Review of "A Preferential Attachment Model for the Stellar Initial Mass Function"
July 18, 2018

1 Summary

The authors develop a preferential attachment model for estimating the stellar IMF. They use Approximate Bayesian Computation to derive posteriors for the model. Use of this methodology (PA + ABC) within astronomy has the potential to improve our understanding of the stellar IMF. Further, the application of this methodology to challenging data sets, such as those studied here, will improve statisticians understanding of the methods and help generate further methodological advances.

I have some questions/suggestions mostly regarding the Preferential Attachment model and ABC.

2 Preferential Attachment Model

- Page 8: More references to statistics papers studying / developing PA models would be helpful. Is PA related to models such as the Chinese Restaurant Process? What sort of computational methods ABC or otherwise have been used for other PA models? Could those methods be used here, as competitors to ABC?

- Page 8: In Equation 3.1 $\pi_t = \min(1, \alpha)$ but two sentences later the range of α is restricted to be $[0, 1)$.

- Page 8: The π_{kt} are the probability that some mass joins existing star k at time t. But the 3.1 definition of π_{kt}, looks like it could be greater than 1. In particular it has units of some mass. Would
$$\frac{\pi_{kt}}{\pi_t + \sum_j \pi_{jt}}$$
make more sense for the probability of joining k?

- Page 8 "The generating process is complete when the total mass of formed stars reaches M_{tot}" The total mass of stars will never equal exactly M_{tot} because the mass additions are from a continuous $exp(\lambda)$ distribution, correct? Do we stop when the total mass first exceeds M_{tot}?

- Figure 2 and 3: Is it reassuring that the PA model can reproduce Kroupa (2001) and Chabrier (2003a,b), given that neither of these models is necessarily correct?

- Page 15: The phenomenon of "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" sounds physically unrealistic. Wouldn't there be some upper limit, such as the star collapsing on itself and becoming a black hole? The PA model proposed seems related to clustering ideas such as Dirichlet process mixture models / Chinese restaurant processes. It has been noted in these models that "it is well known the DPMs favor introducing new components at a log rate as the sample size increases, and tend to produce some large clusters along with many small clusters." (see "Reducing over-clustering via the powered Chinese restaurant process" by Lu, Li, Dunson on arXiv) Is this the same / similar phenomenon happening here?

- Page 25: "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." The proposed model is superior in that we may not know that Chabrier or Kroupa IMF shape is correct but the PA model is flexible and includes both these models as (approximate) submodels. So rather than fitting both Chabrier and Kroupa and doing model selection, we just fit PA and interpret the posteriors. Right?

3 Analysis / Comparison of ABC algorithm

- Section 4: What is the alternative to using ABC in Section 4? Why are these alternatives (such as MCMC or variational Bayes) impractical for this problem? It seems to me that evaluating the likelihood with the measurement errors 3.5 would require convolving the likelihood with normals of different variance and then renormalizing separately for each observation (star). The would require n 1-d integrals to evaluate the likelihood once. Is this true? Is this why one cannot use MCMC?
- Figure 7 a) Degeneracies in posteriors often cause problems for convergence of MCMC algorithms such as Gibbs or Metropolis. Would there be any reason to worry about convergence for sequential ABC? Is it possible to reparameterize to make the posteriors less dependent across these parameters?
- Section 4: How close is the ABC approximation to the actual posterior? Are there any ABC convergence diagnostics available?

4 Other Issues

- Section 3.1.1. I am trying to understand the relevance of this section to the rest of the work. "the power law model is a prevalent assumption in this application" If this assumption is (approximately) correct does the conclusion "the power law fit degrades quickly for γ outside (0.5, 1.5)" imply that priors on γ could/should put most mass in this range? More generally, connecting this section more strongly with the rest of the work would be helpful.
- Page 4: "Focusing on the upper part." the "upper part" is large m? maybe "upper tail". does M_{min} define the left boundary of the upper part? if so, then why would "c" be chosen to make f_M a valid pdf? wouldn't f_M integrate to less than 1?
- Figure 3: I would suggest limiting the x-axes to the support of the posteriors, rather than the support of the priors. The current scaling may be masking differences in the

distributions, especially for the a), the λ^{-1} parameter. I don't think the y-axis need to be the same for all densities. These changes will make comparison across plots more difficult, but I think they are worthwhile because the more important comparison is between the Kroupa and Chabrier model for a particular parameter. Similar comment for Figure 6, especially 6 a).

- Is the completeness function equivalent to some form of probabilistic truncation? (I am using the survival analysis definition of truncation) If so, is there research on this within the survival analysis literature? If so, could the authors provide some citations?
- Is the completeness function 3.4 applied to the data before the measurement error 3.5. If so, why? (I don't see a clear reason for either ordering, perhaps the instrument somehow determines this.)
- In Equation 3.6, is $f_M(m|\theta)$ the stellar initial mass function after accounting for measurement error (3.5) and completeness (3.4)? Shouldn't there be a proportionality, rather than equality due to need for normalization? Is the lhs of 3.6 the mass function (MF) referred to in the subsequent section.
- Page 18: "4.2.1. Simulated data with observational effects" So the simulation in 4.2, before 4.2.1, does not have these effects? This is not explicitly stated and so is rather confusing. Perhaps have 4.2.1 be the simulation without effects and 4.2.2 (currently 4.2.1) the simulation with effects. At the beginning of 4.2 you could let readers know there will be two simulations.
- Effective application of ABC requires a good distance function ρ, tolerances ϵ, kernel, etc. Were there findings, perhaps qualitative, about how to choice these quantities that could be useful for other ABC practitioners? Perhaps the authors could summarize these findings in the conclusions, e.g. a few sentences of the form "In agreement with other studies, selection of a proper distance function was the most challenging . . " or "In contrast to so and so . . ."
- Link / references to code and / or data to reproduce results?