

Modeling Globular Cluster Counts with Bayesian Latent Models

Rafael S. de Souza ^{1,2,3} and Ana L. Chies-Santos ²

¹*Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*

²*Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 90040-060, Brazil*

³*Department of Physics & Astronomy, University of North Carolina at Chapel Hill, NC 27599-3255, USA*

ABSTRACT

We present a Bayesian latent model to describe the scaling relation between globular cluster populations and their host galaxies, updating the framework proposed in [de Souza et al. \(2015a\)](#). GC counts are drawn from a negative-binomial (NB) process linked to host stellar mass, augmented with a newly introduced Gaussian observation layer that enables efficient propagation of measurement errors. The revised formulation preserves the underlying NB process while improving computational tractability. The code snippets, implemented in `Nimble` and `PyMC` are released under the MIT license at [this repository](#) .

Keywords: Interdisciplinary astronomy — Astroinformatics — Astrostatistics

INTRODUCTION

Globular clusters (GCs) serve as valuable tracers of the formation histories and global properties of their host galaxies. Empirical correlations have long been observed between the number of GCs (N_{GC}) and various host-galaxy properties, including stellar mass, halo mass, and central velocity dispersion (e.g. [Blakeslee 1997](#); [Harris et al. 2013](#); [Hudson et al. 2014](#); [Valenzuela et al. 2021](#); [Zaritsky 2022](#); [Saifollahi et al. 2022](#); [Canossa-Gosteinski et al. 2024](#); [Dornan & Harris 2025](#)). For systems beyond the Local Group, N_{GC} are typically more accessible than their total masses, making the scaling relation between the number of clusters and galaxy mass of particular practical and physical interest.

These relations are often fitted using linear regressions intended for continuous responses, despite the inherently discrete nature of count data. Count models—members of the generalized linear model family—provide a principled alternative by linking a linear predictor to a discrete response, ensuring integer-valued outcomes and more realistic uncertainty quantification, especially in the low-count regime. Yet such models remain underutilized in astrophysical research, even though many observables—photon counts, supernova events, and planetary detections among them—are intrinsically discrete.

In [de Souza et al. \(2015a\)](#), we introduced a Bayesian negative-binomial (NB) framework—an exponential family distribution under the umbrella of generalized linear models (see e.g., [de Souza et al. 2015b](#); [Elliott et al. 2015](#))—to describe the dispersion of GC populations across a broad range of galaxy properties. This work, together with [Hilbe et al. \(2017\)](#) and [Hattab et al. \(2019\)](#), in which we implemented a hurdle model to characterize the escape fraction of ionizing radiation during the Epoch of Reionization, has since motivated a series of studies that expanded upon our original approach ([Berek et al. 2023, 2024](#)). The present note revisits that framework, refining the treatment of discrete measurement errors in GC counts. The revised model retains the hierarchical latent-variable treatment for photometric uncertainties but adopts a Gaussian observation layer over the NB process, providing a computationally efficient approximation that remains faithful across most of the observed range.

DATASET

To showcase the new model, we use a public catalog of globular cluster systems (Harris et al. 2013, 2014)¹ comprised by 422 galaxies covering the full Hubble sequence, 247 ellipticals, 94 lenticulars, 55 spirals, and 26 irregulars, for which both N_{GC} and auxiliary photometric quantities are available. For the present analysis, we adopt the stellar mass as the primary predictor, estimated from the V -band absolute magnitude and $(B-V)$ color using the mass-to-light ratio prescription of Bell et al. (2003), and only fitted the elliptical galaxies. The resulting masses span approximately $7.5 \lesssim \log_{10}(M_{\star}/M_{\odot}) \lesssim 12$.

MODEL

The updated framework models the observed globular-cluster counts and stellar masses, for galaxies $i = 1, \dots, N$, as noisy measurements of latent true quantities within a hierarchical Bayesian structure.

The updated framework models the observed globular cluster counts and stellar masses as noisy measurements of latent true quantities within a hierarchical Bayesian structure.

At the core, the intrinsic counts are governed by a discrete negative-binomial (NB) process,

$$\begin{aligned} N_{\text{GC},i} &\sim \text{NB}(p_i, k), & p_i &= \frac{k}{k + \mu_i}, \\ \mu_i &= \exp(\eta_i), & \eta_i &= \beta_0 + \beta_1 x_i, & x_i &\equiv \log_{10}(M_{\star,i}^*/M_{\odot}), \\ x_i^{\text{obs}} &\sim \mathcal{N}(x_i, e_{x,i}^2), & N_{\text{GC},i}^{\text{obs}} &\sim \mathcal{N}(N_{\text{GC},i}, e_{N_{\text{GC},i}}^2). \end{aligned} \quad (1)$$

Here, $x_i = \log_{10}(M_{\star,i}^*/M_{\odot})$ is the latent (true) stellar mass in solar units; $e_{x,i}$ and $e_{N_{\text{GC},i}}$ are the reported observational uncertainties. Although Gaussian terms describe measurement noise (appropriate for continuous quantities), the NB likelihood preserves the discreteness of N_{GC} . The dispersion parameter k controls departures from the Poisson limit. This formulation allows measurement errors in both axes to be explicitly modeled, preserving the discreteness of the intrinsic GC counts while accommodating observational scatter.

ANALYSIS AND DISCUSSION

We fit the hierarchical negative-binomial model using $x = \log_{10}(M_{\star}/M_{\odot})$ as the predictor derived from M_V and color (Section 2). The fitted relation and uncertainties are shown in Figure 1. The NB likelihood accommodates strong extra-Poisson variation while preserving the discreteness of N_{GC} , so zeros and small counts are handled naturally. Across the sampled mass range the trend is positive and approximately log-linear; the 95% credible band encloses the full dynamic range, including systems with $N_{\text{GC}} = 0$. Posterior checks (not shown) reveal no large-scale residual structure, and the inferred dispersion parameter k confirms substantial overdispersion relative to a Poisson model.

Figure 1 shows the empirical relation: the model preserves the lower *bound* of GC counts and properly propagates observational uncertainties through the latent, discrete NB layer. The y axis uses the base-10 pseudo-log transform with $\sigma = 5$, $y' = \text{asinh}(\frac{y/5}{\ln 10})$, which is linear near zero and logarithmic at large y , providing a continuous scale that accommodates both small and zero counts without distortion.

RSS acknowledges support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), process nos. 446508/2024-1 and 315026/2025-1. ACS acknowledges support from FAPERGS (grants 23/2551-0001832-2 and 24/2551-0001548-5), CNPq (grants 314301/2021-6, 312940/2025-4, 445231/2024-6, and 404233/2024-4), and CAPES (grant 88887.004427/2024-00).

Software: Nimble (de Valpine et al. 2017), PyMC (Salvatier et al. 2016).

REFERENCES

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D.

¹ http://www.physics.mcmaster.ca/~harris/GCS_table.txt

2003, ApJS, 149, 289, doi: 10.1086/378847

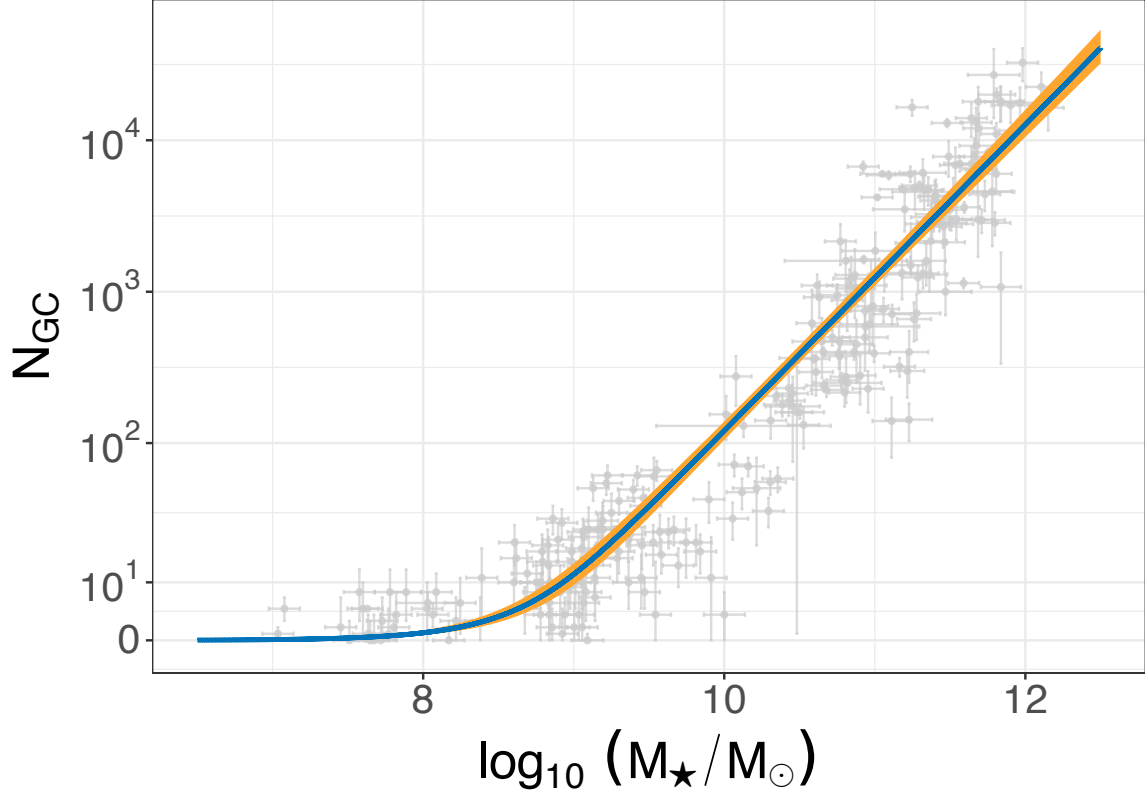


Figure 1. Negative-binomial regression with a latent errors-in-variables treatment for the relation between N_{GC} and $\log_{10}(M_{\star}/M_{\odot})$. Data points are elliptical galaxies from Harris et al. (2013). The solid blue curve shows the posterior median, and the orange band the 95% credible interval. Points include 1σ uncertainties.

- 76 Berek, S. C., Eadie, G. M., Speagle, J. S., & Harris, W. E.
77 2023, *ApJ*, 955, 22, doi: [10.3847/1538-4357/ace7b7](https://doi.org/10.3847/1538-4357/ace7b7)
78 Berek, S. C., Eadie, G. M., Speagle, J. S., & Wang, S. Y.
79 2024, *ApJ*, 972, 104, doi: [10.3847/1538-4357/ad6147](https://doi.org/10.3847/1538-4357/ad6147)
80 Blakeslee, J. P. 1997, *ApJL*, 481, L59, doi: [10.1086/310653](https://doi.org/10.1086/310653)
81 Canossa-Gosteinski, M. A., Chies-Santos, A. L., Furlanetto,
82 C., et al. 2024, *MNRAS*, 534, 1729,
83 doi: [10.1093/mnras/stae2161](https://doi.org/10.1093/mnras/stae2161)
84 de Souza, R. S., Hilbe, J. M., Buelens, B., et al. 2015a,
85 *MNRAS*, 453, 1928, doi: [10.1093/mnras/stv1825](https://doi.org/10.1093/mnras/stv1825)
86 de Souza, R. S., Cameron, E., Killedar, M., et al. 2015b,
87 *Astronomy and Computing*, 12, 21,
88 doi: [10.1016/j.ascom.2015.04.002](https://doi.org/10.1016/j.ascom.2015.04.002)
89 de Valpine, P., Turek, D., Paciorek, C. J., et al. 2017,
90 *Journal of Computational and Graphical Statistics*, 26,
91 403, doi: [10.1080/10618600.2016.1172487](https://doi.org/10.1080/10618600.2016.1172487)
92 Dornan, V., & Harris, W. E. 2025, *ApJ*, 988, 70,
93 doi: [10.3847/1538-4357/ade05e](https://doi.org/10.3847/1538-4357/ade05e)
94 Elliott, J., de Souza, R. S., Krone-Martins, A., et al. 2015,
95 *Astronomy and Computing*, 10, 61,
96 doi: [10.1016/j.ascom.2015.01.002](https://doi.org/10.1016/j.ascom.2015.01.002)
97 Harris, G. L. H., Poole, G. B., & Harris, W. E. 2014,
98 *MNRAS*, 438, 2117, doi: [10.1093/mnras/stt2337](https://doi.org/10.1093/mnras/stt2337)
99 Harris, W. E., Harris, G. L. H., & Alessi, M. 2013, *ApJ*,
100 772, 82, doi: [10.1088/0004-637X/772/2/82](https://doi.org/10.1088/0004-637X/772/2/82)
101 Hattab, M. W., de Souza, R. S., Ciardi, B., et al. 2019,
102 *MNRAS*, 483, 3307, doi: [10.1093/mnras/sty3314](https://doi.org/10.1093/mnras/sty3314)
103 Hilbe, J. M., de Souza, R. S., & Ishida, E. E. O. 2017,
104 *Bayesian Models for Astrophysical Data Using R, JAGS,*
105 *Python, and Stan*, doi: [10.1017/CBO9781316459515](https://doi.org/10.1017/CBO9781316459515)
106 Hudson, M. J., Harris, G. L., & Harris, W. E. 2014, *ApJL*,
107 787, L5, doi: [10.1088/2041-8205/787/1/L5](https://doi.org/10.1088/2041-8205/787/1/L5)
108 Saifollahi, T., Zaritsky, D., Trujillo, I., et al. 2022,
109 *MNRAS*, 511, 4633, doi: [10.1093/mnras/stac328](https://doi.org/10.1093/mnras/stac328)
110 Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, *PeerJ*
111 *Computer Science*, 2, e55, doi: [10.7717/peerj-cs.55](https://doi.org/10.7717/peerj-cs.55)
112 Valenzuela, L. M., Moster, B. P., Remus, R.-S., O'Leary,
113 J. A., & Burkert, A. 2021, *MNRAS*, 505, 5815,
114 doi: [10.1093/mnras/stab1701](https://doi.org/10.1093/mnras/stab1701)
115 Zaritsky, D. 2022, *MNRAS*, 513, 2609,
116 doi: [10.1093/mnras/stac1072](https://doi.org/10.1093/mnras/stac1072)