Statistics for Burst/Transient Signals:

Setting Limits

and

Making Discoveries

Peter Shawhan
Caltech

Statistics for Gravitational Wave Data Analysis (GravStat)
State College, Pennsylvania, 20 May 2005

LIGO-G050266-00-Z

Outline

Gravitational wave transients

Data reduction

- Selection of event candidates
- Summarizing the outcome with a statistic
- Interpretation of the statistic

Experiment / analysis design

- Background strategy considerations

Statistical interpretation

- Frequentist confidence intervals
- Bayesian posterior pdf
- Maximum-SNR analysis
- Distributional analysis

Astrophysical interpretation

Questions for discussion

Gravitational Wave Transients

Signals with “short” duration

Waveform may be known* or unknown

Binary inspiral	Binary merger
Cosmic string cusp / kink	Core collapse supernova
Black hole ringdown	???

** “Known” waveforms depend on one or more physical parameters*

Physically identical *sources* may produce a distribution of observed *signals*

Due to different sky positions, orientations, distances

A *population* of sources may have a range of physical parameters

… as well as a spatial distribution, of course

Gravitational Wave Transients

We normally assume a Poisson process

i.e. events arrive at random times and are independent of each other

A given class of sources/signals is described by a rate

This talk will assume that we are focusing on either discovering GW signals or setting limits on their rate

Eventually, we hope to extract astrophysical information (properties of sources or populations) from observed signals

Selection of Event Candidates

Identify regions of interest in the time series data

- When some sort of linear or non-linear filter exceeds a threshold, record the information as a trigger
- May be a GW signal or a false alarm (*e.g.* noise fluctuation)

Typically use a low threshold initially to produce triggers, then apply additional tests to try to separate GW signals from false alarms

- χ^2 or other detailed waveform tests
- Noise stationarity tests
- Coincidence among two or more detectors
- Cross-correlation of data streams

Cuts define event candidates

- May be mostly (or entirely) background

Summarizing the Outcome

Want to infer something about the number of GW signals in the set of event candidates

Know something about the background from other studies, *e.g.* time-shifted coincidences

Define a statistic, *i.e.* a scalar summary (random variable) which has some power to distinguish background-plus-signal from background alone

Number of event candidates with SNR above some threshold

Maximum observed SNR

Consistency of observed SNR distribution with SNR distribution expected for background alone

Or something else…

Interpretation of the Statistic

The value of the statistic may be used to:

- Calculate the *p*-value (prob. that the statistic would equal or exceed the observed value under the null hypothesis of background alone)
- Assign a frequentist confidence interval on the signal rate
- Calculate a Bayesian posterior pdf, and perhaps derive a credible interval for the signal rate

Cleanest evidence for discovery of gravitational waves is a *p*-value below some threshold

- What should that probability threshold be?
- May depend in part on our prejudice about the event rate

An upper limit is the upper bound of a confidence interval with a lower bound at zero

Background Strategy Considerations

Discovery requires a statistically significant excess of signal above the background

- If background expectation is substantial (say, 0.1 events or higher), then we may need multiple signal events to constitute a statistically significant excess
- If background is very low, then even a single signal event may have high statistical significance

Background estimate typically has some uncertainty

- When setting an upper limit on signal rate, it's conservative to assume that the background is on the *low* side of the estimate, perhaps even zero
- If background is completely unknown, can still set an upper limit on the signal rate

Frequentist Confidence Intervals for Counting Experiments

Set fixed thresholds on all quantities before looking at the data

- *e.g.* to achieve a certain background rate

Count number of event candidates surviving all cuts

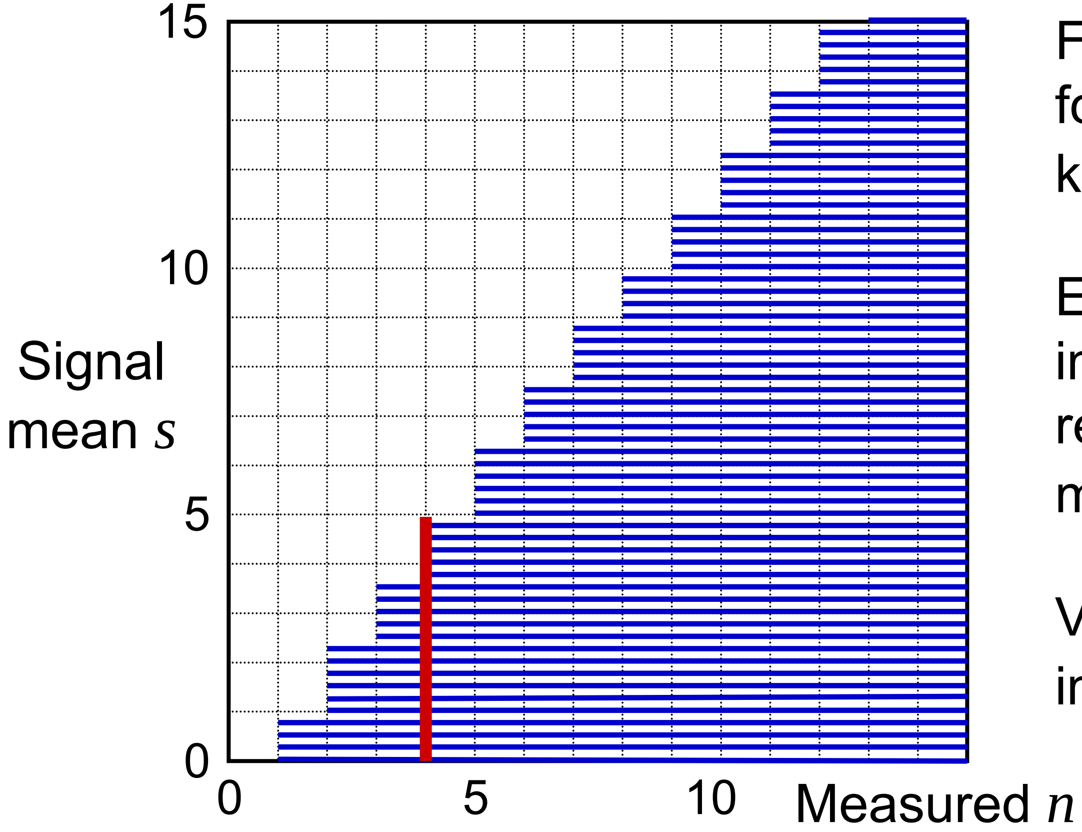
A frequentist confidence interval is a set of models for which the observed outcome was "likely"

- Here, models are parametrized by mean rate

Choice of confidence intervals

- Have freedom (≡ arbitrariness) to choose ordering principle which defines, for each model, what outcomes are "likely" at specified C.L.
- Upper-limit confidence intervals vs. Feldman-Cousins vs. …
- In each case, intervals can be constructed to give specified minimum coverage

Upper-Limit Confidence Intervals



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

Each horizontal blue bar indicates the 90% "likely" results for a given signal mean s

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

If n=4 events are observed, the 90% 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 model**

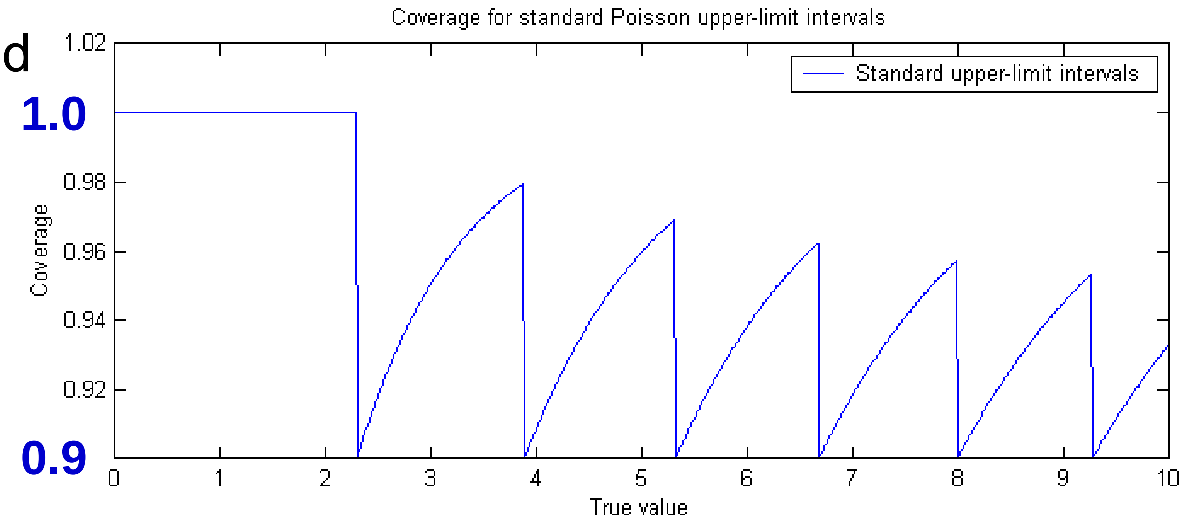
e.g. the true signal mean s

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

Coverage for Upper-Limit Confidence Intervals

Coverage may depend on the true model

(This coverage plot is for the zero-background case)



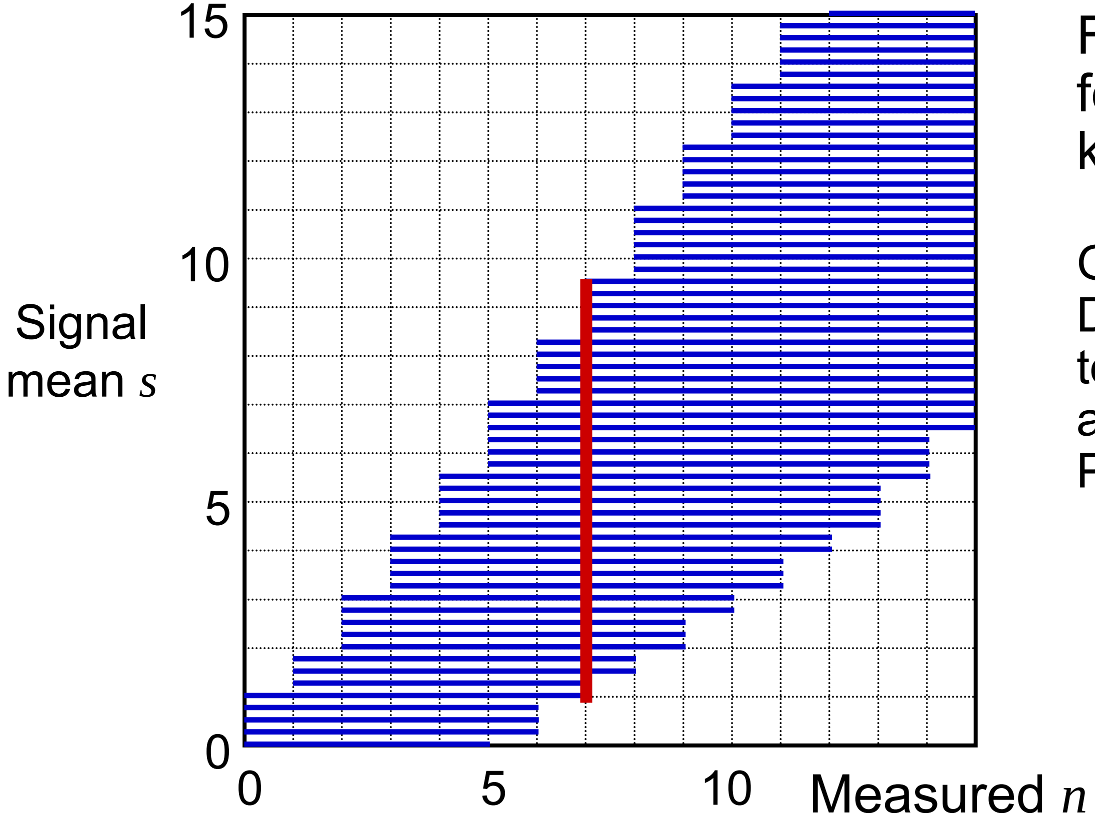
The important thing is the minimum coverage over all possible true models

90% in this case, by construction

IF n=0, there is a clean statement of the meaning of the upper limit $R_{90\%}$:

“For all $R \geq R_{90\%}$, there is at least a 90% chance of observing at least one event in an experiment equivalent to this one”

Feldman-Cousins Confidence Intervals



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

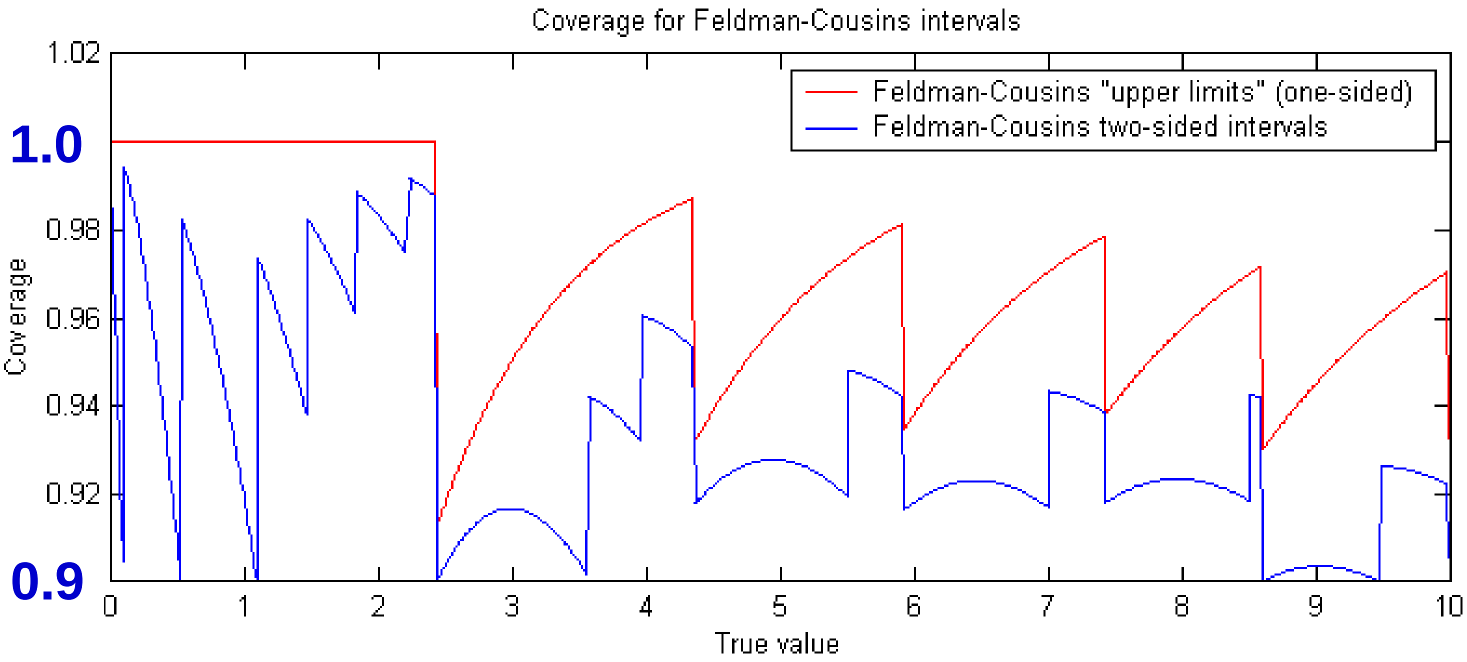
Gary J. Feldman and Robert D. Cousins, “Unified approach to the classical statistical analysis of small signals”, Phys. Rev D 57, 3873 (1998).

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

If n=7 events are observed, the 90% 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 — not in itself a “discovery”

Feldman-Cousins Coverage



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

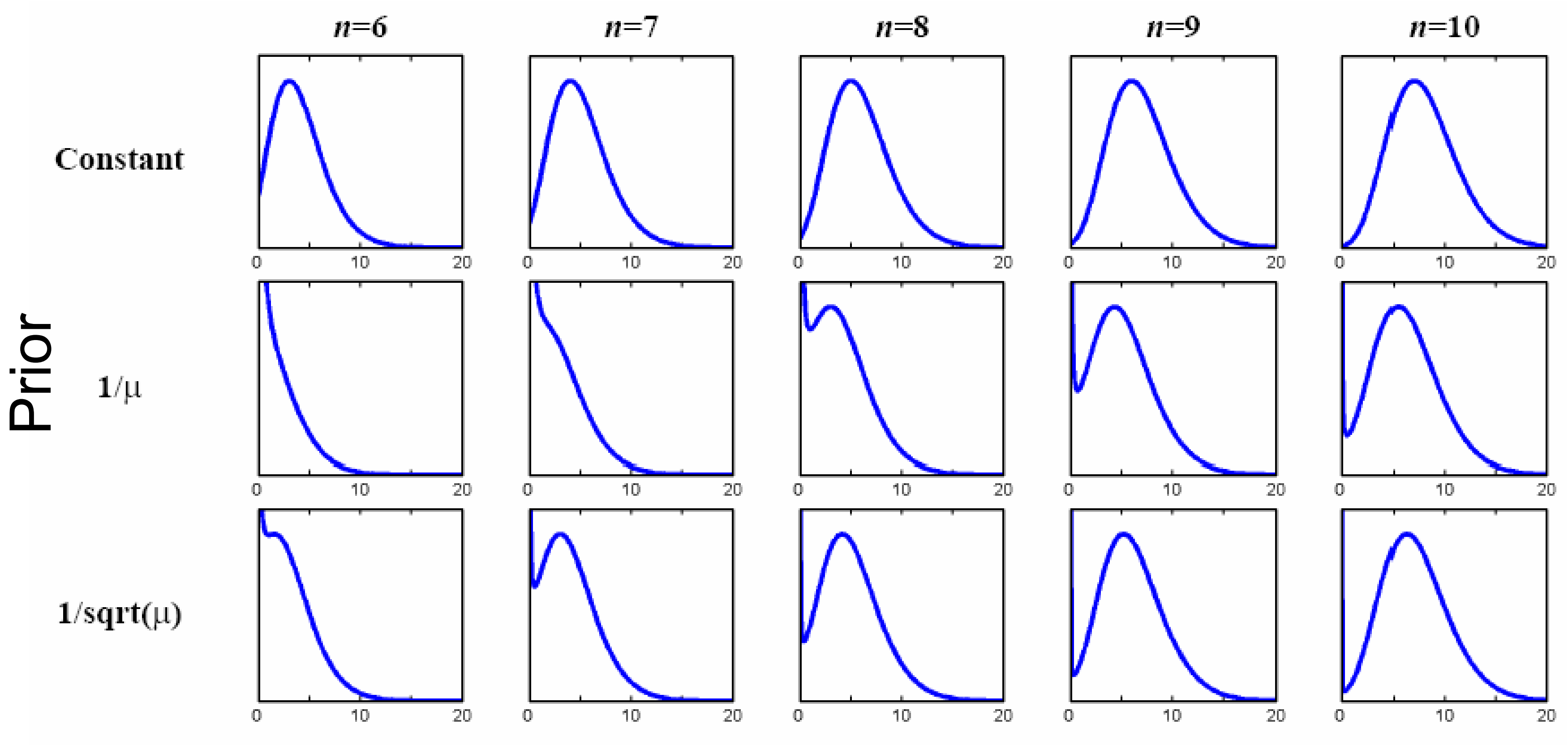
“Feldman-Cousins upper limits” (red curve) over-cover for all true values

- Not the use for which the intervals were constructed

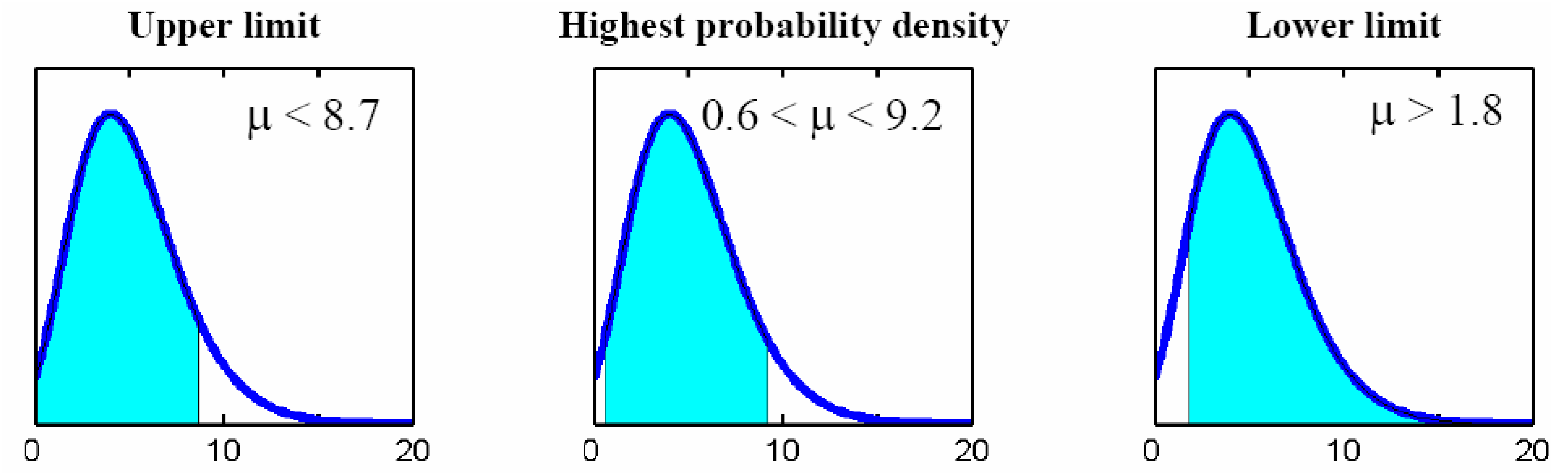
Bayesian Posterior pdfs for Counting Experiments

A discrete measurement (event count) gives a pdf according to the prior

Example: Poisson process with background b=3



Choice of Credible Interval



Maximum-SNR Statistic

Brady, Creighton, and Wiseman, Class. Quant. Grav. 21, S1775 (2004)

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

This is a legitimate frequentist procedure !

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

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

If ignore background (*i.e.* take $\ln P_b = 0$), 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 $R_{90\%} = 0.18 / [T N_G(\rho_{\mathrm{max}})]$

Distributional Analysis

Are the observed distribution and the distribution measured for background alone consistent with being drawn from the same parent distribution?

Need a statistic to evaluate this

- Familiar: Kolmogorov-Smirnov
- Better: Mann-Whitney

Astrophysical Interpretation of a Rate Limit

At least three possible interpretations, depending on what is assumed about the sources

1. Limit on rate of *detectable* signals

A statement about arbitrary signals, not about sources

2. Limit on rate of a particular signal or source as a function of amplitude

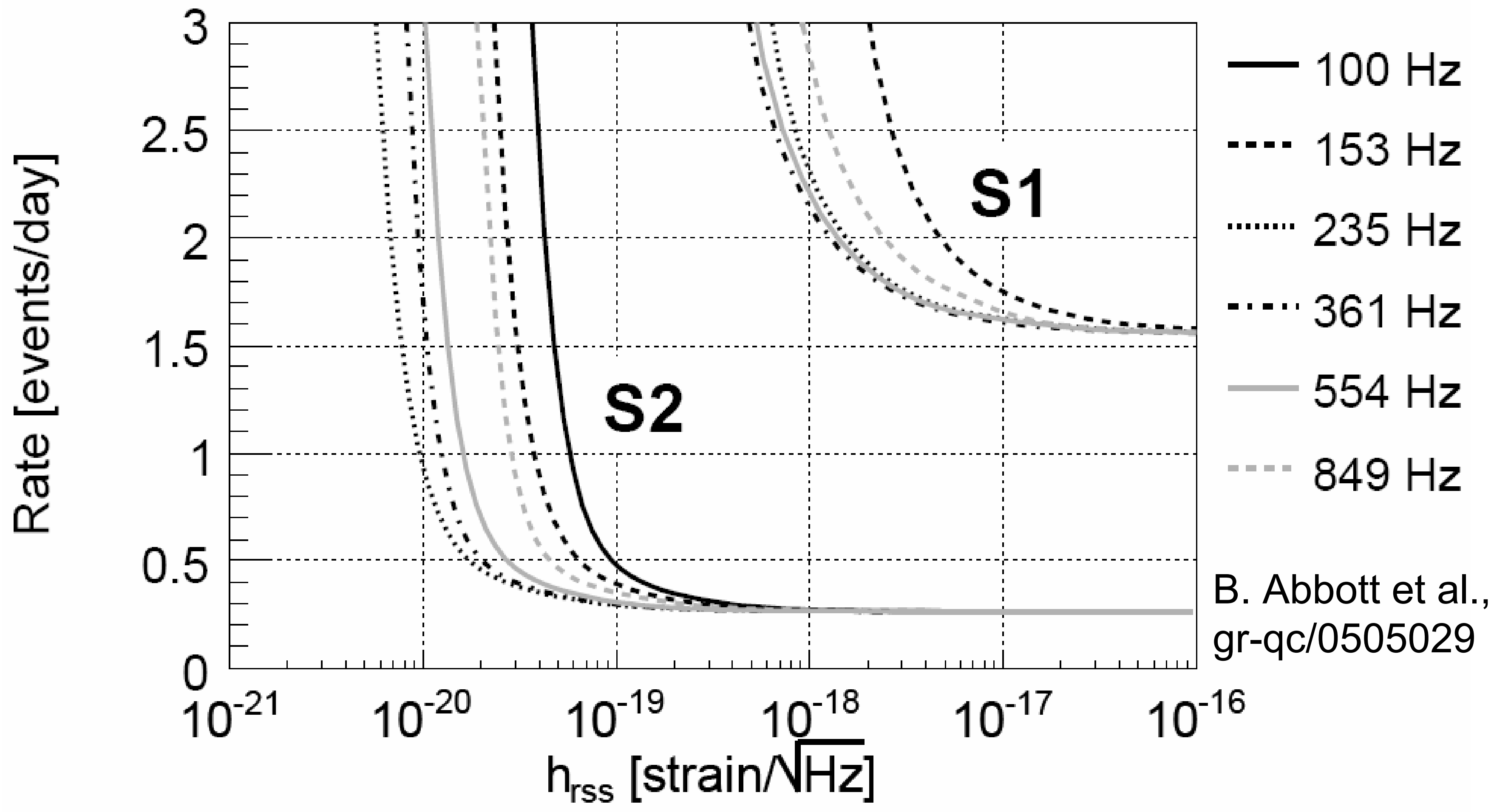
Requires detection efficiency to be evaluated for signal/source in question

Most reliable approach is to use Monte Carlo (simulated signal waveforms added to data)

Monte Carlo may use fixed signals, or may generate sources with random sky position and orientation and project onto detector

Rate-strength diagram

Rate-Strength Exclusion Diagrams



B. Abbott et al., gr-qc/0505029

Limit on rate supposing all signals have some fixed amplitude

Astrophysical Interpretation of a Rate Limit

3. Aggregate rate of a population of sources

Requires a complete model for the population

e.g. LIGO S2 binary neutron star inspiral search

[B. Abbott et al., gr-qc/0505041]

Used:

- Spatial distribution of sources in the Milky Way and other galaxies
- Joint distribution of masses of binary components

Placed limit on rate of binary neutron star inspirals
(with component masses between Msun and 3 Msun)
per Milky Way equivalent galaxy

Questions for Discussion

- **Do we really need to calculate a statistic to learn something from a set of event candidates?**

  What about a Bayesian analysis using all the information associated with the set of event candidates?

- **Is there another good way to discover a GW signal, other than through a *p*-value?**
- **What is an appropriate background rate to aim for?**
- **Why do a frequentist analysis? A Bayesian analysis? A maximum-SNR analysis?
  A distributional analysis?**
- **Should we use one-sided or two-sided frequentist confidence intervals?**