Repartiții comune, marginale și condiționate

a) Cazul discret (continuare)

X, Y discrete, $X \in \{x_1, \dots x_m\}$, $Y \in \{y_1, \dots y_n\}$

$$p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y) \quad \text{(rep. comună)}$$

Repartițiile marginale: ale lui X și Y:

$$p_X(x) = \mathbb{P}(X=x) = \sum_y p_{X,Y}(x,y)$$

$$p_Y(y) = \mathbb{P}(Y=y) = \sum_x p_{X,Y}(x,y)$$

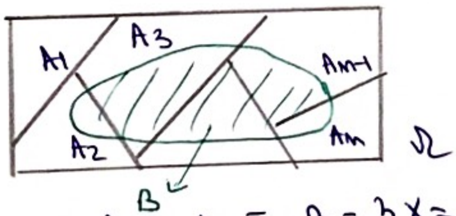
Repartiția condiționată la un eveniment A, $\mathbb{P}(A) > 0$

$$p_{X|A}(x) = \mathbb{P}(X=x \mid A)$$

$$\sum_x p_{X|A}(x) = 1$$

Formula probabilității totale: $A_1, A_2, \dots A_m$ evenimente disjuncte două câte două pe Ω a.î. $\Omega = \bigcup_{i=1}^{m} A_i$, $\mathbb{P}(A_i) > 0$ atunci pt $B \in \mathcal{F}$ avem

$$\mathbb{P}(B) = \sum_{i=1}^{m} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)$$



În particular, dacă $B = \{X = x\}$ atunci

$$p_X(x) = \sum_{i=1}^{m} p_{X|A_i}(x)\,\mathbb{P}(A_i)$$

Dacă în plus, $A_i = \{Y = y_i\}$, $i \in \{1, \dots m\}$ atunci

$$p_X(x) = \sum_{i=1}^{m} p_{X|Y}(x \mid y_i)\, p_Y(y_i)$$

care se poate rescrie sub forma:

$$\boxed{p_X(x) = \sum_y p_{X|Y}(x \mid y)\, p_Y(y)}$$ Formula prob. totale

Rep. condiționată a lui X la $Y=y$ este:

$$\boxed{p_{X|Y}(x \mid y) = \mathbb{P}(X=x \mid Y=y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}}$$

Astfel:

$$\boxed{\begin{aligned} p_{X,Y}(x,y) &= p_{X|Y}(x \mid y)\, p_Y(y) \\ &= p_{X|Y}(y \mid x)\, p_X(x) \end{aligned}}$$

Formula lui Bayes

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)\, p_{Y|X}(y|x)}{p_Y(y)}$$

$$\boxed{p_{X|Y}(x|y) = \frac{p_X(x) \cdot p_{Y|X}(y|x)}{\sum_{x'} p_X(x')\, p_{Y|X}(y|x')}}$$

Exp: O găină depune un număr aleator N de ouă, $N \sim Pois(\lambda)$.

Să presupunem că fiecare ou eclozează cu prob $p \in (0,1)$, independent de celelalte ouă. Fie X nr. de ouă care au eclozat și Y nr de ouă care nu au eclozat.

Adică $N = X+Y$. Vrem să determinăm repartiția comună a (X,Y).

Sol:

Vrem să determinăm probabilitatea $P(X=i, Y=j) = ?$, $i,j \geq 0$.

Din ipoteză știm că $N \sim Pois(\lambda) \Rightarrow \mathbb{P}(N=n) = e^{-\lambda}\frac{\lambda^n}{n!}$

Știm de asemenea că dacă $\{N=n\}$ atunci $X|N=n \sim B(n,p)$ pt că ouăle eclozează de manieră independentă unul față de celălalt. În mod similar,

$Y|N=n \sim B(n, 1-p)$

Putem scrie din formula prob. totale că:

$$\mathbb{P}(X=i, Y=j) = \sum_{n=0}^{\infty} \mathbb{P}(X=i, Y=j | N=n)\, \mathbb{P}(N=n)$$

Dacă $n \neq i+j$ atunci

$$\mathbb{P}(X=i, Y=j | N=n) = 0$$

astfel $\mathbb{P}(X=i, Y=j) = \mathbb{P}(X=i, Y=j | N=i+j)\, \mathbb{P}(N=i+j)$

$$\mathbb{P}(X=i, Y=j | N=i+j) = \mathbb{P}(X=i | N=i+j)$$
$$= \mathbb{P}(Y=j | N=i+j)$$

deci $\mathbb{P}(X=i, Y=j) = \binom{i+j}{i} p^i (1-p)^j \cdot e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}$

$1 = p + (1-p)$

$\Downarrow$

$\lambda = \lambda p + \lambda(1-p)$

$$= \frac{\cancel{(i+j)!}}{i!\,j!} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{\cancel{(i+j)!}}$$

$$= e^{-\lambda p}\frac{\lambda^i}{i!} \cdot e^{-\lambda(1-p)}\frac{\lambda^j}{j!} \cdot p^i \cdot (1-p)^j$$

$$= e^{-\lambda p}\frac{(\lambda p)^i}{i!} \cdot e^{-\lambda(1-p)} \cdot \frac{(\lambda(1-p))^j}{j!} \to Pois(\lambda(1$$

$$P(X=i, Y=j) = e^{-\lambda p}\frac{(\lambda p)^i}{i!} \cdot e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}$$

Dacă vrem să determinăm rep. marginală:

$$P(X=i) = \sum_{j\geq 0} P(X=i, Y=j) = \sum_{j\geq 0} \frac{e^{-\lambda p}(\lambda p)^i}{i!} \cdot e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}$$

$$= e^{-\lambda p}\frac{(\lambda p)^i}{i!} \Rightarrow X \sim Pois(\lambda p)$$

Analog $Y \sim Pois(\lambda(1-p))$

În plus, $P(X=i, Y=j) = P(X=i)\cdot P(Y=j), \forall i,j$ adică $X \perp\!\!\!\perp Y$ (independente)

Media unei funcții de variabile aleatoare

X, Y v.a. discrete, $g: \mathbb{R}^2 \to \mathbb{R}^2$, $Z = g(X,Y)$

ex: a) $g(x,y) = 3x + 7y$

b) $g(x,y) = x\cdot y$

c) $g(x,y) = \max(x,y)$

$$\boxed{\mathbb{E}\{Z\} = \mathbb{E}\{g(X,Y)\} = \sum_x\sum_y g(x,y)\, p_{X,Y}(x,y)}$$

Exp:

rep comună →

X\Y	−1	0	2	Σ
1	1/18	3/18	2/18	6/18
2	2/18	0	3/18	5/18
3	0	4/18	3/18	7/18
Σ	3/18	7/18	8/18	

Vrem să determinăm rep X, Y, rep. cond $X|Y=0$, $Y|X=1$ și calculați media $\mathbb{E}\{XY\}$, $\mathbb{E}\{3X+2Y\}$

Rep. marginală a lui X: (suma liniei)

$$X \sim \begin{pmatrix} 1 & 2 & 3 \\ 6/18 & 5/18 & 7/18 \end{pmatrix}$$

Rep. marginală a lui Y (suma coloane

$$Y \sim \begin{pmatrix} -1 & 0 & 2 \\ 3/18 & 4/18 & 8/18 \end{pmatrix}$$

$X \backslash Y$	-1	0	2	Σ
1	1/18	3/18	2/18	6/18
2	2/18	0	3/18	5/18
3	0	4/18	3/18	7/18
Σ	3/18	7/18	8/18	1

Rep conditionată a lui $X|Y=0 \sim \begin{pmatrix} 1 & 2 & 3 \\ p_{X|Y}(1|0) & p_{X|Y}(2|0) & p_{X|Y}(3|0) \end{pmatrix}$

$$X|Y=0 \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{3/18}{7/18} & \frac{0}{7/18} & \frac{4/18}{7/18} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ \frac{3}{7} & 0 & \frac{4}{7} \end{pmatrix}$$

Analog $Y|X=1$,

$$Y|X=1 \sim \begin{pmatrix} -1 & 0 & 2 \\ \frac{1/18}{6/18} & \frac{3/18}{6/18} & \frac{2/18}{6/18} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 2 \\ \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{pmatrix}$$

$$\mathbb{E}[3X+2Y] = 3\,\mathbb{E}[X] + 2\,\mathbb{E}[Y]$$

$$g(x,y) = 3x+2y$$

$$\mathbb{E}[3X+2Y] = \sum_{(x,y)} (3x+2y)\,\mathbb{P}(X=x, Y=y)$$

$$\mathbb{E}[XY] = \sum_{(x,y)} xy\,\mathbb{P}(X=x, Y=y)$$
$$= 1\cdot(-1)\,1/18 + 0 + 2\cdot 2/18$$
$$= 2\cdot 2/18 + 0 + 4\cdot 3/18 +$$
$$0 + 0 + 6\cdot 3/18$$

Media conditionată

Fie X o v.a. discretă și $A \in \Omega$ cu $\mathbb{P}(A)>0$ atunci definim media condiționată a lui X la A

$$\boxed{\mathbb{E}[X|A] = \sum_{x} x\, p_{X|A}(x)}$$

Pentru o funcție $g(x)$ avem

$$\boxed{\mathbb{E}[g(X)|A] = \sum_{x} g(x)\, p_{X|A}(x)}$$

În particular, $A=\{Y=y\}$ avem media condiționată a lui X la $Y=y$.

$$\boxed{\mathbb{E}[X|Y=y] = \sum_{x} x\, p_{X|Y}(x|y)}$$

Exp: În cazul exemplului anterior vrem să calculăm $\mathbb{E}[X|Y=0]$ și $\mathbb{E}[Y|X=1]$.

Am văzut că $X|Y=0 \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{3}{7} & 0 & \frac{4}{7} \end{pmatrix}$

$$\mathbb{E}\{X \mid Y=0\} = 1 \cdot \frac{3}{7} + 2 \cdot 0 + 3 \cdot \frac{4}{7} = \frac{15}{7}$$

$$Y \mid X=1 \sim \begin{pmatrix} -1 & 0 & 2 \\ \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{pmatrix} \Rightarrow \mathbb{E}\{Y \mid X=1\} = -\frac{1}{6} + \frac{4}{6} = \frac{3}{6}$$

(P) Dacă X și Y v.a. discrete atunci:

$$\boxed{\mathbb{E}\{X\} = \sum_y \mathbb{E}\{X \mid Y=y\} \, \mathbb{P}(Y=y)}$$

Def: Fie X și Y două v.a. discrete. Media condiționată a lui X la Y, notată $\mathbb{E}\{X \mid Y\}$, este o variabilă aleatoare de forma $g(Y)$ pt care $g(y) = \mathbb{E}\{X \mid Y=y\}$, $\forall y$.

Exp: (continuare exp. precedent)

$$X \mid Y=0 \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{3}{7} & 0 & \frac{4}{7} \end{pmatrix} \text{ și } \mathbb{E}\{X \mid Y=0\} = \frac{15}{7}$$

$\frac{1}{18} : \frac{3}{18}$

$$X \mid Y=1 \sim \begin{pmatrix} 1 & 2 & 3 \\ 1/3 & 2/3 & 0 \end{pmatrix} \Rightarrow \mathbb{E}\{X \mid Y=-1\} = \frac{5}{3}$$

$$X \mid Y=2 \sim \begin{pmatrix} 1 & 2 & 3 \\ 2/8 & 3/8 & 3/8 \end{pmatrix} \Rightarrow \mathbb{E}\{X \mid Y=2\} = \frac{17}{8}$$

$$\mathbb{E}\{X \mid Y\} \sim \begin{pmatrix} 5/3 & 15/7 & 15/8 \\ \mathbb{P}(Y=-1) & \mathbb{P}(Y=0) & \mathbb{P}(Y=2) \end{pmatrix}$$

$$\mathbb{E}\{X \mid Y\} \sim \begin{pmatrix} 5/3 & 15/7 & 15/8 \\ 3/18 & 7/18 & 8/18 \end{pmatrix}$$

(P) Media mediei condiționate verifică:

$$\boxed{\mathbb{E}\{\mathbb{E}\{X \mid Y\}\} = \mathbb{E}\{X\}}$$

Obs: $\mathrm{Var}(X \mid A) = \mathbb{E}\{(X - \mathbb{E}\{X \mid A\})^2 \mid A\}$

$$= \mathbb{E}\{X^2 \mid A\} - \mathbb{E}\{X \mid A\}^2$$

$$\mathrm{Var}(X \mid Y=y) = \mathbb{E}\{X^2 \mid Y=y\} - \mathbb{E}\{X \mid Y=y\}^2$$

Cazul variabilelor aleatoare continue (Rep. comună, Rep. marginală, Rep. condiționată)

Fie $(\Omega, \mathcal{F}, \mathbb{P})$ un câmp de probabilitate și fie X și Y două v.a. continue (absolut continue). Spunem că perechea (X,Y) formează un cuplu de v.a. cont. dacă $\exists \, f_{(X,Y)}(x,y) \geq 0$ a.î.

$$\mathbb{P}((X,Y) \in A) = \iint_A f_{(X,Y)}(x,y)\,dx\,dy, \quad \forall A \subseteq \mathbb{R}^2$$
A interval

Funcția $f_{(X,Y)}$ (se mai notează $f_{X,Y}$) se numește densitatea comună a lui (X,Y).

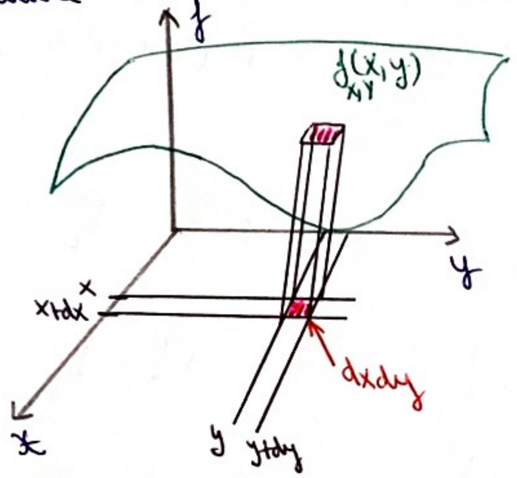
Dacă $A = [a,b] \times [c,d]$ atunci (A = dreptunghi)

$$P((X,Y) \in A) = P(a \le X \le b,\ c \le Y \le d)$$
$$= \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx$$

Dacă în plus $A = \mathbb{R}^2$ atunci

$$P((X,Y) \in \mathbb{R}^2) = 1 \Rightarrow \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

Interpretare:



$$P(X \in (x, x+dx),\ Y \in (y, y+dy)) = \int_x^{x+dx}\int_y^{y+dy} f_{X,Y}(u,v)\,dv\,du \simeq$$
$$\simeq f_{X,Y}(x,y)\,\underbrace{dx\,dy}_{\text{unitatea de arie}}$$

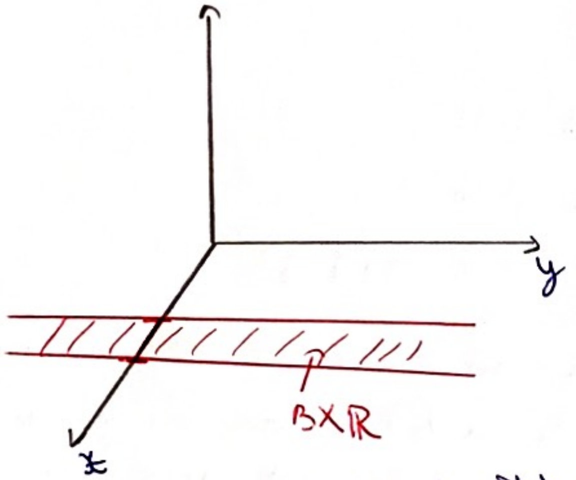
$$f_{X,Y}(x,y) \simeq \frac{P(X \in (x,x+dx),\ Y \in (y, y+dx))}{dx\,dy}$$

$$\simeq \frac{\text{probabilitatea}}{\text{unitatea de arie}}$$

Densitatea $f_{X,Y}(x,y)$ conține toată informația despre X și Y, cunoașterea ei permite calculul oricărei prob. ce implică X și Y: $P((X,Y) \in A)$

În particular, putem calcula și:

$$P(X \in B) = P((X,Y) \in B \times \mathbb{R}) = \int_B \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy\,dx$$



Dacă X este cont și admite densitatea f_X atunci

$$P(X \in B) = \int_B f_X(x)\,dx$$

astfel $\int_B f_X(x)\,dx = \int_B \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy\,dx$

Densitatea marginală a lui X:

$$\boxed{f_X(x)=\int_{\mathbb{R}} f_{X,Y}(x,y)\,dy}$$

Obs: În cazul v.a. discrete:

$$p_X(x)=\sum_y p_{X,Y}(x,y)$$

Analogia dintre discret și cont.: $\left\{\begin{array}{l}\sum \to \int \\ p_{X,Y}(x,y)\to f_{X,Y}(x,y)\,dy\end{array}\right.$

În mod similar, densitatea marginală a lui Y:

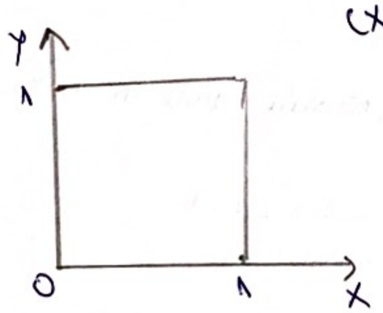
$$\boxed{f_Y(y)=\int_{\mathbb{R}} f_{X,Y}(x,y)\,dx}$$

Exp: Presupunem că Ionel și Maria își dau întâlnire într-o zi specifică la ora 12 și ajung la punctul de întâlnire cu o întârziere (aleatoare) care variază între 0 și 1h.

Notăm cu X și resp Y, durata de întârziere a lui Ionel și resp a Mariei.

Ne interesăm la densitatea comună a lui (X, Y).

$(X,Y)\in[0,1]^2$



$$f_{X,Y}(x,y)=\begin{cases} c, & (x,y)\in[0,1]^2 \\ 0, & \text{altfel}\end{cases}$$

Dacă $f_{X,Y}(x,y)$ este densitate $\Rightarrow$ $c\geq 0$ și

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy=1$$

adică $\int_0^1\int_0^1 c\,dx\,dy = c = 1$

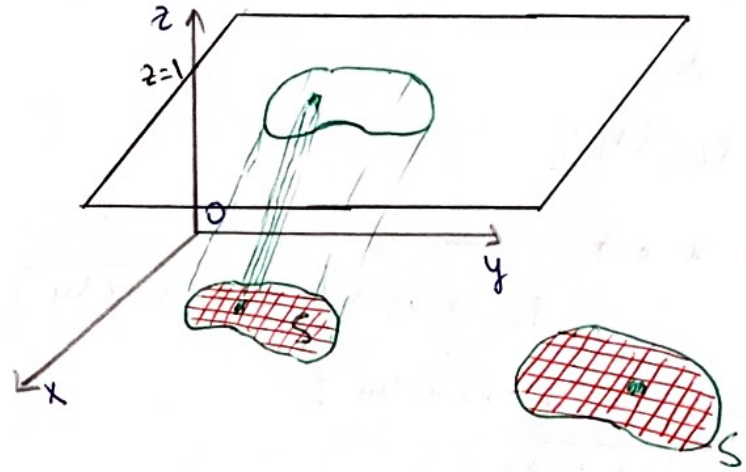
În general, dacă $S\subseteq\mathbb{R}^2$ (triunghi, drept, ...) atunci definim densitatea uniformă pe S, $(X,Y)\sim\mathcal{U}(S)$,

$$f_{X,Y}(x,y)=\begin{cases} c, & (x,y)\in S \\ 0, & \text{altfel}\end{cases}$$

Cum găsim c?

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy=1 \Rightarrow \iint_{\mathbb{R}^2} c\cdot \mathbb{1}_S(x,y)\,dx\,dy=1$$

adică $c \iint_S 1\,dx\,dy = 1$



$$\iint_S 1\,dx\,dy = Aria(S)$$

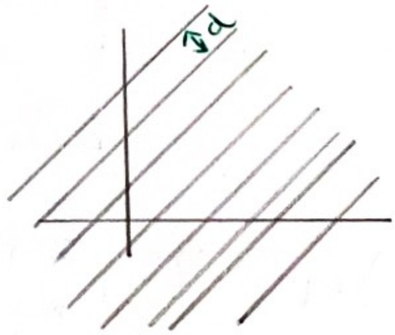
$$\Rightarrow c = \frac{1}{Aria(S)} \Rightarrow f_{X,Y}(x,y) = \begin{cases} 1/Aria(S), & (x,y)\in S \\ 0, & altfel \end{cases}$$

$$\mathbb{P}((X,Y)\in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

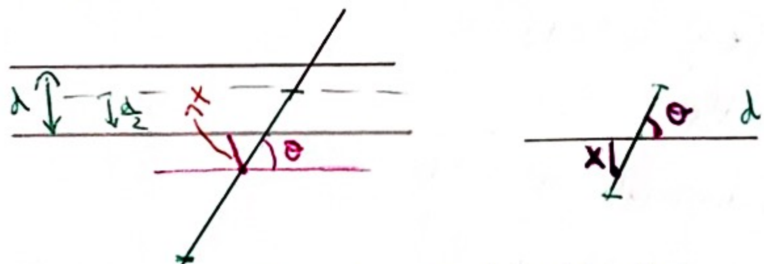
$$= \frac{\iint_A 1_S(x,y)\,dx\,dy}{Aria(S)} = \frac{Aria(A\cap S)}{Aria(S)}$$

Exp: (Problema acului lui Buffon)

O suprafață marcată cu linii paralele aflate la distanța d una față de cealaltă.



Să presupunem că aruncăm un ac de lung. $l<d$.
Care este prob. ca acul să intersecteze una dintre linii?



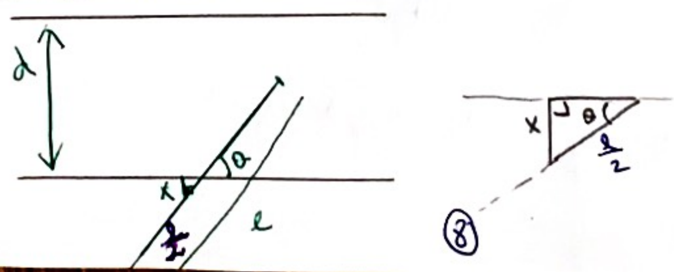
Putem modela poziția acului în funcție de:

- θ unghiul ascuțit format de axa acului cu dreptele paralele
- X distanța de la mijlocul acului la cea mai apropiată dreaptă paralelă

Poziția o vom modela prin (X,θ). Putem presupune că $(X,\theta) \sim \mathcal{U}(S)$ unde

$$S = \left\{(x,\theta) \,\middle|\, \begin{matrix} 0\le x\le d/2 \\ 0\le\theta\le \pi/2 \end{matrix}\right\}$$

Condiția ca acul să intersecteze o linie:



Pt θ fixat

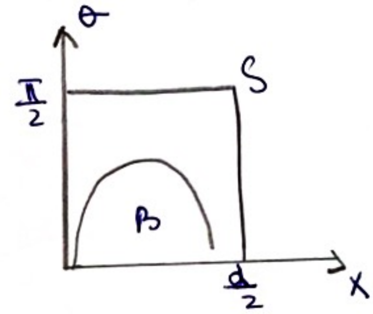
$X \le \frac{l}{2}\sin\theta$

Vrem să calculăm $P(X \le \frac{l}{2}\sin\theta) = ?$

$\underbrace{X,\ \theta}_{v.a.}$

$$P(X \le \frac{l}{2}\sin\theta) = P((X,\theta)\in B), \text{ unde } B = \{(x,\theta)\in S \mid x \le l/2 \sin\theta\}$$

$$= \iint_B f_{X,\theta}(x,\theta)\,dx\,d\theta$$



$$P(X \le \frac{l}{2}\sin\theta) = \iint_B \frac{1}{Aria(S)} 1_S(x,\theta)\,dx\,d\theta$$

$$= \iint_B \frac{4}{\pi d} 1_{[0,\frac{d}{2}]\times[0,\frac{\pi}{2}]}(x,\theta)\,dx\,d\theta$$

$$= \int_0^{\frac{\pi}{2}}\int_0^{l/2\sin\theta} \frac{4}{\pi d}\,dx\,d\theta = \frac{4}{\pi d}\int_0^{\frac{\pi}{2}} l/2\sin\theta\,d\theta = \frac{2l}{\pi d}$$

Def: X și Y două v.a. cont pe $(\Omega, \mathcal{F}, P)$ și $f_{X,Y}(x,y)$ este densitatea comună a (X,Y). Definim funcția de repartiție a (X,Y)

$$\boxed{F_{X,Y}(x,y) = P(X\le x, Y\le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv}$$

Obs: Densitatea comună se determină:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y}F_{X,Y}(x,y)$$

<u>Rep. condiționate</u>

Fie $(\Omega, \mathcal{F}, P)$ c.p., X v.a. cont și $A\in\mathcal{F}$, $P(A)>0$

Definim densitatea cond. a lui X la A, funcția $f_{X|A}(x)\ge 0$ care verifică relația:

$$P(X\in B\mid A) = \int_B f_{X|A}(x)\,dx, \ \forall B\subseteq \mathbb{R}$$

Dacă $B=\mathbb{R}$ atunci $P(X\in\mathbb{R}\mid A)=1$ astfel:

$$\int_{\mathbb{R}} f_{X|A}(x)\,dx = 1 \Rightarrow f_{X|A}(x) \text{ este o densitate de prob.}$$

Putem presupune că $A \to \{X\in A\}$ (A-urile nu coincid) și $P(X\in A)>0$

$$P(X\in B\mid X\in A) = \frac{P(X\in B, X\in A)}{P(X\in A)} = \frac{P(X\in A\cap B)}{P(X\in A)}$$

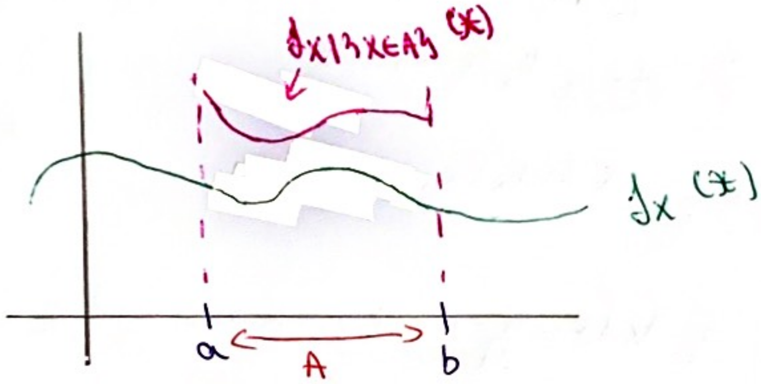
Dacă X admite densitatea de rep f_X atunci

$$P(X\in B\mid X\in A) = \frac{\int_{A\cap B} f_X(x)\,dx}{P(X\in A)}$$

Pe de altă parte,

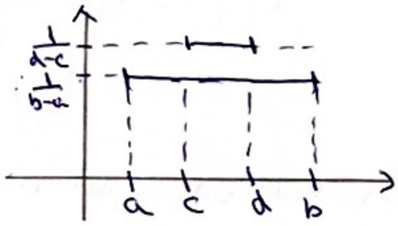
$$\mathbb{P}(X\in B \mid X\in A) = \int_B f_{X\mid\{X\in A\}}(x)\,dx \quad \Big\} \Rightarrow$$

$$\Rightarrow f_{X\mid\{X\in A\}}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}(X\in A)}, & x\in A \\ 0, & \text{altfel} \end{cases}$$



Exp: $X\sim \mathcal{U}([a,b])$ și $[c,d]\subseteq[a,b]$

$$f_{X\mid X\in[c,d]}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}(X\in[c,d])}, & x\in[c,d] \\ 0, & \text{altfel} \end{cases}$$



$$f_X(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$$

$$\mathbb{P}(X\in[c,d]) = \int_c^d f_X(x)\,dx = \frac{d-c}{b-a}$$

$$f_{X\mid X\in[c,d]}(x) = \frac{\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)}{\frac{d-c}{b-a}}\mathbb{1}_{[c,d]}(x) = \frac{1}{d-c}\mathbb{1}_{[c,d]}(x)$$

adică $X\mid X\in[c,d]\sim\mathcal{U}[c,d]$