

Rate Estimation

Will, Jon and Ilya
(Dated: March 24, 2012)

Some notes on the discussion Will and I had today (March 23rd 2012).

I. RATE ESTIMATION

A. Likelihood

The data we observe is a sequence of N events with statistic values x_i , above some threshold x_0 . These events could be drawn from a background described by the rate distribution $\nu_b(x)$, which specifies the rate of events arriving with statistic value greater than x , or from a foreground described by the rate distribution $\nu_f(x)$. We use flags $f_i \in \{0, 1\}$ to denote if an event is foreground ($f_i = 1$) or background ($f_i = 0$). We want to use the data to estimate the f_i 's and some property of the rate distributions (we'll come back to this later).

The probability of seeing data $x = \{x_i\}$ and flags f_i given the rates is the probability that we would draw events $\{x_j : f_j = 1\}$ from the foreground distribution and the events $\{x_j : f_j = 0\}$ from the background distribution. The probability of getting a set of M events $\{y_i\}$ above threshold y_0 and with $y_1 < y_2 < \dots < y_M$ from a Poisson process with loudness distribution $\nu(y)$ is the probability of having 0 events between y_0 and y_1 , which is $e^{-(\nu(y_0) - \nu(y_1))}$, times the probability of having an event at y_1 , which is $-\nu'(y_1)$, times the probability of having no events between y_1 and y_2 , which is $e^{-(\nu(y_1) - \nu(y_2))}$, etc. Overall we find

$$p(\{y_i\}) = e^{-\nu(y_0)} \prod_{i=1}^M [-\nu'(y_i)]. \quad (1.1)$$

Combining the foreground and background distributions we then have

$$p(x, \{f_i\} | \nu_f, \nu_b) = e^{-\nu_f(x_0)} e^{-\nu_b(x_0)} (-1)^N \prod_{\{j:f_j=1\}} \nu'_f(x_j) \prod_{\{k:f_k=0\}} \nu'_b(x_k). \quad (1.2)$$

We want the posterior distribution for the parameters of the rate distributions and the f_i 's, which is

$$p(\nu_f, \nu_b, \{f_i\} | x) \propto p(x, \{f_i\} | \nu_f, \nu_b) p(\nu_f, \nu_b) \quad (1.3)$$

The last term is just the prior probability of the rate parameters.

To relate this to Ilya's parameterisation, we say that the loudness distribution is determined by an overall scale, λ , and a probability distribution of events $p(x)$ with loudness greater than x_0 , which we temporarily assume is known. Then we have

$$\frac{d\nu}{dx} = -\lambda p(x) \quad (1.4)$$

and $\nu(x_0) = \lambda$. The posterior probability then becomes

$$p(\lambda_f, \lambda_b, \{f_i\} | x) \propto p(\lambda_f, \lambda_b) e^{-\lambda_f} e^{-\lambda_b} \lambda_f^{N_f} \lambda_b^{N_b} \prod_{\{j:f_j=1\}} p_f(x_j) \prod_{\{k:f_k=0\}} p_b(x_k) \quad (1.5)$$

where N_f/N_b are the number of foreground/background events and p_f/p_b are the corresponding loudness probability distributions. This is similar to what Ilya was writing down, except possible for the λ_b and λ_f dependent terms.

We can marginalise this probability distribution over the f_i 's which gives

$$p(\lambda_f, \lambda_b | x) \propto p(\lambda_f, \lambda_b) e^{-\lambda_f} e^{-\lambda_b} \prod_{j=1}^N (\lambda_f p_f(x_j) + \lambda_b p_b(x_j)) \quad (1.6)$$

or we could marginalise over the rate parameters. This requires a specification of the prior on the rates. As an illustrative example, we can take a flat prior on the rate parameters and then we have

$$p(\{f_i\} | x) \propto N_f! N_b! \prod_{\{j:f_j=1\}} p_f(x_j) \prod_{\{k:f_k=0\}} p_b(x_k). \quad (1.7)$$

We can go further and work out the marginal distribution of a single f_i , but these distributions are not particularly pleasant. We can also look at 2D probabilities of pairs of f_i 's and would expect to see correlations. Thus, everything can be written down exactly, but given the number of possible combinations of f_i 's that we would have to evaluate, an MCMC will allow us to find the combined distribution more easily.

B. A few more thoughts

The framework above allows us to estimate the form of the rate functions as well, i.e., the $p_f(x)$ and $p_b(x)$, if we assume some model for them. If we parameterize $p_f(x)$ by parameters $\{\mu_f^i\}$ and parameterize $p_b(x)$ by parameters $\{\mu_b^i\}$ then the combined posterior becomes

$$p(\lambda_f, \lambda_b, \{\mu_f^i\}, \{\mu_b^i\}, \{f_i\} | x) \propto p(\lambda_f, \lambda_b, \{\mu_f^i\}, \{\mu_b^i\}) e^{-\lambda_f} e^{-\lambda_b} \lambda_f^{N_f} \lambda_b^{N_b} \prod_{\{j:f_j=1\}} p_f(x_j | \mu_f^i) \prod_{\{k:f_k=0\}} p_b(x_k | \mu_b^i) \quad (1.8)$$

so we can jointly recover the shape of the loudness distribution and the rate, if we assume some model. I think using the power-law or exponential models would be a good illustration even if it was not realistic for LIGO. However, we could also mimic what Drew et al. do. If we divide the statistic space up into bins, we can then model $p_f(x)$ to be flat in each of these bins. The number of free parameters, $\{\mu_f^i\}$, is then the number of bins used. This is essentially what they do when they bin the data and look at the distribution of events from injections and background (timeslides).

Going one step further, we can use the injection and timeslide data directly in this framework. We just assume that we have another set of N_I events, $\{w_i\}$, known to be drawn from the foreground distribution and a set of N_T events $\{z_i\}$ known to be drawn from the background distribution. These events were not drawn with the correct rate parameters, but a fixed number of events were computed and so they do not contribute to the λ_f and λ_b terms, but they do contribute an extra factor

$$\prod_{l=1}^{N_I} p_f(w_l | \mu_f^i) \prod_{m=1}^{N_T} p_b(z_m | \mu_b^i) \quad (1.9)$$

on the right hand side of Eq. (1.8). So, we could simultaneously fit for the shape of the background as well as the rates and the identification of individual sources. If we were using the binned model for the probability distributions we would presumably find that, if there were many more events in the timeslide and injection data set, this would reduce to the analysis that was described above which assumes the shape of the rate distributions are given by the distribution of those events. However, this framework would be more flexible in the limit where the amount of timeslide or injection data was limited.

Another thought was that the injection distribution might not reflect the astrophysical distribution of signals, in which case we might want to see what happens if we only use timeslide data and ignore injection data. Similarly, in the detection era we might not be able to estimate background very easily since there will be too many sources in the data set (we wish!). In that case maybe we would assume that foreground was reliable and background wasn't. All of this is easy enough within this framework, but we might want to think a bit more.