Seminar 10

Vectori aleatori continui

In acest seminar ne intereseaza următoarele noţiuni legate de vectori aleatori continui:

- Densitatea comună de probabilitate a unui vector aleator continuu. Funcţia de repartiţie;
- Calculul probabilităţii unor evenimente;
- Distribuţii marginale. Funcţii de repartiţie marginale;
- Variabile aleatoare condiţionate continue. Independenţă;
- Vector uniform continuu;
- Simulări;

10.1 Probleme rezolvate

1. Se consideră vectorul aleator (X,Y) cu densitatea de probabilitate

$$f_{X,Y}(x,y) = \begin{cases} 4xy, & \text{dacă } x \in [0,1],\ y \in [0,1], \\ 0, & \text{în rest.} \end{cases}$$

a) Să se arate ca f este o densitate de probabilitate.

b) Sa se vizualizeze evenimentul A: ($X < 0.5$ şi $Y > 0.5$) şi să se calculeze probabilitatea $P(A)$.

c) Să se determine funcţia de repartiţie a vectorului (X, Y).

d) Să se determine densităţile marginale $f_X(x)$ şi $f_Y(y)$.

e) Să se determine funcţiile de repartiţie marginale $F_X(x)$şi $F_Y(y)$.

f) Să se studieze dacă cele două variabile X şi Y sunt independente.

g) Să se calculeze $P(X + Y < 1)$.

Rezolvare: a) Din teorie ştim că f este densitate de probabilitate dacă:

- $f_{X,Y}(x,y) \geq 0$, pentru orice $(x,y) \in \mathbb{R}^2$.
- $\iint_{\mathbb{R}^2} f_{X,Y}(x,y)dx\,dy = 1$.

Evident funcţia $f \geq 0, \quad \forall (x,y) \in \mathbb{R}^2$. In plus,

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y)dx\,dy = \int_0^1 \left(\int_0^1 4xy\,dy \right) dx = \int_0^1 \left(4x\frac{y^2}{2} \right) \Big|_0^1 dx = \int_0^1 2xdx = x^2 \Big|_0^1 = 1$$

Deci, f este densitate de probabilitate, având suportul (unde este nenulă) pe pătratul $[0,1] \times [0,1]$.

b) Notăm cu $S = [0, \frac{1}{2}] \times [\frac{1}{2}, 1]$. Avem:

$$P(A) = \iint_S f_{X,Y}(x,y)dx\,dy = \int_0^{1/2} \left(\int_{1/2}^1 4xy\,dy \right) dx = \int_0^{1/2} \left(4x\frac{y^2}{2} \right) \Big|_{1/2}^1 dx = \int_0^{1/2} 2x\frac{3}{4}dx = \frac{3}{16}.$$

Observaţie: Daca vectorul (X,Y) ar fi uniform distribuit pe pătratul $R = [0,1] \times [0,1]$, atunci $f(x,y) = 1, \forall (x,y) \in [0,1] \times [0,1]$. In aceste condiţii

$$P(A) = \iint_S f_{X,Y}(x,y)dx\,dy = \iint_S 1dx\,dy = Aria(S) = \frac{1}{4}.$$

c) Din definiţie avem că pentru $(x,y) \in [0,1]^2$:

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(t,s)\,dt\,ds = \int_0^x \int_0^y 4ts\,dt\,ds = 4\int_0^x tdt \int_0^y s\,ds = x^2y^2.$$

d) Din teorie se ştie că densităţile marginale $f_X(x)$ şi $f_Y(y)$ sunt date de formulele

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Se obţine:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 4xydy = 4x\int_0^1 ydy = 2x, \forall x \in [0,1]$$

Deci,

$$f_X(x) = \begin{cases} 2x, & \text{dacă } x \in [0,1], \\ 0, & \text{în rest.} \end{cases}$$

In mod asemănător,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1 4xydx = 4y\int_0^1 xdx = 2y, \forall y \in [0,1]$$

Deci,

$$f_Y(y) = \begin{cases} 2y, & \text{dacă } y \in [0,1], \\ 0, & \text{în rest.} \end{cases}$$

e) Se ştie că $F_X(x) = \int\limits_{-\infty}^{x} f_X(t)dt$. Pentru $x \in [0,1]$, avem

$$F_X(x) = \int\limits_0^x f_X(t)dt = \int\limits_0^x 2tdt = x^2$$

f) Se observă că $f(x,y) = f_X(x) \cdot f_Y(y)$, deci X şi Y sunt independente.
g)

$$P(X+Y<1) = P((X,Y) \in D) = \iint_D f_{X,Y}(x,y)dx\,dy$$

Obţinem

$$P(X+Y<1) = \int_0^1\int_0^{1-x} 4xydy\,dx = 2\int_0^1 x(1-x)^2 dx = \frac{1}{6}.$$

2. Vectorul aleator (X,Y) are densitatea de probabilitate

$$f_{X,Y}(x,y) = \begin{cases} 6x, & \text{dacă } 0<x<y<1, \\ 0, & \text{în rest.} \end{cases}$$

a) Să se determine densitatea marginală f_X;

b) Să se determine densitatea de probabilitate a variabilei aleatoare $(Y|X=0.25)$;

c) Să se calculeze $P(X>0.5)$ şi $P(Y>0.5|X=0.25)$.

d) Să se determine media şi dispersia variabilei $(Y|X=0.25)$.

Rezolvare: a) Ţinând seama că densitatea de probabilitate $f_{X,Y}$ are suportul (este nenulă) pe domeniul triunghiular haşurat din Fig. **??**, avem:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_x^1 6x\,dy = 6x(1-x),$$

dacă $x \in (0,1)$, şi $f_X(x) = 0$, în rest.
b) Densitatea de probabilitate a variabilei aleatoare $(Y|X=0.25)$ este

$$h(y|0.25) = \frac{f_{XY}(0.25,y)}{f_X(0.25)} = \begin{cases} \frac{6\times 0.25}{6\times 0.25\times 0.75} = \frac{4}{3}, & \text{dacă } y \in (0.25,1), \\ 0, & \text{în rest.} \end{cases}$$

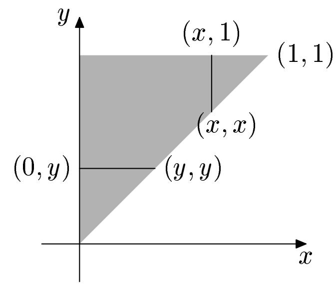

Fig.10.1: Densitatea de probabilitate a vectorului aleator este nenulă pe domeniul triunghiular haşurat.

c) Avem

$$P(X > 0.5) = \int_{0.5}^{\infty} f_X(x)\, dx = \int_{0.5}^{1} 6x(1-x)\, dx = 0.5,$$

iar

$$P(Y > 0.5|X = 0.25) = \int_{0.5}^{\infty} h(y|0.25)\, dy = \int_{0.5}^{1} \frac{4}{3}\, dy = \frac{2}{3}.$$

d) $M((Y|X = 0.25) = \int\limits_{-\infty}^{\infty} y \cdot h(y|0.25)dy = \int_{0.25}^{1} y \cdot \frac{4}{3}dy = \frac{5}{8}$.
Dispersia se va calcula cu formula $\sigma^2(Z) = M(Z^2) - M(Z)^2$, unde $Z = (Y|X = 0.25)$ Folosind LOTUS avem $M(Z^2) = \int\limits_{-\infty}^{\infty} y^2 \cdot h(y|0.25)dy = \int_{0.25}^{1} y^2 \cdot \frac{4}{3}dy = \frac{7}{16}$. Deci, $\sigma^2(Z) = \frac{3}{64}$.

3. Se consideră discul $G = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Se presupune că alegem un punct (x, y) aleator din discul circular G. Acest lucru se poate obţine simulând un vector (X, Y) uniform distribuit pe acest disc.

a) Să se determine densitate de probabilitate a vectorului (X, Y)

b) Să se determine densităţile marginale $f_X(x)$ şi $f_Y(y)$.

c) Să se determine densitatea de probabilitate a variabilei $(X|Y = y_0)$, pentru $-1 \leq y_0 \leq 1$.

d) Variabilele X şi Y sunt independente?

e) Să se scrie un pseudocod de simulare a unei valori de observaţie a vectorului (X, Y).

f) Dacă variabila ce înregistrează numărul de parcurgeri ale buclei do-while din algoritmul de la d) este distribuită geometric de parametru p, să se calculeze probabilitatea p. Care este probabilitatea ca primul număr generat in G să apară la a 3 parcurgere a buclei do-while?

Rezolvare: a) Vectorul (X,Y) este uniform distribuit pe discul G, deci are densitatea de probabilitate

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{Aria(G)} = \frac{1}{\pi}, & \text{dacă } (x,y) \in G, \\ 0, & \text{în rest.} \end{cases}$$

b) Vom determina $f_X(x)$ şi $f_Y(y)$ folosind formulele:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Pentru $-1 \le x \le 1$, avem $f_X(x) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int\limits_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy = \frac{2}{\pi}\sqrt{1-x^2}$ Deci,

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{dacă } -1 \le x \le 1, \\ 0, & \text{în rest.} \end{cases}$$

In mod asemănător,

$$f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2}, & \text{dacă } -1 \le y \le 1, \\ 0, & \text{în rest.} \end{cases}$$

c) Densitatea de probabilitate a variabilei $(X|Y=y_0)\,, -1 \le y_0 \le 1$ este

$$g(x|y_0) = \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)} = \begin{cases} \frac{1}{2\sqrt{1-y_0^2}}, & \text{dacă } -\sqrt{1-y_0^2} \le x \le \sqrt{1-y_0^2}, \\ 0, & \text{în rest.} \end{cases}$$

De exemplu, pentru $y_0 = 1/2$, avem

$$g(x|1/2) = \frac{f_{X,Y}(x,1/2)}{f_Y(1/2)} = \begin{cases} \frac{1}{\sqrt{3}}, & \text{dacă } -\sqrt{3}/2 \le x \le \sqrt{3}/2, \\ 0, & \text{în rest.} \end{cases}$$

Deci, variabila $(X|Y=y_0)$ este uniform distribuită pe intervalul $[-\sqrt{1-y_0^2}, \sqrt{1-y_0^2}]$.
d) Variabilele X şi Y nu sunt independente, pentru că $f_{X,Y}(x,y) \ne f_X(x)\cdot f_Y(y)$.
e) Pentru a determina algoritmul optim ce generează puncte uniform pe G vom determina cel mai mic dreptunghi, cu laturile paralele cu axele de coordonate şi care conţine mulţimea G. Evident, $D = [-1,1]\times[-1,1]$. Vom genera puncte uniform distribuite pe dreptunghiul D. Vectorul (X,Y) ce genereaza puncte uniform distribuite pe D are densitatea de probabilitate

$$g_{(X,Y)}(x,y) = \begin{cases} \frac{1}{Aria(D)} = \frac{1}{4}, & \text{dacă } (x,y) \in D, \\ 0, & \text{în rest.} \end{cases}$$

Rulăm algoritmul urmator in care in interiorul buclei do-while generam puncte uniform distribuite pe D si le retinem doar pe cele care sunt in G, pe baza metodei respingerii:

```
do{
  x = −1 + 2 ∗ urand();
  y = −1 + 2 ∗ urand();
}while(x ∗ x + y ∗ y > 1);
return (x, y);
```

f) Probabilitatea de succes, definită aici ca fiind probabilitatea ca un punct să aparţină lui G, se determină din relaţia

$$p = \iint_G g_{(X,Y)} dxdy = \iint_G \frac{1}{Aria(D)} dxdy = \frac{1}{Aria(D)} \iint_G 1 dxdy = \frac{Aria(G)}{Aria(D)} = \frac{\pi}{4}$$

Variabila N ce înregistrează numărul de parcurgeri ale buclei do-while din algoritmul de mai sus este distribuită geometric, $N \sim Geom(p)$. Deci, $P(N = 3) = (1 - p)^2 p$.

10.2 Probleme propuse

4. Se consideră vectorul aleator (X,Y) cu densitatea de probabilitate

$$f_{X,Y}(x, y) = \begin{cases} cx^2 y(1 + y), & \text{dacă } x \in [0, 3],\ y \in [0, 3], \\ 0, & \text{în rest.} \end{cases}$$

a) Să se determine constanta c astfel încât f să fie o densitate de probabilitate.

b) Sa se vizualizeze evenimentul A: ($1 \le X \le 2$ şi $0 \le Y \le 1$) şi să se calculeze probabilitatea $P(A)$.

c) Să se determine funcţia de repartiţie a vectorului (X, Y).

d) Să se determine densităţile marginale $f_X(x)$ şi $f_Y(y)$.

e) Să se determine funcţiile de repartiţie marginale $F_X(x)$şi $F_Y(y)$.

f) Să se studieze dacă cele două variabile X şi Y sunt independente.

5. Se consideră vectorul aleator (X,Y) cu densitatea de probabilitate

$$f_{X,Y}(x, y) = \begin{cases} x + y, & \text{dacă } x \in [0, 1],\ y \in [0, 1], \\ 0, & \text{în rest.} \end{cases}$$

a) Să se arate că f este o densitate de probabilitate.

b) Fie $F(x, y)$ funcţia de repartiţie a vectorului (X, Y). Să se calculeze $F(1, 1)$.

c) Să se determine densităţile marginale $f_X(x)$ şi $f_Y(y)$.

d) Sunt variabilele X şi Y sunt independente?

e) Să se determine $M(XY)$.

6. Se consideră două variabile aleatoare X şi Y ce dau timpii de execuţie a două procese paralele, independente şi unifom distribuite pe $(0,1)$, respectiv $(0,6)$.
Să se determine probabilitatea ca primul proces să fie executat după cel de-al doilea proces.
Cum se poate determina probabilitatea ca primul proces să fie executat înaintea celui de-al doilea proces (fără a calcula integrala dublă)?
Indicaţie: Se va calcula $P(X > Y) = P((X,Y) \in G)$, unde $G = \{(x,y) \in \mathbb{R}^2 : x > y\}$.

7. Se consideră vectorul aleator (X,Y) este uniform distribuit pe mulţimea $[-1,2] \times [-2,4]$. Să se determine expresia analitică a densităţii de probabilitate şi să se calculeze $P((X,Y) \in G)$, unde G este domeniul triunghiular cu varfurile $A(-1,-2), B(-1,4), C(2,3)$. Să se scrie algoritmul optim de generare a unui punct în triunghiul ABC.

8. Algoritmul de generare de puncte uniform distribuite pe discul eliptic

$$G = \{(x,y) \in \mathbb{R}^2 : \frac{x^2}{16} + y^2 \leq 1\}$$

este

```
do{
  x = -4 + 8 * urand();
  y = -1 + 2 * urand();
}while(x * x/16 + y * y > 1);
return (x, y);
```

a) Să se precizeze vectorul simulat in interiorul buclei `do-while`.

b) De ce acest algoritm este considerat optim pentru a genera puncte in G?

c) Stiind că generarea unui punct în discul eliptic imită un proces Bernoulli, să se precizeze probabilitatea de succes p (succesul este definit aici ca fiind evenimentul să se genereze un punct din G);

d) Să se calculeze probabilitatea ca primul punct generat in mulţimea G să apară la a patra parcurgere a buclei.

e) Să se determine numărul mediu de parcurgeri ale buclei până la generarea unui punct in discul G.