PROIECT LA STATISTICĂ

Membri:
Cernătescu Tiberiu-Florian 312
Florescu Iuliana 312
Ghergu Cristian 312
Madescu Sandra 312

Lider echipă:
Florescu Iuliana

I. **Inegalitatea Berry-Essen** este un rezultat celebru cu ajutorul căruia putem determina acuratețea aproximării pe care o realizează TLC. Cerând în plus față de condițiile din TLC ca $E|X_i|^3 < \infty$, avem că:

$$\sup_x |P(Z_n \le x) - \Phi(x)| \le \frac{33}{4} \frac{E|X_1 - \mu|^3}{\sqrt{n}\sigma^3} \quad ,$$

unde $X_1, X_2 ... X_n$ *i.i.d.*, $\Phi(x)$ este funcția de repartiție a normalei standard, $\mu = E(X_1)$ $\sigma = \sqrt{Var(X_1)}$ și $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$.

1) Calculați $P(Z_n \le x)$ pentru repartițiile:

Binomială, Geometrică, Poisson, Uniformă pe caz discret și respectiv continuu, Exponențială, Gamma și respectiv Beta.

- În cazul repartiției binomială:

```
n <- 10^3
m <- 5
p <- 0.5
X <- rbinom(n, m, p)

#calculam media si varianta
miu <- m*p
sigma <- sqrt(m*p*(1-p))

#calculam v.a. standardizata Z_n
Z_n <- (sqrt(n)*(mean(X)-miu))/sigma

#calculam probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.09661614
```

- În cazul repartiției geometrică:

```
#Generăm eșantioane dintr-o repartiție geometrică
n <- 10^3
p <- 0.5
X <- rgeom(n, p)

#Calculăm media și varianta
miu <- 1 / p
sigma <- sqrt((1 - p) / p^2)

#Calculăm v.a standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  2.74769e-121
```

- În cazul repartiției Poisson:

```r
#Generăm eșantioane dintr-o repartiție Poisson
n <- 10^3
lambda <- 5
X <- rpois(n, lambda)

#Calculăm media și varianta
miu <- lambda
sigma <- sqrt(lambda)

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.9538041
```

- În cazul repartiției uniforme pe caz discret:

```r
#Generăm eșantioane dintr-o repartiție uniformă pe o variabilă discretă
n <- 10^3
a <- 1
b <- 10
X <- sample(a:b, n, replace = TRUE)

#Calculăm media și varianta
miu <- (a + b) / 2
sigma <- sqrt(((b - a + 1)^2 - 1) / 12)

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

# Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.3459245
```

- În cazul repartiției uniforme pe caz continuu:

```r
#Generăm eșantioane dintr-o repartiție uniformă pe un interval continuu
n <- 10^3
a <- 1
b <- 10
X <- runif(n, min = a, max = b)

#Calculăm media și varianta
miu <- (a + b) / 2
sigma <- sqrt((b - a)^2 / 12)

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.07195765
```

- În cazul repartiției exponențiale:

```r
#Generăm eșantioane dintr-o repartiție exponentială
n <- 10^3
lambda <- 0.5
X <- rexp(n, rate = lambda)

#Calculăm media și varianta
miu <- 1 / lambda
sigma <- 1 / lambda

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.06262768
```

- În cazul repartiției Gamma:

```r
#Generăm eșantioane dintr-o repartiție gamma
n <- 10^3
shape <- 2
rate <- 1
X <- rgamma(n, shape = shape, rate = rate)

#Calculăm media și varianta
miu <- shape / rate
sigma <- sqrt(shape) / rate

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

```
P(Z_n <= n) =  0.2884012
```

- În cazul repartiției Beta:

```r
#Generăm eșantioane dintr-o repartiție beta
n <- 10^3
alpha <- 2
beta <- 2
X <- rbeta(n, shape1 = alpha, shape2 = beta)

#Calculăm media și varianta
miu <- alpha / (alpha + beta)
sigma <- sqrt(alpha * beta / ((alpha + beta)^2 * (alpha + beta + 1)))

#Calculăm v.a. standardizată Z_n
Z_n <- (sqrt(n) * (mean(X) - miu)) / sigma

#Calculăm probabilitatea P(Z_n <= x)
P_Z_n <- pnorm(Z_n)

cat("P(Z_n <= n) = ", P_Z_n, "\n")
```

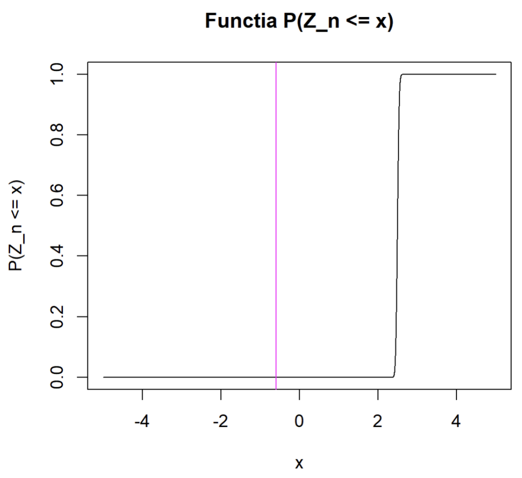
```
P(Z_n <= n) =  0.9007166
```

2) Reprezentați grafic funcțiile obținute la 1).

Pentru a genera grafic funcțiile folosim codul R pentru a genera funcțiile, la fel ca în cazurile de la 1), la care adăugăm reprezentarea grafică.

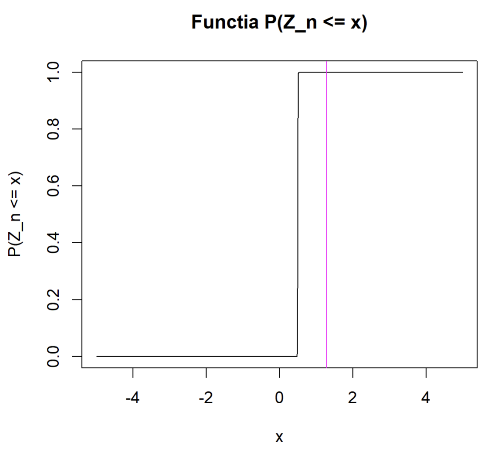
- În cazul repartiției binomială:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n)*(grafic_x-miu))/sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
      main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



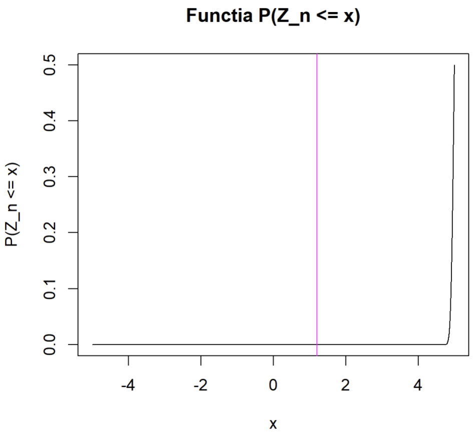
- În cazul repartiției geometrică:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



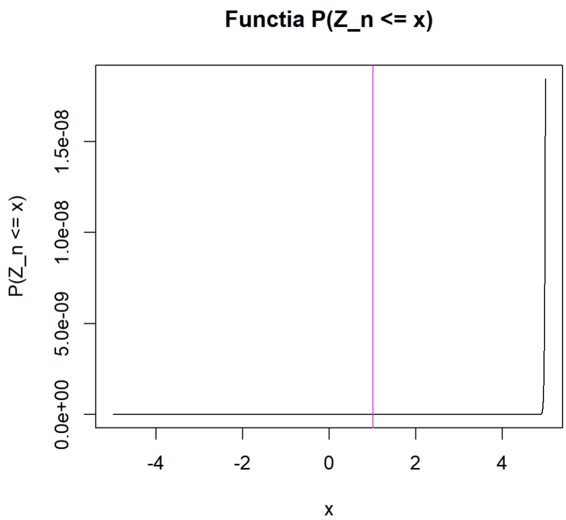
- În cazul repartiției Poisson:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
      main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



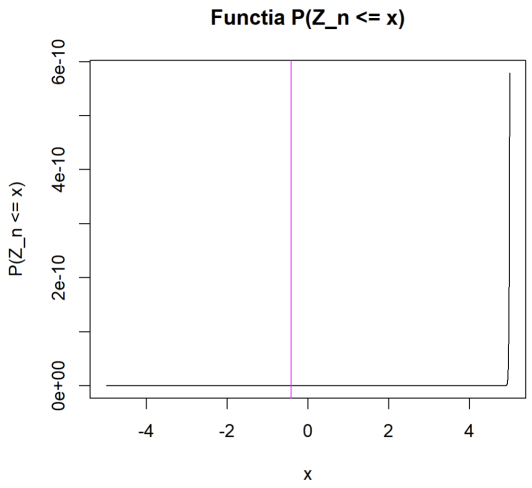
- În cazul repartiției uniforme pe caz discret:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



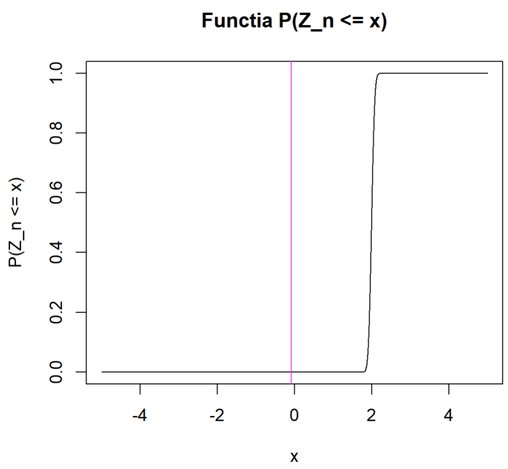
- În cazul repartiției uniforme pe caz continuu:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



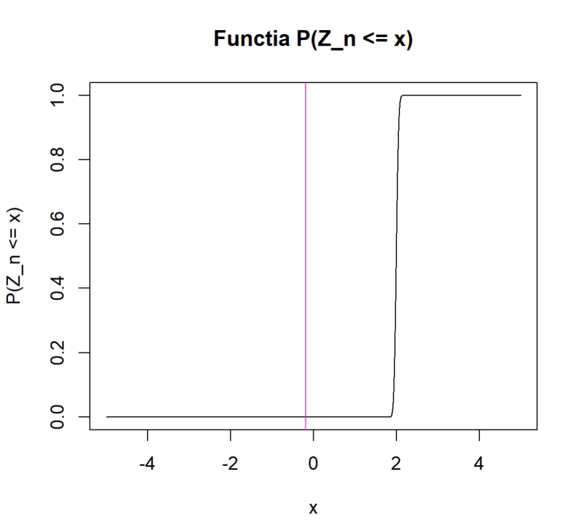
- În cazul repartiției exponențiale:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



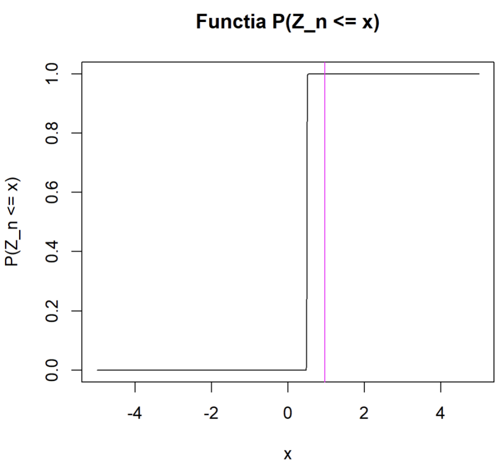
- În cazul repartiției Gamma:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



- În cazul repartiției Beta:

```
grafic_x <- seq(-5, 5, 0.001)
z_n <- (sqrt(n) * (grafic_x - miu)) / sigma
y <- pnorm(z_n)

plot(grafic_x, y, type = 'l', xlab = 'x', ylab = 'P(Z_n <= x)',
     main = 'Functia P(Z_n <= x)')
abline(v = z_n, col = 'magenta')
```



3) Folosind funcția *optimize* aproximați pentru repartițiile de mai sus:

$$\sup_x |P(Z_n \le x) - \Phi(x)|$$

O să folosim codul R scris pentru 1) și 2).

- În cazul repartiției binomială:

```
sup_binomiala <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return (sup)
}

optimizare_binomiala <- optimize(sup_binomiala, c(-n,n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_binomiala$maximum), "\n")
```

- În cazul repartiției geometrică:

```r
sup_geometrica <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_geometrica <- optimize(sup_geometrica, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_geometrica$maximum), "\n")
```

- În cazul repartiției Poisson:

```r
sup_poisson <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_poisson <- optimize(sup_poisson, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_poisson$maximum), "\n")
```

- În cazul repartiției uniforme pe caz discret:

```r
sup_uniforma <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_uniforma <- optimize(sup_uniforma, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_uniforma$maximum), "\n")
```

- În cazul repartiției uniforme pe caz continuu:

```r
sup_uniforma_continua <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_uniforma_continua <- optimize(sup_uniforma_continua, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_uniforma_continua$maximum), "\n")
```

- În cazul repartiției exponențiale:

```r
#Aproximarea repartiției exponențială
sup_exponentiala <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_exponentiala <- optimize(sup_exponentiala, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_exponentiala$maximum), "\n")
```

- În cazul repartiției Gamma:

```r
sup_gamma <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_gamma <- optimize(sup_gamma, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - phi(x)| aproximat cu optimize = ", abs(optimizare_gamma$maximum), "\n")
```

- În cazul repartiției Beta:

```r
sup_beta <- function(x){
  sup <- abs(P_Z_n - pnorm(x))
  return(sup)
}

optimizare_beta <- optimize(sup_beta, c(-n, n), maximum = TRUE)
cat("|P(Z_n <= X) - theta(x)| aproximat cu optimize = ", abs(optimizare_beta$maximum), "\n")
```

4) Construiți câte o funcție în R care să calculeze $E(X)$ și respectiv $Var(X)$, unde tipul repartiției v.a. X este transmis fie printr-o denumire, fie prin funcția de masă/funcția densitate de probabilitate, în cazul discret și respectiv continuu.

```
fct_1_4 <- function(distributie, mass_function = NULL, density_function = NULL){
  if (is.null(mass_function) && is.null(density_function)) {
    if (distributie == "binomiala") {
      n <- 5
      p <- 0.3
      mass_function <- function(k) dbinom(k, size = n, prob = p)
    } else if (distributie == "poisson") {
      lambda <- 2
      mass_function <- function(k) dpois(k, lambda)
    } else if (distributie == "normala") {
      mu <- 0
      sigma <- 1
      density_function <- function(x) dnorm(x, mean = mu, sd = sigma)
    } else if (distributie == "gamma") {
      shape <- 2
      rate <- 1
      density_function <- function(x) dgamma(x, shape = shape, rate = rate)
    } else if (distributie == "beta") {
      alpha <- 2
      beta <- 2
      density_function <- function(x) dbeta(x, shape1 = alpha, shape2 = beta)
    } else if (distributie == "exponentiala") {
      lambda <- 0.5
      density_function <- function(x) dexp(x, rate = lambda)
    } else {
      stop("Nume distributie invalid")
    }
  }


  if(!is.null(mass_function)){
    #Cazul discret pentru o v.a.

    x <- 0:10
    px <- mass_function(x) #functia de masa
    E_X <- sum(x * px) #media lui X
    Var_X <- sum((x-E_X)^2 * px ) #varianta lui X
  }
  else{
    #cazul continua pentru o v.a.

    x <- seq(-3, 3, by = 0.01)
    px <- density_function(x) #densitatea de probabilitate
    E_X <- integrate(density_function, -Inf, Inf)$value #media lui X
    Var_X <- integrate(function(x) (x - E_X)^2 * density_function(x), -Inf, Inf)$value #varianta lui X
  }

  return(list(E_X = E_X, Var_X = Var_X))

}
```

```
print(fct_1_4("binomiala"))
print(fct_1_4("poisson"))
print(fct_1_4("normala"))
print(fct_1_4("gamma"))
print(fct_1_4("beta"))
print(fct_1_4("exponentiala"))

print(fct_1_4(mass_function = function(k) dbinom(k, size = 5, prob = 0.5)))
print(fct_1_4(mass_function = function(k) dpois(k, lambda = 2)))
print(fct_1_4(density_function = function(x) dnorm(x, mean = 0, sd = 1)))
print(fct_1_4(density_function = function(x) dbeta(x, shape1 = 2, shape2 = 2)))
print(fct_1_4(density_function = function(x) dgamma(x, shape = 2, rate = 1)))
print(fct_1_4(density_function = function(x) dexp(x, rate = 0.5)))
```

5) Construiți o funcție în R care să calculeze $E|X_1 - \mu|^3$

```
fct_1_5 <- function(distributie, mass_function = NULL, density_function = NULL){
  if (is.null(mass_function) && is.null(density_function)) {
    if (distributie == "binomiala") {
      n <- 5
      p <- 0.3
      mass_function <- function(k) dbinom(k, size = n, prob = p)
    } else if (distributie == "poisson") {
      lambda <- 2
      mass_function <- function(k) dpois(k, lambda)
    } else if (distributie == "normala") {
      mu <- 0
      sigma <- 1
      density_function <- function(x) dnorm(x, mean = mu, sd = sigma)
    } else if (distributie == "gamma") {
      shape <- 2
      rate <- 1
      density_function <- function(x) dgamma(x, shape = shape, rate = rate)
    } else if (distributie == "beta") {
      alpha <- 2
      beta <- 2
      density_function <- function(x) dbeta(x, shape1 = alpha, shape2 = beta)
    } else if (distributie == "exponentiala") {
      lambda <- 0.5
      density_function <- function(x) dexp(x, rate = lambda)
    } else {
      stop("Nume distributie invalid")
    }
  }


  if(!is.null(mass_function)){
    #Cazul discret pentru o v.a.

    x <- 0:10
    px <- mass_function(x) #functia de masa
    E_X <- sum(x * px) #media lui X
    Var_X <- sum(abs((x-E_X)^3) * px ) #varianta lui X
  }
  else{
    #cazul continua pentru o v.a.

    x <- seq(-3, 3, by = 0.01)
    px <- density_function(x) #densitatea de probabilitate
    E_X <- integrate(density_function, -Inf, Inf)$value #media lui X
    Var_X <- integrate(function(x) abs((x - E_X)^3) * density_function(x), -Inf, Inf)$value #varianta lui X
  }

  return(list(E_X = E_X, Var_X = Var_X))

}
```

```
print(fct_1_5("binomiala"))
print(fct_1_5("poisson"))
print(fct_1_5("normala"))
print(fct_1_5("gamma"))
print(fct_1_5("beta"))
print(fct_1_5("exponentiala"))

print(fct_1_5(mass_function = function(k) dbinom(k, size = 5, prob = 0.5)))
print(fct_1_5(mass_function = function(k) dpois(k, lambda = 2)))
print(fct_1_5(density_function = function(x) dnorm(x, mean = 0, sd = 1)))
print(fct_1_5(density_function = function(x) dbeta(x, shape1 = 2, shape2 = 2)))
print(fct_1_5(density_function = function(x) dgamma(x, shape = 2, rate = 1)))
print(fct_1_5(density_function = function(x) dexp(x, rate = 0.5)))
```

6) Construiți un dataframe în R care să conțină marginile date de inegalitatea Berry-Essen pentru repartițiile: Binomială, Geometrică, Poisson, Uniformă pe caz discret și respectiv continuu, Exponențială, Gamma și respectiv Beta, pentru un volum al eșantionului $n \in \{30,100,1000\}$. Alegerea valorilor parametrilor repartițiilor ilustrate vă revine vouă.

```r
# Definirea funcției pentru distribuția geometrică
dgeom_custom <- function(x, prob) {
  return((1 - prob)^(x - 1) * prob)
}

fct_1_6 <- function(distribution, mass_function = NULL, density_function = NULL){
  if (is.null(mass_function) && is.null(density_function)) {
    if (distribution == "binomiala") {
      n <- 5
      p <- 0.3
      mass_function <- function(k) dbinom(k, size = n, prob = p)
    } else if (distribution == "poisson") {
      lambda <- 2
      mass_function <- function(k) dpois(k, lambda)
    } else if (distribution == "geometrica"){
      prob <- 0.3
      mass_function <- function(k) dgeom_custom(k, prob)
    } else if (distribution == "normala") {
      mu <- 0
      sigma <- 1
      density_function <- function(x) dnorm(x, mean = mu, sd = sigma)
    } else if (distribution == "gamma") {
      shape <- 2
      rate <- 1
      density_function <- function(x) dgamma(x, shape = shape, rate = rate)
    } else if (distribution == "beta") {
      alpha <- 2
      beta <- 2
      density_function <- function(x) dbeta(x, shape1 = alpha, shape2 = beta)
    } else if (distribution == "exponentiala") {
      lambda <- 0.5
      density_function <- function(x) dexp(x, rate = lambda)
    } else {
      stop("Nume distributie invalid")
    }
  }

  if(!is.null(mass_function)){
    #Cazul discret pentru o v.a.

    x <- 0:10
    px <- mass_function(x) #functia de masa
    E_X <- sum(x * px) #media lui X
    Var_X <- sum((x-E_X)^2 * px ) #varianta lui X
    E_abs_X3 <- sum(abs(x-E_X)^3 * px)
  } else{
    #cazul continua pentru o v.a.

    x <- seq(-3, 3, by = 0.01)
    px <- density_function(x) #densitatea de probabilitate
    E_X <- integrate(density_function, -Inf, Inf)$value #media lui X
    Var_X <- integrate(function(x) (x - E_X)^2 * density_function(x), -Inf, Inf)$value #varianta lui X
    E_abs_X3 <- integrate(function(x) (x - E_X)^3 * density_function(x), -Inf, Inf)$value
  }

  return(list(E_X = E_X, Var_X = Var_X, E_abs_X3 = E_abs_X3))

}
```

```r
#Functia Berry-Essen

berry_essen <- function(n, E_X, Var_X, E_abs_X3){
  return(sqrt(Var_X/n) * (1 + (3/8) * (E_abs_X3 /(Var_X^(3/2))) - (1/4) * (E_X^2/(Var_X^2))))
}

#Dataframe

df <- data.frame(distribution = character(),
                 n = numeric(),
                 margine = numeric(),
                 stringsAsFactors = FALSE)
```

```r
#BINOMIALA

params <- fct_1_6("binomiala")
n <- c(30, 100, 1000)
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "binomiala", n = i, margine = margine))
}
print(df)
```

```
  distribution    n    margine
1    binomiala   30 0.19886281
2    binomiala  100 0.10892165
3    binomiala 1000 0.03444405
```

```r
#GEOMETRICA

params <- fct_1_6("geometrica")
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "geometrica", n = i, margine = margine))
}
print(df)
```

```
  distribution    n   margine
1   geometrica   30 0.7508762
2   geometrica  100 0.4112718
3   geometrica 1000 0.1300556
```

```r
#POISSON

params <- fct_1_6("poisson")
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "Poisson", n = i, margine = margine))
}
print(df)
```

```
  distribution    n    margine
1      Poisson   30 0.35453611
2      Poisson  100 0.19418742
3      Poisson 1000 0.06140746
```

```r
#UNIFORMA PE CAZ DISCRET

params <- fct_1_6("uniforma pe caz discret", mass_function = function(k) dunif(k, min = 1, max = 6))
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "uniforma pe caz discret", n = i, margine = margine))
}
print(df)
```

```
             distribution    n    margine
1 uniforma pe caz discret   30 0.44422107
2 uniforma pe caz discret  100 0.24330990
3 uniforma pe caz discret 1000 0.07694135
```

```r
#UNIFORMA PE CAZ CONTINUU

params <- fct_1_6("uniforma pe caz continuu", density_function = function(k) dunif(k, min = 1, max = 6))
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "uniforma pe caz continuu", n = i, margine = margine))
}
print(df)
```

```
              distribution    n   margine
1 uniforma pe caz continuu   30 0.7818936
2 uniforma pe caz continuu  100 0.4282608
3 uniforma pe caz continuu 1000 0.1354279
```

```r
#EXPONENTIALA

params <- fct_1_6("exponentiala", density_function = function(x) dexp(x, rate = 2))
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "exponentiala", n = i, margine = margine))
}
print(df)
```

```
  distribution    n      margine
1 exponentiala   30 -0.034232660
2 exponentiala  100 -0.018750000
3 exponentiala 1000 -0.005929271
```

```r
#GAMMA

params <- fct_1_6("gamma", density_function = function(x) dgamma(x, shape = 2, scale = 2))
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "gamma", n = i, margine = margine))
}
print(df)
```

```
  distribution    n   margine
1        gamma   30 1.2797072
2        gamma  100 0.7009245
3        gamma 1000 0.2216518
```

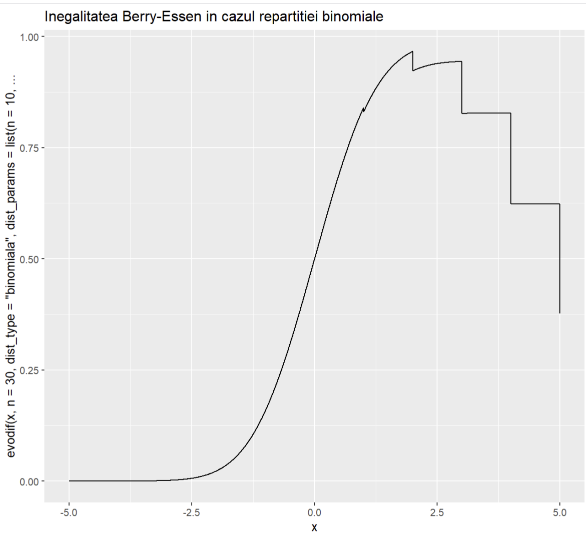
```r
#BETA

params <- fct_1_6("beta", density_function = function(x) dbeta(x, shape1 = 2, shape2 = 2))
for (i in n){
  margine <- berry_essen(i, params$E_X, params$Var_X, params$E_abs_X3)
  df <-  rbind(df, data.frame(distribution = "beta", n = i, margine = margine))
}

print(df)
```

```
  distribution    n     margine
1         beta   30 -0.22342132
2         beta  100 -0.12237290
3         beta 1000 -0.03869771
```

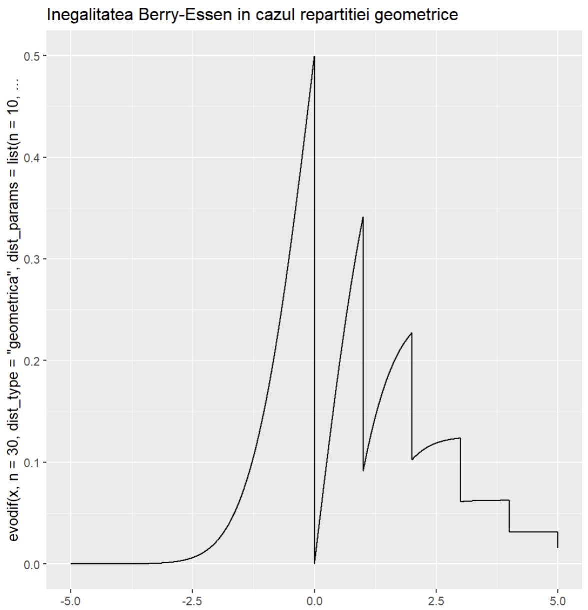
7) Ilustrați grafic pentru $n \in \{30, 100, 1000\}$ pentru un interval pe care îl considerați relevant evoluția diferenței $P(Z_n \leq x) - \Phi(x)$ pentru fiecare din repartițiile de mai sus.

```r
#functie care calculeaza marginea data de inegalitatea Berry-Essen pentru o v.a.
#parametrii functiei sunt X v.a., esantionul n, tipul de distributie si lista cu parametrii distributiilor
functie <- function(X, n, dist_type, dist_params) {
  medie <- E(X)
  var <- Var(X)

  asim <- E((X - medie)^3)/(var^(3/2)) #skewness - masoara asimetria distributiei
  kurt <- E((X - medie)^4)/(var^2) - 3 #kurtosis - masoara sansa ca intr-o distributie sa existe valori extreme

  margine <- (1/(12*n)) * (asim^2 + (kurt - 6)^2/4)

  return(margine)
}

#install.packages("ggplot2")
library(ggplot2)

n <- c(30, 100, 1000)

interval_x <- c(-10, 10)

#Calculul diferentei pentru toate cazurile posibile

evodif <- function(x, n, dist_type, dist_params) {
  phi_x <- pnorm(x, mean = 0, sd = 1) #functia de repartitie standard
  P_zn <-
    if(dist_type == "binomiala") {
      pbinom(x, size = dist_params$n, prob = dist_params$p)
    } else if(dist_type == "geometrica") {
      pgeom(x, prob = dist_params$p)
    } else if(dist_type == "poisson") {
      ppois(x, lambda = dist_params$lambda)
    } else if(dist_type == "uniforma_discreta") {
      punif(x, min = dist_params$a, max = dist_params$b)
    } else if(dist_type == "exponentiala") {
      pexp(x, rate = dist_params$lambda)
    } else if(dist_type == "gamma") {
      pgamma(x, shape = dist_params$shape, rate = dist_params$rate)
    } else if(dist_type == "beta") {
      pbeta(x, shape1 = dist_params$shape1, shape2 = dist_params$shape2)
    }
  return(abs(P_zn - phi_x))
}
```

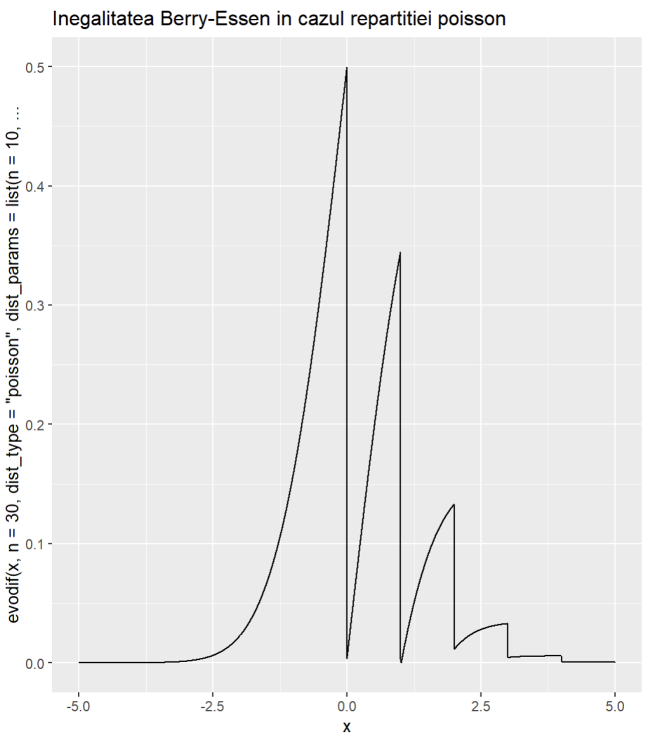
```r
#Repartitie Binomiala

evodif(x = 1, n = 30, dist_type = "binomial", dist_params = list(n = 5, p = 0.5))
library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "binomiala", dist_params = list(n = 10, p = 0.5)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei binomiale")
```



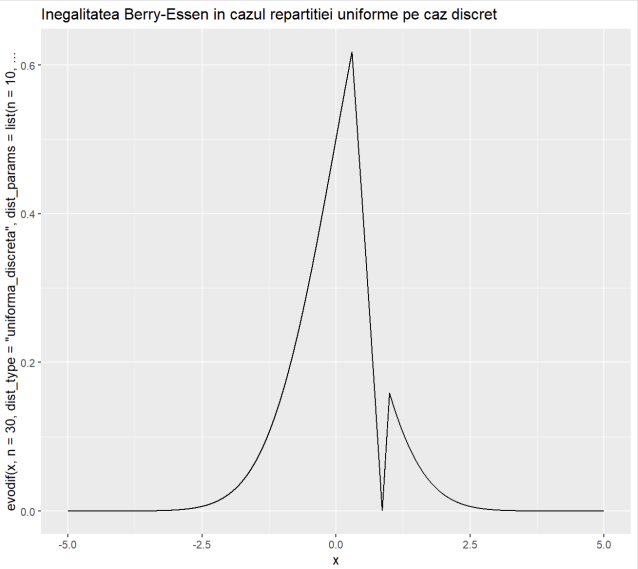
```
#Repartitia Geometrica

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "geometrica", dist_params = list(n = 10, p = 0.5)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei geometrice")

evodif(x = 1, n = 30, dist_type = "geometric", dist_params = list(n = 10, p = 0.5))
```



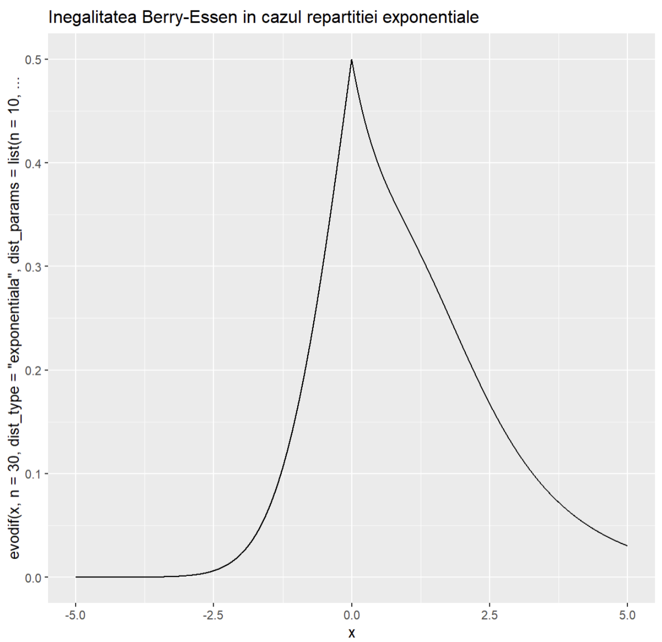

```
#Repartitia Poisson

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "poisson", dist_params = list(n = 10, lambda = 0.7)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei poisson")

evodif(x = 1, n = 30, dist_type = "poisson", dist_params = list(n = 10, lambda = 0.5))
```



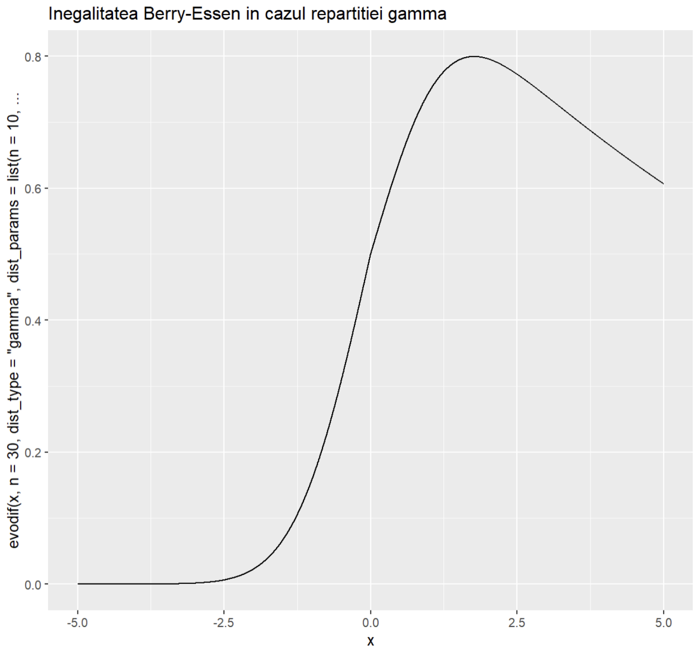
```r
#Repartitia Uniforma - cazul discret

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "uniforma_discreta", dist_params = list(n = 10, a = 0.3, b = 1)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei uniforme pe caz discret")

evodif(x = 1, n = 30, dist_type = "unif_discret", dist_params = list(n = 10, a = 0.5, b = 1))
```



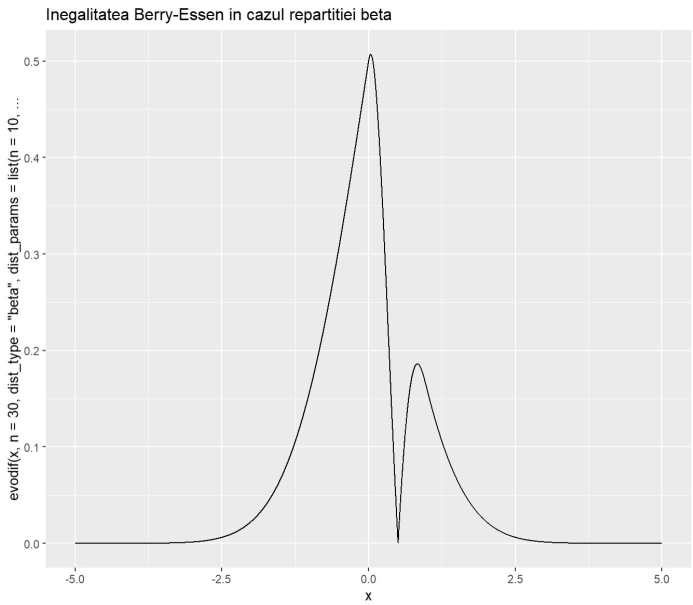

```r
#Repartitia Exponentiala

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "exponentiala", dist_params = list(n = 10, lambda = 0.7)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei exponentiale")

evodif(x = 1, n = 30, dist_type = "exponential", dist_params = list(n = 10, lambda = 0.5))
```



```r
#Repartitia Gamma

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "gamma", dist_params = list(n = 10, shape = 1, rate = 0.1)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei gamma")

evodif(x = 1, n = 30, dist_type = "gamma", dist_params = list(n = 10, shape = 1, rate = 0.1))
```




```r
#Repartitia Beta

library(ggplot2)
val_x <- seq(-5, 5, 0.001)
ggplot(data.frame(x = val_x), aes(x)) +
  geom_line(aes(y = evodif(x, n = 30, dist_type = "beta", dist_params = list(n = 10, shape1 = 2, shape2 = 3)))) +
  ggtitle("Inegalitatea Berry-Essen in cazul repartitiei beta")

evodif(x = 1, n = 30, dist_type = "beta", dist_params = list(n = 10, shape1 = 2, shape2 = 3))
```



8) Construiți o funcție în R care să calculeze marginea dată de inegalitatea Berry-Essen pentru o v.a. pentru care se cunoaște funcția de masă/funcția densitate de probabilitate.

```
#parametrii functiei sunt X v.a., esantionul n, tipul de distributie si lista cu parametrii distributiilor
functie <- function(X, n, dist_type, dist_params) {

  medie <- E(X)
  var <- Var(X)

  asim <- E((X - medie)^3)/(var^(3/2)) #skewness - masoara asimetria distributiei
  kurt <- E((X - medie)^4)/(var^2) - 3 #kurtosis - masoara sansa ca intr-o distributie sa existe valori extreme

  margine <- (1/(12*n)) * (asim^2 + (kurt - 6)^2/4) #calculul marginii folosind inegalitatea Berry-Essen

  return(margine)
}
```

II. Folosiți metoda respingerii pentru a genera observații dintr-o densitate de probabilitate definită prin $f(x) \propto \exp(-x^2/2) \cdot \{\sin(6x)^2 + 3\cos(x)^2 \sin(4x)^2 + 1\}$ parcurgând pașii următori:

(OBS: Notația $\propto$ înseamnă că $f(x)$ este proporțional cu expresia din dreapta)

a) Reprezentați grafic $f(x)$ și arătați că aceasta este mărginită de $Mg(x)$ unde $g(x)$ este densitatea de probabilitate a repartiției normale standard. Determinați o valoare potrivită pentru constanta M, chiar dacă nu este optimă.
(Indiciu: Folosiți funcția optimise din R).

$$f(x) \propto \exp(-x^2/2) \cdot (\sin(6x)^2 - 3\cos(x)^2 \sin(4x)^2 + 1)$$

Fie $h(x) = \exp(-x^2/2) \cdot (\sin(6x)^2 - 3\cos(x)^2 \sin(4x)^2 + 1)$

Dacă $f(x)$ e proporționată cu $h(x)$ atunci: $f(x) = c \cdot h(x)$

$$\Rightarrow h(x) = \frac{f(x)}{c}$$

Căutăm M a.î. $M \geq \frac{f(x)}{g(x)}$, $g(x)$ densitatea repartiției normale de medie 0 și dispersie 1.

$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ dar $\mu = 0$ și $\sigma^2 = 1 \Rightarrow \sigma = \pm 1$.

Adică $g(x) = \frac{1}{\pm\sqrt{2\pi}} \cdot e^{\frac{-(x)^2}{2}}$. Cum densitatea este o funcție pozitivă alegem $\sigma = 1$. Deci $\boxed{g(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x)^2}{2}}}$

Căutăm maximul în raportul $\frac{h(x)}{g(x)}$.

$$\frac{h(x)}{g(x)} = \frac{\cancel{\exp(-x^2/2)} \cdot (\sin(6x)^2 - 3\cos(x)^2 \cdot \sin(4x)^2 + 1)}{\frac{1}{\sqrt{2\pi}} \cdot \cancel{e^{-\frac{(x)^2}{2}}}} =$$

$$= \sqrt{2\pi}(\sin(6x)^2 - 3\cos(x)^2 - \sin(4x)^2 + 1)$$

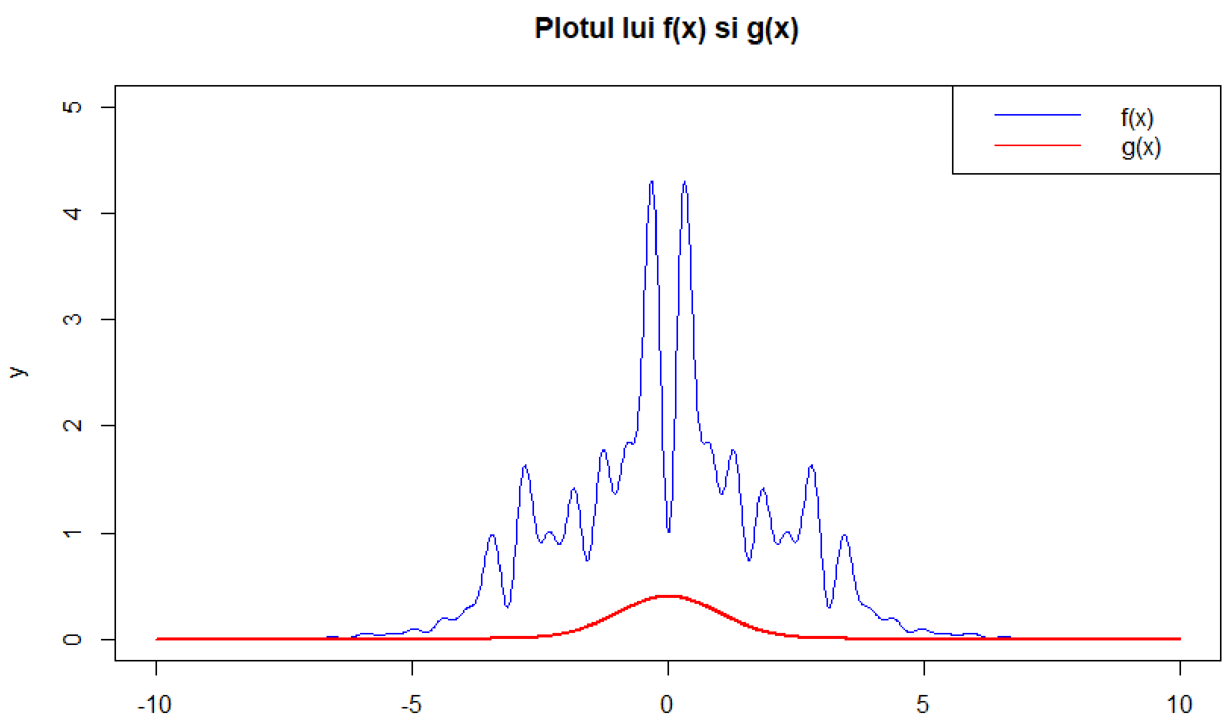
Folosim funcția „optimise" și mărginim graficul lui $h(x)$ cu $g(x) \cdot M$. (M găsit cu „optimise").

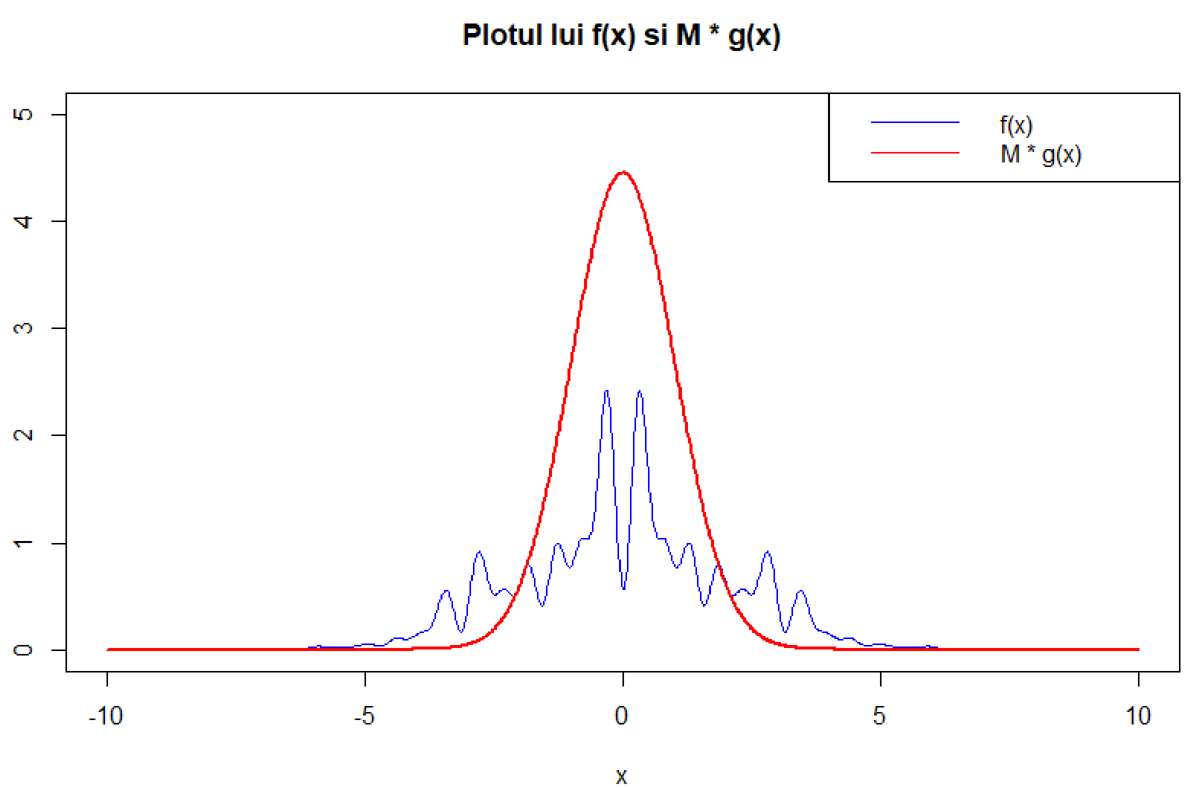
II. Folosiţi **metoda respingerii** pentru a genera observaţii din densitatea de probabilitate definită prin $f(x) \propto \exp(-x^2/2)\{\sin(6x)^2 + 3\cos(x)^2\sin(4x)^2 + 1\}$ parcurgând paşii următori:

(**OBS**: Notaţia "$\propto$" înseamnă că $f(x)$ *este proporţional cu* expresia din dreapta)

a) Reprezentaţi grafic $f(x)$ şi arătaţi că aceasta este mărginită de $Mg(x)$ unde $g(x)$ este densitatea de probabilitate a repartiţiei normale standard. Determinaţi o valoare potrivită pentru constanta M, chiar dacă nu este optimă.
(**Indiciu**: Folosiţi funcţia *optimise* din **R**)

```
#f(x) si g(x)
f <- function(x) {
  exp(-x^2 / 8) * (sin(6 * x)^2 + 3 * cos(x)^2 * sin(4 * x)^2 + 1)
}

g <- function(x) {
  (1 / sqrt(2 * pi)) * exp(-x^2 / 2)
}

# Aflam maximul lui f(x) pe un interval mare
opt_f <- optimise(f, interval = c(-100, 100), maximum = TRUE)

# Aflam maximul lui g(x) pe un interval mare
opt_g <- optimise(g, interval = c(-100, 100), maximum = TRUE)

# Valoarea optimă a lui M este raportul dintre maximul lui f(x) și maximul lui g(x)
M <- opt_f$objective / opt_g$objective

# Generam câteva date din f(x)
x <- seq(-10, 10, length.out = 1000)
fx <- f(x) / opt_f$objective  # Normalize f(x)

# Generam valorile corespunzătoare pentru M * g(x)
Mgx <- M * g(x) / opt_g$objective  # Normalizam M * g(x)

plot(x,f(x),type='l',col = "blue", ylim=c(0,5),main = "Plotul lui f(x) si g(x)", xlab = "x", ylab = "y")

lines(x,g(x), col = "red", lwd = 2)

legend("topright", legend = c("f(x)", "g(x)"), col = c("blue", "red"), lty = 1)


plot(x, fx, type = "l", col = "blue", ylim = c(0, 5), main = "Plotul lui f(x) si M * g(x)", xlab = "x", ylab = "y")

lines(x, Mgx, col = "red", lwd = 2)

legend("topright", legend = c("f(x)", "M * g(x)"), col = c("blue", "red"), lty = 1)
```


Plotul lui f(x) si g(x)
f(x)
g(x)
y
0
1
2
3
4
5
-10
-5
0
5
10


Plotul lui f(x) si M * g(x)
f(x)
M * g(x)
0
1
2
3
4
5
-10
-5
0
5
10
x

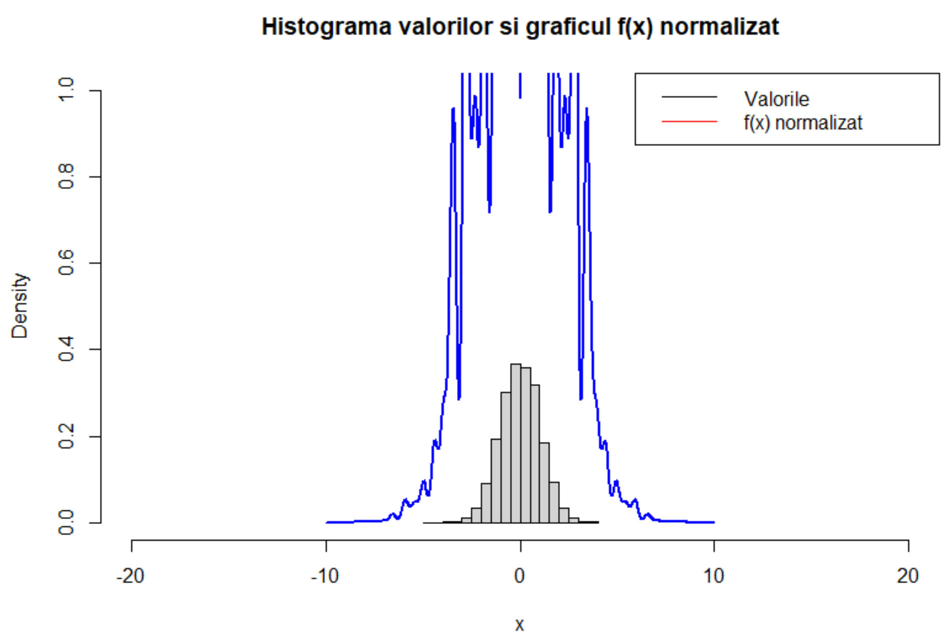
b) Generaţi 25000 de observaţii din densitatea de mai sus folosind metoda respingerii.

```r
# Initializam un vector pentru a stoca observatiile
samples <- numeric()

# Inițializam un contor pentru numărul total
total_samples <- 0

# Generam 25000 de observatii
while (length(samples) < 25000) {
  # Generam x o variabila aleatoare
  x <- rnorm(1)

  total_samples <- total_samples + 1

  # Generam un număr aleator uniform
  u <- runif(1)

  # Acceptam o variabila cu probabilitatea f(x) / (M * g(x))
  if (u <= f(x) / (M * g(x))) {
    samples <- c(samples, x)
  }
}

print(samples)
```

```
> print(samples)
   [1] -2.342544565  0.747514614 -2.241152837  2.399175577  1.332692349  1.097695601  0.601939685  0.522015443 -0.225119288
  [10]  0.022119461 -0.976938017  2.205472823 -0.786322097  1.304180756 -0.676960298  1.246534958 -0.045195358  1.344243283
  [19] -0.356449964 -2.199824599  1.007849503  1.284572656 -1.709147349  0.077441723 -1.370752690  1.374057411  0.822801067
  [28] -0.953101647 -0.429914841 -1.416659755 -1.324135004 -0.236663653  0.829435087 -1.914661275  0.974835152 -0.490787369
  [37] -0.480531022 -1.020129495  1.353413377  0.957424630  1.652510648 -1.339871089  0.526521536  0.398801615  2.377636732
  [46] -0.504902895 -0.774331421 -1.366878865 -0.134480941 -0.825982814  0.611484458  0.440965250  0.685398816 -0.613557359
  [55] -1.139964389 -0.349842278 -1.166969757  0.036435287  0.489432538 -0.807823949 -0.337580104  0.262732546  1.011701155
  [64] -0.975947805 -0.112047362  0.146076413 -0.669098504  1.496243576  0.844400295  1.084929176 -1.116566886  0.107838777
  [73]  0.421772731  1.183048690  0.953664041  0.617545729 -0.983534483 -0.632991749  1.619904757  0.156371808  0.724626941
  [82]  0.437281234  1.052789951  1.894725409 -1.690628317 -0.925400762 -0.213397928  0.953952368 -0.035663186 -0.169792022
  [91] -1.021044921  0.405649001 -0.987101512  1.092944501  0.273068269  0.229228147  1.722412038 -1.641936464 -1.408280293
 [100]  0.757840922  0.574032005 -0.314733672  1.250493954  1.244564400  1.480500559 -1.187509891 -0.924634383 -2.309660213
 [109] -2.566421905 -0.445057440 -0.964146509  0.526072070  0.324484178 -0.093226373 -0.871861564  0.226754285  0.193663501
 [118]  0.388841876  0.677200739  0.734708253 -0.098492390 -0.227273894 -1.295313608  2.237521301  1.202280136 -0.351302043
 [127] -0.330082102 -0.002547687 -0.021627322  1.044185751  1.060758257 -0.116407274 -0.328355867 -1.036799540 -0.129652687
 [136] -0.697781161 -0.176246192 -0.645537427  0.518132805  0.742590394  0.405740953 -1.199546467  1.573124400 -1.527944702
 [145] -0.694886537  0.963542332  0.264125502 -0.316863130 -1.679824406 -0.495347690  1.261311154  1.244289754  0.979942746
 [154] -0.634919870  1.215995596 -0.413980548 -0.336647811  0.987400141 -0.298863148  0.027303948 -1.147226687  0.362426619
 [163]  0.108794954 -1.076180254 -0.958907562  0.277728650  0.890624959 -0.108779112  0.634845016  0.396386999 -0.806886615
 [172]  0.444260702  0.405027103  1.626839187 -0.350507106 -1.722630539  0.228596873 -1.987857444  1.234329085 -1.104723842
 [181]  1.631927320 -1.037984167  1.064845217  0.029914027  0.894784899  0.475649560 -0.475205789  1.971493372 -0.596476804
 [190]  0.925222456  0.258208321  1.866405070 -0.431258000  1.270735143 -0.065323946 -1.995115596  0.378190462  1.086122655
 [199]  0.382247081  1.147012608 -1.923124580 -0.562108674  0.329375168 -0.605815705  1.199930324  0.335577973  1.182487785
 [208]  2.323684734 -0.976277308 -0.043228187 -1.140960569 -0.966218164  0.421865750 -0.256876288  0.266536569 -0.088194288
 [217] -0.067696831  1.252996094  1.214484779 -1.043123802 -0.840936899  0.417033387 -0.353952791 -0.968010423 -1.014008659
 [226]  0.434230826  0.909017032 -1.352969906 -0.616398747 -0.839658551 -1.869304706 -0.575074879  0.186713206  0.545154873
```

c) Deduceţi, pornind de la rata de acceptare a acestui algoritm, o aproximare a *constantei de normalizare* a lui $f(x)$, apoi comparaţi histograma valorilor generate cu reprezentarea grafică a lui $f(x)$ normalizată.

```r
# Calculam  rata de acceptare
acceptance_rate <- length(samples) / total_samples

print(acceptance_rate)

# Aproximam constanta de normalizare a lui f(x)
norm_const <- 1 / acceptance_rate

# Normalizam f(x)
f_norm <- function(x) {
  f(x) / norm_const
}

# Generam câteva valori x pentru plot
x <- seq(-10, 10, length.out = 1000)

# Generam valorile y corespunzătoare pentru f(x) normalizat
y <- f_norm(x)

hist(samples, freq = FALSE, xlim = c(-20, 20), ylim=c(0.0,1), main = "Histograma valorilor si graficul f(x) normalizat", xlab = "x", ylab = "Density")

# f(x) normalizat
lines(x, y, col = "blue", lwd = 2)

legend("topright", legend = c("Valorile", "f(x) normalizat"), col = c("black", "red"), lty = 1)
```



III. Folosind functia check.convergence din pachetul R Convergence Concepts verificati daca pentru urmatoarele exemple sunt verificate convergenta in lege (in distributie), in probabilitate si respectiv convergenta aproape sigura. Interpretati si comentati rezultatele obtinute

1) Fie $X_1, X_2, \dots, X_n$ v.a. i.i.d. $X_i \sim Beta\left(\frac{1}{n}, \frac{1}{n}\right)$ si $X \sim Binomial\left(1, \frac{1}{2}\right)$. Verificati daca $X_n \xrightarrow{\mathcal{D}} X$. Dar in cazul in care $X_i \sim Beta\left(\frac{a}{n}, \frac{b}{n}\right)$, cu $a>0, b>0$?

$X_n \xrightarrow{\Delta} X$ daca $\lim\limits_{n\to\infty} E[X_n] = E[X_n]$.

$\overline{X_n}$ (medio v.a. dist. Beta) $= \frac{1}{n}\sum\limits_{i=1}^{n} X_i \quad \beta(\mu, \nu) = \beta(\frac{1}{n}, \frac{1}{n})$

$$E(X_i) = \frac{\mu}{\mu+\nu} = \frac{\frac{1}{n}}{\frac{1}{n}+\frac{1}{n}} = \frac{\frac{1}{n}}{\frac{2}{n}} = \frac{1}{2}.$$

$$E(\overline{X_n}) = \frac{1}{n}\sum_{1}^{n} E(X_i) = \frac{1}{n}\cdot n \cdot \frac{1}{2} = \frac{1}{2}.$$

$$E(X) = 1\cdot\frac{1}{2} = \frac{1}{2} \text{ (medio v.a. } X).$$

vf. convergentei in distributie: Trebuie sa verificam daca:

$$\underbrace{E(\overline{X_n})}_{\frac{1}{2}} \xrightarrow{n\to\infty} \underbrace{E(X)}_{\frac{1}{2}} \Rightarrow X_n \xrightarrow{\Delta} X$$

$$\beta(\mu, \nu) = \left(\frac{a}{n}, \frac{b}{n}\right)$$

$$E(X_i) = \frac{\frac{a}{n}}{\frac{a}{n}+\frac{b}{n}} = \frac{a}{a+b}$$

$$E(\overline{X_n}) = \frac{1}{n}\cdot\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\cdot\frac{a}{a+b} = \frac{a}{a+b}$$

$$E(X) = 1\cdot\frac{1}{2} = \frac{1}{2}.$$

vf. daca $E(\overline{X_n}) \xrightarrow{n\to\infty} E(X)$ da daca $\frac{a}{a+b} = \frac{1}{2}$ (doar daca $a=b$).

2) Fie $X_1, X_2 \dots X_n$ v.a. i.i.d. uniform distribuite pe mulțimea de valori $\{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$ și $X \sim Unif(0,1)$. Verificați dacă $X_n \xrightarrow{D} X$. Dar $X_n \xrightarrow{\mathbb{P}} X$?

$X_1, \dots, X_n$ uniform distribuite pe mulțimea $\{\frac{1}{n} \dots 1\}$.

$X \sim Unif(a,b)$ La noi $a=0$ $b=1$ $\Rightarrow E(X) = \frac{1}{2}$

$E(X) = \frac{a+b}{2}$

X_i dist. uniform pe mulțimea $\{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$

$$E(X_i) = \frac{1}{n} \sum_{k=1}^{n} x_k = \frac{1}{n}\underbrace{\left(\frac{1}{n} + \frac{2}{n} + \dots + 1\right)}_{\frac{1+2+\dots+n}{n}\ (2)} = \frac{1}{n} \cdot \frac{n+1}{2}$$

$S = 1+2+\dots+n = \frac{n(n+1)}{2}$ (1)

Din (1) și (2) $\Rightarrow E(X_i) = \frac{n+1}{2}$

$$E(\overline{X_n}) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n \cdot \frac{n+1}{2n} = \frac{n+1}{2n}$$

Nu converge către $E(X) = \frac{1}{2} \Rightarrow X_n \not\xrightarrow{D} X$

$X_n \xrightarrow{\mathbb{P}} X$ dacă $\lim\limits_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ $(\forall)\, \varepsilon > 0$.

Adică probabilitatea ca X_n să difere de $X \to 0$ în timp ce $n \to \infty$

Știm $E(\overline{X_n}) = \frac{n+1}{2n}$ și $E(X) = \frac{1}{2}$

$\mathbb{P}\left(\left|\frac{n+1}{2n} - \frac{1}{2}\right| \geq \varepsilon\right)$ $0 < \varepsilon < 1$

$\mathbb{P}\left(\left|\frac{\cancel{n}+1-\cancel{n}}{2n}\right| \geq \varepsilon\right) \Rightarrow \mathbb{P}\left(\left|\frac{1}{2^n}\right| \geq \varepsilon\right) \Rightarrow \mathbb{P}\left(\frac{1}{n^2} \geq \varepsilon\right) \Rightarrow$

$\Rightarrow \mathbb{P}\left(n \leq \frac{1}{2\varepsilon}\right)$

Probabilitatea ca $\overline{X_n}$ să difere cu cel puțin ε de $\frac{1}{2}$ tinde către 0. în timp ce n tinde către ∞. $\Rightarrow X_n \xrightarrow{\mathbb{P}} X$

3) Fie $X_1, X_2, \ldots, X_n$ v.a. i.i.d. Notăm cu m și respectiv M infimumul și respectiv supremumul mulțimii valorilor pe care le poate lua X_1. (i.e. $\mathbb{P}(m \le X \le M) = 1$, $P(X_1 < a) > 0$ și $P(X_1 > b)$ pentru orice $a > m$ și respectiv $b < M$). Verificați că $\min\{X_1, X_2, \ldots, X_n\} \xrightarrow{a.s.} m$ și că $\max\{X_1, X_2, \ldots, X_n\} \xrightarrow{a.s.} M$

$m :=$ infimumul mulțimii valorilor pe care le poate lua X_1
$M :=$ supremumul mulțimii valorilor pe care le poate lua X_1.

$X_n \xrightarrow{a.s.} X$ pt. $n \to \infty$ dacă $\mathbb{P}(\{\lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

$\min\{X_1, \ldots, X_n\} \xrightarrow{a.s.} m$

$\mathbb{P}(\lim_{n\to\infty} \min\{X_1 \ldots X_n\} = m) = 1$

Fie $A_n = \{\omega \in \Omega : \min\{X_1(\omega), \ldots, X_n(\omega)\} > m\}$.

$\mathbb{P}(A_n) = \mathbb{P}(\min\{X_1, X_2, \ldots, X_n\} > m)$

Vreau să arăt că $\lim_{n\to\infty}(\mathbb{P}(A_n)) = 0$ (ad. probabilitatea evenimentelor A_n se apropie de 0 pe măsură ce $n \to \infty$)

Pt. că $\mathbb{P}(X_i > m) > 0$ $(\forall) i$ (infimum este cel mai mic element, iar probabilitatea să obții o valoare mai mare ca m este nenulă).

$A_{n+1} \subset A_n$

$\xrightarrow{\text{cont.}} \lim_{n\to\infty} \mathbb{P}(A_n) = 0 \Rightarrow \mathbb{P}(\ ) = 1$

$\max\{X_1, \ldots, X_n\} \xrightarrow{a.s.} M$.

Fie $B_n = \{\omega \in \Omega : \max\{X_1, \ldots\} < M\}$

Doresc $\lim_{n\to\infty} \mathbb{P}(B_n) = 0$

Pentru că: $\mathbb{P}(X_i < M) > 0$, $(\forall)(i)$

$B_{n+1} \subset B_n$

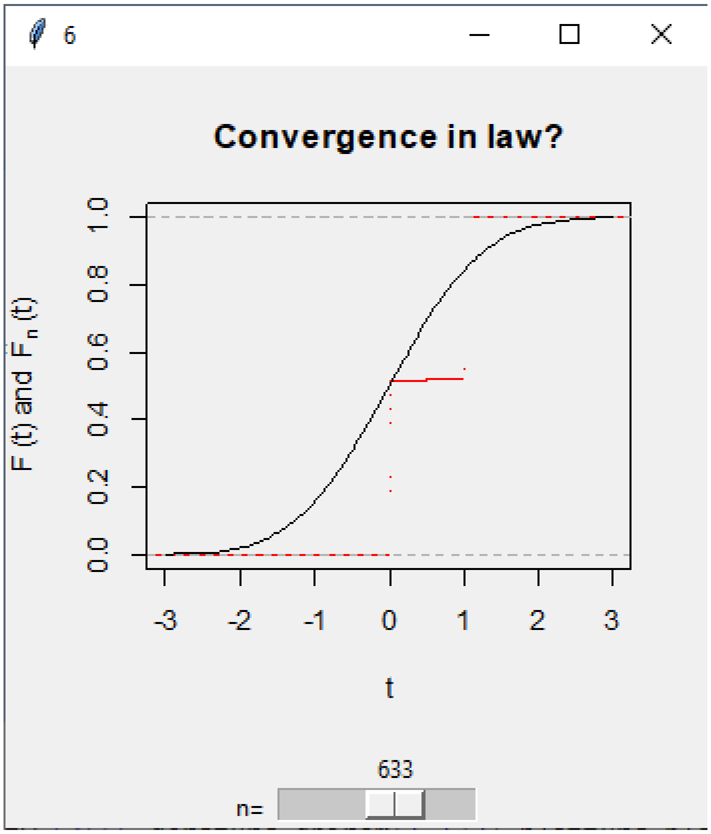
$\Big\} \lim_{n\to\infty} \mathbb{P}(B_n) = 0 \Rightarrow \mathbb{P}(\ ) = 1$

Continuitatea probabilității: Avem o secvență A_n de evenimente $\to$ atunci probabilitatea acestora converge către probabilitatea evenimentelor limită: $\lim \mathbb{P}(A_n) = \mathbb{P}(A)$

III. Folosind funcţia ***check.convergence*** din pachetul R ***ConvergenceConcepts*** verificaţi dacă pentru următoarele exemple sunt verificate convergenţa în lege(în distribuţie), în probabilitate şi respectiv convergenţa aproape sigură. Interpretaţi şi comentaţi rezultatele obţinute.

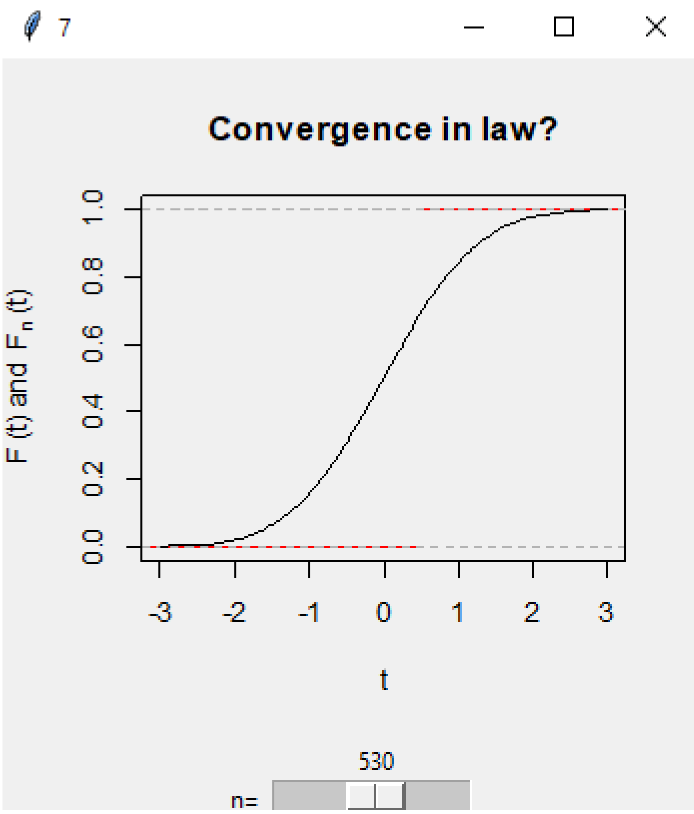
1) Fie $X_1, X_2, \ldots X_n$ v.a. i.i.d $X_i \sim Beta(\frac{1}{n}, \frac{1}{n})$ şi $X \sim Binomial(1, \frac{1}{2})$. Verificaţi dacă $X_n \xrightarrow{D} X$. Dar în cazul în care $X_i \sim Beta(\frac{a}{n}, \frac{b}{n})$, cu $a>0, b>0$?

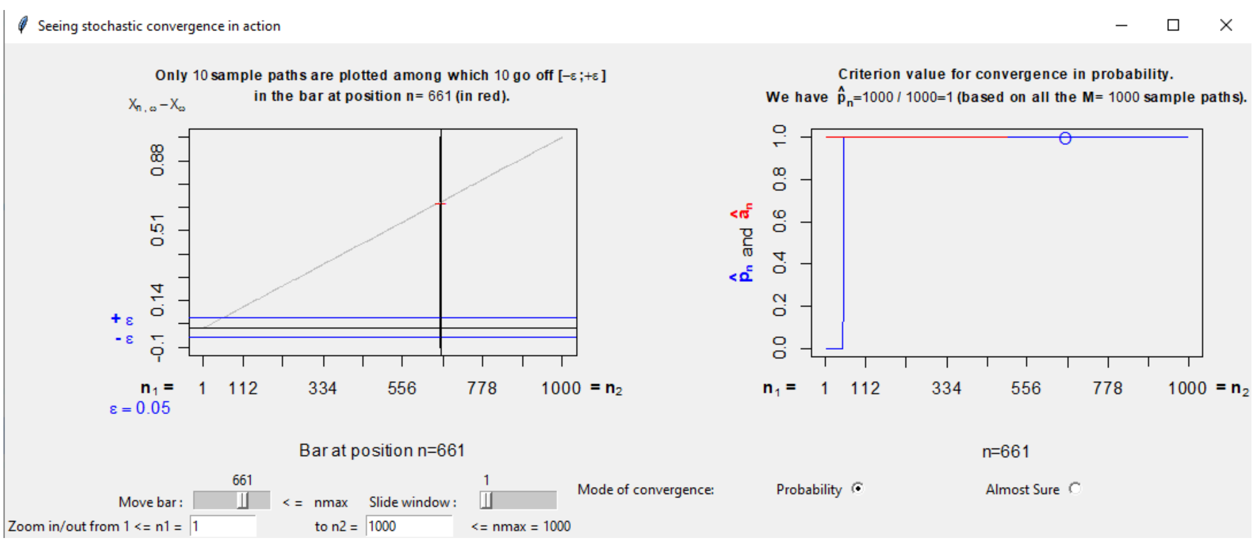
```
library(ConvergenceConcepts)

#1

# Numărul de eșantioane
n <- 1000

# Definirea lui a_n și b_n
a_n <- 2
b_n <- 3

# Generare de eșantioane conform distribuției Beta


betagen1 <- function(n) {
  result <- numeric(n)
  for(i in 1:n) {
    result[i] <- rbeta(1, 1/i, 1/i)
  }
  return(result)
}

betagen2 <- function(n)
{
  rbeta(n,3/n,3/n)
}

check.convergence(nmax = 1000,M=1000,genXn=betagen1,mode="L",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)


check.convergence(nmax = 1000,M=1000,genXn=betagen1,mode="p",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)


check.convergence(nmax = 1000,M=1000,genXn=betagen1,mode="as",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)



check.convergence(nmax = 1000,M=1000,genXn=betagen2,mode="L",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)


check.convergence(nmax = 1000,M=1000,genXn=betagen2,mode="p",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)


check.convergence(nmax = 1000,M=1000,genXn=betagen2,mode="as",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=dbinom(1,1/2),plotfunc=plot)
```



2) Fie $X_1, X_2, \ldots X_n$ v.a. i.i.d uniform distribuite pe mulţimea de valori $\{\frac{1}{n}, \frac{2}{n} \ldots 1\}$ şi $X \sim Unif(0,1)$. Verificaţi dacă $X_n \xrightarrow{D} X$. Dar $X_n \xrightarrow{P} X$?

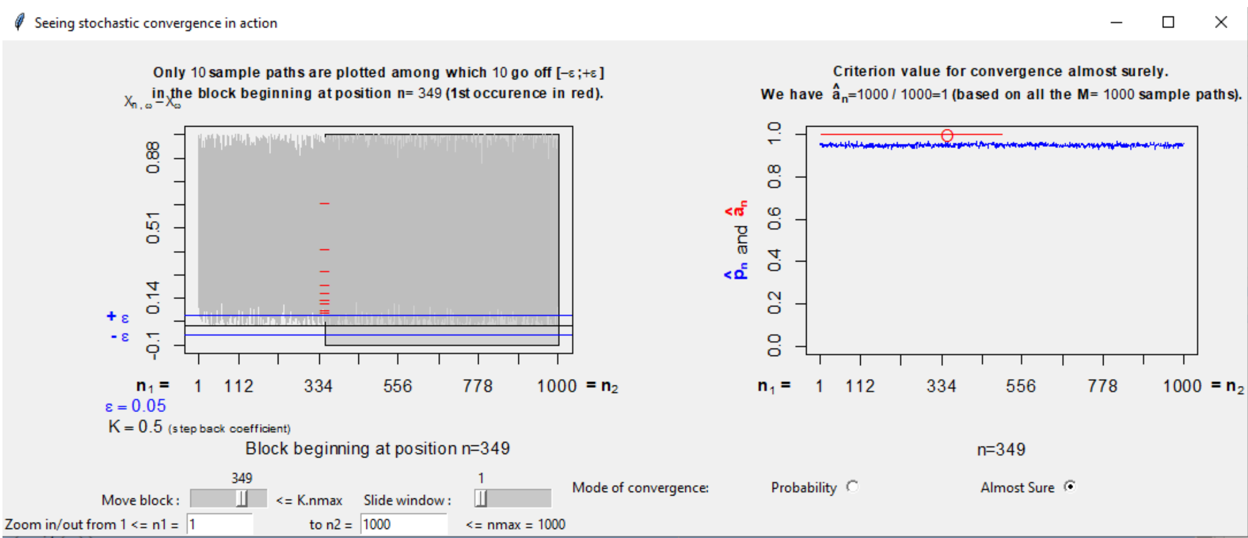
```
#2

func1 <- function(n)
{
  result1 <- numeric(n)
  for(i in 1:n) {
    result1[i] <- i/n
  }
  return(result1)
}


check.convergence(nmax = 1000,M=1000,genXn=func1,mode="L",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=runif(0,1),plotfunc=plot)
check.convergence(nmax = 1000,M=1000,genXn=func1,mode="p",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=runif(0,1),plotfunc=plot)
check.convergence(nmax = 1000,M=1000,genXn=func1,mode="as",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=runif(0,1),plotfunc=plot)
```





3) Fie $X_1, X_2, \ldots X_n$ v.a. i.i.d . Notăm cu m şi respectiv M infimumul şi respectiv supremumul mulţimii valorilor pe care le poate lua X_1.
(i.e. $P(m \le X \le M)=1$, $P(X_1<a)>0$ şi $P(X_1>b)>0$ pentru orice $a>m$ şi respectiv $b<M$).
Verificaţi că $\min\{X_1, X_2 \ldots X_n\} \xrightarrow{a.s.} m$ şi că $\max\{X_1, X_2 \ldots X_n\} \xrightarrow{a.s.} M$.

```r
library(ConvergenceConcepts)

n <- 1000
unif <- function(n) {
  runif(n, min = 0, max = 1)
}

m <- min(unif(n))
M <- max(unif(n))


check.convergence(nmax = 1000,M=1000,genXn=unif,mode="as",epsilon=0.05,r=NULL,nb.sp=10,
                  density=FALSE,densfunc=runif(0,1),plotfunc=plot)
```



IV. O problemă de LOTO

Se consideră date următoarele informații despre un bilet la LOTO 6/49:

- Un bilet simplu, pentru care se pot alege 6 numere din 49 disponibile costă 7 lei.
- În fiecare săptămână există două extrageri.
- În urma extragerii, dacă exact 3 numere din cele alese se regăsesc printre cele extrase se primește suma de 30 de lei, pentru exact 4 numere fondul de câștig este 363 350 lei, pentru exact 5 numere fondul de câștig este 1 090 000 lei. Cu excepția primei situații, în care se primește o sumă fixă, fondul de câștiguri se împarte la numărul de câștigători din categoria respectivă.
- Un bilet asociat unei variante combinate constă în alegerea între 7 și 48 numere (echivalent cu combinări de n luate câte 6 variante simple, $6 < n < 49$). Prețul asociat este calculat înmulțind 7 lei cu câte variante simple sunt echivalente cu acea variantă compusă.

Cerințe:

1) Folosindu-vă de simulare (10^6 valori) estimați următoarele cantități, iar acolo unde este posibil, comparați cu rezultatul teoretic.

a) Probabilitatea de a câștiga din prima încercare, cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.

$$\mathbb{P}(W_3) = \frac{k_3}{t} = \frac{6!}{3!\,(6-3)!} \cdot \frac{43!}{3!\cdot(43-3)!} \cdot \frac{43!\cdot 6!}{49!} =$$

$$= \frac{6!\cdot 43!\cdot 43!\cdot 6!}{3!\cdot 3!\ 3!\cdot 40!\cdot 49!} = \frac{\cancel{4\cdot 5}\cdot 41\cdot \overset{21}{\cancel{42}}\cdot 43\cdot \overset{2}{\cancel{4}}\cdot 5\cdot \cancel{6}}{\underset{11}{\cancel{44}}\cdot \underset{9}{\cancel{45}}\cdot \underset{23}{\cancel{46}}\cdot 47\cdot \underset{\cancel{8}\,2}{\cancel{48}}\cdot 49} =$$

$$= \frac{41\cdot 21\cdot 43\cdot 5}{11\cdot 9\cdot 23\cdot 47\cdot 2\cdot 49} = \frac{8815}{499422} = 0{,}017.$$

$$\mathbb{P}(W_4) = \frac{6!}{4!\,(6-4)!}\cdot\frac{43!}{2!\,(43-2)!}\cdot\frac{43!\cdot 6!}{49!} =$$

$$= \frac{5\cdot\cancel{6}\,3}{\cancel{2}}\cdot\frac{\overset{21}{\cancel{42}}\cdot 43}{1\cdot\cancel{2}}\cdot\frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}{\underset{11}{44}\cdot\underset{9}{45}\cdot\underset{23}{46}\cdot 47\cdot\underset{8}{48}\cdot 49} =$$

$$= \frac{15\cdot\overset{3}{\cancel{21}}\cdot 43\cdot 3}{11\cdot 9\cdot 23\cdot 47\cdot 8\cdot\underset{7}{\cancel{49}}} = \frac{\overset{5}{\cancel{15}}\cdot\cancel{3}\cdot 43\cdot 3}{11\cdot\underset{3}{\cancel{9}}\cdot 23\cdot 47\cdot 8\cdot 7} = \frac{5\cdot 43\cdot 3}{11\cdot 23\cdot 47\cdot 8}$$

$$= \cancel{645}\ \cancel{645}$$

$$= \frac{645}{665896} = 0{,}0009 .$$

$$\mathbb{P}(W_5) = \frac{6!}{5!\,(6-5)!}\cdot\frac{43!}{(43-1)!\cdot 1!}\cdot\frac{43!\cdot 6!}{49!} =$$

$$= \frac{6!}{5!\cdot 1!}\cdot\frac{43!}{42!\cdot 1!}\cdot\frac{6!}{44\cdot 45\cdot 46\cdot 47\cdot 48\cdot 49} =$$

$$= \frac{\cancel{6}\cdot 43\cdot 1\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot\cancel{5}\cdot\cancel{6}\cdot\cancel{2}}{44\cdot\underset{\cancel{3}}{\underset{\cancel{9}}{\cancel{45}}}\cdot\underset{23}{\cancel{46}}\cdot 47\cdot\underset{\cancel{8}_1}{\cancel{48}}\cdot 49} = \frac{43}{44\cdot 23\cdot 47\cdot 49} =$$

$$= \frac{43}{2330636} = 0{,}00001$$

$$\mathbb{P}(W_6) = \frac{K_6}{t} = \frac{6!}{6!\,(6-6)!}\cdot\frac{43!\cdot 43!\cdot 6!}{(6-6)!\cdot(43-0)!\cdot 49!} = \frac{15\ldots}{44\cdot 45\cdot 46\cdot 47\cdot 48}$$

$$= \frac{1}{132\cdot 36\cdot 47\cdot 49} \simeq 0{,}00\ldots 7 .$$

b) Probabilitatea de a câștiga după k încercări (k fiind un parametru), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.

$\mathbb{P}$(câștig într-o încercare din primele k) =

$= 1 - \mathbb{P}$(nimic nu este câștigat în primele k).

$1 - \mathbb{P}(W_3)$ = probabilitatea de a nu câștiga într-o singură încercare.

Probabilitatea de a nu câștiga de-a lungul celor k încercări consecutive:

$\mathbb{P}$(nimic nu este câștigat în primele k) =

$$= \underbrace{(1-\mathbb{P}(W_3))^k \cdot (1-\mathbb{P}(W_4))^k \cdot \ldots \cdot (1-\mathbb{P}(W_6))^k}_{\star}$$

$\mathbb{P}$(câștig în cel puțin o încercare din primele k) =

$= 1 - \star$

$$\Rightarrow 1 - \prod_{i=1}^{k} (1-\mathbb{P}(W_i)).$$

Pentru 3 numere ghicite: $\mathbb{P}$(cel puțin o dată în pr. k)

$$= 1-(1-\mathbb{P}(W_3))^k = 1-(1-0{,}017)^k = 1-(0{,}983)^k$$

Pentru 4 numere ghicite:

$$= 1-(1-\mathbb{P}(W_4))^k = 1-(1-0{,}0009)^k = 1-(0{,}9991)^k$$

Pentru 5 numere găsite:

$$= 1-(1-\mathbb{P}(W_5))^k = 1-(1-0{,}00001)^k = 1-(0{,}99999)^k$$

Pentru 6 numere găsite:

$$= 1-(1-\mathbb{P}(W_6))^k = 1-(1-(0{,}00\ldots7))^k = 1-(0{,}9\ldots3)^k.$$

c) Probabilitatea de a câștiga de r ori din k încercări (r și k fiind parametri), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.

Coeficientul binomial.

p_i = probabilitatea de a câștiga o dată pentru categoria W_i (cu $i = 3, 4, 5, 6$)

Probabilitatea de a câștiga de r ori din k încercări:

$$\mathbb{P}(\text{câștig de } r \text{ ori în } k) = \binom{k}{r} \cdot (p_i)^r (1-p_i)^{k-r}.$$

Această formulă consideră că aceste câștiguri sunt independente și că evenimentul de câștig sau pierdere într-o încercare nu afectează probabilitatea în încercările ulterioare

Pentru 3 numere ghicite:

$$\mathbb{P}(\text{câștig de } r \text{ ori în } k \text{ la } W_3) = \binom{k}{r} \cdot \mathbb{P}(W_3)^r \cdot (1-\mathbb{P}(W_3))^{k-r}$$

$$= \binom{k}{r} \cdot (0{,}017)^r \cdot (0{,}983)^{k-r}$$

$\Rightarrow$ Pentru 4 numere: $\binom{k}{r} \cdot \mathbb{P}(W_4)^r \cdot (1-\mathbb{P}(W_4))^{k-r} =$

$$= \binom{k}{r} \cdot (0{,}0009)^r (0{,}9991)^{k-r}$$

Pentru 5 numere: $\binom{k}{r} \cdot \mathbb{P}(W_5)^r \cdot (1-\mathbb{P}(W_5))^{k-r} =$

$$= \binom{k}{r} \cdot (0{,}00001)^r \cdot (0{,}99999)^{k-r}.$$

Pentru 6 numere: $\binom{k}{r} \cdot \mathbb{P}(W_6)^r \cdot (1-\mathbb{P}(W_6))^{k-r} =$

$= \binom{k}{r} \cdot (0{,}00\ldots7)^r \cdot (0{,}99\ldots3)^{k-r}$.

d) Probabilitatea de a câștiga de r ori după k eșecuri (r și k fiind parametri), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.

$q_i = 1 - p_i$ probabilitatea de a nu câștiga

$(i = 3, \ldots, 6)$

$\mathbb{P}(\text{câștig de } r \text{ ori după } k \text{ eșecuri}) = \binom{k+r-1}{r} (p_i)^r \cdot q_i^k$

$$\boxed{\begin{aligned} q_i &= 1 - \mathbb{P}(W_3) \\ p_i &= \mathbb{P}(W_3) \end{aligned}}$$

$\Rightarrow \mathbb{P}(\text{câștig } 3 \text{ nr}) = \binom{k+r-1}{r} \cdot (0{,}017)^r \cdot (0{,}983)^k$

$\mathbb{P}(\text{câștig } 4 \text{ nr}) = \binom{k+r-1}{r} (0{,}0009)^r \cdot (0{,}9991)^k$

$\mathbb{P}(\text{câștig } 5 \text{ nr}) = \binom{k+r-1}{r} (0{,}00001)^r \cdot (0{,}99999)^k$

$\mathbb{P}(\text{câștig } 6 \text{ nr}) = \binom{k+r-1}{r} (0{,}00\ldots7)^r \cdot (0{,}99\ldots3)^k$.

e) Probabilitatea de a câștiga, în medie, cel puțin o dată pe an, dacă se joacă săptămânal câte un bilet simplu timp de 30 de ani.

$\mathbb{P}$(câștig cel puțin o dată pe an în 30 de ani) =

= 1 − $\mathbb{P}$(niciodată nu câștig în 30 de ani)

Probabilitatea de a nu câștiga într-o săptămână este:

$$\left.\begin{array}{l} 1-\mathbb{P}(\text{câștig}) \\ \text{Se extrage de 2 ori} \end{array}\right\} \Rightarrow [1-\mathbb{P}(\text{câștig})]^2$$

Așadar, probabilitatea de a nu câștiga într-un an este:

$(1-\mathbb{P}(\text{câștig}))^{2\times 52}$ și în 30 de ani: $(1-\mathbb{P}(\text{câștig}))^{2\times 52\times 30}$

$\mathbb{P}(W_3) = 0{,}017$

$1-\mathbb{P}(W_3) = 0{,}983 \Rightarrow (0{,}983)^{2\times 52\times 30}$ (pentru 3 nr)

$\mathbb{P}(W_4) = 0{,}0009$

$1-\mathbb{P}(W_4) = 0{,}9991 \Rightarrow (0{,}9991)^{2\times 52\times 30}$ (pentru 4 nr)

$\mathbb{P}(W_5) = 0{,}00001$

$1-\mathbb{P}(W_5) = 0{,}99999 \Rightarrow (0{,}99999)^{2\times 52\times 30}$ (pentru 5 nr)

$\mathbb{P}(W_6) = 0{,}00\ldots 7$

$1-\mathbb{P}(W_6) = 0{,}9\ldots 3 \Rightarrow (0{,}9\ldots 3)^{2\times 52\times 30}$ (pentru 6 nr)

f) Probabilitatea de a juca săptămânal, timp de un an câte un bilet simplu și a nu câștiga niciodată.

P(nu câștigăm într-o săptămână) pentru un an:

$P(\text{nu câștig})^{2\times52} = (1 - P(W_3))^{2\times52} = (0,983)^{2\times52}$ 3 nr.

$P(\text{nu câștig})^{2\times52} = (1 - P(W_4))^{2\times52} = (0,9991)^{2\times52}$ 4 nr

$P(\text{nu câștig})^{2\times52} = (1 - P(W_5))^{2\times52} = (0,99999)^{2\times52}$ 5 nr.

$P(\text{nu câștig})^{2\times52} = (1 - P(W_6))^{2\times52} = (0,9\ldots3)^{2\times52}$ 6 nr.

g) Probabilitatea de a juca săptămânal, timp de un an câte un bilet simplu și de a câștiga, cumulat, o sumă mai mare decât costul total al biletelor jucate.

$P(\text{câștig într-o săptămână la } W_i)$

$P(\text{câștig într-o săptămână}) = P(W_3) + \ldots + P(W_6)$

$P(\text{câștig cumulat} > \text{costul total}) =$

$= 1 - (1 - P(\text{câștigăm într-o săpt}))^{2\times52} = 1 - (0,983)^{2\times52}$ 3 nr

$P(W_3) = 0,017$

$1 - (0,9991)^{2\times52}$ 4 nr

$1 - (0,99999)^{2\times52}$ 5 nr.

$1 - (0,9\ldots3)^{2\times52}$ 6 nr.

Profit $= 0,017 \cdot 30 + 0,0009 \cdot 363350 + 0,00001 \cdot 390.000 +$
$+ 0,00007 \cdot 1090000 = 0,51 + 32,7015 + 3,9 + 76,3 =$
$= 113,4115$

Total $= 30 + 363350 + 390000 + 1090000 = 1843380$.

2) Folosindu-vă de simulare și / sau rezultate teoretice răspundeți justificat la următoarele întrebări:

a) Care este câștigul mediu anual estimat al Loteriei, știind că există 2 extrageri săptămânale?

Avem două extrageri săptămânale, ceea ce înseamnă că avem $2 * 52 = 104$ extrageri pe an.

Vom calcula câștigul mediu anual estimat pentru fiecare categorie de premii și apoi le vom aduna pentru a obține câștigul total.

Câștigul pentru 3 numere corecte:
30 lei x nr. mediu de câștigători la categoria de 3 nr.

Câștigul pentru 4 numere corecte:
Fondul de câștig pentru 4 numere corecte / numărul mediu de câștigători la categoria de 4 numere: $\frac{363350 \text{ lei}}{\text{nr. câștigători}(4)}$

Câștigul pentru 5 numere corecte:
Fondul de câștig pentru 5 numere / numărul mediu de câștigători la categoria de 5 numere: $\frac{390000 \text{ lei}}{\text{nr. câștigători}(5)}$

Câștigul pentru 6 numere corecte:
Fondul de câștig pentru 6 numere / numărul mediu de câștigători la categoria de 6 numere. $\frac{1090000 \text{ lei}}{\text{nr. câștigători}(6)}$

Pentru a calcula numărul mediu de câștigători la o anumită categorie, putem utiliza formula:

$$\text{Nr. mediu câștigători} = \frac{\text{Nr. total de bilete vândute}}{\text{Șansele de câștig la acea categorie}}$$

Spre exemplu, pentru categoria de 3 numere corecte, unde câștigul este de 30 de lei, șansele de a ghici exact 3 numere sunt date de combinatorică:

$$\binom{6}{3} = \frac{6!}{3!(6-3)!} = 20.$$

Scanned with CamScanner

Apoi putem utiliza această formulă pentru a estima numărul mediu de câștigători la 3 nr: $\frac{\text{Nr. bilete vândute}}{20}$.

4.2b. Care este numărul minim de bilete simple care trebuie cumpărate într-un an pentru ca Loterio să nu fie în pierdere?

Pentru a determina numărul minim de bilete simple care trebuie cumpărate într-un an pentru ca Loterio să nu fie în pierdere, vom lua în considerare costurile și câștigurile asociate fiecărui tip de bilet.

Fie : C = costul unui bilet simplu (7 lei)

N = nr. total de bilete simple vândute într-un an

E = fondul total de câștiguri pentru un an

Vom avea: E = 104 x (30 x nr. mediu câștigători la 3 nr + fondul de câștig pentru 4 nr + fondul de câștig pt. 5 numere + fondul de câștig pentru 6 nr)

În plus, știm că E trebuie să fie mai mare sau cel puțin egal cu C x N, astfel încât să acopere costurile de producție și să permită o anumită marjă de profit.

$C \times N \leq E$

c) Un jucător are un buget lunar de 70 de lei. Care ar fi strategia optimă de joc (bilete simple jucate în săptămâni diferite, un singur bilet cu o variantă complexă jucată o singură dată etc.) care să-i aducă cele mai mari șanse de câștig? Estimați / Calculați aceste șanse!

Simulările sau calculele specifice pot fi necesare pentru a obține estimări precise ale șanselor de câștig. În plus, trebuie să ținem cont de faptul că jocurile de noroc au un grad ridicat de incertitudine și că există șanse de pierdere a întregului buget fără a câștiga.

Ar putea fi util să încerci diferite strategii pe termen scurt și să analizezi rezultatele pentru a determina care se potrivește cel mai bine obiectivelor și preferințelor tale.

Calculul exact al șanselor de câștig depinde de mai mulți factori, cum ar fi numărul exact de bilete simple jucate în fiecare săptămână sau numărul de numere alese pentru varianta complexă. Totuși pot oferi o estimare simplificată pentru 2 scenarii:

Bilete simple în fiecare săptămână:

Cost lunar = $4 \times 7 = 28$ lei

Presupunem că joci un bilet simplu în fiecare săptămână.

Șanse de câștig: depind de numărul de bilete simple jucate. Dacă joci un singur bilet, șansele de a ghici cele 6 numere sunt: $\frac{1}{\binom{49}{6}}$, ceea ce este destul de mic.

Bilet cu o variantă complexă jucată o singură dată:

Cost lunar: o singură dată X lei.

Presupunem că joci un bilet combinat cu 12 numere (de exemplu, $n=12$ pentru $\binom{n}{6} = 924$ combinații posibile)

Șanse de câștig: depind de numărul de numere corecte ghicite din cele 12 alese.

În ambele cazuri, șansele de câștig sunt relativ mici, deoarece este vorba de un joc de noroc. Costul inițial pentru varianta complexă este mai mare, dar șansele de câștig pot fi mai mari decât la biletele simple.

Pentru a face estimări mai precise, ar trebui să folosim date statistice reale și să ținem cont de distribuția istorică a numerelor extrase. În plus, este important să menționăm că jocurile de noroc implică întotdeauna un grad de incertitudine și că rezultatele reale pot să varieze.

~~Calculul~~

d) Care este numărul mediu de câștigători de la fiecare categorie de extragere?

Calculul numărului mediu de câștigători pentru fiecare categorie depinde de șansele de câștig pentru fiecare categorie și de numărul total de bilete vândute. Fără date exacte privind numărul real de bilete vândute sau distribuția numerelor pe bilete, vom face o analiză generală.

Presupunem că N este numărul total de bilete vândute pentru o extragere specifică. Vom utiliza combinatorica pentru a estima șansele de câștig pentru fiecare categorie.

Categoria 3 nr. corecte:

- Șansele de câștig sunt: $\dfrac{\binom{6}{3} \times \binom{43}{3}}{\binom{49}{6}}$ pentru 3 numere
- Numărul mediu de câștigători pentru această categorie este aproximativ $N \times \dfrac{\binom{6}{3} \times \binom{43}{3}}{\binom{49}{6}}$

Categoria 4 nr. corecte:

- șanse: $\dfrac{\binom{6}{4} \times \binom{43}{2}}{\binom{49}{6}}$
- Nr mediu: $N \times$ șanse $= N \times \dfrac{\binom{6}{4} \times \binom{43}{2}}{\binom{49}{6}}$

Categorie 5 nr corecte:

- Șanse: $\dfrac{\binom{6}{5} \times \binom{43}{1}}{\binom{49}{6}}$
- Nr. mediu: $N \times \dfrac{\binom{6}{5} \times \binom{43}{1}}{\binom{49}{6}}$

Categorii 6 nr:

- Șanse: $\dfrac{\binom{6}{6} \times \binom{43}{0}}{\binom{49}{6}}$

- Nr. mediu: $N \times \dfrac{\binom{6}{6} \times \binom{43}{0}}{\binom{49}{6}}$

Aceasta este o analiză simplificată, iar valoarea exactă depinde de multe variabile, inclusiv comportamentul jucătorilor și de numărul real de bilete vândute. Simulările cu date istorice ar putea furniza estimări mai precise

Pentru a calcula șansele de câștig pentru o categorie specifică într-un joc de tip LOTO 6/49 unde trebuie să ghicești un anumit număr de numere corecte dintr-un set de numere, poți utiliza formula combinatorică. Formula generală pentru șansele de câștig în acest context este dată de coeficientul binomial

Formula combinatorică pt. coef. binomial: $\binom{n}{k}$ este:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Aici, pentru a aplica această formulă la un joc LOTO 6/49 unde trebuie să alegi k numere corecte dintr-un set de numere (de exemplu $k=3$ (categoria de 3 nr. corecte)) formula pentru șansele de câștig este:

$$\text{Șanse} = \frac{\binom{6}{k} \times \binom{43}{6-k}}{\binom{49}{6}}$$ unde:

$\binom{6}{k}$ = nr. de moduri în care poți alege k nr. corecte din cele 6 alese.

$\binom{13}{6-k}$ = nr. de moduri în care poți alege restul de 6−k nr. din cele 13 rămase neales e

Aceasta este formula generală pentru calcularea șanselor de câștig pentru o categorie specifică în joc LOTO 6/49. Este important să ajustezi valorile lui k și n în funcție de categoria de câștig pentru care vrei să calculezi șansele.

4.3. Loterio doreşte să introducă un nou tip de bilet, biletul „stânga-dreapta". Pentru acest tip de bilet se consideră câştigătoare situatia în care fie au fost alese exact numerele câştigătoare, fie vecinii săi imediati din stânga sau din dreapta, la nivel de rând, conform aranjării numerelor pe biletul LOTO. De exemplu, dacă este extras 10, dar jucătorul a ales 9 atunci va f. considerat număr câştigata, însă nu se întâmplă acelaşi lucru pentru 11 (pentru că se află pe alt rând).

a) Estimati cât ar trebui să fie pretul unui asemenea bilet pentru ca jucătorii să fie interesaţi să-l cumpere.

Estimarea pretului unui nou tip de bilet, cum ar f biletul „stânga dreapta", poate implica analiza atât a şanselor de câştig, cât şi a costurilor asociate cu dezvoltarea şi implementarea noului joc. Vom încerca să formulăm o estimare simplificată, luând în considerare câteva aspecte important

Presupunem că, pentru acest bilet, jucătorii trebuie să aleagă un anumit număr de elemente (numere) şi ca acestea trebuie să coincidă exact cu numerele extrase sau cu vecinii imediaţi din stânga sau din dreapta, la nivel de rând. Vom presupune că există N nr. în total din care jucătorii pot alege.

Calculul şansela de câştig:

Şansele de a ghici exact numerele extrase sunt $\frac{1}{N}$ (presupunând că jucătorii trebuie să ghicească un singur număr).

Şansele de a ghici vecinii imediaţi din stânga sau din dreapta sunt $\frac{2}{N}$ (presupunând că jucătorii trebuie să generoase unui dintre vecini.

Calculul costului:

Costul unui bilet ar putea fi calculat în funcție de probabilitatea de câștig și de fondul total de câștiguri dorit de organizatorul loteriei.

$$\text{Preț bilet} = \frac{\text{Fondul total de câștiguri}}{\text{Probabilitatea totală de câștig}}$$

Estimarea probabilităților poate depinde de regulile exacte ale jocului, cum ar fi numărul de numere care pot fi alese și modul în care vecinii sunt definiți.

Este important să ținem cont de faptul că prețul biletului trebuie să fie atractiv pentru jucători, dar și să acopere costurile de organizare și să ofere o eventuală marjă de profit.

O analiză mai detaliată și consultarea cu specialiști în jocuri de noroc ar putea oferi o evaluare mai precisă a prețului potențial pentru acest nou tip de bilet.

k) Estimați cât ar trebui să fie prețul minimum al unui asemenea bilet pentru ca Loteria să fie interesată să-l comercializeze.

Estimarea exactă a prețului biletului depinde de informații specifice pe care nu le avem la dispoziție, cum ar fi fondul total de câștiguri dorit (F) și numărul total de numere disponibile pentru alegere (N). Fără aceste date putem face o estimare generală, dar trebuie să ținem cont de faptul că aceasta este o simplificare.

Să luăm un exemplu simplificat:

- Dacă organizatorul loteriei dorește un fond total de câștiguri (F) de, să zicem, 1.000.000 de lei și există un număr total de $N = 49$ numere disponibile, atunci prețul biletului ar putea fi estimat aproximativ astfel:

$$\text{Preț bilet} \approx \frac{N \times F}{3} = \frac{49 \cdot 1.000.000}{3} \approx 16,333,333 \text{ lei}$$

Acesta este un exemplu simplificat și valoarea reală poate varia în funcție de mai mulți factori. Este important să ținem cont de faptul că prețul biletului trebuie să fie atractiv pentru jucători, să acopere costurile de organizare și să permită o eventuală marjă de profit. Pentru o evaluare mai precisă, se recomandă consultarea cu specialiști în domeniu sau analiza datelor de piață relevante.

Scanned with CamScanner

d) Construiți o strategie de a alege numerele cu cele mai mari șanse de câștig cu un asemenea bilet și evaluați care sunt aceste șanse.

Pentru a construi o strategie de alegere a numerelor cu cele mai mari șanse de câștig, în cazul unui bilet „stânga-dreapta", vom analiza probabilitățile asociate cu fiecare scenariu de câștig.

Presupunem că există un total de N numere din care jucătorii pot alege. Având în vedere cele două modalități de câștig: ghicirea exactă a numerelor și ghicirea vecinilor imediați din stânga sau dreapta, putem evalua strategia astfel:

Ghicirea exactă a numerelor extrase:
În acest caz, orice număr ales are o șansă de câștig de $\frac{1}{N}$. Strategia aici ar fi să alegi numerele care să acopere cât mai multe posibilități din cele N posibile.

Ghicirea vecinilor imediați din stânga sau dreapta:
În acest caz, oricare dintre cei doi vecini imediați ai unui număr ales are o șansă de câștig de $\frac{2}{N}$. Strategia aici ar fi să alegi numere care au vecini imediați, care acoperă o gamă cât mai largă de posibilități.

O strategie simplificată ar putea fi să alegi numere în așa fel încât să acoperi cât mai multe opțiuni posibile pentru ambele scenarii. Cu toate acestea, să alegi numere în mod complet aleatoriu poate fi la fel de eficient, deoarece toate nr. au șanse egale de a fi extrase.

Evaluarea exactă a șanselor depinde de nr. total de numere. (N) De exemplu, dacă $N=49$, probabilitatea totală de câștig este de 3 (conform calculului anterior

Astfel, într-un format de stânga, şansele de câştig sunt relativ mici în acest tip de joc, iar alegerea numerelor nu va influenţa semnificativ aceste şanse.

1. Ghicirea exactă a numerelor extrase:

· Fiecare număr ales are o şansă de $\frac{1}{N} = \frac{1}{10}$ de a fi extras exact.

· Pentru un singur număr ales, probabilitatea de ghicire a exact acelui număr este $\frac{1}{10}$!

2. Ghicirea vecinilor imediaţi din stânga sau dreapta:

· Fiecare număr ales are doi vecini imediaţi!

· Probabilitatea de ghicire a unui vecin este $\frac{2}{N} = \frac{2}{10} = \frac{1}{5}$

3. Probabilitatea totală de câştig.

· Probabilitatea totală este suma celor două scenarii

$$\frac{1}{10} + \frac{1}{5} = \frac{3}{10}.$$

$\frac{3}{10}$ este probabilitatea totală de câştig pentru biletul „stânga-dreapta."

e) În ce condiții un asemenea bilet ar aduce pentru un jucător regulat un câștig mediu anual estimat ca fiind superior celui unui bilet simplu?

Pentru a determina condițiile în care un bilet „stânga-dreapta" ar aduce un câștig mediu anual estimat superior celui al unui bilet simplu, vom compara probabilitățile de câștig și fondurile de câștiguri oferite de cele două tipuri de bilete.

Presupunând că un jucător regulat joacă constant și dorește să maximizeze șansele de câștig, va compara probabilitatea totală de câștig pentru biletul „stânga-dreapta" cu probabilitatea de câștig a unui bilet simplu. Aceasta ar putea implica ajustarea strategiei de alegere a numerelor pentru a obține o probabilitate totală de câștig mai mare pe biletul „stânga-dreapta".

Condițiile în care un bilet „stânga-dreapta" poate aduce un câștig mediu anual superior ar putea include:

1. Probabilitatea totală de câștig mai mare: Dacă probabilitatea totală de câștig pentru biletul „stânga-dreapta" este semnificativ mai mare decât probabilitatea de câștig a unui bilet simplu, aceasta ar putea oferi mai multe șanse de câștig.

2. Fondurile de câștiguri pentru fiecare scenariu: Dacă fondurile de câștiguri pentru ghicirea exactă a numerelor sau vecinilor imediați sunt mai mari pe biletul „stânga-dreapta" decât fondurile oferite pentru ghicirea a 3, 4, 5 sau 6 numere, pe un bilet simplu, atunci jucătorul ar putea avea șanse mai mari de a obține un câștig semnificativ.

3. Strategia de alegere a numerelor: Jucătorul ar trebui să adopte o strategie de alegere a numerelor care să maximizeze șansele de câștig pentru biletul „stânga-dreapta". Alegerea unor numere care acoperă cât mai multe opțiuni posibile în ambele scenarii (ghicirea exactă și a vecinilor imediați) ar putea crește probabilitatea totală de câștig.

Este important să analizăm atent regulile exacte ale jocului, probabilitățile de câștig și fondurile de câștiguri pentru a determina în ce condiții biletul „stânga-dreapta" ar putea aduce un câștig mediu-anual mai mare pentru un jucător regulat în comparație cu un bilet simplu.

Să luăm un exemplu concret pentru a ilustra condiții în care un bilet „stânga-dreapta" ar putea aduce un câștig mediu anual mai mare decât un bilet simplu. Vom presupune un fond de câștiguri (¢) de 1.000.000 de lei pentru ambele tipuri de bilete.

Bilet simplu: Probabilitatea de câștig pentru un bilet simplu este de $\frac{1}{13,983,816}$ (probabilitatea de a ghici cele 6 nr)

Fondul de câștig pentru câștigătorii cu 6 nr: 1.090.000 lei

$$Câștig\_mediu\_anual\ (Bilet\ Simplu) = \frac{Prob\_Câștig \times Fond\_Câș}{2}$$

$$= \frac{\frac{1}{13,983,816} \cdot 1,090,000}{2}$$

Bilet „stânga-dreapta": Probabilitatea totală de câștig pentru biletul „stânga-dreapta" este $\frac{3}{10}$ (conform calculelor anterioare).

Fondul de câștig pentru un câștigător cu biletul
"Stația - dreaptă" trebuie să fie ajustat astfel încât
să asigure câștiguri medii comparabile cu cele
ale unui bilet simplu

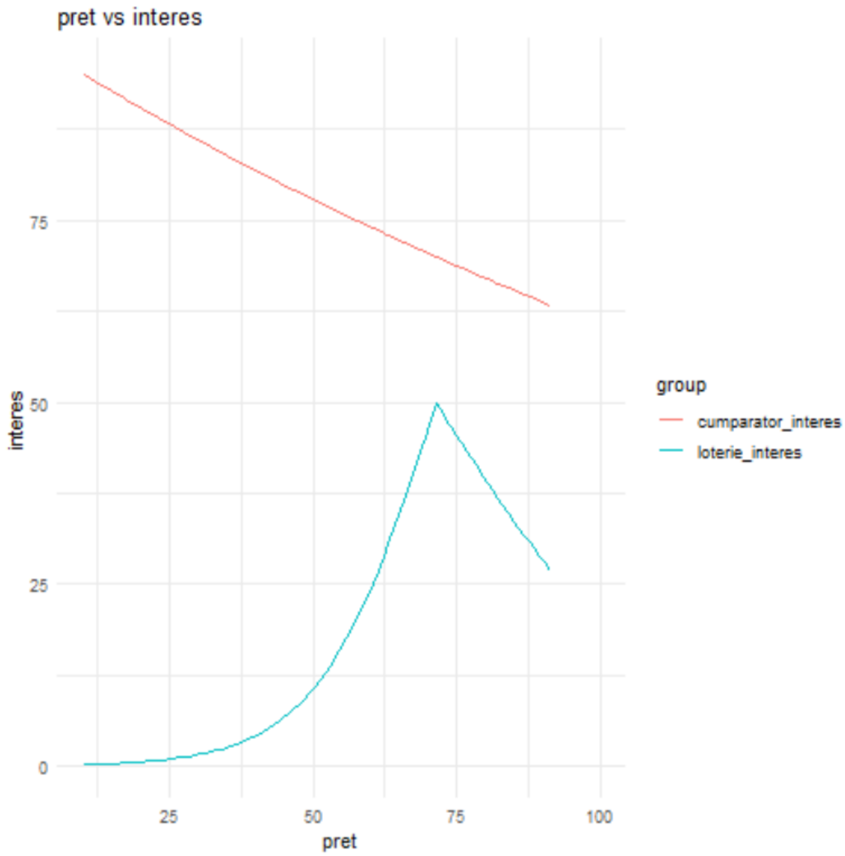
Câștig_mediu_anual (Bilet „St. dr") = $\frac{\text{Prob\_Câștig} \times \text{Fond\_C}}{2}$

$= \frac{\frac{3}{10} \times 1.000.000}{2} = 150.000.$

IV. O problemă de LOTO

c) Pornind de la a) și b) construiți o animație(folosind, de exemplu, pachetul animation https://www.jstatsoft.org/article/view/v053i01) care să arate cum evoluează interesul cumpărătorului și respectiv al Loteriei față de a cumpăra/vinde asemenea bilete în funcție de modificarea prețului acestuia.

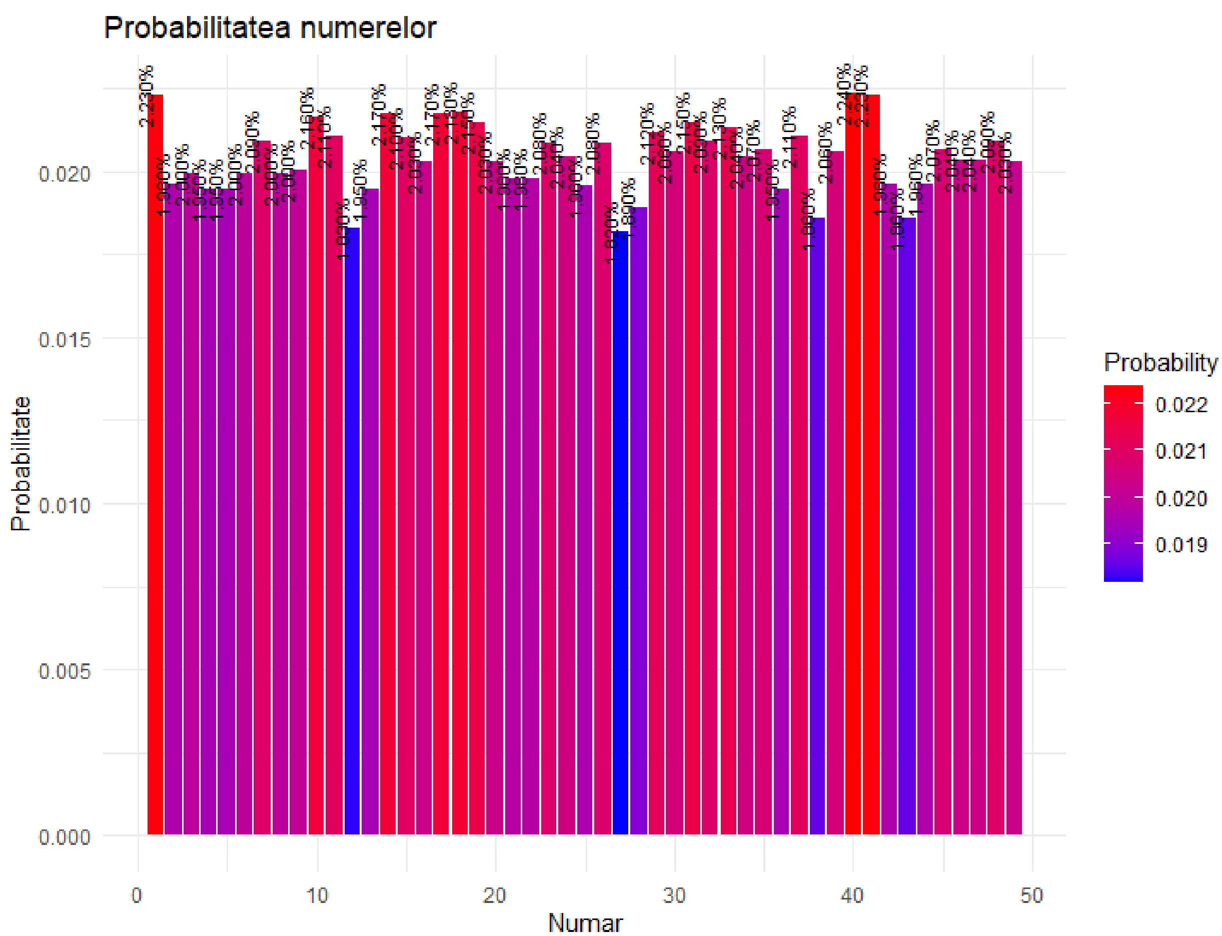
```
#Pentru instalare

#install.packages("ggplot2")
#install.packages("gganimate")
#install.packages("reshape2")

library(ggplot2)
library(gganimate)
library(reshape2)

#Stocam datele pentru preturi
df <- data.frame(
  pret = seq(10, 100, length.out = 1000)
)

# Calculam interesul cumparatorului folosind o
#functie exponentiala

df$cumparator_interes = 100 * exp(-df$pret / 200)


#Pretul la care interesul cumparatorului este de 70%
inflection_point = min(df$pret[df$cumparator_interes <= 70])

#Calculam interesul loteriei folosind
#o funcție logistică care tine cont de interesul cumparatorului
df$loterie_interes = ifelse(df$cumparator_interes < 70,
                            100 - (100 / (1 + exp(-0.05 * (df$pret - inflection_point)))),
                            100 / (1 + exp(-0.1 * (df$pret - inflection_point))))


#Procesarea afisarii
df_long <- melt(df, id.vars = "pret", variable.name = "group", value.name = "interes")

p <- ggplot(df_long, aes(x = pret, y = interes, color = group)) +
  geom_line() +
  labs(title = "pret vs interes", x = "pret", y = "interes") +
  theme_minimal() +
  transition_reveal(pret)

anim_save("animated_plot.gif", animate(p))
browseURL("animated_plot.gif")
```



4) Loteria pune la dispoziție istoricul extragerilor anterioare din ultimii 20 de ani. În ce manieră puteți folosi aceste informații pentru a crește șansele de câștig cu un bilet simplu? Justificați răspunsul și ilustrați-l cu grafice relevante.

```r
library(ggplot2)

set.seed(123)
n_weeks <- 20 * 52

#Generarea numerelor
numbers <- replicate(n_weeks * 2, sample(1:49, 6))

#Calcularea fecventei
frequency <- table(unlist(numbers))

#Calcularea probabilitatii
probability <- frequency / (n_weeks * 2 * 6)

data <- data.frame(Number = as.numeric(names(frequency)), Frequency = as.numeric(frequency), Probability = as.numeric(probability))

#Histograma frecventelor
ggplot(data, aes(x = Number, y = Frequency)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecventa numerelor", x = "Numar", y = "Frecventa")

#Histograma probabilitatiilor
# Create histogram of probabilities with more details
ggplot(data, aes(x = Number, y = Probability)) +
  geom_bar(stat = "identity", aes(fill = Probability)) +
  geom_text(aes(label = scales::percent(round(Probability, 4))), vjust = -0.3, hjust = 0.5, angle = 90, size = 3) +
  scale_fill_gradient(low = "blue", high = "red") +
  theme_minimal() +
  labs(title = "Probabilitatea numerelor", x = "Numar", y = "Probabilitate")


#Cele mai mari probabilitati
data_sorted <- data[order(-data$Probability), ]
top_numbers <- data_sorted$Number[1:6]
top_numbers
```

```
> top_numbers
[1] 40  1 41 18 14 17
```


Probabilitatea numerelor
Probabilitate
0.000
0.005
0.010
0.015
0.020
Numar
0
10
20
30
40
50
Probability
0.022
0.021
0.020
0.019