Максимумът на точките, които можете да получите общо от трите домашни през семестъра е 50, като това кореспондира с бонус от 0.5 към оценката за упражнения.

Ще считаме, че навсякъде работим върху вероятностно пространство $(\Omega, \mathcal{A}, \mathbb{P})$. *За удобство, дефинираме* $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Задача 1. (5 т.)

1. От 10 стандартни тестета от 52 карти се тегли по една карта. Намерете вероятността в получената ръка от 10 карти
   (а) да няма повтарящи се;
   (б) има поне три аса;
   (в) да има четири спатии, три кари, две купи и една пика;
   (г) броят на черните карти да с точно 4 повече от броя на червените, ако е известно, че черните карти са повече от червените.
2. Всички изделия в дадена партида са изправни, а в друга, 1/4 от изделията са за брак. Изделие, взето от случайно избрана партида, се оказва изправно. Да се пресметне вероятността второ случайно избрано изделие от същата партида да се окаже за брак, ако след проверката на първото изделие, то е било върнато обратно в своята партида.
3. Каква е вероятността корените на квадратното уравнение $x^2 + ax + b = 0, a, b \in [0, 1]$ да бъдат реални числа?

Задача 2. (5 т.)

Дете има в левия си джоб четири монети от 1 лв. и три монети от 2 лв., а в десния си джоб две монети от 1 лв. и една монета от 2 лв. Детето прехвърля две монети от левия си в десния си джоб, след това връща обратно две монети от десния в левия. Накрая, детето вади монета от десния си джоб. Каква е вероятността тя да е от 1 лв.?

Задача 3. (5 т.) По две от страните на правилен зар са оцветени в съответно бяло, зелено и червено. Хвърляме този зар два пъти. Нека X е броят на падналите се бели, а Y - на падналите се червени страни. Да се намерят съвместното разпределение на X и Y, независими ли са, ковариацията им, $\mathbb{P}(X = 1|Y = 1)$ и $\mathbb{P}(X > Y)$.

Задача 4. (5 т.) Нека сл. вел. X приема стойности в $\mathbb{N}_0$ и

$$g_X(s) := \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$$

е пораждащата ѝ функция.

1. Нека $Y := 3X$ и $Z = X_1 + X_2$, където $X_1, X_2 \sim X$ са независими. Изразете чрез g_X, пораждащите функции g_Y и g_Z.
2. Нека $X \sim Ge(p)$, т.е. $P(X = k) = p(1 - p)^k$ за $k \in \mathbb{N}_0$. Пресметнете g_X и чрез нейна помощ намерете $\mathbb{E}(X)$ и $D(X)$.

Задача 5. Нека X и Y са независими сл. вел и

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = 2\mathbb{P}(Y = 3) = 2\mathbb{P}(Y = 5) = \frac{1}{2}.$$

Нека $Z_1 := 2X + Y + 1$, $Z_2 := XY$ и $Z_3 = X^Y$. Намерете очакванията и дисперсиите им.

Задача 6 (20 т., залагане на ези/тура)**.** Нека $n \geq 3$ и p са естествени числа. Разглеждаме следната игра с участници $A_1, \ldots, A_n$:

- Всеки играч написва предвижданията си за изходите от $(2p + 1)$ хвърляния на стандартна монета (1/2 вероятност за ези и тура), като пише E за ези и T за тура;
- Монетата се хвърля $(2p + 1)$ пъти и играчите с най-много познати си разделят награда от $n!$ лева.

Пример: $n = 4, p = 1$, предвиждания: ETE, EEE, TTT, EET. При изход TTE, играчи 1 и 3 имат по 2 познати и си разделят 4!, т.е. всеки взима по 12 лева.

$n = 4, p = 1$, предвиждания: ETE, EEE, TTT, EET. При изход TTT, играч 3 има 3 познати и печели $4! = 24$ лева.

За $i \in [n]$ да дефинираме

- $X_i =$ "броят познати от A_i изходи";
- $G_i =$ "печалбата на играч A_i";
- $q_k = \mathbb{P}(X_i = k), r_k = \mathbb{P}(X_i \leq k)$.

Целта ни ще е да пресметнем очакваната печалба на A_1, т.е. $\mathbb{E}(G_1)$ при две различни стратегии.

Част 1. Начални съображения.

1. * Както винаги работим върху вероятностно пространство $(\Omega, \mathcal{F}, \mathbb{P})$. Кое множество бихте предложили за Ω?
2. Какво е разпределението на сл. вел. X_i?
3. Пресметнете r_p.
4. Напомняме, че с $X(\Omega)$ означаваме всички образи на елементи от Ω под функцията X. Например, $X_1(\Omega) = \{0, 1, \ldots, 2p+1\}$, т.к. играчът A_1 може да уцели от 0 до $2p+1$ изхода.

   Ако $\mathbb{P}(A) > 0$, дефинираме условно очакване на сл. вел. X при условие A като

   $$\mathbb{E}(X|A) = \sum_{k \in X(\Omega)} k\mathbb{P}(X = k|A)\,.$$

   Докажете, че ако $B_1, \ldots, B_s$ е пълна система от събития с ненулева вероятност, то

   $$\mathbb{E}X = \sum_{k=1}^{s} \mathbb{E}(X|B_k)\mathbb{P}(B_k)\,.$$

   (Може да разпишете $\mathbb{E}X$ по дефиниция, да приложите формулата за пълната вероятност и да размените реда на сумиране.)

Част 2. Всички играчи играят случайно. В тази част приемаме, че всички играчи избират ези/тура случайно с вероятност 1/2 и независимо един от друг.

1. На какво е равно е $G_1(\Omega)$, т.е. какви са възможните печалби на играча A_1?
2. За всяко $k \in X_1(\Omega), j \in [n]$, докажете, че

   $$\mathbb{P}\left(G_1 = \frac{n!}{j}\,\middle|\, X_1 = k\right) = \binom{n-1}{j-1}(r_k - r_{k-1})^{j-1} r_{k-1}^{n-j}\,.$$

3. Докажете, че за $k > 0$

   $$\mathbb{E}(G_1|X_1 = k) = \frac{(n-1)!(r_k^n - r_{k-1}^n)}{q_k}\,.$$

   Забележете, че горната формула е вярна и за $k = 0$, ако дефинираме $r_{-1} = 0$.

   (Разпишете с помощта на предишната подточка и използвайте бинома на Нютон)

4. Заключете, че $\mathbb{E}(G_1) = (n-1)!$.

   (Разгледайте пълната система от събития $\{X_1 = k\}, k \in \{0, 1, \ldots, 2p+1\}$)

5. Очаквано ли е заключението на горната подточка? Можехме да достигнем по него и по друг начин: можем ли да пресметнем $G_1 + \cdots + G_n$? Използвайте линейността на очакването и симетричността, за да получите (отново) $\mathbb{E}(G_1)$.

Част 3. A_1 и A_2 играят заедно, а всички други играят случайно. В тази част, да приемем, че всички освен A_2 играят случайно. A_2 от своя страна се познава с A_1 и залага на обратното на A_1. Например, ако A_1 залага на EET, то A_2 ще заложи на TTE. Ако някой от двамата спечели, си делят печалбата поравно.

Да означим

- с G', G'_1, G'_2 респективно общата печалба на A_1 и A_2, и печалбите им поотделено; (В такъв случай $G' = G'_1 + G'_2$.)
- Y = "по-големият брой познати изходи от A_1 и A_2"(при ЕЕТ, ТТЕ и изход от хвърлянията EEE, $Y = 2$);
- както преди, за $k \in \{1, 3, 4, \ldots, n\}$, $q_k = \mathbb{P}(X_i = k), r_k = \mathbb{P}(X_i \leq k)$.

1. Докажете, че $Y(\Omega) = \{p+1, \ldots, 2p+1\}$.
2. Докажете, че за $k \in \{p+1, \ldots, 2p+1\}, \mathbb{P}(Y = k) = 2q_k$.
3. Докажете, че за $k \in \{p+1, \ldots, 2p+1\}, j \in \{0, \ldots, n-2\}$

$$\mathbb{P}\left(G' = \frac{n!}{j+1} \,\middle|\, Y = k\right) = \binom{n-2}{j}(r_k - r_{k-1})^j r_{k-1}^{n-2-j}$$

   и като следствие (използвайки бинома на Нютон), че

$$\mathbb{E}(G'|Y = k) = n(n-2)!\frac{r_k^{n-1} - r_{k-1}^{n-1}}{q_k}.$$

4. Заключете, че

$$\mathbb{E}(G') = 2n(n-2)!\left(1 - \frac{1}{2^{n-1}}\right)$$

   и проверете, че тази стратегия е по-добра за A_1 и A_2 от тази в част 2.
5. Пресметнете $\mathbb{E}(G_i)$ за $i \in \{3, \ldots, n\}$.

Задача 7. (15 т., Симулации на случаен граф) Случайните графи са област на вероятностите, която привлича особен интерес през последните години, намирайки връзки със статистическата механика, моделирането на големи мрежи (Facebook, Google Searches, DNA). Проблемът в директното разглеждане на графа на познанствата във Фейсбук например е, че дори да имаме достъп до него, матрицата му на инцидентност би прекалено голяма, за да можем да я манипулираме с днешния хардуер. Затова, ако имаме добър вероятностен модел, това може да ни позволи да правим наблюдения без проверки (например да докажем, че в при всеки граф, който симулираме ще можем да стигнем от всеки до всеки връх с най-много X стъпки - така наречения small world network)

Тук ще се запознаем емпирично с най-лесния модел на случаен граф - този на Ердьош и Рени (1960).

Нека n е броят на върховете на граф G и $p \in [0, 1]$ е параметър. За всяка двойка от върхове, хвърляме монета с вероятност за ези p. Ако се падне ези, свързваме тези върхове, а в противен случай - не.

Целта ни ще е да анализираме вероятността графът да е свързан.

1. Параметрите, които ще меним са n, p и N - броят симулации. При всеки избор на n и p ще трябва да симулирате голям брой графи ($N = 10000, 1000, 100$? изборът на N и n ще зависи от това колко симулации са постижими в обозримо време). При фиксирани n и p, това може да се направи като итерирате по всички двойки върхове и избирате всеки път $Ber(p)$, например чрез *scipy.stats.bernoulli*.

   След като сте избрали кои от $\binom{n}{2}$-те двойки върхове са свързани с ребро, то проверете дали графа е свързан. Ако в M от N-те симулации графът е свързан, оценката ни за вероятността това да е така в общия случай ще е M/N.

   Направете графика на частта свързани графи (M/N) към p - на оста x ще бъде параметъра p, а по y - получената пропорция. Обяснете защо функцията трябва да е растяща.

   Пробвайте с $n = 100, 1000, 5000, 10000, 100000$? $N = 10000$? и разгледайте за p стойностите от 0 до 1 през някаква стъпка (например 0.01).

   Можете да пресметнете броя операции при фиксирани n, p и N, за да знаете какви параметри можете да избирате.

   Забелязвате ли нещо? Разгледайте стойности за p, които са $o(\ln n/n)$, например $\ln(\ln(n))/n$ и такива, които са $\omega(\ln n/n)$, например $(n/2)/n = 1/2$.
2. * Направете същото, но не оценявайте пропорцията на свързаните графи, а големината на най-голямата свързана компонента. В първата част, разглеждахме само дали големината на най-голямата свързана компонента е n или не. Резултат още от първите статии на Ердьош и Рени гласи, че, ако n е голямо, то при $p = o(1/n)$ нямаме свързана компонента с линейна (по n) големина, а ако $p = \omega(1/n)$, т.е. $1/n = o(p)$, то имаме такава компонента. Можете ли да забележите това емпирично?

Задача 8. (10 т., Неравенство на Марков, Граници на Чернов) Нека X е положителна случайна величина с крайно очакване и $a > 0$.

1. Като използвате, че $\mathbb{E}X = \mathbb{P}(X \geq a)\mathbb{E}(X|X \geq a) + \mathbb{P}(X < a)\mathbb{E}(X|X < A)$, докажете неравенството на Марков
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}\,.$$
Докажете чрез него неравенството на Чебишов
$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{DX}{a^2}\,.$$

2. Докажете неравенството на Чернов: за всяко $t > 0$
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}\left(e^{tX}\right)}{e^{ta}}\,.$$
(Каква е връзката между $\{X \geq a\}$ и $\{e^{tX} \geq e^{ta}\}$?)

3. Нека $X_1, \ldots, X_n \sim Ber(1/2)$ са независими и $X = X_1 + \cdots + X_n$. Какво е разпределението, очакването и дисперсията на X? Пресметнете $\mathbb{E}e^{tX}$.

4. Да приемем без доказателство следното следствие от неравенството на Чернов: за $\delta \in (0,1)$ и дефинираното по-горе X
$$\mathbb{P}\left(|X - \mathbb{E}X| \geq \delta n/2\right) \leq 2e^{-\frac{n\delta^2}{6}}\,.$$
*(*Може да направите най-добра оценка за $\mathbb{P}(X \geq a)$ в неравенството на Чернов, като минимизирате $\mathbb{E}(e^{tX})/e^{ta}$ по $t > 0$. Ако $= (1+\delta)\mathbb{E}X$, то може да се провери чрез диференциране, че този минимум се достига за $t = \ln(1+\delta)$ Горната оценка следва от това оптимално неравенство и симетрия.)*

   Колко трябва да изберем δ, така че да сме сигурни, че $X \notin (\mathbb{E}X - \delta n/2, \mathbb{E}X + \delta n/2)$ с вероятност най-много $4/n$ (или еквивалентно $X \in [\mathbb{E}X - \delta n/2, \mathbb{E}X + \delta n/2]$ с вероятност поне $1 - 4/n$).

   А колко трябва да изберем a в неравенството на Чебишов за да сме сигурни, че $X \notin [\mathbb{E}X - a; \mathbb{E}X + a]$ с вероятност най-много $4/n$. Забележете, че неравенството на Чернов дава интервал, който е по порядък по-малък от получения чрез неравенството Чебишов. Сравнете двата интервала за $n = 10000$.