Name and surname: Identification number:

University of Primorska
FAMNIT, Mathematics

Probability

Written examination

August 21$^{\text{th}}$, 2024

Instructions

Read carefully the text of the problems before attempting to solve them. Five problems out of six count for 100%. You are allowed one A4 sheet with formulae and theorems. You have 120 minutes.

Question	a.	b.	c.	d.	Total
1.			•	•	
2.			•	•	
3.			•	•	
4.			•	•	
5.			•	•	
6.			•	•	
Total					

1. (20) We toss n balls into r boxes. Assume the tosses are independent and every box is hit with equal probability. Denote by X the number of empty boxes at the end.

a. (10) Define the events

$$A_k = \{k\text{-th box is empty}\}$$

for $k = 1, 2, \ldots, r$. Express the event $A = \{X = 0\}$ with events A_k.

Solution: $A = \cap_{k=1}^{r} A_k^c$.

b. (10) Compute $P(X = 0)$.

Solution: note that $A^c = \cup_{k=1}^{r} A_k$. *We will use the inclusion-exclusion formula and to do that we need probabilities* $P(A_1 \cap \cdots \cap A_k)$ *for all* k. *In other words, we are computing the probability that on every toss we hit the other* $r - k$ *boxes. By independence we have*

$$P(A_1 \cap \cdots \cap A_k) = \left(\frac{r-k}{r}\right)^n .$$

By symmetry the intersection of any k *events among* $A_1, \ldots, A_r$ *has equal probability, therefore*

$$P(A) = 1 - P(A^c) = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \left(\frac{r-k}{r}\right)^n .$$

Note that for $n < r$ *we have* $P(A) = 1$.

2. (20) We roll a fair die until the first time six appears. We assume that the rolls are independent. Let X_i be the number of appearences of outcome i before the first six for $i = 1, 2, 3, 4, 5$. Let X be the number of rolls up to and including the first six.

a. (10) find the distribution of the pair (X_i, X).

Solution: the possible values of (X_i, X) are the pairs (k, l) with $l \geq 1$ and $0 \leq k \leq l - 1$. The event $\{X_i = k, X = l\}$ occurs in the following disjoint ways: the first $l - 1$ rolls result in exactly k apperances of i and no six, and the l-th roll is a six. We can choose the positions for the outcome i in $\binom{l-1}{k}$ ways. The outcomes on the ramaining $l - 1 - k$ positions are different from i and six. By independence we have

$$P\left(X_i = k, X = l\right) = \binom{l-1}{k} \left(\frac{1}{6}\right)^k \cdot \left(\frac{4}{6}\right)^{l-1-k} \cdot \frac{1}{6}\,.$$

The expression simplifies to

$$P\left(X_i = k, X = l\right) = \binom{l-1}{k} \cdot \frac{4^{l-1-k}}{6^l}\,.$$

We interpret $\binom{0}{0} = 1$.

b. (10) Compute $\operatorname{cov}(X_i, X)$.

Solution: by symmetry the covariances $\operatorname{cov}(X_i, X)$ are equal for $i = 1, 2, 3, 4, 5$. We have

$$\operatorname{cov}(X_1 + X_2 + X_3 + X_4 + X_5, X) = 5 \cdot \operatorname{cov}(X_1, X)\,.$$

On the other hand, $X_1 + X_2 + X_3 + X_4 + X_5 = X - 1$, and therefore

$$\operatorname{cov}(X_1 + X_2 + X_3 + X_4 + X_5, X) = \operatorname{cov}(X - 1, X) = \operatorname{var}(X)\,.$$

Since $X \sim \text{Geom}(\frac{1}{6})$, we have

$$\operatorname{var}(X) = 30\,,$$

and hence

$$\operatorname{cov}(X_1, X) = \operatorname{cov}(X_i, X) = 6\,.$$

3. (20)The *Beta* distribution with parameters $a, b > 0$ is given by the density

$$f(x) = \frac{1}{\mathrm{B}(a,b)}\, x^{a-1}(1-x)^{b-1}$$

for $0 < x < 1$, where

$$\mathrm{B}(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx\,.$$

a. (10) Let X and Y be independent and

$$X \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) \quad \text{and} \quad Y \sim \mathrm{Beta}\left(1, \frac{1}{2}\right)\,.$$

Show that the density of the vector $(U,V) = (X, XY)$ is given by

$$f_{U,V}(u,v) = f_X(u)\, f_Y(v/u)\, \frac{1}{u}$$

for $0 < v < u < 1$. Compute the density XY explicitly. Assume as known

$$\int_v^1 \frac{1}{u}(1-u)^{-1/2}(u-v)^{-1/2}du = \frac{\pi}{\sqrt{v}}\,.$$

Solution: the mapping $\Phi(x,y) = (x, xy)$ is bijective and differentiable on $(0,1)^2$ and maps bijectively onto $\{(u,v)\colon 0 < u < v < 1\}$. We have $J_{\Phi^{-1}}(u,v) = u^{-1}$. The density of (X, XY) follows from the transformation formula. The density $V = XY$ is computed as the marginal density.

$$\begin{aligned}
&f_V(v)\\
&= \int_v^1 f_X(u)\, f_Y(v/u)\, \frac{1}{u}\, du\\
&= \frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{1}{2}\right)\mathrm{B}\left(1, \frac{1}{2}\right)} \int_v^1 u^{-\frac{1}{2}}(1-u)^{\frac{1}{2}-1}\left(1 - \frac{v}{u}\right)^{-\frac{1}{2}} \frac{1}{u}\, du\\
&= \frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{1}{2}\right)\mathrm{B}\left(1, \frac{1}{2}\right)} \int_v^1 u^{-1}(1-u)^{-\frac{1}{2}}(u-v)^{-\frac{1}{2}}\, du\\
&= \frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{1}{2}\right)\mathrm{B}\left(1, \frac{1}{2}\right)} \cdot \frac{\pi}{\sqrt{v}}\\
&= \frac{1}{2\sqrt{v}}\,.
\end{aligned}$$

In the last row the integral is the one given above. The constant equals $\frac{1}{2}$, since the result is a density. We can write $XY \sim \mathrm{Beta}\left(1, \frac{1}{2}\right)$.

b. (10) Let X and Y be independent with

$$X \sim \mathrm{Beta}(a,b) \quad \text{and} \quad Y \sim \mathrm{Beta}(a+b, c)$$

for positive constants a, b and c. Compute the density of XY. Assume as known that

$$\int_v^1 \frac{(1-u)^{b-1}(u-v)^{c-1}du}{u^{b+c}} = \mathrm{B}(b,c)\, v^{-b}(1-v)^{b+c-1}\,.$$

Solution: we have

$$\begin{aligned}
&f_V(v)\\
&= \frac{1}{\mathrm{B}(a,b)\mathrm{B}(a+b,c)} \int_v^1 u^{a-1}(1-u)^{b-1}\left(\frac{v}{u}\right)^{a+b-1}\left(1-\frac{v}{u}\right)^{c-1}\frac{1}{u}du\\
&= \frac{v^{a+b-1}}{\mathrm{B}(a,b)\mathrm{B}(a+b,c)} \int_v^1 \frac{(1-u)^{b-1}(u-v)^{c-1}\,du}{u^{b+c}}\\
&= \frac{\mathrm{B}(b,c)}{\mathrm{B}(a,b)\mathrm{B}(a+b,c)}\, v^{a-1}(1-v)^{b+c-1}\,.
\end{aligned}$$

Therefore $V \sim \mathrm{Beta}(a, b+c)$.

4. (20) Random number generators generate random sequences of zeros and ones. Assume that the generated numbers are independent and that each generated number equals 1 with probability 1/2.

a. (10) When the quality of a random number generator of random numbers is tested, the random variable Y is defined which counts the numbers of appearances of two consecutive ones in the string of n random zeros and ones. The overlapping occurencies are allowed in the sense that in 1011011110111 there are six appearances of two consecutive ones. Compute $E(Y)$.

Solution: define indicators

$$I_k = \begin{cases} 1 & \text{if there are ones in places } k \text{ and } k+1 \\ 0 & \text{otherwise.} \end{cases},$$

for $k = 1, 2, \ldots, n-1$. *We have* $Y = I_1 + I_2 + \cdots + I_{n-1}$, *and*

$$E(I_k) = P(I_k = 1) = \frac{1}{4}.$$

It follows that

$$E(Y) = \frac{n-1}{4}$$

b. (10) Let Z be the number of appearances of the sequence 011 in the set of n generated random numbers, where overlapping occurencies are not allowed. Compute $E(Z)$ and $\mathrm{var}(Z)$.

Solution: two overlapping sequences of the length 3 cannot both be 011, so we define

$$I_k = \begin{cases} 1 & \text{if in places } k,\ k+1 \text{ and } k+2 \text{ there are the numbers 011}, \\ 0 & \text{otherwise.} \end{cases}$$

for $k = 1, 2, \ldots, n-2$. *We have* $Z = I_1 + I_2 + \cdots + I_{n-2}$. *By independence* $E(I_k) = 1/8$, *and hence*

$$E(Z) = \frac{n-2}{8}.$$

We will need covariances for computing the variance. If $l - k \leq 2$, *then* $I_l I_k = 0$ *with probability* 1 *and therefore* $\mathrm{cov}(I_k, I_l) = -E(I_k)E(I_l) = -1/64$. *For* $l-k \geq 3$ *the indicators* I_k *and* I_l *are independent, therefore their covariace equals* 0. *We compute*

$$\begin{aligned} \mathrm{var}(Z) &= \sum_{k=1}^{n-2} \mathrm{var}(I_k) + 2 \sum_{1 \leq k < l \leq n-2} \mathrm{cov}(I_k, I_l) \\ &= \frac{(7(n-2)}{64} - \frac{6(n-4)}{64} \\ &= \frac{n+10}{64}. \end{aligned}$$

5. (20) We deal cards one by one from a well shuffled deck of 52 standard cards Let N by the number of cards up to and including the firdst ace, and let X be the number of Kings appearing before the first ace.

a. (10) Justify that

$$P(X = k|N = n) = \frac{\binom{4}{k}\binom{44}{n-k-1}}{\binom{48}{n-1}}\,.$$

for $n = 1, 2, \ldots, 49$ and $k = 0, 1, \ldots, \min(4, n-1)$. Thus, the conditional distribution of X given $\{N = n\}$ is $\mathrm{HiperGeom}(n-1, 4, 48)$.

Solution: conditional on $\{N = n\}$, the first $n = 1$ cards are a random sample of $n - 1$ cards out of 48 cards that are not aces. There are 4 Kings among tese 48 cards.

b. (10) Express $E(X)$ with $E(N)$.

Solution: by the formula for total expectation we have

$$E(X) = \sum_{n=1}^{49} E(X|N = n)P(N = n)\,.$$

From the first part we have $E(X|N = n) = (n-1) \cdot \frac{4}{48}$. It follows that

$$E(X) = \frac{4}{48}\sum_{n=1}^{49}(n-1)P(N = n) = \frac{E(N) - 1}{12}\,.$$

6. (20) Players A and B take turns tossing a fair coin. We assume that all the tosses are independent. First A tosses the coin until he gets heads. Then B tosses the coin until he gets heads. They repeat the rounds of each tossing the coin until heads.

a. (10) Compute, approximately, the probability that the players will together toss the coin 4081 times or more in 1000 rounds of the game.

Solution: the number of tosses until heads is $\mathrm{Geom}(1/2)$. *If* $X_1 \sim \mathrm{Geom}(1/2)$, *we have* $E(X_1) = 1/p = 2$ *and* $\mathrm{var}(X_1) = q/p^2 = 2$. *In* 1000 *rounds the number of tosses is*

$$S_{2000} = X_1 + X_2 + \cdots + X_{2000}\,,$$

where X_i *are independent, equally distributed random variables. By CLT*

$$\begin{aligned} P(S_{2000} \geq 4081) &= P\left(\frac{S_{2000} - E(S_{2000})}{\sqrt{\mathrm{var}(S_{2000})}} \geq \frac{4081 - E(S_{2000})}{\sqrt{\mathrm{var}(S_{2000})}}\right) \\ &= P\left(\frac{S_{2000} - E(S_{2000})}{\sqrt{\mathrm{var}(S_{2000})}} \geq \frac{4081 - 4000}{\sqrt{4000}}\right) \\ &\approx P(Z \geq 1.28) \\ &= 1 - \Phi(1.28) \\ &\doteq 0.1\,. \end{aligned}$$

b. (10) Player A wins if after 1000 rounds the total number of tosses of Player B exceeds the total number of tosses of Player A by 100 or more. Compute, approximately, the probability that Player A wins.

Hint: look at differences.

Solution: in the notation of the first part we are computing the probability

$$P\left((X_2 + X_4 + \cdots + X_{2000}) - (X_1 + X_3 + \cdots + X_{1999}) \geq 100\right)\,.$$

Denote $Z_k = X_{2k} - X_{2k-1}$, *and rewrite the above probability as*

$$P\left(Z_1 + Z_2 + \cdots + Z_{1000} \geq 100\right)\,.$$

Note that $E(Z_k) = 0$ *and* $\mathrm{var}(Z_k) = 4$. *Denote* $T_n = Z_1 + Z_2 + \cdots + Z_n$. *By CLT we have*

$$\begin{aligned} P(T_{1000} \geq 100) &= P\left(\frac{T_{1000} - E(T_{1000})}{\sqrt{\mathrm{var}(T_{1000})}} \leq \frac{-100 - E(T_{1000})}{\sqrt{\mathrm{var}(T_{1000})}}\right) \\ &\approx P\left(Z \leq \frac{-100}{\sqrt{4000}}\right) \\ &= P(Z \leq -1.58) \\ &\doteq 0.057\,. \end{aligned}$$