Name and surname: ____________________

Identification number: | | | | | | | | |

University of Primorska

FAMNIT, Mathematics

Probability

Midterm 2

June 4th, 2021

## 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, and a handbook of mathematics. Time allowed: 120 minutes.

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

**1.** (20) An urn contains $B$ black and $R$ red balls. We select balls at random without replacement until we pick the first red ball at which point we stop. Let $X$ be the number of black balls selected.

a. (10) Compute $E(X)$.

*Hint: use indicators.*

*Solution: number all the black balls with $k = 1, 2, \ldots, B$. Letting*

$$I_k = \begin{cases} 1 & \textit{if the } k\textit{-th black ball comes before the first red ball} \\ 0 & \textit{else,} \end{cases}$$

*we have*

$$X = \sum_{k=1}^{B} I_k \,.$$

*By symmetry, all indicators have the same expectation. To compute the probability $P(I_1 = 1)$, note that only the relative positions of the $k$-th black ball and the $R$ red balls are relevant. Since all the permutations are equally likely we get*

$$P(I_1 = 1) = \frac{1}{R+1} \,,$$

*and as a consequence*

$$E(X) = \frac{B}{R+1} \,.$$

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

*Solution: we use the indicators from the first part. First observe that*

$$\mathrm{var}(I_k) = \frac{R}{(R+1)^2}$$

*for all $k$. Again by symmetry, all the covariances $\mathrm{cov}(I_k, I_l)$ for $k \neq l$ are also equal. We find $P(I_k = 1, I_l = 1)$ by considering relative positions of the $k$-th and $l$-th black ball, and the $R$ red balls. The probability that the two black balls are the first two is*

$$P(I_k = 1, I_l = 1) = 2 \cdot \frac{1}{R+1} \cdot \frac{1}{R+2} \,.$$

*It follows,*

$$\mathrm{cov}(I_k, I_l) = \frac{2}{(R+1)(R+2)} - \frac{1}{(R+1)^2} \,,$$

*which simplifies to*

$$\mathrm{cov}(I_k, I_l) = \frac{R}{(R+1)^2(R+2)} \,.$$

*By the formula for the variance of sums we have*

$$\mathrm{var}(X) = B \cdot \frac{R}{(R+1)^2} + B(B-1) \cdot \frac{R}{(R+1)^2(R+2)} \,,$$

*which simplifies to*

$$\mathrm{var}(X) = \frac{BR(B+R+1)}{(R+1)^2(R+2)} \,.$$

**2.** (20) Let $X_1, X_2, \ldots$ be independent geometric random variables with the same parameter $p$ and let $q = 1 - p$, that is,

$$P(X_1 = r) = pq^{r-1}\,; \quad r = 1, 2, \ldots\,.$$

Denote $S_k = X_1 + X_2 + \cdots + X_k$.

a. (10) Find $P(S_k = n)$ for $1 \le k \le n$.

*Solution: since sums of independent geometric random variables are negative binomial, we have* $S_k \sim \mathrm{NegBin}(k, p)$. *We have*

$$P(S_k = n) = \binom{n-1}{k-1} p^k q^{n-k}\,.$$

b. (10) For each $n \ge 1$, compute

$$f_n = P\left(S_k = n \text{ for some } k = 1, 2, \ldots, n\right)\,.$$

*Solution: we have*

$$f_n = P\big(\cup_{k=1}^{n}\{S_k = n\}\big)\,.$$

*Observe that the events in the union are disjoint, so we get*

$$f_n = \sum_{k=1}^{n} P(S_k = n)\,.$$

*Using the result form the first part and the binomial formula, we get*

$$f_n = \sum_{k=1}^{n} \binom{n-1}{k-1} p^k q^{n-k} = p\,.$$

**3.** (20) Let the vector $(U, X, Y)$ have the density

$$f(u,x,y) = \frac{x|y|}{2\pi\sqrt{u^3(1-u)^3}}\, e^{-\frac{x^2}{2u}}\, e^{-\frac{y^2}{2(1-u)}}$$

for $u \in (0,1)$, $x > 0$ and $y \in \mathbb{R}$, and zero elsewhere. Define

$$W = \frac{X}{\sqrt{U}} \qquad \text{and} \qquad Z = \frac{Y}{\sqrt{1-U}}\,.$$

a. (10) Find the density of the vector $(U, W, Z)$. Are the random variables $U$, $W$ and $Z$ independent?

*Solution: define*

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

*and observe that the map* $\Phi$ *takes* $(0,1) \times (0,\infty) \times \mathbb{R}$ *bijectively onto itself. We have*

$$\Phi^{-1}(u,w,z) = \big(u,\ w\sqrt{u},\ z\sqrt{1-u}\,\big)\,,$$

*which implies* $J_{\Phi^{-1}}(u,w,z) = \sqrt{u(1-u)}$. *The transformation formula gives*

$$f_{U,W,Z}(u,w,z) = \frac{\sqrt{u(1-u)}\,w|z|}{2\pi\sqrt{u^3(1-u)^3}}\, e^{-\frac{w^2}{2}}\, e^{-\frac{z^2}{2}} \cdot \sqrt{u(1-u)}\,,$$

*which simplifies to*

$$f_{U,W,Z}(u,w,z) = \frac{w|z|}{2\pi\sqrt{u(1-u)}}\, e^{-\frac{w^2}{2}} e^{-\frac{z^2}{2}}\,.$$

*We infer that* $U$, $W$ *and* $Z$ *are independent.*

b. (10) Find the density of $(U, Y) = \big(U, Z\sqrt{1-U}\,\big)$, and compute the density of $Y$.

*Hint: when computing the marginal density, use the new variable*

$$\frac{\sqrt{u}}{\sqrt{1-u}} = v\,.$$

*Solution: noting that* $\int_0^\infty w\, e^{-w^2/2} dw = 1$, *we find that the vector* $(U, Z)$ *has density*

$$f_{U,Z}(u,z) = \frac{|z|}{2\pi\sqrt{u(1-u)}}\, e^{-\frac{z^2}{2}}\,.$$

*Taking* $\Phi(u,z) = \big(u, \sqrt{1-u} \cdot z\big)$, *the transformation formula gives*

$$f_{U,Y}(u,y) = f_{U,Z}\big(u, y/\sqrt{1-u}\,\big) \cdot \frac{1}{\sqrt{1-u}}\,.$$

*Combining both equalities we get*

$$f_{U,Y}(u,y) = \frac{|y|}{2\pi\sqrt{u(1-u)^3}}\, e^{-\frac{y^2}{2(1-u)}} \,.$$

*The density of $Y$ is the marginal density. Integrate the expression with respect to $u$. Introducing the new variable*

$$\frac{\sqrt{u}}{\sqrt{1-u}} = v$$

*we get*

$$\frac{du}{2\sqrt{u(1-u)^3}} = dv\,, \qquad \textit{and} \qquad \frac{1}{1-u} = 1+v^2\,.$$

*Finally,*

$$\begin{aligned} f_Y(y) &= \frac{|y|}{\pi}\int_0^\infty e^{-\frac{y^2(1+v^2)}{2}}dv \\ &= \frac{\sqrt{2}\,|y|}{\sqrt{\pi}}\, e^{-\frac{y^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{y^2v^2}{2}}dv \\ &= \frac{\sqrt{2}\,|y|}{\sqrt{\pi}}\, e^{-\frac{y^2}{2}} \cdot \frac{1}{2|y|} \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} \,. \end{aligned}$$

**4.** (20) In a sequence of independent tosses of a fair coin let $X$ be the number of tosses until the first appearance of the pattern HH, and $Y$ the number of tosses until the second appearance of the pattern HH. Examples:

| | |
|---|---|
| HTTHHTTTHTHH | $X = 5, \quad Y = 12$ |
| HTTHTTHTTHHH | $X = 11, \quad Y = 12$ |

a. (10) Find $E(X)$.

*Solution: define* $B_1 = \{\textit{first toss is a T}\}$, $B_2 = \{\textit{first two tosses are HT}\}$, *and* $B_3 = \{\textit{first two tosses are HH}\}$. *We have*

$$E(X|B_1) = 1 + E(X)\,, \quad E(X|B_2) = 2 + E(X) \quad \textit{and} \quad E(X|B_3) = 2\,.$$

*The formula for total expectation gives*

$$E(X) = \tfrac{1}{2}\big(1 + E(X)\big) + \tfrac{1}{4}\big(2 + E(X)\big) + \tfrac{1}{4} \cdot 2\,.$$

*Solving the linear equation gives* $E(X) = 6$.

b. (10) Find $E(Y - X)$.

*Solution: for* $k = 2, 3, \ldots$ *define*

$$B_k = \{X = k, \textit{the } (k+1)\textit{-th toss is H}\}$$

*and*

$$C_k = \{X = k, \textit{the } (k+1)\textit{-th toss is T}\}\,.$$

*We have*

$$E(Y - X|B_k) = 1 \quad \textit{and} \quad E(Y - X|C_k) = 1 + E(X)\,.$$

*Noting that the events* $B_2, B_3, \ldots, C_2, C_3, \ldots$ *form a partition, the formula for total expectation gives*

$$E(Y - X) = \sum_{k=2}^{\infty} P(B_k) + \sum_{k=2}^{\infty} \big(1 + E(X)\big) P(C_k)\,.$$

*Since* $P(B_k) = P(C_k)$ *and since these events form a partition, we have* $\sum_{k=2}^{\infty} P(B_k) = \sum_{k=2}^{\infty} P(C_k) = \frac{1}{2}$. *As a result, we conclude that*

$$E(Y - X) = 1 + \tfrac{1}{2} E(X) = 4\,.$$

**5.** (20) Let $X$ and $Y$ be independent, non-negative, integer valued random variables with the same distribution. Assume that for $k \geq 1$ we have

$$P(X = k) = \frac{1}{4}\, P(X + Y = k - 1)\,.$$

Let $G(s)$ be the generating function of $X$ and $Y$.

a. (10) Find an equation that is satisfied by $G(s)$.

*Solution: multiply both sides of the above relation by $s^k$ and sum over $k \geq 1$. Denoting $P(X = 0) = p$, we get*

$$\sum_{k=1}^{\infty} P(X = k)s^k = G_X(s) - p$$

*and*

$$\sum_{k=1}^{\infty} \frac{1}{4}\, P(X + Y = k - 1)s^k = \frac{s}{4}\, G_{X+Y}(s)\,.$$

*Since $X$ and $Y$ have the same distribution, we have $G_{X+Y}(s) = G(s)^2$. The desired equation is*

$$G(s) - p = \frac{s}{4}\, G(s)^2\,.$$

b. (10) Find the distribution of $X$.

*Hint: first $G(1) = 1$, and by Newton's expansion we have that for $|x| < 1$*

$$\sqrt{1 - x} = \sum_{k=0}^{\infty} (-1)^k \binom{1/2}{k} x^k\,.$$

*Solution: since $G(1) = 1$, the equation from the first part implies*

$$1 - p = \frac{1}{4}\,.$$

*Solving for $G(s)$ we get*

$$G(s) = \frac{2\left(1 \pm \sqrt{1 - \frac{3s}{4}}\right)}{s}\,.$$

*The coefficients of a generating function must be non-negative. Since $(-1)^k \binom{1/2}{k} < 0$ for all $k = 1, 2, 3, \ldots$, we have to choose the negative sign for the root. Expanding into a power series we get*

$$G(s) = \sum_{k=1}^{\infty} 2 \binom{1/2}{k} (-1)^{k-1} \frac{3^k s^{k-1}}{4^k}\,.$$

*Finally,*

$$P(X = k) = 2 \binom{1/2}{k+1} (-1)^k \left(\frac{3}{4}\right)^{k+1}\,.$$

**6.** (20) In 1999 the patrons of HIT Casinos played the game *Colore* 400,000 times. The probability of winning in the game is $p = 0.00198079$.

a. (10) The number of winning games in the 440,000 games is like a sum of independent equally distributed indicators $X_i$ with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. What, approximately, is the probability that there are 920 or more winning games.

*Solution: using the central limit theorem with the continuity correction and* $n = 440{,}000$ *we have*

$$P(S_n \geq 920) = P(S_n > 919.5) = P\left(\frac{S_n - E(S_n)}{\sqrt{\mathrm{var}(S_n)}} \geq \frac{919.5 - E(S_n)}{\sqrt{\mathrm{var}(S_n)}}\right)$$
$$\approx P\left(Z \geq \frac{919.5 - E(S_n)}{\sqrt{\mathrm{var}(S_n)}}\right),$$

*where* $Z \sim N(0,1)$. *Denoting* $q = 1 - p$, *we have* $E(S_n) = np = 871.55$ *and* $\sqrt{\mathrm{var}(S_n)} = \sqrt{npq} = 29.49$ *so*

$$\frac{919.5 - E(S_n)}{\sqrt{\mathrm{var}(S_n)}} \doteq \frac{919.5 - 871.55}{29.49} \doteq 1.63\,.$$

*The normal table gives*

$$P(S_n \geq 920) \approx 0.052\,.$$

*Remark. Using exact binomial probabilities, we can compute a more accurate value. Within displayed accuracy, we have* $P(S_n \geq 920) \doteq 0.5295$.

b. (10) Suppose the payout for a winning game is $x > 0$. If a patron stakes 1 unit and wins, she gets the stake back along with additional $x$ units. If she loses the game, she loses the stake. Find the largest $x$ such that after 440,000 games the Casino has a loss with probability at most 0.01.

*Solution: in every game the Casino either wins one unit or loses* $x$ *units. The gain is like the sum of* 440,000 *independent equally distributed random variables* $Y_i$ *with* $P(Y_i = 1) = q$ *and* $P(Y_i = -x) = p$. *Letting* $W_n = Y_1 + Y_2 + \cdots + Y_n$, *the largest* $x$ *has to approximately satisfy the equation*

$$P(W_n < 0) = 0.01\,.$$

*We compute* $E(Y_i) = q - px$, $\mathrm{var}(Y_i) = pq(x+1)^2$ *and* $\mathrm{var}(W_n) = npq(x+1)^2$. *By the central limit theorem we have*

$$P(W_n < 0) = P\left(\frac{W_n - E(W_n)}{\sqrt{\mathrm{var}(W_n)}} < -\frac{E(W_n)}{\sqrt{\mathrm{var}(W_n)}}\right)$$
$$= P\left(Z < -\frac{(q - px)\sqrt{n}}{(x+1)\sqrt{pq}}\right)$$
$$\approx 0.01\,.$$

*From the normal table we infer that*

$$-\frac{(q-px)\sqrt{n}}{(x+1)\sqrt{pq}} \approx -2.33\,,$$

*Solving for $x$ we get*

$$x \approx \frac{q\sqrt{n}-2.33\sqrt{pq}}{p\sqrt{n}+2.33\sqrt{pq}} \doteq 466.95\,.$$

*Remark. Exact binomial probabilities again yield more accurate value. Keeping the notation from the preceding part, the gain of the Casino equals $-S_n x + (n - S_n)$. Thus, the Casino has a loss if and only if $S_n > \frac{n}{x+1}$. From exact binomial probabilities, it follows that $P(S_n > 940) > 0.01$ and $P(S_n > 941) < 0.01$. Therefore, $P(S_n > y) \leq 0.01$ if and only if $y \geq 941$. Hence the maximum value of $x$ equals*

$$x = \frac{n}{941} - 1 \doteq 466.5877\,.$$