Universität Bielefeld
Tanja Pasurek
Tutor: Ehsan Abedi WiSe 2024/25

Foundations of Statistics

Homework 3

Part I. Conditional probability and Independence

1. (*Rolling a die*). We roll a die N times. Let A_{ij} be the event that the ith and jth rolls produce the same number. First define a proper probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to model this problem. Then show that the events A_{ij}, $1 \leq i < j \leq N$, are pairwise independent but not independent.

2. (*Friends and random numbers*). Four friends (**A**lex, **B**lake, **C**hris and **D**usty) each choose a random number between 1 and 5.

(a) What is the chance that at least two of them chose the same number? *Hint:* first find the probability of the complement event and for that, use tree diagram for calculating the probabilities.

(b) Perform a computer simulation in `R` playing this game $n = 1000$ rounds and estimating the probability p.
Hint: You can use the function `sample`.

Part II. Random variables and Expectation

3. (*Expectation of integer-valued random variables*).
Let $X : \Omega \to \mathbb{N}$ be an integer-valued random variable. Show that

$$\mathbb{E}(X) = \sum_{n \geq 1}^{\infty} \mathbb{P}(X \geq n).$$

4. (*Indicator random variable*).
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Given an event $A \in \mathcal{A}$, define the indicator random variable

$$\mathbb{I}_A(\omega) := \begin{cases} 1, & \text{if } A \text{ occurs} \ \ (\text{i.e. } \omega \in A), \\ 0, & \text{if } A \text{ does not occur} \ \ (\text{i.e. } \omega \notin A). \end{cases}$$

(a) Prove that for any $A, B \in \mathcal{A}$

$$\mathbb{I}_A^2 = \mathbb{I}_A, \quad \mathbb{I}_{A\cap B} = \mathbb{I}_A \mathbb{I}_B, \quad \mathbb{I}_{A\cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_A \mathbb{I}_B.$$

(b) Show that $\mathbb{I}_A \sim \mathrm{Ber}(p)$ where $p = \mathbb{P}(A)$.

(c) Check the fundamental relation $\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A)$.

(d) Suppose that a random variable $U : \Omega \to [0,1]$ has a uniform distribution, i.e. $U \sim \mathrm{Unif}(0,1)$. For some $0 < p < 1$ define a discrete random variable

$$X(\omega) := \begin{cases} 1, & \text{if } U(\omega) < p, \\ 0, & \text{if } U(\omega) \geq p. \end{cases}$$

Show that $X \sim \mathrm{Ber}(p)$ and that it allows the representation $X = \mathbb{I}_A$.

Part III. Correlation and Independence

5. Let $X_1, ..., X_n$ be a collection of independent random variables, all with mean zero and variance σ^2. Let $a_1, ..., a_n$ be a collection of real numbers such that $\sum_{i=1}^n a_i = 0$. Show that the sum $Y := \sum_{i=1}^n X_i$ and the linear combination $Z := \sum_{i=1}^n a_i X_i$ have zero covariance.

Part IV. Simulation in `R`

6. This exercise is about the casino game *Chuck-a-Luck* (also known as "*Glückswurf*").

This is a game of chance played with 3 standard dice. In the simplest variant, the rules are as follows:

- The player chooses one number, say a, from $\{1, 2, 3, 4, 5, 6\}$.
- The player pays a stake of \$1 and rolls three dice.
- If none of the dice show the number a, the bet is lost.
- If at least one of the dice shows the number a, the player receives the bet back and one additional dollar for each die that shows this number.

(a) Consider a random variable $X =$"player's profit" per game. Determine the probability mass function $f(x) := \mathbb{P}(X = x)$.

(b) Calculate the mean $\mathbb{E}(X)$. Is this game fair?

(c) Now use the `loop` function to simulate the game $n = 10\,000$ and $100\,000$ rounds. In the process we count how much profit we make overall and especially on average per game. You can proceed as follows:

```
nloop<-10000
a<-5
Win<-rep(NA,nloop)
for (k in 1:nloop){
 Dice<-sample(1:6,size=3,replace=TRUE)
 Count_a<-sum(Dice==a)
 Win[k]<-ifelse(Count_a==0,-1,Count_a)
}
sum(Win) ## overall
sum(Win)/nloop  ## on average per game
```

(d) With the following code, you can visualise the development of the average profit over the 100,000 runs.

```
options(scipen=999)
plot(cumsum(Win)/(1:nloop),type="l",bty="n",
ylab="Average Profit",xlab="Number of Rounds")
abline(h=-17/216,col=2,lty=2)
```

Remark: To set the use of *scientific notation* for large numbers ("*e* notation", e.g. `1e+05` instead of `10000`), you can use the `scipen` option. You can turn it off with `options(scipen = 999)` and back on again with `options(scipen = 0)`.