International Journal of Astronomy and Astrophysics, 2020, 10, 39-55
https://www.scirp.org/journal/ijaa
ISSN Online: 2161-4725
ISSN Print: 2161-4717

New Probability Distributions in Astrophysics: II. The Generalized and Double Truncated Lindley

Lorenzo Zaninetti

Physics Department, via P.Giuria 1, Turin, Italy
Email: zaninetti@ph.unito.it

How to cite this paper: 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

Received: February 8, 2020
Accepted: March 24, 2020
Published: March 27, 2020

Abstract

The statistical parameters of five generalizations of the Lindley distribution, such as the average, variance and moments, are reviewed. A new double truncated Lindley distribution with three parameters is derived. The new distributions are applied to model the initial mass function for stars.

Keywords

Stars: Mass Function, Stars: Fundamental Parameters, Methods: Statistical

1. Introduction

The Lindley distribution, after [1] [2], has one parameter. In recent years the Lindley distribution has been the subject of many generalizations, we report some of them among others: one with two parameters [3], a two-parameter weighted one [4], the generalized Poisson-Lindley [5], the extended Lindley [6] and a transmuted Lindley-geometric distribution [7]. Several generalizations of the Lindley distribution can be found in a recent review [8]. The Lindley distribution is useful in modeling biological data from grouped mortality studies [4] [9] and the first application to astrophysics of the Lindley distribution has been done for the initial mass function (IMF) for stars and the luminosity function for galaxies [10]. The IMF is routinely modeled by the lognormal distribution and therefore the following question naturally arises. Can a Lindley distribution or a generalization be an alternative to the lognormal fit for the IMF? In order to answer the above question Section 2 reviews the notion of statistical sample and Lindley distribution, Section 3 reviews five generalizations of the Lindley distribution, Section 4 introduces the double Lindley distribution and Section 5 fits the six new Lindley distributions to four samples for the mass of the stars.

2. Preliminaries

We report some basic information on the adopted sample and on the original Lindley distribution with one parameter.

2.1. The Sample

The experimental sample consists of the data x_i with *i* varying between 1 and *n*; the sample mean, $\overline{x}$, is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

the unbiased sample variance, s^2, is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2, \tag{2}$$

and the sample *r*th moment about the origin, $\overline{x}_r$, is

$$\overline{x}_r = \frac{1}{n}\sum_{i=1}^{n}\left(x_i\right)^r. \tag{3}$$

2.2. The Lindley Distribution with One Parameter

The *Lindley* probability density function (PDF) with one parameter, $f(x)$, is

$$f(x;c) = \frac{c^2 \mathrm{e}^{-cx}(x+1)}{1+c};\ x>0, c>0 \tag{4}$$

where $x>0$ and $c>0$.

The cumulative distribution function (CDF), $F(x)$, is

$$F(x;c) = 1-\left(1+\frac{cx}{1+c}\right)\mathrm{e}^{-cx};\ x>0, c>0. \tag{5}$$

At $x=0$, $f(0)=\frac{c^2}{1+c}$ and is not zero.

The average value or mean, μ, is

$$\mu(c) = \frac{2+c}{c(1+c)}, \tag{6}$$

the variance, σ^2, is

$$\sigma^2(c) = \frac{c^2+4c+2}{c^2(1+c)^2}. \tag{7}$$

The *r*th moment about the origin for the Lindley distribution, μ_r', is

$$\mu_r' = \frac{c^{-r}\Gamma(r+2)+c^{1-r}\Gamma(r+1)}{1+c}, \tag{8}$$

where

$$\Gamma(z) = \int_0^\infty \mathrm{e}^{-t} t^{z-1}\mathrm{d}t, \tag{9}$$

is the gamma function, see [11]. The central moments, μ_r, are

$$\mu_3 = \frac{2c^3 + 12c^2 + 12c + 4}{c^3 (1+c)^3} \tag{10a}$$

$$\mu_4 = \frac{9c^4 + 72c^3 + 132c^2 + 96c + 24}{c^4 (1+c)^4} \tag{10b}$$

More details can be found in [2].

3. Generalizations of the Lindley Distribution

We review the statistics of the Lindley distribution with two parameters, power, generalized, new generalized and new weighted.

3.1. The Lindley Distribution with Two Parameters

The *Lindley* PDF with two parameters TPLD [3] is

$$f(x;b,c) = \frac{c^2 (b+x) e^{-cx}}{bc+1}, \tag{11}$$

where $x > 0$, $c > 0$ and $bc > -1$. The CDF of the TPLD is

$$F(x;c,b) = 1 - \frac{(bc + cx + 1) e^{-cx}}{bc+1}. \tag{12}$$

The average value or mean of the TPLD is

$$\mu(b,c) = \frac{bc+2}{c(bc+1)}, \tag{13}$$

and the variance of the TPLD is

$$\sigma^2(b,c) = \frac{b^2c^2 + 4bc + 2}{c^2 (bc+1)^2}. \tag{14}$$

The mode of the TPLD is at

$$Mode = \frac{1-bc}{c}, \tag{15}$$

see Equation (2.3) in [3]. The *r*th moment about the origin for the TPLD, μ_r', is

$$\mu_r' = \frac{c^{1-r} b \Gamma(r+1) + c^{-r} \Gamma(r+2)}{bc+1}. \tag{16}$$

The two parameters b and c can be obtained by the following match

$$\mu = \overline{x} \tag{17a}$$

$$\sigma^2 = s^2, \tag{17b}$$

which means

$$\hat{b} = \frac{-\left(s^2 + \overline{x}^2\right)\left(\overline{x}\sqrt{-2s^2 + 2\overline{x}^2} - 2s^2\right)}{\left(\overline{x}\sqrt{-2s^2 + 2\overline{x}^2} + \overline{x}^2 - s^2\right)\left(2\overline{x} + \sqrt{-2s^2 + 2\overline{x}^2}\right)}, \tag{18}$$

and

$$\hat{c} = \frac{2\overline{x} + \sqrt{-2s^2 + 2\overline{x}^2}}{s^2 + \overline{x}^2}. \tag{19}$$

3.2. The Power Lindley Distribution

The *power Lindley* PDF with two parameters (PLD) according to [3] is

$$f(x;b,c) = \frac{cb^2 \left(1+x^c\right) x^{c-1} \mathrm{e}^{-bx^c}}{b+1}, \tag{20}$$

where b, c and $x > 0$. The CDF of the PLD is

$$F(x;c,b) = \frac{\left(-bx^c - b - 1\right)\mathrm{e}^{-bx^c} + b + 1}{b+1}. \tag{21}$$

The average value or mean of the PLD is

$$\mu(b,c) = \frac{\left(b^{-c^{-1}} c + b^{\frac{c-1}{c}} c + b^{-c^{-1}}\right)\Gamma\left(\frac{c+1}{c}\right)}{(b+1)c}, \tag{22}$$

and the variance of the PLD is

$$\sigma^2(b,c) = \frac{NA}{DA}, \tag{23}$$

where

$$\begin{aligned} NA = & -b^{-2c^{-1}}\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 c^2 + b^{-2c^{-1}}\Gamma\left(\frac{c+2}{c}\right) bc^2 - b^{\frac{2c-2}{c}}\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 c^2 \\ & -2\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 b^{\frac{-2+c}{c}} c^2 + b^{\frac{-2+c}{c}}\Gamma\left(\frac{c+2}{c}\right) bc^2 - 2b^{-2c^{-1}}\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 c \\ & +2b^{-2c^{-1}}\Gamma\left(\frac{c+2}{c}\right) bc + b^{-2c^{-1}}\Gamma\left(\frac{c+2}{c}\right) c^2 - 2\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 b^{\frac{-2+c}{c}} c \\ & + b^{\frac{-2+c}{c}}\Gamma\left(\frac{c+2}{c}\right) c^2 - b^{-2c^{-1}}\left(\Gamma\left(\frac{c+1}{c}\right)\right)^2 + 2b^{-2c^{-1}}\Gamma\left(\frac{c+2}{c}\right) c, \end{aligned} \tag{24}$$

and

$$DA = (b+1)^2 c^2. \tag{25}$$

The mode of the PLD is at

$$Mode = \frac{-cb + \sqrt{1+\left(b^2+4\right)c^2 + (-2b-4)c} + 2c - 1}{2cb}. \tag{26}$$

The rth moment about the origin for the PLD is

$$\mu_r' = \frac{b^{\frac{-r+c}{c}}\Gamma\left(\frac{r+c}{c}\right) + b^{-\frac{r}{c}}\Gamma\left(\frac{r+2c}{c}\right)}{b+1}. \tag{27}$$

The two parameters b and c of the PLD can be found by numerically solving the nonlinear system given by Equation (17a) and Equation (17b).

3.3. The Generalized Lindley Distribution

The *generalized Lindley* PDF with three parameters (GLD) according to [12] is

$$f(x;a,b,c)=\frac{b^2(bx)^{a-1}(cx+a)\mathrm{e}^{-bx}}{(c+b)\Gamma(a+1)}, \tag{28}$$

where a, b, c and $x>0$. The CDF of the GLD is

$$F(x;a,c,b) = \frac{\mathrm{e}^{-1/2bx}\left(x^{a/2}\left(cb^{a/2}+b^{a/2+1}\right)M_{a/2,a/2+1/2}(bx)+b^{a+1}x^{a}\mathrm{e}^{-1/2bx}(a+1)\right)}{(c+b)\Gamma(a+2)}, \tag{29}$$

where $M_{\mu,\nu}(z)$ is the Whittaker M function, see [11]. The average value or mean of the GLD is

$$\mu(a,b,c)=\frac{ab+ac+c}{b(c+b)}, \tag{30}$$

and the variance of the GLD is

$$\sigma^2(a,b,c)=\frac{ab^2+2cba+c^2a+2cb+c^2}{b^2(c+b)^2}. \tag{31}$$

The hazard rate function, $h(x;a,b,c)$, of the GLD is

$$h(x;a,b,c) = \frac{-b^{a+1}x^{a-1}(cx+a)\mathrm{e}^{-bx}(a+1)}{\mathrm{e}^{-1/2bx}x^{a/2}\left(cb^{a/2}+b^{a/2+1}\right)M_{a/2,a/2+1/2}(bx)+x^{a}b^{a+1}(a+1)\mathrm{e}^{-bx}-(c+b)\Gamma(a+2)}, \tag{32}$$

and **Figure 1** reports an example. Here the CDF, Equation (29), and the hazard rate function, Equation (32), are reported in closed form in contrast to what was asserted by [12]. The mode of the GLD is at

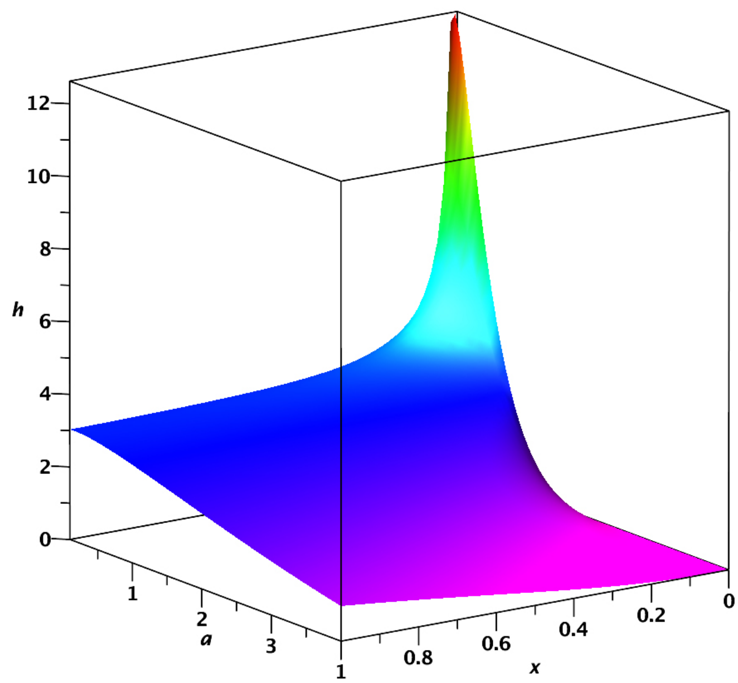

Figure 1. Plot of the three-dimensional surface of the hazard rate function when $b = 3$ and $c = 0.5$.

$$Mode = \frac{-ab + ac + \sqrt{a^2b^2 + 2a^2bc + a^2c^2 - 4abc}}{2bc}. \quad (33)$$

The rth moment about the origin for the GLD is

$$\mu_r' = \frac{\Gamma(r+a)\left(b^{-r}ca + b^{-r}cr + b^{-r+1}a\right)}{(c+b)\Gamma(a+1)}, \quad (34)$$

and in particular the third moment is

$$\mu_3' = \frac{\Gamma(3+a)(ab+ac+3c)}{(c+b)\Gamma(a+1)b^3}. \quad (35)$$

The three parameters a, b and c of the GLD can be obtained by numerically solving the following three non-linear equations

$$\mu = \overline{x} \quad (36a)$$

$$\sigma^2 = s^2 \quad (36b)$$

$$\mu_3' = \overline{x}_3. \quad (36c)$$

3.4. The New Generalized Lindley Distribution

The *new generalized Lindley* PDF with three parameters (NGLD) according to [13] is

$$f(x;a,b,c) = \frac{\left(c^{a+1}x^{a-1}\Gamma(b) + c^b x^{b-1}\Gamma(a)\right)e^{-cx}}{(1+c)\Gamma(a)\Gamma(b)}, \quad (37)$$

where a, b, c and $x > 0$. The CDF of the NGLD is

$$F(x;a,c,b) = \frac{NB}{(1+c)\Gamma(b+2)\Gamma(a+2)}, \quad (38)$$

where

$$\begin{aligned} NB = {} & \Gamma(b+2)x^a c^{a+1}e^{-cx}a + \Gamma(a+2)x^b c^b e^{-cx}b - \Gamma(b+2)c\Gamma(a+1,cx)a \\ & + \Gamma(b+2)x^a c^{a+1}e^{-cx} + \Gamma(a+2)x^b c^b e^{-cx} - \Gamma(b+2)c\Gamma(a+1,cx) \\ & + \Gamma(b+2)\Gamma(a+2)c - \Gamma(a+2)\Gamma(b+1,cx)b + \Gamma(b+2)\Gamma(a+2) \\ & - \Gamma(a+2)\Gamma(b+1,cx) \end{aligned} \quad (49)$$

where $\Gamma(a,z)$ is the incomplete Gamma function, defined by

$$\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}dt, \quad (40)$$

see [11]. The average value of the NGLD is

$$\mu(a,b,c) = \frac{ac+b}{c(1+c)}, \quad (41)$$

and the variance of the NGLD is

$$\sigma^2(a,b,c) = \frac{a^2c - 2abc + ac^2 + b^2c + ac + bc + b}{c^2(1+c)^2}. \quad (42)$$

The rth moment about the origin for the NGLD is

$$\mu_r' = \frac{c^{-r+1}\Gamma(r+a)\Gamma(b)+c^{-r}\Gamma(r+b)\Gamma(a)}{(1+c)\Gamma(a)\Gamma(b)}, \tag{43}$$

and the third moment is

$$\mu_3' = \frac{\Gamma(3+a)\Gamma(b)c+\Gamma(3+b)\Gamma(a)}{c^3(1+c)\Gamma(a)\Gamma(b)}. \tag{44}$$

The three parameters a, b and c of the NGLD are obtained by numerically solving the three non-linear Equation (36a), Equation (36b) and Equation (36a).

3.5. The New Weighted Lindley Distribution

The *new weighted Lindley* PDF with two parameters (NWL) according to [14] is

$$f(x;b,c) = \frac{-c^2(1+b)^2(1+x)\left(-1+\mathrm{e}^{-cbx}\right)\mathrm{e}^{-cx}}{b(cb+b+c+2)}, \tag{45}$$

where b, c and $x>0$. The CDF of the NWL is

$$F(x;c,b) = \frac{NC}{b(cb+b+c+2)}, \tag{46}$$

where

$$\begin{aligned} NC = & -\mathrm{e}^{-cx}b^2cx+\mathrm{e}^{-c(1+b)x}bcx-\mathrm{e}^{-cx}b^2c-2\mathrm{e}^{-cx}bcx+\mathrm{e}^{-c(1+b)x}bc \\ & +\mathrm{e}^{-c(1+b)x}cx-\mathrm{e}^{-cx}b^2-2\mathrm{e}^{-cx}bc-\mathrm{e}^{-cx}cx+b^2c+\mathrm{e}^{-c(1+b)x}c \\ & -2\mathrm{e}^{-cx}b-\mathrm{e}^{-cx}c+b^2+cb+\mathrm{e}^{-c(1+b)x}-\mathrm{e}^{-cx}+2b. \end{aligned} \tag{47}$$

The average value of the NWL is

$$\mu(a,b,c) = \frac{b^2c+2b^2+3cb+6b+2c+6}{(cb+b+c+2)c(1+b)}, \tag{48}$$

and the variance of the NWL is

$$\sigma^2(a,b,c) = \frac{ND}{c^2(bc+b+c+2)^2(1+b)^2}, \tag{49}$$

where

$$\begin{aligned} ND = & b^4c^2+4b^4c+4b^3c^2+2b^4+18b^3c+7b^2c^2+12b^3+32b^2c \\ & +6bc^2+24b^2+30bc+2c^2+24b+12c+12. \end{aligned} \tag{50}$$

The rth moment about the origin for the NWL is

$$\mu_r' = \frac{NE}{b(bc+b+c+2)}, \tag{51}$$

where

$$\begin{aligned} NE = & -\left(c^{1-r}b^{1-r}\left(\frac{1+b}{b}\right)^{-r}+b^{-r}\left(\frac{1+b}{b}\right)^{-r}c^{-r}r-c^{-r}b^2r+c^{1-r}b^{-r}\left(\frac{1+b}{b}\right)^{-r}\right. \\ & -c^{1-r}b^2+b^{-r}\left(\frac{1+b}{b}\right)^{-r}c^{-r}-c^{-r}b^2-2c^{-r}br-2c^{1-r}b-2c^{-r}b-c^{-r}r \\ & \left.-c^{1-r}-c^{-r}\right)\Gamma(1+r). \end{aligned} \tag{52}$$

The two parameters b and c of the NWL can be found by numerically solving the nonlinear system given by Equation (17a) and Equation (17b).

4. The Double Truncated Lindley Distribution

Let X be a random variable defined in $[x_l, x_u]$; the *double truncated* (DTL) version of the Lindley PDF with one parameter, $f_t(x;c,x_l,x_u)$, is

$$f_t(x;c,x_l,x_u)=\frac{c^2\mathrm{e}^{-cx}(x+1)}{\mathrm{e}^{-cx_l}cx_l-\mathrm{e}^{-cx_u}cx_u+\mathrm{e}^{-cx_l}c-\mathrm{e}^{-cx_u}c+\mathrm{e}^{-cx_l}-\mathrm{e}^{-cx_u}}, \tag{53}$$

where the effect of the double truncation increases the parameters from one to three, see [15]. The double truncated Lindley distribution with scale, which has four parameters, was introduced in [10].

Its CDF, $F_t(x;b,c,x_l,x_u)$, is

$$F_t(x;b,c,x_l,x_u)=\frac{NF}{\left((-1+(-x_u-1)c)\mathrm{e}^{cx_l}+(1+(x_l+1)c)\mathrm{e}^{cx_u}\right)^2}, \tag{54}$$

where

$$\begin{aligned}NF=&-\mathrm{e}^{c(x_l+x_u)}\Big(-(1+(x_l+1)c)^2\mathrm{e}^{-c(x_l-x_u)}-(1+(x+1)c)(1+(x_u+1)c)\mathrm{e}^{c(-x+x_l)}\\&+\left((1+(x+1)c)\mathrm{e}^{c(-x+x_u)}+1+(x_u+1)c\right)(1+(x_l+1)c)\Big).\end{aligned} \tag{55}$$

The average value, $\mu_t(c,x_l,x_u)$, is

$$\begin{aligned}&\mu_t(c,x_l,x_u)\\&=\frac{\left(2+(x_u^2+x_u)c^2+(2x_u+1)c\right)\mathrm{e}^{cx_l}-\mathrm{e}^{cx_u}\left(2+(x_l^2+x_l)c^2+(2x_l+1)c\right)}{-c\left((-1+(-x_u-1)c)\mathrm{e}^{cx_l}+\mathrm{e}^{cx_u}(1+(x_l+1)c)\right)}.\end{aligned} \tag{56}$$

The rth moment about the origin for the DTL, $\mu'_r(c,x_l,x_u)$, is

$$\mu'_r(c,x_l,x_u)=\frac{NG}{\left((1+(x_l+1)c)\mathrm{e}^{-cx_l}-(1+(x_u+1)c)\mathrm{e}^{-cx_u}\right)(r+1)}, \tag{57}$$

where

$$\begin{aligned}NG=&-x_l^{r/2}\mathrm{e}^{-1/2cx_l}\left(c^{1-r/2}+c^{-r/2}(r+1)\right)M_{r/2,r/2+1/2}(cx_l)\\&+\left(c^{1-r/2}+c^{-r/2}(r+1)\right)\mathrm{e}^{-1/2cx_u}x_u^{r/2}M_{r/2,r/2+1/2}(cx_u)\\&+c(r+1)\left(\mathrm{e}^{-cx_l}x_l^{r+1}-\mathrm{e}^{-cx_u}x_u^{r+1}\right).\end{aligned} \tag{58}$$

The three parameters which characterize the DTL can be found in the following way. Consider the sample of stellar masses $\mathcal{X}=x_1,x_2,\cdots,x_n$ and let $x_{(1)}\geq x_{(2)}\geq\cdots\geq x_{(n)}$ denote their order statistics, so that $x_{(1)}=\max(x_1,x_2,\cdots,x_n)$, $x_{(n)}=\min(x_1,x_2,\cdots,x_n)$. The first two parameters x_l and x_u are

$$x_l=x_{(n)},\quad x_u=x_{(1)}. \tag{59}$$

The third parameter c can be found by solving the following non-linear equation

$$\mu_t(c,x_l,x_u)=\bar{x}. \tag{60}$$

5. Application to the IMF

We report the adopted statistics for four samples of stars which will be subject of fit, with the lognormal, the Lindley generalizations and the double truncated Lindley.

5.1. The Involved Statistics

The merit function χ^2 is computed according to the formula

$$\chi^2 = \sum_{i=1}^{n} \frac{(T_i - O_i)^2}{T_i}, \tag{61}$$

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

$$T_i = N \Delta x_i p(x), \tag{62}$$

where N is the number of elements of the sample, Δx_i is the magnitude of the size interval, and $p(x)$ is the PDF under examination.

A reduced merit function χ^2_{red} is evaluated by

$$\chi^2_{red} = \chi^2 / NF, \tag{63}$$

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 [16], which involves the degrees of freedom and χ^2. According to [16] p. 658, the fit "may be acceptable" if $Q > 0.001$.

The Akaike information criterion (AIC), see [17], is defined by

$$\mathrm{AIC} = 2k - 2\ln(L), \tag{64}$$

where L is the likelihood function and k the number of free parameters in the model. We assume a Gaussian distribution for the errors and the likelihood function can be derived from the χ^2 statistic $L \propto \exp\left(-\frac{\chi^2}{2}\right)$ where χ^2 has been computed by Equation (65), see [18] [19]. Now the AIC becomes

$$\mathrm{AIC} = 2k + \chi^2. \tag{65}$$

The Kolmogorov-Smirnov test (K-S), see [20] [21] [22], does not require binning the data. The K-S test, as implemented by the FORTRAN subroutine KSONE in [16], finds the maximum distance, D, between the theoretical and the astronomical CDF as well the significance level P_{KS}, see formulas 14.3.5 and 14.3.9 in [16]; if $P_{KS} \geq 0.1$, the goodness of the fit is believable.

5.2. The Selected Sample of Stars

The *first* test is performed on NGC 2362 where the 271 stars have a range $1.47 M_\odot \geq M \geq 0.11 M_\odot$, see [23] and CDS catalog J/MNRAS/384/675/**Table 1**.

Table 1. Numerical values of χ^2_{red}, AIC, probability *Q*, *D*, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the lognormal distribution, see Equation (66), for different mass distributions. The number of linear bins, *n*, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$\sigma = 0.5$, $\mu_{LN} = -0.55$	37.6	1.86	0.014	0.073	0.105
NGC 6611	$\sigma = 1.03$, $\mu_{LN} = -1.26$	71.2	3.73	1.31×10^{-7}	0.093	0.049
γ Velorum	$\sigma = 0.5$, $\mu_{LN} = -1.08$	55.1	2.84	5.08×10^{-5}	0.092	0.033
Berkeley 59	$\sigma = 0.49$, $\mu_{LN} = -0.92$	54.9	2.82	5.49×10^{-5}	0.11	6.46×10^{-5}

The *second* test is performed on the low-mass IMF in the young cluster NGC 6611, see [24] and CDS catalog J/MNRAS/392/1034. This massive cluster has an age of 2 - 3 Myr and contains masses from $1.5M_\odot \geq M \geq 0.02M_\odot$. Therefore the brown dwarfs (BD) region, $\approx 0.2\mathcal{M}_\odot$ is covered.

The *third* test is performed on γ Velorum cluster where the 237 stars have a range $1.31M_\odot \geq M \geq 0.15M_\odot$, see [25] and CDS catalog J/A + A/589/A70/Table 5.

The *fourth* test is performed on young cluster Berkeley 59 where the 420 stars have a range $2.24M_\odot \geq M \geq 0.15M_\odot$, see [26] and CDS catalog J/AJ/155/44/Table 3.

5.3. The Lognormal Distribution

Let *X* be a random variable defined in $[0,\infty]$; the *lognormal* PDF, following [27] or formula (14.2) in [28], is

$$\mathrm{PDF}(x;m,\sigma) = \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}\left(\ln\left(\frac{x}{m}\right)\right)^2}}{x\sigma\sqrt{2\pi}}, \tag{66}$$

where *m* is the median and σ the shape parameter. The CDF is

$$\mathrm{CDF}(x;m,\sigma) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{1}{2}\frac{\sqrt{2}\left(-\ln(m)+\ln(x)\right)}{\sigma}\right), \tag{67}$$

where $\mathrm{erf}(x)$ is the error function, defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t, \tag{68}$$

see [11]. The average value or mean, $E(X)$, is

$$E(X;m,\sigma) = m\mathrm{e}^{\frac{1}{2}\sigma^2}, \tag{69}$$

the variance, $Var(X)$, is

$$Var = \mathrm{e}^{\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right)m^2, \tag{70}$$

the second moment about the origin, $E^2(X)$, is

$$E(X^2;m,\sigma) = m^2\mathrm{e}^{2\sigma^2}. \tag{71}$$

The statistics for the lognormal distribution for these four astronomical samples of stars are reported in Table 1.

5.4. The Generalizations of the Lindley Distribution

The statistics for the Lindley distribution and its generalizations are reported in the following tables: Table 2 for the Lindley distribution with one parameter, Table 3 for the TPLD, Table 4 for the PLD, Table 5 for the GLD, Table 6 for the NGLD and Table 7 for the NWL. The best fit for NGC 2362 is obtained with the PLD, see Figure 2.

The best fit for NGC 6611 is obtained with the Lindley PDF with one parameter, see Figure 3.

Table 2. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the Lindley distribution with one parameter for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$c = 2.05$	95.57	5.03	3.36×10^{-12}	0.248	2.93×10^{-15}
NGC 6611	$c = 2.94$	38.35	2.01	0.0053	0.077	0.161
γ Velorum	$c = 3.18$	90.59	4.66	5.86×10^{-11}	0.322	3.23×10^{-22}
Berkeley 59	$c = 2.76$	149.6	7.76	6.35×10^{-22}	0.323	5.24×10^{-39}

Table 3. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the TPLD distribution with two parameters for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$b = -0.099, c = 4.2$	72.94	3.83	6.8×10^{-8}	0.129	1.76×10^{-4}
NGC 6611	$b = 0.043, c = 4.32$	59.11	3.06	1.23×10^{-5}	0.098	0.033
γ Velorum	$b = -0.035, c = 5.81$	67.74	3.54	5×10^{-7}	0.14	8×10^{-5}
Berkeley 59	$b = -0.032, c = 4.75$	81.47	4.3	2.35×10^{-9}	0.167	8.62×10^{-11}

Table 4. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the PLD distribution with two parameters for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$b = 2.66, c = 2.28$	28.87	1.38	0.128	0.053	0.39
NGC 6611	$b = 3.33, c = 1.27$	53.53	2.75	8.88×10^{-5}	0.087	0.08
γ Velorum	$b = 4.64, c = 1.64$	106.2	5.67	8.59×10^{-14}	0.16	2×10^{-6}
Berkeley 59	$b = 3.48, c = 1.54$	117.1	6.28	8×10^{-16}	0.187	2.37×10^{-13}

Table 5. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the GLD distribution with three parameters for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$a = 4.80, b = 8.38, c = 12.01$	37.63	1.86	0.016	0.064	0.2
NGC 6611	$a = 1.4, b = 4.8, c = 8$	64.34	3.43	1.96×10^{-6}	0.105	0.017
γ Velorum	$a = 2.53, b = 6.5, c = 0.00046$	83.08	4.53	1.25×10^{-9}	0.15	2.8×10^{-5}
Berkeley 59	$a = 2.2, b = 5.09, c = 1$	100.6	5.56	8.6×10^{-13}	0.179	2.93×10^{-12}

Table 6. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the NGLD distribution with three parameters for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$a = 7.34, b = 1.57, c = 10.61$	48.64	2.5	5.4×10^{-4}	0.075	0.086
NGC 6611	$a = 3.14, b = -0.36, c = 6.24$	111.08	6.18	1×10^{-14}	0.225	9.22×10^{-10}
γ Velorum	$a = 4.19, b = 11.51, c = 12.2$	50	2.58	3.4×10^{-4}	0.101	0.014
Berkeley 59	$a = 5.73, b = 19.57, c = 14.46$	54.14	2.83	8.1×10^{-5}	0.086	3.2×10^{-3}

Table 7. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the NWL distribution with two parameters for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$b = 0.008, c = 3.889$	59.72	3.09	9.85×10^{-6}	0.155	3.33×10^{-6}
NGC 6611	$b = 1.57, c = 3.77$	68.46	3.58	3.81×10^{-7}	0.12	4.2×10^{-3}
γ Velorum	$b = 0.0027, c = 5.86$	79	4.16	6.2×10^{-9}	0.195	1.86×10^{-8}
Berkeley 59	$b = 0.007, c = 5.015$	95.13	5.06	9×10^{-12}	0.19	4.73×10^{-15}

The best fit for γ Velorum is obtained with the lognormal PDF, see Figure 4.

The best fit for the young cluster Berkeley 59 is obtained with the NGLD, see Figure 5.

5.5. The Double Truncated Lindley

The statistics for the DTL with three parameters are reported in Table 8. Figure 6 reports the CDF of the DTL for NGC 6611 which is the best fit of the various distributions here analysed for this cluster.

6. Conclusion

In this paper we explored five generalizations of the Lindley distribution as well

Table 8. Numerical values of χ^2_{red}, AIC, probability Q, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test of the DTL for different mass distributions. The number of linear bins, n, is 20.

Cluster	parameters	AIC	χ^2_{red}	Q	D	P_{KS}
NGC 2362	$c = 1.61$, $x_l = 0.12$, $x_u = 1.61$	156.7	8.86	1.75×10^{-23}	0.115	1.2×10^{-3}
NGC 6611	$c = 2.71$, $x_l = 0.019$, $x_u = 1.46$	45.38	2.31	0.0015	0.061	0.395
γ Velorum	$c = 4.81$, $x_l = 0.158$, $x_u = 1.317$	45.89	2.34	1.3×10^{-3}	0.064	0.269
Berkeley 59	$c = 3.93$, $x_l = 0.16$, $x_u = 2.24$	78.57	4.26	7.73×10^{-9}	0.134	4.35×10^{-7}

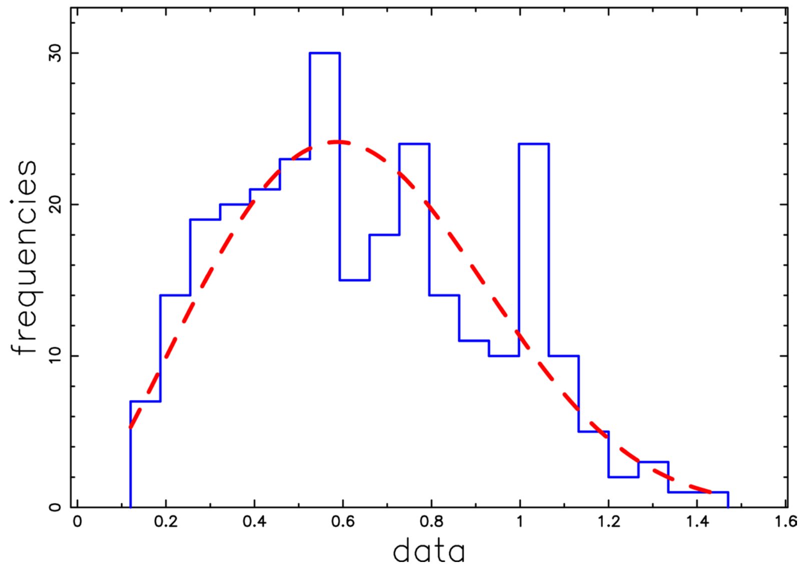

Figure 2. Empirical PDF of mass distribution for NGC 2362 cluster data (273 stars + BDs) when the number of bins, n, is 20 (steps with blue full line) with a superposition of the PLD (red dashed line). Theoretical parameters as in **Table 4**.

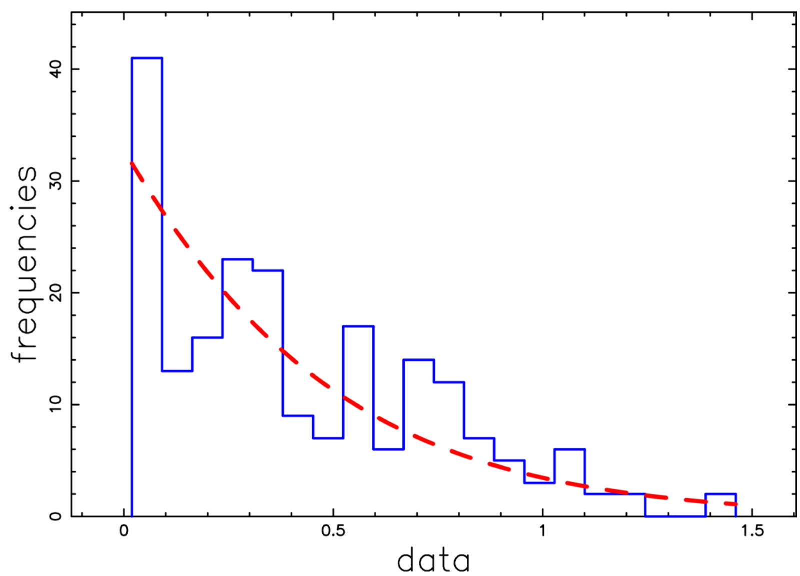

Figure 3. Empirical PDF of mass distribution for NGC 6611 cluster data when the number of bins, n, is 20 (steps with blue full line) with a superposition of the Lindley PDF with one parameter (red dashed line). Theoretical parameters as in **Table 2**.

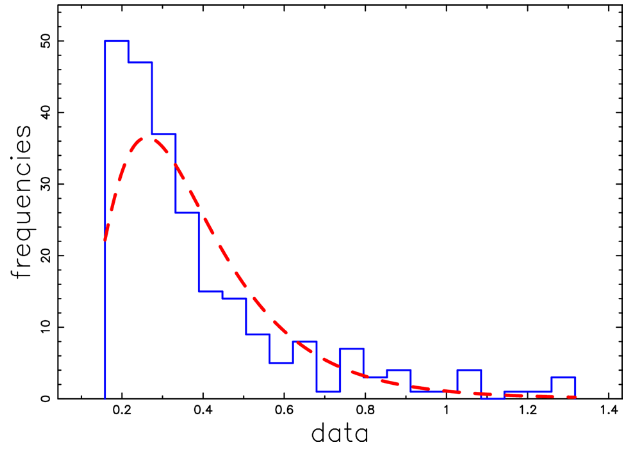

Figure 4. Empirical PDF of mass distribution for γ Velorum cluster data when the number of bins, n, is 20 (steps with blue full line) with a superposition of the lognormal PDF (red dashed line). Theoretical parameters as in Table 1.

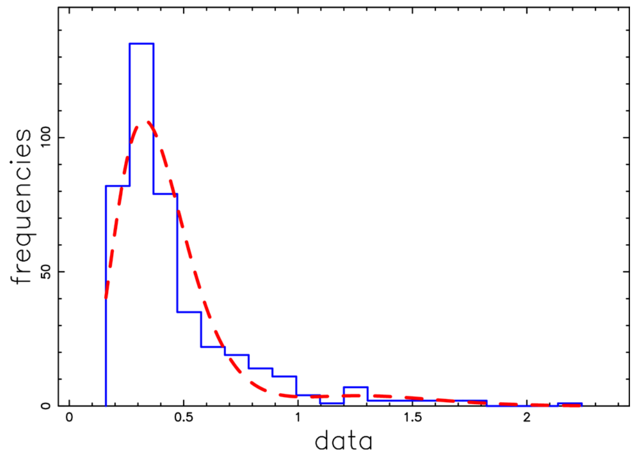

Figure 5. Empirical PDF of mass distribution for the young cluster Berkeley 59 when the number of bins, n, is 20 (steps with blue full line) with a superposition of the NGLD (red dashed line). Theoretical parameters as in Table 6.

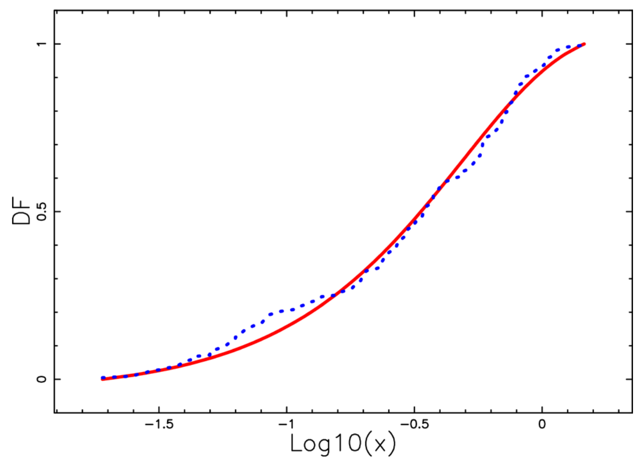

Figure 6. Empirical CDF of mass distribution for NGC 6611 cluster data (blue dotted line) with a superposition of the DTL CDF with one parameter (red line). Theoretical parameters as in Table 8.

Table 9. Best fits: Name of the cluster, name of the distribution, D, the maximum distance between theoretical and observed CDF, and P_{KS}, significance level, in the K-S test.

Cluster	Best fit	D	P_{KS}
NGC 2362	PLD	0.053	0.39
NGC 6611	DTL	0.061	0.395
γ Velorum	DTL	0.064	0.269
Berkeley 59	NWL	0.086	3.2×10^{-3}

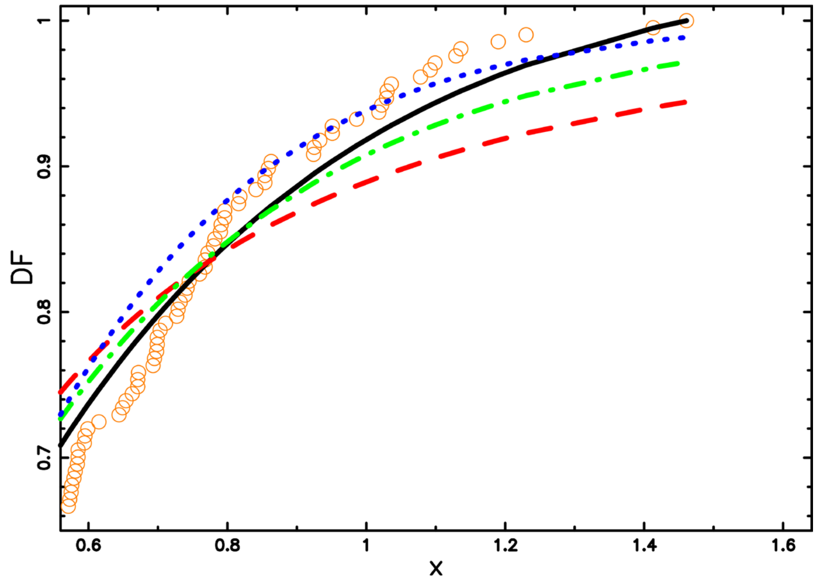

Figure 7. Part of the empirical CDF of mass distribution for NGC 6611 cluster data (orange circles) with a superposition of the DTL CDF with one parameter (black full line), the lognormal (red dashed line), the Lindley with one parameter (green dot-dash-dot-dash line) and the TPLD (blue dot line).

the double truncated Lindley distribution against the lognormal distribution. For each IMF of the four clusters here analysed, the distribution which realizes the best fit is reported in Table 9. The above table allows concluding that the Lindley family here suggested produces better fits than does the lognormal distribution. Figure 7 reports the CDF for NGC 6611 as well as four fitting curves.

Conflicts of Interest

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

References

[1] Lindley, D.V. (1958) Fiducial Distributions and Bayes's Theorem. *Journal of the Royal Statistical Society. Series B* (*Methodological*), **20**, 102-107. https://doi.org/10.1111/j.2517-6161.1958.tb00278.x

[2] Lindley, D.V. (1965) Introduction to Probability and Statistics from a Bayesian Viewpoint. Cambridge University Press, Cambridge.

https://doi.org/10.1017/CBO9780511662973

[3] Shanker, R. and Mishra, A. (2013) A Two-Parameter Lindley Distribution. *Statistics in Transition New Series*, **1**, 45.

[4] Ghitany, M., Alqallaf, F., Al-Mutairi, D.K. and Husain, H. (2011) A Two-Parameter Weighted Lindley Distribution and Its Applications to Survival Data. *Mathematics and Computers in Simulation*, **81**, 1190. https://doi.org/10.1016/j.matcom.2010.11.005

[5] Mahmoudi, E. and Zakerzadeh, H. (2010) Generalized Poisson: Lindley Distribution. *Communications in Statistics Theory and Methods*, **39**, 1785. https://doi.org/10.1080/03610920902898514

[6] Bakouch, H.S., Al-Zahrani, B.M., Al-Shomrani, A.A., Marchi, V.A. and Louzada, F. (2012) An Extended Lindley Distribution. *Journal of the Korean Statistical Society*, **41**, 75. https://doi.org/10.1016/j.jkss.2011.06.002

[7] Merovci, F. and Elbatal, I. (2013) Transmuted Lindley-Geometric Distribution and Its Applications. arXiv Preprint arXiv:1309.3774.

[8] Tomy, L. (2018) A Retrospective Study on Lindley Distribution. *Biometrics and Biostatistics International Journal*, **7**, 163. https://doi.org/10.15406/bbij.2018.07.00205

[9] Shanker, R. and Mishra, A. (2014) A Two-Parameter Poisson-Lindley Distribution. *International journal of Statistics and Systems*, **9**, 79.

[10] Zaninetti, L. (2019) The Truncated Lindley Distribution with Applications in Astrophysics. *Galaxies*, **7**, 61. https://doi.org/10.3390/galaxies7020061

[11] Olver, F.W.J., Lozier, D.W., Boisvert, R.F. and Clark, C.W. (2010) NIST Handbook of Mathematical Functions. Cambridge University Press, Cambridge.

[12] Zakerzadeh, H. and Dolati, A. (2009) Generalized Lindley Distribution. *Journal of Mathematical extension*, **3**, 13.

[13] Ibrahim, E., Merovci, F. and Elgarhy, M. (2013) A New Generalized Lindley Distribution. *Mathematical Theory and Modeling*, **3**, 30.

[14] Asgharzadeh, A., Bakouch, H.S., Nadarajah, S., Sharafi, F., *et al.* (2016) A New Weighted Lindley Distribution with Application. *Brazilian Journal of Probability and Statistics*, **30**, 1. https://doi.org/10.1214/14-BJPS253

[15] Singh, S.K., Singh, U. and Sharma, V.K. (2014) The Truncated Lindley Distribution: Inference and Application. *Journal of Statistics Applications & Probability*, **3**, 219. https://doi.org/10.12785/jsap/030212

[16] 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, UK.

[17] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716. https://doi.org/10.1109/TAC.1974.1100705

[18] Liddle, A.R. (2004) How Many Cosmological Parameters? *Monthly Notices of the Royal Astronomical Society*, **351**, L49. https://doi.org/10.1111/j.1365-2966.2004.08033.x

[19] 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*, *Astronomical Society of the Pacific Conference Series*, Astronomical Society of the Pacific, New York, 508-516

[20] Kolmogoroff, A. (1941) Confidence Limits for an Unknown Distribution Function.

The Annals of Mathematical Statistics, **12**, 461.
https://doi.org/10.1214/aoms/1177731684

[21] Smirnov, N. (1948) Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, **19**, 279.
https://doi.org/10.1214/aoms/1177730256

[22] Massey Frank, J.J. (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, **46**, 68.
https://doi.org/10.1080/01621459.1951.10500769

[23] Irwin, J., Hodgkin, S., Aigrain, S., Bouvier, J., Hebb, L., Irwin, M. and Moraux, E. (2008) The Monitor Project: Rotation of Low-Mass Stars in NGC 2362-Testing the Disc Regulation Paradigm at 5 Myr. *Monthly Notices of the Royal Astronomical Society*, **384**, 675. https://doi.org/10.1111/j.1365-2966.2007.12725.x

[24] Oliveira, J.M., Jeffries, R.D. and van Loon, J.T. (2009) The Low-Mass Initial Mass Function in the Young Cluster NGC 6611. *Monthly Notices of the Royal Astronomical Society*, **392**, 1034. https://doi.org/10.1111/j.1365-2966.2008.14140.x

[25] Prisinzano, L., Damiani, F., *et al.* (2016) The Gaia-ESO Survey: Membership and Initial Mass Function of the γ Velorum Cluster. *Astronomy & Astrophysics*, **589**, A70. https://doi.org/10.1051/0004-6361/201527875

[26] Panwar, N., Pandey, A.K., Samal, M.R., *et al.* (2018) Young Cluster Berkeley 59: Properties, Evolution, and Star Formation. *The Astronomical Journal*, **155**, 44.
https://doi.org/10.3847/1538-3881/aa9f1b

[27] Evans, M., Hastings, N. and Peacock, B. (2000) Statistical Distributions. 3rd Edition, John Wiley & Sons Inc., New York.

[28] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) Continuous Univariate Distributions. Volume 1, 2nd Edition, Wiley, New York.