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 I have just spent some paragraphs on an attempt to describe the essential difference between probability theory and measure theory. But there is a quicker test by which you may discover whether your author is a measure theorist or a probabilist: open any page, and look for the phrases 'measurable function' and 'random variable', and the formulae '$\int f d\mu$' and '$\mathbb{E}(X)$'. The first member of each pair will enable you to diagnose 'measure' and the second 'probability', with little danger of error. So far in this treatise I have firmly used measure theorists' terminology, with a few individual quirks. But in a chapter on probability theory I find that measure-theoretic notation, while perfectly adequate in a formal sense, does such violence to the familiar formulations as to render them unnatural. Moreover, you must surely at some point – if you have not already done so – become familiar with probabilists' language. So in this chapter I will make a substantial step in that direction. Happily, I think that this can be done without setting up any direct conflicts, so that I shall be able, in later volumes, to call upon this work in whichever notation then seems appropriate, without needing to re-formulate it.

(a) So let (Ω,Σ,μ) be a probability space. I take the opportunity given by a new phrase to make a technical move. A **real-valued random variable** on Ω will be a member of $\mathcal{L}^0(\mu)$, as defined in 241A; that is, a real-valued function X defined on a conegligible subset of Ω such that X is measurable with respect to the completion $\hat\mu$ of μ, or, if you prefer, such that $X\restriction E$ is Σ-measurable for some conegligible set $E\subseteq\Omega$.[1]

(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 the sense of Chapter 12 and §133. 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'. 133A says that '$\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y)$ whenever $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ and their sum are defined in $[-\infty,\infty]$', and 122P becomes 'a real-valued random variable X has a finite expectation iff $\mathbb{E}(|X|)<\infty$'.

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

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

(Note that this formula shows that $\mathbb{E}(X)^2\le\mathbb{E}(X^2)$; compare 244Xd(i).) $\mathrm{Var}(X)$ is finite iff $\mathbb{E}(X^2)<\infty$, that is, iff $X\in\mathcal{L}^2(\mu)$ (244A). In particular, $X+Y$ and cX have finite variance whenever X and Y do and $c\in\mathbb{R}$.

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

[1] For an account of how this terminology became standard, see `http://www.dartmouth.edu/~chance/Doob/conversation.html`.

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

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 as usual. There are two points to note here. First, Pr depends on $\hat{\mu}$, not on μ; in effect, the notation automatically directs us to complete the probability space (Ω, Σ, μ). I could, of course, equally well write

$$\Pr(X^2 + Y^2 > 1) = \mu^*\{\omega : \omega \in \operatorname{dom} X \cap \operatorname{dom} Y,\ X(\omega)^2 + Y(\omega)^2 > 1\},$$

taking μ^* to be the outer measure on Ω associated with μ (132B). Secondly, 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 $\mathbb{R}^n$. Part of the reason for this restriction will appear in the next few paragraphs; $\Pr(\psi(X_1, \dots, X_n))$ must be something calculable from knowledge of the joint distribution of $X_1, \dots, X_n$, as defined in 271C. In fact we can safely extend the idea to 'universally measurable' predicates ψ, to be discussed in Volume 4. But it could happen that μ gave a measure to a set of the form $\{\omega : X(\omega) \in A\}$ for some exceedingly irregular set A, and in such a case it would be prudent to regard this as an accidental pathology of the probability space, and to treat it in a rather different way.

(I see that I have rather glibly assumed that the formula above defines $\Pr(\psi(X_1, \dots, X_n))$ for every Borel predicate ψ. This is a consequence of 271Bb below.)

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 $a = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, writing $]-\infty, a]$ for $\prod_{i \le n}]-\infty, \alpha_i]$;

(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 $E \subseteq \mathbb{R}^n$.

proof Let $\hat{\Sigma}$ be the domain of $\hat{\mu}$, and set $D = \bigcap_{i \le n} \operatorname{dom} X_i = \operatorname{dom} \boldsymbol{X}$; then D is conegligible, so belongs to $\hat{\Sigma}$. Let $\hat{\mu}_D = \hat{\mu} \restriction \mathcal{P}D$ be the subspace measure on D (131B, 214B), and ν_0 the image measure $\hat{\mu}_D \boldsymbol{X}^{-1}$ (234D); let T be the domain of ν_0.

Write $\mathcal{B}$ for the algebra of Borel sets in $\mathbb{R}^n$. Then $\mathcal{B} \subseteq \mathrm{T}$. **P** For $i \le n$, $\alpha \in \mathbb{R}$ set $F_{i\alpha} = \{x : x \in \mathbb{R}^n,\ \xi_i \le \alpha\}$, $H_{i\alpha} = \{\omega : \omega \in \operatorname{dom} X_i,\ X_i(\omega) \le \alpha\}$. X_i is $\hat{\Sigma}$-measurable and its domain is in $\hat{\Sigma}$, so $H_{i\alpha} \in \hat{\Sigma}$, and $\boldsymbol{X}^{-1}[F_{i\alpha}] = D \cap H_{i\alpha}$ is $\hat{\mu}_D$-measurable. Thus $F_{i\alpha} \in \mathrm{T}$. As T is a σ-algebra of subsets of $\mathbb{R}^n$, $\mathcal{B} \subseteq \mathrm{T}$ (121J). **Q**

Accordingly $\nu_0 \restriction \mathcal{B}$ is a measure on $\mathbb{R}^n$ with domain $\mathcal{B}$; of course $\nu_0 \mathbb{R}^n = \hat{\mu}D = 1$. By 256C, the completion ν of $\nu_0 \restriction \mathcal{B}$ is a Radon measure on $\mathbb{R}^n$, and $\nu\mathbb{R}^n = \nu_0\mathbb{R}^n = 1$.

For $E \in \mathcal{B}$,

$$\nu E = \nu_0 E = \hat{\mu}_D \boldsymbol{X}^{-1}[E] = \hat{\mu}\boldsymbol{X}^{-1}[E] = \Pr((X_1, \dots, X_n) \in E).$$

More generally, if $E \in \operatorname{dom} \nu$, then there are Borel sets E', E'' such that $E' \subseteq E \subseteq E''$ and $\nu(E'' \setminus E') = 0$, so that $\boldsymbol{X}^{-1}[E'] \subseteq \boldsymbol{X}^{-1}[E] \subseteq \boldsymbol{X}^{-1}[E'']$ and $\hat{\mu}(\boldsymbol{X}^{-1}[E''] \setminus \boldsymbol{X}^{-1}[E']) = 0$. This means that $\boldsymbol{X}^{-1}[E] \in \hat{\Sigma}$ and

$$\hat{\mu}\boldsymbol{X}^{-1}[E] = \hat{\mu}\boldsymbol{X}^{-1}[E'] = \nu E' = \nu E.$$

As for the uniqueness of ν, if ν' is any Radon measure on $\mathbb{R}^n$ such that $\nu'\,]-\infty, a] = \Pr(X_i \le \alpha_i\ \forall\, i \le n)$ for every $a \in \mathbb{R}^n$, then surely

$$\nu'\mathbb{R}^n = \lim_{k\to\infty} \nu'\,]-\infty, k\mathbf{1}] = \lim_{k\to\infty} \nu\,]-\infty, k\mathbf{1}] = 1 = \nu\mathbb{R}^n.$$

Also $\mathcal{I} = \{]-\infty, a] : a \in \mathbb{R}^n\}$ is closed under finite intersections, and ν and ν' agree on $\mathcal{I}$. By the Monotone Class Theorem (or rather, its corollary 136C), ν and ν' agree on the σ-algebra generated by $\mathcal{I}$, which is $\mathcal{B}$ (121J), and are identical (256D).

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. If we think of $\boldsymbol{X}$ as a function from $\bigcap_{i\le n} \operatorname{dom} X_i$ to $\mathbb{R}^n$, then $\nu_{\boldsymbol{X}}E = \Pr(\boldsymbol{X} \in E)$ for every Borel set $E \subseteq \mathbb{R}^n$.

271D Remarks (a) The choice of the Radon probability measure $\nu_{\boldsymbol{X}}$ as 'the' distribution of $\boldsymbol{X}$, with the insistence that 'Radon measures' should be complete, is of course somewhat arbitrary. Apart from the general principle that one should always complete measures, these conventions fit better with some of the work in Volume 4 and with such results as 272G below.

(b) Observe that in order to speak of the distribution of a family $\boldsymbol{X} = (X_1, \dots, X_n)$ of random variables, it is essential that all the X_i should be based on the same probability space.

(c) I see that the language I have chosen allows the X_i to have different domains, so that the family $(X_1, \dots, X_n)$ may not be exactly identifiable with the corresponding function from $\bigcap_{i\le n} \operatorname{dom} X_i$ to $\mathbb{R}^n$. I hope however that using the same symbol $\boldsymbol{X}$ for both will cause no confusion.

(d) It is not useful to think of the whole image measure $\nu_0 = \hat\mu_D \boldsymbol{X}^{-1}$ in the proof of 271B as the distribution of $\boldsymbol{X}$, unless it happens to be equal to $\nu = \nu_{\boldsymbol{X}}$. The 'distribution' of a random variable is exactly that aspect of it which can be divorced from any consideration of the underlying space (Ω, Σ, μ), and the point of such results as 271K and 272G is that distributions can be calculated from each other, without going back to the relatively fluid and uncertain model of a random variable in terms of a function on a probability space.

(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

$$\{\omega : \omega \in \textstyle\bigcap_{i\le n} \operatorname{dom} X_i,\ X_i(\omega) \le \alpha_i\ \forall\ i \le n\} \triangle \{\omega : \omega \in \bigcap_{i\le n} \operatorname{dom} Y_i,\ Y_i(\omega) \le \alpha_i\ \forall\ i \le n\}$$

is negligible, so

$$\begin{aligned}\Pr(X_i \le \alpha_i\ \forall\ i \le n) &= \hat\mu\{\omega : \omega \in \bigcap_{i\le n} \operatorname{dom} X_i,\ X_i(\omega) \le \alpha_i\ \forall\ i \le n\}\\ &= \Pr(Y_i \le \alpha_i\ \forall\ i \le n)\end{aligned}$$

for all $\alpha_0, \dots, \alpha_n \in \mathbb{R}$, and $\nu_{\boldsymbol{X}} = \nu_{\boldsymbol{Y}}$. This means that we can, if we wish, think of a distribution as a measure $\nu_{\boldsymbol{u}}$ where $\boldsymbol{u} = (u_0, \dots, u_n)$ is a finite sequence in $L^0(\mu)$. In the present chapter I shall not emphasize this approach, but it will always be at the back of my mind.

271E Measurable functions of random variables: Proposition Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables (as always in such a context, I mean them all to be on the same probability space (Ω, Σ, μ)); 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.

proof (a)(i) Once again, write $(\Omega,\hat\Sigma,\hat\mu)$ for the completion of (Ω,Σ,μ). Since

$$\Omega\setminus\operatorname{dom}Y\subseteq\bigcup_{i\le n}(\Omega\setminus\operatorname{dom}X_i)\cup\boldsymbol{X}^{-1}[\mathbb{R}^n\setminus\operatorname{dom}h]$$

is negligible (using 271Bb), $\operatorname{dom}Y$ is conegligible. If $a\in\mathbb{R}$, then

$$E=\{x:x\in\operatorname{dom}h,\,h(x)\le a\}\in\mathrm{T}_{\boldsymbol{X}},$$

so

$$\{\omega:\omega\in\Omega,\,Y(\omega)\le a\}=\boldsymbol{X}^{-1}[E]\in\hat\Sigma.$$

As a is arbitrary, Y is $\hat\Sigma$-measurable, and is a random variable.

(ii) Let $\tilde h:\mathbb{R}^n\to\mathbb{R}$ be any extension of h to the whole of $\mathbb{R}^n$. Then $\tilde h$ is $\mathrm{T}_{\boldsymbol{X}}$-measurable, so the ordinary image measure $\nu_{\boldsymbol{X}}\tilde h^{-1}$, defined on $\{F:\tilde h^{-1}[F]\in\operatorname{dom}\nu_{\boldsymbol{X}}\}$, is a Radon probability measure on $\mathbb{R}$ (256G). But for any $A\subseteq\mathbb{R}$,

$$\tilde h^{-1}[A]\triangle h^{-1}[A]\subseteq\mathbb{R}^n\setminus\operatorname{dom}h$$

is $\nu_{\boldsymbol{X}}$-negligible, so $\nu_{\boldsymbol{X}}h^{-1}[F]=\nu_{\boldsymbol{X}}\tilde h^{-1}[F]$ if either is defined.

If $F\subseteq\mathbb{R}$ is a Borel set, then

$$\nu_YF=\hat\mu\{\omega:Y(\omega)\in F\}=\hat\mu(\boldsymbol{X}^{-1}[h^{-1}[F]])=\nu_{\boldsymbol{X}}(h^{-1}[F]).$$

So ν_Y and $\nu_{\boldsymbol{X}}\tilde h^{-1}$ agree on the Borel sets and are equal (256D again).

(b) Now apply Theorem 235E to the measures $\hat\mu$ and $\nu_{\boldsymbol{X}}$ and the function $\phi=\boldsymbol{X}$. We have

$$\int\chi(\boldsymbol{X}^{-1}[F])d\hat\mu=\hat\mu(\boldsymbol{X}^{-1}[F])=\nu_{\boldsymbol{X}}F$$

for every $F\in\mathrm{T}_{\boldsymbol{X}}$, by 271Bb. Because h is $\nu_{\boldsymbol{X}}$-virtually measurable and defined $\nu_{\boldsymbol{X}}$-a.e., 235Eb tells us that

$$\int h(\boldsymbol{X})d\mu=\int h(\boldsymbol{X})d\hat\mu=\int h\,d\nu_{\boldsymbol{X}}$$

whenever either side is defined in $[-\infty,\infty]$, which is exactly the result we need.

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 (on the same probability space!) 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** $\operatorname{Var}(\nu)$ 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 $\operatorname{Var}(X)=\operatorname{Var}(\nu_X)$ 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

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

for every $a\in\mathbb{R}$. (**Warning!** some authors prefer $F_X(a)=\Pr(X<a)$.) Observe that F_X is non-decreasing, that $\lim_{a\to-\infty}F_X(a)=0$, that $\lim_{a\to\infty}F_X(a)=1$ and that $\lim_{x\downarrow a}F_X(x)=F_X(a)$ for every $a\in\mathbb{R}$. By 271Ba, 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, by the n-dimensional version of 271B.

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 $\nu_{\boldsymbol{X}}$; that is, a non-negative function f, integrable with respect to Lebesgue measure μ_L on $\mathbb{R}^n$, such that

$$\int_E f d\mu_L = \nu_{\boldsymbol{X}} E = \Pr(\boldsymbol{X} \in E)$$

for every Borel set $E \subseteq \mathbb{R}^n$ (256J) – if there is such a function, of course.

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 $x \notin \operatorname{dom} h$.

proof (a) Apply 256J to the Radon probability measure $\nu_{\boldsymbol{X}}$.

(b) Of course the condition is necessary. If it is satisfied, then (by B.Levi's theorem)

$$\int f d\mu_L = \lim_{k\to\infty} \int_{]-\infty,k\mathbf{1}]} f d\mu_L = \lim_{k\to\infty} \nu_{\boldsymbol{X}}\,]-\infty, k\mathbf{1}] = 1.$$

So we have a Radon probability measure ν defined by writing

$$\nu E = \int_E f d\mu_L$$

whenever $E \cap \{x : f(x) > 0\}$ is Lebesgue measurable (256E). We are supposing that $\nu\,]-\infty, a] = \nu_{\boldsymbol{X}}\,]-\infty, a]$ for every $a \in \mathbb{R}^n$; by 271Ba, as usual, $\nu = \nu_{\boldsymbol{X}}$, so

$$\int_E f d\mu_L = \nu E = \nu_{\boldsymbol{X}} E = \Pr(\boldsymbol{X} \in E)$$

for every Borel set $E \subseteq \mathbb{R}^n$, and f is a density function for $\boldsymbol{X}$.

(c) By 256E, $\nu_{\boldsymbol{X}}$ is the indefinite-integral measure over μ associated with f. So, writing $G = \{x : f(x) > 0\}$, we have

$$\int h \, d\nu_{\boldsymbol{X}} = \int h \times f d\mu_L$$

whenever either is defined in $[-\infty, \infty]$ (235K). By 234La, h is $\mathrm{T}_{\boldsymbol{X}}$-measurable and defined $\nu_{\boldsymbol{X}}$-almost everywhere, where $\mathrm{T}_{\boldsymbol{X}} = \operatorname{dom} \nu_{\boldsymbol{X}}$, so $\mathbb{E}(h(\boldsymbol{X})) = \int h \, d\nu_{\boldsymbol{X}}$ by 271E.

271J The machinery developed in §263 is sufficient to give a very general result on the densities of random variables of the form $\phi(\boldsymbol{X})$, as follows.

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

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

proof By 262Ia, ϕ is continuous, therefore Borel measurable, so $\phi(\boldsymbol{X})$ is a random variable.

For the moment, fix $k \in \mathbb{N}$ and a Borel set $F \subseteq \mathbb{R}^n$. By 263D(iii), $\phi[D_k]$ is measurable, and by 263D(ii) $\phi[D_k \setminus \operatorname{dom} f]$ is negligible. The function g_k is such that $f(x) = J(x)g_k(\phi(x))$ for every $x \in D_k \cap \operatorname{dom} f$, so by 263D(v) we have

$$\int_F g_k\, d\mu = \int_{\phi[D_k]} g_k \times \chi F\, d\mu = \int_{D_k} J(x)g_k(\phi(x))\chi F(\phi(x))\mu(dx)$$
$$= \int_{D_k \cap \phi^{-1}[F]} f d\mu = \Pr(\boldsymbol{X} \in D_k \cap \phi^{-1}[F]).$$

(The integral $\int_{\phi[D_k]} g_k \times \chi F$ is defined because $\int_{D_k} J \times (g_k \times \chi F)\phi$ is defined, and the integral $\int g_k \times \chi F$ is defined because $\phi[D_k]$ is measurable and g is zero off $\phi[D_k]$.)

Now sum over k. Every g_k is non-negative, so by B.Levi's theorem (123A, 123Xa)

$$\int_F g\, d\mu = \sum_{k=0}^{\infty} \int_F g_k\, d\mu = \sum_{k=0}^{\infty} \Pr(\boldsymbol{X} \in D_k \cap \phi^{-1}[F])$$
$$= \Pr(\boldsymbol{X} \in \phi^{-1}[F]) = \Pr(\phi(\boldsymbol{X}) \in F).$$

As F is arbitrary, g is a density function for $\phi(\boldsymbol{X})$, as claimed.

271K The application of the last theorem to ordinary transformations is sometimes indirect, so I give an example.

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 $\mathbb{R}$.

proof Set $\phi(x, y) = (xy, y)$ for x, $y \in \mathbb{R}^2$. Then ϕ is differentiable, with derivative $T(x, y) = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}$, so $J(x, y) = |\det T(x, y)| = |y|$. Set $D = \{(x, y) : y \neq 0\}$; then D is a conegligible Borel set in $\mathbb{R}^2$ and $\phi\restriction D$ is injective. Now $\phi[D] = D$ and $\phi^{-1}(u, v) = (\frac{u}{v}, v)$ for $v \neq 0$. So $\phi(X, Y) = (X \times Y, Y)$ has a density function g, where

$$g(u, v) = \frac{f(u/v, v)}{|v|} \text{ if } v \neq 0.$$

To find a density function for $X \times Y$, we calculate

$$\Pr(X \times Y \leq a) = \int_{]-\infty,a] \times \mathbb{R}} g = \int_{-\infty}^{a} \int_{-\infty}^{\infty} g(u, v) dv\, du = \int_{-\infty}^{a} h$$

by Fubini's theorem (252B, 252C). In particular, h is defined and finite almost everywhere; and by 271Ib it is a density function for $X \times Y$.

***271L** When a random variable is presented as the limit of a sequence of random variables the following can be very useful.

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 (definition: 245A). 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 $a \in \mathbb{R}$.

proof Set $\gamma = \inf_{b>a} \liminf_{n\to\infty} F_{X_n}(b)$, $\gamma' = \inf_{b>a} \limsup_{n\to\infty} F_{X_n}(b)$.

(a) $F_X(a) \le \gamma$. **P** Take any $b > a$ and $\epsilon > 0$. Then there is an $n_0 \in \mathbb{N}$ such that $\Pr(|X_n - X| \ge b-a) \le \epsilon$ for every $n \ge n_0$ (245F). Now, for $n \ge n_0$,

$$F_X(a) = \Pr(X \le a) \le \Pr(X_n \le b) + \Pr(X_n - X \ge b-a) \le F_{X_n}(b) + \epsilon.$$

So $F_X(a) \le \liminf_{n\to\infty} F_{X_n}(b)+\epsilon$; as ϵ is arbitrary, $F_X(a) \le \liminf_{n\to\infty} F_{X_n}(b)$; as b is arbitrary, $F_X(a) \le \gamma$. **Q**

(b) $\gamma' \le F_X(a)$. **P** Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $F_X(a+2\delta) \le F_X(a) + \epsilon$ (271Ga). Next, there is an $n_0 \in \mathbb{N}$ such that $\Pr(|X_n - X| \ge \delta) \le \epsilon$ for every $n \ge n_0$. In this case, for $n \ge n_0$,

$$\begin{aligned} F_{X_n}(a+\delta) &= \Pr(X_n \le a+\delta) \le \Pr(X \le a+2\delta) + \Pr(X - X_n \ge \delta) \\ &\le F_X(a+2\delta) + \epsilon \le F_X(a) + 2\epsilon. \end{aligned}$$

Accordingly

$$\gamma' \le \limsup_{n\to\infty} F_{X_n}(a+\delta) \le F_X(a) + 2\epsilon.$$

As ϵ is arbitrary, $\gamma' \le F_X(a)$. **Q**

(c) Since of course $\gamma \le \gamma'$, we must have $F_X(a) = \gamma = \gamma'$, as claimed.

271X Basic exercises >(a) Let X be a real-valued random variable with finite expectation, and $\epsilon > 0$. Show that $\Pr(|X - \mathbb{E}(X)| \ge \epsilon) \le \frac{1}{\epsilon^2}\operatorname{Var}(X)$. (This is **Chebyshev's inequality**.)

>(b) Let $F : \mathbb{R} \to [0,1]$ be a non-decreasing function such that (i) $\lim_{a\to-\infty} F(a) = 0$ (ii) $\lim_{a\to\infty} F(a) = 1$ (iii) $\lim_{x\downarrow a} F(x) = F(a)$ for every $a \in \mathbb{R}$. Show that there is a unique Radon probability measure ν in $\mathbb{R}$ such that $F(a) = \nu\,]-\infty, a]$ for every $a \in \mathbb{R}$. (*Hint*: 114Xa.) Hence show that F is the distribution function of some random variable.

>(c) Let X be a real-valued random variable with a density function f. (i) Show that $|X|$ has a density function g_1 where $g_1(x) = f(x) + f(-x)$ whenever $x \ge 0$ and $f(x)$, $f(-x)$ are both defined, 0 otherwise. (ii) Show that X^2 has a density function g_2 where $g_2(x) = (f(\sqrt{x}) + f(-\sqrt{x}))/2\sqrt{x}$ whenever $x > 0$ and this is defined, 0 for other x. (iii) Show that if $\Pr(X = 0) = 0$ then $1/X$ has a density function g_3 where $g_3(x) = \frac{1}{x^2}f(\frac{1}{x})$ whenever this is defined. (iv) Show that if $\Pr(X < 0) = 0$ then $\sqrt{X}$ has a density function g_4 where $g_4(x) = 2xf(x^2)$ if $x \ge 0$ and $f(x^2)$ is defined, 0 otherwise.

>(d) Let X and Y be random variables with a joint density function $f : \mathbb{R}^2 \to \mathbb{R}$. Show that $X + Y$ has a density function h where $h(u) = \int f(u-v, v)dv$ for almost every u.

(e) Let X, Y be random variables with a joint density function $f : \mathbb{R}^2 \to \mathbb{R}$. Show that X/Y has a density function h where $h(u) = \int |v| f(uv, v)dv$ for almost every u.

(f) Devise an alternative proof of 271K by using Fubini's theorem and one-dimensional substitutions to show that

$$\int_a^b \int_{-\infty}^{\infty} \frac{1}{|v|} f(\tfrac{u}{v}, v)dv\,du = \int_{\{(u,v):a\le uv\le b\}} f$$

whenever $a \le b$ in $\mathbb{R}$.

271Y Further exercises (a) Let $\mathfrak{T}$ be the topology of $\mathbb{R}^{\mathbb{N}}$ and $\mathcal{B}$ the σ-algebra of Borel sets (256Yf). (i) Let $\mathcal{I}$ be the family of sets of the form

$$\{x : x \in \mathbb{R}^{\mathbb{N}},\, x(i) \le \alpha_i\ \forall\, i \le n\},$$

where $n \in \mathbb{N}$ and $\alpha_i \in \mathbb{R}$ for each $i \le n$. Show that $\mathcal{B}$ is the smallest family of subsets of $\mathbb{R}^{\mathbb{N}}$ such that (α) $\mathcal{I} \subseteq \mathcal{B}$ (β) $B \setminus A \in \mathcal{B}$ whenever A, $B \in \mathcal{B}$ and $A \subseteq B$ (γ) $\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{B}$ for every non-decreasing sequence

$\langle A_k\rangle_{k\in\mathbb{N}}$ in $\mathcal{B}$. (ii) Show that if μ, μ' are two totally finite measures defined on $\mathbb{R}^{\mathbb{N}}$, and μF and $\mu' F$ are defined and equal for every $F\in\mathcal{I}$, then μE and $\mu' E$ are defined and equal for every $E\in\mathcal{B}$. (iii) Show that if Ω is a set and Σ a σ-algebra of subsets of Ω and $X:\Omega\to\mathbb{R}^{\mathbb{N}}$ is a function, then $X^{-1}[E]\in\Sigma$ for every $E\in\mathcal{B}$ iff π_iX is Σ-measurable for every $i\in\mathbb{N}$, where $\pi_i(x)=x(i)$ for each $x\in\mathbb{R}^{\mathbb{N}}$, $i\in\mathbb{N}$. (iv) Show that if $\boldsymbol{X}=\langle X_i\rangle_{i\in\mathbb{N}}$ is a sequence of real-valued random variables on a probability space (Ω,Σ,μ), then there is a unique probability measure $\nu^{\mathcal{B}}_{\boldsymbol{X}}$, with domain $\mathcal{B}$, such that $\nu^{\mathcal{B}}_{\boldsymbol{X}}\{x:x(i)\le\alpha_i\ \forall\ i\le n\}=\Pr(X_i\le\alpha_i\ \forall\ i\le n)$ for every $\alpha_0,\dots,\alpha_n\in\mathbb{R}$. (v) Under the conditions of (iv), show that there is a unique Radon measure $\nu_{\boldsymbol{X}}$ on $\mathbb{R}^{\mathbb{N}}$ (in the sense of 256Yf) such that $\nu_{\boldsymbol{X}}\{x:x(i)\le\alpha_i\ \forall\ i\le n\}=\Pr(X_i\le\alpha_i\ \forall\ i\le n)$ for every $\alpha_0,\dots,\alpha_n\in\mathbb{R}$.

(b) Let $F:\mathbb{R}^2\to[0,1]$ be a function. Show that the following are equiveridical: (i) F is the distribution function of some pair (X_1,X_2) of random variables (ii) there is a probability measure ν on $\mathbb{R}^2$ such that $\nu\,]-\infty,a]=F(a)$ for every $a\in\mathbb{R}^2$ (iii)(α) $F(\alpha_1,\alpha_2)+F(\beta_1,\beta_2)\ge F(\alpha_1,\beta_2)+F(\alpha_2,\beta_1)$ whenever $\alpha_1\le\beta_1$ and $\alpha_2\le\beta_2$ (β) $F(\alpha_1,\alpha_2)=\lim_{\xi_1\downarrow\alpha_1,\xi_2\downarrow\alpha_2}F(\xi_1,\xi_2)$ for every α_1, α_2 (γ) $\lim_{\alpha\to-\infty}F(\alpha,\beta)=\lim_{\alpha\to-\infty}F(\beta,\alpha)=0$ for all β (δ) $\lim_{\alpha\to\infty}F(\alpha,\alpha)=1$. (*Hint*: for non-empty half-open intervals $]a,b]$, set $\lambda\,]a,b]=F(\alpha_1,\alpha_2)+F(\beta_1,\beta_2)-F(\alpha_1,\beta_2)-F(\alpha_2,\beta_1)$, and continue as in 115B-115F.)

(c) Generalize (b) to higher dimensions, finding a suitable formula to stand in place of that in (iii-α) of (b).

(d) Let (Ω,Σ,μ) be a probability space and $\mathcal{F}$ a filter on $\mathcal{L}^0(\mu)$ converging to $X_0\in\mathcal{L}^0(\mu)$ for the topology of convergence in measure. Show that, writing F_X for the distribution function of $X\in\mathcal{L}^0(\mu)$,

$$F_{X_0}(a)=\inf_{b>a}\liminf_{X\to\mathcal{F}}F_X(b)=\inf_{b>a}\limsup_{X\to\mathcal{F}}F_X(b)$$

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

(e) Let X, Y be non-negative random variables with the same distribution, and $h:[0,\infty[\,\to[0,\infty[$ a non-decreasing function. Show that $\mathbb{E}(X\times hY)\le\mathbb{E}(Y\times hY)$. (*Hint*: in the language of 252Yo, $(Y\times hY)^*=Y^*\times(hY)^*$.)

271 Notes and comments Most of this section seems to have been taken up with technicalities. This is perhaps unsurprising in view of the fact that it is devoted to the relationship between a vector random variable $\boldsymbol{X}$ and the associated distribution $\nu_{\boldsymbol{X}}$, and this necessarily leads us into the minefield which I attempted to chart in §235. Indeed, I call on results from §235 twice; once in 271E, with a $\phi(\omega)=\boldsymbol{X}(\omega)$ and $J(\omega)=1$, and once in 271I, with $\phi(x)=x$ and $J(x)=f(x)$.

Distribution functions of one-dimensional random variables are easily characterized (271Xb); in higher dimensions we have to work harder (271Yb-271Yc). Distributions, rather than distribution functions, can be described for infinite sequences of random variables (271Ya); indeed, these ideas can be extended to uncountable families, but this requires proper topological measure theory, and belongs in Volume 4.

The statement of 271J is elaborate, not to say cumbersome. The point is that many of the most important transformations ϕ are not themselves injective, but can easily be dissected into injective fragments (see, for instance, 271Xc and 263Xd). The point of 271K is that we frequently wish to apply the ideas here to transformations which are singular, and indeed change the dimension of the random variable. I have not given the theorems which make such applications routine and suggest rather that you seek out tricks such as that used in the proof of 271K, which in any case are necessary if you want amenable formulae. Of course other methods are available (271Xf).

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 $i_1,\ldots,i_n$ 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 $\alpha_1,\ldots,\alpha_n\in\mathbb{R}$.

272B Remarks (a) This is perhaps the central contribution of probability theory to measure theory, and as such deserves the most careful scrutiny. The idea of 'independence' comes from outside mathematics altogether, in the notion of events which have independent causes. I suppose that 272G and 272M are the results below which most clearly show the measure-theoretic aspects of the concept. It is not an accident that both involve product measures; one of the wonders of measure theory is the fact that the same technical devices are used in establishing the probability theory of stochastic independence and the geometry of multi-dimensional volume.

(b) In the following paragraphs I will try to describe some relationships between the three notions of independence just defined. But it is worth noting at once the fact that, in all three cases, a family is independent iff all its finite subfamilies are independent. Consequently any subfamily of an independent family is independent. Another elementary fact which is immediate from the definitions is that if $\langle\Sigma_i\rangle_{i\in I}$ is an independent family of σ-algebras, and Σ_i' is a σ-subalgebra of Σ_i for each i, then $\langle\Sigma_i'\rangle_{i\in I}$ is an independent family.

(c) A useful reformulation of 272Ab is the following: A family $\langle\Sigma_i\rangle_{i\in I}$ of σ-subalgebras of Σ is independent iff

$$\mu(\bigcap_{i\in I}E_i)=\prod_{i\in I}\mu E_i$$

whenever $E_i\in\Sigma_i$ for every i and $\{i:E_i\ne\Omega\}$ is finite. (Here I follow the convention of 254F, saying that for a family $\langle\alpha_i\rangle_{i\in I}$ in $[0,1]$ we take $\prod_{i\in I}\alpha_i=1$ if $I=\emptyset$, and otherwise it is to be $\inf_{J\subseteq I,J \text{ is finite}}\prod_{i\in J}\alpha_j$.)

(d) In 272Aa-b I speak of sets $E_i\in\Sigma$ and algebras $\Sigma_i\subseteq\Sigma$. In fact (272Ac already gives a hint of this) we shall more often than not be concerned with $\hat\Sigma$ rather than with Σ, if there is a difference, where $(\Omega,\hat\Sigma,\hat\mu)$ is the completion of (Ω,Σ,μ).

272C The σ-subalgebra defined by a random variable To relate 272Ab to 272Ac we need the following notion. 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. **P**

$$\emptyset=X^{-1}[\emptyset]\in\Sigma_X;$$

if $F \in \mathcal{B}$ then

$$\Omega \setminus X^{-1}[F] = (\Omega \setminus \operatorname{dom} X) \cup X^{-1}[\mathbb{R} \setminus F] \in \Sigma_X,$$

$$\Omega \setminus ((\Omega \setminus \operatorname{dom} X) \cup X^{-1}[F]) = X^{-1}[\mathbb{R} \setminus F] \in \Sigma_X;$$

if $\langle F_k \rangle_{k \in \mathbb{N}}$ is any sequence in $\mathcal{B}$ then

$$\bigcup_{k \in \mathbb{N}} X^{-1}[F_k] = X^{-1}[\bigcup_{k \in \mathbb{N}} F_k],$$

so

$$\bigcup_{k \in \mathbb{N}} X^{-1}[F_k], \quad (\Omega \setminus \operatorname{dom} X) \cup \bigcup_{k \in \mathbb{N}} X^{-1}[F_k]$$

belong to Σ_X. **Q**

Evidently Σ_X is the smallest σ-algebra of subsets of Ω, containing $\operatorname{dom} X$, for which X is measurable. Also Σ_X is a subalgebra of $\hat{\Sigma}$, where $\hat{\Sigma}$ is the domain of the completion of μ (271Aa).

Now we have the following result.

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, as in 272C. Then the following are equiveridical:

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

(ii) whenever $i_1, \dots, i_n$ are distinct members of I and $F_1, \dots, F_n$ 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

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

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

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

proof (a)(i)$\Rightarrow$(ii) Write $\boldsymbol{X} = (X_{i_1}, \dots, X_{i_n})$. Write $\nu_{\boldsymbol{X}}$ for the joint distribution of $\boldsymbol{X}$, and for each $j \le n$ write ν_j for the distribution of X_{i_j}; let ν be the product of $\nu_1, \dots, \nu_n$ as described in 254A-254C. (I wrote §254 out as for infinite products. If you are interested only in finite products of probability spaces, which are adequate for our needs in this paragraph, I recommend reading §§251-252 with the mental proviso that all measures are probabilities, and then §254 with the proviso that the set I is finite.) By 256K, ν is a Radon measure on $\mathbb{R}^n$. (This is an induction on n, relying on 254N for assurance that we can regard ν as the repeated product $(\dots((\nu_1 \times \nu_2) \times \nu_3) \times \dots \nu_{n-1}) \times \nu_n$.) Then for any $a = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, we have

$$\nu\,]-\infty, a] = \nu\big(\prod_{j=1}^n]-\infty, \alpha_j]\big) = \prod_{j=1}^n \nu_j\,]-\infty, \alpha_j]$$

(using 254Fb)

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

(using the condition (i))

$$= \nu_{\boldsymbol{X}}\,]-\infty, a]\,.$$

By the uniqueness assertion in 271Ba, $\nu = \nu_{\boldsymbol{X}}$. In particular, if $F_1, \dots, F_n$ are Borel subsets of $\mathbb{R}$,

$$\Pr(X_{i_j} \in F_j \text{ for every } j \le n) = \Pr(\boldsymbol{X} \in \prod_{j \le n} F_j) = \nu_{\boldsymbol{X}}(\prod_{j \le n} F_j)$$

$$= \nu(\prod_{j \le n} F_j) = \prod_{j=1}^n \nu_j F_j = \prod_{j=1}^n \Pr(X_{i_j} \in F_j),$$

as required.

(b)(ii)⇒(i) is trivial, if we recall that all sets $]-\infty,\alpha]$ are Borel sets, so that the definition of independence given in 272Ac is just a special case of (ii).

(c)(ii)⇒(iv) Assume (ii), and suppose that $i_1,\dots,i_n$ are distinct members of I and $E_j\in\Sigma_{i_j}$ for each $j\le n$. For each j, set $E'_j=E_j\cap\operatorname{dom}X_{i_j}$, so that E'_j may be expressed as $X_{i_j}^{-1}[F_j]$ for some Borel set $F_j\subseteq\mathbb{R}$. Then $\hat\mu(E_j\setminus E'_j)=0$ for each j, so

$$\hat\mu(\bigcap_{1\le j\le n}E_j)=\hat\mu(\bigcap_{1\le j\le n}E'_j)=\Pr(X_{i_1}\in F_1,\dots,X_{i_n}\in F_n)$$
$$=\prod_{j=1}^n\Pr(X_{i_j}\in F_j)$$

(using (ii))

$$=\prod_{i=1}^n\hat\mu E_j.$$

As $E_1,\dots,E_k$ are arbitrary, $\langle\Sigma_i\rangle_{i\in I}$ is independent.

(d)(iv)⇒(ii) Now suppose that $\langle\Sigma_i\rangle_{i\in I}$ is independent. If $i_1,\dots,i_n$ are distinct members of I and $F_1,\dots,F_n$ are Borel sets in $\mathbb{R}$, then $X_{i_j}^{-1}[F_j]\in\Sigma_{i_j}$ for each j, so

$$\Pr(X_{i_1}\in F_1,\dots,X_{i_n}\in F_n)=\hat\mu(\bigcap_{1\le j\le n}X_{i_j}^{-1}[F_j])$$
$$=\prod_{i=1}^n\hat\mu X_{i_j}^{-1}[F_j]=\prod_{j=1}^n\Pr(X_{i_j}\in F_j)$$

.

(e) Finally, observe that (iii) is nothing but a re-formulation of (ii), because if $F_i=\mathbb{R}$ then $\Pr(X_i\in F_i)=1$ and $X_i^{-1}[F_i]\cup(\Omega\setminus\operatorname{dom}X_i)=\Omega$.

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.

proof Writing Σ_i for the σ-algebra defined by X_i, Σ'_i for the σ-algebra defined by $h(X_i)$, $h(X_i)$ is Σ_i-measurable (121Eg) so $\Sigma'_i\subseteq\Sigma_i$ for every i and $\langle\Sigma'_i\rangle_{i\in I}$ is independent, as in 272Bb.

272F Similarly, we can relate the definition in 272Aa to the others.

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.

proof (i)⇒(iii) If $i_1,\dots,i_n$ are distinct members of I and $\alpha_1,\dots,\alpha_n\in\mathbb{R}$, then for each $j\le n$ the set $G_j=\{\omega:X_{i_j}(\omega)\le\alpha_j\}$ is either E_{i_j} or $\emptyset$ or Ω. If any G_j is empty, then

$$\Pr(X_{i_j}\le\alpha_j\text{ for every}j\le n\}=0=\prod_{j=1}^n\Pr(X_{i_j}\le\alpha_j).$$

Otherwise, set $K=\{j:G_j=E_{i_j}\}$; then

$$\Pr(X_{i_j} \le \alpha_j \text{ for every} j \le n\} = \mu(\bigcap_{j\le n} G_j) = \mu(\bigcap_{j\in K} E_{i_j})$$

$$= \prod_{j\in K} \mu E_{i_j} = \prod_{j=1}^{n} \Pr(X_{i_j} \le \alpha_j).$$

As $i_1, \dots, i_n$ and $\alpha_1, \dots, \alpha_n$ are arbitrary, $\langle X_i\rangle_{i\in I}$ is independent.

(iii)⇒(ii) follows from (i)⇒(iii) of 272D, because Σ_i is the σ-algebra defined by X_i.

(ii)⇒(i) is trivial, because $E_i \in \Sigma_i$ for each i.

Remark You will I hope feel that while the theory of product measures might be appropriate to 272D, it is surely rather heavy machinery to use on what ought to be a simple combinatorial problem like (iii)⇒(ii) of this proposition. I suggest that you construct an 'elementary' proof, and examine which of the ideas of the theory of product measures (and the Monotone Class Theorem, 136B) are actually needed here.

272G Distributions of independent random variables I have not tried to describe the 'joint distribution' of an infinite family of random variables. (Indications of how to deal with a countable family are offered in 271Ya and 272Yb. For uncountable families I will wait until §454 in Volume 4.) As, however, the independence of a family of random variables is determined by the behaviour of finite subfamilies, we can approach it through the following proposition.

Theorem Let $\boldsymbol{X} = (X_1, \dots, X_n)$ 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) $X_1, \dots, X_n$ are independent;

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

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

proof (a)(i)⇒(iii) In the proof of (i)⇒(ii) of 272D above I showed that $\nu_{\boldsymbol{X}}$ is the product ν of $\nu_{X_1}, \dots, \nu_{X_n}$.

(b)(iii)⇒(ii) is trivial.

(c)(ii)⇒(i) Suppose that $\nu_{\boldsymbol{X}}$ is expressible as a product $\nu_1 \times \dots \times \nu_n$. Take $a = (\alpha_1, \dots, \alpha_n)$ in $\mathbb{R}^n$. Then

$$\Pr(X_i \le \alpha_i \ \forall\ i \le n) = \Pr(\boldsymbol{X} \in \left]-\infty, a\right]) = \nu_{\boldsymbol{X}}(\left]-\infty, a\right]) = \textstyle\prod_{i=1}^{n} \nu_i \left]-\infty, \alpha_i\right].$$

On the other hand, setting $F_i = \{(\xi_1, \dots, \xi_n) : \xi_i \le \alpha_i\}$, we must have

$$\nu_i \left]-\infty, \alpha_i\right] = \nu_{\boldsymbol{X}} F_i = \Pr(\boldsymbol{X} \in F_i) = \Pr(X_i \le \alpha_i)$$

for each i. So we get

$$\Pr(X_i \le \alpha_i \text{ for every } i \le n) = \textstyle\prod_{i=1}^{n} \Pr(X_i \le \alpha_i),$$

as required.

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.

proof For every distinct $i_1, \dots, i_n \in I$, if we set $\boldsymbol{X} = (X_{i_1}, \dots, X_{i_n})$ and $\boldsymbol{Y} = (Y_{i_1}, \dots, Y_{i_n})$, then $\boldsymbol{X} =_{\text{a.e.}} \boldsymbol{Y}$, so $\nu_{\boldsymbol{X}}$, $\nu_{\boldsymbol{Y}}$ are equal (271De). By 272G, $Y_{i_1}, \dots, Y_{i_n}$ must be independent because $X_{i_1}, \dots, X_{i_n}$ are. As $i_1, \dots, i_n$ are arbitrary, the whole family $\langle Y_i\rangle_{i\in I}$ is independent.

Remark It follows that we may speak of independent families in the space $L^0(\mu)$ of equivalence classes of random variables (241C), saying that $\langle X_i^\bullet\rangle_{i\in I}$ is independent iff $\langle X_i\rangle_{i\in I}$ is.

272I Corollary Suppose that $X_1, \dots, X_n$ are independent real-valued random variables with density functions $f_1, \dots, f_n$ (271H). Then $\boldsymbol{X} = (X_1, \dots, X_n)$ has a density function f given by setting $f(x) = \prod_{i=1}^{n} f_i(\xi_i)$ whenever $x = (\xi_1, \dots, \xi_n) \in \prod_{i\le n} \operatorname{dom}(f_i) \subseteq \mathbb{R}^n$.

proof For $n = 2$ this is covered by 253I; the general case follows by induction on n.

272J The most important theorems of the subject refer to independent families of random variables, rather than independent families of σ-algebras. The value of the concept of independent σ-algebras lies in such results as the following.

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.

proof This is virtually a restatement of 254Fb and 254G. (i) If ϕ is inverse-measure-preserving, $i_1, \dots , i_n \in I$ are distinct and $E_j \in \Sigma_{i_j}$ for each j, then $\bigcap_{j \le n} E_{i_j} = \phi^{-1}[\{x : x(i_j) \in E_j \text{ for every } j \le n\}]$, so that

$$\mu(\bigcap_{j \le n} E_{i_j}) = \lambda\{x : x(i_j) \in E_j \text{ for every } j \le n\} = \prod_{j=1}^n \mu_{i_j} E_{i_j} = \prod_{j=1}^n \mu E_{i_j}.$$

(ii) If $\langle \Sigma_i \rangle_{i \in I}$ is independent, $E_i \in \Sigma_i$ for every $i \in I$ and $\{i : E_i \ne \Omega\}$ is finite, then

$$\mu\phi^{-1}[\prod_{i \in I} E_i] = \mu(\bigcap_{i \in I} E_i) = \prod_{i \in I} \mu E_i = \prod_{i \in I} \mu_i E_i.$$

So the conditions of 254G are satisfied and $\mu\phi^{-1}[W] = \lambda W$ for every $W \in \Lambda$.

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.

proof Let $(\Omega, \hat{\Sigma}, \hat{\mu})$ be the completion of (Ω, Σ, μ). On Ω^I let λ be the product of the measures $\mu\restriction\Sigma_i$, and let $\phi : \Omega \to \Omega^I$ be the diagonal map, as in 272J. ϕ is inverse-measure-preserving for $\hat{\mu}$ and λ, by 272J.

We can identify λ with the product of $\langle \lambda_s \rangle_{s \in S}$, where for each $s \in S$ λ_s is the product of $\langle \mu\restriction\Sigma_i \rangle_{i \in J(s)}$ (254N). For $s \in S$, let Λ_s be the domain of λ_s, and set $\pi_s(x) = x\restriction J(s)$ for $x \in \Omega^I$, so that π_s is inverse-measure-preserving for λ and λ_s (254Oa), and $\phi_s = \pi_s\phi$ is inverse-measure-preserving for $\hat{\mu}$ and λ_s; of course ϕ_s is the diagonal map from Ω to $\Omega^{J(s)}$. Set $\Sigma_s^* = \{\phi_s^{-1}[H] : H \in \Lambda_s\}$. Then Σ_s^* is a σ-subalgebra of $\hat{\Sigma}$, and $\Sigma_s^* \supseteq \tilde{\Sigma}_s$, because

$$E = \phi_s^{-1}[\{x : x(i) \in E\}] \in \Sigma_s^*$$

whenever $i \in J(s)$a dn $E \in \Sigma_i$.

Now suppose that $s_1, \dots , s_n \in S$ are distinct and that $E_j \in \tilde{\Sigma}_{s_j}$ for each j. Then $E_j \in \Sigma_{s_j}^*$, so there are $H_j \in \Lambda_{s_j}$ such that $E_j = \phi_{s_j}^{-1}[H_j]$ for each j. Set

$$W = \{x : x \in \Omega^I,\ x\restriction J(s_j) \in H_j \text{ for every } j \le n\}.$$

Because we can identify λ with the product of the λ_s, we have

$$\lambda W = \prod_{j=1}^n \lambda_{s_j} H_j = \prod_{j=1}^n \hat{\mu}(\phi_{s_j}^{-1}[H_j]) = \prod_{j=1}^n \hat{\mu} E_j = \prod_{j=1}^n \mu E_j.$$

On the other hand, $\phi^{-1}[W] = \bigcap_{j \le n} E_j$, so, because ϕ is inverse-measure-preserving,

$$\mu(\bigcap_{j \le n} E_j) = \hat{\mu}(\bigcap_{j \le n} E_j) = \lambda W = \prod_{j=1}^n \mu E_j.$$

As $E_1, \dots , E_n$ are arbitrary, $\langle \tilde{\Sigma}_s \rangle_{s \in S}$ is independent.

272L I give a typical application of this result as a sample.

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.

proof Let Σ_X, Σ_{X_i} be the σ-algebras defined by X, X_i (272C). Then $\Sigma_X, \Sigma_{X_1}, \dots , \Sigma_{X_n}$ are independent (272D). Let Σ^* be the σ-algebra generated by $\Sigma_{X_1} \cup \ldots \cup \Sigma_{X_n}$. Then 272K (perhaps working in the completion of the original probability space) tells us that Σ_X and Σ^* are independent. But every X_j is Σ^*-measurable so $Y = h(X_1, \dots , X_n)$ is Σ^*-measurable (121Kb); also $\operatorname{dom} Y \in \Sigma^*$, so $\Sigma_Y \subseteq \Sigma^*$ and Σ_X, Σ_Y are independent. By 272D again, X and Y are independent, as claimed.

Remark Nearly all of us, when teaching elementary probability theory, would invite our students to treat this corollary (with an explicit function h, of course) as 'obvious'. In effect, the proof here is a confirmation that the formal definition of 'independence' offered is a faithful representation of our intuition of independent events having independent causes.

272M Products of probability spaces and independent families of random variables We have already seen that the concept of 'independent random variables' is intimately linked with that of 'product measure'. I now give some further manifestations of the connexion.

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}.

proof (a) It is easy to check that each $\tilde\Sigma_i$ is a σ-algebra of sets. The rest amounts just to recalling from 254Fb that if $J\subseteq I$ is finite and $E_i\in\Sigma_i$ for $i\in J$, then

$$\mu(\bigcap_{i\in J}\pi_i^{-1}[E_i])=\mu\{\omega:\omega(i)\in E_i\text{ for every }i\in I\}=\prod_{i\in I}\mu_iE_i$$

if we set $E_i=X_i$ for $i\in I\setminus J$.

(b) We know also that (Ω,Σ,μ) is the product of the completions $(\Omega_i,\hat\Sigma_i,\hat\mu_i)$ (254I). From this, we see that each $\tilde X_{ij}$ is defined μ-a.e., and is Σ-measurable, with the same distribution as X_{ij}. Now apply condition (iii) of 272D. Suppose that $\langle F_{ij}\rangle_{i\in I,j\in J(i)}$ is a family of Borel sets in $\mathbb{R}$, and that $\{(i,j):F_{ij}\ne\mathbb{R}\}$ is finite. Consider

$$E_i=\bigcap_{j\in J(i)}(X_{ij}^{-1}[F_{ij}]\cup(\Omega_i\setminus\operatorname{dom}X_{ij})),$$

$$E=\prod_{i\in I}E_i=\bigcap_{i\in I,j\in J(i)}(\tilde X_{ij}^{-1}[F_{ij}]\cup(\Omega\setminus\operatorname{dom}\tilde X_{ij})).$$

Because each family $\langle X_{ij}\rangle_{j\in J(i)}$ is independent, and $\{j:F_{ij}\ne\mathbb{R}\}$ is finite,

$$\hat\mu_iE_i=\prod_{j\in J(i)}\Pr(X_{ij}\in E_{ij})$$

for each $i\in I$. Because

$$\{i:E_i\ne\Omega_i\}\subseteq\{i:\exists\,j\in J(i),\ F_{ij}\ne\mathbb{R}\}$$

is finite,

$$\mu E=\prod_{i\in I}\hat\mu_iE_i=\prod_{i\in I,j\in J(i)}\Pr(\tilde X_{ij}\in F_{ij});$$

as $\langle F_{ij}\rangle_{i\in I,j\in J(i)}$ is arbitrary, $\langle\tilde X_{ij}\rangle_{i\in I,j\in J(i)}$ is independent.

Remark The formulation in (b) is more complicated than is necessary to express the idea, but is what is needed for an application below.

272N A special case of 272J is of particular importance in general measure theory, and is most useful in an adapted form.

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 $\mu E_i=\frac12$ 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 $\{0,1\}^I$ (254J).

proof I use 254G again. For each $i\in I$ let Σ_i be the algebra $\{\emptyset,E_i,\Omega\setminus E_i,\Omega\}$; then $\langle\Sigma_i\rangle_{i\in I}$ is independent (272F). For $i\in I$ set $\phi_i(\omega)=\phi(\omega)(i)$. Let ν be the usual measure of $\{0,1\}$. Then it is easy to check that

$$\mu\phi_i^{-1}[H]=\frac12\#(H)=\nu H$$

for every $H\subseteq\{0,1\}$. If $\langle H_i\rangle_{i\in I}$ is a family of subsets of $\{0,1\}$, and $\{i:H_i\ne\{0,1\}\}$ is finite, then

$$\mu\phi^{-1}[\bigcap_{i\in I} H_i] = \mu(\bigcap_{i\in I} \phi_i^{-1}[H_i]) = \prod_{i\in J} \mu\phi_i^{-1}[H_i]$$

(because $\phi^{-1}[H_i] \in \Sigma_i$ for each i, and $\langle\Sigma_i\rangle_{i\in I}$ is independent)

$$= \prod_{i\in I} \nu H_i = \lambda(\prod_{i\in I} H_i).$$

As $\langle H_i\rangle_{i\in I}$ is arbitrary, 254G gives the result.

272O Tail σ-algebras and the zero-one law I have never been able to make up my mind whether the following result is 'deep' or not. I think it is one of the many cases in mathematics where a theorem is surprising and exciting if one comes on it unprepared, but is natural and straightforward if one approaches it from the appropriate angle.

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 $E \in \Sigma_\infty^*$.

proof For each n, the family $(\Sigma_0, \dots, \Sigma_n, \Sigma_{n+1}^*)$ is independent, by 272K. So $(\Sigma_0, \dots, \Sigma_n, \Sigma_\infty^*)$ is independent, because $\Sigma_\infty^* \subseteq \Sigma_{n+1}^*$. But this means that every finite subfamily of $(\Sigma_\infty^*, \Sigma_0, \Sigma_1, \dots)$ is independent, and therefore that the whole family is (272Bb). Consequently $(\Sigma_\infty^*, \Sigma_0^*)$ must be independent, by 272K again.

Now if $E \in \Sigma_\infty^*$, then E also belongs to Σ_0^*, so we must have

$$\mu(E \cap E) = \mu E \cdot \mu E,$$

that is, $\mu E = (\mu E)^2$; so that $\mu E \in \{0, 1\}$, as claimed.

272P To support the claim that somewhere we have achieved a non-trivial insight, I give a corollary, which will be fundamental to the understanding of the limit theorems in the next section, and does not seem to be obvious.

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

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

is almost everywhere constant – that is, there is some $u \in [-\infty, \infty]$ such that

$$\limsup_{n\to\infty} \frac{1}{n+1}(X_0 + \dots + X_n) = u$$

almost everywhere.

proof We may suppose that each X_n is Σ-measurable and defined everywhere in Ω, because (as remarked in 272H) changing the X_n on a negligible set does not affect their independence, and it affects $\limsup_{n\to\infty} \frac{1}{n+1}(X_0 + \dots + X_n)$ only on a negligible set. For each n, let Σ_n be the σ-algebra generated by X_n (272C), and Σ_n^* the σ-algebra generated by $\bigcup_{m\ge n} \Sigma_m$; set $\Sigma_\infty^* = \bigcap_{n\in\mathbb{N}} \Sigma_n^*$. By 272D, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is independent, so $\mu E \in \{0, 1\}$ for every $E \in \Sigma_\infty^*$ (272O).

Now take any $a \in \mathbb{R}$ and set

$$E_a = \{\omega : \limsup_{m\to\infty} \frac{1}{m+1}(X_0(\omega) + \dots + X_m(\omega)) \le a\}.$$

Then

$$\limsup_{m\to\infty} \frac{1}{m+1}(X_0 + \dots + X_m) = \limsup_{m\to\infty} \frac{1}{m+1}(X_n + \dots + X_{m+n}),$$

so

$$E_a = \{\omega : \limsup_{m\to\infty} \frac{1}{m+1}(X_n(\omega) + \ldots + X_{n+m}(\omega)) \le a\}$$

belongs to Σ_n^* for every n, because X_i is Σ_n^*-measurable for every $i \ge n$. So $E \in \Sigma_\infty^*$ and

$$\Pr(\limsup_{m\to\infty} \frac{1}{m+1}(X_0 + \ldots + X_m) \le a) = \mu E_a$$

must be either 0 or 1. Setting

$$u = \sup\{a : a \in \mathbb{R},\ \mu E_a = 0\}$$

(allowing $\sup\emptyset = -\infty$ and $\sup\mathbb{R} = \infty$, as usual in such contexts), we see that

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

almost everywhere.

***272Q** I add here a result which will be useful in Volume 5 and which gives further insight into the nature of large independent families.

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 $F \in \mathrm{T}$, $j_1, \ldots, j_r$ are distinct members of J and $E_r \in \Sigma_{j_r}$ for each $r \le n$.

proof (a) As in 272J, give Ω^I the probability measure λ which is the product of the measures $\mu\restriction\Sigma_i$, and let $\phi : \Omega \to \Omega^I$ be the diagonal map, so that ϕ is inverse-measure-preserving for $\hat\mu$ and λ, where $\hat\mu$ is the completion of μ. Write Λ for the domain of λ. Set $\kappa = \max(\omega, \#(\mathcal{E}))$, and let $\mathcal{E}^*$ be the set $\{\bigcap_{r\le n} F_r : n \in \mathbb{N}$, $F_r \in \mathcal{E}$ for every $r \le n\}$. Because $\#(\mathcal{E}^n) \le \kappa$ for each n (2A1Lc), $\#(\mathcal{E}^*) \le \kappa$ (2A1Ld). For each $F \in \mathcal{E}^*$, define $\nu_F : \Lambda \to [0,1]$ by setting $\nu_F W = \hat\mu(F \cap \phi^{-1}[W])$; then ν_F is countably additive and dominated by λ. It therefore has a Radon-Nikodým derivative h_F with respect to λ, so that $\hat\mu(F \cap \phi^{-1}[W]) = \int_W h_F d\lambda$ for every $W \in \Lambda$ (232F). By 254Qc or 254Rb, we can find a function h'_F equal λ-almost everywhere to h_F and determined by coordinates in a countable set J_F, in the sense that $h'_F(w) = h'_F(w')$ whenever w, $w' \in \Omega^I$ and $w\restriction J_F = w'\restriction J_F$. (I am taking it for granted that we chose h'_F to be defined everywhere on Ω^I.)

(b) Set $J = I \setminus \bigcup_{F\in\mathcal{E}^*} J_F$; by 2A1Ld, $I \setminus J = \bigcup_{F\in\mathcal{E}^*} J_F$ has cardinal at most κ. If $F \in \mathcal{E}^*$, $j_1, \ldots, j_r$ are distinct members of J and $E_r \in \Sigma_{j_r}$ for each $r \le n$, $\mu(F \cap \bigcap_{r\le n} E_r) = \mu F \cdot \prod_{r=1}^n \mu E_r$. **P** Set $W = \{w : w \in \Omega^I,\ w(j_r) \in E_r$ for each $r \le n\}$. Then

$$\mu(F \cap \textstyle\bigcap_{r\le n} E_r) = \hat\mu(F \cap \phi^{-1}[W]) = \int_W h'_F d\lambda = \int h'_F \times \chi W\, d\lambda.$$

But observe that W is determined by coordinates in J, while h'_F is determined by coordinates in $J_F \subseteq I \setminus J$; putting 272Ma, 272K and 272R together (or otherwise), we have

$$\mu(F \cap \textstyle\bigcap_{r\le n} E_r) = \int h'_F \times \chi W\, d\lambda = \int h'_F d\lambda \cdot \lambda W = \mu F \cdot \prod_{r=1}^n \mu E_r. \quad \mathbf{Q}$$

(c) Now consider the family $\mathcal{A}$ of those sets $F \in \Sigma$ such that $\mu(F \cap \bigcap_{r\le n} E_r) = \mu F \cdot \prod_{r=1}^n \mu E_r$ whenever $j_1, \ldots, j_n \in J$ are distinct and $E_r \in \Sigma_{j_r}$ for every $r \le n$. It is easy to check that $\mathcal{A}$ is a Dynkin class, and we have just seen that $\mathcal{A}$ includes $\mathcal{E}^*$; as $\mathcal{E}^*$ is closed under $\cap$, $\mathcal{A}$ includes the σ-algebra T of sets generated by $\mathcal{E}^*$ (136B). And this is just what the theorem asserts.

272R I must now catch up on some basic facts from elementary probability theory.

Proposition Let X, Y be independent real-valued random variables with finite expectation (271Ab). Then $\mathbb{E}(X \times Y)$ exists and is equal to $\mathbb{E}(X)\mathbb{E}(Y)$.

proof Let $\nu_{(X,Y)}$ be the joint distribution of the pair (X, Y). Then $\nu_{(X,Y)}$ is the product of the distributions ν_X and ν_Y (272G). Also $\int x\nu_X(dx) = \mathbb{E}(X)$ and $\int y\nu_Y(dy) = \mathbb{E}(Y)$ exist in $\mathbb{R}$ (271F). So

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

(253D). But this is just $\mathbb{E}(X \times Y)$, by 271E with $h(x, y) = xy$.

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

proof (a) Suppose first that all the X_i have finite variance. Set $a_i=\mathbb{E}(X_i)$, $Y_i=X_i-a_i$, $X=X_1+\ldots+X_n$, $Y=Y_1+\ldots+Y_n$; then $\mathbb{E}(X)=a_1+\ldots+a_n$, so $Y=X-\mathbb{E}(X)$ and

$$\begin{aligned}\operatorname{Var}(X)&=\mathbb{E}(Y^2)=\mathbb{E}(\sum_{i=1}^n Y_i)^2\\&=\mathbb{E}(\sum_{i=1}^n\sum_{j=1}^n Y_i\times Y_j)=\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}(Y_i\times Y_j).\end{aligned}$$

Now observe that if $i\ne j$ then $\mathbb{E}(Y_i\times Y_j)=\mathbb{E}(Y_i)\mathbb{E}(Y_j)=0$, because Y_i and Y_j are independent (by 272E) and we may use 272R, while if $i=j$ then

$$\mathbb{E}(Y_i\times Y_j)=\mathbb{E}(Y_i^2)=\mathbb{E}(X_i-\mathbb{E}(X_i))^2=\operatorname{Var}(X_i).$$

So

$$\operatorname{Var}(X)=\textstyle\sum_{i=1}^n\mathbb{E}(Y_i^2)=\sum_{i=1}^n\operatorname{Var}(X_i).$$

(b)(i) I show next that if $\operatorname{Var}(X_1+X_2)<\infty$ then $\operatorname{Var}(X_1)<\infty$. **P** We have

$$\iint(x+y)^2\nu_{X_1}(dx)\nu_{X_2}(dy)=\int(x+y)^2\nu_{(X_1,X_2)}(d(x,y))$$

(by 272G and Fubini's theorem)

$$=\mathbb{E}((X_1+X_2)^2)$$

(by 271E)

$$<\infty.$$

So there must be some $a\in\mathbb{R}$ such that $\int(x+a)^2\mu_{X_1}(dx)$ is finite, that is, $\mathbb{E}((X_1+a)^2)<\infty$; consequently $\mathbb{E}(X_1^2)$ and $\operatorname{Var}(X_1)$ are finite. **Q**

(ii) Now an easy induction (relying on 272L!) shows that if $\operatorname{Var}(X_1+\ldots+X_n)$ is finite, so is $\operatorname{Var}X_j$ for every j. Turning this round, if $\sum_{j=1}^n\operatorname{Var}(X_j)=\infty$, then $\operatorname{Var}(X_1+\ldots+X_n)=\infty$, and again the two are equal.

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 $\nu_X*\nu_Y$ (257A).

proof Set $\nu=\nu_X*\nu_Y$. Take $a\in\mathbb{R}$ and set $h=\chi\,]-\infty,a]$. Then h is ν-integrable, so

$$\nu\,]-\infty,a]=\int h\,d\nu=\int h(x+y)(\nu_X\times\nu_Y)(d(x,y))$$

(by 257B, writing $\nu_X\times\nu_Y$ for the product measure on $\mathbb{R}^2$)

$$=\int h(x+y)\nu_{(X,Y)}(d(x,y))$$

(by 272G, writing $\nu_{(X,Y)}$ for the joint distribution of (X,Y); this is where we use the hypothesis that X and Y are independent)

$$=\mathbb{E}(h(X+Y))$$

(applying 271E to the function $(x,y)\mapsto h(x+y)$)

$$=\Pr(X+Y\le a).$$

As a is arbitrary, $\nu_X*\nu_Y$ is the distribution of $X+Y$.

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 $X+Y$.

proof By 257F, $f*g$ is a density function for $\nu_X*\nu_Y=\nu_{X+Y}$.

272V The following simple result will be very useful when we come to stochastic processes in Volume 4, as well as in the next section.

Etemadi's lemma (Etemadi 96) 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 $\gamma>0$.

proof As in 272P, we may suppose that every X_i is a measurable function defined everywhere on a measure space Ω. Set $\alpha=\max_{m\le n}\Pr(|S_m|\ge\gamma)$. For each $r\le n$, set

$$E_r=\{\omega:|S_m(\omega)|<3\gamma\text{ for every }m<r,\ |S_r(\omega)|\ge 3\gamma\}.$$

Then $E_0,\dots,E_n$ is a partition of $\{\omega:\max_{m\le n}|S_m(\omega)|\ge 3\gamma\}$. Set $E'_r=\{\omega:\omega\in E_r,\ |S_n(\omega)|<\gamma\}$. Then $E'_r\subseteq\{\omega:\omega\in E_r,\ |(S_n-S_r)(\omega)|>2\gamma\}$. But E_r depends on $X_0,\dots,X_r$ so is independent of $\{\omega:|(S_n-S_r)(\omega)|>2\gamma\}$, which can be calculated from $X_{r+1},\dots,X_n$ (272K). So

$$\begin{aligned}\mu E'_r&\le\mu\{\omega:\omega\in E_r,\ |(S_n-S_r)(\omega)|>2\gamma\}=\mu E_r\cdot\Pr(|S_n-S_r|>2\gamma)\\&\le\mu E_r(\Pr(|S_n|>\gamma)+\Pr(|S_r|>\gamma))\le 2\alpha\mu E_r,\end{aligned}$$

and $\mu(E_r\setminus E'_r)\ge(1-2\alpha)\mu E_r$. On the other hand, $\langle E_r\setminus E'_r\rangle_{r\le n}$ is a disjoint family of sets all included in $\{\omega:|S_n(\omega)|\ge\gamma\}$. So

$$\alpha\ge\mu\{\omega:|S_n(\omega)|\ge\gamma\}\ge\sum_{r=0}^n\mu(E_r\setminus E'_r)\ge(1-2\alpha)\sum_{r=0}^n\mu E_r,$$

and

$$\Pr(\sup_{r\le n}|S_r|\ge 3\gamma)=\sum_{r=0}^n\mu E_r\le\min(1,\tfrac{\alpha}{1-2\alpha})\le 3\alpha,$$

(considering $\alpha\le\frac13$, $\alpha\ge\frac13$ separately), as required.

***272W** The next result is a similarly direct application of the ideas of this section. While it will not be used in this volume, it is an accessible and useful representative of a very large number of results on tails of sums of independent random variables.

Theorem (Hoeffding 63) 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 $c\ge 0$.

proof (a) Set $a_i=\mathbb{E}(X_i)$ for each i. If $b\ge 0$ and $i\le n$, then

$$\mathbb{E}(e^{bX_i})\le\exp(ba_i+\frac18 b^2).$$

P Set $\phi(x)=e^{bx}$ for $x\in\mathbb{R}$. Then ϕ is convex, so

$$\phi(x)\le 1+x(e^b-1)$$

whenever $x\in[0,1]$,

$$\phi(X_i)\le_{\text{a.e.}}1+(e^b-1)X_i$$

and

$$\mathbb{E}(e^{bX_i})=\mathbb{E}(\phi(X_i))\le 1+(e^b-1)a_i=e^{h(b)}$$

where $h(t)=\ln(1-a_i+a_ie^t)$ for $t\in\mathbb{R}$. Now $h(0)=0$,

$$h'(t) = \frac{a_i e^t}{1-a_i+a_i e^t} = 1 - \frac{1-a_i}{1-a_i+a_i e^t}, \quad h'(0) = a_i,$$

$$h''(t) = \frac{1-a_i}{1-a_i+a_i e^t} \cdot \frac{a_i e^t}{1-a_i+a_i e^t} \leq \frac{1}{4}$$

because $a_i e^t$ and $1 - a_i$ are both greater than or equal to 0. By Taylor's theorem with remainder, there is some $t \in [0, b]$ such that

$$h(b) = h(0) + bh'(0) + \frac{1}{2}b^2 h_i''(t) \leq ba_i + \frac{1}{8}b^2,$$

and

$$\mathbb{E}(e^{bX_i}) \leq \exp(ba_i + \frac{1}{8}b^2). \quad \mathbf{Q}$$

(b) Take any $b \geq 0$. Then

$$\Pr(S - a \geq c) = \Pr(\sum_{i=0}^{n}(X_i - a_i - c) \geq 0) \leq \mathbb{E}(\exp(b\sum_{i=0}^{n} X_i - a_i - c))$$

(because $\exp(b\sum_{i=0}^{n} X_i - a_i - c) \geq 1$ whenever $\sum_{i=0}^{n} X_i - a_i - c \geq 0$)

$$= e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \mathbb{E}(\prod_{i=0}^{n} \exp(bX_i))$$

$$= e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \prod_{i=0}^{n} \mathbb{E}(\exp(bX_i))$$

(because the random variables $\exp(bX_i)$ are independent, by 272E, so the expectation of the product is the product of the expectations, by 272R)

$$\leq e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \exp(ba_i + \frac{1}{8}b^2)$$

((a) above)

$$= \exp(-(n+1)bc + \frac{n+1}{8}b^2).$$

Now the minimum value of the quadratic $\frac{n+1}{8}b^2 - (n+1)cb$ is $-2(n+1)c^2$ when $b = 4c$, so $\Pr(S - a \geq c) \leq \exp(-2(n+1)c^2)$, as claimed.

272X Basic exercises (a) Let (Ω, Σ, μ) be an atomless probability space, and $\langle \epsilon_n \rangle_{n\in\mathbb{N}}$ any sequence in $[0, 1]$. Show that there is an independent sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ such that $\mu E_n = \epsilon_n$ for every n. (*Hint*: 215D.)

>(b) Let $\langle X_i \rangle_{i\in I}$ be a family of real-valued random variables. Show that it is independent iff

$$\mathbb{E}(h_1(X_{i_1}) \times \ldots \times h_n(X_{i_n})) = \prod_{j=1}^{n} \mathbb{E}(h_j(X_{i_j}))$$

whenever $i_1, \ldots, i_n$ are distinct members of I and $h_1, \ldots, h_n$ are Borel measurable functions from $\mathbb{R}$ to $\mathbb{R}$ such that $\mathbb{E}(h_j(X_{i_j}))$ are all finite.

(c) Write out a proof of 272F which does not use the theory of product measures.

(d) Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a family of real-valued random variables all defined on the same probability space, and suppose that $\boldsymbol{X}$ has a density function f expressible in the form $f(\xi_1, \ldots, \xi_n) = f_1(\xi_1)f_2(\xi_2)\ldots f_n(\xi_n)$ for suitable functions $f_1, \ldots, f_n$ of one real variable. Show that $X_1, \ldots, X_n$ are independent.

(e) Let X_1, X_2 be independent real-valued random variables both with distribution ν and distribution function F. Set $Y = \max(X_1, X_2)$. Show that the distribution of Y is absolutely continuous with respect to ν, with a Radon-Nikodým derivative $F + F^-$, where $F^-(x) = \lim_{t\uparrow x} F(t)$ for every $x \in \mathbb{R}$.

(f) Use 254Sa and the idea of 272J to give another proof of 272O.

(g) 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 Σ_∞ be the σ-algebra generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$. Let T be another σ-subalgebra of Σ such that Σ_n and T are independent for each n. Show that Σ_∞ and T are independent. (*Hint*: apply the Monotone Class Theorem to $\{E : \mu(E \cap F) = \mu E \cdot \mu F$ for every $F \in \mathrm{T}\}$.) Use this to prove 272O.

(h) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables and Y a real-valued random variable such that Y and X_n are independent for each $n \in \mathbb{N}$. Suppose that $\Pr(Y \in \mathbb{N}) = 1$ and that $\sum_{n=0}^\infty \Pr(Y \ge n)\mathbb{E}(|X_n|)$ is finite. Set $Z = \sum_{n=0}^Y X_n$ (that is, $Z(\omega) = \sum_{n=0}^{Y(\omega)} X_n(\omega)$ whenever $\omega \in \operatorname{dom} Y$ is such that $Y(\omega) \in \mathbb{N}$ and $\omega \in \operatorname{dom} X_n$ for every $n \le Y(\omega)$). (i) Show that $\mathbb{E}(Z) = \sum_{n=0}^\infty \Pr(Y \ge n)\mathbb{E}(X_n)$. (*Hint*: set $X'_n(\omega) = X_n(\omega)$ if $Y(\omega) \ge n$, 0 otherwise.) (ii) Show that if $\mathbb{E}(X_n) = \gamma$ for every $n \in \mathbb{N}$ then $\mathbb{E}(Z) = \gamma\mathbb{E}(Y)$. (This is **Wald's equation**.)

>(i) Let $X_1, \dots, X_n$ be independent real-valued random variables. Show that if $X_1 + \ldots + X_n$ has finite expectation so does every X_j. (*Hint*: part (b) of the proof of 272S.)

>(j) Let X and Y be independent real-valued random variables with densities f and g. Show that $X \times Y$ has a density function h where $h(x) = \int_{-\infty}^\infty \frac{1}{|y|} g(y) f(\frac{x}{y}) dy$ for almost every x. (*Hint*: 271K.)

(k) Let $X_0, \dots, X_n$ be independent real-valued random variables such that $d_i \le X_i \le d'_i$ a.e. for every i. (i) Show that if $b \ge 0$ then $\mathbb{E}(e^{bX_i}) \le \exp(ba_i + \frac{1}{8} b^2 (d'_i - d_i)^2)$ for each i, where $a_i = \mathbb{E}(X_i)$. (ii) Set $S = \frac{1}{n+1} \sum_{i=0}^n X_i$ and $a = \mathbb{E}(S)$. Show that

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

for every $c \ge 0$, where $d = \sum_{i=0}^n (d'_i - d_i)^2$.

(l) Suppose that $X_0, \dots, X_n$ are independent real-valued random variables, all with expectation 0, such that $\Pr(|X_i| \le 1) = 1$ for every i. Set $S = \frac{1}{\sqrt{n+1}} \sum_{i=0}^n X_i$. Show that $\Pr(S \ge c) \le \exp(-c^2/2)$ for every $c \ge 0$.

272Y Further exercises (a) Let $X_0, \dots, X_n$ be independent real-valued random variables with distributions $\nu_0, \dots, \nu_n$ and distribution functions $F_0, \dots, F_n$. Show that, for any Borel set $E \subseteq \mathbb{R}$,

$$\Pr(\sup_{i\le n} X_i \in E) = \sum_{i=0}^n \int_E \prod_{j=0}^{i-1} F_j^-(x) \prod_{j=i+1}^n F_j(x) \nu_i(dx),$$

where $F_j^-(x) = \lim_{t\uparrow x} F_j(t)$ for each j, and we interpret the empty products $\prod_{j=0}^{-1} F_j^-(x)$, $\prod_{j=n+1}^n F_j(x)$ as 1.

(b) Let $\boldsymbol{X} = \langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables on a complete probability space (Ω, Σ, μ). Let $\mathcal{B}$ be the Borel σ-algebra of $\mathbb{R}^\mathbb{N}$ (271Ya). Let $\nu_{\boldsymbol{X}}^{(\mathcal{B})}$ be the probability measure with domain $\mathcal{B}$ defined by setting $\nu_{\boldsymbol{X}}^{(\mathcal{B})} E = \mu \boldsymbol{X}^{-1}[E]$ for every $E \in \mathcal{B}$, and write $\nu_{\boldsymbol{X}}$ for the completion of $\nu_{\boldsymbol{X}}^{(\mathcal{B})}$. Show that $\nu_{\boldsymbol{X}}$ is just the product of the distributions ν_{X_n}.

(c) Let $X_1, \dots, X_n$ be real-valued random variables such that for each $j < n$ the family

$$(X_1, \dots, X_j, -X_{j+1}, \dots, -X_n)$$

has the same joint distribution as the original family $(X_1, \dots, X_n)$. Set $S_j = X_1 + \ldots + X_j$ for each $j \le n$. (i) Show that for any $a \ge 0$

$$\Pr(\sup_{1\le j\le n}|S_j|\ge a)\le 2\Pr(|S_n|\ge a).$$

(*Hint*: show that if $E_j=\{\omega:\omega\in\bigcap_{i\le n}\operatorname{dom}X_i,\ |S_i(\omega)|<a \text{ for } i<j,\ |S_j(\omega)|\ge a\}$ then $\mu\{\omega:\omega\in E_j,\ |S_n(\omega)|\ge|S_j(\omega)|\}\ge\frac{1}{2}\mu E_j$.) (ii) Show that $\mathbb{E}(\sup_{j\le n}|S_j|)\le 2\mathbb{E}(|S_n|)$. (iii) Show that $\mathbb{E}(\sup_{i\le n}S_i^2)\le 2\mathbb{E}(S_n^2)$.

(d) Let $\langle X_i\rangle_{i\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and set $S_n=\sum_{i=0}^n X_i$ for each n. Show that if $\langle S_n\rangle_{n\in\mathbb{N}}$ converges to S in $\mathcal{L}^0$ for the topology of convergence in measure, then $\langle S_n\rangle_{n\in\mathbb{N}}$ converges to S a.e.

(e) Let (Ω,Σ,μ) be a probability space.

(i) Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence in Σ. Show that for any real-valued random variable X with finite expectation,

$$\lim_{n\to\infty}\int_{E_n}X\,d\mu-\mu E_n\mathbb{E}(X)=0.$$

(*Hint*: let T_0 be the subalgebra of Σ generated by $\{E_n:n\in\mathbb{N}\}$ and T the σ-subalgebra of Σ generated by $\{E_n:n\in\mathbb{N}\}$. Start by considering $X=\chi E$ for $E\in\mathrm{T}_0$ and then $X=\chi E$ for $E\in\mathrm{T}$. Move from $\mathcal{L}^1(\mu\restriction\mathrm{T})$ to $\mathcal{L}^1(\mu)$ by using conditional expectations.)

(ii) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable independent sequence of real-valued random variables on Ω. Show that for any bounded real-valued random variable Y,

$$\lim_{n\to\infty}\mathbb{E}(X_n\times Y)-\mathbb{E}(X_n)\mathbb{E}(Y)=0.$$

(iii) Suppose that $1<p\le\infty$ and set $q=p/(p-1)$ (taking $q=1$ if $p=\infty$). Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with $\sup_{n\in\mathbb{N}}\|X_n\|_p<\infty$, and Y a real-valued random variable with $\|Y\|_q<\infty$. Show that

$$\lim_{n\to\infty}\mathbb{E}(X_n\times Y)-\mathbb{E}(X_n)\mathbb{E}(Y)=0.$$

(f) Let (Ω,Σ,μ) be a probability space and $\langle Z_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\Pr(Z_n\in\mathbb{N})=1$ for each n, and $\Pr(Z_m=Z_n)=0$ for all $m\ne n$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables on Ω, all with the same distribution ν, and independent of each other and the Z_n, in the sense that if Σ_n is the σ-algebra defined by X_n, and T_n the σ-algebra defined by Z_n, and T is the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\mathrm{T}_n$, then $(\mathrm{T},\Sigma_0,\Sigma_1,\dots)$ is independent. Set $Y_n(\omega)=X_{Z_n(\omega)}(\omega)$ whenever this is defined, that is, $\omega\in\operatorname{dom}Z_n$, $Z_n(\omega)\in\mathbb{N}$ and $\omega\in\operatorname{dom}X_{Z_n(\omega)}$. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of random variables and that every Y_n has the distribution ν.

(g) Show that all the ideas of this section apply equally to complex-valued random variables, subject to suitable adjustments (to be devised).

(h) Develop a theory of independence for random variables taking values in $\mathbb{R}^r$, following through as many as possible of the ideas of this section.

272 Notes and comments This section is lengthy for two reasons: I am trying to pack in the basic results associated with one of the most fertile concepts of mathematics, and it is hard to know where to stop; and I am trying to do this in language appropriate to abstract measure theory, insisting on a variety of distinctions which are peripheral to the essential ideas. For while I am prepared to be flexible on the question of whether the letter X should denote a space or a function, some of the applications of these results which are most important to me are in contexts where we expect to be exactly clear what the domains of our functions are. Consequently it is necessary to form an opinion on such matters as what the σ-algebra defined by a random variable really is (272C).

The point of 272Q is that the family $\mathcal{E}$ does not have to be related in any way to the family $\langle\Sigma_i\rangle_{i\in I}$, except, of course, that we are dealing with measurable sets. All we need to know is that I should be large compared with $\mathcal{E}$; for instance, that $\mathcal{E}$ is countable and I is uncountable. The family $\langle\Sigma_j\rangle_{j\in J}$ is now a kind of 'tail' of $\langle\Sigma_i\rangle_{i\in I}$, safely independent of the 'head' σ-algebra generated by $\mathcal{E}$.

Of course I should emphasize again that such proofs as those in 272R-272S are to be thought of as confirmations that we have a suitable model of probability theory, rather than as reasons for believing the results to be valid in statistical contexts. Similarly, 272T-272U can be approached by a variety of intuitions concerning discrete random variables and random variables with continuous densities, and while the elegant general results are delightful, they are more important to the pure mathematician than to the statistician. But I came to an odd obstacle in the proof of 272S, when showing that if $X_1 + \ldots + X_n$ has finite variance then so does every X_j. We have done enough measure theory for this to be readily dealt with, but the connexion with ordinary probabilistic intuition, both here and in 272Xi, remains unclear to me.

There are four ideas in 272W worth storing for future use. The first is the estimate

$$\mathbb{E}(e^{bX_i}) \leq 1 - a_i + e^b a_i$$

in part (a), a crude but effective way of using the hypothesis that X_i is bounded. The second is the use of Taylor's theorem to show that $1 - a_i + e^b a_i \leq \exp(a_i + \frac{1}{8}b^2)$. The third is the estimate

$$\Pr(Y \geq 0) \leq \mathbb{E}(e^{bY}) \text{ if } b \geq 0$$

used in part (b); and the fourth is 272R. After this one need only be sufficiently determined to reach 272Xk. But even the special case 272W is both striking and useful.

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 It will be helpful to start with an explicit statement of a very simple but very useful lemma.

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 $\omega \in \Omega$.

proof We have

$$\begin{aligned}\mu\{\omega : \{n : \omega \in E_n\} \text{ is infinite}\} &= \mu(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} E_m) = \inf_{n\in\mathbb{N}} \mu(\bigcup_{m\geq n} E_m) \\ &\leq \inf_{n\in\mathbb{N}} \sum_{m=n}^{\infty} \mu E_m = 0.\end{aligned}$$

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.

proof (a) Let (Ω, Σ, μ) be the underlying probability space. If we change each X_n on a negligible set, we do not change the independence of $\langle X_n \rangle_{n\in\mathbb{N}}$ (272H), and the S_n are also changed only on a negligible set; so we may suppose from the beginning that every X_n is a measurable function defined on the whole of Ω.

Because the functional $X \mapsto \mathbb{E}(\min(1, |X|))$ is one of the pseudometrics defining the topology of convergence in measure (245A), $\lim_{m,n\to\infty} \mathbb{E}(\min(1, |S_m - S_n|)) = 0$, and we can find for each $k \in \mathbb{N}$ an $n_k \in \mathbb{N}$ such that $\mathbb{E}(\min(1, |S_m - S_{n_k}|)) \leq 4^{-k}$ for every $m \geq n_k$. So $\Pr(|S_m - S_{n_k}| \geq 2^{-k}) \leq 2^{-k}$ for every $m \geq n_k$. By Etemadi's lemma (272V) applied to $\langle X_i \rangle_{i\geq n_k}$,

$$\Pr(\sup_{n_k\leq m\leq n} |S_m - S_{n_k}| \geq 3 \cdot 2^{-k}) \leq 3 \cdot 2^{-k}$$

for every $n \geq n_k$. Setting

$$H_{kn} = \{\omega : \sup_{n_k\leq m\leq n} |S_m(\omega) - S_{n_k}(\omega)| \geq 3 \cdot 2^{-k}\} \text{ for } n \geq n_k,$$

$$H_k = \bigcup_{n\ge n_k} H_{kn},$$

we have

$$\mu H_k = \lim_{n\to\infty} \mu H_{kn} \le 3\cdot 2^{-k}$$

for each k, so $\sum_{k=0}^{\infty} \mu H_k$ is finite and almost every $\omega\in\Omega$ belongs to only finitely many of the H_k (273A).

Now take any such ω. Then there is some $r\in\mathbb{N}$ such that $\omega\notin H_k$ for any $k\ge r$. In this case, for every $k\ge r$, $\omega\notin\bigcup_{n\ge n_k} H_{kn}$, that is, $|S_n(\omega)-S_{n_k}(\omega)|<3\cdot 2^{-k}$ for every $n\ge n_k$. But this means that $\langle S_n(\omega)\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence, therefore convergent. Since this is true for almost every ω, $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere, as claimed.

(b) Now suppose that $\mathbb{E}(X_n)=0$ for every n and that $\sum_{n=0}^{\infty}\mathbb{E}(X_n^2)<\infty$. In this case, for any $m<n$,

$$\|S_n - S_m\|_1^2 \le \|\chi\Omega\|_2^2\|S_n - S_m\|_2^2$$

(by Cauchy's inequality, 244Eb)

$$= \mathbb{E}(S_n - S_m)^2 = \mathrm{Var}(S_n - S_m)$$

(because $\mathbb{E}(S_n - S_m) = \sum_{i=m+1}^{n}\mathbb{E}(X_i) = 0$)

$$= \sum_{i=m+1}^{n} \mathrm{Var}(X_i)$$

(by Bienaymé's equality, 272S)

$$\to 0$$

as $m\to\infty$. So $\langle S_n^\bullet\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^1(\mu)$ and converges in $L^1(\mu)$, by 242F; by 245G, it converges in measure in $L^0(\mu)$, that is, $\langle S_n\rangle_{n\in\mathbb{N}}$ converges in measure in $\mathcal{L}^0(\mu)$. By (a), $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere, that is, $\sum_{i=0}^{\infty} X_i$ is defined and finite almost everywhere.

Remark The proof above assumes familiarity with the ideas of Chapter 24. However part (b), at least, can be established without any of these; see 273Xa. In 276B there is a generalization of (b) based on a different approach.

273C We now need a lemma (part (b) below) from the theory of summability. I take the opportunity to include an elementary fact which will be useful later in this section and elsewhere.

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 $\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=0}^{n} b_k x_k = 0$.

proof (a) Let $\epsilon>0$. Let m be such that $|x_n - x|\le\epsilon$ whenever $n\ge m$. Let $m'\ge m$ be such that $|\sum_{i=0}^{m-1} x - x_i|\le\epsilon m'$. Then for $n\ge m'$ we have

$$\begin{aligned}|x - \frac{1}{n+1}\sum_{i=0}^{n} x_i| &= \frac{1}{n+1}|\sum_{i=0}^{n} x - x_i|\\ &\le \frac{1}{n+1}|\sum_{i=0}^{m-1} x - x_i| + \frac{1}{n+1}\sum_{i=m}^{n}|x - x_i|\\ &\le \frac{\epsilon m'}{n+1} + \frac{\epsilon(n-m+1)}{n+1} \le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n} x_i = x$.

(b) Let $\epsilon>0$. Write $s_n = \sum_{i=0}^{n} x_i$ for each n, and

$$s = \lim_{n\to\infty} s_n = \sum_{i=0}^{\infty} x_i;$$

set $s^* = \sup_{n\in\mathbb{N}} |s_n| < \infty$. Let $m \in \mathbb{N}$ be such that $|s_n - s| \le \epsilon$ whenever $n \ge m$; then $|s_n - s_j| \le 2\epsilon$ whenever j, $n \ge m$. Let $m' \ge m$ be such that $b_m s^* \le \epsilon b_{m'}$.

Take any $n \ge m'$. Then

$$\begin{aligned}
|\sum_{k=0}^{n} b_k x_k| &= |b_0 s_0 + b_1(s_1 - s_0) + \ldots + b_n(s_n - s_{n-1})| \\
&= |(b_0 - b_1)s_0 + (b_1 - b_2)s_1 + \ldots + (b_{n-1} - b_n)s_{n-1} + b_n s_n| \\
&= |b_0 s_n + \sum_{i=0}^{n-1}(b_{i+1} - b_i)(s_n - s_i)| \\
&\le b_0|s_n| + \sum_{i=0}^{m-1}(b_{i+1} - b_i)|s_n - s_i| + \sum_{i=m}^{n-1}(b_{i+1} - b_i)|s_n - s_i| \\
&\le b_0 s^* + 2s^* \sum_{i=0}^{m-1}(b_{i+1} - b_i) + 2\epsilon \sum_{i=m}^{n-1}(b_{i+1} - b_i) \\
&= b_0 s^* + 2s^*(b_m - b_0) + 2\epsilon(b_n - b_m) \le 2s^* b_m + 2\epsilon b_n.
\end{aligned}$$

Consequently, because $b_n \ge b_{m'}$,

$$|\frac{1}{b_n}\sum_{k=0}^{n} b_k x_k| \le 2\frac{s^* b_m}{b_n} + 2\epsilon \le 4\epsilon.$$

As ϵ is arbitrary,

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{k=0}^{n} b_k x_k = 0,$$

as required.

Remark Part (b) above is sometimes called 'Kronecker's lemma'.

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.

proof As usual, write (Ω, Σ, μ) for the underlying probability space. Set

$$Y_n = \frac{1}{b_n}(X_n - \mathbb{E}(X_n))$$

for each n; then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is independent (272E), $\mathbb{E}(Y_n) = 0$ for each n, and

$$\sum_{n=0}^{\infty} \mathbb{E}(Y_n^2) = \sum_{n=0}^{\infty} \frac{1}{b_n^2} \operatorname{Var}(X_n) < \infty.$$

By 273B, $\langle Y_n(\omega)\rangle_{n\in\mathbb{N}}$ is summable for almost every $\omega \in \Omega$. But by 273Cb,

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

for all such ω. So we have the result.

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.

Remark For most of the rest of this section, we shall take $b_n = n+1$. The special virtue of 273D is that it allows other b_n, e.g., $b_n = \sqrt{n}\ln n$. A direct strengthening of this theorem is in 276C below.

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 $\omega \in \Omega$.

proof In 273D, set $X_n = \chi E_n$, $b_n = n+1$. For almost every ω, we have

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n}(\chi E_i(\omega) - a_i) = 0,$$

writing $a_i = \mu E_i = \mathbb{E}(X_i)$ for each i. (I see that I am relying on 272F to support the claim that $\langle X_n \rangle_{n\in\mathbb{N}}$ is independent.) But for any such ω,

$$\lim_{n\to\infty} \Bigl(\frac{1}{n+1}\#(\{i : i \le n,\, \omega \in E_i\}) - \frac{1}{n+1}\sum_{i=0}^{n} a_i\Bigr)$$
$$= \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n}(\chi E_i(\omega) - a_i) = 0;$$

because we are supposing that $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} a_i = c$, we must have

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

as required.

273G Corollary Let μ be the usual measure on $\mathcal{P}\mathbb{N}$, as described in 254Jb. Then for μ-almost every set $a \subseteq \mathbb{N}$,

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

proof The sets $E_n = \{a : n \in a\}$ are independent, with measure $\frac{1}{2}$.

Remark The limit $\lim_{n\to\infty} \frac{1}{n+1}\#(a \cap \{0, \ldots, n\})$, when it is defined, is called the **asymptotic density** of a.

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.

proof As usual, call the underlying probability space (Ω,Σ,μ); as in 273B we can adjust the X_n on negligible sets so as to make them measurable and defined everywhere on Ω, without changing $\mathbb{E}(X_n)$, $\mathbb{E}(|X_n|)$ or the convergence of the averages except on a negligible set.

(a) For each n, define a random variable Y_n on Ω by setting

$$\begin{aligned}Y_n(\omega) &= X_n(\omega) \text{ if } |X_n(\omega)| \le n,\\ &= 0 \text{ if } |X_n(\omega)| > n.\end{aligned}$$

Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is independent (272E). For each $n \in \mathbb{N}$,

$$\operatorname{Var}(Y_n) \le \mathbb{E}(Y_n^2) \le \mathbb{E}(n^{1-\delta}|X_n|^{1+\delta}) \le n^{1-\delta}K,$$

where $K = \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{1+\delta})$, so

$$\sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \operatorname{Var}(Y_n) \le \sum_{n=0}^{\infty} \frac{n^{1-\delta}}{(n+1)^2} K < \infty.$$

By 273D,

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

is conegligible.

(b) On the other hand, setting

$$E_n = \{\omega : Y_n(\omega) \ne X_n(\omega)\} = \{\omega : |X_n(\omega)| > n\},$$

we have $K \ge n^{1+\delta}\mu E_n$ for each n, so

$$\sum_{n=0}^{\infty} \mu E_n \le 1 + K\sum_{n=1}^{\infty} \frac{1}{n^{1+\delta}} < \infty,$$

and the set $H = \{\omega : \{n : \omega \in E_n\} \text{ is finite}\}$ is conegligible (273A). But of course

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

for every $\omega \in H$.

(c) Finally,

$$|\mathbb{E}(Y_n) - \mathbb{E}(X_n)| \le \int_{E_n} |X_n| \le \int_{E_n} n^{-\delta}|X_n|^{1+\delta} \le n^{-\delta}K$$

whenever $n \ge 1$, so $\lim_{n\to\infty} \mathbb{E}(Y_n) - \mathbb{E}(X_n) = 0$ and

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

(273Ca). Putting these three together, we get

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

whenever ω belongs to the conegligible set $G \cap H$. So

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

almost everywhere, as required.

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.

proof The proof follows the same line as that of 273H, but some of the inequalities require more delicate arguments. As usual, call the underlying probability space (Ω, Σ, μ) and suppose that the X_n are all measurable and defined everywhere on Ω. (We need to remember that changing a random variable on a negligible set does not change its distribution.) Let ν be the common distribution of the X_n.

(a) For each n, define a random variable Y_n on Ω by setting

$$\begin{aligned} Y_n(\omega) &= X_n(\omega) \text{ if } |X_n(\omega)| \le n, \\ &= 0 \text{ if } |X_n(\omega)| > n. \end{aligned}$$

Then $\langle Y_n \rangle_{n\in\mathbb{N}}$ is independent (272E). For each $n \in \mathbb{N}$,

$$\mathrm{Var}(Y_n) \le \mathbb{E}(Y_n^2) = \int_{[-n,n]} x^2 \nu(dx)$$

(271E). To estimate $\sum_{n=0}^\infty \frac{1}{(n+1)^2}\mathbb{E}(Y_n^2)$, set

$$f_n(x) = \frac{x^2}{(n+1)^2} \text{ if } |x| \le n, \ 0 \text{ if } |x| > n,$$

so that $\frac{1}{(n+1)^2}\mathrm{Var}(Y_n) \le \int f_n d\nu$. If $r \ge 1$ and $r < |x| \le r+1$ then

$$\begin{aligned} \sum_{n=0}^\infty f_n(x) &\le \sum_{n=r+1}^\infty \frac{1}{(n+1)^2}(r+1)|x| \\ &\le (r+1)|x| \sum_{n=r+1}^\infty \left(\frac{1}{n} - \frac{1}{n+1}\right) \le |x|, \end{aligned}$$

while if $|x| \le 1$ then

$$\sum_{n=0}^\infty f_n(x) \le \sum_{n=0}^\infty \frac{1}{(n+1)^2} = \frac{\pi^2}{6} \le 2 < \infty.$$

(You do not need to know that the sum is $\frac{\pi^2}{6}$, only that it is finite; but see 282Xo.) Consequently

$$f(x) = \sum_{n=0}^\infty f_n(x) \le 2 + |x|$$

for every x, and $\int f d\nu < \infty$, because $\int |x|\nu(dx)$ is the common value of $\mathbb{E}(|X_n|)$, and is finite. By any of the great convergence theorems,

$$\sum_{n=0}^\infty \frac{1}{(n+1)^2}\mathrm{Var}(Y_n) \le \sum_{n=0}^\infty \int f_n d\nu = \int f d\nu < \infty.$$

By 273D,

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

is conegligible.

(b) Next, setting

$$E_n = \{\omega : X_n(\omega) \ne Y_n(\omega)\} = \{\omega : |X_n(\omega)| > n\},$$

we have

$$E_n = \bigcup_{i\ge n} F_{ni},$$

where

$$F_{ni} = \{\omega : i < |X_n(\omega)| \le i+1\}.$$

Now

$$\mu F_{ni} = \nu\{x : i < |x| \le i+1\}$$

for every n and i. So

$$\sum_{n=0}^{\infty}\mu E_n=\sum_{n=0}^{\infty}\sum_{i=n}^{\infty}\mu F_{ni}=\sum_{i=0}^{\infty}\sum_{n=0}^{i}\mu F_{ni}$$
$$=\sum_{i=0}^{\infty}(i+1)\nu\{x:i<|x|\le i+1\}\le\int(1+|x|)\nu(dx)<\infty.$$

Consequently the set $H=\{\omega:\{n:X_n(\omega)\ne Y_n(\omega)\}$ is finite$\}$ is conegligible (273A). But of course

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

for every $\omega\in H$.

(c) Finally,

$$|\mathbb{E}(Y_n)-\mathbb{E}(X_n)|\le\int_{E_n}|X_n|=\int_{\mathbb{R}\setminus[-n,n]}|x|\nu(dx)$$

whenever $n\in\mathbb{N}$, so $\lim_{n\to\infty}\mathbb{E}(Y_n)-\mathbb{E}(X_n)=0$ and

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

(273Ca). Putting these three together, we get

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

whenever ω belongs to the conegligible set $G\cap H$. So

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

almost everywhere, as required.

Remarks In my own experience, this is the most important form of the strong law from the point of view of 'pure' measure theory. I note that 273G above can also be regarded as a consequence of this form.

For a very striking alternative proof, see 275Yq. Yet another proof treats this result as a special case of the Ergodic Theorem (see 372Xg in Volume 3).

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 $\Omega^{\mathbb{N}}$ (254A-254C).

proof (a) To begin with, suppose that f is integrable. Define functions X_n on $\Omega^{\mathbb{N}}$ by setting

$$X_n(\boldsymbol{\omega})=f(\omega_n)\text{ whenever }\omega_n\in\operatorname{dom}f.$$

Then $\langle X_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of random variables, all with the same distribution as f (272M). So

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f(\omega_i)-\int f d\mu=\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}X_i(\boldsymbol{\omega})-\mathbb{E}(X_i)=0$$

for almost every $\boldsymbol{\omega}$, by 273I, and

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

for almost every $\boldsymbol{\omega}$.

(b) Next, suppose that $f\ge 0$ and $\int f=\infty$. In this case, for every $m\in\mathbb{N}$,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f(\omega_i)&\ge\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\min(m,f(\omega_i))\\&=\int\min(m,f(\omega))\mu(d\omega)\end{aligned}$$

for almost every $\boldsymbol{\omega}$, so

$$\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f(\omega_i)\ge\sup_{m\in\mathbb{N}}\int\min(m,f(\omega))\mu(d\omega)=\infty$$

and

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

for almost every $\boldsymbol{\omega}$.

(c) In general, if $\int f=\infty$, this is because $\int f^+=\infty$ and f^- is integrable, so

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f(\omega_i)&=\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f^+(\omega_i)-\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}f^-(\omega_i)\\&=\infty-\int f^-=\int f\end{aligned}$$

for almost every $\boldsymbol{\omega}$. Similarly,

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

for almost every $\boldsymbol{\omega}$ if $\int fd\mu=-\infty$.

Remark I find myself slipping here into measure-theorists' terminology; this corollary is one of the basic applications of the strong law to measure theory. Obviously, in view of 272J and 272M, this corollary covers 273I. It could also (in theory) be used as a *definition* of integration on a probability space (see 273Ya); it is sometimes called the 'Monte Carlo' method of integration.

273K It is tempting to seek extensions of 273I in which the X_n are not identically distributed, but are otherwise well-behaved. Any such idea should be tested against the following example. I find that I need another standard result, complementing that in 273A.

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.

proof For n, $k\in\mathbb{N}$ set $X_n=\sum_{i=0}^{n}\chi E_i$, $\beta_n=\sum_{i=0}^{n}\mu E_n=\mathbb{E}(X_n)$ and $F_{nk}=\{x:x\in\Omega,\ \#(\{i:i\le n,\ x\in E_i\})\le k\}$. Then

$$\begin{aligned}\mathbb{E}(X_n^2)&=\sum_{i=0}^{n}\sum_{j=0}^{n}\mu(E_i\cap E_j)\le\sum_{i=0}^{n}\mu E_i+\sum_{i=0}^{n}\sum_{j\ne i}\mu E_i\cdot\mu E_j\\&=\beta_n+\beta_n^2-\sum_{i=0}^{n}(\mu E_i)^2,\end{aligned}$$

so

$$\mathrm{Var}(X_n)=\beta_n-\sum_{i=0}^{n}(\mu E_i)^2\le\beta_n.$$

Now if $k<\beta_n$,

$$(\beta_n-k)^2\mu F_{nk}=(\beta_n-k)^2\Pr(X_n\le k)\le\mathbb{E}(X_n-\beta_n)^2=\mathrm{Var}(X_n)\le\beta_n$$

and $\mu F_{nk}\le\frac{\beta_n}{(\beta_n-k)^2}$.

Now recall that we are assuming that $\lim_{n\to\infty}\beta_n=\infty$. So for any $k\in\mathbb{N}$,

$$\mu(\bigcap_{n\in\mathbb{N}}F_{nk})=\lim_{n\to\infty}\mu F_{nk}\le\lim_{n\to\infty}\frac{\beta_n}{(\beta_n-k)^2}=0.$$

Accordingly

$$\mu\{x: x \text{ belongs to only finitely many } E_n\}=\mu(\bigcup_{k\in\mathbb{N}}\bigcap_{n\in\mathbb{N}}F_{nk})=0,$$

and almost every point of Ω belongs to infinitely many E_n.

Remark Of course this result is usually applied to an independent sequence $\langle E_n\rangle_{n\in\mathbb{N}}$. But very occasionally it is of interest to know that it is enough to assume that weaker hypotheses suffice. See also 273Yb.

273L Now for the promised example.

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.

proof Let (Ω,Σ,μ) be a probability space with an independent sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of measurable sets such that $\mu E_n=\frac{1}{(n+3)\ln(n+3)}$ for each n. (I have nowhere explained exactly how to build such a sequence. Two obvious methods are available to us, and another a trifle less obvious. (i) Take $\Omega=\{0,1\}^{\mathbb{N}}$ and μ to be the product of the probabilities μ_n on $\{0,1\}$, defined by saying that $\mu_n\{1\}=\frac{1}{(n+3)\ln(n+3)}$ for each n; set $E_n=\{\omega:\omega(n)=1\}$, and appeal to 272M to check that the E_n are independent. (ii) Build the E_n inductively as subsets of $[0,1]$, arranging that each E_n should be a finite union of intervals, so that when you come to choose E_{n+1} the sets $E_0,\dots,E_n$ define a partition $\mathcal{I}_n$ of $[0,1]$ into intervals, and you can take E_{n+1} to be the union of (say) the left-hand subintervals of length a proportion $\frac{1}{(n+3)\ln(n+3)}$ of the intervals in $\mathcal{I}_n$. (iii) Use 215D to see that the method of (ii) can be used on any atomless probability space, as in 272Xa.)

Set $X_n=(n+3)\ln\ln(n+3)\chi E_n$ for each n; then $\langle X_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of real-valued random variables (272F) and $\mathbb{E}(X_n)=\frac{\ln\ln(n+3)}{\ln(n+3)}$ for each n, so that $\mathbb{E}(X_n)\to0$ as $n\to\infty$. Thus, for instance, $\{X_n:n\in\mathbb{N}\}$ is uniformly integrable and $\langle X_n\rangle_{n\in\mathbb{N}}\to0$ in measure (246Jc); while surely $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{E}(X_i)=0$.

On the other hand,

$$\begin{aligned}\sum_{n=0}^{\infty}\mu E_n=\sum_{n=0}^{\infty}\frac{1}{(n+3)\ln(n+3)}&\ge\int_0^{\infty}\frac{1}{(x+3)\ln(x+3)}dx\\&=\lim_{a\to\infty}(\ln\ln(a+3)-\ln\ln3)=\infty,\end{aligned}$$

so almost every ω belongs to infinitely many of the E_n, by the Borel-Cantelli lemma (273K). Now if we write $Y_n=\frac{1}{n+1}\sum_{i=0}^{n}X_i$, then if $\omega\in E_n$ we have $X_n(\omega)=(n+3)\ln\ln(n+3)$ so

$$Y_n(\omega)\ge\frac{n+3}{n+1}\ln\ln(n+3).$$

This means that

$$\{\omega : \limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n}(X_i(\omega) - \mathbb{E}(X_i)) = \infty\} = \{\omega : \limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i(\omega) = \infty\}$$
$$= \{\omega : \sup_{n\in\mathbb{N}} Y_n(\omega) = \infty\} \supseteq \{\omega : \{n : \omega \in E_n\} \text{ is infinite}\}$$

is conegligible, and the strong law of large numbers does not apply to $\langle X_n\rangle_{n\in\mathbb{N}}$.

Because

$$\lim_{n\to\infty} \|Y_n\|_1 = \lim_{n\to\infty} \mathbb{E}(Y_n) = \lim_{n\to\infty} \mathbb{E}(X_n) = 0$$

(273Ca), $\langle Y_n\rangle_{n\in\mathbb{N}} \to 0$ for the topology of convergence in measure, and $\langle Y_n\rangle_{n\in\mathbb{N}}$ has a subsequence converging to 0 almost everywhere (245K). So

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

for almost every ω. The fact that both $\limsup_{n\to\infty} Y_n$ and $\liminf_{n\to\infty} Y_n$ are constant almost everywhere is of course a consequence of the zero-one law (272P).

***273M** All the above has been concerned with pointwise convergence of the averages of independent random variables, and that is the important part of the work of this section. But it is perhaps worth complementing it with a brief investigation of norm-convergence. To deal efficiently with convergence in $\mathcal{L}^p$, we need the following. (I should perhaps remark that, compared with the general case treated here, the case $p = 2$ is trivial; see 273Xl.)

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 $\mathbb{E}(X) = 0$.

proof (a) Take $\zeta \in]0, 1]$ such that $p\zeta \le 2$ and

$$(1 + \xi)^p \le 1 + p\xi + \frac{p^2}{2}\xi^2$$

whenever $|\xi| \le \zeta$; such exists because

$$\lim_{\xi\to 0} \frac{(1+\xi)^p - 1 - p\xi}{\xi^2} = \frac{p(p-1)}{2} < \frac{p^2}{2}.$$

Observe that

$$(1 + \xi)^p \le (1 + \frac{1}{\zeta})^p + \xi^p + 2p\xi^{p-1}$$

for every $\xi \ge 0$. **P** If $\xi \le \frac{1}{\zeta}$, this is trivial. If $\xi \ge \frac{1}{\zeta}$, then

$$(1 + \xi)^p = \xi^p(1 + \frac{1}{\xi})^p \le \xi^p(1 + \frac{p}{\xi} + \frac{p^2}{2\xi^2})$$
$$\le \xi^p(1 + \frac{p}{\xi} + \frac{p^2\zeta}{2\xi}) = \xi^p + p\xi^{p-1}(1 + \frac{p\zeta}{2}) \le \xi^p + 2p\xi^{p-1}. \text{ **Q**}$$

Define $\eta > 0$ by declaring that $3\eta^{p-1} = \frac{\epsilon}{2}$ (this is one of the places where we need to know that $p > 1$). Let $\delta > 0$ be such that

$$\delta \le \eta\zeta, \quad \frac{p^2}{2\eta^2}\delta + (1 + \frac{1}{\zeta})^p\delta^{p-1} \le \frac{p\epsilon}{2}.$$

(b) Now suppose that S and X are independent random variables with $\|S\|_p = 1$, $\|X\|_p \le \delta$ and $\mathbb{E}(X) = 0$. If $\|X\|_p = 0$ then of course $\|S + X\|_p \le 1 + \epsilon\|X\|_p$, so suppose that X is non-trivial. Write (Ω, Σ, μ) for the underlying probability space and adjust S and X on negligible sets so that they are measurable and defined everywhere on Ω. Set $\alpha = \|X\|_p$, $\gamma = \alpha/\eta$,

$$E = \{\omega : S(\omega) \ne 0\}, \quad F = \{\omega : |X(\omega)| > \gamma|S(\omega)|\}, \quad \beta = \|S \times \chi F\|_p.$$

Then

$$\int |S+X|^p = \int_F |S+X|^p + \int_{E\setminus F} |S+X|^p$$

(because S and X are both zero on $\Omega\setminus(E\cup F)$)

$$= \|(S\times\chi F)+(X\times\chi F)\|_p^p + \int_{E\setminus F} |S|^p|1+\frac{X}{S}|^p$$

$$\le (\|S\times\chi F\|_p + \|X\times\chi F)\|_p)^p + \int_{E\setminus F}|S|^p(1+p\frac{X}{S}+\frac{p^2}{2}\gamma^2)$$

(because $|\frac{X}{S}|\le\gamma\le\frac{\delta}{\eta}\le\zeta\le 1$ everywhere on $E\setminus F$)

$$\le(\beta+\alpha)^p+(1+\frac{p^2}{2}\gamma^2)\int_{E\setminus F}|S|^p+p\int_{E\setminus F}|S|^{p-1}\times\operatorname{sgn}S\times X$$

(writing $\operatorname{sgn}(\xi)=\xi/|\xi|$ if $\xi\ne 0$, 0 if $\xi=0$)

$$=(\beta+\alpha)^p+(1+\frac{p^2}{2}\gamma^2)\int_{\Omega\setminus F}|S|^p+p\int_{\Omega\setminus F}|S|^{p-1}\times\operatorname{sgn}S\times X$$

(because $S=0$ on $\Omega\setminus E$)

$$=\alpha^p(1+\frac{\beta}{\alpha})^p+(1+\frac{p^2}{2}\gamma^2)(1-\beta^p)-p\int_F|S|^{p-1}\times\operatorname{sgn}S\times X$$

(because X and $|S|^{p-1}\times\operatorname{sgn}S$ are independent, by 272L, so $\int|S|^{p-1}\times\operatorname{sgn}S\times X=\mathbb{E}(|S|^{p-1}\times\operatorname{sgn}S)\mathbb{E}(X)=0$)

$$\le\alpha^p\big((1+\frac{1}{\zeta})^p+2p(\frac{\beta}{\alpha})^{p-1}+(\frac{\beta}{\alpha})^p\big)+(1+\frac{p^2}{2}\gamma^2)(1-\beta^p)$$
$$+p\int_F|S|^{p-1}\times|X|$$

(see (a) above)

$$\le\alpha^p(1+\frac{1}{\zeta})^p+\beta^p+2p\beta^{p-1}\alpha+(1+\frac{p^2}{2}\gamma^2)(1-\beta^p)$$
$$+p\int_F\frac{1}{\gamma^{p-1}}|X|^p$$

$$\le\alpha^p(1+\frac{1}{\zeta})^p+2p\frac{\alpha^p}{\gamma^{p-1}}+1+\frac{p^2}{2}\gamma^2+p\frac{\alpha^p}{\gamma^{p-1}}$$

(because $\beta=\|S\times\chi F\|_p\le\frac{1}{\gamma}\|X\times\chi F\|_p\le\frac{\alpha}{\gamma}$)

$$=\alpha^p(1+\frac{1}{\zeta})^p+3p\eta^{p-1}\alpha+1+\frac{p^2\alpha^2}{2\eta^2}$$

$$=1+\big(\alpha^{p-1}(1+\frac{1}{\zeta})^p+3p\eta^{p-1}+\frac{p^2\alpha}{2\eta^2}\big)\alpha$$

$$\le1+\big(\delta^{p-1}(1+\frac{1}{\zeta})^p+3p\eta^{p-1}+\frac{p^2\delta}{2\eta^2}\big)\alpha$$

$$\le1+p\alpha\epsilon\le(1+\epsilon\|X\|_p)^p.$$

So $\|S+X\|_p\le1+\epsilon\|X\|_p$, as required.

***Remark** What is really happening here is that $\phi=\|\;\|_p^p:L^p\to\mathbb{R}$ is differentiable (as a real-valued function on the normed space L^p) and

$$\phi'(S^\bullet)(X^\bullet)=p\int|S|^{p-1}\times\operatorname{sgn}S\times X,$$

so that in the context here

$$\phi((S+X)^\bullet)=\phi(S^\bullet)+\phi'(S^\bullet)(X^\bullet)+o(\|X\|_p)=1+o(\|X\|_p)$$

and $\|S+X\|_p = 1+o(\|X\|_p)$. The calculations above are elaborate partly because they do not appeal to any non-trivial ideas about normed spaces, and partly because we need the estimates to be uniform in S.

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 $\lim_{n\to\infty}\|Y_n\|_p = 0$.

proof (a) Let $\epsilon>0$. Then there is an $M\ge 0$ such that $\mathbb{E}(|X_n|-M)^+\le\epsilon$ for every $n\in\mathbb{N}$. Set

$$X'_n = (-M\chi\Omega)\vee(X_n\wedge M\chi\Omega),\quad \alpha_n=\mathbb{E}(X'_n),\quad \tilde X_n = X'_n-\alpha_n,\quad X''_n = X_n - X'_n$$

for each $n\in\mathbb{N}$. Then $\langle X'_n\rangle_{n\in\mathbb{N}}$ and $\langle\tilde X_n\rangle_{n\in\mathbb{N}}$ are independent and uniformly bounded, and $\|X''_n\|_1\le\epsilon$ for every n. So if we write

$$\tilde Y_n = \frac{1}{n+1}\sum_{i=0}^n\tilde X_i,\quad Y''_n = \frac{1}{n+1}\sum_{i=0}^n X''_i,$$

$\langle\tilde Y_n\rangle_{n\in\mathbb{N}}\to 0$ almost everywhere, by 273E (for instance), while $\|Y''_n\|_1\le\epsilon$ for every n. Moreover,

$$|\alpha_n| = |\mathbb{E}(X'_n-X_n)|\le\mathbb{E}(|X''_n|)\le\epsilon$$

for every n. As $|\tilde Y_n|\le 2M$ almost everywhere for each n, $\lim_{n\to\infty}\|\tilde Y_n\|_1 = 0$, by Lebesgue's Dominated Convergence Theorem. So

$$\begin{aligned}\limsup_{n\to\infty}\|Y_n\|_1 &= \limsup_{n\to\infty}\|\tilde Y_n + Y''_n+\alpha_n\|_1\\ &\le \lim_{n\to\infty}\|\tilde Y_n\|_1 + \sup_{n\in\mathbb{N}}\|Y''_n\|_1+\sup_{n\in\mathbb{N}}|\alpha_n|\\ &\le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\lim_{n\to\infty}\|Y_n\|_1 = 0$, as claimed.

***(b)** Set $M=\sup_{n\in\mathbb{N}}\|X_n\|_p$. For $n\in\mathbb{N}$, set $S_n=\sum_{i=0}^n X_i$. Let $\epsilon>0$. Then 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 $\mathbb{E}(X)=0$ (273M). It follows that $\|S+X\|_p\le\|S\|_p+\epsilon\|X\|_p$ whenever S and X are independent random variables, $\|S\|_p$ is finite, $\|X\|_p\le\delta\|S\|_p$ and $\mathbb{E}(X)=0$. In particular, $\|S_{n+1}\|_p\le\|S_n\|_p+\epsilon M$ whenever $\|S_n\|_p\ge M/\delta$. An easy induction shows that

$$\|S_n\|_p\le\frac{M}{\delta}+M+n\epsilon M$$

for every $n\in\mathbb{N}$. But this means that

$$\limsup_{n\to\infty}\|Y_n\|_p = \limsup_{n\to\infty}\frac{1}{n+1}\|S_n\|_p\le\epsilon M.$$

As ϵ is arbitrary, $\lim_{n\to\infty}\|Y_n\|_p=0$.

Remark There are strengthenings of (a) in 276Xe, and of (b) in 276Ya.

273X Basic exercises (a) In part (b) of the proof of 273B, use Bienaymé's equality to show that $\lim_{m\to\infty}\sup_{n\ge m}\Pr(|S_n-S_m|\ge\epsilon)=0$ for every $\epsilon>0$, so that we can apply the argument of part (a) of the proof directly, without appealing to 242F or 245G or even 244E.

(b) Show that $\sum_{n=0}^\infty\frac{(-1)^{\omega(n)}}{n+1}$ is defined in $\mathbb{R}$ for almost every $\boldsymbol{\omega}=\langle\omega(n)\rangle_{n\in\mathbb{N}}$ in $\{0,1\}^{\mathbb{N}}$, where $\{0,1\}^{\mathbb{N}}$ is given its usual measure (254J).

(c) Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of measurable sets in a probability space, all with the same non-zero measure. Let $\langle a_n\rangle_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers such that $\sum_{n=0}^\infty a_n=\infty$. Show that $\sum_{n=0}^\infty a_n\chi E_n=\infty$ a.e. (*Hint*: Take a strictly increasing sequence $\langle k_n\rangle_{n\in\mathbb{N}}$ such that $d_n=\sum_{i=k_n+1}^{k_{n+1}}a_i\ge 1$ for each n. Set $c_i=\frac{a_i}{(n+1)d_n}$ for $k_n<i\le k_{n+1}$; show that $\sum_{n=0}^\infty c_n^2<\infty=\sum_{n=0}^\infty c_n$. Apply 273D with $X_n=c_n\chi E_n$ and $b_n=\sqrt{\sum_{i=0}^n c_i}$.)

>(d) Take any $q \in [0,1]$, and give $\mathcal{P}\mathbb{N}$ a measure μ such that

$$\mu\{a : I \subseteq a\} = q^{\#(I)}$$

for every $I \subseteq \mathbb{N}$, as in 254Xg. Show that for μ-almost every $a \subseteq \mathbb{N}$,

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

>(e) Let μ be the usual probability measure on $\mathcal{P}\mathbb{N}$ (254Jb), and for $r \ge 1$ let μ^r be the product probability measure on $(\mathcal{P}\mathbb{N})^r$. Show that

$$\lim_{n\to\infty} \frac{1}{n+1}\#(a_1 \cap \ldots \cap a_r \cap \{0,\dots,n\}) = 2^{-r},$$

$$\lim_{n\to\infty} \frac{1}{n+1}\#((a_1 \cup \ldots \cup a_r) \cap \{0,\dots,n\}) = 1 - 2^{-r}$$

for μ^r-almost every $(a_1,\dots,a_r) \in (\mathcal{P}\mathbb{N})^r$.

(f) Let μ be the usual probability measure on $\mathcal{P}\mathbb{N}$, and b any infinite subset of $\mathbb{N}$. Show that $\lim_{n\to\infty} \frac{\#(a\cap b\cap\{0,\dots,n\})}{\#(b\cap\{0,\dots,n\})} = \frac{1}{2}$ for almost every $a \subseteq \mathbb{N}$.

>(g) For each $x \in [0,1]$, let $\epsilon_k(x)$ be the kth digit in the decimal expansion of x (choose for yourself what to do with $0{\cdot}100\ldots = 0{\cdot}099\ldots$). Show that $\lim_{k\to\infty} \frac{1}{k}\#(\{j : j \le k,\, \epsilon_j(x) = 7\}) = \frac{1}{10}$ for almost every $x \in [0,1]$.

(h) Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence of distribution functions for real-valued random variables, in the sense of 271Ga, and F another distribution function; suppose that $\lim_{n\to\infty} F_n(q) = F(q)$ for every $q \in \mathbb{Q}$ and $\lim_{n\to\infty} F_n(a^-) = F(a^-)$ whenever $F(a^-) < F(a)$, where I write $F(a^-)$ for $\lim_{x\uparrow a} F(x)$. Show that $F_n \to F$ uniformly.

>(i) Let (Ω,Σ,μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent identically distributed sequence of real-valued random variables on Ω with common distribution function F. For $a \in \mathbb{R}$, $n \in \mathbb{N}$ and $\omega \in \bigcap_{i\le n} \operatorname{dom} X_i$ set

$$F_n(\omega, a) = \frac{1}{n+1}\#(\{i : i \le n,\, X_i(\omega) \le a\}).$$

Show that

$$\lim_{n\to\infty} \sup_{a\in\mathbb{R}} |F_n(\omega,a) - F(a)| = 0$$

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

(j) Let (Ω,Σ,μ) be a probability space, and λ the product measure on $\Omega^{\mathbb{N}}$. Let $f : \Omega \to \mathbb{R}$ be a function, and set $f^*(\boldsymbol{\omega}) = \limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^n f(\omega_i)$ for $\boldsymbol{\omega} = \langle \omega_n\rangle_{n\in\mathbb{N}} \in \Omega^{\mathbb{N}}$. Show that $\overline{\int} f^* d\lambda = \overline{\int} f d\mu$ whenever the right-hand-side is finite. (*Hint*: 133J(a-i).)

(k) Find an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables with zero expectation such that $\|X_n\|_1 = 1$ and $\|\frac{1}{n+1}\sum_{i=0}^n X_i\|_1 \ge \frac{1}{2}$ for every $n \in \mathbb{N}$. (*Hint*: take $\Pr(X_n \ne 0)$ very small.)

(l) Use 272S to prove 273Nb in the case $p = 2$.

(m) Find an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables with zero expectation such that $\|X_n\|_\infty = \|\frac{1}{n+1}\sum_{i=0}^n X_i\|_\infty = 1$ for every $n \in \mathbb{N}$.

(n) Repeat the work of this section for complex-valued random variables.

(o) Let (X,Σ,μ) be a probability space and $\langle E_n\rangle_{n\in\mathbb{N}}$ an independent sequence in Σ such that $\alpha = \lim_{n\to\infty}\mu E_n$ is defined. For $x\in X$ set $I_x=\{n : x\in E_n\}$. Show that I_x has asymptotic density α for almost every x.

273Y Further exercises (a) Let (Ω,Σ,μ) be a probability space, and λ the product measure on $\Omega^{\mathbb{N}}$. Suppose that f is a real-valued function, defined on a subset of Ω, such that

$$h(\boldsymbol{\omega}) = \lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i)$$

exists in $\mathbb{R}$ for λ-almost every $\boldsymbol{\omega} = \langle\omega_n\rangle_{n\in\mathbb{N}}$ in $\Omega^{\mathbb{N}}$. Show (i) that f has conegligible domain (ii) f is $\hat\Sigma$-measurable, where $\hat\Sigma$ is the domain of the completion of μ (iii) there is an $a\in\mathbb{R}$ such that $h=a$ almost everywhere in $\Omega^{\mathbb{N}}$ (iv) f is integrable, with $\int f d\mu = a$.

(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of random variables with finite variance. Suppose that $\lim_{n\to\infty}\mathbb{E}(X_n)=\infty$ and $\liminf_{n\to\infty}\frac{\mathbb{E}(X_n^2)}{(\mathbb{E}(X_n))^2}\le 1$. Show that $\limsup_{n\to\infty}X_n=\infty$ a.e.

273 Notes and comments I have tried in this section to offer the most useful of the standard criteria for pointwise convergence of averages of independent random variables. In my view the strong law of large numbers, like Fubini's theorem, is one of the crucial steps in measure theory, where the subject changes character. Theorems depending on the strong law have a kind of depth and subtlety to them which is missing in other parts of the subject. I have described only a handful of applications here, but I hope that 273G, 273J, 273Xd, 273Xg and 273Xi will give an idea of what is to be expected. These do have rather different weights. Of the four, only 273J requires the full resources of this chapter; the others can be deduced from the essentially simpler version in 273Xi.

273Xi is the 'fundamental theorem of statistics' or 'Glivenko-Cantelli theorem'. The $F_n(.,a)$ are 'statistics', computed from the X_i; they are the 'empirical distributions', and the theorem says that, almost surely, $F_n\to F$ uniformly. (I say 'uniformly' to make the result look more striking, but of course the real content is that $F_n(.,a)\to F(a)$ almost surely for each a; the extra step is just 273Xh.)

I have included 273N to show that independence is quite as important in questions of norm-convergence as it is in questions of pointwise convergence. It does not really rely on any form of the strong law; in the proof I quote 273E as a quick way of disposing of the 'uniformly bounded parts' X'_n, but of course Bienaymé's equality (272S) is already enough to show that if $\langle X'_n\rangle_{n\in\mathbb{N}}$ is an independent uniformly bounded sequence of random variables with zero expectation, then $\|\frac{1}{n+1}(X_0+\ldots+X_n)\|_p\to 0$ for $p=2$, and therefore for every $p<\infty$.

The proofs of 273H, 273I and 273Na all involve 'truncation'; the expression of a random variable X as the sum of a bounded random variable and a tail. This is one of the most powerful techniques in the subject, and will appear again in §276 and (in a rather different way) in §274. In 273Na I used a slightly different formulation of the method, solely because it matched the definition of 'uniformly integrable' more closely.

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 We need some facts from basic probability theory.

(a) Recall that

$$\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$$

(263G). Consequently, 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 (256E); we call it the **standard normal distribution**. The corresponding distribution function is

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

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

Writing Σ for the algebra of Lebesgue measurable subsets of $\mathbb{R}$, $(\mathbb{R}, \Sigma, \mu_G)$ is a probability space. Note that it is complete, and has the same negligible sets as Lebesgue measure, because $e^{-x^2/2} > 0$ for every x (cf. 234Lc).

(b) A random variable X is **standard normal** if its distribution is μ_G; that is, if the function $x \mapsto \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is a density function for X. The point of the remarks in (a) is that there are such random variables; for instance, take the probability space $(\mathbb{R}, \Sigma, \mu_G)$ there, and set $X(x) = x$ for every $x \in \mathbb{R}$.

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

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

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

by 263H.

(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 $X = \sigma Z + a$ so $\mathbb{E}(X) = \sigma\mathbb{E}(Z) + a = a$, $\mathrm{Var}(X) = \sigma^2\,\mathrm{Var}(Z) = \sigma^2$.

We have, for any $c \in \mathbb{R}$,

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

(substituting $x = a + \sigma y$ for $-\infty < y \le (c-a)/\sigma$)

$$= \Pr(Z \le \tfrac{c-a}{\sigma}) = \Pr(X \le c).$$

So $x \mapsto \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2/2\sigma^2}$ is a density function for X (271Ib). Conversely, of course, 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 Z is standard normal, so is $-Z$, because

$$\Pr(-Z \le a) = \Pr(Z \ge -a) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{\infty} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx.$$

The definition in the first sentence of (d) now makes it obvious that if X is normal, so is $a + bX$ for any $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$.

274B Proposition Let $X_1, \dots, X_n$ 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 $\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$.

proof There are innumerable proofs of this fact; the following one gives me a chance to show off the power of Chapter 26, but of course (at the price of some disagreeable algebra) 272U also gives the result.

(a) Consider first the case $n = 2$. Setting $a_i = \mathbb{E}(X_i)$, $\sigma_i = \sqrt{\mathrm{Var}(X_i)}$, $Z_i = (X_i - a_i)/\sigma_i$ we get independent standard normal variables Z_1, Z_2. Set $\rho = \sqrt{\sigma_1^2 + \sigma_2^2}$, and express σ_1, σ_2 as $\rho\cos\theta$, $\rho\sin\theta$. Consider $U = \cos\theta Z_1 + \sin\theta Z_2$. We know that (Z_1, Z_2) has a density function

$$(\zeta_1, \zeta_2) \mapsto g(\zeta_1, \zeta_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-(\zeta_1^2+\zeta_2^2)/2}$$

(272I). Consequently, for any $c \in \mathbb{R}$,

$$\Pr(U \le c) = \int_F g(z)dz,$$

where $F = \{(\zeta_1, \zeta_2) : \zeta_1 \cos\theta + \zeta_2 \sin\theta \le c\}$. But now let T be the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then it is easy to check that

$$T^{-1}[F] = \{(\eta_1, \eta_2) : \eta_1 \le c\},$$

$$\det T = 1, \quad g(Ty) = g(y) \text{ for every } y \in \mathbb{R}^2,$$

so by 263A

$$\Pr(U \le c) = \int_F g(z)dz = \int_{T^{-1}[F]} g(Ty)dy = \int_{]-\infty,c]\times\mathbb{R}} g(y)dy = \Pr(Z_1 \le c) = \Phi(c).$$

As this is true for every $c \in \mathbb{R}$, U also is standard normal (I am appealing to 271Ga again). But

$$X_1 + X_2 = \sigma_1 Z_1 + \sigma_2 Z_2 + a_1 + a_2 = \rho U + a_1 + a_2,$$

so $X_1 + X_2$ is normal.

(b) Now we can induce on n. If $n = 1$ the result is trivial. For the inductive step to $n + 1 \ge 2$, we know that $X_1 + \ldots + X_n$ is normal, by the inductive hypothesis, and that X_{n+1} is independent of $X_1 + \ldots + X_n$, by 272L. So $X_1 + \ldots + X_n + X_{n+1}$ is normal, by (a).

The computation of the expectation and variance of $X_1 + \ldots + X_n$ is immediate from 271Ab and 272S.

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

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

proof For $0 \le j \le n+1$, set $Z_j = \sum_{i=0}^{j-1} U_i + \sum_{i=j}^n V_i$, taking $Z_0 = \sum_{i=0}^n V_i$ and $Z_{n+1} = \sum_{i=0}^n U_i$, and for $j \le n$ set $W_j = \sum_{i=0}^{j-1} U_j + \sum_{i=j+1}^n V_j$, so that $Z_j = W_j + V_j$ and $Z_{j+1} = W_j + U_j$ and W_j, U_j and V_j are independent (I am appealing to 272K, as in 272L). Then

$$\begin{aligned} |\mathbb{E}\big(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\big)| &= |\mathbb{E}\big(\sum_{i=0}^n h(Z_{i+1}) - h(Z_i)\big)| \\ &\le \sum_{i=0}^n |\mathbb{E}\big(h(Z_{i+1}) - h(Z_i)\big)| \\ &= \sum_{i=0}^n |\mathbb{E}\big(h(W_i + U_i) - h(W_i + V_i)\big)|. \end{aligned}$$

To estimate this sum I turn it into a sum of integrals, as follows. For each i, let ν_{W_i} be the distribution of W_i, and so on. Because $(w, u) \mapsto w + u$ is continuous, therefore Borel measurable, $(w, u) \mapsto h(w, u)$ also is Borel measurable; accordingly $(w, u, v) \mapsto h(w + u) - h(w + v)$ is measurable for each of the product measures $\nu_{W_i} \times \nu_{U_i} \times \nu_{V_i}$ on $\mathbb{R}^3$, and 271E and 272G give us

$$
\begin{aligned}
\bigl|\mathbb{E}\bigl(h(W_i+U_i)-&h(W_i+V_i)\bigr)\bigr| \\
&= \bigl|\int h(w+u)-h(w+v)(\nu_{W_i}\times\nu_{U_i}\times\nu_{V_i})d(w,u,v)\bigr| \\
&= \bigl|\int\Bigl(\int h(w+u)-h(w+v)(\nu_{U_i}\times\nu_{V_i})d(u,v)\Bigr)\nu_{W_i}(dw)\bigr| \\
&\le \int\bigl|\int h(w+u)-h(w+v)(\nu_{U_i}\times\nu_{V_i})d(u,v)\bigr|\nu_{W_i}(dw) \\
&= \int\bigl|\mathbb{E}\bigl(h(w+U_i)-h(w+V_i)\bigr)\bigr|\nu_{W_i}(dw) \\
&\le \sup_{t\in\mathbb{R}}\bigl|\mathbb{E}\bigl(h(t+U_i)-h(t+V_i)\bigr)\bigr|.
\end{aligned}
$$

So we get

$$
\begin{aligned}
|\mathbb{E}\bigl(h(\sum_{i=0}^n U_i)-h(\sum_{i=0}^n V_i)\bigr)| &\le \sum_{i=0}^n|\mathbb{E}\bigl(h(W_i+U_i)-h(W_i+V_i)\bigr)| \\
&\le \sum_{i=0}^n\sup_{t\in\mathbb{R}}|\mathbb{E}\bigl(h(t+U_i)-h(t+V_i)\bigr)|,
\end{aligned}
$$

as required.

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

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

proof (a) The point is that, by Taylor's theorem with remainder,

$$|h(t+x)-h(t)-xh'(t)|\le\frac{1}{2}M_2x^2,$$

$$|h(t+x)-h(t)-xh'(t)-\frac{1}{2}x^2h''(t)|\le\frac{1}{6}M_3|x|^3$$

for every $x\in\mathbb{R}$. So

$$|h(t+x)-h(t)-xh'(t)-\frac{1}{2}x^2h''(t)|\le\min(\frac{1}{6}M_3|x|^3,M_2x^2)\le\frac{1}{6}\epsilon M_3x^2+M_2\psi_\epsilon(x).$$

Integrating with respect to the distribution of U, we get

$$
\begin{aligned}
|\mathbb{E}\big(h(t+U)\big)-h(t)-\tfrac{1}{2}h''(t)\sigma^2\big)| &= |\mathbb{E}(h(t+U))-h(t)-h'(t)\mathbb{E}(U)-\frac{1}{2}h''(t)\mathbb{E}(U^2)| \\
&= |\mathbb{E}\big(h(t+U)-h(t)-h'(t)U-\frac{1}{2}h''(t)U^2\big)| \\
&\le \mathbb{E}\big(|h(t+U)-h(t)-h'(t)U-\frac{1}{2}h''(t)U^2|\big) \\
&\le \mathbb{E}\big(\frac{1}{6}\epsilon M_3U^2+M_2\psi_\epsilon(U)\big) \\
&= \frac{1}{6}\epsilon M_3\sigma^2+M_2\mathbb{E}(\psi_\epsilon(U)),
\end{aligned}
$$

as claimed.

(b) By 274C,

$$
\begin{aligned}
|\mathbb{E}\big(h(\sum_{i=0}^n U_i)-h(\sum_{i=0}^n V_i)\big)| &\le \sum_{i=0}^n \sup_{t\in\mathbb{R}} |\mathbb{E}\big(h(t+U_i)-h(t+V_i)\big)| \\
&\le \sum_{i=0}^n \sup_{t\in\mathbb{R}}\big(|\mathbb{E}(h(t+U_i))-h(t)-\frac{1}{2}h''(t)\sigma_i^2| \\
&\qquad + |\mathbb{E}(h(t+V_i))-h(t)-\frac{1}{2}h''(t)\sigma_i^2|\big),
\end{aligned}
$$

which by (a) above is at most

$$\textstyle\sum_{i=0}^n \frac{1}{3}\epsilon M_3\sigma_i^2+M_2\mathbb{E}(\psi_\epsilon(U_i))+M_2\mathbb{E}(\psi_\epsilon(V_i)),$$

as claimed.

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 $x\ge\epsilon$.

proof Let $f:\,]-\epsilon,\epsilon[\,\to\,]0,\infty[$ be any twice-differentiable function such that

$$\lim_{x\downarrow-\epsilon} f^{(n)}(x)=\lim_{x\uparrow\epsilon} f^{(n)}(x)=0$$

for $n=0$, 1 and 2, writing $f^{(n)}$ for the nth derivative of f; for instance, you could take $f(x)=(\epsilon^2-x^2)^3$, or $f(x)=\exp(-\frac{1}{\epsilon^2-x^2})$. Now set

$$h(x)=1-\textstyle\int_{-\epsilon}^x f/\int_{-\epsilon}^{\epsilon} f$$

for $|x|\le\epsilon$.

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

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

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

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

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

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

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

proof (a) Let $h:\mathbb{R}\to[0,1]$ be a three-times-differentiable function, with continuous third derivative, such that $\chi\,]-\infty,-\epsilon]\le h\le\chi\,]-\infty,\epsilon]$, as in 274E. Set

$$M_2 = \sup_{x\in\mathbb{R}} |h''(x)| = \sup_{|x|\le\epsilon} |h''(x)|,$$

$$M_3 = \sup_{x\in\mathbb{R}} |h'''(x)| = \sup_{|x|\le\epsilon} |h'''(x)|;$$

because h''' is continuous, both are finite. Write $\epsilon' = \epsilon(1-\frac{2}{\sqrt{2\pi}}) > 0$, and let $\eta > 0$ be such that

$$(\frac{1}{3}M_3 + 2M_2)\eta \le \epsilon'.$$

Note that $\lim_{m\to\infty} \psi_m(x) = 0$ for every x, so if X is a random variable with finite variance we must have $\lim_{m\to\infty} \mathbb{E}(\psi_m(X)) = 0$, by Lebesgue's Dominated Convergence Theorem; let $m \ge 1$ be such that $\mathbb{E}(\psi_m(Z)) \le \eta$, where Z is some (or any) standard normal random variable. Finally, take $\delta > 0$ such that $\delta + \delta^2 \le (\eta/m)^2$; note that $\delta \le \eta$.

(I hope that you have seen enough ϵ-δ arguments not to be troubled by any expectation of understanding the reasons for each particular formula here before reading the rest of the argument. But the formula $\frac{1}{3}M_3 + 2M_2$, in association with ψ_δ, should recall 274D.)

(b) Let $X_0, \dots, X_n$ be independent random variables with zero expectation such that $\sum_{i=0}^n \mathrm{Var}(X_i) = 1$ and $\sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i)) \le \delta$. We need an auxiliary sequence $Z_0, \dots, Z_n$ of standard normal random variables to match against the X_i. To create this, I use the following device. Suppose that the probability space underlying $X_0, \dots, X_n$ is (Ω, Σ, μ). Set $\Omega' = \Omega \times \mathbb{R}^{n+1}$, and let μ' be the product measure on Ω', where Ω is given the measure μ and each factor $\mathbb{R}$ of $\mathbb{R}^{n+1}$ is given the measure μ_G. Set $X_i'(\omega, z) = X_i(\omega)$ and $Z_i(\omega, z) = \zeta_i$ for $\omega \in \mathrm{dom}\, X_i$, $z = (\zeta_0, \dots, \zeta_n) \in \mathbb{R}^{n+1}$, $i \le n$. Then $X_0', \dots, X_n', Z_0, \dots, Z_n$ are independent, and each X_i' has the same distribution as X_i (272Mb). Consequently $S' = X_0' + \dots + X_n'$ has the same distribution as $S = X_0 + \dots + X_n$ (using 272T, or otherwise); so that $\mathbb{E}(g(S')) = \mathbb{E}(g(S))$ for any bounded Borel measurable function g (using 271E). Also each Z_i has distribution μ_G, so is standard normal.

(c) Write $\sigma_i = \sqrt{\mathrm{Var}(X_i)}$ for each i, and set $K = \{i : i \le n,\ \sigma_i > 0\}$. Observe that $\eta/\sigma_i \ge m$ for each $i \in K$. **P** We know that

$$\sigma_i^2 = \mathrm{Var}(X_i) = \mathbb{E}(X_i^2) \le \mathbb{E}(\delta^2 + \psi_\delta(X_i)) = \delta^2 + \mathbb{E}(\psi_\delta(X_i)) \le \delta^2 + \delta,$$

so

$$\frac{\eta}{\sigma_i} \ge \frac{\eta}{\sqrt{\delta+\delta^2}} \ge m$$

by the choice of δ. **Q**

(d) Consider the independent normal random variables $\sigma_i Z_i$. We have $\mathbb{E}(\sigma_i Z_i) = \mathbb{E}(X_i') = 0$ and $\mathrm{Var}(\sigma_i Z_i) = \mathrm{Var}(X_i') = \sigma_i^2$ for each i, so that $Z = \sigma_0 Z_0 + \dots + \sigma_n Z_n$ has expectation 0 and variance $\sum_{i=0}^n \sigma_i^2 = 1$; moreover, by 274B, Z is normal, so in fact it is standard normal. Now we have

$$\sum_{i=0}^n \mathbb{E}(\psi_\eta(\sigma_i Z_i)) = \sum_{i\in K} \mathbb{E}(\psi_\eta(\sigma_i Z_i)) = \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_{\eta/\sigma_i}(Z_i))$$

(because $\sigma^2\psi_{\eta/\sigma}(x) = \psi_\eta(\sigma x)$ whenever $x \in \mathbb{R}$, $\sigma > 0$)

$$= \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_{\eta/\sigma_i}(Z)) \le \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_m(Z))$$

(because, by (c), $\eta/\sigma_i \ge m$ for every $i \in K$, so $\psi_{\eta/\sigma_i}(t) \le \psi_m(t)$ for every t)

$$= \mathbb{E}(\psi_m(Z)) \le \eta$$

(by the choice of m). On the other hand, we surely have

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

(e) For any real number t, set

$$h_t(x) = h(x-t)$$

for each $x \in \mathbb{R}$. Then h_t is three-times-differentiable, with $\sup_{x\in\mathbb{R}} |h_t''(x)| = M_2$ and $\sup_{x\in\mathbb{R}} |h'''(x)| = M_3$. Consequently

$$|\mathbb{E}(h_t(S)) - \mathbb{E}(h_t(Z))| \le \epsilon'.$$

$\mathbf{P}$ By 274Db,

$$\begin{aligned}|\mathbb{E}(h_t(S)) - \mathbb{E}(h_t(Z))| &= |\mathbb{E}(h_t(S')) - \mathbb{E}(h_t(Z))| \\ &= |\mathbb{E}(h_t(\sum_{i=0}^n X_i')) - \mathbb{E}(h_t(\sum_{i=0}^n \sigma_i Z_i))| \\ &\le \frac{1}{3}\eta M_3 \sum_{i=0}^n \sigma_i^2 + M_2 \sum_{i=0}^n \mathbb{E}(\psi_\eta(X_i)) + M_2 \sum_{i=0}^n \mathbb{E}(\psi_\eta(\sigma_i Z_i)) \\ &\le \frac{1}{3}\eta M_3 + M_2\eta + M_2\eta \le \epsilon',\end{aligned}$$

by the choice of η. $\mathbf{Q}$

(f) Now take any $a \in \mathbb{R}$. We have

$$\chi\,]-\infty, a-2\epsilon] \le h_{a-\epsilon} \le \chi\,]\infty, a] \le h_{a+\epsilon} \le \chi\,]-\infty, a+\epsilon].$$

Note also that, for any b,

$$\Phi(b+2\epsilon) = \Phi(b) + \frac{1}{\sqrt{2\pi}}\int_b^{b+2\epsilon} e^{-x^2/2}dx \le \Phi(b) + \frac{2\epsilon}{\sqrt{2\pi}} = \Phi(b) + \epsilon - \epsilon'.$$

Consequently

$$\begin{aligned}\Phi(a) - \epsilon \le \Phi(a-2\epsilon) - \epsilon' &= \Pr(Z \le a-2\epsilon) - \epsilon' \le \mathbb{E}(h_{a-\epsilon}(Z)) - \epsilon' \le \mathbb{E}(h_{a-\epsilon}(S)) \\ &\le \Pr(S \le a) \\ &\le \mathbb{E}(h_{a+\epsilon}(S)) \le \mathbb{E}(h_{a+\epsilon}(Z)) + \epsilon' \le \Pr(Z \le a+2\epsilon) + \epsilon' = \Phi(a+2\epsilon) + \epsilon' \\ &\le \Phi(a) + \epsilon.\end{aligned}$$

But this means just that

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

as claimed.

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 $a \in \mathbb{R}$.

proof Given $\epsilon > 0$, take $\delta > 0$ as in Lindeberg's theorem (274F). Then for all n large enough,

$$\frac{1}{s_n^2}\textstyle\sum_{i=0}^n \mathbb{E}(\psi_{\delta s_n}(X_i)) \le \delta.$$

Fix on any such n. Of course we have $s_n > 0$. Set

$$X_i' = \frac{1}{s_n}X_i \text{ for } i \le n;$$

then $X_0', \dots, X_n'$ are independent, with zero expectation,

$$\sum_{i=0}^n \operatorname{Var}(X_i') = \sum_{i=0}^n \frac{1}{s_n^2} \operatorname{Var}(X_i) = 1,$$

$$\sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i')) = \sum_{i=0}^n \frac{1}{s_n^2} \mathbb{E}(\psi_{\delta s_n}(X_i)) \le \delta.$$

By 274F,

$$\big|\Pr(S_n \le a) - \Phi(a)\big| = \big|\Pr(\sum_{i=0}^n X_i' \le a) - \Phi(a)\big| \le \epsilon$$

for every $a \in \mathbb{R}$. Since this is true for all n large enough, we have the result.

274H Remarks (a) The condition

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

is called **Lindeberg's condition**, following Lindeberg 1922.

(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 $\sigma_n = \sqrt{\operatorname{Var}(X_n)}$, $s_n = \sqrt{\sum_{i=0}^n \operatorname{Var}(X_i)}$ 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 $S_n = \frac{1}{s_n}(X_0 + \dots + X_n)$. 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 $\epsilon > 0$. (Feller 66, §XV.6, Theorem 3; Loève 77, §21.2.)

(c) The proof of 274F-274G here is adapted from Feller 66, §VIII.4. It has the virtue of being 'elementary', in that it does not involve characteristic functions. Of course this has to be paid for by a number of detailed estimations; and – what is much more serious – it leaves us without one of the most powerful techniques for describing distributions. The proof does offer a method of bounding

$$|\Pr(S_n \le a) - \Phi(a)|;$$

but it should be said that the bounds obtained are not useful ones, being grossly over-pessimistic, at least in the readily analysable cases. (For instance, a better bound, in many cases, is given by the Berry-Esséen theorem: if $\langle X_n \rangle_{n\in\mathbb{N}}$ is independent and identically distributed, with zero expectation, and the common values of $\sqrt{\mathbb{E}(X_n^2)}$, $\mathbb{E}(|X_n|^3)$ are σ, $\rho < \infty$, then

$$|\Pr(S_n \le a) - \Phi(a)| \le \frac{33\rho}{4\sigma^3\sqrt{n+1}};$$

see Feller 66, §XVI.5, Loève 77, §21.3, or Hall 82.) Furthermore, when $|a|$ is large, $\Phi(a)$ is exceedingly close to either 0 or 1, so that any uniform bound for $|\Pr(S \le a) - \Phi(a)|$ gives very little information; a great deal of work has been done on estimating the tails of such distributions more precisely, subject to special conditions. For instance, if $X_0, \dots, X_n$ are independent random variables with zero expectation, uniformly bounded with $|X_i| \le K$ almost everywhere for each i, $Y = X_0 + \dots + X_n$, $s = \sqrt{\operatorname{Var}(Y)} > 0$, $S = \frac{1}{s}Y$, then for any $\alpha \in [0, s/K]$

$$\Pr(|S| \ge \alpha) \le 2\exp\Big(\frac{-\alpha^2}{2(1 + \frac{\alpha K}{2s})^2}\Big) \simeq 2e^{-\alpha^2/2}$$

if $s \gg \alpha K$ (Rényi 70, §VII.4, Theorem 1). A less precise result of the same kind is in 272Xl.

I now list some of the standard cases in which Lindeberg's condition is satisfied, so that we may apply the theorem.

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 $\sqrt{\operatorname{Var}(X_n)}$, 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 $a \in \mathbb{R}$.

proof In the language of 274G-274H, we have $\sigma_n = \sigma$, $s_n = \sigma\sqrt{n+1}$ and $S_n = \frac{1}{s_n}\sum_{i=0}^n X_i$. Moreover, if ν is the common distribution of the X_n, then

$$\mathbb{E}(\psi_{\epsilon s_n}(X_n)) = \int_{\{x:|x|>\epsilon\sigma\sqrt{n}\}} x^2\nu(dx) \to 0$$

by Lebesgue's Dominated Convergence Theorem; so that

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

by 273Ca. Thus Lindeberg's condition is satisfied and 274G gives the result.

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

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

Set

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

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

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

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

proof The condition

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

means that there are $c > 0$, $n_0 \in \mathbb{N}$ such that $s_n \ge c\sqrt{n+1}$ for every $n \ge n_0$. Let the underlying space be (Ω, Σ, μ), and take ϵ, $\eta > 0$. Writing $\psi_\delta(x) = 0$ for $|x| \le \delta$, x^2 for $|x| > \delta$, as in 274F-274G, we have

$$\mathbb{E}(\psi_{\epsilon s_n}(X_i)) \le \mathbb{E}(\psi_{c\epsilon\sqrt{n+1}}(X_i)) = \int_{F(i,c\epsilon\sqrt{n+1})} X_i^2 d\mu$$

for $n \ge n_0$, $i \le n$, where $F(i,\gamma) = \{\omega : \omega \in \operatorname{dom} X_i,\ |X_i(\omega)| > \gamma\}$. Because $\{X_i^2 : i \in \mathbb{N}\}$ is uniformly integrable, there is a $\gamma \ge 0$ such that $\int_{F(i,\gamma)} X_i^2 d\mu \le \eta c^2$ for every $i \in \mathbb{N}$ (246I). Let $n_1 \ge n_0$ be such that $c\epsilon\sqrt{n_1+1} \ge \gamma$; then for any $n \ge n_1$

$$\frac{1}{s_n^2}\sum_{i=0}^n \mathbb{E}(\psi_{\epsilon s_n}(X_i)) \le \frac{1}{c^2(n+1)}\sum_{i=0}^n \eta c^2 = \eta.$$

As ϵ and η are arbitrary, the conditions of 274G are satisfied and the result follows.

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

Set $s_n = \sqrt{\sum_{i=0}^n \operatorname{Var}(X_i)}$ 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 $a \in \mathbb{R}$.

proof The point is that $\{X_n^2 : n \in \mathbb{N}\}$ is uniformly integrable. **P** Set $K = 1 + \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{2+\delta})$. Given $\epsilon > 0$, set $M = (K/\epsilon)^{1/\delta}$. Then $(X_n^2 - M)^+ \le M^{-\delta}|X_n|^{2+\delta}$, so

$$\mathbb{E}(X_n^2 - M)^+ \le KM^{-\delta} = \epsilon$$

for every $n \in \mathbb{N}$. As ϵ is arbitrary, $\{X_n^2 : n \in \mathbb{N}\}$ is uniformly integrable. **Q**

Accordingly the conditions of 274J are satisfied and we have the result.

274L Remarks (a) All the theorems of this section are devoted to finding conditions under which a random variable S is 'nearly' standard normal, in the sense that $\Pr(S \le a) \doteq \Pr(Z \le a)$ uniformly for $a \in \mathbb{R}$, where Z is some (or any) standard normal random variable. In all cases the random variable S is normalized to have expectation 0 and variance 1, and is a sum of a large number of independent random variables. (In 274G and 274I-274K it is explicit that there must be many X_i, since they refer to a limit as $n \to \infty$. This is not said in so many words in the formulation I give of Lindeberg's theorem, but the proof makes it evident that $n(\delta + \delta^2) \ge 1$, so surely n will have to be large there also.)

(b) I cannot leave this section without remarking that the form of the definition of 'nearly standard normal' may lead your intuition astray if you try to apply it to other distributions. If we take F to be the distribution function of S, so that $F(a) = \Pr(S \le a)$, I am saying that S is 'nearly standard normal' if $\sup_{a\in\mathbb{R}} |F(a) - \Phi(a)|$ is small. It is natural to think of this as approximation in a metric, writing

$$\tilde\rho(\nu, \nu') = \sup_{a\in\mathbb{R}} |F_\nu(a) - F_{\nu'}(a)|$$

for distributions ν, ν' on $\mathbb{R}$, where $F_\nu(a) = \nu\,]-\infty, a]$. In this form, the theorems above can be read as finding conditions under which $\lim_{n\to\infty} \tilde\rho(\nu_{S_n}, \mu_G) = 0$. But the point is that $\tilde\rho$ is not really the right metric to use. It works here because μ_G is atomless. But suppose, for instance, that ν is the Dirac measure on $\mathbb{R}$ concentrated at 0, and that ν_n is the distribution of a normal random variable with expectation 0 and variance $\frac{1}{n}$, for each $n \ge 1$. Then $F_\nu(0) = 1$ and $F_{\nu_n}(0) = \frac{1}{2}$, so $\tilde\rho(\nu_n, \nu) = \frac{1}{2}$ for each $n \ge 1$. However, for most purposes one would regard the difference between ν_n and ν as small, and surely ν is the only distribution which one could reasonably call a limit of the ν_n.

(c) The difficulties here present themselves in more than one form. A statistician would be unhappy with the idea that the ν_n of (b) above were far from ν (and from each other), on the grounds that any measurement involving random variables with these distributions must be subject to error, and small errors of measurement will render them indistinguishable. A pure mathematician, looking forward to the possibility of generalizing these results, will be unhappy with the emphasis given to the values of $\nu\,]-\infty, a]$, for which it may be difficult to find suitable equivalents in more abstract spaces.

(d) These considerations join together to lead us to a rather different definition for a topology on the space P of probability distributions on $\mathbb{R}$. For any bounded continuous function $h : \mathbb{R} \to \mathbb{R}$ we have a pseudometric $\rho_h : P \times P \to [0, \infty[$ defined by writing

$$\rho_h(\nu, \nu') = |\int h\,d\nu - \int h\,d\nu'|$$

for all ν, $\nu' \in P$. The **vague topology** on P is that generated by the pseudometrics ρ_h (2A3F). I will not go into its properties in detail here (some are sketched in 274Yc-274Yf below; see also 285K-285L, 285S and 437J-437T and 454T-454V in Volume 4). But I maintain that the right way to look at the results of this chapter is to say that (i) the distributions ν_S are close to μ_G for the vague topology (ii) the sets $\{\nu : \tilde\rho(\nu, \mu_G) < \epsilon\}$ are open for that topology, and that is why $\tilde\rho(\nu_S, \mu_G)$ is small.

***274M** I conclude with a simple pair of inequalities which are frequently useful when studying normal random variables.

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 $x \ge 1$.

proof (a)

$$\int_x^\infty e^{-t^2/2}dt = \int_0^\infty e^{-(x+s)^2/2}ds \le e^{-x^2/2}\int_0^\infty e^{-xs}ds = \frac{1}{x}e^{-x^2/2}.$$

(b) Set

$$f(t) = e^{-t^2/2} - (1 - x(t-x))e^{-x^2/2}.$$

Then $f(x) = f'(x) = 0$ and $f''(t) = (t^2-1)e^{-t^2/2}$ is positive for $t \ge x$ (because $x \ge 1$). Accordingly $f(t) \ge 0$ for every $t \ge x$, and $\int_x^{x+1/x} f(t)dt \ge 0$. But this means just that

$$\int_x^\infty e^{-t^2/2}dt \ge \int_x^{x+\frac{1}{x}} e^{-t^2/2}dt \ge \int_x^{x+\frac{1}{x}} (1-x(t-x))e^{-x^2/2}dt = \frac{1}{2x}e^{-x^2/2},$$

as required.

274X Basic exercises >(a) Use 272U to give an alternative proof of 274B.

(b) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous on every closed bounded interval, and that $\int_{-\infty}^\infty |f'(x)|e^{-ax^2}dx < \infty$ for every $a > 0$. Let X be a normal random variable with zero expectation. Show that $\mathbb{E}(Xf(X))$ and $\mathbb{E}(X^2)\mathbb{E}(f'(X))$ are defined and equal.

(c) Prove 274D when h'' is M_3-Lipschitz but not necessarily differentiable.

(d) Let $\langle m_k\rangle_{k\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ such that $m_0 = 0$ and $\lim_{k\to\infty} m_k/m_{k+1} = 0$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables such that $\Pr(X_n = \sqrt{m_k}) = \Pr(X_n = -\sqrt{m_k}) = 1/2m_k$, $\Pr(X_n = 0) = 1 - 1/m_k$ whenever $m_{k-1} \le n < m_k$. Show that the Central Limit Theorem is not valid for $\langle X_n\rangle_{n\in\mathbb{N}}$. (*Hint*: setting $W_k = (X_0 + \ldots + X_{m_k-1})/\sqrt{m_k}$, show that $\Pr(W_k \in [\epsilon, 1-\epsilon]) \to 0$ for every $\epsilon > 0$.)

(e) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be any independent sequence of random variables all with the same distribution; suppose that they all have finite variance $\sigma^2 > 0$, and that their common expectation is c. Set $S_n = \frac{1}{\sqrt{n+1}}(X_0 + \ldots + X_n)$ for each n, and let Y be a normal random variable with expectation c and variance σ^2. Show that $\lim_{n\to\infty}\Pr(S_n \le a) = \Pr(Y \le a)$ uniformly for $a \in \mathbb{R}$.

>(f) Show that for any $a \in \mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{2^n}\sum_{r=0}^{\lfloor\frac{n}{2}+a\frac{\sqrt{n}}{2}\rfloor}\frac{n!}{r!(n-r)!} = \lim_{n\to\infty}\frac{1}{2^n}\#(\{I : I \subseteq n,\ \#(I) \le \frac{n}{2} + a\frac{\sqrt{n}}{2}\}) = \Phi(a).$$

(g) Show that 274I is a special case of 274J.

(h) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation. Set $s_n = \sqrt{\sum_{i=0}^n \mathrm{Var}(X_i)}$ and

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

for each $n \in \mathbb{N}$. Suppose that there is some $\delta > 0$ such that

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{i=0}^n \mathbb{E}(|X_i|^{2+\delta}) = 0.$$

Show that $\lim_{n\to\infty}\Pr(S_n \le a) = \Phi(a)$ uniformly for $a \in \mathbb{R}$. (This is a form of **Liapounoff's central limit theorem**; see Liapounoff 1901.)

(i) Let P be the set of Radon probability measures on $\mathbb{R}$. Let $\nu_0 \in P$, $a \in \mathbb{R}$. Show that the map $\nu \mapsto \nu\,]-\infty, a] : P \to [0,1]$ is continuous at ν_0 for the vague topology on P iff $\nu_0\{a\} = 0$.

(j) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent identically distributed sequence of random variables with non-zero finite variance. Let $\langle t_n \rangle_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that $\sum_{n=0}^{\infty} t_n^2 = \infty$. Show that $\sum_{n=0}^{\infty} t_n X_n$ is undefined or infinite a.e. (*Hint*: First deal with the case in which $\langle t_n \rangle_{n\in\mathbb{N}}$ does not converge to 0. Otherwise, use 274G to show that, for any $n \in \mathbb{N}$, $\lim_{m\to\infty} \Pr(|\sum_{i=n}^{m} t_i X_i| \ge 1) \ge \frac{1}{2}$). See also 276Xd.)

(k) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation. Suppose that $M \ge 0$ is such that $|X_n| \le M$ a.e. for every n, and that $\sum_{n=0}^{\infty} \operatorname{Var}(X_n) = \infty$. Set $s_n = \sqrt{\sum_{i=0}^{n} \operatorname{Var}(X_i)}$ for each n, and $S_n = \frac{1}{s_n} \sum_{i=0}^{n} X_i$ when $s_n > 0$. Show that $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ for every $a \in \mathbb{R}$.

274Y Further exercises (a) (Steele 86) Suppose that $X_0, \dots, X_n, Y_0, \dots, Y_n$ are independent random variables such that, for each $i \le n$, X_i and Y_i have the same distribution. Let $h : \mathbb{R}^{n+1} \to \mathbb{R}$ be a Borel measurable function, and set $Z = h(X_0, \dots, X_n)$, $Z_i = h(X_0, \dots, X_{i-1}, Y_i, X_{i+1}, \dots, X_n)$ for each i (with $Z_0 = h(Y_0, X_1, \dots, X_n)$ and $Z_n = h(X_0, \dots, X_{n-1}, Y_n)$, of course). Suppose that Z has finite expectation. Show that $\operatorname{Var}(Z) \le \frac{1}{2} \sum_{i=0}^{n} \mathbb{E}(Z_i - Z)^2$.

(b) Show that for any $\epsilon > 0$ there is a smooth function $h : \mathbb{R} \to [0,1]$ such that $\chi\,]-\infty, -\epsilon] \le h \le \chi\,[\epsilon, \infty[$.

(c) Write P for the set of Radon probability measures on $\mathbb{R}$. For ν, $\nu' \in P$ set

$$\rho(\nu, \nu') = \inf\{\epsilon : \epsilon \ge 0,\ \nu\,]-\infty, a-\epsilon] - \epsilon \le \nu'\,]-\infty, a] \le \nu\,]-\infty, a+\epsilon] + \epsilon \quad \text{for every } a \in \mathbb{R}\}.$$

Show that ρ is a metric on P and that it defines the vague topology on P. (ρ is called **Lévy's metric**.)

(d) Write P for the set of Radon probability measures on $\mathbb{R}$, and let $\tilde\rho$ be the metric on P defined in 274Lb. Show that if $\nu \in P$ is atomless and $\epsilon > 0$, then $\{\nu' : \nu' \in P, \tilde\rho(\nu', \nu) < \epsilon\}$ is open for the vague topology on P.

(e) Let $\langle S_n \rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables, and Z a standard normal random variable. Show that the following are equiveridical:

(i) $\mu_G = \lim_{n\to\infty} \nu_{S_n}$ for the vague topology, writing ν_{S_n} for the distribution of S_n;

(ii) $\mathbb{E}(h(Z)) = \lim_{n\to\infty} \mathbb{E}(h(S_n))$ for every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$;

(iii) $\mathbb{E}(h(Z)) = \lim_{n\to\infty} \mathbb{E}(h(S_n))$ for every bounded function $h : \mathbb{R} \to \mathbb{R}$ such that (α) h has continuous derivatives of all orders (β) $\{x : h(x) \ne 0\}$ is bounded;

(iv) $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ for every $a \in \mathbb{R}$;

(v) $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ uniformly for $a \in \mathbb{R}$;

(vi) $\{a : \lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)\}$ is dense in $\mathbb{R}$.

(See also 285L.)

(f) Let (Ω, Σ, μ) be a probability space and P the set of Radon probability measures on $\mathbb{R}$. Show that $X \mapsto \nu_X : \mathcal{L}^0(\mu) \to P$ is continuous for the topology of convergence in measure on $\mathcal{L}^0(\mu)$ and the vague topology on P.

(g) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables. Suppose that there is an $M \ge 0$ such that $|X_n| \le M$ a.e. for every $n \in \mathbb{N}$, and that $\sum_{n=0}^{\infty} X_n$ is defined, as a real number, almost everywhere. Show that $\sum_{n=0}^{\infty} \operatorname{Var}(X_n) < \infty$.

274 Notes and comments For more than two hundred years the Central Limit Theorem has been one of the glories of mathematics, and no branch of mathematics or science would be the same without it. I suppose it is the most important single theorem of probability theory; and I observe that the proof hardly uses measure theory. To be sure, I have clothed the arguments above in the language of measure and integration. But if you look at their essence, the vital elements of the proof are

(i) a linear combination of independent normal random variables is normal (274Ae, 274B);

(ii) if U, V, W are independent random variables, and h is a bounded continuous function, then $|\mathbb{E}(h(U,V,W))| \le \sup_{t\in\mathbb{R}} |\mathbb{E}(h(U,V,t))|$ (274C);

(iii) if $(X_0,\dots,X_n)$ are independent random variables, then we can find independent random variables $(X_0',\dots,X_n',Z_0,\dots,Z_n)$ such that Z_j is standard normal and X_j' has the same distribution as X_j, for each j (274F).

The rest of the argument consists of elementary calculus, careful estimations and a few of the most fundamental properties of expectations and independence. Now (ii) and (iii) are justified above by appeals to Fubini's theorem, but surely they belong to the list of probabilistic intuitions which take priority over the identification of probabilities with countably additive functionals. If they had given any insuperable difficulty it would have been a telling argument against the model of probability we were using, but would not have affected the Central Limit Theorem. In fact (i) seems to be the place where we really need a mathematical model of the concept of 'distribution', and all the relevant calculations can be done in terms of the Riemann integral on the plane, with no mention of countable additivity. So while I am happy and proud to have written out a version of these beautiful ideas, I have to admit that they are in no essential way dependent on the rest of this treatise.

In §285 I will describe a quite different approach to the theorem, using much more sophisticated machinery; but it will again be the case, perhaps more thoroughly hidden, that the relevance of measure theory will not be to the theorem itself, but to our imagination of what an arbitrary distribution is. For here I do have a claim to make for my subject. The characterization of distribution functions as arbitrary monotonic functions, continuous on the right, and with the correct limits at $\pm\infty$ (271Xb), together with the analysis of monotonic functions in §226, gives us a chance of forming a mental picture of the proper class of objects to which such results as the Central Limit Theorem can be applied.

Theorem 274F is a trifling modification of Theorem 3 of Lindeberg 1922. Like the original, it emphasizes what I believe to be vital to all the limit theorems of this chapter: they are best founded on a proper understanding of finite sequences of random variables. Lindeberg's condition was the culmination of a long search for the most general conditions under which the Central Limit Theorem would be valid. I offer a version of Laplace's theorem (274Xf) as the starting place, and Liapounoff's condition (274Xh) as an example of one of the intermediate stages. Naturally the corollaries 274I, 274J, 274K and 274Xe are those one seeks to apply by choice. There is an intriguing, but as far as I know purely coincidental, parallel between 273H/274K and 273I/274Xe. As an example of an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables, all with expectation zero and variance 1, to which the Central Limit Theorem does *not* apply, I offer 274Xd.

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 $\hat\Sigma$. (Such sequences $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ are called **filtrations**.) 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 $\int_E X_n=\int_E X_m$.

Note that for (ii) it is enough if $\int_E X_{n+1}=\int_E X_n$ whenever $n\in\mathbb{N}$ and $E\in\Sigma_n$.

275B Examples We have seen many contexts in which such sequences appear naturally; here are a few.

(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, as in §233. Subject to the conditions that $\operatorname{dom}X_n\in\Sigma_n$ and X_n is actually Σ_n-measurable for each n (a purely technical point – see 232He), $\langle X_n\rangle_{n\in\mathbb{N}}$ will be a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, because $\int_E X_{n+1}=\int_E X=\int_E X_n$ whenever $E\in\Sigma_n$.

(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 (272C), and set $S_n=X_0+\ldots+X_n$. Then $\langle S_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\tilde\Sigma_n$. (Use 272K to see that $\Sigma_{X_{n+1}}$ is independent of $\tilde\Sigma_n$, so that $\int_E X_{n+1}=\int X_{n+1}\times\chi E=0$ for every $E\in\tilde\Sigma_n$, by 272R.)

(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 (a) It seems appropriate to the concept of a random variable X being 'adapted' to a σ-algebra Σ to require that $\operatorname{dom}X\in\Sigma$ and that X should be Σ-measurable, even though this may mean that other random variables, equal almost everywhere to X, may fail to be 'adapted' to Σ.

(b) Technical problems of this kind evaporate, of course, if all μ-negligible subsets of X belong to Σ_0. But examples such as 275Bb make it seem unreasonable to insist on such a simplification as a general rule.

(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 $E\in\Sigma_i$. The principal case, after $I=\mathbb{N}$, is $I=[0,\infty[$; $I=\mathbb{Z}$ also is interesting, and I think it is fair to say that the most important ideas can already be expressed in theorems about martingales indexed by finite sets I. But in this volume I will generally take martingales to be indexed by $\mathbb{N}$.

(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 $\hat\Sigma$ (the completion of Σ) 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, as in 275Bb, then it is easy to see that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale iff it is a martingale adapted to $\langle\tilde\Sigma_n\rangle_{n\in\mathbb{N}}$.

(e) Continuing from (d), it is also easy to see that 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). (The point is that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, then both $\langle X_n\rangle_{n\in\mathbb{N}}$ and $\langle X'_n\rangle_{n\in\mathbb{N}}$ are adapted to $\langle\hat\Sigma_n\rangle_{n\in\mathbb{N}}$, where

$$\hat\Sigma_n=\{E\triangle F:E\in\Sigma_n,\ F\text{ is negligible}\}.)$$

Consequently we have a concept of 'martingale' as a sequence in $L^1(\mu)$, saying that a sequence $\langle X_n^\bullet\rangle_{n\in\mathbb{N}}$ in $L^1(\mu)$ is a martingale iff $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale.

Nevertheless, I think that the concept of 'martingale adapted to a sequence of σ-algebras' is the primary one, since in all the principal applications the σ-algebras reflect some essential aspect of the problem, which may not be fully encompassed by the random variables alone.

(f) The word 'martingale' originally (in English; the history in French is more complex) referred to a strap used to prevent a horse from throwing its head back. Later it was used as the name of a gambling system in which the gambler doubles his stake each time he loses, and (in French) as a general term for gambling systems. These may be regarded as a class of 'stopped-time martingales', as described in 275L-275P below.

275D A large part of the theory of martingales consists of inequalities of various kinds. I give two of the most important, both due to J.L.Doob. (See also 276Xa-276Xb.)

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 $X^* = \max(X_0, \dots , X_n)$. Then for any $\epsilon > 0$,

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

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

proof Write $\hat\mu$ for the completion of μ, and $\hat\Sigma$ for its domain. Let $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$ to which $\langle X_n \rangle_{n\in\mathbb{N}}$ is adapted. For each $i \le n$ set

$$E_i = \{\omega : \omega \in \operatorname{dom} X_i,\ X_i(\omega) \ge \epsilon\},$$

$$F_i = E_i \setminus \textstyle\bigcup_{j<i} E_j.$$

Then $F_0, \dots , F_n$ are disjoint and $F = \bigcup_{i\le n} F_i = \bigcup_{i\le n} E_i$; moreover, writing H for the conegligible set $\bigcap_{i\le n} \operatorname{dom} X_i$,

$$\{\omega : X^*(\omega) \ge \epsilon\} = F \cap H,$$

so that

$$\Pr(X^* \ge \epsilon) = \hat\mu\{\omega : X^*(\omega) \ge \epsilon\} = \hat\mu F = \textstyle\sum_{i=0}^n \hat\mu F_i.$$

On the other hand, E_i and F_i belong to Σ_i for each $i \le n$, so

$$\textstyle\int_{F_i} X_n = \int_{F_i} X_i \ge \epsilon\hat\mu F_i$$

for every i, and

$$\epsilon\hat\mu F = \epsilon\textstyle\sum_{i=0}^n \hat\mu F_i \le \sum_{i=0}^n \int_{F_i} X_n = \int_F X_n \le \int_F X_n^+ \le \mathbb{E}(X_n^+),$$

as required.

Remark Note that in fact we have $\epsilon\hat\mu F \le \int_F X_n$, where $F = \{\omega : X^*(\omega) \ge \epsilon\}$; this is of great importance in many applications.

275E Up-crossings The next lemma depends on the notion of 'up-crossing'. Let $x_0, \dots , x_n$ be any list of real numbers, and $a < b$ in $\mathbb{R}$. The **number of up-crossings from a to b** in the list $x_0, \dots , x_n$ 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 $j < i < k$. Note that this is also the largest m such that $s_m < \infty$, if we write

$$r_1 = \inf\{i : i \le n,\ x_i \le a\},$$

$$s_1 = \inf\{i : r_1 < i \le n,\ x_i \ge b\},$$

$$r_2 = \inf\{i : s_1 < i \le n,\ x_i \le a\},$$

$$s_2 = \inf\{i : r_2 < i \le n,\ x_i \ge b\}$$

and so on, taking $\inf\emptyset = \infty$.

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

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

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

proof Each individual step in the proof is 'elementary', but the structure as a whole is non-trivial.

(a) The following fact will be useful. Suppose that $x_0,\ldots,x_n$ are real numbers; let u be the number of up-crossings from a to b in the list $x_0,\ldots,x_n$. Set $y_i=\max(x_i,a)$ for each i; then u is also the number of up-crossings from a to b in the list $y_0,\ldots,y_n$. For each $k\le n$, set $c_k=1$ if there is a $j\le k$ such that $x_j\le a$ and $x_i<b$ for $j\le i\le k$, 0 otherwise. Then

$$(b-a)u\le\sum_{k=0}^{n-1}c_k(y_{k+1}-y_k).$$

P I induce on m to show that (defining r_m, s_m as in 275E)

$$(b-a)m\le\sum_{k=0}^{s_m-1}c_k(y_{k+1}-y_k)$$

whenever $m\le u$. For $m=0$ (taking $s_0=-1$) we have $0=0$. For the inductive step to $m\ge 1$, we have $s_{m-1}<r_m<s_m\le n$ (because I am supposing that $m\le u$), and $c_k=0$ if $s_{m-1}\le k<r_m$, $c_k=1$ if $r_m\le k<s_m$. So

$$\sum_{k=0}^{s_m-1}c_k(y_{k+1}-y_k)=\sum_{k=0}^{s_{m-1}-1}c_k(y_{k+1}-y_k)+\sum_{k=r_m}^{s_m-1}(y_{k+1}-y_k)$$
$$\ge(b-a)(m-1)+y_{s_m}-y_{r_m}$$

(by the inductive hypothesis)

$$\ge(b-a)m$$

(because $y_{s_m}\ge b$, $y_{r_m}=a$), and the induction proceeds.

Accordingly

$$\sum_{k=0}^{s_u-1}c_k(y_{k+1}-y_k)\ge(b-a)u.$$

As for the sum $\sum_{k=s_u}^{n-1}c_k(y_{k+1}-y_k)$, we have $c_k=0$ for $s_u\le k<r_{u+1}$, $c_k=1$ for $r_{u+1}\le k<s_{u+1}$, while $s_{u+1}>n$, so if $n\le r_{u+1}$ we have

$$\sum_{k=0}^{n-1}c_k(y_{k+1}-y_k)=\sum_{k=0}^{s_u-1}c_k(y_{k+1}-y_k)\ge(b-a)u,$$

while if $n>r_{u+1}$ we have

$$\sum_{k=0}^{n-1}c_k(y_{k+1}-y_k)=\sum_{k=0}^{s_u-1}c_k(y_{k+1}-y_k)+\sum_{k=r_{u+1}}^{n-1}y_{k+1}-y_k$$
$$\ge(b-a)u+y_n-y_{r_{u+1}}$$
$$\ge(b-a)u$$

because $y_n\ge a=y_{r_{u+1}}$. Thus in both cases we have the required result. **Q**

(b)(i) Now define

$$Y_k(\omega)=\max(a,X_k(\omega))\text{ for }\omega\in\operatorname{dom}X_k,$$

$$F_k=\{\omega:\omega\in\bigcap_{i\le k}\operatorname{dom}X_i,\ \exists\,j\le k,\ X_j(\omega)\le a,\ X_i(\omega)<b\text{ if }j\le i\le k\}$$

for each $k\in\mathbb{N}$. If $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-algebras to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted, then $F_k\in\Sigma_k$ (because if $j\le k$ all the sets $\operatorname{dom}X_j$, $\{\omega:X_j(\omega)\le a\}$, $\{\omega:X_j(\omega)<b\}$ belong to $\Sigma_j\subseteq\Sigma_k$).

(ii) We find that $\int_F Y_k\le\int_F Y_{k+1}$ if $F\in\Sigma_k$. **P** Set $G=\{\omega:X_k(\omega)>a\}\in\Sigma_k$. Then

$$\int_F Y_k=\int_{F\cap G}X_k+a\hat\mu(F\setminus G)$$
$$=\int_{F\cap G}X_{k+1}+a\hat\mu(F\setminus G)$$
$$\le\int_{F\cap G}Y_{k+1}+\int_{F\setminus G}Y_{k+1}=\int_F Y_{k+1}.\ \mathbf{Q}$$

(iii) Consequently $\int_F Y_{k+1} - Y_k \le \int Y_{k+1} - Y_k$ for every $F \in \Sigma_k$.

$$\mathbf{P}\quad \int (Y_{k+1} - Y_k) - \int_F (Y_{k+1} - Y_k) = \int_{\Omega\setminus F} Y_{k+1} - \int_{\Omega\setminus F} Y_k \ge 0. \quad \mathbf{Q}$$

(c) Let H be the conegligible set $\operatorname{dom} U = \bigcap_{i\le n} \operatorname{dom} X_i \in \Sigma_n$. We ought to check at some point that U is Σ_n-measurable; but this is clearly true, because all the relevant sets $\{\omega : X_i(\omega) \le a\}$, $\{\omega : X_i(\omega) \ge b\}$ belong to Σ_n. For each $\omega \in H$, apply (a) to the list $X_0(\omega), \dots , X_n(\omega)$ to see that

$$(b-a)U(\omega) \le \sum_{k=0}^{n-1} \chi F_k(\omega)(Y_{k+1}(\omega) - Y_k(\omega)).$$

Because H is conegligible, it follows that

$$(b-a)\mathbb{E}(U) \le \sum_{k=0}^{n-1} \int_{F_k} Y_{k+1} - Y_k \le \sum_{k=0}^{n-1} \int Y_{k+1} - Y_k$$

(using (b-iii))

$$= \mathbb{E}(Y_n - Y_0) \le \mathbb{E}((X_n - X_0)^+)$$

because $Y_n - Y_0 \le (X_n - X_0)^+$ everywhere on $\operatorname{dom} X_n \cap \operatorname{dom} X_0$. This completes the proof.

275G We are now ready for the principal theorems of this section.

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.

proof (a) Set $H = \bigcap_{n\in\mathbb{N}} \operatorname{dom} X_n$, and for $\omega \in H$ set

$$Y(\omega) = \liminf_{n\to\infty} X_n(\omega), \quad Z(\omega) = \limsup_{n\to\infty} X_n(\omega),$$

allowing $\pm\infty$ in both cases. But note that $Y \le \liminf_{n\to\infty} |X_n|$, so by Fatou's Lemma $Y(\omega) < \infty$ for almost every ω; similarly $Z(\omega) > -\infty$ for almost every ω. It will therefore be enough if I can show that $Y =_{\text{a.e.}} Z$, for then $Y(\omega) = Z(\omega) \in \mathbb{R}$ for almost every ω, and $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ will be convergent for almost every ω.

(b) ? So suppose, if possible, that Y and Z are not equal almost everywhere. Of course both are $\hat{\Sigma}$-measurable, where $(\Omega, \hat{\Sigma}, \hat{\mu})$ is the completion of (Ω, Σ, μ), so we must have

$$\hat{\mu}\{\omega : \omega \in H,\, Y(\omega) < Z(\omega)\} > 0.$$

Accordingly there are rational numbers q, q' such that $q < q'$ and $\hat{\mu}G > 0$, where

$$G = \{\omega : \omega \in H,\, Y(\omega) < q < q' < Z(\omega)\}.$$

Now, for each $\omega \in H$ and $n \in \mathbb{N}$, let $U_n(\omega)$ be the number of up-crossings from q to q' in the list $X_0(\omega), \dots ,$ $X_n(\omega)$. Then 275F tells us that

$$\mathbb{E}(U_n) \le \frac{1}{q'-q}\mathbb{E}((X_n - X_0)^+) \le \frac{1}{q'-q}\mathbb{E}(|X_n| + |X_0|) \le \frac{2M}{q'-q},$$

if we write $M = \sup_{i\in\mathbb{N}} \mathbb{E}(|X_i|)$. By B.Levi's theorem, $U(\omega) = \sup_{n\in\mathbb{N}} U_n(\omega) < \infty$ for almost every ω. On the other hand, if $\omega \in G$, then there are arbitrarily large j, k such that $X_j(\omega) < q$ and $X_k(\omega) > q'$, so $U(\omega) = \infty$. This means that $\hat{\mu}G$ must be 0, contrary to the choice of q and q'. **X**

(c) Thus we must in fact have $Y =_{\text{a.e.}} Z$, and $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω, as claimed.

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 $\mathbb{E}(|X_\infty|) = \lim_{n\to\infty} \mathbb{E}(|X_n|) < \infty$.

proof (i)⇒(ii) By 246D, the set of all conditional expectations of X is uniformly integrable, so $\{X_n : n \in \mathbb{N}\}$ is surely uniformly integrable.

(ii)⇒(iii) If $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable, we surely have $\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|) < \infty$, so 275G tells us that X_∞ is defined almost everywhere. By 246Ja, X_∞ is integrable and $\lim_{n\to\infty} \mathbb{E}(|X_n - X_\infty|) = 0$. Consequently $\mathbb{E}(|X_\infty|) = \lim_{n\to\infty} \mathbb{E}(|X_n|) < \infty$.

(iii)⇒(i) Because $\mathbb{E}(|X_\infty|) = \lim_{n\to\infty} \mathbb{E}(|X_n|)$, $\lim_{n\to\infty} \mathbb{E}(|X_n - X_\infty|) = 0$ (245H(a-ii)). Now take $n \in \mathbb{N}$ and $E \in \Sigma_n$. Then

$$\int_E X_n = \lim_{m\to\infty} \int_E X_m = \int_E X_\infty.$$

As E is arbitrary, X_n is a conditional expectation of X_∞ on Σ_n.

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_∞.

proof By 275G-275H, we know that X_∞ is defined almost everywhere, and, as remarked in the proof of 275H, $\lim_{n\to\infty} \mathbb{E}(|X_\infty - X_n|) = 0$. To see that X_∞ is a conditional expectation of X on Σ_∞, set

$$\mathcal{A} = \{E : E \in \Sigma_\infty, \int_E X_\infty = \int_E X\}, \quad \mathcal{I} = \bigcup_{n\in\mathbb{N}} \Sigma_n.$$

Now $\mathcal{I}$ and $\mathcal{A}$ satisfy the conditions of the Monotone Class Theorem (136B). **P** **(α)** Of course $\Omega \in \mathcal{I}$ and $\mathcal{I}$ is closed under finite intersections, because $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-algebras; in fact $\mathcal{I}$ is a subalgebra of $\mathcal{P}\Omega$, and is closed under finite unions and complements. **(β)** If $E \in \mathcal{I}$, say $E \in \Sigma_n$; then

$$\int_E X_\infty = \lim_{m\to\infty} \int_E X_m = \int_E X,$$

as in (iii)⇒(i) of 275H, so $E \in \mathcal{A}$. Thus $\mathcal{I} \subseteq \mathcal{A}$. **($\gamma$)** If E, $F \in \mathcal{A}$ and $E \subseteq F$, then

$$\int_{F\setminus E} X_\infty = \int_F X_\infty - \int_E X_\infty = \int_F X - \int_E X = \int_{F\setminus E} X,$$

so $F \setminus E \in \mathcal{A}$. **($\delta$)** If $\langle E_k \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $\mathcal{A}$ with union E, then

$$\int_E X_\infty = \lim_{k\to\infty} \int_{E_k} X_\infty = \lim_{k\to\infty} \int_{E_k} X = \int_E X,$$

so $E \in \mathcal{A}$. Thus $\mathcal{A}$ is a Dynkin class. **Q**

Consequently, by 136B, $\mathcal{A}$ includes Σ_∞; that is, X_∞ is a conditional expectation of X on Σ_∞.

Remark I have written '$\lim_{n\to\infty} \mathbb{E}(|X_n - X_\infty|) = 0$'; but you may prefer to say '$X_\infty^\bullet = \lim_{n\to\infty} X_n^\bullet$ in $L^1(\mu)$', as in Chapter 24.

The importance of this theorem is such that you may be interested in a proof based on 275D rather than 275E-275G; see 275Xd.

***275J** As a corollary of this theorem I give an important result, a kind of density theorem for product measures.

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

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

wherever this is defined, where I write '$\int \ldots d(\xi_{n+1}, \ldots)$' 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 $\boldsymbol{\omega} = (\omega_0, \omega_1, \ldots)$ in Ω, and $\lim_{n\to\infty} \mathbb{E}(|X - X_n|) = 0$.

proof For each n, we can identify μ with the product of λ_n and λ'_n, where λ_n is the product measure on $\Omega_0 \times \ldots \times \Omega_n$ (254N). So 253H tells us that X_n is a conditional expectation of X on the σ-algebra $\Lambda_n = \{E \times \prod_{i>n} \Omega_i : E \in \operatorname{dom} \lambda_n\}$. Since (by 254N again) we can think of λ_{n+1} as the product of λ_n and μ_{n+1}, $\Lambda_n \subseteq \Lambda_{n+1}$ for each n. So 275I tells us that $\langle X_n \rangle_{n\in\mathbb{N}}$ converges almost everywhere to a conditional expectation X_∞ of X on the σ-algebra Λ_∞ generated by $\bigcup_{n\in\mathbb{N}}\Lambda_n$. Now $\Lambda_\infty\subseteq\Sigma$ and also $\widehat{\bigotimes}_{n\in\mathbb{N}}\Sigma_n\subseteq\Lambda_\infty$, so every member of Σ is sandwiched between two members of Λ_∞ of the same measure (254Ff), and X_∞ must be equal to X almost everywhere. Moreover, 275I also tells us that

$$\lim_{n\to\infty}\mathbb{E}(|X-X_n|)=\lim_{n\to\infty}\mathbb{E}(|X_\infty-X_n|)=0,$$

as required.

275K Reverse martingales We have a result corresponding to 275I for *decreasing* sequences of σ-algebras. While this is used less often than 275G-275I, it does have very important applications.

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_∞.

proof (a) Set $H=\bigcap_{n\in\mathbb{N}}\operatorname{dom}X_n$, so that H is conegligible. For $n\in\mathbb{N}$, $a<b$ in $\mathbb{R}$, and $\omega\in H$, write $U_{abn}(\omega)$ for the number of up-crossings from a to b in the list $X_n(\omega),X_{n-1}(\omega),\dots,X_0(\omega)$ (275E). Then

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

(275F)

$$\le\frac{1}{b-a}\mathbb{E}(|X_0|+|X_n|)\le\frac{2}{b-a}\mathbb{E}(|X_0|)<\infty.$$

So $\lim_{n\to\infty}U_{abn}(\omega)$ is finite for almost every ω. But this means that

$$\{\omega:\liminf_{n\to\infty}X_n(\omega)<a,\,\limsup_{n\to\infty}X_n(\omega)>b\}$$

is negligible. As a and b are arbitrary, $\langle X_n\rangle_{n\in\mathbb{N}}$ is convergent a.e., just as in 275G. Set $X_\infty(\omega)=\lim_{n\to\infty}X_n(\omega)$ whenever this is defined in $\mathbb{R}$.

(b) By 246D, $\{X_n:n\in\mathbb{N}\}$ is uniformly integrable, so $\mathbb{E}(|X_n-X_\infty|)\to 0$ as $n\to\infty$ (246Ja), and

$$\int_E X_\infty=\lim_{n\to\infty}\int_E X_n=\int_E X_0$$

for every $E\in\Sigma_\infty$.

(c) Now there is a conegligible set $G\in\Sigma_\infty$ such that $G\subseteq\operatorname{dom}X_\infty$ and $X_\infty\restriction G$ is Σ_∞-measurable. **P** For each $n\in\mathbb{N}$, there is a conegligible set $G_n\in\Sigma_n$ such that $G_n\subseteq\operatorname{dom}X_n$ and $X_n\restriction G_n$ is Σ_n-measurable. Set $G'=\bigcup_{n\in\mathbb{N}}\bigcap_{m\ge n}G_m$; then, for any $r\in\mathbb{N}$, $G'=\bigcup_{n\ge r}\bigcap_{m\ge n}G_m$ belongs to Σ_r, so $G'\in\Sigma_\infty$, while of course G' is conegligible. For $n\in\mathbb{N}$, set $X'_n(\omega)=X_n(\omega)$ for $\omega\in G_n$, 0 for $\omega\in\Omega\setminus G_n$; then for $\omega\in G'$, $\lim_{n\to\infty}X'_n(\omega)=\lim_{n\to\infty}X_n(\omega)$ if either is defined in $\mathbb{R}$. Writing $X'_\infty=\lim_{n\to\infty}X'_n$ whenever this is defined in $\mathbb{R}$, 121F and 121H tell us that X'_∞ is Σ_r-measurable and $\operatorname{dom}X'_\infty\in\Sigma_r$ for every $r\in\mathbb{N}$, so that $G''=\operatorname{dom}X'_\infty$ belongs to Σ_∞ and X'_∞ is Σ_∞-measurable. We also know, from (a), that G'' is conegligible. So setting $G=G'\cap G''$ we have the result. **Q**

Thus X_∞ is a conditional expectation of X on Σ_∞.

275L Stopping times In a sense, the main work of this section is over; I have no room for any more theorems of importance comparable to 275G-275I. However, it would be wrong to leave this chapter without briefly describing one of the most fruitful ideas of the subject.

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** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ (also called '**optional time**', '**Markov time**') is a function τ from Ω to $\mathbb{N}\cup\{\infty\}$ such that $\{\omega:\tau(\omega)\le n\}\in\Sigma_n$ for every $n\in\mathbb{N}$.

Remark Of course the condition

$$\{\omega:\tau(\omega)\le n\}\in\Sigma_n\text{ for every }n\in\mathbb{N}$$

can be replaced by the equivalent condition

$$\{\omega:\tau(\omega)=n\}\in\Sigma_n\text{ for every }n\in\mathbb{N}.$$

I give priority to the former expression because it is more appropriate to other index sets (see 275Cc).

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

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

setting $\inf\emptyset=\infty$ as usual. (For by 121Ka the set $E_n=\{\omega:(X_0(\omega),\dots,X_n(\omega))\in H_n\}$ belongs to Σ_n for each n, and $\{\omega:\tau(\omega)\le n\}=\bigcup_{i\le n}E_i$.) In particular, for instance, 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 $\omega\in\Omega$, because

$$\{\omega:(\tau\wedge\tau')(\omega)\le n\}=\{\omega:\tau(\omega)\le n\}\cup\{\omega:\tau'(\omega)\le n\}\in\Sigma_n$$

for every $n\in\mathbb{N}$.

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 $\tilde\Sigma_\tau$.

proof (a) This is elementary. Write $H_n=\{\omega:\tau(\omega)\le n\}$ for each $n\in\mathbb{N}$. The empty set belongs to $\tilde\Sigma_\tau$ because it belongs to Σ_n for every n. If $E\in\tilde\Sigma_\tau$, then

$$(\Omega\setminus E)\cap H_n=H_n\setminus(E\cap H_n)\in\Sigma_n$$

because $H_n\in\Sigma_n$; this is true for for every n, so $X\setminus E\in\tilde\Sigma_\tau$. If $\langle E_k\rangle_{k\in\mathbb{N}}$ is any sequence in $\tilde\Sigma_\tau$ then

$$(\bigcup_{k\in\mathbb{N}}E_k)\cap H_n=\bigcup_{k\in\mathbb{N}}E_k\cap H_n\in\Sigma_n$$

for every n, so $\bigcup_{k\in\mathbb{N}}E_k\in\tilde\Sigma_\tau$.

(b) If $E\in\tilde\Sigma_\tau$ then of course $E\in\Sigma$, and if $n\in\mathbb{N}$ then $\{\omega:\tau'(\omega)\le n\}\subseteq\{\omega:\tau(\omega)\le n\}$, so that

$$E\cap\{\omega:\tau'(\omega)\le n\}=E\cap\{\omega:\tau(\omega)\le n\}\cap\{\omega:\tau'(\omega)\le n\}$$

belongs to Σ_n; as n is arbitrary, $E\in\tilde\Sigma_{\tau'}$.

(c) Set $H_n=\{\omega:\tau(\omega)\le n\}$ for each $n\in\mathbb{N}$. For any $a\in\mathbb{R}$,

$$\begin{aligned}H_n\cap\{\omega:\omega\in\operatorname{dom}\tilde X_\tau,\ \tilde X_\tau(\omega)\le a\}\\ =\bigcup_{k\le n}\{\omega:\tau(\omega)=k,\ \omega\in\operatorname{dom}X_k,\ X_k(\omega)\le a\}\in\Sigma_n.\end{aligned}$$

As n is arbitrary,

$$G_a = \{\omega : \omega \in \operatorname{dom} \tilde{X}_\tau,\ \tilde{X}_\tau(\omega) \le a\} \in \tilde{\Sigma}_\tau.$$

As a is arbitrary, $\operatorname{dom} \tilde{X}_\tau = \bigcup_{m\in\mathbb{N}} G_m \in \tilde{\Sigma}_\tau$ and $\tilde{X}_\tau$ is $\tilde{\Sigma}_\tau$-measurable.

(d) Set $H_k = \{\omega : \tau(\omega) = k\}$ for $k \le m$. Then $\bigcup_{k\le m} H_k$ is conegligible, so

$$\mathbb{E}(\tilde{X}_\tau) = \sum_{k=0}^m \int_{H_k} X_k = \sum_{k=0}^m \int_{H_k} X_m = \int_\Omega X_m = \int_\Omega X_0.$$

(e) Suppose $\tau' \le n$ almost everywhere. Set $H_k = \{\omega : \tau(\omega) = k\}$, $H'_k = \{\omega : \tau'(\omega) = k\}$ for each k; then both $\langle H_k\rangle_{k\le n}$ and $\langle H'_k\rangle_{k\le n}$ are partitions of conegligible subsets of X. Now suppose that $E \in \tilde{\Sigma}_\tau$. Then

$$\int_E \tilde{X}_\tau = \sum_{k=0}^n \int_{E\cap H_k} \tilde{X}_\tau = \sum_{k=0}^n \int_{E\cap H_k} X_k = \sum_{k=0}^n \int_{E\cap H_k} X_n = \int_E X_n$$

because $E \cap H_k \in \Sigma_k$ for every k. By (b), $E \in \tilde{\Sigma}_{\tau'}$, so we also have $\int_E \tilde{X}_{\tau'} = \int_E X_n$. Thus $\int_E \tilde{X}_\tau = \int_E \tilde{X}_{\tau'}$ for every $E \in \tilde{\Sigma}_\tau$, as claimed.

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.

proof As remarked in 275Mb, each $\tau \wedge n$ is a stopping time. If $m \le n$, then $\tilde{\Sigma}_{\tau\wedge m} \subseteq \tilde{\Sigma}_{\tau\wedge n}$ by 275Nb. Each $\tilde{X}_{\tau\wedge m}$ is $\tilde{\Sigma}_{\tau\wedge m}$-measurable, with domain belonging to $\tilde{\Sigma}_{\tau\wedge m}$, by 275Nc, and has finite expectation, by 275Nd; finally, if $m \le n$, then $\tilde{X}_{\tau\wedge m}$ is a conditional expectation of $\tilde{X}_{\tau\wedge n}$ on $\tilde{\Sigma}_{\tau\wedge m}$, by 275Ne.

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 $\inf_{n\in\mathbb{N}} X_n(\omega) = -\infty$.

proof Let $\langle \Sigma_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-algebras to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted. Let H be the conegligible set $\bigcap_{n\in\mathbb{N}} \operatorname{dom} X_n \cap \{\omega : W(\omega) < \infty\}$. For each $m \in \mathbb{N}$, set

$$\tau_m(\omega) = \inf\{n : \omega \in \operatorname{dom} X_n,\ X_n(\omega) > m\}.$$

As in 275Ma, τ_m is a stopping time adapted to $\langle \Sigma_n\rangle_{n\in\mathbb{N}}$. Set

$$Y_{mn} = \tilde{X}_{\tau_m\wedge n},$$

defined as in 275O, so that $\langle Y_{mn}\rangle_{n\in\mathbb{N}}$ is a martingale. If $\omega \in H$, then either $\tau_m(\omega) > n$ and

$$Y_{mn}(\omega) = X_n(\omega) \le m,$$

or $0 < \tau_m(\omega) \le n$ and

$$Y_{mn}(\omega) = X_{\tau_m(\omega)}(\omega) \le W(\omega) + X_{\tau_m(\omega)-1}(\omega) \le W(\omega) + m,$$

or $\tau_m(\omega) = 0$ and

$$Y_{mn}(\omega) = X_0(\omega).$$

Thus

$$Y_{mn}(\omega) \le |X_0(\omega)| + W(\omega) + m$$

for every $\omega \in H$, and

$$|Y_{mn}(\omega)| = 2\max(0, Y_{mn}(\omega)) - Y_{mn}(\omega) \le 2(|X_0(\omega)| + W(\omega) + m) - Y_{mn}(\omega),$$

$$\mathbb{E}(|Y_{mn}|) \le 2\mathbb{E}(|X_0|) + 2\mathbb{E}(W) + 2m - \mathbb{E}(Y_{mn}) = 2\mathbb{E}(|X_0|) + 2\mathbb{E}(W) + 2m - \mathbb{E}(X_0)$$

by 275Nd. As this is true for every $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \mathbb{E}(|Y_{mn}|) < \infty$, and $\lim_{n\to\infty} Y_{mn}$ is defined in $\mathbb{R}$ almost everywhere, by Doob's Martingale Convergence Theorem (275G). Let F_m be the conegligible set on which $\langle Y_{mn}\rangle_{n\in\mathbb{N}}$ converges. Set $H^* = H \cap \bigcap_{m\in\mathbb{N}} F_m$, so that H^* is conegligible.

Now consider

$$E = \{\omega : \omega \in H^*,\ \sup_{n\in\mathbb{N}} X_n(\omega) < \infty\}.$$

For any $\omega \in E$, there must be an $m \in \mathbb{N}$ such that $\sup_{n\in\mathbb{N}} X_n(\omega) \le m$. Now this means that $Y_{mn}(\omega) = X_n(\omega)$ for every n, and as $\omega \in F_m$ we have

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} Y_{mn}(\omega) \in \mathbb{R}.$$

This means that $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω such that $\{X_n(\omega) : n \in \mathbb{N}\}$ is bounded above.

Similarly, $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω such that $\{X_n(\omega) : n \in \mathbb{N}\}$ is bounded below, which completes the proof.

275X Basic exercises >(a) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables with zero expectation and finite variance. Set $s_n = (\sum_{i=0}^n \operatorname{Var}(X_i))^{1/2}$, $Y_n = (X_0 + \ldots + X_n)^2 - s_n^2$ for each n. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale.

>(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale. Show that for any $\epsilon > 0$, $\Pr(\sup_{n\in\mathbb{N}} |X_n|) \ge \epsilon) \le \frac{1}{\epsilon} \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|)$.

(c) Pólya's urn scheme Imagine a box containing red and white balls. At each move, a ball is drawn at random from the box and replaced together with another of the same colour. (i) Writing R_n, W_n for the numbers of red and white balls after the nth move and $X_n = R_n/(R_n + W_n)$, show that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale. (ii) Starting from $R_0 = W_0 = 1$, find the distribution of (R_n, W_n) for each n. (iii) Show that $X = \lim_{n\to\infty} X_n$ is defined almost everywhere, and find its distribution when $R_0 = W_0 = 1$. (See Feller 66 for a discussion of other starting values.)

>(d) 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; for each $n \in \mathbb{N}$ let $P_n : L^1 \to L^1$ be the conditional expectation operator corresponding to Σ_n, where $L^1 = L^1(\mu)$ (242J). (i) Show that $V = \{u : u \in L^1, \lim_{n\to\infty} \|P_n u - u\|_1 = 0\}$ is a $\|\ \|_1$-closed linear subspace of L^1. (ii) Show that $\{E : E \in \Sigma, \chi E^\bullet \in V\}$ is a Dynkin class including $\bigcup_{n\in\mathbb{N}} \Sigma_n$, so includes the σ-algebra Σ_∞ generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$. (iii) Show that if $u \in L^1$ then $v = \sup_{n\in\mathbb{N}} P_n|u|$ is defined in L^1 and is of the form $W^\bullet$ where $\Pr(W \ge \epsilon) \le \frac{1}{\epsilon}\|u\|_1$ for every $\epsilon > 0$. (*Hint*: 275D.) (iv) Show that if X is a Σ_∞-measurable random variable with finite expectation, and for each $n \in \mathbb{N}$ X_n is a conditional expectation of X on Σ_n, then $X^\bullet \in V$ and $X =_{\text{a.e.}} \lim_{n\to\infty} X_n$. (*Hint*: apply (iii) to $u = (X - X_m)^\bullet$ for large m.)

(e) Let (Ω, Σ, μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ, and Σ_∞ the σ-algebra generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$. For each $n \in \mathbb{N} \cup \{\infty\}$ let $P_n : L^1 \to L^1$ be the conditional expectation operator corresponding to Σ_n, where $L^1 = L^1(\mu)$. Show that $\lim_{n\to\infty} \|P_n u - P_\infty u\|_p = 0$ whenever $p \in [1, \infty[$ and $u \in L^p(\mu)$. (*Hint*: 275Xd, 233J/242K, 246Xg.)

(f) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale, and suppose that $p \in]1, \infty[$ is such that $\sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$. Show that $X = \lim_{n\to\infty} X_n$ is defined almost everywhere and that $\lim_{n\to\infty} \|X_n - X\|_p = 0$.

>(g) Let (Ω, Σ, μ) be $[0, 1]$ with Lebesgue measure. For each $n \in \mathbb{N}$ let Σ_n be the finite subalgebra of Σ generated by intervals of the type $[0, 2^{-n}r]$ for $r \le 2^{-n}$. Use 275I to show that for any integrable $X : [0, 1] \to \mathbb{R}$ we must have $X(t) = \lim_{n\to\infty} 2^n \int_{I_n(t)} X$ for almost every $t \in [0, 1[$, where $I_n(t)$ is the interval of the form $[2^{-n}r, 2^{-n}(r + 1)[$ containing t. Compare this result with 223A and 261Yd.

(h) In 275K, show that $\lim_{n\to\infty} \|X_n - X_\infty\|_p = 0$ for any $p \in [1, \infty[$ such that $\|X_0\|_p$ is finite. (Compare 275Xe.)

(i) 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 $\hat\Sigma$. Show that if $\langle\tau_i\rangle_{i\in\mathbb{N}}$ is a sequence of stopping times adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, and we set $\tau(\omega) = \sup_{i\in\mathbb{N}} \tau_i(\omega)$ for $\omega \in \Omega$, then τ is a stopping time adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(j) 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 $\hat\Sigma$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable martingale adapted to Σ_n, and set $X_\infty = \lim_{n\to\infty} X_n$. Let τ be a stopping time adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, and set $\tilde X_\tau(\omega) = X_{\tau(\omega)}(\omega)$ whenever $\omega \in \operatorname{dom} X_{\tau(\omega)}$, allowing ∞ as a value of $\tau(\omega)$. Show that $\tilde X_\tau$ is a conditional expectation of X_∞ on $\tilde\Sigma_\tau$, as defined in 275N.

(k) 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 $\hat{\Sigma}$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale and τ a stopping time, both adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)<\infty$ and that τ is finite almost everywhere. Show that $\tilde{X}_\tau$, as defined in 275Nc, has finite expectation, but that $\mathbb{E}(\tilde{X}_\tau)$ need not be equal to $\mathbb{E}(X_0)$.

(l)(i) Show that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale such that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)$ is finite and $\langle X_n\rangle_{n\in\mathbb{N}}$ is convergent in measure, then $\langle X_n\rangle_{n\in\mathbb{N}}$ is convergent a.e. (ii) Find a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\langle X_{2n}\rangle_{n\in\mathbb{N}}\to 0$ a.e. but $|X_{2n+1}|\ge 1$ a.e. for every $n\in\mathbb{N}$. (iii) Find a martingale which converges in measure but is not convergent a.e.

275Y Further exercises (a) Let (Ω,Σ,μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ an independent sequence of σ-subalgebras of Σ, and X a random variable on Ω with finite variance. Let X_n be a conditional expectation of X on Σ_n for each n. Show that $\lim_{n\to\infty}X_n=\mathbb{E}(X)$ almost everywhere. (*Hint*: consider $\sum_{n=0}^\infty\mathrm{Var}(X_n)$.)

(b) Let (Ω,Σ,μ) be a complete probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ all containing every negligible set, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Let ν be another probability measure with domain Σ which is absolutely continuous with respect to μ, with Radon-Nikodým derivative Z. For each $n\in\mathbb{N}$ let Z_n be a conditional expectation of Z on Σ_n (with respect to the measure μ). (i) Show that Z_n is a Radon-Nikodým derivative of $\nu\restriction\Sigma_n$ with respect to $\mu\restriction\Sigma_n$, for each $n\in\mathbb{N}$. (ii) Defining X_n/Z_n as in 121E, so that its domain is $\{\omega:\omega\in\operatorname{dom}X_n\cap\operatorname{dom}Z_n,\ Z_n(\omega)\ne 0\}$, show that $\langle X_n/Z_n\rangle_{n\in\mathbb{N}}$ is a martingale with respect to the measure ν.

(c) Combine the ideas of 275Cc with those of 275Cd-275Ce to describe a notion of 'martingale indexed by I', where I is an arbitrary partially ordered set.

(d) Let $\langle X_k\rangle_{k\in\mathbb{N}}$ be a martingale on a complete probability space (Ω,Σ,μ), and fix $n\in\mathbb{N}$. Set $X^*=\max(|X_0|,\dots,|X_n|)$. Let $p\in\,]1,\infty[$. Show that $\|X^*\|_p\le\frac{p}{p-1}\|X_n\|_p$. (*Hint*: set $F_t=\{\omega:X^*(\omega)\ge t\}$. Show that $t\mu F_t\le\int_{F_t}|X_n|$. Using Fubini's theorem on $\Omega\times[0,\infty[$ and on $\Omega\times[0,\infty[\,\times[0,\infty[$, show that

$$\mathbb{E}((X^*)^p)=p\int_0^\infty t^{p-1}\mu F_t dt,$$

$$\int_0^\infty t^{p-2}\int_{F_t}|X_n|dt=\frac{1}{p-1}\mathbb{E}(|X_n|\times(X^*)^{p-1}),$$

$$\mathbb{E}(|X_n|\times(X^*)^{p-1})\le\|X_n\|_p\|X^*\|_p^{p-1}.$$

Compare 286A below.)

(e)(i) Show that if $a,\ b\ge 0$ then $a\ln^+b\le a\ln^+a+\frac{b}{e}$, where $\ln^+t=0$ if $t\le 1$, $\ln t$ if $t\ge 1$. (ii) Let (Ω,Σ,μ) be a complete probability space and X, Y non-negative random variables on Ω such that $t\mu F_t\le\int_{F_t}X$ for every $t\ge 0$, where $F_t=\{\omega:Y(\omega)\ge t\}$. Show that $\int_{F_1}Y\le\int_{F_1}X\times\ln^+Y$, and hence that $\mathbb{E}(Y)\le\frac{e}{e-1}(1+\mathbb{E}(X\times\ln^+X))$. (iii) Show that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale on Ω, $n\in\mathbb{N}$ and $X^*=\sup_{i\le n}|X_i|$, then $\mathbb{E}(X^*)\le\frac{e}{e-1}(1+\mathbb{E}(|X_n|\times\ln^+|X_n|))$.

(f) Let (Ω,Σ,μ) be a probability space and $\langle\Sigma_i\rangle_{i\in I}$ a countable family of σ-subalgebras of Σ such that for any $i,\ j\in I$ either $\Sigma_i\subseteq\Sigma_j$ or $\Sigma_j\subseteq\Sigma_i$. Let X be a real-valued random variable on Ω such that $\|X\|_p<\infty$, where $1<p<\infty$, and suppose that X_i is a conditional expectation of X on Σ_i for each $i\in I$. Show that $\|\sup_{i\in I}|X_i|\|_p\le\frac{p}{p-1}\|X\|_p$.

(g) Let (Ω,Σ,μ) be a probability space, with completion $(\Omega,\hat{\Sigma},\hat{\mu})$, and let $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of μ-integrable real-valued functions such that $\operatorname{dom}X_n\in\Sigma_n$ and X_n is Σ_n-measurable for each $n\in\mathbb{N}$. We say that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a **submartingale adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ if $\int_E X_{n+1}\ge\int_E X_n$ for every $n\in\mathbb{N}$ and every $E\in\Sigma_n$. Prove versions of 275D, 275F, 275G, 275Xf for submartingales.

(h) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale, and $\phi:\mathbb{R}\to\mathbb{R}$ a convex function. Show that $\langle\phi(X_n)\rangle_{n\in\mathbb{N}}$ is a submartingale. (*Hint*: 233J.) Re-examine part (b-ii) of the proof of 275F in the light of this fact.

(i) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of non-negative random variables all with expectation 1. Set $W_n = X_0\times\ldots\times X_n$ for every n. (i) Show that $W=\lim_{n\to\infty}W_n$ is defined a.e. (ii) Show that $\mathbb{E}(W)$ is either 0 or 1. (*Hint*: suppose $\mathbb{E}(W)>0$. Set $Z_n=\lim_{m\to\infty}X_n\times\ldots\times X_m$. Show that $\lim_{n\to\infty}Z_n=1$ when $0<W<\infty$, therefore a.e., by the zero-one law, while $\mathbb{E}(Z_n)\le 1$, by Fatou's lemma, so $\lim_{n\to\infty}\mathbb{E}(Z_n)=1$, while $\mathbb{E}(W)=\mathbb{E}(W_n)\mathbb{E}(Z_{n+1})$ for every n.) (iii) Set $\gamma=\prod_{n=0}^\infty\mathbb{E}(\sqrt{X_n})$. Show that $\gamma>0$ iff $\mathbb{E}(W)=1$. (*Hint*: $\Pr(W_n\ge\frac14\gamma^2)\ge\frac14\gamma^2$ for every n, so if $\gamma>0$ then W cannot be zero a.e.; while $\mathbb{E}(\sqrt{W})\le\gamma$.)

(j) Let $\langle(\Omega_n,\Sigma_n,\mu_n)\rangle_{n\in\mathbb{N}}$ be a sequence of probability spaces with product (Ω,Σ,μ). Suppose that for each $n\in\mathbb{N}$ we have a probability measure ν_n, with domain Σ_n, which is absolutely continuous with respect to μ_n, with Radon-Nikodým derivative f_n, and suppose that $\prod_{n=0}^\infty\int\sqrt{f_n}d\mu_n>0$. Let ν be the product of $\langle\nu_n\rangle_{n\in\mathbb{N}}$. Show that ν is an indefinite-integral measure over μ, with Radon-Nikodým derivative f, where $f(\boldsymbol{\omega})=\prod_{n=0}^\infty f_n(\omega_n)$ for μ-almost every $\boldsymbol{\omega}=\langle\omega_n\rangle_{n\in\mathbb{N}}$ in Ω. (*Hint*: use 275Yi to show that $\int f d\mu=1$.)

(k) Let $\langle p_n\rangle_{n\in\mathbb{N}}$ be a sequence in $[0,1]$. Let μ be the usual measure on $\{0,1\}^{\mathbb{N}}$ (254J) and ν the product of $\langle\nu_n\rangle_{n\in\mathbb{N}}$, where ν_n is the probability measure on $\{0,1\}$ defined by setting $\nu_n\{1\}=p_n$. Show that ν is an indefinite-integral measure over μ iff $\sum_{n=0}^\infty|p_n-\frac12|^2<\infty$.

(l) Find a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ such that the sequence ν_{X_n} of distributions (271C) is convergent for the vague topology (274Ld), but $\langle X_n\rangle_{n\in\mathbb{N}}$ is not convergent in measure.

(m) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables such that $\sum_{n=0}^\infty X_n$ is defined in $\mathbb{R}$ almost everywhere. Suppose that there is an $M\ge 0$ such that $|X_n|\le M$ a.e. for every n. Show that $\sum_{n=0}^\infty\mathbb{E}(X_n)$ is defined in $\mathbb{R}$. (*Hint*: 274Yg, 275G.)

(n) 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; set $E_n=\{\omega:\omega\in\operatorname{dom}X_n,\ |X_n(\omega)|>1\}$, $Y_n=X_n\times\chi(\Omega\setminus E_n)$ for each n, and $Z_n(\omega)=\operatorname{med}(-1,X_n(\omega),1)$ for $n\in\mathbb{N}$ and $\omega\in\operatorname{dom}X_n$. Show that the following are equiveridical: (i) $\sum_{n=0}^\infty X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω; (ii) $\sum_{n=0}^\infty\hat\mu E_n<\infty$, $\sum_{n=0}^\infty\mathbb{E}(Y_n)$ is defined in $\mathbb{R}$, and $\sum_{n=0}^\infty\operatorname{Var}(Y_n)<\infty$, where $\hat\mu$ is the completion of μ; (iii) $\sum_{n=0}^\infty\hat\mu E_n<\infty$, $\sum_{n=0}^\infty\mathbb{E}(Z_n)$ is defined in $\mathbb{R}$, and $\sum_{n=0}^\infty\operatorname{Var}(Z_n)<\infty$. (*Hint*: 273K, 275Ym.) (This is a version of the **Three Series Theorem**.)

(o) Let (Ω,Σ,μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\mathbb{E}(\sup_{n\in\mathbb{N}}|X_n|)$ is finite and $X=\lim_{n\to\infty}X_n$ is defined almost everywhere. For each n, let Y_n be a conditional expectation of X_n on Σ_n. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere to a conditional expectation of X on the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\Sigma_n$.

(p) Show that 275Yo can fail if $\langle X_n\rangle_{n\in\mathbb{N}}$ is merely uniformly integrable, rather than dominated by an integrable function.

(q) Let (Ω,Σ,μ) be a complete probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of random variables on Ω, all with the same distribution, and of finite expectation. For each n, set $S_n=\frac{1}{n+1}(X_0+\ldots+X_n)$; let Σ_n be the σ-algebra defined by S_n and Σ_n^* the σ-algebra generated by $\bigcup_{m\ge n}\Sigma_m$. Show that S_n is a conditional expectation of X_0 on Σ_n^*. (*Hint*: assume every X_i defined everywhere on Ω. Set $\phi(\omega)=\langle X_i(\omega)\rangle_{i\in\mathbb{N}}$. Show that $\phi:\Omega\to\mathbb{R}^{\mathbb{N}}$ is inverse-measure-preserving for a suitable product measure on $\mathbb{R}^{\mathbb{N}}$, and that every set in Σ_n^* is of the form $\phi^{-1}[H]$ where $H\subseteq\mathbb{R}^{\mathbb{N}}$ is a Borel set invariant under permutations of coordinates in the set $\{0,\ldots,n\}$, so that $\int_E X_i=\int_E X_j$ whenever $i\le j\le n$ and $E\in\Sigma_n^*$.) Hence show that $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere. (Compare 273I.)

(r) Formulate and prove versions of the results of this section for martingales consisting of functions taking values in $\mathbb{C}$ or $\mathbb{R}^r$ rather than $\mathbb{R}$.

275 Notes and comments I hope that the sketch above, though distressingly abbreviated, has suggested some of the richness of the concepts involved, and will provide a foundation for further study. All the theorems of this section have far-reaching implications, but the one which is simply indispensable in advanced measure theory is 275I, 'Lévy's martingale convergence theorem', which I will use in the proof of the Lifting Theorem in Chapter 34 of the next volume.

As for stopping times, I mention them partly in an attempt to cast further light on what martingales are for (see 276Ed below), and partly because the ideas of 275N-275O are so important in modern probability theory that, just as a matter of general knowledge, you should be aware that there is something there. I add 275P as one of the most accessible of the standard results which may be obtained by this method.

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) 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, then we have

$$\int_E X_{n+1}-X_n=0$$

whenever $E\in\Sigma_n$. 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) Evidently $\langle X_n\rangle_{n\in\mathbb{N}}$ 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, that is, if $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\tilde\Sigma_n\rangle_{n\in\mathbb{N}}$, where $\tilde\Sigma_n$ is the σ-algebra generated by $\bigcup_{i\le n}\Sigma_{X_i}$.

(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. (Compare 275Ce.)

(f) Of course the most important example of 'martingale difference sequence' is that of 275Bb: any independent sequence of random variables with zero expectation is a martingale difference sequence. It turns out that some of the theorems of §273 concerning such independent sequences may be generalized to martingale difference sequences.

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.

proof (a) Let (Ω,Σ,μ) be the underlying probability space, $(\Omega,\hat\Sigma,\hat\mu)$ its completion, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$ such that $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Set $Y_n=\sum_{i=0}^n X_i$ for each $n\in\mathbb{N}$. Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(b) $\mathbb{E}(Y_n \times X_{n+1}) = 0$ for each n. **P** Y_n is a sum of random variables with finite variance, so $\mathbb{E}(Y_n^2) < \infty$, by 244Ba; it follows that $Y_n \times X_{n+1}$ has finite expectation, by 244Eb. Because the constant function $\mathbf{0}$ is a conditional expectation of X_{n+1} on Σ_n,

$$\mathbb{E}(Y_n \times X_{n+1}) = \mathbb{E}(Y_n \times \mathbf{0}) = 0,$$

by 242L. **Q**

(c) It follows that $\mathbb{E}(Y_n^2) = \sum_{i=0}^n \mathbb{E}(X_i^2)$ for every n. **P** Induce on n. For the inductive step, we have

$$\mathbb{E}(Y_{n+1}^2) = \mathbb{E}(Y_n^2 + 2Y_n \times X_{n+1} + X_{n+1}^2) = \mathbb{E}(Y_n^2) + \mathbb{E}(X_{n+1}^2)$$

because, by (b), $\mathbb{E}(Y_n \times X_{n+1}) = 0$. **Q**

(d) Of course

$$\mathbb{E}(|Y_n|) = \int |Y_n| \times \chi\Omega \le \|Y_n\|_2 \|\chi\Omega\|_2 = \sqrt{\mathbb{E}(Y_n^2)},$$

so

$$\sup_{n\in\mathbb{N}} \mathbb{E}(|Y_n|) \le \sup_{n\in\mathbb{N}} \sqrt{\mathbb{E}(Y_n^2)} = \sqrt{\sum_{i=0}^\infty \mathbb{E}(X_i^2)} < \infty.$$

By 275G, $\lim_{n\to\infty} Y_n$ is defined and finite almost everywhere, that is, $\sum_{i=0}^\infty X_i$ 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 $\sum_{n=0}^\infty \frac{1}{b_n^2} \mathrm{Var}(X_n) < \infty$. Then

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

almost everywhere.

proof (Compare 273D.) As usual, write (Ω, Σ, μ) for the underlying probability space. Set

$$\tilde{X}_n = \frac{1}{b_n} X_n$$

for each n; then $\langle \tilde{X}_n\rangle_{n\in\mathbb{N}}$ also is a martingale difference sequence, and

$$\sum_{n=1}^\infty \mathbb{E}(\tilde{X}_n^2) = \sum_{n=1}^\infty \frac{1}{b_n^2} \mathrm{Var}(X_n) < \infty.$$

By 276B, $\langle \tilde{X}_n(\omega)\rangle_{n\in\mathbb{N}}$ is summable for almost every $\omega \in \Omega$. But by 273Cb,

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=0}^n X_i(\omega) = \lim_{n\to\infty} \frac{1}{b_n} \sum_{i=0}^n b_i \tilde{X}_i(\omega) = 0$$

for all such ω. So we have the result.

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 $\sup_{n\ge 1} \frac{1}{n} b_n < \infty$, then $\lim_{n\to\infty} \frac{1}{n} X_n = 0$ a.e.

proof Consider the martingale difference sequence $\langle Y_n\rangle_{n\in\mathbb{N}} = \langle X_{n+1} - X_n\rangle_{n\in\mathbb{N}}$. Then $\mathbb{E}(Y_n \times X_n) = 0$, so $\mathbb{E}(Y_n^2) + \mathbb{E}(X_n^2) = \mathbb{E}(X_{n+1}^2)$ for each n. In particular, $\langle b_n\rangle_{n\in\mathbb{N}}$ must be non-decreasing.

(a) If $\lim_{n\to\infty} b_n = \infty$, take m such that $b_m > 0$; then

$$\sum_{n=m}^\infty \frac{1}{b_{n+1}^2} \mathrm{Var}(Y_n) = \sum_{n=m}^\infty \frac{1}{b_{n+1}^2} (b_{n+1} - b_n) \le \int_{b_m}^\infty \frac{1}{t^2} dt < \infty.$$

By 276C (modifying b_i for $i < m$, if necessary),

$$\lim_{n\to\infty} \frac{1}{b_n} X_n = \lim_{n\to\infty} \frac{1}{b_{n+1}} (X_0 + \sum_{i=0}^n Y_i) = \lim_{n\to\infty} \frac{1}{b_{n+1}} \sum_{i=0}^n Y_i = 0$$

almost everywhere.

(b) If $\gamma = \sup_{n\ge 1} \frac{1}{n} b_n < \infty$, then $\frac{1}{(n+1)^2} \le \min(1, \gamma^2/t^2)$ for $b_n < t \le b_{n+1}$, so

$$\sum_{n=0}^{\infty} \frac{1}{(n+1)^2}(b_{n+1} - b_n) \le \gamma + \gamma^2 \int_{\gamma}^{\infty} \frac{1}{t^2} dt < \infty,$$

and, by the same argument as before, $\lim_{n\to\infty} \frac{1}{n} X_n = 0$ a.e.

276E 'Impossibility of systems' (a) I return to the word 'martingale' and the idea of a gambling system. Consider a gambler who takes a sequence of 'fair' bets, that is, bets which have payoff expectations of zero, but who chooses which bets to take on the basis of past experience. The appropriate model for such a sequence of random events is a martingale in the sense of 275A, taking Σ_n to be the algebra of all events which are observable up to and including the outcome of the nth bet, and X_n to be the gambler's net gain at that time. (In this model it is natural to take $\Sigma_0 = \{\emptyset, \Omega\}$ and $X_0 = \mathbf{0}$.) Certain paradoxes can arise if we try to imagine this model with atomless Σ_n; to begin with it is perhaps easier to work with the discrete case, in which each Σ_n is finite, or is the set of unions of some countable family of atomic events. Now suppose that the bets involved are just two-way bets, with two equally likely outcomes, but that the gambler chooses his stake each time. In this case we can think of the outcomes as corresponding to an independent sequence $\langle W_n \rangle_{n\in\mathbb{N}}$ of random variables, each taking the values ± 1 with equal probability. The gambler's system must be of the form

$$X_{n+1} = X_n + Z_{n+1} \times W_{n+1},$$

where Z_{n+1} is his stake on the $(n+1)$-st bet, and must be constant on each atom of the σ-algebra Σ_n generated by $W_1, \dots, W_n$. The point is that because $\int_E W_{n+1} = 0$ for each $E \in \Sigma_n$, $\mathbb{E}(Z_{n+1} \times W_{n+1}) = 0$, so $\mathbb{E}(X_{n+1}) = \mathbb{E}(X_n)$.

(b) The general result, of which this is a special case, is the following. 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 $\langle Z_n \rangle_{n\ge 1}$ 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 $n \ge 1$, then $W_0, Z_1 \times W_1, Z_2 \times W_2, \dots$ is a martingale difference sequence adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$; the proof that $\int_E Z_{n+1} \times W_{n+1} = 0$ for every $E \in \Sigma_n$ is exactly the argument of (b) of the proof of 276B.

(c) I invited you to restrict your ideas to the discrete case for a moment; but if you feel that you understand what it means to say that a 'system' or **predictable sequence** $\langle Z_n \rangle_{n\ge 1}$ must be adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$, in the sense that every Z_n is Σ_{n-1}-measurable, then any further difficulty lies in the measure theory needed to show that the integrals $\int_E Z_{n+1} \times W_{n+1}$ are zero, which is what this book is about.

(d) Consider the gambling system mentioned in 275Cf. Here the idea is that $W_n = \pm 1$, as in (a), and $Z_{n+1} = 2^n a$ if $X_n \le 0$, 0 if $X_n > 0$; that is, the gambler doubles his stake each time until he wins, and then quits. Of course he is almost sure to win eventually, so we have $\lim_{n\to\infty} X_n = a$ almost everywhere, even though $\mathbb{E}(X_n) = 0$ for every n. We can compute the distribution of X_n: for $n \ge 1$ we have $\Pr(X_n = a) = 1 - 2^{-n}$, $\Pr(X_n = -(2^n - 1)a) = 2^{-n}$. Thus $\mathbb{E}(|X_n|) = (2 - 2^{-n+1})a$ and the almost-everywhere convergence of the X_n is an example of Doob's Martingale Convergence Theorem.

In the language of stopping times (275N), $X_n = \tilde{Y}_{\tau\wedge n}$, where $Y_n = \sum_{k=0}^{n} 2^k a W_k$ and $\tau = \min\{n : Y_n > 0\}$.

***276F** I come now to Komlós's theorem. The first step is a trifling refinement of 276C.

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 $n \ge 1$. Then $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k = 0$ a.e.

proof Making suitable adjustments on a negligible set if necessary, we may suppose that X'_n is Σ_{n-1}-measurable for $n \ge 1$ and that every X_n and X'_n is defined on the whole of Ω. Set $X'_0 = X_0$ and $Y_n = X_n - X'_n$

for $n \in \mathbb{N}$. Then $\langle Y_n \rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$. Also $\mathbb{E}(Y_n^2) \le \mathbb{E}(X_n^2)$ for every n. **P** If $n \ge 1$, X_n' is square-integrable (244M), and $\mathbb{E}(Y_n \times X_n') = 0$, as in part (b) of the proof of 276B. Now

$$\mathbb{E}(X_n^2) = \mathbb{E}(Y_n + X_n')^2 = \mathbb{E}(Y_n^2) + 2\mathbb{E}(Y_n \times X_n') + \mathbb{E}(X_n')^2 \ge \mathbb{E}(Y_n^2). \quad \mathbf{Q}$$

This means that $\sum_{n=0}^{\infty} \frac{1}{(n+1)^2}\mathbb{E}(Y_n^2)$ must be finite. By 276C, $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} Y_i = 0$ a.e. But by 273Ca we also have $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i' = 0$ whenever $\lim_{n\to\infty} X_n' = 0$, which is almost everywhere. So $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i = 0$ 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 $\lim_{k\to\infty} \mathbb{E}(|Y - Y_k|) = 0$.

proof (a) For each k, $|F_k(X_n)| \le_{\text{a.e.}} k\chi\Omega$ for every n, so that $\{F_k(X_n) : n \in \mathbb{N}\}$ is uniformly integrable, and $\{F_k(X_n)^\bullet : n \in \mathbb{N}\}$ is relatively weakly compact in $L^1 = L^1(\mu)$ (247C). Accordingly $v_k = \lim_{n\to\mathcal{F}} F_k(X_n)^\bullet$ is defined in L^1 for the weak topology (2A3Se); take $Y_k : \Omega \to \mathbb{R}$ to be a measurable function such that $Y_k^\bullet = v_k$. For any $E \in \Sigma$,

$$\int_E Y_k = \int v_k \times (\chi E)^\bullet = \lim_{n\to\mathcal{F}} \int_E F_k(X_n).$$

In particular,

$$|\int_E Y_k| \le \sup_{n\in\mathbb{N}} |\int_E F_k(X_n)| \le k\mu E$$

for every E, so that $\{\omega : Y_k(\omega) > k\}$ and $\{\omega : Y_k(\omega) < -k\}$ are both negligible; changing Y_k on a negligible set if necessary, we may suppose that $|Y_k(\omega)| \le k$ for every $\omega \in \Omega$.

(b) Because Y_k is bounded, $Y_k^\bullet \in L^\infty(\mu)$, and

$$\lim_{n\to\mathcal{F}} \int F_k(X_n) \times Y_k = \lim_{n\to\mathcal{F}} \int F_k(X_n)^\bullet \times Y_k^\bullet = \int Y_k^\bullet \times Y_k^\bullet = \int Y_k^2.$$

Accordingly

$$\begin{aligned}\lim_{n\to\mathcal{F}} \int (F_k(X_n) - Y_k)^2 &= \lim_{n\to\mathcal{F}} \int F_k(X_n)^2 - 2\lim_{n\to\mathcal{F}} \int F_k(X_n) \times Y_k + \int Y_k^2 \\ &= \lim_{n\to\mathcal{F}} \int F_k(X_n)^2 - \int Y_k^2 \le \lim_{n\to\mathcal{F}} \int F_k(X_n)^2.\end{aligned}$$

(c) Set $W_0 = Y_0 = \mathbf{0}$, $W_k = Y_k - Y_{k-1}$ for $k \ge 1$. Then $\mathbb{E}(|W_k|) \le \lim_{n\to\mathcal{F}} \mathbb{E}(|F_k(X_n) - F_{k-1}(X_n)|)$ for every $k \ge 1$. **P** Set $E = \{\omega : W_k(\omega) \ge 0\}$. Then

$$\begin{aligned}\int_E W_k &= \int_E Y_k - \int_E Y_{k-1} \\ &= \lim_{n\to\mathcal{F}} \int_E F_k(X_n) - \lim_{n\to\mathcal{F}} \int_E F_{k-1}(X_n) \\ &= \lim_{n\to\mathcal{F}} \int_E F_k(X_n) - F_{k-1}(X_n) \le \lim_{n\to\mathcal{F}} \int_E |F_k(X_n) - F_{k-1}(X_n)|.\end{aligned}$$

Similarly,

$$|\int_{X\setminus E} W_k| \le \lim_{n\to\mathcal{F}} \int_{X\setminus E} |F_k(X_n) - F_{k-1}(X_n)|.$$

So

$$\mathbb{E}(|W_k|) = \int_E W_k - \int_{X\setminus E} W_k \le \lim_{n\to\mathcal{F}} \int |F_k(X_n) - F_{k-1}(X_n)|. \quad \mathbf{Q}$$

It follows that $\sum_{k=0}^{\infty} \mathbb{E}(|W_k|)$ is finite. **P** For any $m \ge 1$,

$$\begin{aligned}\sum_{k=0}^{m} \mathbb{E}(|W_k|) &\le \sum_{k=1}^{m} \lim_{n\to\mathcal{F}} \mathbb{E}(|F_k(X_n) - F_{k-1}(X_n)|)\\ &= \lim_{n\to\mathcal{F}} \mathbb{E}(\sum_{k=1}^{m} |F_k(X_n) - F_{k-1}(X_n)|)\\ &= \lim_{n\to\mathcal{F}} \mathbb{E}(|F_m(X_n)|) \le \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|).\end{aligned}$$

So $\sum_{k=0}^{\infty} \mathbb{E}(|W_k|) \le \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|)$ is finite. **Q**

By B.Levi's theorem (123A), $\lim_{m\to\infty} \sum_{k=0}^{m} |W_k|$ is finite a.e., so that

$$Y = \lim_{m\to\infty} Y_m = \sum_{k=0}^{\infty} W_k$$

is defined a.e.; and moreover

$$\mathbb{E}(|Y - Y_k|) \le \lim_{m\to\infty} \mathbb{E}(\sum_{j=k+1}^{m} |W_j|) \to 0$$

as $k \to \infty$.

***276H Komlós's theorem** (Komlós 67) 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 $\langle X'_n\rangle_{n\in\mathbb{N}}$.

proof Since neither the hypothesis nor the conclusion is affected by changing the X_n on a negligible set, we may suppose throughout that every X_n is measurable and defined on the whole of Ω. In addition, to begin with (down to the end of (e) below), let us suppose that $\mu X = 1$. As in 276G, set $F_k(x) = x$ for $|x| \le k$, 0 for $|x| > k$.

(a) Let $\mathcal{F}$ be any non-principal ultrafilter on $\mathbb{N}$ (2A1O). For $j \in \mathbb{N}$ set $p_j = \lim_{n\to\mathcal{F}} \Pr(|X_n| > j)$. Then $\sum_{j=0}^{\infty} p_j$ is finite. **P** For any $k \in \mathbb{N}$,

$$\begin{aligned}\sum_{j=0}^{k} p_j &= \sum_{j=0}^{k} \lim_{n\to\mathcal{F}} \Pr(|X_n| > j) = \lim_{n\to\mathcal{F}} \sum_{j=0}^{k} \Pr(|X_n| > j)\\ &\le \lim_{n\to\mathcal{F}} (1 + \int |X_n|) \le 1 + \sup_{n\in\mathbb{N}} \int |X_n|.\end{aligned}$$

So $\sum_{j=0}^{\infty} p_j \le 1 + \sup_{n\in\mathbb{N}} \int |X_n|$ is finite. **Q**

Setting

$$p'_j = p_j - p_{j+1} = \lim_{n\to\mathcal{F}} \Pr(j < |X_n| \le j+1)$$

for each j, we have

$$\begin{aligned}\sum_{j=0}^{\infty} (j+1)p'_j &= \lim_{m\to\infty} \bigl(\sum_{j=0}^{m} (j+1)p_j - \sum_{j=0}^{m} (j+1)p_{j+1}\bigr)\\ &= \lim_{m\to\infty} \sum_{j=0}^{m} p_j - (m+1)p_{m+1} \le \sum_{j=0}^{\infty} p_j < \infty.\end{aligned}$$

Next,

$$\lim_{n\to\mathcal{F}}\int F_k(X_n)^2 \le \sum_{j=0}^k (j+1)^2 p'_j$$

for each k. **P** Setting $E_{jn} = \{\omega : j < |X_n(\omega)| \le j+1\}$ for j, $n \in \mathbb{N}$, $F_k(X_n)^2 \le \sum_{j=0}^k (j+1)^2 \chi E_{jn}$, so

$$\lim_{n\to\mathcal{F}}\int F_k(X_n)^2 \le \lim_{n\to\mathcal{F}}\sum_{j=0}^k (j+1)^2 \mu E_{jn} = \sum_{j=0}^k (j+1)^2 p'_j. \quad \mathbf{Q}$$

(b) Define $\langle Y_k\rangle_{k\in\mathbb{N}}$ and $Y =_{\text{a.e.}} \lim_{k\to\infty} Y_k$ from $\langle X_n\rangle_{n\in\mathbb{N}}$ and $\mathcal{F}$ as in Lemma 276G. Then

$$J_k = \{n : n \in \mathbb{N}, \int (F_k(X_n) - Y_k)^2 \le 1 + \sum_{j=0}^k (j+1)^2 p'_j\}$$

belongs to $\mathcal{F}$ for every $k \in \mathbb{N}$. **P** By (a) above and 276Gb,

$$\lim_{n\to\mathcal{F}}\int (F_k(X_n) - Y_k)^2 \le \lim_{n\to\mathcal{F}}\int F_k(X_n)^2 \le \sum_{j=0}^k (j+1)^2 p'_j. \quad \mathbf{Q}$$

Also, of course,

$$K_k = \{n : n \in \mathbb{N}, \Pr(F_j(X_n) \ne X_n) \le p_j + 2^{-j} \text{ for every } j \le k\}$$

belongs to $\mathcal{F}$ for every k.

(c) For n, $k \in \mathbb{N}$ let Z_{kn} be a simple function such that $|Z_{kn}| \le |F_k(X_n) - Y_k|$ and $\int |F_k(X_n) - Y_k - Z_{kn}| \le 2^{-k}$. For $m \in \mathbb{N}$ let Σ_m be the algebra of subsets of Ω generated by sets of the form $\{\omega : Z_{kn}(\omega) = \alpha\}$ for k, $n \le m$ and $\alpha \in \mathbb{R}$. Because each Z_{kn} takes only finitely many values, Σ_m is finite (and is therefore a σ-subalgebra of Σ); and of course $\Sigma_m \subseteq \Sigma_{m+1}$ for every m.

We need to look at conditional expectations on the Σ_m, and because Σ_m is always finite these have a particularly straightforward expression. Let $\mathcal{A}_m$ be the set of 'atoms', or minimal non-empty sets, in Σ_m; that is, the set of equivalence classes in Ω under the relation $\omega \sim \omega'$ if $Z_{kn}(\omega) = Z_{kn}(\omega')$ for all k, $n \le m$. For any integrable random variable X on Ω, define $\mathbb{E}_m(X)$ by setting

$$\begin{aligned}\mathbb{E}_m(X)(\omega) &= \frac{1}{\mu A}\int_A X \text{ if } x \in A \in \mathcal{A}_m \text{ and } \mu A > 0,\\ &= 0 \text{ if } x \in A \in \mathcal{A}_m \text{ and } \mu A = 0.\end{aligned}$$

Then $\mathbb{E}_m(X)$ is a conditional expectation of X on Σ_m.

Now

$$\begin{aligned}\lim_{n\to\mathcal{F}}\int |\mathbb{E}_m(F_k(X_n) - Y_k)| &= \lim_{n\to\mathcal{F}}\sum_{A\in\mathcal{A}_m}\int_A |\mathbb{E}_m(F_k(X_n) - Y_k)|\\ &= \lim_{n\to\mathcal{F}}\sum_{A\in\mathcal{A}_m}|\int_A \mathbb{E}_m(F_k(X_n) - Y_k)|\end{aligned}$$

(because $\mathbb{E}_m(F_k(X_n) - Y_k)$ is constant on each $A \in \mathcal{A}_m$)

$$\begin{aligned}&= \lim_{n\to\mathcal{F}}\sum_{A\in\mathcal{A}_m}|\int_A F_k(X_n) - Y_k|\\ &= \sum_{A\in\mathcal{A}_m}\lim_{n\to\mathcal{F}}|\int_A F_k(X_n) - Y_k| = 0\end{aligned}$$

by the choice of Y_k. So if we set

$$I_m = \{n : n \in \mathbb{N}, \int |\mathbb{E}_m(F_k(X_n) - Y_k)| \le 2^{-k} \text{ for every } k \le m\},$$

then $I_m \in \mathcal{F}$ for every m.

(d) Suppose that $\langle r(n)\rangle_{n\in\mathbb{N}}$ is any strictly increasing sequence in $\mathbb{N}$ such that $r(0) > 0$, $r(n) \in J_n \cap K_n$ for every n and $r(n) \in I_{r(n-1)}$ for $n \ge 1$. Then $\frac{1}{n+1}\sum_{i=0}^n X_{r(i)} \to Y$ a.e. as $n \to \infty$. **P** Express $X_{r(n)}$ as

$$(X_{r(n)} - F_n(X_{r(n)})) + (F_n(X_{r(n)}) - Y_n - Z_{n,r(n)}) + Y_n + Z_{n,r(n)}$$

for each n. Taking these pieces in turn:

(i)

$$\sum_{n=0}^{\infty}\Pr(X_{r(n)}\ne F_n(X_{r(n)}))\le\sum_{n=0}^{\infty}p_n+2^{-n}$$

(because $r(n)\in K_n$ for every n)

$$<\infty$$

by (a). But this means that $X_{r(n)}-F_n(X_{r(n)})\to 0$ a.e., since the sequence is eventually zero at almost every point, and $\frac{1}{n+1}\sum_{i=0}^{n}X_{r(i)}-F_i(X_{r(i)})\to 0$ a.e. by 273Ca again.

(ii) By the choice of the $Z_{n,r(n)}$,

$$\sum_{n=0}^{\infty}\int|F_n(X_{r(n)})-Y_n-Z_{n,r(n)}|\le\sum_{n=0}^{\infty}2^{-n}$$

is finite, so $F_n(X_{r(n)})-Y_n-Z_{n,r(n)}\to 0$ a.e. and $\frac{1}{n+1}\sum_{i=0}^{n}F_i(X_{r(i)})-Y_i-Z_{i,r(i)}\to 0$ a.e.

(iii) By 276G, $Y_n\to Y$ a.e. and $\frac{1}{n+1}\sum_{i=0}^{n}Y_i\to Y$ a.e.

(iv) We know that, for each $n\ge 1$, $r(n)\in I_{r(n-1)}$. So (because $r(n-1)\ge n$) $\int|\mathbb{E}_{r(n-1)}(F_n(X_{r(n)})-Y_n)|\le 2^{-n}$. But as also

$$\int\Bigl|\mathbb{E}_{r(n-1)}\bigl(F_n(X_{r(n)})-Y_n-Z_{n,r(n)}\bigr)\Bigr|\le\int|F_n(X_{r(n)})-Y_n-Z_{n,r(n)}|\le 2^{-n}$$

by 244M and the choice of $Z_{n,r(n)}$,

$$\int|\mathbb{E}_{r(n-1)}Z_{n,r(n)}|=\int\bigl|\mathbb{E}_{r(n-1)}(F_n(X_{r(n)})-Y_n)-\mathbb{E}_{r(n-1)}(F_n(X_{r(n)})-Y_n-Z_{n,r(n)})\bigr|$$

$$\le 2^{-n+1}$$

for every n. Accordingly $\mathbb{E}_{r(n-1)}Z_{n,r(n)}\to 0$ a.e.

On the other hand,

$$\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\int Z_{n,r(n)}^2\le\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\int F_n(X_{r(n)}-Y_n)^2$$

$$\le\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}(1+\sum_{j=0}^{n}(j+1)^2p'_j)$$

(because $r(n)\in J_n$)

$$\le\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}+\sum_{j=0}^{\infty}(j+1)^2p'_j\sum_{n=j}^{\infty}\frac{1}{(n+1)^2}$$

$$\le\frac{\pi^2}{6}+2\sum_{j=0}^{\infty}(j+1)p'_j$$

is finite. (I am using the estimate

$$\sum_{n=j}^{\infty}\frac{1}{(n+1)^2}\le\sum_{n=j}^{\infty}\frac{2}{n+1}-\frac{2}{n+2}=\frac{2}{j+1}.)$$

By 276F, applied to $\langle\Sigma_{r(n)}\rangle_{n\in\mathbb{N}}$ and $\langle Z_{n,r(n)}\rangle_{n\in\mathbb{N}}$, $\frac{1}{n+1}\sum_{i=0}^{n}Z_{i,r(i)}\to 0$ a.e.

(v) Adding these four components, we see that $\frac{1}{n+1}\sum_{i=0}^{n}X_{r(i)}\to 0$, as claimed. **Q**

(e) Now fix any strictly increasing sequence $\langle s(n)\rangle_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $s(0)>0$, $s(n)\in\bigcap_{m\le n}J_m\cap K_n$ for every n and $s(n)\in I_{s(n-1)}$ for $n\ge 1$; such a sequence exists because $J_m\cap K_m\cap I_{s(n-1)}$ belongs to $\mathcal{F}$, so is infinite, for every $n\ge 1$ and $m\in\mathbb{N}$. Set $X'_n=X_{s(n)}$ for every n. If $\langle X''_n\rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X'_n\rangle_{n\in\mathbb{N}}$, then it is of the form $\langle X_{s(r(n))}\rangle_{n\in\mathbb{N}}$ for some strictly increasing sequence $\langle r(n)\rangle_{n\in\mathbb{N}}$. In this case

$$s(r(0))\ge s(0)>0,$$

$$s(r(n))\in J_{r(n)}\cap K_{r(n)}\subseteq J_n\cap K_n \text{ for every } n,$$

$$s(r(n))\in I_{s(r(n)-1)}\subseteq I_{s(r(n-1))} \text{ for every } n\ge 1.$$

So (d) tells us that $\frac{1}{n+1}\sum_{i=0}^{n}X''_i\to Y$ a.e.

(f) Thus the theorem is proved in the case in which (Ω,Σ,μ) is a probability space. Now suppose that μ is σ-finite and $\mu\Omega>0$. In this case there is a strictly positive measurable function $f:\Omega\to\mathbb{R}$ such that $\int f d\mu=1$ (215B(ix)). Let ν be the corresponding indefinite-integral measure (234J), so that ν is a probability measure on Ω, and $\langle\frac{1}{f}\times X_n\rangle_{n\in\mathbb{N}}$ is a sequence of ν-integrable functions such that $\sup_{n\in\mathbb{N}}\int\frac{1}{f}\times X_n d\nu$ is finite (235K). From (a)-(e) we see that there must be a ν-integrable function Y and a subsequence $\langle X'_n\rangle_{n\in\mathbb{N}}$ of $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\frac{1}{n+1}\sum_{i=0}^{n}\frac{1}{f}\times X''_i\to Y$ ν-a.e. for every subsequence $\langle X''_n\rangle_{n\in\mathbb{N}}$ of $\langle X'_n\rangle_{n\in\mathbb{N}}$. But μ and ν have the same negligible sets (234Lc), so $\frac{1}{n+1}\sum_{i=0}^{n}X''_i\to f\times Y$ μ-a.e. for every subsequence $\langle X''_n\rangle_{n\in\mathbb{N}}$ of $\langle X'_n\rangle_{n\in\mathbb{N}}$.

(g) Since the result is trivial if $\mu\Omega=0$, the theorem is true whenever μ is σ-finite. For the general case, set

$$\tilde{\Omega}=\bigcup_{n\in\mathbb{N}}\{\omega:X_n(\omega)\ne 0\}=\bigcup_{m,n\in\mathbb{N}}\{\omega:|X_n(\omega)|\ge 2^{-m}\},$$

so that the subspace measure $\mu_{\tilde{\Omega}}$ is σ-finite. Then there are a $\mu_{\tilde{\Omega}}$-integrable function $\tilde{Y}$ and a subsequence $\langle X'_n\rangle_{n\in\mathbb{N}}$ of $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\frac{1}{n+1}\sum_{i=0}^{n}X''_i\restriction\tilde{\Omega}\to\tilde{Y}$ $\mu_{\tilde{\Omega}}$-a.e. for every subsequence $\langle X''_n\rangle_{n\in\mathbb{N}}$ of $\langle X'_n\rangle_{n\in\mathbb{N}}$. Setting $Y(\omega)=\tilde{Y}(\omega)$ if $\omega\in\tilde{\Omega}$, 0 for $\omega\in\Omega\setminus\tilde{\Omega}$, we see that Y is μ-integrable and that $\frac{1}{n+1}\sum_{i=0}^{n}X''_i\to Y$ μ-a.e. whenever $\langle X''_n\rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X'_n\rangle_{n\in\mathbb{N}}$. This completes the proof.

276X Basic exercises >(a) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale adapted to a sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras. Show that $\int_E X_n^2\le\int_E X_{n+1}^2$ whenever $n\in\mathbb{N}$ and $E\in\Sigma_n$ (allowing ∞ as a value of an integral). (*Hint*: see the proof of 276B.)

>(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale. Show that for any $\epsilon>0$,

$$\Pr(\sup_{n\in\mathbb{N}}|X_n|\ge\epsilon)\le\frac{1}{\epsilon^2}\sup_{n\in\mathbb{N}}\mathbb{E}(X_n^2).$$

(*Hint*: put 276Xa together with the argument for 275D.)

(c) When does 276Xb give a sharper result than 275Xb?

(d) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent identically distributed sequence of random variables with zero expectation and non-zero finite variance, and $\langle t_n\rangle_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}$. Show that (i) if $\sum_{n=0}^{\infty}t_n^2<\infty$, then $\sum_{n=0}^{\infty}t_nX_n$ is defined in $\mathbb{R}$ a.e. (ii) if $\sum_{n=0}^{\infty}t_n^2=\infty$ then $\sum_{n=0}^{\infty}t_nX_n$ is undefined a.e. (*Hint*: 276B, 274Xj.)

(e) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence and set $Y_n=\frac{1}{n+1}(X_0+\ldots+X_n)$ for each $n\in\mathbb{N}$. Show that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is uniformly integrable then $\lim_{n\to\infty}\|Y_n\|_1=0$. (*Hint*: use the argument of 273Na, with 276C in place of 273D, and setting $\tilde{X}_n=X'_n-Z_n$, where Z_n is an appropriate conditional expectation of X'_n.)

(f) Suppose that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a uniformly bounded martingale difference sequence and $\langle a_n\rangle_{n\in\mathbb{N}}\in\ell^2$. Show that $\lim_{n\to\infty}\prod_{i=0}^{n}(1+a_iX_i)$ is defined and finite almost everywhere. (*Hint*: $\langle a_nX_n\rangle_{n\in\mathbb{N}}$ is summable and square-summable a.e.)

>(g) Strong law of large numbers: fifth form A sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables is **exchangeable** if $(X_{n_0},\dots,X_{n_k})$ has the same joint distribution as $(X_0,\dots,X_k)$ whenever $n_0,\dots,n_k$ are distinct. Show that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is an exchangeable sequence of random variables with finite expectation, then $\langle\frac{1}{n+1}\sum_{i=0}^{\infty}X_i\rangle_{n\in\mathbb{N}}$ converges a.e. (*Hint*: 276H.)

(h) In 276B, show that $\mathbb{E}((\sum_{n=0}^{\infty}X_n)^2)\le\sum_{n=0}^{\infty}\mathbb{E}(X_n^2)$.

276Y Further exercises (a) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that $\sup_{n\in\mathbb{N}}\|X_n\|_p$ is finite, where $p\in\,]1,\infty[$. Show that $\lim_{n\to\infty}\|\frac{1}{n+1}\sum_{i=0}^{n}X_i\|_p=0$. (*Hint*: 273Nb.)

(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable martingale difference sequence and Y a bounded random variable. Show that $\lim_{n\to\infty}\mathbb{E}(X_n\times Y)=0$. (Compare 272Ye.)

(c) Use 275Yh to prove 276Xa.

(d) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of random variables such that, for some $\delta>0$, $\sup_{n\in\mathbb{N}}n^{\delta}\mathbb{E}(|X_n|)$ is finite. Set $S_n=\frac{1}{n+1}(X_0+\dots+X_n)$ for each n. Show that $\lim_{n\to\infty}S_n=0$ a.e. (*Hint*: set $Z_k=2^{-k}(|X_0|+\dots+|X_{2^k-1}|)$. Show that $\sum_{k=0}^{\infty}\mathbb{E}(Z_k)<\infty$. Show that $|S_n|\le 2Z_{k+1}$ if $2^k<n\le 2^{k+1}$.)

(e) Strong law of large numbers: sixth form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that, for some $\delta>0$, $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|^{1+\delta})$ is finite. Set $S_n=\frac{1}{n+1}(X_0+\dots+X_n)$ for each n. Show that $\lim_{n\to\infty}S_n=0$ a.e. (*Hint*: take a non-decreasing sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted. Set $Y_n=X_n$ when $|X_n|\le n$, 0 otherwise. Let U_n be a conditional expectation of Y_n on Σ_{n-1} and set $Z_n=Y_n-U_n$. Use ideas from 273H, 276C and 276Yd above to show that $\frac{1}{n+1}\sum_{i=0}^{n}V_i\to 0$ a.e. for $V_i=Z_i$, $V_i=U_i$, $V_i=X_i-Y_i$.)

(f) Show that there is a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ which converges in measure but is not convergent a.e. (Compare 273Ba.) (*Hint*: arrange that $\{\omega:X_{n+1}(\omega)\ne 0\}=E_n\subseteq\{\omega:|X_{n+1}(\omega)-X_n(\omega)|\ge 1\}$, where $\langle E_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of sets and $\mu E_n=\dfrac{1}{n+1}$ for each n.)

(g) Give an example of an identically distributed martingale difference sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\langle\frac{1}{n+1}(X_0+\dots+X_n)\rangle_{n\in\mathbb{N}}$ does not converge to 0 almost everywhere. (*Hint*: start by devising a uniformly bounded sequence $\langle U_n\rangle_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\mathbb{E}(|U_n|)=0$ but $\langle\frac{1}{n+1}(U_0+\dots+U_n)\rangle_{n\in\mathbb{N}}$ does not converge to 0 almost everywhere. Now repeat your construction in such a context that the U_n can be derived from an identically distributed martingale difference sequence by the formulae of 276Ye.)

(h) Construct a proof of Komlós's theorem which does not involve ultrafilters, or any other use of the full axiom of choice, but proceeds throughout by selecting appropriate sub-subsequences. Remember to check that you can prove any fact you use about weakly convergent sequences in L^1 on the same rules.

276 Notes and comments I include two more versions of the strong law of large numbers (276C, 276Ye) not because I have any applications in mind but because I think that if you know the strong law for $\|\;\|_{1+\delta}$-bounded independent sequences, and what a martingale difference sequence is, then there is something missing if you do not know the strong law for $\|\;\|_{1+\delta}$-bounded martingale difference sequences. And then, of course, I have to add 276Yf and 276Yg (which seems to be difficult), lest you be tempted to think that the strong law is 'really' about martingale difference sequences rather than about independent sequences. (Compare 272Yd and 275Xl.)

Komlós's theorem is rather outside the scope of this volume; it is quite hard work and surely much less important, to most probabilists, than many results I have omitted. It does provide a quick proof of 276Xg. However it is relevant to questions arising in some topics treated in Volumes 3 and 4, and the proof fits naturally into this section.

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.

Version of 17.12.12

References for Volume 2

Alexits G. [78] (ed.) *Fourier Analysis and Approximation Theory.* North-Holland, 1978 (Colloq. Math. Soc. Janos Bolyai 19).

Antonov N.Yu. [96] 'Convergence of Fourier series', East J. Approx. 7 (1996) 187-196. [§286 *notes*.]

Arias de Reyna J. [02] *Pointwise Convergence of Fourier Series.* Springer, 2002 (Lecture Notes in Mathematics 1785). [§286 *notes*.]

Baker R. [04] '"Lebesgue measure" on $\mathbb{R}^\infty$, II', Proc. Amer. Math. Soc. 132 (2004) 2577-2591. [254Yb.]

Bergelson V., March P. & Rosenblatt J. [96] (eds.) *Convergence in Ergodic Theory and Probability.* de Gruyter, 1996.

Bogachev V.I. [07] *Measure theory.* Springer, 2007.

du Bois-Reymond P. [1876] 'Untersuchungen über die Convergenz und Divergenz der Fouriersche Darstellungformeln', Abh. Akad. München 12 (1876) 1-103. [§282 *notes*.]

Bourbaki N. [66] *General Topology.* Hermann/Addison-Wesley, 1968. [2A5F.]

Bourbaki N. [87] *Topological Vector Spaces.* Springer, 1987. [2A5E.]

Carleson L. [66] 'On convergence and growth of partial sums of Fourier series', Acta Math. 116 (1966) 135-157. [§282 *notes*, §286 *intro.*, §286 *notes*.]

Clarkson J.A. [1936] 'Uniformly convex spaces', Trans. Amer. Math. Soc. 40 (1936) 396-414. [244O.]

Defant A. & Floret K. [93] *Tensor Norms and Operator Ideals*, North-Holland, 1993. [§253 *notes*.]

Doob J.L. [53] *Stochastic Processes.* Wiley, 1953.

Dudley R.M. [89] *Real Analysis and Probability.* Wadsworth & Brooks/Cole, 1989. [§282 *notes*.]

Dunford N. & Schwartz J.T. [57] *Linear Operators I.* Wiley, 1957 (reprinted 1988). [§244 *notes*, 2A5J.]

Enderton H.B. [77] *Elements of Set Theory.* Academic, 1977. [§2A1.]

Engelking R. [89] *General Topology.* Heldermann, 1989 (Sigma Series in Pure Mathematics 6). [2A5F.]

Etemadi N. [96] 'On convergence of partial sums of independent random variables', pp. 137-144 in Bergelson March & Rosenblatt 96. [272V.]

Evans L.C. & Gariepy R.F. [92] *Measure Theory and Fine Properties of Functions.* CRC Press, 1992. [263Ec, §265 *notes*.]

Federer H. [69] *Geometric Measure Theory.* Springer, 1969 (reprinted 1996). [262C, 263Ec, §264 *notes*, §265 *notes*, §266 *notes*.]

Feller W. [66] *An Introduction to Probability Theory and its Applications,* vol. II. Wiley, 1966. [Chap. 27 *intro.*, 274H, 275Xc, 285N.]

Fremlin D.H. [74] *Topological Riesz Spaces and Measure Theory.* Cambridge U.P., 1974. [§232 *notes*, 241F, §244 *notes*, §245 *notes*, §247 *notes*.]

Fremlin D.H. [93] 'Real-valued-measurable cardinals', pp. 151-304 in Judah 93. [232H.]

Haimo D.T. [67] (ed.) *Orthogonal Expansions and their Continuous Analogues.* Southern Illinois University Press, 1967.

Hall P. [82] *Rates of Convergence in the Central Limit Theorem.* Pitman, 1982. [274H.]

Halmos P.R. [50] *Measure Theory.* Van Nostrand, 1950. [§251 *notes*, §252 *notes*, 255Yn.]

Halmos P.R. [60] *Naive Set Theory.* Van Nostrand, 1960. [§2A1.]

Hanner O. [56] 'On the uniform convexity of L^p and l^p', Arkiv för Matematik 3 (1956) 239-244. [244O.]

Henle J.M. [86] *An Outline of Set Theory.* Springer, 1986. [§2A1.]

Hoeffding W. [63] 'Probability inequalities for sums of bounded random variables', J. Amer. Statistical Association 58 (1963) 13-30. [272W.]

Hunt R.A. [67] 'On the convergence of Fourier series', pp. 235-255 in Haimo 67. [§286 *notes*.]

Jorsbøe O.G. & Mejlbro L. [82] *The Carleson-Hunt Theorem on Fourier Series.* Springer, 1982 (Lecture Notes in Mathematics 911). [§286 *notes*.]

Judah H. [93] (ed.) Proceedings of the Bar-Ilan Conference on Set Theory and the Reals, 1991. Amer. Math. Soc. (Israel Mathematical Conference Proceedings 6), 1993.

Kelley J.L. [55] *General Topology.* Van Nostrand, 1955. [2A5F.]

Kelley J.L. & Namioka I. [76] *Linear Topological Spaces.* Springer, 1976. [2A5C.]

Kirszbraun M.D. [1934] 'Über die zusammenziehenden und Lipschitzian Transformationen', Fund. Math. 22 (1934) 77-108. [262C.]

Kolmogorov A.N. [1926] 'Une série de Fourier-Lebesgue divergente partout', C. R. Acad. Sci. Paris 183 (1926) 1327-1328. [§282 *notes.*]

Komlós J. [67] 'A generalization of a problem of Steinhaus', Acta Math. Acad. Sci. Hung. 18 (1967) 217-229. [276H.]

Körner T.W. [88] *Fourier Analysis.* Cambridge U.P., 1988. [§282 *notes.*]

Köthe G. [69] *Topological Vector Spaces I.* Springer, 1969. [2A5C, 2A5E, 2A5J.]

Krivine J.-L. [71] *Introduction to Axiomatic Set Theory.* D. Reidel, 1971. [§2A1.]

Lacey M.T. [05] *Carleson's Theorem: Proof, Complements, Variations.* `http://arxiv.org/pdf/math/0307008v4.pdf`. [§286 *notes.*]

Lacey M.T. & Thiele C.M. [00] 'A proof of boundedness of the Carleson operator', Math. Research Letters 7 (2000) 1-10. [§286 *intro.*, 286H.]

Lebesgue H. [72] *Oeuvres Scientifiques.* L'Enseignement Mathématique, Institut de Mathématiques, Univ. de Genève, 1972. [Chap. 27 *intro.*]

Liapounoff A. [1901] 'Nouvelle forme du théorème sur la limite de probabilité', Mém. Acad. Imp. Sci. St-Pétersbourg 12(5) (1901) 1-24. [274Xh.]

Lighthill M.J. [59] *Introduction to Fourier Analysis and Generalised Functions.* Cambridge U.P., 1959. [§284 *notes.*]

Lindeberg J.W. [1922] 'Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung', Math. Zeitschrift 15 (1922) 211-225. [274H, §274 *notes.*]

Lipschutz S. [64] *Set Theory and Related Topics.* McGraw-Hill, 1964 (Schaum's Outline Series). [§2A1.]

Loève M. [77] *Probability Theory I.* Springer, 1977. [Chap. 27 *intro.*, 274H.]

Luxemburg W.A.J. & Zaanen A.C. [71] *Riesz Spaces I.* North-Holland, 1971. [241F.]

Mozzochi C.J. [71] *On the Pointwise Convergence of Fourier Series.* Springer, 1971 (Lecture Notes in Mathematics 199). [§286 *notes.*]

Naor A. [04] 'Proof of the uniform convexity lemma', `http://www.cims.nyu.edu/~naor/homepage files/inequality.pdf`, 26.2.04. [244O.]

Rényi A. [70] *Probability Theory.* North-Holland, 1970. [274H.]

Roitman J. [90] *An Introduction to Set Theory.* Wiley, 1990. [§2A1.]

Roselli P. & Willem M. [02] 'A convexity inequality', Amer. Math. Monthly 109 (2002) 64-70. [244Ym.]

Saks S. [1924] 'Sur les nombres dérivés des fonctions', Fund

Schipp F. [78] 'On Carleson's method', pp. 679-695 in Alexits 78. [§286 *notes.*]

Semadeni Z. [71] *Banach spaces of continuous functions I.* Polish Scientific Publishers, 1971. [§253 *notes.*]

Shiryayev A.N. [84] *Probability.* Springer, 1984. [285N.]

Steele J.M. [86] 'An Efron-Stein inequality of nonsymmetric statistics', Annals of Statistics 14 (1986) 753-758. [274Ya.]

Zaanen A.C. [83] *Riesz Spaces II.* North-Holland, 1983. [241F.]

Zygmund A. [59] *Trigonometric Series.* Cambridge U.P, 1959. [§244 *notes*, §282 *notes*, 284Xk.]