A Preferential Attachment Model for the Initial Mass Function

Jessi Cisewski-Kehe*
Department of Statistics and Data Science, Yale University
and
Grant Weller†
Savvysherpa, Minneapolis, MN 55430
and
Chad Schafer
Department of Statistics and Data Science, Carnegie Mellon University
and
David W. Hogg
Center for Cosmology and Particle Physics, New York University

June 29, 2018

Abstract

Accurate specification of a likelihood function is becoming increasingly difficult in many inference problems in astronomy. As sample sizes resulting from astronomical surveys continue to grow, deficiencies in the likelihood function lead to larger biases in key parameter estimates. These deficiencies result from the oversimplification of the physical processes that generated the data, and from the failure to account for observational limitations. Unfortunately, realistic models often do not yield an analytical form for the likelihood. The estimation of a stellar initial mass function (IMF) is an important example. The stellar IMF is the mass distribution of stars initially formed in a given cluster of stars, a population which is not directly observable due to stellar evolution and other disruptions and observational limitations of the cluster. There are several difficulties with specifying a likelihood in this setting since the physical processes and observational challenges result in measurable masses that cannot legitimately be considered independent draws from an IMF. This

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*
†Corresponding author

work improves inference of the IMF by using an approximate Bayesian computation approach that both accounts for observational and astrophysical effects and incorporates a physically-motivated model for star cluster formation. The methodology is illustrated via a simulation study, demonstrating that the proposed approach can recover the true posterior in realistic situations, and applied to observations from astrophysical simulation data.

Keywords: Approximate Bayesian computation, astrostatistics, computational statistics, dependent data

1 Introduction

The Milky Way is home to billions of stars (McMillan 2016), many of which are members of *stellar clusters* - gravitationally bound collections of stars. Stellar clusters are formed from low temperature and high density clouds of gas and dust called *molecular clouds*, though there is uncertainty as to how the stars in a cluster form (Beccari et al. 2017). Each theory of star formation yields a different prediction for the distribution of the masses of stars that initially formed in a cluster. Hence, it is of fundamental interest to estimate this distribution, referred to as the *stellar initial mass function (IMF)*, and assess the validity of these competing theories. In fact, research advances in many areas of stellar, galactic, and extragalactic astronomy are at least somewhat reliant upon accurate understanding of the IMF (Bastian et al. 2010). For example, the IMF is a key component of galaxy and stellar evolution and planet formation (Bally & Reipurth 2005, Bastian et al. 2010, Shetty & Cappellari 2014), along with chemical enrichment and abundance of core-collapse supernovae (Weisz et al. 2013).

There is also ongoing discussion surrounding the *universality* of the IMF, i.e., if a single IMF describes the generative distribution of stellar masses for all star clusters (Bastian et al. 2010). The consensus of the astronomical community is that the IMF is not universal, however, most of the observations had been consistent with universality (Kroupa 2001, Bastian et al. 2010, Ashworth et al. 2017). With further research and growing sample sizes, however, there is increased theoretical (e.g., Bonnell et al. 2006, Dib et al. 2010) and observational (e.g., Treu et al. 2010, van Dokkum & Conroy 2010, Spiniello et al. 2014, Geha et al. 2013, Dib et al. 2017) support for an IMF that can vary cluster to cluster.

Salpeter (1955) studied the evolutionary properties of certain populations of stars, and in the process defined the first IMF (which he called the "original mass function"). This work put forth the now-classic model for the IMF, a power law with a finite upper bound equal to the physical maximum mass of a star that could form in a cluster (Salpeter 1955). More recent studies continue to use this power law form for the IMF, especially for stars of mass greater than half that of our sun (e.g., Massey 2003, Bastian et al. 2010, Da Rio et al. 2012, Lim et al. 2013, Weisz et al. 2013, 2015, Jose et al. 2017). Similar models have been proposed and used in the astronomical literature for inference of the stellar IMF; these will be discussed in the next section. The estimation of the parameters of these proposed models typically relies on the assumption that the observed stars in a stellar cluster form independently; more specifically, the assumption that the masses of the individual stars form independently. The proposed model in this work loosens the assumption of independence in order to explore one of several possible physical formation mechanisms of cluster formation. [[**Jessi:** Chad: I revised the previous sentences]]

Despite this seemingly simple form of the power law model, the statistical challenges of estimating the IMF using *observed* stars from a cluster are significant. Many of the limitations are related to observational issues and to the adequate modeling of the evolution of a star cluster after the initial formation. For example, since stars of greater mass die more rapidly, the upper tail of the IMF is not observed in a cluster of sufficient age. Also, the death of massive stars can trigger additional star formation, contaminating the lower end of the IMF with new stars (Woosley & Heger 2015). There are also issues related to missing lower-mass stars due to the sensitivity of the instruments. The observational astronomers will often estimate the *completeness function* of an observed cluster, which is the probability of observing a star of a particular mass. The completeness function is discussed in more detail below.

The observational limitations and the challenge of modeling cluster evolution make approximate Bayesian computation (ABC) appealing for estimation of the IMF, as ABC allows for relatively easy incorporation of such effects. The difficulty of addressing these limitations implied by the fact that observational effects are often ignored or accounted for in an ad-hoc or unspecified manner (e.g., Da Rio et al. 2012, Ashworth et al. 2017, Jose et al. 2017, Kalari et al. 2018), though Weisz et al. (2013) discuss how some observational limitations can be incorporated into their proposed Bayesian model. A primary appeal of ABC for this application is the ability to incorporate more complex models for cluster formation. Standard IMF models do not specify the process by which a large mass of gas (the molecular cloud) transforms into a gravitationally bound collection of stars. ABC is based on a simple rejection-sampling approach, in which draws of model parameters from a prior distribution are fed through a simulation model to generate a sample of data. If the generated sample is "close" (based on an appropriately chosen metric) to the observed data, the prior draw that produced that generated sample is retained. The collection of accepted parameter values comprise draws from an approximation to the posterior. The simulations (the *forward model*) can include any of the complex processes that make it challenging to derive a likelihood function for the observable data.

This situation is typical of inference challenges that arise in astronomy. See Schafer & Freeman (2012), Akeret et al. (2015), Ishida et al. (2015) for reviews. Recent years have seen a rapid increase in the use of ABC methods for estimation in this field, including specific application to Milky way properties (Robin et al. 2014), strong lensing of galaxies (Killedar et al. 2018, Birrer et al. 2017), large scale structure of the Universe (Hahn, Vakili, Walsh, Hearin, Hogg & Campbell 2017), estimating the redshift distribution (Herbel et al. 2017), galaxy evolution (Hahn, Tinker & Wetzel 2017), weak lensing (Peel et al. 2017, Lin & Kilbinger 2015), exoplanets (Parker 2015), galaxy morphology (Cameron & Pettitt 2012), and supernovae (Weyant et al. 2013).

[[**Jessi:** Goals of the proposed stochastic model and ABC algorithm for inference on the stellar IMF include the following. First, the proposed generative model connects the observable data to a possible star cluster formation mechanism that induces a dependency in the masses of the stars. Second, this model generalizes commonly used IMF models in the sense that it can capture, but also distinguish, two popular competing IMF model shapes. This flexibility can help by not requiring model selection to determine the IMF shape since the proposed model can adapt. And finally, by using an ABC approach, observation effects and uncertainties can easily be incorporated into the analysis]]

This paper is organized as follows. In Section 2, background on the IMF along with inference challenges are presented along with an introduction of ABC. The proposed stochastic model for stellar formation is discussed in Section 3 and the ABC procedure along with a simulation study in Section 4. Section 5 applies the proposed methodology to the estimation of the IMF of a realistic astrophysical simulation (Bate 2012). Finally, Section 6 provides a discussion.

2 Background

2.1 Stellar Initial Mass Function

As noted above, Salpeter (1955) introduced the power law model for the shape of the IMF for masses larger than $0.5M_{\odot}$, where $M_{\odot}$ is the mass of the Sun. Kroupa (2001) extended the range of the IMF by proposing a three-part broken power law model over the range $0.01M_{\odot} < m < M_{\max}$, where $M_{\max}$ is the mass of the largest star that could form with nonzero probability. This model postulates different forms for the IMF for stars of masses $0.01M_{\odot} < m < 0.08M_{\odot}$, $0.08M_{\odot} < m < 0.5$ and $m > 0.5M_{\odot}$. Focusing on the upper part, and defining $\theta = (\alpha, M_{\max})$, the probability density function for mass x in the upper tail of the stellar IMF is assumed to be given by

$$f_M(m \mid \theta) = cm^{-\alpha}, \;\; m \in [M_{\min}, M_{\max}], \tag{1}$$

where the constant c is chosen such that f_M is a valid probability density. Alternative models have been proposed that include log-normal distributions, joint power law and log-normal parts, and truncated exponential distributions (Chabrier 2003*b*,*a*, 2005, Corebelli et al. 2005, Bastian et al. 2010, Offner et al. 2014). The Kroupa (2001) and Chabrier (2003*b*,*a*) models are displayed in Figure 1 along with observational challenges discussed §2.1.1. Power law distributions and log-normal distributions are closely related and may be the result of subtle differences in the underlying formation mechanism (Mitzenmacher

2004). The IMF model we propose will include, as a special case, a family of formation mechanisms that generate power law tails, but also allow for a wider range of tail behaviors (see Section 3.1.1).

2.1.1 Observational Challenges

Observing all stars comprising an IMF is not feasible, as the most massive stars ($m > 10M_{\odot}$) have lifetimes of only a few million years. The lifetime of a star (the time it takes for the star to burn through its hydrogen) depends strongly on its mass: the most massive stars have shorter lives due to the hotter temperatures they must maintain to avoid collapse from the strong gravitational forces. In particular, stellar life is approximately proportional to $m^{-\rho}$ where $\rho \approx 3$ (Hansen et al. 2004, p. 30, Chaisson & McMillan 2011, p. 439). Hence, the mass of the largest star observed in a given cluster is depends on the cluster age.

Furthermore, the IMF will be estimated using a noisy, incomplete view of that cluster. Whether or not a star is observed is dependent on several factors including its mass, its location in the cluster, and its neighbors. Some of these factors are described by a data set's *completeness function*, which quantifies a given star's probability of being observed. This depends on its luminosity (i.e. intrinsic brightness) since it needs to be sufficiently bright to be observable; in particular, completeness depends on stellar flux in comparison with the flux limits of the observations. There are also issues with *mass segregation*: stars with lower mass tend to be on the edge of the cluster, while the most massive stars are often found in the center (Weisz et al. 2013). Due to *stellar crowding* in the center, stars in this region can be more difficult to observe. Additionally, binary stars (star systems consisting of a pair of stars) are difficult to distinguish from a single star, creating the potential for overstating the mass of an object and understating the number of stars in the cluster.

There are additional uncertainties involved in translating the actual observables (e.g. photometric magnitudes) into a mass measurement; that is, the mass values for observable stars are only estimates. The *Hertzsprung-Russell (H-R) Diagram* is a classic visual summary of the distribution of the luminosity and temperature of a collection of stars. A typical H-R Diagram includes a *main sequence* of stars that trace a line from bright and hot

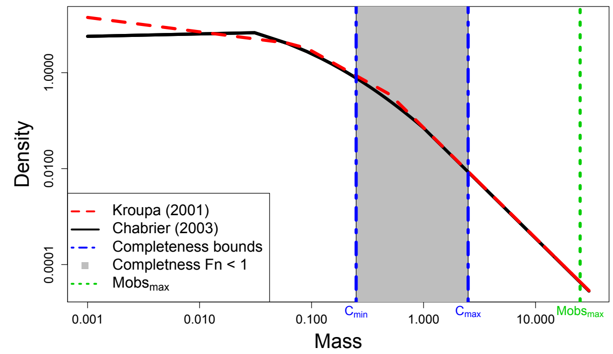


Figure 1: The broken power law model of Kroupa (2001) (red, dashed) and the lognormal with a power law tail model of Chabrier (2003b,a) (black, solid) are displayed along with vertical lines representing several observational challenges. The blue vertical dotted-dashed lines indicate the range of values ($C_{\min}, C_{\max}$) on which the completeness function may be defined, and the vertical green dotted line indicates the maximum observable mass ($Mobs_{\max}$) due to the aging cutoff. The observational challenges are discussed further in §3.2.

stars to dim and cool stars. Stellar mass also evolves along this one-dimensional feature, and since luminosity and temperature are estimable, mass can thus also be estimated. The mass of binary stars can be determined via Kepler's Laws, and hence a *mass-luminosity relationship* can be fit to binaries and then extended to other stars on the main sequence. Unfortunately, luminosity and temperature are nontrivial to estimate, as corrections for effects such as *accretion* and *extinction* are required, along with an accurate estimate of the distance to the stars (Da Rio et al. 2010). The process is further complicated by the dependence of how these transformations are made on the *spectral type* of the star. Careful budgeting of the errors that accumulate is required in order to produce a reasonable error bars on mass estimates; Da Rio et al. (2010) utilize a Monte Carlo approach in which the errors in magnitudes are propagated forward through to uncertainties in the spectral type, the accretion and reddening corrections, and finally to an uncertainty on the mass.

2.2 Approximate Bayesian Computation

Standard approaches to Bayesian inference, either analytical or built on MCMC, require the specification of a likelihood function, $f(m \mid \boldsymbol{\theta})$, with data $m_{\text{obs}} \in \mathcal{D}$, and parameter(s) $\boldsymbol{\theta} \in \Theta$. In many modern scientific inference problems, such as for some emerging models for the stellar IMF, the likelihood is too complicated to be derived or otherwise specified. As noted previously, ABC provides an approximation to the posterior without specifying a likelihood function, and instead relies on forward simulation of the data generating process.

The basic algorithm for sampling from the ABC posterior is attributed to Tavaré et al. (1997) and Pritchard et al. (1999), used for applications to population genetics. The algorithm has three main steps which are repeated until a sufficiently large sample is generated: *Step 1*, Sample $\boldsymbol{\theta}^*$ from the prior; *Step 2*, Generate m_{sim} from forward process assuming truth $\boldsymbol{\theta}^*$; *Step 3*, Accept $\boldsymbol{\theta}^*$ if some distance function $\rho(m_{\text{obs}}, m_{\text{sim}}) \leq \epsilon$, where ϵ is a tuning parameter that should be close to 0. This last step typically consists of comparing low-dimensional summary statistics generated for the observed and simulated datasets. Adequate statistical and computational performance of ABC algorithms depends greatly on the selection of such summary statistics (Joyce & Marjoram 2008, Blum & François 2010, Blum 2010, Fearnhead & Prangle 2012, Blum et al. 2013).

The basic ABC algorithm can be inefficient in cases where the parameter space is of moderate or high dimension. Hence, important adaptations of the basic ABC algorithm incorporate ideas of sequential Monte Carlo (SMC) in order to improve the sampling efficiency (Marjoram et al. 2003, Sisson et al. 2007, Beaumont et al. 2009, Del Moral et al. 2011). A nice overview of ABC can be found in Marin et al. (2012). Here, we use a sequence of decreasing tolerances $\epsilon_{1:T} = (\epsilon_1, \ldots, \epsilon_T)$ with the tolerance ϵ_t shrinking until further reductions do not significantly affect the resulting ABC posterior. The improvement in efficiency is due to the modification that happens after the first time step: instead of sampling from the prior distribution, the proposed $\boldsymbol{\theta}$ are drawn from the previous time step's ABC posterior. Using this adaptive proposal distribution can help to improve the sampling efficiency. The resulting draws, however, are not targeting the correct posterior, and so importance weights, W_t, are used to correct this discrepancy.

3 Forward Model for the IMF

Due to their simple interpretations, mathematical ease, and demonstrated consistency with observations, power law IMFs (or similar variants) have been widely adopted in the astronomy literature (Kroupa et al. 2012); however, open questions remain about stellar formation processes. The proposed forward model is a way to link a possible stellar formation process with the realized mass function (MF). One known underlying mechanism for producing data with power-law tails is based on *preferential attachment* (PA) (Mitzenmacher 2004). The earliest PA model, the Yule-Simon process, was popularized by Simon (1955), and was originally used to model biological genera and word frequencies. Next the proposed data generating model using ideas of preferential attachment are presented for the stellar IMF.

3.1 Preferential attachment for the IMF

The formation of a star cluster is a complicated and turbulent process with different theories on the physical processes that lead to the origin of the stellar IMF (Chabrier 2005, Bate 2012, Offner et al. 2014, Pokhrel et al. 2018). It is generally understood that the molecular cloud fragments and then forms stellar cores with a distribution referred to as the *core mass function* (CMF). Whether evolution from the CMF to the IMF is random, deterministic, or something in between is debated (Offner et al. 2014). In the proposed work, we consider the case where star cores can increase in mass by accreting material from the surrounding cloud and a particular star's final mass can be affected by its neighbors through turbulence or dynamical interactions. That is, rather than assuming that stellar masses in a cluster arise independently of each other, our PA model proposes a resource-limited mass accretion process between stellar cores whose ability to accumulate additional mass is a function of their existing masses. This dependence feature is particularly important for statistical inference, as models that assume independent observations of stellar masses are vulnerable to incorrect and misleading inference. Additionally, the mass of the largest star to form in a cluster is limited by the total cluster mass.

Our proposed stochastic model for stellar formation is as follows: we first fix a total available cluster mass M_{tot}. This quantity can be physically interpreted as the total mass available for stellar formation in a molecular cloud. At each time step $t = 1, 2, \ldots$, a random quantity of mass $m_t \sim Exponential(\lambda)$ enters the collection of stars; $M_{1,1} = m_1$ becomes the mass of the first star. Subsequent masses entering the system form a new star with probability π_t or join existing star $k = 1, \ldots, n_t$ with probability π_{kt}. These probabilities are specified as

$$\pi_t = \min(1, \alpha) \quad \text{and} \quad \pi_{kt} \propto M_{k,t}^{\gamma}. \tag{2}$$

The generating process is complete when the total mass of formed stars reaches M_{tot}. The possible ranges of the three parameters are $\lambda > 0$, $\alpha \in [0, 1)$, and $\gamma > 0$.

The parameter α defines the probability that entering mass forms a new star in a cluster. For the growth component, the model allows for linear ($\gamma = 1$), sublinear ($\gamma < 1$), and superlinear ($\gamma > 1$) behavior; the limiting case of $\gamma \to 0$ gives a uniform attachment model. Finally, the parameter λ acts as a scaling factor which controls the average 'coarseness' of masses joining the forming stellar cores.

The proposed PA mass generation model offers considerable flexibility to approximate existing IMF models in the literature. To illustrate the generality of the proposed model, IMF realizations were drawn assuming the Kroupa (2001) broken power-law model as the true model, defined as

$$f(m) \propto \begin{cases} m^{-0.3}, & m \leq 0.08 \\ k_1 \cdot m^{-1.3}, & 0.08 < m \leq 0.5 \\ k_2 \cdot m^{-2.3}, & m > 0.5, \end{cases} \tag{3}$$

and the Chabrier (2003*b*,*a*) lognormal model, defined as

$$f(m) \propto \begin{cases} \frac{0.158}{m} \times \exp\left(-\frac{(\log_{10}(m) - \log_{10}(0.079))^2}{2(0.69)^2}\right), & m \leq 1 \\ k_3 \cdot m^{-2.3}, & m > 1, \end{cases} \tag{4}$$

where constants k_1, k_2, and k_3 are defined to make the densities continuous. Our proposed PA ABC procedure was then used for inference and Figure 2 displays the resulting posterior predictive IMFs. The proposed model captures the general shape of the true model. Figures 3a - 3c display ABC marginal posteriors for the broken power-law model of Kroupa

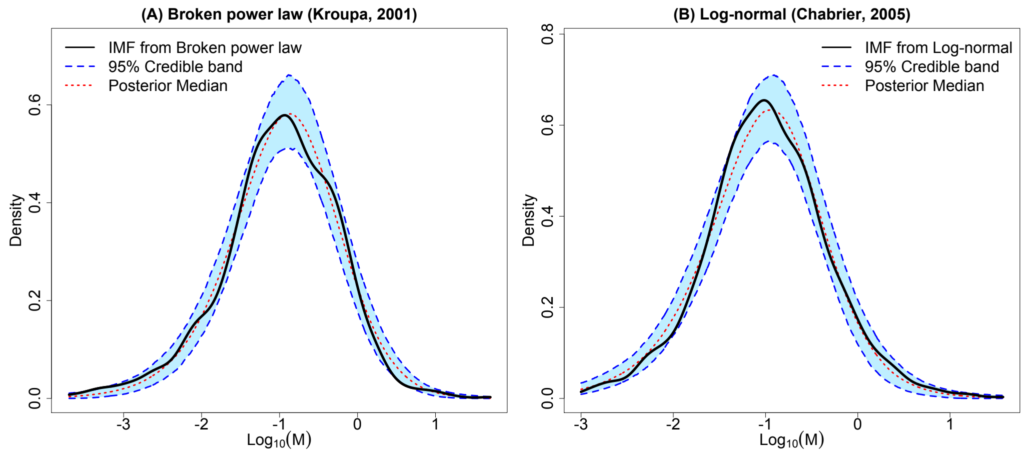
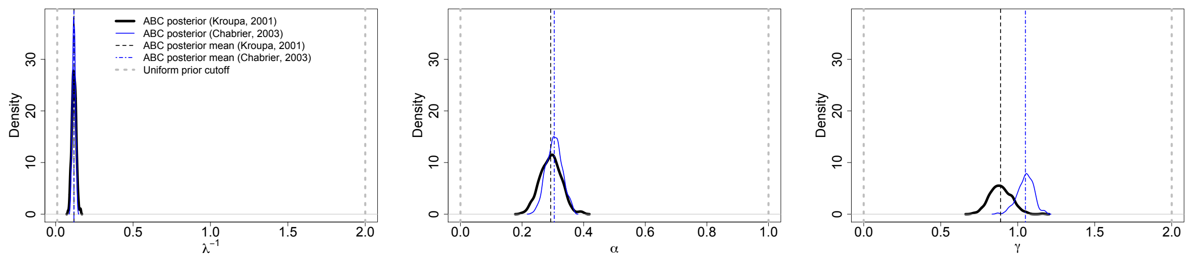


Figure 2: The solid black lines display the IMF with 1,000 stars simulated from (A) the broken power-law model of Kroupa (2001) (see Eq (3)*) and (B) the log-normal model of Chabrier (2003*b,a*) (see Eq* (4)*). The proposed PA ABC model was used with* $N = 1,000$ *particles, and 95% point-wise credible bands are displayed (blue, dashed lines) along with the posterior median (red, dotted) for each data set. The preferential attachment model provides flexibility to approximate both existing models. Both models have similar ABC posterior means for* α *(0.293 and 0.304, respectively). However, the ABC posterior means for* γ *are notably different. The broken power-law model has an ABC posterior mean of 0.889 while the estimate for the lognormal model is 1.050.*

(2001) and the lognormal model of Chabrier (2003*b*,*a*). Both the broken power-law and lognormal models have similar ABC posterior means for α (0.293 and 0.304, respectively). However, the ABC posterior means for γ are notably different. The broken power-law model has an ABC posterior mean of 0.889 while the estimate for the lognormal model is 1.050. Since the Kroupa (2001) and Chabrier (2003*b*,*a*) models use the same power-law slope for masses greater than 0.5 $M_\odot$ and 1 $M_\odot$, respectively, this suggests that the differences in γ are due to differences in the shape of the lower-mass end of the IMF. The proposed model offers an approach for discriminating these models.



(a) ABC posteriors for λ^{-1} *(b) ABC posteriors for* α *(c) ABC posteriors for* γ

Figure 3: Marginal ABC posteriors for data generated from the broken power-law model of Kroupa (2001) (thick black lines) and for data generated from the lognormal model of Chabrier (2003b,a) (thin blue lines). The vertical dashed black lines indicate the ABC posterior mean for the Kroupa (2001) model, vertical dashed and dotted blue lines indicate the ABC posterior mean for the Chabrier (2003b,a) model, and the dotted gray lines indicate the range of the uniform prior for the parameter.

3.1.1 Generating power law tails

As noted above, the PA model with linear evolution (the Yule-Simon Process) is known to generate power law tails (Newman 2005). It is worth exploring the extent to which power law tail behavior is present in cases where $\gamma \neq 1$, as the power law model is such a prevalent assumption in this application. For example, it would be of interest to determine if tests of $H_0\colon \gamma = 1$ would have power to detect deviation from power law tails.

A small simulation study was conducted. Goodness-of-fit was assessed using the standard Kolmogorov-Smirnov statistic, with the empirical distribution of the masses of a collection of stars generated from our PA model compared to the best fitting power law model. As we are only interested in fitting to the upper tail, this analysis is restricted to the region above $1M_\odot$. We fix $\lambda^{-1} = 0.25$ and $M_{\text{tot}} = 1000$, and consider $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$, for values of γ ranging from 0.25 to 5. Fifty data sets are generated for each (α, γ) combination. Results are shown in Figure 4. In order to place the goodness-of-fit on a readily-interpretable scale, the p-value is calculated for each K-S test, and the median across the 50 repetitions is shown. The results support the claim that the tail follows the power law when $\gamma = 1$, but that the power law fit degrades quickly for γ outside $(0.5, 1.5)$. The effect is particularly strong for smaller values of α.

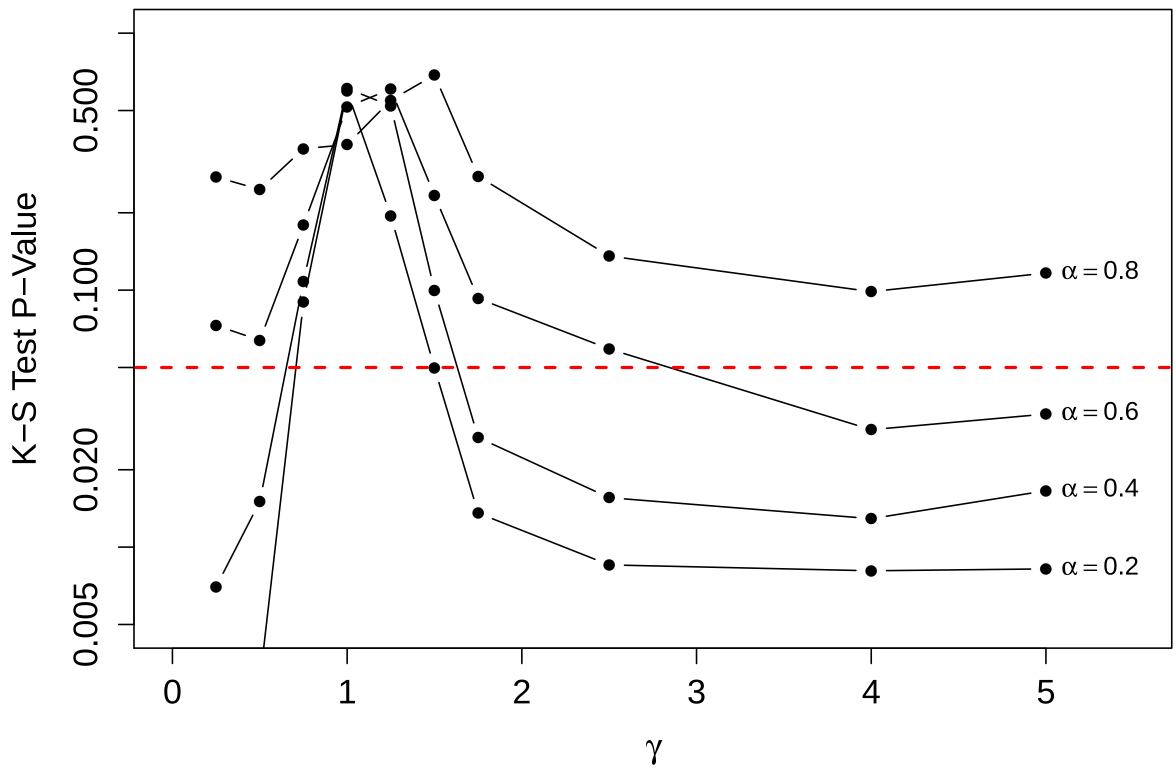


Figure 4: Median p-values from Kolmogorov-Smirnov tests when comparing the tail distribution of masses simulated from the proposed PA model with the best-fitting power law model. Fifty repetitions are done at each (α, γ) *combination. The 0.05 cutoff is shown as a guide. Note that the vertical axis is on the log-scale.*

3.2 Initial Mass Function to the Observed Mass Function

The PA model describes the formation of a star cluster at initial formation. However, we are not generally able to observe the star cluster after initial formation due to observational uncertainties, measurement uncertainties, and aging and dynamical evolution of the cluster. A cluster's present-day observed mass function (MF) is the observed distribution of the stellar masses of a particular cluster.

Observation limitations can be easily incorporated into the ABC framework. For simplicity, we adopt a "linear ramp" completeness function describing the probability of observing a star of mass m:

$$\Pr(\text{observing a star} \mid \text{m}) = \begin{cases} 0, & m < C_{\min} \\ \frac{m - C_{\min}}{C_{\max} - C_{\min}}, & m \in [C_{\min}, C_{\max}] \\ 1, & m > C_{\max}. \end{cases} \tag{5}$$

We assume that the values $C_{\min}$ and $C_{\max}$ are known, though we note that selecting an appropriate completeness function is a difficult process which requires quantification from the observational astronomers for each set of data. Different models for the completeness function could also be considered, including those which allow for spatially-varying observational completeness. A benefit of ABC is the ease at which a new completeness function can be incorporated – it amounts to a simple change in the forward model.

Due to measurement error and practical limitations in translating luminosities into masses, the masses of stars are not perfectly known. This uncertainty can be incorporated in different ways; following Weisz et al. (2013), we assume that the inferred mass of a star m_i is related to its true mass M_i via

$$\log m_i = \log M_i + \sigma_i \eta_i, \tag{6}$$

where η_i is a standard normal random variable, and σ_i is known measurement error. The model for mass uncertainties in (6) is simple and could be extended to account for other sources of uncertainty (e.g. redshift).

As noted previously, the lifecycle of a star depends on certain characteristics such as mass. In the proposed algorithm, stars generated in a cluster are aged using a simple truncation of the largest masses. That is, the distribution of stellar masses for a star cluster of age τ Myr is given by

$$f_M(m \mid \theta, \tau) = f_M(m \mid \theta) \mathbb{I}\{M \leq \tau^{-1/3} \times 10^{4/3}\}, \tag{7}$$

corresponding to stellar lifetimes of $10^4\ M^{-3}$ Myr, where M is the mass of the star (Hansen et al. 2004, Chaisson & McMillan 2011). More sophisticated models that account for effects such as binary stars and stellar wind mass loss can be inserted into this framework.

4 Methods

We propose an ABC framework to make inferences on the IMF given a cluster's present-day observed MF. The proposed ABC algorithm is displayed in Algorithm (1), where N is the desired particle sample size to approximate the posterior distribution, and is motivated by the adaptive ABC algorithm of Beaumont et al. (2009). The forward model, F, in Algorithm (1) is where observational limitations and uncertainties, stellar evolution, and other astrophysical elements can be incorporated as outlined in Section 3.

4.1 Proposed ABC algorithm

Algorithm (1) is initialized using the basic ABC rejection algorithm at time step $t = 1$ using a distance function $\rho(m_{\text{sim}}, m_{\text{obs}})$ to measure the distance between the simulated and observed datasets, m_{sim} and m_{obs}, respectively. The first tolerance, ϵ_1, is adaptively selected by drawing kN particles for some $k > 0$. Then the N particles that have the smallest distance are retained, and ϵ_1 is defined as the largest of those N distances retained. For subsequent time steps ($t > 1$), rather than proposing a draw, θ^*, from the prior, $\pi(\theta)$, the proposed θ^* is selected from the previous time step's ($t - 1$) ABC posterior samples. The selected θ^* is then moved according to some kernel, $K(\theta^*, \cdot)$, to ameliorate degeneracy as the sampler evolves. In order to ensure the true posterior (which requires sampling from the prior) is targeted, the retained draws are weighted according to the appropriate importance weights, W_t – this incorporates the proposal distribution's kernel.

A key step in the implementation of an ABC algorithm is to quantify the distance between the simulated and observed stellar masses. We define a bivariate summary statistic and distance function that captures the shape of the present-day MF and the random number of stars observed, displayed in Equations (8) and (9), respectively. For the shape of the present-day MF, we use a kernel density estimate of the $\log_{10}$ masses (due to the heavy-tailed distribution of the initial masses), and an L_2 distance between the simulated and observed $\log_{10}$ mass function estimates. The number of stars observed is the other summary statistic, with the distance being the absolute value of the difference in the ratio of the counts from 1. The bivariate summary statistic is defined as

$$\rho_1(m_{\text{sim}}, m_{\text{obs}}) = \left[\int \left\{\hat{f}_{\log m_{\text{sim}}}(x) - \hat{f}_{\log m_{\text{obs}}}(x)\right\}^2 dx\right]^{1/2} \tag{8}$$

$$\rho_2(m_{\text{sim}}, m_{\text{obs}}) = \max\left\{\left|1 - n_{\text{sim}}/n_{\text{obs}}\right|, \left|1 - n_{\text{obs}}/n_{\text{sim}}\right|\right\}, \tag{9}$$

where the $\hat{f}$ are kernel density estimates, and n_{sim} and n_{obs} are the number of stars comprising the observed MF. These summary statistics were selected based on performance of a simulation study using the high-mass section of the broken power-law model because the true posterior is known in this setting. Results and additional discussion of the simulation study can be found in Appendix A.

With the bivariate summary statistic, we use a bivariate tolerance sequence, $(\epsilon_{1t}, \epsilon_{2t})$, for $t = 1, \ldots, T$ is such that $\epsilon_{i1} \geq \epsilon_{i2} \geq \cdots \geq \epsilon_{iT}$ for $i = 1, 2$. At time step t, the tolerances are determined based on the empirical distribution of the retained distances from time step $t-1$ (e.g. the $25th$ percentile). As noted previously, the tolerance sequence is initialized adaptively by selecting kN proposals from the prior distributions, then the N proposals that result in the N smallest distances were selected.[1]

In practice, M_{tot} is an unknown quantity of interest. A prior can be assigned to M_{tot} and an additional summary statistic and tolerance sequence can be used. The summary statistics we select in this case is

$$\rho_3(m_{\text{sim}}, m_{\text{obs}}) = \left|\sum_{i=1}^{n_{\text{sim}}} m_{\text{sim},i} - \sum_{j=1}^{n_{\text{obs}}} m_{\text{obs},j}\right|, \tag{10}$$

where $m_{\text{sim},i}$ and $m_{\text{obs},j}$ are the masses of the individual simulated and observed stars, respectively.

[1]The kN sampled distances were scaled, squared, and then added together; the N smallest of these combined distances were retained.

4.2 Simulation study

A short simulation study was carried out to illustrate the performance of the proposed model in the case where the solution is known, using the summary statistics in Equations (8)-(10). Values for the parameters $(\lambda^{-1}, \alpha, \gamma, M_{\text{tot}})$ are displayed in Table 1 along with the number of stars produced, n_{obs}, which are used in the simulation as the observations. The generated IMF's used as the observations for each setting are displayed in Figure 5; number counts rather than densities are displayed in order to emphasize the different number of stars produced under each setting.

One of the more notable differences in the simulated IMFs displayed in Figure 5 used as the observations in the simulation study is the number of stars generated. The α parameter controls the numbers stars that form and so there appears to be two groups of IMFs: one with $\alpha = 0.3$ and one with $\alpha = 0.7$.

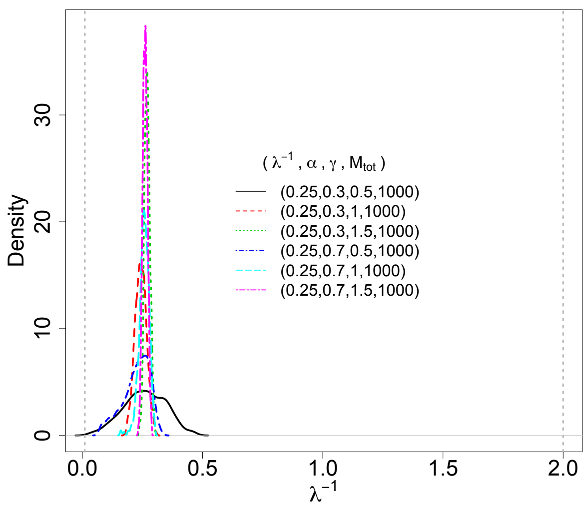
The resulting marginal ABC posteriors for λ^{-1}, α, γ, and M_{tot} are displayed in Figure 6, and the pairwise ABC posterior samples are displayed in Figure 7. For λ^{-1} and M_{tot} (Figure 6a and 6d, respectively) cover the input values well. In general, the marginal ABC posteriors for α and γ (Figure 6b and 6c, respectively) cover the input values of the parameters, but their shapes and widths vary. For example, the width of the marginal posteriors for α appear to depend on the corresponding value of γ, where $\gamma = 0.5$ results in the broadest marginal ABC posteriors for the same α input, and $\gamma = 1.5$ results in the narrowest marginal ABC posteriors for the same α input. Figure 7c provides some incite into this behavior. It appears that the apparent difference in spread of the marginal distributions for α is due to its projection; the joint posteriors for α and γ suggest that their variability changes orientation as the values for α and γ change. There appears to be a degeneracy between λ^{-1} and α, as displayed in Figure 7a. This may be because λ^{-1} controls the average size of the mass pieces and α controls how many of those mass pieces form new stars; a bigger average mass piece (corresponding to *low* values of λ^{-1}) means that fewer new stars can form because the M_{tot} gets used up more quickly.

The marginal posteriors for $\gamma \leq 1$ appear to be narrower with smaller α. A smaller α results in fewer new stars, which means the entering piece of mass has to join an already formed star, at a rate controlled by the value of γ. Hence, more pieces that have to be assigned to existing stars seems to result in more information about the rate of growth of the stars. However, for a $\gamma > 1$, a larger α (producing more new stars) appears to lead to a narrower marginal ABC posterior for γ than when α is smaller; the marginal ABC posterior for $\gamma = 1.5$ corresponding to the smaller value for α (0.3) is wider and pushes against the boundary of the prior. It turns out that this behavior seems to occur because when γ is greater than 1 and α is smaller, all the mass that has to be distributed to the already existing stars (rather than forming a new star) tends to be assigned to the same, most massive star. There ends up a being a runaway effect in this scenario ($\gamma > 1$ and smaller α) where almost all of the incoming mass is assigned to the same star, resulting in one extremely massive star. This means there is essentially only one star for providing information on the $\gamma > 1$, and the higher values of γ are not distinguished, which may be why the ABC posterior pushes against the upper boundary of the prior on γ. Using a summary statistic in the ABC algorithm isolating the most massive star may be a helpful technique in general; however, in the stellar IMF setting considered, the most massive star is generally not observed due to the aging of the cluster (see Section 3.2).

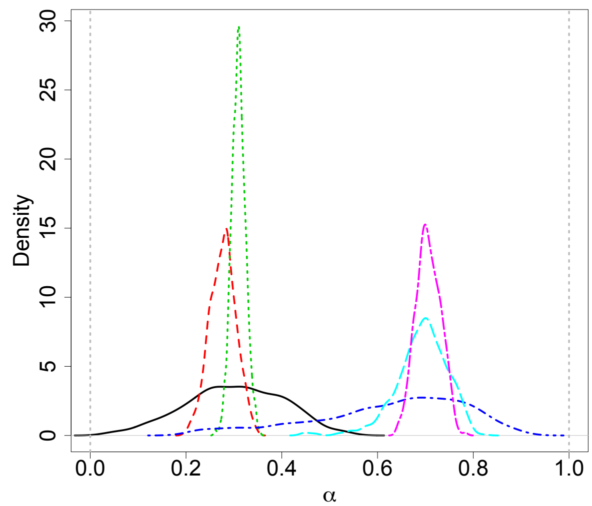
4.2.1 Simulated data with observational effects

[[**Jessi:** running more simulations for this section]]

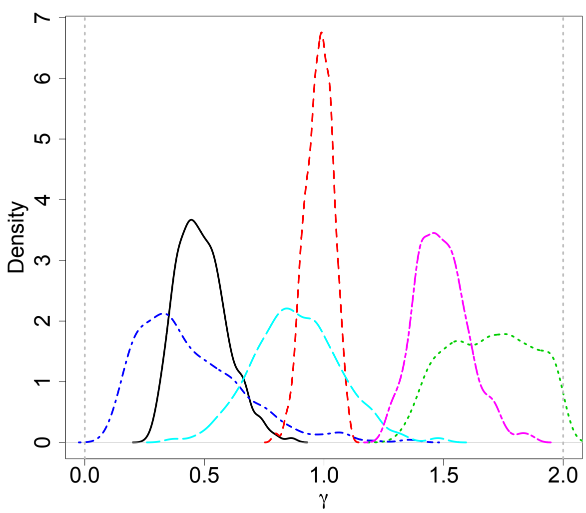
Next we consider a simulation setting where aging, completeness, and measurement error are incorporated in order to analyze these effects on the resulting inference on the IMF. The same IMF is used in the following simulations with $\lambda^{-1} = 0.25$, $\alpha = 0.3$, $\gamma = 1$, and $M_{\text{tot}} = 1000$, but we vary the range of the linear ramp completeness function of Equation (5); $C_{\min}$ is fixed at $0.08M_{\odot}$ and $C_{\max} \in \{0.10, 0.25, 0.5, 0.75, 1\}M_{\odot}$ where low values of $C_{\max}$ result in fewer stars removed from the IMF and, hence, a larger number of stars in the MF. All five sets of MFs are age 30 Myr and have lognormal measurement error with $\sigma = 0.25$.



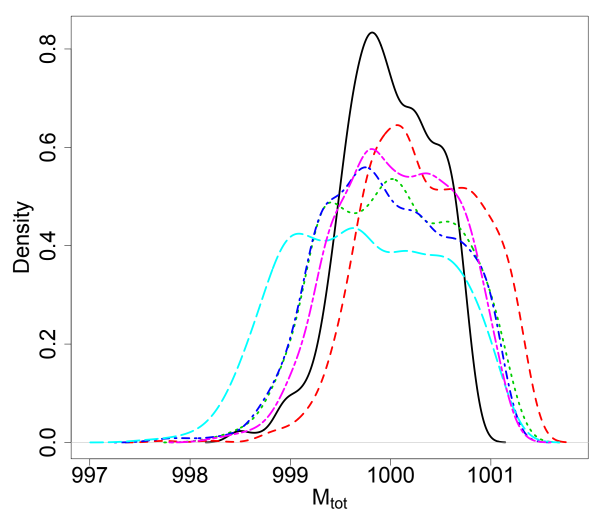
(a) ABC marginal posterior for λ^{-1}



(b) ABC marginal posterior for α

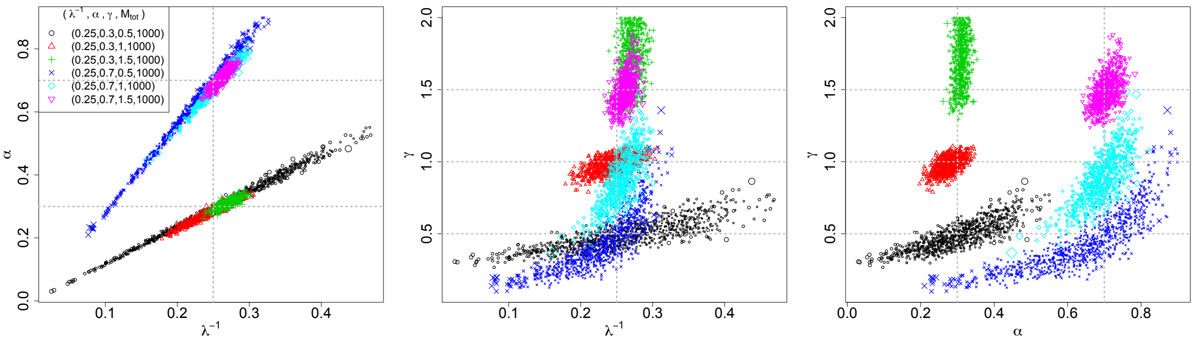


(c) ABC marginal posterior for γ



(d) ABC marginal posterior for M_{tot}

Figure 6: Marginal ABC posteriors for the simulation settings of Table 1. The different color and types of lines indicate the input parameter values corresponding to the weighted kernel density estimates of the marginal ABC posteriors for (a) λ^{-1}, (b) α, (c) γ, and (d) M_{tot}. The vertical dotted gray lines in plots (a) - (c) indicate the range of the uniform prior for the parameter. The prior used for M_{tot} was a Normal distribution with a prior mean of 1200 and a prior standard deviation of 600.



(a) ABC Joint (λ^{-1}, α) *(b) ABC Joint* (λ^{-1}, γ) *(c) ABC Joint* (α, γ)

Figure 7: Pairwise ABC posterior particles samples of (a) (λ^{-1}, α)*, (b)* (λ^{-1}, γ)*, and (c)* (α, γ) *for the simulation settings of Table 1. The different color and types of points indicate the input parameter values, and the size of the plot symbol is scaled with the particle weight.*

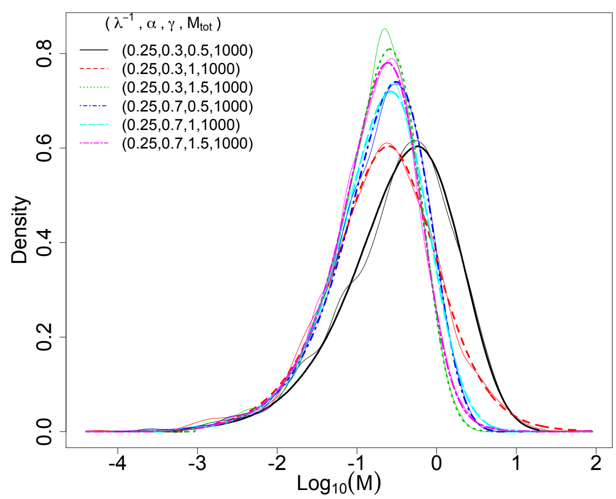
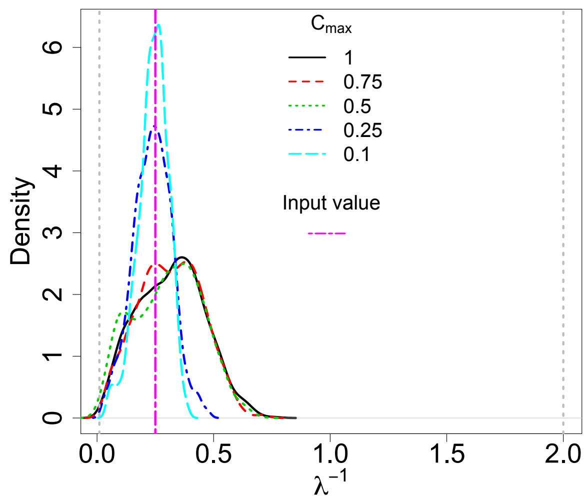
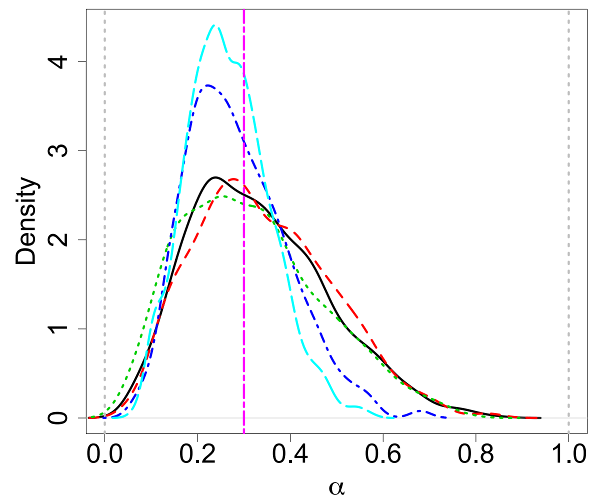


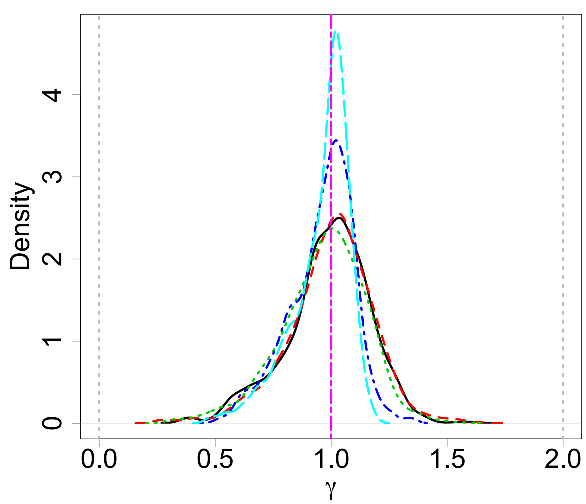
Figure 8: ABC posterior predictive IMFs for the simulation settings of Table 1. The different color and line types indicate the median ABC posterior predictive IMF for the various input parameter values combinations, $(\lambda^{-1}, \alpha, \gamma, M_{tot})$*; the thin solid lines of different colors display the observed IMF associated with the input parameter value combination with the matching color (see Figure 5). The posterior predictive IMFs were derived from taking 1000 draws from the final ABC posteriors in each setting, and then simulating an IMF using the sampled values.*



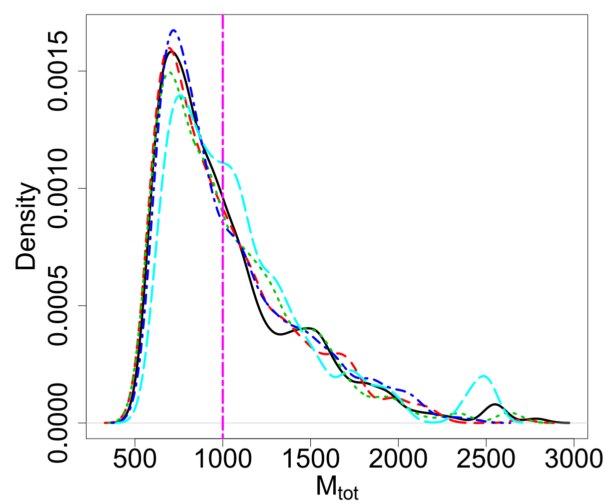
(a) *ABC marginal posterior for* λ^{-1}



(b) *ABC marginal posterior for* α



(c) *ABC marginal posterior for* γ



(d) *ABC marginal posterior for* M_{tot}

*Figure 9: [[**Jessi:** need to update plots]] Marginal ABC posteriors for the simulation setting of Section 4.2.1. The different color and types of lines indicate the differing upper limits of the linear ramp completeness function of Equation* (5), C_{max}, *corresponding to the weighted kernel density estimates of the marginal ABC posteriors for (a)* λ^{-1}, *(b)* α, *(c)* γ, *and (d)* M_{tot}. *The lower limit,* C_{min}, *is fixed at* $0.08M_{\odot}$. *All five datasets started with the same IMF using* $\lambda^{-1} = 0.25$, $\alpha = 0.3$, $\gamma = 1$, *and* $M_{tot} = 1000$, *were aged 30 Myr, and had lognormal measurement error applied with* $\sigma = 0.25$.

*Figure 10: [[**Jessi:** need to update plots]] Pairwise ABC posterior particles samples of (a) (λ^{-1}, α), (b) (λ^{-1}, γ), and (c) (α, γ) for the simulation settings of Section 4.2.1. The different color and types of points indicate the C_{max} values of the linear ramp completeness function, and the size of the plot symbol is scaled with the particle weight. The lower limit, C_{min}, is fixed at $0.08M_{\odot}$. All five datasets started with the same IMF using $\lambda^{-1} = 0.25$, $\alpha = 0.3$, $\gamma = 1$, and $M_{tot} = 1000$, were aged 30 Myr, and had lognormal measurement error applied with $\sigma = 0.25$.*

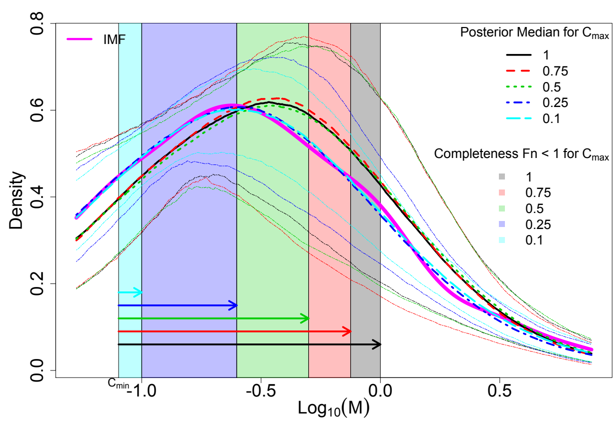


*Figure 11: [[**Jessi:** need to update plots]] Posterior predictive IMF for the simulation settings of Section 4.2.1. The thick magenta line is the true IMF, the thicker lines of varying line type and color are the ABC posterior predictive median IMF for the different values of C_{max}, the shaded regions are the different ranges of completeness (all starting at $C_{min} = 0.08M_{\odot}$), and the thin lines define a 95% credible band for the C_{max} with the matching color. For the posterior predictive IMF, 1000 draws were made from the ABC posteriors of Figure 9 and then 1000 cluster samples were drawn from the proposed forward model.*

5 Astrophysical simulation data

Next we consider a star cluster generated from a radiation hydrodynamical simulation presented in Bate (2012). This simulation resulted in 183 stars and brown dwarfs with a total mass of the resulting objects of about 88.68 $M_\odot$ formed from a 500 $M_\odot$ molecular cloud of uniform density. Understanding that simulations are only an approximation of reality, this astrophysical simulation was implemented to include realistic physics of star cluster formation such as a radiative feedback. The technical details of the simulation are beyond the scope of this work, but can be found in Bate (2012). Figure 12 displays the resulting IMF as a density and histogram.

Validation of the simulated cluster was carried out by comparing its IMF with the model of Chabrier (2005), and was not able to statistically differentiate them using a Kolmogorov-Smirnov test. The Chabrier (2005) IMF is displayed in Figure 12 as a comparison to the simulation data. While the general shape does appear to match well, the Bate (2012) data has a small second mode around $1M_\odot$. The Bate (2012) data seems to have more objects on the lower mass end and fewer between 0.5 and $1M_\odot$ than expected with the Chabrier (2005) IMF model. Additionally, because the shape of the low-mass end of the IMF is not well-constrained observationally, Bate (2012) compares the ratio of number of brown dwarfs to number of stars with masses $< 1M_\odot$ and finds acceptable agreement with observations. Bate (2012) also carryout an analysis of the mechanism(s) behind the shape of the IMF. They found that larger objects have had longer accretion times, while lower mass objects tended to have a dynamical encounter that resulting in the accretion terminating; hence there ended up being, as Bate (2012) described, a "competition between accretion and dynamical encounters." This competition for material seems consistent with the ideas underlying the proposed preferential attachment model.

The 183 objects were used as the observations in the proposed ABC algorithm using 1000 particles, 5 sequential time steps, a kN of 10^4 (for adaptively initializing the algorithm), and the 25th percentile for shrinking the sequential tolerances based on the empirical distribution of the retained distances from the preceding time step. The resulting ABC marginal posteriors are displayed in Figure 13, the pairwise ABC joint posteriors in Figure 14, and the posterior predictive IMF in Figure 15. The ABC posterior means for λ^{-1}, α, and γ are 0.260, 0.537, and 1.091, respectively. The ABC posterior mean of α is notably higher than the ABC posterior means of α for the Kroupa (2001) (0.293) and Chabrier (2003b,a) (0.304) simulated data discussed in Section 3.1 (see Figure 2). The ABC posterior mean of γ is also slightly higher than the 1.050 posterior mean of the Chabrier (2003b,a) data. Though the IMF has a slightly irregular shape with a small second mode around 1 $M_{\odot}$ as noted previously, the proposed ABC method's posterior predictive median and 95% predictive bands generally fit the IMF shape well.

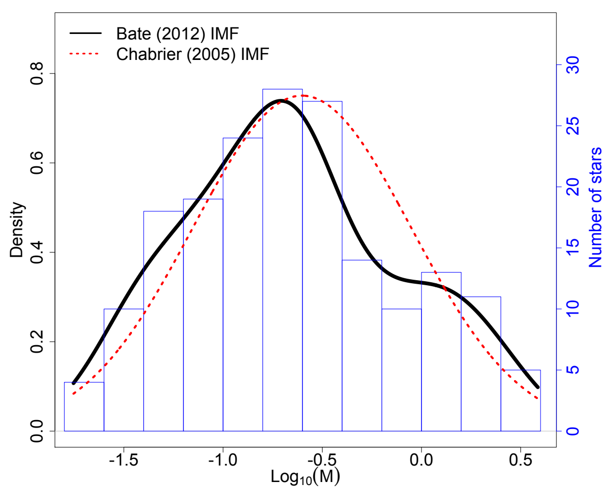
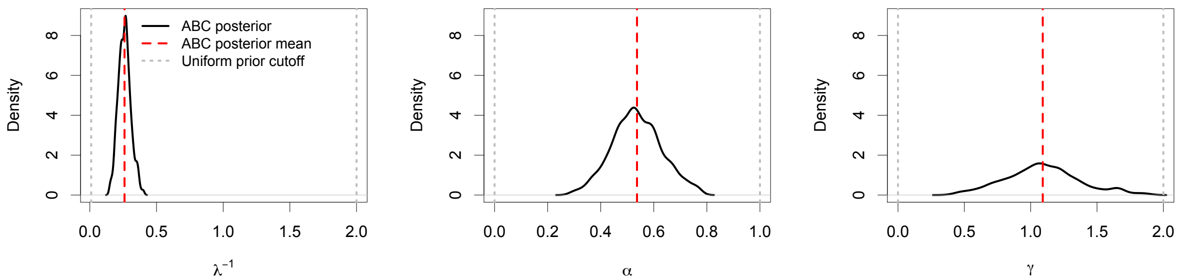


Figure 12: Astrophysical simulation IMF. The black curve displays the IMF of the 183 stars and brown dwarfs simulated from Bate (2012) and the red dotted curve is the IMF of Chabrier (2005). The right axis provides the number of stars (and brown dwarfs) for the histogram (plotted in blue).

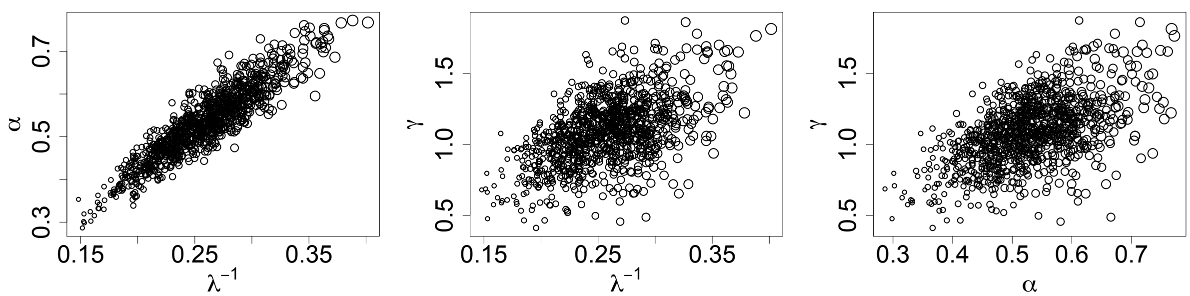
6 Discussion

Accounting for complex dependencies in observations, such as the initial masses of stars forming from a molecular cloud, is a challenging statistical problem. A possible, but unsatisfactory, resolution is simply to proceed as though the dependency is weak enough to assume independence. Instead, we draw from the idea of preferential attachment proposing



(a) ABC posterior for λ^{-1} *(b) ABC posterior for α* *(c) ABC posterior for γ*

Figure 13: Marginal ABC posteriors for astrophysical simulation data from Bate (2012). The vertical dashed red lines indicate the ABC posterior mean, and the dotted gray lines indicate the range of the uniform prior for the parameter.



(a) ABC Joint (λ^{-1}, α) *(b) ABC Joint (λ^{-1}, γ)* *(c) ABC Joint (α, γ)*

Figure 14: Pairwise joint ABC posteriors for astrophysical simulation data from Bate (2012). Pairwise ABC posterior particles samples of (a) (λ^{-1}, α), (b) (λ^{-1}, γ), and (c) (α, γ) for the astrophysical simulation data from Bate (2012). The size of the plot symbol is scaled with the particle weight.

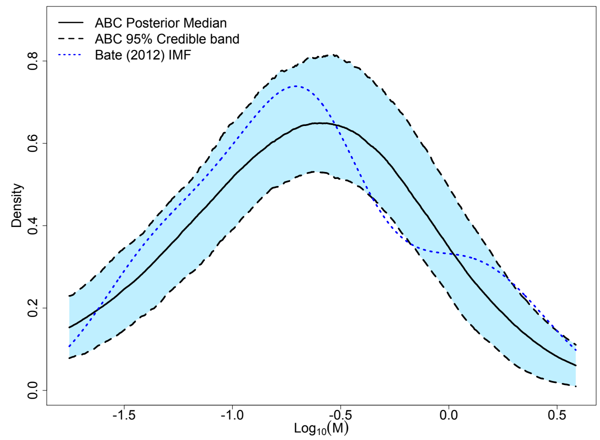


Figure 15: Posterior predictive IMF for astrophysical simulation data from Bate (2012). The median of the posterior predictive IMF (solid black) with a corresponding 95% pointwise predictive band (dashed black) compared to the true IMF (blue dotted). For the posterior predictive IMF, 1000 draws were made from the ABC posteriors of Figure 13 and then 1000 cluster samples were drawn from the proposed forward model.

a new forward model to be used in an ABC algorithm. Though the new generative model was motivated by inference on stellar IMFs, it is generalizable to other applications that operate on a similar principle of dependent data.

The new generative model starts with the total mass of the system and stochastically builds individuals stars of particular mass at a sub-linear, linear, or super-linear rate. A goal of the proposed model and algorithm is to begin making a statistical connection between the observed stellar MF and the formation mechanism of the cluster, not that the proposed model shape is superior to the standard IMF models. Rather, the proposed model is more general in the sense that it that it captures the dependencies among the masses of the stars by connecting the star masses to a possible cluster formation mechanism, and also can accommodate standard models proposed in the astronomical literature. Additionally, by coupling the proposed model with ABC, observational limitations such as the aging and completeness of the observed cluster can easily be accounted for.

While the proposed model is able to account for a particular dependency among the masses during cluster formation, there are several extensions that would be scientifically and statistically interesting. First, the generative model could be extended to capture the spatial dependency among the observations. Intuitively, such an approach could account for a mechanism that limits the formation of multiple very massive stars relatively near to each other. Understanding the spatial distribution of masses of stars during formation would help advance our understanding of stellar formation and evolution. Other effects that could be incorporated into the generative modeling include accounting for binary and other multiple star systems, the possible disturbances to the observed MF as stars die (beyond the censoring of the most massive stars), or spatial completeness functions (i.e. a completeness function that depends on more than the mass of the object, but also its location in the cluster).

Overall, the proposed generative model and ABC methodology provide a useful framework for dealing with complex physical processes that are otherwise difficult to work with in a statistically rigorous fashion. As increasing computational resources allow for greater model complexity in studies of astronomy and other scientific fields, the proposed and other ABC algorithms may open new opportunities for Bayesian inference in challenging problems.

A ABC summary statistic selection

In order to select effective summary statistics for the proposed model, we first employ the ABC methodology in a simplified study that focuses on the posterior of the power law parameter α from Equation (1). We generate a cluster of $n = 10^3$ stars from an IMF with slope $\alpha = 2.35$ (Salpeter 1955), $M_{\text{min}} = 2$, and $M_{\text{max}} = 60$, and a uniform prior distribution for $\alpha \in (0, 6)$. This model was used in order to check the method against the true posterior of α after the observational and aging effects have been incorporated into the forward model. We use the bivariate summary statistic and distance function of Equation (8). Defining the two-dimensional tolerance sequence as $(\epsilon_{1t}, \epsilon_{2t})$ where the subscript t indicates the algorithm time step, and ϵ_{11} and ϵ_{21} were selected using an adaptive start as discussed in Section 4 using an initial number of draws of $10N$ with $N = 10^3$. The algorithm ran for $T = 5$ time steps. At steps $t = 2, \dots, T$, ϵ_{1t} and ϵ_{2t} were set equal to the $25th$ percentile of the distances retained at the previous step from their corresponding distance functions.

The pseudo-data were aged 30 Myr, lognormal measurement error with $\sigma = 0.25$, and observation completeness defined by the linear-ramp function in (5). The simulated IMF and resulting MF (after the noted observational effects were applied) are displayed in Figure 16. The IMF is the object of interest, while the MF contain the actual observations that can be used for analysis.

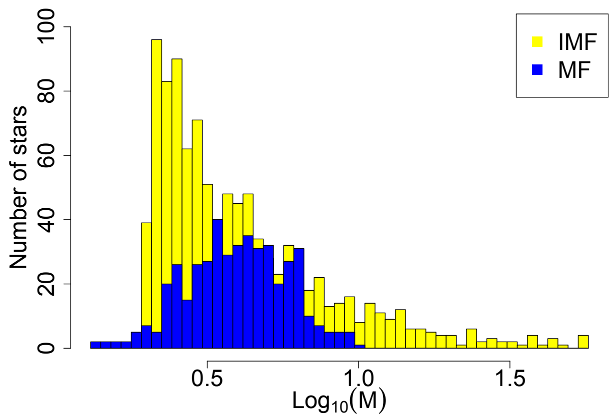
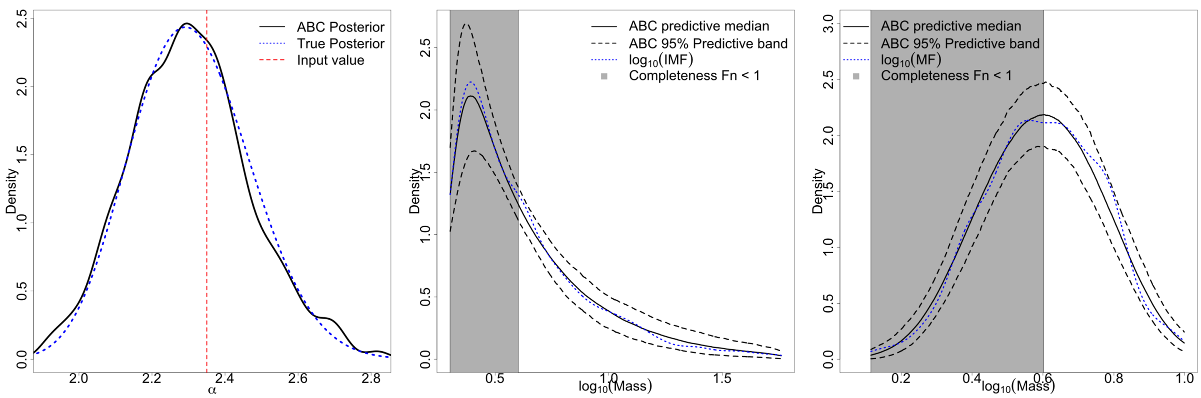


Figure 16: Simulated IMF (yellow) and MF (blue) using power law model. The IMF was simulated with $n = 10^3$ stars using a power law slope of $\alpha = 2.35$. The cluster was aged 30 Myrs, simulated with log-normal measurement error with $\sigma = 0.25$, and had a linear-ramp completeness function applied between 2 and 4 $M_\odot$.

The ABC posterior resulting from the ABC algorithm along with the true posterior for α are displayed in Figure 17a. The ABC posterior matches the true posterior, defined as

$$\begin{aligned} &\pi_F(\alpha \mid m_{\text{obs}}, M_{\text{min}}, M_{\text{max}}, n_{\text{obs}}, n_{tot}, T_{age}) \propto \qquad (11) \\ &\left\{ \Pr(\mathrm{M} > \mathrm{T_{age}}) + \left(\frac{1-\alpha}{\mathrm{M}_{\max}^{1-\alpha} - \mathrm{M}_{\min}^{1-\alpha}} \right) \int_2^4 \mathrm{M}^{-\alpha} \left(1 - \frac{\mathrm{M}-2}{2} \right) \mathrm{dM} \right\}^{n_{tot} - n_{\text{obs}}} \\ &\times \prod_{i=1}^{n_{\text{obs}}} \left\{ \int_2^{T_{age}} (2\pi\sigma^2)^{-\frac{1}{2}} m_i^{-1} e^{-\frac{1}{2\sigma^2}(\log(m_i) - \log(M))^2} \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}} \right) M^{-\alpha} \right. \\ &\left. \times \left(I\{M > 4\} + \left(\frac{M-2}{2} \right) I\{2 \leq M \leq 4\} \right) dM \right\} \end{aligned}$$

where $T_{age} = \tau^{-1/3} \times 10^{4/3}$ is the upper-tail mass cutoff due to aging. The close match between the true and ABC posteriors suggests that the selected summary statistics are useful for carrying out the ABC analysis. Figures 17b and 17c display the ABC posterior predictive IMF and MF. Even in regions where stars are missing due to the observational limitations, the ABC predictive median is still able to recover the shape of the original IMF (though with wider credible bands).



(a) Posterior for α *(b) Posterior predictive IMF* *(c) Posterior predictive MF*

Figure 17: Validation of summary statistics with power law model. (a) The ABC posterior for α (solid black) compared to the true posterior (dotted blue) of Equation (11) *using an input value of 2.35 (dashed vertical red). (b) The median of the posterior predictive IMF (solid black) with a corresponding 95% point-wise predictive band (dashed black) compared to the true IMF (blue dotted) which was the simulated dataset before aging, completeness, or uncertainty were applied, and the gray shaded region indicates where the completeness function was less than 1. (c) The median of the posterior predictive MF (solid black) with a corresponding 95% point-wise predictive band (dashed black) compared to the observed MF (dotted blue) which was the simulated dataset after aging, completeness, and uncertainty were applied. For the posterior predictive IMF, 1000 draws were made from the ABC posterior of (a) and then 1000 cluster samples were drawn from the power law simulation model. For the posterior predictive MF, the 1000 cluster samples used for (b) were then put through the forward model to apply the aging, completeness, and measurement error effects.*

Acknowledgements

We thank the Yale Center for Research Computing for the resources we used for producing this paper. The authors also benefited from feedback from Ewan Cameron. Jessi Cisewski-Kehe and Grant Weller were partially supported by the National Science Foundation under Grant DMS-1043903. Chad Schafer was supported by NSF Grant DMS-1106956. David W. Hogg was partially supported by the NSF (AST-0908357) and the Moore–Sloan Data Science Environment at NYU. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Akeret, J., Refregier, A., Amara, A., Seehars, S. & Hasner, C. (2015), 'Approximate Bayesian computation for forward modeling in cosmology', *arXiv preprint arXiv:1504.07245* .

Ashworth, G., Fumagalli, M., Krumholz, M. R., Adamo, A., Calzetti, D., Chandar, R., Cignoni, M., Dale, D., Elmegreen, B. G., Gallagher, III, J. S., Gouliermis, D. A., Grasha, K., Grebel, E. K., Johnson, K. E., Lee, J., Tosi, M. & Wofford, A. (2017), 'Exploring the IMF of star clusters: a joint SLUG and LEGUS effort', *Monthly Notices of the Royal Astronomical Society* **469**, 2464–2480.

Bally, J. & Reipurth, B. (2005), *The Birth of Stars and Planets*, Cambridge University Press, New York.

Bastian, N., Covey, K. R. & Meyer, M. R. (2010), 'A universal stellar initial mass function? a critical look at variations', *Annu. Rev. Astron. Astr.* **48**, 339–389.

Bate, M. R. (2012), 'Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation', *Monthly Notices of the Royal Astronomical Society* **419**(4), 3115–3146.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. (2009), 'Adaptive approximate Bayesian computation', *Biometrika* **96**(4), 983–990.

Beccari, G., Petr-Gotzens, M. G., Boffin, H. M. J., Romaniello, M., Fedele, D., Carraro, G., De Marchi, G., de Wit, W.-J., Drew, J. E., Kalari, V. M., Manara, C. F., Martin, E. L., Mieske, S., Panagia, N., Testi, L., Vink, J. S., Walsh, J. R. & Wright, N. J. (2017), 'A tale of three cities. OmegaCAM discovers multiple sequences in the color-magnitude diagram of the Orion Nebula Cluster', *Astronomy & Astrophysics* **604**, A22.

Birrer, S., Amara, A. & Refregier, A. (2017), 'Lensing substructure quantification in RXJ1131-1231: a 2 keV lower bound on dark matter thermal relic mass', *Journal of Cosmology and Astroparticle Physics* **5**, 037.

Blum, M. G. (2010), 'Approximate Bayesian computation: a nonparametric perspective', *J. Am. Statist. Assoc.* **105**(491), 1178–1187.

Blum, M. G. & François, O. (2010), 'Non-linear regression models for approximate Bayesian computation', *Stat. Comput.* **20**, 63–73.

Blum, M., Nunes, M., Prangle, D. & Sisson, S. (2013), 'A comparative review of dimension reduction methods in approximate Bayesian computation', *Stat. Sci.* **28**(2), 189–208.

Bonnell, I. A., Clarke, C. J. & Bate, M. R. (2006), 'The Jeans mass and the origin of the knee in the IMF', *Mon. Not. R. Astron. Soc.* **368**(3), 1296–1300.
URL: *http://dx.doi.org/10.1111/j.1365-2966.2006.10214.x*

Cameron, E. & Pettitt, A. N. (2012), 'Approximate Bayesian computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift', *Mon. Not. R. Astron. Soc.* **425**, 44–65.

Chabrier, G. (2003*a*), 'Galactic stellar and substellar initial mass function', *Publications of the Astronomical Society of the Pacific* **115**(809), 763.

Chabrier, G. (2003*b*), 'The galactic disk mass function: reconciliation of the Hubble space telescope and nearby determinations', *The Astrophysical Journal Letters* **586**(2), L133.

Chabrier, G. (2005), The initial mass function: from salpeter 1955 to 2005, *in* 'The Initial Mass Function 50 years later', Springer, The Netherlands, pp. 41–50.

Chaisson, E. & McMillan, S. (2011), *Astronomy today*, 7th edn, Addison-Wesley.

Corebelli, E., Palla, F. & Zinnecker, H., eds (2005), *The initial mass function 50 years later*, Vol. 327 of *Astrophysics and Space Science Library*, Springer.

Da Rio, N., Robberto, M., Hillenbrand, L. A., Henning, T. & Stassun, K. G. (2012), 'The initial mass function of the Orion Nebula Cluster across the H-burning limit', *Astrophys. J.* **748**(1), 1–15.

Da Rio, N., Robberto, M., Soderblom, D. R., Panagia, N., Hillenbrand, L. A., Palla, F. & Stassun, K. G. (2010), 'A multi-color optical survey of the Orion Nebula cluster. II. The HR diagram', *The Astrophysical Journal* **722**(2), 1092.

Del Moral, P., Doucet, A. & Jasra, A. (2011), 'An adaptive sequential Monte Carlo method for approximate Bayesian computation', *Stat. Comput.* **22**(5), 1009–1020.

Dib, S., Schmeja, S. & Hony, S. (2017), 'Massive stars reveal variations of the stellar initial mass function in the Milky Way stellar clusters', *Monthly Notices of the Royal Astronomical Society* **464**, 1738–1752.

Dib, S., Shadmehri, M., Padoan, P., Maheswar, G., Ojha, D. K. & Khajenabi, F. (2010), 'The IMF of stellar clusters: effects of accretion and feedback', *Mon. Not. R. Astron. Soc.* **405**(1), 401–420.
URL: *http://mnras.oxfordjournals.org/content/405/1/401.abstract*

Fearnhead, P. & Prangle, D. (2012), 'Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation', *J. Roy. Stat. Soc.* **B 74**(3), 419–474.

Geha, M., Brown, T. M., Tumlinson, J., Kalirai, J. S., Simon, J. D., Kirby, E. N., VandenBerg, D. A., Muñoz, R. R., Avila, R. J., Guhathakurta, P. & Ferguson, H. C. (2013), 'The

Stellar Initial Mass Function of Ultra-faint Dwarf Galaxies: Evidence for IMF Variations with Galactic Environment', *The Astrophysical Journal* **771**, 29.

Hahn, C., Tinker, J. L. & Wetzel, A. (2017), 'Star Formation Quenching Timescale of Central Galaxies in a Hierarchical Universe', *Astrophysical Journal* **841**, 6.

Hahn, C., Vakili, M., Walsh, K., Hearin, A. P., Hogg, D. W. & Campbell, D. (2017), 'Approximate Bayesian computation in large-scale structure: constraining the galaxy-halo connection', *Mon. Not. R. Astron. Soc.* **469**, 2791–2805.

Hansen, C. J., Kawaler, S. D. & Trimble, V. (2004), *Stellar Interiors: Physical Principles, Structure, and Evolution*, Springer, New York.

Herbel, J., Kacprzak, T., Amara, A., Refregier, A., Bruderer, C. & Nicola, A. (2017), 'The redshift distribution of cosmological samples: a forward modeling approach', *Journal of Cosmology and Astroparticle Physics* **8**, 035.

Ishida, E., Vitenti, S., Penna-Lima, M., Cisewski, J., de Souza, R., Trindade, A., Cameron, E. et al. (2015), 'cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation', *arXiv preprint arXiv:1504.06129* .

Jose, J., Herczeg, G. J., Samal, M. R., Fang, Q. & Panwar, N. (2017), 'The Low-mass Population in the Young Cluster Stock 8: Stellar Properties and Initial Mass Function', *The Astrophysical Journal* **836**, 98.

Joyce, P. & Marjoram, P. (2008), 'Approximately sufficient statistics and Bayesian computation', *Stat. Appl. Genet. Mol.* **7**(1), 1–16.

Kalari, V. M., Carraro, G., Evans, C. J. & Rubio, M. (2018), 'The Magellanic Bridge cluster NGC 796: Deep optical AO imaging reveals the stellar content and initial mass function of a massive open cluster', *ArXiv e-prints* .

Killedar, M., Borgani, S., Fabjan, D., Dolag, K., Granato, G., Meneghetti, M., Planelles, S. & Ragone-Figueroa, C. (2018), 'Simulation-based marginal likelihood for cluster strong lensing cosmology', *Mon. Not. R. Astron. Soc.* **473**, 1736–1750.

Kroupa, P. (2001), 'On the variation of the initial mass function', *Mon. Not. R. Astron. Soc.* **322**(2), 231–246.

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., Thies, I., Dabringhausen, J., Marks, M. & Maschberger, T. (2012), The stellar and sub-stellar IMF of simple and composite populations, *in* I. Oswalt, T. D. with McLean, H. Bond, L. French, P. Kalas, M. Barstow, G. Gilmore & W. Keel, eds, 'Stellar Systems and Galactic Structure', Vol. 5 of *Planets, Stars, and Stellar Systems*, Springer, Dordrecht, pp. 115–142.

Lim, B., Chun, M.-Y., Sung, H., Park, B.-G., Lee, J.-J., Sohn, S. T., Hur, H. & Bessell, M. S. (2013), 'The starburst cluster Westerlund 1: the initial mass function and mass segregation', *Astron. J.* **145**, 46–64.

Lin, C.-A. & Kilbinger, M. (2015), 'A new model to predict weak-lensing peak counts. II. Parameter constraint strategies', *Astronomy & Astrophysics* **583**, A70.

Marin, J.-M., Pudlo, P., Robert, C. P. & Ryder, R. J. (2012), 'Approximate Bayesian computational methods', *Statistics and Computing* **22**(6), 1167–1180.

Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. (2003), 'Markov chain Monte Carlo without likelihoods', *P. Natl. Acad. Sci. USA* **100**(26), 15324–15328.

Massey, P. (2003), 'Massive stars in the local group: implications for stellar evolution and star formation', *Annu. Rev. Astron. Astr.* **41**, 15–56.

McMillan, P. J. (2016), 'The mass distribution and gravitational potential of the Milky Way', *Monthly Notices of the Royal Astronomical Society* p. stw2759.

Mitzenmacher, M. (2004), 'A brief history of generative models for power law and lognormal distributions', *Internet mathematics* **1**(2), 226–251.

Newman, M. E. (2005), 'Power laws, pareto distributions and zipf's law', *Contem. Phys.* **46**(5), 323–351.

Offner, S. S., Clark, P. C., Hennebelle, P., Bastian, N., Bate, M. R., Hopkins, P. F., Moraux, E. & Whitworth, A. P. (2014), 'The origin and universality of the stellar initial mass function', *Protostars and Planets VI* **1**, 53–75.

Parker, A. H. (2015), 'The intrinsic Neptune Trojan orbit distribution: Implications for the primordial disk and planet migration', *Icarus* **247**, 112–125.

Peel, A., Lin, C.-A., Lanusse, F., Leonard, A., Starck, J.-L. & Kilbinger, M. (2017), 'Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey', *Astronomy & Astrophysics* **599**, A79.

Pokhrel, R., Myers, P. C., Dunham, M. M., Stephens, I. W., Sadavoy, S. I., Zhang, Q., Bourke, T. L., Tobin, J. J., Lee, K. I., Gutermuth, R. A. et al. (2018), 'Hierarchical fragmentation in the perseus molecular cloud: From the cloud scale to protostellar objects', *The Astrophysical Journal* **853**(1), 5.

Pritchard, J. K., Seielstad, M. T. & Perez-Lezaun, A. (1999), 'Population growth of human Y chromosomes: A study of Y chromosome microsatellites', *Mol. Bio. Evol.* **16**(12), 1791–1798.

Robin, A., Reylé, C., Fliri, J., Czekaj, M., Robert, C. & Martins, A. (2014), 'Constraining the thick disc formation scenario of the Milky Way', *Astronomy & Astrophysics* **569**, A13.

Salpeter, E. E. (1955), 'The luminosity function and stellar evolution', *Astrophys. J.* **121**, 161–167.

Schafer, C. M. & Freeman, P. E. (2012), Likelihood-free inference in cosmology: potential for the estimation of luminosity functions, *in* 'Statistical Challenges in Modern Astronomy V', Springer, New York, pp. 3–19.

Shetty, S. & Cappellari, M. (2014), 'Salpeter normalization of the stellar initial mass function for massive galaxies at z $\sim$ 1', *Astrophys. J. Let.* **786**, L10.

Simon, H. A. (1955), 'On a class of skew distribution functions', *Biometrika* **42**(314), 425–440.

Sisson, S. A., Fan, Y. & Tanaka, M. M. (2007), 'Sequential Monte Carlo without likelihoods', *P. Natl. Acad. Sci. USA* **104**(6), 1760–1765.

Spiniello, C., Trager, S., Koopmans, L. V. E. & Conroy, C. (2014), 'The stellar IMF in early-type galaxies from a non-degenerate set of optical line indices', *Mon. Not. R. Astron. Soc.* **438**, 1483–1499.

Tavaré, S., Balding, D. J., Griffiths, R. & Donnelly, P. (1997), 'Inferring coalescence times from DNA sequence data', *Genetics* **145**, 505–518.

Treu, T., Auger, M. W., Koopmans, L. V. E., Gavazzi, R., Marshall, P. J. & Bolton, A. S. (2010), 'The initial mass function of early-type galaxies', *Astrophys. J.* **709**, 1195–1202.

van Dokkum, P. G. & Conroy, C. (2010), 'A substantial population of low-mass stars in luminous elliptical galaxies', *Nature* **468**(7326), 940.

Weisz, D. R., Fouesneau, M., Hogg, D. W., Rix, H.-W., Dolphin, A. E., Dalcanton, J. J., Foreman-Mackey, D. T., Lang, D., Johnson, L. C., Beerman, L. C. et al. (2013), 'The Panchromatic Hubble Andromeda Treasury IV. A probabilistic approach to inferring the high-mass stellar initial mass function and other power-law functions', *Astrophys. J.* **762**(2), 123–143.

Weisz, D. R., Johnson, L. C., Foreman-Mackey, D., Dolphin, A. E., Beerman, L. C., Williams, B. F., Dalcanton, J. J., Rix, H.-W., Hogg, D. W., Fouesneau, M. et al. (2015), 'The High-Mass Stellar Initial Mass Function in M31 Clusters', *The Astrophysical Journal* **806**(2), 198.

Weyant, A., Schafer, C. & Wood-Vasey, W. M. (2013), 'Likelihood-free cosmological inference with type Ia supernovae: approximate Bayesian computation for a complete treatment of uncertainty', *Astrophys. J.* **764**, 116–130.

Woosley, S. E. & Heger, A. (2015), The Deaths of Very Massive Stars, *in* J. S. Vink, ed., 'Very Massive Stars in the Local Universe', Vol. 412 of *Astrophysics and Space Science Library*, p. 199.

Data: Observed stellar masses, (m_{obs})

Result: ABC posterior sample of θ

At iteration $t = 1$*:*

for $j = 1, \ldots, kN$ **do**

  Propose $\theta_t^{(j)}$ by drawing $\theta_t^* \sim \pi(\theta)$

  Generate cluster stellar masses m_{sim} from $F(m \mid \theta_t^*)$

  Calculate distance $\rho_t^{(j)} \leftarrow \rho(m_{\text{sim}}, m_{\text{obs}})$

end

$\theta_t^{(j)} \leftarrow \theta_t^{(l)}, l =$ index of N smallest $\rho_t^{(q)}, q = 1, \ldots, kN$

$W_t^{(j)} \leftarrow 1/N, j = 1, \ldots, N$

At iterations $t = 2, \ldots, T$*:*

for $j = 1, \ldots, N$ **do**

  while $\rho(m_{sim}, m_{obs}) > \epsilon_t$ **do**

    Select $\theta^{(j)}$ by drawing from the $\theta_{t-1}^{(i)}$ with probabilities $W_{t-1}^{(i)}, i = 1, \ldots, N$

    Generate $\theta^{*(j)}$ from transition kernel $K(\theta^{(j)}, \cdot)$

    Generate cluster stellar masses m_{sim} from $F(m \mid \theta^{*(j)})$

  end

  $\theta_t^{(j)} \leftarrow \theta^{*(j)}$

  $W_t^{(j)} \leftarrow \frac{\pi(\theta_t^{(j)})}{\sum_{i=1}^N W_{t-1}^{(i)} K(\theta_{t-1}^{(i)}, \theta_t^{(j)})}$

end

$W_t^{(j)} \leftarrow \frac{W_t^{(j)}}{\sum_{l=1}^N W_t^{(l)}}, j = 1, \ldots, N$

Algorithm 1: Stellar IMF ABC algorithm with sequential sampling

Number	λ^{-1}	α	γ	M_{tot}	n_{obs}
1	0.25	0.3	0.5	1000	1166
2	0.25	0.3	1.0	1000	1142
3	0.25	0.3	1.5	1000	1148
4	0.25	0.7	0.5	1000	2801
5	0.25	0.7	1.0	1000	2698
6	0.25	0.7	1.5	1000	2710

Table 1: Parameter values for the simulation study and the number of stars in the data set, n_{obs}.

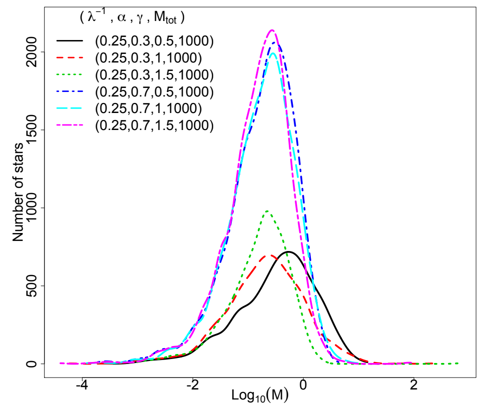


Figure 5: The simulated IMF's used as the observations in the simulation study displayed as number of stars by Log_{10} mass.

$C_{\max}$ ($M_\odot$)	Number of stars in MF (% of IMF)
0.10	
0.25	
0.50	
0.75	
1.0	

Table 2: Number and percentage of stars in the MF compared to the IMF with 1142 stars for the simulation setting of Section 4.2.1 as the upper bound, C_{max}, of the linear ramp completeness function of Equation (5) changes.