MNRAS **000**, ??–?? (2018) Preprint 29 January 2018 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*

29 January 2018

ABSTRACT

Key words: importance sampling – population synthesis – gravitational waves

1 INTRODUCTION

- introduce problem
- binary black holes BBH , LIGO

2 BINARY POPULATION SYNTHESIS

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. 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 (**??**), binary_c (**??**), StarTrack (**?**), SEBA (**??**) (**also add Portegies-zwart** & **Yungelson 1998 and Nelemans 2011??**), COMPAS (**?**) and ComBinE (Kruckow, submitted 2018). These models interpolate between evolutionary tracks of single stars obtained with a detailed stellar structure code (**?**) and rely on an approximate treatment of the physical processes. By doing so, they present a rapid code that can compute the evolution of many stars within a simulation. 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 (**??**) to the end stages as core collapse supernova (**?**), and the more exotic outcomes such as type IA supernovae (**?**), and gravitational wave sources (**?**).

Since it is not known a priori which part of the initial parameter space produces a certain binary population of interest, the parameter space needs to be explored. This is generally done by drawing many samples, i.e. binary systems $\mathbf{x_i}$, of initial binary parameters and evaluating them with the BPS model into their final state $\mathbf{x_f}$,

$$\mathbf{x_f} = u(\mathbf{x_i}), \tag{1}$$

where u here represents the BPS model evaluation.

Often, BPS models are used to study a population of binaries of a certain subtype $\mathbf{X_t}$, e.g. BBHs, and therefore it is useful to define a function,

$$\phi(\mathbf{x_f}|\mathbf{x_i}, u) = \begin{cases} 1 & \text{if } \mathbf{x_f} \in \mathbf{X_t} \\ 0 & \text{else,} \end{cases}$$

that equals unity if $\mathbf{x_i}$ evaluated to the target binary system $\mathbf{X_t}$ and zero if not. We will use this function throughout the paper as it is the main objective to perform the inference on.

The adaptive importance sampling method introduced in this paper can be applied to many different (BPS) codes. In the next section, however, we will introduce the BPS code COMPAS and our settings for the code that is used throughout this paper to demonstrate the strengths of this method.

2.1 Physics in COMPAS

Compact Object Mergers: Population Astrophysics and Statistics (COMPAS), is a BPS code that focuses on GW astrophysics and in particularly is designed to study uncertainties in binary evolution and compact mergers that possibly form the GW progenitors and optimize the information that can be obtained from simulations (**?**). In this work the code is used to analyse binary black holes that merge within Hubble time and therefore form potential progenitors for the observed GWs by the LIGO and VIRGO collaboration. The COMPAS model includes the so-called 'isolated binary evolution channel' for BBH mergers, where both stars initially start with a large separation and thus evolve separately and often a *'commen envelope'* phase is needed to bring the stars closer together such that they can merge within Hubble time. Details of this channel are given in (**?**).

⋆ E-mail: fsbroekgaarden@gmail.com

Like most BPS codes, COMPAS includes many physical processes such as common envelope, mass transfer and kick velocity prescriptions. Throughout this paper we employ the (**?**) fiducial model. Some exceptions on the fiducial model are described below.

- Metallicity, The metallicity of all simulations is set to $Z = 0.00XZ_{\odot}$, which represent the low metallicity environment where massive black holes can be born.
- We focus on BBH mergers that merge within Hubble time. Our target distribution $\mathbf{X_t}$ therefore consists of these binaries that evolve to two black holes. Furthermore we only select the binaries that did not have roche-lobe overflow after the commen envelople phase **(say something about optimistic / pessimistic??)**.
- (add other differences with Stevenson+17 and COMPAS settings I have)

2.2 Initial parameter priors

Since the details of binary evolution are uncertain, a large number of binaries is simulated with BPS models to thoroughly explore the space of initial conditions. The uncertainties are then captured in tunable parameters $\mathbf{x}$ which distribution functions are observably derived. Often between 3 and 20 parameters are used in BPS but for the purpose of this paper we will focus on the three most important variables: the initial mass of the primary star (i.e., the most massive star) $M_{1,i}$, the mass ratio, $q_i = M_{1,i}/M_{2,i}$, between the two stars and the initial separation, a_i. Each initial binary sample can thus be represented by three parameter values

$$\mathbf{x_i} = (M_{1,i},\ a_i,\ q_i). \tag{2}$$

The reason that we focus on these three parameters is (i) because above parameters are the most important parameters in the sense that they dominate the outcome and thereby the uncertainty of the evolution (ref Belczynski & Selma or Moe?) , and (ii) that especially the primary mass and separation have an extremely steep distribution function which can drastically increase the computational costs when simulating rare events (see Section **??**). The distribution functions used in this paper for the three parameters are given below. Although only a subsection of three parameters are used in this work, the adaptive importance sampling method can straightforwardly be extended to all initial parameters.

2.2.1 Primary mass $M_{1,i}$

The distribution of the initial primary mass M_1 follows a power law distribution also known as the initial mass function (IMF) (**?**):

$$p(M_{1,i}) = K_M M_{1,i}^{-\alpha},\ M_{1,i} \in [M_{1,i,\min}, M_{1,i,\max}], \tag{3}$$

where K_M is the normalization constant given by

$$K_M - \frac{\alpha + 1}{M_{1,\max}^{\alpha+1} - M_{1,\min}^{\alpha+1}}.$$

We choose $\alpha = 2.35$ and the primary mass to be in the range $M_{1,i} \in [7, 100]M_{\odot}$, consistent with (**?**).

2.2.2 Mass ratio q_i

The mass ratio q_i is suggested from observations to have a flat distribution (Mazeh et al., 1992; Goldberg & Mazeh, 1994; Tout, 1991), given by

$$p(q_i) = \frac{1}{q_{\max} - q_{\min}},\ q_i \in (q_{\min}, q_{\max}], \tag{4}$$

where $(q_{\min}, q_{\max}] = (0, 1]$ by definition of q. Nevertheless, it is also suggested that there is some dependency of the mass ratio on the period of the system (e.g. Moe & Di Stefano 2016).

2.2.3 Separation a_i

The separation a_i is found to be uniform in log, also known as Opiks law (**?**) and consistent with findings by (**?**) and (**?**). The distribution function is given by

$$p(a_i) = \frac{K_a}{a_i},\ a_i \in [a_{\min}, a_{\max}], \tag{5}$$

where K_a is the normalization constant.

$$K_a = \frac{1}{\log a_{\max} - \log a_{\min}}$$

We choose $[a_{\min}, a_{\max}] = [0.1, 10^3]$ AU consistent with (**?**).

Assuming the parameters $M_{1,i}, q_i, a_i$ to be independent the combined prior distribution of $\mathbf{x_i}$ then becomes:

$$p(\mathbf{x_i}) = p(M_{1,i})\ p(a_i)\ p(q_i). \tag{6}$$

These distributions can now be used to model a population of stars using sampling methods.

3 SAMPLING METHODS

In general BPS, a large number of binaries $\mathbf{x_i}$ are drawn from their birth distribution (Eq. **??**) either (i) randomly, in which case the statistics of the targeted population are estimated using the Monte Carlo estimator or (ii) by creating a uniform grid in parameter space and weighing each sample grid point with the birth distribution. Examples of BPS models that use random draws via Monte Carlo sampling are StarTrack, COMPAS and SEBA, whilst binary_c and ComBinE use grid based sampling **(DOUBLE CHECK THIS)**. In both sampling strategies, simulating rare events (e.g. BBH mergers) becomes computationally expensive and often an intractable problem as a large part of the simulations are spend on evaluating initial binary systems $\mathbf{x_i}$ that don't evolve to a binary system of the target population. In this section we therefore introduce an adaptive importance sampling method that reduces the costs of simulating rare events with BPS by introducing a sampling scheme that aims to adaptively change the sampling distribution towards the target distribution.

3.1 Adaptive importance sampling

The adaptive importance sampling (AIS) method consists of three main steps which are schematically shown in Fig. **??**.

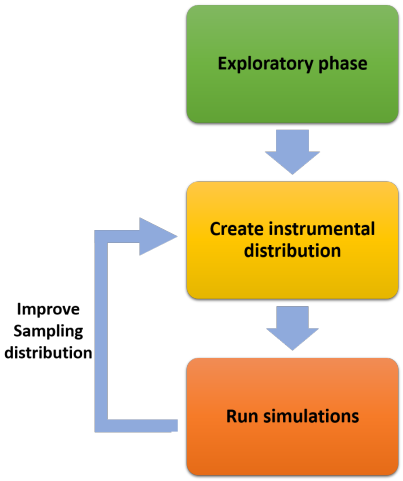

Figure 1. Flowchart of the adaptive importance sampling method that is used in this paper. **should I add an input arrow with input priors and parameters and an output arrow for 'inference'??**

(i) First, there is a so-called *exploratory phase* since it is not known a priori which part of the parameter space produces binaries of the type that is contained in the target distribution. In this phase initial samples are drawn from their birth distribution either randomly or by creating a regular grid (uniform method) and evaluated in the model until a certain number $N_{\rm E,t}$ of binaries of the target type $\mathbf{X_t}$ are found. I.e. the sampling continues until there is a set of $N_{\rm E,t}$ binaries of the target simulation $\mathbf{x_f} \in \mathbf{X_t}$. The choice of $N_{\rm E,t}$ depends mostly on the aimed sampling uncertainty of the simulation and is explained further in Section **??**.

(ii) Secondly, a new sampling distribution, $g(\mathbf{x})$, the so-called *instrumental distribution* is defined by drawing Gaussian distributions in the initial parameter space around each of the $N_{\rm E,t}$ 'successful' binaries $x_{E,k}$ with $k = 1, ..., N_{\rm E,t}$ that evaluated to a binary system of the target population. The instrumental distribution $g(\mathbf{x})$ is thus given by the Gaussian mixture distribution:

$$g(\mathbf{x}) = \sum_{k=1}^{N_{\rm E,t}} g_k(\mathbf{x}) w_k = \sum_{k=1}^{N_{\rm E,t}} \mathcal{N}(\mathbf{x_{t,k}}, \mathbf{\Sigma_k}) w_k \tag{7}$$

where each $\mathcal{N}(\mathbf{x_{t,k}}, \mathbf{\Sigma_k})$ is a multivariate Gaussian distribution with mean $\mathbf{x_{E,k}}$ and covariance matrix $\mathbf{\Sigma_k}$. Furthermore, w_k are the weights of each Gaussian distribution, which equals the weight of drawing the initial sample x_i ($w_k = 1$ in the Monte Carlo random drawing).

(iii) Thirdly, new samples are now drawn from the instrumental distribution $g(\mathbf{x})$ and evaluated with the BPS model. Assuming the outcome function $\phi(\mathbf{x_i})$ to be locally continuous (and not chaotic), a larger fraction of the samples will now produce binaries of the target distribution, and as a result, the simulation will model the target distribution more efficient.

After a certain number of draws, step (ii) and step (iii) can be repeated to even more increase the efficiency.

At the end of a run, the statistics of the target distribution such as the expectation value of $\phi(x)$, i.e. its rate, and distribution functions can be determined by using the AIS estimator

$$\widetilde{\mathbb{E}[\phi(\mathbf{x})]}_{N_t} = \frac{1}{N_t} \sum_{i=1}^{N_t} \phi(\mathbf{x_i}) \frac{p(\mathbf{x_i})}{g(\mathbf{x_i})}, \tag{8}$$

where N_t is the total number of sample drawn from the instrumental distribution. **change this to multiple generations see notes other notebook**

In general, the AIS method is based on easy to implement steps. However, many choices on for example the covariance matrices, the exploratory sampling and other sampling strategies can be made to optimize the AIS method. Below we will therefore first introduce our *'standard'* AIS method which is a minimal working implementation of the AIS method that is used throughout the paper and demonstrated in section **??** to already reduce the simulation costs. Of course many more adjustments can be made, some of which are discussed in Subsection **??**.

3.1.1 Standard implementation

In this subsection we will describe our so-called *standard implementation* of AIS, which is used mostly throughout this paper. Although variations on this standard implementation might make the AIS method more efficient, we choose to first demonstrate the results with this more basic implementation of AIS to make it easier to understand the method and its strengths. Moreover, the choice of variations often varies for each simulation. A few examples of possible variations are discussed in the next section.

(i) We choose $N_{\rm E,t} = 100$ and an exploratory sampling using random draws from the birth distribution (Eq. **??**), i.e. Monte Carlo sampling. We make this choice since Monte Carlo sampling produces samples with weights equal to unity, which is the simplest sampling method for the exploratory phase of AIS. The choice of $N_{\rm E,t} = 100$ implies that we will continue sampling random draws from the birth distribution until we have simulated 100 initial binaries that evolved successfully to a target binary ($\mathbf{x_f} \in \mathbf{X_t}$). Since all samples have equal weight this implies that if we missed in our exploratory phase an area of the parameter space that also produces binaries of our target binary type, that this 'missed island' most likely contributes less than 1/100 to the total integral and thus statistics (such as the BBH fraction and/or the BBH merger rate).

(ii) The instrumental distribution. Since we use random draws the weights w_k in Eq. (**??**) are equal to $1/N_{\rm E,t}$ and each Gaussian distribution contributes equally to the instrumental distribution. For simplicity, we adopt a diagonal covariance matrix for Σ. Moreover, we scale the covariance matrix Σ_k with the average expected distance locally between our initial sample points $\mathbf{x_i}$ in the initial binary parameter space. This is chosen such that samples drawn from a Gaussian $\mathcal{N}(\boldsymbol{\mu_i}, \mathbf{\Sigma_i})$ will generally fall in the space between the successful point x_i and its nearest neighbours. Since the exploratory phase does not place the samples uniformly throughout the initial parameter space, e.g., many more binaries are created with low initial primary mass $M_{1,i}$ since low mass stars are much more common (see Eq. **??**).

Therefore, we first transform the samples to a uniform sampled space, and then draw Gaussians around the space where it is uniform. This implies that the covariance matrix for the Gaussian distribution is given by

$$\Sigma_k = \begin{bmatrix} \sigma_1^2 & 0 & \dots \\ \vdots & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}, \tag{9}$$

where each $\sigma_{j,k}$ is given by

$$\sigma_{j,k} = \frac{\|p_j(x_{k,\max})^{-1} - p_j(x_{k,\min})^{-1}\|}{(N_{\rm E})^{1/d}} \text{ for } j = 1, .., \mathrm{d}, k = 1, ..., N_{\rm E,t} \tag{10}$$

where p_j^{-1} is the inverse of the probability distribution function of the j-th parameter (see Subsection **??**) and N_E is the total number of simulations run for during the exploratory phase.

(iii) *Stopping criteria.* For the stopping criteria of the simulation we use for the standard AIS implementation to stop when a total number of $N_{\rm tot} = XX$ samples have been evaluated. This represents the finite computational time that is often available for simulations due to scarce computational resources. Another possible stopping criteria could be defined by reaching a certain accuracy, this is further discussed in Subsection **??**.

3.1.2 Variations on the standard implementation

- multiple generations of g
- other ways of running exploratory run
- fixed error
- stratified sampling

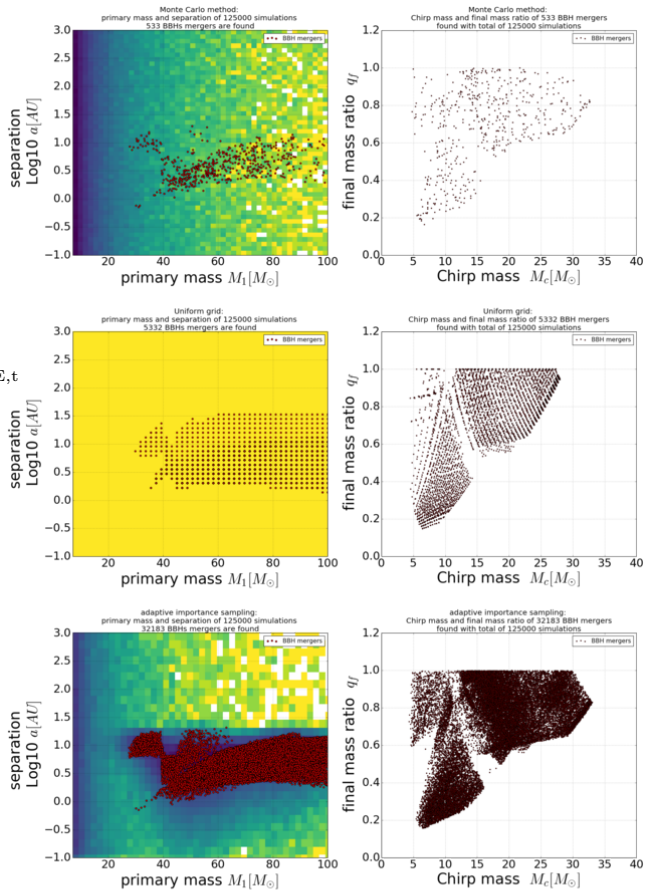

Figure 2. Example of samples that are drawn (left) uniformly distributed over the initial parameter space and (right) follow their birth distribution. The samples in the right plot are centred in the left corner since in the primary mass birth distribution used here low mass stars are much more common that high mass stars.

4 APPLICATION I

4.1 More BBHs in simulation

•

4.2 Rate estimator

•

4.3 Chirp mass distribution

•

5 APPLICATION II

•

6 (...)

7 DISCUSSION

•

8 SUMMARY

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.