Bayesian Rate Estimation

Will M. Farr[*]
Northwestern University and CIERA

Ilya Mandel[†]
University of Birmingham

Jon Gair

We show how to obtain a Bayesian estimate of the rate of signal events from a set of signal and background events indexed by a ranking statistic when the shapes of the signal and background distributions are known, can be estimated, or approximated. We focus on the specific application of estimating astrophysical rates of the coalescence of compact binary black holes or neutron stars from a set of triggers in the LIGO/Virgo gravitational wave detectors, but our framework is fully general. We discuss the systematic effects on the rate estimate due to differences between the assumed and true shapes of the foreground or background distributions, identifying ways these effects can be minimized. Similarly, we discuss the effects of various priors on the rate, including uninformative priors, weakly-informative priors, and the use of priors from previous rate experiments. In the limit where the expected signal rate gives high probability of zero or one signal in the data, our technique reduces to the "loudest event statistic," but it is generally applicable to arbitrarily large signal rates.

I. INTRODUCTION

FIXME: Introduce the necessity of estimating rates, prior work (like [1]), Bayesian inference.

II. MODEL

We assume that we are presented with a data set of N events. Each event may be due to either a signal of interest or an uninteresting background. Each event is associated with a ranking statistic, x. Our data set therefore consists of the ranking statistics for the set of events:

$$d = \{x_i | i = 1, \ldots, N\}. \tag{1}$$

We assume that the events are sorted by ranking statistic, so that $i < j$ implies that $x_i < x_j$.

We assume that both the foreground and background events are samples from an inhomogeneous Poisson process with rates (per unit ranking statistic, x)

$$\frac{dN_f}{dx} = f(x) \tag{2}$$

and

$$\frac{dN_b}{dx} = b(x). \tag{3}$$

The cumulative rates of the two processes are therefore

$$F(x) \equiv \int_{-\infty}^{x} ds\, f(s) \tag{4}$$

and

$$B(x) \equiv \int_{-\infty}^{x} ds\, b(s). \tag{5}$$

The assumption that the foreground and background events form an inhomogeneous Poisson process implies

1. The number of events in any range of ranking statistics, $x \in [x_1, x_2]$ is Poisson distributed with rate $F(x_2) - F(x_1)$ or $B(x_2) - B(x_1)$.

2. The numbers of events in non-overlapping ranges of ranking statistics are independent.

3. The probability of exactly one foreground event between x and $x + h$ is given by

$$P(N = 1 \in [x, x+h]) = f(x) h + \mathcal{O}\left(h^2\right). \tag{6}$$

   and similarly for background events.

4. The probability of two or more events in a small range of ranking statistic is negligable

$$P(N = 2 \in [x, x+h]) = \mathcal{O}\left(h^2\right). \tag{7}$$

The foreground and background rates can in general depend on several parameters; the goal of our analysis is to determine the posterior probability distributions for these parameters that are implied by the data. At the least, we will want to know the overall amplitude of the foreground and background rates. Let

$$f(x) = R_f \hat{f}(x), \tag{8}$$

and

$$b(x) = R_b \hat{b}(x), \tag{9}$$

[*] w-farr@northwestern.edu
[†] ilya@chgk.info

where $\hat{F}(\infty) = \hat{B}(\infty) = 1$. Then R_f and R_b are the total number of foreground and background events expected, respectively. Other parameters may describe the shape of the rate functions, but these will depend on the details of the system being analyzed. In §III we give an example of fitting such shape paremeters.

We do not know a priori which of the events are foreground and which are background. For each event, we introduce a flag, s_i, which is either 0 or 1, indicating a background or foreground event, respectively. These "state" flags are additional parameters in our model. We can marginalize over our uncertainty in the state of any given event by summing all posteriors over $s_i = \{0, 1\}$. Given an identification of each event as foreground or background, and the rates of each, the likelihood of our data is

$$p\left(d\right|\left\{f(x), b(x), \{s_i\}\right\}) = \left[\prod_{\{i|s_i=1\}} f\left(x_i\right)\right]\left[\prod_{\{j|s_j=0\}} b\left(x_j\right)\right] \times \exp\left[-F\left(\infty\right)\right]\exp\left[-B\left(\infty\right)\right]. \quad (10)$$

Written in terms of the rate parameters, this becomes

$$p\left(d\right|\left\{f(x), b(x), \{s_i\}\right\}) = R_f^{N_f}\left[\prod_{\{i|s_i=1\}} \hat{f}\left(x_i\right)\right]\exp\left[-R_f\right] \times R_b^{N_b}\left[\prod_{\{j|s_j=0\}} \hat{b}\left(x_j\right)\right]\exp\left[-R_b\right], \quad (11)$$

where N_f is the number of the s_i that are 1 (i.e. the number of assumed foreground events), and N_b is the number of the s_i that are 0 (i.e. the assumed number of background events). Note that $N_f + N_b = N$, as each event is considered either foreground or background in our model.

A. Priors

FIXME: Discussion of the various priors one might want to impose on the rates, R_f and R_b, such as a flat prior on rate, the Poisson Jeffrey's prior (which is $\propto 1/\sqrt{R}$?), .

A choice of the foreground and background rates implies a prior on the state flags. For fixed foreground and background rates, the probability that any particular event is foreground is given by

$$p\left(s_i = 1 | R_f, R_b\right) = \frac{R_f}{R_f + R_b}. \quad (12)$$

The complete prior is then

$$p\left(R_f, R_b, \{s_i\}\right) = p\left(\{s_i\}\left|R_f, R_b\right.\right) p\left(R_f, R_b\right) = \frac{R_f^{N_f} R_b^{N_b}}{\left(R_f + R_b\right)^N} p\left(R_f, R_b\right). \quad (13)$$

III. GRAVITATIONAL WAVES FROM COMPACT BINARY INSPIRALS

ACKNOWLEDGMENTS

Richard O'Shaughnessy for discussions.

Appendix A: Likelihood for Inhomogeneous Poisson Processes

Consider a set of samples, $\{x_i\}$, from an

[1] R. Biswas, P. R. Brady, J. D. E. Creighton, and S. Fairhurst, Classical and Quantum Gravity **26**, 175009 (2009), arXiv:0710.0465 [gr-qc].