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)****.*

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

$\triangleright$ la anumite tipuri de algoritmi corectitudinea e garantată doar cu o anumită probabilitate

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

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

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