4 Order statistics

Given a sample, its kth order statistic is defined as the kth smallest value of the sample. Order statistics have lots of important applications: indeed, record values (historical maxima or minima) are used in sport, economics (especially insurance) and finance; they help to determine strength of some materials (the weakest link) as well as to model and study extreme climate events (a low river level can lead to drought, while a high level brings flood risk); order statistics are also useful in allocation of prize money in tournaments etc.

☞ In this section we study properties of individual order statistics, that of gaps between consecutive order statistics, and of the range of the sample. We explore properties of extreme statistics of large samples from various distributions, including uniform and exponential. You might find [1, Section 4] useful.

4.1 Definition and extreme order variables

If $\{X_1, X_2, \dots, X_n\}$ is a (random) n-sample from a particular distribution, it is often important to know the values of the largest observation, the smallest observation or the centermost observation (the median). More generally, the variable [1]

$$X_{(k)} \stackrel{\mathsf{def}}{=} \text{“the kth smallest of } X_1,\, X_2,\, \dots,\, X_n\text{”}\,, \qquad k = 1, 2, \dots, n\,, \tag{4.1}$$

is called the kth order variable and the (ordered) collection of observed values $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ is called the order statistic.

In general we have $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$, ie., some of the order values can coincide. If the common distribution of the n-sample $\{X_1, X_2, \dots, X_n\}$ is continuous (having a non-negative density), the probability of such a coincidence vanishes. As in the following we will only consider samples from continuous distributions, we can (and will) assume that all of the observed values are distinct, ie., the following strict inequalities hold:

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}\,. \tag{4.2}$$

It is straightforward to describe the distributions of the extreme order variables $X_{(1)}$ and $X_{(n)}$. Let $F_X(\cdot)$ be the common cdf [2] of the X variables; by independence, we have

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathsf{P}\big(X_{(n)} \le x\big) = \prod_{k=1}^{n} \mathsf{P}(X_k \le x) = \big(F_X(x)\big)^n\,, \\ F_{X_{(1)}}(x) &= \mathsf{P}(X_{(1)} \le x) = 1 - \prod_{k=1}^{n} \mathsf{P}(X_k > x) = 1 - \big(1 - F_X(x)\big)^n\,. \end{aligned} \tag{4.3}$$

[1] it is important to always remember that the distribution of the kth order variable depends on the total number n of observations; in the literature one sometimes specifies this dependence explicitly and uses $X_{k:n}$ to denote the kth order variable.

[2] cumulative distribution function; in what follows, if the reference distribution is clear we will write $F(x)$ instead of $F_X(x)$.

In particular, if X_k have a common density $f_X(x)$, then

$$f_{X_{(1)}}(x) = n\bigl(1 - F_X(x)\bigr)^{n-1} f_X(x)\,, \quad f_{X_{(n)}}(x) = n\bigl(F_X(x)\bigr)^{n-1} f_X(x)\,. \tag{4.4}$$

Example 4.1 *Suppose that the times of eight sprinters are independent random variables with common $\mathcal{U}(9.6, 10.0)$ distribution. Find $\mathsf{P}(X_{(1)} < 9.69)$, the probability that the winning result is smaller than 9.69.*

Solution. Since the density of X is $f_X(x) = 2.5 \cdot \mathbb{1}_{(9.6,10.0)}(x)$, we get, with $x = 9.69$,

$$\mathsf{P}(X_{(1)} < x) = 1 - \bigl(1 - F_X(x)\bigr)^8 \equiv 1 - \bigl(25 - 2.5 \cdot 9.69\bigr)^8 \approx 0.86986\,. \qquad \square$$

Exercise 4.1 (*). Let independent variables $X_1, \dots, X_n$ have $\mathcal{U}(0,1)$ distribution. Show that for every $x \in (0,1)$, we have $\mathsf{P}\bigl(X_{(1)} < x\bigr) \to 1$ and $\mathsf{P}\bigl(X_{(n)} > x\bigr) \to 1$ as $n \to \infty$.

4.2 Intermediate order variables

We now describe the distribution of the kth order variable $X_{(k)}$:

Theorem 4.2 *For $k = 1, 2, \dots, n$,*

$$F_{X_{(k)}}(x) \equiv \mathsf{P}\bigl(X_{(k)} \le x\bigr) = \sum_{\ell=k}^{n} \binom{n}{\ell} \bigl(F_X(x)\bigr)^{\ell} \bigl(1 - F_X(x)\bigr)^{n-\ell}\,. \tag{4.5}$$

Proof. In a sample of n independent values, the number of observations not bigger than x has $\mathsf{Bin}(n, F_X(x))$ distribution. Further, the events

$$A_m(x) \stackrel{\text{def}}{=} \bigl\{\text{exactly } m \text{ of } X_1, X_2, \dots, X_n \text{ are not bigger than } x\bigr\}$$

are disjoint for different m, and $\mathsf{P}\bigl(A_m(x)\bigr) = \binom{n}{m}\bigl(F_X(x)\bigr)^m \bigl(1 - F_X(x)\bigr)^{n-m}$. The result now follows from

$$\mathsf{P}\bigl(X_{(k)} \le x\bigr) \equiv \mathsf{P}\Bigl(\bigcup_{m \ge k} A_m(x)\Bigr) = \sum_{m \ge k} \mathsf{P}(A_m(x))\,. \qquad \square$$

Remark 4.2.1 One can show that

$$\mathsf{P}(X_{(k)} \le x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} y^{k-1}(1-y)^{n-k}\, dy\,. \tag{4.6}$$

We shall derive it below using a different approach.

Exercise 4.2 (**). By using induction or otherwise, prove (4.6).

Corollary 4.3 *If X has density $f(x)$, then*

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} \bigl(F_X(x)\bigr)^{k-1} \bigl(1 - F_X(x)\bigr)^{n-k} f(x)\,. \tag{4.7}$$

Exercise 4.3 (*). Derive (4.7) by differentiating (4.6).

Remark 4.3.1 If X has continuous cdf $F(\cdot)$, then $Y \stackrel{\mathsf{def}}{=} F(X) \sim \mathcal{U}(0,1)$, see Exercise 4.4. Consequently, the diagramm

$$\begin{array}{ccc} \{X_1,\dots,X_n\} & \longrightarrow & \bigl(X_{(1)},\dots,X_{(n)}\bigr) \\ \downarrow & & \downarrow \\ \{Y_1,\dots,Y_n\} & \longrightarrow & \bigl(Y_{(1)},\dots,Y_{(n)}\bigr) \end{array}$$

is commutative with $Y_{(k)} \equiv F\bigl(X_{(k)}\bigr)$. Notice, that if $F(\cdot)$ is strictly increasing on its support, the maps corresponding to $\downarrow$ are one-to-one, ie., one can invert them by writing $X_k = F^{-1}(Y_k)$ and $X_{(k)} = F^{-1}(Y_{(k)})$. We use this fact to give an alternative proof of Corollary 4.3.

Exercise 4.4 (*). If X has continuous cdf $F(\,\cdot\,)$, show that $Y \stackrel{\mathsf{def}}{=} F(X) \sim \mathcal{U}(0,1)$.

Proof of Corollary 4.3. Let $(Y_j)_{j=1}^n$ be a n-sample from $\mathcal{U}(0,1)$, and let $Y_{(k)}$ be the kth order variable. Fix $y \in (0,1)$ and $h > 0$ such that $y+h < 1$, and consider the following events: $C_k^{(y,h)} \stackrel{\mathsf{def}}{=} \{Y_{(k)} \in (y, y+h]\}$, $B_1 \stackrel{\mathsf{def}}{=} \bigl\{\text{exactly one } Y_j \text{ in } (y,y+h]\bigr\}$, and $B_2 \stackrel{\mathsf{def}}{=} \bigl\{\text{at least two } Y_j \text{ in } (y,y+h]\bigr\}$; clearly, $\mathsf{P}(C_k^{(y,h)}) = \mathsf{P}(C_k^{(y,h)} \cap B_1\bigr) + \mathsf{P}(C_k^{(y,h)} \cap B_2\bigr)$. We obviously have

$$\mathsf{P}\bigl(C_k^{(y,h)} \cap B_2\bigr) \le \mathsf{P}(B_2) \le \sum_{m\ge 2} \tbinom{n}{m} h^m = O(h^2) \qquad \text{as } h \to 0\,.$$

On the other hand, on the event $C_k^{(y,h)} \cap B_1$, there are exactly $k-1$ of Y_j's in $(0,y]$, exactly one Y_j in $(y,y+h]$, and exactly $n-k$ of Y_j's in $(y+h,1)$. By independence, this implies that [3]

$$\mathsf{P}\bigl(C_k^{(y,h)} \cap B_1\bigr) = \frac{n!}{(k-1)!\cdot 1!\cdot (n-k)!}\, y^{k-1}\cdot h \cdot (1-y-h)^{n-k}$$

thus implying (4.7) for the uniform distribution. The general case now follows as indicated in Remark 4.3.1. □

Remark 4.3.2 In what follows, we will often use provide proofs for the $\mathcal{U}(0,1)$ case, and refer to Remark 4.3.1 for the appropriate generalizations.
Alternatively, an argument similar to the proof above works directly for $X_{(k)}$ variables, see, eg., [1, Theorem 4.1.2].

4.3 Distribution of the range

For an n-sample $\{X_1,\dots,X_n\}$ with continuous density, let $(X_{(1)},\dots,X_{(n)})$ be the corresponding order statistic. Our aim here is to describe the definition of the range [4] R_n, defined via

$$\mathsf{R}_n \stackrel{\mathsf{def}}{=} X_{(n)} - X_{(1)}\,.$$

We start with the following simple observation.

[3] recall multinomial distributions!

[4] it characterises the spread of the sample.

Lemma 4.4 *The joint density of $X_{(1)}$ and $X_{(n)}$ is given by*

$$f_{X_{(1)},X_{(n)}}(x,y) = \begin{cases} n(n-1)\big(F(y)-F(x)\big)^{n-2} f(x)f(y)\,, & x<y\,,\\ 0 & \textit{else}\,. \end{cases} \tag{4.8}$$

Proof. Since $\mathsf{P}(X_{(n)} \le y) = \big(F(y)\big)^n$, and for $x<y$ we have

$$\mathsf{P}(X_{(1)} > x, X_{(n)} \le y) = \prod_{k=1}^{n} \mathsf{P}(x < X_k \le y) = \big(F(y)-F(x)\big)^n\,,$$

it follows that

$$\mathsf{P}(X_{(1)} \le x, X_{(n)} \le y) = \big(F(y)\big)^n - \mathsf{P}(X_{(1)} > x, X_{(n)} \le y)\,.$$

Now, differentiate w.r.t. x and y. □

Exercise 4.5 (**). In the case of an n-sample from $\mathcal{U}(0,1)$ distribution, derive (4.8) directly from combinatorics (cf. the second proof of Corollary 4.3), and then use the approach in Remark 4.3.1 to extend your result to the general case.

Theorem 4.5 *The density of R_n is given by*

$$f_{\mathsf{R}_n}(r) = n(n-1)\int \big(F(z+r)-F(z)\big)^{n-2} f(z)f(z+r)\,dz\,. \tag{4.9}$$

Exercise 4.6 (*). Prove the density $f_{\mathsf{R}_n}(r)$ formula (4.9).

Example 4.6 *If $X \sim \mathcal{U}(0,1)$, the density (4.9) becomes*

$$f_{\mathsf{R}_n}(r) = n(n-1)\int_0^{1-r} \big(z+r-z\big)^{n-2}\,dz = n(n-1)\,r^{n-2}(1-r) \tag{4.10}$$

for $0<r<1$ and $f_{\mathsf{R}_n}(r)=0$ otherwise, so that[5] $\mathsf{E}\mathsf{R}_n = \frac{n-1}{n+1}$.
Alternatively, *observe that for $X \sim \mathcal{U}(0,1)$, the variables $X_{(1)}$ and $1-X_{(n)}$ have the same distribution with $\mathsf{E}X_{(n)} = \int_0^1 y\,dF_{X_{(n)}}(y) = n\int_0^1 y^n\,dy = \frac{n}{n+1}$.*

Solution. See Exercise 4.7. □

Remark 4.6.1 Recal that the beta distribution $\beta(k,m)$ has density

$$f(x) = \frac{\Gamma(k+m)}{\Gamma(k)\Gamma(m)}\,x^{k-1}(1-x)^{m-1} \equiv \frac{(k+m-1)!}{(k-1)!(m-1)!}\,x^{k-1}(1-x)^{m-1}\,, \tag{4.11}$$

where $0<x<1$ and $f(x)=0$ otherwise.

Exercise 4.7 (*). Let $X \sim \beta(k,m)$, ie., X has beta distribution with parameters k and m. Show that $\mathsf{E}X = \frac{k}{k+m}$ and $\mathsf{Var}X = \frac{km}{(k+m)^2(k+m+1)}$.

[5] in particular, $\mathsf{R}_n \sim \beta(n-1,2)$, where the beta distribution is defined in (4.11).

Example 4.7 *Let X_1, X_2 be a sample from $\mathcal{U}(0,1)$, and let $X_{(1)}$, $X_{(2)}$ be the corresponding order statistics. Find the* pdf *for each of the random variables: $X_{(1)}$, $\mathsf{R}_2 = X_{(2)} - X_{(1)}$ and $1 - X_{(2)}$.*

Solution. By (4.10), $f_{\mathsf{R}_2}(x) = 2(1-x)$ for $0 < x < 1$ and $f_{\mathsf{R}_2}(x) = 0$ otherwise. On the other hand, by (4.4), $f_{X_{(1)}}(x)$ is given by the same expression and, by symmetry (or again by (4.4)), $1 - X_{(2)}$ has the same density as well. □

Example 4.8 *Let X_1, X_2, X_3 be a sample from $\mathcal{U}(0,1)$, and let $X_{(1)}$, $X_{(2)}$, $X_{(3)}$ be the corresponding order statistics. Find the* pdf *for each of the random variables: $X_{(2)}$, $\mathsf{R}_3 = X_{(3)} - X_{(1)}$ and $1 - X_{(2)}$.*

Solution. By (4.10), $f_{\mathsf{R}_2}(x) = 6x(1-x)$ for $0 < x < 1$ and $f_{\mathsf{R}_2}(x) = 0$ otherwise. On the other hand, by (4.7), $f_{X_{(2)}}(x)$ is given by the same expression and, by symmetry, $1 - X_{(2)}$ has the same density as well. Can you explain these findings? □

4.4 Gaps distribution

There are many interesting (and important for applications) questions about order statistics. For example, given an n-sample from $\mathcal{U}(0,1)$ with the order statistic

$$0 \equiv X_{(0)} < X_{(1)} < X_{(2)} < \cdots < X_{(n)} < X_{(n+1)} \equiv 1\,,$$

one is often interested in distribution of the extreme order variables $X_{(1)}$ and $X_{(n)}$, as well as in distribution of the kth gap

$$\Delta_{(k)}X \stackrel{\mathsf{def}}{=} X_{(k)} - X_{(k-1)}\,,$$

of the maximal gap $\max \Delta_{(k)}X = \max\bigl(X_{(k)} - X_{(k-1)}\bigr)$, of the minimal gap $\min \Delta_{(k)}X = \min\bigl(X_{(k)} - X_{(k-1)}\bigr)$ (where k ranges from 1 to $n+1$) among others.

Example 4.9 *Given an n-sample from $\mathcal{U}(0,1)$, we have, for every $a \ge 0$,*

$$\mathsf{P}\bigl(nX_{(1)} > a\bigr) = \mathsf{P}\bigl(X_{(1)} > a/n\bigr) = \mathsf{P}(X > a/n)^n = \bigl(1 - a/n\bigr)^n \approx e^{-a}\,,$$

in other words, the distribution of $nX_{(1)}$ for n large enough is approximately $\mathsf{Exp}(1)$, *ie., the exponential distribution with parameter one.*[6]

Remark 4.9.1 Notice that the previous example describes the distribution of the first gap, $X_{(1)} - X_{(0)} \equiv X_{(1)}$, namely, $\mathsf{P}(X_{(1)} > r) = (1-r)^n$. By symmetry, the last gap $X_{(n+1)} - X_{(n)}$ has the same distribution.

Exercise 4.8 (**). Let $X_{(1)}$ be the first order variable from an n-sample with density $f(\,\cdot\,)$, which is positive and continuous on $[0,1]$, and vanishes otherwise. Let, further, $f(0) = c > 0$. For fixed $y > 0$, show that $\mathsf{P}\bigl(X_{(1)} > \frac{y}{n}\bigr) \approx e^{-cy}$ for large n. Deduce that the distribution of $Y_n \equiv nX_{(1)}$ is approximately $\mathsf{Exp}(c)$ for large enough n.

[6] recall that $X \sim \mathsf{Exp}(\lambda)$ if for every $x \ge 0$ we have $\mathsf{P}(X > x) = e^{-\lambda x}$.

Exercise 4.9 (**). Let $X_1, \dots, X_n$ be independent positive random variables whose joint probability density function $f(\,\cdot\,)$ is right-continuous at the origin and satisfies $f(0) = c > 0$. For fixed $y > 0$, show that

$$\mathsf{P}\big(X_{(1)} > \tfrac{y}{n}\big) \approx e^{-cy}$$

for large n. Deduce that the distribution of $Y_n \stackrel{\mathsf{def}}{=} nX_{(1)}$ is approximately $\mathsf{Exp}(c)$ for large enough n.

Lemma 4.10 *For a given n-sample from $\mathcal{U}(0,1)$, all gaps*

$$\Delta_{(k)}X \stackrel{\mathsf{def}}{=} X_{(k)} - X_{(k-1)}\,, \qquad k = 1, \dots, n+1\,,$$

have the same distribution with $\mathsf{P}(\Delta_{(k)}X > r) = (1-r)^n$ for all $r \in [0,1]$.

Proof. As in the second proof of Corollary 4.3, we deduce that

$$f_{X_{(k-1)},X_{(k)}}(x,y) = \frac{n!}{(k-2)!(n-k)!}x^{k-2}(1-y)^{n-k}\,;$$

this implies that

$$f_{\Delta_{(k)}X}(r) = \int_0^{1-r} f_{X_{(k-1)},X_{(k)}}(x, x+r)\,dx\,.$$

Changing the variables $x \mapsto y$ using $x = y(1-r)$ and comparing the result to the beta distribution (4.11), we deduce that $f_{\Delta_{(k)}X}(r) = n(1-r)^{n-1}\mathbb{1}_{0<r<1}$, equivalently, that $\mathsf{P}(\Delta_{(k)}X \ge r) = (1-r)^n$ for all $r \in [0,1]$. □

Remark 4.10.1 By symmetry, $\mathsf{E}\Delta_{(k)}X = \frac{1}{n+1}$ for all k. The value of $\mathsf{E}R_n$ derived in Example 4.6 is now immediate.

Remark 4.10.2 By the previous example, for every k and n large, the distribution of $n\Delta_{(k)}X$ is approximately[7] $\mathsf{Exp}(1)$.

Remark 4.10.3 As we will see below, for every fixed n, the joint distribution of the vector of gaps

$$\big(\Delta_{(1)}X, \dots, \Delta_{(n+1)}X\big)$$

is symmetric with respect to permutations of individual segments,[8] however, the individual gaps are not independent.

Lemma 4.11 *Let an n-sample from $\mathcal{U}(0,1)$ be fixed. Then for all m, k, satisfying $1 \le m < k \le n+1$ we have*

$$\mathsf{P}\big(\Delta_{(m)}X \ge r_m, \Delta_{(k)}X \ge r_k\big) = (1 - r_m - r_k)^n\,, \qquad (4.12)$$

for all positive r_m, r_k satisfying $r_m + r_k \le 1$.

[7] more precisely, for every sequence $(k_n)_{n\ge1}$ with $k_n \in \{1, 2, \dots, n\}$, the distribution of the rescaled gap $n\Delta_{(k_n)}X$ converges to $\mathsf{Exp}(1)$ as $n \to \infty$.

[8] ie., the gaps $\Delta_{(1)}X, \dots, \Delta_{(n+1)}$ are exchangeable random variables;

Proof. We fix the value of $X_{(m)}$, and use the formula of total probability with integration over the value of $X_{(m)} \in (r_m, 1-r_k)$. Notice, that given $X_{(m)} = x$, the distributions of $(X_{(1)}, \dots, X_{(m-1)})$ and $(X_{(m+1)}, \dots, X_{(n)})$ are conditionally independent. Using the scaled version of Lemma 4.10 we deduce that [9]

$$\mathsf{P}_n\big(\Delta_{(m)}X \ge r_m \mid X_{(m)} = x\big) \equiv \mathsf{P}_{m-1}\Big(\Delta_{(m)}Y \ge \frac{r_m}{x}\Big) = \Big(1 - \frac{r_m}{x}\Big)^{m-1},$$

$$\mathsf{P}_n\big(\Delta_{(k)}X \ge r_k \mid X_{(m)} = x\big) \equiv \mathsf{P}_{n-m}\Big(\Delta_{(k-m)}Y \ge \frac{r_k}{1-x}\Big) = \Big(1 - \frac{r_k}{1-x}\Big)^{n-m}.$$

By the formula of total probability,

$$\mathsf{P}\big(\Delta_{(m)}X \ge r_m, \Delta_{(k)}X \ge r_k\big) = \int_{r_m}^{1-r_k} \Big(1 - \frac{r_m}{x}\Big)^{m-1} \Big(1 - \frac{r_k}{1-x}\Big)^{n-m} f_{X_{(m)}}(x)\,dx,$$

so that, using the explicit expression (4.7) for the density of $X_{(m)}$ and changing the variables via $x = r_m + (1 - r_m - r_k)y$, we get

$$\big(1 - r_m - r_k\big)^n \int_0^1 \frac{n!}{(m-1)!(n-m)!} y^{m-1}(1-y)^{n-m}\,dy \equiv (1 - r_m - r_k)^n,$$

where the last equality follows directly from (4.11). □

Remark 4.11.1 Notice that by (4.12), the gaps $\Delta_{(m)}X$ and $\Delta_{(k)}X$ are not independent. However, for n large enough, we have

$$\begin{aligned}\mathsf{P}\Big(\Delta_{(m)}X \ge \frac{r_m}{n}, \Delta_{(k)}X \ge \frac{r_k}{n}\Big) &\approx e^{-r_m - r_k} \\ &= e^{-r_m}e^{-r_k} \approx \mathsf{P}\Big(\Delta_{(m)}X \ge \frac{r_m}{n}\Big)\mathsf{P}\Big(\Delta_{(k)}X \ge \frac{r_k}{n}\Big),\end{aligned}$$

ie., the rescaled gaps $n\Delta_{(m)}X$ and $n\Delta_{(k)}X$ are asymptotically independent (and asymptotically $\mathsf{Exp}(1)$-distributed).

Exercise 4.10 (**). Let X_1, X_2, X_3, X_4 be a sample from $\mathcal{U}(0,1)$, and let $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, $X_{(4)}$ be the corresponding order statistics. Find the pdf for each of the random variables: $X_{(2)}$, $X_{(3)} - X_{(1)}$, $X_{(4)} - X_{(2)}$, and $1 - X_{(3)}$.

Corollary 4.12 *Let $(X_k)_{k=1}^n$ be a fixed n-sample from $\mathcal{U}(0,1)$. Then for all positive r_k satisfying $\sum_{k=1}^{n+1} r_k \le 1$ we have*

$$\mathsf{P}\big(\Delta_{(1)}X \ge r_1, \dots, \Delta_{(n+1)}X \ge r_{n+1}\big) = \Big(1 - \sum_{k=1}^{n+1} r_k\Big)^n. \qquad (4.13)$$

In particular, the distribution of the vector of gaps $\big(\Delta_{(1)}X, \dots, \Delta_{(n+1)}X\big)$ is exchangeable for every $n \ge 1$.

[9] with Y denoting a sample from $\mathcal{U}(0,1)$;

Remark 4.12.1 Similarly to Remark 4.11.1, one can use (4.13) to deduce that any finite collection[10] of rescaled gaps, say $(n\Delta_{(k_1)}X, \dots, n\Delta_{(k_m)}X)$, becomes asymptotically independent (with individual components $\mathsf{Exp}(1)$-distributed) in the limit of large n.

Exercise 4.11 (***). Prove the asymptotic independence property of any finite collection of gaps stated in Remark 4.12.1.

Exercise 4.12 (***). Using induction or otherwise, prove (4.13).

Example 4.13 *For an n-sample from $\mathcal{U}(0,1)$, let $D_n \stackrel{\mathsf{def}}{=} \min_k \Delta_{(k)}X$. Then (4.13) implies that $\mathsf{P}(D_n \ge r) = \big(1-(n+1)r\big)^n$. In particular, for every $y \ge 0$,*

$$\mathsf{P}_n\big(n^2 D_n > y\big) = \Big(1 - \frac{n+1}{n^2}y\Big)^n \approx e^{-y}\,, \qquad (4.14)$$

ie., for large n the distribution of $Y_n \stackrel{\mathsf{def}}{=} n^2 D_n$ is approximately $\mathsf{Exp}(1)$.

Remark 4.13.1 For an n-sample from $\mathcal{U}(0,1)$, a typical gap is of order n^{-1} (recall Remark 4.10.2), whereas by (4.14) the minimal gap D_n is of order n^{-2}.

4.5 Order statistic from exponential distributions

If $X \sim \mathsf{Exp}(\lambda)$, ie., $\mathsf{P}(X > x) = e^{-\lambda x}$ for all $x \ge 0$, then $Y \stackrel{\mathsf{def}}{=} \lambda X \sim \mathsf{Exp}(1)$. It is thus enough to consider $\mathsf{Exp}(1)$ random variables.

Example 4.14 *If $(X_k)_{k=1}^n$ is an n-sample from $\mathsf{Exp}(1)$, then $X_{(1)} \sim \mathsf{Exp}(n)$.*

Solution. By independence, for every $y \ge 0$, $\mathsf{P}(X_{(1)} > y) = \prod_{k=1}^n \mathsf{P}(X_k > y) = e^{-ny}$. □

Exercise 4.13 (*). Let $X_k \sim \mathsf{Exp}(\lambda_k)$, $k = 1, \dots, n$, be independent with fixed $\lambda_k > 0$. Denote $X_0 = \min\{X_1, \dots, X_n\}$ and $\lambda_0 = \sum_{k=1}^n \lambda_k$. Show that for $y \ge 0$

$$\mathsf{P}\big(X_0 > y, X_0 = X_k\big) = e^{-\lambda_0 y}\,\frac{\lambda_k}{\lambda_0}\,,$$

ie., the minimum X_0 of the sample satisfies $X_0 \sim \mathsf{Exp}(\lambda_0)$ and the probability that it coincides with X_k is proportional to λ_k, independently of the value of X_0.

Recall the following memoryless property of exponential distributions:

Exercise 4.14 (*). If $X \sim \mathsf{Exp}(\lambda)$, show that for all positive a and b we have

$$\mathsf{P}\big(X > a + b \mid X > a\big) = \mathsf{P}(X > b)\,.$$

[10] with a little bit of extra work, one can also allow $m = m_n \to \infty$ slowly enough as $n \to \infty$.

Lemma 4.15 *If $(X_k)_{k=1}^n$ are independent with $X_k \sim \mathsf{Exp}(\lambda_k)$, then for arbitrary fixed positive a and $b_1, \dots, b_n$,*

$$\mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k > a + b_k\big\} \bigm| \min(X_k) > a\Big) = \mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k > b_k\big\}\Big)\,. \tag{4.15}$$

In addition,

$$\mathsf{P}\Big(\bigcap_{k=1}^n \big\{a < X_k \le a + b_k\big\}\Big) = \mathsf{P}\big(\min(X_k) > a\big)\,\mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k \le b_k\big\}\Big)\,. \tag{4.16}$$

Proof. The results follows from independence and the fact $\mathsf{P}(X_k > c) = e^{-\lambda_k c}$. □

Remark 4.15.1 By the uniformity in $a \ge 0$ of the memoryless property (4.15), the latter also holds for random thresholds; eg., if $Y \ge 0$ is a random variable with density $f_Y(y)$ independent of $\{X_1, \dots, X_n\}$, then conditioning on the value of Y and using the formula of total probability, we get

$$\begin{aligned}&\mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k > Y + b_k\big\} \bigm| \min(X_k) > Y\Big)\\ &\qquad = \int f_Y(a)\mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k > a + b_k\big\} \bigm| \min(X_k) > a\Big)\,da = \mathsf{P}\Big(\bigcap_{k=1}^n \big\{X_k > b_k\big\}\Big)\,.\end{aligned}$$

Exercise 4.15 (**). If $X \sim \mathsf{Exp}(\lambda)$, $Y \sim \mathsf{Exp}(\mu)$ and $Z \sim \mathsf{Exp}(\nu)$ are independent, show that for every constant $a \ge 0$ we have $\mathsf{P}(a + X < Y \,|\, a < Y) = \frac{\lambda}{\lambda+\mu}$; deduce that $\mathsf{P}\big(a + X < \min(Y, Z) \,|\, a < \min(Y, Z)\big) = \frac{\lambda}{\lambda+\mu+\nu}$.

Corollary 4.16 *Let $(X_k)_{k=1}^n$ be an n-sample from $\mathsf{Exp}(1)$. Denote*

$$Z_n = X_1 + \tfrac{1}{2}X_2 + \dots + \tfrac{1}{n}X_n\,.$$

Then $X_{(n)}$ and Z_n have the same distribution.[11]

Proof. Write Y_m for the maximum $X_{(m)}$ of an m-sample $(X_k)_{k=1}^m$ from $\mathsf{Exp}(1)$. We obviously have $\mathsf{P}(Z_1 \le a) = 1 - e^{-a}$ and $\mathsf{P}(Y_m \le a) = (1 - e^{-a})^m$ for all $m \ge 1$. Since Z_m has the same distribution as the (independent) sum $Z_{m-1} + U_m$ with $U_m \sim \mathsf{Exp}(m)$,

$$\mathsf{P}\big(Z_m \le a\big) = \mathsf{P}\big(Z_{m-1} + U_m \le a\big) = \int_0^a \mathsf{P}(Z_{m-1} \le a - x)\, m e^{-mx}\, dx\,,$$

for all $m \ge 1$. Next, by symmetry,

$$\mathsf{P}\big(Y_m \le a\big) = m\mathsf{P}\big(X_m < \min(X_1, \dots, X_{m-1}) < Y_{m-1} \le a\big)\,,$$

so that, using (4.16), this becomes

$$m\int_0^a f_{X_m}(x)\,\mathsf{P}\Big(\bigcap_{k=1}^{m-1}\big\{x < X_k \le a\big\}\Big)\,dx = \int_0^a m e^{-mx}\big(1 - e^{-(a-x)}\big)^{m-1}\,dx\,.$$

[11] One can also show that $X_{(k)}$ has the same distribution as $\frac{1}{n+1-k}V_{n+1-k} + \dots + \frac{1}{n}V_n$ with jointly independent $V_k \sim \mathsf{Exp}(1)$.

The result now follows by straightforward induction.[12] □

Exercise 4.16 (**). Carefully prove Corollary 4.16 and compute $\mathsf{E}Y_n$ and $\mathsf{Var}Y_n$.

Exercise 4.17 (**). Let $X_{(n)}$ be the maximum of an n-sample from $\mathsf{Exp}(1)$ distribution. For $x \in \mathbb{R}$, find the value of $\mathsf{P}\big(X_{(n)} \le \log n + x\big)$ in the limit $n \to \infty$.

4.6 Additional problems

Exercise 4.18 (*). Let X_1, X_2 be a sample from a uniform distribution on $\{1,2,3,4,5\}$. Find the distribution of $X_{(1)}$, the minimum of the sample.

Exercise 4.19 (*). Let independent variables $X_1, \dots, X_n$ be $\mathsf{Exp}(1)$ distributed. Show that for every $x > 0$, we have $\mathsf{P}\big(X_{(1)} \le x\big) \to 1$ and $\mathsf{P}\big(X_{(n)} \ge x\big) \to 1$ as $n \to \infty$. Generalise the result to arbitrary distributions on $\mathbb{R}$.

Exercise 4.20 (*). Let $\{X_1, X_2, X_3, X_4\}$ be a sample from a distribution with density $f(x) = e^{7-x}\mathbb{1}_{x>7}$. Find the pdf of the second order variable $X_{(2)}$.

Exercise 4.21 (*). Let X_1 and X_2 be independent $\mathsf{Exp}(\lambda)$ random variables.
a) Show that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent and find their distributions.
b) Compute $\mathsf{E}(X_{(2)} \mid X_{(1)} = x_1)$ and $\mathsf{E}(X_{(1)} \mid X_{(2)} = x_2)$.

Exercise 4.22 (**). Let $(X_k)_{k=1}^n$ be an n-sample from $\mathsf{Exp}(\lambda)$ distribution.
a) Show that the gaps $\big(\Delta_{(k)}X\big)_{k=1}^n$ as defined in Lemma 4.10 are independent and find their distribution.
b) For fixed $1 \le k \le m \le n$, compute the expectation $\mathsf{E}(X_{(m)} \mid X_{(k)} = x_k)$.

Exercise 4.23 (**). If $Y \sim \mathsf{Exp}(\mu)$ and an arbitrary random variable $X \ge 0$ are independent, show that for every $a > 0$, $\mathsf{P}(a + X < Y \mid a < Y) = \mathsf{E}e^{-\mu X}$.

4.6.1 Optional material[13]

Exercise 4.24 (**). In the context of Exercise 4.18, let $\{X_1, X_2\}$ be a sample without replacement from $\{1,2,3,4,5\}$. Find the distribution of $X_{(1)}$, the minimum of the sample.

Exercise 4.25 (*). Let $\{X_1, X_2\}$ be an independent sample from $\mathsf{Geom}(p)$ distribution, $\mathsf{P}(X > k) = (1-p)^k$ for integer $k \ge 0$. Find the distribution of $X_{(1)}$, the minimum of the sample.

Exercise 4.26 (**). Let $\{X_1, X_2, \dots, X_n\}$ be an n-sample from a distribution with density $f(\,\cdot\,)$. Show that the joint density of the order variables $X_{(1)}, \dots, X_{(n)}$ is

$$f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n) = n!\prod_{k=1}^n f(x_k)\,\mathbb{1}_{x_1<\dots<x_n}\,.$$

[12] You might wish to check that the RHS above equals $(1-e^{-a})^m$; to this end, change the variables $x \mapsto y = a - x$, rearrange, then change again $y \mapsto z = e^y$ and integrate the result.

[13] This section contains advanced material for students who wish to dig a little deeper. Problems marked with (****) are strictly optional.

Exercise 4.27 (**). Let $\{X_1, X_2, X_3\}$ be a sample from $\mathcal{U}(0,1)$. Find the conditional density $f_{X_{(1)},X_{(3)}|X_{(2)}}(x,z|y)$ of $X_{(1)}$ and $X_{(3)}$ given that $X_{(2)} = y$. Explain your findings.

Exercise 4.28 (**). Let $\{X_1, X_2, \dots, X_{100}\}$ be a sample from $\mathcal{U}(0,1)$. Approximate the value of $\mathsf{P}(X_{(75)} \le 0.8)$.

Exercise 4.29 (**). Let X_1, X_2, ... be independent random variables with cdf $F(\,\cdot\,)$, and let $N > 0$ be an integer-valued variable with probability generating function $g(\,\cdot\,)$, independent of the sequence $(X_k)_{k\ge1}$. Find the cdf of $\max\{X_1, X_2, \dots, X_N\}$, the maximum of the first N terms in that sequence.

Exercise 4.30 (***). Let $X_{(1)}$ be the first order variable from an n-sample with density $f(\,\cdot\,)$, which is positive and continuous on $[0,1]$, and vanishes otherwise. Let, further, $f(x) \approx cx^\alpha$ for small $x > 0$ and positive c and α. For $y > 0$ and $\beta = \frac{1}{\alpha+1}$, show that the probability $\mathsf{P}(X_{(1)} > yn^{-\beta})$ has a well defined limit for large n. What can you deduce about the distribution of the rescaled variable $Y_n \stackrel{\mathsf{def}}{=} n^\beta X_{(1)}$ for large enough n?

Exercise 4.31 (***). Let X_1, ..., X_n be independent $\beta(k,m)$-distributed random variables whose joint distribution is given in (4.11) (with $k \ge 1$ and $m \ge 1$). Find $\delta > 0$ such that the distribution of the rescaled variable $Y_n \stackrel{\mathsf{def}}{=} n^\delta X_{(1)}$ converges to a well-defined limit as $n \to \infty$. Describe the limiting distribution.

Exercise 4.32 (****). In the situation of Exercise 4.30, let $\alpha < 0$. What can you say about possible limiting distribution of the suitably rescaled fisrt order variable, $Y_n = n^\delta X_{(1)}$, with some $\delta \in \mathbb{R}$?

Exercise 4.33 (****). Denote $Y_n^* = Y_n - \log n$; show that the corresponding cdf, $\mathsf{P}(Y_n^* \le x)$, approaches $e^{-e^{-x}}$, as $n \to \infty$. Deduce that the expectation of the limiting distribution equals γ, the Euler constant,[14] and its variance is $\sum\limits_{k\ge1} \frac{1}{k^2} = \frac{\pi^2}{6}$.

A distribution with cdf $\exp\{-e^{-(x-\mu)/\beta}\}$ is known as Gumbel distribution (with scale and locations parameters β and μ, resp.). It can be shown that its average is[14] $\mu + \beta\gamma$, its variance is $\pi^2\beta^2/6$, and its moment generating function equals $\Gamma(1-\beta t)e^{\mu t}$.

Exercise 4.34 (****). By using the Weierstrass formula,

$$\prod_{k=1}^{\infty}\Bigl(1+\frac{z}{k}\Bigr)^{-1} e^{z/k} = e^{\gamma z}\Gamma(z+1)\,,$$

(where γ is the Euler constant [14] and $\Gamma(\cdot)$ is the classical gamma function, $\Gamma(n) = (n-1)!$ for integer $n > 0$) or otherwise, show that the moment generating function $\mathsf{E}e^{tZ_n^*}$ of $Z_n^* = Z_n - \mathsf{E}Z_n$ approaches $e^{-\gamma t}\Gamma(1-t)$ as $n \to \infty$ (eg., for all $|t| < 1$). Deduce that in that limit $Z_n^* + \gamma$ is asymptotically Gumbel distributed (with $\beta = 1$ and $\mu = 0$).

Exercise 4.35 (***). Let X_1, X_2, ... be independent $\mathsf{Exp}(\lambda)$ random variables; further, let $N \sim \mathsf{Poi}(\nu)$, independent of the sequence $(X_k)_{k\ge1}$, and let $X_0 \equiv 0$. Find the distribution of $Y \stackrel{\mathsf{def}}{=} \max\{X_0, X_1, X_2, \dots, X_N\}$, the maximum of the first N terms of this sequence, where for $N = 0$ we set $Y = 0$.

[14] the Euler constant γ is $\lim_{n\to\infty}\bigl(\sum_{k=1}^n \frac{1}{k} - \log n\bigr) \approx 0.5772156649...$;

5 Coupling

Two random variables, say X and Y, are coupled, if they are defined on the same probablity space. To couple two given variables X and Y, one usually defines a random vector $\bigl(\widetilde{X},\widetilde{Y}\bigr)$ with joint probability $\widetilde{\mathsf{P}}(\,\cdot\,,\,\cdot\,)$ on some probability space[15] so that the marginal distribution of $\widetilde{X}$ coincides with the distribution of X and the marginal distribution of $\widetilde{Y}$ coincides with the distribution of Y.

Example 5.1 *Fix p_1, $p_2 \in [0,1]$ such that $p_1 \le p_2$ and consider the following joint distributions (we write $q_i = 1 - p_i$):*

T_1	0	1	$\widetilde{X}$
0	q_1q_2	q_1p_2	q_1
1	p_1q_2	p_1p_2	p_1
$\widetilde{Y}$	q_2	p_2	

T_2	0	1	$\widetilde{X}$
0	q_2	$p_2 - p_1$	q_1
1	0	p_1	p_1
$\widetilde{Y}$	q_2	p_2	

It is easy to see that in both cases $\widetilde{X} \sim \mathsf{Ber}(p_1)$, $\widetilde{Y} \sim \mathsf{Ber}(p_2)$, though in the first case $\widetilde{X}$ and $\widetilde{Y}$ are independent (and the joint distribution is known as an "independent coupling"), whereas in the second case we have $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$ (and the joint distribution is called a "monotone coupling").

Exercise 5.1 (*). In the setting of Example 5.1 show that every convex linear combination of tables T_1 and T_2, ie., each table of the form $T_\alpha = \alpha T_1 + (1-\alpha)T_2$ with $\alpha \in [0,1]$, gives an example of a coupling of $X \sim \mathsf{Ber}(p_1)$ and $Y \sim \mathsf{Ber}(p_2)$. Can you find all possible couplings for these variables?

5.1 Stochastic order

If $X \sim \mathsf{Ber}(p)$, its tail probabilities $\mathsf{P}(X > a)$ satisfy

$$\mathsf{P}(X > a) = \begin{cases} 1, & a < 0\,, \\ p, & 0 \le a < 1\,, \\ 0, & a \ge 1\,. \end{cases}$$

Consequently, in the setup of Example 5.1, for the variable $X \sim \mathsf{Ber}(p_1)$ and $Y \sim \mathsf{Ber}(p_2)$ with $p_1 \le p_2$ we have $\mathsf{P}(X > a) \le \mathsf{P}(Y > a)$ for all $a \in \mathbb{R}$. The last inequality is useful enough to deserve a name:

✍ **Definition 5.2** [Stochastic domination] *A random variable X is stochastically smaller than a random variable Y (write $X \preccurlyeq Y$) if the inequality*

$$\mathsf{P}\bigl(X > x\bigr) \le \mathsf{P}\bigl(Y > x\bigr) \tag{5.1}$$

holds for all $x \in \mathbb{R}$.

☞ [15] A priori the original variables X and Y can be defined on arbitrary probability spaces, so that one has no reason to expect that these spaces can be "joined" in any way!

Exercise 5.2 (*). Let $X \ge 0$ be a random variable, and let $a \ge 0$ be a fixed constant. If $Y = a + X$, is it true that $X \preccurlyeq Y$? If $Z = aX$, is it true that $X \preccurlyeq Z$? Justify your answer.

Exercise 5.3 (*). If $X \preccurlyeq Y$ and $g(\cdot)$ is an arbitrary increasing function on $\mathbb{R}$, show that $g(X) \preccurlyeq g(Y)$.

☞ **Example 5.3** *Let random variables $X \ge 0$ and $Y \ge 0$ be stochastically ordered, $X \preccurlyeq Y$. If $g(\cdot) \ge 0$ is a smooth increasing function on $\mathbb{R}$ with $g(0) = 0$, then $g(X) \preccurlyeq g(Y)$ and*

$$\mathsf{E}g(X) \equiv \int_0^\infty g'(z)\mathsf{P}(X > z)\,dz \le \int_0^\infty g'(z)\mathsf{P}(Y > z)\,dz \equiv \mathsf{E}g(Y)\,. \tag{5.2}$$

Exercise 5.4 (*). Generalise the inequality in Example 5.3 to a broader class of functions $g(\cdot)$ and verify that if $X \preccurlyeq Y$, then $\mathsf{E}(X^{2k+1}) \le \mathsf{E}(Y^{2k+1})$, $\mathsf{E}e^{tX} \le \mathsf{E}e^{tY}$ with $t > 0$, $\mathsf{E}s^X \le \mathsf{E}s^Y$ with $s > 1$ etc.

Exercise 5.5 (*). Let $\xi \sim \mathcal{U}(0,1)$ be a standard uniform random variable. For fixed $p \in (0,1)$, define $X = \mathbb{1}_{\xi < p}$. Show that $X \sim \mathsf{Ber}(p)$, a Bernoulli random variable with parameter p. Now suppose that $X = \mathbb{1}_{\xi < p_1}$ and $Y = \mathbb{1}_{\xi < p_2}$ for some $0 < p_1 \le p_2 < 1$ and ξ as above. Show that $X \preccurlyeq Y$ and that $\mathsf{P}(X \le Y) = 1$. Compare your construction to the second table in Example 5.1.

Exercise 5.6 (**). In the setup of Example 5.1, show that $X \sim \mathsf{Ber}(p_1)$ is stochastically smaller than $Y \sim \mathsf{Ber}(p_2)$ (ie., $X \preccurlyeq Y$) iff $p_1 \le p_2$. Further, show that $X \preccurlyeq Y$ is equivalent to existence of a coupling $(\widetilde{X}, \widetilde{Y})$ of X and Y in which these variables are ordered with probablity one, $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$.

The next result shows that this connection between stochastic order and existence of a monotone coupling is a rather generic feature:

✍ **Lemma 5.4** *A random variable X is stochastically smaller than a random variable Y if and only if there exists a coupling $(\widetilde{X}, \widetilde{Y})$ of X and Y such that $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$.*

Remark 5.4.1 Notice that one claim of Lemma 5.4 is immediate from

$$\mathsf{P}(x < X) \equiv \widetilde{\mathsf{P}}\big(x < \widetilde{X}\big) = \widetilde{\mathsf{P}}\big(x < \widetilde{X} \le \widetilde{Y}\big) \le \widetilde{\mathsf{P}}\big(x < \widetilde{Y}\big) \equiv \mathsf{P}\big(x < Y\big)\,;$$

the other claim requires a more advanced argument (we shall not do it here!).

Example 5.5 *If $X \sim \mathsf{Bin}(m,p)$ and $Y \sim \mathsf{Bin}(n,p)$ with $m \le n$, then $X \preccurlyeq Y$.*

Solution. Let $Z_1 \sim \mathsf{Bin}(m,p)$ and $Z_2 \sim \mathsf{Bin}(n-m,p)$ be independent variables defined on the same probability space. We then put $\widetilde{X} = Z_1$ and $\widetilde{Y} = Z_1 + Z_2$ so that $\widetilde{Y} - \widetilde{X} = Z_2 \ge 0$ with probability one, $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$, and $X \sim \widetilde{X}$, $Y \sim \widetilde{Y}$. □

Example 5.6 *If $X \sim \mathsf{Poi}(\lambda)$ and $Y \sim \mathsf{Poi}(\mu)$ with $\lambda \le \mu$, then $X \preccurlyeq Y$.*

Solution. Let $Z_1 \sim \mathsf{Poi}(\lambda)$ and $Z_2 \sim \mathsf{Poi}(\mu - \lambda)$ be independent variables defined on the same probability space.[16] We then put $\widetilde{X} = Z_1$ and $\widetilde{Y} = Z_1 + Z_2$ so that $\widetilde{Y} - \widetilde{X} = Z_2 \geq 0$ with probability one, $\widetilde{\mathsf{P}}(\widetilde{X} \leq \widetilde{Y}) = 1$, and $X \sim \widetilde{X}$, $Y \sim \widetilde{Y}$. □

Exercise 5.7 (**). Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be Gaussian r.v.'s.

a) If $\mu_X \leq \mu_Y$ but $\sigma_X^2 = \sigma_Y^2$, is it true that $X \preccurlyeq Y$?

b) If $\mu_X = \mu_Y$ but $\sigma_X^2 \leq \sigma_Y^2$, is it true that $X \preccurlyeq Y$?

Exercise 5.8 (**). Let $X \sim \mathsf{Exp}(\lambda)$ and $Y \sim \mathsf{Exp}(\mu)$ be two exponential random variables. If $0 < \lambda \leq \mu < \infty$, are the variables X and Y stochastically ordered? Justify your answer by proving the result or giving a counter-example.

Exercise 5.9 (**). Let $X \sim \mathsf{Geom}(p)$ and $Y \sim \mathsf{Geom}(r)$ be two geometric random variables, $X \sim \mathsf{Geom}(p)$ and $Y \sim \mathsf{Geom}(r)$. If $0 < p \leq r < 1$, are the variables X and Y stochastically ordered? Justify your answer by proving the result or giving a counter-example.

The gamma distribution $\Gamma(a, \lambda)$ has density $\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{x>0}$ (so that $\Gamma(1, \lambda)$ is just $\mathsf{Exp}(\lambda)$). Gamma distributions have the following additive property: if $Z_1 \sim \Gamma(c_1, \lambda)$ and $Z_2 \sim \Gamma(c_2, \lambda)$ are independent random variables (with the same λ), then their sum is also gamma distributed: $Z_1 + Z_2 \sim \Gamma(c_1 + c_2, \lambda)$.

Exercise 5.10 (**). Let $X \sim \Gamma(a, \lambda)$ and $Y \sim \Gamma(b, \lambda)$ be two gamma random variables. If $0 < a \leq b < \infty$, are the variables X and Y stochastically ordered? Justify your answer by proving the result or giving a counter-example.

5.2 Total variation distance

✍ **Definition 5.7** [Total Variation Distance] *Let μ and ν be two probability measures on the same probability space. The total variation distance between μ and ν is*

$$\mathsf{d_{TV}}(\mu, \nu) \stackrel{\mathsf{def}}{=} \max_A \bigl|\mu(A) - \nu(A)\bigr| . \tag{5.3}$$

If X, Y are two discrete random variables, we write $\mathsf{d_{TV}}(X, Y)$ for the total variation distance between their distributions.

Example 5.8 *If $X \sim \mathsf{Ber}(p_1)$ and $Y \sim \mathsf{Ber}(p_2)$ we have*

$$\mathsf{d_{TV}}(X, Y) = \max\bigl\{|p_1 - p_2|, |q_1 - q_2|\bigr\} = |p_1 - p_2| = \tfrac{1}{2}\bigl(|p_1 - p_2| + |q_1 - q_2|\bigr) .$$

Exercise 5.11 (**). Let measure μ have p.m.f. $\{p_x\}_{x \in \mathcal{X}}$ and let measure ν have p.m.f. $\{q_y\}_{y \in \mathcal{Y}}$. Show that

$$\mathsf{d_{TV}}(\mu, \nu) \equiv \mathsf{d_{TV}}\bigl(\{p\}, \{q\}\bigr) = \tfrac{1}{2} \sum_z \bigl|p_z - q_z\bigr| ,$$

where the sum runs over all $z \in \mathcal{X} \cup \mathcal{Y}$. Deduce that $\mathsf{d_{TV}}(\cdot, \cdot)$ is a distance between probability measures[17](ie., it is non-negative, symmetric, and satisfies the triangle

[16] here and below we assume that $Z \sim \mathsf{Poi}(0)$ means that $\mathsf{P}(Z = 0) = 1$.

[17] so that all probability measures form a **metric space** for this distance!

inequality) such that $\mathsf{d}_{\mathsf{TV}}(\mu,\nu) \le 1$ for all probability measures μ and ν.

Exercise 5.12 (**)**.** In the setting of Exercise 5.11 show that

$$\mathsf{d}_{\mathsf{TV}}\big(\{p\},\{q\}\big) = \sum_z \big(p_z - \min(p_z,q_z)\big) = \sum_z \big(q_z - \min(p_z,q_z)\big)\,. \tag{5.4}$$

An important relation between coupling and the total variation distance is explained by the following fact.

✍ **Example 5.9** [Maximal Coupling] *Let random variables X and Y be such that $\mathsf{P}(X=x)=p_x$, $x\in\mathcal{X}$, and $\mathsf{P}(Y=y)=q_y$, $y\in\mathcal{Y}$, with $\mathsf{d}_{\mathsf{TV}}\big(\{p\},\{q\}\big)>0$. For each $z\in\mathcal{Z}=\mathcal{X}\cup\mathcal{Y}$, put $r_z=\min\big(p_z,q_z\big)$ and define the joint distribution $\widehat{\mathsf{P}}(\,\cdot\,,\cdot\,)$ in $\mathcal{Z}\times\mathcal{Z}$ via*

$$\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=z\big)=r_z\,,\quad \widehat{\mathsf{P}}\big(\widetilde{X}=x,\widetilde{Y}=y\big)=\frac{\big(p_x-r_x\big)\,\big(q_y-r_y\big)}{\mathsf{d}_{\mathsf{TV}}\big(\{p\},\{q\}\big)}\,,\quad x\neq y\,.$$

Then $\widehat{\mathsf{P}}(\,\cdot\,,\cdot\,)$ is a coupling of X and Y.

Solution. By using (5.4) we see that $\sum_z \widehat{\mathsf{P}}(X=x,Y=z)=r_x+(p_x-r_x)=\mathsf{P}(X=x)$, for all $x\in\mathcal{X}$, ie., the first marginal is as expected. Checking correctness of the Y marginal is similar. □

Remark 5.9.1 By (5.4), we have $\mathsf{d}_{\mathsf{TV}}\big(\{p\},\{q\}\big)=0$ iff $p_z=q_z$ for all z. In this case it is natural to define the maximal coupling via $\widehat{\mathsf{P}}\big(\widetilde{X}=x,\widetilde{Y}=y\big)=r_x\mathbb{1}_{x=y}$ for all $x\in\mathcal{X}$, $y\in\mathcal{Y}$. Notice that in this case the coupling is concentrated on the diagonal $x=y$ and all its off-diagonal terms vanish.

☞ **Example 5.10** *Consider $X\sim\mathsf{Ber}(p_1)$ and $Y\sim\mathsf{Ber}(p_2)$ with $p_1\le p_2$. It is a straightforward exercise to check that the second table in Example 5.1 provides the maximal coupling of X and Y. We notice also that in this case*

$$\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=p_2-p_1=\mathsf{d}_{\mathsf{TV}}(X,Y)\,.$$

Exercise 5.13 (**)**.** If $\widehat{\mathsf{P}}(\,\cdot\,,\cdot\,)$ is the maximal coupling of X and Y as defined in Example 5.9 and Remark 5.9.1, show that $\widehat{\mathsf{P}}(\widetilde{X}\neq\widetilde{Y})=\mathsf{d}_{\mathsf{TV}}\big(X,Y\big)$.

Lemma 5.11 *Let $\widehat{\mathsf{P}}(\,\cdot\,,\cdot\,)$ be the maximal coupling of X and Y as defined in Example 5.9. Then for every other coupling $\widetilde{\mathsf{P}}(\,\cdot\,,\cdot\,)$ of X and Y we have*

$$\widetilde{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)\ge\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=\mathsf{d}_{\mathsf{TV}}\big(X,Y\big)\,. \tag{5.5}$$

Proof. Summing the inequalities $\widetilde{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=z\big)\le\min(p_z,q_z)$ we deduce

$$\widetilde{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)\ge 1-\sum_z\min(p_z,q_z)=\sum_z\big(p_z-\min(p_z,q_z)\big)=\mathsf{d}_{\mathsf{TV}}\big(\{p\},\{q\}\big)\,,$$

where the last equality follows from (5.4). The claim follows from Exercise 5.13. □

Remark 5.11.1 Notice that according to (5.5),

$$\widetilde{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)\le\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)=1-\mathsf{d}_{\mathsf{TV}}\big(\widetilde{X},\widetilde{Y}\big)\,,$$

ie., the probability that $\widetilde{X}=\widetilde{Y}$ is maximised under the optimal coupling $\widehat{\mathsf{P}}(\,\cdot\,,\,\cdot\,)$.

☞ **Example 5.12** *Fix $p\in(0,1)$. Then the maximal coupling of $X\sim\mathsf{Ber}(p)$ and $Y\sim\mathsf{Poi}(p)$ satisfies*

$$\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=0\big)=1-p\,,\qquad \widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=1\big)=pe^{-p}\,,$$

and $\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=x\big)=0$ for all $x>1$, so that

$$\mathsf{d}_{\mathsf{TV}}\big(\widetilde{X},\widetilde{Y}\big)\equiv\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=1-\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)=p\big(1-e^{-p}\big)\le p^2\,. \qquad (5.6)$$

Exercise 5.14 (**). For fixed $p\in(0,1)$, complete the construction of the maximal coupling of $X\sim\mathsf{Ber}(p)$ and $Y\sim\mathsf{Poi}(p)$ as outlined in Example 5.12. Is any of the variables X and Y stochastically dominated by another?

5.2.1 The Law of rare events

The following sub-additive property of total variation distance is important in applications.

Example 5.13 *Let $X=X_1+X_2$ with independent X_1 and X_2. Similarly, let $Y=Y_1+Y_2$ with independent Y_1 and Y_2. Then*

$$\mathsf{d}_{\mathsf{TV}}(X,Y)\le\mathsf{d}_{\mathsf{TV}}(X_1,Y_1)+\mathsf{d}_{\mathsf{TV}}(X_2,Y_2)\,. \qquad (5.7)$$

Solution. By (5.5), for any joint distribution $\mathsf{P}(\,\cdot\,)$ of $\{X_1,X_2,Y_1,Y_2\}$ we have

$$\mathsf{d}_{\mathsf{TV}}(X,Y)\le\mathsf{P}(X\neq Y)\le\mathsf{P}(X_1\neq Y_1)+\mathsf{P}(X_2\neq Y_2)\,.$$

Now let $\widehat{\mathsf{P}}_i(\,\cdot\,,\,\cdot\,)$ be the maximal coupling for the pair (X_i,Y_i), $i=1,2$, and let $\mathsf{P}=\widehat{\mathsf{P}}_1\times\widehat{\mathsf{P}}_2$ be the (independent) product measure[18] of $\widehat{\mathsf{P}}_1$ and $\widehat{\mathsf{P}}_2$. Under such P, the variables X_1 and X_2 (similarly, Y_1 and Y_2) are independent; moreover, the RHS of the display above becomes just $\mathsf{d}_{\mathsf{TV}}(X_1,Y_1)+\mathsf{d}_{\mathsf{TV}}(X_2,Y_2)$, so (5.7) follows. □

Remark 5.13.1 An alternativee proof of (5.7) can be obtained as follows. Let Z be the independent sum Y_1+X_2; by using the explicit formula for $\mathsf{d}_{\mathsf{TV}}(\,\cdot\,,\cdot\,)$ in Exercise 5.11 one can show that $\mathsf{d}_{\mathsf{TV}}(X,Z)=\mathsf{d}_{\mathsf{TV}}(X_1,Y_1)$ and $\mathsf{d}_{\mathsf{TV}}(Z,Y)=\mathsf{d}_{\mathsf{TV}}(X_2,Y_2)$; together with the triangle inequality for the total variation distance, $\mathsf{d}_{\mathsf{TV}}(X,Y)\le\mathsf{d}_{\mathsf{TV}}(X,Z)+\mathsf{d}_{\mathsf{TV}}(Z,Y)$, this gives (5.7).

Theorem 5.14 *Let $X=\sum_{k=1}^n X_k$, where $X_k\sim\mathsf{Ber}(p_k)$ are independent random variables. Let, further, $Y\sim\mathsf{Poi}(\lambda)$, where $\lambda=\sum_{k=1}^n p_k$. Then the maximal coupling of X and Y satisfies*

$$\mathsf{d}_{\mathsf{TV}}\big(X,Y\big)\equiv\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)\le\sum_{k=1}^n(p_k)^2\,.$$

[18] ie., s.t. $\mathsf{P}\big(X_i=a_i,Y_j=b_j\big)=\widehat{\mathsf{P}}_1(X_1=a_1,Y_1=b_1)\cdot\widehat{\mathsf{P}}_1(X_2=a_2,Y_2=b_2)$ for all a_i, b_j.

Proof. Write $Y = \sum_{k=1}^n Y_k$, where $Y_k \sim \mathsf{Poi}(p_k)$ are independent rv's, and use the approach of Example 5.13. Of course,

$$\mathsf{P}\Big(\sum_{k=1}^n X_k \neq \sum_{k=1}^n Y_k\Big) \le \sum_{k=1}^n \mathsf{P}(X_k \neq Y_k)$$

for every joint distribution of $(X_k)_{k=1}^n$ and $(Y_k)_{k=1}^n$. Let $\widehat{\mathsf{P}}_k$ be the maximal coupling for the pair $\{X_k, Y_k\}$, and let $\widehat{\mathsf{P}}_0$ be the maximal coupling for two sums. Notice that the LHS above is not smaller than $\mathsf{d_{TV}}(X,Y) \equiv \widehat{\mathsf{P}}_0\big(\widetilde{X} \neq \widetilde{Y}\big)$; on the other hand, using the (independent) product measure $\mathsf{P} = \widehat{\mathsf{P}}_1 \times \dots \times \widehat{\mathsf{P}}_n$ on the right of the display above we deduce that then the RHS becomes just $\sum_{k=1}^n \widehat{\mathsf{P}}_k\big(\widetilde{X}_k \neq \widetilde{Y}_k\big)$. The result now follows from (5.6). □

Exercise 5.15 (**)**.** Let $X \sim \mathsf{Bin}(n, \frac{\lambda}{n})$ and $Y \sim \mathsf{Poi}(\lambda)$ for some $\lambda > 0$. Show that

$$\frac{1}{2}\big|\mathsf{P}(X=k) - \mathsf{P}(Y=k)\big| \le \mathsf{d_{TV}}\big(\widetilde{X},\widetilde{Y}\big) \le \frac{\lambda^2}{n} \qquad \text{for every } k \ge 0. \qquad (5.8)$$

Deduce that for every fixed $k \ge 0$, we have $\mathsf{P}(X=k) \to \frac{\lambda^k}{k!}e^{-\lambda}$ as $n \to \infty$.

☞ **Remark 5.14.1** By (5.8), if $X \sim \mathsf{Bin}(n,p)$ and $Y \sim \mathsf{Poi}(np)$ then for every $k \ge 0$ the probabilities $\mathsf{P}(X=k)$ and $\mathsf{P}(Y=k)$ differ by at most $2np^2$. Eg., if $n = 10$ and $p = 0.01$ the discrepancy between any pair of such probabilities is bounded above by 0.002, ie., they coincide in the first two decimal places.

5.3 Additional problems

Exercise 5.16 (*)**.** Let random variable X have density $f(\,\cdot\,)$, and let $g(\,\cdot\,)$ be a smooth increasing function with $g(x) \to 0$ as $x \to -\infty$. Show that

$$\mathsf{E}g(X) = \int dx \int \mathbb{1}_{z<x}\, g'(z) f(x)\, dz$$

and deduce the integral representation of $\mathsf{E}g(X)$ in (5.2).

Exercise 5.17 (**)**.** Let $X \sim \mathsf{Bin}(1,p_1)$ and $Y \sim \mathsf{Bin}(2,p_2)$ with $p_1 \le p_2$. By constructing an explicit coupling or otherwise, show that $X \preccurlyeq Y$.

Exercise 5.18 (***)**.** Let $X \sim \mathsf{Bin}(2,p_1)$ and $Y \sim \mathsf{Bin}(2,p_2)$ with $p_1 \le p_2$. Construct an explicit coupling showing that $X \preccurlyeq Y$.

Exercise 5.19 (***)**.** Let $X \sim \mathsf{Bin}(3,p_1)$ and $Y \sim \mathsf{Bin}(3,p_2)$ with $0 < p_1 \le p_2 < 1$. Construct an explicit coupling showing that $X \preccurlyeq Y$.

Exercise 5.20 (***)**.** Let $X \sim \mathsf{Bin}(2,p)$ and $Y \sim \mathsf{Bin}(4,p)$ with $0 < p < 1$. Construct an explicit coupling showing that $X \preccurlyeq Y$.

Exercise 5.21 (**). Let $m \le n$ and $0 < p_1 \le p_2 < 1$. Show that $X \sim \mathsf{Bin}(m, p_1)$ is stochastically smaller than $Y \sim \mathsf{Bin}(n, p_2)$.

Exercise 5.22 (**). Let $X \sim \mathsf{Ber}(p)$ and $Y \sim \mathsf{Poi}(\nu)$. Characterise all pairs (p, ν) such that $X \preccurlyeq Y$. Is it true that $X \preccurlyeq Y$ if $p = \nu$? Is it possible to have $Y \preccurlyeq X$?

Exercise 5.23 (**). Let $X \sim \mathsf{Bin}(n, p)$ and $Y \sim \mathsf{Poi}(\nu)$. Characterise all pairs (p, ν) such that $X \preccurlyeq Y$. Is it true that $X \preccurlyeq Y$ if $np = \nu$? Is it possible to have $Y \preccurlyeq X$?

Exercise 5.24 (**). Prove the additive property of $\mathsf{d}_{\mathsf{TV}}(\cdot, \cdot)$ for independent sums, (5.7), by following the approach in Remark 5.13.1.

Exercise 5.25 (***). Generalise your argument in Exercise 5.24 for general sums, and derive an alternative proof of Theorem 5.14.

Exercise 5.26 (**). If X, Y are stochastically ordered, $X \preccurlyeq Y$, and Z is indepedentent of X and Y, show that $(X + Z) \preccurlyeq (Y + Z)$.

Exercise 5.27 (*). If random variables X, Y, Z satisfy $X \preccurlyeq Y$ and $Y \preccurlyeq Z$, show that $X \preccurlyeq Z$.

Exercise 5.28 (**). Let $X = X_1 + X_2$ with independent X_1 and X_2, similarly, $Y = Y_1 + Y_2$ with independent Y_1 and Y_2. If $X_1 \preccurlyeq Y_1$ and $X_2 \preccurlyeq Y_2$, deduce that $X \preccurlyeq Y$.

Exercise 5.29 (**). Let $X_k \sim \mathsf{Ber}(p')$ and $Y_k \sim \mathsf{Ber}(p'')$ for fixed $0 < p' < p'' < 1$ and all $k = 0,1, \dots, n$. Show that $X = \sum_k X_k$ is stochastically smaller than $Y = \sum_k Y_k$.

6 Some non-classical limits

6.1 Convergence to Poisson distribution

In Sec. 5.2.1 we used coupling to derive the "law of rare events" for sums of independent Bernoulli random variables. An alternative approach is to use generating functions:

Exercise 6.1 (*). Let $X_n \sim \mathsf{Bin}(n,p)$, where $p = p(n)$ is such that $np \to \lambda \in (0,\infty)$ as $n \to \infty$. Show that for every fixed $s \in \mathbb{R}$ the generating function $G_n(s) \stackrel{\mathsf{def}}{=} \mathsf{E}\big(s^{X_n}\big) = \big(1+p(s-1)\big)^n$ converges, as $n \to \infty$, to $G_Y(s) = e^{\lambda(s-1)}$, the generating function of $Y \sim \mathsf{Poi}(\lambda)$. Deduce that the distribution of X_n approaches that of Y in this limit.

A time non-homogeneous version of the previous result can be obtained similarly:

Theorem 6.1 *For $n \ge 1$, let $X_k^{(n)}$, $k = 1,\dots,n$, be independent Bernoulli r.v.'s, $X_n \sim \mathsf{Ber}\big(p_k^{(n)}\big)$. Assume that as $n \to \infty$*

$$\max_{1\le k\le n} p_k^{(n)} \to 0 \qquad \text{and} \qquad \sum_{k=1}^n p_k^{(n)} \equiv \mathsf{E}\Big(\sum_{k=1}^n X_k^{(n)}\Big) \to \lambda \in (0,\infty)\,.$$

Then the distribution of $X_n \stackrel{\mathsf{def}}{=} \sum_{k=1}^n X_k^{(n)}$ converges, as $n \to \infty$, to $\mathsf{Poi}(\lambda)$.

A straightforward proof can be derived in analogy with that in Exercise 6.1, by using the following well-known uniform estimate for the logarithmic function:

Exercise 6.2 (*). Show that $|x + \log(1-x)| \le x^2$ uniformly in $|x| \le 1/2$.

Exercise 6.3 (**). Use the estimate in Exercise 6.2 to prove Theorem 6.1.

We now consider a few examples of such convergence for dependent random variables.

Example 6.2 *For a random permutation π of the set $\{1,2,\dots,n\}$, let S_n be the number of fixed points of π. Then, as $n \to \infty$, the distribution of S_n converges to $\mathsf{Poi}(1)$.*

Solution. If the events $(A_m)_{m=1}^n$ are defined via $A_m = \big\{m \text{ is a fixed point of } \pi\big\}$, then $S_n = \sum_{m=1}^n \mathbb{1}_{A_m}$. Notice that for $1 \le m_1 < m_2 < \dots < m_k \le n$ we have

$$\mathsf{P}\big(A_{m_1} \cap A_{m_2} \cap \dots \cap A_{m_k}\big) = \tfrac{(n-k)!}{n!}\,,$$

since the number of permutations which do not move any k given points is $(n-k)!$. By inclusion-exclusion,

$$\begin{aligned}\mathsf{P}(S_n > 0) \equiv \mathsf{P}\big(\cup_{m=1}^n A_m\big) &= \sum_m \mathsf{P}(A_m) - \sum_{m_1<m_2} \mathsf{P}\big(A_{m_1} \cap A_{m_2}\big) + \dots \\ &= \tbinom{n}{1}\tfrac{(n-1)!}{n!} - \tbinom{n}{2}\tfrac{(n-2)!}{n!} + \dots = \sum_{m=1}^n \tfrac{(-1)^{m-1}}{m!}\,,\end{aligned}$$

so that
$$\big|\mathsf{P}(S_n=0)-e^{-1}\big| \equiv \Big|\sum_{m>n}\tfrac{(-1)^m}{m!}\Big| \le \tfrac{1}{(n+1)!}\big(1-\tfrac{1}{n+2}\big)^{-1},$$
ie., $\mathsf{P}(S_n=0)$ converges to e^{-1} as $n\to\infty$. By considering the locations of all fixed points, we deduce that for every integer $k\ge 0$, as $n\to\infty$,
$$\mathsf{P}(S_n=k)=\tbinom{n}{k}\tfrac{(n-k)!}{n!}\mathsf{P}(S_{n-k}=0)=\tfrac{1}{k!}\mathsf{P}(S_{n-k}=0)\to\tfrac{1}{k!}e^{-1}. \qquad \square$$

The next occupancy problem is very important for applications:

Example 6.3 (Balls and boxes) *Let r (distinguishable) balls be placed randomly into n boxes. Write $N_k=N_k(r,n)$ for the number of boxes containing* exactly *k balls. If $r/n\to c$ as $n\to\infty$, then $\mathsf{E}\big(\frac{1}{n}N_k\big)\to\frac{c^k}{k!}e^{-k}$ and $\mathsf{Var}\big(\frac{1}{n}N_k\big)\to 0$ as $n\to\infty$.*

Solution. We consider the case $k=0$; for general $k\ge 0$, see Exercise 6.4. Notice that $N_0=\sum_{j=1}^n \mathbb{1}_{\mathcal{E}_j}$, where $\mathcal{E}_j=\{\text{box } j \text{ is empty}\}$. We have $\mathsf{E}\big(\mathbb{1}_{\mathcal{E}_j}\big)=\mathsf{P}(\mathcal{E}_j)=\big(1-\frac{1}{n}\big)^r$ and $\mathsf{E}\big(\mathbb{1}_{\mathcal{E}_i}\mathbb{1}_{\mathcal{E}_j}\big)=\mathsf{P}(\mathcal{E}_i\cap\mathcal{E}_j)=\big(1-\frac{2}{n}\big)^r$ with $i\ne j$. Consequently, $\mathsf{E}(N_0)=n\big(1-\frac{1}{n}\big)^r$ and $\mathsf{E}\big((N_0)^2\big)=n\big(1-\frac{1}{n}\big)^r+n(n-1)\big(1-\frac{2}{n}\big)^r$. This finally gives, as $n\to\infty$,
$$\mathsf{E}\big(\tfrac{1}{n}N_0\big)\to e^{-c} \quad\text{and}\quad \mathsf{Var}\big(\tfrac{1}{n}N_0\big)=\big(1-\tfrac{2}{n}\big)^r-\big(1-\tfrac{1}{n}\big)^{2r}+O\big(\tfrac{1}{n}\big)\to 0. \qquad \square$$

Remark 6.3.1 The argument implies that when $\frac{r}{n}\to c$, a typical configuration in the occupancy problem contains a positive proportion of empty boxes.

Exercise 6.4 (**). In the setting of Example 6.3, show that for each integer $k\ge 0$, we have $\mathsf{E}\big(\frac{1}{n}N_k\big)\to\frac{c^k}{k!}e^{-k}$ and $\mathsf{Var}\big(\frac{1}{n}N_k\big)\to 0$ as $n\to\infty$.

Exercise 6.5 (**). In the setup of Example 6.3, let X_j be the number of balls in box j. Show that for each integer $k\ge 0$, we have $\mathsf{P}\big(X_j=k\big)\to\frac{c^k}{k!}e^{-c}$ as $n\to\infty$. Further, show that for $i\ne j$, the variables X_i and X_j are not independent, but in the limit $n\to\infty$ they become independent $\mathsf{Poi}(c)$ distributed random variables.

Exercise 6.6 (**). Generalise the result in Exercise 6.5 for occupancy numbers of a finite number of boxes.

Lemma 6.4 *In the occupancy problem, let $ne^{-r/n}\to\lambda\in[0,\infty)$ as $n\to\infty$. Then the distribution of the number $N_0=N_0(r,n)$ of empty boxes approaches $\mathsf{Poi}(\lambda)$ as $n\to\infty$.*

Remark 6.4.1 The probability that a given box is empty equals $(1-\frac{1}{n})^r\approx e^{-r/n}\approx\frac{\lambda}{n}$. If the numbers of balls in different boxes were independent, the result would follow as in the law of rare events. The following argument shows that the actual dependence is rather weak, and the result still holds.

Remark 6.4.2 Notice that in Exercise 6.4 we considered $r\approx cn$, ie., r was of order n while here r is of order $n\log n$.

Proof of Lemma 6.4 The probability that k fixed boxes are empty is

$$\mathsf{P}\big(\{\text{boxes } m_1, \dots, m_k \text{ are empty}\}\big) = \big(1 - \tfrac{k}{n}\big)^r .$$

Denote $\mathsf{p}_k(r,n) \stackrel{\text{def}}{=} \mathsf{P}\big(\{\text{exactly } k \text{ boxes are empty}\}\big)$; then, by fixing the positions of the empty boxes, we can write

$$\mathsf{p}_k(r,n) = \tbinom{n}{k}\big(1 - \tfrac{k}{n}\big)^r \mathsf{p}_0(r, n-k) . \tag{6.1}$$

It is obviously sufficient to show that under the conditions of the lemma,

$$\tbinom{n}{k}\big(1 - \tfrac{k}{n}\big)^r \to \tfrac{\lambda^k}{k!} , \qquad \mathsf{p}_0(r,n) \to e^{-\lambda} . \tag{6.2}$$

We start by checking the first relation in (6.2). To this end, notice that by the well-known inequality $|x + \log(1-x)| \le x^2$ with $|x| \le 1/2$, see Exercise 6.2, the assumption of the lemma, $\log n - \frac{r}{n} - \log\lambda \to 0$, implies that for all fixed $k \ge 0$ with $2k \le n$ we have

$$\log\big[n^k\big(1 - \tfrac{k}{n}\big)^r\big] - k\log\lambda = k\big[\log n - \tfrac{r}{n} - \log\lambda\ r\big] + r\big[\log\big(1 - \tfrac{k}{n}\big) + \tfrac{k}{n}\ r\big] \to 0 ;$$

equivalently, the first asymptotic relation in (6.2) holds.

On the other hand, by inclusion-exclusion,

$$\mathsf{p}_0(r,n) = 1 - \mathsf{P}\Big(\bigcup_{\ell=1}^{n}\{\text{box } \ell \text{ is empty}\}\Big) = \sum_{k=0}^{n}(-1)^k\tbinom{n}{k}\big(1 - \tfrac{k}{n}\big)^r ,$$

so that by dominated convergence,[19] we get $\mathsf{p}_0(r,n) \to \sum_{k\ge0}\frac{(-\lambda)^k}{k!} = e^{-\lambda}$. □

Exercise 6.7 (**). Suppose that each box of cereal contains one of n different coupons. Assume that the coupon in each box is chosen independently and uniformly at random from the n possibilities, and let T_n be the number of boxes of cereal one needs to buy before there is at least one of every type of coupon. Show that $\mathsf{E}T_n = n\sum_{k=1}^{n} k^{-1} \approx n\log n$ and $\mathsf{Var}T_n \le n^2\sum_{k=1}^{n} k^{-2}$.

Exercise 6.8 (**). In the setup of Exercise 6.7, show that for all real a,

$$\mathsf{P}(T_n \le n\log n + na) \to \exp\{-e^{-a}\} , \qquad \text{as } n \to \infty .$$

[**Hint:** If $T_n > k$, then k balls are placed into n boxes so that at least one box is empty.]

Exercise 6.9 (**). In a village of $365k$ people, what is the probability that all birthdays are represented? Find the answer for $k = 6$ and $k = 5$.
[**Hint:** In notations of Exercise 6.7, the problem is about evaluating $\mathsf{P}(T_{365} \le 365k)$.]

Further examples can be found in [2, Sect. 5].

[19] the domination follows from $|\mathsf{p}_0(r,n)| \le \sum_{k=0}^{n}\binom{n}{k}e^{-rk/n} = (1 + e^{-r/n})^n \le e^{ne^{-r/n}}$, which is bounded, uniformly in r and n under consideration; alternatively, one can follow the approach in Example 6.2.

6.2 Convergence to exponential distribution

Geometric distribution is often referred to as the 'discrete exponential' distribution. Indeed, if $X \sim \mathsf{Geom}(p)$, equivalently, $\mathsf{P}(X > k) = (1-p)^k$ for all integer $k \ge 0$, then the distribution of $Y \stackrel{\text{def}}{=} pX$ satisfies

$$\mathsf{P}(Y > x) = \mathsf{P}\bigl(X > \tfrac{x}{p}\bigr) = (1-p)^{x/p}$$

for all $x \ge 0$, which in the limit $p \to 0$ approaches e^{-x}, the tail probability of $\mathsf{Exp}(1)$ distribution.

Exercise 6.10 (*). Show that the moment generating function (MGF) of $X \sim \mathsf{Geom}(p)$ is given by $M_X(t) \equiv \mathsf{E}e^{tX} = pe^t(1-(1-p)e^t)^{-1}$ and that of $Y \sim \mathsf{Exp}(\lambda)$ is $M_Y(t) = \frac{\lambda}{\lambda - t}$ (defined for all $t < \lambda$). Let $Z = pX$ and deduce that, as $p \to 0$, the MGF $M_Z(t)$ converges to that of $Y \sim \mathsf{Exp}(1)$ for each fixed $t < 1$.

Exercise 6.11 (**). The running cost of a car between two consecutive services is given by a random variable C with moment generating function (MGF) $M_C(t)$ (and expectation c), where the costs over non-overlapping time intervals are assumed independent and identically distributed. The car is written off before the next service with (small) probability $p > 0$. Show that the number T of services before the car is written off follows the 'truncated geometric' distribution with MGF $M_T(t) = p\bigl(1-(1-p)e^t\bigr)^{-1}$ and deduce that the total running cost X of the car up to and including the final service has MGF $M_X(t) = p\bigl(1-(1-p)M_C(t)\bigr)^{-1}$. Find the expectation $\mathsf{E}X$ of X and show that for small enough p the distribution of $X^* \stackrel{\text{def}}{=} X/\mathsf{E}X$ is close to $\mathsf{Exp}(1)$.
[**Hint:** Find the MGF of X^* and follow the approach of Exercise 6.10.]

6.2.1 A hitting time problem

Let $(X_\ell)_{\ell \ge 0}$ be a Markov chain in $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ with jump probabilities

$$p_{0,1} = 1\,, \qquad p_{k,k+1} = p\,, \qquad p_{k,k-1} = q \quad {}^\forall k \ge 1\,,$$

where $0 < p < q < 1$ with $p + q = 1$. Assume $X_0 = 0$ and define

$$\tau_k = \inf\bigl\{\ell \ge 0 : X_\ell = k\bigr\}\,. \tag{6.3}$$

One is interested in studying the moment generating function

$$\varphi_0(\varepsilon) = \mathsf{E}^*_0 e^{\varepsilon \tau_n} \tag{6.4}$$

for fixed (large) $n > 1$ and $\varepsilon > 0$ small enough.

Lemma 6.5 *We have* $\mathsf{E}_0\tau_n = \frac{2pq}{(q-p)^2}\bigl[\bigl(\frac{q}{p}\bigr)^n - 1\bigr] - \frac{n}{q-p}$.

Remark 6.5.1 Since $q > p$, Lemma 6.5 implies that the expected hitting time $\mathsf{E}_0\tau_n$ is exponentially large in $n \ge 1$.

Our key result is as follows.

Theorem 6.6 *Let, as before,* $\varphi_0(\varepsilon) = \mathsf{E}_0 e^{\varepsilon\tau_n}$. *Then for all* $|u| < 1$, *with* $\varepsilon = \frac{u}{\mathsf{E}_0\tau_n}$, *we have*

$$\bar\varphi_0(u) \stackrel{\mathsf{def}}{=} \varphi_0\bigl(\tfrac{u}{\mathsf{E}_0\tau_n}\bigr) \equiv \mathsf{E}_0 \exp\bigl\{u\tfrac{\tau_n}{\mathsf{E}_0\tau_n}\bigr\} \to \tfrac{1}{1-t}$$

as $n \to \infty$. *In particular, the distribution of the rescaled hitting time* $\tau_n^* \equiv \frac{\tau_n}{\mathsf{E}_0\tau_n}$ *is approximately* $\mathsf{Exp}(1)$.

Remark 6.6.1 Notice that the distribution of the rescaled hitting time τ_n^* is supported on the whole half-line $(0,\infty)$, ie., for all fixed $0 < a < A < \infty$, both events $\{\tau_n < a(q/p)^n\}$ and $\{\tau_n > A(q/p)^n\}$ have uniformly positive probability. In other words, despite the variable τ_n is of order $(q/p)^n$, its distribution does not concentrate in any way on this exponential scale, even as $n \to \infty$.

Proof of Lemma 6.5. Write $r_k = \mathsf{E}_k\tau_n$; then $r_n = 0$ while

$$r_0 = 1 + r_1 \qquad \text{and} \qquad r_k = 1 + p\,r_{k+1} + q\,r_{k-1} \quad \text{for } 0 < k < n\,.$$

In terms of the differences $\delta_k = r_{k-1} - r_k$ these recurrence relations reduce to $\delta_1 = 1$ and $\delta_{k+1} = \frac{1}{p} + \frac{q}{p}\delta_k$ with $0 < k < n$. By a straightforward induction,

$$\delta_k = \bigl(\delta_1 + \tfrac{1}{q-p}\bigr)\bigl(\tfrac{q}{p}\bigr)^{k-1} - \tfrac{1}{q-p}\,, \qquad 0 < k \le n\,,$$

so that a direct summation gives

$$r_0 = r_0 - r_n = \delta_1 + \delta_2 + \dots + \delta_n = \tfrac{p}{q-p}\bigl(\delta_1 + \tfrac{1}{q-p}\bigr)\bigl[\bigl(\tfrac{q}{p}\bigr)^n - 1\bigr] - \tfrac{n}{q-p}\,,$$

which for $\delta_1 = 1$ coincides with the claim of the lemma. □

The rest of this section is devoted to the proof of Theorem 6.6. We start by deriving a useful expression for the moment generating function of τ_n. Using the renewal decomposition and the strong Markov property, we get, for $|\varepsilon|$ small enough,

$$\mathsf{E}_0 e^{\varepsilon\tau_n} = e^{\varepsilon}\,\mathsf{E}_1\bigl(e^{\varepsilon\tau_n}\mathbb{1}_{\tau_n<\tau_0}\bigr) + e^{\varepsilon}\,\mathsf{E}_1\bigl(e^{\varepsilon\tau_0}\mathbb{1}_{\tau_0<\tau_n}\bigr)\,\mathsf{E}_0 e^{\varepsilon\tau_n}\,,$$

so that

$$\varphi_0(\varepsilon) = \mathsf{E}_0 e^{\varepsilon\tau_n} = \frac{e^{\varepsilon}\vec g_1}{1 - e^{\varepsilon}\overleftarrow{g}_1}\,, \tag{6.5}$$

where for $0 \le m \le n$ we write

$$\vec g_m = \mathsf{E}_m\bigl(e^{\varepsilon\tau_n}\mathbb{1}_{\tau_n<\tau_0}\bigr) \quad \text{and} \quad \overleftarrow{g}_m = \mathsf{E}_m\bigl(e^{\varepsilon\tau_0}\mathbb{1}_{\tau_0<\tau_n}\bigr) \tag{6.6}$$

for the restricted moment generating functions of the hitting times (6.3). The renewal decomposition (6.5) together with appropriate asymptotics of $\vec g_m$ and $\overleftarrow{g}_m$ is key to the proof of Theorem 6.6.

For ε satisfying $4pqe^{2\varepsilon} < 1$, consider the functions

$$\vec\varphi(x) = pe^{\varepsilon}\bigl(1 - qe^{\varepsilon}x\bigr)^{-1} \qquad \text{and} \qquad \overleftarrow{\varphi}(x) = qe^{\varepsilon}\bigl(1 - pe^{\varepsilon}x\bigr)^{-1}\,,$$

define $\vec{x}$, $\overleftarrow{x}$ via

$$\vec{x} = \min\{x > 0 : \vec{\varphi}(x) = x\} = \tfrac{1-\sqrt{1-4pqe^{2\varepsilon}}}{2qe^{\varepsilon}} = \tfrac{2pe^{\varepsilon}}{1+\sqrt{1-4pqe^{2\varepsilon}}}\,,$$
$$\overleftarrow{x} = \min\{x > 0 : \overleftarrow{\varphi}(x) = x\} = \tfrac{1-\sqrt{1-4pqe^{2\varepsilon}}}{2pe^{\varepsilon}} = \tfrac{2qe^{\varepsilon}}{1+\sqrt{1-4pqe^{2\varepsilon}}}\,,$$

and write

$$\rho \equiv \rho(\varepsilon) \stackrel{\mathsf{def}}{=} \vec{x}\,\overleftarrow{x} = \tfrac{1-\sqrt{1-4pqe^{2\varepsilon}}}{1+\sqrt{1-4pqe^{2\varepsilon}}}\,.$$

Notice that $\rho(\varepsilon)$ is a strictly increasing function of real ε, satisfying $0 < \rho(\varepsilon) < 1$ iff $4pqe^{2\varepsilon} < 1$. Further, because of the expansion

$$\sqrt{1-4pqe^{2\varepsilon}} = (q-p)\bigl(1 - \tfrac{4pq}{1-4pq}(e^{2\varepsilon}-1)\bigr)^{1/2} = (q-p) - \tfrac{4pq\varepsilon}{q-p} + O(\varepsilon^2)\,,$$

valid for small ε, we have

$$\rho(\varepsilon) = \frac{2p + \frac{4pq\varepsilon}{q-p} + O(\varepsilon^2)}{2q - \frac{4pq\varepsilon}{q-p} + O(\varepsilon^2)} = \frac{p}{q}\bigl(1 + \tfrac{2\varepsilon}{q-p} + O(\varepsilon^2)\bigr) \approx \frac{p}{q}\,\exp\bigl\{\tfrac{2\varepsilon}{q-p}\bigr\}\,,$$

where the last equality holds up to an error of order $O(\varepsilon^2)$. We similarly obtain

$$\vec{x} = \Bigl(\frac{p}{q}\rho\Bigr)^{1/2} = \frac{p}{q}\exp\Bigl\{\frac{\varepsilon}{q-p}\Bigr\}\bigl(1 + O(\varepsilon^2)\bigr)\,,$$
$$\overleftarrow{x} = \Bigl(\frac{q}{p}\rho\Bigr)^{1/2} = \exp\Bigl\{\frac{\varepsilon}{q-p}\Bigr\}\bigl(1 + O(\varepsilon^2)\bigr)\,.$$

With the above notation we have the following result.

Lemma 6.7 *For ε small enough, for all m, $0 < m < n$, we have*

$$\vec{g}_m = \frac{1-\rho^m}{1-\rho^n}\,(\vec{x})^{n-m}\,, \qquad \overleftarrow{g}_m = \frac{1-\rho^{n-m}}{1-\rho^n}\,(\overleftarrow{x})^m\,.$$

Remark 6.7.1 Actually, $\vec{x} = \mathsf{E}_0 e^{\varepsilon\tau_n}$ so that the first claim of the lemma can be rewritten as

$$\vec{g}_m \equiv \mathsf{E}_m\bigl(e^{\varepsilon\tau_n}\mathbb{1}_{\tau_n<\tau_0}\bigr) = \tfrac{1-\rho^m}{1-\rho^n}\,\mathsf{E}^*_m e^{\varepsilon\tau_n}\,,$$

indicating the explicit probabilitic price of the "finite interval constraint" $\tau_n < \tau_0$, where the expectation E^*_m is taken over the trajectories of the simple random walk on the whole lattice $\mathbb{Z}$ with right and left jumps having probabilities p and q respectively. A similar conection holds between $\overleftarrow{g}_m$ and $(\overleftarrow{x})^m \equiv \mathsf{E}^*_m e^{\varepsilon\tau_0}$.

We postpone the proof of the lemma and first deduce the claim of Theorem 6.6. Using the last result, we rewrite the representation of the moment generating function (6.5) via

$$\varphi_0(\varepsilon) \equiv \mathsf{E}_0 e^{\varepsilon\tau_n} = \frac{e^{\varepsilon}\vec{g}_1}{1 - e^{\varepsilon}\overleftarrow{g}_1} = \frac{e^{\varepsilon}(1-\rho)\vec{x}^{\,n-1}}{(1-\rho^n) - e^{\varepsilon}\overleftarrow{x}(1-\rho^{n-1})}\,, \tag{6.7}$$

and use the small ε expansion of ρ, $\vec{x}$ and $\overleftarrow{x}$ to derive its asymptotics.

First, it follows from expansions of ρ and $\vec{x}$ that

$$1-\rho = \frac{q-p}{q}\exp\Bigl\{-\frac{2p\varepsilon}{(q-p)^2}\Bigr\}\bigl(1+O(\varepsilon^2)\bigr)\,,$$

$$(\vec{x}\,)^{n-1} = \Bigl(\frac{p}{q}\Bigr)^{n-1}\exp\Bigl\{\frac{(n-1)\varepsilon}{q-p}\Bigr\}\bigl(1+O(n\varepsilon^2)\bigr)$$

so that the numerator in (6.7) is

$$\Bigl(1-\frac{p}{q}\Bigr)\Bigl(\frac{p}{q}\Bigr)^{n-1}\bigl(1+O(n\varepsilon)\bigr)\,.$$

Next, rewrite the denominator in (6.7) as $\rho^{n-1}\bigl(e^{\varepsilon}\overleftarrow{x}-\rho\bigr)-\bigl(e^{\varepsilon}\overleftarrow{x}-1\bigr)$ and notice that

$$e^{\varepsilon}\overleftarrow{x}-1 = e^{\varepsilon+\varepsilon/(q-p)}\bigl(1+O(\varepsilon^2)\bigr)-1 = \frac{2q\varepsilon}{q-p}+O(\varepsilon^2)\,,$$

while $\rho^{n-1}=\bigl(\frac{p}{q}\bigr)^{n-1}\bigl(1+O(n\varepsilon)\bigr)$ and $e^{\varepsilon}\overleftarrow{x}-\rho=\frac{q-p}{q}\bigl(1+O(\varepsilon)\bigr)$. Consequently, the denominator in (6.7) equals

$$\Bigl(1-\frac{p}{q}\Bigr)\Bigl(\frac{p}{q}\Bigr)^{n-1}\bigl(1+O(n\varepsilon)\bigr)-\frac{2q\varepsilon}{q-p}+O(\varepsilon^2)\,.$$

Letting $\varepsilon = u/\mathsf{E}_0\tau_n = O\bigl(u(p/q)^n\bigr)$, we get

$$\bar{\varphi}_n(u) = \frac{1+O(n\varepsilon)}{1-\frac{2pq}{(q-p)^2}\bigl(\frac{q}{p}\bigr)^n\frac{u}{\mathsf{E}_0\tau_n}+O(n\varepsilon)} = \frac{1+O(n\varepsilon)}{1-u+O(n\varepsilon)}\,,$$

and the result follows.

Proof of Lemma 6.7. We only prove the first equality,

$$\vec{g}_m \equiv \mathsf{E}_m\bigl(e^{\varepsilon\tau_n}\mathbb{1}_{\tau_n<\tau_0}\bigr) = \frac{1-\rho^m}{1-\rho^n}\,(\vec{x}\,)^{n-m}\,,$$

the other can be verified similarly. Let $\vec{g}_m$ be defined as in (6.6); then $\vec{g}_0=0$, $\vec{g}_n=1$, and the first step decomposition gives

$$\vec{g}_m = e^{\varepsilon}\bigl(p\vec{g}_{m+1}+q\vec{g}_{m-1}\bigr)\qquad\text{for } 0<m<n\,. \tag{6.8}$$

Notice that these equations have a unique finite solution, specified in terms of $\vec{g}_1$ (equivalently, in terms of $\vec{g}_n$). Hence, if we show that $\frac{1-\rho^m}{1-\rho^n}\,(\vec{x}\,)^{n-m}$ solve equations (6.8), it must coincide with $\vec{g}_m$.

This, however, is straightforward, since the definitions of $\vec{x}$, $\overleftarrow{x}$, and ρ imply that

$$pe^{\varepsilon}\,\frac{\rho}{\vec{x}}+qe^{\varepsilon}\,\frac{\vec{x}}{\rho}=1\,,\qquad pe^{\varepsilon}\,\frac{1}{\vec{x}}+qe^{\varepsilon}\vec{x}=1\,,$$

and thus that $\frac{1-\rho^m}{1-\rho^n}\,(\vec{x}\,)^{n-m}$ solve the equations (6.8). □

Remark 6.7.2 A suitable modification of the argument above can be used to prove the result Theorem 6.6 for a different starting point and/or different reflection mechanism at the origin, see Exercises 6.17–6.19.

6.3 Additional problems

Exercise 6.12 (**). Let r (distinguishable) balls be placed randomly into n boxes. Find a constant $c > 0$ such that with $r = c\sqrt{n}$ the probability that no two such balls are placed in the same box approaches $1/e$ as $n \to \infty$.
Find a constant $b > 0$ such that with $r = b\sqrt{n}$ the probability that no two such balls are placed in the same box approaches $1/2$ as $n \to \infty$. Notice that with $n = 365$ and $r = 23$ this is the famous 'birthday paradox'.

Exercise 6.13 (***). In the setup of Example 6.3, let X_j be the number of balls in box j. Show that for every integer $k_j \geq 0$, $j = 1, \ldots, n$, satisfying $\sum_{j=1}^{n} k_j = r$ we have

$$\mathsf{P}\big(X_1 = k_1, \ldots, X_n = k_n\big) = \binom{r}{k_1; k_2; \ldots; k_n} n^{-r} = \frac{r!}{k_1! k_2! \ldots k_n! n^r} \,.$$

Now suppose that $Y_j \sim \mathsf{Poi}\big(\frac{r}{n}\big)$, $j = 1, \ldots, n$, are independent. Show that the conditional probability $\mathsf{P}\big(Y_1 = k_j, \ldots, Y_n = k_n \mid \sum_j Y_j = r\big)$ is given by the expression in the last display.

Exercise 6.14 (***). Find the probability that in a class of 100 students at least three of them have the same birthday.

Exercise 6.15 (****). There are 365 students registered for the first year probability class. Find k such that the probability of finding at least k students sharing the same birthday is about $1/2$.

Exercise 6.16 (***). Let $G_{n,N}$, $N \leq \binom{n}{2}$, be the collection of all graphs on n vertices connected by N edges. For a fixed constant $c \in \mathbb{R}$, if $N = \frac{1}{2}(n \log n + cn)$, show that the probability for a random graph in $G_{n,N}$ to have isolated vertices approaches $\exp\{-e^{-c}\}$ as $n \to \infty$.

Exercise 6.17 (****). Generalise Theorem 6.6 to a general starting point, ie., find the limit of $\varphi_m(\varepsilon) = \mathsf{E}_m e^{\varepsilon \tau_n}$ with $\varepsilon = u/\mathsf{E}_m \tau_n$.

Exercise 6.18 (****). Generalise Theorem 6.6 to a general reflection mechanism at the origin, ie., assuming that $p_{0,1} = 1 - p_{0,0} > 0$.

Exercise 6.19 (****). Generalise Theorem 6.6 to a general reflection mechanism at the origin and a general starting point, cf. Exercise 6.17 and Exercise 6.18.

7 Introduction to percolation

Let $G = (V, E)$ be an (infinite) graph. Suppose the state of each edge $e \in E$ is encoded by the value $\omega_e \in \{0,1\}$, where $\omega_e = 0$ means 'edge e is closed' and $\omega_e = 1$ means 'edge e is open'. Once the state ω_e of each edge $e \in E$ is specified, the configuration $\boldsymbol{\omega} = (\omega_e)_{e\in E} \in \Omega \stackrel{\text{def}}{=} \{0,1\}^E$ describes the state of the whole system. The classical Bernoulli bond percolation studies properties of random configurations $\boldsymbol{\omega} \in \Omega$ under the assumption that $\omega_e \sim \mathsf{Ber}(p)$ for some fixed $p \in [0,1]$, independently for disjoint edges $e \in E$. Write $\mathbb{P}_p$ for the corresponding (product) measure in Ω.

Example 7.1 *Turn the integer lattice $\mathbb{Z}^d$, $d \ge 1$, into a graph $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$, where $\mathbb{E}^d$ contains* all *edges connecting nearest neighbour vertices in $\mathbb{Z}^d$ (ie., those at distance 1). The above construction introduces the Bernoulli bond percolation model on $\mathbb{L}^d$.*

Given a configuration $\boldsymbol{\omega} \in \Omega$, we say that vertices $\mathbf{x}$ and $\mathbf{y} \in V$ are connected (written '$\mathbf{x} \leftrightsquigarrow \mathbf{y}$') if $\boldsymbol{\omega}$ contains a path of open edges connecting $\mathbf{x}$ and $\mathbf{y}$. Then the cluster $C_{\mathbf{x}}$ of $\mathbf{x}$ is $C_{\mathbf{x}} = \big\{\mathbf{y} \in V : \mathbf{y} \leftrightsquigarrow \mathbf{x}\big\}$. Let $|C_{\mathbf{x}}|$ be the number of vertices in the open cluster at $\mathbf{x}$. Then

$$\big\{\mathbf{x} \leftrightsquigarrow \infty\big\} \equiv \big\{|C_{\mathbf{x}}| = \infty\big\}$$

is the event 'vertex $\mathbf{x}$ is connected to infinity', and the percolation probability is

$$\theta_{\mathbf{x}}(p) = \mathbb{P}_p\big(|C_{\mathbf{x}}| = \infty\big) = \mathbb{P}_p\big(\mathbf{x} \leftrightsquigarrow \infty\big)\,.$$

One of the key questions in percolation is whether $\theta_{\mathbf{x}}(p) > 0$ or $\theta_{\mathbf{x}}(p) = 0$.

7.1 Bond percolation in $\mathbb{Z}^d$

For the bond percolation in $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ the percolation probability $\theta_{\mathbf{x}}(p)$ does not depend on $\mathbf{x} \in \mathbb{Z}^d$. One thus writes

$$\theta(p) = \mathbb{P}_p\big(|C_{\mathbf{0}}| = \infty\big) = \mathbb{P}_p\big(\mathbf{0} \leftrightsquigarrow \infty\big)\,; \tag{7.1}$$

this is also known as the order parameter.

Lemma 7.2 *The order parameter $\theta(p)$ of the bond percolation model is a non-decreasing function $\theta : [0,1] \to [0,1]$ with $\theta(0) = 0$ and $\theta(1) = 1$.*

Remark 7.2.1 As will be seen below, a similar monotonicity of the order parameter $\theta(p)$ holds for other percolation models.

By the monotonicity result of Lemma 7.2, the threshold (or critical) value

$$p_{\mathsf{cr}} = \inf\big\{p : \theta(p) > 0\big\} \tag{7.2}$$

is well defined. Clearly, $\theta(p) = 0$ for $p < p_{\mathsf{cr}}$ while $\theta(p) > 0$ for $p > p_{\mathsf{cr}}$; this property is often referred to as the 'phase transition' for the percolation model under consideration. Whether $\theta(p_{\mathsf{cr}}) = 0$ is in general a (difficult) open problem.

Proof of Lemma 7.2. The idea of the argument is due to Hammersley. Let $\{\xi_e\}_{e\in E}$ be independent with $\xi_e \sim \mathcal{U}[0,1]$ for each $e \in E$. Fix $p \in [0,1]$; then

$$\omega_e \stackrel{\mathsf{def}}{=} \mathbb{1}_{\xi_e \le p} \sim \mathsf{Ber}(p)$$

are independent for different $e \in E$. We interpret $\omega_e = \omega_e^p$ as the indicator of the event $\{\text{edge } e \text{ is open}\}$.

For $0 \le p' \le p'' \le 1$, this construction gives $\omega_e^{p'} \le \omega_e^{p''}$ for all $e \in E$, so that

$$\theta(p') \equiv \mathbb{P}_{p'}(\mathbf{0} \leftrightsquigarrow \infty) \le \mathbb{P}_{p''}(\mathbf{0} \leftrightsquigarrow \infty) \equiv \theta(p'')\,.$$

Since obviously $\theta(0) = 0$ and $\theta(1) = 1$, this finishes the proof. □

Example 7.3 *For bond percolation on $\mathbb{L}^1$ we have $p_{\mathsf{cr}} = 1$.*

Solution. Indeed, $\{\mathbf{0} \leftrightsquigarrow \infty\} = \mathcal{A}^+ \cup \mathcal{A}^-$, where $\mathcal{A}^\pm \equiv \{\mathbf{0} \leftrightsquigarrow \pm\infty\}$. For $p < 1$ we have $\mathbb{P}_p(\mathbf{0} \leftrightsquigarrow +\infty) \le \mathbb{P}_p(\mathbf{0} \leftrightsquigarrow n) \le p^n \to 0$ as $n \to \infty$, implying that $\mathbb{P}_p(\mathcal{A}^+) = 0$. Since similarly $\mathbb{P}_p(\mathcal{A}^-) = 0$, we deduce that $\theta(p) = 0$ for all $p < 1$. □

As the next result claims, the phase transition in the Bernoulli bond percolation model on $\mathbb{L}^d$ is non-trivial in dimension $d > 1$.

Theorem 7.4 *For $d > 1$, the critical probability $p_{\mathsf{cr}} = p_{\mathsf{cr}}^{(d)}$ of the bond percolation on $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ satisfies $0 < p_{\mathsf{cr}} < 1$.*

Exercise 7.1 (*). In the setting of Theorem 7.4, show that $p_{\mathsf{cr}}^{(d)} \le p_{\mathsf{cr}}^{(d-1)}$ for all integer $d > 1$.

Exercise 7.2 (*). Let $G = (V, E)$ be a graph of uniformly bounded degree $\deg(v) \le r$, $v \in V$. Show that the number $a_v(n)$ of self-avoiding paths of n jumps starting at v is bounded above by $r(r-1)^{n-1}$.

Exercise 7.3 (**). For an infinite planar graph $G = (V, E)$, its dual $G^* = (V^*, E^*)$ is defined by placing a vertex in each face of G, with vertices u^* and $v^* \in V^*$ connected by a dual edge if and only if the faces corresponding to u^* and v^* share a common boundary edge in G. If G^* has a uniformly bounded degree r^*, show that the number $c_v(n)$ of dual self-avoiding contours of n jumps around $v \in V$ is bounded above by $nr^*(r^*-1)^{n-1}$.

Exercise 7.4 (**). For fixed $A > 0$ and $r \ge 1$ find $x > 0$ small enough so that $A\sum_{n\ge1} nr^n x^n < 1$.

By the result of Exercise 7.1, the claim of Theorem 7.4 follows immediately from the following two observations.

Lemma 7.5 *For each integer $d > 1$ we have $p_{\mathsf{cr}}^{(d)} \ge \frac{1}{2d-1} > 0$.*

Lemma 7.6 *There exists $p < 1$ such that $p_{\mathsf{cr}}^{(2)} \le p < 1$.*

Proof of Lemma 7.5. For integer $n \ge 1$, let $\mathcal{A}_n$ be the event 'there is an open self-avoiding path of n edges starting at $\mathbf{0}$'. Using the result of Exercise 7.2, we get

$$\theta(p) \le \mathbb{P}_p\big(\mathcal{A}_n\big) \le a_{\mathbf{0}}(n)p^n \le \tfrac{2d}{2d-1}\big((2d-1)p\big)^n \to 0$$

as $n \to \infty$ if only $(2d-1)p < 1$. Hence the result. $\square$

Proof of Lemma 7.6. The idea of the argument is due to Peierls. We start by noticing that by construction of Exercise 7.3 the square lattice is self-dual (where the bold edges belong to $\mathbb{L}^2$ and the thin edges belong to its dual):

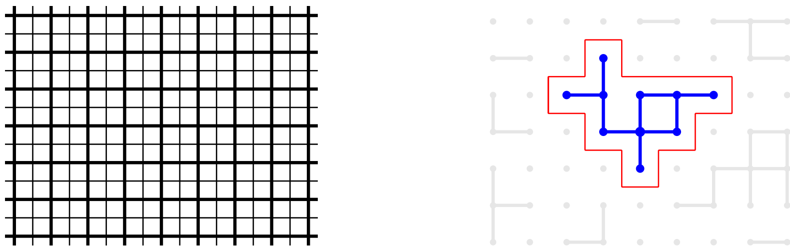

Figure 1: Left: self-duality of the square lattice $\mathbb{Z}^2$. Right: an open dual contour (thin, red online) separating the open cluster at $\mathbf{0}$ (thick, blue online) from ∞.

If the event $\{\mathbf{0} \leftrightsquigarrow \infty\}$ does not occur, there must be a dual contour separating $\mathbf{0}$ from infinity, each (dual) edge of which is open (with probability $1-p$). Let $\mathcal{C}_n$ be the event 'there is an open dual contour of n edges around the origin'. By Exercise 7.3, we have $\mathbb{P}_p\big(\mathcal{C}_n\big) \le c_{\mathbf{0}}(n)\,(1-p)^n \le \frac{4}{3}n\big(3(1-p)\big)^n$ for all $n \ge 1$. Using the subadditive bound in $\big\{\mathbf{0} \leftrightsquigarrow \infty\big\}^{\mathsf{c}} \subseteq \cup_{n\ge1}\mathcal{C}_n$ together with the estimate of Exercise 7.4, we deduce that

$$1-\theta(p) \equiv \mathbb{P}_p\big(\{\mathbf{0} \leftrightsquigarrow \infty\}^{\mathsf{c}}\big) < 1$$

if only $3(1-p) > 0$ is small enough. Hence, there is $p' < 1$ such that $\theta(p) > 0$ for all $p \in (p', 1]$, implying that $p_{\mathsf{cr}}^{(2)} \le p' < 1$, as claimed. $\square$

Further examples of dual graphs are in Fig. 2 below.

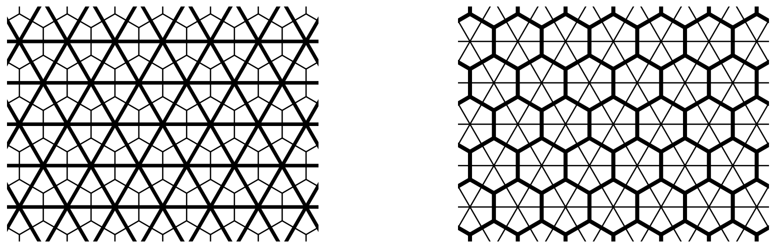

Figure 2: Left: triangular lattice (bold) and its dual hexagonal lattice (thin); right: hexagonal lattice (bold) and its dual triangular lattice (thin).

Exercise 7.5 (**)**.** For bond percolation on the triangular lattice (see Fig. 2, left), show that its critical value p_{cr} is non-trivial, $0 < p_{\mathsf{cr}} < 1$.

Exercise 7.6 (**)**.** For bond percolation on the hexagonal (or honeycomb) lattice (see Fig. 2, right), show that its critical value p_{cr} is non-trivial, $0 < p_{\mathsf{cr}} < 1$.

7.2 Bond percolation on trees

Percolation on regular trees is easier to understand due to their symmetries. We start by considering bond percolation on a rooted tree R_d of index $d > 2$, in which every vertex different from the root has exactly d neighbours; see Fig. 3 showing a finite part of the case $d = 3$.

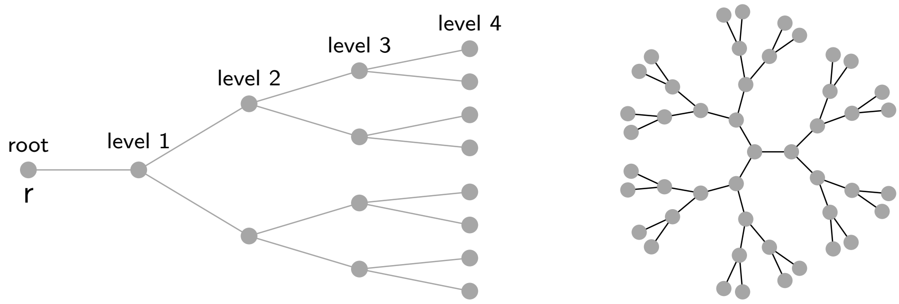

Figure 3: A rooted tree R_3 (left) and a homogeneous tree T_3 (right) up to level 4.

Fix $p \in (0, 1)$ and let every edge in R_d be open with probability p, independently of states of all other edges. Further, write $\theta_{\mathsf{r}}(p) = \mathbb{P}_p(\mathsf{r} \leftrightsquigarrow \infty)$ for the percolation probability of the bond model on R_d. We have

Lemma 7.7 *The critical bond percolation probability on* R_d *is* $p_{\mathsf{cr}} = 1/(d-1)$.

Exercise 7.7 (**). Consider the function $f(x) = (1 - p) + px^{d-1}$. Show that $f(\,\cdot\,)$ is increasing and convex for $x \geq 0$ with $f(1) = 1$. Further, let $0 < x_* \leq 1$ be the smallest positive solution to the fixed point equation $x = f(x)$. Show that $x_* < 1$ if and only if $(d - 1)p > 1$.

Proof. Let $\rho_n = \mathbb{P}_p\big(\{\mathsf{r} \leftrightsquigarrow \text{ level } n\}^{\mathsf{c}}\big)$ be the probability that there is no open path connecting the root r to a vertex at level n. We obviously have $\rho_1 = 1 - p$, and for $k > 1$ the total probability formula gives

$$\rho_k = (1 - p) + p\big(\rho_{k-1}\big)^{d-1} \equiv f(\rho_{k-1})\,,$$

where $f(x) = (1 - p) + px^{d-1}$ is the function from Exercise 7.7. Notice that

$$0 < \rho_1 = 1 - p < \rho_2 = f(\rho_1) < 1\,,$$

which by a straightforward induction implies that $(\rho_n)_{n\geq 1}$ is an increasing sequence of real numbers in $(0, 1)$. Consequently, ρ_n converges to a limit $\bar\rho \in (0, 1]$ such that $\bar\rho = f(\bar\rho)$. Clearly, $\bar\rho$ is the smallest positive solution to the fixed point equation $x = f(x)$. By Exercise 7.7,

$$\bar\rho = \lim_{n\to\infty} \rho_n \equiv 1 - \theta_{\mathsf{r}}(p)$$

is smaller than 1 (equivalently, $\theta_{\mathsf{r}}(p) > 0$) iff $(d - 1)p < 1$. □

An infinite homogeneous tree T_d of index $d > 2$ is a tree every vertex of which has index d. One can think of T_d as a union of d rooted trees with common root, see Fig. 3 (right).

Exercise 7.8 (**). Show that the critical bond percolation probability on T_d is $p_{\mathsf{cr}} = 1/(d - 1)$.

7.3 Directed percolation

The directed (or oriented) percolation is defined on the graph $\vec{\mathbb{L}}^d$ which is the restriction of $\mathbb{L}^d$ to the subgraph whose vertices have non-negative coordinates while the inherited edges are oriented in the increasing direction of their coordinates. In two dimensions $\vec{\mathbb{L}}^2$ thus becomes the 'north-east' graph, see Fig. 4.

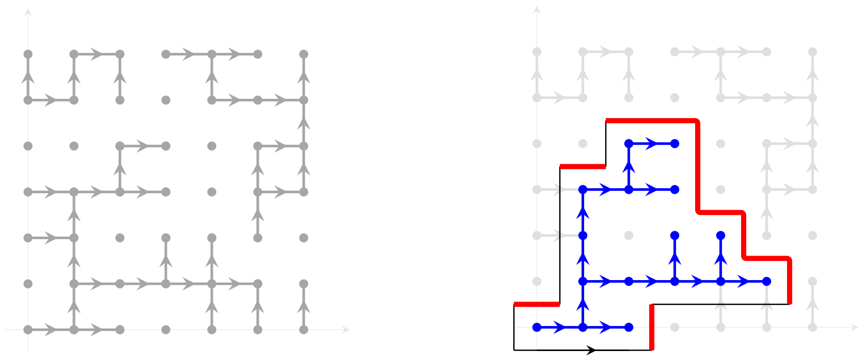

Figure 4: Left: a finite directed percolation configuration in $\vec{\mathbb{L}}^2$. Right: its cluster at the origin (thick, blue online) and the corresponding separating countour; blocking dual edges are thick (red online).

For fixed $p \in [0,1]$ declare each edge in $\vec{\mathbb{L}}^d$ open with probability p (and closed otherwise), independently for all edges in $\vec{\mathbb{L}}^d$. Let $\vec{C}_0$ be the collection of all vertices that may be reached from the origin along paths of open bonds (the blue cluster in Fig. 4). We define the percolation probability of the oriented model by

$$\vec{\theta}(p) = \mathbb{P}_p\big(|\,\vec{C}_0\,| = \infty\big)$$

and the corresponding critical value by

$$\vec{p}_{\mathsf{cr}}(d) = \inf\big\{p : \vec{\theta}(p) > 0\big\}\,.$$

As for the bond percolation, one can show that $\vec{\theta}(p)$ is a non-decreasing function of $p \in [0,1]$, see Exercise 7.9; hence, $\vec{p}_{\mathsf{cr}}(d)$ is well defined.

Exercise 7.9 (**)**.** Show that $\vec{\theta}: [0,1] \to [0,1]$ is a non-decreasing function with $\vec{\theta}(0) = 0$ and $\vec{\theta}(1) = 1$.

As for the bond percolation on $\mathbb{L}^d$, the phase transition for the oriented percolation on $\vec{\mathbb{L}}^d$ is non-trivial:

Theorem 7.8 *For $d > 1$, we have $0 < \vec{p}_{\mathsf{cr}}(d) < 1$.*

Exercise 7.10 (**)**.** Show that for all $d \ge 1$, we have $p_{\mathsf{cr}}(d) \le \vec{p}_{\mathsf{cr}}(d)$.

Exercise 7.11 (**)**.** Show that for all $d > 1$, we have $\vec{p}_{\mathsf{cr}}(d) \le \vec{p}_{\mathsf{cr}}(d-1)$.

Exercise 7.12 (**)**.** Use the path counting idea of Lemma 7.5 to show that for all $d > 1$ we have $\vec{p}_{\mathsf{cr}}(d) > 0$.

In view of Exsercies 7.10–7.12, we have $0 < \vec{p}_{\mathsf{cr}}(d) \le \vec{p}_{\mathsf{cr}}(2) \le 1$. To finish the proof of Theorem 7.8, it remains to show that $\vec{p}_{\mathsf{cr}}(2) < 1$.

As in the (unoriented) bond case, if the cluster $\vec{C}_0$ is finite, there exists a dual contour separating it from infinity, see Fig. 4; we orient it in the anticlockwise direction. Notice that each (dual) bond of the separating contour going northwards or westwards intersects a (closed) bond of the original model. Such blocking dual edges are thick (red online) in Fig. 4; it is easy to see that each closing contour of length $2n$ has exactly n such bonds (indeed, since the contour is closed, it must have equal numbers of eastwards and westwards going bonds; similarly, for vertical edges).

Let $\vec{\mathcal{C}}_n$ be the event that there is an open dual contour of length n separating the origin from infinity. Notice that $\bigl\{|\,\vec{C}_0\,| < \infty\bigr\} \subset \cup_{n\ge1} \vec{\mathcal{C}}_n$.

Exercise 7.13 (**)**.** Let $\vec{c}_0(k)$ be the number of dual contours of length k around the origin separating the latter from infinity. Show that $\vec{c}_0(k) \le k3^{k-1}$.

The standard subadditive bound now implies

$$1 - \vec{\theta}(p) = \mathbb{P}_p\bigl(|\,\vec{C}_0\,| < \infty\bigr) \le \sum_{n\ge1} \vec{c}_0(2n)\,(1-p)^n\,,$$

as each separating dual contour has an even number of bonds and exactly half of them are necessarily open (with probability $1-p$). As in the proof of Lemma 7.6, we deduce that the last sum is smaller than 1 for some $p' < 1$. This implies that $\vec{p}_{\mathsf{cr}}(2) \le p' < 1$, as claimed. This finishes the proof of Theorem 7.8.

7.4 Site percolation in $\mathbb{Z}^d$

In this section we generalise the previous ideas to site percolation in $\mathbb{Z}^d$; it is entirely optional and will not be examined.

The site percolation in $\mathbb{Z}^d$ can be defined similarly. Given $p \in [0,1]$, the vertices are declared '`open`' with probability p, independently of each other. Once all non-'`open`' vertices are removed together with their edges, the configuration decomposes into clusters, similarly to the bond model. As before, one is interested in

$$\theta^{\mathsf{site}}(p) \equiv \mathbb{P}_p\bigl(\mathbf{0} \overset{\mathsf{site}}{\rightsquigarrow} \infty\bigr)\,,$$

the probability that the cluster at the origin is infinite. A similar global coupling shows that the function $\theta^{\mathsf{site}}(p)$ is non-decreasing in $p \in [0,1]$, so it is natural to define the critical value for the site percolation via

$$p^{\mathsf{site}}_{\mathsf{cr}} = \inf\bigl\{p : \theta^{\mathsf{site}}(p) > 0\bigr\}\,. \tag{7.3}$$

The phase transition is again non-trivial in dimension $d > 1$:

Theorem 7.9 *Let $p^{d,\mathsf{site}}_{\mathsf{cr}}$ be the critical value of the site percolation in $\mathbb{Z}^d$ as defined in (7.3). We then have $0 < p^{d,\mathsf{site}}_{\mathsf{cr}} < 1$ for all $d > 1$.*

The argument for the lower bound is easy:

Exercise 7.14 (**). Show that $p_{\mathsf{cr}}^{d,\mathsf{site}} \le p_{\mathsf{cr}}^{d-1,\mathsf{site}}$ for all $d > 1$, while $p_{\mathsf{cr}}^{d,\mathsf{site}} \ge \frac{1}{2d-1}$ for fixed $d > 1$, cf. Lemma 7.5.

To finish the proof of Theorem 7.9 we just need to show that $p_{\mathsf{cr}}^{2,\mathsf{site}} < 1$. To this end, notice that if an open cluster $C_{\mathbf{0}} \subset \mathbb{Z}^2$ of the origin is finite, its external boundary,

$$\partial^{\mathsf{ext}} C_{\mathbf{0}} \stackrel{\mathsf{def}}{=} \{\mathbf{y} \in \mathbb{Z}^2 \setminus C_{\mathbf{0}} : {}^{\exists}\mathbf{x} \in C_{\mathbf{0}} \text{ with } \mathbf{x} \sim \mathbf{y}\},$$

where $\mathbf{x}$, $\mathbf{y} \in \mathbb{Z}^2$ are neigbours (denoted $\mathbf{x} \sim \mathbf{y}$) if they are at Euclidean distance one (equivalently, connected by an edge in $\mathbb{L}^2$).

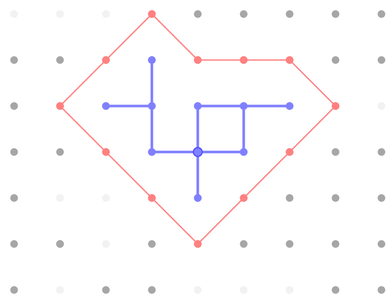

Figure 5: Site cluster at the origin (thick, blue online) and its separating contour (thin, red online).

Argueing as above, it is easy to see that the number $c_{\mathbf{0}}(n)$ of contours of length n separating the origin from infinity is not bigger than $5n \cdot 7^{n-1}$. Consequently, the event $\mathcal{C}_n$ that 'there is a separating contour of n sites around the origin' has probability $\mathbb{P}_p(\mathcal{C}_n) \le c_{\mathbf{0}}(n)(1-p)^n \le \frac{5n}{7}\big(7(1-p)\big)^n$, so that

$$1 - \theta^{\mathsf{site}}(p) \equiv \mathbb{P}_p\big(\{\mathbf{0} \stackrel{\mathsf{site}}{\leftrightsquigarrow} \infty\}^{\mathsf{c}}\big) < 1$$

for all $(1-p) > 0$ small enough. This implies that $p_{\mathsf{cr}}^{2,\mathsf{site}} < 1$, and thus finishes the proof of Theorem 7.9.

Exercise 7.15 (**). Carefully show that the number $c_{\mathbf{0}}(n)$ of contours of length n separating the origin from infinity is not bigger than $5n \cdot 7^{n-1}$.

7.5 Additional problems

Exercise 7.16 (**). Show that the critical site percolation probability on R_d is $p_{\mathsf{cr}} = 1/(d-1)$.

Exercise 7.17 (**). Show that the critical site percolation probability on T_d is $p_{\mathsf{cr}} = 1/(d-1)$.

Exercise 7.18 (****). In the setting of Exercise 7.2, let $b_v(n)$ be the number of connected subgraphs of G on exactly n vertices containing $v \in V$. Show that for some positive A and R, $b_v(n) \le AR^n$.

Exercise 7.19 (****). Consider the Bernoulli bond percolation model on an infinite graph $G = (V, E)$ of uniformly bounded degree, cf. Exercise 7.2. If $p > 0$ is small enough, show that for each $v \in V$ the distribution of the cluster size $|C_v|$ has exponential moments in a neighbourhood of the origin.

For many percolation models, the property in Exercise 7.19 is known to hold for all $p < p_{\mathsf{cr}}$.

References

[1] Allan Gut. *An intermediate course in probability.* Springer Texts in Statistics. Springer, New York, second edition, 2009.

[2] Michael Mitzenmacher and Eli Upfal. *Probability and computing.* Cambridge University Press, Cambridge, 2005. Randomized algorithms and probabilistic analysis.