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

Petr Zavada and Karel Piška

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

Received 2019 March 5; revised 2019 November 7; accepted 2019 November 15; published 2020 January 2

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.

**Key words:** astrometry – binaries: general – methods: statistical – parallaxes – proper motions

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 & Piška 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 \( \mathbf{L} \) of a source is defined by spherical coordinates \( L, l, \) and \( b \) (distance from the Sun, galactic longitude and latitude):

\[
\mathbf{L} = L \mathbf{n}; \quad \mathbf{n} = (\cos b \cos l, \cos b \sin l, \sin b),
\]

\[
-\pi < b < \pi, \quad -\pi < l < \pi.
\]  

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

\[
\mathbf{k}_r = \mathbf{n}_0 = (\cos b_0 \cos l_0, \cos b_0 \sin l_0, \sin b_0),
\]

\[
\mathbf{k}_l = (-\sin l_0, \cos l_0, 0),
\]

\[
\mathbf{k}_b = (-\sin b_0 \cos l_0, -\sin b_0 \sin l_0, \cos b_0),
\]  

where \( \mathbf{k}_r = \mathbf{n}_0(b_0, l_0) \) defines angular position of the event center. Unit vector \( \mathbf{k}_l \) is perpendicular to \( \mathbf{k}_r \) and has direction of increasing \( l \), see Figure 2. Unit vector \( \mathbf{k}_b \) is defined as \( \mathbf{k}_b = \mathbf{k}_r \times \mathbf{k}_l \) and has direction of increasing \( b \). Vector \( \mathbf{k}_l \) has radial direction, perpendicular vectors \( \mathbf{k}_b \) and \( \mathbf{k}_l \) 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 \( \mathbf{n}_i \) inside a circle with the event center \( \mathbf{n}_0 \)
\[ |n_i - n_0| \leq \rho_2, \quad i = 1, \ldots M. \]  

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

\[ x_i = n'_i k_i, \quad y_i = n'_i k_b; \quad n'_i = n_i - n_0. \]  

We define the 2D event as the set:

\[ \{x_i, y_i\}; \quad x_i^2 + y_i^2 \leq \rho_2^2, \quad i = 1, \ldots M. \]  

Since the DR2 catalog involves also data on parallaxes, we can similarly also generate the 3D events—patterns of the \( M \) sources with position \( L_i \) inside the spheres with the center \( L_0 \) and radius \( \rho_3 \):

\[ |L_i - L_0| \leq \rho_3, \quad i = 1, \ldots M. \]  

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

\[ X_i = n_i k_x, \quad Y_i = n_i k_y, \quad Z_i = n_i k_z; \]  

\[ N_i = L_i n_i - L_0, \quad L_i [\text{pc}] = \frac{1000}{p_i [\text{mas}].} \]  

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

\[ \{X_i, Y_i, Z_i\}; \quad X_i^2 + Y_i^2 + Z_i^2 \leq \rho_3^2, \quad i = 1, \ldots M. \]  

### 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 \( (Q = M_1 n) \), for instance a tendency to clustering \( (Q > M_1 n) \) or anticlustering \( (Q < M_1 n) \). Details of the method are described in Zavada & Pîška (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

\[ x_{ij} = |x_i - x_j|, \quad y_{ij} = |y_i - y_j|, \quad d_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2}, \]  

or scaled ones

\[ \hat{x}_{ij} = \frac{x_{ij}}{2\rho_2}, \quad \hat{y}_{ij} = \frac{y_{ij}}{2\rho_2}, \quad \hat{d}_{ij} = \frac{d_{ij}}{2\rho_2}. \]  

The suitable unit of the parameters \( x, y, \rho_2 \) 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

\[ P(\hat{\xi}) = \frac{16\hat{\xi}}{\pi} (\arccos \hat{\xi} - \hat{\xi}\sqrt{1 - \hat{\xi}^2}); \quad \hat{\xi} = \hat{d}_{ij}. \]
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 $\hat{x} < 0.1$. 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

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

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

and/or with the scaled ones

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

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

The suitable unit of the parameters $X_i, Y_i, Z_i, \rho_3$ 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

$$P(\hat{\xi}) = 12\frac{2^4}{5}(1 - \hat{\xi}^2)(1 + 3\hat{\xi}^2 + \hat{\xi}^4); \quad \hat{\xi} = \hat{D}_{ij},$$

$$P(\hat{\xi}) = \frac{9}{2}\lambda\left((\hat{\xi})^2 - \hat{\xi}^4\right)\left(\sqrt{1 - \hat{\xi}^2} + \hat{\xi}^2(4 - \hat{\xi}^2)\right) \times \ln\left(\frac{\hat{\xi}}{1 + \sqrt{1 - \hat{\xi}^2}}\right); \quad \hat{\xi} = \hat{\Delta}_{ij},$$

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

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 $XY, YZ, a n d XZ$. 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 $\rho_2$ or $\rho_3$.

### 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 & Piška (2018), which allows us
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 \( d_{ij} \leq 0.5 \) 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 \( d_{ij} \leq 0.06 \) or equivalently for \( d_{ij} \leq 8.6 \) as. For greater separations inside the event, we observe perfectly uniform distributions of stars. Note the data and curves are equally normalized for \( \hat{x} < 0.1 \). That is why the strong peak in panels (g)–(i) is balanced by a small reduction of distribution outside the peak. Brighter stars \( (G \leq 15, \) 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 \( M \leq 8 \) the slopes are \( \approx 4\% (8\%) \) for the DR1(DR2) data. In Figure 8 we have shown

| 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 \(( -180, 180) \times (\pm 60, \pm 80)\) | 72 | 1290 | 3.21 | 2055,674 |
| C \(( 140, 180) \times ( -10, 10)\) | 18 | 1912 | 2.75 | 3588,183 |

Note. Only sources in distance 1–5000 pc are taken into account. The analysis is done for events \( 2 \leq M \leq 15 \).
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 $G \lesssim 15$, which proves the presence of binaries. Panel (b) again demonstrates the absence of double stars separated by $d_{ij}(d_{ij}) \lesssim 0.5$ as(0.014) due to insufficient resolution. Different scales of $d_{ij}$ in Figures 7(l) and 8(c) are due to different radii $\rho_2$ 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).
Region of events. The analysis is done for events 2 and only sources of positive parallax are included.

Note. The same distributions but for brighter sources G ≤ 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 < \( \xi < 1 \) 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

\[
\Delta_{ij} \lesssim 0.1 - 0.2 \text{ pc.}
\]  

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 (\( \rho_3 = 5 \text{ pc} \)) 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

\[
\mu_\alpha \equiv \mu_\alpha \cos \delta, \quad \mu_\delta
\]

in directions of the R.A. and decl. in the ICRS. So the corresponding transverse 2D velocity \( U \) is given as

\[
U = L(\mu_\alpha, \mu_\delta), \quad U = |U|,
\]

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

\[
\alpha_{ij} = \arccos \left( \frac{U_i \cdot U_j}{U_i U_j} \right), \quad U_{\delta ij} = |U_i + U_j|, \quad v_{ij} = |U_i - U_j|,
\]  

```

**Figure 8.** Panels (a)–(c): distributions of angular separations in the region C for sources \( G \leq 15 \). 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 | Cube Centered at the Origin of the Galactic Reference Frame |
|-----------------------|-------------------------------------------------------------|
| \( \rho_3 \) (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. \( \rho_3 \) 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 \( 2 \leq M \leq 15 \).
where $\alpha_{ij}$ is the angle between both transverse velocities. The corresponding errors read

$$\delta \alpha_{ij} \approx \frac{\sqrt{2} \delta u}{u}, \quad \delta U_{ij} = \delta v_{ij} \approx \sqrt{2} \delta u,$$

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 (b). Panels (j) and (k) are magnified versions of (d) and (h). Unit of separations $X_{ij}, Y_{ij}, Z_{ij}, D_{ij}, \Delta_{ij}$ is 1 pc.

$$u = L\mu, \quad \delta u \approx u \left( \frac{\delta L}{L} \right)^2 + \left( \frac{\delta \mu}{\mu} \right)^2.$$
Note that relative error $\frac{\delta v_{ij}}{v_{ij}}$ can be large, since $v_{ij}$ is small compared with $U_{ij}$ and the errors $\delta v_{ij}$ and $\delta U_{ij}$ 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 $\alpha_{ij}$ is presented in Figures 11(b)–(d) for different regions of $\Delta_{ij}$. In Figure 12(a) we have shown

![Figure 10](image-url) Same distributions as in the previous figure, but for bright pairs, $G \lesssim 15$. 

![Figure 11](image-url)
the correlation of pair transverse separations and angles $\alpha_{ij}$. We observe a very narrow peak in the region of small $\Delta_{ij}$ and $\alpha_{ij}$.\(^1\) The peak is connected with presence of binaries as follows. The transverse velocities of two gravitationally coupled stars are

$$U_i = V + v_i, \quad U_j = V + v_j,$$

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

$$v_i = |v_i| \ll |V|,$$

\(^1\) Input data (23) are related to the ICRS and $\Delta_{ij}$ is calculated in the galactic reference frame. Nevertheless the parameters $\Delta_{ij}$, $\alpha_{ij}$, $U_i$, $v_i$ are invariant under rotation.
so for binaries in our \( \Delta \) window (\( \Delta_{\text{min}} \) is given by resolution of two close sources and \( \Delta_{\text{max}} = 0.1 \text{ pc} \) by (22)) we have

\[
U_{ij} \approx 2V, \quad v_{ij} = v_i + v_j.
\]

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 \( \alpha_{ij} \) and \( \Delta_{ij} \)). The corresponding \( \alpha_{ij} \) peak is also observed in Figures 11(b)–(d) on the background of the collective motion of stars (dominance of \( \alpha_{ij} < 90^\circ \)). The peak is suppressed for \( \Delta_{ij} \gtrsim 1 \text{ pc} \). 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 \( \Delta_{ij} \) in the region of small separations, where the binaries are present. With the use of \( v_{ij} \) and \( \Delta_{ij} \) 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

\[
v_i = -v_j, \quad v_{ij} = v_i + v_j,
\]

Orbital period is

\[
T_A \approx \frac{\pi a}{w},
\]

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

(B) \( M_1 \gg M_2 \), then

\[
v_2 \approx 0, \quad v_{ij} \approx v_1,
\]

but the orbital period is different

\[
T_B \approx \frac{2\pi a}{w},
\]

since the separation \( a \) equals to the orbit radius.

At the same time, Kepler’s law implies for orbital period \( T_g \):

\[
T_g = 2\pi \sqrt{\frac{a^3}{GM_{\text{tot}}}},
\]

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

\[
T_g = 0.937 \times 10^8 \sqrt{\frac{a^3}{M_{\text{tot}}}}.
\]
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:

$$v \rightarrow \frac{v}{w}, \quad v \rightarrow w, \quad \Delta \rightarrow \Delta,$$

and

$$D \rightarrow ai.$$  

Then we get:

$$\rho = \frac{D}{T} = \frac{D}{T0},$$

or

$$\rho = \frac{D}{T} = \frac{D}{T0}.$$  

Similarly, the ratio $a/w$ is distorted as

$$\frac{a}{w} \rightarrow \frac{\Delta}{v}, \quad \frac{a}{w} \rightarrow \frac{a}{w} \varepsilon_2, \quad \varepsilon_2 = \frac{\sqrt{\cos^2 \phi \cos^2 \theta + \sin^2 \phi}}{\sqrt{\sin^2 \phi \cos^2 \theta + \cos^2 \phi}}.$$

Since velocity $w$ is perpendicular to $r$, there is exchange $\cos^2 \phi \equiv \sin^2 \phi$ in denominator. Corresponding distribution of $\varepsilon_2$ is shown in Figure 14(c). The mean values are

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

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 $\Delta_{ij}$ can be obtained by rescaling of these relations:

$$\langle T_0 \rangle = \frac{2\pi}{\langle \varepsilon_2 \rangle} \left( \frac{\Delta_{ij}}{v} \right), \quad \langle T_0 \rangle = \frac{0.937 \times 10^8}{\langle \varepsilon_1 \rangle \sqrt{M_{tot}}} \left( \sqrt{\Delta_{ij}} \right).$$

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,

$$\Delta \lesssim 0.015 \text{ pc}, \quad v_{ij} \lesssim 1.5 \text{ km s}^{-1},$$

then

$$\langle T_0 \rangle \approx 8.0 \times 10^4 \text{ y}$$

and one can check that equality $\langle T_0 \rangle = \langle T_0 \rangle$ implies estimation $M_{tot} \approx 0.8 M_\odot$. 

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, $G \lesssim 15$, $G \lesssim 13$. Blue points with error bars are data, red curves are fits (46).
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 distribution $P_{bg}(\Delta)$ generated by uniformly distributed sources representing background.

\[ P_{bg}(\Delta) d\Delta = N_{bg} P_0(\xi) d\xi; \quad 2\rho \xi = \Delta = d_{ij} \text{ or } \Delta_{ij}, \]

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

\[ P_{bin}(\Delta) = P(\Delta) - P_{bg}(\Delta) \]

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

\[ \beta(\Delta) = \frac{P_{bin}(\Delta)}{P(\Delta)} = 1 - \frac{P_{bg}(\Delta)}{P(\Delta)}. \]

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 $\beta(\Delta)$:

\[ R(\Delta) \approx \frac{1}{1 - \beta(\Delta)}. \]

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:

\[ \beta(\Delta) = \omega \exp \left( -\frac{\Delta^2}{2\sigma^2} \right) + (1 - \omega) \exp \left( -\frac{\Delta}{\sigma_1} \right). \]

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: $\Delta > 0.15$ 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
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 normal. The distribution of velocities is far from being normal. The precision of velocity measurement is lower than separation. The distribution of velocities is far from being normal. The distribution of velocities is far from being normal. The precision of velocity measurement is lower than separation. The distribution of velocities is far from being normal. The distribution of velocities is far from being normal. The precision of velocity measurement is lower than separation. The distribution of velocities is far from being normal. The distribution of velocities is far from being normal. The precision of velocity measurement is lower than separation. The distribution of velocities is far from being normal.

3. Table 3

| All Pairs | All Pairs and Domain AB | G \leq 15 and Domain AB | G \leq 13 and Domain AB |
|-----------|------------------------|------------------------|------------------------|
| $\omega$  | 0.852                  | 0.771                  | 0.505                  |
| $\sigma_1$ (pc) | 0.335                  | 0.396                  | 0.326                  |
| $\sigma_2$ (pc) | 0.0303                 | 0.0573                 | 0.215                  |

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

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

\[ \int_E P_{bg}(\Delta) d\Delta = \int_E P(\Delta) d\Delta \]

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

\[ N_{bg}^D = N_E^D \int_D \frac{P_0(\Delta) d\Delta}{\int_{E(\Delta)} P_0(\Delta) d\Delta}. \]

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

\[ P_{bg}(\Delta, \alpha) \approx P_0(\Delta)P(\alpha), \]  

(50)

then

\[ N_{bin}^D = N_2^E \int_{D(\Delta)} P_0(\Delta)d\Delta \int_{D(\alpha)} P(\alpha)d\alpha. \]  

(51)

If we choose \( D(\alpha) = E(\alpha) \), then the second ratio is 1. In this way, we can calculate \( N_{bin}^D \) 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 \( E(\Delta) \equiv (2, 4) \) pc, where \( N_{bin}^D \) should be zero.

In accordance with Figure 12(a) we observe the highest rate of binaries in the domain \( A \equiv (0, 0.15) \) pc \( \times \) \( (0, 15) \) deg. The binaries are well observable also in the neighboring domain \( B \equiv (0.15, 4) \) pc \( \times \) \( (0, 15) \) deg. For \( G \leq 15 \) 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 \( \beta \) 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 \( \Delta > 1 \) pc, the probability \( \beta \) 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 \( 2 \leq M \leq 15 \) 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

\[ \Delta_{cat} \leq 0.15 \text{ pc}. \]  

(52)

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

\[ \Delta_{cat} = \frac{1}{2}(L_i + L_j)\arccos(n_i.n_j), \]  

(53)

where \( L_{ij}, n_{ij} \) are defined in Section 2.1. The separations \( \Delta \) and \( \Delta_{cat} \) 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 \( \Delta_{cat} \) means only a slightly smeared cut of the distribution of \( \Delta \).

2. Projection of collinearity

   The pairs must meet the condition

\[ \alpha \leq 15^\circ. \]  

(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 \( \beta \) 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 \( \Delta_{ij} \). Additional cuts on inaccurate radial separation would eliminate many of the real binaries and invalidate the function \( \beta \) calculated for \( \Delta_{ij} \). We obtained high \( \beta \) 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 \( (n \geq 2) \) 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 |
| 97–98: Minimum and maximum angular separation of the star from other stars in the group \| 4 |

The summary data from our catalog of binary candidates (I), along with the data extracted from the catalog JEC - Jiménez-Estevez 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:

\[ \text{In the current version of our catalog we omit candidates of systems } n > 2, \text{ so only binaries } (n = 2) \text{ are written.} \]

<http://cdn.gea.esac.esa.int/Gaia/gdr2/ Gaia_source/csv/  
<http://gea.esac.esa.int/archive/documentation/GDR2 Gaia_archive/  
chap_datamodel/sec_dim_main_tables/sec_dim_gaia_source.html>
The Astronomical Journal, 159:33 (15pp), 2020 January

Zavada & Příška

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) | \(\leq 15\) |
| | 9082 (8991) | \(\leq 13\) |
| (II): Total \(N_{bin} = 3055\) | (301) | \(\leq 13\) |

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 pc\(^3\)) (Table 2).

(a) We can only compare sources of magnitude \(G \leq 13\), because (II) does not contain less bright stars. Of the total number 3055 binaries in (II), only 301 lie in the (I) cube (400 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 IEC catalog, apart from the cut \(G \leq 13\) 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 \(\Delta\) 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 \(\Theta_\alpha(M)\), 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 \(5 \times 10^5\) 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, \(G \leq 15\).

The important result of the analysis is probabilistic function \(\beta(\Delta)\), which depends on the separation \(\Delta\) 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 have 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 \(\Delta\) and angles \(\alpha\) between proper motions. From the corresponding domain \(\Delta \times \alpha \equiv (0, 0.15)\) pc \(\times (0, 15)\) deg, we have created the catalog involving 80,560 binary candidates, which represents 67,670 of the true binaries.

This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the Gaia Multilateral Agreement. The work was supported by the project LTT17018 of the MEYS (Czech Republic). Furthermore, we are grateful to J. Grygar for deep interest and many valuable comments and J. Palouš and O. Teryaev for very useful discussions and inspiring comments.

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:

\[
p(l) \sim L - l.
\] (55)

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

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

The probability of interval \(l\) reads

\[
P(l) \sim \int_0^{\gamma_{\max}} (L(y) - l)dy \sim (\arccos l - l\sqrt{1 - l^2});
\]

\[
\gamma_{\max} = \frac{1}{2}\sqrt{1 - l^2}, \quad 0 < l < 1.
\] (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

\[
dl \rightarrow d(\pi l^2) \sim l dl,
\] (58)

which gives distribution

\[
P(l)dl \rightarrow P(l)dl \sim P(l)l dl, \quad 0 < l < 1.
\] (59)
Relation (12) is its normalized form.

(ii) Relation (13)

Distribution (59) can be modified

\[ P(l)dl \sim P(l)dl \sim P(\sqrt{l_1^2 + l_2^2})dl_1dl_2. \]  

Calculation of integral

\[ P(l_x) \sim \int_0^{l_{x \text{max}}} P(\sqrt{l_1^2 + l_2^2})dl_1, \quad l_{x \text{max}} = \sqrt{1 - l_2^2}, \]  

with the use of Wolfram Research, Inc. (2019), and after replacement \( l_x \to \xi \) 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

\[ P(l) \sim \int_0^{y_{\text{max}}} y(L(y) - l)dy \sim \left( \frac{1}{3} - \frac{l}{2} + \frac{l^3}{6} \right); \]

\[ y_{\text{max}} = \frac{1}{2} \sqrt{1 - l^2}, \quad 0 < l < 1, \]  

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

\[ dl \sim d\left(\frac{4}{3} \pi l^3\right) \sim l^2dl, \]  

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

\[ P(l)dl \sim P(l)l^2dl \sim l^2\left( \frac{1}{3} - \frac{l}{2} + \frac{l^3}{6} \right). \]  

which after normalization gives relation (17).

(iv) Relation (19)

Probability \( P(l) \) of random segments \( l = \sqrt{l_1^2 + l_2^2 + l_3^2} \) inside the sphere can be expressed as

\[ P(\sqrt{l_1^2 + l_2^2 + l_3^2})dl_1dl_2dl_3 \sim P(l)l^2dl \sim P(l)dl, \]  

which together with (64) gives

\[ P(\sqrt{l_1^2 + l_2^2 + l_3^2}) = P(l) \sim \left( \frac{1}{3} - \frac{l}{2} + \frac{l^3}{6} \right). \]  

The probability that segment \( l \) has projection \( l_x \) is given as

\[ P(l_x) \sim \int_0^{l_{x \text{max}}} \int_0^{l_{z \text{max}}} P(\sqrt{l_1^2 + l_2^2 + l_3^2})dl_2dl_3, \]

\[ \sim \int_0^{l_{x \text{max}}} P(\sqrt{l_1^2 + l_2^2})dl_1; \quad l_{x \text{max}} = \sqrt{1 - l_2^2}. \]  

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 \( \Delta = \sqrt{l_1^2 + l_2^2} \) is given as

\[ P(\Delta) \sim \int_0^{l_{\Delta \text{max}}} P(\sqrt{l_1^2 + l_2^2})\Delta dl_3; \quad l_{\Delta \text{max}} = \sqrt{1 - \Delta^2}, \]  

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