Characterization of asymmetric data distributions by intergrams instead histograms

Tsvetan B. Georgiev
Institute of Astronomy, Bulgarian Academy of Sciences, Sofia 1784, Bulgaria
tsgeorg@astro.bas.bg
(Submitted on 2.06.2021; Accepted on 20.07.2021)

Abstract. The asymmetric distributions of astronomical values pose complicate shapes. Sometimes they have heavy tails, other times they do not posses event light tails. Then the ordinary statistical parameters (mean, standard deviation, skewness, kurtosis) may not be informative enough and visualization of the distribution by a histogram is necessary. However, in contrast with the empiric cumulative function, which is unique for the available data, the histogram is not unique. Its appearance depends on the user supplied scanning step and binning interval. A better representation is necessary. This work is devoted to the characterization of unimodal asymmetric data distributions by a special figure, intergram and its parameters, introduced here.

The intergram approach is based on the shortest interval among the ordered data (SID), containing specified part p, $0 < p < 1$, of the data, (p-SID). Numerous p-SIDs produce an intergram – a tripod figure containing the bounds of the SIDs plus their middle points as dependences on p of the p-SID (Fig. 2). The intergram is unique for the available data and it may be the alternative of the histogram. The intergram gives specific gradients (Fig. 3) parameters of the morphology of the data distribution. It may prognosticate typical high value for 1% of the studied population (Figs. 3).

This intergram approach is applied on 21 distributions – time series of the Wolf numbers, light variations of irregular stars etc., as well as masses, sizes, luminosities, etc., of astronomical objects. The data distributions collected here have various origin and observing score, but their ordinary and intergram parameters follow common statistical dependences, i.e. these distributions are in principle mutually similar (Figs. 5–7). In particular, an well pronounced quadratic dependence between the skewness and the kurtosis is found (Fig. 6a). Some interesting particularities of the regarded data distributions are briefly commented. The distributions are represented on-line as pdf material.

Key words: Statistic parameters and estimations; Histogram and intergram

Introduction

Many astronomical values show asymmetric distributions (ADs), sometimes even with triangular shape (Fig. 1). Such are the time series of the variations of the Wolf number of the solar spots, the variations of the light flux from an irregular variable star, etc. Such are also the values of physical parameters like the masses of planetary bodies, the luminosities of stars, the diameters of galaxies, etc. These ADs seem to be so fundamental in the nature as the triangles in the geometry. This work is devoted to characterization and comparison of data distributions (DDs) through a special tool – the intergram.

Recently we found quasi-periods in the flickering of the symbiotic star RS Oph (Georgiev, Zamanov, Boeva et al., 2020). The DD of the light fluctuations were compared with the well known DDs of the monthly Wolf number and with the (presumed known) weekly visiting number of the Smolyan Planetarium. (The visiting number does not correlate with the Wolf number.)

These DDs occurrence is very similar. Figure 1 illustrates this by 6 kinds of histograms: along the linear ordinate (a, c, e) and along the logarithmic ordinate (b, d, f), as well as over the linear abscissa (a, b), over the root-squared abscissa (c, d) and over the logarithmic abscissa (e, f). The relevant time series are shown in diagrams (g).

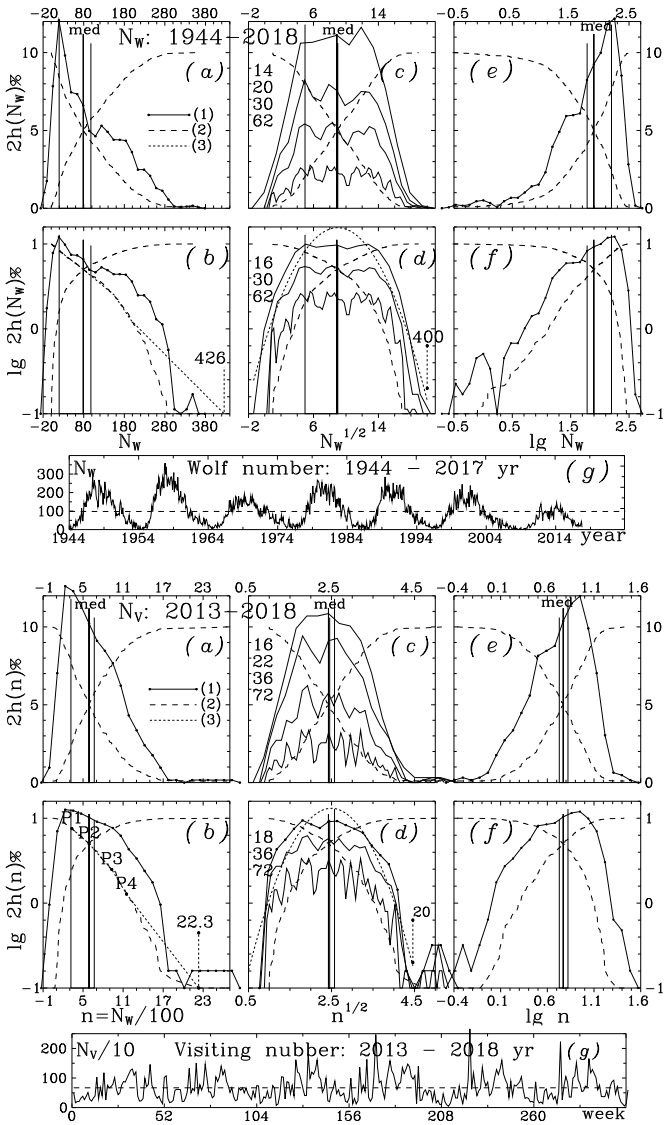

Fig. 1. A curious similarity between the histograms of the Wolf monthly number N_W (top panels, 886 data) and the Planetarium weekly visiting number N_V (bottom panels, 313 data, where $n = N_V/100$) in various coordinates. (See the text.) The lines are: (1) – histograms $h(.)$, in percents $\times 2$ and vertical markers of the mode, the median and the average value; (2) – cumulative functions scaled between 0 and 10, (3) – fits of the reverse cumulative function in (b) and quasi-normal DD in (d). These fits prognosticate the values of N_W or N_V for 1% of the relevant populations. The numbers in (c, d) are the numbers of the few histogram columns (points) in the DD range.

Figure 1 shows that: over linear target parameters both DDs have positive skewness (a, b); over logarithmic parameters the skewness becomes negative (e, f); and over square root of the parameter both DDs show quasi-normal shapes (c, d). These unusual complexities and similarities of these ADs motivated this wok. (See the comments about these DDs in Section 3.)

In comparison with the normal distribution (ND), the AD has a complicated shape. Sometimes it has heavy tail(s) (non-typical large deviations from the mean value) other times it does not event light tail(s). Usually the AD needs of visualization by a histogram. However, the appearance of the histogram depends on the user supplied scanning step and binning interval. So, by contrast with the empiric cumulative function, the empiric histogram is not unique. Better representation is necessary.

The goals of this paper are to describe the specific characterization of the DD through the intergram with its skeleton, functions and gradients (Sections 1 and 2), to represent intergrams and their parameters for 21 DDs (Section 3 plus on-line pdf Appendix), and to compare ordinary and intergram parameters of the selected DDs (Section 4).

Some abbreviations are used in the text, as follows: AD – asymmetric distribution, AV – average value, DD – data distribution; ND – normal distribution, SD – standard deviation, SP – studied population, p%-SID - the shortest interval in the ordered data, containing p% of the data, 1%-PHV – prognostic high value for 1% of the SP.

1. The intergram and its building

Figure 2 contains the intergrams of the distribution of 76 Solar System bodies by their log-diameters. The bodies (planets, satellites and asteroids) are larger than 130 km. The Sun and the giant planets are excluded. Full representation of this case is given in examples #19, #17L and #18L in the on-line Appendix.

In practice, the scientist does not know the density function of the studied population (SP). They have only n sorted data $x_1, x_2, ..., x_n$, considered as a "good sample" from the SP. Such data are shown by dots along the ordinate in Fig. 2a and along the abscissa in Fig. 2b. A relevant histogram that represents approximately the density function of these SP is shown by dashed curve in Fig. 2b.

Another visualization of the DD is possible through the method of the minimal interval, containing p% of the sorted data (p%-SID). This method is mentioned by Rousseeuw & Leroy (1987) in a connection with their "method of the least trimmed squares". The content of every SID may be characterized by all statistical parameters, like the whole DD. But here the goal is the use of all available SIDs for drawing the shape of the DD through the bounds of the p%-SIDs.

Figure 2a contains two vertical segments showing the SIDs above the abscissae $p = 0.25$ and $p = 0.6826$. In this work the middle point of the 25%-SID is used as a robust mode estimator for the DD. The 68.26%-SID corresponds to the $\pm 1\sigma$ interval of the ND. The half of the size of the 68.26%-SID, divided by the median of the DD, is used here as a relative median standard deviation S_M, replacing the ordinary standard deviation (SD) of the DD as more robust

value. The segments fo 2 SIDs in Fig. 2*a* are a part of the "interval skeleton" of the DD, these 4 "ribs" are shown in Fig. 2*b*.

Let x=lg D in Fig. 2*a*. Then the basic parameters of the segment of the 68.26%-SID are the bounds x_1 and x_2, as well as the average value (AV) of the bounds $x_C = (x_1 + x_2)/2$. These 3 parameters, arranged in dependence on p as $x_1(p)$, $x_C(p)$ and $x_2(p)$, for $0 < p < 1$, build the intergram in its definition form (Fig. 2*a*).

The algorithm for the intergram building follows. In the preliminary step the initial minimal part of data, p_0, should be chosen. Here $p_0 = 0.1$, but usually $p_0 = 0.25$. (Often the modal part of the DD is complicated and p_0 must be large enough.) Then the respective initial size of the SID, as number data, $m_1 = p_0 \times n$, should be derived. The first step is derivation of the value of p_1 that corresponds just to the number m_1, $p_1 = m_1/n$, and derivation of the bounds x_l and x_r of the p_1-SID, containing m_1 data, follows. The last action is realized by scanning of the (sorted) data row X by data window with size m_1, by step of one data number. The second step begins by increasing m_1 to $m_2 = m_1 + 1$ and derivation of the respective $p_2 = m_2/n$. After this the bounds x_l and x_r of the p_2-SID are found again by scanning of the data row, etc., until $m_k = n - 1$.

The result is the intergram – a tripod figure containing the SID bound points and their AV in dependence on p. In Fig. 2*a* the number of the used SIDs is $n - m_1 + 1$ points. For the collected $n = 76$ solar bodies, $k = 70$ SIDs are used, containing from 6 to 75 data points.

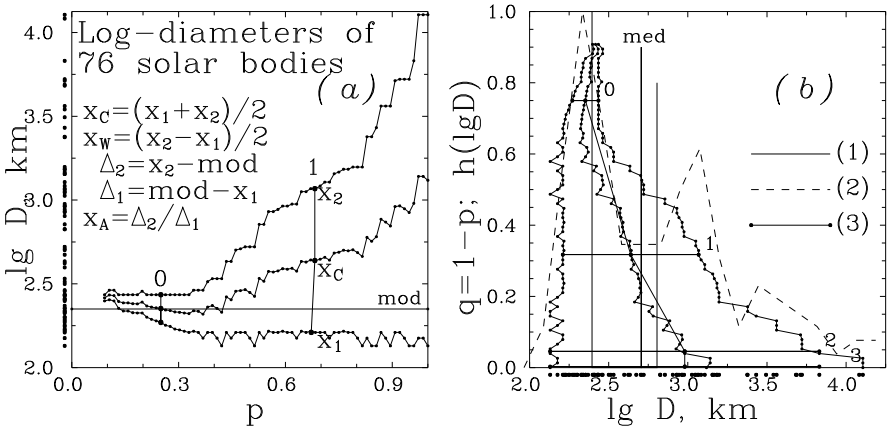

Fig. 2. Intergrams of the log-diameters lg D of Solar bodies. (*a*) – the definition form, in dependence on p ($0.1 < p < 1$); (*b*) – the ordinary form, in dependence on $q = 1 - p$. The lines show: (1) – intergrams and vertical markers of mode, median and AV; (2) – histogram h(lg D), scaled to unit maximal value; (3) – the SIDs ("ribs") No.0 and No.1 in the diagram (*a*), as well as the interval "skeleton" with "ribs" No. 0–3 in the diagram (*b*). The dots in the diagrams *a* and *b* show the ordered available data, lg D. (See the text.)

Figure 2*a* shows the definition form of the intergram together with two used SID points over the abscissa, p. However such reclined tripod is not useful. Figure 2*b* shows the intergram in its ordinary form, beside the ordinate $q = 1 - p$. This form is (i) compatible with the histogram and (ii) it allows logarithmic ordinate (Fig. 3*b*).

Note that in this paper the histograms are more wide than the intergrams because the binning interval is two times larger than the scanning step and because of the scaling of the histograms to unit maximal value. Note that in Fig. 2*b* the histogram has large scale jags, but the intergram is more smooth.

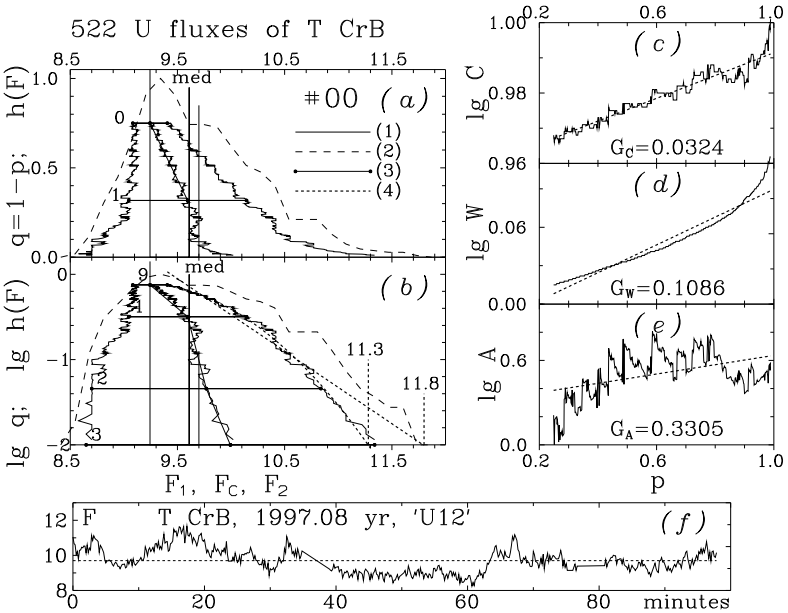

Fig. 3. Intergrams of the flickering flux F of T CrB (f) along linear (a) and logarithmic (b) ordinate. $F_1(q)$, $F_C(q)$ and $F_2(q)$ are "the legs" of the intergram. The quadratic and linear fits of the largest leg of the intergram prognosticate for 1% PHV 11.3 and 11.8. The lines are: (1) – intergrams $h(F)$ and lg $h(F)$, vertical markers of the mode, the median and the AV, (2) – histograms, (3) – skeleton, (4) – fits. Diagrams (c), (d) and (e) represent the morphological functions lg $C(p)$, lg $W(p)$ and lg $A(p)$ (Eq. 1). (See the text.)

Figure 3 is example #00 in this paper. It contains the DD of the U flux (F) of the symbiotic double star T CrB, expressed in units of 10^{-15} erg cm^2 s^{-1} A^{-1} (Georgiev, Boeva, Latev et al., 2021). In diagram 3(a) the intergram and the histogram are shown along linear ordinates, $q = 1-p$ and $h(F)$. The histogram is scaled to unit height in its maximum. In diagram 3(b), the intergram and the histogram are drown along logarithmic ordinates, lg q and lg $h(F)$ respectively. Diagrams 3(c), 3(d) and 3(e) show the intergram functions lg $C(p)$, lg $W(p)$ and lg $A(p)$ (Eq. 1), with their linear fits and gradients G (see below).

2. The skeleton indices, function gradients and intergram prognoses

The intergram in Fig. 2*b* and Fig. 3*b* includes two more special SIDs than the intergram in Fig. 2*a* and Fig. 3*a*. The SID No. 2 contains 95.44% of the data like the $\pm 2\sigma$ interval of the ND. The $\pm 3\sigma$ interval, containing 99.72% of the ND, is not useful because in practice it depends on the presence or absence of tails in the DD. For this reason in this paper the SID No. 3 contains 99% of the data and it is a good 1% zero level for the diagrams with log-ordinates (diagrams (*b*)).

The SIDs 0–3 in diagram (*b*) draw intergram skeleton with 4 ribs. Sometimes the lowest (almost vertical) segment of the skeleton, 2–3, is deviated in respect to the other segments. Such circumstance is seen better beside log-ordinate and this is a direct evidence of excess or deficit of tails in the DD. In Fig. 2*b* the deviated segment 2-3 (seen better in #18) hints to a deficit of one more large planet. In Fig. 3*b* such a circumstance is not seen. The skeletons are present in the diagrams (*b*) in Fig. 3 and in the (*b*) diagrams in the Appendix.

The SIDs No.1 and No.2 diagrams (*b*) may be used naturally for characterization of the intergram body, i.e. for deriving of robust indices of the SP. These two p SID are chosen for easy comparison with the ND. They give robust skeleton indices of the intrgtam shape of the DD – the average width and asymmetry, as well as the average rates of change of the width and asymmetry. These indexes are compatible with ordinary skewness, kurtosis, etc. (Georgiev, 2020, G2020, figs. 6, 8, 9). The mentioned skeleton indexes are independent on the tails of the DD, but they are derived by only four points of the intergram. For this reason the present paper is concentrated on the intergram functions (trends), derivable by (almost) all SIDs of the DD.

One local parameter of every p%-SID is the center of the SID: $x_C = (x_1 + x_2)/2$ (Fig. 2*a*). Other useful parameters are the semi-width $x_W = (x_2 - x_1)/2$ and the asymmetry in respect to the mode of the DD, $x_A = \Delta_2/\Delta_1$, where $\Delta_2 = x_2 - x_{\rm med}$ and $\Delta_1 = x_{\rm med} - x_1$. (In this paper the mode is the center of the 25%-SID.)

Therefore, the morphology of the DDs may be characterized by 3 morphological functions of the intergram: center-function $C(p)$, width-function $W(p)$, and asymmetry-function $A(p)$:

$$C(p) = x_{\rm C}/x_{\rm mod}, \qquad W(p) = x_{\rm W}/x_{\rm mod}, \qquad A(p) = \Delta_2/\Delta_1. \tag{1}$$

Note that while $A(p)$ is dimensionless by definition, $C(p)$ and $W(p)$ may become relative and universal after scaling trough division by the mean position of the DD, here $x_{\rm mod}$.

Logarithms of these functions show quasi-linear dependences on p on the diagrams c, d, e in Fig. 3 and in the Appendix. Comparisons between gradients of these functions with the statistic parameters and by themselves are shown in Figs. 5, 6 and 7. Figure 4 shows schemes of intergrams with different shapes,characterised by different function trends.

Another useful application of the intergram in the case of positive AD is the prognosis of 1%-PHV shown in Fig. 3*b*. Along the logarithmic ordinate the largest led of the intergram (with positive asymmetry), $F_2(\lg\ (q))$, becomes

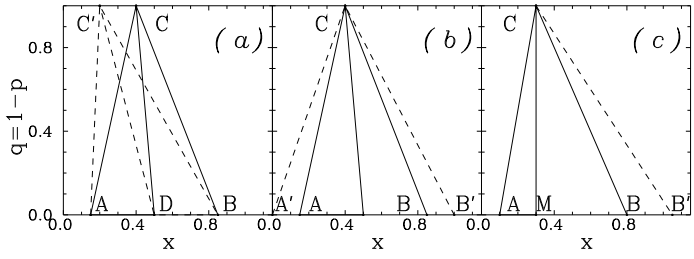

Fig. 4. (a) The intergram ABC' is more tilt (skewed) than ABC and its gradient G_C is larger. (b) The intergram A'B'C has more wide base than ABC and its gradient G_W is larger. (c) The intergram AB'C is more asymmetric than ABC and its gradient G_A is larger.

quasi-quadratic or quasi-linear. The respective fit crosses the abscissa (the 1% level) on $F = 11.3$, or $F = 11.8$. This is the alternative of the prognosis, based on the reverse cumulative function (Fig. 1*b*).

Such prognoses are marked and inserted in numerous diagrams in the Appendix. Generally, Fig. 4*b* shows that the largest positive flickering deviations seem somewhat suppressed like the suppression of the largest Wolf numbers in Fig. 1. Diagrams *c*, *d* and *e* in Fig. 4, as well as in the Appendix, show the intergram functions and their linear fits. The relevant gradients are summarized in Table 1.

3. Examples of data distributions

Figure 3, plus 20 panels in the on-line Appendix represent intergrams, histograms, intergram skeletons and intergram gradients of various DDs. Data about the DDs and their statistical and intergram parameters are summarized in Table 1. The examples in this work are intended mainly for illustrations. However, brief comments of some interesting particularities of the DDs are given below. In the Appendix, the first 10 examples contain linear abscissae (target parameters) and the next 10 – logarithmic abscissae.

The shapes of the collected DDs take 3 types of asymmetry. Positive skewness have 13 DDS: #00, #01, #02, #03, #04, #07, #13, #14, #15, #19, #04l, #17L, #18L. Quasi-symmetric shapes show 5 DDS, 3 NDs – #11, #12, #20 plus 2 log-NDs – , #13L, #14L. Negative skewness have 3 DDs: #01L, #02L, #03L.

The collected DDs include 3 types by the coverage of the distribution through the observations/measurements – full (exhaustive), occasional (provisional) and limited (preliminary bounded). 12 DDs are exhaustively observed. They seem to cover the ranges of the target parameter and the ranges of the density function. The respective 1%-PHV are collected in Table 2.

The distributions of the Wolf number and the Planetarium visits, W and N, are the most interesting among the exhaustively observed DDs. They are similar in all diagrams (Fig. 1, panels #01, #02, #03, #01L, #02L, #03L). The distributions over square root of the numbers are quasi-normal. Another such pair of DDs is not found in this work. The distribution of the Wolf

number has been modeled and simulated by Noble & Wheatland (2011). The largest values of W are suppressed as if some unit area of the solar surface must support limiting Wolf number. The deviated bottom segment 2–3 of the skeletons confirms this. The largest values of N are smaller and clearly suppressed by the capacity of the Planetarium. It seems the Wolf number corresponds to the total area of the spots better than to the number of spots. May be the Planetarium visiting number corresponds to the number of the visiting groups.

Example #04 is based on the B-flux of the LBV AF And in 1917–1989 (Ganchev et al., 2017). The positive tail in the histogram b show the excess of high fluxes, due to the erruption in 1970–1980. For this reason the lowest segment of the skeleton, 2–3, is turned to the right.

The atmosphere extinction over Rozhen NAO in B-band, in 138 nights in 2005–2008 is derived by Dimitrov (2009). The mode, median and AV are 0.13 mag, 0.175 mag and 0.20 mag, respectively (#07).

Six DDs are occasionally observed parameters of disk galaxies – #11, #12, #13, #14, #13L and #14L. The sample (from Hyper LEDA) includes 639 disk (star-forming) galaxies at distances up to 16 Mpc. The galaxies are selected to have morphological types Sab–Ir and inclination angles i 30^o–85^o. Amplitudes of the rotational velocities, which are the results of the log-masses of the galaxies, are badly estimated if $i < 30^o$. The luminosity of the galaxy seems to be badly estimated if $i > 85^o$.

Any sample of galaxies, limited by magnitude (i.e. by observing depth), has bias: the presence of the faint galaxies decreases when the distance increases. Therefore, the apparent distribution of the galaxies may be observed fully, but the intrinsic distribution – not. In result the apparent mean of every distribution is shifted toward the more luminous galaxies. The left (faint) edges of all DDs of the galaxies are badly defined. However, the respective DDs support the correlations and dependences in Section 4.

The 1%-PHV are entered in the diagrams about the galaxies and collected in Table 2 in the Appendix. The 1%-PHV for B surface brightness is -17.2 mag/kpc^2 and 3 galaxies have -17.3 – 16.4 mag/kpc^2. The 1%-PHV for B luminosity is -22.6 abs.mag, but the most luminous galaxy in the sample has -21.6 abs.mag, i.e. it is 2.5 times less luminous. Note the long quasi-linear leg of the intergrams in #13 and #14. The 1%-PHV for B disk diameter is 36–40 kpc and 7 galaxies have 36¡d¡45 kpc. The 1%-PHV for the amplitude of the rotational velocity is 242–316 km/s and 6 galaxies have 245–278 km/s. Generally, the 1%-PHV prognoses correspond well to the observations.

Further, tree DDs, #17L, #18L and #19, contain ADs of masses, diameters and densities of solar bodies, with diameters larger than 130 km. These DDs have big linear ranges of their masses and diameters. So, only logarithms of these parameters may be used for histograms. The local maxima of the DDs about lg $M = 21$ and lg $D = 3$ due to numerous bodies with diameters of about 1000 km and density about 2 g/cm^3. The available DDs about solar bodies are bounded preliminary on the left, but they are cut naturally on the right.

Note the deviated bottom segments 2–3 of the skeletons in #17L, #18L and #19, in the intrgrams (b). The fits prognosticate lg $M = 25.1 - 27$ for 1%-HPV, while the Earth and Uranus have 24.8 and 26.0, respectively. The prognoses for the size are $D = 17800 - 56000$ km, while the Earth has 12756

km and Uranus, with its giant atmosphere, has 51000 km. The intergram prognoses by the linear fits seem to be too high.

The intergram of the density distribution (#19) prognosticates $\rho = 6.1$ g/cm^3. Generally, according to the DDs #17L, #18L and #19 at least one big planet should be present in the Solar system.

4. Correlations between the ordinary and intergram indices

The ordinary indices (of Pearson) that characterize the shape of the DD are the skewness P_1 and the kurtosisP_2. They are based on the statistical central moments of the DD M_k, $k = 1, 2, 3, 4$. The k-th central moment and the second central moment, the dispersion, are:

$$M_k = \Sigma(\Delta x_i)^k/n = \mu_k, \qquad M_2 = \Sigma(\Delta x_i)^2/n = \mu_2 = \sigma^2, \tag{2}$$

where $\Delta x_i = x_i - \mu$, μ is the population mean and σ is the standard deviation. Then the definitions of P_1, P_2 and of the special kurtosis P_{2*} are

$$P_1 = \mu^3/\sigma^3, \quad P_2 = \mu^4/\sigma^4, \quad P_{2*} = \lg\,(P_2/3). \tag{3}$$

Theoretic symmetric distributions, such as the ND (Gaussian), the double exponential distribution (Laplassian) and the uniform distribution, have zero skewness P_1. DDs with heavier right wing/tail have convenient positive skewness and vice versa. However, the mentioned distribution have not convenient kurtosis P_2 – 3, 6 and 6/5, respectively. Therefore, for compatibility with the ND, the modified kurtosis P_{2*} is introduced here. The respective modified kurtosis values P_{2*} become 0, 0.3 and -0.4.

In the parameters in Eq.3 the deviations Δx_i from the mean value participate with their 3-rd or 4-th powers. Therefore, the large deviations are too heavy and these parameters are not robust. The intergram gradient parameters, suggested in this work, are defined by numerous SIDs and may be considered as robust. Moreover, in the most cases the gradients of the intergram functions, in the diagrams (c), (d) and (e), are well defined.

Some indexes of the ADs were compared in the preliminary work G2020, Figs. 6–8. The wide spread formula for estimation of the mode $x_{\rm AV} - x_{\rm mod} = 3(x_{\rm AV} - x_{\rm med})$ was confirmed with coefficient 3.03 by 34 DD, with coefficient of determination 99%.

Figure 5 shows that the intergram gradients gradients $G_{\rm C}$ and $G_{\rm W}$ (Fig. 3) correlate well with the relative median deviation $S_{\rm M}$. The relative median deviation is the half of the 68.26%-SID divided by the median value of the DD. Here $G_{\rm A}$ is usually poor defined and generally it does not correlate with $S_{\rm M}$. DD #13, about the diameters of the galacies, seems to be strongly peculiar in Fig. a.

Figure 6 shows other correlations between the skewness P_1 and kurtosis P_1 (Eqs. 12) and two intergram gradients. The remarkable parabolic dependence between the skewness and the kurtosis in Fig. a where found early by 34 such DDs (G2020, Fig. 7a). Correlation between squared skewness and the kurtosis found by Schlopflocher & Sullivan (2005).

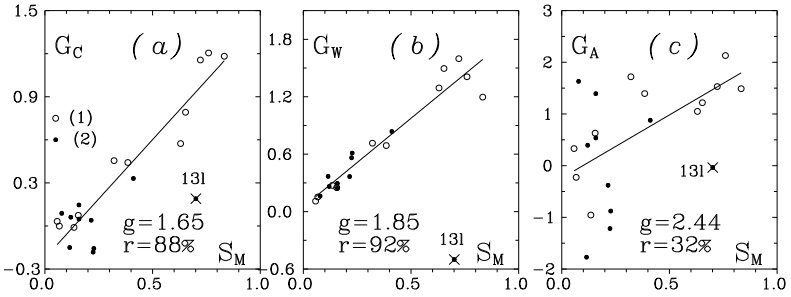

Fig. 5. Correlations between the median deviation S_M and the gradients of the functions in Eq. 1. (1) and (2) – gradients from intergrams with linear and logarithmic abscissae. g – gradient of the linear fit, r – coefficient of determination. The DD #13L is considered peculiar and it is excluded from the fits.

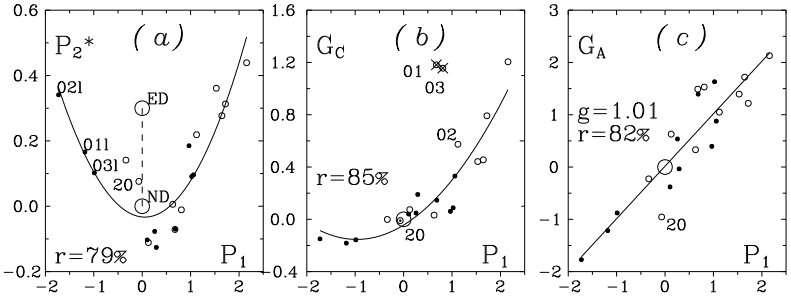

Fig. 6. Juxtapositions of the skewness P_1 with the modified kurtosis $P_{2}*$ and the intergram gradients $G_{\rm C}$ and $G_{\rm W}$. The diagrams in (a) and (b) contain quadratic fits and (a) – linear fit. Circles mark positions of the ND and (only in (a)) the double exponential distribution (ED). In (b) the examples #01 and #03, with Wolf numbers, are excluded from the fits. (See Fig. 4.)

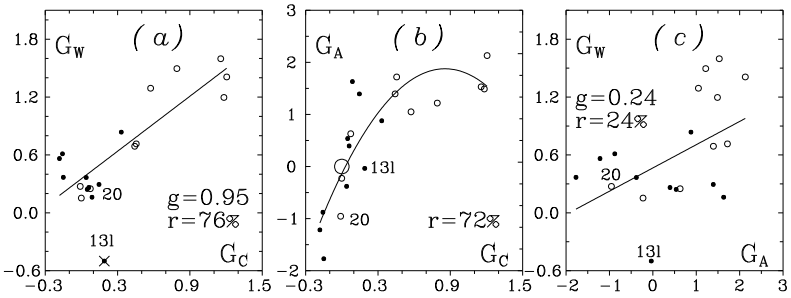

Fig. 7. Mutual correlations between the gradients of the intergram functions defined in Eq. 1. Circles mark the positions of the ND, only in (b). Peculiar is the DDs #13L (See Fig. 4.)

Remarkable peculiarity of the DDs #01 and #03 of the Wolf numbers appear in Fig. 7*b*. They are excluded from the fits. Remarkable correlation is seen in Fig. 7*c* between the asymmetry parameters P_1 and G_A.

Figure 7 shows juxtapositions between the gradients of the intergram functions in Eqs. 1. Fig. 7*a* and 7*b* show some correlations, but the gradients G_A and G_W do not correlate.

All diagrams in Figs. 5–7 may be used for elucidation of peculiar DDs. In the available DDs such are the distribution of the Wolf number and galaxies by diameters.

5. Conclusions

1. The intergram, suggested in the present work, is an unique representation of the DD with the next following valuable merits (Sections 1, 2):

(i) The intergram may replace the histogram well for the visualization of the unimodal AD (Fig. 2, 3);

(ii) The gradients of the intergram functions (Eq. 1) are robust morphological indexes of the intergram body of the AD (Fig. 3, 4);

(iii) Beside a log-ordinate the largest intergram leg of an AD with positive asymmetry, with linear or quadratic fit continuation, may prognosticate the high target value for 1% of the SP (Fig. 3). Table 2 in the Appendix shows such estimation through reverse cumulative function and intergram. In most cases they coincide.

(iv) Beside a log-ordinate the a deviation of the lowest segment intergram skeleton shows clearly excess or deficit of tails of the DD. Examples with significant deficit of large deviations are #01, #03, #19, #17L, #18L. Faint deficit of deviations may be suspected in #02 and #04. Small excess of positive deviations show #02, #04, #13, #014.

2. The DDs collected here have various origin and observing score, but their ordinary and intergram indexes follow common statistical dependences. Generally, these distributions are in principle similar (Figs. 5–7).

3. A quadratic dependence between the indexes of skewness and kurtosis is found here for 21 DDs (Fig. 6*a*). The same dependence was found for 34 DDs in the preliminary work (G2020, Fig. 7*c*). The key reason is the elucidation of DDs with strong negative asymmetry, #01L, #02L, #01L.

4. The distributions of the Wolf number and the visiting number of the Planetarium are very similar over linear, square root and logarithmic target parameter. The distributions over square root of the numbers are quasi-normal. It seems the Wolf number corresponds to the spot area better than to the spot number. (Fig. 1., #01–#03, #01L–#03L.)

5. The B-atmosphere extinction over Rozhen NAO in 138 nights from 2005–2008 yr. was measured with a CCD by Dimitrov (2009). The mode, median and AV are 0.13 mag, 0.175 mag and 0.20 mag, respectively (#07).

6. Generally, the 1% PHV intergram prognoses of the basic parameters of the disk galaxies correspond well to the observations.

7. Generally, the distributions of the masses, diameters and densities of the largest bodies in the Solar System, #17L, #18L and #19 (without the Sun and the giant planets) do not correspond to the observations. They give evi-

dences that at least one more big planet should be present in the Solar System.

Acknowledgements: The author is grateful for the data and recommendations to Boris Komitov (Wolf numbers), Hristina Kalaidzieva (Planetarium visits), Petko Nedialkov (AF And), Orlin Stanchev (disk galaxies) and Dinko Dimitrov (atmosphere extinction). The author is especially grateful for the attention to this wok to Petko Nedialkov and Antonia Valcheva. This work is supported by the grant KΠ-06-H28/2, 08.12.2018 of the Bulgarian National Science Fund.

References

Dimitrov, D. 2009, private communication
Gantchev, G., Valcheva, A., Nedialkov, P., Ovcharov, E. 2017, Bulg. Astron. J., 26, 16
Georgiev, Ts. 2120, Publ. Astron. Soc. Bulgaria, (G2021), http://astro.shu-bg.net/pasb/index
Georgiev, Ts., Zamanov, R., Boeva, S., Latev, G., Spassov, B., Martí, J., Nikolov, G., Ibryamov, S., Tsvetkova, S., Stoyanov, K. A. 2020, Bulg. Astron. J., 32, 35–62
Georgiev, Ts., Boeva, S., Latev, G., Semkov, ED., Stoyanov, K. A., Tsvetkova, S., 2021, Bulg. Astron. J., 34, 10–29
Noble, P. L. & Wheatland, M. S. 2011, Astrophys. J. 732, 5., Fig. 2, Fig. 6
Rousseeuw P.J., Leroy A.M. 1987, *Robust Regression and Outlier Detection*, John Willy & Sons
Schlopflocher, T. P. & Sullivan, P., J. 2005, Boudary-Layer Meteorology, 115. 341