Identification and Significance Assessment of Compact Object Merger Candidates

Kipp Cannon

Presented at CNSSS19, University of Tokyo, Hongo, August 22, 2019
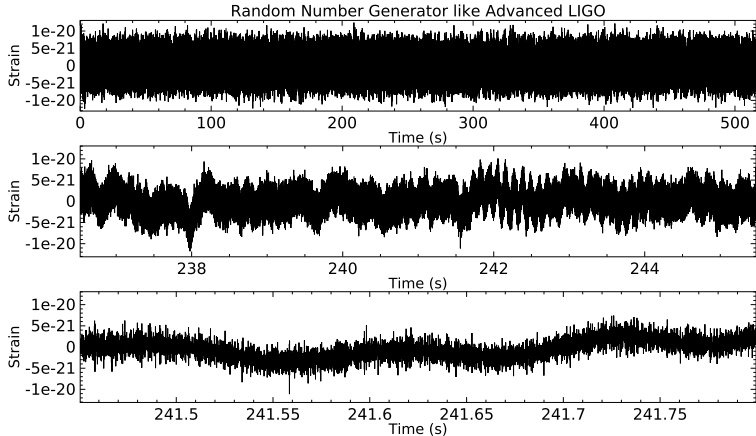
Compact Object Mergers

- ▶ Orbiting bodies lose energy to gravitational radiation.
- ▶ For most of the system's evolution the radiated wave has simple, easily modelled, structure.
- ▶ For compact objects (black holes, neutron stars), orbital speeds get close to c.
- ▶ Near merger, size and deformability of objects becomes significant, dynamics of spacetime, too.
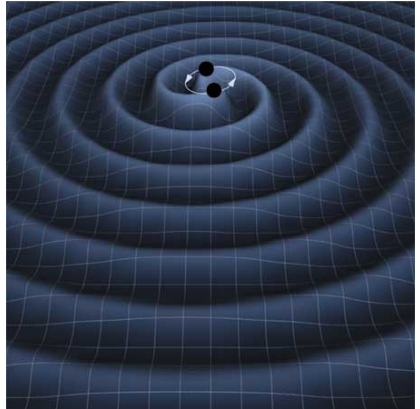
Credit: K. Thorne (Caltech) /
T. Carnahan (NASA GSFC)
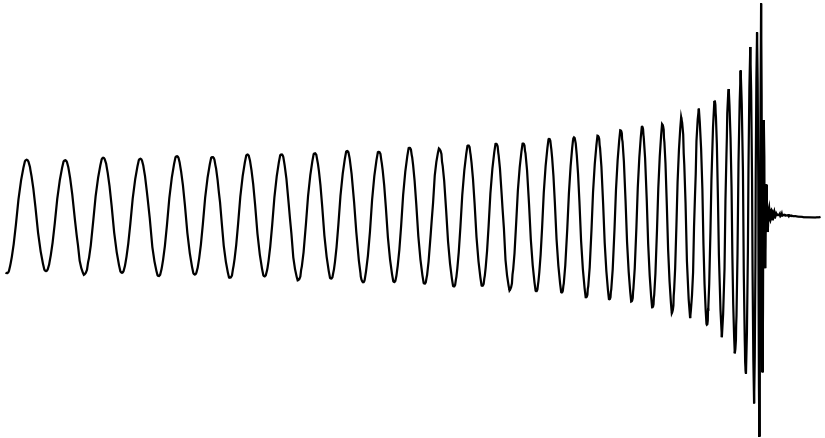
- ▶ Prototypical compact object merger waveform. Vertical axis is strain, horizontal axis is time, scales are irrelevant.
- ▶ If 1.4+1.4 BNS merger: horizontal axis spans about 0.02 s, left edge corresponds to about 400 Hz.

Random Number Generator like Advanced LIGO

▶ Question: do these data contain one or more compact object merger signals?

Plan

Question: do these data contain one or more compact object merger signals?

- ▶ Neyman-Pearson criterion and "optimal" detection statistics.
- ▶ Mathematical description of random fields.
- ▶ Karhunen-Loève theorem and likelihood ratios for correlated variables.
- ▶ The matched filter.
- ▶ The real world.

For More Information

- Erik Vanmarcke. Random Fields: Analysis and Synthesis. The MIT Press, 1983.
- http://relativity.livingreviews.org/Articles/lrr-2012-4/
- https://en.wikipedia.org/wiki/Karhunen%E2%80%93Loeve_expansion
- Cannon *et al.*, A method to estimate the significance of coincident gravitational-wave observations from compact binary coalescence, Phys. Rev. D, **88**, 024025 (2012), arXiv:1209.0718 [gr-qc]
- Cannon *et al.*, Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation, arXiv:1504.04632 [astro-ph.IM]

Neyman-Pearson Criterion

▶ Criterion defines meaning of "optimal" detector.

▶ When performing a hypothesis test between two point hypotheses, choose the descriminant that maximizes the detection efficiency given a fixed false-alarm probability.

▶ Neyman-Pearson Lemma: the likelihood-ratio test satisfies the Neyman-Pearson Criterion.

▶ The "likelihood ratio" is

$$\Lambda(\theta) = \frac{P(\theta|\text{signal})}{P(\theta|\text{no signal})} \tag{1}$$

where θ are your data.

▶ Computing Λ and applying a threshold is the most efficient detector of "signals" at first false-alarm rate.

▶ Example: "Some coins in a batch favour heads over tails by 20%, that is, $P(\text{head}) = 0.6$, while all other coins are fair. You wish to identify biased coins by flipping each coin 3 times. Write a statistic to identify biased coins with the greatest detection efficiency for some choice of false-alarm probability using the outcome of three tosses."

Neyman-Pearson Criterion: Example's Solution

▶ Consider 3 tosses to yield outcomes x_1, x_2, x_3 (e.g., $x_1 =$ head, $x_2 =$ head, $x_3 =$ tail).

▶ Probability that fair coin produces exactly this result:

$$P(x_1, x_2, x_3|\text{not biased}) = 0.5^3 = 1/8. \tag{2}$$

▶ Probability that biased coin produces this result:

$$P(x_1, x_2, x_3|\text{biased}) = \begin{Bmatrix} .4 \text{ if } x_1 \text{ tail} \\ .6 \text{ if } x_1 \text{ head} \end{Bmatrix} \begin{Bmatrix} .4 \text{ if } x_2 \text{ tail} \\ .6 \text{ if } x_2 \text{ head} \end{Bmatrix} \begin{Bmatrix} .4 \text{ if } x_3 \text{ tail} \\ .6 \text{ if } x_3 \text{ head} \end{Bmatrix} \tag{3}$$

▶ Discriminating statistic:

$$\Lambda(x_1, x_2, x_3) = \frac{P(x_1, x_2, x_3|\text{biased})}{P(x_1, x_2, x_3|\text{not biased})} = 8P(x_1, x_2, x_3|\text{biased}). \tag{4}$$

▶ Coin is put into "biased" pile if $\Lambda \geq \Lambda_0$, where Λ_0 is some threshold.

Neyman-Pearson Criterion: Example's Solution

▶ Sometimes a fair coin is classified as biased (false alarm), and sometimes a biased coin classified as fair (false dismissal, efficiency < 1). Each choice of Λ_0 leads to a different false-alarm probability, a different efficiency.

▶ Λ does not tell you the probability that a coin is biased, it tells you what pile to put it in given the rate at which you are prepared to contaminate the pile with incorrectly labelled fair coins.

▶ $8P(x_1, x_2, x_3|\text{biased})$ is the likelihood ratio so we know it satisfies the Neyman-Pearson criterion, but obviously so would $P(x_1, x_2, x_3|\text{biased})$ (without the factor or 8).

▶ Any function that is monotonic in Λ also satisfies the N.-P. criterion, and is equivalent in its suitability for use as a ranking statistic.

▶ Typically $\ln \Lambda$ is used as the ranking statistic because it's often more convenient to work with the logarithms of probabilities.

Time Series
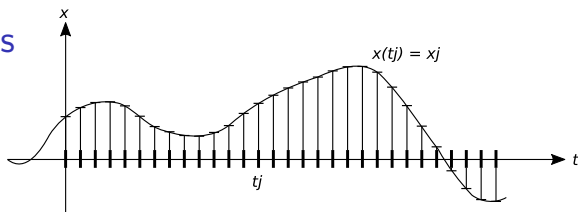
- A time series is a vector. Properties of a vector space:
    - have two operations: "addition" and "scalar multiplication"
    - addition is commutative and associative $\vec{x} + \vec{y} = \vec{y} + \vec{x}$,
      $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$.
    - there exists an element "0" that is the additive identity $\vec{x} + 0 = \vec{x}$.
    - for each element there exists an additive inverse element, $-\vec{x}$.
    - scalar and field multiplication are compatible $(ab)\vec{x} = a(b\vec{x})$.
    - scalar multiplication distributes over field addition $a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$
    - scalar multiplication distributes over scalar addition
      $(a + b)\vec{x} = a\vec{x} + b\vec{x}$.
- It is (I think) self-evident that real- and complex-valued functions of time possess all of these properties.

Inner Product Space

▶ An inner product space is a vector space that possesses one additional property: an "inner product".

▶ An inner product associates each pair of vectors with a scalar and has the following properties:
  ▶ conjugate symmetry: $\langle \vec{x}, \vec{y} \rangle = \overline{\langle \vec{y}, \vec{x} \rangle}$.
  ▶ linear in first argument: $\langle a\vec{x}, \vec{y} \rangle = a \langle \vec{x}, \vec{y} \rangle$,
    $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle$.
  ▶ positive definite: $\langle \vec{x}, \vec{x} \rangle \geq 0$, $\langle \vec{x}, \vec{x} \rangle = 0$ iff $\vec{x} = 0$.

▶ The inner product of a vector with itself can be used to give the space a norm.

Inner Product Space

- It's maybe less obvious, but functions of time possess an inner product as well, for example,

$$\langle f(t), g(t) \rangle = \int f(t) g^*(t) \, \mathrm{d}t, \tag{5}$$

Vector Spaces, Inner Product Spaces: WHO CARES?

▶ Because, silly, now everything you know about vectors and geometry is true about functions as well:
  ▶ a function, like a vector, can be tought of as a point in a space (the point at the end of the vector if placed at the origin);
  ▶ continuous families of functions are surfaces and volumes in that space;
  ▶ you can compute the "distance" between two functions;
  ▶ you can "project" a function onto a family of functions' hypersurface, thereby decomposing it into a component in that family of functions and a component orthogonal to that family.

▶ You can formulate answers to questions like "Does this function of time contain a component that belongs to some given family of functions?"

▶ Do my data contain a binary black hole merger?

The Fourier Transform

$$x_j = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}_k \exp[2\pi \mathrm{i} jk/N] \tag{6}$$

$$\tilde{x}_k = \sum_{j=0}^{N-1} x_j \exp[-2\pi \mathrm{i} jk/N] \tag{7}$$

▶ A linear operator. Example: two-sample time series

$$\begin{bmatrix} \tilde{x}_0 \\ \tilde{x}_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \tag{8}$$

▶ Rows of matrix are orthogonal to each other, and there are as many of them as there are dimensions, so they form a complete orthogonal basis.

▶ Therefore (if suitably normalized) the transformation is unitary, so it preserves inner product.

The Fourier Transform

- Inner-product preserving = norm preserving, "Parseval's theorem".
- Inner-product preserving = angle preserving. $(\vec{a} \cdot \vec{b} = |\vec{a}| \left|\vec{b}\right| \cos \theta)$.
- If all angles between vectors are preserved, the Fourier transform must be (something like) a rotation — it's a "generalized rotation".

(Gaussian) Random Fields

▶ An n-dimensional random field, $f(\vec{r})$, is a set of random variables, one for each point $\vec{r}$.

▶ The field and all of its statistical properties are entirely defined by the set of finite dimensional joint probability distribution functions,

$$P\left[f(\vec{r}_1), \ldots, f(\vec{r}_m)\right] \, \mathrm{d}f(\vec{r}_1) \cdots \mathrm{d}f(\vec{r}_m), \tag{9}$$

giving the probability that for the given $\vec{r}_i$, $i = 1, \ldots, m$, the field f simultaneously takes on values at the points $\vec{r}_i$ in the ranges $f(\vec{r}_i)$ to $f(\vec{r}_i) + \mathrm{d}f(\vec{r}_i)$.

▶ A Gaussian random field is one for which all the m-point probability distributions are Gaussian in the $f(\vec{r}_i)$.

(Gaussian) Random Fields

▶ A Gaussian m-point joint probability distribution for the random variables f_i is given by

$$P(f_1, \ldots, f_m) \, df_1 \cdots df_m = \frac{1}{\sqrt{2\pi}^m \sqrt{\det \mathbb{B}}} \exp\left[-\frac{1}{2} \Delta \vec{f}^{\mathrm{T}} \mathbb{B}^{-1} \Delta \vec{f}\right] df_1 \cdots df_m \tag{10a}$$

where $\mathbb{B}$ is the covariance matrix of the f_i,

$$\mathbb{B}_{ij} \equiv \langle \Delta f_i \Delta f_j \rangle, \tag{10b}$$

and $\Delta \vec{f}$ is the column vector of residuals

$$\Delta f_i \equiv f_i - \langle f_i \rangle. \tag{10c}$$

▶ Note that for Gaussian random variables, any m-point joint distribution requires only the means and two-point covariance.

▶ Sometimes you will see the notation $\langle\langle f_i f_j \rangle\rangle \equiv \langle \Delta f_i \Delta f_j \rangle$.

(Gaussian) Random Fields

▶ For a random field, $f(\vec{r})$, with continuous co-ordinates the covariance matrix generalizes to the two-point covariance function,

$$B(\vec{r}_i, \vec{r}_j) = \langle\!\langle f(\vec{r}_i) f(\vec{r}_j) \rangle\!\rangle, \qquad (11)$$

and, as above, this contains all the information required to construct any m-point joint distribution for the field.

▶ A strictly homogeneous random field is one for which all finite-dimensional joint probability distributions, (9), are left unchanged by a co-ordinate translation. This implies that the joint distribution functions depend on the co-ordinate separations, $r_{ij} = r_j - r_i$, called the lag factors, alone; and that the mean, $\langle f(r) \rangle$, and mean square, $\langle f^2(r) \rangle$, are independent of r.

▶ In this case, the two-point covariance function can be translated to the origin without loss of generality,

$$\mathbb{B}_{ij} = B(r_i, r_j) = \langle\!\langle f(r_i) f(r_j) \rangle\!\rangle = \langle\!\langle f(0) f(r_j - r_i) \rangle\!\rangle = \langle\!\langle f(0) f(r) \rangle\!\rangle = B(r),$$

$$(12)$$

thereby becoming a function of a single co-ordinate.

(Gaussian) Random Fields

▶ If the mean, in addition to being independent of r, is also 0 then the covariance function reduces to

$$B(r) = \langle\!\langle f(0)f(r)\rangle\!\rangle = \langle f(0)f(r)\rangle. \tag{13}$$

In this case, this function is also sometimes denoted as $\xi(0, r) = \xi(r)$, the (two-point) correlation function.

▶ A random field is isotropic if the joint probability distribution functions are left unchanged by rotations. For a homogeneous random field, isotropy means the distribution functions depend only on the magnitudes of the lag factors, $|\vec{r}_{ij}| = |\vec{r}_j - \vec{r}_i|$, not their direction.

▶ Finally, a random field is called ergodic if a single realization of the field contains sufficient information to completely determine all the joint probability distribution functions: ensemble averages and co-ordinate averages are interchangable.

(Gaussian) Random Fields

► It is important to understand that when dealing with random fields one is always discussing the properties of the statistics of the field. This contrasts with classical fields where one characterizes them by the properties of their values. For this reason, the meaning of homogeneous and isotropic when used in the context of a random field is not the same as when those same words are used in the context of classical fields. In particular, a homogeneous classical scalar field is necessarily "isotropic" since a field that is the same everywhere is left unchanged by a rotation. A homogeneous random scalar field, however, can be "anisotropic" as long as the direction dependence of its statistical properties is the same everywhere.
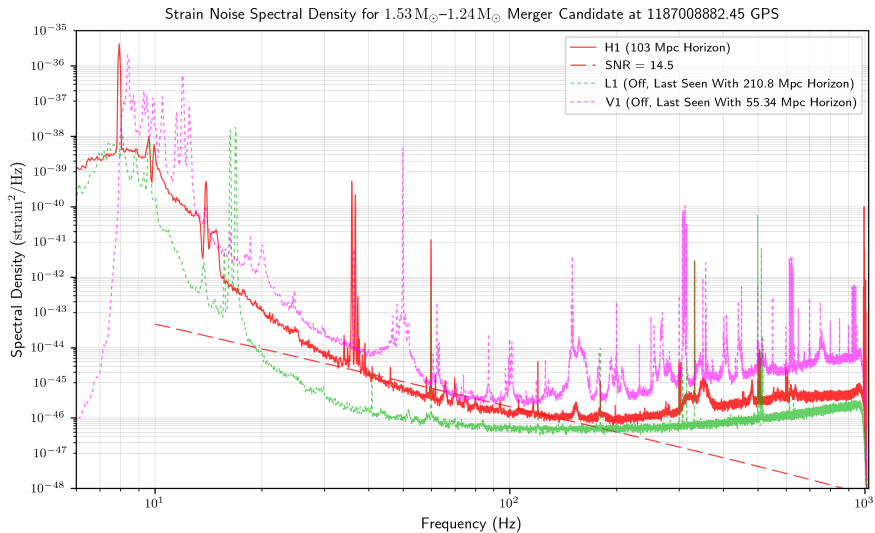
(Gaussian) Random Fields

- The *n*-dimensional Fourier transform of the field is defined as usual
- The integral(sum) is imagined to be evaluated once for each member of the ensemble of field realizations; each evaluation yields a realization of the Fourier transform, that taken together define the ensemble of realizations and from that one can determine the statistical properties of the transform.
- One very important quantity is the two-point spectral correlation function

$$\left\langle \tilde{f}_k \tilde{f}_{k'}^* \right\rangle = \frac{1}{(2\pi)^n} \left\langle \int f(r_1) e^{-ik \cdot r_1} f(r_2) e^{+ik' \cdot r_2} \, dr_1 \, dr_2 \right\rangle \tag{14}$$

- For stationary fields $\left\langle \tilde{f}_k \tilde{f}_{k'}^* \right\rangle \propto \delta(k - k')$ in the continuous case, or $\left\langle \tilde{f}_k \tilde{f}_{k'}^* \right\rangle \propto \delta_{kk'}$ for discrete fields.
- The proportionality constant is the "spectral density function" S_k.

(Gaussian) Random Fields

Strain Noise Spectral Density for $1.53\,\mathrm{M_\odot}$–$1.24\,\mathrm{M_\odot}$ Merger Candidate at 1187008882.45 GPS

(Gaussian) Random Fields

▶ For stationary, one-dimensional, zero-mean Gaussian random field $f(t)$, spectral density function is Fourier transform of two-point correlation function

$$S_k = \left\langle \left| f_{\tilde{k}} \right|^2 \right\rangle = \int \xi(t) \mathrm{e}^{-\mathrm{i}kt} \, \mathrm{d}t \qquad (15)$$

▶ Deriving this is tricky, and for continuous fields the units of the LHS are not what you expect; for discrete fields (sampled time series data) the units of the LHS *are* what you expect.
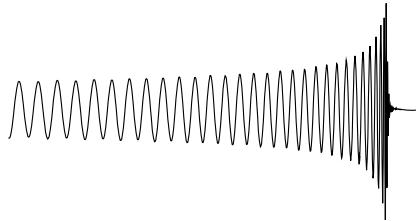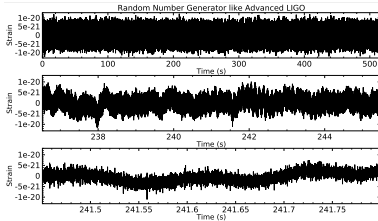
Karhunen-Loève Theorem

▶ Given a zero-mean square integrable stochastic process $n(t_i)$ defined on an interval $[a, b]$ with covariance matrix $\mathbb{B}_{ij} = \langle f(t_i)f(t_j) \rangle$, $t_i, t_j \in [a, b]$, projecting $n(t_i)$ onto the eigenspace of the covariance matrix yields a stochastic process $n'(t_i)$ whose covariance matrix is diagonal.

Karhunen-Loève Theorem: Example

▶ Consider *coloured* zero-mean stationary Gaussian noise $n(t)$ with covariance matrix $\mathbb{B}_{ij} = \langle n(t_i)n(t_j)\rangle$; a signal $s(t_i)$ might or might not be present in the noise, construct a statistic to test for its presence that possesses the highest detection efficiency given a choice of false-alarm probability. Assume the noise is additive, that is with no signal present the data are given by $x(t_i) = n(t_i)$, with the signal present the data are given by $x(t_i) = n(t_i) + s(t_i)$.

Karhunen-Loève Theorem: Examle's Solution

▶ By N.-P. Lemma we know the likelihood ratio, $\Lambda(\vec{x}) = \frac{P(\vec{x}|\text{signal})}{P(\vec{x}|\text{no signal})}$, will suffice as the detection statistic.

▶ Need to calculate $P(\vec{x}|\text{signal})$ and $P(\vec{x}|\text{no signal})$.

▶ In principle this is simple, however, recalling m-point joint distribution for the x_i, we remember that we need to compute $\mathbb{B}^{-1}$, which quickly becomes impractical: e.g., neutron star merger waveform spends 30 minutes in-band in LIGO, at 16384 samples/second $\mathbb{B}$ is 30 million samples square that needs to be inverted. This is *possible* with present day computers, but you can probably count on one hand the number of computers in the world able to perform this calculation.

Karhunen-Loève Theorem: Examle's Solution

- ▶ K.-L. theorem tells us there exist functions $\Phi_k(t_i)$ that are solutions to

$$\sum_j \mathbb{B}_{ij} \Phi_k(t_j) \Delta t = \lambda_k \Phi_k(t_i) \tag{16}$$

  that form an orthonormal basis, and that when projected onto that basis the components of $n(t_i)$ are independent random variables.

- ▶ Because the noise is stationary, $\mathbb{B}_{ij} = B(t_i - t_j)$, it is a circulant matrix. One can show that the eigenvectors, $\Phi_k(t_i)$, of a circulant matrix are Fourier modes with eigenvalues (in this case) $\lambda_k = \left\langle |\tilde{n}_k|^2 \right\rangle = S_k$, the spectral density function.

Karhunen-Loève Theorem: Examle's Solution

▶ Fourier transforming the noise turns it into a set of independent Gaussian random variables, so we can immediately write

$$P(\vec{x}|\text{signal}) = \prod_k \frac{1}{\sqrt{2\pi S_k}} \exp{-\frac{(\tilde{x}_k - \tilde{s}_k)^2}{2S_k}} \qquad (17)$$

$$P(\vec{x}|\text{no signal}) = \prod_k \frac{1}{\sqrt{2\pi S_k}} \exp{-\frac{\tilde{x}_k^2}{2S_k}} \qquad (18)$$

▶ It's convenient to work with the log of the likelihood ratio, $\ln\Lambda$, which is

$$\ln\Lambda = \sum_k \frac{2\tilde{s}_k\tilde{x}_k - \tilde{s}_k^2}{2S_k} \qquad (19)$$

▶ Therefore, thresholding on $F = \sum_k \frac{\tilde{x}_k \tilde{s}_k}{S_k}$ satisfies the Neyman-Pearson criterion.

Geometric Interpretation

$$F = \sum_k \frac{\tilde{x}_k \tilde{s}_k}{S_k} \qquad (20)$$

▶ We perform a rotation-like transformation (F.T.) into a co-ordinate system where the principal axes of the noise distribution are aligned with the co-ordinate axes.

▶ Then rescale each co-ordinate so that the variance of the noise is identical in all directions.

▶ In *that* space, the optimum detection statistic is a simple inner product: the magnitude of the perpendicular projection of the data onto the (normalized) signal vector.

▶ We say the noise, described by S_k, defines a metric for the waveform space, and we are transforming to a co-ordinate system where the metric is flat, then using the normal inner product in that co-ordinate system as the detection statistic.

The Matched Filter

▶ The detector derived above assumed we knew where the signal would be located in the data. What if we don't?

▶ One approach is to treat the time of the signal as a parameter, not as a degree of freedom of the model, and obtain a detection statistic that is a function of that parameter.

$$\tilde{s}_k \to \tilde{s}_k e^{2\pi i j k} \tag{21}$$

$$F \to F(t_j) = \sum_k \frac{\tilde{x}_k \tilde{s}_k}{S_k} e^{2\pi i j k} \tag{22}$$

This can be computed efficiently using an FFT and if rescaled so that $\langle F \rangle = 1$ is called the "matched-filter signal-to-noise ratio".

▶ NOTE: this is not an optimal detector for the presence of a signal (strictly speaking, we've simply not proven it to be, but I can tell you that it's not). This is not answering the question "is there is a signal in the data?", this provides a sequence of answers the questions "is there a signal in the data now? now? now? ...".

The Real World

- ▶ Data do not constitute a stationary Gaussian random field: m-point joint distribution functions are not time-translation invariant, there are "glitches" in the data whose amplitudes are maybe more like power laws than Gaussians (like earthquakes), though nobody really knows what statistics they obey.
- ▶ We are not searching for one signal, but for any member of a family of signals, not all of which we believe are equally likely.

The Real World

- ▶ What do we do? We play a different game.
- ▶ Recall the coin toss test used to find biased coins:
  - ▶ the N-P criterion and lemma told us how to construct an "optimum" discriminant from the outcome of three tosses of the coin,
  - ▶ but it was our choice to use the outcome of **three** tosses as the data from which to construct the discriminant.
  - ▶ Could have chosen to use four tosses,
  - ▶ could have tried to measure the coins' internal mass distribution with an X-ray machine.

The Real World

- ▶ So far we've been imagining that our "data" are the strain time-series samples collected in the instrument. Give up on that.
- ▶ Guided by the knowledge that the matched filter is a good idea for a known signal at a known time in stationary Gaussian noise, run a bank of matched filters over the data, and collect the threshold crossings.
- ▶ Borrowing from the language of particle physics, call these "triggers"; if more than one detector yields a trigger from the same template at nearly the same time called that a "coincidence".
- ▶ Single triggers, and coincidences of triggers, are collectively called "candidates".

The Real World

- ▶ The question redefined: considering each coincidence individually to be the data, is it a GW? Construct a detection statistic to decide this with the highest effency at some fixed false-alarm rate.

- ▶ Alternatively: considering the coincidences collectively to be the data, how many signals are present? This is not a detection problem, it's a problem in statistical inference. See Farr *et al.*, arXiv:1302.5341 [astro-ph.IM].

- ▶ Remember: there is *no reason* for this methodology, there is no derivation that tells you that "template banks" and "triggers" and "coincidences" are the best thing to be doing.

- ▶ This is the end result of a series of choices that have evolved organically as our understanding of the problem has improved while subject to the constraint of getting results published on time.

- ▶ Nevertheless, given this choice, we can apply the N-P lemma to construct the best ranking statistic that we can.

Implementation

- ▶ Each template waveform is one member of the family of waveforms we are looking to find in the data, we construct a matched filter for each template, and from that each "trigger" provides:
  - ▶ The physical properties of the source modelled by the template: component masses, spins, tidal deformabilities, etc., etc..
  - ▶ The time of the signal-to-noise ratio's peak.
  - ▶ The magnitude and phase of the signal-to-noise ratio at its peak.
  - ▶ One or more χ^2 residual values.
  - ▶ The identity of the antenna that produced the trigger (*i.e.*, its location and orientation).
  - ▶ That antenna's strain spectral density at the time of the event.
- ▶ Each "coincidence" is a set of triggers, and so provides all of the information carried by the triggers as well as:
  - ▶ The identities of the antennas that did *not* yield triggers.
  - ▶ Their strain spectral densities at the time of the event.
  - ▶ The time delays between SNR peaks for the antennas that did.
- ▶ The optimum detection statistic in the N-P sense is

$$\mathcal{L}(\text{coincidence}) = \frac{P(\text{parameters}|\text{signal})}{P(\text{parameters}|\text{noise})}.$$

Implementation

- ▶ The different pipelines differ in how they approximate

$$\mathcal{L}(\text{coincidence}) = \frac{P(\text{parameters|signal})}{P(\text{parameters|noise})}.$$

- ▶ ihope/pycbc:
  - ▶ Discard trigger masses, etc., trigger phases, antenna sensitivies, and inter-instrument time delays.
  - ▶ Digest each trigger's SNR and χ^2 properties into an "effective SNR" value, ρ_{eff}.
  - ▶ $P(\text{parameters|signal}) \approx 1$.
  - ▶ $P(\text{parameters|noise}) \approx 1/\sqrt{\sum \rho_{\text{eff}}^2}$.

Implementation

$$\mathcal{L}(\text{coincidence}) = \frac{P(\text{parameters}|\text{signal})}{P(\text{parameters}|\text{noise})}$$

▶ gstlal:
  ▶ See Cannon *et al.*, Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation, arXiv:1504.04632 [astro-ph.IM].

$$\mathcal{L}(\ldots) = \mathcal{L}(\ldots|\bar{\theta})\,\mathcal{L}(\bar{\theta})$$
$$= \frac{P\left(\{\text{H1},\text{L1},\ldots\}, t_{\text{H1}}, \rho_{\text{H1}}, \chi^2_{\text{H1}}, \ldots|\bar{\theta}, \text{signal}\right)}{P\left(\{\text{H1},\text{L1},\ldots\}, t_{\text{H1}}, \rho_{\text{H1}}, \chi^2_{\text{H1}}, \ldots|\bar{\theta}, \text{noise}\right)}\mathcal{L}(\bar{\theta})$$

Implementation

$$\frac{P\left(\{\mathrm{H1},\mathrm{L1},\ldots\},t_{\mathrm{H1}},\rho_{\mathrm{H1}},\chi^2_{\mathrm{H1}},\ldots|\bar{\theta},\mathrm{signal}\right)}{P\left(\{\mathrm{H1},\mathrm{L1},\ldots\},t_{\mathrm{H1}},\rho_{\mathrm{H1}},\chi^2_{\mathrm{H1}},\ldots|\bar{\theta},\mathrm{noise}\right)}\mathcal{L}\left(\bar{\theta}\right)$$

▶ Using disccretely-sampled (histogram) approximations of the joint probability distributions of the remaining parameters, and

▶ Assuming statistical independence of the noise processes in distinct antennas,

▶ Populate denominator PDFs using triggers that fail coincidence combined with a Poisson model for the formation of chance coincidences.

▶ Replace $P(\text{parameters}|\text{signal})$ with $P(\text{parameters}|\text{signal we can publish a detection claim for})$ and compute the latter algebraically assuming the statistics of clean Gaussian data.

Implementation

Numerator is factored like this:

$$
\begin{aligned}
= P\left(t | \text{signal}\right) &\times P\left(\{\text{H1}, \text{L1}, \ldots\} | t, \text{signal}\right) \\
&\times P\left(\rho_{\text{H1}}, \rho_{\text{L1}}, \ldots | t, \{\text{H1}, \text{L1}, \ldots\}, \text{signal}\right) \\
&\times \prod_{\text{inst} \in \{\text{H1}, \text{L1}, \ldots\}} P\left(\chi^2_{\text{inst}} | \rho_{\text{inst}}, \text{signal}\right).
\end{aligned}
$$

▶ Assume that χ^2 residuals are correlated with SNR, and that SNRs are correlated across instruments, but apart from that χ^2 values are statistically independent.

▶ $P(t|\text{signal})$ obtained from detector sensitivities (horizon distances).

Implementation

Denominator is factored like this:

$$= P\left(t|\text{noise}\right) \times P\left(\{\text{H1}, \text{L1}, \ldots\}\,|t, \text{noise}\right)$$
$$\times \prod_{\text{inst}\in\{\text{H1},\text{L1},\ldots\}} P\left(\rho_{\text{inst}}, \chi^2_{\text{inst}}|\text{noise}\right).$$

▶ Assume noise processes in different instruments are statistically independent.
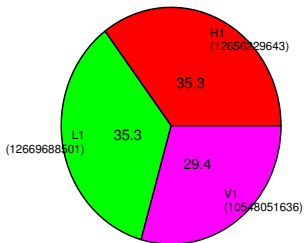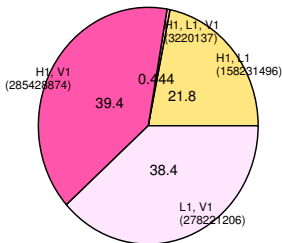▶ $P(t|\text{noise})$ obtained from detector trigger rates.

Implementation

Strain Noise Spectral Density for $1.53\,\mathrm{M_\odot}$–$1.24\,\mathrm{M_\odot}$ Merger Candidate at 1187008882.45 GPS

Legend:
- H1 (103 Mpc Horizon)
- SNR = 14.5
- L1 (Off, Last Seen With 210.8 Mpc Horizon)
- V1 (Off, Last Seen With 55.34 Mpc Horizon)

Spectral Density (strain^2/Hz) vs Frequency (Hz)
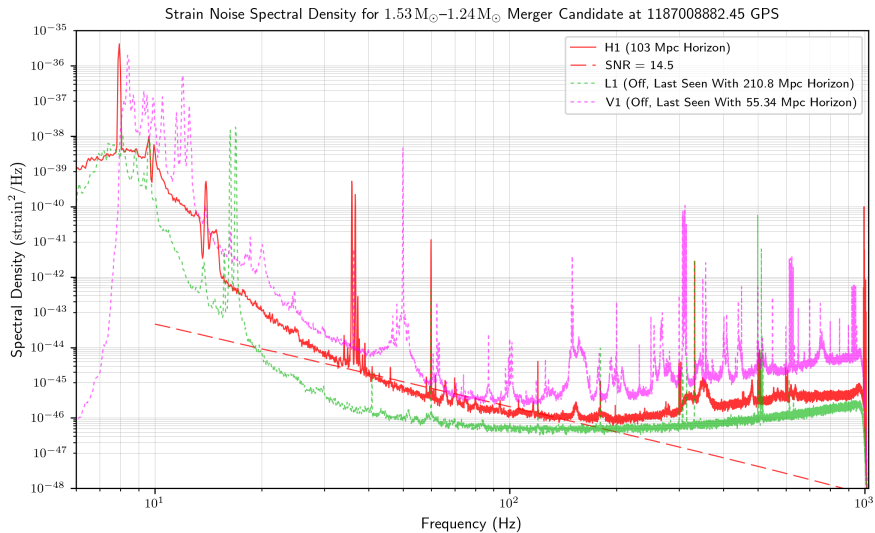
Implementation

Observed Background Event Counts

Projected Background Coincidence Counts
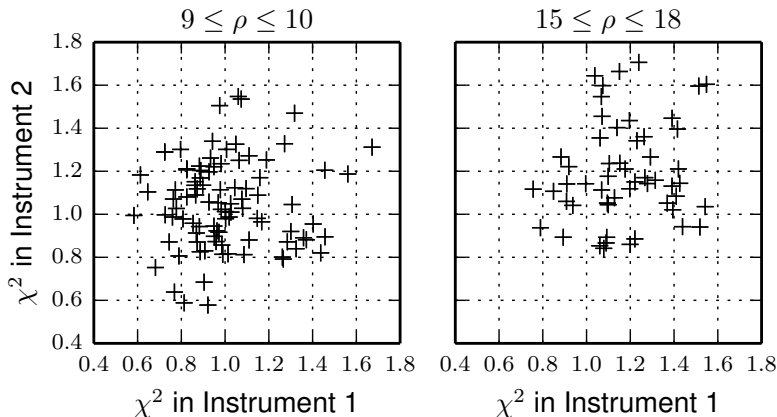
Projected Recovered Signal Distribution
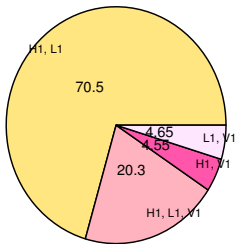
Observed Coincidence Counts
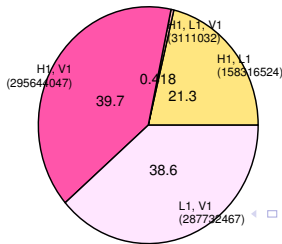
See Cannon *et al.*,
arXiv:1504.04632
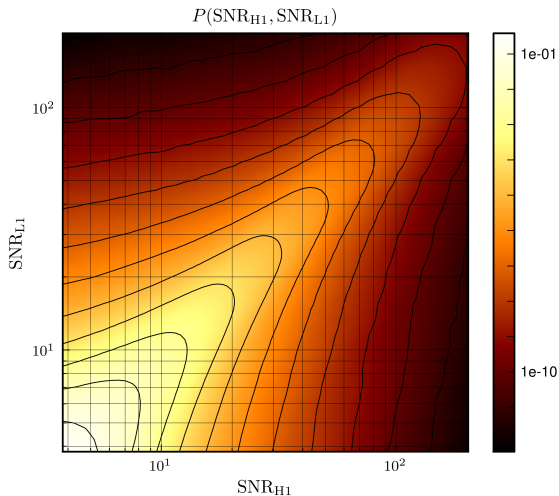[astro-ph.IM].
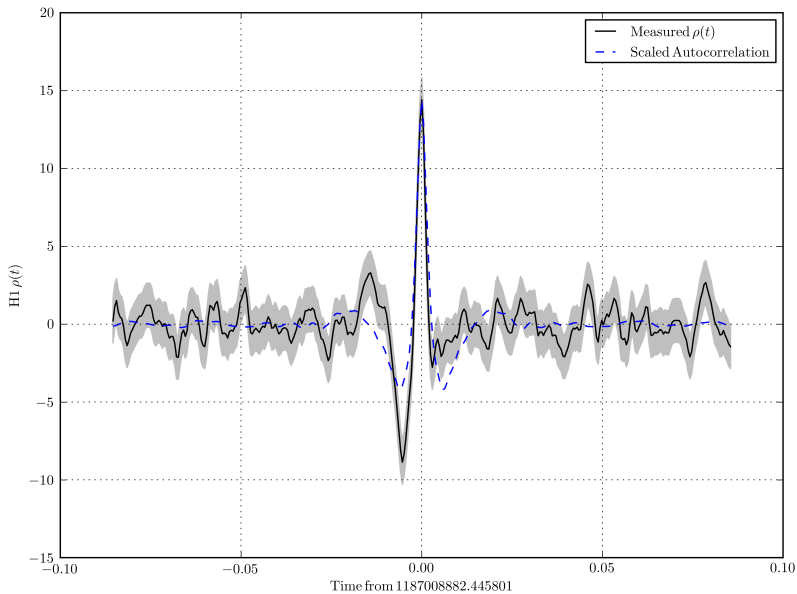
Implementation

Statistical independence of χ^2.
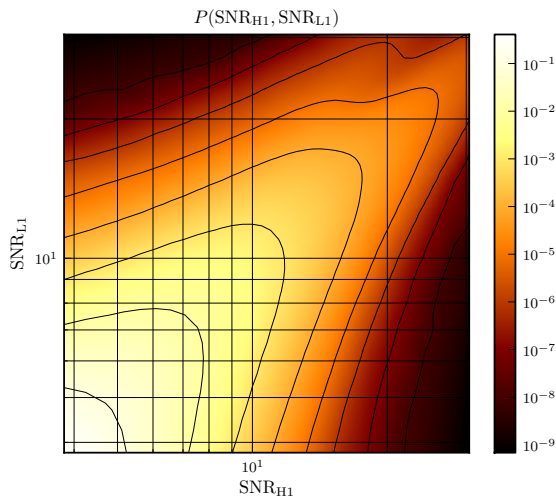

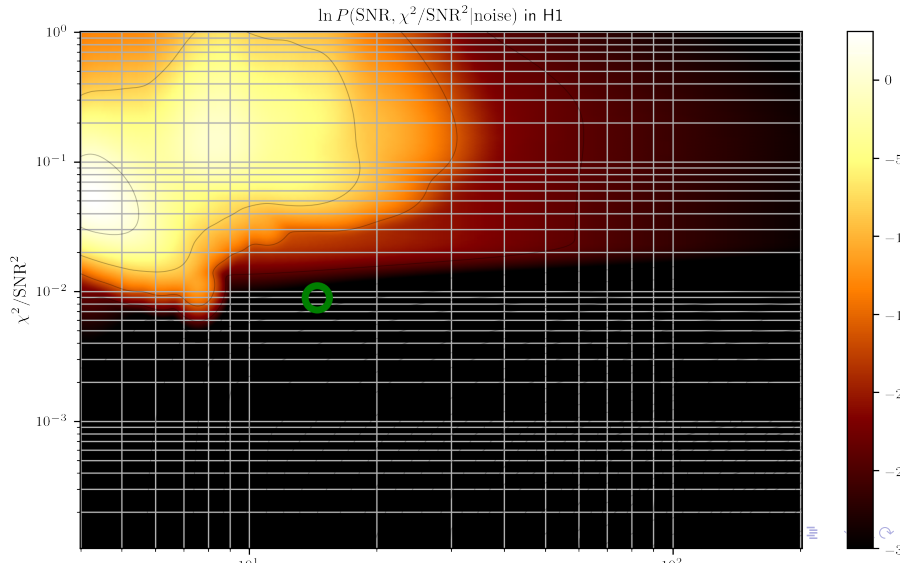See Cannon *et al.*, arXiv:1504.04632 [astro-ph.IM].

Implementation

$P(\mathrm{SNR}_{\mathrm{H1}}, \mathrm{SNR}_{\mathrm{L1}})$

With no Virgo in the network.

Implementation

$P(\text{SNR}_{\text{H1}}, \text{SNR}_{\text{L1}})$

With Virgo in the network. See Cannon *et al.*, arXiv:1504.04632 [astro-ph.IM].

Implementation

$\ln P(\mathrm{SNR}, \chi^2/\mathrm{SNR}^2 | \mathrm{noise})$ in H1

Implementation

$\ln P(\chi^2/\mathrm{SNR}^2|\mathrm{SNR}, \mathrm{signal})$ in H1

Implementation

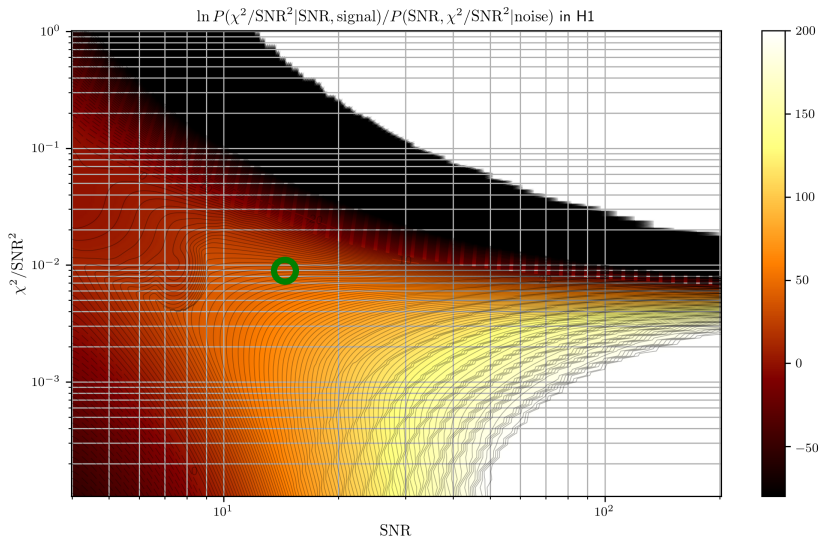$\ln P(\chi^2/\mathrm{SNR}^2|\mathrm{SNR},\mathrm{signal})/P(\mathrm{SNR},\chi^2/\mathrm{SNR}^2|\mathrm{noise})$ in H1
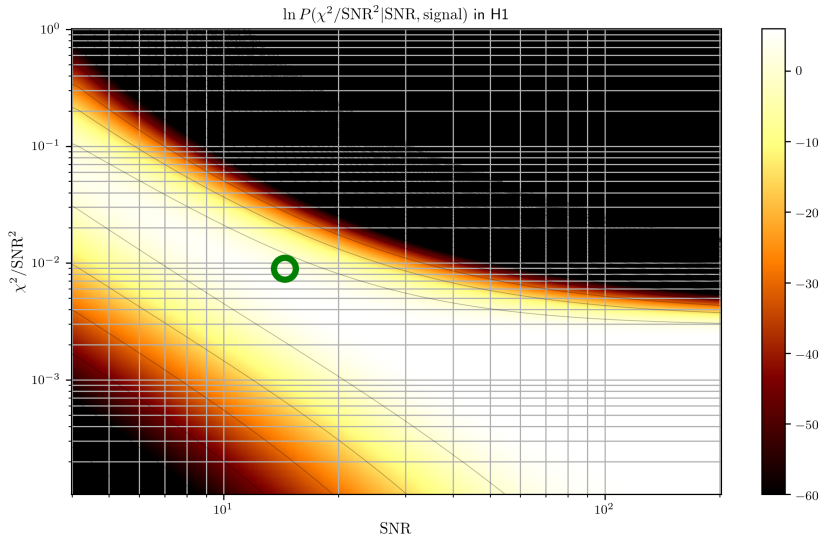
Implementation

Recent improvements:

▶ Selection of density estimation kernel for (ρ, χ^2) PDFs automated.

▶ $P(t, \{\mathrm{H1, L1}, \ldots\} | \mathrm{signal})$ and $P(t, \{\mathrm{H1, L1}, \ldots\} | \mathrm{noise})$ now really depend on time instead of being marginalized over two-week blocks.

▶ $\mathcal{L}(\bar{\theta})$ implemented: ranking statistic can now be provided with an astrophysical merger rate model to take into consideration when assessing candidates.

▶ Phases and time delays and their correlations with SNR have been incorporated. Already had this for two detectors in O2, now generalized to N for O3 with Virgo and hopefully KAGRA.
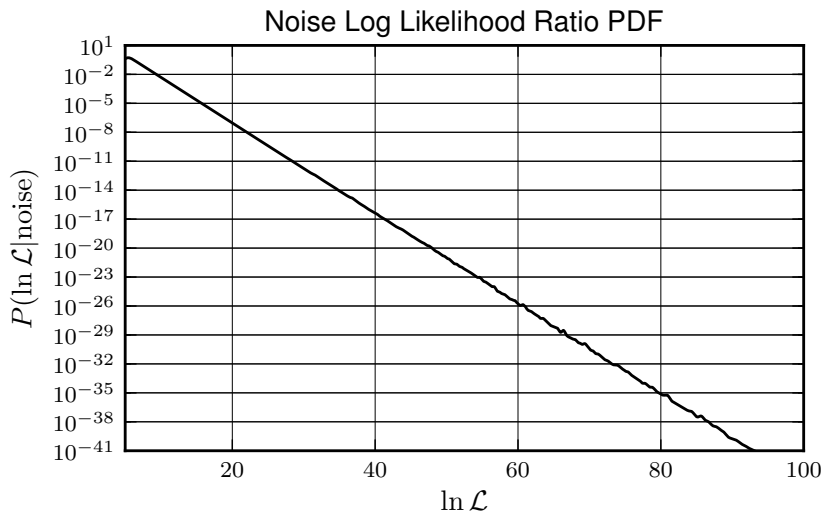
Assessing Significance — False-Alarm Probability

- ▶ pycbc:
  - ▶ Add offsets to trigger times,
  - ▶ construct new coincidences,
  - ▶ histogram ranking statistics,
  - ▶ repeat until tired or until additional statistically independent time slides are no longer possible to construct.
- ▶ gstlal:
  - ▶ Ranking statistic's denominator is the PDF for the parameters of noise coincidences.
  - ▶ Use importance-weighted sampler to generate synthetic noise coincidence parameters,
  - ▶ for each output of the sampler, evaluate ranking statistic,
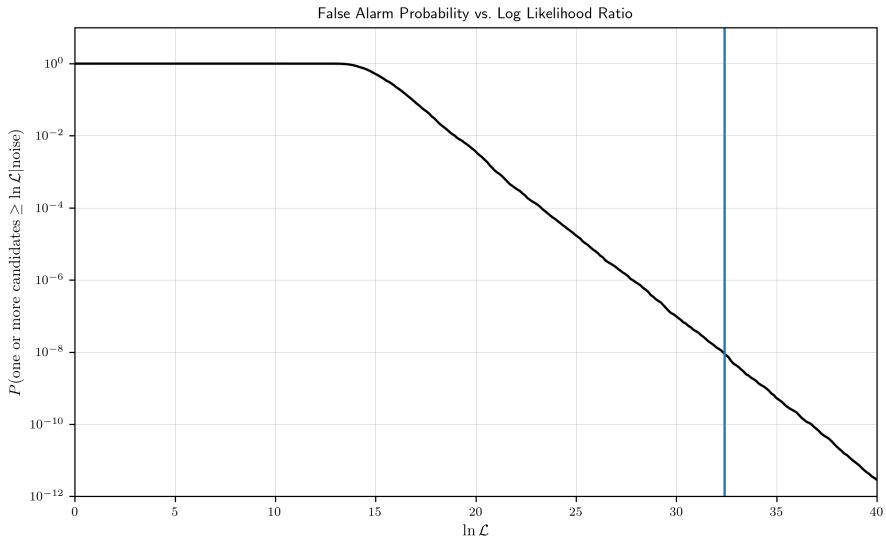 - ▶ histogram the values,
  - ▶ repeat until tired.

Assessing Significance — False-Alarm Probability

Noise Log Likelihood Ratio PDF

Assessing Significance — False-Alarm Probability

False Alarm Probability vs. Log Likelihood Ratio

Take Home Message

- ▶ Not a solved problem.
- ▶ New, better, approach could be devised.
- ▶ Even the current approach is not fully implemented.
    - ▶ Partitioning of template bank into $\bar{\theta}$ blocks not automated.
    - ▶ Auxiliary channels ignored by ranking statistic. ← this is being actively pursued.
- ▶ Lots of room for new ideas and contributions from new people.