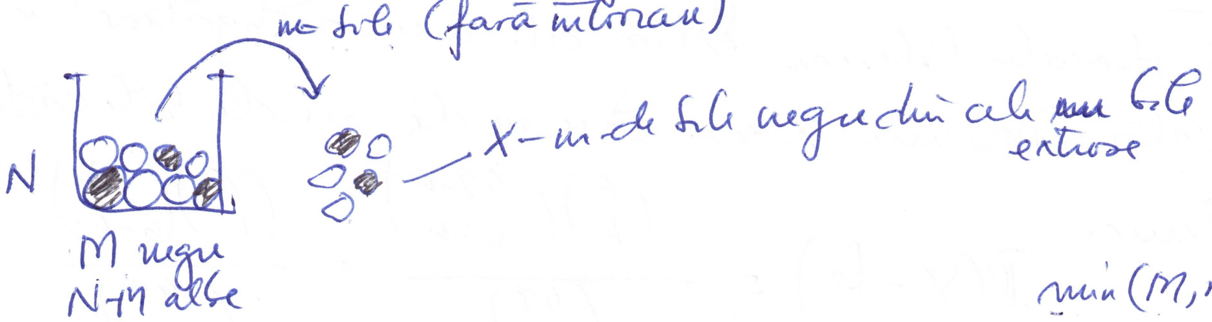
Curs 7.

5) V.a. repartizate hipergeometric

Să presupunem că avem o urnă cu N bile dintre care M sunt de culoare neagră și celelalte N-M sunt albe.

Considerăm v.a. X dată de nr. de bile negre extrase atunci când din urnă am extras n bile fără întoarcere.



Notația: $X \sim HG(n, N, M)$

Vrem să determinăm $\mathbb{P}(X=k) = ?$, $k \in \{0, 1, \dots, \overset{\min(M,n)}{n}\}$

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

$$\mathbb{P}(X=k) \geq 0$$

$$\sum_{k=0}^{\min(M,n)} \mathbb{P}(X=k) = 1$$

Dacă $n \leq M$ atunci $\sum_{k=0}^{n} \mathbb{P}(X=k) = 1$ sau altfel scris

$$\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$$ ← identitatea Vandermonde

Idee: ne uităm la ~~termenul~~ identitatea $(1+x)^M(1+x)^{N-M} = (1+x)^N$ și identificăm coeficientul lui x^n

În cazul în care $Y \sim B(n, \frac{M}{N})$ și $X \sim HG(n, N, M)$ diferența apare în modul de desfășurare a 'experimentului', mai exact în modul în care extragem bilele din urnă:

a) binomial - extragere cu întoarcere

b) hipergeometric - extragere fără întoarcere

Exp: Să considerăm jocul de loto 6 din 49 și presupunem că oficialii loteriei extrag cele 6 nr. câștigătoare în mod aleator. Definim X v.a care ne dă nr. de bile câștigătoare atunci

$$\mathbb{P}(X=k) = \frac{\binom{6}{k}\binom{49-6}{6-k}}{\binom{49}{6}} = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$$

$$X \sim HG(\underset{n}{6}, \underset{N}{49}, \underset{M}{6})$$

Dacă $k=6 \Rightarrow \mathbb{P}(X=6) = \frac{1}{\binom{49}{6}}$

Exp2: (Cum să estimăm nr. de pești dintr-un lac)

Un lac cu $N \geq 1$ pești, N-necunoscut. Mergem la pescuit și prindem $r \geq 1$ pești pe care îi marcăm și apoi îi eliberăm în lac. Adoua zi mergem iar la pescuit și să presupunem că am prins $n \geq 1$ pești dintre care $k \geq 1$ erau marcați.

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

Prin metoda verosimilității maxime găsim că $\hat{N} = \left[\frac{nr}{k}\right]$

6) Repartiția discretă uniformă

$D = \{x_1, x_2, \ldots x_n\}$, $X : \Omega \longrightarrow D \subseteq R$

Spunem că v.a X este repartizată uniform pe D, și scriem $X \sim \mathcal{U}(D)$, dacă

$$\mathbb{P}(X = x) = \frac{1}{|D|} \quad , x \in D$$

Pentru $A \subseteq D$ avem că

$$\mathbb{P}(X \in A) = \frac{|A|}{|D|}$$

Exp: Să presupunem că avem o urnă cu bile numerotate de la 1 la 100. Extragem 5 bile, una după alta:

1) Extragere cu întoarcere

a) Care este rep. v.a care ne dă nr. bilelor ce au o valoare inscripționată ≥ 80?

b) Cum este repartizată v.a care ne dă a j-a extragere $(1 \leq j \leq 5)$?

c) Care este probabilitatea ca nr. 100 să fie extras al patrulea oară?

Sol: a) Prob. să extragem o bilă cu val. ≥ 80 este $\frac{21}{100} = p$

$X \sim B(5, 0.21)$

b) X_j valoarea la a j-a extragere

$X_1 \sim \mathcal{U}(\{1, 2, \ldots 100\})$

$X_j \sim \mathcal{U}(\{1, 2, \ldots 100\})$ $\quad \mathbb{P}(X_j = k) = 1/100$

c) $\mathbb{P}(\{X_1=100\}\cup\{X_2=100\}\cup\{X_3=100\}\cup\{X_4=100\}\cup\{X_5=100\}) =$

$= 1-\mathbb{P}(\{X_1\neq 100\}\cap\{X_2\neq 100\}\cap\dots\cap\{X_5\neq 100\})$

$= 1-\left(\frac{99}{100}\right)^5$

2) Extragerile se fac fără întoarcere

a) $Y \sim HG(5,100,21)$

b) Y_j - valoarea de la a j-a extragere

$Y_1 \sim \mathcal{U}(\{1,2,\dots 100\})$

$Y_j \sim \mathcal{U}(\{1,2,\dots 100\})$

c) $\mathbb{P}\left(\bigcup_{j=1}^{5}\{Y_j=100\}\right)$

Evenimentele $\{Y_1=100\}$ și $\{Y_2=100\}$ sunt disjuncte pt că extragerea se face fără întoarcere, prin urmare $\{Y_j=100\}$ sunt disjuncte 2 câte 2.

$\mathbb{P}\left(\bigcup_{j=1}^{5}\{Y_j=100\}\right) = \mathbb{P}(Y_1=100)+\dots+\mathbb{P}(Y_5=100)$

$= 5/100$

7) v.a repartizată geometric și negativ binomial

Aruncăm în mod repetat cu o monedă a. î. șansa de succes (să obținem cap) este p

V.a X să reprezinte nr. de aruncări până obținem, și incluzînd, primul succes

V.a X este o variabilă aleatoare repartizată geometric de parametru p și notăm $X \sim Geom(p)$

$$X \in \{1,2,3,\dots\} = \mathbb{N}^*$$

$$\mathbb{P}(X=k) = (1-p)^{k-1}p \quad , k \geq 1$$

$$\underbrace{T\ T\ T\ T}_{k-1}\ \underset{k}{H}$$

$\mathbb{P}(X=k) \geq 0$ și

$$\sum_{k\geq 1}^{\infty} \mathbb{P}(X=k) = \sum_{k\geq 1}^{\infty} (1-p)^{k-1}p = p\sum_{i\geq 0}^{\infty} (1-p)^i = p\frac{1}{1-(1-p)} = 1$$

Obs: Există și o versiune alternativă a acestei def: v.a Y reprezintă nr. de eșecuri înaintea primului succes

$$Y \in \{0,1,2,\dots\}$$

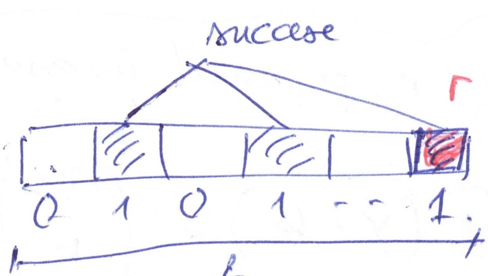
$$Y = X-1. \qquad \mathbb{P}(Y=k) = (1-p)^k p \quad , k \geq 0.$$

Obs: V.a $X \sim Geom(p)$ verifică prop. lipsei de memorie:

$$\mathbb{P}(X = n+k \mid X \geq n) = \mathbb{P}(X=k)$$

Def: V.a. X dată de nr. de aruncări până obținem pt prima oară $r \geq 1$ succese este o variabilă aleatoare rep. negativ binomial și notăm $X \sim NB(r,p)$

$$\mathbb{P}(X=k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$$



prob. să obs. r secv. dată de lung. k în care avem r val de 1 și $k-r$ de 0

(P) Dacă $X \sim NB(r,p)$ atunci putem scrie
$$X = X_1 + X_2 + \dots + X_r, \quad X_i \sim Geom(p) \text{ indep.}$$

8) v.a Poisson

Spunem că v.a X este rep. Poisson de parametru $\lambda > 0$, și scriem $X \sim \mathcal{P}(\lambda)$ (sau $X \sim Pois(\lambda)$) dacă

$$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \{0,1,\dots\}$$

Utilizare:
- modelarea într-un timp specific a sosirilor clienților la ghișeul unui magazin.
- nr. de mașini într-o intersecție (într-un anumit interval de timp)
- nr. de cazuri de boală (rară) dintr-o anumită regiune
- nr. de cuvinte scrise eronat într-o carte
- nr. de email-uri într-un minut

$\mathbb{P}(X=k) \geq 0, \quad k \geq 0$

$$\sum_{k=0}^{\infty} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Aproximarea binomialei prin Poisson:

Fie n, p astfel în așa fel încât $np \to \lambda$ și $X \sim B(n,p)$

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \left(p \simeq \frac{\lambda}{n}\right)$$

$$= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \cdot \frac{1}{n^k} \left(1-\frac{\lambda}{n}\right)^n \cdot \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\frac{\lambda}{n} \to 0 \Rightarrow \left(1-\frac{\lambda}{n}\right)^n \xrightarrow[(1)^\infty]{n\to\infty} \left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda} \Rightarrow e^{-\lambda}$$

$$\frac{n!}{(n-k)!} \cdot \frac{1}{n^k} = \frac{(n-k+1)\cdots(n-1)\cdot n}{n^k} = \frac{n-k+1}{n} \times \frac{n-k+2}{n} \times \cdots \frac{n-1}{n} \times 1$$

pt k fixat avem $\frac{n!}{(n-k)!} \cdot \frac{1}{n^k} \longrightarrow 1.$

$$\frac{\lambda}{n} \to 0 \Rightarrow 1-\frac{\lambda}{n} \to 1 \Rightarrow \left(1-\frac{\lambda}{n}\right)^{-k} \longrightarrow 1.$$

Prin urmare

$$\boxed{\mathbb{P}(X=k) \underset{\substack{n\to\infty \\ np\to\lambda}}{\simeq} e^{-\lambda}\frac{\lambda^k}{k!}}$$

Obs: Folosim această aproximare atunci când n mare, p mic și $np \simeq \lambda$

Exp: $n = 10^6$ potențiali clienți, $p = 2\times 10^{-6}$ din istoric
atunci probabilitatea să avem cel puțin 3 clienți

$\mathbb{P}(X \geq 3) = ?$ $X \sim \mathcal{B}(n,p)$

$$\mathbb{P}(X\geq 3) = 1 - \mathbb{P}(X<3)$$
$$\simeq 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) - \mathbb{P}(X=2)$$
$$\simeq 1 - e^{-2} - 2e^{-2} - \frac{2^2}{2!}e^{-2} = 1 - 5e^{-2}$$

$$\mathbb{P}(X=k) \simeq e^{-\lambda}\frac{\lambda^k}{k!}, \quad \lambda = 2 \quad (\underbrace{2\times 10^{-6}}_{p} \times \underbrace{10^6}_{n})$$

Funcții de v.a. discrete

X v.a. poate suntem interesați de e^X sau $\sin(X)$

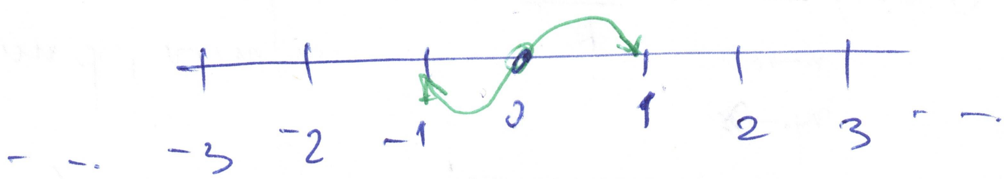
Def: Considerăm $(\Omega, \mathcal{F}, \mathbb{P})$ câmp de prob, $X:\Omega\to\mathbb{R}$ o v.a. (discretă) și $g:\mathbb{R}\to\mathbb{R}$ o funcție. Atunci $g(X):\Omega\to\mathbb{R}$ este o v.a.

$$\begin{tikzcd} \Omega \arrow[r, "X"] \arrow[rr, bend right, "g\circ X"'] & \mathbb{R} \arrow[r, "g"] & \mathbb{R} \end{tikzcd}$$

Dacă X este discretă atunci $g(X)$ este discretă

Exp: (Mersul la întâmplare)

O particulă se mișcă n pași pe dreapta numerelor începând din poziția 0.

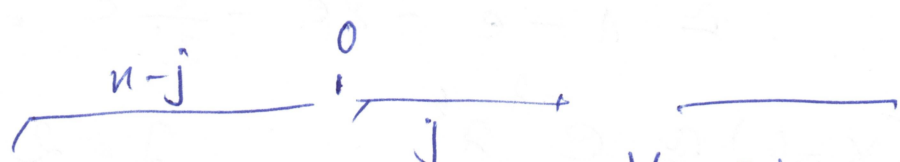


La fiecare pas particula se mișcă spre dreapta sau spre stânga cu prob. $1/2$. Fie Y poziția după n pași.

Vrem să calculăm $\mathbb{P}(Y=k)$

Sol: Nr. de pași pozitivi (spre dreapta), X, este o v.a. rep $B(n, 1/2)$

Dacă $X=j$ atunci avem j pași spre dreapta și $n-j$ pași spre stânga



deci poziția este: $j-(n-j)=2j-n \Rightarrow Y=2X-n$

$$\mathbb{P}(Y=k)=\mathbb{P}(2X-n=k)=\mathbb{P}\left(X=\frac{n+k}{2}\right)$$