(1) **Distribution of the maximum**

Let $X_1, X_2, \dots$ be a sequence of i.i.d. with uniform $(0,1)$ distribution and let $X_{(n)} = \max\limits_{1\le i\le n} X_i$.

Show that the sequence

$$Y_n = n(1 - X_{(n)}), \quad n \in \mathbb{N}$$

converges to an exponential $\exp(1)$ random variable as $n \to \infty$.

•) $$F_n(x) := \mathbb{P}(X_{(n)} \le x) = \mathbb{P}\left(\max_{1\le i\le n} X_i \le x\right) = \mathbb{P}\left(\bigcap_{i=1}^{n} [X_i \le x]\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \le x) = \begin{cases} 0, & \text{if } x<0 \\ x^n, & \text{if } 0\le x<1 \\ 1, & \text{if } 1\le x \end{cases}$$

$$G_n(y) := \mathbb{P}(Y_n \le y) = \mathbb{P}(n(1-X_{(n)}) \le y) = \mathbb{P}\left(1 - X_{(n)} \le \tfrac{y}{n}\right) = \mathbb{P}\left(1 - \tfrac{y}{n} \le X_{(n)}\right)$$

$$= 1 - \mathbb{P}\left(X_{(n)} < 1 - \tfrac{y}{n}\right) = \begin{cases} 0 & , \text{if } y \le 0 \\ 1-(1-\tfrac{y}{n})^n, & \text{if } 0<y\le n \\ 1 & , \text{if } n<y \end{cases}$$

We know from Analysis, that $\left(1+\frac{(-y)}{n}\right)^n \xrightarrow{n\to\infty} e^{-y}$ pointwise in $\mathbb{R}$, hence

$$G_n(y) \xrightarrow{n\to\infty} \begin{cases} 0 & , \text{if } y<0 \\ 1-e^{-y}, & \text{if } y\ge 0 \end{cases},$$ which is the distribution function

of an $\exp(1)$ random Variable, hence $Y_n \xrightarrow{d} Z \sim \exp(1)$

(2) **Coin throws**

An unfair coin is thrown 600 times. The probability of geting a tail in each throw is $\frac{1}{4}$.

(a) Use a Binomial distribution to compute the probability that the number of heads obtained does not differ more than 10 from 450.

(b) Use a Normal approximation without a continuity correction to calculate the probability in (a). How does the result change if the approimation is provided with a continuity correction?

a) Y ... number of tails after 600 tosses, $Y \sim \mathrm{Bin}(n, p)$, $n = 600$, $p = \frac{1}{4}$

$$\mathbb{P}(|(600-Y)-450| \leq 10) = \mathbb{P}(|150-Y| \leq 10) = \mathbb{P}(140 \leq Y \leq 160) \overset{R}{\approx} 0{,}68$$

b) $Y \approx Z \sim N(np, np(1-p))$

$$\mathbb{P}(140 \leq Z \leq 160) = 1 - \mathbb{P}(Z < 140) - \mathbb{P}(Z > 160) \overset{\text{symmetry of normal}}{=} 1 - \mathbb{P}(Z < 140) - \mathbb{P}(Z < 140)$$

$$= 1 - 2\,\mathbb{P}(Z < 140) \overset{R}{\approx} 0{,}65$$

$$np = \frac{600}{4} = 150; \quad np(1-p) = \frac{600}{4}\,\frac{3}{4} = \frac{3 \cdot 150}{4} = \frac{450}{4} = \frac{225}{2}$$

with continuity correction:

$$\mathbb{P}(140 \leq Y \leq 160) \approx \mathbb{P}(140 - \tfrac{1}{2} < Z < 160 + \tfrac{1}{2}) = 1 - 2\,\mathbb{P}(Z < 140 - \tfrac{1}{2}) \overset{R}{\approx} 0{,}68$$

(3) **Simulations**

(a) By applying the R-function `replicate()` generate a sample $X_1, \ldots, X_{10}$ of size 10 from an exponential distribution with a rate parameter 0.2 and sum up its elements. Do this sum 10 000 times and make a histogram of the simulation. Can you say something about the shape of distribution?

(b) Use R to simulate 50 tosses of a fair coin (0 and 1). We call a *run* a sequence of all 1's or all 0's. Estimate the average length of the longest run in 10000 trials and report the result.

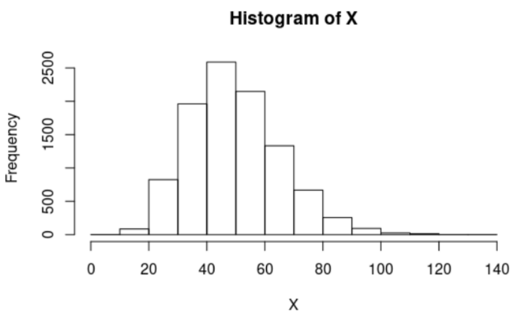
Hint: Use the commands `rbinom` and `rle`. The command `rle()` stands for run length encoding. For example,

```
rle(rbinom(5, 1, 0.5))$lengths
```

is a vector of the lengths of all the different runs in trial of 5 flips of a fair coin.

a) Defining $Y := \sum_{i=1}^{n} X_i$, $n = 10$, $\lambda := \frac{1}{5}$ we expect $Y \sim \text{Gamma}(n, \lambda)$,

where $E(Y) = \frac{n}{\lambda} = 5 \cdot 10 = 50$ and $\text{Var}(Y) = \frac{n}{\lambda^2} = 10 \cdot 25 = 250$



b) Estimation: longes run is approximately 5,94

(4) **Conditional variance**

(a) Show that for any two random variables X and Y the conditional variance identity holds

$$\mathbb{V}ar\,Y = \mathbb{E}\left(Var\,(Y|X)\right) + \mathbb{V}ar\,\left(\mathbb{E}\,(Y|X)\right),$$

provided that the expectations exist. The law of total expectation (the tower property) $\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y))$ should be applied.

(b) Suppose that the distribution of Y conditional on $X = x$ is $\mathcal{N}(x, x^2)$ and that the marginal distribution of X is uniform on $(0, 1)$. Compute $\mathbb{E}Y$, $\mathbb{V}ar\,Y$ and $\mathbb{C}ov(X, Y)$.

a) $$\mathbb{E}(\mathbb{V}ar(Y|X)) + \mathbb{V}ar(\mathbb{E}(Y|X)) = \mathbb{E}\left(\mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2\right) + \mathbb{E}\left((\mathbb{E}(Y|X))^2\right) - \left(\mathbb{E}(\mathbb{E}(Y|X))\right)^2$$

$$= \mathbb{E}(\mathbb{E}(Y^2|X)) \underbrace{- \mathbb{E}\left((\mathbb{E}(Y|X))^2\right) + \mathbb{E}\left((\mathbb{E}(Y|X))^2\right)}_{=0} - \left(\mathbb{E}(\mathbb{E}(Y|X))\right)^2$$

$$= \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{V}ar(Y)$$

b) $$f(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \Leftrightarrow f_{X,Y}(x,y) = f_X(x)\, f(y|x) = \begin{cases} 0 & , \text{ if } x < 0 \\ \frac{1}{\sqrt{2\pi}\,x} \exp\left(\frac{-(y-x)^2}{2x^2}\right) & , \text{ if } 0 \leq x < 1 \\ 0 & , \text{ if } 1 \leq x \end{cases}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\, dx$$

$$Z_x \sim \mathcal{N}(x, x^2)$$

$f_{X,Y}(x,y)$ is normal distribution with Expectation x

$$\mathbb{E}(Y) = \int_{\mathbb{R}} y\, f_Y(y)\, dy \overset{\text{Fubini}}{=} \int_0^1 \int_{\mathbb{R}} y\, f_{X,Y}(x,y)\, dy\, dx = \int_0^1 \mathbb{E}(Z_x)\, dx = \int_0^1 x\, dx = \boxed{\frac{1}{2}}$$

$$\mathbb{E}(Z^2) = \mathbb{V}ar(Z) + (\mathbb{E}(Z))^2 = 2x^2$$

$$\mathbb{E}(Y^2) = \int_{\mathbb{R}} y^2 f_Y(y)\, dy\, dx \overset{\text{Fubini}}{=} \int_0^1 \int_{\mathbb{R}} y^2 f(x,y)\, dy\, dx = \int_0^1 \mathbb{E}(Z_x^2)\, dx = \int_0^1 2x^2\, dx = \frac{2}{3}$$

$$\Rightarrow \mathbb{V}ar(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \frac{2}{3} - \frac{1}{4} = \frac{8-3}{12} = \boxed{\frac{5}{12}}$$

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xy\, f_{X,Y}(x,y)\, dx\, dy = \int_0^1 x \int_{\mathbb{R}} y\, f_{X,Y}(x,y)\, dy\, dx = \int_0^1 x\, \mathbb{E}(Z_x)\, dx = \int_0^1 x^2\, dx = \frac{1}{3}$$

$$\Rightarrow \mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = \frac{1}{3} - \frac{1}{2}\,\frac{1}{2} = \frac{4-3}{12} = \boxed{\frac{1}{12}}$$

(5) (a) **Delta method**
Let $X_1, \dots, X_n$ be i.i.d. from normal distribution with unknown mean μ and known variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Find the limiting distribution of $\sqrt{n}\left(\bar{X}^3 - c\right)$ for an appropriate constant c.

(b) **Logit transformation**
Let $X_n \sim bin(n,p)$. Consider the logit transformation, defined by

$$logit(y) = \ln \frac{y}{1-y}, \qquad 0 < y < 1.$$

Determine the approximate distribution of $logit\left(\frac{X_n}{n}\right)$.

a) By the CLT, $\dfrac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Y \sim \mathcal{N}(0,1)$

$Y_n := \dfrac{\bar{X}_n}{\sigma}$, $\theta := \dfrac{\mu}{\sigma}$, then $\sqrt{n}\,(Y_n - \theta) \xrightarrow{d} Y \sim \mathcal{N}(0,1)$

$g: \mathbb{R} \to \mathbb{R}: y \mapsto (\sigma y)^3$, $\Rightarrow g'(y) = 3\sigma^3 y^2$, and by the delta method,

$\sqrt{n}\left(\bar{X}_n^3 - \mu^3\right) = \sqrt{n}\left((\sigma Y_n)^3 - (\sigma\theta)^3\right) = \sqrt{n}\left(g(Y_n) - g(\theta)\right) \xrightarrow{d} g'(\theta)\,Y = 3\sigma^3\theta^2 Y$

and $3\sigma^3\theta^2 Y = 3\sigma^3 \frac{\mu^2}{\sigma^2} Y = 3\sigma\mu^2 Y \sim \mathcal{N}\left(0, (3\sigma\mu^2)^2\right)$

b) Let $Y_1, \dots, Y_n \sim bin(1,p)$, $X_n = \sum_{i=1}^n Y_i \sim bin(n,p)$, $\mathbb{E}(X_n) = np$, $\mathbb{V}ar(X_n) = np(1-p)$

$\mathbb{E}(Y_i) = p$, $\mathbb{V}ar(Y_i) = p(1-p)$

•) For $p \notin \{0,1\}$

By CLT: $\sqrt{n}\,\dfrac{\frac{X_n}{n} - p}{\sqrt{p(1-p)}} = \dfrac{\sqrt{n}\,(X_n - np)}{n\sqrt{p(1-p)}} = \dfrac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$

We define $Z_n := \dfrac{X_n}{n\sqrt{p(1-p)}}$ and $\theta := \sqrt{\dfrac{p}{1-p}}$, then

$\sqrt{n}\left(Z_n - \theta\right) \xrightarrow{d} \mathcal{N}(0,1)$, we define $g: (0,1) \to \mathbb{R}: z \mapsto logit\left(\sqrt{p(1-p)}\,z\right)$, hence

$$g'(z) = \frac{1 - \sqrt{p(1-p)}\,z}{z}\left(\frac{\sqrt{p(1-p)}}{1 - \sqrt{p(1-p)}\,z} + z\sqrt{p(1-p)}\left(1 - \sqrt{p(1-p)}\,z\right)^{-2}\sqrt{p(1-p)}\right)$$

$$= \frac{1 - \sqrt{p(1-p)}\,z}{\sqrt{p(1-p)}\,z} \quad \frac{\sqrt{p(1-p)}\left(1 - \sqrt{p(1-p)}\,z\right) + z\,p(1-p)}{\left(1 - \sqrt{p(1-p)}\,z\right)^2} = \frac{1}{z\left(1 - \sqrt{p(1-p)}\,z\right)}$$ and $g(\theta) = \log\left(\frac{p}{1-p}\right)$

by application of the delta method we obtain and

$\sqrt{n}\left(logit\left(\frac{X_n}{n}\right) - logit(p)\right) = \sqrt{n}\left(g(Z_n) - g(\theta)\right) \xrightarrow{d} g'(\theta)\,Z = \dfrac{\sqrt{1-p}}{\sqrt{p}\,(1-p)}\,Z = \dfrac{1}{\sqrt{p(1-p)}}\,Z \sim \mathcal{N}\left(0, \dfrac{1}{p(1-p)}\right)$

hence, $logit\left(\dfrac{X_n}{n}\right) \approx \mathcal{N}\left(logit(p), \dfrac{1}{np(1-p)}\right)$