

Open Intro Biostat R Companion:

Chapter 3

Morgan F. Breitmeyer

Contents

1	Normal Distribution	1
1.1	Direct Plot	1
1.2	Through Simulation	2
1.3	A Comparison	3
1.4	Probability Functions	3
2	Binomial Distribution	4
3	Poisson Distribution	5
4	Geometric Distribution	5
5	Negative Binomial Distribution	5

Chapter 3 presents the idea of several known and commonly used distributions. R is a powerful source for accessing and using these distributions.

1 Normal Distribution

We have learned that the **normal distribution** can be written as

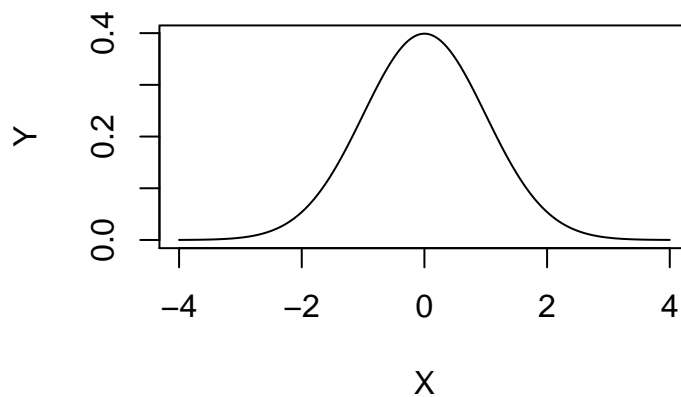
$$N(\mu, \sigma)$$

where μ is the mean of the distribution and σ is the standard deviation. Using this idea, we can create a standard normal distribution in one of two ways:

1.1 Direct Plot

We can use a function in R called the *dnorm* function which for a given possible value of the distribution, returns the probability of the distribution holding that value.

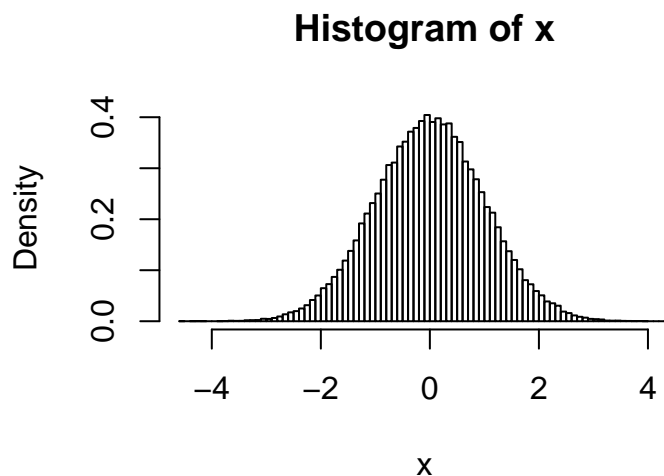
```
## getting a list of values to evaluate distribution
X = seq(-4, 4, 0.01)
## getting normal distribution value of the given X values
Y = dnorm(X, mean = 0, sd = 1)
## plotting results
plot(X, Y, type = "l")
```



1.2 Through Simulation

The second method involves sampling randomly from the known normal distribution to simulate the probabilities that we drew directly in the above method. In this case, we use the function `rnorm`, which takes n number of random samples from the normal distribution with the given mean and standard deviation. In the `hist` command, we include the specification that `freq = FALSE` to make the plot reflect percentages rather than frequency counts.

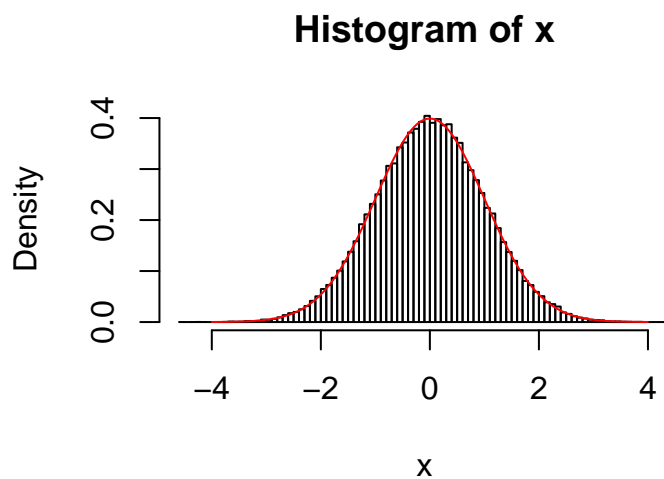
```
set.seed(1)
x <- rnorm(n = 100000, mean = 0, sd = 1)
hist(x, breaks = 100, freq = FALSE)
```



1.3 A Comparison

To see how comparable the two methods are, we can plot them on the same graph as follows.

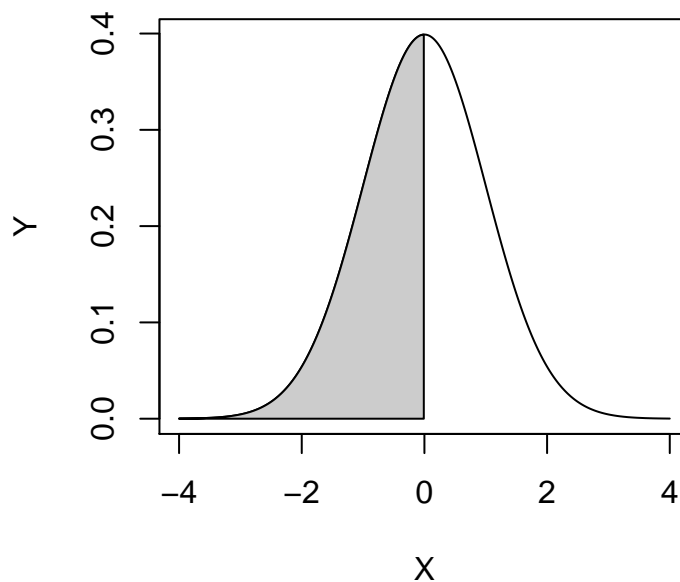
```
hist(x, freq = FALSE, breaks = 100)
lines(X,Y, col = "red")
```



Note that the two show very similar results. For practical purposes, we typically use *dnorm* as it only requires one command and is the asymptotic result of simulations, rather than an approximation.

1.4 Probability Functions

For the normal distribution, there are two more functions that are very useful, *pnorm* and *qnorm*. The *pnorm* tells you the percentage of the normal distribution that is less than or equal to the value you put in. This corresponds to the gray area in the below plot and can be written as $pnorm(X) = P(normal < X)$. On the other hand, the *qnorm* function is the inverse function, instead giving the X value such that the given percentage of the distribution is less than or equal to that value.



An example of using these two functions is as follows

```
pnorm(0, mean = 0, sd = 1)
## [1] 0.5
qnorm(0.5, mean = 0, sd = 1)
## [1] 0
```

2 Binomial Distribution

The functions we saw above for the normal distribution can be similarly used for the **binomial distribution** as follows where the necessary arguments are *size* and *prob*

- *dbinom*

-
- *rbinom*
 - *pbinom*
 - *qbinom*

3 Poisson Distribution

Similarly, we can build out the functions for the **poisson distribution** as the following where the necessary argument is *lambda*

- *dbinom*
- *rbinom*
- *pbinom*
- *qbinom*

4 Geometric Distribution

We can follow the same pattern for the **geometric distribution** to get the following where the necessary argument is *prob*

- *dgeom*
- *rgeom*
- *pgeom*
- *qgeom*

5 Negative Binomial Distribution

Finally, we can follow the same pattern for the **negative binomial distribution** to get the following where the necessary arguments are *prob* and *size*

- *dnbinom*
- *rnbinom*
- *pnbinom*
- *qnbinom*