Curs - Probabilităţi şi Statistică 2023/2024
Secţia Informatică

Facultatea de Matematică şi Informatică
Universitatea Babeş-Bolyai, Cluj-Napoca
Dr. Habil. Hannelore Lisei

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

$\hookrightarrow$ se măsoară *şansele pentru succes* sau *riscul pentru insucces*
al unor evenimente

[Sursa: www.financialmarket.ro]

Fenomene şi procese aleatoare apar, de exemplu, în:

$\rightarrow$ pariuri, loto (6 din 49), jocuri de noroc / jocuri online
$\rightarrow$ previziuni meteo
$\rightarrow$ previziuni economice / financiare, investiţii, cumpărături online (predicţia comportamentului clienţilor)
$\rightarrow$ sondaje de opinie (analiza unor strategii politice), asigurări (evaluarea riscurilor / pierderilor)

$\rightarrow$ **în informatică**:
$\triangleright$ sisteme de comunicare, prelucrarea informaţiei, modelarea traficului în reţea, criptografie;
$\triangleright$ analiza probabilistică a performanţei unor algoritmi, fiabilitatea sistemelor, predicţii în cazul unor sisteme complexe;
$\triangleright$ algoritmi de simulare, machine learning, data mining, recunoaşterea formelor / a vocii;
$\triangleright$ generarea de numere aleatoare (pseudo-aleatoare cu ajutorul calculatorului), algoritmi aleatori
$\triangleright$ `https://www.random.org/randomness/`
se pot genera numere cu ”‘adevarat aleatoare”’ (*true random numbers*), folosind ca sursă un fenomen fizic, ca de exemplu o sursă radioactivă (momentele de timp în care particulele se dezintegrează sunt complet imprevizibile - de exemplu *HotBits service* din Elveţia), sau variaţiile de amplitudine din perturbările atmosferice (atmospheric noise, folosit de Random.org), sau zgomotul de fond dintr-un birou etc.

Exemplu: Generarea de valori aleatoare (în Python)

```
import random
n = random.random()
print("Valoare aleatoare din intervalul (0,1):",n)
N = random.randint(-1,5)
print("Valoare aleatoare din intervalul [-1,5]:",N)
Lista = ["AB", "XY", "EF", "MN","FG"]
```

```
print("O alegere aleatoare din Lista: ", random.choice(Lista))
print("Alegere aleatoare de 6-ori din Lista: ", random.choices(Lista, k=6))
# cu returnare
print("Alegere aleatoare de 4-ori din Lista (fara returnare): ", random.sample(Lista,4))
# fara returnare
```

```
import secrets
x = secrets.randbelow(49)
print("Valoare aleatoare < 49: ",x)
print(secrets.choice("ABCDEF"))
```

```
import numpy as np
N=30
R = np.random.randint(1,7,size=N)
#un vector cu N de elemente, din multimea {1,2,...6}
print("Numere obtinute la aruncarea de",N,"-ori a zarului:\n",R)
total= sum(R==6)
print("De cate ori s-a obtinut 6 in ",N," aruncari ale zarului:",total)
```

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, uneori 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 aleator pentru care rezultatele obţinute sunt corecte *doar* cu o anumită probabilitate se numeşte algoritm **Monte Carlo**.
↪ 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 un vector cu 300 de elemente, din mulţimea $\{0, 1, 2\}$ (ordinea lor este necunoscută; se presupune că şirul conţine cel puţin un 0).
$\longrightarrow$ De care tip este următorul algoritm (scris în Python)?

```
import numpy as np
print('Prima versiune')
N=300
S = np.random.randint(0,3, size = N)
#print(S)
#un vector cu N elemente aleatoare din multimea {0,1,2}
k=1
i= np.random.randint(low=0, high=N)
while S[i] != 0:
    print("iteratia:",k)
    print("S[",i,"]=",S[i])
    i= np.random.randint(low=0, high=N)
    k=k+1
# k = numar iteratii pana se gaseste aleator un 0
if S[i]==0:
    print("iteratia:",k)
    print("S[",i,"]=",S[i])    # i indicele, pentru care S[i]=0
print("La iteratia",k,"s-a gasit aleator un 0.")
```

Răspuns: Algoritm de tip Las Vegas.

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

```
import numpy as np
print("a doua versiune")
N=50
S = np.random.randint(3,size=N)
print(S)
#un vector cu N  elemente, din multimea {0,1,2}
M=3  #nr maxim de iteratii  M>1
a=True
for k in range(M) :
    print("iteratia:",k+1)
    i= np.random.randint(low=0, high=N)
    print("S[",i,"]=",S[i])
    if S[i] == 0:
        print("la iteratia",k+1,"s-a gasit aleator un 0.")
        a=False
        break
if a:
    print("In",k+1,"iteratii nu s-a gasit niciun 0.")
```

$\triangleright$ 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 disjuncte (incompatibile), 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 $A \setminus B = A \cap \bar{B}$.

• Au loc relaţiile: $A \cup \bar{A} = \Omega, \quad A \cap \bar{A} = \emptyset, \quad \bar{\bar{A}} = A.$

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**

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

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*

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

▷ 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: $f_n(A) = \frac{22}{100} = 0.22$ frecvenţa relativă a evenimentului A

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

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

Exemplu istoric - 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. ♠