Simularea pe calculator a repartiţiilor uniforme. Repartiţii discrete clasice

Prezenta lucrare are drept scop simularea pe calculator a unor densităţi de repartiţie discretă, în vederea definirii intuitive a conceptului de probabilitate. Probabilitatea este o funcţie ce asociază fiecărui element dintr-un câmp de evenimente o valoare cuprinsă în intervalul [0,1]. Se va face o corelaţie între conceptul de densitate de probabilitate şi cel de "frecvenţă de apariţie a unui rezultat". Se urmăreşte de asemenea reprezentarea unor repartiţii discrete clasice, cum ar fi repartiţiile binomială si Poisson.

Consideraţii teoretice

Variabila aleatoare este definită ca o funcţie măsurabilă în raport cu un anumit corp de evenimente. Se asigură astfel posibilitatea de a calcula probabilităţile tuturor evenimentelor ce prezintă interes în legătură cu o anumită variabilă aleatoare. Dacă mulţimea valorilor posibile este o mulţime discretă atunci variabila aleatoare este la rândul ei discretă. În caz contrar avem o variabilă continuă.

Fie $\{x_i \mid i\in I\}$ mulţimea de valori posibile a variabilei aleatoare ξ şi mulţimea evenimentelor posibile $\{E_i\}_{i\in I}=\{\xi=x_i\}_{i\in I}$. Şirul de perechi de valori $\begin{pmatrix} x_i \\ P(\xi = x_i) \end{pmatrix}_{i\in I}$ se numeşte *tabel de repartiţie (distribuţie)* al variabilei aleatoare ξ.

Funcţia de repartiţie. Funcţia $F(x):R\to[0,1]$; $F(x)=P[\xi<x]$ se numeşte funcţie de repartiţie a variabilei aleatoare ξ. Pentru o variabilă aleatoare discretă, funcţia de repartiţie este: $F(x)=\sum_{x_i<x} P(\xi=x_i)$ care arată ca F(x) este o funcţie discontinuă, în scară, ale cărei creşteri au loc în funcţie de x_i, cu valorile $p_i=P(E_i)$.

Momentul de ordin k. Dacă variabila ξ este discretă, $\xi=\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $\sum_{i\in I} p_i = 1$, atunci momentul de ordin k este $M[\xi^k]=\sum_{i\in I} x_i^k p_i$. Pentru k=1, $M[\xi]$ se numeşte *media* variabilei ξ.

Repartiţii statistice

Statistica matematică are drept scop cercetarea şi perfecţionarea metodelor de analiză a datelor experimentale referitoare la un anumit fenomen. Rezultatele măsurătorilor efectuate asupra caracteristicii unei populaţii se prezintă sub forma unor serii statistice a căror dimensiune depinde de volumul eşantionului extras din populaţie. Caracteristica examinată, care este în fond o variabilă aleatoare, va fi descrisă corespunzător de o repartiţie spaţială, temporală sau în frecvenţă. În cele ce urmează ne vom referi la repartiţiile în frecvenţă.

Frecvenţa absolută n_i reprezintă numărul de apariţii al unui rezultat în cele n experimente efectuate asupra eşantionului. Frecventa relativa f_i reprezintă frecvenţa absolută raportată la volumul eşantionului. Rezultatele măsurătorilor $(x_1,x_2,..,x_n)$ pot fi organizate în trei tipuri de serii statistice, şi anume:

$$(S1) \quad x_1 < x_2 < ... < x_n; \qquad n_i = 1;\ f_i = 1/n;\ i=\overline{1,n}$$

$$(S2) \quad x_1 < x_2 < ... < x_k; \qquad n_i \neq 1;\ f_i = n_i/n;\ i=\overline{1,k}$$

în care n este volumul eşantionului, n_i sunt frecvenţele absolute de apariţie a valorilor x_i corespunzătoare şi f_i sunt frecvenţele relative.

Rezultatele măsurătorilor pot fi grupate în k intervale de valori, de lungime egală, fiecărui interval corespunzându-i un reprezentant $\hat{x}_i\ (i=\overline{1,k})$, frecvenţele absolute corespunzătoare fiecărui interval (reprezentant) fiind numărul de valori ale caracteristicii măsurate din intervalul respectiv. În acest caz putem construi o serie de tipul:

$$(S3) \quad \hat{x}_1 < \hat{x}_2 < ... < \hat{x}_k;\ n_i \neq 1;\ f_i = n_i/n;\ i=\overline{1,k}$$

care nu se deosebeşte de (S2) decât prin faptul că in serie apar reprezentanţii intervalelor. În toate cazurile suma frecvenţelor relative este unitară. Dacă notăm cu ξ caracteristica examinată, ea poate fi caracterizată prin distribuţia în frecvenţe:

$$\xi = \begin{pmatrix} x_i \\ f_i \end{pmatrix}_{i=\overline{1,n}} \quad \text{unde} \sum_{i=1}^{n} f_i = 1$$

sau de funcţia empirică de repartiţie, notată $F_n^*(x)$, care pentru seriile de tipul S_1 şi S_2 este de forma:

$$F_n^* = \begin{cases} 0 & daca\ x < x_1 \\ \sum_{j=1}^{i-1} f_j & daca\ x_{i-1} \le x < x_i \quad i = \overline{1, k+1} \end{cases}$$

iar pentru seriile de tipul S_3 este de forma:

$$F_n^* = \begin{cases} 0 & daca\ x < l_0 \\ \sum_{j=1}^{i-1} f_j + \frac{x - l_{i-1}}{d} f_i & daca\ l_{i-1} \le x < l_i \quad i = \overline{1, k+1} \end{cases}$$

d fiind lungimea intervalelor *(l_{i-1}, l_i)* şi k numărul acestor intervale.

Funcţia empirică de repartiţie este deci o funcţie scară similară funcţiei de repartiţie a unei variabile aleatoare discrete finite (vezi figura 1).

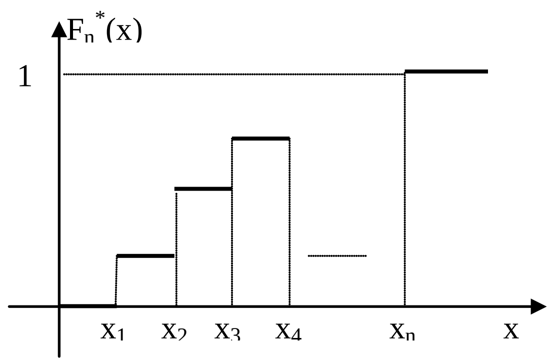

Fig 1 Funcţia empirică de repartiţie

Legi clasice de repartiţie discretă

Repartiţia uniformă

Tabelul de repartiţie al unei variabilei aleatoare ξ uniform distribuite este de forma

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Repartiţia binomială

Tabelul de repartiţie al unei variabilei aleatoare ξ distribuite binomial este de forma:

$$\xi = \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=0,1,\dots,n}$$

ξ este o variabilă aleatoare reprezentând numarul de apariţii ale unui eveniment E în n experimente independente (extragere cu revenire), unde *p* este probabilitatea realizării evenimentului E iar *q* este probabilitatea realizării evenimentului complementar non E (q=1-p).

Următoarele experimente sunt de tip Bernoulli:

1) aruncarea unei monezi de un număr de ori (*n*);
2) un sondaj de opinie efectuat pe un eşantion de n persoane pentru care se solicită doar două răspunsuri la întrebările formulate (da sau nu);

Repartiţia Poisson

Tabelul de repartiţie al unei variabilei aleatoare ξ distribuite Poisson este de forma:

$$\xi = \begin{pmatrix} n \\ \frac{\lambda^n}{n!} e^{-\lambda} \end{pmatrix}_{n \in N}$$

Repartiţia Poisson se obţine ca un caz limită al repartiţiei binomiale în cazul în care n este suficient de mare, p mic (p<0.1). Se foloseşte spre exemplu în controlul statistic al calităţii, când probabilitatea obţinerii unei piese defecte este foarte mică.

Probleme propuse

1. Să se genereze 100 de numere aleatoare uniform distribuite în intervalul (0,100).Rezultatele se grupează în k=10 intervale de lungimi egale. Fiecărui interval îi corespunde un reprezentant care este prima valoare din intervalul respectiv, frecvenţele absolute corespunzătoare fiecărui reprezentant fiind numărul de valori din intervalul respectiv. Să se reprezinte grafic curba de frecvenţe şi funcţia de repartiţie empirică.

2. Sa se calculeze şi să se afişeze densitatea de repartiţie binomială pentru *k* aparţinând intervalului [1,*n*](*p*=1/2, *n* = 20; 40; 80; 160;).

Obs. Distribuţia binomială poate fi simulată pe calculator astfel:

```
% program BINOMIAL
n=input('nr. încercări');
p=input('prob. ev. elem.');
k=input('nr. reuşite solicitate');
q=1-p;
coef=prod(1:n)/(prod(1:k)*prod(1:n-k));
binomial=coef*p^k*q^(n-k);
disp(binomial)
```

3. Se consideră o centrală cu 10 canale telefonice care deserveşte 20 de abonaţi (10 la un capăt, 10 la celălalt). Să se aprecieze eficienţa cu care sunt utilizate aceste canale dacă durata medie a ocupării unei linii este de 12 minute pe oră.

Precizări. Eficienţa centralei se apreciază prin probabilitatea de a găsi k linii ocupate din cele n.

Se defineşte o variabilă aleatoare ξ=”nr de linii ocupate simultan”. Această variabilă aleatoare are o distribuţie binomială cu p=12/60=0.2 – probabilitatea de a găsi o linie ocupată.

4. Să se reprezinte grafic curba de frecvenţă a repartiţiei Poisson pentru λ=0.9 respectiv λ=0.5.

5. Jocul ”Cap sau pajură”

Fie cazul generării unei variabile aleatoare distribuite uniform, ca rezultat al simulării jocului "cap sau pajură". Intuitiv, probabilitatea de a obţine una din feţele monezii este 1/2. Vom simula acest joc, cerând calculatorului să aleagă un număr aleator, uniform distribuit în intervalul (0,1). Dacă numărul ales (probabilitatea de apariţie a numărului) este mai mică decât 1/2, vom atribui ieşirii valoarea "cap"; în caz contrar va fi "pajură". Vom simula pentru:

a. 20 aruncări ale monezii;
b. 1000 de aruncări ale monezii;
c.

```
% Simularea jocului cu monedă "cap sau pajură"
% pentru 20 de aruncări ale monezii.
 p=1/2;
 for nrar=1:20
        if rand<p,
         var(nrar)='c';
        else
         var(nrar)='p';
        end
end
var
```

Observaţie:

Din 20 de aruncări se obţin numere diferite de apariţie ale celor 2 feţe ale monezii, ce ar putea conduce la ideea greşită că probabilitatea de apariţie a fiecărei feţe nu este 1/2.

❖ Să se modifice programul anterior astfel încât să se simuleze, pe baza definiţiei frecvenţiale, densitatea de probabilitate de apariţie a fiecărei feţe în jocul aruncării monezii.

6. Aruncarea zarurilor

Vom experimenta diferitele pariuri ale lui de Mere, simulând pe calculator aruncarea zarurilor. Aceasta implică alegerea unui întreg între 1 şi 6, în mod aleator. Pentru aceasta vom alege mai întâi un număr aleator, uniform distribuit în intervalul (0,1). Apoi îl vom multiplica cu 6, obţinând un număr aleator intre 0 şi 6, vom lua partea întreagă a acestui număr şi îi vom adăuga 1 (se presupune că numărul aleator poate avea valoarea 0, dar nu poate avea niciodată valoarea 1, prevenind astfel ieşirea 7). Vom folosi această metodă pentru a simula un număr mare de experimente, observând în fiecare, dacă apare 6 la 4 aruncări ale zarului:

```
% program de MERE
% ce calculează proporţia de apariţie a lui 6 într-un număr de jocuri,
% fiecare joc  constând din 4 aruncări ale zarului
nr_jocuri=input(' Număr jocuri: ');
noroc=0; out=fix(6*rand)+1;
for n=1:nr_jocuri
    nr_aruncări=1;
   while out~=6 & nr_aruncări~=4
      out=fix(6*rand)+1;    nr_aruncări=nr_aruncări+1;
   end        % end while
   if out==6
     noroc=noroc+1;     out=fix(6*rand)+1;
   end       % end if
end         % end for
disp(noroc/nr_jocuri)
```

Dacă se rulează pentru un numar mic de jocuri, rezultatul afişat va fi mai mic de 0.5. Se observă că simularea trebuie realizată pentru un număr cât mai mare de experimente.

❖ Să se modifice programul anterior astfel încât să se simuleze probabilitatea de reuşită în cel de-al doilea pariu de Mere (probabilitatea de apariţie a unei perechi (6,6) din 24, respectiv 25 de aruncări ale zarurilor).

7. Câştigurile lui Peter şi Paul la aruncarea monezii

Vom considera o problema în care răspunsul exact este dificil de obţinut, dar pentru care, de această dată, simularea oferă rezultate calitative uşor de apreciat:

Peter şi Paul joacă "cap şi pajură"; de fiecare dată când apare "cap" va câştiga Peter un leu, iar când apare "pajura" va câştiga Paul un leu; dorim să determinăm cu ce probabilitate va câştiga Peter j lei (j poate fi orice număr între -nr aruncări şi +nr aruncări) şi de câte ori va fi el în frunte. Se va face următoarea convenţie: dacă câştigul lui Peter este 0, el este conducător, dacă a condus în aruncarea precedentă; se consideră că un joc se desfăşoară pe perioada de 12 aruncări:

```
% programul simulare Peter şi Paul simulează aruncarea monedei
% şi estimează câştigul şi probabilitatea acestuia
 p=1/2;
 nr_jocuri=input('Număr jocuri: ');
 contor_conduceri=zeros(1,13);
 contor_câştiguri=zeros(1,25);
 prob_conduceri=zeros(1,13);
for n=1:nr_jocuri,
    nr_câştiguri=0;
    ultimul_câştig=0;
    nr_conduceri=0;
    for aruncări=1:12
      if rand<p
          nr_câştiguri=nr_câştiguri+1;    % dacă apare cap Peter câştigă
```

```
      else
         nr_câştiguri=nr_câştiguri-1;
      end     % end if
        if (nr_câştiguri>0 | nr_câştiguri==0 & ultimul_câştig>0)
                         nr_conduceri=nr_conduceri+1;
      end    % end if
      ultimul_câştig=nr_câştiguri;
    end    % end for
    % evidenta avantajelor
   contor_conduceri(nr_conduceri+1)= contor_conduceri(nr_conduceri+1)+1;
    % evidenta câştigurilor
   contor_câştiguri(nr_câştiguri+13)= contor_câştiguri(nr_câştiguri+13)+1;
end % end for
for nr_conduceri=0:12
prob_cond(nr_conduceri+1)=contor_conduceri(nr_conduceri+1)/nr_jocuri;
end % end for
prob_cond
```

Simularea sugerează că cel mai improbabil număr de conduceri este 6 şi cele mai probabile sunt 0 sau 12. Peter va rămâne în urma în cea mai mare parte a timpului; dacă jocul ar mai continua, câştigul lui Peter ar reveni la 0, dar într-un timp relativ mare.

❖ Generalizare: Să se scrie un program ce simulează probabilitatea de apariţie a rezultatelor 1,2,...,n ale unui experiment, pentru care probabilităţile corespunzătoare de producere sunt: $p_1, p_2, ..., p_n$.
Afişarea se va face pe 3 coloane ce vor reprezenta: rezultatul, probabilitatea, frecvenţa relativă de apariţie

8. Cursele de cai

Patru cai (Acorn, Balky, Chesnut, Dolby) au participat la numeroase curse. S-a estimat că:
-Acorn a câştigat 30% din curse;
-Balky a câştigat 40% din curse;
-Chesnut a câştigat 20% din curse;
-Dolby a câştigat 10% din curse.

❖ Să se scrie un program care să simuleze rezultatul a 10 curse.

9. Ruleta din Las Vegas

Roata ruletei din Las Vegas are 38 fante numerotate: 0, 00, 1, 2,...36. Fantele 0, 00 sunt verzi. Jumătate din fantele rămase sunt roşii şi jumătate sunt negre. Crupierul învârte roata şi aruncă o minge de fildeş. Dacă pariaţi 1$ pe roşu, veţi câştiga 1$ dacă mingea se opreşte pe roşu; în caz contrar veţi pierde 1$. Dacă mingea ajunge pe una din fantele verzi veţi pierdeţi toată suma deţinută în acel moment .Scrieţi un program ce estimează şi afişează:

- probabilitatea de câştig pentru nr_jocuri=1:1000;
- probabilitatea de a termina jocul pe un câştig de 1$, 0$, respectiv 5$.

10. Se ştie ca dacă w este o variabilă aleatoare discretă repartizată binomial, atunci:

$$P(w < A) = \sum_{k=1}^{[A]} C_n^k \cdot p^k \cdot q^{n-k}$$

Modificaţi programul de calcul al densităţii de probabilitate binomiale BINOMIAL astfel încât acesta să calculeze probabilitatea ca numărul de succese în n încercări independente să fie cuprinsă între *A* şi *B*; folosiţi programul astfel creat pentru determinarea probabilităţii ca dintr-o 100 de aruncări ale unei monezi, numărul de apariţii ale feţei "cap" să fie cuprins :

a) între 35 şi 65;
b) între 40 şi 60;
c) între 45 şi 55.