Bài 5.

Phân phối mẫu và định lý giới hạn trung tâm

5.1. Mẫu ngẫu nhiên phân phối chuẩn

Mô phỏng

Định lý. *Cho* $X_1, X_2, \cdots, X_n$ *là mẫu ngẫu nhiên lấy từ phân phối chuẩn* $N(\mu;\sigma^2)$. *Ta có* $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$ *và* $\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1)$.

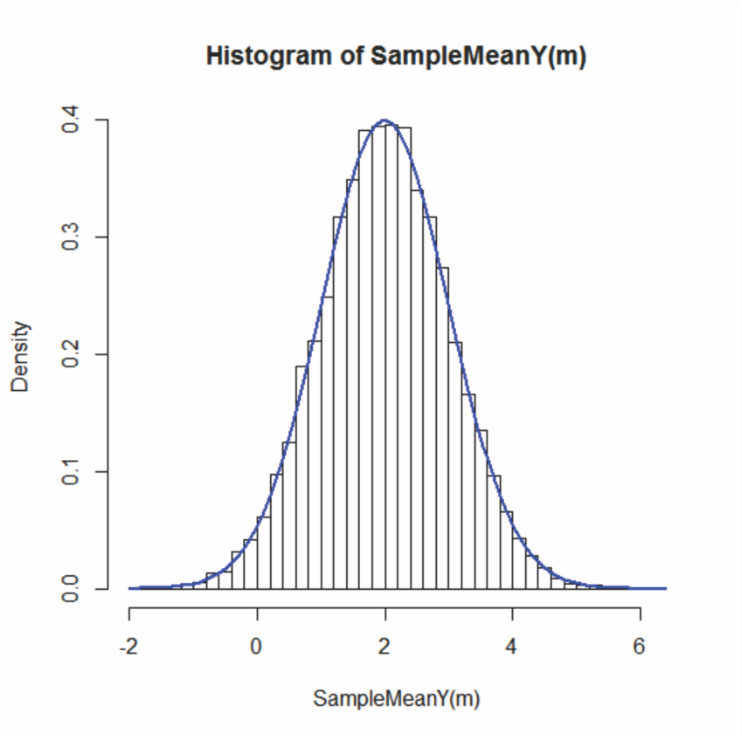
Đầu tiên ta mô phỏng cho phân phối của trung bình mẫu $\bar{X}$:

Đối với mẫu ngẫu nhiên lấy từ phân phối chuẩn, ta thành lập mẫu ngẫu nhiên kích thước n (xem hàm rnorm(n,mu,sigma)),

```
mu <- 2
sigma <- 2
Y <- function() rnorm(1,mu,sigma)
Y()
[1] 4.298447
vecY <- function(n) replicate(n,Y())
n=4
vecY(n)
[1] 3.298608 3.804107 3.272638 2.029872
vecY(n)
[1] 0.3444242 1.7142771 2.5943892 3.1046964
```

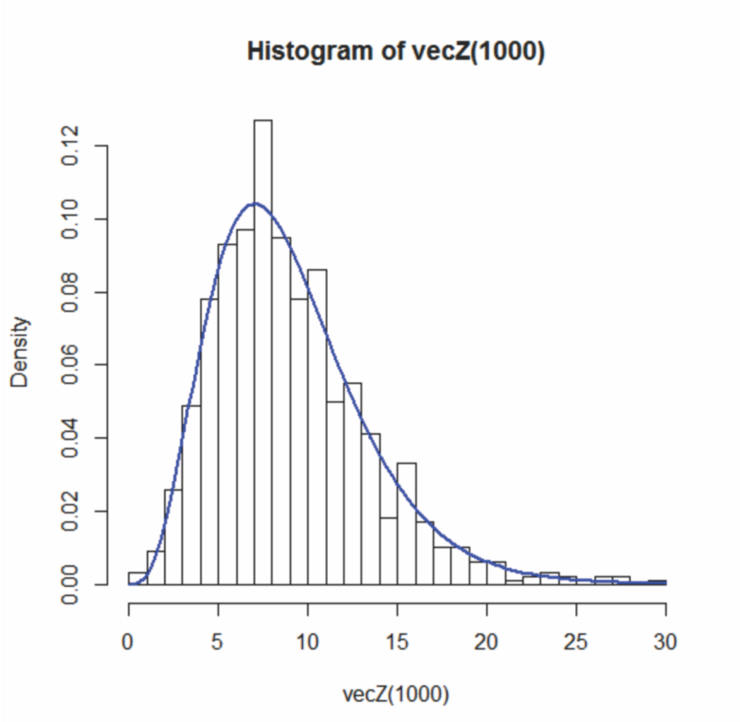
và ta có phân phối trung bình mẫu ngẫu nhiên kích thước n như sau

```
MeanY <- function() mean(vecY(n))
MeanY()
[1] 2.268478
SampleMeanY <- function(m) replicate(m,MeanY())
m=10000
hist(SampleMeanY(m),freq=0,breaks=40)
curve(dnorm(x,mu,sigma/sqrt(n)),col="blue",lty=1,lwd=2,add=TRUE)
```

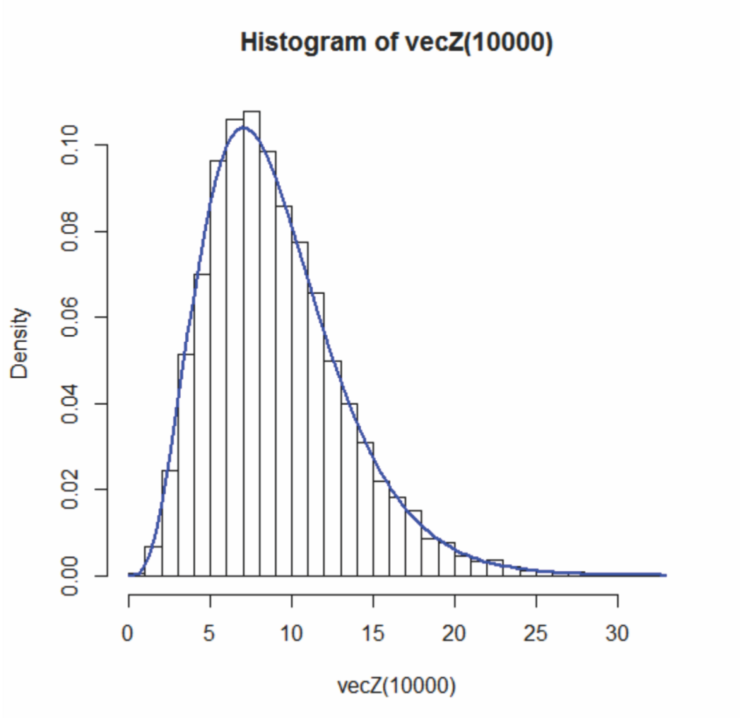


và ta có mô phỏng tương tự cho phân phối của $\frac{(n-1)S_X^2}{\sigma^2}$,

```
mu<-2
sigma<-2
n<-10
Z<-function() {
     x<-rnorm(n,mu,sigma)
     (n-1)*var(x)/sigma^2
}
Z()
[1] 2.074494
Z()
[1] 23.87044
vecZ<-function(m) replicate(m,Z())
hist(vecZ(1000),freq=0,breaks=40)
curve(dchisq(x,df=n-1),col="blue",lty=1,lwd=2,add=TRUE)
```




```
hist(vecZ(10000),freq=0,breaks=40)
curve(dchisq(x,df=n-1),col="blue",lty=1,lwd=2,add=TRUE)
```



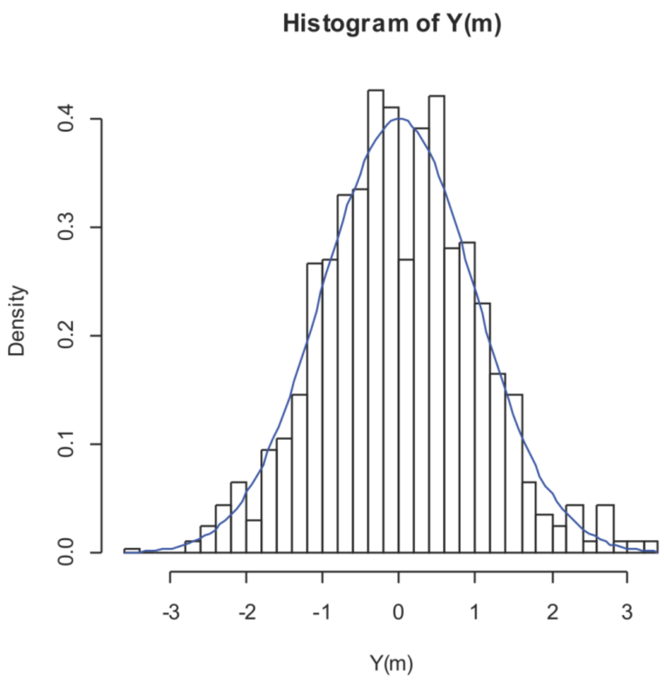
5.2. Mẫu ngẫu nhiên có cùng phân phối

Mô phỏng

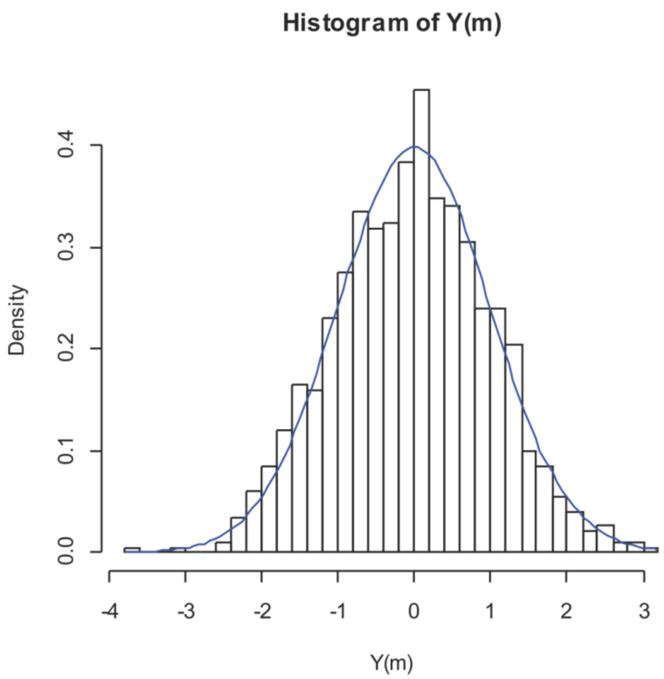
Định lý. *Xét mẫu ngẫu nhiên* $X_1, X_2, \cdots, X_n$ *lấy từ một phân phối có trung bình* μ *hữu hạn và phương sai dương* σ^2. *Ta có biến ngẫu nhiên* $Y_n = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ *có phân phối xấp xỉ với phân phối chuẩn* $N(0;1)$.

bằng cách lần lượt phát sinh các mẫu ngẫu nhiên kích thước $n=100$, 10.000, và 100.000, lấy từ phân phối nhị thức $B(10;0,3)$, để tính giá trị của biến $Y_n=\sqrt{n}\left(\bar{X}-\mu\right)/\sigma$. Lập lại $m=1.000$ lần để nhận được mẫu kích thước 1.000 cho Y_n để so sánh với phân phối chuẩn $N(0,1)$:

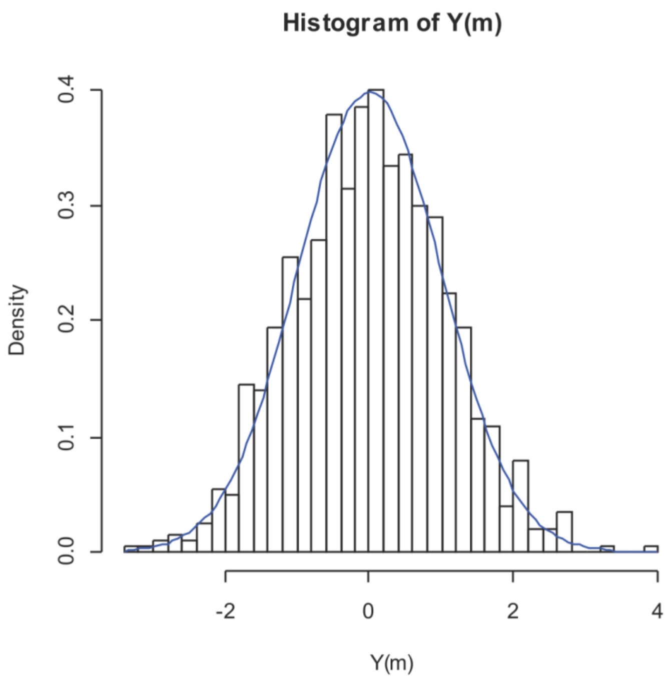
```
size=10
prob=0.3
mauX<-function(n) {
  x<-rbinom(n,size,prob)
  sqrt(n)*(mean(x)-3)/sqrt(2.1)
}
Y<-function(m) replicate(m,mauX(n))
n=100
m=1000
hist(Y(m),freq=0,breaks=40)
curve(dnorm(x),col="blue",lty=1,lwd=2,add=TRUE)
```




```
n=10000
hist(Y(m),freq=0,breaks=40)
curve(dnorm(x),col="blue",lty=1,lwd=2,add=TRUE)
```




```
n=100000
hist(Y(m),freq=0,breaks=40)
curve(dnorm(x),col="blue",lty=1,lwd=2,add=TRUE)
```

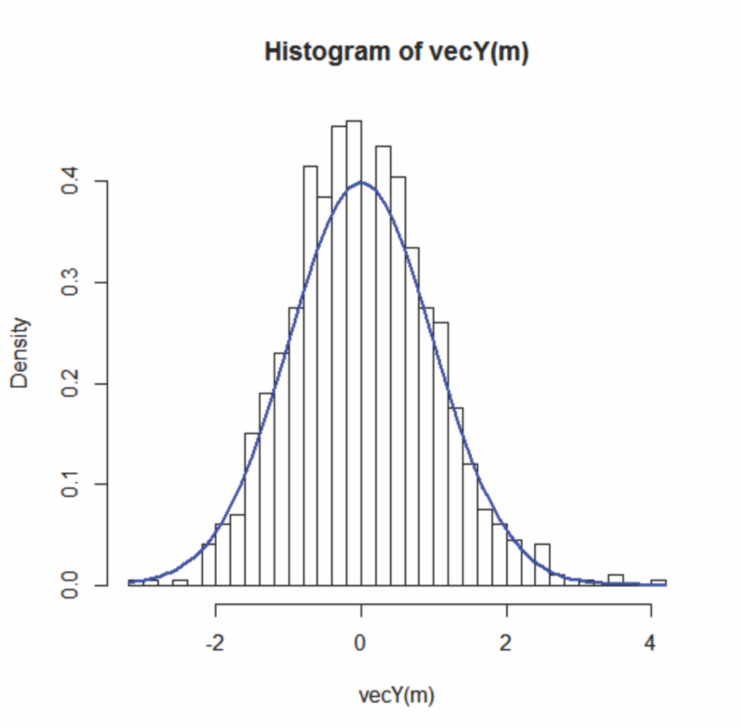


và mô phỏng cho

Định lý. *Xét mẫu ngẫu nhiên* $X_1, X_2, \cdots, X_n$ *lấy từ một phân phối Bernoulli* $B(1;p)$. *Ta có các biến ngẫu nhiên* $\frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ *và* $\frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}}$ *có phân phối xấp xỉ với phân phối chuẩn* $N(0;1)$.

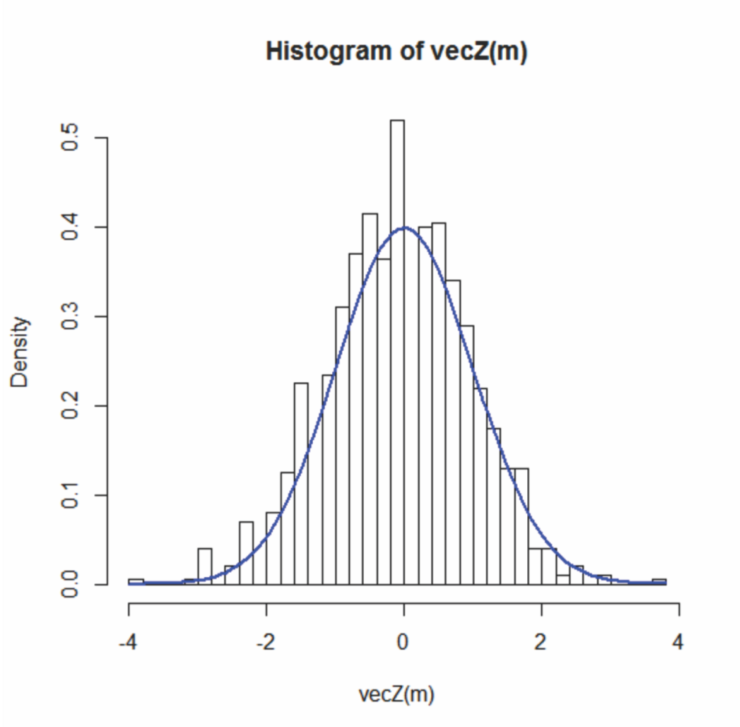
bằng cách phát sinh $m = 1.000$ mẫu ngẫu nhiên kích thước $n = 100$ lấy từ phân phối Bernoulli $B(1;0,3)$. Ứng với mỗi mẫu ngẫu nhiên, tính giá trị $Y = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ để nhận được mẫu ngẫu nhiên kích thước m cho Y và so sánh với phân phối chuẩn $N(0,1)$.

```
prob=0.3
Y<-function(n) {
  x<-rbinom(n,1,prob)
  (mean(x)-prob)*sqrt(n)/sqrt(prob*(1-prob))
}
vecY<-function(m) replicate(m,Y(n))
n=100
m=1000
hist(vecY(m),freq=0,breaks=40)
curve(dnorm(x),col="blue",lty=1,lwd=2,add=TRUE)
```



Tương tự, ứng với mỗi mẫu ngẫu nhiên, tính giá trị $Z = \frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}}$ để nhận được mẫu ngẫu nhiên kích thước m cho Z và ta cũng so sánh histogram của mẫu này so với phân phối chuẩn $N(0,1)$.

```
Z<-function(n) {
  x<-rbinom(n,1,prob)
  (mean(x)-prob)*sqrt(n)/sqrt(mean(x)*(1-mean(x)))
}
vecZ<-function(m) replicate(m,Z(n))
hist(vecZ(m),freq=0,breaks=40)
curve(dnorm(x),col="blue",lty=1,lwd=2,add=TRUE)
```



5.3. Bài tập

Bài 1. Cho X_1, X_2 là mẫu ngẫu nhiên kích thước 2 lấy từ phân phối chuẩn $N(0;1)$. Dùng hàm rnorm() phát sinh X_1, X_2 và $Y = X_1^2 + X_2^2$. Xây dựng hàm MauY phát sinh mẫu ngẫu nhiên kích thước n cho Y. Lần lượt phát sinh mẫu ngẫu nhiên kích thước 100, 1000, 10000 cho Y, vẽ biểu đồ tần suất và đồ thị hàm mật độ xác suất của phân phối Chi-bình phương với 2 bậc tự do cho từng trường hợp. Liên hệ với lý thuyết mẫu.