

# New Probability Distributions in Astrophysics: IX. Truncation for Exponential, Half Gaussian and Sech-Square Distributions with Application to the Galactic Height

Lorenzo Zaninetti

Physics Department, Turin, Italy

Email: l.zaninetti@alice.it

**How to cite this paper:** Zaninetti, L. (2022) New Probability Distributions in Astrophysics: IX. Truncation for Exponential, Half Gaussian and Sech-Square Distributions with Application to the Galactic Height. *International Journal of Astronomy and Astrophysics*, 12, 328-346.  
<https://doi.org/10.4236/ijaa.2022.124019>

**Received:** August 1, 2022

**Accepted:** November 7, 2022

**Published:** November 10, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

What distribution function best fits that of the stars' heights above the Galactic plane? Can truncated distributions improve the fit? In order to answer the above questions, we derive the probability density function, the distribution function, the average value, the  $r$ th moment, the median, an expression to generate random variate and the maximum likelihood estimator, for the truncated exponential, truncated half-normal and the truncated sech-square distributions. The results are applied to the galactic height for open clusters and for Gaia's stars in order to understand whether the truncated distributions are useful or not to astronomers.

## Keywords

Galaxy, Disk Galaxy, Fundamental Parameters Galaxy, Open Clusters and Associations

---

## 1. Introduction

In the field of astrophysics, it is a common practice to estimate the gradients of a given quantity, for example, the vertical galactic height, with an exponential distribution. Another two distributions that have astrophysical interest are the sech-square distribution, corresponding to an isothermal self-gravitating disk, see equation (41) in [1], equation at page 441 in [2], equations (9.8) and (14.6) in [3], equation (2.31) in [4], and the normal or Gaussian distribution, see equation (5) in [5]. In the field of probability, the truncation of a distribution is a common topic of research and we report some of the approaches on the double

truncation of usual interval,  $[0, \infty]$ , distributions. The doubly truncated exponential distribution has been analyzed in [6]. In the field of astrophysics, the following truncated distributions have been analyzed: the Pareto distribution, with application to the masses of stars and asteroids [7], the left truncated beta distribution, with application to the masses of the stars [8], the double truncated gamma distribution, with application to the masses of the stars [9], the double truncated lognormal distribution, with application to the mass of the stars [10], the double truncated Lindley distribution, with applications to the masses of the stars, to the luminosity function for galaxies, and to the photometric maximum in the distribution of galaxies [11], the double truncated generalized gamma distribution, with application to the luminosity functions for galaxies and quasars and to the average magnitude of galaxies as a function of the redshift [12], the double truncated Lindley family, with applications to the luminosity function for galaxies and quasars [13], the truncated Maxwell-Boltzmann distribution, with applications to a numerical relation between the root-mean-square speed and temperature and to a modification of the formula for the Jeans escape flux of molecules from an atmosphere [14], the relativistic Maxwell-Boltzmann distribution, with applications to the synchrotron emission in the presence of a magnetic field and to relativistic electrons [15], the truncated Weibull distribution, with applications to the masses of the stars and to the luminosity functions for galaxies and quasars [16], the truncated two-parameter Sujatha distribution [17], the gamma-Pareto distribution, with application to cosmic rays [18], and the truncated Weibull-Pareto distribution, with applications to the initial mass function for stars, the luminosity function for galaxies of the Sloan Digital Sky Survey, the luminosity function for QSO, and the photometric maximum of galaxies of the 2MASS Redshift Survey. This paper analyses in Section 2 the exponential, the half-normal and the sech-square distributions defined in the interval  $[0, \infty]$ . Section 3 is dedicated to the truncation of these distributions. Section 4 applies the results to the distribution of the vertical galactic height for both open clusters and Gaia's stars.

## 2. Usual Case

In this section, we review the following distributions: the exponential, the Half-Normal, and the Sech-square distributions. The aim is to evaluate which distribution produces the best fit in modeling the galactic heights of open clusters and Gaia's stars. The word "approximate" indicates that the result in question is not an analytical result.

### 2.1. The Exponential

A random variable  $X$  which takes values in  $[0, \infty]$  is said to be exponentially distributed if its distribution function (DF) is

$$F_e(x; b) = 1 - e^{-\frac{x}{b}}, \quad (1)$$

and the probability density function (PDF) is

$$f_e(x; b) = \frac{e^{-\frac{x}{b}}}{b}, \quad (2)$$

where  $b$  is the scale parameter, see [19] [20]. The average value or mean,  $\mu$ , is

$$\mu(b) = b, \quad (3)$$

the variance,  $\sigma^2$ , is

$$\sigma^2(b) = b^2, \quad (4)$$

and the median is at

$$\ln(2)b. \quad (5)$$

Random generation of the exponential variate  $X$  is given by

$$X : b \approx -\ln(1-R)b, \quad (6)$$

where  $R$  is the unit rectangular variate. The parameter  $b$  is the average value of the sample,  $\bar{x}$ ,

$$b = \bar{x}. \quad (7)$$

## 2.2. Half-Normal

Let  $X$  be a random variable defined on  $[0, \infty]$ ; its one-parameter *Half-Normal* PDF is

$$f_N(x; s) = \frac{e^{-\frac{x^2}{2s^2}} \sqrt{2}}{\sqrt{\pi}s}, \quad (8)$$

where  $s$  is the shape parameter [21].

Its DF is

$$F_N(x; s) = \operatorname{erf}\left(\frac{\sqrt{2}x}{2s}\right), \quad (9)$$

where  $\operatorname{erf}(x)$  is the error function, defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (10)$$

The DF has the following power series representation

$$F_N(x; s) = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{-n+\frac{1}{2}} s^{-2n-1} x^{2n+1}}{(2n+1)\sqrt{\pi}n!}. \quad (11)$$

The  $r$ th moment about the origin is

$$\mu_r'(s) = \frac{2^{\frac{r}{2}} s^r \Gamma\left(\frac{r}{2} + \frac{1}{2}\right)}{\sqrt{\pi}}, \quad (12)$$

where  $r$  is an integer and

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt, \quad (13)$$

is the gamma function, see formula (5.2.1) in [22]. The average value or mean,  $\mu$ , is

$$\mu(s) = \frac{\sqrt{2}s}{\sqrt{\pi}}, \quad (14)$$

the variance,  $\sigma^2$ , is

$$\sigma^2(s) = \frac{\sqrt{2}s}{\sqrt{\pi}}. \quad (15)$$

The skewness is

$$\text{skewness}(s) = \frac{\sqrt{2}(\pi-4)}{(\pi-2)^{\frac{3}{2}}}, \quad (16)$$

and the kurtosis

$$\text{kurtosis}(s) = \frac{3\pi^2 - 4\pi - 12}{(\pi-2)^2}. \quad (17)$$

The median does not have an analytical expression but can be expressed approximately. The Winitzki approximation for the median, see Equation (A.2), gives

$$\frac{\sqrt{7} \sqrt{\pi \left( -7\pi \ln\left(\frac{3}{4}\right) + \sqrt{49\pi^2 \ln\left(\frac{3}{4}\right)^2 - 1400\pi^2 \ln\left(\frac{3}{4}\right) + 2800\pi \ln\left(\frac{3}{4}\right) + 40000 - 200} \right)} s}{7\pi}. \quad (18)$$

The Menzel approximation for the median, see Equation (A.1), gives

$$\frac{\sqrt{-2\pi \ln\left(\frac{3}{4}\right)} s}{2}. \quad (19)$$

The median in the case of the Padé Approximant of order (4.2), see Equation (A.3), is the solution of the following approximate equation

$$\tanh\left(\frac{\sqrt{2}x(60\pi^2s^2 - \pi^2x^2 - 240s^2\pi - 40\pi x^2 + 128x^2)}{3s\sqrt{\pi}(20\pi^2s^2 + 3\pi^2x^2 - 80s^2\pi - 40\pi x^2 + 96x^2)}\right) = \frac{1}{2}. \quad (20)$$

A different method reverts the series (11), see page 16 in [23],

$$x = \frac{\sqrt{\pi}s\sqrt{2}y}{2} + \frac{\pi^{\frac{3}{2}}s\sqrt{2}y^3}{24} + \frac{7\pi^{\frac{5}{2}}s\sqrt{2}y^5}{960} + \frac{127\pi^{\frac{7}{2}}s\sqrt{2}y^7}{80640}, \quad (21)$$

where  $y$  is now the approximate DF; the median is obtained by inserting  $y = \frac{1}{2}$ .

**Table 1** reports the percent error of the four methods here implemented.

Random generation of the variate  $X$  for the Half-Normal is obtained with the Box-Muller method, in practice the FORTRAN subroutine gasdev [24], limiting ourselves to the positive values. The parameter  $b$  is obtained from the average

**Table 1.** Percent error,  $\delta$ , for approximating the median of the error function when  $s = 1$ .

Method	$\delta(\%)$
Menzel	0.335
Winitzki	0.00371
Pade	0.00065
Reverted series	0.03329

value of the sample,  $\bar{x}$ ,

$$b = \frac{\bar{x}}{\ln(2)}. \tag{22}$$

### 2.3. Sech-Square Distribution

According to [25], the vertical profile of density in our galaxy can be parameterized by the following density dependence

$$\rho(x) \propto \operatorname{sech}\left(\frac{nx}{b}\right)^2, \tag{23}$$

where  $x$  represents the vertical height,  $b$  is the scale height and  $n$  is an integer. Let  $X$  be a random variable defined on  $[0, \infty]$ ; the PDF corresponding to  $n = 1$  for the above formula is

$$f_s(x; b) = \frac{\operatorname{sech}\left(\frac{x}{b}\right)^2}{b}, \tag{24}$$

and its DF is

$$f_s(x; b) = \tanh\left(\frac{x}{b}\right), \tag{25}$$

where  $b$  is the scale. The average value or mean,  $\mu$ , is

$$\mu(b) = \ln(2)b, \tag{26}$$

the variance,  $\sigma^2$ , is

$$\sigma^2(b) = \frac{b^2(\pi^2 - 12\ln(2)^2)}{12}. \tag{27}$$

The skewness is

$$\text{skewness}(b) = -\frac{6\left(\ln(2)\pi^2 - 8\ln(2)^3 - \frac{9\zeta(3)}{2}\right)\sqrt{3}}{\left(\pi^2 - 12\ln(2)^2\right)^{\frac{3}{2}}}, \tag{28}$$

and the kurtosis

$$\text{kurtosis}(b) = \frac{\frac{21\pi^4}{5} + 72\ln(2)^2 \pi^2 - 432\ln(2)^4 - 648\ln(2)\zeta(3)}{(\pi^2 - 12\ln(2)^2)^2}, \quad (29)$$

where the zeta function is defined by

$$\zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x}, \quad (30)$$

see formula (25.2.1) in [22]. The median is at

$$\operatorname{arctanh}\left(\frac{1}{2}\right)b, \quad (31)$$

and the random generation of the sech-square distribution is obtained by solving the following non-linear equation

$$\tanh\left(\frac{x}{b}\right) = R. \quad (32)$$

The parameter  $b$  is obtained by the average value of the sample,  $\bar{x}$ ,

$$b = \frac{\bar{x}}{\ln(2)}. \quad (33)$$

### 3. Truncated Case

In this section we introduce the truncations of the following distributions: the exponential, the Half-Normal and the Sech-square distribution.

#### 3.1. The Truncated Exponential

Let  $X$  be a random variable defined in  $[x_l, x_u]$ ; the *truncated exponential* PDF,  $f_{et}(x; b, x_l, x_u)$ , is

$$f_{et}(x; b, x_l, x_u) = \frac{e^{-\frac{x}{b}}}{\left(e^{-\frac{x_l}{b}} - e^{-\frac{x_u}{b}}\right)b}, \quad (34)$$

and its DF is

$$F_{et}(x; b, x_l, x_u) = -\frac{-e^{-\frac{x_l}{b}} + e^{-\frac{x}{b}}}{e^{-\frac{x_l}{b}} - e^{-\frac{x_u}{b}}}. \quad (35)$$

The first moment about the origin  $\mu'_1$ , is

$$\mu'_1(b, x_l, x_u) = \frac{e^{-\frac{x_l}{b}}b + e^{-\frac{x_l}{b}}x_l - e^{-\frac{x_u}{b}}b - e^{-\frac{x_u}{b}}x_u}{e^{-\frac{x_l}{b}} - e^{-\frac{x_u}{b}}}, \quad (36)$$

and the second moment about the origin  $\mu'_2$ , is

$$\mu'_2(b, x_l, x_u) = \frac{(2b^2 + 2bx_l + x_l^2)e^{-\frac{x_l}{b}} - 2\left(b^2 + bx_u + \frac{1}{2}x_u^2\right)e^{-\frac{x_u}{b}}}{e^{-\frac{x_l}{b}} - e^{-\frac{x_u}{b}}}. \quad (37)$$

The variance can be evaluated with the usual formula

$$\sigma^2(b, x_l, x_u) = \mu'_2 - (\mu'_1)^2, \tag{38}$$

and the median is at

$$-\ln \left( \frac{e^{-\frac{x_l}{b}}}{2} + \frac{e^{-\frac{x_u}{b}}}{2} \right) b. \tag{39}$$

The parameter  $b$  can be derived by a numerical solution of the following equation, which arises from the maximum likelihood estimator (MLE)

$$\frac{\left( \left( \sum_{i=1}^n x_i \right) + (-b - x_l)n \right) e^{-\frac{x_l}{b}} + e^{-\frac{x_u}{b}} \left( - \left( \sum_{i=1}^n x_i \right) + (b + x_u)n \right)}{\left( e^{-\frac{x_l}{b}} - e^{-\frac{x_u}{b}} \right) b^2} = 0, \tag{40}$$

where the  $x_i$  are the elements of the experimental sample with  $i$  varying between 1 and  $n$ . In the above formula,  $x_l$  represents the minimum of the sample and  $x_u$  the maximum.

### 3.2. The Truncated Half-Normal

Let  $X$  be a random variable defined on  $[0, \infty]$ ; the one-parameter *truncated Half-Normal* PDF is

$$f_{N,t}(x; s) = \frac{e^{-\frac{x^2}{2s^2}} \sqrt{2}}{\sqrt{\pi} s \left( -\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) + \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right) \right)}, \tag{41}$$

and its DF is

$$F_{N,t}(x; s) = - \frac{-\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) + \operatorname{erf} \left( \frac{\sqrt{2} x}{2s} \right)}{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)}, \tag{42}$$

which has the following series representation

$$F_{N,t}(x; s) = \frac{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right)}{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)} + \sum_{n=0}^{\infty} \left[ \frac{(-1)^n 2^{-n+\frac{1}{2}} s^{-2n-1} x^{2n+1}}{(2n+1) \sqrt{\pi} \left( \operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right) \right) n!} \right]. \tag{43}$$

The first moment about the origin  $\mu'_1$ , is

$$\mu'_1(s, x_l, x_u) = -\frac{\sqrt{2}s \left( e^{-\frac{x_l^2}{2s^2}} - e^{-\frac{x_u^2}{2s^2}} \right)}{\sqrt{\pi} \left( \operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right) \right)}, \quad (44)$$

and the second moment about the origin  $\mu'_2$ , is

$$\mu'_2(s, x_l, x_u) = \frac{\left( -\sqrt{2} e^{-\frac{x_l^2}{2s^2}} x_l + \sqrt{2} e^{-\frac{x_u^2}{2s^2}} x_u + s \sqrt{\pi} \left( \operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right) \right) \right) s}{\sqrt{\pi} \left( \operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right) \right)}. \quad (45)$$

The variance can be evaluated by Equation (38). There is no analytical expression for the median; we now present three approximations. The first approximate expression for the median can be obtained by the Menzel approximation for the error function, see Equation (A.1). We report the equation to be solved for  $x$  in order to find the median in the first approximation

$$\frac{-\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) + \sqrt{1 - e^{-\frac{2x^2}{s^2\pi}}}}{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)} = \frac{1}{2}, \quad (46)$$

which has the following approximation

$$\sqrt{\frac{-2\pi \ln \left( \frac{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right)^2}{4} - \frac{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)}{2} - \frac{\operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)^2}{4} + 1 \right) s}{2}}. \quad (47)$$

The second approximation for the median can be obtained by the Winitzki approximation for the error function, see Equation (A.2). We report the equation to be solved for  $x$  in order to find the second median

$$\frac{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \sqrt{1 - e^{-\frac{x^2 \left( \frac{4}{\pi} + \frac{7x^2}{100s^2} \right)}{2s^2 \left( \frac{7x^2}{100s^2} + 1 \right)}}}}{\operatorname{erf} \left( \frac{x_l \sqrt{2}}{2s} \right) - \operatorname{erf} \left( \frac{x_u \sqrt{2}}{2s} \right)} = \frac{1}{2}, \quad (48)$$

which has an omitted complicated solution. The third approximation for the median can be obtained using the Padé Approximant for the error function, see Equation (A.3). We report the equation to be solved for  $x$  in order to find the third median

$$-\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) + \tanh\left(\frac{\left(\frac{(300\pi^2 - 1200\pi)\sqrt{2}x}{2s} + \frac{(-10\pi^2 - 400\pi + 1280)\sqrt{2}x^3}{4s^3}\right)}{150\pi^{\frac{5}{2}} - 600\pi^{\frac{3}{2}} + \frac{15\left(3\pi^{\frac{5}{2}} - 40\pi^{\frac{3}{2}} + 96\sqrt{\pi}\right)x^2}{2s^2}}\right)}{\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) - \operatorname{erf}\left(\frac{x_u\sqrt{2}}{2s}\right)} = \frac{1}{2}, \quad (49)$$

which has an omitted complicated solution. **Table 2** reports the percent error of the approximate median for the truncated Half-Normal.

The parameter  $b$  can be derived by the numerical solution of the following equation, which arises from the MLE:

$$\frac{\sqrt{2}e^{-\frac{x_l^2}{2s^2}}nsx_l - \sqrt{2}e^{-\frac{x_u^2}{2s^2}}nsx_u + \sqrt{\pi}\left(\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) - \operatorname{erf}\left(\frac{x_u\sqrt{2}}{2s}\right)\right)\left(-s^2n + \left(\sum_{i=1}^n x_i^2\right)\right)}{\sqrt{\pi}s^3\left(\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) - \operatorname{erf}\left(\frac{x_u\sqrt{2}}{2s}\right)\right)} = 0. \quad (50)$$

Random generation of the truncated Half-Normal variate  $X$  is given by the numerical solution of the following non-linear equation

$$\frac{-\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) + \operatorname{erf}\left(\frac{\sqrt{2}x}{2s}\right)}{\operatorname{erf}\left(\frac{x_l\sqrt{2}}{2s}\right) - \operatorname{erf}\left(\frac{x_u\sqrt{2}}{2s}\right)} = R, \quad (51)$$

where  $R$  is the rectangular variate.

### 3.3. Truncated Sech-Square Distribution

Let  $X$  be a random variable defined on  $[x_l, x_u]$ ; the PDF for the truncated sech-square distribution is

$$f_{t,s}(x; b, x_l, x_u) = \frac{\operatorname{sech}\left(\frac{x}{b}\right)^2}{b\left(\tanh\left(\frac{x_u}{b}\right) - \tanh\left(\frac{x_l}{b}\right)\right)}, \quad (52)$$

and its DF is

**Table 2.** Percent error,  $\delta$ , of the median obtained by the approximation of the error function when  $x_l = 1, x_u = 10, s = 1$ .

Method	$\delta$ (%)
Menzel	1.37
Winitzki	0.053
Pade	0.036

$$F_{i,s}(x; b, x_l, x_u) = \frac{e^{\frac{x_l}{b}} \tanh\left(\frac{x}{b}\right) + e^{-\frac{x_l}{b}} \tanh\left(\frac{x}{b}\right) - e^{\frac{x_l}{b}} + e^{-\frac{x_l}{b}}}{\left(e^{\frac{x_l}{b}} + e^{-\frac{x_l}{b}}\right) \left(\tanh\left(\frac{x_u}{b}\right) - \tanh\left(\frac{x_l}{b}\right)\right)} \tag{53}$$

The first moment about the origin  $\mu'_1$ , is

$$\begin{aligned} \mu'_1(b, x_l, x_u) &= \frac{1}{2e^{\frac{2x_u}{b}} - 2e^{-\frac{2x_l}{b}}} \left( -b \ln\left(e^{\frac{2x_l}{b}} + 1\right) + b \ln\left(e^{\frac{2x_u}{b}} + 1\right) + 2x_l - 2x_u \right) e^{\frac{2x_u + 2x_l}{b}} \\ &\quad - b \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \ln\left(e^{\frac{2x_l}{b}} + 1\right) + b \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \ln\left(e^{\frac{2x_u}{b}} + 1\right) \\ &\quad + 2e^{\frac{2x_l}{b}} x_l - 2e^{\frac{2x_u}{b}} x_u, \end{aligned} \tag{54}$$

and the second moment about the origin  $\mu'_2$ , is

$$\begin{aligned} \mu'_2(b, x_l, x_u) &= \frac{1}{2e^{\frac{-2x_u + 2x_l}{b}} - 2} \left( -\text{Li}_2\left(-e^{\frac{2x_l}{b}}\right) b^2 + \text{Li}_2\left(-e^{\frac{2x_u}{b}}\right) b^2 - 2bx_l \ln\left(e^{\frac{2x_l}{b}} + 1\right) \right. \\ &\quad + 2bx_u \ln\left(e^{\frac{2x_u}{b}} + 1\right) + 2x_l^2 e^{\frac{-2x_u + 2x_l}{b}} - b^2 \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \text{Li}_2\left(-e^{\frac{2x_l}{b}}\right) \\ &\quad + b^2 \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \text{Li}_2\left(-e^{\frac{2x_u}{b}}\right) - 2bx_l \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \ln\left(e^{\frac{2x_l}{b}} + 1\right) \\ &\quad \left. + 2bx_u \left( e^{\frac{2x_l}{b}} + e^{\frac{2x_u}{b}} + 1 \right) \ln\left(e^{\frac{2x_u}{b}} + 1\right) + (2x_l^2 - 2x_u^2) e^{\frac{2x_l}{b}} - 2x_u^2 \right) \end{aligned} \tag{55}$$

The variance can be evaluated by Equation (38) and the median is at

$$\text{arctanh} \left( \frac{e^{\frac{2x_u + 2x_l}{b}} - 1}{\left(e^{\frac{2x_u}{b}} + 1\right) \left(e^{\frac{2x_l}{b}} + 1\right)} \right) b. \tag{56}$$

The parameter  $b$  can be derived by the numerical solution of the following equation, which arises from the MLE:

$$\begin{aligned} &\frac{1}{\left(-\tanh\left(\frac{x_u}{b}\right) + \tanh\left(\frac{x_l}{b}\right)\right) b^2} \left( 2 \tanh\left(\frac{x_l}{b}\right) - 2 \tanh\left(\frac{x_u}{b}\right) \right) \left( \sum_{i=1}^n \frac{\left(e^{\frac{2x_i}{b}} - 1\right) x_i}{e^{\frac{2x_i}{b}} + 1} \right) \\ &- n \left( x_l \left( \tanh^2\left(\frac{x_l}{b}\right) \right) - x_u \left( \tanh^2\left(\frac{x_u}{b}\right) \right) + \tanh\left(\frac{x_l}{b}\right) b - b \tanh\left(\frac{x_u}{b}\right) - x_l + x_u \right) \\ &= 0. \end{aligned} \tag{57}$$

Random generation of the exponential variate  $X$  is given by

$$X : b, x_l, x_u \approx \ln \left( \frac{\sqrt{N}}{\left( \frac{x_l}{e^b} \right)^2 R - \left( \frac{x_u}{e^b} \right)^2 R + \left( \frac{x_u}{e^b} \right)^2 + 1} \right) b, \tag{58}$$

where  $R$  is the unit rectangular variate and

$$N = - \left( \left( \frac{x_l}{e^b} \right)^2 R - \left( \frac{x_u}{e^b} \right)^2 R + \left( \frac{x_u}{e^b} \right)^2 + 1 \right) \left( -e^{\frac{2x_u + 2x_l}{b}} + \left( \frac{x_l}{e^b} \right)^2 R - \left( \frac{x_u}{e^b} \right)^2 R - \left( \frac{x_l}{e^b} \right)^2 \right). \tag{59}$$

### 4. Astrophysical Applications

This section reviews the adopted statistics as well as some data on open clusters and Gaia’s stars.

#### 4.1. Statistics

The merit function  $\chi^2$  is computed according to the formula

$$\chi^2 = \sum_{i=1}^n \frac{(T_i - O_i)^2}{T_i}, \tag{60}$$

where  $n$  is the number of bins,  $T_i$  is the theoretical value, and  $O_i$  is the experimental value represented by the frequencies. The theoretical frequency distribution is given by

$$T_i = N \Delta x_i p(x), \tag{61}$$

where  $N$  is the number of elements of the sample,  $\Delta x_i$  is the magnitude of the size interval, and  $p(x)$  is the PDF under examination. A reduced merit function  $\chi_{red}^2$  is given by

$$\chi_{red}^2 = \chi^2 / NF, \tag{62}$$

where  $NF = n - k$  is the number of degrees of freedom,  $n$  is the number of bins, and  $k$  is the number of parameters. The goodness of the fit can be expressed by the probability  $Q$ , see equation 15.2.12 in [24], which involves the number of degrees of freedom and  $\chi^2$ . According to [24] p. 658, the fit “may be acceptable” if  $Q > 0.001$ . The Akaike information criterion (AIC), see [26], is defined by

$$AIC = 2k - 2 \ln(L), \tag{63}$$

where  $L$  is the likelihood function and  $k$  the number of free parameters in the model. We assume a Gaussian distribution for the errors. Then the likelihood function can be derived from the  $\chi^2$  statistic  $L \propto \exp\left(-\frac{\chi^2}{2}\right)$  where  $\chi^2$  has been computed by equation (60), see [27], [28]. Now the AIC becomes

$$AIC = 2k + \chi^2. \tag{64}$$

The difference between the evaluation of  $\chi^2$  and the Kolmogorov-Smirnov

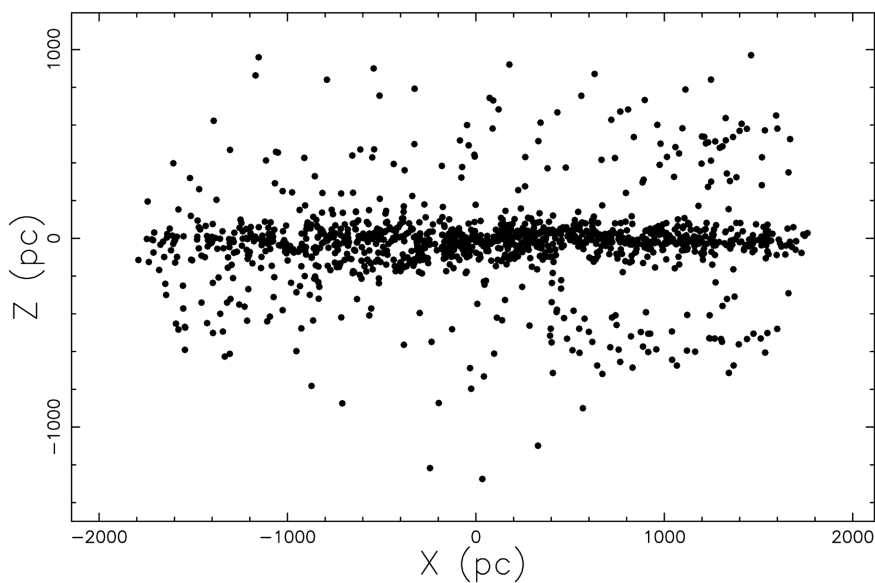
test (K-S), see [29] [30] [31] is that the latter does not require binning the data. The K-S test, as implemented by the FORTRAN subroutine KSONE in [24], finds the maximum distance,  $D$ , between the theoretical and the astronomical DF as well as the significance level  $P_{KS}$ , see formulas 14.3.5 and 14.3.9 in [24]; if  $P_{KS} \geq 0.1$ , the goodness of the fit is believable.

## 4.2. Open Clusters

The open clusters in a radius of 1.8 Kpc were analysed in [32] with a catalog available at CDS. The data can be processed by introducing the rectangular coordinates  $(X, Y, Z)$  and assuming a distance of the Sun from the Galactic centre,  $R_{\odot}$ , of 8 kpc, see **Figure 1**.

We are interested in the distribution of  $Z$ , the distance perpendicular to the galactic plane, and the results for the three distributions here analysed in the interval  $[0, \infty]$  are reported in **Table 3**. and those for the three truncated distributions, analysed in the interval  $[x_l, x_u]$ , are reported in **Table 4**.

A careful examination of **Table 3** and **Table 4** allows concluding that the best results are obtained for the usual exponential distribution, see **Figure 2**, followed



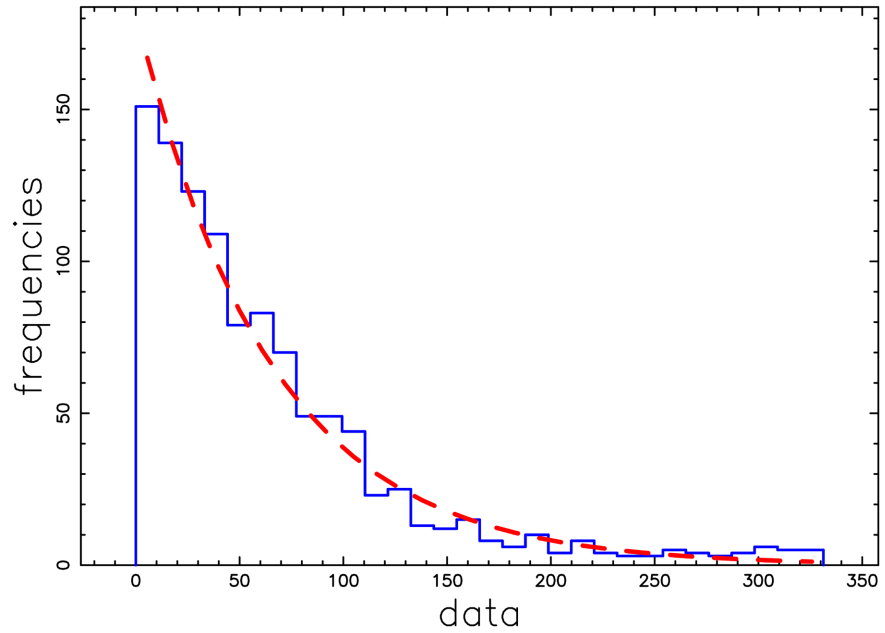
**Figure 1.** Distribution of 1241 open clusters in the  $X$ - $Z$  plane as projected onto the galactic pole.

**Table 3.** Numerical values of  $\chi_{red}^2$ , AIC, probability  $Q$ ,  $D$ , the maximum distance between theoretical and observed DF, and  $P_{KS}$ , significance level, in the K-S test for data from open clusters when  $|Z| \leq 335$  pc. The number of linear bins,  $n$ , is 30.

PDF	parameters	AIC	$\chi_{red}^2$	$Q$	$D$	$P_{KS}$
Exponential	$b = 64.41$ pc	44.16	1.45	0.0542	0.0274	0.39
Half-Normal	$s = 80.73$ pc	110.72	3.74	$3.74 \times 10^{-11}$	$7.68 \times 10^{-2}$	$6.40 \times 10^{-6}$
Sech-square	$b = 92.93$ pc	62.95	2.1	$4.69 \times 10^{-4}$	$5.05 \times 10^{-2}$	$8.47 \times 10^{-3}$

**Table 4.** Numerical values of  $\chi_{red}^2$ , AIC, probability  $Q$ ,  $D$ , the maximum distance between theoretical and observed DF, and  $P_{KS}$ , significance level, in the K-S test for open clusters data when  $|Z| \leq 335$  pc. The imposed parameters are  $n = 30$ ,  $x_i = 2.57 \times 10^{-2}$  pc and  $x_u = 331.28$  pc.

PDF	parameters	AIC	$\chi_{red}^2$	$Q$	$D$	$P_{KS}$
Truncated exponential	$b = 66.71$ pc	53.13	1.74	0.0059	0.032	0.1953
Truncated half-normal	$s = 90.74$ pc	180.86	6.47	$1.35 \times 10^{-23}$	0.119	$1.35 \times 10^{-13}$
Truncated sech-square	$b = 95.09$ pc	73.36	2.49	$2.64 \times 10^{-5}$	0.0582	$1.41 \times 10^{-3}$



**Figure 2.** Empirical PDF of distribution of  $Z$  for open cluster data (blue histogram) with a superposition of the exponential PDF (red line). Theoretical parameters as in **Table 3**.

by the truncated exponential distribution.

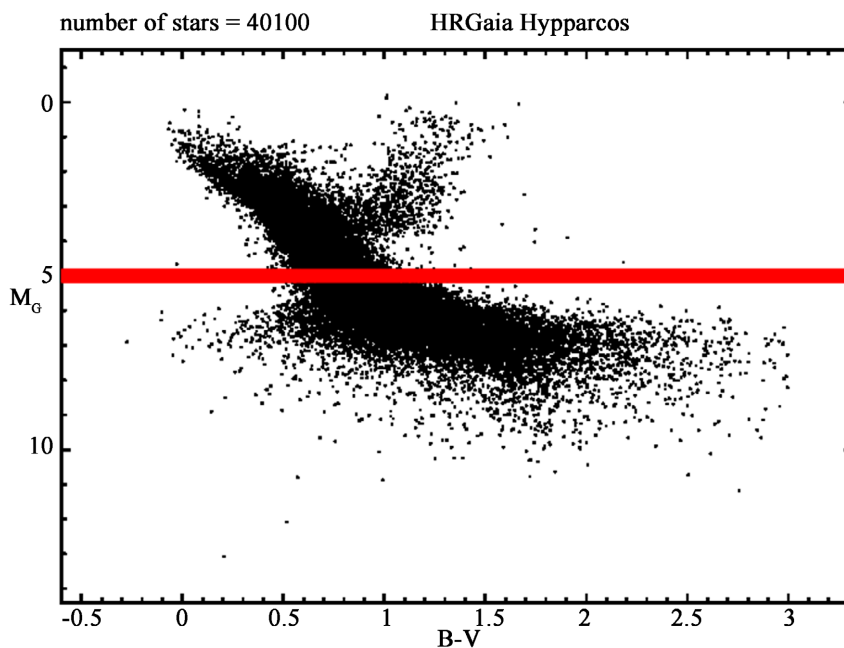
### 4.3. Stars

A great number of stars with mean apparent magnitude in the G-band are available in the Gaia Data Release 1 (Gaia DR1) astrometric catalogs, see [33] [34], with data at <http://vizier.u-strasbg.fr/viz-bin/VizieR> and, specifically, Table I/337/tgasptyc. The absolute magnitude,  $M_G$ , is obtained by the usual formula

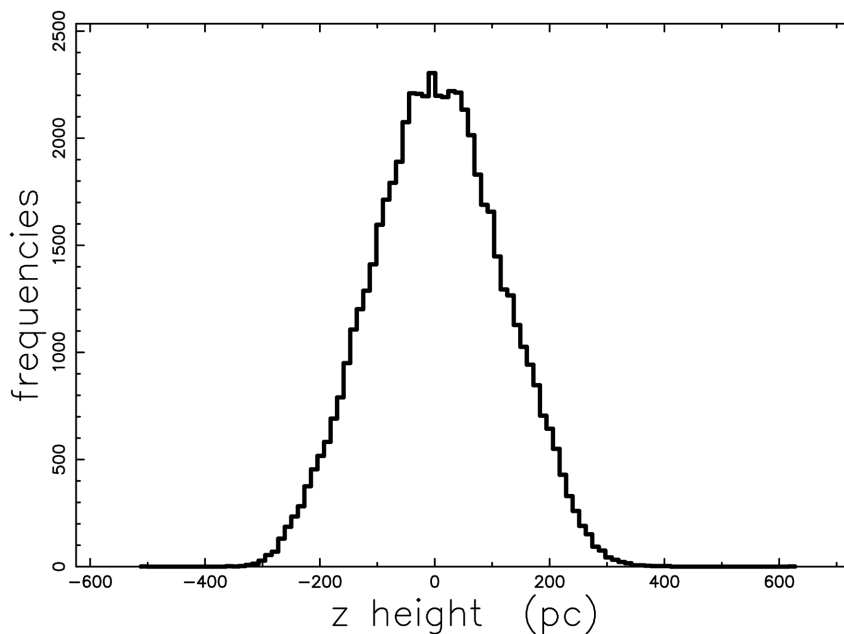
$$M_g = -5 \log(d) + 5 + m_g, \quad (65)$$

where  $m_g$  is the apparent magnitude in the G-band, and  $d$  is the distance in pc. We now select the stars with  $4.8 \leq M_G \leq 5.2$  in the first 1000 pc, amounting to a total of 58027, and we evaluate their rectangular coordinates  $(X, Y, Z)$ . The position of the above slice in absolute magnitude in the H-R diagram is visible in **Figure 3**. The obtained data in  $Z$  of the first 1000 pc are shifted by  $Z_\odot = 5.075$  pc which defines the Sun's position relative to the plane of symmetry; for a review

of the values of  $Z_{\odot}$ , see **Table 1** in [35]. The re-scaled distribution is visible in **Figure 4**. The distribution of  $Z$  has both negative and positive values and in order to increase the statistics we take the absolute values of  $Z$  because the distributions here analysed are defined only for positive values of the random variable. The results for the three distributions here analysed in the interval  $[0, \infty]$  are reported in **Table 5** and those for the three truncated distributions analysed in the interval  $[x_l, x_u]$  are reported in **Table 6**.



**Figure 3.**  $M_G$  against  $(B-V)$ , evaluated as BT-VT, (H-R diagram) in the first 100 pc and selected region in red.



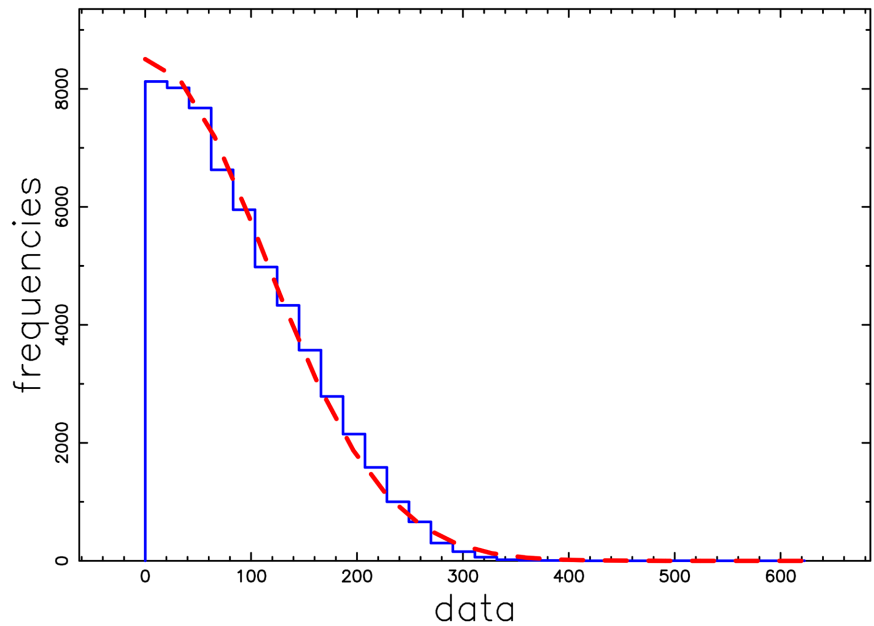
**Figure 4.** Histogram of the re-scaled distribution of  $Z$  for stars with  $4.8 \leq M_G \leq 5.2$ .

**Table 5.** Numerical values of  $\chi^2_{red}$ , AIC, probability  $Q$ ,  $D$ , the maximum distance between theoretical and observed DF, and  $P_{KS}$ , significance level, in the K-S test for the re-scaled distribution of  $Z$  of the Gaia’s star data when  $4.8 \leq M_G \leq 5.2$ . The number of linear bins,  $n$ , is 30.

PDF	parameters	AIC	$\chi^2_{red}$	$Q$	$D$	$P_{KS}$
Exponential	$b = 91.33$ pc	33,073.41	1140	0	$8.79 \times 10^{-2}$	0
Half-Normal	$s = 114.46$ pc	1072.15	36.9	0	$1 \times 10^{-2}$	$4.5 \times 10^{-6}$
Sech-square	$b = 131.76$ pc	9232	318	0	$3.43 \times 10^{-2}$	0

**Table 6.** Numerical values of  $\chi^2_{red}$ , AIC, probability  $Q$ ,  $D$ , the maximum distance between theoretical and observed DF, and  $P_{KS}$ , significance level, in the K-S test for the re-scaled distribution of  $Z$  of the Gaia’s star data when  $4.8 \leq M_G \leq 5.2$ . The imposed parameters are  $n = 30$ ,  $x_l = 1.8 \times 10^{-4}$  pc and  $x_u = 622.53$  pc.

PDF	parameters	AIC	$\chi^2_{red}$	$Q$	$D$	$P_{KS}$
Truncated exponential	$b = 92.051$ pc	34,618	1282	0	$8.59 \times 10^{-2}$	0
Truncated half-normal	$s = 112.97$ pc	883.82	32.51	0	$1.68 \times 10^{-2}$	$8.3 \times 10^{-15}$
Truncated sech-square	$b = 130.87$ pc	8734	323.2	0	$3.73 \times 10^{-2}$	0



**Figure 5.** Empirical PDF of distribution of  $Z$  for Gaia stars (blue histogram) with a superposition of the truncated half-normal PDF (red line). Theoretical parameters as in **Table 6**.

A careful examination of **Table 5** and **Table 6** allows concluding that the best results are obtained for the truncated half-normal distribution, see **Figure 5**.

## 5. Conclusions

### The Truncated Distributions

We derived the PDF, the DF, the average value, the  $r$ th moment, the median, expressions to generate the random variate and the MLE for the truncated exponential, truncated half-normal and the truncated sech-square distributions.

### Astrophysical Applications

We applied both the usual and the truncated three distributions to a sample of galactic-height for open clusters and for Gaia's stars. In the case of open clusters, the best results are obtained by the exponential distribution, followed by the sech-square distribution. In the case of Gaia's stars, the best results are obtained by the truncated half-normal distribution, followed by the sech-square distribution. This means that in the case of Gaia's stars, the truncated half-normal distribution is the best model and should be used by astronomers in the fitting procedure.

### Prospects for the Future

In view of the great importance of the doubly truncated distributions in astrophysics, other distributions can be analysed, for example, the Half-Gumbel distribution [36].

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Spitzer Jr., L. (1942) The Dynamics of the Interstellar Medium. III. Galactic Distribution. *The Astrophysical Journal*, **95**, 329-344. <https://doi.org/10.1086/144407>
- [2] Conti, P.S. and Vacca, W.D. (1990) The Distribution of Massive Stars in the Galaxy. I. The Wolf-Rayet Stars. *The Astrophysical Journal*, **100**, 431-444. <https://doi.org/10.1086/115528>
- [3] Bertin, G. (2000) Dynamics of Galaxies. Cambridge University Press, Cambridge.
- [4] Padmanabhan, P. (2002) Theoretical Astrophysics. Vol. III: Galaxies and Cosmology. Cambridge University Press, Cambridge.
- [5] Maz-Apellániz, J. (2001) The Spatial Distribution of O-B5 Stars in the Solar Neighborhood as Measured by Hipparcos. *The Astrophysical Journal*, **121**, 2737. <https://doi.org/10.1086/320399>
- [6] Balakrishnan, N. and Joshi, P. (1984) Product Moments of Order Statistics from the Doubly Truncated Exponential Distribution. *Naval Research Logistics Quarterly*, **31**, 27-31. <https://doi.org/10.1002/nav.3800310105>
- [7] Zaninetti, L. and Ferraro, M. (2008) On the Truncated Pareto Distribution with Applications. *Central European Journal of Physics*, **6**, 1-6. <https://doi.org/10.2478/s11534-008-0008-2>
- [8] Zaninetti, L. (2013) The Initial Mass Function Modeled by a Left Truncated Beta Distribution. *The Astrophysical Journal*, **765**, Article No. 128. <https://doi.org/10.1088/0004-637X/765/2/128>
- [9] Zaninetti, L. (2013) A Right and Left Truncated Gamma Distribution with Application to the Stars. *Advanced Studies in Theoretical Physics*, **23**, 1139-1147. <https://doi.org/10.12988/astp.2013.310125>

- [10] Zaninetti, L. (2017) A Left and Right Truncated Lognormal Distribution for the Stars. *Advances in Astrophysics*, **2**, 197-213. <https://doi.org/10.22606/adap.2017.23005>
- [11] Zaninetti, L. (2019) The Truncated Lindley Distribution with Applications in Astrophysics. *Galaxies*, **7**, 61-78. <https://doi.org/10.3390/galaxies7020061>
- [12] Zaninetti, L. (2019) New Probability Distributions in Astrophysics: I. The Truncated Generalized Gamma. *International Journal of Astronomy and Astrophysics*, **9**, 393-410. <https://doi.org/10.4236/ijaa.2019.94027>
- [13] Zaninetti, L. (2020) New Probability Distributions in Astrophysics: II. The Generalized and Double Truncated Lindley. *International Journal of Astronomy and Astrophysics*, **10**, 39-55. <https://doi.org/10.4236/ijaa.2020.101004>
- [14] Zaninetti, L. (2020) New Probability Distributions in Astrophysics: III. The Truncated Maxwell-Boltzmann Distribution. *International Journal of Astronomy and Astrophysics*, **10**, 191-202. <https://doi.org/10.4236/ijaa.2020.103010>
- [15] Zaninetti, L. (2020) New Probability Distributions in Astrophysics: IV. The Relativistic Maxwell-Boltzmann Distribution. *International Journal of Astronomy and Astrophysics*, **10**, 302-318. <https://doi.org/10.4236/ijaa.2020.104016>
- [16] Zaninetti, L. (2021) New Probability Distributions in Astrophysics: V. The Truncated Weibull Distribution. *International Journal of Astronomy and Astrophysics*, **11**, 133-149. <https://doi.org/10.4236/ijaa.2021.111008>
- [17] Zaninetti, L. (2021) New Probability Distributions in Astrophysics: VI. The Truncated Sujatha Distribution. *International Journal of Astronomy and Astrophysics*, **11**, 517-529. <https://doi.org/10.4236/ijaa.2021.114028>
- [18] Zaninetti, L. (2022) New Probability Distributions in Astrophysics: VII. The Truncated Gamma-Pareto Distribution Applied to Cosmic Rays. *International Journal of Astronomy and Astrophysics*, **12**, 132-146. <https://doi.org/10.4236/ijaa.2022.121008>
- [19] Jowett, G. (1958) The Exponential Distribution and Its Applications. *The Incorporated Statistician*, **8**, 89-95. <https://doi.org/10.2307/2986561>
- [20] Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011) *Statistical Distributions*. 4th Edition, John Wiley & Sons, Hoboken. <https://doi.org/10.1002/9780470627242>
- [21] Bland, J.M. (2005) The Half-Normal Distribution Method for Measurement Error: Two Case Studies. <https://www-users.york.ac.uk/~mb55/talks/halfnor.pdf>
- [22] Olver, F.W.J., Lozier, D.W., Boisvert, R.F. and Clark, C.W. (2010) *NIST Handbook of Mathematical Functions*. Cambridge University Press, Cambridge.
- [23] Abramowitz, M. and Stegun, I.A. (1965) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York. <https://doi.org/10.1063/1.3047921>
- [24] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in FORTRAN. The Art of Scientific Computing*. Cambridge University Press, Cambridge.
- [25] van der Kruit, P.C. (1988) The Three-Dimensional Distribution of Light and Mass in Disks of Spiral Galaxies. *Astronomy & Astrophysics*, **192**, 117-127.
- [26] Akaike, H. (1974) A New Look at the Statistical Model Identification *IEEE Transactions on Automatic Control*, **19**, 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- [27] Liddle, A.R. (2004) How Many Cosmological Parameters? *MNRAS*, **351**, L49. <https://doi.org/10.1111/j.1365-2966.2004.08033.x>

- [28] Godlowski, W. and Szydowski, M. (2005) Constraints on Dark Energy Models from Supernovae. In: Turatto, M., Benetti, S., Zampieri, L. and Shea, W., Eds., 1604-2004: *Supernovae as Cosmological Lighthouses*, Vol. 342 of Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific, San Francisco, 508-516.
- [29] Kolmogoroff, A. (1941) Confidence Limits for an Unknown Distribution Function. *The Annals of Mathematical Statistics*, **12**, 461-463. <https://doi.org/10.1214/aoms/1177731684>
- [30] Smirnov, N. (1948) Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, **19**, 279-281. <https://doi.org/10.1214/aoms/1177730256>
- [31] Massey Frank, J.J. (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, **46**, 68-78. <https://doi.org/10.1080/01621459.1951.10500769>
- [32] Joshi, Y.C., Dambis, A.K., Pandey, A.K. and Joshi, S. (2016) Study of Open Clusters within 1.8 kpc and Understanding the Galactic Structure. *Astronomy & Astrophysics*, **593**, A116. <https://doi.org/10.1051/0004-6361/201628944>
- [33] Gaia Collaboration, Prusti, T., de Bruijne, J.H.J., Brown, A.G.A., Vallenari, A., Babusiaux, C., Bailer-Jones, C.A.L., Bastian, U., Biermann, M., Evans, D.W., *et al.* (2016) The Gaia Mission. *Astronomy & Astrophysics*, **595**, A1.
- [34] Gaia Collaboration, Brown, A.G.A., Vallenari, A., Prusti, T., de Bruijne, J.H.J., Mignard, F., Drimmel, R., Babusiaux, C., Bailer-Jones, C.A.L., Bastian, U., *et al.* (2016) Gaia Data Release 1. Summary of the Astrometric, Photometric, and Survey Properties. *Astronomy & Astrophysics*, **595**, A2.
- [35] Bobylev, V.V. and Bajkova, A.T. (2016) Analysis of the Z Distribution of Young Objects in the Galactic Thin Disk. *Astronomy Letters*, **42**, 1-9. <https://doi.org/10.1134/S1063773716010023>
- [36] Gumbel, E.J. (1935) Les valeurs extrêmes des distributions statistiques. *Annales de l'institut Henri Poincaré*, **5**, 115-158.
- [37] Howard, R.M. (2022) Arbitrarily Accurate Analytical Approximations for the Error Function. *Mathematical and Computational Applications*, **27**, Article No. 14. <https://doi.org/10.3390/mca27010014>
- [38] Menzel, R. (1975) Approximate Closed Form Solution to the Error Function. *American Journal of Physics*, **43**, 366-367. <https://doi.org/10.1119/1.10085>
- [39] Menzel, R. (1975) Erratum: "Approximate Closed form Solution to the Error Function". *American Journal of Physics*, **43**, Article No. 923. <https://doi.org/10.1119/1.10112>
- [40] Winitzki, S. (2008) A Handy Approximation for the Error Function and Its Inverse. A Lecture Note Obtained through Private Communication.

## Appendix. Approximations for the Error Function

We report two existing approximations for the error function which are invertible, see Table I in [37] for more details, and a new approximation. The Menzel approximation [38] [39] for the error function is

$$\operatorname{erf}(x) = \sqrt{1 - e^{-\frac{4x^2}{\pi}}}. \quad (\text{A.1})$$

The Winitzki approximation [40] for the error function is

$$\operatorname{erf}(x) = \sqrt{1 - e^{-\frac{x^2 \left( \frac{4}{\pi} + \frac{7x^2}{50} \right)}{7x^2 + 50}}}. \quad (\text{A.2})$$

A new approximation for the error function is derived in the framework of the Padé Approximant of order (4,2)

$$\operatorname{erf}(x) = \tanh \left( \frac{\left( 300\pi^2 - 1200\pi \right) x + \left( -10\pi^2 - 400\pi + 1280 \right) x^3}{150\pi^{\frac{5}{2}} - 600\pi^{\frac{3}{2}} + 15 \left( 3\pi^{\frac{5}{2}} - 40\pi^{\frac{3}{2}} + 96\sqrt{\pi} \right) x^2} \right). \quad (\text{A.3})$$

**Table 7** reports the percent error of the three cases here analysed.

**Table 7.** Percent error,  $\delta$ , for approximating the error function in [0,5].

Method	$\delta(\%)$
Menzel	0.7
Winitzki	0.03
Pade	0.05