Probabilistic Parameter Determination of Stars and Galaxies

Research Essay - The Bayesian Method

Author: Billy Harrison[1]

Supervisor: Dr. Ralph Schönrich[2]

Degree Course: MSc Astrophysics

March 22, 2024

[1]University College London, Department of Physics & Astronomy, London, WC1E 6BT, UK

[2]Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK

Contents

1. Introduction

1.1. Bayesian Inference in Astrophysics

The recent increase in abundance and complexity of data sets from satellite missions such as *Gaia* and *Hipparcos* and their related follow-up surveys have gifted astrophysicists with an appropriate set of tools prime for the application of Bayesian methods. Measurements of stellar parallaxes, line-of-sight velocities, proper motions, temperatures, metallicities and stellar abundances from astronomical spectra can be observed through the lens of this statistical framework to explore previously unreliably probed corners of the Galaxy. This school of inference offers advantages of heightened efficiencies when dealing with the problems related with conceptual conclusions drawn from a set of premises in parameter space, and the presence of uncertainty in said data, compared to more traditional frequentist statistical views. We must be very careful in our employment of these measurements however, as systematic biases such as inadequate modelling, noisy or highly uncertain data, selection and temporal biases and thoughtless prior and posterior dependencies can steer us heavily in the wrong direction, and lead us to draw unfounded conclusions about how our universe behaves.

Historically there have been strong links between Bayesian applications and physics - particularly astrophysics - where, due to the nature and scale of what is being measured, data uncertainties are intrinsic. The theoretical and physical understanding, existence of observational constraints (instrument sensitivity, atmospheric wavelength scattering, temporal limitations) and the cumulative updating of new data are all factors in justifying the use of this framework on a large scale. Bayesian analysis cleverly takes advantage of these, what one would instinctively call pitfalls in the context of data interpretation, by incorporating these uncertainties, prior theoretical knowledge and model flexibility in its overarching methods to improve astrophysical parameter estimation. Loredo (1995) is in favour of this approach, offering the argument that an astrophysicist or cosmologist would be much more confident in his statements about our particular Universe when they encompass pre-existing knowledge and data that actually exists rather than via the study of thousands of simulated Universes where the "notion of a statistical ensemble is often extremely contrived and can hence seem irrelevant". Trotta (2008) also backs this Bayesian interpretation and use of data, arguing that the general complexity of both theories and observations will scale with the refinement of the statistical and data analysis skills used to justify them, so ever improving numerical tools are essential in keeping pace. He also puts forward the argument that a "better appreciation of [Bayesian] statistical statements" might help distinguish spurious scientific claims backed by diluted and unfiltered data not considered in the eyes of this framework from actual physical mechanisms that take into account systematic biases and prior knowledge.

1.2. Bayes Theorem & Probability Theory

The entirety of the approach outlined so far hinges on the result of an extremely simple idea known as Bayes theorem, given from the consequences of probability theory and presuming the Bayesian idea of how probability itself is defined - "probability is a measure of the degree of belief about a proposition" (Trotta, 2008), and how it can be manipulated. This is in direct opposition to the classical frequentist approach to statistics which contrastingly treats the definitions of "probability" and "frequency" interchangeably, i.e. probability is encoded by the limit of the long-run frequencies estimated via the use of repeated sampling of a large number of repeatable trials. This approach does not consider exterior information or beliefs about the particular system being designed as a statistical

model, that is however shown in Bayes theorem:

$$P(X|Y,I) = \frac{P(Y|X,I) \cdot P(X|I)}{P(Y|I)} \tag{1}$$

where $P(X|Y,I)$ is the posterior probability: the conditional probability of parameter set X being true given an observation set Y and relevant background information I, $P(Y|X,I)$ is the observed likelihood - the conditional probability of making a set of observations Y given the X and I, $P(X|I)$ is the prior probability, or background knowledge, we assign to X, and $P(Y|I)$ is the probability that the set of observations Y is made which we can easily set to unity (Pont & Eyer, 2004). This can therefore be simplified to

$$P(X|Y,I) = P(Y|X,I) \cdot P(X|I) \tag{2}$$

which is the posterior probability distribution function, or posterior PDF, on the chosen parameter space (Schönrich & Bergemann, 2014). In an astrophysical context, the chosen parameter set X could include a stars effective temperature, chemical composition, age etc, whereas the observation set Y entails only directly detectable measurements as opposed to values that can only be mathematically or empirically inferred, e.g. spectroscopic, astrometric and parallax measurements. The prior effectively captures the initial knowledge on the model parameters before the data is observed (see Section 2.1 for more). Good practice when it comes to priors involves choosing informative ones when there exists sufficient existing knowledge, and adopting caution in this choice when there is little or no prior knowledge available. The likelihood can be framed by asking the question: assuming one knows the parameters exactly, what is the distribution of the data? For a given data/parameter pair, it registers how ”likely” the data is. It must not be misconstrued that the likelihood is a function of the data, or is a probability density function rather than just being *proportional* to a probability. It can be noted that in astrophysical contexts, e.g. exoplanet detection by radial-velocity (Butler et al., 2017; Fischer et al., 2014) and transit (Dragomir et al., 2019) methods, parallax determination (Lindegren et al., 2012), source brightness inference (Tak et al., 2017), and Hubble constant estimations (Hubble, 1929), the likelihood can be assumed to be Gaussian-modelled, with the mean equal to the ”true” model parameter of interest. In other cases, such as describing the brightness of a high-energy source, it can be modelled discretely, for instance by a Poisson distribution.

Qualitatively, the importance and great inferring power of this theorem lies in the fact that the posterior probability that a hypothesis is true given the data can be linked to the probability that one would observe the data given the hypothesis being true - a term that has a much greater chance of assignment (Sivia & Skilling, 2006). Instead of studying how a model should be interpreted based off the introduction of new data and prior knowledge, if we want to isolate a derivation of the probabilities of new parameters given existing probabilities, we can introduce the Bayesian marginalisation theorem:

$$P(X|I) = \int_{-\infty}^{\infty} P(X,Y|I)\, dY \tag{3}$$

which is a powerful device that enables us to integrate out unwanted “nuisance parameters” such as background noise signals, quantities which systematically enter the analysis space but are of no interest.

Both frequentist and Bayesian schools of thought are based on assumptions about how the particular statistical model is framed, how we choose to relate observations with parameters, and if they are treated as fixed unknown values or random variables encoded by probability distributions - they are fundamentally subjective. The key difference lies within their abilities to assist decision making and analysis once the correct context has been established: frequentist theory involves making one’s mind up about an outcome under a shroud of ignorance (a lack of exhibited information) and asking if the

data changes that, whereas Bayesian theory starts with a set of evidence, or an opinion (the prior) and questions whether or not one should change one's mind after the incorporation of observational data. Bayesian theory is objective, easily programmable and moldable, combines varying quantities and is applicable to large sets of observational data, hence is a perfect fit for an astrophysical modelling context.

1.3. Historical Milestones

Although being advocated for long ago by Jeffreys (1939), the pragmatism and conceptual underpinnings of the Bayesian method were largely ignored in the field of astrophysics for many years. It wasn't until the late 80s/early 90s when the realisation that the superiority of this thinking would be well suited to astrophysical problems began to rear its head within the community of contemporary researchers. One of the progenitor advocates was Loredo (1992), who comprehensibly laid forward the case for its implementation, citing the observed interest growth in applied statistics and economics, and "look[ed] forward to a similar revolution occurring in astrophysics". He pioneered many studies in the field, including general event arrival time periodic signal detection (Gregory & Loredo, 1992), various studies on Supernovae (Drell et al., 2000; Loredo & Lamb, 1989), and γ-ray burst distributions sources (Loredo & Wasserman, 1998). Other notable studies of the time include Bailyn et al. (1998)'s Bayesian analysis of stellar black hole mass distributions, Hernandez et al. (2000b)'s inference of star formation histories for local group spheroidal dwarf galaxies and the *Hipparcos* solar neighbourhood (Hernandez et al., 2000a), and predictions of cosmological cold dark matter models by Bunn & Sugiyama (1995).

Focusing on more contemporary stellar and Galactic physics applications, Pont & Eyer (2004) and Jørgensen & Lindegren (2005) present Bayesian methods of stellar age estimates from theoretical isochrones, and divulge into the statistical inference from the resulting age PDFs. Nordström et al. (2004) briefly outline similar approaches to age estimates in their landmark study of the *Geneva-Copenhagen-Survey* (GCS) of the solar neighbourhood. Casagrande et al. (2011) re-analyse the GCS to derive stellar and metallicity parameters where they provide details on the Bayesian scheme adopted for dealing with the observational errors in T_{eff}, metallicities and absolute magnitudes. Bailer-Jones (2011) adopts a Bayesian method for inferring intrinsic stellar parameters, using parallax and photometric data in unison to estimate non-parametric posterior PDFs. Burnett & Binney (2010) employ related techniques in their stellar distance inference determination. Schönrich & Bergemann (2014) similarly present a Bayesian framework/algorithm for stellar parameter derivation that combines all available observable and prior information. Section 2 involves a deeper dive into the individual methodological advantages/disadvantages of these landmark studies, particularly into the statistical techniques and the rise they give to biases and uncertainties.

1.4. Applications in Adjacent Fields

Over the last few years, there have been many studies that employ Bayesian techniques, particularly into the field of black holes and neutron stars. Specifically a variety of studies utilising LIGO and Virgo's observing runs, including Abbott et al. (2021)'s search for an isotropic gravitational-wave background. De Luca et al. (2021) provide evidence supporting the coexistence of merger populations of both astrophysical and primordial origin - Hall et al. (2020) arguing this study by strongly favouring astrophysical models than primordial. Essick & Landry (2020) discriminate between neutron stars and black holes via a Bayesian inference of gravitational-wave source sub-populations. The GRAVITY Collaboration et al. (2020) remark on the first General Relativity Schwarzschild precession in the orbit of the star S2, creating 14 parameter model fits for the distance, mass, position, orbit parameters and

a dimensionless fitting parameter relating to the similarity of the model to Newtonian or GR physics, calculated from posterior fitting and MCMC analysis.

2. Literature Review

This section details in depth the Bayesian method and its historical application to the field of astrophysics. It aims to explore the statistical developments, methodologies, lines of thinking and discoveries, as well as the strengths and limitations associated with the analysis. Key papers and articles (including the contemporary ones mentioned in Section 1.3) that pioneered the field will be rigorously examined.

2.1. Prior Distributions

2.1.1 Double-Counting Bias

One of the main attributes of Bayesian analysis is the application of the prior in combination with observational data, or new evidence, using Bayes Theorem to procure a new posterior. However, to maintain the integrity of this type of inference, it is important to avoid double counting both observational and *a priori* information. Bayesian analysis updates the probabilistic strength of a belief based on new intelligence entering the picture. Hence if the equivalent information is used multiple times in, for example, defining the prior or calculating the likelihood function, an overemphasis focusing on this particular statistic or distribution in parameter space is observed. If this piece of information or data set is considered in multiple stages in the analysis pipeline, it will lead to an intensification of its perceived importance, hence appearing more significantly weighted within the framework. The confidence one would then have in ascertaining probabilities and models of real-world physics becomes increasingly uncertain if an unnoticed overconfidence in particular beliefs consolidates the inference multiple times, applying too much of this analytical weight than is truly encoded. Additionally, from an epistemological viewpoint, the prior distribution represents initial beliefs about a physical system and Bayesian analysis itself rests on the doctrine of updating one's beliefs based on this initial standpoint when new information becomes available. Therefore if a bias is introduced by considering the same information in both stages of the Bayesian method, the very essence of the probabilistic framework comes under scrutiny and we introduce elements of circular reasoning. For these reasons it is important to distinguish between prior beliefs and new evidence, ensuring an accurate and unbiased representation of the posterior probability distribution and mirroring the true impact of said evidence on already established beliefs.

For example, a potential method for measuring the total mass of a galaxy could involve the analysis of a set of individual stars and their velocity dispersion σ to probe the gravitational potential, and hence the overall mass. When applying Bayesian inference the logical steps would be to shape the prior mass distribution $P(M)$ of the target galaxy on an observed sample of similar-type galaxies, and the likelihood function $P(\sigma_{\text{obs}}|M)$ would be informed by the observational target galaxy velocity dispersions for a given mass via the M-σ relation. Here the same observational data is used in both the prior and likelihood components of the posterior, first directly and then inferred, leading to potential bias levels in the target galaxy's posterior mass distribution. If there is a skew observational distribution towards lower masses, informing the posterior of this independent information twice will bias it in the same way, even if the σ_{obs} data would otherwise denote a higher mass.

An example of this double counting of information is seen in Burnett & Binney (2010); here they declare a selection function that catalogues a stars probability of sample inclusion - this explicitly utilizes photometric data which has already been applied in the likelihood function. See Section 2.6

for more detail. Additionally, they employ the estimation of uncertainties in the observable quantities $\boldsymbol{\sigma}_y$ directly from the intrinsic variables or underlying characteristics $\mathbf{x}$, and require an extra term in the PDF, $P(\boldsymbol{\sigma}_y|\mathbf{x})$, to relate them in an analogous Bayes Theorem. Inferring the probability of the quoted errors from $\mathbf{x}$ rather than estimating them directly e.g. Sale (2012) is making multiple applications of the same data when also considering the prior, in addition to the likelihood and selection functions which are conditional of *both* $\mathbf{x}$ and $\boldsymbol{\sigma}_y$.

2.1.2 Non-Informative Priors & Reparameterisation

On the other hand, a prior may be non-informative in that it doesn't make use of information *at all*, and instead hinges on geometrical probabilities of the desired parameter. For instance, the physically-motivated distance prior used by Bailer-Jones (2015); Bailer-Jones et al. (2018a) and visualised in Figure 1 is derived from a theoretical volume dV sweeped by a solid angle $\mathrm{d}\Omega$, and implies a $\mathrm{d}^2\mathrm{e}^{-d}$ relationship on the assumption of an exponential decrease in stellar density as a function of distance as one extends radially outwards in the Galactic disk.

It is important to note the distinction, or lack thereof, between adopting a flat prior, where before data observation all parameter values are considered equally likely, and no prior, where one explicitly states that preconceived intel on the parameters are not imposed. In practice these are often used interchangeably; however, imposing the absence of information could be construed as information itself, thus there are no such things as non-existent priors under the Bayesian purview, only flat ones. One must be careful when dealing with flat priors as they are not invariant under transformation, i.e. what is flat for one parameterisation may be informative in another. Take the astrophysical scenario of stellar parallax estimation. A realistic uniform prior on parallax that is truncated between minimum and maximum values ϖ_{min} and ϖ_{max} could be considered. Equivalently, a uniform distance prior could be chosen where the bounds are coupled via the distance-parallax relationship $d_{min} = \frac{1}{\varpi_{\mathrm{max}}}$ and $d_{max} = \frac{1}{\varpi_{\mathrm{min}}}$. Although both seem to encode similar information, they behave very differently when expressed as a function of d, as shown in Figure 1. This highlights how the choice of parameterisation influences the behaviour of non-informative parameters (strictly speaking, these are weakly-informative because the physical knowledge is reflected in the chosen bounds.) For an objective prior that is invariant under any parameterisation, see Jeffreys prior (Jeffreys, 1946).

2.1.3 Utilising Physical Intelligence

There are many astrophysical scenarios where the priors can be interpreted differently. Studying Lyman-break galaxies, Benítez (2000) found two distinct peaks in the expected redshift distribution. When determining individual redshifts, this knowledge would be absorbed into the priors. In a different field, Huang et al. (2023) study neutrino energy spectra from a core collapse supernova (CCSN) with Bayesian inference. They combine data from multiple detectors requiring a framework that simultaneously considers multiple inputs in its posterior probability distribution. Due to unreliable and limited astrophysical observations for SN1987A (e.g. Morris & Podsiadlowski (2009)), most constraints for the priors have to be obtained from simulations.

Huang et al. (2023) model the neutrino spectrum with a Gaussian distribution. They emphasise that this assumed Gaussian is not physical; rather it just encodes various confidences that the neutrinos will lie within a specific number of standard deviations of the mean Gaussian centre. Such priors will quickly become obsolete in the presence of more and more precise data: then the observational likelihood will become much narrower compared to the priors and dominate the posterior distribution. These processes are a far cry from initial Bayesian neutrino analysis by Loredo & Lamb (1989) where priors are considered uniform. Back to stellar ages: when setting priors in a multi-dimensional

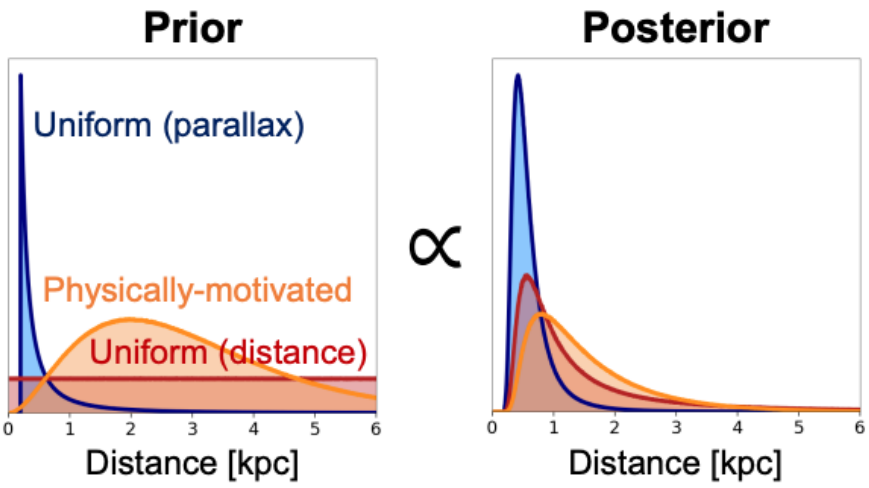

Figure 1: *Left*: Three example priors, one uniform in distance d (red), another corresponding to the reparameterisation in ϖ following the reciprocal relationship (blue), and a third more physically-motivated prior determined by Bailer-Jones et al. (2018b) (orange). *Right*: The respective corresponding posterior PDFs given Gaussian likelihoods in d. Source of figure: Eadie et al. (2023)

parameter space, it is common and good practice to use a flat prior in the main parameter under consideration. For example, Jørgensen & Lindegren (2005) assume a flat prior in age and metallicity when disentangling their information on these two parameters due to the assumption of no *a priori* correlation between them - their goal of determining the existence of an age-metallicity relationship should not assume one exists intrinsically, otherwise they will be proving their hypothesis with circular logic.

Additionally, on a shown quantity any third party can easily multiply on their own prior, in contrast to secondary parameters: it would vice versa be wrong to assume a flat metallicity prior when just providing an age posterior – e.g. for very high observed metallicity ([Fe/H] ~0.3 or similar), the slope of the metallicity distribution will mean that the true expectation for metallicity is significantly lower, thus a flat prior would bias the age estimate. Furthermore, when encountered with data points on the stellar grid that pass through multiple separate isochrones, Jørgensen & Lindegren (2005) examine the choice presented in the absence of any additional information to preferentially opt for the youngest of the isochrone selection as the most reasonable solution. They posit that for a field star on that particular point (e.g. $\log T_{eff} = 3.8$, $M_V = 3.0$) via comparison with exterior information, younger age stars will likely still reside in the slow-evolving core hydrogen burning stage; if higher ages are adopted, there would be an implied assumption that the star is more evolutionarily advanced, which is less probable for your random star. Here, this conscious choice of information could be quantified by the prior distribution rather than being applied *a posteriori* for each individual point in parameter space.

In a similar vein, Bailer-Jones (2011) uses domain knowledge of the Hertzsprung-Russell Diagram (HRD) in the prior to constrain stellar parameters and associated line-of-sight extinctions. This prior, built using stellar evolution simulations from Vallenari et al. (2010), describes the probability distribution of star locations on varying HRD regimes. Advantage is taken of this clear non-uniform population spread to determine parameter value probability ranges for magnitude, in return constraining other parameters such as temperature. Here there are no issues involving the multi-use of data as this information is determined independently of photometric or parallax measurements.

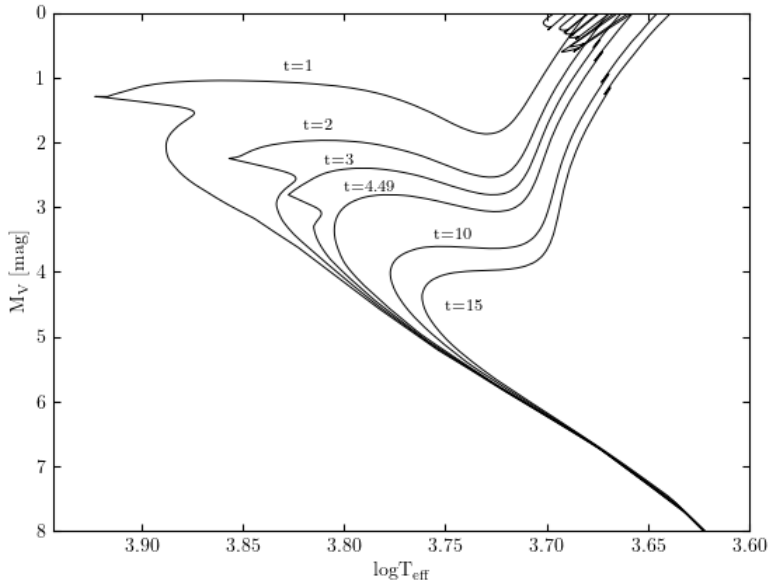

Figure 2: An example of a relatively dense grid of BaSTI stellar isochrone tracks from Pietrinferni et al. (2004); Pietrinferni et al. (2006, 2009) in V-Magnitude M_V and logarithmic effective temperature logT_{eff} parameter space. Shown are tracks for ages t = 1, 2, 3, 4.49 (approximate solar age), 10 and 15 Gyr.

2.2. Isochrone Fitting

To cover the requisite areas of parameter space and to apply appropriate physical constraints on the stellar catalogue, dense grids of theoretical isochrones - paths traced on the colour-magnitude diagram (CMD) representing stellar populations with equivalent ages - can be used to portray stellar models and obtain corresponding PDFs. See Fig 2 for a visual representation of an isochrone grid. Sufficiently dense grids allow for maximal coverage for various combinations of stellar parameters e.g. effective temperature T_{eff}, luminosity L, surface gravity log(g) and iron abundance [Fe/H]. The isochrone library comparison method involves PDF derivations provided by grids of desired stellar parameters for a given star (Casagrande et al., 2011), calculated by summing over these theoretical isochrones.

Landmark studies involving this methodology have evolved in tandem with the advancement of spectroscopic and astrometric catalogues and increasingly refined theoretical isochrone models; Edvardsson et al. (1993) obtained multi-element abundances using derived ages for individual sample stars from initial Vandenberg (1985) isochrones. Bertelli et al. (1994) isochrones consolidated these results with refinements by Ng & Bertelli (1998). Stellar ages have been derived for large numbers of field dwarfs with the use of *Hipparcos* parallaxes (e.g. Asiain et al. (1999); Feltzing et al. (2001); Ibukiyama & Arimoto (2002)).

Given the historical circumstances, isochrone fitting was considered the primary and most reasonable method for parameter inference. When astronomers look up into the night sky at galaxies and field and cluster stars, they see a temporal snapshot of the canonical universe - in other words, they directly see how stars populate the HRD at a given observable and non-changing (on cosmic scales) time. It is hence straightforward for population parameters to be inferred based on these observations in comparison to utilising stellar evolutionary tracks, which require additional assumptions and detailed physical evolutionary constraints. Isochrone methods are also generally computationally less

intensive than stellar tracks, and particularly Bayesian methods which deal with complex sampling algorithms (see Section 2.4), model refinement (Section 2.5), posterior inference (Sections 2.3 and 2.7) and general high parameter space dimensionality.

Some of the forerunners to adopt astrophysical Bayesian methods in combination with isochrone fitting were Pont & Eyer (2004) who present a computation of individual star age PDFs using inverse functions that relate observables with physical parameters, based on the selection and analysis of isochrones nearest the desired object in parameter space. An in depth one-dimensional age PDF derivation example applying these ideas in tandem with the prior and likelihood relationship observations is presented in Section 2.7.

da Silva et al. (2006) use a method in which they consider an isochrone interval between two neighbouring initial masses M_{init}, fixed metallicity [Me/H] and age τ with mean properties in M_V, T_{eff} and the desired parameter. The probability of a target star landing within interval bounds is found by estimating the relative number of said interval stars according to the marginalised initial mass function (IMF), and by summing the probabilities of observed M_V and T_{eff} aligning with theoretical values under Gaussian scattering assumptions. Here, the initial masses are not interpolated over the isochrone to achieve sufficient grid spacing as seen in Schönrich & Bergemann (2014) - such interpolation is important in maximising stellar modelling precision. A dense grid of interpolated isochrone points captures small incremental changes in stellar properties determined by the change in M_{init}. Standard isochrone evolutionary tracks used in e.g. da Silva et al. (2006) may display sections of M_{init} discontinuity where the point spacing gets sparser, hence affecting stellar parameter probing accuracies at these "dead spaces" where there may not be nearby points modelled by the isochrone, for example at the red giant branch (RGB) tip and turnoff.

Jørgensen & Lindegren (2005) employ similar methods, where for each desired parameter value, in this case age τ, the function is also summed over the isochrone set. Here, a continuous interpolation parameter u is introduced along each isochrone to calculate the general parameter function. Characterised by physically significant sections of the isochrone, u encodes equal evolutionary intervals. u is taken to be proportional to the curve length along the track; the linear interpolation ensures the general morphology of the isochrones is preserved, in order to disentangle separate tracks in parameter space from intersecting where they should not.

Schönrich & Bergemann (2014) consolidate the interpolation methods by adding isochrone point weightings calculated in proportion to local parameter space - a sum of the average distance between the core space being assessed e.g ΔM_{init}, $\Delta\tau$, Δ[Me/H]. This volume encodes an average distance between neighbouring isochrone points as well as between neighbouring isochrones themselves. The explored model probability can be updated by multiplying each weighted point with the associated uncertainty - here represented by an n-dimensional Gaussian g with auxiliary dispersion vector σ spanning all relevant space, and summing over the isochrone:

$$P_{\text{mod}}(\mathbf{X}) = \sum_i W_i \, g((\mathbf{X} - \mathbf{X}_i), \sigma) \tag{4}$$

Where $\mathbf{X}_i$ is the parameter space vector ascertained from the model grid and $\mathbf{X}$ represents the parameter space itself. Casagrande et al. (2011) adopt similar methods for Bayesian age estimates, running over isochrone points in [Me/H], τ and mass volume intervals. da Silva et al. (2006) use sum over the isochrones with [Me/H] and τ weightings proportional to a Gaussian mean and standard deviation, and flat age prior respectively.

Including such model probabilities facilitates the task of constraining parameters based on the observational data. It also helps distinguish the likelihood of competing astrophysical models given the parameters obtained from observational data - or conversely separating degenerate solutions where obtained parameters could be explained by several observational models.

These methods of isochrone comparison assume the empirical models encode accurate evolutionary description of path stars of different mass, metallicity and age that distribute across parameter space, particularly along the RGB of the HRD. Thus one must be careful that these assumptions hold for the selected ranges being explored and discontinuities are therefore minimised.

2.3. Maximum-Likelihood & χ^2 methods

Maximum-Likelihood (ML) and Least-Squares (LS), or similarly χ^2 statistics, are commonly used procedures in data analysis and can be justified through the Bayesian approach (Sivia & Skilling, 2006). In the context where the observer is largely ignorant of the stellar mechanism background, the prior PDF is taken to be uniform; thus the resulting posterior PDF transforms to coincide with the likelihood function. Hence the ML estimate value $\mathbf{x_0}$ of the parameter set $\mathbf{x}$ can be inferred from the maximum value of the posterior distribution curve. The solution invariance under monotonic transformations of PDFs or other functions, e.g. the luminosity function $\phi(L) \rightarrow \phi(\log L)$, is one of the many positive features of ML (Bailer-Jones, 2017). In the extensional case where the observer assumes the individual data are independent, further approximations can be made relating to the likelihood function; the overall joint PDF is calculated by the product of probabilities, or the likelihood function, of the individual measurements. Assuming the associated experimental noise is Gaussian-scattered then the likelihood can be approximated as

$$\mathrm{P}(Y|X,I) \propto \exp\left(-\frac{\chi^2}{2}\right) \tag{5}$$

where χ^2 is the sum of the squares of the normalised residuals

$$\chi^2 = \sum_{k=1}^{N} \frac{(F_k - D_k)^2}{\sigma_k^2} \tag{6}$$

where F_k is the ideal noiseless data and σ_k is the expected error bar size of the associated kth datum D_k (Sivia & Skilling, 2006).

The maximum observed likelihood will occur when χ^2 is minimised and the corresponding $\mathbf{x_0}$ is known as the least-squares estimate. χ^2 statistics are only applicable for Normal or Gaussian distributions where sufficiently large data samples are being analysed.

Both of these methods stem from a frequentist approach to analysing data and can be useful in a Bayesian context if the particular criteria outlined above are fulfilled. In practice these criteria are often ill-suited in astrophysics where the specifics of the problem do not align with said requirements. For instance, Benítez (2000) explores the differences in ML and Bayesian estimates via the study of previous cosmological photometric redshift analysis shortcomings - astronomical object recession velocity estimates in the absence of high resolution spectra - and the integration of Bayesian inference techniques to overcome said shortcomings. One basic approach is "Spectral Energy Distribution" (SED) (Bailer-Jones, 2011) fitting where compiled template spectras generated via population synthesis techniques, or empirically through *a priori* data considerations are, after extinction-correction, collated with known galaxy colours to match best conforming redshifts. Another is the training-set technique, where a galaxy's redshift and apparent magnitude/colour relationship can be inferred using multiparametric fits. A common problem of utilising the SED is the possible mismatch between predicted and observed template colours and colour/redshift degeneracies due to lack of Galactic magnitude data. The training-set evaluation suffers from assuming a single functional form of the colour/redshift relationship, and the empirical nature of the analysis encodes a hard limit on the spectroscopic redshift

which is therefore impractical for studying faint galaxy samples.

The Bayesian inference application takes advantage of prior probabilities by weighing them whilst averaging over the individual likelihoods so the probability estimations are not affected by trick peaks (caused by noise etc.). The ML method for SED (which ignores priors) could mistakenly identify these as genuine *desiderata*. Additionally, in this context, Lyman-break galaxies that observe two systematic redshift peaks will make ML methods unworkable due to the lack of monotonicity, compared to the Bayesian thought process. Similarly, the training-set technique's lack of depth of observations can be remedied in a similar sense - the corresponding priors can be calibrated using the data sample from which the redshifts are obtained, including values outside of the hard limit. χ^2 statistics are also frequently considered but ultimately dropped, e.g. in Buchner et al. (2014) due to small sample sizes of X-ray sources (as the bins are cluttered with Poisson noise; re-binning the data into smaller intervals to account for this unwantedly loses information in the energy resolution), and the likelihood function of said sources not following a Gaussian distribution.

2.4. Sampling Algorithms

The objective of Bayesian inference is to obtain information from the posterior PDF, quantified by the distribution of parameters given background information encoded by the prior, and the distribution of experimental data given by the likelihood. However, direct posterior estimation can often be too computationally intensive for a given model (Kelly, 2007) and one may resort to obtaining random draws using Markov-Chain-Monte-Carlo (MCMC) methods.

2.4.1 MCMC Methods

MCMC simulations are mathematical processes that sample the posterior distribution by constructing Markov chains - sequences of possible "events", that in each transition do not depend on the past process state, only the current event. One can model the posterior by utilising these chains and adopting "walkers" which travel the domain of parameter space collecting samples that converge on an approximation of the desired posterior. MCMC methods can explore spaces of high dimensionality, jumping from one parameter configuration to the next and allow for easy estimations of statistically significant properties of the posterior e.g. means, variances, expectation values and standard deviations.

MCMC sampling methods were initially developed in computational science fields towards the turn of the millennium, and Bayesian solutions for astrophysical problems began to emerge soon thereafter. As with general Bayesian methods, the adoption in astrophysical contexts grew hand-in-hand with the exponential growth in computational power. Early studies include cosmological parameter PDF estimations from the Cosmic Microwave Background by Christensen & Meyer (2000) and Knox et al. (2001), and binary system parameter inference using the Gibbs sampler by Christensen & Meyer (2001).

Mann et al. (2015) infer stellar parameters using Markov chains and applied observational constraints. The likelihood function they adopt is a standard sum of Gaussians evaluated at the residual between the model prediction vector for a set of unknown (including both observable and unobservable) parameters, and the observed data. Obtained parameters are tested for unwanted bias from the particular MCMC algorithm, with identical results for a catalogue of well understood, parameter space spanning stars yielded from both a standard single-walker Metropolis-Hastings algorithm and a Quickhull algorithm (Barber et al., 1996). Although the results were analogous, the model uncertainties obtained from the former were larger, perhaps due to a lone walker getting iteratively stuck and repeating covered ground in a local minima well, while also in the global minimum vicinity. For their

model fitting of galaxy cluster lensing, Jullo et al. (2007) attempt to deter this local minima problem by running ten interlinked Markov chains simultaneously.

2.4.2 Reversible-Jump MCMC

Farr et al. (2011) use marginalised posterior probabilities of proposed physical stellar-mass black hole models as a comparison apparatus, using a powerful technique called the reversible-jump MCMC (Green, 1995). This is an extension of the standard MCMC but conducted within the frame of the "super-model" that covers every model and the contained core parameter values, i.e. $\{M_i, \theta_i\}$, with the posterior distribution of the super-model parameters calculated by Bayes Theorem (Equation 1). The probability of a model M_i being "correct" - in other words the distribution compatibility with the data d - is estimated by the model-frequency count in the chain

$$P\left(M_i \mid d\right) = \int dX_i P\left(M_i, X_i \mid d\right) \approx \frac{N_i}{N} \tag{7}$$

with N_i being the number of discrete model parameters samples, and N being the total sample number. For this algorithm to be implemented efficiently, the proposed jump into a model, say a narrow Gaussian, needs to land inside the peak or other important regions of parameter space by chance, which is evidently hard to control. Therefore the output chain must be sufficiently long to ensure the full model comparison. They propose an improved algorithm that increases the efficiency of jump proposals by considering model posterior distribution information explored by the initial single-model regular MCMC samples.

2.4.3 Metropolis-Hastings Algorithm

McMillan (2011) uses the Metropolis-Hastings (M-H) algorithm, an offshoot of the MCMC sampling process in which the next sample in the sequence of "events" is instead evaluated on an accept-reject basis, indicating whether the sample is kept or discarded. Here they employ this to describe model parameters of the MW thin and thick disk components with predetermined constraints. See Appendix A for a detailed overview of the algorithm.

In their neutrino study, Huang et al. (2023) use the M-H algorithm to sample the posterior. An analytical form of this distribution may not exist from the observed data, instead manifesting as a complex form that does not synthesise with any known distribution family. It is therefore necessary to resort to this sampling technique in order to marginalise for the desired parameters. Initial samples generated using this method are prone to heavy influence by starting parameter space location as there is a natural and inevitable correlation between neighbouring samples. Therefore to obtain an appropriate sample set, a number of initial points are disregarded to be free of sampling bias - in the case of this study, the first 200 are dropped. Similarly, if the initial Markov chain lands in high probability regions of the posterior sample, the method may be too computationally expensive to explore all distribution areas - the choice allocated to the proposal density - and can be inefficient if the generated samples have high degrees of correlation. The proposal density hence chooses incrementally small scattering distances due to the covariance matrix dependency.

The implementation of MCMC sampling in an astrophysical context has thus been characterised by a general incremental technique adaptation and refinement when faced with a unique set of problems. The evolution of computational efficiency and the development of software tools, algorithms and sampling packages has given rise to its continued use as a cornerstone of Bayesian inference.

2.5. Model Selection & Quantification

Choosing adequate physical models, characterised by the choice of parameters to be varied and the associated prior distributions, is paramount in the stellar parameter inference process - and hence towards the goal of understanding underlying astrophysical processes. The probability assignment to these models leads to the considerations of weighing their complexity with their observed ability to fit the data; the methodical quandary of model selection when faced with conflicting mathematical frameworks and schools of thought needs to be considered.

2.5.1 Information Criteria

Liddle (2007) explores the obstacle of quantifying competing astrophysical and statistical models using sets of tools based on information theory and Bayesian inference. The balance between the suitability and inferential value of a particular fit to model data and the complexity and reasonable posterior application of said fit must be found. He introduces the Bayesian evidence E - but instead opts for describing two simpler statistics which still encodes this tension, the Akaike Information Criterion (AIC) (Akaike, 1974)

$$\text{AIC} \equiv -2 \ln \mathcal{L}_{\text{max}} + 2k \tag{8}$$

and the Bayesian Information Criterion (BIC) (Schwarz, 1978)

$$\text{BIC} \equiv -2 \ln \mathcal{L}_{\text{max}} + k \ln N \tag{9}$$

The two statistics are similar, both encoding the maximum likelihood $\mathcal{L}_{\text{max}}$ (see Section 2.3) achievable by the models and the number of parameters k, the BIC also including the number of data points N used in the fit. There is general consensus between the two models when applied, but one should be careful when considering the extent to which these assumptions fit real physical situations. It is best to consider, for example, degenerate and unconstrained parameters which are penalised by these criteria but not the evidence E. This problem can be overcome by defining the Deviation Information Criterion (DIC), computed from posterior MCMC samples (Section 2.4), which focuses on determining the number of usefully constrained parameters for a given data set. This has already been introduced into dealing with astrophysical problems by Kunz et al. (2006). It considers the effective number of parameters, p_d, derived from the deviance of the likelihood - a useful statistic relatably proportional to the prior width so that $p_d = N$ dimensions (with a well-defined posterior) corresponding to a well measured parameter, and $p_d = 0$ when the prior width approaches zero and thus an unconstrained parameter. This is useful as the DIC overcomes the failure of the AIC and BIC that do not ignore unconstrained parameters. In many stellar parameter estimation problems, the denominator in Equation 1 is simply taken as a normalisation constant. However in situations where model selection categorisation is crucial, it can be harnessed to reward the synergy of data fit - the likelihood $L(X)$, and model predictiveness - the prior $P(X)$, and is known as the aforementioned evidence E:

$$E \equiv \int L(X)P(X)dX \tag{10}$$

It is thought of as the probability of obtaining the data Y given the assumed model I (again see Equation 1) and quantifies the model complexity. Jullo et al. (2007) offer the analogy to Occam's razor: "All things being equal, the simplest solution tends to be the best one", suggesting that the simplest model that maximises E (i.e. one with minimal parameters and high prior and posterior similarity) is most probable.

2.5.2 Model Selection and Comparison

Farr et al. (2011) explore model selection criteria on the underlying mass distribution of stellar-mass black holes, using observed masses from X-ray binary systems. They consider both parametric models for this distribution: power-law, decaying exponential, Gaussian and two-Gaussian and log-normal, as well as non-parametric models i.e. histograms with carefully chosen bins as to not dilute distribution feature sensitivity. MCMC (Section 2.4) analysis will provide clear posterior distributions for each model, given appropriate priors, but will not quantify the resemblance of the models to the actual physical distribution. They employ the aforementioned reversible-jump MCMC and propose an improved algorithm that increases jump proposal efficiency by considering model PDF information explored by the initial single-model regular MCMC samples. The relative evidence is calculated for both parametric and non-parametric models, with the highest relative probability for describing stellar-mass black hole distributions being for power-laws and single peaked Gaussians. A combination of the AIC (Equation 8) and BIC (Equation 9) could have been used here to address model confidence, facilitating model complexity penalisation via the k terms, and computational simplicity.

In their study of X-ray spectra, Buchner et al. (2014) develop a Bayesian framework to distinguish and compare different physically motivated models in the context of Active Galactic Nuclei (AGN). Common error estimates include: (1) individual parameter error estimation by calculating a one sigma statistic drop value correspondence; (2) estimating local derivatives and inverting the Fisher information matrix which encodes the deviation of non-informative/flat prior distributions according to Jeffreys rule (Robert, 2001); or (3) comparing values of confidence interval contours based on χ^2 distributions (Section 2.3). However these approximations are only valid if the probability distribution of the X-ray spectra closely resembles a single peaked Gaussian, or there is minimal codependency between the studied parameters. They opt for a Bayesian parameter estimation approach, identifying "sub-volumes" in a parameter space deformed by pre-administered priors and using these as proxies for the probability integral over parameter space. Their Bayesian model comparison adopts similar information criteria as in Liddle (2007), namely the BIC and AIC, stating that the model that minimises these values should be preferred.

As well as the methods outlined in Farr et al. (2011), Buchner et al. (2014) explore various traditional Goodness of Fit (GoF) methods e.g. the reduced χ^2 for normal distributions or posterior-based approaches, but conclude that there is no sufficient frequentist or Bayesian statistical method for testing the tension between the fitted model and the examined parameters. Allowing themselves to be less rigorous and instead ask questions about the clarity at which a model "explains" the data in a range covered by the observable, they (in addition to Eadie et al. (2023)) suggest the use of Q-Q plots (Wilk & Gnanadesikan, 1968) which compare two PDFs by plotting their quantiles - or continuously intervaled distribution regions with equal probabilities. An approximate identity line relating these two distributions visualises a greater extent of similarity as it approaches a $y = x$ relationship. These plots compare two model distributions' shape, location in parameter space, scale and skewness, offering surface level visual comparison overviews between competing models.

2.5.3 The Poissonian Distance Metric

In their Milky Way study, Robin et al. (2003) propose model predictions as a tool for estimating Bayesian probabilities; given a set of observable stellar parameters and errors, using suitable simulations the model can establish probability classifications. This assumes that the model correctly reproduces Galactical stellar content, and can probe parameters of an observed star given spectroscopic and astrometric observables. They propose the approach can supplement standard photometric or spectral classification with the limit being model data adequacy - one has to individually verify in each case that the model reproduces the distribution of available observable data. Star formation rates (SFR) of

the local Galactic disk are analysed by Mor et al. (2019) using *Gaia* DR2 data to probe 15-dimensional parameter space. BGM FASt galaxy simulations (Robin et al., 2003) and an approximate Bayesian computation algorithm are used to perform this analysis. They utilise synthetic full sky magnitude-limited stellar samples of colour distributions to compare with the *Gaia* DR2 analogue and subject the parameter set to a threshold calculation relating to the so-called Poissonian distance δ_p (Mor et al., 2018):

$$\delta_p = \frac{1}{N} \sum_{i=1}^{N} q_i \cdot (1 - R_i + \ln(R_i)) \tag{11}$$

where R_i is defined as the quotient $R_i = f_i/q_i$ and q_i and f_i are the number of stars in the data and the model, respectively (see Kendall & Stuart (1973) and Bienayme et al. (1987)). If this metric is smaller than a manually chosen threshold, the parameter set for that particular star is accepted into the posterior, otherwise it is rejected - this ensures set combinations that produce worse results than the best models proposed in previous literature (Mor et al., 2017) are disregarded. δ_p encodes the similarity between the *Gaia* DR2 CMD and the simulated counterpart in terms of star counts and is extended here to evaluate parameter sets such as thick-disk stellar mass densities and age-binned surface stellar mass densities of thin-disk stars.

2.5.4 Model Discrepancies

Mor et al. (2019) also observe discrepancies in the difference in regional (RGB, Main-Sequence (MS) and $3 < M_{\varpi} < 4$) star counts between BGM simulations and the *Gaia* DR2 data. This highlights the importance of being rigorous with the model selection and assumption choices prior to the statistical inference application. The main pitfalls include: poorly refined thick-disk and stellar evolutionary models; falsely assumed stellar parameter relationship distributions, such as radial-metallicity and age-metallicity; an incomplete *Gaia* DR2 catalogue; and the faint star sampling bias. Mass-loss rate underestimations assumed by the stellar evolutionary models could explain star count discrepancy across the entire MS - an excess or deficit of observed stars in particular mass ranges could carry through due to a lack of understanding on distribution behaviour for an entire population. Additionally, the simulations may not have taken into account binary star systems (this particular paper does not mention them), which can influence stellar parameter catalogue calculations.

Jørgensen & Lindegren (2005) attempt to consider the overall statistical effect of unrecognised binaries on stellar age estimates by adding random secondary distributions to their synthetic sample. This leads to effective age overestimates as expected, but luckily with an overall negligible statistical effect. One possible explanation is that typical profiles of star catalogues being probed for age here are those located near the MS turn-off where the age-estimation is insensitive to a shift in V-Magnitude on the y-axis.

2.6. Biases, Errors & Degeneracies

Like any statistical school of thought, Bayesian analysis is susceptible to biases, methodological and systematic errors, and parameter degeneracies. These need to be considered before integration into the inference pipeline. Identifying the genesis of these biases is crucial for stellar parameter inference; handling them early prevents error propagation throughout the analysis - stopping the physical significance and origin of the uncertainties being lost to the noise of the Bayesian process.

2.6.1 Quality Cuts & Selection Functions

Sources of bias include: (1) the choice of prior, as discussed in Section 2.1 with reference to the double counting of experimental data by Burnett & Binney (2010); (2) sampling biases as explored in the context of the MCMC exploration of the posterior by e.g. Huang et al. (2023); and (3) data selection biases from quality cuts and selection functions. In their study of *Gaia* DR2 data, Mor et al. (2019) attempt to avoid white and brown dwarfs in their data set by introducing a quality cut to disregard stars with a $M_{\varpi} < 10$. This could introduce a sampling bias as the fainter population of MS stellar types is being effectively excluded. Hence stellar parameter biases and uncertainties are affected as the full stellar catalogue with an all-encompassing set of photometric properties is not being analytically considered.

Burnett & Binney (2010) deal with similar sampling biases of *Hipparcos* stars which provide reasonably well constrained distances, and restrict parallax errors to below a certain threshold. Therefore one preferentially cuts away stars with low parallaxes and large distances leaving the sample biased towards smaller distances. Instead of quality or selection cuts/choices used to 'clean' a sample of unwanted data intervals, they adopt a selection function into the equation for Bayes Theorem, which encodes the probability of a star being located in the sample given its inherent value and error. This is split into components dependent on the 'intrinsic' variables x and observed parameters of the analysis star $\bar{\boldsymbol{y}}$. However, Schönrich & Bergemann (2014) posit that there is no useful reasoning behind introducing the non-parametric term as sample cut choices are not typically made based on the behaviour of stellar parameters unknown beforehand, and rather on the collective observational survey studied *a priori*. Burnett & Binney (2010) apply quality cuts using this selection function by collapsing it to zero when the observed stellar parallaxes are sufficiently low compared to the associated error. The use of a selection function generally implies there is known observational knowledge necessary to calculate the likelihood distribution. The selection function instead still encodes the constraint against faint stars. The function will arbitrarily discriminate against the tail of distance overestimates resulting in biased error estimates. There is thus no good reason as to why all known parallax information cannot be used as the biases will nevertheless appear encoded by the likelihood if said information is to drop out via the selection function.

Other sources of discrepancy include kinematic biases which are minimised using quality cuts to remove particular data intervals judged to be unreliable or irrelevant, or downright identified as outliers. Suggestions from Schönrich et al. (2019)'s examination of the *Gaia* RV sample involve removing the Galactic mid-plane from the stellar sample by rejecting data with Galactic longitude |b| > 10 deg, or disregarding measurements if there exists a ϖ/σ_{ϖ} ratio > 5. They must be careful not to introduce any selection bias - cuts that reject data relevant to the stellar parameters being studied may lead to an insufficiently representative dataset that does not cover the overall population, in addition to cuts that hinge on well-established information with potential to interfere with the prior PDFs. If the cuts and the prior are founded on the same set of observational measurements, bias in the posterior PDF could be introduced (see Section 2.1). The data removal on a non-random criteria basis for a random data sample may also lead to bias inferences due to the non-uniform likelihood of certain sample-pool groups being potentially more or less likely to be excluded.

2.6.2 Systematic Biases

Other biases are systematic, predictable and constantly deviate from the expected values in a particular and repeatable fashion due to an intrinsic physical source dependence. Schönrich et al. (2019) study the radial velocity sample of *Gaia* DR2 and use Bayesian-derived stellar distances. The statistical method outlined in Schönrich et al. (2012) is employed to validate said distances and test *Gaia* parallaxes to find an associated distance-bias pointing towards a parallax offset. Systematic in-

plane and radial velocity component bias measurements are considered: negligible (typically ignored due to *Gaia* data precision/accuracy) line-of-sight velocity determination errors appear as distance underestimates and the proper motion counterpart conversely appear as overestimates. Biases in this study include that induced by the velocity ellipsoid tilt - a quantity that appears frequently in *N*-body modelling, particularly stellar and Galactic mechanics, and refers to the velocity distribution of stars in different coordinate directions. The radially elongated velocity ellipsoid can distort the calculation matrix used to correlate observed velocity components with the assumed relative distance-bias by introducing a locally varying heliocentric velocity component relationship, impacting measurement accuracy. To correct for this, they assume the velocity ellipsoid always points towards the Galactic Centre, which induces a requirement that there is no inference by the Galactic structure on a correlation between this heliocentric velocity correlation and the Galactic position of the velocity ellipsoid. They posit that such velocity correlations would only physically appear locally due to e.g. passing by stellar streams, and only manifest themselves as distance statistics uncertainties at the small sky patch scale, and outright cancel to the first order at the scale of large-sky coverage such as that of *Gaia*.

Systematic biases are often present on estimated PDFs, such as the metallicity distribution function used by Casagrande et al. (2011) for Bayesian age estimates. A naive implementation of metallicity measurements here could introduce age disparities correlated to the extreme ends of the MDF, with potential to give rise to an artificial age-metallicity correlation. Appropriate priors that reflect the true underlying distribution can be used to reduce this bias, and one would be tempted to adopt the MDF itself as this prior, but this would only reflect instabilities around the distribution peaks. They instead create an analytic tailed-Gaussian function to approximate the distribution and flatten the curve at low metallicities.

Casagrande et al. (2011) also account for reddening on their photometric T_{eff} derivations. They introduce corrections for GCSII stars with E(b-y) > 0.01 mag and at distances greater than 40pc. Biases could emerge here as fainter stars observe greater magnitude uncertainties, thus the previous correction conditions with such sensitive thresholds may falsely flag target stars, breeding potential error propagation in the analysis.

2.6.3 Evading Degeneracies

Bayesian inference in astrophysics often deals with the study of giant stars, e.g. da Silva et al. (2006), which involves an intrinsic so-called age-metallicity degeneracy, i.e. old metal poor stars are located in a similar region in the CMD as young metal rich objects. By measuring the metallicity of sample stars, it is possible to resolve this degeneracy and estimate stellar ages from the CMD. Example PDFs for select observable stars outline parameter errors originating from *Hipparcos* parallax measurement errors, and for red giants, metallicity errors generate from the aforementioned age-metallicity degeneracy. The isochrone fitting method used by da Silva et al. (2006) outlined in Section 2.2 is all-encompassing and accounts for subtle colour and bolometric magnitude variations, thereby reducing final stellar parameter determination errors when observed metallicities are included in the analysis. On the other hand, naive isochrone fits can result in severe biases, for instance, what Pont & Eyer (2004) name a terminal age bias. Some areas of parameter space are more densely populated by the isochrone grid than others. This is because the IMF doesnt account for stellar evolution time-scales while mapping from initial mass to colour or luminosity, hence it may result in an abundance of stars in normally sparse population regions.

In their examination of posterior probability contours in extinction A_0 and effective temperature T_{eff} space, Bailer-Jones (2011) notes that for an overwhelming majority of artificially reddened *Hipparcos* catalogue stars there is a strong A_0-T_{eff} data-intrinsic degeneracy region - potentially due to selective stellar dust absorption at particular wavelengths and subsequent effects on colours at these

temperatures. Varying associated astrophysical parameters corresponding to stars at some A_0 and T_{eff} will result in the same colours being observed, hence encoding a degeneracy along the long narrow contours. As these degeneracies stretch over a range 1000K in T_{eff} and over 1 dex in A_0, the parameter inference here will not be particularly accurate. The addition of extended prior information would be needed to counteract this degeneracy by, for example, introducing the knowledge that the extinction values are near negligible.

Addressing the wide range of areas in which biases such as the ones explored above arise is essential in the implementation of the Bayesian analysis pipeline. As seen, the mitigation of said biases often involves careful study of stellar model specification, prior sensitivity and physical analysis, and data quality and utility assessments via quality cuts or selection functions. Analyses prior to any posterior PDF computation assist in choice justification made to minimise the error and bias that one can expect in any physically-motivated modelling context.

2.7. Posterior Analysis

2.7.1 Confidence Intervals

Although more at home in the frequentist realm, one can employ the use of Confidence Intervals (CIs) as a means to estimate ranges in which population parameters are likely to fall. One interpretation of this is the reliability in which a parameter X can be inferred based on the posterior curve area between two points, say X_1 and X_2, equal to a percentage of the curve that is covered. Since this area is proportional to the extent in the belief that X falls within the given range, this can be considered a sensible uncertainty estimation protocol. Assuming the PDF is normalised to have unit area, the confidence interval is defined as follows to determine X_1 and X_2:

$$\int_{X_1}^{X_2} P(X|Y,I)\mathrm{d}X \approx 0.95, \tag{12}$$

Another possible interpretation: if one were to sample the posterior and construct intervals for each, one would expect a certain percentage to contain the "true" population parameter. A CI of 68% proportional to $\pm 1\sigma$ or 95% ($\pm 2\sigma$) is standard. While constructing the posterior PDF of the stellar age τ, Jørgensen & Lindegren (2005) define the "G-function", or the age component of the likelihood function, marginalised over mass and metallicity using the IMF. This can be thought of as the relative likelihood that a star has age tau after the elimination of mass and metallicity parameters. The G function can be adopted to produce CIs. One proposed calculation involves taking the ratio of the G-function integral with respect to tau over the desired interval τ_1 to τ_2, with the integral over all possible age estimates covered by the isochrones, for example. However this can easily be misinterpreted if the G function is essentially uniform and carries no age information for a particular interval. The sub interval equal to the confidence level times the entire measured age space would hence be technically equatable to a quasi-confidence-interval, but without the value of any intrinsic information stored within its boundaries.

Instead they opt for a different CI derivation from the G-function involving the knowledge (e.g., Casella & Berger (1990)) that the likelihood ratio, or the probability that G is $\geq G_{lim}$, the theoretical limit in G. This implies that $-2\ln G$ experiences a χ^2 distribution, and is compatible with a Gaussian. For the required confidence level, a corresponding value of G_{lim} must be defined from the CI - the confidence interval between τ_1 and τ_2 is the shortest interval such that $G(\tau) < G_{lim}$ is outside said interval. They state that the simple solution of using fixed threshold values G_{lim} relating to the confidence levels is quite reasonable and should produce realistic CIs.

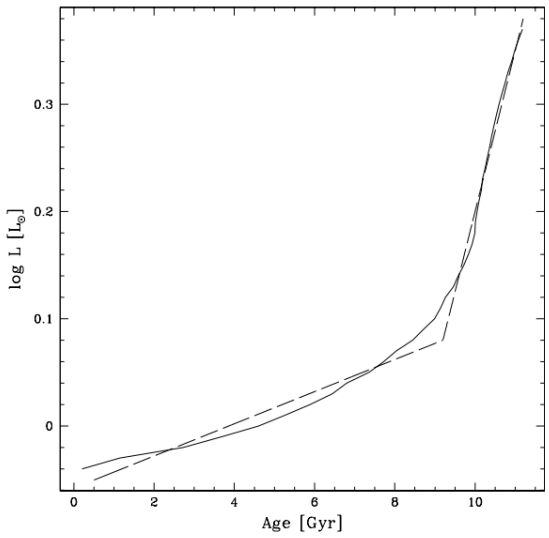

Figure 3: log$L = \mathcal{F}(\tau)$ approximate empirical transformation law (dotted line) i.e. log$L$ as a function of age τ at a fixed solar temperature T (5780K), from Geneva stellar evolution models (Schaller et al., 1992). There is an approximate ratio difference of 10 between the slopes in the MS phase (<9.2Gyr) and the evolved stellar phase (>9.2Gyr) which coincides with the downturn in the posterior (Figure 4). Plot courtesy of Pont & Eyer (2004).

2.7.2 Mean, Median & Mode

Jørgensen & Lindegren (2005) extract estimates on single valued estimates for age, strongly recommending against the use of the mean (i.e. the centre of gravity of the area under the G curve) or median (bisecting the area under the G curve), and tend towards the mode (the position of the global maximum in G) as the method of choice. The mean tends to assign ages in the middle of the permitted range, whereas the mode tends to assign extreme ages in such cases, clearly indicating bad estimates. Additionally, in the case of CI assignments as calculated above, they posit the mode will *always* be located within the interval, which cannot be guaranteed with the mean or median.

In studies where the outcome involves mostly unimodal and symmetrical 1-dimensional marginalised PDFs like a single-peaked Gaussian (e.g. Bailer-Jones (2011)), the mean and the 90% CI are the statistics of choice. For Casagrande et al. (2011), the ML, median, expectation values and 5,16,84 and 95% CIs are extrapolated from the age PDFs.

2.7.3 Posterior Modelling: An Example

The implementation of prior parameterisation, isochrone fitting and empirical parameter relationships and their influence on the posterior are explored via a simple one-dimensional case in Section 2.2 of Pont & Eyer (2004). In their words, by reducing dimensionality, fundamental statistical features of the stellar age determination can easily be illustrated - considering the one dimensional analogue sheds some clarity on the use of Bayesian inference in this context. They explore a standard approach of relating physical parameters X to observable quantities Y via a function $Y = \mathcal{F}(X)$ derived from stellar evolution models - see Figure 3.

One approach to computing age values of individual stars would be to interpolate stellar isochrone tracks to create a grid in (temperature T, luminosity L, iron abundance [Fe/H]) parameter space and match desired stars to points here in which the age and mass value correlate. Interpolation between theoretical isochrones this way is essential to obtain values of $\mathcal{F}$(mass, age, abundance) and hence an

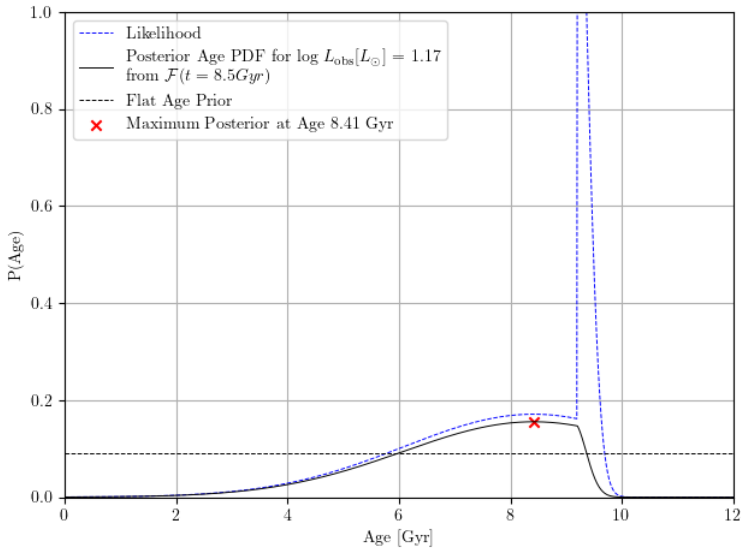

Figure 4: 1D Age PDF calculated from Geneva stellar evolution models to infer a relationship log$L$ = $\mathcal{F}(t)$. The prior is shown by the black dashed line, the likelihood by the blue dashed line, and the posterior PDF by the solid line. The posterior PDF has a maximum close to the value used in the above transformation, backing the use of its implementation.

invertible continuous relation of (m,τ,z) = $\mathcal{F}^{-1}(T, L, [\text{Fe/H}])$ - the desired linking of data and physical variables. For the one dimensional case, they assume that the age τ is computed from a single observable, luminosity L, related by a solid line transformation law log$L$ = $\mathcal{F}(\tau)$ from Geneva stellar evolution models at constant solar temperature and metallicity (Schaller et al., 1992) - the desired posterior probability distribution function in age inferred via Bayes theorem with this relationship encoded within the likelihood function and given an observed value logL_{obs}. The example transformation law between log$L$ and τ is shown below:

$$\log L = \mathcal{F}(\tau) = \begin{cases} \alpha\tau+\gamma \\ \qquad\qquad \text{if } \tau_0 < \tau < \tau_{\text{max}} \end{cases} \tag{13}$$

with $\alpha = 0.015$, $\beta = 0.15$, $\gamma = -0.058$, $\tau_{\text{max}} = 11$Gyr, and τ_0 = 9.2Gyr, the break at τ_0 represents a strong change in the linear relationship due to the star coming off the main sequence, as taken into account by the stellar model. The observational likelihood is assumed to be a single peaked Gaussian centred around the observed log$L$ value, which is expressed in terms of the observations whereas the prior distribution is in terms of the age parameter and must be modified as such to be variably consistent within Bayes Theorem. Without any concrete prior knowledge of the real age, this is assumed to be flat in time and as such becomes a step function in log$L$ - a similar flat to non-flat observation to the distance and parallax prior reparamterisation observed in Figure 1. The resulting posterior PDF is shown in Figure 4 for a singular observed log$L$ value calculated from a measurement made at τ = 8.5Gyr. From this graph it is clear to see that the empirical transformation laws derived from stellar models can be used to modify probability distributions - a Gaussian in log$L$ does not commute evenly to reparameterised age space. It is also evident that a "broken" or unevenly distributed log$L$ prior contributes more heavily to the downturn in the posterior PDF for ages >9.2Gyr compared to the likelihood which sharply peaks in this region.

3. Future Plans & Applications

3.1. Essay Overview

As outlined in the above literature analysis, Bayesian thinking offers a grand and varied selection of analytical techniques to infer a range of stellar parameters from astrometric and spectroscopic data. Seeking to appraise scientific hypotheses in light of observed data and modelling assumptions is one of the main goals. The thought processes thoroughly explored involve: (1) solving problems concerned with reporting probabilities; (2) how statements of uncertainty combine with underlying probability theory; (3) what we actually mean by "background information" encoded within the priors; (4) the virtues of Bayesian methods of constraining parameters; and (5) potential statistical applications in an astrophysical context. The overall Bayesian analysis pipeline should be widely applicable to a range of astrophysical problems, take exhaustive advantage of prior information and reasonable likelihoods, be arbitrarily expandable dependent on available observational stellar clarity and desired physical quantities, and should ideally make use of all available information sources simultaneously. Full parameter PDF estimations should be computationally efficient thanks to a detailed problem overview that identifies and exploits constraints on the observational data, the idealised parameter space, and theoretical stellar models, to minimise overall problem complexity and effective dimensionality (Schönrich & Bergemann, 2014).

For whatever astrophysical or analytical problem we face, it is paramount to construct a robust and adaptable framework that can handle any systematic or unanticipated complications, as well as apply the above conditions to maximum fulfilment. The ultimate intention always involves the gradual chiselling away of the unknown to produce both quantitative and qualitative interpretations of the physical universe. Hence the goal of this project should be to adopt previous Bayesian approaches from the literature with aims of improvement, focusing on refining the following considerations:

- Ensuring the use of a wide range of observation data and background physical stellar theory to be as unbiased as possible.
- Providing stellar parameter distribution estimates that satisfy the imposed observational constraints.
- Evaluating the method viability when faced with incomplete or low quality data sets - i.e. poor or non-existent astrometric or spectroscopic information.
- Similarly, evaluating the method viability and cohesion when analysing data from various surveys.
- Applying priors in a physically motivated outlook with reference to background information or astrophysical phenomena. Extensionally, applying non-informative priors carefully and assessing their properties under reparameterisation.
- Exploring methods of exploiting interpolation and parameter volume weighting in the context of isochrone fitting, and the subsequent PDF estimations.
- Analysing and sampling the posterior with the relevant statistical tools and qualitatively stating their advantages/disadvantages.
- Quantifying and separating the models used to help define the problem via the likelihood function.

- Identifying sources of parameter bias and degeneracies before the Bayesian analysis and eradicating them to the fullest extent via e.g. quality cuts and selection functions.
- Investigating sources of uncertainty in the parameter PDF estimations and model predictions.
- Clearly documenting the analysis methodology and reproducibility with reference to the above specifications.

Now that the Bayesian philosophy, notation, techniques and applications have been thoroughly outlined with reference to the literature, tackling astrophysical problems through this lens will be significantly easier, especially if we are dictated by the operational template outlined above, and throughout the rest of the text.

3.2. Potential Problems

The progress so far has largely involved a thorough examination of the literature. It is thus time to consolidate this gained understanding by looking towards current problems in stellar and Galactic physics where Bayesian analysis could be applied. Recent areas of research in the relevant field which may offer ideas to be explored are outlined below.

In Galactic mechanics, Navarro-Boullosa et al. (2023) infer parameters of l-boson stars using low brightness surface galaxy rotation curves, such as the dark matter component. Carnall et al. (2018) observe relatively rapid star-formation history quenches over 1-2Gyr for a sample of quiescent galaxies with M >10^{10}M$_{\odot}$ - they present a Bayesian analysis computational package for inference and parameter estimation for complex galaxy spectra. On a similar page, Chevallard & Charlot (2016) present a tool to build mock galaxy catalogues to model and interpret their SEDs, incorporating radiation production and its transfer through the interstellar medium (ISM). Rogantini et al. (2021) adopt Bayesian models and inference techniques to examine high-resolution grating spectra and distinguish between hot gas in the ISM and intrinsic gas of an X-ray binary.

In stellar physics, Semenova et al. (2020) use a Bayesian analysis of spectra in an open Galactic cluster to constrain stellar structure and evolution models. Lin et al. (2018) derive stellar masses and ages using Bayesian isochrone fitting and *Gaia* DR1 parallax and 2MASS photometry data.

As is clearly seen, there are many subareas of astrophysics where Bayesian methods are integral to parameter inference, or at least supplemental on the auxiliary. Hence there are many promising potential applications to opt for, all offering streamlined stellar parameter PDF estimates to help probe answers to astrophysical problems. The next steps will be to apply these learned statistical methods on large data sets, deriving unbiased full stellar parameter sets from combined parallax and spectroscopic data, for example stellar ages, which can be used to probe information on the structure of, or the change in SFR of the Milky Way as done in e.g. Mor et al. (2019), or updating methods pioneered by Carnall et al. (2018). Another application could be the Bayesian inference of the α-enhanced RGB population, whose masses imply incompatibly young ages (τ < 6Gyr) against the expected Galactic canon. Stars with these masses are usually formed at early epochs; however, high observed α-abundances may imply material was gained from AGB companions and are hence products of binary evolution (Zhang et al., 2021).

3.3. Future Scope

Here is a prospective overview of the next steps of the project over the coming months. While this is not set in stone, it offers a basic overall guide to use as a benchmark, keeping in mind that I am a part-time student so the timeline of the thesis write-up is stretched over two years.

April:

- Create age PDFs from the model probability using dense isochrone grids with the method outlined above.
- Use marginalisation to integrate out unwanted variables to obtain the moments and hence the expectation value of a parameter set $\langle X_j \rangle$ and its standard deviation $\sigma_{X_j} = \sqrt{\langle {X_j}^2 \rangle - \langle X_j \rangle^2}$.

May to June:

- Expand the application of stellar parameter derivation from observational data to multi-dimensional space.
- Continue conducting research on Bayesian applications in astrophysics, ordering and connecting the historical work, trends and conflicting debates, differing methodology and approaches - all framed within an overview of critical evaluation.

July to August:

- Establish a methodology taking into account those explored in the literature review, create merged local stellar catalogues from spectroscopic, photometric and astrometric data and derive the stellar parameters.
- Major analysis on error propagation derivations and handling, sources of potential bias to determine a degree of statistical confidence in the parameter inference.

September and Beyond:

- Create a synthetic catalogue with a known standard model, or with Monte-Carlo simulations, and scatter the observables with wide Gaussian functions to model the observational errors and compare with the parameters to identify any needed model recalibrations.
- Application of the derived data to learn more about Galactic structure or stellar chemical composition. Potential fields of interest are listed above.

4. Bibliography

Abbott R., et al., 2021, , 104, 022004

Akaike H., 1974, IEEE Transactions on Automatic Control, 19, 716

Asiain R., Figueras F., Torra J., Chen B., 1999, , 341, 427

Bailer-Jones C. A. L., 2011, , 411, 435

Bailer-Jones C. A. L., 2015, , 127, 994

Bailer-Jones C. A. L., 2017, Practical Bayesian Inference: A Primer for Physical Scientists. Cambridge University Press

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018a, , 156, 58

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018b, , 156, 58

Bailyn C. D., Jain R. K., Coppi P., Orosz J. A., 1998, , 499, 367

Barber C. B., Dobkin D. P., Huhdanpaa H., 1996, ACM Trans. Math. Softw., 22, 469–483

Benítez N., 2000, , 536, 571

Bertelli G., Bressan A., Chiosi C., Fagotto F., Nasi E., 1994, , 106, 275

Bienayme O., Robin A. C., Creze M., 1987, , 180, 94

Buchner J., et al., 2014, , 564, A125

Bunn E. F., Sugiyama N., 1995, , 446, 49

Burnett B., Binney J., 2010, , 407, 339

Butler R. P., et al., 2017, The Astronomical Journal, 153, 208

Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018, , 480, 4379

Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, , 530, A138

Casella G., Berger R. L., 1990, Statistical Inference Vol. 70. Duxbury Press Belmont, Ca

Chevallard J., Charlot S., 2016, , 462, 1415

Christensen N., Meyer R., 2000, arXiv e-prints, pp astro–ph/0006401

Christensen N., Meyer R., 2001, , 64, 022001

De Luca V., Franciolini G., Pani P., Riotto A., 2021, , 2021, 003

Dragomir D., et al., 2019, The Astrophysical Journal Letters, 875, L7

Drell P. S., Loredo T. J., Wasserman I., 2000, , 530, 593

Eadie G. M., Speagle J. S., Cisewski-Kehe J., Foreman-Mackey D., Huppenkothen D., Jones D. E., Springford A., Tak H., 2023, arXiv e-prints, p. arXiv:2302.04703

Edvardsson B., Andersen J., Gustafsson B., Lambert D. L., Nissen P. E., Tomkin J., 1993, , 275, 101

Essick R., Landry P., 2020, , 904, 80

Farr W., Sravan N., Cantrell A., Kreidberg L., Bailyn C., Mandel I., Kalogera V., 2011, in APS April Meeting Abstracts. p. H11.002

Feltzing S., Holmberg J., Hurley J. R., 2001, , 377, 911

Fischer D. A., Marcy G. W., Spronck J. F. P., 2014, , 210, 5

GRAVITY Collaboration et al., 2020, , 636, L5

Green P. J., 1995, Biometrika, 82, 711

Gregory P. C., Loredo T. J., 1992, , 398, 146

Hall A., Gow A. D., Byrnes C. T., 2020, , 102, 123524

Hernandez X., Valls-Gabaud D., Gilmore G., 2000a, , 316, 605

Hernandez X., Gilmore G., Valls-Gabaud D., 2000b, , 317, 831

Huang X.-R., Sun C.-L., Chen L.-W., Gao J., 2023, , 2023, 040

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Ibukiyama A., Arimoto N., 2002, , 394, 927

Jeffreys H., 1939, Theory of Probability. Clarendon Press, Oxford, England

Jeffreys H., 1946, Proceedings of the Royal Society of London Series A, 186, 453

Jørgensen B. R., Lindegren L., 2005, , 436, 127

Jullo E., Kneib J. P., Limousin M., Elíasdóttir Á., Marshall P. J., Verdugo T., 2007, New Journal of Physics, 9, 447

Kelly B. C., 2007, , 665, 1489

Kendall M., Stuart A., 1973, The Advanced Theory of Statistics. Vol. 2: Inference and: Relationsship. Griffin, https://books.google.co.uk/books?id=elabQwAACAAJ

Knox L., Christensen N., Skordis C., 2001, , 563, L95

Kunz M., Trotta R., Parkinson D. R., 2006, , 74, 023503

Liddle A. R., 2007, , 377, L74

Lin J., Dotter A., Ting Y.-S., Asplund M., 2018, , 477, 2966

Lindegren L., Lammers U., Hobbs D., O'Mullane W., Bastian U., Hernández J., 2012, , 538, A78

Loredo T. J., 1992, Promise of Bayesian Inference for Astrophysics. Springer New York, New York, NY, pp 275–297, doi:10.1007/978-1-4613-9290-3_31, https://doi.org/10.1007/978-1-4613-9290-3_31

Loredo T. J., 1995, PhD thesis, University of Chicago

Loredo T. J., Lamb D. Q., 1989, Annals of the New York Academy of Sciences, 571, 601

Loredo T. J., Wasserman I. M., 1998, , 502, 75

Mann A. W., Feiden G. A., Gaidos E., Boyajian T., von Braun K., 2015, The Astrophysical Journal, 804, 64

McMillan P. J., 2011, , 414, 2446

Mor R., Robin A. C., Figueras F., Lemasle B., 2017, , 599, A17

Mor R., Robin A. C., Figueras F., Antoja T., 2018, , 620, A79

Mor R., Robin A. C., Figueras F., Roca-Fàbrega S., Luri X., 2019, , 624, L1

Morris T., Podsiadlowski P., 2009, , 399, 515

Navarro-Boullosa A., Bernal A., Vazquez J. A., 2023, , 2023, 031

Ng Y. K., Bertelli G., 1998, , 329, 943

Nordström B., et al., 2004, , 418, 989

Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2004, , 612, 168

Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2006, The Astrophysical Journal, 642, 797

Pietrinferni A., Cassisi S., Salaris M., Percival S., Ferguson J. W., 2009, The Astrophysical Journal, 697, 275

Pont F., Eyer L., 2004, , 351, 487

Robert C., 2001, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer Texts in Statistics, Springer, https://books.google.co.uk/books?id=_ravDT9e8nMC

Robin A. C., Reylé C., Derrière S., Picaud S., 2003, , 409, 523

Rogantini D., Costantini E., Mehdipour M., Kuiper L., Ranalli P., Waters L. B. F. M., 2021, , 645, A98

Sale S. E., 2012, , 427, 2119

Schaller G., Schaerer D., Meynet G., Maeder A., 1992, , 96, 269

Schönrich R., Bergemann M., 2014, , 443, 698

Schönrich R., Binney J., Asplund M., 2012, , 420, 1281

Schönrich R., McMillan P., Eyer L., 2019, , 487, 3568

Schwarz G., 1978, Annals of Statistics, 6, 461

Semenova E., et al., 2020, , 643, A164

Sivia D. S., Skilling J., 2006. https://api.semanticscholar.org/CorpusID:59722824

Tak H., Mandel K., van Dyk D. A., Kashyap V. L., Meng X.-L., Siemiginowska A., 2017, The Annals of Applied Statistics, 11, 1309

Trotta R., 2008, Contemporary Physics, 49, 71

Vallenari A., Chiosi E., Sordo R., 2010, , 511, A79

Vandenberg D. A., 1985, , 58, 711

Wilk M. B., Gnanadesikan R., 1968, Biometrika, 55, 1

Zhang M., et al., 2021, The Astrophysical Journal, 922, 145

da Silva L., et al., 2006, , 458, 609

A. Metropolis-Hastings Algorithm

For a given choice for the parameter X_i and data d, the posterior PDF $P\left(X_i \mid d\right)$ is found as usual, and an iterative trial parameter sequence X' is defined by varying the original parameter equidistantly in all directions in parameter space, governed by a "proposal density" Q(X_1,X_2). This is initially a Gaussian distribution in each parameter individually, aligned to a reasonable extent with the PDF to ensure the only constraints in Q are the symmetry when swapping X_1 and X_2, the first two parameter sets in the iteration. The second phase of the algorithm is updated such that Q is constructed from the covariance matrix between X' and X_i, with the step size in any given direction conditioned by the eigenvectors of said matrix. Once the trial parameter set has been scattered by this density, the new posterior $P\left(X' \mid d\right)$ is determined - this proposed sample either accepted or rejected based on the ratio of this new distribution with the original, and its magnitude compared to a dimensionless random variable picked from a uniform distribution. After the trial parameter set has been accepted X_{i+1} is equated to X', otherwise the iterative process is begun again starting X_{i+1}.