Curs - Probabilităţi şi Statistică 2021/2022

Facultatea de Matematică şi Informatică
Universitatea Babeş-Bolyai, Cluj-Napoca

Teoria Probabilităţilor

Teoria probabilităţilor este o disciplină a matematicii care se ocupă de studiul fenomenelor aleatoare.

- *aleator* = care depinde de o împrejurare viitoare şi nesigură; supus întâmplării
- provine din latină: *aleatorius*; *alea* (lat.) = zar; joc cu zaruri; joc de noroc; şansă; risc

↪ se măsoară *şansele pentru succes* sau *riscul pentru insucces* al unor evenimente

Fenomene şi procese aleatoare apar, de exemplu, în:

[Sursa: www.financialmarket.ro]

→ jocuri de noroc, pariuri, loto (6 din 49)
→ previziuni meteo
→ previziuni economice / financiare
→ sondaje de opinie, asigurări (evaluarea riscurilor, pierderilor)

→ **în informatică**:
▷ sisteme de comunicare, prelucrarea informaţiei, modelarea traficului în reţea;
▷ analiza probabilistică a unor algoritmi, fiabilitatea sistemelor;
▷ algoritmi de simulare, machine learning, data mining, recunoaşterea formelor sau a vocii;
▷ generarea de numere aleatoare, algoritmi aleatori: de tip Monte-Carlo, de tip Las Vegas etc.

Octave online: `https://octave-online.net`

Exemplu: Generarea de valori aleatoare (în Octave/Matlab)

```
a=rand % valoare aleatoare între 0 şi 1
v1=rand(1,10)    %vector cu 10 valori aleatoare între 0 şi 1
a=4; b=10;
v2=a+(b-a)*rand(1,15) %vector cu 15 valori  aleatoare între 4 şi 10
z=randi(6,1,20)
%vector cu 20 de valori aleatoare din mulţimea {1,2,3,4,5,6}
```

Exerciţiu: Generaţi un vector cu 100 de valori aleatoare `0` şi `1`, în care `0` şi `1` au aceleaşi şanse de apariţie.

Răspuns: `floor(2*rand(1,100))` sau `randi(2,1,100)-1`

Algoritmi aleatori

Def. 1. *Un algoritm pe cursul executării căruia se iau anumite decizii aleatoare este numit* ***algoritm aleator (randomizat)****.*

▷ durata de execuţie, spaţiul de stocare, rezultatul obţinut sunt variabile aleatoare (chiar dacă se folosesc aceleaşi valori input)

▷ la anumite tipuri de algoritmi corectitudinea e garantată doar cu o anumită probabilitate

▷ în mod paradoxal, incertitudinea ne poate oferi mai multă eficienţă

Exemplu: Random QuickSort, în care elementul pivot este selectat aleator

• Algoritm de tip **Las Vegas** este un algoritm aleator, care returnează la fiecare execuţie rezultatul corect (independent de alegerile aleatoare făcute); durata de execuţie este o variabilă aleatoare.
Exemplu: Random QuickSort

• Un algoritm aleatoriu pentru care rezultatele obţinute sunt corecte *doar* cu o anumită probabilitate se numeşte algoritm **Monte Carlo**.
$\hookrightarrow$ se examinează probabilitatea cu care rezultatul este corect; probabilitatea de eroare poate fi scăzută semnificativ prin execuţii repetate, independente;
Exemplu:
▷ testul Miller-Rabin, care verifică dacă un număr natural este prim sau este număr compus; testul returnează fie răspunsul "numărul este sigur un număr compus" sau răspunsul "numărul este probabil un număr prim";
Exerciţiu: Fie S(1),...,S(300) un vector cu 300 de elemente, din mulţimea $\{0, 1, 2\}$ (ordinea lor este necunoscută). $\longrightarrow$ De care tip este următorul algoritm (scris în Octave)?

```
S=randi(3,1,300)-1;
k=0;
do
    k=k+1;
    i=randi(300);
until (S(i) == 0)
i  %  indicele, pentru care S(i)=0
k  % număr iteraţii până se găseşte aleator un 0
```

Răspuns: Algoritm de tip Las Vegas.

Versiunea Monte Carlo a problemei formulate anterior: se dă M numărul maxim de iteraţii.

```
M=3;
S=randi(3,1,300)-1;
k=0;
do
    k=k+1 ;
    i=randi(300);
until ( (S(i) == 0) || (k==M) )
i  % indicele, pentru care S(i)=0  sau pentru care k==M
k
% număr iteraţii până se găseşte
% aleator un 0 sau programul s-a oprit
S(i)
```

▷ dacă 0 este găsit, atunci algoritmul se încheie cu rezultatul corect, altfel algoritmul nu găseşte niciun 0.

Noţiuni introductive:

• Experienţa aleatoare este acea experienţă al cărei rezultat nu poate fi cunoscut decât după încheierea ei.

• Evenimentul este rezultatul unui experiment.

Exemple:

▷ Experiment: aruncarea a două zaruri, eveniment: ambele zaruri indică 1

▷ experiment: aruncarea unei monede, eveniment: moneda indică pajură

▷ experiment: extragerea unei cărţi de joc, eveniment: s-a extras as

▷ experiment: extragerea unui număr la loto, eveniment: s-a extras numărul 27

• evenimentul imposibil, notat cu $\emptyset$, este evenimentul care nu se realizează niciodată la efectuarea experienţei aleatoare

• evenimentul sigur este un eveniment care se realizează cu certitudine la fiecare efectuare a experienţei aleatoare

• spaţiul de selecţie, notat cu Ω, este mulţimea tuturor rezultatelor posibile ale experimentului considerat

◇ spaţiul de selecţie poate fi finit sau infinit

• dacă A este o submulţime a lui Ω atunci A se numeşte eveniment aleator, iar dacă A are un singur element atunci A este un eveniment elementar.

▷ O *analogie între evenimente şi mulţimi* permite o scriere şi o exprimare mai comode ale unor idei şi rezultate legate de conceptul de eveniment aleator.

Exemplu: Experimentul: aruncarea unui zar, spaţiul de selecţie: $\Omega = \{e_1, e_2, e_3, e_4, e_5, e_6\}$,
e_i: s-a obţinut numărul i ($i = 1, \ldots, 6$) ; $e_1, e_2, e_3, e_4, e_5, e_6$ sunt evenimente elementare

A: s-a obţinut un număr par $\Rightarrow A = \{e_2, e_4, e_6\}$

$\bar{A}$: s-a obţinut un număr impar $\Rightarrow \bar{A} = \{e_1, e_3, e_5\}$ ♣

Operaţii cu evenimente

• dacă $A, B \subseteq \Omega$, atunci evenimentul reuniune $A \cup B$ este un eveniment care se produce dacă cel puţin unul din evenimentele A sau B se produce

• dacă $A, B \subseteq \Omega$, atunci evenimentul intersecţie $A \cap B$ este un eveniment care se produce dacă cele două evenimente A şi B se produc în acelaşi timp

• dacă $A \subseteq \Omega$ atunci evenimentul contrar sau complemetar $\bar{A}$ este un eveniment care se realizează atunci când evenimentul A nu se realizează

• $A, B \subseteq \Omega$ sunt evenimente incompatibile (disjuncte), dacă $A \cap B = \emptyset$

• dacă $A, B \subseteq \Omega$, atunci evenimentul diferenţă $A \setminus B$ este un eveniment care se produce dacă A are loc şi B nu are loc, adică

$$A \setminus B = A \cap \bar{B}$$

Relaţii între evenimente

• dacă $A, B \subseteq \Omega$, atunci A implică B, dacă producerea evenimentului A conduce la producerea evenimentului B: $A \subseteq B$

• dacă A implică B şi B implică A, atunci evenimentele A şi B sunt egale: $A = B$

Proprietăţi ale operaţiilor între evenimente $A, B, C \subseteq \Omega$

Operaţiile de reuniune şi intersecţie sunt operaţii **comutative**:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A\,,$$

asociative

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)\,,$$

şi **distributive**

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C);$$

satisfac legile lui De Morgan

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Are loc $\bar{\bar{A}} = A$.

Frecvenţa relativă şi frecvenţa absolută

Def. 2. *Fie A un eveniment asociat unei experienţe, repetăm experienţa de n ori (în aceleaşi condiţii date) şi notăm cu $r_n(A)$ numărul de realizări ale evenimentului A; frecvenţa relativă a evenimentului A este numărul*

$$f_n(A) = \frac{r_n(A)}{n}$$

$r_n(A)$ este frecvenţa absolută a evenimentului A.

Definiţia clasică a probabilităţii

Def. 3. *Într-un experiment în care cazurile posibile sunt finite la număr şi au aceleaşi şanse de a se realiza,* ***probabilitatea*** *unui eveniment A este numărul*

$$P(A) = \frac{\textit{numărul de cazuri favorabile apariţiei lui } A}{\textit{numărul total de cazuri posibile}}.$$

$\triangleright$ Prin repetarea de multe ori a unui experiment, în condiţii practic identice, frecvenţa relativă $f_n(A)$ de apariţie a evenimentului A este aproximativ egală cu $P(A)$

$$f_n(A) \approx P(A), \text{ dacă } n \to \infty.$$

Exemplu: Experiment: Se aruncă 4 monede. Evenimentul A: *(exact) 3 din cele 4 monede indică pajură*; experimentul s-a repetat de $n = 100$ de ori şi evenimentul A a apărut de 22 de ori.

$$f_n(A) = ?, \qquad P(A) = ?$$

Răspuns: $f_n(A) = \frac{22}{100} = 0.22$

$\Omega = \{(c, c, c, c), (c, p, p, p), \ldots, (p, p, p, c), (p, p, p, p)\}$

$A = \{(c, p, p, p), (p, c, p, p), (p, p, c, p), (p, p, p, c)\} \Rightarrow P(A) = \frac{4}{2^4} = 0.25$ ♠

Exemplu - Joc de zaruri (sec. XVII): Un pasionat jucător de zaruri, cavalerul de Méré, susţinea în discuţiile sale cu B. Pascal că a arunca un zar de 4 ori pentru a obţine cel puţin o dată faţa şase, este acelaşi lucru cu a arunca de 24 ori câte două zaruri pentru a obţine cel puţin o dublă de şase. Cu toate acestea, cavalerul de Méré a observat că jucând în modul al doilea (cu două zaruri aruncate de 24 ori), pierdea faţă de adversarul său, dacă acesta alegea primul mod (aruncarea unui singur zar de 4 ori). Pascal şi Fermat au arătat că probabilitatea de câştig la jocul cu un singur zar aruncat de 4 ori este $p_1 \approx 0.5177$, iar probabilitatea $p_2 \approx 0.4914$ la jocul cu două zaruri aruncate de 24 de ori. Deşi diferenţa dintre cele două probabilităţi este mică, totuşi, la un număr mare de partide, jucătorul cu probabilitatea de câştig p_1 câştigă în faţa jucătorului cu probabilitatea de câştig p_2. Practica jocului confirmă astfel justeţea raţionamentului matematic, contrar credinţei lui de Méré.

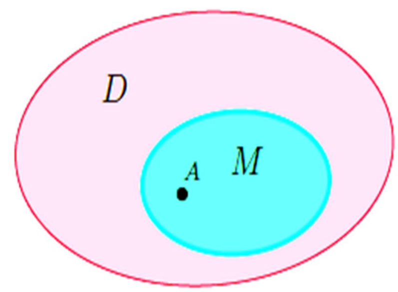
♠

Definiţia axiomatică a probabilităţii

Definiţia clasică a probabilităţii poate fi utilizată numai în cazul în care numărul cazurilor posibile este finit. Dacă numărul evenimentelor elementare este infinit, atunci există evenimente pentru care probabilitatea în sensul clasic nu are nici un înţeles.

Probabilitatea geometrică: Măsura unei mulţimi corespunde lungimii în $\mathbb{R}$, ariei în $\mathbb{R}^2$, volumului în $\mathbb{R}^3$. Fie $M \subset D \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$, mulţimi cu măsură finită.
Alegem aleator un punct $A \in D$ (în acest caz spaţiul de selecţie este D). Probabilitatea geometrică a evenimentului "$A \in M$" este

$$P(A \in M) := \frac{\text{măsura}(M)}{\text{măsura}(D)}.$$



$M \subset D \subset \mathbb{R}^2$

O teorie formală a probabilităţii a fost creată în anii '30 ai secolului XX de către matematicianul rus **Andrei Nikolaevici Kolmogorov**, care, în anul **1933**, a dezvoltat teoria axiomatică a probabilităţii în lucrarea sa *Conceptele de bază ale Calculului Probabilităţii*.

$\twoheadrightarrow$ $P : \mathcal{K} \to \mathbb{R}$ este o funcţie astfel încât oricărui eveniment aleator $A \in \mathcal{K}$ i se asociază valoarea $P(A)$, ***probabilitatea de apariţie a evenimentului*** A

$\hookrightarrow$ $\mathcal{K}$ este o mulţime de evenimente şi are structura unei σ-algebre (vezi Def. 4)
$\hookrightarrow$ P satisface anumite axiome (vezi Def. 5)

Def. 4. *O familie $\mathcal{K}$ de evenimente din spaţiul de selecţie Ω se numeşte* ***σ-algebră*** *dacă sunt satisfăcute condiţiile:*

(i) *$\mathcal{K}$ este nevidă;*

(ii) *dacă $A \in \mathcal{K}$, atunci $\bar{A} \in \mathcal{K}$;*

(iii) *dacă $A_n \in \mathcal{K}$, $n \in \mathbb{N}^*$, atunci* $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{K}$.

Perechea $(\Omega, \mathcal{K})$ se numeşte ***spaţiu măsurabil***.

Exemple: 1) Dacă $\emptyset \neq A \subset \Omega$ atunci $\mathcal{K} = \{\emptyset, A, \bar{A}, \Omega\}$ este o σ-algebră.
2) $\mathcal{P}(\Omega)$:= mulţimea tuturor submulţimilor ale lui Ω este o σ-algebră.
3) Dacă $(\Omega, \mathcal{K})$ este un spaţiu măsurabil şi $\emptyset \neq B \subseteq \Omega$, atunci

$$B \cap \mathcal{K} = \{B \cap A : A \in \mathcal{K}\}$$

este o σ-algebră pe mulţimea B, iar $(B, B \cap \mathcal{K})$ este un spaţiu măsurabil.

P. 1. *Proprietăţi ale unei σ-algebre: Dacă $\mathcal{K}$ este o σ-algebră în Ω, atunci au loc proprietăţile:*

(1) $\emptyset, \Omega \in \mathcal{K}$*;*

(2) $A, B \in \mathcal{K} \Longrightarrow A \cap B,\ A \setminus B \in \mathcal{K}$;

(3) $A_n \in \mathcal{K}$*,* $n \in \mathbb{N}^* \Longrightarrow \bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{K}$*.*

Def. 5. *Fie $\mathcal{K}$ o σ-algebră în Ω. O funcţie $P : \mathcal{K} \to \mathbb{R}$ se numeşte **probabilitate** dacă satisface axiomele:*

(i) $P(\Omega) = 1$;

(ii) $P(A) \geq 0$ *pentru orice* $A \in \mathcal{K}$;

(iii) pentru orice şir $(A_n)_{n \in \mathbb{N}^*}$ de evenimente două câte două disjuncte (adică $A_i \cap A_j = \emptyset$ pentru orice $i \neq j$) din $\mathcal{K}$ are loc
$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n)\,.$$

*Tripletul $(\Omega, \mathcal{K}, P)$ format din spaţiul măsurabil $(\Omega, \mathcal{K})$ şi probabilitatea $P : \mathcal{K} \to \mathbb{R}$ se numeşte **spaţiu de probabilitate**.*

P. 2. *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate. Au loc proprietăţile:*

(1) $P(\bar{A}) = 1 - P(A)$ *şi* $0 \leq P(A) \leq 1$*;*

(2) $P(\emptyset) = 0$*;*

(3) $P(A \setminus B) = P(A) - P(A \cap B)$*;*

(4) $A \subseteq B \Longrightarrow P(A) \leq P(B)$*, adică P este monotonă;*

(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*.*

Exerciţiu: a) Să se arate că pentru $\forall\, A, B, C \in \mathcal{K}$ are loc:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

b) Pentru $A_1, ..., A_n \in \mathcal{K}$ care e formula similară de calcul pentru $P(A_1 \cup A_2 \cup ... \cup A_n)$?

Exemplu: Dintr-un pachet de 52 de cărţi de joc se extrage o carte aleator. Care este probabilitatea p de a extrage a) un as sau o damă de pică? b) o carte cu inimă sau un as?
R.: a) A: s-a extras un as; D: s-a extras damă de pică; A şi D sunt două evenimente disjuncte (incompatibile)

$$p = P(A \cup D) = P(A) + P(D) = \frac{4+1}{52};$$

b) I: s-a extras o carte cu inimă; I şi A *nu* sunt evenimente incompatibile

$$p = P(I \cup A) = P(I) + P(A) - P(I \cap A) = \frac{13+4-1}{52} = \frac{4}{13}.$$

♡

Evenimente independente

Def. 6. *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate. Evenimentele $A, B \in \mathcal{K}$ sunt* ***evenimente independente*** *dacă*

$$P(A \cap B) = P(A)P(B).$$

Observaţie: Fie evenimentele $A, B \in \mathcal{K}$. Evenimentele A şi B sunt **independente**, dacă **apariţia evenimentului A, nu influenţează apariţia evenimentului B şi invers**. Două evenimente se numesc **dependente** dacă probabilitatea realizării unuia dintre ele depinde de faptul că celălalt eveniment s-a produs sau nu.
Exerciţiu: Se aruncă un zar de două ori.
A: primul număr este 6; B: al doilea număr este 5; C: primul număr este 1.
Sunt A şi B evenimente independente?
Sunt A şi C evenimente independente? Sunt A şi C evenimente disjuncte?
Sunt B şi C evenimente dependente? ♣

P. 3. *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate şi fie $A, B \in \mathcal{K}$. Sunt echivalente afirmaţiile:*

(1) *A şi B sunt independente.*

(2) *$\bar{A}$ şi B sunt independente.*

(3) *A şi $\bar{B}$ sunt independente.*

(4) *$\bar{A}$ şi $\bar{B}$ sunt independente.*

Def. 7. *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate. $B_1, \ldots, B_n$ sunt n* ***evenimente independente (în totalitate)*** *din $\mathcal{K}$ dacă*

$$P(B_{i_1} \cap \cdots \cap B_{i_m}) = P(B_{i_1}) \cdot \ldots \cdot P(B_{i_m})$$

pentru orice submulţime finită $\{i_1, \ldots, i_m\} \subseteq \{1, 2, ..., n\}$.

Exemplu: Se dă algoritmul de tip Monte-Carlo

```
M=input('M=') % numar maxim de iteratii; M >= 1
M=3 % maximale Anzahl Iterationen; M >= 1
S=[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3];
% 0,1,2,3 apar fiecare cu probabilitatea  5/20=1/4
S=S(randperm(length(S))) % permutare aleatoare a lui S
k=0;
do
    k=k+1;
    i=randi(20);
    % se alege o valoare aleatoare S(i)
until ( (S(i) == 0) || (k==M) )
fprintf('k: %d \n',k)
fprintf('S(%d): %d \n',i,S(i))
```

Se calculează probabilităţile următoarelor evenimente (din punct de vedere teoretic):

$$P(\text{“primul 0 este găsit la a } M\text{-a iteraţie”}) = \left(\frac{3}{4}\right)^{M-1} \cdot \frac{1}{4},$$

$$P(\text{“0 nu este găsit în } M \text{ iteraţii”}) = \left(\frac{3}{4}\right)^{M},$$

probabilitatea evenimentului complementar este

$$P(\text{“cel puţin un 0 este găsit în } M \text{ iteraţii”}) = 1 - \left(\frac{3}{4}\right)^{M} \longrightarrow 1, \text{ când } M \to \infty.$$

♣

Exemplu: 1) $A, B, C \in \mathcal{K}$ sunt trei evenimente independente (în totalitate), dacă

$$P(A \cap B) = P(A)P(B),\ P(A \cap C) = P(A)P(C),\ P(B \cap C) = P(B)P(C),$$

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

2) Cele 4 feţe ale unui tetraedru regulat sunt vopsite astfel: una este roşie, una este albastră, una este verde şi una este colorată având cele trei culori. Se aruncă tetraedrul şi se consideră evenimentele: R: tetraedrul cade pe o parte ce conţine culoarea roşie; A: tetraedrul cade pe o parte ce conţine culoarea albastră; V: tetraedrul cade pe o partea ce conţine culoarea verde.
Sunt cele 3 evenimente independente în totalitate?

R.: Nu, cele 3 evenimente nu sunt independente în totalitate pentru că $P(R \cap A \cap V) = \frac{1}{4} \neq P(R)P(A)P(V) = \frac{1}{8}$.
3) Pentru a verifica dacă n evenimente distincte $B_1, \ldots, B_n$ sunt independente în totalitate câte relaţii trebuie verificate?
R.: $C_n^2 + C_n^3 + ... + C_n^n = 2^n - C_n^0 - C_n^1 = 2^n - 1 - n$. ♦

Probabilitate condiţionată

Def. 8. *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate şi fie $A, B \in \mathcal{K}$.* ***Probabilitatea condiţionată a evenimentului** A **de evenimentul** B este $P(\cdot|B) : \mathcal{K} \to [0,1]$ definită prin*

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

dacă $P(B) > 0$. $P(A|B)$ este probabilitatea apariţiei evenimentului A, ştiind că evenimentul B s-a produs.

Observaţie: Fie evenimentele $A, B \in \mathcal{K}$ astfel încât $P(A) > 0$ şi $P(B) > 0$. Evenimentele A şi B sunt **independente** (a se vedea Def. 6), dacă apariţia evenimentului A, nu influenţează apariţia evenimentului B şi invers, adică

$$P(A|B) = P(A) \text{ şi } P(B|A) = P(B)\,.$$

Exemplu: Se extrag succesiv fără returnare două bile dintr-o urnă cu 4 bile albe şi 5 bile roşii.
a) Ştiind că prima bilă este roşie, care este probabilitatea (condiţionată) ca a doua bilă să fie albă?
b) Care este probabilitatea ca ambele bile să fie roşii?
R.: pentru $i \in \{1, 2\}$ fie evenimentele
R_i: la a i-a extragere s-a obţinut o bilă roşie;
$A_i = \bar{R}_i$: la a i-a extragere s-a obţinut o bilă albă;
a) $P(A_2|R_1) = \frac{4}{8}$. **b)** $P(R_1 \cap R_2) = P(R_2|R_1)P(R_1) = \frac{4}{8} \cdot \frac{5}{9}$. ♣

P. 4. *Pentru $A, B \in \mathcal{K}, P(A) > 0, P(B) > 0$ au loc:*

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)\,,$$

$$P(\bar{A}|B) = 1 - P(A|B).$$

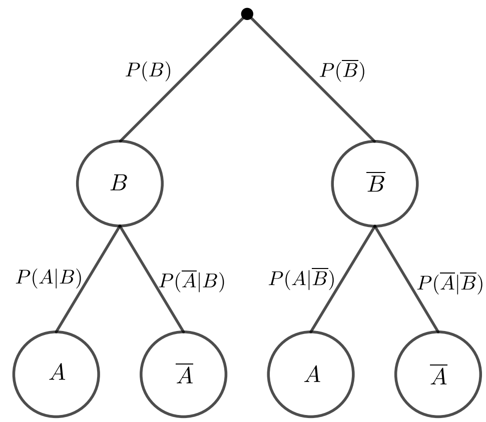


Fig.1. Probabilităţi condiţionate

Def. 9. *O familie $\{H_1, \ldots H_n\} \subset \mathcal{K}$ de evenimente din Ω se numeşte* ***partiţie*** *sau* ***sistem complet de evenimente*** *a lui Ω, dacă $\bigcup\limits_{i=1}^{n} H_i = \Omega$ şi pentru fiecare $i, j \in \{1, \ldots, n\}$, $i \neq j$, evenimentele H_i şi H_j sunt disjuncte, adică $H_i \cap H_j = \emptyset$.*

Exemplu: Dacă $B \subset \Omega$ atunci $\{B, \bar{B}\}$ formează o partiţie a lui Ω. ♠

P. 5. ***(Formula probabilităţii totale)*** *Într-un spaţiu de probabilitate $(\Omega, \mathcal{K}, P)$ considerăm partiţia $\{H_1, ..., H_n\}$ a lui Ω cu $H_i \in \mathcal{K}$ şi $P(H_i) > 0\ \forall\, i \in \{1, \ldots, n\}$, şi fie $A \in \mathcal{K}$. Atunci are loc*

$$P(A) = P(A|H_1)P(H_1) + ... + P(A|H_n)P(H_n).$$

Exemplu: Într-o urnă sunt 7 bile albe, notate cu 1,2,3,4,5,6,7, şi 6 bile roşii notate cu 8,9,10,11,12,13. Se extrage o bilă. **a)** Ştiind că bila extrasă este roşie, care este probabilitatea (condiţionată) p_1, ca numărul înscris să fie divizibil cu 4? **b)** Ştiind că prima bilă este roşie, care este probabilitatea (condiţionată) p_2, ca o a doua bilă extrasă să fie un număr impar? (Prima bilă nu s-a returnat în urnă!)

R.: Se consideră evenimentele:

A_1: prima bilă extrasă are înscris un număr divizibil cu 4;

B_1: prima bilă extrasă este roşie;

C_1: prima bilă extrasă are înscris un număr impar;

C_2: a doua bilă extrasă are înscris un număr impar.

a) $p_1 = P(A_1|B_1) = \frac{2}{6}\,.$

b) $p_2 = P(C_2|B_1) =?$ Folosim Def.8 şi P.4, scriem succesiv

$$
\begin{aligned}
p_2 = P(C_2|B_1) &= \frac{P(C_2 \cap B_1)}{P(B_1)} = \frac{P(C_2 \cap B_1 \cap C_1) + P(C_2 \cap B_1 \cap \bar{C}_1)}{P(B_1)} \\
&= \frac{P(C_2|B_1 \cap C_1)P(B_1 \cap C_1) + P(C_2|B_1 \cap \bar{C}_1)P(B_1 \cap \bar{C}_1)}{P(B_1)} = \frac{\frac{6}{12} \cdot \frac{3}{13} + \frac{7}{12} \cdot \frac{3}{13}}{\frac{6}{13}} = \frac{13}{24}\,.
\end{aligned}
$$

♡

Exemplu: Ce probabilităţi calculează programul de mai jos? Ce tip de algoritm aleator este?
► `randi(imax,n,m)` generează o `n`×`m` matrice cu valori întregi aleatoare (pseudoaleatoare) între `1` şi `imax`.

```
clear all
ci=0;
cp=0;
c=0;
a=0;
b=0;
N=1000;
A=[1:20];
for i=1:N
  r= randi(length(A));
 v=A(r);
 ci=ci+mod(v,2);
 cp=cp+(mod(v,2)==0);
  c=c+ mod(v,2)*(mod(v,3)==0);
  a=a+ mod(v,2)*(6<=v && v<=10);
  b=b+ (mod(v,2)==0)*(v>=14);
end
p1=c/ci
p2=a/ci
p3=b/cp
```

R.: Se extrage aleator un număr din şirul `A=[1,2,...,20]`.
►`p1` estimează probabilitatea condiţionată ca numărul ales aleator să fie divizibil cu 3, *ştiind* că s-a extras un număr impar;
► `p2` estimează probabilitatea condiţionată ca numărul ales aleator să provină din mulţimea $\{6, 7, 8, 9, 10\}$, *ştiind* că s-a extras un număr impar;

► `p3` estimează probabilitatea condiţionată ca numărul ales aleator să provină din mulţimea $\{14, 15, 16, 17, 18, 19, 20\}$, *ştiind* că s-a extras un număr par.
Algoritmul este de tip Monte-Carlo! ♣

Exerciţiu: Să se calculeze valorile teoretice pentru probabilităţile `p1`, `p2`, `p3` din exemplul anterior! ♠

P. 6. ***(Regula de înmulţire)*** *Fie $(\Omega, \mathcal{K}, P)$ un spaţiu de probabilitate şi fie $A_1, \dots, A_n \in \mathcal{K}$ astfel încât $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Atunci,*

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)\dots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Exemplu: Într-o urnă sunt 2 bile verzi şi 3 bile albastre. Se extrag 2 bile succesiv, fără returnare. Care este probabilitatea ca
a) prima bilă să fie verde, iar cea de-a doua albastră?
b) cele 2 bile să aibă aceeaşi culoare?
c) a doua bilă să fie albastră?
d) prima bilă să fie verde, *ştiind* că a doua este albastră?
e) se mai extrage o a treia bilă; se cere probabilitatea ca prima bilă să fie verde, cea de-a doua albastră şi a treia tot albastră.
R.: Notăm pentru $i \in \{1, 2, 3\}$ evenimentele:
A_i: la a i-a extragere s-a obţinut bilă albastră; V_i: la a i-a extragere s-a obţinut bilă verde;
a) folosim P.4: $P(V_1 \cap A_2) = P(A_2|V_1)P(V_1) = \frac{3}{4} \cdot \frac{2}{5}$
b) $P((V_1 \cap V_2) \cup (A_1 \cap A_2)) = P(V_1 \cap V_2) + P(A_1 \cap A_2) = P(V_2|V_1)P(V_1) + P(A_2|A_1)P(A_1) = \frac{1}{4} \cdot \frac{2}{5} + \frac{2}{4} \cdot \frac{3}{5}$
c) folosim formula probabilităţii totale P.7:
$P(A_2) = P(A_2|V_1)P(V_1) + P(A_2|A_1)P(A_1) = \frac{3}{4} \cdot \frac{2}{5} + \frac{2}{4} \cdot \frac{3}{5}$
d) folosim P.4: $P(V_1|A_2) = \frac{P(V_1 \cap A_2)}{P(A_2)} = \frac{P(A_2|V_1)P(V_1)}{P(A_2)} = \frac{\frac{3}{4} \cdot \frac{2}{5}}{\frac{3}{4} \cdot \frac{2}{5} + \frac{2}{4} \cdot \frac{3}{5}}$
e) formula de înmulţire a probabilităţilor P.6:
$P(V_1 \cap A_2 \cap A_3) = P(V_1) \cdot P(A_2|V_1) \cdot P(A_3|V_1 \cap A_2) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3}$.

♣

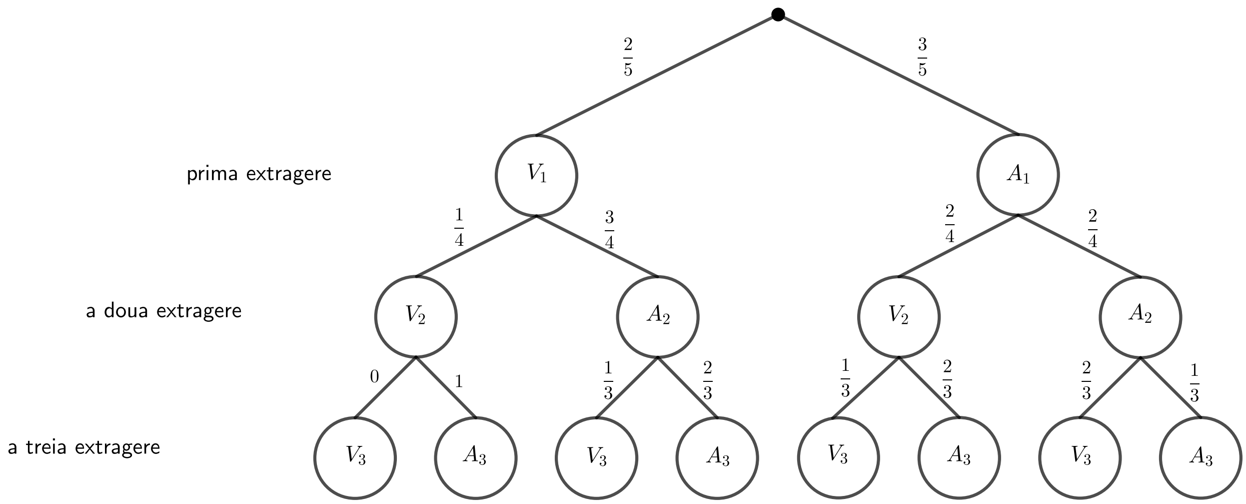


Fig. 3. Extragere fără returnare

Formula lui Bayes

Formula lui Bayes este o metodă de a "corecta" (a revizui, a îmbunătăţi) pe baza unor noi date (informaţii) disponibile o probabilitate determinată apriori. Se porneşte cu o estimare pentru probabilitatea unei anumite ipoteze H (engl. *hypothesis*). Dacă avem noi date (date de antrenare, dovezi, informaţii, evidenţe - engl. *evidence*) E, ce privesc ipoteza H, se poate calcula o probabilitate "corectată" pentru ipoteza H, numită probabilitate posterioară (a-posteriori).

$\hookrightarrow$ $P(H)$ probabilitatea ca ipoteza H să fie adevărată, numită şi *probabilitatea apriori*;

$\hookrightarrow$ probabilitatea condiţionată $P(H|E)$ este *probabilitatea posterioară* (corectată de cunoaşterea noilor date / informaţii);

$\hookrightarrow$ $P(E|H)$ probabilitatea ca să apară datele (informaţiile), ştiind că ipoteza H este adevarată;

$\hookrightarrow$ $P(E|\bar{H})$ probabilitatea ca să apară datele (informaţiile), ştiind că ipoteza H este falsă (ipoteza $\bar{H}$ este adevarată).

Folosind P.5 are loc:

$$P(E) = P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H}) = P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot (1 - P(H)).$$

Formula lui Bayes este în acest caz

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H})}.$$

P. 7. (Formula lui Bayes)

Într-un spaţiu de probabilitate $(\Omega, \mathcal{K}, P)$ *considerăm partiţia* $\{H_1, \ldots, H_n\}$ *a lui* Ω *cu* $H_i \in \mathcal{K}$

şi $P(H_i) > 0 \; \forall \; i \in \{1, ...n\}$, *şi fie* $E \in \mathcal{K}$ *astfel încât* $P(E) > 0$. *Atunci,*

$$P(H_j|E) = \frac{P(E|H_j)P(H_j)}{P(E)} = \frac{P(E|H_j)P(H_j)}{P(E|H_1)P(H_1) + ... + P(E|H_n)P(H_n)} \quad \forall \, j \in \{1, 2, ..., n\}.$$

▷ pentru $i \in \{1, 2, ..., n\}$ $P(H_i)$ sunt **probabilităţi apriori** pentru H_i, numite şi ipoteze (aserţiuni; engl. *hypothesis*)
▷ E se numeşte **evidenţă** (dovadă, premisă, informaţie; engl. *evidence*);
▷ cu formula lui Bayes se calculează probabilităţile pentru ipoteze, cunoscând evidenţa: $P(H_j|E)$, $j \in \{1, 2, ..., n\}$, care se numesc **probabilităţi posterioare** (ulterioare);
▷ $P(E|H_i)$, $i \in \{1, 2, ..., n\}$, reprezintă verosimilitatea (engl. *likelihood*) datelor observate.
▷ Se pot calcula probabilităţile *cauzelor*, date fiind (cunoscând / ştiind) *efectele*; formula lui Bayes ne ajută să diagnosticăm o anumită situaţie sau să testăm o ipoteză.

Exemplu: Considerăm evenimentele (în teste clinice):
H: o persoană aleasă aleator dintr-o populaţie are o anumită alergie $\mathcal{A}$
E: testul clinic returnează `pozitiv` privind alergia $\mathcal{A}$
$\bar{E}$: testul clinic returnează `negativ` privind alergia $\mathcal{A}$
▷ din statistici anterioare sunt cunoscute:
$p = P(H)$, probabilitatea ca o persoană selectată aleator din populaţie să sufere de alergia $\mathcal{A}$;
sensibilitatea testului $s_1 = P(E|H)$;
specificitatea testului $s_2 = P(\bar{E}|\bar{H})$;
▷ probabilitatea de a obţine răspuns `fals pozitiv` este $P(E|\bar{H}) = 1 - s_2$;
▷ un test clinic bun implică valori apropiate de 1 pentru s_1 şi s_2;
▶ cunoscând p, s_1, s_2 se doreşte a se determina *valoarea predictivă* $P(H|E)$:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H})} = \frac{s_1 \cdot p}{s_1 \cdot p + (1 - s_2) \cdot (1 - p)}.$$

♣

Variable aleatoare

Exemplu: Un jucător aruncă două monede $\Rightarrow \Omega = \{(c, p), (c, c), (p, c), (p, p)\}$ (c=cap; p=pajură)

X indică de câte ori a apărut pajură: $\Rightarrow X : \Omega \to \{0, 1, 2\}$

$\Rightarrow P(X = 0) = P(X = 2) = \frac{1}{4}, P(X = 1) = \frac{1}{2}$ ■

Notaţie 1. *variabilă/variabile aleatoare* $\to$ *v.a.*

O variabilă aleatoare este:
▶ **discretă**, dacă ia un număr finit de valori $(x_1, \ldots, x_n)$ sau un număr infinit numărabil de valori

$(x_1, \ldots, x_n, \ldots)$

► **continuă**, dacă valorile sale posibile sunt nenumărabile şi sunt într-un interval (sau reunine de intervale) sau în $\mathbb{R}$

V.a. discrete: exemple de v.a. numerice discrete: suma numerelor obţinute la aruncarea a 4 zaruri, numărul produselor defecte produse de o anumită firmă într-o săptămână; numărul apelurilor telefonice într-un call center în decursul unei ore; numărul de accesări ale unei anumite pagini web în decursul unei anumite zile (de ex. duminica); numărul de caractere transmise eronat într-un mesaj de o anumită lungime; exemple de v.a. categoriale ($\rightarrow$ se clasifică în categorii): prognoza meteo: ploios, senin, înnorat, ceţos; calitatea unor servicii: nesatisfăcătoare, satisfăcătoare, bune, foarte bune, excepţionale . . .)

V.a. continue sunt v.a. numerice: timpul de funcţionare până la defectare a unei piese electronice, temperatura într-un oraş, viteza înregistrată de radar pentru maşini care parcurg o anumită zonă . . .

Variabile aleatoare numerice - definiţie formală

Def. 10. *Fie* $(\Omega, \mathcal{K}, P)$ *spaţiu de probabilitate.* $X : \Omega \rightarrow \mathbb{R}$ *este o variabilă aleatoare, dacă*

$$\{\omega \in \Omega \,:\, X(\omega) \leq x\} \in \mathcal{K} \text{ pentru fiecare } x \in \mathbb{R}.$$

Variabile aleatoare discrete $X : \Omega \rightarrow \{x_1, x_2, \ldots, x_i, \ldots\}$

Def. 11. ***Distribuţia de probabilitate*** *a v.a. discrete* X

$$X \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_i & \ldots \\ p_1 & p_2 & \ldots & p_i & \ldots \end{pmatrix} = \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

$I \subseteq \mathbb{N}$ *(mulţime de indici nevidă);* $p_i = P(X = x_i) > 0$, $i \in I$, *cu* $\sum\limits_{i \in I} p_i = 1$.

▷ O variabilă aleatoare discretă X este caracterizată de **distribuţia de probabilitate** $P[X]$:

(1) $\quad P[X] : \{x_1, x_2, \ldots\} \rightarrow [0,1]$, definită prin $P[X](x) = P(X = x) \; \forall \, x \in \{x_1, x_2, \ldots\}$.

▷ Notăm $\{X = x_i\} = \{\omega \in \Omega \,:\, X(\omega) = x_i\}$; acesta este un eveniment din $\mathcal{K}$ pentru fiecare $i \in I$.

Distribuţii discrete clasice

Distribuţia discretă uniformă: $X \sim Unid(n),\ n \in \mathbb{N}^*$

$$X \sim \begin{pmatrix} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}$$

Exemplu: Se aruncă un zar, fie X v.a. care indică numărul apărut

$$\Rightarrow X \sim \begin{pmatrix} 1 & 2 & \dots & 6 \\ \frac{1}{6} & \frac{1}{6} & \dots & \frac{1}{6} \end{pmatrix}$$

Matlab/Octave: `unidrnd`$(n, ...)$, `randi`$(n, ...)$ generează valori aleatoare; `unidpdf`(x, n) calculează $P(X = x)$, iar $X \sim Unid(n)$.

Distribuţia Bernoulli: $X \sim Bernoulli(p)$, $p \in (0, 1)$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}$$

Exemplu: în cadrul unui experiment poate să apară evenimentul A (*succes*) sau $\bar{A}$ (*insucces*)
$X = 0 \Leftrightarrow$ dacă $\bar{A}$ apare; $X = 1 \Leftrightarrow$ dacă A apare
$\Rightarrow X \sim Bernoulli(p)$ cu $p := P(A)$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1 - P(A) & P(A) \end{pmatrix}$$

♣

generare în Matlab/Octave:

```
n=1000;
p=0.3;
nr=rand(1,n);
X=(nr<=p) % vector de date avand distributia Bernoulli(p)
%%%%%%%%
Y=floor(rand(1,n)+p)% vector de date avand distributia Bernoulli(p)
%%%%%%%%
```

Distribuţia binomială: $X \sim Bino(n, p), n \in \mathbb{N}^*, p \in (0, 1)$
în cadrul unui experiment poate să apară evenimentul A (*succes*) sau $\bar{A}$ (*insucces*)
• $A =$ succes cu $P(A) = p$, $\quad \bar{A} =$ insucces $P(\bar{A}) = 1 - p$
• se repetă experimentul de n ori
• v.a. X= numărul de succese în n repetări independente ale experimentului $\Rightarrow$ valori posibile: $X \in \{0, 1, \dots, n\}$

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k \in \{0, \dots, n\}.$$

$$X \sim Bino(n, p) \iff X \sim \begin{pmatrix} k \\ C_n^k p^k (1 - p)^{n-k} \end{pmatrix}_{k \in \{0, \dots, n\}}$$

Exemplu: Un zar se aruncă de 10 ori, fie X v.a. care indică de câte ori a apărut numărul 6 $\Rightarrow X \sim Bino(10, \frac{1}{6})$.

$\rightarrow$ are loc formula binomială

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}$$

pentru $a = p$ şi $b = 1 - p$ se obţine

$$1 = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k}.$$

Matlab/Octave: `binornd`$(n, p, ...)$ generează valori aleatoare; `binopdf`(x, n, p) calculează $P(X = x)$, iar $X \sim Bino(n, p)$.

$\triangleright$ Distribuţia binomială corespunde **modelului cu extragerea bilelor dintr-o urnă cu bile de două culori şi cu returnarea bilei după fiecare extragere**:
Într-o urnă sunt n_1 bile albe şi n_2 bile negre. Se extrag cu returnare n bile; fie v.a. X_1 = numărul de bile albe extrase; X_2 = numărul de bile negre extrase
$\Rightarrow X_1 \sim Bino(n, p_1)$ cu $p_1 = \frac{n_1}{n_1+n_2}$, $X_2 \sim Bino(n, p_2)$ cu $p_2 = \frac{n_2}{n_1+n_2}$.

Exemplu: Fie un canal de comunicare binară care transmite cuvinte codificate de N biţi fiecare. Probabilitatea transmiterii cu succes a unui singur bit este p, iar probabilitatea unei erori este $1 - p$. Presupunem, de asemenea, că un astfel de cod este capabil să corecteze până la m erori (într-un cuvânt), unde $0 \leq m \leq N$. Se ştie că transmiterea biţilor succesivi este independentă, atunci probabilitatea transmiterii cu succes a unui cuvânt este $P(A)$, unde
A: cel mult m erori apar în transmiterea celor N biţi

$$P(A) = \sum_{k=0}^{m} C_N^k p^{N-k} (1-p)^k.$$

♠

Exerciţii: 1) Un client accesează o dată pe zi o anumită pagină web, care oferă produse bio, cu probabilitatea 0.4. Cu ce probabilitate clientul accesează această pagină în total de 3 ori în următoarele 6 zile?
2) O reţea de laborator este compusă din 15 calculatoare. Reţeaua a fost atacată de un virus nou, care atacă un calculator cu o probabilitatea 0.4, independent de alte calculatoare. Care este probabilitatea ca virusul a atacat a) cel mult 10 computere; b) cel puţin 10 calculatoare; c) exact 10 calculatoare? ♣

Distribuţia hipergeometrică: $X \sim Hyge(n, n_1, n_2),\ n, n_1, n_2 \in \mathbb{N}^*$
Într-o urnă sunt n_1 bile albe şi n_2 bile negre. Se extrag **fără returnare** n bile.
Fie v.a. X = numărul de bile albe extrase ⇒ valori posibile pentru X sunt $\{0, 1, \ldots, n^*\}$ cu

$$n^* = \min(n_1, n) = \begin{cases} n_1 & \text{dacă } n_1 < n \text{ (mai puţine bile albe decât numărul de extrageri)} \\ n & \text{dacă } n_1 \geq n \text{ (mai multe bile albe decât numărul de extrageri)} \end{cases}$$

Fie $n_1, n_2, n \in \mathbb{N}$ cu $n \leq n_1 + n_2$ şi notăm $n^* = \min(n_1, n)$.

$$\Rightarrow\ P(X = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n}, \quad k \in \{0, \ldots, n^*\}.$$

Matlab/Octave: `hygernd`$(n_1 + n_2, n_1, n, ...)$ generează valori aleatoare;
`hygepdf`$(x, n_1 + n_2, n_1, n)$ calculează $P(X = x)$, iar $X \sim Hyge(n, n_1, n_2)$.

Exemplu: 1) Într-o urnă sunt $n_1 = 2$ bile albe şi $n_2 = 3$ bile negre. Se extrag fără returnare $n = 3$ bile. Fie v.a. X = numărul de bile albe extrase. Vom calcula $P(X = 1)$ cu două metode:
Prima metodă: Pentru $i \in \{1, 2, 3\}$ fie evenimentele
A_i: la a i-a extragere s-a obţinut bilă albă
$N_i = \bar{A}_i$: la a i-a extragere s-a obţinut bilă neagră.
Scriem

$$P(X = 1) = P(A_1 \cap N_2 \cap N_3) + P(A_1 \cap N_2 \cap N_3) + P(A_1 \cap N_2 \cap N_3),$$
$$P(A_1 \cap N_2 \cap N_3) = P(A_1)P(N_2|A_1)P(N_3|A_1 \cap N_2) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{5}$$
$$P(N_1 \cap A_2 \cap N_3) = P(N_1)P(A_2|N_1)P(N_3|N_1 \cap A_2) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{1}{5}$$
$$P(N_1 \cap N_2 \cap A_3) = P(N_1)P(N_2|N_1)P(A_3|N_1 \cap N_2) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{1}{5}$$
$$\Rightarrow P(X = 1) = \frac{3}{5}.$$

A doua metodă: O bilă albă din două se poate alege în $C_2^1 = 2$ moduri, două bile neagre din trei se pot alege în $C_3^2 = 3$ moduri, trei bile din cinci se pot alege în $C_5^3 = 10$ moduri

$$\Rightarrow P(X = 1) = \frac{C_2^1 \cdot C_3^2}{C_5^3} = \frac{2 \cdot 3}{10} = \frac{3}{5}.$$

2) Loto 6 din 49 → Care este probabilitatea de a nimeri exact 4 numere câştigătoare?
R.: Între cele 49 de bile exact $n_1 = 6$ sunt câştigătoare ("bilele albe") şi $n_2 = 43$ necâştigătoare ("bilele negre"). Care este probabilitatea ca din $n = 6$ extrageri fără returnare, exact $k = 4$

numere să fie câştigătoare?

$$\Rightarrow \ P(X=4)=\frac{C_6^4 C_{43}^2}{C_{49}^6}$$

$\diamondsuit$

Distribuţia geometrică $X \sim Geo(p), p \in (0,1)$
În cadrul unui experiment poate să apară evenimentul A (*succes*) sau $\bar{A}$ (*insucces*)
• $A=$ succes cu $P(A)=p$, $\quad \bar{A}=$ insucces $P(\bar{A})=1-p$
• se repetă (independent) experimentul până apare prima dată A ("succes")
• v.a. X arată de câte ori apare $\bar{A}$ (numărul de "insuccese") *până* la apariţia primului A ("succes") $\Rightarrow$ valori posibile: $X \in \{0,1,\dots\}$

$$P(X=k)=p(1-p)^k \quad \text{pentru } k \in \{0,1,2,\dots\}.$$

Matlab/Octave: $\texttt{geornd}(p,...)$ generează valori aleatoare; $\texttt{geopdf}(x,p)$ calculează $P(X=x)$, iar $X \sim Geo(p)$.

Exemplu: X v.a. ce indică numărul de retransmisii printr-un canal cu zgomot (canal cu perturbări) până (înainte de) la prima recepţionare corectă a mesajului; X are distribuţie geometrică. $\spadesuit$

Variabile aleatoare independente

Def. 12. *Variabilele aleatoare discrete X (care ia valorile $\{x_i, i \in I\}$) şi Y (care ia valorile $\{y_j, j \in J\}$) sunt **independente**, dacă şi numai dacă*

$$P(X=x_i, Y=y_j)=P(X=x_i)P(Y=y_j) \quad \forall\, i \in I, j \in J,$$

unde $P(X=x_i, Y=y_j)=P(\{X=x_i\}\cap\{Y=y_j\})\ \forall\, i \in I, j \in J.$

Observaţie: Fie evenimentele $A_i=\{X=x_i\}, i \in I$, şi $B_j=\{Y=y_j\}, j \in J$.
V.a. X şi Y sunt independente $\Longleftrightarrow \forall\,(i,j) \in I \times J$ evenimentele A_i şi B_j sunt independente (a se vedea Def. 6).

Exemplu: Se aruncă o monedă de 10 ori. Fie X v.a. care indică de câte ori a apărut pajură în primele cinci aruncări ale monedei; fie Y v.a. care indică de câte ori a apărut pajură în ultimele cinci aruncări ale monedei. X şi Y sunt v.a. independente. Care este distribuţia de probabilitate a lui X, respectiv Y?

P. 8. *Fie variabilele aleatoare discrete X (care ia valorile $\{x_i, i \in I\}$) şi Y (care ia valorile $\{y_j, j \in J\}$). Sunt echivalente afirmaţiile:*

(1) X *şi* Y *sunt v.a. sunt independente;*
(2) $P(X=x|Y=y)=P(X=x) \quad \forall x\in\{x_i, i\in I\}, y\in\{y_j, j\in J\}$;
(3) $P(Y=y|X=x)=P(Y=y) \quad \forall x\in\{x_i, i\in I\}, y\in\{y_j, j\in J\}$;
(4) $P(X\leq x, Y\leq y)=P(X\leq x)\cdot P(Y\leq y) \quad \forall\, x,y\in\mathbb{R}$.

Def. 13. $\mathbb{X}=(X_1,\ldots,X_m)$ *este un* ***vector aleator discret*** *dacă fiecare componentă a sa este o variabiă aleatoare discretă.*
Fie $K\subseteq\mathbb{N}$ *o mulţime de indici şi fie date* $\mathbb{x}_k := (x_{1,k},...,x_{m,k})\in\mathbb{R}^m, k\in K$.
Dacă $\mathbb{X}:\Omega\to\{\mathbb{x}_k, k\in K\}$ *este un vector aleator discret, atunci*

$$P(\mathbb{X}=\mathbb{x}_k) := P\big(\{\omega\in\Omega : \mathbb{X}(\omega)=\mathbb{x}_k\}\big),\ k\in K,$$

determină ***distribuţia de probabilitate a vectorului aleator discret*** $\mathbb{X}$

$$\mathbb{X}\sim\begin{pmatrix} \mathbb{x}_k \\ P(\mathbb{X}=\mathbb{x}_k)\end{pmatrix}_{k\in K}.$$

$\triangleright$ Vectorii aleatori sunt caracterizaţi de distribuţiile lor de probabilitate! De exemplu, un vector aleator cu 2 componente:

$$\mathbb{X}=(X,Y)\sim\begin{pmatrix}(x_i,y_j)\\ p_{ij}\end{pmatrix}_{(i,j)\in I\times J}$$

unde I, $J\subseteq\mathbb{N}$ sunt mulţimi de indici,
$p_{ij} := P((X,Y)=(x_i,y_j))=P(\{X=x_i\}\cap\{Y=y_j\}), p_{ij}>0\ \forall\, i\in I, j\in J$,
iar $\sum\limits_{(i,j)\in I\times J} p_{ij}=1$.

$\triangleright$ Uneori distribuţia vectorului (X,Y) se dă sub formă tabelară:

$X \backslash Y$	...	y_j	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	...	p_{ij}	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Exemplu: Fie vectorul aleator discret (X,Y) cu distribuţia dată de

următorul tabel:

$X \backslash Y$	0	1
-1	$\frac{1}{4}$	$\frac{1}{2}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$\Longrightarrow P(X=-1,Y=0)=\frac{1}{4}, P(X=-1,Y=1)=\frac{1}{2}$, etc.

a) Să se determine $P(X=-1)$, $P(X\leq 3)$, respectiv $P(Y=1)$, $P(Y\leq -1)$.
b) Sunt X şi Y v.a. independente?

Observaţie: Dacă X şi Y sunt v.a. independente, atunci

(2) $$p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall\, i \in I, j \in J.$$

$\triangleright$ Dacă X şi Y sunt v.a. independente, şi se ştiu distribuţiile lor, atunci distribuţia vectorului aleator (X, Y) se determină pe baza formulei (2).
$\triangleright$ Dacă se cunoaşte distribuţia vectorului aleator (X, Y) distribuţiile lui X şi Y se determină astfel:

$$P(X = x_i) = \sum_{j \in J} p_{ij} \quad \forall i \in I$$

$$P(Y = y_j) = \sum_{i \in I} p_{ij} \quad \forall j \in J.$$

Exemplu:
$\triangleright$ **Modelul urnei cu r culori cu returnarea bilei după fiecare extragere:** fie p_i probabilitatea de a extrage o bilă cu culoarea i, $i = \overline{1, r}$ dintr-o urnă; fie X_i v.a. ce indică numărul de bile de culoarea i, $i = \overline{1, r}$ după n extrageri *cu returnarea bilei extrase*, iar ordinea de extragere a bilelor de diverse culori nu contează

$$\begin{aligned} P(X_1 = k_1, \ldots, X_r = k_r) &= \text{probabilitatea de a obţine } k_i \text{ bile cu culoarea } i,\ i = \overline{1, r}, \\ &\quad \text{din } n = k_1 + \ldots + k_r \text{ extrageri } \textit{cu returnarea} \text{ bilei extrase} \\ &= \frac{n!}{k_1! \ldots k_r!} \cdot p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}, \end{aligned}$$

$\triangleright$ $(X_1, ..., X_r)$ este un vector aleator discret.
$\triangleright$ cazul $r = 2$ corespunde distribuţiei binomiale (modelul binomial cu bile de două culori într-o urnă, a se vedea pg. 18): (X_1, X_2) este un vector aleator discret, iar $X_1 + X_2 = n$; X_1 şi X_2 *nu* sunt v.a. independente.