Media și momentele variabilelor aleatoare

Repetăm un experiment de N ori și urmărim rezultatul unei v.a. X de interes, e.g binomial → aruncăm cu banul de 10 ori și fie X v.a care ne dă nr de H în cele 10 aruncări.

La sf. celor N repetiții avem $x_1, \dots, x_N$
$1, 2, 2, \dots, 7$ $(N = 1000)$

Media aritmetică a acestor valori:

$$m = \frac{\sum_{i=1}^{N} x_i}{N}$$

Exp: $N = 8$: $\underset{x_1}{1}, \underset{x_2}{1}, 1, 2, 3, 3, 4, \underset{x_8}{5}$

$$m = \frac{1+1+1+2+3+3+4+5}{8} = \frac{20}{8}$$

$$m = \frac{3\times 1 + 1\times 2 + 2\times 3 + 1\times 4 + 1\times 5}{8}$$

Dacă v.a. X discretă cu fct de masă $f(x) = P(X = x)$ atunci am văzut că $f(x) \simeq \frac{N(x)}{N}$ → nr de ori în care ev $\{X = x\}$ s-a realizat

$\Rightarrow N(x) \simeq f(x) \cdot N$

$$m \simeq \frac{1}{N} \sum_x x N(x) = \frac{1}{\cancel{N}} \sum_x x f(x) \cancel{N} = \sum_x x f(x)$$

suma ponderată

Def: Fie X o v.a discretă. Definim media lui X prin:

$$\boxed{E[X] = \sum_x x f(x)}$$ unde $f(x)$ e fct de masă

ori de câte ori seria $\sum_x |x| f(x) < \infty$

Dacă seria asta e divergentă atunci spunem că media nu e definită

Exp: Dacă X v.a. care reprezintă rezultatul aruncării unui zar $\Rightarrow X \in \{1,2,3,4,5,6\}$

$f(x) = 1/6$

$E[X] = \frac{1}{6}(1+2+3+4+5+6) = 3.5.$

Exp: Dacă $X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$ atunci

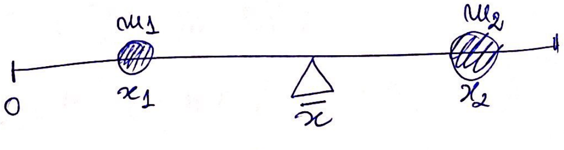
$E[X] = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n$

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

$E[X] = (-2) \times 1/4 + (-1) \times 1/8 + 1 \times 1/2 + 3 \times 1/8$

$= \frac{-4-1+4+3}{8} = \frac{1}{4}$

Obs: Interpretare fizică $\Rightarrow$ centrul de greutate (masă) a unui sistem finit de corpuri.



$\bar{x} M = m_1 x_1 + m_2 x_2$

$\bar{x} = \frac{m_1}{M} x_1 + \frac{m_2}{M} x_2$

unde $M = m_1 + m_2$

Proprietăți ale mediei:

a) Dacă X este o v.a. constantă $X=c$ atunci: $\mathbb{E}[X]=c$

b) Dacă X este o v.a pozitivă, $X \geq 0$ atunci $\mathbb{E}[X] \geq 0$

c) Dacă X și Y sunt 2 v.a. a.î. $X \geq Y$. Atunci $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

d) Dacă X și Y sunt 2 variabile aleatoare discrete și $a, b \in \mathbb{R}$ atunci $\boxed{\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y].}$

e) Fie $A \subseteq \Omega$ un eveniment și $\mathbb{1}_A(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \text{altfel}\end{cases}$

Atunci $\mathbb{E}[\mathbb{1}_A]=\mathbb{P}(A)$.

f) Dacă X și Y sunt 2 v.a independente atunci

$$\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$$

Obs: În general, $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$

(P) Dacă X e o v.a discretă și $g: \mathbb{R}\to\mathbb{R}$ o fct reală atunci $Y=g(X)$ este o v.a discretă de medie

$$\mathbb{E}[g(X)]=\sum_x g(x)\mathbb{P}(X=x)$$

Exp: $X \sim \begin{pmatrix} -2 & -1 & 1 & 3 \\ 1/4 & 1/8 & 1/2 & 1/8 \end{pmatrix}$ $Y=X^2$

Met 1:

$Y \in \{1, 4, 9\}$

$Y \sim \begin{pmatrix} 1 & 4 & 9 \\ 5/8 & 1/4 & 1/8 \end{pmatrix}$

$$\mathbb{E}[Y]=1\times 5/8+4\times 1/4+9\times 1/8=\frac{22}{8}$$

Met 2:

$g(x)=x^2$

$$\mathbb{E}[Y]=\mathbb{E}[X^2]=\sum_{x\in\{-2,-1,1,3\}} x^2\,\mathbb{P}(X=x)=22/8$$

Def: Fie X o v.a discretă. Numim moment de ordin k media $E[X^k]$ și moment centrat de ordin k în a media $E[(X-a)^k]$.

Dacă $a = E[X]$ atunci $E[(X-E(X))^k]$ se numește moment centrat de ordin k.

Def: Momentul centrat de ordin 2 se numește varianța sau dispersia v.a X și se notează

$$\boxed{Var(X) = E[(X-E[X])^2]}$$

Obs: Varianța este o măsură care surprinde gradul de împrăștiere al datelor în jurul mediei.

Definim abaterea standard a unei v.a X prin

$$SD(X) = \sqrt{Var(X)}$$

standard deviation

Obs: Dacă X e o măsură în u.m. atunci $Var(X)$ se măsoară în $(u.m)^2$ iar $SD(X)$ în u.m.

Prop ale varianței:

Dacă X este o v.a const atunci $Var(X)=0$

a) Dacă X este o v.a (discretă), atunci $Var(X) \geq 0$

b) Dacă X este o v.a și $a \in \mathbb{R}$ atunci $\boxed{Var(X+a) = Var(X)}$

c) Dacă X este o v.a și $a \in \mathbb{R}$ atunci

d) Dacă X este o v.a.

$$\boxed{Var(aX) = a^2 Var(X)}$$

Obs: $a=-1 \Rightarrow Var(-X) = Var(X)$

Obs: $a, b \in \mathbb{R}$ atunci $\boxed{Var(aX+b) = a^2 Var(X)}$

e) Dacă X este o v.a (discretă) atunci

$$\boxed{Var(X) = E[X^2] - E[X]^2}$$

7) Dacă X și Y sunt 2 v.a. independente atunci:

$$\boxed{Var(X+Y) = Var(X) + Var(Y)}$$

Calculul mediei și al varianței pt repartiții cunoscute

1) Bernoulli $X \sim B(p)$ $X \in \{0, 1\}$ $P(X=1) = p$

$\Rightarrow P(X=0) = 1-p$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} \Rightarrow \boxed{E[X] = p}$$

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

$$\boxed{Var[X] = p - p^2 = p(1-p)}$$

2) Rep binomială $X \sim B(n,p)$

$X \in \{0, 1, 2, \dots, n\}$ $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$E[X] = \sum_{x \in \{0..n\}} x \, P(X=x) = \sum_{k=0}^{n} k \, P(X=k)$$

$$= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np$$

Reamintim că $X = X_1 + \ldots + X_n$ unde $X_i \sim B(p)$ iar $X_1, X_2, \ldots, X_n$ sunt indep.

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + \ldots + X_n] \overset{\text{liniaritate}}{=} \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]$$

$$\overset{\text{simetrie}}{=} n\,\mathbb{E}[X_1] = np$$

$$Var(X) = Var(X_1 + \ldots + X_n) \overset{\text{indep}}{=} Var(X_1) + \ldots + Var(X_n)$$

$$\overset{\text{simetrie}}{=} n\,Var(X_1) = np(1-p)$$

3.) <u>Rep Hipergeometrică</u> $X \sim HG(n, N, M)$

N → nr total de bile

M → bile negre

n → nr bile extrase fără întoarcere

$$\mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Fie X_j v.a care ia val 1 dacă bila j extrasă este aleatoare neagră și 0 altfel.

Atunci $X_j \sim B(p)$ cu $p = \frac{M}{N}$. Mai mult, X este

$$X = X_1 + X_2 + \ldots + X_n$$

Var a $X_1, X_2, \ldots, X_n$ <u>nu sunt independente</u>:

$$\boxed{\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = n\,\mathbb{E}[X_1] = n\frac{M}{N}}$$

Scanned with CamScanner

4) Rep Poisson $\quad X \sim Pois(\lambda), \quad X \in \{0,1,2,\dots\}$

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

$$\boxed{E[X]} = \sum_{k=0}^{\infty} k\, P(X=k) = \sum_{k=0}^{\infty} k\, e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k\, e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^{\ell+1}}{\ell!} = \lambda e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^\ell}{\ell!}$$

$$\boxed{=\lambda}$$

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

5) Rep Geometrică $\quad X \sim Geom(p)$

$X \in \{1,2,3,\dots\}$

$$P(X=k) = (1-p)^{k-1} p \qquad \text{Notăm } q = 1-p$$

$$\boxed{E[X]} = \sum_{k=1}^{\infty} k\, P(X=k) = \sum_{k=1}^{\infty} k\, q^{k-1} p = p\sum_{k=1}^{\infty} k\, q^{k-1}$$

$$= p\sum_{k=1}^{\infty}(q^k)' \longleftarrow \text{derivarea se face după } q$$

$$= p\left(\sum_{k=1}^{\infty} q^k\right)' = p\frac{d}{dq}\left(\sum_{k=0}^{\infty} q^k - 1\right)$$

$$= p\frac{d}{dq}\left(\frac{1}{1-q} - 1\right) = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}$$

$$\boxed{Var(X) = \frac{q}{p^2}}$$

Scanned with CamScanner

Variabile aleatoare continue (absolut continue)

Def: Fie $(\Omega, \mathcal{F}, P)$ un câmp de probabil și $X: \Omega \to \mathbb{R}$ o v.a.
Spunem că v.a. X este continuă (absolut continuă) dacă există o fct pozitivă $f(x) \geq 0$ cu prop că

$$P(X \in A) = \int_A f(x)dx, \quad \forall A \subseteq \mathbb{R}$$

(intervale sau reuniuni cel mult numărabile de intervale)

Funcția f se numește densitate de repartiție.

Reamintim: Dacă X v.a discretă atunci pt $A \subset \mathbb{R}$

$$P(X \in A) = \sum_{x \in X(\Omega) \cap A} P(X = x) = \sum_{x \in X(\Omega) \cap A} f(x)$$

fct de masă

Obs: $A = [a, b]$ atunci

$$P(X \in A) = P(a \leq X \leq b) = \int_{[a,b]} f(x)dx = \int_a^b f(x)dx$$

dacă $A = \mathbb{R}$ atunci $P(X \in \mathbb{R}) = P(\Omega) = 1$

$$P(X \in \mathbb{R}) = \int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{+\infty} f(x)dx$$

astfel $\boxed{\int_{-\infty}^{+\infty} f(x)dx = 1}$

(P) Densitatea de repartiție f a unei v.a continue trebuie să satisfacă urm. proprietăți:

a) $f(x) \geq 0, \ \forall x \in \mathbb{R}$

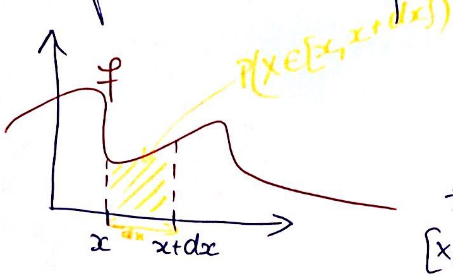
b) $\int_{-\infty}^{+\infty} f(x)dx = 1$

Obs. În def v.a continue dacă luăm $A = \{a\}$

atunci $\boxed{P(X = a) = \int_A f(x)dx = \int_a^a f(x)dx = 0}$

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X < b)$$

Interpretare: De ce f este densitate?



Vrem să calculăm
$\mathbb{P}(X \in [x, x+dx])$

$= \int_{[x, x+dx]} f(t)\,dt = \int_x^{x+dx} f(t)\,dt$

Dacă dx e f mic:
$$\mathbb{P}(X \in [x, x+dx]) \simeq f(x)(x+dx-x) = f(x)dx$$

Astfel
$$\underset{\text{densitate}}{f(x)} \simeq \frac{\mathbb{P}(X \in [x, x+dx])}{dx} \quad \left(\frac{\text{probabilitatea}}{\text{unit. de lungime}}\right)$$

Obs: Dacă X este măsurat în m (u.m) atunci f e ~~va~~ măsurată în $\frac{1}{m}$ (1/u.m).