MNRAS **000**, 1–4 (2017) Preprint 30 September 2017 Compiled using MNRAS LaTeX style file v3.0

Efficiently sampling rare events in population synthesis models

Floor Broekgaarden et al. ,[1,2]⋆ (TBD)
[1] *Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100, København ø, Denmark*
[2] *Astronomical Institute Anton Pannekoek, University of Amsterdam, P.O. Box 94249, 1090 GE, Amsterdam, The Netherlands*

30 September 2017

ABSTRACT
We present an adaptive importance sampling method that can significantly enhance stellar evolution simulations, especially when considering rare events. Simulations often involve calculating integrals over the initial parameter space, e.g. when calculating the fraction of binary black hole mergers. The method presented here estimates the wanted outcome by drawing samples from an instrumental distribution that is adaptively build-up from the function output. We test the performance of the method on rapid binary population synthesis models to estimate (i) the fraction of BBH mergers and (ii) the chirp mass distribution. We find that this method reduces the costs of the simulation up to a factor $\sim Y$.

Key words: importance sampling – population synthesis – gravitational waves

1 INTRODUCTION

Binary population synthesis models are a versatile tool in astrophysics to make predictions for populations of stars and the rate of astrophysical transient events of stellar origin. Binary population synthesis models have now been used to study a variety of astrophysical problems ranging from the characteristics of young stellar populations and how they are affected by the products of binary interaction (de Mink 2013, Schneider+14) to the end stages as core collapse supernova (Zapartas+17), and the more exotic outcomes such as type IA supernovae (Toonen), and gravitational wave sources (Stevenson+2017).

The models include a large variety of physical processes that can take place during the evolution of a star in a binary system such as super nova explosions, stellar winds, mass transfer and common envelope evolution. Examples of binary population synthesis models are BSE (Hurley+2000, Hurley+2002 and references within),binary_c (Izzard+2004, Izzard+2009), StarTrack (Belczynski+2005), SEBA (Portegies-Zwart 1996, Verbunt+1996) and COMPAS (Stevenson+2017). These models interpolate between evolutionary tracks of single stars obtained with a detailed stellar structure code (Pols+98) and rely on an approximate treatment of the physical processes. Therefore, they can present a rapid code that can evaluate the evolution of many stars and populations of stars. However, due to the multi-scale nature, complex processes involved and many initial parameters of the evolution, modelling a full population of binary systems is still computationally expensive and simulations are often limited by the scarce computational resources. The computational cost therefore can be a limiting factor in our exploration of the parameter space and hence understanding of the model outcome.

Especially when simulating a process that involves rare events, many simulations are needed before a simulation outcome is determined with certain precision. An example of a rare event are the initial binary systems that eventually produce gravitational waves. To evolve to a binary black hole (BBH) that produces gravitational waves (GWs) that can be observed by GW detectors LIGO and Virgo, the initial binary system has to start with massive stars and survive processes such as mass transfer, common envelope evolution and supernova kicks. Therefore, only a very small fraction of the initial parameter space of the population of binary systems eventually produces BBHs that can merge within Hubble time and produce GWs. Nevertheless, knowing which part of the parameter space produces the BBHs and GWs and their properties and comparing this with the recent observations of GWs (Abbott+2016) help improve our understanding of the binary population synthesis models and physical processes included.

In this paper we describe a method that aims to reduce the computational cost of the simulation of rare events in binary population synthesis models by using a method called importance sampling. We investigate the computational benefit of this method over the techniques that are traditionally applied such as Monte Carlo sampling. We test

---

⋆ E-mail: fsbroekgaarden@gmail.com

our method for the binary population synthesis model COMPAS to efficiently calculate the fraction of binary stars that eventually produce GWs within Hubble time, and the chirp mass distribution of these BBHs.

2 METHODS

Calculating informative properties of the outcome of simulations such as the fraction of BBHs that produce GWs or their chirp mass distribution often involves calculating integrals. However, since in a simulation only a finite number of simulations are run, in practice it comes down to estimating the value of the integral from the evaluated simulations. The most used method to do this is the Monte Carlo method. The aim of this paper is to introduce a method that improves this estimation compared to the Monte Carlo method. Therefore, we will first introduce the Monte Carlo method before explaining the proposed method.

2.1 Monte Carlo estimator

Let $\boldsymbol{x} = x_1, x_2, ...x_d$ be a random variable of dimension d and let $p(\boldsymbol{x})$ be the initial probability distribution function of $\boldsymbol{x}$. Let $\phi(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ be a function that evaluates the input variable $\boldsymbol{x}$ in the model $u(\mathbf{x})$ and maps it to an output of interest. For example, $u(\mathbf{x})$ can represent the population synthesis model and $\phi(\boldsymbol{x})$ can be the function that maps the initial mass of a star to its final mass.

The basic principle of the Monte Carlo method is to generate a finite number of random samples $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N$ that are identically and independently distributed from $p(\mathbf{x})$ and represent the distribution. The expectation value of $\phi(\mathbf{x})$ can then be estimated with the Monte Carlo method by

$$\hat{M}[\phi(\mathbf{x})] = \frac{1}{N}\sum_{k=1}^{N}\phi(\mathbf{x_k}) \approx \mathbb{E}[\phi(\mathbf{x})]. \tag{1}$$

The Monte Carlo method estimator has a convergence rate of $O(\frac{1}{\sqrt{N}})$, which can be derived from the central limit theorem. Although the convergence rate is independent of the number of dimensions, it is also relatively slow: to decrease the error of (1) by a factor of ten, one needs to increase the number of samples by a factor of hundred. The goal of this paper is to try to improve the estimator (1) focusing on simulating rare events with binary population synthesis models.

2.2 Importance sampling

Importance sampling can reduce the error of estimation obtained after a given number of evaluations (and therefore the costs of the simulation), by taking the random variables $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N$ from a so-called *instrumental distribution* $g(\mathbf{x})$, which is a different distribution than the prior distribution $p(\mathbf{x})$ but acts on the same parameter space. The idea is to take an instrumental distribution that samples more sample points $\mathbf{x_k}$ in the part of the initial parameter space that contributes most to our output parameter space of interest (e.g. BBH mergers). Especially when simulating a process where a small part of the initial space produces the output of interest (i.e. a rare event), changing the sampling distribution can significantly reduce the costs of the simulation. The intuitive idea of importance sampling is shown in Fig. (1).

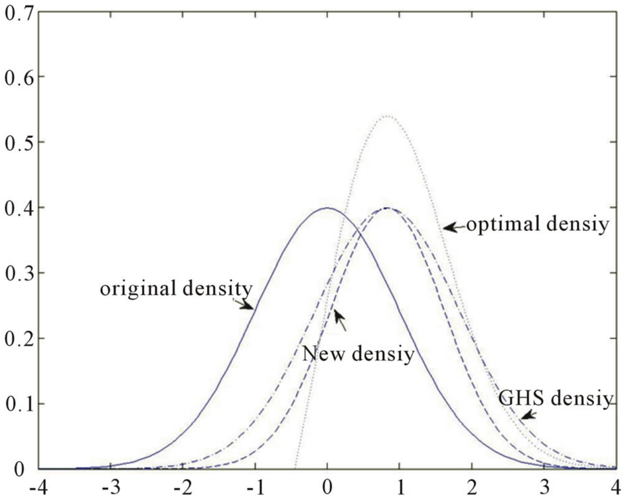

Figure 1. The intuitive idea of importance sampling is that often in the initial distribution function only very few of the drawn samples will contribute to calculating the integral of the output function (especially in rare events). The idea of importance sampling is to change the distribution from which the random variables are drawn such that a higher fraction of the drawn samples contribute to the outcome.

However, since the sampling distribution is changed, weights are introduced that correct for this in the estimation of the expectation of $\phi(\mathbf{x})$. If $\mathbf{x}$ is initially distributed by $p(\mathbf{x})$ and the instrumental distribution $g(\mathbf{x})$ is used, the estimator for the expectation value of $\phi(x)$ via importance sampling is given by

$$\hat{I}[\phi(x)] = \frac{1}{N}\sum_{k=1}^{N}\phi(x_k)\frac{p(x_k)}{g(x_k)}, \tag{2}$$

where $p(\mathbf{x_k})/g(\mathbf{x_k}) = w_k$ are the weights.

2.3 Adaptive importance sampling

Often the output function $\phi(\mathbf{x})$ is not known before running any simulations, hence the instrumental distribution $g(\mathbf{x})$ cannot be determined on beforehand. Instead, we use an adaptive sampling scheme that adaptively samples from an instrumental distribution $g_i(\mathbf{x})$ that is based on earlier model outcomes. The basic algorithm of the method works as follows:

(i) Sample random variables $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N$ from the initial distribution $p(\mathbf{x})$ and evaluate $\phi(\mathbf{x_k})$ for all $k = 1, 2, ..N_{\rm ini}$ until a certain threshold is reached. This threshold can for instance be a number of successful evaluations e.g. "*when* 100 *binary black hole mergers are simulated*" or "*when* 100 *binary black holes with chirp mass above* 20 $M_\odot$ *are simulated*". In each case there is an initial number $N_{\rm ini}$ of sample points in parameter space needed to produce N_s initial successes of the model (where $N_s = 100$ in our examples).

(ii) Now define the instrumental distribution $g_1(\mathbf{x})$ as a mixture of N_s Gaussian distributions around the N_s successful sample points in the initial parameter space. The instrumental distribution is then described by

$$g_1(\mathbf{x}) = \sum_{i=1}^{N_\mathrm{s}} \frac{1}{N_\mathrm{s}} \mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}), \tag{3}$$

where each Gaussian distribution $\mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ is equally weighted with $1/N_\mathrm{s}$ in the mixture distribution. The idea is thus that each Gaussian distribution $\mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ as part of the mixture $g(\mathbf{x})$ is drawn around one "successful" sample point $\mathbf{x_i} \in \{\mathbf{x_i}\}_{i=1}^{i=N_\mathrm{s}}$. This implies that the means $\mu_\mathbf{i}$ of the individual Gaussians in Equation (3) are given by

$$\mu_\mathbf{i} = \mathbf{x_i} \qquad \text{for } i = 1, 2, ..., N_\mathrm{s}. \tag{4}$$

The covariance matrix Σ is chosen to scale with the average expected distance between two sample points $\{\mathbf{x_k}\}_{k=1}^{k=N_\mathrm{ini}}$ in our initial parameter space. This is chosen such that samples drawn from a Gaussian $\mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ will generally fall in between the successful point (and mean of the Gaussian) x_i and its nearest neighbour. For simplicity we choose $\Sigma_i = \Sigma$ for all i and also a diagonal covariance matrix for Σ given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots \\ \vdots & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}, \tag{5}$$

where each σ_k is given by

$$\sigma_k = \frac{\|\mathrm{max}_k - \mathrm{min}_k\|}{(N_\mathrm{ini})^{1/d}} \quad \text{for } k = 1, .., \mathrm{d}. \tag{6}$$

In Equation (6) max_i and min_i are the maximum and minimum range of x_i. (NB: for later generations of $g(\mathbf{x})$ N_ini is changed into N_tot, the total number of samples that were drawn and evaluated in the simulation at this stage).

(iii) New samples $\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_K}$ are drawn from the instrumental distribution $g_1(\mathbf{x})$ given by Equation (3) and evaluated in the function $\phi(\mathbf{x})$. Since we are now drawing samples from the instrumental distribution that is focused in the initial parameter space around the initial samples that produced an outcome of interest, the samples that we draw from the instrumental distribution will more often also produce the rare event - as long as the output space does not behave too chaotic or stochastic.

(iv) From these evaluations the expected value of $\phi(\mathbf{x})$ can be estimated using Equation (2) and the uncertainty of the estimation.

(v) Increase the number of sample points, or repeat steps (ii) to (iv) to update the instrumental distribution (which will converge to the distribution of the rare event) until a certain error is reached.

3 A TEST CASE APPLICATION: BBH MERGERS

Consider for $u(\mathbf{x})$ the binary population synthesis model COMPAS that simulates the evolution of binary systems and focuses on the evolution to compact objects such as neutron stars and black holes. Suppose the initial parameter space is 3-dimensional (d= 3) with parameters $\mathbf{x} = (M_1, a, q)$, where M_1 is the initial mass in solar mass $M_\odot$ of the most massive star (the primary) in the binary system, a is the initial separation of the binary given in AU and q the initial mass ratio of the binary, i.e. $q = M_2/M_1$ where M_2 is the mass of the secondary. Suppose the outcome of interest is the fraction of binary black hole mergers $f_\mathrm{BBH\ merger}$, our output function $\phi(M_1, a, q)$ can then be given by

Table 1. Summary of properties initial parameters

parameter	pdf	range
M_1	$p(M_1) \propto M_1^{-2.35}$	$[7, 100]\mathrm{M}_\odot$
a	$p(a) \propto 1/a$	$[0.1, 10^3]\mathrm{AU}$
q	1	$[0, 1]$

$$\phi(M_1, a, q) = \begin{cases} 1 \text{ if } u(M_1, a, q) \text{ produces a BBH merger} \\ 0 \text{ else} \end{cases} \tag{7}$$

The fraction of the initial parameter space that will produce BBH mergers when evaluated in the model (i.e. $f_\mathrm{BBH\ merger}$) can then be estimated by using Equation (2) by simulating N binary systems with the adaptive importance sampling method. In other words

$$f \approx \hat{I}[\phi(M_1, a, q)] = \frac{1}{N} \sum_{i=1}^{N} \phi(M_1, a, q) \frac{p(M_1, a, q)}{g(M_1, a, q)}. \tag{8}$$

Assuming $p(M_1, a, q) = p(M_1)\ p(a)\ p(q)$ the prior is given by the product of the individual probability distribution functions which are summarized in Table 1. Using these distributions, we find

$$p(M_1, a, q) \propto \frac{M_1^{-2.35}}{a} \tag{9}$$

Following the algorithm described in Section (2.3) we define the instrumental distribution

$$g_1(M_1, a, q) = \sum_{i=1}^{N_\mathrm{s}} \frac{1}{N_\mathrm{s}} \mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma}). \tag{10}$$

By filling in Equations (7), (9) and (10) into Equation (8) we have all the ingredients for the importance sampling estimator.

4 PRELIMINARY RESULTS

To test how well the adaptive importance sampling method works we run a large Monte Carlo simulation with more than 10^7 sample points to estimate the fraction of the BBH mergers within the mentioned initial parameter space up to error $1.4 \cdot 10^{-5}$. We find $f = 0.002797 \pm 14$. We then run the simulation using the adaptive importance sampling method with different number of total samples N_tot and estimate the fraction of BBH mergers with Equation (8) and compare this with the value for the fraction found by the large Monte Carlo run. We also run the simulation with N_tot samples using the crude Monte Carlo method as given in Equation (1) and add the estimated error from the true fraction to the plot.

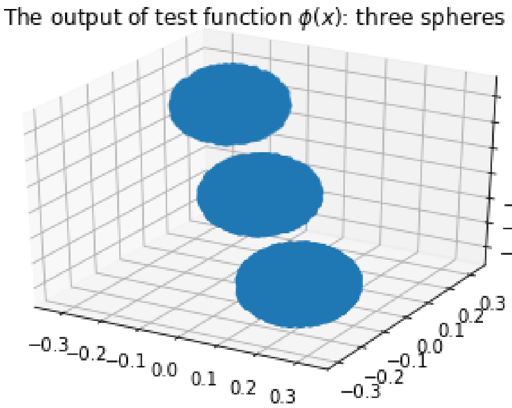

Figure 2. Plot of the test function $\phi(x)$ that is used to obtain the preliminary results. $\phi(x)$ maps all x within the sphere (blue dots) to 1 and the rest to zero. The volume of the spheres is chosen such that the fraction of successful samples (blue dots) is representative for COMPAS.

At the moment of writing, the full code that works for the binary population synthesis model COMPAS is still in progress. Nevertheless, we have tested the adaptive importance sampling on a test problem and show in this section the results of these tests.

For the preliminary tests, we created a 3-dimensional parameter space $\Omega_3 = [-1,1]^3$. We defined $\phi(x)$ to be a function on Ω_3 such that it maps a fraction $f \sim 0.00345$ of the full parameter space to 1 whilst the rest maps to zero (i.e. $\phi(x) = 1$ on three spheres in Ω_3 and 0 else). This value for the fraction is chosen such that it is representative for the fraction of BBHs that merges within Hubble time in simulations run with population synthesis model COMPAS. The regions that map to zero are defined as three spheres within the parameter space, which are shown in Figure 2.

We initially draw samples from a 3D uniform distribution and after thirty of the initial samples fall in the volume of the spheres (and thus evaluate to 1 instead of 0) the initial sampling is stopped and we change to the adaptive importance sampling scheme. To test the performance of the method, we run the simulation several times for multiple N_{tot} and compute the error. We compare the results with the Monte Carlo method and the true value of the fraction, which in this case can be determined analytically by the volume of the spheres. The results are shown in Figure 3. From this Figure it can be seen that:

- The error of the estimation for both methods is smaller for larger N_{tot}. This is expected as more simulation runs, and thus more computational cost, will usually give better results.
- The errors of the adaptive importance sampling method are always smaller than the errors from the Monte Carlo method. This means that with the same number of runs, the adaptive importance sampling method gives better results than the Monte Carlo method.
- The adaptive importance sampling method is a factor Y more efficient: the same error is obtained with Y times less sample points.

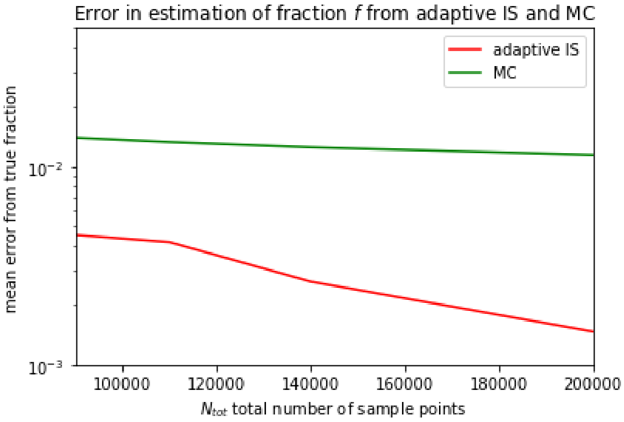

Figure 3. Mean absolute error of the adaptive importance method (red) and the Monte Carlo method (green) as a function of the number of simulations run N_{tot} (i.e. the computational costs). These results are still preliminary as it obtained from tests with the toy population synthesis model. We are working on performing a similar test when the code is fully adapted for binary population synthesis codes like COMPAS.

ACKNOWLEDGEMENTS

We thank the the Kavli Foundation, Niels Bohr institute and DARK Cosmology Centre in Copenhagen for their hospitality and for organising the Kavli summer school in gravitational waves astrophysics 2017. This work could not have been done without their support.

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