Introduction to Data Science (1MS041)
Uppsala University – Autumn 2024
Report for Assignment 1

Per Vincent Ankarbåge, Jonathan Falk, Madhur Gupta,
Henrik Jonasson, Adam Rokah

September 27, 2024

All group members contributed equally by individually taking on assigned problems (each name is specified in the title for their respective question), following up by collective discussion for fine-tuning the solutions and reporting.

1 Madhur

To show that the complements A^c and B^c are independent events, we need to prove that:

$$P(A^c \cap B^c) = P(A^c) \cdot P(B^c)$$

Given: A and B are independent events. By definition, independence means:

$$P(A \cap B) = P(A) \cdot P(B)$$

The complement of A (denoted A^c) is the event that A does not happen. Similarly, B^c is the event that B does not happen. Using the rule of complements:

$$P(A^c) = 1 - P(A) \quad \text{and} \quad P(B^c) = 1 - P(B)$$

By using De Morgan's Law, which states that the complement of the union is the intersection of the complements:

$$A^c \cap B^c = (A \cup B)^c$$

Therefore:

$$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B)$$

Using the formula for the union of two events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Since A and B are independent, we know that:

$$P(A \cap B) = P(A) \cdot P(B)$$

So:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

Now we can calculate $P(A^c \cap B^c)$:

$$P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A) \cdot P(B))$$

Simplifying:

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A) \cdot P(B)$$

We know:

$$P(A^c) = 1 - P(A) \quad \text{and} \quad P(B^c) = 1 - P(B)$$

Therefore:

$$P(A^c) \cdot P(B^c) = (1 - P(A)) \cdot (1 - P(B))$$

Expanding the product:

$$P(A^c) \cdot P(B^c) = 1 - P(A) - P(B) + P(A) \cdot P(B)$$

Notice that:

$$P(A^c \cap B^c) = P(A^c) \cdot P(B^c)$$

Therefore, the complements A^c and B^c are independent.

2 Adam

We are given that $P(Br) = 1/4$ and are given that there is independence between the three children involved.

2a

Define X as the number of brown haired children following the binomial probability distribution: $X \sim \text{Binomial}(n = 3, p = \frac{1}{4})$.

Moreover,

$$P(X \geq 2 \mid X \geq 1) = \frac{P(X \geq 2 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 2)}{P(X \geq 1)}$$

Hence, we would like to calculate:

$$\begin{aligned}
P(X = 2) &= \binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{9}{64} \\
P(X = 3) &= \left(\frac{1}{4}\right)^3 = \frac{1}{64} \\
\text{Therefore, } P(X \geq 2) &= \frac{10}{64} = \frac{5}{32} \\
P(X \geq 1) &= 1 - P(\text{no brown hair}) = 1 - \left(\frac{3}{4}\right)^3 = \frac{37}{64} \\
\text{Thus, } P(X \geq 2 \cap X \geq 1) &= \frac{P(X \geq 2)}{P(X \geq 1)} = \frac{10}{37}
\end{aligned}$$

2b

If it is known that the oldest child has brown hair, what is the probability that at least two children have brown hair?

Let the outcomes of each trial be T or F, where T is the outcome that a child has brown hair, and F be the outcome that a child does not have brown hair.

The sample space is $\Omega = \{TFF, FTF, FFT, TTF, TFT, FTT, FFF, TTT\}$ We want to find the conditional probability,

$$P(\text{at least two children have brown hair} \mid \text{the oldest child has brown hair})$$

$$= P(X \geq 2 \mid B)$$

. Where B denotes the event that the oldest child has brown hair.

To do this start by examining $P(B)$. From the sample space Ω we see that

$$\begin{aligned}
P(B) &= P(TTT, TFT, FTT, TTT) \\
&= P(TTT) + P(TFT) + P(FTT) + P(TTT) \\
&= \left(\frac{1}{4}\right)^3 + \left(2 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4}\right) + \left(\left(\frac{3}{4}\right)^2 \cdot \frac{1}{4}\right) \\
&= \frac{1}{4}
\end{aligned}$$

Now let's look at $P(X \geq 2 \cap B)$ that is, the probability that the oldest child has brown hair and at least two children have brown hair.

$$\begin{aligned}
P(X \geq 2 \cap B) &= P(TFT) + P(FTT) + P(TTT) \\
&= \left(\frac{1}{4}\right)^3 + \left(2 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4}\right) \\
&= \frac{7}{64}
\end{aligned}$$

Now we can calculate the conditional probability.

$$\begin{aligned}
P(X \geq | B) &= \frac{P(X \geq \cap B)}{P(B)} \\
&= 4 \cdot \frac{7}{64} \\
&= \frac{7}{16}
\end{aligned}$$

3 Per

Let (X, Y) be uniformly distributed on the unit disc, $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. Set $R = \sqrt{X^2 + Y^2}$ what is the CDF and PDF of R?

Let

$$F_R(r) = P(R \leq r)$$

.

Where $F_R(r)$ is the CDF of R. All points (x, y) are uniformally distributed on the unit circle, meaning that every point has an equal probability of occuring on the circle. Meaning that $P(R \leq r)$ is thus the probability that a point is shorter than r on the unit circle, yielding

$$\begin{aligned} &= \frac{\text{points with shorter distance from origin than } r}{\text{area of the unit circle}} \\ &= \frac{\pi r^2}{\pi 1^2} = \frac{\pi r^2}{\pi} \\ &= r^2 \end{aligned}$$

Since R is a continous random variable, we may find the PDF $f_R(r)$ by differentiating the CDF.

$$f_R(r) = \frac{d}{dr}(r^2) = 2r$$

4 Henrik

Let the coin yield heads with probability p and tails with probability $1 - p$. The probability that $X = x$ tosses are required until the first head appears is equal to the first $x - 1$ tosses yielding tails, and the last toss a head. As each coin toss is independent of the others, we simply multiply the probabilities of each toss to get:

$$P(X = x) = p(1 - p)^{x-1}$$

Now the expected value of our discrete variable X is given by the following sum over all possible values x of X:

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} xP(X = x) = \sum_{x=1}^{\infty} xp(1 - p)^{x-1}$$

To calculate the expected value, note that $px(1 - p)^{x-1} = p(-\frac{d}{dp}(1 - p)^x)$ if we view $x(1 - p)^{x-1}$ as a function over p. Using the linearity of differentiation and the fact that we can move the common factor p outside of the sum, we get that:

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} p(-\frac{d}{dp}(1 - p)^x) = p(-\frac{d}{dp}\sum_{x=1}^{\infty}(1 - p)^x)$$

Rewriting $(1 - p)^x$ as $(1 - p)(1 - p)^{x-1}$ we can use the sum of an infinite geometric series to get:

$$\sum_{x=1}^{\infty}(1 - p)^x = \sum_{x=1}^{\infty}(1 - p)(1 - p)^{x-1} = \sum_{x=0}^{\infty}(1 - p)(1 - p)^x = \frac{(1 - p)}{1 - (1 - p)} = \frac{(1 - p)}{p}$$

Now the quotient rule yields that:

$$-\frac{d}{dp}\frac{(1 - p)}{p} = -\frac{(-p - (1 - p))}{p^2} = \frac{1}{p^2}$$

Putting it all together, we get that:

$$\mathbb{E}(X) = p\frac{1}{p^2} = \frac{1}{p}$$

Finally, as the coin used was a fair coin, we have that $p = \frac{1}{2}$, which inserted in the expression above yields that $\mathbb{E}(X) = 2$.

5 Jonathan

5.1

Let $X_1, \ldots, X_n$ be i.i.d. from Bernoulli(P). Let $\alpha > 0$ be fixed and define

$$\varepsilon_n = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$$

Let

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and define the confidence interval $I_n = [\hat{p}_n - \varepsilon_n, \hat{p}_n + \varepsilon_n]$. Use Hoeffding's inequality to show that

$$P(p \in I_n) \geq 1 - \alpha.$$

We have:

$$P(p \in I_n) = P(\hat{p}_n - \varepsilon_n \leq p \leq \hat{p}_n + \varepsilon_n)$$

$$= P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) = 1 - P\left(|\hat{p}_n - p| \geq \varepsilon_n\right).$$

By Hoeffding's inequality: $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(P) with $P(X_i \in [0, 1]) = 1$, then for any $\varepsilon_n > 0$, we get for

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

that

$$P\left(|\hat{p}_n - \mathbb{E}[\hat{p}_n]| \geq \varepsilon_n\right) \leq 2e^{-2n\varepsilon_n^2}.$$

We know that

$$\mathbb{E}[\hat{p}_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} np = p,$$

so

$$P\left(|\hat{p}_n - p| \geq \varepsilon_n\right) \leq 2e^{-2n\varepsilon_n^2}.$$

We have

$$\varepsilon_n = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}.$$

Thus,

$$P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) \leq 2e^{-2n\left(\frac{1}{2n}\log\frac{2}{\alpha}\right)}.$$

Simplifying,

$$P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) \leq 2e^{-\log\frac{2}{\alpha}}$$

$$P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) \leq 2\frac{\alpha}{2}$$

$$P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) \leq \alpha.$$

Thus,

$$P(p \in I_n) = 1 - P\left(|\hat{p}_n - p| \leq \varepsilon_n\right) \geq 1 - \alpha. \quad \square$$

5.2

The code used for all parts of the question is

```
import random
import matplotlib.pyplot as plt
import math
import numpy as np

p_old = 0.4
p_new = 0.5
alpha = 0.05

def toss(n,p):
    num_of_heads = 0
    for tosses in range (n):
        rand = random.random()
        if (rand < p):
            num_of_heads = num_of_heads + 1
    return num_of_heads

def coverage(n):
    num_times_p_old_in_I = 0
    num_times_p_new_in_I = 0
    epsilon_n = math.sqrt((1/(2*n))*math.log(2/alpha))
    for i in range(n):
        p_hat = toss(n,p_old)/n
        I_n = np.linspace(p_hat - epsilon_n, p_hat + epsilon_n, 100000)
        if I_n[0] <= p_old <= I_n[-1]:
            num_times_p_old_in_I += 1
        if I_n[0] <= p_new <= I_n[-1]:
            num_times_p_new_in_I += 1
    return num_times_p_old_in_I, I_n[-1]-I_n[0], num_times_p_new_in_I
```

```
N = (10,100,1000,10000)
C = [0]*len(N)
L = [0]*len(N)
P = [0]*len(N)
i = 0
for n in N:
    (A,B,D) = coverage(n)
    C[i] = A/n
    L[i] = B
    P[i] = D/n
    i = i+1

plt.subplot(3, 1, 1)
plt.plot(N, C)
plt.title("Coverage of old p")

plt.subplot(3, 1, 2)
plt.plot(N, L)
plt.title("Size of Interval")

plt.subplot(3, 1, 3)
plt.plot(N, P)
plt.title("Coverage of new p")

plt.tight_layout()
```

Evaluating this code gives us the following graph of the coverage for $p = 0.4$ and $\alpha = 0.05$.

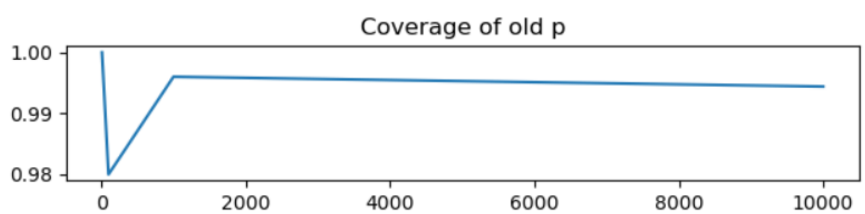

Figure 1: Coverage of $p = 0.4$.

As we can see, for small n (e.g. $n = 10$), the coverage is practically 1 as the confidence interval I_n is larger and thus very likely to include p. For $n = 100$ we see that the coverage decreases to about 0.98, indicating that $p \in I_n$ about 98% of the time. As we increase n from here on out, the coverage seems to plateau at around 99.5%.

5.3

The size of the confidence interval plotted over n results in the following.

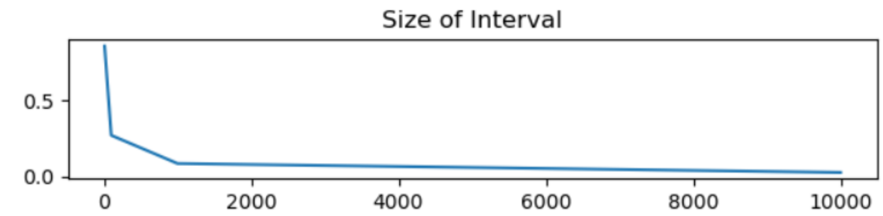

Figure 2: Size of I_n.

As we can see, The size of the interval is large for small n and decreases as n increases. For $n = 100$ we see that the size of I_n is around 0.25 and around 0.1 at $n = 1000$. It then slowly approaches 0 as n gets large.

5.4

If we don't know that p has changed from 0.4 to 0.5, what is the probability that our new confidence interval I_n will contain $p = 0.4$? That is, what is the probability that our decision that $p = 0.4$ is in I_n is correct? We get the following graph.

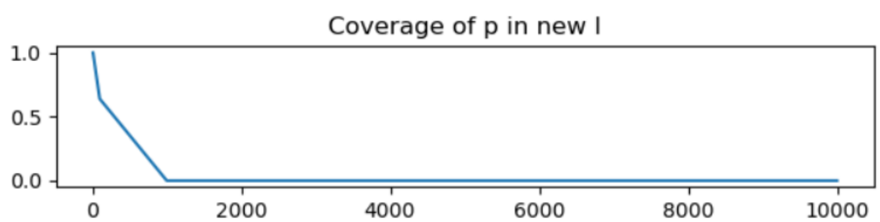

Figure 3: Coverage of $p = 0.4$ in new I_n.

As we can see, for small n, $p = 0.4$ is often in I_n since the interval itself is large. However, as n increases, the likelihood that $p = 0.4$ is in I_n decreases. For $n = 100$, the probability is quite large at around 0.6. This means that even at $n = 100$, we cannot be certain that our decision is correct or not. That being said, at $n = 1000$, the probability that $p = 0.4 \in I_n$, i.e. the probability that our decision is correct, is practically 0.