University of Primorska

FAMNIT, Mathematics

Probability

Exam

August 20th, 2018

Name and surname: ______________________ Identification number: | | | | | | | | |

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 two hours.

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

1. (20) We select a subset of size n out of a set of size N with replacement and repeatedly r-times. Every subset is equally likely to be selected. Subsequent selections are independent of each other..

a. (5) What is the probability that a given fixed k elements are contained in every subset selected?

Solution: $\left[\frac{\binom{N-k}{n-k}}{\binom{N}{n}}\right]^r$ *(with agreement that* $\binom{m}{s} = 0$*, when* $s < 0$ *or* $s > m$*).*

b. (15) What is the probability that none of the elements are contained in every subset selected ? You do not need to simplify the expression obtained.

Hint: define the events

$$A_i = \{\textit{the } i\textit{-th element is contained in every subset selected }\}.$$

Solution: For $i = 1, 2, \dots, N$ *let* A_i *be an event, where* i*-th element is the chosen element in every subset. It was computed above that for arbitrary distinguish* $i_1, i_2, \dots, i_k$ *holds*

$$P\big(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\big) = \left[\frac{\binom{N-k}{n-k}}{\binom{N}{n}}\right]^r .$$

With usage of the inclusions and exclusions one gets

$$\begin{aligned}
&P\big(A_1^c \cap A_2^c \cap \cdots \cap A_N^c\big) \\
&\quad = 1 - P\big(A_1 \cup A_2 \cup \cdots \cup A_N\big) \\
&\quad = 1 - \sum_{i_1=1}^{N} P(A_{i_1}) + \sum_{1 \le i_1 < i_2 \le N} P(A_{i_1} \cap A_{i_2}) \\
&\qquad - \sum_{1 \le i_1 < i_2 < i_3 \le N} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \cdots \\
&\qquad + (-1)^n \sum_{1 \le i_1 < i_2 < \cdots < i_n \le N} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) \\
&\quad = \sum_{k=0}^{n} (-1)^k \binom{N}{k} \left[\frac{\binom{N-k}{n-k}}{\binom{N}{n}}\right]^r .
\end{aligned}$$

2. (20) Let consider a scyscraper with infinitely many floors numbered by $k = 0, 1, \ldots$. On floor 0 X_0 people enter the elevator where $X_0 \sim \mathrm{Po}(\lambda)$. On every subsequent floors the elevator stops and every individual who is still in the elevator exits with probability $\frac{1}{2}$, independently of the others, independently of previous exits and independently of X_0. Let X_k be the number of people in the elevator after the elevator leaves the k-th floor for $k = 0, 1, 2, \ldots$.

a. (10) Find the distribution of X_1.

Hint: we have

$$\begin{aligned} & P(X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n) \\ = \quad & P(X_0 = k_0, \ldots, X_{n-1} = k_{n-1}) P(X_n = k_n | X_0 = k_0, \ldots, X_{n-1} = k_{n-1}) \,. \end{aligned}$$

Solution: By assumptions

$$X_n|_{X_0=k_0,\ldots,X_{n-1}=k_{n-1}} \sim \mathrm{Bin}\left(k_{n-1}, \frac{1}{2}\right) \,,$$

where for $k_{n-1} = 0$ *we can interprate binomial distribution as a constant* 0. *For* $k = 1$ *we get*

$$\begin{aligned} P(X_1 = k_1) &= \sum_{l=k_1}^{\infty} P(X_0 = l) P(X_1 = k_1 | X_0 = l) \\ &= \sum_{l=k_1}^{\infty} \frac{e^{-\lambda}\lambda^l}{l!} \cdot \binom{l}{l-k_1} \left(\frac{1}{2}\right)^l \\ &= \frac{e^{-\lambda}}{k_1!} \left(\frac{\lambda}{2}\right)^{k_1} \sum_{l=k_1}^{\infty} \frac{(\lambda/2)^{l-k_1}}{(l-k_1)!} \\ &= \frac{e^{-\lambda}}{k_1!} \left(\frac{\lambda}{2}\right)^{k_1} e^{\lambda/2} \\ &= \frac{e^{-\lambda/2}(\lambda/2)^{k_1}}{k_1!} \,. \end{aligned}$$

It follows that $X_1 \sim \mathrm{Po}(\lambda/2)$.

b. (10) Let M be the number of the floor on which the last individuals in the elevator exit. Find the distribution of M.

Hint: find the distribution of X_m by induction.

Solution: It holds $P(M \leq m) = P(X_m = 0)$ for $m = 0, 1, \ldots$. Similarly as in the first part the induction can be used and it holds $X_m \sim \mathrm{Po}\,(2^{-m}\lambda)$, which means

$$P(X_m = 0) = e^{-2^{-m}\lambda}\,.$$

It follows

$$P(M = 0) = e^{-\lambda}$$

and

$$P(M = m) = P(M \leq m) - P(M \leq m-1) = e^{-2^{-m}\lambda} - e^{-2^{-m+1}\lambda}$$

for $m = 1, 2, \ldots$.

3. (20) Let random variables U and V be independent and uniformly distributed on the interval $(0, 1)$.

a. (10) Compute the density of the random vector $(X, Y) = (U, V(1 - U))$.

Solution: In transformation formula we take

$$\Phi(u, v) = (u, v(1 - u)) .$$

We compute

$$\Phi^{-1}(x, y) = \left(x, \frac{y}{1 - x}\right)$$

and

$$J_{\Phi^{-1}}(x, y) = \frac{1}{1 - x} .$$

It follows

$$f_{X,Y}(x, y) = \frac{1}{1 - x}$$

for $x, y \in (0, 1)$ *and* $x + y < 1$.

b. (10) Compute the density of the random variable $Z = U + V(1 - U)$.

Solution: We can use the formula

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)dx .$$

In our case we integrate just on the interval where the density is nonzero. It follows

$$\begin{aligned} f_Z(z) &= \int_0^z \frac{dx}{1 - x} \\ &= -\log(1 - z) . \end{aligned}$$

The formula for density holds for $0 < z < 1$.

4. (20) A group of $n \geq 3$ gamblers are sitting around a round table. All of them roll their own dice once; all dice are standard (1 to 6 dots), fair (every number of dots has equal probability) and the rolls are independent. Denote by W the number of pairs of neighbouring gamblers at the table who roll a neigboring number of dots. The numbers from the set $\{1, 2, 3, 4, 5, 6\}$ are neighbouring numbers if their difference is 1 (3 and 4 are neighboring numbers, but 6 and 1 are not, and also 3 and 3 are not).

a. (10) Compute $E(W)$ and $\mathrm{var}(W)$.

Hint: indicators.

Solution: We can write $W = I_1 + I_2 + \cdots + I_n$, where I_i is indicator for the event where i-th gambler and his right neighbour roll a neighbouring numbers. The probability od+f this event is:

$$E(I_i) = \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{6} = \frac{5}{18},$$

and

$$E(W) = \frac{5n}{18}.$$

For computing the variance, there are two standard approaches. We can start by

$$\mathrm{var}(W) = E(W^2) - \big(E(W)\big)^2$$

and

$$E(W^2) = \sum_{i=1}^{n} \sum_{j=1}^{n} E(I_i I_j).$$

The random variable $I_i I_j$ is an indicator of the event where for i-th and j-th gambler holds that they and their right neighbours roll a neighbouring numbers. For $i = j$ the probability of this event is equal to $5/18$; there are n such terms in the above double sum. If the i-th and j-th gambler are neighbours, we are computing the probability that the rolls of three neighbouring gamblers forms a chain of neighbouring numbers. The probability of this event equals to

$$E(I_i I_j) = \frac{2}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{1}{36} = \frac{1}{12};$$

there are $2n$ such terms in the above sum. If i and j are neither equal nor neighbouring, the probability of this event equals to $(5/18)^2 = 25/324$; there are $n^2 - 3n$ such terms in above sum (here we need the assumption that $n \geq 3$). We sum up and get:

$$E(W^2) = n \cdot \frac{5}{18} + 2n \cdot \frac{1}{12} + (n^2 - 3n) \cdot \frac{25}{324} = \frac{25n^2 + 69n}{324}.$$

We substract and get:

$$\mathrm{var}(W) = \frac{23n}{108}\,.$$

The result can be derived also by covariances:

$$\mathrm{var}(W) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(I_i, I_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(I_i I_j) - E(I_i)E(I_j)\,.$$

For $i = j$ *je* $\mathrm{cov}(I_i, I_j) = 5/18 - (5/18)^2 = 65/324$. *If* i*-th and* j*-th gambler are neighbouring, the covariance equals* $\mathrm{cov}(I_i, I_j) = 1/12 - 25/324 = 1/162$. *In every other case it holds* $\mathrm{cov}(I_i, I_j) = 0$*: the events that the neigbours of* i*-th and the neigbours of* j*-th gambler roll neighboring numbers are independent. We sum up and get:*

$$\mathrm{var}(W) = n \cdot \frac{65}{324} + 2n \cdot \frac{1}{162} = \frac{23n}{108}\,,$$

which is the same as before.

b. (10) Let S be the number of gamblers who roll a six. Compute $\mathrm{cov}(W, S)$.

Solution: We write $S = J_1 + J_2 + \cdots + J_n$*, where* J_j *is the indicator of the event, that* j*-th gambler rolls six dots. For computing the covariance of* W *and* S *there are two standard ways again. We can set:*

$$\mathrm{cov}(W, S) = E(WS) - E(W)\,E(S)$$

and

$$E(WS) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(I_i J_j)\,.$$

The random vatiable $I_i I_j$ *is an indicator of the event where* i*-th gambler and his rught neigbour roll neigbouring number of dots and*j*-th gambler rollssix dots. For*$i = j$ *this is an event where* i*-th gambler rolls six dots and his right neigbour rolls five dots; the probability of this event is equal to* $1/36$ *and there are* n *such term in the above double sum. If the* j*-th gambler is the right neigbour of the* i*-th gambler, this is the event where* j*-th gambler roll six and the* i*-th gambler rolls five; the probability of this event is* $1/36$ *again and there are* n *such term in the above double sum. If the* j*-th gambler is neither equal to the* i*-th and nor is his right neighbour, the probability of this event is equal to* $(1/6) \cdot (5/18) = 5/108$*; there are* $n^2 - 2n$ *such term in the above sum. We sum up and get:*

$$E(WS) = 2n \cdot \frac{1}{36} + (n^2 - 2n) \cdot \frac{5}{108} = \frac{5n^2 - 4n}{108}\,.$$

It is obvious that $E(J_j) = 1/6$*and consecuently* $E(S) = n/6$. *We substract and get:*

$$\mathrm{var}(W) = -\frac{n}{27}\,.$$

We can aigain use the approach with covariances:

$$\mathrm{cov}(W) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(I_i, J_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} E(I_i J_j) - E(I_i)E(J_j)\,.$$

If $i = j$ *or the* j*-th gambler is the right neighbour of the* i*-th, it holds* $\mathrm{cov}(I_i, J_j) = 1/36 - 5/108 = -1/54$, *otherwise* $\mathrm{cov}(I_i, J_j) = 0$. *We sum up and get*

$$\mathrm{cov}(W, S) = -\frac{n}{27}\,,$$

which is the same as before.

5. (20) Let X be a random variable with values in $\{2, 3, \ldots\}$ and with distribution

$$P(X = k) = (k-1)p^2(1-p)^{k-2}$$

for $p \in (0,1)$. Assume as known, that for $|x| < 1$

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}\,.$$

a. (10) Show that the probability generating function of the random variable X equals

$$G_X(s) = \left(\frac{ps}{1-(1-p)s}\right)^2\,.$$

Solution: We compute

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} s^k P(X = k) \\ &= \sum_{k=2}^{\infty} s^k \cdot (k-1)p^2(1-p)^{k-2} \\ &= \frac{p^2 s}{1-p} \sum_{k=2}^{\infty} (k-1)\left(s(1-p)\right)^{k-1} \\ &= \frac{p^2 s}{1-p} \cdot \frac{s(1-p)}{(1-(1-p)s)^2} \\ &= \left(\frac{ps}{1-(1-p)s}\right)^2\,. \end{aligned}$$

b. (10) Compute $\mathrm{var}(X)$.

Solution: We compute

$$G_X'(s) = 2\left(\frac{ps}{1-(1-p)s}\right) \cdot \frac{p(1-(1-p)s) + ps(1-p)}{(1-(1-p)s)^2} = \frac{2p^2 s}{(1-(1-p)s)^3}\,.$$

It follows

$$E(X) = G_X'(1) = \frac{2}{p}\,.$$

We compute

$$G_X''(s) = 2p^2 \cdot \frac{(1-(1-p)s)^3 + 3s(1-p)(1-(1-p)s)^2}{(1-(1-p)s)^6} = \frac{2p^2(1+2(1-p)s)}{(1-(1-p)s)^4}\,.$$

It follows

$$E\left(X(X-1)\right) = G_X''(1) = \frac{2(1+2(1-p))}{p^2}\,.$$

We get

$$\begin{aligned}
\mathrm{var}(X) &= E(X^2) - E(X)^2 \\
&= E(X(X-1)) + E(X) - E(X)^2 \\
&= \frac{2(1+2(1-p))}{p^2} + \frac{2}{p} - \frac{4}{p^2} \\
&= \frac{2(1+2(1-p)) + 4p - 4}{p^2} \\
&= \frac{2+4-4p+2p-4}{p^2} \\
&= \frac{2(1-p)}{p^2}\,.
\end{aligned}$$

6. (20) Let $f\colon [0,1] \to \mathbb{R}$ be integrable function and denote $I = \int_0^1 f(x)\,dx$ and $v^2 = \int_0^1 f^2(x)\,dx$. The idea of the Monte-Carlo method for computing integrals is to generate independent random variables $X_1, X_2, \ldots$ with uniform distribution on $[0,1]$ by computer and to compute the sum

$$A_n = \frac{1}{n}\sum_{k=1}^{n} f(X_k)\,.$$

a. (10) Let $f(x) = x$. Compute $E(A_n)$ and $\mathrm{var}(A_n)$.

Solution: We compute

$$\begin{aligned} E(A_n) &= \frac{1}{n}\sum_{k=1}^{n} E(X_k) \\ &= \frac{1}{n}\sum_{k=1}^{n}\int_0^1 x\,\mathrm{d}x \\ &= \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}(A_n) &= \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{var}(X_k) \\ &= \frac{1}{n}\mathrm{var}(X_1) \\ &= \frac{1}{n}\left(\int_0^1 x^2\,\mathrm{d}x - \left(\int_0^1 x\,\mathrm{d}x\right)^2\right) \\ &= \frac{1}{n}\left(\frac{1}{3}-\frac{1}{4}\right) \\ &= \frac{1}{12n} \end{aligned}$$

b. (15) Let $f(x) = x$. Denote $I = 1/2$. How large should n be so that $P(I - 0,01 \leq A_n \leq I + 0,01) \geq 0,99$ will hold?

Solution: We apply the central limit theorem. Let denote $S_n = \sum_{k=1}^{n} X_k$ *and* $\sigma^2 = \mathrm{var}(X_k)$. *We know that it holds* $E(X_k) = I = 1/2$ *and* $\sigma^2 = \mathrm{var}(X_k) =$

$1/12$. *We can evaluate*

$$\begin{aligned} P(I-0,01 \leq A_n \leq I+0,01) &= P(-0,01 \leq A_n - I \leq 0,01) \\ &= P(-0,01 \leq \frac{S_n - nI}{n} \leq 0,01) \\ &= P\left(-0,01 \cdot \frac{\sqrt{n}}{\sigma} \leq \frac{S_n - nI}{\sqrt{n}\sigma} \leq 0,01 \cdot \frac{\sqrt{n}}{\sigma}\right) \\ &\approx P(0,01 \cdot \sqrt{12n} \leq Z \leq 0,01 \cdot \sqrt{12n})\,. \end{aligned}$$

It must hold $0,01 \cdot \sqrt{12n} \geq 2,56$. *It follows* $n = 5.461$.