1, \ldots, n\} \Longrightarrow \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

(h) Collection of sets $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$ is called independent if

$A_i \in \mathcal{A}_i$ and $I \subset \{1, \ldots, n\}$

$$\Longrightarrow \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Theorem 4.1 (Change of variables Theorem)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(S, \mathcal{S})$ be a measurable space. Let $f : \Omega \to S$ be measurable and $g : S \to [0, \infty]$ be a Borel measurable function. Then the change of variable formula holds:

(9) $$\int_S g d\nu = \int_\Omega (g \circ f) d\mathbb{P},$$

where ν is the push forward measure of $\mathbb{P}$ by f. That is

$$\nu(B) = \mathbb{P}(f^{-1}(B)), \quad \forall B \in \mathcal{S}.$$

An application:

Corollary 4.2

Given $(X_1, \ldots, X_n)$ random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let ν be the distribution of the vector valued random variable $(X_1, \ldots, X_n)$. That is

$$\nu := \mathbb{P}_{(X_1, \ldots, X_n)}.$$

Further, let $g : \mathbb{R}^d \to \mathbb{R}$ be a Borel measurable function which is either non-negative or bounded and $f : \Omega \to \mathbb{R}^n$ is defined by $f := (X_1, \ldots, X_n)$ then the expectation of $g(X_1, \ldots, X_n)$:

(10) $$\mathbb{E}\left[g(X_1, \ldots, X_n)\right] = \int_{\mathbb{R}^n} g d\nu.$$

Definition 4.3 (continuous r.v.)

Let X be a r.v. defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X is a continuous r.v. if there exists a non-negative function $f : \mathbb{R} \to [0, \infty)$ such that

$$\mathbb{P}(X \leqslant x) = \int_{-\infty}^{\infty} f(t) dt, \quad \forall x \in \mathbb{R}.$$

Then f is the density function of X.

Distribution functions

Let X be a random variable (r.v.) on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The cumulative distribution function (CDF) or simply distribution function is

(11) $$F(x) = F_X(x) := \mathbb{P}(X \leqslant x).$$

Remark 4.4

In some books instead of " $\leqslant$ " they write " $<$ " in (11). This does not matter when we deal with continuous r.v. however, the proper use of tables of discrete distributions is effected by this convention!

Distribution functions (cont.)

Theorem 4.5

Every CDF (cumulative distribution function) F has the following properties:

1. *F is non-decreasing*
2. *F is right continuous. That is $\lim_{y \downarrow x} F(y) = F(x)$.*
3. *$\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.*

*Conversely, if F is a function satisfying (1)-(3) then F is the CDF of a r.v. (see Homework **??**).*

Stieltjes measure functions

Definition 4.6

We say that $F : \mathbb{R} \to \mathbb{R}$ is a Stieltjes measure function if

(i) F is nondecreasing and
(ii) F is right continuous that is $\lim_{y \downarrow x} F(y) = F(x)$.

Theorem 4.7

Let F be a Stieltjes measure function. Then there exists a measure $\mu = \mu_F$ on $(\mathbb{R}, \mathfrak{R})$ such that $\mu((a, b]) = F(b) - F(a)$.

Stieltjes measure functions (cont.)

The idea of the proof.

Let $\mathcal{S}$ be the collection of semiopen intervals of the form $(a, b]$ on $\mathbb{R}$ $-\infty \leqslant a < b \leqslant \infty$. We define $\mu(a, b] = F(b) - F(a)$. Then $\mathcal{S}$ is a semialgebra ($F(-\infty) = \lim_{x\to-\infty} F(x)$ and $F(\infty) = \lim_{x\to\infty} F(x)$) and μ is a pre-measure on the generated algebra $\mathcal{A}$. Let μ^* be defined as in (2). Then μ^* is a metric outer measure so the measure μ generated in Theorem 2.1 is a Borel measure on $\mathbb{R}$ such that $\mu((a, b]) = F(b) - F(a)$. □

Stieltjes measure functions (cont.)

Note that this theorem implies that the CDF of a random variable uniquely determines the distribution of a the random variable.

Conditional Expectation

Assume that a random vector (X, Y) has the joint density function $f_{X,Y}(x, y)$. Assume that for an y the marginal probability density function we have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx > 0.$$

Then we can introduce the conditional density

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

Conditional Expectation (cont.)

Using this we can define

(12) $$\mathbb{P}(X \in A | Y = y) = \int_A f_{X|Y}(x)dx,$$

although the condition $\mathbb{P}(Y = y) = 0$. The corresponding conditional expectation is

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

Conditional Expectation (cont.)

This is a random variable which is a function of Y. In general: We learned in the course Stochastic Processes that for any r.v. $X, Y_1, \ldots, Y_n$ there exists a Borel function g s.t.

(14) $$\mathbb{E}[X|Y_1, \ldots, Y_n] = g(Y_1, \ldots, Y_n).$$

As we have seen earlier this means that

(15) $$\mathbb{E}[X|Y_1, \ldots, Y_n] \in \sigma(Y_1, \ldots, Y_n).$$

In the lights of the previous comments, we define the conditional expectation as follows:

Conditional Expectation (cont.)

Definition 5.1 (Conditional Expectation)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$ and let X be an L^1 r.v.. We say that Z is a version of the conditional expectation of X w.r.t. $\mathcal{G}$, $\mathbb{E}[X|\mathcal{G}]$ if:

(a) $Z \in \mathcal{G}$ and

(b) $\int_A X d\mathbb{P} = \int_A Z d\mathbb{P}$ for every $A \in \mathcal{G}$.

Conditional Expectation (cont.)

We have seen that

Theorem 5.2

(i) *There exists a conditional expectation $\mathbb{E}[X|\mathcal{G}]$ for any L^1 r.v. X and $\mathcal{G} \subset \mathcal{F}$ sub-σ-algebra.*

(ii) *Any two versions of $\mathbb{E}[X|\mathcal{G}]$ are equal $\mathbb{P}$-a.s..*

The construction of $\mathbb{E}[X|\mathcal{G}]$ by the Radon Nikodym theorem: Suppose that X is an L^1 r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. We introduce the signed measure ν on $(\Omega, \mathcal{F})$:

$$\nu(B) := \int_B X d\mathbb{P}.$$

Conditional Expectation (cont.)

Then

(16) $\nu \ll \mathbb{P}$.

Let $\nu^{\mathcal{G}}$ be the restriction of the measure ν from $\mathcal{F}$ to $\mathcal{G}$ and similarly let $\mathbb{P}^{\mathcal{G}}$ be the restriction of $\mathbb{P}$ from $\mathcal{F}$ to $\mathcal{G}$. Then $\mu^{\mathcal{G}}$ and $\mathbb{P}^{\mathcal{G}}$ are measures on $(\Omega, \mathcal{G})$ and by (16)

$$\nu^{\mathcal{G}} \ll \mathbb{P}^{\mathcal{G}}.$$

Conditional Expectation (cont.)

Let Z be the Radon-Nikodym derivative

$$Z = \frac{d\nu^{\mathcal{G}}}{d\mathbb{P}^{\mathcal{G}}} \in L^1(\Omega, \mathcal{G}, \mathbb{P}^{\mathcal{G}}).$$

Then $\forall A \in \mathcal{G}$:

(17) $$\int_A X d\mathbb{P} = \nu(A) = \nu^{\mathcal{G}}(A) = \int_A Z d\mathbb{P}^{\mathcal{G}} = \int_A Z d\mathbb{P},$$

Conditional Expectation (cont.)

since $A \in \mathcal{G}$ and Z is $\mathcal{G}$-measurable. If Z_1 and Z_2 satisfy (18) the $Z_1(\omega) = Z_2(\omega)$ for $\mathbb{P}$ almost all $\omega \in \Omega$. In this way a r.v. Z satisfying (18) is a version of $\mathbb{E}[X|\mathcal{G}]$.

Conditional Expectation (cont.)

Example 5.3 (This Example is from [17])

Let $\Omega := \{a, b, c, d, e, f\}$, $\mathcal{F} = 2^{\Omega}$ and $\mathbb{P}$ is the uniform distribution on Ω. The r.v. X, Y, Z are defined by

$$X \sim \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 5 & 5 & 7 \end{pmatrix}, Y \sim \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix}$$

$$Z \sim \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 2 & 2 \end{pmatrix}$$

Then $\mathbb{E}[X|\sigma(Y)]$ and $\mathbb{E}[X|\sigma(Z)]$ are given on the next slides.

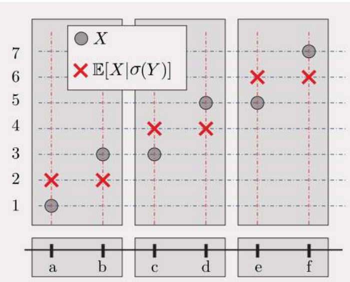

Figure: Figure for Example 5.3. The Figure is from [17]

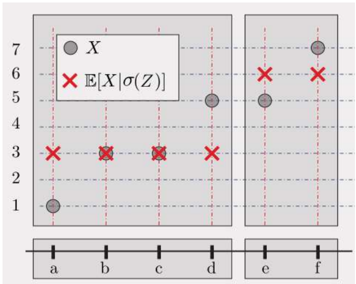

Figure: Figure for Example 5.3. The Figure is from [17]

Properties of the conditional expectation Here we follow the Zitkovicz's Lecture notes [17]. All the proofs are available either there or in [5].

Let $X, Y, \{X_n\}_{n=1}^{\infty}$ be r.v. on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Further, let $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ be sub-σ-algebras of $\mathcal{A}$.
(a) Linearity: $\mathbb{E}[a \cdot X + b \cdot Y|\mathcal{F}] = \mathbb{E}[a \cdot X|\mathcal{F}] + \mathbb{E}[b \cdot Y|\mathcal{F}]$
(b) Monotnicity If $X \leqslant Y$ then $\mathbb{E}[X|\mathcal{F}] \leqslant \mathbb{E}[Y|\mathcal{F}]$ a.s.
(c) If $X \in \mathcal{F}$ then $\mathbb{E}[X|\mathcal{F}] = X$.
(d) Conditional Jensen: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and $\mathbb{E}[|\varphi(X)|] < \infty$. Then

$$\mathbb{E}[\varphi(X)|\mathcal{F}] \geqslant \varphi(\mathbb{E}[X|\mathcal{F}]), \text{ a.s.}$$

(e) L^p-non-expansive: Let $p \in [1, \infty]$. If $X \in L^p$ then $\mathbb{E}[X|\mathcal{F}] \in L^p$ and

$$\|\mathbb{E}[X|\mathcal{F}]\|_{L^p} \leqslant \|\mathbb{E}[X]\|_{L^p}$$

(f) Pulling out what is known: Let $Y \in \mathcal{F}$ and $XY \in L^1$ then

(18) $$\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}].$$

(g) L^2-projection Assume that $X \in L^2(\mathcal{A})$. Then minimum of

$$\min_{Z \in L^2(\mathcal{F})} \mathbb{E}\left[(X - Z)^2\right]$$

is attained at $Z = \mathbb{E}[X|\mathcal{F}]$. That is $\mathbb{E}[X|\mathcal{F}]$ is the orthogonal projection of X to $L^2(\mathcal{F})$ if $X \in L^2(\mathcal{A})$.

(h) Tower property If $\mathcal{F} \subset \mathcal{G}$ then

(19) $$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}.]$$

(i) Irrelevance of independent information If $\mathcal{F}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then

(20) $$\mathbb{E}[X|\sigma(\mathcal{F}, \mathcal{G})] = \mathbb{E}[X|\mathcal{G}]$$

In particular

(21) If X is independent of $\mathcal{F}$ then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$ a.s.

(j) Conditional monotone convergence theorem If $0 \leqslant X_n \leqslant X_{n+1}$ a.s. for all n and $X_n \to X \in L^1$ a.s. then

$$\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}].$$

(k) Conditional Fatau Lemma Let $X_n \geqslant 0$ a.s. for $\forall n$ and assume that $\liminf_{n\to\infty} X_n \in L^1$. Then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n|\mathcal{F}\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{F}] \text{ a.s.}$$

(l) Cond. dominated convergence Theorem Assume that

- $\exists Z \in L^1$ s.t. $\forall n$, $|X_n| \leqslant Z$ a.s.

- $X_n \to X$ a.s.

Then

(22) $$\mathbb{E}[X_n|\mathcal{F}] \to \mathbb{E}[X|\mathcal{F}]$$ both in L^1 and a.s.

(m) Cond. expectation for countable partition generated sub-σ-algebra Let $\{\Omega_1, \Omega_2, \dots\}$ be a partition of Ω. We define $\mathcal{F} := \sigma(\Omega_1, \Omega_2, \dots)$. Then

(23) $$\mathbb{E}[X|\mathcal{F}](\omega) = \frac{\mathbb{E}[X; \Omega_i]}{\mathbb{P}(\Omega_i)}, \text{ for } \omega \in \Omega_i.$$

If $\mathcal{F} = \{\emptyset, \Omega\}$ then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$.

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \le p \le 1$	$\binom{n}{x} p^x(1-p)^{n-x}$, $x = 0, 1, \dots, n$	$(pe^t + 1 - p)^n$	np	$np(1 - p$
Poisson with parameter $\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \dots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$, $x = 1, 2, \dots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \le p \le 1$	$\binom{n-1}{r-1} p^r(1-p)^{n-r}$, $n = r, r+1, \dots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Figure: Figure is from [15]

	Probability mass function, $f(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

Figure: Figure is from [15]

[1] P. Billingsley
Convergence of probability measures
Wiley, 1968

[2] B. Driver
Analysis tools with examples
Lecturenotes, 2012. Click here.

[3] R. Durrett
Essentials of Stochastic Processes, Second edition
Springer, 2012. Click here

[4] R. Durrett
Probability: Theory with examples, 4th edition
Cambridge University Press, 2010.

[5] R. Durrett
Probability: Theory and Examples
Click here

[6] D.H. Fremlin
Measure Theory Volume I
Click here

[7] D.H. Fremlin
Measure Theory Volume II
Click here

[8] O. van Gaans
Probability measures on metric spaces
Click here

[9] S. Karlin, H.M. Taylor
A first course in stochastic processes
Academic Press, New York, 1975

[10] S. Karlin, H.M. Taylor
Sztochasztikus Folyamatok
Gondolat, Budapest, 1985

[11] S. Karlin, H.M. Taylor
A second course in stochastic processes
, Academic Press, 1981

[12] G. Lawler
Intoduction to Stochastic Processes
Chapmann & Hall 1995.

[13] D.A. Levin, Y. Peres,E.L. Wilmer
Markov chains and mixing times
American Mathematical Society, 2009.

[14] P. Mattila Geometry of sets and measure in Euclidean spaces. Cambridge, 1995.

[15] S. Ross
A First Course in Probability, 6th ed.
Prentice Hall, 2002

[16] D. Williams
Probability with Martingales
Cambridge 2005

[17] G. Zitkovic *Theory of Probabilty I, Lecture 7*
Click here