Chapter 2
Independence

The measure theory from the preceding chapter is a linear theory that could not describe the dependence structure of events or random variables. We enter the realm of probability theory exactly at this point, where we define independence of events and random variables. Independence is a pivotal notion of probability theory, and the computation of dependencies is one of the theory's major tasks.

In the following, $(\Omega, \mathcal{A}, \mathbf{P})$ is a probability space and the sets $A \in \mathcal{A}$ are the events. As soon as constructing probability spaces has become routine, the concrete probability space will lose its importance and it will be only the random variables that will interest us. The bold font symbol $\mathbf{P}$ will then denote the universal object of a probability measure, and the probabilities $\mathbf{P}[\cdot]$ with respect to it will always be written in square brackets.

2.1 Independence of Events

We consider two events A and B as (stochastically) independent if the occurrence of A does not change the probability that B also occurs. Formally, we say that A and B are independent if

$$\mathbf{P}[A \cap B] = \mathbf{P}[A] \cdot \mathbf{P}[B]. \tag{2.1}$$

Example 2.1 (Rolling a die twice) Consider the random experiment of rolling a die twice. Hence $\Omega = \{1, \dots, 6\}^2$ endowed with the σ-algebra $\mathcal{A} = 2^{\Omega}$ and the uniform distribution $\mathbf{P} = \mathcal{U}_{\Omega}$ (see Example 1.30(ii)).

(i) Two events A and B should be independent, e.g., if A depends only on the outcome of the first roll and B depends only on the outcome of the second roll. Formally, we assume that there are sets $\tilde{A}, \tilde{B} \subset \{1, \dots, 6\}$ such that

$$A = \tilde{A} \times \{1, \dots, 6\} \quad \text{and} \quad B = \{1, \dots, 6\} \times \tilde{B}.$$

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_2,

Now we check that A and B indeed fulfill (2.1). To this end, we compute $\mathbf{P}[A] = \frac{\#A}{36} = \frac{\#\tilde{A}}{6}$ and $\mathbf{P}[B] = \frac{\#B}{36} = \frac{\#\tilde{B}}{6}$. Furthermore,

$$\mathbf{P}[A \cap B] = \frac{\#(\tilde{A} \times \tilde{B})}{36} = \frac{\#\tilde{A}}{6} \cdot \frac{\#\tilde{B}}{6} = \mathbf{P}[A] \cdot \mathbf{P}[B].$$

(ii) Stochastic independence can occur also in less obvious situations. For instance, let A be the event where the sum of the two rolls is odd,

$$A = \big\{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \in \{3, 5, 7, 9, 11\}\big\},$$

and let B be the event where the first roll gives at most a three

$$B = \big\{(\omega_1, \omega_2) \in \Omega : \omega_1 \in \{1, 2, 3\}\big\}.$$

Although it might seem that these two events are entangled in some way, they are stochastically independent. Indeed, it is easy to check that $\mathbf{P}[A] = \mathbf{P}[B] = \frac{1}{2}$ and $\mathbf{P}[A \cap B] = \frac{1}{4}$. ◊

What is the condition for *three* events A_1, A_2, A_3 to be independent? Of course, any of the pairs (A_1, A_2), (A_1, A_3) and (A_2, A_3) has to be independent. However, we have to make sure also that the simultaneous occurrence of A_1 *and* A_2 does not change the probability that A_3 occurs. Hence, it is not enough to consider pairs only.

Formally, we call three events A_1, A_2 and A_3 (stochastically) independent if

$$\mathbf{P}[A_i \cap A_j] = \mathbf{P}[A_i] \cdot \mathbf{P}[A_j] \quad \text{for all } i, j \in \{1, 2, 3\},\ i \neq j, \tag{2.2}$$

and

$$\mathbf{P}[A_1 \cap A_2 \cap A_3] = \mathbf{P}[A_1] \cdot \mathbf{P}[A_2] \cdot \mathbf{P}[A_3]. \tag{2.3}$$

Note that (2.2) does not imply (2.3) (and (2.3) does not imply (2.2)).

Example 2.2 (Rolling a die three times) We roll a die three times. Hence $\Omega = \{1, \ldots, 6\}^3$ endowed with the discrete σ-algebra $\mathcal{A} = 2^\Omega$ and the uniform distribution $\mathbf{P} = \mathcal{U}_\Omega$ (see Example 1.30(ii)).

(i) If we assume that for any $i = 1, 2, 3$ the event A_i depends only on the outcome of the ith roll, then the events A_1, A_2 and A_3 are independent. Indeed, as in the preceding example, there are sets $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3 \subset \{1, \ldots, 6\}$ such that

$$A_1 = \tilde{A}_1 \times \{1, \ldots, 6\}^2,$$
$$A_2 = \{1, \ldots, 6\} \times \tilde{A}_2 \times \{1, \ldots, 6\},$$
$$A_3 = \{1, \ldots, 6\}^2 \times \tilde{A}_3.$$

The validity of (2.2) follows as in Example 2.1(i). In order to show (2.3), we compute

$$\mathbf{P}[A_1 \cap A_2 \cap A_3] = \frac{\#(\tilde{A}_1 \times \tilde{A}_2 \times \tilde{A}_3)}{216} = \prod_{i=1}^{3} \frac{\#\tilde{A}_i}{6} = \prod_{i=1}^{3} \mathbf{P}[A_i].$$

(ii) Consider now the events

$$A_1 := \{\omega \in \Omega : \omega_1 = \omega_2\},$$
$$A_2 := \{\omega \in \Omega : \omega_2 = \omega_3\},$$
$$A_3 := \{\omega \in \Omega : \omega_1 = \omega_3\}.$$

Then $\#A_1 = \#A_2 = \#A_3 = 36$; hence $\mathbf{P}[A_1] = \mathbf{P}[A_2] = \mathbf{P}[A_3] = \frac{1}{6}$. Furthermore, $\#(A_i \cap A_j) = 6$ if $i \neq j$; hence $\mathbf{P}[A_i \cap A_j] = \frac{1}{36}$. Hence (2.2) holds. On the other hand, we have $\#(A_1 \cap A_2 \cap A_3) = 6$, thus $\mathbf{P}[A_1 \cap A_2 \cap A_3] = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$. Thus (2.3) does not hold and so the events A_1, A_2, A_3 are not independent. ◊

In order to define independence of larger families of events, we have to request the validity of product formulas, such as (2.2) and (2.3), not only for pairs and triples but for all finite subfamilies of events. We thus make the following definition.

Definition 2.3 (Independence of events) Let I be an arbitrary index set and let $(A_i)_{i \in I}$ be an arbitrary family of events. The family $(A_i)_{i \in I}$ is called *independent* if for any finite subset $J \subset I$ the product formula holds:

$$\mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j].$$

The most prominent example of an independent family of infinitely many events is given by the perpetuated independent repetition of a random experiment.

Example 2.4 Let E be a finite set (the set of possible outcomes of the individual experiment) and let $(p_e)_{e \in E}$ be a probability vector on E. Equip (as in Theorem 1.64) the probability space $\Omega = E^{\mathbb{N}}$ with the σ-algebra $\mathcal{A} = \sigma(\{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E, n \in \mathbb{N}\})$ and with the product measure (or Bernoulli measure) $\mathbf{P} = \left(\sum_{e \in E} p_e \delta_e\right)^{\otimes \mathbb{N}}$; that is where $\mathbf{P}[[\omega_1, \ldots, \omega_n]] = \prod_{i=1}^{n} p_{\omega_i}$. Let $\tilde{A}_i \subset E$ for any $i \in \mathbb{N}$, and let A_i be the event where $\tilde{A}_i$ occurs in the ith experiment; that is,

$$A_i = \left\{\omega \in \Omega : \omega_i \in \tilde{A}_i\right\} = \biguplus_{(\omega_1, \ldots, \omega_i) \in E^{i-1} \times \tilde{A}_i} [\omega_1, \ldots, \omega_i].$$

Intuitively, the family $(A_i)_{i \in \mathbb{N}}$ should be independent if the definition of independence makes any sense at all. We check that this is indeed the case. Let $J \subset \mathbb{N}$ be finite and $n := \max J$. Formally, we define $B_j = A_j$ and $\tilde{B}_j = \tilde{A}_j$ for $j \in J$ and $B_j = \Omega$ and $\tilde{B}_j = E$ for $j \in \{1, \ldots, n\} \setminus J$. Then

$$\mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \mathbf{P}\left[\bigcap_{j \in J} B_j\right] = \mathbf{P}\left[\bigcap_{j=1}^{n} B_j\right]$$

$$= \sum_{e_1 \in \tilde{B}_1} \cdots \sum_{e_n \in \tilde{B}_n} \prod_{j=1}^{n} p_{e_j} = \prod_{j=1}^{n} \Big(\sum_{e \in \tilde{B}_j} p_e \Big) = \prod_{j \in J} \Big(\sum_{e \in \tilde{A}_j} p_e \Big).$$

This is true in particular for $\#J = 1$. Hence $\mathbf{P}[A_i] = \sum_{e \in \tilde{A}_i} p_e$ for all $i \in \mathbb{N}$, whence

$$\mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j]. \tag{2.4}$$

Since this holds for all finite $J \subset \mathbb{N}$, the family $(A_i)_{i \in \mathbb{N}}$ is independent. ◊

If A and B are independent, then A^c and B also are independent since $\mathbf{P}[A^c \cap B] = \mathbf{P}[B] - \mathbf{P}[A \cap B] = \mathbf{P}[B] - \mathbf{P}[A]\mathbf{P}[B] = (1 - \mathbf{P}[A])\mathbf{P}[B] = \mathbf{P}[A^c]\mathbf{P}[B]$. We generalize this observation in the following theorem.

Theorem 2.5 *Let I be an arbitrary index set and let $(A_i)_{i \in I}$ be a family of events. Define $B_i^0 = A_i$ and $B_i^1 = A_i^c$ for $i \in I$. Then the following three statements are equivalent.*

(i) *The family $(A_i)_{i \in I}$ is independent.*
(ii) *There is an $\alpha \in \{0, 1\}^I$ such that the family $(B_i^{\alpha_i})_{i \in I}$ is independent.*
(iii) *For any $\alpha \in \{0, 1\}^I$, the family $(B_i^{\alpha_i})_{i \in I}$ is independent.*

Proof This is left as an exercise. □

Example 2.6 (Euler's prime number formula) The *Riemann zeta function* is defined by the Dirichlet series

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s} \quad \text{for } s \in (1, \infty).$$

Euler's prime number formula is a representation of the Riemann zeta function as an infinite product

$$\zeta(s) = \prod_{p \in \mathcal{P}} \left(1 - p^{-s}\right)^{-1}, \tag{2.5}$$

where $\mathcal{P} := \{p \in \mathbb{N} : p \text{ is prime}\}$.

We give a probabilistic proof for this formula. Let $\Omega = \mathbb{N}$, and for fixed $s > 1$ define $\mathbf{P}$ on 2^Ω by

$$\mathbf{P}\big[\{n\}\big] = \zeta(s)^{-1} n^{-s} \quad \text{for } n \in \mathbb{N}.$$

Let $p\mathbb{N} = \{pn : n \in \mathbb{N}\}$ and $\mathcal{P}_n = \{p \in \mathcal{P} : p \le n\}$. We consider $p\mathbb{N} \subset \Omega$ as an event. Note that $\mathbf{P}[p\mathbb{N}] = p^{-s}$ and that $(p\mathbb{N}, p \in \mathcal{P})$ is independent. Indeed, for $k \in \mathbb{N}$ and mutually distinct $p_1, \ldots, p_k \in \mathcal{P}$, we have $\bigcap_{i=1}^k (p_i\mathbb{N}) = (p_1 \ldots p_k)\mathbb{N}$. Thus

$$\begin{aligned}
\mathbf{P}\left[\bigcap_{i=1}^k (p_i\mathbb{N})\right] &= \sum_{n=1}^\infty \mathbf{P}\big[\{p_1 \ldots p_k n\}\big] \\
&= \zeta(s)^{-1}(p_1 \ldots p_k)^{-s} \sum_{n=1}^\infty n^{-s} \\
&= (p_1 \ldots p_k)^{-s} = \prod_{i=1}^k \mathbf{P}[p_i\mathbb{N}].
\end{aligned}$$

By Theorem 2.5, the family $((p\mathbb{N})^c, p \in \mathcal{P})$ is also independent, whence

$$\begin{aligned}
\zeta(s)^{-1} = \mathbf{P}\big[\{1\}\big] &= \mathbf{P}\left[\bigcap_{p \in \mathcal{P}} (p\mathbb{N})^c\right] \\
&= \lim_{n\to\infty} \mathbf{P}\left[\bigcap_{p \in \mathcal{P}_n} (p\mathbb{N})^c\right] \\
&= \lim_{n\to\infty} \prod_{p \in \mathcal{P}_n} \big(1 - \mathbf{P}[p\mathbb{N}]\big) = \prod_{p \in \mathcal{P}} \big(1 - p^{-s}\big).
\end{aligned}$$

This shows (2.5). ◊

If we roll a die infinitely often, what is the chance that the face shows a six infinitely often? This probability should equal one. Otherwise there would be a last point in time when we see a six and after which the face only shows a number one to five. However, this is not very plausible.

Recall that we formalized the event where infinitely many of a series of events occur by means of the limes superior (see Definition 1.13). The following theorem confirms the conjecture mentioned above and also gives conditions under which we *cannot* expect that infinitely many of the events occur.

Theorem 2.7 (Borel–Cantelli lemma) *Let $A_1, A_2, \ldots$ be events and define $A^* = \limsup_{n\to\infty} A_n$.*

(i) *If $\sum_{n=1}^\infty \mathbf{P}[A_n] < \infty$, then $\mathbf{P}[A^*] = 0$. (Here $\mathbf{P}$ could be an arbitrary measure on $(\Omega, \mathcal{A})$.)*
(ii) *If $(A_n)_{n\in\mathbb{N}}$ is independent and $\sum_{n=1}^\infty \mathbf{P}[A_n] = \infty$, then $\mathbf{P}[A^*] = 1$.*

Proof (i) **P** is upper semicontinuous and σ-subadditive; hence, by assumption,

$$\mathbf{P}\big[A^*\big] = \lim_{n\to\infty} \mathbf{P}\left[\bigcup_{m=n}^{\infty} A_m\right] \le \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbf{P}[A_m] = 0.$$

(ii) De Morgan's rule and the lower semicontinuity of **P** yield

$$\mathbf{P}\big[\big(A^*\big)^c\big] = \mathbf{P}\left[\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} A_n^c\right] = \lim_{m\to\infty} \mathbf{P}\left[\bigcap_{n=m}^{\infty} A_n^c\right].$$

However, for every $m \in \mathbb{N}$ (since $\log(1-x) \le -x$ for $x \in [0,1]$), by upper continuity of **P**

$$\mathbf{P}\left[\bigcap_{n=m}^{\infty} A_n^c\right] = \lim_{N\to\infty} \mathbf{P}\left[\bigcap_{n=m}^{N} A_n^c\right] = \prod_{n=m}^{\infty}\big(1-\mathbf{P}[A_n]\big)$$
$$= \exp\left(\sum_{n=m}^{\infty} \log\big(1-\mathbf{P}[A_n]\big)\right) \le \exp\left(-\sum_{n=m}^{\infty} \mathbf{P}[A_n]\right) = 0. \qquad \square$$

Example 2.8 We throw a die again and again and ask for the probability of seeing a six infinitely often. Hence $\Omega = \{1,\ldots,6\}^{\mathbb{N}}$, $\mathcal{A} = (2^{\{1,\ldots,6\}})^{\otimes\mathbb{N}}$ is the product σ-algebra and $\mathbf{P} = \big(\sum_{e\in\{1,\ldots,6\}} \frac{1}{6}\delta_e\big)^{\otimes\mathbb{N}}$ is the Bernoulli measure (see Theorem 1.64). Furthermore, let $A_n = \{\omega \in \Omega : \omega_n = 6\}$ be the event where the nth roll shows a six. Then $A^* = \limsup_{n\to\infty} A_n$ is the event where we see a six infinitely often (see Remark 1.14). Furthermore, $(A_n)_{n\in\mathbb{N}}$ is an independent family with the property $\sum_{n=1}^{\infty} \mathbf{P}[A_n] = \sum_{n=1}^{\infty} \frac{1}{6} = \infty$. Hence the Borel–Cantelli lemma yields $\mathbf{P}[A^*] = 1$. ◊

Example 2.9 We roll a die only once and define A_n for any $n \in \mathbb{N}$ as the event where in this one roll the face showed a six. Note that $A_1 = A_2 = A_3 = \ldots$. Then $\sum_{n\in\mathbb{N}} \mathbf{P}[A_n] = \infty$; however, $\mathbf{P}[A^*] = \mathbf{P}[A_1] = \frac{1}{6}$. This shows that in Part (ii) of the Borel–Cantelli lemma, the assumption of independence is indispensable. ◊

Example 2.10 Let $\Lambda \in (0,\infty)$ and $0 \le \lambda_n \le \Lambda$ for $n \in \mathbb{N}$. Let X_n, $n \in \mathbb{N}$, be Poisson random variables with parameters λ_n. Then

$$\mathbf{P}[X_n \ge n \text{ for infinitely many } n] = 0.$$

Indeed,

$$\sum_{n=1}^{\infty} \mathbf{P}[X_n \ge n] = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \mathbf{P}[X_n = m] = \sum_{m=1}^{\infty}\sum_{n=1}^{m} \mathbf{P}[X_n = m]$$
$$= \sum_{m=1}^{\infty}\sum_{n=1}^{m} e^{-\lambda_n}\frac{\lambda_n^m}{m!} \le \sum_{m=1}^{\infty} m\frac{\Lambda^m}{m!} = \Lambda e^{\Lambda} < \infty. \qquad ◊$$

Note that in Theorem 2.7 in the case of independent events, only the probabilities $\mathbf{P}[A^*] = 0$ and $\mathbf{P}[A^*] = 1$ could show up. Thus the Borel–Cantelli lemma belongs to the class of so-called 0–1 laws. Later we will encounter more 0–1 laws (see, for example, Theorem 2.37).

Now we extend the notion of independence from families of events to families of classes of events.

Definition 2.11 (Independence of classes of events) Let I be an arbitrary index set and let $\mathcal{E}_i \subset \mathcal{A}$ for all $i \in I$. The family $(\mathcal{E}_i)_{i\in I}$ is called *independent* if, for any finite subset $J \subset I$ and any choice of $E_j \in \mathcal{E}_j$, $j \in J$, we have

$$\mathbf{P}\left[\bigcap_{j\in J} E_j\right] = \prod_{j\in J} \mathbf{P}[E_j]. \tag{2.6}$$

Example 2.12 As in Example 2.4, let $(\Omega, \mathcal{A}, \mathbf{P})$ be the product space of infinitely many repetitions of a random experiment whose possible outcomes e are the elements of the finite set E and have probabilities $p = (p_e)_{e\in E}$. For $i \in \mathbb{N}$, define

$$\mathcal{E}_i = \big\{\{\omega \in \Omega : \omega_i \in A\} : A \subset E\big\}.$$

For any choice of sets $A_i \in \mathcal{E}_i$, $i \in \mathbb{N}$, the family $(A_i)_{i\in\mathbb{N}}$ is independent; hence $(\mathcal{E}_i)_{i\in\mathbb{N}}$ is independent. ◊

Theorem 2.13

(i) *Let I be finite, and for any $i \in I$ let $\mathcal{E}_i \subset \mathcal{A}$ with $\Omega \in \mathcal{E}_i$. Then*

$$(\mathcal{E}_i)_{i\in I} \text{ is independent} \iff (2.6) \text{ holds for } J = I.$$

(ii) *$(\mathcal{E}_i)_{i\in I}$ is independent $\iff$ $((\mathcal{E}_j)_{j\in J}$ is independent for all finite $J \subset I)$.*

(iii) *If $(\mathcal{E}_i \cup \{\emptyset\})$ is $\cap$-stable, then*

$$(\mathcal{E}_i)_{i\in I} \text{ is independent} \iff \big(\sigma(\mathcal{E}_i)\big)_{i\in I} \text{ is independent.}$$

(iv) *Let K be an arbitrary set and let $(I_k)_{k\in K}$ be mutually disjoint subsets of I. If $(\mathcal{E}_i)_{i\in I}$ is independent, then $(\bigcup_{i\in I_k} \mathcal{E}_i)_{k\in K}$ is also independent.*

Proof (i) "$\Longrightarrow$" This is trivial.

(i) "$\Longleftarrow$" For $J \subset I$ and $j \in I \setminus J$, choose $E_j = \Omega$.

(ii) This is trivial.

(iii) "$\Longleftarrow$" This is trivial.

(iii) "$\Longrightarrow$" Let $J \subset I$ be finite. We will show that for any two finite sets J and J' with $J \subset J' \subset I$,

$$\mathbf{P}\left[\bigcap_{i\in J'} E_i\right] = \prod_{i\in J'} \mathbf{P}[E_i]$$
$$\text{for any choice } \begin{cases} E_i \in \sigma(\mathcal{E}_i), & i \in J, \\ E_i \in \mathcal{E}_i, & i \in J' \setminus J. \end{cases} \tag{2.7}$$

The case $J' = J$ is exactly the claim we have to show.

We carry out the proof of (2.7) by induction on $\#J$. For $\#J = 0$, the statement (2.7) holds by assumption of this theorem.

Now assume that (2.7) holds for every J with $\#J = n$ and for every finite $J' \supset J$. Fix such a J and let $j \in I \setminus J$. Choose $J' \supset \tilde{J} := J \cup \{j\}$. We show the validity of (2.7) with J replaced by $\tilde{J}$. Since $\#\tilde{J} = n + 1$, this verifies the induction step.

Let $E_i \in \sigma(\mathcal{E}_i)$ for any $i \in J$, and let $E_i \in \mathcal{E}_i$ for any $i \in J' \setminus (J \cup \{j\})$. Define two measures μ and ν on $(\Omega, \mathcal{A})$ by

$$\mu : E_j \mapsto \mathbf{P}\left[\bigcap_{i\in J'} E_i\right] \quad \text{and} \quad \nu : E_j \mapsto \prod_{i\in J'} \mathbf{P}[E_i].$$

By the induction hypothesis (2.7), we have $\mu(E_j) = \nu(E_j)$ for every $E_j \in \mathcal{E}_j \cup \{\emptyset, \Omega\}$. Since $\mathcal{E}_j \cup \{\emptyset\}$ is a π-system, Lemma 1.42 yields that $\mu(E_j) = \nu(E_j)$ for all $E_j \in \sigma(\mathcal{E}_j)$. That is, (2.7) holds with J replaced by $J \cup \{j\}$.

(iv) This is trivial, as (2.6) has to be checked only for $J \subset I$ with

$$\#(J \cap I_k) \le 1 \quad \text{for any } k \in K. \qquad \square$$

2.2 Independent Random Variables

Now that we have studied independence of events, we want to study independence of random variables. Here also the definition ends up with a product formula. Formally, however, we can also define independence of random variables via independence of the σ-algebras they generate. This is the reason why we studied independence of classes of events in the last section.

Independent random variables allow for a rich calculus. For example, we can compute the distribution of a sum of two independent random variables by a simple convolution formula. Since we do not have a general notion of an integral at hand at this point, for the time being we restrict ourselves to presenting the convolution formula for integer-valued random variables only.

Let I be an arbitrary index set. For each $i \in I$, let $(\Omega_i, \mathcal{A}_i)$ be a measurable space and let $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$ be a random variable with generated σ-algebra $\sigma(X_i) = X_i^{-1}(\mathcal{A}_i)$.

Definition 2.14 (Independent random variables) The family $(X_i)_{i\in I}$ of random variables is called *independent* if the family $(\sigma(X_i))_{i\in I}$ of σ-algebras is independent.

As a shorthand, we say that a family $(X_i)_{i\in I}$ is "i.i.d." (for "independent and identically distributed") if $(X_i)_{i\in I}$ is independent and if $\mathbf{P}_{X_i} = \mathbf{P}_{X_j}$ for all $i, j \in I$.

Remark 2.15

(i) Clearly, the family $(X_i)_{i\in I}$ is independent if and only if, for any finite set $J \subset I$ and any choice of $A_j \in \mathcal{A}_j$, $j \in J$, we have

$$\mathbf{P}\left[\bigcap_{j\in J}\{X_j \in A_j\}\right] = \prod_{j\in J}\mathbf{P}[X_j \in A_j].$$

The next theorem will show that it is enough to request the validity of such a product formula for A_j from an $\cap$-stable generator of $\mathcal{A}_j$ only.

(ii) If $(\tilde{\mathcal{A}}_i)_{i\in I}$ is an independent family of σ-algebras and if each X_i is $\tilde{\mathcal{A}}_i - \mathcal{A}_i$-measurable, then $(X_i)_{i\in I}$ is independent. This is a direct consequence of the fact that $\sigma(X_i) \subset \tilde{\mathcal{A}}_i$.

(iii) For each $i \in I$, let $(\Omega_i', \mathcal{A}_i')$ be another measurable space and assume that $f_i : (\Omega_i, \mathcal{A}_i) \to (\Omega_i', \mathcal{A}_i')$ is a measurable map. If $(X_i)_{i\in I}$ is independent, then $(f_i \circ X_i)_{i\in I}$ is independent. This statement is a special case of (ii) since $f_i \circ X_i$ is $\sigma(X_i) - \mathcal{A}_i'$-measurable (see Theorem 1.80). ◊

Theorem 2.16 (Independent generators) *For any $i \in I$, let $\mathcal{E}_i \subset \mathcal{A}_i$ be a π-system that generates $\mathcal{A}_i$. If $(X_i^{-1}(\mathcal{E}_i))_{i\in I}$ is independent, then $(X_i)_{i\in I}$ is independent.*

Proof By Theorem 1.81, $X_i^{-1}(\mathcal{E}_i)$ is a π-system that generates the σ-algebra $X_i^{-1}(\mathcal{A}_i) = \sigma(X_i)$. Hence the statement follows from Theorem 2.13. □

Example 2.17 Let E be a countable set and let $(X_i)_{i\in I}$ be random variables with values in $(E, 2^E)$. In this case, $(X_i)_{i\in I}$ is independent if and only if, for any finite $J \subset I$ and any choice of $x_j \in E$, $j \in J$,

$$\mathbf{P}[X_j = x_j \text{ for all } j \in J] = \prod_{j\in J}\mathbf{P}[X_j = x_j].$$

This is obvious since $\{\{x\} : x \in E\} \cup \{\emptyset\}$ is a π-system that generates 2^E, thus $\{X_i^{-1}(\{x_i\}) : x_i \in E\} \cup \{\emptyset\}$ is a π-system that generates $\sigma(X_i)$ (Theorem 1.81). ◊

Example 2.18 Let E be a finite set and let $p = (p_e)_{e\in E}$ be a probability vector. Repeat a random experiment with possible outcomes $e \in E$ and probabilities p_e for $e \in E$ infinitely often (see Example 1.40 and Theorem 1.64). Let $\Omega = E^{\mathbb{N}}$ be the infinite product space and let $\mathcal{A}$ be the σ-algebra generated by the cylinder sets (see (1.8)). Let $\mathbf{P} = (\sum_{e\in E} p_e\delta_e)^{\otimes\mathbb{N}}$ be the Bernoulli measure. Further, for any $n \in \mathbb{N}$, let

$$X_n : \Omega \to E, \quad (\omega_m)_{m\in\mathbb{N}} \mapsto \omega_n,$$

be the projection on the nth coordinate. In other words: For any simple event $\omega \in \Omega$, $X_n(\omega)$ yields the result of the nth experiment. Then, by (2.4) (in Example 2.4), for $n \in \mathbb{N}$ and $x \in E^n$, we have

$$\begin{aligned}
&\mathbf{P}[X_j = x_j \text{ for all } j = 1, \ldots, n] \\
&\quad = \mathbf{P}\big[[x_1, \ldots, x_n]\big] = \mathbf{P}\left[\bigcap_{j=1}^{n} X_j^{-1}\big(\{x_j\}\big)\right] \\
&\quad = \prod_{j=1}^{n} \mathbf{P}\big[X_j^{-1}\big(\{x_j\}\big)\big] = \prod_{j=1}^{n} \mathbf{P}[X_j = x_j],
\end{aligned}$$

and $\mathbf{P}[X_j = x_j] = p_{x_j}$. By virtue of Theorem 2.13(i), this implies that the family $(X_1, \ldots, X_n)$ is independent and hence, by Theorem 2.13(ii), $(X_n)_{n\in\mathbb{N}}$ is independent as well. ◊

In particular, we have shown the following theorem.

Theorem 2.19 *Let E be a finite set and let $(p_e)_{e\in E}$ be a probability vector on E. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and an independent family $(X_n)_{n\in\mathbb{N}}$ of E-valued random variables on $(\Omega, \mathcal{A}, \mathbf{P})$ such that $\mathbf{P}[X_n = e] = p_e$ for any $e \in E$.*

Later we will see that the assumption that E is finite can be dropped. Also one can allow for different distributions in the respective factors. For the time being, however, this theorem gives us enough examples of interesting families of independent random variables.

Our next goal is to deduce simple criteria in terms of distribution functions and densities for checking whether a family of random variables is independent or not.

Definition 2.20 For any $i \in I$, let X_i be a real random variable. For any finite subset $J \subset I$, let

$$\begin{aligned}
F_J := F_{(X_j)_{j\in J}} &: \mathbb{R}^J \to [0, 1], \\
&x \mapsto \mathbf{P}[X_j \le x_j \text{ for all } j \in J] = \mathbf{P}\left[\bigcap_{j\in J} X_j^{-1}\big((-\infty, x_j]\big)\right].
\end{aligned}$$

Then F_J is called the *joint distribution function* of $(X_j)_{j\in J}$. The probability measure $\mathbf{P}_{(X_j)_{j\in J}}$ on $\mathbb{R}^J$ is called the *joint distribution* of $(X_j)_{j\in J}$.

Theorem 2.21 *A family $(X_i)_{i\in I}$ of real random variables is independent if and only if, for every finite $J \subset I$ and every $x = (x_j)_{j\in J} \in \mathbb{R}^J$,*

$$F_J(x) = \prod_{j\in J} F_{\{j\}}(x_j). \tag{2.8}$$

Proof The class of sets $\{(-\infty, b], b \in \mathbb{R}\}$ is an $\cap$-stable generator of the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ (see Theorem 1.23). Equation (2.8) says that, for any choice of real numbers $(x_i)_{i\in I}$, the events $(X_i^{-1}((-\infty, x_i]))_{i\in I}$ are independent. Hence Theorem 2.16 yields the claim. □

Corollary 2.22 *In addition to the assumptions of Theorem 2.21, we assume that any F_J has a continuous* density $f_J = f_{(X_j)_{j\in J}}$ *(the* joint density *of $(X_j)_{j\in J}$). That is, there exists a continuous map $f_J : \mathbb{R}^J \to [0, \infty)$ such that*

$$F_J(x) = \int_{-\infty}^{x_{j_1}} dt_1 \dots \int_{-\infty}^{x_{j_n}} dt_n \, f_J(t_1, \dots, t_n) \quad \textit{for all } x \in \mathbb{R}^J$$

(where $J = \{j_1, \dots, j_n\}$). In this case, the family $(X_i)_{i\in I}$ is independent if and only if, for any finite $J \subset I$

$$f_J(x) = \prod_{j\in J} f_j(x_j) \quad \textit{for all } x \in \mathbb{R}^J. \tag{2.9}$$

Corollary 2.23 *Let $n \in \mathbb{N}$ and let $\mu_1, \dots, \mu_n$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and an independent family of random variables $(X_i)_{i=1,\dots,n}$ on $(\Omega, \mathcal{A}, \mathbf{P})$ with $\mathbf{P}_{X_i} = \mu_i$ for each $i = 1, \dots, n$.*

Proof Let $\Omega = \mathbb{R}^n$ and $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$. Let $\mathbf{P} = \bigotimes_{i=1}^n \mu_i$ be the product measure of the μ_i (see Theorem 1.61). Further, let $X_i : \mathbb{R}^n \to \mathbb{R}$, $(x_1, \dots, x_n) \mapsto x_i$ be the projection on the ith coordinate for each $i = 1, \dots, n$. Then, for any $i = 1, \dots, n$,

$$\begin{aligned} F_{\{i\}}(x) &= \mathbf{P}[X_i \le x] = \mathbf{P}\big[\mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{n-i}\big] \\ &= \mu_i\big((-\infty, x]\big) \cdot \prod_{j\neq i} \mu_j(\mathbb{R}) = \mu_i\big((-\infty, x]\big). \end{aligned}$$

Hence indeed $\mathbf{P}_{X_i} = \mu_i$. Furthermore, for all $x_1, \dots, x_n \in \mathbb{R}$,

$$\begin{aligned} F_{\{1,\dots,n\}}\big((x_1, \dots, x_n)\big) &= \mathbf{P}\left[\bigtimes_{i=1}^n (-\infty, x_i]\right] = \prod_{i=1}^n \mu_i\big((-\infty, x_i]\big) \\ &= \prod_{i=1}^n F_{\{i\}}(x_i). \end{aligned}$$

Hence Theorem 2.21 (and Theorem 2.13(i)) yields the independence of $(X_i)_{i=1,\dots,n}$. □

Example 2.24 Let $X_1, \ldots, X_n$ be independent exponentially distributed random variables with parameters $\theta_1, \ldots, \theta_n \in (0, \infty)$. Then

$$F_{\{i\}}(x) = \int_0^x \theta_i e^{-\theta_i t}\, dt = 1 - e^{-\theta_i x} \quad \text{for } x \geq 0$$

and hence

$$F_{\{1,\ldots,n\}}\big((x_1, \ldots, x_n)\big) = \prod_{i=1}^n \big(1 - e^{-\theta_i x_i}\big).$$

Consider now the random variable $Y = \max(X_1, \ldots, X_n)$. Then

$$\begin{aligned} F_Y(x) &= \mathbf{P}[X_i \leq x \text{ for all } i = 1, \ldots, n] \\ &= F_{\{1,\ldots,n\}}\big((x, \ldots, x)\big) = \prod_{i=1}^n \big(1 - e^{-\theta_i x}\big). \end{aligned}$$

The distribution function of the random variable $Z := \min(X_1, \ldots, X_n)$ has a nice closed form:

$$\begin{aligned} F_Z(x) &= 1 - \mathbf{P}[Z > x] \\ &= 1 - \mathbf{P}[X_i > x \text{ for all } i = 1, \ldots, n] \\ &= 1 - \prod_{i=1}^n e^{-\theta_i x} = 1 - \exp\big(-(\theta_1 + \ldots + \theta_n)x\big). \end{aligned}$$

In other words, Z is exponentially distributed with parameter $\theta_1 + \ldots + \theta_n$. ◊

Example 2.25 Let $\mu_i \in \mathbb{R}$ and $\sigma_i^2 > 0$ for $i \in I$. Let $(X_i)_{i \in I}$ be real random variables with joint density functions (for finite $J \subset I$)

$$f_J(x) = \prod_{j \in J} \big(2\pi\sigma_j^2\big)^{-\frac{1}{2}} \exp\Bigg(-\sum_{j \in J} \frac{(x_j - \mu_j)^2}{2\sigma_j^2}\Bigg) \quad \text{for } x \in \mathbb{R}^J.$$

Then $(X_i)_{i \in I}$ is independent and X_i is normally distributed with parameters (μ_i, σ_i^2).

For any finite $I = \{i_1, \ldots, i_n\}$ (with mutually distinct $i_1, \ldots, i_n$), the vector $Y = (X_{i_1}, \ldots, X_{i_n})$ has the n-dimensional normal distribution with $\mu = \mu^I := (\mu_{i_1}, \ldots, \mu_{i_n})$ and with $\Sigma = \Sigma^I$ the diagonal matrix with entries $\sigma_{i_1}^2, \ldots, \sigma_{i_n}^2$ (see Example 1.105(ix)). ◊

Theorem 2.26 *Let K be an arbitrary set and I_k, $k \in K$, arbitrary mutually disjoint index sets. Define $I = \bigcup_{k \in K} I_k$.*
If the family $(X_i)_{i \in I}$ is independent, then the family of σ-algebras $(\sigma(X_j, j \in I_k))_{k \in K}$ is independent.

Proof For $k \in K$, let

$$\mathcal{Z}_k = \left\{ \bigcap_{j \in I_k} A_j : A_j \in \sigma(X_j), \#\{j \in I_k : A_j \neq \Omega\} < \infty \right\}$$

be the semiring of finite-dimensional rectangular cylinder sets. Clearly, $\mathcal{Z}_k$ is a π-system and $\sigma(\mathcal{Z}_k) = \sigma(X_j, j \in I_k)$. Hence, by Theorem 2.13(iii), it is enough to show that $(\mathcal{Z}_k)_{k \in K}$ is independent. By Theorem 2.13(ii), we can even assume that K is finite.

For $k \in K$, let $B_k \in \mathcal{Z}_k$ and $J_k \subset I_k$ be finite with $B_k = \bigcap_{j \in J_k} A_j$ for certain $A_j \in \sigma(X_j)$. Define $J = \bigcup_{k \in K} J_k$. Then

$$\mathbf{P}\left[\bigcap_{k \in K} B_k\right] = \mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j] = \prod_{k \in K} \prod_{j \in J_k} \mathbf{P}[A_j] = \prod_{k \in K} \mathbf{P}[B_k].$$

□

Example 2.27 If $(X_n)_{n \in \mathbb{N}}$ is an independent family of real random variables, then also $(Y_n)_{n \in \mathbb{N}} = (X_{2n} - X_{2n-1})_{n \in \mathbb{N}}$ is independent. Indeed, for any $n \in \mathbb{N}$, the random variable Y_n is $\sigma(X_{2n}, X_{2n-1})$-measurable by Theorem 1.91, and $(\sigma(X_{2n}, X_{2n-1}))_{n \in \mathbb{N}}$ is independent by Theorem 2.26. ◊

Example 2.28 Let $(X_{m,n})_{(m,n) \in \mathbb{N}^2}$ be an independent family of Bernoulli random variables with parameter $p \in (0, 1)$. Define the waiting time for the first "success" in the mth row of the matrix $(X_{m,n})_{m,n}$ by

$$Y_m := \inf\{n \in \mathbb{N} : X_{m,n} = 1\} - 1.$$

Then $(Y_m)_{m \in \mathbb{N}}$ are independent geometric random variables with parameter p (see Example 1.105(iii)). Indeed,

$$\{Y_m \le k\} = \bigcup_{l=1}^{k+1} \{X_{m,l} = 1\} \in \sigma(X_{m,l}, l = 1, \ldots, k+1) \subset \sigma(X_{m,l}, l \in \mathbb{N}).$$

Hence Y_m is $\sigma(X_{m,l}, l \in \mathbb{N})$-measurable and thus $(Y_m)_{m \in \mathbb{N}}$ is independent. Furthermore,

$$\mathbf{P}[Y_m > k] = \mathbf{P}[X_{m,l} = 0, l = 1, \ldots, k+1] = \prod_{l=1}^{k+1} \mathbf{P}[X_{m,l} = 0] = (1-p)^{k+1}.$$

Concluding, we get $\mathbf{P}[Y_m = k] = \mathbf{P}[Y_m > k-1] - \mathbf{P}[Y_m > k] = p(1-p)^k$. ◊

Definition 2.29 (Convolution) Let μ and ν be probability measures on $(\mathbb{Z}, 2^{\mathbb{Z}})$. The *convolution* $\mu * \nu$ is defined as the probability measure on $(\mathbb{Z}, 2^{\mathbb{Z}})$ such that

$$(\mu * \nu)\big(\{n\}\big) = \sum_{m=-\infty}^{\infty} \mu\big(\{m\}\big)\nu\big(\{n-m\}\big).$$

We define the nth convolution power recursively by $\mu^{*1} = \mu$ and

$$\mu^{*(n+1)} = \mu^{*n} * \mu.$$

Remark 2.30 The convolution is a symmetric operation: $\mu * \nu = \nu * \mu$. ◊

Theorem 2.31 *If X and Y are independent $\mathbb{Z}$-valued random variables, then* $\mathbf{P}_{X+Y} = \mathbf{P}_X * \mathbf{P}_Y$.

Proof For any $n \in \mathbb{Z}$,

$$\begin{aligned}\mathbf{P}_{X+Y}\big(\{n\}\big) &= \mathbf{P}[X+Y=n] \\ &= \mathbf{P}\Bigg[\biguplus_{m\in\mathbb{Z}}\big(\{X=m\}\cap\{Y=n-m\}\big)\Bigg] \\ &= \sum_{m\in\mathbb{Z}} \mathbf{P}\big[\{X=m\}\cap\{Y=n-m\}\big] \\ &= \sum_{m\in\mathbb{Z}} \mathbf{P}_X\big[\{m\}\big]\mathbf{P}_Y\big[\{n-m\}\big] = (\mathbf{P}_X * \mathbf{P}_Y)\big[\{n\}\big].\end{aligned}$$

□

Owing to the last theorem, it is natural to define the convolution of two probability measures on $\mathbb{R}^n$ (or more generally on an Abelian group) as the distribution of the sum of two independent random variables with the corresponding distributions. Later we will encounter a different (but equivalent) definition that will, however, rely on the notion of an integral that is not yet available to us at this point (see Definition 14.17).

Definition 2.32 (Convolution of measures) Let μ and ν be probability measures on $\mathbb{R}^n$ and let X and Y be independent random variables with $\mathbf{P}_X = \mu$ and $\mathbf{P}_Y = \nu$. We define the *convolution* of μ and ν as $\mu * \nu = \mathbf{P}_{X+Y}$.

Recursively, we define the convolution powers μ^{*k} for all $k \in \mathbb{N}$ and let $\mu^{*0} = \delta_0$.

Example 2.33 Let X and Y be independent Poisson random variables with parameters μ and $\lambda \geq 0$. Then

$$\begin{aligned}\mathbf{P}[X+Y=n] &= e^{-\mu}e^{-\lambda}\sum_{m=0}^{n}\frac{\mu^m}{m!}\frac{\lambda^{n-m}}{(n-m)!} \\ &= e^{-(\mu+\lambda)}\frac{1}{n!}\sum_{m=0}^{n}\binom{n}{m}\mu^m\lambda^{n-m} = e^{-(\mu+\lambda)}\frac{(\mu+\lambda)^n}{n!}.\end{aligned}$$

Hence $\mathrm{Poi}_\mu * \mathrm{Poi}_\lambda = \mathrm{Poi}_{\mu+\lambda}$. ◊

Exercise 2.2.1 Let X and Y be independent random variables with $X \sim \exp_\theta$ and $Y \sim \exp_\rho$ for certain $\theta, \rho > 0$. Show that

$$\mathbf{P}[X < Y] = \frac{\theta}{\theta + \rho}.$$

Exercise 2.2.2 (Box–Muller method) Let U and V be independent random variables that are uniformly distributed on [0, 1]. Define

$$X := \sqrt{-2\log(U)}\cos(2\pi V) \quad \text{and} \quad Y := \sqrt{-2\log(U)}\sin(2\pi V).$$

Show that X and Y are independent and $\mathcal{N}_{0,1}$-distributed.

Hint: First compute the distribution of $\sqrt{-2\log(U)}$ and then use the transformation formula (Theorem 1.101) as well as polar coordinates.

Exercise 2.2.3 (Multinomial distribution) Let $m \in \mathbb{N}$ and let $p = (p_1, \ldots, p_m)$ be a probability vector on $\{1, \ldots, m\}$. Let $X_1, \ldots, X_n$ be independent random variables with values in $1, \ldots, m$ and distribution p. We define an $\mathbb{N}_0^m$-valued random variable $Y = (Y_1, \ldots, Y_m)$ by

$$Y_i := \#\{k = 1, \ldots, n : X_k = i\} \quad \text{for } i = 1, \ldots, m.$$

Show that for $k = (k_1, \ldots, k_m) \in \mathbb{N}_0^m$ with $k_1 + \ldots + k_m = n$, we have

$$\mathbf{P}[Y = k] = \mathrm{Mul}_{n,p}\big(\{k\}\big) := \binom{n}{k} p^k. \tag{2.10}$$

Here

$$\binom{n}{k} = \binom{n}{k_1, \ldots, k_m} = \frac{n!}{k_1! \ldots k_m!}$$

is the *multinomial coefficient* and $p^k = p_1^{k_1} \ldots p_m^{k_m}$. The distribution $\mathrm{Mul}_{n,p}$ on $\mathbb{N}_0^m$ is called multinomial distribution with parameters n and p.

2.3 Kolmogorov's 0–1 Law

With the Borel–Cantelli lemma, we have seen a first 0–1 law for independent events. We now come to another 0–1 law for independent events and for independent σ-algebras. To this end, we first introduce the notion of the tail σ-algebra.

Definition 2.34 (Tail σ-algebra) Let I be a countably infinite index set and let $(\mathcal{A}_i)_{i\in I}$ be a family of σ-algebras. Then

$$\mathcal{T}\big((\mathcal{A}_i)_{i\in I}\big) := \bigcap_{\substack{J\subset I\\ \#J<\infty}} \sigma\bigg(\bigcup_{j\in I\setminus J} \mathcal{A}_j\bigg)$$

is called the *tail σ-algebra* of $(\mathcal{A}_i)_{i\in I}$. If $(A_i)_{i\in I}$ is a family of events, then we define

$$\mathcal{T}\big((A_i)_{i\in I}\big) := \mathcal{T}\big(\big(\{\emptyset, A_i, A_i^c, \Omega\}\big)_{i\in I}\big).$$

If $(X_i)_{i\in I}$ is a family of random variables, then we define $\mathcal{T}((X_i)_{i\in I}) := \mathcal{T}((\sigma(X_i))_{i\in I})$.

The tail σ-algebra contains those events A whose occurrence is independent of any fixed finite subfamily of the X_i. To put it differently, for any finite subfamily of the X_i, we can change the values of the X_i arbitrarily without changing whether A occurs or not.

Theorem 2.35 *Let $J_1, J_2, \ldots$ be finite sets with $J_n \uparrow I$. Then*

$$\mathcal{T}\big((\mathcal{A}_i)_{i\in I}\big) = \bigcap_{n=1}^{\infty} \sigma\bigg(\bigcup_{m\in I\setminus J_n} \mathcal{A}_m\bigg).$$

In the particular case $I=\mathbb{N}$, this reads $\mathcal{T}((\mathcal{A}_n)_{n\in\mathbb{N}}) = \bigcap_{n=1}^{\infty} \sigma(\bigcup_{m=n}^{\infty} \mathcal{A}_m)$.

The name "tail σ-algebra" is due to the interpretation of $I=\mathbb{N}$ as a set of times. As is made clear in the theorem, any event in $\mathcal{T}$ does not depend on the first finitely many time points.

Proof "$\subset$" This is clear.

"$\supset$" Let $J_n \subset I$, $n\in\mathbb{N}$, be finite sets with $J_n \uparrow I$. Let $J\subset I$ be finite. Then there exists an $N\in\mathbb{N}$ with $J\subset J_N$ and

$$\begin{aligned}\bigcap_{n=1}^{\infty} \sigma\Big(\bigcup\nolimits_{m\in I\setminus J_n} \mathcal{A}_m\Big) &\subset \bigcap_{n=1}^{N} \sigma\Big(\bigcup\nolimits_{m\in I\setminus J_n} \mathcal{A}_m\Big)\\ &= \sigma\Big(\bigcup\nolimits_{m\in I\setminus J_N} \mathcal{A}_m\Big) \subset \sigma\Big(\bigcup\nolimits_{m\in I\setminus J} \mathcal{A}_m\Big).\end{aligned}$$

The left-hand side does not depend on J. Hence we can form the intersection over all finite J and obtain

$$\bigcap_{n=1}^{\infty} \sigma\Big(\bigcup\nolimits_{m\in I\setminus J_n} \mathcal{A}_m\Big) \subset \mathcal{T}\big((\mathcal{A}_i)_{i\in I}\big).$$

□

Maybe at first glance it is not evident that there are any interesting events in the tail σ-algebra at all. It might not even be clear that we do not have $\mathcal{T} = \{\emptyset, \Omega\}$. Hence we now present simple examples of tail events and tail σ-algebra measurable random variables. In Section 2.4, we will study a more complex example.

Example 2.36

(i) Let $A_1, A_2, \ldots$ be events. Then the events $A_* := \liminf_{n\to\infty} A_n$ and $A^* := \limsup_{n\to\infty} A_n$ are in $\mathcal{T}((A_n)_{n\in\mathbb{N}})$. Indeed, if we define $B_n := \bigcap_{m=n}^{\infty} A_m$ for $n \in \mathbb{N}$, then $B_n \uparrow A_*$ and $B_n \in \sigma((A_m)_{m\geq N})$ for any $n \geq N$. Thus $A_* \in \sigma((A_m)_{m\geq N})$ for any $N \in \mathbb{N}$ and hence $A_* \in \mathcal{T}((A_n)_{n\in\mathbb{N}})$. The case A^* is similar.

(ii) Let $(X_n)_{n\in\mathbb{N}}$ be a family of $\overline{\mathbb{R}}$-valued random variables. Then the maps $X_* := \liminf_{n\to\infty} X_n$ and $X^* := \limsup_{n\to\infty} X_n$ are $\mathcal{T}((X_n)_{n\in\mathbb{N}})$-measurable. Indeed, if we define $Y_n := \sup_{m\geq n} X_m$, then for any $N \in \mathbb{N}$, the random variable $X^* = \inf_{n\geq 1} Y_n = \inf_{n\geq N} Y_n$ is $\mathcal{T}_N := \sigma(X_n, n \geq N)$-measurable and hence also measurable with respect to $\mathcal{T}((X_n)_{n\in\mathbb{N}}) = \bigcap_{n=1}^{\infty} \mathcal{T}_n$.
The case X_* is similar.

(iii) Let $(X_n)_{n\in\mathbb{N}}$ be real random variables. Then the *Cesàro limits*

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$$

are $\mathcal{T}((X_n)_{n\in\mathbb{N}})$-measurable. In order to show this, choose $N \in \mathbb{N}$ and note that

$$X_* := \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=N}^{n} X_i$$

is $\sigma((X_n)_{n\geq N})$-measurable. Since this holds for any N, X_* is $\mathcal{T}((X_n)_{n\in\mathbb{N}})$-measurable. The case of the limes superior is similar. ◊

Theorem 2.37 (Kolmogorov's 0–1 Law) *Let I be a countably infinite index set and let $(\mathcal{A}_i)_{i\in I}$ be an independent family of σ-algebras. Then the tail σ-algebra is* **P**-*trivial, that is,*

$$\mathbf{P}[A] \in \{0, 1\} \quad \textit{for any } A \in \mathcal{T}\big((\mathcal{A}_i)_{i\in I}\big).$$

Proof It is enough to consider the case $I = \mathbb{N}$. For $n \in \mathbb{N}$, let

$$\mathcal{F}_n := \left\{\bigcap_{k=1}^{n} A_k : A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n\right\}.$$

Then $\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n$ is a semiring and $\sigma(\mathcal{F}) = \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n)$. Indeed, for any $n \in \mathbb{N}$ and $A_n \in \mathcal{A}_n$, we have $A_n \in \mathcal{F}$; hence $\sigma(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n) \subset \sigma(\mathcal{F})$. On the other hand, we have $\mathcal{F}_m \subset \sigma(\bigcup_{n=1}^{m} \mathcal{A}_n) \subset \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n)$ for any $m \in \mathbb{N}$; hence $\mathcal{F} \subset \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n)$.

Let $A \in \mathcal{T}((\mathcal{A}_n)_{n\in\mathbb{N}})$ and $\varepsilon > 0$. By the approximation theorem for measures (Theorem 1.65), there exists an $N \in \mathbb{N}$ and mutually disjoint sets $F_1, \ldots, F_N \in \mathcal{F}$ such that $\mathbf{P}[A \triangle (F_1 \cup \ldots \cup F_N)] < \varepsilon$. Clearly, there is an $n \in \mathbb{N}$ such that $F_1, \ldots, F_N \in \mathcal{F}_n$ and thus $F := F_1 \cup \ldots \cup F_N \in \sigma(\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_n)$. Obviously, $A \in \sigma(\bigcup_{m=n+1}^{\infty} \mathcal{A}_m)$; hence A is independent of F. Thus

$$\varepsilon > \mathbf{P}[A \setminus F] = \mathbf{P}\big[A \cap (\Omega \setminus F)\big] = \mathbf{P}[A]\big(1 - \mathbf{P}[F]\big) \geq \mathbf{P}[A]\big(1 - \mathbf{P}[A] - \varepsilon\big).$$

Letting $\varepsilon \downarrow 0$ yields $0 = \mathbf{P}[A](1 - \mathbf{P}[A])$. □

Corollary 2.38 *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of independent events. Then*

$$\mathbf{P}\Big[\limsup_{n\to\infty} A_n\Big] \in \{0, 1\} \quad \textit{and} \quad \mathbf{P}\Big[\liminf_{n\to\infty} A_n\Big] \in \{0, 1\}.$$

Proof Essentially this is a simple conclusion of the Borel–Cantelli lemma. However, the statement can also be deduced from Kolmogorov's 0–1 law as limes superior and limes inferior are in the tail σ-algebra. □

Corollary 2.39 *Let $(X_n)_{n\in\mathbb{N}}$ be an independent family of $\overline{\mathbb{R}}$-valued random variables. Then $X_* := \liminf_{n\to\infty} X_n$ and $X^* := \limsup_{n\to\infty} X_n$ are almost surely constant. That is, there exist $x_*, x^* \in \overline{\mathbb{R}}$ such that $\mathbf{P}[X_* = x_*] = 1$ and $\mathbf{P}[X^* = x^*] = 1$.*

If all X_i are real-valued, then the Cesàro limits

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \quad \textit{and} \quad \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i$$

are also almost surely constant.

Proof Let $X_* := \liminf_{n\to\infty} X_n$. For any $x \in \overline{\mathbb{R}}$, we have $\{X_* \leq x\} \in \mathcal{T}((X_n)_{n\in\mathbb{N}})$; hence $\mathbf{P}[X_* \leq x] \in \{0, 1\}$. Define

$$x_* := \inf\big\{x \in \mathbb{R} : \mathbf{P}[X_* \leq x] = 1\big\} \in \overline{\mathbb{R}}.$$

If $x_* = \infty$, then evidently

$$\mathbf{P}[X_* < \infty] = \lim_{n\to\infty} \mathbf{P}[X_* \leq n] = 0.$$

If $x_* \in \mathbb{R}$, then

$$\mathbf{P}[X_* \leq x_*] = \lim_{n\to\infty} \mathbf{P}\bigg[X_* \leq x_* + \frac{1}{n}\bigg] = 1$$

and

$$\mathbf{P}[X_* < x_*] = \lim_{n\to\infty} \mathbf{P}\bigg[X_* \leq x_* - \frac{1}{n}\bigg] = 0.$$

If $x_* = -\infty$, then

$$\mathbf{P}[X_* > -\infty] = \lim_{n\to\infty} \mathbf{P}[X_* > -n] = 0.$$

The cases of the limes superior and the Cesàro limits are similar. □

Exercise 2.3.1 Let $(X_n)_{n\in\mathbb{N}}$ be an independent family of $\text{Rad}_{1/2}$ random variables (i.e., $\mathbf{P}[X_n = -1] = \mathbf{P}[X_n = +1] = \frac{1}{2}$) and let $S_n = X_1 + \ldots + X_n$ for any $n \in \mathbb{N}$. Show that $\limsup_{n\to\infty} S_n = \infty$ almost surely.

2.4 Example: Percolation

Consider the d-dimensional integer lattice $\mathbb{Z}^d$, where any point is connected to any of its $2d$ nearest neighbors by an edge. If $x, y \in \mathbb{Z}^d$ are nearest neighbors (that is, $\|x - y\|_2 = 1$), then we denote by $e = \langle x, y\rangle = \langle y, x\rangle$ the edge that connects x and y. Formally, the set of edges is a subset of the set of subsets of $\mathbb{Z}^d$ with two elements:

$$E = \Big\{\{x, y\} : x, y \in \mathbb{Z}^d \text{ with } \|x - y\|_2 = 1\Big\}.$$

Somewhat more generally, an undirected *graph* G is a pair $G = (V, E)$, where V is a set (the set of "vertices" or nodes) and $E \subset \{\{x, y\} : x, y \in V, x \neq y\}$ is a subset of the set of subsets of V of cardinality two (the set of *edges* or *bonds*).

Our intuitive understanding of an edge is a connection between two points x and y and not an (unordered) pair $\{x, y\}$. To stress this notion of a connection, we use a different symbol from the set brackets. That is, we denote the edge that connects x and y by $\langle x, y\rangle = \langle y, x\rangle$ instead of $\{x, y\}$.

Our graph (V, E) is the starting point for a stochastic model of a porous medium. We interpret the edges as tubes along which water can flow. However, we want the medium not to have a homogeneous structure, such as $\mathbb{Z}^d$, but an amorphous structure. In order to model this, we randomly destroy a certain fraction $1 - p$ of the tubes (with $p \in [0, 1]$ a parameter) and keep the others. Water can flow only through the remaining tubes. The destroyed tubes will be called "closed", the others "open". The fundamental question is: For which values of p is there a connected infinite system of tubes along which water can flow? The physical interpretation is that if we throw a block of the considered material into a bathtub, then the block will soak up water; that is, it will be wetted inside. If there is no infinite open component, then the water may wet only a thin layer at the surface. See Fig. 2.1 for a computer simulation of the percolation model.

We now come to a formal description of the model. Choose a parameter $p \in [0, 1]$ and an independent family of identically distributed random variables $(X_e^p)_{e\in E}$ with $X_e^p \sim \text{Ber}_p$; that is, $\mathbf{P}[X_e^p = 1] = 1 - \mathbf{P}[X_e^p = 0] = p$ for any $e \in E$. We define the set of *open* edges as

$$E^p := \Big\{e \in E : X_e^p = 1\Big\}. \tag{2.11}$$

Consequently, the edges in $E \setminus E^p$ are called *closed*. Hence we have constructed a (random) subgraph $(\mathbb{Z}^d, E^p)$ of $(\mathbb{Z}^d, E)$. We call $(\mathbb{Z}^d, E^p)$ a percolation model (more precisely, a model for *bond percolation*, in contrast to *site percolation*, where vertices can be open or closed). An (open) path (of length n) in this subgraph is a sequence $\pi = (x_0, x_1, \ldots, x_n)$ of points in $\mathbb{Z}^d$ with $\langle x_{i-1}, x_i\rangle \in E^p$ for all $i = 1, \ldots, n$. We say that two points $x, y \in \mathbb{Z}^d$ are connected by an open path if there is an $n \in \mathbb{N}$ and an open path $(x_0, x_1, \ldots, x_n)$ with $x_0 = x$ and $x_n = y$. In this case, we write $x \longleftrightarrow_p y$. Note that "$\longleftrightarrow_p$" is an equivalence relation; however, a random one, as it depends on the values of the random variables $(X_e^p)_{e\in E}$. For every $x \in \mathbb{Z}^d$, we define the (random) open cluster of x; that is, the connected component of x in the graph $(\mathbb{Z}^d, E^p)$:

$$C^p(x) := \big\{y \in \mathbb{Z}^d : x \longleftrightarrow_p y\big\}. \tag{2.12}$$

Lemma 2.40 *Let $x, y \in \mathbb{Z}^d$. Then $\mathbb{1}_{\{x\longleftrightarrow_p y\}}$ is a random variable. In particular, $\#C^p(x)$ is a random variable for any $x \in \mathbb{Z}^d$.*

Proof We may assume $x = 0$. Let $f_{y,n} = 1$ if there exists an open path of length at most n that connects 0 to y, and $f_{y,n} = 0$ otherwise. Clearly, $f_{y,n} \uparrow \mathbb{1}_{\{0\longleftrightarrow_p y\}}$ for $n \to \infty$; hence it suffices to show that each $f_{y,n}$ is measurable. Let $B_n := \{-n, -n+1, \ldots, n-1, n\}^d$ and $E_n := \{e \in E : e \cap B_n \neq \emptyset\}$. Then $Y_n := (X_e^p : e \in E_n) : \Omega \to \{0,1\}^{E_n}$ is measurable (with respect to $2^{(\{0,1\}^{E_n})}$) by Theorem 1.90. However, $f_{y,n}$ is a function of Y_n, say $f_{y,n} = g_{y,n} \circ Y_n$ for some map $g_{y,n} : \{0,1\}^{E_n} \to \{0,1\}$. By the composition theorem for maps (Theorem 1.80), $f_{y,n}$ is measurable.

The additional statement holds since $\#C^p(x) = \sum_{y\in\mathbb{Z}^d} \mathbb{1}_{\{x\longleftrightarrow_p y\}}$. □

Definition 2.41 We say that percolation occurs if there exists an infinitely large open cluster. We call

$$\begin{aligned}\psi(p) &:= \mathbf{P}[\text{there exists an infinite open cluster}]\\ &= \mathbf{P}\Bigg[\bigcup_{x\in\mathbb{Z}^d} \big\{\#C^p(x) = \infty\big\}\Bigg]\end{aligned}$$

the probability of percolation. We define

$$\theta(p) := \mathbf{P}\big[\#C^p(0) = \infty\big]$$

as the probability that the origin is in an infinite open cluster.

By the translation invariance of the lattice, we have

$$\theta(p) = \mathbf{P}\big[\#C^p(y) = \infty\big] \quad \text{for any } y \in \mathbb{Z}^d. \tag{2.13}$$

The fundamental question is: How large are $\theta(p)$ and $\psi(p)$ depending on p? We make the following simple observation.

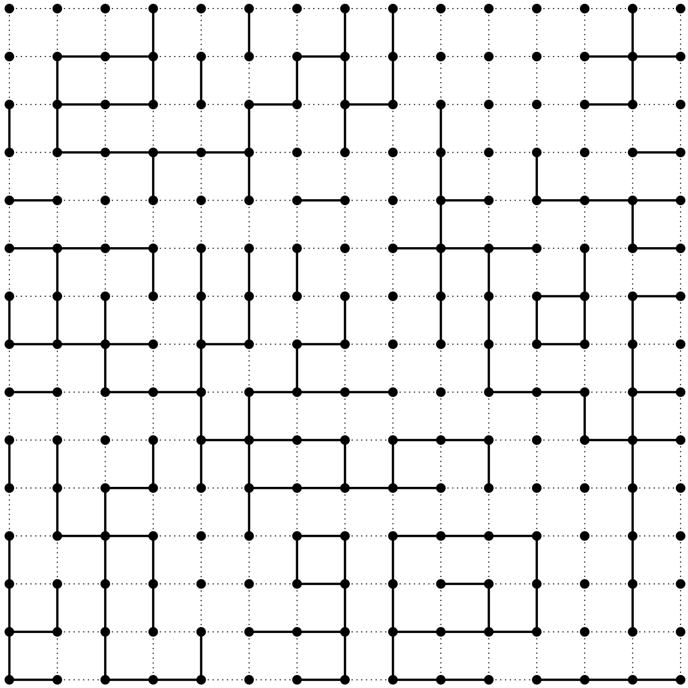

Fig. 2.1 Percolation on a 15×15 grid, $p = 0.42$

Theorem 2.42 *The map* $[0, 1] \to [0, 1]$, $p \mapsto \theta(p)$ *is monotone increasing.*

Proof Although the statement is intuitively so clear that it might not need a proof, we give a formal proof in order to introduce a technique called *coupling*. Let $p, p' \in [0, 1]$ with $p < p'$. Let $(Y_e)_{e \in E}$ be an independent family of random variables with $\mathbf{P}[Y_e \leq q] = q$ for any $e \in E$ and $q \in \{p, p', 1\}$. At this point, we could, for example, assume that $Y_e \sim \mathcal{U}_{[0,1]}$ is uniformly distributed on $[0, 1]$. Since we have not yet shown the existence of an independent family with this distribution, we content ourselves with Y_e that assume only three values $\{p, p', 1\}$. Hence

$$\mathbf{P}[Y_e = q] = \begin{cases} p, & \text{if } q = p, \\ p' - p, & \text{if } q = p', \\ 1 - p', & \text{if } q = 1. \end{cases}$$

Such a family $(Y_e)_{e \in E}$ exists by Theorem 2.19. For $q \in \{p, p'\}$ and $e \in E$, we define

$$X_e^q := \begin{cases} 1, & \text{if } Y_e \leq q, \\ 0, & \text{else.} \end{cases}$$

Clearly, for any $q \in \{p, p'\}$, the family $(X_e^q)_{e \in E}$ is independent (see Remark 2.15(iii)) and $X_e^q \sim \mathrm{Ber}_q$. Furthermore, $X_e^p \leq X_e^{p'}$ for any $e \in E$. The procedure of defining two families of random variables that are related in a specific way (here "$\leq$") on one probability space is called a *coupling*.

Clearly, $C^p(x) \subset C^{p'}(x)$ for any $x \in \mathbb{Z}^d$; hence $\theta(p) \leq \theta(p')$. $\square$

With the aid of Kolmogorov's 0–1 law, we can infer the following theorem.

Theorem 2.43 *For any* $p \in [0,1]$, *we have*

$$\psi(p) = \begin{cases} 0, & \text{if } \theta(p) = 0, \\ 1, & \text{if } \theta(p) > 0. \end{cases}$$

Proof If $\theta(p) = 0$, then by (2.13)

$$\psi(p) \leq \sum_{y \in \mathbb{Z}^d} \mathbf{P}\big[\#C^p(y) = \infty\big] = \sum_{y \in \mathbb{Z}^d} \theta(p) = 0.$$

Now let $A = \bigcup_{y \in \mathbb{Z}^d} \{\#C^p(y) = \infty\}$. Clearly, A remains unchanged if we change the state of finitely many edges. That is, $A \in \sigma((X_e^p)_{e \in E \setminus F})$ for every finite $F \subset E$. Hence A is in the tail σ-algebra $\mathcal{T}((X_e^p)_{e \in E})$ by Theorem 2.35. Kolmogorov's 0–1 law (Theorem 2.37) implies that $\psi(p) = \mathbf{P}[A] \in \{0, 1\}$. If $\theta(p) > 0$, then $\psi(p) \geq \theta(p)$ implies $\psi(p) = 1$. $\square$

Due to the monotonicity, we can make the following definition.

Definition 2.44 The critical value p_c for percolation is defined as

$$\begin{aligned} p_c &= \inf\big\{p \in [0,1] : \theta(p) > 0\big\} = \sup\big\{p \in [0,1] : \theta(p) = 0\big\} \\ &= \inf\big\{p \in [0,1] : \psi(p) = 1\big\} = \sup\big\{p \in [0,1] : \psi(p) = 0\big\}. \end{aligned}$$

We come to the main theorem of this section.

Theorem 2.45 *For* $d = 1$, *we have* $p_c = 1$. *For* $d \geq 2$, *we have* $p_c(d) \in [\frac{1}{2d-1}, \frac{2}{3}]$.

Proof First consider $d = 1$ and $p < 1$. Let $A^- := \{X^p_{\langle n,n+1\rangle} = 0 \text{ for some } n < 0\}$ and $A^+ := \{X^p_{\langle n,n+1\rangle} = 0 \text{ for some } n > 0\}$. Let $A = A^- \cap A^+$. By the Borel–Cantelli lemma, we get $\mathbf{P}[A^-] = \mathbf{P}[A^+] = 1$. Hence $\theta(p) = \mathbf{P}[A^c] = 0$.

Now assume $d \geq 2$.

Lower bound. First we show $p_c \geq \frac{1}{2d-1}$. Clearly, for any $n \in \mathbb{N}$,

$$\mathbf{P}\big[\#C^p(0) = \infty\big] \leq \mathbf{P}\big[\text{there is an } x \in C^p(0) \text{ with } \|x\|_\infty = n\big].$$

We want to estimate the probability that there exists a point $x \in C^p(0)$ with distance n from the origin. Any such point is connected to the origin by a path without self-intersections π that starts at 0 and has length $m \geq n$. Let $\Pi_{0,m}$ be the set of such paths. Clearly, $\#\Pi_{0,m} \leq 2d \cdot (2d-1)^{m-1}$ since there are $2d$ choices for the first step and at most $2d - 1$ choices for any further step. For any $\pi \in \Pi_{0,m}$, the probability that π uses only open edges is

$$\mathbf{P}[\pi \text{ is open}] = p^m.$$

Hence, for $p < \frac{1}{2d-1}$,

$$\begin{aligned}\theta(p) &\le \sum_{m=n}^{\infty} \sum_{\pi \in \Pi_{0,m}} \mathbf{P}[\pi \text{ is open}] \\ &\le \frac{2d}{2d-1} \sum_{m=n}^{\infty} \big((2d-1)p\big)^m \\ &= \frac{2d}{(2d-1)(1-(2d-1)p)} \big((2d-1)p\big)^n \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

We conclude that $p_c \ge \frac{1}{2d-1}$.

Upper bound. We can consider $\mathbb{Z}^d$ as a subset of $\mathbb{Z}^d \times \{0\} \subset \mathbb{Z}^{d+1}$. Hence, if percolation occurs for p in $\mathbb{Z}^d$, then it also occurs for p in $\mathbb{Z}^{d+1}$. Hence the corresponding critical values are ordered $p_c(d+1) \le p_c(d)$.

Thus, it is enough to consider the case $d = 2$. Here we show $p_c \le \frac{2}{3}$ by using a contour argument due to Peierls [127], originally designed for the Ising model of a ferromagnet, see Example 18.16 and (18.9).

For $N \in \mathbb{N}$, we define (compare (2.12) with $x = (i, 0)$)

$$C_N := \bigcup_{i=0}^{N} C^p\big((i,0)\big)$$

as the set of points that are connected (along open edges) to at least one of the points in $\{0, \dots, N\} \times \{0\}$. Due to the subadditivity of probability (and since $\mathbf{P}[\#C^p((i,0)) = \infty] = \theta(p)$ for any $i \in \mathbb{Z}$), we have

$$\theta(p) = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{P}\big[\#C^p\big((i,0)\big) = \infty\big] \ge \frac{1}{N+1} \mathbf{P}[\#C_N = \infty].$$

Now consider those closed contours in the dual graph $(\tilde{\mathbb{Z}}^2, \tilde{E})$ that surrounds C_N if $\#C_N < \infty$. Here the dual graph is defined by

$$\begin{aligned}\tilde{\mathbb{Z}}^2 &= \left(\frac{1}{2}, \frac{1}{2}\right) + \mathbb{Z}^2, \\ \tilde{E} &= \big\{\{x, y\} : x, y \in \tilde{\mathbb{Z}}^2, \|x - y\|_2 = 1\big\}.\end{aligned}$$

An edge $\tilde{e}$ in the dual graph $(\tilde{\mathbb{Z}}^2, \tilde{E})$ crosses exactly one edge e in $(\mathbb{Z}^2, E)$. We call $\tilde{e}$ open if e is open and closed otherwise. A circle γ is a self-intersection free

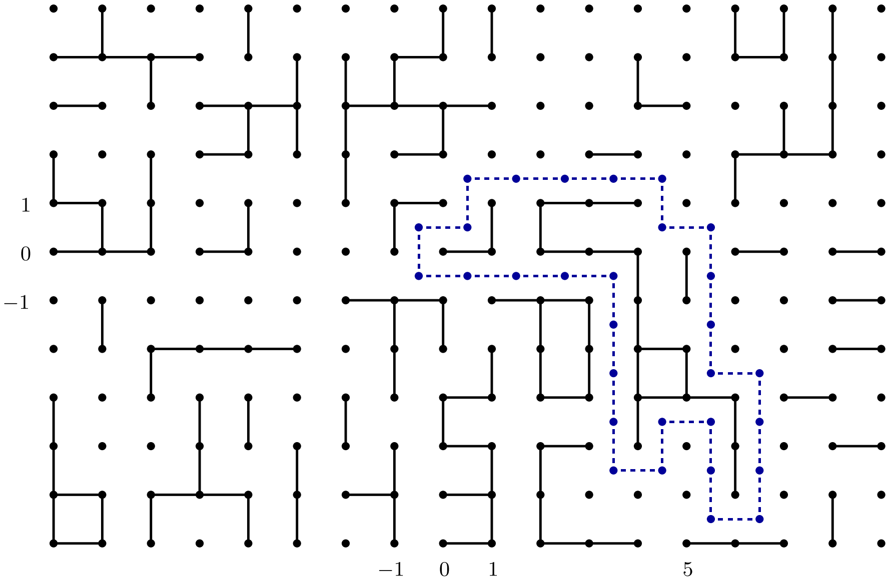

Fig. 2.2 Contour of the cluster C_5

path in $(\tilde{\mathbb{Z}}^2, \tilde{E})$ that starts and ends at the same point. A contour of the set C_N is a minimal circle that surrounds C_N. Minimal means that the enclosed area is minimal (see Fig. 2.2). For $n \geq 2N$, let

$$\Gamma_n = \big\{\gamma : \gamma \text{ is a circle of length } n \text{ that surrounds } \{0, \ldots, N\} \times \{0\}\big\}.$$

We want to deduce an upper bound for $\#\Gamma_n$. Let $\gamma \in \Gamma_n$ and fix one point of γ. For definiteness, choose the upper point $(m + \frac{1}{2}, \frac{1}{2})$ of the rightmost edge of γ that crosses the horizontal axis (in Fig. 2.2 this is the point $(5 + \frac{1}{2}, \frac{1}{2})$). Clearly, $m \geq N$ and $m \leq n$ since γ surrounds the origin. Starting from $(m + \frac{1}{2}, \frac{1}{2})$, for any further edge of γ, there are at most three possibilities. Hence

$$\#\Gamma_n \leq n \cdot 3^n.$$

We say that γ is closed if it uses only closed edges (in $\tilde{E}$). A contour of C_N is automatically closed and has a length of at least $2N$. Hence for $p > \frac{2}{3}$

$$\begin{aligned}\mathbf{P}[\#C_N < \infty] &= \sum_{n=2N}^{\infty} \mathbf{P}[\text{there is a closed circle } \gamma \in \Gamma_n] \\ &\leq \sum_{n=2N}^{\infty} n \cdot \big(3(1-p)\big)^n \stackrel{N\to\infty}{\longrightarrow} 0.\end{aligned}$$

We conclude $p_c \leq \frac{2}{3}$. □

In general, the value of p_c is not known and is extremely hard to determine. In the case of bond percolation on $\mathbb{Z}^2$, however, the exact value of p_c can be determined due to the self-duality of the planar graph $(\mathbb{Z}^2, E)$. (If $G = (V, E)$ is a planar graph; that is, a graph that can be embedded into $\mathbb{R}^2$ without self-intersections, then the vertex set of the dual graph is the set of faces of G. Two such vertices are connected by exactly one edge; that is, by the edge in E that separates the two faces. Evidently, the two-dimensional integer lattice is isomorphic to its dual graph. Note that the contour in Fig. 2.2 can be considered as a closed path in the dual graph.) We cite a theorem of Kesten [95].

Theorem 2.46 (Kesten 1980) *For bond percolation in* $\mathbb{Z}^2$, *the critical value is* $p_c = \frac{1}{2}$ *and* $\theta(p_c) = 0$.

Proof See, for example, the book of Grimmett [63, pp. 287ff]. □

It is conjectured that $\theta(p_c) = 0$ holds in any dimension $d \geq 2$. However, rigorous proofs are known only for $d = 2$ and $d \geq 19$ (see [67]).

*Uniqueness of the Infinite Open Cluster**

Fix a p such that $\theta(p) > 0$. We saw that with probability one there is *at least* one infinite open cluster. Now we want to show that there is *exactly* one.

Denote by $N \in \{0, 1, \ldots, \infty\}$ the (random) number of infinite open clusters.

Theorem 2.47 (Uniqueness of the infinite open cluster) *For any* $p \in [0, 1]$, *we have* $\mathbf{P}_p[N \leq 1] = 1$.

Proof This theorem was first proved by Aizenman, Kesten and Newman [2, 3]. Here we follow the proof of Burton and Keane [23] as described in [63, Section 8.2].

The cases $p = 1$ and $\theta(p) = 0$ (hence in particular the case $p = 0$) are trivial. Hence we assume now that $p \in (0, 1)$ and $\theta(p) > 0$.

Step 1. We first show that

$$\mathbf{P}_p[N = m] = 1 \quad \text{for some } m = 0, 1, \ldots, \infty. \tag{2.14}$$

We need a 0–1 law similar to that of Kolmogorov. However, N is not measurable with respect to the tail σ-algebra. Hence we have to find a more subtle argument. Let $u_1 = (1, 0, \ldots, 0)$ be the first unit vector in $\mathbb{Z}^d$. On the edge set E, define the translation $\tau : E \to E$ by $\tau(\langle x, y\rangle) = \langle x + u_1, y + u_1\rangle$. Let

$$E_0 := \big\{\big\langle (x_1, \ldots, x_d), (y_1, \ldots, y_d)\big\rangle \in E : x_1 = 0, y_1 \geq 0\big\}$$

be the set of all edges in $\mathbb{Z}^d$ that either connect two points from $\{0\} \times \mathbb{Z}^{d-1}$ or one point of $\{0\} \times \mathbb{Z}^{d-1}$ with one point of $\{1\} \times \mathbb{Z}^{d-1}$. Clearly, the sets $(\tau^n(E_0), n \in \mathbb{Z})$

are disjoint and $E = \biguplus_{n\in\mathbb{Z}} \tau^n(E_0)$. Hence the random variables $Y_n := (X^p_{\tau^n(e)})_{e\in E_0}$, $n \in \mathbb{Z}$, are independent and identically distributed (with values in $\{0, 1\}^{E_0}$). Define $Y = (Y_n)_{n\in\mathbb{Z}}$ and $\tau(Y) = (Y_{n+1})_{n\in\mathbb{Z}}$. Define $A_m \in \{0, 1\}^E$ by

$$\{Y \in A_m\} = \{N = m\}.$$

Clearly, the value of N does not change if we shift *all* edges simultaneously. That is, $\{Y \in A_m\} = \{\tau(Y) \in A_m\}$. An event with this property is called *invariant* or shift invariant. Using an argument similar to that in the proof of Kolmogorov's 0–1 law, one can show that invariant events (defined by i.i.d. random variables) have probability either 0 or 1 (see Example 20.26 for a proof).

Step 2. We will show that

$$\mathbf{P}_p[N = m] = 0 \quad \text{for any } m \in \mathbb{N} \setminus \{1\}. \tag{2.15}$$

Accordingly, let $m = 2, 3, \ldots$. We assume that $\mathbf{P}[N = m] = 1$ and show that this leads to a contradiction.

For $L \in \mathbb{N}$, let $B_L := \{-L, \ldots, L\}^d$ and denote by $E_L = \{e = \langle x, y\rangle \in E : x, y \in B_L\}$ the set of those edges with both vertices lying in B_L. For $i = 0, 1$, let $D^i_L := \{X^p_e = i \text{ for all } e \in E_L\}$. Let N^1_L be the number of infinite open clusters if we consider all edges e in E_L as open (independently of the value of X^p_e). Similarly define N^0_L where we consider all edges in E_L as closed. Since $\mathbf{P}_p[D^i_L] > 0$, and since $N = m$ almost surely, we have $N^i_L = m$ almost surely for $i = 0, 1$.

Let

$$A^2_L := \bigcup_{x^1,x^2\in B_L\setminus B_{L-1}} \big\{C^p(x^1) \cap C^p(x^2) = \emptyset\big\} \cap \big\{\#C^p(x^1) = \#C_p(x^2) = \infty\big\}$$

be the event where there exist two points on the boundary of B_L that lie in different infinite open clusters. Clearly, $A^2_L \uparrow \{N \geq 2\}$ for $L \to \infty$.

Define $A^2_{L,0}$ in a similarly way to A^2_L; however, we now consider all edges $e \in E_L$ as closed, irrespective of whether $X^p_e = 1$ or $X^p_e = 0$. If A^2_L occurs, then there are two points x^1, x^2 on the boundary of B_L such that for any $i = 1, 2$, there is an infinite self-intersection free open path π_{x^i} starting at x^i that avoids x^{3-i}. Hence $A^2_L \subset A^2_{L,0}$. Now choose L large enough for $\mathbf{P}[A^2_{L,0}] > 0$.

If $A^2_{L,0}$ occurs and if we open all edges in B_L, then at least two of the infinite open clusters get connected by edges in B_L. Hence the total number of infinite open clusters decreases by at least one. We infer $\mathbf{P}_p[N^1_L \leq N^0_L - 1] \geq \mathbf{P}_p[A^2_{L,0}] > 0$, which leads to a contradiction.

Step 3. In Step 2, we have shown already that N does not assume a *finite* value larger than 1. Hence it remains to show that almost surely N does not assume the value ∞. Indeed, we show that

$$\mathbf{P}_p[N \geq 3] = 0. \tag{2.16}$$

This part of the proof is the most difficult one. We assume that $\mathbf{P}_p[N \geq 3] > 0$ and show that this leads to a contradiction.

We say that a point $x \in \mathbb{Z}^d$ is a *trifurcation point* if

- x is in an infinite open cluster $C^p(x)$,
- there are exactly three open edges with endpoint x, and
- removing all of these three edges splits $C^p(x)$ into three mutually disjoint infinite open clusters.

By T we denote the set of trifurcation points, and let $T_L := T \cap B_L$. Let $r := \mathbf{P}_p[0 \in T]$. Due to translation invariance, we have $(\#B_L)^{-1}\mathbf{E}_p[\#T_L] = r$ for any L. (Here $\mathbf{E}_p[\#T_L]$ denotes the expected value of $\#T_L$, which we define formally in Chapter 5.) Let

$$A_L^3 := \bigcup_{x^1,x^2,x^3 \in B_L \setminus B_{L-1}} \left(\bigcap_{i \neq j} \{C^p(x^i) \cap C^p(x^j) = \emptyset\} \right) \cap \left(\bigcap_{i=1}^{3} \{\#C^p(x^i) = \infty\} \right)$$

be the event where there are three points on the boundary of B_L that lie in different infinite open clusters. Clearly, $A_L^3 \uparrow \{N \geq 3\}$ for $L \to \infty$.

As for $A_{L,0}^2$, we define $A_{L,0}^3$ as the event where there are three distinct points on the boundary of B_L that lie in different infinite open clusters if we consider all edges in E_L as closed. As above, we have $A_L^3 \subset A_{L,0}^3$.

For three distinct points $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$, let F_{x^1,x^2,x^3} be the event where for any $i = 1, 2, 3$, there exists an infinite self-intersection free open path π_{x^i} starting at x^i that uses only edges in $E^p \setminus E_L$ and that avoids the points x^j, $j \neq i$. Then

$$A_{L,0}^3 \subset \bigcup_{\substack{x^1,x^2,x^3 \in B_L \setminus B_{L-1} \\ \text{mutually distinct}}} F_{x^1,x^2,x^3}.$$

Let L be large enough for $\mathbf{P}_p[A_{L,0}^3] \geq \mathbf{P}_p[N \geq 3]/2 > 0$. Choose three pairwise distinct points $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$ with $\mathbf{P}_p[F_{x^1,x^2,x^3}] > 0$.

If F_{x^1,x^2,x^3} occurs, then we can find a point $y \in B_L$ that is the starting point of three mutually disjoint (not necessarily open) paths π_1, π_2 and π_3 that end at x^1, x^2 and x^3. Let G_{y,x^1,x^2,x^3} be the event where in E_L exactly those edges are open that belong to these three paths (that is, all other edges in E_L are closed). The events F_{x^1,x^2,x^3} and G_{y,x^1,x^2,x^3} are independent, and if both of them occur, then y is a trifurcation point. Hence

$$r = \mathbf{P}_p[y \in T] \geq \mathbf{P}_p[F_{x^1,x^2,x^3}] \cdot \big(p \wedge (1-p)\big)^{\#E_L} > 0.$$

Now we show that r must equal 0, which contradicts the assumption $\mathbf{P}_p[N \geq 3] > 0$. Let K_L be the set of all edges which have at least one endpoint in B_L. We consider two edges in K_L as equivalent if there exists a path in B_L

along open edges that does not hit any trifurcation point and which joins at least one endpoint of each of the two edges. We denote the equivalence relation by R and let $U_L = K_L/R$ be the set of equivalence classes. (Note that the three neighboring edges of a trifurcation point are in different equivalence classes.) We turn the set $H_L := U_L \cup T_L$ into a graph by considering two points $x \in T_L$ and $u \in U_L$ as neighbors if there exists an edge $k \in u$ which is incident to x. Note that each point $x \in T_L$ has exactly three neighbors which are in U_L. The points in U_L can be isolated (that is, without neighbors) or can be joined to arbitrarily many points in T_L but not in U_L.

A circle is a self-avoiding (finite) path that ends at its starting point. Note that the graph H_L has no circles. To show this assume there was a self-avoiding path $(h_0, h_1, \dots, h_n)$ starting and ending in some point $h_0 = h_n = x \in T_L$. Then $h_1, h_{n-1} \in U_L$ are distinct but connected in K^p even if we remove x. However, by the definition of the trifurcation point x, this is impossible. On the other hand, if there was a self-avoiding path $(g_0, \dots, g_m)$ starting and ending in some point $g_0 = g_m = u \in U_L$, then $(g_1, g_2, \dots, g_m, g_1)$ is a self-avoiding path starting and ending in $g_1 \in T_L$. However, we have just shown that such a path could not exist.

Write $\deg_{H_L}(h)$ for the degree of $h \in H_L$; that is, the number of neighbors of h in H_L. A point h with $\deg_{H_L}(h) = 1$ is called a *leaf* of H_L. Obviously, only points of U_L can be leaves. Let Z be a connected component of H_L that contains at least one point $x \in T_L$. Since Z is a tree (that is, it is connected and contains no circles), we have

$$\#Z - 1 = \frac{1}{2} \sum_{h \in Z} \deg_{H_L}(h).$$

Rearranging this formula yields an expression for the number of leaves:

$$\begin{aligned}\#\{u \in Z : \deg_{H_L}(u) = 1\} &= 2 + \sum_{h \in Z} \big(\deg_{H_L}(h) - 2\big)^+ \\ &\geq 2 + \#\{h \in Z : \deg_{H_L}(h) \geq 3\} \\ &= 2 + \#(Z \cap T_L).\end{aligned}$$

Summing over the connected components Z of H_L with at least one point in T_L, we obtain

$$\#\{u \in H_L : \deg_{H_L}(u) = 1\} \geq \#T_L.$$

Observe that any leaf $u \in H_L$ contains an edge that is incident to a point $x \in T_L$. Hence the edges of u lie in an infinite open cluster of K^p and there is at least one edge $k \in u$ incident to a point at the boundary $B_L \setminus B_{L-1}$ of B_L. For distinct leaves these are distinct points since the leaves belong to *disjoint* open clusters. Hence we get the bound

$$\#T_L \leq \#(B_L \setminus B_{L-1})$$

and thus

$$\frac{\#T_L}{\#B_L} \leq \frac{\#(B_L \setminus B_{L-1})}{\#B_L} \leq \frac{d}{L} \overset{L\to\infty}{\longrightarrow} 0.$$

Now $r = (\#B_L)^{-1}\mathbf{E}_p[\#T_L] \leq d/L$ implies $r = 0$. (Note that in the argument we used the notion of the expected value $\mathbf{E}_p[\#T_L]$ that will be formally introduced only in Chapter 5.) □