

# HW5

Christian Sallinger

22.4.2021

## 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  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ . Show that the sequence

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

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

**Solution:** We will show convergence in distribution:

$$\begin{aligned} F_n(y) &= \mathbb{P}(Y_n \leq y) = \mathbb{P}(n(1 - X_{(n)}) \leq y) \\ &= \mathbb{P}(X_{(n)} \geq 1 - \frac{y}{n}) \\ &= 1 - \mathbb{P}(\max_{1 \leq i \leq n} X_i \leq 1 - \frac{y}{n}) \\ &= 1 - \mathbb{P}(X_1 \leq 1 - \frac{y}{n})^n \\ &= \begin{cases} 1, & 1 - \frac{y}{n} \leq 0 \\ 1 - (1 - \frac{y}{n})^n, & 0 < 1 - \frac{y}{n} < 1 \\ 0, & 1 \leq 1 - \frac{y}{n} \end{cases} \end{aligned}$$

Here we used the fact that  $\mathbb{P}(\max_{1 \leq i \leq n} X_i \leq x) = \mathbb{P}(X_1 \wedge X_2 \wedge \dots \wedge X_n \leq x)$  as well as the independence of the  $X_i$ . We can now reformulate the boundary:

$$1 - \frac{y}{n} \leq 0 \iff n \leq y$$

So for  $n \rightarrow \infty$  this case will never happen. The second one is

$$\begin{aligned} 0 < 1 - \frac{y}{n} < 1 &\iff 1 > \frac{y}{n} > 0 \iff n > y > 0 \\ 1 \leq 1 - \frac{y}{n} &\iff y \leq 0 \end{aligned}$$

If we now use our knowledge from calculus, namely that  $\lim_{n \rightarrow \infty} (1 + \frac{y}{n})^n = e^y$  we now get

$$\lim_{n \rightarrow \infty} F_n(y) = \begin{cases} 1 - e^{-y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

which is exactly the cdf of  $\exp(1)$ .

## 2. Coin throws

An unfair coin is thrown 600 times. The probability of getting 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 approximation is provided with a continuity correction?

**Solution:**

- (a) Suppose  $X \sim \text{bin}(600, \frac{3}{4})$ , what we want to calculate is

$$\mathbb{P}(440 \leq X \leq 460) = \mathbb{P}(X \leq 460) - \mathbb{P}(X \leq 439)$$

We can do this very easily in R. We note that this is not the same as

$$\mathbb{P}(X \leq 460) - \mathbb{P}(X \leq 440)$$

which we would get if we were working with continuous distributions. Nevertheless we will calculate this value, to compare in (b).

```
pbinom(460, 600, 3/4) - pbinom(439, 600, 3/4)
```

```
## [1] 0.6778428
```

```
pbinom(460, 600, 3/4) - pbinom(440, 600, 3/4)
```

```
## [1] 0.6540917
```

- (b) We want to use the CLT for  $X_i \sim \text{bernoulli}(\frac{3}{4})$ , with  $\mathbb{E}(X_i) = \frac{3}{4}$  and  $\text{Var}(X_i) = \frac{3}{16}$ . We however do not want to approximate a sample mean, but just the sum  $S_n = \sum X_i$ . If we plug this into the CLT we see that

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \approx \mathcal{N}(0, 1)$$

So all in all we get (without continuity correction)

$$\mathbb{P}(440 \leq S_n \leq 460) = \mathbb{P}\left(\frac{440 - n\mu}{\sqrt{n\sigma^2}} \leq \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{460 - n\mu}{\sqrt{n\sigma^2}}\right) \approx \Phi\left(\frac{460 - n\mu}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{440 - n\mu}{\sqrt{n\sigma^2}}\right)$$

In our case  $n = 600$  and the expectation and variance we have already stated above. With continuity correction we get almost the same, just  $460 + 0.5$  on the upper bound and  $440 - 0.5$  on the lower one. We use R to calculate the values.

```
pnorm((460-600*0.75)/sqrt(600*(3/16))) - pnorm((440-600*0.75)/sqrt(600*(3/16)))
```

```
## [1] 0.6542214
```

```
pnorm((460+ 0.5-600*0.75)/sqrt(600*(3/16))) - pnorm((440-0.5-600*0.75)/sqrt(600*(3/16)))
```

```
## [1] 0.6778012
```

We see that the value with continuity correction is a better approximation. What we can also observe that the value without continuity correction is a good approximation to the second value in (a), which is not surprising since there we assumed that we deal with a continuous distribution there.

### 3. Simulations

- (a) By applying the R- function `replicate()` generate a sample  $X_1, \dots, X_{10}$  of size 10 from an exponential distribution with a rate parameter 0.2 and sum up its elements. Do this sum 10000 times and make a histogram of the simulation. Can you say something about the shape of the 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,

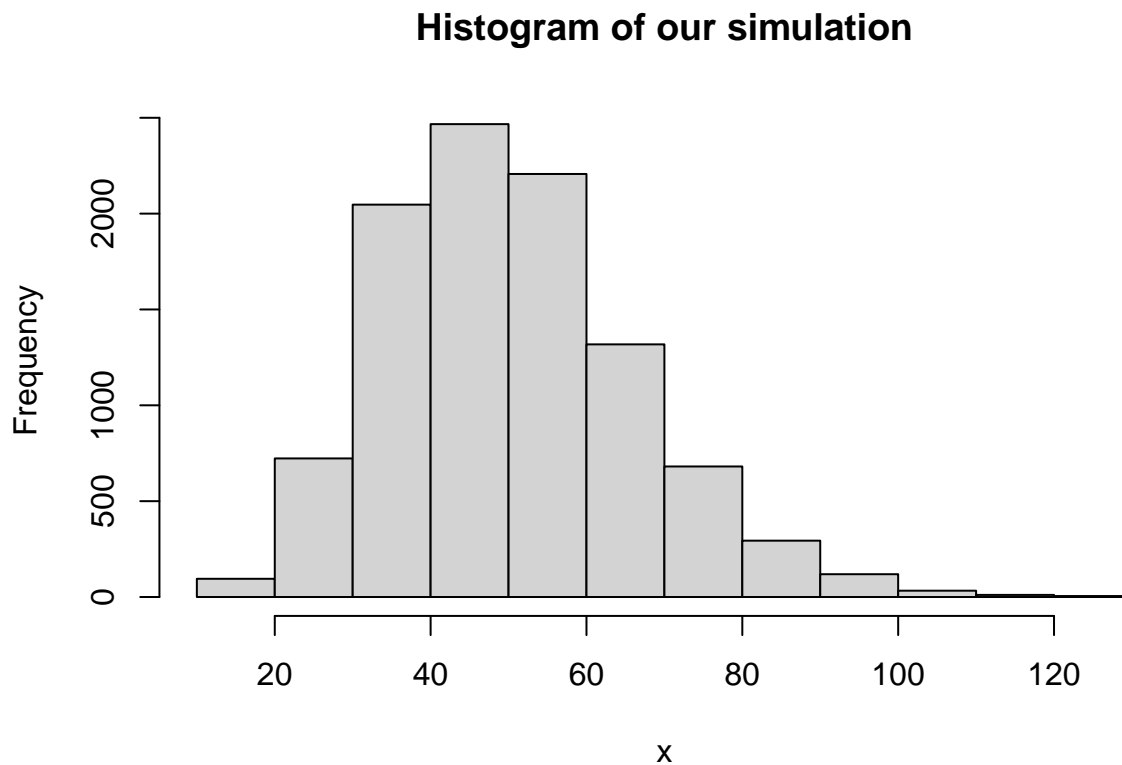
```
rle(rbinom(5, 1, 0.5))$lengths
```

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

**Solution:**

(a)

```
x = replicate(10000, sum(rexp(10,0.2)))
hist(x, main = "Histogram of our simulation")
```



(b)

```
mean(replicate(10000, max(rle(rbinom(50, 1, 0.5))$lengths)))
```

```
## [1] 5.993
```

#### 4. Conditional variance

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

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

provided that the expectation exists. 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  $\mathbb{C}ov(X, Y)$ .

**Solution:** (a) We first remind of the definition

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

We will use the conditional parallel axis theorem

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

With the law of total expectation we now get

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

(b) We use the law of total expectation and get

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(X) = \frac{1}{2}$$

To calculate the variance we use the formula we derived in (a) as well as  $\mathbb{E}(X^2) = \int_0^1 x^2 dx = 1/3$  to get

$$\mathbb{V}ar(Y) = \mathbb{E}(\mathbb{V}ar(Y|X)) + \mathbb{V}ar(\mathbb{E}(Y|X)) = \mathbb{E}(X^2) + \mathbb{V}ar(X) = \frac{1}{3} + \frac{1}{12} = \frac{5}{12}$$

To calculate the covariance we again use the law of total expectation as well as the fact that  $\mathbb{E}(XY) = \mathbb{E}(X\mathbb{E}(Y|X))$  and finally get

$$\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X\mathbb{E}(Y|X)) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{V}ar(X) = \frac{1}{12}$$

#### 5. (a) Delta method

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

## (b) Logit transformation

Let  $X_n \sim \text{bin}(n, p)$ . Consider the logit transformation, defined by

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

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

**Solution:** (a) We aim to use the lemma on page 10 of the slides to lecture five. We know from the CLT that

$$\sqrt{n}(\bar{X} - \mu) \rightarrow \mathcal{N}(0, \sigma^2)$$

We use the lemma with  $g(x) = x^3$ , so we get

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \rightarrow \mathcal{N}(0, \sigma^2 (g'(\mu))^2)$$

So the limiting distribution is

$$\mathcal{N}(0, 9\sigma^2\mu^4)$$

with constant  $c := \mu^3$ .

(b) We define the new i.i.d. random variables  $Y_1, \dots, Y_n \sim \text{bernoulli}(p)$ , then it holds that  $\frac{X_n}{n} = \bar{Y}_n$ . We again aim to use the same lemma as in (a) with  $g(x) := \text{logit}(x)$  and derivative

$$g'(x) = \frac{1}{x(1-x)}$$

We know from the CLT that

$$\sqrt{n}(\bar{Y}_n - p) \rightarrow \mathcal{N}(0, p(1-p))$$

With our lemma we get

$$\sqrt{n}(\text{logit}(\bar{Y}_n) - \text{logit}(p)) \rightarrow \mathcal{N}\left(0, \frac{1}{p(1-p)}\right)$$

With the results from last week we get, for large  $n$ , that

$$\text{logit}(\bar{Y}_n) \approx \mathcal{N}\left(\text{logit}(p), \frac{1}{np(1-p)}\right).$$