Bayesian Rate Estimation

Will M. Farr[*]
Center for Interdisciplinary Exploration and Research in Astrophysics
Department of Physics and Astronomy
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208

Jonathan R. Gair[†]
Institute of Astronomy
University of Cambridge
Madingley Road, Cambridge CB3 0HA
United Kingdom

Ilya Mandel[‡]
School of Physics and Astronomy
University of Birmingham
Edgbaston B15 2TT Birmingham
United Kingdom

Curt Cutler[§]
Jet Propulsion Lab, 4800 Oak Grove Dr., Pasadena, CA 91109 and
Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125

We show how to obtain a Bayesian estimate of the rate of signal events from a set of signal and background events when the shapes of the signal and background distributions are known, can be estimated, or approximated; our method works well even if the foreground and background event distributions overlap significantly. We give examples of determining the rates of gravitatonal-wave events in the presence of background triggers from a template bank when noise parameters are known and/or can be fit from the trigger data. We also give an example of determining globular-cluster shape and location parameters from an observation of a stellar field that contains a non-uniform background density of stars superimposed on the cluster stars.

[*] w-farr@northwestern.edu; http://faculty.wcas.northwestern.edu/will-farr/
[†] jrg23@cam.ac.uk
[‡] imandel@star.sr.bham.ac.uk; http://www.sr.bham.ac.uk/~imandel
[§] Curt.J.Cutler@jpl.nasa.gov

I. INTRODUCTION

The task of estimating a rate of foreground events when a mixture of foreground and background events is present in the data is a common one in physical and astrophysical applications. This problem comes up, among others, in gravitational-wave data analysis [e.g., **? ?**] and in astronomical observations of a field of objects of mixed provenance [, GC cluster paper].. In this paper, we introduce a robust formalism for estimating event rates from the data when the shape of foreground and background distributions are known (but may have additional free parameters), but the provenance of individual events as either background or foreground is unknown.

We use a Bayesian approach and consider all available data to ensure that the inferred rates are both unbiased and maximally constrained in the presence of limited observations. Bayes' theorem yields the *posterior* probability density function on a set of parameters, $\vec{\theta}$, given the observed data, d, under a model M:

$$p(\vec{\theta}|d, M) = \frac{p(\vec{\theta}|M)p(d|\vec{\theta}, M)}{p(d|M)}, \tag{1}$$

where $p(\vec{\theta}|M)$ are the *prior* probabilities of the model parameters, $p(d|\vec{\theta}, M)$ is the *likelihood* of obtaining the data given a particular choice of parameters, and the normalizing factor $p(d|M)$ is known as the *evidence*.

Two alternative approaches have been suggested in the past. One, known as the *loudest-event statistic* [**?**], uses only the information from the highest-ranked event in the data to infer the rate distribution. This approach has been used successfully [**? ?**] when the number of loud foreground events is small (typically zero or one) to obtain upper limits on foreground rates. However, the loudest-event statistic ignores all events except the loudest one, and so suffers from an unnecessary loss of information; therefore, we expect it to yield a much larger variance than strictly necessary when multiple events are present in the data.

Another possible approach is based on the use of only loud, "gold-plated" events, ones which are certain (or nearly certain) to come from the foreground, to derive rates. We refer to this approach as the *foreground-dominated statistic*. The foreground-dominated statistic may yield accurate results when the foreground and background are cleanly separated, at least for the loudest events, and the number of such loud events is sufficiently large. However, it cannot properly account for marginal events. While either the loudest-event statistic or the foreground-dominated statistic can approach the accuracy of our proposed method in specific regimes, both are suboptimal in a general case.

In order to demonstrate our method, we consider three different examples. The first two come from the field of gravitational-wave data analysis, but could equally arise in any application that employs matched filtering [**?**] to extract weak signals with known shapes from the data. The last example considers the case of a globular cluster on a background of field stars. Throughout, we compare the results obtained with our technique to the loudest-event and foreground-dominated statistics, which make use of a limited subset of the available information.

This paper is organized as follows. In section II we introduce our method for Bayesian rate estimation, and describe loudest-event and foreground-dominated statistics for comparison. Section III includes three examples of the application of these techniques. We conclude with a discussion of important remaining issues in section IV.

II. MODEL

We assume that we are presented with a data set of N events that exceed a pre-specified threshold in ranking statistic, x_{min}. 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:

$$d = \{x_i | i = 1, \dots, N\}. \tag{2}$$

The number of events N is also part of the observed data, but we separate out N and the observed ranking statistics, d, for convenience. We can choose how to label our events. Ultimately we will label the events in order of ranking statistic, i.e., $x_1 < x_2 < \cdots < x_N$, but some of the derivations that follow are simpler if the events are ordered by time of arrival. We will use d to denote ranking statistic-ordered events, and d_{to} to denote time-ordered events.

We assume that both the foreground and background events are samples from an inhomogeneous Poisson process with respective differential rates

$$\frac{dN_f}{dx} = f(x, \theta) \tag{3}$$

and

$$\frac{dN_b}{dx} = b(x, \theta), \tag{4}$$

where the θ argument represents additional "shape" parameters that may affect the distribution, and for which we will eventually fit. The cumulative rates of the two processes are therefore

$$F(x, \theta) \equiv \int_{-\infty}^{x} ds\, f(s, \theta) \tag{5}$$

and

$$B(x, \theta) \equiv \int_{-\infty}^{x} ds\, b(s, \theta). \tag{6}$$

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 $F(x_2, \theta) - F(x_1, \theta)$ or $B(x_2, \theta) - B(x_1, \theta)$.

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

$$P(n = 1 \in [x, x + h]) = f(x, \theta)h + \mathcal{O}\left(h^2\right). \tag{7}$$

   and similarly for background events.

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

$$P(n = 2 \in [x, x + h]) = \mathcal{O}\left(h^2\right). \tag{8}$$

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

$$f(x, \theta) = R_f \hat{f}(x, \theta'), \tag{9}$$

and

$$b(x, \theta) = R_b \hat{b}(x, \theta'), \tag{10}$$

where $\hat{F}(\infty, \theta') = \hat{B}(\infty, \theta') = 1$, and $\theta' = \theta \setminus \{R_f, R_b\}$. Then $R_f \equiv F(\infty, \theta)$ and $R_b \equiv B(\infty, \theta)$ are the total number of foreground and background events expected and $\hat{f}(x, \theta')$ and $\hat{b}(x, \theta')$ are the likelihood of obtaining an event with ranking statistic x under the foreground and background distributions. In what follows, we will drop the prime, using θ to denote all parameters of the rate distributions except R_f and R_b.

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 (background) or 1 (foreground). These "state" flags are parameters in our model, along with R_f, R_b, and θ. We can marginalize over our uncertainty in the state of any given event by summing posteriors over $f_i = \{0, 1\}$.

Assuming time-ordered data, d_{to}, in the following, Bayes' theorem relates the posterior probability of the state flags, rates, and shape parameters, $p\left(\{f_i\}, R_f, R_b, \theta | d_{\mathrm{to}}, N\right)$, the likelihood of the data, $p\left(d_{\mathrm{to}} | \{f_i\}, N, R_f, R_b, \theta\right)$, and the prior probability of state flags, rates and shape parameters before any data are obtained, $p\left(\{f_i\}, N, R_f, R_b, \theta\right)$:

$$p\left(\{f_i\}, R_f, R_b, \theta | d_{\mathrm{to}}, N\right) = \frac{p\left(d_{\mathrm{to}} | \{f_i\}, N, R_f, R_b, \theta\right) p\left(\{f_i\}, N, R_f, R_b, \theta\right)}{p(d_{\mathrm{to}}, N)}. \tag{11}$$

The normalization constant, called the evidence, $p(d_{\mathrm{to}}, N)$, is independent of the state flags, rates, and shape parameters.

Each foreground event is drawn from the probability distribution $\hat{f}$ and each background event is drawn from the probability distribution $\hat{b}$. The events are independent of each other. Therefore, the likelihood of the data is

$$p\left(d_{\mathrm{to}} | \{f_i\}, N, R_f, R_b, \theta\right) = \left[\prod_{\{i | f_i = 1\}} \hat{f}\left(x_i, \theta\right)\right] \left[\prod_{\{i | f_i = 0\}} \hat{b}\left(x_i, \theta\right)\right]. \tag{12}$$

This is the probability that the first observed event is a fore/background event (if $f_1 = 1, 0$) with ranking statistic x_1 and the second observed event is a fore/background event (if $f_2 = 1, 0$) with ranking statistic x_2 etc. If the events are ordered by ranking statistic the corresponding expression is more complicated, since x_1 is now the event from foreground or background with the smallest ranking statistic, etc. We will return to the statistic-ordered case later.

The prior distribution can be factorized as

$$p\left(\left\{f_i\right\}, N, R_f, R_b, \theta\right) = p\left(\left\{f_i\right\}|N, R_f, R_b\right) p\left(N|R_f, R_b\right) p\left(R_f, R_b, \theta\right) = p\left(\left\{f_i\right\}, N|R_f, R_b\right) p\left(R_f, R_b, \theta\right). \tag{13}$$

The probability that the i'th state flag is $f_i = 1$ is given by $R_f/(R_f + R_b)$, while the probability that it is zero is $R_b/(R_f + R_b)$, provided the data are time-ordered as we have assumed. Then

$$p\left(\left\{f_i\right\}|N, R_f, R_b\right) = \prod_{\{i|f_i=1\}} \left(\frac{R_f}{R_f + R_b}\right) \prod_{\{i|f_i=0\}} \left(\frac{R_b}{R_f + R_b}\right) = \left(\frac{R_f}{R_f + R_b}\right)^{N_f} \left(\frac{R_b}{R_f + R_b}\right)^{N_b}, \tag{14}$$

where N_f and N_b are the numbers of foreground and background flags, $N_f + N_b = N$. Meanwhile,

$$p\left(N|R_f, R_b\right) = \frac{\left(R_f + R_b\right)^N}{N!} e^{-\left(R_f + R_b\right)}, \tag{15}$$

since the distribution of total event number is a Poisson process with rate $R_f + R_b$. Combining these yields the conditional probability of the flags on the rates:

$$p\left(\left\{f_i\right\}, N|R_f, R_b\right) = \frac{R_f^{N_f} R_b^{N_b}}{N!} \exp\left[-\left(R_f + R_b\right)\right]. \tag{16}$$

The second term in Eq. (13) is a traditional prior. Because the rate parameters enter the posterior in the same form as Poisson rates, we choose here the Poisson Jeffreys prior on rates [**?**], independent of the shape parameters

$$p\left(R_f, R_b, \theta\right) = \alpha \frac{1}{\sqrt{R_f R_b}} p(\theta), \tag{17}$$

where α is a normalisation constant, but of course other choices are possible. This choice has the advantage that the prior is normalizable as $R_f, R_b \to 0$, and the exponentials in Eq. (16) regularize the posterior as $R_f, R_b \to \infty$.

Putting everything together, the posterior is

$$p\left(\left\{f_i\right\}, R_f, R_b, \theta|d_{\mathrm{to}}, N\right) = \frac{\alpha}{p(d_{\mathrm{to}}, N)\, N!} \left[\prod_{\{i|f_i=1\}} R_f \hat{f}\left(x_i, \theta\right)\right] \left[\prod_{\{i|f_i=0\}} R_b \hat{b}\left(x_i, \theta\right)\right] \exp\left[-\left(R_f + R_b\right)\right] \frac{p(\theta)}{\sqrt{R_f R_b}} \tag{18}$$

When sampling the posterior, the first term, which is independent of the parameters of interest, can be omitted and the equals sign replaced by proportionality; however, we have kept this term explicitly so that we can see the equivalence to ranking-statistic ordered data. Once data have been observed, there is a unique loudness ordering and time ordering of those events, and so there is a one to one correspondence between a time-ordered posterior $p\left(\left\{f_i\right\}, R_f, R_b, \theta|d_{\mathrm{to}}, N\right)$ and the corresponding statistic-ordered posterior $p\left(\left\{f_i\right\}, R_f, R_b, \theta|d, N\right)$, which means $p\left(\left\{f_i\right\}, R_f, R_b, \theta|d, N\right) = p\left(\left\{f_i\right\}, R_f, R_b, \theta|d_{\mathrm{to}}, N\right)$. However, the evidence $p(d, N) = N!\, p(d_{\mathrm{to}}, N)$, since there are $N!$ ways in which N events can be ordered in time and would have the same set of ranking statistics.

Not surprisingly, this ranking-statistic ordered posterior can be computed directly by assuming that the flags, $\{f_i\}$, are un-observed data and treating the sets $\{x_i|f_i = 1\}$ and $\{x_i|f_i = 0\}$ as samples from an inhomogeneous Poisson process. For an inhomogeneous Poisson process with rate function $r(y)$ (cumulative rate $R(y)$), the likelihood of a set of samples $\{y_i\}$ is given by

$$\begin{aligned} p\left(\left\{y_i\right\}|r\right) \mathrm{d}^N y_i &= P\left(\text{zero events below } y_1\right) \times P\left(\text{one event between } y_1 \text{ and } y_1 + \mathrm{d}y_1\right) \\ &\quad \times P\left(\text{zero events between } y_1 + \mathrm{d}y_1 \text{ and } y_2\right) \ldots \\ p\left(\left\{y_i\right\}|r\right) &= \lim_{\delta y_i \to 0} \exp\left[-R\left(y_1\right)\right]\left[r\left(y_1\right) + \mathcal{O}\left(\delta y_1\right)\right] \times \exp\left[-\left[R\left(y_2\right) - R\left(y_1 + \delta y_1\right)\right]\right] \times \ldots \\ &= \left[\prod_i r\left(y_i\right)\right] \exp\left[-R\left(\infty\right)\right]. \end{aligned} \tag{19}$$

Applying this once to the foreground samples, once to the background samples and taking the product, we obtain $p(d, \{f_i\}, N|R_f, R_b, \theta)$ and thence $p(\{f_i\}, R_f, R_b, \theta|d, N) = p(d, \{f_i\}, N|R_f, R_b, \theta)\ p(R_f, R_b, \theta) \,/\, p(d, N)$. With the identification $p(d, N) = N!\, p(d_{\mathrm{to}}, N)$, as justified above, we reproduce Eq. (18).

We can marginalize the posterior over the flags, f_i, obtaining

$$p\left(R_f, R_b, \theta|d, N\right) = \sum_{\{f_i\}\in\{0,1\}^N} p\left(\{f_i\}, R_f, R_b, \theta|d, N\right) \propto \prod_i \left[R_f \hat{f}\left(x_i, \theta\right) + R_b \hat{b}\left(x_i, \theta\right)\right] \exp\left[-\left(R_f + R_b\right)\right] \frac{p(\theta)}{\sqrt{R_f R_b}}. \tag{20}$$

This expression is useful if we are only interested in rates and not the probability that any particular event is foreground or background. Unlike the full posterior (Eq. (18)), Eq. (20) contains only continuous parameters. We note that the terms that depend on the overall rate parameters, R_b or R_f, are of the form $R_b^{n-1/2}\exp(-R_b)$ and so marginalisation over either R_b or R_f can be achieved analytically using

$$I_n = \int_0^\infty x^{n-\frac{1}{2}}\, \mathrm{e}^{-x} \mathrm{d}x = \frac{(2n-1)!!}{2^n}\sqrt{\pi} \tag{21}$$

using the usual notation $(2n-1)!! \equiv (2n-1)(2n-3)\cdots 1$.

Eq. (18) is unchanged if the ranking statistic is multi-dimensional; in this case, the rates are

$$R_f = \int d^k\vec{x}\, f(x, \theta) \tag{22}$$

and

$$R_b = \int d^k\vec{x}\, b(x, \theta), \tag{23}$$

where f and b are rate densities on the k-dimensional space of ranking statistics. We give an example of fitting for multi-dimensional rate densities in § III D.

It is informative to relate these results to two other methods for estimating the foreground rate parameter — the loudest event statistic and the foreground-dominated statistic.

1. *Loudest event statistic*

If we were to include only the k loudest events in the posterior distribution, rather than all observed events, the posterior (Eq. (18)) would be modified by an additional factor of $\exp[R_f \hat{F}(x_{N-k+1}, \theta) + R_b \hat{B}(x_{N-k+1}, \theta)]$, where we have assumed events are ordered by loudness, so that x_{N-k+1} is the k-th loudest event. This term accounts for the data-dependent threshold that a loudest event statistic employs.

For the usual $k = 1$ case [**?**], the marginalised posterior (Eq. (20)) becomes

$$p_{\mathrm{LE}}\left(R_f, R_b, \theta|d\right) \propto \left(R_f \hat{f}(x_N, \theta) + R_b \hat{b}(x_N, \theta)\right) \exp\left[-\left(R_f(1 - \hat{F}(x_N, \theta)) + R_b(1 - \hat{B}(x_N, \theta))\right)\right] \frac{p(\theta)}{\sqrt{R_f R_b}}. \tag{24}$$

where x_N denotes the loudness of the loudest event. Marginalising over R_b we obtain

$$p_{\mathrm{LE}}\left(R_f, \theta|d\right) \propto \left(\frac{\hat{b}(x_N, \theta)}{2(1 - \hat{B}(x_N, \theta))} + R_f \hat{f}(x_N, \theta)\right) \frac{\sqrt{\pi}\, p(\theta)}{\sqrt{1 - \hat{B}(x_N, \theta)}\,\sqrt{R_f}} \exp\left(-R_f(1 - \hat{F}(x_N, \theta))\right). \tag{25}$$

This posterior has a maximum in R_f at

$$R_f = \frac{\hat{f}(x_N, \theta) - (1 - \hat{F}(x_N, \theta))\tilde{b}(x_N, \theta) + \sqrt{\left(\hat{f}(x_N, \theta) - (1 - \hat{F}(x_N, \theta))\tilde{b}(x_N, \theta)\right)^2 - 4\tilde{b}(x_N, \theta)(1 - \hat{F}(x_N, \theta))\hat{f}(x_N, \theta)}}{4\hat{f}(x_N, \theta)(1 - \hat{F}(x_N, \theta))}. \tag{26}$$

where $\tilde{b}(x_N, \theta) = \hat{b}(x_N, \theta)/(1 - \hat{B}(x_N, \theta))$. If $\tilde{b}(x_N, \theta) \ll 1$, we obtain the result $(1 - \hat{F}(x_N, \theta))\, R_f \approx 1/2$. This can be understood as the statement that the rate of foreground events with ranking statistic greater than x_N, $(1 - \hat{F}(x_N, \theta))\, R_f$, is of order 1, as expected. However, $\tilde{b}(x_N, \theta) = -\mathrm{d}[\ln(1 - \hat{B}(x, \theta))]/\mathrm{d}x$ and $(1 - \hat{B}(x, \theta)) \to 0$ as

$x \to \infty$, so this term will be divergent and for many reasonable examples, we will find $\tilde{b}(x_N, \theta) >> \hat{f}(x_N, \theta)$, in which case the posterior on R_f is peaked at 0. This issue highlights the problem with using a loudest-event statistic with an improper prior on the background rate R_b. No matter how improbable an event with $x = x_N$ is under the background distribution, it can become likely that the event at x_N is from the background distribution by taking the background rate to be sufficiently large. Although this predicts many more events with $x < x_N$, by using only the loudest event we do not incorporate the information that no such events are seen. This problem is avoided in the new framework described here, since we use all events detected above threshold and combined rates, $R_f + R_b$, significantly greater than the total number of observed events are strongly disfavoured.

This problem can be avoided in the loudest event framework, by including an upper limit on the rate, $R_{\max}$, in the prior for R_b. The marginalised distribution for the foreground rate then becomes

$$p_{\rm LE}\left(R_f, \theta | d\right) \propto \left(\frac{\hat{b}(x_N,\theta)}{(1-\hat{B}(x_N,\theta))^{\frac{3}{2}}} \left[\frac{\sqrt{\pi}}{2} \mathrm{erf}\left(\sqrt{(1-\hat{B}(x_N,\theta))\,R_{\max}}\right) - \sqrt{(1-\hat{B}(x_N,\theta))\,R_{\max}}\, \mathrm{e}^{-(1-\hat{B}(x_N,\theta))\,R_{\max}} \right] \right.$$
$$\left. + R_f \hat{f}(x_N,\theta) \frac{\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{(1-\hat{B}(x_N,\theta))\,R_{\max}}\right)}{\sqrt{(1-\hat{B}(x_N,\theta))}} \right) \frac{p(\theta)}{\sqrt{R_f}} \exp\left(-R_f(1-\hat{F}(x_N,\theta))\right), \qquad (27)$$

where $\mathrm{erf}(x)$ is the error function, defined in the usual way $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-u^2) \mathrm{d}u$. If $(1-\hat{B}(x_N,\theta))\,R_{\max} \ll 1$, Eq. (27) can be approximated by

$$p_{\rm LE}\left(R_f, \theta | d\right) \propto \left(\frac{R_{\max}}{3} \hat{b}(x_N,\theta) + R_f \hat{f}(x_N,\theta) \right) \frac{p(\theta)}{\sqrt{R_f}} \exp\left(-R_f(1-\hat{F}(x_N,\theta))\right) \qquad (28)$$

and if $\hat{f}(x_N,\theta) \gg R_{\max}\hat{b}(x_N.\theta)$ we find the same result as before, $(1-\hat{F}(x_N,\theta))\,R_f \approx 1/2$.

2. Foreground dominated statistic

If we set the threshold for including an event, $x_{\min}$, sufficiently high, we can ensure that $\hat{f}(x_i,\theta) \gg \hat{b}(x_i,\theta)$ for all ranking statistics x_i in the data set. The posterior can then be approximated by

$$p_{\rm FD}\left(R_f, R_b, \theta | d\right) \propto \prod_i \left[\hat{f}\left(x_i,\theta\right)\right] R_f^N \exp\left[-\left(R_f + R_b\right)\right] \frac{p(\theta)}{\sqrt{R_f R_b}}. \qquad (29)$$

Normalisation over R_b gives a constant factor and the posterior on the foreground rate becomes

$$p_{\rm FD}\left(R_f, \theta | d\right) \propto \prod_i \left[\hat{f}\left(x_i,\theta\right)\right] R_f^{N-\frac{1}{2}} \exp\left[-R_f\right] p(\theta). \qquad (30)$$

Ignoring the dependence on θ, this is peaked at a rate $R_f = N - 1/2$, so we have the expected result that, in the foreground dominated regime, the rate is approximately equal to the number of events observed. We note that the rate R_f is defined as the rate of events occurring above the specified threshold and therefore if the foreground dominated statistic were used to estimate rates $R_{f,1}$ and $R_{f,2}$ for thresholds $x_{\min} = x_1$, $x_{\min} = x_2$ respectively, these rates should be compared by equating $R_{f,1}(1-\hat{F}(x_2,\theta))$ and $R_{f,2}(1-\hat{F}(x_1,\theta))$.

III. EXAMPLES

In this section we present several examples of the application of our framework to various rate estimation problems in the presence of background.

A. Gravitational Waves with Non-Overlapping Templates

Suppose we attempt to detect gravitational wave signals in a data stream by matched filtering in the frequency domain against a set of N template waveforms [e.g., **?** **?**]. In our simplistic model, we suppose the data stream consists of stationary Gaussian noise with a power spectral density $S(f)$ combined additively with some number of gravitational wave signals. We assume that the signals are sufficiently rare that they do not overlap in the data stream. The signal to noise ratio (SNR) of a template, $h(f)$, given data, $d(f)$, is

$$\rho_h \equiv \frac{\langle h, d\rangle}{\sqrt{\langle h, h\rangle}}, \tag{31}$$

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

$$\langle a, b\rangle \equiv 4\Re \int_0^\infty df\, \frac{a^*(f) b(f)}{S(f)}. \tag{32}$$

We suppose for simplicity that the templates are sufficiently distinct that

$$\langle h_i, h_j\rangle \simeq \delta_{ij}. \tag{33}$$

In the following subsection, we will generalize the model to overlapping templates. We rank candidate events by their maxmim SNR over the entire template bank,

$$x \equiv \max_h \rho_h, \tag{34}$$

and consider only events that have a maximum SNR above some threshold, $x > x_{\mathrm{min}}$.

For a data stream of pure noise, $d(f) = n(f)$, the SNR of a each template follows a $N(0,1)$ distribution. The background ranking statistic (i.e. the maximum SNR over the template bank) then has a cumulative distribution without thresholding of

$$\hat{B}(x) = \left(\frac{1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)}{2}\right)^N \tag{35}$$

where $\mathrm{erf}(x)$ is the error function as before. Imposing the threshold, $x > x_{\mathrm{min}}$, the cumulative distribution of the background becomes

$$\hat{B}(x) = \frac{\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right)^N - \left(1 + \mathrm{erf}\left(\frac{x_{\mathrm{min}}}{\sqrt{2}}\right)\right)^N}{2^N - \left(1 + \mathrm{erf}\left(\frac{x_{\mathrm{min}}}{\sqrt{2}}\right)\right)^N} \tag{36}$$

for $x > x_{\mathrm{min}}$, 0 otherwise.

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

$$\hat{F}(x) = 1 - \frac{x_{\mathrm{min}}^3}{x^3}. \tag{37}$$

Note that this scenario has no shape parameters θ for the foreground and background distributions.

To demonstrate the effectiveness of our formalism, we applied it to a synthetic data set with foreground and background distributions drawn from Eqs. (36) and (37) with $R_f^{\mathrm{true}} = 10.4$ and $R_b^{\mathrm{true}} = 95.1$, using 1000 templates. **[Specify x_{min}?]** The synthetic data consisted of 13 foreground events and 85 background events; the cumulative distribution for the ranking statistic of the synthetic data appears in Figure 1. We used a Markov Chain Monte Carlo simulation to draw samples of state flags and rates from the joint posterior (Eq. (18)).

In Figure 2, we show the marginalized posterior densities for the foreground and background rates. (Refer to Eq. (20).) Figure 3 shows the posterior foreground probability for each event marginalized over all other events' types and the foreground and background rates.

We can compare these results to results obtained using the two approximations described earlier, the loudest-event statistic and the foreground-dominated statistic. The marginalised distribution for the foreground rate using these alternatives are shown in Figure 4. In this case, the loudest event had $x_N = 9.5$**(correct this)**. The loudest-event statistic depends on a specification of the maximum, R_{max}, for the background rate. We show results for $R_{\mathrm{max}} = \infty$, i.e., the improper prior, and $R_{\mathrm{max}} = 10000$. The results for other reasonable choices of $R_{\mathrm{max}} = 100, 1000, 100000$ etc.

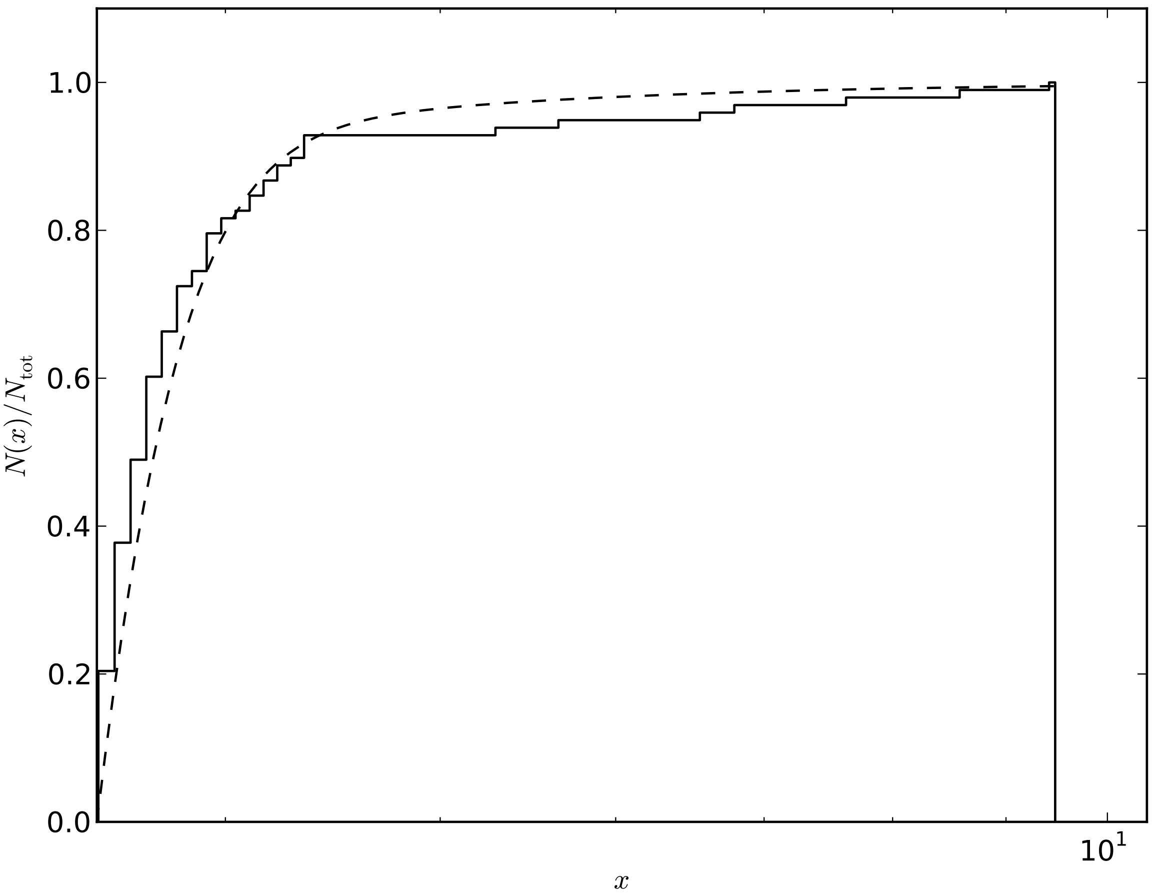

FIG. 1. The cumulative distribution of the ranking statistics for the synthetic data used to test the formalism on the model from §III A. The solid line gives the cumulative distribution of the synthetic data; the dashed line gives the theoretical cumulative distribution for the models in Eqs. (36) and (37) combined with $R_f = 10.4$ and $R_b = 95.1$.

gave exactly the same posterior, since $\hat{b}(x_N)R_{\max} \ll \hat{f}(x_N)$ for all these choices and we are therefore in the regime where the posterior is insensitive to $R_{\max}$. To apply the foreground-dominated statistic we must specify a threshold above which we assume all events are foreground. It is reasonable to do this based on a specification for the relative probability of an event being fore/background, $\hat{f}(x)/\hat{b}(x) = p_{\rm thresh}$. Setting $p_{\rm thresh} = 0.99$ gives $x_{\min} = 4.07$ and there are $N = 13$ events exceeding that threshold. Setting $p_{\rm thresh} = 0.5$ gives $x_{\min} = 3.82$ and there are $N = 15$ events exceeding that threshold **(correct these numbers using Will's data)**. We show results for both of these choices of $x_{\min}$ in Figure 4. We also show similar results for two extreme cases — setting $x_{\min} = 3.5$ so that we assume all 98 observed events are foreground; and setting $x_{\min} = 9$ so that only the loudest event is classified as foreground.

The loudest event statistic with the improper prior gives, as expected, a poor approximation to the foreground rate. The peak is more accurately located when a prior maximum rate is defined, but the distribution is much wider than using the full analysis described here. This is to be expected as much of the information is being thrown away. The foreground-dominated statistic gives a reasonable approximation to the true foreground rate, and a distribution that is not much wider than the full analysis, for the non-extreme choices of $x_{\min}$. For $x_{\min} = 9$, the foreground dominated statistic has equivalent performance to the loudest event statistic, as we might expect, while when $x_{\min} = 3.5$, it performs poorly since we are approximating the foreground rate by the total foreground plus background rate. This indicates that, provided the threshold is chosen appropriately, the foreground dominated statistic can perform quite well at estimating the rate. The fact that it reproduces the posterior from the full analysis so well is indicative of the fact that most of the information about the foreground comes from the loudest events. Indeed, for $x_{\min} \approx 4$, the set of events consists mostly of foreground sources and that set contains a decent number of events (13), so it is not

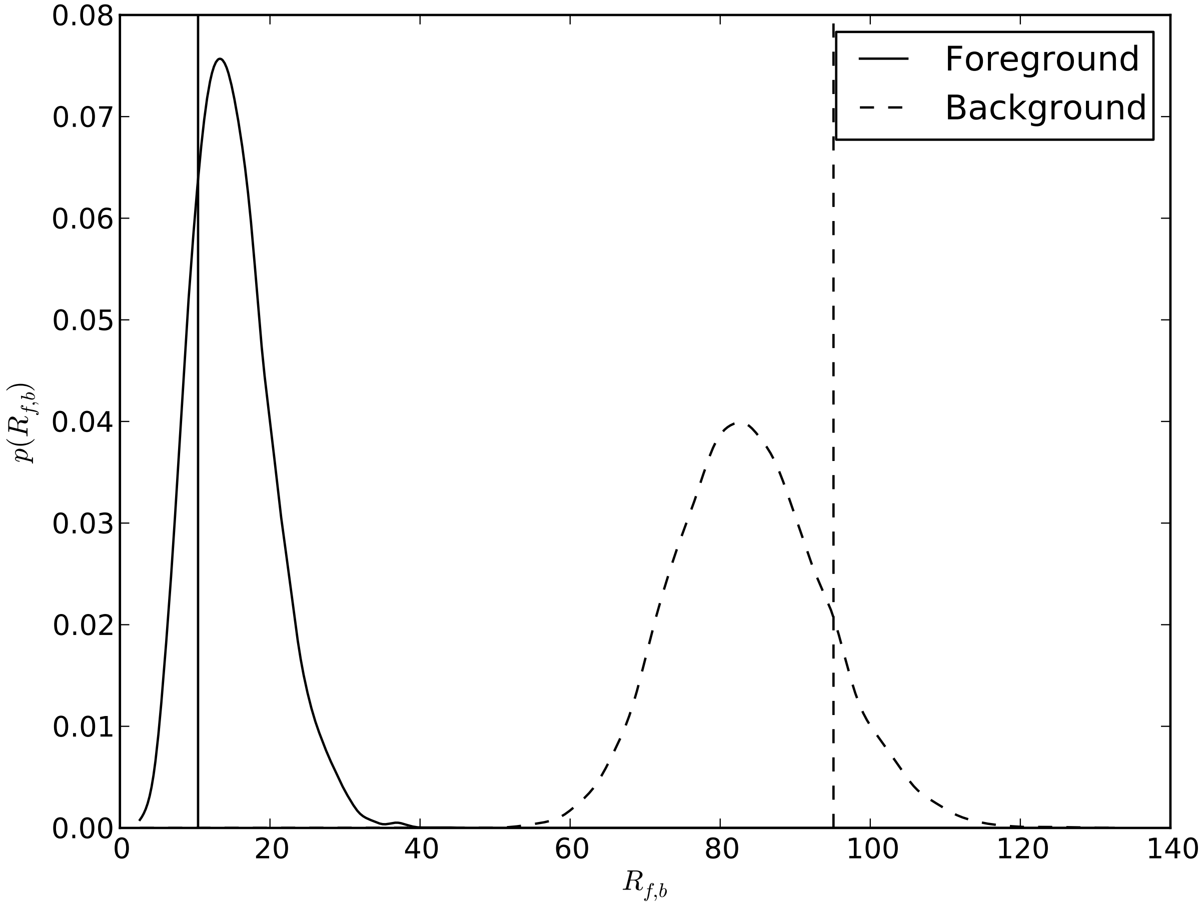

FIG. 2. The marginalized posterior densities for R_f (solid line) and R_b (dashed line) for the analytic model discussed in §III A. The vertical lines indicate the "true" values used to generate the synthetic data set. Both the true foreground and background rates lie well within the probability envelope for R_f and R_b.

surprising that this statistic can estimate the foreground rate well. The choice of $x_{\min}$ was motivated by knowledge of $\hat{b}(x,\theta)$ and $\hat{f}(x,\theta)$, which we have in this case, but this might not be possible if the parameters θ were simultaneously being estimated. The full analysis naturally incorporates inference about the rate parameters θ and the background rate R_b along with the foreground rate and incorporates maximum information from the data set and should therefore lead to narrower posteriors in general.

B. Gravitational Waves With Overlapping Templates

In §III A we assumed that the overlap between different templates in the template bank was negligible, so the SNRs recovered by different templates are independent random variables. In fact, template banks are not constructed in this way [e.g., **? ?**], because signals could fall in the gaps between the non-overlapping templates. We can model this effect by assuming that a template bank of N actual templates will behave as if it had $N_{\rm eff}$ *independent* templates. Rather than pre-computing $N_{\rm eff}$, we can fit for it as a shape parameter. That is, we assume that $\theta = \{N_{\rm eff}\}$ is a shape parameter for the background cumulative distribution:

$$\hat{B}\left(x, N_{\rm eff}\right) = \frac{\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)^{N_{\rm eff}} - \left(1 + \operatorname{erf}\left(\frac{x_{\min}}{\sqrt{2}}\right)\right)^{N_{\rm eff}}}{2^{N_{\rm eff}} - \left(1 + \operatorname{erf}\left(\frac{x_{\min}}{\sqrt{2}}\right)\right)^{N_{\rm eff}}}. \tag{38}$$

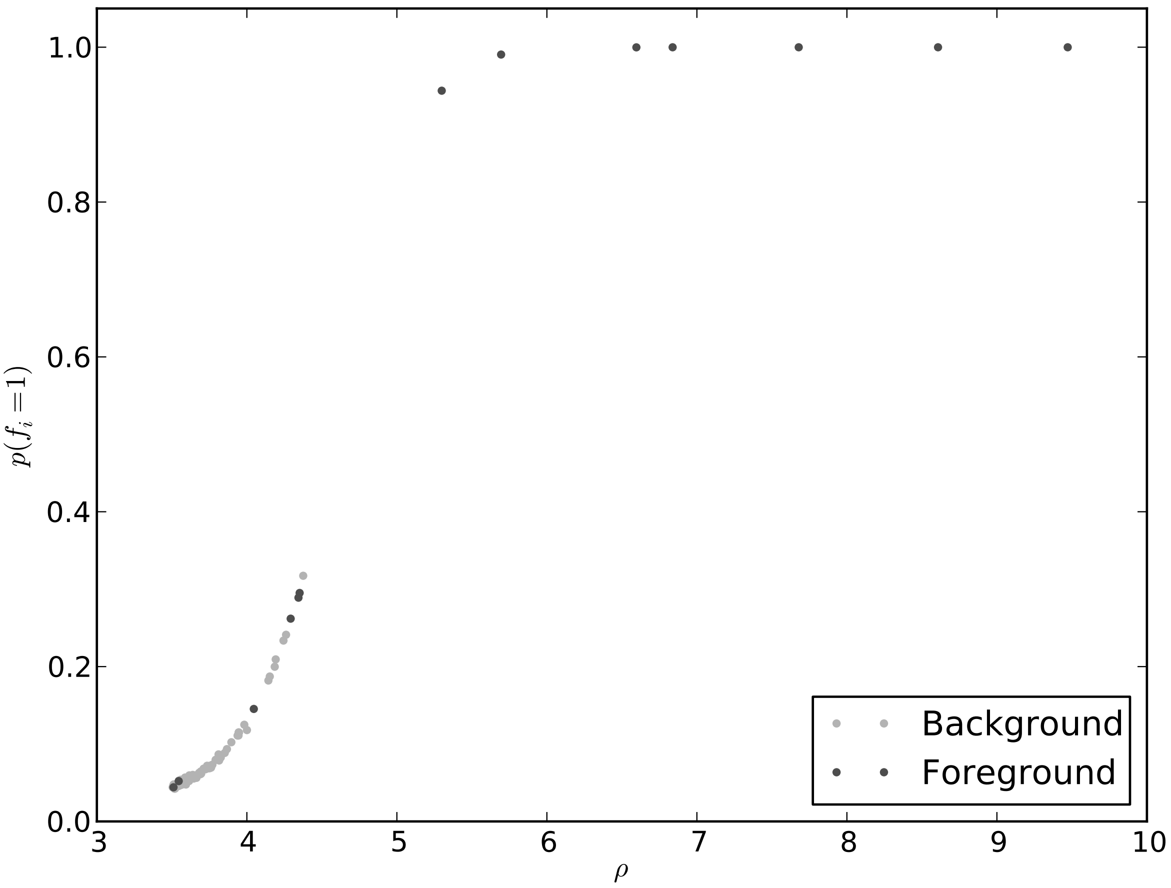

FIG. 3. Foreground probability for each event in the synthetic data set of §III A marginalized over all other parameters (error bars are one standard deviation). True foreground events are in dark grey, background events in light grey. Even though our method cannot identify the status of most events with confidence, it can still correctly estimate the rates (Figure 2).

Results from such an analysis appear in Figures 5 and 6. We use the same parameters and data set as in §III A, with $R_f = 10.4$, $R_b = 95.1$, and $N_{\mathrm{eff}} = 1000$, but now allow N_{eff} to be a parameter of the background distribution, with a flat prior. Both the rates and the number of effective templates are recovered without significant loss of accuracy relative to the fixed N_{eff} situation in §III A.

If we consider the two alternative methods, the loudest event and foreground dominated statistics, we will recover the same foreground distributions as in the case with N_{eff} fixed, which are shown in Figure 4. This is because the parameter N_{eff} affects only the background distribution, to which the foreground-dominated statistic is insensitive, and in the loudest event case, after marginalisation over N_{eff} we find $\int_0^{N_{\max}} \hat{b}(x_N, N_{\mathrm{eff}}) \mathrm{d}N_{\mathrm{eff}} \ll 3N_{\max}/R_{\max}\hat{f}(x_N, N_{\mathrm{eff}})$ and so we are still in the foreground-dominated regime in which the loudest event tells us nothing about the background. Similarly, neither of these alternative methods can inform us about the value of N_{eff}, a property of the background. **Technically, if we allowed really large values of N_{eff} we would get into a regime where the loudest event could be background and we'd get something here. But, we would need $N_{\max} \sim 10^{20}$ and so I don't think we should consider these alternatives further in this case. On the other hand, this ties in with the comment at the end of the previous subsection about the difficulty of choosing x_{min} for the foreground-dominated statistic when there are other parameters θ to be estimated simultaneously (N_{eff} in this case), so perhaps it's worthwhile to add some comment about this?**

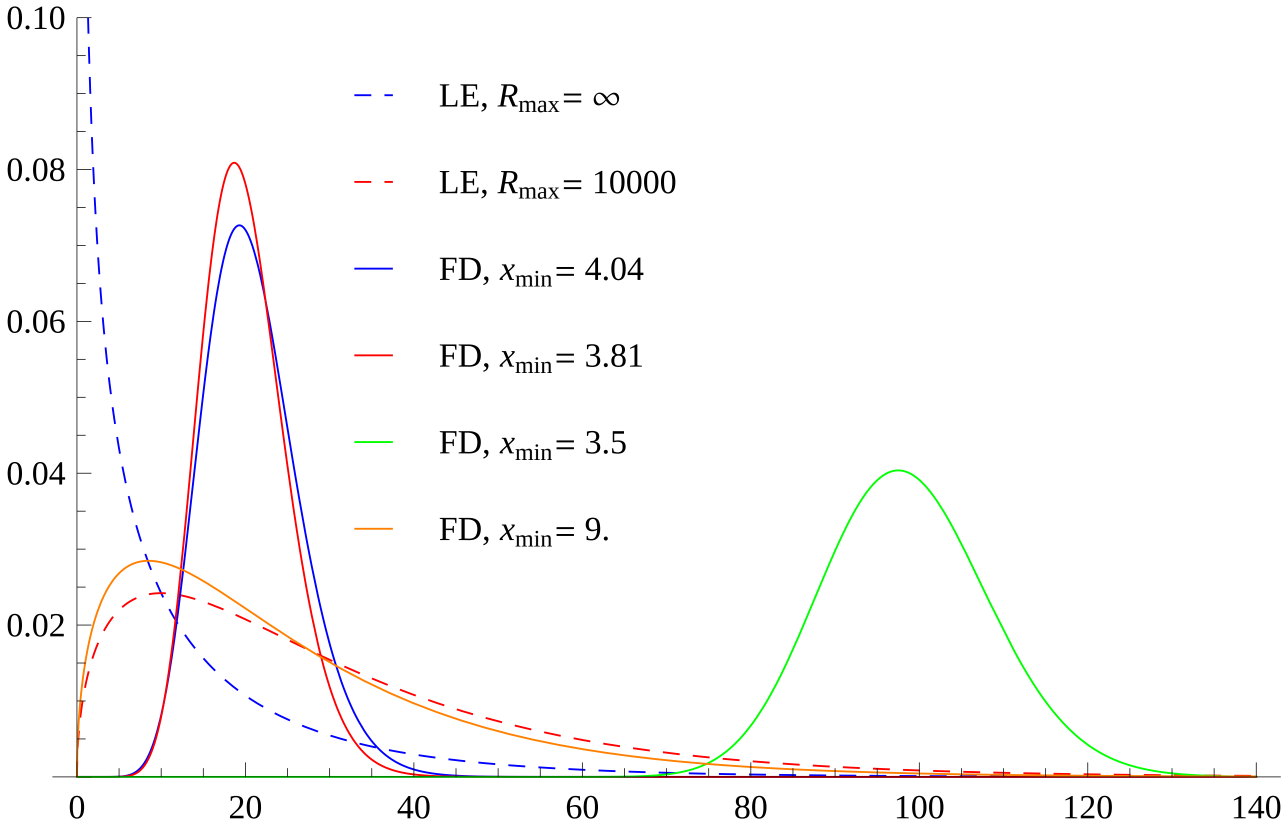

FIG. 4. Posteriors on foreground rate obtained using alternative methods — the loudest event statistic and the foreground dominated statistic. Results are shown for two different choices of the maximum a priori background rate, $R_{\rm max}$, in the case of the loudest event ("LE") statistic, and for several different choices of the threshold, $x_{\rm min}$, in the case of the foreground dominated ("FD") statistic. **I can see the figure on the screen, but can't print it...**

C. Dependence of the Bayesian Rate Estimate on the Value of the Threshold

This paper is concerned with Bayesian rate estimates based on a lists of events (which are a certain kind of summary of the complete data set): either the loudest event, or a list of all events whose SNR exceeds a certain, pre-set threshold. In this subsection we address the question of how the estimate depends on the threshold value.

To begin with, we recall the well-known fact that the Bayesian estimator is unbiased, in the following sense. For simplicity, assume that the model consists of a single rate parameter R, with prior distribution $p(R)$. Consider an ensemble of data sets whose distribution is consistent with that prior; i.e., such that $p(data)$ is given by

$$p(data) = \int p(data|R)\, p(R)\, dR\,. \tag{39}$$

For each data set in the ensemble, compute the Bayesian estimator $R_B = \int Rp(R|data)dR$. Then it is immediate that

$$\int R_B(data)\, p(data)\, d(data) = \int R\, p(R)\, dR\,. \tag{40}$$

I.e, the data-weighted average of the Bayesian estimate R_B equals the prior-weighted average R. Therefore (for any fixed prior) all threshold values will yield, on average, the same estimate of the rate. However this equality of averages does *not* imply that all threshold values yield the same information. In general, as the threshold is lowered to include more events, the error bar on the estimate shrink. In this subsection we give quantitative illustrations of how the error bar shrinks when the threshold is lowered.

Consider the following model problem. Let $p(\rho)\Delta\rho$ be the probability that data yields a putative event with SNR in the range $[\rho - \Delta\rho/2, \rho + \Delta\rho/2]$ We write $p = p_n + p_s$, where p_n is the distribution of noise events and p_s is the

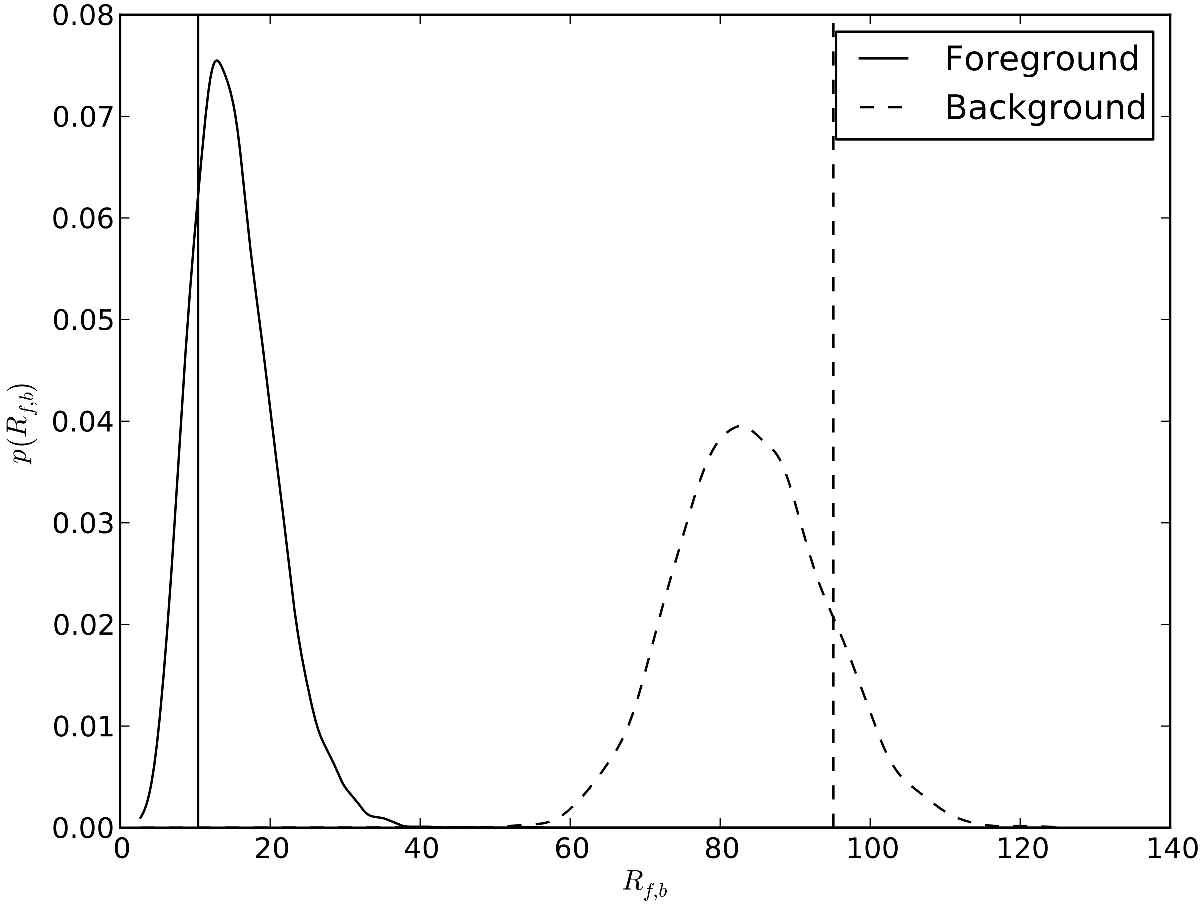

FIG. 5. The foreground (solid lines) and background (dashed lines) rate posterior, marginalized over all flags and the $N_{\rm eff}$ parameter, for the gravitational wave template detection scenario with overlapping templates discussed in §III B. The true values of the rates, $R_f = 10.4$ and $R_b = 95.1$, are indicated with vertical lines. The distributions are not significantly wider than those of Figure 2, in spite of the extra parameter.

distribution of events from actual astrophysical sources. Both p_n and p_s are the product of a rate (of noise or true events) and the observation time. Here we will assume that the noise distribution is Gaussian, so that p_n has the general form

$$p_n(\rho) = \Lambda_n \exp\big[-\rho^2/2\big]\,. \tag{41}$$

We find it useful to define ρ_1 as the SNR such that a data set will have on average a single noise event louder than ρ_1; i.e., such that

$$\int_{\rho_1}^{\infty} P_n(\rho)\,d\rho = 1\,. \tag{42}$$

For this example we will assume that the spatial distribution of true events can be approximated as uniform in Euclidean space. Then we can write the full $p(\rho)$ as

$$p(\rho) = \left[\sqrt{\pi/2}\mathrm{erfc}(\rho_1/\sqrt{2})\right]^{-1}\, exp[-\rho^2/2] + 3R_t\frac{\rho_1^3}{\rho^4}\,. \tag{43}$$

where R_t is the mean number of true events with $\rho > \rho_1$.

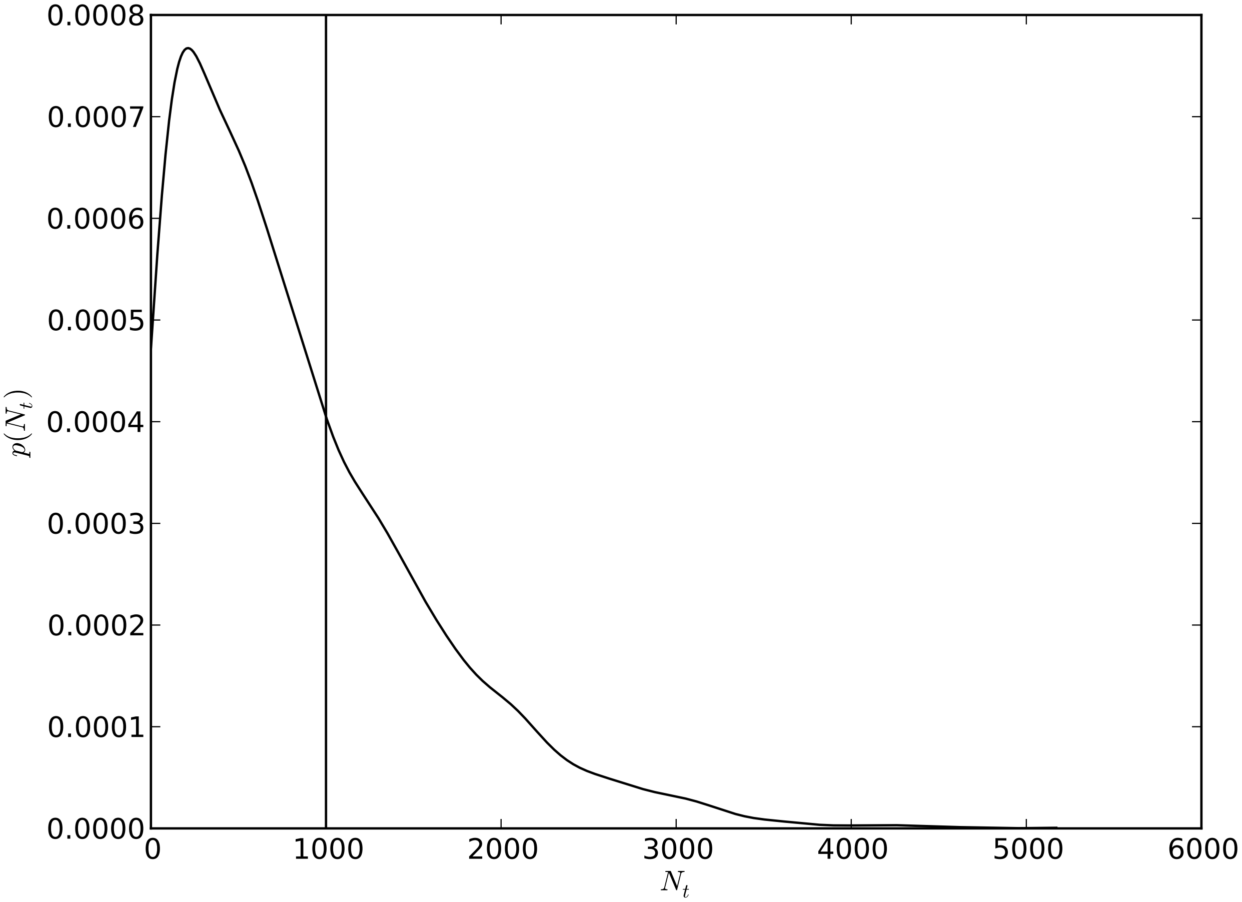

FIG. 6. The posterior on the number of effective templates, $N_{\rm eff}$, for the model and data discussed in §III B, marginalized over all state flags and rates. The true value, $N_{\rm eff} = 1000$, is indicated by the vertical line.

For any pair (ρ_1, R_t), it is straightforward to construct random event lists drawn from the corresponding $p(\rho)$, and straightforward to apply a threshold by "throwing away" all events with ρ less than some threshold value ρ_{th}. For any thresholded event list, we use Eqs. ..... to construct the probability density $p(R)$ (except that in this example, instead of the Jeffrey's prior, we assume a "hat-shaped" prior; see below). For that event list, we define ΔR by

$$(\Delta R_B)^2 = \int (R - R_B)^2 \, p(R) \, dR \,. \tag{44}$$

Figs. 7 –10 illustrate how ΔR varies with the threshold value ρ_{th}. In all cases we assumed that $\rho_1 = 8$. For Fig. **??**, we adopted a flat prior on R between 1 and 10 (i.e, we took $p(R) = 1/9$ for $1 < R < 10$, and zero outside that interval). We drew 10^3 random realizations $(R_t, data)$, where R_t was drawn from that flat prior and the data–a list of ρ values– was drawn from the distribution $p(data|R_t)$ given by Eq. 43 . For any given threshold value, ρ_{th}, we created an event list by taking the subset of the ρ values greater than ρ_{th} For each data set, we used a sequence of increasing threshold values ρ_{th} to create a corresponding sequence of (ever shorter) event lists. For each event list, we calculated ΔR and $R_B - R_t$. As a code check, we verified that the mean of $R_B - R_t$ (over all realizations) was consistent with zero for all ρ_{th}. Fig. **??** displays both the mean value of ΔR and the mean number of noise events (as opposed to actual signals) as a function of ρ_{th}. Note the relative insensitivity of ΔR_t to ρ_{th}: as the mean number of noise events on the event list shrinks from 8 to $\sim 10^{-6}$, the average ΔR_B increases only $\sim 14\%$, from 0.42to 0.48.

Figs. 8 –10 were constructed similarly, but with the following differences. First, we took the prior $p(R)$ to be $p(R) \propto 1/R$ for $0.03 < R < 30$, and zero outside that interval (i.e., $p(R)$ is the Jeffries prior, but restricted to some range). Second, we regarded R_t as fixed (as of course it would be in Nature), and only varied the data–randomly

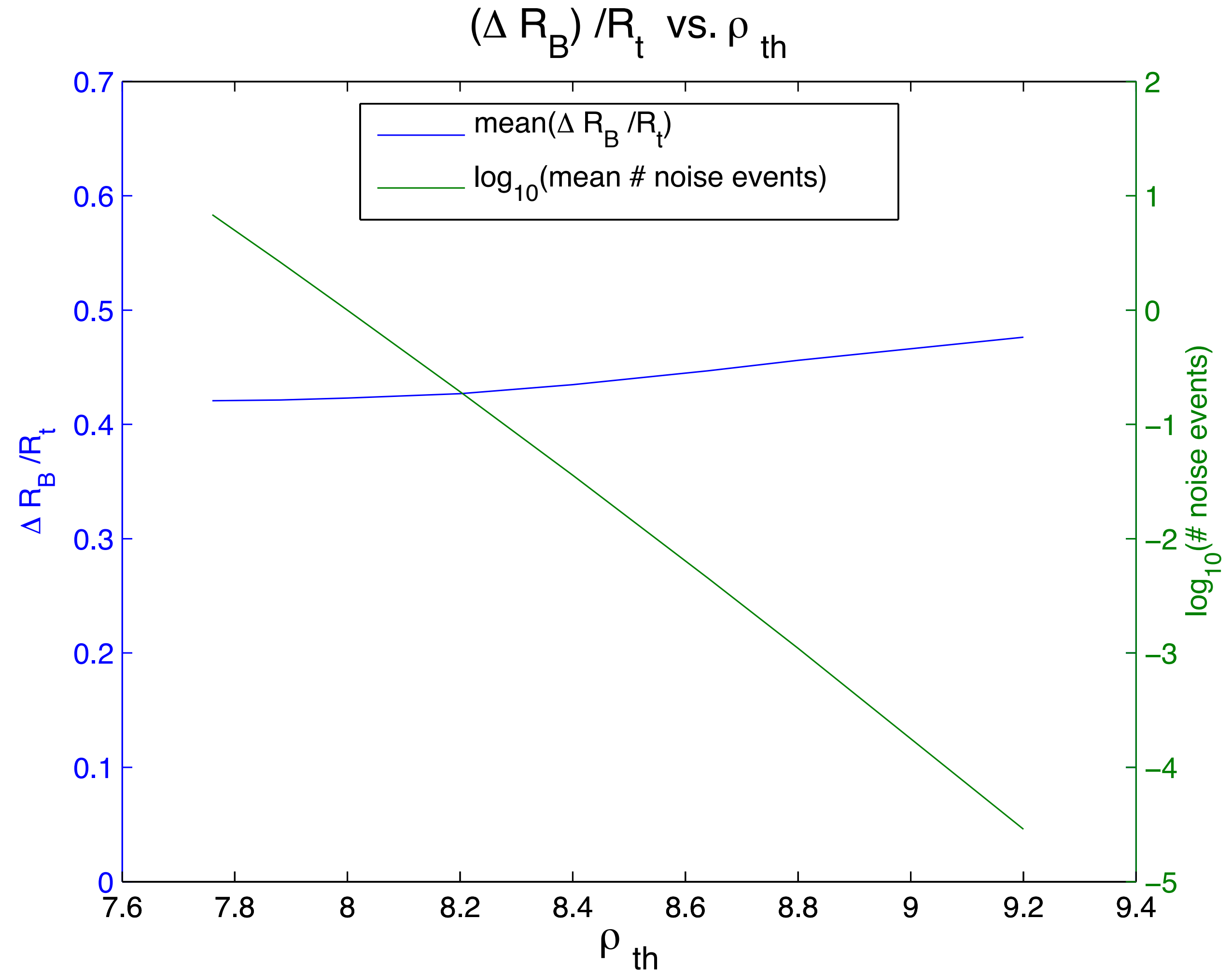

FIG. 7. Displays the mean "error bar" ΔR_B as function of the threshold value ρ_{th}. In this case, the values of R_t were randomly drawn from a flat prior on the interval $[1, 10]$. We also show how the (mean) size of the event list varies with ρ_{th}.

drawing 10^3 realizations from $p(data|R_t)$. For each realization we calculated ΔR_B and $(R_B - R_t)$, and then we calculated the mean over all realizations. The mean of $(R_B - R_t)$ was generally *not* zero, since the prior biases the results. (Of course, if one *updated* the prior after each realization, the bias would decrease towards zero, but that is not what we did: we evaluated each realization with the same Jeffries prior.)

The number of noise events is the same as for the first case, shown in Fig. **??**. Note that again, for all three values of R_t, the mean uncertainty ΔR_B increases by only $\sim 20\%$ as the mean number of noise events on the event list shrinks from 8 to $\sim 10^{-6}$. And as one would expect, the relative uncertainty shrinks as R_t increases.

D. Star Cluster Parameters With Background Contamination

Look up von Hippel paper about Bayes 9, the cluster isochrone fitting software—we should cite, since our approach to cluster membership.

Our final example concerns fitting for the location and shape parameters of a cluster of stars observed on top of a stellar background with a density gradient. In this example, stars are either members of the cluster (i.e. foreground) or background contamination, with a spatially varying density (i.e. our rate functions are two-dimensional). We assume that a star cluster has a Plummer surface-density profile [**? ?**],

$$\hat{f}(\vec{x}, \theta) = \frac{1}{\pi r_0^2 \left(1 + \frac{|\vec{x} - \vec{x}_0|^2}{r_0^2}\right)^2}, \tag{45}$$

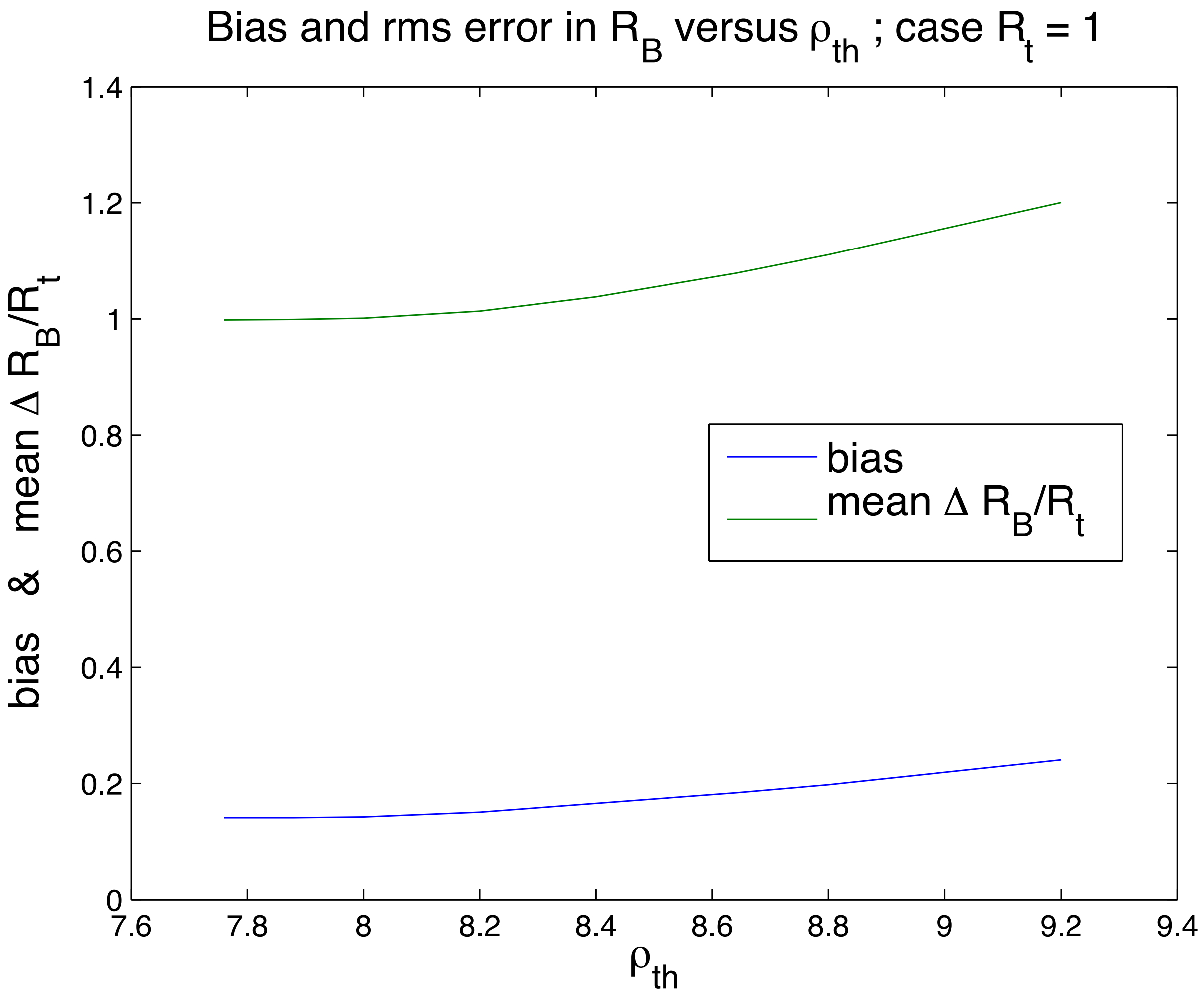

FIG. 8. Displays the mean rms error ΔR_B and mean $(R_B - R_t)$ (i.e., the bias) as functions of the threshold value ρ_{th}. In this case, the actual value of R_t was 1.0, and $p(R|data)$ was computed using the Jeffries prior on the interval $[0.03, 30]$.

where $\vec{x}_0$ is the location on the sky of the center of the cluster, r_0 is a radial scale parameter, and $\vec{x} = (x, y)$ is the position on the sky. We assume a square observational domain[**?**], $\vec{x} \in [0,1]^2$, and a background that has a density gradient at an arbitrary orientation with respect to the observational axes:

$$\hat{b}\left(\vec{x}, \theta\right) = 1 + \vec{\gamma} \cdot \left(\vec{x} - \vec{x}_{1/2}\right), \tag{46}$$

where $\vec{\gamma}$ is the gradient, and $\vec{x}_{1/2} = [1/2, 1/2]$ is the centroid of the observational domain. **[I think that if the domain is finite, the previous equation is only normalized for all $\vec{\gamma}$ if $\vec{x}_{1/2}$ is in the center of a symmetric domain (as it happens to be in this case).]** **Yes, as you say, this function is only normalized for $\vec{x}_{1/2}$ is the center, and also for small enough $\vec{\gamma}$; both of these conditions obtain here.**

We use simulated data drawn from our model with parameters

$$\theta_0 \equiv \{x_0, y_0, r_0, \gamma_x, \gamma_y\} = \left\{\frac{1}{2}, \frac{1}{2}, 0.18, -\frac{1}{2}, \frac{1}{2}\right\}, \tag{47}$$

with $R_f = 1000$ and $R_b = 10000$. For this set of parameters, the average density of the background and the peak density of the cluster are comparable; there are an order of magnitude more background stars than cluster stars in the field. Figure 11 shows the density of stars (cluster and background) on the sky for our particular data set.

To analyze our synthetic data set, we analytically marginalized over the state flags (i.e. cluster membership), using the likelihood in Eq. (20). We did this to take advantage of the *emcee* sampler of **?**], which requires all parameters to be in $\mathbb{R}$. We applied a prior on the shape parameters that is flat in $\vec{x}_0$ and $\vec{\gamma}$, and an (approximately) Jeffreys prior

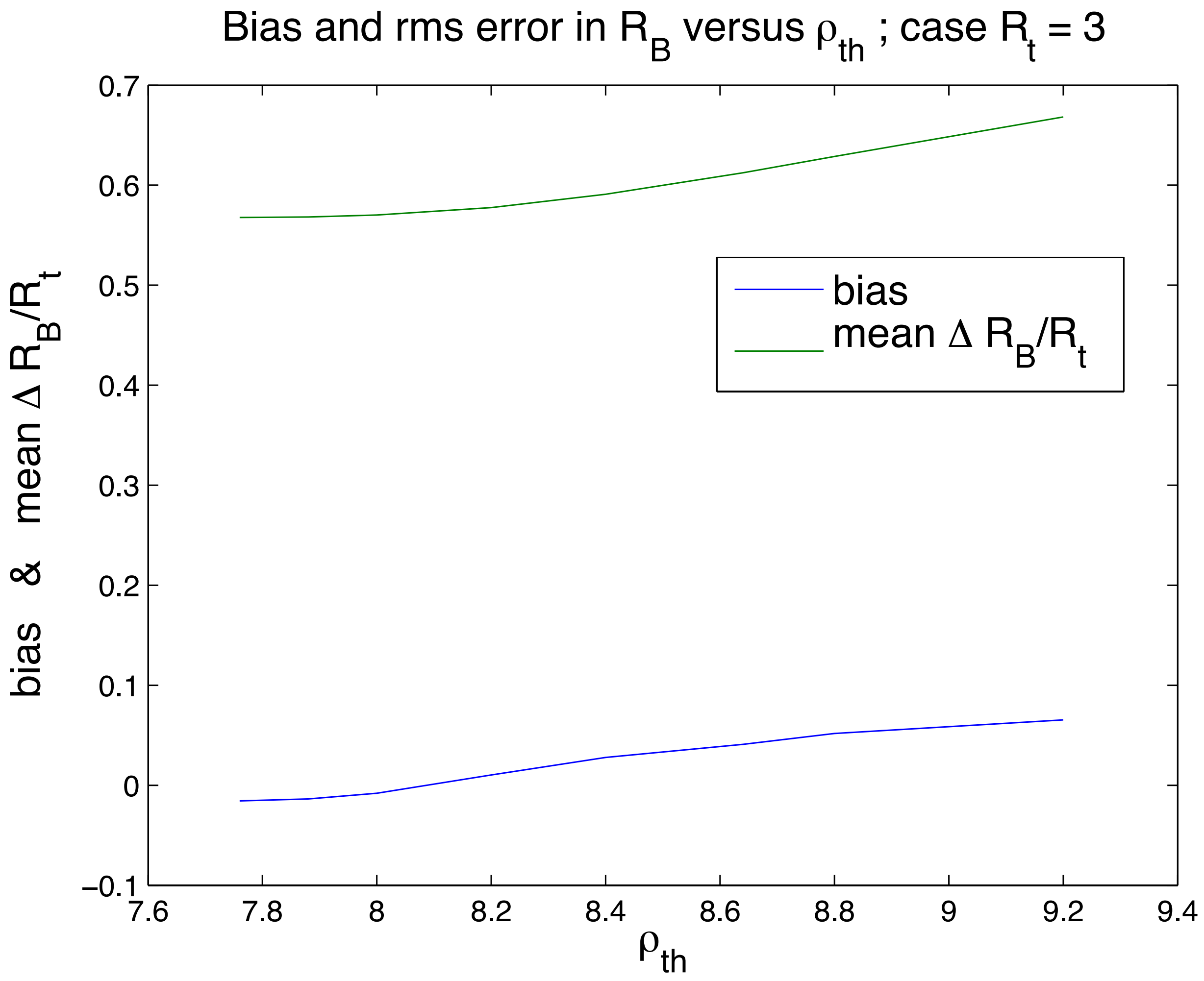

FIG. 9. Same as the previous figure, but displaying the size of ΔR_B relative to R_t, with the actual R_t equal to 3.0.

on r_0,

$$p\left(r_0\right) = \frac{\sqrt{R_f}}{r_0}. \tag{48}$$

[Why is this the Jeffreys prior (BTW, it's Jeffreys, or perhaps Jeffreys', but definitely not Jeffrey's ;))? It's not obvious why the prior on r_0 should depend on R_f?] Here is the calculation:

$$p\left(r_0\right) = \sqrt{\left\langle\left(\frac{\partial \log \mathcal{L}}{\partial r_0}\right)^2\right\rangle} = \sqrt{R_f\left\langle\left(\frac{\partial \log \hat{f}}{\partial r_0}\right)^2\right\rangle} = \sqrt{R_f\left\langle\left(\frac{2}{r_0} - \frac{4r_0}{r^2+r_0^2}\right)^2\right\rangle} = \sqrt{\frac{4R_f}{3r_0^2}} \propto \frac{\sqrt{R_f}}{r_0}. \tag{49}$$

Sorry, I still don't get it. Isn't $\mathcal{L} = R_f\hat{f} + R_b\hat{b}$? If so, how do you get the second equality above? Note that the calculation of the prior uses the likelihood before analytical marginalization over the flags, f_i. We are ignoring the fact that R_f and r_0 are correlated. In truth, the Jeffreys prior would be the determinant of the matrix built out of terms like $\partial \log \mathcal{L}/\partial\theta_i \times \partial \log \mathcal{L}/\partial\theta_j$, but let's not get carried away. . . . As another aside, note that we could have predicted the $1/r_0$ dependence because r_0 has units, so the only unit-invariant prior is to be flat in $\log r_0$; I'm don't have a similar argument for the appearance of the $\sqrt{R_f}$ factor. (Note that this factor of $\sqrt{R_f}$ cancels with the Jeffreys prior on the rate, $1/\sqrt{R_f}$; we have verified that the priors on these parameters are irrelevant to our results, as would be expected from the measurement of $\sim$ 1000 foreground stars.)

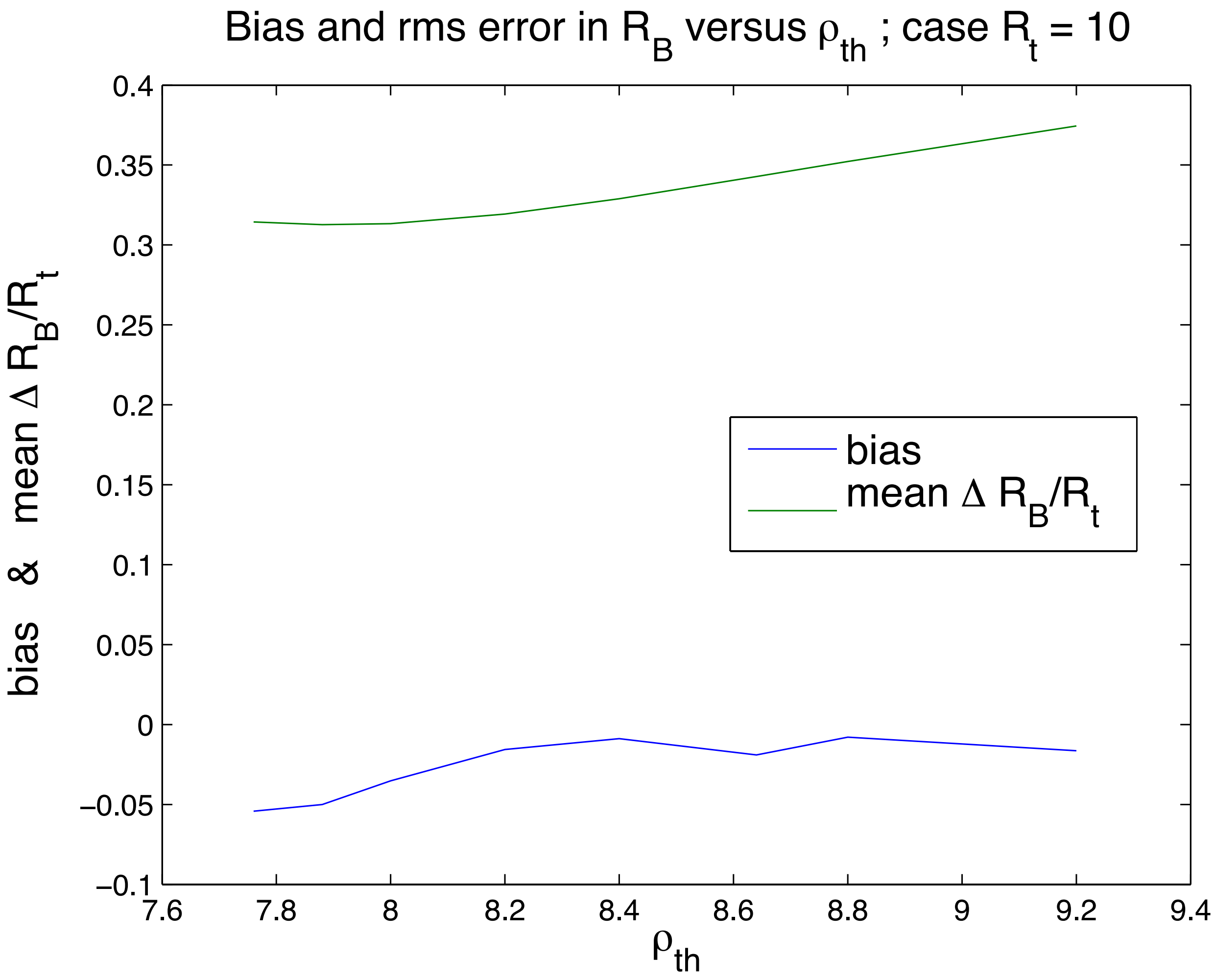

FIG. 10. Same as the previous figure, but with the actual R_t equal to 10.0.

Figure 12 shows the posterior for the location parameters, $\vec{x}_0$; the center of the cluster is localized to within about 10% of the cluster scale. Figure 13 shows the posteriors inferred on the cluster and background numbers, R_f and R_b, and Figure 14 shows the posterior for the cluster's scale parameter. In spite of the significant background, the cluster scale and total number are accurately recovered by our analysis.

IV. DISCUSSION

In this paper, we have developed a Bayesian framework for rate estimation when the data consists of a mixture of foreground and background events. We demonstrated the application of this framework using several examples from gravitational-wave data analysis in the presence of signatures of binary mergers and noise triggers, and astronomical image analysis in the presence of several populations of stars. We showed that this framework is generally superior to both the loudest-event statistic and and the foreground-dominated statistic.

Throughout this paper, we have assumed that the shape of the foreground and background distributions is known, or at least can be modeled with several additional parameters. This is not necessarily easy to do. For example, in the case of gravitational-wave data analysis, the shape of the foreground distribution of events may depend on the details of a complex data-analysis pipeline as well as the astrophysical source distribution, while the background event distribution depends on data quality and may deviate significantly from the simple Gaussian-noise behavior modeled in section III. Several approaches have been developed to accurately model both distributions, e.g., through the use of injected signals to model the foreground distribution [**? ?**]. However, this is a difficult problem (e.g., because of the need to estimate the background at the very tails of the distribution), and will require significant future work.

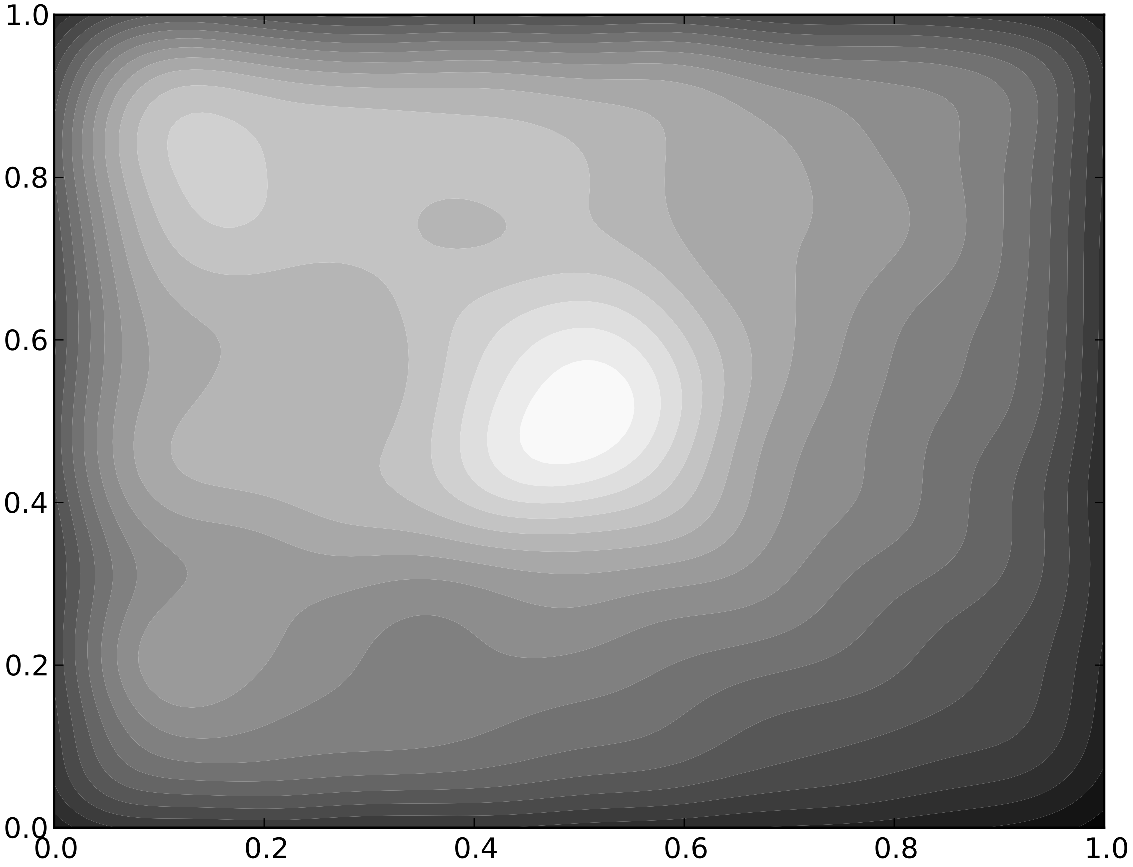

FIG. 11. The density contours on the sky of our synthetic data set of cluster and background stars for the parametrs in Eq. (47). There are a factor of 10 more stars (10000) in the background than in the cluster (1000), but the average density of cluster and background stars is similar.

A further complication is that we have considered the rate of events in the data as products of some analysis pipeline. This rate may be different from the physical rate of interest, such as the rate of compact-binary mergers per unit time per unit volume which generate gravitational waves, or the physical numbers of stars in the cluster and field populations which produce the observed luminosities. Again, the conversion between the two will depend on the details of the data-analysis algorithm and ranking statistic, including any selection effects, and would need to be determined on a case-by-case basis.

Furthermore, in a practical application there could be multiple classes of events, not just foreground and background. For example, we are not necessarily interested in the rate of gravitational-wave signals per se, but separately in the rate of signals from mergers of binary neutron stars and binary black holes – populations that may sometimes be difficult to distinguish. Our approach is readily extendable to this particular complication, however. Note that it is symmetric with respect to foreground and background events (as expected, since one physicist's background is another physicist's foreground). We could relabel foreground and background events into other competing event classes, and further classes could be added in a straightforward way. However, the ability to distinguish classes relies on different distributions of their statistics. In general, rankings may need to be extended to include other statistics in addition to the signal "loudness" statistic in order to indicate both event significance and the probability of event attribution to a particular class.

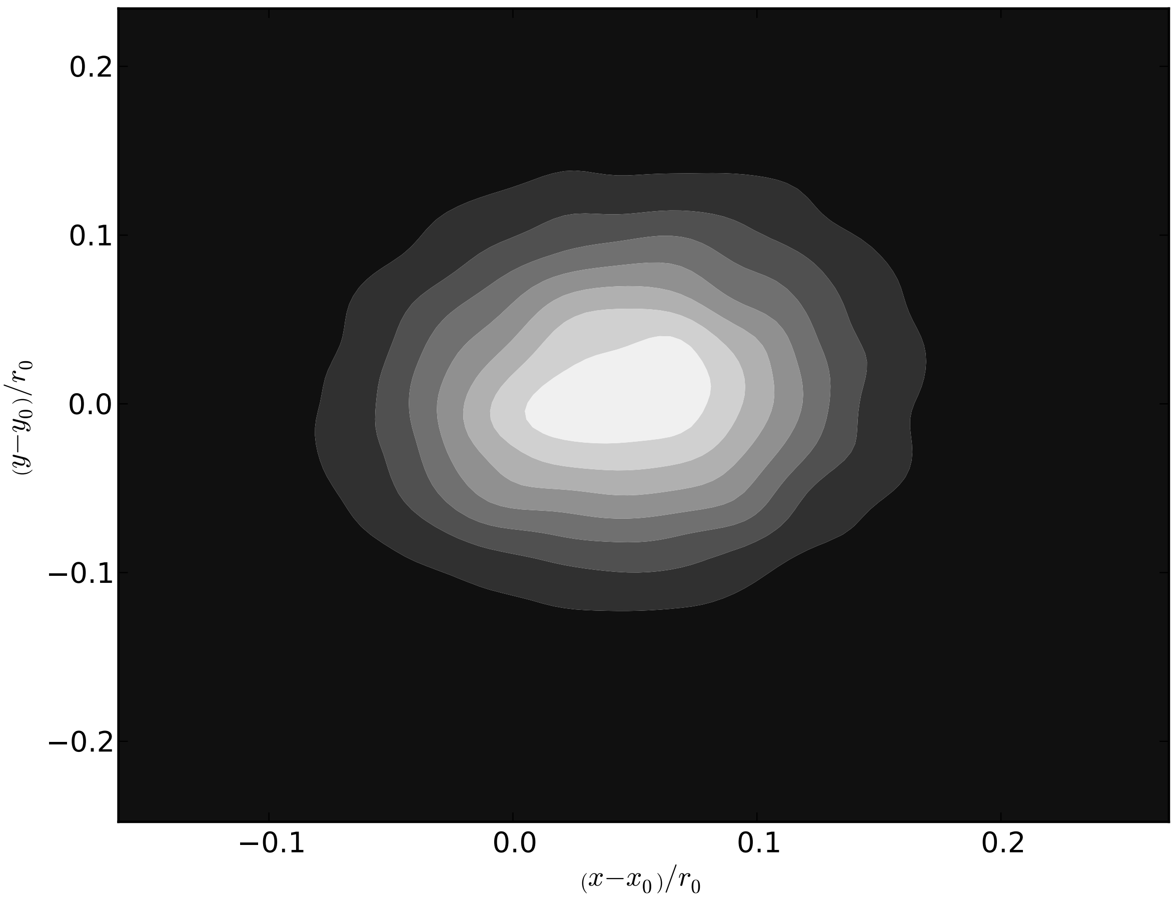

FIG. 12. Contours of the posterior probability distribution for the center of the cluster, $\vec{x}_0$, in the example from § III D. The center $(x, y) = (x_0, y_0)$ is determined to within about 5% of the structural radius of the cluster, r_0 (see Eq. (47)).

ACKNOWLEDGMENTS

We thank Kipp Cannon, Chad Hanna, Drew Keppel, and Richard O'Shaughnessy for discussions and suggestions about this manuscript. IM and WMF thank KITP.

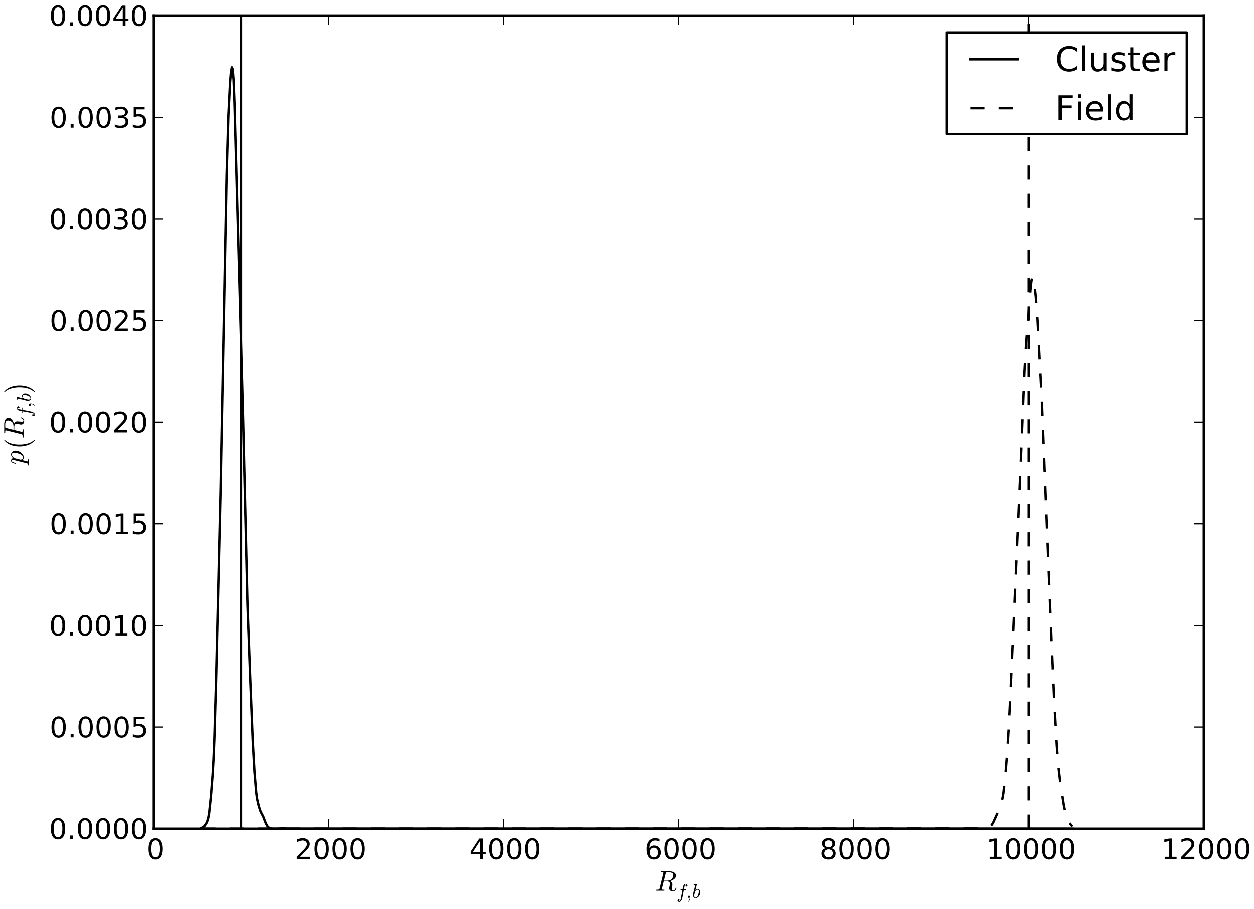

FIG. 13. Posterior densities for the number of stars in the cluster (R_f) and in the field (R_b) in the example from § III D. Vertical lines indicate the true values (see Eq. (47)).

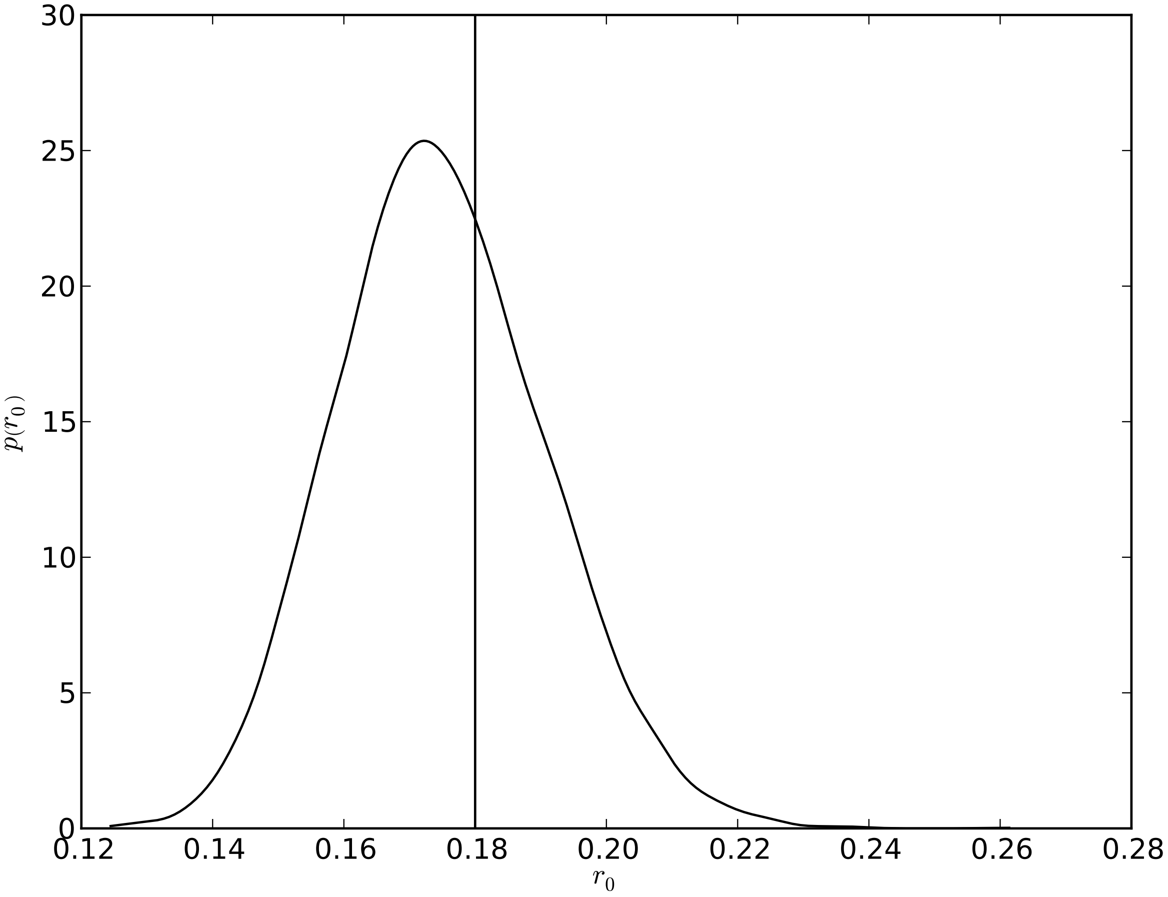

FIG. 14. Posterior density for the scale parameter for the cluster, r_0, from the example in § III D. The true value is indicated by the vertical line (see Eq. 47). **[Perhaps combine Figs. 8 and 9 (put an extra scale for r_0 at the top of 8?]**