

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 $E(X_{(2)} | X_{(1)} = x_1)$ and $E(X_{(1)} | X_{(2)} = x_2)$.

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 $E(X_{(m)} | X_{(k)} = x_k)$.

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

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, $P(X > k) = (1 - p)^k$ 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 $P(X_{(75)} \leq 0.8)$.

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 $P(X_{(1)} > yn^{-\beta})$ has a well defined limit for large n . What can you deduce about the distribution of the rescaled variable $Y_n \stackrel{\text{def}}{=} n^\beta X_{(1)}$ 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, $P(Y_n^* \leq x)$, 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 $\Gamma(1 - \beta t)e^{\mu t}$.

Ex. 4.34 (****). By using the Weierstrass formula, $\prod_{k=1}^{\infty} (1 + \frac{z}{k})^{-1} e^{z/k} = e^{z\gamma} \Gamma(z + 1)$, (where γ is the Euler constant¹ 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 $Y = 0$. 

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

Coupling

Ex. 5.1 (*). In the setting of Example 5.1 show that every convex linear combination of tables T_1 and T_2 , i.e., 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 $\tilde{X} = \mathbb{1}_{\xi < p_1}$ and $\tilde{Y} = \mathbb{1}_{\xi < p_2}$ for some $0 < p_1 \leq p_2 < 1$ and ξ as above. Show that $\tilde{X} \preceq \tilde{Y}$ and that $P(\tilde{X} \leq \tilde{Y}) = 1$. 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 $Y \sim \text{Ber}(p_2)$ (i.e., $X \preceq Y$) 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, $\tilde{P}(\tilde{X} \leq \tilde{Y}) = 1$.

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 (i.e., it is non-negative, symmetric, and satisfies the triangle inequality) such that $d_{\text{TV}}(\cdot, \cdot) \leq 1$ 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

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

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

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


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


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 $Y \preceq X$? 