Lectures on Probability Theory

General recommendations.

- These lectures assume that the audience is familiar with measure theory.
- The videos do not replace the books. I suggest to choose one among the many listed at the end of these notes and to read the corresponding sections before or after the videos.
- After the statement of a result, interrupt the video and try to prove the assertion. It is the only way to understand the difficulty of the problem, to differentiate simple steps from crucial ones, and to appreciate the ingenuity of the solution. Sometimes you find an alternative proof of the result.
- You can speed-up or slow-down the video. By pressing settings at the bottom-right corner, you can modify the playback speed.
- Exercises highlighted in blue present results which will be used later in the lectures and are highly recommended, as well as those indicated with $*$.
- Send me an e-mail if you find a mistake which is not reported in these notes.
- If you typed in latex, with no personal definitions nor the use of special packages, solutions to some exercises proposed below, send the file. Hopefully, I'll create a note with solutions to the exercises, acknowledging the authors of the solutions.
- A note about the methodology. I ask the students to view the video(s) before the class. In the first part of the lecture, I recall the content of the video. Sometimes, I ask one of the students to replace me. Occasionally, the student is randomly chosen. This is the opportunity for the students to ask questions on the content of the class. In the second part of the lecture, I present some of the applications included in the "Further Readings" topic.

December 2, 2020

Lecture 1: Introduction

Summary. This lecture is based on Sections 3.1 and 3.2 of [Chung].

Content and Comments.

- 0:00 Definition of random variables, probability distribution measures and distribution functions.
- 8:14 One-to-one correspondance between probability distribution measures and distribution functions.
- 12:32 Definition of discrete random variables and discrete distribution functions.
- 15:38 Definition of absolutely continuous and singular distributions. The Cantor distribution is constructed at the end of [Chung, Section 1.3].
- 20:24 $F = F_d + F_{\rm ac} + F_s$. [Chung, Theorem 1.3.2]
- 22:09 Definition of expectation
- 24:28 For a non-negative random variable,

$$\sum_{n\ge 1} P[\, X \ge n\,] \;\le\; E[X] \;\le\; 1 + \sum_{n\ge 1} P[\, X \ge n\,]$$

  This is [Chung, Theorem 3.2.1].
- 37:05 $\int_\Omega f(X)\, dP \,=\, \int_{\mathbb{R}} f(x)\, \mu_X(dx)$, [Chung, Theorem 3.2.2].
- 40:32 Jensen's inequality, [Chung, Section 3.2]. See comment below on convex functions.
- 47:28 Cebyshev's inequality, [Chung, Section 3.2].

On convex functions. Let I be an open interval of $\mathbb{R}$ (which may be equal to $\mathbb{R}$). Consider a real-valued convex function $\varphi : I \to \mathbb{R}$. Show that φ is continuous on I and that it has left and right-derivatives at every point. Denote by $(D_+\varphi)(x)$ the right-derivative of φ at x. Show that for all $x_0 \in I$, $(D_+\varphi)(x_0)\,(x - x_0) + \varphi(x_0) \le \varphi(x)$. This bound is used in the proof of Jensen's inequality.

Further Readings.

A. [Varadhan, Chapter 1] presents a review on measure theory.
B. [Breiman, Chapter 2] examines the properties of the distribution functions of random vectors and presents Kolmogorov's extension theorem. Be aware that Breiman defined the distribution function as $F_X(x) = P[X < x]$, while we adopt in these lectures the convention that $F_X(x) = P[X \le x]$.
C. [Durrett, Sections 1.1 – 1.3] has many examples.

Recommended exercises.

a. In Section 3.1 of [Chung], prove Theorems 1 to 6 (that is Theorems 3.1.1 to 3.1.6).
b. Section 3.1 of [Chung], exercises 3, 4, 5, 11.
c. In Section 3.2 of [Chung], prove Theorems 2 and 3.
d. Section 3.2 of [Chung], exercises 2, 5, 6, 7, 11, 12, 13, 14, 16, 19

Suggested exercises.

a. Section 3.1 of [Chung], exercises 6, 10
b. Section 3.2 of [Chung], exercises 1, 4, 8, 10, 15, 17, 18

Lecture 2: Independence

Summary. This lecture is based on Section 3.3 of [Chung].

Content.

0:00 Definition of independent random variables
4:47 Subfamilies of independent random variables are independent
8:12 Definition of distribution function and probability measure of a random vector
12:43 Lemma. A finite set of random variables is independent if and only if the distribution function of the random vector is equal to the product of the distribution functions.
18:15 Lemma. A finite set of random variables is independent if and only if the probability measure of the random vector is equal to the product of the probability measures.
19:12 Theorem. Let $X_1, \dots, X_N$ be independent random variables and $f_1, \dots, f_N$, $f_j : \mathbb{R} \to \mathbb{R}$, measurable functions. Then, $f_1(X_1), \dots, f_N(X_N)$ are independent random variables.
22:52 Theorem. Let $X_1, \dots, X_N$ be independent random variables, $n_0 = 0$, $1 \le n_1 < n_2 < \cdots < n_p = N$ and $f_1, \dots, f_p$, $f_j : \mathbb{R}^{n_j - n_{j-1}} \to \mathbb{R}$, measurable functions. Then, $f_1(X_1, \dots, X_{n_1}), \dots, f_p(X_{n_{p-1}+1}, \cdots, X_{n_p})$ are independent random variables.
26:19 Theorem. Let X, Y be independent random variables such that $E[\,|X|\,] < \infty$, $E[\,|Y|\,] < \infty$. Then, $E[\,X\,Y\,] = E[\,X\,]\,E[\,Y\,]$.
43:17 $\int_\Omega f(X)\, dP = \int_{\mathbb{R}} f(x)\, \mu_X(dx)$.
45:53 Second proof of the identity $E[\,X\,Y\,] = E[\,X\,]\,E[\,Y\,]$.
52:03 Construction of a product measure on an infinite product space.

Comments and References.

17:40 Lemma. Assume that $n = 2$. Fix x_1. Denote by $\mathcal{M}$ the class of sets B which satisfies the identity

$$P\Big[\, X_1 \le x_1 \;\, X_2 \in B\,\Big] \;=\; P\big[\, X_1 \le x_1\big]\; P\big[X_2 \in B\,\big]\;.$$

Show that $\mathcal{M}$ is a monotone class and contains the algebra generated by the intervals $(-\infty, a]$. Apply the monotone class theorem, Theorem 1.5 of [Taylor, Section 1.5], to conclude that the previous identity holds for all Borel sets B. Fix a Borel set B_0. Denote by $\mathcal{M}$ the class of sets B which satisfies the identity

$$P\Big[\, X_1 \in B \;\, X_2 \in B_0\,\Big] \;=\; P\big[\, X_1 \in B\big]\; P\big[X_2 \in B_0\,\big]\;.$$

Show that $\mathcal{M}$ is a monotone class and contains the algebra generated by the intervals $(-\infty, a]$. Apply the monotone class theorem, Theorem 1.5 of [Taylor, Section 1.5], to conclude that the previous identity holds for all Borel sets B.
28:00 Theorem. The construction of the integral is presented from Section 5.1 to 5.3 of [Taylor]. Fubini's and Tonelli's theorems can be found in Section 6.3 of [Taylor].

52:03 Details of the construction can be found in Lecture 17 of the course on measure theory and in Section 6.6 of [Taylor].

Further readings.

A. [Taylor] for all results on measure theory used in this lecture.
B. [Chung, Section 3.3] provides many details skipped in the lecture and further examples.
C. [Breiman, Section 3.1] presents independence from a slightly different point of view.
D. [Durrett, Section 1.4] gives many examples.

Suggested exercises.

a. [Chung, Section 3.3], exercises 4, 8, 9, 10, 14, 15.
b. [Varadhan], exercise 23 of Chapter 1 and exercise 4 of chapter 3
c. Give and example of three random variables X, Y and Z defined on the same probability space and such that X and Y are independent, Y and Z are independent, X and Z are independent, but X, Y and Z are not independent.
d. [Breiman, Section 3.1], problems 1 and 2.
e. [Durrett, Section 1.4] exercises 2, 4, 5, 6, 12, 13, 16, 17, 19.

Lecture 3: Applications of Independence

Summary. The first two applications of this lecture can be found in Section 1.5 of [Durrett], the last one in Chapter 3 of [Varadhan-LD].

Content and Comments.

0:00 Weak law of large numbers.
8:06 Convergence in L^p implies convergence in probability.
14:25 Bernstein polynomials approximate uniformly continuous functions.
33:57 An upper bound for large deviations. This result is known as Cramer's theorem.

Further readings.

A. Interesting examples (coupon collector, random permutation, occupancy problems, St. Petersburg paradox) can be found in Section 1.5 of [Durrett].
B. More details on large deviations of i.i.d. random variables can be found in [Varadhan-LD, Deuschel-Stroock, Dembo-Zeitouni].

Suggested exercises.

a. Exercises in Section 1.5 of [Durrett].

Lecture 4: Convergence of random variables

Summary. This lecture is based on [Chung, Section 4.1].

Content and Comments.

- 0:00 Definition of almost sure convergence
- 2:09 A necessary and sufficient condition for almost sure convergence. [Chung, Theorem 4.1.1]
- 12:24 Definition of convergence in probability
- 13:19 Almost sure convergence implies convergence in probability. [Chung, Theorem 4.1.2]
- 15:18 A sequence which converges in probability admits a subsequence which converges almost surely. [Chung, Theorem 4.2.3]
- 25:17 Definition of convergence in L^p
- 26:03 Convergence in L^p implies convergence in probability. [Chung, Theorem 4.1.4]
- 27:55 A sequence dominated by a random variable in L^p and which converges in probability also Converges in L^p. [Chung, Theorem 4.1.4]
- 35:07 The Corollary of [Taylor, Theorem 5.6] is applied here.
- 37:33 An example of a sequence which converges almost surely and does not converge in L^p. Note that this sequence converges almost surely to 0 and not only in probability. See Example 2 of [Chung, Section 4.1].
- 42:11 An example of a sequence which converges in L^p and does not converge almost surely. See Example 1 of [Chung, Section 4.1]

Further readings.

A. [Breiman, Section 2.8] for the definition of Cauchy sequences and their properties.

Recommended exercises.

a. [Chung, Section 4.1], exercises 4, 7, 8, 9, 10, 15, 18, 20
b. [Breiman, Section 2.8], problems 12, 13 and 14. Problem 14 asks to prove a result similar to the one used in the lecture at time 0:00.

Suggested exercises.

a. [Chung, Section 4.1], exercises 1, 3, 5, 6, 11, 12, 19

Lecture 5: Borel-Cantelli lemma

Summary. This lecture is based on [Chung, Section 4.2] and [Durrett, Section 1.6].

Content and Comments.

- 0:00 Definition of $\limsup_n E_n$, $\liminf_n E_n$
- 1:50 $P[\limsup_n E_n] = \lim_n P[\cup_{m\geq n} E_m]$
- 5:14 $\limsup_n E_n = \{\, E_n \text{ i. o. } \} := \{\omega : \omega \in E_n \text{ i. o. } \}$
- 15:07 [Chung, Theorem 4.2.1]. $\sum_{n\geq 1} P[E_n] < \infty \Rightarrow P[\, E_n \text{ i. o. }] = 0$.
- 17:28 Application. [Durrett, Theorem 1.6.5] X_n i.i.d., $E[X_1^4] < \infty \Rightarrow (X_1 + \cdots + X_n)/n \to E[X_1]$ a. e.
- 29:46 [Chung, Theorem 4.2.4]. $(E_n : n \geq 1)$, independent, $\sum_{n\geq 1} P[E_n] = \infty \Rightarrow P[E_n \text{ i. o. }] = 1$.
- 35:21 Remark: $(E_n : n \geq 1)$, independent. Then, $P[E_n \text{ i. o. }] = 1$ or 0. Morever, $P[E_n \text{ i. o. }] = 1$ if and only if $\sum_{n\geq 1} P[E_n] = \infty$.
- 37:05 Application. [Durrett, Theorem 1.6.7]. X_n i.i.d., $E[\,|X_1|\,] = \infty \Rightarrow P[\,|X_n| \geq n \text{ i. o. }] = 1$.
- 38:57 We are using here [Chung, Theorem 3.2.1] (cf. Lecture 1, time 0:00)
- 39:39 In particular, $P\big[\lim_n (X_1 + \cdots + X_n)/n \text{ exists and belongs to } \mathbb{R}\big] = 0$.
- 40:36 To prove that the set $\{\omega \in \Omega : \lim_n (X_1+\cdots+X_n)/n \text{ exists and belongs to } \mathbb{R}\}$ is an element of the σ-algebra $\mathcal{F}$, recall that this set corresponds to the set $\{\omega \in \Omega : (X_1 + \cdots + X_n)/n \text{ is a Cauchy sequence } \}$.
- 50:45 Thus, the hypothesis $E[\,|X_1|\,] < \infty$ is needed for a strong law of large numbers for i.i.d random variables.

Further readings.

A. [Chung, Theorem 4.2.5]. Only pairwise independence is needed in [Chung, Theorem 4.2.4]

B. [Durrett, Examples 1.6.2 and 1.6.3] [If R_n represents the number of records up to time n, $R_n/\log n \to 1$ a. e.] and [if L_n represents the size of largest sequence of consecutive 1's in a Bernoulli sequence, $L_n/\log_2 n \to 1$ a. e.]

C. [Breiman, Propositions 3.16-3.18] investigates the number of returns to the origin in a coin-tossing problem.

Recommended exercises.

a. [Chung, Section 4.2] 2, 5, 6, 7, 10, 12, 14, 16,

b. [Durrett, Section 1.6] Exercises $2-8$, 10

c. [Breiman, Section 3.3] Problems 6 [we presented a proof of this result earlier. Use Borel-Cantelli to derive a second proof], 7, 9, 10

Suggested exercises.

a. [Chung, Section 4.2] 1, 3, 4, 8, 9, 11, 13, 15, 18, 19, 20

b. [Durrett, Section 1.6] Exercises 12, 14, 17, 18

Lecture 6: Weak convergence: Helly's selection theorem and tightness

Summary. This lecture is based on [Breiman, Sections 8.1 and 8.2] and [Varadhan, Section 2.3].

Content and Comments.

- 0:00 Definition of weak convergence and convergence in distribution.
- 8:55 The space of distributions $\mathcal{N}$ and the space of generalized distributions $\mathcal{M}$.
- 13:38 [Breiman, Theorem 8.6] Helly's selection Theorem, This result correspond to steps 1, 2, 3 of [Varadhan, Theorem 2.4]
- 49:18 Examples where the limit is a generalized distribution and not a distribution.
- 51:54 Uniqueness o limit points yields convergence. [Breiman, Corollary 8.8]
- 1:00:18 Tightness of probability measures [Breiman, Definition 8.9]
- 1:05:52 A set of distribution funtions $\{F_\alpha : \alpha \in I\}$ is tight if and only if the following statement holds $[F_{\alpha_n} \to G \Rightarrow G \in \mathcal{N}]$. [Breiman, Proposition 8.10]
- 1:21:11 Let me clarify. By hypothesis, $F_{\alpha_n}(n) - F_{\alpha_n}(-n) \leq 1-\epsilon$ for all $n \geq 1$. We introduced a subsequence $\alpha_n^{(1)}$. This means that $\alpha_n^1 = \alpha_{p(n)}$, where $p(n) \geq n$ and $p(n+1) > p(n)$. By definition of the sequence $p(n)$ and the above inequality, $F_{\alpha_n^{(1)}}(p(n)) - F_{\alpha_n^{(1)}}(-p(n)) = F_{\alpha_{p(n)}}(p(n)) - F_{\alpha_{p(n)}}(-p(n)) \leq 1-\epsilon$. Hence, instead of writing $F_{\alpha_n^{(1)}}(n) - F_{\alpha_n^{(1)}}(-n)$, I should have written $F_{\alpha_{p(n)}}(p(n)) - F_{\alpha_{p(n)}}(-p(n))$. This is what I meant when saying that n is the one which corresponds to the subsequence $\alpha_n^{(1)}$.

Further readings.

A. [Chung, Sections 4.3 and 4.4] provides examples and an alternative view on convergence in distribution.

B. [Durrett, Section 2.2] presents many examples of sequences of random variables which converge in distribution.

Recommended exercises.

a. Prove [Chung, Theorems 4.3.1 and 4.3.2]

b. [Breiman, Chapter 8] Problems 1, 2, 4, 5, 6, 7, 9, 10.

c. [Durrett, Section 2.2] Exercise 6

Suggested exercises.

a. [Breiman, Chapter 8] Problems 3, 8, 11.

b. [Chung, Section 4.3] Exercises 3, 8

c. [Durrett, Section 2.2] Exercises 2, 3, 7

Lecture 7: Weak convergence: Helly-Bray's theorem

Summary. This lecture is based on [Breiman, Section 8.3], [Chung, Section 4.4] and [Varadhan, Section 2.3].

Content and Comments.

0:00 Helly-Bray's theorem, [Breiman, Proposition 8.12] presents a stronger version of the first part of the theorem, and [Varadhan, Theorem 2.3].

32:26 Convergence in probability implies convergence in distribution. [Chung, Theorem 4.4.5].

41:18 Convergence in distribution to a Dirac mass implies convergence in probability. [Chung, Section 4.4, Exercise 4].

46:15 $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} y \in \mathbb{R} \Rightarrow X_n + Y_n \xrightarrow{d} X + y$ and $X_n Y_n \xrightarrow{d} Xy$. [Chung, Theorem 4.4.6].

Further readings.

A. [Breiman, Section 8.3] presents a stronger version of Helly-Bray's theorem [continuity is replaced by the hypothesis that the set of discontinuity points has measure 0 for the limiting probability measure.

B. [Chung, Theorem 4.4.1] states that to prove convergence in distribution it is enough to show that $E[f(X_n)] \to E[f(X)]$ for all continuous functions f with compact support. The corollary of [Chung, Theorem 4.4.6] provides necessary and sufficient conditions for convergence in distribution in terms of open and closed sets.

Recommended exercises.

*a. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function which is continuous everywhere, except at a finite number of points, represented by $D = \{x_1, \dots, x_p\}$. Let μ_n be a sequence of probability measures which converges weakly to μ. If $\mu(D) = 0$, then $\int f\, d\mu_n \to \int f\, d\mu$.

b. [Breiman, Section 8.3], problems 13, 14

c. [Chung, Section 4.4], exercises 1, 4, 6, 7, 9, 11

d. [Varadhan, Chapter 2], exercises 9, 10, 11

e. Prove [Breiman, Propositions 8.12, 8.15, 8.17, 8.19]

Remark 0.1. *The assertion of the recommended exercise* [a] *will be used several times in the next lectures.*

Suggested exercises.

a. [Chung, Section 4.4], exercises 2, 3, 8, 10, 12

Lecture 8: Characteristic functions

Summary. This lecture is based on [Breiman, Section 8.7]

Content and Comments.

- 0:00 Definition of a characteristic function.
- 2:58 Elementary properties of characteristic functions. [Breiman, Proposition 8.27]. We also prove that the characteristic function is positive-definite.
- 13:25 Let μ and ν be two probability measures on $\mathbb{R}$ such that $\varphi_\mu = \varphi_\nu$. Then, $\mu = \nu$. [Breiman, Theorem 8.24]. The theorem on approximation by trigonometric polynomials is [Rudin, Theorem 8.15].
- 44:28 There exists $0 < K < \infty$ such that for all $a > 0$ and probability measures μ on $\mathbb{R}$,

$$\mu\Big(\Big[-\frac{1}{a}\,,\, \frac{1}{a} \Big]^c \Big) \;\le\; \frac{K}{a} \int_0^a \big[\, 1 - \Re\varphi_\mu(t) \,\big]\, dt\;.$$

  This is [Breiman, Proposition 8.29]. $\sin(t)/t < 1$ because $\sin(t) = \int_0^t \cos(r)\, dr$.
- 56:47 Let $(X_n : n \ge 1)$ be a sequence of random variables. Denote by $\varphi_n(t)$ the associated characteristic functions. Assume that there exists $\delta > 0$ and $\varphi : [-\delta, \delta] \to \mathbb{C}$ such that $\varphi_n(t) \to \varphi(t)$ for all $|t| \le \delta$. Assume that φ is continuous at 0. Then, the sequence is tight. This is part of [Breiman, Theorem 8.28].

Further readings.

A. [Varadhan, Section 2.1] has many interesting comments and examples. It provides the inversion formula: a formula for the distribution function in terms of the characteristic function. This is an alternative way to prove that the characteristic functions identifies the distribution measure.
B. [Chung, Sections 6.1 and 6.2] provide further examples of characteristic functions.

Recommended exercises.

a. [Varadhan, Section 2.1], exercises 2, 3, 4, 5.
b. [Chung, Section 6.1]. Prove properties (iii), (iv) and (v) of characteristic functions, the corollary of Theorem 6.1.4 and Theorem 6.1.5.
c. [Chung, Sections 6.1], exercise 16.

Suggested exercises.

a. [Varadhan, Section 2.1], exercise 1.
b. [Chung, Section 6.1], exercise 11, 12.
c. [Chung, Section 6.2], exercises 3, 7, 9.

Lecture 9: The Lévy continuity theorem

Summary. This lecture is based on [Breiman, Sections 8.6, 8.7, 8.9 and 8.11] and on [Varadhan, Section 3.6].

Content and Comments.

- 0:00 Lévy continuity theorem. [Breiman, Theorem 8.28].
- 7:22 Expansion of the characteristic function. [Breiman, Proposition 8.44].
- 27:30 The characteristic function of the sum of two independent random variables. [Breiman, Proposition 8.33].
- 30:32 Application: The central limit theorem for i.i.d. random variables with finite second moments. This is [Varadhan, Theorem 3.17]. See [Breiman, Theorem 8.20] for another proof of this result. We use here the expansion of $\log(1+z)$ for $z \in \mathbb{C}$. See [Breiman, Proposition 8.46].

Further readings.

- A. Read [Breiman, Sections 8.12] and [Varadhan, Section 2.2]. The starting question is: does there exist two distributions with the same moments or does the convergence of moments entail the convergence in distribution.
- B. Read [Breiman, Sections 8.13]. It is proved there that the Laplace transforms characterize the distribution of positive random variables.

Recommended exercises.

- *a. Let $(X_n : n \geq 1)$ be a sequence of random variables which converges in distribution to X. Denote by $\varphi_n(t)$, $\varphi(t)$ the characteristic functions of X_n, X, respectively. Show that φ_n converges to φ uniformly on bounded intervals. This is [Breiman, Proposition 8.31] and will be used many times below.
- b. Prove [Breiman, Propositions 8.30, 8.33, 8.37]
- c. [Breiman, Chapter 8], exercises 16, 17 (see [Varadhan, Theorem 2.7]), 21
- d. Prove [Varadhan, Theorem 2.6]
- e. [Chung, Section 6.4], exercises 4, 7, 11, 24

Suggested exercises.

- a. [Chung, Section 6.3], exercises 6, 8.
- b. Prove [Chung, Theorem 6.4.6]
- c. [Chung, Section 6.4], exercise 6.

Lecture 10: Weak law of large numbers

Summary. This lecture is based on [Varadhan, Sections 3.2 – 3.4].

Content and Comments.

- 0:00 Weak law of large numbers, first proof based on truncation. [Varadhan, Theorem 3.3].
- 16:40 Weak law of large numbers, second proof based on characteristic functions. [Varadhan, Theorem 3.3]. See [Chung, Lemma of Section 6.4] for the proof that $(1 + z_n/n)^n \to e^z$ if $z_n \to z \in \mathbb{C}$.
- 24:32 Comments on the strong law of large numbers.
- 25:47 Kolmogorov inequality. [Varadhan, Lemma 3.7].
- 44:47 Lévy inequality. [Varadhan, Lemma 3.8].

Further readings.

A. [Chung, Sections 5.1 – 5.3] covers substantially the same material. It provides a proof of the strong law of large numbers under thy hypothesis of a finite second moment.
B. Read the example of [Chung, Section 5.2]

Recommended exercises.

a. [Varadhan, Chapter 3], exercises 5, 6, 7.
b. Prove [Chung, Theorems 5.1.1 – 5.1.3].
c. [Chung, Section 5.1], exercises 1, 2, 8, 9.
d. Prove [Chung, Theorems 5.2.1 – 5.2.3].
e. [Chung, Section 5.2], exercises 2, 5, 6, 9, 10, 13

Lecture 11: Convergence of series

Summary. This lecture is based on [Varadhan, Section 3.4].

Content and Comments.

- 0:00 Theorem: Let $(X_j : j \geq 1)$ be a sequence of independent random variables and $S_n = \sum_{1 \leq j \leq n} X_j$, Then S_n converges in distribution if and only it converges in probability if and only if it converges a.s. [Varadhan, Theorem 3.9].
- 2:22 Lemma 1: A Cauchy sequence in probability converges in probability. This is exercise 3.11 of [Varadhan, Section 3.4].
- 18:39 Lemma 2: For all $\epsilon > 0$, $\lim_{m,n\to\infty} P[\, \max_{m<k\leq n} |X_k - X_m| > \epsilon\,] = 0 \Longrightarrow \exists\, X$ s.t. $X_n \to X$ a. s. This is exercise 3.12 of [Varadhan, Section 3.4].
- 40:56 Proof of the theorem, divided in several claims.
- 46:18 Claim 1: $\varphi_{S_n - S_m}(t) \to 1$ for all $|t| \leq t_0$.
- 52:18 Claim 2: $\varphi_{S_n - S_m}(t) \to 1$ for all $t \in \mathbb{R}$. To prove the exercise, show that $1 - \cos(2\,t) \leq 4\,[\,1 - \cos(t)\,]$ for all $t \in \mathbb{R}$.
- 57:12 Claim 3: $S_n - S_m \to 0$ in probability.
- 1:03:09 Claim 4: There exists a r.v. S such that $S_n \to S$ in probability.
- 1:04:25 Claim 5: $F_S = F$.
- 1:06:54 S_n converges a.s. to S.
- 1:12:38 Actually, $k = m + 1$ and runs from $p + 1$ to q. This is corrected at time [1:14:10].
- 1:14:54 Some steps have been skipped. Here is a complete argument. To apply Lévy's inequality, exactly as stated in the previous lecture, set $Y_i = X_{p+i}$, $N = q - p$, $M = k - p$. Note that M varies from 1 to N and that the bound can be written as

$$P\Big[\, \big|\, Y_M \,+\, \cdots \,+\, Y_N \,\big| \,>\, \frac{\epsilon}{2} \,\Big] \;\leq\; \delta \qquad \text{for } \; 1 \leq M \leq N\,.$$

Lévy's inequality yields that

$$P\Big[\, \max_{1\leq M\leq N} \big|\, Y_1 \,+\, \cdots \,+\, Y_M \,\big| \,>\, \epsilon \,\Big] \;\leq\; \frac{\delta}{1-\delta}\,.$$

Rewriting this in terms of the variables X_i yields that

$$P\Big[\, \max_{1\leq M\leq N} \big|\, X_{p+1} \,+\, \cdots \,+\, X_{p+M} \,\big| \,>\, \epsilon \,\Big] \;\leq\; \frac{\delta}{1-\delta}\,.$$

That is

$$P\Big[\, \max_{p<k\leq q} \big|\, S_k - S_p \,\big| \,>\, \epsilon \,\Big] \;=\; P\Big[\, \max_{1\leq M\leq N} \big|\, S_{p+M} - S_p \,\big| \,>\, \epsilon \,\Big] \;\leq\; \frac{\delta}{1-\delta}\,.$$

.

Recommended exercises.

a. Prove [Chung, Theorems 5.3.2].
b. [Chung, Section 5.3], exercises 1, 2, 3, 6.
c. [Durrett, Section 1.8], exercises 9, 11.

Lecture 12: Kolmogorov's three series theorem

Summary. This lecture is based on [Varadhan, Section 3.4].

Content and Comments.

- 0:00 Kolmogorov's one series theorem, first proof. [Varadhan, Theorem 3.10]
- 4:50 One series theorem, second proof.
- 11:12 Kolmogorov's two series theorem, [Varadhan, Theorem 3.11]
- 15:48 Kolmogorov's three series theorem, direct statement [Varadhan, Theorem 3.12]
- 28:01 Application: the convergence of the random series $\sum_{1\le j\le n}(Z_j/j^\theta)$, where Z_j is a sequence of i.i.d. random variables such that $P[Z_1=1]=1/2=P[Z_1=-1]$.
- 34:10 Kolmogorov's three series theorem, the converse statement [Varadhan, Theorem 3.12]
- 36:28 Lemma: $(Y_j)_{j\ge1}$ sequence of independent variables such that $|Y_j|\le C$, $E[Y_j]=0$, $\sum_{1\le j\le n} Y_j$ converges a.s. Then, $\sum_{j\ge1}\mathrm{Var}\,[Y_j]<\infty$.
- 1:11:00 Kolmogorov's Three series theorem, proof of the converse statement.

Recommended exercises.

a. [Breiman, Chapter 3], problem 11.
b. [Durrett, Section 1.8], exercise 1.
c. [Durrett, Section 1.8], solve example 3.

Lecture 13: The strong law of large numbers

Summary. This lecture is based on [Varadhan, Section 3.5].

Content and Comments.

- 0:00 Kronecker's lemma. [Chung, Section 5.4].
- 12:27 The strong law of large numbers for mean-zero random variables. [Varadhan, Theorem 3.14].
- 32:21 Exercise: Let $(X_j : j \geq 1)$ be a sequence of i.i.d. random variables such that $E[\, X_1^+ \,] = \infty$, $E[\, X_1^- \,] < \infty$. Then, $\lim_N S_N/N = +\infty$ almost surely.
- 33:18 Let $(X_j : j \geq 1)$ be a sequence of i.i.d. random variables such that $E[\,|X_1|\,] = \infty$. Then, $\limsup_N |S_N|/N = +\infty$ almost surely. [Chung, Theorem 5.4.2].
- 42:00 Kolmogorov's 0-1 law. [Varadhan, Theorem 3.15].

Further readings.

- A. [Chung, Section 5.4] states more general results on the almost sure convergence of series
- B. [Durrett, Sections 1.7 and 1.8] also states more general results and presents instructive examples.

Recommended exercises.

- *a. Let $(X_j : j \geq 1)$ be a sequence of i.i.d. random variables such that $E[\,|X_1|\,] < \infty$. Prove that $(X_1 + \cdots + X_N)/N \to E[X_1]$ almost surely.
- *b. Prove [Varadhan, Corollary 3.16] and solve exercises 3.16 and 3.15.
- *c. Fill the gaps left in the proof of Kolmogorov's 0-1 law.
- d. Prove [Chung, Theorem 5.4.1] and its corollary.
- e. Prove [Chung, Theorem 5.4.3] and its corollary.
- f. [Chung, Section 5.4], exercises 1, 5, 7, 9.
- g. Fill the details of Examples 1.8.1 − 1.8.3 in [Durrett]
- h. Prove [Durrett, Theorem 1.8.7 and 1.8.8]
- i. [Durrett, Section 1.8], exercises 4, 5, 6, 9, 10, 11, 12.

Suggested exercises.

- a. Prove [Durrett, Theorem 1.7.3]
- b. [Durrett, Section 1.7], exercises 1, 2, 3, 4.
- c. Prove [Durrett, Theorem 1.7.4]
- d. Fill the details of [Durrett, Example 1.7.3]
- e. [Durrett, Section 1.8], exercises 1, 2, 3, 7, 8.

Lecture 14: Law of large numbers II.

Summary. This lecture is based on [Chung, Sections 5.4 and 5.5].

Content and Comments.

- 0:00 Let $(X_j : j \geq 1)$ be mean zero, independent random variables. Let $(a_n : n \geq 1)$ be an increasing sequence of real numbers diverging to $+\infty$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be an even, non-negative function. Assume that $\psi_1 : (0, \infty) \to \mathbb{R}$, defined by $\psi_1(x) = \varphi(x)/x$, is increasing and $\psi_2 : (0, \infty) \to \mathbb{R}$, defined by $\psi_2(x) = \varphi(x)/x^2$, is decreasing. Then, $\sum_{1 \leq j \leq n}(X_j/a_j)$ converges almost surely if $\sum_{j \geq 1} E[\varphi(X_j)/\varphi(a_j)] < \infty$. This is [Chung, Theorem 5.4.1].
- 24:55 Corollary: Under the hypotheses of the previous theorem, $(X_1 + \cdots + X_N)/a_N \to 0$ almost surely. Kronecker lemma was stated at the beginning of the previous lecture.
- 26:20 Example 1: $\varphi(x) = |x|^p$ for $1 \leq p \leq 2$.
- 29:29 Example 2: Let $(X_j : j \geq 1)$ be mean zero, independent random variables. Assume that there exist $1 < p \leq 2$ and $M < \infty$ such that $E[\,|X_j|^p\,] \leq M$ for all $j \geq 1$. Then, $(X_1 + \cdots + X_N)/N \to 0$ almost surely. Note that if there exist $q > 2$ and $A < \infty$ such that $E[\,|X_j|^q\,] \leq A$ for all $j \geq 1$, then, by Hölder inequality, $E[\,|X_j|^2\,] \leq A^{2/q}$ for all $j \geq 1$. In particular, the thesis holds if there exist $p > 1$ and $M < \infty$ such that $E[\,|X_j|^p\,] \leq M$ for all $j \geq 1$.
- 32:55 Let $(X_j : j \geq 1)$ be mean zero, i.i.d. random variables. Assume that $\sigma^2 = E[X_1^2] < \infty$. Then, for all $\epsilon > 0$, $(X_1 + \cdots + X_N)/\sqrt{N}\,(\log N)^{(1/2)+\epsilon} \to 0$ almost surely.
- 38:57 Let F be a distribution function and $(X_j : j \geq 1)$ be a sequence of i.i.d. random variables whose distribution function is F. Then, the empirical distribution function converges, uniformly in $\mathbb{R}$, to the distribution function almost surely. This is [Chung, Theorem 5.5.1].

Recommended exercises.

a. [Chung, Section 5.4], exercises 2, 3, 4, 6, 10, 12, 13.
b. [Chung, Section 5.5], exercises 1, 2, 3.

Lecture 15: Applications of the Law of large numbers.

Summary. This lecture is based on [Chung, Section and 5.5] and [Durrett-4th, Sections 2.4].

Content and Comments.

0:00 Shannon's entropy. This is [Durrett-4th, Example 2.4.3]
7:41 Renewal process. I followed [Chung, Section and 5.5]. This is also [Durrett-4th, Example 2.4.1].

Lecture 16: Central Limit Theorem

Summary. This lecture is based on [Varadhan, Section 3.6].

Content and Comments.

0:00 Statement of Lindeberg's theorem. [Varadhan, Theorem 3.18].
6:10 Claim 1: $\max_{1\le j\le n}(\sigma_j^2/s_n^2) \longrightarrow 0$.
9:20 Claim 2: For all $\epsilon > 0$, $\lim_{n\to\infty} \sum_{1\le j\le n} P\big[\,|X_{n,j}| \,>\, \epsilon\,\big] \;=\; 0$.
13:17 Claim 3: For all $T > 0$, $\sup_{|t|\le T} \max_{1\le j\le n} \big|\,\varphi_{n,j}(t) \,-\, 1\,\big| \longrightarrow 0$.
20:39 Proof of Lindeberg's theorem, Part 1:

$$\lim_{n\to\infty} \sup_{|t|\le T} \sum_{j=1}^{n} \big|\,\varphi_{n,j}(t) \,-\, 1\,\big|^2 \;=\; 0\;.$$

28:03 Part 1A: $\sup_{|t|\le T} \max_{1\le j\le n} \big|\,\varphi_{n,j}(t) \,-\, 1\,\big| \longrightarrow 0$.
32:04 Part 1B: There exists a constant C_T such that $\sup_{|t|\le T} \sum_{1\le j\le n} \big|\,\varphi_{n,j}(t) - 1\,\big| \;\le\; C_T$.
33:38 Proof of Lindeberg's theorem, Part 2:

$$\lim_{n\to\infty} \sup_{|t|\le T} \Big|\, \sum_{j=1}^{n} \big[\,\varphi_{n,j}(t) \,-\, 1\,\big] \,+\, \frac{t^2}{2}\,\Big| \;=\; 0\;.$$

42:41 Lyapounov's condition implies Lindeberg's.
46:32 The same proof yields the following result. For each $n \ge 1$, let $(X_{n,j} : 1 \le j \le k_n)$ be independent random variables. Assume that $k_n \to \infty$, $E[X_{n,j}] = 0$, $s_n^2 = \sum_{1\le j\le k_n} E[X_{n,j}^2] \;=\; 1$. If, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2\, \mu_{n,j}(dx) \;=\; 0\;.$$

Then, $\sum_{1\le j\le k_n} X_{n,j}$ converges in distribution to a mean-zero Gaussian random variable with variance equal to 1.

Further readings.

A. [Chung, Sections 7.1 and 7.2].
B. [Breiman, Sections 9.1 – 9.3].
C. [Durrett-4th, Section 3.4].
D. In [Durrett-4th, Section 3.4], read example 8.

Recommended exercises.

a. For each $n \ge 1$, let $(X_{n,j} : 1 \le j \le k_n)$ be independent random variables. Assume that $k_n \to \infty$, $E[X_{n,j}] = 0$, $\sum_{1\le j\le k_n} E[X_{n,j}^2] \;=\; 1$. Assume, furthermore, that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2\, \mu_{n,j}(dx) \;=\; 0\;.$$

Then, $\sum_{1\le j\le k_n} X_{n,j}$ converges in distribution to a mean-zero Gaussian random variable with variance equal to 1.
b. Prove [Chung, Theorem 7.1.1].
c. Fill the gaps of all examples in [Durrett-4th, Section 3.4].

d. [Varadhan, Section 3.6], exercises 17, 19.
e. [Chung, Section 7.1], exercise 1.
v. [Durrett-4th, Section 3.4], exercises 5, 6, 7.

Suggested exercises.

a. [Chung, Section 7.1], exercises 2, 4.
b. [Durrett-4th, Section 3.4], exercises 2, 3, 4, 8.

Lecture 17: Central Limit Theorem, II.

Summary. This lecture is based on [Chung, Section 7.2].

Content and Comments.

0:00 Comments on sums of small independent random variables.

11:27 The converse of Lindeberg's theorem. This is part of [Chung, Theorem 7.2.1].

12:14 Lemma:

$$\lim_{n\to\infty} \max_{1\le j\le n} P\big[\,|X_{n,j}| \,>\, \epsilon\,\big] \;=\; 0 \quad \text{for all } \epsilon > 0$$

if and only if

$$\lim_{n\to\infty} \sup_{|t|\le T} \max_{1\le j\le n} \big|\,\varphi_{n,j}(t) \,-\, 1\,\big| \;=\; 0 \quad \text{for all } T > 0.$$

24:00 Proof of the theorem, initial considerations.

30:12 Proof of the theorem, part 1: We have that

$$\lim_{n\to\infty} \sum_{1\le j\le n} \big[\,\varphi_{n,j}(t) \,-\, 1\,\big] \;=\; -\,(t^2/2)\ .$$

39:51 Proof of the theorem, part 2: conclusion.

Further readings.

A. [Chung, Section 7.2]
B. In [Durrett-4th, Section 3.4], read subsection 3.4.3.

Recommended exercises.

*a. [Chung, Section 7.2], exercise 3.
b. [Chung, Section 7.2], exercise 7, 10.
c. [Durrett-4th, Section 3.4], exercise 9

Suggested exercises.

a. [Chung, Section 7.2], exercise 5, 8, 9, 12.
b. [Durrett-4th, Section 3.4], exercises 10, 13

Lecture 18: Infinitely Divisible Laws.

Summary. This lecture is based on [Breiman, Sections 9.4 and 9.5]

Content and Comments.

0:00 Statement of the problem: Let $(X_{n,j} : 1 \le j \le k_n)$, $k_n \to \infty$, be an array of independent random variables. Assume that they are uniformly negligible: For all $\epsilon > 0$,

$$\lim_{n\to\infty} \max_{1\le j\le k_n} P\big[\,|X_{n,j}| > \epsilon\,\big] \;=\; 0\;.$$

Let $S_n = \sum_{1\le j\le k_n} X_{n,j}$. 1. What are the possible limits (in distribution) of S_n? 2. Give necessary conditions on the sequence to guarantee the convergence.

2:41 Poisson convergence. Let $(X_{n,j} : 1 \le j \le n)$ be an array of i.i.d. random variables such that $P[X_{n,1} = 1] = p_n = 1 - P[X_{n,1} = 0]$. Then, they are uniformly negligible if $p_n \to 0$. Moreover, $S_n = \sum_{1\le j\le k_n} X_{n,j}$ converges in distribution to a Poisson law if the sequence $n\,p_n$ converges to some $\lambda \in [0,\infty)$.

9:57 Proposition. S_n converges in distribution if and only if $n\,p_n \to \lambda \in [0,\infty)$. In this case the limit is the Poisson distribution with parameter λ. This is [Breiman, Theorem 9.4]. The index n in sometimes denoted by N.

25:50 The characteristic function of a Poisson distribution. The one of a Poisson distribution with jump size $a > 0$ ($P[X = ak] = (\lambda^k/k!)\,e^{-\lambda}$, $k \ge 0$. The characteristic function of finite sums of independent Poisson distributions with different parameters and jump sizes.

32:10 Definition of infinitely divisible laws.

38:20 Proposition. Let $(X_{n,j} : 1 \le j \le k_n)$ be an array of i.i.d. random variables. Suppose that $k_n \to \infty$ and that S_n converges in distribution to S. Then, S has an infinitely divisible law. This is [Breiman, Proposition 9.9]. Note that the converse statement is trivial. That is, if S has an infinitely divisible law, then there exists an array of i.i.d. random variables $(X_{n,j} : 1 \le j \le n)$ such that $S_n = \sum_{1\le j\le n} X_{n,j}$ converges in distribution to S.

47:57 Examples of IDL: Dirac, Gaussian, Poisson and independent Poisson sums of i.i.d. random variables.

Further readings.

A. [Chung, Section 7.6]
B. [Durrett-4th, Sections 3.6 and 3.8]

Recommended exercises.

a. Prove [Durrett-4th, Theorem 3.6.1].
b. [Durrett-4th, Section 3.6], exercises 1, 2, 3, 4, 5, 6, 7.
c. Prove [Durrett-4th, Lemmata 3.6.2 – 3.6.4].
d. Prove [Durrett-4th, Theorems 3.6.6 and 3.6.7].
e. [Durrett-4th, Section 3.8], exercise 2.
f. Prove [Chung, Theorem 7.6.1]
g. [Chung, Section 7.6], exercises 1, 3.

Lecture 19: Accompanying laws.

Summary. This lecture is based on [Varadhan, Section 3.7].

Content and Comments.

0:00 Formulation of the one-dimensional central limit problem. Recall that $k_n \to \infty$.

2:01 The Poisson transformation of a law into an infinite divisible law.

4:51 The uniform negligibility is equivalent to the convergence to 1 of the characteristic functions, uniformly over bounded intervals.

6:00 Definition of the accompanying laws.

10:18 Statement of the theorem on accompanying laws. This is [Varadhan, Theorem 3.19].

12:33 Lemma 1: The sequence $a_{n,j}$ converges uniformly to 0:

$$\lim_{n\to\infty} \max_{1\le j\le k_n} |\, a_{n,j} \,| \;=\; 0 \;.$$

16:20 Corollary: The sequence $\widetilde{X}_{n,j}$ is uniformly negligible.

20:04 Lemma 2: There exists a finite constant C_0 such that

$$|\, \widetilde{a}_{n,j} \,| \;\le\; C_0 P[\, |\, \widetilde{X}_{n,j} \,| \;\ge\; 1/2 \,] \;.$$

34:41 Recollection of the statement of the theorem and of the results proved so far.

36:52 Lemma 3: Let $B_n = \sum_j a_{n,j}$. Then, $\sum_j X_{n,j} - A_n$ converges in distribution to S if and only if $\sum_j \widetilde{X}_{n,j} + B_n - A_n$ converges in distribution to S. A similar statement holds with Y in place of X. At time 38:17, I write $B_n + A_n$ instead of $B_n - A_n$.

Conclusion: It is enough to prove the theorem with $X_{n,j}$, $Y_{n,j}$ replaced by $\widetilde{X}_{n,j}$, $\widetilde{Y}_{n,j}$, respectively.

Further readings.

A. A different approach to the one-dimensional central limit problem is presented in [Breiman, Section 9.5 – 9.7].

Lecture 20A: Proof of the Accompanying Laws theorem, Part 1.

Summary. This lecture is based on [Varadhan, Section 3.7].

Content and Comments.

0:00 Statement of the first main proposition. This is [Varadhan, Theorem 3.19] with $X_{n,j}$, $Y_{n,j}$ replaced by $\widetilde{X}_{n,j}$, $\widetilde{Y}_{n,j}$, respectively.

1:44 Step 0: It is enough to prove that for all sequences $(A_n : n \geq 1)$, the difference $\varphi_{\sum_j \widetilde{X}_{n,j} - A_N}(t) - \varphi_{\sum_j \widetilde{Y}_{n,j} - A_N}(t) \to 0$ for all $t \in \mathbb{R}$.

3:59 Step 1: It is enough to prove that for all $T > 0$, there exists a finite constant C_T such that $\sum_j |\, \varphi_{n,j}(t) - 1 \,| \leq C_T$ for all $|t| \leq T$.

14:14 Step 2: It is enough to prove that

(a) There exists a finite constant C_0 such that $\sum_j P[\, |\, \widetilde{X}_{n,j} \,| \geq 1\,] \leq C_0$ for all $n \geq 1$,

(b) There exists a finite constant C_0 such that $\sum_j E[\, \widetilde{X}^2_{n,j} \chi_{|\widetilde{X}_{n,j}| \leq 1}\,] \leq C_0$ for all $n \geq 1$.

27:25 Proposition: Assume that $\sum_j \widetilde{Y}_{n,j} - A_n$ converges in distribution to some law, then the previous estimates (a) and (b) hold.

28:08 Step 1: Under the hypothesis of the proposition, for all $T > 0$, there exists a finite constant C_T such that

$$\sum_{j=1}^{k_n} E\Big[\, \{\, 1 - \cos(t\widetilde{X}_{n,j})\,\}\,\Big] \;\leq\; C_T \tag{0.1}$$

for all $|t| \leq T$ and $n \geq 1$. Note that the estimate

$$\sum_{j=1}^{k_n} E\Big[\, \{\, 1 - \cos(t\widetilde{X}_{n,j})\,\}\,\Big] \;\leq\; C_0$$

obtained at time [37:00] holds for all $|t| \leq t_0$ and all n large, say $n \geq n_0$. By changing the value of the constant C_0 this inequality can be extended to $1 \leq n \leq n_0$. This explains why (0.1) is in force for all $n \geq 1$.

38:31 Step 2: Condition (0.1) implies (a) and (b).

48:34 Actually, we proved the following result. Let $(Z_{n,j} : 1 \leq j \leq k_n)$ be an array of independent random variables such that $k_n \to \infty$. Assume that for all $T > 0$, there exists a finite constant C_T such that

$$\sum_{j=1}^{k_n} E\Big[\, \{\, 1 - \cos(tZ_{n,j})\,\}\,\Big] \;\leq\; C_T \tag{0.2}$$

for all $|t| \leq T$ and $n \geq 1$. Then,

(a) There exists a finite constant C_0 such that $\sum_j P[\, |\, Z_{n,j} \,| \geq 1\,] \leq C_0$ for all $n \geq 1$, and

(b) There exists a finite constant C_0 such that $\sum_j E[\, Z^2_{n,j} \chi_{|Z_{n,j}| \leq 1}\,] \leq C_0$ for all $n \geq 1$.

Recommended exercise.

(*a) Assume that for all $T > 0$, there exists a finite constant C_T such that

$$\sum_{j=1}^{k_n} E\Big[\,\{\,1 - \cos(t\widetilde{X}_{n,j})\,\}\,\Big] \;\leq\; C_T$$

for all $|t| \leq T$ and $n \geq 1$. Show that for all $\delta > 0$, there exists a finite constant C_0 such that $\sum_j P[\,|\,\widetilde{X}_{n,j}\,| \,\geq \delta\,] \leq C_\delta$ for all $n \geq 1$.

Lecture 20B: Proof of the Accompanying Laws theorem, Part 2.

Summary. This lecture is based on [Varadhan, Section 3.7].

Content and Comments.

- 0:00 Statement of the second main result: Assume that $\sum_j \widetilde{X}_{n,j} - A_n$ converges in distribution to some law, then conditions (a) and (b) of the first part of the lecture are in force.
- 2:23 Remarks on symmetric random variables. $\varphi_{X-X'}(t) = |\,\varphi_X(t)\,|^2$.
- 4:45 Remark 2: Let $(Z_n : n \geq 1)$ be a sequence of random variables which converges in distribution to some random variable S, and let $(Z'_n : n \geq 1)$ be an independent copy of $(Z_n : n \geq 1)$. Then, $Z_n - Z'_n$ converges in distribution to $S - S'$, where S' is an independent copy of S.
- 7:09 Remark 3: Let $U_{n,j} = \widetilde{X}_{n,j} - \widetilde{X}'_{n,j}$, where $(\widetilde{X}'_{n,j} : 1 \leq j \leq k_n)$ is an independent copy of $(\widetilde{X}_{n,j} : 1 \leq j \leq k_n)$. The sequence $\sum_j U_{n,j}$ converges in distribution.
- 11:35 Strategy of the proof,
- 12:32 Step 1: The conditions (a) and (b) hold for the sequence $U_{n,j}$. In view of [48:34] of the previous lecture, it is enough to prove (0.2) for $U_{n,j}$.
- 27:30 Step 2(a): Condition (a) holds for $\widetilde{X}_{n,j}$.
- 34:02 Step 2(b): Condition (b) holds for $\widetilde{X}_{n,j}$.

Claim: Condition (0.1) implies that for every $m > 0$ there exists a finite constant C_m such that $\sum_j E[\,\widetilde{X}^2_{n,j}\chi_{|\widetilde{X}_{n,j}|\leq m}\,] \leq C_m$. Note that 1 has been replaced by m in the indicator function.

Proof. Since $1 - \cos(x) \geq 0$ for all $x \in \mathbb{R}$, it follows from (0.1) that

$$\sum_j E\Big[\,\{\,1 - \cos(t\widetilde{X}_{n,j})\,\}\chi_{|t\widetilde{X}_{n,j}|\leq \pi/4}\,\Big] \;\leq\; C_T$$

for all $|t| \leq T$. Since there exists $a > 0$ such that $1 - \cos(x) \geq ax^2$ for $|x| \leq \pi/4$,

$$a\,t^2 \sum_j E\Big[\widetilde{X}^2_{n,j}\,\chi_{|t\widetilde{X}_{n,j}|\leq \pi/4}\Big] \;\leq\; C_T$$

for $|t| \leq T$. Choosing $t = T = \pi/4m$ yields that

$$a\left(\frac{\pi}{4m}\right)^2 \sum_j E\Big[\widetilde{X}^2_{n,j}\,\chi_{|\widetilde{X}_{n,j}|\leq m}\Big] \;\leq\; C_{\pi/4m}\;,$$

as claimed $\square$

Lecture 21: The Lévy–Khintchine theorem.

Summary. This lecture is based on [Varadhan, Section 3.8].

Content and Comments.

0:00 Let $(X_{n,j} : 1 \le j \le k_n)$ be an array of independent random variables. According to the previous lecture, in order to examine the convergence in distribution of the sequence of random variables $\sum_j X_{n,j} - A_n$, we have to consider the same problem for the sequence $\sum_j \widetilde{Y}_{n,j} - A'_n$. That is, to examine the convergence of the sequence of characteristic functions $\exp\{\int [e^{itx} - 1]\, M_n(dx) - i a_n t\}$, where $M_n = \sum_j \widetilde{\mu}_{n,j}$, $\widetilde{\mu}_{n,j}$ being the distribution measure of the random variable $\widetilde{X}_{n,j}$.

6:22 The measure M_N is not a probability measure. Definition of Lévy measure. The measures M_n are Lévy measures. Introduction of the function $\theta(x)$.

19:50 Theorem: For every Lévy measure M, $b \in \mathbb{R}$, $\sigma^2 \ge 0$, the function

$$\varphi_{M,b,\sigma^2}(t) \;:=\; \exp\Big\{ \int \big[\, e^{itx} - 1 - it\theta(x) \,\big] M(dx) \,+\, itb \,-\, \frac{\sigma^2 t^2}{2} \Big\} \qquad (0.3)$$

is the characteristic function of an infinite divisible law. This is [Varadhan, Theorem 3.20].

23:45 Step 1: For all $\delta > 0$, the claim holds for

$$\exp\Big\{ \int_{|x|\ge\delta} \big[\, e^{itx} - 1 - it\theta(x) \,\big] M(dx) \,+\, itb \Big\}$$

33:24 Step 2: The claim holds for

$$\exp\Big\{ \int \big[\, e^{itx} - 1 - it\theta(x) \,\big] M(dx) \,+\, itb \Big\}$$

42:52 Proof of the theorem.

46:23 Theorem: (uniqueness of the representation). Assume that $\varphi_{M_1,b_1,\sigma_1^2}(t) = \varphi_{M_2,b_2,\sigma_2^2}(t)$ for all $t \in \mathbb{R}$ and that $M_1(\{0\}) = M_2(\{0\}) = 0$. Then, $M_1 = M_2$, $b_1 = b_2$, $\sigma_1^2 = \sigma_2^2$. In the statement of the theorem, I forgot to add the hypothesis $M_1(\{0\}) = M_2(\{0\}) = 0$. We can always assume that $M(\{0\}) = 0$ for a Lévy measure M because we only integrate continuous functions which vanish at the origin..

48:18 Step 1: $\sigma_1^2 = \sigma_2^2$.

51:07 Step 2: $M_1 = M_2$.

59:24 Proof of the exercise left in the lecture. We showed that the bounded Borel measures $M_{1,s}(dx) = [1 - \cos(sx)]\, M_1(dx)$ and $M_{2,s}(dx) = [1 - \cos(sx)]\, M_2(dx)$ are equal for all $s \in \mathbb{R}$. To prove that $M_1 = M_2$, it is enough to show that $\int F dM_1 = \int F dM_2$ for all continuous function with compact support in $\mathbb{R} \setminus \{0\}$. Fix $0 < a < b < \infty$ and a continuous function F with support contained in $[a, b]$. Choose δ small so that $\delta b < (2\pi - \delta)a$. Choose s so that $(\delta/a) < s < (2\pi - \delta)/b$. It follows from this choice that $sx \in [\delta, 2\pi - \delta]$ for all $x \in [a, b]$. In particular, there exists $c_\delta > 0$ such that $1 - \cos(sx) \ge c_\delta$ for all $x \in [a, b]$. Thus, $F(x)/[1 - \cos(sx)]$ is a continuous function with support contained in $[a, b]$. Hence, as $M_{1,s} = M_{2,s}$, $\int F\, dM_1 = \int F(x)/[1 - \cos(sx)]\, M_{1,s}(dx) = \int F(x)/[1 - \cos(sx)]\, M_{2,s}(dx) = \int F\, dM_2$, as claimed.

Recommended exercises.

a. [Varadhan, Section 3.8], exercises 21, 22, 23, 24, 25.

Lecture 22A: The one-dimensional central limit problem, part 1.

Summary. This lecture is based on [Varadhan, Section 3.8].

Remark 0.1 of Lecture 7 is used many times in this lecture.

Content and Comments.

0:00 Theorem: The sequence of random variables X_{M_n,σ_n^2,a_n} converges to a random variable X if and only if there exists a Lévy measure M, $\sigma^2 \geq 0$ and $a \in \mathbb{R}$ such that $X = X_{M,\sigma^2,a}$ and

(i) For all $f \in C_b(\mathbb{R})$ for which there exists $\delta > 0$ such that $f(x) = 0$ for all $|x| \leq \delta$, we have that $\int f \, dM_n \to \int f \, dM$.

(ii) There exists $x_0 > 0$ such that $M(\{-x_0, x_0\}) = 0$ and

$$\lim_{n\to\infty} \left\{ \int_{-x_0}^{x_0} x^2 \, M_n(dx) \; + \; \sigma_n^2 \right\} \; = \; \int_{-x_0}^{x_0} x^2 \, M(dx) \; + \; \sigma^2 \, . \tag{0.4}$$

(iii) $\lim_{n\to\infty} a_n \to a$.

This is [Varadhan, Theorem 3.21].

12:17 If there exists $x_0 > 0$ such that $M(\{x_0\} \cup \{-x_0\}) = 0$ and (0.4) holds, then for all $x_1 > 0$ such that $M(\{x_1\} \cup \{-x_1\}) = 0$ we have that

$$\lim_{n\to\infty} \left\{ \int_{-x_1}^{x_1} x^2 \, M_n(dx) \; + \; \sigma_n^2 \right\} \; = \; \int_{-x_1}^{x_1} x^2 \, M(dx) \; + \; \sigma^2 \, .$$

20:25 Proof that the conditions are sufficient.

38:27 Comments on the proof

Lecture 22B: The one-dimensional central limit problem, part 2.

Summary. This lecture is based on [Varadhan, Section 3.8].

Content and Comments.

0:00 Statement of the theorem.

1:53 Strategy of the proof.

3:30 Step A1: $\lim_{t_0\to 0}\sup_n\{\int[1 - \cos(tx)]\,dM_n + \sigma_n^2\,t^2/2\} = 0$.

12:04 Step A2: For all $T > 0$, there exists a finite constant C_T such that

$$\int\{1-\cos(tx)\}\,dM_n \le C_T$$

for all $|t| \le T$ and $n \ge 1$. This has been proved in Lecture 20A, see equation (0.1). It follows from this bound (see Lecture 20A, time [38:31] and exercise (*a)) that there exists a finite constant C_0 such that

$$\text{(a)} \quad \int_{|x|\ge\delta} M_n(dx) \le C_0 \quad \text{and} \quad \text{(b)} \quad \int_{|x|\le 1} x^2\,dM_n \le C_0$$

for all $n \ge 1$,

16:04 Step A3: For all $\epsilon > 0$, there exists $A_\epsilon > 0$ such that $M_n([-A, A]^c) \le \epsilon$ for all $n \ge 1$, $A \ge A_\epsilon$. I refer here to the result proved at time [44:28] in Lecture 8.

23:01 Step A4: Let $\omega(x) = x^2/(1 + x^2)$. Then, there exists $C_0 < \infty$ such that $\alpha_n = \int\omega(x)\,M_n(dx) \le C_0$ for all $n \ge 1$.

26:42 Step B: Claim: Given a subsequence $(n_k : k \ge 1)$, there exist a sub-subsequence $(n_{k_j} : j \ge 1)$ and a Lévy measure M such that

$$\int f(x)\,M_{n_{k_j}}(dx) \to \int f(x)\,M(dx) \quad \text{and} \quad \int_{-x_0}^{x_0} x^2\,M_{n_{k_j}}(dx) \to \int_{-x_0}^{x_0} x^2\,M(dx)$$

for all $f \in C_b^*(\mathbb{R})$ and $x_0 > 0$ such that $M(\{-x_0, x_0\}) = 0$. Here, $C_b^*(\mathbb{R})$ is the set of bounded, continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which there exists $\delta > 0$ such that $f(x) = 0$ for all $|x| \le \delta$.

29:10 Step B1: Assume that $\alpha_n \to 0$. Then, the claim formulated in Step B holds and $M = 0$.

37:00 Step B2: Assume that $\alpha_n \to \alpha > 0$. Then, the claim formulated in Step B holds.

52:57 Step C: Claim: Given a subsequence $(n_k : k \ge 1)$, there exist a sub-subsequence $(n_{k_j} : j \ge 1)$ and $\sigma^2 \ge 0$ such that $\sigma^2_{n_{k_j}} \to \sigma^2$. In particular, given a subsequence $(n_k : k \ge 1)$, there exist a sub-subsequence $(n_{k_j} : j \ge 1)$ and a triple $(M, \sigma^2, 0)$ such that $(M_{n_{k_j}}, \sigma^2_{n_{k_j}}, 0) \to (M, \sigma^2, 0)$.

56:56 Step D: Lemma: Assume that $X_n \to X$ and that $X_n + a_n \to Y$ in distribution. Then $a_n \to a$ for some a and $Y = X + a$ in distribution.

1:01:32 Step E: The sequence $a_{n_{k_j}}$ converges to some $a \in \mathbb{R}$.

1:03:58 Conclusion of the tightness part of the proof: Given a subsequence $(n_k : k \ge 1)$, there exist a sub-subsequence $(n_{k_j} : j \ge 1)$ and a triple (M, σ^2, a) such that $(M_{n_{k_j}}, \sigma^2_{n_{k_j}}, a_{n_{k_j}}) \to (M, \sigma^2, a)$.

1:05:15 Uniqueness of limits.

Further readings.

A. The introduction of the measures $d\nu_n = [x^2/(1+x^2)]\, dM_n$ in the previous proof is taken from [Breiman, Theorem 9.17].

Lecture 23: Applications.

Summary. This lecture is based on [Varadhan, Section 3.8].

Content and Comments.

0 A reminder of the result proved in the previous lectures. The convergence of the sum $\sum_j X_{n,j} - A_n$ is reduced to the convergence of a triple $(\widetilde{M}_n, 0, b_n)$.

0 Poisson convergence: Assume that $(X_{n,j} : 1 \le j \le k_n)$ is an array of independent random variables such that $P[X_{n,j} = 1] = p_{n,j} = 1 - P[X_{n,j} = 0]$. Suppose, further that $\max_{1\le j\le k_n} p_{n,j} \to 0$. Then, the sequence $\sum_j X_{n,j} - A_n$ converges in distribution to a random variable S if and only if $p_n = \sum_j p_{n,j} \to p$ and $A_n \to A$. Moreover, $S = N - A$, where N is a Poisson distribution of parameter p and $A \in \mathbb{R}$ the limit of A_n. Note that when $p = 0$, N is the degenerate Poisson distribution, that is, $P[N = 0] = 1$.

0 Let $\varphi_{M,\sigma^2,b}(t)$ be the characteristic function associated to a triple (M, σ^2, b). Then, the distribution is Gaussian if and only if $M = 0$.

0 Let $(X_{n,j} : 1 \le j \le k_n)$ be an array of independent, uniformly negligible random variables. Assume that $\sum_j X_{n,j}$ converges in distribution to a random variable S. Then, S is Gaussian if and only if for all $\delta > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} P[\, |X_{n,j}| > \delta\,] \;=\; 0 \tag{0.5}$$

0 Under the previous hypotheses, the mean μ and variances σ^2 of S are given by

$$\sigma^2 \;=\; \lim_{n\to\infty} \sum_{j=1}^{k_n} E\Big[\, (X_{n,j} - a_{n,j})^2 \, \chi_{|X_{n,j}-a_{n,j}|\le x_0}\,\Big] \tag{0.6}$$

$$\mu \;=\; \lim_{n\to\infty} \sum_{j=1}^{k_n} \Big\{ a_{n,j} \;+\; E\big[\,\theta(X_{n,j} - a_{n,j})\,\big] \Big\}$$

Note that in the first line the limit does not depend on x_0 because $\widetilde{M}_n \to 0$.

0 Assume, further, that $E[X_{n,j}] = 0$, $\sigma^2_{n,j} = E[X^2_{n,j}] < \infty$, $\sigma^2_n = \sum_j \sigma^2_{n,j} \to \sigma^2$ and $S \sim N(0, \sigma^2)$. Then, Lindeberg's condition holds. This (re)proves the assertion that Lindeberg's condition are not only sufficient for convergence to a Gaussian random variable, but also necessary.

0 Claim 1: $\sum_j a_{n,j} \to 0$. This claim is not necessary for the argument. For this reason I skipped its proof in the lecture. It is presented below.

0 Claim 2: $\sum_j a^2_{n,j} \to 0$.

0 Conclusion: Lindeberg's condition holds. For all $\delta > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} E\big[\, X^2_{n,j} \, \chi_{|X_{n,j}|>\delta}\,\big] \;=\; 0$$

Proof that $\sum_j a_{n,j} \to 0$. Since $\mu = 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \Big\{ a_{n,j} \;+\; E\big[\,\theta(X_{n,j} - a_{n,j})\,\big] \Big\} \;=\; 0$$

Fix $0 < \delta < 1/2$. By (0.5) and since θ is a bounded function, we may introduce the indicator of $|X_{n,j}| \le \delta$ inside the expectation to get that

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \Big\{ a_{n,j} \;+\; E\big[\, \theta(X_{n,j} - a_{n,j})\, \chi_{|X_{n,j}|\le\delta} \,\big] \Big\} \;=\; 0 \tag{0.7}$$

We claim that there exists a finite constant C_0 such that

$$\limsup_{n\to\infty} \sum_{j=1}^{k_n} E\Big[\,\big|\, \theta(X_{n,j} - a_{n,j}) \;-\; (X_{n,j} - a_{n,j}) \,\big|\, \chi_{|X_{n,j}|\le\delta} \Big] \;\le\; C_0\, \delta$$

By definition of θ, there exists a finite constant C_0 such that $|\, \theta(x) - x \,| \le C_0 |x|^3$ for $|x| \le 1$. Hence, as $\max_j a_{n,j} \to 0$ and $\delta < 1/2$, the previous sum is bounded by

$$C_0 \sum_{j=1}^{k_n} E\Big[\,\big|\, X_{n,j} - a_{n,j} \,\big|^3 \chi_{|X_{n,j}|\le\delta} \Big] \;\le\; C_0\, \delta \sum_{j=1}^{k_n} E\Big[\,\big|\, X_{n,j} - a_{n,j} \,\big|^2 \chi_{|X_{n,j}|\le\delta} \Big] \;,$$

where the value of the constant C_0 may change from line to line. The last sum is bounded by the sum appearing in (0.6) provided we choose x_0 sufficiently large. This proves the claim.

It follows from this claim and (0.7) that there exists a finite constant C_0 such that

$$\limsup_{n\to\infty} \Big| \sum_{j=1}^{k_n} \Big\{ a_{n,j} \;+\; E\big[\, (X_{n,j} - a_{n,j})\, \chi_{|X_{n,j}|\le\delta} \,\big] \Big\} \Big| \;\le\; C_0\, \delta \;.$$

Note that the function θ disappeared. Since $\max_j a_{n,j}$ is bounded, (0.5) and the previous equation yield that

$$\limsup_{n\to\infty} \Big| \sum_{j=1}^{k_n} E\big[\, X_{n,j}\, \chi_{|X_{n,j}|\le\delta} \,\big] \Big| \;\le\; C_0\, \delta \;.$$

By (0.5) again,

$$\limsup_{n\to\infty} \Big| \sum_{j=1}^{k_n} E\big[\, X_{n,j}\, \chi_{\delta<|X_{n,j}|\le 1} \,\big] \Big| \;=\; 0 \;.$$

Hence,

$$\limsup_{n\to\infty} \Big| \sum_{j=1}^{k_n} E\big[\, X_{n,j}\, \chi_{|X_{n,j}|\le 1} \,\big] \Big| \;\le\; C_0\, \delta \;.$$

To complete the proof, it remains to recall that $a_{n,j} = E[\, X_{n,j}\, \chi_{|X_{n,j}|\le 1} \,]$ and to let $\delta \to 0$.

Lecture 24: Conditional expectation.

Summary. This lecture is based on [Chung, Section 9.1].

Content and Comments.

0:0 Let $(\Omega, \mathcal{F}, P)$ be a probability space, fixed throughout this lecture. Definition of $P[\,B\,|\,A\,]$ and of $P_A[\,\cdot\,]$.

2:10 Definition of $E[\,f\,|\,A\,]$.

6:02 Example with a deck of cards

12:38 Let $(A_n : n \geq 1)$ be a partition of Ω such that $A_n \in \mathcal{F}$, $P[A_n] > 0$. Then,

$$P[B] \;=\; \sum_{n\geq 1} P[\,B\,|\,A_n\,]\; P[A_n]$$

15:12 For all f integrable,

$$E[f] \;=\; \sum_{n\geq 1} E[\,f\,|\,A_n\,]\; P[A_n]$$

17:11 Let $\mathcal{G} = \sigma(A_j : j \geq 1)$. Exercise: show that B belongs to $\mathcal{G}$ if and only if $B = \cup_{j\in M} A_j$ for some $M \subset \mathbb{N}$.

18:31 A function $h : \Omega \to \mathbb{R}$ is measurable with respect to $\mathcal{G}$ if and only if it is constant on each set A_j. Proof that it is measurable if it is constant.

26:18 Proof that it is constant if it is measurable.

33:44 Let $E[\,f\,|\,\mathcal{G}\,] \;=\; \sum_{n\geq 1} E[\,f\,|\,A_n\,]\,\chi_{A_n}$. Then, $E[\,f\,|\,\mathcal{G}\,]$ is $\mathcal{G}$-measurable.

37:35 For every bounded function h, measurable with respect to $\mathcal{G}$,

$$\int E[\,f\,|\,\mathcal{G}\,]\; h\; dP \;=\; \int_B f\; h\; dP$$

43:27 Step 1: For every set B measurable with respect to $\mathcal{G}$,

$$\int_B E[\,f\,|\,\mathcal{G}\,]\; dP \;=\; \int_B f\; dP$$

48:17 Step 2: Extension to bounded functions, measurable with respect to $\mathcal{G}$.

51:07 Summary of the properties of $E[\,f\,|\,\mathcal{G}\,]$ in the case where $\mathcal{G} = \sigma(A_j : j \geq 1)$.

56:44 Definition of the conditional expectation $E[\,f\,|\,\mathcal{G}\,]$ for an integrable function $f \in \mathcal{F}$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$.

58:30 Existence and uniqueness of the conditional expectation.

1:04:06 Uniquenes of the conditional expectation.

1:06:20 The conditional expectation $E[\,f\,|\,\mathcal{G}\,]$ is integrable.

1:10:56 If f is $\mathcal{G}$-measurable, then $E[\,f\,|\,\mathcal{G}\,] = f$.

1:12:32 Let f be an integrable function. Then, f can be decomposed as $f = f_1 + f_2$, where f_1 is $\mathcal{G}$-measurable and $f_2 \perp \mathcal{G}$ in the sense that $E[\,f\,h\,] = 0$ for all bounded functions h which are $\mathcal{G}$-measurable.

1:17:10 Definition of $E[\,f\,|\,X\,]$ and of $P[\,A\,|\,\mathcal{G}\,]$.

Further readings.

A. The proof of the Radon-Nykodim theorem can be found in [Taylor, Section 6.4] or in my lectures on measure theory.
B. [Varadhan, Section 4.1] provides a short proof of Radon-Nykodim theorem, which relies on the Hahn-Jordan decomposition of a measure.
C. [Durrett-4th, Section 5.1], examples $1 - 6$
D. [Breiman, Section 4.1] proposes a slightly different approach to conditional expectation.

Recommended exercises.

a. Show that $E\big[\,E[\,Z\,|\,\mathcal{G}\,]\,\big] = E[Z]$ if Z is integrable.
b. [Varadhan, Section 4.1], exercises $1 - 8$ recall important facts from measure theory which are used in the proofs of the properties of conditional expectation.
c. [Chung, Section 9.1], exercises $1 - 6$

Suggested exercises.

a. [Breiman, Section 4.1], problems 2, 3, 5, 7, 8.

Lecture 25: Properties of conditional expectation.

Summary. This lecture is based on [Chung, Section 9.1].

Throughout this lecture, $(\Omega, \mathcal{F}, P)$ is a fixed probability space, $\mathcal{G}$ is a σ-algebra contained in $\mathcal{F}$ and X, Y, $(X_n : n \geq 1)$ are integrable random variables which are measurable with respect to $\mathcal{F}$.

Content and Comments.

- 0:0 The conditional expectation is linear: $E[\, X + Y \,|\, \mathcal{G}\,] = E[\, X \,|\, \mathcal{G}\,] + E[\, Y \,|\, \mathcal{G}\,]$
- 5:14 The conditional expectation is monotone: $E[\, X \,|\, \mathcal{G}\,] \leq E[\, Y \,|\, \mathcal{G}\,]$ if $X \leq Y$.
- 10:06 If Y is $\mathcal{G}$-measurable and X, XY are integrable, then $E[\, X\, Y \,|\, \mathcal{G}\,] = Y\, E[\, X \,|\, \mathcal{G}\,]$
- 19:54 $\big| E[\, X \,|\, \mathcal{G}\,] \big| \leq E[\, |X| \,|\, \mathcal{G}\,]$.
- 22:50 Conditional monotone convergence theorem
- 29:40 Conditional Fatou's lemma
- 34:37 Conditional dominated convergence theorem
- 44:07 Jensen's conditional inequality
- 1:01:46 Schwarz' conditional inequality. The argument could be slightly simpler to prove that the integral of $X^2/E[X^2 \,|\, \mathcal{G}\,]$ is equal to 1. We have that

$$E\Big[\frac{X^2}{E[\, X^2 \,|\, \mathcal{G}\,]}\, \chi_{\epsilon \leq E[\, X^2 \,|\, \mathcal{G}\,] \leq \epsilon^{-1}} \,\Big|\, \mathcal{G} \Big] \;=\; \frac{1}{E[\, X^2 \,|\, \mathcal{G}\,]}\, \chi_{\epsilon \leq E[\, X^2 \,|\, \mathcal{G}\,] \leq \epsilon^{-1}} E\big[\, X^2 \,\big|\, \mathcal{G}\,\big]$$

because X^2 and $(X^2/E[\, X^2 \,|\, \mathcal{G}\,])\, \chi_{\epsilon \leq E[\, X^2 \,|\, \mathcal{G}\,] \leq \epsilon^{-1}}$ are integrable. Taking expectation on both sides and simplifying the right-hand side yields that

$$\int \frac{X^2}{E[\, X^2 \,|\, \mathcal{G}\,]}\, \chi_{\epsilon \leq E[\, X^2 \,|\, \mathcal{G}\,] \leq \epsilon^{-1}}\, dP \;=\; \int \chi_{\epsilon \leq E[\, X^2 \,|\, \mathcal{G}\,] \leq \epsilon^{-1}}\, dP$$

because $E\big[\, E[\, Z \,|\, \mathcal{G}\,] \big] = E[Z]$ if Z is integrable. It remains to let $\epsilon \to 0$ and invoke the monotone convergence theorem.

- 1:15:26 If $\mathcal{G}_1 \subset \mathcal{G}_2$, then

$$E\Big[\, E[\, X \,|\, \mathcal{G}_2\,] \,\Big|\, \mathcal{G}_1 \Big] \;=\; E[\, X \,|\, \mathcal{G}_1\,] \;=\; E\Big[\, E[\, X \,|\, \mathcal{G}_1\,] \,\Big|\, \mathcal{G}_2 \Big]$$

Further readings.

A. [Varadhan, Section 4.2], specially Remark 4.5.
B. [Breiman, Section 4.2]
C. [Durrett-4th, Section 5-1]

Exercises a, b, c and d below are strongly recommended.

Recommended exercises.

*a. [Varadhan, Section 4.2], exercise 9
*b. Prove [Durrett-4th, Section 5-1], Theorems 4 and 8
*c. [Durrett-4th, Section 5-1], exercises 8, 9
*d. Prove [Breiman, Section 4.2] Proposition 4.20.(4)
e. [Chung, Section 9.1], exercises 7, 8, 9, 12
f. [Durrett-4th, Section 5-1], exercises 3, 4, 6

Suggested exercises.

a. [Chung, Section 9.1], exercises 10, 11, 13
b. [Durrett-4th, Section 5-1], exercises 7, 10, 11, 12
c. [Breiman, Section 4.2], problems 12, 14, 15

Lecture 26: Regular conditional probability.

Summary. This lecture is based on [Varadhan, Section 4.3].

Content and Comments.

Throughout this section, $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra.

0:00 Definition of the conditional probability $\mu(\omega, A) := P[A \,|\, \mathcal{G}\,]$, for $A \in \mathcal{F}$.

3:30 Properties of the conditional probability $\mu(\,\cdot\,, A)$, $A \in \mathcal{F}$.

9:03 Definition of a regular conditional probability (RCP) given a σ-algebra.

13:54 Let $\mu(\omega, A)$ be a RCP on $(\Omega, \mathcal{F}, P)$ given $\mathcal{G}$. Then, for all integrable function $f : \Omega \to \mathbb{R}$ ($\mathcal{F}$-measurable), $\omega \mapsto \int f \, d\,\mu_\omega$ is a version of the conditional expectation $E[\, f \,|\, \mathcal{G}\,]$.

20:30 Theorem: Let P be a probability measure on $([0,1], \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra, and let $\mathcal{G} \subset \mathcal{B}$. Then, there exists a RCP given $\mathcal{G}$. This is [Varadhan, Theorem 4.7].

29:18 Step 1: For $x \in \mathbb{Q}$, define $F_x : \Omega \to [0,1]$ as $F_x = P[\,(-\infty, x] \,|\, \mathcal{G}\,]$. Main properties of F_x.

37:46 Step 2: On a set of full measure N^c, define $G_x = \inf\{\, F_y : y > x \,,\, y \in \mathbb{Q}\,\}$. For $\omega \in N^c$, $y \mapsto G_y(\omega)$ is a distribution function.

44:53 Step 3: For $\omega \in N^c$, let $Q(\omega, \cdot)$ be the measure on $([0,1], \mathcal{B})$ associated to the distribution function $G_\cdot(\omega)$. For $\omega \in N$, define $Q(\omega, \cdot)$ as the Lebesgue measure on $([0,1], \mathcal{B})$. For all $A \in \mathcal{B}$, $\omega \mapsto Q(\omega, A)$ is $\mathcal{G}$-measurable.

54:04 Step 4: For all $A \in \mathcal{B}$, $B \in \mathcal{G}$, $\int_B Q(\,\cdot\,, A)\, dP = P[A \cap B]$. This concludes the proof. Keep in mind that the definition of Q is different on N, but this set has measure 0 and is measurable with respect to $\mathcal{G}$. We may therefore replace B by $B \cap N^c$ in the previous argument.

Further readings.

A. [Breiman, Section 4.3] defines regular conditional distributions (RCD) and proves the existence of regular conditional distributions for random vectors taking values in Borel spaces. In the notes of this chapter, there is an example of space where a RCP does not exist.

C. [Durrett-4th, Section 5.1.3] proves the existence of RCD's on Borel spaces. Exercise 15 illustrates the interest of the existence of RDP's.

Suggested exercises.

a. [Varadhan, Section 4.3], exercise 11

b. [Durrett-4th, Section 5-1], exercises 13, 15.

c. [Breiman, Section 4.2], problem 17

References

[Breiman] L. Breiman, *Probability.* Corrected reprint of the 1968 original. Classics in Applied Mathematics, 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. xiv+421 pp. ISBN: 0-89871-296-3

[Chung] K. L. Chung, *A course in Probability Theory.* Third edition. Academic Press, Inc., San Diego, CA, 2001. xviii+419 pp. ISBN: 0-12-174151-6

[Dembo-Zeitouni] A. Dembo, O. Zeitouni, *Large deviations techniques and applications. Corrected reprint of the second (1998) edition.* Stochastic Modelling and Applied Probability, 38. Springer-Verlag, Berlin, 2010. ISBN: 978-3-642-03310-0

[Deuschel-Stroock] J-D. Deuschel, D. W. Stroock, *Large deviations.* Pure and Applied Mathematics, 137. Academic Press, Inc., Boston, MA, 1989. ISBN: 0-12-213150-9

[Durrett] R. Durrett, *Probability: Theory and Examples.* Second edition. Duxbury Press, Belmont, CA, 1996. xiii+503 pp. ISBN: 0-534-24318-5

[Durrett-4th] R. Durrett, *Probability: Theory and Examples.* Fourth edition. Cambridge Series in Statistical and Probabilistic Mathematics, 31. Cambridge University Press, Cambridge, 2010. ISBN: 978-0-521-76539-8

[Rudin] W. Rudin, *Principles of Mathematical Analysis.* Third edition. McGraw-Hill, Inc. New York (1964).

[Taylor] S. J. Taylor, *Introduction to Measure and Integration*, Cambridge University Press, 1973. ISBN 978-0-521-09804-5

[Varadhan-LD] *Large deviations and applications.* CBMS-NSF Regional Conference Series in Applied Mathematics, 46. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984. ISBN: 0-89871-189-4

[Varadhan] S. R. S. Varadhan, *Probability Theory.* Courant Lecture Notes in Mathematics, 7. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001. ISBN: 0-8218-2852-5