

## 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 numărului ales. Scrieți distribuția lui  $X$ , apoi calculați valoarea sa medie  $E(X)$ .

2. Considerăm următoarea problemă de *clasificare naivă Bayes* a unor restaurante (R), în

• clasele: **recomandat** sau **nerecomandat**,  
în funcție de următoarele *atribute* cu valorile lor posibile:

- cost (C): **ieftin**, **mediu**, **scump**;
- timp de așteptare (T): **puțin**, **mediu**, **îndelungat**;
- mâncare (M): **fadă**, **acceptabilă**, **bună**, **delicioasă**.

R, C, T, M sunt variabilele aleatoare (categoriale) și r, n, i, m, s, p, m, î, f, a, b, d valorile de mai sus, în ordinarea în care sunt menționate.

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

	<i>Cost</i>	<i>Timp de așteptare</i>	<i>Mâncare</i>	<i>Restaurant</i>
1	mediu	îndelungat	acceptabilă	nerecomandat
2	scump	puțin	bună	recomandat
3	ieftin	îndelungat	delicioasă	recomandat
4	mediu	puțin	bună	recomandat
5	ieftin	mediu	acceptabilă	nerecomandat
6	ieftin	puțin	fadă	nerecomandat
7	mediu	puțin	acceptabilă	nerecomandat
8	mediu	mediu	delicioasă	recomandat
9	scump	puțin	delicioasă	recomandat
10	ieftin	îndelungat	bună	nerecomandat
11	scump	puțin	acceptabilă	nerecomandat
12	mediu	mediu	bună	recomandat
13	mediu	îndelungat	fadă	nerecomandat
14	scump	mediu	delicioasă	recomandat
15	ieftin	mediu	fadă	nerecomandat
16	mediu	puțin	delicioasă	recomandat
17	ieftin	puțin	acceptabilă	recomandat
18	scump	îndelungat	bună	nerecomandat
19	ieftin	puțin	fadă	recomandat
20	scump	îndelungat	delicioasă	nerecomandat

i) Folosind datele din tabel, determinați probabilitățile claselor și probabilitățile condiționate ale atributelor, știind clasa.

ii) Considerăm evenimentul dat de *vectorul de atribută*:  $E = (C = s) \cap (T = m) \cap (M = b)$ . Alegeti o clasă pentru  $E$ , stabilind care din următoarele probabilități este mai mare:  $P(\mathbf{R} = \mathbf{r}|E)$  sau  $P(\mathbf{R} = \mathbf{n}|E)$ .

iii) Determinați  $P(E)$ .

## 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 numărului ales. Scrieți distribuția lui  $X$ , apoi calculați valoarea sa medie  $E(X)$ .

2. Considerăm următoarea problemă de *clasificare naivă Bayes* a unor restaurante (R), în

• clasele: **recomandat** sau **nerecomandat**,  
în funcție de următoarele *atribute* cu valorile lor posibile:

- cost (C): **ieftin**, **mediu**, **scump**;
- timp de așteptare (T): **puțin**, **mediu**, **îndelungat**;
- mâncare (M): **fadă**, **acceptabilă**, **bună**, **delicioasă**.

R, C, T, M sunt variabilele aleatoare (categoriale) și r, n, i, m, s, p, m, î, f, a, b, d valorile de mai sus, în ordinarea în care sunt menționate.

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

	<i>Cost</i>	<i>Timp de așteptare</i>	<i>Mâncare</i>	<i>Restaurant</i>
1	mediu	îndelungat	acceptabilă	nerecomandat
2	scump	puțin	bună	recomandat
3	ieftin	îndelungat	delicioasă	recomandat
4	mediu	puțin	bună	recomandat
5	ieftin	mediu	acceptabilă	nerecomandat
6	ieftin	puțin	fadă	nerecomandat
7	mediu	puțin	acceptabilă	nerecomandat
8	mediu	mediu	delicioasă	recomandat
9	scump	puțin	delicioasă	recomandat
10	ieftin	îndelungat	bună	nerecomandat
11	scump	puțin	acceptabilă	nerecomandat
12	mediu	mediu	bună	recomandat
13	mediu	îndelungat	fadă	nerecomandat
14	scump	mediu	delicioasă	recomandat
15	ieftin	mediu	fadă	nerecomandat
16	mediu	puțin	delicioasă	recomandat
17	ieftin	puțin	acceptabilă	recomandat
18	scump	îndelungat	bună	nerecomandat
19	ieftin	puțin	fadă	recomandat
20	scump	îndelungat	delicioasă	nerecomandat

i) Folosind datele din tabel, determinați probabilitățile claselor și probabilitățile condiționate ale atributelor, știind clasa.

ii) Considerăm evenimentul dat de *vectorul de atribută*:  $E = (C = s) \cap (T = m) \cap (M = b)$ . Alegeti o clasă pentru  $E$ , stabilind care din următoarele probabilități este mai mare:  $P(\mathbf{R} = \mathbf{r}|E)$  sau  $P(\mathbf{R} = \mathbf{n}|E)$ .

iii) Determinați  $P(E)$ .

$X: \Omega \rightarrow \mathbb{R}$

$X: \Omega \rightarrow \{x_1, \dots, x_n\}$

v. a. discretă finită

$X \sim (p_1, p_2, \dots, p_n)$

$$p_k = P(X = x_k)$$

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

$$S = \{ \overline{ab} \in \mathbb{N} \mid \overline{ab} \leq 50, a+b \text{ impar} \}$$

$$X \sim \left( \begin{array}{ccccccc} 1 & 3 & 5 & 7 & 9 & 11 & 13 \\ \frac{1}{21} & \frac{2}{21} & \frac{5}{21} & \frac{6}{21} & \frac{6}{21} & \frac{3}{21} & \frac{1}{21} \end{array} \right)$$

$$a+b=1 \quad \overline{ab} \in \{10\}$$

$$a+b=3 \quad \overline{ab} \in \{12, 21, 30\}$$

$$a+b=5 \quad \overline{ab} \in \{14, 23, 32, 41, 50\}$$

$$a+b=7 \quad \overline{ab} \in \{16, 25, 34, 43\}$$

$$a+b=9 \quad \overline{ab} \in \{18, 27, 36, 45\}$$

$$a+b=11 \quad \overline{ab} \in \{29, 38, 47\}$$

$$a+b=13 \quad \overline{ab} \in \{49\}$$

Scrieți distribuția lui  $X$ , apoi calculați valoarea sa medie  $E(X)$ .

2. Considerăm următoarea problemă de *clasificare naivă Bayes* a unor restaurante (**R**), în

- **clasele:** recomandat sau nerecomandat,

în funcție de următoarele *attribute* cu valorile lor posibile:

- **cost (C):** ieftin, mediu, scump;
- **temp de aşteptare (T):** puțin, mediu, îndelungat;
- **mâncare (M):** fadă, acceptabilă, bună, delicioasă.

- Timp de aşteptare ( $\text{r}$ ,  $\text{n}$ ,  $i$ ,  $m$ ,  $s$ ,  $p$ ,  $m$ ,  $\hat{i}$ ,  $f$ ,  $a$ ,  $b$ ,  $d$ ):  $\text{fadă}$ ,  $\text{acceptabilă}$ ,  $\text{bună}$ ,  $\text{delicioasă}$ .

$\text{R}$ ,  $\text{C}$ ,  $\text{T}$ ,  $\text{M}$  sunt variabilele aleatoare (categoriale) și  $\text{r}$ ,  $\text{n}$ ,  $i$ ,  $m$ ,  $s$ ,  $p$ ,  $m$ ,  $\hat{i}$ ,  $f$ ,  $a$ ,  $b$ ,  $d$  valorile de mai sus, în ordinarea în care sunt menționate.

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

	<i>Cost</i>	<i>Timp de așteptare</i>	<i>Mâncare</i>	<i>Restaurant</i>
1	mediu	îndelungat	acceptabilă	nerecomandat
2	scump	puțin	bună	recomandat
3	ieftin	îndelungat	delicioasă	recomandat
4	mediu	puțin	bună	recomandat
5	ieftin	mediu	acceptabilă	nerecomandat
6	ieftin	puțin	fadă	nerecomandat
7	mediu	puțin	acceptabilă	nerecomandat
8	mediu	mediu	delicioasă	recomandat
9	scump	puțin	delicioasă	recomandat
10	ieftin	îndelungat	bună	nerecomandat
11	scump	puțin	acceptabilă	nerecomandat
12	mediu	mediu	bună	recomandat
13	mediu	îndelungat	fadă	nerecomandat
14	scump	mediu	delicioasă	recomandat
15	ieftin	mediu	fadă	nerecomandat
16	mediu	puțin	delicioasă	recomandat
17	ieftin	puțin	acceptabilă	recomandat
18	scump	îndelungat	bună	nerecomandat
19	ieftin	puțin	fadă	recomandat
20	scump	îndelungat	delicioasă	nerecomandat

10 n

i) Folosind datele din tabel, determinați probabilitățile claselor și probabilitățile condiționate ale atributelor, știind clasa.

ii) Considerând evenimentul dat de existența rețelei de socializare  $R = \text{r}$ , să se calculeze  $P(C = \dots | R = \text{r})$  și  $P(T = \dots | R = \text{r})$ .

$R = \text{r}$	$R = \text{m}$	$P(R = \text{r})$	$P(R = \text{m})$
10	10	$\frac{1}{2}$	$\frac{1}{2}$

$C$	$R = \text{r}$	$R = \text{m}$	$P(C = \dots   R = \text{r})$	$P(C = \dots   R = \text{m})$
$i$	3	4	$3/10$	$4/10$
$m$	6	3	$6/10$	$3/10$
$\Delta$	3	3	$3/10$	$3/10$

$T$	$R = \text{r}$	$R = \text{m}$	$P(T = \dots   R = \text{r})$	$P(T = \dots   R = \text{m})$
$P$	6	3	$6/10$	$3/10$
$m$	3	2	$3/10$	$2/10$
$\Delta$	1	5	$1/10$	$5/10$

M	R=n	R=m	P(M=... R=n)	P(M=... R=m)
f	1	3	1/10	3/10
a	1	4	1/4	4/10
b	3	2	3/10	2/10
d	5	1	5/10	1/10

ii) Considerăm evenimentul dat de *vectorul de atrbute*:  $E = (C=s) \cap (T=m) \cap (M=b)$ . Alegeți o clasă pentru  $E$ , stabilind care din următoarele probabilități este mai mare:  $P(\mathbf{R}=\mathbf{r}|E)$  sau  $P(\mathbf{R}=\mathbf{n}|E)$ .

$$E = (C=d) \cap (T=m) \cap (M=b)$$

$$P(R=n|E) = \frac{P(t|R=n)}{P(E)} \cdot P(R=n)$$

$$= \frac{P((C=d) \cap (T=m) \cap (M=b) | R=n)}{P(E)} \cdot P(R=n)$$

$$= \frac{3}{10} \cdot \frac{3}{10} \cdot \frac{3}{10} \cdot \frac{1}{2} \cdot \frac{1}{P(E)} = \frac{24}{2000} \cdot \frac{1}{P(E)}$$

$$P(R=m|E) = \frac{3}{10} \cdot \frac{2}{10} \cdot \frac{2}{10} \cdot \frac{1}{2} \cdot \frac{1}{P(E)} = \frac{12}{2000} \cdot \frac{1}{P(E)}$$

$$P(R=n|E) > P(R=m|E) \quad \text{recomandat}$$

$$P(R=n|E) + P(R=m|E) = 1$$

Determinați  $P(E)$ .

$$\frac{24}{2000} \cdot \frac{1}{P(E)} + \frac{12}{2000} \cdot \frac{1}{P(E)} = 1$$

$$\Rightarrow P(E) = \frac{39}{2000} = \frac{19,5}{1000}$$

3. Ce valoare teoretică estimează programul următor? Calculați valoarea teoretică corepunzătoare.

```
[ ]: import numpy as np
```

$N=2000$

$S = np.concatenate((np.zeros(50), np.ones(70), 2*np.ones(80)))$

$X = []$

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("Val. medie:", np.mean(X))

\* distribuția geometrică

$\times \sim Geo(p)$

$Geo(p) \sim \left( \frac{p}{q(1-p)} \right)^k$

$$p = 5\% / 200 = 1/4$$

$\underbrace{0 \dots 0}_{50} \underbrace{1 \dots 1}_{70} \underbrace{2 \dots 2}_{80}$

{ până la apariția valorii 0

$$E(x) = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k \cdot p(1-p)^k$$

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

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

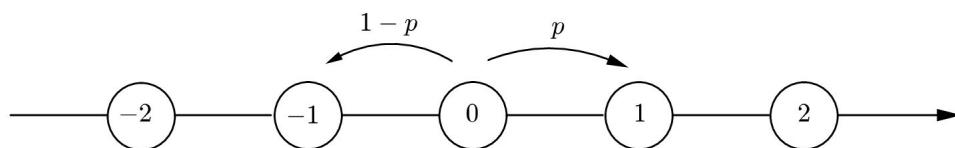
a) Determinați valoarea lui  $c$ .

b) Pentru  $a = 3$  și  $b = 21$ , calculați

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

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

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 și cu probabilitatea  $1 - p$  la stânga. Nodurile sunt centrate în numerele întregi:



Fie  $X$  variabila aleatoare care indică poziția finală a punctului material după  $n \in \mathbb{N}$  pași ai unei deplasări ce pornește din nodul 0. Determinați distribuția și valoarea medie lui  $X$ .

6. Considerăm vectorul aleator discret  $(X, Y)$  cu distribuția dată sub formă tabelară:

$$\textcircled{5} \quad p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 + \dots$$

$$(1-p) + p + 2p(1-p) + 3p(1-p)^2 + \dots$$

$$(1-p) \sum_{k=0}^{\infty} (k+1)p(1-p)^k = (1-p) \left[ \sum_{k=0}^{\infty} kp(1-p)^k + \sum_{k=0}^{\infty} p(1-p)^k \right]$$

$$E(X) = (1-p) E(X) + (1-p) p \sum_{k=0}^{\infty} (1-p)^k$$

$$p E(X) = p(1-p) \cdot \frac{1}{1-(1-p)} \Rightarrow p \cdot E(X) = 1-p$$

$$E(X) = \frac{1-p}{p} = 3$$

$$1-p = q \in (0,1)$$

$\downarrow$

$$\frac{3}{4}$$

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

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

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

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

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

$$\text{a) } c(b-a+1)=1$$

$$c = \frac{1}{b-a+1}$$

$$\sum_{k=a}^b P(X=k) = 1$$

$$\text{b) } P\left(\left\{X \leq \frac{a+b}{2}\right\} \cup \left\{\frac{a+b}{6} \leq X\right\}\right)$$

$$\frac{a+b}{2} = 12$$

$$P\left(\left\{X \leq \frac{a+b}{2}\right\} \cap \left\{\frac{a+b}{6} \leq X\right\}\right)$$

$$\frac{a+b}{6} = 4$$

$$X \sim \begin{pmatrix} 3 & \dots & 21 \\ \frac{1}{12} & \dots & \frac{1}{12} \end{pmatrix}$$

$$P = \left( \left\{ x \leq \frac{a+b}{2} \right\} \cap \left\{ \frac{a+b}{6} \leq x \right\} \right) =$$

$$= P(\{x \leq 12\} \cap \{4 \leq x\}) = P\{x \in (4, 12)\}$$

$$= \frac{12-4+1}{12} = \frac{9}{12}$$

c)  $P(x=a) = \frac{1}{3}, E(x)=1$

$$E(x) = a \cdot c + (a+1)c + \dots + b \cdot e$$

$$P(x=a) = \frac{1}{3} \Rightarrow c = \frac{1}{3} = \frac{1}{b-a+1}$$

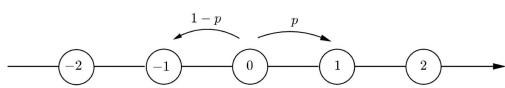
$$E(x) = -c[a+a+1+a+2]$$

$$= \frac{1}{3}(3a+3) = a+1 = 1 \Rightarrow a=0, b=2$$

$$\Rightarrow b-a=2 \Rightarrow$$

$$\Rightarrow b=a+2$$

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 și cu probabilitatea  $1-p$  la stânga. Nodurile sunt centrate în numerele întregi:



Fie  $X$  variabila aleatoare care indică poziția finală a punctului material după  $n \in \mathbb{N}$  pași ai unei deplasări ce pornește din nodul 0. Determinați distribuția și valoarea medie a lui  $X$ .

$$Y_1 \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix} \sim 2X_1 - 1$$

$$\text{Bernoulli}(p) \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} = X_1 = X_i$$

$$\text{Binomiala}(p) \sim \binom{n}{k} p^k (1-p)^{n-k} \sim X_1 + X_2 + \dots + X_n$$

$$\text{pașul } m : Y_m \sim 2X_{m-1} \quad X = Y_1 + \dots + Y_m$$

$$X \sim \binom{2k-m}{m} p^k (1-p)^{m-k}$$

$$= (2X_1 - 1) + \dots + (2X_{m-1} - 1)$$

$$= 2(X_1 + \dots + X_{m-1}) - m = 2 \text{Bino}(p) - m$$

$$E(\text{Bino}(p)) = m \cdot p$$

$$E(X) = E(2 \cdot \text{Bino}(p) - m) = 2E(\text{Bino}(p)) - m = 2mp - m$$



**6.** Considerăm vectorul aleatoar discret  $(X, Y)$  cu distribuția 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 distribuțiile de probabilitate ale variabilelor aleatoare  $X$  și  $Y$ .
- b) Calculați probabilitatea ca  $|X - Y| = 1$ , știind că  $Y > 0$ .
- c) Sunt evenimentele  $X = 2$  și  $Y = 1$  independente?
- d) Sunt variabilele aleatoare  $X$  și  $Y$  independente?
- e) Sunt evenimentele  $X = 1$  și  $Y = 1$  condițional independente, cunoscând  $X + Y = 2$ ?
- f) Este variabila aleatoare  $X$  condițional independentă de  $Y$ , cunoscând  $X + Y$ ?
- g) Calculați valoarea medie a variabilei aleatoare  $2X + Y^2$ .

**7.** O monedă este aruncată de 10 ori. Fie  $X$  variabila aleatoare care indică diferența dintre numărul de capete și numărul de pajuri obținute. Determinați:

- i) distribuția de probabilitate a lui  $X$ ;
- ii) valoarea medie a lui  $X$ .