Statistical Analysis of Binary Stars from the *Gaia* Catalog Data Release 2

Petr Zavada and Karel Píška
Institute of Physics of the Czech Academy of Sciences, Na Slovance 2, 182 21 Prague 8, Czech Republic; zavada@fzu.cz

Abstract

We have developed a general statistical procedure for analysis of 2D and 3D finite patterns, which is applied to the data from recently released *Gaia*-ESA catalog DR2. The 2D analysis clearly confirms our former results on the presence of binaries in the former DR1 catalog. Our main objective is the statistical 3D analysis of DR2. For this, it is essential that the DR2 catalog includes parallaxes and data on the proper motion. The analysis allows us to determine for each pair of stars the probability that it is the binary star. This probability is represented by the function $\beta(\Delta)$ depending on the separation. Furthermore, a combined analysis of the separation with proper motion provides a clear picture of binaries with two components of the motion: parallel and orbital. The result of this analysis is an estimate of the average orbital period and mass of the binary system. The catalog we have created involves 80,560 binary candidates.

1. Introduction

In this paper we analyze the recent data from the new catalog DR2 Gaia Collaboration et al. (2018) obtained by the *Gaia*-ESA mission. If compared with the previous DR1 catalog Gaia Collaboration et al. (2016a, 2016b), the DR2 contains the cleaner data complemented with parallaxes and data on the 2D proper motion. Parallaxes allow us to determine the distance of the stars, so we can substantially enlarge our former DR1 analysis (Zavada & Píška 2018) and work with the 3D patterns of moving stars.

In the present study, we focus on the statistical analysis of the presence of binaries. This topic is related to the recent studies of various aspects of binaries with the use of the catalogs DR1 (Oelkers et al. 2017; Oh et al. 2017) and DR2 (Ziegler et al. 2018; Jiménez-Esteban et al. 2019). These authors, apart from their own results, present an up-to-date overview of important findings on binary stars. Other important papers exploring wide binaries can be cited from the era before *Gaia*, for example, Caballero (2009) and Close et al. (1990). However, our approach and objectives are rather different, so the results obtained are complementary.

Methodology for 2D analysis has been described in detail in our above-quoted paper. In Section 2 we repeat its essence and perform generalization for the 3D case. For 2D analysis in Section 3 we take the same region in the DR2 catalog as we used in the DR1, so we can compare results from both corresponding data sets.

Principal results are obtained from the 3D analysis of a sample of DR2 data and are presented in Section 4. This part deals with two issues: the analysis of 3D separations and the analysis of proper motion of pairs of sources. The combination of both insights provides essential information about the statistical set of binaries. Obtained results are discussed in Section 5. Here we define the probabilistic function $\beta(\Delta)$, which is important for discussion on the occurrence of binaries. Our present catalog of binary candidates is described in

Section 6, where we also shortly discuss its content and overlap with the catalog JEC—Jiménez-Esteban et al. (2019).

A brief summary of the paper is presented in Section 7. The Appendix is devoted to the derivation of some relations important for our statistical approach. The most important are distributions of separations of random sources uniformly distributed inside circles or spheres of unit diameter. The significant role of these functions for our approach is explained in Section 2.

2. Methodology

The methods are designed for analysis of the distribution of stars inside circles or spheres covering the chosen region of the sky, as sketched in Figure 1. These 2D and 3D star patterns are called events. Input data for the generation of the event grids are supposed in the galactic reference frame. So, the position L of a source is defined by spherical coordinates L, l, and b (distance from the Sun, galactic longitude and latitude):

$$L = Ln; \qquad n = (\cos b \cos l, \cos b \sin l, \sin b),$$
$$-\frac{\pi}{2} \leqslant b \leqslant \frac{\pi}{2}, \qquad -\pi < l \leqslant \pi. \qquad (1)$$

In the center of circles or spheres we define a local orthonormal frame defined by the basis:

$$k_r = n_0 = (\cos b_0 \cos l_0, \cos b_0 \sin l_0, \sin b_0),$$
$$k_l = (-\sin l_0, \cos l_0, 0),$$
$$k_b = (-\sin b_0 \cos l_0, -\sin b_0 \sin l_0, \cos b_0), \qquad (2)$$

where $k_r = n_0(b_0, l_0)$ defines angular position of the event center. Unit vector k_l is perpendicular to k_r and has direction of increasing l, see Figure 2. Unit vector k_b is defined as $k_b = k_r \times k_l$ and has direction of increasing b. Vector k_r has radial direction, perpendicular vectors k_b and k_l lies in the transverse plain.

2.1. Definition of Events

The 2D event of the multiplicity M is a set of stars with angular positions n_i inside a circle with the event center n_0
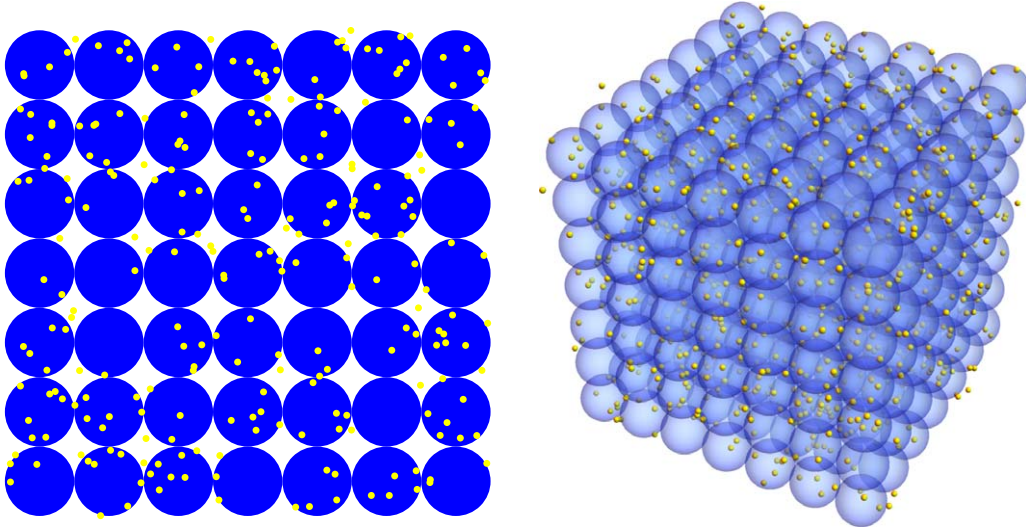
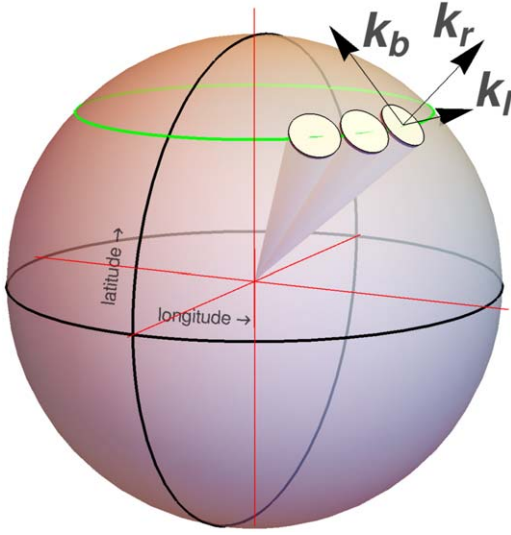**Figure 1.** Grids of 2D (left) and 3D (right) events with uniform distributions of the stars.

Figure 2. Galactic reference frame defined by galactic longitude and latitude with an event local frame defined by orthonormal basis $\boldsymbol{k}_r, \boldsymbol{k}_l, \boldsymbol{k}_b$.

and a small angular radius ρ_2:

$$|\boldsymbol{n}_i - \boldsymbol{n}_0| \leqslant \rho_2, \qquad i = 1, \dots M. \qquad (3)$$

With the use of event local basis (2), the local coordinates are defined as

$$x_i = \boldsymbol{n}'_i.\boldsymbol{k}_l, \qquad y_i = \boldsymbol{n}'_i.\boldsymbol{k}_b; \qquad \boldsymbol{n}'_i = \boldsymbol{n}_i - \boldsymbol{n}_0. \qquad (4)$$

We define the 2D event as the set:

$$\{x_i, y_i\}; \quad x_i^2 + y_i^2 \leqslant \rho_2^2, \quad i = 1, \dots M. \qquad (5)$$

Since the DR2 catalog involves also data on parallaxes, we can similarly also generate the 3D events—patterns of the M sources with position $\boldsymbol{L}_i$ inside the spheres with the center $\boldsymbol{L}_0$ and radius ρ_3:

$$|\boldsymbol{L}_i - \boldsymbol{L}_0| \leqslant \rho_3, \qquad i = 1, \dots M. \qquad (6)$$

With the use of the star positions (1) and local basis (2) we define local coordinates $\{X_i, Y_i, Z_i\}$ as

$$X_i = \boldsymbol{N}_i.\boldsymbol{k}_l, \qquad Y_i = \boldsymbol{N}_i.\boldsymbol{k}_b, \qquad Z_i = \boldsymbol{N}_i.\boldsymbol{k}_r; \qquad (7)$$

$$N_i = L_i \boldsymbol{n}_i - \boldsymbol{L}_0, \qquad L_i[\text{pc}] = \frac{1000}{p_i[\text{mas}]}, \qquad (8)$$

where p_i is parallax, and L_i is distance of the star. The 3D event is defined as the set:

$$\{X_i, Y_i, Z_i\}; \quad X_i^2 + Y_i^2 + Z_i^2 \leqslant \rho_3^2, \quad i = 1, \dots M. \qquad (9)$$

2.2. 2D Methods

The first method is based on the Fourier analysis of 2D events, where we have introduced characteristic functions $\Theta_n(M)$ depending on the event multiplicity M. These functions are generated by a set of events and measure statistical deviations from uniform distribution of stars ($\Theta_n(M) = 1$), for instance a tendency to clustering ($\Theta_n(M) > 1$) or anticlustering ($0 < \Theta_n(M) < 1$). Details of the method are described in Zavada & Píška (2018), in this paper we will present only the result.

With the use of a second complementary method, we analyze distributions of angular separations of sources inside the 2D events (5). Distribution is generated from the set of events. We use either absolute separations

$$x_{ij} = |x_i - x_j|, \quad y_{ij} = |y_i - y_j|, \quad d_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2}, \qquad (10)$$

or scaled ones

$$\hat{x}_{ij} = \frac{x_{ij}}{2\rho_2}, \quad \hat{y}_{ij} = \frac{y_{ij}}{2\rho_2}, \quad \hat{d}_{ij} = \frac{d_{ij}}{2\rho_2}. \qquad (11)$$

The suitable unit of the parameters x_i, y_i, ρ_2 will be for our purpose 1 as ($1''$). Distribution of scaled separations generated by Monte Carlo (MC) for a uniform distribution of stars in the sky is shown in Figure 3. The exact shape of normalized MC distributions reads

$$P(\hat{\xi}) = \frac{16\hat{\xi}}{\pi}(\arccos\hat{\xi} - \hat{\xi}\sqrt{1 - \hat{\xi}^2}); \quad \hat{\xi} = \hat{d}_{ij}, \qquad (12)$$
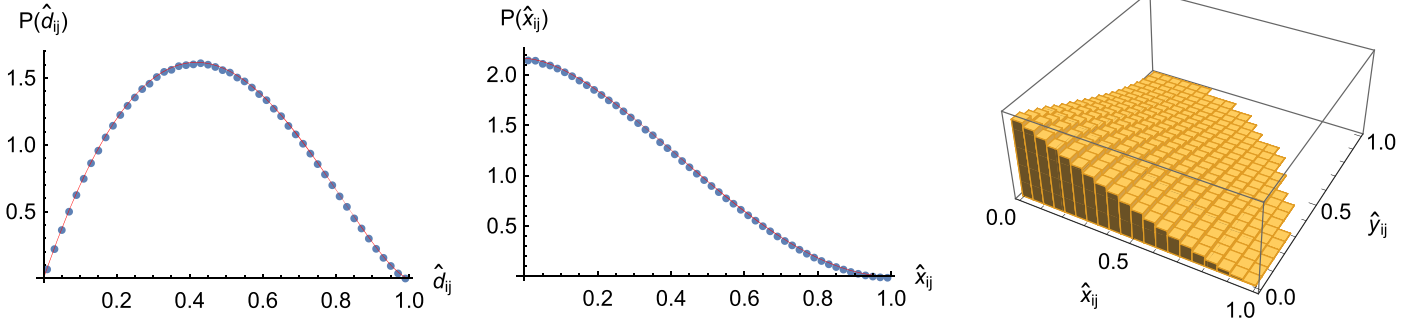
Figure 3. MC distributions of separations of uniformly generated stars (points). The red curves represent the functions (12) and (13). The MC statistics correspond to 5×10^5 events of multiplicity $M = 5$.
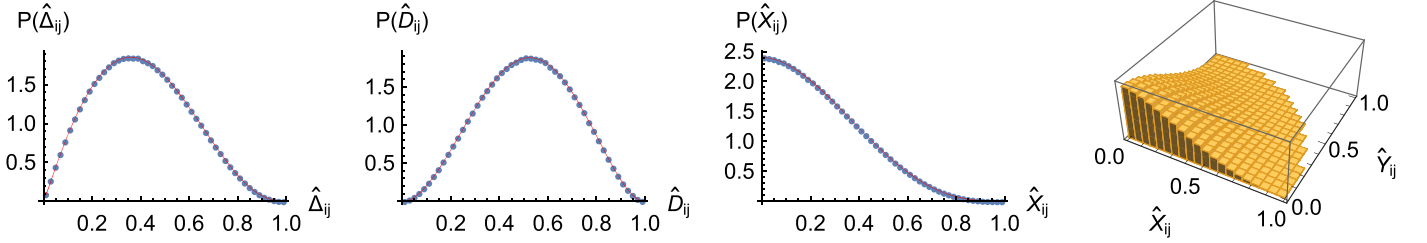

Figure 4. MC distributions of separations between uniformly generated stars (points) in 3D. The red curves represent functions (17)–(19). The MC statistics is represented by 5×10^5 events of multiplicity $M = 5$.

$$P(\hat{\xi}) = \frac{64}{3\pi^2}((1 + \hat{\xi}^2)\text{EllipticE}(1 - \hat{\xi}^2)$$
$$- 2\hat{\xi}^2\text{EllipticK}(1 - \hat{\xi}^2)); \quad \hat{\xi} = \hat{x}_{ij}, \hat{y}_{ij}, \quad (13)$$

where the functions EllipticK (EllipticE) are complete elliptic integrals of the first (second) kind. The proof is given in the Appendix. These distributions do not depend on the event multiplicity and radius, this is an advantage of the scaled separations. Obviously, we have always $0 < \hat{\xi} < 1$. These exact functions replace their approximations resulting from the MC calculation applied in the previous paper.

2.3. 3D Methods

Similarly, as in the 2D case, we shall work with absolute separations

$$X_{ij} = |X_i - X_j|, \quad Y_{ij} = |Y_i - Y_j|, \quad Z_{ij} = |Z_i - Z_j|, \quad (14)$$

$$D_{ij} = \sqrt{X_{ij}^2 + Y_{ij}^2 + Z_{ij}^2}, \quad \Delta_{ij} = \sqrt{X_{ij}^2 + Y_{ij}^2}, \quad (15)$$

and/or with the scaled ones

$$\hat{X}_{ij} = \frac{X_{ij}}{2\rho_3}, \quad \hat{Y}_{ij} = \frac{Y_{ij}}{2\rho_3}, \quad \hat{Z}_{ij} = \frac{Z_{ij}}{2\rho_3},$$

$$\hat{D}_{ij} = \frac{D_{ij}}{2\rho_3}, \quad \hat{\Delta}_{ij} = \frac{\Delta_{ij}}{2\rho_3}. \quad (16)$$

The suitable unit of the parameters X_i, Y_i, Z_i, ρ_3 is for our purpose 1 pc. Distribution of scaled separations generated by MC from the uniform distribution of stars in the 3D region of the sky is shown in Figure 4. Exact shapes of these normalized MC distributions read

$$P(\hat{\xi}) = 12\hat{\xi}^2(2 - 3\hat{\xi} + \hat{\xi}^3); \quad \hat{\xi} = \hat{D}_{ij}, \quad (17)$$

$$P(\hat{\xi}) = \frac{9}{2}\hat{\xi}\left((2 + \hat{\xi}^2)\sqrt{1 - \hat{\xi}^2} + \hat{\xi}^2(4 - \hat{\xi}^2) \right.$$

$$\left. \times \ln\frac{\hat{\xi}}{1 + \sqrt{1 - \hat{\xi}^2}}\right); \quad \hat{\xi} = \hat{\Delta}_{ij}, \quad (18)$$

$$P(\hat{\xi}) = \frac{12}{5}(1 - \hat{\xi})^3(1 + 3\hat{\xi} + \hat{\xi}^2); \quad \hat{\xi} = \hat{X}_{ij}, \hat{Y}_{ij}, \hat{Z}_{ij}, \quad (19)$$

as proved in the Appendix. Shapes of these distributions similarly to (12) and (13) do not depend on event multiplicity and radius. The analysis with the use of characteristic functions $\Theta_n(M)$ could be in the 3D case done separately in the plains XY, YZ, and ZX. However, such analysis is not the aim of the present paper.

2.4. Aims

In Section 3 using the DR2 data set we shall obtain the characteristic functions $\Theta_n(M)$, afterwards we check distributions (12) and (13). The distributions (17)–(19) will be used for the data analysis in Section 4.1. All these distributions are of key importance for the analysis of real data. They represent the templates, which can reveal a violation of uniformity in the star distributions. Binary (and multiple) star systems are an example of such a violation, which manifests as the peaks in the distributions of angular or space separations in the region of close sources. In general, the scale of expected structure violating uniformity should be less than the event radius ρ_2 or ρ_3.

3. Analysis of 2D Events

Here we present the results obtained from regions of the DR2 catalog listed in Table 1. The regions are shown in Figure 5. The corresponding events are created with the same angular radius as in Zavada & Píška (2018), which allows us
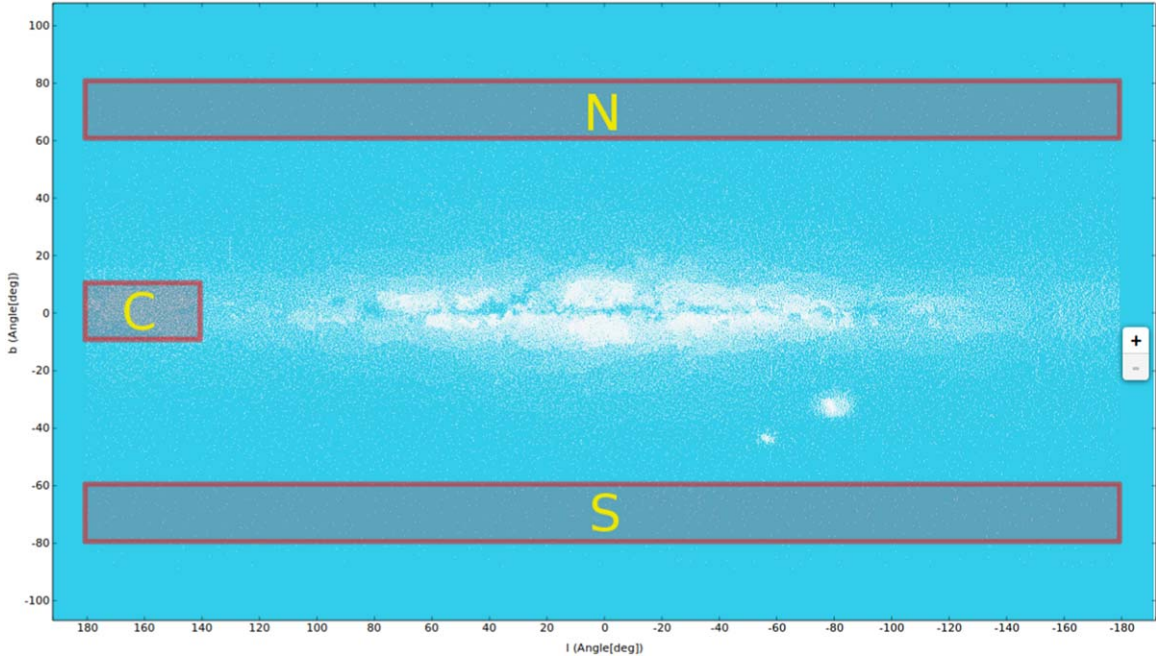
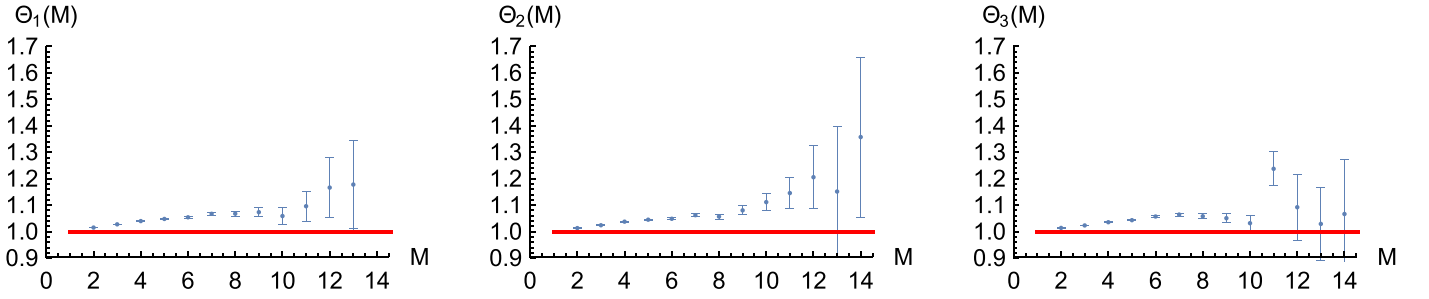**Figure 5.** Analyzed regions in the DR2 catalog.

Figure 6. Characteristic functions $\Theta_n(M)$, $n = 1, 2, 3$ for events in the area N and S without any cut on magnitudes. Red line corresponds to the uniform distribution of sources.

Table 1
Analyzed Regions in the DR2 Catalog, Where ρ_2 Is the Angular Radius of the Events, $\langle L \rangle$, $\langle M \rangle$ Are Average Distance and Event Multiplicity, and N_e Is the Total Number of Events

	2D Region: $l \times b(\text{deg}^2)$	$\rho_2(\text{as})$	$\langle L \rangle (\text{pc})$	$\langle M \rangle$	N_e
N and S	$\langle -180, 180 \rangle \times \langle \pm 60, \pm 80 \rangle$	72	1290	3.21	2,055,674
C	$\langle 140, 180 \rangle \times \langle -10, 10 \rangle$	18	1912	2.75	3,588,183

Note. Only sources in distance 1–5000 pc are taken into account. The analysis is done for events $2 \leqslant M \leqslant 15$.

the consistent comparison of results from the DR1 and DR2 catalogs. First, we checked the events covering the regions N and S. Their nonuniformity defined by the characteristic functions $\Theta_n(M)$ is demonstrated in Figure 6. The clear result $\Theta_n(M) > 1$ indicates the presence of clustering.

Corresponding distributions of angular separations are shown in Figure 7 together with curves (12) and (13). These results can be compared with those in Figures 7 (lower panels), 10, and 11 in the former paper. We observe:

(i) The peaks at small angular separations in the DR2 corresponding to binaries are clearer and more pronounced than in the DR1 catalog. Panels (e) and (k) in Figure 7 demonstrate the double stars separated by $d_{ij} \lesssim 0.5$ as are absent because such close pairs are not resolved in the DR2 data set as reported

in Arenou et al. (2018). In both catalogs, we observe an excess of binaries in the region N and S for $\hat{d}_{ij} \lesssim 0.06$ or equivalently for $d_{ij} \lesssim 8.6$ as. For greater separations inside the event, we observe perfectly uniform distributions of stars. Note the data and curves are equally normalized for $0 < \hat{\xi} < 1$. That is why the strong peak in panels (g)–(i) is balanced by a small reduction of distribution outside the peak. Brighter stars ($G \leqslant 15$, panels (g)–(l)) show evidently stronger peaks than the sample without any cut on magnitude (panels (a)–(f)). A similar tendency was already observed in the catalog DR1.

(ii) A more pronounced presence of binaries is demonstrated also in Figure 6. The slopes of lines in DR2 are greater than in DR1—clustering is more obvious. For $M \leqslant 8$ the slopes are $\approx 4\%(8\%)$ for the DR1(DR2) data. In Figure 8 we have shown
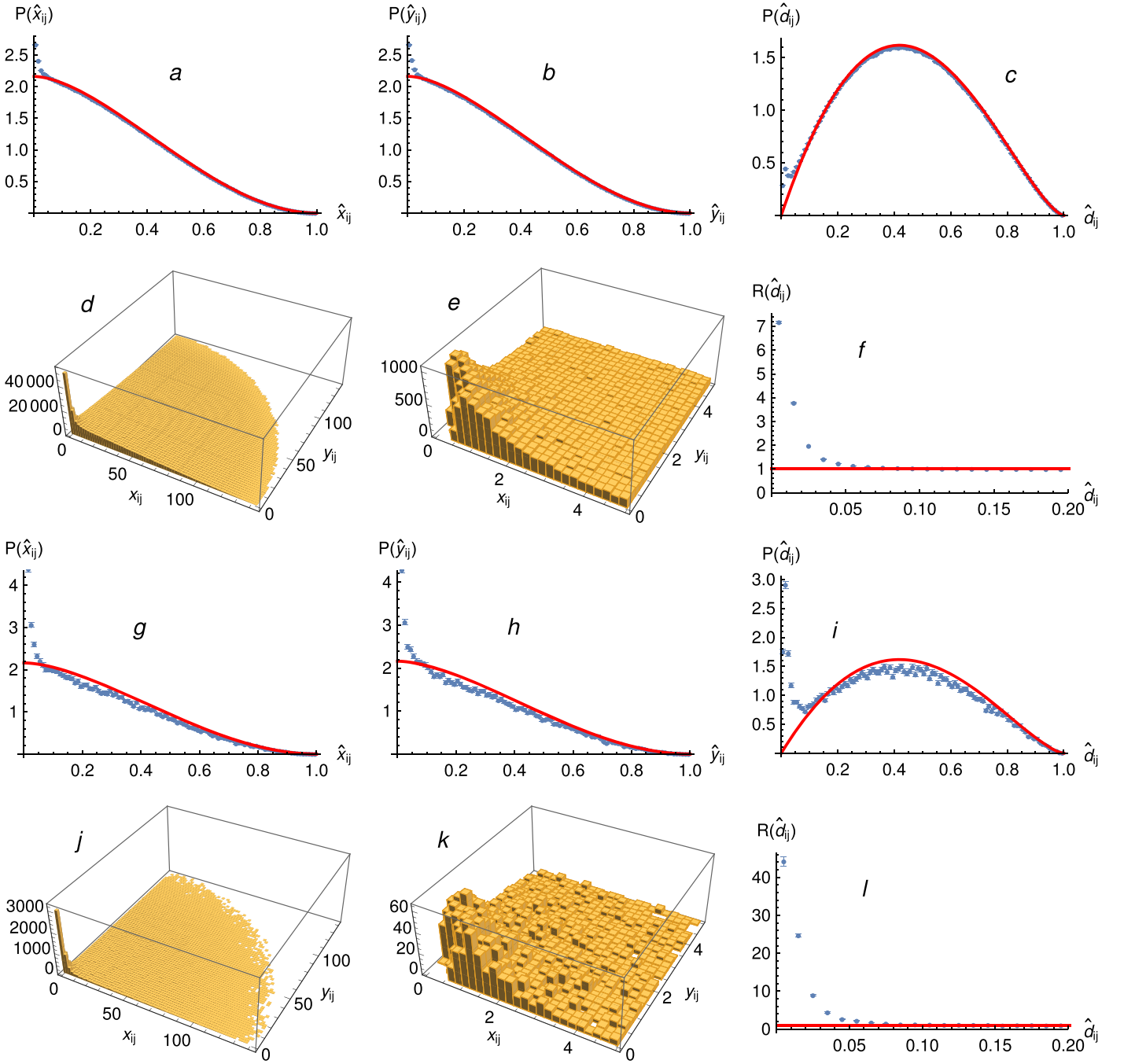
Figure 7. Panels (a)–(f): distributions of angular separations in the region N and S for all *G*. The blue points in the panels (a)–(c) represent the data on scaled separations $\hat{x}_{ij}$, $\hat{y}_{ij}$, $\hat{d}_{ij}$ and the red curves are functions (12) and (13) representing uniform simulation. Panel (f) is the ratio of data to simulation from panel (c). Panels (d) and (e) represent 3D plot of separations x_{ij}, y_{ij} in different scales (unit is 1 as). Panels (g)–(l): the same for sources $G \leqslant 15$.

some results obtained in a more populated region C. Also here we can observe a clear peak at small angular separations of sources of the magnitude $G \leqslant 15$, which proves the presence of binaries. Panel (b) again demonstrates the absence of double stars separated by $d_{ij}(\hat{d}_{ij}) \lesssim 0.5$ as(0.014) due to insufficient resolution. Different scales of $\hat{d}_{ij}$ in Figures 7(l) and 8(c) are due to different radii ρ_2 of events from N and S and C regions. Similar plots could be presented for whole spectrum of magnitudes in region C; however, elevation above the red line due to binaries is much less than that in Figure 8(c). The reason

can be that denser region C with all magnitudes generates higher background of the optical doubles and a consequently lower relative rate of binaries.

4. Binaries in 3D Events

We present the results obtained from the 3D region defined in Table 2. The parallax and angular components of the star proper motion are the parameters, which substantially enrich the recent *Gaia* data. We work with the 3D events (9).
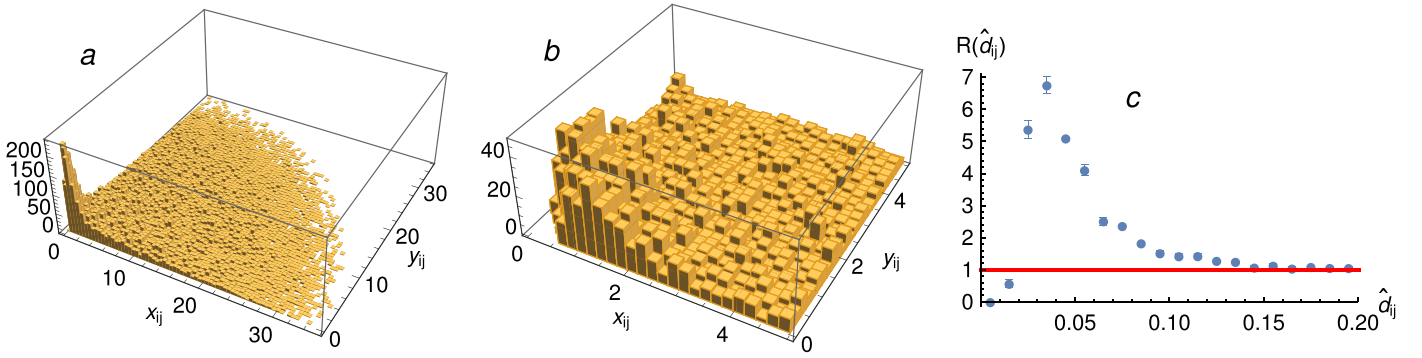
Figure 8. Panels (a)–(c): distributions of angular separations in the region C for sources $G \leqslant 15$. Panels (a) and (b) represent a 3D plot of separations x_{ij}, y_{ij} in different scales (unit is 1 as). Panel (c) is the ratio of data distribution $P(\hat{d}_{ij})$ to the function (12), like panels (f) and (l) in previous Figure 7.

Table 2
Region of 3D Analysis is Cube Centered at the Origin of the Galactic Reference Frame

Region	ρ_3(pc)	$\langle L \rangle$ (pc)	$\langle M \rangle$	N_e
Cube of edge 400 pc	2	188	6.8	727,744

Note. Only sources of positive parallax are included. ρ_3 is the radius of events, $\langle L \rangle$, $\langle M \rangle$ are average distance and event multiplicity, and N_e is the total number of events. The analysis is done for events $2 \leqslant M \leqslant 15$.

4.1. Analysis of 3D Separations

The summary results of the analysis obtained from all magnitudes G are shown in Figure 9. Distributions of scaled separations in panels (a)–(c), (g), and (h) perfectly match the uniform distribution of sources, but with the exception of the first bin in (a), (b), and (h). Apparent excess of very close pairs in planes XY, YZ, and ZX is seen also in panels (d)–(f). However, the sharp peak can be observed only in the plane XY (panel (d) and its magnified version (j)). Smearing in the direction of Z (difference of radial positions, panels (c), (e), and (f)) is due to lower accuracy in measuring of parallaxes. The errors of local coordinates depend on the errors of separations and with the use of definitions (7) and (8) are calculated as

$$\delta X_{ij} = \sqrt{\left(\frac{\partial X_{ij}}{\partial L_i}\delta L_i\right)^2 + \left(\frac{\partial X_{ij}}{\partial L_j}\delta L_j\right)^2}$$

$$\approx \delta L \sqrt{(\boldsymbol{n}_i \boldsymbol{k}_l)^2 + (\boldsymbol{n}_j \boldsymbol{k}_l)^2} \leqslant \sqrt{2}\,\rho_3 \frac{\delta L}{L} \quad (20)$$

and similarly

$$\delta Y_{ij} \leqslant \sqrt{2}\,\rho_3 \frac{\delta L}{L}, \quad \delta Z_{ij} \approx \delta L \sqrt{(\boldsymbol{n}_i \boldsymbol{k}_r)^2 + (\boldsymbol{n}_j \boldsymbol{k}_r)^2} \leqslant \sqrt{2}\,\delta L. \quad (21)$$

Note that $\delta L/L = \delta p/p$, where p is parallax. Obviously $\delta Z_{ij} \gg \delta X_{ij}$, δY_{ij} since $\boldsymbol{n}_i \boldsymbol{k}_l$, $\boldsymbol{n}_i \boldsymbol{k}_b \ll \boldsymbol{n}_i \boldsymbol{k}_r \lesssim 1$. That is why we prefer distributions of Δ_{ij} to D_{ij} for obtaining precise results. Maximum values of separations X, Y, and Z is 4 pc, which follows from the event radius $\rho_3 = 2$ pc. The excess of close pairs is obvious most explicitly from the distribution of Δ_{ij} in panel (k) that represents a magnified version of (h). The important ratio of the distribution $P(\Delta_{ij})$ to the uniform

simulation (red curve in panel (h) rescaled to $\Delta_{ij} = 2\rho_3 \hat{\Delta}_{ij}$) is shown in logarithmic scale of Δ_{ij} in panel (i).

The same distributions but for brighter sources $G \leqslant 15$ are shown in Figure 10. Similarly, as in the 2D case, the peaks are stronger for brighter sources and distributions outside the peaks confirm uniformity of the star distribution. Again due to equal normalization of data and red curves for $0 < \hat{\xi} < 1$ the strong peak in panels (g) and (h) is balanced by a small reduction of distribution beyond the peak. The excess of close pairs observed in both figures again convincingly indicates the presence of binaries.

For quantitative estimates, the important panel is (i), which displays the ratio data/simulation. This is more accurate than only displaying peaks with some undefined background. Panel (i) suggests that separations of binary systems in the analyzed region meet very approximately

$$\Delta_{ij} \lesssim 0.1 - 0.2 \text{ pc.} \quad (22)$$

We observe only a tail of distribution corresponding to more separated binaries. Closer pairs are absent due to the limited angular resolution in DR2 data. This result is compatible with the older data reported in Close et al. (1990). In Section 5 further discussion is devoted to the probability of the binary separation above this limit. We have checked that sampling with events generated by spheres of different radius ($\rho_3 = 5$ pc) does not change the approximate result (22).

4.2. Proper Motion of Binaries

The proper motion of the stars in DR2 is defined by two angular velocities

$$\mu_{\alpha*}(\equiv \mu_\alpha \cos\delta), \qquad \mu_\delta \quad (23)$$

in directions of the R.A. and decl. in the ICRS. So the corresponding transverse 2D velocity $\boldsymbol{U}$ is given as

$$\boldsymbol{U} = L(\mu_{\alpha*}, \mu_\delta), \qquad U = |\boldsymbol{U}|, \quad (24)$$

where L is distance of the star calculated from the parallax (8). For the pair of stars we can define:

$$\alpha_{ij} = \arccos\frac{\boldsymbol{U}_i \cdot \boldsymbol{U}_j}{U_i U_j}, \quad U_{ij} = |\boldsymbol{U}_i + \boldsymbol{U}_j|, \quad v_{ij} = |\boldsymbol{U}_i - \boldsymbol{U}_j|,$$
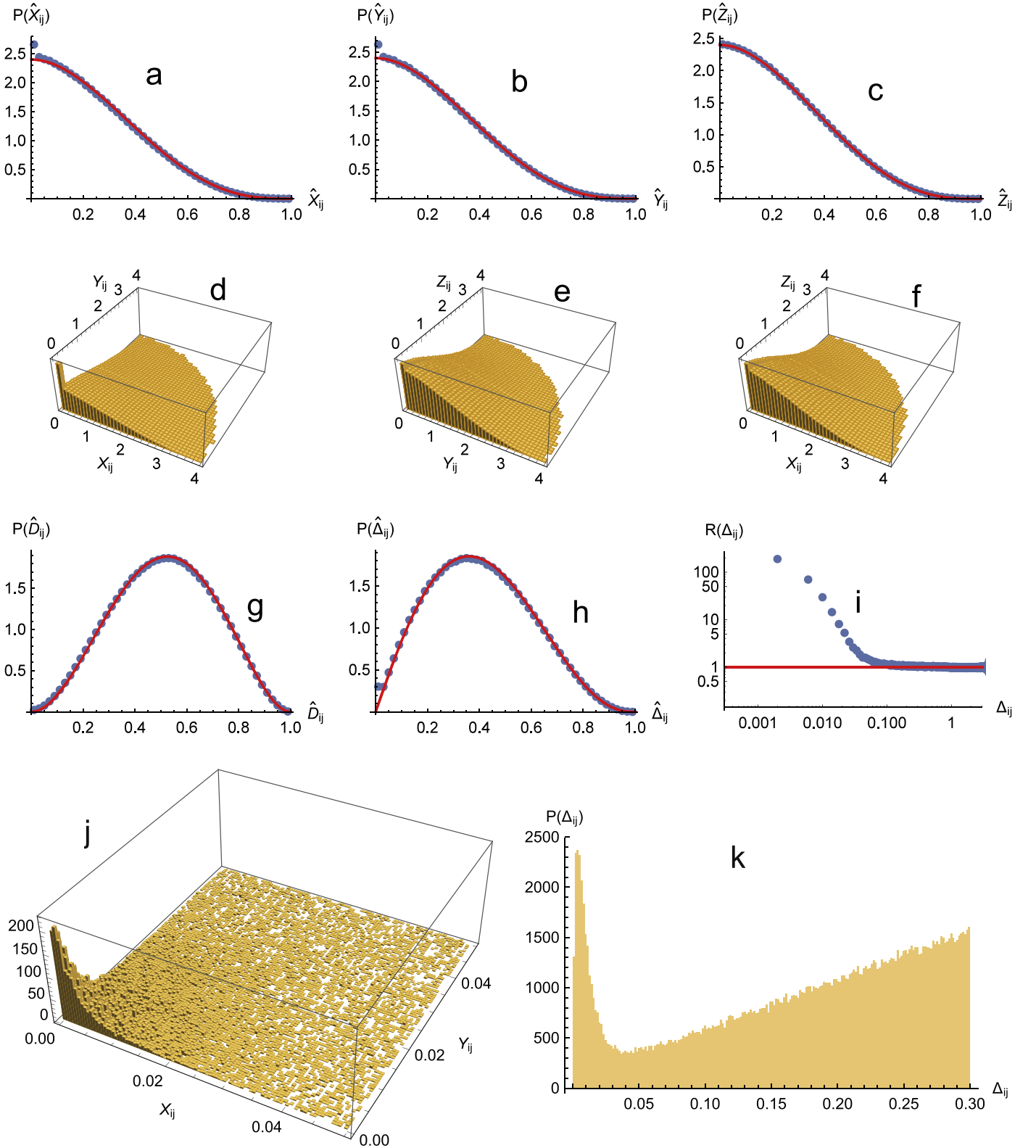
$$(25)$$

Figure 9. Distributions of separations (14)–(16) for all G. The blue points in the panels (a)–(c), (g), and (h) represent the data on scaled separations $\hat{X}_{ij}$, $\hat{Y}_{ij}$, $\hat{Z}_{ij}$, $\hat{D}_{ij}$, $\hat{\Delta}_{ij}$ and the red curves are functions (17)–(19) representing uniform simulation. Panels (d)–(f) show 2D projections of separations. Panel (i) is the ratio of data to simulation from panel (h). Panels (j) and (k) are magnified versions of (d) and (h). Unit of separations X_{ij}, Y_{ij}, Z_{ij}, D_{ij}, Δ_{ij} is 1 pc.

where α_{ij} is the angle between both transverse velocities. The corresponding errors read

$$\delta\alpha_{ij} \approx \sqrt{2}\frac{\delta u}{u}, \qquad \delta U_{ij} = \delta v_{ij} \approx \sqrt{2}\,\delta u, \qquad (26)$$

where

$$u = L\mu, \qquad \delta u \approx u\sqrt{\left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta \mu}{\mu}\right)^2}. \qquad (27)$$
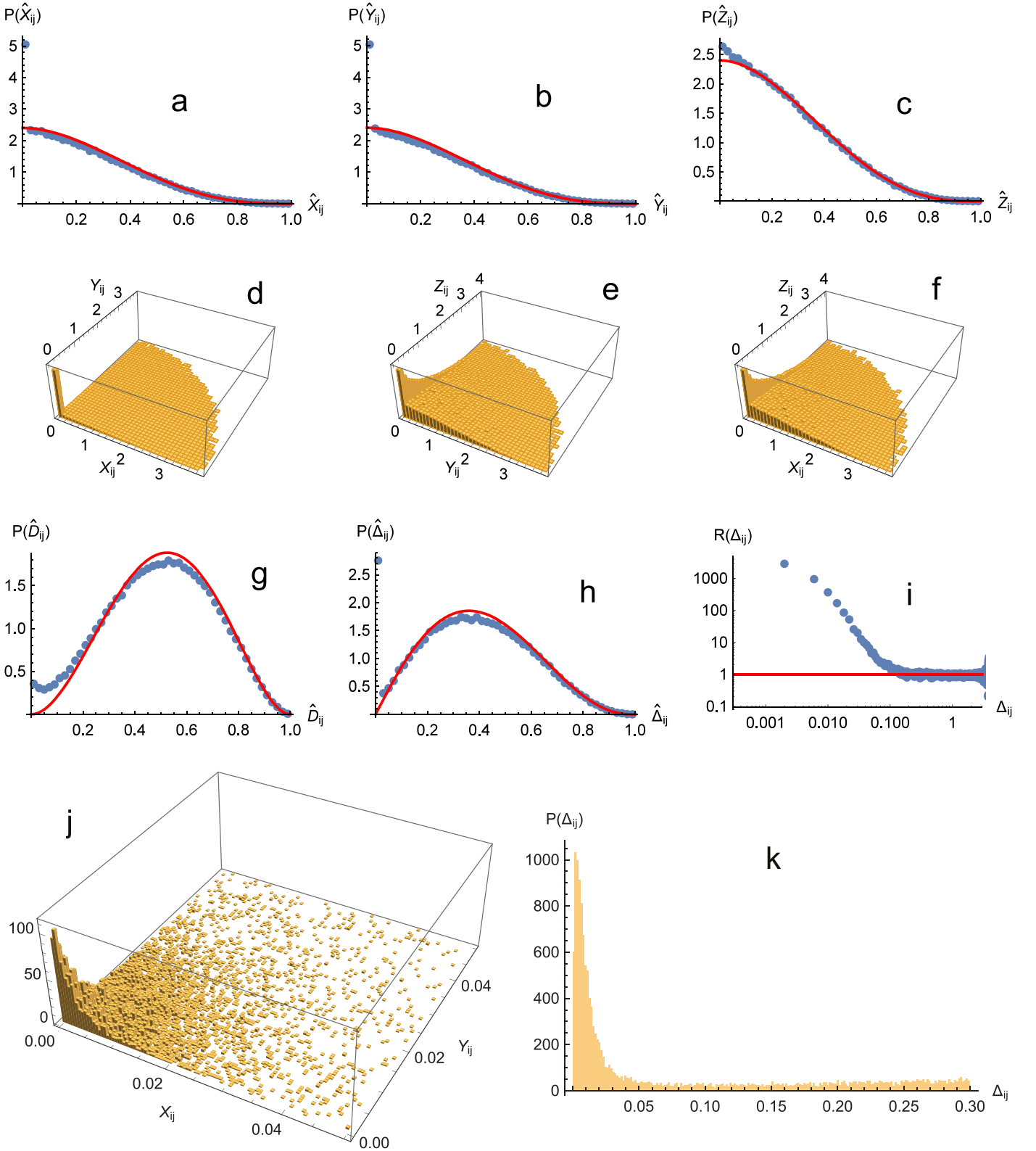
Figure 10. Same distributions as in the previous figure, but for bright pairs, $G \leqslant 15$.

Note that relative error $\delta v_{ij}/v_{ij}$ can be large, since v_{ij} is small compared with U_{ij} and the errors δv_{ij} and δU_{ij} are the same. In Figure 11(a) we show the distribution of the velocities U of the stars from the region defined in Table 2. Distribution of the corresponding pair angles α_{ij} is presented in Figures 11(b)–(d) for different regions of Δ_{ij}. In Figure 12(a) we have shown
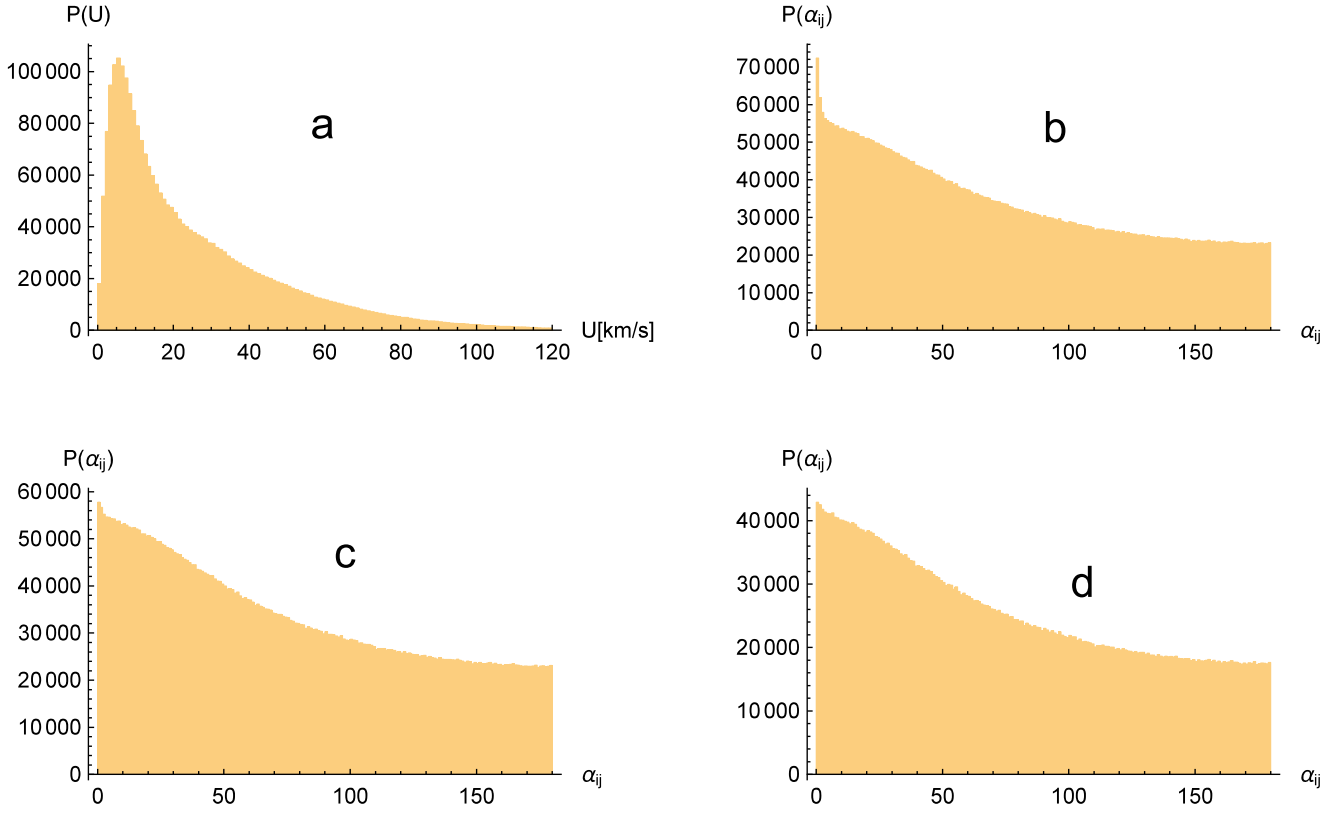
Figure 11. Panel (a): distribution of the star transverse velocities. Panels (b)–(d): distribution of the corresponding pair angles α_{ij} for $\Delta_{ij} > 0$, 0.15 and 1 pc.
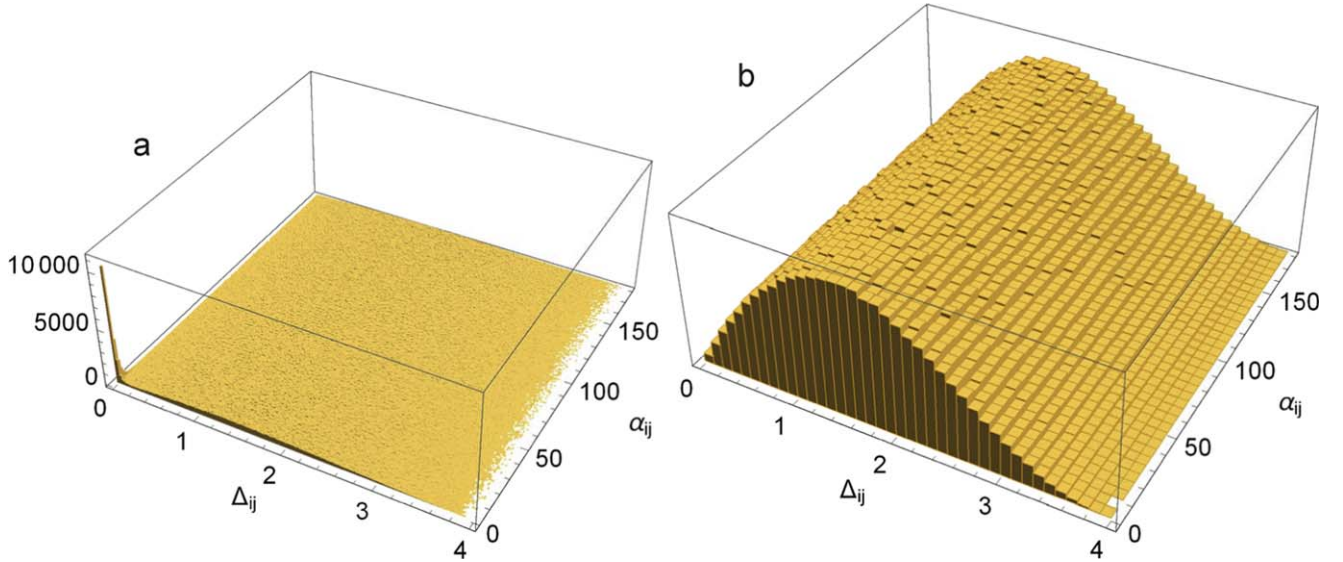

Figure 12. Panel (a): correlation of the pair transverse separation (unit is 1 pc) with angle of corresponding transverse velocities. Panel (b): corresponding MC simulation generated from uniform distributions of positions and velocity directions.

the correlation of pair transverse separations and angles α_{ij}. We observe a very narrow peak in the region of small Δ_{ij} and α_{ij}.[1] The peak is connected with presence of binaries as follows. The transverse velocities of two gravitationally

coupled stars are

$$\boldsymbol{U}_i = \boldsymbol{V} + \boldsymbol{v}_i, \qquad \boldsymbol{U}_j = \boldsymbol{V} + \boldsymbol{v}_j, \qquad (28)$$

where $\boldsymbol{V}$ is transverse velocity of their center of gravity and $\boldsymbol{v}_i$, $\boldsymbol{v}_j$ are transverse projections of instantaneous orbital velocities, they have always opposite direction. Dominance of very small α_{ij} means that

$$v_i = |\boldsymbol{v}_i| \ll |\boldsymbol{V}|, \qquad (29)$$

---

[1]   Input data (23) are related to the ICRS and Δ_{ij} is calculated in the galactic reference frame. Nevertheless the parameters Δ_{ij}, α_{ij}, U_{ij}, v_{ij} are invariant under rotation.
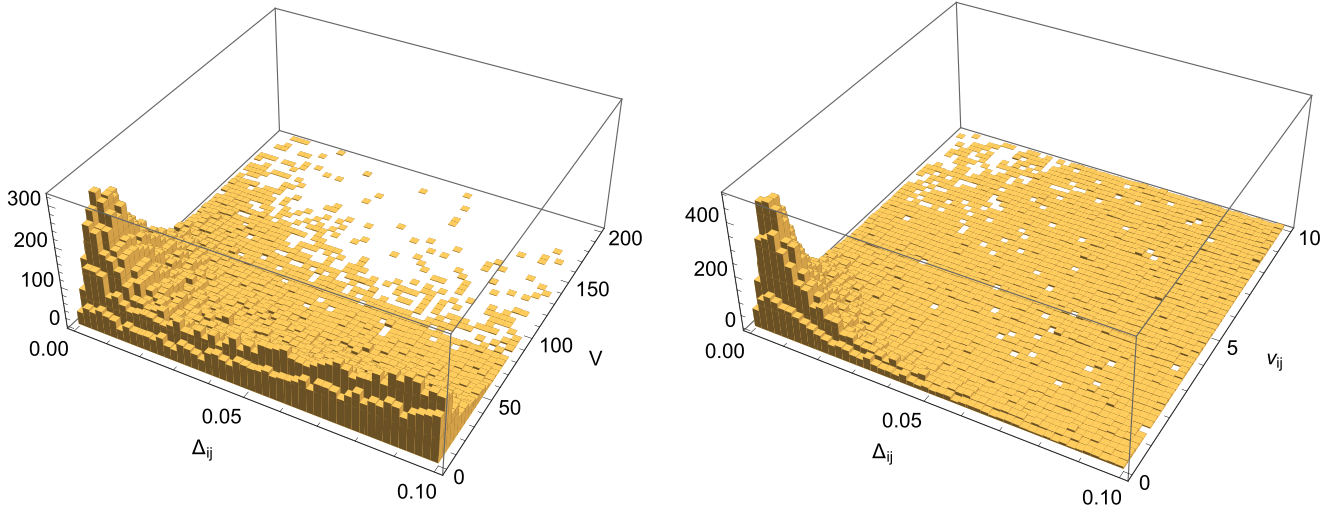
Figure 13. Correlation of the pair transverse separation Δ_{ij} with the transverse velocity V of the star system and with the transverse velocity v_{ij} of periodic motion. Unit of velocity is km s^{-1}.
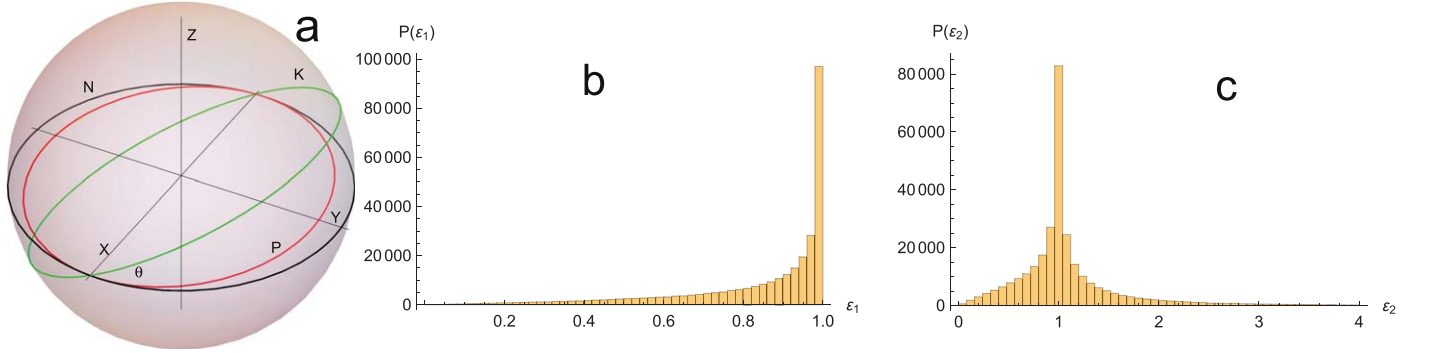

Figure 14. Panel (a): real orbit (K) and its deformed projection on the sky (P). Real orbit (N) in the plane xy. Panels (b) and (c): distributions of random parameters ε_1 and ε_2.

so for binaries in our Δ window ($\Delta_{\min}$ is given by resolution of two close sources and $\Delta_{\max} = 0.1$ pc by (22)) we have

$$U_{ij} \approx 2V, \qquad v_{ij} = v_i + v_j. \tag{30}$$

For comparison, we have generated an MC plot from uniform distributions of positions and velocity directions, which is shown in Figure 12(b). Obviously, uniform distribution contradicts Figure 12(a), which reflects the presence of binaries (peak at small α_{ij} and Δ_{ij}). The corresponding α_{ij} peak is also observed in Figures 11(b)–(d) on the background of the collective motion of stars (dominance of $\alpha_{ij} < 90°$). The peak is suppressed for $\Delta_{ij} \gtrsim 1$ pc. Selection of comoving systems is a basis of the methodology applied in the catalog JEC.

In Figure 13 we show correlations of the velocities V and v_{ij} with Δ_{ij} in the region of small separations, where the binaries are present. With the use of v_{ij} and Δ_{ij} we try to roughly estimate the orbital period of the binary star. To simplify the calculation, we assume the space binary orbits are circular and the star 3D separation is a (semimajor axis). There are the extreme cases:

(A) $M_1 \approx M_2$, then

$$\boldsymbol{v}_i \approx -\boldsymbol{v}_j, \qquad v_{ij} \approx v_i + v_j. \tag{31}$$

Orbital period is

$$T_A \approx \frac{\pi a}{w}, \tag{32}$$

where w is the space orbital velocity and a equals the diameter of the orbit.

(B) $M_1 \gg M_2$, then

$$v_2 \approx 0, \qquad v_{ij} \approx v_1, \tag{33}$$

but the orbital period is different

$$T_B \approx \frac{2\pi a}{w}, \tag{34}$$

since the separation a equals to the orbit radius.

At the same time, Kepler's law implies for orbital period T_g:

$$T_g = 2\pi \sqrt{\frac{a^3}{GM_{\text{tot}}}}, \tag{35}$$

where G is gravitational constant and $M_{\text{tot}} = M_1 + M_2$ is mass of the star system. For units $T_g[\text{y}]$, $a[\text{pc}]$, and $M_{\text{tot}}[M_\odot]$ we have

$$T_g = 0.937 \times 10^8 \sqrt{\frac{a^3}{M_{\text{tot}}}}. \tag{36}$$
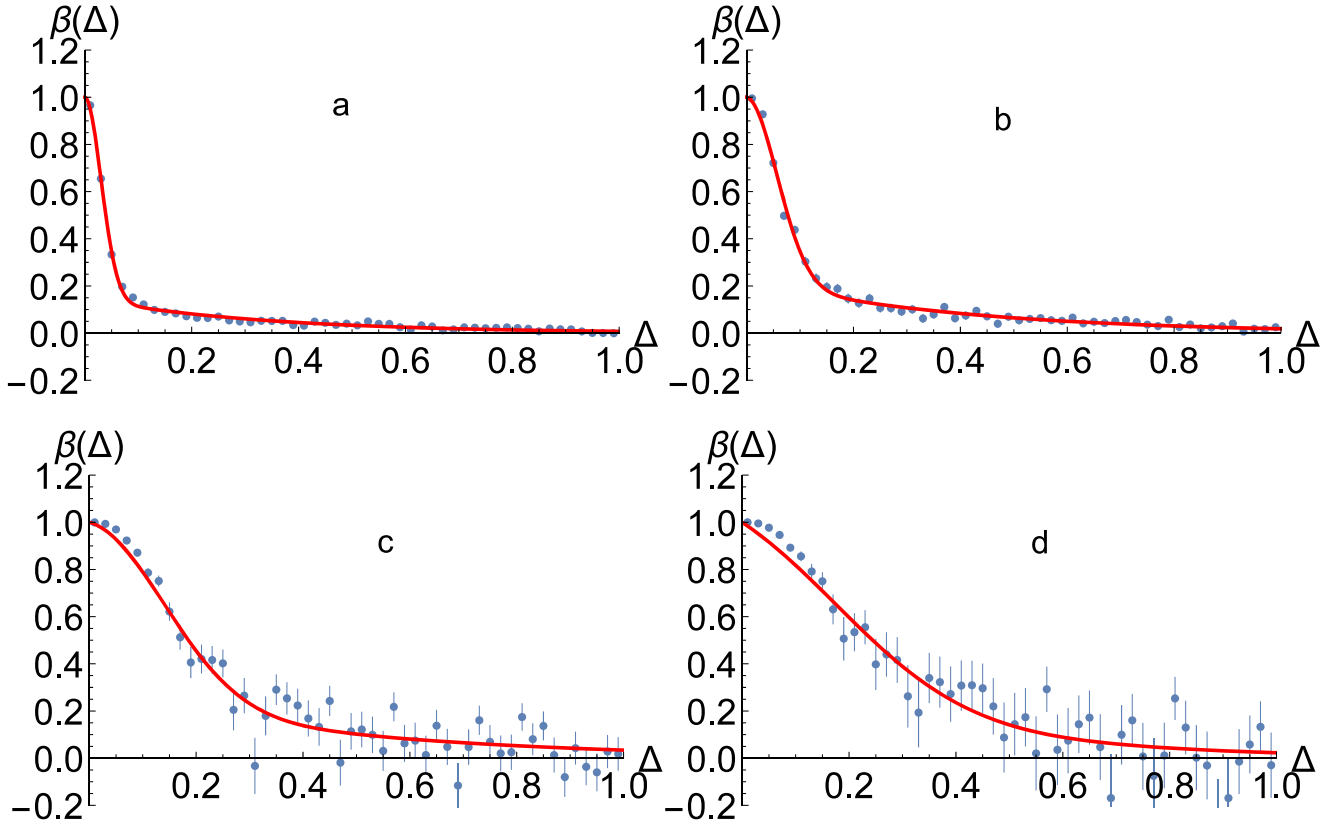
10

Figure 15. Probabilistic function $\beta(\Delta)$ for different domains $\Delta \times \alpha$ and intervals of magnitude G. Panel (a): all pairs and magnitudes. Panels (b)–(d): united domain **AB** for magnitude intervals G all, $G \leqslant 15$, $G \leqslant 13$. Blue points with error bars are data, red curves are fits (46).

This relation also allows us to estimate the period. If the plane of orbit is perpendicular to the line of sight (axis Z), like the orbit N in Figure 14(a), then it is possible to simply substitute: $v_{ij}/2 \to w$, in (32), $v_{ij} \to w$, in (34) and $\Delta_{ij} \to a$ in (36), (32), and (34).

Then we get:

$$T_g = 0.937 \times 10^8 \sqrt{\frac{\Delta_{ij}^3}{M_{\text{tot}}}}, \qquad T_v = T_A = T_B = \frac{2\pi\Delta_{ij}}{v_{ij}}. \quad (37)$$

The orbit reference frame in the figure is similar to the local 3D event frame defined by the basis (2) and coordinates (7). But its origin L_0 is defined by the actual position of the center of mass of the binary. How does one deal with the orbits whose plane is not perpendicular to the line of sight like the orbit K in the same figure? The orbits in the figure are defined as follows:

$$N : X = r\cos\phi, \qquad Y = r\sin\phi, \, Z = 0$$
$$K : X = r\cos\phi\cos\theta, \;\; Y = r\sin\phi, \;\; Z = r\cos\phi\sin\theta$$
$$P : X = r\cos\phi\cos\theta, \;\; Y = r\sin\phi, \;\; Z = 0$$

where $r = a/2$ for the case (A), $r = a$ for (B), and ϕ is azimuthal angle in the plane XY. The orbit K inclined at an angle θ is observed only in its projection P. Corresponding observed separation between the stars is Δ_{ij}:

$$a \to \Delta_{ij}, \;\; \sqrt{a^3} \to \sqrt{\Delta_{ij}^3} = \sqrt{a^3}\,\varepsilon_1,$$
$$\varepsilon_1 = (\cos^2\phi\cos^2\theta + \sin^2\phi)^{3/4}. \quad (38)$$

Random angles ϕ, θ generate distribution of ε_1 shown in Figure 14(b). The MC distribution demonstrates smearing of

the real separation a due to random ϕ and slope θ of the orbit. Similarly, the ratio a/w is distorted as

$$\frac{a}{w} \to \frac{\Delta_{ij}}{v_{ij}} = \frac{a}{w}\varepsilon_2, \qquad \varepsilon_2 = \frac{\sqrt{\cos^2\phi\cos^2\theta + \sin^2\phi}}{\sqrt{\sin^2\phi\cos^2\theta + \cos^2\phi}}. \quad (39)$$

Since velocity w is perpendicular to r, there is exchange $\cos^2\phi \rightleftarrows \sin^2\phi$ in denominator. Corresponding distribution of ε_2 is shown in Figure 14(c). The mean values are

$$\langle \varepsilon_1 \rangle = 0.791, \qquad \langle \varepsilon_2 \rangle = 1.21$$

and represent a scale of distortion of real orbital periods, if replaced by relations (37). More accurate estimate of the periods in some region of Δ_{ij} can be obtained by rescaling of these relations:

$$\langle T_v \rangle = \frac{2\pi}{\langle \varepsilon_2 \rangle}\left\langle \frac{\Delta_{ij}}{v_{ij}} \right\rangle, \qquad \langle T_g \rangle = \frac{0.937 \times 10^8}{\langle \varepsilon_1 \rangle \sqrt{M_{\text{tot}}}}\langle \sqrt{\Delta_{ij}^3} \rangle. \quad (40)$$

We have estimated the average periods from the maximum in Figure 13. If we take the sources roughly in the region of half width of the maximum,

$$\Delta \leqslant 0.015 \text{ pc}, \qquad v_{ij} \leqslant 1.5 \text{ km s}^{-1}, \quad (41)$$

then

$$\langle T_v \rangle \approx 8.0 \times 10^4 \text{ y} \quad (42)$$

and one can check that equality $\langle T_v \rangle = \langle T_g \rangle$ implies estimation $M_{\text{tot}} \approx 0.8\,M_\odot$.
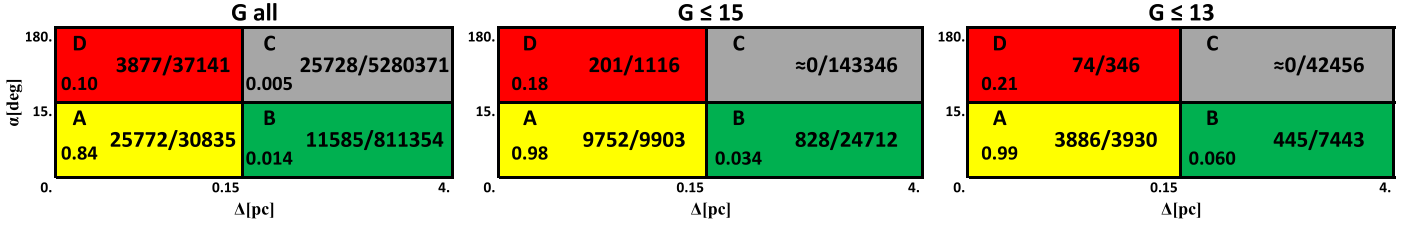
Figure 16. Ratio N_{bin}/N_2 of the number of binaries and the total number of pairs in different domains $\Delta \times \alpha$ and intervals of magnitude G.

5. Discussion

The separation of a pair of stars and the similarity of their movements can serve as two signatures of the binaries. We can compare them:

(i) Distributions of the 2D and 3D separations are studied in Sections 3 and 4.1. The procedure is simple, the distribution of separations $P(\Delta)$ (within the defined circles or spheres) is compared with the corresponding separations $P_{bg}(\Delta)$ generated by uniformly distributed sources representing background

$$P_{bg}(\Delta)d\Delta = N_{bg}P_0(\hat{\xi})d\hat{\xi}; \qquad 2\rho\hat{\xi} = \Delta = d_{ij} \text{ or } \Delta_{ij}, \quad (43)$$

where P_0 is given by (12) or (18) and N_{bg} is the corresponding number of the background pairs in the data events. Its accurate calculation is described below, see Equations (47)–(49). The binary distribution reads

$$P_{\mathrm{bin}}(\Delta) = P(\Delta) - P_{bg}(\Delta)$$

and the probability $\beta(\Delta)$ that the pair is a real binary is given as

$$\beta(\Delta) = \frac{P_{\mathrm{bin}}(\Delta)}{P(\Delta)} = 1 - \frac{P_{bg}(\Delta)}{P(\Delta)}. \quad (44)$$

The function $R(\Delta)$ displayed in Figures 7(f) and (l), 8(c), 9(i), and 10(i) is another representation of the probabilistic function $\beta(\Delta)$:

$$R(\Delta) \approx \frac{1}{1 - \beta(\Delta)}. \quad (45)$$

Consider the distribution $P(\Delta)$ from Figures 9(k) and 10(k). The corresponding function $\beta(\Delta)$ is shown in Figure 15(a), where the red curve represent fit of the function:

$$\beta(\Delta) = \omega \exp\left(-\frac{\Delta^2}{2\sigma_2^2}\right) + (1 - \omega)\exp\left(-\frac{\Delta}{\sigma_1}\right). \quad (46)$$

The result of the fit is shown in the first row (all pairs) of Table 3. The longer tail corresponding to the second term describes the small probability of bound pair at greater separations: $\Delta \gtrsim 0.15$ pc. This tail is not visible in panels (i) in Figures 9 and 10. Two exponential terms in $\beta(\Delta)$ may correspond to two different classes of binaries. The question is to what extent the excess of wide binaries consists of stable bound systems. Part of the excess may be an image of widening pairs that were less separated but weakly bound in the past. The accuracy of the method is based on three conditions:

1. precise separation measurement in a suitably selected statistical set of events that generates $P(\Delta)$,
2. precise modeling of the background defining $P_{bg}(\Delta)$, and

Table 3
Values of the Parameters ω, σ_1, σ_2

	ω	σ_1(pc)	σ_2(pc)
All Pairs	0.852	0.335	0.0303
All Pairs and Domain **AB**	0.771	0.396	0.0573
G ⩽ 15 and Domain **AB**	0.705	0.464	0.143
G ⩽ 13 and Domain **AB**	0.505	0.326	0.215

3. relatively high peak and low background giving probability $\beta(\Delta)$ close to 1 in the peak region.

(ii) In principle, a similar approach could be applied to comoving pairs. However, it is obvious that meeting the conditions above is more difficult for velocities or their differences. The precision of velocity measurement is lower than separation. The distribution of velocities is far from being simply uniform, we do not know the exact form of the velocity background. Figure 11(b) suggests a more complicated collective motion in the background and a relatively low peak of binaries. That is why we prefer the primary signature based on the spatial separation, where we work with exactly defined background and relatively high peaks, like panels (i) and (k) in Figures 9 and 10. On the other hand, one can expect that a combination of position and velocity data will suppress background and improve the selection of binaries in terms of the function $\beta(\Delta)$. A selection based on such a combination is described below. In Section 4.2 we have shown that combination of the space separation with proper motion allows us to estimate the orbital period.

We denote by N_{bin}^D and N_{bg}^D the numbers of real and false (background) binaries in some domain D of separations Δ. Correspondingly we denote $N_2^D = N_{\mathrm{bin}}^D + N_{bg}^D$, where N_2^D is the number of pairs in a given data set. We have

$$N_{bg}^D = \int_D P_{bg}(\Delta)d\Delta, \quad (47)$$

where P_{bg} is distribution defined by (18) and renormalized in such a way that

$$\int_E P_{bg}(\Delta)d\Delta = \int_E P(\Delta)d\Delta \quad (48)$$

where the integration is over the domain E safely outside the peak of binaries. So we calculate

$$N_{bg}^D = N_2^E \frac{\int_{D(\Delta)} P_0(\Delta)d\Delta}{\int_{E(\Delta)} P_0(\Delta)d\Delta}, \quad (49)$$

where P_0 is defined by (18) with the substitution $\hat{\xi} = \Delta/2\rho$. The same procedure can be applied for the domains $D(\Delta) \times D(\alpha)$. Since we can assume that Δ and α are not

Table 4
The Number of Binaries N_{bin} with the Total Number of Sources N_s in Different Intervals of the Magnitude G

	N_s	N_{bin}	N_{bin}/N_s
All G	2734,222	66,962	0.024
$G \leqslant 15$	474,689	10,781	0.023
$G \leqslant 13$	256,787	4405	0.017

correlated in the background distribution

$$P_{bg}(\Delta, \alpha) \approx P_0(\Delta)P(\alpha), \qquad (50)$$

then

$$N_{bg}^D = N_2^E \frac{\int_{D(\Delta)} P_0(\Delta)d\Delta}{\int_{E(\Delta)} P_0(\Delta)d\Delta} \frac{\int_{D(\alpha)} P(\alpha)d\alpha}{\int_{E(\alpha)} P(\alpha)d\alpha}. \qquad (51)$$

If we choose $D(\alpha) = E(\alpha)$, then the second ratio is 1. In this way, we can calculate N_{bg}^D without the knowledge of $P(\alpha)$. The selected domains are shown, together with the results $\beta \approx N_{bin}/ N_2$ in Figure 16. For the calculation, we take the common interval $E(\Delta) \equiv \langle 2, 4\rangle$ pc, where N_{bin}^E should be zero. In accordance with Figure 12(a) we observe the highest rate of binaries in the domain $A \equiv \langle 0, 0.15\rangle$ pc $\times \langle 0, 15\rangle$ deg. The binaries are well observable also in the neighboring domain $B \equiv \langle 0.15, 4\rangle$ pc $\times \langle 0, 15\rangle$ deg. For $G \leqslant 15$ the number of binaries in the domains C is zero within statistical errors. Presence of binaries in the domain D requires further analysis.

The probabilistic function $\beta(\Delta)$ corresponding to the united domain AB is shown in panels (b)–(d) in Figure 15 for different intervals of the magnitude G. Obviously, the function (46) fitted with the parameters listed in Table 3 can be rather approximate. We observe that function β is getting wider with decreasing G. This may suggest that brighter and therefore statistically more massive stars can form stable bound systems even at greater separations. Obviously, for $\Delta \gtrsim 1$ pc, the probability β is compatible with zero. This also corresponds to an absence of a binary peak in Figure 11(d), which relates to $\Delta > 1$ pc. Total numbers of binaries are listed in Table 4. The results suggest that in the *Gaia* DR2 data in the region defined by Table 2, the number of binaries can represent $\approx 2\%$ of all stars in this region.

6. Catalog

In this section, we describe the catalog of binary candidates, which we have created from the events of multiplicity $2 \leqslant M \leqslant 15$ defined by Table 2. For the first version of the catalog we accept only the candidates from domain A shown in the first panel in Figure 16. So, we do not accept all candidates, but only candidates with a high probability to be the true binary. The candidates meet the following conditions:

(1) Projection of separation
We accept the pairs, which satisfy

$$\Delta_{cat} \leqslant 0.15 \text{ pc}. \qquad (52)$$

In general, the projection of separation Δ depends on the reference frame. In the paper, we worked with the local reference frame defined by the event center, where projection Δ into the local plane XY is given by (15). In the catalog we do not use local frames. The cut (52) is applied to Δ_{cat}, which is

defined as the length of the arc

$$\Delta_{ij,cat} = \frac{1}{2}(L_i + L_j)\arccos \boldsymbol{n}_i\boldsymbol{n}_j, \qquad (53)$$

where L_α, $\boldsymbol{n}_\alpha$ are defined in Section 2.1. The separations Δ and Δ_{cat} are not exactly equal, but we have checked that in our conditions their difference is small, $\langle\delta\Delta\rangle \approx 0.007$ pc. Then the sharp cut on Δ_{cat} means only a slightly smeared cut of the distribution of Δ.

(2) Projection of collinearity
The pairs must meet the condition

$$\alpha \leqslant 15°. \qquad (54)$$

Both conditions define the domain A in the first panel of Figure 16. The panel shows that the average probability of a binary star is $\beta \approx 84\%$. If the stars are brighter, (second and third panel), then the average β is almost 100%.

(3) Radial separation
In fact, the radial separation is not explicitly used in our algorithm for selection of binaries. The reason is a rather low precision of radial separations, as explained in the discussion of panels (c), (e), and (f) in Figures 9 and 10. The only constraint is given by the diameter of our events (4 pc). Separation selection is based solely on Δ_{ij}. Additional cuts on inaccurate radial separation would eliminate many of the real binaries and invalidate the function β calculated for Δ_{ij}. We obtained high β even without a cut on radial separation.

Furthermore, it is evident that spherical events fill the space only partially ($\approx 52\%$). In this way, half the stars are lost for analysis. We also lose binaries between two adjacent events when each star falls into another. In order to recover these losses, we work with the modified coverage:

(i) The event spheres are replaced by cubes of edge 4 pc with no gaps between them. In each cube, we search for the pairs meeting the conditions (52) and (54).
(ii) The procedure is repeated with the same cubes centered in the corners of the former cubes and the search results are merged.

The catalog is represented by a matrix that is defined as follows. Each line represents one star and the following data are in the columns:

1–2: Group ID and Group size ($n \geqslant 2$) to match stars with the group they belong to.[2]

3–96: Copy of the original entry for the star from *Gaia*-DR2 archive,[3] according to the documentation.[4]

97–98: Minimum and maximum angular separation of the star from other stars in the group [as].

99–100: Minimum and maximum projected physical separation of the star from other stars in the group [pc].

The summary data from our catalog of binary candidates (I), along with the data extracted from the catalog JEC - Jiménez-Esteban et al. (2019) (II) is shown in Table 5. The number of candidates N_2 in this table correspond to N_2 in the domains A in Figure 16 after increasing with the repeated covering. The comparison of the two catalogs shows the following:

---

[2] In the current version of our catalog we omit candidates of systems n > 2, so only binaries (n = 2) are written.
[3] http://cdn.gea.esac.esa.int/Gaia/gdr2/gaia_source/csv/
[4] http://gea.esac.esa.int/archive/documentation/GDR2/Gaia_archive/chap_datamodel/sec_dm_main_tables/ssec_dm_gaia_source.html

Table 5
Summary Table of Binaries from the Catalogs (I) and (II), See the Text

Catalog	N_2	(N_{bin})	G
(I): Domain A	80,560	(67,670)	All
	22,674	(22,201)	$\leqslant 15$
	9082	(8991)	$\leqslant 13$
(II): Total $N_{bin} = 3055$		(301)	$\leqslant 13$

Note. N_2 is the number of binary candidates, N_{bin} is the real expected number of binaries. The data in the second column are related to the full cube region $(400\ \text{pc})^3$ (Table 2).

(a) We can only compare sources of magnitude $G \leqslant 13$, because (II) does not contain less bright stars. Of the total number 3055 binaries in (II), only 301 lie in the (I) cube $(400\ \text{pc})^3$. Increasing the edge of this cube by the factor 2.15 would increase volume 10 times with the number of binaries comparable to the total (II).

(b) The number 301(II) could be compared with the corresponding number of candidates $C_2 = 9082$(I). However, in the JEC catalog, apart from the cut $G \leqslant 13$ many other restrictions and selections are made. Definition of the binary is not the same in both catalogs. In our opinion, this is the reason for the difference between (I) and (II).

(c) (I) and (II) have 108 common binary candidates.

(d) 86(II) candidates are absent in (I) since the separation Δ exceeds 0.15 pc. These candidates do not contradict our general criteria, but the corresponding probability $\beta(\Delta)$ can be lower as shown in Figure 15.

(e) 54(II) candidates are absent in (I) since their spatial separation exceeds 4 pc (event diameter). Such candidates may not contradict our criteria; however, $\beta(\Delta)$ due to a great background can be extremely low.

(f) 18(II) candidates are absent in (I) since these candidates are located in a dense area generating the high multiplicity events, which exceed our currently set limit.

(g) The last 35(II) candidates are absent in (I), mainly due to the fact that even after the second coverage some couples (II) remain separated in two neighboring event cubes.

The total number of the binary candidates of all *Gaia* magnitudes in the catalog (I) is 80,560, which corresponds to the expected real number of binaries 67,670. The full current catalog (I) in the csv form is available on the website https://www.fzu.cz/~piska/Catalogue/. We plan to further develop and optimize our catalog methodology.

7. Summary and Conclusion

We have proposed a general statistical method for analysis of finite 2D and 3D patterns. In the present study, the method has been applied to the analysis of binary star systems in different regions of the *Gaia* catalog DR2.

Results on 2D statistical analysis were compared with our former results obtained from the previous catalog DR1. The new results give in the distribution of angular separations more clear evidence of binaries. Independent signature follows from the characteristic functions $\Theta_n(M)$, which clearly indicate a tendency to clustering. However, the most important results are obtained from the 3D analysis introduced in the present paper. We have analyzed about 5×10^5 of events inside the cube of edge 400 pc centered at the origin of the galactic reference frame. In distributions of pair separations, we observe the sharp peaks at small separations corresponding to binaries, which are more striking for brighter sources, $G \leqslant 15$.

The important result of the analysis is probabilistic function $\beta(\Delta)$, which depends on the separation Δ of a pair of stars and indicates the probability that the pair constitutes a bound system. The function suggests that brighter, more massive binary stars have on average a greater separation. With increasing separation the function falls rapidly. We obtained the ratio binaries/singles $\approx 2\%$.

Furthermore, we had shown that a combined analysis of 3D separations with the proper motion of the pairs of sources gives a clear picture of the binaries with two components of the motion: parallel and orbital. The analysis allowed us to estimate the average orbital period and mass of the binary star system in the chosen statistical ensemble.

The highest probability of the binary is observed at smallest separations Δ and angles α between proper motions. From the corresponding domain $\Delta \times \alpha \equiv \langle 0, 0.15 \rangle\ \text{pc} \times \langle 0, 15 \rangle\ \text{deg}$, we have created the catalog involving 80,560 binary candidates, which represents 67,670 of the true binaries.

Appendix
Proof of Relations (12), (13), and (17)–(19)

(i) Relation (12)

First we consider two random points on a segment L. The probability that the points are separated by interval l reads:

$$p(l) \sim L - l. \tag{55}$$

Furthermore, we suppose a circle of diameter $2R = 1$ with a chords $L(y)$ involving random segment l (Figure 17). We have

$$L(y) = 2\sqrt{(1/2)^2 - y^2} = \sqrt{1 - (2y)^2}. \tag{56}$$

The probability of interval l reads

$$P_{\parallel}(l) \sim \int_0^{y_{max}} (L(y) - l)dy \sim (\arccos l - l\sqrt{1 - l^2});$$

$$y_{max} = \frac{1}{2}\sqrt{1 - l^2}, \quad 0 < l < 1. \tag{57}$$

This distribution is generated by random pairs on the chords parallel to axis x. For arbitrary random pairs separated by l inside the circle, we integrate distributions (57) over all directions in 2D and replace

$$dl \rightarrow d(\pi l^2) \sim ldl, \tag{58}$$

which gives distribution

$$P_{\parallel}(l)dl \rightarrow P(l)dl \sim P_{\parallel}(l)ldl, \quad 0 < l < 1. \tag{59}$$
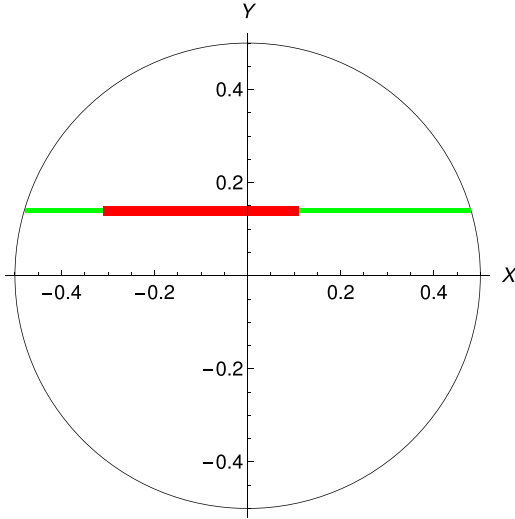
Figure 17. Random segment l (red) on the chord $L(y)$ (green).

Relation (12) is its normalized form.

(ii) Relation (13)

Distribution (59) can be modified

$$P(l)dl \sim P_{\parallel}(l)ldl \sim P_{\parallel}(\sqrt{l_x^2 + l_y^2})dl_x dl_y. \quad (60)$$

Calculation of integral

$$P(l_x) \sim \int_0^{l_{y\,\max}} P_{\parallel}(\sqrt{l_x^2 + l_y^2})dl_y, \quad l_{y\,\max} = \sqrt{1 - l_x^2} \quad (61)$$

with the use of Wolfram Research, Inc. (2019), and after replacement $l_x \rightarrow \hat{\xi}$ and normalization gives relation (13).

(iii) Relation (17)

Now instead of circle $2R = 1$, we consider the sphere of the same radius. The procedure is a modification of the case *(i)*. Now instead of integral (57) we get

$$P_{\parallel}(l) \sim \int_0^{y_{\max}} y(L(y) - l)dy \sim \left(\frac{1}{3} - \frac{l}{2} + \frac{l^3}{6}\right);$$
$$y_{\max} = \frac{1}{2}\sqrt{1 - l^2}, \quad 0 < l < 1, \quad (62)$$

where y means radius of a cylinder of parallel chords. The additional y in the integral means that we integrate chords on surfaces of cylinders of different radii. Then instead of (58) we use

$$dl \rightarrow d\left(\frac{4}{3}\pi l^3\right) \sim l^2 dl, \quad (63)$$

since the integration of chords is over all directions in 3D. The resulting distribution reads

$$P(l)dl \sim P_{\parallel}(l)l^2 dl \sim l^2\left(\frac{1}{3} - \frac{l}{2} + \frac{l^3}{6}\right), \quad (64)$$

which after normalization gives relation (17).

(iv) Relation (19)

Probability $\boldsymbol{P}(l)$ of random segments $l = \sqrt{l_x^2 + l_y^2 + l_z^2}$ inside the sphere can be expressed as

$$\boldsymbol{P}(\sqrt{l_x^2 + l_y^2 + l_z^2})dl_x dl_y dl_z \sim \boldsymbol{P}(l)l^2 dl \sim \boldsymbol{P}(l)dl, \quad (65)$$

which together with (64) gives

$$\boldsymbol{P}(\sqrt{l_x^2 + l_y^2 + l_z^2}) = \boldsymbol{P}(l) \sim \left(\frac{1}{3} - \frac{l}{2} + \frac{l^3}{6}\right). \quad (66)$$

The probability that segment l has projection l_x is given as

$$P(l_x) \sim \int_0^{l_{z\,\max}} \int_0^{l_{y\,\max}} \boldsymbol{P}(\sqrt{l_x^2 + l_y^2 + l_z^2})dl_y dl_z \quad (67)$$

$$\sim \int_0^{t_{\max}} \boldsymbol{P}(\sqrt{l_x^2 + t^2})t\,dt; \quad t_{\max} = \sqrt{1 - l_x^2}. \quad (68)$$

The last integral (equally for l_y, l_z) can be after inserting from (66) easily calculated, and after normalization gives relation (19).

(v) Relation (18)

In a similar way, the probability that segment l has projection $\Delta = \sqrt{l_x^2 + l_y^2}$ is given as

$$P(\Delta) \sim \int_0^{l_{z\,\max}} \boldsymbol{P}(\sqrt{\Delta^2 + l_z^2})\Delta dl_z; \quad l_{z\,\max} = \sqrt{1 - \Delta^2}, \quad (69)$$

which after inserting from (66) and integration with the use of Wolfram Research, Inc. (2019) implies relation (18).

ORCID iDs

Petr Zavada ⓘ https://orcid.org/0000-0002-8296-2128

References

Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17
Caballero, J. A. 2009, A&A, 507, 251
Close, L. M., Richer, H. B., & Crabtree, D. R. 1990, AJ, 100, 1968
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016b, A&A, 595, A2
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016a, A&A, 595, A1
Jiménez-Esteban, F. M., Solano, E., & Rodrigo, C. 2019, AJ, 157, 78
Oelkers, R. J., Stassun, K. G., & Dhital, S. 2017, AJ, 153, 259
Oh, S., Price-Whelan, A. M., Hogg, D. W., Morton, T. D., & Spergel, D. N. 2017, AJ, 153, 257
Wolfram Research, Inc 2019, Mathematica, Version 12.0, Champaign, IL
Zavada, P., & Píška, K. 2018, A&A, 614, A137
Ziegler, C., Law, N. M., Baranec, C., et al. 2018, AJ, 156, 259