Cairo University
Faculty of Economics and Political Science
Statistics Department

Beta Distribution on "R"

A Probability Project Report

Second Level - Arabic Section
Group (2)
Spring 2024

Students:
Mohamed Amir Mohamed El Samahy
Ezz Eldin Ahmed Mohamed Fatouh
Abdelrahman Ahmed Youssef
Abdelrahman Mostafa Kamel

Under Supervision of:
Dr. Reda Mazloum
Dr. Abeer Abdelaal

Introduction:

As time goes on, it becomes increasingly more important to utilize technology to its full extent. As statisticians, this means learning how to use programs, packages and programming languages -alongside knowing the theory behind all statistics-. So, this project aims to:

I. Improve the students' computer skills in general, and working with R specifically (which is one of the more important programming languages for a statistician).
II. Improve the students' report writing skills.
III. Train students on teamwork through working in groups.
IV. Make the students aware of the differences and similarities of theoretical and observational results.
V. Illustrate the differences in the P.D.F of the distribution as the parameters change.

The distribution showcased in this report is the ***Beta distribution***.

The Beta distribution is a family of continuous probability distributions set on the interval [0, 1] having two positive shape parameters, expressed by α and β.[1]

It that can be used to represent proportion or probability outcomes. Some examples would be[2]:

- The likelihood of the audience rating the new movie release.
- To find how likely it is that your preferred candidate for mayor will receive x% of the vote.
- The click-through rate of the website, which is the proportion of visitors.

The work organization:

- All students worked on the R code, the introduction and general comments.
- Student Mohamed Amir (a= 2, b=5) worked on coordinating the group's efforts and his part with the assigned parameters (Pages: 3,4).
- Student Ezz Eldin Ahmed (a=3, b=3) worked on formatting, checking and correcting the report as well as his part with the assigned parameters (Pages: 5,6).
- Student Abdelrahman Ahmed (a=9, b=9) provided the theoretical computations alongside his part of the project with the assigned parameters (Pages: 7,8).
- Student Abdelrahman Mostafa (a=3, b=1) also provided the theoretical calculations alongside his part of the project with the assigned parameters (Pages: 9,10).

[1] Beta Distribution - Definition, Formulas, Properties, Applications (byjus.com)

[2] Beta Distribution - Examples, Formula and Applications (vedantu.com)

Characteristics of any distribution:

Characteristic	Formula
Mean	$E(X) = \int_{-\infty}^{\infty} x \ f_X(x)\,dx\ ,\ \ x \in \mathbb{R}$
Expected Value of X^2	$E(X^2) = \int_{-\infty}^{\infty} x^2 \ f_X(x)\,dx\ ,\ \ x \in \mathbb{R}$
Variance	$Var(X) = E(X^2) - [E(X)]^2$
Coefficient of skewness	$\alpha = \frac{E(X-\mu)^3}{(E(X-\mu)^2)^{\frac{3}{2}}} = \frac{\mu_3}{\sigma^3}$
Coefficient of kurtosis	$\beta = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{\mu_4}{\sigma^4}$

Then, the characteristics of the beta distribution:

Characteristic	Formula
Mean	$E(X) = \int_0^1 x \ \frac{1}{B(a,b)} \ x^{a-1}\,(1-x)^{b-1}dx\ ,(0 < x < 1\ , a, b > 0) = \frac{a}{a+b}$
Variance	$Var(X) = \left[\int_0^1 x^2 \ \frac{1}{B(a,b)} \ x^{a-1}\,(1-x)^{b-1}dx\ \right] - \left[\int_0^1 x \ \frac{1}{B(a,b)} \ x^{a-1}\,(1-x)^{b-1}dx\ \right]^2,$ $(0 < x < 1\ , a, b > 0) = \frac{ab}{(a+b)^2(a+b+1)}$
Coefficient of skewness	$\alpha = \frac{E(X-\mu)^3}{(E(X-\mu)^2)^{\frac{3}{2}}} = \frac{\mu_3}{\sigma^3} = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$
Coefficient of kurtosis	$\beta = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{\mu_4}{\sigma^4} = \frac{3(a+b+1)[ab(a+b-6)+2(a+b)^2]}{ab(a+b+2)(a+b+3)}$

The probability density function is given by:

$$f(x) = \begin{cases} \frac{1}{\beta\,(a,b)} x^{a-1}(1-x)^{b-1}, & 0 < x < 1\ ,\ a, b > 0 \\ 0 & ,\ o.w \end{cases}$$

Another Form:

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, & 0 < x < 1\ ,\ a, b > 0 \\ 0 & ,\ \ o.w \end{cases}$$

The data was generated using Beta distribution parameters: β(a = 2, b= 5),

Sample Size = 1000. (See Appendix for the code used to generate the data)

$$Probability\ Density\ Function: f(x) = \begin{cases} \dfrac{1}{\beta\ (2{,}5)} x^{2-1}(1-x)^{5-1}, & 0 < x < 1\ ,\ \ a, b > 0 \\ 0 & ,\qquad o.w \end{cases}$$

The theoretical mean, variance, coefficients of skewness and kurtosis of the Beta distribution

$\mu = \frac{\alpha}{\alpha+\beta}$ $\qquad\qquad$ $\mu = \frac{2}{2+5} = \frac{2}{7}$

$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\qquad\qquad$ $\sigma^2 = \frac{2\cdot 5}{(2+5)^2(2+5+1)} = \frac{10}{395} \approx 0.025$

$$coefficient\ of\ skewness(\alpha) = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$$

$$coefficient\ of\ skewness(\alpha) = \frac{2(5-2)\sqrt{2+5+1}}{(2+5+2)\sqrt{2\cdot 5}} = \frac{6\sqrt{8}}{9\sqrt{10}} \approx 0.63$$

$$coefficient\ of\ kurtosis(\beta) = \frac{3(a+b+1)[ab(a+b-6)+2(a+b)^2]}{ab(a+b+2)(a+b+3)}$$

$$coefficient\ of\ kurtosis(\beta) = \frac{3(2+5+1)[(2)(5)(2+5-6)+2(2+5)^2]}{(2)(5)(2+5+2)(2+5+3)} = 2.88$$

Observed mean, variance, coefficient of skewness and the coefficient of kurtosis of sample size (n =1000)

$$\mu = 0.284097671368828$$

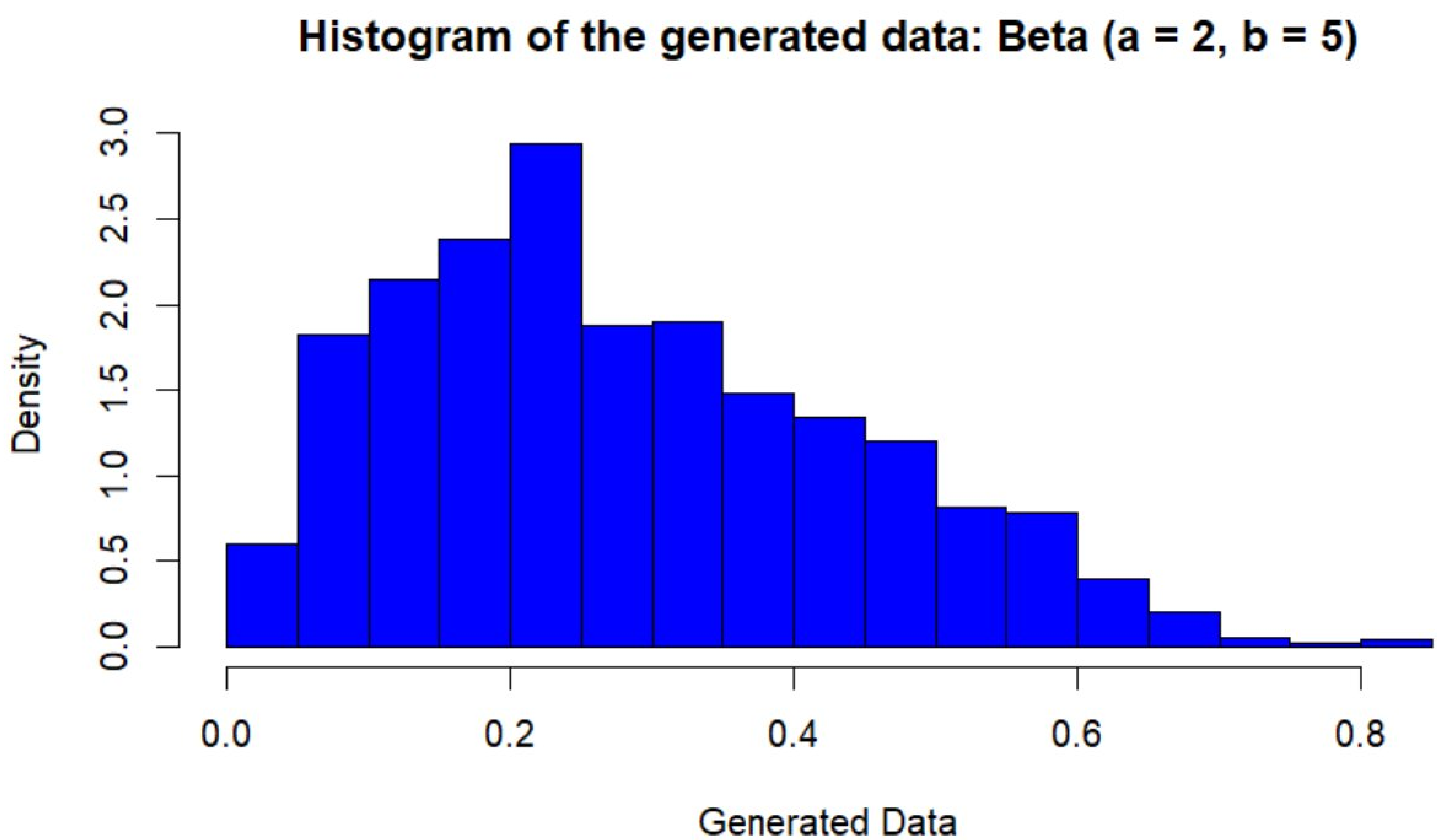
$$\sigma^2 = 0.0258867083152307$$

$$coefficient\ of\ skewness(\alpha) = 0.551445475085262$$

$$coefficient\ of\ kurtosis(\beta) = 2.64028857831321$$

Comment:

- The mean in the theoretical and the observed are so close as the theoretical is bigger than the observed by (0.0016166143454577)
- The variance in the theoretical and the observed are so close as the theoretical is less than the observed by (≈ 0.003)
- The coefficient of skewness (α) in the theoretical and the observed, the theoretical is greater than the observed by (≈ 0.1)
- The coefficient of kurtosis (β) in the theoretical and the observed, the theoretical is greater than the observed by (0.2397114217)



The data was generated from the Beta distribution with parameters: B(a = 3, b=3). With a Sample Size = 1000. (see Appendix for the code used to generate the data)

$$Probability\ Density\ Function: f(x) = \begin{cases} \frac{1}{\beta\ (3,3)} x^{3-1}(1-x)^{3-1}, & 0 < x < 1\ ,\ a, b > 0 \\ 0 & ,\quad o.w \end{cases}$$

The theoretical mean, variance, coefficients of skewness and kurtosis of the Beta distribution are obtained as follows:

Theoretical mean (E(X)) = $\frac{a}{a+b} = \frac{3}{3+3} = \frac{3}{6} = \frac{1}{2} = 0.5$

Theoretical variance (V(X)) = $\frac{ab}{(a+b)^2(a+b+1)} = \frac{(3)(3)}{(3+3)^2(3+3+1)} = \frac{9}{(6)^2(7)} = \frac{9}{252} \approx 0.0357$

Theoretical coefficient of skewness (α) $= \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}} = \frac{2(3-3)\sqrt{3+3+1}}{(3+3+2)\sqrt{(3)(3)}} = \frac{2(0)\sqrt{7}}{(8)(9)} = 0$

Theoretical coefficient of kurtosis (β) $= \frac{3(a+b+1)[ab(a+b-6)+2(a+b)^2]}{ab(a+b+2)(a+b+3)} =$

$\frac{3(3+3+1)(3)(3)(3+3-6)+2(3+3)^2]}{(3)(3)(3+3+2)(3+3+3)} = 2.33$

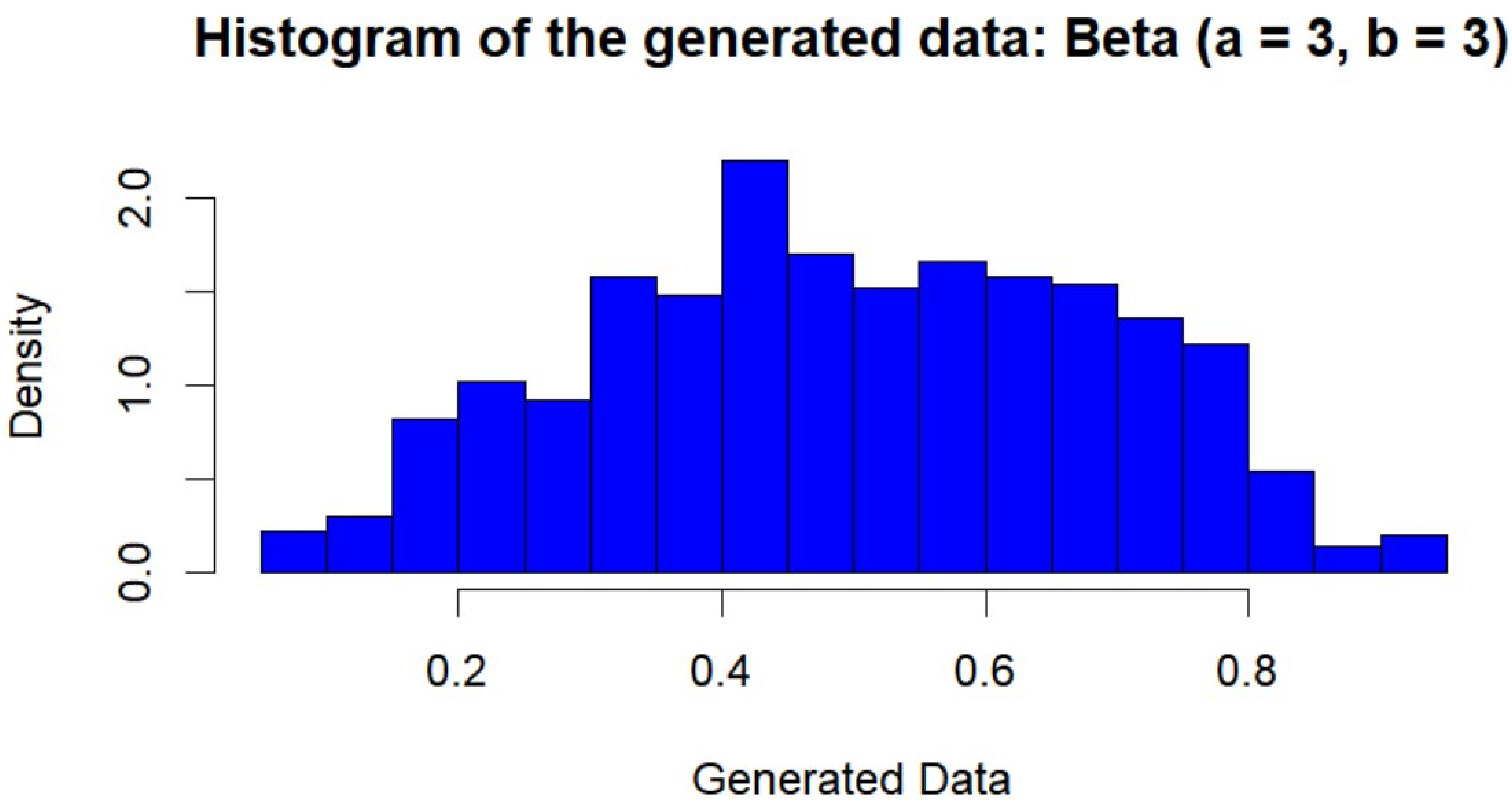
Results:

- The observations mean was equal to 0.4978796, which is close to the previously calculated mean (0.5). (about 0.0021 difference)
- The observations variance was equal to 0.03622391, which was also close to the calculated variance (0.0357). (about 0.000524 difference)
- While the calculated coeff. of skewness was (0), the observations coeff. of skewness was (-0.05667316). (about 0.05667316 difference)
- Finally, the calculated coeff. of kurtosis was (2.33), however, the observations coeff. of kurtosis was around (2.214944). (0.115056 difference)

Conclusion:

1. The observations both mean and variance were very close to their theoretical counterpart. As the sample size got bigger, the observations mean and variance got closer and closer to 0.5 and 0.0357 respectively.

2. The theoretical coeff. of skewness (0) implies that the data is symmetrical, but the value of the observations coeff. of skewness was (-0.05667316). So, it's a little skewed to the left and not perfectly symmetrical. Once again, as the sample size got bigger, the coeff. of skewness got closer to 0.
3. The coeff. of kurtosis was -as the theoretical calculation resulted- a value of 2.33, while the observations result was 2.214944. Both values are under 3 (which is the value of the normal dist.'s beta coefficient), making the distribution "Platykurtic" which means "Less peaked around the center".



X has a beta distribution with (a=9, b=9) **(see Appendix for the code used to generate data)**

$$Probability\ Density\ Function: f(x) = \begin{cases} 218790\, x^8\, (1-x)^8 & ,\ 0 < x < 1\ ,\ \ a, b > 0 \\ 0 & ,\ \ o.w \end{cases}$$

The Theoretical mean, variance, coefficient of skewness and the coefficient of kurtosis.

- **Mean**

$$E(X)= \frac{a}{a+b} = \frac{9}{9+9} = 0.5$$

- **Variance**

$$V(X)= \frac{ab}{(a+b+1)*(a+b)^2} = \frac{9*9}{(9+9+1)*(9+9)^2} = 0.0131578$$

- **Coefficient of skewness**

$$\alpha = \frac{E(X-\mu)^3}{(E(X-\mu)^2)^{\frac{3}{2}}} = \frac{0}{(0.0131578)^{\frac{3}{2}}} = 0$$

- **Coefficient of kurtosis**

$$\beta = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{0.000469}{(0.0131578)^2} = 2.7142857$$

The mean, variance, coefficient of skewness and the coefficient of kurtosis computed by (R) with sample size {n =1000}

- **Mean**

E(X) = 0.5005589

- **Variance**

V(X)= 0.01289419

- **Coefficient of skewness**

α = -0.03569049

- **Coefficient of kurtosis**

β = 2.596176

Comment:

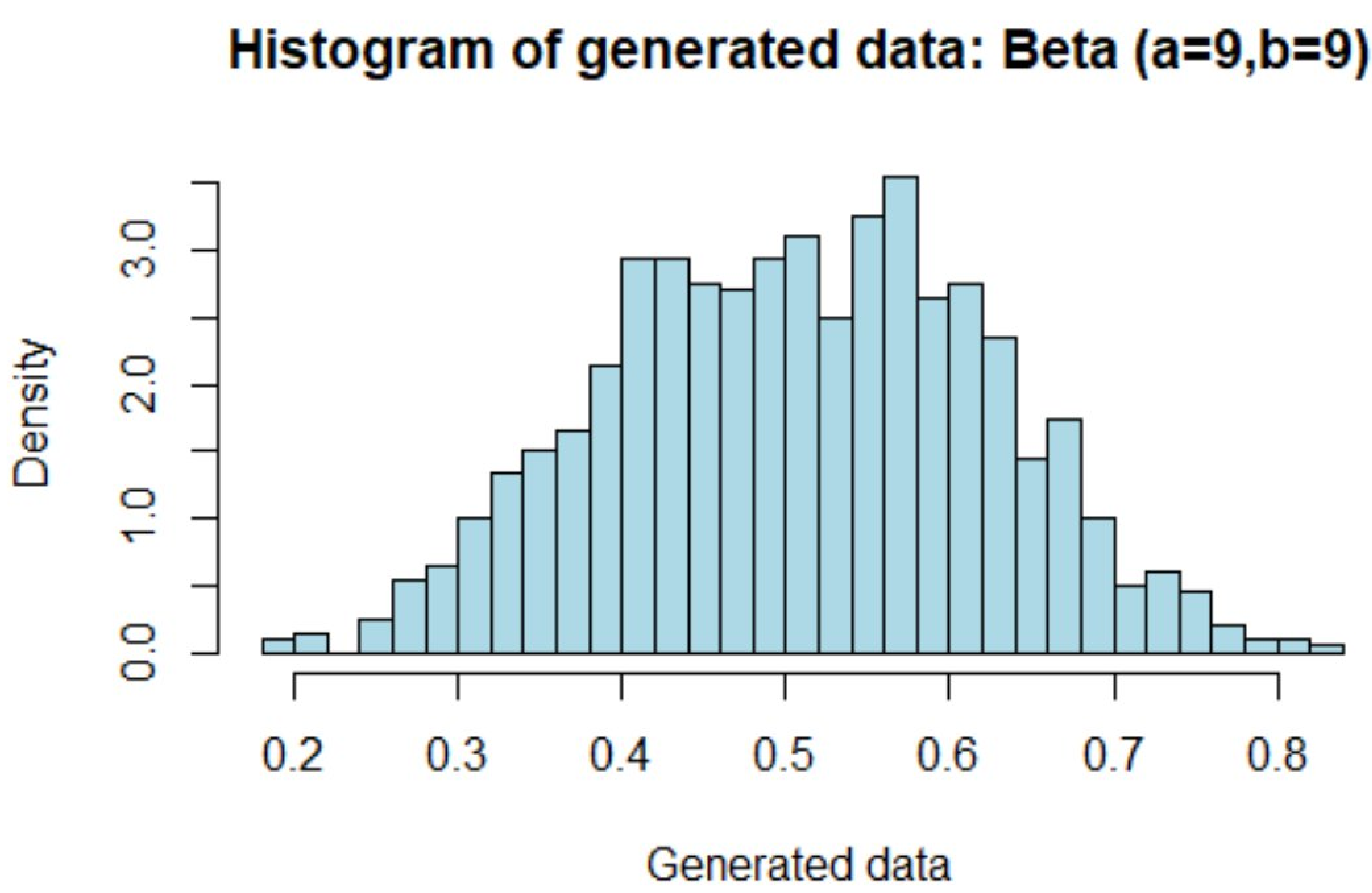
- The difference between the observations mean and the theoretical mean is (0.0005598) “so close”

- The difference between the observations variance and the theoretical variance is (-0.00026361) “so close”
- The difference between the observations Coefficient of skewness and the theoretical Coefficient of skewness is (-0.03569049) “the same as the observational one because the theoretical =0”
- The difference between the observations Coefficient of kurtosis and the theoretical Coefficient of kurtosis is (-0.1181097) “so close”

Conclusion:

1. The mean and the variance in the observations are so close to the theoretical because when we take a big observation the result become closer to the theoretical one in our case {n=1000}.
2. In the observation the Coefficient of skewness = -0.03569049 “skewed to the left” and theoretical= 0 “symmetric”.
3. In the observation the Coefficient of kurtosis = 2.596176 “less peaked than the normal curve” and theoretical=2.7142857 “less peaked than the normal curve”.

Graph of the generated data:



X ~β (3,1) with sample size = 1000 (see Appendix for the code used for the generated data)

$$Probability\ Density\ Function: f(x) = \begin{cases} \frac{1}{\beta\ (3,1)} x^{3-1}(1-x)^{1-1}, & 0 < x < 1\ ,\ a, b > 0 \\ 0 & ,\quad o.w \end{cases}$$

Theoretical part

Mean $= \frac{a}{a+b} = \frac{3}{3+1} = \frac{3}{4} = 0.75$

Variance $= \frac{ab}{(a+b)^2(a+b+1)} = \frac{3}{(3+1)^2(3+4+1)} = \frac{3}{128} = 0.0234375$

Before calculating the coefficients of skewness and kurtosis, we need a general form for raw moments:

For β(k+a, b), E(X^k)= $\frac{\Gamma k+a.\Gamma a+b}{\Gamma k+a+b.\Gamma a}$
then, using the relationship between central and raw moments:

- When k = 2, raw and central moments
  - $\mu_2 = \mu'_2 - \mu'^2_1$
- When k = 3, raw and central moments
  - $\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1$
- When k = 4, raw and central moments
  - $\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$

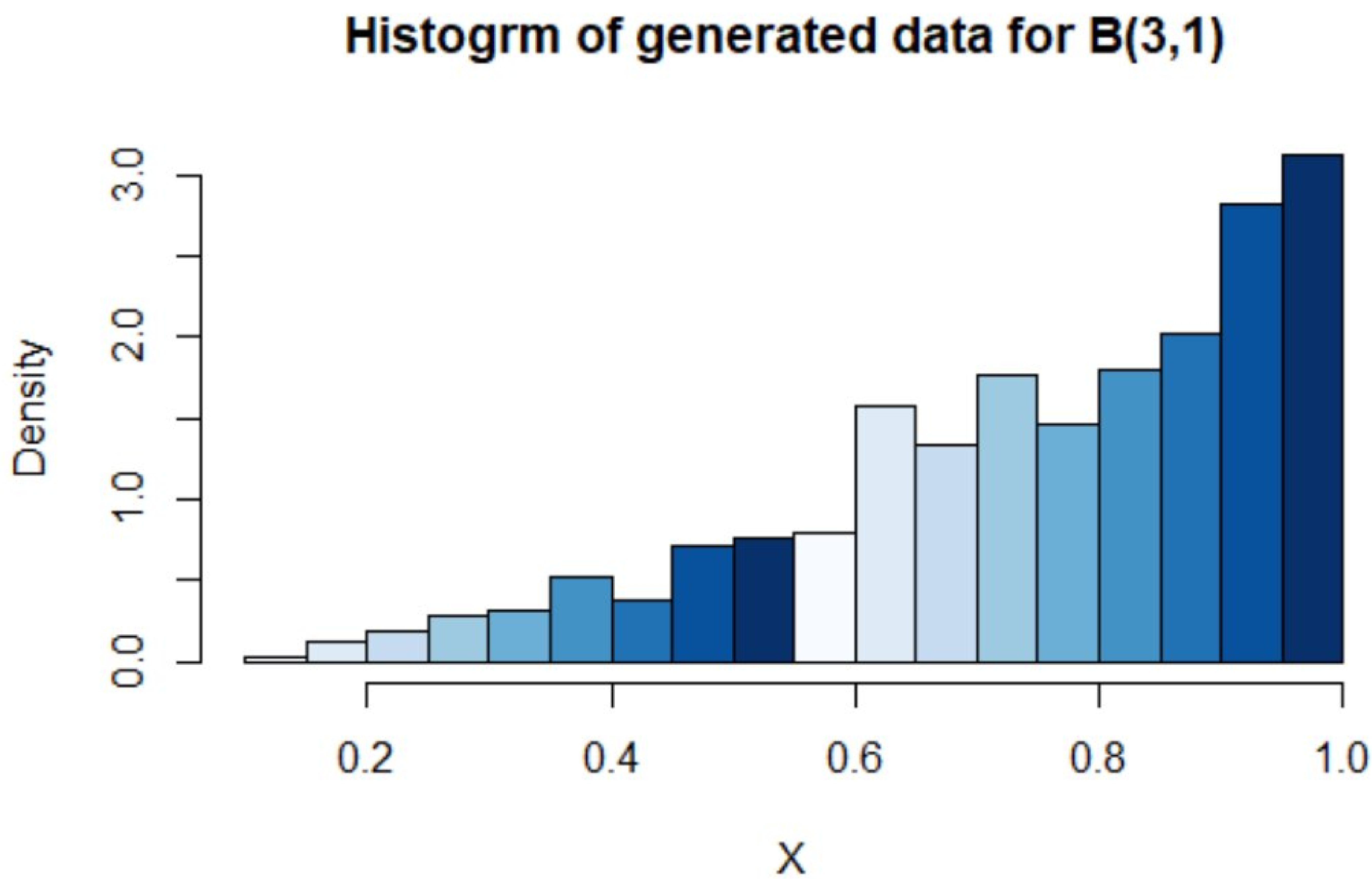
α (the coefficient of skewness) $= \frac{E(x-\mu)^3}{(E(x-\mu)^2)^{\frac{3}{2}}} = \frac{\mu^3}{\sigma^3} = \frac{-0.00625}{(0.0375)^{\frac{3}{2}}}$ = -0.860

β (the coefficient of kurtosis) $= \frac{E(x-\mu)^4}{(E(x-\mu)^2)^2} = \frac{\mu^4}{\sigma^4} = \frac{0.0037}{(0.0375)^2} = 2.631$

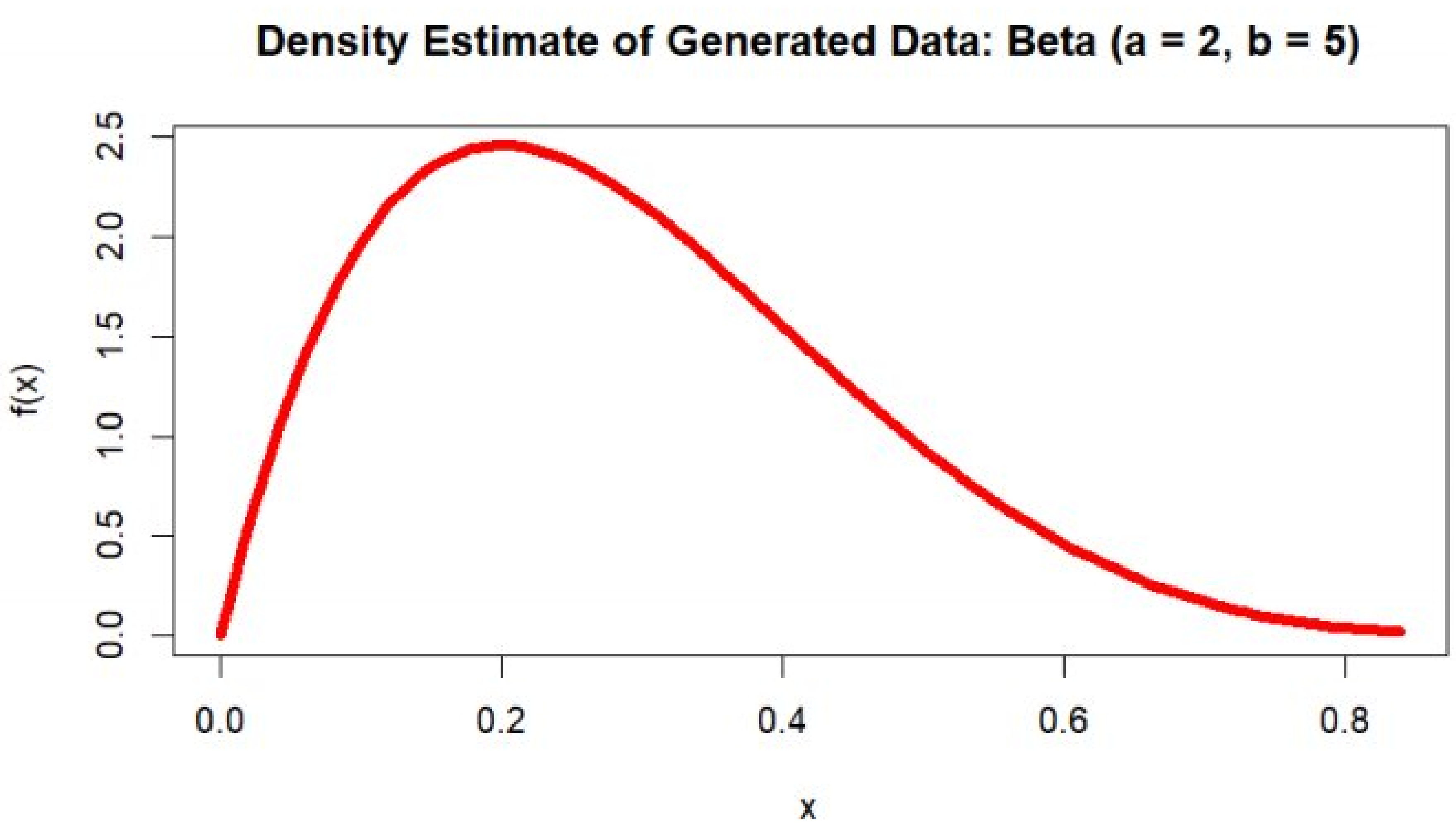
Applied part

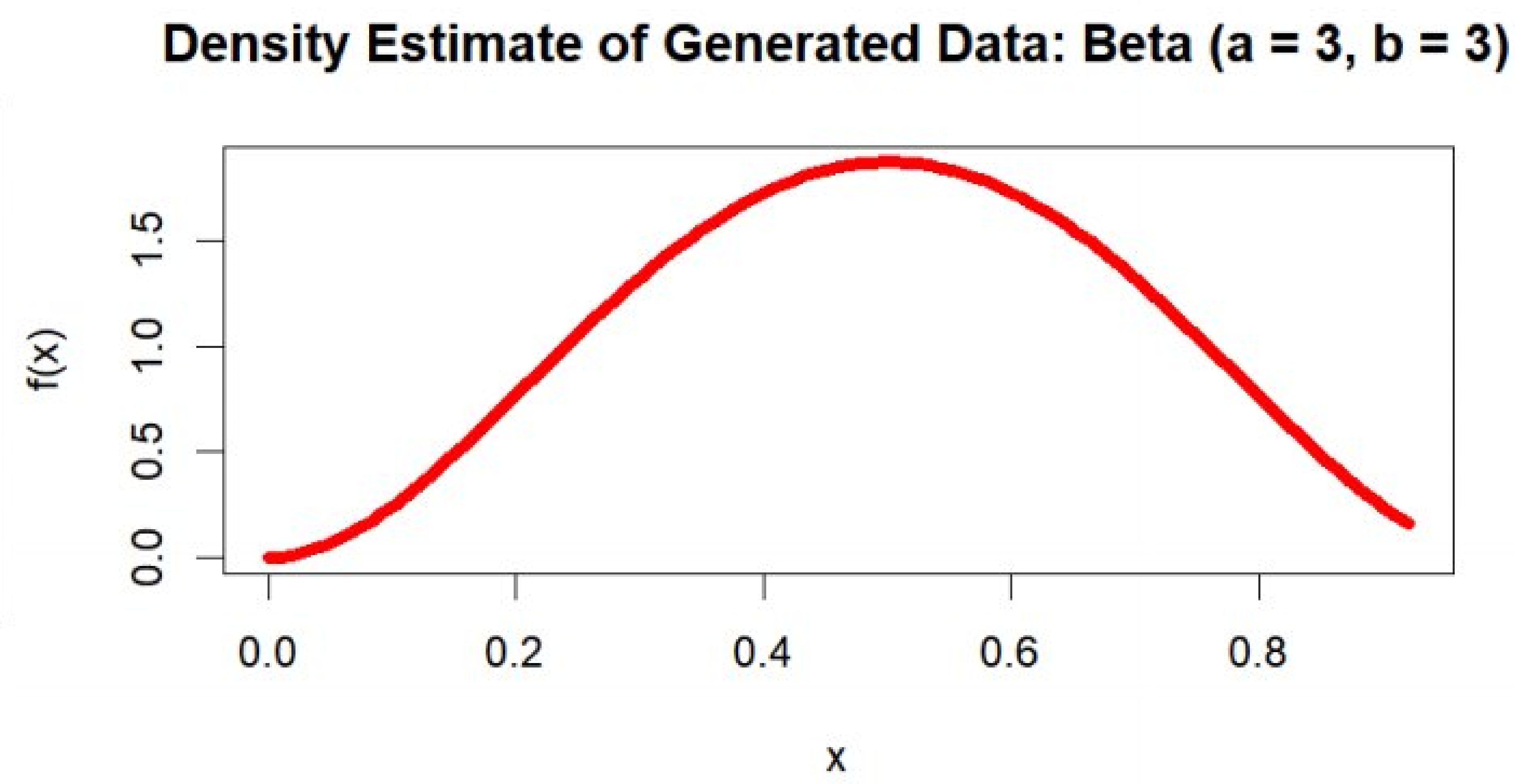
- Observed mean: 0.756035 > theoretical mean: 0.75
- Observed variance: 0.038466> theoretical variance: 0.0234375
- Observed coefficient of kurtosis: 2.9179> theoretical kurtosis: 2.6311
- Observed coefficient of skewness: -0.816> theoretical skewness: -0.860

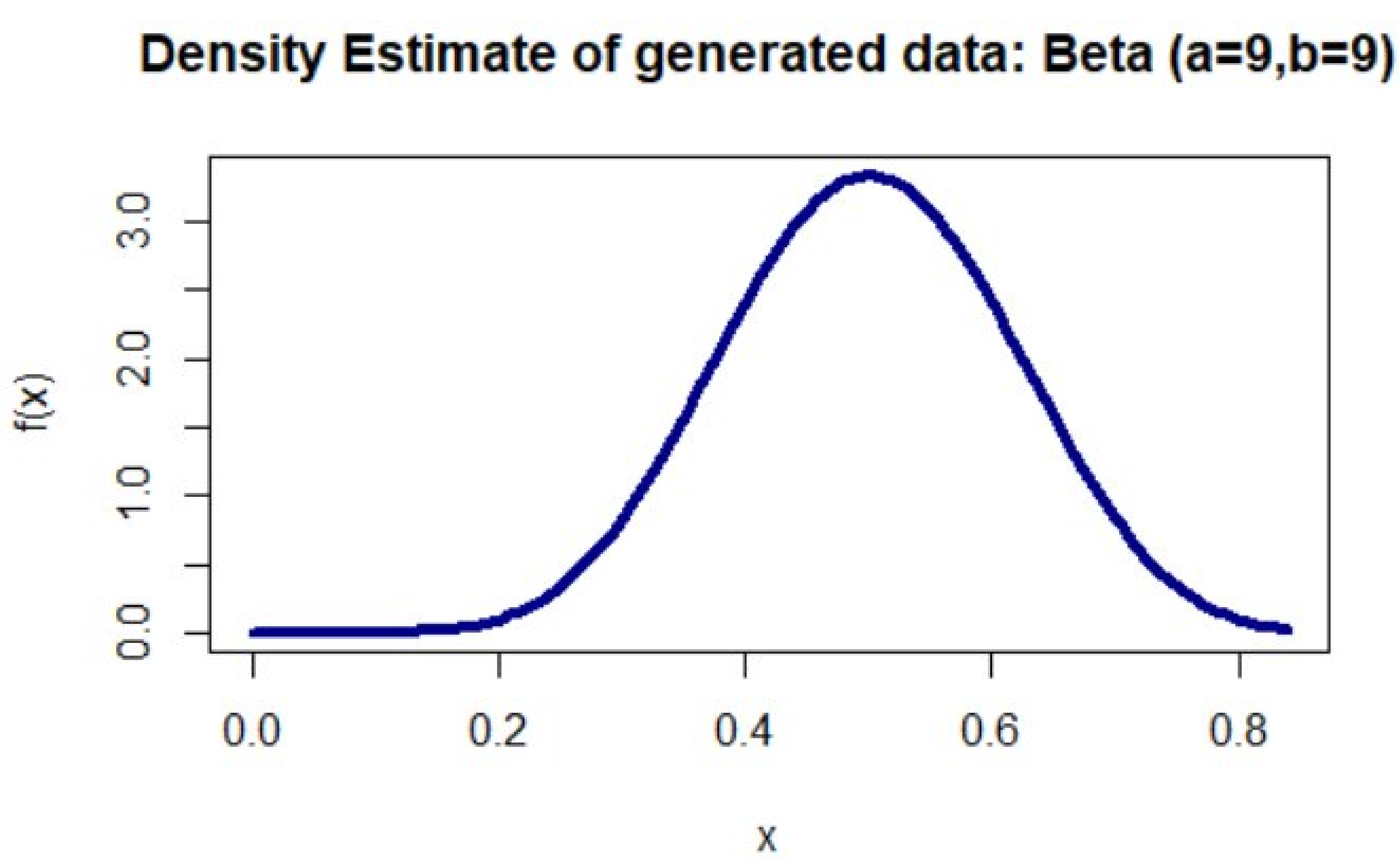
It is obvious that the difference between observed characteristics and theoretical characteristics that when number of observations get bigger the difference will gradually disappear.

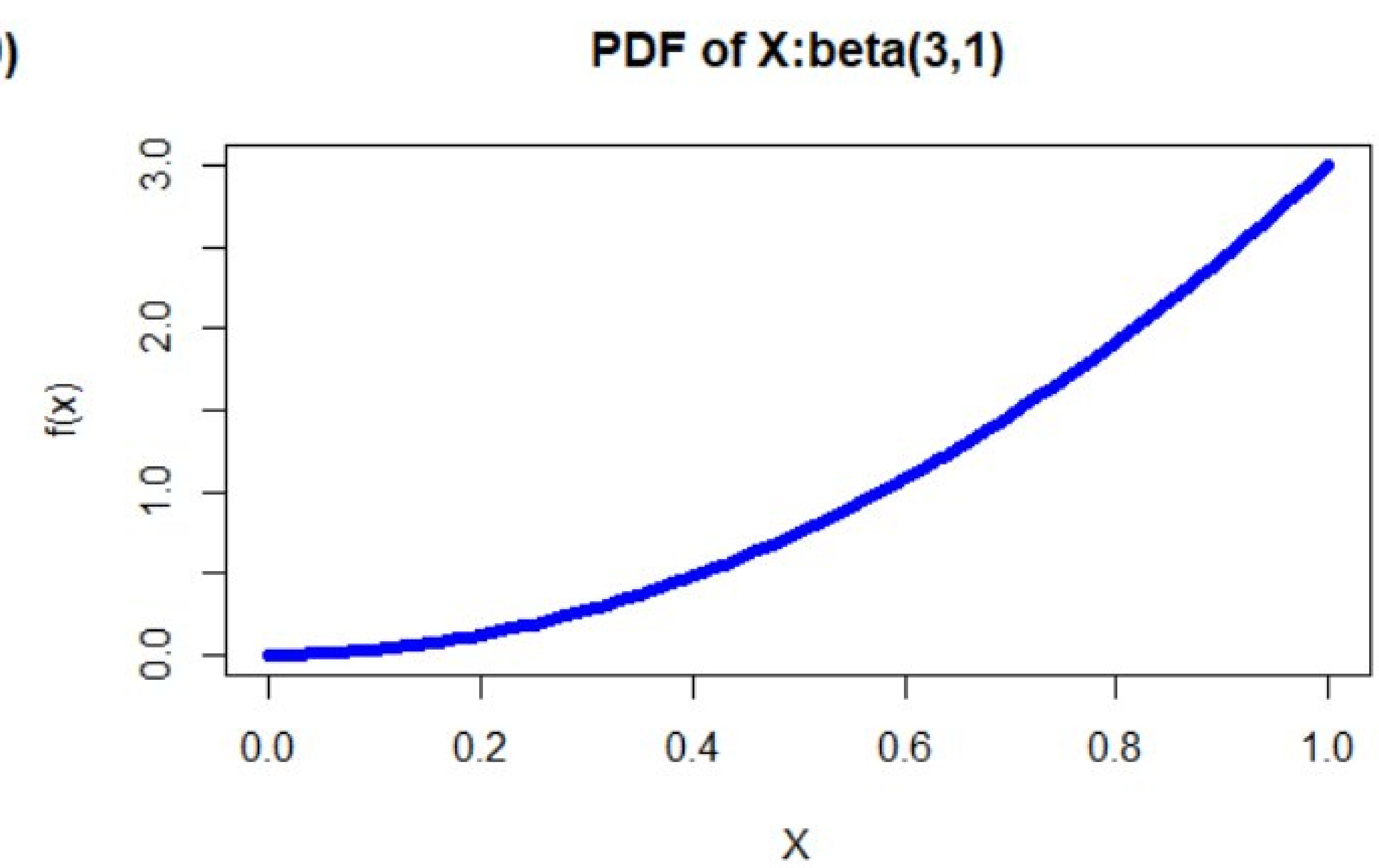


Graphs of the beta distribution for various choices of the parameters









General Comments:

It is noted that there's a difference in the characteristics and graphs depending on the parameters (a,b).

- When a = b, the graphs were skewed to the left and less peaked around the center.
- When a > b, the graph was skewed to the right and less peaked around the center.
- When a < b, the graph was again skewed to the left and less peaked around the center.

It's also noted that the bigger the sample size got, the closer the observations results got to the theoretical calculations, which confirms the accuracy of the calculations.

Appendix

The code used (with α = [2,3,9] and β = [5,3,9])[3]

```
setwd('C:/Users/MSI/Documents/COLLEGE')
set.seed(10)

#data generation from beta(a=α,b=β) & w = the data generated
a <- α
b <- β
w <- rbeta(1000,shape1 = a,shape2 = b)

#Calculating mean & variance
#Theoretical mean: (a/a+b)
#Theoretical var.: ab/((a+b)^2)(a+b+1)

mu <- mean(w)
sigma_sq <- var(w)
minimum <- min(w)
maximum <- max(w)

#Calculating skewness & kurtosis
#Theoretical skewness: mu3/((sigma)^2)^3
#Theoretical kurtosis: mu4/((sigma)^2)^2
skew_coeff <- skewness(w)
kurt_coeff <- kurtosis(w)

#Graphing generated data
hist(w, freq = FALSE, breaks = 25,
     main = "Histogram of the generated data: Beta (a = α, b = β)",
     xlab = "Generated Data", col = "blue")

#P.D.F of beta
x <- w
curve(dbeta(x,a,b), from=0 , to=max(w),
      main= "Density Estimate of Generated Data: Beta (a = α, b = β)",
      xlab= 'x', ylab = 'f(x)',lwd= 5, col ='red', add = FALSE)
```

[3] ***note that it is needed to install the "moments" package in order to get the skewness() and kurtosis() functions to work.**