International Journal of Statistical Distributions and Applications
2025; 11(3): 85-97
http://www.sciencepublishinggroup.com/j/ijsda
doi: 10.11648/j.ijsda.20251103.11
ISSN: 2472-3487 (Print); ISSN: 2472-3509 (Online)

Properties and Approximations of a Bessel Distribution for Data Science Applications

Massimiliano Bonamente*

Department of Physics and Astronomy, University of Alabama in Huntsville, Huntsville, USA

Email address:
bonamem@uah.edu (Massimiliano Bonamente)
*Corresponding author

To cite this article:
Massimiliano Bonamente. (2025). Properties and Approximations of a Bessel Distribution for Data Science Applications. *International Journal of Statistical Distributions and Applications, 11*(3), 85-97. https://doi.org/10.11648/j.ijsda.20251103.11

Received: 9 March 2025; **Accepted:** 21 March 2025; **Published:** 5 August 2025

Abstract: This paper presents properties and approximations of a random variable based on the zero–order modified Bessel function that results from the compounding of a zero-mean Gaussian with a χ^2_1-distributed variance. This family of distributions is a special case of the McKay family of Bessel distributions and of a family of generalized Laplace distributions. It is found that the Bessel distribution can be approximated with a null-location Laplace distribution, which corresponds to the compounding of a zero-mean Gaussian with a χ^2_2-distributed variance. Other useful properties and representations of the Bessel distribution are discussed, including a closed form for the cumulative distribution function that makes use of the modified Struve functions. Another approximation of the Bessel distribution that is based on an empirical power-series approximation is also presented. The approximations are tested with the application to the typical problem of statistical hypothesis testing. It is found that a Laplace distribution of suitable scale parameter can approximate quantiles of the Bessel distribution with better than 10% accuracy, with the computational advantage associated with the use of simple elementary functions instead of special functions. It is expected that the approximations proposed in this paper be useful for a variety of data science applications where analytic simplicity and computational efficiency are of paramount importance.

Keywords: Bessel Distribution, Laplace Distribution, Generalized Laplace Distribution, Gaussian Distribution, Variance-gamma Distribution, Compounding of Random Variables, Hypothesis Testing

1. Introduction

Many experiments and natural phenomena are described by random variables whose parameters are themselves variable according to another distribution. This *compounding* of random variables is well documented in the literature, such as the Poisson distribution with a gamma-distributed rate parameter, which results in a negative binomial distribution [e.g. 4, 18, 23]. Compound distributions are sometimes referred to as *contagious* distribution [e.g. 19], based on the fact that certain models of diseases feature a (positive) correlation between the number and the probability of occurrence of certain outcomes [see, e.g., 18]. More recently, this type of distributions are referred to as *mixtures*, and they are commonly used in a variety of machine learning applications [e.g., the EM algorithm 3, 22].

For continuous variables, the compounding of a normal distributions is a topic of general interest. While the compounding of a normal distribution with mean distributed according to another normal distribution is known to retain the normal shape, this is not the case when compounding over the variance parameter, leading to substantially more complex distributions. For example, the *variance gamma* distribution results from the compounding of a normal distribution with a gamma–distributed variance, and it is often used for daily stock market returns [e.g. 32]. Similarly, a scaled t distribution results from the compounding with another empirical distribution for the variance [40]. A recent review of theory and applications for the compounding of two normal variable is provided by [15].

This paper is concerned with the case of a random variable $Y \sim N(0, X^2)$ with variance distributed as $X \sim N(0, \sigma^2)$, a problem first studied nearly a century ago by [48] and [10]. The distribution of Y can also be described as the compounding of a normal variable with a χ^2_1-distributed variance. [1] The renewed interest in this compounded distribution is motivated by the occurrence of this variable in a model of systematic errors for the physical sciences that was recently proposed by the author [5–7]. The Y variable also admits the equivalent representation $Y = \sigma Z_1 Z_2$, where Z_1 and Z_2 are two independent standard normal variables [10]. This variable is distributed according to the Bessel distribution [e.g. 34], and it is a special case of a family of generalized Laplace distributions [for a review, see 29].

The goal of this paper is twofold. First, it presents a review of the properties of the Bessel random variable, with emphasis on equivalent representations as a function of other known variables with simpler analytic properties. Second, it provides approximations to its distribution function that overcome the singularity and complexity of the Bessel function $K_0(x)$. This paper is therefore aimed at improving the use of the Bessel distribution by means of simple and statistically-motivated approximations. Such approximations are convenient for a variety of statistics and data science applications, in that analytic simplicity leads to ease of use and interpretation for the broader data science community, and to the reduction of computational resources, which is important in large-scale (e.g., machine-learning) applications.

2. A Bessel Distribution from the Compounding of Two Zero–mean Normal Distributions

2.1. Distribution Functions and Equivalent Representation

Let $Y \sim N(0, a^2X^2)$ and $X \sim N(0, \sigma^2)$, with a^2 and σ^2 two non–negative constants, be two independent normal random variables. Conditionally on $X = x$, Y has a normal distribution with zero mean and variance a^2x^2. The probability density function (pdf) of Y is given by conditioning on the values of the X variable,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y/X}(y) \cdot f_X(x)\, dx, \tag{1}$$

where $f_{Y/X}(y)$ and $f_X(x)$ are respectively the conditional distribution of Y given X (hereafter represented as Y/X) and that of X. The distribution of the Y variable is said to result from from the compounding or mixing of two normal distributions. For a textbook reference of compounding distributions, see e.g. Sec. 5.13 of [27] or Sec. 4.3 of [36]. The Y variable admits the equivalent representation

$$Y = a\sigma Z_1 Z_2 \tag{2}$$

where Z_1 and Z_2 are two standard normal distributions, as used by [10], see also Lemma 4 in Appendix 6. Given that the parameters a and σ are completely degenerate, one may set $a = 1$ without loss of generality, and the random variable will depend only on the parameter σ. This compound random variable has therefore the same distribution as described by [10]; for a more recent review of the product of two random variables, see e.g. [44].

The probability density function of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{\pi\sigma}\int_0^{\infty} e^{-\dfrac{y^2}{2x^2}} \cdot e^{-\dfrac{x^2}{2\sigma^2}} \cdot \frac{dx}{x} \\ &= \frac{1}{\pi\sigma} K_0\left(\left|\frac{y}{\sigma}\right|\right) \text{ for } y \in \mathbb{R} \end{aligned} \tag{3}$$

(see Lemma 3 and Corollary 2 in App. 6). The integral is of a form than can be described via modified Bessel function of the second kind, [e.g., see 3.478.4 in 17]. This distribution is referred to as a *Bessel distribution* (of order zero), and it is referred to as $K(\sigma)$, where $\sigma > 0$ will be referred to as the *scale parameter* of the Bessel distribution. This distribution was obtained by [48] and [10] using the product of two standard normal distributions. The distribution in Eq. 3 is a special case of the family of Bessel distributions proposed by McKay [34], for $c = m = 0$ in their Eq. 1, and it is used in a variety of applications [e.g. 16, 30, 35]. As noted already, the Bessel distribution is also a special case of a variance-gamma distribution that is used in business applications [e.g. 32].

The distribution is illustrated in Fig. 1; the singularity at $y = 0$ is one of the reasons that make this distribution notoriously difficult to simulate [e.g. 44, 49]. The reason for this singularity can be seen as the compounding of the variance of the Y variable with a normal distribution with fixed variance σ^2. For small random values of X, the associated variance of Y is correspondingly small, leading to an unbounded peak in the density at $y = 0$. There is therefore practical interest to provide statistically motivated approximations to this distribution that have simpler analytic properties.

It is also possible to give the cumulative distribution function (CDF) of a $Y \sim K(\sigma)$ via modified Struve functions $L_\nu(x)$ as

$$\begin{aligned} F_Y(y) =& \frac{1}{2} + \frac{y}{2\sigma}\left(K_0\left(\left|\frac{y}{\sigma}\right|\right) L_{-1}\left(\left|\frac{y}{\sigma}\right|\right) + \right. \\ & \left. K_1\left(\left|\frac{y}{\sigma}\right|\right) L_0\left(\left|\frac{y}{\sigma}\right|\right)\right) \text{ for } y \in \mathbb{R}. \end{aligned} \tag{4}$$

Proof. The result is obtained by use of the indefinite integral

$$\int K_0(x)dx = \frac{\pi x}{x}(K_0(x)L_{-1}(x) + K_1(x)L_0(x)) \tag{5}$$

which is an immediate consequence of integral 10.43.2 in [38], using $K_{-1}(x) = K_1(x)$ and with elementary integral calculus. □

This closed form for the CDF is particularly convenient in numerical applications, in that it enables the use of available numerical evaluations for the modified Struve and

[1] Given that the chi-squared distribution belongs to the gamma family, the Bessel distribution is therefore also a special case of a variance–gamma distribution.

Bessel functions directly, without the need for further ad-hoc numerical integration of the function $K_0(x)$, which can be computationally more challenging, in particular with respect to its singularity. [2]

2.2. Approximation via a Laplace Variable

Although a number of numerical approximations to the Bessel function, and therefore to the Bessel distribution, are available in the literature [e.g. 33, 44], it is interesting to investigate an approximation that is both numerically accurate and statistically motivated. For this purpose, this paper explores approximations of the Bessel distribution via the class of Laplace distributions and that of generalized Laplace distributions [e.g. 29].

The Laplace distribution, also known as double exponential distribution, has density

$$f_L(x) = \frac{1}{2s} e^{-\dfrac{|x-\theta|}{s}} \quad \text{for } x \in \mathbb{R} \tag{6}$$

with $s > 0$ a scale parameter, and θ a location parameter that will be hereafter set to zero for this application. In the notation of [29], this is referred to as a *classical* Laplace distribution $\mathrm{CL}(\theta, s)$, and this notation will be used throughout. The Laplace distribution of Eq. 6 has mean θ and variance $2s^2$. [3] Its characteristic function (hereafter ch. f.) is

$$\psi_L(t) = \frac{1}{1 + s^2 t^2} \quad \text{for } t \in \mathbb{R}. \tag{7}$$

A random variable $X \sim \mathrm{CL}(0, s)$ admits the representation

$$X = s\sqrt{2W} Z \tag{8}$$

where $W \sim \mathrm{Exp}(\alpha = 1)$ is a standard exponential with rate $\alpha = 1$, and Z the usual standard normal variable; for a proof, see Proposition 2.3.1 of [29]. Since the gamma distribution is a scale family, whereas $\forall c \in \mathbb{R}$ and with $T \sim \mathrm{gamma}(\alpha, r)$, the random variable cT retains the same shape parameter r and has rate parameter α/c, $cT \sim \mathrm{gamma}({}^{\alpha}/_{c}, r)$. An exponential distribution is a gamma distribution with shape parameter $r = 1$, thus $W \sim \mathrm{gamma}(\alpha = 1, r = 1)$. Using the scale property of the gamma family of distributions, then $2W := S_2 \sim \mathrm{gamma}({}^{1}/_{2}, 1) \sim \chi^2_2$. Thus,

$$X = s\sqrt{S_2} Z. \tag{9}$$

On the other hand, the representation of a Bessel variable $Y \sim N(0, X^2)$ with $X \sim N(0, \sigma^2)$, i.e. $Y \sim K(\sigma)$ as in Eq. 2, is

$$Y = \sigma\sqrt{S_1} Z \tag{10}$$

where $S_1 \sim \chi^2_1$, and Z is an independent standard normal.

Remark 1. A comparison between the two representations in Eqns. 9 and 10 suggests an approximation to the Bessel distribution with a Laplace distribution. The approximation consists of replacing the assumption of a χ^2_1–distributed variance for the Bessel distribution with a χ^2_2–distributed variance for the Laplace distribution.

2.3. Representation as a Generalized Laplace Variables

A class of generalized asymmetric Laplace (GAL) distributions is defined by the ch. f.

$$\psi_{\mathrm{GAL}}(t) = \frac{e^{i\theta t}}{\left(1 + \dfrac{\sigma^2 t^2}{2} - i\mu t)\right)^{\tau}} \tag{11}$$

with $t \in \mathbb{R}$, and it is referred to as $\mathrm{GAL}(\theta, \mu, \sigma, \tau)$ in [29]. In this paper, it will always be true that $\theta = \mu = 0$, and therefore the distribution will be symmetrical about the origin, same as the classical Laplace.

A key property of the generalized Laplace distribution is that its probability distribution function can be described via Bessel functions of the second type and of order $\tau - {}^{1}/_{2}$ [see Eq. 4.1.30 of 29]. For $\tau = {}^{1}/_{2}$, the pdf is

$$f_{\mathrm{GAL}}(x) = \frac{1}{\pi(\sigma/\sqrt{2})} K_0\left(\frac{|x|}{\sigma/\sqrt{2}}\right) \tag{12}$$

which is the same as Eq. 3 when $s = \sigma/\sqrt{2}$ is used. It is therefore clear that a $\mathrm{GAL}(0, 0, \sigma/\sqrt{2}, {}^{1}/_{2})$ random variable has the same distribution as the Bessel variable $K(\sigma)$. On the other hand, when the order of the Bessel function is semi-integer ($r + {}^{1}/_{2}$ with $r \in \mathbb{N}$), the Bessel functions of the second kind are known to have a closed form for its pdf with elementary functions (see App. 1.1). This property is used when considering the sum and average of Laplace variables in Sec. 3.

The same result can be obtained using the fact that a random variable $U \sim \mathrm{GAL}(0, \mu, \sigma, \tau)$ admits the representation

$$U = \mu W + \sigma\sqrt{W} Z \tag{13}$$

where W is a gamma random variable with rate $\alpha = 1$ and shape $r = \tau$, $W \sim \mathrm{gamma}(1, \tau)$ [for a proof, see Sec. 4.1.2 of 29].

Remark 2. The representation of Eq. 13 shows that a general GAL variable is equivalent to the variance gamma family of distributions introduced by [32], as also pointed out by [29].

This paper is concerned with a zero–mean variable, $\mu = 0$, thus $U = \sigma\sqrt{W} Z$. Therefore, compounding a standard normal with a $\mathrm{gamma}(1, \tau)$ distributed variance leads to a Bessel distribution with parameter $\sigma/\sqrt{2}$, or equivalently a $\mathrm{GAL}(0, 0, \sigma, \tau)$. According to the scale property of the gamma distribution, $2W = S_1 \sim \chi^2_1$. Thus, a $U \sim \mathrm{GAL}(0, 0, \sigma, \tau =$

[2] In `python`, the `modstruve` and `kn` functions in `scipy` lead to order-of-magnitude reduction in computational speed for the evaluation of $F_Y(y)$, compared to a direct numerical integration of the `kn` function with, e.g., `quad` or similar routines.

[3] In the notation of [29], this distribution is referred to as the classical Laplace distribution, whereas a *standard* Laplace distribution is a re-parameterization of Eq. 6 with $s = \sigma/\sqrt{2}$.

1/2) variable admits the representation

$$U = \left(\frac{\sigma}{\sqrt{2}}\right)\sqrt{S_1}Z, \tag{14}$$

which is the same as the usual representation (Eq. 10) of the Bessel variable $K(\sigma/2)$.

Remark 3. The representation in Eq. 14 shows how the Bessel distribution $K(\sigma)$ can be obtained as a special case of the family of generalized Laplace distributions $\mathrm{GAL}(0, 0, \sigma/\sqrt{2}, 1/2)$.

2.4. Properties of the Zero–order Bessel Distribution

Key properties of the Bessel distribution are summarized in this section.

Property 1. The ch. f. of a $Y \sim K(\sigma)$ random variable is

$$\psi_K(t) = \frac{1}{\sqrt{1+\dfrac{\sigma^2 t^2}{2}}},\ t \in \mathbb{R} \tag{15}$$

and the moment generating function(m.g.f.) is

$$M_K(t) = \frac{1}{\sqrt{1-\dfrac{\sigma^2 t^2}{2}}},\ \text{ for } -\frac{2}{\sigma^2} < t < \frac{2}{\sigma^2}. \tag{16}$$

Proof. This is an immediate consequence of the fact that $Y \sim \mathrm{GAL}(0, 0, \sqrt{2}\sigma, 1/2)$ and the ch. f. of the generalized Laplace distribution in Eq. 11. An equivalent proof is also provided in Corollary 3 in Appendix 6. The m.g.f. is then simply obtained as $M_K(t) = \psi_K(t/i)$ [see, e.g., Eq. 5.1.12 of 47]. □

Moments of $Y \sim K(\sigma)$ can also be given by means of the following integral, without use of the m.g.f.:

$$\int_0^\infty t^{\mu-1}K_0(t)dt = 2^{\mu-1}\Gamma(\mu/2)^2, \tag{17}$$

which is a special case of integral 10.53.19 in [38].

Property 2 (Mean and variance of the zero–order Bessel distribution). The mean is $\mathrm{E}[Y] = 0$ and the variance is $\mathrm{Var}(Y) = \sigma^2$.

Proof. A proof follows immediately from the m.g.f. of Y given in Eq. 16. Alternatively, a proof can be given using the compounding representation of Y. Indicate with G the normal distribution of $X \sim N(0, \sigma^2)$ and with H the normal distribution of $Y/X \sim N(0, X^2)$. The mean and variance of Y are therefore

$$\begin{cases} \mathrm{E}[Y] = \mathrm{E}_G[E_H[Y/X]] = 0 \\ \mathrm{Var}(Y) = E_G[\mathrm{Var}_H(Y/X)] + \mathrm{Var}_G(\mathrm{E}_H[Y/X]) \\ = \sigma^2, \end{cases} \tag{18}$$

where the usual conditional variance formula was used [e.g., Sec. 4.3 of 36]. □

Now, with $Y \sim K(\sigma)$ and $X \sim CL(0, s)$, and since $\mathrm{Var}(Y) = \sigma^2$ and $\mathrm{Var}(X) = 2s^2$, it is clear that the use of $s = \sigma/\sqrt{2}$ would results in an approximating Laplace distribution with same variance as the given Bessel distribution $K(\sigma)$. In applications where σ is a parameter of physical interest, such as the model of systematic errors in [6, 7], it is therefore reasonable to parameterize the Laplace distribution as $K(\sigma/\lambda)$ with $\lambda = \sqrt{2}$. Further discussion of this approximation and methods to estimate the best scale factor for the approximating Laplace distribution are presented in Sec. 4.

Property 3. The σ parameter of a $K(\sigma)$ distribution is a genuine scale parameter.

Proof. This property can be immediately seen from the form of Eq. 4 for the CDF, which depends only on the ratio y/σ. □

The same scale property is also shared by the Laplace distribution $\mathrm{CL}(0, \sigma)$.

Property 4 (Laplace variable as sum of two Bessel variables). Let Y_1, Y_2 be two independent and identically distributed (i.i.d.) $K(\sigma)$. Then $X = Y_1 + Y_2$ is a Laplace variable $\mathrm{CL}(0, \sigma/\sqrt{2})$.

Proof. The proof is an immediate consequence of the ch. f. of a Bessel distribution in Eq. 15, and of the fact that a Bessel distribution is a $\mathrm{GAL}(0, 0, \sqrt{2}\sigma, 1/2)$ with ch. f. as in Eq. 7. □

Remark 4. Properties 3 and 4 further underline the relationship between the Laplace and Bessel distributions. In particular, the singularity at the origin of the pdf for a Bessel variable disappears when two identical Bessel variables are summed, resulting in a Laplace variable which has a bounded pdf at the origin. This result can be clearly seen as the first step towards the "normalization" of the sum of Bessel variables, according to the central limit theorem, when many i.i.d. variables are summed.

3. Sum and Average of Laplace Variables

Given the approximation of a Bessel distribution with a Laplace that was suggested in Sec. 2, it is further investigated whether it is possible to approximate a Bessel distribution with the sum of Laplace variables. The characteristic functions of the two families of distributions already shows that this will not be possible in an exact sense. In fact, as also discussed in Prop. 4, it is the Laplace distribution that can be obtained as the sum of two Bessel variables, and not vice-versa. Nonetheless, it is useful to present key results about the sum and average of Laplace variables, which can be of interest in certain associated data science applications.

First the result provided by [29] for the sum of standard classical Laplace variables is generalized in the following lemma.

Lemma 1. The sum of n i.i.d. standard Laplace random variables $Y_i \sim \mathrm{CL}(0, s)$ has the representation

$$\sum_{i=1}^{n} Y_i = \sqrt{2} s \sqrt{G_n} Z \tag{19}$$

where $G_n \sim \mathrm{gamma}(1, n)$ is a gamma distribution with rate parameter 1 and shape parameter n.

Proof. First, the ch. f. of $Y_i \sim \mathrm{CL}(0, s)$ is

$$\psi_L(t) = \frac{1}{1 + s^2 t^2}$$

(see, e.g., 2.4.2 in [29] and Eq. 7), and the ch. f. for the sum of n identical copies is

$$\psi(t) = \left(\frac{1}{1 + s^2 t^2} \right)^n .$$

According to Lemma 6, the characteristic function for the compounding of a normal variable Z with a variance distributed like $s^2\, U$, with s a constant, is

$$\psi(t) = M_U \left(\frac{s^2 t^2}{2} \right) \tag{20}$$

where M_U is the m.g.f. of the U distribution. Since the m.g.f. of a $\mathrm{gamma}(1, n)$ variable is

$$M_\gamma(t) = \left(\frac{1}{1 - t} \right)^n$$

it follows that Eq. 20 is satisfied by $U = \mathrm{gamma}(1, n)$ for a Laplace parameter $\sqrt{2}s$, and therefore

$$\sum_{i=1}^{n} Y_i = \sqrt{2}\, s\, \sqrt{G_n} Z.$$

which extends the proof of [29], see their Proposition 2.2.10, to any zero–mean Laplace distribution with parameter s. □

Remark 5. Lemma 1 shows that the sum of Laplace variables with parameter s is equal to the compounding of a normal variable with variance that is distributed as $2s^2 G_n$, with $G_n \sim \mathrm{gamma}(1, n)$. On the other hand, the Bessel variable in Eq. 10 has a variance that is distributed as $\sigma^2 S_1$, with $S_1 \sim \chi^2_1 = \mathrm{gamma}(1/2, 2)$. It is therefore not possible to obtain a Bessel variable (of order zero) as the sum of any number of classical Laplace variables, as already noted.

A more useful representation of the sum of Laplace variables, and one that has a simple analytic density, makes use of the representation of a Laplace variable $Y_i \sim \mathrm{CL}(0, s)$ as the difference of two standard exponential variables,

$$Y_i = s(W_1 - W_2) \tag{21}$$

where W_1 and W_2 are two iid standard exponential variables [Proposition 2.2.1 ff., 29]. The representation of the average of n standard Laplace variables is provided by Eq. 2.3.24 of [29], which is report in the following with a generalization to the sum of zero–mean Laplace variables with parameter s.

Lemma 2. Let $Y_i \sim \mathrm{CL}(0, s)$, $i = 1, \dots, n$ be n i.i.d. Laplace variables. Then,

$$T = n\overline{Y}_n = \sum_{i=1}^{n} Y_i = s(G_1 - G_2) \tag{22}$$

where G_1 and G_2 are two i.i.d. $\mathrm{gamma}(1, n)$ variables.

Proof. This is an immediate consequence of Eq. 21 and the closedness of the gamma family of distributions with respect to the shape parameter. □

Corollary 1. The density of T is

$$f_T(y) = \frac{(|y|/(2s))^{n-1/2}}{\Gamma(n)\sqrt{\pi} s} K_{n-1/2} \left(\frac{|y|}{s} \right). \tag{23}$$

Proof. Another representation of a $U \sim \mathrm{GAL}(0, \mu, \sigma, \tau)$ is

$$U = \frac{\sigma}{\sqrt{2}} \left(\frac{1}{\kappa}\, G_1 - \kappa\, G_2 \right), \tag{24}$$

with G_1, G_2 two iid $\mathrm{gamma}(1, \tau)$ variables, see Proposition 4.1.3 of [29], and the associated $\kappa(\sigma)$ parameter simply becomes $\kappa = 1$ for our application with $\mu = 0$. Now, this representation is such that $T \sim \mathrm{GAL}(0, 0, \sigma = \sqrt{2}s, \tau = n)$. Accordingly, the density is given by Eq. 23, which generalizes the density of the sum of n iid Laplace variables in the presence of the $s \neq 1$ parameter [see Eq. 2.3.25, 29]. □

The results of Lemma 2 and Corollary 1 lead to a simple representation of the density of the sum and average of Laplace variables. In fact, with Eq. 23 containing a Bessel function of second kind and of semi–integer order, the density can be written as a finite sum of elementary functions, see Eq. 36 in App. 1.1, as:

$$\begin{aligned} f_T(y) =& \frac{(|y|/(2s))^{n-1}}{2s\,\Gamma(n)} e^{-|y|/s}. \\ & \sum_{k=0}^{n-1} \frac{(n-1+k)!}{(n-1-k)!k!} \left(\frac{|y|}{s} \right)^{-k} \frac{1}{2^k}. \end{aligned} \tag{25}$$

The distribution of $\overline{X}_n$ for the average is immediately related to that of the sum T via

$$f_{\overline{Y}_n}(y) = n\, f_T(ny),\, \forall y \in \mathbb{R},$$

leading to simple analytic forms for the density of the average of n Laplace variables with the same parameter s. The first four densities are reported in Table 1, which generalize the results of [29]. It is clear that $\mathrm{E}[\overline{Y}_n] = 0$ and $\mathrm{Var}(\overline{Y}_n) = 2s^2/n$. The distributions of $\overline{Y}_n$ for $n = 1$ through 4 are shown in Figure 1, where for a given n we chose Laplace distributions $\mathrm{CL}(0, s\sqrt{n})$, i.e., with an increasing variance such that asymptotically $\overline{Y}_n \overset{a}{\sim} N(0, 2s^2)$, according to the central limit theorem (see green curve in the figure). Of course, the average of n Laplace distributions with fixed scale parameter s and variance $2s^2$ would converge to zero (i.e., per the law of large numbers), and it is not shown in the figure.

Remark 6. It is clear that, within the family of averages of n Laplace variables $\overline{Y}_n$ with same variance, the best approximation to the Bessel distribution $K(\sigma)$ is provided for the case $n = 1$, i.e., by the single Laplace variable $\mathrm{CL}(0, \sigma/\sqrt{2})$.

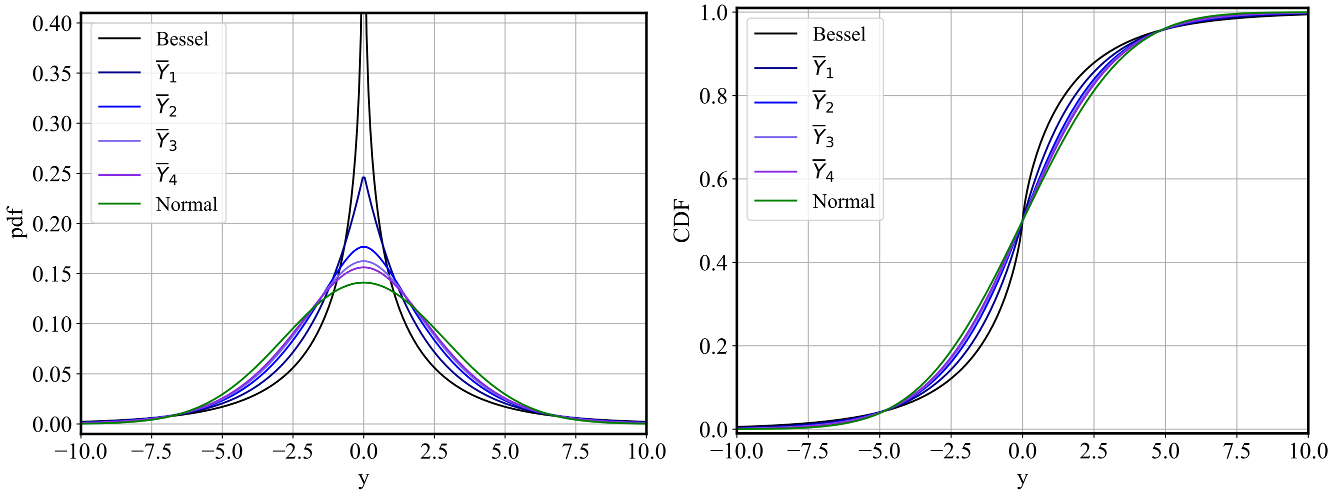

Figure 1. *Distributions of the average $\overline{Y}_n$ of n =1, 2, 3, and 4 Laplace variables $\mathrm{CL}(0, s\sqrt{n})$ with $s = 2$, the Bessel distribution $K(\sigma = s\sqrt{2})$ and the normal distribution $N(0, 2s^2)$, all of same variance.*

Table 1. *Probability density function of the average of n Laplace variables, i.e. the $\overline{Y}_n$ variable for $n = 1, 2, 3, 4$, according to Eq. 25.*

n	**Density of $\overline{Y}_n$**
1	$f(y) = \dfrac{1}{2s} e^{-\lvert y\rvert/s}$
2	$f_2(y) = \dfrac{1}{2s} e^{-2\lvert y\rvert/s} \left(1 + 2\dfrac{\lvert y\rvert}{s}\right)$
3	$f_3(y) = \dfrac{9}{16s} e^{-3\lvert y\rvert/s} \left(1 + 3\dfrac{\lvert y\rvert}{s} + 3\left(\dfrac{\lvert y\rvert}{s}\right)^2\right)$
4	$f_4(y) = \dfrac{1}{24s} e^{-4\lvert y\rvert/s} \left(15 + 60\dfrac{\lvert y\rvert}{s} + 96\left(\dfrac{\lvert y\rvert}{s}\right)^2 + 64\left(\dfrac{\lvert y\rvert}{s}\right)^3\right)$

4. Simple Approximations of the Bessel Distributions

Based on the theoretical results described in Sec. 2 and 3, simple approximations for the Bessel distribution are now discussed and tested. These approximations can be useful in data science applications where a tractable density and overall simple analytic properties are convenient.

4.1. The Normal Approximation

A normal distribution of same mean and variance as the $K(\sigma)$ is shown in Figure 2, illustrating that the normal distribution $N(0, \sigma^2)$ is generally a poor approximation to the Bessel distribution. The cross–over points between the two distributions, illustrated in Figure 2, also show that the Bessel distribution has harder tails than the normal distribution of same mean and variance. This is in fact a practical feature that makes the Bessel distribution useful in certain applications where harder–than–normal tails are required [37, 39, 41].

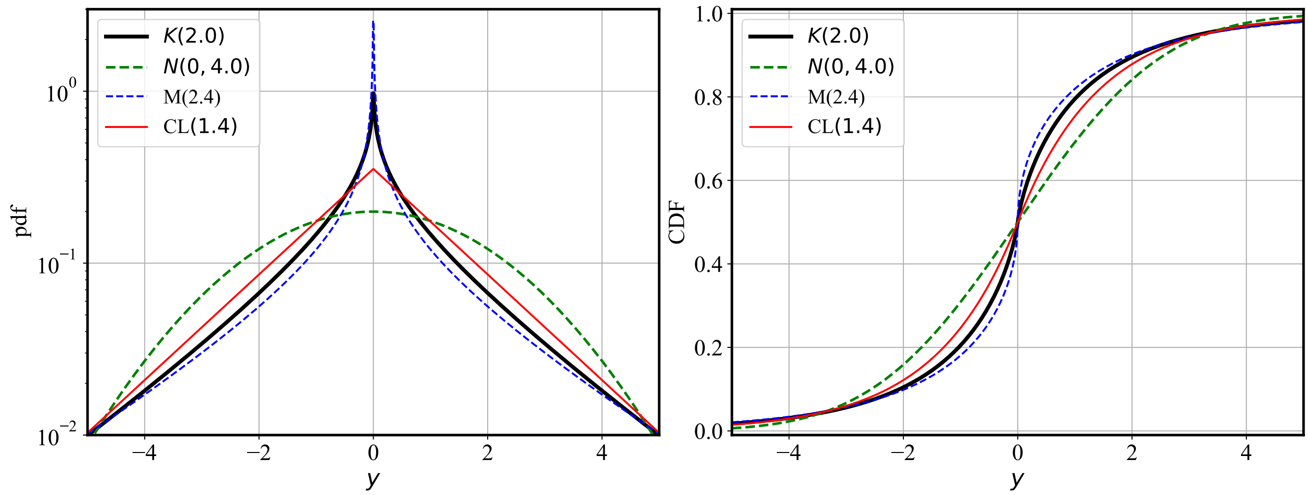

Figure 2. *Probability distribution function and CDF of the Bessel distribution $K(\sigma = 2)$ according to Eq. 3. For comparison, a Laplace distribution $CL(0, \sigma/\lambda)$ with $\lambda = \sqrt{2}$, a normal distribution $N(0, \sigma^2)$ and a distribution based on the Martin–Maas $M(s = 2.4)$ approximation are also shown.*

Remark 7. The normal distribution is equivalent to ignoring the compounding that leads to the singularity in the Bessel distribution at $y = 0$, and it is therefore not recommended as an approximation for the Bessel distribution. As also discussed in Sec. 3, a normal distribution $N(0, 2\sigma^2)$ is naturally the asymptotic distribution of the average of n iid $\mathrm{CL}(0, \sigma\sqrt{n})$ variables, according to the central limit theorem [e.g., Sec. 17.4 of 11].

4.2. The Zero–order Power–series Approximation, i.e., the Martin–Maas Distribution

It is possible to describe the modified Bessel function of the second kind and of order zero via a power series. For example, the recent work by [33] provides a power series that is suitable as an approximation; retaining the first term in their expansion, a simple approximation is provided by

$$K_0(x) \simeq \sqrt{\frac{\pi}{2x}}\, e^{-x} \quad \text{for } x \geq 0.$$

This approximation can be used in Eq. 3 using a scale parameter s, and normalized to unity to yield the following density:

$$f_M(y) = \frac{1}{2\sqrt{\pi s}} \frac{e^{-|y|/s}}{\sqrt{|y|}}, \quad \text{for } y \in \mathbb{R}. \tag{26}$$

This pdf is referred to as the *Martin–Maas* approximation to the modified Bessel distribution, and the corresponding family of distributions is indicated as $M(s)$. [4] This approximation is shown in Figure 2 as a dashed blue curve. Despite its singularity at $y = 0$, this approximation has a convenient closed form for its CDF:

$$F_M(y) = \begin{cases} \frac{1}{2}\left(1 + \mathrm{erf}(\sqrt{y/s})\right) & \text{for } y \geq 0 \\ 1 - \frac{1}{2}\left(1 + \mathrm{erf}(\sqrt{-y/s})\right) & \text{for } y < 0 \end{cases} \tag{27}$$

where $\mathrm{erf}(x)$ is the usual error function. This CDF is also illustrated in Figure 2 as a dashed blue curve. From the form of its CDF, it is evident that s is a genuine scale parameter for this family of distributions.

The m.g.f. for the distribution is

$$M_M(t) = \frac{1}{2\sqrt{s}}\left(\frac{1}{\sqrt{t + 1/s}} + \frac{1}{\sqrt{1/s - t}}\right), \tag{28}$$

which can be easily obtained via direct integration of $\mathrm{E}[e^{tY}]$ with the aid of elementary functions. The m.g.f. shows that $\mathrm{E}[X] = 0$, as expected, and $\mathrm{Var}(X) = {}^3\!/_2\, s^2$. Compared to a Bessel distribution $K(s)$ of same parameter, the Martin–Maas distribution $M(s)$ has a larger variance for the same scale parameter.

4.3. The Laplace or Double–exponential Approximation

An approximation that overcomes the singularity at the origin is provided by the Laplace distribution, as discussed in Secs. 2.2 and 3. For the purpose of applications, it may be convenient to re-parameterize the $CL(0, \sigma)$ distributions as $CL(0, \sigma/\lambda)$, i.e., with a density

$$f_L(x) = \frac{\lambda}{2\sigma} e^{-\dfrac{\lambda\,|x|}{\sigma}}, \quad \text{for } x \in \mathbb{R} \tag{29}$$

where λ is an ancillary parameter that is completely degenerate with σ. This λ parameter can be introduced to adjust the approximating Laplace distribution, when comparing it to a

[4] There appears to be no previously named distribution of this form in the literature. This distribution could also be referred to as the *inverse-square-root modified exponential* distribution.

$K(\sigma)$ Bessel distribution. The associated CDF also has a simple analytic form,

$$F_L(y) = \begin{cases} 1 - \dfrac{1}{2}\, e^{-\dfrac{\lambda}{\sigma} y} & \text{for } y \geq 0 \\ \dfrac{1}{2}\, e^{+\dfrac{\lambda}{\sigma} y} & \text{for } y < 0. \end{cases} \tag{30}$$

As illustrated in Figure 2, the Laplace distribution has the advantages of a peaked shape at the origin that mimics the Bessel distribution but without the singularity, and the flexibility of adjusting of the scale parameter to match certain properties of the Bessel distribution (see Sec. 5 for further discussion). This distribution is therefore suggested as a possible approximation to the Bessel distribution for certain applications where a simple distribution is preferable.

As discussed in Sec. 2.2, a value $\lambda = \sqrt{2}$ results in a Laplace variables with variance σ^2, same as a $K(\sigma)$ Bessel distribution it intends to approximate. The parameter λ, however, can be chosen empirically, based on the intended application. For example, the choice of $\lambda = 1.5$ was suggested in [7] for the purpose of approximating p–values of the distribution for hypothesis testing.

It is useful to determine in a more formal way a quantitative "distance" between the probability distributions of two $K(\sigma)$ and $\mathrm{CL}(0, \sigma/\lambda)$ variables, and how the parameter λ can be chosen to minimize such distance. For this purpose two distances were used: a Kolmogorov–Smirnov distance [e.g. 28] defined as

$$D_{KS} = \sup_{x \in \mathbb{R}} |F_L(x) - F_K(x)|, \tag{31}$$

and a Wasserstein distance [e.g. 46] defined as

$$D_W = \int_{\mathbb{R}} |F_L(x) - F_K(x)| dx. \tag{32}$$

Distances between distributions are commonly used to test the agreement between a sample distribution and a parent distribution, or the agreement between two sample distributions [e.g. 1, 2, 12, 28, 45]; for recent reviews, see e.g. [20, 42]. In this application, the same measures are used but for two fully specified parent distributions, with the goal of finding the parameter λ that minimizes the distance between the two distributions.

For this purpose, several values of the σ parameter in the range $\sigma = 0.1$ to 10 were considered, and a simple numerical least-squares method was used to find the best-fit value of λ that minimizes the distances D_{KS} and D_W. For the Kolmogorov-Smirnov distance, it was found that $\lambda_{KS} = 1.83$ minimizes the distance between the two distribution, for a distance $D_{KS} = 0.0253$ for all values of σ. This means that with this choice of the ancillary parameter λ, the two CDFs differ by less than approximately 0.025 for all values of the random variable. For the Wasserstein distance, it was found that $\lambda_W = 1.54$ minimizes the distance, for all values of σ. Moreover, the Wasserstein distance for this value of λ is such that $D_W/\sigma = 0.083$ for all values of σ. [5]

The findings are summarized in Figure 3, using $\sigma = 1$ as a reference and without loss of generality due to the scale-family nature of the distributions. First, the fact that the same distance parameter λ was found, and the same distance (scaled by the parameter σ for the integral Wasserstein distance) for all values of the parameter σ, are precisely a result of the fact that both the Laplace and the Bessel distributions are scale families. Second, the maximum distance between the two cumulative distributions can be kept at the percent-level for the entire range of values of the random variable.

Figure 3 also shows that the differences in the CDF of the $CL(0, \sigma/\lambda)$, compared to the target $K(\sigma)$, depend on the choice of λ. For the λ_W value obtained from the Wasserstein distance optimization, the largest differences occur near the median, and in larger magnitude than when λ_{WD} is used. However, for larger (and smaller) quantiles, λ_W results lead to a distribution that is in a better agreement, compared to the use of λ_{WD}. It is therefore suggested that the choice of λ depend on the application at hand. An application that illustrates this process of analysis is provided in the next section.

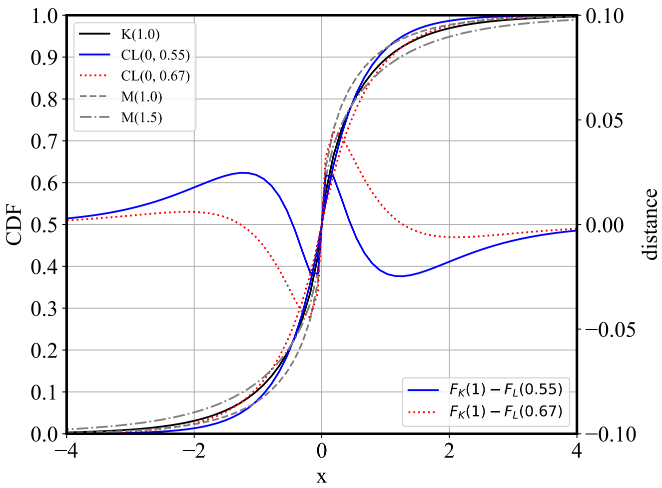

Figure 3. CDF of $K(\sigma)$ for $\sigma = 1$, and two Laplace distributions with scale parameter σ/λ, with two choices of $\lambda_W = 1.54$ and $\lambda_{KS} = 1.83$. Wasserstein and Kolmogorov-Smirnov distances are also reported in the right-hand vertical axis.

5. Application to Hypothesis Testing

The methods discussed in this paper are now used to approximate quantiles of the Bessel distribution, which are a key component of hypothesis testing. Given that hypothesis testing is a fundamental process for a variety of statistics and data science applications [e.g. 31], accurate and efficient calculation of quantiles is of broad relevance for practical

[5] There is no immediate interpretation of the Wasserstein distance in terms of difference in the values of the distributions, as is the case for the Kolmogorov-Smirnov distance. Rather, the Wasserstein distance is often referred to as the earth-mover's distance [e.g., based on the Kantorovich-Rubinstein theorem 21, 26], and it can be interpreted as the minimum amount of work required to transform one distribution into the other.

applications. For this purpose, a $K(\sigma)$ Bessel distribution with $\sigma = 1$ is used as a reference distribution, and a choice of parameters for the $\mathrm{CL}(0, \sigma/\lambda)$ Laplace distribution and for the $M(s)$ Martin-Maas distribution are used to evaluate the $p = 0.5$ (median), 0.683, 0.9, 0.95 and 0.99 quantiles, which correspond to critical values of same probability for one–sided hypothesis testing. Given that the Bessel distribution is symmetric about the origin, identical considerations are used to calculate $p \leq 0.5$ quantiles, which can be used for two–sided hypothesis testing.

Results are shown in Table 2, with some of the distributions illustrated also in Figure 3.

Table 2. *$(1-\alpha)$–quantiles, or one–sided critical values with probability α to exceed ($x = F^{-1}(1-\alpha)$) of the Bessel, Laplace and Martin–Maas distributions, for representative choices of the probability and of the parameters of the distributions. In parenthesis are the percent deviations from values of the Bessel distribution.*

α	$K(\sigma = 1)$	λ of Laplace distribution $\mathcal{L}(\sigma/\lambda)$					s of Martin–Maas distribution $M(s)$		
		1.0	$\sqrt{2}$	1.54	1.83	2	1	1.2	1.5
(prob.)		$(1-\alpha)$-quantiles or one-sided critical values with probability to exceed α							
0.500	0.00	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (-0)	0.00 (0)
0.317	0.22	0.46 (111)	0.32 (49)	0.30 (37)	0.25 (15)	0.23 (5)	0.11 (-48)	0.14 (-37)	0.17 (-21)
0.100	1.03	1.61 (56)	1.14 (10)	1.05 (1)	0.88 (-15)	0.80 (-22)	0.82 (-21)	0.99 (-5)	1.23 (19)
0.050	1.60	2.30 (44)	1.63 (2)	1.50 (-6)	1.26 (-21)	1.15 (-28)	1.35 (-15)	1.62 (2)	2.03 (27)
0.010	2.98	3.91 (31)	2.77 (-7)	2.54 (-15)	2.14 (-28)	1.96 (-34)	2.71 (-9)	3.25 (9)	4.06 (36)

It is found that for the $1 - p = \alpha = 0.1$, 0.05 and 0.01 residual probabilities, also 90%, 95% and 99% one-sided critical values, a choice of $\lambda \simeq 1.5$ provides the most accurate approximation of the quantiles when using the Laplace distribution. For the Martin–Maas distribution, a value of the scale parameter $s \simeq 1.2$ is the most appropriate to estimate these quantiles; this value corresponds to a distribution with larger variance than the approximating Bessel distribution (see Sec. 4.2). For both distributions, it is found that those quantiles can be approximated to better than 10% accuracy, compared to the reference $K(\sigma)$ distribution. The Laplace distribution has the advantage of having both pdf and CDF with simple closed forms, and without the singularity at the origin in its density, providing a computational advantage for numerical applications.

Remark 8. The analysis reported in this section shows that the Laplace distribution is a suitable approximation for the Bessel distribution for use in hypothesis testing, with a value of the ancillary parameter $\lambda \simeq 1.5$ being preferred when the goal is the estimate of large (e.g., $p \geq 0.9$) quantiles. This value is also similar to the $\lambda = \sqrt{2}$ value that results in the same variance for the two distributions (see Sec. 2.2), and to the $\lambda_W = 1.54$ value that minimizes the Wasserstein distance (see Sec. 4.3). On the other hand, for estimates of the distribution near the median, a larger value of λ might be appropriate, e.g. a value that is near $\lambda_{KS} = 1.83$ which minimizes the Kolmogorov-Smirnov distance between the two distributions.

A notable application of the Bessel distribution is for the treatment of systematic errors in regression problems with Poisson data [5, 6], which has motivated the author's interest in this distribution. In that application, the statistic that was used to assess the significance of a nested model component [e.g. 8, 9] is obtained as the convolution of a χ^2_ν distribution with a $K(\sigma)$ distribution. When the Bessel distribution is approximated with a Laplace, it is in fact possible to provide a closed form for the pdf of the statistic under the null hypothesis, with advantages in both computational time and ease of interpretation of the regression [7]. There is therefore a significant incentive for the use of a simple approximation to the Bessel distribution, such as the one provided by the Laplace distribution. The results of this paper show that a choice of $\lambda \simeq 1.5$ for the approximating Laplace distribution is reasonable, as was in fact suggested in that application.

6. Conclusions

The Bessel distribution results from the compounding of two zero-mean normal distributions, and it is therefore of use in a variety of statistics applications [e.g. 7, 16, 30, 35]. One of the limitations to the use of the Bessel distribution for practical data science applications is the singularity at the origin, and accordingly a variety of approximations have been proposed in the literature [for recent results, see e.g. 33, 44]. This paper has reviewed key properties of the Bessel distribution, presented a simple closed form for its CDF via modified Bessel and Struve functions, and investigated and tested approximations via the Laplace distribution (also known as double-exponential) and with other distributions.

This paper has shown that a suitable and statistically-motivated approximation to a Bessel variable $K(\sigma)$ is the Laplace variable $\mathrm{CL}(0, \sigma/\sqrt{2})$, also referred to as the double-exponential distribution. In fact, such Laplace distribution results from the compounding of a zero-mean normal distribution with a χ^2_2-distributed variance, instead of a χ^2_1-distributed variance for the Bessel distribution. Moreover, the two distributions have the same mean and variance, and both belong to a scale-family of distributions. It was also shown that it is not possible to represent a Bessel distribution as the sum or average of any number of Laplace distributions, and therefore the single-Laplace approximation appears a simple and viable alternative with a density that only uses elementary functions.

For hypothesis testing, which is a common task in statistics and in data science applications, it was shown that the Laplace distribution can approximate the Bessel distribution with an accuracy of better than 10% for the calculation of typical quantiles. For this class of applications, it was shown that it is convenient to adjust the scale parameter of the Laplace distribution by means of an auxiliary parameter λ such that a $K(\sigma)$ variable is approximated by a $\mathrm{CL}(0, \sigma/\lambda)$ variable. To minimize the difference between the exact and the approximating quantiles, it was shown that the auxiliary parameter λ can typically be chosen as $\lambda \simeq 1.5$, which also corresponds to the minimum Wasserstein distance between the Bessel and Laplace distributions.

The closed form for both the pdf and the CDF of the Laplace distribution are also of advantage for other applications, such as the ability to evaluate analytically certain convolution integrals, as in the case of a model for the goodness-of-fit statistic in the presence of systematic errors in the regression of integer-count data [e.g. 7]. Another class of applications that may benefit from these approximations is to the modeling of stock and option pricing, which was in fact one of the applications for which the variance gamma distribution was initially proposed [32]. Given that the Bessel distribution is a particular case of the variance gamma [e.g. 14, 29, 32], certain applications that use zero-mean variance-gamma distributions [e.g. 13, 24] may benefit from the approximations described in this paper. It is therefore believed that the representations and approximations of the Bessel variable presented in this paper can be of use in a variety of practical applications, where simple analytic tools are preferrable to the complexity of the Bessel function.

Appendix

Appendix I: Distributions and Useful Mathematical Formulae

A random variable is said to have a *gamma* distribution $\mathrm{gamma}(\alpha, r)$ with rate parameter α and shape parameter r if it follows the density distribution

$$f_{\gamma}(x) = \frac{\alpha^r x^{r-1}}{\Gamma(r)} e^{-\alpha x}, \tag{33}$$

and with $\mathrm{E}[X] = r/\alpha$ and $\mathrm{Var}(X) = r/\alpha^2$. A special case is the *chi–squared* χ^2_ν variable, which is gamma variable $\mathrm{gamma}(\alpha = 1/2, r = \nu/2)$ with density

$$f_{\chi^2}(x) = \frac{x^{\nu/2-1}}{\Gamma(\nu/2) 2^{\nu/2}} e^{-x/2}. \tag{34}$$

Another special case is the *exponential* variable $\mathrm{Exp}(\alpha)$, which is a gamma variable $\mathrm{gamma}(\alpha, 1)$, with density

$$f_E(x) = \alpha e^{-\alpha x}. \tag{35}$$

The Bessel function of second kind and of semi–integer order has the following representation as a sum of elementary functions:

$$K_{r+1/2}(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \sum_{k=0}^{r} \frac{(r+k)!}{(r-k)!k!} (2x)^{-k}, \tag{36}$$

(see e.g. 8.468 of [17]).

Appendix II: Review of Compounding of Distributions

This appendix provides a review of elementary methods for the compounding of two random variables. It begins with a simple general result for the distribution of the product $Z = X\,Y$ of two random variables, and then provides results for the case of normal distributions of direct relevance to this paper.

Lemma 3. Let X and Y be two continuous random variables with joint pdf $h(x, y)$. The distribution of $Z = X\,Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} h(x, z/x)\, \frac{dx}{|x|} \tag{37}$$

[see, e.g., 43, p.141].

Proof. This is immediately seen by

$$F_Z(z) = P(Z \leq z) = P(Y \leq z/x \mid X = x);$$

separating the probabilities according to the sign of x, to give

$$F_Z(z) = \int_0^{\infty} f_X(x)dx \int_0^{z/x} f_Y(y)dy + \int_{-\infty}^{0} f_X(x)dx \int_{z/x}^{-\infty} f_Y(y)dy.$$

From this, using a change of variables and the fundamental theorem of calculus, Eq. 37 is obtained. □

Lemma 3 is the most general case for the product of two continuous random variables. The application to independent variables follows.

Corollary 2. Let X and Y be independent continous random variables respectively with pdf $f_X(x)$ and $f_Y(y)$. Then the distribution function of $Z = X\,Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z/x)\, \frac{dx}{|x|} \tag{38}$$

Proof. This is an immediate consequence of Lemma 3 and the independence of random variables. □

The compounding of two random variables are zero–mean normal distributions can be described as the product of two random variables. This notation makes for an easy interpretation of the operation of compounding, and it is presented as the following lemma.

Lemma 4. Let $Y \sim N(0, X^2)$ with $X \sim N(0, \sigma^2)$, with σ^2 a non–negative number, i.e., the Y variable is the result of the compounding of two zero–mean normal variables. Then $Y = \sigma Z_1 Z_2$ is an equivalent representation for the compounded distribution, with Z_1 and Z_2 two independent standard normal variables.

Proof. Conditional on $X = x$, the distribution of Y is

$$f_{Y/X}(y) = \frac{1}{\sqrt{2\pi x^2}} e^{-y^2/(2x^2)}.$$

According to the definition of a compound distribution, the marginal distribution of Y is the integral of the conditional distribution times the distribution of X, which can be thought of as a prior:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y/X}(y) f_X(x) dx = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-y^2/(2x^2)} \right) f_X(x) \frac{dx}{|x|}.$$

Now, the first integrand is $f_Z(y/x)$, where Z is a standard normal variable, and $X = \sigma Z$ is the usual representation of X as a function of another (independent) standard normal distribution. According to Eq. 38, the variable Y is therefore distributed as the product of Z_1 and of an independent σZ_2 variable. □

Lemma 4 lets us refer to the compounded variable Y as the product $Y = \sigma Z_1 Z_2$ of two zero–mean normal distributions, without loss of generality. It will be noticed that in Lemma 4 the assumption of normality for X was never used. This implies that a more general result can be stated as follows.

Lemma 5. Let $Y \sim N(0, \sigma^2 U)$ with U a random variable and σ^2 a constant. Then the random variable Y admits the equivalent representation

$$Y = \sigma \sqrt{U}\, Z \tag{39}$$

where Z is a standard normal variable.

Proof. The proof is a direct consequence of Corollary 2, and the conditional normal distribution of Y/σ, following the same steps as Lemma 4. □

It is also possible to give a general expression for the characteristic function of a compounded distribution of the type of Eq. 39, following [25].

Lemma 6. Let $Y \sim N(0, \sigma^2 U)$ with U a random variable and σ^2 a constant. Then the random variable Y has the characteristic function

$$\psi_Y(t) = E_U[e^{-(U^2\sigma^2 t^2)/2}] = M_U\left(\frac{\sigma^2 t^2}{2}\right), \tag{40}$$

where the expectation is with respect to the U distribution, and $M_U(t)$ is the moment generating function of the U distribution.

Proof. The proof of the first equality is a simple exercise (see, e.g., [25]), and it is based on the ch. f. of a zero–mean normal variable,

$$\psi(t) = e^{-(\sigma^2 t^2)/2}. \tag{41}$$

The second equality follows from the definition of the ch. f.. □

Corollary 3. Let $Y \sim N(0, X^2)$ with $X \sim N(0, \sigma^2)$, with σ^2 a non–negative number. Then $Y = \sigma Z_1 Z_2$, and its characteristic function is

$$\psi_Y(t) = \frac{1}{\sqrt{1 + \sigma^2 t^2}} \tag{42}$$

Proof. This is a consequence of Lemma 6, and the simple proof is provided in [25]. See also [29], Sec. 4.1.1, for the characteristic function of the generalized asymmetric Laplace distribution, of which the Bessel–distributed Y is a special case. □

Another useful consequence of Lemma 6 is a simple proof of the algebraic property that $Y = UZ = \sqrt{a}\sqrt{U/a}\, Z$ applies to the representation of a zero–mean normal variable Z with variance distributed according to U.

Corollary 4. A random variable $Y \sim N(0, U)$ is equivalent to the compounding of $Y \sim N(0, aV)$ with $V = U/a$, where $a > 0$ is a constant, i.e., the Y variable can be equivalently represented as

$$Y = \sqrt{U} Z \iff Y = \sqrt{a}\sqrt{\frac{U}{a}}\, Z, \text{ with } a > 0, \tag{43}$$

where Z is a standard normal random variable.

Proof. The proof is an immediate consequence of the characteristic function of Lemma 6, which depends on the U variable and on the variance σ^2 only via the product $U^2\sigma^2$. □

ORCID

0000-0002-8597-9742 (Massimiliano Bonamente)

Author Contributions

Massimiliano Bonamente is the sole author. The author read and approved the final manuscript.

Funding

This work was supported in part by the NASA Astrophysics Data Analysis Program (ADAP2018B) grant *Closing the Gap on the Missing Baryons at Low Redshift with multi–wavelength observations of the Warm–Hot Intergalactic Medium*, and by the NSF EPSCoR grant *WP-SEEDS: Workshop and Partnership for Southern Excellence and Education in Data Science*, awarded to the University of Alabama in Huntsville.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Anderson TW (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics*, 33(3): 1148-1159.

[2] Anderson TW, Darling DA (1952). Asymptotic Theory ofCertain "GoodnessofFit" Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, 23(2): 193-212.

[3] Baum LE, Petrie T, Soules GW, Weiss N (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Annals of Mathematical Statistics*, 41: 164-171.

[4] Bliss CI, Fisher RA (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2): 176-200.

[5] Bonamente M (2023). Systematic errors in the maximum-likelihood regression of Poisson count data: introducing the overdispersed χ^2 distribution. *Monthly Notices of the Royal Astronomical Society*, 522(2): 1987-2001.

[6] Bonamente M, Chen Y, Zimmerman D (2025). Maximum-likelihood Regression with Systematic Errors for Astronomy and the Physical Sciences. I. Methodology and Goodness-of-fit Statistic of Poisson Data. *The Astrophysical Journal*, 980(1): 139.

[7] Bonamente M, Zimmerman D, Chen Y (2025). Maximum-likelihood Regression with Systematic Errors for Astronomy and the Physical Sciences. II. Hypothesis Testing of Nested Model Components for Poisson Data. *The Astrophysical Journal*, 980(1): 140.

[8] Cash W (1976). Generation of Confidence Intervals for Model Parameters in X-ray Astronomy. *Astronomy and Astrophysics*, 52: 307.

[9] Cash W (1979). Parameter estimation in astronomy through application of the likelihood ratio. *Astrophys. J.*, 228: 939.

[10] Craig CC (1936). On the Frequency Function of xy. *Annals of Mathematical Statistics*, 7: 1-15.

[11] Cramer H (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton.

[12] Darling DA (1957). The Kolmogorov-Smirnov, Cramér-von Mises Tests. The Annals of Mathematical Statistics, 28(4): 823-838.

[13] Drahokoupil J (2020). Variance Gamma process in the option pricing model. FFA Working Papers 3.002, Prague University of Economics and Business.

[14] Fischer A, Gaunt RE, Sarantsev A (2023). The variance-gamma distribution: A review. arXiv 2303.05615.

[15] Gaunt RE (2022). The basic distributional theory for the product of zero mean correlated normal random variables. Statistica Neerlandica, 76(4): 450-470.

[16] Górska K, Horzela A, Jankov Masirević D, Pogány TK (2022). Observations on the McKay Iv Bessel distribution. Journal of Mathematical Analysis and Applications, 516(1): 126481.

[17] Gradshteyn IS, Ryzhik IM (2007). Table of integrals, series, and products. Elsevier/Academic Press, Amsterdam, seventh edition. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

[18] Greenwood M, Yule GU (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83(2): 255-279.

[19] Gurland J (1958). A generalized class of contagious distributions. Biometrics, 14(2): 229-249.

[20] Hallin M, Mordant G, Segers J (2021). Multivariate goodness-of-fit tests based on wasserstein distance. arXiv 2003.06684.

[21] Hanin LG (1992). Kantorovich-Rubinstein Norm and Its Application in the Theory of Lipschitz Spaces. *Proceedings of the American Mathematical Society*, 115(2): 345-352.

[22] Hastie T, Tibshirani R, Friedman J (2009). The elements of statistical learning: data mining, inference and prediction. Springer, 2 edition.

[23] Hilbe JM (2011). Negative Binomial Regression. Cambridge University Press, 2 edition.

[24] Hoyyi A, Rosadi, D (2021). Daily stock prices prediction using variance gamma model. J. Math. Comput. Sci., 11: 1888-1903.

[25] Jiang X, Nadarajah S (2019). On characteristic functions of products of two random variables. *Wireless Personal Communications*, 108: 1157-1177.

[26] Kantorovich LV, Rubinshtein S (1958). On a space of totally additive functions. Vestnik of the St. Petersburg University: Mathematics, 13(7): 52-59.

[27] Kendall M, Stuart A (1977). The Advanced Theory of Statistics. Vol.1: Distribution Theory. London: Griffin, 1977, 4th ed.

[28] Kolmogorov A (1933). Sulla determinazione empirica di una legge di distribuzione. Giornale dell' Istituto Italiano degli Attuari, 4: 1-11.

[29] Kotz S, Kozubowski T, Podgorski K (2001). The Laplace Distribution and Generalizations. *Birkaeuser*.

[30] Laha R (1954). On some properties of the bessel function distributions. Bull. Calcutta Math. Soc., 46: 59-72.

[31] Lehmann E, Romano JP (2022). Testing Statistical Hypotheses. *Springer*, Cham.

[32] Madan DB, Seneta E (1990). The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63(4): 511-524.

[33] Martin P, Maass F (2022). Accurate analytic approximation to the Modified Bessel function of Second Kind K0(x). *Results in Physics*, 35: 105283.

[34] McKay A (1932). A Bessel function distribution. Biometrika, 39-44.

[35] McLeish DL (1982). A robust alternative to the normal distribution. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 10(2): 89-102.

[36] Mood AM, Graybill FA, Boes DC (1974). Introduction to the theory of statistics, Third Ed. McGraw Hill.

[37] Nadarajah S (2007). A Modified Bessel Distribution of the Second Kind. Statistica, 67: 405-413.

[38] Olver FWJ, Daalhuis ABO, D W Lozier BIS, Boisvert RF, Clark CW, Miller BR, et al. (2024). NIST Digital Library of Mathematical Functions. https://dlmf.nist.gov/, Release 1.2.3 of 2024-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[39] Park S, Lim J (2022). An overview of heavy-tail extensions of multivariate gaussian distribution and their relations. *Journal of Applied Statistics*, 49: 3477-3494.

[40] Praetz PD (1972). The distribution of share price changes. *The Journal of Business*, 45(1): 49-55.

[41] Progri I (2016). Generalized bessel function distributions. *Journal of Geolocation, Geo-information, and Geo-intelligence*, 2016: 20-34.

[42] Ramdas A, Trillos NG, Cuturi M (2017). On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests. Entropy, 19(2).

[43] Rohatgi V (1976). An introduction to probability theory and mathematical statistics. *Wiley*, New York.

[44] Seijas-Macías A, Oliveira A, Oliveira TA, Leiva V (2020). Approximating the distribution of the product of two normally distributed random variables. *Symmetry*, 12(8).

[45] Smirnov NV (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bull. Math. Univ. Moscou, 2: 2.

[46] Villani C (2008). Optimal Transport: Old and New. *Springer Nature*.

[47] Wilks S (1962). Mathematical Statistics, Sec. Ed. J. Wiley and Sons.

[48] Wishart J, Bartlett MS (1932). The distribution of second order moment statistics in a normal system. *Proceedings of the Cambridge Philosophical Society*, 28(4): 455.

[49] Yuan L, Kalbfleisch J (2000). On the Bessel Distribution and Related Problems. *Annals of the Institute of Statistical Mathematics*, 52(3): 438-447.