BI Norwegian Business School

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

WITH: **Emil Aas Stoltenberg**

DAY OF EXAMINATION: **June 4, 2024**

EXAMINATION HOURS: **Five hours**

PERMITTED AIDS: **A bilingual dictionary**

INSTRUCTIONS: **Write with a pen, NOT a pencil. Be concise.**

This exam set contains three exercises and comprises five pages. An appendix with results that can be pointed to in solving the exercises is included on pages 4 and 5.

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

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

and define the function

$$\Pr(A) = \frac{|A|}{6}, \quad \text{for all } A \subset \Omega.$$

Here $|A|$ denotes the number of elements in A.

(a). Show that Pr is a probability function.

(b). Consider the event $A = \{\text{the die shows an odd number}\}$. Find the probability of the event A.

(c). Consider the event $B = \{\text{die shows a number bigger than or equal to } 5\}$. Show that A and B are independent events.

(d). With $A = \{\text{the die shows an odd number}\}$ as in (b), introduce the random variable

$$X(\omega) = \begin{cases} \omega - 3, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in \Omega \setminus A, \end{cases}$$

Find the probability mass function of X. In other words, compute the probabilities $\Pr(X = x)$ for all $x \in \mathbb{R}$.

(e). Let X be as in (d). Make a sketch of the cumulative distribution function $F(x) = \Pr(X \leq x)$. Clearly indicate the relevant numbers on the x-axis and on the y-axis.

(f). Compute the expectation and the variance of X.

(g). Let the event B be as defined in (c), and introduce the random variable Y defined by

$$Y(\omega) = \begin{cases} 1, & \text{if } \omega \in B, \\ -1, & \text{if } \omega \in \Omega \setminus B. \end{cases}$$

With X as defined in (d), show that X and Y are dependent random variables.

(h). With X and Y as above, compute the covariance

$$\text{Cov}(X, Y) = \text{E}\{(X - \text{E}[X])(Y - \text{E}[Y])\}.$$

Exercise 2. Consider the function

$$f(x) = \frac{a}{\theta^a}\, x^{a-1}, \quad \text{for } x \in [0, \theta],$$

and $f(x) = 0$ for x outside of $[0, \theta]$, where $\theta > 0$ is an unknown parameter that we want to estimate, while $a > 0$ is some known number (so we do not need to estimate it).

(a). Verify that $f(x)$ is a probability density function (p.d.f.).

(b). Let X be a random variable with distribution determined by the p.d.f. $f(x)$. Find the cumulative distribution function (c.d.f.) $F(x)$ of X.

(c). Please do not show that for any $k > 0$ the expectation of X^k is

$$\mathrm{E}\,(X^k) = \frac{a}{a+k}\theta^k,$$

but do find the expectation and variance of X.

(d). Let $X_1, \dots, X_n$ be $n \geq 2$ independent and identically distributed random variables, each with the same distribution as X. Let $\bar{X}_n = (X_1 + \cdots + X_n)/n$ be the empirical mean of these, and consider the estimator of θ given by

$$\widehat{\theta}_n = \frac{a+1}{a}\bar{X}_n.$$

Find the expectation and the variance of this estimator.

(e). Let Y_n be the maximum of the n random variables $X_1, \dots, X_n$, that is

$$Y_n = \max(X_1, \dots, X_n).$$

Since $X_1, \dots, X_n$ are independent random variables, the events $\{X_1 \leq y\}, \dots, , \{X_n \leq y\}$ are also independent, for any y. Show that the c.d.f. of Y is given by

$$G(y) = \begin{cases} 0, & \text{for } y < 0 \\ (y/\theta)^{an}, & \text{for } 0 \leq y < \theta \\ 1, & \text{for } y \geq \theta. \end{cases}$$

(f). Find an expression for the probability density function, say $g(y)$, of Y.

(g). Find expressions for the expectation and the variance of Y_n. To get a full score, you should be able to find these without doing the integration.

(h). Let $\widetilde{\theta}_n = c_n Y_n$ be an estimator for θ, where c_n is a constant (i.e., not a random variable) so that $\mathrm{E}\,(\widetilde{\theta}_n) = \theta$. Please find an expression for c_n.

(i). In terms of mean squared error, which estimator is best for estimating θ?

(j). The Python code below generates a random sample of size n, contained in `xx`, from the distribution with p.d.f. $f(x)$ and parameter θ. Explain why these simulations can set us on track of an answer to Ex. 2(i). Comment also on the limitations of this approach.

```python
import numpy as np
n = 23
theta = 2.34
a = 2. # just some positive number
sims = 10**3
thetahats = np.zeros(sims)
thetatildes = np.zeros(sims)

for jj in range(0,sims):
        uu = np.random.uniform(0,1,n)
        xx = theta*uu**(1/a) # the data
        thetahats[jj] = ((a+1)/a)*np.mean(xx)
        thetatildes[jj] = cn*np.max(xx)

mse_hat = np.mean((thetahats - theta)**2)
mse_tilde = np.mean((thetatildes - theta)**2)
```

(k). Let $\varepsilon > 0$ be some number, and Y_n be as defined in Ex. (e). Show that

$$\Pr(|Y_n - \theta| \geq \varepsilon) \to 0,$$

as n tends to infinity.

Exercise 3. There are two coins in a hat, a coin whose probability of heads is one half (the fair coin), and a coin whose probability of heads is two thirds. To the naked eye they look identical.

(a). You pick a coin at random (meaning that both coins in the hat have equal probability of being picked) and flip it. What is the probability that it lands heads?

(b). You pick a coin at random, flip it, and it comes up heads. What is the probability that you picked the fair coin?

(c). You decide to flip the same coin once more, and yet again it comes up heads. What is now the probability that you picked the fair coin? Please also state the very reasonable assumption you are making in this computation.

(d). Consider the event $H_1 = \{\text{heads on first toss}\}$ and the event $H_2 = \{\text{heads on second toss}\}$. Are these two events independent? Why, or why not?

Appendix

(A.0) 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; and (iii) if A and B are disjoint events (i.e. $A \cap B = \emptyset$), then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

(A.1) For any probability function and events A and B, we have

(a) $\Pr(\emptyset) = 0$.

(b) $\Pr(A) \leq 1$.

(c) $\Pr(A) = 1 - \Pr(A^c)$.

(d) $\Pr(B \setminus A) = \Pr(B) - \Pr(A \cap B)$.

(e) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

(f) If $A \subset B$ then $\Pr(A) \leq \Pr(B)$.

(A.2) The conditional probability of A given B is

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad \text{provided } \Pr(B) > 0.$$

This definition gives the multiplication rule $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)$. A conditional probability function $B \mapsto \Pr(B \mid A)$ is a probability function.

(A.3) Let A and B be two events, such that $0 < \Pr(B) < 1$. The law of total probability says that

$$\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c),$$

and if $B_1, \ldots, B_k$ is a partition of the sample space Ω, i.e., $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $B_1 \cup \cdots \cup B_k = \Omega$, then $\Pr(A) = \sum_{j=1}^{k} \Pr(A \mid B_j)\Pr(B_j)$.

(A.4) 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.5) Two events A and B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$. The events $A_1, \ldots, A_n$ are mutually independent (or just independent) if

$$\Pr(A_{j_1} \cap \ldots \cap A_{j_k}) = \Pr(A_{j_1}) \cdots \Pr(A_{j_k}),$$

for any subset $\{j_1, \ldots, j_k\}$ of $\{1, \ldots, n\}$.

(A.6) The expectation of a discrete random variable W taking the values $\{w_1, w_2, \ldots\}$ is

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

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

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

The variance of any random variable Z is

$$\mathrm{Var}(Z) = \mathrm{E}\,\{(Z - \mathrm{E}\,[Z])^2\} = \mathrm{E}\,\{Z^2\} - (\mathrm{E}\,\{Z\})^2.$$

(A.7) The indicator function of an event A, denoted I_A, is

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{otherwise.} \end{cases}$$

The expectation of an indicator function is $\mathrm{E}\,(I_A) = \Pr(A)$. For any two events A and B, the product of their indicator functions is $I_A I_B = I_{A\cap B}$.

(A.8) The cumulative distribution function of a random variable X is the function $F(x) = \Pr(X \leq x)$. If X is a discrete random variable taking the values $\{x_1, x_2, \ldots\}$, then

$$F(x) = \sum_{j:x_j \leq x} \Pr(X = x_j).$$

If X is a continuous random variable with p.d.f. $f(x)$, then

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y.$$

In particular, the p.d.f. of a continuous random variable can be found by differentiating its c.d.f., i.e.,

$$f(x) = F'(x).$$

(A.9) If $\widehat{\theta}_n$ is an estimator of θ, the mean squared error of $\widehat{\theta}_n$ is

$$\text{mse} = \mathrm{E}\,\{(\widehat{\theta}_n - \theta)^2\}.$$

It is always true that $\text{mse} = \mathrm{Var}(\widehat{\theta}_n) + \{\mathrm{E}\,(\widehat{\theta}_n) - \theta\}^2$. Notice that the mse is a function of θ. Small mse is good.

(A.10) Markov's inequality: Let X be a nonnegative random variable. For any $\varepsilon > 0$, we have that $\Pr(X \geq \varepsilon) \leq \mathrm{E}\,(X)/\varepsilon$. We deduce that for any random variable Y

$$\Pr(|Y| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\mathrm{E}\,(Y^2).$$

for any any $\varepsilon > 0$.