Article:

Akter, S. and Goodwin, S.P. (2019) Finding binary star fractions in any distribution. Monthly Notices of the Royal Astronomical Society, 488 (3). pp. 3446-3451. ISSN: 0035-8711

https://doi.org/10.1093/mnras/stz1939

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY
MNRAS **488,** 3446–3451 (2019) doi:10.1093/mnras/stz1939

Finding binary star fractions in any distribution

Suraiya Akter[1] and Simon P. Goodwin[1,2★]
[1]Department of Physics and Astronomy, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield, S3 7RH
[2]Humanitas College, Kyung Hee University, 1732 Deogyeong-daero, Yongin-si, Gyeonggi-do 17104, South Korea

Accepted 2019 July 9. Received 2019 July 9; in original form 2018 November 1

ABSTRACT

Candidate visual binary systems are often found by identifying two stars that are closer together than would be expected by chance. However, in regions with non-trivial density distributions, the 'random' distances between stars varies because of the background distribution, as well as the presence of binaries. We show that when no binaries are present, the distribution of the ratios of the distances to the nearest and tenth nearest neighbours, d_1/d_{10}, is always well approximated by a Gaussian with mean 0.2–0.3 and variance 0.16–0.19 for any underlying density distribution. The introduction of binaries causes some (or all) nearest neighbours to become closer than expected by random chance, introducing a component to the distribution where d_1/d_{10} is much lower than expected. We show how a simple single or double Gaussian fit to the distribution of d_1/d_{10} can be used to find the binary fraction in any underlying density distribution quickly and simply.

Key words: methods: statistical – stars: binaries.

1 INTRODUCTION

Many stars – quite possibly most stars – form in binary and multiple systems (Duchêne & Kraus 2013; Reipurth et al. 2014). As multiple systems are a typical outcome of star formation, the properties of these systems are a clue to how stars form, and a strong constraint on star formation theories (Goodwin 2010).

Many multiple systems are discovered through visual searches, by looking for two stars that are closer than would be expected through random chance (this has been done since Michell 1767). A fundamental problem with this method is quantifying at what distance(s) a pair is close enough to be considered a possible binary, and at what distance(s) it is probably a random alignment. This is often based on the Struve (1852) formula,

$$P = \frac{N(N-1)(\pi s^2)}{2A}, \tag{1}$$

where P is the probability that two stars will be closer than a separation s if they are two of N stars within an area A. The problem with this approach is that the surface densities of the stars almost always vary significantly across the region of interest in a non-trivial (substructured, centrally condensed, fractal, filamentary) way. A similar problem occurs when using more sophisticated methods such as the angular two-point correlation function (e.g. Bahcall & Soneira 1981; Dhital et al. 2010), which requires a model of the background distribution. This means that while it is useful for the field, for example, in complex (young) regions, it is unclear if the method is finding binaries or underlying structure. The same problem applies to the Larson (1995) approach (i.e. the 'Larson plot').

In this paper, we introduce a very simple method to analyse a region with any underlying density structure and to find the likely total binary fraction. This method is 'Struve-like', although it sets the area of interest as being that to the tenth nearest neighbour and considers the distribution of possible distances to the nearest neighbour as a ratio. This avoids variations in local density across the whole region.

Whilst we formulate and discuss this method in the context of finding binary stars in two-dimensional spatial data, it can also be applied to the search for the fraction of pairing above random chance in any data set.

2 METHODOLOGY

There are many ways of defining the proportion of stars that are multiples (see Reipurth & Zinnecker 1993 for a list). In this paper, we consider that stars are only in binary systems and we quantify the fraction as the 'binary fraction', $F_{\rm bin} = B/(S + B)$, where B is the number of binary systems and S is the number of single systems.

Stars in young regions have a wide variety of underlying density distributions, from substructured/fractal to spherical and centrally condensed, containing tens to thousands of stars; for example, see Cartwright & Whitworth (2004), who quantify the structure of various young regions with the Q-parameter.

In this paper, we analyse fractal regions (created using the method of Goodwin & Whitworth 2004) with fractal dimensions $D = 1.6$ (very clumpy), $D = 2$ (moderately clumpy) and $D = 3$ (smooth, not

★ E-mail: s.goodwin@sheffield.ac.uk

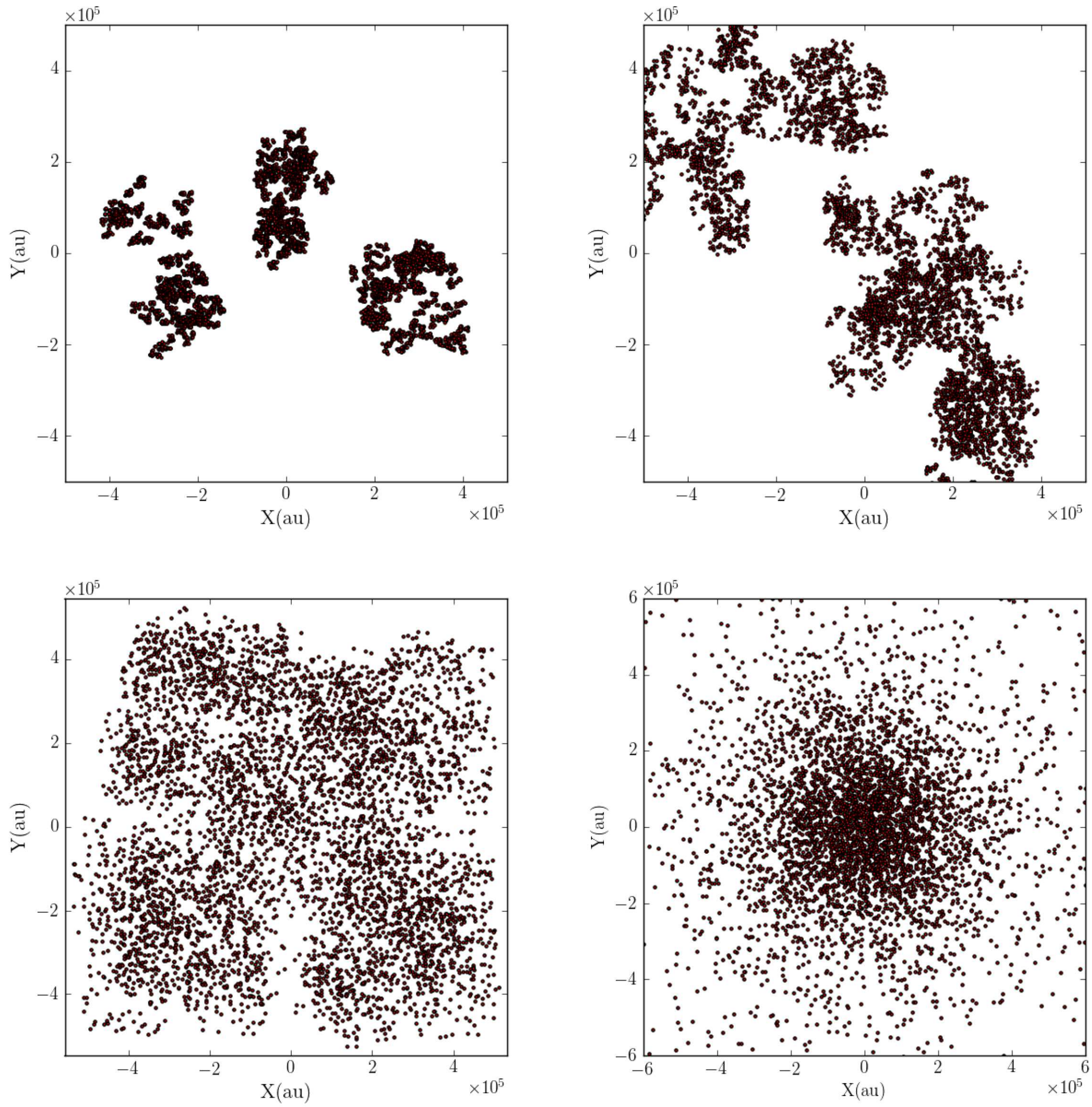

Figure 1. Projected density distributions of $N = 5000$ stars with no binaries. Top-left panel, a $D = 1.6$ fractal; top-right panel, a $D = 2.0$ fractal; bottom-left panel, a $D = 3.0$ fractal; bottom-right panel, a Plummer sphere. The X–Y size scale is ∼2 pc to mimic realistic regions, but this is not important (see text).

far from a Poisson distribution), and also Plummer (1911) spheres (created using the method of Aarseth, Hénon & Weilen 1974).

Fig. 1 illustrates one realization of each of our semi-realistic young regions,[1] each with 5000 stars. The top-left panel shows a fractal with fractal dimension $D = 1.6$, the top-right panel shows a $D = 2$ fractal, the bottom-left panel shows a $D = 3$ (roughly uniform density) fractal and the bottom-right panel shows a Plummer sphere. Each of these regions contains no binaries (just 5000 single stars).

Each distribution is created in three dimensions and projected into two dimensions (in each case looking along an arbitray z-axis). Each projection is 5×10^5 au on a side (roughly 2 pc, to mimic a realistic young region – although, as we shall see, the absolute scale is unimportant).

In most simulated young regions (all except the $D = 3$ roughly uniform distribution), the projected surface densities of stars vary significantly across the field of view (dense clumps in the $D = 1.6$ and $D = 2$ fractals, and a central concentration in the Plummer sphere). The core of our problem in searching for binaries is to determine in such complex distributions whether two objects are closer together than would be expected by chance, in particular when we have no a priori model of the underlying distribution (and where the presence of an unknown binary population might bias any attempt to quantify the underlying distribution).

Note that the slightly box-like appearance of the $D = 3$ fractal is an artefact of how it was generated (see Goodwin & Whitworth 2004). It is possible to 'jiggle' the distribution to remove these artefacts, but we have decided to keep them in order to make the distribution more complex.

In order to see if some stars are closer than would be expected, we need a model of the distribution as if there were no binaries. To do this, we use the ratio of the distance of the nearest neighbour, d_1, to the distance of the tenth nearest neighbour, d_{10}. We take the tenth nearest neighbour because this is enough neigbours distant to go beyond local multiple systems, but close enough to avoid

[1]It does not matter if these are particularly realistic models of real star-forming regions, only that they are non-trivial density distributions.

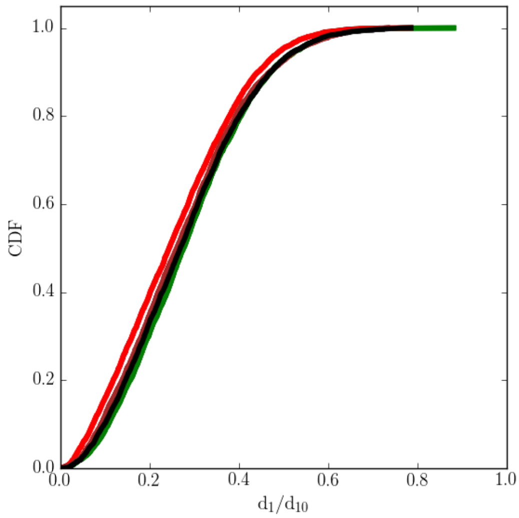

Figure 2. CDFs of the d_1/d_{10} ratios for the four $N = 5000$ density distributions in Fig. 1. Fractal dimension $D = 1.6$ is the red line (top-left panel in Fig. 1), $D = 2.0$ is the black line (top-left panel in Fig. 1), $D = 3.0$ is the brown line (bottom-left panel in Fig. 1) and the Plummer sphere is the green line (bottom-right panel in Fig. 1).

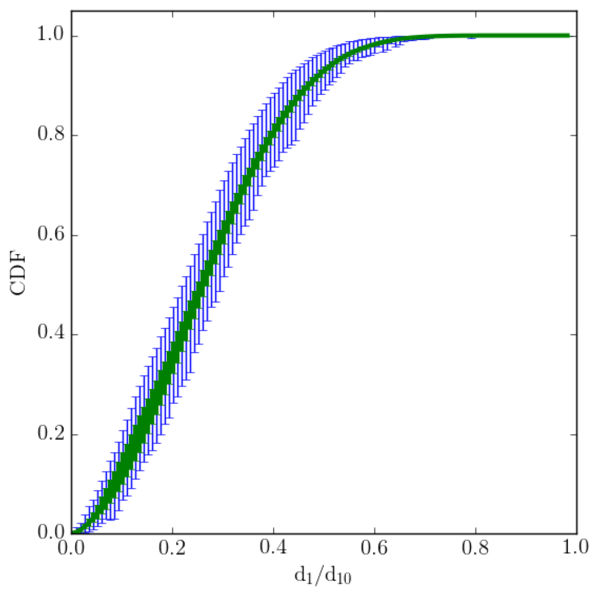

Figure 3. The spread of the CDFs of *the* d_1/d_{10} ratios for 1000 realizations (250 of each type). The green error bars show the spread of the central 50 per cent of distributions, and the blue error bars show the extremes of all 1000 realizations. All realizations can be fitted well with Gaussians with means 0.20–0.29 and standard deviations 0.16–0.19.

large-scale (surface) density variations/structure. It is also a large enough number to avoid too much 'noise' in the distance of the tenth nearest neighbour (i.e. if the distance to the tenth nearest neighbour is changed too much, then it becomes either the ninth or the eleventh nearest neighbour).

If we assume that the ten nearest neighbours are distributed randomly (i.e. there are no binary companions, and no density variations on the scale out to the tenth nearest neighbour), then we expect the nearest neighbour to be, on average, at a distance of $1/\sqrt{10} \sim 0.3d_{10}$. To test this, we perform Monte Carlo simulations of Poisson distributions. We find that the ratio of d_1/d_{10} is always close to a Gaussian with mean 0.20–0.29 and standard deviation 0.16–0.19 when $N = 500$ (when $N = 5000$ these ranges decrease to 0.22–0.26 and 0.17–0.18, respectively). The fact that the distribution of d_1/d_{10} would be close to a Gaussian is not obvious a priori, but appears always to be the case (at least for reasonably large N).

In Fig. 2, we plot the cumulative distribution functions (CDFs) of d_1/d_{10} for all four distributions plotted in Fig. 1. As we can see, despite all four projected distributions looking very different visually, the distributions of d_1/d_{10} are all very similar. All four distributions are close to a Gaussian (which, again, is not obvious a priori). As we can see in Fig. 2, the mean and median values of d_1/d_{10} are $\sim$0.3 for all distributions. Fitting a Gaussian to each CDF, the means and standard deviations of d_1/d_{10} are 0.22 ± 0.18 (red line, top left $D = 1.6$ fractal), 0.24 ± 0.17 (black line, top right $D = 2$ fractal), 0.25 ± 0.17 (brown line, bottom left $D = 3$ fractal), and 0.26 ± 0.16 (green line, bottom right Plummer sphere).

There are slight variations between the CDFs of d_1/d_{10} depending on the underlying distributions. The $D = 1.6$, very clumpy fractal (shown by the red line) has a slightly lower mean (0.22) than the rest as it is clumpy enough that the tenth nearest neighbour is sensitive to the large-scale fractal structure. All the standard deviations of the d_1/d_{10} ratios are also slightly larger than that for a Poisson distribution, because of the non-uniform nature of the distributions. However, we will see that these differences are negligible when compared with the effect of binaries.

In Fig. 3, we show the range of d_1/d_{10} distributions found in a large ensemble of 250 different realizations of each of our four different density distributions (1000 different realizations in total, each with $N = 5000$ single stars). The inner green error bars show the range of the middle 50 per cent of realizations, and the blue error bars show the maximum range of all 1000 realizations.

Across all realizations, the mean is always in the range 0.20–0.29, and the standard deviations are in the range 0.16–0.19 for the best-fitting (single) Gaussian.

To summarize, the distribution of the dimensionless ratio d_1/d_{10} is always very similar without binaries for a wide variety of different morphologies: roughly, a Gaussian with mean $\sim$0.20–0.29 and standard deviation $\sim$0.16$-$0.19.

2.1 Adding binaries

We now investigate the effect of adding a population of binaries to our distributions. We take regions with a total number of stars, either $N = 500$ or $N = 5000$. The total number of stars is fixed to within ±1, and the binary fraction sets what fraction of the total are either single stars or binaries (e.g. with $N = 500$ and a binary fraction of 50 per cent, there are 167 single stars and 166 binary systems for a total of $167 + (2 \times 166) = 499$ stars). Single stars and the centres of mass of binary systems are placed according to the desired density distribution (Plummer or fractal).

We include binary systems with an Opik law (log-flat) separation distribution between 1 and 2000 au in three dimensions, which we randomly project into two dimensions (assuming zero eccentricity).

2.2 Recovering binary fractions

The left column of Fig. 4 shows three Plummer spheres with $N = 5000$ stars and binary fractions of 0 per cent (top), 100 per cent (middle) and 50 per cent (bottom). Each panel is very similar visually. The density of points seems higher in the top panel (no binaries) and lower in the middle panel (all binaries), but this is only because we require that N is the same in both. With no prior

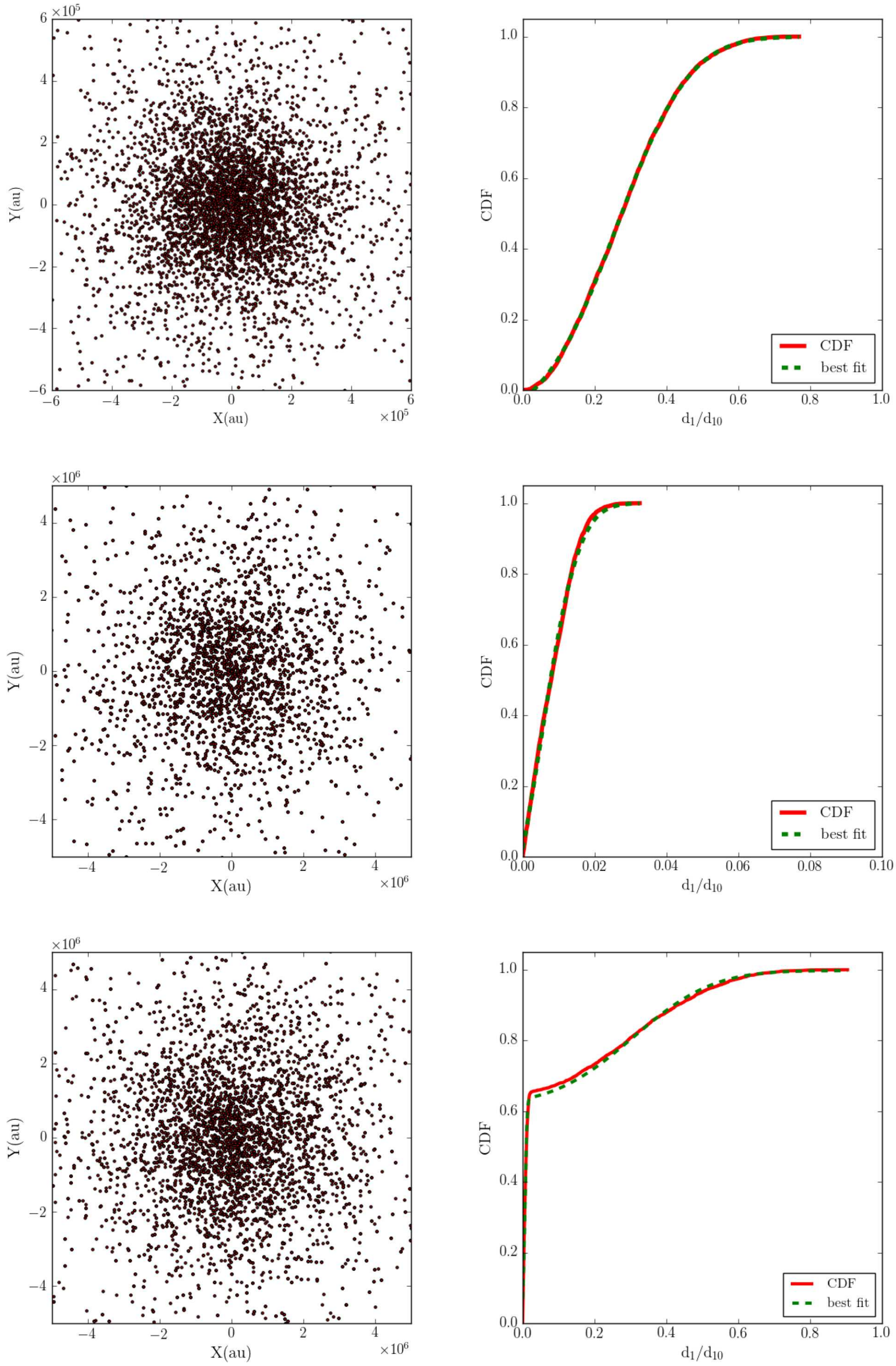

Figure 4. Projected views of stars in Plummer spheres with $N = 5000$ stars alongside the d_1/d_{10} CDFs and their best-fitting Gaussians: the top, middle and bottom rows show 0, 100 and 50 per cent binary fractions, respectively.

knowledge of N, it would be easy to think that the middle panel had far fewer stars in total. Looking closely, it is possible to convince oneself that the middle panel has more close pairs, but this is only really visible in the outer regions, because in the centre crowding confuses the eye.

The right column of Fig. 4 shows the CDFs of the d_1/d_{10} ratio, determined as above, for each of the distributions to the left.

The top row is for no binaries. The CDF of d_1/d_{10} is essentially the same as that of the Plummer sphere in Fig. 2 (with slight differences because it is a different realization). The dashed green line is a best-fitting Gaussian with mean 0.26 and variance 0.18, of the form expected for a distribution with no objects closer than expected by random chance.

The middle row is for all binaries. The CDF of d_1/d_{10} is again fitted by a single Gaussian, but the mean is only 0.005 (standard deviation 0.008), much lower than that expected for any distribution with no binaries. The most distant nearest neighbours only have a d_1/d_{10} of $\sim$0.03, which means that almost all stars have a

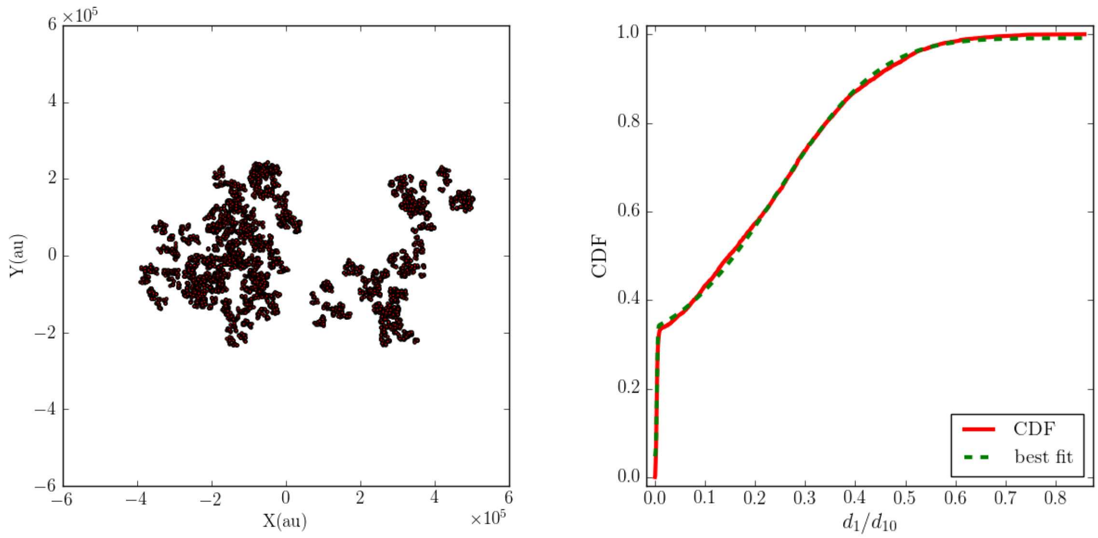

Figure 5. Left panel: a projected view of an $N = 5000$ $D = 1.6$ fractal with a 20 per cent binary fraction. Right panel: the CDF of d_1/d_{10} (red line) with a two Gaussian best fit (green dashed line).

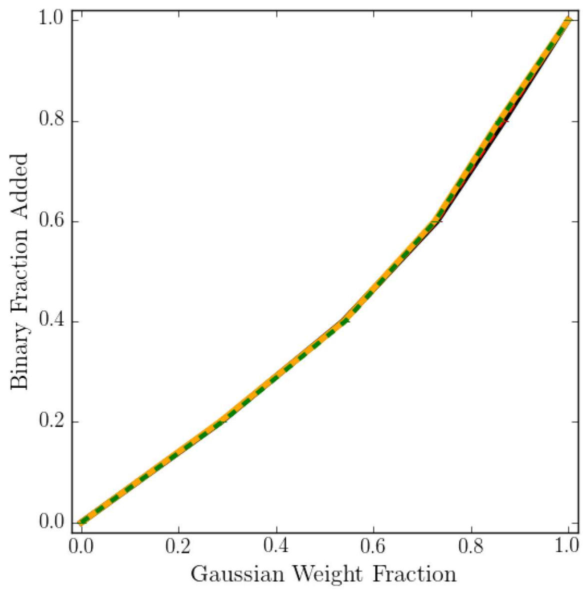

Figure 6. The binary fraction against the relative weight of the binary Gaussian fit. The solid lines are $N = 500$ star regions, and the dashed lines are $N = 5000$ star regions. Black lines are fractal distributions and red lines are Plummer spheres. All four lines lie almost on top of each other.

companion much closer than expected by random chance, whatever the underlying distribution.

The bottom row is for half binaries/half single stars. The CDF of d_1/d_{10} is now a much more complex shape and a single Gaussian is an extremely poor fit. There is a rapid rise from zero to ∼0.03 for around 60 per cent of pairings, then a gentler rise towards a ratio of unity for the last roughly 40 per cent of pairings.

Unsurprisingly, this form of the distribution is a mixture of the random CDF in the top panel and the CDF for all binaries in the middle panel (half of our pairings are true binaries and half are random). The best fit to the CDF is two Gaussians with relative weights 0.62 : 0.38, one Gaussian with mean 0.005 and standard deviation 0.005 (the binaries) and the second Gaussian with mean 0.27 and standard deviation 0.15 (the random pairings).

As another example, in the left panel of Fig. 5, we show a $D = 1.6$ fractal with $N = 5000$ stars and a 20 per cent binary fraction. Because of the underlying fractal nature of this distribution, it is not possible to see whether there are some stars closer together than might be expected. However, the right panel of Fig. 5 shows that the CDF of the d_1/d_{10} ratio shows a similar form to the bottom panel of Fig. 4: a sharp rise at low d_1/d_{10} (the binaries) and then a broader distribution out to large d_1/d_{10} (the random pairings).

The best fit to the CDF on the right of Fig. 5 is two Gaussians with weighting 0.29 : 0.71, one Gaussian with mean 0.004 and standard deviation 0.0015 (the binaries) and the second Gaussian with mean 0.25 and standard deviation 0.16 (the random pairings).

The parameters of the CDF with two Gaussians, required to fit the 50 per cent binary fraction Plummer sphere in the bottom panel of Fig. 4 and the 20 per cent binary fraction $D = 1.6$ fractal in Fig. 5, are very similar. The only significant difference is the relative weightings, which are directly related to the binary fractions in each case. Also, the binary component has a slightly higher weighting than the binary fraction: 0.62 for a 50 per cent binary fraction and 0.29 for a 20 per cent binary fraction. The difference between the Gaussian weightings and the underlying binary fractions is due to a small number of random pairings with a low-enough separation to be considered as possible binaries, which over-weights the 'binary' Gaussian.

To test the relationship between the weighting factor and true binary fractions, we create 250 realizations of each of region with $N = 500$ and 5000 stars, with binary fractions of 0, 20, 40, 60, 80 and 100 per cent, with density distributions that are Plummer spheres and $D = 1.6$, 2 and 3 fractals. In each case, we test whether a single or two Gaussian model is a better fit to the data.

If a single Gaussian is the best fit and if the mean and standard deviation are in the ranges 0.2–0.3 and 0.16–0.19, respectively, then we can be confident we are observing a random (no binary) population (or one in which binaries are so wide that they are indistinguishable from single stars). If the mean is much lower than 0.2, then we can be confident we are observing a 100 per cent binary population.

If the data are fitted best by two Gaussians, then we expect one of these Gaussians to have a mean of 0.2–0.3 and a standard deviation of 0.16–0.19 (the single population), and the other to have a significantly lower mean and standard deviation (the binary population).

In Fig. 6, we show the weighting of the binary Gaussian (i.e. the Gaussian that has a mean less than 0.20) against the binary fraction. As expected, the relative weight of this Gaussian increases from zero when the binary fraction is zero, to a weighting of unity when the binary fraction is unity.

For the reasons discussed above, the increase is not quite linear and the offset between the weighting and binary fraction is greatest when they are both roughly a half (when 'contamination' is worst).

The different colours and line styles in Fig. 6 represent different N and different density distributions (see caption for details). However, they are all very similar to each other, showing that the weighting of the binary does not depend on the underlying density distribution. As a fairly good rule of thumb, the Gaussian weighting is typically 0.05–0.10 higher than the true binary fraction for most intermediate binary fractions.

There are cases where it is possible to imagine an unclear or ambiguous result from this analysis. For example, if the mean of a best-fitting single Gaussian were around 0.2, it would be difficult to decide if this was due to wide binaries or the particular density distribution. It is also possible to imagine particular density distributions and/or binary separations that would be impossible to distinguish. One could also imagine a situation where three (or more) Gaussians were required to fit a particularly unusual binary separation distribution (especially if embedded in a complex density distribution).

3 CONCLUSION

We have developed a method to estimate the fraction of objects that are closer than expected by chance in an arbitrary distribution. We have developed this to search for binary stars (i.e. gravitationally bound pairs) in astrophysical data, but the method is applicable to many possible data sets.

We summarize the method, as follows.

(i) Find the distances to the nearest and tenth nearest neighbours of every object and construct a CDF of d_1/d_{10}.

(ii) Fit both single and double Gaussian models to the CDF.

(iii) If a single Gaussian is the best fit and the mean and standard deviation are in the range 0.20–0.30 and 0.16–0.19, respectively, then the data show no evidence of binaries.

(iv) If a single Gaussian is the best fit with a mean much lower than 0.2, then the data suggest a 100 per cent binary fraction.

(v) If a double Gaussian is the best fit, we expect one Gaussian to have the mean and standard deviation of the single population (roughly 0.20–0.30 and 0.16–0.19, respectively), and the other to model the binary population. The underlying binary fraction is related to the relative weightings of the Gaussians (where the binary Gaussian tends to be 0.05–0.10 higher than the true binary fraction).

It is worth noting that our example data sets are 'perfect' and we observe every object. Real data are subject to biases and selection effects and, in particular, observational data will have a resolution limit below which no two objects can be distinguished, which will mean that the d_1/d_{10} ratio is truncated at this limit. Obviously, the particular limitations of any data set should be included when analysing that data set.

In this paper, we only aim to find the binary fraction in a data set, and we do not go beyond this. It is possible to extend this analysis to find the probability that the nearest neighbour is a true binary, and to estimate the (two-dimensional) distribution of binary separations. However, we leave this for future work.

REFERENCES

Aarseth S. J., Hénon M., Weilen R., 1974, A&A, 37, 183
Bahcall J. N., Soneira R. M., 1981, ApJ, 246, 122
Cartwright A., Whitworth A. P., 2004, MNRAS, 348, 589
Dhital S., West A. A., Stassun K. G., Bochanski J. J., 2010, AJ, 139, 2566
Duchêne G., Kraus A., 2013, ARA&A, 51, 269
Goodwin S. P., 2010, Phil. Trans. R. Soc. A, 368, 851
Goodwin S. P., Whitworth A. P., 2004, A&A, 413, 929
Larson R. B., 1995, MNRAS, 272, 213
Michell J., 1767, Phil. Trans. R. Soc., 57, 234
Plummer H. C., 1911, MNRAS, 71, 460
Reipurth B., Zinnecker H., 1993, ApJ, 278, 81
Reipurth B. et al., 2014, in Beuther H. et al., eds, Protostars and Planets VI. Univ. of Arizona Press, Tucson, p. 267
Struve F. G. W., 1852, Stellarum Fixarum Imprimus Compositarum Positiones Mediae pro Epocha 1830.0. Academia Caesareae Petropolitanae, Petrograd

This paper has been typeset from a TEX/LATEX file prepared by the author.