Detections Limits, Upper Limits, and Confidence Intervals in High-Energy Astrophysics

David van Dyk

Department of Statistics
University of California, Irvine

Joint work with
Vinay Kashyap, The CHAS Collaboration, and SAMSI SaFeDe Working Group

Outline of Presentation

This talk has two parts.

1. **Soap Box**
    - What do short Confidence Intervals Mean?
    - Is there an all-purpose statistical solution?
    - Why compute Confidence Intervals at all?
2. Illustrations from High-Energy Astro-statistics
    - Detection and Detection Limits
    - Upper Limits for undetected sources
    - Quantifying uncertainty for detected sources
    - Nuisance parameters

What is a Frequency Confidence Interval?

Background contaminated Poisson

$$N \sim \text{Poisson}(\mu + b),$$

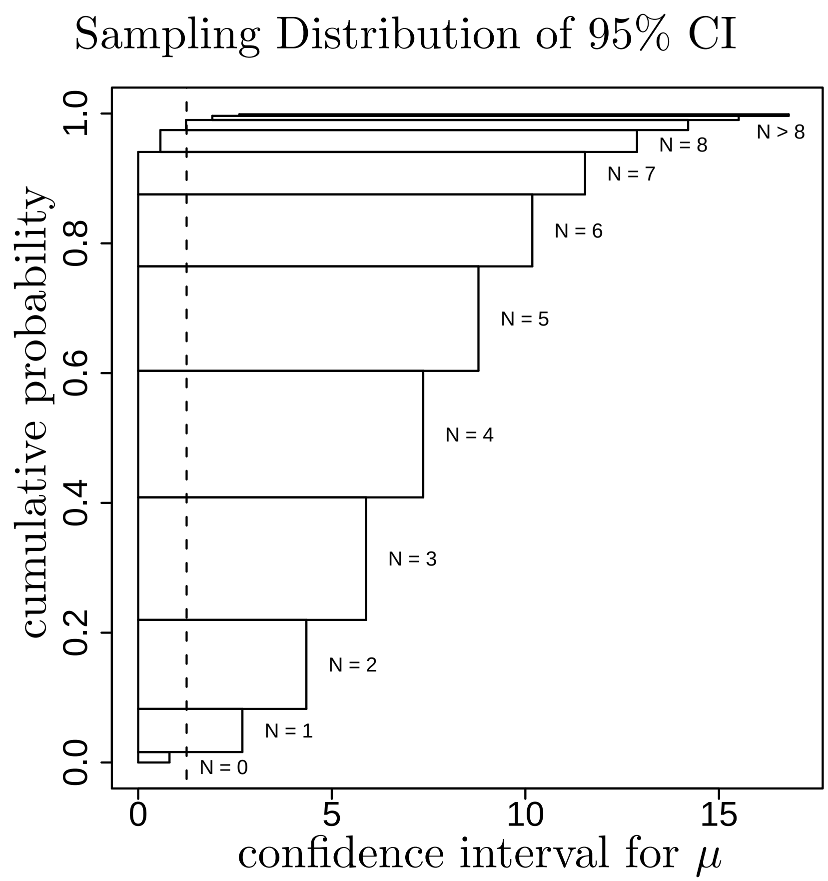
with $b = 2.88$ and $\mu = 1.25$.

- Confidence Interval:

$$\{\mu : N \in \mathcal{I}(\mu)\},$$

  where $\Pr(N \in \mathcal{I}(\mu)|\mu) \geq 95\%$.

- Values of μ with propenstity to generate observed N.



The CI gives values of μ that are plausible given the observed data.

Short or Empty Frequency Confidence Intervals

What do they mean?

There are few values of μ that are plausible given the observed data.

What they do not mean?

The experimental uncertainty is small. (Std Error or Risk of $\hat{\mu}$??)

What if CI are repeatedly short or empty?

Short intervals should be an uncommon occurrence. If they are not the statistical model is likely wrong, regardless of the strength of the subjective prior belief in the model.

Pre-Data and Post-Data Probability

- A Frequency Confidence Interval has (at least) a 95% chance of covering the true value of μ.
- What is the chance that $\emptyset$ contains μ?

There is a 95% chance that Confidence Intervals *computed in this way* *contain the true value of* μ.

- Frequency-based probability says nothing about a particular interval.
- Bayesian methods are better suited to quantify this type of probability.
- Precise probability statements may not be the most relevant probability statements.

Our intuition leads us to condition on the observed data.

Leaning on the Science Rather than on the Statistics

Desirable Properties of Confidence Intervals Include (Mandelkern, 2002):

1. Based on well-define principles that are neither arbitrary nor subjective
2. Do not depend on prior or **outside knowledge** about the parameter
3. **Equivariant**
4. Convey experimental uncertainty
5. Correspond to precise probability statements

But....

1. Mightn't there be physical reasons to prefer a parameterization over another?
2. Why so much concern over using outside information regarding the *values* of the parameter but none over the *form* of the statistical model?

Is it easier to make a final decision on the choice of CI than on the choice of parameterization?

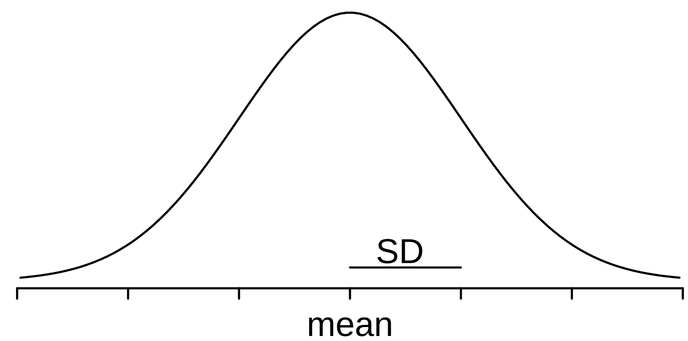
Say What You Mean, Mean What You Say

- Confidence Intervals are not designed to represent experimental uncertainty.
- Rather than trying to use Confidence Intervals for this purpose, an estimator that is designed to represent experimental uncertainty should be used.
- For example, we might estimate the risk of an estimator.

$$\mathrm{Risk}(\mu, \hat{\mu}) = \mathrm{E}_{\mu}\big[\mathrm{Loss}(\mu, \hat{\mu})\big] = \sum_{N=0}^{\infty} \mathrm{Loss}(\mu, \hat{\mu}(N))\, f(N|\mu)$$

Why Use Confidence Intervals?

- The Guassian distribution can be fully represented by *two* parameters, the mean and the standard deviation.
- No information is lost if we report a simple central 95% probability interval.
- The CLT ensures that for large samples many likelihoods and posterior distributions are approximately Gaussian.

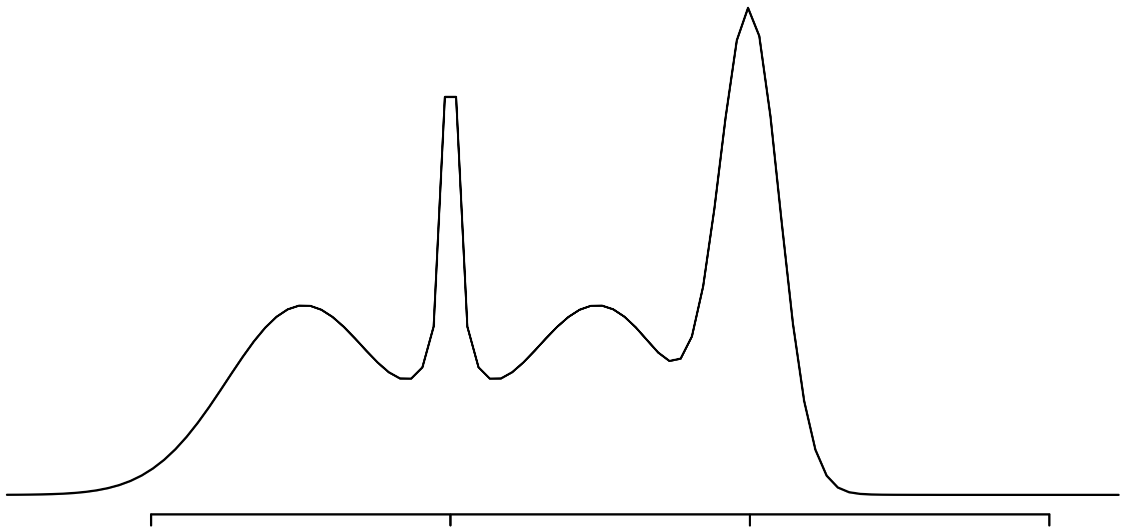


- The bell-shape of the Gaussian distribution gives a simple interpretation to such intervals.
- Analogous reasoning yields a large sample interpretation for Frequency Confidence Intervals.

The 95% Confidence Interval:

A story of Gaussian sampling and posterior distributions.

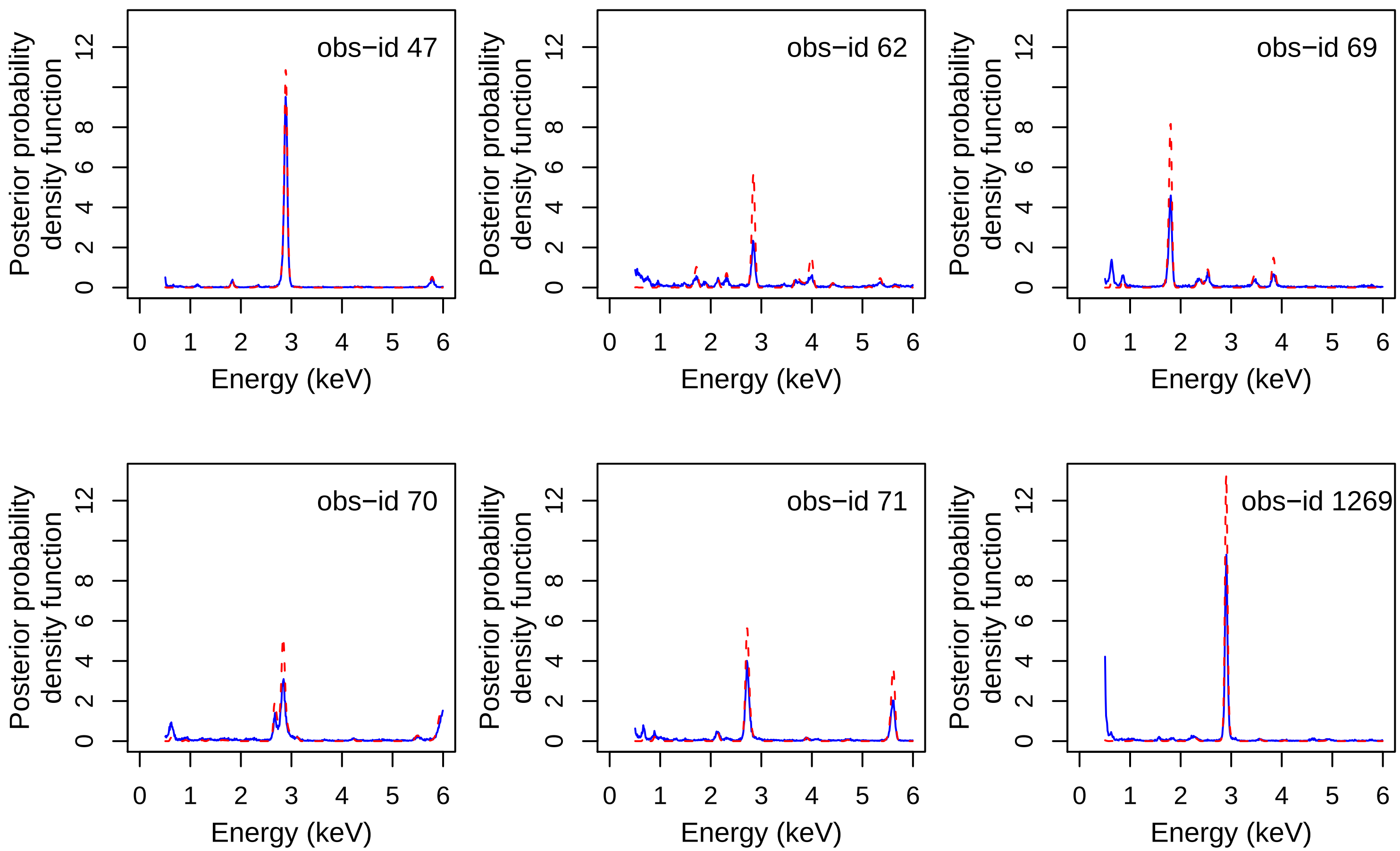
But what if....



- How do we summarize this distribution?
- Why do we want to? What is the benefit?
- With more complex distributions, simple summaries are not available.

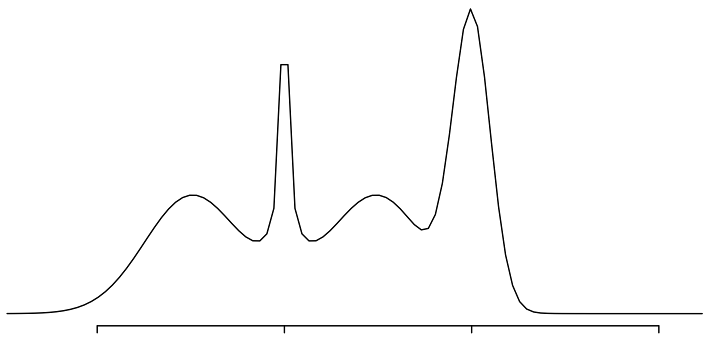
Solution: Do Not Summarize!!

Multiple Modes



An Interval is not a helpful summary!

A False Sense of Confidence



versus **[2.03, 7.87]**

- Confidence Intervals *appear* to give a definite range of plausible values.
- Plots seem wishy-washy.
- But the likelihood (or posterior distribution) contains full information.

Confidence Intervals Give a <u>False</u> Sense of Confidence.

Outline of Presentation

This talk has two parts.

1. Soap Box
   - What do short Confidence Intervals Mean?
   - Is there an all-purpose statistical solution?
   - Why compute Confidence Intervals at all?
2. **Illustrations from High-Energy Astro-statistics**
   - Detection and Detection Limits
   - Upper Limits for undetected sources
   - Quantifying uncertainty for detected sources
   - Nuisance Parameters

A Survey

- We observe a field of optical sources with the *Chandra X-ray Observatory*.
- There are a total of $n = 1377$ potential X-ray sources.
- Some correspond to multiple optical sources.
- All observations are background contaminated.
- An independent background only observation is available.

Model:

$$\begin{aligned} Y_i^S &\sim \text{Poisson}\left(\kappa_i^S(\lambda_i^S + \lambda_{j(i)}^B)\right), \\ Y_j^B &\sim \text{Poisson}\left(\kappa_j^B \lambda_j^B\right), \\ \lambda_i^S &\sim \text{Gamma}(\alpha, \beta). \end{aligned}$$

We specify non-informative prior distributions for $(\lambda_1^B, \ldots, \lambda_J^B, \alpha, \beta)$.

Detection Limits and P-values

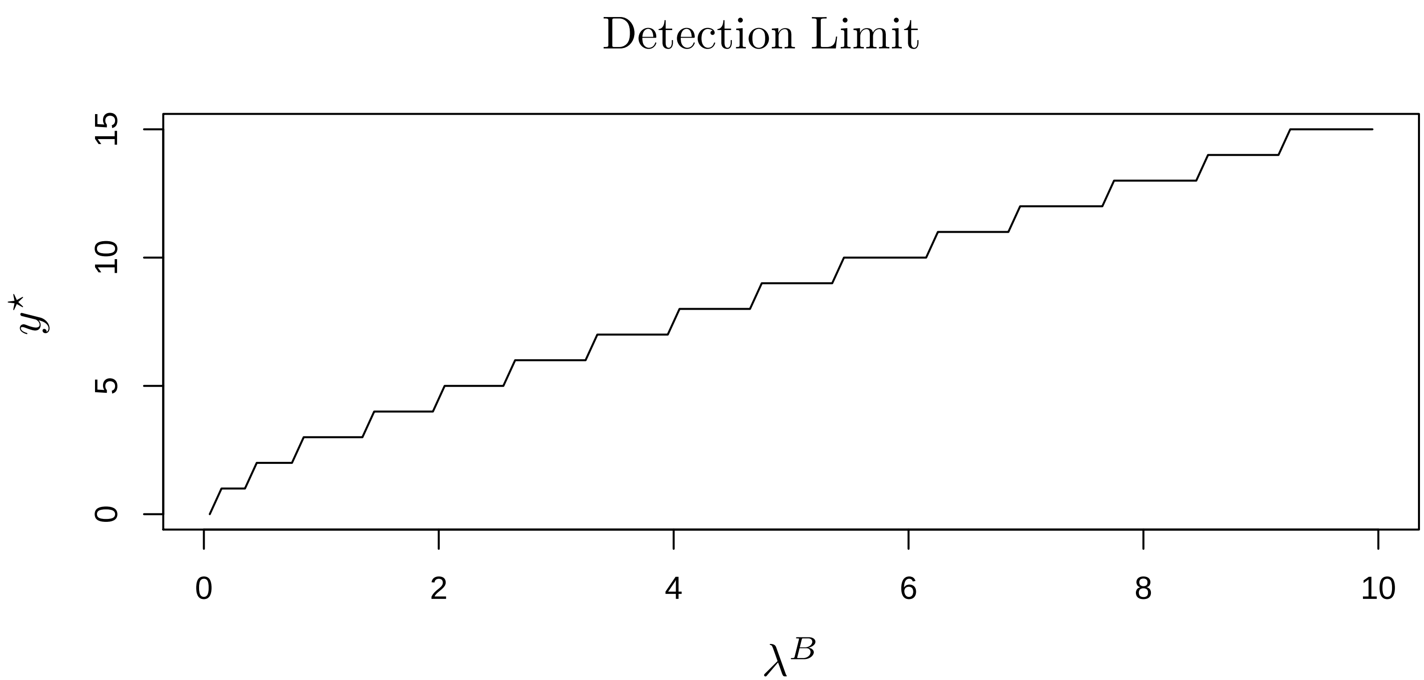
How large must y_i^S be in order to conclude there is an X-ray source?

Procedure:

1. Suppose $\lambda_i^S = 0$.
2. Use Y_j^B to compute prior (gamma) distribution of $\lambda_{j(i)}^B$.
3. Compute 95th percentile of marginal (negative-binomial) distribution of Y_i^S.
4. Set the *detection limit*, $y_i^\star$ to this percentile.
5. If $y_i^S > y_i^\star$ conclude there is an X-ray source.

A similar strategy can be used to compute a p-value.

Illustrating Detection Limits



Here $\kappa^S = \kappa^B = 1$, and λ^B is treated as known.

Upper Limits

If there is no detection, how large might λ_i^S be?

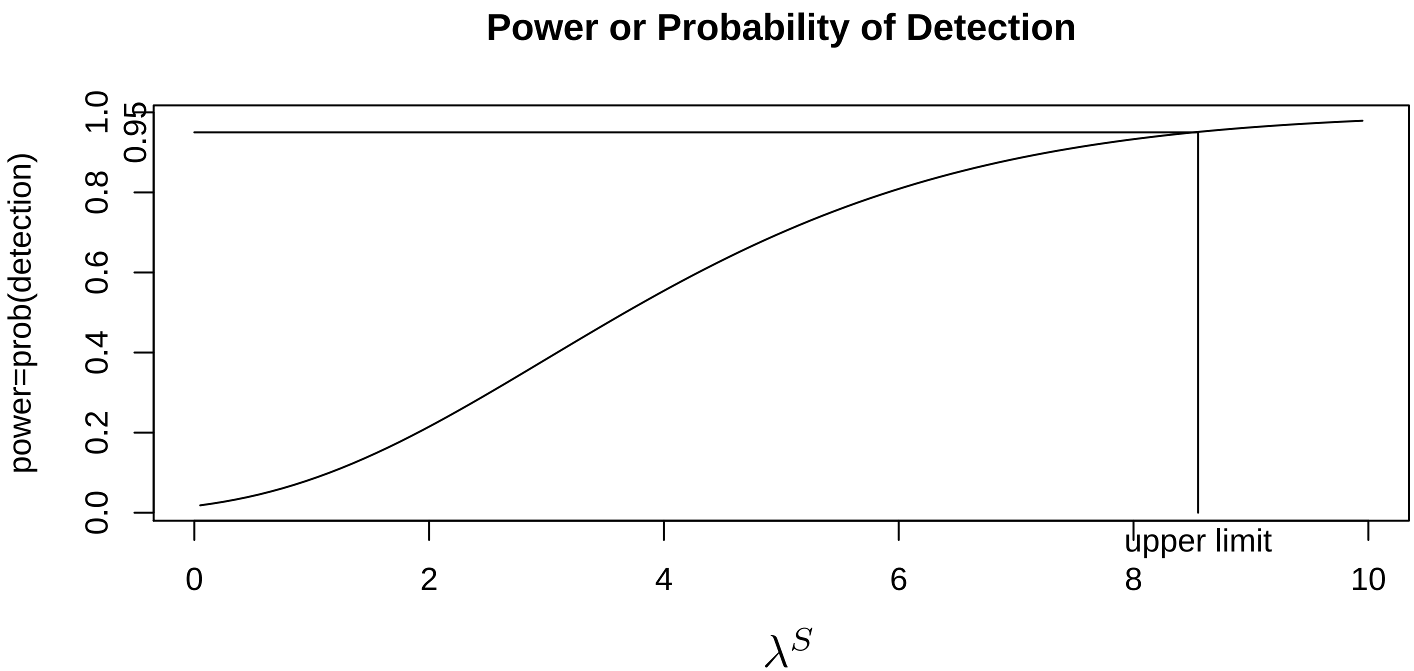
Procedure:

1. Suppose $Y_i^S \sim \text{Poisson}\left(\kappa_i^S(\lambda_i^S + \lambda_i^B)\right)$.
2. We *detect* a source if $y_i^S > y_i^\star$.
3. Let the *upper limit*, U_i^S, be the *smallest* λ_i^S such that

$$\Pr(Y_i^S > y_i^\star | \lambda_i^S) > 95\%.$$

4. Note: $Y_i^S \mid \lambda_i^S$ is the convolution of a Poisson and a negative binomial.

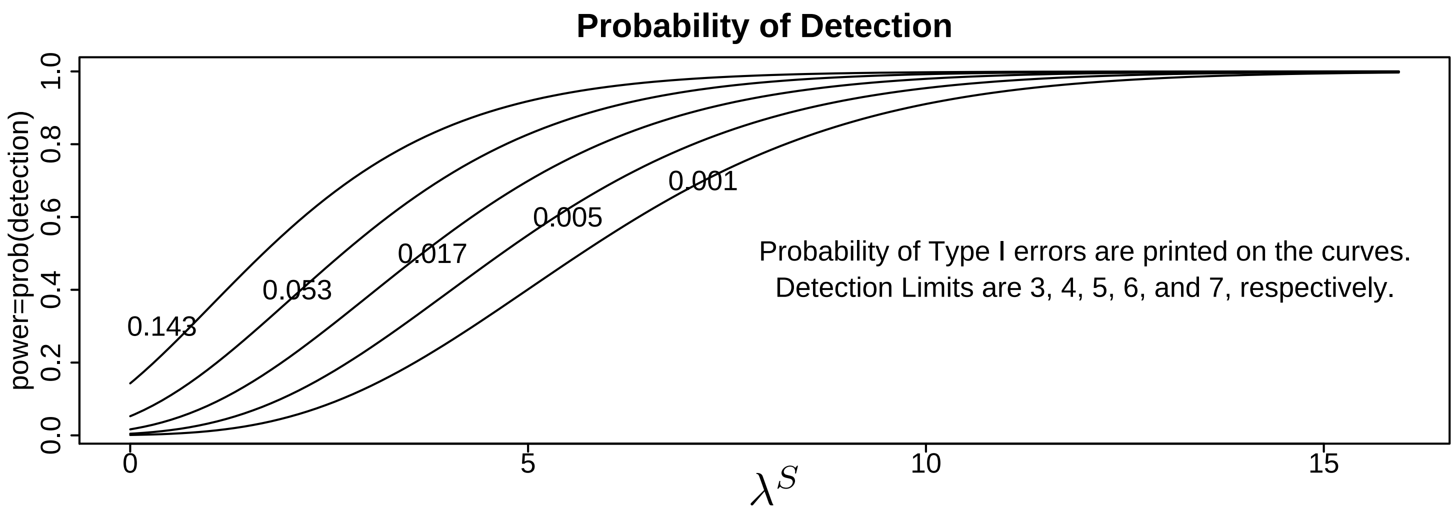
Illustrating Upper Limits

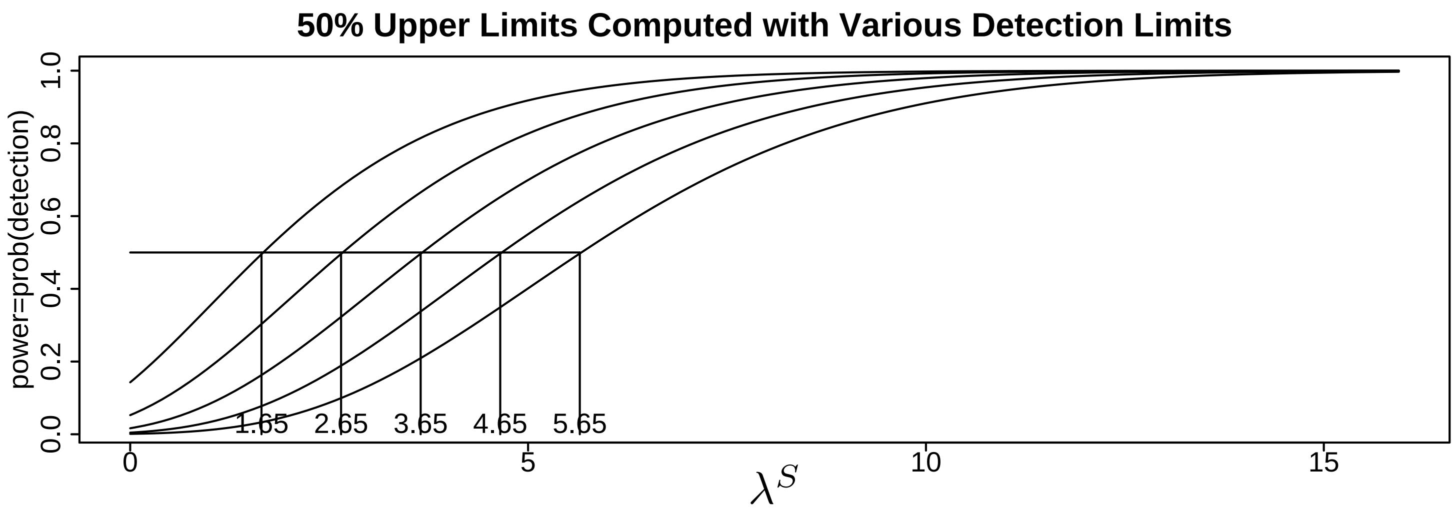


Here we assume $\lambda^B = 2$.

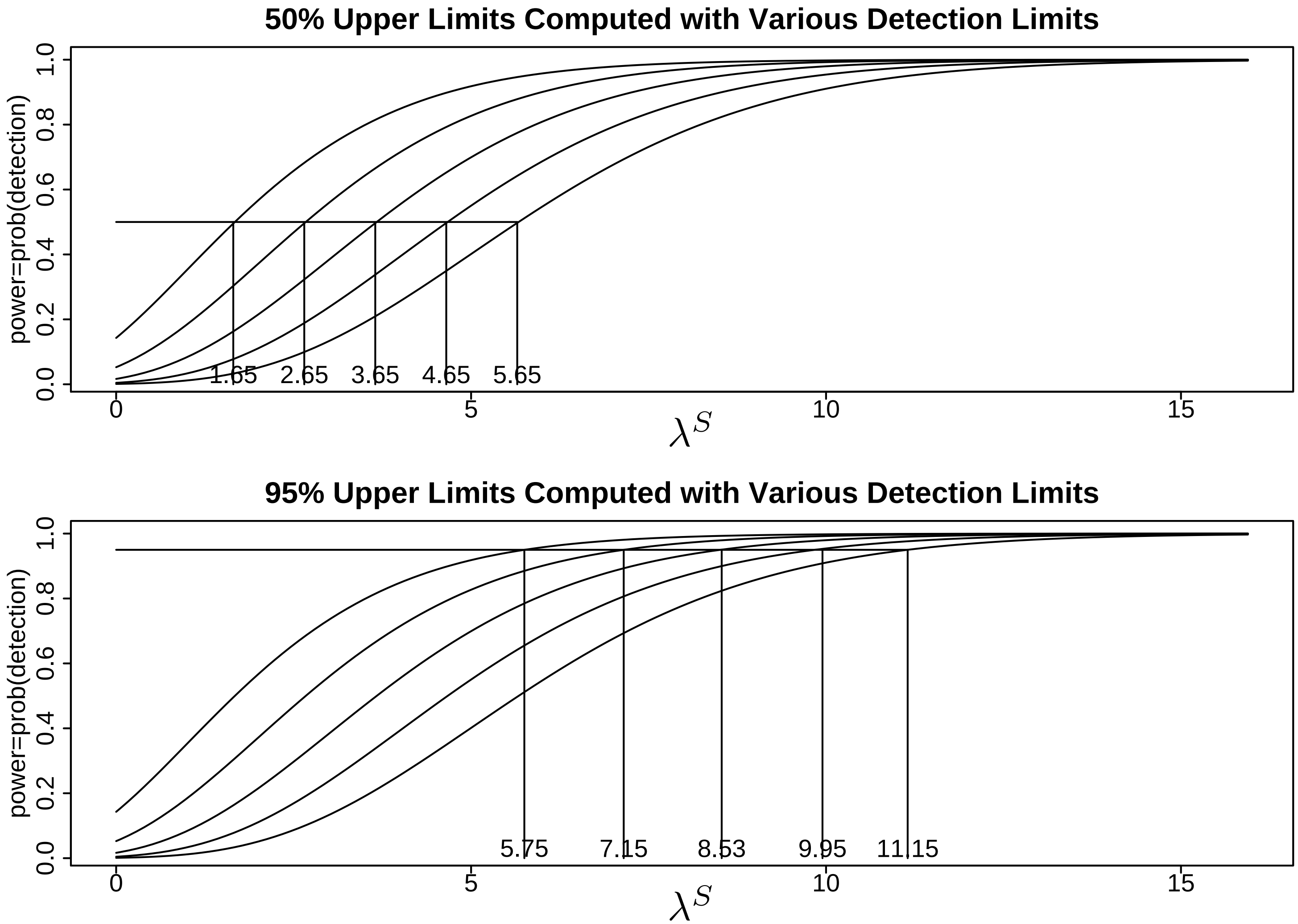
The 95% Upper Limit computed with a 95% Detection Limit.

The Effect of Choice of Detection Limit on Upper Limits





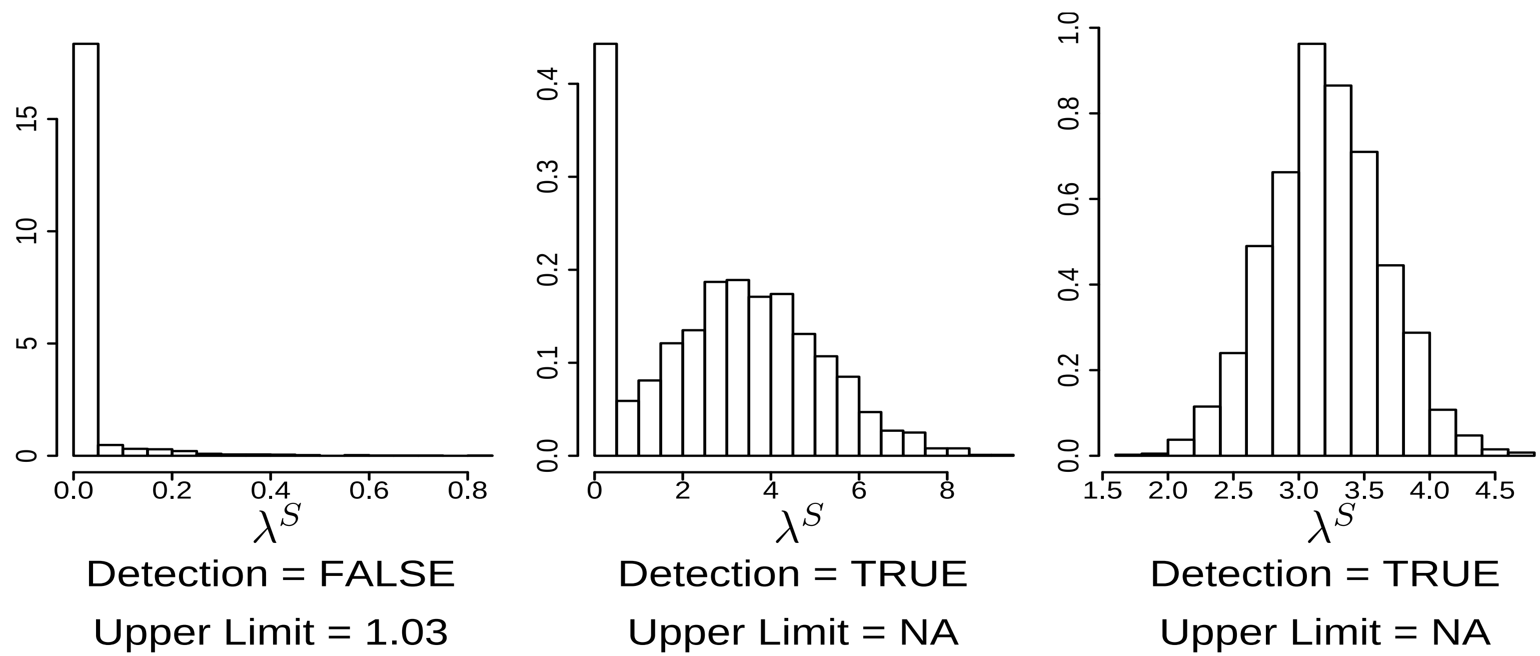
The Effect of Upper Limit Probability on Upper Limits



Interval Estimates

For detected or non-detected sources,
we may wish to quantify uncertainty for λ_i^S.

Three typical posterior distributions, computed using an MCMC sampler:



Confidence Intervals Do Not Adequately Represent All Distributions.

Nuisance Parameters

Typical Strategy:

- Average (or perhaps optimize) over nuisance parameters.
- Check frequency properties via simulation studies.
- If necessary, adjust procedures and priors to improve frequency properties.

Sometimes (partially) ancillary statistics are available:

$$
\begin{aligned}
T = Y^S + Y^B &\sim \text{Poisson}\left(\lambda^S + 2\lambda^B\right) = \text{Poisson}\left((\rho+1)\lambda^B\right) \\
Y^S \mid T &\sim \text{Binomial}\left(T, \frac{\lambda^S + \lambda^B}{\lambda^S + 2\lambda^B}\right) = \text{Binomial}\left(T, \frac{\rho}{1+\rho}\right)
\end{aligned}
$$

With $\rho = (\lambda^S + \lambda^B)/\lambda^B$ and $\kappa^S = \kappa^B = 1$.

We use the Binomial Dist'n to compute Detection Limits and Upper Limits.

1. Both procedures appear to detect *exactly* the same sources.
2. The Binomial Procedure produces more conservative Upper Limits (for ρ).
   - If the Detection Limit is T, the Upper Limit for ρ is $+\infty$.

Concluding Recommendations

1. *Plot the (profile) likelihood or (marginal) posterior distribution!*
2. Don't expect too much of confidence intervals.
3. Remember frequency confidence intervals & p-values use *pre-data* probability.
   - Rejecting may not mean the alternative is more likely than the null.
4. Remember that subjective model choices can be just as influential as a subjective prior distribution.
5. Be willing to rethink subjective model choices based on prior information.
6. Be willing to use your data to access the likely values of nuisance parameters. Use methods that take advantage of this information.