

Order statistics

Ex. 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 $\mathbb{P}(X_{(1)} < x) \rightarrow 1$ and $\mathbb{P}(X_{(n)} > x) \rightarrow 1$ as $n \rightarrow \infty$.

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

$$\mathbb{P}(X_{(k)} \leq x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} y^{k-1}(1-y)^{n-k} dy.$$

Ex. 4.3 (*). Derive the density $f_{X_{(k)}}(x)$ from (4.7) by differentiating (4.6).

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

Ex. 4.5 (**). In the case of an n -sample from $\mathcal{U}(0, 1)$ distribution, derive (4.8),

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)(F(y) - F(x))^{n-2} f(x)f(y) \mathbb{1}_{x < y},$$

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.

Ex. 4.6 (*). Prove the density $f_{R_n}(r)$ formula (4.9), $f_{R_n}(r) = n(n-1) \int (F(z+r) - F(z))^{n-2} f(z)f(z+r) dz$.

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

Ex. 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 $\mathbb{P}(X_{(1)} > \frac{y}{n}) \approx e^{-cy}$ for large n . Deduce that the distribution of $Y_n \equiv nX_{(1)}$ is approximately $\text{Exp}(c)$ for large enough n .

Ex. 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 $\mathbb{P}(X_{(1)} > \frac{y}{n}) \approx e^{-cy}$ for large n . Deduce that the distribution of $Y_n \stackrel{\text{def}}{=} nX_{(1)}$ is approximately $\text{Exp}(c)$ for large enough n .

Ex. 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 $1 - X_{(3)}$.

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

Ex. 4.12 (***). Using induction or otherwise, prove (4.13), describing the joint gaps distribution:

$$\mathbb{P}(\Delta_{(1)}X \geq r_1, \dots, \Delta_{(n+1)}X \geq r_{n+1}) = \left(1 - \sum_{k=1}^{n+1} r_k\right)^n,$$

for all positive r_k satisfying $\sum_{k=1}^{n+1} r_k \leq 1$.

Ex. 4.13 (*). Let $X_k \sim \text{Exp}(\lambda_k)$, $k = 1, \dots, n$, 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 $y \geq 0$ we have $\mathbb{P}(X_0 > y, X_0 = X_k) = e^{-\lambda y} \frac{\lambda_k}{\lambda_0}$, ie., the minimum X_0 of the sample satisfies $X_0 \sim \text{Exp}(\lambda_0)$ and the probability that it coincides with X_k is proportional to λ_k , independently of the value of X_0 .

Ex. 4.14 (*). If $X \sim \text{Exp}(\lambda)$, show that for all positive a and b we have $\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b)$.

Ex. 4.15 (**). If $X \sim \text{Exp}(\lambda)$, $Y \sim \text{Exp}(\mu)$ and $Z \sim \text{Exp}(\nu)$ are independent, show that for every constant $a \geq 0$ we have $\mathbb{P}(a + X < Y \mid a < Y) = \frac{\lambda}{\lambda + \mu}$; deduce that $\mathbb{P}(a + X < \min(Y, Z) \mid a < \min(Y, Z)) = \frac{\lambda}{\lambda + \mu + \nu}$.

Ex. 4.16 (**). Carefully prove Corollary 4.16 and compute $\mathbb{E}Y_n$ and $\text{Var}Y_n$.

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

Ex. 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.

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

Ex. 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 $X_{(2)}$.

Ex. 4.21 (*). Let X_1 and X_2 be independent $\text{Exp}(\lambda)$ random variables.

a) Show that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent and find their distributions.

b) Compute $\mathbb{E}(X_{(2)} | X_{(1)} = x_1)$ and $\mathbb{E}(X_{(1)} | X_{(2)} = x_2)$.

Ex. 4.22 (**). Let $(X_k)_{k=1}^n$ be an n -sample from $\text{Exp}(\lambda)$ distribution.

a) Show that the gaps $(\Delta_{(k)} X)_{k=1}^n$ as defined in Lemma 4.10 are independent and find their distribution.

b) For fixed $1 \leq k \leq m \leq n$, compute the expectation $\mathbb{E}(X_{(m)} | X_{(k)} = x_k)$.

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

Order statistics: optional problems

Ex. 4.24 (**). In the context of Ex. 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.

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

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


Ex. 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.


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

Ex. 4.29 (**). Let X_1, X_2, \dots 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 $(X_k)_{k \geq 1}$. Find the cdf of $\max\{X_1, X_2, \dots, X_N\}$, the maximum of the first N terms in that sequence.

Ex. 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 $\mathbb{P}(X_{(1)} > y n^{-\beta})$ has a well defined limit for large n . What can you deduce about the distribution of the rescaled variable $Y_n \stackrel{\text{def}}{=} n^\beta X_{(1)}$ for large enough n ?


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


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

Ex. 4.33 (****). Denote $Y_n^* = Y_n - \log n$; show that the corresponding cdf, $\mathbb{P}(Y_n^* \leq x)$, approaches $e^{-e^{-x}}$, as $n \rightarrow \infty$. Deduce that the expectation of the limiting distribution equals γ , the Euler constant,¹ and its variance is $\sum_{k \geq 1} \frac{1}{k^2} = \frac{\pi^2}{6}$. 

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 $\mu + \beta\gamma$, its variance is $\pi^2\beta^2/6$, and its moment generating function equals $\Gamma(1 - \beta t)e^{\mu t}$.

¹the Euler constant γ is $\lim_{n \rightarrow \infty} (\sum_{k=1}^n \frac{1}{k} - \log n) \approx 0.5772156649\dots$

Ex. 4.34 (****). By using the Weierstrass formula, $\prod_{k=1}^{\infty} (1 + \frac{z}{k})^{-1} e^{z/k} = e^{\gamma z} \Gamma(z+1)$, (where γ is the Euler constant ¹ and $\Gamma(\cdot)$ is the classical gamma function, $\Gamma(n) = (n-1)!$ for integer $n > 0$) or otherwise, show that the moment generating function $Ee^{tZ_n^*}$ of $Z_n^* = Z_n - EZ_n$ approaches $e^{-\gamma t} \Gamma(1-t)$ as $n \rightarrow \infty$ (eg., for all $|t| < 1$). Deduce that in that limit $Z_n^* + \gamma$ is asymptotically Gumbel distributed (with $\beta = 1$ and $\mu = 0$). 

Ex. 4.35 (***). Let X_1, X_2, \dots be independent $\text{Exp}(\lambda)$ random variables; further, let $N \sim \text{Poi}(\nu)$, independent of the sequence $(X_k)_{k \geq 1}$, and let $X_0 \equiv 0$. Find the distribution of $Y \stackrel{\text{def}}{=} \max\{X_0, X_1, X_2, \dots, X_N\}$, the maximum of the first N terms of this sequence, where for $N = 0$ we set $Y = 0$. 

Coupling

Ex. 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 $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$. Can you find all possible couplings for these variables?

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

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

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

Ex. 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 $X \sim \text{Ber}(p)$, 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 $0 < p_1 \leq p_2 < 1$ and ξ as above. Show that $X \preceq Y$ and that $P(X \leq Y) = 1$. Compare your construction to the second table in Example 5.1.

Ex. 5.6 (**). In the setup of Example 5.1, show that $X \sim \text{Ber}(p_1)$ is stochastically smaller than $Y \sim \text{Ber}(p_2)$ (ie., $X \preceq Y$) iff $p_1 \leq p_2$. Further, show that $X \preceq Y$ is equivalent to existence of a coupling (\tilde{X}, \tilde{Y}) of X and Y in which these variables are ordered with probability one, $\hat{P}(\tilde{X} \leq \tilde{Y}) = 1$.

Ex. 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 $X \preceq Y$?

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

Ex. 5.8 (**). Let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$ 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.

Ex. 5.9 (**). Let $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(r)$ be two geometric random variables, $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(r)$. 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 $\text{Exp}(\lambda)$). 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: $Z_1 + Z_2 \sim \Gamma(c_1 + c_2, \lambda)$.

Ex. 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.

Ex. 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 $d_{\text{TV}}(\mu, \nu) \equiv d_{\text{TV}}(\{p\}, \{q\}) \stackrel{\text{def}}{=} \max_A |\mu(A) - \nu(A)| = \frac{1}{2} \sum_z |p_z - q_z|$, where the sum runs over all $z \in \mathcal{X} \cup \mathcal{Y}$. Deduce that $d_{\text{TV}}(\cdot, \cdot)$ is a distance between probability measures (ie., it is non-negative, symmetric, and satisfies the triangle inequality) such that $d_{\text{TV}}(\mu, \nu) \leq 1$ for all probability measures μ and ν .

Ex. 5.12 (**). In the setting of Ex. 5.11 show that $d_{\text{TV}}(\{p\}, \{q\}) = \sum_z (p_z - \min(p_z, q_z)) = \sum_z (q_z - \min(p_z, q_z))$.

Ex. 5.13 (**). If $\hat{P}(\cdot, \cdot)$ is the maximal coupling of X and Y as defined in Example 5.9 and Remark 5.9.1, show that $\hat{P}(\tilde{X} \neq \tilde{Y}) = d_{\text{TV}}(X, Y)$.

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

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

$$\frac{1}{2} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \leq d_{\text{TV}}(\tilde{X}, \tilde{Y}) \leq \lambda^2/n \quad \text{for every } k \geq 0.$$

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

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

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


and deduce the integral representation of $\text{E}g(X)$ in (5.2), $\text{E}g(X) \equiv \int_0^\infty g'(z) \mathbb{P}(X > z) dz$.


Ex. 5.17 (**). Let $X \sim \text{Bin}(1, p_1)$ and $Y \sim \text{Bin}(2, p_2)$ with $p_1 \leq p_2$. By constructing an explicit coupling or otherwise, show that $X \preceq Y$.


Ex. 5.18 (***). Let $X \sim \text{Bin}(2, p_1)$ and $Y \sim \text{Bin}(2, p_2)$ with $p_1 \leq p_2$. Construct an explicit coupling showing that $X \preceq Y$.


Ex. 5.19 (***). Let $X \sim \text{Bin}(3, p_1)$ and $Y \sim \text{Bin}(3, p_2)$ with $0 < p_1 \leq p_2 < 1$. Construct an explicit coupling showing that $X \preceq Y$.


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


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


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


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


Ex. 5.24 (**). Prove the additive property of $d_{\text{TV}}(\cdot, \cdot)$ for independent sums, (5.7), by following the approach in Remark 5.13.1. 

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

Ex. 5.26 (**). If X, Y are stochastically ordered, $X \preceq Y$, and Z is independent of X and Y , show that $(X + Z) \preceq (Y + Z)$. 

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

Ex. 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 $X_1 \preceq Y_1$ and $X_2 \preceq Y_2$, deduce that $X \preceq Y$. 

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

Some non-classical limits

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

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

Ex. 6.3 (**). Use the estimate in Ex. 6.2 to prove Theorem 6.1: if $X_k \sim \text{Ber}(p_k^{(n)})$, $k = 1, \dots, n$, are independent with $\max_{1 \leq k \leq n} p_k^{(n)} \rightarrow 0$ and $\sum_{k=1}^n p_k^{(n)} \equiv \mathbb{E}(\sum_{k=1}^n X_k^{(n)}) \rightarrow \lambda \in (0, \infty)$ as $n \rightarrow \infty$. Show that the generating function of $Y_n = \sum_{k=1}^n X_k$ converges pointwise to that of $Y \sim \text{Poi}(\lambda)$ in this limit.

Ex. 6.4 (**). 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 $r/n \rightarrow c$ as $n \rightarrow \infty$, then $\mathbb{E}(\frac{1}{n} N_k) \rightarrow \frac{c^k}{k!} e^{-c}$ and $\text{Var}(\frac{1}{n} N_k) \rightarrow 0$ as $n \rightarrow \infty$.

Ex. 6.5 ().** In the setup of Ex. 6.4, let X_j be the number of balls in box j . Show that for each integer $k \geq 0$, we have $P(X_j = k) \rightarrow \frac{c^k}{k!} e^{-c}$ as $n \rightarrow \infty$. Further, show that for $i \neq j$, the variables X_i and X_j are not independent, but in the limit $n \rightarrow \infty$ they become independent $\text{Poi}(c)$ distributed random variables.

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

Ex. 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 $ET_n = n \sum_{k=1}^n k^{-1} \approx n \log n$ and $\text{Var}T_n \leq n^2 \sum_{k=1}^n k^{-2}$.

Ex. 6.8 ().** In the setup of Ex. 6.7, show that for all real a , $P(T_n \leq n \log n + na) \rightarrow \exp\{-e^{-a}\}$, as $n \rightarrow \infty$. [Hint: If $T_n > k$, then k balls are placed into n boxes so that at least one box is empty.]

Ex. 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 Ex. 6.7, the problem is about evaluating $P(T_{365} \leq 365k)$.]

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


Ex. 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 $M_T(t) = p(1 - (1-p)e^t)^{-1}$ and deduce that the total running cost X of the car up to and including the final service has MGF $M_X(t) = p(1 - (1-p)M_C(t))^{-1}$. Find the expectation EX of X and show that for small enough p the distribution of $X^* \stackrel{\text{def}}{=} X/EX$ is close to $\text{Exp}(1)$. [Hint: Find the MGF of X^* and follow the approach of Ex. 6.10.]


Ex. 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 $n \rightarrow \infty$. 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 $n \rightarrow \infty$. Notice that with $n = 365$ and $r = 23$ this is the famous 'birthday paradox'.


Ex. 6.13 (*)**. In the setup of Ex. 6.4, let X_j be the number of balls in box j . Show that for every integer $k_j \geq 0$, $j = 1, \dots, n$, satisfying $\sum_{j=1}^n k_j = r$ we have


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


Now suppose that $Y_j \sim \text{Poi}(\frac{r}{n})$, $j = 1, \dots, n$, are independent. Show that $P(Y_1 = k_1, \dots, Y_n = k_n \mid \sum_j Y_j = r)$ is given by the expression in the last display.


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

Ex. 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 $1/2$. 

Ex. 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 $n \rightarrow \infty$. 

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

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

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

Introduction to percolation

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

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

Ex. 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 $nr^*(r^*-1)^{n-1}$.

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

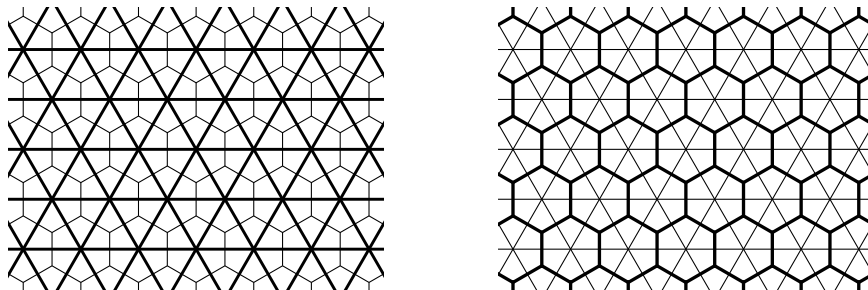


Figure 1: Left: triangular lattice (bold) and its dual hexagonal lattice (thin); right: hexagonal lattice (bold) and its dual triangular lattice (thin).

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

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

Ex. 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 $(d-1)p > 1$.

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

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

Ex. 7.10 (**). Show that for all $d \geq 1$, we have $p_{cr}(d) \leq \vec{p}_{cr}(d)$.

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

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

Ex. 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 $\vec{c}_0(k) \leq k3^{k-1}$.

Ex. 7.14 (**). Show that $p_{cr}^{d,\text{site}} \leq p_{cr}^{d-1,\text{site}}$ for all $d > 1$, while $p_{cr}^{d,\text{site}} \geq \frac{1}{2d-1}$ for fixed $d > 1$, cf. Lemma 7.5. 🚶

Ex. 7.15 (**). Carefully show that the number $c_0(n)$ of contours of length n separating the origin from infinity is not bigger than $5n \cdot 7^{n-1}$.

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

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

Ex. 7.18 (****). In the setting of Ex. 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 , $b_v(n) \leq AR^n$. 🚶

Ex. 7.19 (****). Consider the Bernoulli bond percolation model on an infinite graph $G = (V, E)$ of uniformly bounded degree, cf. Ex. 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 Ex. 7.19 is known to hold for all $p < p_{cr}$.