LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type **LIGO-T030090-00-Z** 2003/05/11

Notes on analysis in presence of variable foreground and background rates

Lee Samuel Finn

Distribution of this draft:

LIGO I Collaboration

California Institute of Technology
LIGO Project - MS 51-33
Pasadena CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project - MS 20B-145
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

WWW: http://www.ligo.caltech.edu/

Processed with LaTeX on 2003/05/11

Abstract

The S1 data set has strongly variable noise and calibration, leading to a non-stationary background rate and a non-stationary foreground rate even in the presence of a stationary gravitational wave source population. This variability should either be accommodated in the analysis or treated as a source of systematic error. In this note we first derive an analysis of event data that takes this variability into account. From this analysis we we allow ourselves to "forget" what we know of the variability, finding a statistical analysis that averages over the variability. From this latter analysis Iwe discover what averaging means and get a sense of the error committed when the bulk of the events are contributed by a single, short segment of data with a very high noise.

This is exactly the case that arise for the S1 data set and the SLOPE events: the detector noise, as measured by the background event rate, is extremely variable and all the recorded events occur in a data segment with exceptionally large noise. In this specific case, and more generally whenever the detector noise is highly variable and the analysis does not take that variability into account, the results are dominated by the noisiest data. In the case of a fixed threshold event trigger generator this significantly weakens the upper limit that can be set on the foreground event rate; more dangerously, in the case of a floating threshold event trigger generator (such as TFCLUSTERS and POWER) it can lead to an upper limit that is over-restrictive.

$Id: T030090.tex,v 1.4 2003/05/11 15:12:16 lsf Exp $

Contents

1 Introduction

The S1 data set has strongly variable noise and calibration. In the case of the SLOPE ETG, burst group data segment 21, which is less than 1% of the total duration of data analysed, contributes all of the observed events and determines the background rate as well. This note was provoked by the question "What errors do we committ if we treat segment 21 on par with all other data segments and set aside our knowledge that the all the observed events and all the background are contributed by this segment?"

Knowledge of the variability of the noise and calibration, which affect background and efficiency, can be accommodated in the statistical analysis of event data. Beginning from a statistical analysis that takes this variability into account we allow ourselves to "forget" what we know, finding a statistical analysis that averages over the variability. In the process we discover what averaging means and get a sense of the error committed when the bulk of the events are contributed by a single, short segment of data with a very high noise.

2 Analysis in the presence of variability

A data processing pipeline produces candidate gravitational wave events. Some fraction of these are foreground events (i.e., of gravitational wave origin), the remaining are background events. In a fixed threshold ETG like SLOPE the background event rate depends on the threshold and the noise. In an interpreted analysis the number of foreground events depends on the constant rate of source events and the efficiency, which is a function of the noise level, the calibration, and the source event characteristics.

Consider the analysis of a data stream of livetime T, during which the detector noise and calibration are varying. Partition the livetime T into a set of M shorter epochs during which the detector calibration and noise are approximately constant. Introduce the following nomenclature

$$\mathcal{K}_j = \left(\begin{array}{l}\text{noise and calibration information for epoch } j\\ \text{which determines the expected background } \mu_{B,j}\\ \text{and detection efficiency } \epsilon_j \text{ in the interval}\end{array}\right) \tag{1a}$$

$$\Delta_j = \left(\text{ Duration of interval } j \right) \tag{1b}$$

$$\mu_{B,j} = \left(\text{ Expected number of background events in interval } j \right) \tag{1c}$$

$$\dot{n}_S = \left(\text{ source event rate } \right) \tag{1d}$$

$$\epsilon_j = \left(\begin{array}{l}\text{efficiency — i.e., the probability that a source}\\ \text{event becomes a foreground event — in the interval } j\end{array}\right) \tag{1e}$$

Note that the expected number of foreground events in the interval Δ_j depends on the source rate and the efficiency:

$$\mu_{F,j} = \left(\text{ Expected number of foreground events in interval } j \right) \tag{1f}$$

$$= \epsilon_j \Delta_j \dot{n}_S \tag{1g}$$

Consider an observation of N events and suppose we know how many events we observe in each interval Δ_j,

$$N_j = \left(\text{ Number of events observed in interval } j \right) \tag{2}$$

$$N = \sum_{j=1}^{M} N_j. \tag{3}$$

The probability of the particular observation $\mathcal{N}$,

$$\mathcal{N} = \{N_k : k = 1 \ldots M\}, \tag{4}$$

is

$$P(\mathcal{N}|\mathcal{K}) = \prod_{j=1}^{M} P(N_k|\mu_k) \tag{5a}$$

$$P(n|\mu) = \frac{\mu^n}{n!}e^{-\mu} \tag{5b}$$

$$\mu_k = \mu_{B,k} + \mu_{F,k}. \tag{5c}$$

From the observation $\mathcal{N}$ and knowledge of the $\mu_{B,k}$ and ϵ_k the likelihood in equation 5a is a function of $\dot{n}_S$ and we can determine bounds on $\dot{n}_S$ by any of the usual techniques.[1]

3 Ignoring noise and calibration variability

Now let's suppose that we forget (or never knew), in which interval each of the N events took place: i.e., we know only the total number of observed events N in the total observation of livetime T. Now the likelihood is

$$P(N|\mathcal{K}) = \sum_{N_1=0}^{M}\sum_{N_2=0}^{M-N_1}\sum_{N_3=0}^{M-N_1-N_2}\cdots\prod_{j=1}^{M} P(N_k|\mu_k) \tag{6a}$$

$$= \left[\sum_{N_1=0}^{M}\sum_{N_2=0}^{M-N_1}\sum_{N_3=0}^{M-N_1-N_2}\cdots\frac{\mu_k^{N_k}}{N_k!}\right]e^{-\sum_k \mu_k} \tag{6b}$$

$$= \frac{\left(\sum_k \mu_k\right)^N}{N!}e^{-\sum_k \mu_k} \tag{6c}$$

$$= P(\mu|N), \tag{6d}$$

where

$$\mu = \sum_{k=1}^{M} \mu_k \tag{6e}$$

$$= \sum_{k=1}^{M} \left(\mu_{B,k} + \mu_{F,k}\right). \tag{6f}$$

Thus if we ignore arrival time we return to a Poisson distribution, where the expected number of events is the time-averaged background rate times the live-time:

$$\mu_B = \sum_{k=1}^{M} \mu_{B,k} \tag{7a}$$

$$= T\sum_{k=1}^{M} \frac{\Delta_k}{T}\frac{\mu_{B,k}}{\Delta_k}. \tag{7b}$$

[1] The usual techniques, however, don't include classical confidence interval or upper limit constructions, which require that the observation be one-dimensional: e.g., a total number of observed events, not a number of observed events in each of several intervals.

What about our estimate of the source rate $\dot{n}_S$? From knowledge of μ_B and our observation N we have a bound on μ_F, where

$$\mu_F = \sum_{k=1}^{M} \mu_{F,k}. \tag{8a}$$

Recall that

$$\mu_{F,k} = \epsilon_k \Delta_k \dot{n}_S. \tag{8b}$$

Hence

$$\mu_F = \sum_{k=1}^{M} \mu_{F,k} \tag{8c}$$

$$= \dot{n}_S \sum_{k=1}^{M} \epsilon_k \Delta_k \tag{8d}$$

or

$$\dot{n}_S = \frac{\mu_F}{T\overline{\epsilon}} \tag{8e}$$

where $\overline{\epsilon}$ is the time-averaged efficiency

$$\overline{\epsilon} = \sum_{k=1}^{M} \frac{\Delta_k}{T} \epsilon_k. \tag{8f}$$

4 Discussion

What does all this mean? In the case of the SLOPE events in the S1 data set we have an exceptionally high background in one segment, which is also the segment where all our final events arise. Let's do a two segment example, modeled on the S1 data set. Let the first segment be 99.5% of the livetime with an expected background of 0.05 events and unit efficiency and let the second segment be 0.5% of the livetime with an expected background of 2 events and also unit efficiency. Suppose our observation is of 5 events, but all occured in the second segment. What is the bound on the rate?

First consider what happens if we keep in mind where the events came from. In a Frequentist analysis we find that there is no good answer to the question of what the foreground rate is: i.e., the first thing we discover is that the observation is so anomalous that we have to regard it as a statistical fluke. We can ask within what interval 90% of the likelihood lies: in this case, 90% of the likelihood lies within a rate of $2.33/T$, where T is the total duration of the observation. (This would also be the Bayesian upper limit assuming a flat prior; however, that is the answer to a different question.)

If we forget the when of the 5 observed events, then we have 5 observed events in a background of 2.05 events and 90% of the likelihood lies in the interval $[0.18, 7.4]/T$. (Again, this would be the smallest Bayesian credible set with a flat prior.) The great difference between these two analyses should be a warning that the approximations involved in second are poor ones.

Finally, if we ignore the noisy second segment entirely, 90% of the likelihood lies within a rate of $2.31/T$, where T is the livetime (now in the absence of the second segment). The difference between 2.31 and 2.33 is closer to what we expect: our upper limit should not be changed by a very short (0.5% of the livetime) but noisy segment of data that is, viewed by itself, not too unusual.

TFCLUSTERS and POWER both adjust their threshold to maintain an approximately constant background rate. Since the detector noise varies strongly over the S1 data set this variable threshold guarantees a variable efficiency, with lower than average efficiency where the noise amplitude is high and higher than average efficiency where the noise amplitude is low. Focus again on a two segment example. Let each segment be 50% of the data and have the same background expectation of 1 event. Let the first segment have unit efficiency and the second have an efficiency of 0.5. Suppose our observation is of 3 events, but all occured in the second segment. What is the bound on the rate?

Keeping track of the when of the events, 90% of the likelihood lies within a rate of $6.6/T$. If we forget the when of the events, however, then we find that 90% of the likelihood lies within a rate of $6.2/T$, which is — knowing that the second analysis is ignoring an important feature of the data that the first analysis takes account of — an overly strong conclusion to draw from the data.

The SLOPE example is very close to the actual case of the S1 data: the analysis ignores that variability of the detector noise and the rate derived from SLOPE events is dominated by segment 21, which is approximate 0.5% of the data and where all 5 observed events occurred though only 2 were expected from the estimated background. The variable threshold example is contrived: the burst group did not calculate the efficiency in different data segments.[2]

5 Conclusion

In the particular case of SLOPE events in the S1 data set the burst group analysis ignores the very evident strong variability of the detector noise and calibration, leading to a result that differs significantly from an analysis that takes account of that variability. An analysis based on an average background rate (and presumably, average efficiencies) applied to the S1 SLOPE events is, in any meaningful sense of the word, incorrect.

Data do not "speak for themselves": what light they shed on nature is observed through the prism of statistical analysis. While always true, this observation is especially important in the limit of weak signals in strong noise, or low foreground and high background rates. Incorrect results are readily obtained if the analysis does not correctly reflect the experimental reality. Ignoring the relationship of the statistic analysis to the data imperils the work of the collaboration: at the very least it can lead to results that are overly weak and thus can rightly be said to waste the potential of the data; of even greater concern is the possibility that an ill-considered analysis can lead to truly incorrect conclusions.

[2] Nor did it calculate the actual average efficiency: instead, it calculated an average over a select subset of data segments. It is not known how representative those subsegments are.