MNRAS **000**, 1–9 (2018) Preprint 2 May 2019 Compiled using MNRAS LATEX style file v3.0

StarSampler: A Code for Sampling Phase Space Coordinates of Stellar Systems

Mao-Sheng Liu,[1]★ Matthew G. Walker[1], and Jorge Peñarrubia[2]
[1]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
[2]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We introduce `StarSampler`, a Python module that generates random samples from any user-defined, stationary distribution function (DF) that specifies the probability density of stellar coordinates within six-dimensional phase space. The sampled DF can describe stellar systems that are self-gravitating and/or embedded within a potential generated by other sources (e.g., a dark matter halo). The user can choose from two types of sampling technique: an exact, but often inefficient, rejection sampling method or an approximate, but faster, importance sampling technique. We compare the performance of both sampling methods using three different DFs that are included as examples: those of King (1966), Osipkov (1979) and Merritt (1985), and Strigari, Frenk & White (2017). Generally, the additional sampling noise incurred from the importance sampling is greatly diminished with a larger sample size (>10,000). For smaller sample size the approximated sampling can be improved with proposal functions that better resembles the target DF, and with a larger number of proposal sample points.

Key words: methods: statistical – galaxies: dwarf – galaxies: kinematics and dynamics

1 INTRODUCTION

The dynamics of collisionless stellar systems are specified by the phase-space distribution function (DF), $f(\vec{x}, \vec{v}, t)$ (Binney & Tremaine 2008). For a system in steady state ($df/dt = 0$), f is a function of phase-space coordinates through quantities that are conserved along a given orbit—e.g., energy $E = v^2/2 + \Phi(\vec{r})$, where $\Phi(\vec{r})$ is the gravitational potential, and angular momentum $J = \vec{r} \times \vec{v}$. Thus by expressing $f(E, J)$, one can build dynamical models for equilibrium stellar systems.

In many cases it is desirable to obtain random samples from a DF. For example, one might sample the DF in order to obtain initial conditions for numerical simulations. Alternatively, sampling from a DF can generate mock catalogs for testing and comparing the reliability of various methods for inferring $\Phi(\vec{r})$ from observations of stellar kinematics (e.g., Walker & Peñarrubia 2011; Richardson & Fairbairn 2014, *The Gaia Challenge*[1] series of workshops). Furthermore, sampling from a DF can itself become part of the procedure for inferring $\Phi(\vec{r})$, as with 'Approximate Bayesian Computation' (ABC; Beaumont et al. 2002), which compares models according to their ability to generate synthetic data sets that mimic observed ones. Similarly, as suggested by Diggle & Gratton (1984), samples from DFs can be smoothed in order to estimate model likelihoods that are otherwise difficult to compute—e.g., in the common case that the available data consist of observations only of the line-of-sight projection of phase space.

There are many methods for drawing random samples from a DF. In general, the optimal choice of sampling method depends on the DF itself. In the simple case of a univariate DF, $f(x)$, for which the inverse of the cumulative distribution function (CDF), $\mathrm{Prob}(X \leq x)$, can be calculated, then Inverse Transform Sampling (e.g., Devroye 1986) generates random samples of $f(x)$ with 100% efficiency (i.e., no rejection). For a multivariate DF, $f(x, y)$, one can use Inverse Transform Sampling with conditional probability to sample the variables one at a time; however efficient sampling requires that the multivariate distribution can be conveniently marginalized.

Here we consider the more general case in which the DF 1) can be multivariate, 2) perhaps cannot be marginalized analytically (or is expensive to marginalize numerically), and 3) has arbitrary shape. In order to sample DFs that can have these characteristics, we present `StarSampler`, a Python module that allows sampling of an arbitrary user-defined DF, $f(\vec{x}, \vec{v})$. The user can specify either rejection sampling or importance sampling. In principle, rejection sampling is applicable to DFs of arbitrary shape; in practice, however, sampling efficiency can be extremely low and generally depends on the DF itself. Alternatively, importance sampling draws approximate samples with efficiency that can be made relatively insensitive to the shape of the DF.

`StarSampler` requires the user to specify a DF, where the user can choose to include a gravitational potential in cases where the stars are not necessarily self-gravitating. As examples, the package

★ E-mail: maoshenl@andrew.cmu.edu
[1] http://astrowiki.ph.surrey.ac.uk/dokuwiki/doku.php

includes three types of pre-defined DFs that we use here to illustrate performance and highlight the relative merits/drawbacks of the two sampling methods. First, we provide a sampler for the self-consistent, isotropic DF of King (1966), which is commonly used to fit photometric and kinematic data for globular clusters. We provide a second sampler for the DF proposed by Osipkov (1979) and Merritt (1985), which specifies a velocity dispersion ellipsoid that changes from isotropic at the center to radial orbits at large radius. Finally, we provide a sampler for the relatively flexible DF recently proposed by Strigari et al. (2017, 'SFW' hereafter) to model dwarf galaxies in the context of ΛCDM cosmology.

We describe our general sampling methods in section 2. We then illustrate performance of the sampler using each of the three specific pre-defined models, comparing sampling results with analytic calculations (section 4). Finally, we compare the performance and efficiency of rejection and importance sampling techniques for a few example models (section 5). In a separate contribution, we explore how `StarSampler` can be used with stellar-kinematic data to make inferences about gravitational potentials.

2 SAMPLING METHODS

In order to draw random samples from multi-dimensional DFs, we employ rejection and importance sampling methods. In this section, we briefly summarize both techniques.

2.1 Rejection Sampling

Rejection sampling, or the acceptance-rejection method, is one of the basic techniques used to generate observations from a distribution. Suppose that we want to sample from a distribution, $p(x)$, that may or may not be normalized. Let $q(x)$ be a proposal distribution that we already know how to sample efficiently (e.g., a uniform or normal distribution), and assume that there is a constant $1 \leq C < \infty$, such that $p(x) \leq Cq(x)$ for all x. The rejection sampling steps are:

(i) Draw sample $x_i \sim q(x)$
(ii) Draw $u_i \sim$ Uniform(0,1)
(iii) If $u_i < \frac{p(x_i)}{Cq(x_i)}$, accept x_i; otherwise, reject.

The sampling efficiency (the acceptance rate) depends on how closely $q(x)$ envelops $p(x)$. However, without prior knowledge of the DF, finding a good envelope function is difficult. For simplicity and to allow sampling from arbitrary DFs with unknown shapes, we choose the proposal distribution to be uniform and equal to the global maximum of $p(x)$. With this choice, we effectively set $C = 1$. A alternative approach to rejection sampling is to use an iterative partitioning algorithm, such as the adaptive rejection sampling in Vasiliev (2018).

2.2 Importance Sampling

In cases where the proposal distribution does not closely envelop the true distribution, rejection sampling can become inefficient—especially when the DF is narrowly peaked. One alternative approach is importance sampling (or Sampling Importance Resampling), which is an approximate sampling method. Introduction and overview the method can be found in Rubin (1988); Smith & Gelfand (1992); Gelman et al. (2013).

Given a distribution to be sampled, $p(x)$, and a proposal distribution, $q(x)$, now define weights $w(x) \equiv \frac{p(x)}{q(x)}$, such that $p(x) = \frac{p(x)}{q(x)} q(x) = w(x)q(x)$. To draw N samples, the importance sampling steps are:

(i) Draw M ≥ N samples $x_1, x_2, ..., x_M \sim q(x)$
(ii) Calculate importance weight $w_i = p(x_i)/q(x_i)$ for $i = 1, 2, ..., M$
(iii) Redraw x_i^*, with (or without) replacement, from $\{x_1, x_2, ..., x_M\}$ with probability $w_i / \sum_{j=1}^{M} w_j$ for $i = 1, 2, ..., N$

The distinct advantage of importance sampling is that it is nonitterative. One retains most of the samples drawn from the proposal $q(x)$, and the draws can be made inexpensive by the choice of $q(x)$. Further more, the method doesn't require that the proposal density envelopes the target DF, and regardless whether the DF has local maxima.

The justification for this algorithm is that as $M/N \to \infty$, assuming a univariate $p(x)$ for simplicity, the cumulative distribution function (CDF) of x^* is given as:

$$\Pr(x^* \leq a) = \left(\frac{1}{M} \sum_i^M w_i 1_{(-\infty,a)}(x_i) \middle/ \frac{1}{M} \sum_i^M w_i \right) \xrightarrow[M\to\infty]{}$$
$$\frac{\mathbb{E}_q p(x) 1_{(-\infty,a)}(x)/q(x)}{\mathbb{E}_q p(x)/q(x)} = \frac{\int_{-\infty}^{a} p(x)dx}{\int_{-\infty}^{\infty} p(x)dx} = \int_{-\infty}^{a} p(x)dx$$

Therefore the distribution of x^* approximate the target DF $p(x)$ with large M. We can think of $p(x)$ being approximated by the discrete distribution supported on the x_i with probabilities proportional to the weights w_i. In practice the less the target $p(x)$ resembles the proposal $q(x)$ the larger M will be needed to make a closer approximation. For example, if $p(x) = q(x)$ or $w_i = 1$, M/N=1 is suffice. In fact, if that's the case, we can obtain exact samples by sampling without replacement (note the step 3 of the sampling algorithm). Sampling with replacement will lead to repeated samples of x_i, which has zero probability for a continuous $p(x)$. However, if $q(x)$ is a poor approximation of $p(x)$ sampling with replacement is preferred as it avoids forced choose of some unlikely values of x_i. In this work we use sampling with replacement by default, except in section 4.4 where we compare sampling with and without replacement on a toy model. We will also explore the effects of the size of the proposal points later in section 5.

2.3 Proposal Densities

For both sampling techniques, proposal distributions must be normalizable and nonzero wherever the true density is nonzero. Clearly, an appropriate choice of proposal distribution can improve the rejection sampling efficiency and/or improve the accuracy of importance sampling. In order to obtain reasonable efficiency in the general case of a multivariate DF of unknown shape, one can adopt step functions for proposal densities, as illustrated in Figure 1 and described below.

In general we anticipate the DF to be a smooth function with a global maximum that we can locate using standard numerical procedures. Therefore, the simplest proposal density is $q_1 = \max(p)$, a constant function that we set equal to the maximum of the DF (blue line in Figure 1). The uniform function $q_1(x) \geq p(x)$ is easy to sample from, but will be prone to inefficiency, particularity for the DFs that's sharply peaked. We can obtain more efficient proposal densities, $q_2(x)$, by evaluating $p(x)$ on an n-step grid within the range of the domain of x (red curve in Figure 1). However, while $q_2(x) \geq p(x)$ is true for univariate DFs, this condition is not

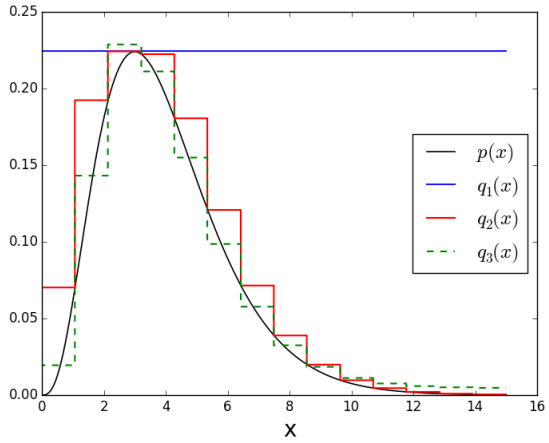

Figure 1. Three proposal distributions q_1, q_2, and q_3 for density $p(x)$. q_1 is the maximum of $p(x)$. Functions $q_2 > p(x)$ and $q_3(x)$ are step functions built from evaluation of $p(x)$ on a grid. Note that $q_3(x)$ is lifted by 2% of the function maximum to guarantee that $q_3(x) > 0$ wherever $p(x) > 0$.

guaranteed to hold for multivariate DFs. Therefore, for rejection sampling, which requires $q(x) \geq p(x)$, we adopt proposal density q_1. We use the Nelder-Mead method (Nelder & Mead 1965) to find the maximum, we find that it works well for the three types of the stellar DF describe in this paper. However, in the case where the DF has local maxima the method might get stuck in one of them, therefore the code also has the option to find the maximum by brute force.

The importance sampling requires only that the proposal $q(x) > 0$ where the target $p(x) > 0$. Previous works had attempted iterative procedures to build and improve the proposal function (e.g. Givens & Raftery 1996; Ascasibar 2008). For speed and simplicity here we adopt proposal density $q_3(x)$, which we build once by evaluating the DF at the mid-points of the grid used to build $q_2(x)$, thereby improving again the correspondence between proposal and target density. However, because the DF evaluated at the mid-point can be zero even when the DF has nonzero values between the same two grid points, we artificially lift the proposal by 5% of the maximum of the function values at the grids to guarantee that $q_3(x) > 0$ where $p(x) > 0$.

Our use of step functions allows the proposal density for importance sampling to trace the original DF and adapt according to changes in parameters that specify the DF. However, the disadvantage is that the number of function evaluations needed to build the step function scales exponentially with dimensionality of the DF, potentially dominating the budget of sampling time.

3 INPUT/OUTPUT

The `StarSampler` package allows the user to sample a distribution function that can depend on up to six phase-space coordinates (note that a spherically symmetric DF generically depends on just three such coordinates, e.g., radius, radial velocity and tangential velocity (r, v_r, v_t).

The `StarSampler` requires a user-defined python model class, under which the DF of the model is defined. The model parameters (`model_param`) specify how the density depends on the phase-space coordinates (see examples below). They are pass to the python class initialization function, and necessary one-time calculations for model setup are performed there. Note that model parameters also includes the gravitational potential parameters that allow for DFs in which the tracer particles are not self-gravitating.

In order to define the domain of the sampling region, the user must also identify the extrema of the spatial and velocity coordinates (e.g., for spherical models, a limiting radius, $r_{\rm lim}$, and the corresponding escape velocity $v_{\rm lim}$). Escape velocity imposes a limit on the range of velocity, but it is possible that the DF extend to infinity spatially. In that case a physical limit can be imposed. For example, in our implementation of Osipkov-Merritt-Zhao models, r_{200} is used as the limiting radius. It is also possible to make variable transformation to transform the unbounded range to a bounded one; however without prior knowledge of the DFs this is difficult. Finally, the user also needs to provide the number of variables that specify the spatial and velocity coordinates to help construct the appropriate dimension of the proposal points, (e.g. (r, v_r, v_t) has one spatial and two velocity coordinates).

Once the model class is defined and called, the class object can be pass on to either the rejection or the importance sampling function to draw samples. The rejection sampling method only needs the user to supply the model class and specify the number of samples, N, to be drawn. For importance sampling, the user must further specify the number of steps $N_{\rm step}$, to use in constructing the proposal density, and the resample factor.

`StarSampler` then returns an array of N samples from the specified phase space coordinates. In the standard case of a spherically symmetric DF, `StarSampler` optionally lets the user specify that the sampled points be transformed to 6D Cartesian phase space (x, y, z, v_x, v_y, v_z), based on the assumption of uniform angular distributions.

4 EXAMPLE STELLAR DENSITY FUNCTIONS

In order to test our methodology and provide concrete examples to the user, we have implemented the sampling of three DFs from the literature that represent a range of complexity. Here we illustrate the performance of `StarSampler` in generating random samples from each of these DFs. Previous methods that sample some of the specific models can be find in (e.g., Zemp et al. 2008; Pelupessy et al. 2013; Vasiliev 2018; McMillan & Dehnen 2007; Gieles & Zocchi 2015). Our purpose here is not to reinvent these wheels, but merely to demonstrate the easy applicability of StarSampler to different cases.

4.1 King model

The model of King (1966) imposes an exponential cutoff to the standard isothermal sphere. As such, the model is well suited to describe systems that have isotropic velocity dispersions and truncated configurations. The King DF is (e.g., Binney & Tremaine 2008)

$$f(\mathcal{E}) = \begin{cases} \rho_0 (2\pi\sigma^2)^{(-3/2)} e^{(\mathcal{E}/\sigma^2 - 1)} & \text{if } \mathcal{E} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where σ is a scale velocity dispersion and $\mathcal{E}(r, v) \equiv \Psi(r) - v^2/2$ is the relative energy defined by relative potential $\Psi \equiv -\Phi + \Phi_0$, where Φ_0 is constant.

The King model is fully specified by free parameters σ and $\Psi(0)$, where $\Psi(0)$ is the central relative potential. By integrating the DF over velocity space, one obtains the mass density of stars as a

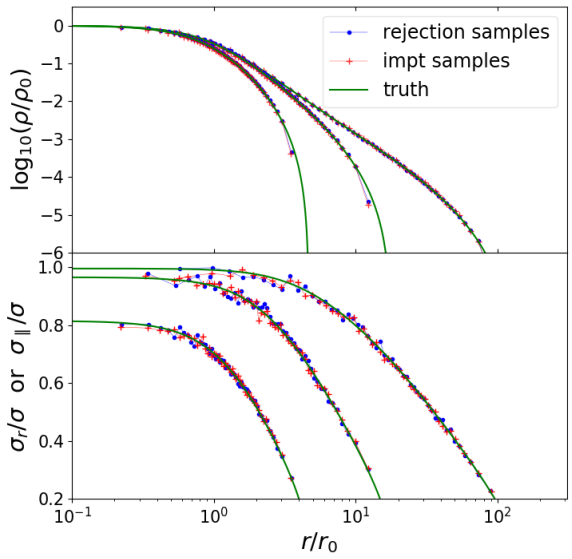

Figure 2. Mass-density (top) and radial velocity dispersion (bottom) profiles for King models with $\Psi(0)/\sigma^2 = 9, 6, 3$ (in order of most to least extended). Solid curves are calculated directly from the DF, while data points are estimates obtained by binning samples of size $N = 10^5$, obtained using rejection (blue) and importance (red) sampling methods.

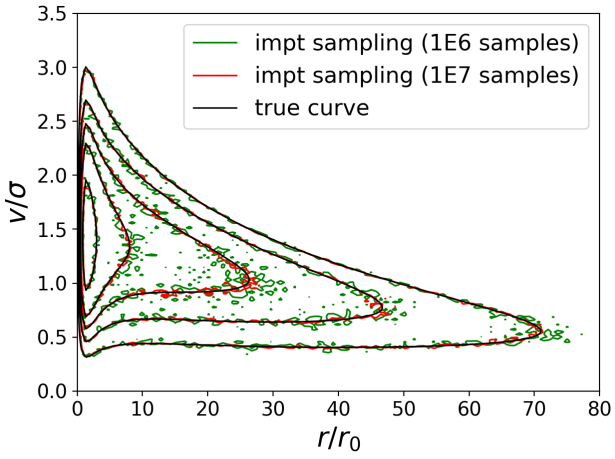

Figure 3. Contours of constant phase-space density for a King DF with $\Psi(0)/\sigma^2 = 9$. Solid curves are calculated directly from the DF, while green and red curves are estimates obtained by binning random samples of size $N = 10^6$ and 10^7 using importance sampling.

function of relative potential, $\rho_*(\Psi)$. Assuming the stellar system is self-gravitating, a second relationship between these quantities is given by Poisson's equation, $\nabla^2\Psi = -4\pi G\rho_*(r)$, which can be solved to obtain $\Psi(r)$ (see Binney & Tremaine (2008) for details), allowing the DF to be specified in terms of r and v.

Figure 2 shows stellar mass-density (top) and radial velocity dispersion (bottom) profiles that we obtain for King models with $\Psi(0)/\sigma^2 = 9$, 6, and 3. The profiles are normalized by core radii and core densities that are related by $r_0 \equiv \sqrt{9\sigma^2/(4\pi G\rho_0)}$. Smooth curves are calculated directly from model parameters, while discrete data points represent estimates that we obtain by binning samples of size $N = 10^5$ that we obtain by rejection (blue) and importance (red) sampling. For the importance sampling, our proposal density is a 2D step function of $N_{\rm step} = 30$ steps in both position and velocity dimensions, with the proposal sample size equals $M = 3N$. Reassuringly, the estimated profiles approximate the calculated ones. For the case of $\Phi/\sigma^2 = 9$, Figure 3 displays contours of the DF in the plane of position and velocity, with sample size at $N = 10^6$ and 10^7. Again we find that estimates obtained by binning our samples, (here, for clarity, we show only the estimates from our samples obtained via importance sampling) correspond to the contours calculated directly from the DF.

4.2 Osipkov-Merritt-Zhao model

The Osipkov-Merritt (OM) DF, proposed by Osipkov (1979) and Merritt (1985), depends on angular momentum (J) as well as relative energy ($\mathcal{E} = \Psi - v^2/2$) per unit mass through the variable $Q \equiv \mathcal{E} - J^2/(2r_a^2)$, where r_a sets the scale for a transition from isotropic velocity dispersion at small radii to radially-biased dispersion at large radii: $\beta(r) \equiv 1 - (\sigma_\theta^2 + \sigma_\phi^2)/(2\sigma_r^2) = r^2/(r_a^2 + r^2)$. Thus the OM DF can be expressed as $f(\mathcal{E}, J^2) = f(Q)$, with $f(Q) = 0$ where $Q < 0$. Integrating this DF over velocity space, the stellar mass-density profile is

$$\rho_*(r) = \frac{4\pi}{1 + r^2/r_a^2} \int_0^\Psi dQ\, f(Q)\sqrt{2(\Psi - Q)}. \tag{2}$$

Via Abel transform, the DF is obtained as

$$f(Q) = \frac{1}{2\pi\sqrt{2}} \frac{dG(Q)}{dQ} \tag{3}$$

with

$$G(Q) = -\int_0^Q \frac{d\rho_Q}{d\Psi} \frac{d\Psi}{\sqrt{Q - \Psi}} \tag{4}$$

and

$$\rho_Q(r) = \left(1 + \frac{r^2}{r_a^2}\right) \rho_*(r). \tag{5}$$

In our OM-Zhao model that we provide as an example in StarSampler, we consider a stellar population that is described by a generalized Hernquist profile (Hernquist 1990; Zhao 1996),

$$\rho_*(r) = \rho_s \left(\frac{r}{r_*}\right)^{-\gamma_*} \left[1 + \left(\frac{r}{r_*}\right)^{\alpha_*}\right]^{(\gamma_* - \beta_*)/\alpha_*} \tag{6}$$

and embedded within a dark matter halo that has density profile of the same general form (but allows for different parameters),

$$\rho_{DM}(r) = \rho_d \left(\frac{r}{r_d}\right)^{-\gamma_d} \left[1 + \left(\frac{r}{r_d}\right)^{\alpha_d}\right]^{(\gamma_d - \beta_d)/\alpha_d} \tag{7}$$

We assume that the system is dark matter dominated, such that the stellar mass contributes negligibly to the potential. The general form of the DM halo profile does not guarantee that the enclosed-mass profile, $M(r) = 4\pi \int_0^\infty r^2 \rho_{DM}(r) dr$, converges to a finite value; therefore, we truncate the DM density and set $\rho_{DM} = 0$ for $r > r_{200}$, where r_{200} is the radius where the average mass within is 200 times the critical density of the universe.

To calculate the DF, we first tabulate the function $G(Q)$ according to equation 4. Since $G(Q)$ is monotonic and the DF depends on its derivative, we use monotone cubic interpolation to interpolate the

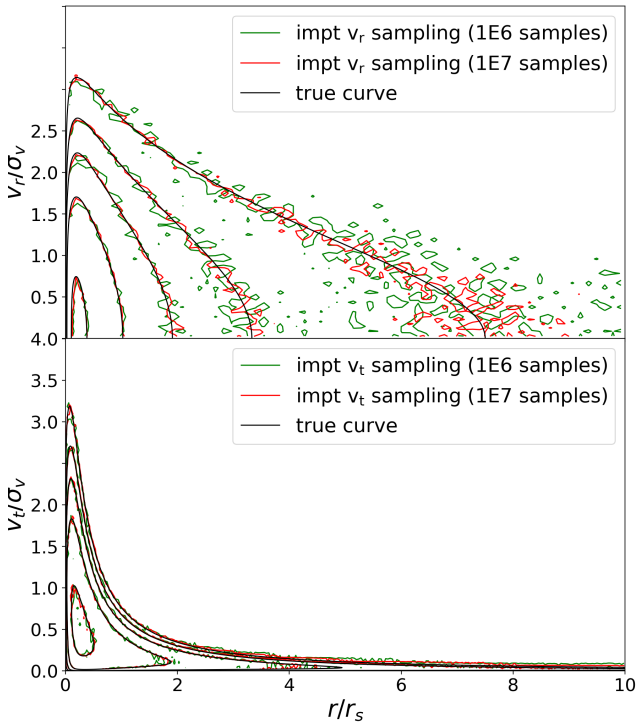

Figure 4. Contours of constant radial (top) and tangential (bottom) phase-space densities following the DF of Osipkov-Merritt-Zhao model. Black curves are calculated directly from the DF (see Section 4.2 for parameter values), while green and red profiles are estimated from samples of $N = 10^6$ and $N = 10^7$ stars drawn via importance sampling, respectively. Notice the characteristic change from isotropic central velocity dispersion to radially anisotropic dispersion at large radius.

table and then calculate $f(Q)$. Note that angular momentum $J = rv_t$ and speed $v = \sqrt{v_r^2 + v_t^2}$, where v_r and v_t are radial and tangential velocity, respectively. The final density function is expressed as $f(Q) = f(r, v_r, v_t)$.

Our implementation of the OM-Zhao model is specified by a set of parameters `model_param = [`r_a`,` r_*`,` α_*`,` β_*`,` γ_*`,` ρ_d`,` r_d`,` α_d`,` β_d`,` γ_d`]`. The first five parameters give the anisotropy radius and the stellar density profile, while the latter five parameters set the density profile of the dark matter halo. Figure 4 displays the contour plots of radial and tangential velocity densities for an OM model with `model_param = [.25 (kpc), .25 (kpc), 2, 5, .1, 4e8 (`$M_\odot$`/kpc`^{3}`), 1 (kpc), 1, 3, 1]`. Black curves are calculated directly from the DF, while green and red contours are estimates obtained from random samples of size $N = 10^6$ and $N = 10^7$, respectively. Samples are drawn using importance sampling. The proposal distribution is a 3D step function with 30 steps in each of the (r, v_r, v_t) dimensions with a proposed sample size of $M = 3N$. We see that with larger sample size, curves converges to the truth.

4.2.1 *Sampling from the conditional probability distribution*

For some combinations of OM-Zhao model parameters sampling can become prohibitively inefficient for the purpose of drawing large samples. For example, specifying a cosmologically-motivated limiting radius of $r_{\rm lim} = r_{200}$, where r_{200} at which the mean enclosed density is 200 times that of the cosmic background, implies large escape velocities and thus wide velocity ranges, increasing the sampling domain and lowering acceptance rates. In the case of the spherical OM-Zhao model, however, the DF can be conveniently marginalized, enabling a two-stage rejection sampling procedure that significantly increases efficiency.

Specifically, we can sample first from the 1D probability distribution for position, and then from the conditional distribution of velocity given position. Since the first step involves sampling in just one dimension, we can build an efficient proposal function, q_2, that satisfies the rejection sampling requirement that $q_2(r) \geq p(r)$. The steps in this alternative procedure, which is available in StarSampler's implementation of the OM-Zhao model, are

(i) Build proposal functions $q_2(r)$, $q_2(Q)$, based on $\rho_*(r)$ and $f(Q)$.
(ii) Draw radius samples $r^1, r^2, ..., r^N$, where $r^i \sim \rho_*(r)$ with rejection sampling using proposal density $q_2(r)$.
(iii) For each r^i, draw Q^i, where $Q^i \sim f(Q)$ with rejection sampling, using proposal density $q_2(Q)$.
(iv) For each r^i and Q^i, draw $v_r^i \sim \text{Uniform}(0, v_{\rm escape})$, and calculate v_t^i from v_r^i, r^i and Q^i.

We emphasise that `StarSampler`'s primary purpose is to provide a sampler that works for arbitrary spherical DFs. While the two-step sampling routine above enables more efficient rejection sampling in the specific case of the OM DF, some DFs will be sufficiently complicated that the conditional probability distribution cannot be so easily calculated. In such cases, `StarSampler`'s original sampling of the multivariate DF provides a robust option. It is to an example of such a complicated DF that we now turn.

4.3 Strigari-Frenk-White (SFW) model

Strigari et al. (2017, 'SFW' hereafter) introduce a flexible DF that is a separable function of energy and angular momentum: $f(E, J) = h(E)g(J)$. The SFW model allows a wide range of energy distributions and velocity dispersion anisotropies. The distribution of angular momentum is defined as

$$g(J) = \left[\left(\frac{J}{J_\beta} \right)^{\frac{b_0}{\alpha}} + \left(\frac{J}{J_\beta} \right)^{\frac{b_1}{\alpha}} \right]^\alpha , \tag{8}$$

describing the transition, at scale angular momentum J_β, from a power law of index b_0 to one with index b_1. The parameter α controls the rapidity of the transition, and is required to be positive when $b_1 > b_0$ and negative when $b_1 \leq b_0$.

The energy distribution function is

$$h(E) = \begin{cases} NE^a(E^q + E_c^q)^{d/q}(\Phi_{lim} - E)^e & \text{for } E < \Phi_{lim} \\ 0 & \text{for } E \geq \Phi_{lim} \end{cases} . \tag{9}$$

where N is a normalization constant. The index a determines the behaviour of the function at small energies. Near the characteristic energy E_c and beyond, the power law has slope $a + d$, where q specifies the rapidity of the transition. By definition, orbits with $E > \Phi(r_{\rm lim})$ are unbound, effectively truncating the stellar density at limiting radius $r = r_{\rm lim}$.

As with the OM model, our implementation of the SFW model assumes that the gravitational potential is dominated by dark matter. While SFW assumed that the dark matter halo follows the density profile suggested by Navarro et al. (1997), here we allow for the more general form given by Equation

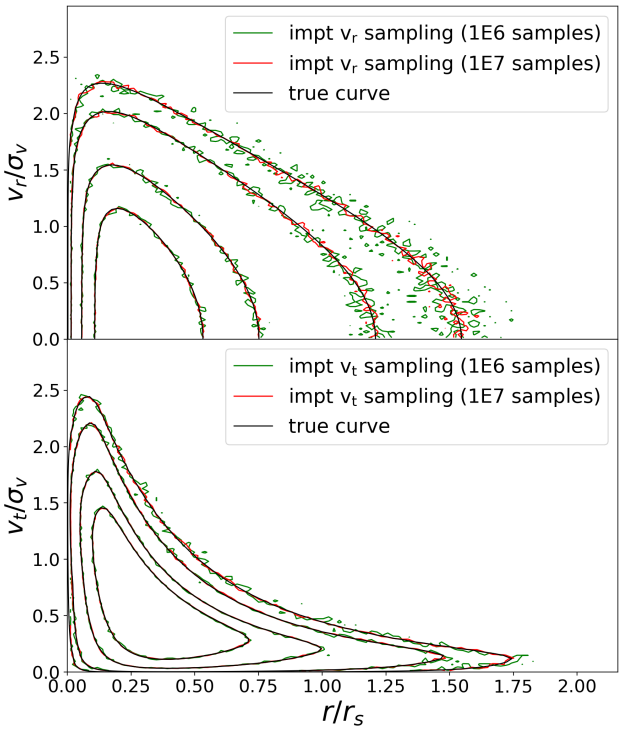

Figure 5. Contour plots of radial and tangential densities. Same as Figure 4, except that here we use the model of Strigari et al. (2017). See Section 4.3 for parameter values.

7. Therefore the full parameter set for our implementation of the SFW model is `model_param = [a, d, e, Ec, r_lim, b_0, b_1, |α|, q, J_β, ρ_d, r_d, α_d, β_d, γ_d]`

Figure 5 displays the radial and tangential velocity distributions for the model that SFW propose to describe the 'metal-rich' stellar population of the Sculptor dwarf galaxy (see their Table 1), with parameters `model_param = [2, -5.3, 2.5, .16(Φ_d), 1.5 (kpc), 0, -9, 1, 6.9, .086 (r_d√Φ_d, 8e7 (M_⊙kpc^-3), 0.694 (kpc), 1., 3., 1.]`, where $\Phi_d \equiv 4\pi G \rho_d r_d^2$ (in units of km^2 s^{-2}). As in Figure 4, we generate samples of size $N = 10^6$ and $N = 10^7$ using importance sampling (with proposal 3D step function $N_{\rm step} = 30$ in each dimension, and a proposal sample size of $M = 3N$). Binning these samples to obtain the contour plots of velocities in both radial and tangential direction. We recover the expected curves in both directions.

4.4 A toy model

As an example of an arbitrary DF we consider a 4D probability function. The density has two "spatial" coordinates and two "velocity" coordinates: $f(x_1, x_2, v_1, v_2) = h(x_1, x_2)g(v_1, v_2)$, where $h(x_1, x_2)$ is a two-dimensional mixture of five Gaussians investigated by Shaw et al. (2007). The X_1, X_2 positions of the peaks and the widths σ and amplitudes A of the Gaussians are summarized in Table 1. The velocity distribution $g(v_1, v_2)$ takes the following form,

$$g(v_1, v_2) = \frac{1}{18\pi} e^{\frac{-(v_1^2+v_2^2)}{18}} \cos^2\left(\frac{v_1}{2}\right) \cos^2\left(\frac{v_2}{2}\right) \qquad (10)$$

We sample from the toy DF and the histograms of the sampled x_2 and v_2 are plotted in Figure 6, together with the corresponding

Peak	X_1	X_2	A	σ
1	-0.400	-0.400	0.500	0.010
2	-0.350	0.200	1.000	0.010
3	-0.200	0.150	0.800	0.030
4	0.100	-0.150	0.500	0.020
5	0.450	0.100	0.600	0.050

Table 1. The set of parameters X_1, X_2, A, σ defining the 5 Gaussians for the toy model spatial density $h(x_1, x_2)$.

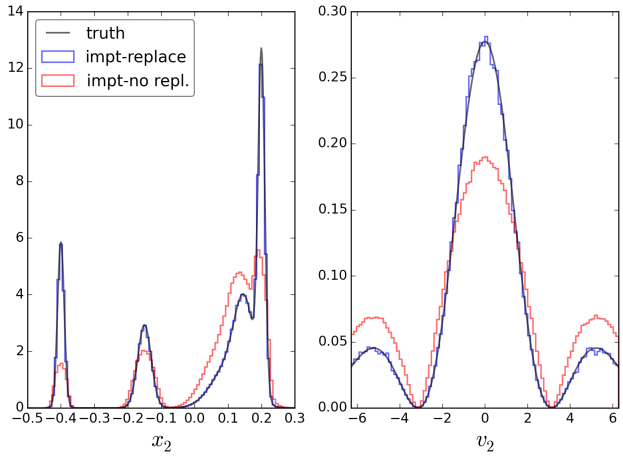

Figure 6. Marginal distributions of x_2 and v_2 of the toy model $f(x_1, x_2, v_1, v_2)$ are shown in left and right panels, respectively. The model is specified by the spatial parameters in Table 1 and velocity distribution in equation 10. The true distributions are the black curves. The blue curves are the histograms of the sample points drawn with replacement using importance sampling, and the red curves are the sample points drawn without replacement. The number of steps in the proposal function for both methods is set to be $N_{\rm step} = 30$, however, proposal sample size $M = 30N$ for sampling with replacement and $M = 150N$ for sampling without replacement.

marginalized distributions in black. We made two samples of 10^5 points using importance sampling methods: one with replacement in blue, and one without replacement in red. The step numbers in the proposal function for both methods are set to be $N_{\rm step} = 30$, however $M = 30N$ for sampling with replacement and $M = 150N$ for sampling without replacement. It is clear from the plots that sampling without replacement gives a very biased density estimate for both x_2 and v_2 distributions, the density is especially undersampled near the peak of the distribution. The sampling can be improved with larger $N_{\rm step}$ and M, but that is computationally expensive, especially to compute a 4D proposal function. On the other hand, the blue curves shows that sampling with replacement gives superior results even when it has lower resample factor. Therefore, given our proposal density we advise our users to use the importance sampling method with replacement as we do in all of the other test cases in this paper.

5 IMPORTANCE VS REJECTION SAMPLING

Importance sampling is generally faster than rejection sampling, in the sense that it requires fewer draws from the proposal distribution. With a uniform proposal distribution, depending on the shape of the

density function, the efficiency (i.e., the acceptance rate) of rejection sampling can be frustratingly low. On the other hand, once the proposal step function is built, the number of draws required by importance sampling can be just a few times the desired sample size. The tradeoff is the additional sampling variance introduced because we are sampling from an approximated density, which approaches the truth with increasing number of proposal points. The variance can be reduced by using a larger number of steps in proposal function; however, this can be costly, as the number of function evaluations for the proposal scales exponentially with sample dimensions. For the 3-dimensional (r, v_r, v_t) OM and SFW models considered here, it requires $N_{\rm step}^3$ function evaluations just to build the proposal density. Another way to reduce sampling variance is by drawing a larger number of proposal points, which is a more efficient approach with higher dimensional DFs. Overall, for higher dimensional DFs ($> 4D$), rejection sampling might be the preferred method.

Furthermore, sample points from importance sampling can have multiple copies because of the weighted resampling of the proposal is performed with replacement. On the other hand, in some cases it may be desirable to use the importance samples to obtain rough estimates of, say, the surface density and velocity dispersion profiles. The top two panels of Figure 7 show the root mean square (RMS) percentage error between importance sample-estimated projected stellar density profiles and the theoretical projected profiles, versus the proposal function step numbers ($N_{\rm step}$), for sample sizes $N = \{10^3, 10^4, 10^5, 10^6\}$ and proposal sample size of $M = 3N$. For consistency, all profiles are represented by 50 bins, with each bins contains equal number of stars. The top left panel is the result of OM model and the top right panel shows the SFW model. In addition, the RMS percentage error based on rejection sampling are plotted in dash lines. We see a few unsurprising trends. First, we see that as number of steps in the proposal function increases, the percentage error decreases. The effect is more prominent with smaller sample size. Second, overall error reduced with larger sample size, and the effect $N_{\rm step}$ diminishes with larger N. Third, as expected, rejection sampling generally gives lower RMS errors, however, with large sample size or large $N_{\rm step}$, two sampling methods become comparable. Lower two panels of Figure 7 shows the RMS percentage errors for the velocity dispersion profiles. We see the same trends as in the top two panels, except that the error is much less compared to the error in the projected density profiles. It is worth noting that the SFW model presented has generally lower RMS errors compared to the OM model, and that the RMS error is almost having no dependence on the $N_{\rm step}$. Therefore the optimal choice of $N_{\rm step}$ depends on the specific DF, as expected.

To explore the impact of the proposal sample size in the importance sampling, RMS percentage error is calculated as a function of resample factor (rf), the multiple of the sample size that sets the number of the proposal points. The proposal density is set with a fixed $\rm N_{step} = 20$. The result is shown in Figure 8 in solid lines; dash lines are based on rejection sampling for reference. We see similar trends as in Figure 7, larger resample factor results in lower RMS errors in OM model, for both projected density profile and the velocity dispersion profile on the left panels, and the reduction is most significant for smaller sample size. As the sample size increases, the proposal points also increases, and the effect of resample factor diminishes. For this particular SFW model, a resample factor of a few is sufficient.

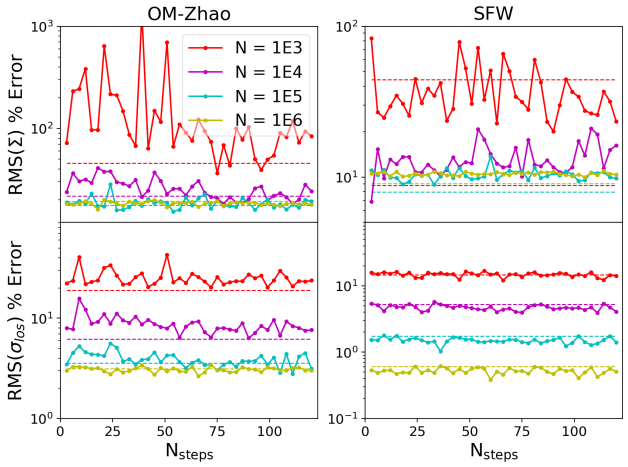

Figure 7. RMS percentage errors of the estimated projected stellar density profiles, based on importance sampling, versus number of step-function pieces used to generate the proposal density, are show in the top two panels, in solid lines. On the left (right) is OM (SFW) model, with same parameter as in Figure 4 (5). Dash lines shows the RMS % error based on rejection sampling. Sample sizes of $N = 10^3, 10^4, 10^5, 10^6$ are plotted. The lower two panels shows the same plots, but for the RMS % error of velocity dispersion profiles.

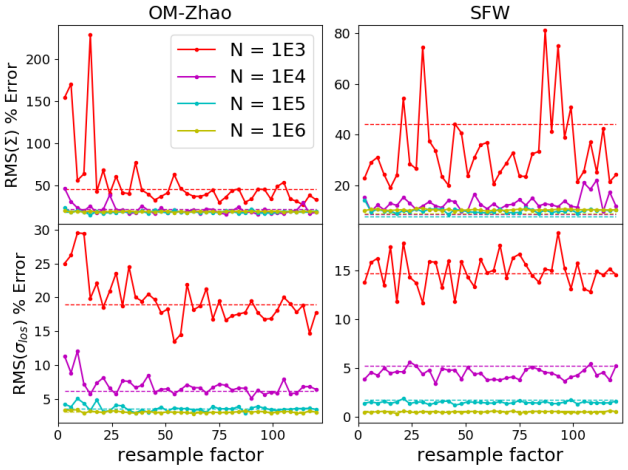

Figure 8. Similar to Figure 7 excepts that RMS % errors is calculated against the resample factor, with a fixed $\rm N_{step} = 20$. Dash lines are results from rejection sampling, for reference. Different colors show the RMS error of different sample sizes. Top panels are the RMS % errors in projected density profiles and bottoms two shows the error in velocity dispersion profiles. The OM model is shown on the left and SFW model is on the right.

6 IMPORTANCE SAMPLING APPLICATIONS

With importance sampling with replacement, the samples most likely contains repeated points, which is not appropriate for some applications such as for initialization of N-body simulations, where all the particles should be unique. However, there are some applications where one can take advantages of the importance sampling, in particular to make inference from a dataset where likelihood calculation is intractable. For example, one can use importance sampling to generate synthetic training data for machine learning methods to

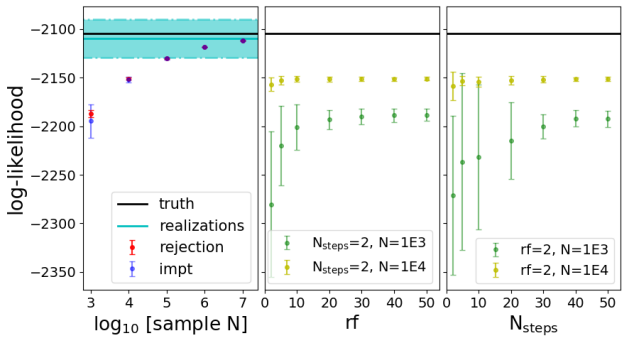

Figure 9. Likelihood estimate based on Gaussian Kernel Density estimate of the sample points. Left panel shows the estimated log-likelihood as a function of sample size N. The red points are based on rejected sampling and blue points are based on importance sampling with $N_{steps} = 20$ and rf = 5. Each point shows the mean and standard deviation of a hundred repeated estimates. The truth is the black horizontal line. The shaded region and the cyan line are the standard deviation and the mean of the true log-likelihood of 100 realization of the mock dataset. The middle panels shows the dependence of the log-likelihood as a function of the resample factor (rf), at a constant $N_{steps} = 2$ for sample $N = 1,000$ and 10,000 in green and yellow, respectively. In the same manner, the right panel shows the log-likelihood versus N_{steps} with constant rf = 2.

recover model parameters (e.g., Ntampaka et al. 2015; Ravanbakhsh et al. 2016; Levasseur et al. 2017). The input features could be a set of summary statistics that are insensitive to the repeated points, for example, the surface density and the velocity dispersion profiles. Another application is to use the sampling technique in conjunction with Approximate Bayesian Computation (ABC; Beaumont et al. 2002; Turner & Van Zandt 2012) inference method. ABC bypasses the likelihood evaluation by simulating artificial data on the fly from the model under consideration, which is compared with the real data through a set of summary statistics. ABC is used widely in cosmology (e.g., Hahn et al. 2017; Cameron & Pettitt 2012; Kacprzak et al. 2018; Weyant et al. 2013) and examples of the ABC implementations can be found in (Jennings & Madigan 2017; Akeret et al. 2015; Alsing et al. 2018).

The distribution of large number of sample points is an unbiased representation of the underlying density, and can be used to approximate the original density function. The idea is not new; it was also introduced by Diggle & Gratton (1984), who proposed to use simulated realizations of an implicit model to estimate the log-likelihood. Here we provide an example using the SFW model, with parameters `model_param = [5.30, -3.74, .948, .268(`Φ_d`), 3.91 (kpc), 0, -14.0, 1, 6.43, .286 (`$r_d\sqrt{\Phi_d}$`, .703 (`$M_\odot pc^{-3}$`), 0.183 (kpc), 1., 3., 1.17]`. Because it is a spherical model and typically we observe only 3D (x, y, v_z) out of the 6D coordinates, we are interested in estimating the density $P(R, v_z)$, where the projected radius $R = \sqrt{x^2 + y^2}$. Note that calculating $P(R, v_z)$ directly from $f(E, J)$ requires computationally expensive triple-integral. Specifically we first generate N samples $\{(x_j, y_j, z_j, v_{xj}, v_{yj}, v_{zj})|j = 1, ..., N\}$ from the model, then we keep only the $\{(R_j, v_{zj})|j = 1, ..., N\}$ points, and 2D Gaussian Kernel Density Estimate is applied to make a smoothed approximation of the density, $\hat{P}(R, v_z)$. Here we use the kernel bandwidth based on Silverman's rule of thumb. The likelihood of a mock dataset of a thousand stars (generated under the same model) is calculated using $\hat{P}(R, v_z)$, and the result is presented in Figure 9. The left panel shows the estimated log-likelihood as a function of the number of points N used to estimate the density. The red is the density estimate based on the rejection sampling, and the blue is based on the importance sampling (with rf = 5 and proposal $N_{steps} = 20$). Each point shows the mean and the standard deviation of a hundred likelihood estimates. We see that as N increases the estimated likelihood approaches the truth which is the solid black line. As a comparison of the estimation accuracy, we generate a hundred realizations of the mock dataset and calculate their true likelihood, their mean and standard deviation is plotted as cyan line and the width of the shaded region. We see that for this particular model, with a sample size of about 10^5 the estimated likelihood approaches a level comparable to the likelihood variation from the dataset. We also see that as N increases the there is virtually no differences between the density estimate using rejection sampling versus the importance sampling. For smaller sample size, importance sampling shows larger variances in the likelihood estimate, but that can be mitigated with larger resample factor (rf) and the larger number of proposal function steps. The middle panel of Figure 9 shows the improvement with larger resample factor (at a fixed $N_{steps} = 2$) for $N = 1,000$ and $10,000$, in green and yellow, respectively. Similarly the right panel shows the improvement with larger proposal function steps (at a fixed rf = 2) for $N = 1,000$ and $10,000$. But overall the likelihood estimate improvement depends mostly on the sample size N, with sensitivity to the choice of resample factor and proposal steps at a smaller sample size.

7 SUMMARY

We have introduced a Python software module that draws random samples from an arbitrary, user-defined DF The user can choose either rejection or importance sampling. The returned sample elements correspond to the function variables, which can be phase-space coordinates or functions thereof. The package also includes three pre-defined spherical stellar models that return 6-D phase-space coordinates corresponding to King, Osipkov-Merritt, and SFW models. The package is available at `https://github.com/maoshenl/StarSampler`.

Without prior knowledge of the shape of a general DF, a uniform function equal to the maximum of the function is used as proposal density for rejection sampling. Therefore the sampling efficiency can vary wildly with the same model under different parameter sets. Importance sampling can ameliorate this issue. It is an approximate sampling method, and a step function is built as a proposal density based on the true DF. The importance sampling speed doesn't depend on the shape of the DFs, but has an added cost of higher sampling variance. The sampling variance could be reduced with larger number of steps in the proposal function and/or larger number of proposal sample points. The function evaluations needed to construct proposal function scales to the power of the DF dimensions, which could be very computationally expensive. Samples from importance methods could also be unrealistic due to resampling of the proposal samples with replacement, although it is mitigated by adding random angular position and velocity when there is symmetry in the phase space coordinates.

We compared the difference between two sampling methods on some of the DFs, and suggested a few applications of the importance sampling. Although these quantitative results cannot be generalized to arbitrary DFs, robust qualitative results include that variance due to low sample size dominates the additional variance from the

importance sampling, and that with large sample points ($> 10^6$), importance sampling also converge to the truth, with little or no dependence on the $N_{\rm step}$ and the resample factor. The technique can be useful to quickly generate approximate samples, and/or to allow inferences about model parameters and gravitational potentials.

ACKNOWLEDGEMENTS

We thank Jessi Cisewski and Brendan McVeigh for helpful discussions. M.G.W. is supported by National Science Foundation grants AST-1313045 and AST-1412999.

REFERENCES

Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, Journal of Cosmology and Astroparticle Physics, 2015, 043
Alsing J., Wandelt B., Feeney S., 2018, arXiv preprint arXiv:1801.01497
Ascasibar Y., 2008, Computer Physics Communications, 179, 881
Beaumont M. A., Zhang W., Balding D. J., 2002, Genetics, 162, 2025
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Cameron E., Pettitt A., 2012, Monthly Notices of the Royal Astronomical Society, 425, 44
Devroye L., 1986, in Proceedings of the 18th conference on Winter simulation. pp 260–265
Diggle P. J., Gratton R. J., 1984, Journal of the Royal Statistical Society. Series B (Methodological), pp 193–227
Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, Bayesian data analysis. CRC press
Gieles M., Zocchi A., 2015, MNRAS, 454, 576
Givens G. H., Raftery A. E., 1996, Journal of the American Statistical Association, 91, 132
Hahn C., Vakili M., Walsh K., Hearin A. P., Hogg D. W., Campbell D., 2017, Monthly Notices of the Royal Astronomical Society, 469, 2791
Hernquist L., 1990, ApJ, 356, 359
Jennings E., Madigan M., 2017, Astronomy and computing, 19, 16
Kacprzak T., Herbel J., Amara A., Réfrégier A., 2018, Journal of Cosmology and Astroparticle Physics, 2018, 042
King I. R., 1966, AJ, 71, 64
Levasseur L. P., Hezaveh Y. D., Wechsler R. H., 2017, The Astrophysical Journal Letters, 850, L7
McMillan P. J., Dehnen W., 2007, MNRAS, 378, 541
Merritt D., 1985, AJ, 90, 1027
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nelder J. A., Mead R., 1965, The computer journal, 7, 308
Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, ApJ, 803, 50
Osipkov L. P., 1979, Pisma v Astronomicheskii Zhurnal, 5, 77
Pelupessy F. I., van Elteren A., de Vries N., McMillan S. L. W., Drost N., Portegies Zwart S. F., 2013, A&A, 557, A84
Ravanbakhsh S., Oliva J. B., Fromenteau S., Price L., Ho S., Schneider J. G., Póczos B., 2016, in ICML. pp 2407–2416
Richardson T., Fairbairn M., 2014, MNRAS, 441, 1584
Rubin D. B., 1988, Bayesian statistics, 3, 395
Shaw J. R., Bridges M., Hobson M. P., 2007, MNRAS, 378, 1365
Smith A. F., Gelfand A. E., 1992, The American Statistician, 46, 84
Strigari L. E., Frenk C. S., White S. D. M., 2017, ApJ, 838, 123
Turner B. M., Van Zandt T., 2012, Journal of Mathematical Psychology, 56, 69
Vasiliev E., 2018, preprint, (arXiv:1802.08255)
Walker M. G., Peñarrubia J., 2011, ApJ, 742, 20
Weyant A., Schafer C., Wood-Vasey W. M., 2013, The Astrophysical Journal, 764, 116
Zemp M., Moore B., Stadel J., Carollo C. M., Madau P., 2008, MNRAS, 386, 1543
Zhao H., 1996, MNRAS, 278, 488

This paper has been typeset from a TEX/LATEX file prepared by the author.