

Curs 5: Distribuții clasice de probabilitate

Conf.dr. Maria Jivulescu

Departamentul de Matematică
UPT



O varb. aleatoare este descrisă prin tabelul de distribuție

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}, \quad \sum_{i=1}^n p_i = 1$$

unde: x_i -valorile lui X , iar , $p_i = P(X = x_i), 0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1$

sau p.m.f.(probability mass function)

$$p(x) = \begin{cases} p_1, & \text{daca } x = x_1 \\ p_2, & \text{daca } x = x_2 \\ \dots & \\ p_n, & \text{daca } x = x_n \\ 0, & \text{in rest} \end{cases}$$

sau c.d.f.(cumulative distrb. function=funcția de repartitie)

$$F_X(x) := P(X \leq x)$$

Există tabele de distribuție ce pot descrie exp. aleatoare de un anumit tip?

- Distribuția Bernoulli
- Distribuția binomială
- Distribuția geometrică
- Distribuția Poisson

Experiment Bernoulli = aruncarea monedei: succesul fiind, de exemplu apariția feței ban, iar eșecul: apariția stemei.

- fiecare încercare are doar două rezultate mutual exclusive, unul numit succes (notat cu 1) și celălalt eșec (notat cu 0);
- X - variabila aleatoare ce înregistrează rezultatul unei încercări;
- p - probabilitatea succesului ($0 < p < 1$), iar q -prob eșecului

$$p = P(X = 1), \quad q = 1 - p = P(X = 0);$$

$$\text{■ } X = \begin{pmatrix} 1 & 0 \\ p & 1 - p \end{pmatrix} \text{ sau } p(x) = \begin{cases} p, & \text{daca } x = 1 \\ 1 - p, & \text{daca } x = 0 \\ 0, & \text{in rest} \end{cases}$$

- $M(X) = p; \sigma^2(X) = p(1 - p)$.
- notație: $X \sim \text{Bernoulli}(p)$
- în cazul aruncării monedei: $X \sim \text{Bernoulli}(1/2)$

Distribuția Bernoulli se folosește în generarea de șiruri de biți aleatori.

Run-uri de biți=o succesiune de biți identici într-un șir de n biți rezultați din simularea de n ori a unei variabilei aleatoare X distribuită Bernoulli.

De exemplu în șirul de biți 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, avem următoarele run-uri de biți 1:

11, 1, 111

În analiza șirurilor de biți aleatori folosiți în probleme de securitate (criptografie) este foarte util numărul mediu de run-uri de biți din șir. Se poate demonstra¹

$$M(N) = p + (n - 1)(1 - p)p$$

¹vezi notite curs 5

Experiment Bernoulli repetat de n ori \Rightarrow **variabila aleatoare binomială**

- O încercare este o etapă a experimentului, ce are două rezultate mutual exclusive: unul numit *succes* și celălalt *eșec*.
- Încercările sunt independente;
- Pentru fiecare încercare probabilitatea succesului este aceeași, p ;

Exemplul clasic: aruncarea de n ori a unei monede;

- succesul: cade banul (B), eșecul, cade stema (S).
- $p = 1/2$
- X - înregistrează numărul de fețe=ban ce se obțin din n aruncări.
- X se spune că este distribuită binomial de parametrii n, p .
- $D_X = \{0, 1, 2, \dots, n\}$
- Ne interesează $p_k = P(X = k)$

Cazul $n = 3$, $\Omega = \{(\text{Aruncarea 1}, \text{Aruncarea 2}, \text{Aruncarea 3})\}$, avem:

- nu se obține **banul la nicio aruncare** \leftrightarrow se obține (S,S,S);

$$P(X = 0) = \frac{1}{2 \cdot 2 \cdot 2} = C_3^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^3$$

- se obține **banul la o aruncare** din cele trei \leftrightarrow (B,S,S) sau (S,B,S) sau (S, S, B);

$$P(X = 1) = \frac{3}{2 \cdot 2 \cdot 2} = C_3^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^2$$

- se obține **banul la două aruncări** din cele trei \leftrightarrow (B,B,S) sau (S,B,B) sau (B, S, B);

$$P(X = 2) = \frac{3}{2 \cdot 2 \cdot 2} = C_3^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1$$

- se obține **banul la toate cele 3 aruncări** \leftrightarrow (B,B,B)

$$P(X = 3) = \frac{1}{2 \cdot 2 \cdot 2} = C_3^3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^0$$

Deci,

$$X = \left(C_3^k p^k (1-p)^{n-k} \right), \quad k = 0, 1, 2, 3, p = 1/2$$

Definiție

Variabilă aleatoare binomială este o variabilă discretă având distribuția de probabilitate: $X = \left(C_n^k p^k (1-p)^{n-k} \right), \quad k = 0, 1, \dots, n$

- $\text{Bin}(n, p)$ -clasa variabilelor aleatoare binomiale, asociate unor experimente Bernoulli cu n încercări și probabilitatea p a unui succes într-o încercare
- Notăție $X \sim \text{Bin}(n, p)$
- n și p se numesc parametrii distribuției binomiale.

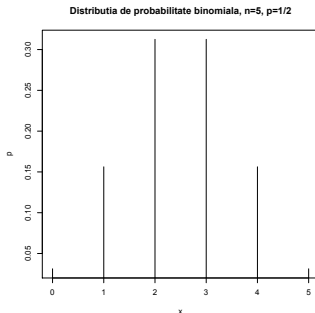


Figure: Vizualizarea distribuției binomiale

Ce se întâmplă dacă am o monedă falsă, în care **Banul** apare cu $p = 2/3$?

Pentru mai multe vizualizări ale distribuției binomiale vizitați
<https://mathlets.org/mathlets/>

- Mulțimea valorilor $D = \{0, 1, 2, \dots, n\}$
- $P(X = k) = ?$, $k \in \{0, 1, 2, \dots, n\}$ (probabilitatea ca în n încercări să înregistrăm k succese)
- Fie X_i variabila aleatoare Bernoulli, ce înregistrează rezultatul încercării i . Variabilele X_i sunt identic distribuite și independente:

$$X_i = \begin{pmatrix} 1 & 0 \\ p & 1 - p \end{pmatrix}, \quad i \in \{1, \dots, n\}$$

- Vectorul aleator (X_1, X_2, \dots, X_n) are distribuția de probabilitate:
$$\begin{aligned} P((X_1, X_2, \dots, X_n) = (i_1, i_2, \dots, i_n)) &= P(X_1 = i_1, \dots, X_n = i_n) \\ &= P(X_1 = i_1)P(X_2 = i_2) \cdots P(X_n = i_n) \\ &= p^k (1 - p)^{n-k}, \end{aligned}$$
$$k \text{ este numărul de valori } 1 \text{ în } n\text{-lista de biți } (i_1, i_2, \dots, i_n).$$

- Caz particular: $n = 3$
- Numărul de succese în $n = 3$ încercări este $X = X_1 + X_2 + X_3$.
- suma variabilelor Bernoulli $X_1 + X_2 + X_3$ are valori în $\{0, 1, 2, 3\}$.
- De exemplu, pentru cazul $k = 2$

$$P(X_1 + X_2 + X_3 = 2) = P[(X_1 = 1, X_2 = 1, X_3 = 0) \cup (X_1 = 1, X_2 = 0, X_3 = 1) \cup (X_1 = 0, X_2 = 1, X_3 = 1)] = C_3^2 p^2 (1 - p)$$

- Probabilitatea ca în n încercări să avem k succese este:

$$\begin{aligned} P(X_1 + X_2 + \dots + X_n = k) &= P\left(\bigcup_{i_1 + i_2 + \dots + i_n = k} (i_1, i_2, \dots, i_n) (X_1 = i_1, \dots, X_n = i_n)\right) \\ &= \sum_{i_1 + i_2 + \dots + i_n = k} \sum_{(i_1, i_2, \dots, i_n)} P(X_1 = i_1, \dots, X_n = i_n) \end{aligned}$$

unde suma se calculează după toate n -listele de biți (i_1, i_2, \dots, i_n) pentru care suma biților este k (există C_n^k astfel de liste!).

- Deci,

$$P(X_1 + X_2 + \dots + X_n = k) = C_n^k p^k (1 - p)^{n-k}$$

De exemplu, cazul în care aruncăm o monedă până la primul succes (apariția banului).

- Efectuăm o succesiune de încercări independente ce pot avea ca rezultat, succes sau eșec;
- Fie Y variabila aleatoare ce înregistrează numărul de încercări până la primul succes, inclusiv.
- Mulțimea valorilor variabilei Y este $\mathbb{N}^* = \{1, 2, \dots, n, \dots\}$;

Distribuția geometrică se definește în contextul unei experiment Bernoulli în care numărul de încercări nu este fixat apriori.

- Dorim să calculăm $P(Y=k), \forall k \in \mathbb{N}^*$;
- Evenimentul $(Y = k)$ se exprimă astfel

$$(Y = k) = (X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1)$$

- Dacă notăm cu p probabilitatea succesului, atunci:
$$P(Y = k) = P(X_1 = 0)P(X_2 = 0) \cdots P(X_{k-1} = 0)P(X_k = 1) = (1 - p)^{k-1}p$$
- Fie X_i rezultatul celei de-a i -a încercări. X_i ia valoarea 1, respectiv 0 după cum în a i -a încercare rezultatul este un succes, respectiv eșec.

Definiție

O variabilă aleatoare Y ce are distribuția geometrică, dă numărul de încercări într-un proces Bernoulli, până se obține primul succes, *inclusiv*. Y ia o mulțime numărabilă de valori.

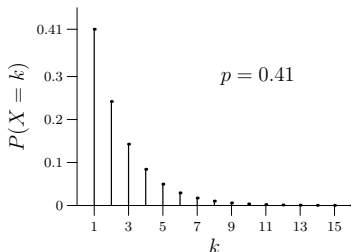
$$Y = \left(\binom{k}{(1-p)^{k-1}p} \right), \quad k \in \mathbb{N}^*, p \in (0, 1)$$

Geom(p)- clasa varb. a. cu distribuția geometrică de parametru p .

Proprietăți

Dacă $X \sim \text{Geom}(p)$, atunci $M(X) = \frac{1}{p}$ și $\sigma^2(X) = \frac{1-p}{p^2}$.

Distribuția de probabilitate geometrică, $\text{Geom}(p = 0.41)$,



Observație importantă

Distribuția geometrică^a se poate defini și ca variabila W ce înregistrează numărul de eșecuri, înaintea primului succes, adică $W=Y-1$.

În acest caz, $D_W = \{0, 1, 2, \dots\}$, iar

$$P(W = k) = P(Y - 1 = k) = P(Y = K + 1) = (1 - p)^k p$$

^ahttps://en.wikipedia.org/wiki/Geometric_distribution

Se consideră o mulțime de N obiecte, dintre care a obiecte clasificate cu succes și b obiecte clasificate cu eșec. Din cele N produse se selectează aleator n obiecte $0 \leq n \leq N$, fără înlocuire. Prin variabila X vom măsura câte produse din cele n selectate sunt din categoria a .

Variabila X are tabelul de distribuție

$$X = \left(\frac{C_a^k C_b^{n-k}}{C_N^n} \right)$$

Notăm $X \sim \text{Hipgeom}(N, n, p = \frac{a}{N})$.

Avem $M(X) = np$, $\sigma^2(X) = \frac{N-n}{N-1} npq$.

Acest model de distribuție poate fi folosit în diverse situații, cum ar fi: populația poate fi un grup de indivizi cu aceeași opinie sau orice altă caracteristică comună.

Alte distribuții clasice de probabilitate?

Variabilă aleatoare distribuită Poisson = numărul de produceri ale unui eveniment rar, într-un interval de timp fixat.

Exemple clasice pentru distribuția Poisson :

- numărul de cereri de acces pe minut la un server al unei baze de date;
- număr de apeluri la o centrală telefonică într-o oră

Observații:

- evenimentele ce se pot produce în orice moment al intervalului de timp fixat sunt independente;
- presupunem că evenimentul se produce cu "intensitate" constantă: în medie, se produc λ evenimente în intervalul fixat de timp,

Rezultat

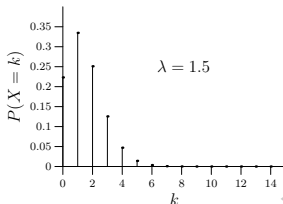
Probabilitatea ca evenimentul rar, A , să se producă de k ori într-un interval de timp fixat, este: $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$.

Definiție

Distribuția de probabilitate a variabilei aleatoare Poisson este:

$$X = \left(\frac{k}{\frac{e^{-\lambda} \lambda^k}{k!}} \right) \quad k = 0, 1, 2, \dots, n, \dots,$$

- $\text{Pois}(\lambda)$ clasa variabilelor aleatoare Poisson de parametru λ .
- $X \sim \text{Pois}(\lambda)$. , atunci $M(X) = \lambda$ și $\sigma^2(X) = \lambda$;
- Variabilele aleatoare Poisson sunt singurele cunoscute, pentru care media este egală cu dispersia.



Distribuția Poisson este o aproximare a distribuției binomiale
(n – foarte mare, p – foarte mic, $\lambda = np$ constantă pozitivă)

Teoremă

Fie $X \sim B(n, p = \frac{\lambda}{n})$, cu $\lambda > 0$ fixat. Atunci, $\forall k \in \{0, 1, 2, \dots\}$, avem

$$\lim_{n \rightarrow \infty} P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Important: În cazuri speciale (n -mare și p -mic) putem folosi distribuția Poisson, care este mult mai simplă decât distribuția binomială.

- Există milioane de distribuții ce apar în diferite contexte;
- Se poate folosi Wikipedia pentru alte informații (de exemplu, **distribuția hipergeometrică** https://en.wikipedia.org/wiki/Hypergeometric_distribution sau **distribuția uniformă** https://en.wikipedia.org/wiki/Discrete_uniform_distribution)
- important este să reținem ce putem modela cu fiecare distribuție și să ne dezvoltăm intuiție legata de ele (chiar si prin graficul lor)

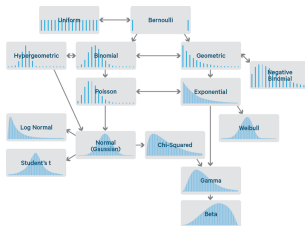


Figure: Distribuții discrete și continue