A brief introduction to the gamma distribution

Radoslav Harman

May 11, 2022

1 Introduction

I did not use any particular book or paper to prepare this classroom material; everything here is simple to derive from the basic theorems of calculus and probability. However, it is possible to find much published information on the gamma distribution; see, for instance, the book by Balakrishnan and Nevzorov (ISBN: 978-0-471-42798-8) for an introduction and the book by Johnson, Kotz and Balakrishnan (ISBN: 978-0-471-58495-7) for details.

For the definition of the gamma distribution we will need the gamma function $\Gamma : (0,\infty) \to \mathbb{R}$:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

It is simple to show that $\Gamma(z) = (z-1)!$ for all $z \in \mathbb{N}$. That is, the gamma function can be viewed as an extension of the function $z \to (z-1)!$, $z \in \mathbb{N}$, to $(0,\infty)$[1]. Some properties of the gamma function used in probability and statistics:

- $\Gamma(z) = \prod_{j=1}^{k}(z-j)\Gamma(z-k)$ for $z > k \in \mathbb{N}$, in particular
- $\Gamma(z) = (z-1)\Gamma(z-1)$ for $z > 1$;
- $\Gamma(z + (1/2)) = (2z-1)!!\sqrt{\pi}/2^z$ for $z \in \mathbb{N}$, in particular
- $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(3/2) = \sqrt{\pi}/2$;
- $\Gamma(z+1) \approx \sqrt{2\pi z}(z/e)^z$ for large z, called Stirling's approximation;
- $\frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)} = \int_0^1 x^{z_1-1}(1-x)^{z_2-1}\mathrm{d}x$ for any $z_1, z_2 > 0$.[2]

Also, since probability and statistics is closely related to geometry, we often use the formula for the volume of the unit d-dimensional ball:

- $V_d = \pi^{d/2}/\Gamma(1 + (d/2))$.

[1] The definition can be further extended to all *complex* arguments z (except $0, -1, -2, \ldots$), but in this text we only need the gamma function defined on $(0,\infty)$.

[2] This provides the relation of the gamma function to the so-called beta function.

2 Definition

The random variable X is said to follow the gamma distribution with shape $k > 0$ and scale $\theta > 0$,[3] denoted $X \sim Gam(k, \theta)$, if X is a continuous random variable with density

$$f(x) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta}, \text{ for } x > 0,$$

and $f(x) = 0$ for $x \leq 0$.

The most important special cases of $Gam(k, \theta)$ are

- the exponential distribution[4] $Gam(1, \theta)$,
- the Erlang distribution $Gam(n, \theta)$, $n \in \mathbb{N}$, and
- the χ_n^2 distribution $Gam(n/2, 2)$, $n \in \mathbb{N}$.

The gamma distribution is used to model phenomena that produce random positive real numbers. This can be, for instance, a random time (such as the lifetime of a component of a larger system), monetary costs (such as insurance claims), size (such as the size of a randomly selected tissue cell), particle speed, magnitude of rainfalls and so on.

3 The shape and the scale

The parameter k strongly influences the "shape" of the density:

- for $k < 1$ the density is unbounded with $\lim_{x \to 0_+} f(x) = +\infty$,
- for $k = 1$ it is an exponential curve with $\lim_{x \to 0_+} f(x) = 1/\theta$, and
- for $k > 1$ the density is a "skewed hump" with $\lim_{x \to 0_+} f(x) = 0$.

Increasing the parameter θ makes the density "stretched" and "squat", but does not influence the overall shape. Also, it is possible to show that the maximum of f, i.e., the mode, is[5] $\theta(k-1)$.

Suppose that we measure some $Gam(k, \theta)$-distributed random quantity. What is the distribution of the same quantity measured in different units? For instance, let $X \sim Gam(k, \theta)$ be an inter-arrival time between two consecutive

[3] In this text we will always assume that both the shape parameter and the scale parameter are positive, without stating it explicitly.

[4] Therefore, this text also covers the exponential distribution. For instance, we will obtain the higher-order moments of the exponential distribution as a special case of the gamma distribution. However, some interesting properties of the exponential distribution, such as its "memorylessness", do not generalize to the geometric distribution.

[5] If the mode exists, that is, if $k > 1$. Incidentally, the median of $Gam(k, \theta)$ cannot be expressed as a simple function of k and θ. For the special case of the exponential distribution $Gam(1, \theta)$, the median is $\sqrt{2}\theta$.

customers, measured in minutes. What is the distribution of the random variable $60X$, which is the same measurement expressed in seconds? It turns out that it is again gamma-distributed, but with a change in the scale parameter:

Theorem 1. Let $X \sim Gam(k, \theta)$ and let $c > 0$. Then $cX \sim Gam(k, c\theta)$.

Technically, the previous theorem is a direct consequence of the well-known formula for the density f_{cX} of a linearly transformed continuous random variable: $f_{cX}(x) = \frac{1}{c} f_X(x/c)$, where $c > 0$ and f_X is the density of X.

4 Moments

Theorem 2. The moment generating function[6] of $X \sim Gam(k, \theta)$ is

$$M_X(t) = \frac{1}{(1-\theta t)^k} \text{ for } t < 1/\theta.$$

The proof is straightforward: $M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta} \mathrm{d}x = \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^{k-1} e^{-(1/\theta - t)x/\theta} \mathrm{d}x =^* \frac{1}{\theta^k \Gamma(k)} \int_0^\infty \frac{y^{k-1}}{(1/\theta - t)^{k-1}} e^{-y} \frac{1}{1/\theta - t} \mathrm{d}y =^{**} \frac{1}{(1-\theta t)^k}$, where in the equality $*$ we used the substitution $y = (1/\theta - t)x$ and in the inequality $**$ we used some basic algebra and the definition of $\Gamma(k)$.

Based on Theorem 2 we could calculate the absolute moments of $X \sim Gam(k, \theta)$. However, the moments are simple to obtain in the classical way: First, let $\tilde{X} \sim Gam(k, 1)$ and $n \in \mathbb{N}$. Then $E(\tilde{X}^n) = \int_0^\infty x^n \frac{1}{\Gamma(k)} x^{k-1} e^{-x} \mathrm{d}x = \Gamma(n+k)/\Gamma(k)$, where we used the definition of $\Gamma(n+k)$. If $X \sim Gam(k, \theta)$ for some $\theta > 0$, then $X = \theta \tilde{X}$ for $\tilde{X} \sim Gam(k, 1)$ (see Theorem 1), therefore $E(X^n) = E((\theta \tilde{X})^n) = \theta^n E(\tilde{X}^n) = \theta^n \Gamma(n+k)/\Gamma(k)$. We obtained

Theorem 3. If $X \sim Gam(k, \theta)$ and $n \in \mathbb{N}$ then

$$\alpha_n = E(X^n) = \theta^n \frac{\Gamma(n+k)}{\Gamma(k)}.$$

It is worth mentioning that we can use the basic properties of Γ to simplify the ratio

$$\frac{\Gamma(n+k)}{\Gamma(k)} = (n+k-1)(n+k-2) \cdots k.$$

In particular, if $k = 1$, i.e. for an exponential random variable, we have $E(X^n) = \theta^n n!$. From the equations above we directly have

Theorem 4. If $X \sim Gam(k, \theta)$ then $\alpha_1 = EX = k\theta$ and $\beta_2 = DX = k\theta^2$.

[6] Note that for a continuous random variable, the moment generating function of X is in fact the so-called Laplace transform of the density of X.

We can obtain higher order central moments β_n, $n > 2$ from the general relations between the central and the absolute moments:

$$\beta_k = E((X - EX)^k) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (EX)^j \alpha_{k-j}.$$

After tiresome but straightforward calculations we could obtain

Theorem 5. The skewness of $X \sim Gam(k, \theta)$ is $\gamma_1 = 2/\sqrt{k}$ and the kurtosis is $\gamma_2 = 3 + 6/k$.

We see that the gamma distribution is positively skewed and leptokurtic (i.e., $\gamma_2 > 3$, where 3 is the kurtosis of any $Y \sim N(\mu, \sigma^2)$). For $k \to \infty$ the skewness diminishes, and the kurtosis converges to the one of the normal distribution. In fact, for increasing k the standardized gamma distribution converges to $N(0, 1)$.

5 The summation

Theorem 6. Let $X_1, \ldots, X_n$ be independent random variables, $X_i \sim Gam(k_i, \theta)$, $i = 1, \ldots, n$. Then $X = X_1 + \cdots + X_n \sim Gam(\sum_{i=1}^{n} k_i, \theta)$.

This theorem will be obvious once we prove that for two independent random variables $X_1 \sim Gam(k_1, \theta)$ and $X_2 \sim Gam(k_2, \theta)$ we have $Y := X_1 + X_2 \sim Gam(k_1 + k_2, \theta)$. But this is simple: for $y > 0$ we have

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y - x) \mathrm{d}x \\
&= \frac{e^{-y/\theta}}{\theta^{k_1+k_2}\Gamma(k_1)\Gamma(k_2)} \int_0^y x^{k_1-1}(y - x)^{k_2-1} \mathrm{d}x \\
&= \frac{e^{-y/\theta}}{\theta^{k_1+k_2}\Gamma(k_1)\Gamma(k_2)} y^{k_1+k_2-1} \int_0^1 z^{k_1-1}(1 - z)^{k_2-1} \mathrm{d}x \\
&= \frac{1}{\theta^{k_1+k_2}\Gamma(k_1 + k_2)} y^{k_1+k_2-1} e^{-y/\theta} \sim Gam(k_1 + k_2, \theta),
\end{aligned}$$

where the first equality is the well-known formula for the convolution of two continuous probability distributions,[7] the second one uses the form of the densities of X_1 and X_2,[8] the third one is based on the substitution $z = x/y$, and the fourth one follows from the relation of the gamma and the beta functions mentioned in the first section of this study material.

Note that Theorem 6 implies that the sum of $n \in \mathbb{N}$ independent random variables with the exponential distribution $Gam(1, \theta)$ is a random variable with

[7] That is, it is the formula for the density of the sum Y of two independent continuous random variables X_1 and X_2.

[8] Note also the change of limits in the integral, which is the consequence of the fact that both f_1 and f_2 are 0 on $(-\infty, 0)$.

the Erlang distribution $Gam(n, \theta)$. In addition, Theorems 6 and 1 imply that the artihmetic mean of n independent $Gam(k_i, \theta)$ variables is distributed according to $Gam(\sum_i k_i, \theta/n)$.

6 Limit cases

Since for $k \in \mathbb{N}$ the distribution of $X \sim Gam(k, \theta)$ is the same as the distribution of the sum of k independent $Gam(1, \theta)$ random variables, the central limit theorem implies that for $k \to \infty$ the standardized $Gam(k, \theta)$, i.e., the Erlang distribution, converges to $N(0, 1)$:

Theorem 7. Let $X_k \sim Gam(k, \theta)$, $k = 1, 2, \ldots$, and let $x \in \mathbb{R}$. Then

$$\lim_{k\to\infty} P\left[\frac{X_k - EX_k}{\sqrt{DX_k}} < x\right] = \Phi(x), \tag{1}$$

where Φ is the distribution function of $N(0, 1)$.

We could replace EX_k with $k\theta$ and DX_n with $k\theta^2$ in (1). Loosely put, for a large k the distribution $Gam(k, \theta)$ is approximately the $N(k\theta, k\theta^2)$ distribution.