A short discussion about statistical methods

Shi Xiaodong, Peng Haiping
State Key Laboratory of Particle Detection and Electronics
University of Science and Technology of China

2017.11.13

Outline

- Introduction
- Bayes Method
- Frequentist Method
- Wald approximation for profile likelihood ratio

Introduction

In particle physics experiments, results' statistical significance can be quantified by p-value or its equivalent Gaussian significance. When the significance is not strong, the upper limit is expected to describe the sensitivity.

There are two basic statistic method: Bayes and Frequentist.

Bayes Method

- Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on **prior** knowledge of conditions that might be related to the event. —Wikipedia

Bayes' law: $$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

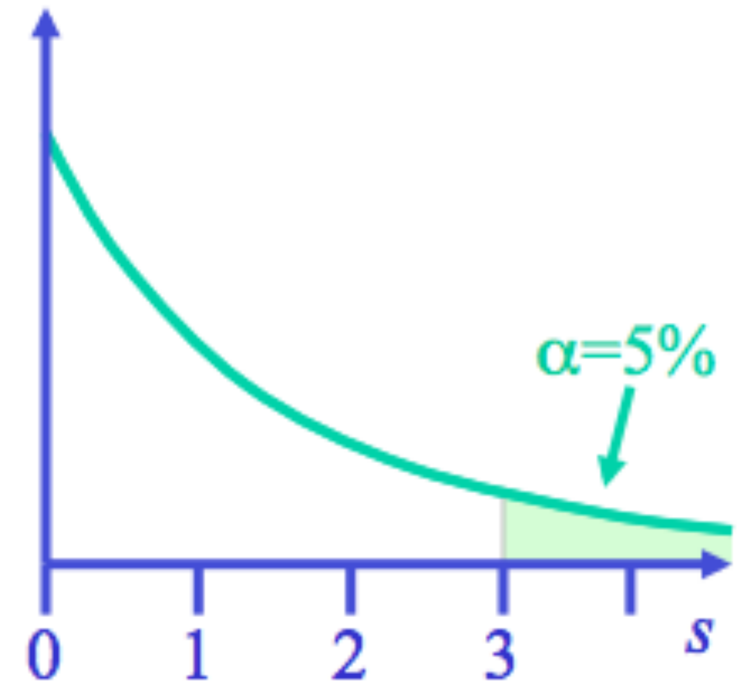
$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{p(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)\mathrm{d}\theta}$$

Bayes Method

- To get the upper limit. One needs to assume varying signal strengths, then get the likelihood values by fitting, then integrate the likelihood curve to make the formula equals to (1-alpha).

$$\int_0^{s^{\mathrm{up}}} P(s|n)\mathrm{d}s = \frac{\int_0^{s^{\mathrm{up}}} L(n;s)\pi(s)\mathrm{d}s}{\int_0^{\infty} L(n;s)\pi(s)\mathrm{d}s}$$

A uniform prior, $\pi(s) = 1$



Likelihood curve

Frequentist Method

- Frequentist inference is a type of statistical inference that draws conclusions from sample data by emphasizing the **frequency** or proportion of the data. An alternative name is frequentist statistics. —Wikipedia

$$P(A) = Limit_{N \to \infty} \frac{N_A}{N}$$

Do experiment with infinite times.

Frequentist Method

Hypothesis Testing (Frequentist Technique)

Null hypothesis H_0 : hypothesis which you try to falsify / reject
(one can not verify / approve hypothesis)

Always just bkg.

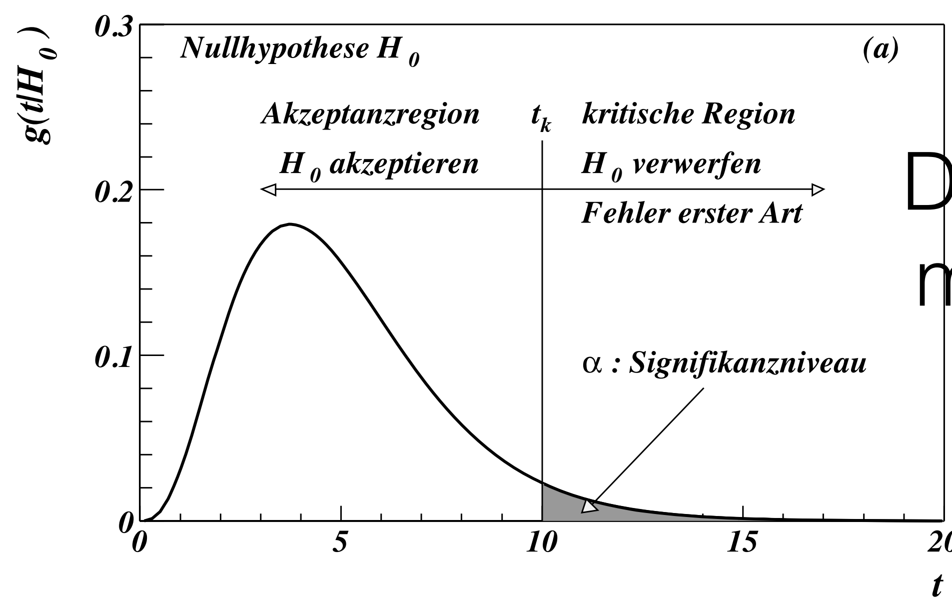
Test statistic t: any function of your data which is used to quantify (dis-)agreement with H_0

g(t|H0): probability density function PDF for test statistics under null hypothesis H_0

Critical region: range of test statistic for which H_0 is rejected

α: significance (level)
size of test
error of 1st kind.
probability to reject H_0,
if H_0 is true

$$\boxed{\alpha = \int_{t_k}^{\infty} g(t|H_0)dt.}$$



Do experiment many times to get $g(t|H_0)$

- *p*-value: probability to observe at least n_{obs} events if the null hypothesis H_0 ($s = 0$) is true

Frequentist Method

Hypothesis Testing

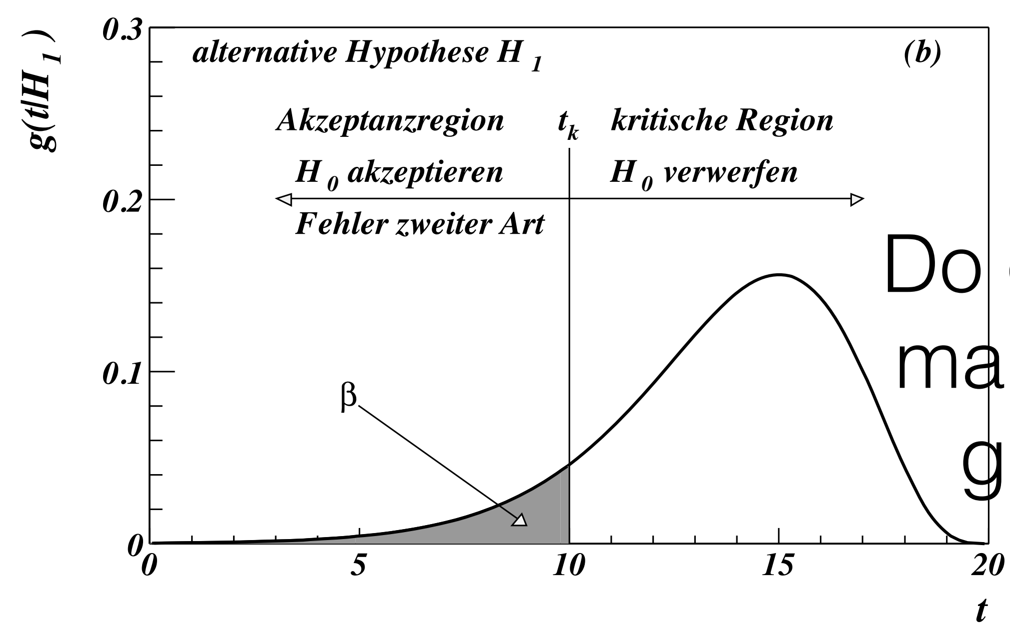
In principle: infinity many possibilities to choose critical region for given α
(especially for one sided tests you need an alternative hypothesis to decide what you call inconsistent with null hypothesis)

Include bkg & signal.

Alternative hypothesis H_1 : hypothesis which you would like to approve

$g(t|H_1)$: probability density function for test statistics under alternative hypothesis H_1

$$\boxed{\beta = \int_{-\infty}^{t_k} g(t|H_1)dt.}$$



Do experiment many times to get $g(t|H_0)$

β: error of 2nd kind
$M=1-\beta$: power

β prob. to reject H_1, if H_1 is true
$1-\beta$ prob to "accept" H_1, if H_1 is true

Frequentist Method

How to do experiment many times?

One way is:
Sample the number observed with Poisson function many times. For one time, get one t value. After sample enough times(do experiment), one can get $g(t|H0)$. Then one can get p-value.
If the result is a distribution, split into several bins then sample in every bin.

In order to get upper limit, one can get a p_μ-value for each H_μ(μ is POI, parameter of interest).
Then 90% CLupper limit on μ is highest value for p_μ-value is not less than 0.1.

Frequentist Method

One technology problem:

In order to get p.d.f. of test statistic(t), one need to sample for many times, calculate t for many times. To get the upper limit, one need much more cpu time.

One solution: use asymptotic formulae for special t.

1007.1727

Wald approximation for profile likelihood ratio

Base on the profile likelihood ratio:

maximizes L for
Specified μ

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

maximize L

the parameter μ determines the strength of the signal process

use $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_{\text{tot}})$ to denote all of the nuisance parameters.

We can use $t_\mu = -2 \ln \lambda(\mu)$ as test statistic.

Wald approximation for profile likelihood ratio

Use approximation due to Wald (1943)

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$

$$\hat{\mu} \sim \text{Gaussian}(\mu', \sigma)$$

sample size

μ' H_0

If we can neglect the $O(1/\sqrt{N})$ term, $-2\ln\lambda(\mu)$ follows a noncentral chi-square distribution for one degree of freedom with noncentrality parameter

$$\Lambda = \frac{(\mu-\mu')^2}{\sigma^2}$$

$$f(t_\mu;\Lambda) = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\sqrt{\Lambda}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\sqrt{\Lambda}\right)^2\right)\right]$$

Wald approximation for profile likelihood ratio

The Asimov data set

To estimate sigma, consider special data set where all statistical fluctuations suppressed and parameters are replaced by their expectation values. $\rightarrow$ $\hat{\mu} = \mu'$ $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$

$$\lambda_{\mathrm{A}}(\mu) = \frac{L_{\mathrm{A}}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_{\mathrm{A}}(\hat{\mu}, \hat{\boldsymbol{\theta}})} = \frac{L_{\mathrm{A}}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_{A}(\mu', \boldsymbol{\theta})}$$

$$-2\ln\lambda_{\mathrm{A}}(\mu) = \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda$$

Asimov value of $-2\ln\lambda(\mu)$ gives non-centrality param. Λ, or equivalently, σ.

Then by using Asimov set, one can get p.d.f. of t, then the calculation is much easier.

Wald approximation for profile likelihood ratio

About error band

$$\hat{\mu} \sim \text{Gaussian}(\mu', \sigma)$$

It is convenient to calculate error bands for the median significance corresponding to the $\pm N\sigma$ variation of $\hat{\mu}$. As $\hat{\mu}$ is Gaussian distributed, these error bands on the significance are simply the quantiles that map onto the variation of $\hat{\mu}$ of $\pm N\sigma$ about μ'.

Wald approximation for profile likelihood ratio

About the package used in one group of Atlas.

Input:

- data distribution; bkg distribution; signal distribution.(all in histogram form)
- Systematic error.
- Control region if need.

Could combine different channels or even different experiments(just no correlation case).
The Likelihood is based on poisson function.

$$\mathcal{L}(\mu,\vec{\theta}) = \left\{ \prod_{k=e\mu,\mu e} \prod_{j=0}^{N_{\text{category}}} \prod_{i=1}^{N_{\text{bins}}} P(N_{ijk}|\mu s_{ijk} + \sum_{m}^{N_{\text{bkg}}} b_{ijkm}) \right\} \times \prod_{i=1}^{N_{\theta}} N(\tilde{\theta}|\theta)$$

Wald approximation for profile likelihood ratio

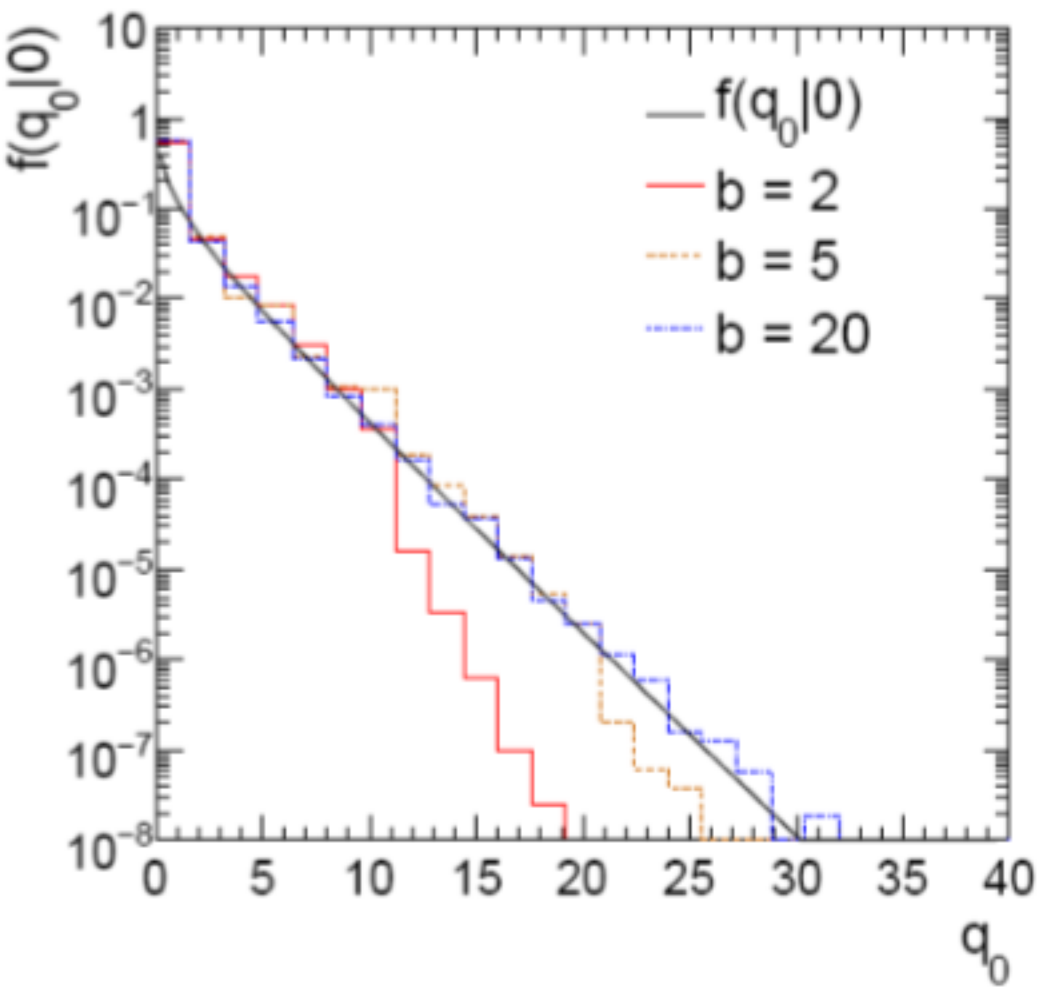
Monte Carlo test of asymptotic formula

$$n \sim \text{Poisson}(\mu s + b)$$

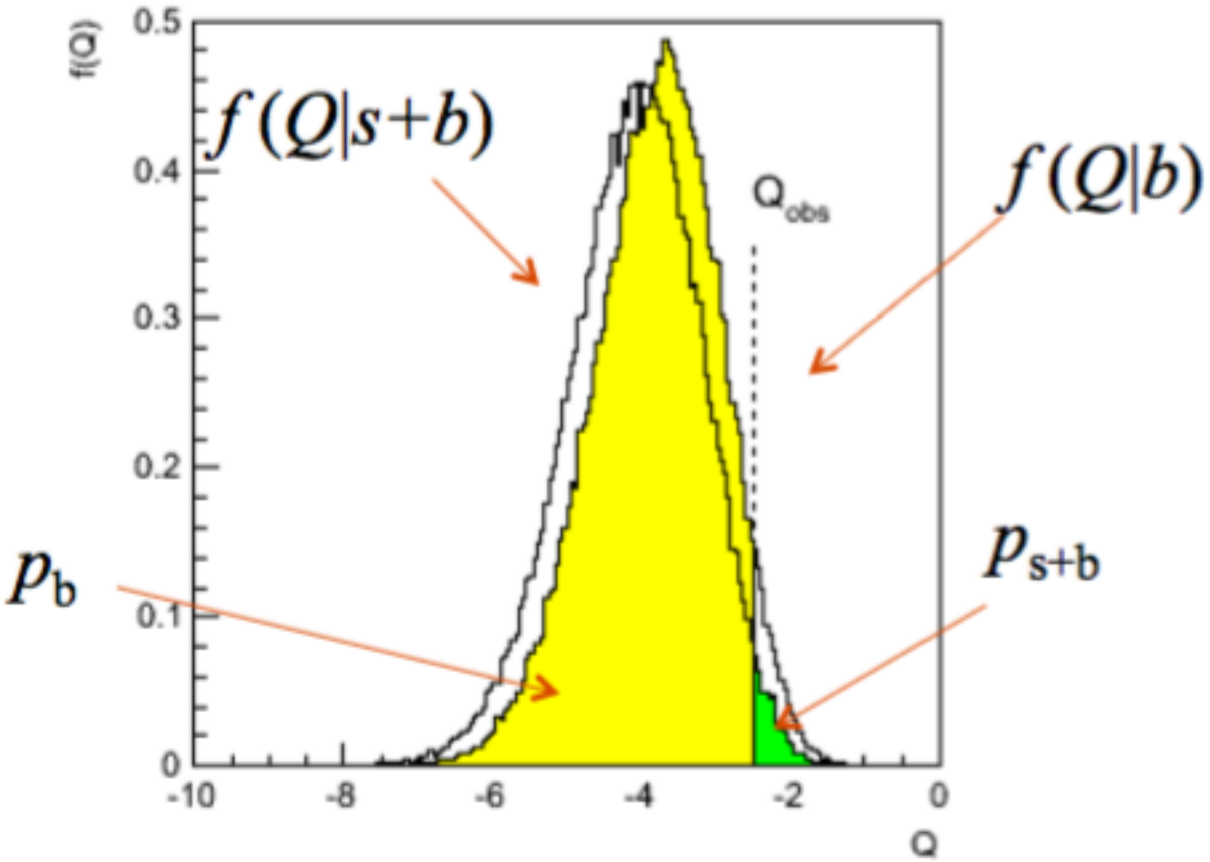
$$m \sim \text{Poisson}(\tau b)$$

Here take $\tau = 1$.

Asymptotic formula is good approximation to 5σ level ($q_0 = 25$) already for $b \sim 20$.



One more issue



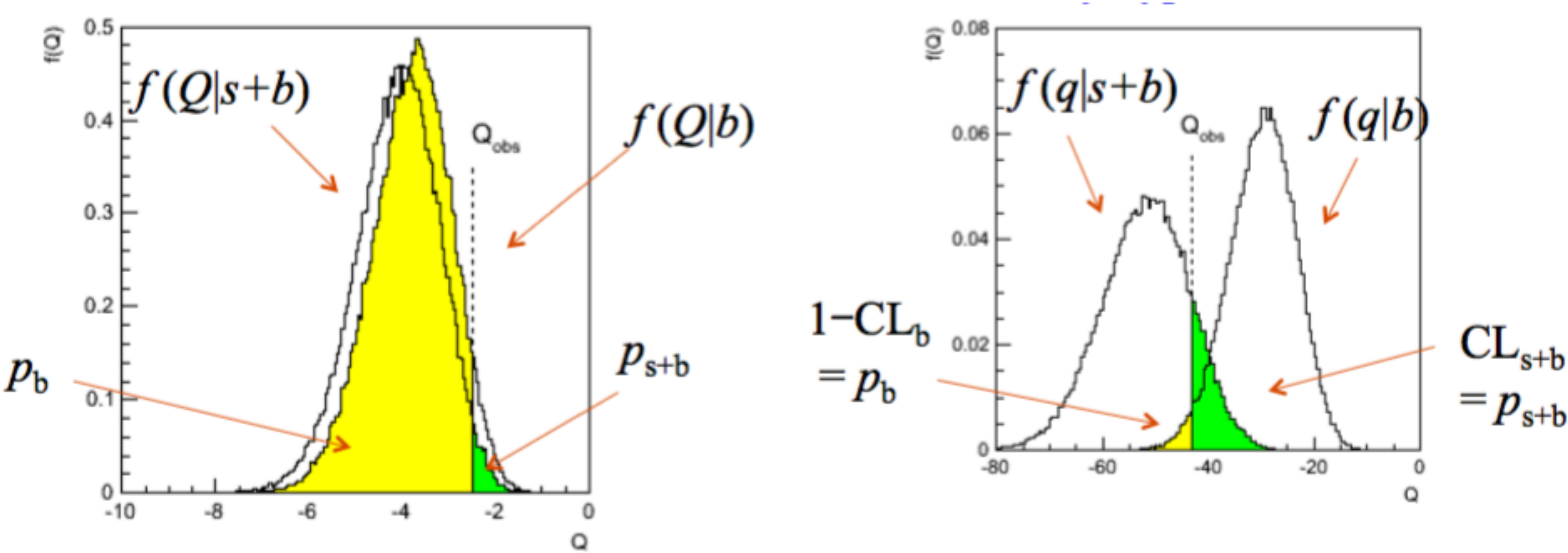
Sometimes these two distributions are close to each other.

Even if p_{s+b} is little, we should not regard a model as excluded.

solution

In the CLs method the p-value is reduced according to the recipe

$$p_\mu \to \frac{p_\mu}{1 - p_b}$$



Summary

Brief discussion about Bayes and Frequentist method on upper limit.

Introduction about Wald approximation for Frequentist method.

Debate your preferred statistical technique in a statistics forum, not a physics result publication!

We should debate and choose statistic method. Comparing their results may be meaningless.

Back-up