IMS Lecture Notes–Monograph Series
Volume Title
Vol. 0 (0000) 1–8

DOI: 0000/0000.0000

Statistical Models for Globular Cluster Luminosity Distribution

Max-Louis G. Buot* **and Donald St. P. Richards**[†,1]

Xavier University and Penn State University

Abstract: We consider statistical models which have been proposed for luminosity distributions for the globular clusters in the Milky Way and M31. Although earlier research showed that the cluster luminosity functions in those two galaxies were well fit by Gaussian distributions, subsequent investigations suggested that their luminosities were better fit by t-, rather than Gaussian, distributions. By applying the Bayesian Information Criterion, we do not find overwhelming statistical evidence that the t-distribution is superior to the Gaussian distribution as a model of luminosity distribution for the Milky Way. In the case of M31, we find moderate evidence that the Gaussian distribution is superior to the t-distribution. In neither case do we find strong evidence to support the use of one distribution over the other as a statistical model for the luminosities of the globular clusters in the Milky Way and M31. Consequently, we recommend that the Gaussian be retained as the statistical model for luminosity distribution. Moreover, we urge caution in the use of the Kolmogorov-Smirnov statistic to justify the choice of statistical models for globular cluster luminosity functions.

1. Introduction

Globular clusters are some of the oldest objects in the Universe, are innately luminous, and there is strong evidence that they are formed during periods of major star formation (Larsen and Richtler [9]). For these and other reasons, globular clusters have played an important role in research on the formation of galaxies (Harris [6]; van den Bergh [16]).

The study of the globular cluster luminosity function (GCLF) of a galaxy is motivated by the distribution of luminosities of the globular clusters within the galaxy. Precisely, the GCLF of a galaxy is the relative number of its globular clusters at a given luminosity. In the Milky Way, empirical evidence suggests that the corresponding GCLF is usually unimodal and nearly symmetric, and the peak of the luminosity function occurs at a magnitude which varies little from galaxy to galaxy (Harris [5]). For these reasons, it has been found that the GCLF peak in the Milky Way is sometimes appropriate as a standard candle for distance measurement (Whitmore [17]).

In early work on luminosity functions, much attention was paid to plausible analytical forms of the GCLF. In several instances, including globular clusters in

*Department of Mathematics & Computer Science, Xavier University, Cincinnati, Ohio 45207, U.S.A.; e-mail: buotm@xavier.edu.

†Department of Statistics, The Pennsylvania State University, University Park, PA 16802, U.S.A.; e-mail: richards@stat.psu.edu.

[1]Supported in part by the National Science Foundation, grant DMS-0705210.

AMS 2000 subject classifications: Primary 85A35, 62G07; Secondary 62G10.

Keywords and phrases: Bayesian Information Criterion, Galaxies, The Galaxy, Gaussian distribution, Globular cluster systems, Kolmogorov-Smirnov statistic, Luminosity distribution, M31, The Milky Way, Maximum likelihood estimation, Nonparametric statistics, Schwarz criterion, t-distribution.

the Milky Way and in M31, it was shown that a Gaussian distribution was a good analytical fit to the empirically observed distribution of luminosities (Racine and Shara [11]; van den Bergh [15]; Harris, *et al.* [4]). A subsequent analysis (Secker [14]) argued that a t-distribution provided a better fit, and that claim has led to wide acceptance of the t-distribution in subsequent research on luminosity functions (Barmby, *et al.* [2]).

As measured by the maximum likelihood procedure, the t-distribution has been shown to fit the empirical data more closely than the Gaussian distribution (Secker [14]). However, this comparison is complicated by the fact that the analytical form of the Gaussian distribution is based on only two parameters, the mean and standard deviation, whereas the t-distribution has an additional index, or shape, parameter. This raises the issue of whether or not the increase in the likelihood function for the t-distribution over the Gaussian may be caused by the presence of an additional parameter.

The issue of the goodness-of-fit of a hypothesized distribution *vis-a-vis* the number of parameters in the underlying analytical form of that distribution is precisely the *raison d'etre* of the Bayesian Information Criterion (BIC) or Schwarz criterion [13]. Simply put, the BIC is designed to ascertain the extent to which an improved fit is due to an increase in the number of parameters in the analytical form of the distribution.

In the sequel, we apply the BIC to the data provided by Secker [14]. We do not find strong statistical evidence in support of the t-distribution over the Gaussian distribution, or *vice versa*, as a statistical model for the luminosity distribution of globular clusters in the Milky Way or M31. Consequently, given the fewer number of parameters in the Gaussian distribution and the many well-known attractive features of that distribution, we recommend that the Gaussian be utilized as a statistical model for luminosity distribution. Moreover, we urge caution in the use of the classical Kolmogorov-Smirnov statistic as justification for the choice of statistical models for globular cluster luminosity distributions.

2. The Bayesian Information Criterion

Suppose that two statistical distributions are plausible models for an observed data set. These models may be fit to the data using a variety of statistical procedures, e.g., residual sums of squares, the method of moments, or the method of maximum likelihood.

In polynomial regression, for example, the residual sum of squares can be reduced simply by increasing the degree of the polynomial regression function. In general, the more complex the mathematical form of a model, the better a model will be seen to fit the data. Therefore, it is clear that the choice of a statistical model should not be assessed entirely by measures such as residual sums of squares or likelihood function values for, by increasing the number of parameters in the hypothesized model, we can obtain a relentless reduction in the residual sums of squares or an increase in the values of the likelihood function.

The Bayesian Information Criterion (BIC) constitutes a standard approach to assessing the relative plausibility of two competing statistical models which are being fit to data with *large* sample sizes (Schwarz [13]). To balance any difference in the number of parameters between two statistical models, the BIC penalizes a model which has a larger number of free parameters. To illustrate this approach, consider the situation of two competing models, $f_1(x; \theta_1, \ldots, \theta_{m_1})$ and

$f_2(x;\phi_1,\ldots,\phi_{m_2})$. Here, $\theta_1,\ldots,\theta_{m_1}$ and $\phi_1,\ldots,\phi_{m_2}$ are parameters of the corresponding density functions f_1 and f_2, respectively. On being given a random sample $X_1,\ldots,X_n$, we construct the likelihood functions

$$L_1(\theta_1,\ldots,\theta_{m_1}) = \prod_{i=1}^{n} f_1(x_i;\theta_1,\ldots,\theta_{m_1})$$

and

$$L_2(\phi_1,\ldots,\phi_{m_2}) = \prod_{i=1}^{n} f_2(x_i;\phi_1,\ldots,\phi_{m_2}),$$

and the corresponding BIC are, respectively,

$$\mathrm{BIC}_1 = -2\ln L_1(\theta_1,\ldots,\theta_{m_1}) + m_1 \ln n$$

and

$$\mathrm{BIC}_2 = -2\ln L_2(\phi_1,\ldots,\phi_{m_2})] + m_2 \ln n.$$

Given the parameters $\theta_1,\ldots,\theta_{m_1}$ and $\phi_1,\ldots,\phi_{m_2}$, and explicit formulas for the density functions f_1 and f_2, the relative superiority of the model $f_1(x;\theta_1,\ldots,\theta_{m_1})$ over the model $f_2(x;\phi_1,\ldots,\phi_{m_2})$ is measured by the difference in BICs, *viz.*,

$$\begin{aligned}\Delta_{\mathrm{BIC}} &= \mathrm{BIC}_2 - \mathrm{BIC}_1 \\ &= 2[\ln L_1(\theta_1,\ldots,\theta_{m_1}) - \ln L_2(\phi_1,\ldots,\phi_{m_2})] - (m_1 - m_2)\ln n.\end{aligned}$$

The first term in this expression is a measure of the increase in the likelihood function values of the first model over the second, and the second term is a penalty term reflecting the difference in the numbers of parameters in the models. Thus, Δ_{BIC} assesses any increase in the likelihood in light of the additional number of parameters necessary to achieve such an increase.

In practice, the values of $\theta_1,\ldots,\theta_{m_1}$ and $\phi_1,\ldots,\phi_{m_2}$ are unknown and need to be estimated from the data. Thus, we calculate the corresponding maximum likelihood estimates $\widehat{\theta}_1,\ldots,\widehat{\theta}_{m_1}$ and $\widehat{\phi}_1,\ldots,\widehat{\phi}_{m_2}$ and use those results to calculate

$$\widehat{\mathrm{BIC}}_1 = -2\ln L_1(\widehat{\theta}_1,\ldots,\widehat{\theta}_{m_1}) + m_1 \ln n$$

and

$$\widehat{\mathrm{BIC}}_2 = -2\ln L_1(\widehat{\phi}_1,\ldots,\widehat{\phi}_{m_2}) + m_2 \ln n,$$

the corresponding maximum likelihood estimates of BIC_1 and BIC_2, respectively. Finally, we calculate

$$\begin{aligned}\widehat{\Delta}_{\mathrm{BIC}} &= \widehat{\mathrm{BIC}}_2 - \widehat{\mathrm{BIC}}_1 \\ &= 2[\ln L_1(\hat{\theta}_1,\ldots,\hat{\theta}_{m_1}) - \ln L_2(\hat{\phi}_1,\ldots,\hat{\phi}_{m_2})] - (m_1 - m_2)\ln n,\end{aligned} \tag{2.1}$$

the maximum likelihood estimate of Δ_{BIC}.

General rules for using $\widehat{\Delta}_{\mathrm{BIC}}$ to assess the relative goodness-of-fit of the models f_1 and f_2 are as follows (Jeffreys [7, Appendix B]; Kass and Raftery [8]; Mukherjee, *et al.* [10]):

- $0 \le \widehat{\Delta}_{\mathrm{BIC}} < 2$: Weak evidence that Model 1 is superior to Model 2
- $2 \le \widehat{\Delta}_{\mathrm{BIC}} \le 6$: Moderate evidence that Model 1 is superior to Model 2
- $6 < \widehat{\Delta}_{\mathrm{BIC}} \le 10$: Strong evidence that Model 1 is superior to Model 2
- $\widehat{\Delta}_{\mathrm{BIC}} > 10$: Very strong evidence that Model 1 is superior to Model 2

For the case in which $\widehat{\Delta}_{\text{BIC}} < 0$, one simply reverses the roles of L_1 and L_2 in order to apply these general rules.

We now apply these procedures to compare three statistical models for GCLF data.

3. Models for GCLF distribution in the Milky Way and in M31

Consider the following competing models for GCLF in the Milky Way: A Gaussian model (van den Bergh [15]),

$$f_1(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\,\exp\Big[-\frac{(x-\mu)^2}{2\sigma^2}\Big], \tag{3.1}$$

and a t-distribution model (Secker [14]),

$$f_2(x;\mu,\sigma,\delta) = \frac{\Gamma\big((\delta+1)/2\big)}{\sqrt{\pi\delta}\,\sigma\,\Gamma(\delta/2)}\Big[1+\frac{(x-\mu)^2}{\delta\sigma^2}\Big]^{-(\delta+1)/2}, \tag{3.2}$$

where, in each model, $-\infty < x < \infty$ and the permissible ranges of the parameters are $-\infty < \mu < \infty$, $\sigma > 0$, and $\delta > 0$. In each model, μ represents the population mean; σ is a measure of variability; and, in the case of the t-distribution model, δ is a shape parameter.

Under the Gaussian model (3.1), the likelihood function corresponding to a random sample $X_1,\ldots,X_n$ is

$$\begin{aligned} L_1(\mu,\sigma) &= \prod_{i=1}^n f(X_i;\mu,\sigma) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Big[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\Big]. \end{aligned}$$

It is well-known that the maximum likelihood estimator of the parameter μ is $\widehat{\mu} = \bar{X}$, the sample mean. Further, denoting by S the sample standard deviation, it is well-known that the maximum likelihood estimator of σ^2 is $\widehat{\sigma}^2 = (n-1)S^2/n$. Because $(n-1)/n \approx 1$ for large values of n and because the estimator S^2 has numerous desirable statistical properties, including unbiasedness, it is common practice to estimate σ^2 by S^2.

Using the 100 observations comprising the Milky Way data from [14, Table 1], it is found [14, p. 1475] that the likelihood function L_1 for the Gaussian model is maximized at $(\widehat{\mu},\widehat{\sigma}) = (-7.14, 1.41)$, and then we obtain [14, p. 1476]

$$\ln L_1(-7.14, 1.41) = -176.4, \tag{3.3}$$

consequently, $\widehat{\text{BIC}}_1 = 362.0$.

In the case of the t-distribution model (3.2), the likelihood function is

$$\begin{aligned} L_2(\mu,\sigma,\delta) &= \prod_{i=1}^n f_2(X_i;\mu,\sigma,\delta) \\ &= \prod_{i=1}^n \frac{\Gamma(\frac{\delta+1}{2})}{\sqrt{\pi\delta}\,\sigma\,\Gamma(\frac{\delta}{2})}\Big[1+\frac{(X_i-\mu)^2}{\delta\sigma^2}\Big]^{-(\delta+1)/2}. \end{aligned}$$

Unlike the Gaussian case, no algebraic formulas for $\widehat{\mu}$, $\widehat{\sigma}$, or $\widehat{\delta}$ are available for this model, and therefore the maximization of L_2 and the corresponding likelihood estimates are obtained numerically. Calculating again from the Milky Way data [14, p. 1476] leads to the maximum likelihood estimates $(\widehat{\mu}, \widehat{\sigma}, \widehat{\delta}) = (-7.31, 1.03, 3.55)$.[1] Then, the corresponding value of L_2 is given by

$$\ln L_2(-7.31, 1.03, 3.55) = -173.0, \tag{3.4}$$

consequently, $\widehat{\mathrm{BIC}}_2 = 359.8$. The maximum likelihood calculations in (3.3) and (3.4) suggest that the Gaussian model (3.1) is inferior to the t-model as a distribution for GCLF data in the Milky Way. However, there remains the issue of whether the increase in likelihood in (3.4) over (3.3) is due to the larger number of parameters in (3.2).

On applying the formula for $\widehat{\Delta}_{\mathrm{BIC}}$ in (2.1) with the values derived from (3.3) and (3.4), we obtain

$$\widehat{\Delta}_{\mathrm{BIC}} = \widehat{\mathrm{BIC}}_2 - \widehat{\mathrm{BIC}}_2 = -2.2\ .$$

By the general rules on application of $\widehat{\Delta}_{\mathrm{BIC}}$ for model comparison, we have only moderate evidence that the t-distribution is superior to the Gaussian distribution as a model for the Galactic data. In particular, we do not have overwhelming evidence in support of one model over the other.

In the case of 82 globular clusters from M31 [14, Table 2], similar calculations [14, p. 1478] for the Gaussian model lead to the maximum likelihood estimates $(\widehat{\mu}, \widehat{\sigma}) = (16.98, 0.99)$ and the corresponding likelihood function value is given by

$$\ln L_1(16.98, 0.99) = -115.4, \tag{3.5}$$

hence, $\widehat{\mathrm{BIC}}_1 = 239.6$. In the case of the t- model, the parameters are found [14, p. 1478] to have maximum likelihood estimates $(\widehat{\mu}, \widehat{\sigma}, \widehat{\delta}) = (17.0, 0.90, 11.02)$,[2] and the corresponding likelihood function value is

$$\ln L_2(17.0, 0.90, 11.02) = -115.1, \tag{3.6}$$

hence, $\widehat{\mathrm{BIC}}_2 = 243.4$. As before, the maximum likelihood calculations in (3.5) and (3.6) suggest that the Gaussian model (3.1) is inferior to the t-model as a distribution for the GCLF data in M31.

On applying the formula for $\widehat{\Delta}_{\mathrm{BIC}}$ in (2.1) with the values derived from (3.5) and (3.6), we obtain

$$\widehat{\Delta}_{\mathrm{BIC}} = \widehat{\mathrm{BIC}}_2 - \widehat{\mathrm{BIC}}_2 = 3.8\ .$$

Consequently, we have moderate evidence that the Gaussian distribution is superior to the t-distribution as a statistical model for the GCLF of M31. Again, we find no overwhelming evidence in support of one model over the other.

To infer whether the observed values of $\widehat{\Delta}_{\mathrm{BIC}}$, *viz.*, -2.2 (in the case of the Milky Way data) and 3.8 (in the case of the M31 data), are significantly small, we estimated the standard error of $\widehat{\Delta}_{\mathrm{BIC}}$ using bootstrap methods, with all calculations

[1] We remark that optimization routines in the statistical package R [12] returned the estimates $(\widehat{\mu}, \widehat{\sigma}, \widehat{\delta}) = (-7.30, 1.07, 4.28)$, which are different from those reported in [14, *loc. cit.*]. However, to afford direct comparison with the results of [14], and because our overall conclusions are the same for both reported estimates, we opt to work with the values given in [14].

[2] In this case, optimization routines in the package R [12] returned the estimates $(\widehat{\mu}, \widehat{\sigma}, \widehat{\delta}) = (16.99, 0.91, 12.53)$.

being done with the statistical package R [12]. In each instance, we simulated from the respective models a set of observations equal in number to the sample sizes of the actual data sets in [14]. From those observations, we then computed the corresponding values of $\widehat{\Delta}_{\mathrm{BIC}}$, and then calculated the estimated standard error of $\widehat{\Delta}_{\mathrm{BIC}}$ from its simulated values based on $20,000$ bootstrap samples.

To fit a parametric bootstrap, we used the t-distribution to model the Milky Way data. The estimated standard error of $\widehat{\Delta}_{\mathrm{BIC}}$ is 21.9, which is extremely large. Hence, the conclusion that the observed value of -2.2 indicates moderate evidence in favor of the t- over the Gaussian distribution should be accepted cautiously.

In the case of the M31 data, we used the Gaussian distribution to model the data. The estimated standard error of $\widehat{\Delta}_{\mathrm{BIC}}$ is 0.9, a small value. Consequently, we infer moderate evidence in favor of the Gaussian over the t-distribution.

A Cauchy model was also examined [14, p. 1474] as a possible statistical model for GCLF in the Milky Way and M31. In this case, the analytical form of the density function is

$$f_3(x;\mu,\sigma) = \frac{\sigma}{\pi[\sigma^2 + (x-\mu)^2]} \ , \tag{3.7}$$

where $-\infty < x < \infty$, and the parameter ranges are $-\infty < \mu < \infty$ and $\sigma > 0$. The parameter μ represents the median and σ a measure of the spread of the distribution. Here again, we apply maximum likelihood calculations [14, pp. 1476–1478] to compare the Cauchy with the Gaussian or t-distribution models. When the Gaussian or t- models are compared to the Cauchy model as a fit to the Milky Way GCLF data, we obtain $\widehat{\Delta}_{\mathrm{BIC}} > 20$ in both cases. When the Gaussian or t- models are compared to the Cauchy model for the M31 data, we obtain $\widehat{\Delta}_{\mathrm{BIC}} = 17.4$ in the Gaussian case and 24.2 in the t- case. In all instances, there is strong or very strong evidence that the Gaussian and t- distributions each are superior to the Cauchy distribution as a fit to the GCLF data from the Milky Way or from M31.

In comparing the Gaussian or the t- distributions to the Cauchy, we again calculated bootstrap estimates of the standard errors of the statistic $\widehat{\Delta}_{\mathrm{BIC}}$. To construct bootstrap samples, we utilized the same simulation procedures as before, sampling from the Gaussian model because of the prior evidence that it is a superior fit to the data. The estimated standard errors for the data sets are 7.0 and 6.4 for the Milky Way and M31 data, respectively, so we have even stronger evidence in favor of the Gaussian over the Cauchy distribution as a fit for either data set.

Similarly, in comparing the t- model to the Cauchy in both data sets, we calculated bootstrap estimates of the standard errors of $\widehat{\Delta}_{\mathrm{BIC}}$, using the t-distribution as the model. In this case, the estimated standard errors of $\widehat{\Delta}_{\mathrm{BIC}}$ for the data sets are 6.4 and 5.9 for the Milky Way and M31 data, respectively. Consequently, we have further evidence that the t-distribution is a better fit for the data than the Cauchy distribution.

As regards the fitting of more complex models, with higher numbers of parameters, to GCLF data, it is now clear that such models must be compared to the Gaussian model not only on the basis of the likelihood function but also with regard to the values of the BIC. As a final comment in support of the simpler Gaussian model, we point out that Harris [6, p. 293] also commented that although models with greater numbers of free parameters will fit the data more accurately, such models do not appear to provide "any immediate insight about the astrophysical processes governing the cluster luminosities and masses ..."

4. Concluding Remarks

The BIC is only one of many procedures for testing the goodness-of-fit of a statistical model and, as with any procedure, it should be used carefully. In particular, it is not a panacea and, in general, should be applied in conjunction with other information criteria.

There are also drawbacks with the BIC. Findley [3] notes that under specific circumstances, the BIC will tend to select the model which has fewer parameters. Precisely, Findley proves that "if the log-likelihood-ratio sequence of two models with different numbers of estimated parameters is bounded in probability then the BIC will, with asymptotic probability 1, select the model having fewer parameters." At least heuristically, this result can be deduced *via* (2.1), as follows: Suppose that $m_1 < m_2$ and suppose also that the first term in (2.1) is bounded above and below by universal constants, i.e., constants which do not depend on n. Then $\widehat{\Delta}_{\text{BIC}}$ in (2.1) is dominated by $(m_2 - m_1)\ln n$ as n increases; consequently, $\widehat{\Delta}_{\text{BIC}}$ will become positive. By the general rules for application of the BIC, we would then infer strong evidence in favor of $f_1(x;\theta_1,\ldots,\theta_{m_1})$, the model having the fewer number of parameters.

In the context of astrophysical applications of comparing two statistical models, $f_1(x;\theta_1,\ldots,\theta_{m_1})$ and $f_2(x;\phi_1,\ldots,\phi_{m_2})$, for globular cluster luminosity functions, an application of Findley's theorem requires a sequential calculation of the corresponding log-likelihood ratios

$$\ln \frac{L_1(\hat{\theta}_1,\ldots,\hat{\theta}_{m_1})}{L_2(\hat{\phi}_1,\ldots,\hat{\phi}_{m_2})}$$

for successively increasing sample sizes n. If it is *believed* that the observed sequence of log-likelihood ratios will be bounded between two universal constants, and such a belief necessarily will have to be justified on intrinsic astrophysical arguments, then it be necessary to look for alternatives to the BIC in measuring the relative plausibility of a statistical model.

In fitting models to GCLF data, it is important to bear in mind that the BIC is *consistent*: As $n \to \infty$, the probability that the BIC will determine the correct statistical model converges to 1 (Jeffreys [7]). The property of consistency does not hold, in general, for model-fitting information criteria.

There is also the issue of utilizing the Kolmogorov-Smirnov statistic to measure the goodness-of-fit of a statistical model. In the case of the Milky Way and M31 data considered in this paper, the Kolmogorov-Smirnov statistic was also applied [14, p. 1467, *ff.*] to support the choice of the t-distribution over the Gaussian and Cauchy models. Here again, though, there is cause for concern. The importance of statistics based on empirical distribution functions, such as the Kolmogorov-Smirnov statistic, stems from the fact that they are distribution-free in the case of continuous data, such as luminosity measurements. However, *these statistics are no longer distribution-free when the model parameters need to be estimated from the data*. Consequently, in model fitting contexts, goodness-of-fit levels of significance derived from the Kolmogorov-Smirnov statistic are usually incorrect when applied with estimated parameters [1].

We urge that caution be exercised in using the Kolmogorov-Smirnov statistic to justify the fitting of the models (3.1), (3.2), or (3.7) to GCLF data. The basic problem with the Kolmogorov-Smirnov statistic is that it is not penalized; it is possible that a penalized form of that statistic may work well, and indeed a main point of this paper is that penalized statistics generally are better for fitting models to luminosity data. Moreover, any parameter estimation in a penalized Kolmogorov-Smirnov statistic can be dealt with by bootstrap methods.

Acknowledgments

We are grateful to Tom Hettmansperger and Mercedes Richards for their comments on earlier drafts of the manuscript.

References

[1] Babu, G. J., and Rao, C. R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhyā*, **66**, 63–74.
[2] Barmby, P., Huchra, J. P., and Brodie J. P. (2001). The M31 globular cluster luminosity function. *Astronom. J.*, **121**, 1482–1496.
[3] Findley, D. F. (1991). Counterexamples to parsimony and BIC. *Ann. Inst. Statist. Math.*, **43**, 505–514.
[4] Harris, W. E., Allwright, J. W. B., Pritchet, C. J., and van den Bergh, S. (1991). The luminosity distribution of globular clusters in three giant Virgo ellipticals. *Astrophys. J. Suppl.*, **76**, 115–151.
[5] Harris, W. E. (1998). Globular cluster systems as distance indicators. In: ASP Conf. Ser. 4, *The Extragalactic Distance Scale*, ed. C. Pritchet and S. van den Bergh (San Francisco, ASP), 231–254.
[6] Harris, W. E. (2000). Globular cluster systems. *Star Clusters, Saas-Fee Advanced Course 28*. Lecture Notes 1998, Swiss Society for Astrophysics and Astronomy (Berlin, Springer-Verlag).
[7] Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. (Oxford, Clarendon Press).
[8] Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.*, **90**, 773–795.
[9] Larsen, S. S., and Richtler, T. (1999). Young massive star clusters in nearby galaxies . I. Identification and general properties of the cluster systems. *Astron. & Astrophys.*, **345**, 59–72.
[10] Mukherjee, S., Feigelson, E. D., Babu, G. J., Murtagh, F., Fraley, C., and Raftery, A. (1998). Three types of gamma-ray bursts. *Astrophys. J.*, **508**, 314–327.
[11] Racine, R., and Shara, M. (1979). The luminosity distribution of globular clusters in M31. *Astronom. J.*, **84**, 1694–1696.
[12] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005.
[13] Schwarz, G. E. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
[14] Secker, J. (1992). A statistical investigation into the shape of the globular cluster luminosity distribution. *Astronom. J.*, **104**, 1472–1481.
[15] van den Bergh, S. (1985). The luminosity function of globular clusters. *Astrophys. J.*, **297**, 361–364.
[16] van den Bergh, S. (2000). Some musings on globular cluster systems. *Publ. Astronom. Soc. Pacific*, **112**, 932–941.
[17] Whitmore, B. C. (1997). Globular clusters as distance indicators. In: it The Extragalactic Distance Scale, ed. M. Livio, M. Donahue, & N. Panagia. (Cambridge, Cambridge Univ. Press), 254–272.