**The Beta Distribution:**

**A Classic in Mathematical Statistics for Social Science Researchers**

## ABSTRACT

The objective of thismethodological-academic chapter is to provide a clear and illustrative introduction to the Beta distribution, with examples applied to the social and health sciences using data randomly generated from Beta distributions. Although this continuous distribution is well-known in mathematical statistics, it remains relatively unfamiliar to social scientists. The chapter begins with a historical overview to provide context, followed by an explanation of its support, parameters, and key functions, including density, cumulative distribution, quantile, moment-generating, and characteristic functions. Descriptive measures such as central tendency, variation, and shape are also discussed. The chapter then covers parameter estimation, focusing on the method of moments and maximum likelihood. The fit of a random sample to a Beta distribution with unknown parameters is tested using the G test and quantile-quantile plot. Additionally, the chapter presents the associated limiting distributions and generalizations. Four examples illustrate applications of the Beta distribution. The first example shows how to compute probabilities and descriptive measures when the two parameters of the distribution are known. The second example estimates the two shape parameters (alpha and beta) using the method of moments, calculates descriptive measures, and performs probability calculations. The third example estimates alpha and beta, along with their standard errors and confidence intervals, using likelihood methods. The fourth example tests the goodness of fit of a sample to a Beta distribution using the G-test and the QQ plot. Four appendices with R scripts for these calculations are included. The chapter closes with conclusions and suggestions for using the Beta distribution.

*Keywords:Continuous Distribution, Parameter Estimation, Goodness of Fit, Social and Health Sciences, Mathematical Statistics, R program.*

**1. INTRODUCTION**

The Beta distribution has multiple applications in the social sciences due to its ability to model probabilities and proportions, as well as its flexibility characteristics, given its ability to take on multiple forms(Kaplan, 2023) [1]. Although its range is bounded from 0 to 1 in its classic expression, this can be extended in the form of four parameters: two shape and two threshold (Yin & Yin, 2024) [2]. This distribution is ideal for describing data restricted to the range [0, 1], such as proportions of success, approval rates, electoral participation or fractions of time dedicated to certain activities (Geissinger et al., 2022) [3]. In a public opinion survey, the Beta distribution can model the proportion of respondents who approve of certain policies (Pokhriyal, Valentino, & Vosoughi, 2023) [4]. In Bayesian estimation, it is often employed as a non-informative prior in estimating a probability or a quantile (Papas, 2021) [5]. For example, in a clinical trial to measure the effectiveness of a psychotherapeutic intervention, the Beta distribution may represent the initial uncertainty among therapists about the probability of success (Stahl et al., 2022) [6].

The Beta distribution is also valuable for generating simulated data in scenarios where variables are expected to range between 0 and 1 or between a minimum *a* and a maximum *b* (in the case of the four-parameter Beta distribution). Its shapes can vary widely, from unimodal to bimodal or no mode, symmetrical to skewed (positive or negative), and from platykurtic to mesokurtic or leptokurtic profiles. This versatility makes it an excellent tool for studying kurtosis and for understanding this elusive concept of the shape of a distribution (Braouezec & Cagnol, 2023) [7]. Additionally, it can serve as a model for calculating probabilities in phenomena that conform to this distribution, such as those evaluated in terms of success or failure (Santana-e-Silva, Cribari-Neto, & Vasconcellos, 2022) [8].

Despite its flexibility and applicability in modeling independent trials with binary outcomes (e.g., clinical case vs. no case, therapeutic success vs. failure, or fit vs. unfit student for clinical practice), the Beta distribution remains relatively underutilized in social sciences, such as psychology (Bertelsen et al., 2022) [9]. However, it has been widely studied and applied since the early development of mathematical statistics(Aljohani, 2024) [10]. Therefore, the purpose of this chapter is to familiarize social science researchers with this versatile distribution by providing a clear and illustrative introduction to the Beta distribution with examples applied to the social and health sciences.

**2.Characterization of the distribution**

The Beta distribution is a continuous probability distribution that offers an alternative approach to solving issues in probability calculations associated with the discrete binomial distribution. Specifically, it models the proportion of successes in *n* independent Bernoulli trials under replenishment sampling, where the probability of success remains constant (Hahn, 2022) [11].

**2.1 historical note**

The Beta distribution was introduced by the English mathematician Thomas Bayes (1701‑1761). During the final years of his life, Bayes became interested in probability theory, developing new ideas that he did not publish (Kaplan, 2023) [1]. His friend, the Welsh philosopher and mathematician Richard Price (1723‑1791), was aware of Bayes' work and decided to present it to the Royal Society of London(Gelman & Vehtari, 2021) [12]. At the request of the society, Price authored the essay An Essay towards Solving a Problem in the Doctrine of Chances, which was published in 1763 in the Philosophical Transactions of the Royal Society of London [13]. In this work, based on Bayes' writings, the Beta distribution appears, although it is not explicitly named. It is worth noting that Bayes did not formalize the moment-generating function or the characteristic function of this distribution, nor did he develop its moments (García-García et al., 2022) [14].

The English statistician Karl Pearson (1857‑1936) explored this distribution in mathematical detail(Agresti, 2021) [15], referring to it as the Type I distribution (Pearson, 1895) [16]. It is worth noting that the correspondence between Pearson's Type I distribution and the Beta distribution is not direct and requires rescaling (Pearson, 1916) [17]. Additionally, the actuary William Palin Elderton (1877‑1962), a collaborator of Pearson, contributed to the tabulation of the Pearson’s Type I distribution, focusing on both the mode (Elderton, 1906) [18] and the mean (Elderton, 1927, 1938) [19‑20]. In English-language literature prior to the Second World War, the distribution was commonly referred to as Pearson's Type I distribution. However, the term "Beta distribution" has since been widely adopted(Mathieson & Terhorst, 2022) [21].

The estimation of the parameters of the Beta distribution was the subject of a significant controversy between Karl Pearson and Ronald Aylmer Fisher (1890‑1962). Pearson advocated for the method of moments, while Fisher championed the method of maximum likelihood. This debate was ultimately resolved in favor of Fisher with the advent of computers, as the maximum likelihood method involves an iterative calculation process that is too complex to perform manually(Bowman & Shenton, 2007) [22]. The iterative process typically begins with an initial estimation using the method of moments, which is straightforward and serves as a starting point for obtaining a more efficient solution with reduced error and narrower confidence intervals (Fujita, Okada, & Katahira, 2022) [23].

**2.2Support and parameters**

The Beta distribution is defined by two shape parameters, α and β, with a parameter space in the interval (0, ). The support of a random variable X following this distribution ranges from 0 to 1. The value 0 is included in the support if α > 0, and the value 1 is included if β > 1. The distribution is denoted as Beta(α, β)(Pascucci, 2024). [24]. Refer to Equation 1.

|  |  |  |
| --- | --- | --- |
| *Distribution**Support**Shape parameters* |  | (1) |

**2.3 key functions**

This subsection of the Beta distribution characterization defines six key functions: the probability density function, the cumulative distribution function, the quantile function, the moment generating function, and the characteristic function(Ghahramani, 2024) [25].

*Probability Density Function*: The definite integral of the density function enables the determination of the probability that a value *x* of the continuous random variable X falls within a specified interval of integration, provided that the interval lies within the distribution's domain. This is represented by *fX*(*x*).Refer to Equation 2 for its analytic expression.

|  |  |
| --- | --- |
|  | (2) |

In Equation 2, B(α, β) represents the beta function of the parameters α and β, which is equivalent to a quotient where the numerator is the product of the gamma functions of α and β, and the denominator is the gamma function of the sum of α and β. The analytical formula for this function is shown in Equation 3.The beta function can be computed with the R program using the following script (Singh, 2020) [26].

|  |  |
| --- | --- |
|  | (3) |

# R script to calculate the beta function. Define the parameters.

alpha <- 2 # Change this value to the desired α parameter.

beta <- 3 # Change this value to the desired β parameter.

result <- beta(alpha, beta)

cat("The value of B(α =", alpha, ", β =", beta, ") is", result, ".", "\n")

# The value of B(α = 2, β = 3) is 0.08333333.

The gamma function, denoted as Γ(α), can also be computed in R (Singh, 2020; R Core Team, 2024a)[26‑27], with its analytical formula shown in Equation 4. The gamma function of 0.5 is the square root of pi, the gamma function of 1 is 1, the gamma function of 1.5 is the square root of pi divided by 2, the gamma function of 2 is 2, and the gamma function of 2.5 is three-quarters of the square root of pi, illustrating some of its values.

|  |  |
| --- | --- |
|  | (4) |

# R script to calculate the gamma function. Define the parameter.

alpha <- 3 # Change this value to the desired α value, which has to be a real positive number.

result <- gamma(alpha)

cat("The value of gamma(α =", alpha, ") is", result, ".", "\n")

# The value of gamma(α = 3) is 2.

The Beta distribution is known as a multiform distribution because it exhibits a wide variety of profiles (Elderton, 1938) [20]. When both parameters α and β are less than one, the curve is convex and bimodal. Typically, its profile is smooth or curvilinear, but if one of the parameters equals 1, the curve becomes angular. When α = β = 1, the distribution corresponds to the standard continuous uniform distribution, which has a rectangular shape (Figure 1). If both parameters are greater than one, the curve is concave. When α and β are equal, the profile is symmetric; if the values of α and β differ, the distribution becomes asymmetric. Positive (right-tailed) skewness occurs when β > α, while negative (left-tailed) skewness occurs when α > β. As α and β approach equality and increase toward infinity, the distribution converges to a normal distribution (Figure 2).



**Figure 1. Probability density functions of four beta-distributed random variables**



**Figure 2. Probability density functions of four distinct beta-distributed random variables**

*Cumulative Distribution Function*: This function calculates the probability that a continuous or discrete random variable X falls within the range from the lower limit of the distribution to a given value x, where x lies within the distribution's domain. It is written as *FX*(*x*).Refer to Equation 5 for its analytic expression, where *Ix*(α, β) represents the regularized incomplete beta function of the parameters α and β evaluated at *x*.

|  |  |
| --- | --- |
|  | (5) |

The analytical formula for the function *Ix*(α, β) is given in Equation 6.It can be computed in R using the following script (R Core Team, 2024b) [27].

|  |  |
| --- | --- |
|  | (6) |

# R script to calculate the regularized incomplete beta function.

# Define the parameters.

alpha <- 2 # Change this value to the desired α parameter.

beta <- 3 # Change this value to the desired β parameter.

x <- 0.5 # Change this value to the upper limit x (in the interval [0, 1]).

result <- pbeta(q = x, shape1 = alpha, shape2 = beta)

cat("The valueof I\_x(x =", x, ",α =", alpha, ",β =", beta, ") is", result,".", "\n")

# The value of I\_x(x = 0.5, α = 2, β = 3) is 0.6875.

*Quantile Function*: This function is the inverse of the cumulative distribution function. It determines the value of X corresponding to a specified cumulative probability, with the input argument ranging between 0 and 1. It is represented as *QX*(x) or *F-1X*(*x*).Refer to Equation 7 for its analytic expression, where *I-1x*(α, β) denotes the inverse function of the regularized incomplete beta function of the parameters α and β evaluated at *p* (quantile order or cumulative probability up to *x*).

|  |  |
| --- | --- |
|  | (7) |

There is no analytical formula for this function. However, it can be computed with the R program using the following script (R Core Team, 2024b) [28].

# R script to calculate the inverse function of the regularized incomplete beta function. Define the parameters.

alpha <- 2 # Change this value to the desired α parameter.

beta <- 3 # Change this value to the desired β parameter.

p <- 0.6875 # Change this value to the desired cumulative probability value p or order of quantile.

result <- qbeta(p, shape1 = alpha, shape2 = beta)

cat("The value of I\_x(p =", p, "|", "α =" , alpha, ",β =", beta, ")", "is", result,".", "\n")

# The value of I\_x(p = 0.7| α = 2, β = 3) is 0.5.

Moment Generating Function: The k-th derivative of this function, when evaluated at zero, yields the k-th order moments of the distribution, which is how it gets its name. These moments represent the expected value of the variable raised to the k-th power. The function is denoted as *MX*(*t*).Refer to Equation 8 for its analytic expression.

|  |  |
| --- | --- |
|  | (8) |

1F1(α, α+β, *t*) is the confluent hypergeometric functionthat can be computed with the R program using the following script (R Core Team, 2024c).

# R script to calculate theconfluent hypergeometric function.

# Load package hypergeo.

library(hypergeo)

# Define the parameters.

alpha <- 2 # Change this value to the desired α parameter.

beta <- 3 # Change this value to the desired β parameter.

t <- 1.5 # Change this value to the desired t value,which is a positive real number greater than or equal to 1.

result <- genhypergeo(U = alpha, L = alpha + beta, z = t)

cat("The value of 1\_F\_1(α =", alpha, ", β =", alpha + beta, ", t =", t, ") is", result, ".", "\n")

# The value of 1\_F\_1(α = 2, β = 5, t = 1.5) is 1.907989.

*Characteristic Function*: The characteristic function is derived by applying the Fourier transform (with an inverted sign) to the mathematical expression of the probability density function. It is used to study the behavior of the moments of a distribution and is represented as *CX*(*t*).Refer to Equation 9 for its analytic expression, where *i*denotes the unit imaginary number(√-1) and 1F1(α, α+β, it) represents the confluent hypergeometric function with parametersα and α+β, evaluated at *i* × *t*.

|  |  |
| --- | --- |
|  | (9) |

**2.4 Descriptive measures**

Three types of descriptive measures are presented: measures of central tendency, including the arithmetic mean, median, mode, geometric mean, and harmonic mean; measures of variability, including variance, standard deviation, mean absolute deviation, and entropy; and measures of shape, including skewness and kurtosis based on standardized central moments. Additionally, the calculation of the k-th non-central moment is shown (Pascucci, 2024) [24].

**2.4.1 Measures of central tendency**

*Mathematical expectation* (*E*) *or arithmetic mean*(*M*): This measure of central tendency is the first (non-central) moment, and in this distribution, it is calculated as the definite integral from 0 to 1 of each value multiplied by its corresponding density (Equation 10).

|  |  |
| --- | --- |
|  | (10) |

In the estimation of shape parameters α and β using the maximum likelihood method, the mathematical expectation of the logarithm of the values (Equation 11) and the mathematical expectation of the complement of the logarithm of the values (Equation 12), where the values of X range from 0 to 1, are also important.

|  |  |
| --- | --- |
|  | (11) |
|  | (12) |

In equations 11 and 12, ψ(α) denotes the digamma function of α which is the partial derivative in α of the natural logarithm of the gamma function evaluated in α. Its analytical expression is shown in Equation 13.

|  |  |
| --- | --- |
|  | (13) |

When α is a natural number, the digamma function of α can be calculated using the mathematical formula in Equation 14, where the Euler and Mascheroni constants, denoted by e and γ, respectively, appear (Equation 15).

|  |  |
| --- | --- |
|  | (14) |
|  | (15) |

When α is the average of two consecutive natural values, the formula for calculating the digamma function of α is shown in Equation 16.

|  |  |
| --- | --- |
|  | (16) |

*Median* (*Mdn*): This measure of central tendency is the 0.5 quantile, obtained by evaluating the function in Equation 7 at 0.5, as shown in Equation 17. If the parameters α and β are greater than one, the median is given by (α - 1/3) / (α + β - 2/3), as shown in Equation 18.

|  |  |
| --- | --- |
|  | (17) |
|  | (18) |

*Mode* (*Mo*):This measure of central tendency corresponds to the peak of the distribution. If the parameters of alpha and beta form are less than 1, the distribution is bimodal (asymptotic values 0 and 1). If alpha is less than or equal to 1 and beta is greater than 1, the mode is the value 0. If alpha is greater than 1 and beta is less than or equal to 1, the mode is 1. If alpha and beta are equal to 1, it has no mode, since it becomes a standard continuous uniform distribution. If alpha and vein are greater than one, the mode is the quotient between alpha minus (numerator) and α + β – 2 (denominator). Refer to Equation 19.

|  |  |
| --- | --- |
|  | (19) |

*Geometric Mean* (*G*)This measure of central tendency is obtained by taking the natural exponential of the definite integral from 0 to 1 of the product of the natural logarithm of each value and its corresponding density. If the shape parameters α and β are greater than 1, the geometric mean is given by the quotient (α - 1/2) / (α + β - 1/2). Refer to Equation 20.

|  |  |
| --- | --- |
|  | (20) |

*Harmonic Mean* (*H*)This measure of central tendency is obtained as the inverse of the definite integral from 0 to 1 of the product of the inverse of each value and its corresponding density. When the shape parameter α is greater than 1 and the shape parameter β is greater than 0, the harmonic mean is given by the quotient (α - 1) / (α + β - 1). Refer to Equation 21.

|  |  |
| --- | --- |
|  | (21) |

**2.4.2 Measure of the k-th order tendency**

*K-th Non-Central Moment*: This is obtained through the definite integral from 0 to 1 of the product of each value raised to the k-th power and its corresponding density. It is equal to the product from 1 to k of the quotient (α + *i*) / (α + β + *i*), where *i* is the index of the moment's order. Refer to Equation 22.Its application to the non-central moment of second order can be seen in Equation 23.

|  |  |
| --- | --- |
|  | (22) |
|  | (23) |

**2.4.3 Measures of variation**

*Variance*(σ²): This measure of variation is obtained through the definite integral from 0 to 1 of the product of the squared deviation of each value from the arithmetic mean and its corresponding density (Equation 24). It is equivalent to the second non-central moment minus the square of the first central moment (Equation 25).

|  |  |
| --- | --- |
|  | (24) |
|  | (25) |

In the estimation of the shape parameters α and β using the maximum likelihood method, the variance of the logarithm of the values (Equation 26) and the variance of the complement of the logarithm of the values (Equation 27) play an important role. These variances are computed using the trigamma function.

|  |  |
| --- | --- |
|  | (26) |
|  | (27) |

In Equations 26 and 27, the trigamma function of α, denoted as ψ₁(α), appears. This function corresponds to the second partial derivative of the gamma function with respect to α, or equivalently, the first partial derivative of the digamma function with respect to α (Equation 28).It can be computed either by a definite integral from 0 to 1 (Equation 29) or by an infinite series with terms 1 / (α + n)2, where *n*is the index of the series, ranging from 1 to infinity (Equation 30).The digamma function of 0.5 is half of pi squared, the digamma function of 1 is one-sixth of pi squared, the digamma function of 1.5 is half of pi squared minus 4, and the digamma function of 2 is one-sixth of pi squared, as illustrated with some values of this function (Equation 31).

|  |  |
| --- | --- |
|  | (28) |
|  | (29) |
|  | (30) |
|  | (31) |

The trigamma function of α can be computed with the R program using the following script (R Core Team, 2024a) [27].

# R script to calculate the trigamma function. Define the parameters.

alpha <- 3 # Change this value to the desired α value, which has to be a real positive number.

result <- trigamma(alpha)

cat("The value of trigamma(α =", alpha, ") is", result, ".", "\n")

# The value of trigamma(α = 3) is 0.3949341.

*Standard Deviation*(*σ*): This is the square root of the variance. As a measure of absolute variation, it returns the value to the original units, counteracting the quadratic units generated by the variance.Refer to Equation 32.

|  |  |
| --- | --- |
|  | (32) |

*Mean Absolute Deviation*(*MAD*): This is the definite integral from 0 to 1 of the product of each absolute deviationof each value from the arithmetic mean and its corresponding density. Refer to Equation 33

|  |  |
| --- | --- |
|  | (33) |

*Shannon's Entropy*: Entropy, or the average amount of information, denoted by H (Shannon, 1948) [30]. For the Beta distribution, this measure of variability is computed using the negative definite integral from 0 to 1 of the product of the natural logarithm of each density and the corresponding density (Equation 34). For discrete distributions, it is calculated as the negative sum of the product of the natural logarithm of each probability density and the corresponding density. In such distributions, the maximum entropy value corresponds to a uniform distribution, where all outcomes contribute equally to the total information, while an entropy of 0 corresponds to a degenerate distribution, where a single value contains all the information. In these distributions, the probability mass or point probability is always less than or equal to 1. If a particular value x of the random variable X has a unit point probability (*p*(X = *x*) = 1), the associated information is zero. In all other cases, the information is greater than 0.

|  |  |
| --- | --- |
|  | (34) |

For discrete distributions with bounded support, the entropy reaches a maximum corresponding to the natural logarithm of the number of values. This maximum occurs when the values are uniformly distributedand provide the same information (Saupe et al., 2024) [31]. In the case of a distribution with bounded support between a minimum value *a* and a maximum value *b*, the normalized efficiency or entropy can be calculated as the ratio of the entropy to its maximum, using the natural logarithm of the cardinality of the support: Η(x)/*max*(Η) = Η/ln(*k*), where *k* = #{*a*, *b*} and # denotes the cardinality of a finite set.

However, for continuous distributions, entropy can be either positive or negative and does not have a maximum. When distributions have density values greater than 1, entropy can be negative, as is the case for beta and exponential distributions (Wang, Zhong, Li, & Peng, 2021)[32]. Unlike Shannon's entropy, relative entropy or Kullback-Leibler divergence is bounded between 0 and 1 for continuous distributions [33]. However, it is not a measure of variability but of goodness of fit or agreement between two models (Kullback, 1959) [34].

The Kullback-Leibler divergence or relative entropy between two variables X and Y with Beta distribution, *DKL*(X‖Y), is calculated using the formula in Equation 35.

|  |  |
| --- | --- |
|  | (35) |

**2.4.4 Measures of shape**

*Pearson’s Measure of Skewness* (based on the standardized third central moment): This measure was developed by Karl Pearson as a parameter within his system of distributions (Pearson, 1916) [17] and is denoted as ∓√β1, which is often simplified to β1. It characterizes the asymmetry of the shape of a distribution. A value of 0 indicates that the two sides of the distribution are equal by locating the axis of symmetry or partition at the arithmetic mean. A negative value signifies that the left side is longer than the right side, while a positive value indicates that the right side is longer than the left. For a Beta distribution, it is computed using the definite integral from 0 to 1 of the product of each standardized value cubed and its corresponding density. Refer to Equation 36.

|  |  |
| --- | --- |
|  | (36) |

*Pearson’s Excess Kurtosis*: This is a measure of kurtosis based on the fourth standardized central moment (one of the parameters within Pearson’s distribution system), adjusted by subtracting 3, which is the kurtosis value for a normal distribution. Pearson developed it in 1905 and denotedas β2‑3 [35].It quantifies the redistribution of density from the shoulders to the tails relative to a normal distribution, particularly in continuous symmetric distributions.In this type of distribution, a value of 0 indicates mesocurtosis, with a distribution of density between the tails and the shoulders corresponding to a normal distribution; a negative value indicates shortened tails versus thickened shoulders, and a positive value indicates elongated tails versus thinned shoulders. It can be expressed as the ratio of the fourth central moment to the square of the second central moment. For a Beta distribution, it is calculated using the definite integral from 0 to 1 of the product of each standardized value raised to the fourth power and its corresponding density. Refer to Equation 37.

|  |  |
| --- | --- |
|  | (37) |

**3. Parameter estimation**

In this third section of the chapter, the estimation of the two shape parameters of the Beta distribution is discussed using two methods: the method of moments, which provides estimates directly from the sample mean and variance, and maximum likelihood, which requires an initial estimate of the alpha and beta parameters and yields the final estimates after an iterative computational process (Ali, Akanihu, & Felix, 2023) [36].

**3.1 Moment method**

Consider a random sample of size *n* from a variable X that follows a Beta distribution, where the shape parameters alpha and beta are unknown. In this method, the population mean is equated to the sample mean (Equation 38), and the population variance is equated to the sample variance (Equation 39). This results in a system of two equations with two unknowns (alpha and beta). On the left side are the two unknown parameters, while on the right side are two statistics that depend only on the sample and have known values (Equation 40).

|  |  |
| --- | --- |
|  | (38) |
|  | (39) |
|  | (40) |

We isolate the parameter beta in the first equation of the system shown in Equation 40. See Equation 41 for further details on the process of thisisolation.

|  |  |
| --- | --- |
|  | (41) |

Substitute beta into the second equation of the system shown in Equation 40 and solve for alpha,resulting in the alpha estimator. See Equation 42 for further details on the process of this substitution.

|  |  |
| --- | --- |
|  | (42) |

The value obtained for α is substituted into the above clearance of βshown in the Equation 41, resulting in the beta estimator.See Equation 43 for more details on this new substitution.

|  |  |
| --- | --- |
|  | (43) |

The circumflex accent is added to the parameters alpha and beta and they are equal to the results of the processes shown in Equations 42 and 43, respectively. In this manner, the formulas for the estimators derived using the method of moments are obtained (Equation 44). These two estimators are straightforward to calculate from the sample data of size *n* and become more accurate as *n* increases (Prucha, 2021) [37].

|  |  |
| --- | --- |
|  | (44) |

**3.2** **Maximum likelihood method**

Consider the likelihood function presented in Equation 45, which represents the probability or likelihood that specific values of the two shape parameters of a Beta distribution (α and β) correspond to the distribution of the sample x, randomly drawn from the population of values in the variable X (Ali et al., 2023) [36]. This method assumes that the sample x is a sequence of independent and identically distributed random variables, requiring a large sample size. The objective is to determine the two parameter values that maximize the likelihood function.

|  |  |
| --- | --- |
|  | (45) |

Before maximizing the function, taking the natural logarithm of both sides of Equation 45 yields the log-likelihood function presented in Equation 46. This process is referred to as the log-likelihood transformation (Rahman & Amin, 2024) [38].

|  |  |
| --- | --- |
|  | (46) |

Maximization is performed by taking partial derivatives, setting them equal to zero, and solving the resulting system of equations. Specifically, the logarithm of the likelihood function shown in Equation 46 is differentiated with respect to α and set to zero. Similarly, it is differentiated with respect to β and set to zero. This process results in a system of two equations with two unknowns (α and β) that lacks a straightforward analytical solution. Refer to Equation 47 for more details on this procedure (Millard, 2024) [39].

|  |  |
| --- | --- |
|  | (47) |

The resulting system of two equations with two unknowns (α and β), which has no closed-form solution, is solved through an iterative numerical computation process. This process continues until there is no significant change in the values of α and β or the change is smaller than a predefined convergence criterion. As an initial solution , the estimates obtained via the method of moments can be used. Using the formulas in Equation 48, new estimatesare computed iteratively. Starting from the initial estimates provided by the method of moments, the process continues until the k-th iteration, where the estimates remain unchanged in the first six decimal places compared to the previous iteration (*j* = 0, 1, ..., *k*-1)(Zaiontz, 2024) [40].

|  |  |
| --- | --- |
|  | (48) |

If the values of α and β are expected to be greater than 1 (indicating a concave density curve profile), the values given in Equation 49 can be used as initial estimates in Equation 48, based on the geometric mean (Fisher, 1971) [41].

|  |  |
| --- | --- |
|  | (49) |

Among the iterative procedures used to maximize the likelihood function, the limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm for bounded constrained optimization, developed by Byrd et al. (1995) [42], is the most commonly recommended approach (R Core Team, 2024b) [28].

To obtain the standard errors of the two estimates and the covariance between them, the Fisher information matrix is first computed. For the two parameters of the Beta distribution, this is a 2×2 square matrix, as shown in Equation 50 (Chattamvelli & Shanmugam, 2022) [43].

|  |  |
| --- | --- |
|  | (50) |

The inverse of Fisher's information matrix provides the covariance matrix of the maximum likelihood estimates. It should be noted that the two diagonal elements of the covariance matrix correspond to the Cramér-Rao lower bounds (CRLB) for the α and β estimates (Aharon & Tabrikian, 2024) [44]. See Equation 51.

|  |  |
| --- | --- |
|  | (51) |

Maximum likelihood estimators are asymptotically efficient and normally distributed, meaning that their variance coincides with the Cramér-Rao lower bound and their sampling distribution converges to a normal distribution as the sample size tends to infinity (Tarima & Flournoy, 2019) [45]. The asymptotic standard errors of the estimates for α and β can be obtained by taking the square root of the diagonal elements of the inverse Fisher information matrix (Equation 52 for the α parameter and Equation 54 for the β parameter). These asymptotic standard errors can be used to calculate a Wald-type confidence interval (Equation 53 for the α parameter and Equation 55 for the β parameter), which, being asymptotic, requires a large sample size.

|  |  |
| --- | --- |
|  | (52) |
|  | (53) |

|  |  |
| --- | --- |
|  | (54) |
|  | (55) |

It should be noted that the theoretical asymptotic formulas in Equations 53 and 55 are not typically used in practice by statistical programs to obtain confidence intervals. Instead, a likelihood profile is used. This method does not assume that the likelihood function is symmetric or that standard errors adequately capture the uncertainty in the estimate. Rather, it uses the likelihood function itself to determine the bounds of the confidence interval (Efford, 2020) [46].

The maximum likelihood method, employing an iterative numerical approach, provides a robust framework for estimating the parameters of the Beta distribution. By leveraging numerical techniques, initial estimates based on the method of moments, and asymptotic properties, accurate parameter estimates and their associated standard errors can be obtained. This methodology is widely used with the Beta distribution and is particularly recommended when the assumption of a large sample of independent and identically distributed data holds (Ali et al., 2023) [36].

**4. Testing the Fit of Sample Data to a Beta Distribution with Unknown Parameters**

**4.1 G-test of goodness of fit**

Following McDonald (2014) and Moore (1986) [47‑48], the G-test (Woolf, 1957) [49], along with Williams' continuity correction (Williams, 1976) [50], can be used to evaluate the fit of data from a random sample of size *n* to a Beta distribution with unknown shape parameters (α and β). The number of class intervals can be determined using Moore's rule, which requires a minimum sample size of 45 observations (Moore, 1986) [48]. As this is an asymptotic test, its accuracy increases with larger sample sizes (Berrett & Samworth, 2021) [51]. The G-test is a classic and straightforward method that provides reliable results (McDonald, 2014; Moore, 1986) [47‑48], although more recent approaches, such as the L2-type goodness-of-fit test for the Beta distribution family proposed by Ebner and Liebenberg (2021) [52], have also been developed.

The sample data collected from the continuous quantitative variable X must range from 0 to 1; otherwise, normalization is required. To normalize the data, the sample minimum is subtracted from each value, and the result is divided by the difference between the sample maximum and minimum. Alternatively, the potential minimum and maximum values of variable X can also be used (Equation 56). Both approaches ensure that the data range from 0 to 1. Using the potential minimum and maximum values of variable X assumes that these values are known and well-defined, which may pose a limitation. However, this approach has the advantage of avoiding bias, especially in small samples, ensuring consistency in analyses involving multiple samples, and facilitating the replication of results. Therefore, it is the preferred method, particularly when limited support is available in the data, which follow a Beta distribution (Nikseresht & Amindavar, 2024) [53].

|  |  |
| --- | --- |
|  | (56) |

The null hypothesis of the test states that the sample data for variable X follow a Beta distribution with unknown parameters: H₀: X ~ Beta(α, β). The alternative hypothesis is that the data follow a different distribution: H₁: X ≁ Beta(α, β).

We begin by creating a table with *k* class intervals (CIi, where *i* = 1, 2, ..., *k*) for the *n* sample data points. Following Moore's (1986) rule [48], the frequency is uniform (∀ni = *nCI*), while the amplitude (*ai*) varies. The number of class intervals (*k*) depends on the sample size and is calculated as twice the sample size raised to the power of two-fifths, rounded to the nearest integer (Equation 57). If the sample size divided by*k* is an integer, uniform frequency per class interval is obtained, which should be at least 5. For a sample size of 45, this rule yields 9 class intervals, each with a uniform frequency of 5.Therefore, 45 is the minimum sample size according to this rule.

|  |  |
| --- | --- |
|  | (57) |

If twice the sample size raised to the power of two-fifths is rounded to the nearest integer (*k*), and the sample size divided by theresulting number (*n* / *k*)is a non-integer, then thisquotient is rounded down (*nCI*), leading to an excess frequency: *nexc* = *n* - *k* × ⌊*n*/*k*⌋ (Equation 58). This means that the *n* - *k* × ⌊*n*/*k*⌋ central intervals receive an additional element (*nCI* + 1). If *k* is odd, the central interval is located at position (*k* + 1)/2. To determine which class intervals receive an additional element, start with this central interval and alternate between the left and right intervals. If *k* is even, there are two central intervals, located at positions *k*/2 and *k*/2 + 1. Begin with the central interval in the left half, proceed to the central interval in the right half, and continue alternating left and right until all central intervals that receive an additional element are identified (Moral, 2024) [54].Alternatively, *k* can be the nearest integer greater than or equal to 5 that, when divided by the sample size, results in an integer corresponding to the homogeneous frequency.

|  |  |
| --- | --- |
|  | (58) |

The frequency of each class interval constitutes the observed frequency (*oi*). The alpha and beta parameters of the assumed Beta distribution are then estimated using either the maximum likelihood method or the method of moments. Next, the cumulative probability up to the upper limit of each interval is calculated using the cumulative distribution function of the Beta distribution, based on the estimated shape parameters.

We then calculate the probability of each interval. The probability of the first interval corresponds to the cumulative probability from 0 to its upper limit. For the second through the penultimate intervals, the probability is obtained by subtracting the probability of the previous interval from the cumulative probability up to the upper limit. The probability of the last interval is calculated by subtracting the cumulative probability of the penultimate interval from 1. Finally, the expected frequency (*ei*) is obtained by multiplying the probabilities of the intervals by the sample size.

The test statistic, denoted by G, is calculated from the observed and expected frequencies using the formula presented in Equation 59 [47‑49].

|  |  |
| --- | --- |
|  | (59) |

The continuity correction proposed by Williams (1976)[50], denoted by *q* (Equation 60), is applied to the g-statistic. For this, the g-statistic is divided by *q*, which depends on the sample size and the degrees of freedom (Equation 61).

|  |  |
| --- | --- |
|  | (60) |
|  | (61) |

The sampling distribution of the test statistic (*G* or *GCC*) follows a chi-square distribution with *k* - 3 degrees of freedom. The decision is made based on the right tail. At a significance level of 5%, if the test statistic is less than or equal to the critical value (the 0.95th quantile of a chi-square distribution with *k* - 3 degrees of freedom), the null hypothesis of fit to a Beta distribution is retained; otherwise, it is rejected. Alternatively, the right-tailed probability value of the test statistic can be calculated using a chi-square distribution with *k* - 3 degrees of freedom. If the probability value is greater than or equal to the significance level of 0.05, the null hypothesis is retained; otherwise, it is rejected (McDonald, 2014; Williams, 1976) [47, 50].

The statistical power of the test can be obtained by the complement of the cumulative distribution function of a non-central chi-square distribution. The non-centrality parameter is the value of the test statistic (*g or gcc*), the degrees of freedom are *k* - 3, and it is evaluated at the critical value, which is the 0.95th quantile of a chi-square distribution with *k* - 3 degrees of freedom. This probability value represents the Type II error or the probability of maintaining the null hypothesis when the alternative is false. Its complement provides the power or the probability of rejecting the null hypothesis when the alternative is false(Equation 62). If the null hypothesis is maintained, the power should be less than 0.5 and ideally 0.2 (Rolke & Gongora, 2021) [55].

|  |  |
| --- | --- |
|  | (62) |

If the null hypothesis is rejected, the effect size can be obtained by taking the square root of the quotient between the test statistic (*G or Gcc*) and the product of the sample size and the degrees of freedom, which is known as Cramer's *V* coefficient [56], ranging from 0 to 1. The cutoff points for interpreting effect size are as follows: values of *V* less than 0.1/√*df* indicate a trivial effect size; from 0.1/√df to 0.29/√*df*, small; from 0.3/√*df* to 0.49/√*df*, medium; and greater than or equal to 0.5/√*df*, large, where df represents the degrees of freedom, which in this case are *k*−3 (Cohen, 1988; Fey, Hu, & Delios, 2023) [57‑58]. Refer to Equation 63 for more details.

|  |  |
| --- | --- |
|  | (63) |

**4.2 Q-Q plot**

Another option is to use the quantile-quantile plot (Wickham et al., 2024) [59]. The *n* sample data are sorted in ascending order and assigned an order from 1 to *n*. These values constitute the empirical quantiles, denoted by *x*(*i*). The theoretical quantile order (*pi*) is calculated using the formula shown in Equation 64, which applies the plotting position constant of one-third (Millard and Kowarik, 2024) [60], as recommended by Tukey (1977) [61] in his exploratory data analysis and by the simulation study on the calculation of sample quantiles in statistical packages by Hyndman and Fann (1992) [62].

|  |  |
| --- | --- |
|  | (64) |

This formula is based on the median of the *i*-th order statistic of a random sample of size *n* drawn from a standard uniform distribution U[0, 1]. The sampling distribution of this order statistic follows a Beta distribution with parameters of the form: α = *i* and β = *n* + 1 - *i*. Both parameters are never lower than 1, and when α = β, this value is greater than 1. Therefore, the median of the statistic is: *mdn*(u) = (α - 1/3) / (α + β - 2/3) = (*i* - 1/3) / (*n* + 1/3). We estimate the parameters α and β of the assumed Beta distribution using the *n* sample data, applying either the maximum likelihood method or the method of moments. Using the quantile function of a Beta distribution, with the estimates of α and β as the values of its shape parameters, we evaluate the orders of the theoretical quantiles (*pi*) to obtain the theoretical quantiles, which can be denoted as*xt*(*i*). See Equation 65.

|  |  |
| --- | --- |
|  | (65) |

The theoretical quantiles are plotted on the horizontal (abscissa) axis, and the empirical quantiles are plotted on the vertical (ordinate) axis of a Cartesian plane. A straight line is drawn at a 45-degree angle from the origin. If the coordinate points align along this line, the fit is perfect.

This fit can be quantified by calculating the Pearson product-moment correlation between the empirical and theoretical quantiles. If the correlation is equal to 1, the fit is perfect. The asymptotic 95% confidence interval for the correlation coefficient can be calculated using Fisher's transformation (Pinelis, 2020) [63]. If this interval includes the value 1, the fit is considered perfect in a two-tailed test with a significance level of 5%. Refer to Equation 66 for more details.

|  |  |
| --- | --- |
|  | (66) |

Values of correlations above 0.9 are generally considered very good, indicating that the distribution model fits the data accurately. Therefore, if the confidence interval does not include 1 but does include 0.90, the fit can be interpreted as good (Cohen, 1988) [57], provided that the estimated density curve follows the expected profile of the Beta distribution corresponding to the estimated parameters. See Figures 1 and 2.

**5.Related and limiting distributions**

A Beta distribution with parameters α = β = 1 corresponds to a standard continuous uniform distribution: Beta(1, 1) ≡ U[0, 1] (Gupta&Nadarajah, 2004). [64].

A Beta distribution with parameters α = β = 1/2 corresponds to a standard arcsine distribution with threshold parameters *a* = 0 and *b* = 1: Beta(1/2, 1/2) ≡ Arcsine(0, 1)(Jeffreys, 1946) [65].

A Beta distribution with parameters α = β = 0 corresponds to a Bernoulli distribution with parameter *p* = 1/2 (probability of success): Beta(0, 0) ≡ B(*p* = 1/2). It should be noted that α and β lie outside the parametric space, making this a case of distributional convergence. Its significance lies in its use by Haldane (1932) as a prior distribution to estimate the binomial proportion, ensuring that the Bayesian estimator coincides with the maximum likelihood estimator (Haldane, 1932) [66].

The cumulative distribution function of a binomial distribution with parameters *n* (number of independent trials) and *p* (probability of success) is equivalent to the complementary cumulative distribution function of a Beta distribution with parameters α = *x* + 1 and β = *n* - *x*: *FX*(*x*) = *FY*(*y* = *p*), where X ~ B(*n*, *p*) and Y ~ Beta(α = *x* + 1, β = *n* - *x*). See Equation 67. Additionally, the product of *n* + 1 and the p-th quantile of a Beta distribution with parameters α = *x* + 1 and β = *n* + 1 - *x*, rounded to the nearest integer, corresponds to the value *x* of a binomial distribution with parameters *n* and *p*. See Equation 68. Conversely, the quotient between the value *x* of the binomial distribution B(*n*, *p*) and *n* + 1 approximates the p-th quantile of a Beta distribution with parameters α = *x* + 1 and β = *n* + 1 - *x*. See Equation 69 (Gupta & Kapoor, 2020) [67]. For more details, refer to Table 1, where these equivalences are illustrated using a binomial distribution: B(*n* = 10, *p* = 1/3).

|  |  |
| --- | --- |
|  | (67) |
|  | (68) |
|  | (69) |

**Table 1. Equivalences between the Binomial and Beta distributions applied to an example**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x*X ~B(*n*=10,*p*=1/3 | *FX*(*x|n,p*)X ~B(*n*=10,*p*=1/3) | 1−*FY*(*p*|α,β)Y ~ Beta(α = *x*+1,β = *n*−*x*) | *x* =⌊(*n*+1) ×*QY*(*p* = 1/3|α = *x*+1,β = *n*+1−*x*)⌉ | *y =**QY*(*p* = 1/3|α = *x*+1,β = *n*+1−*x*) | *x*/(*n*+1)X ~B(*n* = 10,*p* = 1/3 |
| 0 | 0.028248 | 0.028248 | ⌊0.350954⌉ = 0 | 0.031905 | ≈ 0 |
| 1 | 0.149308 | 0.149308 | ⌊1.093178⌉ = 1 | 0.099380 | ≈ 0.090909 |
| 2 | 0.382783 | 0.382783 | ⌊1.924608⌉ = 2 | 0.174964 | ≈ 0.181818 |
| 3 | 0.649611 | 0.649611 | ⌊2.802831⌉ = 3 | 0.254803 | ≈ 0.272727 |
| 4 | 0.849732 | 0.849732 | ⌊3.714699⌉ = 4 | 0.337700 | ≈ 0.363636 |
| 5 | 0.952651 | 0.952651 | ⌊4.654983⌉ = 5 | 0.423180 | ≈ 0.454545 |
| 6 | 0.989408 | 0.989408 | ⌊5.622114⌉ = 6 | 0.511101 | ≈ 0.545455 |
| 7 | 0.998410 | 0.998410 | ⌊6.617227⌉ = 7 | 0.601566 | ≈ 0.636364 |
| 8 | 0.999856 | 0.999856 | ⌊7.644767⌉ = 8 | 0.694979 | ≈ 0.727273 |
| 9 | 0.999994 | 0.999994 | ⌊8.715563⌉ = 9 | 0.792324 | ≈ 0.818182 |
| 10 | 1 | - | ⌊9.859576⌉ = 10 | 0.896325 | ≈ 0.909091 |

Note. Beta(α, β) = Beta distribution with shape parameters α and β; B(*n*, *p*) = binomial distribution with parameters *n* (number of independent trials) and *p* (probability of success); *FX(x*) = *P*(X ≤ *x*) = cumulative distribution function of the variable X evaluated at x; *QY*(*p*) = quantile function of the variable Y evaluated at the cumulative probability *p*, or the p-th quantile of Y; ~ denotes "follows"; ⌊⌉ denotes rounding to the nearest integer; ≈ denotes approximately.

If the variable X follows a Beta distribution with shape parameters α and β, its complement follows a Beta distribution with parameters β and α. In other words, it is the mirror-symmetric counterpart: X ~ Beta(α, β)⟹ 1 - X ~ Beta(β, α)(Gupta & Nadarajah, 2004)[64].

If the variable X follows a Beta distribution with shape parameters α and β, then the ratio of X to its complement follows a Beta prime distribution with shape parameters α and β, denoted as Beta'(α, β): X ~ Beta(α, β) ⟹ X / (1 - X) ~ Beta'(α, β). The density function (Equation 70), cumulative distribution function (Equation 71), and descriptive measures including the mathematical expectation or arithmetic mean (Equation 72), mode (Equation 73), variance (Equation 74), and the coefficients of skewness (Equation 75) and excess kurtosis (Equation 76), based on standardized central moments, are provided below for a continuous variable X defined on the interval [0, ∞) that follows a Beta prime distribution (Gupta & Nadarajah, 2004) [64].

|  |  |
| --- | --- |
|  | (70) |
|  | (71) |
|  | (72) |
|  | (73) |
|  | (74) |
|  | (75) |
|  | (76) |

If the variable X follows a Beta distribution with shape parameters α and β, then the ratio of the complement of X to X follows a Beta prime distribution with shape parameters β and α: X ~ Beta(α, β) ⟹ (1 - X) / X ~ Beta'(β, α)(Johnson, Kotz, & Balakrishnan, 1995)[68].

If the variable X follows a Beta distribution with shape parameters α/2 and β/2, then the ratio, where the numerator is the product βX and the denominator is the product α(1 − X), follows a Snedecor-Fisher F distribution (Snedecor & Cochran, 1989) [69] with degrees of freedom α and β: X ~ Beta(α/2, β/2) ⟹ βX / [α(1 − X)] ~ F(α, β).

If X follows a Snedecor-Fisher F distribution with degrees of freedom α and β, then the ratio of the product αX to the linear combination αX + β follows a Beta distribution with shape parameters α/2 and β/2: X ~ F(α, β) ⟹ αX / (αX + β) ~ Beta(α/2, β/2)(Chattamvelli & Shanmugam, 2022) [43].

If the variable X follows a Beta distribution with shape parameters α and β, then the negative natural logarithm of X follows an exponential distribution with rate parameter λ = α, which is the inverse of the scale: X ~ Beta(α, β) ⟹ -ln(X) ~ Exp(λ = α)(Chattamvelli & Shanmugam, 2022) [43].

If X follows an exponential distribution with rate parameter λ (the inverse of the scale), then the exponential function with base e and exponent −X follows a Beta distribution with parameters α = λ and β = 1: X ~ Exp(λ) ⟹ e-X ~ Beta(α = λ, β = 1) (Chattamvelli & Shanmugam, 2022) [43].

If X follows a standard uniform distribution (with threshold parameters *a* = 0 and *b* = 1) and the constant*a* is a positive real number, then the power of X raised to the inverse of a follows a Beta distribution with parameters α = *a* and β = 1: X ~ U[0, 1] and *a*> 0 ⇒ X1/*a* ~ Beta(α = *a*, β = 1) (Gupta & Kapoor, 2020)[67].

If the variable X follows a Beta distribution with parameters α = β = 3/2, then the three-factor product 2 × *R ×* (X − 1) follows a semicircular distribution with parameter *R* (Wigner, 1958) [70]: X ~ Beta(α = 3/2, β = 3/2) ⟹ 2*R*(X − 1) ~ Semicircular(*R*).

If the random variable X follows a Cauchy distribution with location parameter x₀ and scale parameter γ, then the inverse of 1 + X² follows a Beta distribution with parameters α = 1/2 and β = 1/2: X ~ Cauchy(0, 1) ⇒ 1 / (1 + X²) ~ Beta(α = 1/2, β = 1/2)(Chattamvelli & Shanmugam, 2022) [43].

Let X be a random variable of parameters α = 1 and β = *n*. When *n* tends to infinity, the variable X converges to an exponential distribution of rate parameter or inverse of scale λ = 1 (Gupta & Nadarajah, 2004) [64].

Let X be a random variable following a Beta distribution with parameters α = 1 and β = *n*. As *n* approaches infinity, the variable *n*X converges to an exponential distribution with rate parameter λ = 1 (the inverse of the scale parameter): X ~ Beta(1, *n*), limn→∞(*n*X) = Y ~ Exp(λ = 1)(Chattamvelli & Shanmugam, 2022) [43].

Let X be a random variable following a Beta distribution with shape parameter α and shape parameter β = *n*. When *n* tends to infinity, the variable X converges to a Gamma distribution with shape parameter α and rate parameter λ = 1 (the inverse of the scale parameter): X ~ Beta(α, *n*), limn→∞(nX) = Y~ Gamma(α, λ = 1)(Chattamvelli & Shanmugam, 2022) [43].

Let X be a random variable following a Beta distribution with shape parameters α and β. As both parameters approach infinity and are approximately equal, X converges to a normal distribution with location parameter μ = α / (α+β) and scale parameter σ = √[αβ / ((α+β)(α+β+1))]. Consequently, the standardized variable converges to a standard normal distribution *N*(0, 1), under these conditions (Gupta & Kapoor, 2020) [67]. See Equation 77.

|  |  |
| --- | --- |
|  | (77) |

This distributional convergence is faster when α = β = *n*. In this case, the variable X converges to a normal distribution with location parameter μ = 1/2 and scale parameter σ = 1 / √(8*n* + 4): X ~ Beta(α = n, β = n), limn→∞(X) = Y ~ N(1/2, 1/(8n+4)) (Gupta & Kapoor, 2020) [67].

Let X be a variable with a *U*[0, 1] distribution. A random sample of size *n* is drawn: {*u₁*, *u₂*, ..., *uᵢ*, ..., *uₙ*}. The *n* data points are ordered in ascending order and assigned an index *i*, ranging from 1 to *n*. The statistic at the i-th order, denoted by *uᵢ*, follows a Beta distribution with parameters α = *i* and β = *n* + 1 - *i*. Thus, the sample minimum follows a Beta(1, *n*) distribution, and the sample maximum follows a Beta(*n*, 1) distribution(Stuart & Ord, 2010) [71]. Refer to Equation 78 for more details.

|  |  |
| --- | --- |
|  | (78) |

Let X and Y be two independent random variables. If X follows a Gamma distribution with shape parameter α and rate parameter λ, and Y follows a Gamma distribution with shape parameter β and rate parameter λ, then the ratio X / (X + Y) follows a Beta distribution with shape parameters α and β: X ⊥⊥ Y, X ~ Gamma(α, λ), and Y ~ Gamma(β, λ) ⟹ X / (X + Y) ~ Beta(α, β)(Chattamvelli & Shanmugam, 2022) [43].

Let X and Y be two independent random variables. If X follows a chi-squared distribution with α degrees of freedom and Y follows a chi-squared distribution with β degrees of freedom, then the ratio X / (X + Y) follows a Beta distribution with parameters α/2 and β/2: X ⊥⊥ Y, X ~ χ²(α), and Y ~ χ²(β) ⟹ X / (X + Y) ~ Beta(α/2, β/2)(Chattamvelli & Shanmugam, 2022) [43].

Let X and Y be two independent random variables. If X follows a non-central chi-squared distribution with α degrees of freedom and non-centrality parameter λ, and Y follows a chi-squared distribution with β degrees of freedom, then the ratio X / (X + Y) follows a non-central Beta distribution with parameters α/2, β/2, and non-centrality parameter λ: X ⊥⊥ Y, X ~ NCχ²(α, λ), and Y ~ χ²(β) ⟹ X / (X + Y) ~ NCBeta(α/2, β/2, λ) (Orsi, 2017)[72].

If the random variable X follows a binomial distribution with parameters *n* (number of trials) and *p* (probability of success), where *p* is itself a random variable following a Beta distribution with parameters α and β, then X follows a Beta-Binomial distribution with parameters *n*, α, and β: X ~ B(*n*, *p*) and *p* ~ Beta(α, β) => X ~ BetaBin(n, α, β). The probability mass function (Equation 79), cumulative distribution function (Equation 80), descriptive measures such as the expected value or arithmetic mean (Equation 81), non-central second moment (Equation 82), variance (Equation 83), skewness coefficient (Equation 84), and excess kurtosis (Equation 85) based on standardized central moments, as well as method-of-moments estimators for the parameters alpha (Equation 86) and beta (Equation 87), are provided below for the discrete variable X. This variable is supported on the finite set {0, 1, ..., *n*} and follows a Beta-Binomial distribution with parameters *n*, α, and β (Navarro & Perfors, 2005)[73].

|  |  |
| --- | --- |
|  | (79) |
|  | (80) |
|  | (81) |
|  | (82) |
|  | (83) |
|  | (84) |
|  | (85) |
|  | (86) |
|  | (87) |

As we can see in Equation 80, the generalized hypergeometric function 3F2(*a*; *b*; *z*), also known as the pFq-type hypergeometric function, appears in the cumulative distribution function of a discrete variable X that follows a Beta-Binomial distribution BetaBin(*n*, α, β) (Navarro & Perfors, 2005)[73]. This function can be computed using R's hypergeo package. In this context, *a* is the vector of three numerator parameters (*a1*, *a2,a3*), *b* is the vector of two denominator parameters (*b1*, *b2*), and *z* is the argument of the function. The script for its calculation is provided below (R Core Team, 2024c) [29].

# R script to calculate thegeneralized hypergeometric function.

# Load the package hypergeo.

library(hypergeo)

# Define parameters.

n <- 5 # Number of trials.

alpha <- 2 # Alpha parameter of the Beta-Binomial distribution.

beta <- 3 # Beta parameter of the Beta-Binomial distribution.

x <- 0.4 # Variable x for the hypergeometric function.

# Define the parameters for 3F2(a; b; z).

a <- c(1, -x, n - x + beta) # Numerators.

b <- c(n - x + 1, 1 - x - alpha) # Denominators.

z <- x

# Compute 3F2(a; b; z) and display the result.

result <- genhypergeo(U = a, L = b, z = z)

cat("The value of 3F2(a = 1,", -x, ",", n - x + beta, "; b =",

n - x + 1, ",", 1 - x - alpha, "; z = 1) is", result, ".\n")

# The value of 3F2(a = 1, -0.4 , 7.6 ; b = 5.6 , -1.4 ; z = 1) is 0.5340251.

If the random variable X follows a negative binomial distribution with stopping parameter *r* (number of failures) and success probability parameter *p*, where *p* itself is a random variable with a Beta distribution characterized by parameters α and β, then X follows a Beta negative binomial distribution with parameters r, α and β: X ~ BN(*r*, *p*) and *p* ~ Beta(α, β) ⟹ X ~ BetaNB(*r*, α, β). The probability mass function (Equation 88), cumulative distribution function (Equation 89), expectation or arithmetic mean (Equation 90), variance (Equation 91), and coefficient of skewness based on the third standardized central moment (Equations 92) are provided below for the discrete variable X, which is supported on the infinite set {0, 1, ...} and follows a beta negative binomial distribution with parameters *r*, α, and β (Wanas, & Al-Ziadi, 2021) [74].

|  |  |
| --- | --- |
|  | (88) |
|  | (89) |
|  | (90) |
|  | (91) |
|  | (92) |

The Dirichlet distribution is the multivariate generalization of the Beta distribution, just as the Wishart distribution is the multivariate generalization of the gamma distribution (Ng, Tian, & Tang (2011)[75].

Another generalization of the Beta distribution is the four-parameter Beta distribution, which introduces two threshold parameters, *a* and *c*, in addition to the shape parameters α and β. Consequently, the support of the distribution shifts from the interval [0, 1] to the bounded interval [*a*, *c*], where *a*, *c*∈ R and *c*>*a*. This distribution is denoted as Beta(α, β, *a*, *c*)(Johnson, Kotz, & Balakrishnan, 1995) [68].

Within the family of four-parameter Beta distributions, the PERT (Program Evaluation and Review Technique) distribution, developed by Clark (1962) [76], stands out. This distribution is defined by three parameters: the threshold parameter for the minimum value *a*, the modal value *b* (the peak of the distribution), and the threshold parameter for the maximum value *c*. The shape parameters α and β are functions of these three parameters, as the distribution is constrained such that its mathematical expectation is (*a* + 4*b* + *c*) / 6, and its standard deviation is approximately one-sixth of the range (Clark, 1962) [76].

The PERT distribution was created as an alternative to the triangular distribution for risk analysis when only the most likely value and an estimate of the possible range of a variable are known. Compared to the triangular distribution, the PERT distribution exhibits a smoother profile and assigns greater weight to the mode in the calculation of the mean (Khan, Bickel, & Hammond, 2023)[77].

The following are provided: the probability density function (Equation 93), the cumulative distribution function (Equation 94), quantile function (Equation 95), and descriptive measures, including the mathematical expectation or arithmetic mean (Equation 96), median (Equation 97), mode (Equation 98), variance (Equation 99), skewness coefficient based on the third standardized central moment (Equation 100), and excess kurtosis based on the fourth standardized central moment (Equation 101) of a PERT distribution with minimum value *a*, modal value *b*, and maximum value *c*(Clark, 1962; Khan et al., 2023) [76‑77]

|  |  |
| --- | --- |
|  | (93) |
|  | (94) |
|  | (95) |
|  | (96) |
|  | (97) |
|  | (98) |
|  | (99) |
|  | (100) |
|  | (101) |

**6. Applications of the Beta distribution**

The Beta distribution is highly valuable as a probability model for binomial proportions. Even bounded data that are not originally proportions can be transformed into proportions using the min-max transformation (Gupta & Kapoor, 2020) [64], as shown in Equation56: *zi* = (*xi* - *min*(X)) / (*max*(X) - *min*(X)).Examples of its applications include traffic analysis of anonymous communications in mobile networks (Wang, Pei, & Wang, 2013) [78], energy demand forecasting within a dynamic structural break detection framework (Nikseresht & Amindavar, 2024) [53], diabetes classification (Wee et al., 2023) [79], and psychological disorder diagnosis (Heck et al., 2023) [80].

One of the most important applications of the Beta distribution is estimating the binomial proportion through a Bayesian approach (Kaplan, 2023; García-García et al. 2022) [1,14]. After conducting *n* independent repetitions of a random experiment with two possible outcomes (0 = failure and 1 = success) and observing *x* successes, this method adopts a general framework. The framework specifies three options for the two shape parameters of the Beta distribution, which serves as the prior distribution: Bayes-Laplace: α = β = 1, Jeffreys: α = β = 1/2, and Haldane: α = β = 0(Seo & Kim, 2022)[81].

Prior distribution: Beta(αprior, βprior) and prior point estimator: αprior / (αprior + βprior).

Posterior distribution: *p*∈*P*∼Beta(αposterior = *x* + αprior, βposterior = *n* - *x* + βprior).

Posterior estimator:= *E*(*P*) = αposterior / (αposterior + βposterior).

Confidence interval at (1 - α) × 100. Refer to Equation 102, where I-1α/2(αposterior, βposterior) representsregularized incomplete Beta function evaluated at α/2 with parameters αposterior and βposterior, and I-11-α/2(αposterior, βposterior) denotes this function evaluated at 1 - α/2.

|  |  |
| --- | --- |
|  | (102) |

Prior distribution of Bayes (1863) [13] and Laplace (1820) [82]: Beta(1, 1) ≡ *U*[0, 1]and prior point estimator: 1/2.

Posterior distribution: *p*∈*P*∼ Beta(*x* + 1, *n* - *x* + 1).

Posterior estimator: = *E*(*P*) = (*x* + 1) / (*n* + 2).

Confidence interval at (1 - α) × 100 (Equation 103).

|  |  |
| --- | --- |
|  | (103) |

Jeffreys' (1946) [65] prior distribution: Beta(α = 1/2, β = 1/2) ≡ Arcsine(0, 1) and prior point estimator: 1/2.

Posterior distribution: *p*∈*P*∼ Beta(*x* + 0.5, *n* - *x* + 0.5).

Posterior estimator: = *E*(*P*) = (*x* + 0.5) / (*n* + 1).

Confidence interval at (1 - α) × 100 (Equation 104).

|  |  |
| --- | --- |
|  | (104) |

Haldane's (1932) [66] prior distribution: Beta(α = 0, β = 0) ≡ B(*p* = 1/2)and prior point estimator: *E*(X) = 1/2, when X ∼B(*p* = 1/2).

Posterior distribution: Beta(x, *n* - *x*).

Posterior estimator: = *E*(X) = *x* / *n*.

Confidence interval at (1 - α) × 100 (Equation 105).

|  |  |
| --- | --- |
|  | (105) |

The calculation of sample quantiles represents the other side of the coin in probability estimation, as it seeks the value of X that accumulates the probability *p* (the quantile order) in the (unknown) distribution of X (Prendergast, Dedduwakumara, & Staudte, 2024) [83]. Probability spans a continuous range from 0 to 1. If the p-value is disregarded and random values are drawn uniformly from this range, these probability values follow a standard uniform distribution or a Beta distribution with α = 1 and β = 1. Refer to Equation 106 for further details.

|  |  |
| --- | --- |
|  | (106) |

If the probability values are arranged in ascending order, the probability at position *i* follows a Beta distribution with parameters α = *i* and β = *n* + 1 - *i*.The mathematical expectation, or arithmetic mean, of this distribution serves as the point estimator for the quantile of order *p* (*cp*). By isolating *i*, the position of the quantile can be identified among the *n* sample data points arranged in ascending order (*x*(*i*)). Consequently, the value of *i* must be an integer (Pham, 2023) [84]. Refer to Equation 107 for further details.

|  |  |
| --- | --- |
|  | (107) |

If the result is a decimal number, it is rounded down to ⌊*i*⌋, and a linear interpolation rule is applied to determine the quantile of X between the values *x*(⌊*i*⌋) and *x*(⌊*i*⌋+1) values (Equation 108).

|  |  |
| --- | --- |
|  | (108) |

**7. ExampleS of Calculation**

**7.1EXample 1. CALCULATING PROBABILITIES AND DESCRIPTIVE MEASURES of a beta distribution with known parameters**

A condom brand analyzed the distribution of the annual proportion of accidental breakage when using its standard condom and concluded that it follows a Beta distribution with parameters alpha = 2 and beta = 7. Based on this model, calculate the following: 1) the probability that the proportion of accidental breakage is less than 0.1, between 0.15 and 0.22, and greater than 0.25; 2) the mathematical expectation (arithmetic mean), geometric mean, harmonic mean, median, and mode of this distribution; 3) the variance, standard deviation, mean absolute deviation, and entropy of the distribution; and 4) the skewness and excess kurtosis based on standardized central moments. Additionally, plot the density and cumulative distribution functions of the distribution on a graph.

X = proportion of accidental breakage per year of the standard condom: X ~ Beta(α = 2, β = 7) and x ∈ [0, 1]. Appendix 1 provides the script to perform all the calculations.

The probability that the proportion of accidental breakage is less than 0.1 is calculated using Equation 5.

The probability that the proportion of accidental breakage is in the range [0.15, 0.22] is calculated by subtracting the cumulative distribution function value at the larger point from the value at the smaller point, as obtained using Equation 5.

The probability that the proportion of accidental breakage is less than greater than 0.25 is calculated using Equation 5.

The mathematical expectation, or arithmetic mean, denoted as *E*(X) and calculated using Equation 10; the geometric mean, denoted as *G*(X) and calculated using Equation 20; the harmonic mean, denoted as *H*(X) and calculated using Equation 21; the median, denoted as *Mdn*(X) and calculated using Equation 18; and the mode, denoted as *Mo*(X) and calculated using Equation 19, are obtained as measures of central tendency for the Beta(α = 2, β = 7) distribution.

The variance, denoted as σ2(X) and calculated using Equation 24, the standard deviation, denoted as σ(X) and calculated using Equation 32, the mean absolute deviation, denoted as *MAD*(X) and calculated using Equation 33, and the Shannon’s entropy, denoted as Η(X) and calculated using Equation 34, are obtained as measures of variability.

Pearson’s coefficient of skewness based on the third standardized central moment, denoted as β1 and calculated using Equation 36, and Pearson’s excess kurtosis based on the fourth standardized central moment, denoted as β2−3 and calculated using Equation 37, were obtained as measures of shape.

The probability density function (PDF) and cumulative distribution function (CDF) are plotted together. The PDF of this unimodal distribution exhibits positive skewness and represents leptokurtosis. See Figure 3.



**Figure 3. Probability density function and cumulative distribution function for the proportion of accidental breakage per year of the standard condom**

**7.2ExampLe 2. Parameter Estimation Using the Method of Moments, Descriptives, and ProbabilitYCALCULATION**

Twenty random samples were collected from 100 participants across different cities in a country (cluster sampling), and the following proportions of accidental condom breakage per year were obtained (Table 2). Respondents were asked whether they had experienced a condom break during vaginal or anal penetration in the past year. The data are known to follow a Beta distribution. Estimate its parameters using the method of moments. Calculate the mathematical expectation, median, and mode as descriptive measures of central tendency. Obtain the variance as a measure of variability. Additionally, calculate the skewness and excess kurtosis based on standardized central moments as measures of shape. Plot the probability density and cumulative distribution functions on a graph. Finally, indicate the probability that there will be less than one-fifth accidental breakage, between one-twentieth and one-tenth accidental breakage, and more than one-fourth accidental breakage.

Appendix 2 contains the R script for performing the calculations for this example, either by estimating parameters using the method of moments or by using these estimates as initial values to obtain the maximum likelihood estimates (R Core Team, 2024b)[28].

**Table 2. Data in random order, squared differential scores with respect to the mean, and logarithmic transformation of the data and its complement.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *i* | *xi* | (*xi*-*mx*)2 | ln(*xi*) | ln(1-*xi*) |
| 1 | 0.12354423 | 0.01556620 | -2.09115605 | -0.13186904 |
| 2 | 0.44261002 | 0.03775296 | -0.81506621 | -0.58449014 |
| 3 | 0.08595943 | 0.02635733 | -2.45387984 | -0.08988032 |
| 4 | 0.42195273 | 0.03015221 | -0.86286199 | -0.54809963 |
| 5 | 0.24214223 | 0.00003803 | -1.41823000 | -0.27725955 |
| 6 | 0.42195046 | 0.03015142 | -0.86286737 | -0.54809570 |
| 7 | 0.18675247 | 0.00378918 | -1.67797123 | -0.20671975 |
| 8 | 0.21126755 | 0.00137206 | -1.55462994 | -0.23732812 |
| 9 | 0.17104856 | 0.00596915 | -1.76580779 | -0.18759370 |
| 10 | 0.16972457 | 0.00617548 | -1.77357833 | -0.18599779 |
| 11 | 0.43533068 | 0.03497718 | -0.83164935 | -0.57151499 |
| 12 | 0.23879670 | 0.00009048 | -1.43214272 | -0.27285481 |
| 13 | 0.11629033 | 0.01742888 | -2.15166537 | -0.12362670 |
| 14 | 0.15349581 | 0.00898951 | -1.87408201 | -0.16664013 |
| 15 | 0.14122295 | 0.01146738 | -1.95741543 | -0.15224594 |
| 16 | 0.19140420 | 0.00323814 | -1.65336786 | -0.21245612 |
| 17 | 0.29961921 | 0.00263276 | -1.20524291 | -0.35613111 |
| 18 | 0.33053753 | 0.00676156 | -1.10703507 | -0.40128017 |
| 19 | 0.19892422 | 0.00243884 | -1.61483133 | -0.22179973 |
| 20 | 0.38360256 | 0.01830440 | -0.95814826 | -0.48386333 |
| Σ | 4.96617644 | 0.26365313 | -30.06162905 | -5.95974676 |
| *M* | 0.24830882 | 0.01318266 | -1.50308145 | -0.29798734 |

Note. *i* = random order of the data; xᵢ = datum in random order *i*; (xᵢ - mₓ)² = squared differential score with respect to the mean; ln(xᵢ) = logarithmic transformation of the data; and ln(1 - xᵢ) = logarithmic transformation of the complement of the data.

Estimation of parameters α and β by the method of momentsusing Equation 44.

The calculations for the mathematical expectation, denoted as *E*(X) and derived using Equation 10; the median, denoted as *Mdn*(X) and derived using Equation 18; the mode, denoted as *Mo*(X) and derived using Equation 19; the variance, denoted as σ²(X) and derived using Equation 24; Pearson’s coefficient of skewness based on the third standardized central moment, denoted as β₁ and derived using Equation 36; and Pearson’s excess kurtosis based on the fourth standardized central moment, denoted as β₂-3 and derived using Equation 37, for the Beta distribution with estimated parameters: α = 3.0917 and β = 9.3593, are presented below.

Figure 4 shows the probability density function (PDF) and cumulative distribution function (CDF) of the random variable X, representing the percentage of accidental condom breakage. The variable follows a Beta distribution with estimated parameters: α = 3.092 and β = 9.3593.



**Figure 4. Plot of probability density function (PDF) and cumulative distribution function (CDF) of X ~ Beta(α = 3.0917, β = 9.3593)**

Probability that accidental breaks are less than one-fifth (Equation)is calculated using Equation 5.

The probability of accidental breakage ranging from one-twentieth to one-tenth is calculated by subtracting the cumulative distribution function value at the larger point from the value at the smaller point, as obtained using Equation 5.

Probability that accidental breaks are greater than one quarteris calculated using Equation 5.

**7.3 example 3. ESTIMATING PARAMETERS USING MAXIMUM LIKELIHOOD**

Using the data from Example 2, estimate the two shape parameters of the Beta distribution both pointwise and intervalwise at the 95% confidence level using maximum likelihood. Additionally, provide the maximum value of the natural logarithm of the likelihood function, as well as the information and covariance matrices of the parameters.

See Appendix 3 for the R script to perform the maximum likelihood estimation. The limited-memory BFGS algorithm is used for the iterative optimization procedure (Byrd et al., 1995) [42], as it is the most widely recommended approach (R Core Team, 2024b) [28]. The resulting output is as follows:

Alpha estimate: α = 3.550721.

Beta estimate: β = 10.72127.

Standard errors:

se(alpha) = 1.07633.

se(beta) = 3.407696.

95% confidence interval for alpha: [1.8509, 6.1159]

95% confidence interval for beta: [5.3476, 18.848]

The maximum value of the natural logarithm of the likelihood function can be obtained using Equation 46.

The information matrix (Equation 50) provided by R for this estimation is as follows:

The inverse of the information matrix is the parameter covariance matrix (Equation 51) shown below:

The standard error is obtained using Equation 52, and the asymptotic confidence interval is calculated using Equation 54 for the α estimate.

The standard error is obtained using Equation 53 and the asymptotic confidence interval is calculated using Equation 55 for theβ estimate.

The 95% asymptotic confidence intervals for the α and β parameters (Equations 53 and 55, respectively) differ slightly from the estimates obtained through the likelihood profile procedure in R.

**7.4 Testing the Fit Using the G-Test and QQ Plot**

Test whether the distribution of sample data representing the success rates of 45 psychotherapy centers treating depression follows a Beta distribution with unknown parameters, using Woolf's (1957) G-test [49] with Williams' (1976) [50] continuity correction and a quantile-quantile plot.

x <- (0.8395, 0.7899, 0.3563, 0.7369, 0.7256, 0.6556, 0.9187, 0.8562, 0.8232, 0.6691, 0.7292, 0.8388, 0.6472, 0.9662, 0.5982, 0.8271, 0.8999, 0.7879, 0.8956, 0.8615, 0.8482, 0.9506, 0.5631, 0.7208, 0.9069, 0.4467, 0.6634, 0.8003, 0.5479, 0.5524, 0.5674, 0.6014, 0.6341, 0.7612, 0.8020, 0.8135, 0.8572, 0.7862, 0.9187, 0.4601, 0.9054, 0.8169, 0.8262, 0.5782, 0.7650)

Running the R script listed in Appendix 4 yields the following results: Table 3, which shows the observed and expected frequencies by class interval for the G-test calculation; Figure 5, which presents the QQ plot; and Figure 6, which shows the estimated density curve. Density estimation was performed using Epanechnikov’s parabolic kernel [85] with the Sheather-Jones bandwidth [86], as it provides the curve closest to the distribution of the empirical data (Chen, 2017) [87]. In addition to the goodness-of-fit test using Woolf's G-test [49] with Williams' continuity correction [50], the correlation between the theoretical and empirical quantiles, along with its 95% asymptotic confidence interval [41], is also shown.

Estimation of the parameters alpha and beta using the method of moments:

Alpha estimate:α = 5.9224 (Equation 44).

Beta estimate: β = 2.0292 (Equation 44).

Moore's rule:

Number of intervals: k = 9 (Equation 57).

Common frequency per interval: n\_CI = 5 (Equation 58).

Excess frequency (one per central interval): n\_exc = 0 (Equation 58).

**Table 3. Observed and expected frequencies by class interval for the G-test calculation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index*i* | Classinterval | Observedfrequency (Oi) | Expectedfrequency (Ei) | Oi × ln(Oi/Ei) |
| 1 | [0.3563, 0.5619] | 5 | 4.973538 | 0.0265324 |
| 2 | (0.5619, 0.6268] | 5 | 3.835350 | 1.3258858 |
| 3 | (0.6268, 0.7036] | 5 | 6.458694 | -1.2799463 |
| 4 | (0.7036, 0.7633] | 5 | 6.419845 | -1.2497800 |
| 5 | (0.7633, 0.8011] | 5 | 4.525576 | 0.4984657 |
| 6 | (0.8011, 0.8265] | 5 | 3.157137 | 2.2988615 |
| 7 | (0.8265, 0.8564] | 5 | 3.706254 | 1.4970810 |
| 8 | (0.8564, 0.9056] | 5 | 5.650358 | -0.6114047 |
| 9 | (0.9056, 0.9662] | 5 | 4.895224 | 0.1058897 |

Note. *i* = class interval index, *Oi* = observed frequency, *Ei* = expected frequency, and O\_LN\_O\_E = *Oi* × ln(*Oi* / *Ei*)

G-test result:

G statistic without continuity correction: g = 5.2232 (Equation 59).

Williams’ continuity correction: q = 1.0494 (Equation 60).

Williams’ continuity-corrected G-test statistic: g\_cc = 4.9774 (Equation 61).

P-value = 0.546718.

The null hypothesis that the data fit a Beta distribution is not rejected at a significance level of 0.05 using Woolf’s G-test.

The right-tailed statistical power for the alternative hypothesis of non-normality for the G-test with continuity correction: ϕ = 0.3347 (Equation 62).

Effect size through Cramer’s V coefficient: V = 0.1358 (Equation 63).

Interpretation of effect sizebased on Cohen (1988) [57] (Equation 63):

- Trivial effect size: V < 0.0408

- Small effect size: V from 0.0408 to < 0.1184

- Medium effect size: V from 0.1184 to < 0.2

- Large effect size: V >= 0.2

Based on the calculated V = 0.1358, the effect size is classified as medium.

Consistent with the result of the test, Figure 5 shows a 45-degree line plot of the ordered pairs of theoretical quantiles (x-axis) and empirical quantiles (y-axis), indicating a good fit to the theoretical distribution model (Beta distribution with estimated parameters: α = 5.9224 and β = 2.0292). Specifically, Pearson’s correlation is very close to 1. Although its 95% asymptotic confidence interval does not include 1, it does encompass the value 0.9. This confidence interval was calculated using Equation 66, and the order of theoretical quantiles (*pi*) was determined using Equation 64, which corresponds to the median of the ith quantile of a random sample of size *n* drawn from a *U*[0, 1] distribution.

Pearson’s correlation coefficient between empirical and theoretical quantiles: r = 0.9945.

The 95% confidence interval using Fisher's transformation: [0.990, 0.997].



**Figure 5. QQ plot**

The curve with densities estimated using the Epanechnikov's kernel [85] and the Sheather-Jones bandwidth [86] shows a concave profile (α and β > 1) with asymmetry toward the left tail (α > β), corresponding to a Beta distribution with an alpha estimate of 5.9224 and a beta estimate of 2.0292, fitting well with the expected density profile. See Figure 6.



**Figure 6. Theorical and empirical density curves**

**8. Conclusion**

Apart from its application in binomial ratio and quantile estimation, the Beta distribution is a versatile model well-suited for calculating probabilities in independent experiments involving the evaluation of a dichotomous variable, where the success or case ratio is recorded. Examples include multiple clinical trials or epidemiological registries. Additional applications are found in subjective logic, Bayesian decision-making, and cognitive psychology. The Beta distribution is also applicable when data can be transformed into a continuum ranging from 0 to 1 using the min-max transformation. This transformation involves subtracting the variable's minimum value from each data point and dividing the result by the range (the difference between the maximum and minimum values). This approach is commonly used in contexts such as income distributions, hazard functions, stock returns, and insurance losses.

When applying this probability model, it is necessary to estimate its two shape parameters, which determine the distribution's skewness and kurtosis. In addition, the central tendency and variability also depend on these two parameters. Both shape parameters must be greater than 0. Estimation is straightforward using the method of moments, which relies on the sample mean and variance, but somewhat more complex—yet more efficient—when employing the maximum likelihood method. Since the mean and variance are finite, the distribution satisfies the central limit theorem, making asymptotic confidence intervals valid for estimating its descriptive statistics and parameters.

The fit of a continuous variable's data to this probability model can be easily assessed using the chi-square test and the likelihood ratio test, in addition to visually inspecting the density curve and the quantile-quantile plot. The density curve should exhibit a continuous profile and be bounded within the range [0, 1]. If the bounds are wider or narrower, the data may correspond to the four-parameter Beta distribution, which includes two threshold parameters in addition to the two shape parameters.

When its two shape parameters are equal and greater than 0, the Beta distribution's density curve is symmetric and concave, converging to the normal distribution as the parameters increase. If the parameters are equal but less than 1, the density curve is symmetric and convex. When both parameters are equal to 1, the density profile is rectangular. If the parameters are very close to 0, the distribution approximates the Bernoulli distribution with a success probability of one-half.

When alpha is greater than beta and both parameters are greater than 1, the density curve is concave with left-tailed asymmetry. However, if beta is less than 1, the concave curve exhibits a vertical asymptote at 1. Conversely, when alpha is less than beta and both parameters are greater than 1, the density curve is convex with right-tailed asymmetry. If alpha is less than 1, the convex curve has a vertical asymptote at 0, resembling an exponential distribution.

For this reason, the Beta distribution is also referred to as the multiform distribution, reflecting its versatility, particularly in its generalized four-parameter form. Using the min-max transformation, this form simplifies to the two-parameter Beta distribution.

**References**

1. Kaplan, D. (2023). Bayesian statistics for the social sciences. New York: Guilford Publications.
2. Yin, X., & Yin, D. S. (2024). Transformer-based parameter estimation in statistics. Mathematics, 12(7), 1040.

<https://doi.org/10.3390/math12071040>

1. Geissinger, E. A., Khoo, C. L., Richmond, I. C., Faulkner, S. J., & Schneider,D. C. (2022). A case for beta regression in the natural sciences. Ecosphere, 13(2), e3940.

<https://doi.org/10.1002/ecs2.3940>

1. Pokhriyal, N., Valentino, B. A., & Vosoughi, S. (2023). Quantifying participation biases on social media. EPJ Data Science, 12(1), 26.
<https://doi.org/10.1140/epjds/s13688-023-00405-6>
2. Papas, E. B. (2021). The global prevalence of dry eye disease: A Bayesian view. Ophthalmic and Physiological Optics, 41(6), 1254-1266.
[https://doi.org/10.1111/opo.12888
PMid:34545606](https://doi.org/10.1111/opo.12888.%20PMid%3A34545606)
3. Stahl, A., Sukgen, E. A., Wu, W. C., Lepore, D., Nakanishi, H., Mazela, J., ... & FIREFLEYE Study Group. (2022). Effect of intravitreal aflibercept vs laser photocoagulation on treatment success of retinopathy of prematurity: the FIREFLEYE randomized clinical trial. Journal of the American Medical Association, 328(4), 348-359.

<https://doi.org/10.1001/jama.2022.10564>.
PMid:35881122
PMCid:PMC9327573

1. Braouezec, Y., & Cagnol, J. (2023). A lattice approach to the beta distribution induced by stochastic dominance: Theory and applications. Journal of the Operational Research Society, 74(6), 1424-1442.
<https://doi.org/10.1080/01605682.2022.2096500>
2. Santana-e-Silva, J. J., Cribari-Neto, F., & Vasconcellos, K. L. (2022). Beta distribution misspecification tests with application to Covid-19 mortality rates in the United States. Plos One, 17(9), e0274781.

https://doi.org/10.1371/journal.pone.0274781
PMid:36126077
PMCid:PMC9488837

1. Bertelsen, T. B., Hoffart, A., Rekdal, S. S., & Zahl-Olsen, R. (2022). Bayes factor benefits for clinical psychology: review of child and adolescent evidence base. F1000Research, 11.

<https://doi.org/10.12688/f1000research.76842.2>
PMid:37809055
PMCid:PMC10558984

1. Aljohani, H. M. (2024). Advances in medical data modeling: A new logarithmic beta generated family of distributions with theory and inference. Alexandria Engineering Journal, 102, 339-360.

https://doi.org/10.1016/j.aej.2024.05.058

1. Hahn, E. D. (2022). The tilted beta-binomial distribution in overdispersed data: Maximum likelihood and bayesian estimation. Journal of Statistical Theory and Practice, 16(3), 43.

<https://doi.org/10.1007/s42519-022-00271-x>

1. Gelman, A., & Vehtari, A. (2021). What are the most important statistical ideas of the past 50 years?. Journal of the American Statistical Association, 116(536), 2087‑2097.

<https://doi.org/10.1080/01621459.2021.1938081>

1. Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical Transitions of the Royal Society of London, 53(1763), 370‑418.

<https://doi.org/10.1098/rstl.1763.0053>

1. García-García, J. I., Fernández Coronado, N. A., Arredondo, E. H., & Imilpán Rivera, I. A. (2022). The binomial distribution: Historical origin and evolution of its problem situations. Mathematics, 10(15), 2680.
<https://doi.org/10.3390/math10152680>
2. Agresti, A. (2021). The foundations of statistical science: A history of textbook presentations. Brazilian Journal of Probability and Statistics, 35(4), 657-698.
<https://doi.org/10.1214/21-BJPS518>
3. Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material. Philosophical Transactions of the Royal Society, Series A: Mathematical, Physical and Engineering Sciences,186, 343‑414.

https://doi.org/10.1098/rsta.1895.0010

1. Pearson, K. (1916). Mathematical contributions to the theory of evolution. XIX. Second supplement to a memoir on skew variation. Philosophical Transition of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences, 216, 429‑457.

<https://doi.org/10.1098/rsta.1916.0009>

1. Elderton, W. P. (1906). Frequency-curves and correlation. London, UK: Charles and Edwin Layton.
2. Elderton, W. P. (1927). Frequency-curves and correlation (2th ed.). Cambridge, UK: Cambridge University Press.
3. Elderton, W. P. (1938). Frequency-curves and correlation (3th ed.). Cambridge, UK: Cambridge University Press.
4. Mathieson, I., & Terhorst, J. (2022). Direct detection of natural selection in Bronze Age Britain. Genome Research, 32(11-12), 2057-2067.
<https://doi.org/10.1101/gr.276862.122>
PMid:36316157
PMCid:PMC9808619
5. Bowman, K. O., & Shenton, L. R. (2007). The beta distribution, moment method, Karl Pearson and RA Fisher. Far East Journal of Theoretical Statistics, 23(2), 133‑164.
6. Fujita, K., Okada, K., & Katahira, K. (2022). The Fisher information matrix: A tutorial for calculation for decision making models.

<https://doi.org/10.31234/osf.io/hdwut>

1. Pascucci, A. (2024). Probability theory: Random variables and distributions. I. Berlin: Springer Nature.

<https://doi.org/10.31234/osf.io/hdwut>

1. Ghahramani, S. (2024). Fundamentals of probability. Boca Raton, FL: CRC Press.
2. Singh, D. (2020). Beta and gamma function implementation in R.
<https://www.pluralsight.com/resources/blog/guides/beta-and-gamma-function-implementation-in-r>
3. R Core Team. (2024a). Special {base}. Special functions of mathematics. In R Documentation.

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/Special.html>

1. R Core Team. (2024b). stats (version 3.6.2). Beta: The beta distribution. In R Documentation.
https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Beta
2. R Core Team. (2024c).hypergeo (version 1.2-13). hypergeo: The hypergeometric function. In R Documentation.

https://www.rdocumentation.org/packages/hypergeo/versions/1.2-13/topics/hypergeo

1. Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379‑423.

https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

1. Saupe, D., Rusek, K., Hägele, D., Weiskopf, D., & Janowski, L. (2024). Maximum entropy and quantized metric models for absolute category ratings. *IEEE Signal Processing Letters, 31*, 2970-2974.

https://doi.org/10.1109/LSP.2024.3480832

1. Wang, D., Zhong, J., Li, C., & Peng, Z. (2021). Box-Cox sparse measures: A new family of sparse measures constructed from kurtosis and negative entropy. Mechanical Systems and Signal Processing, 160, 107930.

https://doi.org/10.1016/j.ymssp.2021.107930

1. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics, 22(1), 79‑86.

https://doi.org/10.1214/aoms/1177729694

1. Kullback, S. (1959). Information theory and statistics. Hoboken, NJ: John Wiley and Sons.
2. Pearson, K. (1905). Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A rejoinder.Biometrika, 4, 169-212.
https://doi.org/10.1093/biomet/4.1-2.169
3. Ali, H., Akanihu, C. N., & Felix, J. (2023). Investigating the parameters of the beta distribution World Journal of Advanced Research and Reviews, 19(01), 815‑830.
https://doi.org/10.30574/wjarr.2023.19.1.1409
4. Prucha, I. R. (2021). Instrumental variables/method of moments estimation. In M. M. Fischer & P. Nijkamp (Eds.), Handbook of regional science (pp. 2097‑2116). Berlin: SpringerNature.

https://doi.org/10.1007/978-3-662-60723-7\_90

1. Rahman, M., & Amin, M. I. (2024). A note on parameter estimation in beta distribution. *Far East Journal of Theoretical Statistics, 68*(2), 255‑262.
https://doi.org/10.17654/0972086324015
2. Millard, S. P. (2024) ebeta {EnvStats}. Estimate quantiles of a beta distribution. InR Documentation.

https://search.r-project.org/CRAN/refmans/EnvStats/html/ebeta.html

1. Zaiontz, C. (2024). Fitting beta distribution parameters via MLE.
https://real-statistics.com/distribution-fitting/distribution-fitting-via-maximum-likelihood/fitting-beta-distribution-parameters-mle/
2. Fisher, R. A. (1971). Statistical methods and scientific inference (3rd ed.). New York: Hafner.
3. Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 1190‑1208.

https://doi.org/10.1137/0916069

1. Chattamvelli, R., & Shanmugam, R. (2022). Continuous distributions in engineering and the applied sciences - Part I. Berlin: Springer Nature.
https://doi.org/10.1007/978-3-031-02435-1
2. Aharon, O., & Tabrikian, J. (2024). Asymptotically tight Bayesian Cramér-Rao bound. IEEE Transactions on Signal Processing,72, 3333-3346.
https://doi.org/10.1109/TSP.2024.3421900
3. Tarima, S., & Flournoy, N. (2019). Asymptotic properties of maximum likelihood estimators with sample size recalculation. Statistical Papers 60, 373–394.
https://doi.org/10.1007/s00362-019-01095-x
PMid:31827313
PMCid:PMC6905624
4. Efford, M. G. (2020). Confint.secr {secr}. Profile likelihood confidence intervals. R Documentation.

https://www.otago.ac.nz/density/html/confint.secr.html

1. McDonald, J. H. (2014). G-test of goodness-of-fit. In Handbook of biological statistics (3rd ed.). Baltimore, Maryland: Sparky House Publishing. https://www.biostathandbook.com/gtestgof.html.
2. Moore, D. (1986). Tests of chi-squared type. In R. B. D’Agostino & M. A. Stephens (Eds.). Goodness-of-fit techniques (pp. 63–95). New York, NY: Marcel Dekker.

https://doi.org/10.1201/9780203753064-3
PMCid:PMC5897041

1. Woolf, B. (1957). The log likelihood ratio test (the G-test); methods and tables for tests of heterogeneity in contingency tables. Annals of Human Genetics, 21(4), 397−409.

https://doi.org/10.1111/j.1469-1809.1972.tb00293.x
PMid:13435648

1. Williams, D. A. (1976). Improved likelihood ratio test for complete contingency tables. Biometrika, 63(1), 33−37.

https://doi.org/10.1093/biomet/63.1.33

1. Berrett, T. B., & Samworth, R. J. (2021). USP: an independence test that improves on Pearson’s chi-squared and the G-test. Proceedings of the Royal Society,Series A: Mathematical, Physical and Engineering Sciences, 477(2256), 20210549.

https://doi.org/10.1098/rspa.2021.0549
PMid:35153605
PMCid:PMC8652272

1. Ebner, B., & Liebenberg, S. C. (2021). On a new test of fit to the beta distribution. Stat, 10(1), e341.

https://doi.org/10.1002/sta4.341

1. Nikseresht, A., & Amindavar, H. (2024). Energy demand forecasting using adaptive ARFIMA based on a novel dynamic structural break detection framework. Applied Energy, 353(1), 122069.

https://doi.org/10.1016/j.apenergy.2023.122069

1. Moral de la Rubia, J. M. (2024). Determination of the number and width of class intervals using R. Annals of Environmental Science and Toxicology, 8(1), 22-42.

https://dx.doi.org/10.17352/aest.000077

1. Rolke, W., & Gongora, C.G. (2021). A chi-square goodness-of-fit test for continuous distributions against a known alternative. Computational Statistics, 36, 1885‑1900.

https://doi.org/10.1007/s00180-020-00997-x

1. Cramér, H. (1946). Mathematical methods of statistics. Princeton, NJ: Princeton University Press.

https://doi.org/10.1515/9781400883868

1. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
2. Fey, C. F., Hu, T., Delios, A. (2023) The measurement and communication of effect sizes in management research. Management and Organization Review, 19(1), 176-197.

<https://doi.org/10.1017/mor.2022.2>

1. Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., …, & van den Brand, T. (2024) "ggplot2": Create elegant data visualisations using the grammar of graphics.

https://doi.org/10.32614/CRAN.package.ggplot2

1. Millard, S. P., & Kowarik, A. (2024). Package for environmental statistics, including US EPA guidance.

https://github.com/alexkowa/EnvStats

1. Tukey, J. W. (1977). Exploratory data analysis. Readings, MA: Addison-Wesley.
2. Hyndman, R. J., & Fan, Y. (1996) Sample quantiles in statistical packages. American Statistician, 50(4), 361−365.

https://doi.org/10.2307/2684934

1. Pinelis, I. (2020). An asymptotically optimal transform of Pearson’s correlation statistic. Mathematical Methods of Statistics, 28(1), 307-318.
https://doi.org/10.3103/S1066530719040057
2. Gupta, A. K., & Nadarajah, S. (Eds.). (2004). Handbook of beta distribution and its applications. Boca Ratón: Fl: CRC Press.

<https://doi.org/10.1201/9781482276596>

1. Jeffreys, H. (1946). An invariant form of the prior probability in estimation problems. Proceedings Royal Society of London, Series A(Mathematical, Physical and Engineering Sciences), 186, 453‑461.

http://doi.org/10.1098/rspa.1946.0056
PMid:20998741

1. Haldane, J. B. S. (1932). A note on inverse probability. Mathematical Proceedings of the Cambridge Philosophical Society, 28(1), 55‑61.
<https://doi.org/10.1017/s0305004100010495>
2. Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of mathematical statistics. New Delhi, India: Sultan Chand & Sons.
3. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Beta distributions. En Continuous univariate distributions, 2nd ed. (Vol. 2, pp. 210−275). Hoboken, NJ: John Wiley & Sons.
4. Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods (8th ed.). Ames: Iowa State University Press.
5. Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. Annals of Mathematics, 67(2), 325‑327.

<https://doi.org/10.2307/1970008>

1. Stuart, A., & Ord, K. (2010). Kendall's advanced theory of statistics, distribution theory (Vol. 1). Hoboken, NJ: John Wiley & Sons.
2. Orsi, C. (2017). New insights into non-central beta distributions. arXiv preprintarXiv:1706.08557.

https://doi.org/10.48550/arXiv.1706.08557

1. Navarro, D., & Perfors, A. (2005). An introduction to the beta-binomial model. Adelaide: University of Adelaide.
2. Wanas, A. K., & Al-Ziadi, N. A. J. (2021). Applications of beta negative binomial distribution series on holomorphic functions. *Earthline Journal of Mathematical Sciences*, *6*(2), 271-292.

https://doi.org/10.34198/ejms.6221.271292

1. Ng, K. W., Tian, G. L., & Tang, M. L. (2011). Dirichlet and related distributions: Theory, methods and applications. Hoboken: John Wiley & Sons.
https://doi.org/10.1002/9781119995784
2. Clark, C. E. (1962) The PERT model for the distribution of an activity. Operations Research, 10(3), 405−406.

https://doi.org/10.1287/opre.10.3.405

1. Khan, I. A., Bickel, J. E., & Hammond, R. K. (2023). Determining the accuracy of the triangular and PERT distributions. Decision Analysis, 20(2), 151-165.
https://doi.org/10.1287/deca.2022.0464
2. Wang, Z. J., Pei, H. R., & Wang, Y. (2013, December). Sampling traffic analysis of anonymous communications in mobile ad hoc networks. In 2013 IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks (pp. 233-239). New York: IEEE Publisher.

https://doi.org/10.1109/10.1109/MSN.2013.38

1. Wee, B. F., Hwong, H. H., Sivakumar, S., Lim, K. H., Wong, W. K., Chew, M., & Tiong, M. C. (2023). Optimizing diabetes classification with support vector machine and SMOTEENN-based feature selection. In 2023 International Conference on Digital Applications, Transformation & Economy (ICDATE) (pp. 1-5). New York: IEEE Publisher.

<https://doi.org/10.1109/ICDATE58146.2023.10248600>

1. Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., … . Hoijtink, H. (2023). A review of applications of the Bayes factor in psychological research. Psychological Methods, 28(3), 558‑579.
https://doi.org/10.1037/met0000454
PMid:35298215
2. Seo, J. I., & Kim, Y. (2022). Nonparametric prior elicitation for a binomial proportion. Communications in Statistics-Simulation and Computation, 51(6), 2809-2821.
https://doi.org/10.1080/03610918.2019.1702210
3. Laplace, P. S. (1820). Théorie analytique des probabilités (3th ed.). Paris: Courcier.
4. Prendergast, L. A., Dedduwakumara, S., & Staudte, R. G. (2024). rquest: An R package for hypothesis tests and confidence intervals for quantiles and summary measures based on quantiles. arXiv Preprint arXiv:2410.11093.
https://doi.org/10.48550/arXiv.2410.11093
5. Pham, H. (2023). Basic statistics. InSpringer handbook of engineering statistics, 2nd ed. (pp. 3‑41). Berlin: Springer Nature.

https://doi.org/10.1007/978-1-4471-7503-2\_1

1. Epanechnikov, V.A. (1969). Nonparametric estimation of a multidimensional probability density. Theory of Probability and its Applications, 14(1), 156‑161.
https://doi.org/10.1137/1114019
2. Sheather, S.J.,& Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B: Statistical Methodology, 53(3), 683‑690.
https://doi.org/10.1111/j.2517-6161.1991.tb01857.x
3. Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology, 1(1), 161‑187.
https://doi.org/10.1080/24709360.2017.1396742

**# APPENDIX 1. R Script for Example 1. Beta Distribution with Known Parameters.**

# Load package.

library(modeest)

# Enter the values for the Beta distribution parameters.

alpha <- 2

beta <- 7

#1) Calculation of the specific probabilities of the example.

p1 <- pbeta(0.1, alpha, beta) # P(X <= 0.1)

p2 <- pbeta(0.22, alpha, beta) - pbeta(0.15, alpha, beta) # P(0.15 <= X <= 0.22)

p3 <- 1 - pbeta(0.25, alpha, beta) # P(X >= 0.25)

#2) Distribution statistics: measures of central tendency.

mean\_arithmetic <- alpha / (alpha + beta)

mean\_geometric <- exp((digamma(alpha) - digamma(alpha + beta)))

mean\_harmonic <- (alpha - 1) / (alpha + beta - 1) #If alpha >=1 and beta > 0

median <- qbeta(0.5, alpha, beta)

mode<- betaMode(shape1 = alpha, shape2 = beta)

#3) Distribution statistics: measures of variation.

variance <- (alpha \* beta) / ((alpha + beta)^2 \* (alpha + beta + 1))

std\_dev <- sqrt(variance)

mad <- integrate(function(x) abs(x - mean\_arithmetic) \* dbeta(x, alpha, beta), 0, 1)$value

entropy <- lgamma(alpha + beta) - lgamma(alpha) - lgamma(beta) +

(alpha - 1) \* (digamma(alpha) - digamma(alpha + beta)) +

(beta - 1) \* (digamma(beta) - digamma(alpha + beta))

# 4) Skewness and excess kurtosis based on standardized central moments.

skewness <- (2 \* (beta - alpha) \* sqrt(alpha + beta + 1)) / ((alpha + beta + 2) \* sqrt(alpha \* beta))

excess\_kurtosis <- 6 \*(((beta - alpha)^2 \* (alpha + beta + 1)) - (alpha \* beta \* (alpha + beta + 2))) / (alpha \* beta \* (alpha + beta + 2) \* (alpha + beta + 3))

#5) The probability density function and the cumulative distribution function are displayed together on a graph.

x <- seq(0, 1, length.out = 500)

density <- dbeta(x, alpha, beta)

cdf <- pbeta(x, alpha, beta)

# jpeg("Figure 3\_Functions.jpg", width = 2500, height = 1500, res = 500) # To save the plot as a JPEG file, remove the hashtag symbol in front of the command. The plot will then be saved to the Files folder instead of being displayed in the console.

par(mar = c(4, 4, 1, 1) + 0.1)

plot(x, density, type = "l", col = "blue", lwd = 2, ylab = "Density / Cumulative Probability", xlab = "x", main = "")

lines(x, cdf, col = "red", lwd = 2)

legend("topright", legend = c("PDF", "CDF"), col = c("blue", "red"), lwd = 2)

dev.off()

# Print results.

cat("Probabilities:\n")

cat("P(X < 0.1):", p1, "\n")

cat("P(0.15 < X < 0.22):", p2, "\n")

cat("P(X > 0.25):", p3, "\n\n")

cat("Measures of central tendency:\n")

cat("Arithmetic mean: M(X) =", round(mean\_arithmetic, 4), "\n")

cat("Geometric mean: G(X) =", round(mean\_geometric, 4), "\n")

cat("Harmonic mean: H(X) =", round(mean\_harmonic, 4), "\n")

cat("Median: Mdn(X) =", round(median, 4), "\n")

cat("Mode: Mo(X) =", round(mode, 4), "\n\n")

cat("Measures of variation:\n")

cat("Variance: σ^2 =", round(variance, 4), "\n")

cat("Standard deviation: σ", round(std\_dev, 4), "\n")

cat("Mean absolute deviation: MAD(X) =", round(mad, 4), "\n")

cat("Entropy: Η(x) =", round(entropy, 4), "\n\n")

cat("Shape measures based on standardized central moments:\n")

cat("Person’s coefficient of skewness: β\_1 =", round(skewness, 4), "\n")

cat("Pearson’s excess kurtosis: β\_2 - 3 =", round(excess\_kurtosis, 4), "\n")

**# APPENDIX 2. R Script for Example 2. Estimation Using Method of Moments.**

# Vector of X-values.

x <- c(0.12354423, 0.44261002, 0.08595943, 0.42195273, 0.24214223, 0.42195046, 0.18675247, 0.21126755, 0.17104856, 0.16972457, 0.43533068, 0.23879670, 0.11629033, 0.15349581, 0.14122295, 0.19140420, 0.29961921, 0.33053753, 0.19892422, 0.38360256)

# 1a) Estimation of the alpha and beta parameters by applying the method of moments.

mean\_x <- mean(x)

var\_x <- var(x)

alpha <- mean\_x \* ((mean\_x \* (1 - mean\_x)) / var\_x - 1)

beta <- (1 - mean\_x) \* ((mean\_x \* (1 - mean\_x)) / var\_x - 1)

cat("Parameter estimation using the method of moments:\n")

cat("Alpha estimate: α =", round(alpha, 4), "\n")

cat("Beta estimate: β =", round(beta, 4), "\n")

# 2) Descriptive statistics.

mean\_arithmetic <- alpha / (alpha + beta)

median <- if (alpha > 1 && beta > 1) (alpha - 1/3) / (alpha + beta - 2/3) else NA

mode <- if (alpha < 1 && beta < 1) {c(0, 1)

} else if (alpha <= 1 && beta > 1) {0

} else if (alpha > 1 && beta <= 1) {1

} else if (alpha > 1 && beta > 1) {(alpha - 1) / (alpha + beta - 2)

} else if (alpha = 1 && beta = 1) {"There is no mode"}

variance <- (alpha \* beta) / ((alpha + beta)^2 \* (alpha + beta + 1))

std\_dev <- sqrt(variance)

skewness <- (2 \* (beta - alpha) \* sqrt(alpha + beta + 1)) / ((alpha + beta + 2) \* sqrt(alpha \* beta))

excess\_kurtosis <- 6 \*(((beta - alpha)^2 \* (alpha + beta + 1)) - (alpha \* beta \* (alpha + beta + 2))) / (alpha \* beta \* (alpha + beta + 2) \* (alpha + beta + 3))

cat("Measures of central tendency:\n")

cat("Arithmetic mean: M(X) =", round(mean\_arithmetic, 4), "\n")

cat("Median: Mdn(X) =", round(median, 4), "\n")

cat("Mode: Mo(X) =", round(mode, 4), "\n\n")

cat("Measures of variation:\n")

cat("Variance: σ^2 =", round(variance, 4), "\n")

cat("Standard deviation: σ", round(std\_dev, 4), "\n")

cat("Measures of shape:\n")

cat("Person’s coefficient of skewness: β\_1 =", round(skewness, 4), "\n")

cat("Pearson’s excess kurtosis: β\_2 - 3 =", round(excess\_kurtosis, 4), "\n")

# 3) The probability density function and the cumulative distribution function are displayed together on a graph.

x <- seq(0, 1, length.out = 500)

density <- dbeta(x, alpha, beta)

cdf <- pbeta(x, alpha, beta)

# jpeg("Figure4\_Functions.jpg", width = 2500, height = 1500, res = 500) # To save the plot as a JPEG file, remove the hashtag symbol in front of the command. The plot will then be saved to the Files folder instead of being displayed in the console.

par(mar = c(4, 4, 1, 1) + 0.1)

plot(x, density, type = "l", col = "blue", lwd = 2, ylab = "Density / Cumulative Probability", xlab = "x", main = "")

lines(x, cdf, col = "red", lwd = 2)

legend("topright", legend = c("PDF", "CDF"), col = c("blue", "red"), lwd = 2)

dev.off()

# 4) Calculation of the specific probabilities of the example.

p1 <- pbeta(0.2, alpha, beta) # P(X < 0.2)

p2 <- pbeta(0.1, alpha, beta) - pbeta(0.05, alpha, beta) # P(0.1 <= X <= 0.05)

p3 <- 1 - pbeta(0.25, alpha, beta) # P(X > 0.25)

cat("Probabilities:\n")

cat("P(X < 1/5):", p1, "\n")

cat("P(1/20 <= X <= 1/10):", p2, "\n")

cat("P(X > 1/4):", p3, "\n\n")

**# APPENDIX 3. R Script for Example 3. Estimation Using Maximum Likelihood.**

# 1b) Estimation of the alpha and beta parameters using maximum likelihood. To apply this method, replace 1a with 1b in the script of the APPENDIX 2.

# Vector of X-values.

x <- c(0.12354423, 0.44261002, 0.08595943, 0.42195273, 0.24214223, 0.42195046, 0.18675247, 0.21126755, 0.17104856, 0.16972457, 0.43533068, 0.23879670, 0.11629033, 0.15349581, 0.14122295, 0.19140420, 0.29961921, 0.33053753, 0.19892422, 0.38360256)

# Load package.

library(stats4)

# Negative log-likelihood function.

neg\_log\_likelihood <- function(alpha, beta) {

if (alpha <= 0 || beta <= 0) return(Inf) # Ensure parameters are positive

-sum(dbeta(x, alpha, beta, log = TRUE)) }

# Initial parameters.

alpha\_start <- mean(x) \* ((mean(x) \* (1 - mean(x))) / var(x) - 1)

beta\_start <- (1 - mean(x)) \* ((mean(x) \* (1 - mean(x))) / var(x) - 1)

# Maximum likelihood estimation.

mle\_fit <- mle(neg\_log\_likelihood, start = list(alpha = alpha\_start, beta = beta\_start),

method = "L-BFGS-B", lower = c(0.001, 0.001))

# Extract and display results.

mle\_summary <- summary(mle\_fit)

alpha <- mle\_summary@coef["alpha", "Estimate"] # alpha estimate

beta <- mle\_summary@coef["beta", "Estimate"] # beta estimate

cat("Maximum Likelihood estimation results:\n")

cat("Alpha estimate: α =", alpha, "\n")

cat("Beta estimate: β =", beta, "\n")

cat("Standard errors:\n")

cat("se(alpha) =", mle\_summary@coef["alpha", "Std. Error"], "\n")

cat("se(beta) =", mle\_summary@coef["beta", "Std. Error"], "\n")

# Confidence intervals for the parameters.

ci <- confint(mle\_fit)

cat("95% confidence interval for alpha:", "[", round(ci["alpha", ], 4), "]", "\n")

cat("95% confidence interval for beta:", "[", round(ci["beta",], 4), "]", "\n")

# Additional information about the estimate.

# Maximum value of the natural logarithm of the likelihood function.

log\_likelihood\_max <- logLik(mle\_fit)

cat("Maximum value of the natural logarithm of the likelihood function: ln\_L(α, β| x) =", log\_likelihood\_max, ".", "\n")

# Information and covariance matrices of parameters.

hessian\_matrix <- attr(mle\_fit, "details")$hessian

cat("Information matrix", "\n")

print(hessian\_matrix)

det\_IM <- det(hessian\_matrix)

cat("Determinant of the information matrix: |I(α, β)| =", det\_IM, ".", "\n")

cov\_matrix <- vcov(mle\_fit)

cat("Parameter covariance matrix", "\n")

print(cov\_matrix)

# Calculating standard errors from the information matrix.

sd\_alpha <- sqrt(hessian\_matrix[2, 2]/det\_IM)

sd\_beta <- sqrt(hessian\_matrix[1, 1]/det\_IM)

cov\_alpha\_beta <- -hessian\_matrix[1, 2]/det\_IM

cat("Standard deviation of alpha:", sd\_alpha, "\n")

cat("Standard deviation of beta:", sd\_beta, "\n")

cat("covariance between alpha and beta:", cov\_alpha\_beta, "\n")

# 95% asymptotic confidence intervals for alpha and beta.

LL\_alpha <- alpha - qnorm(0.975) \* sd\_alpha

UL\_alpha <- alpha + qnorm(0.975) \* sd\_alpha

cat("The 95% asymptotic confidence interval for alpha:", "[", round(LL\_alpha, 4), ",", round(UL\_alpha, 4), "]", ".", "\n")

LL\_beta <- beta - qnorm(0.975) \* sd\_beta

UL\_beta <- beta + qnorm(0.975) \* sd\_beta

cat("The 95% asymptotic confidence interval for beta:", "[", round(LL\_beta, 4), ",", round(UL\_beta, 4), "]", ".", "\n")

# **APPENDIX 4. R Script for Example 4. Woolf’s G-Test and QQ Plot.**

# Generating the random sample in Example 4.

set.seed(123)

x <- round(rbeta(45, shape1 = 6, shape2 = 2), 4)

print(x)

# Estimation of the alpha and beta parameters by applying the method of moments.

alpha <- mean(x) \* ((mean(x) \* (1 - mean(x))) / var(x) - 1)

beta <- (1 - mean(x)) \* ((mean(x) \* (1 - mean(x))) / var(x) - 1)

cat("Estimation of the parameters alpha and beta using the method of moments:\n")

cat("Alpha estimate: α =", round(alpha, 4), "\n")

cat("Beta estimate: β =", round(beta, 4), "\n")

# Moore's rule to determine the number of intervals.

n <- length(x)

k <- round(2 \* n^0.4, 0)

n\_CI <- floor(n / k)

n\_exc <- n - k \* n\_CI

cat("Moore's rule:\n")

cat("Number of intervals: k =", k, "\n")

cat("Common frequency per interval: n\_CI =", n\_CI, "\n")

cat("Excess frequency (one per central interval): n\_exc =", n\_exc, "\n")

# Class intervals.

x\_sorted <- sort(x)

interval\_limits <- numeric(k + 1)

for (i in 1:(k + 1)) {if (i == 1) {interval\_limits[i] <- min(x\_sorted)

} else if (i == (k + 1)) {interval\_limits[i] <- max(x\_sorted)

} else {interval\_limits[i] <- quantile(x\_sorted, probs = (i - 1) / k)}}

# Frequency table.

frequency\_table <- data.frame(Class\_interval = character(k),

Observed\_frequency = numeric(k),

Expected\_frequency = numeric(k),

O\_LN\_O\_E = numeric(k),

stringsAsFactors = FALSE)

for (i in 1:k) {lower\_limit <- interval\_limits[i]

upper\_limit <- interval\_limits[i + 1]

interval\_label <- paste0("(", round(lower\_limit, 4), ", ", round(upper\_limit, 4), "]")

n\_o <- sum(x >= lower\_limit & x <= upper\_limit)

F\_z\_LS\_prev <- pbeta(lower\_limit, alpha, beta)

F\_z\_LS <- pbeta(upper\_limit, alpha, beta)

n\_e <- n \* (F\_z\_LS - F\_z\_LS\_prev)

g\_i <- ifelse(n\_e == 0, 0, n\_o \* log(n\_o / n\_e))

frequency\_table[i, ] <- c(interval\_label, n\_o, n\_e, g\_i)}

frequency\_table$Observed\_frequency <- as.numeric(frequency\_table$Observed\_frequency)

frequency\_table$Expected\_frequency <- as.numeric(frequency\_table$Expected\_frequency)

frequency\_table$O\_LN\_O\_E <- as.numeric(frequency\_table$O\_LN\_O\_E)

total\_g <- sum(frequency\_table$O\_LN\_O\_E, na.rm = TRUE)

cat("Table. Observed and expected frequencies by class interval for the G-test calculation\n")

print(frequency\_table)

cat("Note. i = class interval index, Oi = observed frequency, Ei = expected frequency, and O\_LN\_O\_E = Oi × ln(Oi / Ei)\n")

# G-test.

alpha\_significance <- 0.05

g <- 2 \* total\_g

q <- 1 + (k^2 - 1) / (6 \* n \* (k - 3))

g\_cc <- g / q

p\_value <- pchisq(g\_cc, df = k - 3, lower.tail = FALSE)

cat("\nG-test result:\n")

cat("G statistic without continuity correction: g =", round(g, 4), "\n")

cat("Williams’ continuity correction: q =", round(q, 4), "\n")

cat("Williams’ continuity-corrected G-test statistic: g\_cc =", round(g\_cc, 4), "\n")

cat("P-value =", round(p\_value, 6), "\n")

if (p\_value < alpha\_significance) {

cat("The null hypothesis that the data fit a Beta distribution is rejected at a significance level of 0.05 using Woolf’s G-test.\n")

} else {cat("The null hypothesis that the data fit a Beta distribution is not rejected at a significance level of 0.05 using Woolf’s G-test.\n")}

# Statistical power.

power\_g\_cc <- 1 - pchisq(qchisq(alpha\_significance, df = k - 3, lower.tail = FALSE), df = k - 3, ncp = g\_cc, lower.tail = TRUE, log.p = FALSE)

cat("The right-tailed statistical power for the alternative hypothesis of non-normality for the G-test with continuity correction: ϕ =", round(power\_g\_cc, 4), "\n")

# Effect size.

v <- sqrt(g\_cc/ (n \* (k - 3)))

cat("Effect size through Cramer’s V coefficient: V =", round(v, 4), "\n")

interpret\_cramers\_v <- function(v, k) {trivial\_cutoff <- 0.1 / sqrt(k - 3)

small\_cutoff <- 0.29 / sqrt(k - 3)

medium\_cutoff <- 0.49 / sqrt(k - 3)

large\_cutoff <- 0.5 / sqrt(k - 3)

effect\_size <- ifelse(v < trivial\_cutoff, "trivial", ifelse(v < small\_cutoff, "small",

ifelse(v < medium\_cutoff, "medium", "large")))

cat("Interpretation of effect size based on Cohen, 1988:\n")

cat("- Trivial effect size: V <", round(trivial\_cutoff, 4), "\n")

cat("- Small effect size: V from", round(trivial\_cutoff, 4), "to <", round(small\_cutoff, 4), "\n")

cat("- Medium effect size: V from", round(small\_cutoff, 4), "to <", round(medium\_cutoff, 4), "\n")

cat("- Large effect size: V >=", round(medium\_cutoff, 4), "\n")

cat("\nBased on the calculated V =", round(v, 4), ", the effect size is classified as", effect\_size, ".\n")}

interpret\_cramers\_v(v, k)

# QQ plot.

# jpeg("Figure 5\_QQ\_Plot.jpg", width = 1200, height = 800, res = 150) # To save the plot as a JPEG file, remove the hashtag symbol in front of the command. The plot will then be saved to the Files folder instead of being displayed in the console.

par(mar = c(5, 5, 1, 1) + 0.1)

p <- (1:n - 1/3) / (n + 1/3)

theoretical\_quantiles <- qbeta(p, alpha, beta)

empirical\_quantiles <- sort(x)

plot(theoretical\_quantiles, empirical\_quantiles, main = "", xlab = "Theoretical Quantiles (Beta Distribution)", ylab = "Empirical Quantiles", pch = 16, col = "blue", cex = 1.2)

abline(0, 1, col = "red", lwd = 2)

# dev.off()

# Calculate Pearson's correlation coefficient between empirical and theoretical quantiles and 95% confidence interval using Fisher's hyperbolic arctangent transformation.

library(psych)

correlation <- cor.test(theoretical\_quantiles, empirical\_quantiles, method = "pearson")

Fisher\_ci <- r.con(r = correlation$estimate, n = n, p = 0.95)

cat("Pearson’s correlation coefficient between empirical and theoretical quantiles: r =", round(correlation$estimate, 4), ".", "\n")

cat("The 95% confidence interval using Fisher's transformation:", "[", round(Fisher\_ci[1], 4), ",", round(Fisher\_ci[2], 4), "]", ".", "\n")

# Plot with the estimated and theoretical density curves.

estimated\_d <- density(x, kernel = "epa", bw = "SJ")

x\_t <- seq(0, 1, length.out = 1000)

theoretical\_d <- dbeta(x\_t, alpha, beta)

# jpeg("Figure 6\_Density\_curves.jpg", width = 2500, height = 1500, res = 500)

par(mar = c(4, 4, 1, 1) + 0.1) # To save the plot, remove the hashtag symbol.

plot(estimated\_d, xlim = c(0, 1), type = "l", col = "blue", lwd = 2, ylab = "Density", xlab = "x", main = "")

lines(x\_t, theoretical\_d, col = "red", lwd = 2)

legend("topleft", legend = c("Empirical", "Theoretical"), col = c("blue", "red"), lwd = 2)

dev.off()