A Very Basic Introduction to R – Part VI

Simulation of Normal Random Variables

The `rnorm()` function can be used to simulate N independent normal random variables.

For example, we can generate 5 standard normal random numbers as follows:

```
> rnorm(5)
```

```
[1]  0.6636134  0.7429042  1.1421494 -0.3520732 -1.4082090
```

Recall that a standard normal random variable has mean 0 and standard deviation 1. We compare these values with sample values in the next example, where 500 standard normal numbers are generated:

```
> x <- rnorm(500)
> mean(x)  # close to 0
```

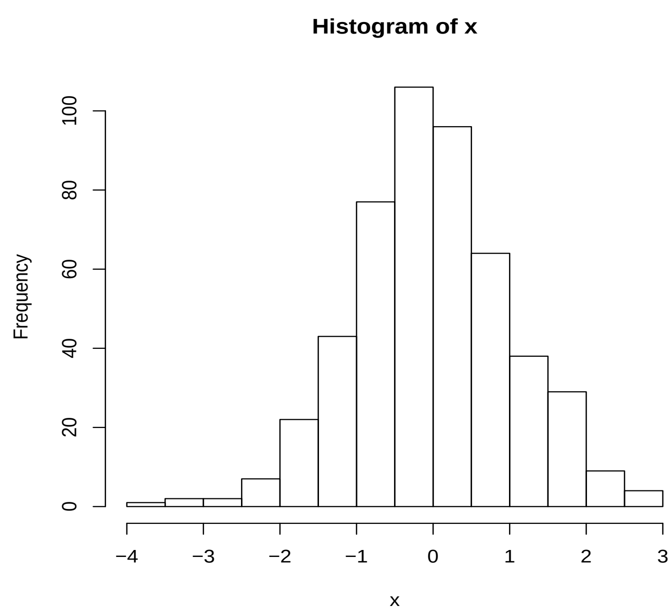
```
[1] -0.01183159
```

```
> sd(x) # close to 1
```

```
[1] 1.016111
```

Look at the histogram of the sample contained in `x`:

```
> hist(x)
```

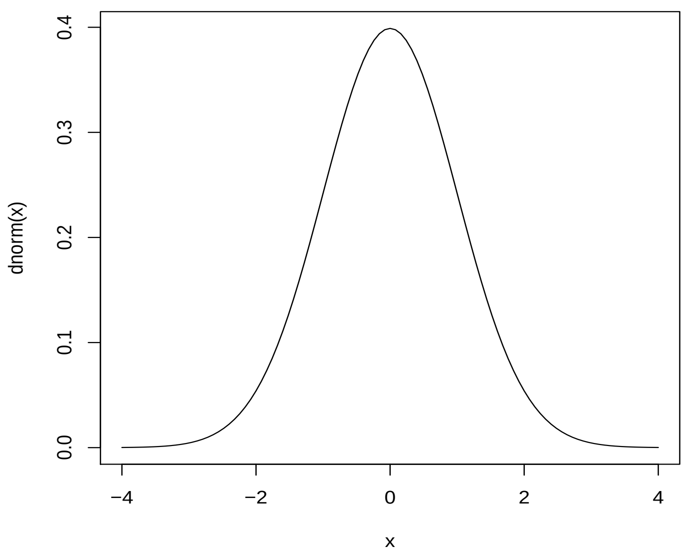


It has the typical shape for the normal distribution - similar to a bell. The histogram is an estimate of the standard normal probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

We can plot this function using the `curve()` function:

```
> curve(dnorm, -4, 4)  # outside [-4,4] the function is very close to 0
```



Note the similarity between the above histogram and the density curve.

The general normal probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Here, μ denotes the mean or expected value, and σ denotes the standard deviation. In other words,

$$E[X] = \mu$$

and

$$\text{Var}(X) = \sigma^2.$$

To obtain a normal sample with nonzero mean and arbitrary standard deviation, use the `mean` and `sd` arguments. For example, simulating 25 normals with mean 5 and standard deviation 6, we type

```
> y <- rnorm(25, mean=5, sd=2)
```

The sample mean is an estimate of the population mean (or expected value) and the sample standard deviation is an estimate of the theoretical standard deviation:

```
> mean(y) # close to 5
```

```
[1] 4.227553
```

```
> sd(y) # close to 2
```

```
[1] 1.717066
```

```
> var(y) # close to 4
```

```
[1] 2.948316
```

We can estimate the probability that a normal random variable Y is less than a given value, x:

$$P(Y \leq x)$$

by simulating a large number of Y values and finding the proportion that are less than x:

```
> Y <- rnorm(100000) # a large number of standard normals
> sum(Y < 2)/100000 # the proportion less than 2
```

```
[1] 0.97768
```

The theoretical value can be calculated using the `pnorm()` function:

```
> pnorm(2) # close to the sample value
```

```
[1] 0.9772499
```

For another example, estimate the probability that a normal random variable is less than 3.4, if its true mean is 7 and its standard deviation is 4, and compare with the theoretical value.

```
> Y <- rnorm(100000, 7, 4)
> sum(Y < 3.4)/100000 # estimated proportion
```

```
[1] 0.18339
```

```
> pnorm(3.4, 7, 4) # theoretical value
```

```
[1] 0.1840601
```

Exercises

1. Generate 20 normal random numbers with mean 15 and standard deviation 4, and plot their histogram.
2. Calculate the expected value and variance of the above numbers.
3. Assign 2000 normal random numbers with parameters $\mu = 4.5$ and $\sigma = 3$ to a vector called `U`.
4. Calculate the mean value of `U`. Compare with the theoretical value.
5. Calculate the variance of `U`. Compare with the theoretical value.
6. Plot the histogram of the sample in `U`. Compare with a plot of the theoretical probability distribution.
7. Use simulation to estimate the expected value of U^3.
8. Estimate the probability that U is less than 6.0. Compare with the theoretical value.
9. Suppose V is a normal random variable with mean 10 and variance 25. Simulate 10000 values of V and compute their mean and variance, comparing with the theoretical values.
10. Estimate $P(V > 2)$ and compare with the theoretical value.