HW4

Christian Sallinger

20 4 2021

1. The mean of independent normal distributions

(a) Show that the moment generating function (mgf) of $X \sim \mathcal{N}(\mu, \sigma^2)$ is of the form

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

(b) Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $Y = aX + b$ with fixed real constants a and b. Show that $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
(c) Let $X_1, \ldots, X_n$ be independent identically distributed random variables with $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Show that the mean $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is also normally distributed and $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Solution:

The mgf is given via $M_X(t) = \mathbb{E}(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$. We will show this first for $Z \sim \mathcal{N}(0, 1)$:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(2tx - x^2)/2} dx = \cdots$$

We can substitute $y = x - t$ and get

$$\cdots = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(2t(y+t) - (y+t)^2)/2} dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(t^2 - y^2)/2} dy = e^{t^2/2}$$

Where the last equality follows from the well known calculus result $\int_{\mathbb{R}} e^{x^2/2} dx = \sqrt{2\pi}$.

We will show in (b), that $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$. We can now use the properties of the mfg to show the wanted result:

$$M_{\sigma Z + \mu}(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$$

which is exactly what we wanted to show.

(b) To show this we use our transformation theorem: The function

$$\begin{aligned} g :& (-\infty, \infty) \to (-\infty, \infty) \\ & x \mapsto ax + b \end{aligned}$$

is invertible with differentiable inverse $h(x) = \frac{x-b}{a}$ as long as $a \neq 0$. Since Y would be constant for $a = 0$ we will not consider this case anyway. For the pdf of Y we now get:

$$
\begin{aligned}
f_Y(y) = f_X(h(y))|h'(y)| &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left((y-b)/a-\mu\right)^2}{2\sigma^2}} \frac{1}{a} \\
&= \frac{1}{\sqrt{2\pi a^2\sigma^2}} e^{-\frac{(y-b-\mu a)^2}{2\sigma^2 a^2}} = \frac{1}{\sqrt{2\pi a^2\sigma^2}} e^{-\frac{(y-(\mu a+b))^2}{2\sigma^2 a^2}}
\end{aligned}
$$

So we can see that $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ really holds.

(c) Since the random variables are independent we know from the lecture that the mgf of $\overline{X}$ is given by

$$
M_{\overline{X}}(t) = \left(M_{X_1}(\frac{t}{n})\right)^n .
$$

We have already seen in (a) what the mfg of X_1 looks like, so we can plug it in here:

$$
\left(M_{X_1}(\frac{t}{n})\right)^n = \left(e^{t\mu/n + \frac{\sigma^2 t^2}{2n^2}}\right)^n = e^{t\mu + \frac{\sigma^2 t^2}{n}}
$$

Since the mgf exists in a neighbourhood of zero the distribution is uniquely determined by its mgf and as we can see this mgf corresponds to a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, which is exactly what we wanted to show.

2. Sum of two independet distributions

(a) Let $X \sim \mathcal{P}(\lambda_1)$ and $Y \sim \mathcal{P}(\lambda_2)$ be two independent Poisson random variables. Show that

$$
X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2).
$$

(b) Let U and V be two independet random variables with exponential distribution $\exp(\lambda)$. Show that

$$
\begin{aligned}
U + V &\sim Gamma(2, \lambda) \quad \text{and} \\
\min\{U, V\} &\sim \exp(2\lambda).
\end{aligned}
$$

Hint: It is useful to use moment generating functions. Recall, the pdf of a random variable $X \sim Gamma(\alpha, \frac{1}{\beta})$ is

$$
f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}, & x > 0 \\ 0, & x \leq 0 \end{cases}
$$

and its mgf is of the form $\left(\frac{1}{1-\beta t}\right)^\alpha$ for $t \leq \frac{1}{\beta}$. Particularly, the pdf of a random variable $X \sim \exp(\lambda) = Gamma(1, \frac{1}{\lambda})$ is of the form

$$
f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}
$$

Solution: (a) We recall that the pmf of a random variable $Z \sim \mathcal{P}(\lambda)$ is given by

$$
f_Z(k) = \frac{\lambda^k}{k!} e^{-\lambda}.
$$

Now we can use the convolution formula for discrete random variables:

$$
\begin{aligned}
f_{X+Y}(k) &= \sum_{y\in\{0,\dots,k\}} f_Y(y)f_X(k-y) = \sum_{y\in\{0,\dots,k\}} \frac{\lambda_2^y}{y!}e^{-\lambda_1}\frac{\lambda_1^{k-y}}{(k-y)!}e^{-\lambda_2} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{y\in\{0,\dots,k\}} \frac{\lambda_2^y\lambda_1^{k-y}}{y!(k-y)!} = e^{-(\lambda_1+\lambda_2)}\frac{1}{k!}\sum_{y\in\{0,\dots,k\}}\binom{k}{y}\lambda_2^y\lambda_1^{k-y} \\
&= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}
\end{aligned}
$$

(b) Here we use the convolution formula for continous random variables:

$$
\begin{aligned}
f_{U+V}(z) &= \int_{\mathbb{R}} f_U(z-y)f_Y(y)dy = \lambda^2\int_{\mathbb{R}} e^{-\lambda(z-y)}e^{-\lambda y}1_{(0,\infty)}(z-y)1_{(0,\infty)}(y)dy \\
&= \lambda^2\int_0^z e^{-\lambda z}dy = \lambda^2 z e^{-\lambda z}, \quad \text{z} > 0
\end{aligned}
$$

This is what we wanted to show, since $\Gamma(2) = 1$. For the second part we use the general formula for the cdf of the minimum of two independent, idendtically distributed random variables from last week. First we define $Z = \min\{U, V\}$. Also recall that the cdf for a random variable $x \sim \exp(\lambda)$ is given by

$$
F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}
$$

So we get:

$$
F_Z(z) = 2F_U(z) - F_U(z)^2 = 2(1-e^{-\lambda z}) - (1-e^{-\lambda z})^2 = 2 - 2e^{-\lambda z} - 1 + 2e^{-\lambda z} - e^{-2\lambda z} = 1 - e^{-2\lambda z}, \quad z \geq 0
$$

Comparing with the cdf above we see that $\min\{U, V\} \sim \exp(2\lambda)$ really holds.

3. Real roots

Let A, B and C be independent ramdom variables, uniformly distributed on $(0, 1)$.

(a) What is the probability that the quadratic equation $Ax^2 + Bx + C = 0$ has real roots?
(b) Consider the following code in R. What does it do and how is it related to your solution in part (a)?

```
n = 10000
a = runif(n)
b = runif(n)
c = runif(n)
sum(b^2>4*a*c)/n
```

```
## [1] 0.2529
```

Hint: In HW2/ex. 3(b) we showed that if X has uniform $(0, 1)$ distribution then $-\log X$ has exponential distribution $\exp(1)$. In an analogue way, one can prove that $-s\log X \sim \exp(\frac{1}{s})$ for any $s > 0$. Also, in HW4/ex. 2(b) we proved that the sum of two independent exponential distributions is a gamma distribution. Namely, if $X \sim \exp(1)$ and $Y \sim \exp(1)$ are independent then $X + Y \sim \mathit{Gamma}(2, 1)$.

Solution:

Using the "Grosse Loesungsformel" we see that the quadratic equation has real roots iff $B^2 \geq 4AC$ holds. By use of the monotony of the logarithm as well as random variables $Z \sim \exp\left(\frac{1}{2}\right)$, $U, V \sim \exp(1)$ and $G \sim \mathit{Gamma}(2, 1)$ we follow the hint:

$$
\begin{aligned}
\mathbb{P}(B^2 \geq 4AC) &\iff \mathbb{P}(\log(B^2) \geq \log(4AC)) \\
&\iff \mathbb{P}(-2\log(B) \leq -\log(4AC)) \iff \mathbb{P}(-2\log(B) \leq -\log(4) - \log(A) - \log(C)) \\
&\iff \mathbb{P}(Z - (-\log(A) - \log(C)) \leq -\log(4)) \iff \mathbb{P}(Z - (U + V) \leq -\log(4)) \\
&\iff \mathbb{P}(Z - G \leq -\log(4))
\end{aligned}
$$

To get the pdf of $Z - G$ we use the convolution formula:

$$
\begin{aligned}
f_{Z+(-G)}(z) &= \int_{\mathbb{R}} f_Z(z-y) f_G(-z) dy = -\frac{1}{2} \int_{\mathbb{R}} e^{-\frac{z-y}{2}} y e^y \ 1_{(-\infty,0)}(y) \ 1_{[0,\infty)}(z-y) dy = -\frac{1}{2} \int_{-\infty}^{\min\{z,0\}} e^{-\frac{z-y}{2}} y e^y dy \\
&= \begin{cases} -\frac{(6z-4)e^z}{18}, & z < 0 \\ \frac{2e^{\frac{z}{2}}}{9}, & z \geq 0 \end{cases}
\end{aligned}
$$

Now we can calculate the probability:

$$
\mathbb{P}(Z - G \leq -\log(4)) = F_{Z-G}(-\log(4)) = \int_{-\infty}^{-\log(4)} f_{Z-G}(z) dz \approx 0.2544
$$

(b) The command "runif" creates a vector of length n with samples taken from an uniformly distributed random variable. In the sum we get 1 if $b_j^2 > 4a_j c_j$ for the j-th entry of the vectors and 0 otherwise. So in the sum we have the total number of times the inequality holds for the n-samples. Dividing by n then gives the mean. The value gives us a good approximation of the real probability because by the strong law of large numbers it converges to it. To see this we define a random variable X that is 1 if $B^2 > 4AC$ and 0 otherwise. The expectation is

$$
\mathbb{E}(X) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = \mathbb{P}(B^2 > 4AC)
$$

4. Sum and average

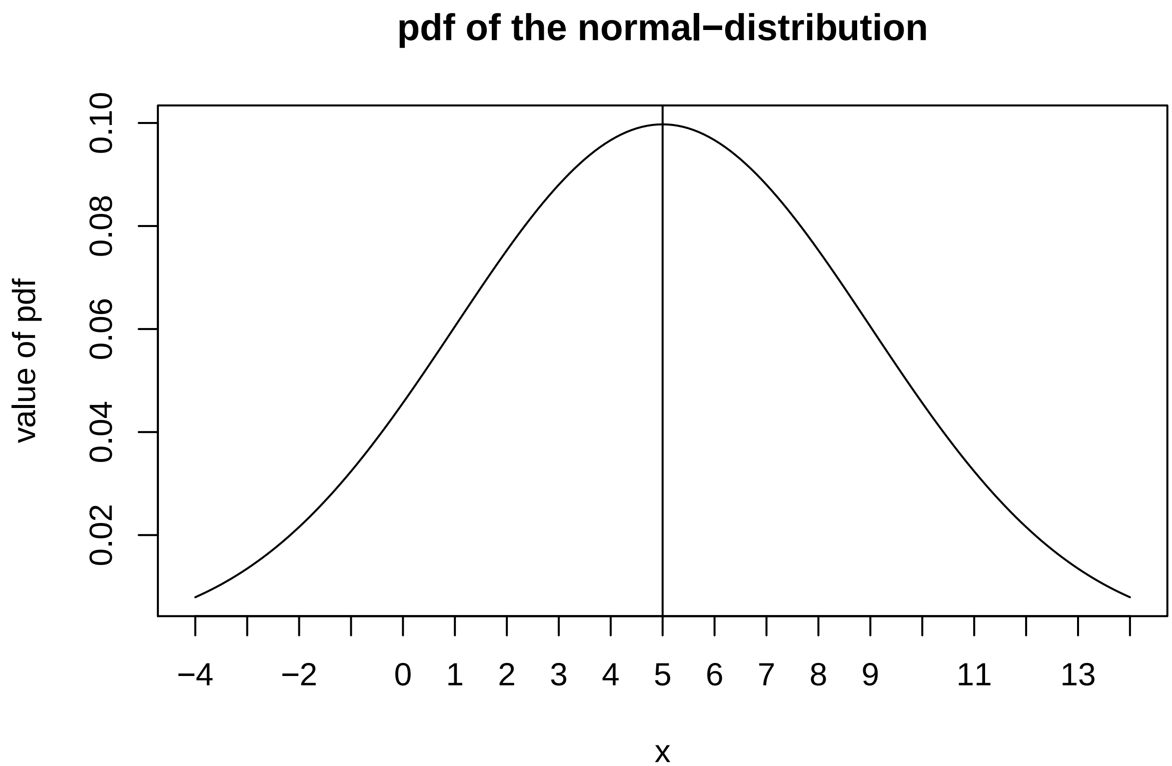
Let X be a random variable with $\mathcal{N}(5, 2^2)$. Let $X_1, X_2, \ldots, X_{50}$ be independent identically distributed copies of X. Let S be their sum and $\overline{X}$ their average, i.e.

$$
S = X_1 + \cdots + X_{50} \quad \text{and} \quad \overline{X} = \frac{1}{50}(X_1 + \cdots + X_{50}).
$$

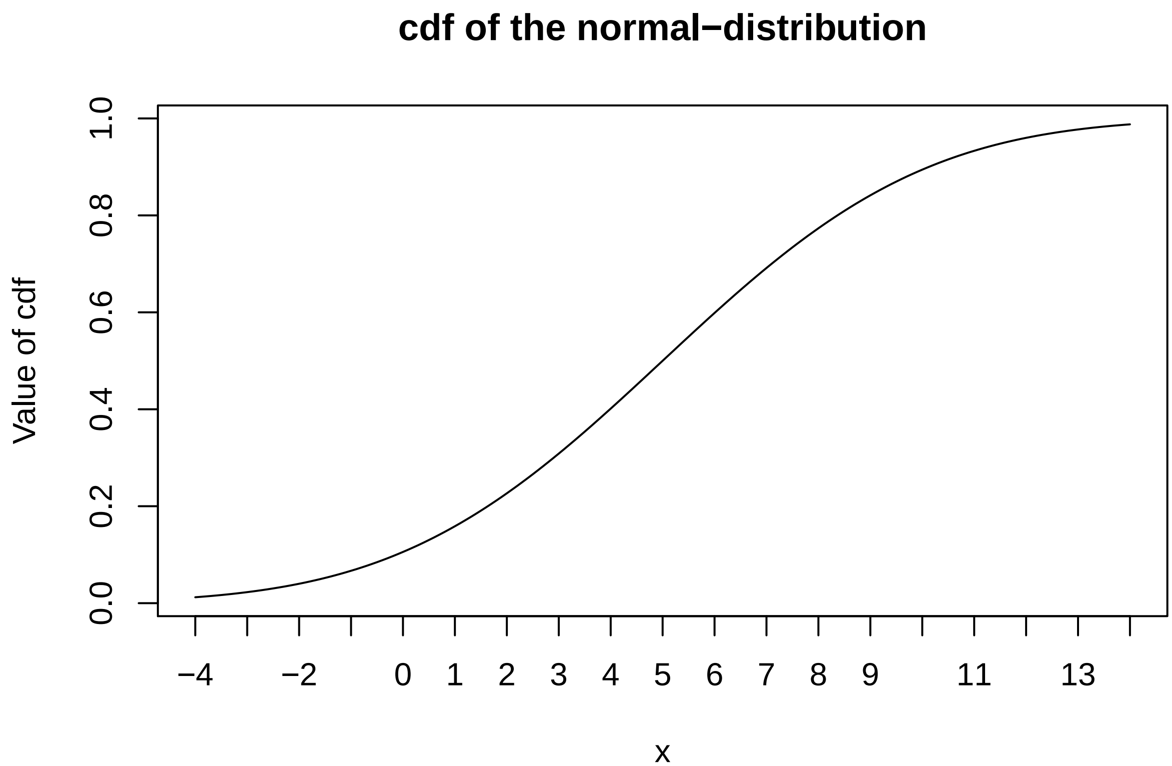
(a) Plot the density and the distribution function for X using R.
(b) What are the expectation and the standard deviation of S and $\overline{X}$?
(c) Generate a sample of 50 numbers from $\mathcal{N}(5, 2^2)$. Plot the histogram for this sample. Do the same for a sample of 500 numbers from $\mathcal{N}(5, 2^2)$.

Solution: (a)

```
x <- seq(-4,14,0.1)
y <- dnorm(x,5,4)
plot(x,y,type="l", main = "pdf of the normal-distribution",xlab = "x", ylab = "value of pdf"
     ,xaxp  = c(-4, 14, 18))
abline(v = 5)
```




```
z <- pnorm(x,5,4)
plot(x,z,type = "l", main = "cdf of the normal-distribution",xlab = "x", ylab = "Value of cdf"
     ,xaxp  = c(-4, 14, 18))
```



(b) We first calculate the expectations:

$$\mathbb{E}(S) = \sum_{i=1}^{50} \mathbb{E}(X_i) = 50 \cdot 5 = 250$$

$$\mathbb{E}(\overline{X}) = \frac{1}{50}\mathbb{E}(S) = 5$$

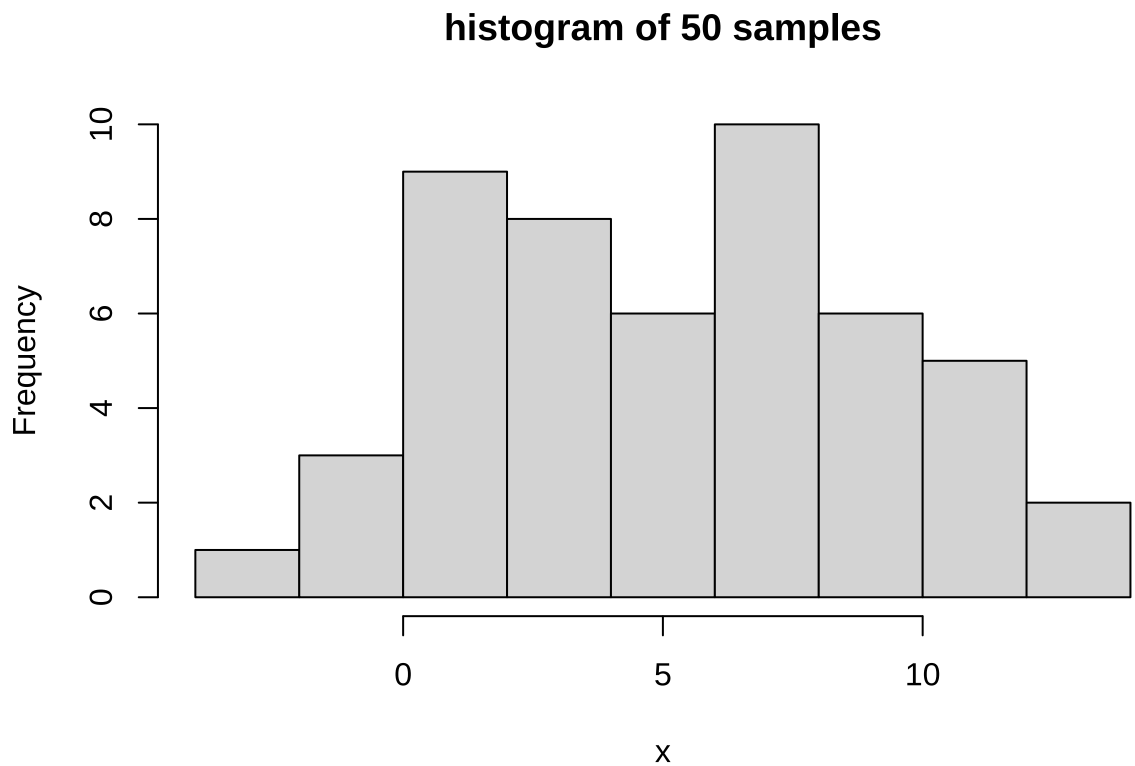
For the variances, since the random variables are independent, we get:

$$\mathbb{V}ar(S) = \sum_{i=1}^{50} \mathbb{V}ar(X_i) = 200$$

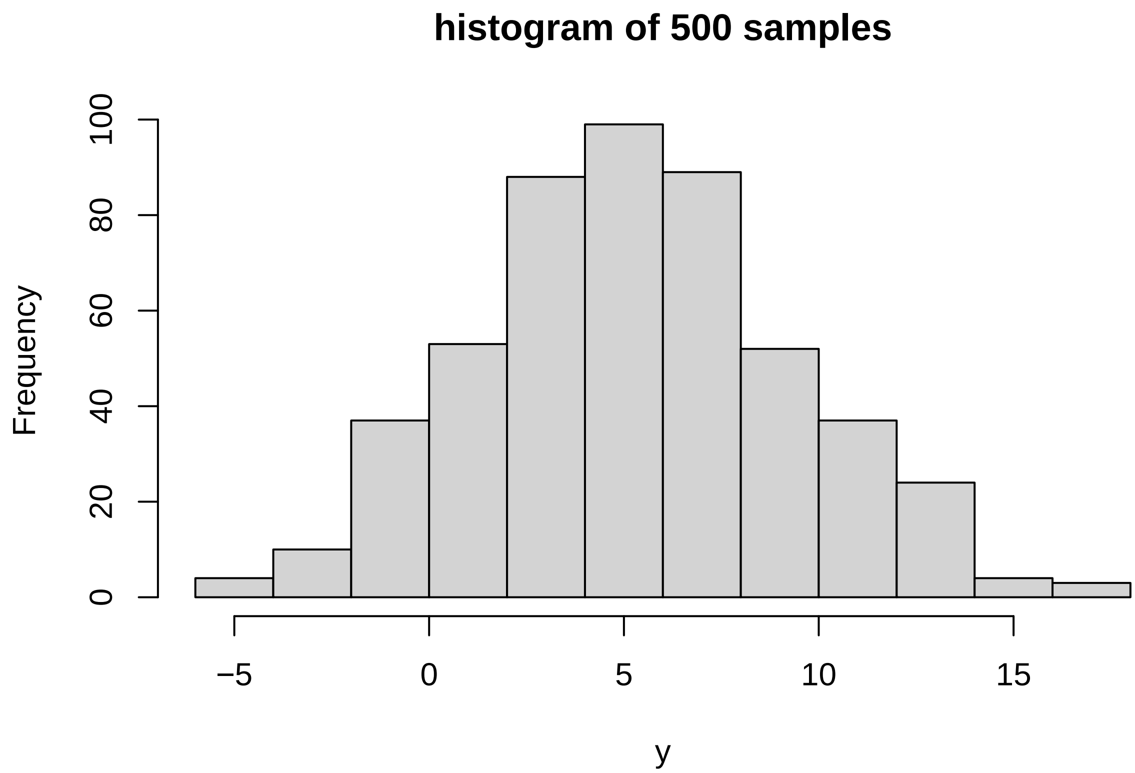
$$\mathbb{V}ar(\overline{X}) = \left(\frac{1}{50}\right)^2 \mathbb{V}ar(S) = \frac{200}{2500} = \frac{8}{100}$$

(c)

```
x = rnorm(50,5,4)
hist(x, main = "histogram of 50 samples")
```




```
y = rnorm(500,5,4)
hist(y, main = "histogram of 500 samples")
```



5. Central Limit Theorem

Let $\overline{X}_1$ and $\overline{X}_2$ be the means of two independent samples of size n from the same population with variance σ^2. Use the Cental limit theorem to find a value for n so that

$$P(|\overline{X}_1 - \overline{X}_2| < \frac{\sigma}{50}) \approx 0.99.$$

Justify your calculations.

Solution:

We first define $\overline{X}_n := \overline{X_1} - \overline{X_2} = \frac{1}{n}\sum_{i=1}^{n}(X_{1,i} - X_{2,i})$. We now want to calculate the mean of X_i for some i:

$$\mathbb{E}(X_i) = \mathbb{E}(X_{1,i}) - \mathbb{E}(X_{2,i}) = \mu - \mu = 0$$

Since they are independent samples we can also calculate the variance rather easily.

$$\mathbb{V}ar(X_i) = \mathbb{V}ar(X_{1,i} - X_{2,i}) = \mathbb{V}ar(X_{1,i}) + \mathbb{V}ar(X_{2,i}) = 2\sigma^2$$

The cental limit theorem now states that the cdf of $\sqrt{n} \cdot \frac{\overline{X}_n}{\sqrt{2}\sigma}$ converges to the cdf of the standard normal distribution. So we can use Φ as an approximation.

$$\mathbb{P}(|\overline{X}_1 - \overline{X}_2| < \frac{\sigma}{50}) = \mathbb{P}(-\frac{\sigma}{50} < \overline{X}_n < \frac{\sigma}{50}) = \mathbb{P}(-\frac{1}{50} < \frac{\sqrt{2}}{\sigma\sqrt{2}}\overline{X}_n < \frac{1}{50})$$
$$= \mathbb{P}(-\frac{\sqrt{n}}{50\sqrt{2}} < \frac{\sqrt{n}}{\sigma\sqrt{2}}\overline{X}_n < \frac{\sqrt{n}}{50\sqrt{2}}) \approx \Phi\Big(\frac{\sqrt{n}}{50\sqrt{2}}\Big) - (1 - \Phi\Big(\frac{\sqrt{n}}{50\sqrt{2}}\Big)) = 2\Phi\Big(\frac{\sqrt{n}}{50\sqrt{2}}\Big) - 1$$

Now we want to find n so that

$$2\Phi\Big(\frac{\sqrt{n}}{50\sqrt{2}}\Big) - 1 \approx 0.99 \iff \Phi\Big(\frac{\sqrt{n}}{50\sqrt{2}}\Big) \approx 0.995 \iff \frac{\sqrt{n}}{50\sqrt{2}} = 2.58 \iff n = 33282$$