MAT 5171 Probability Theory II - Lecture Notes

April 11, 2020

1

$$(V_n, W_n) \xrightarrow{\text{d}} (V, W)$$

iff

$$aV_n + bW_n \xrightarrow{\text{d}} aV + bW$$

for any choice of a, b For example, $V_n = \sqrt{n}(\bar{X} - \mu)$ and V is normal: $P(V_n \leq x) \to \Phi(x)$ Here: $P(V_n \leq x, W_n \leq y) \to F(x, y)$

2 Laws of Large Numbers. Maximal Inequalities. Convergence of Random Series (Chapter 22)

Make sure that you are familiar with the following topics:

- Markov and Chebyshev inequality;
- Basic properties of the expectation;
- Borel-Cantelli lemma;

What we covered?

- In class I discussed material related to Section 22 in the textbook (Patrick Billingsley, *Probability and Measure* - I am using the anniversary edition).
- More specifically, I discussed in class:
  - A simple version of the weak law of large numbers: if $\{X_i, i \geq 1\}$ are i.i.d. random variables with finite variance and $S_n = \sum_{i=1}^n X_i$, then S_n/n converges in probability to $\mathrm{E}[X_1]$. The proof is simple: we need to show that

    $$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mathrm{E}[X_1]\right| > \epsilon\right) = 0$$

    for each $\epsilon > 0$. In order to do this we apply Chebyshev inequality.

- A simple version of the strong law of large numbers: if $\{X_i, i \geq 1\}$ are i.i.d. random variables with finite fourth moment and $S_n = \sum_{i=1}^n X_i$, then S_n/n converges in almost surely to $\mathrm{E}[X_1]$. The proof:
  * Assume for simplicity that $\mathrm{E}[X_1] = 0$;
  * Calculate $\mathrm{E}[S_n^4]$;
  * Use Markov inequality to get $P(|S_n| > n\epsilon) \leq C/n^2$, where C is a constant;
  * Since $\sum_{n=1}^\infty P(|S_n| > n\epsilon) < \infty$, use the Borel-Cantelli lemma to get that

    $$P\left(\left|\frac{S_n}{n}\right| > \epsilon \text{ infinitely often}\right) = 0 \ .$$

    This means that S_n/n converges almost surely to 0.
  * Note that this method will not work by assuming finite variance only. Indeed, then we can only obtain $P(|S_n| > n\epsilon) \leq C/n$ and the Borel-Cantelli lemma is not applicable.
- Proof of Theorem 22.1 - strong law of large numbers, assuming only that the mean is finite. Method of proof:
  * Introduce truncated variables $Y_k = 1\{X_k \leq k\}$. These random variables are independent, but have different distribution. In particular, $\lim_{k\to\infty} \mathrm{E}[Y_k] = \mathrm{E}[X_1]$;
  * Consider the truncated sum $S_n^* = \sum_{i=1}^n Y_i$. Calculate its variance (it is finite since Y_i's are bounded!!!), apply Chebyshev inequality and the Borel-Cantelli lemma to obtain almost sure convergence of the truncated sum;
  * Next,

    $$\sum_{n=1}^\infty P(X_n \neq Y_n) = \sum_{n=1}^\infty P(X_1 > n) \leq \mathrm{E}[X_1] < \infty \ .$$

    Use the Borel-Cantelli lemma to conclude that $(S_n^* - S_n)/n$ converge to zero almost surely.
- Proof of Theorems 22.4. Important tool: Define the sets $A_k = \{|S_k| > \varepsilon, |S_j| < \varepsilon, j = 1, \ldots, k-1\}$. The sets are disjoint and

  $$\{\max_{1\leq k\leq n} |S_k| > \varepsilon\} = \bigcup_{k=1}^n A_k \ .$$

  Also, split $S_n = S_k + (S_n - S_k)$, $k < n$, to use independence.

Additional material:

- Rick Durrett, *Probability. Theory and Examples. Fourth Edition.* (Available in the library). Theorems 2.2.1, 2.2.3, 2.2.6, 2.2.7, 2.3.5, 2.5.2, 2.5.3; Lemmas 2.2.2, 2.4.3. All those theorems and lemmas are either repetitions of results I proved in class or extensions of laws of large numbers and maximal inequalities.

3 Convergence in Distribution (Chapter 25)

What we covered?

- Definitions, Examples 25.1, 25.2;
- Convergence in Distributions, Example 25.5 (convergence of maxima for exponential random variables), also convergence of maxima for Pareto random variables;
- Convergence in Probability, eelation between different types of convergence: Theorem 25.2. February 10
- Properties of Convergence in Distribution: Theorem 25.4. February 10
- Skorokhod's theorem: Theorem 25.6 - how we cam make weak convergence and almost sure convergence equivalent? February 10 - presentation; see relevant preliminary result in Lemma 1 below.
- Mapping theorems: Theorems 25.7, 25.8. February 10
- Integration to the limit. February 10

Additional material

- Rick Durrett, *Probability. Theory and Examples. Fourth Edition.* (Available in the library). Section 3.2.2 - Theorem 3.2.2, 3.2.3, 3.2.4, 3.2.5. All those theorems and lemmas are either repetitions of results I proved in class or extensions

Lemma 1 *Let X and Y be random variables with continuous and strictly increasing distributions functions F and G. We say that X is stochastically smaller than Y if $F(x) \geq G(x)$ for all x (the inequality is correct, there is no mistake). Then there exists a probability space and random variables $\tilde{X}$, $\tilde{Y}$, such that $\tilde{X}$ has the same distribution as X, $\tilde{Y}$ has the same distribution as Y and $\tilde{X} \leq \tilde{Y}$ almost surely.*

Proof: $\Omega = [0,1]$; $\mathcal{F}$ - Borel sigma field; $P = \lambda$, the Lebesgue measure. Let $U : \Omega \to [0,1]$ be defined as $U(\omega) = \omega$. Then for $x \in [0,1]$,

$$P(U \leq x) = P(\{\omega : U(\omega) \leq x\}) = \lambda(\{\omega : \omega \leq x\}) = x \ .$$

Define $\tilde{X}(\omega) = F^{\leftarrow}(\omega)$, $\tilde{Y}(\omega) = G^{\leftarrow}(\omega)$. Clearly, $P(\tilde{X}(\omega) \leq x) = F(x)$.

Now, since $F(x) \geq G(x)$, we also have $\{x : G(x) > \omega\} \subseteq \{x : F(x) > \omega\}$ and thus

$$\inf_x \{x : G(x) > \omega\} \subseteq \inf_x \{x : F(x) > \omega\} \ .$$

This means that $G^{\leftarrow}(\omega) \geq F^{\leftarrow}(\omega)$ and $\tilde{Y} \geq \tilde{X}$ almost surely.

4 Characteristic functions

Material related to Section 26 in the textbook. Outline:

- Definition;
- Moments and Derivatives, Theorem 26.1
- Independence
- Uniqueness, proof of Theorem 26.2 - presentation
- Continuity, proof of Theorem 26.3.
- Additional material:
  - Rick Durrett, *Probability. Theory and Examples. Fourth Edition.* (Available in the library). Section 3.3. Look especially at Theorem 3.3.4 - this is the inversion theorem in case when μ has possibly some mass. Theorem 3.3.6, 3.3.8

5 Central Limit Theorem

Material related to Section 27 in the textbook. Outline:

- Theorems 27.1, 72.2, 27.3
- Some inequalities to remember:

$$|\prod_{i=1}^{d} z_i - \prod_{i=1}^{d} w_i| \leq \sum_{i=1}^{d} |z_i - w_i|$$

$$|e^z - 1 - z| \leq |z|^2 e^{|z|}$$

$$|e^{itx} - (1 + itx - \frac{1}{2}t^2x^2)| \leq |tx|^2 \wedge |tx|^3$$

- Additional material:
  - Theorem 27.5, CLT for dependent variables;
  - Rick Durrett, *Probability. Theory and Examples. Fourth Edition.* (Available in the library). Section 3.4.

6 Conditional Expectation

Material related to Sections 32-34 in the textbook.

6.1 Some Measure Theory

In what follows, $(S, \mathcal{G}, \mu)$ is a measurable space and $g : S \to \mathbb{R}_+$ is a nonnegative function. We recall several properties and definitions.

- A measure μ is finite of $\mu(S) < \infty$. A measure μ is σ-finite if we can write $S = \bigcup_{i=1}^{\infty} A_i$ such that $\mu(A_i) < \infty$ for each $i \geq 1$. For example, the Lebesgue measure on $[0, 1]$ is finite. The Lebesgue measure on $\mathbb{R}$ is σ-finite, but not finite.

- Notation: $\mu(g) = \int g \, d\mu$. For example, if $A \in \mathcal{G}$ and $g = 1_A$, then

$$\mu(g) = \int g \, d\mu = \int_A d\mu = \mu(A).$$

  $\mu(g)$ is a real number!!!

- Let μ be a measure on $(S, \mathcal{G})$. For a function $f : S \to \mathbb{R}_+$ we define a new measure $\nu = f\mu$ by

$$\nu(A) = (f\mu)(A) = \int_A f \, d\mu \,, \qquad A \in \mathcal{G} \,.$$

  Note that we can write $(f\mu)(A) = \mu(f 1_A)$. $f\mu$ is a measure!!!

- Assume additionally that f is bounded. Then

$$\nu(A) \leq \mu(A) \sup_{x \in S} f(x) \,.$$

  Hence, if $\mu(A) = 0$ then also $\nu(A) = 0$.

- Let $(S, \mathcal{G}) = ([0, 1], \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-field. Let λ be a Lebesque measure. Let F be a distribution function and we assume that $f = F'$ exists and is bounded. Set

$$\nu((a, b]) = F(b) - F(a) \,, \qquad a < b \,.$$

  Then

$$\nu((a, b]) = \int_a^b f(x) \, dx \leq |b - a| \sup_{x \in [0,1]} f(x) \,.$$

  Hence, if $A \in \mathcal{B}$ is such that $\lambda(A) = 0$ then also $\nu(A) = 0$.

  Here: $\nu = f\lambda$, where f is the density and λ is the Lebesque measure. Since F is differentiable, F is absolutely continuous. This explains the name absolute continuity.

The last two examples lead to *absolute continuity of measures.*

Definition 1 *Assume that* $(S, \mathcal{G})$ *is a measurable space. Let* μ, ν *be two measures. We say that* ν *is absolutely continuous with respect to* μ *if* $\mu(A) = 0$ *implies* $\nu(A) = 0$ *for any* $A \in \mathcal{G}$*. We write* $\nu \ll \mu$*.*

One of the most important statements in the probability theory is **Radon-Nikodym** theorem.

Theorem 1 (Radon-Nikodym) *Assume that* $(S, \mathcal{G})$ *is a measurable space. Let* μ, ν *be two* σ*-finite measures such that* $\nu \ll \mu$*.*

There exists a function $f : S \to \mathbb{R}_+$ *such that*

$$\nu(A) = \int_A f \, d\mu \,, \qquad \text{for all } A \in \mathcal{G} \,.$$

The meaning is: If the measures are absolutely continuous, then $\nu = f\mu$*.*

- Notation:
$$f = \frac{d\nu}{d\mu} \,.$$
- In the example above, $\nu = F$, $\mu = \lambda$ and f is just standard derivative.
- If $h : S \to \mathbb{R}_+$, then we have the following formula
$$\int_A h \, d\nu = \int_A hf \, d\mu \,.$$
  The above formula is just change of variables

Goal: Prove Theorem 1.

- Theorem 1 is valid for σ-finite measures, but I will prove it for finite measures only.

In order to do this, we introduce a concept of *singular measures* and prove *Lebesgue decomposition theorem.*

Definition 2 (Singular measures) *Assume that* $(S, \mathcal{G})$ *is a measurable space. Let* μ, ν *be two measures.*

The measures are mutually singular (written as $\mu \perp \nu$*) if there exists* $A \in \mathcal{G}$ *such that* $\mu(A) = 0 = \nu(A^c)$*.*

- Note: the above property does not need to hold for all sets $A \in \mathcal{G}$. One set is enough.

Theorem 2 (Lebesgue decomposition) *Assume that* $(S, \mathcal{G})$ *is a measurable space. Let* μ, ν *be two* σ*-finite measures such that* $\nu \ll \mu$*. Then* $\nu = \nu_a + \nu_s$*, where* $\nu_s \perp \mu$ *and* $\nu_a = f\mu$ *for some function* $f : S \to \mathbb{R}_+$*.*

- We will not prove this theorem, but to get some intuition, assume that S is countable so that $\mathcal{G} = 2^S$. Define

$$S_\mu = \{s \in S : \mu(\{s\}) = 0\} \ .$$

Then clearly $\mu(S_\mu) = 0$ (it would not be true if the space is uncountable. Take for example real line and the Lebesque measure. Then $\mu(\{s\}) = 0$ for all $s \in \mathbb{R}$, but $\mu(\mathbb{R}) = \infty$. The countability of the space is very important here). We can take

$$\nu_s(A) = \nu(A \cap S_\mu) \ , \qquad \nu_a(A) = \nu(A \cap S_\mu^c) \ , \qquad A \in \mathcal{G} \ .$$

Choose $A = S_\mu^c$, then $\nu_s(A) = \nu(S_\mu^c \cap S_\mu) = 0$. Hence, $\nu_s \perp \mu$. Furthermore, the function f can be chosen as

$$f(s) = \frac{\nu(\{s\})}{\mu(\{s\})}$$

for all s such that $\mu(\{s\}) > 0$. To see this you need to check that $\nu_a = f\mu$. For this start evaluating

$$\int_A f d\mu = \sum_{s \in A} \frac{\nu(\{s\})}{\mu(\{s\})} \mu(\{s\}) = \sum_{s \in A} \nu(\{s\}) = \nu(A) = \nu(A \cap S_\mu^c) = \nu_a(A)$$

The integral becomes the sum because we have countable space. Also, since $\mu(S_\mu) = 0$ and $\nu \ll \mu$, $\nu(S_\mu) = 0$, hence $\nu(A) = \nu(A \cap S_\mu^c)$

- Note that the meaning of Theorem 2 is that f is the Radon-Nikodym derivative $\frac{d\nu_a}{d\mu}$.

Proof of Theorem 1:

1. Assume for simplicity that the space S is countable.

2. From Theorem 2 we know that $\nu = \nu_a + \nu_s$ and $\nu_a = f\mu$ for some function f.

3. The proof will be finished if we are able to show that $\nu_s \equiv 0$, so that there is no singular part, so that $\nu = f\mu$.

4. From Theorem 2 we also know that there exists a set $A \in \mathcal{G}$ such that $\nu_s(A^c) = \mu(A) = 0$. Indeed, we can choose $A = S_\mu$, then $\nu_s(A^c) = \nu_s(S_\mu^c \cap S_\mu) = 0$ and from the explanation to Theorem 2, $\mu(S_\mu) = 0$.

5. We also assumed that $\nu \ll \mu$. Hence, from the previous step, for the selected set A, we have $\nu(A) = 0$. This also means that $\nu_s(A) = 0$.

6. We combine the last two steps. We have $\nu_s(A^c) = 0$ and $\nu_s(A) = 0$, which implies $\nu(A \cup A^c) = \nu_s(S) = 0$.

$\square$

6.2 Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be a random variable defined on it. We say that $X \in L^1(\Omega, \mathcal{F}, P)$ if $E[|X|] = \int |X| dP < \infty$.

Theorem 3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X \in L^1(\Omega, \mathcal{F}, P)$. Given $\mathcal{H} \subseteq \mathcal{F}$ there exists a random variable Y such that for all $H \in \mathcal{H}$ we have*

$$E[X 1_H] = E[Y 1_H] \, , E[X 1_H] = E[E[X \mid \mathcal{H}] 1_H] \quad . \tag{1}$$

The random variable Y is called the conditional expectation of X given $\mathcal{H}$ and is denoted by $Y = E[X|\mathcal{H}]$. Note that Y is $\mathcal{H}$-measurable.

- Note that if
$$E[X 1_F] = E[Y 1_F]$$
for all $F \in \mathcal{F}$, then $X = Y$ almost surely.

- If X and Z are random variables defined on $(\Omega, \mathcal{F}, P)$, then the notation $E[X \mid Z]$ stands for $E[X \mid \sigma(Z)]$, where $\sigma(Z)$ is the sigma-field generated by Z. If $Z = X$, then $E[X \mid \sigma(X)] = X$.

- This is very important to understand that the conditional expectation is a random variable. Intuitively, in the context above, the value of the conditional expectation depends on the outcome of the random variable Z. The outcome of the latter changes, then the conditional expectation changes.

- If X and Z are independent, then $E[X \mid Z] = E[X]$.

Proof of Theorem 3: Assume first that X is nonnegative. Let μ denote the probability measure obtained by restriction of P to $(\Omega, \mathcal{H})$, that is $\mu(H) = P(H)$ for all $H \in \mathcal{H}$ and $\mu(\Omega) = 1$.

Recall that XP denotes the measure on $(\Omega, \mathcal{F})$ such that $(XP)(A) = \int_A X \, dP$ for all $A \in \mathcal{F}$ (recall the notation $f\mu$ from the previous section - here $f = X$, $P = \mu$). Let ν be the restriction of XP to $(\Omega, \mathcal{H})$. Note that ν is a finite measure since $\nu(\Omega) = E[X] < \infty$.

If $H \in \mathcal{H}$ is such that $\mu(H) = P(H) = 0$ then $\nu(H) = 0$. Therefore, $\nu \ll \mu$. By Theorem 1 there exists a function $Y : \Omega \to \mathbb{R}_+$ such that $\nu = Y\mu$. This implies that for all $H \in \mathcal{H}$ we have

$$E[X 1_H] = \int_H X \, dP = (XP)(H) = \nu(H) = (Y\mu)(H) = \int_H Y \, d\mu = \int_H Y \, dP = E[Y 1_H] \, .$$

This finishes the proof. The proof for an arbitrary random variable follows by splitting X into the positive and the negative part.

Example 1 Assume that $X(\omega) = \sum_{i=1}^m x_i 1_{\omega \in A_i}$, $Z(\omega) = \sum_{j=1}^n z_j 1_{\omega \in B_j}$, where $A_1, \dots, A_m$ and $B_1, \dots, B_n$ are two disjoint partitions of Ω. From classical probability,

$$E[X|Z = z_j] = \sum_{i=1}^m x_i P(X = x_i \mid Z = z_j) \; .$$

Then $Y(\omega) = E[X|Z = z_j]$ whenever $Z(\omega) = z_j$ is our conditional expectation.

Indeed, let $\mathcal{H} = \sigma(Z)$. If $H \in \mathcal{H}$ then $H = \bigcup_{j \in I} B_j$ for $I \subseteq \{1, \dots, n\}$. Then

$$\begin{aligned} E[Y 1_H] &= \sum_{j \in I} E[Y 1_{B_j}] = \sum_{j \in I} E[E[X|Z = z_j] 1_{B_j}] \\ &= \sum_{j \in I} E[X|Z = z_j] \times E[1_{B_j}] = \sum_{j \in I} E[X|Z = z_j] \times P(B_j) = E[X 1_H] \; . \end{aligned}$$

Example 2 Assume that $\mathcal{H}$ is generated by a finite collection $H_1, \dots, H_n$. We claim that

$$Y(\omega) = E[X \mid \mathcal{H}](\omega) = \frac{1}{P(H_i)} \int_{H_i} X \, dP \; , \qquad \omega \in H_i \; .$$

Note that the right hand side is

$$\frac{E[X 1_{H_i}]}{P(H_i)}$$

if $\omega \in H_i$. The above expression is a random variable (since it depends on ω, but once ω is fixed this is just a number).

Indeed, we will verify (1). Any set in $H \in \mathcal{H}$ is a finite union of sets $H_1, \dots, H_n$. Thus, (1) will hold for any H if we will verify it for any of the sets H_j, $j = 1, \dots, n$. We have

$$E[Y 1_{H_j}] = E\left[\frac{1}{P(H_j)} E[X 1_{H_j}] 1_{H_j}\right] = E[X 1_{H_j}] E\left[\frac{1}{P(H_j)} 1_{H_j}\right] = E[X 1_{H_j}] \; .$$

Example 3 In what follows,

- V, W are integrable random variable;
- $\mathcal{H} \subseteq \mathcal{F}$,
- X_0 is independent of $\mathcal{H}$ and integrable;
- X_1 is $\mathcal{H}$-measurable and integrable;
- Z is a random variable. If $\mathcal{H} = \sigma(Z)$ then X_1 is $\mathcal{H}$-measurable if and only if $X_1 = f(Z)$ for a measurable function f. Moreover, X_0 is independent of $\mathcal{H}$ if and only if X_0 is independent of Z.

(a) For constants a, b we have

$$E[aV + bW \mid \mathcal{H}] = a\,\underbrace{E[V \mid \mathcal{H}]}_{=V_0} + b\,\underbrace{E[W \mid \mathcal{H}]}_{=W_0} .$$

Note that (1) means for example that

$$E[V1_H] = E[V_0 1_H] = E[E[V \mid \mathcal{H}]1_H] .$$

For any $H \in \mathcal{H}$:

$$\begin{aligned} E[(aV + bW)1_H] &= E[aV1_H] + E[bW1_H] = aE[V1_H] + bE[W1_H] \\ &= aE[V_0 1_H] + bE[W_0 1_H] \\ &= E[(aV_0 + bW_0)1_H] . \end{aligned}$$

This means that $aV_0 + bW_0$ is the conditional expectation of $(aV + bW)$ given $\mathcal{H}$.

(b) It holds:

$$E[\psi(X_0) \mid \mathcal{H}] = E[\psi(X_0)] =: \mu .$$

In order to prove it, you have to verify the identity (1), following the same steps as in Exercise 3 in the last Assignment. We need to show that for each $H \in \mathcal{H}$

$$E[\psi(X_0)1_H] = E[\mu 1_H] .$$

Since X_0 is independent of $\mathcal{H}$, $E[\psi(X_0)1_H] = E[\psi(X_0)]E[1_H] = \mu \times P(H)$. End of the proof.

(c) It holds:

$$E[\phi(X_1) \mid \mathcal{H}] = \phi(X_1) . \tag{2}$$

In order to prove it, you have to verify the identity (1), following the same steps as in Exercise 3 in the last Assignment. We need to show

$$E[\phi(X_1)1_H] = E[\phi(X_1)1_H] .$$

There is nothing to prove here.

Assume additionally that the random variable X_0 has mean zero. Can we take

$$E[\phi(X_1) \mid \mathcal{H}] = \phi(X_1) + X_0?$$

We evaluate

$$E[(\phi(X_1) + X_0)1_H] = E[\phi(X_1)1_H] + E[X_0 1_H] = E[\phi(X_1)1_H] + E[X_0]P(H) = E[\phi(X_1)1_H] .$$

In the first equation we used part (a), the next one is part (b). Thus, $\phi(X_1) + X_0$ fulfills (1). But, $\phi(X_1) + X_0$ is not $\mathcal{H}$-measurable!

(d) We have

$$E[X_1 V \mid \mathcal{H}] = X_1 E[V \mid \mathcal{H}] \ .$$

Note that our candidate for the conditional expectation (the random variable on the right hand side) is $\mathcal{H}$-measurable.

We start with the left hand side. We need to evaluate $E[X_1 V 1_H]$ for $H \in \mathcal{H}$. Take first $X_1 = 1_{H_0}$, $H_0 \in \mathcal{H}$. Then

$$E[X_1 V 1_H] = E[V 1_{H \cap H_0}] \ .$$

Let us denote $V_0 = E[V \mid \mathcal{H}]$. By (1),

$$E[V 1_{H \cap H_0}] = E[V_0 1_{H \cap H_0}] = E[1_{H_0} V_0 1_H] = E[X_1 V_0 1_H] \ .$$

Thus, we have

$$E[X_1 V 1_H] = E[X_1 V_0 1_H] \ .$$

But this means that

$$E[X_1 V \mid \mathcal{H}] = X_1 V_0 = X_1 E[V \mid \mathcal{H}] \ .$$

(e) Assume that (X, W) is a bivariate normal vector, such that both components are standard normal. the correlation is assumed to be ρ. What is $E[W \mid X]$? Here we will not prove equality (1), rather we will use the properties (a), (b), (c) proven above.

Recall that W can be written as $W = \rho X + \sqrt{1-\rho^2} Z$, where Z is standard normal, independent of everything else. Then

$$\begin{aligned} E[W \mid X] &= E[\rho X + \sqrt{1-\rho^2} Z \mid X] = E[\rho X \mid X] + E[\sqrt{1-\rho^2} Z \mid X] \\ &= \rho X + \sqrt{1-\rho^2} E[Z] = \rho X \ . \end{aligned}$$

If $X = W$, then $E[W \mid X] = E[W \mid W] = W$

Example 4 (a) If $Y = \mathrm{E}[X \mid \mathcal{H}]$ then

$$\mathrm{E}[Y] = \mathrm{E}[X] \tag{3}$$

Indeed, (1) can be written as

$$\int_H X dP = \int_H Y dP$$

for all $H \in \mathcal{H}$. Take $H = \Omega$ to get

$$\int_\Omega X dP = \int_\Omega Y dP$$

which can be recognized as (3). We can re-write (3) as

$$\mathrm{E}[\mathrm{E}[X \mid \mathcal{H}]] = \mathrm{E}[X] \ . \tag{4}$$

(b) We know that

$$\mathrm{E}[|X|] \geq |\mathrm{E}[X]| \ .$$

We have

$$\mathrm{E}[|X| \mid \mathcal{H}] \geq |\mathrm{E}[X \mid \mathcal{H}]| \ . \tag{5}$$

Additional material:

- Properties of conditional expectations: Theorem 34.2, 34.3, 34.4.

7 Martingales

A martingale is a model for a fair game. Suppose we have a probability spaces $(\Omega, \mathcal{F}, P)$ and a sequence of σ-algebras

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots .$$

Such an increasing sequence $\{\mathcal{F}_n, n \geq 1\}$ of σ-fields is called *filtration*. Intuitively, $\mathcal{F}_n$ represents the information up to time n (including time n).

Definition 3 *Let $M_1, M_2, \ldots,$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. The sequence $\{(M_n, \mathcal{F}_n), n \geq 1\}$ is a martingale if*

(i) *$\{\mathcal{F}_n\}$ is a filtration;*

(ii) *M_n is $\mathcal{F}_n$-measurable;*

(iii) *$\mathrm{E}[|M_n|] < \infty$;*

(iv)

$$\mathrm{E}[M_{n+1} \mid \mathcal{F}_n] = M_n \, . \tag{6}$$

Alternatively, we say that the sequence $\{M_n\}$ is a martingale w.r.t the filtration $\{\mathcal{F}_n\}$

Natural filtration. Let $\mathcal{G}_n = \sigma(M_1, \ldots, M_n)$. Then $\{\mathcal{G}_n, n \geq 1\}$ is a *natural filtration* of the sequence $\{M_n\}$. Then (6) is equivalently written as

$$M_n = \mathrm{E}[M_{n+1} \mid \mathcal{G}_n] = \mathrm{E}[M_{n+1} \mid \sigma(M_1, \ldots, M_n)] = \mathrm{E}[M_{n+1} \mid M_1, \ldots, M_n] \, .$$

Martingale difference. Since M_n is $\mathcal{F}_n$-measurable, $\mathrm{E}[M_n \mid \mathcal{F}_n] = M_n$ and hence the martingale property (6) can be written equivalently as

$$\mathrm{E}[M_{n+1} \mid \mathcal{F}_n] = \mathrm{E}[M_n \mid \mathcal{F}_n] \, ,$$
$$\mathrm{E}[M_{n+1} - M_n \mid \mathcal{F}_n] = 0 \, . \tag{7}$$

The last expression leads to the definition of the *martingale difference* sequence: $\{(X_n, \mathcal{F}_n)\}$ is a martingale difference if

$$\mathrm{E}[X_{n+1} \mid \mathcal{F}_n] = 0 \, .$$

Above, $X_{n+1} = M_{n+1} - M_n$. Hence:

- If $\{M_n\}$ is a martingale, then the sequence $\{X_n\}$ defined by $X_{n+1} = M_{n+1} - M_n$ is a martingale difference;
- If $\{X_n\}$ is a martingale difference, then the sequence $\{M_n\}$ defined by $M_n = X_1 + \cdots + X_n$ is a martingale.

Of course, each time we need to remember about the filtration. We note that

$$\sigma(X_1, \ldots, X_n) = \sigma(M_1, \ldots, M_n) \, .$$

7.1 Properties

- Why a martingale is a fair game? Let's take (6) and re-write it using the definition of the conditional expectation

$$\int_A M_{n+1} dP = \int_A M_n dP$$

for all $A \in \mathcal{F}_n$. Take $A = \Omega$. Then the above property reads

$$\mathrm{E}[M_{n+1}] = \mathrm{E}[M_n] \ .$$

- Let now $\{X_n\}$ be a martingale difference. Then $\mathrm{E}[X_n] = 0$. Furthermore, by (4)

$$\mathrm{E}[X_n X_{n+1}] = \mathrm{E}[\mathrm{E}[X_n X_{n+1} \mid \mathcal{F}_n]] = \mathrm{E}[X_n \mathrm{E}[X_{n+1} \mid \mathcal{F}_n]] = 0 \ .$$

Thus, the martingale difference has covariance zero, but $\{X_n\}$ are not independent. Note also that we do not need a finite variance for the covariance to exists.

- A function of a martingale is not necessary a martingale. Indeed, let M_n be a martingale and consider $\widetilde{M}_n = |M_n|$. Then

$$\mathrm{E}[\widetilde{M}_{n+1} \mid \mathcal{F}_n] = \mathrm{E}[|M_{n+1}| \mid \mathcal{F}_n] \geq |\mathrm{E}[M_{n+1} \mid \mathcal{F}_n]| = |M_n| = \widetilde{M}_n \ .$$

In fact, $|M_n|$ is a submartingale.

7.2 Examples

(1) Assume that $\{X_n\}$ are i.i.d with mean zero. Let $\{\mathcal{G}_n\}$ be a natural filtration, that is $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Then $\{X_n\}$ is a martingale difference and $M_n = X_1 + \cdots + X_n$ is a martingale. Indeed,

$$\begin{aligned}
&\mathrm{E}[M_{n+1} \mid M_1, \ldots, M_n] = \mathrm{E}[M_n + X_{n+1} \mid M_1, \ldots, M_n] \\
&= \mathrm{E}[M_n \mid M_1, \ldots, M_n] + \mathrm{E}[X_{n+1} \mid M_1, \ldots, M_n] \\
&= M_n + \mathrm{E}[X_{n+1} \mid X_1, \ldots, X_n] = M_n + \mathrm{E}[X_{n+1}] \\
&= M_n + 0 \ .
\end{aligned}$$

Note further that if $\mathrm{E}[X_n] \neq 0$, then $\{M_n\}$ is not a martingale.

(2) Assume that $\{X_n\}$ are i.i.d with mean zero and variance σ^2. Let $\{\mathcal{G}_n\}$ be a natural filtration. Let $S_n = X_1 + \cdots + X_n$. Then

$$M_n = S_n^2 - n\sigma^2$$

is a martingale.

(3) Let Z be an integrable random variable and let $\{\mathcal{F}_n\}$ be a filtration. Define

$$M_n = \mathrm{E}[Z \mid \mathcal{F}_n] \ .$$

Then $\{M_n\}$ is a martingale. Indeed, M_n is $\mathcal{F}_n$-measurable and by (4)

$$\mathrm{E}[|M_n|] = \mathrm{E}[|\mathrm{E}[Z \mid \mathcal{F}_n]|] \leq \mathrm{E}[\mathrm{E}[|Z| \mid \mathcal{F}_n]] = \mathrm{E}[|Z|]$$

(4) Assume that $\{X_n\}$ are i.i.d with mean zero. Let $\{\mathcal{G}_n\}$ be a natural filtration. For each n, let B_n be a bounded random variable which is measurable with respect to $\mathcal{G}_{n-1}$. We think of B_n as being the "bet" on the game X_n; we can see the results of $X_1, \ldots, X_{n-1}$ before choosing a bet but one cannot see X_n. The total fortune by time n is given by $M_0 = 0$ and

$$M_n = B_1 X_1 + \cdots + B_n X_n \ .$$

Then $\{M_n\}$ is a martingale w.r.t $\{\mathcal{G}_n\}$.

(5) Let $\{Z_n\}$ be a sequence of i.i.d. random variables with mean zero. Define

$$X_n = \sigma_n Z_n$$

and let $\mathcal{G}_n$ be the natural filtration of $\{X_n\}$. Here: $\{Z_n\}$ be a sequence of i.i.d. random variables with mean zero and variance 1 such that Z_{n+1} is independent of $\mathcal{G}_n$ and σ_n is assumed to be $\mathcal{G}_{n-1}$-measurable. Then

$$\begin{aligned}
&\mathrm{E}[X_{n+1} \mid \mathcal{G}_n] = \mathrm{E}[\sigma_{n+1} Z_{n+1} \mid \mathcal{G}_n] = \sigma_{n+1}\mathrm{E}[Z_{n+1} \mid \mathcal{G}_n] \\
&= \sigma_{n+1}\mathrm{E}[Z_{n+1}] = 0 \ .
\end{aligned}$$

Hence, $\{X_n\}$ is a martingale difference. On the other hand

$$\mathrm{E}[X_{n+1}^2 \mid \mathcal{G}_n] = \sigma_{n+1}^2 \ .$$

(6) Assume that on the probability space $(\Omega, \mathcal{F}, P)$ we have a probability measure Q. Consider a sequence of random variables $\{Y_n\}$ and let $\mathcal{G}_n$ be its natural filtration. Assume that $(Y_1, \ldots, Y_n)$ has a density p_n under measure P and density q_n under measure Q. Define

$$M_n = \frac{q_n(Y_1, \ldots, Y_n)}{p_n(Y_1, \ldots, Y_n)} \ .$$

Note that an element of $\mathcal{G}_n$ is $\{(Y_1, \ldots, Y_n) \in H\}$, where H is a "nice" set in R^n. Thus

$$\begin{aligned}
&\mathrm{E}[M_n 1\{(Y_1, \ldots, Y_n) \in H\}] = \int_{(Y_1,\ldots,Y_n)\in H} M_n dP \\
&= \int_{(Y_1,\ldots,Y_n)\in H} \frac{q_n(Y_1, \ldots, Y_n)}{p_n(Y_1, \ldots, Y_n)} dP \\
&\int_H \frac{q_n(y_1, \ldots, y_n)}{p_n(y_1, \ldots, y_n)} p_n(y_1, \ldots, y_n) dy_1 \cdots dy_n \\
&\int_H q_n(y_1, \ldots, y_n) dy_1 \cdots dy_n = Q(H) \ .
\end{aligned}$$

Furthermore, $\{M_n\}$ is a martingale.

7.3 Stopping times

Assume that $\{\mathcal{F}_n, n \geq 0\}$ is a filtration. An integer-valued random variable τ is called a *stopping time* (relative to the filtration) if for all $n \geq 1$,

$$\{\tau = n\} \in \mathcal{F}_n \; .$$

Equivalently,

$$\{\tau \leq n\} \in \mathcal{F}_n \; .$$

Indeed,

$$\{\tau \leq n\} = \bigcup_{i=0}^{n} \{\tau = i\} \in \mathcal{F}_n$$

Furthermore,

$$\{\tau \geq n+1\} \in \mathcal{F}_n \tag{8}$$

since $\{\tau \geq n+1\}$ and $\{\tau \leq n\}$ are complementary events.
Examples:

- Assume that $\{X_i\}$ is a sequence of i.i.d. random variables. Let $\{\mathcal{G}_n\}$ be its natural filtration. Let $S_n = X_1 + \cdots + X_n$ and $A \subset \mathbb{R}$. Then

  $$\tau = \min\{j : S_j \in A\}$$

  is a stopping time.

- More generally, if $\{M_n\}$ is a martingale, then

  $$\tau = \min\{j : M_j \in A\}$$

  is a stopping time.

- However, $\tau = \max\{j : M_j \in A\}$ is not a stopping time.

- If τ is a stopping time, then $\tau \wedge n$ is also a stopping time.

Theorem 4 *Assume that $\{(M_n, \mathcal{F}_n), n \geq 0\}$ is a martingale. Then*

$$\widetilde{M}_n = M_{\tau \wedge n} = \begin{cases} M_\tau & if\ \tau < n \\ M_n & if\ \tau \geq n \end{cases}$$

is also a martingale and

$$\mathrm{E}[\widetilde{M}_n] = \mathrm{E}[M_n] = \mathrm{E}[M_0] \; .$$

Re-phrasing: "A stopped martingale is again a martingale". See Theorem 35.2 for a generalization.

Proof: Note that

$$\widetilde{M}_n = M_{\tau\wedge n} = M_n 1\{\tau \geq n\} + \sum_{j=0}^{n-1} M_j 1\{\tau = j\}$$

$$\widetilde{M}_{n+1} = M_{\tau\wedge(n+1)} = M_{n+1} 1\{\tau \geq (n+1)\} + \sum_{j=0}^{n} M_j 1\{\tau = j\} \ .$$

Clearly, $\widetilde{M}_n$ is $\mathcal{F}_n$-measurable and integrable. Now, we calculate

$$\mathrm{E}\left[\widetilde{M}_{n+1} \mid \mathcal{F}_n\right] = \mathrm{E}\left[M_{n+1} 1\{\tau \geq (n+1)\} \mid \mathcal{F}_n\right] + \sum_{j=0}^{n} \mathrm{E}\left[M_j 1\{\tau = j\} \mid \mathcal{F}_n\right]$$

$$= 1\{\tau \geq (n+1)\} \mathrm{E}\left[M_{n+1} \mid \mathcal{F}_n\right] + \sum_{j=0}^{n} M_j 1\{\tau = j\}$$

$$= 1\{\tau \geq (n+1)\} M_n + \sum_{j=0}^{n} M_j 1\{\tau = j\}$$

$$= \{1\{\tau \geq n\} - 1\{\tau = n\}\} M_n + \sum_{j=0}^{n} M_j 1\{\tau = j\} = \widetilde{M}_n \ .$$

Examples:

- Assume that $\{X_n\}$ is a sequence of i.i.d. random variables such that $P(X_i = 1) = P(X_i = -1) = 1/2$. Let $M_n = a + X_1 + \cdots + X_n$. Define $\tau = \inf\{j : M_j = 0\}$. Then

  $$\mathrm{E}[M_{\tau\wedge n}] = \mathrm{E}[M_0] = a \ .$$

  We note at the same time that $\mathrm{E}[M_\tau] = 0$.

- Now, consider $\tau = \inf\{j : M_j = 0 \textit{ or } M_n = N\}$ for some integer N. Then again

  $$\mathrm{E}[M_{\tau\wedge n}] = \mathrm{E}[M_0] = a \ .$$

  At the same time

  $$\mathrm{E}[M_{\tau\wedge n}] = N P(M_\tau = N) \ ,$$

  hence

  $$P(M_\tau = N) = a/N \ .$$

7.4 Martingale convergence theorem

Theorem 5 *Assume that $\{M_n\}$ is a martingale such that $K := \sup_{n\geq 1} \mathrm{E}[|M_n|] < \infty$. Then $M_n \to M$ with probability 1 and $\mathrm{E}[|M|] \leq K$.*