# SEMINAR 4

Modelul urnei cu bile de 2 culori și bilă nereturnată: fie $n_1, n_2, n \in \mathbb{N}$ cu $n \le n_1 + n_2$ și fie $k \in \mathbb{N}$ a.î. $k \le n_1$ și $n - k \le n_2$; considerăm o urnă, care are inițial $n_1$ bile albe și $n_2$ bile negre, avem

$p(k; n)$ = prob. de a obține $k$ bile albe din $n$ extrageri fără returnare, în care ordinea de extragere nu contează

$$= \frac{C_{n_1}^{k} \cdot C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}}$$

DISTRIBUȚIE HIPERGEOMETRICĂ

① Dintr-un set de 52 cărți de joc se extrag aleator, pe rând, fără returnare, 13 cărți. Calculați probabilitatea: a) A: „nu s-a extras nicio treflă"

b) B: „s-au obținut 5 inimi"

c) C: „s-a obținut cel mult un as"

pachet 52 cărți

13 fiecare culoare

a) $\frac{C_{13}^{0} \cdot C_{39}^{13}}{C_{52}^{13}}$ → aplicare formulă propusă

b) $p(5, 13) = \frac{C_{13}^{5} \cdot C_{39}^{8}}{C_{52}^{13}}$

c) unul sau niciunul: $p(1, 13) + p(0, 13) = \frac{C_4^1 \cdot C_{48}^{12}}{C_{52}^{13}} + \frac{C_4^0 \cdot C_{48}^{13}}{C_{52}^{13}}$

sunt 4 ași în pachet

Modelul urnei cu $n$ culori și bila nereturnată: fie $n_i$ = nr. inițial de bile de culoarea $i$ din urnă, $i = \overline{1,r}$

$p(k_1, \dots, k_r; n)$ = prob. de a obține $k_i$ bile de culoarea $i$, $i = \overline{1,r}$ din $n = k_1 + \dots + k_r$ extrageri fără returnare, în care ordinea de extragere nu contează

$$= \frac{C_{n_1}^{k_1} \cdot \ldots \cdot C_{n_r}^{k_r}}{C_{n_1+\dots+n_r}^{n}}$$

$n = 2$ corespunde distribuției hipergeometrice

② O echipă formată din 4 cercetători este aleasă aleator dintr-un grup de 4 matematicieni, 3 informaticieni și 5 fizicieni. Care e prob ca echipa să fie formată din 2 mat., 1 info., 3 fizicieni?

$$p(2, 1, 3) = \frac{C_4^2 \cdot C_3^1 \cdot C_5^3}{C_{12}^4}$$ → doar formula

Clasificarea naivă Bayes:

Clasificatorii bayesieni naivi sunt o familie de clasificatori probabilistici simpli, bazați pe aplicarea formulei lui Bayes cu ipoteze "naive" de independență condiționată între atribute, cunoscând clasificarea. În aplicațiile practice pt. modelele bayesiene naive se folosește metoda prob. maxime. Noțiunea folosită în acest context este conditional independența între variabilele aleatoare.

**Def**: Fie $U, X, Y, Z$ v.a. discrete, care iau valori în mulțimile $\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$. V.a. $U, X, Y$ sunt condițional independente, cunoscând (știind) v.a. $Z$ dacă pt. fiecare $u \in \mathcal{U}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ are loc

$$P(U = u, X = x, Y = y \mid Z = z) = P(U = u \mid Z = z) P(X = x \mid Z = z) P(Y = y \mid Z = z).$$

③ Considerăm urm. problemă cu clasificare naivă Bayes a unor restaurante(R) în clasele: recomandat (r), nerecomandat (n), în funcție de urm. atribute cu valorile lor posibile:

cost (c): ieftin, mediu, scump

timp de așteptare (T): puțin, mediu, îndelungat

mâncare (M): fadă, acceptabilă, bună, delicioasă

i) folosind datele din tabel, det. prob. claselor și prob. condiționate ale atributelor ~~lor~~, știind clasa

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

| C | R=n | R=r | P(C=...\|R=n) | P(C=...\|R=r) |
|---|---|---|---|---|
| m | 3 | 4 | 3/10 | 4/10 |
| s | 3 | 3 | 3/10 | 3/10 |
| i | 4 | 3 | 4/10 | 3/10 |

| T | R=n | R=r | P(T=...\|R=n) | P(T=...\|R=r) |
|---|---|---|---|---|
| p | 6 | 3 | 6/10 | 3/10 |
| m | 3 | 2 | 3/10 | 2/10 |
| i | 1 | 5 | 1/10 | 5/10 |

| M | R=n | R=r | P(M=...\|R=n) | P(M=...\|R=r) |
|---|---|---|---|---|
| a | 4 | 1 | 4/10 | 1/10 |
| b | 2 | 3 | 2/10 | 3/10 |
| d | 1 | 5 | 1/10 | 5/10 |
| f | 3 | 1 | 3/10 | 1/10 |

ii) considerăm evenimentul dat de vectorul de atribute: $E=(C=s)\cap(T=m)\cap(M=b)$. Alegeți o clasă pt E, stabilind care din urm. prob. este mai mare: $P(R=r|E)$ sau $P(R=n|E)$.

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

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

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

Clasa aleasă pt R este R=r (e mai mare)!

iii) Determinați $P(E)$ → sistem complet

$$1=P(R=n|E)+P(R=r|E)=\frac{39}{2000}\cdot\frac{1}{P(E)}\Rightarrow P(E)=\frac{39}{2000}$$ (prob. ca restaurantul să aibă caracteristicile de sus)

Un mesaj este transmis printr-un canal de comunicare cu perturbări. Probabilitatea ca mesajul să fie recepționat este de 10%. Dacă mesajul nu este recepționat, atunci se reia transmisia mesajului, independent de transmisiile anterioare. Fie X variabila aleatoare care indică nr. de transmisii până la prima în care mesajul e recepționat. Det. valoarea medie a lui X.

primul succes după x ratări → distribuția geometrică

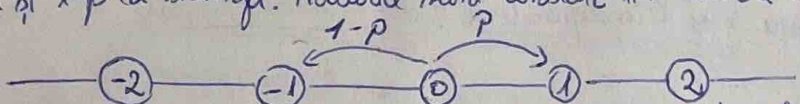
$X\begin{pmatrix} k \\ p(1-p)^k\end{pmatrix}_{k=0,1,2,\dots}$ $\quad p=\frac{1}{10}$ → dacă era continuu făceam $\int_{\mathbb{R}}$

$$E(X)=\sum_{k=0}^{\infty} k\cdot p(1-p)^k \underset{k=j+1}{=} (1-p)\sum_{k=1}^{\infty} k\,p(1-p)^{k-1}$$ = (primul de sus · primul de jos + al doilea sus · ...)

$$=(1-p)\sum_{j=0}^{\infty} (j+1)p(1-p)^j = (1-p)\sum_{j=0}^{\infty} j\,p(1-p)^j + (1-p)\sum_{j=0}^{\infty} p(1-p)^j =$$

$$=(1-p)E(X) + (1-p)p\cdot\underbrace{\sum_{j=0}^{\infty} (1-p)^j}_{\text{progresie geometrică}=\frac{1}{p}} = (1-p)E(X)+(1-p) \Rightarrow E(X)=\frac{1-p}{p}=9$$

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




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

dacă $Y_i$ repr. pasul i, atunci $Y_i \sim \begin{pmatrix}-1 & 1\\ 1-p & p\end{pmatrix}$ → distribuția unui pas $\Rightarrow Y_i = 2X_i - 1$ cu $X_i \sim$ Bernoulli$(p)$, $i\in\{1,\dots,n\}$

$X = Y_1+\dots+Y_n = (2X_1-1)+\dots+(2X_n-1)$, $X_1+\dots+X_n\sim$ Bino$(n,p)$ $\Rightarrow$

$\Rightarrow X\sim\begin{pmatrix}2k-n\\ C_n^k p^k(1-p)^{n-k}\end{pmatrix}_{k=\overline{0,n}}$ și $E(X)=2np-n$

Σ Bernoulli urmează Bino

doar am încercat să facem legătura cu Bernoulli pe care-l știm afla

⑦ O monedă e aruncată de 10 ori. Fie X var. a. care indică nr. capete - nr. pajuri. Det:

i) distribuția de probabilitate a lui X ii) valoarea medie a lui X

i) C, P → nr. de capete și pajuri $\Rightarrow$ C, P ~ Bino$(10,\frac{1}{2})$, $P=10-C$ și $X=C-P=2C-10 \Rightarrow X\sim\begin{pmatrix}2k-10\\ C_{10}^k\frac{1}{2^{10}}\end{pmatrix}_{k=\overline{0,10}}$

$X\sim$ Bino$(n,p) \Leftrightarrow X\sim\begin{pmatrix}k\\ C_n^k p^k(1-p)^{n-k}\end{pmatrix}_{k\in\{0,\dots,n\}}$ DISTRIBUȚIA BINOMIALĂ

ii) $E(X)=E(C-P)=E(C)-E(P)=0$ pt. că C și P au aceeași distribuție

au ace val medie

~~tulle una~~

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

| X \ Y | -2 | 1 | 2 |
|---|---|---|---|
| 1 | 0,2 | 0,1 | 0,2 |
| 2 | 0,1 | 0,1 | 0,3 |

a) să se det. distribuțiile de probabilitate ale var. aleatoare X și Y

$P(X=x_i)=\sum_{j\in J} p_{ij}, \forall i\in J$ vector aleator discret

$X \sim \begin{pmatrix} 1 & 2 \\ 0,5 & 0,5 \end{pmatrix}$ $\quad Y = \begin{pmatrix} -2 & 1 & 2 \\ 0,3 & 0,2 & 0,5 \end{pmatrix}$ $\quad Y^2 = \begin{pmatrix} 4 & 1 & 4 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$

b) calculați prob. ca $|X-Y|=1$, știind că $Y>0$

$P(|X-Y|=1 \mid Y>0) = \frac{P(|X-Y|=1 \cap Y>0)}{P(Y>0)} = \frac{P(X=1,Y=2)+P(X=2,Y=1)}{P(Y>0)} = \frac{0,2+0,1}{0,7} = \frac{0,3}{0,7}$

c) sunt evenimentele $X=2$ și $Y=1$ independente?

$P(X=2, Y=1) = 0,1$
$P(X=2)\cdot P(Y=1) = 0,5\cdot 0,2 = 0,1$ $\Big\} \Rightarrow$ sunt independente dacă $P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$ atunci sunt independente

d) Sunt var. aleatoare X și Y independente? dăm contraexemplu

$P(X=2,Y=2) = 0,3$
$P(X=2)\cdot P(Y=2) = 0,5\cdot 0,5 = 0,25$ $\Big\}$ nu sunt independente

e) Sunt even. $X=1$ și $Y=1$ condițional independente, cunoscând $X+Y=2$?

$P(X=1, Y=1 \mid X+Y=2) = 1$
$P(X=1 \mid X+Y=2) = 1$ (nu există altă posibilitate pt $X+Y=1$ decât $X=1$) $\Big\} \Rightarrow$ sunt condițional independente
$P(Y=1 \mid X+Y=2) = 1$

f) este var. a. X condițional independentă de X știind $X+Y$? contraexemplu

3 poate fi 2+1 sau 1+2

$P(X=1, Y=2 \mid X+Y=3) = \frac{P(X=1,Y=2)}{P(X+Y=3)} = \frac{0,2}{0,2+0,1}$

$P(X=1 \mid X+Y=3) = \frac{0,5}{0,2+0,1}$ ← de unde 0,2 ?? la el

$P(Y=2 \mid X+Y=3) = \frac{0,5}{0,3}$

g) calculați val. medie a v.a. $2X+Y^2$

$E(2X+Y^2) = 2E(X) + E(Y^2) = 2(1\cdot 0,5 + 2\cdot 0,5) + (-2)^2\cdot 0,3 + 1^2\cdot 0,2 + 2^2\cdot 0,5 = 6,4$

↓

ridicarea la putere nu poate fi trecută înafară