LIGO Technical Report T020064-00-Z

Optimizing thresholds for rate estimation using coincidence in detectors of unequal sensitivity

Lee Samuel Finn

Penn State

Abstract

Both the burst and inspiral LIGO analyses will use coincident events in two or three detectors to estimate a source rate above a background. In these notes we describe the optimization problem involved in setting the threshold for identifying events in each detector, prior to determination of the coincident events.

I. INTRODUCTION

Analysis of gravitational wave data in the LIGO detector will make extensive use of "coincidence": candidate gravitational wave events are required to be observed separately in each detector with "start times" that differ by no more than the light travel time between the detectors (and the uncertainty in the definition of the event start time). The coincidence procedure may also involve other cuts on the events registered independently at each detector: e.g., their relative amplitude, spectrum, etc. The generation of the list of events in each detector will involve a set of cuts on the data that are unique to each detector, its sensitivity, and the types of analyses undertaken on the detector data. None of these need be the same at each detector. The coincidence step itself will also involve a set of cuts on the relative characteristics of different events.

Some of the cuts ρ_k at each detector reduce the detector's "live time": i.e., the integrated duration of the data that is considered analyzable. For example, we may require of any analyzed data that the contemporaneous seismic noise in a given band at a site be below a certain threshold. Other cuts may affect the fraction of the M_k events identified at detector k that are expected to be contributed by an unavoidable background (i.e., the false alarm probability). These same cuts will also affect the fraction of actual gravitational wave events that will be missed by the detection procedure (i.e., the false dismissal probability, which is 1 minus the efficiency).

Our intuition tells us that, for the "best" analysis, the detailed choice of cuts must simultaneously maximize the live time and the efficiency, while minimizing the false rate. Here we make our intuition precise by addressing the question of how these cuts should be chosen to make best use of the data from the multiple detectors to determine an event rate.

II. NOMENCLATURE

a. The Detectors. Assume that we have N independent detectors. We assume nothing here about the similarity or differences between the detectors, but we do require that they operate independently of each other: i.e., the sensitivity of one detector does not affect the sensitivity of any other detector.

b. Events. At detector k an analysis characterized by a set of cuts ρ_k produces a set of events M_k over a live time T_k. Each event is characterized by at least a start time and perhaps by other information. A coincidence procedure, which operates on the events in the different detectors and their start times, introduces its own set of cuts ρ_{N+1} and produces a final set of M events over a live time T. We identify the collection of cuts — from the analyses at the individual detectors and the final coincidence step — by ρ.

c. Detector Event Model. The "events" identified at each detector arise from additive background and foreground components. The background component is assumed to be i.i.d. at each detector, and independent between detectors. The foreground component is assumed to be of gravitational wave origin and, thus, correlated between detectors with a time-lag between detectors consistent with a single plane wave incident on the detector array.

Note that, because the background events are independent and identically distributed (i.i.d.) the background in each detector is a Poisson process. We also assume that the foreground contribution arises from a Poisson process, with an intrinsic rate θ.

d. Detector Sensitivity. The efficiency of detector k to the foreground is $\epsilon_k(\rho_k)$. (The efficiency ϵ_k obviously depends on the cuts ρ_k.) *Setting aside the uncertainty in the assignment of foreground start time* the overall efficiency of the analysis pipeline is thus

$$\epsilon = \prod_k \epsilon_k(\rho_k). \tag{2.1}$$

III. THE OPTIMIZATION PROBLEM

A. Bayesian or Frequentist?

In a Bayesian analysis, the goal is to update our prejudice about θ, represented by $P(\theta)$, based on our observation of the number of coincident events M in the observation time (i.e., the detector "live time") T. Our final degree of belief that θ takes on a particular value is given by the probability distribution $P\left(\theta|M,T,\rho,b\right)$, which is related to $P(\theta)$ by

$$P\left(\theta|M,T,\rho,b\right) \propto P\left(M|\theta,T,\rho,b\right) P(\theta) \tag{3.1}$$

where b is the background rate after coincidence.

In a Bayesian analysis all inferences are based on $P\left(\theta|M,T,\rho,b\right)$; on the other hand, in a frequentist analysis inferences are based on the sampling distribution $P\left(M|\theta,T,\rho,b\right)$, which is used in constructing $P(\theta|M, T, \rho, b)$. In either case we may choose a particular inference that we regard as critical and ask that the cuts ρ be chosen in such a way that some measure of strength of the inference is extremized. Thus, our first task, to which we devote the rest of this section, is to determine sampling distribution $P(M|\theta, T, \rho, b)$ and the background rate b.

B. The background rate b

Estimation of the background rate is difficult when the foreground cannot be "turned-off" or otherwise modulated independently of the background. *Under the assumption that the background in each detector is independent of the background in all other detectors,* and that foreground events are rare, we can effectively "turn-off" the foreground by adding a time delay to the different single-detector event lists such that no foreground event would pass our coincidence cut when the coincidence cut is carried out on the shifted-time detector event lists. Since the background is i.i.d. in each detector and independent between distinct detectors, the background event rate after coincidence is unaffected by the delay.[3]

C. Determining the coincidence sampling distribution

The known background is i.i.d. with an event rate b. We have assumed that the underlying source events that give rise to the foreground are i.i.d. with an unknown rate θ. The sampling distribution $P(M|\theta, T, \rho, b)$, which describes the probability that M coincident events are observed, is thus Poisson,

$$P(M|\theta, T, \rho, b) = \frac{\phi^M}{M!} e^{-\phi} \tag{3.2}$$

where the Poisson parameter ϕ is the product of the observation duration T and the sum of the background rate b and the rate of source events θ reduced by the efficiency of the coincidence procedure:

$$\phi = T(b + \theta\epsilon). \tag{3.3}$$

In the next section we explore a Bayesian analysis for the event rate and recommend a measure on the posterior probability distribution $P(\theta|M, \rho, T, b)$ that can be used to optimize the cuts ρ; in the subsection following the next we explore a Frequentist analysis and recommend a measure on the sampling distribution that can be used to optimize the cuts ρ.

IV. OPTIMIZING CUTS FOR A BAYESIAN ANALYSIS

We would like to choose a prior $p(\theta)$ that presumes as little as possible about the unknown parameter θ. Following [1] we interpret this to mean that the priori $P(\theta)$ should share the same symmetries as the sampling distribution. The sampling distribution is scale invariant on the Poisson parameter ϕ, which is $\epsilon\theta + b$, with b and ϵ known. Introducing the scale transformation

$$\phi' = \alpha\phi \tag{4.1a}$$

we thus require that the prior satisfy

$$P(\phi')d\phi' \; = \; P(\phi)d\phi; \tag{4.1b}$$

or

$$P(\alpha\phi)\alpha \; = \; P(\phi) \tag{4.1c}$$

$$P(\phi) \; \propto \; \phi^{-1} \tag{4.1d}$$

$$P(\theta) \; \propto \; (b + \epsilon\theta)^{-1} \, . \tag{4.1e}$$

As long as the expected number of background events Tb is non-zero the posterior is normalizable:

$$P(\theta|M, \epsilon, T, b) = \frac{\epsilon T \phi^{M-1}}{\Gamma(M, bT)} e^{-\phi} \tag{4.2}$$

for $\theta \geq 0$ where $\Gamma(M, Tb)$ is the incomplete gamma function

$$\Gamma(M, Tb) = \int_{Tb}^{\infty} x^{M-1} e^{-x} dx. \tag{4.3}$$

This posterior characterizes our degree of belief that θ takes on a particular value given an observation of M events over an integrated time T in our analysis chain characterized by an efficiency ϵ and a background rate b. In our analysis the efficiency, background rate and live time T are all functions of the cuts ρ.

Given an observation M the posterior describes our degree of belief that θ takes on a particular value. We can characterize the posterior by the mean and standard deviation of ϕ, which is linearly related to θ:

$$\mu_\phi := \overline{\phi} \; = \; \Gamma(M+1, Tb)/\Gamma(M, Tb) \tag{4.4}$$

$$\sigma_\phi^2 := \overline{\left[\left(\phi - \overline{\phi}\right)^2\right]} \; = \; \frac{\Gamma(M+2, Tb)}{\Gamma(M, Tb)} - \left(\frac{\Gamma(M+1, Tb)}{\Gamma(M, Tb)}\right)^2 . \tag{4.5}$$

Since ϕ is linearly related to θ we can express μ_θ and σ_θ^2 in terms of μ_ϕ and σ_ϕ^2:

$$\mu_\theta = \frac{1}{T\epsilon}\left(\frac{\Gamma(M+1,Tb)}{\Gamma(M,Tb)} - Tb\right) \tag{4.6a}$$

$$\sigma_\theta^2 = \frac{\sigma_\phi^2}{T^2\epsilon^2} \tag{4.6b}$$

To determine the "best" choice of cuts we establish a measure of "goodness" and then ask for the cuts that extremize the goodness. As an example, let us make our measure of goodness the mean μ_θ when there is no signal. For any given M we get a different upper limit; so, let us focus on the expectation value of μ_θ over an ensemble of observations where the event rate is Tb:

$$\overline{\mu_\theta} := \sum_{M=0}^{\infty} \mu_\theta(M,Tb)\frac{(Tb)^M}{M!}e^{-Tb} \tag{4.7}$$

$$= -\frac{b}{\epsilon} + \frac{1}{T\epsilon}\sum_{M=0}^{\infty}\frac{\Gamma(M+1,Tb)}{\Gamma(M,Tb)}\frac{e^{-Tb}}{M!} \tag{4.8}$$

The best choice of cuts is the choice that mimizes $\overline{\mu_\theta}$. Changes in the cuts can affect both the live time T and the background rate b. In this expression the choice of cuts enters through the live time T, the background rate b, and the efficiency ϵ. (Other choices can certainly be made; this is one, hopefully sensible, example.)

V. OPTIMIZING CUTS FOR A FREQUENTIST ANALYSIS

To determine the "best" choice of cuts in a Frequentist analysis we establish a measure of "goodness" and then ask for the cuts that maximize the goodness. As an example, let us focus on the concept of the unified upper limit/confidence interval on θ [2]. From the sampling distribution we can determine the upper limit on θ that is associated with an observation of M events. Now suppose that θ is actually zero. In an ensemble of observations the M observed are Poisson distributed with Poisson parameter Tb. Associated with each M is confidence interval and each of these intervals has a mid-point. Focus attention on the ensemble mean of these mid-points. Better analyses will have smaller ensemble means of the mid-point and we choose our cuts to minimize the expectation value of the mid-point of the unified upper limit/confidence interval. The mean associated with any given analysis will depend on the cuts through the live time T, the background rate b, and the efficiency

ϵ. (Other choices can certainly be made; this is one, hopefully sensible, example.)

[1] E. T. Jaynes, IEEE Transactions on Systems Science and Cybernetics **4**, 227 (1968).

[2] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).

[3] By choosing the delay small, but larger than the light travel time between the detectors, longer timescale non-stationarity in the background event rates are not an issue.