

# Search Procedures in High Energy Physics

Luc Demortier

*The Rockefeller University*

O-Bayes09, Philadelphia, USA, June 5–9, 2009

## Outline

- 1 Experimental high energy physics
- 2 Reference priors with partial information
- 3 Upper limits and central intervals
- 4 Coverage studies
- 5 Intrinsic intervals
- 6 Search procedures
- 7 Open questions

## Experimental High Energy Physics

One of the main goals of experimental high energy physics is to discover new elementary particles among the products of the particle collisions created at large accelerator centers such as those at Fermilab, near Chicago, IL, and at CERN, near Geneva, Switzerland.

Although accelerators typically produce millions of collisions per second, *interesting* collisions (“signal”) are very rare. Thus, the process of selecting such collisions is very stringent, and the number of collision events selected for further study obeys Poisson statistics to a very good approximation:

$$N \sim \frac{(\mu + \epsilon\sigma)^n}{n!} e^{-\mu - \epsilon\sigma},$$

where

- 1  $\mu$  is the background contamination: this is the expected number of events that are not signal but are essentially indistinguishable from it.
- 2  $\epsilon\sigma$  is the expected number of signal events, written as a product of the **effective luminosity**  $\epsilon$  and a quantity of intrinsic physics interest, the **signal cross section**  $\sigma$ .

Usually there is independent information available about  $\mu$  and  $\epsilon$ , either from simulation studies or from subsidiary measurements. The parameter of interest is  $\sigma$ .

## The Standard Search Procedure

For mainly historical reasons, physicists tend to favor a frequentist approach to statistical inference. This includes their standard search procedure:

- 1 Choose three confidence levels  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ .
- 2 Find a test statistic  $T$  that is sensitive to the presence of signal.
- 3 Compute the  $p$  value corresponding to the observed value of  $T$ .
- 4 If  $p \leq 1 - \alpha_1$ , claim discovery and compute an  $\alpha_2$  confidence level two-sided interval on the signal cross section  $\sigma$ . The purpose of this interval is to provide an estimate of the magnitude of the observed effect.
- 5 If  $p > 1 - \alpha_1$ , make no claim, and compute an  $\alpha_3$  confidence level upper limit on  $\sigma$ . The purpose of this upper limit is to provide an upper bound on values of  $\sigma$  that the experiment cannot convincingly distinguish from zero.

Typical choices of confidence levels are  $\alpha_1 = 1 - 2.8 \times 10^{-7}$ ,  $\alpha_2 = 0.68$ , and  $\alpha_3 = 0.95$ . The stringency of the  $p$  value test (discovery requires a  $5\sigma$  deviation from the background level) is due to the so-called look-elsewhere effect: high energy physicists examine large numbers of independent histograms every year, so that one expects several random background fluctuations at the  $4\sigma$  level on that time scale.

## Problems with the Standard Search Procedure (1/2)

### 1 Frequentist coverage.

The type of interval that is reported (two-sided or upper limit) depends on the result of the  $p$  value test, but the interval is constructed without taking the test into account, i.e. as if the interval type had been chosen *before* looking at the data. As a consequence, the true frequentist coverage of the interval is below the stated value for some values of  $\sigma$ .

### 2 Empty intervals.

For some values of the background  $\mu$  and the observed number of events  $n$ , the resulting interval for  $\sigma$  is empty. While this is correct from a frequentist point of view, it is unsatisfying from a scientific one, since an empty interval cannot possibly contain the true value of  $\sigma$ .

### 3 Sensitivity to expected background.

For a given observed number of events, the upper limit tends to decrease very fast as a function of the expected background. Thus, large background contamination could in principle allow one to exclude more signal than small contamination.

### 4 Nuisance parameters.

The proper frequentist treatment of problems with many nuisance parameters can be challenging.

## Problems with the Standard Search Procedure (2/2)

Within a frequentist framework, the empty-interval problem of the standard search procedure can be solved by using a likelihood-ratio ordering rule for constructing intervals, instead of a central or upper-limit ordering rule. However, this approach still does not quite solve the other problems.

What we propose to do instead is to develop a solution based on **Bayesian reference analysis**. There are many advantages to such an approach: inferences are invariant under parameter transformations and behave well under measurement replication, the method is very general and computationally tractable, it is “objective Bayesian” while avoiding the marginalization paradoxes, and it can be embedded in a subjective Bayesian framework.

In contrast with many applications of reference analysis, we will not require the full power of the method with regard to nuisance parameters. In high energy physics we usually have partial information about these, which simplifies some of the calculations. As we will see however, this also brings up some interesting new issues.

## Reference Priors with Partial Information

Suppose that  $\phi$  labels the nuisance parameter(s) and  $\theta$  the parameter of interest. If we have prior information about  $\phi$ , there are two ways to proceed (Sun & Berger 1998):

**Method 1:** We are given a marginal prior  $\pi(\phi)$  for  $\phi$ .  
In this case we need the conditional reference prior  $\pi_R(\theta | \phi)$  for  $\theta$  given a fixed value of  $\phi$ .

**Method 2:** We are given a conditional prior  $\pi(\phi | \theta)$  for  $\phi$  given  $\theta$ .  
In this case we can marginalize the probability model  $p(x|\theta, \phi)$  with respect to  $\phi$  in order to obtain  $p(x|\theta) = \int p(x|\theta, \phi) \pi(\phi|\theta) d\phi$ , and we can then compute the reference prior  $\pi_R(\theta)$  for the marginalized model.

Often there are physics reasons for assuming that prior information about  $\phi$  is independent of  $\theta$ . An experimenter who is willing to make such an assumption can declare that  $\pi(\phi | \theta) = \pi(\phi)$  and use Method 2. Otherwise Method 1 will be preferable.

An important practical advantage of Method 1 is that the conditional reference prior is computed once and for all, for a given model, and can be used with any subjective nuisance prior. In contrast, for Method 2 the reference prior must be recomputed every time the subjective prior changes.

## Method 1 Applied to a Single-Count Measurement (1/3)

The single-count measurement is characterized by the likelihood:

$$p(n|\sigma, \epsilon, \mu) = \frac{(\epsilon\sigma + \mu)^n}{n!} e^{-\epsilon\sigma - \mu} \quad \text{with } \sigma \geq 0 \quad \text{and } \epsilon, \mu > 0.$$

In Method 1 we construct the conditional reference prior  $\pi_R(\sigma | \epsilon, \mu)$ . The first step consists in calculating Jeffreys' prior for  $\sigma$  while holding  $\epsilon$  and  $\mu$  fixed:

$$\pi_J(\sigma | \epsilon, \mu) \propto \left\{ \mathbb{E} \left[ -\frac{\partial^2}{\partial \sigma^2} \ln p(n | \sigma, \epsilon, \mu) \right] \right\}^{\frac{1}{2}} \propto \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}}.$$

This prior is improper with respect to  $\sigma$  however, so that an additional step, known as the “compact support argument,” is required:

Choose a nested sequence  $\Lambda_1 \subset \Lambda_2 \subset \dots$  of compact subsets of the parameter space  $\Lambda$  for  $(\sigma, \epsilon, \mu)$ , such that  $\cup_i \Lambda_i = \Lambda$  and the integral  $K_i(\epsilon, \mu)$  of  $\pi_J(\sigma | \epsilon, \mu)$  over  $\Omega_i \equiv \{\sigma : (\sigma, \epsilon, \mu) \in \Lambda_i\}$  is finite. Then, on  $\Omega_i$ :

$$\pi_{R,i}(\sigma | \epsilon, \mu) = \frac{\pi_J(\sigma | \epsilon, \mu)}{K_i(\epsilon, \mu)} I_{\Omega_i}(\sigma),$$

and on the whole parameter space:

$$\pi_R(\sigma | \epsilon, \mu) = \lim_{i \rightarrow \infty} \frac{\pi_{R,i}(\sigma | \epsilon, \mu)}{\pi_{R,i}(\sigma_0 | \epsilon_0, \mu_0)} \quad \text{with } (\sigma_0, \epsilon_0, \mu_0) \text{ a fixed point.}$$

## Method 1 Applied to a Single-Count Measurement (2/3)

The theory of reference priors provides few guidelines for choosing the compact sets  $\Lambda_i$ . Try:

$$\Lambda_i = \left\{ (\sigma, \epsilon, \mu) : \sigma \in [0, u_i], \epsilon \in [0, v_i], \mu \in [0, w_i] \right\},$$

where  $\{u_i\}$ ,  $\{v_i\}$ , and  $\{w_i\}$  are increasing sequences of positive constants. Then:

$$\pi_{R,i}(\sigma | \epsilon, \mu) = \frac{1}{K_i(\epsilon, \mu)} \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}} I_{[0, u_i]}(\sigma),$$

where:

$$K_i(\epsilon, \mu) \equiv \int_{\Omega_i} \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}} d\sigma = 2 [\sqrt{\epsilon u_i + \mu} - \sqrt{\mu}],$$

and therefore:

$$\pi_R(\sigma | \epsilon, \mu) \propto \sqrt{\frac{\epsilon}{\epsilon\sigma + \mu}}.$$

This is improper with respect to  $\sigma$ , but the  $\epsilon$  dependence differs from that of the conditional Jeffreys' prior. There is a problem however. If

$$\pi(\epsilon, \mu) = \frac{e^{-\epsilon}}{\sqrt{\pi\epsilon/2}} \pi(\mu),$$

then the resulting posterior  $p(n | \sigma, \epsilon, \mu) \pi_R(\sigma | \epsilon, \mu) \pi(\epsilon, \mu)$  is improper!

## Method 1 Applied to a Single-Count Measurement (3/3)

The cause of this problem is the choice of compact sets. Note that the Jeffreys prior for this problem,  $\pi_J(\sigma | \epsilon, \mu) d\sigma$ , is invariant under scale transformations  $\epsilon \rightarrow c\epsilon$ ,  $\sigma \rightarrow \sigma/c$ , where  $c$  is constant. Our initial choice of compact sets does not share this invariance, so we try instead:

$$\Lambda_i = \left\{ (\sigma, \epsilon, \mu) : \sigma \in [0, u_i/\epsilon], \epsilon \in [1/v_i, v_i], \mu \in [0, w_i] \right\},$$

where  $u_i$ ,  $v_i$ , and  $w_i$  are as before. Repeating the same calculation as before, we now find:

$$\pi_{R1}(\sigma | \epsilon, \mu) \propto \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}},$$

which is identical to Jeffreys' prior for this case and yields well-behaved posteriors (thanks to Jim Berger and Dongchu Sun for help with this problem!).

## Evidence-Based Prior for Background and Effective Luminosity

To do further calculations with the Method 1 conditional reference prior we must specify a subjective prior for  $\epsilon$  and  $\mu$ . Here we take:

$$\pi(\epsilon, \mu) = \frac{a(a\epsilon)^{x-1/2} e^{-a\epsilon}}{\Gamma(x+1/2)} \frac{b(b\mu)^{y-1/2} e^{-b\mu}}{\Gamma(y+1/2)},$$

A typical situation is that information about  $\mu$  and  $\epsilon$  is obtained from Monte Carlo simulations, subsidiary measurements, and theoretical beliefs... This information is summarized by point estimates ( $\bar{\epsilon}$ ,  $\bar{\mu}$ ) and relative uncertainties ( $\delta\epsilon$ ,  $\delta\mu$ ), which are then identified with the corresponding means and coefficients of variation of the two component distributions of  $\pi(\epsilon, \mu)$ :

$$\bar{\epsilon} = \frac{x + \frac{1}{2}}{a}, \quad \delta\epsilon = \frac{1}{\sqrt{x + \frac{1}{2}}}, \quad \text{or} \quad x = \frac{1}{\delta\epsilon^2} - \frac{1}{2}, \quad a = \frac{1}{\bar{\epsilon} \delta\epsilon^2};$$
$$\bar{\mu} = \frac{y + \frac{1}{2}}{b}, \quad \delta\mu = \frac{1}{\sqrt{y + \frac{1}{2}}}, \quad \text{or} \quad y = \frac{1}{\delta\mu^2} - \frac{1}{2}, \quad b = \frac{1}{\bar{\mu} \delta\mu^2}.$$

Another possible interpretation of  $\pi(\epsilon, \mu)$  is that it is the joint posterior of two Poisson measurements  $x$  and  $y$  of the effective luminosity and background, respectively, and where Jeffreys' prior was used for both  $\epsilon$  and  $\mu$ .

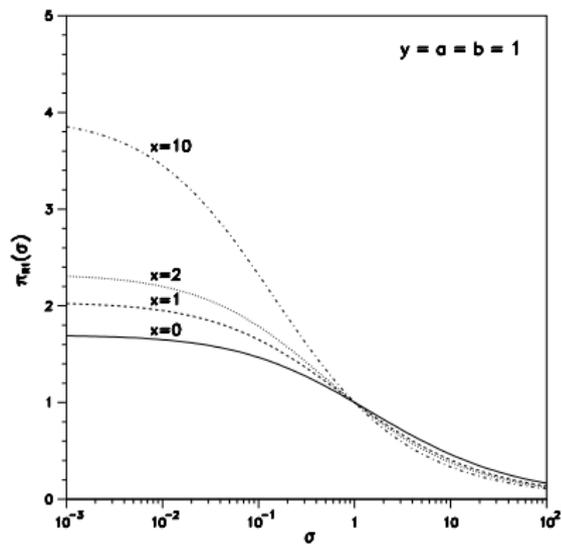
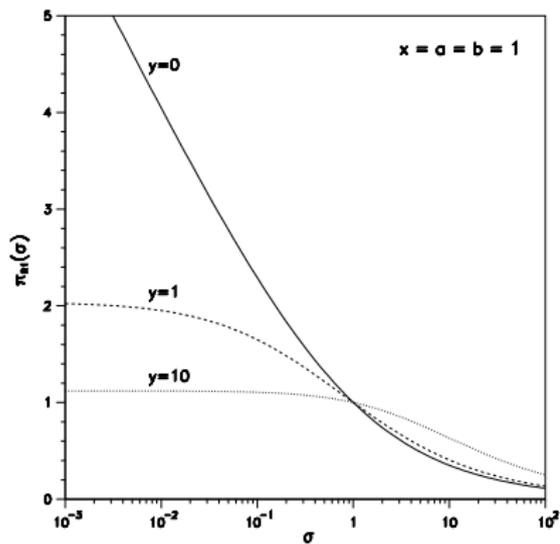
## The Marginal Method-1 Prior (1/2)

The marginal Method-1 cross section prior is:

$$\begin{aligned}\pi_{R1}(\sigma) &\propto \int_0^\infty d\epsilon \int_0^\infty d\mu \frac{\epsilon}{\sqrt{\epsilon\sigma + \mu}} \frac{a(a\epsilon)^{x-1/2} e^{-a\epsilon}}{\Gamma(x + \frac{1}{2})} \frac{b(b\mu)^{y-1/2} e^{-b\mu}}{\Gamma(y + \frac{1}{2})}, \\ &\propto {}_2F_1\left(\frac{1}{2}, x + \frac{3}{2}; x + y + 2; 1 - \frac{b}{a}\sigma\right), \\ &\propto \sqrt{\frac{a}{b\sigma}} {}_2F_1\left(\frac{1}{2}, y + \frac{1}{2}; x + y + 2; 1 - \frac{a}{b\sigma}\right),\end{aligned}$$

where  ${}_2F_1$  is the Gauss hypergeometric function. The last expression shows that  $\pi_{R1}(\sigma)$  is improper and varies as  $1/\sqrt{\sigma}$  at large  $\sigma$ .

## The Marginal Method-1 Prior (2/2)



## The Marginal Method-1 Posterior (1/2)

For the marginal Method-1 posterior we find:

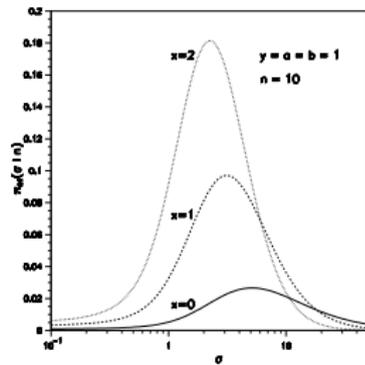
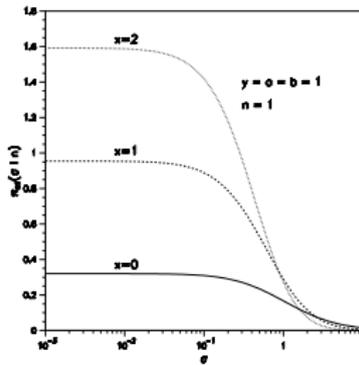
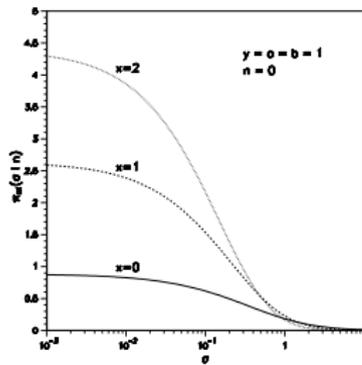
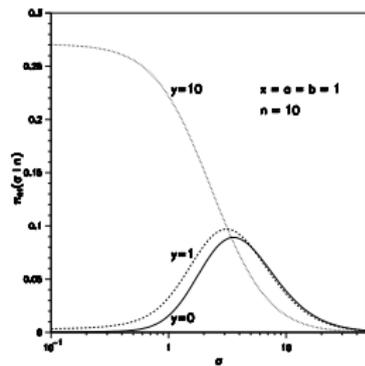
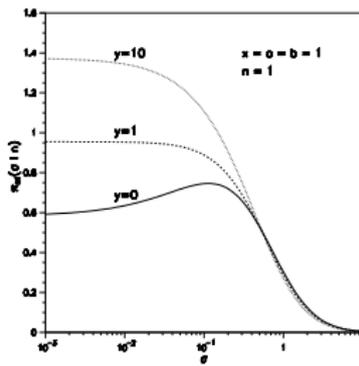
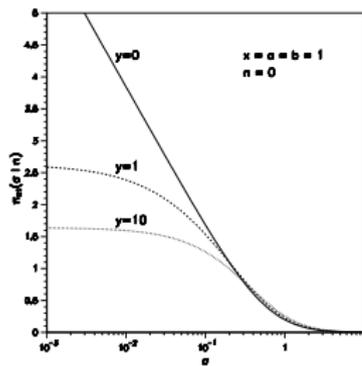
$$\begin{aligned} \pi_{R1}(\sigma | n) &= K' \int_0^\infty d\epsilon \int_0^\infty d\mu \frac{\epsilon(\epsilon\sigma + \mu)^{n-\frac{1}{2}} e^{-\epsilon\sigma - \mu}}{n!} \\ &\quad \frac{a(a\epsilon)^{x-\frac{1}{2}} e^{-a\epsilon}}{\Gamma(x + \frac{1}{2})} \frac{b(b\mu)^{y-\frac{1}{2}} e^{-b\mu}}{\Gamma(y + \frac{1}{2})}, \\ &= K \frac{{}_2F_1(\frac{1}{2} - n, x + \frac{3}{2}; x + y + 2; \frac{1-b\sigma/a}{1+\sigma/a})}{(1 + \sigma/a)^{x+\frac{3}{2}}}, \end{aligned}$$

where

$$K = \binom{x + y + n + \frac{1}{2}}{n - \frac{1}{2}} \frac{x + \frac{1}{2}}{a} \frac{b^{y+\frac{1}{2}}}{(1+b)^{y+n}} \left[ I_{\frac{b}{1+b}} \left( y + \frac{1}{2}, n + \frac{1}{2} \right) \right]^{-1},$$

$$I_z(u, v) \equiv \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} \int_0^z t^{u-1} (1-t)^{v-1} dt.$$

## The Marginal Method-1 Posterior (2/2)



## Method 2 Applied to a Single-Count Measurement (1/2)

The first step in Method 2 is the calculation of the marginal data pdf. We use the same  $(\epsilon, \mu)$  prior as before, i.e. we assume that  $\pi(\epsilon, \mu | \sigma) = \pi(\epsilon, \mu)$ .

Thus:

$$\begin{aligned} p(n | \sigma) &= \iint p(n | \sigma, \epsilon, \mu) \pi(\epsilon, \mu | \sigma) d\epsilon d\mu, \\ &= \iint \frac{(\epsilon\sigma + \mu)^n}{n!} e^{-\epsilon\sigma - \mu} \frac{a(a\epsilon)^{x-1/2}}{\Gamma(x+1/2)} e^{-a\epsilon} \frac{b(b\mu)^{y-1/2}}{\Gamma(y+1/2)} e^{-b\mu} d\epsilon d\mu, \\ &= \left[ \frac{a}{a+\sigma} \right]^{x+\frac{1}{2}} \left[ \frac{b}{b+1} \right]^{y+\frac{1}{2}} \sum_{k=0}^n u_{nk} \left[ \frac{\sigma}{a+\sigma} \right]^k, \end{aligned}$$

where

$$u_{nk} = \binom{x - \frac{1}{2} + k}{k} \binom{y - \frac{1}{2} + n - k}{n - k} \left[ \frac{1}{b+1} \right]^{n-k},$$

and we used generalized binomial coefficients:

$$\binom{v}{w} \equiv \frac{\Gamma(v+1)}{\Gamma(w+1) \Gamma(v-w+1)}.$$

## Method-2 Applied to a Single-Count Measurement (2/2)

We then compute Jeffreys' prior from  $p(n|\sigma)$ :

$$\pi_{R2}(\sigma) \propto \sqrt{\mathbb{E} \left\{ \left[ \frac{d}{d\sigma} \ln p(n|\sigma) \right]^2 \right\}} \propto \sqrt{\sum_{n=0}^{\infty} \frac{[(x+1/2) S_n^0 - (a/\sigma) S_n^1]^2}{(a+\sigma)^{x+5/2} S_n^0}},$$

with

$$S_n^m \equiv \sum_{k=0}^n k^m u_{nk} \left[ \frac{\sigma}{a+\sigma} \right]^k \quad \text{for } m = 0, 1.$$

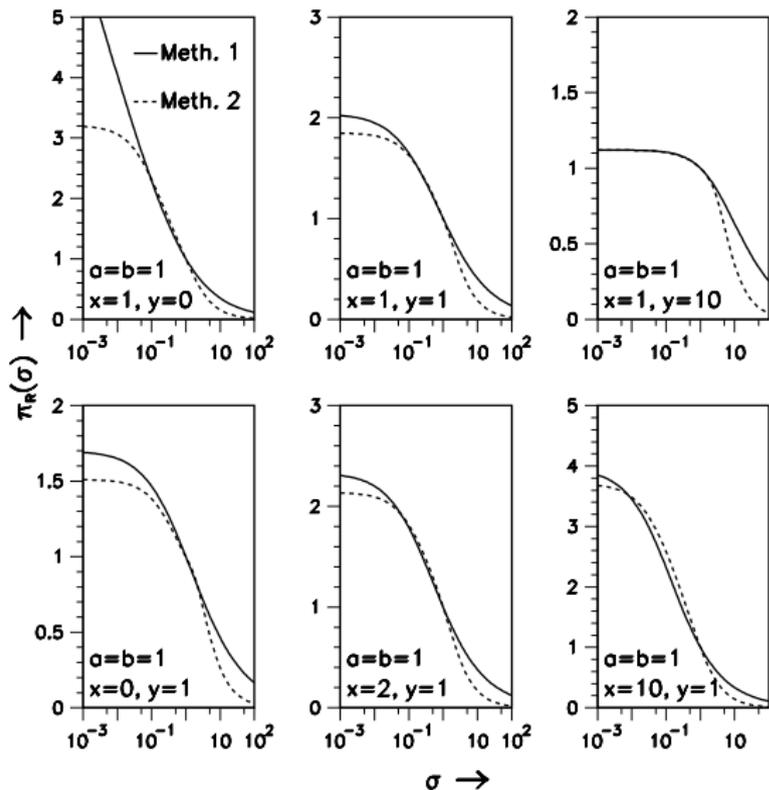
The posterior is simply:

$$\pi_{R2}(\sigma | n) \propto p(n|\sigma) \pi_{R2}(\sigma).$$

Its normalization must be obtained numerically.

Note that Method 2 does not require a compact support argument.

## Comparison of Method-1 and Method-2 Priors



## Generalizations of the Single-Count Model

The model for single-count measurements can be generalized in a number of ways (all of which are important in high energy physics):

- 1 The multi-bin model, based on the likelihood:

$$p(\vec{n} | \sigma, \vec{\epsilon}, \vec{\mu}) = \prod_{i=1}^M \frac{(\mu_i + \epsilon_i \sigma)^{n_i}}{n_i!} e^{-\mu_i - \epsilon_i \sigma}.$$

The Method-1 reference prior for this model is:

$$\pi_J(\sigma | \vec{\epsilon}, \vec{\mu}) = \sqrt{\sum_{i=1}^M \frac{\epsilon_i^2}{\mu_i + \epsilon_i \sigma}}.$$

- 2 Unbinned likelihoods

Together with collaborators H. Prosper and S. Jain we are preparing software to handle these more complicated problems. However we have not found a way to implement the compact support argument in all generality.

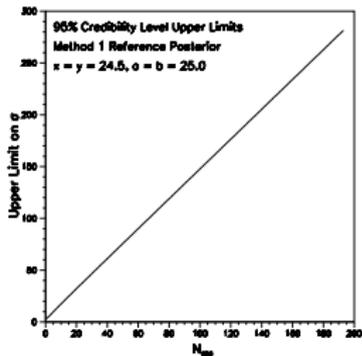
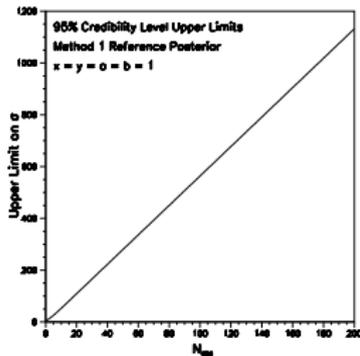
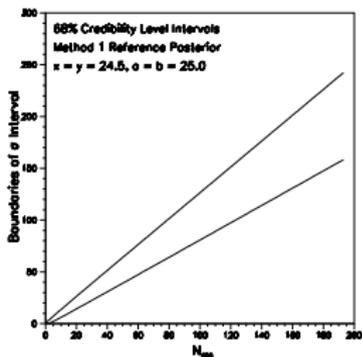
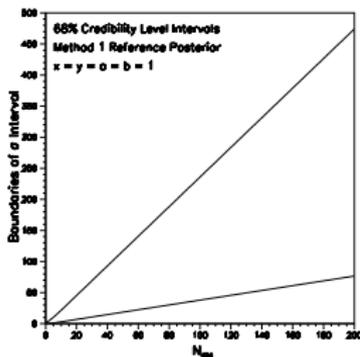
## Upper Limits (1/5)

A common way to summarize posterior distributions is by computing intervals. Here we first look at upper limits and central intervals. We are interested in the following properties:

- 1 How do the interval boundaries vary with  $N_{obs}$ , the observed number of events?
- 2 What is the difference between Method-1 and Method-2 reference posterior intervals?
- 3 How do upper limits vary as a function of the mean of the background prior?
- 4 How do upper limits vary with the uncertainties on background and effective luminosity?

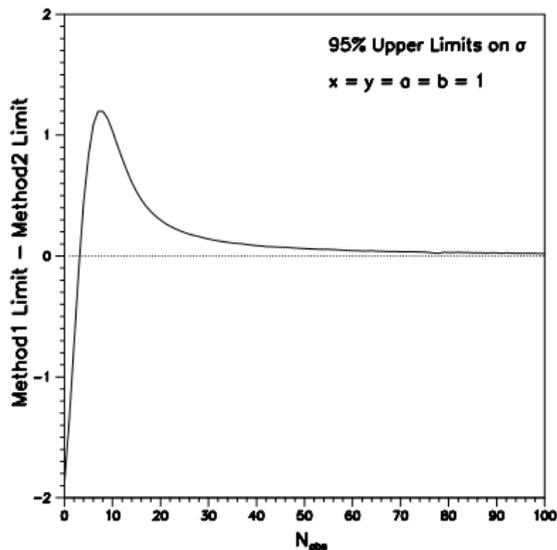
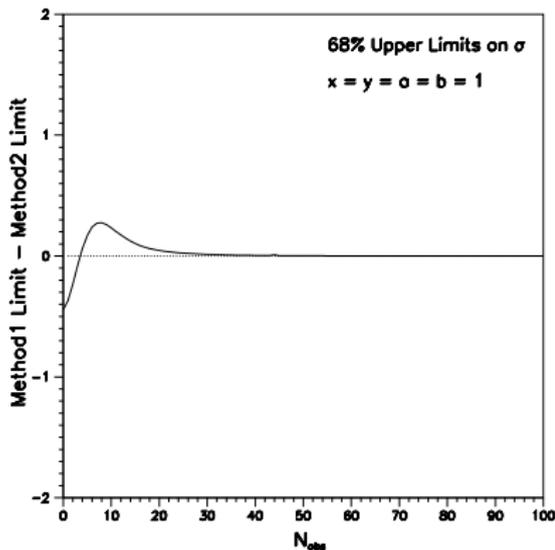
## Upper Limits (2/5)

Dependence on  $N_{obs}$  for  $\delta\epsilon = \delta\mu = 0.82$  (left) and for  $\delta\epsilon = \delta\mu = 0.20$  (right):



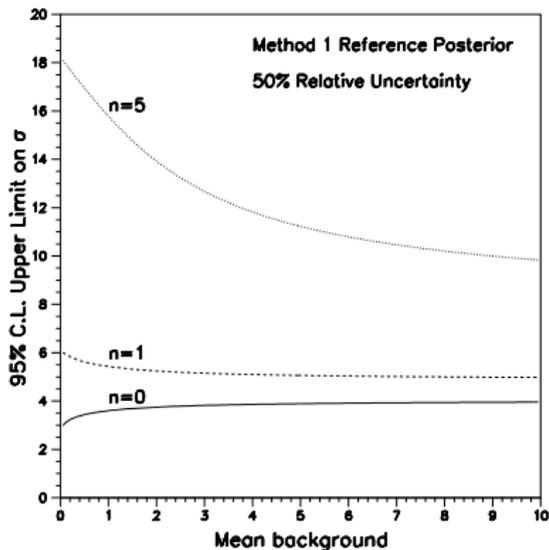
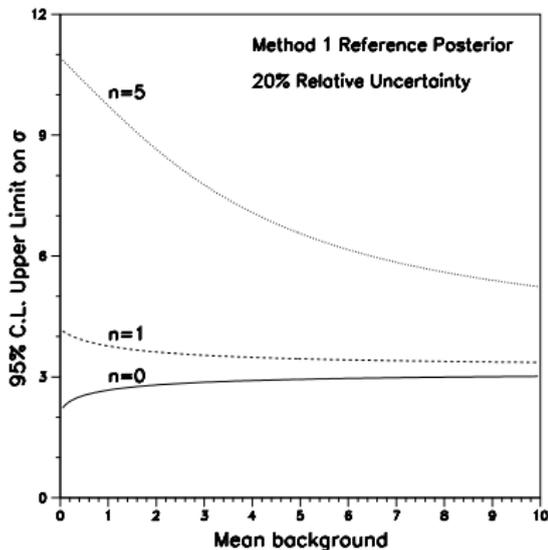
## Upper Limits (3/5)

Difference between Method-1 and Method-2 upper limits at the 68% (left) and 95% (right) credibility levels:



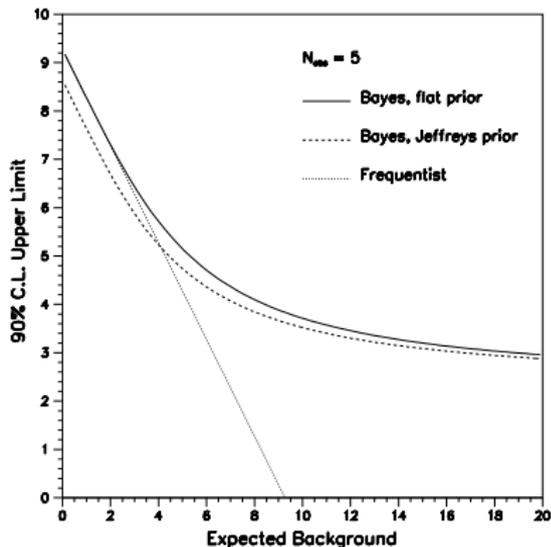
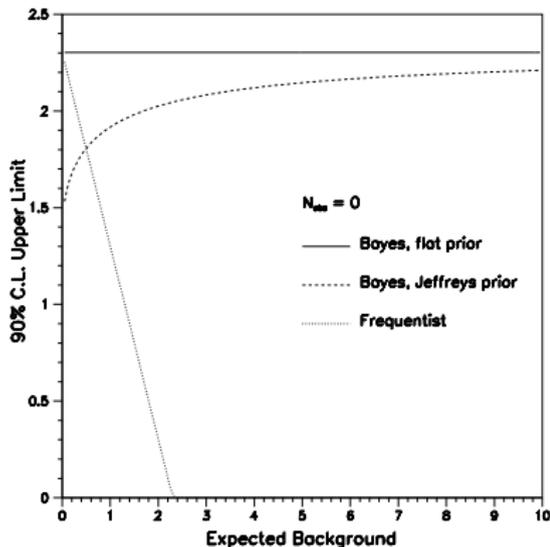
## Upper Limits (4/5)

Variation of 95% C.L. Method-1 upper limit with mean background, for  $\delta\epsilon = \delta\mu = 20\%$  (left) and  $\delta\epsilon = \delta\mu = 50\%$  (right):



## Upper Limits (5/5)

Dependence of various 90% C.L. upper limit constructions on expected background, when there are no uncertainties:



## Behavior under Measurement Replication (1/2)

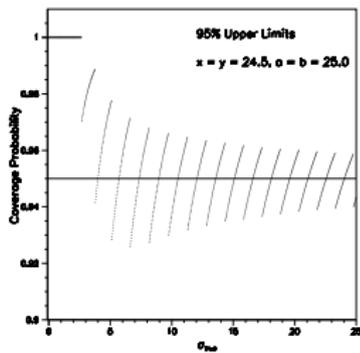
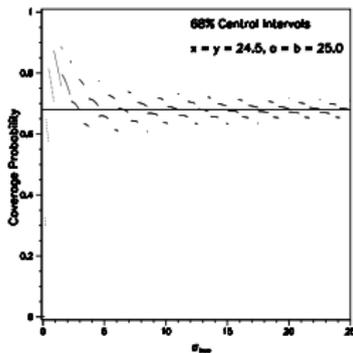
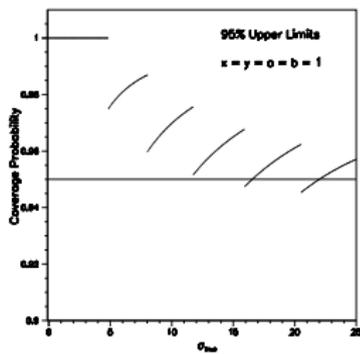
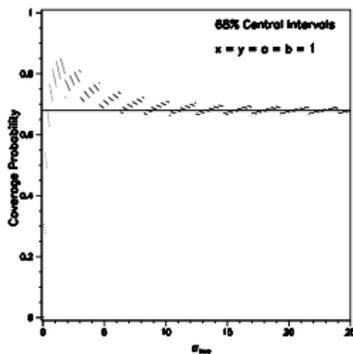
- Bayesian intervals based on a subjective prior satisfy an “average coverage” theorem, according to which the average of the frequentist coverage over the prior equals the credibility of the interval.
- When the prior is objective, there are some technical and conceptual issues with the calculation of a prior-averaged coverage, especially if the prior is improper. In this case, an often useful criterion is the *pointwise* coverage of the intervals.
- When the prior is a mixture of objective and subjective components, a natural approach is to average the coverage over the subjective prior components and check pointwise coverage with respect to the remaining parameters.

In our case, adopting the latter approach means that we will calculate the coverage with respect to the marginalized data pdf:

$$p(n|\sigma) = \iint p(n|\sigma, \epsilon, \mu) \pi(\epsilon, \mu|\sigma) d\epsilon d\mu.$$

## Behavior under Measurement Replication (2/2)

For the Method-1 reference prior (Method-2 plots are very similar):



## Intrinsic Intervals (1/3)

Suppose we have one observation  $x$  from a model  $p(x | \theta)$  and are interested in the true value of  $\theta$ . In a decision-theoretic approach one considers the loss  $\ell(\theta_0, \theta)$  suffered if the value  $\theta_0$  is used as a proxy for the unknown true value of  $\theta$  in some application of our measurement result. The posterior expected loss from using  $\theta_0$  is then:

$$l(\theta_0 | x) = \mathbb{E}_{\theta | x}[\ell(\theta_0, \theta)] = \int d\theta \ell(\theta_0, \theta) \pi(\theta | x).$$

The idea is to use the  $\theta$  value with lowest expected loss as a point estimate of  $\theta$ , and regions of lowest expected loss as interval estimates.

Reference analysis proposes to use intrinsic discrepancy to define the loss function (Bernardo 2005). In its general form, the intrinsic discrepancy loss function is:

$$\delta\{\theta_0, \theta\} = \min \left\{ \int dx p(x | \theta_0) \ln \frac{p(x | \theta_0)}{p(x | \theta)}, (\theta \leftrightarrow \theta_0) \right\}.$$

Compared with other loss functions, intrinsic loss has the advantage of being invariant under one-to-one transformations of the parameter(s), under one-to-one transformations of the data, and under reduction of the data by sufficiency.

## Intrinsic Intervals (2/3)

If nuisance parameters  $\nu$  are present, the definition of  $\delta\{\theta_0, \theta\}$  includes an additional minimization:

$$\delta\{\theta_0, (\theta, \nu)\} = \inf_{\nu_0} \left\{ \min \left\{ \int dx p(x | \theta_0, \nu_0) \ln \frac{p(x | \theta_0, \nu_0)}{p(x | \theta, \nu)}, \left( \begin{array}{l} \theta \leftrightarrow \theta_0 \\ \nu \leftrightarrow \nu_0 \end{array} \right) \right\} \right\}.$$

However, this technique for dealing with nuisance parameters is not appropriate for the single-count measurement because

- 1 the parameter of interest  $\sigma$  is not identifiable via the likelihood (only the combination  $\mu + \epsilon\sigma$  is);
- 2 the nuisance parameters  $\epsilon, \mu$  are constrained by subjective priors.

In the single-count measurement, minimization over  $(\mu_0, \epsilon_0)$  would yield zero loss because of the non-identifiability issue.

## Intrinsic Intervals (3/3)

In principle there are two ways of solving this problem:

### 1 Conditional Method:

- Evaluate the loss  $\ell((\theta, \nu), (\theta_0, \nu))$  from the full model  $p(x | \theta, \nu)$  by assuming that the nuisance parameters  $\nu$  are exactly known.
- Obtain the expected loss by averaging  $\ell((\theta, \nu), (\theta_0, \nu))$  over the posterior  $\pi(\theta, \nu | x)$ .

### 2 Marginal Method:

- Evaluate the loss  $\ell(\theta, \theta_0)$  from the marginal model  $p(x | \theta) = \int d\nu p(x | \theta, \nu) \pi(\nu)$ .
- Obtain the expected loss by integrating the loss function over the marginal posterior  $\pi(\theta | x) = \int d\nu \pi(\theta, \nu | x)$ .

A possible objection against the conditional approach is that we may not want the loss suffered from using the wrong value of  $\theta$  to depend on nuisance parameters that are conceptually independent of  $\theta$ , and that are, furthermore, unknown.

## Intrinsic Intervals for the Single-Count Measurement (1/4)

In the conditional approach we work with the Poisson model:

$$p(n | \sigma, \mu, \epsilon) = \frac{(\mu + \epsilon\sigma)^n e^{-\mu - \epsilon\sigma}}{n!},$$

where  $\epsilon$  and  $\mu$  are initially assumed known. The intrinsic loss is:

$$\delta\{\sigma_0, \sigma | \epsilon, \mu\} = \min\{\kappa\{\sigma_0 | \sigma\}, \kappa\{\sigma | \sigma_0\}\},$$

where:

$$\kappa\{\sigma_0 | \sigma\} = \sum_{n=0}^{\infty} \frac{(\mu + \epsilon\sigma)^n e^{-\mu - \epsilon\sigma}}{n!} \ln \frac{(\mu + \epsilon\sigma)^n e^{-\mu - \epsilon\sigma}}{(\mu + \epsilon\sigma_0)^n e^{-\mu - \epsilon\sigma_0}}.$$

After some simple algebra we find that

$$\delta\{\sigma_0, \sigma | \epsilon, \mu\} = \epsilon |\sigma_0 - \sigma| - [\mu + \epsilon \min(\sigma, \sigma_0)] \left| \ln \frac{\mu + \epsilon\sigma}{\mu + \epsilon\sigma_0} \right|.$$

The posterior expected intrinsic loss is then

$$d(\sigma_0 | n) = \int d\sigma \int d\mu \int d\epsilon \delta\{\sigma_0, \sigma | \mu, \epsilon\} \pi(\sigma, \mu, \epsilon | n).$$

Here we will not pursue this approach further.

## Intrinsic Intervals for the Single-Count Measurement (2/4)

In the marginal approach we use the marginal pdf of the data:

$$\begin{aligned} p(n|\sigma) &= \iint p(n|\sigma, \epsilon, \mu) \pi(\epsilon, \mu|\sigma) d\epsilon d\mu, \\ &= \left[ \frac{a}{a+\sigma} \right]^{x+\frac{1}{2}} \left[ \frac{b}{b+1} \right]^{y+\frac{1}{2}} \sum_{k=0}^n u_{nk} \left[ \frac{\sigma}{a+\sigma} \right]^k, \end{aligned}$$

The intrinsic loss is

$$\delta\{\sigma_0, \sigma\} = \min \left\{ \sum_{n=0}^{\infty} p(n|\sigma_0) \ln \frac{p(n|\sigma_0)}{p(n|\sigma)}, \sum_{n=0}^{\infty} p(n|\sigma) \ln \frac{p(n|\sigma)}{p(n|\sigma_0)} \right\},$$

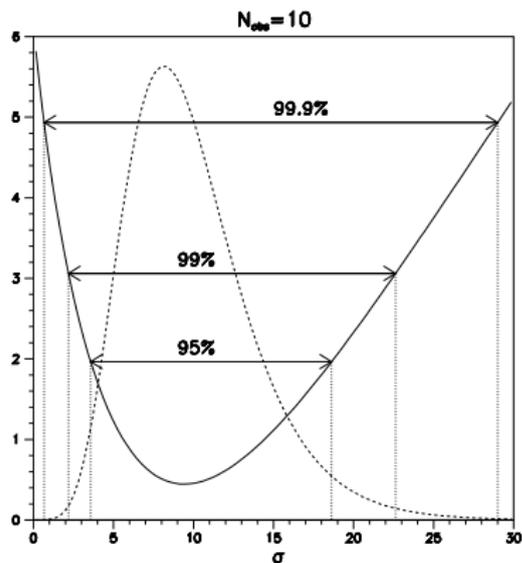
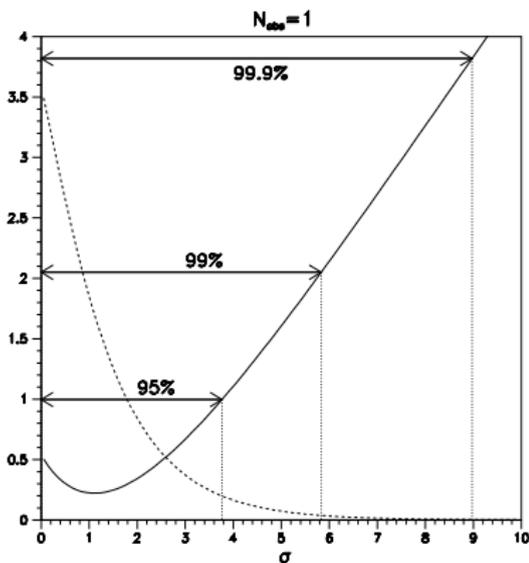
and the posterior expected intrinsic loss is

$$d(\sigma_0|n) = \int d\sigma \delta\{\sigma_0, \sigma\} \pi(\sigma|n),$$

where  $\pi(\sigma|n)$  is the marginal posterior, obtained via either Method 1 or Method 2. Neither of the above two expressions can be further simplified, but they can be coded numerically.

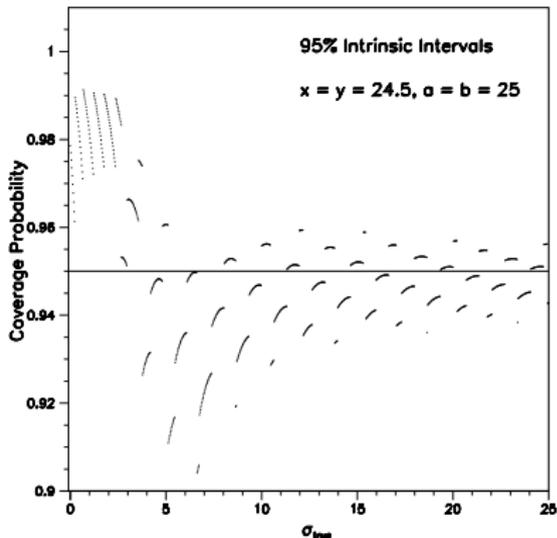
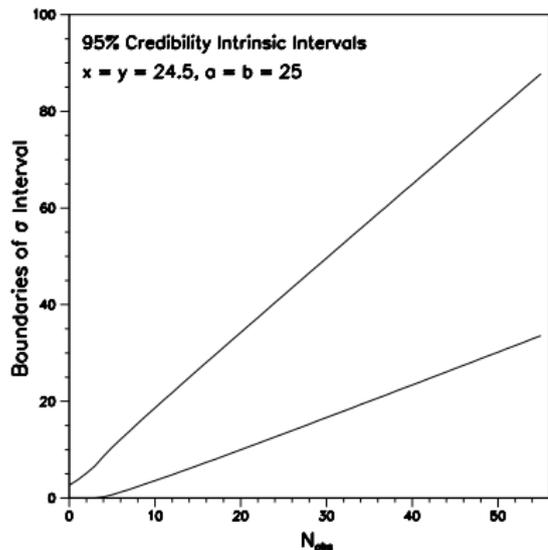
## Intrinsic Intervals for the Single-Count Measurement (3/4)

Construction of intrinsic intervals for 1 (left) and 10 (right) observed events, using the Method-1 reference posterior (dashed curves) with 20% uncertainties on prior background and effective luminosity. The solid curves show the posterior expected intrinsic loss.



## Intrinsic Intervals for the Single-Count Measurement (4/4)

Left: boundaries of minimum posterior expected intrinsic loss intervals with 95% credibility as a function of the observed number of events. Right: coverage probability of these intervals as a function of the true value of the cross section  $\sigma$ .



## Reference Analysis Search Procedures (1/4)

Here is an objective Bayesian version of the search procedure described at the beginning of this talk:

- 1 Choose three credibility levels  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ .
- 2 Compute the marginal reference posterior for the parameter of interest, say  $\sigma$ .
- 3 Compute an  $\alpha_1$ -credibility intrinsic interval  $S$  for  $\sigma$ .
- 4 If interval  $S$  does not include the value  $\sigma = 0$ , claim discovery and compute an  $\alpha_2$ -credibility intrinsic interval on  $\sigma$ .
- 5 If interval  $S$  includes  $\sigma = 0$ , make no claim and compute an  $\alpha_3$ -credibility upper limit on  $\sigma$ .

This Bayesian procedure avoids all the problems of the frequentist version: effect of test on subsequent interval construction, empty intervals, sensitivity to expected background, and handling of nuisance parameters.

## Reference Analysis Search Procedures (2/4)

In addition to working with fixed discovery thresholds, physicists like to quote “significances” ( $p$  values in classical parlance). Reference analysis provides the so-called **Bayes Reference Criterion (BRC)**, which is the reference posterior expected intrinsic loss evaluated at the null hypothesis (Bernardo 1999, Bernardo & Rueda 2002):

$$\text{BRC} \equiv d(0|n) = \int_0^\infty d\sigma \pi_{R1}(\sigma|n) \min\{\kappa\{\sigma|0\}, \kappa\{0|\sigma\}\},$$

where

$$\kappa\{\sigma_1|\sigma_2\} \equiv \sum_{n=0}^{\infty} p(n|\sigma_2) \ln \frac{p(n|\sigma_2)}{p(n|\sigma_1)},$$

where  $p(n|\sigma)$  is, as before, the marginal pdf for the data. The BRC can be interpreted as the minimum posterior expected log-likelihood ratio in favor of the model that generated the data, and it therefore provides its own calibration of the evidence against the null hypothesis:  $\text{BRC}=\ln(10) \approx 2.3$  is mild evidence,  $\text{BRC}=\ln(100) \approx 4.6$  is strong, and  $\text{BRC}=\ln(1000) \approx 6.9$  is decisive. In contrast with  $p$  values, the BRC does not need adjusting for the dimensionality of the problem and the size of the sample.

## Reference Analysis Search Procedures (3/4)

There are other “candidate” Bayesian measures of significance:

- 1 Remove the symmetrization in BRC; this gives:

$$d_1 \equiv \int_0^{\infty} d\sigma \pi_{R1}(\sigma | n) \kappa\{0 | \sigma\} \quad \text{or} \quad d_2 \equiv \int_0^{\infty} d\sigma \pi_{R1}(\sigma | n) \kappa\{\sigma | 0\}.$$

- 2 Find the credibility  $\gamma$  of the widest HPD interval that does not contain  $\sigma = 0$ , and define significance as  $p_{HPD} = 1 - \gamma$ .
- 3 Find the credibility  $\gamma$  of the widest intrinsic interval that does not contain the value  $\sigma = 0$ , and define significance as  $p_{BRC} = 1 - \gamma$ ; this is essentially the posterior tail probability of BRC.
- 4 Not fully Bayesian, but also interesting, is the prior-predictive  $p$  value

$$p_{PP} = \int_0^{\infty} d\mu \sum_{n=n_{obs}}^{\infty} p(n | \sigma = 0, \mu) \pi(\mu).$$

For illustration purposes it is convenient to convert probabilities such as  $p_{HPD}$ ,  $p_{BRC}$ , and  $p_{PP}$  into  $Z$ -values, according to:

$$p = \int_Z^{+\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \quad \text{i.e.} \quad Z(p) = \sqrt{2} \operatorname{erf}^{-1}(1 - 2p).$$

## Reference Analysis Search Procedures (4/4)

Examples of significance calculations for the single-count model. (To help identify trends, line 2 is repeated at lines 5, 8, and 12.)

	Background	Eff. Lum.	$N_{obs}$	$d_1$	$d_2$	BRC	$Z_{BRC}$	$Z_{HPD}$	$Z_{PP}$
1	$1.0 \pm 0.0$	$1.0 \pm 0.0$	10	15.68	7.20	7.20	3.76	4.92	5.18
2	$1.0 \pm 0.2$	$1.0 \pm 0.2$	10	14.93	6.12	6.12	3.46	4.50	4.94
3	$1.0 \pm 0.2$	$1.0 \pm 0.4$	10	18.56	4.47	4.47	3.03	4.25	4.94
4	$1.0 \pm 0.2$	$1.0 \pm 0.6$	10	28.54	3.30	3.30	2.86	4.18	4.94
5	$1.0 \pm 0.2$	$1.0 \pm 0.2$	10	14.93	6.12	6.12	3.46	4.50	4.94
6	$1.0 \pm 0.2$	$2.0 \pm 0.8$	10	18.56	4.47	4.47	2.96	4.16	4.94
7	$1.0 \pm 0.2$	$3.0 \pm 1.8$	10	28.54	3.30	3.30	2.63	3.95	4.94
8	$1.0 \pm 0.2$	$1.0 \pm 0.2$	10	14.93	6.12	6.12	3.46	4.50	4.94
9	$1.0 \pm 0.4$	$1.0 \pm 0.2$	10	11.62	6.08	6.08	3.45	4.04	4.45
10	$1.0 \pm 0.6$	$1.0 \pm 0.2$	10	8.85	6.04	6.04	3.44	3.56	4.03
11	$1.0 \pm 0.8$	$1.0 \pm 0.2$	10	6.87	6.01	6.01	3.42	3.17	3.70
12	$1.0 \pm 0.2$	$1.0 \pm 0.2$	10	14.93	6.12	6.12	3.46	4.50	4.94
13	$2.0 \pm 0.4$	$1.0 \pm 0.2$	10	8.66	4.55	4.55	2.96	3.26	3.66
14	$3.0 \pm 0.6$	$1.0 \pm 0.2$	10	5.52	3.39	3.39	2.53	2.45	2.83
15	$4.0 \pm 0.8$	$1.0 \pm 0.2$	10	3.65	2.52	2.52	2.15	1.84	2.20

## Open Questions

- 1 The compact support argument in the calculation of reference priors is not trivial. One would welcome a more deterministic approach to the construction of the required compact sets. This would be especially useful for coding purposes.
- 2 Partially known nuisance parameters are common in high energy physics, and we understand how to incorporate them in the calculation of reference priors. It is less clear what approach to take in the calculation of the posterior expected intrinsic loss: conditional, marginal, other? This affects both the construction of intrinsic intervals and tests based on the BRC.
- 3 How should we calibrate BRC? Can we just take a “standard” example and compare with a  $p$  value test? However, BRC depends on the signal model, whereas  $p$  values do not! In addition, physicists are worried about the “look-elsewhere” effect, this is what motivates their  $5\sigma$  discovery threshold. How can we address this concern with BRC? Finally, our numerical examples illustrate that BRC behaves correctly with respect to uncertainties in the background and signal model. Can we believe this in all generality?