The Gamma Distribution
Properties, Proofs and Applications

Etienne Collin, Marie Ouellet & Rania Yahyaoui

Probability and Statistics - 201-BNM-LW

Section 20108

Professor Vincent Carrier

St. Lawrence Cégep Champlain

Due on May 25, 2022 at 23:59

Table of Contents

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.

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 modeling 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 modeled according to a gamma distribution.

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 -shape and scale- give it sufficient flexibility to fit various climates.

Service time can also be modeled using the gamma distribution. For example, if one is waiting in line 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 multi-step manufacturing and distribution processes, the gamma distribution provides very useful applications in everyday life.

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 goal to model the time in between incoming calls at a call center, along with the expected number of calls, thus allowing call centers to know what their staffing capacity should be depending on the time of day.

Beyond waiting lines, the erlang distribution is often used by retailers to model the frequency of interpurchase 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.

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 distribution. This may be due to the fact that, more broadly, the erlang distribution can be used to model cell cycle time distribution.

$$Y = A \cdot x^{k-1} \frac{e^{\frac{-x}{b}}}{b^k} \cdot \Gamma(k) \tag{1}$$

In this case, the shape parameter α predicts the number of carcinogenic driver events, whereas the shape parameter λ 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. 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!

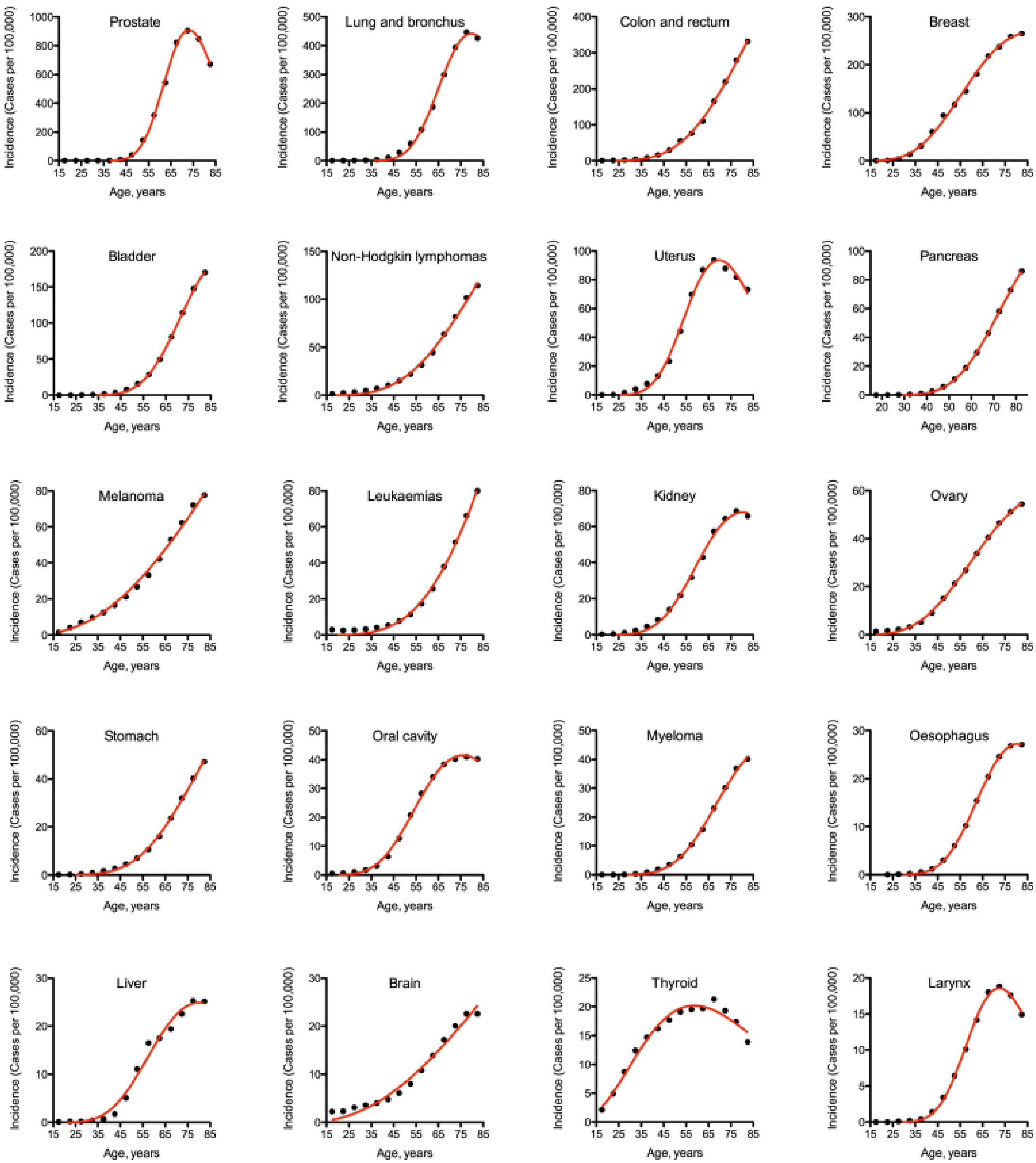

FIGURE 1: "The Erlang distribution approximates cancer incidence by age for 20 most prevalent cancer types. Dots indicate actual data for 5-year age intervals, curves indicate the PDF of the Erlang distribution fitted to the data [...]. The middle age of each age group is plotted. Cancer types are arranged in the order of decreasing incidence" [1] .

Introduction

The gamma distribution is part of the two-parameters family of continuous probability distributions. Indeed, it may be parameterized with two different parameterizations[2] :

Parameterization 1:

$$\text{Shape: } \alpha > 0 \qquad\qquad \text{Rate: } \lambda > 0 \tag{2}$$

Parameterization 2:

$$\text{Shape: } k > 0 \qquad\qquad \text{Scale: } \theta > 0 \tag{3}$$

Explain why there are two parameterizations and why the λ is sometimes replaced with a β.

That being said, in this document, as the two parameterizations only exist for the sake of convenience and are identical in their results, only the parameterization 1 presented in Equation 2 will be considered and used for proofs.

Relation to other distributions

The gamma distribution has many common parameterizations. Hence, in order to make working with them easier, they were given a specific name such as the chi-square distribution, the exponential distribution and more. Let us prove their relationship

Exponential Distribution

First, the exponential distribution is known to be a special case with parameters $\alpha = 1$ and λ such that:

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

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

$$\begin{aligned} f(x) &= \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \\ &= \frac{\lambda^1 x^0}{\Gamma(1)} e^{-\lambda x} \\ &= \lambda e^{-\lambda x} \end{aligned} \tag{4}$$

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

Chi-Squared Distribution

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

$$\mathcal{G}\text{ammma}({}^{n}/_{2}, {}^{1}/_{2}) \equiv \chi^2_n$$

Proof. First, to prove that $\mathcal{G}\text{ammma}({}^{n}/_{2}, {}^{1}/_{2}) \equiv \chi^2_n$, let's prove that $\mathcal{G}\text{ammma}({}^{1}/_{2}, {}^{1}/_{2}) \equiv \chi^2_1$, a chi-squared distribution with one (1) degree of freedom, using a change of variables with $Y = Z^2$ for $Y \sim \chi^2_1$.

$$\begin{aligned}
G(y) &= P(Y \leq y) \\
&= P(Z^2 \leq y) \\
&= P(-\sqrt{y} \leq z \leq \sqrt{y}) \\
&= \Phi(\sqrt{y}) - \Phi(-\sqrt{y})
\end{aligned}$$

$$\begin{aligned}
g(y) &= \frac{f(\sqrt{y})}{2\sqrt{y}} + \frac{f(-\sqrt{y})}{2\sqrt{y}} \\
&= \frac{e^{-y/2}}{2\sqrt{2\pi y}} + \frac{e^{-y/2}}{2\sqrt{2\pi y}} \\
&= \frac{\cancel{2}e^{-y/2}}{\cancel{2}\sqrt{2\pi y}} \\
&= \frac{}{\sqrt{2}\Gamma(\frac{1}{2})} y^{-1/2} e^{-y^{1/2}} \\
&= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} y^{1/2-1} e^{-(1/2)y}
\end{aligned} \tag{5}$$

$$\therefore Y \sim \mathcal{G}\text{ammma}({}^{1}/_{2}, {}^{1}/_{2}) \equiv \chi^2_1$$

One may generalize this result by using Z_i are i.i.d. $\mathcal{N}(0,1)$ to find the distribution of

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

Hence, using the moment generating function,

$$\begin{aligned} M_{\sum_{i=1}^{n}(Z_i^2)}(t) &= \Pi_{i=1}^{n}(M_{Z_i^2}(t)) \\ &= M(t)^n \\ &= \left(\frac{1/2}{1/2 - t}\right)^{n/2} \\ &\sim \mathcal{G}\text{amma}(n/2, 1/2) \end{aligned} \tag{6}$$

$$\therefore Y \sim \mathcal{G}\text{amma}(n/2, 1/2) \equiv \chi_n^2$$

QED

Beta Distribution

Normal Distribution

Wishart Distribution

Properties to prove

Support: $x \in (0, \infty)$

Probability density function:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \tag{7}$$

Cumulative distribution function:

$$F(x) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \lambda x) \tag{8}$$

Expected value, also known as the theoretical mean:

$$\mu = E(x) = \frac{\alpha}{\lambda} \tag{9}$$

There is no simple closed form equation for the median of a gamma distribution.

Mode:

$$\text{Mode} = \frac{(\alpha - 1)}{\lambda} \text{ for } \alpha \geq 1 \tag{10}$$

Variance:

$$\text{Var}(x) = \frac{\alpha}{\lambda^2} \tag{11}$$

Skewness:

$$\text{Skewness} = \frac{2}{\sqrt{\alpha}} \tag{12}$$

Excess kurtosis:

$$\text{Kurtosis} = \frac{6}{\alpha} \tag{13}$$

Entropy:

$$\text{Entropy} = \alpha + \ln \lambda + \ln \Gamma(\alpha) + (1 - \alpha)\psi(\alpha) \tag{14}$$

Moment generating function:

$$M(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha} \text{ for } t < \lambda \tag{15}$$

Characteristic function:

$$\text{CF} = \left(1 - \frac{it}{\lambda}\right)^{-\alpha} \tag{16}$$

Methods of moments:

$$\begin{aligned} \alpha &= \frac{E(X)^2}{\text{Var}(X)} \\ \lambda &= \frac{E(X)}{\text{Var}(X)} \end{aligned} \tag{17}$$

List of Figures

List of Tables

References

[1] Aleksey V. Belikov. "The Number of Key Carcinogenic Events Can Be Predicted from Cancer Incidence". In: *Sci Rep* 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).

[2] 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).