Seminarul 4

1. Fie S mulțimea tuturor numerelor naturale cel mult egale cu 50, cu exact două cifre de parități diferite. Un număr este ales aleator din S. Fie X suma cifrelor num?rului ales. Scrieți distribuția lui X, apoi calculați valoarea sa medie $E(X)$.

R: Fie Y numărul ales aleator. Avem

- $X = 1$, dacă $Y \in \{10\}$.
- $X = 3$, dacă $Y \in \{12, 21, 30\}$.
- $X = 5$, dacă $Y \in \{14, 41, 23, 32, 50\}$.
- $X = 7$, dacă $Y \in \{16, 25, 34, 43\}$.
- $X = 9$, dacă $Y \in \{18, 27, 36, 45\}$.
- $X = 11$, dacă $Y \in \{29, 38, 47\}$.
- $X = 13$, dacă $Y \in \{49\}$.

Deci, $X \sim \begin{pmatrix} 1 & 3 & 5 & 7 & 9 & 11 & 13 \\ \frac{1}{21} & \frac{3}{21} & \frac{5}{21} & \frac{4}{21} & \frac{4}{21} & \frac{3}{21} & \frac{1}{21} \end{pmatrix}$, $E(X) = 1 \cdot \frac{1}{21} + 3 \cdot \frac{3}{21} + 5 \cdot \frac{5}{21} + 7 \cdot \frac{4}{21} + 9 \cdot \frac{4}{21} + 11 \cdot \frac{3}{21} + 13 \cdot \frac{1}{21} = \frac{1+9+25+28+36+33+13}{21} = \frac{145}{21} \approx 6{,}9$.

2. Considerăm următoarea problemă de *clasificare naivă Bayes* a unor restaurante (**R**), în
• *clasele*: **r**ecomandat sau **n**erecomandat,
în funcţie de următoarele *atribute* cu valorile lor posibile:
• cost (C): ieftin, mediu, scump;
• timp de aşteptare (T): puţin, mediu, îndelungat;
• mâncare (M): fadă, acceptabilă, bună, delicioasă.

R, C, T, M sunt variabelele aleatoare (categoriale) şi **r**, **n**, i, m, s, p, m, $\hat{\imath}$, f, a, b, d valorile de mai sus, în ordinarea în care sunt menţionate.

Considerăm următorul *tabel de date* furnizat de clienţii unor restaurante:

	Cost	*Timp de aşteptare*	*Mâncare*	**Restaurant**
1	mediu	îndelungat	acceptabilă	**nerecomandat**
2	scump	puţin	bună	**recomandat**
3	ieftin	îndelungat	delicioasă	**recomandat**
4	mediu	puţin	bună	**recomandat**
5	ieftin	mediu	acceptabilă	**nerecomandat**
6	ieftin	puţin	fadă	**nerecomandat**
7	mediu	puţin	acceptabilă	**nerecomandat**
8	mediu	mediu	delicioasă	**recomandat**
9	scump	puţin	delicioasă	**recomandat**
10	ieftin	îndelungat	bună	**nerecomandat**
11	scump	puţin	acceptabilă	**nerecomandat**
12	mediu	mediu	bună	**recomandat**
13	mediu	îndelungat	fadă	**nerecomandat**
14	scump	mediu	delicioasă	**recomandat**
15	ieftin	mediu	fadă	**nerecomandat**
16	mediu	puţin	delicioasă	**recomandat**
17	ieftin	puţin	acceptabilă	**recomandat**
18	scump	îndelungat	bună	**nerecomandat**
19	ieftin	puţin	fadă	**recomandat**
20	scump	îndelungat	delicioasă	**nerecomandat**

i) Folosind datele din tabel, determinaţi probabilităţile claselor şi probabilităţile condiţionate ale atributelor, ştiind clasa.
ii) Considerăm evenimentul dat de *vectorul de atribute*: $E = (C = s) \cap (T = m) \cap (M = b)$. Alegeţi o clasă pentru E, stabilind care din următoarele probabilităţi este mai mare: $P(\mathbf{R} = \mathbf{r}|E)$ sau $P(\mathbf{R} = \mathbf{n}|E)$.
iii) Determinaţi $P(E)$.

R.:

i)

$\mathbf{R} = \mathbf{r}$	$\mathbf{R} = \mathbf{n}$	$P(\mathbf{R} = \mathbf{r})$	$P(\mathbf{R} = \mathbf{n})$
10	10	$\frac{1}{2}$	$\frac{1}{2}$

C	$\mathbf{R} = \mathbf{r}$	$\mathbf{R} = \mathbf{n}$	$P(C = ...\|\mathbf{R} = \mathbf{r})$	$P(C = ...\|\mathbf{R} = \mathbf{n})$
i	3	4	$\frac{3}{10}$	$\frac{4}{10}$
m	4	3	$\frac{4}{10}$	$\frac{3}{10}$
s	3	3	$\frac{3}{10}$	$\frac{3}{10}$

T	$\mathbf{R}=\mathbf{r}$	$\mathbf{R}=\mathbf{n}$	$P(T=...\mid\mathbf{R}=\mathbf{r})$	$P(T=...\mid\mathbf{R}=\mathbf{n})$
p	6	3	$\frac{6}{10}$	$\frac{3}{10}$
m	3	2	$\frac{3}{10}$	$\frac{2}{10}$
$\hat{\imath}$	1	5	$\frac{1}{10}$	$\frac{5}{10}$

M	$\mathbf{R}=\mathbf{r}$	$\mathbf{R}=\mathbf{n}$	$P(M=...\mid\mathbf{R}=\mathbf{r})$	$P(M=...\mid\mathbf{R}=\mathbf{n})$
f	1	3	$\frac{1}{10}$	$\frac{3}{10}$
a	1	4	$\frac{1}{10}$	$\frac{4}{10}$
b	3	2	$\frac{3}{10}$	$\frac{2}{10}$
d	5	1	$\frac{5}{10}$	$\frac{1}{10}$

ii) Pe baza formulei lui Bayes şi a ipotezei de independenţă condiţionată, deducem că:

$$P(\mathbf{R}=\mathbf{r}|E) = \frac{P(E|\mathbf{R}=\mathbf{r})P(\mathbf{R}=\mathbf{r})}{P(E)} = \frac{P(C=s,T=m,M=b|\mathbf{R}=\mathbf{r})P(\mathbf{R}=\mathbf{r})}{P(E)}$$

$$= \frac{P(C=s|\mathbf{R}=\mathbf{r})P(T=m|\mathbf{R}=\mathbf{r})P(M=b|\mathbf{R}=\mathbf{r})P(\mathbf{R}=\mathbf{r})}{P(E)} = \frac{\frac{3}{10}\cdot\frac{3}{10}\cdot\frac{3}{10}\cdot\frac{1}{2}}{P(E)} = \frac{1}{P(E)}\cdot\frac{27}{2000}$$

şi

$$P(\mathbf{R}=\mathbf{n}|E) = \frac{P(E|\mathbf{R}=\mathbf{n})P(\mathbf{R}=\mathbf{n})}{P(E)} = \frac{P(C=s,T=m,M=b|\mathbf{R}=\mathbf{n})P(\mathbf{R}=\mathbf{n})}{P(E)}$$

$$= \frac{P(C=s|\mathbf{R}=\mathbf{n})P(T=m|\mathbf{R}=\mathbf{n})P(M=b|\mathbf{R}=\mathbf{n})P(\mathbf{R}=\mathbf{n})}{P(E)} = \frac{\frac{3}{10}\cdot\frac{2}{10}\cdot\frac{2}{10}\cdot\frac{1}{2}}{P(E)} = \frac{1}{P(E)}\cdot\frac{12}{2000}.$$

Deoarece $P(\mathbf{R}=\mathbf{r}|E) > P(\mathbf{R}=\mathbf{n}|E)$, asociem vectorului de atribute E clasa $\mathbf{R}=\mathbf{r}$.

iii) Din ii) rezultă

$$1 = P(\mathbf{R}=\mathbf{r}|E) + P(\mathbf{R}=\mathbf{n}|E) = \frac{1}{P(E)}\cdot\frac{27+12}{2000},$$

deci

$$P(E) = \frac{19,5}{1000} = 0,0195.$$

3. Ce valoare teoretică estimează programul următor? Calculaţi valoarea teoretică corespunzătoare.

```
[ ]: import numpy as np

N=2000
```

```python
S = np.concatenate((np.zeros(50),np.ones(70),2*np.ones(80)))
X=[]
for _ in range(N):
    k=0
    i= np.random.randint(len(S))
    while S[i] != 0:
        i= np.random.randint(len(S))
        k=k+1
    X.append(k)

print(" . . . . . :",np.mean(X))
```

R: Programul generează numere aleatoare pentru o variabilă aleatoare $X \sim Geo(p)$, unde $p = \frac{50}{50+70+80} = \frac{1}{4}$ =probabilitatea de a alege un element nul din vectorul S:

$$P(X = k) = p(1-p)^k, k = 0, 1, 2, \ldots$$

Programul estimează valoarea medie $E(X)$. Valoarea sa teoretică este $E(X) = \sum_{k=0}^{\infty} kp(1-p)^k$. Pe baza criteriului raportului, seria cu termeni pozitivi $\displaystyle\sum_{k=0}^{\infty} kp(1-p)^k$ este convergentă.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kp(1-p)^k = (1-p)\sum_{k=1}^{\infty} kp(1-p)^{k-1} \\
&\overset{k=j+1}{=\!=} (1-p)\sum_{j=0}^{\infty}(j+1)p(1-p)^j = (1-p)\sum_{j=0}^{\infty} jp(1-p)^j + (1-p)\sum_{j=0}^{\infty} p(1-p)^j \\
&= (1-p)E(X) + (1-p) \implies E(X) = \frac{1-p}{p}.
\end{aligned}$$

$\implies E(X) = \frac{\frac{3}{4}}{\frac{1}{4}} = 3$, deci în medie sunt necesare 3 iteraţii până apare primul 0.

4. Fie $a, b \in \mathbb{Z}$, $a \leq b$, și $c \in (0, 1)$. Spunem că variabila aleatore X are o distribuție uniform discretă dacă

$$X \sim \begin{pmatrix} a & a+1 & \ldots & b \\ c & c & \ldots & c \end{pmatrix}.$$

a) Determinați valoarea lui c.
b) Pentru $a = 3$ și $b = 21$, calculați

$$P\Big(\Big\{X \leq \frac{a+b}{2}\Big\} \cup \Big\{\frac{a+b}{6} \leq X\Big\}\Big) \text{ și } P\Big(\Big\{X \leq \frac{a+b}{2}\Big\} \cap \Big\{\frac{a+b}{6} \leq X\Big\}\Big).$$

c) Determinați a și b, știind că $P(X = a) = \frac{1}{3}$ și $E(X) = 1$.

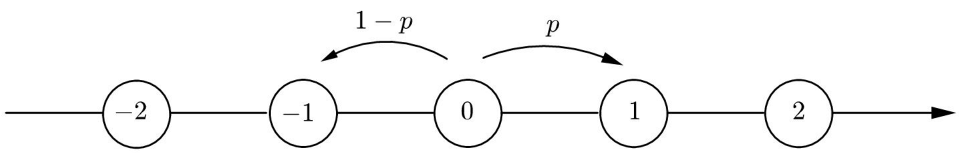
R: a) Deoarece $\sum_{k=a}^{b} P(X=k)=1$ are $b-a+1$ termeni (egali cu c), avem $c=\dfrac{1}{b-a+1}$.

b) $a=3$ and $b=21 \Longrightarrow \dfrac{a+b}{2}=12, \dfrac{a+b}{6}=4$. Calculăm

$$P\big(\{X\leq 12\}\cup\{4\leq X\}\big)=1-P\big(\{12<X\}\cap\{X<4\}\big)=1-0=1.$$
$$P\big(\{X\leq 12\}\cap\{4\leq X\}\big)=P\big(X\in\{4,5,...,12\}\big)=\frac{12-4+1}{19}=\frac{9}{19}.$$

c) $c=\frac{1}{b-a+1}=\frac{1}{3} \Longrightarrow b-a=2$. $E(X)=\frac{a+(a+1)+\dots+b}{b-a+1}=\frac{a(b-a+1)+1+2+\dots+(b-a)}{b-a+1}=a+\frac{(b-a)(b-a+1)}{2(b-a+1)}=a+\frac{b-a}{2}=\frac{b+a}{2}=1 \Longrightarrow b+a=2$. Deci, $a=0, b=2$.

5. Un punct material se deplasează pe axa reală dintr-un nod spre un nod vecin, la fiecare pas, cu probabilitatea $p\in(0,1)$ la dreapta şi cu probabilitea $1-p$ la stânga. Nodurile sunt centrate în numerele întregi:



Fie X variabila aleatoare care indică poziţia finală a punctului material după $n\in\mathbb{N}$ paşi ai unei deplasări ce porneşte din nodul 0. Determinaţi distribuţia şi valoarea medie lui X.

R: Dacă Y_i reprezintă pasul i, atunci $Y_i \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix} \Longrightarrow Y_i=2X_i-1$ cu $X_i \sim Bernoulli(p)$, $i\in\{1,\dots,n\}$. $X=Y_1+\ldots+Y_n=(2X_1-1)+\ldots+(2X_n-1)$, $X_1+\ldots+X_n \sim Bino(n,p) \Longrightarrow X\sim \begin{pmatrix} 2k-n \\ C_n^k p^k(1-p)^{n-k} \end{pmatrix}_{k=\overline{0,n}}$ şi $E(X)=2np-n$.

6. Considerăm vectorul aleatoar discret (X,Y) cu distribuţia dată sub formă tabelară:

$X \backslash Y$	-2	1	2
1	0,2	0,1	0,2
2	0,1	0,1	0,3

a) Să se determine distribuţiile de probabilitate ale variabilelor aleatoare X şi Y.
b) Calculaţi probabilitatea ca $|X-Y|=1$, ştiind că $Y>0$.
c) Sunt evenimentele $X=2$ şi $Y=1$ independente?
d) Sunt variabilele aleatoare X şi Y independente?
e) Sunt evenimentele $X=1$ şi $Y=1$ condiţional independente, cunoscând $X+Y=2$?

f) Este variabila aleatoare X condiţional independentă de Y, cunoscând $X+Y$?
g) Calculaţi valoarea medie a variabilei aleatoare $2X+Y^2$.

R: a) $X \sim \begin{pmatrix} 1 & 2 \\ 0{,}5 & 0{,}5 \end{pmatrix}, Y \sim \begin{pmatrix} -2 & 1 & 2 \\ 0{,}3 & 0{,}2 & 0{,}5 \end{pmatrix}$.
b) $P(|X-Y|=1 \big| Y>0) = \frac{P(|X-Y|=1, Y>0)}{P(Y>0)} = \frac{P(X=1,Y=2)+P(X=2,Y=1)}{P(Y>0)} = \frac{0{,}3}{0{,}7} = \frac{3}{7}$.
c) $P(X=2, Y=1) = 0{,}1 = 0{,}5 \cdot 0{,}2 = P(X=2) \cdot P(Y=1) \implies X=2$ şi $Y=1$ sunt independente.
d) $P(X=2, Y=2) = 0{,}3 \neq 0{,}25 = 0{,}5 \cdot 0{,}5 = P(X=2) \cdot P(Y=2) \implies X$ şi Y nu sunt independente.
e) $P(X=1, Y=1 | X+Y=2) = 1 = P(X=1 | X+Y=2) \cdot P(Y=1 | X+Y=2) \implies$ $X=1$ şi $Y=1$ sunt condiţional independente, cunoscând $X+Y=2$.
f) $P(X=1, Y=2 | X+Y=3) = \frac{P(X=1,Y=2)}{P(X+Y=3)} = \frac{0{,}2}{0{,}3} \neq \frac{0{,}2}{0{,}3} \cdot \frac{0{,}2}{0{,}3} = P(X=1 | X+Y=3) \cdot P(Y=2 | X+Y=3) \implies X$ şi Y nu sunt condiţional independente, cunoscând $X+Y$.
g) $E(2X+Y^2) = 2E(X) + E(Y^2) = 2(1 \cdot 0{,}5 + 2 \cdot 0{,}5) + (-2)^2 \cdot 0{,}3 + 1^2 \cdot 0{,}2 + 2^2 \cdot 0{,}5 = 6{,}4$.

7. O monedă este aruncată de 10 ori. Fie X variabila aleatoare care indică diferenţa dintre numărul de capete şi numărul de pajuri obţinute. Determinaţi:
i) distribuţia de probabilitate a lui X;
ii) valoarea medie a lui X.
R: i) Dacă C şi P indică numărul de capete, respectiv de pajuri, atunci $C, P \sim \mathrm{Bino}(10, \frac{1}{2})$, $P = 10 - C$ şi $X = C - P = 2C - 10 \implies X \sim \begin{pmatrix} 2k-10 \\ C_{10}^k \frac{1}{2^{10}} \end{pmatrix}_{k=\overline{0,10}}$.
ii) $E(X) = E(C-P) = E(C) - E(P) = 0$, deoarece C şi P au aceeaşi distribuţie.