

# Efficiently sampling rare events in population synthesis models

Floor Broekgaarden et al. <sup>1,2\*</sup> (TBD)

<sup>1</sup>*Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100, København ø, Denmark*

<sup>2</sup>*Astronomical Institute Anton Pannekoek, University of Amsterdam, P.O. Box 94249, 1090 GE, Amsterdam, The Netherlands*

21 January 2018

## 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 by an order of magnitude.

**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 et al. 2013; Schneider et al. 2014) to the end stages as core collapse supernova (Zapartas et al. 2017), and the more exotic outcomes such as type IA supernovae (Toonen et al. 2012), and gravitational wave sources (Stevenson et al. 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 et al. 2000, 2002), binary\_c (Izzard et al. 2004, 2009), StarTrack (Belczynski et al. 2008), SEBA (Portegies Zwart & Verbunt 1996; Verbunt et al. 1996) and COMPAS (Stevenson et al. 2017). These models interpolate between evolutionary tracks of single stars obtained with a detailed stellar structure code (Pols et al. 1998) 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 et al. 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.

BPS can be a expensive tool, especially when calculating rare events or when solving for specific parameter outcomes when the initial parameter space is complex and high dimensional.

## 2 METHOD

In this section we will introduce the main ingredients of BPS and explain our adaptive importance sampling algorithm. In Section 2.1 the general idea of BPS is introduced where after in Section 2.3 the BPS parameters and prior distributions are defined. In Section 2.2 the adaptive importance sampling method is described. This method is tested in Section 3.

### 2.1 Simulating populations of stars

In BPS, the evolution of a large population of binary stars is simulated by evolving a set of initial conditions through rapid binary evolution algorithms. The initial parameter space is then explored by placing samples of binaries  $\mathbf{x}_i$  in the initial parameter space and evaluating them in the BPS model. There are two methods commonly used to sample the parameter space: (i) drawing binaries  $\mathbf{x}_i$  randomly from their birth distributions  $p(\mathbf{x}_i)$  or by (ii) creating an (equally spaced) grid of binaries in the initial parameter space and weighing each sample point by its birth distribution. Fig (1) shows an example of how the samples can be distributed in the initial parameter space for both methods.

Each initial binary system can thus be characterized by a set of initial parameters  $\mathbf{x}_i$ . For the purpose of this paper, the initial conditions are characterized with the initial mass  $M_{1,i}$ , the separation  $a_i$  and the mass ratio  $q_i$  that are described in more detail in Section 2.3. However, in general many more parameters are used in BPS depending on the simulation. For the demonstration of the adaptive sampling method however, we will focus on these three most important parameters. Sampling from the birth distributions it follows that

$$\mathbf{x}_i \sim p(\mathbf{x}_i), \quad (1)$$

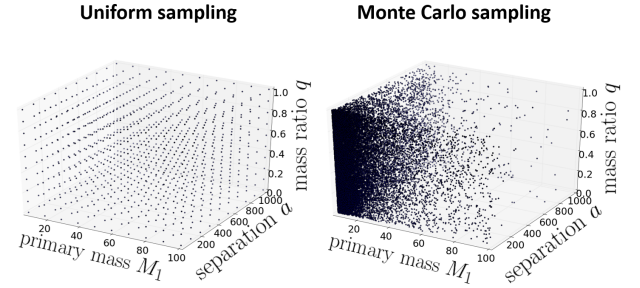
with  $\mathbf{x}_i = (M_{1,i}, a_i, q_i)$ . By evaluating every binary system  $\mathbf{x}_i$  then through the BPS model, a simulated population of binary stars is created, which can be used to predict observations and compare observations and theory.

BPS can be defined as a function  $u$  that outputs a final outcome,  $\mathbf{x}_f$ , of a binary system by evolving the initial conditions using rapid stellar evolution algorithms.

$$\mathbf{x}_f = u(\mathbf{x}_i)$$

In addition, since we are often interested in simulating specific target outcomes within the population of  $\mathbf{x}_f$ , for example only BBH mergers, we can introduce a function  $\phi(\mathbf{x}_f|\mathbf{x}_i, u)$  that equals unity if  $\mathbf{x}_f$  represents a binary from the population that was targeted  $\mathbf{x}_T$  and zero if it doesn't, i.e.

$$\phi(\mathbf{x}_f|\mathbf{x}_i, u) = \begin{cases} 1, & \text{if } \mathbf{x}_f \in \mathbf{x}_T \\ 0, & \text{else} \end{cases}$$



**Figure 1.** Example of samples that are drawn (left) uniformly distributed over the initial parameter space and (right) follow their birth distribution. The samples in the right plot are centred in the left corner since in the primary mass birth distribution used here low mass stars are much more common than high mass stars.

However, since it is not known a priori which initial conditions will result in the outcome of interest (e.g. a BBH merger) a large part of the initial parameter space needs to be explored. Especially when the event is rare, such as in the case of a BBH mergers, this exploration becomes very expensive and often is an intractable problem. Adaptive importance sampling is a sampling method that can reduce this cost.

### 2.2 Adaptive importance sampling

In the adaptive importance sampling (aIS) method, the distribution from which the initial binaries are sampled is adaptively changed such that the computational time is spent more on the part of the initial parameter space that also contributes most to the information. In other words, the adaptive importance sampling method reduces the sampling variance by sampling from a so-called *instrumental distribution* that is more similar to the target distribution than the prior distribution, this is shown in Fig. (2). Especially when simulating a process where a small part of the initial space produces the output of interest, changing the sampling distribution can significantly reduce the costs of the simulation.

The aIS method consists of three main steps which are shown in the scheme in Fig. (3).

- **Initial guess**

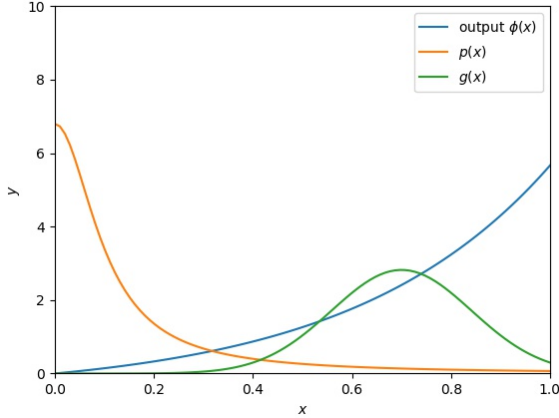
Since the initial parameters that will produce systems of a certain population are not known a priori, the initial parameter space has to be explored first. Therefore, only for the first run, a first guess of where to draw the binary samples  $\mathbf{x}_f$  in the initial parameter space has to be made. The most simplified choice would be to draw the first samples using the Monte Carlo sampling method.

- **Run simulation**

Then, initial conditions for the binary systems are drawn from the initial guess distribution and the samples are run through the BPS model until a certain threshold number of binaries of a specific target population have been found. (e.g. sample until 100 BBH mergers are found).

- **Improve sampling distribution**

Now that a first set of initial conditions  $\{\mathbf{x}_i\}_{i=1}^N$  is known to produce binary systems of the specific population of interest, i.e.,  $\phi(\mathbf{x}_i) = 1$  for all  $i = 1, \dots, N$ , we can improve



**Figure 2.** The intuitive idea of importance sampling is that often in the initial distribution function  $p(x)$  only very few of the drawn samples will contribute to the integral of the output function  $\phi(x)$  if  $p(x)$  and  $\phi(x)$  are dissimilar (especially in rare events). The idea of importance sampling is to sample instead from an instrumental distribution  $g(x)$  such that a higher fraction of the drawn samples contribute to the estimate of the integral of  $\phi(x)$ .

the sampling distribution. This is done by defining the new sampling distribution  $g(\mathbf{x}_i)$  as the mixture of  $N$  multivariate Gaussian distributions, where each Gaussian distribution  $g_i$  is centered around a successful outcome  $x_i$ . I.e.:

$$g(\mathbf{x}) = \sum_{i=1}^K \frac{1}{K} g_i(\mathbf{x}) = \frac{1}{K} \sum_{i=1}^K \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (2)$$

where each Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  is equally weighted with  $1/K$  in the mixture distribution. Since the Gaussians are distributed around the  $\mathbf{x}_f$  that belonged to the target population, the means  $\boldsymbol{\mu}_i$  are given by  $\boldsymbol{\mu}_i = \mathbf{x}_i$  for  $i = 1, \dots, K$ .

Furthermore, we scale the covariance matrix  $\boldsymbol{\Sigma}$  with the average expected distance between our initial sample points  $\mathbf{x}_i$  in the initial binary parameter space. This is chosen such that samples drawn from a Gaussian  $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  will generally fall in the space between the successful point  $x_i$  and its nearest neighbour. For simplicity we choose  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$  for all  $i$  and also adopt a diagonal covariance matrix for  $\boldsymbol{\Sigma}$  given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \dots \\ \vdots & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}, \quad (3)$$

where each  $\sigma_k$  is given by

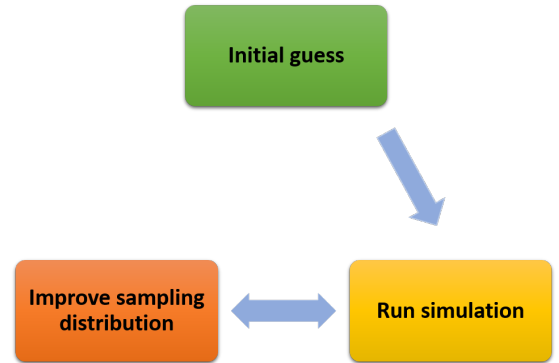
$$\sigma_k = \frac{\|x_{k,\max} - x_{k,\min}\|}{(N_{\text{ini}})^{1/d}} \quad \text{for } k = 1, \dots, d. \quad (4)$$

where  $N_{\text{ini}}$  is the total number of initial samples used and  $d$  is the dimension of the parameter space.

To account for the change of the sampling distribution, weights are introduced that correct for the new distribution when calculating the distributions and moments of  $\phi(\mathbf{x})$ . The weights are given by

$$w_k = p(\mathbf{x}_i)/g(\mathbf{x}_i). \quad (5)$$

### Adaptive importance sampling:



**Figure 3.**

And thus, for example, the expectation value of  $\phi(\mathbf{x})$  using adaptive importance sampling is given by

$$\hat{I}[\phi(\mathbf{x})] = \frac{1}{K} \sum_{i=1}^K \phi(\mathbf{x}_i) \frac{p(\mathbf{x}_i)}{g(\mathbf{x}_i)}. \quad (6)$$

### 2.3 Initial binary parameter space

Each binary system is characterized in these simulations by the three initial parameters  $\mathbf{x}_i = (M_{1,i}, a_i, q_i)$ . such that when assuming independent parameters the prior of  $\mathbf{x}_i$  is given by:

$$p(\mathbf{x}_i) = p(M_{1,i}) p(a_i) p(q_i). \quad (7)$$

The distributions and parameter spaces of each of the three parameters is introduced below.

The initial primary mass  $M_{1,i}$  is given by a power law  $p(M_1) = K_M M_1^{-\alpha}$ ,  $M_{1,i} \in [M_{1,\min}, M_{1,\max}]$  (8) where  $\alpha = 2.35$  is chosen to follow the Salpeter law and  $K_M$  is the normalization constant. We choose the primary mass to be in the range  $M_1 \in [7, 100] M_\odot$  for astrophysical arguments.

The separation  $a_i$  is chosen to be uniform in log and its prior is given by

$$p(a_i) = \frac{K_a}{a_i}, \quad a_i \in [a_{\min}, a_{\max}] \quad (9)$$

where we choose  $[a_{\min}, a_{\max}] = [0.1, 10^3]$  AU and  $K_a$  is the normalization constant.

The mass ratio  $q$  is chosen to be flat and is given by

$$p(q) = \frac{1}{q_{\max} - q_{\min}} \quad (10)$$

### 2.4 More details about the code?

## 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

**Table 1.** Summary of properties initial parameters

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

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  $f_{\text{BBH merger}}$ , our output function  $\phi(M_1, a, q)$  can then be given by

$$\phi(M_1, a, q) = \begin{cases} 1 & \text{if } u(M_1, a, q) \text{ produces a BBH merger} \\ 0 & \text{else} \end{cases} \quad (11)$$

The fraction of the initial parameter space that will produce BBH mergers when evaluated in the model (i.e.  $f_{\text{BBH merger}}$ ) can then be estimated by using Equation (6) by simulating  $N$  binary systems with the adaptive importance sampling method. In other words

$$f \approx \hat{I}[\phi(M_1, a, q)] = \frac{1}{N} \sum_{i=1}^N \phi(M_1, a, q) \frac{p(M_1, a, q)}{g(M_1, a, q)}. \quad (12)$$

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

$$p(M_1, a, q) \propto \frac{M_1^{-2.35}}{a} \quad (13)$$

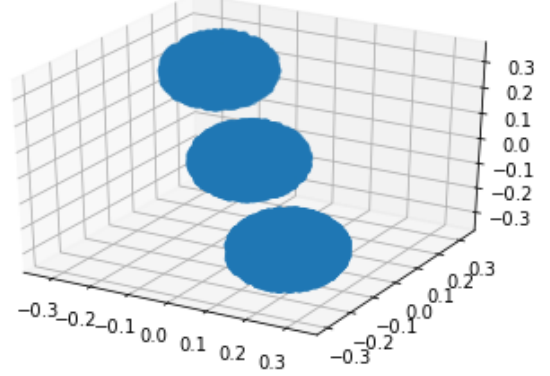
Following the algorithm described in Section (??) we define the instrumental distribution

$$g_1(M_1, a, q) = \sum_{i=1}^{N_s} \frac{1}{N_s} \mathcal{N}(\mu_i, \Sigma). \quad (14)$$

By filling in Equations (11), (13) and (14) into Equation (12) 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  $N_{\text{tot}}$  and estimate the fraction of BBH mergers with Equation (12) and compare this with the value for the fraction found by the large Monte

**The output of test function  $\phi(x)$ : three spheres**

**Figure 4.** 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.

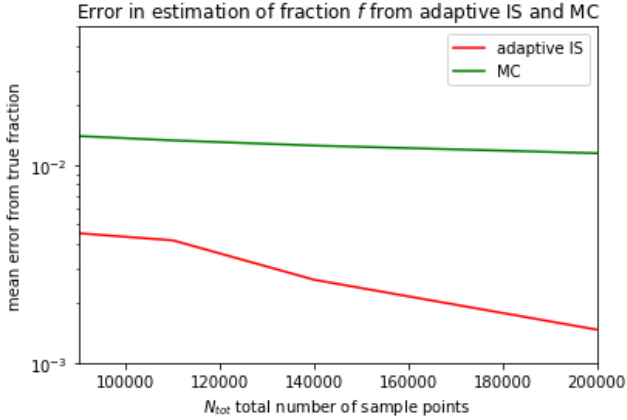
Carlo run. We also run the simulation with  $N_{\text{tot}}$  samples using the crude Monte Carlo method as given in Equation (??) and add the estimated error from the true fraction to the plot.

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  $\Omega_3$  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  $\Omega_3$  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 4.

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  $N_{\text{tot}}$  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 5. From this Figure it can be seen that:

- The error of the estimation for both methods is smaller for larger  $N_{\text{tot}}$ . 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.



**Figure 5.** Mean absolute error of the adaptive importance method (red) and the Monte Carlo method (green) as a function of the number of simulations run  $N_{\text{tot}}$  (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.

- The adaptive importance sampling method is a factor Y more efficient: the same error is obtained with Y times less sample points.

## 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.

## REFERENCES

- Abbott B. P., et al., 2016, Physical review letters, 116, 061102
- Belczynski K., Kalogera V., Rasio F. A., Taam R. E., Zezas A., Bulik T., Maccarone T. J., Ivanova N., 2008, The Astrophysical Journal Supplement Series, 174, 223
- De Mink S., Langer N., Izzard R., Sana H., de Koter A., 2013, The Astrophysical Journal, 764, 166
- Hurley J. R., Pols O. R., Tout C. A., 2000, Monthly Notices of the Royal Astronomical Society, 315, 543
- Hurley J. R., Tout C. A., Pols O. R., 2002, Monthly Notices of the Royal Astronomical Society, 329, 897
- Izzard R. G., Tout C. A., Karakas A. I., Pols O. R., 2004, Monthly Notices of the Royal Astronomical Society, 350, 407
- Izzard R. G., Glebbeek E., Stancliffe R. J., Pols O., 2009, Astronomy & Astrophysics, 508, 1359
- Pols O. R., Hurley J. R., Tout C. A., Eggleton P. P., et al., 1998, Monthly Notices of the Royal Astronomical Society, 298, 525
- Portegies Zwart S., Verbunt F., 1996, Astronomy and Astrophysics, 309, 179
- Schneider F., Langer N., de Koter A., Brott I., Izzard R. G., Lau H. H., 2014, Astronomy & Astrophysics, 570, A66
- Stevenson S., Vigna-Gómez A., Mandel I., Barrett J. W., Neijssel C. J., Perkins D., de Mink S. E., 2017, Nature Communications, 8

- Toonen S., Nelemans G., Zwart S. P., 2012, Astronomy & Astrophysics, 546, A70
- Verbunt F., Kuiper L., Belloni T., Johnston H. M., De Bruyn A., Hermesen W., van der Klis M., 1996, Astronomy and Astrophysics, 311, L9
- Zapartas E., et al., 2017, Astronomy & Astrophysics, 601, A29

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.