Binary fraction estimation in open clusters with *Gaia*

Author: Judit Donada Oliu
Advisors: Friedrich Anders and Carme Jordi
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Abstract: In this study, we estimate the binary fraction of unresolved binaries with high mass ratios for 191 open clusters between 8.9 Myr and 3 Gyr old. Our first approach is based on assuming the colour-magnitude diagram of the cluster main-sequence members to be a mixture of two Gaussians, associated with single stars and unresolved binaries. We find that the binary fraction follows the reported general trend of increasing for higher masses, and it also behaves as expected with open cluster's distance; while we find no statistically significant correlation between the binary fraction and the open cluster's age.

I. INTRODUCTION

The characterization of the binary fraction in open clusters (OCs) is relevant to several branches of astrophysics. Correctly accounting for unresolved binary stars is essential in determining the total mass and stellar mass function of OCs ([1] and [2]). They also influence the OC's dynamical evolution, providing relevant information about the outcome of star forming processes in different environments and constraints on the initial OC's state ([3]). Furthermore, binaries are responsible for high rates of stellar collisions and black-hole mergers ([4]), thus being relevant for high-energy astrophysics and stellar population studies.

Stars tend to be born in pairs and higher-order hierarchies, and it has been historically asserted that most of them form in clusters, though according to recent studies the percentage of stars born in bound clusters might be significantly lower ([5] and [6]). In a first approximation, stars in OCs can be considered to lie at the same distance and to have the same age, initial chemical composition and reddening; which makes OCs highly suitable for studying their binary systems. Most of their binaries are unresolvable in images, but can be identified in the colour-magnitude diagram (CMD): depending on their mass ratio they appear as secondary sequences above the main sequence (MS), the equal mass binaries being 0.75 mag brighter than their MS counterparts.

OCs display significantly different binary fractions, which both at primordial and late times are not very well constrained. Since typically Galactic OCs are disrupted on a time scale of a few hundred Myrs ([7]), they tend to be relatively young. For young massive clusters in the local universe, the measured binary fractions are close to those of field stars ([4]).

In order to estimate the binary fraction of unresolved binaries, precise knowledge of OC's age, distance modulus and reddening is needed. This study uses the catalogue published by [8], which comprises 2017 OCs identified with *Gaia* Data Release 2 (*Gaia* DR2) astrometry, with a total number of members of ~230,000 (brighter than G=18 mag). OC's distances, ages and extinctions were accurately estimated combining parallaxes with *Gaia* DR2 photometry.

II. DATA PROCESSING AND ANALYSIS

The already developed Jupyter notebook distributed by Nigel Hambly and Mark Taylor (DPAC) called "NAM 2019: API access with python to Gaia DR2" ([9]) has been adapted and extended in order to obtain a code suited for handling the [8] catalogue data and studying the unresolved binary fraction determination of its OCs.

A. System selection for each cluster

First of all, absolute G magnitude (M_G) and intrinsic colour index ($(BP-RP)_0$) were calculated for each member of the [8] catalogue, excluding the ones with missing BP and/or RP photometry:

$$M_G = G - \mu - 0.89 \cdot A_V \quad (1)$$

$$(BP-RP)_0 = (BP-RP) - \frac{0.89}{1.85} \cdot A_V \quad (2)$$

where G is the G-band mean apparent magnitude, μ is the OC distance modulus in magnitudes, and A_V is the OC visual extinction (its applied factor is to obtain the extinction in the G-band).

With both magnitudes the Hertzsprung-Russell diagram (HRD) for the 2017 OCs was represented, and used to derive restrictions upon M_G and $(BP-RP)_0$ for a member of a particular OC to be considered in our study. The followed procedure is described below and it is illustrated for two particular OCs in Fig. 1.

As our binary fraction determination involves only MS stars, two straight parallel lines were drawn in the HRD which lay above and below the MSs of all clusters, and a first rough selection was made excluding red clump stars, white dwarfs, and some outliers by imposing:

$$2.9 \cdot (BP-RP)_0 - 1.4 < M_G < 2.9 \cdot (BP-RP)_0 + 3.4 \quad (3)$$

Another restriction was required in order to exclude members which are already evolving towards giant stars. This happens at a turn off point, dependant on OC's age (τ). To estimate its corresponding intrinsic colour index ($(BP-RP)_{0_{t-o}}$), first an estimate of the mass of a star at the turn off was calculated for each OC:

$$M_{t-o}[M_\odot] = \left[10^{(\log_{10}(\tau[yr])-10)}\right]^{-\frac{1}{2.5}} \quad (4)$$

We interpolated $(BP-RP)_0$ for any mass within the interpolation range using a solar metallicity young isochrone from PARSEC ([10]) and fitting an interpolation polynomial $f(M[M_\odot])$. So, for each OC, the turn off intrinsic colour index was obtained as $(BP-RP)_{0_{t-o}} = f(M_{t-o}[M_\odot])$, and the imposed restriction was:

$$(BP-RP)_0 \geq (BP-RP)_{0_{t-o}} \quad (5)$$

(i.e., members of $M < M_{t-o}$ were selected, meaning that their MS lifetime is $\tau_{MS} > \tau$).

Finally, the redder MS members of each OC were excluded to avoid binary fraction overestimation due to the

* Electronic address: jdonadol7@alumnes.ub.edu

limit of G=18 mag. It was done by fitting a polynomial $p_{MS}((BP-RP)_0)$ to the MS of its CMD, composed of the MS members selected for the moment. A third degree polynomial was fitted through the method of least squares, then the sigma clipping technique was applied to down-weight systems farther than 0.2 mag from the MS, and finally the $p_{MS}((BP-RP)_0)$ quartic polynomial was refitted.

Using the points $[(BP-RP)_0]_1 = [(BP-RP)_0]_{max} - 0.4$ and $[(BP-RP)_0]_2 = [(BP-RP)_0]_{max} - 0.2$, and their respective M_G values interpolated using $p_{MS}((BP-RP)_0)$, we estimated the linear slope of the lower MS (s) and used it to determine the quantity:

$$\Delta[(BP-RP)_0] = \frac{0.75}{s} \quad (6)$$

This was subtracted from the value $[(BP-RP)_0]_{max}$ of each OC in order to establish the highest intrinsic colour index that could have its selected members. Thus, the last requirement for a member to be considered was:

$$(BP-RP)_0 \leq [(BP-RP)_0]_{max} - \Delta[(BP-RP)_0] \quad (7)$$

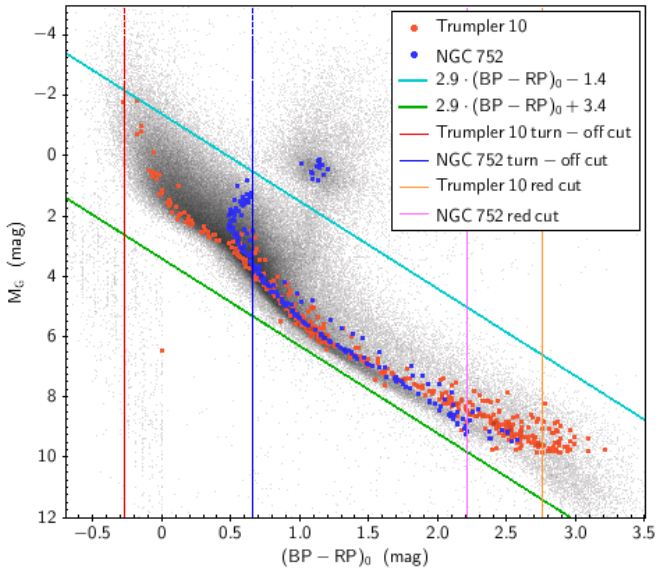

FIG. 1: Hertzsprung-Russell diagram of the 2017 OCs. Two examples of studied OCs are highlighted: Trumpler 10 ($\log(\tau) = 7.51$), and NGC 752 ($\log(\tau) = 9.07$). The blue and green inclined lines are cuts common to all OCs, while the vertical cuts which exclude the turn off and redder members are calculated particularly for each OC. All four lines form a specific parallelogram for each OC containing its selected MS members.

Having selected the MS members for each OC as the ones which verify conditions (3), (5) and (7), further requirements were imposed on studied OCs:

1. The selected MS members must be greater than or equal to 50.
2. The range of the intrinsic colour index must be greater than or equal to 1 mag.

As a result, the initial sample of 2017 OCs was reduced to a sample of 434 studied OCs.

B. Gaussian mixture

A polynomial $p_{MS^{selected}}((BP-RP)_0)$ was again fitted to each OC's CMD applying the sigma clipping technique as before, but the CMD being now composed of all the OC's selected MS members (see an example in Fig. 2). This polynomial was used to calculate the following quantity for each member:

$$\Delta G = M_G - p_{MS^{selected}}((BP-RP)_0) \quad (8)$$

Then, ΔG histograms were represented for each OC for the following samples:

(a) All selected MS members (see an example in Fig. 3).
(b) Selected MS members in four different intrinsic colour index ranges (or, equivalently, mass ranges; detailed in Table I) which have at least 10 members.

For all histograms it was assumed that all their data points were generated from a mixture of two Gaussian distributions with unknown parameters. We used the fit method of GaussianMixture option of sklearn.mixture package to find the best mixture of two Gaussians for the histogram data through an expectation-maximization approach. The parameters obtained for each Gaussian (mean μ, standard deviation σ and weight w) were written in an output file, as well as relevant OC parameters and the number of selected MS members (in the whole intrinsic colour index range for sample (a) and in the studied range for samples (b)).

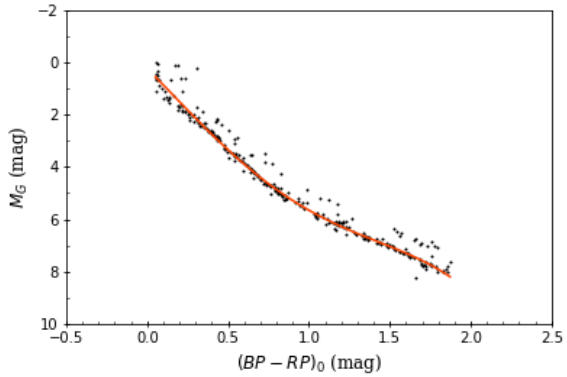

FIG. 2: Entire colour-magnitude diagram with the polynomial fitted to its MS (orange curve) for the OC NGC 2527, which has 303 selected MS members and $\log(\tau) = 8.84$.

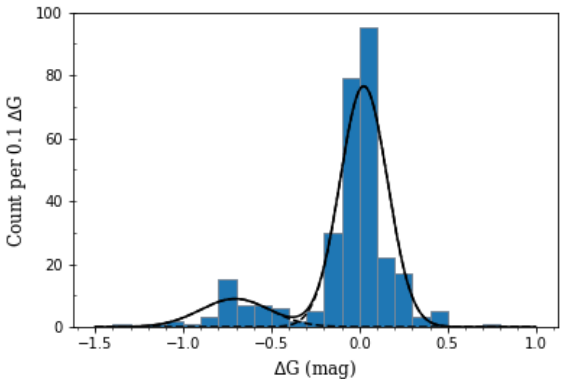

FIG. 3: ΔG histogram of all selected MS stars for NGC 2527. The MS fitted Gaussian has $\mu_{MS} = 0.025$ mag, $\sigma_{MS} = 0.135$ mag and $w_{MS} = 0.854$; while the binaries' one has $\mu_B = -0.706$ mag, $\sigma_B = 0.197$ mag, $w_B = 0.146$. Its estimated unresolved binary fraction is $f_{bin} \pm \delta[f_{bin}] = (14 \pm 2)\%$.

The highest weight Gaussian (w_{MS}) corresponds to the one fitted tracing the MS, while the one with the smallest weight (w_B) corresponds to the binaries; and they are normalized: $w_{MS} + w_B = 1$.

C. Clusters selection

The CMD with its fitted $p_{MS^{selected}}((BP-RP)_0)$ and the (a) ΔG histogram with its two fitted Gaussians were visually inspected for every OC. The OCs which had a CMD so sparse that a well-delineated MS could not be identified were excluded. OCs having a polynomial $p_{MS^{selected}}((BP-RP)_0)$ which poorly fitted the MS were also disregarded.

By means of this manual selection 243 OCs were discarded, remaining a final sample of 191 OCs for which the unresolved binary fraction was estimated.

D. Binary fraction estimation

The multiplicity fraction f_{bin} of a stellar population, often referred to as the binary fraction, is defined as:

$$f_{bin} = \frac{B+T+\dots}{S+B+T+\dots} \quad (9)$$

where S is the number of single stars, B the number of binary systems, T the number of triple systems, and so on.

The aim of this study is to estimate the unresolved binary fraction of OCs, which in this approximation equals the weight of the binaries' Gaussian: $f_{bin} = \frac{w_B}{w_{MS}+w_B} = w_B$.

It must be noted that the fitted Gaussian associated with the MS does not only include single stars and resolved components of binary systems, but also unresolved binary systems composed of stars of such different masses (low mass ratio) that the joint signal detected lies very close to the MS of single stars. Therefore, it is expected that our f_{bin} will be underestimated. Thus, rather than an estimation of the total unresolved binary fraction, our calculated f_{bin} is an estimation of the unresolved binary fraction of systems which have a mass ratio close to 1 (i.e., systems near the equal-mass binary sequence). Moreover, triples and higher-order systems cannot be distinguished from binaries by our method, so the estimated binary fraction includes their contribution.

For the estimation of the uncertainty of f_{bin}, we consider that the probability of any system out of the N selected MS members being a binary system follows a binomial distribution with a probability of success w_B. The Wald formula ([11]) is applied to estimate its uncertainty:

$$\delta[f_{bin}] = \delta[w_B] = z\sqrt{\frac{w_B(1-w_B)}{N}} \quad (10)$$

where $z = 0.99$ (it is given for a 68% confidence level).

This uncertainty is underestimated, as it only involves a statistical estimation of the error derived from the Gaussian fitted to the binaries. It does not take into account intrinsic photometry errors or uncertainties associated with the method applied to fit the Gaussians (involving the accuracy of the MS members selection and polynomial fitting).

For each OC, several values of $f_{bin} \pm \delta[f_{bin}]$ were calculated: one using w_B derived applying the Gaussian mixture model to all its N selected MS members (sample (a)), and the others using w_B derived applying the Gaussian mixture model to each of the mass ranges with a population $N \geq 10$ (samples (b)).

III. RESULTS

Analysis of the Gaussian mixture algorithm

In order to evaluate the behaviour of the Gaussian mixture algorithm from which f_{bin} is derived, Fig. 4 is represented.

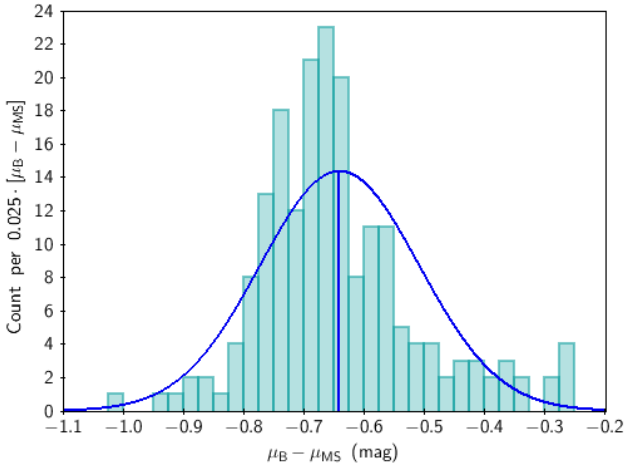

FIG. 4: Histogram of the separation between the mean of the binaries' Gaussian (μ_B) and the MS one (μ_{MS}), both Gaussians being fitted to the complete member sample (a) for each OC. The solid curve is a fitted Gaussian function, and the vertical line indicates its mean.

Both the most frequent value and the mean of the fitted Gaussian fall below the expected value $\mu_B - \mu_{MS} = -0.75$ mag. This can be explained by μ_{MS}'s tendency to fall slightly above the MS along with μ_B's tendency to fall below the equal-mass binary sequence.

μ_{MS} does not correspond to the MS lower envelope when there is a considerable proportion of low mass ratio binary systems (regarded in this method as simple systems) or when the polynomial deviates from the MS curvature at one of the two edges (or both). This happens in the example of Figs. 2 and 3: the MS Gaussian is almost centred at the polynomial, which does not always exactly trace the MS lower envelope and clearly deviates from it at the bluest end, where the estimation of the turn off cut has turned out to be inaccurate.

As for μ_B, the equal-mass binary sequence is not in general clearly delineated, what causes the binaries' Gaussian to tend to be centred at higher magnitudes. The example of NGC 2527 (Figs. 2 and 3) has one of the most well-delineated equal-mass binary sequence and even in this case we obtain $\mu_B - \mu_{MS} = -0.73$ mag.

The values $\mu_B - \mu_{MS} < -0.75$ mag could be explained due to selected members whose membership probability has been wrongly determined and whose contribution is relevant.

Dependence of binary fraction on mass

For the group of f_{bin} derived from the 191 histograms of sample (a) and the four groups of f_{bin} derived from the histograms of samples (b), we calculated their weighted mean as follows:

$$\overline{f_{bin}} \pm \delta\left[\overline{f_{bin}}\right] = \overline{f_{bin}} \pm \frac{\sigma_{f_{bin}}}{\sqrt{N}} = \\ = \frac{\sum_{i=1}^{N} \alpha_i \left(f_{bin}\right)_i}{\sum_{i=1}^{N} \alpha_i} \pm \frac{\sqrt{\frac{N}{\sum_{i=1}^{N} \alpha_i} \cdot \left[\sum_{i=1}^{N} \left\{\alpha_i \cdot \left[\left(f_{bin}\right)_i - \overline{f_{bin}}\right]^2\right\}\right]}}{\sqrt{N}} \quad (11)$$

being $\alpha_i = \frac{1}{\delta[f_{bin}]}$ the weight of each $\left(f_{bin}\right)_i$ value.

The results are shown in Table I.

Mass range ($M_\odot$)	N	$\overline{f_{bin}} \pm \delta[\overline{f_{bin}}]$ (%)	$\sigma_{f_{bin}}$ (%)
Entire	191	17.0 ± 0.5	7
[0.0, 0.5]	32	14.3 ± 0.9	5
[0.5, 1.5]	140	15.6 ± 0.6	7
[1.5, 3.0]	190	16.9 ± 0.5	7
> 3	3	20.9 ± 1.8	3

TABLE I: Mass range considered for the f_{bin} calculation for each OC, number of OCs for which f_{bin} can be calculated in this range (N), weighted mean of the unresolved binary fraction in this range and its dispersion (standard deviation: $\sigma_{f_{bin}}$).

The weighted mean of the unresolved binary fraction clearly shows a tendency to increase as mass increases, which is in agreement with published studies regardless of the mass ratio range of the systems they consider to be binaries. In [12] it is found that for young dense star clusters there is an increase in the binary frequency with primary mass. It is explained as a pure dynamical effect: the more massive a primary star, the lower the probability that the binary is destroyed by gravitational interactions. In [2] and [3] it is noted, however, that the fact that binary fraction increases at increasing primary mass might not be a universal feature of all OCs; and that it is required further study of large OC samples that cover broader ranges of mass, age, metallicity, and environment. But this general increase of non-single stars as a function of primary mass is actually considered a very solid result for Population I field stars in the solar neighbourhood ([13]). So, our result is in much agreement with these previous studies.

Dependence of binary fraction on OC's distance

Fig. 5 shows that the distant OCs have larger f_{bin} than the near ones, on average. This reflects that near multiple systems can be resolved, while farther ones appear as binaries. As the members of a close multiple system can be told apart, they are seen as individual points in the CMD, located at the MS (not contributing, therefore, to the unresolved binary fraction). But, as distance increases, multiple systems which were formerly resolved at shorter distances no longer are. Thus, the number of unresolved systems susceptible to contributing to f_{bin} increases, because the system appears as a single point in the CMD at a certain distance from the MS depending on its mass ratio.

Moreover, another effect can be seen: having a lower limit on the magnitude (in this case, G=18 mag) causes that for distant OCs only their more massive (and, thus, more luminous) members can be detected. As a result, their f_{bin} obtained with the complete member sample (a) tends to be dominated by the most massive stars, which have the larger f_{bin} (see Table I). Then, the trend in Fig. 5 can be partially explained by the bias of our sample introduced by the limiting magnitude cut.

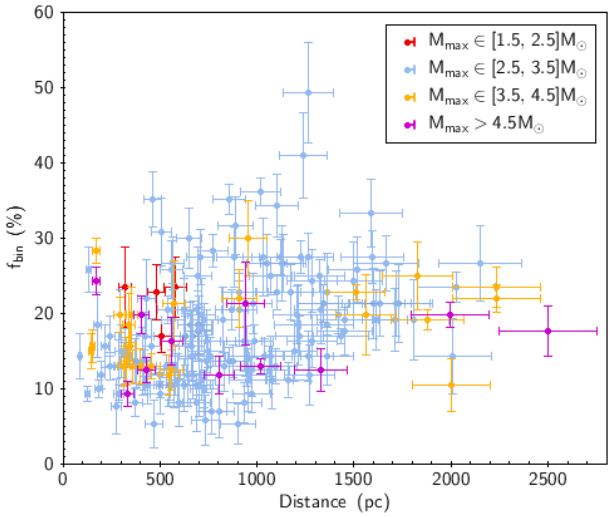

FIG. 5: The unresolved binary fractions of the 191 OCs, derived from their complete member sample (a), plotted against their distance, colour-coded by the mass range where the most massive system of each OC lies. f_{bin} error bars are calculated using (10).

Dependence of binary fraction on OC's age

The unresolved binary fraction derived from sample (a) is represented as a function of OC's age in Fig. 6 only for the OCs nearer than 1 kpc, thus avoiding to include the selection effect the distance causes for distant OCs.

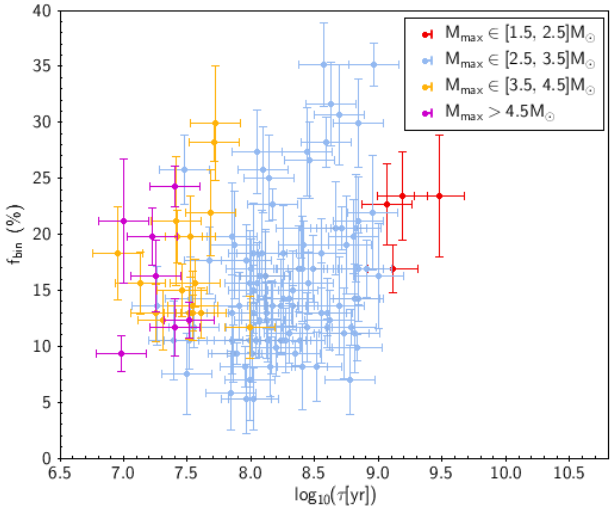

FIG. 6: The unresolved binary fractions for studied OCs nearer than 1 kpc, derived from their complete member sample (a), plotted against age; colour-coded and with the same $\delta[f_{bin}]$ as in Fig. 5.

According to the quite large sample in Fig. 6, there does not seem to be any clear correlation between f_{bin} and OC's age. On the one hand, since binaries are on average more massive than single stars, they trace the process of dynamical mass segregation: over time, the binaries migrate toward the cluster centre increasing the binary fraction in the OC's core ([14]). As a consequence, on average, older OCs tend to have larger f_{bin} than young OCs. On the other hand, it is clearly seen in Fig. 6 that young OCs have more massive members than older OCs. As the greater the mass is, the shorter is its MS lifetime, for old OCs the most massive stars have already left the MS. Thus, being dominated by less massive stars as

they age, according to Table I their f_{bin} diminishes over time. Therefore, the observed dependence in Fig. 6 might be the result of the blending of both opposite effects, as well as of the effect derived from the MS completeness depending on OC's distance.

IV. CONCLUSIONS

Through an approximate model of the mixture distribution of single stars and binaries in an OC's CMD, the unresolved binary fraction has been calculated for 191 OCs. From the study of the applied method and its estimated values, the following conclusions are drawn:

- Our estimated unresolved binary fraction f_{bin} refers to binary systems near the equal-mass binary sequence (of high mass ratio) because the fitted MS Gaussian always includes a portion of low mass ratio binary systems. Therefore, our values and their statistical error $\delta[f_{bin}]$ are underestimated with respect to those found in the literature referred to the entire binary population. This prevents a direct comparison between our values and the ones reported by other studies, besides having to calculate f_{bin} for the same mass range (which may be not sufficiently populated to apply our method). So, rather than comparing absolute values, data trends are studied.
- The number of OCs for which the f_{bin} has been estimated is higher than in any existing study in the literature, but the reliability of the method applied to calculate it varies greatly depending on each particular case. Its main drawbacks are:
  1. The ill delimitation of the OC's MS members, especially due to the vertical cuts in intrinsic colour index, that causes a poor fitting of the MS polynomial at its ends. This is the case at the turn off point in Fig. 2 and for Trumpler 10 in Fig. 1, as opposed to NGC 752 exemplary case in Fig. 1.
  2. The restricted reliable applicability of the Gaussian mixture model. The histogram on which is based depends on the quality of the fitted MS polynomial, that might be deficient in some cases. Moreover, the derived f_{bin} has a strong dependency on the OC's CMD peculiarities. The applied method works properly for CMDs with a well-delineated and enough populated equal-mass binary sequence (e.g. NGC 752 in Fig. 1). When it is not clearly distinguishable and/or there is a considerable proportion of low mass ratio binary systems, the means of the fitted Gaussians get closer and both Gaussians overlap significantly. As a result, f_{bin} tends to be overestimated because the binaries' Gaussian, centred closer to the MS, accounts for MS members which should not be included in it.
- Our estimated statistical error $\delta[f_{bin}]$ behaves coherently: it decreases as a greater number of members is involved in the f_{bin} calculation; and for a certain number of members, increases as f_{bin} raises.
- Even though a thorough analysis of the sample's completeness should be carried out, f_{bin}'s tendency to increase for higher masses is in agreement with its general reported behaviour ([2], [3] and [12]). f_{bin}'s dependence on OC's distance is also reasonable. No significant correlation can be inferred between f_{bin} and OC's age, which might be the result of both mass segregation and the effects of age and distance on the completeness of the studied MS.
- Further sophistications of our applied method could be accounting for differential extinction and the particular distance of every OC member, as well as considering the specific metallicity of each OC instead of a common isochrone. Otherwise, more refined methods than this first approach could be developed to select MS members and derive the unresolved binary fraction without presupposing a Gaussian mixture model, as done in [3] and [15].

Acknowledgments

I would like to express my gratitude to my advisors, Dr. Friedrich Anders and Dr. Carme Jordi, for their valuable guidance and dedication. I also wish to thank Dr. Eduard Masana for his collaboration.

---

[1] Borodina, O. I., Carraro, G., Seleznev, A. F., & Danilov, V. M. **2019**, ApJ, *908*, 60.

[2] Borodina, O. I., Seleznev, A. F., Carraro, G., & Danilov, V. M. **2019**, ApJ, *874*, 127.

[3] Li, L., Shao, Z., Li, Z.-Z., et al. **2020**, ApJ, *901*, 49.

[4] González, E., Kremer, K., Chatterjee, S., et al. **2021**, ApJ, *908*, L29.

[5] Ward, J. L., Kruijssen, J. M. D., & Rix, H.-W. **2020**, MNRAS, *495*, 663.

[6] Anders, F., Cantat-Gaudin, T., Quadrino-Lodoso, I., et al. **2021**, A&A, *645*, L2.

[7] Sheikhi, N., Hasheminia, M., Khalaj, P., et al. **2016**, MNRAS, *457*, 1028.

[8] Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. **2020**, A&A, *640*, A1.

[9] https://www.gaia.ac.uk/science/workshops/gaia-dr2-NAM2019-jul19

[10] http://stev.oapd.inaf.it/cmd

[11] Agresti, A., & Coull, B., «Approximate is better than 'exact' for interval estimation of binomial proportions», *The American Statistician*, *52*, 119-126, 1998.

[12] Kaczmarek, T., Olsczak, C., & Pfalzner, S. **2011**, A&A, *528*, A144.

[13] Fuhrmann, K., Chini, R., Kaderhandt, L., & Chen, Z. **2017**, ApJ, *836*, 139.

[14] De Grijs, R., Li, C., & Geller, A., «The dynamical importance of binary systems in young massive star clusters», *Proceedings of the International Astronomical Union*, 12(S316), 222-227, 2015.

[15] Niu, H., Wang, J., & Fu, J. **2020**, ApJ, *903*, 93.