Efficiently sampling rare events in population synthesis models

Floor Broekgaarden,[1⋆] Stephen Justham,[1,2] Ilya Mandel[1,3] Jonathan Gair[1,4]
Tassos Fragos[1,5]
[1] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100, København ø, Denmark
[2] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing, China
[3] School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK
[4] School of Mathematics, University of Edinburgh, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK
[5] Geneva Observatory, University of Geneva, Chemin des Maillettes 51, 1290 Sauverny, Switzerland

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 estimates the wanted property by drawing samples from an instrumental distribution that is adaptively build-up from the function output. We test 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 Y.

Key words: importance sampling – population synthesis – gravitational waves

1 INTRODUCTION

Rapid binary population synthesis models are a versatile tool in astrophysics to study the evolution of populations of stars and compare theory and observations. The models include a large variety of physical processes that can take place during the evolution such as super nova explosions, stellar winds, mass transfer and common envelope evolution. Examples of binary population synthesis models are binary_c (REFs), Startreck (REF), SEBA (REF) and COMPAS (REF). These models ~~are 1D~~ ~~and often~~ interpolate physics from more detailed simulations and can therefore present a rapid code that can evaluate the evolution of many stars and populations of stars. However, due to the multi-scaleness, complex processes involved and many initial parameters of the evolution, modelling many stars is still computationally expensive and simulations are often limited by scarce computational resources.

The computational cost ~~therefore limits our~~ exploration of the parameter space and hence our understanding of the model outcome.

They include a large variety of underlying binary interaction processes that are challenging to model and induce uncertainties in the outcome of the model. Hence, to fully understand the simulations outcome and subsequently the underlying physics, it is important to incorporate uncertainty from the beginning of the model instead of as an afterthought~~, to reduce its computational cost.~~ 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 ~~eventually~~ evolve to a binary black hole (BBH) that produces gravitational waves that can be observed by LIGO and Virgo, the initial binary system has to start with massive stars and survive processes such as kicks. Therefore, only a very small fraction of the initial population of binary systems space forms BBHs that can merge and produce GWs.

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. (...)

2 METHODS

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

⋆ E-mail: mn@ras.org.uk (KTS)

© 2017 The Authors

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 (Ref). 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 (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).

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

~~However,~~ 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, an adaptive sampling scheme is therefore used 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 following:

(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 $k = 1, 2, ..N_{\rm ini}$

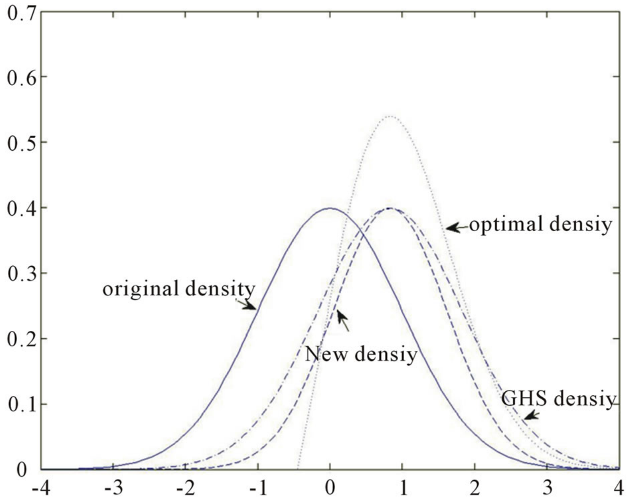

Figure 1. NEED TO CHANGE THIS FIG. Here we will display the intuitive idea of why sampling from an instrumental distribution can reduce the costs of integrating (i.e. improve it)

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 was 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 $N_{\rm s}$ Gaussian distributions around the $N_{\rm s}$ successful sample points in the initial parameter space. This idea is schematically shown in Figure (2). The instrumental distribution is then described by

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

where each Gaussian distribution $\mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ is equally weighted with $1/N_{\rm s}$ 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 $\mathbf{x_i} \in \{\mathbf{x_i}\}_{i=1}^{i=N_{\rm s}}$. This implies that the means $\mu_{\rm i}$ of the individual Gaussians in Equation (3) are given by

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

The covariance matrix Σ is chosen to scale with the average expected distance between two sample points $\{\mathbf{x_k}\}_{k=1}^{k=N_{\rm ini}}$ 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

$$\sigma_k = \frac{\|\max_k - \min_k\|}{(N_{\rm ini})^{1/d}} \quad \text{for } k = 1, .., \text{d}. \tag{6}$$

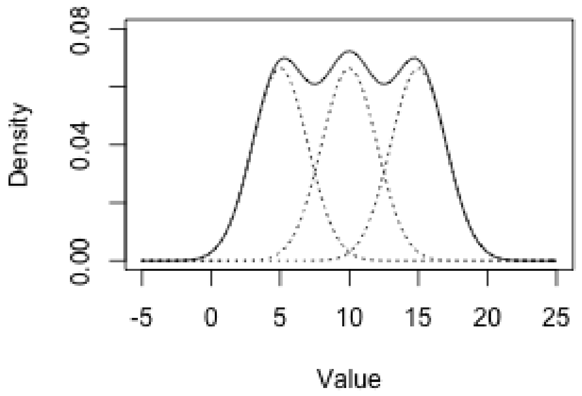

Figure 2. NEED TO CHANGE THIS FIG. Here we will show the intuitive idea of why we sample from a mixture of Gaussians around our hits of BBH mergers)

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) **(TO CHANGE)**: I should add that we reject samples when they are drawn outside of parameter space, and the fact that we don't have to normalize our distribution for this at all as in the importance sampling estimator the normalizations of the instrumental distribution cancels out.

(v) From these evaluations the expected value of $\phi(\mathbf{x})$ can be estimated using Equation (2). And the uncertainty is estimated by Equation() **TO DO**: [Add equations for uncertainties].

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

2.4 Example: fraction of BBH mergers

EXPLANATION OF WHY FRACTION BBH MERGER IS INTERESTING.

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	$p(M_1) \propto M_1^{-2.35}$	$[7, 100]M_\odot$
a	$p(a) \propto 1/a$	$[0.1, 10^3]AU$
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

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

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

2.5 Test

To test how well the adaptive importance sampling method works we run a large Monte Carlo simulation to estimate the fraction of the BBH mergers within the mentioned initial parameter space up to error X. 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 fraction and error to the plot. The results are plotted in Figure (3)

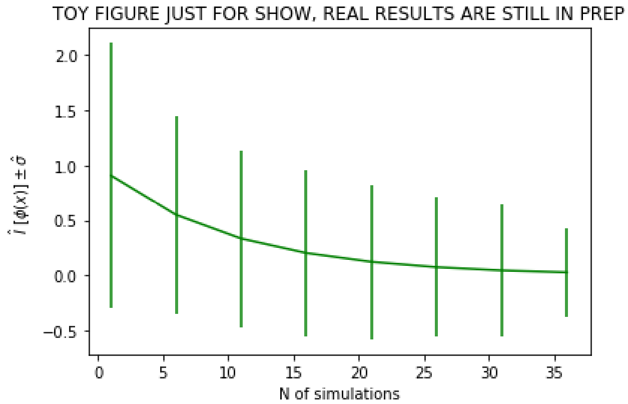

Figure 3. STILL NEED TO CHANGE. Our main result is a plot that shows the estimators with importance sampling and the estimated error with it as a function of the number of simulations run. We can then compare this with the "true fraction of gravitational waves" and the Monte Carlo sampling method.

3 RESULTS

From Fig (3) it can be seen that

- the method converges
- the estimated errors of the method decrease and are always consistent with the "true fraction"
- The Adaptive importance sampling method is a factor Y more efficient: the same error is obtained with Y times less sample points.

4 CONCLUSIONS

ACKNOWLEDGEMENTS

We thank the Kavli institute and Niels Bohr institute.

REFERENCES

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