Gravitational-Wave Data Analysis: Lecture 4

Peter S. Shawhan

Gravitational Wave Astronomy Summer School
May 31, 2012

Outline for Today

- Probability
- Statistical inference
- Confidence intervals and upper limits
- How to evaluate the sensitivity of a search

Different Meanings of "Probability"

Probability of a given result from some random process (theory)

e.g. rolling two dice and getting a certain sum

or, getting 2 background events in a GW search when the average background is known to be 0.4 events

These are really statements about the expected *frequency* of each outcome of the random process, according to a given theory

Probability of a given theory to be the correct description of the random process

The basis for Bayesian statistics – really about *belief*

e.g. the probability that detectable GW signals occur with an average rate of 1 per year, considering an observation that 2 events were found in a GW search that had an average background of 0.4 events

Note: Theories related by an adjustable parameter are different theories; have to work with probability *density* in such cases

Basic Mathematics of Probability

Definitions:

$p(X) \equiv$ the probability that X is true

$0 \leq p(X) \leq 1$

$p(X|Y) \equiv$ the probability that X is true, given that Y is true

Basic rules:

Sum rule: $p(X,Y) \equiv p(X \text{ and } Y) = p(X|Y)\, p(Y)$

Marginalization: $p(X) = \int dY\, p(X,Y) = \int dY\, p(X|Y) p(Y)$

Bayes' theorem:

Likelihood

Posterior probability

Prior probability

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

"Evidence" $= \int dX\ p(Y\,|\,X)\,p(X)$

Remarks about Likelihood

The same likelihood function, $p(Y|X)$, governs both frequentist and Bayesian statistics

- It describes the random aspects of the random process
- Must be known in either case

Subtle difference of interpretation

- Frequentist view considers all possible outcomes Y for one or more theories X
- Bayesian view takes the outcome Y to be fixed, so that the likelihood is a function of the theory X
- Note that the evidence $p(Y)$ involves the prior and likelihood for all theories (but only for the one outcome that was actually obtained)

Even in the frequentist view, we don't actually have to be able to repeat the experiment; we just have to understand what is random in the experiment

The Importance of the Prior

The Bayesian approach requires making an assumption about the prior probabilities of the different theories

- Even theories which differ only by the adjustment of one or more parameters

This can be troublesome

- Discrete set of theories: assign all of them equal prior probabilities?
- Continuum of theories: assign uniform prior probability density for all values of the parameter that relates the theories?

Conclusions drawn from an experiment can be strongly influenced by the prior

- Sometimes just look at the likelihood ratio (or Bayes factor) for two theories, to see how much the experimental data favors one vs. the other, without involving the priors

Example
(originally posed by Graham Woan)

A gravitational wave detector may have detected a gravitational wave burst from a Type II supernova. But since burst-like signals in a detector can also be produced by instrumental glitches—in fact, only 1 out of 10,000 bursts in the data are really due to a supernova—the data are checked for glitches using an auxiliary veto channel test.

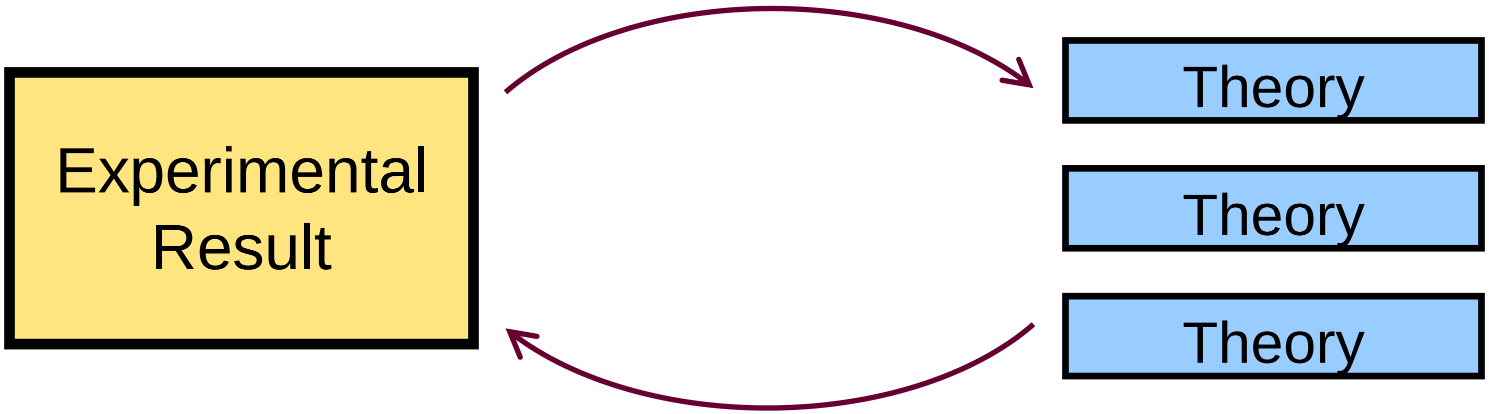
From Monte Carlo simulations, one finds that the veto channel test confirms that the burst is due to a supernova 95% of the time if there really was a GW burst in the data; but falsely claims the that the burst is due to a supernova 1% of the time, when there was no GW burst in the data.

It turns out that the measured burst passes the veto channel test. What is the probability that it's due to a supernova?

What to Conclude from an Experiment

► Bayesian ►

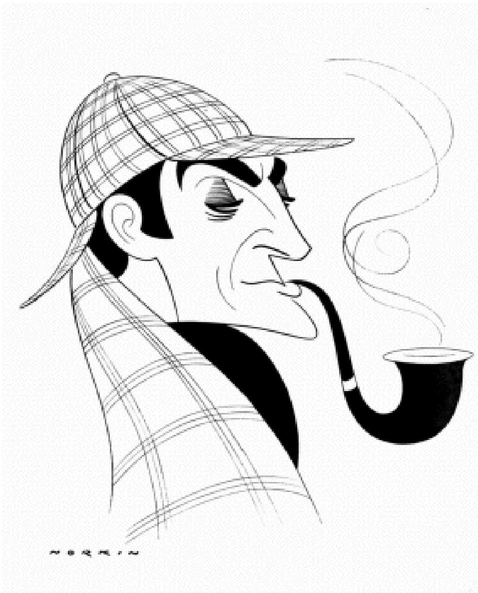
How does this result change my belief about what theories are the most probable?



For what theories is this a likely result?

◄ Frequentist ◄

Why Be a Frequentist?



“We know that he did not come through the door, the window, or the chimney. We also know that he could not have been concealed in the room, as there is no concealment possible. When, then, did he come?”

— Sherlock Holmes, in *The Sign of the Four*

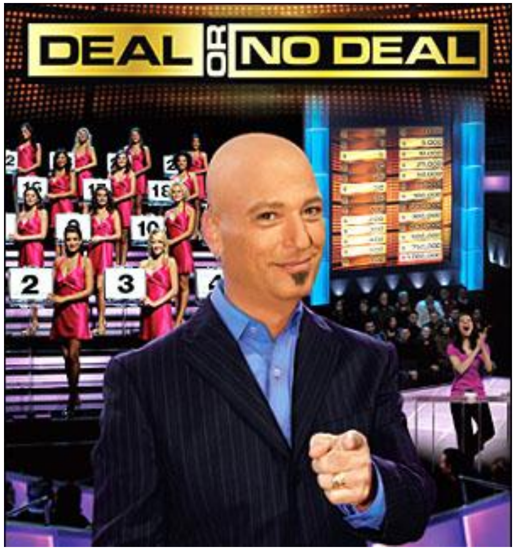
Print by Sam Norkin, www.samnorkin.com

The frequentist approach allows you to **rule out** (with some confidence) **theories which are unlikely to have produced the observed result**, **each theory judged independently**

You’re not required (or allowed) to assign a probability to any given theory being the correct one

You’re not required to consider *all* possible theories

Why Be a Bayesian?



In various situations, we maintain a certain level of belief in a variety of theories, and our beliefs change as we gather new information

What is the probability that a certain suitcase, selected in advance, contains $1,000,000 ?

The Bayesian approach allows you to **judge the probability of each theory to be the correct theory**, using data along with prior judgment

Very natural to incorporate results from a sequence of experiments

Bayes' theorem tells you how to update your beliefs, but not how to assign probabilities to different theories *a priori*

To get an *absolute* probability, you must consider *all viable* theories

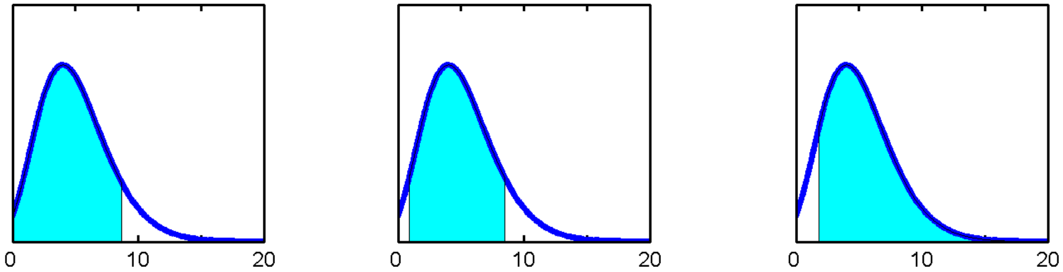
Frequentist Confidence Intervals

Summarize experimental result with a scalar *statistic*

Number of events above a fixed amplitude threshold, amplitude of the "loudest" event, … In principle, could choose anything

Determine what values of this statistic are **likely vs. unlikely** to be produced in the case of each candidate theory

Have a choice about what to focus on: unusually high values, lowest-probability values, unusually low values, Feldman-Cousins ordering principle, …

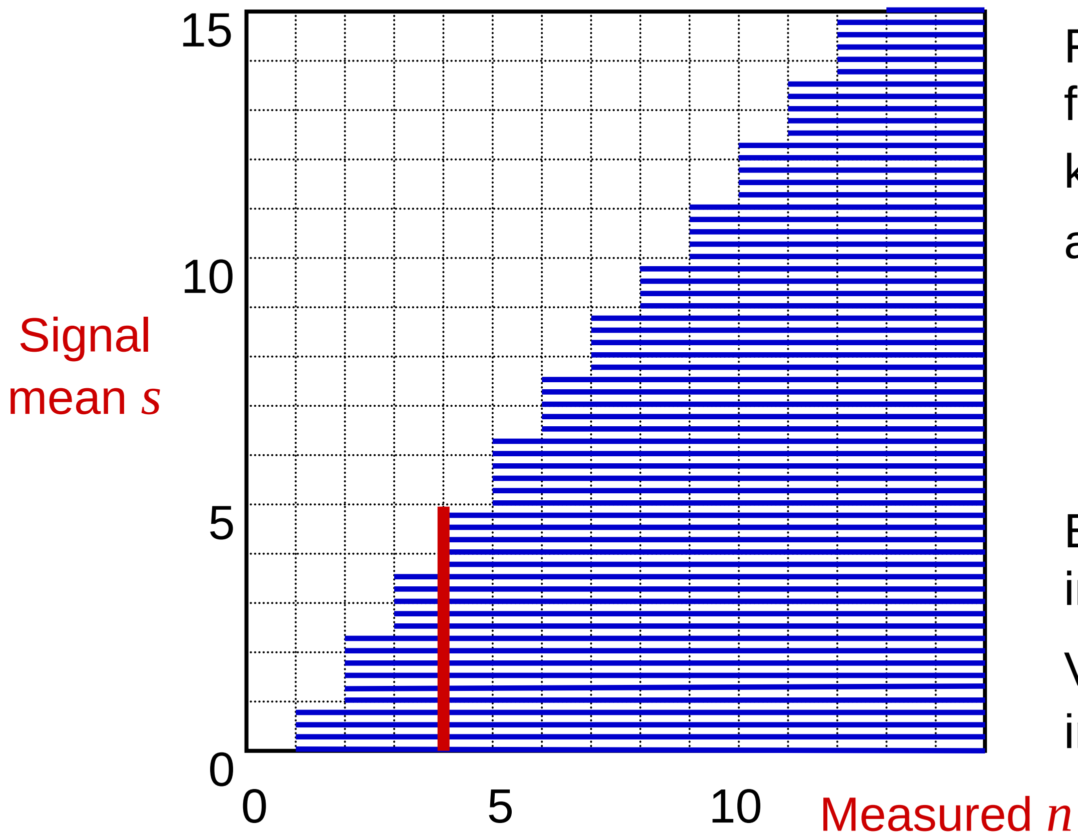


Typically group 90% as "likely", other 10% as "unlikely" (or 95% / 5%, or 99% / 1%)

A frequentist confidence interval for a given experimental result is a set of theories (i.e., range of parameter values) for which that result was likely

Upper Limit Confidence Intervals

(high values are considered likely)



For a "counting experiment" for a Poisson process with known background b = 3 and signal mean s

$$P(n) = \mu^n e^{-\mu} / n!$$

where $\boldsymbol{\mu = s + b}$

Each horizontal blue bar indicates the 90% "likely" results

Vertical bar is the confidence interval to use for a given n

If n=4 events are observed, the confidence interval is [0.00,4.99]

If n=0 events are observed, the confidence interval is **empty**,
i.e. n=0 is not a likely (at 90%) outcome for *any* signal mean s when b=3 !

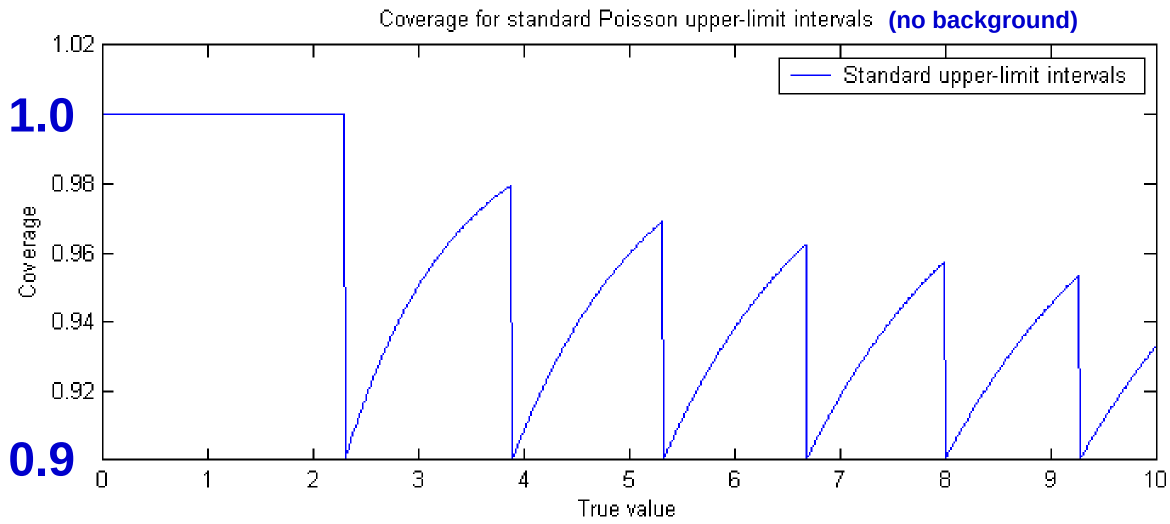
Coverage

***Coverage* is the fraction of the time that the interval assigned contains the true theory**

e.g. the true signal mean *s*

Coverage is a property of the interval-setting procedure, not of any particular experimental result

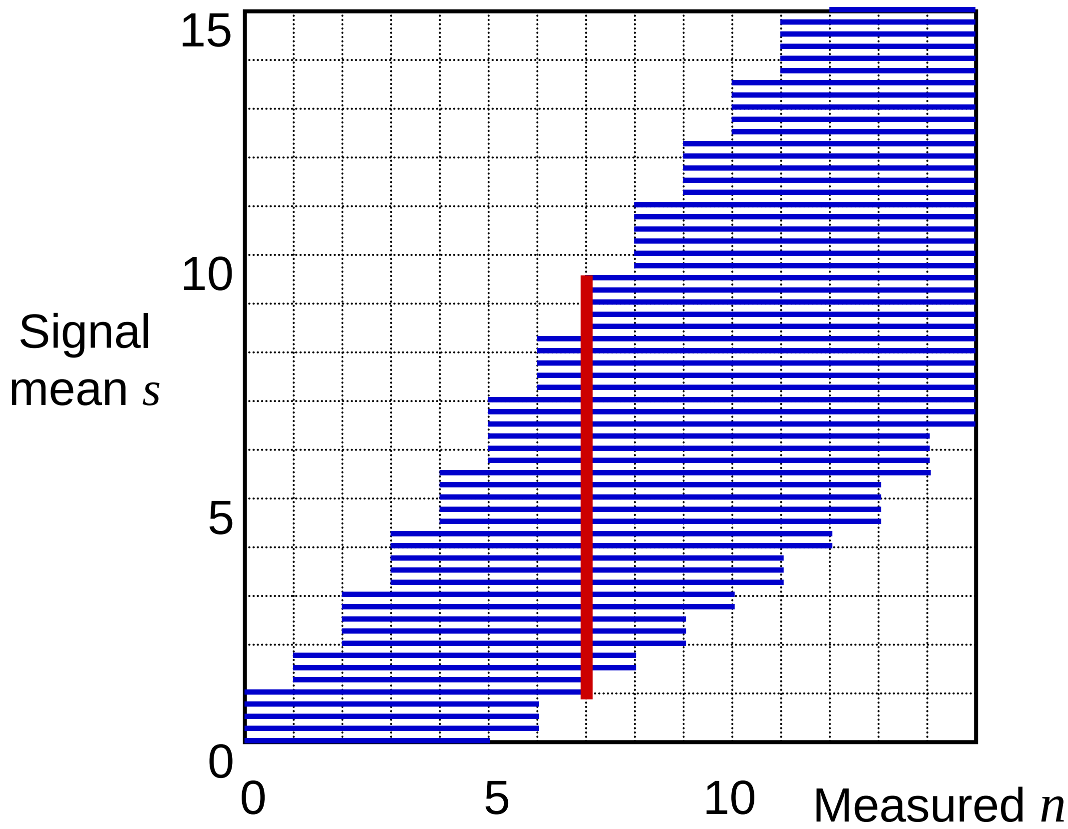
Coverage may depend on the true theory



The important thing is the minimum coverage over all possible true theories

90% in this case, by construction

Feldman-Cousins Confidence Intervals
("most likely" values are considered likely)



For a counting experiment for a Poisson process with known background b = 3

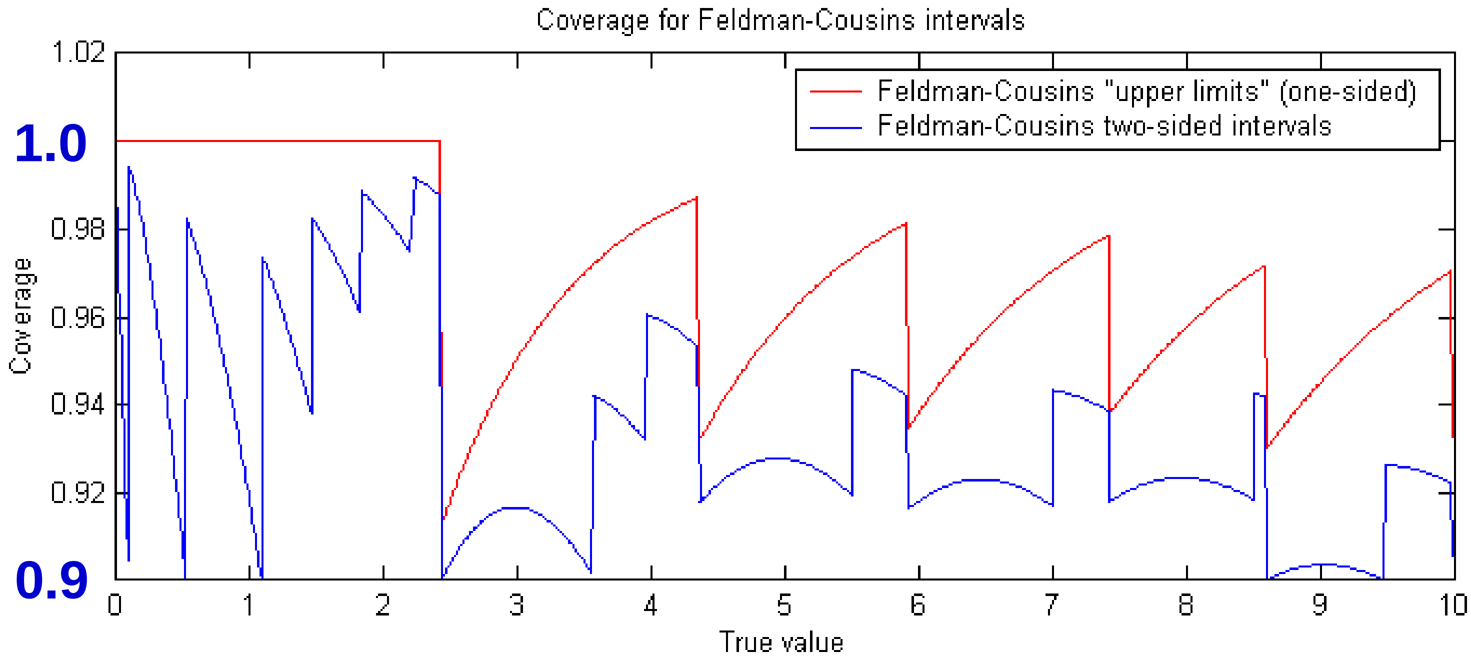
Gary J. Feldman and Robert D. Cousins, "Unified approach to the classical statistical analysis of small signals", Phys. Rev D 57, 3873 (1998).

This ordering principle groups some unusually high and some unusually low values together as "unlikely"

If n=7 events are observed, the confidence interval is [0.89,9.53]

Note that even if the signal mean s is zero, this experiment will produce an interval excluding zero 8.4% of the time

Feldman-Cousins Coverage



Feldman-Cousins intervals (blue curve) satisfy 90% minimum coverage, by construction

Sometimes people use "Feldman-Cousins upper limits", using only the upper end of the interval even if the lower end is nonzero

Those over-cover for all true values (red curve) !

Loudest Event Statistic

Essentially a counting experiment with the threshold dynamically set to be infinitesimally above the amplitude of the highest-amplitude event ($\rho_{\max}$)

This is a legitimate frequentist procedure !

Rate per galaxy:

$$\mathcal{R} < \mathcal{R}_{90\%} = \frac{2.303 + \ln P_b}{T N_G(\rho_{\max})}$$

Probability that all background events have $\rho < \rho_{\max}$

Observation time

Number of galaxies within range

If ignore background (*i.e.* take P_b=1), then limit is conservative

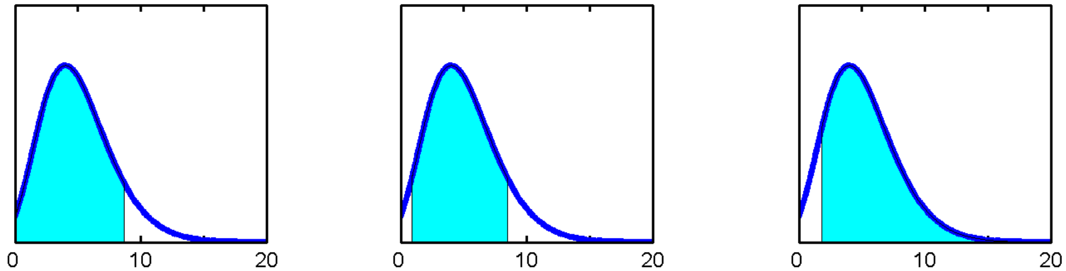
If include background, then there is some chance of getting:

- An empty interval, if $P_b < 0.10$
- An upper limit which is misleadingly low
  e.g. if P_b = 0.12, then $\mathcal{R}_{90\%} = 0.18 / [T\, N_G(\rho_{\max})]$

Bayesian Parameter Estimation and Confidence Intervals

In a family of theories, the "best" theory is taken to be the one which maximizes the posterior probability

- If the prior was uniform, then this is also the theory with a maximum likelihood
- Sometimes involves marginalizing over "nuisance parameters"



A Bayesian confidence interval is a set of theories which has a specified probability (e.g. 90%) of containing the true theory

Compare to:

A frequentist confidence interval for a given experimental result is a set of theories for which that result was likely

Notes on Upper Limits, etc.

Always based on a *population* of sources

- Parametrized in a physical (or non-physical) way
- Confidence intervals / limits are placed on regions of the parameter space

Study using Monte Carlo simulations

- Add simulated signals to real data and re-run the analysis to see how many are detected

Desirable to do the analysis “blind” until the analysis details are frozen

- Study background and simulated-signal samples, but not the *real* sample
- Avoids the possibility of human bias that could, in principle, make the stated upper limit invalid
- If human judgment is involved, it’s difficult to predict how the human(s) would have behaved if some other outcome had occurred

Evaluating Detection Efficiency

Test / tune searches using simulated signals

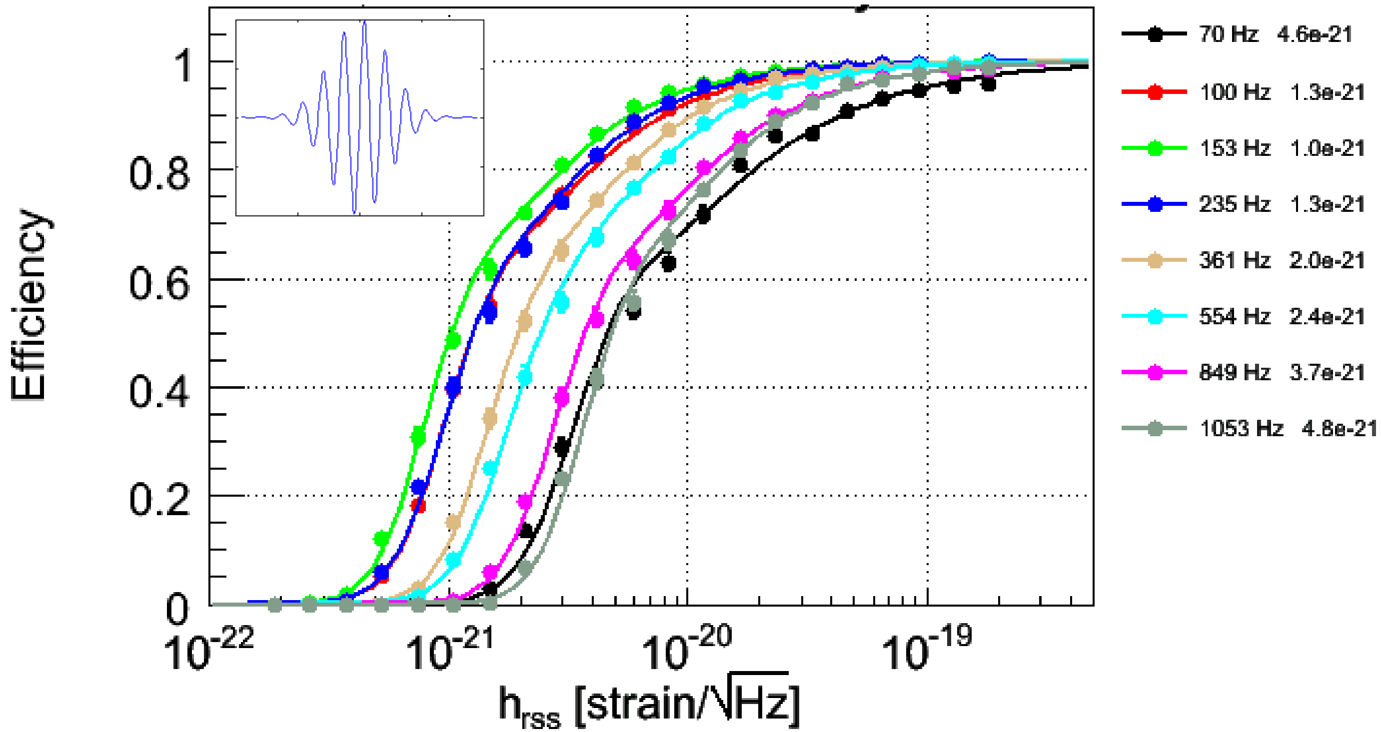
Astrophysically modeled...

or *ad hoc*, e.g. “Sine-Gaussians”

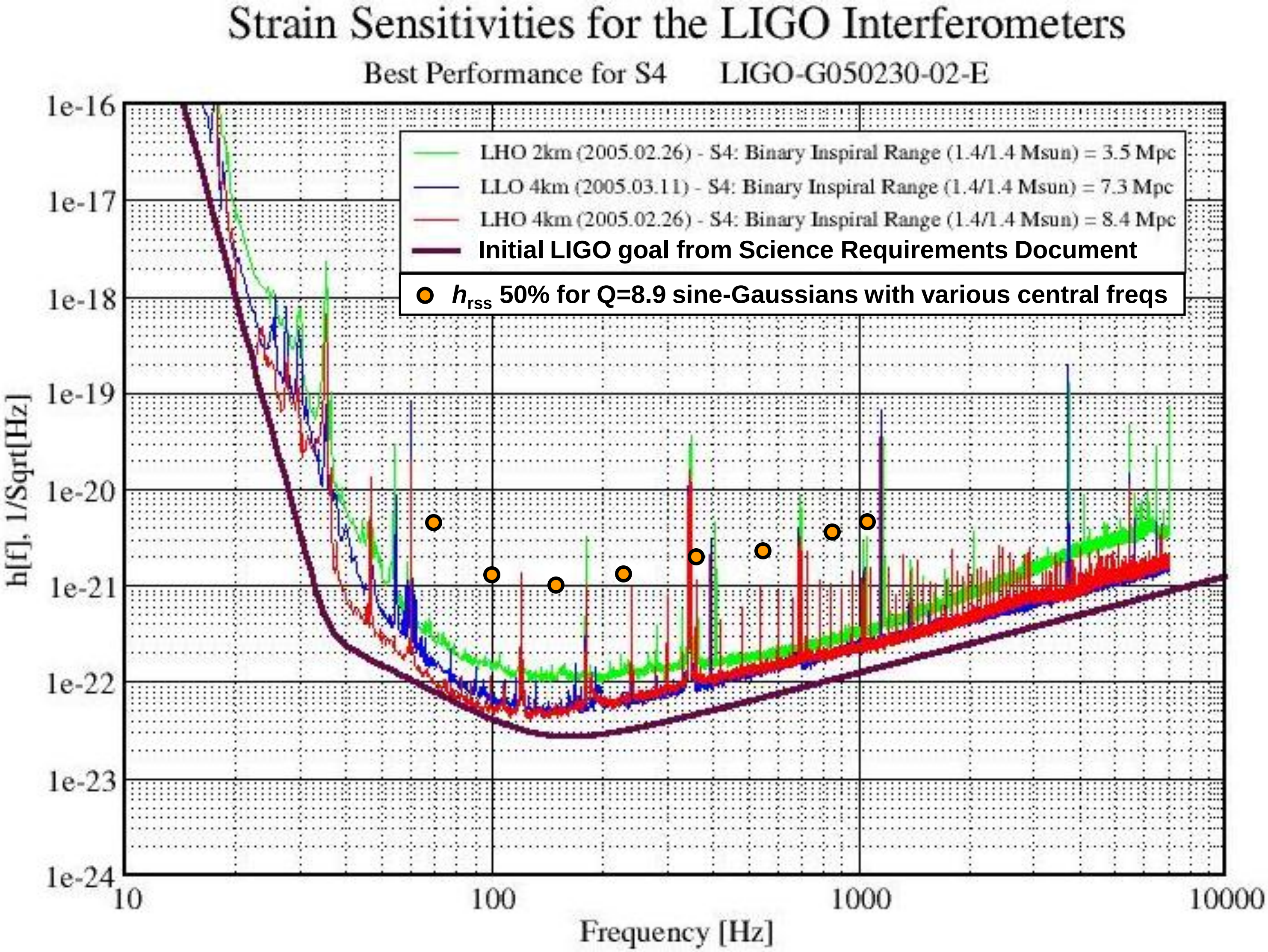
$h(t) = h_0 \sin(2\pi ft) \exp(-2(\pi ft/Q)^2)$ $\quad$ $h_{rss} = h_0 (Q/4f)^{1/2} / \pi^{1/4}$

Linearly polarized; random sky position & polarization angle

f =



Frequency Dependence of Sensitivity



Exclusion Regions

Example from LIGO-Virgo all-sky burst search:
Parameter space is rate vs. signal strength