

Statistics for HEP

Lecture 3: More on Discovery and Limits

<http://indico.cern.ch/conferenceDisplay.py?confId=173728>



Academic Training Lectures

CERN, 2–5 April, 2012



Glen Cowan

Physics Department

Royal Holloway, University of London

g.cowan@rhul.ac.uk

www.pp.rhul.ac.uk/~cowan

Outline

Lecture 1: Introduction and basic formalism

Probability, statistical tests, parameter estimation.

Lecture 2: Discovery and Limits

Quantifying discovery significance and sensitivity

Frequentist and Bayesian intervals/limits

→ **Lecture 3: More on discovery and limits**

Bayesian limits

The Look-Elsewhere Effect

Dealing with nuisance parameters

Expected discovery significance

Lecture 4: Unfolding (deconvolution)

Correcting distributions for effects of smearing

The Bayesian approach to limits

In Bayesian statistics need to start with ‘**prior pdf**’ $\pi(\theta)$, this reflects degree of belief about θ before doing the experiment.

Bayes’ theorem tells how our beliefs should be updated in light of the data x :

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int L(x|\theta')\pi(\theta') d\theta'} \propto L(x|\theta)\pi(\theta)$$

Integrate **posterior pdf** $p(\theta | x)$ to give interval with any desired probability content.

For e.g. $n \sim \text{Poisson}(s+b)$, 95% CL upper limit on s from

$$0.95 = \int_{-\infty}^{s_{\text{sup}}} p(s|n) ds$$

Bayesian prior for Poisson parameter

Include knowledge that $s \geq 0$ by setting prior $\pi(s) = 0$ for $s < 0$.

Could try to reflect ‘prior ignorance’ with e.g.

$$\pi(s) = \begin{cases} 1 & s \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Not normalized but this is OK as long as $L(s)$ dies off for large s .

Not invariant under change of parameter — if we had used instead a flat prior for, say, the mass of the Higgs boson, this would imply a non-flat prior for the expected number of Higgs events.

Doesn’t really reflect a reasonable degree of belief, but often used as a point of reference;

or viewed as a recipe for producing an interval whose frequentist properties can be studied (coverage will depend on true s).

Bayesian interval with flat prior for s

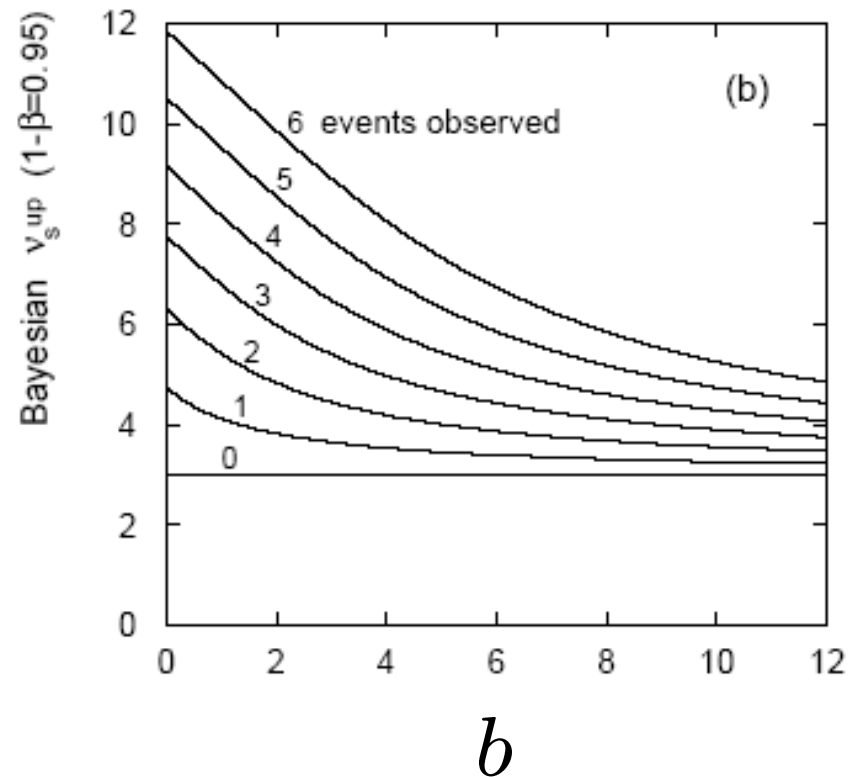
Solve numerically to find limit s_{up} .

For special case $b = 0$, Bayesian upper limit with flat prior numerically same as one-sided frequentist case ('coincidence').

Otherwise Bayesian limit is everywhere greater than the one-sided frequentist limit, and here (Poisson problem) it coincides with the CLs limit.

Never goes negative.

Doesn't depend on b if $n = 0$.



Priors from formal rules

Because of difficulties in encoding a vague degree of belief in a prior, one often attempts to derive the prior from formal rules, e.g., to satisfy certain invariance principles or to provide maximum information gain for a certain set of measurements.

Often called “objective priors”

Form basis of Objective Bayesian Statistics

The priors do not reflect a degree of belief (but might represent possible extreme cases).

In Objective Bayesian analysis, can use the intervals in a frequentist way, i.e., regard Bayes’ theorem as a recipe to produce an interval with certain coverage properties.

Priors from formal rules (cont.)

For a review of priors obtained by formal rules see, e.g.,

Robert E. Kass and Larry Wasserman, *The Selection of Prior Distributions by Formal Rules*, J. Am. Stat. Assoc., Vol. 91, No. 435, pp. 1343-1370 (1996).

Formal priors have not been widely used in HEP, but there is recent interest in this direction, especially the reference priors of Bernardo and Berger; see e.g.

L. Demortier, S. Jain and H. Prosper, *Reference priors for high energy physics*, Phys. Rev. D 82 (2010) 034002, arXiv:1002.1111.

D. Casadei, *Reference analysis of the signal + background model in counting experiments*, JINST 7 (2012) 01012; arXiv:1108.4270.

Jeffreys' prior

According to *Jeffreys' rule*, take prior according to

$$\pi(\boldsymbol{\theta}) \propto \sqrt{\det(I(\boldsymbol{\theta}))}$$

where

$$I_{ij}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \ln L(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] = - \int \frac{\partial^2 \ln L(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} L(\mathbf{x}|\boldsymbol{\theta}) dx$$

is the Fisher information matrix.

One can show that this leads to inference that is invariant under a transformation of parameters.

For a Gaussian mean, the Jeffreys' prior is constant; for a Poisson mean μ it is proportional to $1/\sqrt{\mu}$.

Jeffreys' prior for Poisson mean

Suppose $n \sim \text{Poisson}(\mu)$. To find the Jeffreys' prior for μ ,

$$L(n|\mu) = \frac{\mu^n}{n!} e^{-\mu} \qquad \frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\mu^2}$$

$$I = -E \left[\frac{\partial^2 \ln L}{\partial \mu^2} \right] = \frac{E[n]}{\mu^2} = \frac{1}{\mu}$$

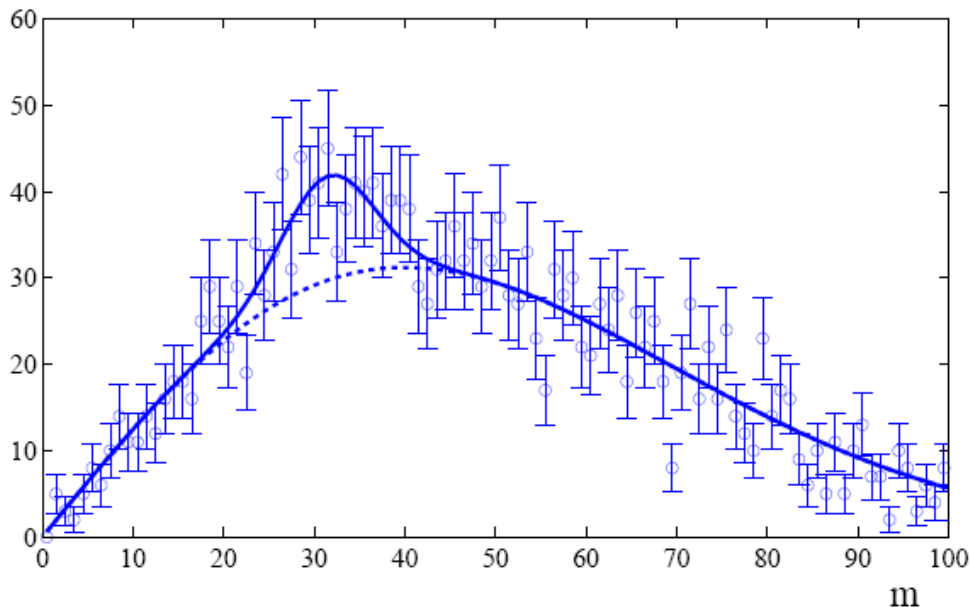
$$\pi(\mu) \propto \sqrt{I(\mu)} = \frac{1}{\sqrt{\mu}}$$

So e.g. for $\mu = s + b$, this means the prior $\pi(s) \sim 1/\sqrt{(s + b)}$, which depends on b . Note this is not designed as a degree of belief about s .

The Look-Elsewhere Effect

Suppose a model for a mass distribution allows for a peak at a mass m with amplitude μ .

The data show a bump at a mass m_0 .



How consistent is this with the no-bump ($\mu = 0$) hypothesis?

p -value for fixed mass

First, suppose the mass m_0 of the peak was specified a priori.

Test consistency of bump with the no-signal ($\mu = 0$) hypothesis with e.g. likelihood ratio

$$t_{\text{fix}} = -2 \ln \frac{L(0, m_0)}{L(\hat{\mu}, m_0)}$$

where “fix” indicates that the mass of the peak is fixed to m_0 .

The resulting p -value

$$p_{\text{fix}} = \int_{t_{\text{fix,obs}}}^{\infty} f(t_{\text{fix}}|0) dt_{\text{fix}}$$

gives the probability to find a value of t_{fix} at least as great as observed **at the specific mass m_0** .

p -value for floating mass

But suppose we did not know where in the distribution to expect a peak.

What we want is the probability to find a peak at least as significant as the one observed **anywhere** in the distribution.

Include the mass as an adjustable parameter in the fit, test significance of peak using

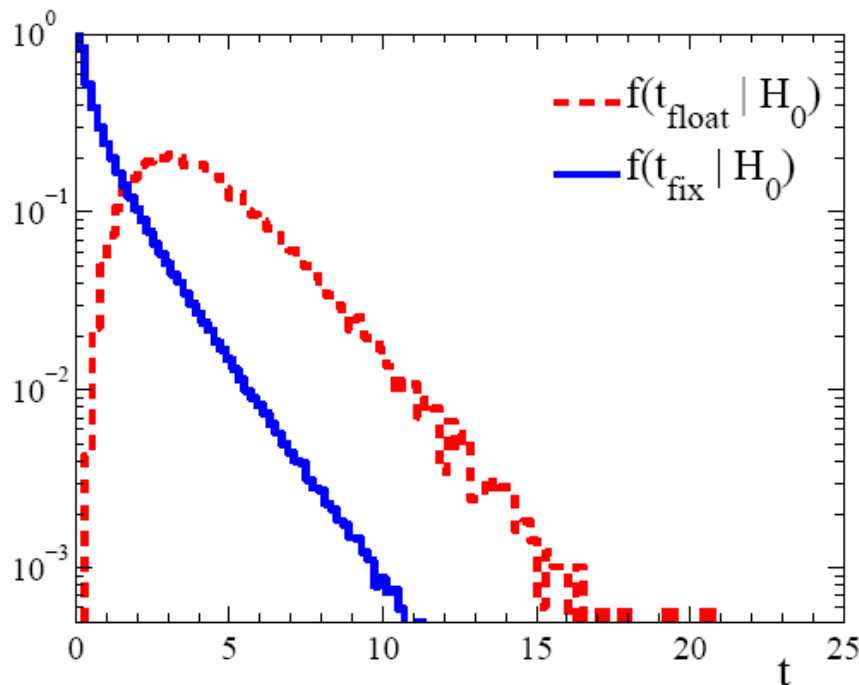
$$t_{\text{float}} = -2 \ln \frac{L(0)}{L(\hat{\mu}, \hat{m})} \quad (\text{Note } m \text{ does not appear in the } \mu = 0 \text{ model.})$$

$$p_{\text{float}} = \int_{t_{\text{float,obs}}}^{\infty} f(t_{\text{float}}|0) dt_{\text{float}}$$

Distributions of t_{fix} , t_{float}

For a sufficiently large data sample, $t_{\text{fix}} \sim \text{chi-square}$ for 1 degree of freedom (Wilks' theorem).

For t_{float} there are two adjustable parameters, μ and m , and naively Wilks theorem says $t_{\text{float}} \sim \text{chi-square}$ for 2 d.o.f.



In fact Wilks' theorem does not hold in the floating mass case because one of the parameters (m) is not defined in the $\mu = 0$ model.

So getting t_{float} distribution is more difficult.

Approximate correction for LEE

We would like to be able to relate the p -values for the fixed and floating mass analyses (at least approximately).

Gross and Vitells show the p -values are approximately related by

$$p_{\text{float}} \approx p_{\text{fix}} + \langle N(c) \rangle$$

where $\langle N(c) \rangle$ is the mean number “upcrossings” of $-2\ln L$ in the fit range based on a threshold

$$c = t_{\text{fix}} = Z_{\text{fix}}^2$$

and where Z_{fix} is the significance for the fixed mass case.

So we can either carry out the full floating-mass analysis (e.g. use MC to get p -value), or do fixed mass analysis and apply a correction factor (much faster than MC).

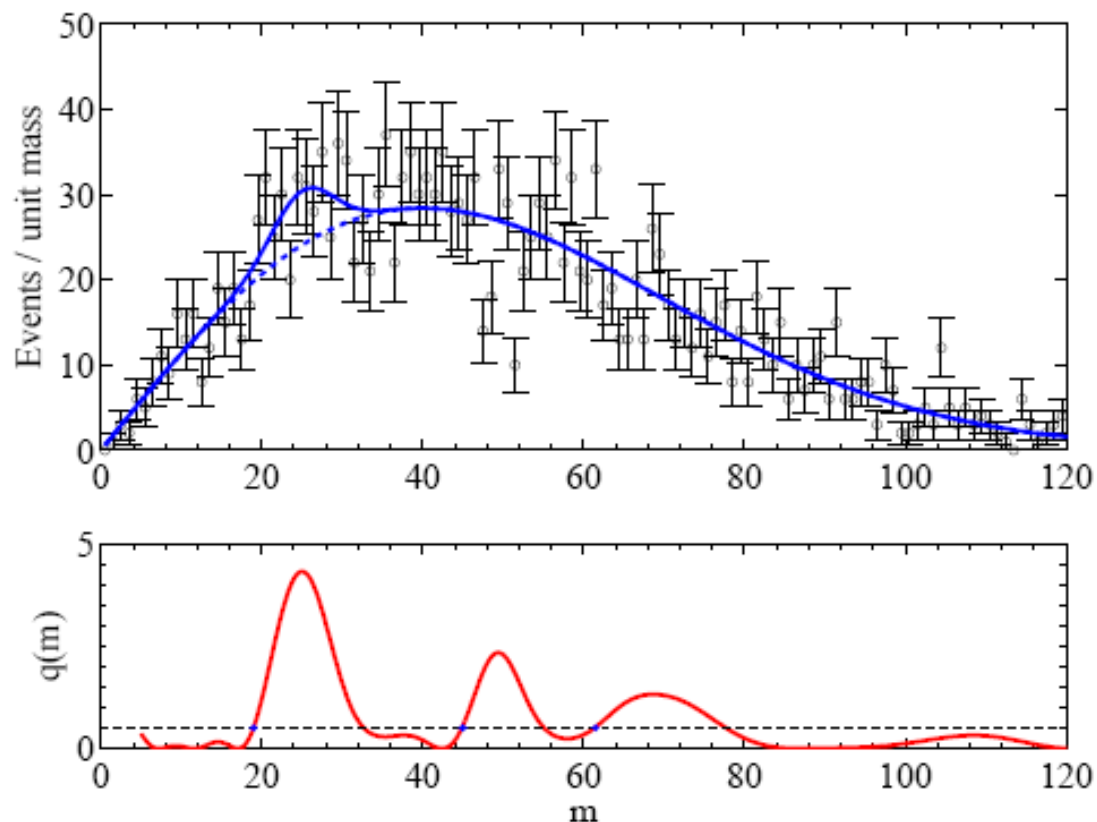
Upcrossings of $-2\ln L$

The Gross-Vitells formula for the trials factor requires $\langle N(c) \rangle$, the mean number “upcrossings” of $-2\ln L$ in the fit range based on a threshold $c = t_{\text{fix}} = Z_{\text{fix}}^2$.

$\langle N(c) \rangle$ can be estimated from MC (or the real data) using a much lower threshold c_0 :

$$\langle N(c) \rangle \approx \langle N(c_0) \rangle e^{-(c-c_0)/2}$$

In this way $\langle N(c) \rangle$ can be estimated without need of large MC samples, even if the the threshold c is quite high.

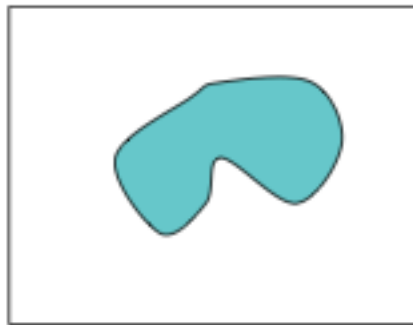


Multidimensional look-elsewhere effect

Generalization to multiple dimensions: number of upcrossings replaced by expectation of Euler characteristic:

$$E[\varphi(A_u)] = \sum_{d=0}^n \mathcal{N}_d \rho_d(u)$$

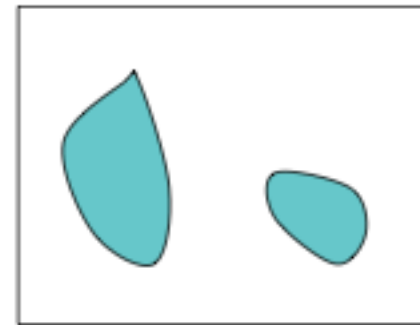
- Number of disconnected components minus number of 'holes'



$\varphi=1$



$\varphi=0$



$\varphi=2$

Applications: astrophysics (coordinates on sky), search for resonance of unknown mass and width, ...

Summary on Look-Elsewhere Effect

Remember the Look-Elsewhere Effect is when we test a single model (e.g., SM) with multiple observations, i.e., in multiple places.

Note there is no look-elsewhere effect when considering exclusion limits. There we test specific signal models (typically once) and say whether each is excluded.

With exclusion there is, however, the analogous issue of testing many signal models (or parameter values) and thus excluding some even in the absence of signal (“spurious exclusion”)

Approximate correction for LEE should be sufficient, and one should also report the uncorrected significance.

“There's no sense in being precise when you don't even know what you're talking about.” — John von Neumann

Why 5 sigma?

Common practice in HEP has been to claim a discovery if the p -value of the no-signal hypothesis is below 2.9×10^{-7} , corresponding to a significance $Z = \Phi^{-1}(1 - p) = 5$ (a 5σ effect).

There a number of reasons why one may want to require such a high threshold for discovery:

- The “cost” of announcing a false discovery is high.

- Unsure about systematics.

- Unsure about look-elsewhere effect.

- The implied signal may be a priori highly improbable (e.g., violation of Lorentz invariance).

Why 5 sigma (cont.)?

But the primary role of the p -value is to quantify the probability that the background-only model gives a statistical fluctuation as big as the one seen or bigger.

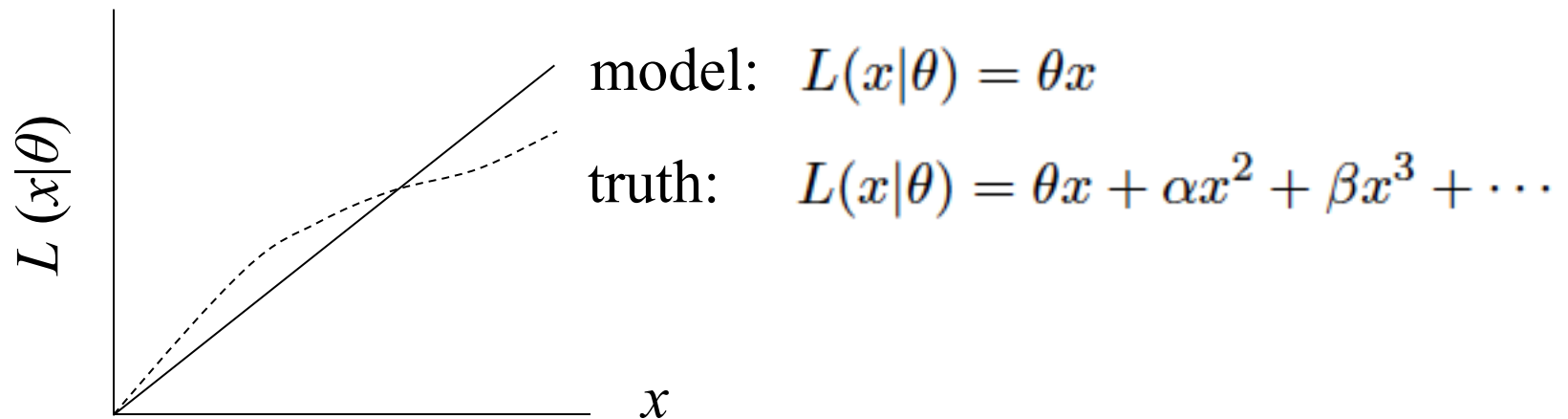
It is not intended as a means to protect against hidden systematics or the high standard required for a claim of an important discovery.

In the processes of establishing a discovery there comes a point where it is clear that the observation is not simply a fluctuation, but an “effect”, and the focus shifts to whether this is new physics or a systematic.

Providing LEE is dealt with, that threshold is probably closer to 3σ than 5σ .

Nuisance parameters

In general our model of the data is not perfect:



Can improve model by including additional adjustable parameters.

$$L(x|\theta) \rightarrow L(x|\theta, \nu)$$

Nuisance parameter \leftrightarrow systematic uncertainty. Some point in the parameter space of the enlarged model should be “true”.

Presence of nuisance parameter decreases sensitivity of analysis to the parameter of interest (e.g., increases variance of estimate).

p -values in cases with nuisance parameters

Suppose we have a statistic q_θ that we use to test a hypothesized value of a parameter θ , such that the p -value of θ is

$$p_\theta = \int_{q_{\theta, \text{obs}}}^{\infty} f(q_\theta | \theta, \nu) dq_\theta$$

But what values of ν to use for $f(q_\theta | \theta, \nu)$?

Fundamentally we want to reject θ only if $p_\theta < \alpha$ for all ν .

→ “exact” confidence interval

Recall that for statistics based on the profile likelihood ratio, the distribution $f(q_\theta | \theta, \nu)$ becomes independent of the nuisance parameters in the large-sample limit.

But in general for finite data samples this is not true; one may be unable to reject some θ values if all values of ν must be considered, even those strongly disfavoured by the data (resulting interval for θ “overcovers”).

Profile construction (“hybrid resampling”)

K. Cranmer, PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics, 2008.
oai:cds.cern.ch:1021125, cdsweb.cern.ch/record/1099969.

Compromise procedure is to reject θ if $p_\theta \leq \alpha$ where the p -value is computed assuming the value of the nuisance parameter that best fits the data for the specified θ :

$$\hat{\hat{v}}(\theta)$$

“double hat” notation means value of parameter that maximizes likelihood for the given θ .

The resulting confidence interval will have the correct coverage for the points $(\theta, \hat{\hat{v}}(\theta))$.

Elsewhere it may under- or overcover, but this is usually as good as we can do (check with MC if crucial or small sample problem).

“Hybrid frequentist-Bayesian” method

Alternatively, suppose uncertainty in ν is characterized by a Bayesian prior $\pi(\nu)$.

Can use the marginal likelihood to model the data:

$$L_{\text{m}}(x|\theta) = \int L(x|\theta, \nu)\pi(\nu) d\nu$$

This does not represent what the data distribution would be if we “really” repeated the experiment, since then ν would not change.

But the procedure has the desired effect. The marginal likelihood effectively builds the uncertainty due to ν into the model.

Use this now to compute (frequentist) p -values \rightarrow result has hybrid “frequentist-Bayesian” character.

The “ur-prior” behind the hybrid method

But where did $\pi(\nu)$ come from? Presumably at some earlier point there was a measurement of some data y with likelihood $L(y|\nu)$, which was used in Bayes’ theorem,

$$\pi(\nu|y) \propto L(y|\nu)\pi_0(\nu)$$

and this “posterior” was subsequently used for $\pi(\nu)$ for the next part of the analysis.

But it depends on an “ur-prior” $\pi_0(\nu)$, which still has to be chosen somehow (perhaps “flat-ish”).

But once this is combined to form the marginal likelihood, the origin of the knowledge of ν may be forgotten, and the model is regarded as only describing the data outcome x .

The (pure) frequentist equivalent

In a purely frequentist analysis, one would regard both x and y as part of the data, and write down the full likelihood:

$$L(x, y|\theta, \nu) = L(x|\theta, \nu)L(y|\nu)$$

“Repetition of the experiment” here means generating both x and y according to the distribution above.

In many cases, the end result from the hybrid and pure frequentist methods are found to be very similar (cf. Conway, Roever, PHYSTAT 2011).

More on priors

Suppose we measure $n \sim \text{Poisson}(s+b)$, goal is to make inference about s .

Suppose b is not known exactly but we have an estimate b_{meas} with uncertainty σ_b .

For Bayesian analysis, first reflex may be to write down a Gaussian prior for b ,

$$\pi(b) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(b-b_{\text{meas}})^2/2\sigma_b^2}$$

But a Gaussian could be problematic because e.g.

$b \geq 0$, so need to truncate and renormalize;

tails fall off very quickly, may not reflect true uncertainty.

Bayesian limits on s with uncertainty on b

Consider $n \sim \text{Poisson}(s+b)$ and take e.g. as prior probabilities

$$\pi(s, b) = \pi_s(s)\pi_b(b) \quad (\text{or include correlations as appropriate})$$

$$\pi_s(s) = \text{const}, \sim 1/\sqrt{s+b} \dots$$

$$\pi_b(b) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(b-b_{\text{meas}})^2/2\sigma_b^2} \quad (\text{or whatever})$$

Put this into Bayes' theorem,

$$p(s, b|n) \propto L(n|s, b)\pi(s, b)$$

Marginalize over the nuisance parameter b ,

$$p(s|n) = \int p(s, b|n) db$$

Then use $p(s|n)$ to find intervals for s with any desired probability content.

Gamma prior for b

What is in fact our prior information about b ? It may be that we estimated b using a separate measurement (e.g., background control sample) with

$$m \sim \text{Poisson}(\tau b) \quad (\tau = \text{scale factor, here assume known})$$

Having made the control measurement we can use Bayes' theorem to get the probability for b given m ,

$$\pi(b|m) \propto P(m|b)\pi_0(b) \propto \frac{(\tau b)^m}{m!} e^{-\tau b} \pi_0(b)$$

If we take the ur-prior $\pi_0(b)$ to be constant for $b \geq 0$, then the posterior $\pi(b|m)$, which becomes the subsequent prior when we measure n and infer s , is a **Gamma distribution** with:

$$\text{mean} = (m + 1) / \tau$$

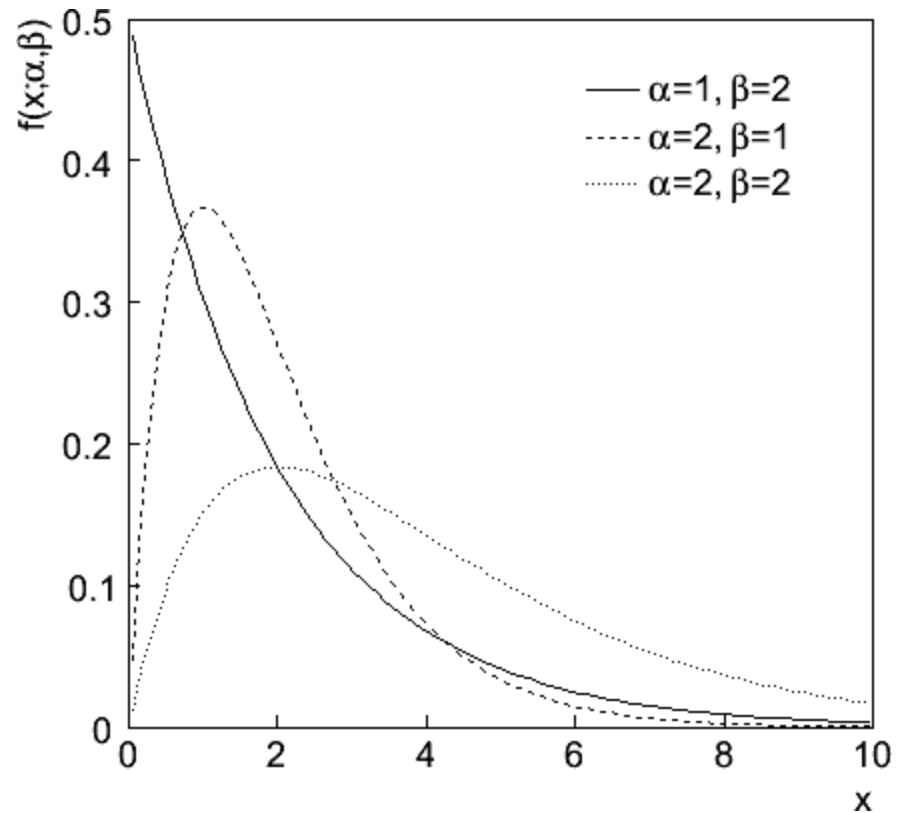
$$\text{standard dev.} = \sqrt{(m + 1) / \tau}$$

Gamma distribution

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

$$E[x] = \alpha\beta$$

$$V[x] = \alpha\beta^2$$



Frequentist test with Bayesian treatment of b

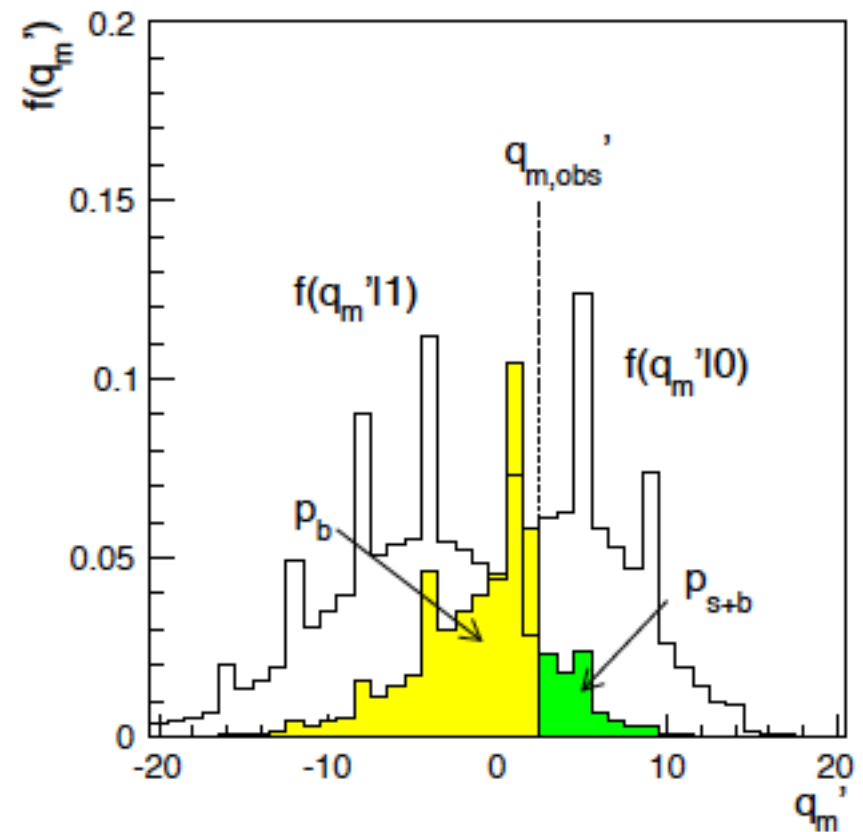
Distribution of n based on marginal likelihood (gamma prior for b):

$$P_m(n|\mu) = \int P(n|\mu, b)\pi(b) db$$

and use this as the basis of a test statistic:

$$q_m = -2 \ln \frac{P_m(n|1)}{P_m(n|0)}$$

p -values from distributions of q_m under background-only (0) or signal plus background (1) hypotheses:



Frequentist approach to same problem

In the frequentist approach we would regard both variables

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

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

as constituting the data, and thus the full likelihood function is

$$L(s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-\tau b}$$

Use this to construct test of s with e.g. profile likelihood ratio

$$\lambda(s) = \frac{L(s, \hat{\hat{b}})}{L(\hat{s}, \hat{b})}$$

Note here that the likelihood refers to both n and m , whereas the likelihood used in the Bayesian calculation only modeled n .

Test based on fully frequentist treatment

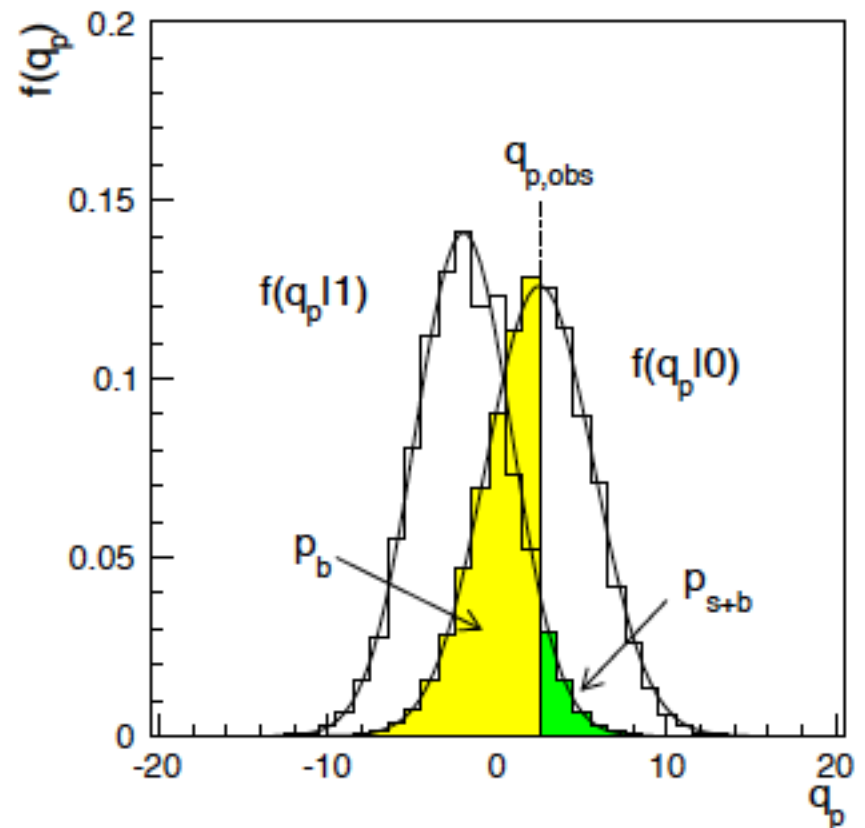
Data consist of both n and m , with distribution

$$P(n, m|\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \frac{(\tau b)^m}{m!} e^{-\tau b}$$

Use this as the basis of a test statistic based on ratio of profile likelihoods:

$$q_p = -2 \ln \frac{P(n, m|1, \hat{b}(1))}{P(n, m|0, \hat{b}(0))}$$

Here combination of two discrete variables (n and m) results in an approximately continuous distribution for q_p .



Log-normal prior for systematics

In some cases one may want a log-normal prior for a nuisance parameter (e.g., background rate b).

$$\pi_b(b) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{b} \exp \left[-\frac{(\ln(b/b_0))^2}{2\sigma^2} \right]$$

This would emerge from the Central Limit Theorem, e.g., if the true parameter value is uncertain due to a large number of multiplicative changes, and it corresponds to having a Gaussian prior for $\beta = \ln b$.

$$\pi_\beta(\beta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(\beta - \beta_0)^2}{2\sigma^2} \right]$$

where $\beta_0 = \ln b_0$ and in the following we write σ as σ_β .

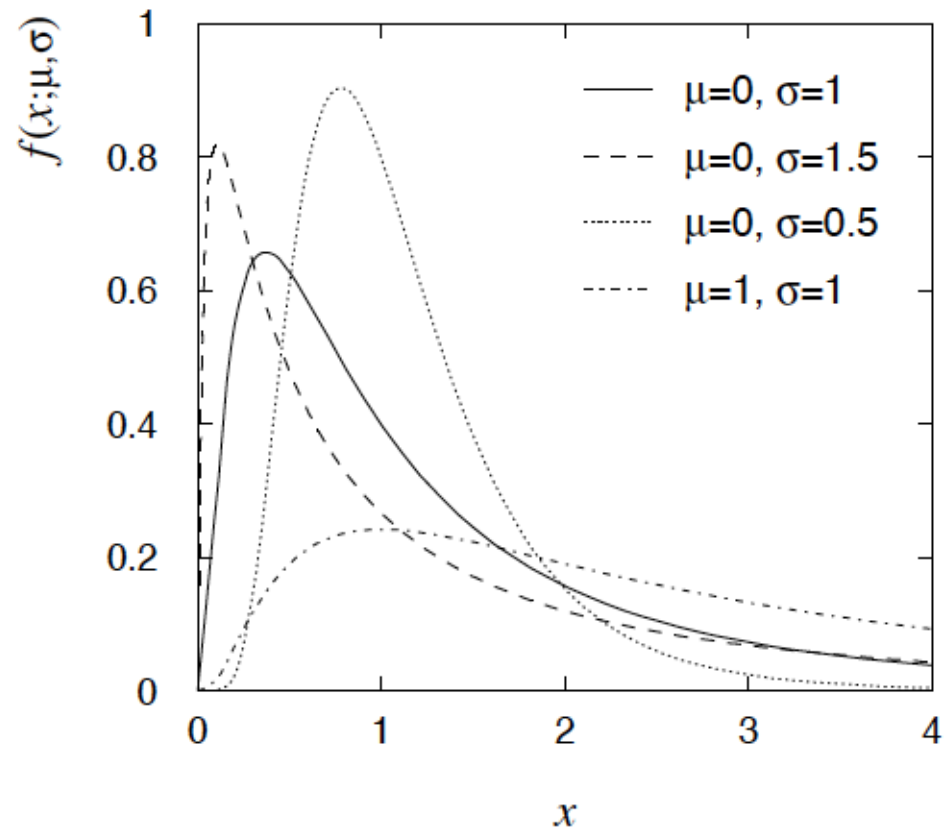
The log-normal distribution

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(\frac{-(\log x - \mu)^2}{2\sigma^2}\right)$$

$$E[x] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

$$V[x] =$$

$$\exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$



Frequentist-Bayes correspondence for log-normal

The corresponding frequentist treatment regards the best estimate of b as a measured value b_{meas} that is log-normally distributed, or equivalently has a Gaussian distribution for $\beta_{\text{meas}} = \ln b_{\text{meas}}$:

$$p(\beta_{\text{meas}}|\beta) = \frac{1}{\sqrt{2\pi}\sigma_{\beta}} e^{-(\beta_{\text{meas}}-\beta)/2\sigma_{\beta}^2}$$

To use this to motivate a Bayesian prior, one would use Bayes' theorem to find the posterior for β ,

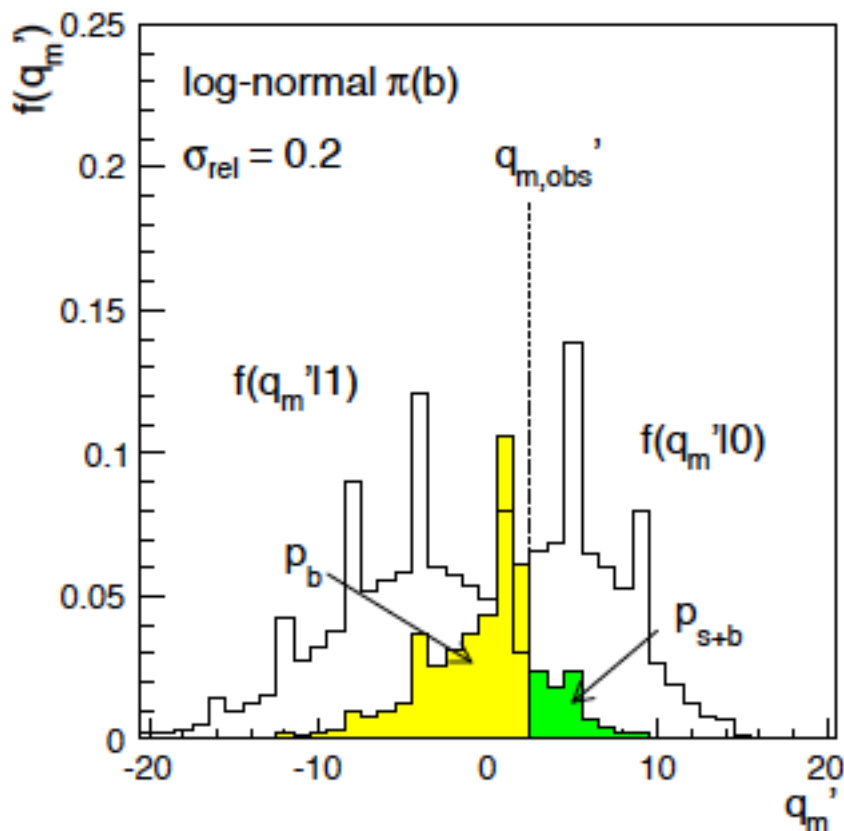
$$p(\beta|\beta_{\text{meas}}) \propto p(\beta_{\text{meas}}|\beta)\pi_{0,\beta}(\beta)$$

If we take the ur-prior $\pi_{0,\beta}(\beta)$ constant, this implies an ur-prior for b of

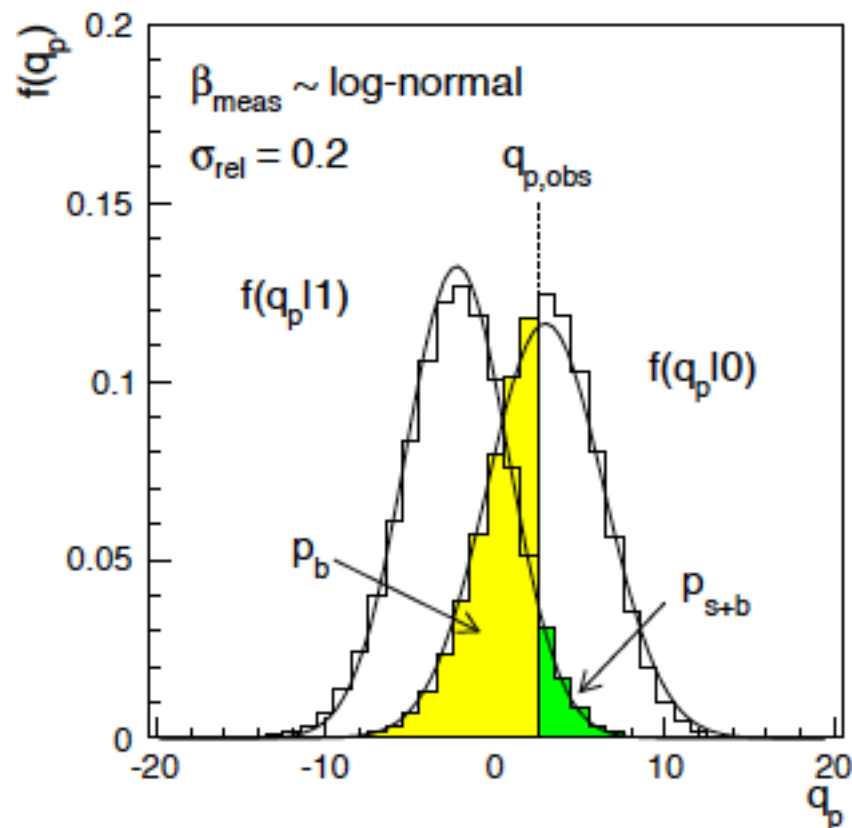
$$\pi_{0,b}(b) = \pi_{0,\beta}(\beta) \left| \frac{d\beta}{db} \right| \propto \frac{1}{b}$$

Example of tests based on log-normal

Bayesian treatment of b :



Frequentist treatment of b_{meas} :



Final result similar but note in Bayesian treatment, marginal model is only for n , which is discrete, whereas in frequentist model both n and continuous b_{meas} are treated as measurements.

Discovery significance for $n \sim \text{Poisson}(s + b)$

Consider again the case where we observe n events, model as following Poisson distribution with mean $s + b$ (assume b is known).

- 1) For an observed n , what is the significance Z_0 with which we would reject the $s = 0$ hypothesis?
- 2) What is the expected (or more precisely, median) Z_0 if the true value of the signal rate is s ?

Gaussian approximation for Poisson significance

For large $s + b$, $n \rightarrow x \sim \text{Gaussian}(\mu, \sigma)$, $\mu = s + b$, $\sigma = \sqrt{s + b}$.

For observed value x_{obs} , p -value of $s = 0$ is $\text{Prob}(x > x_{\text{obs}} | s = 0)$,:

$$p_0 = 1 - \Phi\left(\frac{x_{\text{obs}} - b}{\sqrt{b}}\right)$$

Significance for rejecting $s = 0$ is therefore

$$Z_0 = \Phi^{-1}(1 - p_0) = \frac{x_{\text{obs}} - b}{\sqrt{b}}$$

Expected (median) significance assuming signal rate s is

$$\text{median}[Z_0 | s + b] = \frac{s}{\sqrt{b}}$$

Better approximation for Poisson significance

Likelihood function for parameter s is

$$L(s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

or equivalently the log-likelihood is

$$\ln L(s) = n \ln(s+b) - (s+b) - \ln n!$$

Find the maximum by setting $\frac{\partial \ln L}{\partial s} = 0$

gives the estimator for s : $\hat{s} = n - b$

Approximate Poisson significance (continued)

The likelihood ratio statistic for testing $s = 0$ is

$$q_0 = -2 \ln \frac{L(0)}{L(\hat{s})} = 2 \left(n \ln \frac{n}{b} + b - n \right) \quad \text{for } n > b, \text{ 0 otherwise}$$

For sufficiently large $s + b$, (use Wilks' theorem),

$$Z_0 \approx \sqrt{q_0} = \sqrt{2 \left(n \ln \frac{n}{b} + b - n \right)} \quad \text{for } n > b, \text{ 0 otherwise}$$

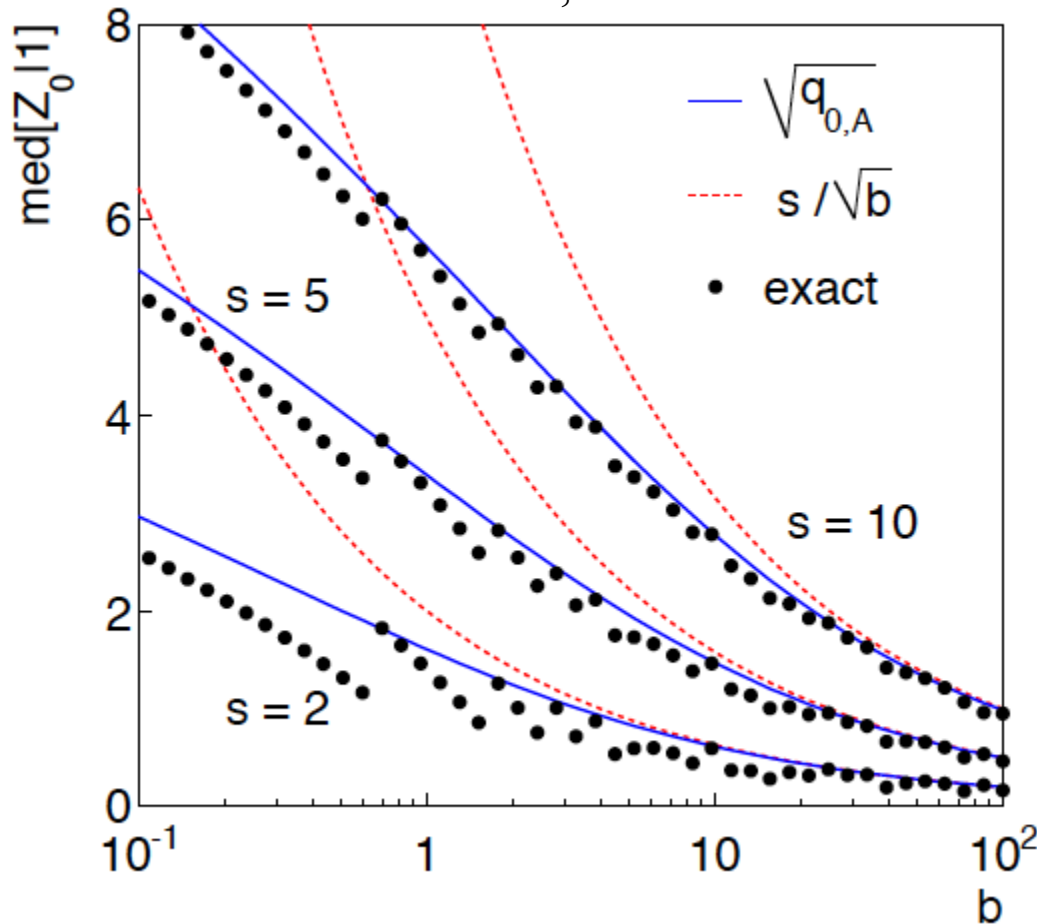
To find $\text{median}[Z_0|s+b]$, let $n \rightarrow s + b$ (i.e., the Asimov data set):

$$\text{median}[Z_0|s + b] \approx \sqrt{2 \left((s + b) \ln(1 + s/b) - s \right)}$$

This reduces to s/\sqrt{b} for $s \ll b$.

$n \sim \text{Poisson}(\mu s + b)$, median significance, assuming $\mu = 1$, of the hypothesis $\mu = 0$

CCGV, arXiv:1007.1727



“Exact” values from MC, jumps due to discrete data.

Asimov $\sqrt{q_{0,A}}$ good approx. for broad range of s, b .

s/\sqrt{b} only good for $s \ll b$.

Summary of Lecture 3

Bayesian treatment of limits is conceptually easy (integrate posterior pdf); appropriate choice of prior not obvious.

Look-Elsewhere Effect

Need to give probability to see a signal as big as the one you saw (or bigger) anywhere you looked. Hard to define precisely; approximate correction should be adequate.

Why 5 sigma? If LEE taken in to account, one is usually convinced the effect is not a fluctuation much earlier (at 3 sigma?)

Nuisance parameters

Need enough in model so that for at least some point in parameter space it is correct.

Profile or marginalize. (Profiling allows use of asymptotic formulae.)

Extra slides

Maximize the expected Kullback–Leibler divergence of posterior relative to prior:

$$D[\pi, p] \equiv \int p(\theta|x) \ln \frac{p(\theta|x)}{\pi(\theta)} d\theta$$

This maximizes the expected posterior information about θ when the prior density is $\pi(\theta)$.

Finding reference priors “easy” for one parameter:

Theorem 1 *Let $\mathbf{z}^{(k)} = \{z_1, \dots, z_k\}$ denote k conditionally independent observations from \mathcal{M}_z . For sufficiently large k*

$$\pi_k(\theta) \propto \exp \{E_{\mathbf{z}^{(k)}|\theta}[\log p_h(\theta | \mathbf{z}^{(k)})]\}$$

where $p_h(\theta | \mathbf{z}^{(k)}) \propto \prod_{i=1}^k p(z_i | \theta) h(\theta)$ is the posterior which corresponds to any arbitrarily chosen strictly positive prior function $h(\theta)$ which makes the posterior proper for any $\mathbf{z}^{(k)}$.

Reference priors (2)

J. Bernardo,
L. Demortier,
M. Pierini

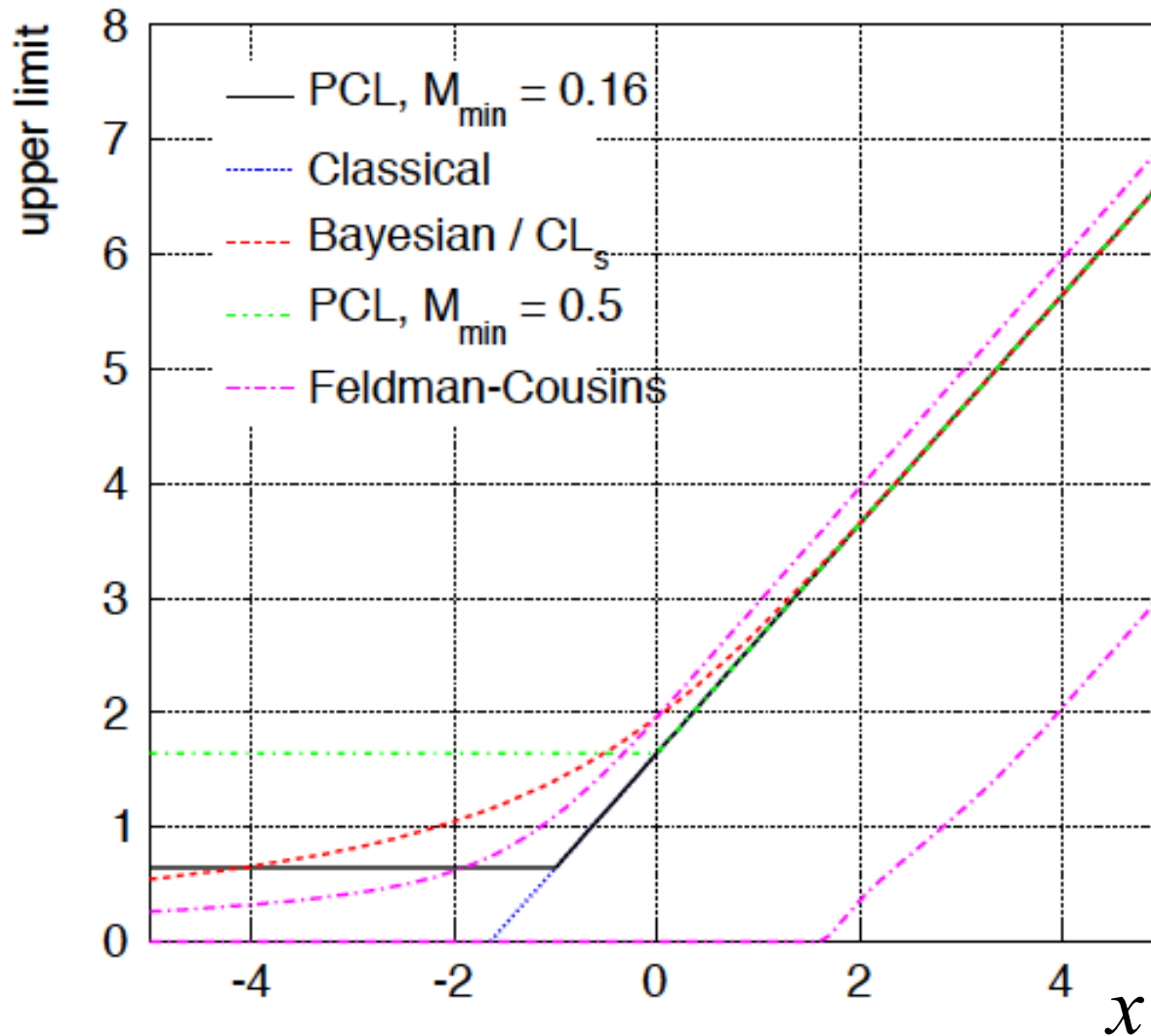
Actual recipe to find reference prior nontrivial;
see references from Bernardo's talk, website of
Berger (www.stat.duke.edu/~berger/papers) and also
Demortier, Jain, Prosper, PRD 82:33, 34002 arXiv:1002.1111:

$$\pi_R(\theta) = \lim_{k \rightarrow \infty} \frac{\pi_k(\theta)}{\pi_k(\theta_0)},$$

$$\text{with } \pi_k(\theta) = \exp \left\{ \int p(x_{(k)} | \theta) \ln \left[\frac{p(x_{(k)} | \theta) h(\theta)}{\int p(x_{(k)} | \theta) h(\theta) d\theta} \right] dx_{(k)} \right\}$$

Prior depends on order of parameters. (Is order dependence important? Symmetrize? Sample result from different orderings?)

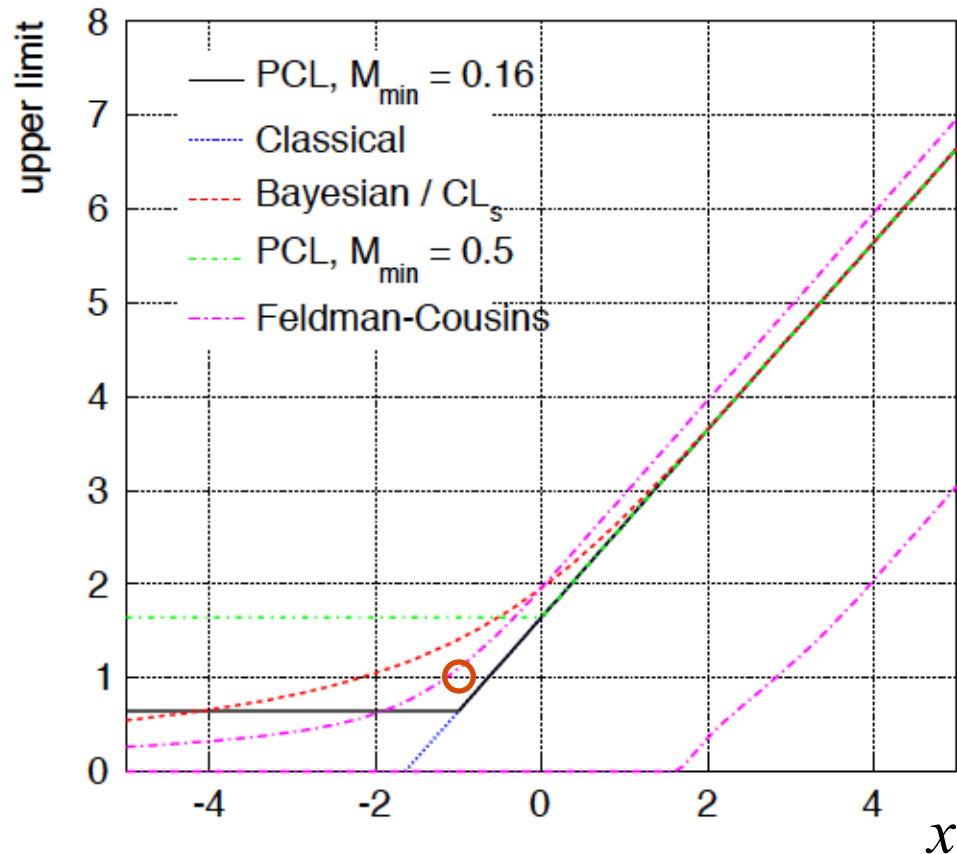
Upper limit on μ for $x \sim \text{Gauss}(\mu, \sigma)$ with $\mu \geq 0$



Comparison of reasons for (non)-exclusion

Suppose we observe $x = -1$.

$\mu = 1$ excluded by diag. line,
why not by other methods?

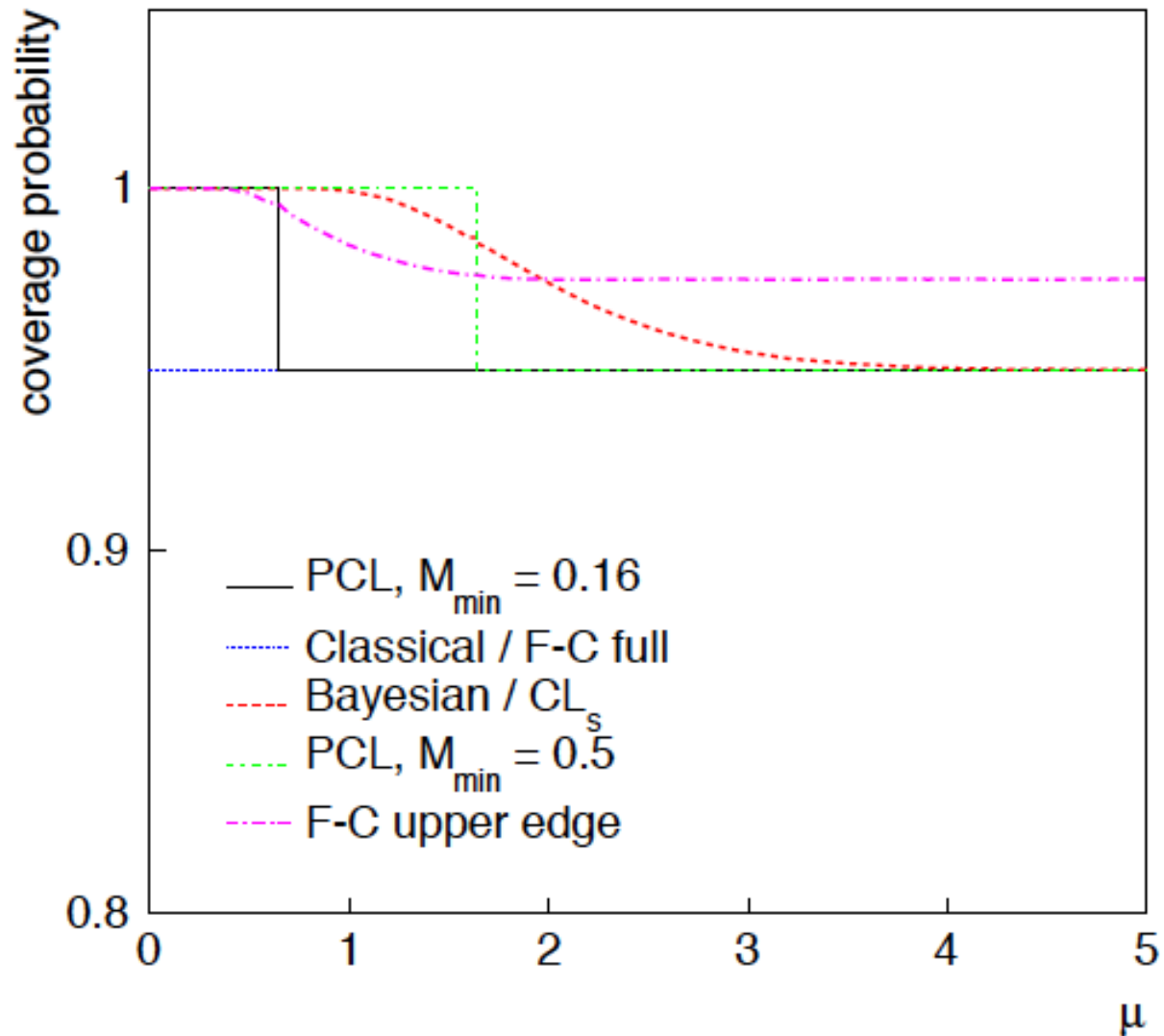


PCL ($M_{\min}=0.5$): Because the power of a test of $\mu = 1$ was below threshold.

CLs: Because the lack of sensitivity to $\mu = 1$ led to reduced $1 - p_b$, hence CL_s not less than α .

F-C: Because $\mu = 1$ was not rejected in a test of size α (hence coverage correct). But the critical region corresponding to more than half of α is at high x .

Coverage probability for Gaussian problem



Flip-flopping

F-C pointed out that if one decides, based on the data, whether to report a one- or two-sided limit, then the stated coverage probability no longer holds.

The problem (flip-flopping) is avoided in unified intervals.

Whether the interval covers correctly or not depends on how one defines repetition of the experiment (the ensemble).

Need to distinguish between:

- (1) an idealized ensemble;
- (2) a recipe one follows in real life that resembles (1).

Flip-flopping

One could take, e.g.:

Ideal: always quote upper limit (∞ # of experiments).

Real: quote upper limit for as long as it is of any interest, i.e., until the existence of the effect is well established.

The coverage for the idealized ensemble is correct.

The question is whether the real ensemble departs from this during the period when the limit is of any interest as a guide in the search for the signal.

Here the real and ideal only come into serious conflict if you think the effect is well established (e.g. at the 5 sigma level) but then subsequently you find it not to be well established, so you need to go back to quoting upper limits.

Flip-flopping

In an idealized ensemble, this situation could arise if, e.g., we take $x \sim \text{Gauss}(\mu, \sigma)$, and the true μ is one sigma below what we regard as the threshold needed to discover that μ is nonzero.

Here flip-flopping gives undercoverage because one continually bounces above and below the discovery threshold. The effect keeps going in and out of a state of being established.

But this idealized ensemble does not resemble what happens in reality, where the discovery sensitivity continues to improve as more data are acquired.