The frequency and mass-ratio distribution of binaries in clusters – I. Description of the method and application to M67

Michael D. Albrow ⓘ★ and Isaac H. Ulusele

School of Physical and Chemical Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

ABSTRACT

We present a new method for probabilistic generative modelling of stellar colour–magnitude diagrams (CMDs) to infer the frequency of binary stars and their mass-ratio distribution. The method invokes a mixture model to account for overlapping populations of single stars, binaries, and outliers in the CMD. We apply the model to *Gaia* observations of the old open cluster, M67, and find a frequency $f_B(q > 0.5) = 0.258 \pm 0.019$ for binary stars with mass ratio greater than 0.5. The form of the mass-ratio distribution function rises towards higher mass ratios for $q > 0.3$.

Key words: methods: data analysis – methods: statistical – open clusters and associations: general – open clusters and associations: M67.

1 INTRODUCTION

Studies have shown that perhaps 40–90 per cent of field stars in the Milky Way are born in clusters or associations (Bressert et al. 2010; Ward, Kruijssen & Rix 2020). Binary stars play an important role in the evolution of such systems. Through three-body or four-body gravitational interactions, binaries can enhance the dispersal of a star cluster through ejection of stars. In dense clusters, binaries provide an energy reservoir that prevents runaway core collapse through the conversion of binding energy to kinetic energy in three-body encounters. Binary stars in a cluster shape its structure and evolution, which in turn shapes the structure and evolution of the binary systems (Hut et al. 1992; Hurley et al. 2005).

Counts of binary stars in clusters have been done by various authors using either radial velocities, photometric colour–magnitude diagrams (CMDs), or photometric time series analysis.

CMD photometric methods generally apply some criterion to separate single from binary stars, count the binary stars in particular CMD locations (sometimes comparing this to models), and apply some mass-ratio function to compute the total binary frequency. For some examples of this general approach, see Milone et al. (2012) and Ji & Bregman (2013) for globular clusters, Elson et al. (1998) and Li, de Grijs & Deng (2013) for young massive Large Magellanic Cloud (LMC) clusters, and Sollima et al. (2010) for Galactic open clusters.

Radial velocities can be used to detect binary stars through their orbital motion, and can be used as a cluster membership criterion. The detected radial velocity binary population must be corrected statistically for declining detection efficiency for low orbital inclination to derive the cluster binary frequency. Radial velocity studies, e.g. the WIYN Open Cluster Study (Mathieu 2000), have generally been limited to stars near or above the top of the main sequence.

Time series analysis to detect variable stars, including binaries, has been used by Albrow et al. (2001) for the globular cluster 47

Tucanae. Inference of a binary frequency is, however, fairly indirect, relying on assumptions such as the contact binary lifetime.

The mass-ratio distribution of field binary stars has been determined from radial velocity surveys by several authors. Duquennoy & Mayor (1991) found a distribution that peaks around $q = 0.25$ and declines towards higher mass ratios. In contrast, Fisher, Schröder & Smith (2005) infer a flat distribution with a sharp upturn towards higher mass ratios. For globular clusters, Milone et al. (2012) find flat or gently rising distributions with q (above $q = 0.5$). The model that we will construct below allows these possibilities. As noted by Milone et al. (2012), all the measured mass-ratio distributions for field stars and clusters are fundamentally at odds with random draws of secondaries from the initial mass function.

Overall, it has been found that the frequency of binary stars in open clusters can range from 25–70 per cent. For globular clusters, the binary frequency is much lower, certainly less than 25 per cent, and possibly less than 5 per cent. Prior to this paper, all estimates rely on assumptions about the form of the mass ratio (q) distribution at low q. Estimates are also complicated further by the well-established phenomenon that binaries are more centrally concentrated than single stars.

M67 (NGC 2682) is a rich old Milky Way open cluster located at coordinates RA = 08:51:23, Dec. = +11:49:02 (J2000) and a distance of 860 pc. It has been extensively studied photometrically (e.g Yadav et al. 2008; Sarajedini, Dotter & Kirkpatrick 2009; Gao 2018), and has been the target or several spectroscopic surveys (Pasquini et al. 2012; Geller et al. 2021). The binary star frequency has been estimated as >38 per cent (Montgomery, Marschall & Janes 1993), >26 per cent (Gao 2018), >45 per cent (Davenport & Sandquist 2010), ~50 per cent (Fan et al. 1996), 34 ± 3 per cent (Geller et al. 2021).

The layout of the remainder of this paper is as follows. In Section 2, we discus *Gaia* observations of M67, and our filtering to achieve a clean CMD. In Section 3, we develop a parametrized generative model for the CMD, which we solve in Section 4.

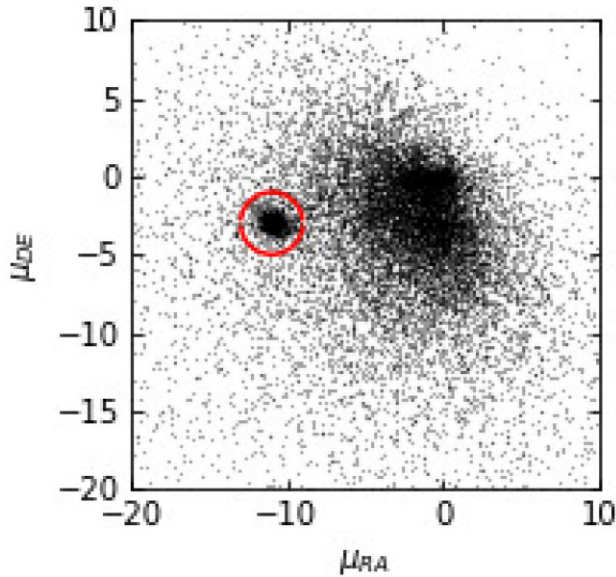★ E-mail: michael.albrow@canterbury.ac.nz

Figure 1. Proper motion of stars in a 1 deg radius cone centred on M67. The red circle encloses our initial selection of cluster stars.

2 *GAIA* PHOTOMETRY

2.1 Selection

Standard *Gaia* EDR3 photometry is available in three photometric passbands, G_{BP} and G_{RP} integrated from low-resolution spectra, and G measured from the astrometric field CCDs (Evans et al. 2018; Riello et al. 2021). Hereafter, we refer to these as *B*, *R*, and *G*. We first made the flux-excess corrections to the *B* and *R* magnitudes as recommended by Riello et al. (2021). For our data, these corrections were almost negligible. There is a known problem with overestimated *Gaia B* flux for faint red stars (see section 8.1 in Riello et al. 2021) so for our analysis we use CMDs in $(G − R, G)$.

Initially, we selected all *Gaia* EDR3 data from a cone of radius 1 deg centred on M67 (NGC 2682). These data included the cluster as well as many foreground and background stars. We next made a coarse cut on proper motion, selecting those stars that had a proper motion (μ_α, μ_δ) that lay within a circle of radius 2 mas yr^{-1} centred on $(−11, −3)$ mas yr^{-1} (Fig. 1). We next filtered out stars with a parallax uncertainty greater than 20 per cent and fitted a 5D Gaussian in Galactic latitude (*l*), Galactic longitude (*b*), μ_α, μ_δ, and parallax (ϖ) to the remainder. By trial and error, we imposed a cut on this 5D distribution at a level for which we were satisfied with a moderately clean CMD, Fig. 2. The parameter boundaries for the adopted cuts are listed in Table 1. This process has resulted in a very clean CMD, with less contamination than that of Yadav et al. (2008), albeit not as deep as that study. It appears similar in quality to the M67 *Gaia* CMD derived by Gao (2018) using machine learning methods.

The final CMD shows a main sequence with an obvious population of binary stars distributed up to 0.75 mag above the main-sequence ridge line, plus a few outliers that may be triple systems, or evolved binaries. Binary systems with two equal-mass main-sequence stars appear 0.75 mag above the main sequence. Binaries with lower mass ratios appear redder and fainter than the equal-mass case, see for instance the tracks in fig. 3 of Elson et al. (1998).

There is some evidence for mass segregation in M67. Fan et al. (1996) found a half-mass radius of 7 arcmin for blue stragglers, 10 arcmin for upper main-sequence stars, and 12 arcmin for the
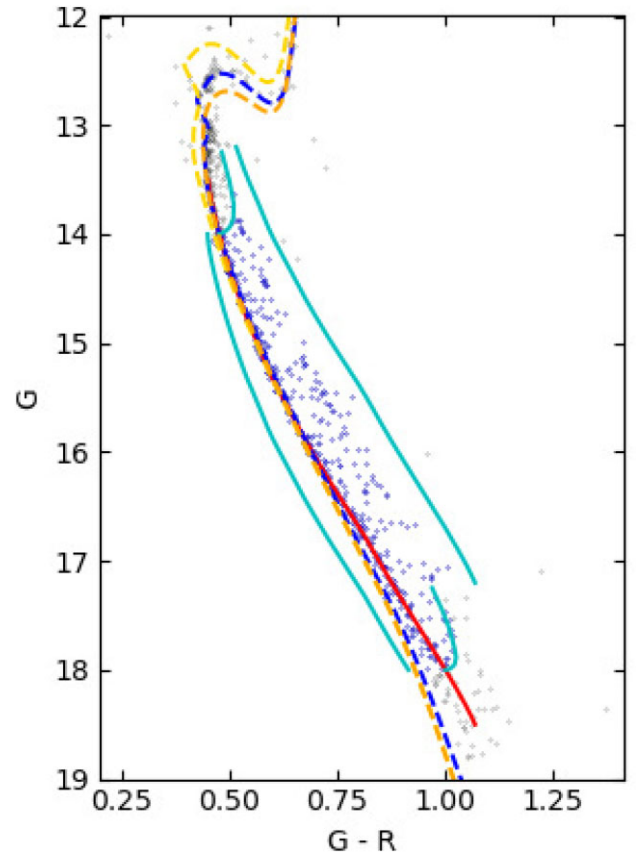

Figure 2. CMD for stars passing the filters as described in Section 2.1. The dashed blue line is the MESA isochrone for an age of 5 Gyr. The red line shows the colour-corrected isochrone. Gold and orange dashed lines show isochrones for 4 and 6 Gyr, respectively. Cyan curves show the lower and upper cuts applied so that binary-star primary masses are drawn from the same population as single stars, and some broader cuts to reject obvious outliers.

Table 1. 5D selection cuts applied to the *Gaia* data.

	Minimum	Maximum
l	215.18	216.20
b	31.48	32.37
ϖ (mas)	1.006	1.314
μ_α (mas yr^{-1})	−11.77	−10.06
μ_δ (mas yr^{-1})	−3.98	−2.04

lower main sequence. This is confirmed from radial velocity studies (Mathieu & Latham 1986; Geller et al. 2021). We might therefore expect main-sequence binaries to be more centrally concentrated than single stars. Our analysis covers stars drawn from a radius of ∼30 arcmin, so effectively averages over any mass segregation of binaries. The tidal radius of M67 is estimated to be 42 arcmin (Fan et al. 1996; Francic 1989), and its core radius is 8.24 arcmin (Davenport & Sandquist 2010).

For the analysis presented below, we restricted ourselves to considering stars that are clearly on the main sequence or are binaries consisting of two main-sequence stars. We have manually removed the few stars that are clearly above the equal-mass-binary main sequence. Additionally, we have retained only stars where the primary (or only) component lies within $13.5 < G < 18$. These cuts are shown as magenta curves in Fig. 2, and were found necessary in

order that our binary star distribution can be modelled below with the same primary-star mass function as the single stars.

2.2 Covariant uncertainties

Each 2D data point on the CMD, $(G - R, G)$ is formed from G and R measurements with independent uncertainties, but the $G - R$ combination is not independent of G. Consequently, a $(G - R, G)$ datum has an associated covariance matrix

$$\mathsf{S} = \begin{bmatrix} \sigma_{G-R}^2 & \sigma_{G-R,G} \\ \sigma_{G-R,G} & \sigma_G^2 \end{bmatrix}, \tag{1}$$

where $\sigma_{G-R}^2 = \sigma_G^2 + \sigma_R^2$ and it can be shown (see Appendix A) that the off-diagonal elements, $\sigma_{G-R,G} = \sigma_G^2$. As viewed in the $(G - R, G)$ plane, each data point can be considered as having a tilted bivariate Gaussian probability distribution.

2.3 Isochrone fitting

To model the stellar population of the cluster, we rely on having an isochrone that accurately describes the main-sequence magnitude and colour as a function of mass. We used isochrones from the MESA Isochrones and Stellar Tracks project[1] (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016). After some trial and error, we adopted a 5 Gyr isochrone with [Fe/H] = +0.06 as being the best fit to the upper main sequence and turnoff. This is consistent with the metallicty determinations of Hobbs & Thorburn (1991) ([Fe/H] = −0.04 ± 0.12) and Önehag, Gustafsson & Korn (2014) ([Fe/H] = 0.06). The age, however, is long compared with most recent determinations, which are generally close to 4 Gyr. The age inferred for M67, however, depends on the adopted theoretical isochrone calculations (see fig. 10 of Yadav et al. 2008), in particular the treatment of convective overshooting. In Fig. 2, we have shown the M67 *Gaia* CMD with MESA isochrones for 4, 5, and 6 Gyr and [Fe/H] = +0.06. We based our fit on the main-sequence turnoff region, and found that only the 5 Gyr isochrone provided a correct width for the subgiant branch and matched the 'hook' at the top of the main sequence. The specific choice of isochrone is inconsequential for the purpose of this paper since we will eventually marginalize over all parameters associated with the mass distribution.

The MESA isochrones deviate from the observed main-sequence ridge line near the bottom of the main sequence. We introduce a colour correction to the isochrone below $G = 16.5$, implicitly adopting the G versus mass relation from the isochrone.

2.4 Limits on binary star detection

Depending on the quality of the photometry and the chosen passbands, a given CMD will have a threshold in q below which binaries are practically indistinguishable from single stars. In Fig. 3, we show sections of our final trimmed CMD data, along with the colour-corrected isochrone for the main sequence and tracks representing binary stars with a fixed primary mass and different values of q. We indicate the locations of binary stars with $q = (0.4, 0.5, 0.6)$. For our data, $q = 0.5$ is a sensible threshold to adopt, as we perceive that such binary stars are sufficiently displaced from the main sequence to be recognized as such. The model we will develop in the following section does not depend on this threshold, but it is important for the interpretation of the model outputs.
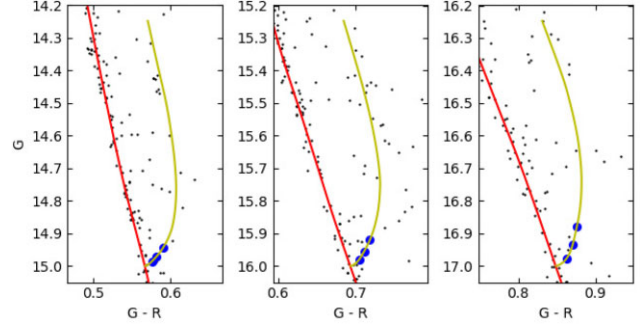
Figure 3. Cutout sections of the final CMD near different G magnitudes. The red lines are the adopted isochrone with colour correction applied. The yellow lines are tracks for binary stars with the same primary mass and varying mass ratio. The blue dots mark the locations $q = (0.4, 0.5, 0.6)$.

3 MODEL

In this section, we develop a probabilistic generative model for the main-sequence region of a CMD. Our aim is to develop a probability for any location on the CMD as a function of a set of numerical parameters, and then to sample that probability distribution for the observed data in order to learn the probability distributions for the parameters. Our method is strongly influenced by the mixture modelling approach of Hogg, Bovy & Lang (2010). We refer the reader to Taylor et al. (2015) for a comprehensive example of constructing a generative probabilistic model, in that case for the bimodal colour–colour distribution for field galaxies from the Galaxy And Mass Assembly (GAMA) survey.

Our *Gaia* data are the set $\boldsymbol{D} = (\boldsymbol{D}_k)$ for stars k, with each $\boldsymbol{D}_k = (G_k - R_k, G_k)^T$. Assuming that our model for the data can be described by a vector of parameters $\boldsymbol{\theta}$ (to be described), then from Bayes theorem the probability distribution for $\boldsymbol{\theta}$,

$$P(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{P(\boldsymbol{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{Z}, \tag{2}$$

where $P(\boldsymbol{D}|\boldsymbol{\theta})$ is the likelihood function, $P(\boldsymbol{\theta})$ is the prior for $\boldsymbol{\theta}$ and Z is the marginal likelihood, a constant for a given $\boldsymbol{D}$ and model.

We recognize that the majority of CMD data points are either single stars or binaries, but that there may be some 'bad' data points that would not be represented by a combination of single and binary star distributions in the CMD. We adopt a generative mixture model (Hogg et al. 2010), where each star has a probability f_B of being a binary, f_O of being an outlier, and $(1 - f_B - f_O)$ of being a single star. This approach is a natural treatment for the overlapping populations close to the main-sequence ridge line, and explicitly does not require the labelling of an individual star as being a member of a specific population. By design, our data sample contains few outliers, but we retain the f_O term for generality of the model description.

The likelihood function,

$$P(\boldsymbol{D}_k|\boldsymbol{\theta}) = (1 - f_B - f_O)P_S(\boldsymbol{D}_k|\boldsymbol{\theta}) + f_B P_B(\boldsymbol{D}_k|\boldsymbol{\theta}) \\ + f_O P_O(\boldsymbol{D}_k|\boldsymbol{\theta}), \tag{3}$$

where $P_S(\boldsymbol{D}|\boldsymbol{\theta})$, $P_B(\boldsymbol{D}|\boldsymbol{\theta})$, $P_O(\boldsymbol{D}|\boldsymbol{\theta})$ are respectively the likelihood functions for single stars, binaries, and outliers. We describe each of these in turn below. The total likelihood is the product of the individual star likelihoods so that

$$\ln P(\boldsymbol{D}|\boldsymbol{\theta}) = \sum_k \ln P(\boldsymbol{D}_k|\boldsymbol{\theta}). \tag{4}$$

3.1 Error bar rescaling

It is common in observational studies for measurement uncertainties to be underestimated (or overestimated). In practice, this is equivalent to assuming that the underlying true distribution has an intrinsic width. Such a width could arise from effects such as differential reddening or small systematic errors in the photometry. We allow for the possibility of an intrinsic width or incorrect error bars in our model by adopting a free parameter, h, that scales all data error bars. In general, this manifests as $\mathbf{S}_{G,G-R} \rightarrow h^2 \mathbf{S}_{G,G-R}$ in the probability calculations that follow. From examination of Fig. 2, it appears that the data may become more noisy towards the lower main sequence, so we allow h to vary linearly along the main sequence as $h = h_0 + h_1(G - G_0)$, where we adopt $G_0 = 16.0$. Error bar scaling thus introduces two parameters to our model.

3.2 Outliers

We begin with the distribution of outliers, since that is the simplest of the three individual likelihoods. In what follows, we adopt the notation

$$\mathcal{N}(\boldsymbol{x}, \mathbf{S}) \equiv \frac{1}{2\pi \sqrt{\det(\mathbf{S})}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \mathbf{S}^{-1}\boldsymbol{x}\right) \qquad (5)$$

for a normalized bivariate Gaussian centred at (0,0) with covariance matrix $\mathbf{S}$, evaluated at x. We also note that a probability density in x should strictly be denoted by the derivative, dP/dx. For brevity, we adopt the looser notation, $P(x)$.

From the standard rules of probability, we can expand the outlier likelihood function

$$P_{\mathrm{O}}(\boldsymbol{D}_k|\boldsymbol{\theta}) = \int P(\boldsymbol{D}_k|\boldsymbol{D}')P_{\mathrm{O}}(\boldsymbol{D}'|\boldsymbol{\theta})\,\mathrm{d}^2\boldsymbol{D}'. \qquad (6)$$

The first term in the integral is the (Gaussian) probability of the given data point $\boldsymbol{D}_k$ given a 'true' CMD location $\boldsymbol{D}'$. The second term is the probability distribution on the CMD for outliers, for which we adopt a very broad bivariate Gaussian. Consequently, the integral becomes a convolution

$$P_{\mathrm{O}}(\boldsymbol{D}_k|\boldsymbol{\theta}) = \int \mathcal{N}(\boldsymbol{D}_k - \boldsymbol{D}', h^2 \mathbf{S}_k)\mathcal{N}(\boldsymbol{D}' - \boldsymbol{D}_O, \mathbf{S}_O)\,\mathrm{d}^2\boldsymbol{D}' \qquad (7)$$

$$= \mathcal{N}(\boldsymbol{D}' - \boldsymbol{D}_k, h^2 \mathbf{S}_k) * \mathcal{N}(\boldsymbol{D}' - \boldsymbol{D}_O, \mathbf{S}_O) \qquad (8)$$

$$= \mathcal{N}(\boldsymbol{D}_k - \boldsymbol{D}_O, h^2 \mathbf{S}_k + \mathbf{S}_O), \qquad (9)$$

and the likelihood is thus a bivariate Gaussian. In principle, $\boldsymbol{D}_O$ and $\mathbf{S}_O$ could be regarded as elements of $\boldsymbol{\theta}$, but in practice it is sufficient to set them as constants. Here, we adopt $\boldsymbol{D}_O = (0.75, 16)$ and $\mathbf{S}_O = \mathrm{diag}[(0.75)^2, (4.0)^2]$. Generally, $\mathbf{S}_k$ is negligible compared to $\mathbf{S}_O$ and can be removed from the final covariance with no effect.

The above description is certainly an imprecise representation of the outlier distribution but, as noted by Hogg et al. (2010), much of the power of a mixture model such as this one comes from simply including an outlier distribution, even if it is inaccurate in detail.

3.3 Single stars

We assume that the true locations of single stars in the CMD lie along the isochrone line, and are drawn from some mass distribution. This mass distribution is truncated at the top of the main sequence, and fades away more gradually at the bottom of the main sequence due to declining *Gaia* detection efficiency with increasing G magnitude.

We model this observational mass distribution as

$$P(M|\gamma, k, M_0, M_{\mathrm{max}}) = C_M M^{-\gamma} \times \frac{1}{1 + e^{-k(M-M_0)}}$$
$$\times H(M_{\mathrm{max}} - M). \qquad (10)$$

Here, the first term is a power law in M. The second term is a logistic function that acts as a smoothed step, centred at M_0 and with characteristic width $1/k$, to control the bottom of the mass distribution. The top of the distribution is subject to a sharp cutoff by the third term, a Heaviside step function at mass $M_{\mathrm{max}} = 1.186\,\mathrm{M_\odot}$, corresponding to G = 13.5. The normalization parameter C_M is computed numerically so that $\int_{-\infty}^{\infty} P(M|\gamma, k, M_0, M_{\mathrm{max}})\mathrm{d}M = 1$. For our M67 data, we finally opted to apply a hard cut near the bottom of the main sequence, so could have replaced the logistic function with a second step function, however we retain this form for generality of the model.

Since there is a one-to-one mapping, $D'(M)$, from a given M to a location on the CMD the single-star likelihood function can be expanded as

$$P_{\mathrm{S}}(\boldsymbol{D}_k|\boldsymbol{\theta}) = \int P(\boldsymbol{D}_k|\boldsymbol{D}')P(\boldsymbol{D}'|M)P(M|\boldsymbol{\theta})\,\mathrm{d}^2\boldsymbol{D}'\,\mathrm{d}M \qquad (11)$$

$$= \int P(\boldsymbol{D}_k|\boldsymbol{D}'(M))P(M|\boldsymbol{\theta})\,\mathrm{d}M \qquad (12)$$

$$= \int \mathcal{N}(\boldsymbol{D}_k - \boldsymbol{D}'(M), h^2 \mathbf{S}_k)P(M|\boldsymbol{\theta})\,\mathrm{d}M. \qquad (13)$$

We can perform this integration by breaking the mass distribution function into a sum of discrete mass steps of width ΔM and height $P(M_l|\boldsymbol{\theta})$ located at masses M_i,

$$P_{\mathrm{S}}(\boldsymbol{D}_k|\boldsymbol{\theta}) = \sum_l \mathcal{N}(\boldsymbol{D}_k - \boldsymbol{D}'(M_l), h^2 \mathbf{S}_k)P(M_l|\boldsymbol{\theta})\Delta M. \qquad (14)$$

In practice, results from using this expression are equivalent to employing the expression we will develop in the following section for binary stars, with a mass ratio set to be close to zero.

3.4 Binary stars

We assume that the more massive star, M_1, in each binary system is drawn from the same distribution as the single-star mass function, equation (10). Each secondary star has a mass $M_2 = qM_1$, where q is drawn from a parametrized probability density. After testing various forms for this distribution, we adopted a general cubic polynomial,

$$P(q|a_k, \dot{a}_k, M; k = 1..3) = \tilde{P}_0(q) +$$
$$\Sigma_{k=1}^3 (a_k + \dot{a}_k(M - M_{\mathrm{ref}}))\tilde{P}_k(q), \qquad (15)$$

where $\tilde{P}_k(q)$ are the shifted Legendre polynomials, an orthogonal basis set over the range $0 < q < 1$. This general form allows the distribution to bend up or down at either or both ends. The form of the distribution as written requires no normalization since

$$\int_0^1 \tilde{P}_k(x)\mathrm{d}x = \begin{cases} 1, & k = 0 \\ 0, & k > 0. \end{cases} \qquad (16)$$

By including the derivative parameters $\dot{a}_k \equiv \partial a/\partial M$, the shape of the distribution is allowed to vary with mass, around some reference value M_{ref}, which we choose as the centre of the mass range covered by the CMD main sequence.

We consider this general model for the distribution as described, and also the more restricted versions with $\dot{a}_k$ set to zero.

To compute the likelihood, we represent the probability densities for M_1 and q as linear combinations of 50 predefined Gaussian basis functions,

$$P(M_1) = \sum_i a_i \mathcal{N}(M_1 - M_{1,i}, \sigma_M) \tag{17}$$

$$P(q) = \sum_j b_j \mathcal{N}(q - q_j, \sigma_q), \tag{18}$$

where the centroids, $M_{1,i}$ and q_j are equally spaced across the mass and mass-fraction ranges, $0.1 < = M_1/M_\odot < = 1.1$ and $0 < q < = 1$. The Gaussian widths are set to $\sigma_{M_1} = 0.01$ and $\sigma_q = 0.01$, which we found by trial-and-error to produce a good representation of the underlying functions. The choice of Gaussians for these basis functions will become apparent below.

It is a requirement of the formalism that follows that all a_i and b_j are zero or positive. The coefficients a_i and b_j can be computed analytically as linear fits to numerical representations of equations (10) and (15). However, the optimal linear fit coefficients are not guaranteed to be positive. Instead, we use the non-negative least squares algorithm from Lawson & Hanson (1974), which iterates to a solution.

A model magnitude for a binary star is obtained by adding the fluxes of the two components in the appropriate bandpass. This is done by interpolating magnitudes for masses M_1 and qM_1 from the isochrones, converting magnitudes to fluxes, adding the fluxes, and reconverting to magnitudes. This procedure is performed separately for G and R, then the resulting values are combined to form $(G - R, G)$.

Each combination of (M_i, q_j) forms a normalized bivariate Gaussian in (M, q) space with covariance matrix, $\mathbf{S}_{M,q} = \text{diag}(\sigma_M^2, \sigma_q^2)$. This transforms to a covariance matrix in $(G - R, G)$ space, $\mathbf{S}_{G-R,G} = \mathbf{J}\mathbf{S}_{M,q}\mathbf{J}^T$, where the Jacobian

$$\mathbf{J} = \begin{bmatrix} \frac{\partial G - R}{\partial M} & \frac{\partial G - R}{\partial q} \\ \frac{\partial G}{\partial M} & \frac{\partial G}{\partial q} \end{bmatrix}. \tag{19}$$

The partial derivatives in the Jacobian are not dependent on the data or model parameters and can be pre-computed for any given (M, q) by interpolating colours and magnitudes from the isochrone. The combination of all the (M, q) bivariate Gaussians map to a set of tilted overlapping bivariate Gaussians in $(G - R, G)$ space.

Similar to what we have done previously, we expand the binary-star likelihood

$$P_B(\mathbf{D}_k|\boldsymbol{\theta}) = \int P(\mathbf{D}_k|\mathbf{D}')P_B(\mathbf{D}'|\boldsymbol{\theta})\,\mathrm{d}^2\mathbf{D}'. \tag{20}$$

This can be further decomposed as

$$P_B(\mathbf{D}_k|\boldsymbol{\theta}) = \sum_{i,j} a_i b_j \int P(\mathbf{D}_k|\mathbf{D}')P_B(\mathbf{D}'|M_i, q_j)\,\mathrm{d}^2\mathbf{D}' \tag{21}$$

$$= \sum_{i,j} a_i b_j \int \mathcal{N}(\mathbf{D}_k - \mathbf{D}', h^2\mathbf{S}_k)\mathcal{N}$$
$$\times (\mathbf{D}' - \mathbf{D}_{ij}, \mathbf{S}_{ij})\,\mathrm{d}^2\mathbf{D}', \tag{22}$$

where $\mathbf{D}_{ij} \equiv \mathbf{D}(M_i, q_j)$ and $\mathbf{S}_{ij} \equiv \mathbf{J}\mathbf{S}_{M_i,q_j}\mathbf{J}^T$. Here, the choice of Gaussian bases for M and q becomes apparent. The integral is again a convolution that can be evaluated analytically, so that

$$P_B(\mathbf{D}_k|\boldsymbol{\theta}) = \sum_{i,j} a_i b_j \mathcal{N}(\mathbf{D}_k - \mathbf{D}_{ij}, h^2\mathbf{S}_k + \mathbf{S}_{ij}). \tag{23}$$

3.5 Total likelihood

Expanding equation (4),

$$\ln P(\mathbf{D}|\boldsymbol{\theta}) = \sum_k \ln \Big\{ (1 - f_B - f_O)P_S(\mathbf{D}_k|\boldsymbol{\theta})$$
$$+ f_B P_B(\mathbf{D}_k|\boldsymbol{\theta}) + f_O P_O(\mathbf{D}_k|\boldsymbol{\theta}) \Big\}. \tag{24}$$

Inserting the individual likelihoods developed above, we can write this in a compact nested form,

$$\ln P(\mathbf{D}|\boldsymbol{\theta}) = \sum_k \text{LSE}\{A_k, B_k, C_k\}, \tag{25}$$

where LSE is the log sum exp function,

$$\text{LSE}(a, b, ...) \equiv \ln\left(e^a + e^b + ...\right), \tag{26}$$

and

$$A_k = -\frac{1}{2}(\mathbf{D}_k - \mathbf{D}_O)^T \mathbf{S}_0^{-1}(\mathbf{D}_k - \mathbf{D}_O) + \ln\frac{f_O}{2\pi\sqrt{\det(\mathbf{S}_O)}}, \tag{27}$$

$$B_k = \underset{l}{\text{LSE}}\Big\{ -\frac{1}{2}(\mathbf{D}_k - \mathbf{D}(M_l))^T (h^2\mathbf{S}_k)^{-1}(\mathbf{D}_k - \mathbf{D}(M_l))$$
$$+ \ln\frac{(1 - f_B - f_O)P(M_l|\boldsymbol{\theta})\Delta M}{2\pi\sqrt{\det(h^2\mathbf{S}_k)}} \Big\}, \tag{28}$$

$$C_k = \underset{ij}{\text{LSE}}\Big\{ -\frac{1}{2}(\mathbf{D}_k - \mathbf{D}_{ij})^T (h^2\mathbf{S}_k + \mathbf{S}_{ij})^{-1}(\mathbf{D}_k - \mathbf{D}_{ij})$$
$$+ \ln\frac{f_B a_i b_j}{2\pi\sqrt{\det(h^2\mathbf{S}_k + \mathbf{S}_{ij})}} \Big\}. \tag{29}$$

3.6 Priors

As described in the preceding sections, our general model has 13 parameters, $\boldsymbol{\theta} = (\gamma, k, M_0, a_1, a_2, a_3, \dot{a}_1, \dot{a}_2, \dot{a}_3, f_B, f_O, h_0, h_1)$. To compute the posterior probability distribution for $\boldsymbol{\theta}$, we must define a prior distribution, $P(\boldsymbol{\theta})$. We assume that this distribution is separable for most parameters, except a_k and $\dot{a}_k$, which we couple to limit the total change in shape of $P(q)$ along the main sequence. Additionally, we require h_1 to be positive, and scale its potential distribution with h_0. In the absence of any compelling previous knowledge, and after some trial and error, we generally adopt sensible uniform, normal or truncated-normal distributions as follows:

$$\begin{aligned} P(\log_{10} k) &= \mathcal{N}(2, 0.3) \\ P(M_0) &= \mathcal{N}(M_{\min} + \Delta M/2, \Delta M/2) \\ P(\gamma) &= \mathcal{N}(0, 1) \\ P(a_i) &= \mathcal{N}(0, 2) \\ P(\dot{a}_i) &= \mathcal{N}(0, 0.1/\Delta M)) \\ P(f_B) &= U(0.02, 0.95) \\ P(f_O) &= U(0, 0.05) \\ P(\log_{10} h_0) &= \mathcal{N}(0.1, 0.3) \\ P(h_1) &= \mathcal{N}_{\mathcal{T}}(0, 0.4h_0, 0, 2), \end{aligned} \tag{30}$$

where $M_{\min}$ is the minimum main-sequence mass, ΔM is half of the mass range of the CMD main sequence, $U(a, b)$ is the uniform distribution between a and b, and $\mathcal{N}_{\mathcal{T}}(a, b, c, d)$ is the normal distribution centred at a with standard deviation b truncated below c and above d.

4 RESULTS AND CONCLUSIONS

We have investigated four different models that use or do not use $\dot{a}_k$ and h_1 in various combinations.

Table 2. Posterior parameter values and 1 σ uncertainties.

	A	B	C	D
$\log_{10} k$	$2.37^{+0.18}_{-0.16}$	$2.35^{+0.18}_{-0.16}$	$2.40^{+0.19}_{-0.18}$	$2.37^{+0.18}_{-0.16}$
M_0	$0.549^{+0.0031}_{-0.0034}$	$0.549^{+0.0032}_{-0.0034}$	$0.5490^{+0.0030}_{-0.0034}$	$0.549^{+0.0031}_{-0.0032}$
γ	$0.04^{+0.25}_{-0.25}$	$0.04^{+0.24}_{-0.27}$	$0.03^{+0.25}_{-0.24}$	$0.03^{+0.24}_{-0.24}$
a_1	$0.4^{+0.5}_{-0.5}$	$0.4^{+0.5}_{-0.6}$	$0.4^{+0.5}_{-0.5}$	$0.4^{+0.5}_{-0.6}$
a_2	$0.71^{+0.34}_{-0.30}$	$0.7^{+0.36}_{-0.33}$	$0.71^{+0.35}_{-0.30}$	$0.7^{+0.39}_{-0.33}$
a_3	$-0.1^{+0.4}_{-0.4}$	$-0.1^{+0.4}_{-0.4}$	$-0.1^{+0.4}_{-0.4}$	$-0.1^{+0.4}_{-0.4}$
$\dot{a}_1$	0	$-0.1^{+0.4}_{-0.4}$	0	$-0.1^{+0.4}_{-0.4}$
$\dot{a}_2$	0	$0.0^{+0.4}_{-0.4}$	0	$0.0^{+0.4}_{-0.4}$
$\dot{a}_3$	0	$0.0^{+0.4}_{-0.4}$	0	$0.0^{+0.4}_{-0.4}$
f_B	$0.43^{+0.10}_{-0.07}$	$0.43^{+0.11}_{-0.07}$	$0.43^{+0.11}_{-0.07}$	$0.42^{+0.11}_{-0.07}$
f_O	$0.0060^{+0.0044}_{-0.0028}$	$0.0061^{+0.0043}_{-0.0029}$	$0.0062^{+0.0045}_{-0.0030}$	$0.0063^{+0.0044}_{-0.0030}$
h_0	$1.36^{+0.11}_{-0.10}$	$1.36^{+0.11}_{-0.10}$	$1.69^{+0.08}_{-0.08}$	$1.69^{+0.08}_{-0.07}$
h_1	$0.17^{+0.05}_{-0.05}$	$0.17^{+0.05}_{-0.05}$	0	0
$\Delta \ln P$	0.00	-0.01	-4.49	-4.49
$\Delta \log_{10} Z$	0.00	-0.04	-3.08	-3.25

To sample the posterior distributions (equation 25), we have used the affine-invariant ensemble sampler EMCEE (Foreman-Mackey et al. 2013), and the nested sampler (Skilling 2004) DYNESTY (Higson et al. 2019; Speagle 2020). Fundamentally, these are different approaches to sampling the probability space. EMCEE, a Markov Chain Monte Carlo sampler, starts with a number of points (walkers) distributed close to a guess at the posterior maximum probability. It then uses a quasi-downhill iteration on $-\ln P$ to roughly locate the maximum. The walkers converge to a stationary distribution that mirrors the probability distribution. In contrast, the nested sampler begins with points distributed throughout the entire prior hypervolume, iteratively replacing points that are outside a rising probability contour. Despite the different methods, both samplers converged to the same results. Following this process, the Nelder–Mead method was used to locate the maximum posterior probability point for each model.

In Table 2, we list the parameter medians and 1σ uncertainties for our four models that include or exclude the combinations of $\dot{a}_k$ and h_1. Posterior parameter distributions for all parameters for the most general model (model B) are shown as marginalized corner plots and histograms in Fig. 4. The parameters are well defined but, as might be expected, there are covariances between f_B and the shape parameters for the q distribution (i.e. we would expect q distributions that rise towards $q = 0$ to result in a higher value of f_B). Corner plots for the more restricted models are similar.

Also listed in Table 2 are two extra values for each model. First, we quote the relative maximum probability, $\Delta \ln P(\boldsymbol{\theta}|\boldsymbol{D})$, for each model. This is a measure of how well the model fits the data. For readers more familiar with the χ^2 statistic, $\Delta \ln P$ is equivalent to $\Delta\chi^2/2$. Secondly, we quote the Bayes factor, $\Delta \log_{10} Z$, which indicates the overall relative probability of the different models given the data, compensating implicitly for the extra fit freedom that is introduced with additional parameters. An interpretation of $\Delta \log_{10} Z$ is given by Kass & Raftery (1995), that $\Delta \log_{10} Z >$ (0.5, 1, 2) represents ('substantial', 'strong', 'decisive') evidence for a proposition. The quoted $\Delta \log_{10} Z$ values have an estimated uncertainty of 0.2.

In Fig. 5, we summarize the implied results for the mass ratios and binary fractions. For each model, we use random samples from the posterior parameter distribution to show the mass-ratio distribution

function and the fraction of binary stars with a mass ratio greater than a given q,

$$f_B(q' > q) = f_B \int_q^1 P(q|a_k, \dot{a}_k, M; k = 1..3)\, dq. \tag{31}$$

This latter quantity is tabulated in Table 3 for equally spaced values of q. Readers may use this table to find binary fractions with mass ratios above any particular desired limit. The distributions implied by the prior, equation (30), are shown for comparison, demonstrating that it is the likelihood, not the prior, that is driving the results.

In Fig. 6, we show our final data, along with three random realizations of model A with the posterior maximum probability parameters.

From these results we conclude the following:

(i) There is decisive evidence that models A and B, with error bar scaling that increases towards the lower main sequence (h_1), are better than those with constant scaling. We remind the reader that an error bar scaling greater than 1 is equivalent to allowing an intrinsic width to the main sequence, so that models A and B can also be interpreted as allowing for an increasing main-sequence width towards fainter magnitudes.

(ii) All of the models have a rising mass-ratio distribution function with q for $q > 0.5$ (Fig. 5). This shows that more binary stars (per unit q) exist with high-mass ratios than intermediate-mass ratios, but not with the extreme peak found by Fisher et al. (2005).

(iii) Below $q \approx 0.3$, there is little constraint on the form of the mass-ratio distribution function. This is to be expected, since such binary stars lie on or very close to the main sequence and cannot be distinguished from single stars.

(iv) The implied binary fraction is well constrained for $q \gtrsim 0.3$, but becomes increasingly less so as $q \to 0$.

(v) The 'best-fitting' model is model A, which has a constant mass-ratio distribution function shape. However, the improvement of fit over the other model (B) that includes h_1 is negligible ($\Delta \ln P = 0.01$).

(vi) The model with the strongest evidence, Z, is also model A with a constant shape to the q distribution. However, the evidence that this model is better than model B (which has a mass-varying shape) is insubstantial. Having fewer parameters, we adopt model A as our favoured option.

(vii) Since the fraction of outliers, f_O, is negligible for each model, we can interpret f_B as being the binary frequency, without need for adjustment.

(viii) Formally, the favoured models A and B imply an overall binary fraction, $f_B = 0.43^{+0.11}_{-0.07}$, but this result may be unreliable. The lower limit is sensible, and is imposed by the reliable detection of binaries with $q > 0.5$, coupled with the realistic expectation that the binary frequency below this threshold is not zero. However, the reason for the upper limit is not obvious, yet is none the less required by the data. Larger values of f_B could eventuate if the binary mass-ratio distribution function were to rise more steeply towards $q = 0$ than is apparent in Fig. 5. Such behaviour is permitted by the prior (see the final column of Fig. 5), yet is not favoured by the (data-driven) likelihood. We are unable to pinpoint a particular physical cause. Restricting ourselves to higher mass-ratio binaries, we find $f_B(q > 0.5) = 0.258 \pm 0.019$.

5 SUMMARY

We have presented a new generative model for star-cluster CMD that allows the measurement of the binary star mass-ratio distribution as
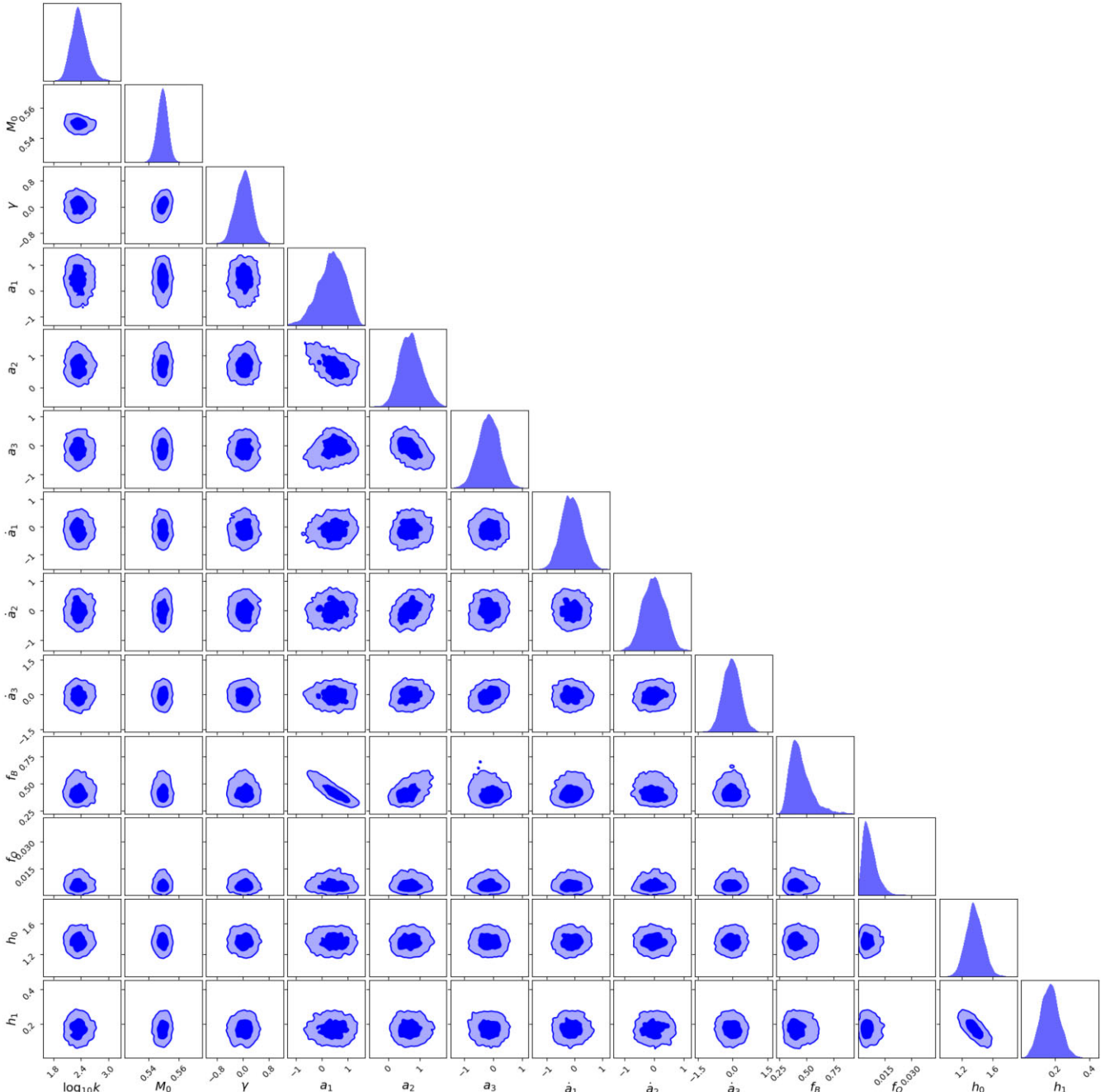
Figure 4. Corner plot for the most general model, model B, showing the $1\,\sigma$ and $2\,\sigma$ contours of the posterior probability distribution projected against each pair of parameters (marginalized over the remaining parameters). The 1D histograms show the marginalized probability distributions for each individual parameter.

well as the binary star frequency. The model naturally accounts for the locations of stars on the CMD without individual classification of stars as single or binaries.

The first application of the model is presented using *Gaia* photometry of the old open cluster, M67. By considering a 5D phase-space distribution of stars in the direction of M67, we have obtained a very clean sample of cluster stars that cover the low-main sequence through to just above the main-sequence turnoff.

We find the frequency of binary stars in M67 with mass ratios greater than 0.4 to be $f_B(q > 0.5) = 0.258 \pm 0.019$. The mass-ratio distribution is found to rise gently towards higher mass ratios. There

is no compelling evidence that the form of the distribution varies with primary mass along the main sequence.

DATA AND CODE AVAILABILITY

Gaia Early Data Release 3 (EDR3) data are publicly available via the *Gaia* archive, https://gea.esac.esa.int/archive/, and the Centre de Données astronomiques de Strasbourg(CDS) catalogue service, https://vizier.cds.unistra.fr/viz-bin/VizieR.

The code used for this analysis is written in PYTHON and CUDA (via the PYCUDA PYTHON library). CUDA is an extension to C/C++ that uses an NVIDIA graphical processing unit to per-
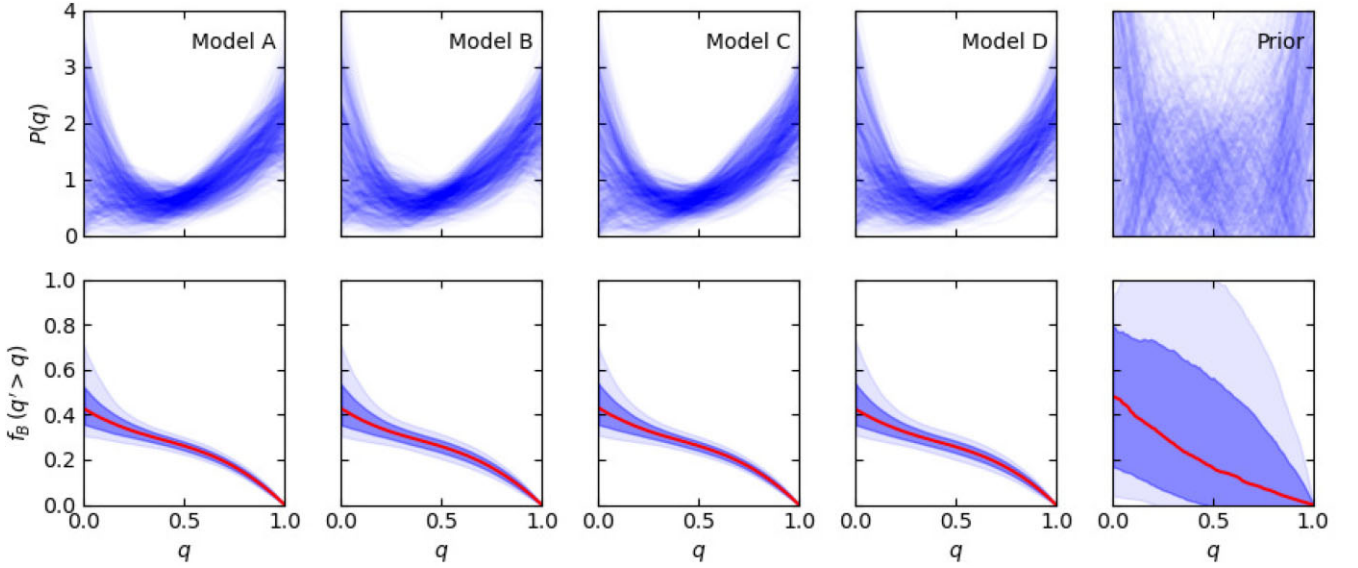
Figure 5. Summary of the mass-ratio distribution function and binary fraction for the different models. Each of the first four columns represents a different model, corresponding to the tabulated parameters in Table 2. The upper row shows the mass-ratio distribution function for 1000 random weighted samples from the posterior distribution. The lower row shows the fraction of binary stars with mass ratio greater than q for each model, with the red line representing the median of the posterior distribution and the shaded regions indicating the 1σ and 2σ uncertainties in this quantity. The final column shows the same information, but obtained from random sampling of the prior distribution.

Table 3. Fraction of binary stars with mass ratio greater than a given q for each model.

q	A	B	C	D
0.1	$0.37^{+0.05}_{-0.05}$	$0.37^{+0.06}_{-0.05}$	$0.38^{+0.06}_{-0.05}$	$0.37^{+0.06}_{-0.05}$
0.2	$0.34^{+0.034}_{-0.033}$	$0.34^{+0.039}_{-0.034}$	$0.34^{+0.037}_{-0.033}$	$0.33^{+0.04}_{-0.04}$
0.3	$0.310^{+0.024}_{-0.024}$	$0.307^{+0.030}_{-0.029}$	$0.310^{+0.027}_{-0.025}$	$0.305^{+0.031}_{-0.029}$
0.4	$0.284^{+0.020}_{-0.020}$	$0.282^{+0.027}_{-0.027}$	$0.285^{+0.021}_{-0.022}$	$0.280^{+0.027}_{-0.026}$
0.5	$0.258^{+0.019}_{-0.019}$	$0.256^{+0.026}_{-0.026}$	$0.258^{+0.020}_{-0.020}$	$0.254^{+0.026}_{-0.026}$
0.6	$0.227^{+0.018}_{-0.018}$	$0.224^{+0.025}_{-0.025}$	$0.228^{+0.019}_{-0.019}$	$0.224^{+0.025}_{-0.026}$
0.7	$0.188^{+0.016}_{-0.016}$	$0.185^{+0.023}_{-0.022}$	$0.188^{+0.017}_{-0.017}$	$0.186^{+0.023}_{-0.023}$
0.8	$0.136^{+0.014}_{-0.013}$	$0.134^{+0.019}_{-0.019}$	$0.137^{+0.015}_{-0.014}$	$0.136^{+0.019}_{-0.019}$
0.9	$0.071^{+0.010}_{-0.009}$	$0.070^{+0.013}_{-0.012}$	$0.072^{+0.010}_{-0.010}$	$0.071^{+0.013}_{-0.012}$

form parallel calculations. We use CUDA to perform the likelihood calculation from Section 3.5. The CUDA language exposes GPU hardware cores as groups of threads arranged into blocks. In our code, GPU threads are used to sum over basis functions, and blocks to sum over data points. The code is being developed publicly at https://github.com/MichaelDAlbrow/CMDF itter.
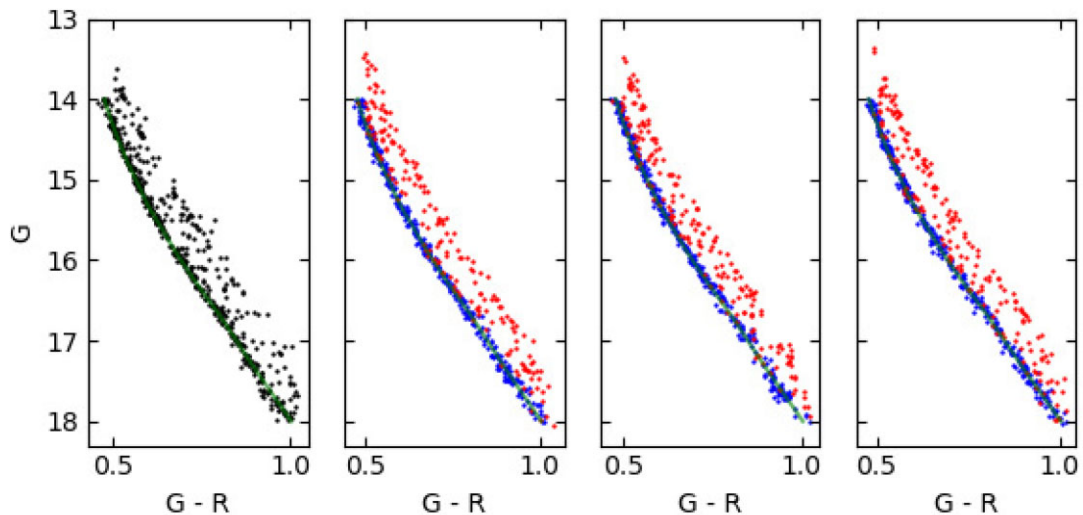
ACKNOWLEDGEMENTS

Figure 6. CMDs for the data (left-hand panel) and three random realizations of the posterior maximum for model A. Single stars are shown in blue, binaries in red.

REFERENCES

Albrow M. D., Gilliland R. L., Brown T. M., Edmonds P. D., Guhathakurta P., Sarajedini A., 2001, ApJ, 559, 1060
Bressert E. et al., 2010, MNRAS, 409, L54
Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Davenport J. R. A., Sandquist E. L., 2010, ApJ, 711, 559
Dotter A., 2016, ApJS, 222, 8
Duquennoy A., Mayor M., 1991, A&A, 500, 337
Elson R. A. W., Sigurdsson S., Davies M., Hurley J., Gilmore G., 1998, MNRAS, 300, 857
Evans D. W. et al., 2018, A&A, 616, A4
Fan X. et al., 1996, AJ, 112, 628
Fisher J., Schröder K.-P., Smith R. C., 2005, MNRAS, 361, 495
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Francic S. P., 1989, AJ, 98, 888
Gao X., 2018, ApJ, 869, 9
Geller A. M., Mathieu R. D., Latham D. W., Pollack M., Torres G., Leiner E. M., 2021, AJ, 161, 190
Higson E., Handley W., Hobson M., Lasenby A., 2019, Stat. Comput., 29, 891
Hobbs L. M., Thorburn J. A., 1991, AJ, 102, 1070
Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Hurley J. R., Pols O. R., Aarseth S. J., Tout C. A., 2005, MNRAS, 363, 293
Hut P. et al., 1992, PASP, 104, 981
Ji J., Bregman J. N., 2013, ApJ, 768, 158
Kass R. E., Raftery A. E., 1995, J. Am. Stat. Assoc., 90, 773
Lawson C. L., Hanson R. J., 1974, Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, NJ
Li C., de Grijs R., Deng L., 2013, MNRAS, 436, 1497
Mathieu R. D., 2000, in Pallavicini R., Micela G., Sciortino S., eds, ASP Conf. Ser. Vol. 198, Stellar Clusters and Associations: Convection, Rotation, and Dynamos. Astron. Soc. Pac., San Francisco, p. 517
Mathieu R. D., Latham D. W., 1986, AJ, 92, 1364
Milone A. P. et al., 2012, A&A, 540, A16
Montgomery K. A., Marschall L. A., Janes K. A., 1993, AJ, 106, 181
Önehag A., Gustafsson B., Korn A., 2014, A&A, 562, A102
Pasquini L. et al., 2012, A&A, 545, A139
Paxton B. et al., 2013, ApJS, 208, 4
Paxton B. et al., 2015, ApJS, 220, 15
Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3
Riello M. et al., 2021, A&A, 649, A3
Sarajedini A., Dotter A., Kirkpatrick A., 2009, ApJ, 698, 1872
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Proc. Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 395
Sollima A., Carballo-Bello J. A., Beccari G., Ferraro F. R., Pecci F. F., Lanzoni B., 2010, MNRAS, 401, 577
Speagle J. S., 2020, MNRAS, 493, 3132
Taylor E. N. et al., 2015, MNRAS, 446, 2144
Ward J. L., Kruijssen J. M. D., Rix H.-W., 2020, MNRAS, 495, 663
Yadav R. K. S. et al., 2008, A&A, 484, 609

APPENDIX A: CMD COVARIANCE

The covariance, $\sigma_{G-R,G}$ can be calculated from the definition of covariance and the algebra of expectation values (E),

$$
\begin{aligned}
\sigma_{G-R,G} &= \mathrm{Cov}(G-R, G) \\
&= E[(G-\mu_G)((G-R)-\mu_{G-R}] \\
&\quad - E[G-\mu_G]E[(G-R)-\mu_{G-R}] \\
&= E[(G-\mu_G)((G-\mu_G)-(R-\mu_R))] - 0 \times 0 \\
&= E[(G-\mu_G)^2 - (G-\mu_G)(R-\mu_R)] \\
&= E[(G-\mu_G)^2] - E[(G-\mu_G)(R-\mu_R)] \\
&= E[(G-\mu_G)^2] - E[(G-\mu_G)]E[(R-\mu_R)] \\
&= \sigma_G^2 - 0 \times 0 \\
&= \sigma_G^2 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (A1)
\end{aligned}
$$

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