Journal of Physics: Conference Series

PAPER • OPEN ACCESS

An Application of The Markov Chain Monte Carlo (MCMC) Method to Open Cluster Membership Determination

To cite this article: Y A Hidayat *et al* 2019 *J. Phys.: Conf. Ser.* **1231** 012029

An Application of The Markov Chain Monte Carlo (MCMC) Method to Open Cluster Membership Determination

Y A Hidayat, M I Arifyanto, Aprilia and M I Hakim

[1]Department of Astronomy, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia

Email: yudarhd@gmail.com

Abstract. The determination of membership in open clusters will be more difficult than determination membership of the globular clusters. This is due to the location of open clusters which located on the galactic disk. From that location, when observing open clusters, it will be observed foreground and background stars also. Determination of membership in cluster could be done by calculating the probability of a star becoming a cluster member. We used double elliptic bivariate Gaussian function to represent the population of cluster and field stars in the region. The function has eleven parameters and solved by using Markov Chain Monte Carlo (MCMC). The Metropolis-Hastings algorithm is used in this works to obtain the probability of open cluster membership. We used astrometric data from GAIA Data Release 2 for two open clusters which their membership are to be determined i.e. IC 1590 and NGC 2262. Our result is nearly in accordance with the MWSC II catalogue. After the cluster membership is obtained, the physical parameters of the two clusters using isochrones fitting from [1] are the distance(pc) 2200 and 2366, reddening(mag) 0.728 and 0.830, distance modulus(mag) 12.210 and 12.438 also log age 6.895 and 8.995, respectively for IC 1590 and NGC 2262.

Keywords: *Markov Chain Monte Carlo (MCMC) and Open Cluster Membership Determination*

1. Introduction

The Study determining the membership of clusters is really important to learn about stellar evolution and Galaxy. A star cluster is any physically related group of stars that formed together, they remain together due to their mutual gravitational attraction. Star clusters are divided into two types which are open cluster and globular cluster. This paper will be focused on determining of open clusters because open clusters are located on the galactic disk. From that information, the field stars; foreground and background stars will be observed as well. So determining the membership of open clusters becomes important for identifying the actual members of the cluster. The determination of membership of open clusters can be done by calculating the probability of a star becoming a member of the cluster. To calculate the probability, we use double elliptic bivariate Gaussian function from [2] which contains have eleven parameters. In this work, Markov Chain Monte Carlo (MCMC) is applied to separate member stars from the field stars using Metropolis-Hastings algorithm. To ensure the parameters obtained by MCMC, quality of fit (D) parameter should be calculated. If the obtained value of D is close to 0, then the value from eleven parameters fit to the model. This probability is compared from MWSC II (Milky Way Stellar Cluster II) from [3].

IOP Conf. Series: Journal of Physics: Conf. Series **1231** (2019) 012029 doi:10.1088/1742-6596/1231/1/012029

2. Data and Methods

Proper motion data from GAIA Data Release 2 is used as the astrometric data to determine the membership of IC 1590 and NGC 2262. As for the photometry data, we use magnitude of the stars in G band from GAIA Data Release 2 and also J, H band from 2MASS. These photometry data is used to construct CMD. Data of probability from MWSC II catalogue is used as the comparation to the result. The data from this research has been selected with the criteria:

i) Parallax error < 0.15 (mas)
ii) $-15.0 < \mu_x < 10.0$ with error $\mu_x < 0.3$ (mas/yr)
iii) $-15.0 < \mu_y < 20.0$ with error $\mu_y < 0.3$ (mas/yr)

The stars of the cluster can be identified using double elliptic bivariate Gaussian equation

$$\begin{aligned}\Phi(\mu_x, \mu_y) &= (1-n_f)\Phi_c(\mu_x, \mu_y) + n_f\Phi_f(\mu_x, \mu_y) \\ &= \frac{1-n_f}{2\pi\sigma_{xc}\sigma_{yc}\sqrt{1-\rho_c^2}}\exp\left\{-\frac{1}{2(1-\rho_c^2)}\left[\begin{matrix}\left(\frac{\mu_x-\mu_{x0c}}{\sigma_{xc}}\right)^2+\left(\frac{\mu_y-\mu_{y0c}}{\sigma_{yc}}\right)^2 \\ -2\rho_c\left(\frac{\mu_x-\mu_{x0c}}{\sigma_{xc}}\right)+\left(\frac{\mu_y-\mu_{y0c}}{\sigma_{yc}}\right)\end{matrix}\right]\right\} \\ &+ \frac{n_f}{2\pi\sigma_{xf}\sigma_{yf}\sqrt{1-\rho_f^2}}\exp\left\{-\frac{1}{2(1-\rho_f^2)}\left[\begin{matrix}\left(\frac{\mu_x-\mu_{x0f}}{\sigma_{xf}}\right)^2+\left(\frac{\mu_y-\mu_{y0f}}{\sigma_{yf}}\right)^2 \\ -2\rho_f\left(\frac{\mu_x-\mu_{x0f}}{\sigma_{xf}}\right)\left(\frac{\mu_y-\mu_{y0f}}{\sigma_{yf}}\right)\end{matrix}\right]\right\}. \qquad (1)\end{aligned}$$

This equation depends on eleven parameters namely n_f is the proportion of field stars, $\mu_{x0c}, \mu_{y0c}, \sigma_{xc}$ and $\sigma_{yc}$are the proper motion centroids and the dispersions or standard deviations for the cluster, respectively; $\mu_{x0f}, \mu_{y0f}, \sigma_{xf}$ and σ_{yf} are the proper motion centroids and the dispersions or standard deviations for the field, respectively; ρ_c and ρ_f are the cluster and the field proper motion correlation coefficients. To derive the eleven parameters, the MCMC method calculate a random probability of posterior that is

$$P(\theta|D, M) \propto P(D|\theta, M)P(\theta|M), \qquad (2)$$

Where $P(D|\theta, M)$ is likelihood function and $P(\theta|M)$ is prior or range value of eleven parameters. Each of parameter, data and model is denoted as θ, D and $M.$ Bayes rules is used to determine the probability of member stars,

$$P_i(C|\mu_{xi}, \mu_{yi}) = \frac{(1-n_f)\Phi_c(\mu_{xi}, \mu_{yi})}{n_f\Phi_f(\mu_{xi}, \mu_{yi}) + (1-n_f)\Phi_c(\mu_{xi}, \mu_{yi})}, \qquad (3)$$

if the value of membership probability $> 50\%$ then it is identified as a member star and if the value of membership probability $< 50\%$ then it is identified as a field star. The MCMC method is very depended on the number of iterations, the greater number of iterations performed, the closer the value of eleven parameters to convergence of each parameters. Therefore, the value of each parameters obtained from mean and standard deviation. On the results obtained, then after the value obtained from eleven parameters, based on [4] to confirm from value of each parameter can be calculated from D parameter or quality of fit

$$D = \frac{(\mu_{x0cMC}-\mu_{x0cLIT})^2}{(\sigma_{xcMC})^2+(\sigma_{xcLIT})^2} + \frac{(\mu_{y0cMC}-\mu_{y0cLIT})^2}{(\sigma_{ycMC})^2+(\sigma_{ycLIT})^2}, \qquad (4)$$

where MC is value from MCMC method and LIT is value from references.

If the value is very small or that means close to 0, the value obtained is considered good. To confirm that the probability value of the parameters is greater than 50% Vector Point Diagram (VPD) and Color-Magnitude Diagram (CMD) are done by performing isochrone fitting of [1] on physical parameters of MWSC II. The greater the probability value, the smaller the number of clusters members. After getting the probability value > 50% for each star (member stars), we choose the data near the peak of parallax value for further analysis.

3. Results and Analysis

The MCMC method used in this work is applied for the two open clusters. The initial and prior values are listed in the table 1. The best parameters values are obtained after iterating 1,000,000 times with step of 0.001 for IC 1590 and NGC 2262. The value obtained is the mean value with standard deviation for eleven parameters. The parameters values are listed in the table 2.

Table 1. Initial and prior for IC 1590[a] and NGC 2262[b]

Parameter	Initial[a]	Prior[a]	Initial[b]	Prior[b]
n_f	0.5	$0.0 < n_f < 1.0$	0.5	$0.0 < n_f < 1.0$
μ_{x0c}	-1.736 ± 0.081	$-4.0 < \mu_{x0c} < 0.5$	0.189 ± 0.031	$-0.4 < \mu_{x0c} < 0.7$
μ_{y0c}	-0.954 ± 0.048	$-2.5 < \mu_{y0c} < 0.5$	0.118 ± 0.017	$-0.4 < \mu_{y0c} < 0.4$
σ_{xc}	0.345 ± 0.031	$0.0 < \sigma_{xc} < 9.0$	0.061 ± 0.001	$0.0 < \sigma_{xc} < 2.0$
σ_{yc}	0.279 ± 0.012	$0.0 < \sigma_{yc} < 6.0$	0.053 ± 0.006	$0.0 < \sigma_{yc} < 1.3$
ρ_c	0.015	$-1.0 < \rho_c < 1.0$	0.015	$-1.0 < 1\rho_c < 1.0$
μ_{x0f}	-1.634 ± 0.037	$-4.0 < \mu_{x0f} < 2.0$	0.149 ± 0.013	$-0.6 < \mu_{x0f} < 0.8$
μ_{y0f}	-0.999 ± 0.016	$-3.0 < \mu_{y0f} < 1.0$	0.124 ± 0.006	$-0.5 < \mu_{y0f} < 0.8$
σ_{xf}	3.552 ± 0.075	$0.0 < \sigma_{xf} < 12.0$	0.824 ± 0.03	$0.0 < \sigma_{xf} < 4.0$
σ_{yf}	2.614 ± 0.031	$0.0 < \sigma_{yf} < 8.0$	1.292 ± 0.024	$0.0 < \sigma_{yf} < 3.0$
ρ_f	0.015	$-1.0 < \rho_f < 1.0$	0.015	$-1.0 < \rho_f < 1.0$

[a] the values for IC 1590
[b] the values for NGC 2262

Table 2. Parameter value using MCMC method for IC 1590[a] and NGC 2262[b]

Parameter	IC 1590	NGC 2262
N	2874	908
n_f	0.304 ± 0.016	0.429 ± 0.022
μ_{x0c}	-1.613 ± 0.046	0.064 ± 0.026
μ_{y0c}	-0.975 ± 0.018	0.101 ± 0.029
σ_{xc}	1.692 ± 0.048	0.506 ± 0.027
σ_{yc}	1.133 ± 0.029	0.487 ± 0.036
ρ_c	0.015 ± 0.029	-0.114 ± 0.061
μ_{x0f}	-0.317 ± 0.186	-0.415 ± 0.112
μ_{y0f}	-0.972 ± 0.022	-0.902 ± 0.082
σ_{xf}	4.667 ± 0.089	2.678 ± 0.086
σ_{yf}	3.634 ± 0.122	3.913 ± 0.132
ρ_f	0.030 ± 0.034	-0.107 ± 0.047

The results of fitting using MCMC after cross-match between MWSC II and 2MASS are 30 for IC 1590 and 49 for NGC 2262. The number is the star with probability value $> 50\%$. The comparation number of member stars of probability between MCMC method and MWSC II Catalogue are listed in the table 3. The kinematic and photometric membership probabilities for stars in a cluster region of MWSC II catalogue is obtained from data-processing pipeline on the basis of stellar data from PPMXL and 2MASS. Based on the table 3, we can see the comparation number of member stars of each probability between MCMC method and MWSC II Catalogue. They differ due to the difference in the centre of the clusters used and the radius they encompass.

Table 3. Comparation number of member stars of each probability between MCMC method and MWSC II Catalogue after cross-match

	MCMC Method		MWSC II Catalogue	
Probability (%)	**IC 1590**	**NGC 2262**	**IC 1590**	**NGC 2262**
50	30	49	18	52
70	25	45	11	34
90	19	36	5	36

The obtained value of each parameter is calculated by using parameter D based on the equation (4). The value obtained in parameter D has the greatest value on NGC 2262 and the smallest value on IC 1590 written on the table 4. The value obtained has been in the order less than 10^{-1} that gives the meaning that the value obtained by the method is considered very good for the determination of open clusters membership. Then for the member stars, we obtained CMD by performing isochrone fitting of [4] on physical parameters are listed on table 5, the J and H band are used for the isochrone fitting. The distribution probability of membership for IC 1590 and NGC 2262 are described on the figure 1. Parallax value of each clusters are displayed on the figure 2 and we use the value near the peak.

Table 4. D value for IC 1590 and NGC 2262

	IC 1590	**NGC 2262**
Parameter D	0.0184	0.0489

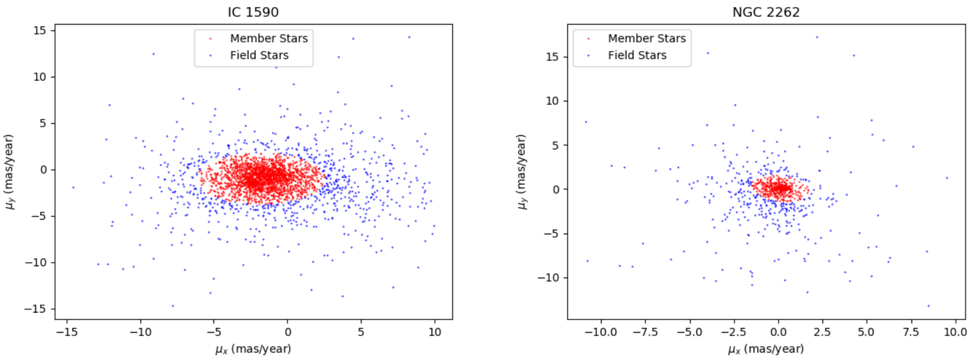

Figure 1. Vector Point Diagram (VPD) for IC 1590 (left) and NGC 2262 (right), the red color is representing member stars and the blue color is representing field stars

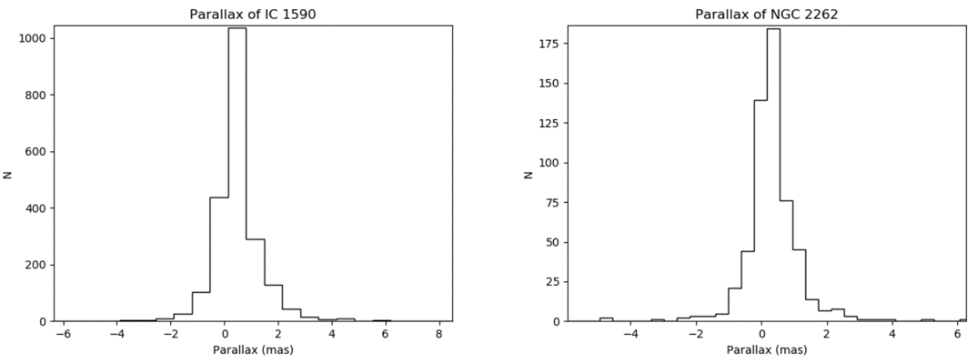

Figure 2. Histogram of parallax for IC 1590 (left) and NGC 2262 (right), the value near the peak is used for analysis

Table 5. The basic parameters for IC 1590 and NGC 2262 using isochrone fitting

Name	Distance (pc)	Reddening (mag)	Distance Modulus (mag)	Log Age
IC 1590	2200	0.728	12.210	6.895
NGC 2262	2366	0.830	12.438	8.995

The CMD of G, J and H band are very good and we can see the CMD from each open cluster is clearly on main sequences described by figure 3. Figure 4 describes the corner plot from each parameter. The parameter value obtained is the mean of the iteration after burn-in. The number of burn-in used in this work is 80% from the whole iteration. We check each value by evaluating the peak of the corner in every parameter and we get the value near the peak.

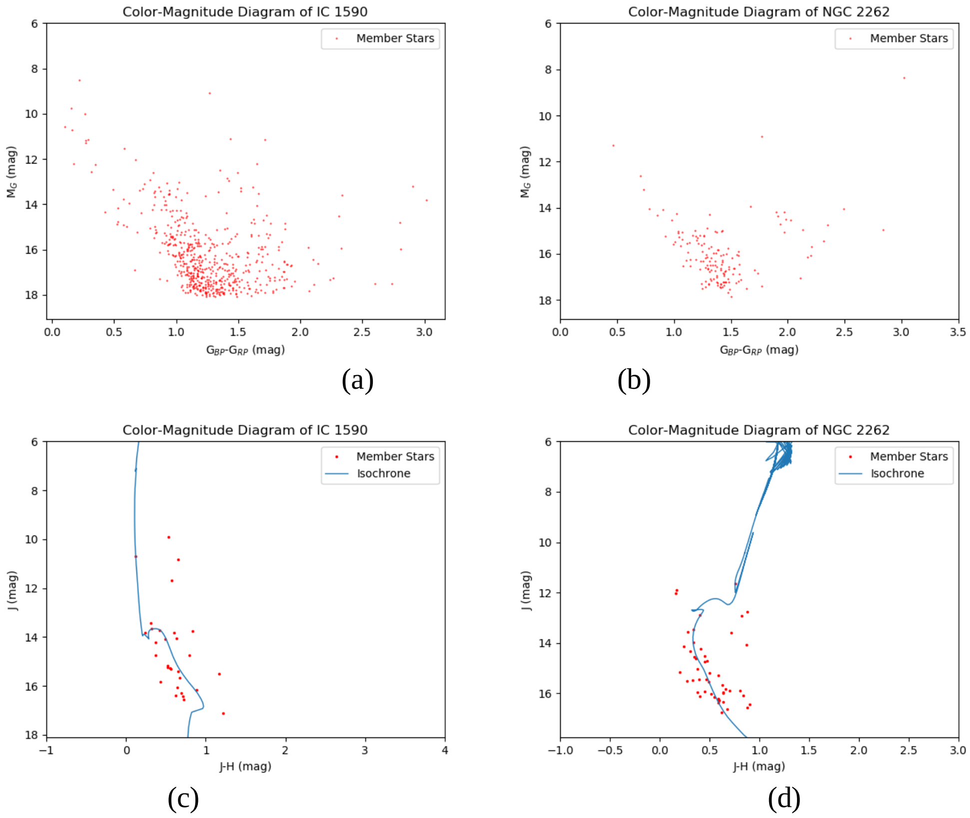

Figure 3. Color-Magnitude Diagram (CMD) using G band from GAIA Data Release 2 for member stars; probability value > 50% for (a) IC 1590 and (b) NGC 2262. Color-Magnitude Diagram (CMD) using J and H band from MWSC II after cross-match for member stars each open clusters. The red dots are member stars and the blue line is isochrone fitting from [4] for (c) IC 1590 and (d) NGC 2262

4. Conclusion and Discussion

The MCMC method can determine the open cluster membership by separating the stars around the star cluster into cluster and field stars. Using this method, the value of parameter D or quality of fit is obtained, with the largest value are 0.0489 for NGC 2262 and 0.0184 for IC 1595, that gives the meaning that the value obtained by the method is very good for the determination of open clusters membership. The amount of data and iteration is very influential on the time and convergence of the required value. The MCMC method is very good at determining the parameter value of double elliptic bivariate Gaussian function for a star cluster; it can be open clusters and globular clusters whose membership has not been determined. Furthermore, a precise of range value or prior is required to obtain a convergent parameter value. The advantages of MCMC method are usable for complicated distributions in high-dimensional spaces even when we don't know where the regions of high probability are, relatively easy to implement and fairly reliable to determine the parameter value from some functions.

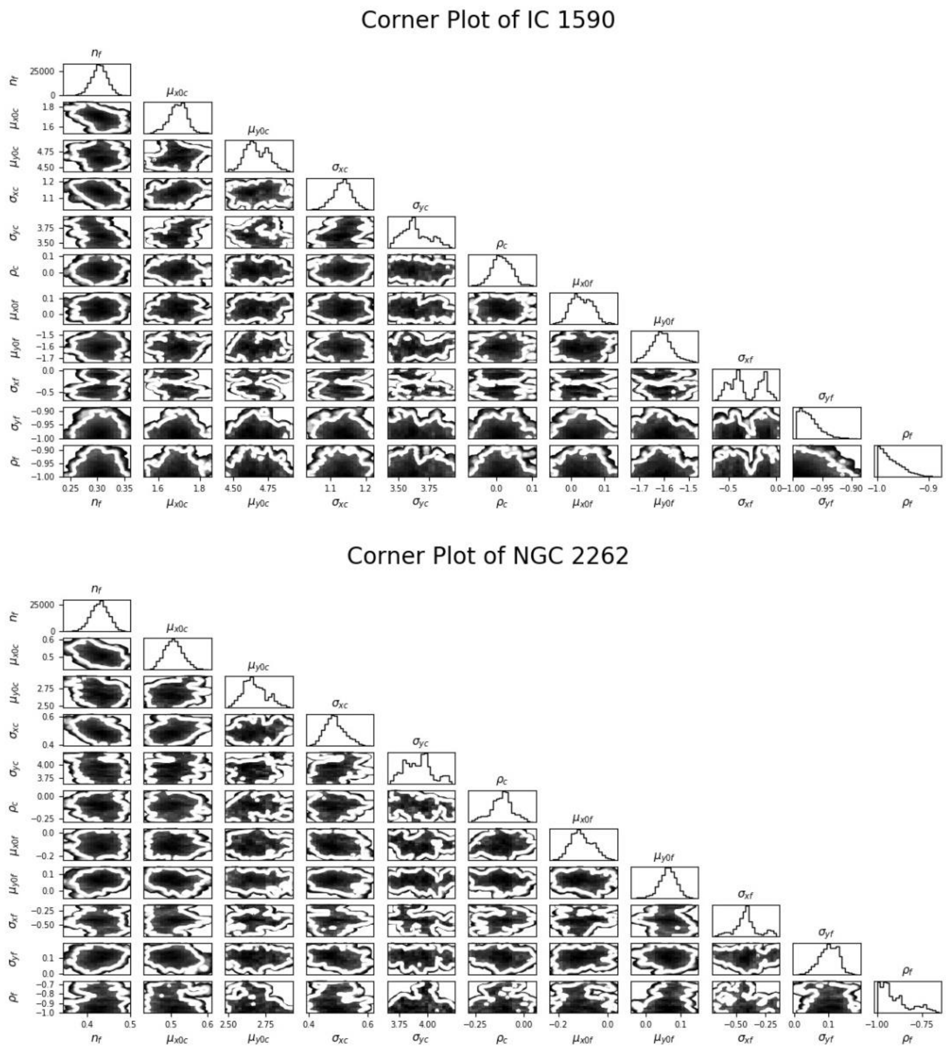

Figure 4. Corner plot from each parameters value for IC 1590 (upper) and NGC 2262 (lower). The parameter value obtained is the mean of steps after burn-in and the value is near the peak

Acknowledgments

We acknowledged Program Penelitian, Pengabdian Kepada Masyarakat, dan Inovasi (P3MI) for funding travel grant to SEAAN 2018

References

[1] Girardi L et al. 2000 *A&A* **141** 371-383
[2] Sabogal-Martinez B E *et al*. 2001 *Revista Mexicana de Astronomia y Astrofisica* **37** 105-114
[3] Kharchenko N V *et al*. 2013 *A&A* **53** 558
[4] Dias W S *et al*. 2014 *A&A* **79** 564
[5] Fitriana I K and M Ikbal Arifyanto 2016 *ISSEL. JoPCS* **771**
[6] Gaia Collaboration Brown A G A *et al*. 2018 *A&A* manuscript no. aa201833051
[7] Sharma S 2017 *Annu. Rev. Astron. Astrophys.* **55** 1-49