Confidence Intervals and Upper Bounds for Small Signals in the Presence of Noise.

Dr. Wolfgang Rolke

Workshop on Confidence Limits, 27-28

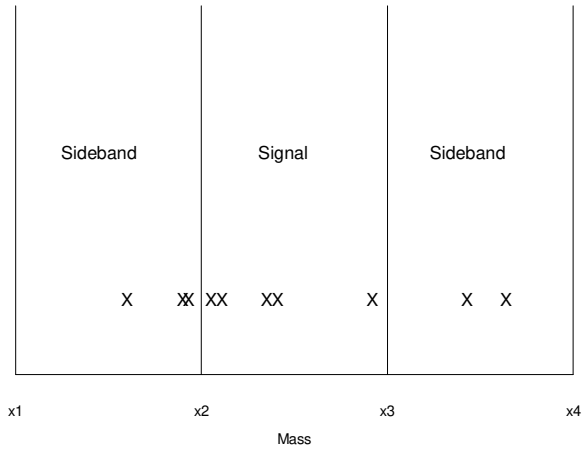
March, 2000

This work was supported by the
DOE-EPSCoR program

Table of Contents:

- Statement of Problem
- Some general ideas about the construction of confidence intervals
- A confidence region in two dimensions
- A confidence interval for the signal rate
- Performance of the new method
- Extensions

1 Statement of Problem



Events is Signal and Background Range

$$\tau = \frac{(x2 - x1) + (x4 - x3)}{x3 - x2}$$

$$X = 5, \qquad Y = 5$$

$$X \sim Pois(\mu + b), \qquad Y \sim Pois(\tau b)$$

2 Constructing Confidence Intervals

Popular method:

1. Find Hypothesis test

2. Invert the test

Hypothesis Test: Make assumption about parameter, see whether data agrees with it.

Say $H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$

Decision is based on the acceptance region:

$$A(\mu_0) = \{\mathbf{x} : \text{Test accepts null hypothesis}\}$$

Turn this around to find confidence region:

$$C(\mathbf{x}) = \{\mu_0 : \mathbf{x} \in A(\mu_0)\}$$

3 A confidence region in two dimensions

We will find a hypothesis test for $H_0 : \mu = \mu_0, b = b_0$

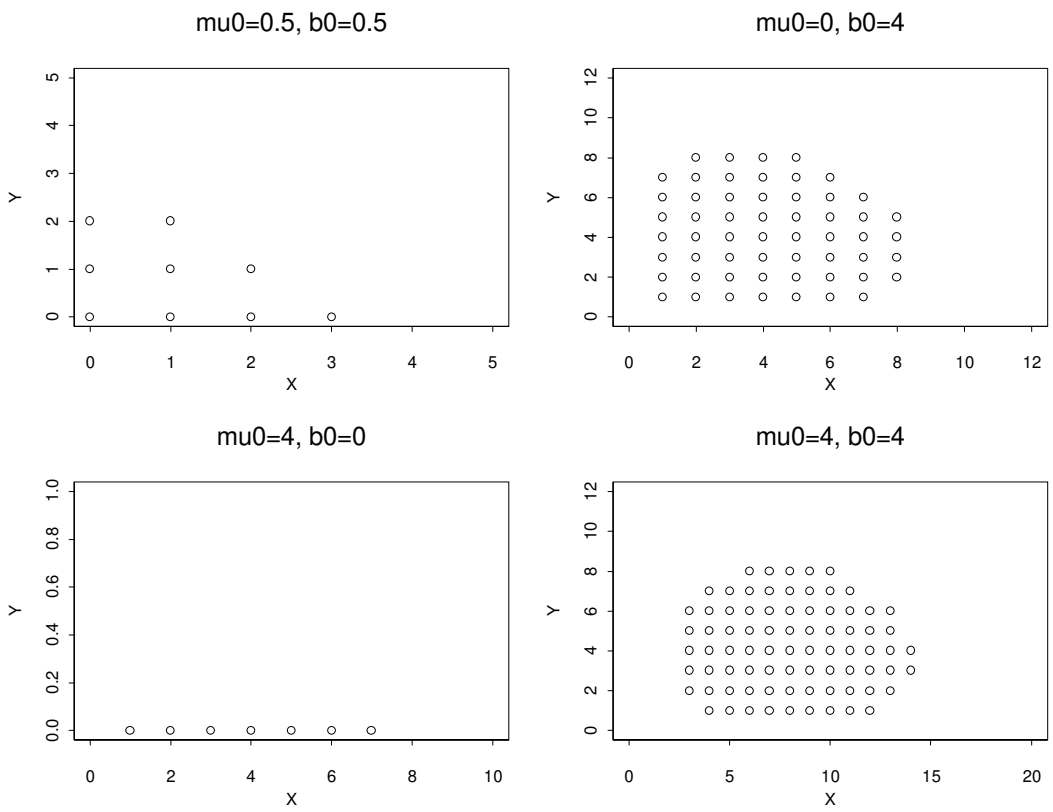
1. List (almost) all possible observations (X, Y) assuming H_0 is true

2. Find their probabilities by

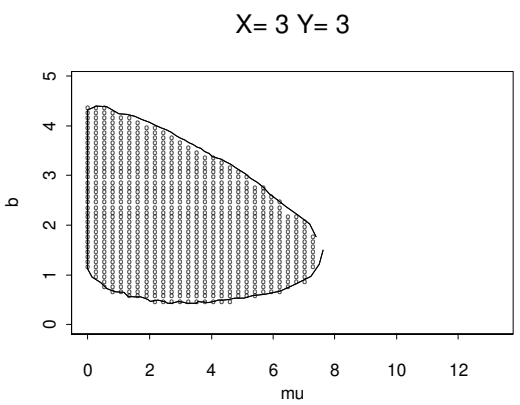
$$p(X, Y; \mu_0, b_0) = \frac{(\mu_0 + b_0)^X}{X!} e^{-(\mu_0 + b_0)} \cdot \frac{(\tau b_0)^Y}{Y!} e^{-\tau b_0}$$

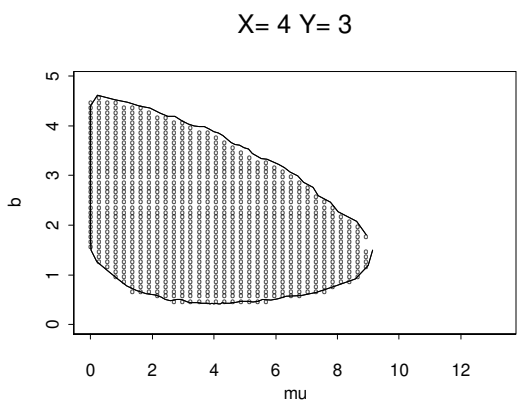
3. Sort all observations from the most likely to the most unlikely

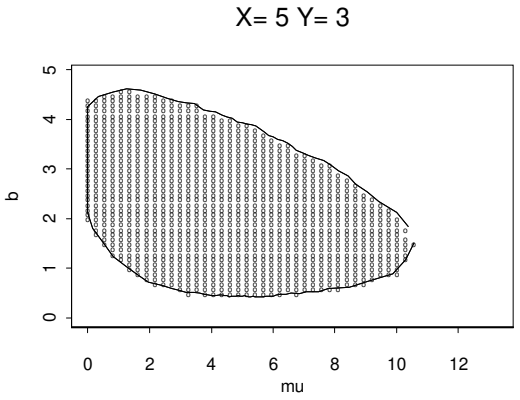
4. For a test with significance level a find the partial sums from the largest to the k^{th} observation until you reach $1 - \alpha$



Acceptance Region for Hypothesis tests,
tau=1,alpha=0.9







tau= 2

alpha= 0.9

Confidence Regions for mu and b

4 A confidence interval for the signal rate

Again, start with a hypothesis test:$H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$

Test Statistic: Likelihood Ratio test:

$$\Lambda(\mu_0; \mathbf{x}) = \frac{\max\{l(\mu_0, b) : b \geq 0\}}{\max\{l(\mu, b) : \mu \geq 0, b \geq 0\}}$$

$$= \frac{\text{Best explanation for data if } H_0 \text{ is true}}{\text{Best explanation for data, period}}$$

Here $l(\mu, b) = p(X, Y; \mu, b)$ is the likelihood function of μ and b.

So we need to find the maximum likelihood estimator of the background rate b for a fixed signal rate μ_0.

$$\frac{\partial}{\partial b} \log p(X, Y; \mu_0, b)$$

$$= \frac{\partial}{\partial b}\{X \cdot \log(\mu_0 + b) - (\mu_0 + b) +$$

$$Y \cdot \log(\tau b) - \tau b + \log(X! \cdot Y!)\}$$

$$= \frac{X}{\mu_0 + b} - 1 + \frac{Y}{b} - \tau = 0$$

$$\widehat{b} = \frac{\left\{X + Y - (1+\tau)\mu_0 + \sqrt{(X + Y - (1+\tau)\mu_0)^2 + 4(1+\tau)Y\mu_0}\right\}}{/\,(2(1+\tau))}$$

$l(\mu, \widehat{b})$ is called the profile likelihood of μ

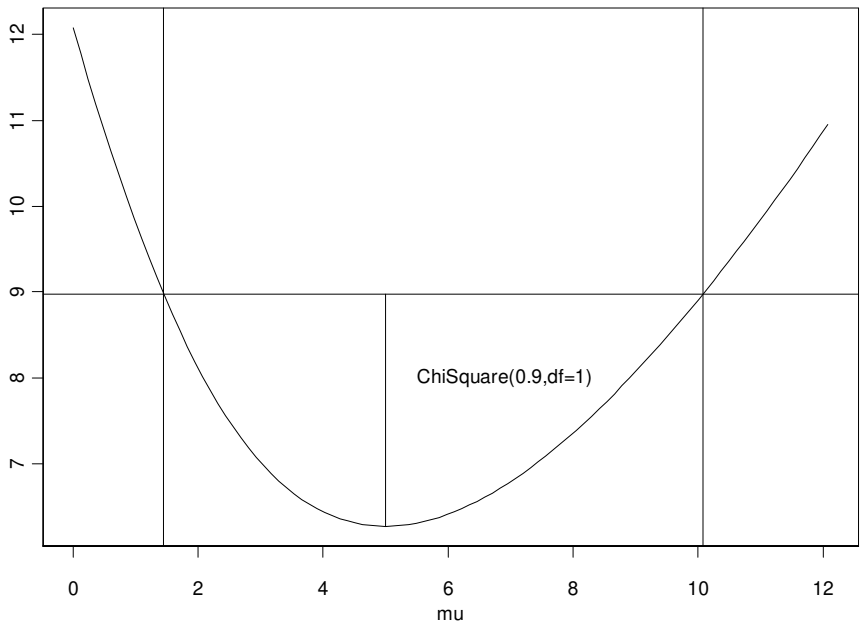
How can we find confidence limits from this?

BIG THEOREM: For a big enough sample,

$$-2 \log \Lambda(\mu_0; \mathbf{x}) \sim \chi^2$$

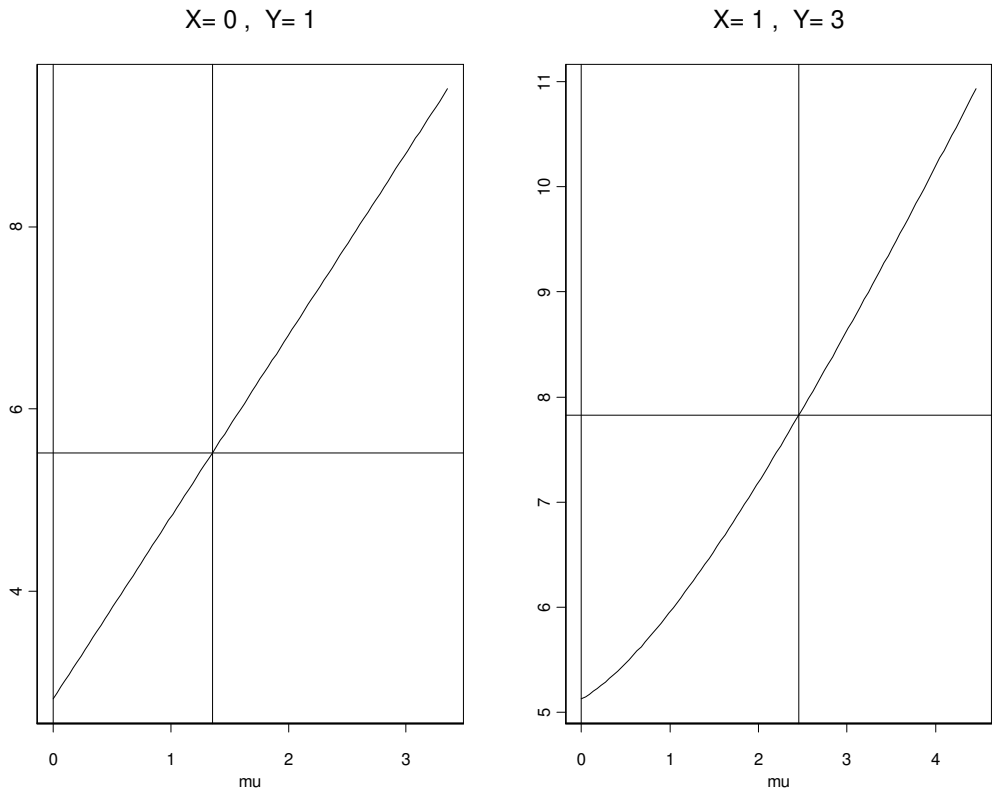
Not new: MINUIT and MINOS

But: we have a sample of size 1!



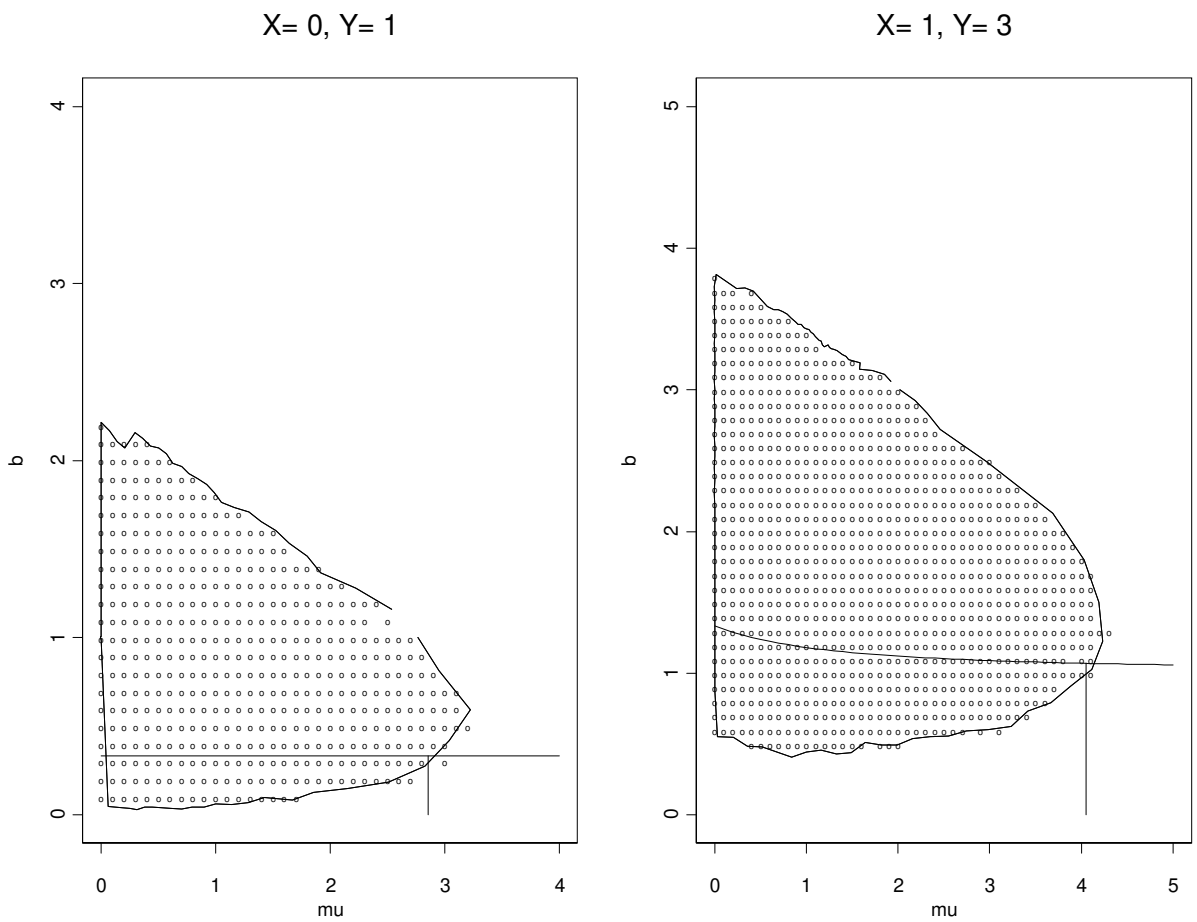
Profile Likelihood function with two-sided limits

Some Problems:



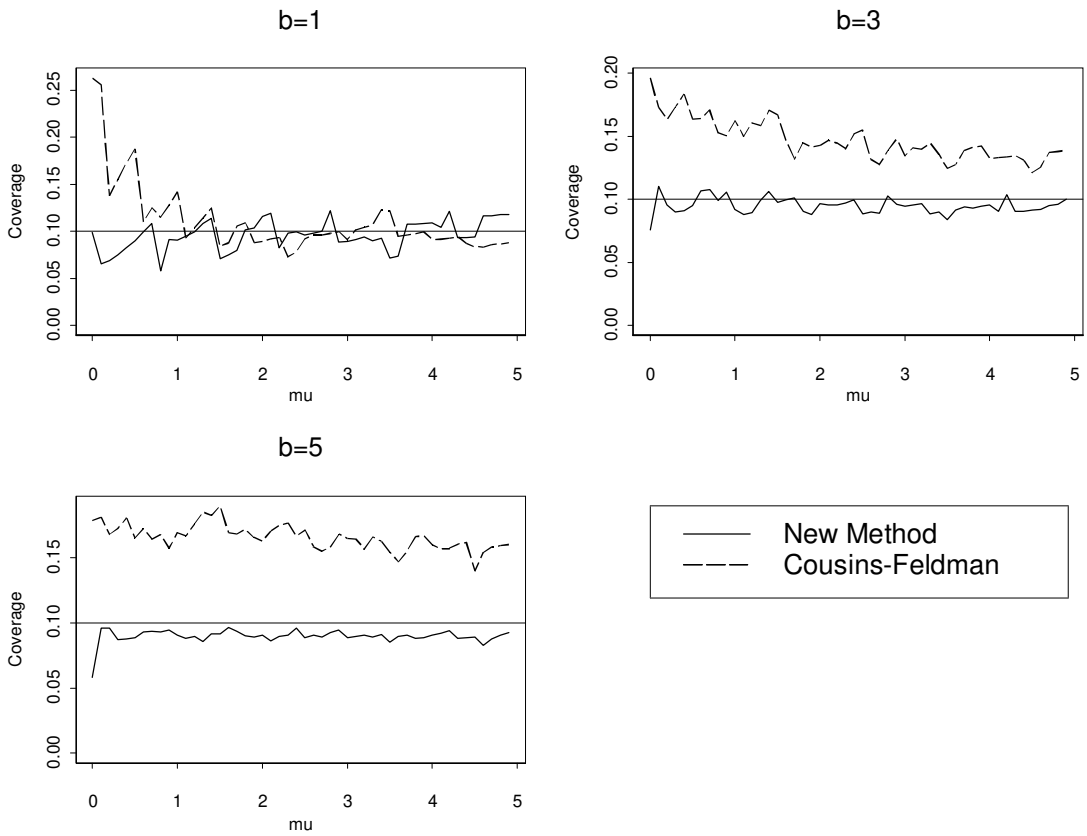
Profile Likelihood function for some extreme cases

Solution: Use a variation of first method for these cases:



Confidence Region with profile likelihood curve and upper bound.

5 Performance of the new Method



True coverage rates for our method and for Cousins-Feldman. Background rates are b=1,3 and 5, signal rates range from 0 to 5. Here we used $\tau = 1$ and we found 90% confidence intervals

6 Extensions of Method

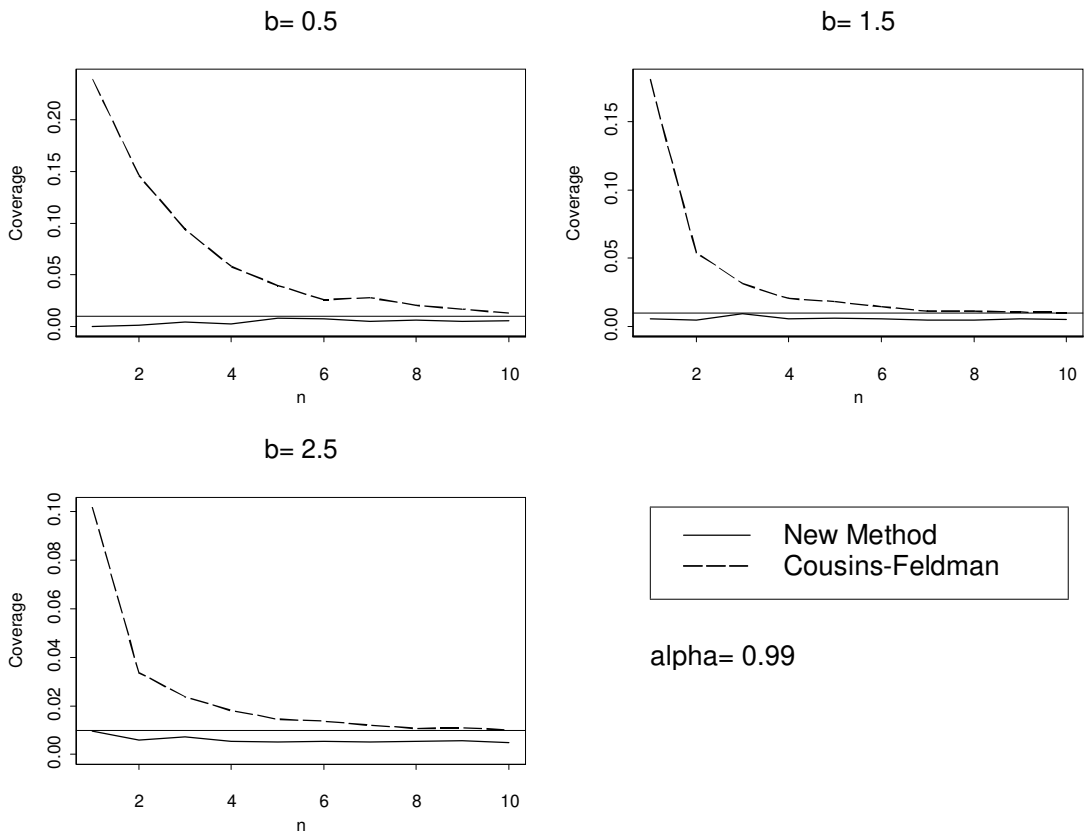
6.1 Estimating backgroud from Monte Carlo

New model:

$$X \sim Pois(\mu + b), Y \sim Pois(nb)$$

where Y is the sum of the events observed from the Monte Carlo and n is the number of times the Monte Carlo was run.

Same as before, just replace τ by n



Estimation of background by Monte Carlo, for mu=0. n is the number of times the Monte Carlo was run.

6.2 Include a second background source.

For example: mis id of pion for muon

New model:

$$X \sim Pois(\mu + b + \eta), Y \sim Pois(\tau b)$$
$$Z \sim Pois(n\eta)$$

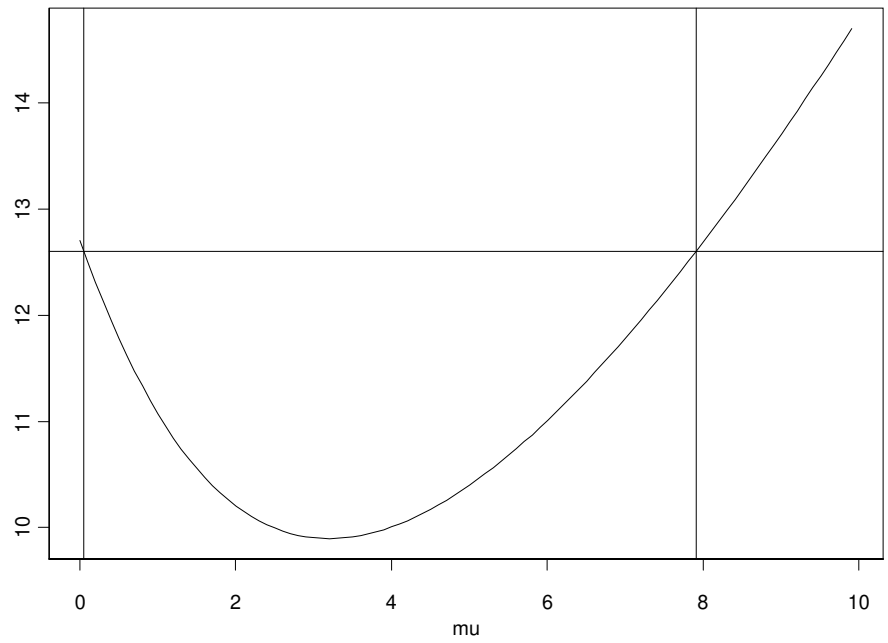
where Z is the sum of the events observed from the Monte Carlo and n is the number of times the Monte Carlo was run.

$$p(x, y; \mu, b, \eta) =$$
$$\frac{(\mu+b+\eta)^x}{x!}e^{-(\mu+b+\eta)}\frac{(\tau b)^y}{y!}e^{-\tau b}\frac{(n\eta)^z}{z!}e^{-n\eta}$$

Profile Likelihood:

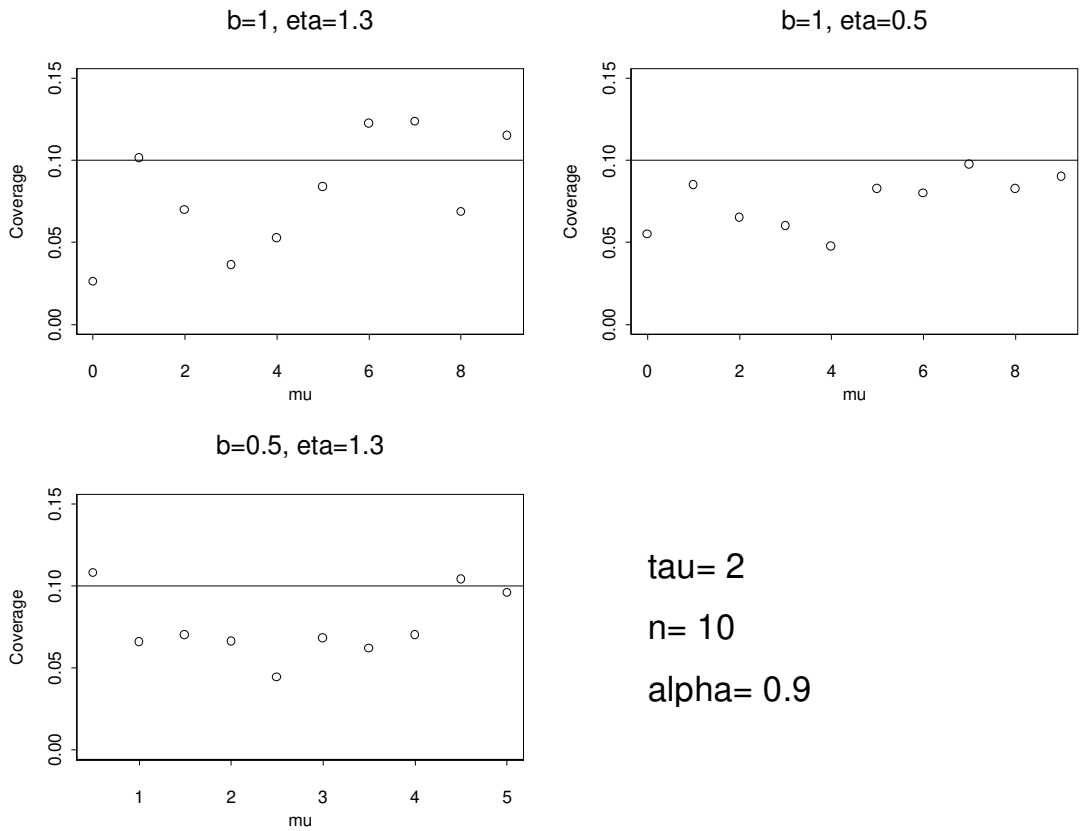
$$\frac{\partial \log p}{\partial b} = \frac{X}{\mu + b + \eta} - 1 + \frac{Y}{b} - \tau = 0$$

$$\frac{\partial \log p}{\partial \eta} = \frac{X}{\mu + b + \eta} - 1 + \frac{Z}{\eta} - n = 0$$



Profile Likelihood function for X=5, Y=1, Z=13, tau=2, n=10

Same problems as before with near-linear profile likelihoods if X is small compared to Y and Z.



Performance of method with two background sources. Coverage is estimated by Monte Carlo and has some uncertainty

7 Conclusions

- Method treats uncertainty in background as statistical error
- "Unified" treatment of upper limits and two-sided confidence interals
- (Almost) correct coverage
- Works just as well if background is estimated by Monte Carlo
- Extends easily to second background source