Чернівецький національний університет імені Юрія Федьковича

Інститут фізико-технічних та комп’ютерних наук

Кафедра Математичних проблем управління і кібернетики

Лабораторна робота № 3

“Тема: Оцінювання функції розподілу”

з дисципліни

“Інтелектуальний аналіз”

Варіант - 2

Виконав:

ст. гр. 341СК Гаваньо Дмитро

Прийняв:

професор Дрінь Я. М.

Чернівці – 2025

**Мета:** Ознайомитися з методами оцінювання функції розподілу 𝐹𝜉 (𝑥) випадкової величини, про яку відомо, що вона є неперервною. Дослідження вибірки значень випадкової величини із заданим законом розподілу.

**Завдання 1:** Побудуйте для вибірки, сформованої в завданні 1.1, 95%-й “коридор” для функції розподілу випадкової величини, яка досліджується.

1. Прочитайте файл, в якому знаходяться дані.
2. Визначте статистику Колмогорова – функцію K(z) і побудуйте її графік.
3. Задайте значення величини α. Розв'яжіть графічно рівняння 1-K(z)=α .
4. Побудуйте “коридор” для теоретичної функції розподілу



**Завдання 2:** Згенеруйте вибірку значень випадкової величини із вказаним неперервним розподілом і виконайте повний попередній аналіз для вказаних значень об'єму вибірки, числа інтервалів групування і надійної ймовірності. Побудуйте графіки щільності ймовірності і функції розподілу і порівняйте їх із одержаними графіками відповідних вибіркових функцій.

1. Надайте змінній n значення , яке дорівнює 100.
2. Побудуйте для заданого розподілу графіки щільності ймовірності і функції розподілу.
3. Знайдіть математичне сподівання, дисперсію, середнє квадратичне відхилення, медіану, моменти 3- і 4-го порядку, асиметрію і ексцес заданого розподілу.
4. Згенеруйте вибірку об'єму n значень випадкової величини, яка має заданий розподіл.
5. Запишіть як функції змінної n вибіркові значення середнього, середнього квадратичного відхилення, моментів 3- і 4-го порядку, асиметрії і ексцесу.
6. Побудуйте гістограму, полігон частот, графік накопичених відносних частот.
7. Побудуйте 95 %-й “коридор” для теоретичної функції розподілу і зобразіть на цьому ж графіку функцію заданого в умові розподілу ймовірностей.
8. Порівняйте обчислені теоретичні і вибіркові значення параметрів.
9. Виконайте обчислення пп. 4-7 для n=150,200,300,500

**Хід роботи:**

**Завдання 1:**

Програмний код:

# 1. Дані з геометричного розподілу

set.seed(123)

x <- rgeom(50, prob = 0.3)   # вибірка 50 елементів

# 2. Функція Колмогорова K(z)

K <- function(z, terms=100) {

  s <- 0

  for (k in 1:terms) {

    s <- s + (-1)^(k-1) \* exp(-2 \* k^2 \* z^2)

  }

  1 - 2 \* s

}

png("Graphic\\3.1lab\_Kolmogorov\_funcion.png", width = 800, height = 600)

# (b) Функція Колмогорова K(z)

z <- seq(0.02, 2, length=200)

plot(z, K(z), type="l", lwd=2, col="darkgreen",

     main="Функція Колмогорова K(z)",

     xlab="z", ylab="K(z)")

dev.off()

alpha <- 1 - K(z)

png("Graphic/3.1lab\_alpha\_equals\_Kz.png", width=800, height=600)

plot(z, alpha, type="l", lwd=2, col="blue",

     main="Alpha = 1 - K(z)",

     xlab="D (max |F\_n - F|)",

     ylab="Alpha")

grid()

abline(h=0.05, col="red", lty=2)  # наприклад рівень значущості 0.05

dev.off()

get\_Dn <- function(x, F\_theor) {

  # x - вектор вибірки

  F\_emp <- ecdf(x)

  z\_vals <- sort(unique(x))  # усі унікальні точки для оцінки

  D\_n <- max(abs(F\_emp(z\_vals) - F\_theor(z\_vals)))

  return(D\_n)

}

plot\_KS\_corridor\_norm <- function(file\_path, col\_name, mu=150, sigma=10, alpha=0.05) {

  # Завантажуємо дані

  data <- read.csv("d:\\Study\\R\\Lab3\\numbers\_list.csv", stringsAsFactors = FALSE)

  print(data)

  x <- data[[col\_name]]

  print(x)

  n <- length(x)

  # Емпірична CDF

  F\_emp <- ecdf(x)

  # Теоретична CDF для N(mu, sigma)

  F\_theor <- function(z) pnorm(z, mean=mu, sd=sigma)

  # Критичне значення D для 95% коридору (наближення)

  D\_n <- get\_Dn(x, F\_theor)

  # Малюємо графік

  max\_z <- max(x, mu + 4\*sigma)  # для охоплення всіх значень

  png("Graphic/3.1lab\_KS\_corridor\_normal.png", width=800, height=600)

  plot(F\_emp, main="ECDF + 95% KS Corridor (Normal)", xlab="z", ylab="F(z)", lwd=2, col="blue")

  curve(F\_theor(x), from=0, to=max\_z, add=TRUE, col="red", lwd=2)

  # 95% коридор

  curve(F\_theor(x) + D\_n, from=0, to=max\_z, add=TRUE, col="darkgreen", lty=2)

  curve(F\_theor(x) - D\_n, from=0, to=max\_z, add=TRUE, col="darkgreen", lty=2)

  legend("bottomright", legend=c("ECDF","Theoretical CDF","95% KS Corridor"),

         col=c("blue","red","darkgreen"), lwd=c(2,2,1), lty=c(1,1,2))

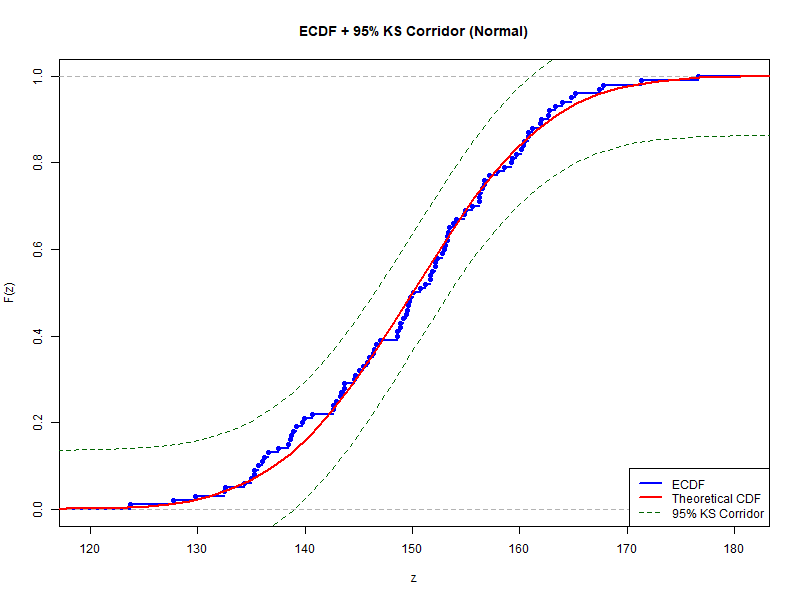
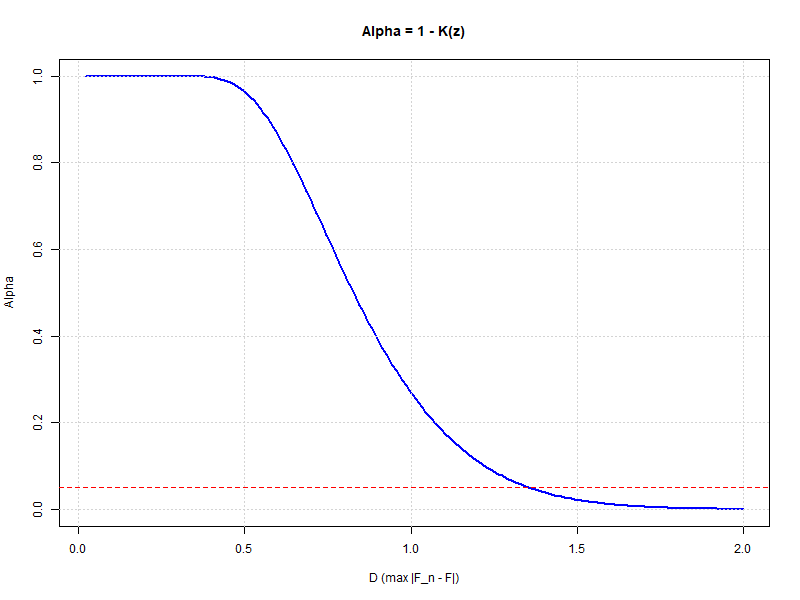
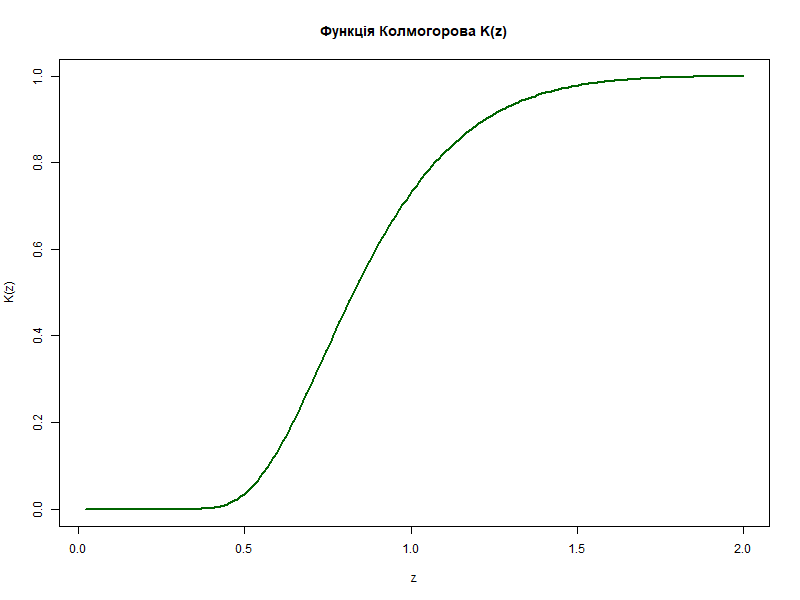
  dev.off()

}

plot\_KS\_corridor\_norm("numbers\_list.csv", "Value")

Результат виконання:

Критичне значення D для 95% коридору: 0.05204544



**Завдання 2:**

Програмний код:

setwd("D:\\Study\\R\\Lab3") # set working directory

seed = 124

# ------------- FUNCTIONS ----------------

# --- 1. Графік щільності ймовірності (PMF) ---

plot\_geom\_density <- function(prob=0.2, max\_k=20) {

  x <- 0:max\_k

  y <- dgeom(x, prob=prob)

  png("Graphic\\3.2lab\_geom\_PMF.png", width=800, height=600)

  plot(x, y, type="h", lwd=2, col="blue",

       main=paste("PMF (Geometric, p =", prob, ")"),

       xlab="k", ylab="P(X=k)")

  points(x, y, pch=19, col="red")

  dev.off()

}

# --- 2. Графік функції розподілу (CDF) ---

plot\_geom\_cdf <- function(prob=0.2, max\_k=20) {

  x <- 0:max\_k

  y <- pgeom(x, prob=prob)

  png("Graphic\\3.2lab\_geom\_CDF.png", width=800, height=600)

  plot(x, y, type="s", lwd=2, col="blue",

       main=paste("CDF (Geometric, p =", prob, ")"),

       xlab="k", ylab="F(k)")

       dev.off()

}

# Theoretical stats

theoretical\_stats\_geom <- function(p) {

  mu <- (1 - p) / p

  var <- (1 - p) / (p^2)

  sd <- sqrt(var)

  m3 <- (2 - p) \* (1 - p) / (p^3)   # central 3rd moment

  m4 <- (6 + p^2) \* (1 - p) / (p^4) # central 4th moment

  skew <- (2 - p) / sqrt(1 - p)

  excess <- (6 + p^2) / (1 - p)

  list(mean = mu, variance = var, sd = sd, median = qgeom(0.5, p),

       m3 = m3, m4 = m4, skewness = skew, excess = excess)

}

# geometric dist sample

generate\_geom\_sample <- function(N, prob=0.2) {

  set.seed(seed)

  sample <- rgeom(N, prob=prob)

  return(sample)

}

# Sample stats

sample\_moment\_central <- function(x, k) mean((x - mean(x))^k)

sample\_skewness <- function(x) sample\_moment\_central(x, 3) / sd(x)^3

sample\_excess <- function(x) sample\_moment\_central(x, 4) / sd(x)^4 - 3

sample\_stats <- function(x) {

  list(

    mean = mean(x),

    median = median(x),

    variance = var(x),

    sd = sd(x),

    m3 = sample\_moment\_central(x, 3),

    m4 = sample\_moment\_central(x, 4),

    skewness = sample\_skewness(x),

    excess = sample\_excess(x)

  )

}

# --- 1. Полігон частот ---

plot\_frequency\_polygon <- function(x, bins=20, id=50) {

  # Створюємо гістограму без малювання

  hist\_info <- hist(x, breaks=bins, plot=FALSE)

  # Середини бінів

  mids <- hist\_info$mids

  counts <- hist\_info$counts

  # Полігон частот

  png(paste0("Graphic\\3.2lab\_Frequency\_polygon\_",id,".png"), width = 800, height = 600)

  plot(mids, counts, type="b", lwd=2, col="blue",

       main="Frequency polygon",

       xlab="x", ylab="Frequency", pch=19)

  grid()

  dev.off()

}

# --- 2. Графік накопичених відносних частот ---

plot\_cumulative\_relative <- function(x, bins=20, id=50) {

  hist\_info <- hist(x, breaks=bins, plot=FALSE)

  mids <- hist\_info$mids

  rel\_counts <- hist\_info$counts / sum(hist\_info$counts)

  cum\_rel <- cumsum(rel\_counts)

  png(paste0("Graphic\\3.2lab\_Cumulative\_Relative\_Frequency\_",id,".png"), width = 800, height = 600)

  plot(mids, cum\_rel, type="s", lwd=2, col="darkgreen",

       main="Cumulative Relative Frequency",

       xlab="x", ylab="Cumulative Relative Frequency")

  grid()

  abline(h=seq(0,1,0.1), col="gray", lty=2)

  dev.off()

}

# Compare theoretical vs sample

plot\_geom\_CDF\_corridor <- function(x, p, id) {

  n <- length(x)

  x\_sorted <- sort(x)

  Kz <- ecdf(x)

  kgrid <- min(x\_sorted):max(x\_sorted)

  Fk <- pgeom(kgrid, prob = p)

  # 95% Kolmogorov corridor

  band\_radius <- 1.36 / sqrt(n)

  upper <- pmin(1, Fk + band\_radius)

  lower <- pmax(0, Fk - band\_radius)

  # Save plot to PNG

  png(paste0("Graphic\\3.2lab\_geom\_CDF\_corridor",id,".png"), width = 800, height = 600)

  # Plot theoretical CDF

  plot(kgrid, Fk, type = "s", lwd = 2, col = "blue",

       main = "Geometric CDF with 95% corridor",

       xlab = "k", ylab = "F(X ≤ k)")

  # Plot empirical CDF

  lines(Kz, verticals = TRUE, do.points = FALSE, col = "black", lwd = 2)

  # Plot corridor

  lines(kgrid, upper, col = "red", lty = 2)

  lines(kgrid, lower, col = "red", lty = 2)

  # Legend

  legend("bottomright",

         legend = c("Empirical CDF", "Theoretical CDF", "95% corridor"),

         col = c("black", "blue", "red"),

         lty = c(1,1,2),

         lwd = c(2,2,1),

         bty = "n")

  grid()

  dev.off()

}

# Analyze one sample

analyze\_one <- function(x, p, plot=FALSE, id=0) {

  kgrid <- min(x):max(x)

  stats <- sample\_stats(x)

  if (plot) {

    plot\_cumulative\_relative(x, id = id)

    plot\_frequency\_polygon(x, id = id)

  }

  stats

}

# ----------------------------------------

p <- 0.2

n <- 50

kgrid <- 0:50

alpha = 0.1

# 2

plot\_geom\_density(prob=p, max\_k=n)

plot\_geom\_cdf(prob=p, max\_k=n)

# 3

cat("=== Theoretical stats (Geometric, p=", p, ") ===\n", sep="")

pmf <- dgeom(kgrid, prob = p)

print(sample\_stats(pmf))

# 4

cat("\n=== Sample stats ===\n")

x <- generate\_geom\_sample(n)

# 5

res50 <- analyze\_one(x, p, id = n)

print(res50)

# 6

plot\_cumulative\_relative(x)

plot\_frequency\_polygon(x)

# 7

plot\_geom\_CDF\_corridor(x, p, id = n)

# 8.

cat("\n=== Comparison table ===\n")

theor <- theoretical\_stats\_geom(p)

table <- data.frame()

table <- rbind(table, data.frame(

  type = "sample",

  n = n,

  mean = res50$mean,

  sd = res50$sd,

  skew = res50$skewness,

  excess = res50$excess

))

table <- rbind(table, data.frame(

  type = "theor",

  n = n,

  mean = theor$mean,

  sd = theor$sd,

  skew = theor$skewness,

  excess = theor$excess

))

numeric\_cols <- sapply(table, is.numeric)

table\_rounded <- table

table\_rounded[, numeric\_cols] <- round(table\_rounded[, numeric\_cols], 4)

print(table\_rounded)

# 9

ns\_all <- c(50,150,200,300,500)

for (ni in ns\_all) {

  x <- generate\_geom\_sample(ni)

  res <- analyze\_one(x, p, plot= TRUE, id=ni)

  plot\_geom\_CDF\_corridor(x, p, id = ni)

}

Результат виконання:

=== Theoretical stats (Geometric, p=0.2) ===

$mean

[1] 4

$variance

[1] 20

$sd

[1] 4.472136

$median

[1] 3

$m3

[1] 180

$m4

[1] 3020

$skewness

[1] 2.012461

$excess

[1] 7.55

=== Sample stats ===

9 0 6 2 4 2 1 0 8 2 2 2 2 7 12 5 8 0 4 0 8 0 4 0 3 29 2 3 0 12 2 0 4 3 3 3 3 10 1 3 2 2 6 1 1 0 0 0 5 3

$mean

[1] 3.78

$median

[1] 2.5

$variance

[1] 23.23633

$sd

[1] 4.820407

$m3

[1] 341.5799

$m4

[1] 8392.555

$skewness

[1] 3.049588

$excess

[1] 12.54387

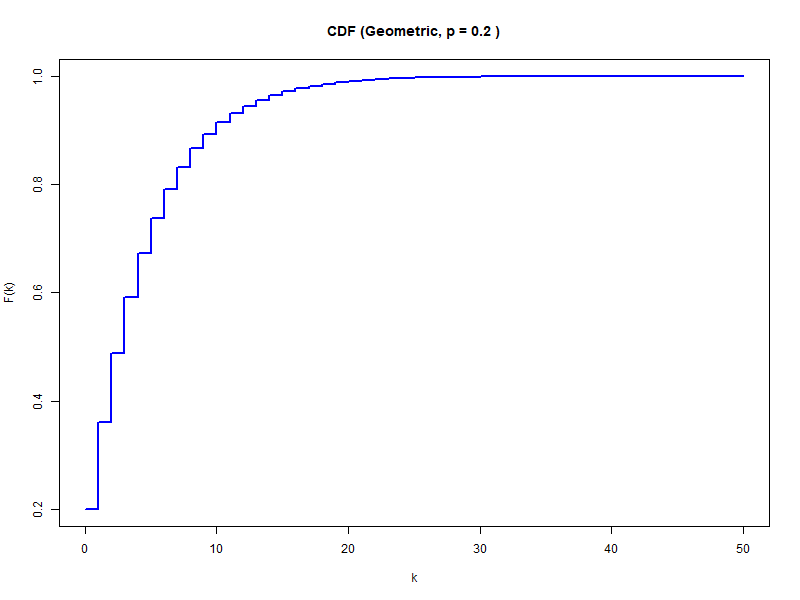
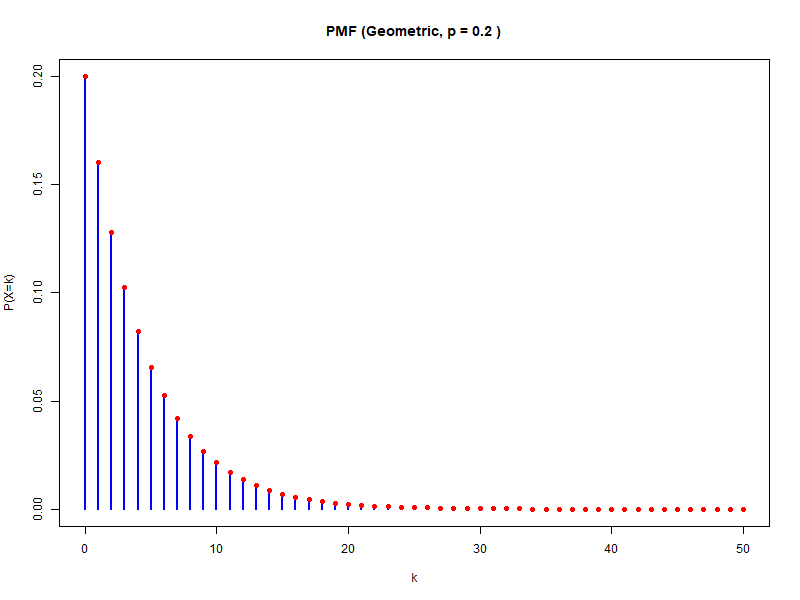
=== Comparison table ===

type n mean sd skew excess

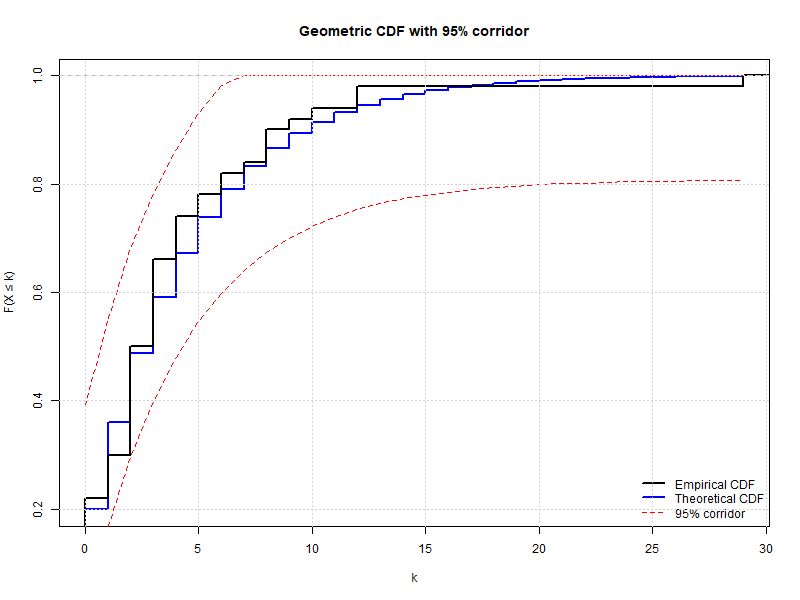
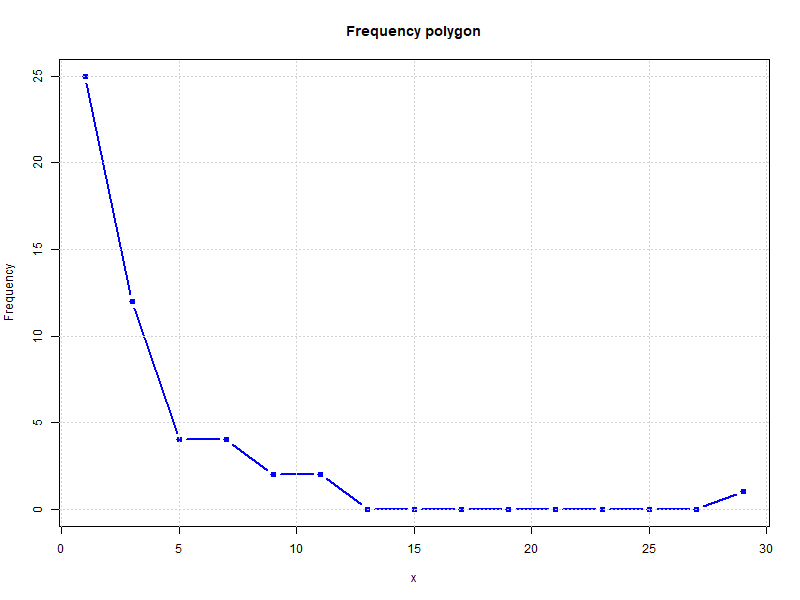
1 sample 50 3.78 4.8204 3.0496 12.5439

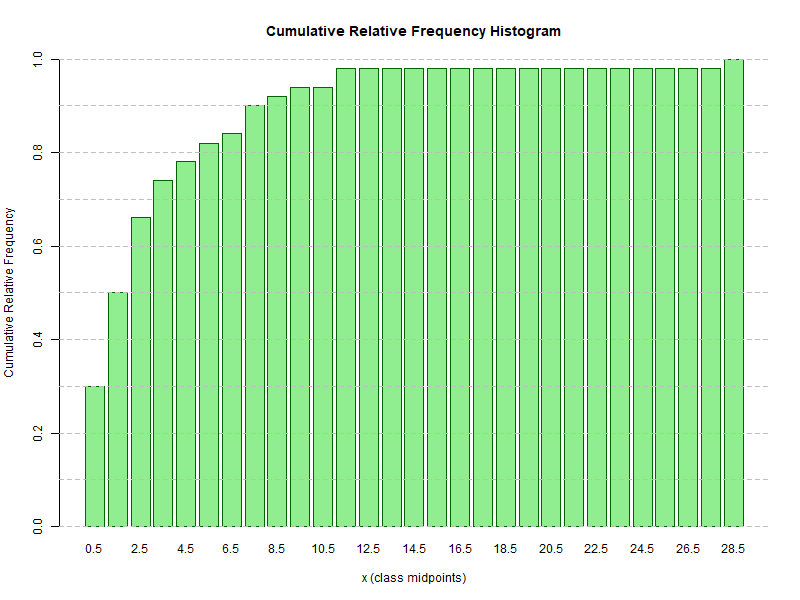
2 theor 50 4.00 4.4721 2.0125 7.5500

Теоретичні графіки для n = 50:

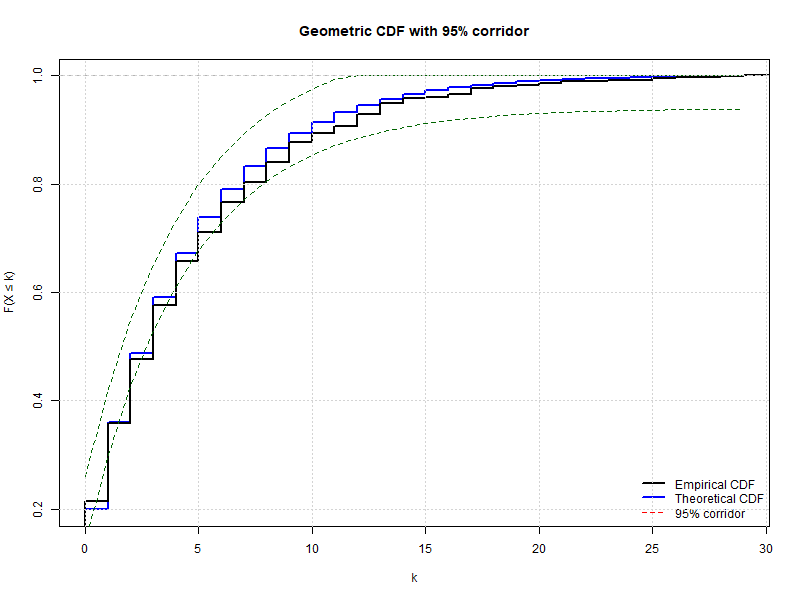
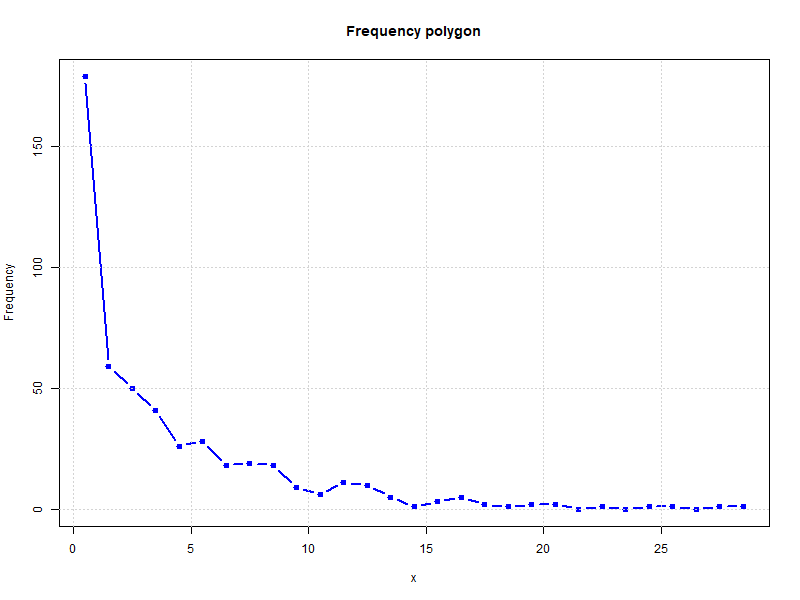
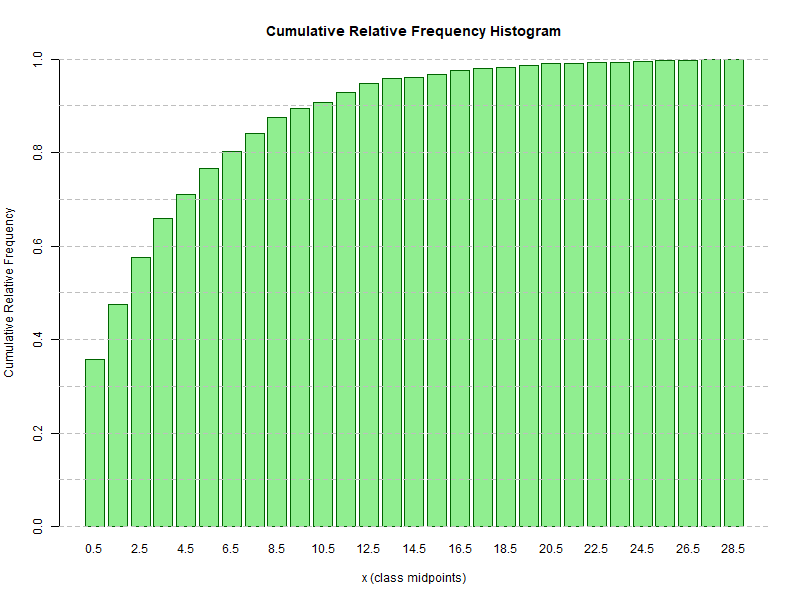


Графіки для n = 50:





Графіки для n = 500:



**Висновок:** я ознайомився з методами оцінювання функції розподілу 𝐹𝜉 (𝑥) випадкової величини, про яку відомо, що вона є неперервною. Дослідив вибірки значень випадкової величини із заданим законом розподілу.