Contents

Discrete probability is an adequate framework for Sections 1, 2, 3. Continuous probability space is needed for Sect. 4; the interval $(0, 1)$ with Lebesgue measure can be used. Sect. 5 needs some more measure theory.

1 Combinatorial setup: a simple random walk, n steps

A path (of length n) of the one-dimensional simple random walk is a sequence $(s_0, \dots, s_n)$ of integers such that the increments $x_k = s_k - s_{k-1}$ are ± 1 (for $k = 1, \dots, n$), and $s_0 = 0$. Every path is endowed with the probability $1/2^n$.

In other words, the one-dimensional simple random walk (of length n) is the sequence $(S_0, \dots, S_n)$ of sums $S_k = X_1 + \dots + X_k$ of random signs $X_1, \dots, X_n$, that is, independent random variables, each taking on two values $-1, +1$ with probabilities $1/2$, $1/2$.

The three-dimensional simple random walk is defined similarly, but each X_k takes on six values $(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$ with equal probabilities $(1/6)$. The same for any dimension.

1a Extremal values, etc.

Let $(S_0, \dots, S_n)$ be the one-dimensional simple random walk, and $M_n = \max(S_0, \dots, S_n)$. Clearly, $\frac12 (S_n + n)$ has the binomial distribution, $\mathbb{P}\big(S_n = -n + 2m \big) = 2^{-n} \binom{n}{m}$ for $m = 0, 1, \dots, n$. And what about the distribution of M_n?

1a1 Proposition. For every $m \ge 0$,

$$\mathbb{P}\big(M_n = m \big) = \mathbb{P}\big(S_n = m \big) + \mathbb{P}\big(S_n = m+1 \big) = \\ = 2^{-n} \cdot \begin{cases} \binom{n}{\frac n2 \pm \frac m2} & \text{for } m+n \text{ even,} \\ \binom{n}{\frac n2 \pm \frac{m+1}2} & \text{for } m+n \text{ odd.} \end{cases}$$

1a2 Proposition. For every s, m such that $m \ge 0$ and $m \ge s$,

$$\mathbb{P}\big(S_n = s, M_n = m \big) = \mathbb{P}\big(S_n = 2m - s \big) - \mathbb{P}\big(S_n = 2m - s + 2 \big) \, .$$

1a3 Proposition. For every $a, b \ge 0$ such that $a + b > 0$,

$$\mathbb{P}\big(S_1 > 0, \dots, S_{a+b} > 0 \,\big|\, S_{a+b} = a - b \big) = \frac{a-b}{a+b} \, .$$

The latter is well-known as 'the ballot theorem': "Suppose that in an election candidate A gets a votes and candidate B gets b votes where $b < a$. Then the (conditional) probability that throughout the counting A always beats B is $(a-b)/(a+b)$."

The proof uses Reflection Principle.

See Durrett, Sect. 3.3, especially, Item (3.2).

Note that the probability does not decay if $a, b \to \infty$, $a/b \to \text{const}$. In fact, the expected number of zeros remains bounded...

Here is another use of reflection. Let us say that k is a point of increase if

$$\begin{aligned} S_l &< S_k \quad \text{for } l = 0, \dots, k-1 \,, \\ S_l &\ge S_k \quad \text{for } l = k+1, \dots, n \,. \end{aligned}$$

1a4 Proposition. The expected number of points of increase is equal to 1.

However, it is well-known that for large n the walk typically has no points of increase. A paradox! What do you think? A clue: I tried 1000 paths of length $n = 100$ and got the following empirical distribution for the number of points of increase:

value	0	1	2	3	4	5	6	7	8	9	10	11	12	14	19	21
occurs	722	63	45	41	34	24	20	9	14	8	7	1	4	4	2	2

1b Return probability

Denote by $p_n^{(d)}$ the probability of the event $S_n = 0$ for the d-dimensional simple random walk $(S_0, \dots, S_n)$. Clearly, $p_n^{(d)} = 0$ for odd n.

1b1 Proposition.

$$\begin{aligned} p_{2n}^{(1)} &= 2^{-2n} \binom{2n}{n} \,; \\ p_{2n}^{(2)} &= (p_{2n}^{(1)})^2 = 4^{-2n} \binom{2n}{n}^2 \,; \\ p_{2n}^{(3)} &= 6^{-2n} \binom{2n}{n} \sum_{k+l+m=n} \binom{n}{k,l,m}^2 \,. \end{aligned}$$

Note that $p_{2n}^{(3)} \ne (p_{2n}^{(1)})^3$.

See Durrett, Sect. 3.2.

2 Asymptotic combinatorics: a simple random walk, $n \to \infty$ steps

2a Weak law of large numbers

Note that the probability $\mathbb{P}\big(|S_n| \le a\big)$ depends on the dimension of the simple random walk $(S_k)_k$, on a and n, but not on the total number of steps (be it n or more).

2a1 Proposition. For every $\varepsilon > 0$,

$$\mathbb{P}\big(\tfrac{1}{n}|S_n| \le \varepsilon\big) \to 1 \quad \text{as } n \to \infty\,.$$

This is a special case of Weak Law of Large Numbers. See also Sect. 3a.

2b Asymptotic normality

Note that the probability $\mathbb{P}\big(S_n = k\big)$ for the one-dimensional simple random walk $(S_k)_k$ depends on k and n, but not on the total number of steps (be it n or more). Clearly, the probability vanishes if $n + k$ is odd.

Here is the corresponding Local Limit Theorem.

2b1 Proposition.

$$\mathbb{P}\big(S_n = k\big) = \frac{2}{\sqrt{2\pi n}} \exp\Big(-\frac{k^2}{2n}\Big) \cdot \bigg(1 + \alpha_n\Big(\frac{k}{\sqrt{n}}\Big)\bigg) \quad \text{for } k+n \text{ even}\,,$$

where $\alpha_n(\cdot) \to 0$ uniformly on bounded intervals.

In contrast, the next result is global.

2b2 Theorem. (*De Moivre-Laplace*)

$$\mathbb{P}\big(a\sqrt{n} < S_n < b\sqrt{n}\big) \to \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\,\mathrm{d}x \quad \text{as } n \to \infty$$

whenever $-\infty \le a \le b \le \infty$.

See Durrett, Sect. 2.1. See also Sect. 3b (below).

2c Large deviations

2c1 Proposition. For the one-dimensional simple random walk, for every $c \in (-1, 1)$,

$$\frac{1}{n} \ln \mathbb{P}\big(S_n > cn\big) \to -\gamma(c) \quad \text{as } n \to \infty\,,$$

where $\gamma(c) = \frac{1}{2}(1+c)\ln(1+c) + \frac{1}{2}(1-c)\ln(1-c)$.

See Durrett, Sect. 1.9, p. 76; also Sect. 2.1, Exercise 1.3.

Proposition 2c1 suggests the approximation (for large c and n)

$$\mathbb{P}\big(S_n > c\big) \approx \mathrm{e}^{-n\gamma(c/n)} = \frac{n^n}{\sqrt{(n-c)^{n-c}(n+c)^{n+c}}}\,.$$

However, Theorem 2b2 suggests another approximation,

$$\mathbb{P}\big(S_n > c\big) \approx \frac{1}{\sqrt{2\pi}} \int_{c/\sqrt{n}}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x \approx \exp\Big(-\frac{c^2}{2n}\Big)\,.$$

A paradox! What do you think? A clue: for $n = 200$,

c	0	30	60	90	120	150	180
$2^{-n}\binom{n}{(n+c)/2}$	$6 \cdot 10^{-2}$	$6 \cdot 10^{-3}$	$6 \cdot 10^{-6}$	$5 \cdot 10^{-11}$	$1 \cdot 10^{-18}$	$3 \cdot 10^{-29}$	$1 \cdot 10^{-44}$
$\frac{2}{\sqrt{2\pi n}} \exp(-\frac{c^2}{2n})$	$6 \cdot 10^{-2}$	$6 \cdot 10^{-3}$	$7 \cdot 10^{-6}$	$9 \cdot 10^{-11}$	$1 \cdot 10^{-17}$	$2 \cdot 10^{-26}$	$4 \cdot 10^{-37}$
$\exp(-n\gamma(\frac{c}{n}))$	1	$1 \cdot 10^{-1}$	$1 \cdot 10^{-4}$	$8 \cdot 10^{-10}$	$2 \cdot 10^{-17}$	$3 \cdot 10^{-28}$	$1 \cdot 10^{-43}$

2d Mean number of returns

Recall the probability $p_n^{(d)}$ of Sect. 1b. Note that $p_1^{(d)} + \dots + p_n^{(d)}$ is the expected number of returns. What happens as $n \to \infty$?

2d1 Proposition. The series $\sum_{k=1}^{\infty} p_k^{(d)}$ diverges for $d = 1$ and $d = 2$, but converges for $d = 3$.

See Durrett, Sect. 3.2. See also Sect. 4i (below).

3 Long sequences of random variables: more general random walks, etc.

3a Weak Law of Large Numbers (again)

Note that the distribution

$$\mu^{*n} = \underbrace{\mu * \dots * \mu}_{n} \qquad \text{(convolution of measures\dots)}$$

of the sum $S_n = X_1 + \dots + X_n$ of n independent random variables, each distributed μ, depends on n and μ only. Below, S_n means just a random variable distributed μ^{*n}, where μ is the distribution of X_1. This way we may avoid *infinite* sequences of independent random variables.

3a1 Theorem. If $\mathbb{E}\,|X_1| < \infty$ then

$$\mathbb{P}\big(|\tfrac{1}{n} S_n - \mathbb{E}\, X_1| \le \varepsilon \big) \to 1 \quad \text{as } n \to \infty\,.$$

See Durrett, Sect. 1.5, Corollary (5.8).

Clearly, 3a1 implies 2a1.

Rather unexpectedly, 3a1 can be used for proving Weierstrass's approximation theorem: polynomials are dense in $C[0,1]$. See Durrett, Sect. 1.5, Example 5.1.

3b Asymptotic normality (again)

Here is the Central Limit Theorem for independent identically distributed random variables; X_1 and S_n are treated as in Sect. 3a.

3b1 Theorem. Let $\mathbb{E}\, X_1 = 0$ and $\mathbb{E}\, X_1^2 = 1$. Then

$$\mathbb{P}\big(a\sqrt{n} < S_n < b\sqrt{n} \big) \to \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2} \, \mathrm{d}x \quad \text{as } n \to \infty$$

whenever $-\infty \le a \le b \le \infty$.

See Durrett, Sect. 2.4, Theorem (4.1).

Clearly, 2b2 is a special case.

3c Random walks on graphs

Assume that a connected finite oriented graph has m vertices, and each vertex has k outgoing edges and k incoming edges (the same k for all vertices). In addition, we assume *aperiodicity:* there exists no $p \in \{2, 3, \dots\}$ such that every loop length is divisible by p. A random walk started at a given vertex. Denote by S_n the position of the walk after n steps.

3c1 Proposition. For each vertex x of the graph,

$$\mathbb{P}\big(S_n = x \big) \to \frac{1}{m} \quad \text{as } n \to \infty\,.$$

This fact follows from a convergence theorem for Markov chains. See Durrett, Sect. 5.4, Example 4.5; also Sect. 5.5(a).

Another example: brother-sister mating.

Two animals are mated and among their direct descendants two individuals of opposite sex are selected at random. These animals are mated and the process continues.

Each individual can be one of three genotypes AA, Aa, aa. The genotype of the offspring is determined by selecting a letter from each parent.

3c2 Proposition. If the two initial animals are of genotypes Aa, aa then, after n steps, A disappears with probability that tends to 3/4 (as $n \to \infty$), while a disappears with probability that tends to 1/4.

The proof uses Markov chains and martingales.
See Durrett, Sect. 5.1, Exercise 1.5. See also Sect. 4j (below).

3d Borel-Cantelli lemmas

Given $p_1, \dots, p_n \in [0,1]$, we may consider independent events $A_1, \dots, A_n$ such that $\mathbb{P}\big(A_1\big) = p_1, \dots, \mathbb{P}\big(A_n\big) = p_n$, their indicators $X_k = \mathbf{1}_{A_k}$ (0–1 valued random variables), and the (random) number of events occurred,

$$S_n = X_1 + \dots + X_n \, .$$

The distribution of S_n is the convolution of two-point measures... Clearly, $\mathbb{E}\, S_n = p_1 + \dots + p_n$.

3d1 Proposition. (a) If $\sum_{k=1}^\infty p_k < \infty$ then

$$\lim_{M \to \infty} \lim_{n \to \infty} \mathbb{P}\big(S_n \le M\big) = 1 \, .$$

In other words, $\mathbb{P}\big(S_n \le M\big)$ tends to 1 as $M \to \infty$, *uniformly* in n.

(b) If $\sum_{k=1}^\infty p_k = \infty$ then

$$\forall M \quad \lim_{n \to \infty} \mathbb{P}\big(S_n \le M\big) = 0 \, .$$

See also Sect. 4a (below).

Part (a) can be generalized to dependent events and applied to the number of returns of the three-dimensional walk (see 2d1 for $d = 3$). However, Part (b) fails for dependent events; thus, we cannot combine it with 2d1 for $d = 1, 2$. (However, see Sect. 4i.)

4 Infinite sequences of random variables

4a Borel-Cantelli lemmas (again)

Given events $A_1, A_2, \dots \subset \Omega$, we consider the (random) number of events occurred: $S : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$,

$$S = \sum_{k=1}^\infty \mathbf{1}_{A_k} \, .$$

4a1 Theorem. (a) If $\sum_k \mathbb{P}\big(A_k\big) < \infty$ then $S < \infty$ with probability 1.

(b) If $A_1, A_2, \dots$ are independent and $\sum_k \mathbb{P}\big(A_k\big) = \infty$ then $S = \infty$ with probability 1.

See Durrett, Sect. 1.6.

For independent events, $\mathbb{P}\big(S < \infty\big)$ is either 0 or 1, which is a special case of a Zero-One Law (see Sect. 5b below).

4a2 Exercise. Let $U_1, U_2, \dots$ be independent random variables, each distributed uniformly on $(-1, 1)$. Prove that the sequence $(nU_n)_{n=1}^\infty$ is dense in $\mathbb{R}$ a.s., but the sequence $(n^2 U_n)_{n=1}^\infty$ is not.

If A_k are independent, $\mathbb{P}\big(A_k\big) \to 0$ but $\sum_k \mathbb{P}\big(A_k\big) = \infty$, then the two-valued random variables (indicators) $X_k = \mathbf{1}_{A_k}$ converge to 0 in $L_2(\Omega)$, and nevertheless, $\limsup_k X_k(\omega) = 1$ for almost all $\omega \in \Omega$. (Find a simpler, non-probabilistic example on $\Omega = (0, 1)$.)

4b Number of records

Consider independent random variables $X_1, X_2, \dots$ distributed uniformly on $(0, 1)$, events

$$A_k = \{X_k > \max(X_1, \dots, X_{k-1})\}$$

('a record occurs at time k'), and random variables

$$R_n = \mathbf{1}_{A_1} + \dots + \mathbf{1}_{A_n}$$

('the number of records at time n').

4b1 Proposition.

$$\frac{R_n}{\ln n} \to 1 \quad \text{almost surely as } n \to \infty \,.$$

Proposition 4b1 follows from Proposition 4b2.

4b2 Proposition. If $A_1, A_2, \dots$ are pairwise independent events such that $\sum \mathbb{P}\big(A_n\big) = \infty$ then

$$\frac{\mathbf{1}_{A_1} + \dots + \mathbf{1}_{A_n}}{\mathbb{P}\big(A_1\big) + \dots + \mathbb{P}\big(A_n\big)} \to 1 \quad \text{almost surely as } n \to \infty \,.$$

Proposition 4b2 is evidently stronger than the second Borel-Cantelli lemma. See Durrett, Sect. 1.6, Theorem 6.8 and Example 6.2.

4c Strong Law of Large Numbers

4c1 Theorem. Let $X_1, X_2, \dots$ be independent identically distributed random variables. If $\mathbb{E}\,|X_1| < \infty$ then

$$\frac{X_1 + \dots + X_n}{n} \to \mathbb{E}\,X_1 \quad \text{a.s. as } n \to \infty\,.$$

See Durrett, Sect. 1.7, Items (7.1) and (8.6).

Similarly to Borel-Cantelli lemmas, the Strong Law (and many other results) can be reformulated in the framework of Sect. 3; however, such a reformulation is rather cumbersome: for every $\varepsilon > 0$,

$$\mathbb{P}\Big(\max_{k\in[n,n+m]} \Big| \frac{X_1 + \dots + X_k}{k} - \mathbb{E}\,X_1 \Big| \le \varepsilon \Big) \to 1 \quad \text{as } n \to \infty$$

uniformly in m.

Note that $\mathbb{P}\big(\max \dots > \varepsilon\big)$ is usually much larger than $\max \mathbb{P}\big(\dots > \varepsilon\big)$.

Compare the Strong Law with the Weak Law (Theorem 3a1):

$$\mathbb{P}\Big(\Big| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}\,X_1 \Big| \le \varepsilon \Big) \to 1 \quad \text{as } n \to \infty\,.$$

4d Normal numbers

A real number $x \in (0,1)$ is called 10-normal, if its decimal digits $\alpha_1, \alpha_2, \dots$ defined by

$$x = \frac{\alpha_1}{10} + \frac{\alpha_2}{10^2} + \dots; \qquad \alpha_1, \alpha_2, \dots \in \{0,1,2,3,4,5,6,7,8,9\}$$

have equal frequencies, that is,

$$\frac{\#\{k \in [1,n] : \alpha_k = a\}}{n} \to \frac{1}{10} \quad \text{as } n \to \infty \quad \text{(for all } a)$$

and moreover, their combinations have equal frequencies, that is,

$$\frac{\#\{k \in [1,n] : \alpha_k = a_1, \alpha_{k+1} = a_2, \dots, \alpha_{k+l-1} = a_l\}}{n} \to \frac{1}{10^l} \quad \text{as } n \to \infty$$

for all $a_1, \dots, a_l$ and all l. Similarly, p-normal numbers are defined for any $p = 2, 3, \dots$. Finally, x is called normal, if it is p-normal for all p.

4d1 Proposition. Normal numbers exist.

It is usually said that we have no explicit example of a normal number. However, we have a (terribly slow) algorithm that constructs rational approximations to a normal number.

Proposition 4d1 follows from Proposition 4d2.

4d2 Proposition. (*Borel's normal number theorem, 1909*) Almost all numbers are normal.

That is, the set of all normal numbers is Lebesgue measurable, and its Lebesgue measure is equal to 1.

Proposition 4d2 follows from the Strong Law of Large Numbers.

See Durrett, Sect. 6.2, Example 2.5.

Do not think that the normality exhausts probabilistic properties of (digits of) real numbers. For another example see Sect. 4f (below).

4e Empirical distribution function

Let $X_1, X_2, \dots$ be independent identically distributed random variables, F their (common) (cumulative) distribution function, that is, $F(x) = \mathbb{P}\big(X_1 \le x \big)$, and

$$F_n(x) = \frac{1}{n} \# \{ k \in [1,n] : X_k \le x \} \,,$$

the so-called empirical distribution function.

4e1 Theorem. (*Glivenko-Cantelli*)

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \quad \text{amost surely, as } n \to \infty \,.$$

It follows from the Strong Law of Large Numbers.

See Durrett, Sect. 1.7, (7.4).

4f Series with random signs

Let $\beta_1, \beta_2, \dots$ be binary digits of a real number $x \in (0,1)$,

$$x = \frac{\beta_1}{2} + \frac{\beta_2}{2^2} + \dots ; \qquad \beta_1, \beta_2, \dots \in \{0,1\} \,.$$

4f1 Proposition. The series

$$\sum_{n=1}^{\infty} \frac{(-1)^{\beta_n}}{n}$$

converges for almost all x.

Convergence in L_2 is rather evident, but almost everywhere convergence is not.

Proposition 4f1 follows from Proposition 4f2.

4f2 Proposition. (*Kolmogorov*) Suppose $X_1, X_2, \dots$ are independent random variables with $\mathbb{E}\, X_n = 0$. If $\sum \operatorname{Var}(X_n) < \infty$ then the series $\sum X_n$ converges almost surely.

Proposition 4f2 follows from Kolmogorov's Three Series Theorem.

See Durett, Sect. 1.8, (8.3).

By the way, the sum of the series, $f(x) = \sum \frac{(-1)^{\beta_n}}{n}$, is a terrible (but measurable) function. Especially,

$$\operatorname{mes}\{x \in (a,b) : f(x) \in (c,d)\} > 0$$

for *all* intervals $(a,b) \subset (0,1)$, $(c,d) \subset \mathbb{R}$ (here 'mes' stands for the Lebesgue measure). Surely we cannot draw its graph!

4g Martingale convergence

Given a function $f \in L_2(0,1)$, we consider its orthogonal projection f_n to the 2^n-dimensional subspace of step functions constant on the intervals $\bigl(2^{-n}(k-1), 2^{-n}k\bigr)$. That is,

$$f_n(x) = 2^n \int_{2^{-n}(k-1)}^{2^{-n}k} f(u)\,\mathrm{d}u \quad \text{for } x \in \bigl(2^{-n}(k-1), 2^{-n}k\bigr)\,.$$

4g1 Proposition. $f_n(x) \to f(x)$ (as $n \to \infty$) for almost all $x \in (0,1)$.

Convergence in L_2 is rather evident, but almost everywhere convergence is not.

See also Sect. 5d (below).

4h Backwards martingale convergence

Given a function $f \in L_2(0,1)$, we consider its orthogonal projection f_n to the subspace of 2^{-n}-periodic functions, that is, functions $g \in L_2(0,1)$ satisfying $g(x) = g(x+2^{-n})$ whenever $0 < x < x + 2^{-n} < 1$. Thus,

$$f_n(x) = 2^{-n} \sum_{k \in \mathbb{Z}: 0 < x + 2^{-n}k < 1} f(x + 2^{-n}k) \quad \text{for } x \in (0,1)\,.$$

4h1 Proposition. $f_n(x) \to \int_0^1 f(u)\,\mathrm{d}u$ (as $n \to \infty$) for almost all $x \in (0,1)$.

Convergence in L_2 is rather evident, but almost everywhere convergence is not.

See also Sect. 5d (below).

4i Random walk: recurrence

4i1 Theorem. The simple d-dimensional random walk returns to the origin (almost sure) infinitely many times if $1 \le d \le 2$ (recurrence), but only finitely many times if $d \ge 3$ (transience).

'A drunk man will find his way home but a drunk bird may get lost forever' (Kakutani).

The proof uses Propositions 2d1 and 4i2.

4i2 Proposition. The following three conditions are equivalent for every d-dimensional random walk $(S_n)_n$:
(a) $S_n = 0$ for at least one $n \ge 1$, almost surely;
(b) $S_n = 0$ for infinitely many n, almost surely;
(c) $\sum_{n=1}^\infty \mathbb{P}\big(S_n = 0\big) = \infty$.

See Durrett, Sect. 3.2, (2.2) and (2.3).

4j Random walks on graphs (again)

Assume that a connected finite oriented graph has m vertices, and each vertex has k outgoing edges and k incoming edges (the same k for all vertices). A random walk started at a given vertex, and will stop on the first return to this point.

4j1 Proposition. The expected number of moves is equal to m (the number of vertices).

The proof uses Markov chains.
See Durrett, Sect. 5.4, (4.6) and Example 4.5.

For the 'brother-sister mating' example of Sect. 3c we can reformulate Prop. 3c2 as follows.

4j2 Proposition. If the two initial animals are of genotypes Aa, aa then finally, either A disappears (which happens with probability 3/4) or a disappears (which happens with probability 1/4).

5 Power tools: measures and sub-σ-fields

5a Conditioning as disintegration of measure

5a1 Proposition. Every probability measure μ on $\mathbb{R}^2$ is of the form

$$\iint f(x,y)\,\mu(\mathrm{d}x\mathrm{d}y) = \int \bigg(\int f(x,y)\,\mu_x(\mathrm{d}y) \bigg) \nu(\mathrm{d}x)$$

for some probability measure ν on $\mathbb{R}$ and some family $(\mu_x)_{x\in\mathbb{R}}$ of probability measures μ_x on $\mathbb{R}$; the equality holds for all bounded μ-measurable functions $f : \mathbb{R}^2 \to \mathbb{R}$. (The internal integral is a ν-measurable function of x.)

Naturally, μ_x is interpreted as the conditional distribution of y given x.

Proposition 5a1 is rather evident for discrete measures and for absolutely continuous measures (recall Fubini theorem), but not for singular measures.

Proposition 5a1 is a special case of an existence theorem of regular conditional probabilities.

See Durrett, Sect. 4.1, Theorem (1.6).

5b Zero-one laws

5b1 Proposition. For every sequence $(X_n)_n$ of independent random variables,

$$\limsup_{n\to\infty} X_n = \text{const} \quad \text{a.s.}$$

In other words: there exists $a \in (-\infty, +\infty]$ such that $\mathbb{P}\big(\limsup X_n = a\big) = 1$.

Proposition 5b1 is a special case of Proposition 5b2.

5b2 Proposition. Let a Borel function $f : \mathbb{R}^\infty \to \mathbb{R}$ satisfy

$$(x_n = y_n \text{ for all } n \text{ large enough}) \quad \Longrightarrow \quad (f(x) = f(y)) \,.$$

Then

$$f(X_1, X_2, \dots) = \text{const} \quad \text{a.s.}$$

for every sequence $(X_n)_n$ of independent random variables.

Prop. 5b2 follows from Kolmogorov's 0–1 law.

For independent, *identically distributed* random variables X_n we have two more 0–1 laws; the same conclusion under different assumptions on f. One assumption (related to ergodicity) is shift-invariance:

$$f(x_1, x_2, \dots) = f(x_2, x_3, \dots) \,.$$

Another assumption (related to the Hewitt-Savage 0–1 law) is permutation invariance:

$$f(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_{l-1}, x_l, x_{l+1}, x_{l+2}, \dots) = \\ f(x_1, \dots, x_{k-1}, x_l, x_{k+1}, \dots, x_{l-1}, x_k, x_{l+1}, x_{l+2}, \dots) \,.$$

See Durrett, Sect. 1.8, Item (8.1); Sect. 3.1, Item (1.1); Sect. 4.6, Example 6.3.

5c Exchangeability

5c1 Theorem. (*de Finetti*) The following two conditions on a probability measure μ on the space $\{0,1\}^\infty$ (of all infinite sequences of numbers 0 and 1) are equivalent.

(a) μ is invariant under permutations of coordinates;

(b) μ is a mixture of products of identical measures.

The proof uses the Hewitt-Savage 0–1 law.

The same holds for $\mathbb{R}^\infty$.

See Durrett, Sect. 4.6, Example 6.4.

5d Martingale convergence (again)

5d1 Proposition. For all random variables $Y, X_1, X_2, \dots$ such that $\mathbb{E}\,|Y|^2 < \infty$, the sequence of conditional expectations $\mathbb{E}\big(Y \,\big|\, X_1, \dots, X_n\big)$ converges almost surely (as $n \to \infty$).

See Durrett, Sect. 4.2, Item (2.10); Sect. 4.4, Theorem (4.9).

5d2 Proposition. For all random variables $Y, X_1, X_2, \dots$ such that $\mathbb{E}\,|Y|^2 < \infty$, the sequence of conditional expectations $\mathbb{E}\big(Y \,\big|\, X_n, X_{n+1}, \dots\big)$ converges almost surely (as $n \to \infty$).

See Durrett, Sect. 4.6, Theorem (6.1).

Convergence in L_2 is rather evident, but almost sure convergence is not.

Note the great generality: dependence between the random variables $Y, X_1, X_2, \dots$ is arbitrary.

Clearly, 5d1 implies 4g1, and 5d2 implies 4h1.

Rather unexpectedly, 5d2 implies also Theorem 4c1 (Strong Law of Large Numbers); see Durrett, Sect. 4.6, Example 6.1.