

THE GAMMA DISTRIBUTION
PROPERTIES, PROOFS AND APPLICATIONS

ETIENNE COLLIN, MARIE OUELLET & RANIA YAHYAOU
PROBABILITY AND STATISTICS - 201-BNM-LW

Section 20108
PROFESSOR VINCENT CARRIER

ST. LAWRENCE CÉGEP CHAMPLAIN

Due on May 13, 2022 at 23:59



St.Lawrence
CEGEP CHAMPLAIN

Table of Contents

Table of Contents	1
1 Introduction	3
1.1 Historical Context & Parameterization	3
1.2 Probability Density Function	4
1.3 Cumulative Distribution Function	6
2 Properties	8
2.1 Moment Generating Function	8
2.2 Expected Value	9
2.2.1 Finding the Equation	9
2.2.1.1 Proof 1	9
2.2.1.2 Proof 2	10
2.2.1.3 Proof 3	10
2.2.2 Special Case	11
2.2.2.1 Expected Value	11
2.2.2.2 Variance	12
2.3 Variance	13
2.3.0.1 Proof 1	13
2.3.0.2 Proof 2	15
2.4 Quantiles	16
2.5 Mode	17
2.5.1 Special Cases	18
2.5.1.1 Case 1:	18
2.5.1.2 Case 2:	18
2.6 Inflection Points	19
2.7 Median	20
2.8 Memoryless Property	23
3 The Gamma Function	25
3.1 Definition	25
3.1.1 Special Case of $\Gamma(1/2)$	26
3.1.1.1 Initial Setup	26
3.1.1.2 The Proof	28
3.1.2 Negative Values of α	29
3.2 Relationship to Factorial Notation	30
3.3 Plotting the Gamma Function	31
4 Special Cases of the Gamma Distribution & Related Proofs	32
4.1 The Exponential Distribution	32
4.1.1 Discovery of the Gamma Probability Density Function	33
4.1.2 Sum of Exponential Random Variables	36

4.1.2.1	Using Convolution	36
4.1.2.2	Using a Change of Variables	38
4.2	The Beta Distribution	40
4.2.1	Quotient of Gamma Random Variables	42
4.2.1.1	Using a Jacobian	42
4.2.1.2	Using a Change of Variables	44
4.3	The Chi-Squared Distribution	46
4.4	The Normal Distribution	49
4.5	The Erlang Distribution	51
4.6	The Wishart Distribution	52
5	Advanced Properties	53
5.1	Skewness	53
5.2	Information Entropy	53
5.3	Characteristic Function	54
5.4	Methods of Moments	54
5.4.1	Higher Moments	55
5.5	Excess kurtosis	55
6	General Applications and Family	56
6.1	Insurance Companies	56
6.2	Natural Events Prediction	56
6.3	Customer Satisfaction	56
6.3.1	Service Time	56
6.3.2	Call Centers	57
6.3.3	Retail	57
6.4	Oncology	57
A	Graphs - Python Code	60
A.1	Probability Density Function	60
A.2	Cumulative Distribution Function	61
A.3	Central Limit Theorem	62
B	Simulations - Python Code	63
B.1	Simulation 1: $X + Y$	63
B.2	Simulation 2: $\frac{X}{X+Y}$	64
	List of Figures	65
	References	67

1 Introduction

1.1 Historical Context & Parameterization

The gamma distribution^[1] is a distribution of continuous random variables X with two nonnegative parameters α and λ . A gamma distributed random variable X is represented as:

$$X \sim \mathcal{Gamma}(\alpha, \lambda)$$

or, alternatively, as:

$$X \sim \Gamma(\alpha, \lambda)$$

The gamma probability distribution gave rise to many other functions and distributions, such as the gamma function, the exponential distribution, the Erlang distribution or the chi-square distribution. Originally, it was first presented by Leon Hard Euler, a Swiss mathematician and physician. As Euler gained popularity, the gamma distribution was further researched by important figures in the domain of mathematics such as Karl Weierstrass, Carl Friedrich Gauss, Charles Hermite, and many more.^{[2], [3], [4]}

The gamma distribution is part of the two-parameter family of continuous probability distributions. Indeed, it may be parameterized with two different parameterizations:^[3]

Parameterization 1:

Shape: $\alpha > 0$

Rate: $\lambda > 0$

Parameterization 2:

Shape: $k > 0$

Scale: $\theta > 0$

In the second parameterization, the scale parameter θ corresponds to the inverse of the rate parameter λ , allowing flexibility in the modelization of the distribution. Indeed, the application of the scale parameter is pertinent when one wants to use the value of the mean time between events in a Poisson process, as opposed to the mean rate of occurrence during one unit of time.

That being said, in this document, as the two parameterizations only exist for the sake of convenience and are identical in their results, only parameterization 1 will be considered and used for proofs.

1.2 Probability Density Function

The probability density function of a gamma distributed random variable X corresponds to

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ for } x \in (0, \infty) \quad (1)$$

In statistics, it is often used to model the time between independent events that occur at a constant average rate in a Poisson process. Indeed, the parameter α indicates the number of events that are modelled, and the parameter λ specifies the average rate at which they occur. Here is a graph of the probability density function of a gamma distribution with parameters $\alpha = 4$ and $\lambda = 3$. As an example, this probability density function could be used to evaluate the time until four events that occur at a rate of three events per unit time take place:

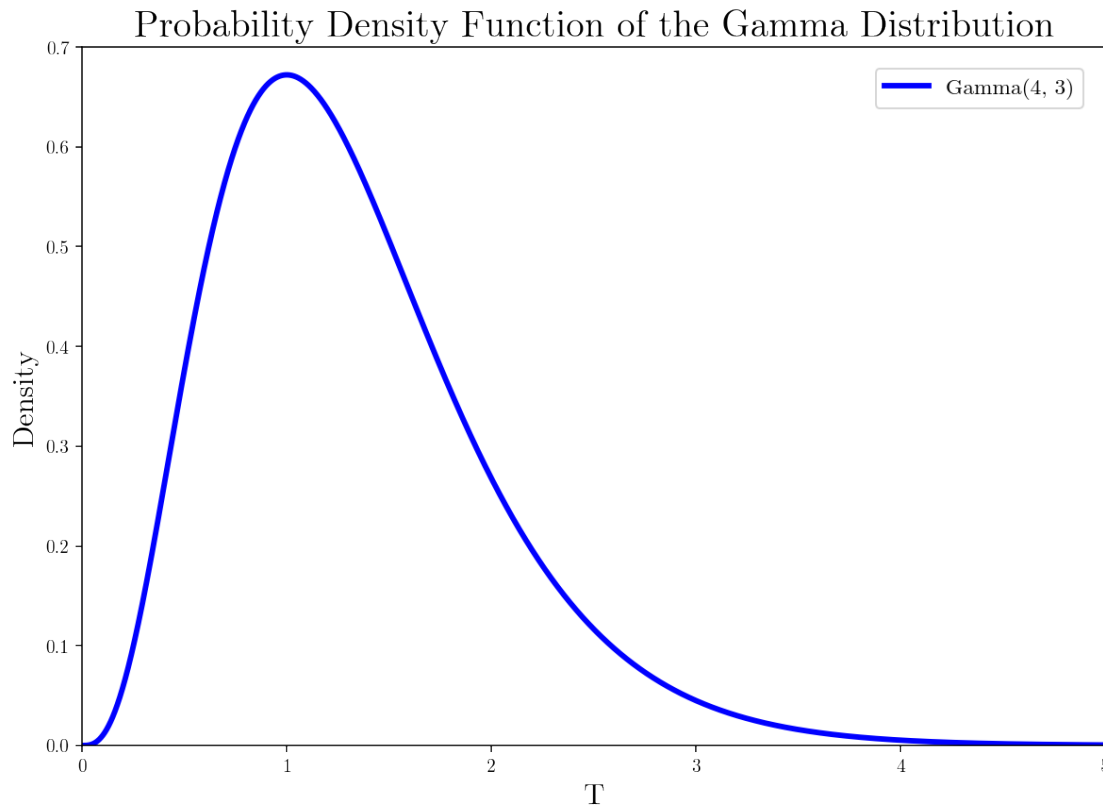


FIGURE 1: Plot of the probability density function of a gamma distribution.

Let's now prove that this equation is a probability density function. That is, let's show that:

$$\int_{D_X} f(x) \, dx = 1 \quad (2)$$

Proof. Let's use the following substitution to solve the integral:

$u = \lambda x; \qquad \qquad \qquad du = \lambda \, dx$
--

$$\begin{aligned}
 \int_{D_X} f(x) \, dx &= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \\
 &= \frac{1}{\lambda} \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{u}{\lambda}\right)^{\alpha-1} e^{-u} \, du \\
 &= \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} \, du \\
 &= 1
 \end{aligned} \quad (3)$$

QED

It is now clear that Equation 1 is a probability density function.

1.3 Cumulative Distribution Function

The cumulative distribution function (CDF) of the gamma distribution is defined as

$$F(x) = \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\lambda u} du \quad (4)$$

However, this integral has no definite form, that is, the integrand has no antiderivative. Hence, this CDF is often represented in its simplified form using the lower incomplete gamma function:^[5]

$$\gamma(\alpha, \lambda x) \equiv \int_0^x u^{\alpha-1} e^{-u} du \quad (5)$$

such that

$$F(x) = \frac{\gamma(\alpha, \lambda x)}{\Gamma(\alpha)} \quad (6)$$

Interestingly, there exists a special case of the CDF: when α is a positive integer, that is, when the distribution is an Erlang distribution (see subsection 4.5), the CDF follows the following series expansion:^[3]

$$F(x) = 1 - \sum_{i=0}^{\alpha-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} = e^{-\lambda x} \sum_{i=\alpha}^{\infty} \frac{(\lambda x)^i}{i!} \quad (7)$$

That being said, generally, although lacking a closed form, the cumulative distribution function of the gamma distribution may be graphed using software:

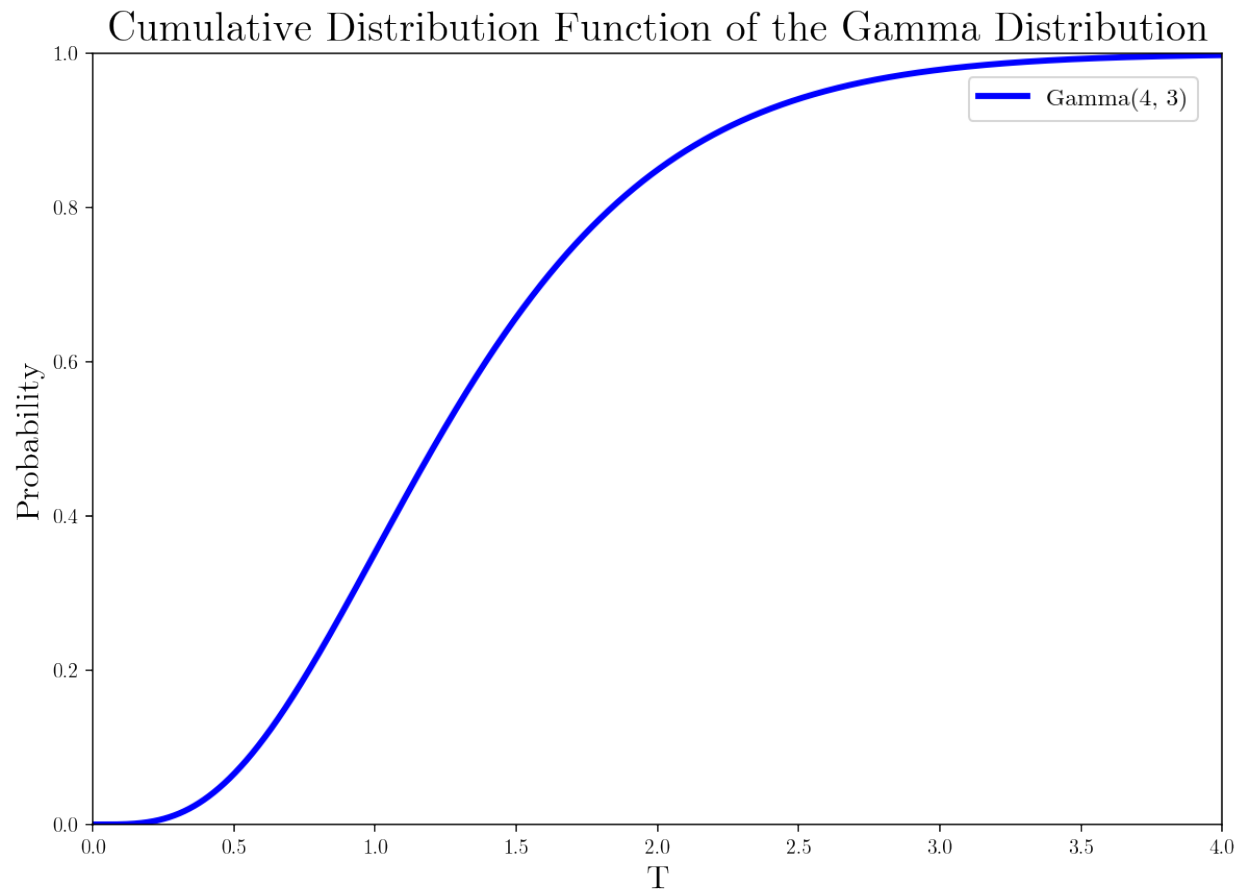


FIGURE 2: Plot of the cumulative distribution function of a gamma distribution.

2 Properties

Now that the fundamentals of the gamma distribution are set, a few properties stemming from the probability density function and the cumulative distribution function will be presented and proven. These properties, along with being extremely intriguing, are essential to multiple computations and statistical analyses.

2.1 Moment Generating Function

The moment generating function (MGF) of a real number, vector or matrix random variable X is used to compute a distribution's moment. It provides an extremely useful tool for distribution analysis and proofs related to many other properties. Unlike the characteristic function, which will be briefly explored in subsection 5.3, the MGF does not always exist for real values of a distribution.^[6] Here is the MGF of the gamma distribution:

$$M(t) = \left(\frac{\lambda}{\lambda - t} \right)^\alpha \quad \text{for } t < \lambda \quad (8)$$

This function may be found using the fact that $M(t) = E(e^{tx})$. Hence:

$$\begin{aligned} M(t) &= E(e^{tx}) \\ &= \int_0^\infty e^{tx} f(x) \, dx \\ &= \int_0^\infty e^{tx} \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \, dx \\ &= \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-(\lambda-t)x} \, dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-(\lambda-t)x} \, dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} (1) \\ &= \left(\frac{\lambda}{\lambda-t} \right)^\alpha \end{aligned} \quad (9)$$

2.2 Expected Value

2.2.1 Finding the Equation

The expected value, also known as the theoretical mean, may be found for the gamma distribution using the following formula:

$$\mu = E(X) = \frac{\alpha}{\lambda} \quad (10)$$

Interestingly, there are multiple ways to prove this result. Three different proofs will be explored in this paper.

2.2.1.1 Proof 1

The first technique uses the gamma distribution probability density function presented in subsection 1.2 as a means to get the expected value. In fact:

$$E(X) = \int_0^{\infty} x f(x) dx = \frac{\alpha}{\lambda} \quad (11)$$

Proof.

$$\begin{aligned} E(X) &= \int_0^{\infty} x f(x) dx \\ &= \int_0^{\infty} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha} e^{-\lambda x} dx \\ &= \int_0^{\infty} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{(\alpha+1)-1} e^{-\lambda x} dx \\ &= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \int_0^{\infty} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\lambda x} dx \\ &= \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} (1) \\ &= \frac{\alpha}{\lambda} \end{aligned} \quad (12)$$

QED

2.2.1.2 Proof 2

Using the moment generating function obtained in subsection 1.2, the expected value of a gamma distribution may be found. That is:

$$E(X) = M'(0) = \frac{\alpha}{\lambda} \quad (13)$$

Proof.

$$\begin{aligned} M'(t) &= \frac{d}{dt} \left(\frac{\lambda}{\lambda - t} \right)^\alpha \\ &= \lambda \alpha \left(\frac{\lambda}{\lambda - t} \right)^{\alpha-1} \frac{d}{dt} \left(\frac{1}{\lambda - t} \right) \\ &= \frac{\lambda \alpha}{(\lambda - t)^2} \left(\frac{\lambda}{\lambda - t} \right)^{\alpha-1} \end{aligned} \quad (14)$$

$$\begin{aligned} M'(0) &= \frac{\lambda \alpha}{\lambda^2} \left(\frac{\lambda}{\lambda} \right)^{\alpha-1} \\ &= \frac{\alpha}{\lambda} \end{aligned} \quad (15)$$

QED

2.2.1.3 Proof 3

A third way to demonstrate that $E(X) = \frac{\alpha}{\lambda}$, is by performing the following integral:

$$\int_0^\infty 1 - F(x) \, dx \quad (16)$$

However, this technique is not possible in the case of the gamma distribution, as there is no explicit formula for $F(x)$ as seen in subsection 1.3.

2.2.2 Special Case

The expected value may be used to derive several other properties. For example, consider i.i.d random variable $X \sim \mathcal{G}\text{amma}(\alpha, \lambda)$ such that $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$. Then:

$$\begin{aligned}
 E(X^\beta) &= \int_0^\infty x^\beta f(x) \, dx \\
 &= \int_0^\infty x^\beta \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \\
 &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+\beta)-1} e^{-\lambda x} \, dx \\
 &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\beta)}{\lambda^{\alpha+\beta}} \int_0^\infty \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} x^{(\alpha+\beta)} \, dx \\
 &= \frac{\Gamma(\alpha+\beta)}{\lambda^\beta \Gamma(\alpha)} \text{ for } \beta \in (-\alpha, \infty)
 \end{aligned} \tag{17}$$

The expected value and the variance can be determined using the provided formula above:

2.2.2.1 Expected Value

If $E(X^\beta) = \frac{\Gamma(\alpha+\beta)}{\lambda^\beta \Gamma(\alpha)}$, then:

$$\begin{aligned}
 E(X^1) &= \frac{\Gamma(\alpha+1)}{\lambda^1 \Gamma(\alpha)} \\
 E(X^1) &= \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} \\
 &= \frac{\alpha}{\lambda}
 \end{aligned} \tag{18}$$

2.2.2.2 Variance

In order to find the variance, $E(X^2)$ will need to be found using the same technique used to determine the formula for $E(X)$. If $E(X^\beta) = \frac{\Gamma(\alpha+\beta)}{\lambda^\beta \Gamma(\alpha)}$, then:

$$\begin{aligned} E(X^2) &= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \\ &= \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\lambda^2 \Gamma(\alpha)} \\ &= \frac{\alpha^2 + \alpha}{\lambda^2} \end{aligned} \tag{19}$$

The variance can now be found with this technique:

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{\alpha^2 + \alpha}{\lambda^2} - \left[\frac{\alpha}{\lambda}\right]^2 \\ &= \frac{\alpha^2 + \alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \\ &= \frac{\alpha}{\lambda^2} \end{aligned} \tag{20}$$

In the next section, we will examine other techniques that may be used to get the variance of the gamma distribution.

2.3 Variance

The variance is a measure of how "spread out" the data is. Indeed, the lower the variance is, the sharper the probability density function graph will be. More specifically, the variance is the square of the standard deviation of the random variable examined.^[7]

This property may be calculated using the following equation:

$$\text{Var}(X) = \frac{\alpha}{\lambda^2} \quad (21)$$

and this last equation may be proven in two different ways using either the moment generating function or an integral.

2.3.0.1 Proof 1

This first proof, featuring the moment generating function obtained in subsection 2.1, will prove the following relationship.

$$\text{Var}(X) = M''(0) - [M'(0)]^2 = \frac{\alpha}{\lambda^2} \quad (22)$$

Proof. Consider the following result from Equation 14 in subsection 2.2:

$$M'(t) = \frac{\lambda\alpha}{(\lambda - t)^2} \left(\frac{\lambda}{\lambda - t} \right)^{\alpha-1} \quad (23)$$

Then $M''(t)$ is derived to be:

$$\begin{aligned}
 M''(t) &= \frac{d}{dt} M'(t) \\
 &= \frac{d}{dt} \left[\frac{\lambda\alpha}{(\lambda-t)^2} \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} \right] \\
 &= \lambda\alpha \left(\frac{d}{dt} \left[\frac{1}{(\lambda-t)^2} \right] \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} + \left(\frac{1}{(\lambda-t)^2} \right) \frac{d}{dt} \left[\left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} \right] \right) \\
 &= \lambda\alpha \left(\left(\frac{2}{(\lambda-t)^3} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} + \left(\frac{\lambda(\alpha-1)}{(\lambda-t)^2} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-2} \frac{d}{dt} \left[\frac{1}{\lambda-t} \right] \right) \\
 &= \lambda\alpha \left(\left(\frac{2}{(\lambda-t)^3} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} + \left(\frac{\lambda(\alpha-1)}{(\lambda-t)^2} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-2} \left(\frac{1}{(\lambda-t)^2} \right) \right) \\
 &= \lambda\alpha \left(\left(\frac{2}{(\lambda-t)^3} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-1} + \left(\frac{\lambda(\alpha-1)}{(\lambda-t)^4} \right) \left(\frac{\lambda}{\lambda-t} \right)^{\alpha-2} \right)
 \end{aligned} \tag{24}$$

Let's not even try to simplify further and directly substitute $t = 0$:

$$\begin{aligned}
 M''(0) &= \lambda\alpha \left(\left(\frac{2}{(\lambda-0)^3} \right) \left(\frac{\lambda}{\lambda-0} \right)^{\alpha-1} + \left(\frac{\lambda(\alpha-1)}{(\lambda-0)^4} \right) \left(\frac{\lambda}{\lambda-0} \right)^{\alpha-2} \right) \\
 &= \lambda\alpha \left(\frac{2}{\lambda^3}(1) + \frac{\lambda(\alpha-1)}{\lambda^4}(1) \right) \\
 &= \lambda\alpha \left(\frac{2}{\lambda^3} + \frac{\alpha-1}{\lambda^3} \right) \\
 &= \lambda\alpha \left(\frac{\alpha+1}{\lambda^3} \right) \\
 &= \frac{\alpha^2 + \alpha}{\lambda^2}
 \end{aligned} \tag{25}$$

Finally:

$$\begin{aligned}
 \text{Var}(X) &= M''(0) - [M'(0)]^2 \\
 &= \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda} \right)^2 \\
 &= \frac{\alpha^2 + \alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \\
 &= \frac{\alpha^2 + \alpha - \alpha^2}{\lambda^2} \\
 &= \frac{\alpha}{\lambda^2}
 \end{aligned} \tag{26}$$

QED

2.3.0.2 Proof 2

This time, the proof features an integral to obtain the variance. That is:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\alpha}{\lambda^2} \quad (27)$$

Proof. In the following equations, $E(X)$ is the expected value obtained in subsection 2.2 and $E(X^2)$ is obtained using a process similar to Equation 11 of Proof 1 in subsection 2.2. That is:

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 f(x) \, dx \\ &= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\lambda x} \, dx \\ &= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{(\alpha+2)-1} e^{-\lambda x} \, dx \\ &= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} x^{(\alpha+2)-1} e^{-\lambda x} \, dx \\ &= \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\lambda^2 \Gamma(\alpha)} (1) \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} \end{aligned} \quad (28)$$

Then,

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 \\ &= \frac{\alpha^2 + \alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \\ &= \frac{\alpha^2 + \alpha - \alpha^2}{\lambda^2} \\ &= \frac{\alpha}{\lambda^2} \end{aligned} \quad (29)$$

QED

Conclusively, Equation 21 is confirmed.

2.4 Quantiles

Usually, the quantile function is defined as the inverse of the cumulative distribution function. Since the gamma distribution has no closed form for its CDF (see subsection 1.3) and, therefore, the latter cannot be inverted, it has no quantile function. In other words, there is no explicit form for the quantile function:

$$q(p) = F^{-1}(p) \quad (30)$$

because $F^{-1}(x)$ is not defined, since $f(x)$ has no antiderivative.

Visually, quantiles may be represented as equal subdivisions of a probability density function graph starting from the median. That is, in the following figure, the first quantile is represented by the section weighing 34% of the whole PDF, the second one represents the section weighing 13.5% and the last one represents the section weighing 2%. As may be noticed, these subdivisions are equally divided both sides of the median.

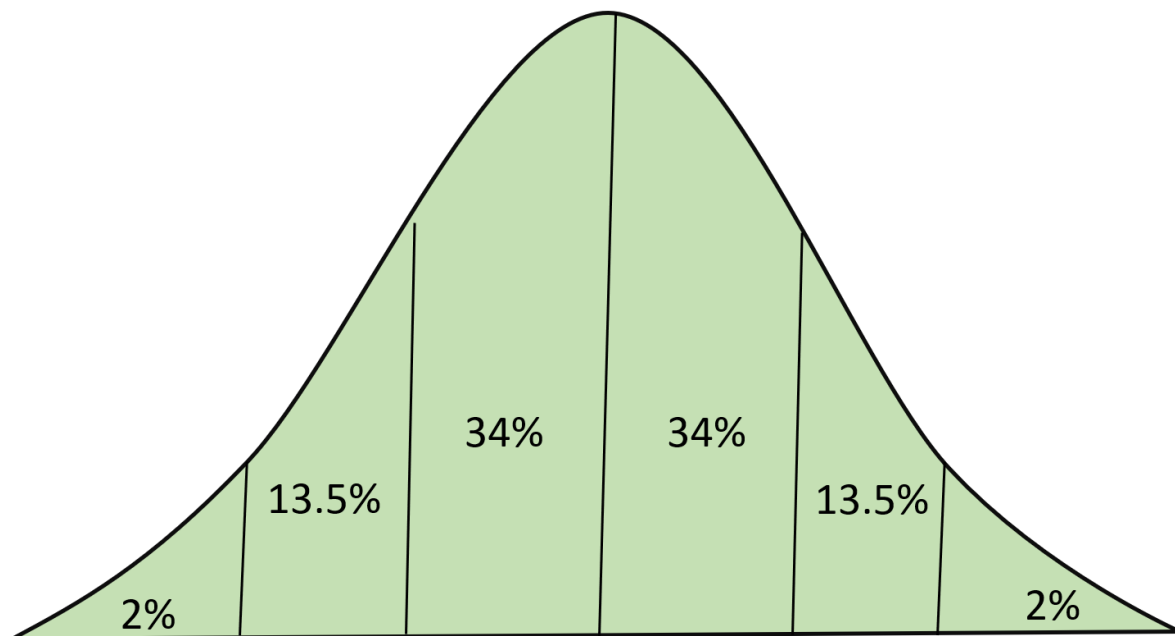


FIGURE 3: Visual representation of quantiles in a probability density distribution. Provided by GeeksforGeeks^[8]

2.5 Mode

The mode is defined as the value appearing the most through a certain set of data values, that is, the location of the highest peak on a probability density function plot. On Figure 4, the mode represents the value of T where the density is at its peak.

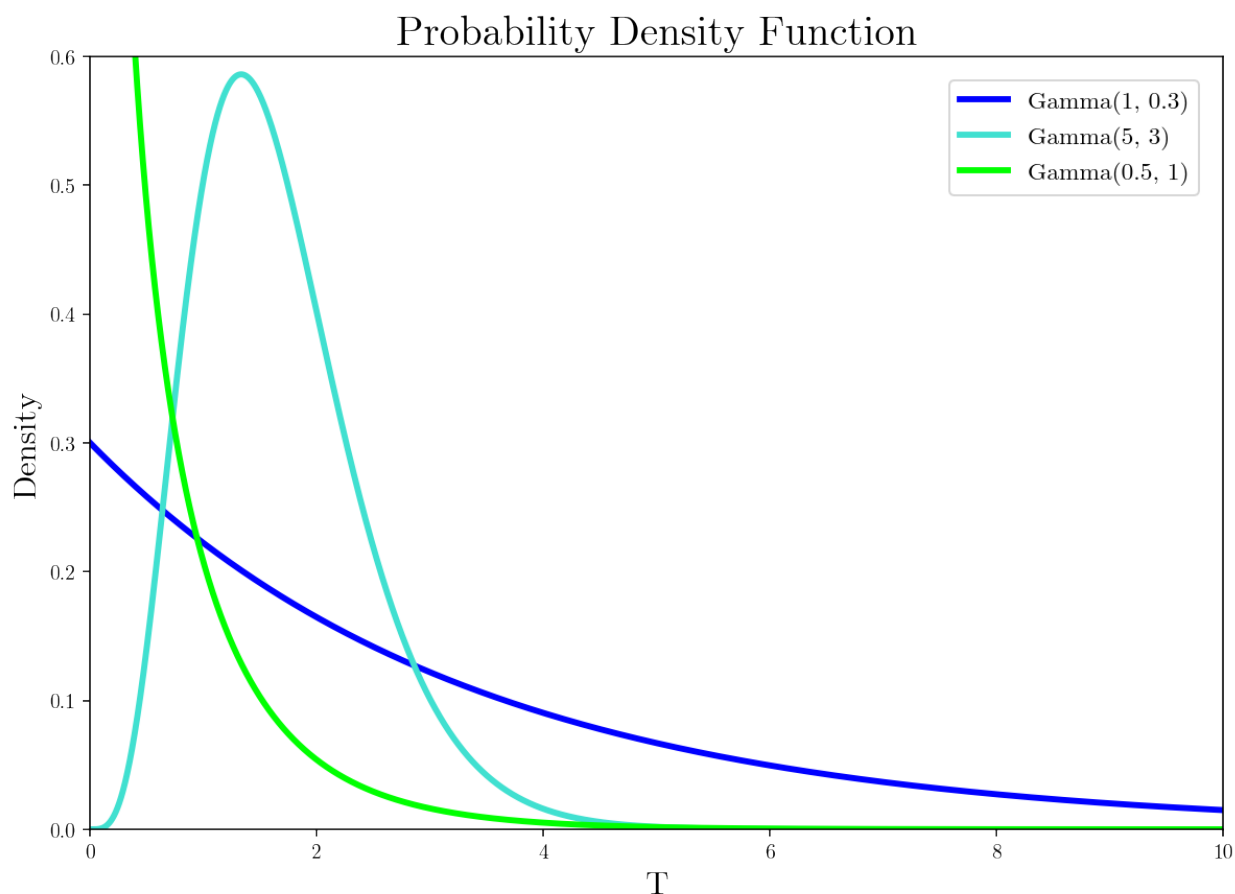


FIGURE 4: Plot of the probability density function of three distinct gamma distributions with different modes.

As may be observed in the plot of $\mathcal{G}\text{amma}(1, 0.3)$, the highest value of the density on the curve is 0.3 at $T = 0$ which corresponds to the mode of this specific parameterization of the distribution. The mode of the density function of $\mathcal{G}\text{amma}(1, 0.3)$ is thus 0.3. To be more precise and avoid the need for a graph, the following equation may be used to calculate the

mode of a gamma distribution:

$$\text{Mode} = \frac{(\alpha - 1)}{\lambda} \text{ for } \alpha \geq 1 \quad (31)$$

Ergo, for the plot of $\mathcal{G}\text{amma}(5, 3)$, the mode can be determined to be:

$$T = \frac{(5 - 1)}{3} = \frac{4}{3} \quad (32)$$

Accordingly, on Figure 4, the highest peak of the plot of $\mathcal{G}\text{amma}(5, 3)$ is at $\frac{4}{3}$.

2.5.1 Special Cases

There are two special cases concerning the mode of the gamma distribution: when the density of the PDF goes to infinity (the graph has a vertical asymptote) and when the parameter $\alpha = 1$.

2.5.1.1 Case 1:

The first special case is when the mode of the PDF never reaches its maximum value, that is, when there is a vertical asymptote on its plot. The plot of $\mathcal{G}\text{amma}(0.5, 1)$ on Figure 4 features this case. Indeed, if $f(x)$ is the probability density function, then:

$$\lim_{T \rightarrow 0^+} f(x) = \infty \quad (33)$$

Notice how this case arises when the condition stated in Equation 31 is broken, that is, that $\alpha \geq 1$. In fact, in the case of the purple plot, $\alpha = 0.5 < 1$

2.5.1.2 Case 2:

In a gamma distribution $\mathcal{G}\text{amma}(1, \lambda) \equiv \mathcal{E}\text{xp}(\lambda)$, the highest point of the plot of the PDF will be located at $T = 0$ and will correspond to a density value of λ . In Figure 4, the plot of $\mathcal{G}\text{amma}(1, 0.3)$ is an example of this special case as it has $\alpha = 1$ as a parameter. Hence, the mode of this latter parameterization of the gamma distribution is of $\lambda = 0.3$ at $T = 0$.

2.6 Inflection Points

The more the value of α increases, the narrower the density function graph of $\mathcal{G}\text{amma}(\alpha, \lambda)$ becomes narrower. For example, when $\alpha = 1$, the density function graph does not contain any inflection point. However, its y -intercept is equal to λ . As the value of α increases to 5, two inflection points, on each side of the mode, can be found on the graph. Inflection points are obtained by performing the second derivative of the curve. Below are found different possible shapes of the density function graph in function of the value of α . By increasing the value of α , the inflection points become closer to each other.

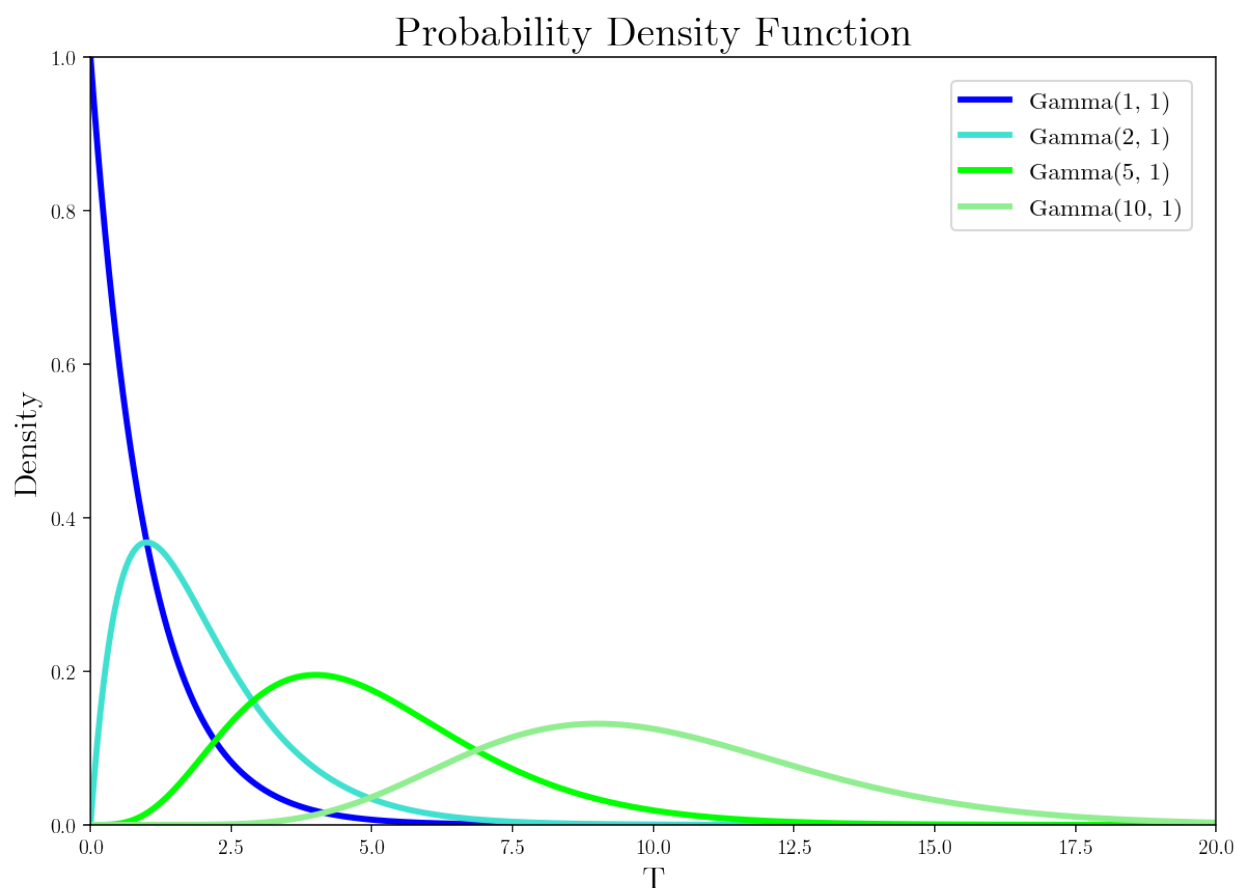


FIGURE 5: Plot of the probability density function of four distinct gamma distributions with different inflection points.

2.7 Median

The median of a probability distribution is the value that separates the area under the curve of its distribution function in two equal parts. It follows that, in the case of the gamma distribution, the median is defined as the value m for which:

$$\frac{1}{2} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^m x^{\alpha-1} e^{-\lambda x} dx \quad (34)$$

However, it is interesting to note that the median of the gamma distribution has no simple closed form : due to its highly varying shape, it cannot be represented in an equation as a function of its parameters. Interestingly, some have noticed that an approximation of the median can be performed in cases where the parameter $\lambda = 1$:^[3]

$$m = \alpha \left(1 - \frac{1}{9\alpha}\right) \quad (35)$$

Let's take the case of a distribution $\mathcal{G}\text{amma}(5, 1)$ to examine this claim. The approximation of the median using Equation 35 is as follows:

$$\begin{aligned} m &= \alpha \left(1 - \frac{1}{9\alpha}\right) \\ &= 5 \left(1 - \frac{1}{9 \cdot 5}\right) \\ &= \frac{44}{9} \\ &= 4.\bar{8} \end{aligned} \quad (36)$$

Let's confirm the validity of the answer using the graph of the probability density function of the distribution $\mathcal{G}\text{amma}(5, 1)$.

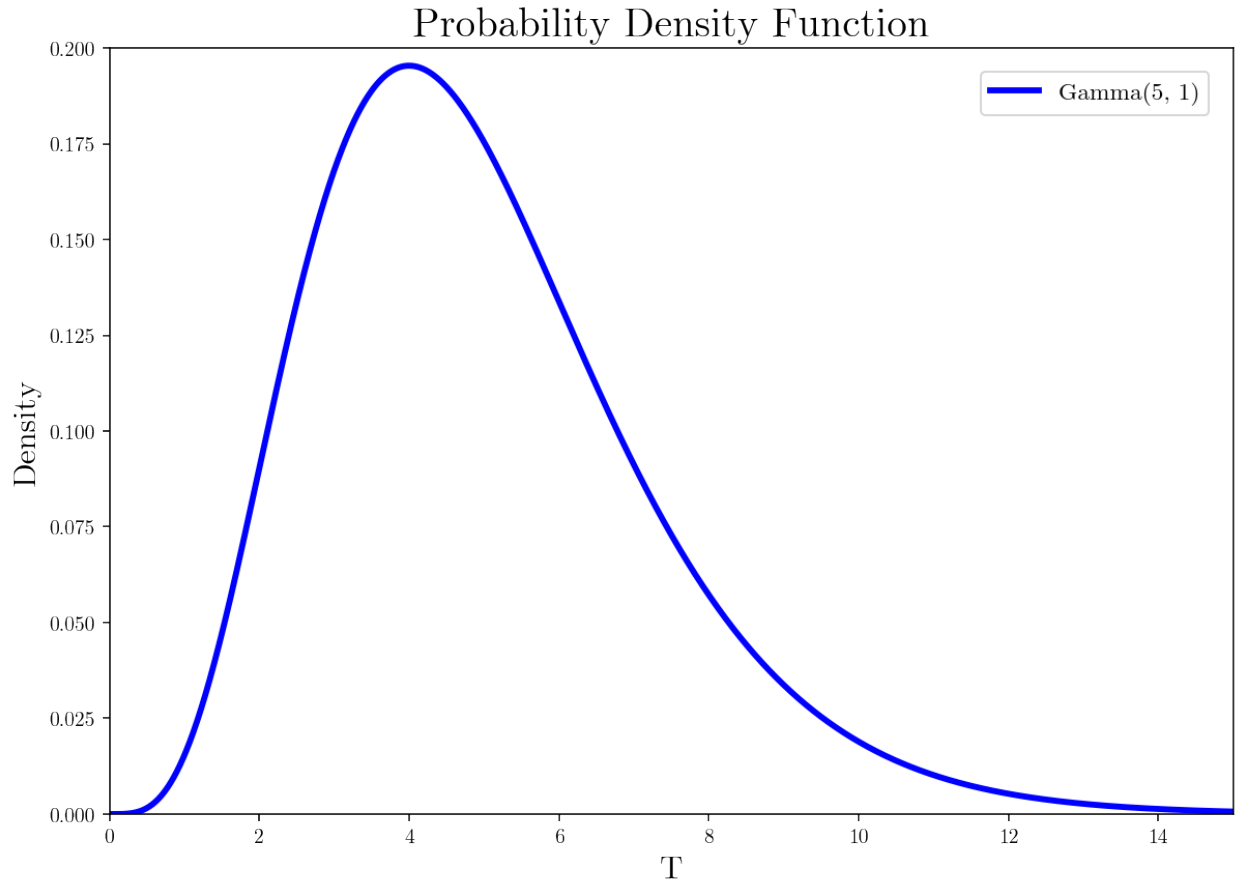


FIGURE 6: Plot of the probability density function of a gamma distribution with parameters $\alpha = 5$ and $\lambda = 1$.

The approximation seems to, in fact, be valid as the PDF is "weighted" more towards the right and its median seems to be between 4.5 and 5.

Still, there is comfort in the fact that some special cases of the gamma distribution do have a median, such as the exponential distribution. Indeed, the median of a continuous random variable following the gamma distribution such that $X \sim \mathcal{Gamma}(1, \lambda) \equiv \mathcal{Exp}(\lambda)$

may be obtained through a simple integration process:

$$\begin{aligned}\frac{1}{2} &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^m x^{\alpha-1} e^{-\lambda x} dx \\ \Rightarrow \frac{1}{2} &= -e^{-\lambda m} + 1 \\ \Rightarrow \frac{1}{2} &= e^{-\lambda m} \\ \Rightarrow \ln\left(\frac{1}{2}\right) &= -\lambda m \\ \Rightarrow -\ln(2) &= -\lambda m \\ \Rightarrow m &= \frac{\ln(2)}{\lambda}\end{aligned}\tag{37}$$

Hence, the median of an exponential distribution is defined as $\frac{\ln(2)}{\lambda}$.

2.8 Memoryless Property

The memoryless property indicates that a given probability distribution is independent of its history: the likelihood of something happening in the future has no relation to past events.

In statistics, a continuous random variable X satisfies the memoryless property if:

$$P(X > x + a | X > a) = P(X > x) \text{ for } a, x \geq 0 \quad (38)$$

It is easy to show that the exponential distribution (see subsection 4.1) displays the memoryless property.

$$\begin{aligned} P(X > x + a | X > a) &= \frac{P(X > x + a, X > a)}{P(X > a)} \\ &= \frac{P(X > x + a)}{P(X > a)} \\ &= \frac{1 - F_X(x + a)}{1 - F_X(a)} \\ &= \frac{e^{-\lambda(x+a)}}{e^{-\lambda a}} \\ &= e^{-\lambda x} \\ &= P(X > x) \end{aligned} \quad (39)$$

As a consequence, in a sequence of exponentially distributed events, the expected time until the next event always corresponds to $\frac{1}{\lambda}$, no matter how much time has passed since the last event.

However, it is interesting to note that when $\alpha > 1$, the future behaviour of the probability distribution is not independent of its past. Therefore, in general, the gamma distribution does not possess the memoryless property. This can be understood intuitively when looking at a case involving $\text{Gamma}(100, 1)$, which models the total time until 100 events separated by an average waiting time of 1 minute occur. It can be shown by brute force that if one has already waited $x = 110$ minutes, the probability of having to wait at least another

ten minutes for the hundredth event to occur is much lower than the probability of having to wait more than ten minutes in total.^{[9],[10]}

$$\begin{aligned}
 P(X > 110 + 10 | X > 110) &= \frac{P(X > 110 | X > 120) \cdot P(X > 120)}{P(X > 110)} \\
 &\approx 0.1760
 \end{aligned} \tag{40}$$

$$P(X > 10) \approx 1$$

$$P(X > 110 + 10 | X > 110) \neq P(X > 10)$$

Clearly, the memoryless property is not satisfied.

3 The Gamma Function

Studied by Leonhard Euler (1707-1783), the gamma function is related to the gamma distribution. This extremely handy function allows the generalization of the factorial notation. As a matter of fact, while the factorial $\alpha!$ is restricted to $\alpha \in \mathbb{N}$, the gamma function $\Gamma(\alpha)$ allows the computation of fractional and negative numbers such that $\alpha \in \mathbb{Q}$. By abusing the notation, the gamma function may even be used to compute $\Gamma(\alpha)$ such that $\alpha \in \mathbb{R}$. The relationship between the gamma function and the factorial notation is as follows:

$$\Gamma(\alpha) = (\alpha - 1)! \text{ for } \alpha \in \mathbb{N} \quad (41)$$

3.1 Definition

First of all, the gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \text{ for } \alpha \in \mathbb{R}^+ \quad (42)$$

Assuming that $\alpha > 1$, the closed form of the gamma function may be found with integration by parts:

$u = x^{\alpha-1};$ $du = (\alpha - 1)x^{\alpha-2} dx;$	$dv = e^{-x} dx$ $v = -e^{-x}$
--	-----------------------------------

$$\begin{aligned}
 \Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \\
 &= - \left[x^{\alpha-1} e^{-x} \right]_0^\infty + (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} dx \\
 &= 0 + (\alpha - 1) \Gamma(\alpha - 1) \\
 &= (\alpha - 1) \Gamma(\alpha - 1)
 \end{aligned} \quad (43)$$

Therefore,

$$\Gamma(\alpha) = (\alpha - 1) \Gamma(\alpha - 1) \text{ for } \alpha \in (1, \infty) \quad (44)$$

Furthermore, $\Gamma(1)$ is defined separately as being:

$$\Gamma(1) = \int_0^\infty e^{-x} dx = \left[-e^{-x} \right]_0^\infty = 1 \quad (45)$$

3.1.1 Special Case of $\Gamma(1/2)$

An extremely interesting case of the gamma function is $\Gamma(1/2)$. In fact, this result allows the computation of fractional values of α such that $\alpha \in \mathbb{Q}$. This special case is as follows:

Claim 3.1.1.

$$\Gamma(1/2) = \sqrt{\pi}$$

3.1.1.1 Initial Setup

First, to prove Claim 3.1.1, let's start by setting up some equations in preparation for the proof.

$$I = \int_0^\infty e^{-x^2} dx \tag{46}$$

From this result, one obtains that:

$$\begin{aligned} I^2 &= \left(\int_0^\infty e^{-x^2} dx \right) \left(\int_0^\infty e^{-y^2} dy \right) \\ &= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy \end{aligned} \tag{47}$$

To solve this integral easily, one may use the following change of variables:

$$x = r \cos \theta \qquad y = r \sin \theta$$

This change of variable requires the Jacobian of the transformation to be computed:

$$\mathcal{J} = \det \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \det \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r \tag{48}$$

Using the result obtained in Equation 48, the integral in Equation 47 may be computed using polar coordinates:

$$\begin{aligned}
 I^2 &= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy = \int_0^{\frac{\pi}{2}} \int_0^\infty r e^{-r^2} dr d\theta \\
 &= \frac{\pi}{2} \int_0^\infty r e^{-r^2} dr \\
 &= \frac{\pi}{2} \left[-\frac{e^{-r^2}}{2} \right]_0^\infty \\
 &= \frac{\pi}{2} \frac{1}{2} \\
 &= \frac{\pi}{4}
 \end{aligned} \tag{49}$$

It then follows that Equation 46 may be solved:

$$I = \sqrt{I^2} = \int_0^\infty e^{-x^2} dx = \sqrt{\frac{\pi}{4}} = \frac{\sqrt{\pi}}{2} \tag{50}$$

Accordingly, Corollaries 3.1.1 and 3.1.2 logically follow:

Corollary 3.1.1.

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi} \tag{51}$$

Corollary 3.1.2.

$$\int_{-\infty}^\infty e^{-x^2/2} dx = \sqrt{2\pi} \tag{52}$$

Let's prove these corollaries:

Proof. Consider this substitution of variables:

$$z = \sqrt{2}x \qquad dz = \sqrt{2} dx$$

Using this substitution, Corollaries 3.1.1 and 3.1.2 may be proven:

$$\int_{-\infty}^\infty e^{-x^2} dx = \frac{1}{\sqrt{2}} \int_{-\infty}^\infty e^{-z^2/2} dz = \sqrt{\pi} \tag{53}$$

$$\int_{-\infty}^\infty e^{-x^2/2} dx = \sqrt{2} \sqrt{\pi} = \sqrt{2\pi} \tag{54}$$

QED

3.1.1.2 The Proof

The proof for Claim 3.1.1 may now be articulated:

Proof. Consider, once again, this other substitution of variables:

$$x = \sqrt{z} \qquad dx = \frac{1}{2\sqrt{z}} dz = \frac{z^{-1/2}}{2} dz$$

Using it, the following integral may be solved:

$$\begin{aligned} \int_0^\infty e^{-x^2} dx &= \frac{1}{2} \int_0^\infty z^{-1/2} e^{-z} dz \\ &= \frac{1}{2} \Gamma(1/2) \\ &= \frac{\sqrt{\pi}}{2} \end{aligned} \tag{55}$$

Conclusively, Claim 3.1.1 is proven:

$$\Gamma(1/2) = \sqrt{\pi} \tag{56}$$

QED

3.1.2 Negative Values of α

As was mentioned earlier, by abusing the identity shown in Equation 44, one may extend the gamma function to negative non-integer values of α such that $\alpha \in \mathbb{R} \setminus \mathbb{Z}^-$. For example, to find $\Gamma(-9/2)$, one may compute:

$$\begin{aligned}
 \Gamma(1/2) &= (-1/2)\Gamma(-1/2) \\
 &= (-1/2)(-3/2)\Gamma(-3/2) \\
 &= (3/4)(-5/2)\Gamma(-5/2) \\
 &= (-15/8)(-7/2)\Gamma(-7/2) \\
 &= (105/16)(-9/2)\Gamma(-9/2) \\
 &= (-945/32)\Gamma(-9/2)
 \end{aligned} \tag{57}$$

$$\therefore \Gamma(-9/2) = -\frac{32}{945}\Gamma(1/2) = -\frac{32}{945}\sqrt{\pi} \tag{58}$$

Hence, to calculate the value of a gamma function with a negative argument, one might modify a bit the result obtained in Equation 44 such that:

$$\Gamma(\alpha) = \frac{\Gamma(\alpha + 1)}{\alpha} \text{ for } \alpha \in \mathbb{R}^- \setminus \mathbb{Z} \tag{59}$$

for example, take the same example of $\Gamma(-9/2)$

$$\begin{aligned}
 \Gamma(-9/2) &= \frac{\Gamma(-7/2)}{(-9/2)} \\
 &= \frac{\Gamma(-5/2)}{(-9/2)(-7/2)} \\
 &= \frac{\Gamma(-3/2)}{(63/4)(-5/2)} \\
 &= \frac{\Gamma(-1/2)}{(-315/8)(-3/2)} \\
 &= \frac{\Gamma(1/2)}{(945/16)(-1/2)} \\
 &= \frac{\Gamma(1/2)}{(-945/32)}
 \end{aligned} \tag{60}$$

$$\therefore \Gamma(-9/2) = -\frac{32}{945}\Gamma(1/2) = -\frac{32}{945}\sqrt{\pi} \tag{61}$$

3.2 Relationship to Factorial Notation

Using a proof by induction, gamma functions can be shown to be related to the factorial notation.

Proof. Let us prove Equation 41, that is, that:

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx = (n-1)! \text{ for } n \in \mathbb{N} \quad (62)$$

For $n = 1$:

$$\begin{aligned} \Gamma(1) &= \int_0^\infty x^{1-1} e^{-x} dx \\ &= \int_0^\infty e^{-x} dx \\ &= 1 \\ &= (1-1)! \\ &= 0! \end{aligned} \quad (63)$$

Assume that the equation holds for $n = k$ such that:

$$\Gamma(k) = (k-1)! \quad (64)$$

Then, let us prove that the equation is valid for $n = k + 1$ using integration by parts:

$u = x^k;$ $du = kx^{k-1} dx;$	$dv = e^{-x} dx$ $v = -e^{-x}$
-----------------------------------	-----------------------------------

$$\begin{aligned} \Gamma(k+1) &= \int_0^\infty x^{(k+1)-1} e^{-x} dx \\ &= \left[-x^k e^{-x} \right]_0^\infty + k \int_0^\infty x^{k-1} e^{-x} dx \\ &= 0 + k\Gamma(k) \\ &= k(k-1)! \\ &= ((k+1)-1)! \\ &= k! \end{aligned} \quad (65)$$

Therefore, Equation 41 is proven by mathematical induction.

QED

3.3 Plotting the Gamma Function

Using everything that was computed previously, the graph of $\Gamma(\alpha)$ may be plotted. It includes both positive, negative and fractional values of α . Notice how the asymptotes of the graph are found at every negative integer value; the function, as was mentioned previously, is not defined for negative integers.

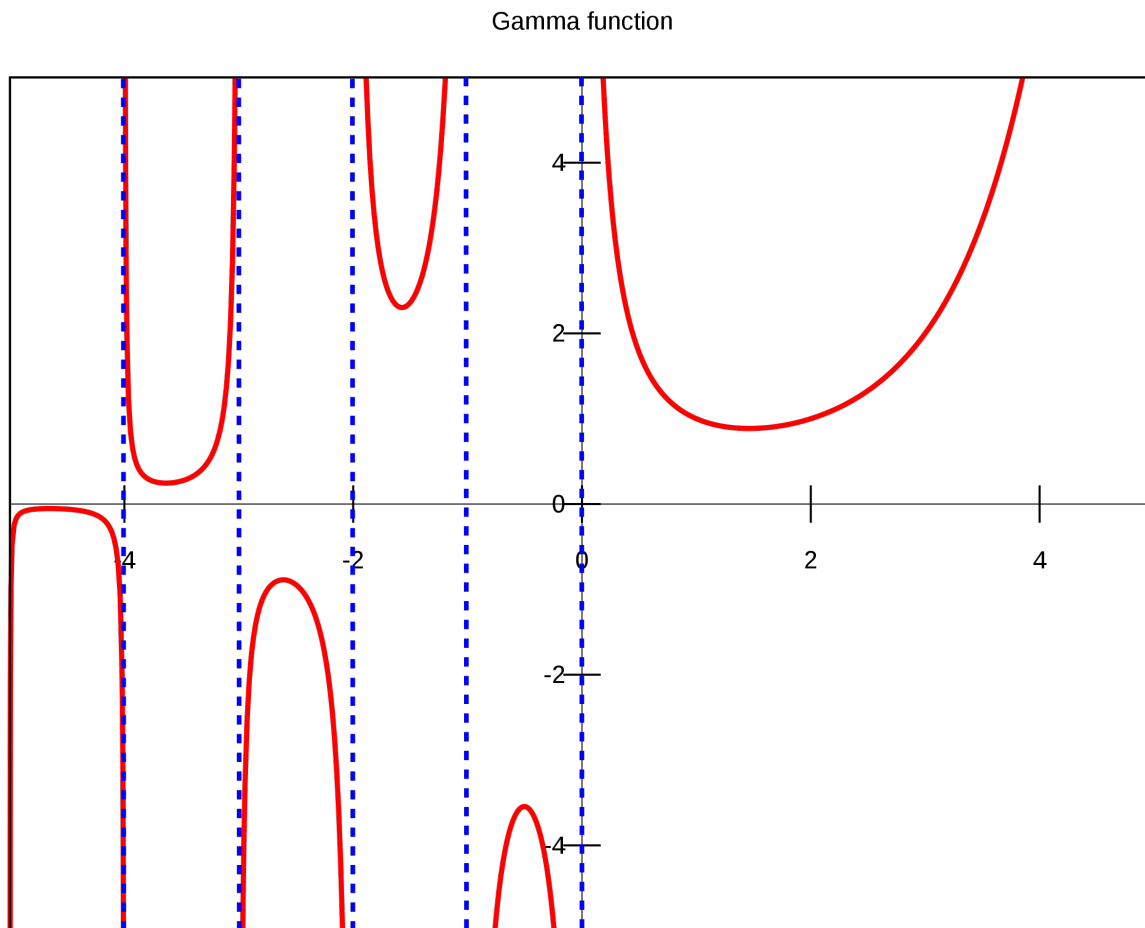


FIGURE 7: Plot of the function $\Gamma(\alpha)$ in red and the asymptotes of the function in dotted blue lines.

4 Special Cases of the Gamma Distribution & Related Proofs

The gamma distribution has many common parameterizations. Hence, in order to make working with them easier, they were each given a specific name such as the chi-square distribution, the exponential distribution and more. In this section, we will present some of these special cases and prove their relationship.

4.1 The Exponential Distribution

The gamma distribution can model the elapsed time between random and independent events. However, what should one do when one wants to strictly model the time to wait before a single new event?

If a unique event is modelled, it follows that the shape parameter is limited to $\alpha = 1$. In such a case, the gamma distribution is said to follow an exponential distribution: more specifically, if $X \sim \mathcal{G}\text{amma}(1, \lambda)$, then it can also be said that $X \sim \mathcal{E}\text{xp}(\lambda)$. The continuous random variable X thus follows an exponential distribution with rate parameter λ , where the rate parameter represents how quickly events occur. More concretely:

$$X \sim \mathcal{G}\text{amma}(1, \lambda) \equiv \mathcal{E}\text{xp}(\lambda) \quad (66)$$

Indeed, using a simple substitution of the variables, one gets that:

$$\begin{aligned} f(x) &= \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \\ &= \frac{\lambda^1 x^0}{\Gamma(1)} e^{-\lambda x} \\ &= \lambda e^{-\lambda x} \end{aligned} \quad (67)$$

$$\therefore X \sim \mathcal{G}\text{amma}(1, \lambda) \equiv X \sim \mathcal{E}\text{xp}(\lambda) \quad (68)$$

Accordingly, it can be demonstrated that the sum of exponential random variables is a gamma random variable. That is, if X_1, X_2, \dots, X_n are i.i.d random variables following $\mathcal{Exp}(\lambda)$, then:

$$Y = \sum_{i=1}^k [X_i] \sim \mathcal{Gamma}(k, \lambda) \quad (69)$$

4.1.1 Discovery of the Gamma Probability Density Function

Here, we propose that the PDF of the gamma distribution was found by adding multiple i.i.d random variables following an exponential distribution $\mathcal{Exp}(\lambda)$. Here is a proof by induction demonstrating this supposition:^[11]

Proof. First, let us find the sum of two exponential random variables, which we will denote S_2 :

$$\begin{aligned} f_{S_2} &= f_{X_1+X_2}(t) \\ &= \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s} \, ds \\ &= \lambda^2 e^{-\lambda t} \int_0^t 1 \, ds \\ &= \lambda^2 t e^{-\lambda t} \\ &= \frac{\lambda^2 t^{2-1} e^{-\lambda t}}{(2-1)!} \end{aligned} \quad (70)$$

With this result, we can now determine the probability distribution function of the sum of three, as well as four exponential random variables. Let's start with the sum of three

variables:

$$\begin{aligned}
 f_{X_1+X_2+X_3}(t) &= f_{S_2+X_3}(t) \\
 &= \int_{-\infty}^{\infty} f_{S_2}(t-s)f_{X_3}(s) \, ds \\
 &= \int_{-\infty}^{\infty} f_{X_3}(t-s)f_{S_2}(s) \, ds \\
 &= \int_0^t \lambda e^{-\lambda(t-s)} \lambda^2 s e^{-\lambda s} \, ds \\
 &= \lambda^3 e^{-\lambda t} \int_0^t s \, ds \\
 &= \frac{\lambda^3 t^2 e^{-\lambda t}}{2!} \\
 &= \frac{\lambda^3 t^{3-1} e^{-\lambda t}}{(3-1)!}
 \end{aligned} \tag{71}$$

Then, let's sum four variables:

$$\begin{aligned}
 f_{X_1+X_2+X_3+X_4}(t) &= f_{S_3+X_4}(t) \\
 &= \int_{-\infty}^{\infty} f_{S_3}(t-s)f_{X_4}(s) \, ds \\
 &= \int_{-\infty}^{\infty} f_{X_4}(t-s)f_{S_3}(s) \, ds \\
 &= \int_0^t \lambda e^{-\lambda(t-s)} \frac{\lambda^3 s^2 e^{-\lambda s}}{2!} \, ds \\
 &= \frac{\lambda^4 e^{-\lambda t}}{2!} \int_0^t s^2 \, ds \\
 &= \frac{\lambda^4 t^3 e^{-\lambda t}}{3!} \\
 &= \frac{\lambda^4 t^{4-1} e^{-\lambda t}}{(4-1)!}
 \end{aligned} \tag{72}$$

A pattern can clearly be seen emerging from those results. Hence, we make the claim that:

$$f_{S_n}(t) = f_{\sum_{i=1}^n X_i}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} \tag{73}$$

To prove this claim, let us assume that:

$$\begin{aligned} f_{S_{n-1}}(t) &= f_{\sum_{i=1}^{n-1} X_i}(t) \\ &= \frac{\lambda^{n-1} t^{n-2} e^{-\lambda t}}{(n-2)!} \end{aligned} \quad (74)$$

By convolution, the probability distribution function of an infinite sum of exponential random variables can now be found.

$$\begin{aligned} f_{S_n}(t) &= f_{\sum_{i=1}^n X_i}(t) \\ &= f_{\sum_{i=1}^{n-1} [X_i] + X_n}(t) \\ &= f_{S_{n-1} + X_n}(t) \\ &= \int_{-\infty}^{\infty} f_{S_{n-1}}(t-s) f_{X_n}(s) \, ds \\ &= \int_{-\infty}^{\infty} f_{X_n}(t-s) f_{S_{n-1}}(s) \, ds \\ &= \int_0^t \lambda e^{-\lambda(t-s)} \frac{\lambda^{n-1} s^{n-2} e^{-\lambda s}}{(n-2)!} \, ds \\ &= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \int_0^t s^{n-2} \, ds \\ &= \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} \\ &= \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)} \end{aligned} \quad (75)$$

This corresponds, obviously, to the probability distribution function of a gamma distribution.

Therefore:

$$S_n \sim \text{Gamma}(n, \lambda) \quad (76)$$

Hence, we get a density function whose parameter n can correspond to all real values instead of being limited to natural numbers. QED

4.1.2 Sum of Exponential Random Variables

It is interesting to know that the gamma distribution is infinitely divisible. That is, any random variable following a gamma distribution $\mathcal{G}\text{amma}(\alpha, \lambda)$ may be expressed as a sum of n other random variables following a gamma distribution as long as their shape parameters add up to α and their rate parameter is λ .

4.1.2.1 Using Convolution

Using the notion of convolution, then

$$\begin{aligned}
 f(z) &= \int_0^z f_1(z-y)f_2(y) \, dy \\
 &= \int_0^z \frac{\lambda^{\alpha_1} e^{-\lambda(z-y)} (z-y)^{\alpha_1-1}}{\Gamma(\alpha_1)} \cdot \frac{\lambda^{\alpha_2} e^{-\lambda y} y^{\alpha_2-1}}{\Gamma(\alpha_2)} \, dy \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z e^{-\lambda z} (z-y)^{\alpha_1-1} y^{\alpha_2-1} \, dy
 \end{aligned} \tag{77}$$

$y = zt; \qquad \qquad \qquad dy = z \, dt; \qquad \qquad \qquad t = \frac{y}{z}$

Therefore,

$$\begin{aligned}
 f(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 e^{-\lambda z} (z-zt)^{\alpha_1-1} (zt)^{\alpha_2-1} z \, dt \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 z^{\alpha_1+\alpha_2-1} e^{-\lambda z} (1-t)^{\alpha_1-1} t^{\alpha_2-1} \, dt \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} \int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} \, dt \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} B(\alpha_2, \alpha_1)
 \end{aligned} \tag{78}$$

Here, let us now use an alternative definition of the beta distribution to obtain our final

result:

$$\begin{aligned} f(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} \cdot \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} \text{ for } z > 0, \alpha_1 + \alpha_2 - 1 > 0 \end{aligned} \quad (79)$$

Since $f(z)$ is in fact a probability density function, then:

$$\begin{aligned} 1 &= \int_0^\infty \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} B(\alpha_2, \alpha_1) dz \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} B(\alpha_2, \alpha_1) \int_0^\infty e^{-\lambda z} z^{\alpha_1+\alpha_2-1} dz \end{aligned} \quad (80)$$

Finally, using the definition of the gamma distribution, we arrive to:

$$\begin{aligned} 1 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} B(\alpha_2, \alpha_1) \frac{\Gamma(\alpha_1+\alpha_2)}{\lambda^{\alpha_1+\alpha_2}} \\ &= \frac{B(\alpha_2, \alpha_1)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \end{aligned} \quad (81)$$

Thus, as a by-product, we have found that:

$$B(\alpha_2, \alpha_1) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \quad (82)$$

thereby confirming our initial assumption about the alternative definition of the beta distribution.

4.1.2.2 Using a Change of Variables

We will now prove this same result using a change of variables technique. First, let $Z = X + Y$. We will now find the distribution function of Z :

$$\begin{aligned}
 F_{X+Y}(z) &= P(Z \leq z) \\
 &= P(X + Y \leq z) \text{ for } x, y \in (0, \infty) \\
 &= P(X \leq z - y) \\
 &= \int_0^\infty \int_0^{z-y} f_X(x) f_Y(y) dx dy \\
 &= \int_0^\infty F_x(z - y) f_Y(y) dy
 \end{aligned} \tag{83}$$

The associated probability density function can now be found.

$$\begin{aligned}
 f_{x+y}(z) &= \int_0^\infty f_X(z - y) f_Y(y) dy \text{ for } z \in (0, \infty) \\
 &= \int_0^\infty \int_0^{z-y} f_x(x) f_y(y) dx dy \\
 &= \int_0^\infty \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} (z - y)^{\alpha_1-1} e^{-\lambda(z-y)} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} y^{\alpha_2-1} e^{-\lambda y} dy \\
 &= \frac{\lambda^{\alpha_1} \lambda^{\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^\infty (z - y)^{\alpha_1-1} e^{-\lambda(z-y)} y^{\alpha_2-1} e^{-\lambda y} dy \\
 &= \frac{\lambda^{\alpha_1} \lambda^{\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^\infty (z - y)^{\alpha_1-1} e^{\lambda(-z+y-y)} y^{\alpha_2-1} dy
 \end{aligned} \tag{84}$$

$y = zu;$	$dy = z du$
-----------	-------------

Let us now make a simple substitution so that the final answer can be revealed.

$$\begin{aligned}
 f_{X+Y}(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} e^{-\lambda z} \int_0^\infty (z - zu)^{\alpha_1-1} (zu)^{\alpha_2-1} z du \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} e^{-\lambda z} \int_0^\infty z^{\alpha_1+\alpha_2} (1 - u)^{\alpha_1-1} u^{\alpha_2-1} du \\
 &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} \int_0^\infty u^{\alpha_2-1} (1 - u)^{\alpha_1-1} du
 \end{aligned} \tag{85}$$

We will note here that the integral part of this equation corresponds to the Beta distribution function. Hence, we get:

$$f_{X+Y}(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda z} z^{\alpha_1+\alpha_2-1} B(\alpha_1, \alpha_2) \quad (86)$$

Using an alternative definition of the beta distribution, which was proven in the last demonstration, we arrive to our final answer:

$$f_{X+Y}(z) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} e^{-\lambda z} z^{(\alpha_1+\alpha_2)-1} \quad (87)$$

This corresponds, most clearly, to the probability density function of a gamma distribution.

$$Z \sim \mathcal{G}\text{amma}(\alpha_1 + \alpha_2, \lambda) \quad (88)$$

Here is a graph showing this result:

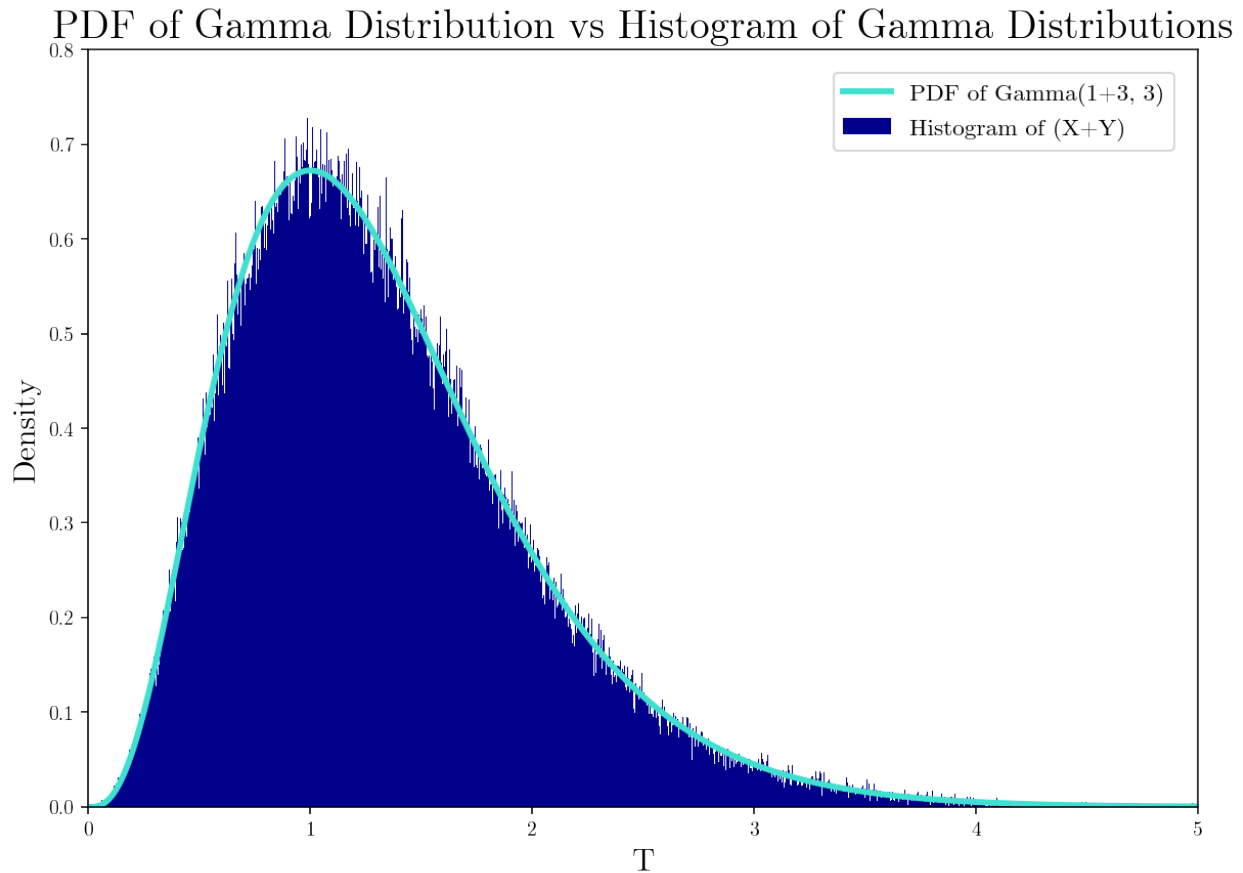


FIGURE 8: Comparison of the histogram of a sum of gamma random variables with the probability density function of a gamma distribution.

4.2 The Beta Distribution

A random variable X is said to follow a beta distribution $\text{Beta}(\alpha, \beta)$ with two parameters α and β if its probability distribution function is defined by:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \text{ for } 0 \leq x \leq 1 \text{ and } \alpha, \beta > 0 \quad (89)$$

where, from Equation 82:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (90)$$

Here is a graph of the probability density function:

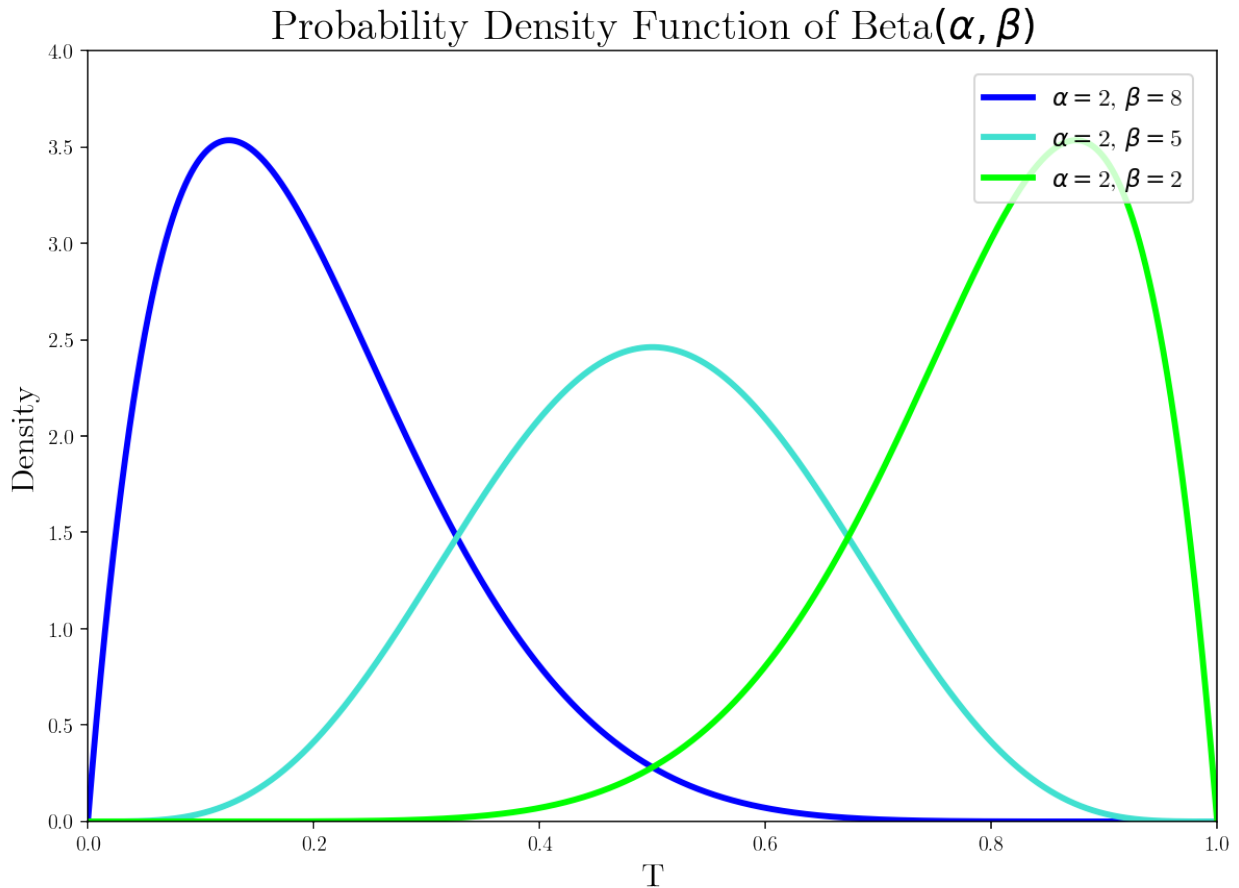


FIGURE 9: Plot of the probability distribution function of a beta distribution for various parameterizations.

When looking at the numerator of the beta distribution function, similarities with the distribution function of the binomial distribution for independently distributed random events are evident.

$$p(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ for } x \in \mathbb{N} \cup \{0\} \quad (91)$$

Indeed, the two distribution functions are analogous to a certain extent, for while the binomial distribution models the number of successes, the beta distribution models the probability of success. Since $\alpha - 1$ represents the number of successes and $\beta - 1$ represents the number of failures, modifying those parameters changes the shape of the probability distribution to model varying probabilities. As α increases (more successful events), the bulk of the probability distribution will shift towards the right, whereas an increase in β moves the distribution towards the left (more failures). Finally, the distribution will narrow if both α and β increase, for this increases the certainty regarding possible outcomes.^{[12], [13]}

4.2.1 Quotient of Gamma Random Variables

Let's show how this distribution is related to the gamma distribution. Indeed, for $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$, then $\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$. Let's prove it using two different techniques:^[11]

4.2.1.1 Using a Jacobian

Proof. First, let:

$$\boxed{Y_1 = X_1 + X_2; \quad Y_2 = \frac{X_1}{X_1 + X_2}}$$

Therefore:

$$\boxed{x_1 = y_1 y_2; \quad x_2 = y_1(1 - y_2)}$$

$$\mathcal{J} = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \det \begin{bmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{bmatrix} = -y_1 \quad (92)$$

$$\begin{aligned} f(y_1, y_2) &= f(x_1, x_2) \cdot \mathcal{J} \\ &= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} e^{-\lambda(x_1 + x_2)} \text{ for } x_1, x_2 > 0 \\ &= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (y_1 y_2)^{\alpha_1 - 1} (y_1(1 - y_2))^{\alpha_2 - 1} e^{-\lambda y_1} y_1 \\ &= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 + \alpha_2 - 1} y_2^{\alpha_2 - 1} (1 - y_2)^{\alpha_2 - 1} e^{-\lambda y_1} \end{aligned} \quad (93)$$

Hence:

$$\begin{aligned} f(y_2) &= \int_0^\infty f(y_1, y_2) dy_1 \\ &= \int_0^\infty \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 + \alpha_2 - 1} y_2^{\alpha_2 - 1} (1 - y_2)^{\alpha_2 - 1} e^{-\lambda y_1} dy_1 \end{aligned} \quad (94)$$

The beta distribution may now be obtained through the following calculations:

$$y = \lambda y_1 \Rightarrow y_1 = \frac{y}{\lambda}$$

$$\begin{aligned}
 \frac{1}{\lambda} dy &= dy_1 \\
 &= \frac{\lambda^{\alpha_1+\alpha_2} y_2^{\alpha_2-1} (1-y_2)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty \frac{e^{-y}}{\lambda} \left(\frac{y}{\lambda}\right)^{\alpha_1+\alpha_2-1} dy \\
 &= \frac{\lambda^{\alpha_1+\alpha_2} y_2^{\alpha_2-1} (1-y_2)^{\alpha_2-1}}{\lambda^{\alpha_1+\alpha_2} \Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty e^{-y} y^{\alpha_1+\alpha_2-1} dy \\
 &= \frac{y_2^{\alpha_2-1} (1-y_2)^{\alpha_2-1} \Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \\
 &= \frac{y_2^{\alpha_2-1} (1-y_2)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} \text{ for } 0 \leq y \leq 1 \text{ and } \alpha_1, \alpha_2 > 0
 \end{aligned} \tag{95}$$

This last result is exactly the equation of the beta distribution presented in Equation 89, but with different variables. QED

4.2.1.2 Using a Change of Variables

Proof. Alternatively, using a change of variables, consider the following random variable:

$$Z = \frac{X_1}{X_1 + X_2} \quad (96)$$

then:

$$\begin{aligned} P(Z \leq z) &= P\left(\frac{X_1}{X_1 + X_2} \leq z\right) \\ &= P(X_1 \leq z(x_1 + x_2)) \\ &= P((1 - z)X_1 \leq zx_2) \\ &= P\left(X_1 \leq \frac{zx_2}{1 - z}\right) \text{ for } X_2 \in (0, \infty) \\ &= \int_0^\infty \int_0^{\frac{zx_2}{1-z}} f(x_1) dx_1 f(x_2) dx_2 \\ &= \int_0^\infty F_{X_1}\left(\frac{zx_2}{1-z}\right) f(x_2) dx_2 \\ h(z) &= \int_0^\infty \frac{x_2}{(1-z)^2} f_{x_1}\left(\frac{zx_2}{1-z}\right) f(x_2) dx_2 \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)(1-z)^{\alpha_1+1}} \int_0^\infty x_2^{\alpha_1+\alpha_2-1} e^{-\lambda \frac{x_2}{1-z}} dx_2 \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)(1-z)^{\alpha_1+1}} \int_0^\infty x_2^{\alpha_1+\alpha_2-1} e^{-\lambda \frac{x_2}{1-z}} dx_2 \end{aligned} \quad (97)$$

$u = \frac{\lambda x_2}{1-z}; \quad x_2 = \frac{u(1-z)}{\lambda}; \quad dx_2 = \frac{(1-z)}{\lambda} du$
--

$$\begin{aligned} h(z) &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)(1-z)^{\alpha_1+1}} \int_0^\infty \left(\frac{u(1-z)}{\lambda}\right)^{\alpha_1+\alpha_2-1} e^{-u} \left(\frac{1-z}{\lambda}\right) du \\ &= \frac{\lambda^{\alpha_1+\alpha_2} (1-z)^{\alpha_1+\alpha_1}}{\lambda^{\alpha_1+\alpha_2} \Gamma(\alpha_1)\Gamma(\alpha_2)(1-z)^{\alpha_1+1}} \int_0^\infty u^{\alpha_1+\alpha_2-1} e^{-u} du \\ &= \frac{z^{\alpha_1-1} (1-z)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty u^{\alpha_1+\alpha_2-1} e^{-u} du \\ &= \frac{z^{\alpha_1-1} (1-z)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \Gamma(\alpha_1 + \alpha_2) \\ &= \frac{z^{\alpha_1-1} (1-z)^{\alpha_2-1}}{B(\alpha_1 + \alpha_2)}; \alpha_1, \alpha_2 > 0, \text{ for } 0 < z < 1 \end{aligned} \quad (98)$$

This last result is clearly the beta distribution found in Equation 89.

QED

It follows from the previous results that the beta distribution is, in fact, a proportion of two gamma distributions!

Here is a comparison of the histogram of the random variable $Z = \frac{X}{X+Y}$, where $X \sim \mathcal{G}\text{amma}(2, 1)$ and $Y \sim \mathcal{G}\text{amma}(5, 1)$, with the plot of the probability density function of the beta distribution with parameters $\mathcal{B}\text{eta}(2, 5)$:

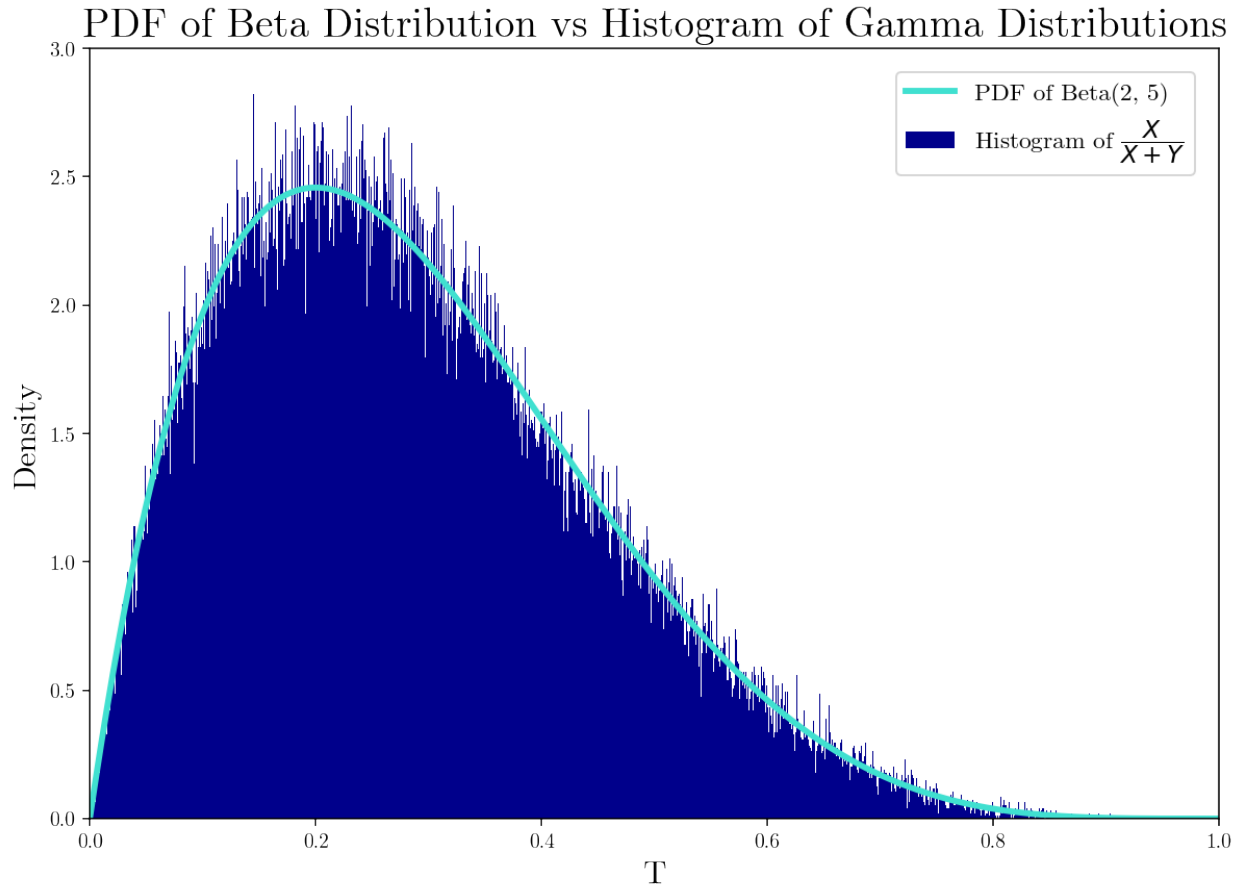


FIGURE 10: Comparison of the histogram of a quotient of gamma random variables with the probability density function of a beta distribution.

4.3 The Chi-Squared Distribution

Then, a gamma distribution with parameters $\alpha = n/2$ and $\lambda = 1/2$ is called the chi-squared distribution with n degrees of freedom such that

$$X \sim \mathcal{G}\text{amma}(n/2, 1/2) \equiv \chi_n^2 \quad (99)$$

Proof. First, to show that $\mathcal{G}\text{amma}(n/2, 1/2) \equiv \chi_n^2$, let's prove that $\mathcal{G}\text{amma}(1/2, 1/2) \equiv \chi_1^2$, a chi-squared distribution with one (1) degree of freedom, using a change of variables $Y = Z^2$ with $Z \sim \mathcal{N}(0, 1)$.

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(Z^2 \leq y) \\ &= P(-\sqrt{y} \leq z \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ g(y) &= \frac{f(\sqrt{y})}{2\sqrt{y}} + \frac{f(-\sqrt{y})}{2\sqrt{y}} \\ &= \frac{e^{-y/2}}{2\sqrt{2\pi y}} + \frac{e^{-y/2}}{2\sqrt{2\pi y}} \\ &= \frac{2e^{-y/2}}{2\sqrt{2\pi y}} \\ &= \frac{1}{\sqrt{2}\Gamma(\frac{1}{2})} y^{-1/2} e^{-y/2} \\ &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} y^{1/2-1} e^{-(1/2)y} \end{aligned} \quad (100)$$

$$\therefore Y \sim \mathcal{G}\text{amma}(1/2, 1/2) \equiv \chi_1^2 \quad (101)$$

One may generalize this result by using Z_1, Z_2, \dots, Z_n as i.i.d. random variables following $\mathcal{N}(0, 1)$ to find the distribution of

$$Y = \sum_{i=1}^n Z_i^2 \quad (102)$$

Hence, using the moment generating function,

$$\begin{aligned}
 M_{\sum_{i=1}^n Z_i^2}(t) &= \prod_{i=1}^n M_{Z_i^2}(t) \\
 &= M(t)^n \\
 &= \left(\frac{1/2}{1/2 - t} \right)^{n/2}
 \end{aligned} \tag{103}$$

This last line is equal to the moment generating function of a random variable following the distribution $\mathcal{Gamma}(n/2, 1/2)$. Hence,

$$\therefore Y \sim \mathcal{Gamma}(n/2, 1/2) \equiv \chi_n^2 \tag{104}$$

QED

Using the previously proven result, the density function of the chi-squared distribution may be determined by substituting the parameters $\alpha = n/2$ and $\lambda = 1/2$ into the density function of the gamma distribution:

$$\begin{aligned}
 f(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \\
 &= \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} \\
 &= \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2}
 \end{aligned} \tag{105}$$

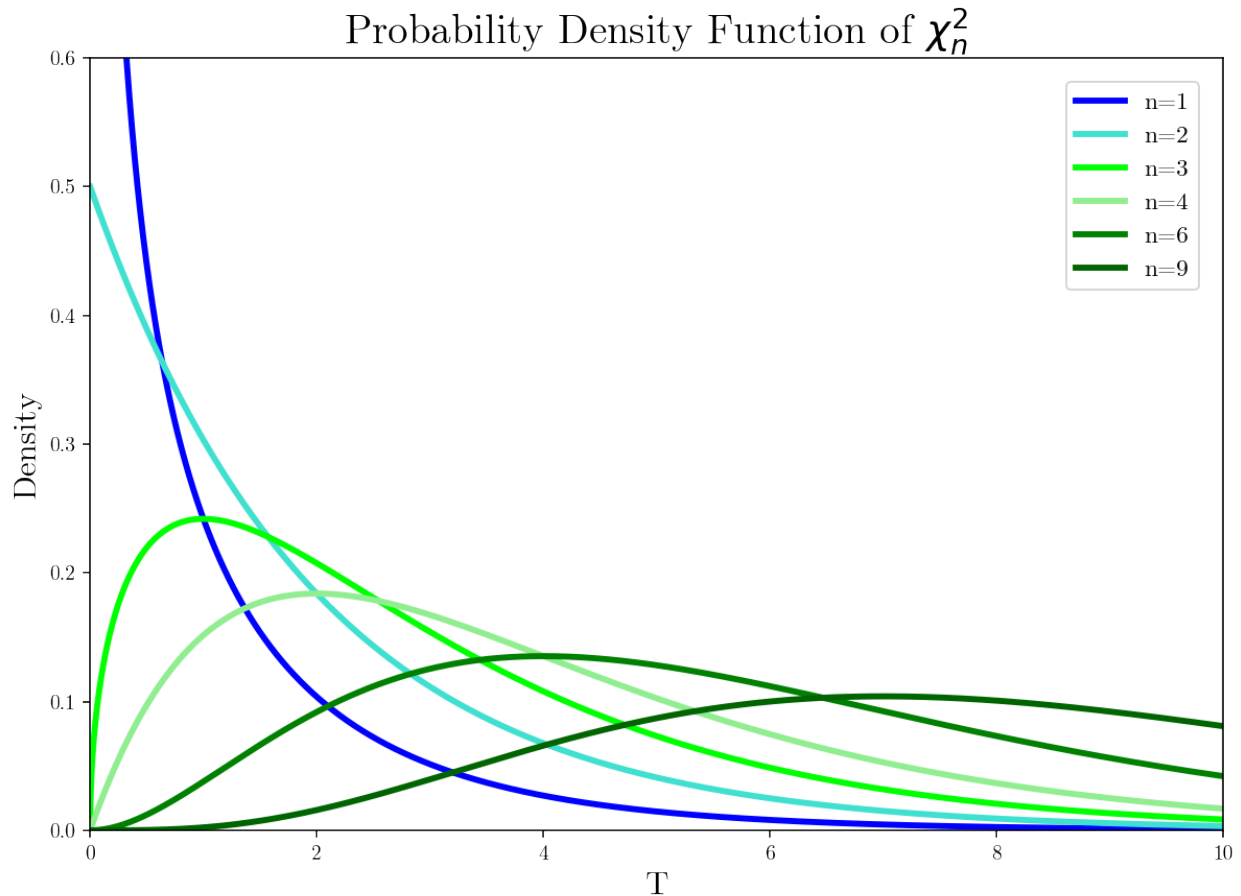


FIGURE 11: Plot of the probability distribution function of a χ_n^2 distribution for various values of n .

The chi-square distribution is often used to describe the distribution of a sum of squared random variables, to test the fit of a distribution of data, or to determine whether data series are independent or dependent.

4.4 The Normal Distribution

As was shown previously in subsection 4.3, the normal distribution is related to the gamma distribution as it may be used to prove that the latter is related to the chi-squared distribution. That is:

$$Z \sim \mathcal{N}(\text{ormal}(0, 1) \Rightarrow Z^2 \sim \mathcal{G}\text{amma}(1/2, 1/2) \equiv \chi_1^2 \quad (106)$$

Furthermore, since the gamma distribution $\mathcal{G}\text{amma}(\alpha, \lambda)$ may be viewed as a sum of α exponential distributions, then by the central limit theorem, the density function of the gamma distribution will be extremely similar to that of a normal distribution when α approaches ∞ :

$$\mathcal{G}\text{amma}(\alpha, \lambda) \approx \mathcal{N}(\alpha/\lambda, \alpha/\lambda^2) \text{ for } \alpha \rightarrow \infty \quad (107)$$

The central limit theorem will not be proven as it is quite intuitive; the more data samples one has, the more centred around the expected value the probability density function graph will be. That being said, here is a graph representing this concept by comparing the probability density function graph of a $\mathcal{G}\text{amma}(60, 1)$ to that of a $\mathcal{N}(60, 60)$. Because $\mathcal{G}\text{amma}(\alpha, \lambda)$ is in essence a sum of α i.i.d. random variables following a distribution $\mathcal{E}\text{xp}(\lambda)$, then it makes sense that having a distribution $\mathcal{G}\text{amma}(60, 1)$ is like having 60 data samples following a $\mathcal{E}\text{xp}(1)$ distribution, which fits the central limit theorem.

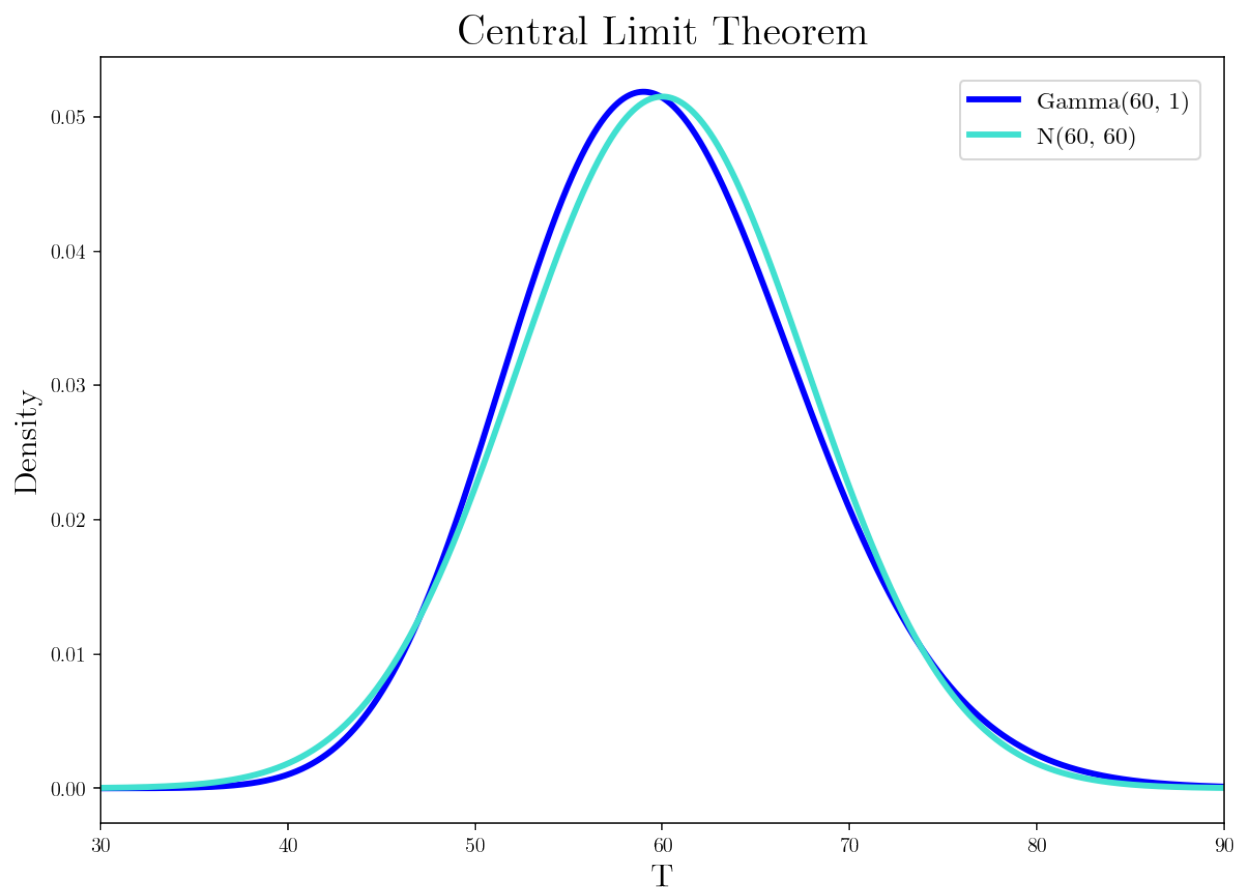


FIGURE 12: Graph showcasing the central limit theorem by comparing the probability density function plot of a $\mathcal{G}\text{amma}(60, 1)$ distribution to that of $\mathcal{N}(60, 60)$ distribution.

4.5 The Erlang Distribution

In real-world applications, it is often useful to limit the number of occurrences to a positive integer (it may not be pertinent to determine, for example, the waiting time until 2.665 events take place). In this context, the special name “erlang distribution” has been attributed to the specific gamma distribution in which the shape parameter α only takes positive integer values and where the rate parameter λ is a positive real number.

$$f_{X_\alpha}(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \alpha \in \mathbb{N}; x \geq 0 \\ 0 & x < 0 \end{cases} \quad (108)$$

This distribution takes its name from A. K. Erlang, the man who initially popularized its use in order to examine the number of telephone calls which can be made at the same time to the operators of the switching stations, in the field of phone traffic engineering. Indeed, the Erlang distribution’s original purpose was to measure the time between incoming calls, a statistic which can be used along with the expected duration of incoming calls to analyze phone traffic loads. His work has since been expanded to consider queuing systems in general, as well as the load on web servers, inter-purchase times and cancer incidence (see section 6).^[14]

4.6 The Wishart Distribution

The Wishart distribution, named in honour of the Scottish mathematician John Wishart who first formulated it in 1928, is a generalization to multiple dimensions of the gamma distribution: it is a family of distributions defined over symmetric positive matrices. Namely, if X_1, \dots, X_n are independent and form a $p \times n$ matrix $X = (X_1, \dots, X_n)$, then the distribution of a $p \times p$ random matrix $M = XX' = \sum_{i=1}^n X_i X_i'$ is said to have a Wishart distribution with n degrees of freedom.

If $n \geq p$, such a matrix is denoted $\mathbf{M} \sim W_p(n, \Sigma)$ and has the following probability density function, where $\Gamma_p(\alpha)$ is the multivariate gamma function and where Σ is a fixed symmetric positive matrix of size $p \times p$.

$$f(\mathbf{M}) = \frac{1}{2^{np/2} \Gamma_p(\alpha) |\Sigma|^{n/2}} |\mathbf{M}|^{n-p-1/2} e^{-\frac{\text{trace}(\Sigma^{-1}\mathbf{M})}{2}} \quad (109)$$

Interesting similarities with the probability density function of the gamma distribution can clearly be noticed in the form of this equation. While this generalization of the gamma distribution to matrices is a rather intriguing topic that was important to mention, this paper will not explore the Wishart distribution in greater depth. ^[15]

5 Advanced Properties

5.1 Skewness

In statistics, the skewness of a distribution is defined as the degree of asymmetry of the shape of the probability distribution with respect to the normal distribution. Distributions can exhibit right (positive) or left (negative) skewness; the normal distribution is qualified as exhibiting zero skewness. In the case of the gamma distribution, the skewness is only dependent upon the shape parameter α and corresponds to:^{[16],[3]}

$$\text{Skewness} = \frac{2}{\sqrt{\alpha}} \quad (110)$$

5.2 Information Entropy

The entropy of a random variable is defined as the average level of uncertainty associated with the possible outcomes of its probability distribution. In the case of continuous probability distributions such as the gamma distribution, one usually refers to differential entropy (also called continuous entropy), defined as:^{[17],[3]}

$$h(X) = E[-\log(f(X))] \quad (111)$$

More specifically, the entropy of the gamma distribution corresponds to:

$$H(X) = \alpha + \ln(\lambda) + \ln(\Gamma(\alpha)) + (1 - \alpha)\psi(\alpha) \quad (112)$$

Following is the proof of this preceding equation:

Proof.

$$\begin{aligned}
 H(X) &= E[-\ln(p(X))] \\
 &= E[-\alpha \ln(\lambda) + \ln(\Gamma(\alpha)) - (\alpha - 1) \ln(X) + \lambda X] \\
 &= \alpha + \ln(\lambda) + \ln(\Gamma(\alpha)) + (1 - \alpha)\psi(\alpha)
 \end{aligned} \tag{113}$$

QED

5.3 Characteristic Function

In statistics, any random variable associated with a probability density function is said to have a characteristic function, which corresponds to the Fourier transformation of its probability density function. The characteristic function thus simply corresponds to an alternative way of describing a random variable: it fully determines the behaviour and properties of a given probability distribution to the same extent as the distribution function, but can provide different insights to the features of the random variable. In the case of the gamma distribution, the characteristic function is defined as:^[18]

$$\text{CF} = \left(1 - \frac{it}{\lambda}\right)^{-\alpha} \tag{114}$$

5.4 Methods of Moments

The "methods of moments" of a probability distribution are methods used to estimate population parameters, that is, to describe various aspects of a given group or population. The methods of moments of the gamma distribution are defined as:^{[19], [20]}

$$\begin{aligned}
 \alpha &= \frac{E(X)^2}{\text{Var}(X)} \\
 \lambda &= \frac{E(X)}{\text{Var}(X)}
 \end{aligned} \tag{115}$$

5.4.1 Higher Moments

To describe or estimate further shape parameters of a probability distribution, the higher-order moments of a function can be used. Moments beyond 4th order are often increasingly hard to estimate, and larger sample sizes are often required to obtain better quality of estimation.^[21]

The n th raw moment is given by:

$$E[X^n] = \frac{\Gamma(n + \alpha)}{\lambda^n \Gamma(\alpha)} \quad (116)$$

5.5 Excess kurtosis

Kurtosis, the fourth standardized moment, is a characteristic used to describe the shape, or the “tailedness” of a probability distribution. The normal distribution is used as the standard, displaying an excess kurtosis of zero with which one can compare the excess kurtosis of other distributions.^[22]

The exponential distribution, as a special case of the gamma distribution, has a positive excess kurtosis and is hence called leptokurtic. This, in short, indicates that it exhibits fatter tails than the normal distribution. Following is the equation characterizing the excess kurtosis of the gamma distribution:^[22]

$$\text{Kurtosis} = \frac{6}{\alpha} \quad (117)$$

6 General Applications and Family

The gamma distribution is used in many fields, since it is directly related to the Erlang, normal and exponential distributions whose contributions extend to several disciplines. Here, we will present some of those applications.

6.1 Insurance Companies

First, the gamma distribution is of great use in the field of insurance services, given its direct relation to the exponential distribution. For example, an analyst could use this distribution to specify the amount of time a product lasts if one uses it at a constant average rate, thus modelling how reliable it is (and how much insurance should be charged for it). Then, the size of loan defaults and the cost of insurance claims are also often modelled according to a gamma distribution.^[23]

6.2 Natural Events Prediction

Then, the gamma distribution can also be used to model the amount of rainfall accumulated in a given reservoir. Indeed, this distribution fits positive data, represents rainfall distribution well and its two parameters give it sufficient flexibility to fit various climates.^[24]

6.3 Customer Satisfaction

6.3.1 Service Time

Service time can also be modelled using the gamma distribution. For example, if one is the sole person in line waiting for a meal, the waiting time until one receives the long-awaited food can be modelled using an exponential distribution. Using the same principle, the Erlang distribution can allow one to determine the total length of a process, that is, the duration of a sequence of independent events. For instance, if a large number of people are waiting

in line to be served, the distribution of each of their individual waiting times (the sum of several independent exponentially distributed variables) will correspond to the time it takes for the employee to serve everyone in it. Therefore, the gamma distribution lies at the heart of what is called queuing theory: the mathematical study of the congestion of waiting lines. Since waiting lines are found in countless places such as banks, restaurants and hospitals, as well as on web servers or multistep manufacturing and distribution processes, the gamma distribution provides very useful applications in everyday life.

6.3.2 Call Centers

One of the most famous of those applications concerns phone queuing, on which A. K. Erlang famously worked. The Erlang distribution has indeed been developed in the objective to model the time between incoming calls at a call centre, along with the expected number of calls, thus allowing call centres to know what their staffing capacity should be depending on the time of day.

6.3.3 Retail

Beyond waiting lines, the Erlang distribution is often used by retailers to model the frequency of inter-purchase times by consumers, which gives them an idea of how often a given consumer is expected to purchase a product from them and helps them control inventory and staffing.

6.4 Oncology

Finally, in the field of oncology, the age distribution of cancer incidence also follows a gamma distribution. Although the factors underlying cancer development are not yet fully understood, it has been hypothesized that cancers arise after several successive “driver events”, that is, after some number of mutations occurs in a cell. Analyses of cancer statistics suggest that the incidence of the most prevalent cancer types with respect to the patients’ age closely follows the gamma probability distribution or, more specifically, the Erlang distribu-

tion. This may be due to the fact that, more broadly, the Erlang distribution can be used to model the cell cycle time distribution.^[25] The age distribution of cancer can be modelled using the following equation:

$$Y = A \cdot x^{\alpha-1} \frac{e^{-\frac{x}{\beta}}}{\beta^{\alpha}} \cdot \Gamma(k) \quad (118)$$

In this case, the shape parameter α predicts the number of carcinogenic driver events, whereas the scale parameter β predicts the average time between those events for each cancer type. Using an additional amplitude parameter A , the maximal population susceptibility to a given type of cancer can even be predicted.^[25] Given that experimental research on cancer development is crucial for the lives of many people, numerical references such as that provided by the gamma distribution are of paramount importance in our society. The gamma distribution can save lives, if it is used wisely!

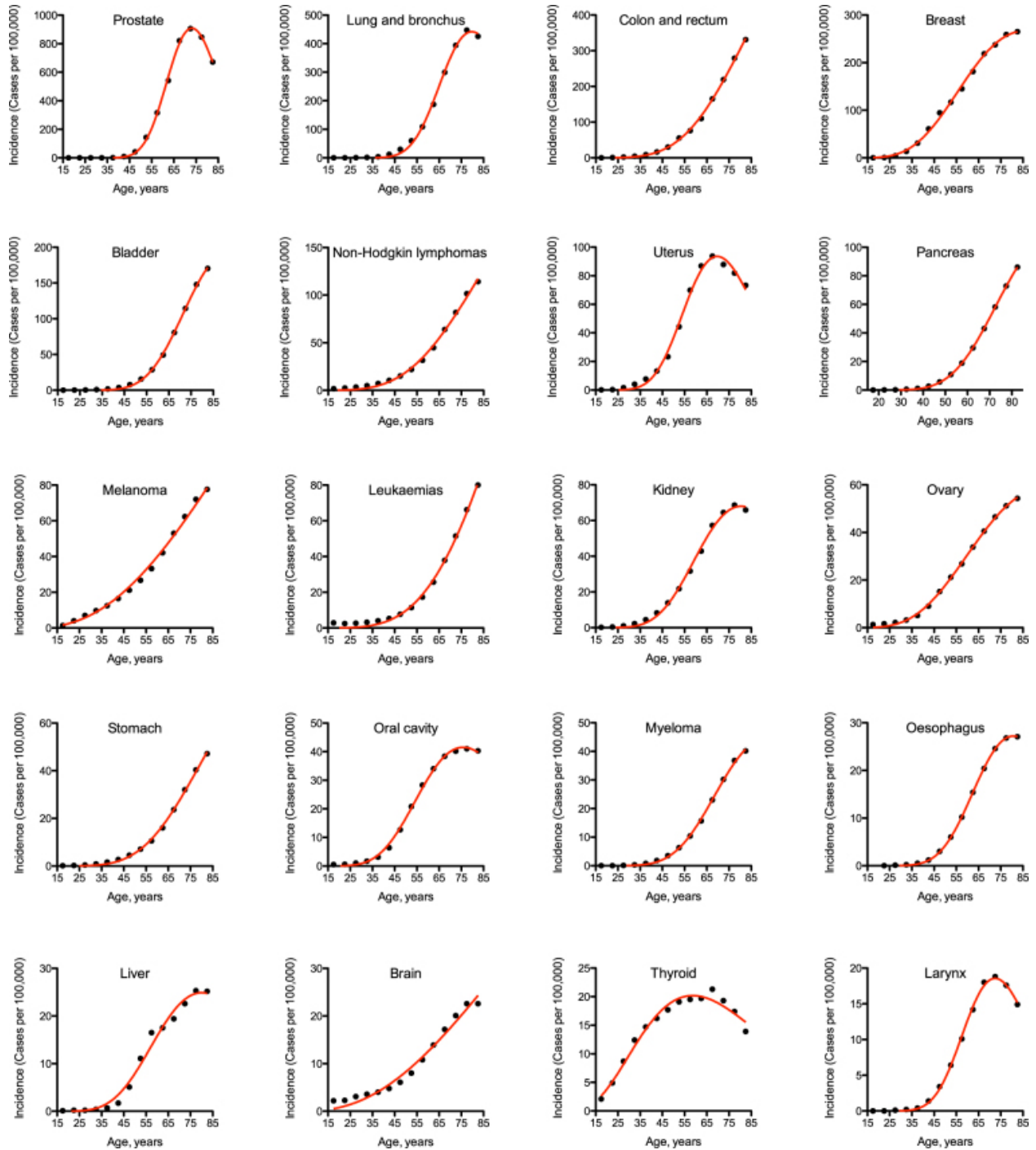


FIGURE 13: "The Erlang distribution approximates cancer incidence by age for 20 most prevalent cancer types. Dots indicate actual data for 5-year age intervals, curves indicate the PDF of the Erlang distribution fitted to the data [...]. The middle age of each age group is plotted. Cancer types are arranged in the order of decreasing incidence".^[25]

A Graphs - Python Code

A.1 Probability Density Function

Here is the code used to generate the probability density function graph of the gamma distribution in subsection 1.2:

```
0 import math
1 import numpy as np          # pip install numpy
2 import matplotlib.pyplot as plt  # pip install matplotlib
3
4 # Settings for matplotlib figure size, font and dpi
5 plt.rcParams["figure.figsize"] = [10, 7]
6 plt.rcParams["font.family"] = "CMU Serif"
7 plt.rcParams['figure.dpi'] = 140
8
9
10 def gamma_pdf(alpha_=4, lambda_=3):
11     x = np.linspace(0, 5, 100000)
12     # PDF of Gamma(alpha_, lambda_)
13     y1 = sp.stats.gamma.pdf(x, alpha_, scale=1/lambda_)
14
15     # Generating graph using matplotlib
16     plt.title("Probability Density Function of the Gamma Distribution", fontsize=20)
17     plt.xlabel("T", fontsize=16)
18     plt.ylabel("Density", fontsize=16)
19     plt.plot(x, y1, label=f"Gamma({alpha_}, {lambda_})", linewidth=3, color="blue")
20     plt.legend(bbox_to_anchor=(1, 1), loc="upper right", borderaxespad=1, fontsize=12)
21     plt.ylim([0, 0.7])
22     plt.xlim([0, 5])
23     plt.show()
24
25 gamma_pdf(4, 3)
```

A.2 Cumulative Distribution Function

Here is the code used to generate the cumulative distribution function graph of the gamma distribution in subsection 1.3:

```
0 import math
1 import itertools
2 import numpy as np          # pip install numpy
3 import matplotlib.pyplot as plt  # pip install matplotlib
4
5 # Settings for matplotlib figure size, font and dpi
6 plt.rcParams["figure.figsize"] = [10, 7]
7 plt.rcParams["font.family"] = "CMU Serif"
8 plt.rcParams['figure.dpi'] = 140
9
10
11 def gamma_cdf(alpha_=4, lambda_=3):
12     x = np.linspace(0, 4, 1000000)
13     # PDF of Gamma(alpha_, lambda_)
14     y1 = sp.stats.gamma.cdf(x, alpha_, scale=1/lambda_)
15
16     # Generating graph using matplotlib
17     plt.title("Cumulative Distribution Function of the Gamma Distribution", fontsize=20)
18     plt.xlabel("T", fontsize=16)
19     plt.ylabel("Probability", fontsize=16)
20     plt.plot(x, y1, label=f"Gamma({alpha_}, {lambda_})", linewidth=3, color="blue")
21     plt.legend(bbox_to_anchor=(1, 1), loc="upper right", borderaxespad=1, fontsize=12)
22     plt.ylim([0, 1])
23     plt.xlim([0, 4])
24     plt.show()
25
26 gamma_cdf()
```

A.3 Central Limit Theorem

Here is the code used to generate the central limit theorem graph in subsection 4.4:

```

0 import math
1 import numpy as np          # pip install numpy
2 import matplotlib.pyplot as plt # pip install matplotlib
3 import scipy as sp          # pip install scipy
4 from scipy import stats      # Used to generate gamma pdf array
5
6 # Settings for matplotlib figure size, font and dpi
7 plt.rcParams["figure.figsize"] = [10, 7]
8 plt.rcParams["font.family"] = "CMU Serif"
9 plt.rcParams['figure.dpi'] = 140
10
11 def central_limit_theorem(alpha_=60, lambda_=1, expected_=60, variance_=60):
12     x = np.linspace(30, 90, 100000)
13     # PDF of Gamma(alpha_, lambda_)
14     y1 = sp.stats.gamma.pdf(x, alpha_, scale=1/lambda_)
15     # PDF of N(expected_, variance_)
16     y2 = sp.stats.norm.pdf(x, expected_, math.sqrt(variance_))
17
18     # Generating graph using matplotlib
19     plt.title("Central Limit Theorem", fontsize=20)
20     plt.xlabel("T", fontsize=16)
21     plt.ylabel("Density", fontsize=16)
22     plt.plot(x, y1, label=f"Gamma({alpha_}, {lambda_})", linewidth=3, color="blue")
23     plt.plot(x, y2, label=f"N{expected_, variance_}", linewidth=3, color="turquoise")
24     plt.legend(bbox_to_anchor=(1, 1), loc="upper right", borderaxespad=1, fontsize=12)
25     plt.xlim([30, 90])
26     plt.show()
27
28 central_limit_theorem(60, 1, 60, 60)

```

B Simulations - Python Code

B.1 Simulation 1: $X + Y$

Comparison of the graph of $\mathcal{Gamma}(\alpha + \beta, \lambda)$ and the histogram of $X + Y$ with $X \sim \mathcal{Gamma}(\alpha, \lambda)$ and $Y \sim \mathcal{Gamma}(\beta, \lambda)$. Featured in ??.

```

0 import numpy as np                # pip install numpy
1 import matplotlib.pyplot as plt   # pip install matplotlib
2 import scipy as sp                # pip install scipy
3 from scipy import stats           # Used to generate gamma pdf array
4
5 # Settings for matplotlib figure size, font and dpi
6 plt.rcParams["figure.figsize"] = [10, 7]
7 plt.rcParams["font.family"] = "CMU Serif"
8 plt.rcParams['figure.dpi'] = 140
9
10
11 def simulation_sum(alpha_=1, beta_=3, lambda_=3):
12     # x and y are vectors of 1000000 i.i.d. Gamma(alpha_, lambda_) r.v.
13     x1 = np.random.gamma(alpha_, 1/lambda_, 1000000)
14     y1 = np.random.gamma(beta_, 1/lambda_, 1000000)
15     n = np.linspace(0, 5, 1000000)
16     # PDF of Gamma(alpha_+ beta_, lambda_)
17     y2 = sp.stats.gamma.pdf(n, (alpha_+beta_), scale=1/lambda_)
18
19     # r.v. of interest
20     u = x1 + y1
21
22     plt.title("PDF of Gamma Distribution vs Histogram of Gamma Distributions", fontsize=20)
23     plt.xlabel("T", fontsize=16)
24     plt.ylabel("Density", fontsize=16)
25     plt.plot(n, y2, label=f"PDF of Gamma({alpha_}+{beta_}, {lambda_})", linewidth=3,
26              color="turquoise")
27     plt.hist(u, label="Histogram of (X+Y)", bins=10000, density=True, color="darkblue")
28     plt.legend(bbox_to_anchor=(1, 1), loc="upper right", borderaxespad=1, fontsize=12)
29     plt.ylim([0, 0.8])
30     plt.xlim([0, 5])
31     plt.show()
32 simulation_sum(1, 3, 3)

```

B.2 Simulation 2: $\frac{X}{X+Y}$

Comparison of the graph of $\text{Beta}(\alpha, \beta)$ and the histogram of $\frac{X}{X+Y}$ with $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$. Featured in Figure 10.

```

0 import numpy as np                # pip install numpy
1 import matplotlib.pyplot as plt   # pip install matplotlib
2 import scipy as sp                # pip install scipy
3 from scipy import stats           # Used to generate gamma pdf array
4
5 # Settings for matplotlib figure size, font and dpi
6 plt.rcParams["figure.figsize"] = [10, 7]
7 plt.rcParams["font.family"] = "CMU Serif"
8 plt.rcParams['figure.dpi'] = 140
9
10
11 def simulation_quotient(alpha_=2, beta_=5, lambda_=1):
12     # x and y are vectors of 1000000 i.i.d. Gamma(alpha_, lambda_) r.v.
13     x1 = np.random.gamma(alpha_, 1/lambda_, 1000000)
14     y1 = np.random.gamma(beta_, 1/lambda_, 1000000)
15     n = np.linspace(0, 1, 10000)
16     # PDF of Beta(alpha_, beta_)
17     y2 = sp.stats.beta.pdf(n, alpha_, beta_)
18
19     # r.v. of interest
20     v = x1/(x1 + y1)
21
22     # Generating graph using matplotlib
23     plt.title("PDF of Beta Distribution vs Histogram of Gamma Distributions", fontsize=20)
24     plt.xlabel("T", fontsize=16)
25     plt.ylabel("Density", fontsize=16)
26     plt.plot(n, y2, label=f"PDF of Beta({alpha_}, {beta_})", linewidth=3, color="turquoise")
27     plt.hist(v, label=r"Histogram of $\dfrac{X}{X+Y}$", bins=10000, density=True,
28             color="darkblue")
29     plt.legend(bbox_to_anchor=(1, 1), loc="upper right", borderaxespad=1, fontsize=12)
30     plt.ylim([0, 3])
31     plt.xlim([0, 1])
32     plt.show()
33 simulation_quotient(2, 5, 1)

```

List of Figures

1	Plot of the probability density function of a gamma distribution.	4
2	Plot of the cumulative distribution function of a gamma distribution.	7
3	Visual representation of quantiles in a probability density distribution. Provided by GeeksforGeeks ^[8]	16
4	Plot of the probability density function of three distinct gamma distributions with different modes.	17
5	Plot of the probability density function of four distinct gamma distributions with different inflection points.	19
6	Plot of the probability density function of a gamma distribution with parameters $\alpha = 5$ and $\lambda = 1$	21
7	Plot of the function $\Gamma(\alpha)$ in red and the asymptotes of the function in dotted blue lines.	31
8	Comparison of the histogram of a sum of gamma random variables with the probability density function of a gamma distribution.	39
9	Plot of the probability distribution function of a beta distribution for various parameterizations.	40
10	Comparison of the histogram of a quotient of gamma random variables with the probability density function of a beta distribution.	45
11	Plot of the probability distribution function of a χ_n^2 distribution for various values of n	48
12	Graph showcasing the central limit theorem by comparing the probability density function plot of a $\mathcal{G}\text{amma}(60,1)$ distribution to that of $\mathcal{N}(60,60)$ distribution.	50

-
- 13 "The Erlang distribution approximates cancer incidence by age for 20 most prevalent cancer types. Dots indicate actual data for 5-year age intervals, curves indicate the PDF of the Erlang distribution fitted to the data [...]. The middle age of each age group is plotted. Cancer types are arranged in the order of decreasing incidence".^[25] 59

References

- [1] Vincent Carrier. *Probability and Statistics 201-BNM-05*. URL: <https://sites.google.com/view/vincentcarrier/probability-and-statistics-201-bnm-05> (visited on 04/13/2022).
- [2] William L. Hosch. *Gamma Distribution*. In: *Encyclopedia Britannica*. Britannica, 2017-09-15. URL: <https://www.britannica.com/science/gamma-distribution> (visited on 04/06/2022).
- [3] Wikipedia. *Gamma Distribution*. In: *Wikipedia*. 2022-02-23. URL: https://en.wikipedia.org/w/index.php?title=Gamma_distribution&oldid=1073512326 (visited on 03/02/2022).
- [4] Pascal Sebah and Xavier Gourdon. *Introduction to the Gamma Function*. 2002-02-04. URL: <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.527.8519>.
- [5] Eric W. Weisstein. *Incomplete Gamma Function*. URL: <https://mathworld.wolfram.com/> (visited on 04/22/2022).
- [6] Wikipedia. *Moment-Generating Function*. In: *Wikipedia*. 2022-04-18. URL: https://en.wikipedia.org/w/index.php?title=Moment-generating_function&oldid=1083357569 (visited on 04/22/2022).
- [7] Wikipedia. *Variance*. In: *Wikipedia*. 2022-04-13. URL: <https://en.wikipedia.org/w/index.php?title=Variance&oldid=1082503093> (visited on 04/22/2022).
- [8] GeeksforGeeks. *Quantile Quantile Plots*. GeeksforGeeks. 2021-01-19. URL: <https://www.geeksforgeeks.org/quantile-quantile-plots/> (visited on 04/29/2022).
- [9] Daniel Myers. *CS 547 Lecture 9: Conditional Probabilities and the Memoryless Property*.
- [10] ProbabilityCourse. *Exponential Distribution*. Probability Course. URL: https://www.probabilitycourse.com/chapter4/4_2_2_exponential.php (visited on 04/29/2022).

-
- [11] Carolynne K. Ayienda. “Gamma and Related Distributions”. Thesis. University of Nairobi, 2013. URL: <http://erepository.uonbi.ac.ke/handle/11295/58770> (visited on 02/28/2022).
- [12] Aerin Kim. *Beta Distribution — Intuition, Examples, and Derivation*. Medium. 2020-07-30. URL: <https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af> (visited on 02/28/2022).
- [13] Matt DosSantos DiSorbo. “Beta and Gamma”. In: *Probability!* bookdown. URL: <https://bookdown.org/probability/beta/beta-and-gamma.html#gamma> (visited on 05/09/2022).
- [14] Zach. *What Is the Erlang Distribution?* Statology. 2020-12-20. URL: <https://www.statology.org/erlang-distribution/> (visited on 03/22/2022).
- [15] Wikipedia. *Wishart Distribution*. In: *Wikipedia*. 2022-04-13. URL: https://en.wikipedia.org/w/index.php?title=Wishart_distribution&oldid=1082546446 (visited on 04/22/2022).
- [16] James Chen. *Learn About Skewness*. Investopedia. 2021-04-30. URL: <https://www.investopedia.com/terms/s/skewness.asp> (visited on 04/25/2022).
- [17] Wikipedia. *Differential Entropy*. In: *Wikipedia*. 2022-02-16. URL: https://en.wikipedia.org/w/index.php?title=Differential_entropy&oldid=1072236352 (visited on 04/25/2022).
- [18] Wikipedia. *Characteristic Function (Probability Theory)*. In: *Wikipedia*. 2022-04-29. URL: [https://en.wikipedia.org/w/index.php?title=Characteristic_function_\(probability_theory\)&oldid=1085265595](https://en.wikipedia.org/w/index.php?title=Characteristic_function_(probability_theory)&oldid=1085265595) (visited on 04/29/2022).
- [19] Wikipedia. *Method of Moments (Statistics)*. In: *Wikipedia*. 2021-06-18. URL: [https://en.wikipedia.org/w/index.php?title=Method_of_moments_\(statistics\)&oldid=1029190833](https://en.wikipedia.org/w/index.php?title=Method_of_moments_(statistics)&oldid=1029190833) (visited on 04/29/2022).
-

-
- [20] Alane Lim. *What Is a Population Parameter?* ThoughtCo. 2019-05-14. URL: <https://www.thoughtco.com/population-parameter-4588247> (visited on 04/29/2022).
- [21] Wikipedia. *Moment (Mathematics)*. In: *Wikipedia*. 2022-04-02. URL: [https://en.wikipedia.org/w/index.php?title=Moment_\(mathematics\)&oldid=1080682562](https://en.wikipedia.org/w/index.php?title=Moment_(mathematics)&oldid=1080682562) (visited on 04/29/2022).
- [22] Wikipedia. *Kurtosis*. In: *Wikipedia*. 2022-04-18. URL: <https://en.wikipedia.org/w/index.php?title=Kurtosis&oldid=1083383594> (visited on 04/25/2022).
- [23] Ajay Tiwari. *Modeling Insurance Claim Severity*. The Startup. 2020-04-10. URL: <https://medium.com/swlh/modeling-insurance-claim-severity-b449ac426c23> (visited on 03/22/2022).
- [24] Gregory J. Husak, Joel Michaelsen, and Chris Funk. “Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications”. In: *International Journal of Climatology* 27.7 (2007), pp. 935–944. ISSN: 1097-0088. DOI: 10.1002/joc.1441. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.1441> (visited on 03/22/2022).
- [25] Aleksey V. Belikov. “The Number of Key Carcinogenic Events Can Be Predicted from Cancer Incidence”. In: *Scientific Reports* 7 (2017-09-22), p. 12170. ISSN: 2045-2322. DOI: 10.1038/s41598-017-12448-7. pmid: 28939880. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5610194/> (visited on 03/22/2022).