Repartiţii continue clasice

1. Repartiţia normală

Dacă o variabilă aleatoare $\xi \in N(m,\sigma^2)$ are densitatea de repartiţie de forma: $f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi})}\cdot e^{\frac{-(x-m)^2}{2\cdot\sigma^2}}$, cu $x\in R$, atunci **repartiţia** se numeşte **normală**.

- Graficul acestei repartiţii este celebrul **clopot al lui Gauss**.
- Parametrul m este **media** variabilei ξ ; densitatea de repartiţie este simetrică faţă de dreapta x = m.
- Parametrul σ este **abaterea medie pătratică** a variabilei ξ şi reprezintă o măsură a gradului de împrăştiere a densităţii.
- Pătratul abaterii medii pătratice (σ^2) este **dispersia variabilei aleatoare**.
- Funcţia de repartiţie corespunzătoare densităţii normale este:

$$F(x)=\int_{-\infty}^{x} f(u)du = P[\xi < x],$$

unde F(x) reprezintă, de fapt, suprafaţa mărginită de f(x) şi axa x între limitele: $u_{min}= -\infty$ şi $u_{max} = x$.

Se ştie că: $F(+\infty)=1$ şi $F(-\infty) = 0$. Se pot concepe programe Matlab de reprezentare a densităţii de repartiţie, respectiv a funcţiei de repartiţie normale (sau gaussiene). Pentru reprezentarea funcţiei de repartiţie F(x) se poate înlocui, practic, limita $-\infty$ cu cantitatea ($m-4\sigma$), deoarece peste această limită F(x) tinde să se anuleze.

- Realizaţi un program care să reprezinte grafic densitatea de repartiţie normală pentru m=0,1,2,3 şi σ = 0.5, 1, 2.
- Scrieţi un program care să determine aria cuprinsă între curba densităţii de repartiţie normale (m=0, σ=1), axa x şi verticalele corespunzătoare punctelor $x_{min}=0$ şi $x_{max}=\sigma$. Analog pentru $x_{max} = 2\sigma$ respectiv 3σ. Ce semnificaţie are această arie?

Indicaţie
Se va aproxima suprafaţa de calcul cu sume de dreptunghiuri cu laturile dx=0.01 şi înălţimea f(x).

- Pe baza programului anterior, scrieţi în Matlab funcţia arienormala(0,x_{max}). Tabelaţi valorile returnate de funcţie pentru dx_afisare=0.1.
- Modificaţi funcţia arienormala(0,x_{max}) pentru a genera funcţiile:
  a) arienormala($-x_{max}$,x_{max})
  b) arienormala($-\infty$,x_{max})

Ce reprezintă valorile returnate de funcţiile de la punctele a) şi b) ?
Cum folosiţi funcţiile de mai sus pentru a determina: $P[\xi > 50]$, $P[60 < \xi < 80]$, unde $\xi\in N(0,1)$?

Din forma analitică a funcţiei de repartiţie normale rezultă imposibilitatea soluţionării explicite a unei ecuaţii de forma F(x)=y. Acest fapt a impus dezvoltarea unor metode speciale de evaluare a soluţiei $x = F^{-1}(y)$.
Una din aceste metode are la baza propoziţia:

Dacă U şi V sunt 2 variabile aleatoare independente, distribuite uniform în intervalul [0,1], atunci variabilele aleatoare:

$$X=\sqrt{-2\cdot\ln(U)}\cdot\cos(2\pi\cdot V) \quad Y=\sqrt{-2\cdot\ln(U)}\cdot\sin(2\pi\cdot V)$$

sunt independente şi distribuite normal, de medie m = 0 şi dispersie $\sigma^2=1$. Pentru X şi Y definim funcţiile de repartiţie:

$$\Phi(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du \qquad \Phi(Y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{u^2}{2}}du$$

Astfel, pentru a obţine o variabilă aleatoare Z, normal distribuită, dar cu $m\neq 0$ şi $\sigma \neq 1$ se va folosi relaţia $Z = \sigma X + m$ ce va determina:

$$F(z) = P[Z \le z] = P[s \cdot X + m \le z] = P[X \le \frac{z-m}{\sigma}] = \Phi(\frac{z-m}{\sigma})$$

$$f(z) = \frac{d\Phi(z)}{dz} = \frac{d\Phi(\frac{z-m}{\sigma})}{d(\frac{z-m}{\sigma})} \cdot \frac{d(\frac{z-m}{\sigma})}{dz} = f_{N(0,1)}(\frac{z-m}{\sigma}) \cdot \frac{1}{\sigma} = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\frac{-(z-m)^2}{2 \cdot \sigma^2}}$$

Observaţii:

1. F este cunoscută sub denumirea generică de "repartiţie normală redusă".
2. Relaţia $P[m - k\sigma < Z < m + k\sigma] = P[-k < X < k] = 2\Phi(k) - 1$ pentru k=3 devine:
   $P[m-3\sigma < Z < m+3\sigma] = 0 .997$, deci 99,7% din valorile variabilei aleatoare Z aparţin intervalului.

2. Repartiţii ce au la baza repartiţia normală (Maxwell, Gamma, χ^2)

Pornind de la o variabilă repartizată normal, în unele experimente se poate ajunge la variabile cu distribuţii diferite de cea normală (Gamma, Maxwell, χ^2), care au următoarele densităţi de repartiţie:

Denumire	**Domeniu**	**Densitate de repartiţie**
Gamma $\Gamma(\alpha,\lambda)$:	$[0,\infty)$	$f(x) = \frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1} e^{-\lambda x}$, $\alpha,\lambda>1$
Maxwell	$[0,\infty)$	$f(x) = x \cdot e^{-\frac{x^2}{2}}$
$\chi^2(n,\sigma)$	$[0,\infty)$	$f(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\sigma^n \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2\sigma^2}}$

unde $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$

De exemplu, presupunem că lăsăm să cadă un corp pe o suprafaţă plană (punctul din care aruncăm se află deasupra originii). În aceste considerente coordonatele x şi y ale poziţiei de intersecţie a corpului cu suprafaţa plană pot fi considerate normal distribuite de medie m = 0 şi dispersie data (să o presupunem pentru moment σ = 1). Ne punem întrebarea ce distribuţie are distanţa de la punctul de cădere la originea planului?

O problemă similară apare în fizică atunci când presupunem mişcarea plană a unei particule, ce are coordonatele vitezei v_x, respectiv v_y normal distribuite şi pentru care dorim să stabilim distribuţia vitezei particulei:

$v = \sqrt{v_x{}^2 + v_y{}^2}$. Se obţine că variabila v este repartizată Maxwell.

Dacă luăm în considerare cazul unei ţinte circulare pentru săgeţi şi un experiment de aruncare la ţintă, putem alege în mod aleator perechea de coordonate (x, y) (fiecare variabilă fiind repartizată normal de medie m=0 şi dispersie σ = 1) ca reprezentând punctul de intersecţie al vârfului unei săgeţi cu suprafaţa circulară a ţintei. Distanţa $r = \sqrt{x^2 + y^2}$ a punctului (x, y) faţă de origine este o variabilă repartizată Maxwell.

Observaţii:

- dacă x, y∈N(0,1), atunci $R^2 = x^2 + y^2 \in \Gamma(\frac{1}{2},1)$.

- dacă ξ , $\eta \in$N(0,1), atunci $\xi^2 \in \chi^2(1,1)$, $\eta^2 \in \chi^2(1,1)$ şi $R^2 = \xi^2 + \eta^2 \in \chi^2(2,1) = \Gamma(\frac{1}{2},1)$.
- dacă R^2 este repartizată Gamma, atunci R este repartizată Maxwell.

Funcţia de repartiţie în variabila R:

$$F_R(x) = P[R < x] = P[R^2 < x^2] = \int_0^{x^2} \frac{1}{2\Gamma(1)} \cdot e^{-\frac{u}{2}} du = \frac{1}{2}\int_0^{x^2} e^{-\frac{u}{2}} du = 1 - e^{-\frac{x^2}{2}}$$

$$f_R(x) = \frac{dF_R(x)}{dx} = \begin{cases} x \cdot e^{-\frac{x^2}{2}}, \text{pentru } x \geq 0 \\ 0, \quad \text{pentru } x < 0 \end{cases},$$

care este densitatea de repartiţie Maxwell.

Programul următor reprezintă implementarea repartiţiei Maxwell:

```
 nr_aruncari=input(' Număr aruncari : ')
 dens=zeros(1,60);
 for i=1:nr_aruncari
        x=randn;  y=randn;
        d=sqrt(x^2+y^2);
        for j=1:60
        if(d<=j*0.05)
          dens(j)=dens(j)+1;
          break;
        end
        end
 end
subplot(2,1,1), plot(dens/nr_aruncari,'w'), title('repartiţie reala')
 r=0.0:0.05:3.0; subplot(2,1,2)
 dens=r.*exp((-1/2)*r.^2);
 plot(dens,'w'), title('repartiţie teoretica')
```

Se va rula programul pentru număr de aruncări = 15; 20; 100.

Problemă propusă: Să se realizeze o implementare în Matlab pentru problema aruncării la ţintă din exemplul expus anterior.

3. Repartiţia exponenţial negativă

Dacă o variabilă aleatoare ξ are densitatea de repartiţie de forma: $f(x) = \lambda e^{-\lambda x}$, $x \in [0,\infty)$, $\lambda > 0$, sau funcţia de repartiţie $F(x) = P[\xi < x] = \begin{cases} 1 - e^{-\lambda \cdot x}, & x \geq 0 \\ 0, & \text{in rest} \end{cases}$, atunci repartiţia se numeşte exponenţial negativă.

Acest tip de variabilă aleatoare apare, de regulă, în descrierea experimentelor ce implică o întrebare de forma: "Care este intervalul de timp scurs până la apariţia unui anume rezultat?".

Ex. 1: Să se simuleze şi să se vizualizeze densitatea de repartiţie şi funcţia de repartiţie exponenţial-negativă pentru λ = 1/2; 1/10; 1/50 şi x= 0:200.

Ex. 2: Un experiment, frecvent întâlnit, bazat pe repartiţia exponenţial-negativă este următorul:

Presupunem că dorim să luăm un autobuz ce trebuie să sosească în staţie la ora 12:00. Acest autobuz soseşte în staţie la aproximativ fiecare 30 de minute, dar de obicei întârzie. Am putea considera că vom ajunge în staţie în mijlocul unui astfel de interval de timp, astfel încât să nu avem de aşteptat decât 15 minute sau, pe de altă parte, am putea considera că autobuzul circulă aşa de aleator aşa încât nu contează când ajungem în staţie şi că vom avea de aşteptat o perioadă de 30 de minute. Aceste două conjuncturi aparent rezonabile, dar şi contradictorii în acelaşi timp au definit problema BUS PARADOX.

Ţinând cont că autobuzul este aşteptat să ajungă din minut în minut avem toate motivele să ne întrebăm cât vom avea de aşteptat.
Presupunem că în fiecare minut după ora 12:00, dacă autobuzul nu a ajuns încă, el va ajunge cu probabilitate 1/30. Pentru a scrie un program de simulare, vom alege un număr aleator în intervalul [0,1] şi îl vom testa dacă este mai mic decât 1/30; dacă nu este vom genera altă valoare până vom obţine prima valoare ce îndeplineşte condiţia (atunci autobuzul va sosi). Vom crea un contor i, ce ţine evidenţa numărului de minute pe care îl avem de aşteptat până la sosirea autobuzului:

```
%Experimentul "Bus paradox"
 % Câte minute, după ora 12:00 va trebui să aşteptăm autobuzul
 % dacă,  în  fiecare  minut el va ajunge cu probabilitatea 1/30
 nr_experimente= input('Număr experimente(<=1000) : ');
 contor=zeros(1,nr_experimente);
 for n=1:nr_experimente,
   x=rand;    c=1;
   while x>=(1/30),
            c=c+1;    x=rand;
   end
   contor(n)=c;
 end;
 f=zeros(1,120);
 for i=1:120
 for n=1:nr_experimente
 if contor(n) == i,
            f(i)=f(i)+1;
 end;  end;  end;
 plot((1/nr_experimente)*f);
```

Experimentul "Bus-paradox" pune în evidenta o modalitate de simulare a repartiţiei exponenţial-negative, în care legea de repartiţie continuă este înlocuită cu o aproximaţie discretă (distribuţia geometrică).

Ex. 3: Scrieţi o funcţie Matlab: ariexp(l,y) ce returnează F(y) corespunzătoare repartiţiei exponenţial negative, folosind o aproximare discretă pentru densitatea de repartiţie corespunzătoare (vezi programul "Bus paradox").

Folosiţi această funcţie pentru a rezolva următoarele probleme:

a) presupunem că timpul de bună funcţionare al unui autoturism este repartizat exponenţial-negativ cu parametrul λ=1/4; cu ce probabilitate posesorul unui astfel de autovehicul îl va avea în stare de funcţiune după 4 ani?
b) presupunem că timpul de reparaţie (în ore) a unei anumite componente a unui sistem este repartizat exponenţial-negativ cu λ=1/2; cu ce probabilitate timpul de reparaţie va depăşi 4 ore?

Extinzând repartiţia exponenţial negativă la suma a 2 variabile:

$$Z = X + Y, \text{ cu: } f_X(x) = f_Y(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & 0 \le x \\ 0 & x < 0 \end{cases}$$

vom obţine, conform teoremei de convoluţie a densităţilor de probabilitate:

$$f_Z(x) = \int_{-\infty}^{+\infty} f_X(x-y)\cdot f_Y(y)\cdot dy = \int_{-\infty}^{+\infty} \lambda \cdot e^{-\lambda\cdot(x-y)}\cdot \lambda \cdot e^{-\lambda y}\cdot dy = \begin{cases} \int_0^x \lambda^2 \cdot e^{-\lambda x} dy = \dfrac{\lambda^2}{\Gamma(2)}\cdot e^{-\lambda x}\cdot x, \text{ pentru } x \ge 0 \\ 0, \text{ pentru } x < 0 \end{cases}, \text{ adica } Z \in \Gamma(2,\lambda)$$

Exerciţiu 5: Să se vizualizeze $f_Z(x)$, pentru x = 0:10.

Exerciţiu 6: Să se determine şi să se reprezinte grafic:

a) $f_{Z1}(x)$, $Z_1 = X - Y$, unde $f_X(x) = f_Y(x) = \lambda e^{-\lambda x}$, $x \ge 0$.
b) $f_{Z2}(z)$, $Z_2 = X + Y$, $f_X(x) = \lambda e^{-\lambda x}$, $f_Y(x) = me^{-mx}$.