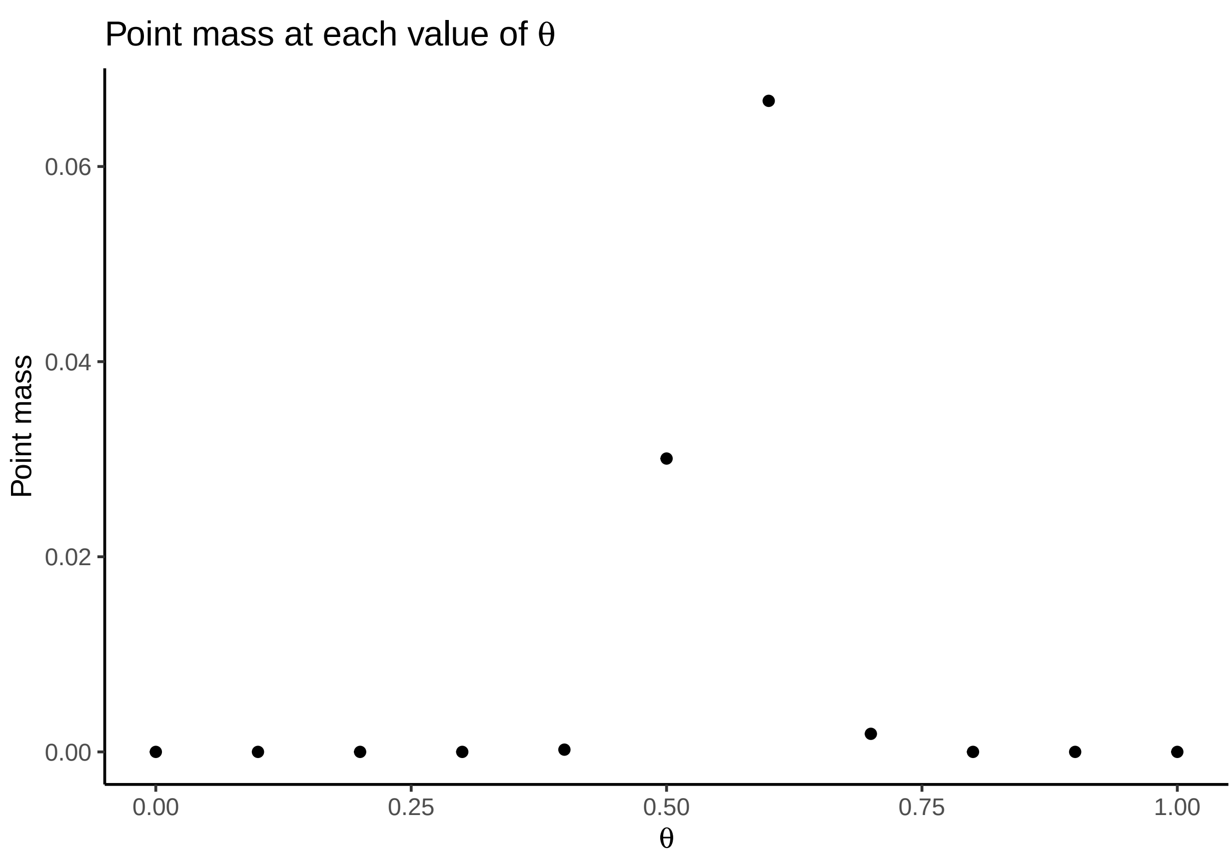
Other exercises 1

(b)

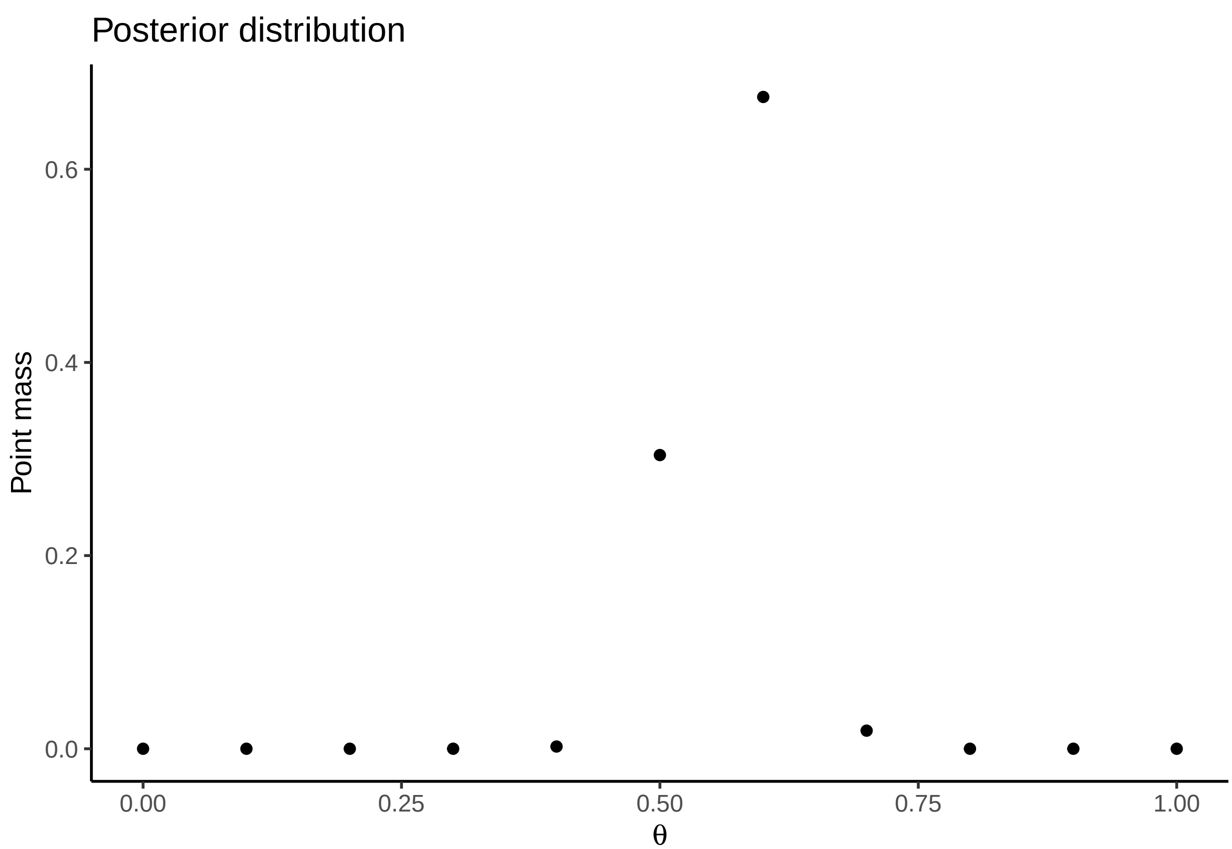
```
theta <- seq(0,1,0.1)
y <- 57
n <- 100

bin_fall <- function(theta) {
  choose(n,y)*theta^y*(1-theta)^(n-y)
}

results <- tibble(theta, bin = sapply(theta, bin_fall))

results %>%
  ggplot(aes(theta, bin)) +
  geom_point() +
  labs(x=TeX("$\\theta$"),
       y = "Point mass",
       title = TeX("Point mass at each value of $\\theta$")) +
  theme_classic()
```



(c)

Using Bayes rule we get that the posterior distribution for each theta is the point mass calculated in part (*b*) divided by the sum of the beta values. This gives

```
# sum up previous results
sum_bin <- results %>%
  select(bin) %>% sum()

bin_posterior <- function(bin) {bin/sum_bin}

results_posterior <- tibble(theta, bin_post = sapply(results$bin, bin_posterior))

results_posterior %>%
  ggplot(aes(theta, bin_post)) +
  geom_point() +
  theme_classic() +
  labs(x=TeX("$\\theta$"),
       y = "Point mass",
       title = TeX("Posterior distribution"))
```

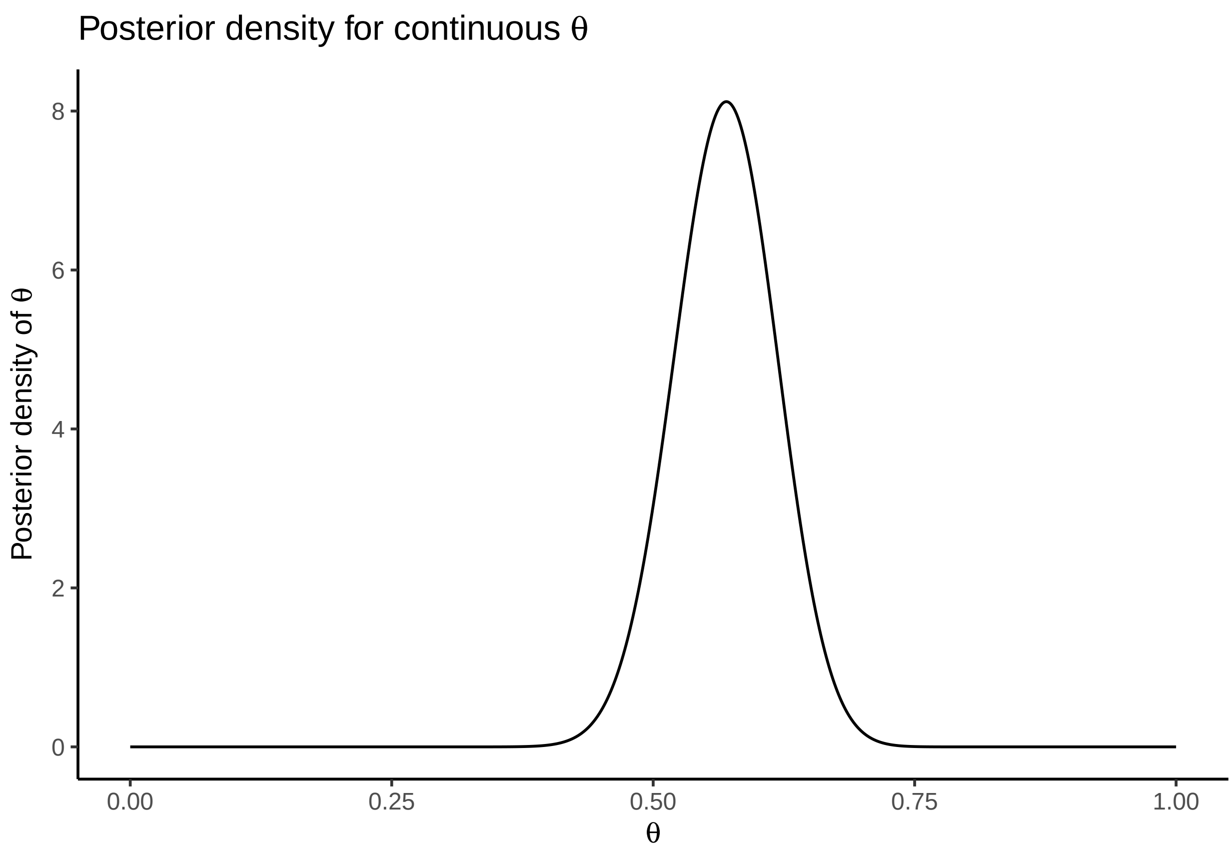


(d)

Now suppose you allow θ to be any value in the interval $[0, 1]$. Using the uniform prior density for θ, so that $p(\theta) = 1$, plot the posterior density of θ as a function of θ. According to the posterior density, what is the probability of $\theta > 0.5$?

The uniform prior density is a special case of the beta distribution which is conjugate for the class of binomial sampling distributions The uniform prior density can be set up as $\theta^{1-1}(1-\theta)^{1-1}$ and we had the binomial likelihood $\theta^y(1-\theta)^{n-y}$ Bayes theorem then gives the parameters for the beta distribution $\alpha = y + 1$ and $\beta = n - y + 1$

```r
# plot the beta distribution
theta <- seq(0, 1, 0.0001)

tibble(x = theta, y = dbeta(theta, shape1 = 57+1, shape2 = 100-57+1)) %>%
  ggplot(aes(x=x, y = y)) +
  geom_line() +
  theme_classic() +
  labs(x=TeX("$\\theta$"),
       y=TeX("Posterior density of $\\theta$"),
       title=TeX("Posterior density for continuous $\\theta$"))
```



According to the posterior density what is the prob. of $\theta > 0.5$?

```r
1-pbeta(0.5, 58, 44)
```

```
## [1] 0.9183604
```

(e)

Why are the heights of posterior densities in (c) and (d) not the same?

The continuous θ values reach 0.57, the maximum height, which the discrete θ does not, therefore it goes higher.

Other excercises 2

Random numbers, probability density functions (pdf) and cumulative density functions(cdf)

The goal of this exercise is to generate random numbers, plot the histogram, the empirical pdf and cdf for these numbers, and see how they compare to the theoretical pdf and cdf. The goal is also to compare the sample mean and standard deviation to the theoretical mean and standard deviation.

(a)

Generate $B = 2000$ numbers from the gamma distribution with parameters $\alpha = 3$, $\beta = 0.5$ (Matlab:gamrnd, R:rgamma). Compute the sample mean and the sample standard deviation and compare to the theoretical mean and standard deviation. (Matlab:mean,std, R:mean,sd).

```
set.seed(2699)

n <- 2000
alpha <- 3
beta <- 0.5

B <- rgamma(2000, shape = alpha, rate = 0.5)

# sample mean and sd
mean_sample <- mean(B)
sd_sample <- sd(B)

# theoretical mean and sd

mean_theo <- alpha/beta
sd_theo <- sqrt(alpha/beta^2)
```

Table 1: Comparison of means

Sample mean	Theoretical mean
5.954974	6

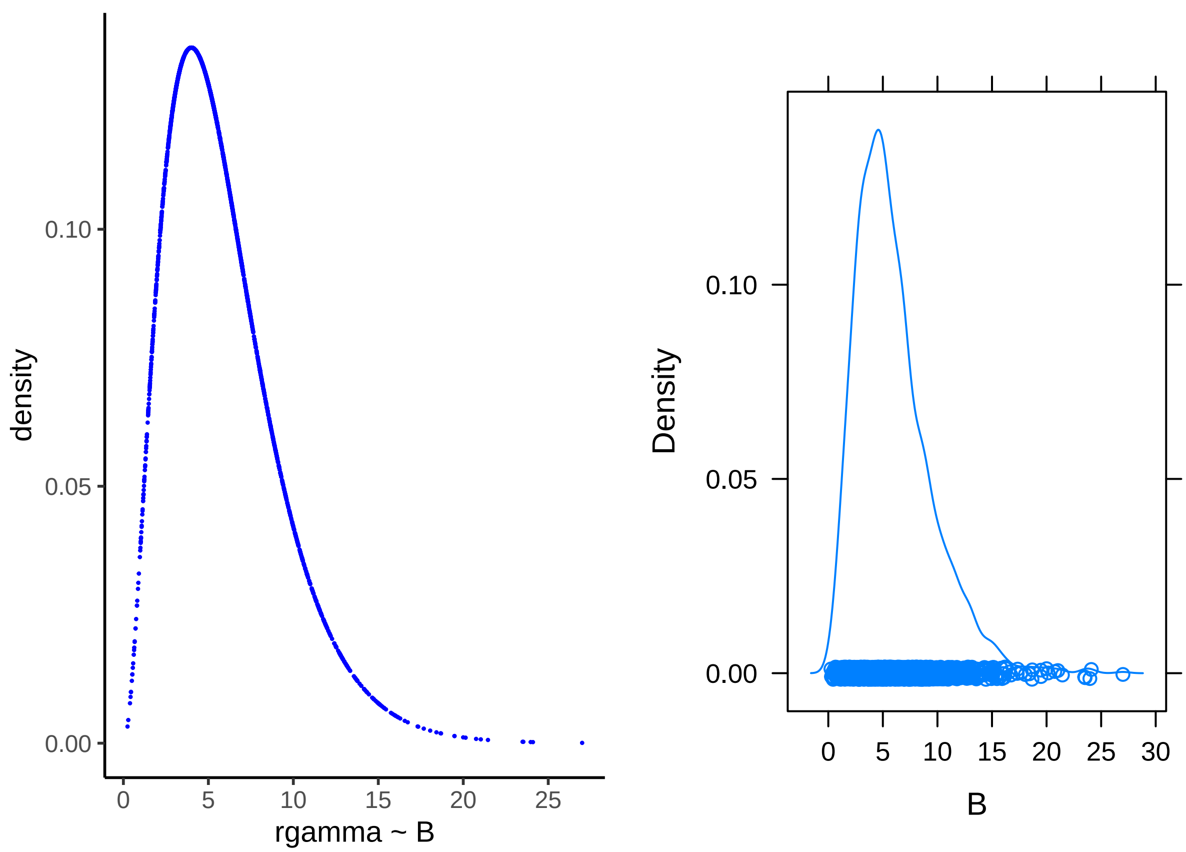
Table 2: Comparison of standard deviations

Sample sd	Theoretical sd
3.469081	3.464102

(b)

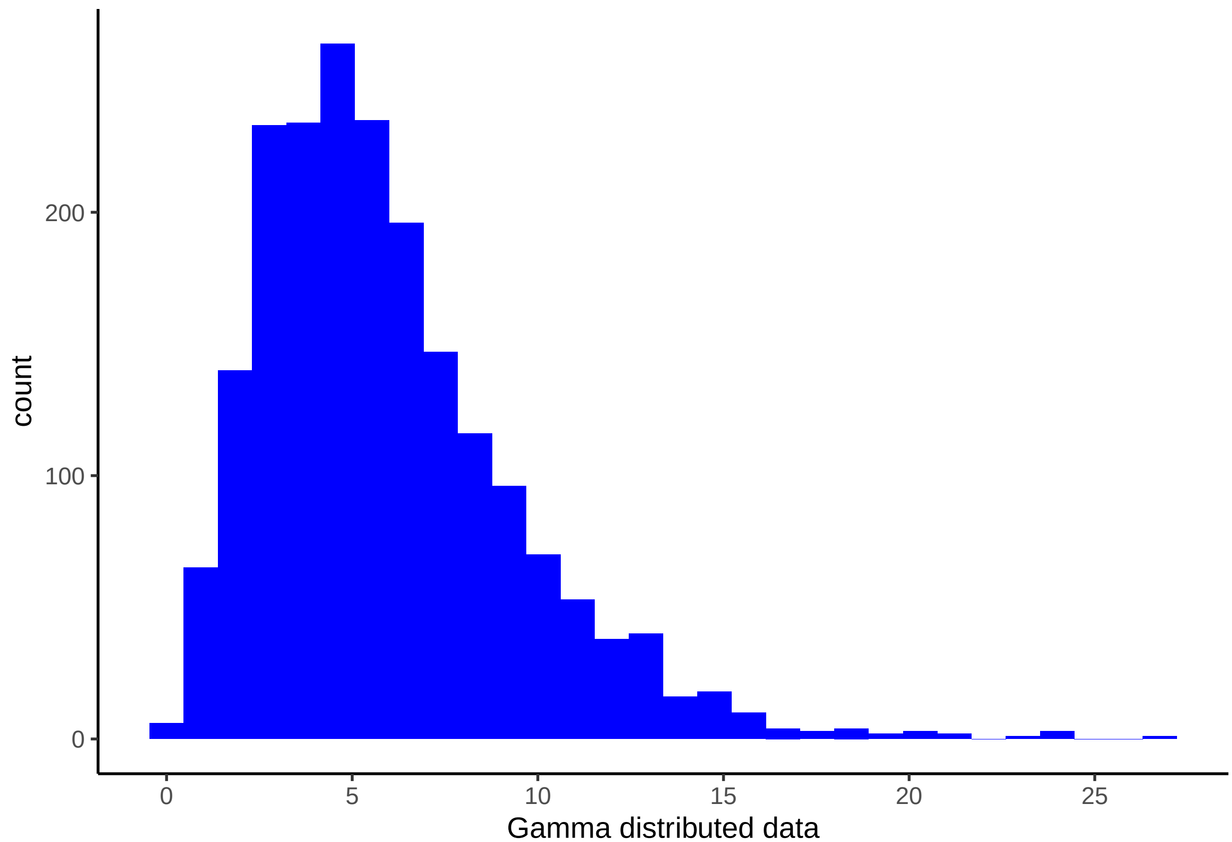
Plot the theoretical density (pdf) of the gamma distribution (Matlab:gampdf, R:dgamma). Plotempirical density based on the data on the same graph. In Matlabksdensitycan be used. In R one can use densityplot(y) given that y are the data. Plot the histogram of the data on another graph (Matlab:hist, R:histogram).

```
set.seed(2699)

# theoretical density and empirical density
plot <- tibble(x = B, y = dgamma(B,shape = 3, rate = 0.5))

plot %>%
  ggplot(aes(x=x, y = y)) +
  geom_point(size = 0.1, color = "blue") +
  theme_classic() +
  labs(x = "rgamma ~ B",
       y = "density") +
  scale_x_continuous(breaks = seq(0, 30, 5)) -> plot1

# empirical density based on the data
densityplot(B) -> plot2

plot_grid(plot1, plot2)
```



```r
# Plot the histogram of the data on another graph
plot %>%
  ggplot(aes(x=x)) +
  geom_histogram(fill = "blue") +
  theme_classic() +
  labs(x = "Gamma distributed data",
       y = "count") +
  scale_x_continuous(breaks = seq(0, 30, 5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

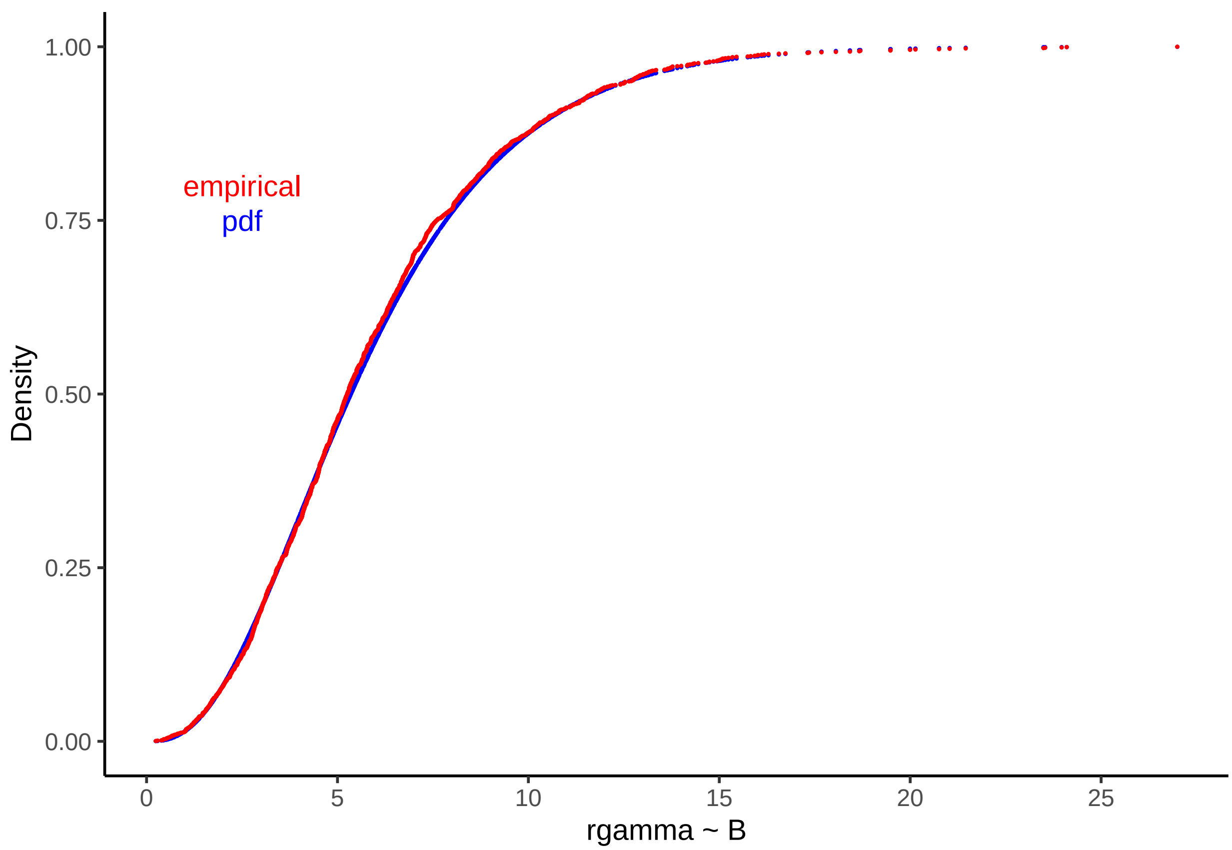


(c)

Plot the theoretical cumulative density function (cdf) of the gamma distribution (Matlab:gamcdf, R:pgamma). Plot empirical cumulative density based on the data on the same graph.

```r
set.seed(2699)

# heoretical cumulative density function (cdf) of the gamma distribution
plot <- tibble(x = B,
               y = pgamma(B,
                          shape = 3,
                          rate = 0.5))
# empirical density based on the data
n <- length(B)
B_sort <- sort(B)
```

```
emp_plot <- tibble(x=B_sort,
                   y=(1:n)/n)

plot %>%
  ggplot(aes(x=x, y = y)) +
  geom_point(size = 0.1,
             color = "blue") +
  theme_classic() +
  labs(x = "rgamma ~ B",
       y = "Density") +
  scale_x_continuous(breaks = seq(0, 30, 5)) +
  geom_point(data = emp_plot,
             aes(x=x,
                 y=y),
             color = "red",
             size = 0.1) +
  annotate(geom="text", x=2.5, y=0.8, label="empirical", color="red") +
  annotate(geom="text", x=2.5, y=0.75, label="pdf", color="blue")
```



rgamma produces a well gamma distrubuted sample.