Order statistics

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

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

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

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 $Y \overset{\mathsf{def}}{=} F(X) \sim \mathcal{U}(0,1)$.

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

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

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

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

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

Ex. 4.9 (**). Let X_1, …, 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 $\mathsf{P}\big(X_{(1)} > \frac{y}{n}\big) \approx e^{-cy}$ for large n. Deduce that the distribution of $Y_n \overset{\mathsf{def}}{=} nX_{(1)}$ is approximately $\mathsf{Exp}(c)$ 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:

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

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

Ex. 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 $y \ge 0$ we have $\mathsf{P}\big(X_0 > y, X_0 = X_k\big) = e^{-\lambda y}\,\frac{\lambda_k}{\lambda_0}$, 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.

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

Ex. 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 $\mathsf{P}\big(a + X < \min(Y,Z) \,|\, a < \min(Y,Z)\big) = \frac{\lambda}{\lambda+\mu+\nu}$.

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

Ex. 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 $n \to \infty$.

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, …, 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 $\mathbb{R}$.

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 Exp(λ) 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 $\mathsf{E}(X_{(1)} \mid X_{(2)} = x_2)$.

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

a) Show that the gaps $\bigl(\Delta_{(k)}X\bigr)_{k=1}^n$ 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 $\mathsf{E}(X_{(m)} \mid X_{(k)} = x_k)$.

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

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 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.

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

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 $\mathsf{P}(X_{(75)} \le 0.8)$.

Ex. 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.

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

Ex. 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 $Y_n \overset{\mathsf{def}}{=} n^\delta X_{(1)}$ converges to a well-defined limit as $n \to \infty$. 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 fisrt 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, $\mathsf{P}(Y_n^* \le x)$, approaches $e^{-e^{-x}}$, as $n \to \infty$. Deduce that the expectation of the limiting distribution equals γ, the Euler constant,[1] 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[1] $\mu + \beta\gamma$, 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, $\prod_{k=1}^\infty \bigl(1 + \frac{z}{k}\bigr)^{-1} e^{z/k} = e^{\gamma z}\Gamma(z + 1)$, (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 $\mathsf{E}e^{tZ_n^*}$ 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 $\mu = 0$).

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

---

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