Version of 26.8.13

Chapter 27

Probability theory

Lebesgue created his theory of integration in response to a number of problems in real analysis, and all his life seems to have thought of it as a tool for use in geometry and calculus (LEBESGUE 72, vols. 1 and 2). Remarkably, it turned out, when suitably adapted, to provide a solid foundation for probability theory. The development of this approach is generally associated with the name of Kolmogorov. It has so come to dominate modern abstract probability theory that many authors ignore all other methods. I do not propose to commit myself to any view on whether σ-additive measures are the only way to give a rigorous foundation to probability theory, or whether they are adequate to deal with all probabilistic ideas; there are some serious philosophical questions here, since probability theory, at least in its applied aspects, seeks to help us to understand the material world outside mathematics. But from my position as a measure theorist, it is incontrovertible that probability theory is among the central applications of the concepts and theorems of measure theory, and is one of the most vital sources of new ideas; and that every measure theorist must be alert to the intuitions which probabilistic methods can provide.

I have written the preceding paragraph in terms suggesting that 'probability theory' is somehow distinguishable from the rest of measure theory; this is another point on which I should prefer not to put forward any opinion as definitive. But undoubtedly there is a distinction, rather deeper than the elementary point that probability deals (almost) exclusively with spaces of measure 1. M.Loève argues persuasively (LOÈVE 77, §10.2) that the essence of probability theory is the artificial nature of the probability spaces themselves. In measure theory, when we wish to integrate a function, we usually feel that we have a proper function with a domain and values. In probability theory, when we take the expectation of a random variable, the variable is an 'observable' or 'the result of an experiment'; we are generally uncertain, or ignorant, or indifferent concerning the factors underlying the variable. Let me give an example from the theorems below. In the proof of the Central Limit Theorem (274F), I find that I need an auxiliary list $Z_0,\dots,Z_n$ of random variables, independent of each other and of the original sequence $X_0,\dots,X_n$. I create such a sequence by taking a product space $\Omega\times\Omega'$, and writing $X'_i(\omega,\omega')=X_i(\omega)$, while the Z_i are functions of ω'. Now the difference between the X_i and the X'_i is of a type which a well-trained analyst would ordinarily take seriously. We do not think that the function $x\mapsto x^2:[0,1]\to[0,1]$ is the same thing as the function $(x_1,x_2)\mapsto x_1^2:[0,1]^2\to[0,1]$. But a probabilist is likely to feel that it is positively pedantic to start writing X'_i instead of X_i. He did not believe in the space Ω in the first place, and if it turns out to be inadequate for his intuition he enlarges it without a qualm. Loève calls probability spaces 'fictions', 'inventions of the imagination' in Larousse's words; they are necessary in the models Kolmogorov has taught us to use, but we have a vast amount of freedom in choosing them, and in their essence they are nothing so definite as a set with points.

A probability space, therefore, is somehow a more shadowy entity in probability theory than it is in measure theory. The important objects in probability theory are random variables and distributions, particularly joint distributions. In this volume I shall deal exclusively with random variables which can be thought of as taking values in some power of $\mathbb{R}$; but this is not the central point. What is vital is that somehow the *codomain*, the potential set of values, of a random variable, is much better defined than its *domain*. Consequently our attention is focused not on any features of the artificial space which it is convenient to use as the underlying probability space – I write 'underlying', though it is the most superficial and easily changed aspect of the model – but on the distribution on the codomain induced by the random variable. Thus the Central Limit Theorem, which speaks only of distributions, is actually more important in applied probability than the Strong Law of Large Numbers, which claims to tell us what a long-term average will almost certainly be.

W.Feller (Feller 66) goes even farther than Loève, and as far as possible works entirely with distributions, setting up machinery which enables him to go for long stretches without mentioning probability spaces at all. I make no attempt to emulate him. But the approach is instructive and faithful to the essence of the subject.

Probability theory includes more mathematics than can easily be encompassed in a lifetime, and I have selected for this introductory chapter the two limit theorems I have already mentioned, the Strong Law of Large Numbers and the Central Limit Theorem, together with some material on martingales (§§275-276). They illustrate not only the special character of probability theory – so that you will be able to form your own judgement on the remarks above – but also some of its chief contributions to 'pure' measure theory, the concepts of 'independence' and 'conditional expectation'.

Version of 11.12.08

271 Distributions

I start this chapter with a discussion of 'probability distributions', the probability measures on $\mathbb{R}^n$ defined by families $(X_1,\dots,X_n)$ of random variables. I give the basic results describing the circumstances under which two distributions are equal (271G), integration with respect to a distribution (271E), and probability density functions (271H-271K).

271A Notation (a) Let (Ω,Σ,μ) be a probability space. A **real-valued random variable** on Ω will be a member of $\mathcal{L}^0(\mu)$.

(b) If X is a real-valued random variable on a probability space (Ω,Σ,μ), write $\mathbb{E}(X)=\int X\,d\mu$ if this is defined in $[-\infty,\infty]$. In this case I will call $\mathbb{E}(X)$ the **mean** or **expectation** of X. Thus we may say that 'X **has a finite expectation**' in place of 'X is integrable'.

(c) If X is a real-valued random variable with finite expectation, the **variance** of X is

$$\mathrm{Var}(X)=\mathbb{E}(X-\mathbb{E}(X))^2=\mathbb{E}(X^2-2\mathbb{E}(X)X+\mathbb{E}(X)^2)=\mathbb{E}(X^2)-(\mathbb{E}(X))^2.$$

(d) I shall allow myself to use such formulae as

$$\Pr(X>a),\quad \Pr(X-\epsilon\le Y\le X+\delta),$$

where X and Y are random variables on the same probability space (Ω,Σ,μ), to mean respectively

$$\hat\mu\{\omega:\omega\in\operatorname{dom}X,\ X(\omega)>a\},$$

$$\hat\mu\{\omega:\omega\in\operatorname{dom}X\cap\operatorname{dom}Y,\ X(\omega)-\epsilon\le Y(\omega)\le X(\omega)+\delta\},$$

writing $\hat\mu$ for the completion of μ. I will use this notation *only for predicates corresponding to Borel measurable sets*; that is to say, I shall write

$$\Pr(\psi(X_1,\dots,X_n))=\hat\mu\{\omega:\omega\in\textstyle\bigcap_{i\le n}\operatorname{dom}X_i,\ \psi(X_1(\omega),\dots,X_n(\omega))\}$$

only when the set

$$\{(\alpha_1,\dots,\alpha_n):\psi(\alpha_1,\dots,\alpha_n)\}$$

is a Borel set in $\mathbb{R}^n$.

271B Theorem Let (Ω,Σ,μ) be a probability space, and $X_1,\dots,X_n$ real-valued random variables on Ω. Set $\boldsymbol{X}(\omega)=(X_1(\omega),\dots,X_n(\omega))$ for $\omega\in\bigcap_{i\le n}\operatorname{dom}X_i$.

(a) There is a unique Radon measure ν on $\mathbb{R}^n$ such that

$$\nu\,]-\infty,a]=\Pr(X_i\le\alpha_i\text{ for every }i\le n)$$

whenever $a=(\alpha_1,\dots,\alpha_n)\in\mathbb{R}^n$;

(b) $\nu\mathbb{R}^n=1$ and $\nu E=\hat\mu(\boldsymbol{X}^{-1}[E])$ whenever νE is defined, where $\hat\mu$ is the completion of μ; in particular, $\nu E=\Pr((X_1,\dots,X_n)\in E)$ for every Borel set $E\subseteq\mathbb{R}^n$.

271C Definition Let (Ω, Σ, μ) be a probability space and $X_1, \dots, X_n$ real-valued random variables on Ω. By the **(joint) distribution** or **law** $\nu_{\boldsymbol{X}}$ of the family $\boldsymbol{X} = (X_1, \dots, X_n)$ I shall mean the Radon probability measure ν of 271B.

271D Remarks (e) If $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ are such that $X_i =_{\text{a.e.}} Y_i$ for each i, then $\nu_{\boldsymbol{X}} = \nu_{\boldsymbol{Y}}$.

271E Measurable functions of random variables: Proposition Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables ; write $\mathrm{T}_{\boldsymbol{X}}$ for the domain of the distribution $\nu_{\boldsymbol{X}}$, and let h be a $\mathrm{T}_{\boldsymbol{X}}$-measurable real-valued function defined $\nu_{\boldsymbol{X}}$-a.e. on $\mathbb{R}^n$. Then we have a random variable $Y = h(X_1, \dots, X_n)$ defined by setting

$$h(X_1, \dots, X_n)(\omega) = h(X_1(\omega), \dots, X_n(\omega)) \text{ for every } \omega \in \boldsymbol{X}^{-1}[\operatorname{dom} h].$$

The distribution ν_Y of Y is the measure on $\mathbb{R}$ defined by the formula

$$\nu_Y F = \nu_{\boldsymbol{X}} h^{-1}[F]$$

for just those sets $F \subseteq \mathbb{R}$ such that $h^{-1}[F] \in \mathrm{T}_{\boldsymbol{X}}$. Also

$$\mathbb{E}(Y) = \int h \, d\nu_{\boldsymbol{X}}$$

in the sense that if one of these exists in $[-\infty, \infty]$, so does the other, and they are then equal.

271F Corollary If X is a single random variable with distribution ν_X, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \, \nu_X(dx)$$

if either is defined in $[-\infty, \infty]$. Similarly

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \, \nu_X(dx)$$

(whatever X may be). If X, Y are two random variables then we have

$$\mathbb{E}(X \times Y) = \int xy \, \nu_{(X,Y)} d(x,y)$$

if either side is defined in $[-\infty, \infty]$.

Remark If ν is the distribution of a real-valued random variable, that is, a Radon probability measure on $\mathbb{R}$, I will say that the **expectation** $\mathbb{E}(\nu)$ of ν is $\int_{-\infty}^{\infty} x \, \nu(dx)$ if this is defined; if ν has finite expectation, then its **variance** $\mathrm{Var}(\nu)$ will be $\int x^2 \, \nu(dx) - (\mathbb{E}(\nu))^2$. Thus if X is a real-valued random variable with distribution ν_X, $\mathbb{E}(X) = \mathbb{E}(\nu_X)$ and $\mathrm{Var}(X) = \mathrm{Var}(\nu_X)$ whenever these are defined.

271G Distribution functions (a) If X is a real-valued random variable, its **distribution function** is the function $F_X : \mathbb{R} \to [0, 1]$ defined by setting

$$F_X(a) = \Pr(X \le a) = \nu_X \left]-\infty, a\right]$$

for every $a \in \mathbb{R}$. (**Warning!** some authors prefer $F_X(a) = \Pr(X < a)$.) X and Y have the same distribution iff $F_X = F_Y$.

(b) If $X_1, \dots, X_n$ are real-valued random variables on the same probability space, their **(joint) distribution function** is the function $F_{\boldsymbol{X}} : \mathbb{R}^n \to [0, 1]$ defined by writing

$$F_{\boldsymbol{X}}(a) = \Pr(X_i \le \alpha_i \ \forall \ i \le n)$$

whenever $a = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution function, they have the same distribution.

271H Densities Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables, all defined on the same probability space. A **density function** for $(X_1, \dots, X_n)$ is a Radon-Nikodým derivative, with respect to Lebesgue measure, for the distribution $\nu_{\boldsymbol{X}}$.

271I Proposition Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables, all defined on the same probability space. Write μ_L for Lebesgue measure on $\mathbb{R}^n$.

(a) There is a density function for $\boldsymbol{X}$ iff $\Pr(\boldsymbol{X} \in E) = 0$ for every Borel set E such that $\mu_L E = 0$.

(b) A non-negative Lebesgue integrable function f is a density function for $\boldsymbol{X}$ iff $\int_{]-\infty,a]} f d\mu_L = \Pr(\boldsymbol{X} \in]-\infty, a])$ for every $a \in \mathbb{R}^n$.

(c) Suppose that f is a density function for $\boldsymbol{X}$, and $G = \{x : f(x) > 0\}$. Then if h is a Lebesgue measurable real-valued function defined almost everywhere in G,

$$\mathbb{E}(h(\boldsymbol{X})) = \int h\,d\nu_{\boldsymbol{X}} = \int h \times f d\mu_L$$

if any of the three integrals is defined in $[-\infty, \infty]$, interpreting $(h \times f)(x)$ as 0 if $f(x) = 0$ and $x \notin \operatorname{dom} h$.

271J Theorem Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables, and $D \subseteq \mathbb{R}^n$ a Borel set such that $\Pr(\boldsymbol{X} \in D) = 1$. Let $\phi : D \to \mathbb{R}^n$ be a function which is differentiable relative to its domain everywhere in D; for $x \in D$, let $T(x)$ be a derivative of ϕ at x, and set $J(x) = |\det T(x)|$. Suppose that $J(x) \neq 0$ for each $x \in D$, and that $\boldsymbol{X}$ has a density function f; and suppose moreover that $\langle D_k \rangle_{k \in \mathbb{N}}$ is a disjoint sequence of Borel sets, with union D, such that $\phi_k = \phi \restriction D_k$ is injective for every k. Then $\phi(\boldsymbol{X})$ has a density function $g = \sum_{k=0}^{\infty} g_k$ where

$$\begin{aligned} g_k(y) &= \tfrac{f(\phi_k^{-1}(y))}{J(\phi_k^{-1}(y))} \text{ for } y \in \phi[D_k \cap \operatorname{dom} f], \\ &= 0 \text{ for } y \in \mathbb{R}^n \setminus \phi[D_k]. \end{aligned}$$

271K Proposition Let X, Y be two random variables with a joint density function f. Then $X \times Y$ has a density function h, where

$$h(u) = \int_{-\infty}^{\infty} \frac{1}{|v|} f(\tfrac{u}{v}, v) dv$$

whenever this is defined in $\mathbb{R}$.

***271L Proposition** Let $\langle X_n \rangle_{n \in \mathbb{N}}$ be a sequence of real-valued random variables converging in measure to a random variable X. Writing F_{X_n}, F_X for the distribution functions of X_n, X respectively,

$$F_X(a) = \inf_{b>a} \liminf_{n\to\infty} F_{X_n}(b) = \inf_{b>a} \limsup_{n\to\infty} F_{X_n}(b)$$

for every $a \in \mathbb{R}$.

Version of 3.4.09

272 Independence

I introduce the concept of 'independence' for families of events, σ-algebras and random variables. The first part of the section, down to 272G, amounts to an analysis of the elementary relationships between the three manifestations of the idea. In 272G I give the fundamental result that the joint distribution of a (finite) independent family of random variables is just the product of the individual distributions. Further expressions of the connexion between independence and product measures are in 272J, 272M and 272N. I give a version of the zero-one law (272O), and I end the section with a group of basic results from probability theory concerning sums and products of independent random variables (272R-272W).

272A Definitions Let (Ω, Σ, μ) be a probability space.

(a) A family $\langle E_i \rangle_{i \in I}$ in Σ is **(stochastically) independent** if

$$\mu(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_n}) = \prod_{j=1}^{n} \mu E_{i_j}$$

whenever $i_1, \dots, i_n$ are distinct members of I.

(b) A family $\langle\Sigma_i\rangle_{i\in I}$ of σ-subalgebras of Σ is **(stochastically) independent** if

$$\mu(E_1\cap E_2\cap\ldots\cap E_n)=\prod_{j=1}^n \mu E_j$$

whenever $i_1,\ldots,i_n$ are distinct members of I and $E_j\in\Sigma_{i_j}$ for every $j\le n$.

(c) A family $\langle X_i\rangle_{i\in I}$ of real-valued random variables on Ω is **(stochastically) independent** if

$$\Pr(X_{i_j}\le\alpha_j \text{ for every } j\le n)=\prod_{j=1}^n \Pr(X_{i_j}\le\alpha_j)$$

whenever $i_1,\ldots,i_n$ are distinct members of I and $\alpha_1,\ldots,\alpha_n\in\mathbb{R}$.

272C The σ-subalgebra defined by a random variable Let (Ω,Σ,μ) be a probability space and X a real-valued random variable defined on Ω. Write $\mathcal{B}$ for the σ-algebra of Borel subsets of $\mathbb{R}$, and Σ_X for

$$\{X^{-1}[F]:F\in\mathcal{B}\}\cup\{(\Omega\setminus\operatorname{dom}X)\cup X^{-1}[F]:F\in\mathcal{B}\}.$$

Then Σ_X is a σ-algebra of subsets of Ω.

Σ_X is the smallest σ-algebra of subsets of Ω, containing $\operatorname{dom}X$, for which X is measurable. Σ_X is a subalgebra of $\hat\Sigma$, where $\hat\Sigma$ is the domain of the completion of μ.

272D Proposition Let (Ω,Σ,μ) be a probability space and $\langle X_i\rangle_{i\in I}$ a family of real-valued random variables on Ω. For each $i\in I$, let Σ_i be the σ-algebra defined by X_i. Then the following are equiveridical:

(i) $\langle X_i\rangle_{i\in I}$ is independent;

(ii) whenever $i_1,\ldots,i_n$ are distinct members of I and $F_1,\ldots,F_n$ are Borel subsets of $\mathbb{R}$, then

$$\Pr(X_{i_j}\in F_j \text{ for every } j\le n)=\prod_{j=1}^n \Pr(X_{i_j}\in F_j);$$

(iii) whenever $\langle F_i\rangle_{i\in I}$ is a family of Borel subsets of $\mathbb{R}$, and $\{i:F_i\ne\mathbb{R}\}$ is finite, then

$$\hat\mu\bigl(\bigcap_{i\in I}(X_i^{-1}[F_i]\cup(\Omega\setminus\operatorname{dom}X_i))\bigr)=\prod_{i\in I}\Pr(X_i\in F_i),$$

where $\hat\mu$ is the completion of μ;

(iv) $\langle\Sigma_i\rangle_{i\in I}$ is independent with respect to $\hat\mu$.

272E Corollary Let $\langle X_i\rangle_{i\in I}$ be an independent family of real-valued random variables, and $\langle h_i\rangle_{i\in I}$ any family of Borel measurable functions from $\mathbb{R}$ to $\mathbb{R}$. Then $\langle h_i(X_i)\rangle_{i\in I}$ is independent.

272F Proposition Let (Ω,Σ,μ) be a probability space, and $\langle E_i\rangle_{i\in I}$ a family in Σ. Set $\Sigma_i=\{\emptyset,E_i,\Omega\setminus E_i,\Omega\}$, the ($\sigma$-)algebra of subsets of Ω generated by E_i, and $X_i=\chi E_i$, the indicator function of E_i. Then the following are equiveridical:

(i) $\langle E_i\rangle_{i\in I}$ is independent;

(ii) $\langle\Sigma_i\rangle_{i\in I}$ is independent;

(iii) $\langle X_i\rangle_{i\in I}$ is independent.

272G Distributions of independent random variables: Theorem Let $\boldsymbol{X}=(X_1,\ldots,X_n)$ be a finite family of real-valued random variables on a probability space. Let $\nu_{\boldsymbol{X}}$ be the corresponding distribution on $\mathbb{R}^n$. Then the following are equiveridical:

(i) $X_1,\ldots,X_n$ are independent;

(ii) $\nu_{\boldsymbol{X}}$ can be expressed as a product of n probability measures $\nu_1,\ldots,\nu_n$, one for each factor $\mathbb{R}$ of $\mathbb{R}^n$;

(iii) $\nu_{\boldsymbol{X}}$ is the product measure of $\nu_{X_1},\ldots,\nu_{X_n}$, writing ν_{X_i} for the distribution of the random variable X_i.

272H Corollary Suppose that $\langle X_i\rangle_{i\in I}$ is an independent family of real-valued random variables on a probability space (Ω,Σ,μ), and that for each $i\in I$ we are given another real-valued random variable Y_i on Ω such that $Y_i=_{\text{a.e.}}X_i$. Then $\langle Y_i\rangle_{i\in I}$ is independent.

272I Corollary Suppose that $X_1,\ldots,X_n$ are independent real-valued random variables with density functions $f_1,\ldots,f_n$. Then $\boldsymbol{X}=(X_1,\ldots,X_n)$ has a density function f given by setting $f(x)=\prod_{i=1}^n f_i(\xi_i)$ whenever $x=(\xi_1,\ldots,\xi_n)\in\prod_{i\le n}\operatorname{dom}(f_i)\subseteq\mathbb{R}^n$.

272J Proposition Let (Ω, Σ, μ) be a complete probability space, and $\langle \Sigma_i \rangle_{i \in I}$ a family of σ-subalgebras of Σ. For each $i \in I$ let μ_i be the restriction of μ to Σ_i, and let $(\Omega^I, \Lambda, \lambda)$ be the product probability space of the family $\langle (\Omega, \Sigma_i, \mu_i) \rangle_{i \in I}$. Define $\phi : \Omega \to \Omega^I$ by setting $\phi(\omega)(i) = \omega$ whenever $\omega \in \Omega$ and $i \in I$. Then ϕ is inverse-measure-preserving iff $\langle \Sigma_i \rangle_{i \in I}$ is independent.

272K Proposition Let (Ω, Σ, μ) be a probability space and $\langle \Sigma_i \rangle_{i \in I}$ an independent family of σ-subalgebras of Σ. Let $\langle J(s) \rangle_{s \in S}$ be a disjoint family of subsets of I, and for each $s \in S$ let $\tilde{\Sigma}_s$ be the σ-algebra of subsets of Ω generated by $\bigcup_{i \in J(s)} \Sigma_i$. Then $\langle \tilde{\Sigma}_s \rangle_{s \in S}$ is independent.

272L Corollary Let $X, X_1, \dots, X_n$ be independent real-valued random variables and $h : \mathbb{R}^n \to \mathbb{R}$ a Borel measurable function. Then X and $h(X_1, \dots, X_n)$ are independent.

272M Products of probability spaces and independent families of random variables: Proposition Let $\langle (\Omega_i, \Sigma_i, \mu_i) \rangle_{i \in I}$ be a family of probability spaces, and (Ω, Σ, μ) their product.

(a) For each $i \in I$ write $\tilde{\Sigma}_i = \{\pi_i^{-1}[E] : E \in \Sigma_i\}$, where $\pi_i : \Omega \to \Omega_i$ is the coordinate map. Then $\langle \tilde{\Sigma}_i \rangle_{i \in I}$ is an independent family of σ-subalgebras of Σ.

(b) For each $i \in I$ let $\langle X_{ij} \rangle_{j \in J(i)}$ be an independent family of real-valued random variables on Ω_i, and for $i \in I$, $j \in J(i)$ write $\tilde{X}_{ij}(\omega) = X_{ij}(\omega(i))$ for those $\omega \in \Omega$ such that $\omega(i) \in \operatorname{dom} X_{ij}$. Then $\langle \tilde{X}_{ij} \rangle_{i \in I, j \in J(i)}$ is an independent family of random variables, and each $\tilde{X}_{ij}$ has the same distribution as the corresponding X_{ij}.

272N Proposition Let (Ω, Σ, μ) be a complete probability space, and $\langle E_i \rangle_{i \in I}$ an independent family in Σ such that $\mu E_i = \frac{1}{2}$ for every $i \in I$. Define $\phi : \Omega \to \{0,1\}^I$ by setting $\phi(\omega)(i) = 1$ if $\omega \in E_i$, 0 if $\omega \in \Omega \setminus E_i$. Then ϕ is inverse-measure-preserving for the usual measure λ on $\{0,1\}^I$.

272O Tail σ-algebras and the zero-one law: Proposition Let (Ω, Σ, μ) be a probability space and $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$ an independent sequence of σ-subalgebras of Σ. Let Σ_n^* be the σ-algebra generated by $\bigcup_{m \ge n} \Sigma_m$ for each n, and set $\Sigma_\infty^* = \bigcap_{n \in \mathbb{N}} \Sigma_n^*$. Then μE is either 0 or 1 for every $E \in \Sigma_\infty^*$.

272P Corollary Let (Ω, Σ, μ) be a probability space, and $\langle X_n \rangle_{n \in \mathbb{N}}$ an independent sequence of real-valued random variables on Ω. Then

$$\limsup_{n \to \infty} \frac{1}{n+1}(X_0 + \ldots + X_n)$$

is almost everywhere constant.

***272Q Theorem** Let (Ω, Σ, μ) be a probability space, and $\langle \Sigma_i \rangle_{i \in I}$ an independent family of σ-subalgebras of Σ. Let $\mathcal{E} \subseteq \Sigma$ be a family of measurable sets, and T the σ-algebra generated by $\mathcal{E}$. Then there is a set $J \subseteq I$ such that $\#(I \setminus J) \le \max(\omega, \#(\mathcal{E}))$ and T, $\langle \Sigma_j \rangle_{j \in J}$ are independent, in the sense that $\mu(F \cap \bigcap_{r \le n} E_r) = \mu F \cdot \prod_{r=1}^n \mu E_r$ whenever $F \in \mathrm{T}$, $j_1, \dots, j_r$ are distinct members of J and $E_r \in \Sigma_{j_r}$ for each $r \le n$.

272R Proposition Let X, Y be independent real-valued random variables with finite expectation. Then $\mathbb{E}(X \times Y)$ exists and is equal to $\mathbb{E}(X)\mathbb{E}(Y)$.

272S Bienaymé's Equality Let $X_1, \dots, X_n$ be independent real-valued random variables. Then $\operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n)$.

272T The distribution of a sum of independent random variables: Theorem Let X, Y be independent real-valued random variables on a probability space (Ω, Σ, μ), with distributions ν_X, ν_Y. Then the distribution of $X + Y$ is the convolution $\nu_X * \nu_Y$.

272U Corollary Suppose that X and Y are independent real-valued random variables, and that they have densities f and g. Then the convolution $f*g$ is a density function for $X+Y$.

272V Etemadi's lemma Let $X_0,\dots,X_n$ be independent real-valued random variables. For $m\le n$, set $S_m=\sum_{i=0}^m X_i$. Then

$$\Pr(\sup_{m\le n}|S_m|\ge 3\gamma)\le 3\max_{m\le n}\Pr(|S_m|\ge\gamma)$$

for every $\gamma>0$.

***272W Theorem** Let $X_0,\dots,X_n$ be independent real-valued random variables such that $0\le X_i\le 1$ a.e. for every i. Set $S=\frac{1}{n+1}\sum_{i=0}^n X_i$ and $a=\mathbb{E}(S)$. Then

$$\Pr(S-a\ge c)\le\exp(-2(n+1)c^2)$$

for every $c\ge 0$.

Version of 2.12.09

273 The strong law of large numbers

I come now to the first of the three main theorems of this chapter. Perhaps I should call it a 'principle', rather than a 'theorem', as I shall not attempt to enunciate any fully general form, but will give three theorems (273D, 273H, 273I), with a variety of corollaries, each setting out conditions under which the averages of a sequence of independent random variables will almost surely converge. At the end of the section (273N) I add a result on norm-convergence of averages.

273A Lemma Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be a sequence of measurable sets in a measure space (Ω,Σ,μ), and suppose that $\sum_{n=0}^\infty\mu E_n<\infty$. Then $\{n:\omega\in E_n\}$ is finite for almost every $\omega\in\Omega$.

273B Lemma Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and set $S_n=\sum_{i=0}^n X_i$ for each $n\in\mathbb{N}$.

(a) If $\langle S_n\rangle_{n\in\mathbb{N}}$ is convergent in measure, then it is convergent almost everywhere.

(b) In particular, if $\mathbb{E}(X_n)=0$ for every n and $\sum_{n=0}^\infty\mathbb{E}(X_n^2)<\infty$, then $\sum_{n=0}^\infty X_n$ is defined, and finite, almost everywhere.

273C Lemma (a) If $\lim_{n\to\infty}x_n=x$, then $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n x_i=x$.

(b) Let $\langle x_n\rangle_{n\in\mathbb{N}}$ be such that $\sum_{i=0}^\infty x_i$ is defined in $\mathbb{R}$, and $\langle b_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence in $[0,\infty[$ diverging to ∞. Then $\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=0}^n b_kx_k=0$.

273D The strong law of large numbers: first form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and suppose that $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $]0,\infty[$, diverging to ∞, such that $\sum_{n=0}^\infty\frac{1}{b_n^2}\operatorname{Var}(X_n)<\infty$. Then

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=0}^n(X_i-\mathbb{E}(X_i))=0$$

almost everywhere.

273E Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables such that $\mathbb{E}(X_n)=0$ for every n and $\sup_{n\in\mathbb{N}}\mathbb{E}(X_n^2)<\infty$. Then

$$\lim_{n\to\infty}\frac{1}{b_n}(X_0+\ldots+X_n)=0$$

almost everywhere whenever $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of strictly positive numbers and $\sum_{n=0}^{\infty}\frac{1}{b_n^2}$ is finite. In particular,

$$\lim_{n\to\infty}\frac{1}{n+1}(X_0+\ldots+X_n)=0$$

almost everywhere.

273F Corollary Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of measurable sets in a probability space (Ω,Σ,μ). and suppose that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\mu E_i=c.$$

Then

$$\lim_{n\to\infty}\frac{1}{n+1}\#(\{i:i\le n,\,\omega\in E_i\})=c$$

for almost every $\omega\in\Omega$.

273G Corollary Let μ be the usual measure on $\mathcal{P}\mathbb{N}$. Then for μ-almost every set $a\subseteq\mathbb{N}$,

$$\lim_{n\to\infty}\frac{1}{n+1}\#(a\cap\{0,\ldots,n\})=\tfrac{1}{2}.$$

273H Strong law of large numbers: second form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|^{1+\delta})<\infty$ for some $\delta>0$. Then

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(X_i-\mathbb{E}(X_i))=0$$

almost everywhere.

273I Strong law of large numbers: third form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with finite expectation, and suppose that they are **identically distributed**, that is, all have the same distribution. Then

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(X_i-\mathbb{E}(X_i))=0$$

almost everywhere.

273J Corollary Let (Ω,Σ,μ) be a probability space. If f is a real-valued function such that $\int f d\mu$ is defined in $[-\infty,\infty]$, then

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f(\omega_i)=\int f d\mu$$

for λ-almost every $\boldsymbol{\omega}=\langle\omega_n\rangle_{n\in\mathbb{N}}\in\Omega^{\mathbb{N}}$, where λ is the product measure on $\Omega^{\mathbb{N}}$.

273K Borel-Cantelli lemma Let (Ω,Σ,μ) be a probability space and $\langle E_n\rangle_{n\in\mathbb{N}}$ a sequence of measurable subsets of Ω such that $\sum_{n=0}^{\infty}\mu E_n=\infty$ and $\mu(E_m\cap E_n)\le\mu E_m\cdot\mu E_n$ whenever $m\ne n$. Then almost every point of Ω belongs to infinitely many of the E_n.

273L Example There is an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of non-negative random variables such that $\lim_{n\to\infty}\mathbb{E}(X_n)=0$ but

$$\limsup_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{\infty}X_i-\mathbb{E}(X_i)=\infty,$$

$$\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{\infty}X_i-\mathbb{E}(X_i)=0$$

almost everywhere.

***273M Lemma** For any $p \in]1, \infty[$ and $\epsilon > 0$, there is a $\delta > 0$ such that $\|S+X\|_p \le 1+\epsilon\|X\|_p$ whenever S and X are independent random variables, $\|S\|_p = 1$, $\|X\|_p \le \delta$ and $\mathbb{E}(X) = 0$.

273N Theorem Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and set $Y_n = \frac{1}{n+1}(X_0 + \ldots + X_n)$ for each $n \in \mathbb{N}$.

(a) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is uniformly integrable, then $\lim_{n\to\infty} \|Y_n\|_1 = 0$.

*(b) If $p \in]1, \infty[$ and $\sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$, then $\lim_{n\to\infty} \|Y_n\|_p = 0$.

Version of 13.4.10

274 The central limit theorem

The second of the great theorems to which this chapter is devoted is of a new type. It is a limit theorem, but the limit involved is a limit of *distributions*, not of functions (as in the strong limit theorem above or the martingale theorem below), nor of equivalence classes of functions (as in Chapter 24). I give three forms of the theorem, in 274I-274K, all drawn as corollaries of Theorem 274G; the proof is spread over 274C-274G. In 274A-274B and 274M I give the most elementary properties of the normal distribution.

274A The normal distribution (a) If we set

$$\mu_G E = \frac{1}{\sqrt{2\pi}}\int_E e^{-x^2/2}dx$$

for every Lebesgue measurable set E, μ_G is a Radon probability measure; we call it the **standard normal distribution**. The corresponding distribution function is

$$\Phi(a) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-x^2/2}dx$$

for $a \in \mathbb{R}$; for the rest of this section I will reserve the symbol Φ for this function.

(b) A random variable X is **standard normal** if its distribution is μ_G.

(c) If X is a standard normal random variable, then

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty xe^{-x^2/2}dx = 0,$$

$$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty x^2e^{-x^2/2}dx = 1.$$

(d) More generally, a random variable X is **normal** if there are $a \in \mathbb{R}$ and $\sigma > 0$ such that $Z = (X-a)/\sigma$ is standard normal. In this case $\mathbb{E}(X) = a$, $\mathrm{Var}(X) = \sigma^2$. $x \mapsto \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2/2\sigma^2}$ is a density function for X. Conversely, a random variable with such a density function is normal, with expectation a and variance σ^2. The **normal distributions** are the distributions with these density functions.

(e) If X is normal, so is $a + bX$ for any $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$.

274B Proposition Let $X_1, \ldots, X_n$ be independent normal random variables. Then $Y = X_1 + \ldots + X_n$ is normal, with $\mathbb{E}(Y) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)$ and $\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$.

274C Lemma Let $U_0, \ldots, U_n, V_0, \ldots, V_n$ be independent real-valued random variables and $h : \mathbb{R} \to \mathbb{R}$ a bounded Borel measurable function. Then

$$|\mathbb{E}\big(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\big)| \le \sum_{i=0}^n \sup_{t\in\mathbb{R}} |\mathbb{E}\big(h(t+U_i) - h(t+V_i)\big)|.$$

274D Lemma Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded three-times-differentiable function such that $M_2 = \sup_{x\in\mathbb{R}} |h''(x)|$, $M_3 = \sup_{x\in\mathbb{R}} |h'''(x)|$ are both finite. Let $\epsilon > 0$.

(a) Let U be a real-valued random variable with zero expectation and finite variance σ^2. Then for any $t \in \mathbb{R}$ we have

$$|\mathbb{E}(h(t+U)) - h(t) - \frac{\sigma^2}{2}h''(t)| \le \frac{1}{6}\epsilon M_3\sigma^2 + M_2\mathbb{E}(\psi_\epsilon(U))$$

where $\psi_\epsilon(x) = 0$ if $|x| \le \epsilon$, x^2 if $|x| > \epsilon$.

(b) Let $U_0, \dots, U_n, V_0, \dots, V_n$ be independent random variables with finite variances, and suppose that $\mathbb{E}(U_i) = \mathbb{E}(V_i) = 0$ and $\mathrm{Var}(U_i) = \mathrm{Var}(V_i) = \sigma_i^2$ for every $i \le n$. Then

$$\begin{aligned}|\mathbb{E}\bigl(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\bigr)| \\ \le \frac{1}{3}\epsilon M_3 \sum_{i=0}^n \sigma_i^2 + M_2 \sum_{i=0}^n \mathbb{E}\bigl(\psi_\epsilon(U_i)\bigr) + M_2 \sum_{i=0}^n \mathbb{E}\bigl(\psi_\epsilon(V_i)\bigr).\end{aligned}$$

274E Lemma For any $\epsilon > 0$, there is a three-times-differentiable function $h : \mathbb{R} \to [0,1]$, with continuous third derivative, such that $h(x) = 1$ for $x \le -\epsilon$ and $h(x) = 0$ for $x \ge \epsilon$.

274F Lindeberg's theorem Let $\epsilon > 0$. Then there is a $\delta > 0$ such that whenever $X_0, \dots, X_n$ are independent real-valued random variables such that

$$\mathbb{E}(X_i) = 0 \text{ for every } i \le n,$$

$$\textstyle\sum_{i=0}^n \mathrm{Var}(X_i) = 1,$$

$$\textstyle\sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i)) \le \delta$$

(writing $\psi_\delta(x) = 0$ if $|x| \le \delta$, x^2 if $|x| > \delta$), then

$$\bigl|\Pr(\textstyle\sum_{i=0}^n X_i \le a) - \Phi(a)\bigr| \le \epsilon$$

for every $a \in \mathbb{R}$.

274G Central Limit Theorem Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables, all with zero expectation and finite variance; write $s_n = \sqrt{\sum_{i=0}^n \mathrm{Var}(X_i)}$ for each n. Suppose that

$$\lim_{n\to\infty} \frac{1}{s_n^2}\textstyle\sum_{i=0}^n \mathbb{E}(\psi_{\delta s_n}(X_i)) = 0 \text{ for every } \delta > 0,$$

writing $\psi_\delta(x) = 0$ if $|x| \le \delta$, x^2 if $|x| > \delta$. Set

$$S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for each $n \in \mathbb{N}$ such that $s_n > 0$. Then

$$\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

274H Remarks (a) The condition

$$\lim_{n\to\infty} \frac{1}{s_n^2}\textstyle\sum_{i=0}^n \mathbb{E}(\psi_{\epsilon s_n}(X_i)) = 0 \text{ for every } \epsilon > 0$$

is called **Lindeberg's condition**.

(b) Lindeberg's condition is necessary as well as sufficient, in the following sense. Suppose that $\langle X_n \rangle_{n\in\mathbb{N}}$ is an independent sequence of real-valued random variables with zero expectation and finite variance; write

$\sigma_n = \sqrt{\text{Var}(X_n)}$, $s_n = \sqrt{\sum_{i=0}^{n} \text{Var}(X_i)}$ for each n. Suppose that $\lim_{n\to\infty} s_n = \infty$, $\lim_{n\to\infty} \frac{\sigma_n}{s_n} = 0$ and that $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ for each $a \in \mathbb{R}$, where $S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$. Then

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=0}^{n} \mathbb{E}(\psi_{\epsilon s_n}(X_i)) = 0$$

for every $\epsilon > 0$.

274I Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, all with the same distribution, and suppose that their common expectation is 0 and their common variance is finite and not zero. Write σ for the common value of $\sqrt{\text{Var}(X_n)}$, and set

$$S_n = \frac{1}{\sigma\sqrt{n+1}}(X_0 + \ldots + X_n)$$

for each $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

274J Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and suppose that $\{X_n^2 : n \in \mathbb{N}\}$ is uniformly integrable and that

$$\liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \text{Var}(X_i) > 0.$$

Set

$$s_n = \sqrt{\sum_{i=0}^{n} \text{Var}(X_i)}, \quad S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for large $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

274K Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and suppose that

(i) there is some $\delta > 0$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{2+\delta}) < \infty$,

(ii) $\liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \text{Var}(X_i) > 0$.

Set $s_n = \sqrt{\sum_{i=0}^{n} \text{Var}(X_i)}$ and

$$S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for large $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

***274M Lemma** (a) $\int_x^\infty e^{-t^2/2} dt \le \frac{1}{x} e^{-x^2/2}$ for every $x > 0$.

(b) $\int_x^\infty e^{-t^2/2} dt \ge \frac{1}{2x} e^{-x^2/2}$ for every $x \ge 1$.

Version of 3.12.12

275 Martingales

This chapter so far has been dominated by independent sequences of random variables. I now turn to another of the remarkable concepts to which probabilistic intuitions have led us. Here we study evolving systems, in which we gain progressively more information as time progresses. I give the basic theorems on pointwise convergence of martingales (275F-275H, 275K) and a very brief account of 'stopping times' (275L-275P).

275A Definition Let (Ω, Σ, μ) be a probability space with completion $(\Omega, \hat{\Sigma}, \hat{\mu})$, and $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. A **martingale adapted to** $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$ is a sequence $\langle X_n \rangle_{n \in \mathbb{N}}$ of integrable real-valued random variables on Ω such that (i) $\operatorname{dom} X_n \in \Sigma_n$ and X_n is Σ_n-measurable for each $n \in \mathbb{N}$ (ii) whenever $m \le n \in \mathbb{N}$ and $E \in \Sigma_m$ then $\int_E X_n = \int_E X_m$.

275B Examples (a) Let (Ω, Σ, μ) be a probability space and $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Let X be any real-valued random variable on Ω with finite expectation, and for each $n \in \mathbb{N}$ let X_n be a conditional expectation of X on Σ_n. Subject to the conditions that $\operatorname{dom} X_n \in \Sigma_n$ and X_n is actually Σ_n-measurable for each n, $\langle X_n \rangle_{n \in \mathbb{N}}$ will be a martingale adapted to $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$.

(b) Let (Ω, Σ, μ) be a probability space and $\langle X_n \rangle_{n \in \mathbb{N}}$ an independent sequence of random variables all with zero expectation. For each $n \in \mathbb{N}$ let $\tilde{\Sigma}_n$ be the σ-algebra generated by $\bigcup_{i \le n} \Sigma_{X_i}$, writing Σ_{X_i} for the σ-algebra defined by X_i, and set $S_n = X_0 + \ldots + X_n$. Then $\langle S_n \rangle_{n \in \mathbb{N}}$ is a martingale adapted to $\tilde{\Sigma}_n$.

(c) Let (Ω, Σ, μ) be a probability space and $\langle X_n \rangle_{n \in \mathbb{N}}$ an independent sequence of random variables all with expectation 1. For each $n \in \mathbb{N}$ let $\tilde{\Sigma}_n$ be the σ-algebra generated by $\bigcup_{i \le n} \Sigma_{X_i}$, writing Σ_{X_i} for the σ-algebra defined by X_i, and set $W_n = X_0 \times \ldots \times X_n$. Then $\langle W_n \rangle_{n \in \mathbb{N}}$ is a martingale adapted to $\langle \tilde{\Sigma}_n \rangle_{n \in \mathbb{N}}$.

275C Remarks (c) The concept of 'martingale' can readily be extended to other index sets than $\mathbb{N}$; indeed, if I is any partially ordered set, we can say that $\langle X_i \rangle_{i \in I}$ is a martingale on (Ω, Σ, μ) adapted to $\langle \Sigma_i \rangle_{i \in I}$ if (i) each Σ_i is a σ-subalgebra of $\hat{\Sigma}$ (ii) each X_i is an integrable real-valued Σ_i-measurable random variable such that $\operatorname{dom} X_i \in \Sigma_i$ (iii) whenever $i \le j$ in I, then $\Sigma_i \subseteq \Sigma_j$ and $\int_E X_i = \int_E X_j$ for every $E \in \Sigma_i$.

(d) Given just a sequence $\langle X_n \rangle_{n \in \mathbb{N}}$ of integrable real-valued random variables on a probability space (Ω, Σ, μ), we can say simply that $\langle X_n \rangle_{n \in \mathbb{N}}$ is a **martingale** on (Ω, Σ, μ) if there is some non-decreasing sequence $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$ of σ-subalgebras of $\hat{\Sigma}$ such that $\langle X_n \rangle_{n \in \mathbb{N}}$ is a martingale adapted to $\langle \Sigma_n \rangle_{n \in \mathbb{N}}$. If we write $\tilde{\Sigma}_n$ for the σ-algebra generated by $\bigcup_{i \le n} \Sigma_{X_i}$, where Σ_{X_i} is the σ-algebra defined by X_i, then $\langle X_n \rangle_{n \in \mathbb{N}}$ is a martingale iff it is a martingale adapted to $\langle \tilde{\Sigma}_n \rangle_{n \in \mathbb{N}}$.

(e) Continuing from (d), if $\langle X_n \rangle_{n \in \mathbb{N}}$ is a martingale on (Ω, Σ, μ), and $X'_n =_{\text{a.e.}} X_n$ for every n, then $\langle X'_n \rangle_{n \in \mathbb{N}}$ is a martingale on (Ω, Σ, μ). Consequently we have a concept of 'martingale' as a sequence in $L^1(\mu)$, saying that a sequence $\langle X^\bullet_n \rangle_{n \in \mathbb{N}}$ in $L^1(\mu)$ is a martingale iff $\langle X_n \rangle_{n \in \mathbb{N}}$ is a martingale.

275D Lemma Let (Ω, Σ, μ) be a probability space, and $\langle X_n \rangle_{n \in \mathbb{N}}$ a martingale on Ω. Fix $n \in \mathbb{N}$ and set $X^* = \max(X_0, \ldots, X_n)$. Then for any $\epsilon > 0$,

$$\Pr(X^* \ge \epsilon) \le \frac{1}{\epsilon} \mathbb{E}(X_n^+),$$

writing $X_n^+ = \max(0, X_n)$.

275E Up-crossings Let $x_0, \ldots, x_n$ be any list of real numbers, and $a < b$ in $\mathbb{R}$. The **number of up-crossings from** a **to** b in the list $x_0, \ldots, x_n$ is the number of pairs (j, k) such that $0 \le j < k \le n$, $x_j \le a$, $x_k \ge b$ and $a < x_i < b$ for $j < i < k$.

275F Lemma Let (Ω,Σ,μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale on Ω. Suppose that $n\in\mathbb{N}$ and that $a<b$ in $\mathbb{R}$. For each $\omega\in\bigcap_{i\le n}\operatorname{dom}X_i$, let $U(\omega)$ be the number of up-crossings from a to b in the list $X_0(\omega),\dots,X_n(\omega)$. Then

$$\mathbb{E}(U)\le\frac{1}{b-a}\mathbb{E}((X_n-X_0)^+),$$

writing $(X_n-X_0)^+(\omega)=\max(0,X_n(\omega)-X_0(\omega))$ for $\omega\in\operatorname{dom}X_n\cap\operatorname{dom}X_0$.

275G Doob's Martingale Convergence Theorem Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale on a probability space (Ω,Σ,μ), and suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)<\infty$. Then $\lim_{n\to\infty}X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω in Ω.

275H Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Then the following are equiveridical:

(i) there is a random variable X, of finite expectation, such that X_n is a conditional expectation of X on Σ_n for every n;

(ii) $\{X_n:n\in\mathbb{N}\}$ is uniformly integrable;

(iii) $X_\infty(\omega)=\lim_{n\to\infty}X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω, and $\mathbb{E}(|X_\infty|)=\lim_{n\to\infty}\mathbb{E}(|X_n|)<\infty$.

275I Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ; write Σ_∞ for the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\Sigma_n$. Let X be any real-valued random variable on Ω with finite expectation, and for each $n\in\mathbb{N}$ let X_n be a conditional expectation of X on Σ_n. Then $X_\infty(\omega)=\lim_{n\to\infty}X_n(\omega)$ is defined almost everywhere; $\lim_{n\to\infty}\mathbb{E}(|X_\infty-X_n|)=0$, and X_∞ is a conditional expectation of X on Σ_∞.

***275J Proposition** Let $\langle(\Omega_n,\Sigma_n,\mu_n)\rangle_{n\in\mathbb{N}}$ be a sequence of probability spaces with product (Ω,Σ,μ). Let X be a real-valued random variable on Ω with finite expectation. For each $n\in\mathbb{N}$ define X_n by setting

$$X_n(\boldsymbol{\omega})=\int X(\omega_0,\dots,\omega_n,\xi_{n+1},\dots)d(\xi_{n+1},\dots)$$

wherever this is defined, where I write '$\int\dots d(\xi_{n+1},\dots)$' to mean integration with respect to the product measure λ_n' on $\prod_{i\ge n+1}\Omega_i$. Then $X(\boldsymbol{\omega})=\lim_{n\to\infty}X_n(\boldsymbol{\omega})$ for almost every $\boldsymbol{\omega}=(\omega_0,\omega_1,\dots)$ in Ω, and $\lim_{n\to\infty}\mathbb{E}(|X-X_n|)=0$.

275K Reverse martingales: Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-increasing sequence of σ-subalgebras of Σ, with intersection Σ_∞. Let X be any real-valued random variable with finite expectation, and for each $n\in\mathbb{N}$ let X_n be a conditional expectation of X on Σ_n. Then $X_\infty=\lim_{n\to\infty}X_n$ is defined almost everywhere and is a conditional expectation of X on Σ_∞.

275L Stopping times: Definition Let (Ω,Σ,μ) be a probability space, with completion $(\Omega,\hat\Sigma,\hat\mu)$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$. A **stopping time adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a function τ from Ω to $\mathbb{N}\cup\{\infty\}$ such that $\{\omega:\tau(\omega)\le n\}\in\Sigma_n$ for every $n\in\mathbb{N}$.

275M Examples (a) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to a sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras, and H_n is a Borel subset of $\mathbb{R}^{n+1}$ for each n, then we have a stopping time τ adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ defined by the formula

$$\tau(\omega)=\inf\{n:\omega\in\textstyle\bigcap_{i\le n}\operatorname{dom}X_i,\ (X_0(\omega),\dots,X_n(\omega))\in H_n\},$$

setting $\inf\emptyset=\infty$. In particular, the formulae

$$\inf\{n:X_n(\omega)\ge a\},\quad \inf\{n:|X_n(\omega)|>a\}$$

define stopping times.

(b) Any constant function $\tau:\Omega\to\mathbb{N}\cup\{\infty\}$ is a stopping time. If τ, τ' are two stopping times adapted to the same sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras, then $\tau\wedge\tau'$ is a stopping time adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, setting $(\tau\wedge\tau')(\omega)=\min(\tau(\omega),\tau'(\omega))$ for $\omega\in\Omega$.

275N Lemma Let (Ω,Σ,μ) be a complete probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Suppose that τ and τ' are stopping times on Ω, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale, all adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(a) The family

$$\tilde\Sigma_\tau=\{E:E\in\Sigma,\,E\cap\{\omega:\tau(\omega)\le n\}\in\Sigma_n\text{ for every }n\in\mathbb{N}\}$$

is a σ-subalgebra of Σ.

(b) If $\tau(\omega)\le\tau'(\omega)$ for every ω, then $\tilde\Sigma_\tau\subseteq\tilde\Sigma_{\tau'}$.

(c) Now suppose that τ is finite almost everywhere. Set

$$\tilde X_\tau(\omega)=X_{\tau(\omega)}(\omega)$$

whenever $\tau(\omega)<\infty$ and $\omega\in\operatorname{dom}X_{\tau(\omega)}$. Then $\operatorname{dom}\tilde X_\tau\in\tilde\Sigma_\tau$ and $\tilde X_\tau$ is $\tilde\Sigma_\tau$-measurable.

(d) If τ is essentially bounded, that is, there is some $m\in\mathbb{N}$ such that $\tau\le m$ almost everywhere, then $\mathbb{E}(\tilde X_\tau)$ exists and is equal to $\mathbb{E}(X_0)$.

(e) If $\tau\le\tau'$ almost everywhere, and τ' is essentially bounded, then $\tilde X_\tau$ is a conditional expectation of $\tilde X_{\tau'}$ on $\tilde\Sigma_\tau$.

275O Proposition Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale and τ a stopping time, both adapted to the same sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras. For each n, set $(\tau\wedge n)(\omega)=\min(\tau(\omega),n)$ for $\omega\in\Omega$; then $\tau\wedge n$ is a stopping time, and $\langle\tilde X_{\tau\wedge n}\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\tilde\Sigma_{\tau\wedge n}\rangle_{n\in\mathbb{N}}$, defining $\tilde X_{\tau\wedge n}$ and $\tilde\Sigma_{\tau\wedge n}$ as in 275N.

275P Corollary Suppose that (Ω,Σ,μ) is a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale on Ω such that $W=\sup_{n\in\mathbb{N}}|X_{n+1}-X_n|$ is finite almost everywhere and has finite expectation. Then for almost every $\omega\in\Omega$, either $\lim_{n\to\infty}X_n(\omega)$ exists in $\mathbb{R}$ or $\sup_{n\in\mathbb{N}}X_n(\omega)=\infty$ and $\inf_{n\in\mathbb{N}}X_n(\omega)=-\infty$.

Version of 16.4.13

276 Martingale difference sequences

Hand in hand with the concept of ‘martingale’ is that of ‘martingale difference sequence’ (276A), a direct generalization of the notion of ‘independent sequence’. In this section I collect results which can be naturally expressed in terms of difference sequences, including yet another strong law of large numbers (276C). I end the section with a proof of Komlós’s theorem (276H).

276A Martingale difference sequences (a) Let us say that if (Ω,Σ,μ) is a probability space, with completion $(\Omega,\hat\Sigma,\hat\mu)$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$, then a **martingale difference sequence adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of real-valued random variables on Ω, all with finite expectation, such that (i) $\operatorname{dom}X_n\in\Sigma_n$ and X_n is Σ_n-measurable, for each $n\in\mathbb{N}$ (ii) $\int_E X_{n+1}=0$ whenever $n\in\mathbb{N}$ and $E\in\Sigma_n$.

(b) $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ iff $\langle\sum_{i=0}^n X_i\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(c) Just as in 275Cd, we can say that a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ is in itself a **martingale difference sequence** if $\langle\sum_{i=0}^n X_i\rangle_{n\in\mathbb{N}}$ is a martingale.

(d) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence then $\langle a_nX_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence for any real a_n.

(e) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence and $X'_n=_{\text{a.e.}}X_n$ for every n, then $\langle X'_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence.

276B Proposition Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that $\sum_{n=0}^\infty \mathbb{E}(X_n^2) < \infty$. Then $\sum_{n=0}^\infty X_n$ is defined, and finite, almost everywhere.

276C The strong law of large numbers: fourth form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence, and suppose that $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $]0,\infty[$, diverging to ∞, such that $\sum_{n=0}^\infty \frac{1}{b_n^2}\operatorname{Var}(X_n) < \infty$. Then

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=0}^n X_i = 0$$

almost everywhere.

276D Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale such that $b_n = \mathbb{E}(X_n^2)$ is finite for each n.

(a) If $\sup_{n\in\mathbb{N}} b_n$ is infinite, then $\lim_{n\to\infty}\frac{1}{b_n}X_n = 0$ a.e.

(b) If $\sup_{n\ge1}\frac{1}{n}b_n < \infty$, then $\lim_{n\to\infty}\frac{1}{n}X_n = 0$ a.e.

276E 'Impossibility of systems' (b) If $\langle W_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, and $\langle Z_n\rangle_{n\ge1}$ is a sequence of random variables such that (i) Z_n is Σ_{n-1}-measurable (ii) $Z_n\times W_n$ has finite expectation for each $n\ge1$, then $W_0, Z_1\times W_1, Z_2\times W_2,\dots$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

***276F Lemma** Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Suppose that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a sequence of random variables on Ω such that (i) X_n is Σ_n-measurable for each n (ii) $\sum_{n=0}^\infty\frac{1}{(n+1)^2}\mathbb{E}(X_n^2)$ is finite (iii) $\lim_{n\to\infty}X_n' = 0$ a.e., where X_n' is a conditional expectation of X_n on Σ_{n-1} for each $n\ge1$. Then $\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^n X_k = 0$ a.e.

***276G Lemma** Let (Ω,Σ,μ) be a probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)$ is finite. For $k\in\mathbb{N}$ and $x\in\mathbb{R}$ set $F_k(x) = x$ if $|x|\le k$, 0 otherwise. Let $\mathcal{F}$ be an ultrafilter on $\mathbb{N}$.

(a) For each $k\in\mathbb{N}$ there is a measurable function $Y_k:\Omega\to[-k,k]$ such that $\lim_{n\to\mathcal{F}}\int_E F_k(X_n) = \int_E Y_k$ for every $E\in\Sigma$.

(b) $\lim_{n\to\mathcal{F}}\mathbb{E}((F_k(X_n)-Y_k)^2)\le\lim_{n\to\mathcal{F}}\mathbb{E}(F_k(X_n)^2)$ for each k.

(c) $Y=\lim_{k\to\infty}Y_k$ is defined a.e. and $\lim_{k\to\infty}\mathbb{E}(|Y-Y_k|)=0$.

***276H Komlós's theorem** Let (Ω,Σ,μ) be any measure space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of integrable real-valued functions on Ω such that $\sup_{n\in\mathbb{N}}\int|X_n|$ is finite. Then there are a subsequence $\langle X_n'\rangle_{n\in\mathbb{N}}$ of $\langle X_n\rangle_{n\in\mathbb{N}}$ and an integrable function Y such that $Y =_{\text{a.e.}} \lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n X_i''$ whenever $\langle X_n''\rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X_n'\rangle_{n\in\mathbb{N}}$.

Version of 8.4.09

Concordance

I list here the section and paragraph numbers which have (to my knowledge) appeared in print in references to this chapter, and which have since been changed.

272S Distribution of a sum of independent random variables This result, referred to in the 2002 and 2004 editions of Volume 3, and the 2003 and 2006 editions of Volume 4, is now 272T.

272U Etemadi's lemma This result, referred to in the 2003 and 2006 editions of Volume 4, is now 272V.

272Yd This exercise, referred to in the 2002 and 2004 editions of Volume 3, is now 272Ye.

273Xh This exercise, referred to in the 2006 edition of Volume 4, is now 273Xi.

276Xe This exercise, referred to in the 2003 and 2006 editions of Volume 4, is now 276Xg.