Variabile aleatoare

9. **(Variabile aleatoare: proprietăţi de bază pentru medii, varianţă, covarianţă)**

Fie variabila aleatoare $X : \Omega \to R$, cu funcţia de probabilitate P.

Dacă X este variabilă aleatoare *discretă*, atunci prin definiţie $P(x) \stackrel{not.}{=} P(X = x) \stackrel{not.}{=} P(\{\omega \mid X(\omega) = x\}) \geq 0$ pentru orice $x \in \mathbb{R}$, şi $\sum_{x_i \in Val(X)} P(x_i) = 1$, unde $Val(X)$ este mulţimea valorilor variabilei aleatoare X.

Dacă X este variabilă aleatoare *continuă*, având funcţia densitate de probabilitate p, atunci prin definiţie $p(X = x) \geq 0$ pentru orice $x \in \mathbb{R}$, şi $\int_{-\infty}^{\infty} p(X = x)dx = 1$ (sau, scris mai simplu: $\int_{-\infty}^{+\infty} p(x)dx = 1$).

a. ■ ***CMU, 2010 spring, E. Xing, T. Mitchell, A. Singh, HW1, pr. 1.1***

Dacă X este variabilă aleatoare discretă, *media* sa se defineşte ca fiind numărul real $E[X] = \sum_{x_i \in Val(X)} x_i \cdot P(X = x_i)$. Dacă X este variabilă aleatoare continuă, media sa este $E[X] = \int_{-\infty}^{\infty} x \cdot p(X = x)dx$.

Arătaţi că pentru orice două variabile aleatoare W şi Z de acelaşi tip (adică fie ambele discrete fie ambele continue), având acelaşi domeniu de definiţie (Ω), avem

$$E[W + Z] = E[W] + E[Z].$$

b. ***CMU, 2010 spring, E. Xing, T. Mitchell, A. Singh, HW1, pr. 1.3***

Fie X o variabilă aleatoare. Notăm $\bar{X} = E[X]$. *Varianţa* lui X se defineşte ca fiind $Var(X) = E[(X - \bar{X})^2]$. Arătaţi că:

$$Var(X) = E[X^2] - (E[X])^2.$$

***Indicaţie*: Nu este necesar să faceţi demonstraţia separat pentru cele două cazuri, discret şi respectiv continuu.**

c. ***CMU, 2009 spring, Tom Mitchell, HW2, pr. 1.3.1***

***Covarianţa* a două variabile aleatoare X şi Y care au acelaşi domeniu de definiţie (Ω) se defineşte astfel: $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$, unde $E[X]$ este media lui X.**

Demonstraţi egalitatea:

$$Cov(X, Y) = E[XY] - E[X] \cdot E[Y].$$

Răspuns:

a. Pentru cazul discret vom folosi o formă echivalentă a formulei pentru media unei variabile aleatoare şi anume $E[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$. Prin urmare, putem

scrie:

$$\begin{aligned}
E[W+Z] &= \sum_{u \in Val(W+Z)} u \cdot P(W+Z=u) \\
&= \sum_{w \in Val(W), z \in Val(Z)} (w+z) \cdot P(W+Z=w+z) \\
&= \sum_{w \in Val(W)} \sum_{z \in Val(Z)} (w+z) \cdot P(\{\omega \in \Omega \mid (W+Z)(\omega) = w+z\}) \\
&= \sum_{w \in Val(W)} \sum_{z \in Val(Z)} w \cdot P(\{\omega \in \Omega \mid (W+Z)(\omega) = w+z\}) + \\
& \quad \sum_{w \in Val(W)} \sum_{z \in Val(Z)} z \cdot P(\{\omega \in \Omega \mid (W+Z)(\omega) = w+z\}) \\
&= \sum_{w \in Val(W)} w \sum_{z \in Val(Z)} P(\{\omega \in \Omega \mid W(\omega) = w, Z(\omega) = z\}) + \\
& \quad \sum_{z \in Val(Z)} z \sum_{w \in Val(W)} P(\{\omega \in \Omega \mid W(\omega) = w, Z(\omega) = z\}) \\
&= \sum_{w \in Val(W)} wP(\{\omega \in \Omega \mid W(\omega) = w\}) + \sum_{z \in Val(Z)} zP(\{\omega \in \Omega \mid Z(\omega) = z\}) \\
&= E[W] + E[Z]
\end{aligned}$$

Pentru cazul continuu:

$$\begin{aligned}
E[W+Z] &= \int_w \int_z (w+z)\, p_{WZ}(w,z) dz dw \\
&= \int_w \int_z w\, p_{WZ}(w,z) dz dw + \int_w \int_z z\, p_{WZ}(w,z) dz dw \\
&= \int_w w \int_z p_{WZ}(w,z) dz dw + \int_z z \int_w p_{WZ}(w,z) dw dz \\
&= \int_w w p_W(w) dw + \int_z z p_Z(z) dz \\
&= E[W] + E[Z]
\end{aligned}$$

b. Pentru a demonstra această proprietate vom ţine cont de liniaritatea mediei unei variabile aleatoare — $E[aX+b] = aE[X]+b$, o proprietate care se demonstrează imediat —, precum şi de faptul că $E[X+Y] = E[X]+E[Y]$, unde X, Y sunt variabile aleatoare iar a şi b sunt constante reale:

$$\begin{aligned}
Var(X) &= E[(X-\bar{X})^2] = E[X^2 - 2X\bar{X} + \bar{X}^2] \\
&= E[X^2] - E[2X\bar{X}] + E[\bar{X}^2] = E[X^2] - 2\bar{X}E[X] + \bar{X}^2 \\
&= E[X^2] - 2\bar{X}^2 + \bar{X}^2 = E[X^2] - \bar{X}^2 = E[X^2] - (E[X])^2
\end{aligned}$$

c. Se foloseşte acelaşi gen de raţionament ca la punctul precedent:

$$\begin{aligned}
Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\
&= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]
\end{aligned}$$

10. **(Rezultat teoretic: covarianţa oricăror 2 variabile aleatoare independente este 0)**

■ ***CMU, 2010 spring, E. Xing, T. Mitchell, A. Singh, HW1, pr. 1.2***

În mod intuitiv, două variabile aleatoare X şi Y sunt *independente* atunci când cunoaşterea valorii uneia dintre ele (de exemplu X) nu furnizează niciun indiciu despre valoarea celeilalte variabile (Y, în acest caz).

Formal, dacă X şi Y sunt variabile aleatoare discrete, independenţa lor revine la egalitatea $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ pentru orice $x \in Val(X)$ şi orice $y \in Val(Y)$.

Similar, dacă X şi Y sunt variabile aleatoare continue, atunci $p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$ pentru orice valori x şi y posibile.

Arătaţi că dacă X şi Y sunt variabile aleatoare independente de acelaşi tip (adică fie discret fie continuu), atunci

$$E[XY] = E[X] \cdot E[Y].$$

Echivalent: X, Y independente $\Rightarrow Cov(X, Y) = 0$. (A se vedea problema 9 punctul c.)

***Observaţie*: Reciproca implicaţiei de mai sus nu este în general adevărată. A se vedea problema 11.**

<u>Răspuns:</u>

Pentru cazul în care variabilele aleatoare independente X şi Y sunt discrete, putem scrie:

$$\begin{aligned}
E[XY] &= \sum_{x \in Val(X)} \sum_{y \in Val(Y)} xyP(X = x, Y = y) \\
&\stackrel{indep.}{=} \sum_{x \in Val(X)} \sum_{y \in Val(Y)} xyP(X = x) \cdot P(Y = y) \\
&= \sum_{x \in Val(X)} \left(xP(X = x) \sum_{y \in Val(Y)} yP(Y = y) \right) \\
&= \sum_{x \in Val(X)} xP(X = x)E[Y] = \left(\sum_{x \in Val(X)} xP(X = x) \right) E[Y] \\
&= E[X] \cdot E[Y]
\end{aligned}$$

Pentru cazul continuu demonstraţia este similară:

$$\begin{aligned}
E[XY] &= \int_x \int_y xy\, p(X = x, Y = y) dy dx \\
&\stackrel{indep.}{=} \int_x \int_y xy\, p(X = x) \cdot p(Y = y) dy dx \\
&= \int_x x\, p(X = x) \int_y y\, p(Y = y) dy\, dx = \int_x x\, p(X = x) E[Y] dx \\
&= E[Y] \cdot \int_x x\, p(X = x) dx = E[Y] \cdot E[X] = E[X] \cdot E[Y]
\end{aligned}$$

11. **(Covarianţa nulă nu implică în mod neapărat independenţa variabilelor aleatoare)**

CMU, 2009 spring, Tom Mitchell, HW2, pr. 1.3.2
CMU, 2009 fall, Geoff Gordon, HW1, pr. 3.1

X	Y	$P(X,Y)$
0	0	1/3
1	0	0
2	0	1/3
0	1	0
1	1	1/3
2	1	0

a. Reciproca afirmaţiei din problema 10 nu este în general adevărată, deci $Cov(X,Y) = 0 \not\Rightarrow X$ şi Y sunt independente. Arătaţi aceasta folosind ca exemplu variabilele aleatoare ale căror distribuţii sunt date în tabelul alăturat.

b. Totuşi, dacă X şi Y sunt variabile aleatoare binare luând valori în mulţimea $\{0,1\}$, iar $E[XY] = E[X] \cdot E[Y]$, atunci X şi Y sunt independente. Justificaţi.

(Aşadar, două variabile aleatoare binare cu valori în mulţimea $\{0,1\}$ sunt independente atunci şi numai atunci când au covarianţa nulă.)

<u>Răspuns:</u>

a. Din datele furnizate în exemplul a rezultă următoarele probabilităţi:

$$
\begin{array}{lll}
P(X=0) = 1/3 & P(Y=0) = 2/3 & P(XY=0) = 2/3 \\
P(X=1) = 1/3 & P(Y=1) = 1/3 & P(XY=1) = 1/3 \\
P(X=2) = 1/3 & & P(XY=2) = 0
\end{array}
$$

Putem calcula mediile acestor variabile aleatoare folosind formula de definiţie pentru variabile discrete $E[X] = \sum_x x \cdot P(X=x)$:

$$
E[X] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1
$$
$$
E[Y] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}
$$
$$
E[XY] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} + 2 \cdot 0 = \frac{1}{3}
$$

Conform formulei care a fost demonstrată la problema 10, covarianţa variabilelor X şi Y este: $Cov(X,Y) = E[XY] - E[X] \cdot E[Y] = \frac{1}{3} - 1 \cdot \frac{1}{3} = 0$. Cu toate acestea, variabilele X şi Y nu sunt independente. Într-adevăr, pentru $X = 0$ şi $Y = 0$ se observă că $P(X=0, Y=0) = 1/3$ dar $P(X=0)P(Y=0) = 1/3 \cdot 2/3 = 2/9 \neq \frac{1}{3}$.

Prin acest contra-exemplu am arătat că implicaţia „$Cov(X,Y) = 0 \Rightarrow X$ şi Y sunt independente“ nu este în general adevărată.

b. În continuare vom demonstra că în cazul în care X şi Y iau valori în mulţimea $\{0,1\}$ implicaţia de mai sus este adevărată.

Dacă X şi Y sunt variabile aleatoare binare, atunci:

$$
\begin{aligned}
E[X] &= 0 \cdot P(X=0) + 1 \cdot P(X=1) = P(X=1) \\
E[Y] &= P(Y=1) \\
E[XY] &= P(X=1, Y=1)
\end{aligned}
$$

Covarianţa nulă înseamnă — conform problemelor 9 şi 10 — că $E[XY] = E[X] \cdot E[Y]$, adică:

$$
P(X=1, Y=1) = P(X=1)P(Y=1)
$$

Însă, a demonstra independenţa variabilelor aleatoare X şi Y revine la a arăta că $P(X,Y) = P(X)P(Y)$ pentru toate combinaţiile posibile de valori ale variabilelor. Unul din cazuri ($X = 1, Y = 1$) este deja demonstrat, deci mai există încă 3 cazuri. Pentru acestea vom utiliza formulele: $P(A \cap B) = P(A) - P(A \cap \bar{B})$ şi $P(\bar{A}) = 1 - P(A)$.

Cazul $X = 1, Y = 0$:

$$\begin{aligned} P(X=1,Y=0) &= P(X=1) - P(X=1,Y=1) \\ &= P(X=1) - P(X=1)P(Y=1) \\ &= P(X=1)(1-P(Y=1)) \\ &= P(X=1)P(Y=0) \end{aligned}$$

Cazul $X = 0, Y = 1$ se tratează similar cu cazul anterior:

$$\begin{aligned} P(X=0,Y=1) &= P(Y=1) - P(X=1,Y=1) \\ &= P(Y=1) - P(X=1)P(Y=1) \\ &= P(Y=1)(1-P(X=1)) \\ &= P(Y=1)P(X=0) \end{aligned}$$

Cazul $X = 0, Y = 0$:

$$\begin{aligned} P(X=0,Y=0) &= P(X=0) - P(X=0,Y=1) \\ &= P(X=0) - P(X=0)P(Y=1) \\ &= P(X=0)(1-P(Y=1)) \\ &= P(X=0)P(Y=0) \end{aligned}$$

Prin urmare, egalitatea $P(X,Y) = P(X)P(Y)$ este adevărată pentru toate cazurile, deci variabilele X şi Y sunt independente.

12. **(Variabile aleatoare: calcul de medii)**

Fie X o variabilă aleatoare pentru care $E(X) = \mu$ şi $Var(X) = \sigma^2$.

a. ***CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, HW1, pr. 2.4***

Cât este $E[X(X-1)]$ în funcţie de μ şi σ?

b. ***CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, midterm, pr. 1.c***

Fie $c \in R$. Care din următoarele variante sunt adevărate?

A. $E[(X-c)^2] = (\mu-c)^2 + \sigma^2$ **D.** $E[(X-c)^2] = (\mu-c)^2 + 2\sigma^2$
B. $E[(X-c)^2] = (\mu-c)^2$ **E.** $E[(X-c)^2] = \mu^2 + c^2 + 2\sigma^2$
C. $E[(X-c)^2] = (\mu-c)^2 - \sigma^2$ **F.** $E[(X-c)^2] = \mu^2 + c^2 - 2\sigma^2$.

Răspuns:

a. De la problema 9.a ştim că media sumei a două variabile aleatoare este suma mediilor variabilelor respective. De asemenea, este imediat demonstrabil că

$E[c \cdot X] = c \cdot E[X]$, unde X, Y sunt variabile aleatoare iar c este o constantă. Ţinând cont că $Var(X) = E[X^2] - (E[X])^2$ (vezi problema 9.b), putem scrie:

$$\begin{aligned} E[X(X-1)] &= E[X^2 - X] = E[X^2] - E[X] \\ &= E[X^2] - (E[X])^2 + (E[X])^2 - E[X] \\ &= Var(X) + (E[X])^2 - E[X] \\ &= \sigma^2 + \mu^2 - \mu \end{aligned}$$

b. Pentru a găsi varianta adevărată vom calcula $E[(X-c)^2]$:

$$\begin{aligned} E[(X-c)^2] &= E[X^2 - 2cX + c^2] = E[X^2] - 2cE[X] + c^2 \\ &= E[X^2] - (E[X])^2 + (E[X])^2 - 2cE[X] + c^2 \\ &= \sigma^2 + \mu^2 - 2c\mu + c^2 \\ &= \sigma^2 + (\mu - c)^2 \end{aligned}$$

Deci varianta A este adevărată.

13. **(Variabile aleatoare discrete – distribuţia Bernoulli – şi evenimente aleatoare identic distribuite: calcul de medii)**

CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, HW1, pr. 2.2

Un iepuraş se joacă „de-a săritelea“. Poziţia lui iniţială coincide cu originea axei reale. După aceea, iepuraşul face câte un salt de-a lungul axei, fie la stânga fie la dreapta. Pentru a determina în ce direcţie să sară, iepuraşul dă cu banul. Dacă obţine stema, va sări spre dreapta, iar dacă obţine banul, va sări spre stânga. Probabilitatea de a obţine stema este p. Se presupune că toate salturile iepuraşului au aceeaşi lungime, şi anume 1.

Care va fi poziţia (medie) la care ne aşteptăm să fie iepuraşul după ce face n salturi?

Răspuns:

Fiecare săritură a iepuraşului este modelată de o variabilă aleatoare. Un salt spre dreapta înseamnă o deplasare cu $+1$ pe axă, iar un salt spre stânga -1. Să notăm cu X_i variabila aleatoare corespunzătoare săriturii i. Aceasta este:

$$X_i : \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$$

Media acestei variabile aleatoare este $E[X_i] = -1 \cdot (1-p) + 1 \cdot p = 2p - 1$.

Ţinând cont de proprietatea de liniaritate a mediilor, poziţia iepuraşului după n salturi este de aşteptat să fie:

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = n(2p-1).$$

De exemplu, pentru $p = 1/2$, se va obţine poziţia 0 pe axa reală (aşa cum este de aşteptat dacă n este număr par), în vreme ce pentru $p = 2/3$ va rezulta poziţia $n/3$ (dacă $n/3 \in \mathbb{N}$), iar pentru $p = 1/3$ poziţia $-n/3$ (similar).

14. **(Distribuţii de probabilitate condiţionale: regula de multiplicare)**

CMU, 2009 fall, Geoff Gordon, HW2, pr. 2.1

Arătaţi că pentru orice valori x, y şi z ale variabilelor aleatoare X, Y şi Z respectiv, avem:

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

În notaţie simplificată: $P(X, Y \mid Z) = P(X \mid Y, Z) \cdot P(Y \mid Z)$.

***Indicaţie*: Folosiţi regula de înlănţuire pentru evenimente aleatoare.**

Răspuns:

Folosind definiţia probabilităţii condiţionale şi regula de înlănţuire (cu termenii ordonaţi în mod convenabil), egalitatea cerută se obţine astfel:

$$\begin{aligned} P(X, Y \mid Z) &\stackrel{def.}{=} \frac{P(X, Y, Z)}{P(Z)} \stackrel{not.}{=} \frac{P(Z \cap Y \cap X)}{P(Z)} \\ &= \frac{P(Z)P(Y \mid Z)P(X \mid Y, Z)}{P(Z)} = P(Y \mid Z) \cdot P(X \mid Y, Z) \end{aligned}$$

***Observaţie*: O altă metodă de rezolvare constă în a aplica pentru fiecare membru al egalităţii din enunţ definiţia probabilităţii condiţionale, după simplificări obţinându-se pentru ambii membri aceeaşi valoare:**

$$\begin{aligned} P(X, Y \mid Z) &= \frac{P(X, Y, Z)}{P(Z)} \\ P(X \mid Y, Z) \cdot P(Y \mid Z) &= \frac{P(X, Y, Z)}{P(Y, Z)} \cdot \frac{P(Y, Z)}{P(Z)} = \frac{P(X, Y, Z)}{P(Z)} \end{aligned}$$

15. **(Variabile aleatoare discrete: distribuţii corelate, marginale, condiţionale)**

CMU, 2002 fall, Andrew Moore, final exam, pr. 4.a

Considerăm un set de date definit cu ajutorul a 3 variabile aleatoare cu valori booleene X, Y şi Z. Care dintre seturile de informaţii de mai jos sunt suficiente pentru a specifica distribuţia corelată $P(x, y, z)$?

A.	**B.**	**C.**	**D.**
$P(\neg X \mid Z)$	$P(\neg X \mid \neg Z)$	$P(X \mid Z)$	$P(X \mid Z)$
$P(\neg X \mid \neg Z)$	$P(X \mid \neg Z)$	$P(X \mid \neg Z)$	$P(X \mid \neg Z)$
$P(\neg Y \mid X, Z)$	$P(Y \mid X, Z)$	$P(Y \mid X, Z)$	$P(Y \mid X, Z)$
$P(\neg Y \mid X, \neg Z)$	$P(Y \mid X, \neg Z)$	$P(Y \mid X, \neg Z)$	$P(Y \mid X, \neg Z)$
$P(\neg Y \mid \neg X, Z)$	$P(Y \mid \neg X, Z)$	$P(Y \mid \neg X, Z)$	$P(\neg Y \mid \neg X, \neg Z)$
$P(\neg Y \mid \neg X, \neg Z)$	$P(Y \mid \neg X, \neg Z)$	$P(\neg Y \mid \neg X, \neg Z)$	$P(Y \mid \neg X, \neg Z)$
$P(Z)$	$P(Z)$	$P(\neg Z)$	$P(Z)$

Răspuns:

Pentru a calcula distribuţia corelată a mai multor variabile aleatoare se poate aplica regula de înlănţuire.[11] În cazul nostru putem scrie că:

$$P(X, Y, Z) = P(Z) \cdot P(X \mid Z) \cdot P(Y \mid X, Z)$$

Deoarece variabilele aleatoare X, Y şi Z au valori booleene, pentru a specifica distribuţia corelată $P(X, Y, Z)$ este nevoie să se calculeze valoarea acesteia în fiecare din cele 8 cazuri posibile:

$$\begin{array}{llll} P(X, Y, Z) & P(X, Y, \neg Z) & P(X, \neg Y, Z) & P(X, \neg Y, \neg Z) \\ P(\neg X, Y, Z) & P(\neg X, Y, \neg Z) & P(\neg X, \neg Y, Z) & P(\neg X, \neg Y, \neg Z) \end{array}$$

Cunoaştem de asemenea relaţii de calcul de forma:

$$\begin{aligned} P(\neg X) &= 1 - P(X) \\ P(\neg X \mid Y) &= 1 - P(X \mid Y) \end{aligned}$$

Aşadar, pentru a aplica regula de înlănţuire de mai sus vom avea nevoie de

$P(Z)$ **sau** $P(\neg Z)$;

$P(X \mid Z)$ **sau** $P(\neg X \mid Z)$;
$P(X \mid \neg Z)$ **sau** $P(\neg X \mid \neg Z)$;

$P(Y \mid X, Z)$ **sau** $P(\neg Y \mid X, Z)$;
$P(Y \mid \neg X, Z)$ **sau** $P(\neg Y \mid \neg X, Z)$;
$P(Y \mid X, \neg Z)$ **sau** $P(\neg Y \mid X, \neg Z)$;
$P(Y \mid \neg X, \neg Z)$ **sau** $P(\neg Y \mid \neg X, \neg Z)$.

Cu aceste precizări, putem specifica pentru fiecare dintre seturile de informaţii din enunţ dacă sunt suficiente pentru a calcula distribuţia corelată $P(X, Y, Z)$.

Cazul A. Da. Se observă că putem calcula $P(X \mid Z)$ şi $P(X \mid \neg Z)$ din primele două probabilităţi din enunţ. De asemenea, utilizând următoarele 4 probabilităţi se pot deduce: $P(Y \mid X, Z)$, $P(Y \mid X, \neg Z)$, $P(Y \mid \neg X, Z)$ şi $P(Y \mid \neg X, \neg Z)$. Iar din $P(Z)$ se obţine $P(\neg Z)$. Prin urmare, există toate informaţiile necesare distribuţiei corelate $P(X, Y, Z)$.

Cazul B. Nu, informaţiile din enunţ nu sunt suficiente. Nu putem deduce valoarea pentru $P(X \mid Z)$.

Cazul C. Da, informaţiile din enunţ sunt suficiente. Din $P(X \mid Z)$ se obţine $P(\neg X \mid Z)$, iar din $P(X \mid \neg Z)$ se obţine $P(\neg X \mid \neg Z)$. Din următoarele 4 probabilităti se obţin celelalte 4 necesare pentru $P(Y \mid X, Z)$, şi anume: $P(\neg Y \mid X, Z)$, $P(\neg Y \mid X, \neg Z)$, $P(\neg Y \mid \neg X, Z)$ şi respectiv $P(Y \mid \neg X, \neg Z)$.

Cazul D. Nu, informaţiile din enunţ nu sunt suficiente. Nu putem deduce valoarea pentru $P(Y \mid \neg X, Z)$.

16. **(Variabile aleatoare discrete: independenţă condiţională)**

CMU, 2005 fall, T. Mitchell, A. Moore, midterm, pr. 4

Fie variabilele aleatoare discrete A, B şi C având distribuţia corelată conform tabelului de mai jos.

[11] Pentru variabile aleatoare, regula de înlănţuire

$$P(A_1, A_2, \ldots, A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_n \mid A_1, A_2, \ldots, A_{n-1})$$

se demonstrează imediat pornind de la regula de înlănţuire pentru (probabilităţi de) evenimente aleatoare. A se vedea problema 14.

a. Este variabila A independentă condiţional de B în raport cu variabila C?

b. Dacă aţi răspuns afirmativ la întrebarea a, faceţi o schimbare în primele două linii ale tabelului de mai sus pentru a obţine o distribuţie pentru care răspunsul la aceeaşi întrebare să devină negativ. Invers, dacă aţi răspuns negativ la întrebarea a, faceţi o schimbare în primele două linii ale tabelului încât răspunsul să devină afirmativ.

A	B	C	$P(A,B,C)$
0	0	0	1/8
0	0	1	1/8
0	1	0	1/8
0	1	1	1/8
1	0	0	1/8
1	0	1	1/8
1	1	0	1/8
1	1	1	1/8

Răspuns:

a. Faptul că variabila A este independentă condiţional de B în raport cu variabila C se mai notează prin $A \perp B \mid C$ şi poate fi demonstrat prin una din următoarele două relaţii:

$$P(A=a,B=b \mid C=c) = P(A=a \mid C=c)\cdot P(B=b \mid C=c) \ \textbf{ sau}$$

$$P(A=a \mid B=b,C=c) = P(A=a \mid C=c),\ \textbf{dacă } P(B=b,C=c) \neq 0.$$

pentru orice $a \in Val(A), b \in Val(B), c \in Val(C)$. Deşi în general se foloseşte prima relaţie, pentru acest exerciţiu este mai uşor să utilizăm cea de-a doua relaţie. Conform tabelului dat în enunţ, rezultă imediat că $P(B=b,C=c)\neq 0$ pentru orice $b,c \in \{0,1\}$. Vom demonstra că pentru orice $a,b,c \in \{0,1\}$ este adevărat că $P(A=a \mid B=b,C=c) = P(A=a \mid C=c)$. Cele două probabilităţi condiţionate vor fi calculate folosind datele din tabel:

Cazul $(0,0,0)$: $P(A=0 \mid B=0,C=0) = \dfrac{1\cdot\frac{1}{8}}{2\cdot\frac{1}{8}} = \dfrac{1}{2}$ **şi** $P(A=0 \mid C=0) = \dfrac{2\cdot\frac{1}{8}}{4\cdot\frac{1}{8}} = \dfrac{1}{2}$.

Cazul $(0,0,1)$: $P(A=0 \mid B=0,C=1) = \dfrac{1\cdot\frac{1}{8}}{2\cdot\frac{1}{8}} = \dfrac{1}{2}$ **şi** $P(A=0 \mid C=1) = \dfrac{2\cdot\frac{1}{8}}{4\cdot\frac{1}{8}} = \dfrac{1}{2}$.

Se observă că pentru toate celelalte cazuri se obţin aceleaşi valori, deci

$$P(A=a \mid B=b,C=c) = P(A=a \mid C=c),\ \forall a,b,c \in \{0,1\}.$$

Aşadar, variabila A este independentă condiţional de B în raport cu variabila C.

b. Schimbarea trebuie făcută în aşa fel încât să se păstreze relaţia $\sum P(A,B,C) = 1$. O variantă posibilă este:

A	B	C	$P(A,B,C)$
0	0	0	1/4
0	0	1	0
	...		...

Pentru aceste noi valori se observă că: $P(A=0 \mid B=0,C=0) = \dfrac{1\cdot 1/4}{1\cdot 1/4+1\cdot 1/8} = \dfrac{2}{3}$ **şi** $P(A=0 \mid C=0) = \dfrac{1\cdot 1/4+1\cdot 1/8}{1\cdot 1/4+3\cdot 1/8} = \dfrac{3}{5}$. **Este suficient un singur caz în care probabilităţile respective nu sunt egale, prin urmare variabila A nu este independentă condiţional de variabila B în raport cu a treia variabilă, C.**

17. **(Variabile aleatoare continue: funcţia densitate de probabilitate)**

CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, HW1, pr. 1.5

Fie variabila aleatoare continuă X a cărei funcţie densitate de probabilitate (în limba engleză: "probability density functions"; scris, sub formă abreviată, pdf) este:

$$p(x) = \begin{cases} cx^2 & \textbf{pentru } 1 \le x \le 2 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

a. Ţinând cont de proprietăţile funcţiei densitate de probabilitate (a se vedea notiţele de la curs), cât trebuie să fie valoarea constantei c?

b. Desenaţi graficul funcţiei de mai sus.

c. Calculaţi $P(X > 3/2)$.

<u>Răspuns:</u>

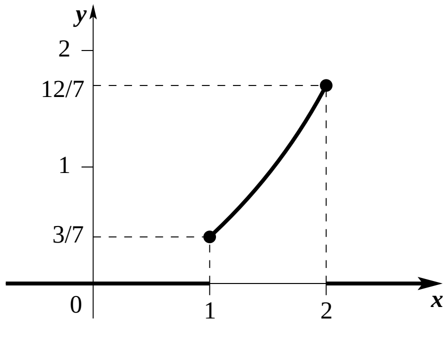
a. Faptul că $p(x)$ este funcţie densitate de probabilitate pentru variabila aleatoare continuă X înseamnă că $p(x) \ge 0, \forall x$ şi că $\int_{-\infty}^{+\infty} p(x)dx = 1$. Pentru a afla constanta c vom calcula valoarea integralei:

$$\begin{aligned} \int_{-\infty}^{+\infty} p(x)dx &= \underbrace{\int_{-\infty}^{1} p(x)dx}_{=0} + \int_{1}^{2} p(x)dx + \underbrace{\int_{2}^{+\infty} p(x)dx}_{=0} \\ &= \int_1^2 cx^2 dx = c \cdot \int_1^2 x^2 dx = c \cdot \left.\frac{x^3}{3}\right|_1^2 = c\left(\frac{2^3}{3} - \frac{1^3}{3}\right) = c \cdot \frac{7}{3} \end{aligned}$$

Prin urmare, $c \cdot \dfrac{7}{3} = 1 \Rightarrow c = \dfrac{3}{7}$.

b. Trebuie să reprezentăm grafic funcţia:

$$p(x) = \begin{cases} \dfrac{3}{7}x^2 & \textbf{pentru } 1 \le x \le 2 \\ 0 & \textbf{în caz contrar}. \end{cases}$$



În intervalul $[1, 2]$, această funcţie este un fragment din parabola corespunzătoare funcţiei de gradul al doilea: $\dfrac{3}{7}x^2$, parabolă care are vârful în punctul $(0, 0)$. Putem calcula $p(1) = 3/7 \approx 0.42$ şi $p(2) = 12/7 \approx 1.71$.

c. Valoarea probabilităţii cerute se poate calcula astfel:

$$\begin{aligned}
P(X > 3/2) &\overset{not.}{=} P(\{\omega \mid X(\omega) > \frac{3}{2}\}) \overset{def.}{=} \int_{X(\omega)=x>3/2} p(x)dx = \int_{3/2}^{+\infty} p(x)dx \\
&= \int_{3/2}^{2} p(x)dx + \underbrace{\int_{2}^{+\infty} p(x)dx}_{=0} \\
&= \int_{3/2}^{2} \frac{3}{7}x^2 dx = \frac{3}{7} \cdot \int_{3/2}^{2} x^2 dx = \frac{3}{7} \cdot \left.\frac{x^3}{3}\right|_{3/2}^{2} = \frac{1}{7} \cdot x^3\Big|_{3/2}^{2} \\
&= \frac{1}{7} \cdot \left(8 - \frac{27}{8}\right) = \frac{1}{7} \cdot \frac{64-27}{8} = \frac{37}{56}
\end{aligned}$$

O altă variantă de rezolvare este bazată pe folosirea *funcţiei cumulative de distribuţie*, care, după cum ştim, se defineşte prin relaţia $F(x) \overset{def.}{=} P(X \leq x)$, pentru orice $x \in \mathbb{R}$:

$$\begin{aligned}
P(X > 3/2) &= 1 - P(X \leq 3/2) = 1 - F\left(\frac{3}{2}\right) = 1 - \int_{-\infty}^{3/2} p(x)dx \\
&= 1 - \left(\int_{-\infty}^{1} \underbrace{p(x)}_{0} dx + \int_{1}^{3/2} p(x)dx\right) = 1 - \left(0 + \int_{1}^{3/2} p(x)dx\right) \\
&= 1 - \int_{1}^{3/2} p(x)dx = 1 - \int_{1}^{3/2} \frac{3}{7}x^2 dx = 1 - \frac{1}{7} \cdot x^3\Big|_{1}^{3/2} \\
&= 1 - \frac{1}{7} \cdot \left(\frac{27}{8} - 1\right) = 1 - \frac{1}{7} \cdot \frac{19}{8} \\
&= 1 - \frac{19}{56} = \frac{37}{56}
\end{aligned}$$

18. (Variabile aleatoare continue: funcţia densitate de probabilitate)

CMU, 2008 spring, Eric Xing, HW1, pr. 1.1.b

Fie funcţia

$$p(x) = \begin{cases} cx^{-d} & \textbf{pentru } x > 1 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

Care sunt valorile posibile pentru c şi d în aşa fel ca p să poată reprezenta o funcţie densitate de probabilitate?

Răspuns:

O funcţie p poate reprezenta o funcţie densitate de probabilitate dacă $p(x) \geq 0$ pentru $\forall x$, şi $\int_{-\infty}^{\infty} p(x)dx = 1$. Vom folosi aceste două condiţii pentru a calcula valorile posibile pentru c şi d.

Prima condiţie $p(x) \geq 0$ implică faptul că $c \geq 0$, asupra lui d neimpunând nicio restricţie. Mai mult, ţinând cont de forma lui p şi de cea de-a doua condiţie, vom avea chiar $c > 0$, fiindcă $c = 0$ ar implica $\int_{-\infty}^{\infty} p(x)dx = 0 \neq 1$.

Pentru a aplica cea de-a doua condiţie, calculăm integrala:

$$\int_{-\infty}^{\infty} p(x)dx = \int_{1}^{\infty} cx^{-d}dx = c\int_{1}^{\infty} x^{-d}dx$$

Vom avea de tratat două cazuri:

$$\textbf{Cazul 1:} \quad d = 1 \quad \Rightarrow c\int_{1}^{\infty} x^{-d}dx = c\int_{1}^{\infty} \frac{1}{x}dx = c \cdot \ln x|_{1}^{\infty} = \infty$$

$$\textbf{Cazul 2:} \quad d \neq 1 \quad \Rightarrow c\int_{1}^{\infty} x^{-d}dx = c \cdot \left.\frac{x^{-d+1}}{-d+1}\right|_{1}^{\infty} = \frac{c}{1-d} \cdot x^{1-d}|_{1}^{\infty}$$

Subcazul $d > 1$**:** $x^{1-d}|_{1}^{\infty} = 0 - 1 = -1$**. Deci în acest caz** $c\int_{1}^{\infty} x^{-d}dx = \dfrac{c}{d-1}$**.**

Subcazul $d < 1$**:** $x^{1-d}|_{1}^{\infty} = +\infty$**. Deci în acest caz** $c\int_{1}^{\infty} x^{-d}dx = +\infty$**.**

Prin urmare, $\int_{-\infty}^{\infty} p(x)dx = 1 \Rightarrow d > 1$ **şi** $\dfrac{c}{d-1} = 1$**.**

În concluzie, p **poate reprezenta o funcţie densitate de probabilitate dacă**

$$c > 0,\ d > 1 \textbf{ şi } c = d - 1.$$

Referitor la aceste trei restricţii, se observă că ultimele două o implică pe cea dintâi, deci o fac superfluă.

19. **(Variabile aleatoare uniforme continue, independente; funcţii densitate de probabilitate)**

CMU, 2008 spring, Eric Xing, HW1, pr. 1.5

O persoană pleacă la servici între orele 8:00 şi 8:30, iar timpul necesar deplasării este între 40 şi 50 de minute. Considerăm X **variabila aleatoare care reprezintă timpul de plecare exprimat în minute scurse după ora 8:00 şi** Y **variabila aleatoare care reprezintă durata deplasării. Presupunând că aceste două variabile sunt independente şi uniform distribuite, indicaţi:**

a. funcţiile densitate de probabilitate $p(x) \stackrel{not.}{=} P_X(X = x)$, $p(y) \stackrel{not.}{=} P_X(Y = y)$ **şi** $p(x, y) \stackrel{not.}{=} P_{\boldsymbol{XY}}(X = x, Y = y)$**.**

b. probabilitatea ca persoana respectivă să ajungă la servici înainte de ora 9.

<u>Răspuns:</u>

a. Conform enunţului $Val(X) = [0, 30]$ **şi** $Val(Y) = [40, 50]$**. Pentru a determina funcţia densitate de probabilitate pentru** X **şi respectiv** Y **vom impune restricţia ca integrala valorilor pe domeniul de definiţie să fie** 1**:**

$$\int_{-\infty}^{\infty} p(x)dx = 1 \Leftrightarrow \int_{0}^{30} p(x)dx = 1 \stackrel{p-unif.}{\Leftrightarrow} p(x) = \begin{cases} \dfrac{1}{30} & \textbf{pentru } 0 \leq x \leq 30 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

$$\int_{-\infty}^{\infty} p(y)dy = 1 \Leftrightarrow \int_{40}^{50} p(y)dy = 1 \stackrel{p-unif.}{\Leftrightarrow} p(y) = \begin{cases} \dfrac{1}{10} & \textbf{pentru } 40 \leq y \leq 50 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

Deoarece variabilele X şi Y sunt independente, funcţia densitate de probabilitate corelată este:

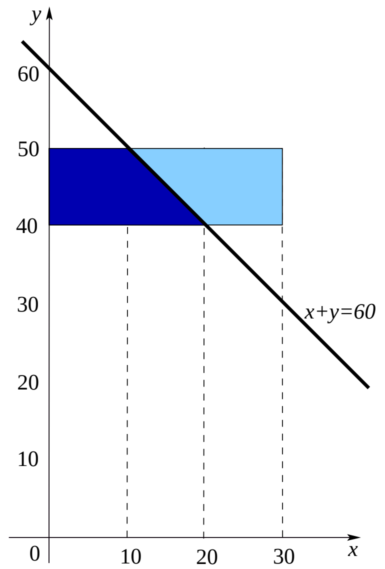
$$p(x,y) = p(x)\cdot p(y) = \begin{cases} \dfrac{1}{300} & \textbf{pentru } 0 \le x \le 30 \textbf{ şi } 40 \le y \le 50 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

b. Probabilitatea ca persoana respectivă să ajungă la servici înainte de ora 9 este $P(X+Y \le 60)$.

***Grafic*, aşa cum se observă din figura alăturată, această probabilitate este foarte simplu de găsit: $p(x+y \le 60) = \frac{1}{2}$. Într-adevăr, ştim că aria întregii zone dreptunghiulare pe care $p(x,y) \neq 0$ este**

$$\int_{x=0}^{30}\int_{y=40}^{50} p(x,y)dydx = 1.$$

Probabilitatea urmărită este aria zonei din acest dreptunghi care se găseşte „sub" dreapta de ecuaţie $x+y=60$. Această zonă corespunde perechilor de valori (x,y) care satisfac condiţia ca persoana sa ajungă la servici înainte de ora 9.



***Alternativ*, această arie se poate obţine prin calcul direct:**

$$\begin{aligned} P(X+Y\le 60) &= \int_{x=0}^{10}\int_{y=40}^{50}\frac{1}{300}dydx + \int_{x=10}^{20}\int_{y=40}^{60-x}\frac{1}{300}dydx \\ &= \int_{x=0}^{10}\frac{1}{300}\cdot y|_{40}^{50}dx + \int_{x=10}^{20}\frac{1}{300}\cdot y|_{40}^{60-x}dx \\ &= \int_{x=0}^{10}\frac{1}{300}\cdot 10dx + \int_{x=10}^{20}\frac{1}{300}\cdot(20-x)dx \\ &= \frac{1}{30}\cdot x|_0^{10} + \frac{1}{300}\cdot\left(20x-\frac{1}{2}x^2\right)\Big|_{10}^{20} = \frac{10}{30}+\frac{50}{300} = \frac{1}{2}. \end{aligned}$$

20. **(Variabile aleatoare uniforme continue: p.d.f. condiţională; independenţă)**

CMU, 2003 fall, T. Mitchell, A. Moore, midterm, pr. 2

Fie X şi Y variabile aleatoare continue având funcţia densitate de probabilitate corelată definită astfel:

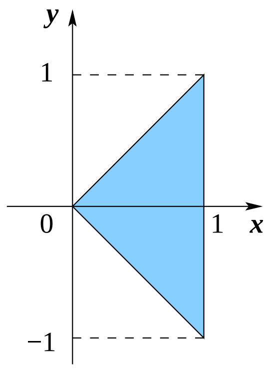
$$p(x,y) = \begin{cases} 1 & \textbf{pentru } 0 < x < 1, \ |\, y \,| < x \\ 0 & \textbf{în caz contrar}. \end{cases}$$

a. Cât este $p(y \mid x = 0.5)$?

b. Este variabila X independentă de variabila Y?

Răspuns:

Putem reprezenta grafic funcţia densitate de probabilitate corelată a variabilelor aleatoare X şi Y. În figura alăturată porţiunea triunghiulară reprezintă domeniul pentru care valoarea funcţiei este 1, în rest valoarea acesteia fiind 0.



a. Conform definiţiei p.d.f. condiţionale, avem

$$p(y \mid x = 0.5) = \frac{p(x{=}0.5, y)}{p(x{=}0.5)}.$$

Conform definiţiei p.d.f. marginale,

$$p(x = 0.5) = \int_{-\infty}^{+\infty} p(x = 0.5, y)dy = \int_{-0.5}^{+0.5} 1dy = 1$$

Deci $p(y \mid x = 0.5) = p(y, x = 0.5)$. Aşadar, pentru a calcula valoarea cerută vom înlocui în definiţia funcţiei densitate de probabilitate corelată valoarea cunoscută $x = 0.5$.

$$p(y \mid x = 0.5) = p(x = 0.5, y) = \begin{cases} 1 & \textbf{pentru } 0 < 0.5 < 1, \mid y \mid < 0.5 \\ 0 & \textbf{în caz contrar.} \end{cases}$$
$$= \begin{cases} 1 & \textbf{pentru } -0.5 < y < 0.5 \\ 0 & \textbf{în caz contrar.} \end{cases}$$

b. Variabilele X şi Y nu sunt independente. Intuitiv, se observă că în definiţia funcţiei densitate de probabilitate corelată există o dependenţă între valorile lor, şi anume: $\mid y \mid < x$.

Pentru a demonstra riguros faptul că variabilele X şi Y nu sunt independente trebuie să arătăm că relaţia $p(x, y) = p(x) \cdot p(y)$ nu este adevărată pentru $\forall x, y$ din domeniul de definiţie.

Se observă că $p(X = \frac{1}{4}, Y = 0) = 1$ (conform definiţiei), dar probabilităţile marginale corespunzătoare sunt:

$$p_X(X = \frac{1}{4}) = \int_{-\frac{1}{4}}^{\frac{1}{4}} 1dy = y|_{-\frac{1}{4}}^{\frac{1}{4}} = \frac{1}{2}$$

şi

$$p_Y(Y = 0) = \int_0^1 1dx = 1.$$

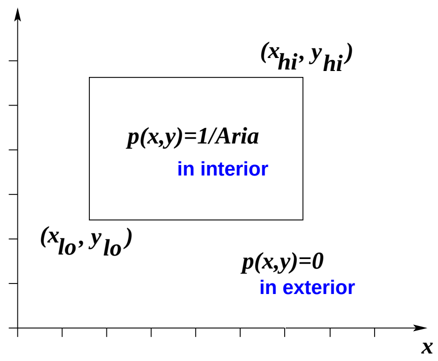
Aşadar,

$$p(X = \frac{1}{4}, Y = 0) = 1 \neq \frac{1}{2} = p_X(X = \frac{1}{4}) \cdot p_Y(Y = 0).$$

21. **(Variabile aleatoare uniforme continue: funcţia densitate de probabilitate; distribuţii corelate, marginale, condiţionale; formula lui Bayes)**

CMU, 2002 fall, Andrew Moore, midterm, pr. 2

Figura alăturată ilustrează o clasă simplă de funcţii densitate de probabilitate (pdf) definite peste perechi de variabile continue reale x, y.
Numim această clasă de funcţii *Rectangle-PDF*. O funcţie din această clasă este desemnată în mod generic prin $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$.
Aşadar, *parametrii* clasei *Rectangle-PDF* sunt cele 4 coordonate: x_{lo}, y_{lo}, x_{hi} şi y_{hi}.



Definiţia funcţiei densitate de probabilitate $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$ **este:**

$$p(x,y) = \begin{cases} \dfrac{1}{(x_{hi} - x_{lo})(y_{hi} - y_{lo})} & \textbf{dacă } x_{lo} \leq x \leq x_{hi} \textbf{ şi } y_{lo} \leq y \leq y_{hi} \\ 0 & \textbf{în caz contrar.} \end{cases}$$

***Observaţie*: Se poate arăta imediat că** $\int_x \int_y p(x, y) dx\, dy = 1$.

a. Pentru $Rectangle(0, 0, 1/2, 2)$, **calculaţi:** $p(x = 1/4, y = 1/4)$, $p(y = 1/4)$, $p(x = 1/4)$, $p(x = 1/4 \mid y = 1/4)$.

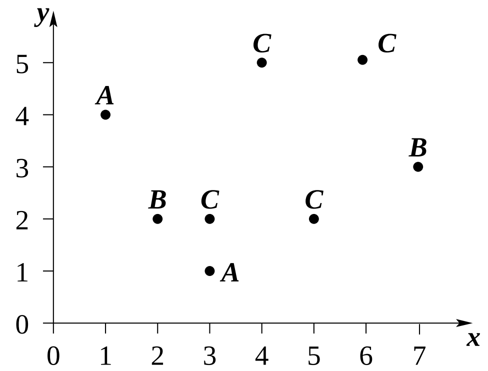
b. Fie D o mulţime de R puncte, $(x_1, y_1), (x_2, y_2), \ldots, (x_R, y_R)$, **selectate din** $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$ **în mod independent unele de altele.**[12] **Pentru simplitate, vom nota** $\theta = (x_{lo}, y_{lo}, x_{hi}, y_{hi})$. **Prin definiţie, verosimilitatea datelor D este** $P(D \mid \theta)$.

Dacă vrem să găsim $\theta^{MLE} \stackrel{not.}{=} (x_{lo}^{MLE}, y_{lo}^{MLE}, x_{hi}^{MLE}, y_{hi}^{MLE})$, **valorile parametrilor care maximizează *verosimilitatea* datelor D, atunci este *evident* că vom lua**

$$x_{lo}^{MLE} = \min_k\{x_k\},\ y_{lo}^{MLE} = \min_k\{y_k\},\ x_{hi}^{MLE} = \max_k\{x_k\},\ y_{hi}^{MLE} = \max_k\{y_k\}.$$

În continuare, ca date concrete, vom folosi punctele $(x_i, y_i), i = 1, ..., 6$, **clasificate în trei „concepte“ A, B, C care fac parte din clasa *Rectangle-PDF*:**

x	y	*Concept*
1	4	A
3	1	A
2	2	B
7	3	B
3	2	C
4	5	C
5	2	C
6	5	C



[12] În mod riguros, în acest context conceptul $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$ desemnează mulţimea punctelor din plan pentru care funcţia omonimă $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$ care a fost definită mai sus ia valori nenule. Această mulţime este exact dreptunghiul reprezentat în imaginea de mai sus şi punctele din interiorul lui. Altfel spus, cu suprapunere de notaţie, funcţia $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$ desemnează conceptul geometric $Rectangle(x_{lo}, y_{lo}, x_{hi}, y_{hi})$.

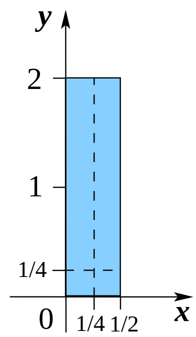
Vom avea, de *exemplu*, $p(2.5, 2.5 \mid A) \stackrel{not.}{=} p_{X,Y}(2.5, 2.5 \mid A) = 1/6$**, fiindcă** A **este estimat în sensul verosimilităţii maxime (MLE) ca fiind** *Rectangle*$(1, 1, 3, 4)$**, deci cele două laturi ale dreptunghiului** A **sunt de mărime 2 şi respectiv 3.**

Folosind formula lui Bayes, calculaţi:

– $P(Concept = A \mid x = 1.5, y = 3)$

– $P(Concept = A \mid x = 2.5, y = 2.5)$

– $P(Concept = A \mid y = 5)$.



Răspuns:

a. În figura alăturată avem reprezentarea grafică a funcţiei *Rectangle*$(0, 0, 1/2, 2)$**. Cum punctul** $(1/4, 1/4)$ **se găseşte în interiorul dreptunghiului** ($0 < \frac{1}{4} < \frac{1}{2}$ **şi** $0 < \frac{1}{4} < 2$)**, rezultă că valoarea p.d.f. corelate cerute este:**

$$p\left(x = \frac{1}{4}, y = \frac{1}{4}\right) = \frac{1}{\left(\frac{1}{2} - 0\right)(2 - 0)} = \frac{1}{\frac{1}{2} \cdot 2} = 1$$

Valorile p.d.f. marginale cerute sunt:

$$\begin{aligned}
p\left(y = \frac{1}{4}\right) &= \int_{-\infty}^{\infty} p(x, y = \frac{1}{4})\, dx \\
&= \int_{-\infty}^{0} p(x, y = \frac{1}{4})\, dx + \int_{0}^{1/2} p(x, y = \frac{1}{4})\, dx + \int_{1/2}^{\infty} p(x, y = \frac{1}{4})\, dx \\
&= \int_{-\infty}^{0} 0\, dx + \int_{0}^{1/2} \frac{1}{\left(\frac{1}{2} - 0\right)(2 - 0)}\, dx + \int_{1/2}^{\infty} 0\, dx \\
&= 0 + \int_{0}^{1/2} 1\, dx + 0 = \frac{1}{2} - 0 = \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
p\left(x = \frac{1}{4}\right) &= \int_{-\infty}^{\infty} p(x = \frac{1}{4}, y)\, dy \\
&= \int_{-\infty}^{0} p(x = \frac{1}{4}, y)\, dy + \int_{0}^{2} p(x = \frac{1}{4}, y)\, dy + \int_{2}^{\infty} p(x = \frac{1}{4}, y)\, dy \\
&= \int_{-\infty}^{0} 0\, dy + \int_{0}^{2} \frac{1}{\left(\frac{1}{2} - 0\right)(2 - 0)}\, dy + \int_{2}^{\infty} 0\, dy = 0 + \int_{0}^{2} 1\, dy + 0 \\
&= 2 - 0 = 2
\end{aligned}$$

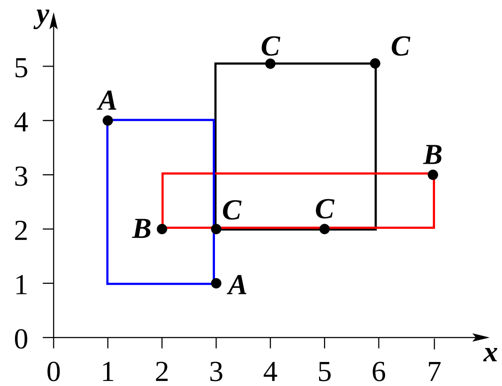
În sfârşit, valoarea p.d.f. condiţionale cerute este:

$$p\left(x = \frac{1}{4} \mid y = \frac{1}{4}\right) = \frac{p(x = \frac{1}{4}, y = \frac{1}{4})}{p(y = \frac{1}{4})} = \frac{1}{\frac{1}{2}} = 2$$

***Observaţie*: Aceeaşi valoare pentru $p(x = \frac{1}{4} \mid y = \frac{1}{4})$ se obţine şi dacă observăm că variabilele x şi y sunt independente, deci $p(x = \frac{1}{4} \mid y = \frac{1}{4}) = p(x = \frac{1}{4}) = 2$.**

b. Ţinând cont de date (vezi tabelul din enunţ), cele trei concepte A, B, C vor fi estimate în sensul verosimilităţii maxime ca:

$$Rectangle_A(1, 1, 3, 4)$$
$$Rectangle_B(2, 2, 7, 3)$$
$$Rectangle_C(3, 2, 6, 5)$$



Aşadar, vom avea:

$$p(x, y \mid A) = \begin{cases} \frac{1}{(3-1)(4-1)} = \frac{1}{6}, & \textbf{dacă } 1 \leq x \leq 3 \textbf{ şi } 1 \leq y \leq 4 \\ 0, & \textbf{altfel;} \end{cases}$$

$$p(x, y \mid B) = \begin{cases} \frac{1}{(7-2)(3-2)} = \frac{1}{5}, & \textbf{dacă } 2 \leq x \leq 7 \textbf{ şi } 2 \leq y \leq 3 \\ 0, & \textbf{altfel;} \end{cases}$$

$$p(x, y \mid C) = \begin{cases} \frac{1}{(6-3)(5-2)} = \frac{1}{9}, & \textbf{dacă } 3 \leq x \leq 6 \textbf{ şi } 2 \leq y \leq 5 \\ 0, & \textbf{altfel.} \end{cases}$$

Acum putem calcula cele trei probabilităţi condiţionale cerute, folosind formula lui Bayes. Probabilităţile a priori $P(A) = \frac{2}{8}$, $P(B) = \frac{2}{8}$ şi $P(C) = \frac{4}{8}$ care intervin în aplicarea acestei formule au fost estimate din setul de date D din

enunţ, în sensul verosimilităţii maxime. (A se vedea *Observaţia* de mai jos.)

$$
\begin{aligned}
P & \quad (Concept = A \mid x = 1.5, y = 3) \\
& = \frac{P(x = 1.5, y = 3 \mid A) \cdot P(A)}{P(1.5, 3 \mid A) \cdot P(A) + P(1.5, 3 \mid B) \cdot P(B) + P(1.5, 3 \mid C) \cdot P(C)} \\
& = \frac{\frac{1}{6} \cdot \frac{2}{8}}{\frac{1}{6} \cdot \frac{2}{8} + 0 \cdot \frac{2}{8} + 0 \cdot \frac{4}{8}} = 1 \\
P & \quad (Concept = A \mid x = 2.5, y = 2.5) \\
& = \frac{P(x = 2.5, y = 2.5 \mid A) \cdot P(A)}{P(2.5, 2.5 \mid A) \cdot P(A) + P(2.5, 2.5 \mid B) \cdot P(B) + P(2.5, 2.5 \mid C) \cdot P(C)} \\
& = \frac{\frac{1}{6} \cdot \frac{2}{8}}{\frac{1}{6} \cdot \frac{2}{8} + \frac{1}{5} \cdot \frac{2}{8} + 0 \cdot \frac{4}{8}} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{5}} = \frac{5}{5+6} = \frac{5}{11} \\
P & \quad (Concept = A \mid y = 5) = \frac{P(y = 5 \mid A) \cdot P(A)}{P(y = 5)} \\
& = \frac{[\int_{-\infty}^{\infty} P(x = t, y = 5 \mid A) dt] \cdot P(A)}{P(y = 5)} = \frac{\int_{-\infty}^{\infty} 0 dt \cdot P(A)}{P(y = 5)} = \frac{0 \cdot P(A)}{P(y = 5)} = 0
\end{aligned}
$$

***Observaţie*: Aşa cum am precizat deja, valorile probabilităţilor a priori $P(A) = \frac{2}{8}$, $P(B) = \frac{2}{8}$ şi $P(C) = \frac{4}{8}$ au fost estimate (în sensul verosimilităţii maxime) din tabelul de date din enunţ. Pentru a înţelege mai bine de ce s-a procedat aşa, ar fi fost mai explicit dacă făceam condiţionarea (şi) în funcţie de D, setul de date din enunţ:**

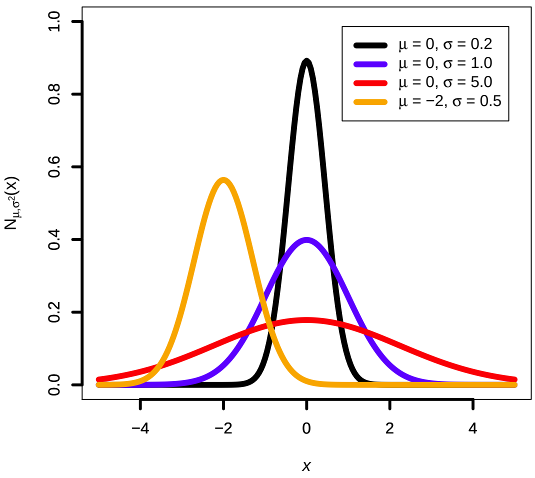
$$
P(Concept = A \mid x = 1.5, y = 3, D) = \frac{P(x = 1.5, y = 3 \mid A, D) \cdot P(A \mid D)}{P(1.5, 3 \mid A, D) \cdot P(A \mid D) + P(1.5, 3 \mid B, D) \cdot P(B \mid D) + P(1.5, 3 \mid C, D) \cdot P(C \mid D)}
$$

Totuşi, din motive legate de simplitate, am optat pentru notaţia folosită mai sus.

22. **(Distribuţia gaussiană uni-variată: calculul mediei şi al varianţei)**

CMU, 2010 spring, T. Mitchel, E. Xing, A. Singh, HW1, pr. 1.3.2

Distribuţia gaussiană: p.d.f.



Considerăm o variabilă aleatoare X care are distribuţia normală (gaussiană):

$$N(X=x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

unde σ poate fi orice număr real pozitiv, iar μ orice număr real.
Arătaţi că:

a. $E[X] = \mu$

b. $Var[X] = \sigma^2$

Răspuns:

a. Vom calcula media variabilei aleatoare X folosind formula de definiţie:

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Facem schimbarea de variabilă $v = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma v + \mu$ **şi** $dx = \sigma dv$ **şi obţinem:**

$$\begin{aligned}
E[X] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (\sigma v+\mu) e^{-\frac{v^2}{2}}(\sigma dv) = \frac{\sigma}{\sqrt{2\pi}\sigma}\left(\sigma\int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}} dv + \mu\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(-\sigma\int_{-\infty}^{\infty}(-v)e^{-\frac{v^2}{2}} dv + \mu\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(\underbrace{-\sigma\, e^{-\frac{v^2}{2}}\Big|_{-\infty}^{\infty}}_{=0} + \mu\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right) = \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv
\end{aligned}$$

Pentru a calcula integrala din partea dreaptă vom folosi următorul artificiu:

$$\begin{aligned}
\left(\int_{v=-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right)^2 &= \left(\int_{x=-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right)\cdot\left(\int_{y=-\infty}^{\infty} e^{-\frac{y^2}{2}} dy\right) \\
&= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dy\, dx = \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dy\, dx
\end{aligned}$$

Schimbând variabilele x, y în coordonatele polare r, θ obţinem:

$$\left(\int_{v=-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right)^2 = \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}} (r\,dr\,d\theta) = \int_{r=0}^{\infty} re^{-\frac{r^2}{2}} \left(\int_{\theta=0}^{2\pi} d\theta\right) dr$$

$$= \int_{r=0}^{\infty} re^{-\frac{r^2}{2}} \theta|_0^{2\pi} dr = 2\pi \int_{r=0}^{\infty} re^{-\frac{r^2}{2}} dr = 2\pi \left(-e^{-\frac{r^2}{2}}\right)\Bigg|_0^{\infty} = 2\pi(0-(-1)) = 2\pi$$

***Notă*: Schimbarea în coordonate polare a fost realizată utilizând formulele $x = r\cos\theta$ şi $y = r\sin\theta$, cu $r \geq 0$ şi $r \in [0, 2\pi)$. Aşadar, $x^2 + y^2 = r^2$, iar matricea jacobiană corespunzătoare acestei transformări de coordonate este:**

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r \geq 0,$$

de unde a rezultat $dx\,dy = r\,dr\,d\theta$.

Revenind la calculul mediei variabilei aleatoare $X \sim N(\mu, \sigma)$, obţinem:

$$E[X] = \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv = \frac{\mu}{\sqrt{2\pi}} \sqrt{2\pi} = \mu.$$

b. Trebuie să arătăm că $Var[X] = \sigma^2$. Vom calcula varianţa lui X utilizând formula $Var[X] = E[X^2] - (E[X])^2$. Întrucât am calculat $E[X]$, trebuie să mai calculăm valoarea mediei lui X^2.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 p(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Făcând aceeaşi schimbare de variabilă $v = \dfrac{x-\mu}{\sigma} \Rightarrow x = \sigma v + \mu$, cu $dx = \sigma dv$, obţinem:

$$\begin{aligned} E[X^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\sigma v + \mu)^2 e^{-\frac{v^2}{2}} (\sigma dv) = \frac{\sigma}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\sigma^2 v^2 + 2\sigma\mu v + \mu^2)\, e^{-\frac{v^2}{2}} dv \\ &= \frac{1}{\sqrt{2\pi}} \left(\sigma^2 \int_{-\infty}^{\infty} v^2\, e^{-\frac{v^2}{2}} dv + 2\sigma\mu \int_{-\infty}^{\infty} v\, e^{-\frac{v^2}{2}} dv + \mu^2 \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv\right) \end{aligned}$$

Ştim că $\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv = \sqrt{2\pi}$ (de la punctul a) şi calculăm uşor $\int_{-\infty}^{\infty} ve^{-\frac{v^2}{2}} dv = -e^{-\frac{v^2}{2}}\Bigg|_{-\infty}^{\infty} = 0$.

Rămâne de calculat prima integrală. Pentru aceasta vom utiliza metoda de integrare prin părţi:

$$\begin{aligned}\int_{-\infty}^{\infty} v^2 e^{-\frac{v^2}{2}} dv &= \int_{-\infty}^{\infty} (-v) \left(-v e^{-\frac{v^2}{2}} \right) dv = \int_{-\infty}^{\infty} (-v) \left(e^{-\frac{v^2}{2}} \right)' dv \\ &= (-v)\, e^{-\frac{v^2}{2}} \Bigg|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1) e^{-\frac{v^2}{2}} dv = 0 + \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv = \sqrt{2\pi}.\end{aligned}$$

Pentru a calcula valoarea integralei definite am folosit următoarea proprietate:

$$\lim_{v\to\infty} v\, e^{-\frac{v^2}{2}} = \lim_{v\to\infty} \frac{v}{e^{\frac{v^2}{2}}} \stackrel{l'Hôpital}{=} \frac{1}{v\, e^{\frac{v^2}{2}}} = 0 = \lim_{v\to-\infty} v\, e^{-\frac{v^2}{2}}.$$

Aşadar,

$$E[X^2] = \frac{1}{\sqrt{2\pi}} \left(\sigma^2 \sqrt{2\pi} + 2\sigma\mu \cdot 0 + \mu^2 \sqrt{2\pi} \right) = \sigma^2 + \mu^2.$$

Deci varianţa variabilei aleatoare $X \sim N(\mu, \sigma)$ este:

$$Var[X] = E[X^2] - (E[X])^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2.$$

23. **(Distribuţia normală: reducerea cazului nestandard la cazul standard)**

CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, HW1, pr. 3

Considerăm variabila aleatoare X cu distribuţia normală de medie $\mu = 1$ şi varianţă $\sigma^2 = 4$. Calculaţi următoarele probabilităţi:

a. $P(X \leq 3)$.

b. $P(|X| \leq 2)$.

***Indicaţie*: Pentru a calcula aceste probabilităţi, soluţia este să transformaţi variabila aleatoare X în distribuţia normală standard Z după formula: $Z = \dfrac{X-\mu}{\sigma}$. Pentru distribuţia probabilistă normală standard ($\mu = 0$ şi $\sigma = 1$), valorile funcţiei de distribuţie cumulativă (engl., cumulative distribution function), notată Φ şi definită prin relaţia $\Phi(x) \stackrel{not.}{=} P(Z \leq x)$, se consideră că sunt deja calculate. Puteţi folosi următorul tabel de valori:**

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
−3.4	0.0003	−1.7	0.0446	0.0	0.5000	1.7	0.9554
−3.3	0.0005	−1.6	0.0548	0.1	0.5398	1.8	0.9641
−3.2	0.0007	−1.5	0.0668	0.2	0.5793	1.9	0.9713
−3.1	0.0010	−1.4	0.0808	0.3	0.6179	2.0	0.9772
−3.0	0.0013	−1.3	0.0968	0.4	0.6554	2.1	0.9821
−2.9	0.0019	−1.2	0.1151	0.5	0.6915	2.2	0.9861
−2.8	0.0026	−1.1	0.1357	0.6	0.7257	2.3	0.9893
−2.7	0.0035	−1.0	0.1587	0.7	0.7580	2.4	0.9918
−2.6	0.0062	−0.9	0.1841	0.8	0.7881	2.5	0.9938
−2.5	0.0062	−0.8	0.2119	0.9	0.8159	2.6	0.9953
−2.4	0.0082	−0.7	0.2420	1.0	0.8413	2.7	0.9965
−2.3	0.0107	−0.6	0.2743	1.1	0.8643	2.8	0.9974
−2.2	0.0139	−0.5	0.3085	1.2	0.8849	2.9	0.9981
−2.1	0.0179	−0.4	0.3446	1.3	0.9032	3.0	0.9987
−2.0	0.0228	−0.3	0.3821	1.4	0.9192	3.1	0.9990
−1.9	0.0287	−0.2	0.4207	1.5	0.9332	3.2	0.9993
−1.8	0.0359	−0.1	0.4602	1.6	0.9452	3.3	0.9995

Răspuns:

a. $P(X \leq 3) = P\left(\frac{X-\mu}{\sigma} \leq \frac{3-\mu}{\sigma}\right) = P\left(Z \leq \frac{3-1}{\sqrt{4}}\right) = P(Z \leq 1) = \Phi(1) = 0.8413$.

b. Vom proceda analog, descompunând probabilitatea $P(|X| \leq 2) = P(-2 \leq X \leq 2)$ într-o diferenţă de două probabilităţi:

$$\begin{aligned}
P(|X| \leq 2) &= P(-2 \leq X \leq 2) = P(X \leq 2) - P(X \leq -2) \\
&= P(X-1 \leq 2-1) - P(X-1 \leq -2-1) \\
&= P\left(\frac{X-1}{\sqrt{4}} \leq \frac{2-1}{\sqrt{4}}\right) - P\left(\frac{X-1}{\sqrt{4}} \leq \frac{-2-1}{\sqrt{4}}\right) \\
&= P\left(Z \leq \frac{2-1}{2}\right) - P\left(Z \leq \frac{-2-1}{2}\right) = \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{3}{2}\right) \\
&= 0.6915 - 0.0668 = 0.6247
\end{aligned}$$

24. **(O proprietate: matricea de covarianţă a oricărui vector de variabile aleatoare este simetrică şi pozitiv semi-definită)**

■ □ ***prelucrare de Liviu Ciortuz după “The Multivariate Gaussian Distribution”, Chuong B. Do, 2008***

Fie variabilele aleatoare $X_1, \ldots, X_n$, cu $X_i : \Omega \to \mathbb{R}$ pentru $i = 1, \ldots, n$. *Matricea de covarianţă* a *vectorului de variabile aleatoare* $X = (X_1, \ldots, X_n)$ este o matrice pătratică de dimensiune $n \times n$, ale cărei elemente se definesc astfel: $[Cov(X)]_{ij} \stackrel{def.}{=} Cov(X_i, X_j)$, pentru orice $i, j \in \{1, \ldots, n\}$.

Arătaţi că $\Sigma \stackrel{not.}{=} Cov(X)$ este matrice simetrică şi pozitiv semi-definită, cea de-a doua proprietate însemnând că pentru orice vector $z \in \mathbb{R}^n$ are loc inegalitatea

$z^\top \Sigma z \geq 0$. **(Vectorii $z \in \mathbb{R}^n$ sunt consideraţi vectori-coloană, iar simbolul $\top$ reprezintă operaţia de transpunere de matrice.)**

Răspuns:

Faptul că matricea Σ este simetrică decurge imediat din definiţia ei: dacă $X = (X_1, \ldots, X_n)$, atunci $[Cov(X)]_{i,j} \stackrel{def.}{=} Cov(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])] = E[(X_j - E[X_j])(X_i - E[X_i])] = Cov(X_j, X_i) = [Cov(X)]_{j,i}$, pentru orice $i, j \in \{1, \ldots, n\}$.
Apoi, pentru orice vector $z \in \mathbb{R}^n$ de forma $z = (z_1, \ldots, z_n)$, avem:

$$
\begin{aligned}
z^\top \Sigma z &= \sum_{i=1}^n z_i (\sum_{j=1}^n \Sigma_{ij} z_j) = \sum_{i=1}^n \sum_{j=1}^n (z_i \Sigma_{ij} z_j) = \sum_{i=1}^n \sum_{j=1}^n (z_i \, Cov[X_i, X_j] \, z_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n (z_i \, E[(X_i - E[X_i])(X_j - E[X_j])] \, z_j) \\
&= E\left[\sum_{i=1}^n \sum_{j=1}^n z_i \, (X_i - E[X_i])(X_j - E[X_j]) \, z_j\right]
\end{aligned}
$$

Ultima dintre egalităţile de mai sus derivă din proprietatea de liniaritate a mediilor. Mai departe,

$$
\begin{aligned}
z^\top \Sigma z &= E\left[\left(\sum_{i=1}^n z_i \, (X_i - E[X_i])\right)\left(\sum_{j=1}^n (X_j - E[X_j]) \, z_j\right)\right] \\
&= E\left[\left(\sum_{i=1}^n (X_i - E[X_i]) \, z_i\right)\left(\sum_{j=1}^n (X_j - E[X_j]) \, z_j\right)\right]
\end{aligned}
$$

Pentru a finaliza demonstraţia, să mai observăm că ultima expresie obţinută mai sus se scrie sub formă vectorială astfel:

$$
E[((X - E[X])^\top \cdot z)^2],
$$

ceea ce evident, reprezintă o cantitate ne-negativă. Aşadar, $z^\top \Sigma z \geq 0$.

25. (Distribuţii gaussiene multi-variate: o proprietate importantă, în cazul în care matricea de covarianţă este diagonală)

■ □ • ***prelucrare de Liviu Ciortuz după "The Multivariate Gaussian Distribution", Chuong B. Do, 2008***

Fie o variabilă aleatoare $X : \Omega \to \mathbb{R}^d$. În cele ce urmează elementele din $\mathbb{R}^d$ vor fi considerate vectori-coloană. Vom nota cu $\mathbb{S}^d_+$ spaţiul matricilor simetrice pozitiv definite[13] de dimensiune $d \times d$.

Vom spune că variabila X, reprezentată sub forma $X = [X_1 \ldots X_d]^\top$, urmează o distribuţie gaussiană (sau, *normală*) multi-variată, având media $\mu \in \mathbb{R}^d$ şi

[13] Prin definiţie, o matrice $A \in \mathbb{R}^{d \times d}$ este pozitiv definită dacă $x^\top A x > 0$ pentru orice $x \neq 0$ din $\mathbb{R}^d$, unde simbolul $\top$ reprezintă operaţia de transpunere de matrice.

matricea de covarianţă $\Sigma \in \mathbb{S}_+^d$, dacă funcţia ei de densitate [de probabilitate] are forma analitică următoare:

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)\right),$$

unde notaţia $|\Sigma|$ desemnează discriminantul matricii Σ, iar $\exp(\ )$ desemnează funcţia exponenţială având baza e.[14] Pe scurt, vom nota această proprietate de definiţie a lui X sub forma $X \sim \mathcal{N}(\mu, \Sigma)$.

Observaţii:
1. Se poate demonstra că orice matrice pozitiv definită Σ este inversabilă (deci $|\Sigma| \neq 0$), iar Σ^{-1} este de asemenea matrice pozitiv definită. Aşadar pentru orice vector nenul z, vom avea $z^\top \Sigma^{-1} z > 0$. Aceasta implică faptul că pentru orice vector $x \neq \mu$, vom avea:

$$(x-\mu)^\top \Sigma^{-1} (x-\mu) > 0, \text{ deci } -\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu) < 0.$$

Similar cazului uni-variat, graficul acestei funcţii exponenţiale va avea forma unui clopot cu deschiderea cuadratică, îndreptată în jos.
2. Coeficientul funcţiei exponenţiale, adică $\dfrac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}$, nu depinde de x, prin urmare el este doar un factor de normalizare folosit pentru a ne asigura că

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left((x-\mu)^\top \Sigma^{-1} (x-\mu)\right) dx_1\, dx_2 \ldots dx_d = 1.$$

Considerăm cazul simplu, în care $d = 2$ şi matricea de covarianţă Σ este diagonală,[15] deci

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Arătaţi că într-un astfel de caz, expresia funcţiei de densitate [de probabilitate] gaussiană multi-variată este identică cu produsul a două funcţii de densitate de tip gaussian, uni-variate şi independente, prima funcţie având media μ_1 şi varianţa σ_1^2, iar cea de-a doua având media μ_2 şi varianţa σ_2^2.

Răspuns:

În condiţiile de mai sus, funcţia de densitate [de probabilitate] gaussiană multi-variată are forma

$$p(x; \mu, \Sigma) \;=\; \frac{1}{2\pi \begin{vmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{vmatrix}^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right)$$

[14] În cazul general al unui vector X format din d variabile aleatoare, elementul generic (i, j) al matricii de covarianţă asociate lui X este prin definiţie $Cov(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])]$ pentru $i, j \in \{1, \ldots, d\}$. Pentru demonstraţia faptului că matricea de covarianţă Σ este întotdeauna simetrică şi pozitiv semi-definită vezi problema 24.

[15] În aceste condiţii, se poate demostra uşor că elementele de pe diagonala matricii Σ sunt strict pozitive. Într-adevăr, este suficient să se particularizeze vectorul z din formalizarea proprietăţii de ***pozitiv-definire*** a matricii Σ la valoarea $z = (1, 0)$ şi respectiv $z = (0, 1)$.

Aplicând mai întâi formula pentru determinantul de ordin 2,[16] şi calculând apoi inversa matricei Σ, obţinem:

$$\begin{aligned}
p(x;\mu,\Sigma) &= \frac{1}{2\pi\,\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}^\top \left(\frac{1}{\sigma_1^2\sigma_2^2}\begin{bmatrix}\sigma_2^2 & 0 \\ 0 & \sigma_1^2\end{bmatrix}\right)\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}\right) \\
&= \frac{1}{2\pi\,\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}^\top \begin{bmatrix}\frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2}\end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}\right) \\
&= \frac{1}{2\pi\,\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}^\top \begin{bmatrix}\frac{1}{\sigma_1^2}(x_1-\mu_1) \\ \frac{1}{\sigma_2^2}(x_2-\mu_2)\end{bmatrix}\right) \\
&= \frac{1}{2\pi\,\sigma_1\sigma_2}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2\right) \\
&= \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2\right)\frac{1}{\sqrt{2\pi}\,\sigma_2}\exp\left(-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2\right) \\
&= p(x_1;\mu_1,\sigma_1^2)\;p(x_2;\mu_2,\sigma_2^2).
\end{aligned}$$

Generalizând, este imediat că orice variabilă gaussiană d-dimensională de medie $\mu\in\mathbb{R}^d$ şi matrice de covarianţă diagonală $\Sigma \stackrel{not.}{=} diag(\sigma_1^2,\sigma_2^2,\ldots,\sigma_d^2)$, se comportă ca o colecţie de d variabile aleatoare gaussiene uni-variate independente, având respectiv mediile μ_i şi varianţele σ_i^2.

26. **(Distribuţia gaussiană bi-variată; o proprietate: distribuţia condiţională a unei componente în raport cu cealaltă componentă este tot de tip gaussian; identificarea parametrilor acestei distribuţii condiţionale)**

■□ · ***formulare de Liviu Ciortuz, după "Pattern Classification" (2nd ed.), [Appendix A. Mathematical foundations] R. Duda, P. Hart, D. Stork. John Wiley & Sons Inc., 2001***

Fie X o variabilă aleatoare care urmează o distribuţie gaussiană bi-variată de parametri μ (vectorul de medii) şi Σ (matricea de covarianţă). Aşadar, $\mu=(\mu_1,\mu_2)\in\mathbb{R}^2$, iar $\Sigma\in\mathcal{M}_{2\times 2}(\mathbb{R})$.

Prin definiţie, $\Sigma = Cov(X,X)$, unde $X \stackrel{not.}{=} (X_1,X_2)$, aşadar $\Sigma_{ij} = Cov(X_i,X_j)$ pentru $i,j\in\{1,2\}$. De asemenea, $Cov(X_i,X_i) = Var[X_i] \stackrel{not.}{=} \sigma_i^2 \geq 0$ pentru $i\in\{1,2\}$, în vreme ce pentru $i\neq j$ avem $Cov(X_i,X_j) = Cov(X_j,X_i) \stackrel{not.}{=} \sigma_{ij}$. În sfârşit, dacă introducem „coeficientul de corelare" $\rho \stackrel{def.}{=} \dfrac{\sigma_{12}}{\sigma_1\sigma_2}$, rezultă că putem scrie matricea de covarianţă Σ sub forma următoare:

[16] Anume, $\begin{vmatrix} a & b \\ c & d\end{vmatrix} = ad-bc$.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{1}$$

Demonstraţi că ipoteza $X \sim \mathcal{N}(\mu, \Sigma)$ implică faptul că distribuţia condiţională $X_2|X_1$ este de tip gaussian, şi anume $X_2|X_1 = x_1 \sim \mathcal{N}(\mu_{2|1}, \sigma_{2|1}^2)$, cu $\mu_{2|1} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$ şi $\sigma_{2|1}^2 = \sigma_2^2(1-\rho^2)$.

***Observaţie*: Pentru $X_1|X_2$, rezultatul este similar: $X_1|X_2 = x_2 \sim \mathcal{N}(\mu_{1|2}, \sigma_{1|2}^2)$, cu $\mu_{1|2} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$ şi $\sigma_{1|2}^2 = \sigma_1^2(1-\rho^2)$.**

Răspuns:

Conform definiţiei distribuţiei condiţionale, $X_2|X_1 = x_1$ are funcţia de densitate de probabilitate

$$p(x_2|x_1) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}, \tag{2}$$

unde p_{X_1,X_2} este densitatea distribuţiei gaussiene bi-variate de parametri μ şi Σ, iar p_{X_1} este distribuţia marginală a variabilei X_1. Aşadar,

$$\begin{aligned} p_{X_1,X_2}(x_1,x_2) &= \frac{1}{(\sqrt{2\pi})^2\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right) \text{ şi} \\ p_{X_1}(x_1) &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2\right). \end{aligned} \tag{3}$$

Se constată imediat din relaţia (1) că *determinantul* matricei Σ este $\sigma_1^2\sigma_2^2(1-\rho^2)$. Întrucât din expresia lui p_{X_1,X_2} de mai sus rezultă că Σ trebuie să fie matrice inversabilă şi că trebuie să existe radicalul din $|\Sigma|$, vom avea $\rho \in (-1, 1)$. În sfârşit, pentru că se consideră $\sigma_1,\ \sigma_2 > 0$, rezultă că $\sqrt{|\Sigma|} = \sigma_1\sigma_2\sqrt{1-\rho^2}$.
Inversa matricei Σ se calculează astfel:

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\Sigma^* = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \\ &= \frac{1}{(1-\rho^2)}\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \end{aligned}$$

Prin urmare,

$$\begin{aligned} & p_{X_1,X_2}(x_1,x_2) = \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}(x_1-\mu_1)\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \\ &\quad \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right) \end{aligned} \tag{4}$$

Observaţie: **Din expresia pe care tocmai am obţinut-o se observă uşor că $p_{X_1,X_2}(x_1, x_2)$ îşi atinge maximul atunci când $(x_1, x_2) = (\mu_1, \mu_2)$. Curbele de ecuaţie $p_{X_1,X_2}(x_1, x_2) = c$, pentru diverse valori ale lui $c \in \mathbb{R}^+$, se numesc *curbe de izocontur*. Ele au formă elipsoidală.**

Înlocuind expresiile (3) şi (4) în definiţia (2), vom obţine:

$$
\begin{aligned}
p(x_2|x_1) &= \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)} \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \\
&\quad \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right) \\
&\quad \cdot\sqrt{2\pi}\sigma_1 \exp\left(\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_2-\mu_2}{\sigma_2} - \rho\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left(\frac{x_2-[\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)]}{\sigma_2\sqrt{1-\rho^2}}\right)^2\right]
\end{aligned}
$$

Din această expresie se observă că $X_2|X_1 = x_1$ urmează o distribuţie gaussiană de parametri $\mu_{2|1} \stackrel{not.}{=} \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$ şi $\sigma^2_{2|1} \stackrel{not.}{=} \sigma^2_2(1-\rho^2)$.

27. **(Variabile aleatoare discrete: distribuţia Bernoulli, distribuţia normală standard; intervale de încredere, teorema limită centrală: aplicaţie la calculul erorii reale a unui clasificator)**

■• *CMU, 2008 fall, Eric Xing, HW3, pr. 3.3*

Nu demult, Chris s-a decis să folosească un nou clasificator (binar) care filtrează emailurile spam. Ulterior, el a vrut să evalueze cât de bun este acest clasificator. În acest scop, el a testat clasificatorul pe un mic set de date constituit din 100 de emailuri alese în mod aleatoriu dintre toate emailurile sale. Rezultatul pe care l-a obţinut a fost următorul: 83 de emailuri au fost clasificate corect. Aşadar rata erorii produse (sau: eroarea medie produsă) de clasificator pe acest mic set de date este de 17%. Este evident însă că eroarea aceasta este mai mică sau mai mare decât *eroarea reală*, pur şi simplu datorită alegerii aleatoare a celor 100 de emailuri.

Dacă se consideră un nivel de încredere de 95%, în ce interval se va situa eroarea reală (ţinând cont de acest experiment)?

Răspuns:

Vom nota cu X_i variabila aleatoare care reprezintă producerea unui mesaj email ($i = 1, \ldots, n = 100$) şi vom considera că $X_i = 1$ dacă emailul respectiv este clasificat eronat şi 0 în cazul contrar.

Notăm cu $\mu \stackrel{not.}{=} e_{real}$ media variabilei X_i şi cu σ^2 varianţa variabilei X_i. (Valorile lui μ şi σ nu depind de i).

Conform Teoremei Limită Centrală, eroarea la eşantionare,

$$e_{sample} \stackrel{not.}{=} \frac{X_1 + \ldots + X_n}{n},$$

văzută ca variabilă aleatoare, este aproximată de $\mathcal{N}(\mu, \sigma)$, distribuţia normală de medie μ şi varianţă σ^2. În consecinţă, notând

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\,\sigma},$$

rezultă că $Z_n \sim \mathcal{N}(0,1)$, deci pentru orice $a \in \mathbb{R}$ vom avea $P(Z_n \leq a) \to \Phi(a)$, unde $\Phi(x) \stackrel{def.}{=} P(Z \leq x)$ desemnează funcţia de distribuţie cumulativă (engl., cumulative distribution function, c.d.f.) pentru distribuţia normală standard (vezi problema 23).

Folosind notaţiile de mai sus, vom scrie următoarele echivalenţe, care au loc pentru orice $a \geq 0$:

$$\begin{aligned}
\mid Z_n \mid \leq a \quad & \Leftrightarrow \quad \left|\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\,\sigma}\right| \leq a \Leftrightarrow \left|\frac{X_1 + \ldots + X_n - n\mu}{n\sigma}\right| \leq \frac{a}{\sqrt{n}} \\
& \Leftrightarrow \quad \left|\frac{X_1 + \ldots + X_n - n\mu}{n}\right| \leq \frac{a\sigma}{\sqrt{n}} \Leftrightarrow \left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \leq \frac{a\sigma}{\sqrt{n}} \\
& \Leftrightarrow \quad |e_{sample} - e_{real}| \leq \frac{a\sigma}{\sqrt{n}} \Leftrightarrow |e_{real} - e_{sample}| \leq \frac{a\sigma}{\sqrt{n}} \\
& \Leftrightarrow \quad -\frac{a\sigma}{\sqrt{n}} \leq e_{real} - e_{sample} \leq \frac{a\sigma}{\sqrt{n}} \Leftrightarrow e_{sample} - \frac{a\sigma}{\sqrt{n}} \leq e_{real} \leq e_{sample} + \frac{a\sigma}{\sqrt{n}} \\
& \Leftrightarrow \quad e_{real} \in [e_{sample} - \frac{a\sigma}{\sqrt{n}}, e_{sample} + \frac{a\sigma}{\sqrt{n}}] \qquad (5)
\end{aligned}$$

Pentru a determina efectiv invervalul de încredere cerut, rămâne să mai aflăm a şi σ şi apoi să le înlocuim în relaţia (5).

Constanta a se determină uşor ţinând cont de faptul că în enunţ se precizează *nivelul de încredere* (95%) pentru apartenenţa erorii reale la intervalul specificat. Această restricţie corespunde relaţiei $P(\mid Z_n \mid \leq a) = 0.95$. Pe de altă parte, întrucât $\Phi(-a) + \Phi(a) = 1$, avem:

$$P(\mid Z_n \mid \leq a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1,$$

deci

$$P(\mid Z_n \mid \leq a) = 0.95 \Leftrightarrow 2\Phi(a) - 1 = 0.95 \Leftrightarrow \Phi(a) = 0.975 \Leftrightarrow a \cong 1.97 \text{ (vezi pr. 23)}.$$

Ne-a mai rămas de calculat valoarea lui σ. Vom ţine cont de faptul că

$$\sigma^2 \stackrel{not.}{=} Var_{real} = e_{real}(1 - e_{real}), \qquad (6)$$

ultima egalitate având loc fiindcă toate variabilele X_i sunt de tip Bernoulli. În relaţia (6) vom putea aproxima e_{real} cu e_{semple}, întrucât

$$E[e_{sample}] = e_{real} \text{ şi } Var_{sample} = \frac{1}{n} Var_{real} \to 0 \text{ pentru } n \to +\infty,$$

conform raţionamentelor din problema 5.ab de la capitolul *Estimarea probabilităţilor*.

În sfârşit, putem scrie:

$$\frac{a\sigma}{\sqrt{n}} = \frac{\sqrt{0.17(1-0.17)}}{\sqrt{100}} \cong 0.07,$$

deci relaţia (5) devine $e_{real} \in [0.10, 0.24]$.

Sumarizând, putem afirma cu încredere de 95% că eroarea reală a clasificatorului folosit de Chris este mai mare sau egală cu 0.1 şi mai mică sau egală cu 0.24.

28. **(Variabile aleatoare: Adevărat sau Fals?)**

CMU, 2006 fall, E. Xing, T. Mitchell, final exam, pr. 1.b
CMU, 2008 fall, Eric Xing, midterm exam, pr. 1.1, 1.2, 1.3

a. Dacă o variabilă aleatoare continuă X are funcţia densitate de probabilitate p diferită de zero pe tot domeniul de definiţie, atunci probabilitatea ca X să ia o valoare oarecare x (notaţie: $P(X = x)$) este egală cu $p(x)$.

b. $E[X + Y] = E[X] + E[Y]$ pentru orice două variabile aleatoare X şi Y.

c. $Var[X + Y] = Var[X] + Var[Y]$ pentru orice două variabile aleatoare X şi Y.

d. $E[XY] = E[X] \cdot E[Y]$ pentru orice două variabile aleatoare X şi Y.

<u>Răspuns:</u>

a. Fals (în general). Dacă variabila aleatoare continuă X are funcţia densitate de probabilitate p, atunci $P(a \leq X \leq b) = \int_a^b p(x)dx$. Prin urmare, $P(X = x) = 0$ pentru orice $x \in \mathbb{R}$. Enunţul afirmă pe de o parte că $p(x) = P(X = x) = 0$ pentru un anume $x \in \mathbb{R}$, iar pe de altă parte că p este diferită de zero pe tot domeniul de definiţie, ceea ce este absurd.

b. Adevărat. Demonstraţia este făcută în problema 9 punctul a.

c. Fals (în general). Se poate considera situaţia $Y = -X$, caz în care $Var[X + Y] = 0$, dar $Var[X] + Var[Y] = E[X^2] - (E[X])^2 + E[(-X)^2] - (E[-X])^2 = 2\,Var[X]$. Aşadar, pentru orice variabilă aleatoare X cu $Var[X] \neq 0$, luând $Y = -X$, rezultă că $Var[X + Y] = 0$ iar $Var[X] + Var[Y] \neq 0$. Un exemplu de variabilă aleatoare cu varianţa nenulă este distribuţia gaussiană.

Mai *general*, se poate demonstra că $Var[X + Y] = Var[X] + Var[Y] + 2\,Cov[X, Y]$ astfel:

$$\begin{aligned}
Var[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - (E[X])^2 - 2E[X] \cdot E[Y] - (E[Y])^2 \\
&= (E[X^2] - (E[X])^2) + (E[Y^2] - (E[Y])^2) + (2E[XY] - 2E[X] \cdot E[Y]) \\
&= Var[X] + Var[Y] + 2\,Cov[X, Y]
\end{aligned}$$

Prin urmare, egalitatea din enunţ este adevărată pentru orice două variabile aleatoare X şi Y pentru care $Cov[X, Y] = 0$ (vezi problema 9.c), dar este falsă în rest. Egalitatea $Cov[X, Y] = 0$ este adevărată de exemplu atunci când X şi Y sunt variabile independente (vezi problema 10).

d. Fals, în general. Afirmaţia din enunţ este echivalentă cu $Cov[X, Y] = 0$. Aşa cum am menţionat la punctul c, ea este adevărată dacă, de exemplu X şi Y sunt variabile aleatoare independente. Dacă, în schimb, vom considera de pildă cazul $Y = X$, cu X variabilă aleatoare binară care ia valoarea 1 cu probabilitatea p şi valoarea 0 cu probabilitatea $1-p$, se poate arăta imediat că $E[X^2] \neq (E[X])^2$ pentru orice $p \in (0, 1)$.

Variabile aleatoare

46. **(Variabile aleatoare: independenţa condiţională; Formula lui Bayes)**

◦ ***CMU, 2005 fall, T. Mitchell, A. Moore, midterm, pr. 1.1***

a. Fie variabilele aleatore H, E_1 şi E_2. Presupunem că vrem să calculăm $P(H \mid E_1, E_2)$, dar nu avem informaţii referitoare la independenţa condiţională. Care din următoarele seturi de numere sunt suficiente pentru acest scop?

i. $P(E_1, E_2)$, $P(H)$, $P(E_1 \mid H)$, $P(E_2 \mid H)$
ii. $P(E_1, E_2)$, $P(H)$, $P(E_1, E_2 \mid H)$
iii. $P(H)$, $P(E_1 \mid H)$, $P(E_2 \mid H)$.

b. Presupunem acum că $P(E_1|H, E_2) = P(E_1|H)$ pentru toate valorile posibile ale lui H, E_1 şi E_2. În acest caz, care dintre seturile de numere de la punctul a sunt suficiente pentru a calcula $P(H \mid E_1, E_2)$?

47. **(Independenţa condiţională a variabilelor aleatoare: o proprietate)**

∗ ***CMU, 2009 fall, Geoff Gordon, HW2, pr. 2.2***

Arătaţi că dacă variabilele aleatoare X şi Y sunt independente în raport cu Z (adică: $P(X, Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z)$), iar distribuţia de probabilitate corelată a lui X şi Y este independentă de W în raport cu Z, atunci X şi W sunt independente în raport cu Z.

În notaţie simplificată: $X \perp Y \mid Z$ şi $(X, Y) \perp W \mid Z$ implică $X \perp W \mid Z$ (şi similar $Y \perp W \mid Z$).

48. **(Variabile aleatoare: medii şi varianţe; exemplificări ale unor proprietăţi)**

CMU, 2016 fall, N. Balcan, M. Gormley, HW2, pr. 1.4

Fie X o variabilă aleatoare având media $E[X] = 1$ şi varianţa $Var(X) = 1$. Calculaţi:

i. $E[3X]$;
ii. $Var(3X)$;
iii. $Var(X + 3)$.

49. **(Variabila *indicator* pentru un eveniment aleator: calculul mediei)**

• ◦ ***CMU, 2015 spring, T. Mitchell, N. Balcan, HW2, pr. 1.c***

Fie un eveniment aleatoriu oarecare A, iar X o variabilă aleatoare definită astfel:

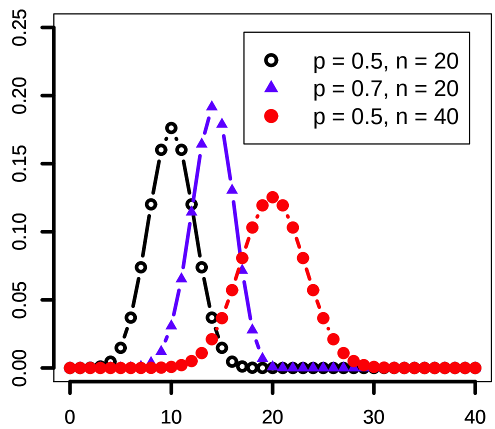
$$X = \begin{cases} 1 & \text{dacă evenimentul } A \text{ se realizează} \\ 0 & \text{în caz contrar.} \end{cases}$$

Uneori, X este numită variabila aleatoare *indicator* pentru evenimentul A. Arătaţi că $E[X] = P(A)$, unde $E[X]$ reprezintă *valoarea medie* a lui X.

50. **(Variabile aleatoare discrete: distribuţia binomială)**

CMU, 2009 spring, Ziv Bar-Joseph, HW1, pr. 1.4

Distribuţia binomială: p.m.f.



Un marinar încearcă să meargă pe o punte alunecoasă, însă datorită mişcărilor navei, el poate face exact un pas la fiecare interval de timp egal cu unitatea, şi anume: fie un pas înainte (cu probabilitatea p), fie un pas înapoi (cu probabilitatea $1-p$). Marinarul nu poate rămâne imobil.
Poziţia marinarului la momentul de timp $i \in \{0, 1, \ldots\}$ va fi exprimată cu ajutorul unei variabile aleatoare $X_i \in \{-\infty, \ldots, -2, -1, 0, 1, 2, \ldots, +\infty\}$, astfel:

$X_0 = 0$ **cu probabilitate** 1;
$P(X_{t+1} = x_i + 1 \mid X_t = x_i) = p,$ **pentru** $t \geq 0$;
$P(X_{t+1} = x_i - 1 \mid X_t = x_i) = 1 - p,$ **pentru** $t \geq 0$.

a. Cât este $P(X_{16} = 8)$, adică probabilitatea ca marinarul să se afle în poziţia $+8$ după exact 16 unităţi de timp?

b. Generalizare: Cât este $P(X_n = r)$, adică probabilitatea ca marinarul să se afle în poziţia r după exact n unităţi de timp (considerând $n \geq r$ şi $n - r$ număr par)?

c. Bazat pe formula de la punctul precedent, calculaţi $P(X_{32} = 16)$.

d. Este probabilitatea de la punctul c aceeaşi cu $P(X_{32} = 16 \mid X_{16} = 8)$, adică probabilitatea ca marinarul să se afle la poziţia $+16$ după exact 32 de unităţi de timp ştiind că la momentul de timp 16 se afla la poziţia $+8$?

51. **(Distribuţia binomială: calculul mediei şi al varianţei)**

■ □ ∗ ***Liviu Ciortuz, 2015***

Calculaţi media şi varianţa distribuţiei binomiale de parametri r, n şi p. (Notaţie: $b(r; n, p) = C_n^r p^r (1-p)^{n-r}$.)

Vă reamintim că această distribuţie reprezintă numărul (r) de apariţii ale feţei cu stema (corespondent engl., *head*) obţinute la efectuarea a n aruncări independente ale unei monede. Veţi presupune că probabilitatea de apariţie a stemei la o aruncare oarecare a acestei monede este p.

52. **(Variabile aleatoare discrete – distribuţia categorială – şi evenimente independente şi identic distribuite; calcul de medii)**

◦ ***CMU, 2009 fall, Geoff Gordon, HW1, pr. 4***

Presupunem că avem n coşuri şi m mingi. Aruncăm mingile în coşuri în mod independent şi aleatoriu, aşa încât fiecare minge este la fel de probabil să cadă în oricare dintre coşuri. (Pentru simplitate, vom presupune că la orice aruncare mingea cade într-un coş oarecare .)

a. Care este probabilitatea ca prima minge să cadă în primul coş?

b. Care este, în medie, numărul de mingi care au căzut în primul coş?

***Sugestii*:**
1. Puteţi defini o variabilă aleatoare care să reprezinte (similar unei *funcţii-indicator*) faptul că mingea i a căzut în primul coş:

$$X_i = \begin{cases} 1 & \textbf{dacă mingea } i \textbf{ a căzut în primul coş}, \\ 0 & \textbf{altfel.} \end{cases}$$

2: Ţineţi cont de liniaritatea mediei (adică, media unei combinaţii liniare de variabile este...).

c. Care este probabilitatea ca primul coş să rămână gol după aruncarea celor m mingi?

d. Care este, în medie, numărul de coşuri care rămân goale după aruncarea celor m mingi?

53. **(Variabile aleatoare discrete: distribuţii corelate, distribuţii marginale; medii, independenţă)**

CMU, 2009 fall, Carlos Guestrin, HW1, pr. 1.2

Fie două variabile aleatoare discrete cu distribuţia (i.e., funcţia masă de probabilitate) corelată dată în tabelul de mai jos (partea dreaptă).

X	Y	$P(X,Y)$
0	0	α
0	1	β
1	0	γ
1	1	δ

a. Calculaţi valorile medii ale variabilelor X şi Y în funcţie de α, β, γ şi δ.

b. Găsiţi condiţiile necesare şi suficiente astfel încât variabilele X şi Y să fie independente.

54. **(Covarianţa nulă vs. independenţa variabilelor aleatoare binare)**

∗ ***CMU, 2009 fall, Geoff Gordon, HW1, pr. 3.2***

Se ştie că atunci când covarianţa a două variabile aleatoare este nulă nu rezultă în mod neapărat că variabilele respective sunt independente (vezi pr. 11.a). Însă, dacă X şi Y sunt variabile aleatoare binare luând valori în mulţimea $\{0,1\}$ şi covarianţa lor este nulă, rezultă că X şi Y sunt independente (vezi pr. 11.b).

Arătaţi că, în cazul în care doar variabila aleatoare X este binară, nu rezultă în mod neapărat că X şi Y sunt independente, deşi covarianţa lor este nulă.

***Indicaţie*: Folosiţi ca exemplu variabilele aleatoare ale căror distribuţii sunt date în tabelul alăturat.**

X	Y	$P(X,Y)$
0	−1	0.1
0	0	0.4
0	1	0.1
1	−1	0
1	0	0.4
1	1	0

55. **(Variabile aleatoare discrete: distribuţii corelate, distribuţii marginale, distribuţii condiţionale; independenţă, independenţă condiţională)**

• ⋆ ***CMU, 2016 fall, N. Balcan, M. Gormley, HW2, pr. 1.4***

Fie trei variabile aleatoare X, Y şi Z care iau valori în mulţimea $\{0,1\}$. În tabelul de mai jos este dată distribuţia probabilistă corelată a acestor trei variabile, $P(X,Y,Z)$.

	$Z=0$		$Z=1$	
	$X=0$	$X=1$	$X=0$	$X=1$
$Y=0$	$1/24$	$1/12$	$1/12$	$5/24$
$Y=1$	$1/12$	p	q	$7/24$

a. Considerând că X şi Y sunt independente, găsiţi valorile lui p şi q.

***Indicaţie*: Este util (deşi nu obligatoriu) să calculaţi mai întâi $P(X,Y)$, distribuţia corelată a variabilelor X şi Y, completând tabel următor, după care veţi calcula şi distribuţiile (marginale) pentru X şi Y, de preferinţă ca o linie şi respectiv o coloană suplimentară la acest tabel.**

	$X=0$	$X=1$
$Y=0$		
$Y=1$		

b. Considerând valorile lui p şi q determinate la punctul a, sunt X şi Y independente condiţional în raport cu Z? De ce?

56. **(O mixtură de distribuţii categoriale: calculul mediei şi al varianţei)**

■ □ • ∘ ***CMU, 2010 fall, Aarti Singh, HW1, pr. 2.2.1-2***

Presupunem că avem două zaruri cu şase feţe. Un zar este perfect, iar celălalt zar este măsluit şi are următoarea funcţie masă de probabilitate (p.m.f.):

$$P_2(x) = \begin{cases} \dfrac{1}{2} & \text{pentru } x = 6; \\ \dfrac{1}{10} & \text{pentru } x \in \{1, 2, 3, 4, 5\}. \end{cases}$$

Vom folosi o monedă pentru a decide ce zar să aruncăm. Dacă la aruncarea monedei obţinem stema (engl., head), vom arunca apoi zarul perfect; în caz contrar vom arunca zarul măsluit. Probabilitatea ca la aruncarea monedei să obţinem stema este $p \in (0, 1)$.

a. Calculaţi în funcţie de p media variabilei aleatoare (X) care reprezintă „mixtura“ de distribuţii [categoriale] descrisă mai sus.

b. Calculaţi în funcţie de p varianţa variabilei aleatoare X.

57. **(Variabile aleatoare continue: funcţia densitate de probabilitate)**

∗ ***CMU, 2008 spring, Eric Xing, HW1, pr. 1.1.a***

Fie funcţia $f(x) = ce^{-|x|}, -\infty < x < \infty$.

Ce valoare/valori poate avea constanta reală c în aşa fel ca f să poată reprezenta o funcţie densitate de probabilitate?

58. **(Variabile aleatoare continue: calculul unei probabiltăţi, folosind funcţia densitate de probabilitate)**

∗ ***CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, midterm, pr. 1.b***

Presupunem că funcţia densitate de probabilitate a unei variabile aleatoare continue X este definită astfel:

$$p(x) = \begin{cases} \dfrac{4}{3}(1 - x^3) & \textbf{pentru } 0 \leq x \leq 1 \\ 0 & \textbf{în caz contrar}. \end{cases}$$

Cât este $P(X < 0)$?

59. **(Variabile aleatoare discrete vs. variabile aleatoare continue; distincţia dintre p.m.f. şi p.d.f.)**

□ • ∘ ***CMU, 2012 spring, Ziv Bar-Joseph, HW1, pr. 1.1***

Mickey nu este încă bine iniţiat în teoria probabilităţilor şi se confruntă cu chestiunea următoare, despre care el este înclinat să creadă că este un paradox:

Fie X o variabilă aleatoare cu distribuţia de probabilitate uniformă, definită pe intervalul $[0, \frac{1}{2}]$, şi anume $f(x) = 2$ pentru $x \in [0, \frac{1}{2}]$. Întrucât suma probabilităţilor care corespund unei distribuţii aleatoare trebuie să fie 1, Mickey este nedumerit de ce valoarea lui $f(x)$ este mai mare decât suma totală.

Explicaţi acest paradox. Altfel spus, arătaţi ce anume nu ştie Mickey.

60. **(Distribuţia Poisson: calculul mediei şi al varianţei; distribuţia Gamma: calculul mediei)**

Liviu Ciortuz, 2017

a. Distribuţia *Poisson* este o distribuţie discretă de parametru $\lambda > 0$, a cărei funcţie masă de probabilitate este dată de expresia

$$p(x \mid \lambda) = \frac{1}{e^{\lambda}} \cdot \frac{\lambda^x}{x!}, \text{ pentru orice } x \in \mathbb{N}.$$

Prin convenţie, se consideră că $0! = 1$. Factorul $\frac{1}{e^{\lambda}}$, care nu depinde de x, este aşa-numitul *factor de normalizare*.

Demonstraţi că media acestei distribuţii este λ, iar varianţa ei este tot λ.

b. Distribuţia *Gamma* este o distribuţie continuă, de parametri $r > 0$ (care dă *forma* distribuţiei, engl., shape) şi $\alpha > 0$ (numit *rata*, engl., rate), cu funcţia densitate de probabilitate definită pe $\mathbb{R}^+$ prin expresia următoare:

$$p(x) \stackrel{not.}{=} Gamma(x|r, \alpha) = \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x},$$

unde *factorul de normalizare* este $\frac{\alpha^r}{\Gamma(r)}$. Funcţia Γ (a lui Euler) este o generalizare în $\mathbb{R}^+$ a definiţiei numerelor factoriale din $\mathbb{N}$ (şi anume, $\Gamma(r) = (r-1)!$ pentru orice $r \in \mathbb{N}^*$).

Demonstraţi că media distribuţiei Gamma este $\frac{r}{\alpha}$.

***Sugestie*: Veţi ţine cont de una din proprietăţile funcţiei Γ a lui Euler: $\Gamma(r+1) = r \cdot \Gamma(r)$ pentru orice $r > 0$.**

61. **(Distribuţia gaussiană, cazul bi-variat: explicitarea p.d.f. într-un caz particular; matrici de covarianţă pentru vectori de variabile aleatoare: o proprietate)**

□ • ○ ***MIT, 2006 fall, Tommi Jaakkola, HW1, pr. 5***

a. Fie X un vector de variabile aleatoare de tip gaussian, cu

$$E[X] = \begin{pmatrix} 10 \\ 5 \end{pmatrix} \quad \text{şi} \quad Cov(X) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Scrieţi expresia funcţiei densitate de probabilitate (p.d.f.) pentru X, fără a folosi notaţia matriceală. Aşadar, considerând $X = (x_1, x_2)$, scrieţi funcţia sa de densitate de probabilitate corelată ca un $P(x_1, x_2)$.

b. Fie A, B matrice [de numere reale] de dimensiune $p \times q$, iar x un vector aleator de dimensiune $q \times 1$. Demonstraţi egalitatea următoare

$$Cov(Ax, Bx) = A\ Cov(x)\ B^\top,$$

unde prin $Cov(u, v) \stackrel{def.}{=} E[(u - E[u])(v - E[v])^\top]$ am notat matricea de cros-covarianţă pentru doi vectori aleatori oarecare u şi v, iar prin $Cov(u) \stackrel{def.}{=} E[(u - E[u])(u - E[u])^\top]$ am notat matricea de covarianţă pentru vectorul u.

62. **(Variabile aleatoare continue, variabile aleatoare discrete; independenţă, variabile aleatoare condiţionate, medii)**

◦ ***CMU, 2009 spring, Ziv Bar-Joseph, final exam, pr. 1.1***

Fie $X : \Omega \to [0, 1]$ o variabilă aleatoare continuă având distribuţia uniformă (notaţie: $X \sim Uniform(0, 1)$). Fie a şi b două numere reale, astfel încât $0 < a < b < 1$. Definim (tot pe Ω) variabilele aleatoare discrete Y şi Z în funcţie de valorile lui X, astfel:

$$Y(\omega) = \begin{cases} 1 & \textbf{dacă } 0 \leq X(\omega) \leq a, \\ 0 & \textbf{altfel} \end{cases}$$

$$Z(\omega) = \begin{cases} 1 & \textbf{dacă } b \leq X(\omega) \leq 1, \\ 0 & \textbf{altfel.} \end{cases}$$

a. Sunt variabilele Y şi Z independente? Justificaţi răspunsul.

***Indicaţie*: Pentru a găsi răspunsul corect vă sugerăm că ar fi util să completaţi tabelul de mai jos.**

y	z	$P(Y = y)$	$P(Z = z)$	$P(Y = y)P(Z = z)$	$P(Y = y, Z = z)$
0	0				
0	1				
1	0				
1	1				

b. Pentru fiecare din valorile z ale variabilei Z, calculaţi media variabilei condiţionate $Y \mid Z = z$. (Notaţie: $E_Y[Y \mid Z = z]$) .

63. **(Distribuţii probabiliste discrete şi distribuţii probabiliste continue)**

□ • ○ ***CMU, 2015 spring, T. Mitchell, N. Balcan, HW1, pr. 2.2***

Faceţi corespondenţa dintre numele de distribuţii probabiliste din coloana din stânga cu funcţiile [masă, respectiv densitate] de probabilitate din coloana din dreapta.

		f.	$p^{1-x}(1-p)^x$ **cu** $x \in \{0,1\}$
a.	**gaussiană multi-variată**	**g.**	$\dfrac{1}{b-a}$ **pentru** $a \leq x \leq b$; 0 **în caz contrar**
b.	**exponenţială**		
c.	**uniformă**	**h.**	$C_n^x p^x (1-p)^{n-x}$ **pentru** $x \in \{0,\ldots,n\}$
d.	**Bernoulli**	**i.**	$\lambda e^{-\lambda x}$ **pentru** $x \geq 0$; 0 **în caz contrar**
e.	**binomială**	**j.**	$\dfrac{1}{\sqrt{(2\pi)^d \lvert\Sigma\rvert}} \exp\left(-\frac{1}{2} - (x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$

64. **(Variabile aleatoare: Adevărat sau Fals?)**

□ • ○ ***CMU, 2005 fall, T. Mitchell, A. Moore, midterm, pr. 1.2***

Fie X şi Y două variabile aleatoare. Care dintre următoarele afirmaţii sunt adevărate?

a. Dacă X şi Y sunt independente, atunci $E[2XY] = 2E[X]\,E[Y]$ şi $Var[X+2Y] = Var[X] + Var[Y]$.

b. Dacă X şi Y sunt independente, iar $X > 1$, atunci $Var[X+2Y^2] = Var[X] + 4\,Var[Y^2]$ şi $E[X^2 - X] \geq Var[X]$.

c. Dacă X şi Y nu sunt independente, atunci $Var[X+Y] = Var[X] + Var[Y]$.

d. Dacă X şi Y sunt independente, atunci $E[XY^2] = E[X]\,E[Y]^2$ şi $Var[X+Y] = Var[X] + Var[Y]$.

e. Dacă X şi Y nu sunt independente şi $f(X) = X^2$, atunci $E[f(X)\,E[Y] = E[f(X)]\,E[Y]$ şi $Var[X+2Y] = Var[X] + 4\,Var[Y]$.