Proceedings of the XIII Bulgarian-Serbian Astronomical Conference (XIII BSAC) Velingrad, Bulgaria, October 3-7, 2022
Editors: Evgeni Semkov, Milan S. Dimitrijević, Momchil Dechev and Zoran Simić
Publ. Astron. Soc. "Rudjer Bošković" No 25, 2023, 63-67

INTERGRAM INSTEAD HISTOGRAM: PROGNOSIS OF THE EXTREME POPULATION VALUE

TSVETAN B. GEORGIEV

Institute of Astronomy and NAO, Bulgarian Academy of Sciences, 72 Tsarigradsko Shaussee Blvd, Sofia 1784, Bulgaria
E-mail: tsgeorg@astro.bas.bg

Abstract. The intergram is a proxy of the histogram. In contrast to the histogram and similar to the cumulative distribution, the intergram is unique for the studied data sample. The intergram is built by the method of the shortest intervals containing specified part among the sorted data. The extrapolation of the bright leg of interrgram in log-log coordinates may prognosticate the extreme value of the studied sample for 1% of the population. Four examples of extrapolation and comparison between the prognosis and observation are shown. The distribution of the diameters and masses of 76 bodies of the Solar system give evidences about one missing big planet. The estimated maximal rotation velocities among of 639 disk galaxies, 242 km/s, is in good agreement with the observations. The habitant numbers of 297 Bulgarian towns shows absent of town with about 525 000 habitants. The excess is the metropolis, containing 2.5 times more habitahts. Also, the brightest stars of the cluster Praesepe pose apparently about 6.2 mag. However the intergram of 196 cluster members prognosticates shows that in the beginning 1% of the stars had to had about 4 mag. About 5.6% of the sample is already out of the MS.

1. INTRODUCTION

Many astronomical values show asymmetric distributions, sometimes with triangular shape. Such are the Wolf number of the solar spots, the light fluxes of flickering stars (Georgiev et al. 2022). Such are also the values of some physical parameters – masses and diameters of planetary bodies, luminosities and diameters of galaxies, etc. These asymmetric distributions seem to be so fundamental in the nature like the triangles in the geometry. That's why we proposed a special proxy of the histogram, called intergram, with 21 applications (Georgiev 2022).

Figure 1 explain the building of intergram of the distribution of 76 bodies of the Solar System by their log-diameters. These bodies (planets, satellites, asteroids) are selected to be larger than 180 km. The Sun and 4 giant planets are excluded. The intergram is shown by its two appearances in Figs.1(a) and 1(b).

In practice, the scientist does not know the density function of the studied population. He has only a "good sample" of *n* data, sorted by increasing. Such data are shown by dots along the ordinate in Fig. 1(a) and along the abscissa in Fig. 1(b). A relevant histogram, that represents approximately the density function of this sample is shown in Fig. 1(b) by dashed curve. The appearance of the histogram depends on the step and binning interval, specified by the user. Another representation of the sample density, the inerGram, is possible through the method of the Minimal Interval of Data, containing *p*% of the sorted data (*p*%-MID). This method is mentioned by Rousseeuw & Leroy (1987) in a connection with their "method of the least trimmed squares". Here we draw the shape of the SP through the bounds of the *p*%-MIDs.

2. BUILDING THE INTERGRAM

Figures 1 showS details in the building of the intergram.

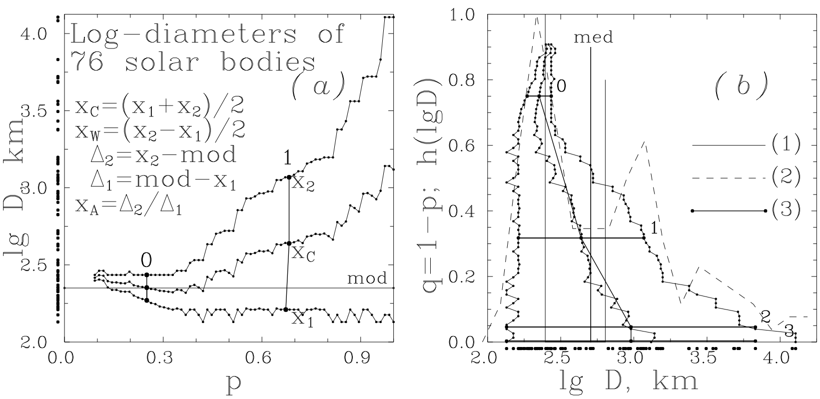

Figure 1: Intergram for solar bodies. (a) Definition form of the intergram tripod $f(\lg D_k)=f(p)$ with localparameters: x_1 and x_2 – bounds, x_C – center, x_W – width, x_A – asymmetry; (b) Working form of the intergram $q_k=f(\lg D)$ where $q=1-P$, D is the diameter in the sorted sample. (1) – intergram; (2) – histogram; (3) – intergram skeleton (see Georgiev 2022).

Let us to have a sorted data sample $\{X\}$: x_1, x_2, ..., $x_{n.}$ Preliminary Step 0 is a choice of the initial minimal part of data, p_1, for omitting the top of the inergram. Often the modal part of the sample dencity is complicate and p_1 must be enough large. In Fig. 1 p_1=0.1 (6 data). A better value may be p_1=0.25 (Figs. 2, 3). In Fig.4 p_1=0.5. Step 1 is the deriving of the 1st minimal interval $[x_{1,}\, x_2]$ among the sorted data, containing $m_1=p_{1\times}n$ data. Step 2 is deriving the 2nd minimal interval, containing $m_2=m_1+1$, data with relevant part of data $p_2=m_2/n$. Step 3 concerns by the same way p_3=m_2+1 etc., till m_{k-1}=n-1. In the end the case m_k=n may be added for p_k=1. The number of the steps (of used MIDs) is n-m_1+1.

The result in Fig.1(a) is the intergram by definition. It is a "lie" tripod graph with abscissa p and 3 ordinates – MID bound points x_1 and x_2, as well as their average value x_C. Here 70 abscissa points are used.

Figure 1(a) contains two vertical segments which showthe MIDs above the abscissae p=0.25 (0) and p=0.6826 (1). The middle point of the 25%-MID is used as a robust mode estimator for the syudied distribution. The 68.26%-MID corres-ponds to the part p for 1 sigma interval of the normal distribution. The vertical segments for 2 MIDs in Fig. 1(a) are a part of the interval skeleton of the sample. The full skeleton contains also segments for 2 sigma and 3 sigma intervals of the samle in Fig. 1(b). The skeleton gives morphological parameters of the dis-tribution (see Georgiev 2022).

It is convenient to represent the standing intergram tripod, as in Fig. 1(b), simi-lar to histogram. For this purpose we change p with q=1 − p and yye use q as ordinate. In Fig.1(b) the dashed curve is the histogram. The verticals show the positions of the mode, median and average.

In principle the distribution with positive skewnes has exponentially decreasing right leg. Under a log-ordinate this leg becomes quasi-linear. Thus the relevant fit may prognosticate the quantity of the population for given abscissa value, for example for 1% of the population (see below).

3. APPLICATIONS OF THE INTERGRAM PROGNOSIS

Figures 2 show intergrams of Solar System bodies larger than 180 km.

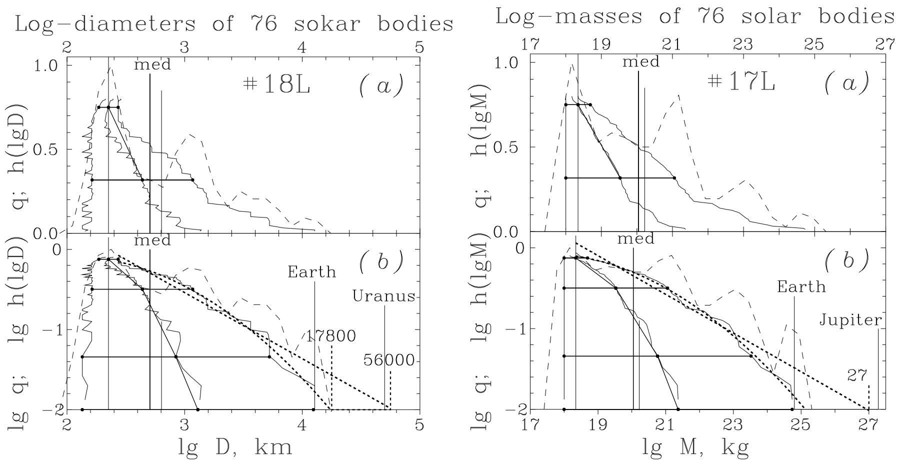

Figure 2: Intergrams of the solar bodies by diameters (left) and masses (right). Thin curves and lines show the integrams and their skeletons, dashed curves show histograms, short dashed curves show linear and quadratic fits of the intrgrams in log-log coordinates.

The relevant histograms (dashed curves) show excess of bodies with sizes about 1000 km and masses about 10^{21} kg. Such bodies are similar to the satelites of Ura-nus. The intergrams do not show suc details. The right linear extrapolations of the down intergrams (in the left without Sun and giant planets, in the right without Sun) prognodtiecte that 1% of the population must contains one big planet arge as Uranus and as massive as Jupiter, respectively. The intergram of the densities predict a planet with mean density about 8.6 g/cm^3 Georgiev (2022).

Figure 3, left diagrams, shows intetgrams of the the rotational velocities of 639 disk galaxies. This example concerns a distribution, depending on numerous na-tural factors. In (b) the right leg sows linearity from 10 to 200 km/s. The prog-nostic value for 1% MID is V_R=242 km/s. This prognosis is in good agreement with the observations. Only three more massive galaxies are present in the sample, including 1 with V_R =277 km/s.

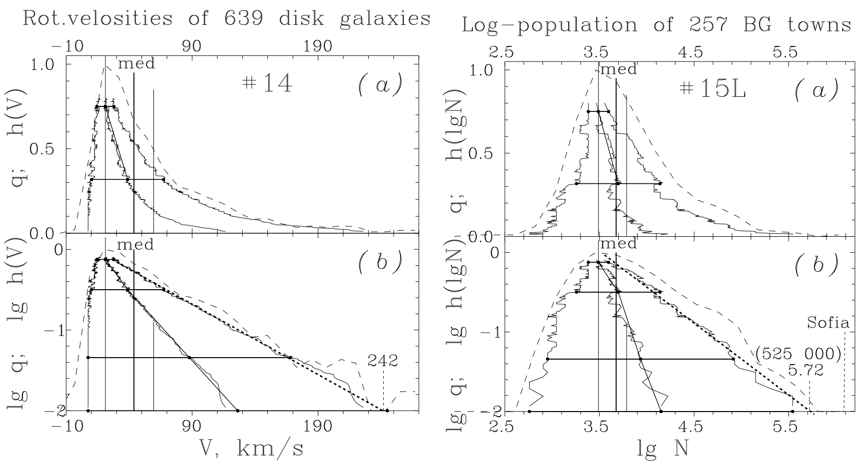

Figure 3: Left diagrams: Intergrams of the rotation velocities of disk galaxies; Right diagrams: Intergrams of the habitants of Bulgarian towns. Thin curves and lines – integrams and skeletons; Dashed curves – histograms; Short dashed curves – linear fits and extrapolations of the bright legs of the intrgrams in log-log coordinates.

Figure 3, right diagrams, shows the intergrams of the habitants of 257 Bulga-rian towns. This is not natural distribution It depends mainly on the human factors. The prognostic value for 1% MID must be 525 000, more thah Plovdiv or Varna, which have about 330 000. However, the metropolis with 1 243 000 habitants.is about 2.5 times larger in respect with the prognosis.

Figure 4 shows an application of the intergram for prognosis of the magnitude of the brightest stars in the beginning time of the open cluster Praesepe (M 44). In the catalogue of Wang et al. (1995) 196 stars are members of the cluster. Their CMD is represented in Fig. 4(a). V-intergram of these stars is shown in Frig.4(b).

Thh3 intrgram in Fig. 4(b) is natural sample distribution, depending on the evolutionary status of the cluster. A part of the brightest stars are leaved the main sequence. The prognostic value for 1% MID, given by the linear fit is about 4 mag. The gradient of the fit is -0.276. The breakdown level of the fit is lg q(V) = -1.25,, i.e. 5.6% of the studied sample sre leaved the MS of the CMD and 94.4% remain. The use of B-magnitudes gives the same result, 4 mag but the gradient is 0.259.

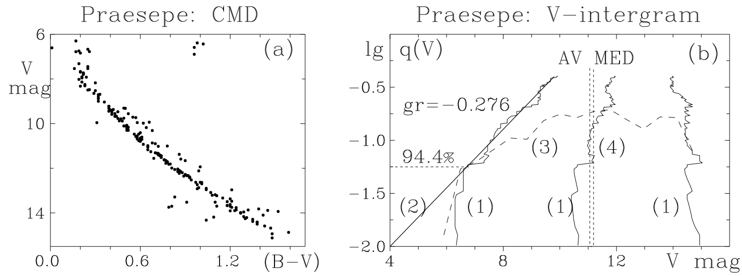

Figure 4: (a) Apparent color-magnitude diagram of 196 members of the cluster Praesepe; (b) Log-intergram of the V-mag-nitudes: 1 – legs of the intergram; (2) fit and extrapolation of the bright leg; (3) histogram; (4) –average abd median of the distribution.

4. CONCLUSIONS

The intergram as a procy of the histogram is an useful unique represent-tation of the data distribution. The parameters implemented in that are implement-ted in Fig. 1(a) may be used for morphological functions of the intergram – beha-viors of the center $C(p)$, width $W(p)$ and asymmetry $A(p)$. The gradients of these func-tions give classifications of the intergrams (see for details Georgiev, 2022).

In this work the intergram is used as tool for prognosticate the extreme values in the studied population.

Acknowledgements

The author is grateful to O. Stanchev and A. Mutafov for the attention and help.

References

Georgiev Ts.: 2022, *Bulg. Astron. J.*, 36, 70-81.

Georgiev,Ts., Boeva S., Latev G., Semkov E., Stoyanov, K. A., Tsvetkova S.: 2021, *Bulg. Astron. J.*, 34, 10-29.

Rousseeuw P. J., Leroy A. M.: 1987, *Robust Regression and Outlier Detection,* John Willy & Sons.

Wang J.J., Chen L., Zhao J.H., Jiang P.F.:1995, *Astron. Astrophys. Suppl.*, 113, 419-430.