Statistics: Tutorial sheet 1 - Solutions to Practice Exercises

Exercise 1. Let $X_1, \dots, X_n$ be iid Uniform$(0, \theta)$ distributed for some $\theta > 0$. Then $f_X(x \mid \theta) = 1/\theta$ for $0 \leq x \leq \theta$.

a. Derive the cdf of $X_{(n)} = \max\{X_1, \dots, X_n\}$.

b. Use your answer to a. to derive the first two moments of $X_{(n)}$.

Solution.

a. We derive the cdf and pdf using the independence assumption:

$$
\begin{aligned}
F(x \mid \theta) &= \mathbb{P}_\theta(X_{(n)} \leq x) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i \leq x) = \frac{x^n}{\theta^n} \\
f(x \mid \theta) &= n\frac{x^{n-1}}{\theta^n}
\end{aligned}
$$

b. We can solve this exercise in two different ways. The first way is using traditional integration techniques. From the pdf we obtain the moments

$$
\begin{aligned}
\mathbb{E}X_{(n)} &= \int_0^\theta x\frac{n}{\theta^n}x^{n-1}dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta. \\
\mathbb{E}X_{(n)}^2 &= \int_0^\theta x^2\frac{n}{\theta^n}x^{n-1}dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n}{n+2}\theta^2.
\end{aligned}
$$

The second technique relies on the fact that we know that $f(x) = n\frac{x^{n-1}}{\theta^n}$ is a pdf for all $n \in \mathbb{N}$. Using that fact we get

$$
\begin{aligned}
\mathbb{E}X_{(n)} &= \int_0^\theta x\frac{n}{\theta^n}x^{n-1}dx = \frac{n}{n+1}\theta\int_0^\theta \frac{n+1}{\theta^{n+1}}x^n dx = \frac{n}{n+1}\theta. \\
\mathbb{E}X_{(n)}^2 &= \int_0^\theta x^2\frac{n}{\theta^n}x^{n-1}dx = \frac{n}{n+2}\theta^2\int_0^\theta \frac{n+2}{\theta^{n+2}}x^{n+1} dx = \frac{n}{n+2}\theta^2.
\end{aligned}
$$

Exercise 2. Let (X, Y) be a random vector with joint pdf

$$f(x, y) = 15x^2y \qquad \text{if } 0 < x < y < 1.$$

a. Show that $f(x, y)$ is indeed a pdf, that is, show that the integral of $f(x, y)$ over its domain is equal to one.

b. Derive the univariate pdf of X.

SOLUTION.

a. We check whether $\int\int f(x,y)dxdy = 1$.

$$\int_0^1 \int_x^1 x^2 y dy dx = \int_0^1 x^2 \left[y^2/2\right]_x^1 dx = \frac{1}{2}\int_0^1 x^2(1-x^2)dx = \frac{1}{2}\left[x^3/3 - x^5/5\right]_0^1 = \frac{1}{15}.$$

b.

$$f(x) = \int_0^1 f(x,y)dy = \int_x^1 15x^2 y dy = 15x^2 \int_x^1 y dy = 15x^2\left[y^2/2\right]_x^1 = \frac{15}{2}x^2(1-x^2).$$

Exercise 3 (4.31). Suppose Y has a binomial distribution with n trials and success probability X, where $X \sim \text{Uniform}(0,1)$. Recall that a Binomial$(n,p)$ distribution has mean np and variance $np(1-p)$.

a. Find $\mathbb{E}Y$ and $\mathbb{V}\text{ar}\, Y$.
   Hint: use the laws of total expectation that were discussed.

SOLUTION.

a.

$$\begin{aligned}
\mathbb{E}Y &= \mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(nX) = n\mathbb{E}(X) = n/2.\\
\mathbb{V}\text{ar}\, Y &= \mathbb{E}(\mathbb{V}\text{ar}(Y \mid X)) + \mathbb{V}\text{ar}(\mathbb{E}(Y \mid X)) = \mathbb{E}(nX(1-X)) + \mathbb{V}\text{ar}(nX)\\
&= n\mathbb{E}X - n\mathbb{E}X^2 + n^2\,\mathbb{V}\text{ar}(X) = n/2 - n/3 + n^2/12\\
&= n(n+2)/12.
\end{aligned}$$

Exercise 4 (5.34). Let $X_1, \ldots, X_n$ be an iid sequence of random variables with $\mathbb{E}(X_1) = \mu$ and $\mathbb{V}\text{ar}(X_1) = \sigma^2$. Show that

$$\mathbb{E}\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} = 0 \qquad \text{and} \qquad \mathbb{V}\text{ar}\,\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} = 1.$$

It follows that the normalisation of $\overline{X}_n$ in the CLT has the same first two moments as the limiting Normal$(0,1)$ distribution.

SOLUTION.

$$\begin{aligned}
\mathbb{E}\overline{X}_n &= \mathbb{E}\frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}n\mathbb{E}X_1 = \mathbb{E}X_1 = \mu.\\
\mathbb{E}\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} &= \frac{\sqrt{n}}{\sigma}\mathbb{E}(\overline{X}_n - \mu) = \frac{\sqrt{n}}{\sigma}(\mathbb{E}(\overline{X}_n) - \mu) = \frac{\sqrt{n}}{\sigma} \times 0 = 0.\\
\mathbb{V}\text{ar}\,\overline{X}_n &= \mathbb{V}\text{ar}\,\frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\text{ar}\,X_i = \frac{1}{n^2}n\,\mathbb{V}\text{ar}\,X_1 = \frac{1}{n}\,\mathbb{V}\text{ar}\,X_1 = \frac{\sigma^2}{n}.\\
\mathbb{V}\text{ar}\,\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} &= \frac{n}{\sigma^2}\,\mathbb{V}\text{ar}(\overline{X}_n - \mu) = \frac{n}{\sigma^2}\,\mathbb{V}\text{ar}\,\overline{X}_n = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1.
\end{aligned}$$

Exercise 5. Let $X_1, \ldots, X_n$ be iid Exponential(λ) distributed random variables, so $f_X(x) = \lambda e^{-\lambda x}$.

a. Derive the expectation and variance of X_1.

b. Show that $\sqrt{n}(\lambda \overline{X}_n - 1) \xrightarrow{d}$ Normal$(0,1)$ as n goes to infinity.

SOLUTION.

a. The easiest way to do this is to use integration by parts.

$$\begin{aligned}
\mathbb{E}(X_1) &= \int_0^\infty x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}. \\
\mathbb{E}(X_1^2) &= \int_0^\infty x^2\lambda e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x}\right]_0^\infty - \int_0^\infty 2x \times -e^{-\lambda x} dx \\
&= \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}. \\
\mathbb{V}\mathrm{ar}\, X_1 &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.
\end{aligned}$$

b. By the CLT we have

$$\sqrt{n}\frac{\overline{X}_n - 1/\lambda}{\sqrt{1/\lambda^2}} \xrightarrow{d} \mathrm{N}(0,1).$$

Multiply both the numerator and denominator by λ to obtain the result.

Exercise 6. Suppose $X_n \sim$ Binomial(n, p), then

$$f_{X_n}(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The binomial distribution is a very useful distribution that describes many experiments. It originates as the sum of n independent $Y_i \sim$ Bernoulli(p) random variables, that is, $X_n = \sum_{i=1}^n Y_i$. The unfortunate part about the binomial distribution is that it contains binomial coefficients, which are hard to compute and take a lot of computing time. Use the central limit theorem to show that the distribution of X_n can be approximated by a Normal$(np, np(1-p))$ distribution as n goes to infinity. We can therefore nicely approximate the binomial distribution, which is discrete, by a normal distribution, which is continuous, as n goes to infinity.

SOLUTION. Note that we can rewrite

$$X_n = \sum_{i=1}^n Y_i = n\frac{1}{n}\sum_{i=1}^n Y_i = n\overline{Y}_n.$$

We have $\mathbb{E}(Y_1) = p$ and $\mathbb{V}\mathrm{ar}(Y_1) = p(1-p)$. Therefore we obtain by the CLT that

$$\frac{n\overline{Y}_n - np}{\sqrt{np(1-p)}} = \sqrt{n}\frac{\overline{Y}_n - p}{\sqrt{p(1-p)}} \xrightarrow{d} \mathrm{N}(0,1).$$

For large n we thus have approximately

$$X_n = n\overline{Y}_n \sim \sqrt{np(1-p)}\mathrm{N}(0,1) + np = \mathrm{N}(np, np(1-p)).$$

Therefore, for large n we have approximately $X_n \approx \mathrm{N}(np, np(1-p))$.