

Laborator 8 – Probabilități și Statistică Matematică

GENERAREA UNEI VARIABLE ALEATOARE DISCRETE

Definiți o funcție care să genereze un esantion de talie n dintr-o distribuție discretă definită pe multimea $\{x_1, \dots, x_N\}$ cu probabilitățile $\{p_1, \dots, p_N\}$.

Avem următoarea funcție:

```
GenerateDiscrete = function(n = 1, x, p, err = 1e-15){  
  # talia esantionului  
  
  # x alfabetul  
  
  # p probabilitatile  
  
  lp = length(p)  
  lx = length(x)  
  
  # verify if x and p have the same size  
  if(abs(sum(p)-1)>err | sum(p>=0)!=lp){  
    stop("suma probabilitatilor nu este 1 sau probabilitatile sunt mai mici decat 0")  
  }else if(lx!=lp){  
    stop("x si p ar trebui sa aiba aceeasi marime")  
  }else{  
    out = rep(0, n)  
  
    indOrderProb = order(p, decreasing = TRUE) # index  
    pOrdered = p[indOrderProb] # rearrange the values of the probabilities  
    xOrdered = x[indOrderProb] # rearrange the values of x  
  
    # u = runif(n) # generate n uniforms  
  
    pOrderedCS = cumsum(pOrdered)  
  
    for (i in 1:n){
```

```

u = runif(1)

k = min(which(u<=pOrderedCS))

out[i] = xOrdered[k]

}

}

return(out)

}

```

Pentru a testa această funcție să considerăm următoarele două exemple:

1. Ne propunem să generăm observatii din $X \sim \begin{pmatrix} 1 & 2 & 3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$, în acest caz:

$x = [1,2,3]$ si $p = [0.2,0.3,0.5]$. Începem prin generarea a $n = 10$ observatii din repartitia lui X:

```

GenerateDiscrete(10, c(1,2,3), c(0.2,0.3,0.5))
[1] 3 2 3 3 3 1 2 3 1 2

```

Plecând de la un esantion de $n = 10000$ de observatii vrem să comparăm, cu ajutorul diagramei cu bare verticale (barplot), repartitia esantionului cu cea teoretică :

```

n = 10000

x = GenerateDiscrete(n, c(1,2,3), c(0.2,0.3,0.5))

# cate observatii din fiecare valoare unica a lui x

pX = table(x)/n

pT = c(0.2,0.3,0.5)

indX = c(1,2,3)

barplot(rbind(pX, pT),

        beside = T,

        space = c(0.1, 1),

        col = c(myblue, myred),

```

```

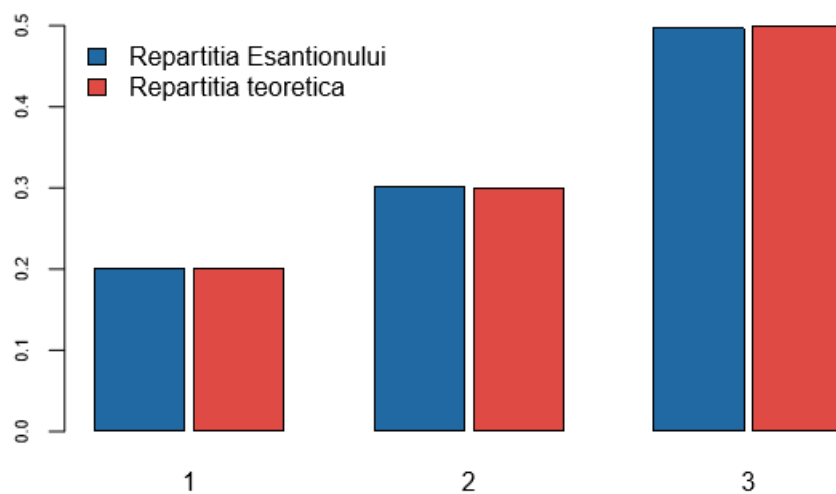
names.arg = indX,

cex.axis = 0.7,

legend.text = c("Repartitia Esantionului", "Repartitia teoretica"),

args.legend = list(x = "topleft", bty = "n"))

```



2. În acest caz considerăm variabila aleatoare $X \sim \begin{pmatrix} a & b & c & d \\ 0.15 & 0.25 & 0.15 & 0.45 \end{pmatrix}$, deci

$x = [a, b, c, d]$ și $p = [0.15, 0.25, 0.15, 0.45]$. Mai jos generăm $n = 15$ observatii din repartitia variabilei aleatoare X:

```
GenerateDiscrete(15, c('a','b','c','d'), c(0.15,0.25,0.15,0.45))
```

```
[1] "d" "c" "c" "d" "d" "b" "c" "c" "b" "b" "d" "a" "d" "d" "d"
```

Ca și în cazul primului exemplu, vom compara repartitia teoretică cu cea a unui esantion de

$n = 10000$ de observatii:

```
n = 10000
```

```
x = GenerateDiscrete(n, c('a','b','c','d'), c(0.15,0.25,0.15,0.45))
```

```
# cate observatii din fiecare valoare unica a lui x
```

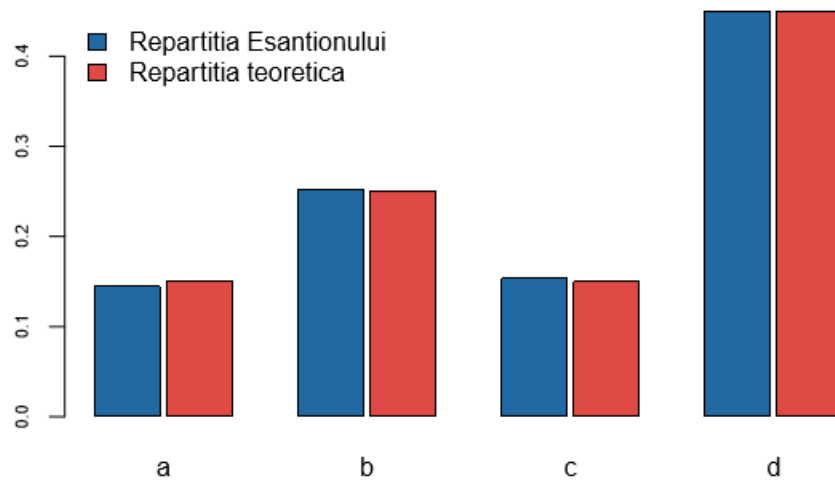
```
pX = table(x)/n
```

```
pT = c(0.15,0.25,0.15,0.45)
```

```

indX = c('a','b','c','d')
barplot(rbind(pX, pT),
        beside = T,
        space = c(0.1, 1),
        col = c(myblue, myred),
        names.arg = indX,
        cex.axis = 0.7,
        legend.text = c("Repartitia Esantionului", "Repartitia teoretica"),
        args.legend = list(x = "topleft", bty = "n"))

```



APLICATIE:

Pentru variabila aleatoare X cu distributia urmatoare $X \sim \begin{pmatrix} \frac{1}{2} & 1 & 2 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$

a) Generati $n = 12$ observatii din repartitia variabilei aleatoare X .

b) Plecând de la un esantion de $n = 10000$ de observatii comparati repartitia esantionului cu cea teoretică.

FUNCTIA DE REPARTITIE PENTRU VARIABILE ALEATOARE

Scrieti o functie în R care să traseze graficul functiei de repartitie a unei distributii date. Verificati si documentatia functiei ecdf.

Definim următoarea functie:

```
cdfPlot = function(dist, title, err = 1e-10){  
  # dist - repartitia discreta (sau discretizata)  
  
  lp = length(dist)  
  if (abs(sum(dist)-1)>err | sum(dist>=0)!=lp){  
    stop("Eroare: vectorul de probabilitati nu formeaza o repartitie")  
  }else{  
    x = 0:(lp-1)  
    # ia valori in 1:lp  
    cp = cumsum(dist)  
    plot(x, cp, type = "s", lty = 3,  
         xlab = "x",  
         ylab = "F",  
         main = paste("Functia de repartitie:", title),  
         ylim = c(0,1), col = "grey", bty = "n")  
    abline(h = 0, lty = 2, col = "grey")  
    abline(h = 1, lty = 2, col = "grey")  
    for(i in 1:(lp-1)){  
      lines(c(x[i], x[i+1]), c(cp[i], cp[i]),
```

```
col = myblue,
lwd = 2)
} points(x,cp, col = myred, pch = 20, cex = 0.85)}}}
```

APLICATIE:

Fie variabila aleatoare discreta simpla X cu distributia urmatoare $X \sim$

1	2	3	4	5	6
0.1	0.2	0.3	0.1	0.2	0.1

Determinați funcția de repartiție a variabilei aleatoare X și reprezentați grafic variabila și funcția ei de repartiție.