Wald's theorem and the "Asimov" data set

Eilam Gross & Ofer Vitells

ATLAS statistics forum , Dec. 2 2009

Outline

- We have previously "guessed" that the median significance of many toy MC experiments could be obtained simply by using the "Asimov" data set, i.e. the one data set in which all observed quantities are set equal to their expected values.
- This guess was supported by MC evidence but not proven.
- Here we show that due to a theorem by Wald [1943] , this relation can indeed be mathematically proven, in the same limiting approximation as that of Wilks theorem (the "χ^2 approximation").

Introduction

- The well known theorem by Wilks [1939] states that the profile likelihood ratio -2logλ distributes asymptotically as χ^2, *when the null hypothesis is true.*
- Wald' theorem [1943] generalizes this result to the *non-null hypothesis:* In that case the asymptotic distribution of -2logλ is a *non-central* χ^2*.* The approx. is essentially

$$-2\log\lambda(\theta) \simeq (\hat{\theta}-\theta)^T \mathbf{V}^{-1}(\hat{\theta}-\theta)$$

Where $\hat{\theta}$ are the MLE's , which are normally distributed with

covariance matrix $\mathbf{V}_{ij} = -E[\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}]$ (Fisher Information)

- The *non-centrality* parameter is

$$\Lambda = (\theta_0 - \theta)^T \mathbf{V}^{-1}(\theta_0 - \theta)$$

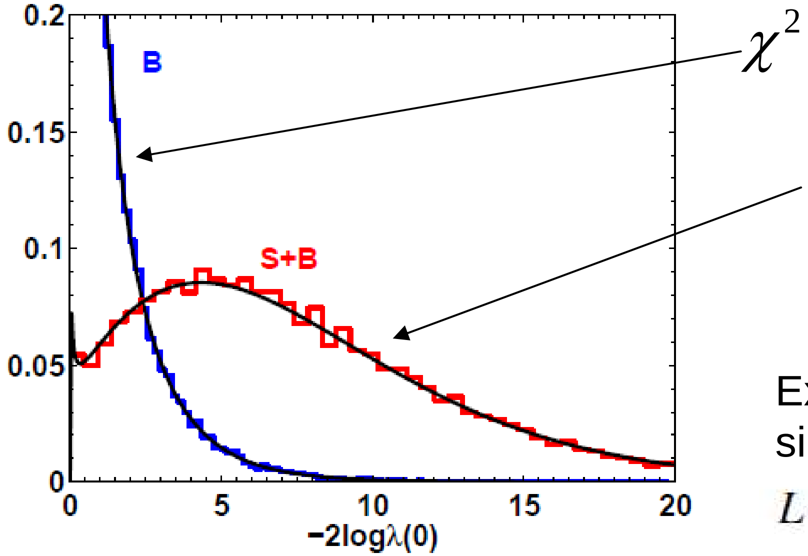
where θ_0 is the true parameter **(Kendall & Stuart, 6th edition, §22.7, p.246)**

Introduction

- For the "Asimov" data set, the MLE's assume their "true" values, $\hat{\theta}_{asimov} = \theta_0$ so we have in this limit

$$-2\log\lambda_{asimov} = \Lambda$$

- i.e., the "Asimov" data set produces the *non-centrality parameter* of the distribution of -2logλ, under the non-null hypothesis



Non-central χ^2

With $\Lambda = -2\log\lambda_{asimov}$

Example: counting experiment with a sideband measurement

$$L = Poiss(n|\mu s + b)Poiss(m|\tau b)$$

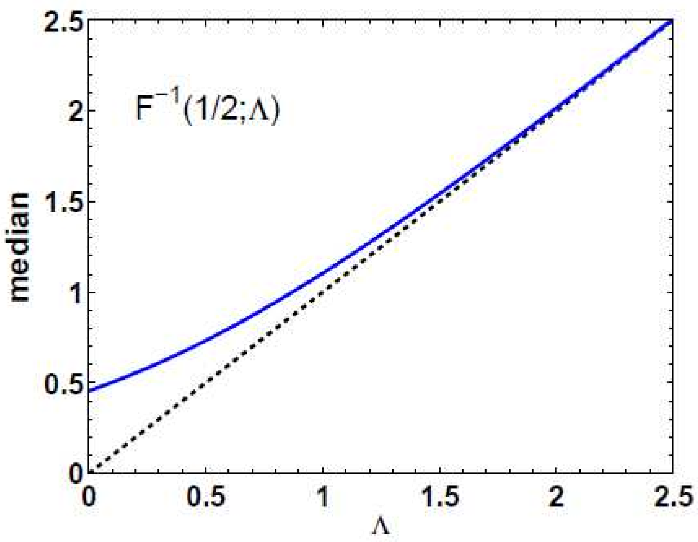
S=50 , b = 100 , τ = 0.5

Relation to median significance

- Since the non-null distribution is known, the median significance can be related the the Asimov value of -2logλ by

$$Z^2_{median} = -2\log\lambda_{median} = F^{-1}(1/2; -2\log\lambda_{asimov})$$

where $F(x;\Lambda)$ is the cumulative distribution function of a non-central χ^2 with non-centrality parameter Λ .



The median converges to the Asimov value for $\Lambda \gg 1$

We have not yet specified the null hypothesis, i.e. this relation holds for both discovery & exclusion

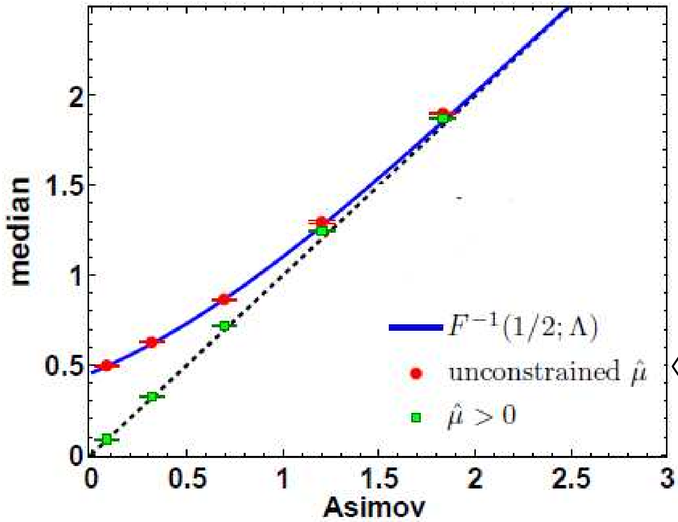
Physical constraints

- Usually we would not like to allow the signal strength μ to be negative, since this has no physical meaning
- For a single parameter of interest (μ) we have

$$-2\log\lambda(\mu) \simeq \frac{(\hat{\mu}-\mu)^2}{\sigma_\mu^2} \quad \text{with} \quad -2\log\lambda(0)_{asimov} \simeq \frac{1}{\sigma_\mu^2} = \Lambda$$

- $\hat{\mu}$ is normally distributed with variance σ_μ^2 and $\langle\hat{\mu}\rangle = 1$ (under the non-null hypothesis, for discovery)
- If we require $\hat{\mu} > 0$, then $P(\hat{\mu} > 1) = P(\hat{\mu}^2 > 1) = P(\hat{\mu}^2/\sigma_\mu^2 > 1/\sigma_\mu^2) = 1/2$
  ➔ the median is given exactly by the Asimov value

$$Z_{median} = \sqrt{-2\log\lambda_{asimov}}$$



⇐ Toy MC with different values of S

$$L = Poiss(n|\mu s + b)Poiss(m|\tau b)$$

Exclusion

- For exclusion, we test $\mu = 1$, and $\langle \hat{\mu} \rangle = 0$ under the non-null
- Similarly to the previous case, we need to require $\hat{\mu} < 1$ in order for the equality between the median and Asimov to hold :

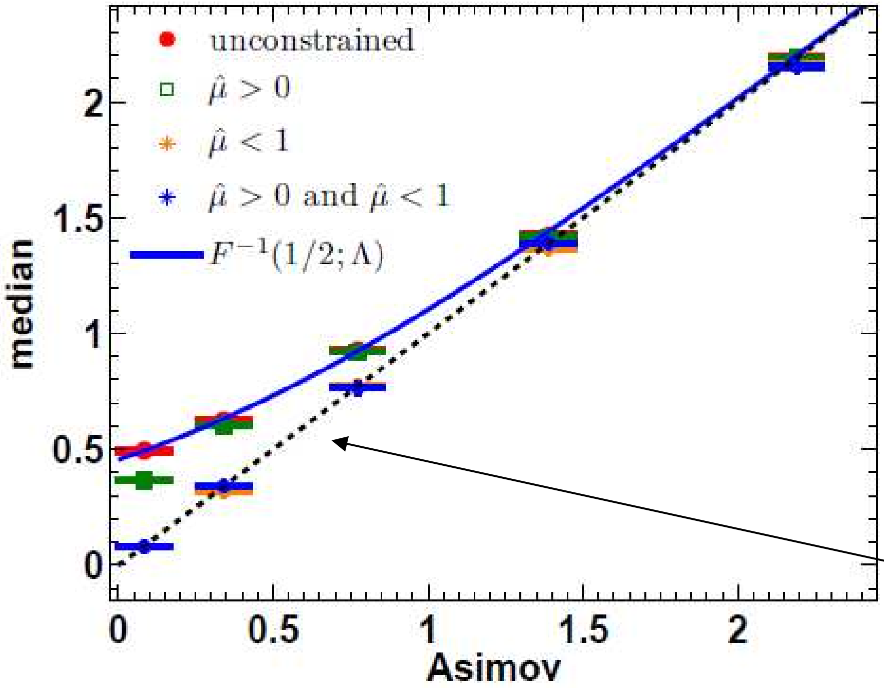
$$P(1-\hat{\mu} > 1) = P((1-\hat{\mu})^2 > 1) = P((1-\hat{\mu})^2 / \sigma_\mu^2 > 1/\sigma_\mu^2) = 1/2 \quad \text{if} \quad \hat{\mu} < 1$$

so $Z_{median} = \sqrt{-2\log\lambda_{asimov}}$ if $\hat{\mu} < 1$

- This also makes sense physically: we would not like to reject the null hypothesis (s+b) if the signal rate is higher then expected

- Notice that in this case negative values of $\hat{\mu}$ always result in -2logλ(1) being larger then the median, since $(1-\hat{\mu})^2 > 1$ if $\hat{\mu} < 0$. Therefore, the median will not change whether we require $\hat{\mu}$ to be positive or not.

Exclusion



Toy MC with different values of S

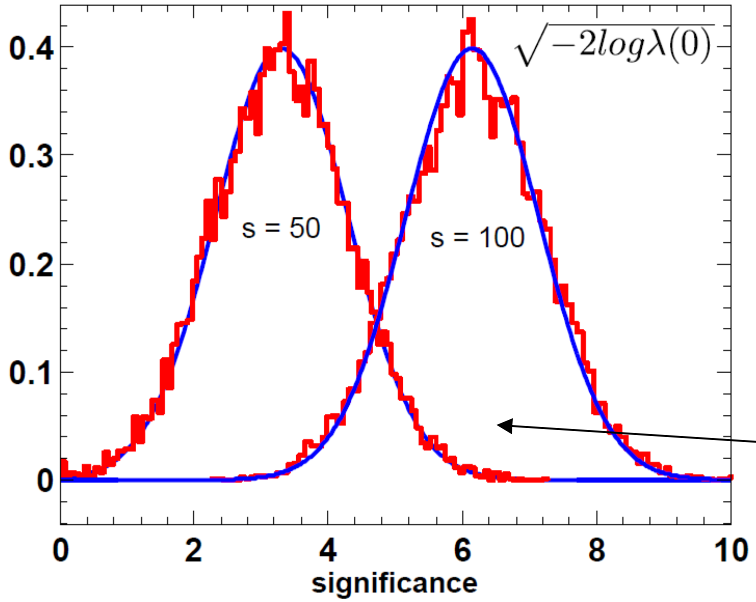
$$L = Poiss(n|\mu s + b)Poiss(m|\tau b)$$

When $\hat{\mu} < 1$,The Asimov and median values agree , regardless of the additional requirement $\hat{\mu} > 0$

Significance error bands

$$-2\log\lambda(\mu) \simeq \frac{(\hat{\mu}-\mu)^2}{\sigma_\mu^2}$$

- Notice that the significance (e.g. $\sqrt{-2\log\lambda(0)}$) is normally distributed with variance **1,**
  i.e. the "error bands" of the significance are always just ±1



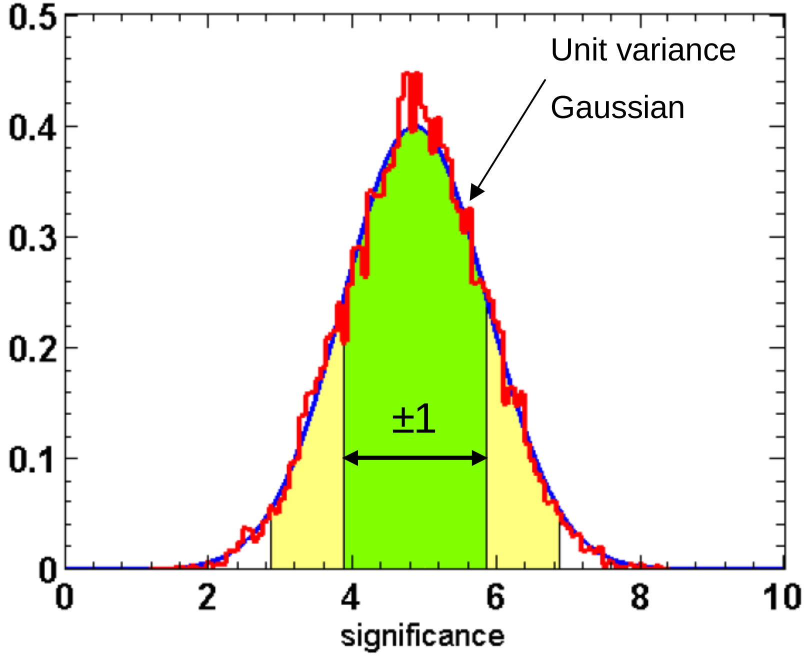
e.g. if the expected median significance is 4.8 , then the ±1σ band (68% CL interval) is 4.8±1

unit variance gaussians

Combination example

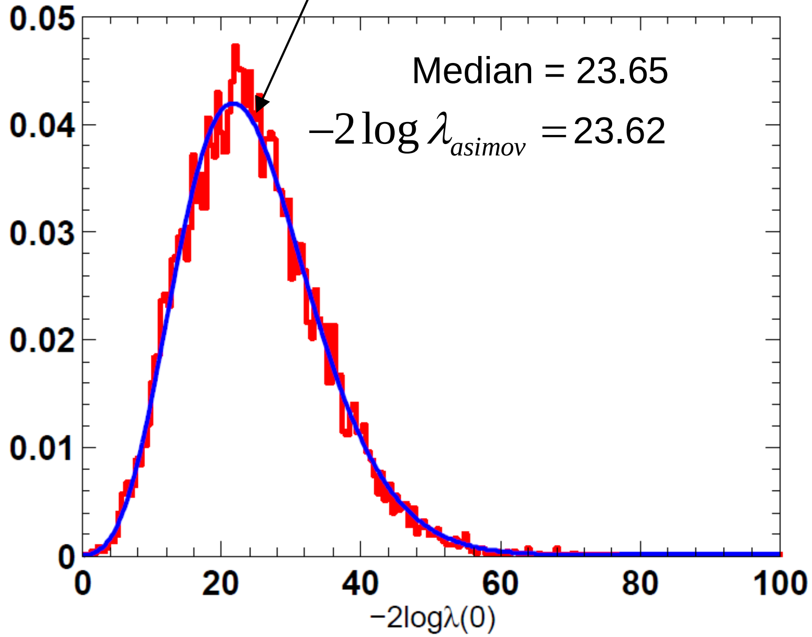
- We extend the example to a combination of 5 channels, each one is a counting experiment with a sideband measurement

$$L(\mu) = \prod_i Poiss(n_i \mid \mu s_i + b_i) Poiss(m_i \mid \tau_i b_i)$$



Non-central χ^2

With $\Lambda = -2\log\lambda_{asimov}$



In this example:
s_i = {20,10, 5, 40, 25}
b_i = {10, 30, 10, 60, 40} ,
τ_i = {0.5, 1 , 2 , 0.7 , 1}

Conclusions

- Wald's theorem puts the Asimov "guess" on solid ground. we have shown that the Asimov data set can be used to obtain the median significance for both discovery & exclusion.

- Under "physical" constraints , $\hat{\mu} > 0$ for discovery and $\hat{\mu} < 1$ for exclusion, the Asimov value is equal to the median significance, as was previously conjectured.

- Wald's theorem gives the asymptotic distribution of the profile likelihood ratio under the non-null (alternative) hypothesis, from which one can get both the median significance and the associated error bands, without generating a single toy MC experiment.