Stochastic processes

Károly Simon

This lecture is based on
Essentials of Stochastic processes
book of Rick Durrett

Department of Stochastics
Institute of Mathematics
Technical University of Budapest
www.math.bme.hu/~simonk

2020 File E

Review: conditional distributions

In the course Probability I some of you have studied conditional distributions in [1, Chapter 6.5.]. For example, we have jointly continuous random variables X, Y, whose joint density function $f(x, y)$. Then the density functions of the marginals are:

(1)

$$f_Y(y_0) = \int_{x=-\infty}^{\infty} f(x, y_0)dx, \quad f_X(x_0) = \int_{y=-\infty}^{\infty} f(x_0, y)dx.$$

Review: conditional distributions (cont.)

The conditional density function of X with respect to the event $\{Y = y\}$ (of zero probability):

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

Conditional expectation (cont.)

This comes from that

$$F_{X|Y}(x|y) \approx \mathbb{P}(X < x | Y \in [y, y + \Delta y)) = \frac{F(x, y + \Delta y) - F(x, y)}{\mathbb{P}(Y \in [y, y + \Delta y))} = \frac{\frac{F(x,y+\Delta y) - F(x,y)}{\Delta y}}{\frac{\mathbb{P}(Y \in [y, y+\Delta y))}{\Delta y}} \approx \frac{F'_y(x, y)}{f_Y(y)}.$$

We get $f_{X|Y}(x|y)$ from this by differentiating $F_{X|Y}(x|y)$ with respect to x:

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

Conditional expectation (cont.)

Accordingly:

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

if $f_Y(y)>0$. Hungarian students learnt it in [1, Chapter 7.3].
If we do not fix Y, $\mathbb{E}\left[X|Y\right]$ is a r.v. too.

Conditional expectation (cont.)

Lemma 1.1

(See [1, chapter 7.3]) Let $u, v : \mathbb{R} \to \mathbb{R}$ *be Borel measurable functions. Then*

(a) $\mathbb{E}\left[u(X) \cdot v(Y) | Y\right] = v(Y) \cdot \mathbb{E}\left[u(X) | Y\right]$, *where* u, v *are Borel measurable func.*

(b) *For Borel measurable func.* $g : \mathbb{R}^2 \to \mathbb{R}$*:*

$$\mathbb{E}\left[g(X, Y)\right] = \mathbb{E}\left[\mathbb{E}\left[g(X, Y) | Y\right]\right] \tag{2}$$

Conditional expectation (cont.)

This is the tower property. In special case:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]. \tag{3}$$

Example 2.1

Let us compute $\mathbb{E}[X|Y]$, if

$$f(x,y) = \frac{e^{-x/y}e^{-y}}{y}, \text{ if } 0 < x, y < \infty.$$

Solution:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{(1/y)e^{-x/y}e^{-y}}{\int_{-\infty}^{\infty} (1/y)e^{-x/y}e^{-y}\,dx} = \frac{e^{-x/y}}{y}.$$

So $\mathbb{E}[X|Y=y] = \int_0^\infty \frac{x}{y} e^{-x/y} dx = y$. From here $\mathbb{E}[X|Y] = Y$.

Example 2.2

Let $T := [0,1] \times [0,2]$.

$$f(x,y) = \begin{cases} \frac{1}{4}(2x+y), & \text{if } (x,y) \in T; \\ 0, & \text{otherwise.} \end{cases}$$

Let us compute function $g : \mathbb{R} \to \mathbb{R}$, for which $\mathbb{E}[X|Y] = g(Y)$.

Solution: $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{4}(1+y)$.

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2x+y}{1+y} \text{ if } (x,y) \in T.$$

So

$$\begin{aligned}
\mathbb{E}[X|Y=y] &= \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx \\
&= \int_{-\infty}^{\infty} x \cdot \frac{2x+y}{1+y}dx \\
&= \frac{1}{6} \cdot \frac{4+3y}{1+y}.
\end{aligned}$$

Let $g(y) := \frac{1}{6} \cdot \frac{4+3y}{1+y}$. Then from above:

$$\mathbb{E}[X|Y] = g(Y).$$

Example 2.3

Let $X \sim \mathrm{Uniform}(0,1)$ and $Y|X = x \sim \mathrm{Uniform}(0,x)$, if $0 < x < 1$. Then $\mathbb{E}[X|Y] = \frac{Y-1}{\ln Y}$.

Namely, we have learnt that $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$. Hence,

$$f(x,y) = f_{Y|X}(y|x) \cdot f_X(x) = \frac{1}{x} \cdot 1, \quad \text{if } 0 < y < x < 1$$

Let $0 < y < 1$. Then

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x = \int_y^1 \frac{1}{x}\mathrm{d}x = -\ln y.$$

If $y \notin (0,1)$ then $f_Y(y) = 0$. Then for $0 < y < x < 1$:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1/x}{-\ln y}$$

If $y \notin (0,1)$ then the conditional density function $f_{X|Y}(x|y)$ does not make sense. If $y \in (0,1)$ but $x \notin (0,1)$ then $f_{X|Y}(x|y) = 0$. Hence,

$$\mathbb{E}[X|Y = y]$$

$$= \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\mathrm{d}x = \int_y^1 x \cdot \frac{1/x}{-\ln y}\mathrm{d}x = \frac{y-1}{\ln y}.$$

That is $\mathbb{E}[X|Y] = \frac{Y-1}{\ln Y}$.

Lemma 2.4

Let $X, Y_1, \ldots, Y_n$ be random variables. Then there exists a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$, such that

$$\mathbb{E}[X|Y_1, \ldots, Y_n] = g(Y_1, \ldots, Y_n). \tag{4}$$

We have seen this for case $n = 1$ and (X, Y) is jointly continuous. But in the general case, we can use the same argument to prove the statement.

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Some measure theory

Definition 2.5

- Let $\mathcal{B}^n$ be the Borel σ-algebra on $\mathbb{R}^n$-en. If $n = 1$, then we simply write $\mathcal{B}$.
- Let $\xi : \Omega \to \mathbb{R}^n$. If $\xi^{-1}(\mathcal{B}^n) \subset \mathcal{A}$ then we say that ξ is measurable with respect to (w.r.t.) $\mathcal{A}$ and we also say that ξ is a random variable,.
- σ-algebra generated by r.v. $\xi_1, \dots, \xi_n$ (denoted by $\sigma(\xi_1, \dots, \xi_n)$), is the smallest σ-algebra, for which all the r.v. $\xi_1, \dots, \xi_n$ are measurable.
- For a r.v. η, $\eta \in \mathcal{A}$ means that η is measurable with respect to $\mathcal{A}$.

Some measure theory

Theorem 2.6

- η *and* $\xi_1, \dots, \xi_n$ *r.v.* $(\Omega, \mathcal{A}, \mathbb{P})$.
- $\mathcal{F} := \sigma(\xi_1, \dots, \xi_n)$.

$\eta \in \mathcal{F} \iff \exists g : \mathbb{R}^n \to \mathbb{R}$, *Borel measurable function, for which*

$$\eta(\omega) = g(\xi_1(\omega), \dots, \xi_n(\omega)). \tag{5}$$

Some measure theory (cont.)

The proof can be found in [4, Chapter 3.6]. For further readings on measure theory I suggest to click on the next line:
Durrett, Probability: Theory and Examples, Apendix
or type into an Internet browser: `https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf`

Some measure theory (cont.)

Remark 2.7

Now we use the notation of Theorem 2.6. Let $A \in \mathcal{A}$ be an event. Then

$$
\begin{aligned}
\text{(6)}\quad A \in \mathcal{F} \quad &\Longleftrightarrow \quad \mathbb{1}_A \in \mathcal{F} \\
&\Longleftrightarrow \quad \exists g,\ \mathbb{1}_A(\omega) = g(\xi_1(\omega), \dots, \xi_n(\omega)),
\end{aligned}
$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel measurable function.

A property of conditional expected value

Notation: $\mathbb{E}[X;A] := \mathbb{E}[X \cdot \mathbb{1}_A]$, where $A \in \mathcal{A}$.

Theorem 2.8

(a) $\mathbb{E}[X|Y] \in \sigma(Y)$

(b) $\forall A \in \sigma(Y)$, $\mathbb{E}[X;A] = \mathbb{E}[\mathbb{E}[X|Y];A]$

Part (a) comes from Theorem 2.6.

Proof of Part (b)

Let us fix arbitrary real numbers $a < b$ and let $A = Y^{-1}([a,b])$. Obviously it is enough to prove part (b) for this kind of sets.

A property of conditional expected value (cont.)

Proof of Part (b) (cont.)

Let $g(x,y) := x \cdot \mathbb{1}_{[a,b]}(y)$. Below we apply parts (b) and then (a) of Lemma 1.1:

$$\begin{aligned} \mathbb{E}[X;A] &= \mathbb{E}[g(X,Y)] \\ &= \mathbb{E}[\mathbb{E}[g(X,Y)|Y]] \\ &= \mathbb{E}\left[\mathbb{E}\left[X \cdot \mathbb{1}_{[a,b]}(Y)|Y\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{[a,b]}(Y) \cdot \mathbb{E}[X|Y]\right] \\ &= \mathbb{E}[\mathbb{E}[X|Y];A]. \end{aligned}$$

Conditioning for σ-algebras

We would like to define the conditional expectation for σ-algebras. We can imagine this as that the conditional expectation value for the r.v. Y (e.g. in Theorem 2.8) is a conditional expectation for $\sigma(Y)$-algebra.

Aim: To extend this definition to an arbitrary (so not only continuous) r.v's conditional expectation for an arbitrary σ-algebra $\mathcal{F} \subset \mathcal{A}$.

Example 2.9 (This Example is from [13])

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.

$\mathbb{E}[X|\sigma(Y)]$

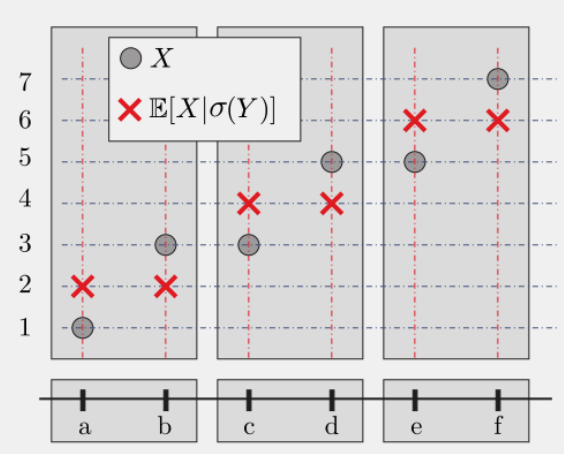


Figure: Figure for Example 2.9. The Figure is from [13]

$\mathbb{E}[X|\sigma(Y)]$

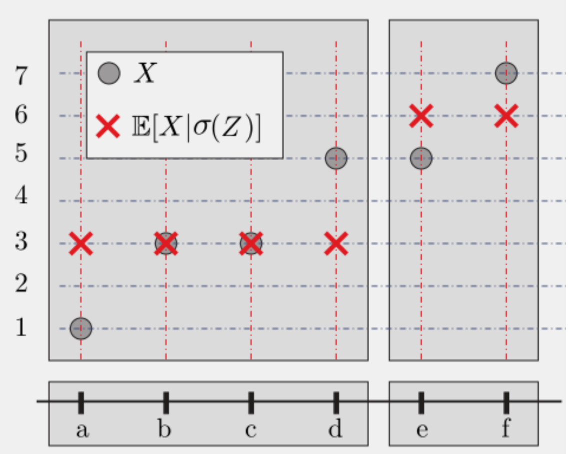
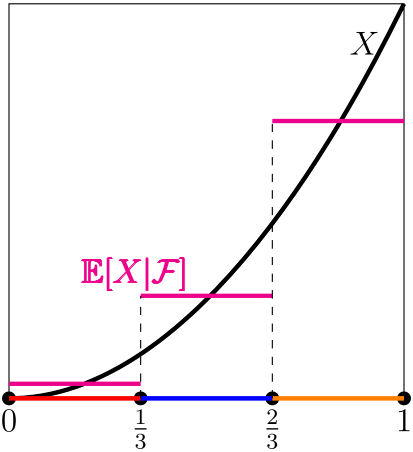


Figure: Figure for Example 2.9. The Figure is from [13]

$\mathcal{F}$ is generated by $\left\{[0, \frac{1}{3}), [\frac{1}{3}, \frac{2}{3}), [\frac{2}{3}, 1)\right\}$



Definition 2.10 (Conditional expectation with respect to a σ-algebra)

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let

- X be a r.v. for which: $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty$.
- $\mathcal{F}$ be sub-σ-algebra of $\mathcal{A}$.

Conditional expectation of X with respect to $\mathcal{F}$ (denoted by $\mathbb{E}[X|\mathcal{F}]$) is a r.v. Z which satisfies:

(a) $Z \in \mathcal{F}$, (Z is measurable for $\mathcal{F}$) and

(b) $\forall A \in \mathcal{F}$:

$$\int_A X d\mathbb{P} = \int_A Z d\mathbb{P}. \tag{7}$$

Remark 2.11

- Parts (a) and (b) from above are generalizations of Theorem 2.8's parts (a) and (b), in such sense, that in Theorem 2.8 $\mathcal{F} = \sigma(Y)$. So $\mathbb{E}[X|\mathcal{F}]$ is the generalization of $\mathbb{E}[X|Y]$. (Cf. Theorem 2.8.)
- If a r.v. Z satisfies conditions (a) and (b) above then we say that Z is a version of $\mathbb{E}[X|\mathcal{F}]$.
- Our first aim is to prove, that $\mathbb{E}[X|\mathcal{F}]$ exists and unique (up to measure zero). We do this by applying the Radon-Nikodym Theorem. But for this, we need some review from measure theory. (Now we follow book [3, A.8].)

Definition 3.1

On measurable space $(\Omega, \mathcal{F})$:

1. μ is a measure, if
   - $\mu : \mathcal{F} \to [0, \infty], \mu(\emptyset) = 0$.
   - If $E = \bigcup_{i=1}^{\infty} E_i$ disjoint union, then $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$.
2. ν is a σ-finite measure if there exist sets $A_n \in \mathcal{F}$, s.t.
   - $\Omega = \bigcup_{n=1}^{\infty} A_n$
   - $\nu(A_n) < \infty$.

Definition 3.2 (Signed measure)

Given a measurable space $(\Omega, \mathcal{F})$ (Ω is a set, on which $\mathcal{F}$ is a σ-algebra). α is a signed measure on $(\Omega, \mathcal{F})$, if

- $\alpha(E) \in (-\infty, \infty]$, $\forall E \in \mathcal{F}$.
- $\alpha(\emptyset) = 0$.
- If $E = \cup E_i$ is disjoint union, then $\alpha(E) = \sum_i \alpha(E_i)$, in such sense, that
  1. If $\alpha(E) < \infty$, then there is absolute convergence,
  2. If $\alpha(E) = \infty$, then $\sum_i \alpha(E_i)^- < \infty$ and $\sum_i \alpha(E_i)^+ = \infty$.

Jordan's Theorem: $\exists \alpha_1, \alpha_2$ are positive measures, that $\alpha_1 \perp \alpha_2$ and $\alpha = \alpha_1 - \alpha_2$.

Absolute continuity of measures

Let

- μ be a finite or σ-finite measure on $\mathcal{F}$
- ν be a finite, signed measure on $\mathcal{F}$.

We say that measure ν is absolute continuous for μ ($\nu \ll \mu$), if

$$\forall C \in \mathcal{A} : \mu(C) = 0 \Rightarrow \nu(C) = 0.$$

Review of measure theory

Theorem 3.3 (Radon-Nikodym)

1. $(\Omega, \mathcal{F})$ *probability space.*
2. μ σ*-finite,* ν *a signed measure on* $\mathcal{F}$*.*
3. $\nu \ll \mu$ *on* $\mathcal{F}$*.*

Then $\exists f \in \mathcal{F}$*, s.t.*

(a) $\int |f(\omega)| d\mu(\omega) < \infty$,

(b) $\nu(C) = \int_C f(\omega) d\mu(\omega), \forall C \in \mathcal{F}$.

(c) *If* $f_1, f_2 \in \mathcal{F}$ *satisfy (a) and (b), then* $f_1(\omega) = f_2(\omega)$*, a.e.* $\omega \in \Omega$*.*

We denote function f by

$$f = \frac{d\nu}{d\mu}$$

Function f is called Radon-Nikodym derivative of measure ν with respect to μ.

Definition of conditional expectation

Let ξ be an integrable r.v. on $(\Omega, \mathcal{A}, \mathbb{P})$ ($\int |\xi(\omega)| d\omega < \infty$), and let $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra.

Now we define conditional expectation of ξ with respect to σ-algebra $\mathcal{F}$, $\mathbb{E}\left[\xi|\mathcal{F}\right]$.

In most cases $\mathcal{F}$ gives the information we have. (Recall Theorem 2.6.) Assuming $\mathcal{F}$ means that based on the information we have, the best estimate for the value of X is the to-be-defined $\mathbb{E}\left[\xi|\mathcal{F}\right]$.

Definition of conditional expectation

To define $\mathbb{E}[\xi|\mathcal{F}]$, first let us introduce the signed measure μ_ξ on $\mathcal{A}$:

$$\mu_\xi(B) := \int_B \xi(\omega) d\mathbb{P}(\omega), \quad B \in \mathcal{A}. \tag{8}$$

Obviously μ_ξ is a signed measure. From the definition:

$$\mu_\xi \ll \mathbb{P}. \tag{9}$$

If we restrict both μ_ξ and $\mathbb{P}$ to $\mathcal{F}$, we get measures $\mu|_\mathcal{F}$ and $\mathbb{P}|_\mathcal{F}$. Absolute continuity of formula (9) is also true for restricted measures:

$$\mu_\xi|_\mathcal{F} \ll \mathbb{P}|_\mathcal{F}. \tag{10}$$

Definition of conditional expectation

Consider the following Radon-Nikodym derivative

$$f := \frac{d\mu|_{\mathcal{F}}}{d\mathbb{P}|_{\mathcal{F}}}$$

Then

(a) $f \in \mathcal{F}$ and

(b) $\forall E \in \mathcal{F}$: $\int_E f d\mathbb{P} = \mu_\xi(E) = \int_E \xi d\mathbb{P}$.

Observe that: conditions (a) and (b) above are the same as conditions (a) and (b) in Definition 2.10. Hence,

Definition of conditional expectation

- conditional expected value $\mathbb{E}\left[\xi|\mathcal{F}\right]$ exists,
- $\mathbb{E}\left[\xi|\mathcal{F}\right]$ a.e. equals to Radon-Nikodym derivative $\frac{d\mu|_{\mathcal{F}}}{d\mathbb{P}|_{\mathcal{F}}}$
- From mod 0 uniqueness of Radon-Nikodym derivative $\mathbb{E}\left[\xi|\mathcal{F}\right]$ is unique in the same sense.
- Radon-Nikodym derivative $\frac{d\mu|_{\mathcal{F}}}{d\mathbb{P}|_{\mathcal{F}}}$ is a version of conditional expected value $\mathbb{E}\left[\xi|\mathcal{F}\right]$.

Definition of conditional expectation

Definition 4.1 (Conditional probability)

Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. For every $A \in \mathcal{A}$ conditional probability of A with respect to (w.r.t.) the σ-algebra $\mathcal{F}$:

$$\mathbb{P}\left(A|\mathcal{F}\right) := \mathbb{E}\left[\mathbb{1}_A|\mathcal{F}\right]. \tag{11}$$

Properties of conditional expectation I

(a) **Linearity**:
$\mathbb{E}\left[aX+bY|\mathcal{F}\right]=a\mathbb{E}\left[X|\mathcal{F}\right]+b\mathbb{E}\left[Y|\mathcal{F}\right]$

(b) **Monotonity**:
If $X\leq Y$, then $\mathbb{E}\left[X|\mathcal{F}\right]\leq\mathbb{E}\left[Y|\mathcal{F}\right]$.

(c) **Csebisev inequality**:

$$\mathbb{P}\left(|X|\geq a|\mathcal{F}\right)\leq a^{-2}\mathbb{E}\left[X^2|\mathcal{F}\right]. \tag{12}$$

(d) **Monoton convergence theorem**: Let us assume, that $X_n\geq 0,\ X_n\uparrow X,\ \mathbb{E}\left[X\right]<\infty$ then

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

Properties of conditional expectation II

(e) Applying the above for $Y_1 - Y_n$: If $Y_n \downarrow Y$, $\mathbb{E}\left[|Y_1|\right], \mathbb{E}\left[|Y|\right] < \infty$, then $\mathbb{E}\left[X_n|\mathcal{F}\right] \downarrow \mathbb{E}\left[X|\mathcal{F}\right].$

(f) **Jensen inequality**: If φ is convex, $\mathbb{E}\left[|X|\right], \mathbb{E}\left[|\varphi(X)|\right] < \infty$, then

$$\varphi(\mathbb{E}\left[X|\mathcal{F}\right]) \leq \mathbb{E}\left[\varphi(X)|\mathcal{F}\right]. \tag{13}$$

(g) **Conditional Cauchy Schwarz:**

$$\mathbb{E}\left[XY|\mathcal{F}\right]^2 \leq \mathbb{E}\left[X^2|\mathcal{F}\right]\mathbb{E}\left[Y^2|\mathcal{F}\right]. \tag{14}$$

Properties of conditional expectation III

(h) $X \to \mathbb{E}[X|\mathcal{F}]$ is a contraction on L^p, if $p \geq 1$:

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{F}]|^p] \leq \mathbb{E}[|X|^p]$$

(i) If $\mathcal{F}_1 \subset \mathcal{F}_2$, then

1. $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]|\mathcal{F}_2] = \mathbb{E}[X|\mathcal{F}_1]$
2. $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1] = \mathbb{E}[X|\mathcal{F}_1]$

So always the more primitive σ-algebra wins.

(j) If $X \in \mathcal{F}$, $\mathbb{E}[|Y|], \mathbb{E}[|XY|] < \infty$, then

$$\mathbb{E}\left[X \cdot Y|\mathcal{F}\right] = X \cdot \mathbb{E}[Y|\mathcal{F}]. \tag{15}$$

Properties of conditional expectation IV

(k) **$\mathbb{E}[X|\mathcal{F}]$ as projection**: Let us assume, that $\mathbb{E}\left[X^2\right] < \infty$. Then $\mathbb{E}[X|\mathcal{F}]$ is the orthogonal projection of X to $L^2(\Omega, \mathcal{F}, \mathbb{P})$. In other words:

$$\mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{F}])^2\right] = \min_{Y \in \mathcal{F}} \mathbb{E}\left[(X - Y)^2\right].$$

(l) **$X \to \mathbb{E}[X|\mathcal{F}]$ is self-adjoint on $L^2(\Omega, \mathcal{A}, \mathbb{P})$:**

$$\begin{aligned} \mathbb{E}[X \cdot \mathbb{E}[Y|\mathcal{F}]] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}] \cdot \mathbb{E}[Y|\mathcal{F}]] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}] \cdot Y]. \end{aligned} \tag{16}$$

Properties of conditional expectation V

Let us define conditional variation w.r.t. σ-algebra (see [1, Def. 7.35] and [1, Statement 7.36]):

$$\mathrm{Var}(X|\mathcal{F}) := \mathbb{E}\left[X^2|\mathcal{F}\right] - \mathbb{E}\left[X|\mathcal{F}\right]^2.$$

Then

(m) $\mathrm{Var}\,(X) = \mathbb{E}\left[\mathrm{Var}(X|\mathcal{F})\right] + \mathrm{Var}\left(\mathbb{E}\left[X|\mathcal{F}\right]\right).$

(n) $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$ is disjoint union and $\mathbb{P}(\Omega_i) > 0$. Let $\mathcal{F}$ be the σ-algebra generated by $\{\Omega_i\}_{i=1}^{\infty}$. Then for a r.v. X:

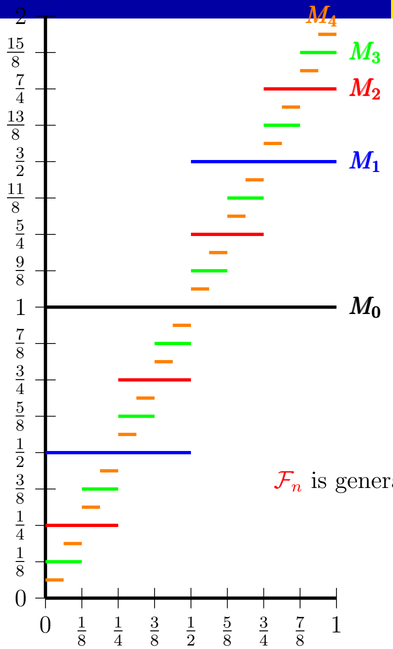
$$\mathbb{E}\left[X|\mathcal{F}\right] = \sum_i \frac{\mathbb{E}\left[X;\Omega_i\right]}{\mathbb{P}(\Omega_i)} \cdot \mathbb{1}_{\Omega_i}.$$

Properties of conditional expectation VI

(p) **Bayes's formula:** Let $F \in \mathcal{F}$ and $A \in \mathcal{A}$. Then

$$\mathbb{P}(F|A) = \frac{\int\limits_F \mathbb{P}(A|\mathcal{F})}{\int\limits_\Omega \mathbb{P}(A|\mathcal{F})}. \tag{17}$$

Is is easy to see, that this statement gives Bayes-theorem, in the case, when $\mathcal{F}$ is generated by a partition.



$\Omega = [0,1]$ and $\mathbb{P} = \mathcal{L}\mathrm{eb}|_{[0,1]}$

if $x = \sum_{n=1}^{\infty} x_n 2^{-n}$, $x_n \in \{0,1\}$

let $\theta_k = \begin{cases} 1, & \text{if } x_k = 1; \\ -1, & \text{if } x_k = 0. \end{cases}$

From this, $M_n(x) = 1 + \sum_{k=1}^{n} \theta_k 2^{-k}$

$\mathcal{F}_n$ is generated by $\left\{ \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) : i = 0, \dots 2^n - 1 \right\}$

$M_n \in \mathcal{F}_n$

$\mathbb{E}\left[M_{n+1} | \mathcal{F}_n \right] = M_n$

Definition 5.1

- An increasing sequence of σ-algebras $\mathcal{F}_n$ is called **filtration**.
- X_n is **adapted** to $\mathcal{F}_n$, if $X_n \in \mathcal{F}_n$, $\forall n$.
- (X_n) is a **martingale** for filtration $\mathcal{F}_n$, if
  - (a) $\mathbb{E}\left[|X_n|\right] < \infty$
  - (b) X_n is adapted to $\mathcal{F}_n$,
  - (c) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$, $\forall n \geq 1$.

If (a) and (b) are satisfied, but $=$ of (c) is replaced by

(c') $\leq$, then (X_n) is a **supermartingale**,

(c'') $\geq$, then (X_n) is a **submartingale**.

Example 5.2

Let us imagine a player, who plays a fair game (with expected value 0) very many times. Let M_n be his/her winning after the n^{th} game (or losing if M_n is negative) and let Y_n be the outcome of the n^{th} game and let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Then (M_n) is a martingale for $\mathcal{F}_n$.

Example 5.3

We throw a regular coin many times. Let the outcome of the n^{th} throw be $\xi_n = 1$ if it's head and $\xi_n = -1$ if it's tail. Let $X_n := \xi_1 + \cdots + \xi_n$ and $\mathcal{F}_n := \sigma\{\xi_1, \ldots, \xi_n\}$ if $n \geq 1$ and $X_0 = 0$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \underbrace{\mathbb{E}[X_n|\mathcal{F}_n]}_{X_n} + \underbrace{\mathbb{E}[\xi_{n+1}|\mathcal{F}_n]}_{0} = X_n.$$

So X_n is a martingale for $\mathcal{F}_n$.

Example 5.4

Let $X_1, \dots, X_n$ i.i.d. $\mathbb{E}[X_i] = \mu$ and $S_n := S_0 + X_1 + \cdots + X_n$ be a random walk. Then $M_n := S_n - n\mu$ is a martingale for $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$. Namely: $M_{n+1} - M_n = X_{n+1} - \mu$ is independent of $X_n, \dots, X_1, S_0$, so

$$\mathbb{E}[M_{n+1} - M_N | \mathcal{F}_n] = \mathbb{E}[X_{n+1}] - \mu = 0.$$

So

$$\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n.$$

If $\mu \leq 0$, then S_n supermartingale and if $\mu \geq 0$, then S_n submartingale.

Theorem 5.5

Let X_n be a MC, whose transition matrix is $\mathbf{P} = (p(x,y))_{x,y}$. Let us assume, that for a function $f : S \times \mathbb{N} \to \mathbb{R}$:

$$f(x,n) = \sum_y p(x,y) f(y,n+1). \tag{18}$$

Then $M_n = f(X_n, n)$ is a martingale for $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. In the special, when

$$h(x) = \sum_y p(x,y) h(y), \tag{19}$$

then $h(X_n)$ is martingale for $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

Proof.

$$\begin{aligned}\mathbb{E}\left[f(X_{n+1}, n+1)|\mathcal{F}_n\right] &= \sum_y p(X_n, y) f(y, n+1) \\ &= f(X_n, n).\end{aligned}$$

□

Example 5.6 (Gambler's Ruin)

Let $X_1, X_2, \dots$ i.i.d. s.t. for some $p \in (0,1)$, $p \neq 1/2$:

$$\mathbb{P}(X_i = 1) = p \text{ and } \mathbb{P}(X_i = -1) = q = 1 - p.$$

Let $S_n = S_0 + X_1 + \cdots X_n$. Then

$$M_n := \left(\frac{q}{p}\right)^{S_n}$$

is a martingale.

This comes from that $h(x) = \left(\frac{q}{p}\right)^x$ satisfies condition (19). Hence we can apply Theorem 5.5.

Example 5.7 (Simple symmetric random walk)

$Y_1, Y_2, \ldots$ i.i.d. $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = 1/2$.
$S_n = S_0 + Y_1 + \cdots + Y_n$. Then $M_n := S_n^2 - n$ is a martingale for $\sigma(Y_1, \ldots, Y_n)$.

Namely: we must show, that for $f(x, n) = x^2 - n$ the equality in (18) is satisfied. In other words, that

$$x^2 - n = \frac{1}{2}((x-1)^2 - (n+1)) + \frac{1}{2}\left((x+1)^2 - (n+1)\right).$$

And this is given by a trivial computation.

Example 5.8 (product of independent r.v.s)

Given are $X_1, X_2, \dots \geq 0$ i.i.d. and $\mathbb{E}[X_i] = 1$. Then $M_n = M_0 \cdot X_1 \cdots X_n$ is a martingale for $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$.

Namely:

$$\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n] = M_n \cdot \mathbb{E}[X_{n+1} - 1 | \mathcal{F}_n] = 0.$$

This latter is because X_{n+1} is independent of $X_1, \dots, X_n$, hence X_{n+1} is also independent of the σ-algebra $\mathcal{F}_n$ generated by them. So, $\mathbb{E}[X_{n+1} - 1 | \mathcal{F}_n] = \mathbb{E}[X_{n+1} - 1] = 0$.

Theorem 5.9

Let $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ a convex function and ψ be increasing.

(a) *If M_n is a martingale, then $\varphi(M_n)$ is a submartingale.*

(b) *If M_n submartingale, then $\psi(M_n)$ is a submartingale also.*

This is an immediate corollary of Jensen's inequality (formula (13)) and the definition.
So, if M_n is a martingale, then e.g. $|M_n|$ and M_n^2 are submartingale.

Theorem 5.10

Let M_n be a martingale. Then

$$\mathbb{E}\left[M_{n+1}^2 | \mathcal{F}_n\right] - M_n^2 = \mathbb{E}\left[(M_{n+1} - M_n)^2 | \mathcal{F}_n\right]. \tag{20}$$

Proof.

$$\mathbb{E}\left[(M_{n+1} - M_n)^2 | \mathcal{F}_n\right] = \mathbb{E}\left[M_{n+1}^2 | \mathcal{F}_n\right] - 2M_n \underbrace{\mathbb{E}\left[M_{n+1} | \mathcal{F}_n\right]}_{M_n} + M_n^2 = \mathbb{E}\left[M_{n+1}^2 | \mathcal{F}_n\right] - M_n^2. \tag{21}$$

□

Now we prove the orthogonality of the increments of the martingale.

Theorem 5.11

Let M_n be a martingale and let $0 \le i \le j \le k < n$. Then

$$\mathbb{E}\left[(M_n - M_k) \cdot M_j\right] = 0. \tag{22}$$

and its obvious corollary:

$$\mathbb{E}\left[(M_n - M_k) \cdot (M_j - M_i)\right] = 0. \tag{23}$$

Proof.

Proof of (22):

$$\begin{aligned} \mathbb{E}\left[(M_n - M_k)M_j\right] &= \mathbb{E}\left[\mathbb{E}\left[(M_n - M_k)M_j | \mathcal{F}_k\right]\right] \\ &= \mathbb{E}\Big[M_j \cdot \underbrace{\mathbb{E}\left[(M_n - M_k) | \mathcal{F}_k\right]}_{0}\Big] = 0 \end{aligned}$$

Corollary 5.12

Using notation of Theorem 5.11:

$$\mathbb{E}\left[(M_n - M_0)^2\right] = \sum_{k=1}^{n} \mathbb{E}\left[(M_k - M_{k-1})\right]^2 .$$

Proof.

By using formula (23):

$$\mathbb{E}\left[(M_n - M_0)^2\right] = \mathbb{E}\left[\left(\sum_{k=1}^{n} M_k - M_{k-1}\right)^2\right]$$
$$= \sum_{k=1}^{n} (M_k - M_{k-1})^2$$
$$+ 2 \sum_{1 \le j < k \le n} \underbrace{\mathbb{E}\left[(M_k - M_{k-1})(M_j - M_{j-1})\right]}_{0}.$$

□

Let $m \leq n$, then from the definition:

Lemma 5.13

- *If M_n is martingale, then* $\mathbb{E}[M_m] = \mathbb{E}[M_n]$,
- *If M_n is submartingale, then* $\mathbb{E}[M_m] \leq \mathbb{E}[M_n]$,
- *If M_n is supermartingale, then* $\mathbb{E}[M_m] \geq \mathbb{E}[M_n]$.

The next example is about a famous betting strategy. Then we will see that

(24) "you can't beat an unfavorable game."

Doubling strategy

In every round of a fair game Charlie bets by the so-called doubling strategy: If he wins in a game, then he bets $1 in the next one. But if he loses, in the next one he doubles his previous bet. The following table shows what happens if Charlie wins first after four lost game:

bet	1	2	4	8	16
outcome of the game	L	L	L	L	W
profit	-1	-3	-7	-15	1

If he wins in the $(k+1)^{st}$ game after k losses, then his loss is: $1+2+\cdots+2^{k-1}=2^k-1$. His winning in the $(k+1)^{st}$ game: 2^k, so his profit is: 1\$.

Generalization

- X_i is the outcome of the i^{th} game (e.g. ± 1).
- M_n is a supermartingale with respect to $X_0, X_1, \ldots,$ that is with respect to $\mathcal{F}_n := \sigma(X_0, , \ldots, X_n)$. That is $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$, $M_n \in \mathcal{F}_n$, $\mathbb{E}[|M_n|] < \infty$.
- H_n is a betting strategy, which depends on the outcome of the first $n-1$ games, so $H_n \in \mathcal{F}_{n-1} = \sigma(M_0, X_1, \ldots, X_{n-1})$. We say that H_n is *predictable*. $H_n \geq 0$. (Distinguish the bettor from the house.)
- W_n is the net profit using betting strategy H_n. That is $W_n = W_0 + \sum_{m=1}^{n} H_m \cdot (M_m - M_{m-1})$.

Examples

- Let $X_i = 1$ with probability $1/2$ and $X_i = -1$ with probability $1/2$ and $M_n = X_1 + \cdots + X_n$ and the strategy can be $H_n = 1$ for all n.
- Doubling strategy: X_n, M_n as above but H_m is 2^{k-1} if the last win happened k steps before.
- H_m is the amount of stocks we have between time $m-1$ and m and M_m the price of stocks at time m.

Theorem 5.14

Let us assume, that

- *M_n is a supermartingale for $\mathcal{F}_n$.*
- $\exists c_n > 0 : 0 \leq H_n \leq c_n$,

Then W_n is a supermartingale also.

We need $H_n \geq 0$ to ensure that the player does not become the house.

$H_n \leq c_n$ is needed for the expectation to exist. For the applications it is a handy condition.

Proof.

The change of the winning from moment n to $n+1$:

$$W_{n+1} - W_n = H_{n+1}\left(M_{n+1} - M_n\right).$$

Because $H_{n+1} \in \mathcal{F}_n$:

$$\begin{aligned}\mathbb{E}\left[W_{n+1} - W_n | \mathcal{F}_n\right] &= \mathbb{E}\left[H_{n+1}\left(M_{n+1} - M_n\right) | \mathcal{F}_n\right] \\ &= H_{n+1}\mathbb{E}\left[M_{n+1} - M_n | \mathcal{F}_n\right] \leq 0.\end{aligned}$$

So W_n is supermartingale for $\mathcal{F}_n$. □

Theorem 5.15

Using the above notaion: let us assume, that $0 < c_n$ *exists s.t.* $|H_n| < c_n$. *Then*

(a) *If* M_n *is a martingale, then* W_n *is also a martingale (for* $\mathcal{F}_n$*).*

(b) *If* M_n *is a supermartingale, then* W_n *is also a supermartingale (for* $\mathcal{F}_n$*).*

Similar to the proof of Theorem 5.14.

Stopping time or optional random variable

We have defined stopping times for Markov Chains in File A. Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ be the information we know in moment n.

Definition 5.16

A r.v. N, which takes values from the set $\{1, 2, \dots\} \cup \{\infty\}$, is a stopping time, if $\{N = n\} \in \mathcal{F}_n$, $\forall n < \infty$.

Stopping time or optional random variable (cont.)

Example 5.17 ("hitting time")

$X_1, X_2, \ldots$ i.i.d., $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$,
$S_n := X_1 + \cdots + X_n$. Hitting time of set A is
$N := \min \{n : S_n \in A\}$.

Lemma 5.18

Sum, max, min of stopping times are also stopping time.

This easily comes from the definition.

Stopping time or optional random variable (cont.)

Now we define σ-algebra $\mathcal{F}_T$ at stopping time T, which mainly represent the information we know at time T.

Definition 5.19 (σ-algebra at stopping time)

$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T = n\} \in \mathcal{F}_n\} .$$

Stopping time or optional random variable (cont.)

Lemma 5.20

Let N, T be stopping times. Then

- $\{T \leq n\} \in \mathcal{F}_T$, *in other words* $T \in \mathcal{F}_T$.
- $X_1, X_2, \ldots$ *i.i.d.,* $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, $S_n := X_1 + \cdots + X_n$, $M_n := \max\{S_m : m \leq n\}$. *Then* $S_N, M_N \in \mathcal{F}_N$.
- *In general: if* $Y_n \in \mathcal{F}_n$, *then* $Y_T \in \mathcal{F}_T$.
- *If* $N \leq T$, *then* $\mathcal{F}_N \subset \mathcal{F}_T$.

Stopping time or optional random variable (cont.)

Proving the above statements is homework.

Theorem 5.21

Let $X_1, X_2, \ldots$ i.i.d., $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$, N a stopping time (independent of $\{X_i\}$). Conditionally for $\{T < \infty\}$: $\{X_{N+n}, n \geq 1\}$ are independent of $\mathcal{F}_N$ and have the same distribution as X_n.

bet= \$1 till a stopping time

Given a stopping time T and in every game the bet is only \$1. We stop the game at time T. Let

$$H_m := \begin{cases} 1, & \text{if } m \leq T; \\ 0, & \text{if } m > T. \end{cases}$$

We claim that $H_m \in \mathcal{F}_{m-1}$, so H_m is predictable by definition on slide 66. Namely,

$$\{H_m = 0\} = \bigcup_{k=1}^{m-1} \{T = k\} \in \mathcal{F}_{m-1}.$$

So, we can use Theorem 5.14: Hence we cannot win much with this strategy either.

Theorem 5.22

Let us assume, that M_n is martingale, supermartingale or submartingale for σ-algebra $\mathcal{F}_n$ and let T be a stopping time. Then the stopped process $M_{n\wedge T}$ is also martingale, supermartingale or submartingale for M_n, where

$$T \wedge n := \min\{T, n\}.$$

Furthermore,

(a) M_n is martingale $\Longrightarrow \mathbb{E}[M_{T\wedge n}] = \mathbb{E}[M_0]$,

(b) M_n is supermartingale $\Longrightarrow$ $\mathbb{E}[M_{T\wedge n}] \leq \mathbb{E}[M_0]$,

(b) M_n submartingale $\Longrightarrow \mathbb{E}[M_{T\wedge n}] \geq \mathbb{E}[M_0]$.

Proof

Let $W_0 := M_0$. Then by definition of W_n

$$W_n = M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) = M_{T\wedge n}.$$

Namely,

- if $T \geq n$, then $W_n = M_n$ and
- if $T \leq n$, then $W_n = M_T$.

Using this, Theorems 5.14 and 5.15 we get the statement. Parts (a), (b), (c) come from Lemma 5.13.

Exit distributions

Now we are going to see an application of Theorem 5.22 and examine in the general case, that when can we substitute $M_{T\wedge n}$ in part (a) of Theorem 5.22 into M_T.

Given: $a, b \in \mathbb{Z}$, $a < b$, $X_1, X_2, \dots$ i.i.d. and

$$\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = \tfrac{1}{2}.$$

Let $S_n := S_0 + X_1 + \cdots + X_n$ and

$$\tau := \min\{n : S_n \in (a, b)\}.$$

Exit distributions (cont.)

Obviously: S_n is martingale and τ is stopping time. If we want to compute $\mathbb{E}_x[\tau]$, then we can use the following heuristic:

(25) $x \stackrel{?}{=} \mathbb{E}_x[S_\tau] = a \cdot \mathbb{P}_x(S_\tau = a) + b \cdot (1 - \mathbb{P}_x(S_\tau = a)).$

If this is true, then:

(26) $$\mathbb{P}_x(S_\tau = a) = \frac{b-x}{b-a}.$$

Exit distributions (cont.)

The argument above is just a heuristic because Theorem 5.22 only guarantees $x = S_{\tau \wedge n}$ instead of the first equality in formula (25). When can we omit $\wedge n$? First let us see an example, when we cannot:

Let $V_a := \min\{n : S_n = a\}$. Recall that we have proven in file A, that $\forall N > 0$:

$$\mathbb{P}_1 (V_N < V_0) = \frac{1}{N}. \tag{27}$$

Exit distributions (cont.)

So $\mathbb{P}_1(V_0 < \infty) = 1$. For some $n \in \mathbb{N}$:

$$T := V_0 \text{ and } \widetilde{T}_n := \min\{V_0, V_n\}.$$

Then T and $\widetilde{T}_n$ are obviously stopping times. It can be seen from formula (27), that

$$\mathbb{E}_1\left[S_{\widetilde{T}_n}\right] = 0 \cdot \mathbb{P}_1(V_0 < V_n) + n \cdot \underbrace{\mathbb{P}_1(V_n < V_0)}_{1/n} = 1.$$

Exit distributions (cont.)

So here we could leave $\wedge n$. But

$$1 \neq 0 = \mathbb{E}_1 \left[S_T \right].$$

So we could not cancell $\wedge n$ of T. The next theorem shows us when we can leave $\wedge n$.

Exit distributions (cont.)

Theorem 5.23

Let us assume, that M_n is a martingale and T is a stopping time, for which

- $\mathbb{P}(T < \infty) = 1$ *and*
- $\exists K : |M_{T \wedge n}| \leq K$.

Then $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

Exit distributions (cont.)

Proof

From Theorem 5.22:

$$\mathbb{E}[M_0] = \mathbb{E}[M_{T\wedge n}] = \mathbb{E}[M_T; T \leq n] + \mathbb{E}\Big[\underbrace{M_n}_{\leq |M_{T\wedge n}| \leq K}; T > n\Big].$$

So

$$|\mathbb{E}[M_0] - \mathbb{E}[M_T; T \leq n]| \leq K\mathbb{P}(T > n) \to 0. \tag{28}$$

Theorem 5.24 (Doob's Optional Stopping Theorem)

Let X be a supermartingale and T be a stopping time. If any of the following conditions holds

(i) T is bounded.

(ii) X is bounded and $T < \infty$ a.s..

(iii) $\mathbb{E}[T] < \infty$ and X has bounded increments.

then

(a) *$X_T \in L^1$ and $\mathbb{E}(X_T) \le \mathbb{E}[X_0]$.*

(b) *If X is a martingale then $\mathbb{E}(X_T) = \mathbb{E}[X_0]$.*

Proof (cont.)

On the other hand,

$$\mathbb{E}\left[M_T\right] - \mathbb{E}\left[M_T; T \leq n\right] = \mathbb{E}\left[M_T; T > n\right]. \tag{29}$$

Using that

$$\begin{aligned} \left|\mathbb{E}\left[M_T; T > n\right]\right| &\leq \sum_{k=n+1}^{\infty} \left|\mathbb{E}\left[M_k; T = k\right]\right| \\ &= \sum_{k=n+1}^{\infty} \left|\mathbb{E}\left[M_{k\wedge T}; T = k\right]\right| \\ &\leq K \cdot \mathbb{P}\left(T > n\right) \to 0. \end{aligned}$$

By combining formulas (28) and (29) completes the proof.

Wald equality

Let $X_1, X_2, \ldots$ be i.i.d., $\mathbb{E}[X_i] = \mu$. Let $S_n := S_0 + X_1 + \cdot + X_n$. We know, that then $M_n - n\mu$ is a martingale for X_n.

Theorem 5.25 (Wald's equation)

If T is a stopping time with $\mathbb{E}[T] < \infty$ then

$$\mathbb{E}[S_T - S_0] = \mu \mathbb{E}[T].$$

Proof can be found in (see [3]).

Convergence

Theorem 5.26 (Convergence theorem)

If $X_n \geq 0$ is a supermartingale, then $X_\infty := \lim_{n\to\infty} X_n$ exists and $\mathbb{E}[X_\infty] \leq \mathbb{E}[X_0]$.

Before the proof of the theorem, we need the following lemma, which is called Doob's martingale inequality.

Lemma 5.27

Let $X_n \geq 0$ be a supermartingale and $\lambda > 0$. In this case:

$$\mathbb{P}\left(\max_{n\geq 0} X_n > \lambda\right) \leq \mathbb{E}[X_0]/\lambda. \tag{30}$$

Proof of the lemma

Let $T := \min\{n : X_n > \lambda\}$. Observe that

$$(31) \qquad \{T < \infty\} = \left\{\max_{n\geq 0} X_n > \lambda\right\}$$

Let $A_n := \{\omega \in \Omega : T(\omega) < n\}$. Then

$$(32) \qquad X_{T(\omega)\wedge n}(\omega) = X_{T(\omega)}(\omega) > \lambda \text{ if } \omega \in A_n$$

It comes from Theorem 5.22, that

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_{T\wedge n}] \geq \mathbb{E}[X_T; A] \geq \lambda \mathbb{P}(A_n). \text{ So },$$

$$\forall n : \mathbb{P}(T < n) = \mathbb{P}(A_n) \leq \mathbb{E}[X_0]/\lambda.$$

Hence $\mathbb{P}(T < \infty) \leq \mathbb{E}[X_0]/\lambda$. And this completes the proof of the lemma by (31). ■

Draft of the proof of Theorem 5.26

Let $S_0 := 0$, $a < b$ and let us define the following stopping times:

$$R_k := \min\{n \geq S_{k-1} : X_n \leq a\}$$
$$S_k := \min\{n \geq R_k : X_n \geq b\}.$$

By a similar reasoning as in the proof of the previous lemma can we get that:

$$\mathbb{P}(S_k < \infty | R_k < \infty) \leq \frac{a}{b}.$$

Draft of the proof of Theorem 5.26 (cont.)

Iterating this

$$\mathbb{P}\left(S_k < \infty\right) \leq \left(\frac{a}{b}\right)^k \to 0 \text{ exponentially fast.}$$

So X_n only cuts interval $[a, b]$ from under finitely many times. Let

$$Y := \liminf_{n\to\infty} X_n \text{ and } Z := \limsup_{n\to\infty} X_n$$

Draft of the proof of Theorem 5.26 (cont.)

If $\mathbb{P}(Y < Z) > 0$ was true, then for some $a < b$ it would also be:

$$\mathbb{P}(Y < a < b < Z) > 0.$$

In this case X_n would cross the interval $[a, b]$ from below a to above b infinitely many times, which is not possible, so limit $X_\infty = \lim_{n\to\infty} X_n$ exists. Moreover, for all n, M:

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_n] \geq \mathbb{E}[X_n \wedge M] \to \mathbb{E}[X_\infty \wedge M].$$

Draft of the proof of Theorem 5.26 (cont.)

So

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_\infty \wedge M] \uparrow \mathbb{E}[X_\infty].$$

Polya's Urn,

Given an urn with initially contains: $r > 0$ red and $g > 0$ green balls.

(a) draw a ball from the urn randomly,

(b) observe its color,

(c) return the ball to the urn along with c new balls of the same color.

- If $c = 0$ this is sampling with replacement.
- If $c = -1$ sampling without replacement.

From now we assume that $c \geq 1$. After the n-th draw and replacement step is completed:

Polya's Urn, (cont.)

- the number of green balls in the urn is: G_n.
- the number of red balls in the urn is: R_n.
- the fraction of green balls in the urn is X_n.
- Let $Y_n = 1$ if the n-th ball drawn is green. Otherwise $Y_n := 0$.
- Let $\mathcal{F}_n$ be the filtration generated by $Y = (Y_n)$.

Polya's Urn, (cont.)

Claim 1

X_n is a martingale w.r.t. $\mathcal{F}_n$.

Proof Assume that

$$R_n = i \text{ and } G_n = j$$

Then

$$\mathbb{P}\left(X_{n+1} = \frac{j+c}{i+j+c}\right) = \frac{j}{i+j},$$

and

$$\mathbb{P}\left(X_{n+1} = \frac{j}{i+j+c}\right) = \frac{i}{i+j}.$$

Polya's Urn, (cont.)

Hence

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = \frac{j+c}{i+j+c}\cdot\frac{j}{i+j} + \frac{j}{i+j+c}\cdot\frac{i}{i+j} = \frac{j}{i+j} = X_n. \tag{33}$$

□

Corollary 5.28

There exists an X_∞ s.t. $X_n \to X_\infty$ a.s..

This is immediate from Theorem 5.26.

Polya's Urn, (cont.)

In order to find the distribution of X_∞ observe that:

- The probability $p_{n,m}$ of getting green on the first m steps and getting red in the next $n-m$ steps is the same as the probability of drawing altogether m green and $n-m$ red balls in any particular redescribed order.
- $$p_{n,m} = \prod_{k=0}^{m-1} \frac{g+kc}{g+r+kc} \cdot \prod_{\ell=0}^{n-m-1} \frac{r+\ell c}{g+r+(m+\ell)c}$$

Polya's Urn, (cont.)

If $c = g = r = 1$ then

$$\mathbb{P}(G_n = m+1) = \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1}.$$

That is X_∞ is uniform on $(0,1)$: In the general case X_∞ has density

$$\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)} x^{(g/c)-1}(1-x)^{(r/c)-1}.$$

That is the distribution of X_∞ is $\mathrm{Beta}\left(\frac{g}{c}, \frac{r}{c}\right)$

Review

Recall that $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$.

Density function of Gamma distribution with parameter , (α, λ):

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

For $\alpha, \beta > 0$ parameters the β-distribution $\mathrm{Beta}(\alpha, \beta)$ is

$$(34) \quad f_{\alpha,\beta}(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } x \in [0,1]; \\ 0, & \text{otherwise,} \end{cases}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Application

Let $U_1, \ldots, U_n$ be i.i.d. $U_i \sim \mathrm{Uni}(0,1)$. Let $U_{(k)}$ be the k-th smallest of them. Then

$$U_{(k)} \sim \mathrm{Beta}(k, n+1-k).$$

[1] Balázs Márton, Tóth Bálint
Valószínűségszámítás 1. jegyzet matematikusoknak and fizikusoknak
Bálázs Márton Honlapja, 2012. Az internetes változatért kattintson ide.

[2] R. Durrett
Essentials of Stochastic Processes, Second edition
Springer, 2012. A majdnem kész változatért kattintson ide.

[3] R. Durrett
Probability Theory with examples, Second edition
Duxbury Press, 1996 . második kiadás.

[4] I.I. Gihman, A.V. Szkorohod
Bevezetés a sztochasztikus folyamatok elméletébe
Műszaki Könyvkiadó1975, Budapest, 1985

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

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

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

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

[9] Major Péter
Folytonos idejű Markov láncok
http://www.renyi.hu/~major/debrecen/debrecen2008a/markov3.html

[10] PIET VAN MIEGHEM
The Poisson process
http://www.nas.its.tudelft.nl/people/Piet/CUPbookChapters/PACUP_Poisson.pdf

[11] RÉNYI ALFRÉD
Valószínűségszámítás, (negyedik kiadás)
Tankönyvkiadó Budapest, 1981.

[12] TÓTH BÁLINT *Sztochasztikus folyamatok jegyzet*
Tóth Bálint Jegyzetért kattintson ide

[13] G. ZITKOVIC *Theory of Probabilty I, Lecture 7*
Click here

Example

This is an example for conditional expectation.

Example

We define the probability pace $(\Omega, \mathcal{A}, \mathbb{P})$ as follows:

- $\Omega := [0,1]^2$
- $\mathcal{A}$ is the σ-algebra of Borel sets on $[0,1]^2$
- $\mathbb{P} := \mathcal{L}_2|_{[0,1]^2}$. The two-dimensional Lebesgue measure (area on the plane) restricted to the unit square.

So, an element ω of the sample space Ω is of the form $\omega = (x, y) \in [0,1]^2$.

Example (cont.)

- Let S be the random variable defined by $S(x,y) := x + y$. This is a random variable (r.v.) since this is a measurable function from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}$.
- Let $\mathcal{F} \subset \mathcal{A}$ be the σ-algebra defined by $\mathcal{B} \times [0,1]$, where $\mathcal{B}$ the Borel σ-algebra on the unit interval $[0,1]$.

Let $Z := \mathbb{E}[S|\mathcal{F}]$. Then

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

(b) $\int_A S d\mathbb{P} = \int_A Z d\mathbb{P}$ for all $A \in \mathcal{F}$.

Example (cont.)

The meaning of condition (a) is as follows: Clearly the function $Z:[0,1]^2 \to \mathbb{R}$ is measurable with respect to $\mathcal{F}$ (that is $Z \in \mathcal{F}$) if both of the following two conditions hold:

(i) $Z(x,y_1) = Z(x,y_2)$ holds for all $y_1, y_2 \in [0,1]$, ($Z(x,y)$ is constant on vertical lines)

(ii) $x \mapsto Z(x,0)$ is Borel measurable.

The meaning of condition (b) is:

$$\int_A Z(x,y)dxdx = \int_A S(x,y)dxdy, \quad \forall A \in \mathcal{F}. \tag{35}$$

Example (cont.)

If $A \in \mathcal{F}$ then A is of the form: $A = B \times [0,1]$, where $B \subset [0,1]$ Borel set. It is enough to check that (35) holds only for the sets of the form $[a,b] \times [0,1]$. For these sets (35) reads like

(36)
$$\int_a^b \int_0^1 Z(x,y)dydx = \int_a^b \int_0^1 S(x,y)dydx = \int_a^b \int_0^1 (x+y)dydx.$$

Example (cont.)

Using (i) from the one but last slide:

$$\int_a^b \int_0^1 Z(x,y)\,dy\,dx = \int_a^b Z(x,0)\,dx.$$

On the other hand, using that $\int_0^1 (x+y)\,dy = \frac{x}{2} + \frac{1}{6}$ we obtain that

$$\int_a^b \int_0^1 (x+y)\,dy\,dx = \int_a^b \left(\frac{x}{2} + \frac{1}{6}\right) dx. \tag{37}$$

That is by (36) the two yellow formulas are equal for all $0 \leq a < b \leq 1$. We use this for $b = a + \Delta x$:

Example (cont.)

$$\int_{a}^{a+\Delta x} Z(x,0)dx = \int_{a}^{a+\Delta x} \left(\frac{x}{2} + \frac{1}{6}\right) dx \tag{38}$$

We divide by Δx on both sides and we let $\Delta x \to 0$ we get from Newton-Leibnitz formula that

$$Z(x,y) = Z(x,0) = \frac{x}{2} + \frac{1}{6} \qquad \forall (x,y) \in [0,1]^2. \tag{39}$$