University of Liege

2022-2023

MATH2022–1
2022-2023

Project :
Monte Carlo Methods in Statistics.

Wilfried Mvomo Eto - s226625

June 9, 2023

2. Plot the distribution $Be(\alpha, \beta)$ for different values of its shape parameters:

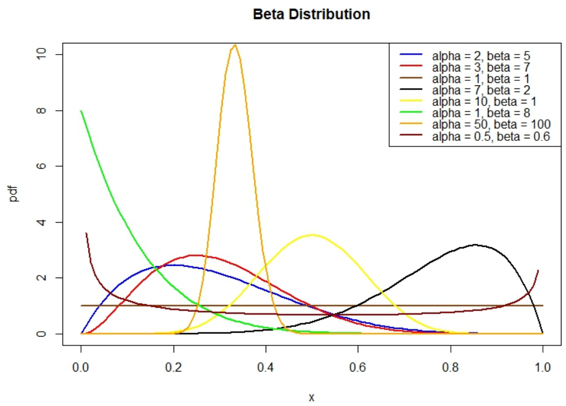

Figure 1: Simulation results draws from Be(α, β)

- When $\alpha = 1$ and $\beta = 1$: this corresponds to the uniform distribution on the interval $[0, 1]$. All values in this range have an equal probability of occurring.
- When $\alpha = 10$ and $\beta = 1$: the distribution is skewed towards the right. Higher values of α result in a sharper peak and a heavier tail on the right side.
- When $\alpha > 1$ and $\beta > 1$: the distribution is generally bell-shaped and symmetric when $\alpha = \beta$. Higher values of α and β result in a more concentrated distribution around the mean.
- When $\alpha = 0.5$ and $\beta = 0.6$: the distribution is U-shaped, with modes at both extremes (0 and 1).

Indeed, when the α and β parameters are equal to 1, the distribution is uniform. This is illustrated in the figure 1.

3. We cannot apply the inverse transform method to sample random values of a beta distribution. Indeed, as we have :
$f(x|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0.$

Afterwards we can deduce the cumulative density function of the beta distribution according to its density function f :
$F(x) = \int_0^x \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} dt.$

However, this integral does not have a closed-form solution in terms of elementary functions. Thus we are not able to have $F^{-1}(u)$ analytically.

Simulate 10 000 values of a Be(4, 6) distribution (on the R code).

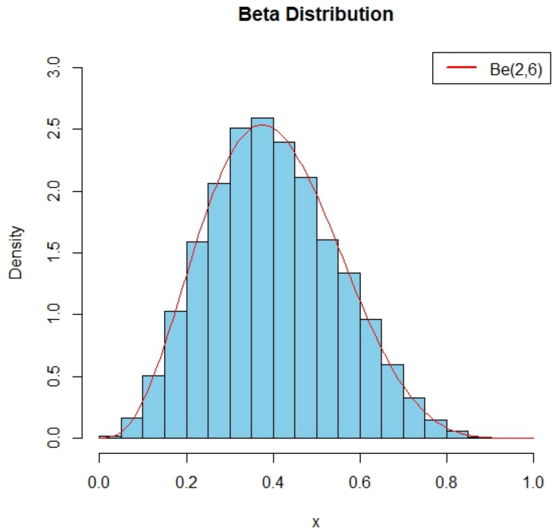

Figure 2: Simulation result of 10 000 draws from a Be(4,6) distribution

4. For the transformation method algorithm, we should consider a family of random variables U_i that are independently and identically distributed for each i to $U_{[0,1]}$, such that :

$$Y = \frac{\sum_{j=1}^{\alpha} log(U_j)}{\sum_{j=1}^{\beta+\alpha} log(U_j)} \sim Beta(\alpha, \beta)$$

with α and β two non-zero natural numbers.

Applying it to simulate 10, 000 values of a Be(2, 6) distribution, we get the figure 3.

5. a. Since a gamma distribution Ga(α, β) is defined by the density function $f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{[0,+\infty[}(x)$ for $\alpha, \beta > 0$, we can determine the cumulative distribution function of X denoted $F_X(x)$. Indeed, we have :

$$F_X(x) = P(X \leq x) = P(\frac{Y_1}{Y_1 + Y_2} \leq x)$$

, $f_{Y1}(y_1) = \frac{1}{\Gamma(\alpha)} \cdot y_1^{\alpha-1} \cdot e^{-y_1}$ and $f_{Y2}(y_2) = \frac{1}{\Gamma(\beta)} \cdot y_2^{\beta-1} \cdot e^{-y_2}$ with $y_1, y_2 \in [0, +\infty[$ (1).

To find the probability distribution function of X, we differentiate the $F_X(x)$ with respect to x:

$f_X(x) = \frac{dF_X(x)}{dx}$.

Using the fact that Y_1 and Y_2 are independent, we can write the joint distribution function of Y_1 and Y_2 as the product of their individual distribution functions:

$P\left(\frac{Y_1}{Y_1+Y_2} \leq x\right) = \int\int [f_{Y_1}(y_1) \cdot f_{Y_2}(y_2)]\, dy_1\, dy_2$.

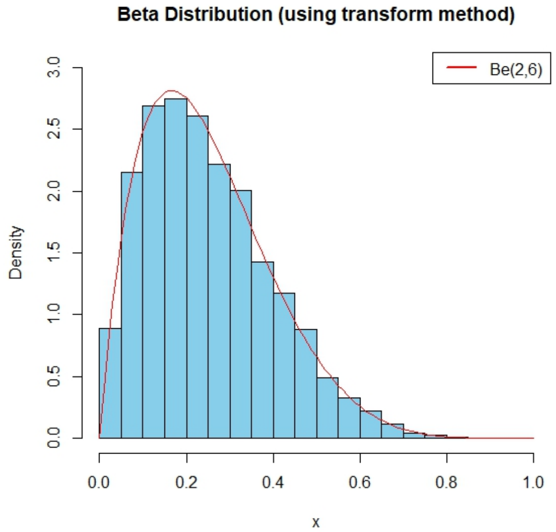

Figure 3: Simulation result of 10 000 draws from a Be(4, 6) distribution (with transformation method algorithm)

To simplify the integral, we use the change of variables: u $= y_1$ and v $= y_1 + y_2$.

The Jacobian determinant of the transformation is 1. Indeed, we have :

$\frac{\partial u}{\partial y_1} = 1$, $\frac{\partial u}{\partial y_2} = 0$, $\frac{\partial v}{\partial y_1} = 1$ and $\frac{\partial v}{\partial y_2} = 1$.

$|J(x)| = \left| \left(\frac{\partial u}{\partial y_1} \cdot \frac{\partial v}{\partial y_2} \right) - \left(\frac{\partial u}{\partial y_2)} \cdot \frac{\partial v}{\partial y_1} \right) \right| = 1 - 0 = 1$

Applying the change of variables, the integral becomes: $\int \int [f_{Y_1}(u) \cdot f_{Y_2}(v-u)] \, du \, dv$, knowing that the limits of integration for u and v are: $0 \leq u \leq \infty$ and $u \leq v \leq \infty$.

Substituting the probability density functions of Y_1 and Y_2 (1), we have:

$F_X(x) = \int \int \left(\frac{1}{\Gamma(\alpha)} u^{\alpha-1} e^{-u} \cdot \frac{1}{\Gamma(\beta)} (v-u)^{\beta-1} e^{-(v-u)} \right) du \, dv.$

Afterwards simplifying we get:

$F_X(x) = \left(\frac{1}{\Gamma(\alpha) \cdot \Gamma(\beta)} \right) \int \int u^{\alpha-1} (v-u)^{\beta-1} e^{-v} \, du \, dv.$

Since this integral represents the Beta distribution with parameters α and β, we can conclude that $X = \frac{Y_1}{Y_1+Y_2}$ follows a Beta distribution with parameters α and β: $X \sim \text{Be}(\alpha, \beta)$.

b. Use this relation to construct an algorithm to generate a beta random variable and simulate 10, 000 values from a Be(2, 6) (via the R code).

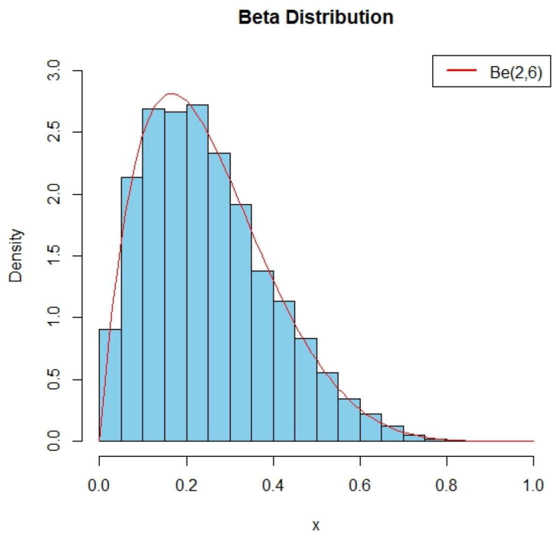

Figure 4: Simulation result of 10 000 draws from Be(α, β) a Be(4, 6) distribution distribution using Ga(2,1) and Ga(6,1) distributions

6. We can gather the different results from applying the both approaches of generating Be(2,6) in the table below :

Table 1: Comparison of Runtimes and Precisions

Approach	Runtime (seconds)	Precision
First Approach (transformation method algorithm)	0.1480029	1
Second Approach (answer 5)	0.003436089	1

Therefore the second approach of generating the beta distribution Be(2,6) (answer 5) has the smaller runtime than the first one (transformation method algorithm). But the both approaches have the same precision.

7. a. First the probability density function of the uniform distribution $U_{[0,1]}$ is constant within the interval $[0, 1]$ and zero elsewhere :

$$f_U(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Moreover, we can find $C \geq 1$ such that :

$$C \geq \frac{f(x|\alpha, \beta)}{f_U(x)}$$

for all x. Hence, the assumptions of the Acceptance-Rejection theorem are observed.

We define acceptance probability by :

$$P = \frac{f(x|\alpha, \beta)}{f_U(x)}$$

.

Then this ratio simplifies to :

$$P = f(x|\alpha, \beta)$$

as the probability density function of $U_{[0,1]}$ is constant within the interval $[0, 1]$. Thus, the acceptance probability P is equal to the probability density function of the Beta distribution with parameters α and β. To calculate P, you can use the formula:

$$P = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1}$$

with $0 < x < 1, \quad \alpha > 1, \quad \beta > 1$.

Since this acceptance probability P is a function of x and depends on the specific values of α and β. Thus we can define the limiting constant C such as C is equal to P.

b. Determine (using R) this limiting constant C used in the Accept-Reject algorithm if the target distribution is Be(1.6, 5.8). We get C = 3.063.

c. Implement the algorithm and perform 10 000 simulations to obtain values from a Be(1.6, 5.8) distribution (via the R code).

d. Create a plot that shows all simulated values, distinguishing between the accepted (in green) and rejected values (in red) (figure 5). Also indicate the instrumental density and the true density function of the Be(1.6, 5.8) distribution. The value of the acceptance is : 0.324.

e. Show, in a general setting, that the probability of acceptance in an Accept-Reject algorithm with limiting constant C on the density ratio $\frac{f}{g}$ is equal to $\frac{1}{C}$.

In an Accept-Reject theorem, we have two probability density functions $f(x)$ and $g(x)$, where $f(x)$ is the target distribution we want to sample from, and $g(x)$ is a known proposal distribution, checking out the existence of the real number $C \geq 1$ with : $Cg(x) \geq f(x)$.

Let's define the acceptance probability P as the probability that a sample generated from $g(x)$ is accepted and used as a sample from $f(x)$. This acceptance probability is directly related to the constant C and the density ratio $f(x)/g(x)$.

To show that the probability of acceptance is equal to $1/C$, we'll start by considering the condition for acceptance in the Accept-Reject algorithm. Typically, a sample from $g(x)$,

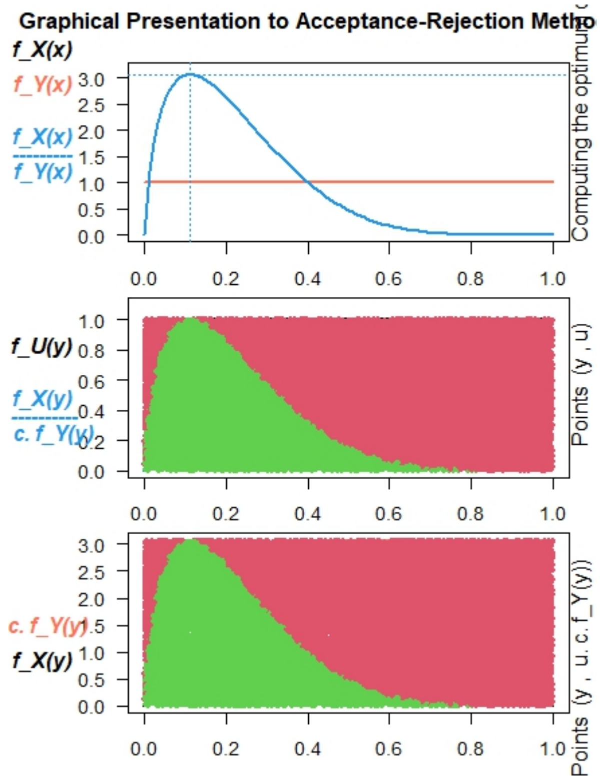

Figure 5: Be(1.6,5.8) distribution from $U_{[0,1]}$ distribution using Accept-Reject algorithm

denoted as X, is accepted if a uniform random variable U in the range $[0, 1]$ is less than or equal to $f(X)/(C \cdot g(X))$:

$$P = P(U \leq \frac{f(X)}{C \cdot g(X)}) = \int_0^1 P(U \leq \frac{f(X)}{C \cdot g(X)})\, dU = \int_0^1 \frac{f(X)}{C \cdot g(X)}\, dU$$

$$P = \frac{1}{C} \int_0^1 \frac{f(X)}{g(X)}\, dU$$

Since U is uniformly distributed in the range $[0, 1]$, its probability density function is constant at 1 over that range.

Afterwards:

$$P = \frac{1}{C} \frac{E[g(X)]}{E[f(X)]}$$

Since $\int_0^1 \frac{f(X)}{g(X)}\, dU$ is equal to the expectation of the density ratio $\frac{f(X)}{g(X)}$ with respect to $g(X)$.

Therefore:

$$P = \frac{1}{C} \frac{E[g(X)]}{E[f(X)]} = \frac{1}{C}$$

, since $\int_0^1 g(X)\, dX = 1$ and $\int_0^1 f(X)\, dX = 1$.

Compare this result with the calculated value in the previous question. Using the R code, the both result are sensitively equal. We can illustrate:

from answer 7.d, we have C = 0.32786, applying the result of the answer 7.e, we have C' = 1/3.063 = 0.3264773.

8. a. Show formally that, for the ratio $\frac{f}{g}$ to be bounded when f is a Beta(α, β) density and g is a Beta(a, b) density, we must have both $a \leq \alpha$ and $b \leq \beta$. Determine the maximal ratio in terms of α, β, a, and b.

We have :

$\frac{f(x)}{g(x)} = \frac{B(a,b) \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}}{B(\alpha,\beta) \cdot x^{a-1} \cdot (1-x)^{b-1}} = \frac{B(a,b)}{B(\alpha,\beta)} \cdot x^{\alpha-a} \cdot (1-x)^{\beta-b}$.

To analyze the boundedness of the ratio, we can consider the behavior of the terms $x^{\alpha-a}$ and $(1-x)^{\beta-b}$ separately:

- for the term $x^{\alpha-a}$:
  - if $\alpha \geq a$, the exponent $(\alpha - a)$ is positive, and the term remains bounded;
  - if $\alpha = a$, the exponent $(\alpha - a)$ becomes 1, and the term remains bounded;
  - If $\alpha < a$, the exponent $(\alpha - a)$ is negative, and the term can become unbounded as x approaches 0;
- for the term $(1-x)^{\beta-b}$:
  - if $\beta \geq b$, the exponent $(\beta - b)$ is positive, and the term remains bounded;
  - if $\beta = b$, the exponent $(\beta - b)$ becomes 1, and the term remains bounded;
  - if $\beta < b$, the exponent $(\beta - b)$ is negative, and the term can become unbounded as x approaches 1.

Thus the ratio $\frac{f}{g}$ is bounded when f is a Beta(α, β) density and g is a Beta(a, b) density, have both $a \leq \alpha$ and $b \leq \beta$.

To determine the maximal ratio, we consider the case when $a = \alpha$ and $b = \beta$:

$\frac{f(x)}{g(x)} = \frac{B(a,b)}{B(\alpha,\beta)} \cdot x^0 \cdot (1-x)^0 = \frac{B(a,b)}{B(\alpha,\beta)}$.

Hence, the maximal ratio of $\frac{f(x)}{g(x)}$ occurs when $a = \alpha$ and $b = \beta$, and the maximal ratio is equal to $\frac{B(a,b)}{B(\alpha,\beta)} = 1$.

b. Redo questions 7(b) to (d) with the appropriate beta distribution as the instrumental distribution, using the knowledge that $a = \lfloor \alpha \rfloor = 1$ and $b = \lfloor \beta \rfloor = 5$ (using the R code). In this case, we can consider the limiting constant C = 1.549 according to the R code. The acceptance rate using the Be(1,5) as proposal distribution, is 0.641. This is value is too greater than one of using $U_{[0,1]}$ which is 0.32816. In term of runtime, that of the first approach (using the uniform distribution $U_{[0,1]}$) is 32.36964 secs, whereas that of the second approach (using the $Be(1,5)$ distribution) is just 4.198068 secs. Thus in term of runtime and acceptance rate, the Accept-Reject algorithm using the $Be(1,5)$ is the best among the both algorithms.

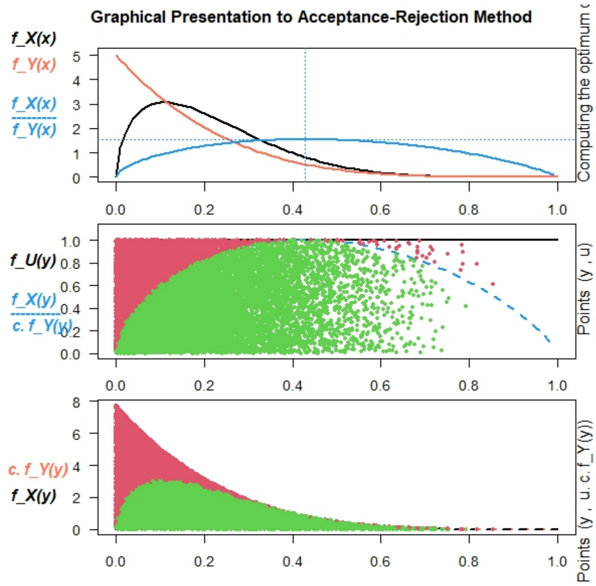

Figure 6: Be(1.6,5.8) distribution from Be(1,5) distribution using Accept-Reject algorithm

9. Show that, for X an observation from the negative binomial distribution NB(r, p), the family of beta distributions Be(α, β) is a family of conjugate priors.

Let's denote the prior distribution as Be(α, β) and the posterior distribution as Be(α', β') after observing data X from the negative binomial distribution NB(r, p). Using the Bayes's theorem, we have: $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$ (1) .

As $X \sim$ NB(r, p), its mass probability density function is defined by:

$\pi(X = k) = \binom{k+r-1}{k} \cdot q^r \cdot (1-q)^k$, where q represents the probability of a success in a single trial in the negative binomial distribution.

From (1), we can write: $\pi(\theta|X = k) \propto f(X = k|\theta) * \pi(\theta)$

$f(X = k|\theta) \cdot \pi(\theta) = (k + r - 1)\binom{k}{r-1} \cdot q^r \cdot (1-q)^k \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$.

Then integrating: $\int f(X = k|\theta) \cdot \pi(\theta) = \int \left[(k + r - 1)\binom{k}{r-1} \cdot q^r \cdot (1-q)^k \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}\right] d\theta$.

Afterwards, we normalize the posterior distribution with $B(\alpha, \beta)$ and we get: $\pi(\theta|X = k) = \frac{(k+r-1)\binom{k}{r-1}\cdot q^r \cdot (1-q)^k \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}}{B(\alpha,\beta)}$.

Therefore:

$\pi(\theta|X = k) \propto \theta^{\alpha'-1} \cdot (1-\theta)^{\beta'-1}$, where $\alpha' = \alpha + n$ and $\beta' = \beta + m$, m and n being two real numbers.

Hence, for X an observation from the negative binomial distribution $\mathrm{NB}(r, p)$, the family of beta distributions $\mathrm{Be}(\alpha, \beta)$ is a family of conjugate priors.

10. a. Show that for the gamma function, the following property holds: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0$. As such, the gamma function can be seen as a generalization of factorials.

The gamma function is defined as:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

.

Using integration by parts, we have:

$$\Gamma(\alpha + 1) = \int_0^\infty t^\alpha e^{-t} dt = \left[-t^\alpha e^{-t}\right]_0^\infty + \int_0^\infty e^{-t} \alpha t^{\alpha-1} dt = \alpha \int_0^\infty t^{\alpha-1} e^{-t} dt$$

, as $\lim_{t\to\infty} t^\alpha e^{-t} = 0$.

Therefore, we obtain:

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

(1) The result (1), show that the gamma function is a generalization of factorials.

b. Determine the moment of order t of a $\mathrm{Be}(\alpha, \beta)$ distribution denoted $\mathrm{E}[X^t]$:

We have:

$\mathrm{E}[X^t] = \int_0^1 x^t \cdot f(x|\alpha, \beta)\, dx = \int_0^1 x^t \cdot \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha+t-1}(1-x)^{\beta-1} dx$

c. Use this formula to obtain the expected value $\mathrm{E}[X]$ and variance Var(X) for $X \sim \mathrm{Be}(\alpha, \beta)$:

The expected value $\mathrm{E}[X]$ is to equal to the moment of order one. As we have determined $\mathrm{E}[X^t]$ in the previous answer, we can infer that :

$\mathrm{E}[X] = \int_0^1 x \cdot f(x|\alpha, \beta)\, dx = \int_0^1 x \cdot \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = \int_0^1 \frac{1}{B(\alpha,\beta)} x^\alpha (1-x)^{\beta-1} dx$

Since $B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(a+1) = a\cdot\Gamma(a)$, then:

$E(X) = \int_0^1 x \cdot \frac{\Gamma(\alpha+\beta)\cdot\Gamma(\alpha)}{\Gamma(\alpha)\cdot\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx$

$= \frac{\Gamma(\alpha+\beta)\cdot\Gamma(\alpha)\cdot\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)} \int_0^1 \frac{\Gamma(\alpha+1+\beta)\cdot\Gamma(\alpha+1)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx.$

$= \frac{\Gamma(\alpha+\beta)\cdot\Gamma(\alpha)\cdot\alpha\cdot\Gamma(\alpha)}{(\alpha+\beta)\cdot\Gamma(\alpha+\beta)} \int_0^1 \frac{\Gamma(\alpha+1+\beta)\cdot\Gamma(\alpha+1)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx$

$E(X) = \frac{\alpha}{\alpha+\beta} \int_0^1 \frac{\Gamma(\alpha+1+\beta)\cdot\Gamma(\alpha+1)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx$

Therefore, the expected value of the $\mathrm{Be}(\alpha,\beta)$ distribution is :

$E(X) = \frac{\alpha}{\alpha+\beta}\int_0^1 \mathrm{Beta}(x;\alpha+1,\beta)dx = \frac{\alpha}{\alpha+\beta}$ since $\int_0^1 \mathrm{Beta}(x;\alpha+1,\beta)dx = 1.$

For the variance $\mathrm{Var}(X)$ of $\mathrm{Be}(\alpha,\beta)$, we have :$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$. (**)

We use the previous answer (10.b), $E(X) = \frac{\alpha}{\alpha+\beta}$ and $\Gamma(a+1) = a\cdot\Gamma(a)$ to determine the moment of order two:

$\mathrm{E}[X^2] = \int_0^1 x^2 \cdot f(x|\alpha,\beta)\ dx = \int_0^1 x^2 \cdot \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{(\alpha+\beta)(\alpha+\beta+1)}{\alpha(\alpha+1)}.$

Using (**), we can conclude that the variance of $\mathrm{Be}(\alpha,\beta)$ distribution is:

$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = (\frac{(\alpha+\beta)(\alpha+\beta+1)}{\alpha(\alpha+1)}) - (\frac{\alpha}{\alpha+\beta})^2 = \frac{(\alpha+\beta)^2}{(\alpha+\beta+1)\alpha\beta}.$

d. Use the Monte Carlo approach to estimate the shape parameters of a $\mathrm{Be}(25,7)$ distribution (set n = 10 000)(via the R code):

From the answer 10.c, we have the sample mean and the sample variance are defined as follows: $\hat{E}(X) = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}$ (1) and $\hat{Var}(X) = \frac{(\hat{\alpha}+\hat{\beta})^2}{(\hat{\alpha}+\hat{\beta}+1)\hat{\alpha}\hat{\beta}}$.

- set the number of samples $n = 10000$ for the Monte Carlo simulation;
- generate n random samples from the Beta(25,7) distribution using an appropriate random number generator;
- calculate the empirical mean of the generated samples: $\hat{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$, where X_i is the i-th sample;
- since the exact expression for variance is not known, we cannot use it directly. However, we can use the empirical variance of the generated samples: $\hat{\mathrm{Var}}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{E}(X))^2$;
- now, we can estimate the parameters α and β using the following equations:

$\hat{\alpha} = \hat{E}(X) \cdot \left(\frac{1-\hat{E}(X)}{\hat{\text{Var}}(X)} - 1 \right)$

$\hat{\beta} = (1 - \hat{E}(X)) \cdot \left(\frac{1-\hat{E}(X)}{\hat{\text{Var}}(X)} - 1 \right)$

Thus we obtained via the R code the estimates $\hat{\alpha} = 32.97069$ and $\hat{\beta} = 8.854324$ obtained from the Monte Carlo approach.

11. a. Use the previous implementation for the MOM and use the function ebeta contained in the package EnvStats for the MLE estimator for the beta distribution to create histograms illustrating the distribution of each of the shape parameters (using R code):

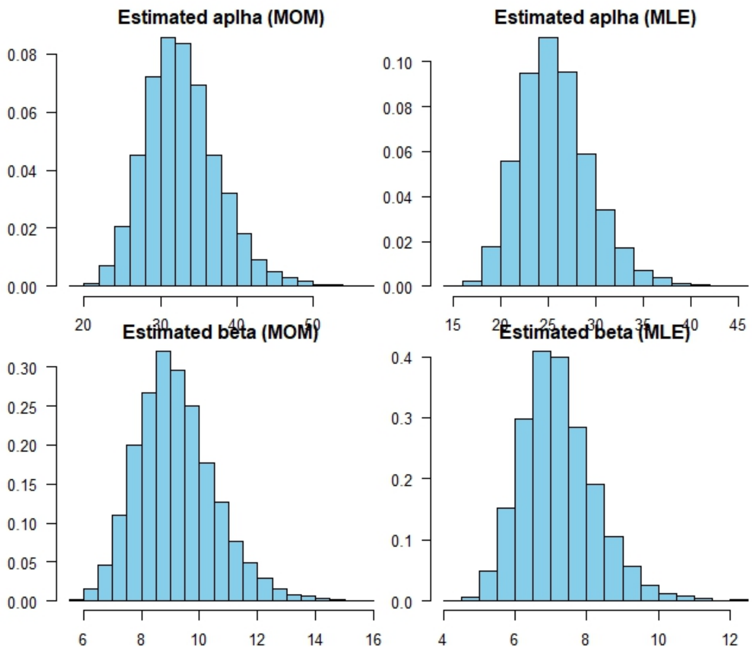

Figure 7: Histograms illustrating the distribution of each of the shape parameters

b. Evaluate the efficiency of each method in terms of bias and variance:

Based on the given results, the best approach among Maximum Likelihood Estimation (MLE) and Method of Moments (MOM) depends on the specific criteria being considered.

For the parameter α, the MLE approach has a lower bias (0.752558) compared to the MOM approach (7.85832). Additionally, the MLE approach has a lower variance (14.23337) compared to the MOM approach (23.52876). Therefore, based on these criteria, the MLE approach is preferable for estimating the α parameter.

For the parameter β, the MLE approach again has a lower bias (0.2042686) compared to the MOM approach (2.196341). The variance of the β parameter is also lower with the MLE approach (1.054457) compared to the MOM approach (1.820871). Thus, considering these criteria, the MLE approach is also preferred for estimating the β parameter.

In summary, based on the provided bias and variance values, the Maximum Likelihood Estimation (MLE) approach appears to outperform the Method of Moments (MOM) approach for both the α and β parameters.

11. Illustrate the asymptotic normality of the obtained MLE estimators:

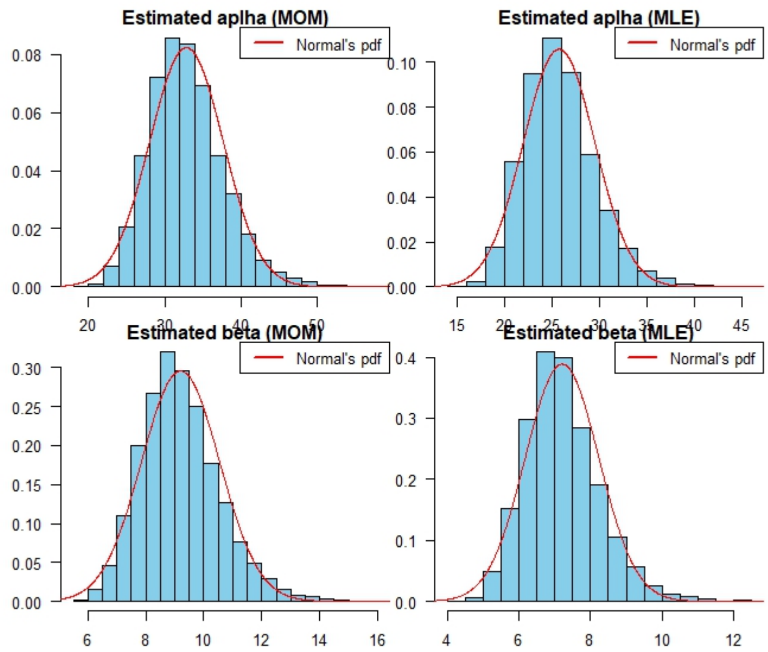

Figure 8: Illustration the asymptotic normality of the obtained MLE estimators

Remark: Upon the figures 5 and 6, $f_X(x)$ and $f_Y(x)$ represent respectively the pdfs target and proposal distributions relatively to the given answer. An additional, f_U is the pdf of the uniform distribution $U_{[0,1]}$.