AIMS Mathematics, 10(10): 25033–25048.
DOI: 10.3934/math.20251109
Received: 12 June 2025
Revised: 18 September 2025
Accepted: 30 September 2025
Published: 31 October 2025

https://www.aimspress.com/journal/Math

*Research article*

# Generalized Chi distribution for non-integer degrees of freedom in modelling biological collectivities

**Juan Carlos Castro-Palacio[1], Pedro Fernández de Córdoba[2], Álvaro González-Cortés[2,*], J. M. Isidro[2], A. Noverques[3], Marcos Orellana-Panchame[1,4], Sarira Sahu[5] and Enrique A. Sánchez-Pérez[2]**

[1] Centro de Tecnologías Físicas, Universitat Politècnica de València, Camino de Vera, s/n, 46022 València, Spain

[2] Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain

[3] Institute for Industrial, Radiophysical and Environmental Safety (ISIRYM), Universitat Politècnica de València, Camino de Vera, s/n, 46022, València, Spain

[4] Departamento de Matemática, Universidad Nacional Autónoma de Honduras (UNAH), 21102 San Pedro Sula, Honduras

[5] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C.U., A. Postal 70-543, Mexico DF 04510, Mexico

* **Correspondence:** Email: agoncor@alumni.upv.es; Tel: 0034963877000; Fax: 00343877669.

**Abstract:** This article aims to expand the practical potential of the Chi and Chi-square distributions by extending their applicability to non-integer degrees of freedom, $k$. The study has two main objectives: first, to explore and leverage the relationship between the Chi, Chi-square, and Gamma distributions to interpret non-integer degrees of freedom; and second, to apply this framework to real survival-time data in the context of biological and health sciences. Three specific cases are analyzed: the germination times of *Pinus tropicalis* Morelet seeds cultivated in western Cuba; the interval between infection by virulent tuberculous bacilli and death from tuberculosis in guinea pigs; and the period from the onset of symptoms to death due to COVID-19 in patients in Mexico during 2020. A parameter estimation was performed by maximum likelihood (MLE), and the optimal value of $k$ was found to be non-integer in all cases.

**Keywords:** Chi distribution; Gamma distribution; non-integer degrees of freedom; survival time data; statistical modelin

## 1. Introduction

The Chi ($\chi$) and Chi-squared distributions are continuous probability distributions widely used in various scientific fields, such as econometrics [1], and have deep roots in classical goodness-of-fit tests (see for example the works of Gumbel [3] and Fisher [4]). Two particularly relevant cases are worth mentioning. A Chi distribution with $k = 2$ degrees of freedom leads to a Rayleigh distribution, which is commonly employed in meteorology to study two-dimensional wind speeds [5]. Additionally, when $k = 3$, a Maxwell-Boltzmann distribution is obtained, which is frequently used to describe the velocities of independent particles in an ideal gas in thermodynamic equilibrium [6].

In addition to the Chi and Chi-square distributions, another significant continuous probability distribution in statistics is the Gamma distribution, which is characterized by the shape parameter $\alpha$ and the scale parameter $\beta$, both positive. This distribution is particularly useful to model survival times, Poisson processes, and events rates [7–9]. When $2\alpha$ assumes a positive integer value $k$, a Chi-square distribution with $k$ degrees of freedom is obtained, which will be especially useful for the purposes of this article. Random variables following a Chi and Chi-square distribution with $k$ degrees of freedom can be generated using $k$ independent and identically distributed (iid) Gaussian variables, commonly referred to as generating functions.

In a previous study [10], the practical utility limits of the Chi distribution were explored when the generating functions did not have an identical distribution. For this purpose, Gaussian variables with a mean of 0 and different variances were generated. Along the same line, in [11], the robustness of this model was assessed, even when the generating functions were not symmetric, by replacing the Gaussian generators with ex-Gaussian functions. However, in both works, the degrees of freedom remained restricted to positive integers. In the present study, was focus on extending the use of the Chi distribution to non-integer degrees of freedom, thereby leveraging the existing connections between the Chi, Chi-square, and Gamma distributions. Although some of the ideas presented herein can be found in [12], the present study focuses on further exploring the relationships between different probability distributions and its contextualisation within biological processes. In particular, we focus on the practical applications of generalized degrees of freedom in real-world, therby demonstrating its suitability to modelling. As we will see, this approach is particularly useful in complex systems, such as biological systems, where responses are influenced by a variety of factors. Using non-integer values for $k$ helps capture this variability more accurately without adding extra variables to the model.

This article presents an analysis of the moment-generating functions (mgf) of the Chi-square and Gamma distributions, thereby aiming to provide an interpretation for non-integer degrees of freedom in terms of the shape parameter $\alpha$ of the Gamma distribution. Subsequently, the relationship between the Chi and Chi-square distributions will be used to extend this interpretation to the Chi distribution. Finally, real data in the fields of biological and health sciences will be used to assess the practical potential of the Chi distribution with non-integer degrees of freedom.

Although the extension of the $\chi$ distribution to non-integer values of $k$ was considered in classical works [12], we recall their lack of an explicit interpretation in biological contexts (mainly focused on applications in telecommunications and signal theory). In contrast, this work introduces two main contributions: (i) we show how the $\chi$ distribution with non-integer $k$ can be rigorously obtained from sums of Gamma distributions with a common scale parameter, thus providing a precise mathematical interpretation of fractional degrees of freedom; and (ii) we apply this extension to biological and

health phenomena, thereby interpreting fractional $k$ as an effective dimensionality parameter that simultaneously accounts for population heterogeneity and the partial correlation structure of the underlying processes.

## 2. Materials and methods

The purpose of this section is to present some definitions and results related to the probability distributions that will be discussed. For this part, the reader can refer to [13]. We will include a discussion about the suitability to model some biological processes with these distributions.

### *2.1. Preliminaries of the Gamma distribution*

The Gamma distribution is a continuous probability distribution whose probability density function (pdf) is given by

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}\exp\left\{\frac{-x}{\beta}\right\}}{\beta^{\alpha}\Gamma(\alpha)}, \tag{2.1}$$

where

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt, \tag{2.2}$$

and $\alpha$ and $\beta$ are positive parameters referred to as shape and scale parameters, respectively.

The mgf of the Gamma distribution is

$$M_G(t) = (1-\beta t)^{-\alpha}. \tag{2.3}$$

If a set of $m$ independent Gamma random variables $G_1, G_2, \ldots, G_m$ with shape parameters $\alpha_i$ for $i = 1, 2, \ldots, m$ and a common scale parameter $\beta$ are considered, then the random variable $G = \sum_{i=1}^{m} G_i$ follows a Gamma distribution with a shape parameter $\alpha = \sum_{i=1}^{m} \alpha_i$ and a scale parameter $\beta$. This can be represented as

$$\sum_{i=1}^{m} G_i(\alpha_i,\beta) = G\left(\sum_{i=1}^{m}\alpha_i,\beta\right). \tag{2.4}$$

### *2.2. Preliminaries of Chi and Chi-square distributions*

A random variable distributed according to a Chi-square distribution with $k$ degrees of freedom is defined and denoted as

$$\chi^2(k) = \sum_{i=1}^{k} X_i^2, \tag{2.5}$$

where, each of the $X_i$, $i = 1, 2, ..., k$ are independent and identically distributed (iid) Gaussian random variables with a mean of 0 and a variance of 1. Its pdf is given by

$$f(x;k) = \frac{x^{\frac{k}{2}-1}\exp\left\{\frac{-x}{2}\right\}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} \tag{2.6}$$

and its mgf is

$$M_{\chi^2}(t) = (1-2t)^{-\frac{k}{2}}. \tag{2.7}$$

Recall that if two random variables have the same mgf, then they have the same probability distribution [13, Theorem 6.1]. In particular, it should be noted that the expression (2.7) is a particular case of (2.3) when considering the parameters $\alpha = k/2$ and $\beta = 2$. Therefore, a random variable $X$ follows a Chi-square distribution if, and only if, $X$ is a random variable with a Gamma distribution with parameters $\alpha = k/2$ and $\beta = 2$. Furthermore, if in expression (2.5) we restrict to the case $k = 1$ and consider a Gaussian random variable with an arbitrary variance $\sigma^2$, then the corresponding mgf can be calculated and a relationship with the Gamma distribution can be established in a similar way. Indeed, note that

$$\begin{aligned} M_{\chi^2}(t) &= E\left[e^{tX^2}\right] \\ &= \int_{-\infty}^{\infty} \exp\left\{tx^2\right\} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\left(1-2\sigma^2 t\right)\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2\left(1-2\sigma^2 t\right)^{-1}}\right\} dx. \end{aligned}$$

If $S^2 = \sigma^2\left(1-2\sigma^2 t\right)^{-1}$, and $1-2\sigma^2 t > 0$, then it is obtained that

$$\begin{aligned} M_{\chi^2}(t) &= \frac{S}{\sigma}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}S} \exp\left(-\frac{x^2}{2S^2}\right) dx \\ &= \frac{S}{\sigma}\cdot 1 \\ &= \left(1-2\sigma^2 t\right)^{-\frac{1}{2}}. \end{aligned} \tag{2.8}$$

Now, consider a variable $\chi^2(k)$, $k > 1$:

$$\chi^2(k) = \sum_{i=1}^{k} X_i^2,$$

where each of the $X_i$ are iid Gaussian variables with mean 0 and variance $\sigma^2$. Then, its mgf can be calculated as

$$\begin{aligned} M_{\chi^2}(t) &= E\left(e^{t\sum_{i=1}^{k} X_i^2}\right) = E\left(e^{tX_1^2} e^{tX_2^2} ... e^{tX_k^2}\right) \\ &= E\left(e^{tX_1^2}\right) E\left(e^{tX_2^2}\right) ... E\left(e^{tX_k^2}\right) \\ &= \left(1-2\sigma^2 t\right)^{-\frac{k}{2}}. \end{aligned} \tag{2.9}$$

When comparing the mgf obtained in Eq (2.9) with the mgf of the Gamma distribution, it is easy to notice that the $\chi^2(k)$ distribution with generators that has a mean of 0 and a variance of $\sigma^2$ is essentially a Gamma distribution with

$$\alpha = \frac{k}{2} \text{ and } \beta = 2\sigma^2. \tag{2.10}$$

Therefore, the pdf of the $\chi^2(k)$ distribution with Gaussian generators with a mean of 0 and a variance of $\sigma^2$ can be generalized as

$$f(x;\sigma,k) = 2^{-\frac{k}{2}}\sigma^{-k}\left[\Gamma\left(\frac{k}{2}\right)\right]^{-1} x^{\frac{k}{2}-1}\exp\left\{\frac{-x}{2\sigma^2}\right\}. \tag{2.11}$$

Likewise, a random variable following a Chi distribution with $k$ degrees of freedom is typically defined as:

$$\chi(k) = \left(\chi^2(k)\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{k} X_i^2\right)^{\frac{1}{2}}, \tag{2.12}$$

where each of the $X_i$, $i = 1, 2, ..., k$ is an iid Gaussian random variable with a mean of 0 and a variance of $\sigma^2$. Its pdf is given by

$$g(x;B,k) = 2^{1-\frac{k}{2}}B^{-\frac{k}{2}}\left[\Gamma\left(\frac{k}{2}\right)\right]^{-1} x^{k-1}\exp\left\{-\frac{x^2}{2B}\right\},\ x > 0 \tag{2.13}$$

where $B = \sigma^2$ denotes the variance of the generating Gaussian variables. It acts as the scale parameter of the $\chi$ distribution, thereby directly influencing its variance, and is related to the Gamma distribution scale parameter through $B = \beta/2$.

Based on Eqs (2.9), (2.10), and (2.12), and using Property (2.4), when the degrees of freedom $k = 2\alpha$ are not a positive integer, the $\chi(k)$ distribution can be defined and interpreted as

$$\chi(k) = \left(\sum_{i=1}^{n} G(\alpha_i,\beta)\right)^{1/2} = \left(G\left(\sum_{i=1}^{n}\alpha_i,\beta\right)\right)^{1/2} = (G(\alpha,\beta))^{1/2}, \tag{2.14}$$

where $\alpha$ is the sum of the corresponding $\alpha_i$ values.

The requirement $\alpha > 0$ is sufficient for the Gamma function to absolutely converge and thus be well-defined (2.1), but it is not necessary. In fact, since $\Gamma$ is meromorphic on the complex plane, it can be analytically continued to $\mathbb{C}$ except for its poles (0 and the negative integers). Therefore, we can eventually have $k$ defined for negative or complex numbers. Nevertheless, for the purposes of this study, we restrict our attention to $k \in \mathbb{R}_+$. In this case, the generalized Chi distribution can be generated as indicated in Eq (2.14), and its probability density function is given by Eq (2.13), valid for $x > 0$ and $k > 0$.

### *2.3. Chi Distribution in dynamical biological contexts*

One of the key advantages of modeling with the Chi distribution is its strong physical foundation, which underpins numerous natural processes. This is particularly evident in the mentioned Rayleigh distribution for 2 degrees of freedom and the Maxwell-Boltzmann distribution for 3 degrees of freedom. Moreover, being single-parameter models for $B$, these distributions are often preferred for their parsimony. However, when these distributions are applied outside their theoretical framework to real-world contexts, their underlying assumptions are often not met. For example, the assumption

of independence between the latent Gaussian variables frequently proves to be an oversimplification. Although these assumptions are often close enough to reality for the chi distribution to provide a good fit, in some situations, the fit can be poor even when the model's fundamental principles remain applicable. For this reason, our work proposes the introduction of an additional, non-integer parameter in the degrees of freedom. This approach offers a greater flexibility to fit the model to data while retaining the valuable interpretability of the chi distribution's original framework.

Nevertheless, the non-integer value of $k$ is not just a quantitative adjustment to improve the fit, but it may also have biological significance. In biological systems, the fractional degree of freedom can be understood as an effective measure of the number of mechanisms which drive the phenomenon under study. When the underlying processes are not fully independent, the effective dimensionality of the system may change, yielding non-integer values of $k$. Moreover, population heterogeneity contributes to this effect: individuals often exhibit different sensitivities or transition rates, and the aggregation of such heterogeneous sub-processes produces a fractional $k$ that reflects the weighted superposition of multiple biological mechanisms.

It is important to note that models such as the Gamma, Weibull, or Lognormal distributions also allow for fractional parameters and are well established in survival analysis, like the ones we are interested in. However, they are typically linked to general stochastic processes such as waiting times, hazard rates, or multiplicative effects. By contrast, the Chi distribution provides a complementary perspective: its degree of freedom parameter $k$ is grounded in the number of underlying system dimensions, thereby retaining a clear physical and geometric interpretation connected to classical results in statistical physics. Extending $k$ to fractional values preserves this structural link while simultaneously offering a way to represent the effective number of mechanisms in biological systems. This combination of statistical flexibility and mechanistic interpretability constitutes the distinctive advantage of the generalized Chi distribution framework proposed here.

To further explain the interpretability of our proposal, we present the following example for illustrative purposes. An exponential decay manifests itself in various natural phenomena, such as radioactive disintegration. Recall that if $N(t)$ is the number of radionuclides at time $t$ and $N_0$ is the initial number of radionuclides, then

$$N(t) = N_0 e^{-\lambda t},$$

where $\lambda$ is the radioactive decay constant, which is the probability of disintegration per unit of time. Therefore, the variation rate of the number of radioactive nuclides per unit of time is given by

$$A(t) = -\frac{dN(t)}{dt} = N_0 \lambda e^{-\lambda t}.$$

In this case, it is well-known the suitability of the exponential distribution. However, the probability $\lambda$ may not be constant over time, for example if there is a time dependence such as $\lambda(t) = \lambda_0 t$. Then, $A(t) = N_0 \lambda_0 t e^{-\lambda_0 t^2}$. Note that this expression corresponds to the pdf of the $\chi$ distribution with $k = 2$ according to (2.13). This situation is natural and typical of living systems, such as the ones we will study, which change over time. This is in contrast to inert systems, such as radioactive disintegration, which do not change. Biological phenomena, such as seed germination and disease progression, exhibit changing behaviour over time. For instance, the probability of a seed germinating on the first day differs from that a week later. Moreover, we explained earlier, using a non-integer value of $k$ helps to reflect this changing behaviour more accurately, thereby improving the quality of the results while retaining their physical interpretability.

### 2.4. Parameter estimation

We analyzed three datasets: (i) germination times of *Pinus tropicalis* seeds cultivated in western Cuba; (ii) the interval between infection by virulent tuberculous bacilli and death from tuberculosis in guinea pigs; and (iii) the period from symptom onset to death from COVID-19 in patients in Mexico during 2020.

For each dataset, the parameters of the different distributions we compared were estimated by *maximum likelihood*. Optimization was performed via numerical optimization using SciPy (`minimize`, Powell method), with convergence tolerances set to $xtol = 10^{-9}$ and $ftol = 10^{-9}$, and a maximum of 40,000 iterations. For the generelized Chi distribution, we set $k = 5$ as a reference initialization, while the parameter $B$ was initially approximated from the sample variance through the relationship

$$\mathrm{Var}(X) = B\left[k - 2\left(\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\right)^2\right]. \tag{2.15}$$

Once the parameters had been fitted, we reported the log-likelihood (LL), the corrected Akaike information criterion (AICc), the Bayesian information criterion (BIC) and the Kolmogorov–Smirnov p-value (KS p) in order to compare the different distributions. The KS p-value was used as a measure of goodness of fit. The results were stable with respect to the initialisation and no numerical instabilities were observed during the fitting process.

## 3. Results and discussion

After having argued that it makes sense to use the Chi distribution even when the degree of freedom is not an integer number, we are in a position to apply it to real data.

As the first case, we analyze the germination times of *Pinus tropicalis* Morelet seeds grown under uniform conditions in western Cuba.

As discussed in the previous section, the Rayleigh model is suitable to analyze seed germination times under the assumptions that the seeds are identical and that the germination of one seed does not affect the others. In this first approach, the fitting yields a parameter estimate of $B = 120.15 \pm 12.24$. The Kolmogorov–Smirnov p-value (0.07) indicates that there is insufficient statistical evidence to reject the Rayleigh distribution as a theoretical model, thus indicating that the model captures the fundamentals of the phenomenon. However, in order to improve the overall fit and motivated by the exploration of a more flexible model, while retaining the same theoretical framework, we considered the Chi distribution with free parameters $k$ and $B$. Under this approach, a significant improvement in the fit was achieved, thus reaching a $p$-value of 0.14, with estimates of $k = 1.74$ and $B = 138.43$. The corresponding fit is shown in Figure 1.

A comparative summary of the fitted models is presented in Table 1. The Weibull distribution provides the best overall fit, as indicated by its lower log-likelihood, AICc, and BIC values, together with a relatively high KS $p$-value (0.25). The Lognormal distribution shows a moderate performance, thereby yielding intermediate information criteria and the highest KS $p$-value (0.43), although its log-likelihood and information criteria are less favorable than those of the Weibull. In turn, the Chi distribution, yields parameter estimates ($k = 1.74 \pm 0.22$, $B = 138.43 \pm 23.28$) with a KS $p$-value of 0.14, thus confirming that it achieves a competitive fit despite presenting higher information criteria.

Given its theoretical relevance within our framework, the Chi distribution represents a well-grounded and practically valid alternative to the Weibull and Lognormal models. We emphasize that, unlike the Lognormal or Weibull, the Chi distribution is not merely a numerical adjustment: it preserves a direct connection with the system's degrees of freedom, thus offering a geometric and physical interpretation that enriches the analysis.

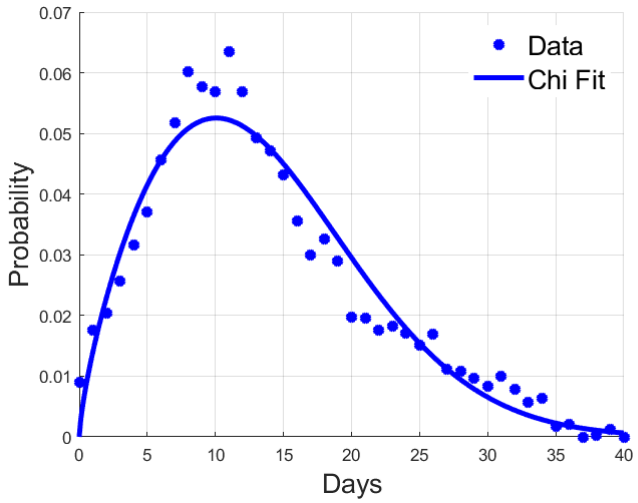


**Figure 1.** Fit of the Chi distribution to the germination times of *Pinus tropicalis* Morelet seeds grown under uniform conditions in western Cuba. Estimated parameters are $k = 1.74$ and $B = 138.43$, with the goodness-of-fit assessed by the Kolmogorov–Smirnov statistic (KS p = 0.14).

**Table 1.** Comparative summary of probability density functions fitted to the germination-time data. Parameter estimates are reported with their standard errors. The last columns present: LL, AICc, BIC, and KS p.

| Model | Parameters (±SE) | LL | AICc | BIC | KS p |
|---|---|---|---|---|---|
| Chi | $k = 1.74 \pm 0.22,\ B = 138.43 \pm 23.28$ | -359.86 | 723.84 | 728.92 | 0.14 |
| Weibull | $c = 1.75 \pm 0.14,$ scale $= 14.95 \pm 0.90$ | -340.34 | 684.81 | 689.90 | 0.25 |
| Lognormal | $\sigma = 0.75 \pm 0.05,$ scale $= 10.77 \pm 0.81$ | -352.40 | 708.92 | 714.01 | 0.43 |

Second, we analyzed a dataset describing the time (in days) elapsed between inoculation with virulent tuberculous bacilli and death from tuberculosis in five cohorts of guinea pigs. These cohorts were labeled 1, 2, 3, 4, and 5. Each cohort was administered an increasing dose of bacilli, with the first receiving the lowest dose and the fifth the highest. The parameter estimates for $k$ and $B$, together with the corresponding values of LL, AICc, BIC, and KS $p$, are reported in Table 2.

The results show that the Lognormal distribution generally provides the best fit across most cohorts, consistently achieving the highest KS $p$-values (0.95, 0.99, and 0.50 in cohorts 1, 2, and 5) and the lowest information criteria. The Weibull distribution performs moderately well, particularly in cohorts 4 and 5, where it offers a reasonable alternative despite being outperformed by the Lognormal. In turn,

the Chi distribution, achieves an especially good fit in cohort 3 (KS $p = 0.53$), thus outperforming both the Weibull and the Lognormal in terms of the goodness-of-fit, although its performance is clearly inferior in cohorts 4 and 5. Taken together, these results indicate that, although the Lognormal tends to dominate in terms of statistical fit, the Chi distribution achieves competitive results in certain cohorts and contributes an interpretative component that makes it especially valuable.

**Table 2.** Comparative summary of different probability density functions, showing the estimated values of their respective parameters along with standard errors for each cohort. The last columns present: LL, AICc, BIC, and KS p for the goodness of fit.

| Cohort | Model | Parameters (±SE) | LL | AICc | BIC | KS p |
|---|---|---|---|---|---|---|
| 1 | Chi | $k = 2.64 \pm 0.44,\ B = 27086.55 \pm 5397.02$ | -365.16 | 734.52 | 738.50 | 0.47 |
| | Weibull | $c = 2.22 \pm 0.21,\ \text{scale} = 273.45 \pm 16.86$ | -365.91 | 736.03 | 740.01 | 0.68 |
| | Lognormal | $\sigma = 0.47 \pm 0.04,\ \text{scale} = 216.45 \pm 13.02$ | -361.95 | 728.11 | 732.09 | 0.95 |
| 2 | Chi | $k = 2.72 \pm 0.43,\ B = 17926.58 \pm 3369.69$ | -394.16 | 792.52 | 796.74 | 0.44 |
| | Weibull | $c = 2.23 \pm 0.20,\ \text{scale} = 226.15 \pm 13.12$ | -395.23 | 794.65 | 798.88 | 0.58 |
| | Lognormal | $\sigma = 0.45 \pm 0.04,\ \text{scale} = 180.02 \pm 9.96$ | -389.93 | 784.04 | 788.26 | 0.99 |
| 3 | Chi | $k = 1.84 \pm 0.27,\ B = 22689.87 \pm 4322.63$ | -427.78 | 859.74 | 864.12 | 0.53 |
| | Weibull | $c = 1.83 \pm 0.16,\ \text{scale} = 199.60 \pm 13.63$ | -427.36 | 858.90 | 863.28 | 0.38 |
| | Lognormal | $\sigma = 0.63 \pm 0.05,\ \text{scale} = 149.05 \pm 11.05$ | -429.09 | 862.36 | 866.74 | 0.32 |
| 4 | Chi | $k = 2.23 \pm 0.34,\ B = 9127.16 \pm 1729.37$ | -380.84 | 765.85 | 770.14 | 0.03 |
| | Weibull | $c = 1.93 \pm 0.16,\ \text{scale} = 141.15 \pm 9.39$ | -380.98 | 766.14 | 770.43 | 0.13 |
| | Lognormal | $\sigma = 0.47 \pm 0.04,\ \text{scale} = 110.47 \pm 6.19$ | -369.74 | 743.67 | 747.96 | 0.40 |
| 5 | Chi | $k = 1.19 \pm 0.17,\ B = 14153.73 \pm 2930.90$ | -401.89 | 807.95 | 812.33 | 0.01 |
| | Weibull | $c = 1.38 \pm 0.12,\ \text{scale} = 111.22 \pm 10.10$ | -398.02 | 800.22 | 804.60 | 0.07 |
| | Lognormal | $\sigma \pm 0.06,\ \text{scale} = 77.25 \pm 6.51$ | -391.03 | 786.23 | 790.61 | 0.50 |

As mentioned earlier, Chi distributions with two and three degrees of freedom correspond to the Rayleigh and Maxwell–Boltzmann distributions, respectively. These distributions provide particularly clear interpretations of the parameter $k$ and its link to the system dimensionality. In the present context, $k$ can be understood as an effective dimensionality of the system. Fixing $k$ at an average value of 2.12 reflects the assumption that the same structural dimensions govern the phenomenon across all cohorts, thus illustrating one of the advantages of the Chi distribution: beyond providing a statistical fit, it allows for a mechanistic interpretation in terms of the effective degrees of freedom.

Thus the observed variability is thus entirely captured by changes in the parameter $B$. From a biological perspective, increasing the inoculated dose leads to progressively shorter and more homogeneous survival times, thus directly affecting the variance of the distribution. Within this framework, $B$ becomes the key parameter modulating the variability of the system. This decomposition into a structural component ($k$) and a scale component ($B$) highlights the analytical power of the Chi distribution, as it allows us to link the statistical performance with its biological interpretation. The results are summarized in Table 3, and the corresponding fits are displayed in Figure 2.

**Table 3.** Summary of the estimated values for the parameter $B$ of the $\chi$ probability density function, along with its respective standard error for each cohort. In all fits, the value of $k$ was fixed at 2.12. The last column shows the KS p value for the goodness of fit.

| Cohort | $B$ | KS p |
|---|---|---|
| 1 | 33766.23 ± 4234.03 | 0.73 |
| 2 | 23033.52 ± 2733.19 | 0.45 |
| 3 | 19726.12 ± 2257.99 | 0.22 |
| 4 | 9599.81 ± 1122.50 | 0.05 |
| 5 | 7952.78 ± 910.33 | 0.00 |

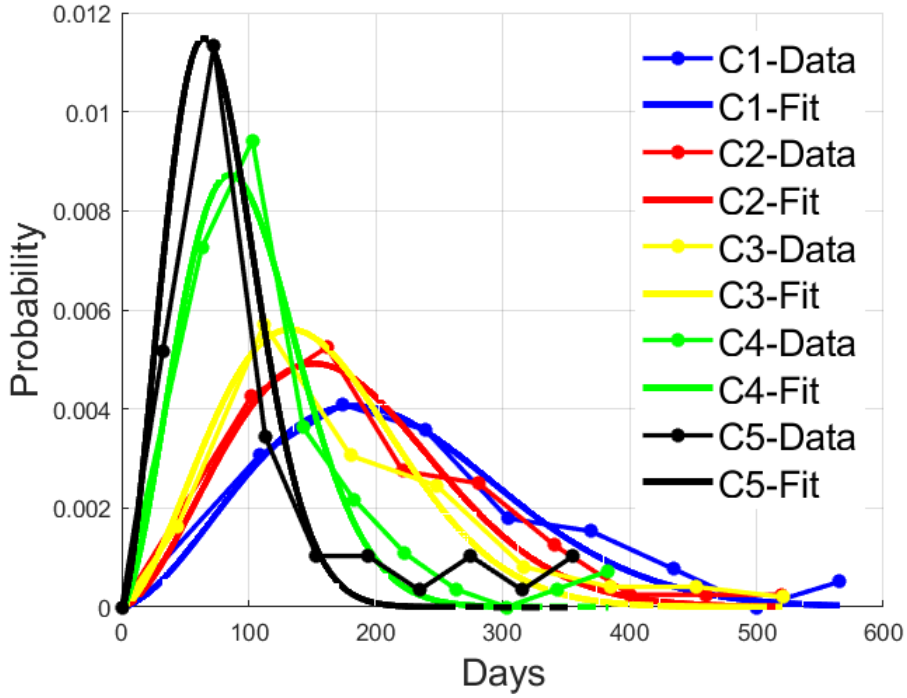


**Figure 2.** Survival time distribution of guinea pigs inoculated with virulent tuberculous bacilli. The curves show the experimental data and the $\chi$ distribution fits for the five cohorts with different doses. For all Chi distribution fits, $k$ was fixed at the average value of 2.12.

Observing the obtained values of $B$, an inverse relationship between the initial dose of tuberculous bacilli administered and the value of the parameter $B$ is evident. Since $B$ is directly related to the variance of the distribution, this implies that higher doses of virulent bacilli result in lower variabilities in the times elapsed from infection to death. This relationship is more clearly illustrated in Figure 3.

The third dataset corresponds to the times elapsed between the onset of symptoms and death from COVID-19 in eight groups of individuals, categorized by age and gender. Tables 4 and 5 present a comparative summary of the fits obtained with the Chi, Lognormal, and Weibull distributions for groups of men and women aged between 20 and 100 years.

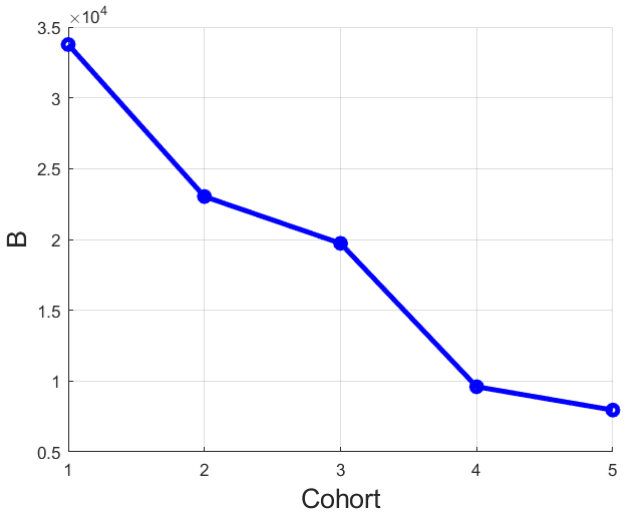


**Figure 3.** Fitted values of $B$ as a function of the initial dose of virulent tuberculous bacilli administered. The numbers 1, 2, 3, 4, and 5 on the x-axis represent the different cohorts of guinea pigs.

**Table 4.** Comparative summary of probability density functions fitted to the COVID-19 symptom-to-death time data (women, 20–99 years). Parameter estimates are reported with their standard errors. The last columns present: LL, AICc, BIC, and KS p.

| Model | Parameters (±SE) | LL | AICc | BIC | KS p |
|---|---|---|---|---|---|
| | **20–39 years** | | | | |
| Chi | $k = 1.13 \pm 0.12,\ B = 238.24 \pm 36.50$ | -298.24 | 600.56 | 606.64 | 0.33 |
| Weibull | $c = 1.40 \pm 0.09,\ \text{scale} = 14.45 \pm 0.88$ | -295.16 | 594.40 | 600.48 | 0.57 |
| Lognormal | $\mu = 2.34 \pm 0.06,\ \sigma = 0.72 \pm 0.05$ | -295.34 | 594.76 | 600.83 | 0.62 |
| | **40–59 years** | | | | |
| Chi | $k = 1.20 \pm 0.09,\ B = 239.90 \pm 24.32$ | -783.65 | 1571.33 | 1578.83 | 0.16 |
| Weibull | $c = 1.43 \pm 0.06,\ \text{scale} = 15.06 \pm 0.63$ | -776.36 | 1556.76 | 1564.26 | 0.59 |
| Lognormal | $\mu = 2.36 \pm 0.04,\ \sigma = 0.76 \pm 0.03$ | -780.74 | 1565.51 | 1573.01 | 0.28 |
| | **60–79 years** | | | | |
| Chi | $k = 1.28 \pm 0.08,\ B = 214.22 \pm 19.93$ | -893.59 | 1791.22 | 1799.01 | 0.14 |
| Weibull | $c = 1.48 \pm 0.06,\ \text{scale} = 14.91 \pm 0.56$ | -884.69 | 1773.41 | 1781.20 | 0.59 |
| Lognormal | $\mu = 2.36 \pm 0.04,\ \sigma = 0.73 \pm 0.03$ | -887.43 | 1778.88 | 1786.68 | 0.24 |
| | **> 80 years** | | | | |
| Chi | $k = 1.40 \pm 0.13,\ B = 138.28 \pm 16.89$ | -461.81 | 927.67 | 934.31 | 0.54 |
| Weibull | $c = 1.58 \pm 0.09,\ \text{scale} = 12.99 \pm 0.60$ | -459.87 | 923.79 | 930.43 | 0.84 |
| Lognormal | $\mu = 2.24 \pm 0.05,\ \sigma = 0.70 \pm 0.04$ | -463.77 | 931.61 | 938.24 | 0.46 |

**Table 5.** Comparative summary of probability density functions fitted to the COVID-19 symptom-to-death time data (men, 20–99 years). Parameter estimates are reported with their standard errors. The last columns present: LL, AICc, BIC, and KS p.

| Model | Parameters (±SE) | LL | AICc | BIC | KS p |
|---|---|---|---|---|---|
| | **20–39 years** | | | | |
| Chi | $k = 1.11 \pm 0.10,\ B = 253.27 \pm 32.22$ | -472.58 | 949.21 | 955.94 | 0.27 |
| Weibull | $c = 1.38 \pm 0.07,\ \text{scale} = 14.62 \pm 0.77$ | -467.67 | 939.39 | 946.12 | 0.70 |
| Lognormal | $\mu = 2.33 \pm 0.05,\ \sigma = 0.76 \pm 0.04$ | -468.88 | 941.81 | 948.54 | 0.56 |
| | **40–59 years** | | | | |
| Chi | $k = 1.36 \pm 0.08,\ B = 203.80 \pm 16.69$ | -1267.79 | 2539.61 | 2547.80 | 0.20 |
| Weibull | $c = 1.54 \pm 0.06,\ \text{scale} = 15.35 \pm 0.50$ | -1262.22 | 2528.48 | 2536.67 | 0.68 |
| Lognormal | $\mu = 2.39 \pm 0.04,\ \sigma = 0.74 \pm 0.03$ | -1276.80 | 2557.63 | 2565.82 | 0.14 |
| | **60–79 years** | | | | |
| Chi | $k = 1.39 \pm 0.08,\ B = 206.60 \pm 16.64$ | -1297.64 | 2599.30 | 2607.53 | 0.28 |
| Weibull | $c = 1.57 \pm 0.06,\ \text{scale} = 15.76 \pm 0.50$ | -1292.95 | 2589.92 | 2598.16 | 0.67 |
| Lognormal | $\mu = 2.41 \pm 0.03,\ \sigma = 0.73 \pm 0.03$ | -1309.19 | 2622.40 | 2630.64 | 0.12 |
| | **> 80 years** | | | | |
| Chi | $k = 1.21 \pm 0.09,\ B = 216.92 \pm 24.07$ | -639.75 | 1283.54 | 1290.70 | 0.05 |
| Weibull | $c = 1.43 \pm 0.07,\ \text{scale} = 14.25 \pm 0.65$ | -629.42 | 1262.88 | 1270.04 | 0.30 |
| Lognormal | $\mu = 2.32 \pm 0.04,\ \sigma = 0.72 \pm 0.03$ | -624.86 | 1253.76 | 1260.92 | 0.43 |

Observing the values of $k$ obtained in each fit, it is evident that they fall consistently between 1 and 2; which is the midpoint between a Chi distribution with one degree of freedom (a half-normal distribution) and a Rayleigh distribution. Consistent with the previous dataset, and to characterize each cohort using a single parameter B, the degrees of freedom will be fixed at their average value $k = 1.26$. The results obtained are summarized in Table 6, and their respective graphs are presented in Figure 4.

**Table 6.** Summary of the estimated values for the parameter $B$ of the $\chi$ pdf, along with its respective standard error and the KS $p$-value for each cohort. In all fits, $k$ was set to 1.26.

| | Age (years) | $B$ | Standard error of $B$ | KS p |
|---|---|---|---|---|
| Women | 20–39 | 213.80 | 21.54 | 0.33 |
| | 40–59 | 228.56 | 16.15 | 0.13 |
| | 60–79 | 217.33 | 14.30 | 0.16 |
| | > 80 | 154.10 | 13.47 | 0.45 |
| Men | 20–39 | 223.71 | 19.12 | 0.15 |
| | 40–59 | 220.09 | 13.09 | 0.26 |
| | 60–79 | 228.55 | 13.45 | 0.20 |
| | > 80 | 208.47 | 16.11 | 0.05 |

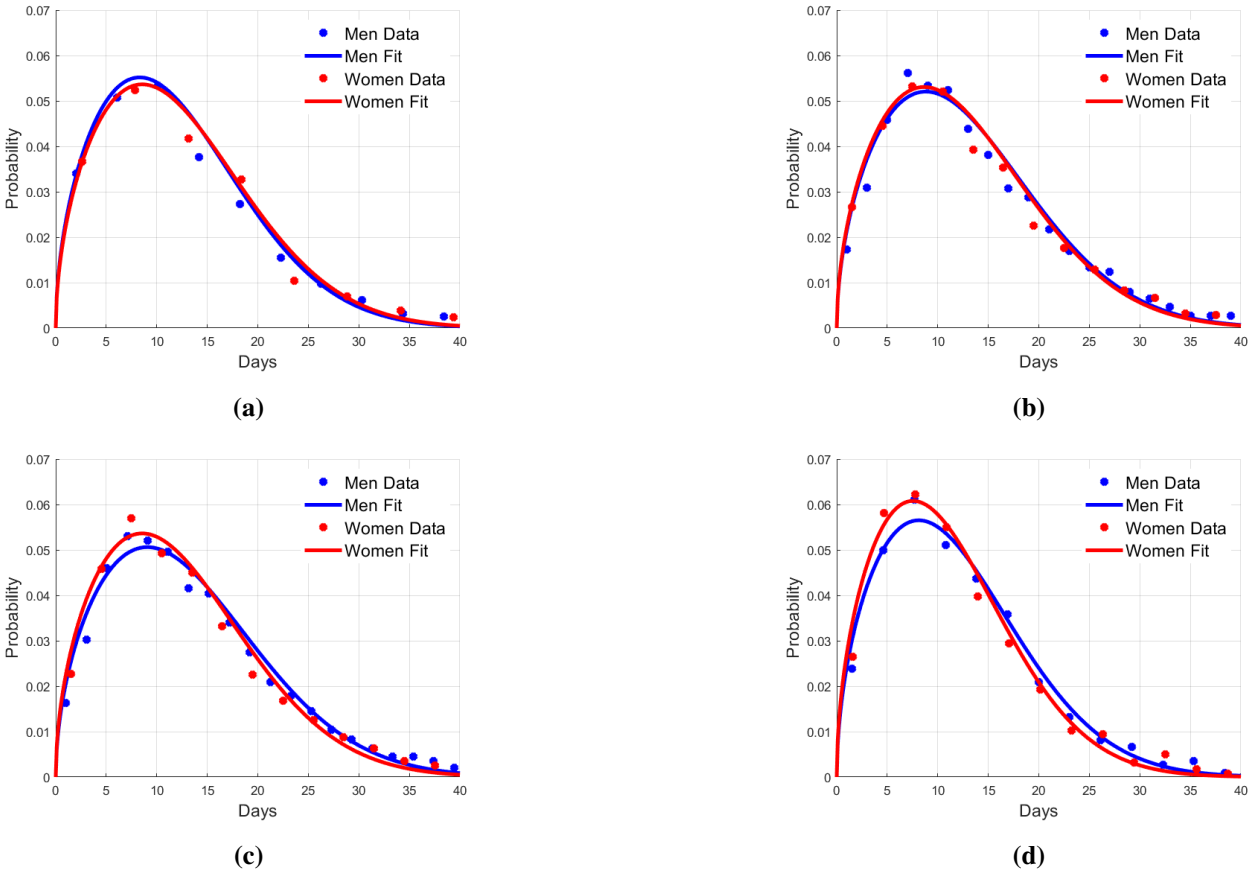


**Figure 4.** In panels (a), (b), (c), and (d), the $\chi$ distribution fittings for the cohorts of individuals aged between 20-40, 40-60, 60-80, and 80-100 years, respectively, are displayed.

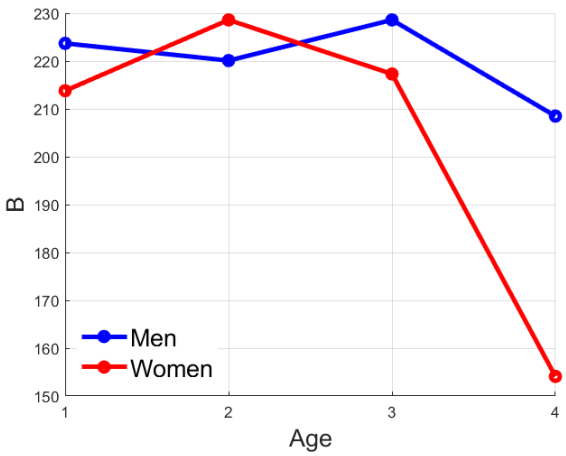


**Figure 5.** Fitted values of $B$ as a function of the age of individuals divided by sex. The numbers 1, 2, 3, and 4 on the x-axis represent individuals whose ages are 20–40, 40–60, 60–80, and 80–100 years, respectively.

Figure 5 shows the evolution of the parameter B as a function of age. These results suggest a

increase in the variability of the times between the onset of symptoms and death from the disease as age advances. However, this variability decreases drastically as individuals approach 80 years of age. This decrease suggests greater homogeneity in individuals' responses to the disease, which could be related to the natural increase in health fragility at these stages of life.

This study highlights the practical potential of the Chi distribution in analyzing survival times while expanding its applicability to non-integer degrees of freedom. This advancement is particularly significant as it allows us to leverage its advantages, such as an ease of simulation and data representation using a single parameter. These findings not only enhance our understanding of the Chi distribution but also open new possibilities for future research in various scientific contexts.

## 4. Conclusions

The Chi distribution has been widely used in various scientific fields. Although its probability density function allows for non-integer degrees of freedom, its interpretation in such cases may not be completely clear. In this paper, an interpretation of the Chi distribution with non-integer degrees of freedom is provided. In Eq 2.14, it has been established that the Chi distribution, in its most general version, that is, when $k$ is not an integer, can be generated by summing Gamma distributions with the same scale parameter $\beta$ and shape parameters $\alpha_i$, whose values sum to $k$.

Furthermore, the present study demonstrated that the Chi distribution constitutes a robust, versatile and interpretable framework to model biological data when extended to non-integer degrees of freedom. One of its major strengths is that improvements in the statistical fit can be made without compromising the theoretical structure or physical interpretability of the classical Chi distribution. Allowing fractional values of $k$ gives the model the flexibility needed to accommodate the deviations inherent in the hypothesis required for the chi model, which characterises biological systems, while retaining the clear geometric and mechanistic meaning associated with the degrees of freedom. Overall, this extension to non-integer degrees of freedom opens new possibilities for research in various scientific contexts, which allows for a better understanding and modeling of complex phenomena in areas such as biology, physics, engineering, and social sciences. This methodological expansion represents a significant advance in the researchers' ability to address complex problems with robust and versatile statistical tools.

As a demonstration of the practical potential of these ideas, the Chi distribution was applied to three datasets. For the seed germination times of *Pinus tropicalis Morelet*, it was determined that the optimal fit is at $k = 1.74$ and $B = 138.43$.

For the data concerning the interval between inoculation with virulent tuberculous bacilli and death from tuberculosis in various cohorts of guinea pigs, a comprehensive analysis was performed owing to the dataset's richness. By setting $k = 2.12$, each cohort was described by a single parameter $B$, with $B$ being directly linked to the variance of the Chi distribution; it was found that higher initial doses of bacilli were associated with reduced variabilities in the time to death from the disease.

In the context of survival times related to COVID-19, the intervals between the onset of symptoms and death were examined across different age and gender cohorts. For all groups, using $k = 1.26$, high-quality fits were achieved. These findings underscore the effectiveness of the Chi distribution in modeling complex survival data. Additionally, it was observed that the variability in the time from symptom onset to death increases with age in both sexes. However, in individuals over 80 years old,

this variability decreases significantly, thus suggesting a more homogeneous response to the disease among the elderly.

Taken together, the results obtained show that the Chi distribution with non-integer degrees of freedom not only constitutes a valid alternative to traditional models such as Weibull and Lognormal, but also provides a unique interpretative perspective through the parameter $k$. This methodological framework can be extended to other contexts where survival or waiting times play a central role, offering a balance between statistical flexibility and biological interpretability. Nevertheless, future research should further investigate the interpretation of the parameter $k$ as an effective measure of the underlying mechanisms. These directions open a promising path to broaden the scope and practical utility of the Chi distribution across different scientific disciplines.

## Author contributions

Juan Carlos Castro-Palacio: Formal analysis, Data curation, Methodology; Pedro Fernández de Córdoba: Conceptualization, Supervision, Writing – review & editing; Álvaro González Cortés: Investigation, Writing – review & editing, Visualization; J. M. Isidro: Conceptualization, Investigation, Methodology; A. Noverques: Formal analysis, Methodology; Marcos Orellana-Panchame: Data curation, Investigation, Writing – original draft; Sarira Sahu: Formal analysis, Investigation; Enrique A. Sánchez Pérez: Supervision, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

We would like to acknowledge funding from the Generalitat Valenciana (Spain) through the PROMETEO CIPROM/2023/32 grant, and from the Agencia Estatal de Investigación (Spain) through the PID2022-142407NB-I00 and PID2024-162480OB-I00 grants. The work of S.S. is partially supported by DGAPA-UNAM (México) Project No. IN103522. The author A.G.C. was supported by a contract of the Programa de Ayudas de Investigación y Desarrollo (PAID-01-24), Universitat Politècnica de València.

## Conflict of interest

Dr. Pedro Fernández de Córdoba and Dr. A. Noverques are the Guest Editors of the special issue "Mathematical Modelling and Numerical Simulation of Physical & Engineering Problems" for AIMS Mathematics. Dr. Pedro Fernández de Córdoba and Dr. A. Noverques were not involved in the editorial review and in the decision to publish this article. The authors declare no conflict of interest.

## References


1. D. W. K. Andrews, Chi-square diagnostic tests for econometric models, *J. Econometrics*, **37** (1988), 135–156. https://doi.org/10.1016/0304-4076(88)90079-6
2. K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, **50** (1900), 157–175.
3. E. J. Gumbel, On the reliability of the classical chi-square test, *Ann. Math. Stat.*, **14** (1943), 253–263.
4. R. A. Fisher, The conditions under which chi-squared measures the discrepancey between observation and hypothesis, *J. R. Stat. Soc.*, **87** (1924), 442–450. https://doi.org/10.1093/bjps/34.2.105
5. B. McWilliams, M. M. Newmann, D. Sprevak, The probability distribution of wind velocity and direction, *Wind Eng.*, **3** (1979), 269–273.
6. R. C. Tolman, *The principles of statistical mechanics*, New York: Dover Publications, 1980.
7. C. W. Zhang, M. Xie, J. Y. Liu, T. N. Goh, A control chart for the gamma distribution as a model of time between events, *Int. J. Prod. Res.*, **45** (2007), 5649–5666.
8. D. W. Coit, T. Jin, Gamma distribution parameter estimation for field reliability data with missing failure times, *IIE Trans.*, **32** (2000), 1161–1166.
9. R. Winkelmann, A count data model for gamma waiting times, *Stat. Pap.*, **37** (1996), 177–187.
10. J. C. Castro-Palacio, J. M. Isidro, E. Navarro-Pardo, L. Velázquez-Abad, P. Fernández-de Córdoba, Monte carlo simulation of a modified chi distribution with unequal variances in the generating gaussians. A discrete methodology to study collective response times, *Mathematics*, **9** (2020), 77.
11. N. Ortigosa, M. Orellana-Panchame, J. C. Castro-Palacio, P. Córdoba, J. M. Isidro, Monte carlo simulation of a modified chi distribution considering asymmetry in the generating functions: Application to the study of health-related variables, *Symmetry*, **13** (2021), 924.
12. M. Nakagami, The m-distribution—a general formula of intensity distribution of rapid fading, In: *Statistical methods in radio wave propagation*, Belfast: The Universities Press, 1960, 3–36. https://doi.org/10.1016/B978-0-08-009306-2.50005-4
13. D. D. Wackerly, W. Mendenhall, R. L. Scheaffer, *Mathematical statistics with applications*, 7 Eds., Belmont: Thomson Higher Education, 2008.