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 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, f_i, which is either 0 or 1, indicating a background or foreground event, respectively. These "state" flags are additional (unobserved) data in our model. We can marginalize over our uncertainty in the state of any given event by summing all posteriors over $f_i = \{0, 1\}$. Given an identification of each event as foreground or background, and the rates of each, the likelihood[2] of our data is

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

Written in terms of the rate parameters, this becomes

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

where N_f is the number of the f_i that are 1 (i.e. the number of assumed foreground events), and N_b is the number of the f_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.

It will occasionally be convenient to work with the ratio, $\hat{f}(x)/\hat{b}(x)$ instead of the values of $\hat{f}$ and $\hat{b}$ separately. This ratio can be interpreted as a likelihood ratio between the foreground and background models for an event that has occurred with ranking statistic x:

$$\frac{\hat{f}(x)}{\hat{b}(x)} = \frac{p(x|\text{foreground})}{p(x|\text{background})}. \quad (12)$$

In some circumstances, it is possible to compute this ratio empirically (for example, in Cannon *et al.* [3], the authors give a practical method for estimating the distribution of loudnesses for known-background events and known-foreground events in a mock data stream from the Advanced LIGO and Advanced Virgo gravitational wave detectors). Dividing Eq. (11) by $\prod_{\{i|f_i=0\}} \hat{b}(x_i)$, we obtain

$$p\left(d, \{f_i\} \left| f(x), b(x)\right.\right) \propto \left[\prod_{\{i | f_i = 1\}} \frac{\hat{f}(x_i)}{\hat{b}(x_i)}\right] R_f^{N_f} R_b^{N_b} \exp\left[-\left(R_f + R_b\right)\right]. \quad (13)$$

A. Priors

Because the overall rates enter the likelihood in Eq. (11) in the same form as in the constant-rate Poisson likelihood, a reasonable prior for R_f and R_b is the Poisson Jeffrey's prior[4],

$$p(R) \propto \frac{1}{\sqrt{R}}. \quad (14)$$

One advantage of the Jeffrey's prior in this circumstance is that it has finite probability mass as $R \to 0$, and the likelihood in Eq. (11) controls its behavior as $R \to \infty$ to ensure a normalizable posterior.

Of course, other priors could be used; in particular, previous experiments may have placed constraints on the rate that can be used as priors in subsequent experiments. Throughout the remainder of this paper, however, we use the Jeffrey's prior, Eq. (14).

B. Posteriors

Putting everything together, the posterior probability of each event's state and the overall rate is given by

$$p(R_f, R_b, \{f_i\} | d) \propto R_f^{N_f} \left[\prod_{\{i | f_i = 1\}} \hat{f}\left(x_i\right)\right] \exp\left[-R_f\right] \times R_b^{N_b} \left[\prod_{\{j | f_j = 0\}} \hat{b}\left(x_j\right)\right] \exp\left[-R_b\right] p(R_f, R_b). \quad (15)$$

If we marginalize over all the flags, f_i, then we obtain

$$p(R_f, R_b | d) \propto \prod_i \left[R_f \hat{f}\left(x_i\right) + R_b \hat{b}\left(x_i\right)\right] \times \exp\left[-\left(R_f + R_b\right)\right] p\left(R_f, R_b\right). \quad (16)$$

If, instead, we marginalize over all but one flag, f_k, we obtain that the posterior ratio of signal to noise probability for event k given the rates R_f and R_b is

$$\frac{p(R_f, R_b, f_k = 1 | d)}{p(R_f, R_b, f_k = 0 | d)} = \frac{R_f \hat{f}\left(x_k\right)}{R_b \hat{b}\left(x_k\right)} p\left(R_f, R_b\right). \quad (17)$$

Note that the *posterior* estimate of the ratio of signal to noise probabilities for event i differs from the likelihood ratio, Eq. (12), because it depends on the overall rates of foreground and background events.

III. GRAVITATIONAL WAVES FROM COMPACT BINARY INSPIRALS

A. Analytic Example

Suppose we attempt to detect gravitational wave signals in a data stream by matched filtering against a set of N templates. The data stream consists of stationary Gaussian noise with a power spectral density $S(f)$ combined additively with some number of signals. The signal to noise ratio (SNR) of a template, $h(f)$, given data, $d(f)$, is given by

$$\rho_h \equiv \frac{\langle h, d\rangle}{\sqrt{\langle h, h\rangle}}, \tag{18}$$

where $\langle\cdot\rangle$ denotes the noise-weighted inner product:

$$\langle a, b\rangle \equiv \int_0^\infty df\, \frac{a^*(f)b(f)}{S(f)}. \tag{19}$$

We suppose for simplicity that the templates are sufficiently distinct that

$$\langle h_1, h_2\rangle \sim 0. \tag{20}$$

For a data stream of pure noise, $d(f) = n(f)$, the SNR of a each template follows a $N(0,1)$ distribution. If our ranking statistic for events in each interval of data is their maximum SNR over the N templates,

$$x \equiv \max_h \rho_h, \tag{21}$$

then the background distribution is

$$\hat{B}(x) = \left(\frac{1+\mathrm{erf}\left(\frac{x}{2}\right)}{2}\right)^N \tag{22}$$

If, in addition, we impose a threshold ranking statistic, x_{min}, to be considered an "event," then $\hat{B}$ becomes

$$\hat{B}(x) = \frac{\left(1+\mathrm{erf}\left(\frac{x}{2}\right)\right)^N - \left(1+\mathrm{erf}\left(\frac{x_{\mathrm{min}}}{2}\right)\right)^N}{2^N - \left(1+\mathrm{erf}\left(\frac{x_{\mathrm{min}}}{2}\right)\right)^N} \tag{23}$$

The SNR of a gravitational wave signal in an interferometric detector scales as $1/d$, where d is the distance to the source. Ignoring possible cosmological effects, the number of sources will scale as d^3. Thus, we expect that the foreground cumulative distribution of events will follow

$$\hat{F}(x) = 1 - \frac{x_{\mathrm{min}}^3}{x^3}. \tag{24}$$

To demonstrate the effectiveness of our formalism, we applied it to a synthetic data set with foreground and background distributions drawn from Eqs. (23) and (24) with $R_f^{\mathrm{true}} = 10.4$ and $R_b^{\mathrm{true}} = 95.1$. The cumulative distribution for the ranking statistic of the synthetic data we used appears in Figure 1.

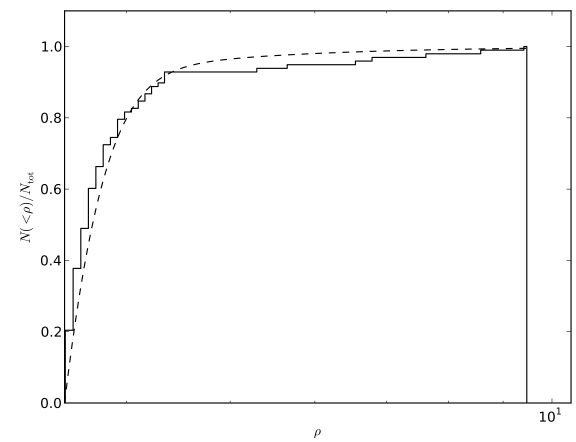

FIG. 1. The cumulative distribution of the ranking statistics for the synthetic data used to test the the formalism on the model from §III A. The solid line gives the cumulative distribution of the synthetic data; the dashed line gives the theoretical cumulative distribution for the models in Eqs. (23) and (24) combined with $R_f = 10.4$ and $R_b = 95.1$.

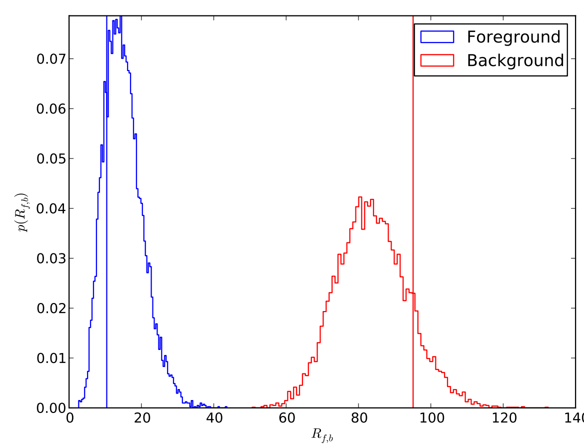

FIG. 2. The posterior densities for R_f (in blue) and R_b (in black) for the analytic model discussed in §III A. The vertical lines indicate the "true" used to generate the synthetic data set.

In Figure 2, we show the marginalized posterior densities for the foreground and background rates recovered from a synthetic data set drawn from the distributions in Eqs. (23) and (24) with . (Refer to Eq. (16).) Figure 3 shows the posterior foreground probability for each event marginalized over all other events' types and the foreground and background rates. (Refer to Eq. (17).)

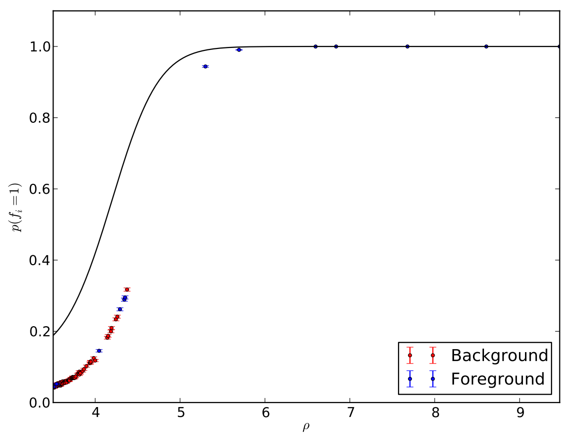

FIG. 3. Foreground probability for each event in the synthetic data set of §III A marginalized over all other parameters. True foreground events are in blue, background events in red. The solid line is the likelihood ratio $p(x|\text{foreground})/p(x|\text{background})$ (see Eq. (12)) for this model; this exceeds the marginalized foreground probability for many of the events because in this data set there are approximately nine times as many background events as foreground events (see Eq. (17)).

ACKNOWLEDGMENTS

Kipp Cannon, Chad Hanna, and Drew Keppel for providing the data set from Cannon *et al.* [3] and assisting us in its use. Richard O'Shaughnessy for discussions.

Appendix A: Likelihood for Inhomogeneous Poisson Processes

FIXME: To do.

[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].
[2] See §A for a derivation of Eq. (10).
[3] K. Cannon, C. Hanna, and D. Keppel, *A method to estimate the significance of coincident gravitational-wave observations from compact binary coalescence*, Tech. Rep. P1200031 (LVC, https://dcc.ligo.org/DocDB/0088/P1200031/003/far.pdf, 2012).
[4] Recall that the Jeffrey's prior for a parameter θ with likeilhood function, $p(d|\theta)$, is

$$p(\theta) \propto \sqrt{\left\langle \left(\frac{\partial p(d|\theta)}{\partial \theta} \right)^2 \right\rangle}, \tag{A1}$$

where the expectation is taken over all data, d. The Jeffrey's prior assigns prior weight to locations in parameter space proportional to the Fisher information that is expected to result from a measurement at that parameter value.