Article

Percentile Study of χ Distribution. Application to Response Time Data

**Juan Carlos Castro-Palacio [1], Pedro Fernández-de-Córdoba [1] , J. M. Isidro [1] ,
Esperanza Navarro-Pardo [2],* and Romeo Selvas Aguilar [3]**

[1]   Grupo de Modelización Interdisciplinar, InterTech, Instituto Universitario de Matemática Pura y Aplicada,
    Universitat Politècnica de València, E-46022 Valencia, Spain; juancas@upvnet.upv.es (J.C.C.-P.);
    pfernandez@mat.upv.es (P.F.-d.-C.); joissan@mat.upv.es (J.M.I.)
[2]   Grupo de Modelización Interdisciplinar, InterTech, Departamento de Psicología Evolutiva y de la Educación,
    Universitat de València, E-46010 Valencia, Spain
[3]   Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Nuevo
    León, San Nicolás de los Garza 66455, Nuevo León, Mexico; romeo.selvasag@uanl.edu.mx
*   Correspondence: esperanza.navarro@uv.es

check for
updates

Abstract: As a continuation of our previous work, where a Maxwell–Boltzmann distribution was
found to model a collective's reaction times, in this work we will carry out a percentile study of the χ
distribution for some freedom ranging from $k = 2$ to $k = 10$. The most commonly used percentiles in
the biomedical and behavioral sciences have been included in the analysis. We seek to provide a
look-up table with percentile ratios, taken symmetrically about the median, such that this distribution
can be identified in practice in an easy way. We have proven that these ratios do not depend upon the
variance chosen for the k generating Gaussians. In general, the χ probability density, generalized
to take any value of the variance, represents an ideal gas in a k-dimensional space. We also derive
an approximate expression for the median of the generalized χ distribution. In the second part of
the results, we will focus on the practical case of $k = 3$, which represents the ideal gas in physics,
and models quite well the reaction times of a human collective. Accurately, we will perform a more
detailed scrutiny of the percentiles for the reaction time distribution of a sample of 50 school-aged
children (7200 reaction times).

Keywords: χ distribution; ideal gas model; Maxwell–Boltzmann distribution; response times

1. Introduction

The χ distribution is a well-known probability density function (PDF) for its use in applied
statistics [1], and for its appearance in several contexts, such as the Rayleigh distribution ($k = 2$) and
the Maxwell–Boltzmann (MB) distribution in the ideal gas model [2]. This PDF is also interesting
because it is built from Gaussian-distributed random variables combined through a Euclidean distance
operation. This fact makes the χ distribution easily applicable to model the velocity distribution of a
multidimensional ideal gas.

In a recent work [3], we discovered the MB distribution ($k = 3$) when studying the reaction times
of a collective of 168 children. A correspondence between a system of brains and a system of identical,
independent particles forming the ideal gas was evidenced.

We are particularly interested in applications to biomedical and collective behavior contexts.
In this respect, and in line with our previous works [3,4], we will carry out here a percentile study of
the χ distribution. Taking as our reference the most commonly used percentiles, we will report on the
percentile ratios, such that the χ PDF can be easily identifiable when exploring data samples.

This work also includes an MB distribution representation of the reaction times of a collective of 50 school-aged children. The participants were asked to respond to visual stimuli within a few hundred milliseconds. We have proven in our previous work [3] that the collective behavior of individuals at short times is ruled by an MB distribution which depends on a single parameter B, namely, a measure of the dispersion of the distribution.

The outline of this work is as follows. In Section 2, the methodology for the percentile study and the analysis of the experimental response time data of 50 school-aged children is explained. Subsequently, in Section 3, the main results are discussed in two subsections: consequences for the percentile study of the χ distribution (PDF and CDF), and the case of $k = 3$, as applied to describe collective reaction times. Finally, in Section 4, some conclusions are drawn.

2. Methodology

2.1. Percentile Analysis of the χ Distribution (from k = 2 to k = 10)

The χ probability density distribution can be written as,

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

where k is the number of the degrees of freedom of the distribution, and its variance is related to the parameter B. The corresponding cumulative distribution function (CDF) is expressed as

$$F(x;k) = 1 - \left(\Gamma\!\left(\frac{k}{2}\right)\right)^{-1} \Gamma\!\left(\frac{k}{2}, \frac{x^2}{2\,B}\right),$$ (2)

where $\Gamma\!\left(\frac{k}{2}, \frac{x^2}{2B}\right)$ is the upper incomplete gamma function and $\Gamma\!\left(\frac{k}{2}\right)$ is the gamma function. The χ distribution as written in Equation (1) describes a multidimensional ideal gas of particles.

Based on the CDF, the following percentiles combinations were determined: **1** (p1-p50-p99), **2** (p3-p50-p97), **3** (p5-p50-p95), **4** (p10-p50-p90), **5** (p20-p50-p80), **6** (p25-p50-p75), and **7** (p30-p50-p70). The preceding number is just a label to identify each combination. The three numbers between brackets are the positions of the symmetric percentiles about the median (p 50). For each value of the degree of freedom from $k = 2$ to $k = 10$, a ratio was calculated as follows for each percentile combination,

$$r = \frac{x_2 - x_{\text{median}}}{x_{\text{median}} - x_1}$$ (3)

where x_1 and x_2 are the positions of symmetric percentiles on the left and right ends of a given percentile combination, and x_{median}, the position of the median (e.g., for (p1-p50-p99): $x_1 = $ p1, $x_2 = $ p99 and $x_{\text{median}} = $ p50). It is important to remark that we have numerically proven that the ratio r is independent of the value of the parameter B in Equation (1), which is easy to demonstrate analytically:

Let $x_p(B)$ be the point where the CDF takes the value of the percentile p for a given value of the parameter B, that is,

$$F\!\left(x_p(B); k\right) = 1 - \left(\Gamma\!\left(\frac{k}{2}\right)\right)^{-1} \Gamma\!\left(\frac{k}{2}, \frac{x_p^2(B)}{2B}\right) = p$$ (4)

Then, $x_p(B)$ satisfies

$$\Gamma\!\left(\frac{k}{2}, \frac{x_p^2(B)}{2B}\right) = (1 - p)\, \Gamma\!\left(\frac{k}{2}\right)$$ (5)

If we denote $\Gamma_{k/2}(y) \equiv \int_y^\infty t^{k/2-1} e^{-t} dt$, we can write the former expression as:

$$\Gamma_{k/2}\left(\frac{x_p^2(B)}{2B}\right) = (1-p)\,\Gamma\left(\frac{k}{2}\right) \tag{6}$$

Thus, $x_p^2(B) = 2\,B\,\Gamma_{\frac{k}{2}}^{-1}\left((1-p)\,\Gamma\left(\frac{k}{2}\right)\right)$, where $\Gamma_{\frac{k}{2}}^{-1}\left((1-p)\,\Gamma\left(\frac{k}{2}\right)\right)$ denotes the inverse function, and therefore $x_p\,(B)$ depends on the parameter B only through the factor $\sqrt{B}$. Then we can establish that: $x_p\,(B) = \sqrt{B}\cdot x_p\,(B=1)$. From the last relation, it is straightforward to demonstrate that r is independent of the parameter B as the factor $\sqrt{B}$ is common to $x_2(B)$, $x_{\text{median}}(B)$, and $x_1(B)$.

2.2. Response Time Experiments. A case for k = 3

In the second part of this work we will focus on a percentile study of the MB distribution ($k = 3$), which is in line with our previous works on reaction times to visual stimuli [3,4]. In addition to the aforementioned percentile combinations (labeled as **1** to **7** in Section 2.1), the following ones covering a broader area under the curve will be determined: **8** (p0.5-p50-p99.5), **9** (p0.7-p50-p99.3), **10** (p2-p50-p98), **11** (p7-p50-p93), **12** (p15-p50-p85), and **13** (p40-p50-p60).

As a case study, we will analyze the response times of a sample of 50 school-aged children between 8 and 10 years of age from Valencia, Spain. The details of the computer-based experiments and for the design of the visual stimuli can be found in our previous work [4], also the methodology for the data processing. However, we will recall in what follows a few expressions with relevance for this work.

From the first three moments (mean: M_i, standard deviation: S_i and skewness: λ_i) of the reaction time distributions for each participant (denoted by the index i), the following vector can be defined

$$\vec{v}_i = \left(\left(M_i - \overline{M}\right)/\overline{M}, \left(S_i - \overline{S}\right)/\overline{S}, \left(\lambda_i - \overline{\lambda}\right)/\overline{\lambda}\right) = \left(v_i^{(1)}, v_i^{(2)}, v_i^{(3)}\right), \tag{7}$$

where $\overline{M}$, $\overline{S}$ and $\overline{\lambda}$ are the mean values of the moments over the sample of children. Then, the Euclidean norm of the vector, $\vec{v}_i$, is calculated as

$$\left|\vec{v}_i\right| = \sqrt{\left(v_i^{(1)}\right)^2 + \left(v_i^{(2)}\right)^2 + \left(v_i^{(3)}\right)^2} \tag{8}$$

Each participant can now be represented by a single scalar, $\left|\vec{v}_i\right|$, which has been evidenced to be distributed following an MB distribution [4]. The following expression will be used to fit the experimental data,

$$Q\left(\left|\vec{v}_i\right|\right) = \sqrt{\frac{2}{\pi}} B^{-3/2} \left|\vec{v}_i\right|^2 \exp\left(-\frac{\left|\vec{v}_i\right|^2}{2B}\right) \tag{9}$$

where the parameter B is the only free parameter.

3. Results and Discussions

3.1. Multidimensional, Anisotropic Ideal Gas (χ of k Degrees of Freedom)

The PDF curves of the χ distribution for values from $k = 2$ to $k = 10$ are depicted in Figure 1a and the CDF in Figure 1b for $B = 0.5$. As the values of k increase, the maxima are shifted to the right, and the curves become more symmetric. Similarly, the differences among contiguous curves decrease as the values of k increase. The maximum value of skewness is given for $k = 2$ (Rayleigh distribution).
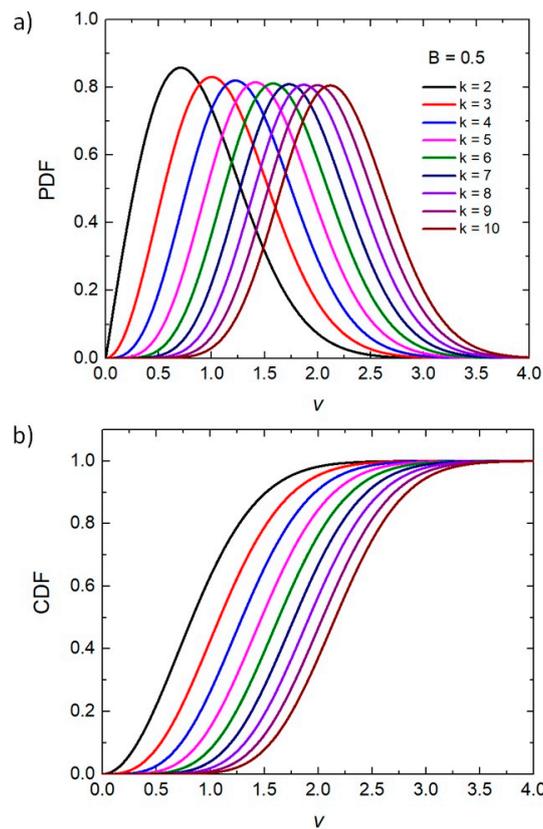
Figure 1. Probability (panel **a**) and cumulative (panel **b**) density functions of the χ distribution for $k = 2$ to $k = 10$. In both panels the value of B is 0.5 (see Equations (1) and (2)).

In Table 1, the ratios r for different combinations of percentiles Equation (3) are shown. The percentiles were taken symmetrically about the median. The biggest differences across the χ distributions for different k are observed for the ratios involving the smallest values of x_1 and the largest values of x_2 (see Figure 2), thus allowing us to better discriminate among the different curves.
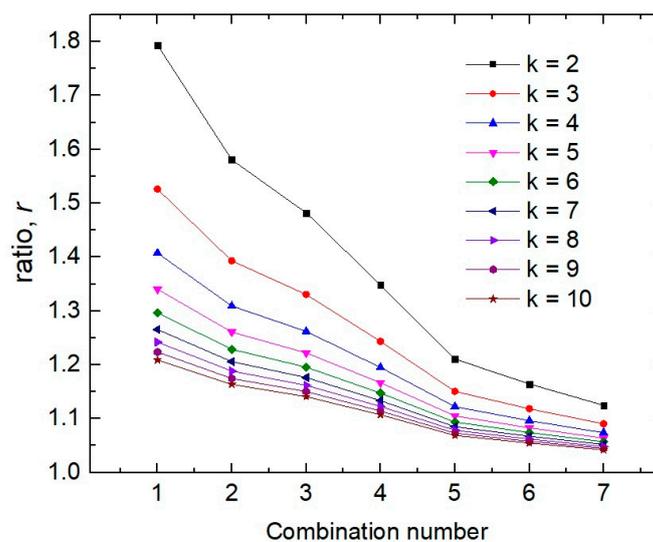
Figure 2. Percentile ratio, r, versus the percentile combination number. The legend for the combination number is: **1** (p1-p50-p99), **2** (p3-p50-p97), **3** (p5-p50-p95), **4** (p10-p50-p90), **5**(p20-p50-p80), **6** (p25-p50-p75), and **7** (p30-p50-p70).

Table 1. Ratios (r) for different combinations of percentile. As demonstrated above, these ratios are independent of the values of B chosen in Equations (1) and (2).

Combinations	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
1 (p1-p50-p99)	1.7935	1.5259	1.4075	1.3404	1.2966	1.2655	1.2421	1.2237	1.2088
2 (p3-p50-p97)	1.5805	1.3930	1.3090	1.2607	1.2287	1.2058	1.1884	1.1747	1.1635
3 (p5-p50-p95)	1.4821	1.3305	1.2619	1.2220	1.1955	1.1764	1.1618	1.1502	1.1408
4 (p10-p50-p90)	1.3483	1.2435	1.1951	1.1666	1.1474	1.1335	1.1228	1.1143	1.1073
5 (p20-p50-p80)	1.2108	1.1506	1.1221	1.1051	1.0935	1.0850	1.0784	1.0731	1.0688
6 (p25-p50-p75)	1.1643	1.1183	1.0963	1.0831	1.0740	1.0673	1.0622	1.0580	1.0547
7 (p30-p50-p70)	1.1248	1.0904	1.0738	1.0638	1.0570	1.0519	1.0480	1.0448	1.0422

One interesting finding is that the percentile ratios are independent of the parameter B, which has been also proven analytically in Section 2.1. On the other hand, using that $x_{median}(B) = \sqrt{B} \cdot x_{median}(B = 1)$, the well-known approximate value of the median of a χ distribution of k degrees of freedom and parameter $B = 1$, ($x_{median} \approx \sqrt{k\left(1 - \frac{2}{9k}\right)^3}$) can be generalized in order to include the explicit dependence of the parameter B,

$$x_{median}(B) \approx \sqrt{B} \sqrt{k\left(1 - \frac{2}{9k}\right)^3} \tag{10}$$

3.2. Reaction Times and Ideal Gas. Appearance of the Golden Ratio

Figure 3 shows the probability distribution function of the $\left|\vec{v}_i\right|$ as defined in Equation (8). The experimental data are fitted to the MB function in Equation (9) by using the nonlinear fitting algorithm of Levenberg–Marquardt [5,6]. The fitted value of B is 0.123 ± 0.010 with a coefficient of determination of $R^2 = 0.90$. A total of 7200 reactions were involved in building the graph in Figure 3.

Using the cumulative distribution for $k = 3$ Equation (2), the percentile ratio r was determined for each of the following combinations of percentiles (beyond those presented in Table 1): $r = 1.6098$, **8** (p0.5-p50-p99.5), $r = 1.5690$, **9** (p0.7-p50-p99.3), $r = 1.4421$, **10** (p2-p50-p98), $r = 1.3305$, **11** (p7-p50-p93), $r = 1.1902$, **12** (p15-p50-p85), and $r = 1.0424$, **13** (p40-p50-p60).

When we look at the combination labelled as **8** (p0.5-p50-p99.5), we can see that a ratio very close to the Golden Ratio is obtained. In Table 2 we present the values of x_1 (where $CDF(x_1) = 0.005$) and x_2 (where $CDF(x_2) = 0.995$), the value of r and the relative error with respect to the Golden Ratio $\Phi = \frac{1 + \sqrt{5}}{2}$ [7].

Table 2. Values of x_1, x_2, x_{median}, the ratio r, and the relative error with respect to Φ ($k = 3$).

k	x_1	$CDF(x_1)$	x_2	$CDF(x_2)$	x_{median}	r	Φ	Relative Error
3	0.26789	0.005	3.5831	0.995	1.5382	1.6098	1.6180	0.5%

If we impose $x_1 = 0$ and $r = \Phi$, that is, $x_2 = x_{median}(1 + \Phi) = 4.027$, a value of $CDF(x_2) = 0.999$ is obtained. Therefore, the total probability is recovered within a precision of 0.001. From Equation (10), if we consider a χ distribution of k degrees of freedom and a parameter B, the previous expression can be written in general terms as $x_2 = \sqrt{B} \sqrt{k\left(1 - \frac{2}{9k}\right)^3}(1 + \Phi)$. Specifically, for the data summarized in Figure 3, that is, 7200 reaction times (50 participants times 144 reaction times each) and parameter $B = 0.123 \pm 0.010$, a value of $x_2 = 1.4169$ was found.
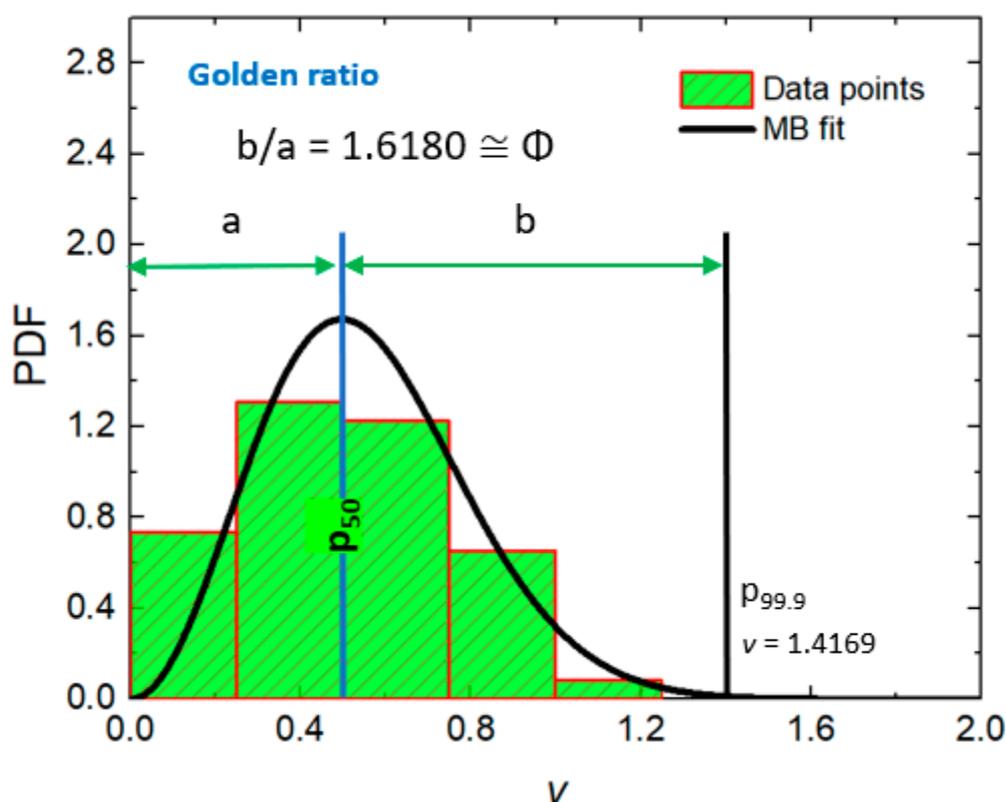
Figure 3. MB-like distribution (χ for $k = 3$ and $B = 0.123$) obtained from response time data with a coefficient of determination of $R^2 = 0.90$. The position of the median is indicated by a blue vertical line and the label "p50". A total of 7200 reaction times were used. The appearance of the Golden Ratio can be observed.

4. Conclusions

A percentile study of the χ distribution for values of the degree of freedom between $k = 2$ and $k = 10$ has been carried out. The ratios between combinations of percentiles, taken symmetrically about the median, have been reported. The independence of ratios with respect to the parameter B (proportional to the variance) has been shown both numerically and analytically. A general expression of the median as a function of the free parameter B and the degree of freedom (k) has also been reported for the χ distribution. Results show that those percentile combinations comprising bigger areas can be more reliable to identify this distribution (for different k values), as the most significant differences among the ratios are given. The combination of percentiles: p0.5-p50-p99.5 in the case of the MB distribution yielded a ratio, r, close to the Golden Ratio Φ, within a 0.5% of relative error. When the ratio r is taken equal to the Golden Ratio, a value of the cumulative density function of 0.999 is reached, that is, it approaches 1 with a precision of 0.001. This is a curious practical observation for the χ distribution of k = 3 degrees of freedom.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Doob, J.L. *Stochastic Processes*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1966.
2. Tolman, R. *The Principles of Statistical Mechanics*; Dover Publications Inc.: New York, NY, USA, 2003.
3. Hernaiz-Guijarro, M.; Castro-Palacio, J.C.; Navarro-Pardo, E.; Isidro, J.M.; Fernández de Córdoba, P. A Probabilistic Classification Procedure Based on Response Time Analysis towards a Quick Pre–diagnosis of Student's Attention Deficit. *Mathematics* **2019**, *7*, 473. [CrossRef]
4. Castro-Palacio, J.C.; Fernández de Córdoba, P.; Isidro, J.M.; Navarro-Pardo, E. Brain reaction times: Linking Individual and Collective Behaviour through Physics modelling. *arXiv* **2019**, arXiv:1910.12725.
5. Levenberg, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [CrossRef]
6. Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]
7. Meisner, G.B. *The Golden Ratio: The Divine Beauty of Mathematics*; Race Point Publishing: London, UK, 2018.
8. World Medical Association. Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *J. Am. Med. Assoc.* **2013**, *310*, 2191–2194. [CrossRef] [PubMed]