

Analyzing a Population Segment in Gravitational-Wave Catalogs

SOURMENDRA KISHORE ROY ^{1,2} LIEKE A. C. VAN SON ² AND WILL M. FARR ^{1,2}

¹*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*

²*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*

ABSTRACT

Maths for analyzing a population segment in gravitational-wave catalogs (this is not a paper)

Keywords: gravitational waves — statistics

1. NO SELECTION CUT & NO NOISE TRIGGER

Our gravitational-wave catalog is described by the strain time-series \vec{D} , the single event parameter θ , and the population hyperparameters $\vec{\lambda}$. Let's introduce a cut in the latent parameter space at $\theta = \theta_{cut}$. We aim to analyze the catalogs for $\theta < \theta_{cut}$ and $\theta > \theta_{cut}$ separately.

We divide the data \vec{D} into two parts: $\vec{D} = \bigcup_{j=1}^{N_{tr1}} d_{1j} \cup \bigcup_{k=1}^{N_{tr2}} d_{2k}$. For \vec{d}_1 , the 1-cdf of the parameter estimation (PE) posterior $P(\theta > \theta_0 | d_i) < p_0$ and for \vec{d}_2 , the 1-cdf $P(\theta > \theta_0 | d_i) > p_0$, based on a sufficiently small preassigned value p_0 . Correspondingly, the total number of triggers is the sum of the triggers found in \vec{d}_1 and \vec{d}_2 : $N_{tr} = N_{tr1} + N_{tr2}$.

The population is:

$$\frac{dN}{d\theta} = \frac{dN_1}{d\theta}(\lambda_1)\Theta(\theta_{cut} - \theta) + \frac{dN_2}{d\theta}(\lambda_2)\Theta(\theta - \theta_{cut}) \quad (1)$$

With fixed λ_1 , the posterior of λ_2 given N_{tr2} number of triggers in \vec{d}_2 :

$$P(\lambda_2 | \vec{d}_2, N_{tr2}, \lambda_1) = z^{-1} \pi(\lambda_2) \mathcal{L}(\vec{d}_2, N_{tr2} | \lambda_2, \lambda_1) \quad (2)$$

z is the evidence. In the data segment \vec{d}_2 , events can either appear from the $\theta < \theta_0$ region due to PE errors, or they originate from the $\theta > \theta_0$ region. Thus, the rate in the hierarchical likelihood \mathcal{L} , modeled by an inhomogeneous Poisson process (Mandel et al. 2019), has contributions from both the $\theta < \theta_0$ and $\theta > \theta_0$ regions:

$$\mathcal{L}(\vec{d}_2, N_{tr2} | \lambda_2, \lambda_1) = \frac{e^{-N_1(\lambda_1) - N_2(\lambda_2)}}{N_{tr2}!} \prod_{k=1}^{N_{tr2}} \int_{\theta} d\theta P(d_k | \theta) \frac{dN}{d\theta}(\lambda_1, \lambda_2) \quad (3)$$

With the population model (1), this simplifies to,

$$\mathcal{L}(\vec{d}_2, N_{tr2} | \lambda_2, \lambda_1) = \left(\prod_{k=1}^{N_{tr2}} \pi(d_k) \right) \frac{e^{-N_1(\lambda_1) - N_2(\lambda_2)}}{N_{tr2}!} \prod_{k=1}^{N_{tr2}} \left[\int_{-\infty}^{\theta_{cut}} d\theta P(\theta | d_k) \frac{\frac{dN_1}{d\theta}(\lambda_1)}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)} \right] \quad (4)$$

where $\pi(\theta)$ is the PE prior, $\pi(d_k)$ is the prior on the data, $N_1(\lambda_1)$ and $N_2(\lambda_2)$ are the number of events in the $\theta < \theta_0$ and $\theta > \theta_0$ regions, respectively:

$$N_1(\lambda_1) = \int_{-\infty}^{\theta_{cut}} d\theta P_{det}(\theta) \frac{dN_1}{d\theta}(\lambda_1), \quad N_2(\lambda_2) = \int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) \frac{dN_2}{d\theta}(\lambda_2) \quad (5)$$

soumendrakis.roy@stonybrook.edu

skishoreroy@flatironinstitute.org

lvanson@flatironinstitute.org

will.farr@stonybrook.edu

wfarr@flatironinstitute.org

- **Corollary:** If θ is measured perfectly for $\theta \in (-\infty, \theta_{cut}]$, $P(\theta > \theta_{cut}|d_i) > p_0$ never occurs. Hence, $P_{det}(\theta) = 0$. This results in

$$N_1(\lambda_1) = 0 \quad (6)$$

On the contrary, for $\theta \in [\theta_{cut}, \infty)$, $P(\theta > \theta_{cut}|d_i) > p_0$ always occurs. Hence, $P_{det}(\theta) = 1$. This implies,

$$N_2(\lambda_2) = \int_{\theta_{cut}}^{\infty} d\theta \frac{dN_2}{d\theta}(\lambda_2) = N_{tr2} \quad (7)$$

Putting (4) in (2), the full posterior is:

$$P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1) = z^{-1} \pi(\lambda_2) \left(\prod_{k=1}^{N_{tr2}} \pi(d_k) \right) \frac{e^{-N_1(\lambda_1) - N_2(\lambda_2)}}{N_{tr2}!} \prod_{k=1}^{N_{tr2}} \left[\int_{-\infty}^{\theta_{cut}} d\theta P(\theta|d_k) \frac{\frac{dN_1}{d\theta}(\lambda_1)}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)} \right] \quad (8)$$

Before proceeding further, let's define:

$$A_{1,k}(\lambda_1) = \int_{-\infty}^{\theta_{cut}} d\theta P(\theta|d_k) \frac{\frac{dN_1}{d\theta}(\lambda_1)}{\pi(\theta)}, \quad A_{2,k}(\lambda_2) = \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)} \quad (9)$$

Now with fixed λ_1 and λ_{20}

$$\begin{aligned} P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1) &\propto \frac{P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1)}{P(\lambda_{20}|\vec{d}_2, N_{tr2}, \lambda_1)} = \frac{\pi(\lambda_2)}{\pi(\lambda_{20})} \frac{e^{-N_2(\lambda_2)}}{e^{-N_2(\lambda_{20})}} \prod_{k=1}^{N_{tr2}} \left[\frac{A_{1,k}(\lambda_1) + A_{2,k}(\lambda_2)}{A_{1,k}(\lambda_1) + A_{2,k}(\lambda_{20})} \right] \\ &= \frac{\pi(\lambda_2)}{\pi(\lambda_{20})} \frac{e^{-N_2(\lambda_2)}}{e^{-N_2(\lambda_{20})}} \prod_{k=1}^{N_{tr2}} \left[\frac{A_{1,k}(\lambda_1)}{A_{1,k}(\lambda_1) + A_{2,k}(\lambda_{20})} + \frac{A_{2,k}(\lambda_2)}{A_{2,k}(\lambda_{20})} \frac{A_{2,k}(\lambda_{20})}{A_{1,k}(\lambda_1) + A_{2,k}(\lambda_{20})} \right] \end{aligned}$$

Keeping the λ_2 dependent terms,

$$P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1) \propto \pi(\lambda_2) e^{-N_2(\lambda_2)} \prod_{k=1}^{N_{tr2}} \left[(1 - p_{astro,m,k}) + \frac{A_{2,k}(\lambda_2)}{A_{2,k}(\lambda_{20})} p_{astro,m,k} \right] \quad (10)$$

Here, $p_{astro,m,k}$ is the probability that a certain trigger in \vec{d}_2 falls in $\theta > \theta_0$ region, given fiducial λ_1 and λ_{20} :

$$p_{astro,m,k} = \frac{A_{2,k}(\lambda_{20})}{A_{1,k}(\lambda_1) + A_{2,k}(\lambda_{20})} \quad (11)$$

- **Corollary:** For $p_{astro,m,k} \rightarrow 1$ (no contamination from $\theta < \theta_0$ region),

$$P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1) \propto \pi(\lambda_2) \exp\left(-\int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) \frac{dN_2}{d\theta}(\lambda_2)\right) \prod_{k=1}^{N_{tr2}} \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)} \quad (12)$$

In this case, we can also marginalize over the total rate R_2 [with $\frac{dN_2}{d\theta}(\lambda_2) = R_2 f_2(\theta|\lambda'_2)$], and focus on the shape parameters λ'_2 in the $\theta > \theta_0$ region,

$$P(\lambda'_2|\vec{d}_2, N_{tr2}, \lambda_1) = \int_{R_2} dR_2 P(\lambda_2|\vec{d}_2, N_{tr2}, \lambda_1)$$

For $\pi(R_2, \lambda'_2) = \frac{1}{R_2} \pi(\lambda'_2)$, it converges to a nice looking expression,

$$P(\lambda'_2|\vec{d}_2, N_{tr2}, \lambda_1) \propto \Gamma(N_{tr2}) \pi(\lambda'_2) \prod_{k=1}^{N_{tr2}} \left[\frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_{2k}) \frac{f_2(\theta|\lambda'_2)}{\pi(\theta)}}{\int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) f_2(\theta|\lambda'_2)} \right] \propto \pi(\lambda'_2) \prod_{k=1}^{N_{tr2}} \left[\frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_{2k}) \frac{f_2(\theta|\lambda'_2)}{\pi(\theta)}}{\int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) f_2(\theta|\lambda'_2)} \right]$$

Switching to more conventional symbols:

$$N_2(\lambda_2) = R_2 \overline{VT}(\lambda'_2), \quad \frac{A_{2,k}(\lambda_2)}{A_{2,k}(\lambda_{20})} = \frac{R_2}{R_{20}} w_{2k}(\lambda'_2, \lambda'_{20}) \quad (13)$$

The posterior in (10) becomes,

$$P(\lambda_2 | \vec{d}_2, N_{tr2}, \lambda_1) \propto \pi(\lambda_2) e^{-R_2 \overline{VT}(\lambda'_2)} \prod_{k=1}^{N_{tr2}} \left[(1 - p_{astro,m,k}) + \frac{R_2}{R_{20}} w_{2k}(\lambda'_2, \lambda'_{20}) p_{astro,m,k} \right] \quad (14)$$

We need to compute \overline{VT} , $w_{2k}(\lambda'_2, \lambda'_{20})$ and $p_{astro,m,k}$ to obtain (14).

$$\overline{VT} = \int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) f_2(\theta | \lambda_2) \Rightarrow \overline{VT} \approx \frac{1}{N_{draw}} \sum_{\substack{\theta \sim P(\theta|det, draw) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_2)}{p_{draw}(\theta)} \quad (15)$$

$$w_{2k}(\lambda'_2, \lambda'_{20}) = \frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{f_2(\theta | \lambda'_2)}{\pi(\theta)}}{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{f_2(\theta | \lambda'_{20})}{\pi(\theta)}} \Rightarrow \mathcal{W}_{2k}(\lambda'_2, \lambda'_{20}) \approx \frac{\frac{1}{N_{samp}} \sum_{\substack{\theta \sim P(\theta|d_k) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_2)}{\pi(\theta)}}{\frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta|d_k) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_{20})}{\pi(\theta)}} \quad (16)$$

$$\begin{aligned} p_{astro,m,k}(\lambda_1, \lambda_{20}) &= \frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}}{\int_{-\infty}^{\theta_{cut}} d\theta P(\theta | d_k) \frac{\frac{dN_1}{d\theta}(\lambda_1)}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}} \\ &\Rightarrow \mathcal{P}_{astro,m,k}(\lambda_1, \lambda_{20}) \approx \frac{\frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta|d_k) \\ \theta > \theta_{cut}}} \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}}{\frac{1}{N_{samp,1}} \sum_{\substack{\theta \sim P(\theta|d_k) \\ \theta < \theta_{cut}}} \frac{\frac{dN_1}{d\theta}(\lambda_1)}{\pi(\theta)} + \frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta|d_k) \\ \theta > \theta_{cut}}} \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}} \end{aligned} \quad (17)$$

2. WITH SELECTION CUT AND NOISE TRIGGER

Here, we select our data segment \vec{d}_2 based on two conditions:

1. **det1:** The false alarm rate (FAR) of a trigger is less than a preassigned value (FAR_0) which is sufficiently large.
2. **det2:** The $1 - CDF$ of the PE posterior $P(\theta > \theta_0 | d_i)$ is greater than a preassigned value p_0 , where p_0 is sufficiently small.

The posterior distribution of λ_2 given the number of triggers N_{tr2} in the strain time series \vec{d}_2 , with λ_{10} fixed,

$$P(\lambda_2 | \vec{d}_2, N_{tr2}, \lambda_{10}) \propto \pi(\lambda_2) \mathcal{L}(\vec{d}_2, N_{tr2} | \lambda_2, \lambda_{10}) \quad (18)$$

\mathcal{L} is modeled by an inhomogeneous Poisson process that includes both foreground and background components (Roulet et al. 2020),

$$\mathcal{L}(\vec{d}_2, N_{tr2} | \lambda_2, \lambda_{10}) = e^{-N_a(\lambda_2, \lambda_{10})} \prod_{k=1}^{N_{tr2}} \left[\frac{\frac{dN_a}{dd_k}(\lambda_2, \lambda_{10})}{\frac{dN_a}{dd_k}(\lambda_{20}, \lambda_{10})} p_{astro,k}(\lambda_{20}, \lambda_{10}) + (1 - p_{astro,k}(\lambda_{20}, \lambda_{10})) \right]$$

For the population

$$\frac{dN}{d\theta} = \frac{dN_1}{d\theta}(\lambda_{10}) \Theta(\theta_{cut} - \theta) + \frac{dN_2}{d\theta}(\lambda_2) \Theta(\theta - \theta_{cut})$$

the expected number of astrophysical triggers is divided into two integrals:

$$N_a(\lambda_2, \lambda_{10}) = \int_{-\infty}^{\theta_{cut}} d\theta P_{det1,det2}(\theta) \frac{dN_1}{d\theta}(\lambda_{10}) + \int_{\theta_{cut}}^{\infty} d\theta P_{det1,det2}(\theta) \frac{dN_2}{d\theta}(\lambda_2)$$

and the ratio:

$$\frac{\frac{dN_a}{dd_k}(\lambda_2, \lambda_{10})}{\frac{dN_a}{dd_k}(\lambda_{20}, \lambda_{10})} = \frac{\int_{-\infty}^{\theta_{cut}} d\theta P(\theta|d_k) \frac{\frac{dN_1}{d\theta}(\lambda_{10})}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)}}{\int_{-\infty}^{\theta_{cut}} d\theta P(\theta|d_k) \frac{\frac{dN_1}{d\theta}(\lambda_{10})}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}} = \frac{A_{1k}(\lambda_{10}) + A_{2k}(\lambda_2)}{A_{1k}(\lambda_{10}) + A_{2k}(\lambda_{20})}$$

where

$$\begin{aligned} A_{1k}(\lambda_{10}) &= \int_{-\infty}^{\theta_{cut}} d\theta P(\theta|d_k) \frac{\frac{dN_1}{d\theta}(\lambda_{10})}{\pi(\theta)} \\ A_{2k}(\lambda_{20}) &= \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)} \\ A_{2k}(\lambda_2) &= \int_{\theta_{cut}}^{\infty} d\theta P(\theta|d_k) \frac{\frac{dN_2}{d\theta}(\lambda_2)}{\pi(\theta)} \end{aligned}$$

$\frac{dN_a}{dd_k}(\lambda_2, \lambda_{10}) / \frac{dN_a}{dd_k}(\lambda_{20}, \lambda_{10})$ can be further simplified to:

$$\frac{\frac{dN_a}{dd_k}(\lambda_2, \lambda_{10})}{\frac{dN_a}{dd_k}(\lambda_{20}, \lambda_{10})} = \frac{A_{2k}(\lambda_2)}{A_{2k}(\lambda_{20})} p_{astro,m,k}(\lambda_1, \lambda_{20}) + (1 - p_{astro,m,k}(\lambda_1, \lambda_{20}))$$

with

$$p_{astro,m,k}(\lambda_1, \lambda_{20}) = \frac{A_{2k}(\lambda_{20})}{A_{1k}(\lambda_1) + A_{2k}(\lambda_{20})}$$

Putting everything together, the posterior in equation (18) becomes:

$$\begin{aligned} P(\lambda_2 | \vec{d}_2, N_{tr2}, \lambda_1) &\propto \pi(\lambda_2) e^{-R_2 \overline{VT}(\lambda'_2)} \prod_{k=1}^{N_{tr2}} \left[\left(\frac{R_2}{R_{20}} w_{2k}(\lambda'_2, \lambda'_{20}) p_{astro,m,k}(\lambda_{10}, \lambda_{20}) \right. \right. \\ &\quad \left. \left. + (1 - p_{astro,m,k}(\lambda_{10}, \lambda_{20})) p_{astro,k}(\lambda_{10}, \lambda_{20}) + (1 - p_{astro,k}(\lambda_{10}, \lambda_{20})) \right) \right] \end{aligned} \quad (19)$$

To compute (19), we need 4 quantities:

$$\overline{VT} = \int_{\theta_{cut}}^{\infty} d\theta P_{det}(\theta) f_2(\theta | \lambda_2) \Rightarrow \overline{VT} \approx \frac{1}{N_{draw}} \sum_{\substack{\theta \sim P(\theta | \det1, \det2, draw) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_2)}{p_{draw}(\theta)} \quad (20)$$

$$w_{2k}(\lambda'_2, \lambda'_{20}) = \frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{f_2(\theta | \lambda'_2)}{\pi(\theta)}}{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{f_2(\theta | \lambda'_{20})}{\pi(\theta)}} \Rightarrow \mathcal{W}_{2k}(\lambda'_2, \lambda'_{20}) \approx \frac{\frac{1}{N_{samp}} \sum_{\substack{\theta \sim P(\theta | d_k) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_2)}{\pi(\theta)}}{\frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta | d_k) \\ \theta > \theta_{cut}}} \frac{f_2(\theta | \lambda_{20})}{\pi(\theta)}} \quad (21)$$

$$\begin{aligned} p_{astro,m,k}(\lambda_{10}, \lambda_{20}) &= \frac{\int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}}{\int_{-\infty}^{\theta_{cut}} d\theta P(\theta | d_k) \frac{\frac{dN_1}{d\theta}(\lambda_{10})}{\pi(\theta)} + \int_{\theta_{cut}}^{\infty} d\theta P(\theta | d_k) \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}} \\ &\Rightarrow \mathcal{P}_{astro,m,k}(\lambda_{10}, \lambda_{20}) \approx \frac{\frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta | d_k) \\ \theta > \theta_{cut}}} \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}}{\frac{1}{N_{samp,1}} \sum_{\substack{\theta \sim P(\theta | d_k) \\ \theta < \theta_{cut}}} \frac{\frac{dN_1}{d\theta}(\lambda_{10})}{\pi(\theta)} + \frac{1}{N_{samp,0}} \sum_{\substack{\theta \sim P(\theta | d_k) \\ \theta > \theta_{cut}}} \frac{\frac{dN_2}{d\theta}(\lambda_{20})}{\pi(\theta)}} \end{aligned} \quad (22)$$

and,

$$p_{astro,k} = \frac{dN_a(\lambda_{10}, \lambda_{20})}{dN_a(\lambda_{10}, \lambda_{20}) + dN_b} \quad (23)$$

- **Corollary:** If our FAR cut is low enough to exclude any background triggers, then

$$p_{astro,k} \rightarrow 1$$

This implies:

$$P(\lambda_2 | \vec{d}_2, N_{tr2}, \lambda_1) \propto \pi(\lambda_2) e^{-R_2 \overline{VT}(\lambda'_2)} \prod_{k=1}^{N_{tr2}} \left[\frac{R_2}{R_{20}} w_{2k}(\lambda'_2, \lambda'_{20}) p_{astro,m,k}(\lambda_{10}, \lambda_{20}) + (1 - p_{astro,m,k}(\lambda_{10}, \lambda_{20})) \right] \quad (24)$$

Notice that, unlike analyzing the full population, rate marginalization here is not so simple.

REFERENCES

- | | |
|--|--|
| <p>Mandel, I., Farr, W. M., & Gair, J. R. 2019, Monthly Notices of the Royal Astronomical Society, 486, 1086, doi: 10.1093/mnras/stz896</p> | <p>Roulet, J., Venumadhav, T., Zackay, B., Dai, L., & Zaldarriaga, M. 2020, Physical Review D, 102, doi: 10.1103/physrevd.102.123022</p> |
|--|--|