**Gamma probability distribution**

This PDF is often used in survival and reliability analysis. For example, these PDFs are used for modeling lifetimes of individuals or manufactured products.

The PDF contains the gamma function defined by



for α > 0.

Notes:

* When α is a positive integer, Γ(α) = (α-1)!; for example, Γ(3) = (3-1)! = 2! = 2×1 = 2
* Through integrating by parts, one can show Γ(α) =   
  (α-1)Γ(α-1)
* Γ(1/2) = 
* R function: gamma()
* Sage function: gamma()

The gamma PDF for a random variable Y is

****

where α > 0 and β > 0 are parameters.

Notes:

* In most realistic applications, α and β will not be known and we will need to estimate them. How to do this will be discussed in future sections.
* α controls the shape of the PDF because it mostly influences the “peakedness” of the PDF.
* β controls the scale of the PDF because it mostly influences the spread of the PDF.
* E(Y) = αβ, Var(Y) = αβ2

Proof for E(Y)







Notice that  is a gamma PDF with α + 1 and β as its parameters!

Thus,  = 1 and

.

A similar proof can be performed for E(Y2), which will lead to Var(Y).

Example: Basic gamma PDF calculations (gamma.R)

The dgamma() function finds f(y) in R. For example,

Code change from what was shown in the video

> pdf <- function(y, alpha, beta) {

dgamma(x = y, shape = alpha, scale = beta)

}

> integrate(f = pdf, lower = 0, upper = Inf, alpha = 1,

beta = 1)

1 with absolute error < 5.7e-05

for α = 1 and β = 1.

The pgamma() function evaluates F(y) in R. For example,

> # P(Y < 1)

> integrate(f = pdf, lower = 0, upper = 1, alpha = 1,

beta = 1)

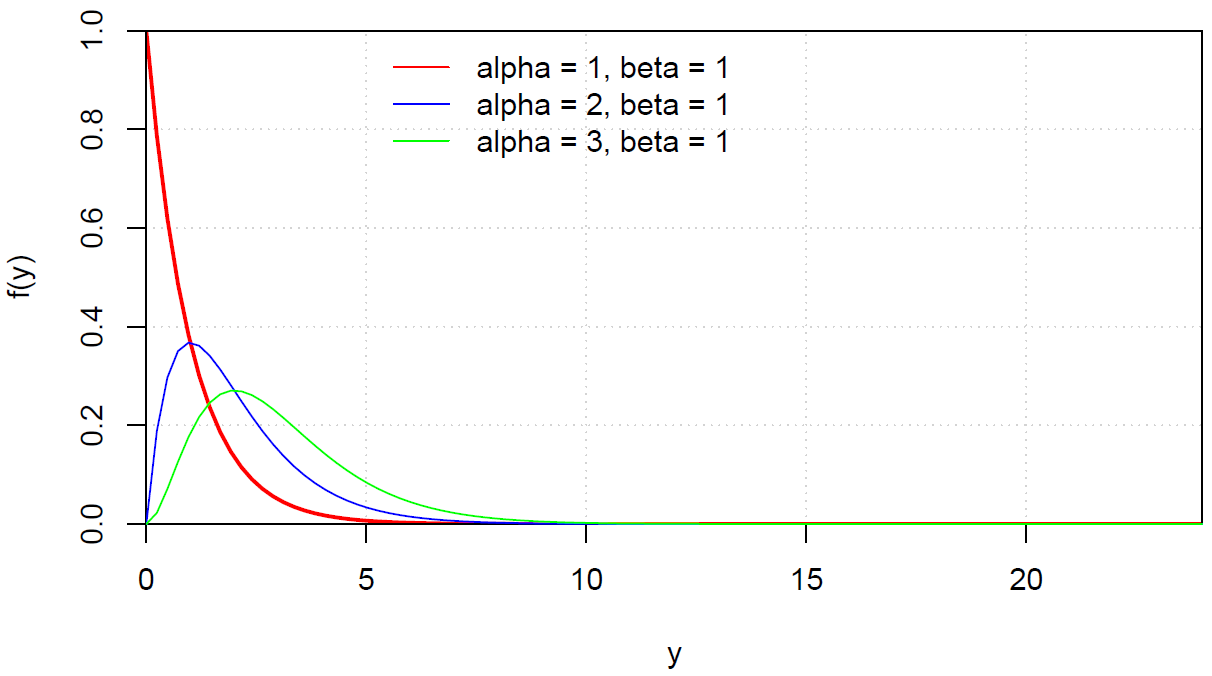
0.6321206 with absolute error < 7e-15

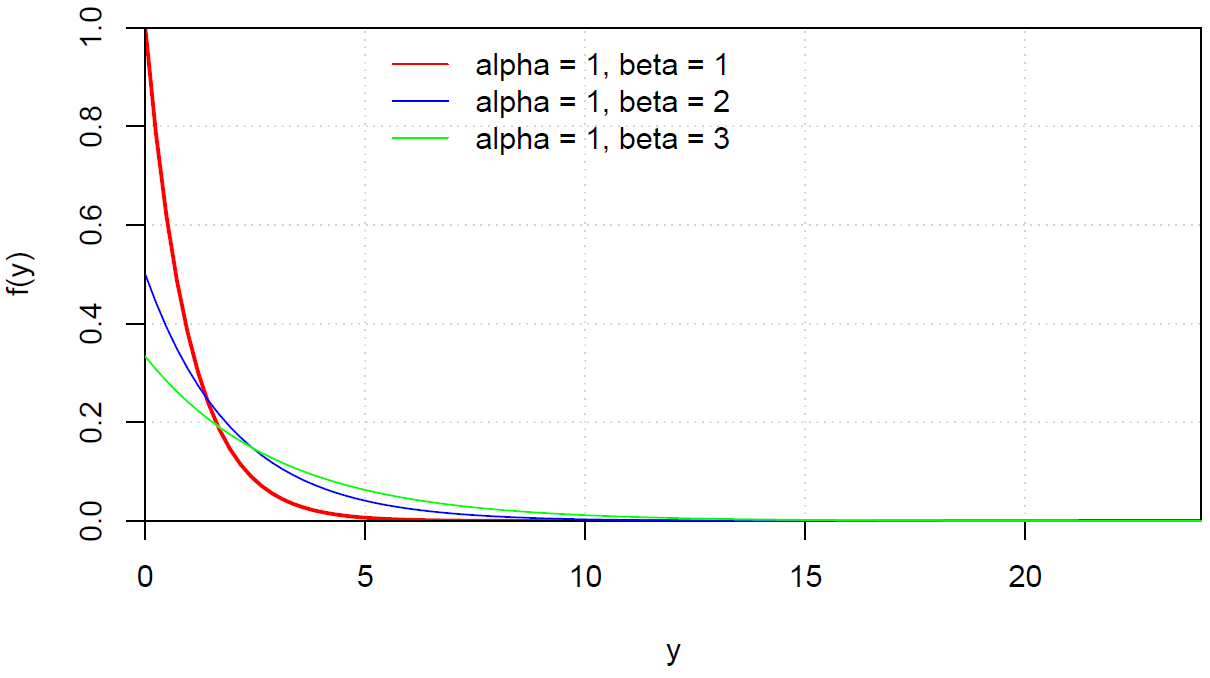
> pgamma(q = 1, shape = 1, rate = 1)

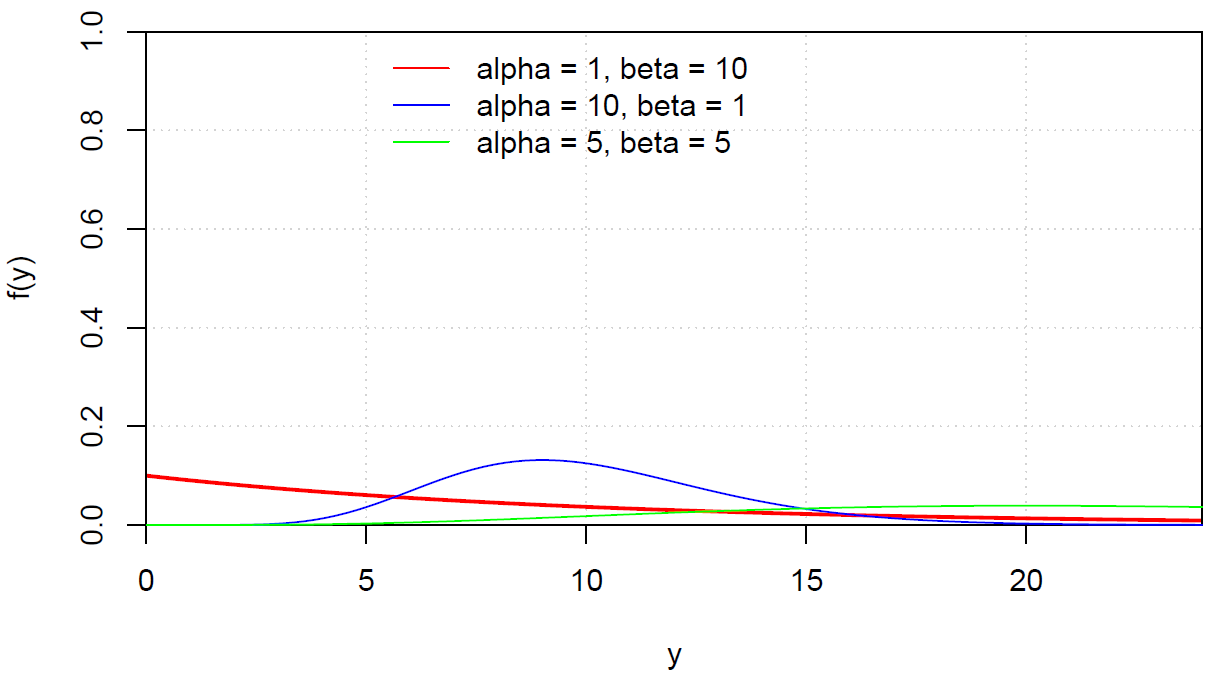
[1] 0.6321206

for α = 1 and β = 1.

Below are a few plots for comparison purposes. The x- and y-axis scales are fixed for comparison purposes.







Summary of the parameter values in the plots:

|  |  |  |  |
| --- | --- | --- | --- |
| α | β | μ | σ2 |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 |
| 3 | 1 | 3 | 3 |
| 1 | 2 | 2 | 4 |
| 1 | 3 | 3 | 9 |
| 10 | 1 | 10 | 10 |
| 1 | 10 | 10 | 100 |
| 5 | 5 | 25 | 125 |

Questions:

* What happens if α and/or β are increased?
* What happens if α and/or β are decreased?
* Why would someone want to use different values of α and/or β?

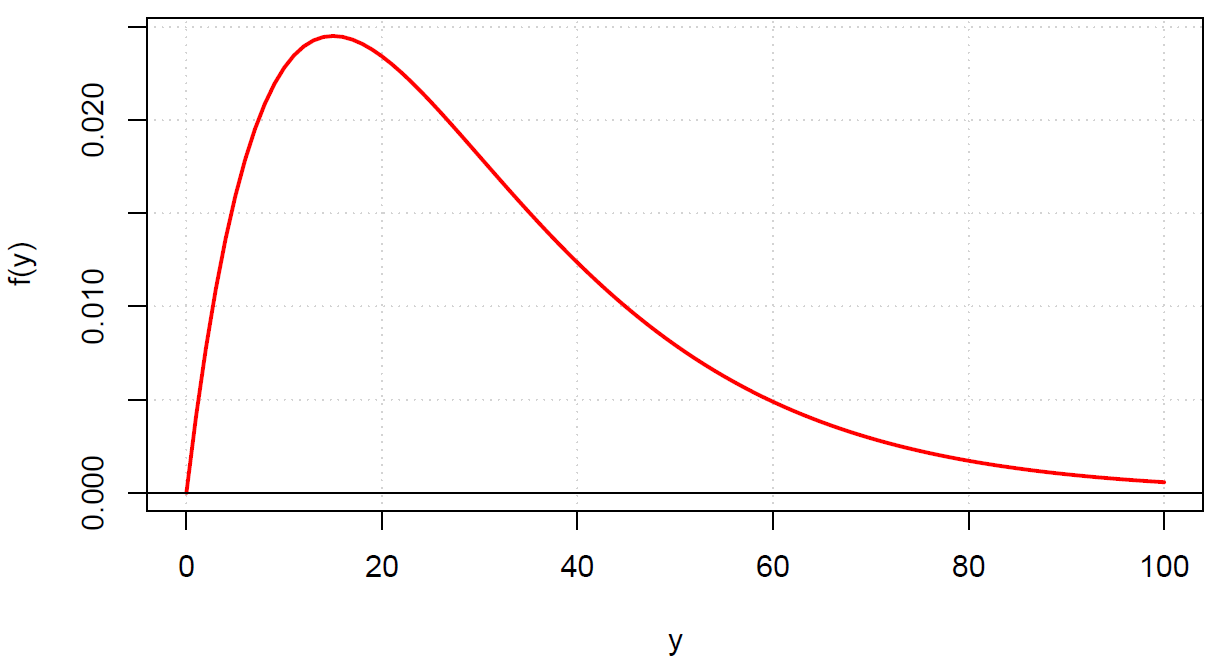
Examine values of μ and σ2 too.

Example: Distribution of lifetimes (lifetimes.R)

Let Y be a random variable denoting the lifetime in months of a particular type of animal in a population. Suppose the following gamma PDF is appropriate for these lifetimes.

 for y > 0 and f(y) = 0 otherwise

For this example, β = 15 and α = 2.



Questions:

* What are the mean and variance?   
    
  The mean and variance are E(Y) = αβ = 2×15 = 30 and Var(X) = 2×152 = 450. Thus, one would expect the animals to live 30 months on average for this population.
* What is the probability an animal in the population lives longer than 80 months?

The probability can be found from

P(Y > 80) = ****. Notice that integration by parts would be needed here.

> # P(Y > 80) = 1 - F(80)

> 1 - pgamma(q = 80, shape = 2, scale = 15)

[1] 0.03057702

* What is the median lifetime?

The value c needs to found such that the probability of living less than c years is 0.5. Then we could use

****

and solve for c.

> # Median lifetime

> qgamma(p = 0.5, shape = 2, scale = 15)

[1] 25.1752

* A sample of size 1,000 from a population characterized by this PDF can be simulated using the rgamma() function. Compare the sample summary measures to what we would expect based on the gamma PDF.

> # Sample

> set.seed(5627)

> y <- rgamma(n = 1000, shape = 2, scale = 15)

> head(y)

[1] 50.945624 23.781652 61.251459 32.560420

8.267238 11.028743

> mean(y)

[1] 30.45355

> var(y)

[1] 465.3751

> # Estimate of P(Y > 80)

> mean(y > 80)

[1] 0.03

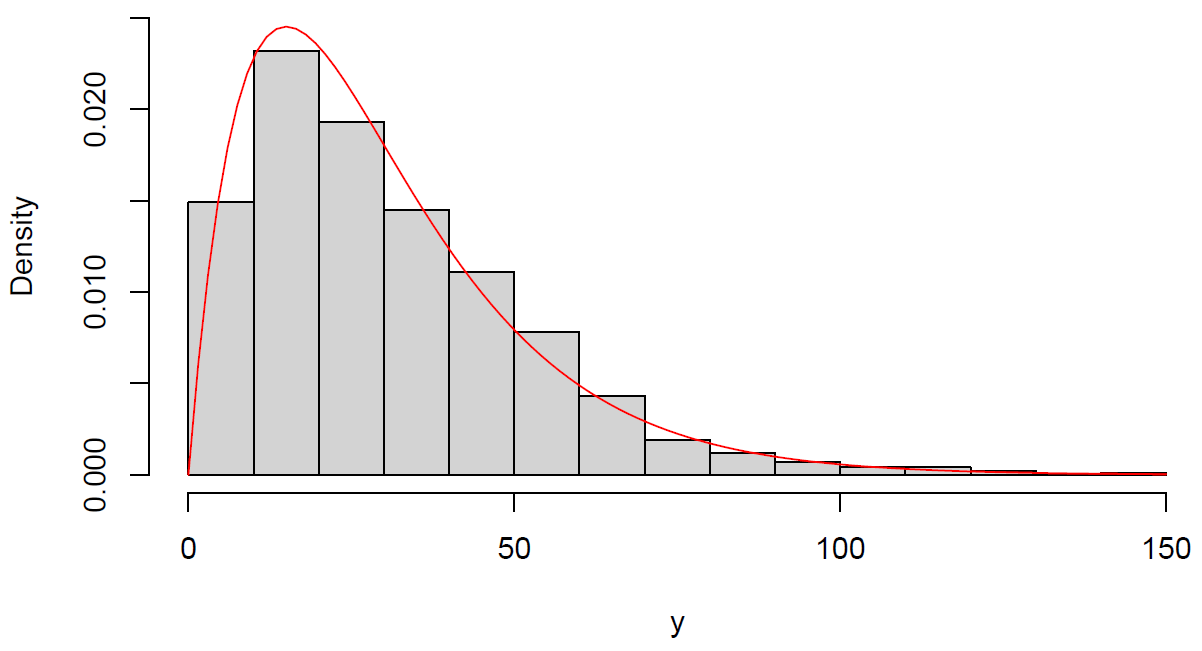
> # PDF

> hist(y, main = "", freq = FALSE, ylim =

c(0,0.025))

> curve(expr = dgamma(x = x, shape = 2, scale =

15), col = "red", add = TRUE)



> # CDF

> plot.ecdf(x = y, lwd = 2, panel.first = grid(),

ylab = "Probability", xlab = "y", col = "blue",

main = "")

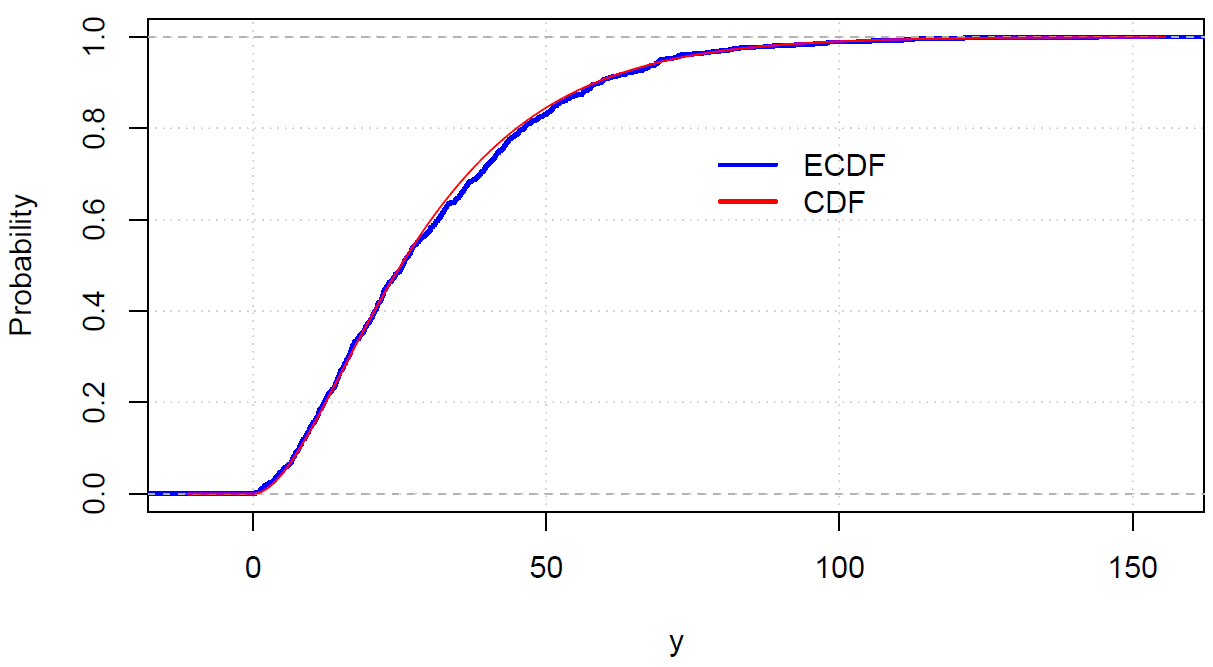
> curve(expr = pgamma(q = x, shape = 2, scale =

15), col = "red", add = TRUE, n = 1000)

> legend(x = 75, y = 0.8, legend = c("ECDF",

"CDF"),lty = 1, col = c("blue", "red"), lwd =

2, bty = "n")



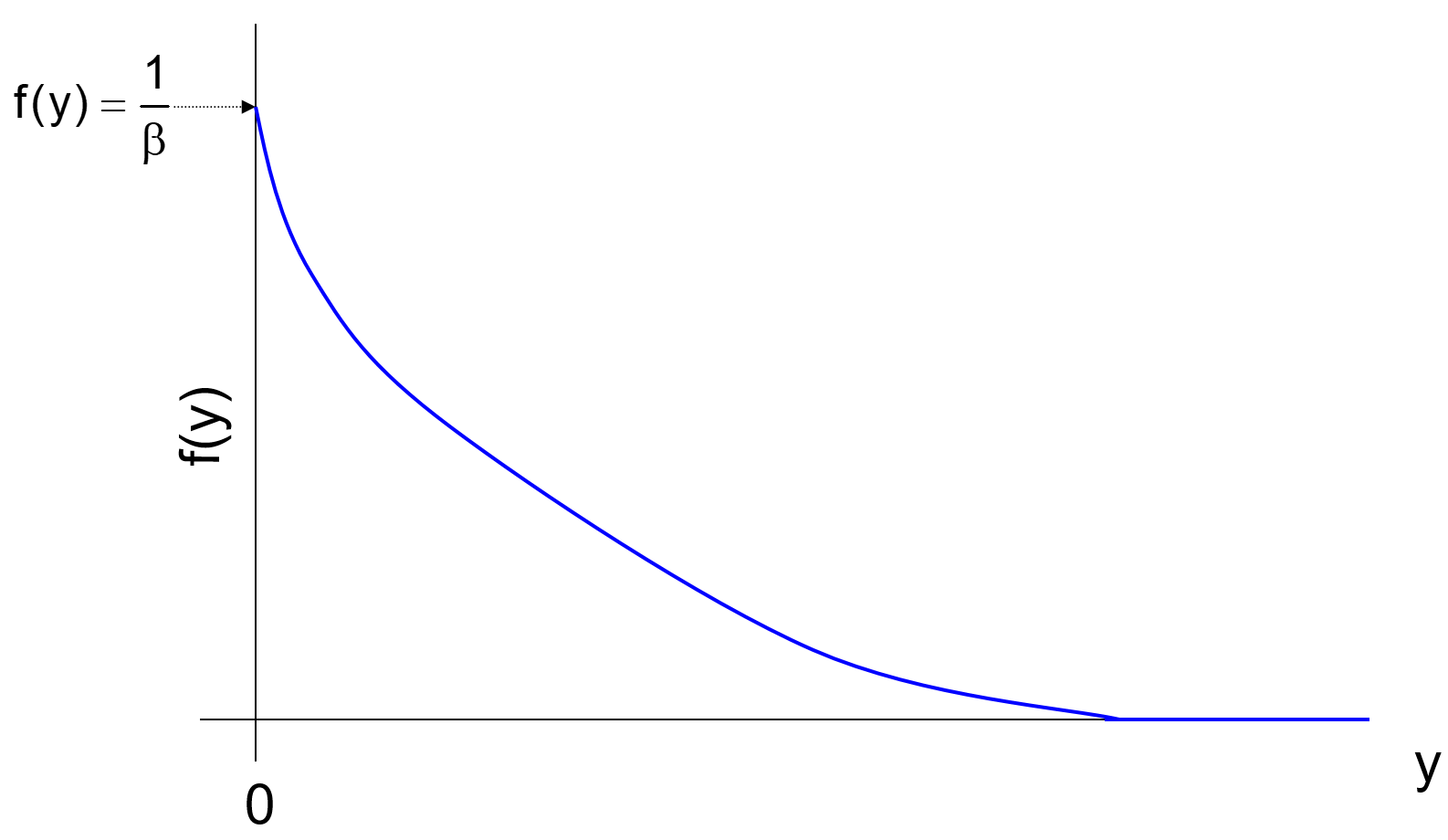
There are a few important special cases of the gamma PDF. One of them is the exponential PDF. For a random variable Y, this PDF is

****

where β > 0.

Notes:

* This is the gamma PDF with α = 1.
* β controls the scale of the PDF because it mostly influences the spread of the PDF.
* In general, this is what a plot of the PDF looks like.



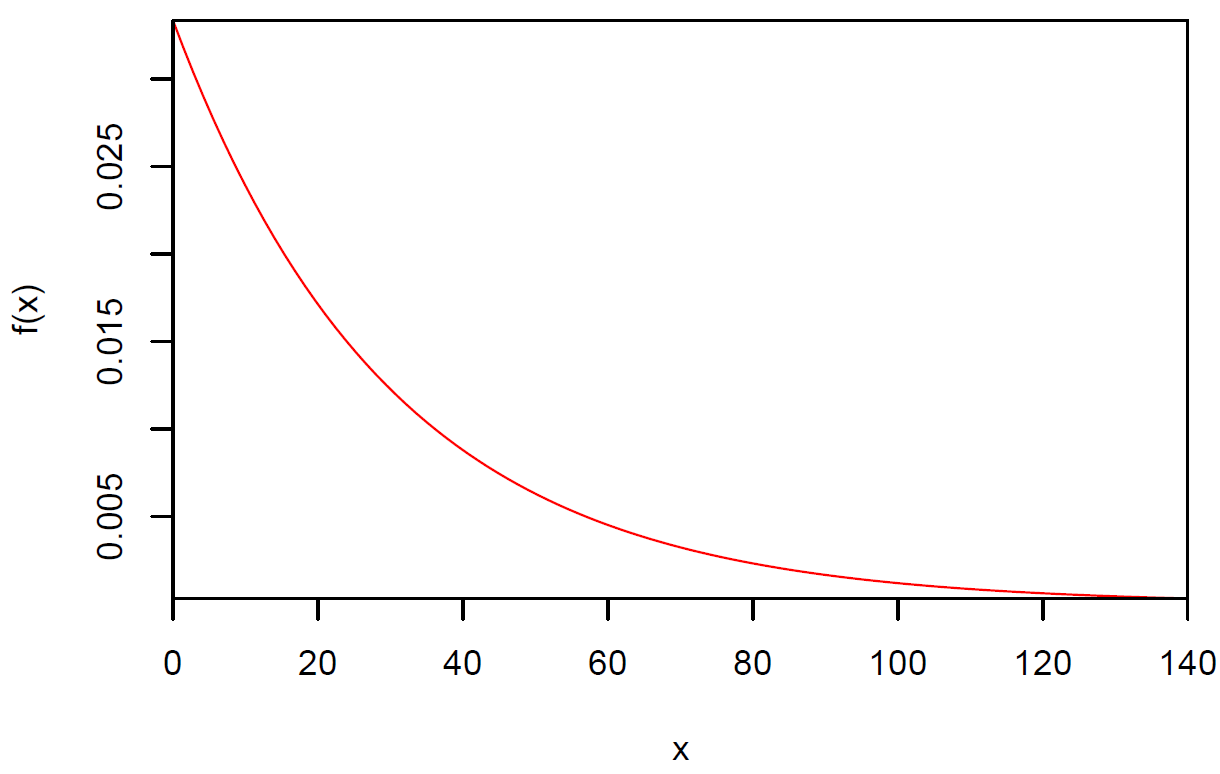
The height of the curve at a point yo is . Notice that when yo = 0,  because e0 = 1.

* The mean and variance for the exponential PDF are E(Y) = μ = β and Var(Y) = σ2 = β2.

Example: Transaction time

The number of seconds between transactions (e.g., purchases) on a website can be represented by a PDF. Let X be a random variables representing the seconds. Suppose the PDF for X is





Review this example from previous sections. While the gamma functions in R can be used here, there are also a set of exponential PDF functions too. For example, dexp() calculates f(y). Note that the PDF is parameterized a little different in R. These functions use 1/β for the rate argument.