Detection statistic for multiple algorithms, networks and data sets.

S.Klimenko, C.Pankow

LIGO note T-1200169

1. Introduction

A central problem of any counting experiment is the estimation of the probability distribution function of false events. It is common to approximate it by the Poisson distribution with a mean defined by an underlying background process. The Poisson mean is estimated with the simulation of the background process or by isolation of a sample of events which is representative of the background process. In the gravitational wave searches with networks of detectors, the former is not practically possible due to a non-stationary (and often unknown) behavior of the detector noise. Therefore the latter method is used by generating a background sample with the time-shift analysis **??**. The counting rates, both for signal and noise, depend on the experimental selection cuts, which are very desirable to set prior to looking at the foreground data. It is also important to rank events in the observed data samples. Such ranking statistics (denoted below as ρ) are usually produced by the search algorithms and used to control the false alarm rate (FAR) and the search sensitivity to the expected GW signals. By measuring rates and detection efficiency as a function of ρ, a well defined statistical procedure can be used to answer the following questions:

- How to estimate a significance of observed foreground events.
- How to estimate astrophysical rates in case of detection.
- How to set upper limits on astrophysical rates in case of no detection.

This procedure has been already addressed in the literature **??** and in Section 2.1 we describe in details its frequentist version.

However the problem becomes more challenging if several searches are conducted and they are used to produce a combined result. The purpose of this note is to address a problem common both for the burst and inspiral all-sky searches: - define a statistical procedure to combine multiple searches with different network configurations and data sets into a single measurement. This problem has been partially addressed by introducing a fixed threshold statistic, which was used for calculation of upper limits in the combined S5 burst search (P.Sutton, [1]). A more general likelihood ratio ranking statistic is under development (R.Vaulin et al [2]). It addresses the calculation of upper limits, but it is not clear yet how the significance and the rate posterior distribution can be calculated. Below we describe the Inverse False Alarm rate Density (IFAD) statistic which adresses all three questions listed above. It is designed for the all-sky searches and provides a simple method for combination of results obtained with multiple searches.

2. Search for binary sources

2.1. Single search

Lets consider first a single search for binary sources with one network and one data set of duration T corresponding to a well defined state of instruments (single data epoch). The search produces a detection statistic ρ which is used to rank both the foreground and background events. The significance of triggers is defined by their Inverse False Alarm Rate (IFAR) estimated from the time shift analysis. The sensitivity of the search (at a given threshold on ρ) is characterized by a detection volume V estimated for a given population of simulated events. The Poisson probability that n loudest foreground events may be produced by the background process is

$$P_s(n, T_n) = 1 - \sum_{i=0}^{i=n-1} \exp\left(-T/T_n\right) \frac{(T/T_n)^i}{i!} \tag{1}$$

where T_n is the IFAR of the least significant event (out of n events) with the detection statistic ρ_n. If a detection statement is made, the rate posterior probability is

$$P_r(n, R) = \frac{(RTV_n)^n}{n!} \exp\left(-RTV_n\right) \tag{2}$$

where R is the estimator of the astrophysical rate and the V_n is the search volume estimated at the threshold ρ_n. If no detection statement is made, the rate upper limit is set from the loudest event statistic [3]

$$R_{90\%} = \frac{2.3}{V_1 T} \tag{3}$$

where V_1 is the search volume associated with the loudest event.

2.2. Multiple searches

A more complicated case is when several searches are performed. It can be one or several algorithms applied to different data sets and networks with different strain sensitivities, state of detectors and their false alarm rates (different data epochs). Given an algorithm applied to a selected data epoch we refer to this as a single search. Lets consider several searches that need to be combined in a single measurement. To do this, the events in different searches need to be ranked agains each other. This problem can be solved with the IFAD ranking statistic, which is introduced below.

Given an event i from a search j, it is characterized by its IFAR T_{ij} and by the search volume V_{ij}. For simplicity we assume that searches do not overlap in time. The product $T_{ij}V_{ij}$ is the Inverse False Alarm rate Density (IFAD) estimator, which can be used as the universal rainking statistic. Namely, given the values of this statistic for two arbitrary events (possibly from different searches), we can say which event is more significant. Respectively, the False Alarm rate Density (FAD) estimator is

$$r_b = [T_{ij}V_{ij}]^{-1}. \tag{4}$$

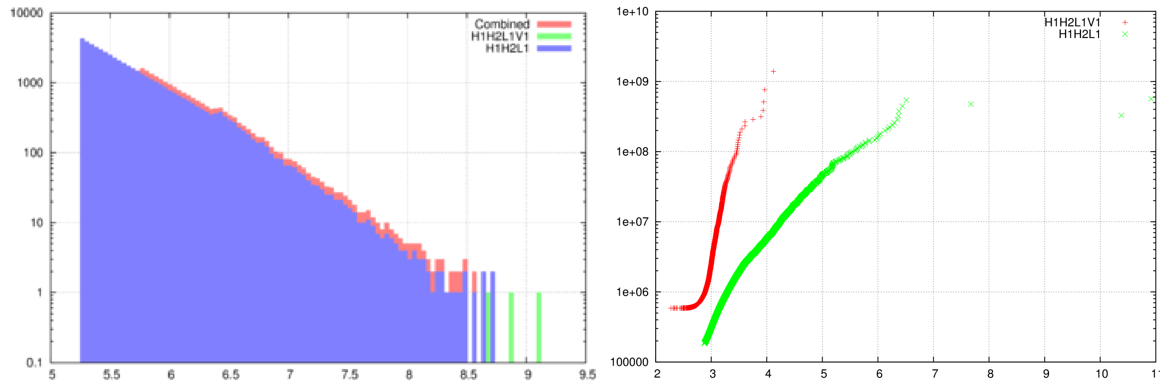

Figure 1. Left plot: number of background event (vertical axis) vs $\log_{10}(r_b^{-1})$ (horizontal axis) for two network configurations: L1H1H2 (blue), L1H1H2V1 (green), combined (red). Right plot: dependence r_b^{-1} vs ρ for L1H1H2 (green) and L1H1H2V1 (red).

For example, Figure 2 shows a distribution of the r_b^{-1} statistic for the background events from the IMBH search. There is an obvious problem with the proposed IFAD statistic (the same as for the likelihood ranking statistics). As seen from the Figure 2, the r_b^{-1} may not be a monotonous function of the search statistic ρ and therefore it may change ranking of events based on ρ. However, lets note that this is not the only possible FAD estimator. A more conservative and robust estimator is

$$r_b[i,j] = T_b^{-1} \sum_{\rho_n > \rho_i} \frac{1}{V_{nj}}. \tag{5}$$

where T_b is the accumulated background live time and the sum is taken over all background events louder than the event i. As Figure **??** shows this statistic is a monotonous function of ρ and events in all searches can be ranked by their IFAD: $r_b^{-1}[n+1] < r_b^{-1}[n]$, where n is the rank of events in the combined event set. Below we omit the explicit dependence of r_b on the rank n.

Given a threshold on r_b, the time-volume productivity of each search k can be calculated

$$\nu_k = V_k(r_b)T_k \tag{6}$$

where the V_k is the search volume and the T_k is the observation time. The ν_k a measure of the integrated sensitivity of the search at a given false alarm rate. The time-volume productivity of the combined search is given by the sum over all searches: $\nu = \sum \nu_k$. It determines the number of expected detections ($R\nu$, where R is the astrophysical rate) and the contribution of the background ($r_b\nu$).

The significance of n loudest events observed in the foreground sample is given by the Poisson probability

$$P_s(n) = 1 - \sum_{i=0}^{i=n-1} \exp\left(-r_b\nu\right) \frac{r_b^i \nu^i}{i!}. \tag{7}$$

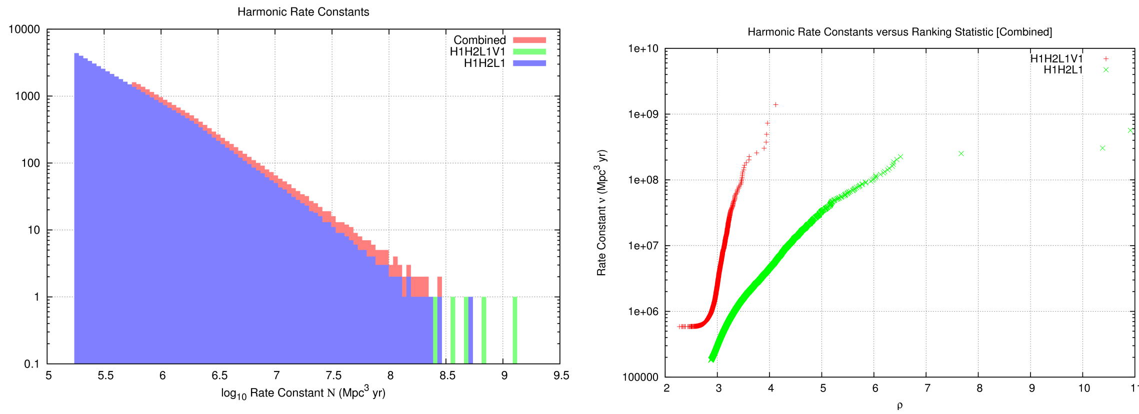

Figure 2. Corrected IFAD statistic. Left plot: number of background event (vertical axis) vs $\log_{10}(r_b^{-1})$ (horizontal axis) for two network configurations: L1H1H2 (blue), L1H1H2V1 (green), combined (red). Right plot: dependence r_b^{-1} vs ρ for L1H1H2 (green) and L1H1H2V1 (red).

If a detection statement is made, the rate posterior probability is

$$P_r(n, R) = \frac{R^n \nu^n}{n!} \exp(-R\nu). \tag{8}$$

For no detection statement, the rate upper limit is estimated from the loudest event statistic

$$R_{90\%} = \frac{2.3}{\nu[1]}. \tag{9}$$

where $\nu[1]$ is the time-volume productivity of the combined search calculated for the highest rank event in the foreground data set.

Note, applied to a single search this procedure results in a more conservative detection statement and the same rate upper limit as in Section 2.1 In case of several searches with very different properties, the combined measurement is dominated by the search with the largest integrated time-volume product ν_k. In principle, all searches which have a significant ν_k product should be included into the analysis. Events from all searches are ranked by their IFAD statistic which weights out less sensitive searches with the large false alarm rates.

3. References

1. P.Sutton, arXiv:0905.4089v2 [physics.data-an] 7 apr 2010
2. R. Vaulin, Likelihood method for trigger combination, LIGO note G1000330
3. R.Biswas et al, arXiv:0710.0465v2 [gr-qc] 20 oct 2007