Максимумът на точките, които можете да получите 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* $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Част 1

Задача 1. (10 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. 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.
3. 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. Дефинираме $Y := 3X$ и $Z = X_1 + X_2$, където $X_1, X_2 \sim X$ са независими. Изразете пораждащите функции g_Y и g_Z чрез g_X.
4. Студенти влизат последователно на изпит, показвайки личната си карта. Преди изпита е обявено, че първият студент, чийто рожден ден съвпада с рождения ден на вече влязъл студент, ще получи единица бонус към оценката си. На кое място трябва да застанете в редицата от студенти, за да имате най-голям шанс да сте печелившия студент?
5. Човек се намира на числовата ос в точката $n \in \mathbb{N}$ и последователно прави стъпка към $(n+1)$ с вероятност $p > 1/2$ и към $(n-1)$ с вер. $(1-p)$. Нека $p_n = \mathbb{P}(\text{човекът достига } 0, \text{ тръгвайки от } n)$. Изразете p_1, p_2 и p_3 чрез p.

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 (10 т.)**.** Точка (x, y) се избира случайно в квадрата $[-1,1] \times [-1,1]$.

1. Дефинираме двата региона:
$$A := \{x^2 + y^2 \le 1\} \qquad \text{и} \qquad B := \{|x| + |y| \le 1.2\}.$$
   Пресметнете $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$, $\mathbb{P}(A|B)$ и $\mathbb{P}(B|A)$.
2. Напишете програма, която генерира N случайни точки (x, y) в квадрата, като ще оценим $p = \mathbb{P}(x^2 + y^2 \le 1)$ чрез
$$\hat{p} = \frac{\#\{(x_i, y_i) : x_i^2 + y_i^2 \le 1\}}{N}.$$

(а) За всяко $N \in \{10^3, 10^4, 10^5, 10^6\}$, отчетете $\hat{p}$, абсолютната грешка $|\hat{p} - p|$ и приблизителен 95% доверителен интервал, използвайки $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$. Проверете дали аналитичната стойност на p попада в интервала. Как се променя ширината на интервала с N?

(б) На логаритмична скала начертайте графика на $\hat{p}$ спрямо N. Опишете какво наблюдавате.

(в) Измерете времето за изпълнение за всяко N и обсъдете компромиса между точност и време за изчисление.

Задача 5. (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?

Част 2

Задача 6. (15 т.)

1. Заек тръгва от точката 0 на числовата ос и прави независими равномерно разпределени в интервала $[0, 1]$ скокове в положителна посока. Участъкът $[1 - x, 1]$ е капан с дължина $x \in [0, 1]$. Каква е вероятността заекът да прескочи капана?

2. Нека $X \sim \text{Unif}(0, 1)$. Дефинираме $Y := X(1 - X)$. Намерете D_Y, F_Y, f_Y, и $\mathbb{E}[Y]$.

3. Нека сл.в. X и Y имат съвместна плътност

$$f_{XY}(x, y) := \begin{cases} kxe^{-y}, & 0 \le x \le 1, 0 \le y \le \frac{1}{1-x}, \\ 0, & \text{иначе.} \end{cases}$$

   Намерете константата k, $\mathbb{P}(Y - X \le 1.5)$ и $\mathbb{E}[Y|X = 0.5]$. Независими ли са X и Y?

4. Нека сл.в. X и Y имат съвместна плътност

$$f_{XY}(x, y) := \begin{cases} 3(x + y), & 0 \le x \le y \le 1, \\ 0, & \text{иначе.} \end{cases}$$

Дефинираме сл.в.

$$U := X + Y, \qquad V := \frac{X}{X+Y}.$$

Намерете $f_{U,V}$, f_U и f_V. Независими ли са U и V? Пресметнете $\mathbb{E}[X|U]$, $\mathbb{E}[Y|U]$ и $\mathbb{E}[V]$.

5. Софтуерна задача се изпълнява на сървър. Времето T (в секунди), необходимо за изпълнение на задачата, ако тя се изпълнява без прекъсване, е експоненциална случайна величина с честота $\lambda > 0$. Сървърни сривове се случват според независим експоненциален часовник F с честота $\mu > 0$. Ако сривът се случи преди задачата да приключи, тя се рестартира отначало, след като сървърът бъде незабавно ресетнат. Този процес се повтаря, докато задачата най-накрая приключи. Нека S бъде общото време до окончателното приключване на задачата. Намерете вероятността работата да приключи преди да възникне срив в сървъра и очакваното време за завършване на задачата. Обяснете защо $\mathbb{E}[S]$ е крайно, въпреки че работата може да бъде рестартирана безброй много пъти. Какво става с $\mathbb{E}[S]$ при $\mu \to 0$ и $\mu \to \infty$?

Задача 7. (10 т., Неравенство на Марков, Граници на Чернов) Нека X е положителна случайна величина с крайно очакване и $a > 0$.

1. 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 $\{e^{tX} \geq e^{ta}\}$?)

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

3. 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 $n = 10000$.

Задача 8. (30 т., Симулации на случайни величини) Предполагаме, че можем да симулираме числа $\text{Unif}(0,1)$ безпроблемно — тази част е например математическа, ако искаме да конструираме детерминистични редици, които да „изглеждат като” извадка от независими $\text{Unif}(0,1)$ числа (така наречените псевдо-случайно числа) или чисто технологична - чрез използването на физически феномени.

В тази задача ще разгледаме 3 класически алгоритъма за симулиране на реализация X от конкретно разпределение - чрез използване на обратната функция на F_X, чрез метод на отхърлянето и чрез метода на Бокс и Мюлер.

Метод 1:

1. Нека $U \sim \text{Unif}(0,1)$. Какво е разпределението на $F_X^{-1}(U)$?, където F_X^{-1} е обратната на F_X функция? [1] Използвайте получения резултат, за да предложите процедура за симулиране на:

   (а) $X \sim Exp(\lambda)$ за $\lambda > 0$;

   (б) $X \sim Cauchy(c)$ за $c > 0$, т.е.

$$f_X(x) = \frac{c}{\pi(c^2 + x^2)};$$

[1] Формално, тъй като F_X може да не е строго растяща, F_X^{-1} може да не е добре дефинирана навсякъде. Затова обикновено се работи с обобщената обратна: $F_X^{-1}(t) := \inf\{x \in \mathbb{R} : F_X(x) \geq t\}$. За нашите цели може да считате, че работим с нормална обратна.

(в) $X \sim Pareto(\theta)$ за $\theta > 0$, т.е.

$$f_X(x) = \frac{\theta}{x^{1+\theta}} \mathbb{1}_{\{x \geq 1\}}.$$

2. За моделиране на какви феномени се използват горните разпределения и защо? Генерирайте извадки от тях и илюстрирайте закона за големите числа и централната гранична теорема в тези 3 случая.

3. Какви са проблемите на този метод, ако искаме да го използваме за нормално или гама разпределение?

Метод 2: Следващият алгоритъм, който ще разгледаме[2] се използва при симулирането на X, която има плътност f_X:

- Намираме подходяща сл. вел. Y, от можем да симулираме лесно, например чрез метод 1, която има има плътност f_Y и $f_X(x)/f_Y(x) \leq c$ за всички x и някаква константа $c > 0$;

- Генерираме представител от Y;

- Генерираме (независимо от Y) $U \sim \text{Unif}(0, 1)$;

- Ако
$$U \leq \frac{f_X(Y)}{cf_Y(Y)},$$
приемаме Y като извадка за X, в противен случай го отхвърляме и повтаряме процедурата отначало.

1. Докажете коректността на алгоритъма.

2. Генерирайте $X \sim N(\mu, \sigma^2)$. *(една възможност е след подходяща трансформация да използвате за Y експоненциалното разпределение)*

3. Нека $X \sim \Gamma(\alpha, 1)$, т.е.
$$f_X(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \mathbb{1}_{\{x>0\}}.$$

   (а) Нека $\alpha < 1$ и Y е сл.вел. с плътност
   $$f_Y(x) = \frac{\alpha e}{\alpha + e} \left(x^{\alpha-1} \mathbb{1}_{\{0<x<1\}} + e^{-x} \mathbb{1}_{\{x \geq 1\}} \right).$$

   (б) Докажете, че горното наистина е плътност. Как бихте симулирали Y чрез метод 1?

   (в) Можем ли да използваме метод 2 за генериране на $X \sim \Gamma(\alpha, 1)$ за $\alpha < 1$?

   (г) Как бихте симулирали $\Gamma(\alpha, \beta)$, използвайки горния метод и свойствата на гама разпределнието?

Метод 3:

1. Нека $X^2 \sim Exp(1/2)$ и $\Theta \sim \text{Unif}(0, 2\pi)$ са независими. Намерете рапределенията на $Z_1 := R \cos \Theta$ и $Z_2 := R \sin \Theta$. Независими ли са?

2. Как бихте използвали горното, за да симулирате n независими $N(\mu, \sigma)$ сл. величини?

Често използвано разпределение (защо?) е лог-нормалното: $X \sim Lognormal(\mu, \sigma^2)$, ако $\ln(X) \sim N(\mu, \sigma)$. Предложете 3 начина да симулирате от него.

[2] чийто идеи се свързват с идеи на фон Нойман от 1949-1950