Reference priors for high energy physics

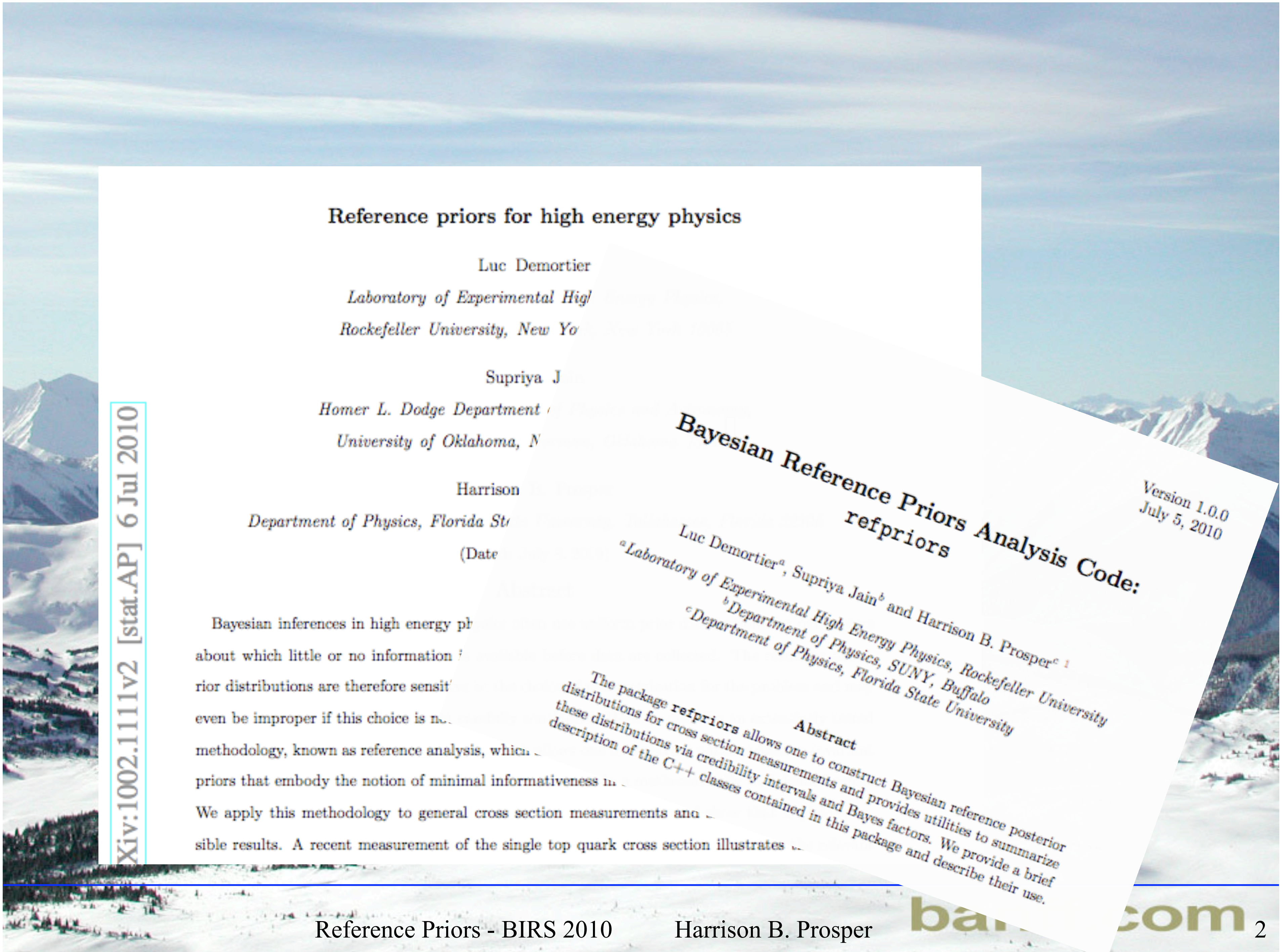
Luc Demortier[1], Supriya Jain[2] , and Harrison B. Prosper[3]

Statistical issues relevant to significance of discovery claims

Banff International Research Center

15 July, 2010

1. Rockefeller University
2. State University of New York at Buffalo
3. Florida State University



Reference priors for high energy physics

Luc Demortier

Laboratory of Experimental Hig

Rockefeller University, New Yo

Supriya J

Homer L. Dodge Department

University of Oklahoma, N

Harrison

Department of Physics, Florida St

(Date

Bayesian inferences in high energy ph
about which little or no information
rior distributions are therefore sensit
even be improper if this choice is n
methodology, known as reference analysis, which
priors that embody the notion of minimal informativeness in
We apply this methodology to general cross section measurements and
sible results. A recent measurement of the single top quark cross section illustrates

arXiv:1002.1111v2 [stat.AP] 6 Jul 2010

Bayesian Reference Priors Analysis Code: refpriors

Version 1.0.0
July 5, 2010

Luc Demortier[a], Supriya Jain[b] and Harrison B. Prosper[c 1]

[a] *Laboratory of Experimental High Energy Physics, Rockefeller University*

[b] *Department of Physics, SUNY, Buffalo*

[c] *Department of Physics, Florida State University*

Abstract

The package **refpriors** allows one to construct Bayesian reference posterior distributions for cross section measurements and provides utilities to summarize these distributions via credibility intervals and Bayes factors. We provide a brief description of the C++ classes contained in this package and describe their use.

Outline

- Introduction
- Single-Count Model
- Multi-Count Model
- Studies
- refpriors
- Conclusions & Summary

Introduction (1)

Some Terminology

evidence-based priors

proper priors that incorporate pertinent information

formal priors

priors derived using formal rules

statistical model

1. probability distribution of data
2. the sampling space
3. the stopping rule

Introduction (2)

The Platonic Bayesian

uses evidence-based priors for *every* parameter for *every* problem…and is happy not to have a life.

The Non-Platonic Bayesian

acknowledging the impossibility of eliciting *every* detail of *every* prior for *every* problem chooses either:

1. to "Abandon all hope, ye who enter here."
2. or uses formal priors as needed and checks that results are reasonable.

Introduction (3)

Formal Priors

High energy physicists (who are pragmatic to a fault) use the simple formal rule:

"when in doubt, make it flat!"

The goal of our paper is to initiate a distancing of the field from this rule to one that has a *foundation*, namely, the rule by Bernardo (1979) and Berger and Bernardo (1989, 1992, 2009), which yields so-called *reference priors*. They are the basis of *reference analysis*, an integrated set of Bayesian methods that include methods for parameter and interval estimation and hypothesis testing.

Introduction (4)

Reference Priors

Their properties ought to appeal to physicists:

1. *generality* – there is a well-defined algorithm for creating a reference prior for almost any problem
2. *invariance* – reference priors are invariant under one-to-one transformations, $\pi(\phi \mid x) = \pi(\theta \mid x) \mid \partial\theta / \partial\phi \mid$
3. *consistency* – the posterior densities over the sampling space cluster around the true values of the parameters
4. *coherence* – avoidance of marginalization paradoxes: posterior densities that can be arrived at in two ways that ought to agree but do not.

Introduction (5)

Reference Priors

Basic intuition: the information provided by observations is contingent on what is already known.

Therefore, since the posterior $\pi(\theta \mid x)$ encodes what one knows about the parameter θ after the observation x, while the prior $\pi(\theta)$ encodes what one knows before, one expects that the greater the discrepancy between the posterior and the prior, the greater the information gain.

This is made precise using ideas from information theory.

Introduction (6)

Expected Information (I)

The information expected from one observation x is quantified using the *expected intrinsic information*

$$I\{\pi\} = \int D[\pi, \pi_x] m(x) dx, \quad \pi \equiv \text{ prior}, \quad \pi_x \equiv \text{posterior}$$

where

$$m(x) \equiv \int p(x \mid \theta) \pi(\theta) d\theta$$

is the *marginal distribution* (or integrated likelihood) and

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

is the *Kullback-Leibler divergence* between the densities $p(\theta)$ and $q(\theta)$.

Introduction (7)

Expected Information (II)

Suppose we have k observations, $x_{(k)} = (x_1, x_2, \ldots, x_k)$. The expected intrinsic information generalizes in a natural way:

$$I_k\{\pi\} = \int D[\pi, \pi_{x(k)}]\, m(x_{(k)})\, dx_{(k)}$$

with

$$m(x_{(k)}) \equiv \int p(x_{(k)} \mid \theta)\pi(\theta)d\theta = \int \left[\prod_{i=1}^{k} p(x_i \mid \theta)\right]\pi(\theta)d\theta$$

As k grows to infinity, we expect to recover all the information that can possibly be had about the parameter θ.

In principle, $I_\infty\{\pi\}$ (the *missing information*) is the quantity to be maximized with respect to $\pi(\theta)$ to obtain the reference prior $\pi_R(\theta)$.

Introduction (8)

Reference Priors for One-Parameter Models (I)

$I_\infty\{\pi\}$ typically diverges. Therefore, its maximization is defined by a limit: one maximizes $I_k\{\pi\}$ to find π_k. (NB: since π_k must integrate to 1, it may be necessary to restrict its domain to a compact set $\Theta_k \subset \Theta_{k+1}$.)

This procedure yields the following constructive definition of the reference prior:

$$\pi_R(\theta) = \lim_{k\to\infty} \pi_k(\theta) / \pi_k(\theta_0)$$

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

where the fixed point θ_0 and $h(\theta)$ may be freely chosen.

Introduction (9)

Reference Priors for One-Parameter Models (II)

If the posterior $\pi(\theta \mid x)$ is asymptotically normal, a very useful result obtains, namely, that the reference prior for models with one continuous parameter reduces to the well-known Jeffreys prior (the square-root of the Fisher information F):

$$\pi_R(\theta) = \sqrt{F}$$

$$= \sqrt{\int p(x \mid \theta) \left[-\frac{d^2}{d\theta^2} \ln p(x \mid \theta) \right] dx}$$

Introduction (10)

Nuisance Parameters

When nuisance parameters are present, there are two plausible ways to proceed depending on what prior information is available (Sun, Berger):

1. Method 1: Assume that

   $$\pi(\theta, \phi) = \pi_R(\theta \mid \phi)\, \pi(\phi),$$ where $\pi_R(\theta \mid \phi)$ is computed from the model $p(x \mid \theta, \phi)$ for *fixed* ϕ.

2. Method 2: Assume that

   $$\pi(\theta, \phi) = \pi(\phi \mid \theta)\, \pi_R(\theta),$$ where $\pi_R(\theta)$ is computed from the marginalized model

   $$p(x \mid \theta) = \int p(x \mid \theta, \phi)\, \pi(\phi \mid \theta)\, d\phi$$

Count Models – Some Exact Results

The Single-Count Model (1)

Let

n be the observed count (the number of events)

σ the cross section (the signal strength)

ε the effective integrated luminosity and

μ the mean background count.

The likelihood for the model is

$$p(n \mid \sigma, \varepsilon, \mu) = \text{Poisson}(n \mid \varepsilon\sigma + \mu), \quad 0 \le \sigma < \infty \quad \text{and} \quad 0 < \varepsilon, \mu < \infty$$

ε and μ are nuisance parameters.

The Single-Count Model (2)

The Evidence-based Prior

We assume that ε and μ are independent *a priori* and both are independent of the cross section σ.

We further assume that the priors for ε and μ can be modeled with gamma densities:

$$\pi(\varepsilon, \mu \mid \sigma) = \pi(\varepsilon, \mu) = \text{Gamma}(a\varepsilon \mid x + 1/2) \times \text{Gamma}(b\mu \mid y + 1/2)$$

where a, b, x, and y are known constants.

The Single-Count Model (3)

Reference Prior – Method 1 (I)

In this method, we find the conditional reference prior $\pi_R(\sigma \mid \varepsilon, \mu)$ using

$$\pi_R(\sigma \mid \varepsilon,\mu) = \lim_{k\to\infty} \pi_k(\sigma \mid \varepsilon,\mu) / \pi_k(\sigma_0 \mid \varepsilon,\mu)$$

where σ_0 is any fixed point.

Since the single-count model is asymptotically normal, we can use Jeffreys' rule to compute

$$\pi_k(\sigma \mid \varepsilon,\mu) \propto \frac{\varepsilon}{\sqrt{\varepsilon\sigma + \mu}}$$

The Single-Count Model (4)

Reference Prior – Method 1 (II)

As a function of σ, the function

$$\pi_k(\sigma \mid \varepsilon, \mu) \propto \frac{\varepsilon}{\sqrt{\varepsilon \sigma + \mu}}$$

does not integrate to 1, so we must restrict its domain to a compact set. Let's try

$$\Theta_k = \{(\sigma, \varepsilon, \mu) : \sigma \in [0, u_k], \varepsilon \in [0, v_k], \mu \in [0, w_k]\}$$

where $u_k < u_{k+1}$, $v_k < v_{k+1}$, $w_k < w_{k+1}$. We obtain:

$$\pi_R(\sigma \mid \varepsilon, \mu) \propto \frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon \sigma + \mu}}$$

The Single-Count Model (5)

Reference Prior – Method 1 (III)

This prior was found to yield an *improper* posterior when the evidence-based prior for ε is ~ exp(-ε) $\varepsilon^{-1/2}$. This problem can be traced to our choice of compact sets (Berger).

Noting that ε and σ enter as a product in the model, a better choice is

$$\Theta_k = \left\{(\sigma,\varepsilon,\mu) : \sigma \in [0, u_k / \varepsilon], \varepsilon \in [1/v_k, v_k], \mu \in [0, w_k]\right\}$$

This yields a result *identical* to the Jeffreys prior

$$\pi_{R1}(\sigma \mid \varepsilon, \mu) \propto \frac{\varepsilon}{\sqrt{\varepsilon\sigma + \mu}}$$

and produces well-behaved posterior densities.

The Single-Count Model (6)

Reference Prior – Method 2 (I)

In this method, the reference prior $\pi_{R2}(\sigma)$ is computed from the marginalized model

$$p(n \mid \sigma) = \iint p(n \mid \sigma, \varepsilon, \mu) \pi(\varepsilon, \mu \mid \sigma) \, d\varepsilon \, d\mu$$

$$= \left[\frac{a}{a+\sigma} \right]^{x+1/2} \left[\frac{b}{b+1} \right]^{y+1/2} S_n^0(\sigma)$$

where $$S_n^m(\sigma) = \sum_{k=0}^{n} \left[\frac{a}{a+\sigma} \right]^k C_{mk}$$

C_{mk} are constants

We find

$$\pi_{R2}(\sigma) \propto \sqrt{\sum_{k=0}^{\infty} \frac{\left[(x+1/2) S_k^0(\sigma) - a S_k^1(\sigma)/\sigma \right]^2}{(a+\sigma)^{x+5/2} S_k^0(\sigma)}}$$

The Multiple-Count Model

Reference Prior – Method 1

An important generalization is to the multi-count model comprising M counts:

$$p(\vec{n} \mid \sigma, \vec{\varepsilon}, \vec{\mu}) = \prod_{i=1}^{M} \text{Poisson}(n_i \mid \varepsilon_i \sigma + \mu_i)$$

The reference prior is again identical to the Jeffreys prior and is given by

$$\pi_{R1}(\sigma) \propto \sqrt{\sum_{i=1}^{M} \frac{\varepsilon_i^2}{\varepsilon_i \sigma + \mu_i}}$$

(For Method 2, the marginalized model $p(\vec{n} \mid \sigma)$ can be computed exactly, but we compute $\pi_{R2}(\sigma)$ numerically.)

Numerical Algorithms

Numerical Algorithms

Multi-count Model – Method 1

The reference prior is computed using the algorithm:

$$
\begin{array}{l}
\vec{n}_0 = \text{ array of observed counts} \\
\text{for } i = 1,\ldots,I: \\
\quad (\sigma, \vec{\varepsilon}_i, \vec{\mu}_i) \sim p(\vec{n}_0 \mid \sigma, \vec{\varepsilon}, \vec{\mu})\, \pi(\vec{\varepsilon}, \vec{\mu}) \\
\quad \text{for } j = 1,\ldots,J: \\
\quad\quad \vec{n}_j \sim p(\vec{n} \mid \sigma, \vec{\varepsilon}_i, \vec{\mu}_i) \\
\quad\quad d_j = d^2[-\ln p(\vec{n}_j \mid \sigma_i, \vec{\varepsilon}_i, \vec{\mu}_i)] / d\sigma_i^2 \text{ numerically} \\
\quad \pi_{R1}(\sigma_i \mid \vec{\varepsilon}_i, \vec{\mu}_i) = \sqrt{\frac{1}{J}\sum_{j=1}^{J} d_j} \\
\pi_{R1}(\sigma_i \mid \vec{n}_0) = \text{histogram(fill } \sigma_i, \text{weight} = \pi(\sigma_i \mid \vec{\varepsilon}_i, \vec{\mu}_i))
\end{array}
$$

Numerical Algorithms

Multi-count Model – Method 2

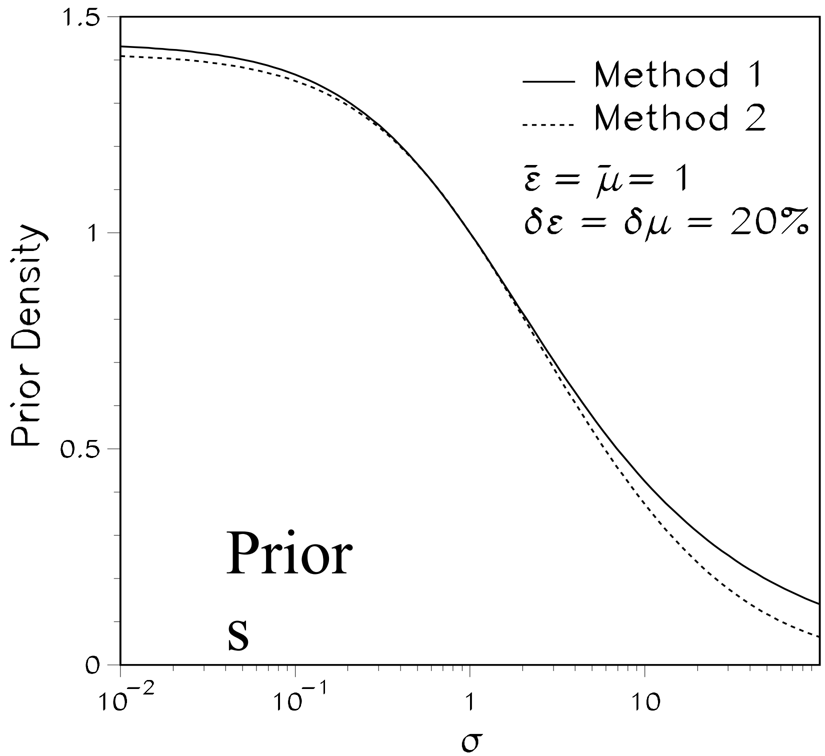
The reference prior is computed using the algorithm:

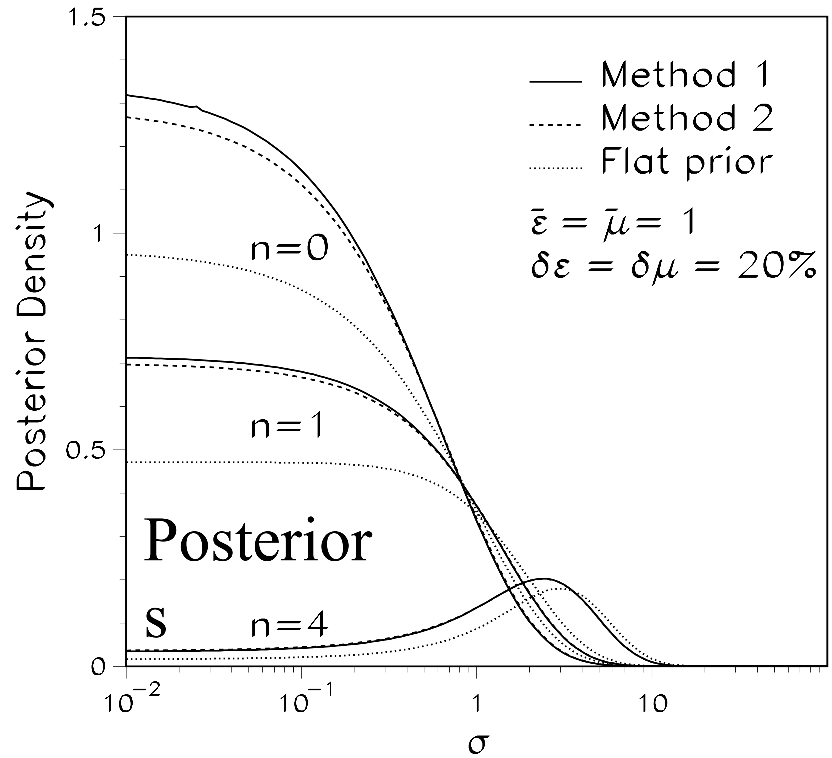
$\vec{n}_0 =$ array of observed counts
for $i = 1,\ldots,I$:
    specify σ_i
    for $j = 1,\ldots,J$:
        $\vec{n}_j \sim p(\vec{n} \mid \sigma_i)$
        $d_j = d^2[-\ln p(\vec{n}_j \mid \sigma_i)] / d\sigma_i^2$ numerically
    $\pi_{R2}(\sigma_i) = \sqrt{\frac{1}{J}\sum_{j=1}^{J} d_j}$
$\pi_{R2}(\sigma_i \mid \vec{n}_0) = \text{integral}(\{p(\vec{n}_0 \mid \sigma_i)\}, \{\pi_{R2}(\sigma_i)\})$

Studies

Studies (1)

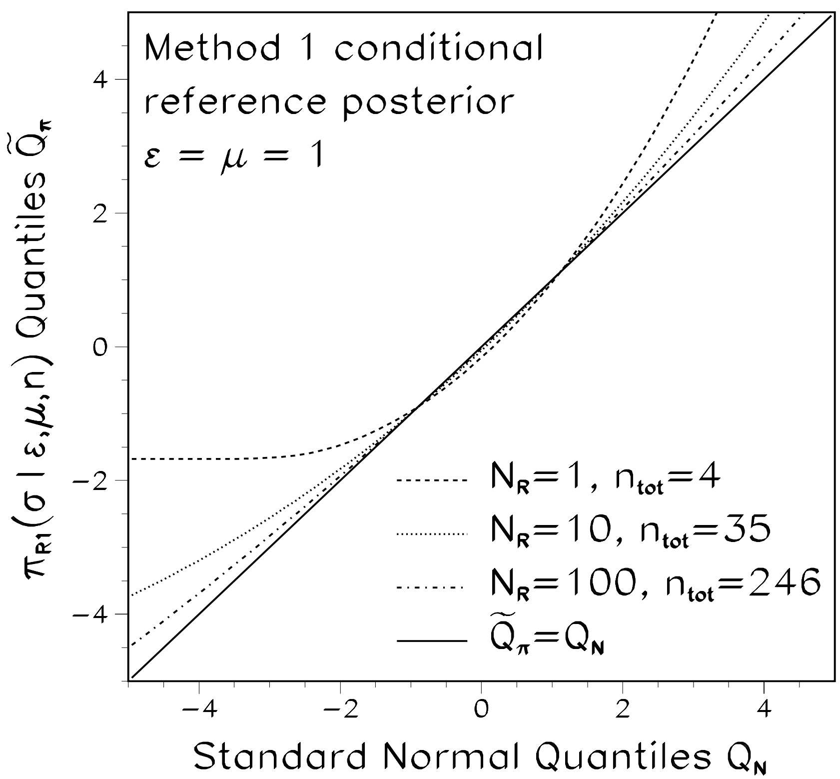
Prior and posterior densities

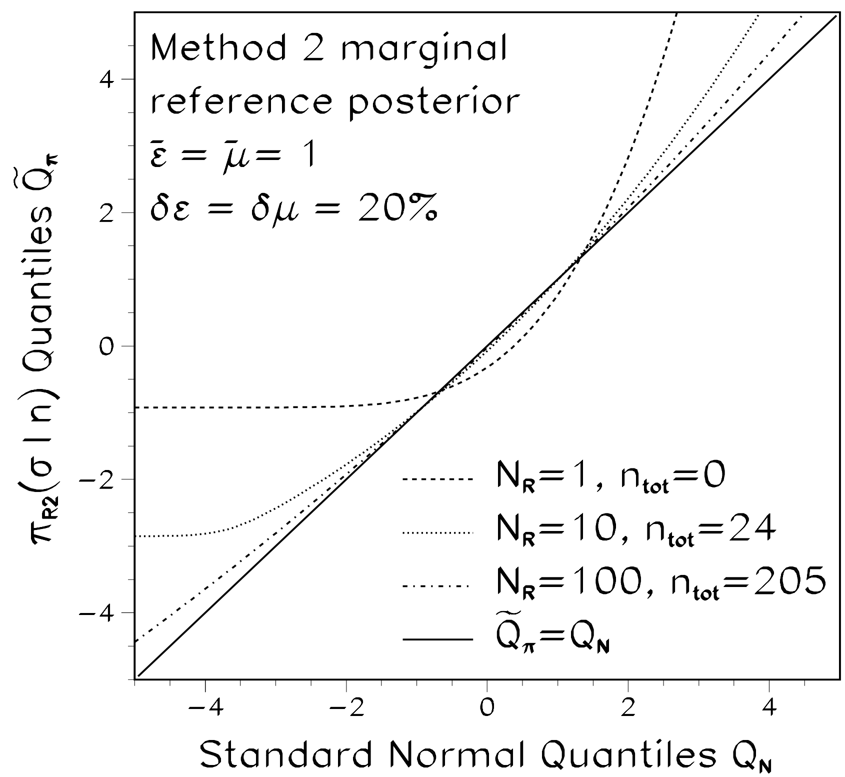




Studies (2)

Asymptotic behavior of posterior



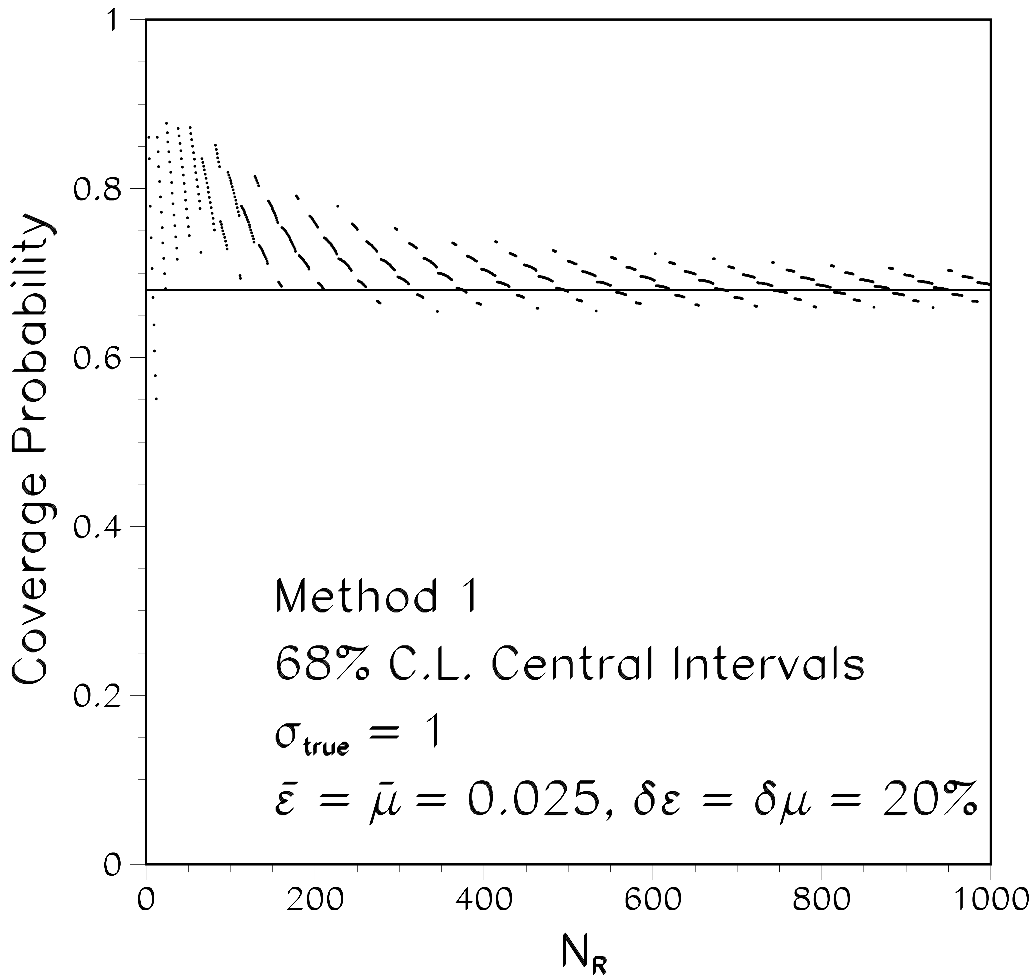


N_R – number of replications, n_{tot} – summed counts

Studies (3)

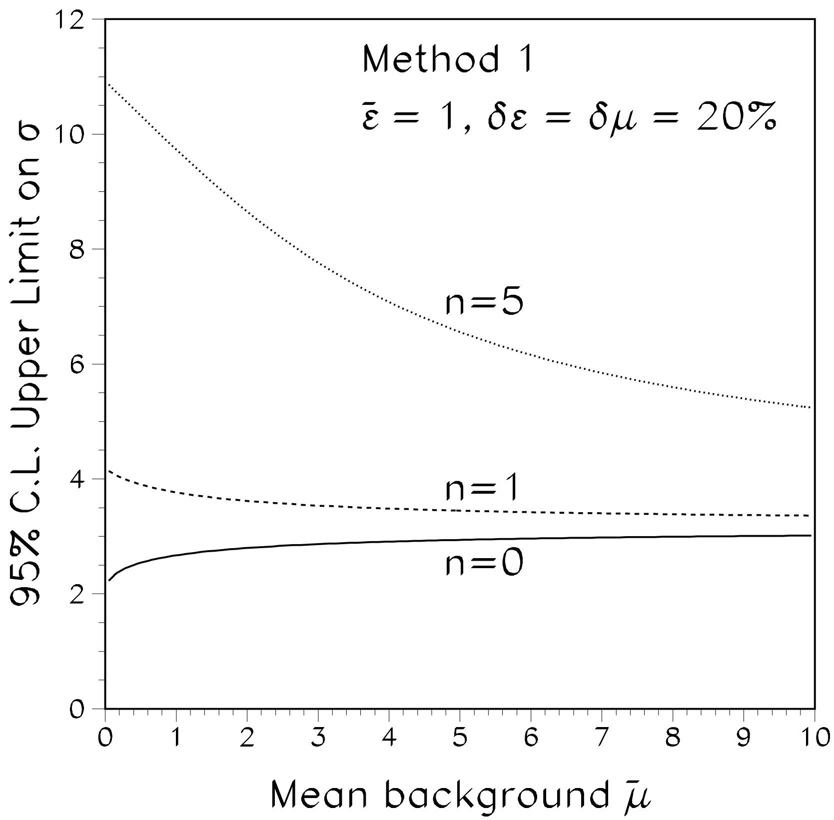
Sampling Consistency

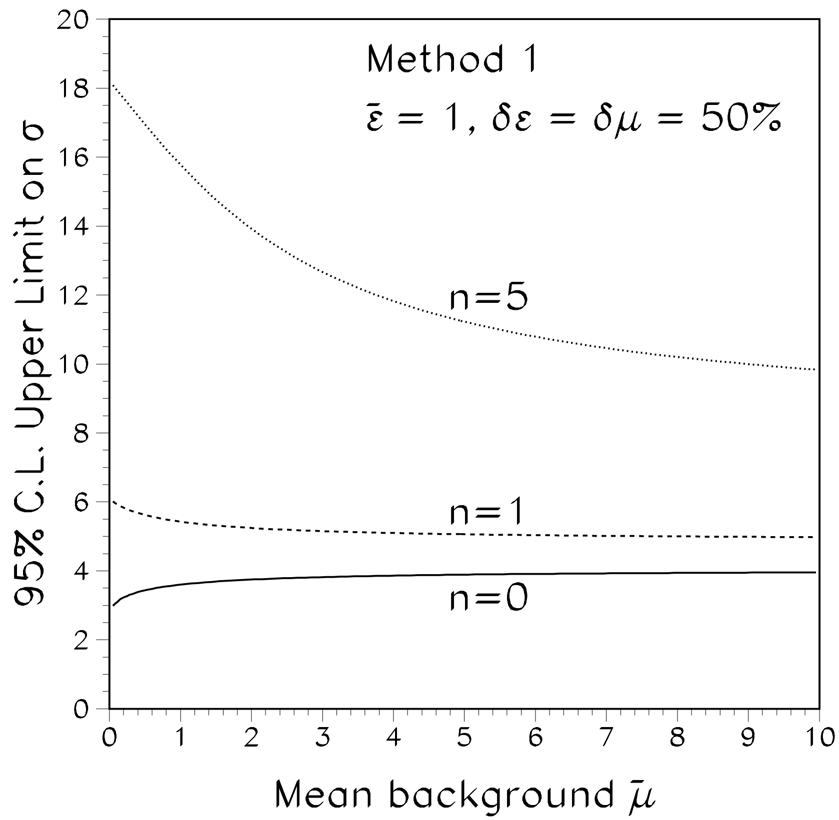
Coverage for a fixed value of σ, but averaged over the nuisance parameters, as a function of the number of replications, N_R, of the experiment.



Studies (4)

Dependence of upper limit on count and relative uncertainty on the background.

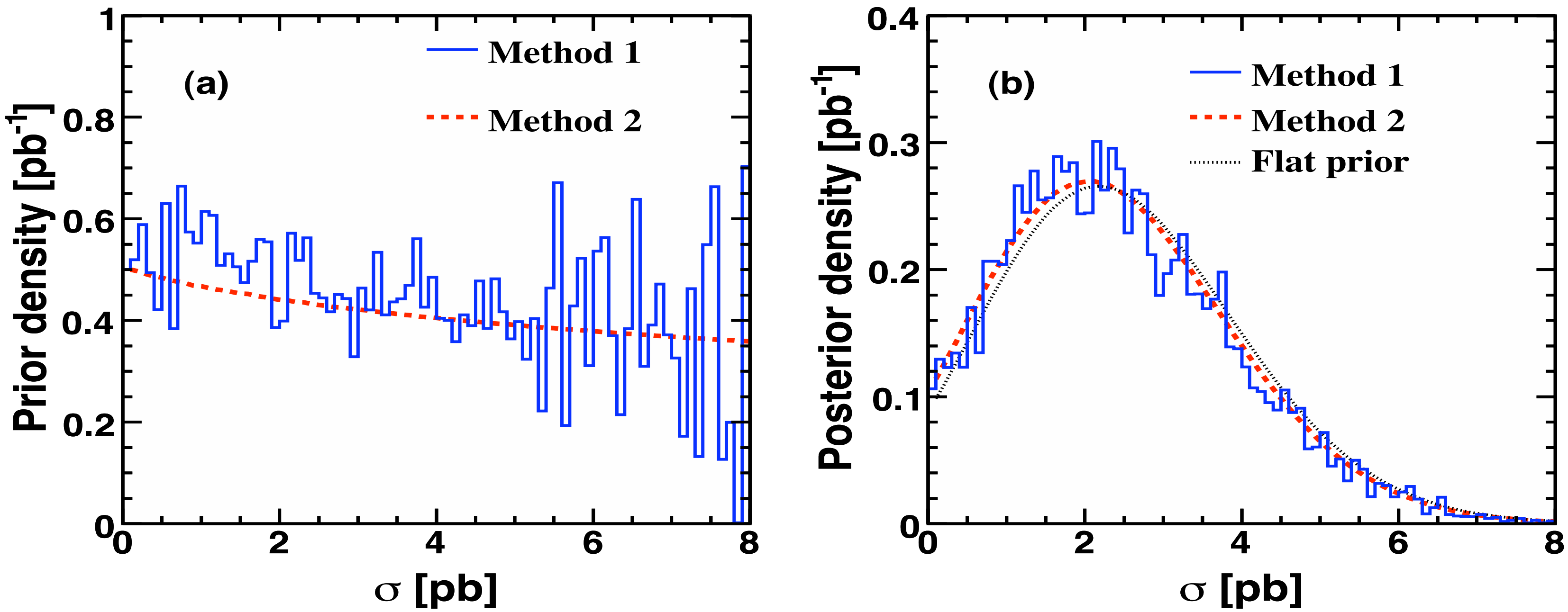




Studies (5)

D0 Single Top Measurement

Channel: electron, 1-tag, 2-jet, ~ 500 counts spread over $M = 50$ bins.



Courtesy D0 Collaboration

refpriors

This package (written in C++) implements the single and multi-count models and provides classes to calculate reference priors for *any* model (binned or un-binned) that depends on a single parameter of interest.

The code has been released to the Physics Statistics Code Repository (phystat.org).

Our near-term plan is to incorporate the classes into RooStats.

The longer term plan is to (have someone) implement *reference analysis* methods in RooStats, building on the work done in the refpriors package.

Conclusions & Summary

- Incorporating reference priors into high energy physics analyses seems feasible (see arXiv:1002.1111v2).
- We have released code to perform the crucial first step of reference analysis: the construction of reference priors.
- We have implemented two construction methods depending on what prior information is available.
- Could this be the beginning of the end of flat prior mania?