BI Norwegian Business School

FINAL EXAM: **EBA 2904 – Statistics with programming**

WITH: **Emil Aas Stoltenberg**

DAY OF EXAMINATION: **June 2, 2022**

EXAMINATION HOURS: **09:00–14:00**

PERMITTED AIDS: **A bilingual dictionary**

This exam set contains three exercises and comprises three pages. An appendix with results that can be pointed to in solving the exercises is included at the end of this document.

Exercise 1. Consider the sample space associated with rolling a die once,

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

For any subset A of Ω, let $|A|$ be the number of elements in A. For all subsets A of Ω and also for the empty set $\emptyset$, define the function

$$\Pr(A) = \frac{|A|}{6},$$

(a). Show that $\Pr(\cdot)$ is a probability function.

(b). Let $A = \{\text{odd outcome}\} = \{1, 3, 5\}$. Find the probability of A occurring.

(c). Let A be as in (b), and define the random variable

$$X = I_A,$$

with I_A the indicator function of the event A. Show that $\mathrm{E}(X) = 1/2$.

(d). Let X be as in (c). Show that $\mathrm{Var}(X) = 1/4$.

(e). Let A be as in (b), and define the event $B = \{\text{five or higher}\} = \{5, 6\}$. Show that A and B are independent events.

(f). Let A and B be as defined in (b) and (e). Show that A and B^c are independent events.

(g). Consider the function

$$Q(D) = 2\,\mathrm{E}(I_D X),$$

defined for all subsets D of Ω and also for the empty set. With B as defined in (e), show that

$$Q(B) = \Pr(B).$$

(h). Show that $Q(\cdot)$ and $\Pr(\cdot)$ are not the same function. To show this it suffices to find an event, say G, for which $Q(G) \neq \Pr(G)$.

(i). The function Q is as defined in (g). Show that $Q(\cdot)$ is a probability function.

Exercise 2. Consider the function

$$f(x) = \begin{cases} 1/\theta, & \text{for } x \in [0, \theta], \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $\theta > 0$ is some unknown parameter.

(a). Show that $f(x)$ is a probability density function (pdf).

(b). Let X be a random variable with the pdf $f(x)$ given in (1). Find the cumulative distribution function (cdf)

$$F(x) = \Pr(X \leq x),$$

of the random variable X.

(c). Find an expression for the expectation of X.

(d). Find an expression for the variance of X.

(e). Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with the same distribution as the X from (b). Show that[1]

$$\widehat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i, \tag{2}$$

is an unbiased estimator for θ. That $\widehat{\theta}$ is unbiased for θ means that $\mathrm{E}\,(\widehat{\theta}) = \theta$.

(f). Find an expression for the variance of $\widehat{\theta}$.

(g). Let the $X_1, \ldots, X_n$ be as in (e) and define $Y = \max(X_1, \ldots, X_n)$. The pdf of Y is then

$$g(y) = \frac{n}{\theta^n} y^{n-1}, \quad \text{for } y \in [0, \theta],$$

with $g(y) = 0$ for y outside of $[0, \theta]$. Use the random variable Y to construct another unbiased estimator for θ. Denote your estimator $\widetilde{\theta}$.

(h). When choosing between different estimators of the same parameter, your boss wants your company to use the one that minimises risk. Your boss defines the risk of an estimator as its average squared distance from what is being estimated. Since we are estimating θ, this means that the risk of an estimator δ is

$$\mathrm{risk}(\delta, \theta) = \mathrm{E}\,\{(\delta - \theta)^2\},$$

Which of the two estimators do you recommend to your boss: the estimator $\widehat{\theta}$ given in (2), or the estimator $\widetilde{\theta}$ that you found in (g)?

(i). In Python, to sample independent realisations from the distribution determined by the pdf $f(x)$ of (1), you can write

```
import numpy as np
xx = np.random.uniform(0,theta,n)
```

provided `theta` and `n` are assigned meaningful values. Sketch a Python script where you answer the question in (h) by way of simulations instead of mathematics.

Exercise 3. Kaia is a meticulous investor that keeps track of all her investment decisions and correlates them with entries in her diary. She has calculated that on average half of her decisions are good decisions, while the other half are bad decisions. On days Kaia feels happy and confident, however, sixty percent of her decisions are good decisions, and, luckily for her, she feels happy and confident eight out of ten days.

(a). What is the probability of Kaia making a good decision on a day she is not happy or not confident?

(b). Today Kaia only had time to make one decision, and it was a bad one. What is the probability that she had a one of her happy and confident days?

---

[1] *Hint*: If you find that $\mathrm{E}\,(\widehat{\theta})$ does not equal θ, then you might need to check your answer to (c), which in turn has implications for your answer to (d).

Appendix

(A.0) For all events A, B and C we have $A \cup B = B \cup A$ and $A \cap B = B \cap A$; and $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$, as well as

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$$

$$(A \cup B)^c = A^c \cap B^c, \quad \text{and} \quad (A \cap B)^c = A^c \cup B^c.$$

(A.1) The function $\Pr(\cdot)$ taking events as its arguments, is a **probability function** if
- (i) $\Pr(A) \geq 0$ for all events A;
- (ii) $\Pr(\Omega) = 1$ for the sample space Ω;
- (iii) if A and B are disjoint events (i.e. $A \cap B = \emptyset$), then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

(A.2) For any event A, $\Pr(A) = 1 - \Pr(A^c)$. For all events F and G, we define the set minus by $F \setminus G = F \cap G^c$. Also, for any probability function $\Pr(F \setminus G) = P(F) - P(F \cap G)$.

(A.3) Let A and B be two events such that $\Pr(B) > 0$. The **conditional probability** of A given B is $\Pr(A \mid B) = \Pr(A \cap B)/\Pr(B)$.

(A.4) Let A and B be two events, such that $0 < \Pr(B) < 1$. The **Law of total probability** says that $\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c)$.

(A.5) Let A and B be two events such that $\Pr(A) > 0$ and $\Pr(B) > 0$. **Bayes' formula** reads

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)}.$$

(A.6) Two events F **and** G **are independent** if $\Pr(F \cap G) = \Pr(F)\Pr(G)$.

(A.7) The **expectation** of a discrete random variable W taking the values $\{w_1, \ldots, w_k\}$ is

$$\mathrm{E}(W) = \sum_{\omega \in \Omega} W(\omega)\Pr(\omega) = \sum_{j=1}^{k} w_j \Pr(W = w_j).$$

The expectation of a continuous random variable Z with probability density function (pdf) $f(z)$ is

$$\mathrm{E}(Z) = \int_{-\infty}^{\infty} z f(z)\,\mathrm{d}z.$$

(A.8) The **indicator function** of an event G, denoted I_G, is

$$I_G(\omega) = \begin{cases} 1, & \text{if } \omega \in G, \\ 0, & \text{otherwise,} \end{cases}$$

The expectation of an indicator function is $\mathrm{E}(I_G) = \Pr(G)$. For any two events F and G, the product of their indicator functions is $I_F I_G = I_{F \cap G}$. For any two events F and G, we have that $I_{F \cup G} = I_F + I_G - I_{F \cap G}$. In particular, if F and G are disjoint, $I_{F \cup G} = I_F + I_G$.

(A.9) Let X be a random variable taking the values $\{x_1, \ldots, x_k\}$, and Y be a random variable taking the values $\{y_1, \ldots, y_m\}$. The random variables X and Y are independent if the events

$$\{X = x_i\} \quad \text{and} \quad \{Y = y_j\},$$

are independent for all $i = 1, \ldots, k$ and $j = 1, \ldots, m$.

(A.10) If X and Y are two independent random variables, then $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

(A.11) The **cumulative distribution function (cdf)** of a random variable X is the function $F(x) = \Pr(X \leq x)$. If X is a discrete random variable taking the values $\{x_1, \ldots, x_k\}$, then $F(x) = \sum_{j: x_j \leq x} \Pr(X = x_j)$. If X is a continuous random variable with pdf $f(x)$, then $F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$. In particular, $f(x) = F'(x)$, where $F'(x)$ is the derivative of $F(x)$.