

Modele liniowe - raport 1

Łukasz Rębisz

23.10.2022

Zadanie 1

Zgodnie z poleceniem mamy za zadanie wygenerować 100 wektorów losowych z rozkładu dwuwymiarowego normalnego $N(0, I)$, wykorzystując funkcję *rnorm*.

Funkcja *rnorm*(*n*) zwraca wektor *n* liczb pochodzących ze standardowego rozkładu normalnego.

Zauważmy, że w przypadku dwuwymiarowego rozkładu normalnego o macierzy kowariancji $\Sigma = I$ współrzędne rozkładu są niezależne i pochodzą ze standardowego, jednowymiarowego rozkładu normalnego.

Możemy zatem wylosować $2n$ wektorów z rozkładu $N(0, 1)$ i zapisać je w macierzy o wymiarach $n \times 2$. Wiersze macierzy będą wówczas wektorami losowymi z rozkładu dwuwymiarowego normalnego $N(0, I)$.

W celu uzyskania powtarzalności wyników losowania ustalmy ziarno generatora liczb pseudolosowych.

```
set.seed(1)
```

i) $n = 100$

Losujemy 100 wektorów z rozkładu dwuwymiarowego normalnego $N(0, I)$ zgodnie z opisaną powyżej procedurą.

```
m <- 2          # wymiar rozkładu
n <- 100        # liczba wektorów losowych
X_100 <- matrix(rnorm(n*m), nrow=n)
```

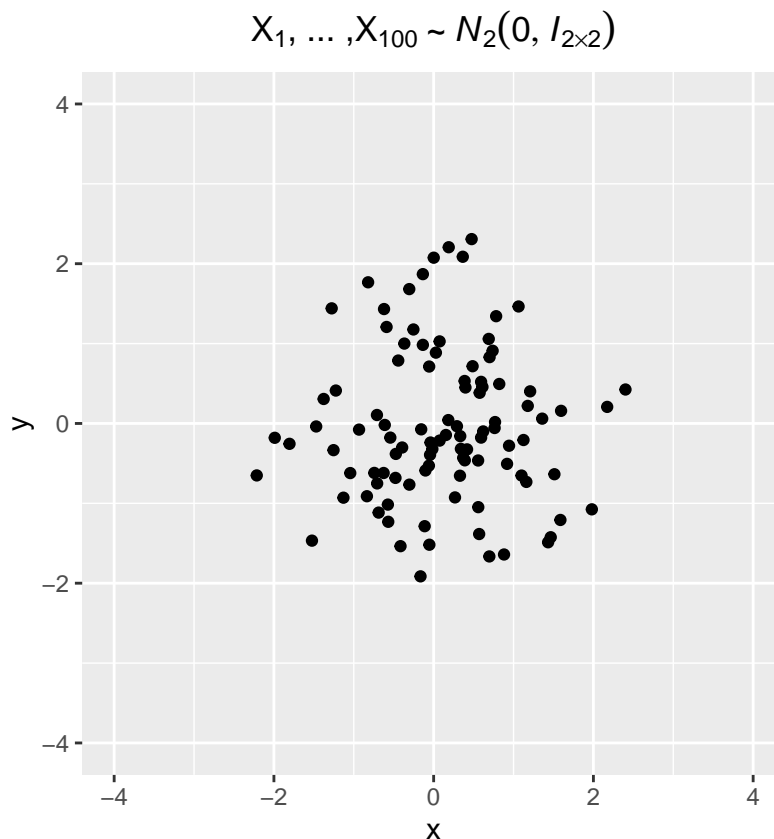
Zaznaczmy wylosowane wektory na płaszczyźnie.

```
library(ggplot2)
```

```
## Warning: pakiet 'ggplot2' został zbudowany w wersji R 4.1.3
```

```
X_100 <- as.data.frame(X_100)
colnames(X_100) <- c('x', 'y')
```

```
ggplot(X_100, aes(x, y)) +
  geom_point() +
  xlim(-4, 4) + ylim(-4, 4) +
  labs(title = expression(paste(X[1], " , ... ,", X[100] %~% italic(N[2])(0, italic(I[2]*%2))))),
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)
```



Na podstawie powyższego wykresu trudno stwierdzić, czy otrzymane punkty układają się w pewien określony kształt na płaszczyźnie.

Sprawdźmy, czy średnie wartości x i y odpowiadają w przybliżeniu wartościom oczekiwany (w tym przypadku równym zero).

Średnia próbkowa $\bar{x} = 0.1089$, natomiast $\bar{y} = -0.0378$.

W celu sprawdzenia, jak dokładnie zachowuje się badany rozkład, zwiększmy wielkość próby.

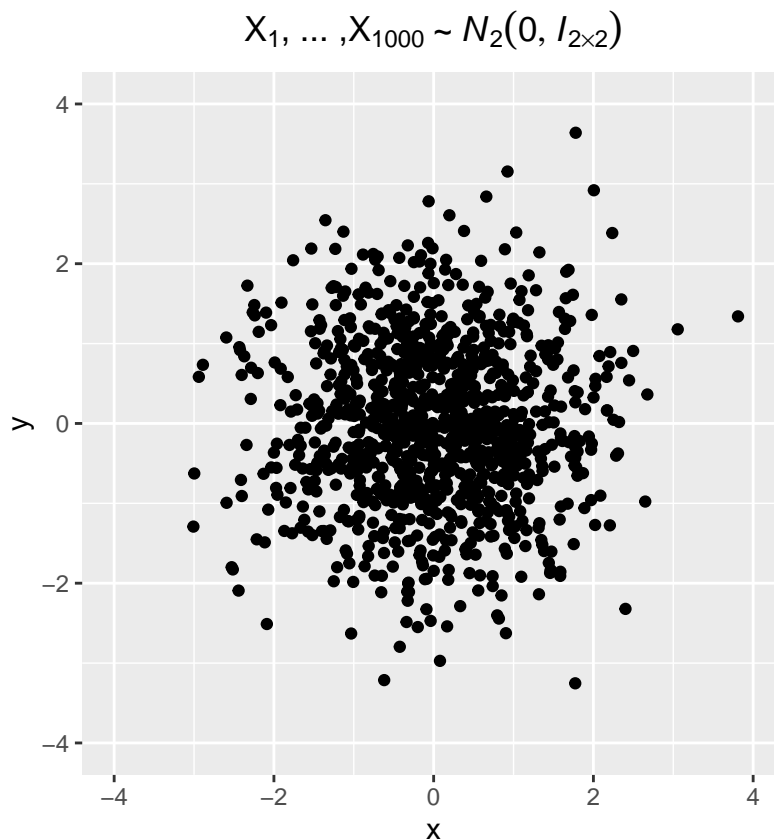
ii) $n = 1000$

W analogiczny sposób losujemy tym razem 1000 wektorów z dwuwymiarowego rozkładu normalnego $N(0, I)$ oraz zaznaczamy wylosowane wektory na płaszczyźnie.

```
n <- 1000      # liczba wektorów losowych
set.seed(1)
X_1000 <- matrix(rnorm(n*m), nrow=n)

X_1000 <- as.data.frame(X_1000)
colnames(X_1000) <- c('x', 'y')

ggplot(X_1000, aes(x, y)) +
  geom_point() +
  xlim(-4,4)+ylim(-4,4)+
  labs(title = expression(paste(X[1], " , ... ,", X[1000] %~% italic(N[2])(0, italic(I[2*%2]))))) +
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)
```



Na podstawie powyższego wykresu możemy stwierdzić, że wylosowane punkty koncentrują się wokół punktu $(0, 0)$. Trudno jednak stwierdzić, jaki kształt na płaszczyźnie ma badany rozkład.

Obliczmy średnie próbkowe dla współrzędnych: $\bar{x} = -0.0116$, $\bar{y} = -0.0163$. Wartości średnie są bliskie wartościom oczekiwanym (dla obu współrzędnych równym 0).

iii) $n = 10\ 000$

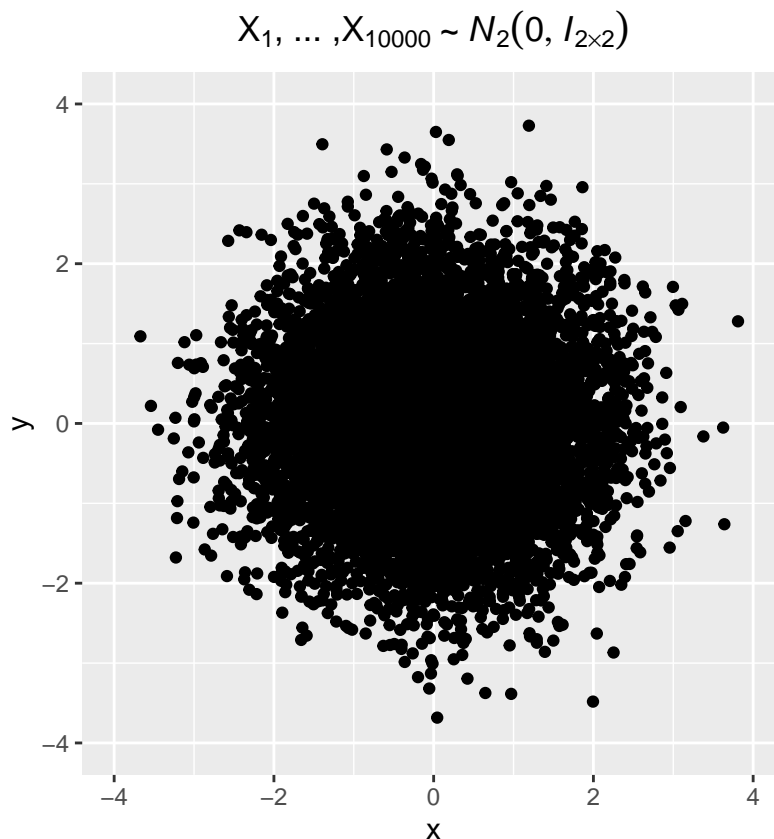
Wylosujmy 10 000 wektorów z badanego rozkładu i zaznaczmy otrzymane punkty na płaszczyźnie.

```
n <- 10000          # liczba wektorów losowych
set.seed(1)
X_10000 <- matrix(rnorm(n*m), nrow=n)

X_10000 <- as.data.frame(X_10000)
colnames(X_10000) <- c('x', 'y')

ggplot(X_10000, aes(x, y)) +
  geom_point() +
  xlim(-4,4)+ylim(-4,4)+
  labs(title = expression(paste(X[1], "...", X[10000] %~% italic(N[2])(0, italic(I[2*%2])))))
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



Analiza powyższego wykresu pozwala wysunąć wniosek, że wektory losowe z rozkładu dwuwymiarowego normalnego $N(0, I)$ koncentrują się wokół punktu $(0, 0)$ symetrycznie, tzn. prawdopodobieństwo wylosowania punktu płaszczyzny o danych współrzędnych maleje wraz ze wzrostem odległości punktu od początku układu współrzędnych. Z tego powodu otrzymana chmura punktów ma kształt kolisty.

Sprawdźmy, że rzeczywiście średnie próbkowe współrzędnych są bliskie zera: $\bar{x} = -0.0065$, $\bar{y} = -0.0042$.

Sprawdźmy, że próbkowe odchylenia standardowe są bliskie wartościom teoretycznym równym 1 dla obu współrzędnych: $s_x = 1.0249$, $s_y = 0.9816$.

Ogólnie, macierz kowariancji równa

```
##      x      y
## x 1.025 0.005
## y 0.005 0.982
```

jest w przybliżeniu równa macierzy $I_{2 \times 2}$.

Analiza powyższych statystyk potwierdza, że otrzymane wektory dobrze przybliżają badany rozkład $N_2(0, I_{2 \times 2})$.

Zadanie 2

Chcąc przekształcić otrzymane w poprzednim zadaniu chmury punktów pochodzące z rozkładu $N_2(0, I)$ w chmury punktów z rozkładu $N_2(\mu, \Sigma)$, skorzystamy z następującego faktu:

- Dla dowolnej ustalonej macierzy $A_{k \times p}$ i wektora $B \in \mathbb{R}^k$ definiujemy wektor losowy $Y = AX + B$. Zachodzi $\mu^Y = A\mu^X + B$ i $\Sigma^Y = A\Sigma^X A^T$.

Niech $X \sim N_2(0, I)$, $Y \sim N_2(\mu, \Sigma)$.

Wówczas $\mu = \mu^Y = A\mu^X + B = A \cdot \bar{0} + B = B$, czyli $B = \mu$.

Zgodnie z powyższym faktem $\Sigma = \Sigma^Y = A\Sigma^X A^T = A I A^T = A A^T$. Zatem w celu wyznaczenia macierzy przejścia A chcielibyśmy przedstawić macierz Σ w postaci iloczynu pewnej macierzy i jej transpozycji.

Zauważmy, że macierz Σ to macierz kowariancji zmiennych X i Y :

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{Y,X} & \sigma_Y^2 \end{pmatrix}$$

Powyższa macierz jest symetryczna, bo

$$\sigma_{X,Y} = \text{Cov}(X, Y) = E((X - EX)(Y - EY)) = E((Y - EY)(X - EX)) = \text{Cov}(Y, X) = \sigma_{Y,X}.$$

Macierz Σ jest także dodatnio określona (w przypadku niezdegenerowanych wektorów losowych X i Y). Z kryterium Sylwestera:

$$M_1 = \sigma_X^2 > 0,$$

$$M_2 = \det(\Sigma) = \sigma_X^2 \cdot \sigma_Y^2 - \sigma_{X,Y}^2 > 0.$$

Skoro macierz kowariancji Σ jest dodatnio określona i symetryczna, to można ją przedstawić w postaci iloczynu macierzy dolnotrójkątnej A i jej transpozycji (macierz spełnia założenia rozkładu Choleskiego).

Rozkład Choleskiego macierzy Σ :

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{Y,X} & \sigma_Y^2 \end{pmatrix} = A A^T = \begin{pmatrix} a_{1,1} & 0 \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{2,1} \\ 0 & a_{2,2} \end{pmatrix} = \begin{pmatrix} a_{1,1}^2 & a_{1,1} \cdot a_{2,1} \\ a_{1,1} \cdot a_{2,1} & a_{2,1}^2 + a_{2,2}^2 \end{pmatrix}.$$

Stąd:

$$\sigma_X^2 = a_{1,1}^2,$$

$$\sigma_{X,Y} = a_{1,1} \cdot a_{2,1},$$

$$\sigma_Y^2 = a_{2,1}^2 + a_{2,2}^2.$$

Zatem:

$$a_{1,1} = \sqrt{\sigma_X^2},$$

$$a_{2,1} = \frac{\sigma_{X,Y}}{a_{1,1}} = \frac{\sigma_{X,Y}}{\sqrt{\sigma_X^2}},$$

$$a_{2,2} = \sqrt{\sigma_Y^2 - a_{2,1}^2} = \sqrt{\sigma_Y^2 - \frac{\sigma_{X,Y}^2}{\sigma_X^2}}.$$

Wyprowadziwszy wzory na macierze A i B pozwalające na przejście ze zmiennej $X \sim N_2(0, I)$ do $Y \sim N_2(\mu, \Sigma)$, zaimplementujmy rozwiązanie i narysujmy wykresy otrzymanych punktów.

```
vectors_I_to_sigma <- function(n, mu, sigma){
  # n - liczba wektorów losowych
  # mu - wartość oczekiwana rozkładu Y
  # sigma - macierz kowariancji rozkładu Y

  # Y = A*X+B

  B <- matrix(mu, ncol = 2, nrow = n, byrow=TRUE)
  A <- matrix(c(sqrt(sigma[1,1]), 0, sigma[2,1]/sqrt(sigma[1,1]), sqrt(sigma[2,2]-(sigma[2,1]^2/sigma[1,1]))), nrow = 2, ncol = n, byrow=TRUE)
```

```

set.seed(1)
X <- matrix(rnorm(2*n), nrow=n)

Y <- matrix(nrow=n, ncol = 2)
Y <- X %*%t(A) + B

return(Y)
}

```

i)

Zmienna Y ma rozkład $N_2(\mu, \Sigma)$, gdzie

$$\mu = (4, 2),$$

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

Wygenerujmy 10 000 wektorów z rozkłady Y i zaznaczmy je na płaszczyźnie.

```

sigma09 <- matrix(c(1, 0.9, 0.9, 1), byrow = TRUE, nrow = 2)
mu1 <- c(4,2)

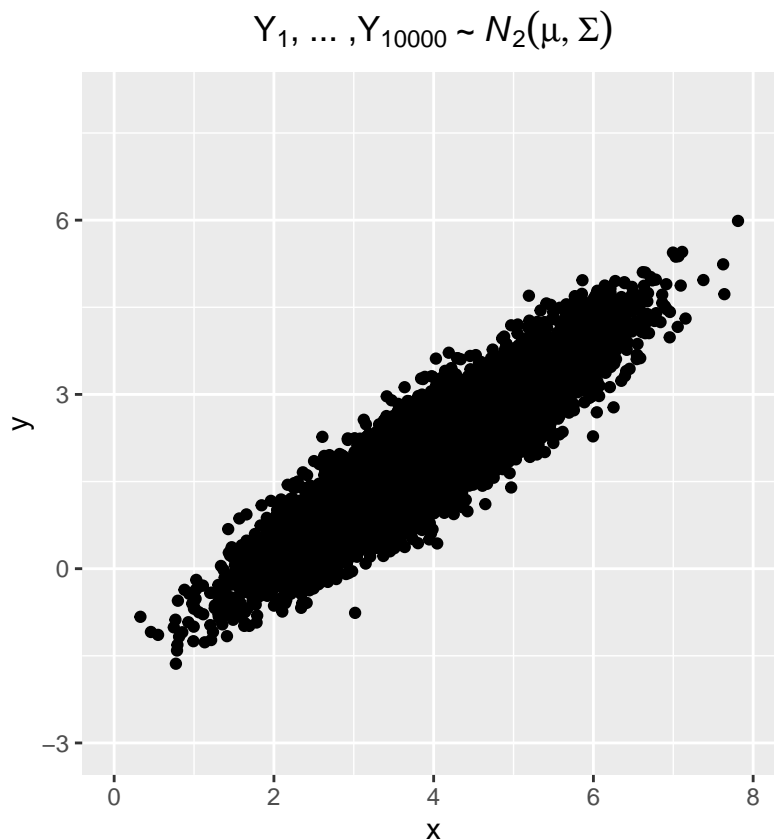
Y09 <- vectors_I_to_sigma(n=10000, mu1, sigma = sigma09)

Y09 <- as.data.frame(Y09)

colnames(Y09) <- c('x','y')

ggplot(Y09, aes(x, y)) +
  geom_point() +
  xlim(0,8)+ylim(-3,8)+
  labs(title = expression(paste(Y[1], "...", Y[10000] %~% italic(N[2])(mu, italic(Sigma))))) +
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)

```



Analiza wykresu pozwala stwierdzić, że punkty badanego rozkładu koncentrują się wokół punktu (4,2). Chmura punktów ma w przybliżeniu kształt eliptyczny.

Zauważmy, że średnie próbkowe współrzędnych są bliskie wartości oczekiwanej równej (4, 2):

$$\bar{x} = 3.993463 \approx 4,$$

$$\bar{y} = 1.9922903 \approx 2.$$

Zauważmy, że próbkowe kowariancje są bliskie wartościom teoretycznym. Wartości teoretyczne i wyestymowane różnią się o:

```
round(abs(cov(Y09)-sigma09),4)
```

```
##          x          y
## x 0.0249 0.0245
## y 0.0245 0.0205
```

Analiza powyższych statystyk potwierdza, że otrzymane wektory dobrze przybliżają badany rozkład $N_2(\mu, \Sigma_{2 \times 2})$, gdzie:

$$\mu = (4, 2),$$

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

ii)

W tym przypadku Y ma rozkład $N_2(\mu, \Sigma)$, gdzie

$$\mu = (4, 2),$$

$$\Sigma = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}.$$

Wygenerujmy 10 000 wektorów z rozkłady Y i zaznaczmy je na płaszczyźnie.

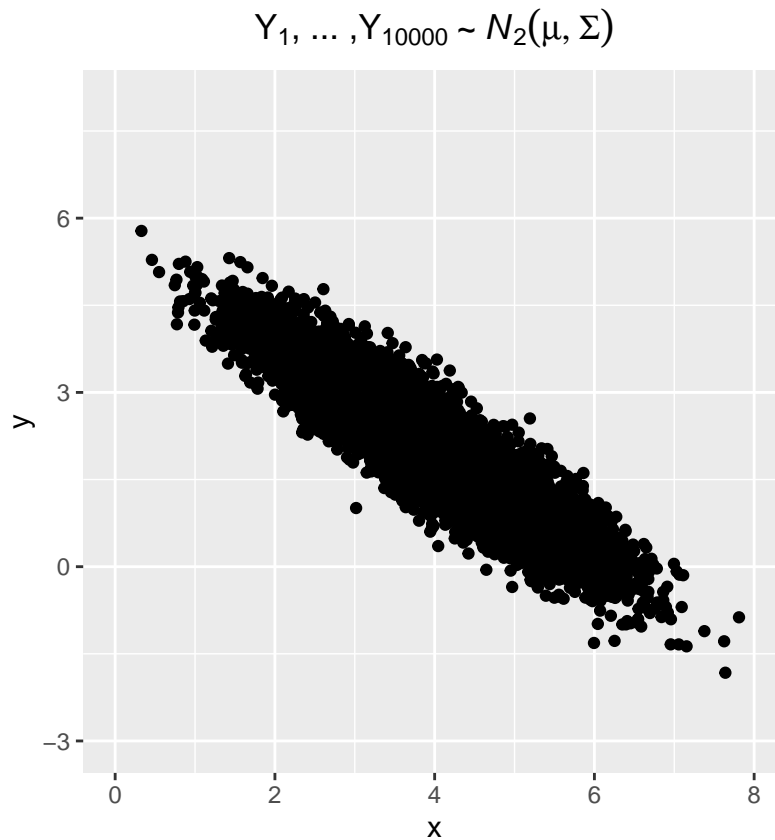
```
sigma_09 = matrix(c(1, -0.9, -0.9, 1), byrow = TRUE, nrow = 2)

Y_09 <- vectors_I_to_sigma(n=10000, mu1, sigma = sigma_09)

Y_09 <- as.data.frame(Y_09)

colnames(Y_09) <- c('x', 'y')

ggplot(Y_09, aes(x, y)) +
  geom_point() +
  xlim(0,8)+ylim(-3,8)+
  labs(title = expression(paste(Y[1], "...", Y[10000] %~% italic(N[2])(mu, italic(Sigma))))) +
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)
```



Analiza wykresu pozwala stwierdzić, że otrzymane punkty koncentrują się wokół punktu (4,2). Kształt wykresu jest w przybliżeniu eliptyczny i symetryczny względem otrzymanej w poprzednim podpunkcie chmury punktów.

Zauważmy, że średnie próbkowe współrzędnych są bliskie wartości oczekiwanej równej (4, 2):

$$\bar{x} = 3.993463 \approx 4,$$

$$\bar{y} = 2.0040569 \approx 2.$$

Zauważmy, że próbkowe kowariancje są bliskie wartościom teoretycznym. Wartości teoretyczne i wyestymowane różnią się o:

```
round(abs(cov(Y_09)-sigma_09),4)
```

```
##           x           y
## x 0.0249 0.0203
## y 0.0203 0.0128
```

Analiza powyższych statystyk potwierdza, że otrzymane wektory dobrze przybliżają badany rozkład $N_2(\mu, \Sigma_{2 \times 2})$, gdzie:

$$\mu = (4, 2),$$

$$\Sigma = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}.$$

iii)

W tym przypadku Y ma rozkład $N_2(\mu, \Sigma)$, gdzie

$$\mu = (4, 2),$$

$$\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}.$$

Wygenerujmy 10 000 wektorów z rozkładu Y i zaznaczmy je na płaszczyźnie.

```
sigma0 = matrix(c(9, 0, 0, 1), byrow = TRUE, nrow = 2)

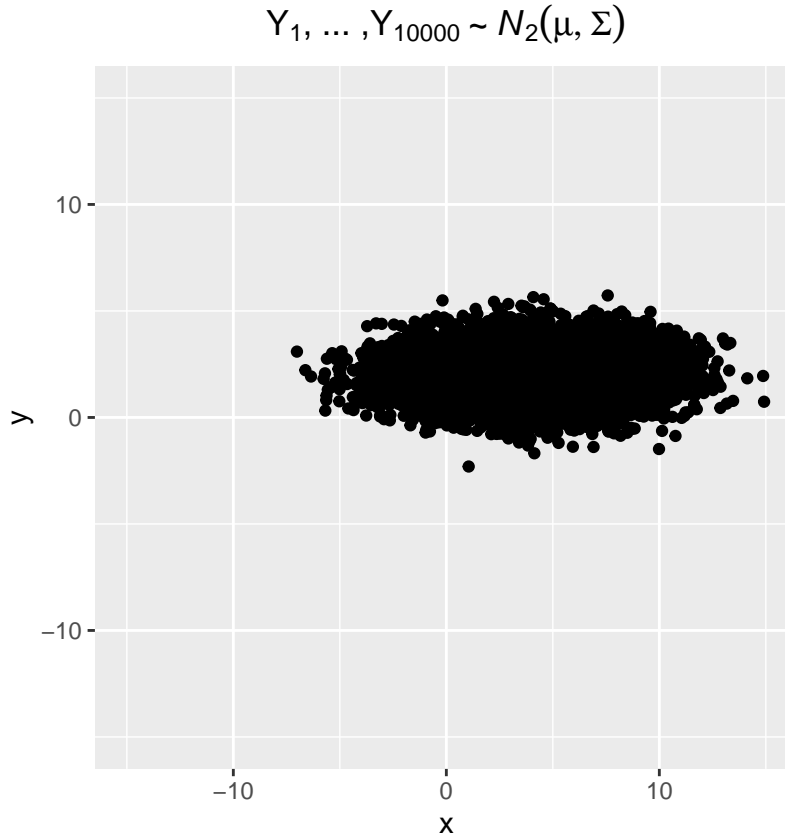
Y0 <- vectors_I_to_sigma(n=10000, mu1, sigma = sigma0)

Y0 <- as.data.frame(Y0)

colnames(Y0) <- c('x', 'y')

ggplot(Y0, aes(x, y)) +
  geom_point() +
  xlim(-15,15) + ylim(-15,15)+
  labs(title = expression(paste(Y[1], "...", Y[10000] %~% italic(N[2])(mu, italic(Sigma))))) +
  theme(plot.title = element_text(hjust = 0.5), aspect.ratio=1)

## Warning: Removed 1 rows containing missing values (geom_point).
```



Analiza wykresu pozwala stwierdzić, że otrzymana chmura punktów koncentruje się wokół punktu (4,2) i ma kształt eliptyczny.

Zauważmy, że średnie próbkowe współrzędnych są bliskie wartości oczekiwanej równej (4, 2):

$$\bar{x} = 3.9803889 \approx 4,$$

$$\bar{y} = 1.9958099 \approx 2.$$

Zauważmy, że próbkowe kowariancje są bliskie wartościom teoretycznym. Wartości teoretyczne i wyestymowane różnią się o:

```
round(abs(cov(Y0)-sigma0),4)
```

```
##           x           y
## x 0.2238 0.0146
## y 0.0146 0.0184
```

Analiza powyższych statystyk potwierdza, że otrzymane wektory dobrze przybliżają badany rozkład $N_2(0, \Sigma_{2 \times 2})$, gdzie:

$$\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}.$$

Zadanie 3

Naszym zadaniem jest wygenerowanie n wektorów losowych z rozkładu $N_{100}(0, I_{100 \times 100})$, a następnie przekształcenie otrzymanych wektorów w wektory z rozkładu $N_{100}(0, \Sigma_{100 \times 100})$, gdzie $\Sigma(i, i) = 1$ i $\Sigma(i, j) = 0.9$ dla $i \neq j$.

Zauważmy, że w przypadku rozkładu normalnego o macierzy kowariancji $\Sigma = I$ współrzędne rozkładu są niezależne i pochodzą ze standardowego, jednowymiarowego rozkładu normalnego.

Możemy zatem wylosować $100n$ wektorów z rozkładu $N(0,1)$ i zapisać je w macierzy o wymiarach $n \times 100$. Wiersze macierzy będą wówczas wektorami losowymi z rozkładu normalnego $N_{100}(0, I_{100 \times 100})$.

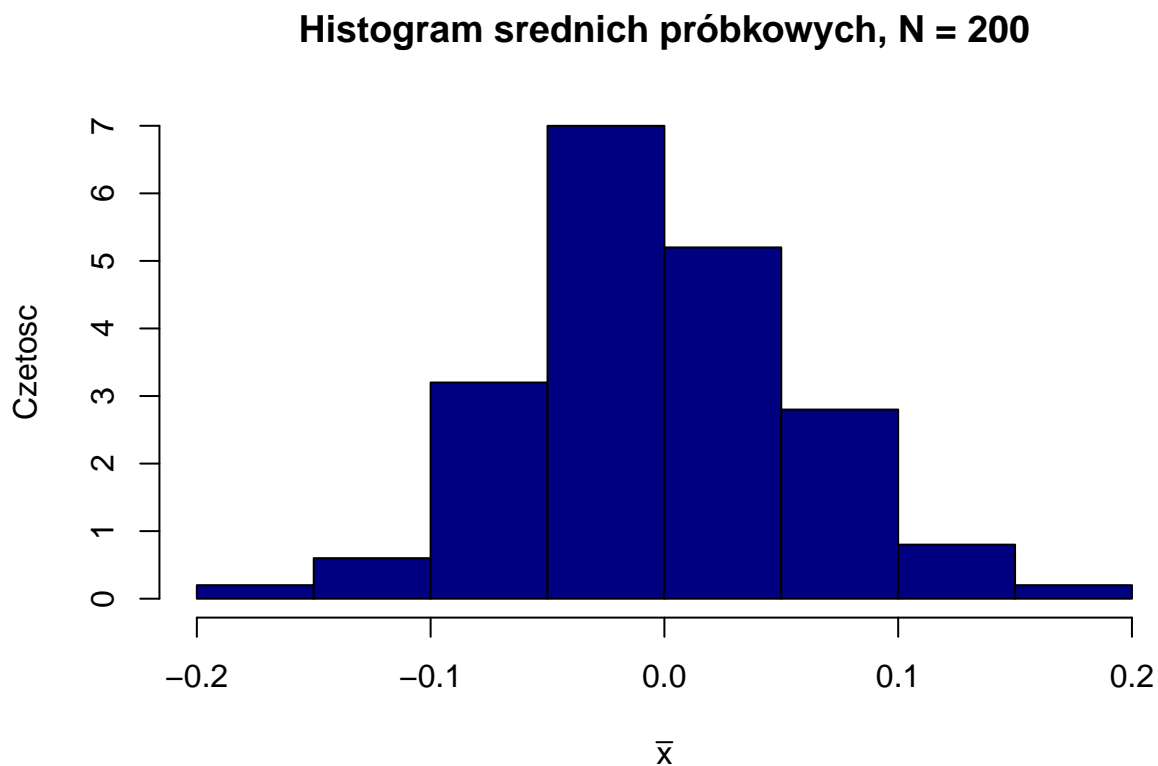
Wylosujemy w ten sposób $n=200$ wektorów z rozkładu $N_{100}(0, I_{100 \times 100})$.

```
k <- 100 # wymiar przestrzeni
N <- 200
set.seed(1)
X <- matrix(rnorm(k*N), nrow = N, ncol = k)
```

Sprawdźmy, czy otrzymane punkty dobrze estymują rozkład $N_{100}(0, I_{100 \times 100})$.

Porównajmy średnią próbkową poszczególnych współrzędnych z wartością oczekiwaną równą zero (dla wszystkich współrzędnych), rysując histogram średnich próbkowych:

```
hist(apply(X,2,mean), freq = FALSE, col = 'navyblue', main = "Histogram średnich próbkowych, N = 200", ylab = "Czestosc", xlab = "x̄")
```

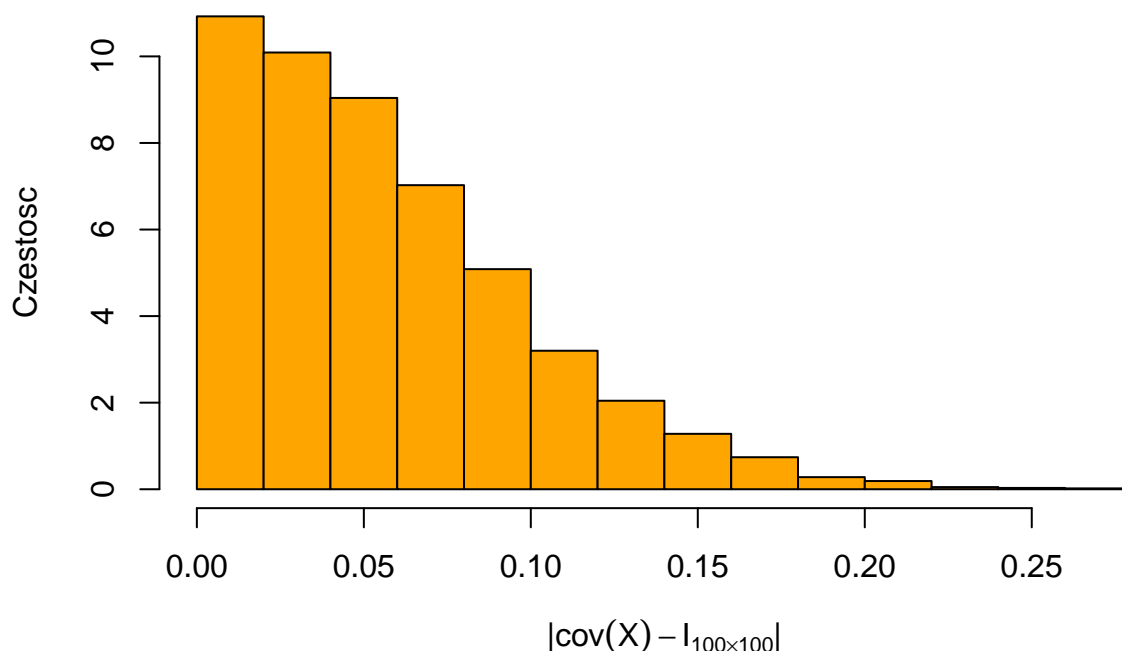


Powyższy histogram wskazuje na dobre przybliżenie wartości oczekiwanej równej zero.

Porównajmy także próbkowe kowariancje dla poszczególnych współrzędnych z wartościami teoretycznymi, rysując histogram różnic pomiędzy próbkowymi a teoretycznymi kowariancjami.

```
I <- diag(1, nrow=100)
hist(abs(cov(X)-I), freq=FALSE, col = 'orange', main = "Histogram różnic dla kowariancji, N = 200", ylab = "Czestosc", xlab = "różnica")
```

Histogram różnic dla kowariancji, N = 200



Powyższy histogram wskazuje na dobre przybliżenie próbkowych kowariancji (przeważają niewielkie różnice).

Analiza powyższych histogramów pozwala stwierdzić, że otrzymane punkty dobrze przybliżają badany rozkład $N_{100}(0, I_{100 \times 100})$.

Chcąc przekształcić otrzymane wektory w wektory z rozkładu $N_{100}(0, \Sigma_{100 \times 100})$, możemy analogicznie do przypadku dwuwymiarowego skorzystać z faktu, że macierz kowariancji Σ jest symetryczna i dodatnio określona. Dzięki temu możemy dokonać rozkładu Choleskiego macierzy Σ .

Rozkład Choleskiego macierzy Σ :

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_{n,n} \end{pmatrix} = AA^T = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ 0 & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{pmatrix}$$

Otrzymujemy zatem następujący układ równań:

$$\begin{cases} \sigma_{1,1} = a_{1,1}^2, \\ \sigma_{2,1} = a_{1,1} \cdot a_{2,1}, \\ \sigma_{2,2} = a_{2,2}^2, \\ \sigma_{3,2} = a_{3,1} \cdot a_{2,1} + a_{3,2} \cdot a_{2,2}, \\ \vdots \end{cases}$$

Analiza powyższego układu równań pozwala wyprowadzić ogólny wzór:

$$\sigma_{i,j} = \sum_{k=1}^{\min(i,j)} a_{i,k} \cdot a_{j,k} .$$

Z powyższego układu równań wyznaczamy wartości $a_{i,j}$:

$$\begin{cases} a_{1,1} = \sqrt{\sigma_{1,1}}, \\ a_{2,1} = \frac{\sigma_{2,1}}{a_{1,1}}, \\ a_{2,2} = \sqrt{\sigma_{2,2} - a_{2,1}^2}, \\ a_{3,2} = \frac{\sigma_{3,2} - a_{3,1} \cdot a_{2,1}}{a_{2,2}}, \\ \vdots \end{cases}$$

Analiza powyższego układu równań pozwala wyprowadzić ogólny wzór:

$$a_{i,i} = \sqrt{\sigma_{i,i} - \sum_{k=1}^{i-1} a_{i,k}^2} .$$

$$a_{j,i} = \frac{\sigma_{j,i} - \sum_{k=1}^{i-1} a_{j,k} a_{i,k}}{a_{i,i}}$$

Zaimplementujmy funkcję rozkładu Choleskiego, obliczając wartości macierzy dolnotrójkątnej wiersz po wierszu zgodnie z powyższymi wzorami:

```
Cholesky <- function(X){
  n <- dim(X)[1]
  A <- matrix(0, nrow = n, ncol =n)

  for(i in 1:n){
    for(j in 1:i){
      sum <- 0
      for(k in 1:(j-1)){
        if(j > 1) sum <- sum + A[i,k] * A[j,k]

        if(i==j) {A[i,j] <- sqrt(X[i,i] - sum)} else{
          A[i,j] <- (1/A[j,j])*(X[i,j]-sum)}
        }
      }
    }
  }
  return(A)
}
```

Wektory losowe z rozkładu $X \sim N_{100}(0, I_{100 \times 100})$ chcemy przekształcić w wektory z rozkładu $\tilde{X} \sim N_{100}(0, \Sigma_{100 \times 100})$, gdzie $\Sigma(i, i) = 1$ i $\Sigma(i, j) = 0.9$ dla $i \neq j$.

Wiemy (analogicznie do sytuacji dwuwymiarowej), że

$$Y := \tilde{X} = AX + B,$$

$$0 = \mu^{\tilde{X}} = A\mu^X + B = A \cdot \bar{0} + B = B,$$

$$\Sigma = AIA^T = AA^T.$$

Implementacja wykorzystująca napisaną powyżej funkcję $Cholesky(X)$:

```
N <- 200
k <- 100

set.seed(1)
X <- matrix(rnorm(k*N), nrow = N, ncol = k)

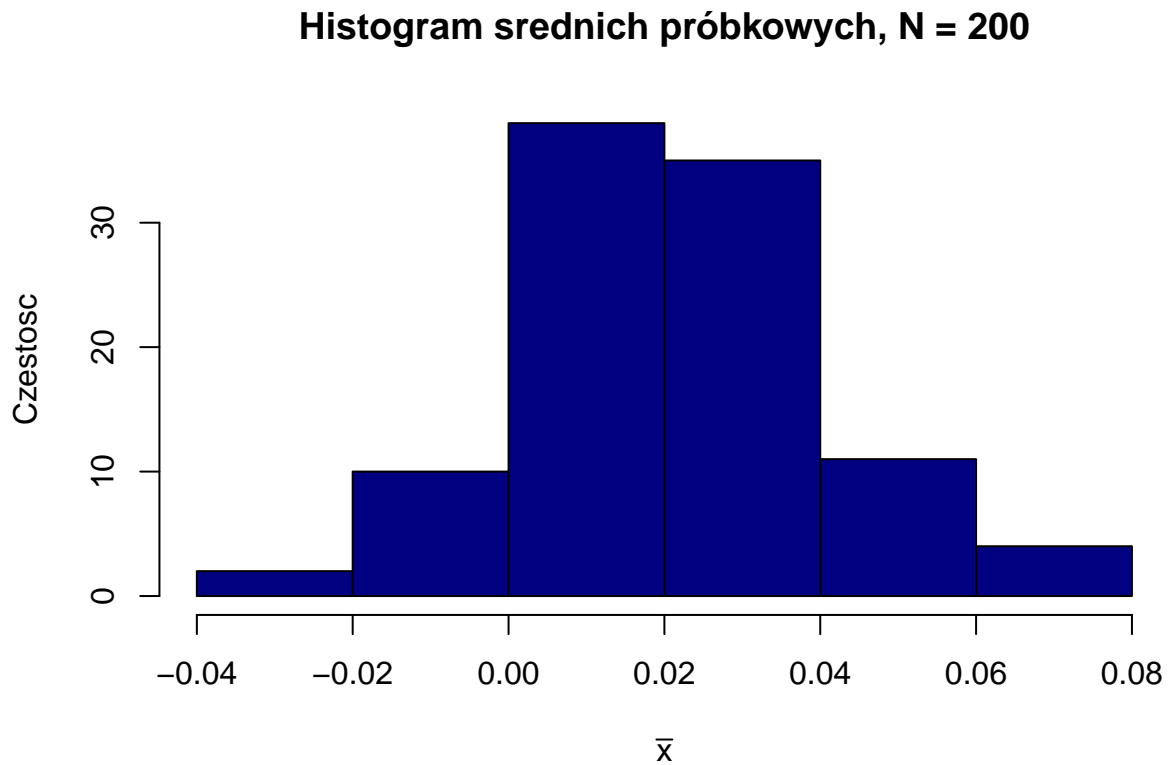
sigma <- matrix(0.9, nrow = k, ncol = k) + diag(0.1, k)

A <- Cholesky(sigma)

Y <- X %*% t(A)
```

Sprawdźmy, czy otrzymane wektory dobrze przybliżają rozkład $N_{100}(0, \Sigma_{100 \times 100})$.

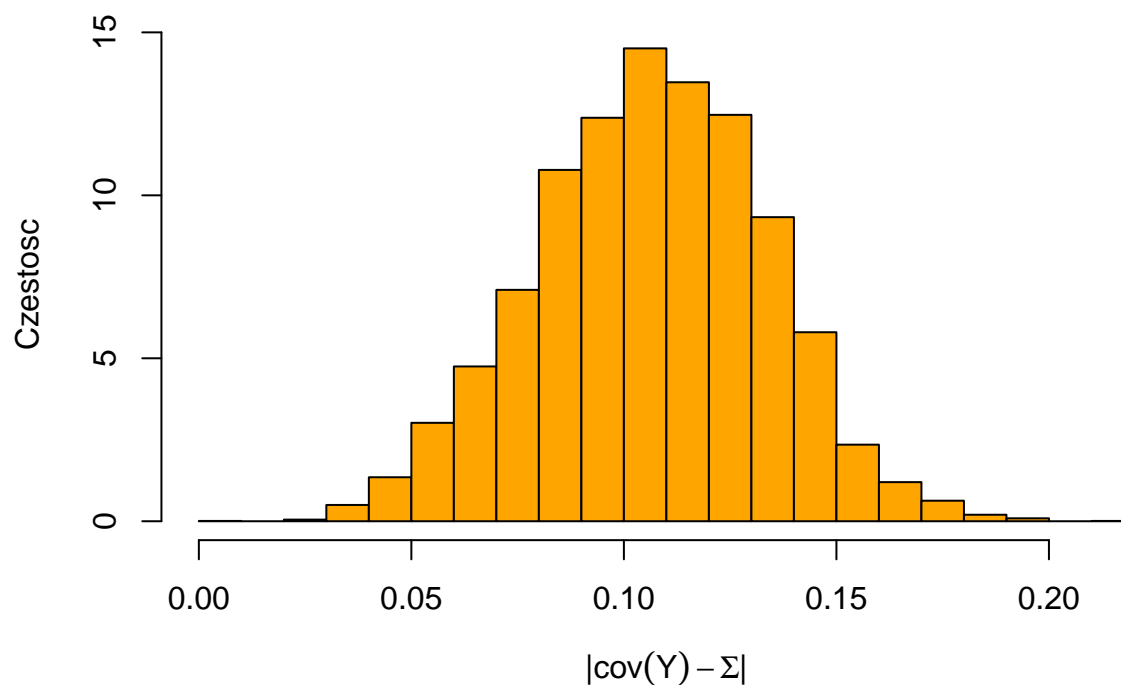
Narysujmy histogram średnich próbkowych dla poszczególnych współrzędnych.



Na podstawie powyższego histogramu możemy stwierdzić, że otrzymane punkty dobrze przybliżają badany rozkład (o wartości oczekiwanej równej zero).

Porównajmy także próbkowe kowariancje dla poszczególnych współrzędnych z wartościami teoretycznymi, rysując histogram różnic pomiędzy próbkowymi a teoretycznymi kowariancjami.

Histogram różnic dla kowariancji, N = 200



Powyższy histogram wskazuje na małe różnice pomiędzy wystymowanymi a teoretycznymi wartościami kowariancji (dominująca różnica wynosi ok. 0.107).

Analiza powyższych histogramów pozwala stwierdzić, że otrzymane punkty dobrze przybliżają badany rozkład $N_{100}(0, \Sigma_{100 \times 100})$.

Zwiększmy rozmiar próby do 10 000 i powtórzmy losowanie.

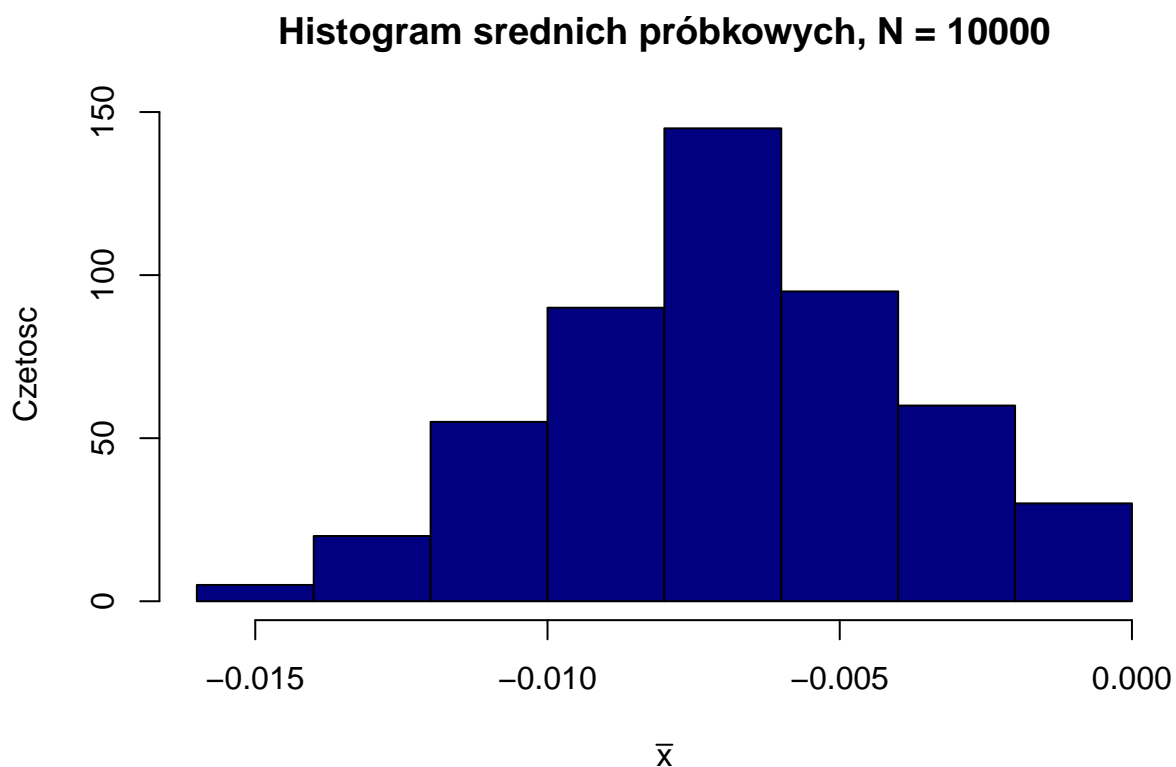
```
N <- 10000
k <- 100

set.seed(1)
X <- matrix(rnorm(k*N), nrow = N, ncol = k)

sigma <- matrix(0.9, nrow = k, ncol = k) + diag(0.1, k)

A <- Cholesky(sigma)

Y <- X %*% t(A)
```

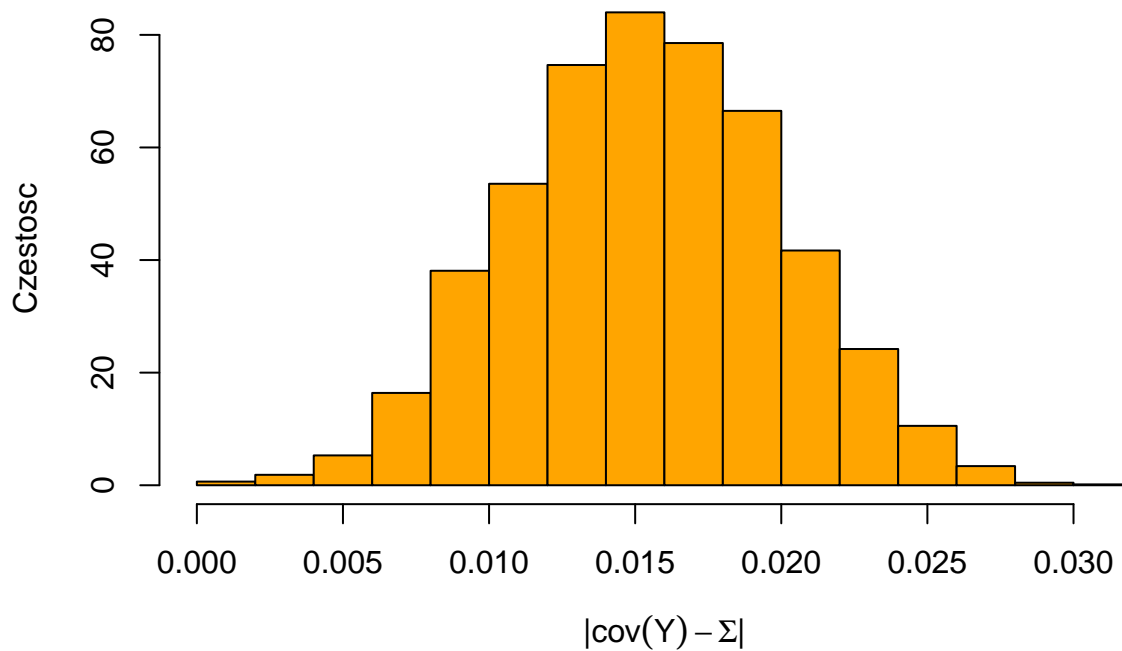


Powyższy histogram wskazuje, że średnie próbkowe mają wartości w przybliżeniu równe zero.

Zatem otrzymane punkty dobrze przybliżają badany rozkład (o wartości oczekiwanej równej zero).

Porównajmy także próbkowe kowariancje dla poszczególnych współrzędnych z wartościami teoretycznymi, rysując histogram różnic pomiędzy próbkowymi a teoretycznymi kowariancjami.

Histogram różnic dla kowariancji, N = 10000



Powyższy histogram wskazuje na małe różnice pomiędzy wystymowanymi a teoretycznymi wartościami kowariancji (dominująca różnica wynosi ok. 0.015).

Analiza powyższych histogramów pozwala stwierdzić, że otrzymane punkty dobrze przybliżają badany rozkład $N_{100}(0, \Sigma_{100 \times 100})$.