Lecture 6 : Distributions

Theorem 6.0.1 (Hölder's Inequality) *If* $p, q \in [1, \infty]$ *with* $1/p + 1/q = 1$ *then*

$$\mathbb{E}(|XY|) \leq ||X||_p ||Y||_q \tag{6.1}$$

Here $||X||_r = (\mathbb{E}(|X|^r))^{1/r}$ *for* $x \in [1, \infty)$*; and* $||X||_\infty = \inf\{M : \mathbb{P}(|X| > M) = 0\}$.

Proof: See the proof of (5.2) in the Appendix of Durrett. ■

Example 6.0.2 *If* $|Y| \leq b$ *then*

$$\mathbb{E}(|XY|) \leq b\mathbb{E}(|X|)$$

Theorem 6.0.3 (Cauchy-Schwarz Inequality) *The special case* $p = q = 2$ *is the Cauchy-Schwarz inequality.*

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(X^2)\mathbb{E}(Y^2))^{1/2} \tag{6.2}$$

Proof: Apply Hölder's inequality for $p = q = 2$. ■

Example 6.0.4 *For* $X \geq 0$ *take* $Y = \mathbf{1}_X$. *Then we get*

$$\mathbb{E}(X) = \mathbb{E}(|XY|) \leq \mathbb{E}(X^2)^{1/2}\mathbb{E}(Y^2)^{1/2} = \mathbb{E}(X^2)^{1/2}\mathbb{E}(Y)^{1/2} = \mathbb{E}(X^2)^{1/2}\mathbb{P}(X > 0)^{1/2}$$

so squaring both sides we get

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}.$$

Theorem 6.0.5 (Minkowski's Inequality (Triangle inequality for L^p))

$$||X + Y||_p \leq ||X||_p + ||Y||_p$$

6.1 Probability distribution on the real line

If $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ is a $\mathcal{F} \setminus \mathcal{S}$-measuable r.v. , then X induces a probability measure defined on $\mathcal{S}$. Set $\mu(S) \overset{\text{def}}{=} \mathbb{P}(X \in S) = \mathbb{P}(X^{-1}(B))$. It is easy to check that μ defined thusly is a probability measure. For example, observe that for countably many disjoint S_i's,

$$\mu(\cup_i S_i) = \mathbb{P}(X^{-1}(\cup_i S_i)) = \mathbb{P}(\cup_i X^{-1}(S_i)) = \sum_i \mathbb{P}(X^{-1}(S_i)) = \sum_i \mu(S_i).$$

The other properties of a probability measure can be checked in a similar manner.

In the case that the r.v. X is real-valued, we say that that the induced measure is the *distribution of* X, and describe this measure by its *cumulative distribution function* (cdf), $F_X(s) = \mathbb{P}(X \leq s)$.

Theorem 6.1.1 *A cdf F of some probability measure on $\mathbb{R}$ has the following properties:*

1. *F is an increasing function of x.*
2. $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$
3. *F is right continous, i.e.,* $\lim_{y\downarrow x} F(y) = F(x)$
4. *If $F(x-) = \lim y \uparrow xF(y)$ then $F(x-) = P(X > x)$*
5. *$P(X = x) = F(x) - F(x-)$*

Proof: Refer to Theorem 1.1. in Durrett on page 4. ■

Theorem 6.1.2 *If F satisfies the first three properties of Theorem 6.1.1, then it is the distribution function of some r.v. and there is a unique probability measure on $(\mathbb{R}, \mathcal{R})$ that has $\mu((a, b]) = F(b) - F(a)$ for all a, b.*

Proof: This follows Durrett, pg. 5. Let $F : \mathbb{R} \to (0, 1)$ have properties $1, 2, 3$ in Theorem 2.5. We will construct a r.v. defined on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}((0, 1)), \lambda)$, where λ denotes Lebesgue measure, and show that it has distribution function F.

Let

$$X(\omega) = \sup\{y : F(y) < \omega\}.$$

$$\begin{tikzcd}
(\Omega, \mathcal{F}, \mathbb{P}) \arrow[rr, "Y \circ X"] \arrow[dr, "X"'] & & (T, \mathcal{T}, \mathbb{P}_{Y \circ X}) \\
& (S, \mathcal{S}, \mathbb{P}_X) \arrow[ur, "Y"'] &
\end{tikzcd}$$

Figure 6.1: An illustration of the transitivity of the image laws.

If we show that

$$\{\omega : X(\omega) \le x\} = \{\omega : \omega \le F(x)\}$$

then the desired result follows immediately since $\mathbb{P}(\omega : \omega \le F(x)) = F(x)$. (Recall that $\mathbb{P}$ is Lebesgue measure on (0,1) and that F is increasing, so that $\mathbb{P}(\omega : \omega \le F(x)) = \sup\{\omega : \omega \le F(x)\}$.) To check the set equality above, observe that if $\omega \le F(x)$ then $X(\omega) \le x$, since $x \notin \{y : F(y) < \omega\}$. On the other hand, if $\omega > F(x)$, then since F is right continuous, there is an $\epsilon > 0$ so that $F(x + \epsilon) < \omega$ and $X(\omega) \ge x + \epsilon > x$. ■

6.2 Change of Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to S$ a $(\mathcal{F} \backslash \mathcal{S})$-measurable random variable. X induces a new probability measure $\mathbb{P}_X$ on $(S, \mathcal{S})$.

Definition 6.2.1 $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ *is called the* $\mathbb{P}$ *law of* X *or the* $\mathbb{P}$ *distribution of* X.

Let $(T, \mathcal{T})$ be another measurable space and $Y : S \to T$ a measurable map. Then we have transitivity of the image laws.

Theorem 6.2.2 (Transitivity of the image laws) *The* $\mathbb{P}$ *distribution of* $Y \circ X$ *is equal to the* $\mathbb{P}_X$ *distribution of* Y.

Theorem 6.2.3 (Change of variable formula) *Let* Y *be a real-valued r.v. on* $(S, \mathcal{S})$*.* Y *is* $\mathbb{P}_X$*-integrable iff* $Y \circ X$ *is* $\mathbb{P}$*-integrable, and then*

$$\int_S Y d\mathbb{P}_X = \int_\Omega (Y \circ X) d\mathbb{P}$$

Proof: Fix X and vary Y. For indicators Y the identity is the transitivity of image laws, and this passes to simple r.v.'s Y, then all r.v.'s Y. See Durrett [1.3, pp. 17] ■

Lecture 7 : Product Spaces

7.3 Product spaces and Fubini's Theorem

Definition 7.3.1 *If* $(\Omega_i, \mathcal{F}_i)$ *are measurable spaces,* $i \in I$ *(index set), form* $\prod_i \Omega_i$. *For simplicity,* $\Omega_i = \Omega_1$.

$\prod_i \Omega_i$ *(write* Ω *for this) is the space of all maps:* $I \to \Omega_1$. *For* $\omega \in \prod_i \Omega_i$, $\omega = (\omega_i : i \in I, \omega_i \in \Omega_i)$. Ω *is equipped with* projections, $X_i : \Omega \to \Omega_i$, $X_i(\omega) = \omega_i$.

[picture of square, Ω_1 on one side, Ω_2 on other, point ω in middle, maps under projection to each side]

Definition 7.3.2 *A product* σ*-field on* Ω *is that generated by the projections:* $\mathcal{F} = \sigma\left((X_i \in F_i) \mid F_i \in \mathcal{F}_i\right)$.

$F_1 \times F_2 = (\Omega \mid \Omega_1 \in F_1 \wedge \Omega_2 \in F_2) \in \mathcal{F} = (X_1 \in F_1) \cap (X_2 \in F_2)$.

We now seek to construct the *product measure*. We start with the $n = 2$ case. $(\Omega, \mathcal{F}) = (\Omega_1, \mathcal{F}_1) \times (\Omega_2, \mathcal{F}_2)$. Suppose we have probability measures P_i on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$. Then there is a natural way to construct $P = P_1 \times P_2$ (product measure) on $(\Omega, \mathcal{F})$.

Key idea: $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$, $P(F_1 \times F_2) = P_1(F_1) \times P_2(F_2)$.

Theorem 7.3.3 (Existence of product measure and Fubini theorem) *: (notation* $\omega = (\omega_1, \omega_2)$*) There is a unique probability measure* P *on* $(\Omega_1, \mathcal{F}_1) \times (\Omega_2, \mathcal{F}_2)$ *such that for every non-negative product-measurable [measurable w.r.t. the product* σ*-field] function* $f : \Omega_1 \times \Omega_2 \to [0, \infty)$,

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1 \times \omega_2) P(d\omega) = \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) P_2(d\omega_2) \right] P_1(d\omega_1) =$$
$$\int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) P_1(d\omega_1) \right] P_2(d\omega_2).$$

Note that this really tells very explicitly what P is:

$$P(A) = \int_{\Omega_2} 1_A(\omega) P(d\omega) \qquad (*)$$

Fix ω_2, look at $A\omega_2 = \{\omega_1 \mid (\omega_1, \omega_2) \in A\}$.

$$\int P_1(A\omega_2)P_2(d\omega_2)$$

Look at the level of sets. $f = 1_A$. Look at formula $(*)$. Look at the collection $\mathcal{C}$ of all $A \subset \Omega_1 \times \Omega_2$ such that $*$ makes sense, i.e.:

1. $A\omega_2 \in \mathcal{F}_1$ for every $\omega_2 \in \Omega_2$.
2. $\omega_2 \to P_1(A\omega_1)$ is measurable.

Observe:

1. $\mathcal{C}$ contains all $A_1 \times A_2$ (π-system, closed under intersection), and get $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$
2. $\mathcal{C}$ is a λ-system.

Thus, $\mathcal{C} \supset \sigma(A_1 \times A_2)$, which is the product of σ-fields.

(Checking λ-system uses monotone convergence theorem.)

We know that the order of integration was not relevant, because the other way gives us a measure which agrees on rectangles (by commutativity of addition), and rectangles generate (and we get a π-system out of them), so we can apply the π-λ-theorem.

Extend our measure on indicator functions to simple functions, additively, and so on.

7.4 Independence

Random variables X_1 and X_2 with values in $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ are called *independent* iff $\mathbb{P}(X_1 \in F_1, X_2 \in F_2) = \mathbb{P}(X_1 \in F_1)\mathbb{P}(X_2 \in F_2)$ for all $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$.

Observe:

1. If we take $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ and $\mathbb{P} = P_1 \times P_2$, and $X_i(\omega) = \omega_i$ projections as before, then X_1 and X_2 are independent random variables with distributions P_1 and P_2.
2. If X_1 and X_2 are independent random variables, defined on any background space $(\Omega, \mathcal{F})$, then the *joint distribution* of (X_1, X_2) is the product measure $P_1 \times P_2$, $P_i = P_{X_i}$ which is the $\mathbb{P}$ distribution of X_i.

$\omega \to X_1(\omega)$, $\omega \to X_2(\omega)$. Look at $\omega \to (X_1(\omega), X_2(\omega))$ as a map from Ω to $\Omega_1 \times \Omega_2$. We check that this map is product-measurable. If $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$, then

$$\{(X_1, X_2) \in A_1 \times A_2\} = (X_1 \in A_1) \cap (X_2 \in A_2) \in \mathcal{F}.$$

Now $P_{(X_1,X_2)} = \mathbb{P}$ distribution of (X_1, X_2) makes sense.

Corollary 7.4.1 (Fubini) *If X_1 and X_2 are independent random variables,*

$$\mathbb{E}\left[f(X_1, X_2)\right] = \int \left[\int f(x_1, x_2) P_1(dx_1)\right] P_2(dx_2).$$

This justifies formulas for distribution of $X_1 + X_2$, $X_1 X_2$ for real r.v.'s. Example: If X_1 has density f_1, X_2 has density f_2, $\mathbb{P}(X_1 \in A) = \int_{A_1} f(x_1) dx_1$. Then $X_1 + X_2$ has a density

$$f(z) = (f_1 \times f_2)(z) = \int_{\mathbb{R}} f_1(x) f_2(z - x) dx.$$

Check this by application of Fubini theorem.

$\mathbb{E}[g(X_1 + X_2)] = \int g(z) f(z) dz.$

Useful fact: For real random variables, X_1 and X_2 are independent if and only if $\mathbb{P}(X_1 \le x_1, X_2 \le x_2) = \mathbb{P}(X_1 \le x_1)\mathbb{P}(X_2 \le x_2)$ for all real x_1, x_2.

Fix $F_1 = (-\infty, x_1]$ first. Consider all sets with $\mathbb{P}(X_1 \in F_1, X_2 \in F_2) = \mathbb{P}(X_1 \in F_1)\mathbb{P}(X_2 \in F_2)$.

Extend to n variables. $X_1, \ldots, X_n$ are independent iff

$$\mathbb{P}\left(\bigcap_i (X_i \in F_i)\right) = \prod_i \mathbb{P}(X_i \in F_i).$$

Same discussion with product spaces. Most proofs work by induction on n, reduces to $n = 2$. For example, X_1, X_2, X_3 are independent iff X_1 and X_2 are independent and (X_1, X_2) and X_3 are independent.

Intuitive properties: e.g. if $X_1, \ldots, X_5$ are independent, then $X_1 + X_3 + X_5$ and $X_2 + X_4$ are independent.

To check this, we need to check that we can factor the probabilities as above, which goes for simple functions, and then extends.

Proposition 7.4.2 *If X and Y are independent and $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y|) < \infty$ then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Proof: Use Fubini. Be careful: we only showed Fubini for non-negative functions.

First, check in case $X \geq 0$, $Y \geq 0$. $\mathbb{E}(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy \mathbb{P}(X \in dx)\mathbb{P}(Y \in dy) = \mathbb{E}(X)\mathbb{E}(Y)$, by Fubini.

In general, $X = X^+ - X^-$, $Y = Y^+ - Y^-$, and get four pieces, then put back together. ■

Note that the most general form of Fubini's theorem gives

$$\mathbb{E}(f(X,Y)) = \int \int f$$

provided this $\int \int f$ formula is finite when f is replaced by $|f|$.]

Recall $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ always, provided they are finite. $\mathbf{V}ar(X+Y) = \mathbf{V}ar(X) + \mathbf{V}ar(Y) + 2\mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]$ (last term is *covariance*), and last term is 0 if X and Y are independent.

Lecture 8 : Weak Law of Large Numbers

References: Durrett [Sections 1.4, 1.5]

The Weak Law of Large Numbers is a statement about sums of independent random variables. Before we state the WLLN, it is necessary to define convergence in probability. We say Y_n *converges in probability* to Y and write $Y_n \xrightarrow{\mathbb{P}} Y$ if, $\forall \epsilon > 0$,

$$P(\omega : |Y_n(\omega) - Y(\omega)| > \epsilon) \to 0, \quad n \to \infty.$$

Theorem 8.4.3 (Weak Law of Large Numbers) *Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbb{E}|X|^2 < \infty$ and define $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X.$$

Proof: In this proof, we employ the common strategy of first proving the result under an L^2 condition (i.e. assuming that the second moment is finite), and then using truncation to get rid of the extraneous moment condition.

First, we assume $\mathbb{E}(X^2) < \infty$. Because the X_i are iid,

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(X_i) = \frac{\operatorname{Var}(X)}{n}.$$

By Chebychev's inequality, $\forall\ \epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{\operatorname{Var}(X)}{n\epsilon^2} \to 0.$$

Thus, $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X$ under the finite second moment condition. To transition from L^2 to L^1, we use truncation. ■

To transition from L^2 to L^1, we use truncation.

Theorem 8.4.4 (Weak Law of Large Numbers) *Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbb{E}|X|^2 < \infty$ and define $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X.$$

Proof: For $0 < x < \infty$ let

$$\begin{aligned} X_{xk} &= X_k \mathbf{1}_{(|X_k| \le x)} \\ Y_{xk} &= X_k \mathbf{1}_{(|X_k| > x)} \end{aligned}$$

Then, we have $X_k = X_{xk} + Y_{xk}$ and

$$\begin{aligned} \frac{S_n}{n} &= \frac{1}{n}\sum_{k=1}^{n} X_{xk} + \frac{1}{n}\sum_{k=1}^{n} Y_{xk} \\ &= U_{xn} + V_{xn} \end{aligned}$$

Applying Jensen's inequality, we have

$$E\left|\frac{1}{n}\sum_{k=1}^{n} Y_{xk}\right| \le \frac{1}{n}\sum_{k=1}^{n} E|Y_{xk}| = E(|X|\mathbf{1}_{(|X|>x)})$$

and by DCT,

$$\mathbb{E}(|X|\mathbf{1}_{(|X|>x)}) \to 0, \qquad x \to \infty.$$

Fix $1 > \epsilon > 0$ and choose x such that

$$\mathbb{E}\left(|X|\mathbf{1}_{(|X|>x)}\right) = \mathbb{E}|Y_{x1}| < \epsilon^2.$$

Let $\mu_x = \mathbb{E}(X_{x1})$ and $\mu = \mathbb{E}(X)$. Then, we also have

$$|\mu_x - \mu| \le |\mathbb{E}(Y_{x1})| < \epsilon^2 < \epsilon.$$

Let $B_n = \{|U_{xn} - \mu_x| > \epsilon\}$ and $C_n = \{|V_{xn}| > \epsilon\}$. Noting that $\mathbb{E}(X_{xk}^2) \le x^2 < \infty$, we can apply the Weak Law of Large Numbers to U_{xn}. Thus, we choose $N > 0$ such that $\forall\ n > N$,

$$\mathbb{P}(B_n) = \mathbb{P}(|U_{nx} - \mu_x| > \epsilon) < \epsilon.$$

Now, by Chebyshev's inequality, we also have

$$\mathbb{P}(C_n) = \mathbb{P}(|V_{xn}| > \epsilon) \le \frac{\mathbb{E}|V_{xn}|}{\epsilon} \le \frac{\mathbb{E}|Y_{x1}|}{\epsilon} \le \epsilon$$

But on $B_n^c \cap C_n^c = (B_n \cup C_n)^c$, we have $|U_{xn} - \mu_x| \le \epsilon$ and $|V_{xn}| \le \epsilon$, and therefore

$$\left|\frac{S_n}{n} - \mu\right| \le |U_{xn} - \mu_x| + |V_{xn}| + |\mu_x - \mu| \le 2\epsilon + \epsilon^2 \le 3\epsilon.$$

Thus, $\forall\ n > N$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X\right| > 3\epsilon\right) \le \mathbb{P}(B_n \cup C_n) \le 2\epsilon.$$

$\blacksquare$

Lecture 9 : Convergence of random variables

(These notes are a revision of the work of Jin Kim, 2002.)

9.5 Convergence of random variables

First significant example: the *weak law of large numbers (WLLN).* We want to state that with a general notion of convergence in probability.

Definition 9.5.1 *Given a sequence of r.v's X_n defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, say X_n* converges in probability *to X, $X_n \xrightarrow{\mathbb{P}} X$, if X is a r.v. on $(\Omega, \mathcal{F})$, and for all $\epsilon > 0$,*

$$\lim_{m \to \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

Theorem 9.5.2 (Weak Law of Large Numbers) *Let $X, X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X| < \infty$. Then*

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}(X).$$

Other notions of convergence of r.v.'s:

Pointwise Convergence: $X_n(\omega) \longrightarrow X(\omega)$ for all $\omega \in \Omega$. This is a very strong notion: too strong for many purposes.

Almost Sure Convergence: We say $X_n \xrightarrow{a.s.} X$ if $X_n(\omega) \longrightarrow X(\omega)$ for all $\omega \notin N$, with $\mathbb{P}(N) = 0$, or equivalently $\mathbb{P}(\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty) = 1$.

Convergence in L^p ($p \geq 1$): We say $X_n \xrightarrow{\mathbf{L^p}} X$ if $\|X_n - X\|_p \longrightarrow 0$, i.e. $\lim_{n \to \infty} \mathbb{E}|X_n - X|^p = 0$.

Convergence in moments: We say $X_n \longrightarrow X$ in moments if $\|X_n\|_p \longrightarrow \|X\|_p$, for all $p \in \mathbb{N}$.

Convergence in Distribution: (Not really a notion of convergence of r.v.) A notion of convergence of a probability distribution on $\mathbb{R}$ (or more general space). We say $X_n \xrightarrow{d} X$ if $\mathbb{P}(X_n \leq x) \longrightarrow \mathbb{P}(X \leq x)$ for all x at which the RHS is continuous.

This weak convergence appears in the central limit theorem.

Fact 9.5.3 *(See text)* $X_n \xrightarrow{d} X \iff \mathbb{E}f(X_n) \longrightarrow \mathbb{E}f(X)$ *for all bounded and continuous function* f.

Properties in Common for $\xrightarrow{\mathbb{P}}, \xrightarrow{p.w.}, \xrightarrow{a.s.}, \xrightarrow{\mathbf{L}^p}$:

a) $X_n \longrightarrow X$, $Y_n \longrightarrow Y \Longrightarrow X_n + Y_n \longrightarrow X + Y$, $X_nY_n \longrightarrow XY$.

b) $X_n \longrightarrow X \iff (X_n - X) \longrightarrow 0$ (useful and common reduction).

c) For all of $\xrightarrow{\mathbb{P}}, \xrightarrow{a.s.}$, and $\xrightarrow{\mathbf{L}^p}$ the limit X is unique up to a.s. equivalence.

d) Cauchy sequences are convergent (completeness). (Need a metric to metrize $\xrightarrow{\mathbb{P}}$, but that is easily provided. See text.)

Theorem 9.5.4 *The following property holds among the types of convergence.*

$$\begin{tikzcd}
\boxed{X_n \xrightarrow{a.s.} X} \arrow[dr, Rightarrow] & & \boxed{X_n \xrightarrow{\mathbf{L}^p} X} \arrow[dl, Rightarrow, "(*)"] \\
& \boxed{X_n \xrightarrow{\mathbb{P}} X} \arrow[d, Rightarrow, "(**)"] & \\
& \boxed{X_n \xrightarrow{d} X} &
\end{tikzcd}$$

Proof: $(*)$ can be proved by Chebyshev's inequality (with usually $p = 2$):

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}(|X_n - X|^p)}{\epsilon^p}$$

$(**)$ is proved in the text. ■

Example 9.5.5 (Moving blip) *(An example showing that almost sure convergence is a stronger condition than convergence in probability.) Let* $\Omega = [0, 1]$ *with Lebesgue measure. For* $i \in \mathbb{N}$ *choose* n *such that* $2^n \leq i < 2^{n+1}$. *Let* $y_i = i/2^n$. *Define*

$$X_i(t) = \begin{cases} 1, & \text{if } t \in [y_i, y_i + 1/2^n]; \\ 0, & \text{else.} \end{cases}$$

Then $X_n \xrightarrow{\mathbb{P}} 0$, *but for no* $t \in [0, 1]$ *does* $X_n(t) \longrightarrow 0$.

Example 9.5.6 *Suppose that* $X_1, X_2, \ldots$ *are r.v.'s that have mean* 0*, have finite variances, and are uncorrelated. Let* $S_n = X_1 + \cdots + X_n$*. If* $\sum_{k=1}^{\infty} \mathbb{E}(X_k^2) < \infty$*, then* S_n *converges in* L^2 *to a limit* S_∞*, hence* $S_n \xrightarrow{\mathbb{P}} S_\infty$*, i.e.* $\lim_{n\to\infty} \mathbb{P}(|S_n - S_\infty| > \epsilon) = 0$ *for all* $\epsilon > 0$*.*

Proof: Look at the Cauchy criterion. Take $m > n$:

$$\mathbb{E}(S_m - S_n)^2 = \mathbb{E}\left(\sum_{k=n+1}^{m} X_k\right)^2 = \sum_{k=n+1}^{m} \mathbb{E}(X_k^2) \to 0$$

as $m, n \to \infty$. Therefore,

$$\sum_{k=1}^{\infty} \mathbb{E}(X_k^2) < \infty.$$

■

Fact 9.5.7 *If the* X_n *are independent (or more generally, martingale distributions), then* $S_n \xrightarrow{a.s.} S_\infty$*.*

The proof of this fact is deferred.

Fact 9.5.8 (Stout's Almost Sure Convergence) *There are examples of uncorrelated sequences with* $\sum_n X_n^2 < \infty$ *where a.s. convergence fails.*

9.6 Preliminaries for Study of a.s. Convergence

Definition 9.6.1 *Let* q_n *be some statement, true or false for each* n*. We say* q_n infinitely often *or (*q_n *i.o.) if for all* n *there is* $m \geq n$ *such that* q_m *is true, and (*q_n *ev.) if there exists* n *such that for all* $m \geq n$*,* q_m *is true. Now let* q_n *depend on* ω*, giving events*

$$A_n = \{\omega : q_n(\omega) \text{ is true}\}.$$

We now have new events,

$$\{A_n \text{ i.o.}\} = \{\omega : \omega \in A_n \text{ i.o.}\} = \bigcap_n \bigcup_{m \geq n} A_m,$$

and

$$\{A_n \text{ ev.}\} = \bigcup_n \bigcap_{m \geq n} A_m.$$

In analysis, $1_{(A_n \text{ i.o.})} = \lim_{n\to\infty} \sup_{m\geq n} 1_{A_m}$ and $1_{(A_n \text{ ev.})} = \lim_{n\to\infty} \inf_{m\geq n} 1_{A_m}$.

Given a sequence of events A_n for each $\omega \in \Omega$, consider $1_{A_n(\omega)}$ as a function of n, $\omega \longmapsto (1, 0, 0, 1, \dots)$.

Notice (de Morgan) that $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ ev.}\}$ and $\{A_n \text{ ev.}\}^c = \{A_n^c \text{ i.o.}\}$

Observe $X_n \xrightarrow{a.s.} X \Longleftrightarrow \forall\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon \text{ i.o.}) = 0$.

Argue this (Facts about convergence) $X_n \longrightarrow X \Longleftrightarrow \forall\epsilon > 0$, $|X_n - X| < \epsilon$ ev., so

$$\begin{aligned} X_n \xrightarrow{a.s.} X &\Longleftrightarrow \forall\epsilon > 0,\ \mathbb{P}(|X_n - X| \leq \epsilon \text{ ev.}) = 1 \\ &\Longleftrightarrow \forall\epsilon > 0,\ \mathbb{P}(|X_n - X| > \epsilon \text{ i.o.}) = 0. \end{aligned}$$

Because

$$X_n \longrightarrow X \text{ a.s.} \qquad \Longleftrightarrow \qquad X_n - X \longrightarrow 0 \text{ a.s.}\ ,$$

it is enough to prove for the case of convergence to 0.

Proposition 9.6.2 *The following are equivalent:*

1. $X_n \xrightarrow{a.s.} 0$
2. $\forall\epsilon > 0,\ \mathbb{P}(|X_n| > \epsilon \textit{ i.o.}) = 0$
3. $M_n \xrightarrow{\mathbb{P}} 0$ *where* $M_n := \sup_{n\leq k} |X_k|$
4. $\forall\ \epsilon_n \downarrow 0\ :\ \mathbb{P}(|X_n| > \epsilon_n \textit{ i.o.}) = 0$

Note: "$\forall$" in Proposition 4 cannot be replaced by "$\exists$". For example, Let $X_n = (1/\sqrt{n})U_n$, where $U_1, U_2, ...$ are independent $U[0, 1]$.

Take $\epsilon_n = 1/2/\sqrt{n}$. Then, $\mathbb{P}(X_n > \epsilon_n) = \mathbb{P}(U_n > 1/2) = 1/2$. So, $\mathbb{P}(X_n > \epsilon_n \text{ i.o.}) = 1$.

But if we take $\epsilon_n = \frac{1}{\sqrt{n}}$. Then, $\mathbb{P}(X_n > \epsilon_n) = \mathbb{P}(U_n > 1) = 0$.

Proof: (only for the equivalence of 1 and 3)

Suppose Proposition 1 holds. If $X_n(\omega) \to 0$ a.s., then $\sup_{n\leq k} |X_k(\omega)| \to 0$ a.s. But this implies that $M_n \to 0$ a.s. Thus, $M_n \xrightarrow{\mathbb{P}} 0$.

Conversely, if $M_n \downarrow$ as $n \uparrow$, then we know in advance that M_n has a almost-surely-limit in $[0, \infty]$. ■

Lemma 9.6.3 *If* $X_n \xrightarrow{\mathbb{P}} X$*, then there exists a subsequence* n_k *such that* $X_{n_k} \to X$ *a.s.*

Proof: It is enough to show that there exists $\epsilon_k \downarrow 0$ such that $\sum_k \mathbb{P}(|X_{n_k} - X| > \epsilon_k) < \infty$. We can take $\epsilon_k = 1/k$ and choose n_k so that $\mathbb{P}(|X_{n_k} - X| > 1/k) \leq 1/2^k$. Then, $\sum_k \mathbb{P}(|X_{n_k} - X| > \epsilon_k) < \infty$, and by **BCL I** we can conclude that $X_{n_k} \to X$ a.s. ■

Lecture 10 : Borel-Cantelli lemmas

10.7 Borel-Cantelli Lemmas

Recall that for real valued random variables X_n and X,

$$\begin{aligned}\{X_n \to X\} &= \{\omega : X_n(\omega) \to X(\omega)\} \\ &= \{\forall \epsilon > 0, |X_n - X| \leq \epsilon \;\text{ eventually}\}\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}(X_n \to X) = 1 \;&\Leftrightarrow \forall \epsilon > 0,\; \mathbb{P}(|X_n - X| \leq \epsilon \; ev.) = 1 \\ &\Leftrightarrow \forall \epsilon > 0,\; \mathbb{P}(|X_n - X| > \epsilon \; i.o.) = 0\end{aligned}$$

Let the event $A_n := \{|X_n - X| > \epsilon\}$. Then, we are motivated by consideration of a.s. convergence to find useful conditions for $\mathbb{P}(A_n \text{ i.o.}) = 0$.

Recall that $\{A_n \text{ i.o.}\} = \bigcap_n \bigcup_{m \geq n} A_m$.

Theorem 10.7.1 (Borel-Cantelli Lemmas) *Let $(\Omega, F, \mathbb{P})$ be a probability space and let (A_n) be a sequence of events in F. Then,*

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

2. *If $\sum_n \mathbb{P}(A_n) = \infty$ and A_n are independent, then $\mathbb{P}(A_n \; i.o.) = 1$.*

There are many possible substitutes for independence in BCL II, see Kochen-Stone Lemma.

Before proving BCL, notice that

- $\mathbf{1}_{A_n}$ i.o. $= \limsup_{n\to\infty} \mathbf{1}_{A_n}$
- $\mathbf{1}_{A_n}$ ev. $= \liminf_{n\to\infty} \mathbf{1}_{A_n}$
- $\{A_n \text{ i.o.}\} = \lim_{m\to\infty}(\cup_{n>m} A_n)$ (note: as $m \uparrow$, $\cup_{n\geq m} A_n \downarrow$)

- $\{A_n \text{ ev.}\} = \lim_{m\to\infty}(\cap_{n>m} A_n)$ (note: as $m \uparrow$, $\cap_{n\geq m} A_n \uparrow$).

Therefore,

$$\begin{aligned}
\mathbb{P}(A_n \text{ ev.}) &\leq \liminf_{n\to\infty} \mathbb{P}(A_n) && \text{by Fatou's lemma}\\
&\leq \limsup_{n\to\infty} \mathbb{P}(A_n) && \text{obvious from definition}\\
&\leq \mathbb{P}(A_n \text{ i.o.}) && \text{dual of Fatou's lemma (i.e. apply to } -\mathbb{P})
\end{aligned}$$

Proof: (Of BCL I)

$$\begin{aligned}
\mathbb{P}(A_n \text{ i.o.}) &= \lim_{m\to\infty} \mathbb{P}(\cup_{n\geq m} A_n)\\
&\leq \lim_{m\to\infty} \sum_{n\geq m}^{\infty} \mathbb{P}(A_n) \;=\; 0 \quad \text{since } \sum_{i=1}^{\infty} \mathbb{P}(A_n) < \infty.
\end{aligned}$$

■

Proof: (Of BCL II) Assume that $\Sigma\mathbb{P}(A_n) = \infty$ and the A_n's are independent. We will show that $\mathbb{P}(A_n^c \text{ ev.}) = 0$.

$$\begin{aligned}
\mathbb{P}(A_n^c \text{ ev.}) &= \lim_{n\to\infty} \mathbb{P}(\cap_{m\geq n} A_m^c) \;=\; \lim_{n\to\infty} \prod_{m\geq n} \mathbb{P}(A_m^c) && (10.3)\\
&= \lim_{n\to\infty} \prod_{m\geq n} (1 - \mathbb{P}(A_m)) \;\leq\; \lim_{n\to\infty} \prod_{m\geq n} \exp\left(-\mathbb{P}(A_m^c)\right) && (10.4)\\
&= \lim_{n\to\infty} \exp\left(-\sum_{m\geq n} \mathbb{P}(A_m^c)\right) = 0
\end{aligned}$$

since $(-\sum_{m\geq n} \mathbb{P}(A_m^c)) \to \infty,$ as $n \to \infty$

For (10.3), we used the following fact (due to the independence of A_n):

$$\mathbb{P}(\cap_{m\geq n} A_m^c) = \lim_{N\to\infty} \mathbb{P}(\cap_{n\leq m\leq N} A_m^c) = \lim_{N\to\infty} \prod_{n\leq m\leq N} \mathbb{P}(A_m^c) = \prod_{n\leq m} \mathbb{P}(A_m^c).$$

For (10.4), $1 - x \leq \exp(-x)$ was used. ■

For an example in which the theorem cannot be applied, consider $A_n = (0, 1/n)$ in $(0, 1)$. Then, $\mathbb{P}(A_n) = 1/n$, $\sum \mathbb{P}(A_n) = \infty$, but $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(\emptyset) = 0$.

Example 10.7.2 *Consider random walk in* $\mathbb{Z}^d$*,* $d = 0, 1, \cdots$ $S_n = X_1 + \cdots + X_n,$*,* $n = 0, 1, \cdots$ *where* X_i *are independent in* $\mathbb{Z}^d$*. In the simplest case, each* X_i *has uniform distribution on* 2^d *possible strings. i.e., if* $d = 3$*, we have* $2^3 = 8$ *neighbors*

$$\left\{ \begin{array}{c} (+1, +1, +1) \\ \vdots \\ (-1, -1, -1) \end{array} \right\}.$$

Note that each coordinate of S_n *does a simple coin-tossing walk independently. We can prove that*

$$\mathbb{P}(S_n = 0 \text{ } i.o.) = \left\{ \begin{array}{lll} 1 & \text{if } d = 1 \text{ or } 2 & (recurrent) \\ 0 & \text{if } d \geq 3 & (transient)\,. \end{array} \right. \tag{10.5}$$

Proof Sketch: (of (10.5))
Let us start with $d = 1$, then

$$\begin{aligned} \mathbb{P}(S_{2n} = 0) &= \mathbb{P}(n \text{ ``+'' signs and } n \text{ ``}-\text{'' signs}) && (10.6) \\ &= \binom{2n}{n} 2^{-2n} && (10.7) \\ &\sim \frac{c}{\sqrt{n}} \text{ as } n \longrightarrow \infty. && (10.8) \end{aligned}$$

where we used the facts that $n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$, and that $a_n \sim b_n$ iff $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

Note

$$\sum \left(\frac{1}{\sqrt{n}}\right)^d \left\{ \begin{array}{ll} = \infty & d = 1, 2 \\ < \infty & d = 3, 4, \cdots \end{array} \right. \tag{10.9}$$

Thus, $\sum_n \mathbb{P}(S_{2n} = 0) = \infty$, and BC II and (10.9) together gives (10.5). ■

Example 10.7.3 (for the case $d = 1$**)** $\{S_2 = 0\}$ *is the event of ending up back to the origin at step 2 when we started at the origin.* $\mathbb{P}(S_2 = 0) = 1/2$*. Note:*

$$\mathbb{P}(S_{10{,}000} = 0) \sim \frac{c}{\sqrt{n}} \approx 1/100,$$

$$\mathbb{P}(S_{10{,}002} = 0) \approx 1/100,$$

$$\mathbb{P}(S_{10{,}000} = 0, S_{10{,}002} = 0) = \mathbb{P}(S_{10{,}000} = 0)\mathbb{P}(S_{10{,}002} = 0 | S_{10{,}000} = 0) \approx 1/100 \cdot 1/2,$$

Later in the course, we will show that for the case $d = 1$*, even when the* $(S_{2n} = 0)$ *are dependent, it is still true that* $\mathbb{P}(S_{2n} = 0 \text{ } i.o. \text{ }) = 1$*.*

The same result holds for the case $d = 2$.

In general,

$$\mathbb{P}(S_{2n} = 0) = \frac{\binom{2n}{n}}{2^{2n}}^{d} \approx \frac{c^d}{n^{d/2}}.$$

For $d = 2$, this is $\sim \frac{c^2}{n}$ which is not summable. Thus, $\mathbb{P}(S_{2n} = 0 \text{ i.o.}) = 1$. For $d \geq 3$, this is $\sim \frac{c^3}{n^{3/2}}$ which is summable. Then, by **BCL I**, $\mathbb{P}(S_{2n} = 0 \text{ i.o.}) = 0$.

10.8 Almost sure convergence

Lecture 11 : Strong Law of Large Numbers

11.9 Showing almost sure convergence via subsequences

General settings/notation: let $S_n = X_1 + ... + X_n$. The X_i's are assumed independent, all defined on some probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. Sometimes, we assume the X_i are identically distributed.

An important technique for showing almost sure convergence from convergence in probability is to consider subsequences. We first note a few general facts about the various types of convergence we know:

1. If $Y_n \to Y$ a.s. then $Y_n \to Y$ in $\mathbb{P}$.
2. If $Y_n \to Y$ in $\mathbb{P}$ then there exists a fixed increasing subsequence n_k such that $Y_{n_k} \to Y$ a.s.
3. $Y_n \to Y$ in $\mathbb{P}$ iff for every subsequence n_k there exists a further subsequence n'_k so that $Y_{n'_k} \to Y$ a.s.

Proofs of 2 and 3 are in the textbook. We first begin with a technique which uses the information about almost sure convergence of a subsequence of a sequence of random variables, and then somehow getting control over a maximum. Let us now describe the technique.

One can prove $Y_n \to Y$ a.s. by first showing $Y_{n_k} \to Y$ a.s. for some n_k (we choose n_k) and then getting control over

$$M_k = \max_{n_k \leq m < n_{k+1}} |X_m - X_{n_k}|$$

In particular we must be able to show that $M_k \to 0$ a.s. because if $\omega \in \Omega$ is such that both $X_{n_k}(\omega) \to 0$ and $M_k(\omega) \to 0$ then we get (using the triangle inequality and th fact that the max is greater than the elements of set over which maximum is taken)

$$X_m(\omega) \to X(\omega),$$

so if $M_k \xrightarrow{a.s.} 0$, then the above holds almost surely. To illustrate how to use the technique, we start with the example of SLLN for bounded random variables.

Theorem 11.9.1 *If $X, X_1, X_2, ...$ are i.i.d. random variables with $\mathbb{E}(X) = \mu$, and there exists K such that $\mathbb{P}(|X| < K) = 1$, and $S_n := X_1 + X_2 + ... + X_n$, then*

$$\frac{S_n}{n} \to \mathbb{E}(X) \text{ a.s.} \tag{11.10}$$

Proof: First we find a subsequence converging almost surely to the mean. For that we use two tools:

- convergence in probability; and
- the Borel-Cantelli lemma.

Without loss of generality, we can assume that $E(X) = 0$. Since the X are bounded they have finite variance. From Chebyshev's inequality we get

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \frac{E(X^2)}{n\epsilon^2}.$$

This means that $\frac{S_n}{n} \to 0$ in $\mathbb{P}$. Notice that $\sum_k \frac{1}{k^2}$ converges to a finite value, therefore for the subsequence $n_k = k^2$ we get, using the Borel-Cantelli lemma,

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \epsilon \text{ i.o.}\right) = 0,$$

which means that $\frac{S_{n^2}}{n^2} \to 0$ a.s.

For convenience we define

$$D_n := \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$$

for $n^2 \leq k < (n+1)^2$. Notice that $|D_n| \leq K(2n+1)$

We have $|S_k| \leq |S_{n^2}| + D_n$ and $\frac{1}{k} \leq \frac{1}{n^2}$, so we have the following inequality:

$$\begin{aligned} \left|\frac{S_k}{k}\right| &\leq \left|\frac{S_{n^2}}{n^2}\right| + \frac{D_n}{n^2} \\ &\leq \left|\frac{S_{n^2}}{n^2}\right| + \frac{(2n+1)K}{n^2}. \end{aligned}$$

Since both of the terms on the left hand side converge to zero a.s. we have that the left hand side converge to 0 a.s. ■

Now we show how we can extend this to random variables with bounded variance.

Theorem 11.9.2 *If $X, X_1, X_2, ...$ are i.i.d. random variables with $\mathbb{E}(X) = \mu$, and there exists K such that $\mathbb{P}(|X| < K) = 1$, and $S_n := X_1 + X_2 + ... + X_n$, then*

$$\frac{S_n}{n} \to \mathbb{E}(X) \textit{ a.s.} \tag{11.11}$$

Proof: First we find a subsequence converging almost surely to the mean. For that we use two tools:

- convergence in probability; and
- the Borel-Cantelli lemma.

Without loss of generality, we can assume that $E(X) = 0$. Since the X are bounded they have finite variance. From Chebyshev's inequality we get

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \frac{E(X^2)}{n\epsilon^2}.$$

This means that $\frac{S_n}{n} \to 0$ in $\mathbb{P}$. Notice that $\sum_k \frac{1}{k^2}$ converges to a finite value, therefore for the subsequence $n_k = k^2$ we get, using the Borel-Cantelli lemma,

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \epsilon \text{ i.o.}\right) = 0,$$

which means that $\frac{S_{n^2}}{n^2} \to 0$ a.s.

Now let us try to control M_k as defined above. For convenience we define

$$D_n := \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$$

for $n^2 \le k < (n+1)^2$. We have $|S_k| \le |S_{n^2}| + D_n$ and $\frac{1}{k} \le \frac{1}{n^2}$, so we have the following inequality:

$$\left|\frac{S_k}{k}\right| \le \left|\frac{S_{n^2}}{n^2}\right| + \frac{D_n}{n^2}.$$

Finally, using the definition of M_k, we get the following:

$$\begin{aligned} M_k &\le \max_{n^2 \le k < (n+1)^2} \left|\frac{S_k}{k}\right| + \left|\frac{S_{n^2}}{n^2}\right| \\ &\le 2\left|\frac{S_{n^2}}{n^2}\right| + \frac{D_n}{n^2}. \end{aligned}$$

So all we need to prove is that $\frac{D_n}{n^2} \to 0$ a.s. Let us define a new quantity $T_m = S_{n^2+m} - S_{n^2}$. Therefore,

$$\begin{aligned} D_n^2 &= \max_{1 \le m \le 2n} T_m^2 \\ &\le \sum_{m=1}^{2n} T_m^2. \end{aligned}$$

Taking expectations on both sides, we get that

$$\begin{aligned} E(D_n^2) &\le \sum_{m=1}^{2n} m\sigma^2 = n(2n+1)\sigma^2 \\ &\le 4n^2\sigma^2, \end{aligned}$$

where $E(X^2) = \sigma^2$. Hence we get that

$$\begin{aligned} \mathbb{P}\left(\left|\frac{D_n}{n^2}\right| > \epsilon\right) &\le \frac{E\left(\left(\frac{D_n}{n^2}\right)^2\right)}{\epsilon^2} \\ &\le \frac{4\sigma^2}{n^2\epsilon^2}. \end{aligned}$$

Applying the Borel-Cantelli lemma with the fact

$$\sum_n \mathbb{P}\left(\left|\frac{D_n}{n^2}\right| > \epsilon\right) < \infty$$

we get that $\frac{D_n}{n^2} \to 0$ a.s., which completes the proof. ■

11.10 Kolmogorov's Maximal Inequality

Now we proceed to Kolmogorov's inequality. We formally state it as follows.

Theorem 11.10.1 (Kolmogorov's Inequality) *Let* $X_1, X_2, ...$ *be independent with* $E(X_i) = 0$ *and* $\sigma_i^2 = E(X_i^2) < \infty$, *and define* $S_k = X_1 + X_2 + ... + X_k$. *Then*

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge \epsilon\right) \le \frac{E(S_n^2)}{\epsilon^2}. \tag{11.12}$$

Proof: Decompose the event according to when we escape from the $\pm\epsilon$ strip. Let

$$A_k = \{|S_m| < \epsilon \text{ for } 1 \le m < k;\ |S_k| \ge \epsilon\}$$

In words, A_k is the event that the first escape out of the ϵ strip occurs at the kth step. Also notice that all these events are disjoint, and that $\bigcup_{k=1}^{n} A_k \left\{\max_{1\leq k\leq n} |S_k| \geq \epsilon\right\}$. Then,

$$E(S_n^2) \geq E\left(S_n^2 1\left(\bigcup_{k=1}^{n} A_k\right)\right)\sum_{k=1}^{n} E\left(S_n^2 1_{A_k}\right).$$

We can split $S_n^2 = S_k^2 + (S_n - S_k)^2 + 2S_k(S_n - S_k)$, and write

$$\begin{aligned} E\left(S_n^2 1_{A_k}\right) &= E\left(S_k^2 1_{A_k}\right) + E\left((S_n - S_k)^2 1_{A_k}\right) + E\left(2(S_n - S_k)S_k 1_{A_k}\right) \\ &\geq \epsilon^2 \mathbb{P}(A_k), \end{aligned}$$

where the first term is larger than ϵ^2, the second term is always positive, and the third term is the expectation of a product of two independent random variables witn mean 0.

Finally, we put this into the summation to get

$$E(S_n^2) \geq \sum_{k=1}^{n} \mathbb{P}(A_k)\epsilon^2 = \mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \epsilon\right)\epsilon^2,$$

which easily leads to the result. This completes the proof. ■

We observe that the inequality is valid for any sequence of r.v.'s $(X_1, ...X_n)$ such that

$$E\left(2(S_n - S_k)S_k 1_{A_k}\right) = 0.$$

This will lead to the definition in future lectures of a *martingale difference sequence.*

Lecture 12 : Basic $\mathcal{L}^2$ Convergence Theorem

12.11 Basic $\mathcal{L}^2$ Convergence Theorem

Theorem 12.11.1 (Basic $\mathcal{L}^2$ Convergence Theorem) *Let* X_1 X_2, ... *be independent random variables with* $\mathbb{E}(X_i) = 0$ *and* $\mathbb{E}(X_i^2) = \sigma_i^2 < \infty$, $i = 1, 2, \ldots$, *and* $S_n = X_1 + X_2 + \cdots + X_n$. *If* $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$, *then* S_n *converges a.s. and in* $\mathcal{L}^2$ *to some* S_∞ *with* $\mathbb{E}(S_\infty^2) = \sum_{i=1}^{\infty} \sigma_i^2$.

Recall: We have done this before, with the conclusion for the $\mathcal{L}^2$ case with the weaker assumption that $\mathbb{E}(X_i X_j) = 0$ for $i \neq j$. The only new thing is the conclusion of a.s. convergence for the independent case. In fact, the proof just uses Kolmogorov's inequality from the last lecture. Thus the conclusion is valid for a martingale $\{S_n\}$ with $\mathbb{E}[X_{n+1} f(X_1, \ldots, X_n)] = 0$ for all bounded measurable $f : \mathbb{R}^n \to \mathbb{R}$.

Proof: First note that $\mathcal{L}^2$ convergence and existence of S_∞ is implied by the orthogonality of the X_i's: since $\mathbb{E}(X_i X_j) = 0$ for $i \neq j$,

$$\begin{aligned} \mathbb{E}(S_n^2) &= \sum_{i=1}^{n} \sigma_i^2, \text{ and} \\ \mathbb{E}((S_n - S_m)^2) &= \sum_{i=m+1}^{n} \sigma_i^2 \to 0 \text{ as } m, n \to \infty, \end{aligned}$$

so S_n is Cauchy in $\mathcal{L}^2$. Since $\mathcal{L}^2$ is complete, there is a unique S_∞ (up to a.s. equivalence) such that $S_n \to S_\infty$ in $\mathcal{L}^2$.

Turning to a.s. convergence, the method is to show the sequence (S_n) is a.s. Cauchy. The limit of S_n then exists a.s. by completeness of the set of real numbers. The same argument applies more generally to martingale differences X_i. Note that this method gives S_∞ more explicitly, and does not appeal to completeness of $\mathcal{L}^2$.

Recall that S_n is Cauchy a.s. means $M_n := \sup_{p,q \geq n} |S_p - S_q| \to 0$ a.s. Note that $0 \leq M_n(\omega) \downarrow$ implies that $M_n(\omega)$ converges to a limit in $[0, \infty]$. So, if $\mathbb{P}(M_n > \epsilon) \to 0$ for all $\epsilon > 0$, then $M_n \downarrow 0$ a.s.

Let $M_n^* := \sup_{p \geq n} |S_p - S_n|$. By the triangle inequality,

$$|S_p - S_q| \leq |S_p - S_n| + |S_q - S_n| \Rightarrow M_n^* \leq M_n \leq 2M_n^*,$$

so it is sufficient to show that $M_n^* \xrightarrow{P} 0$.

For all $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{p\geq n}|S_p - S_n| > \epsilon\right) = \lim_{N\to\infty}\mathbb{P}\left(\max_{n\leq p\leq N}|S_p - S_n| > \epsilon\right)$$
$$\leq \lim_{N\to\infty}\sum_{i=n+1}^{N}\frac{\sigma_i^2}{\epsilon^2} = \sum_{i=n+1}^{\infty}\frac{\sigma_i^2}{\epsilon^2}$$

where we applied Kolmogorov's inequality in the second step. Since $\sum_{i=1}^{\infty}\sigma_i^2 < \infty$,

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{p\leq n}|S_p - S_n| > \epsilon\right) = 0$$

■

Remark: Just orthogonality rather than independence of the X_is is not enough to get an a.s. limit. Counterexamples are hard. According to classical results of Rademacher-Menchoff, for orthogonal X_i the condition $\sum_i(\log^2 i)\sigma_i^2 < \infty$ is enough for a.s. convergence of S_n, whereas if $b_i \uparrow$ with $b_i = o(\log^2 i)$ there exist orthogonal X_i such that $\sum_i b_i\sigma_i^2 < \infty$ and S_n diverges almost surely.

12.12 Kolmogorov's Three-Series Theorem

An easy consequence of the Basic $\mathcal{L}^2$ Convergence Theorem is the sufficiency part of Kolmogorov's three-series theorem:

Theorem 12.12.1 (Kolmogorov) *Let $X_1, X_2, \ldots$ be independent. Fix $b > 0$. Convergence of the following three series*

- $\sum_n \mathbb{P}(|X_n| > b) < \infty$
- $\sum_n \mathbb{E}(X_n\mathbf{1}_{(|X_n|<b)})$ *converges to a finite limit*
- $\sum_n \mathbf{V}ar(X_n\mathbf{1}_{(|X_n|<b)}) < \infty$

is equivalent to $\mathbb{P}(\sum_n X_n$ *converges to a finite limit*$) = 1$

Note: If any one of the three series diverges then

$$\mathbb{P}\left(\sum_n X_n \text{ converges to a finite limit}\right) = 0$$

by Kolmogorov's zero-one law (will be shown later). Note also that if one or more of the series diverges for some b, then one or more of the series must diverge for every b, but exactly which of the three series diverge may depend on b. Examples can be given of 8 possible combinations of convergence/divergence.

Proof: [Proof of sufficiency] That is, convergence of all 3 series implies $\sum_n X_n$ converges a.s. Let $X'_n = X_n 1_{(|X_n| \le b)}$. Since $\sum_n \mathbb{P}(X'_n \neq X_n) = \sum_n \mathbb{P}(|X_n| > b) < \infty$, the Borel-Cantelli lemma gives $\mathbb{P}(X'_n \neq X_n \text{ i.o.}) = 0$ which implies $\mathbb{P}(X'_n = X_n \text{ ev.}) = 1$. Also if $X'_n(\omega) = X_n(\omega)$ ev., then $\sum_n X_n(\omega)$ converges $\Leftrightarrow \sum_n X'_n(\omega)$ converges.

Therefore it is enough to show that

$$\mathbb{P}\left(\sum_n X'_n \text{ converges to a finite limit}\right) = 1$$

Now

$$\sum_{n=1}^{N} X'_n = \sum_{n=1}^{N} (X'_n - \mathbb{E}(X'_n)) \; + \; \sum_{n=1}^{N} \mathbb{E}(X'_n).$$

$\sum_{n=1}^{N} \mathbb{E}(X'_n)$ has a limit as $N \to \infty$ by hypothesis, and

$$\sum_{n=1}^{\infty} \mathbb{E}((X'_n - \mathbb{E}(X'_n))^2) = \sum_{n=1}^{\infty} \mathrm{var}(X'_n) < \infty$$

implies that $\sum_{n=1}^{\infty}(X'_n - \mathbb{E}(X'_n))$ converges a.s. by the basic $\mathcal{L}^2$ convergence theorem. ■

12.13 Kolmogorov's 0-1 Law

$X_1, X_2, \ldots$ are independent random variables (not necessarily real valued). Let $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots) =$ the future after time $n =$ the smallest σ-field with respect to which all the $X_m, m \geq n$ are measurable. Let $\mathcal{T} = \cap_n \mathcal{F}'_n =$ the remote future, or *tail* $\sigma-$field.

Example 12.13.1 *$\{\omega : S_n(\omega)$ converges$\} \in \mathcal{T}$.*

Theorem 12.13.2 (Kolmogorov's 0-1 Law) *If $X_1, X_2, \ldots$ are independent and $A \in \mathcal{T}$ then $\mathbb{P}(A) = 0$ or 1.*

Proof: The idea is to show that A is independent of itself, that is, $\mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$, so $\mathbb{P}(A) = \mathbb{P}(A)^2$, and hence $\mathbb{P}(A) = 0$ or 1. We will prove this in two steps:

(a) $A \in \sigma(X_1, \ldots, X_k)$ and $B \in \sigma(X_{k+1}, X_{k+2}, \ldots)$ are independent.

Proof of (a): If $B \in \sigma(X_{k+1}, \ldots, X_{k+j})$ for some j, this follows from (4.5) in chapter 1 of [?]. Since $\sigma(X_1, \ldots, X_k)$ and $\cup_j \sigma(X_{k+1}, \ldots, X_{k+j})$ are π-systems that contains Ω (a) follows from (4.5) in chapter 1 of [?]).

(b) $A \in \sigma(X_1, X_2, \ldots)$ and $B \in \mathcal{T}$ are independent.

Proof of (b): Since $\mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, \ldots)$, if $A \in \sigma(X_1, \ldots, X_k)$ for some k, this follows from (a). $\cup_k \sigma(X_1, \ldots, X_k)$ and $\mathcal{T}$ are π-systems that contain Ω, so (b) follows from (4.2) in chapter 1 of [?].

Since $\mathcal{T} \subset \sigma(X_1, X_2, \ldots)$, (b) implies that $A \in \mathcal{T}$ is independent of itself and the theorem follows.

■

Recall *Kronecker's lemma*: If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} X_n/a_n$ converges a.s., then $(\sum_{m=1}^{n} X_m)/a_n \xrightarrow{a.s.} 0$.

Let $X_1, X_2, \ldots$ be independent with mean 0 and $S_n = X_1 + X_2 + \cdots + X_n$. If $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2)/a_n^2 < \infty$, then by the basic $\mathcal{L}^2$ convergence theorem $\sum_{n=1}^{\infty} X_n/a_n$ converges a.s. Then $S_n/a_n \to 0$ a.s.

Example 12.13.3 *Let $X_1, X_2, \ldots$ be i.i.d., $\mathbb{E}(X_i) = 0$, and $\mathbb{E}(X_i^2) = \sigma^2 < \infty$.*
Take $a_n = n$:

$$\sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty \Rightarrow \frac{S_n}{n} \overset{a.s.}{\to} 0.$$

Now take $a_n = n^{\frac{1}{2}+\epsilon}$, $\epsilon > 0$:

$$\sum_{n=1}^{\infty} \frac{\sigma^2}{n^{1+2\epsilon}} < \infty \Rightarrow \frac{S_n}{n^{\frac{1}{2}+\epsilon}} \overset{a.s.}{\to} 0.$$

Lecture 10 : Kolmogorov's Law of Large Numbers

(These notes are a revision of the work of Vinod Prabhakaran, 2002.)

10.14 Law of the Iterated Logarithm

Let X_1, X_2, ... be i.i.d. with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = \sigma^2$, $S_n = X_1 + ... + X_n$. We know

$$\frac{S_n}{n^{\frac{1}{2}+\varepsilon}} \xrightarrow{a.s} 0 \text{ as } n \to \infty.$$

We will show later

$$\frac{S_n}{\sigma n^{\frac{1}{2}}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

For general interest, we state, without proof, the *Law of the Iterated Logarithm*:

$$\limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = 1 \text{ a.s.}$$

$$\liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = 1 \text{ a.s.}$$

$$\mathbb{P}(S_n > (1+\varepsilon)\sigma\sqrt{2n\log(\log n)} \text{ i.o.}) = 0$$

$$\mathbb{P}(S_n > (1-\varepsilon)\sigma\sqrt{2n\log(\log n)} \text{ i.o.}) = 0$$

10.15 Kolmogorov's Law of Large Numbers

Theorem 10.15.1 *Let X_1, X_2, ... be i.i.d. with $\mathbb{E}(|X_i|) < \infty$, $S_n = X_1 + ... + X_n$. Then $S_n/n \to \mathbb{E}(X)$ a.s. as $n \to \infty$.*

Note that the theorem is true with just pairwise independence instead of the full independence assumed here [[**?**], p.55 (7.1)]. The theorem also has an important generaliztion to stationary sequences (the *ergodic theorem*, [[**?**], p.337 (2.1)]).

Proof: *Step 1*: Replace X_i by $\widetilde{X}_i = X_i - \mathbb{E}X$ (note $\mathbb{E}X_i = \mathbb{E}X$). Then

$$\frac{\widetilde{S}_n}{n} = \frac{S_n}{n} - \mathbb{E}X.$$

So it's enough to consider $\mathbb{E}X = 0$.

Step 2: Now we assume $\mathbb{E}X = 0$. Introduce truncated variables

$$\widehat{X}_n := X_n I\left(|X_n| \le n\right).$$

Observe that

$$\mathbb{P}(X_n = \widehat{X}_n \ ev.) = 1.$$

(To see this, check

$$\mathbb{P}(X_n \ne \widehat{X}_n \ i.o.) = \mathbb{P}(|X_n| > n \ i.o.)$$

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n\ ) = \sum_{n=1}^{\infty} \mathbb{P}(|X| > n\ ) = \mathbb{E}\left(\sum_{n=1}^{\infty} I\left(|X| > n\right)\right) < \infty$$

since

$$\sum_{n=1}^{\infty} I\left(|X| > n\right) = \sum_{1 \le n < X} 1 \le |x| + 1.$$

Compare this to the tail sum formula for a random variable$X$ with values in $0, 1, 2, ...$

$$\mathbb{E}X = \sum_{n=1}^{\infty} n\mathbb{P}(X = n) = \sum_{n=1}^{\infty} \mathbb{P}(X \ge n).$$

Step 3: Center the truncated variables. Define $\widetilde{X}_n := \widehat{X}_n - \mathbb{E}\left(\widehat{X}_n\right)$.
We will show that

$$\left(\frac{S_n}{n} \to 0\right) \overset{a.s.}{\underset{\text{(a)}}{=}} \left(\frac{\hat{S}_n}{n} \to 0\right) \overset{a.s.}{\underset{\text{(b)}}{=}} \left(\frac{\tilde{S}_n}{n} \to 0\right),$$

where $\hat{S}_n = \hat{X}_1 + \hat{X}_2 + \cdots + \hat{X}_n$ and $\tilde{S}_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n$. Then using Kronecker's lemma we will show that $\mathbb{P}\left(\tilde{S}_n / n \to 0\right) = 1$.

(a) comes from the fact that if $\omega \in \left\{\omega : X_n(\omega) = \widehat{X}_n(\omega) \ \ ev.\right\}$ (which has probablity 1), then $S_n(\omega) - \widehat{S}_n(\omega)$ is eventually not dependent on n. So

$$\frac{S_n(\omega) - \widehat{S}_n(\omega)}{n} \to 0 \ \text{ for such } \omega.$$

(b) comes from

$$\frac{\widehat{S}_n}{n} - \frac{\widetilde{S}_n}{n} = \frac{\mathbb{E}\widehat{X}_1 + \mathbb{E}\widehat{X}_2 + ... + \mathbb{E}\widehat{X}_n}{n} \to 0 \text{ as } n \to \infty \ \ (\text{By analysis and } \mathbb{E}\widehat{X}_i \to 0)$$

But

$$\mathbb{E}\widehat{X}_n = \mathbb{E}[X_n I\left(|X_n| \leq n\right)] = \mathbb{E}[X I\left(|X| \leq n\right)] \rightarrow \mathbb{E}X \text{ as } n \rightarrow \infty.$$

as the integrand is dominated by $|X|$ and note
$E(|X|) < \infty$.

Now, we use Kronecker's lemma and the $\mathcal{L}^2$ convergence theorem to show that

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\left(\widetilde{X}_n^2\right)}{n^2} < \infty.$$

$$\begin{aligned}\mathbb{E}\left(\widetilde{X}_n^2\right) = \mathbb{E}\left[\left(\widehat{X}_n - \mathbb{E}\left(\widehat{X}_n\right)\right)^2\right] &= \mathbb{E}\left[\left(X\mathbf{1}\left(|X| \leq n\right) - \mathbb{E}(X\mathbf{1}\left(|X| \leq n\right))\right)^2\right] \\ &\leq \mathbb{E}(X\mathbf{1}\left(|X| \leq n\right))^2 .\end{aligned}$$

So

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathbb{E}\left(\widetilde{X}_n^2\right)}{n^2} \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}X^2 I\left(|X| \leq n\right)}{n^2} &= \mathbb{E}\left(X^2 \sum_{n=1}^{\infty} \frac{I\left(|X| \leq n\right)}{n^2}\right) \\ &\approxeq \mathbb{E}\left(\frac{X^2}{|X|}\right) = \mathbb{E}(|X|) < \infty.\end{aligned}$$

This came from

$$\sum_{n=1}^{\infty} \frac{x^2 1_{(|x| \leq n)}}{n^2} \approxeq x^2 \sum_{n=1}^{\infty} \frac{1}{n^2} \approxeq x^2 \frac{1}{|x|} \approxeq |x| .$$

■

10.16 Convergence in distribution

Definition 10.16.1 *$X_n \xrightarrow{d} X$ if $\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x)$ for all x at which $x \rightarrow \mathbb{P}(X \leq x)$ is continuous. We call this* convergence in distribution *or* weak convergence*.*

Note. This is really a notion of convergence of probability measures rather than of convergence of random variables. Now the limit random variable X is only unique in distribution, not unique almost surely. Obviously, if $X \overset{d}{=} Y$ and $X_n \xrightarrow{d} X$, then $X_n \xrightarrow{d} Y$.

Theorem 10.16.2 (Skorokhod) *$X_n \xrightarrow{d} X \Longleftrightarrow$ there exists a probability with space random variables Y_n with $Y_n \overset{d}{=} X_n$, $Y \overset{d}{=} X$ and $Y_n \xrightarrow{a.s.} Y$.*

Proof: Take a single uniform variable U and use it to create the Y_n and Y. Let

$$F_n(x) = \mathbb{P}(X_n \leq x)\,,$$

$$F(x) = \mathbb{P}(X \leq x)\,,$$
$$Y_n = F_n^{-1}(U),$$
$$Y = F^{-1}(U).$$

Check that

$$F^{-1}(U) = \inf\left\{x : f(x) > u\right\}.$$

■

The following proposition is an application of the above.

Proposition 10.16.3 *If $X_n \xrightarrow{d} X$, then for every bounded continuous function $f : R \to R$*

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]\,.$$

Proof: Without loss of generality, $X_n \overset{a.s.}{\to} X$. Then $f(X_n) \to f(X)$ is bounded, so we can take expectations and use the bounded convergence theorem. ■