Statistics for Twenty-first Century Astrometry

2000 Heinrich K. Eichhorn Memorial Lecture

William H. Jefferys (bill@astro.as.utexas.edu)
University of Texas at Austin, Austin, TX USA

Abstract. H.K. Eichhorn had a lively interest in statistics during his entire scientific career, and made a number of significant contributions to the statistical treatment of astrometric problems. In the past decade, a strong movement has taken place for the reintroduction of Bayesian methods of statistics into astronomy, driven by new understandings of the power of these methods as well as by the adoption of computationally-intensive simulation methods to the practical solution of Bayesian problems. In this paper I will discuss how Bayesian methods may be applied to the statistical discussion of astrometric data, with special reference to several problems that were of interest to Eichhorn.

Keywords: Eichhorn, astrometry, Bayesian statistics

1. Introduction

Bayesian methods offer many advantages for astronomical research and have attracted much recent interest. The Astronomy and Astrophysics Abstracts website (http://adsabs.harvard.edu/) lists 117 articles with the keywords 'Bayes' or 'Bayesian' in the past 5 years, and the number is increasing rapidly (there were 33 articles in 1999 alone). At the June, 1999 meeting of the American Astronomical Society, held in Chicago, there was a special session on Bayesian and Related Likelihood Techniques. Another session at the June, 2000 meeting also featured Bayesian methods. A good introduction to Bayesian methods in astronomy can be found in Loredo (1990).

Bayesian methods have many advantages over frequentist methods, including the following: it is simple to incorporate prior physical or statistical information into the analysis; the results depend only on what has actually been observed and not on observations that might have been made but were not; it is straightforward to compare models and average over both nested and unnested models; and the interpretation of the results is very natural, especially for physical scientists.

Bayesian inference is a systematic way of approaching statistical problems, rather than a collection of ad hoc techniques. Very complex problems (difficult or impossible to handle classically) are straightforwardly analyzed within a Bayesian framework. Bayesian analysis is coherent: we will not find ourselves in a situation where the analysis tells us that two contradictory things are simultaneously likely to be true.

With proposed astrometric missions (e.g., FAME) where the signal can be very weak, analyses based on normal approximations may not be adequate. In such situations, Bayesian analysis that explicitly assumes the Poisson nature of the data may be a better choice than a normal approximation.

2. Outline of Bayesian Procedure

In a nutshell, Bayesian analysis entails the following systematic steps: (1) Choose prior distributions (priors) that reflect your knowledge about each parameter and model prior to looking at the data. (2) Determine the likelihood function of the data under each model and parameter value. (3) Compute and normalize the full posterior distribution, conditioned on the data, using Bayes' theorem. (4) Derive summaries of quantities of interest from the full posterior distribution by integrating over the posterior distribution to produce marginal distributions or integrals of interest (e.g., means, variances).

2.1. Priors

The first ingredient of the Bayesian recipe is the prior distribution. Eichhorn was acutely aware of the need to use *all* available information when reducing data, and often criticized the common practice of throwing away useful information either explicitly or by the use of suboptimal procedures. The Bayesian way of preventing this is to use priors properly. The investigator is *required* to provide *all* relevant prior information that he has before proceeding with the analysis. Moreover, there is always prior information. For example, we cannot count a negative number of photons, so in photon-counting situations that may be presumed as known. Parallaxes are greater than zero. We now know that the most likely value of the Hubble constant is in the ballpark of 60-80 km/sec/mpc, with smaller probabilities of its being higher or lower. Prior information can be statistical in nature, e.g., we may have statistical knowledge about the spatial or velocity distribution of stars, or the variation in a telescope's plate scale.

In Bayesian analysis, our knowledge about a parameter θ is encoded by a prior probability distribution on the parameter, e.g., $p(\theta \mid B)$, where B is background information. Where prior information is vague or uninformative, a vague prior generally recovers results similar to a classical analysis. However, in model selection and model averaging situations, Bayesian analysis usually gives quite different results, being more conservative about introducing new parameters than is typical of frequentist approaches.

Sensitive dependence of the result on reasonable variations in prior information should be tested, and if present indicates that no analysis, Bayesian or other, can give reliable results. Since frequentist analyses do not use priors and therefore are incapable of sounding such a warning, this can be considered a strength of the Bayesian approach.

The problem of prior information of a statistical or probabilistic nature was addressed in a classical framework by Eichhorn (1978) and by Eichhorn and Standish (1981). They considered adjusting astrometric data given prior knowledge about some of the parameters in the problem, e.g., that the plate scale values only varied within a certain dispersion. For the cases studied in these papers (multivariate normal distributions), the result is similar to the Bayesian one, although the interpretation is different.

In another example, Eichhorn and Smith (1996) studied the Lutz-Kelker bias. The classical way to understand the Lutz-Kelker bias is that it is more likely that we have observed a star slightly farther away with a negative error that brings it closer in to the observed distance, than that we have observed a slightly nearer star with a positive error that pushes it out to the observed distance, because the number of stars increases with increasing distance. The Bayesian notes that it is more likely a priori that a star of unknown distance is farther away than that it is nearer, which dictates the use of a prior that increases with distance. The mathematical analysis gives a similar result, but the Bayesian approach, by demanding at the outset that we think about prior information, inevitably leads us to consider this phenomenon, which classical astrometrists missed for a century.

2.2. The Likelihood Function

The likelihood function $\mathcal{L}$ is the second ingredient in the Bayesian recipe. It describes the statistical properties of the mathematical model of our problem. It tells us how the statistics of the observations (e.g., normal or Poisson data) are related to the parameters and to any background information. It is proportional to the sampling distribution for observing the data Y, given the parameters, but we are interested in its functional dependence on the parameters:

$$\mathcal{L}(\theta; Y, B) \propto p(Y \mid \theta, B)$$

The likelihood is known up to a constant but arbitrary factor which cancels out in the analysis.

Like Bayesian estimation, maximum likelihood estimation (upon which Eichhorn based many of his papers) is founded upon using the likelihood function. This is good, because the likelihood function is always a sufficient statistic for the parameters of the problem. Furthermore, according to the important *Likelihood Principle* (Berger, 1985), it can be shown that under very general and natural conditions, the likelihood function contains all of the information in the data that can be used for inference. However, the likelihood is not the whole story. Maximum likelihood by itself does not take prior information into account, and it fails badly in some notorious situations, like errors-in-variables problems (i.e., both x and y have error), when the variance of the observations is estimated. Bayesian analysis gets the right answer in this case; classical analysis relies on a purely *ad hoc* factor of 2 correction. A purely likelihood approach presents other problems as well.

2.3. Posterior Distribution

The third part of the Bayesian recipe is to use Bayes' theorem to calculate the posterior distribution. The posterior distribution encodes what we know about the parameters and model *after* we observe the data. Thus, Bayesian analysis models a process of learning from experience.

Bayes' theorem says that

$$p(\theta \mid Y, B) = \frac{p(Y \mid \theta, B) p(\theta \mid B)}{p(Y \mid B)} \tag{1}$$

It is a trivial result of probability theory. The denominator

$$p(Y \mid B) = \int p(Y \mid \theta, B) p(\theta \mid B) d\theta \tag{2}$$

is just a normalization factor and can often be dispensed with.

The posterior distribution after observing data Y can be used as the prior distribution for new data Z, which makes it easy to incorporate new data into an analysis based on earlier data. It can be shown that any coherent model of learning is equivalent to Bayesian learning. Thus in Bayesian analysis, results take into account all known information, do not depend on the order in which the data (e.g, Y and Z) are obtained, and are consistent with common sense inductive reasoning as well as with standard deductive logic. For example, if A entails B, then observing B should support A (inductively), and observing $\neg B$ should refute A (logically).

2.4. Summarizing Results

The fourth and final step in our Bayesian recipe is to use the posterior distribution we have calculated to give us summary information about the quantities we are interested in. This is done by integrating over the posterior distribution to produce marginal distributions or integrals of interest (e.g., means, variances). Bayesian methodology provides a simple and systematic way of handling nuisance parameters required by the analysis but which are of no interest to us. We simply integrate them out (marginalize them) to obtain the marginal distribution of the parameter(s) of interest:

$$p(\theta_1 \mid Y, B) = \int p(\theta_1, \theta_2 \mid Y, B) d\theta_2 \tag{3}$$

Likewise, computing summary statistics is simple. For example, posterior means and variances can be calculated straightforwardly:

$$\bar{\theta}_1 \mid Y, B = \int \theta_1 p(\theta_1 \mid Y, B) d\theta_1 \tag{4}$$

3. Model Selection and Model Averaging

Eichhorn and Williams (1963) studied the problem of choosing between competing astrometric models. Often the models are empirical, e.g., polynomial expansions in the coordinates. The problem is to avoid the Scylla of underfitting the data, resulting in a model that is inadequate, and the Charybdis of overfitting the data (i.e., fitting noise as if it were signal). Navigating between these hazards is by no means trivial, and standard statistical methods such as the F-test and stepwise regression are not to be trusted, as they too easily reject adequate models in favor of overly complex ones.

Eichhorn and Williams proposed a criterion based on trading off the decrease in average residual against the increase in the average error introduced through the error in the plate constants. The Bayesian approach reveals how these two effects should be traded off against each other, producing a sort of Bayesian Ockham's razor that favors the simplest adequate model. The basic idea behind the Bayesian Ockham's razor was discussed by Jefferys and Berger (1992). Eichhorn and Williams' basic notion is sound; but in my opinion the Bayesian approach to this problem is simpler and more compelling, and unlike standard frequentist approaches, it is not limited to nested models. Moreover, it allows for *model averaging*, which is unavailable to any classical approach.

3.1. Bayesian Model Selection

Given models M_i, which depend on a vector of parameters θ, and given data Y, Bayes' theorem tells us that

$$p(\theta, M_i \mid Y) \propto p(Y \mid \theta, M_i) p(\theta \mid M_i) p(M_i) \tag{5}$$

The probabilities $p(\theta \mid M_i)$ and $p(M_i)$ are the prior probabilities of the parameters given the model and of the model, respectively; $p(Y \mid \theta, M_i)$ is the likelihood function, and $p(\theta, M_i \mid Y)$ is the joint posterior probability distribution of the parameters and models, given the data. Note that some parameters may not appear in some models, and there is no requirement that the models be nested.

Assume for the moment that we have supplied priors and performed the necessary integrations to produce a normalized posterior distribution. In practice this is often done by simulation using Markov Chain Monte Carlo (MCMC) techniques, which will be described later. Once this has been done, it is simple in principle to compute posterior probabilities of the models:

$$p(M_i \mid Y) = \int p(\theta, M_i \mid Y) d\theta \tag{6}$$

The set of numbers $p(M_i \mid Y)$ summarizes our degree of belief in each of the models, after looking at the data. If we were doing model selection, we would choose the model with the highest posterior probability. However, we may wish to consider another alternative: model averaging.

3.2. Bayesian Model Averaging

Suppose that one of the parameters, say θ_1, is common to all models and is of particular interest. For example, θ_1 could be the distance to a star. Then instead of choosing the distance as inferred from the most probable model, it may be better (especially if the models are empirical) to compute its marginal probability density over all models and other parameters. This in essence weights the parameter as inferred by from each model by the posterior probability of the model. We obtain

$$p(\theta_1 \mid Y) = \sum_i \int p(\theta_1, \theta_2, \ldots, \theta_n, M_i \mid Y) d\theta_2 \ldots d\theta_n \tag{7}$$

Then, if we are interested in summary statistics on θ_1, for example its posterior mean and variance, we can easily calculate them by integration:

$$\begin{aligned} \bar{\theta}_1 &= \textstyle\int \theta_1 p(\theta_1 \mid Y) d\theta_1 \\ \mathrm{Var}(\theta_1) &= \textstyle\int (\theta_1 - \bar{\theta}_1)^2 p(\theta_1 \mid Y) d\theta_1 \end{aligned} \tag{8}$$

4. Simulation

Until recently, a major practical difficulty has been computing the required integrals, limiting Bayesian inference to situations where results can be obtained exactly or with analytic approximations. In the past decade, considerable progress has been made in solving the computational difficulties, particularly with the development of Markov Chain Monte Carlo (MCMC) methods for simulating a random sample (draw) from the full posterior distribution, from which marginal distributions and summary means and variances (as well as other averages) can be calculated conveniently (Dellaportas et al., 1998; Tanner, 1993; Müller, 1991). These have their origin in physics. Metropolis-Hastings and Gibbs sampling are two popular schemes that originated in early attempts to solve large physics problems by Monte Carlo methods.

The basic idea is this: Starting from an arbitrary point in the space of models and parameters, and following a specific set of rules—which depend only on the *unnormalized* posterior distribution—we generate a random walk in model and parameter space, such that the distribution of the generated points converges to a sample drawn from the underlying posterior distribution. The random walk is a Markov chain: That is, each step depends only upon the immediately previous step, and not on any of the earlier steps. Many rules for generating the transition from one state to the next are possible. All converge to the same distribution. One attempts to choose a rule that will give efficient sampling with a reasonable expenditure of effort and time.

4.1. The Gibbs Sampler

The Gibbs sampler is a scheme for generating a sample from the full posterior distribution by sampling in succession from the conditional distributions. Thus, let the parameter vector θ be decomposed into a set of subvectors $\theta_1, \theta_2, \ldots \theta_n$. Suppose it is possible to sample from the full conditional distributions

$$\begin{gathered} p(\theta_1 \mid \theta_2, \theta_3, \ldots, \theta_n) \\ p(\theta_2 \mid \theta_1, \theta_3, \ldots, \theta_n) \\ \vdots \\ p(\theta_n \mid \theta_1, \theta_2, \ldots, \theta_{n-1}) \end{gathered}$$

Starting from an arbitrary initial vector $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_n^0)$, generate in succession vectors $\theta^1, \theta^2, \ldots, \theta^k$ by sampling in succession from the conditional distributions

$$p(\theta_1^k \mid \theta_2^{k-1}, \theta_3^{k-1}, \ldots, \theta_n^{k-1})$$

$$p(\theta_2^k \mid \theta_1^k, \theta_3^{k-1}, \ldots, \theta_n^{k-1})$$
$$\vdots$$
$$p(\theta_n^k \mid \theta_1^k, \theta_2^k, \ldots, \theta_{n-1}^k)$$

with $\theta^k = (\theta_1^k, \theta_2^k, \ldots, \theta_n^k)$. In the limit of large k, the sample thus generated will converge to a sample drawn from the full posterior distribution.

4.2. Example of Gibbs Sampling

Suppose we have normally distributed observations $X_i, i = 1, \ldots, N$, of a parameter x, with unknown variance σ^2. The likelihood is

$$p(X \mid x, \sigma^2) \propto \sigma^{-N} \exp\left(-\sum_i (X_i - x)^2/2\sigma^2\right) \tag{9}$$

Assume a flat (uniform) prior for x and a "Jeffreys" prior $1/\sigma^2$ for σ^2. The posterior is proportional to the prior times the likelihood:

$$p(x, \sigma^2 \mid X) \propto \sigma^{-(N+2)} \exp\left(-\sum_i (X_i - x)^2/2\sigma^2\right) \tag{10}$$

The full conditional distributions are: for x, a normal distribution with mean equal to the average of the X's and variance equal to σ^2/N (which is known at each Gibbs step); and $-\sum_i (X_i - x)^2/\sigma^2$ has a chi-square distribution with N degrees of freedom. Those familiar with least squares will find this result comforting.

4.3. Metropolis-Hastings Step

The example is simple because the conditional distributions are all standard distributions from which samples can easily be drawn. This is not usually the case, and we would have to replace Gibbs steps with another scheme. A Metropolis-Hastings step involves proposing new value of θ^* by drawing it from a suitable *proposal distribution* $q(\theta^* \mid \theta)$, where θ is the value at the previous step. Then a calculation is done to see whether to accept the proposed θ^* as the new step, or to keep the old θ as the new step. If we retain the old value, the Metropolis sampler does not "move" the parameter θ at this step. If we accept the new value, it will move. We choose $q(\theta^* \mid \theta)$ so that we can easily and efficiently generate random samples from it, and with other characteristics that we hope will yield efficient sampling and rapid convergence to the target distribution.

Specifically, if $p(\theta)$ is the target distribution from which we wish to sample, first generate θ^* from $q(\theta^* \mid \theta)$. Then calculate

$$\alpha = \min\left[1, \frac{p(\theta^*)q(\theta \mid \theta^*)}{p(\theta)q(\theta^* \mid \theta)}\right] \tag{11}$$

Then generate a random number r uniform on $[0,1]$. Accept the proposed θ^* if $r \leq \alpha$, otherwise keep θ. Note that if $p(\theta^*) = q(\theta^* \mid \theta)$ for all θ, θ^*, then we will always accept the new value. In this case the Metropolis-Hastings step becomes an ordinary Gibbs step. Although the Metropolis-Hastings steps are guaranteed to produce a Markov chain with the right limiting distribution, one often gets better performance the more closely $q(\theta^* \mid \theta)$ approximates $p(\theta^*)$.

5. A Model Selection/Averaging Problem

With T.G. Barnes of McDonald Observatory and J.O. Berger and P. Müller of Duke University's Institute for Statistics and Decision Sciences, I have been working on a Bayesian approach to the problem of estimating distances to Cepheid variables using the surface-brightness method. We use photometric data in several colors as well as Doppler velocity data on the surface of the star to determine the distance and absolute magnitude of the star. Although this problem is not astrometric *per se*, it is nonetheless a good example of the application of Bayesian ideas to problems of this sort and illustrates several of the points made earlier (prior information, model selection, model averaging).

We model the radial velocity and V-magnitude of the star as Fourier polynomials of unknown order. Thus, for the velocities:

$$v_r = \bar{v}_r + \Delta v_r \tag{12}$$

where v_r is the observed radial velocity and $\bar{v}_r$ is the mean radial velocity. With τ denoting the phase and M_i the order of the polynomial for a particular model we have

$$\Delta v_r = \sum_{j=1}^{M_i} (a_j \cos j\tau + b_j \sin j\tau) \tag{13}$$

This becomes a model selection/averaging problem because we want to use the optimal order M_i of Fourier polynomial and/or we want to average over models in an optimal way. For example, as can be seen in Figures 1-3—which show fits of the velocity data for the star T

Monocerotis by Fourier polynomials of orders 4 through 6—to the eye the fourth order fit is clearly inadequate, whereas a sixth-order fit seems to be introducing artifacts and appears to be overfitting the data. The question is, what will the Bayesian analysis tell us?

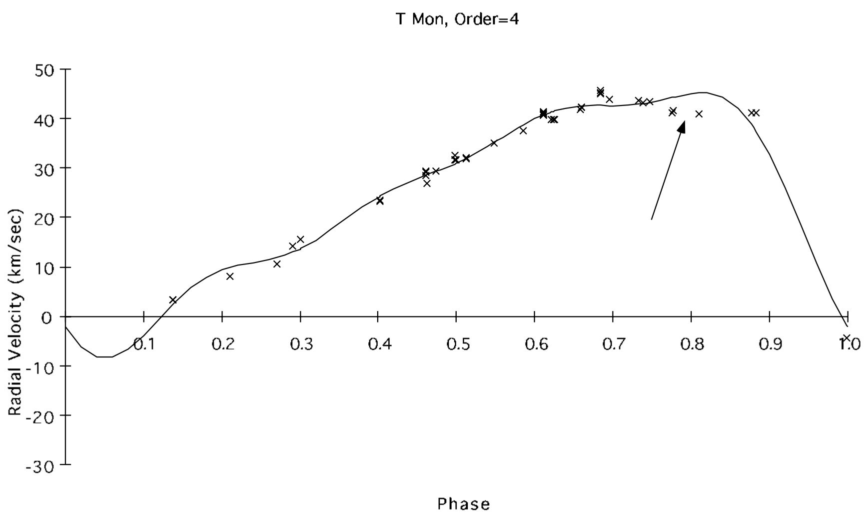

Figure 1. The radial velocity data for T Mon fitted with a fourth-order trigonometric polynomial. The arrow points to a physically real "glitch" in the velocity. This fit is clearly inadequate.

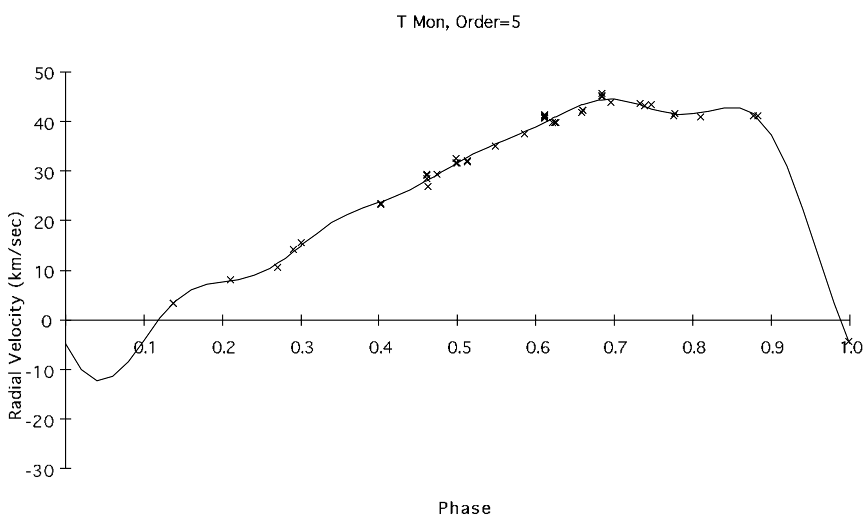

Figure 2. The radial velocity data for T Mon fitted with a fifth-order trigonometric polynomial. This fit seems quite adequate to the data, including the fit to the "glitch" of Figure 1.

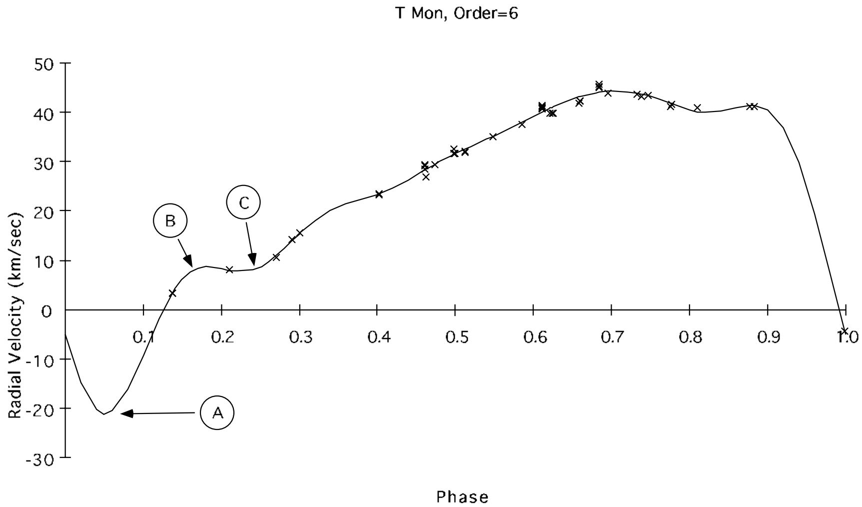

Figure 3. The radial velocity data for T Mon fitted with a sixth-order trigonometric polynomial. This fit is not clearly better than the fit of Figure 2, and shows some evidence of overfitting, as indicated by the arrows $A - C$; these bumps are not supported by any data (*cf.* Figure 2). Bump A, in particular, is much larger than in the lower order fit; Bumps B and C are probably a consequence of the algorithm attempting to force the curve nearly through the adjacent points.

The Δ-radius of the star is proportional to the integral of the Δ-radial velocity:

$$\Delta r = -f \sum_{j=1}^{M_i} (a_j \sin j\tau - b_j \cos j\tau)/j \tag{14}$$

where f is a positive numerical factor.

The relationship between the radius and the photometry is given by

$$V = 10(C - (A + B(V - R) - 0.5 \log_{10}(\phi_0 + \Delta r/s))) \tag{15}$$

where the V and R magnitudes are corrected for reddening, A, B, and C are known constants, ϕ_0 is the angular diameter of the star and s is the distance to the star.

The resulting model is fairly complex, simultaneously estimating a number of Fourier coefficients and nuisance parameters (up to 40 variables) for a large number of distinct models (typically 50), along with the parameters of interest (e.g., distance and absolute magnitudes). The Markov chain provides a sample drawn from the posterior distribution for our problem as a function of all of these variables, including model specifier. From it we obtain very simply the marginal distributions of parameters of interest as the marginal distributions of the sample, and means and variances of parameters (or any other desired quantities) as sample means and sample variances based on the sample.

Selected results from the MCMC simulation for T Monocerotis can be seen in Figures 4-7. The velocity simulation (Figure 4) confirms what our eyes already saw in Figures 1-3, namely, that the fifth-order velocity model is clearly the best. Nearly all the posterior probability for the velocity models is assigned to the fifth-order model, with just a few percent to the sixth-order model. Perhaps more interestingly, Figure 5 shows that the third and fourth-order *photometry* models get nearly equal posterior probability. This means that the posterior marginal distribution for the parallax of T Mon (Figure 6) is actually averaged over models, with nearly equal weight coming from each of these two photometry models. The simulation history of the parallax is shown in Figure 7; one can follow how the simulation stochastically samples the parallax.

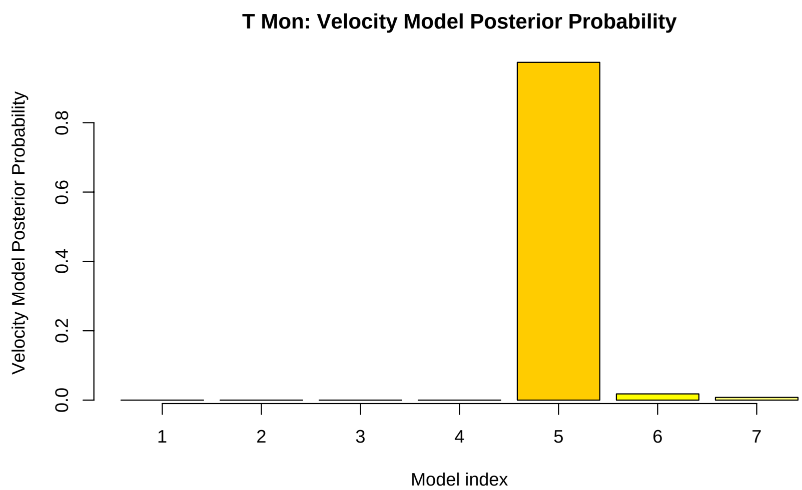

Figure 4. Posterior marginal distribution of velocity models for T Mon.

5.1. Significant Issues on Priors

Cepheids are part of the disk population of the galaxy, and for low galactic latitudes are more numerous at larger distances s. So distances calculated by maximum likelihood or with a flat prior will be affected by Lutz-Kelker bias, which can amount to several percent. The Bayesian solution is to recognize that our prior distribution on the distance of stars depends on the distance. For a uniform distribution it would be

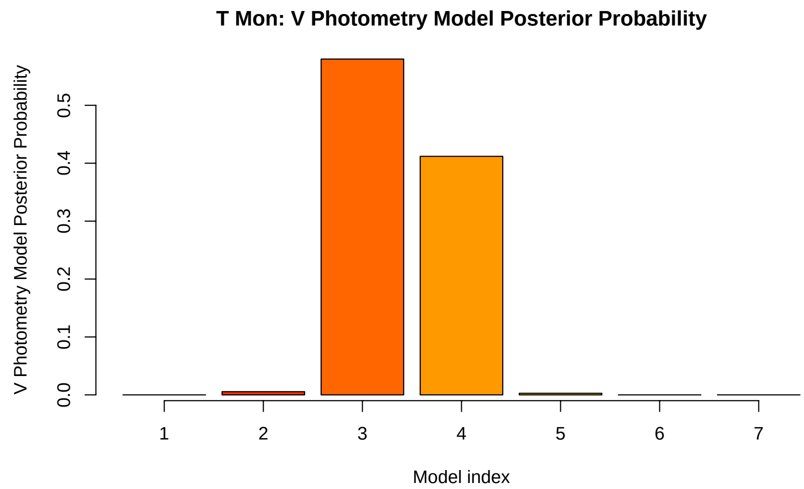

Figure 5. Posterior marginal distribution of photometry models for T Mon.

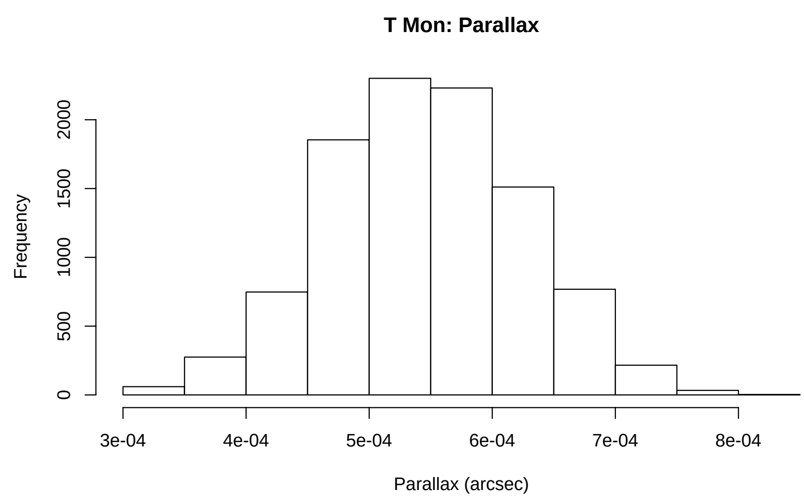

Figure 6. Posterior marginal distribution of the parallax of T Mon.

proportional to $s^2 ds$, which although an improper distribution, gives a reasonable answer if the posterior distribution is normalizable.

In our problem we have information about the spatial distribution of Cepheid variable stars that would make such a simple prior inappropriate. Since Cepheids are part of the disk population, their density decreases with distance from the galactic plane. Therefore we chose a spatial distribution of stars that is exponentially stratified as we go

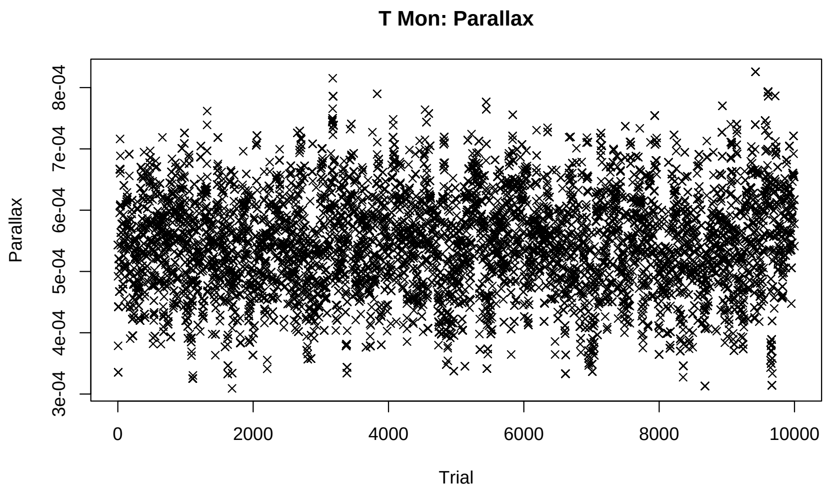

Figure 7. Simulation history of the parallax of T Mon.

away from the galactic plane. We adopted a scale height of 97 ± 7 parsecs, and sampled the scale height as well. Our prior on the distance is

$$p(s) = \rho(s)s^2 ds$$

where $\rho(s)$ is the spatial density of stars. For our spatial distribution of stars we have

$$\rho(s) = \exp(-z/|z_0|) \tag{16}$$

where z_0 is the scale height, $z = s \sin \beta$, and β is the latitude of the star.

The priors on the Fourier coefficients must also be chosen carefully. If they are too vague and spread out, significant terms may be rejected. If they are too sharp and peaked, overfitting may result. For our problem we have used a maximum entropy prior, of the form

$$p(c) \propto \exp(-c'X'Xc/2\sigma^2) \tag{17}$$

where $c = (a, b)$ is the vector of Fourier coefficients, X is the design matrix of the sines and cosines for the problem, and σ is a parameter to be estimated (which itself needs its own vague prior). This maximum entropy prior expresses the proper degree of ignorance about the Fourier coefficients. It has been recommended by Gull (1988) in the context of maximum entropy analysis and is also a standard prior for this sort of problem known to statisticians as a Zellner G-prior.

6. Summary

Bayesian analysis is a promising statistical tool for discussing astrometric data. It suggests natural approaches to problems that Eichhorn considered during his long and influential career. It requires us to think clearly about prior information, e.g., it naturally forces us to consider the Lutz-Kelker phenomenon from the outset, and guides us in building it into the model using our knowledge of the spatial distribution of stars. It effectively solves the problem of accounting for competing astrometric models by Bayesian model averaging. We can expect Bayesian and quasi-Bayesian methods to play important roles in missions such as FAME and SIM, which challenge the state of the art of statistical technology.

Acknowledgements

I thank my colleagues Thomas G. Barnes, James O. Berger and Peter Müller for numerous valuable discussions, Ivan King and an anonymous referee for their comments on the manuscript, and the organizers of the Fifth Alexander von Humboldt Colloquium for giving me the opportunity to honor my friend and colleague Heinrich K. Eichhorn with this paper.

References

Berger, J. O.: 1985, *Statistical Decision Theory and Bayesian Analysis, Second Edition*, pp. 27–33. New York: Springer Verlag.

Dellaportas, P., J. J. Forster, and I. Ntzoufras: 1998, 'On Bayesian Model and Variable Selection Using MCMC'. Technical report, Department of Statistics, Athens University of Economics and Business.

Eichhorn, H.: 1978, 'Least-squares adjustment with probabilistic constraints'. *Mon. Not. Royal Astron. Soc.* **182**, 355–360.

Eichhorn, H. and H. Smith: 1996, 'On the estimation of distances from trigonometric parallaxes'. *Mon. Not. Royal Astron. Soc.* **281**, 211–218.

Eichhorn, H. and E. M. Standish: 1981, 'Remarks on nonstandard least-squares problems'. *Astron. J.* **86**, 156–159.

Eichhorn, H. and C. A. Williams: 1963, 'On the Systematic Accuracy of Photographic Astrometric Data'. *Astron. J.* **68**, 221–231.

Gull, S. F.: 1988, 'Bayesian inductive inference and maxiumum entropy'. In: G. J. Erickson and C. R. Smith (eds.): *Maximum-Entropy and Bayesian Methods in Science and Engineering.* Dordrecht: Kluwer, pp. 153–74.

Jefferys, W. H. and J. O. Berger: 1992, 'Occam's razor and Bayesian statistics'. *American Scientist* **80**, 74–72.

Loredo, T.: 1990, 'From Laplace to Supernova 1987A: Bayesian inference in astrophysics'. In: P. Fogère (ed.): *Maximum Entropy and Bayesian Methods.* Dordrecht: Kluwer Academic Publishers, pp. 81–142.

Müller, P.: 1991, 'A generic approach to posterior integration and Bayesian sampling'. Technical report 91-09, Statistics Department, Purdue University.

Tanner, M. A.: 1993, *Tools for Statistical Inference.* New York: Springer-Verlag.

Address for Offprints: William H. Jefferys, Dept. of Astronomy, University of Texas, Austin, TX 78712