Discussion of The Small n Problem
By Glen Cowan
and
Bayesian Methods in Particle Physics
From Small N to Large
By Harrison Prosper

Michael Woodroofe and Bodhisattva Sen

University of Michigan

Themes

Both Papers

- Confidence Limits
- Discovery

Prosper: The Classificaton Problem

- Marks: Variables observed with events
- Different distributions for background and signal

Brief Outline: Confidence; Discovery; (try to) connect Discovery to Classification.

Confidence Limits for s

The Small n Problem

The Simplest Case: Given

$$N = N_b + N_s \sim \text{Poisson}(b + s)$$

where b is known and s is unknown, find an upper limit for s.

Litmus Tests:

- Equivariance under transformations, $s' = \phi(s)$.
- No dependence on b when $N = 0$.

Note: If $N = 0$, then $N_b = 0$ and $N_s = 0$.

Confidence: Continued

Classical Solutions

- Invert UMP Tests of $H_0 : s \geq s_0$.
- Regions of high relative likelihood, the one-sided case.

Litmus Tests: Equivariant, but can depend on b when $N = 0$.

Example: KARMEN: When $b = 3$ and $N = 0$, the UMP limit is negative and the RHL is 1.08, for a 90% interval.

The Uniform Bayesian Solution

Posterior Distributions: If $s \sim \text{Uniform}(0, \infty)$, then

$$P[s > s_0 | n] = \frac{\int_{s_0}^{\infty} (b+s)^n e^{-(b+s)} ds/n!}{\int_0^{\infty} (b+s)^n e^{-(b+s)} ds/n!} = \frac{\int_{b+s_0}^{\infty} t^n e^{-t} dt}{\int_b^{\infty} t^n e^{-t} dt}$$

Credible Sets: A level $1 - \gamma$ credible set is $s \leq s_{\text{up}}(n)$, where

$$P[s > s_{\text{up}}(n) | n] = \gamma.$$

Example: KARMEN: When $b = 3$ and $N = 0$, $s_{\text{up}} = 2.31$.

Litmus Tests: Independent of b when $n = 0$, but the uniform prior is not invariant.

The Uniform Bayesian Solution: Continued
AKA: The Conditional Frequentist Solution

Find:

$$P[s > s_0|n] = P_{s_0}[N \leq n|N_b \leq n].$$

Solve

$$P_{s_{\mathrm{up}}(n)}[N \leq n|N_b \leq n] = \gamma. \qquad (*)$$

Then

$$P_s[s \leq s_{\mathrm{up}}(N)] = \cdots \geq 1 - \gamma.$$

Notes 1: The construction in (*) is equivariant under parameter transformations.

2. Forget the Bayesian derivation; keep the answer. [Roe and W. (1999, *Phys. Rev.*, D)]

Complications

Intervals: Both Bayes and RHL extend

Nuisance Paramters: Unknown b, but $M \sim \text{Poisson}(\tau b)$.

- Prosper: conjugate priors.
- Heinrich: confidence limits with non-informative priors.

Frequentist

- RHL: Approximate, $P_{s,\hat{b}_s}[s \leq s_{\text{up}}] \geq 1 - \gamma$.
- Hybrid resampling [Chuang and Lai (2000, *Statistica Sinica*)].

Marks: Can be included, except for conditional frequentist [Roe and W. (2000), *Phys. Rev*, D.)].

Foundations

Bayesian Methods in Particle Physics

Points of Agreements

- (Subjective) Bayes is coherent
- Bayes is interesting, and useful.
- Both Bayesian and Frequentist approaches are needed.

Other Points

- May not have a prior, especially in high dimensional problem
- Mixtures of conjugate priors.

A Toy Example

Model: $X \sim \text{Normal}(\mu.\sigma^2)$

Goal: Estimate $\theta = \mu^2$

Estimators

MLE $\hat{\theta} = X^2$.

Unbiased $\hat{\theta} = X^2 - \sigma^2$

Uniform Bayes $\hat{\theta} = X^2 + \sigma^2$

Properties

- Unbiased can be negative.
- Bayes is inadmissible.

The Discovery Problem

Marks: Suppose:

- $N = N_b + N_s$, as above
- With each event a mark X is observed; and

$$f(x) = \frac{bf_b(x) + sf_s(x)}{b + s}.$$

Classification: For a single event,

$$P[\text{Signal}|X = x] = \frac{sf_s(x)}{bf_b(x) + sf_s(x)}.$$

Unknown Parameters

- Bayesian: Conjugate priors or Simulation.
- Frequentist: Maximum Likelihood, using EM.

An Application to Astronomy

Observtions from Sextans

- Members or foreground stars.
- X = position and velocity.
- Velocity is normal given position.
- Interest in velocity dispersions.

Analysis

- (Almost) a full likelihood analysis, using EM.
- Uses estimated probability of membership in a weighted log-likelihood.

Ref: Sen and Walker, Poster.

The Discovery Problem

Recall: $N \sim \text{Poisson}(b+s)$ and $X_1, \cdots, X_n \sim^{\text{ind}} f|n$.

Likelihood

$$L(s|\text{Data}) = \frac{1}{n!} e^{-(b+s)} \prod_{i=1}^{n} [b f_b(x_i) + s f_s(x_i)].$$

Prior Distribution: Let

$$\Pi(t) = P[s \leq t]$$

$$\pi_0 = P[s = 0],$$

and

$$\pi(t)dt = P[t \leq s \leq t + dt | s > 0].$$

Conventional Formulation

Test $H_0 : s = 0$

Posterior Probability: If $N = n,\; X_1 = x_1, \cdots$, and $X = x_n$, then

$$\pi_0^* = P[S = 0|\text{Data}] = \frac{\pi_0}{\pi_0 + (1 - \pi_0)B},$$

where

$$B = \int_0^\infty \frac{L(s|\text{Data})}{L(0|\text{Data})} \pi(s) ds$$

Note: Depends crucially on π_0.

If $\pi_0 = 0$, then $\pi_0^* = 0$ for any n and $x_1, \cdots, x_n$.

In More Detail

$$B = \sum_{k=0}^{n} \frac{1}{b^k} C_{n,k} \int_0^{\infty} s^k e^{-s} \pi(s) ds,$$

where

$$C_{n,k} = \sum_{j_1+\cdots+j_n=k} \prod_{i=1}^{n} \left(\frac{f_s(x_i)}{f_b(x_i)} \right)^{j_i}.$$

and $j_i = 0$ or 1.

The Uniform Case. If $\pi(s) \equiv 1$, then

$$B = \sum_{k=0}^{n} \frac{k!}{b^k} C_{n,k}.$$

Alternative Formulation

Have we Seen a Signal Yet: Is $N_s > 0$?

Find:

$$P[N_s = 0|\text{Data}] = \frac{\int_0^\infty e^{-s} d\Pi(s)}{\sum_{k=0}^{n} \binom{n}{k} b^{-k} C_{n,k} \int_0^\infty s^k e^{-s} d\Pi(s)}.$$

Dependence on π_0; Apparently, less.

Notes 1: Can have $\pi_0 = 0$ and $P[N_s = 0|\text{Data}] \approx 1$.

2. Simplifies in the Uniform Case.

3. Implicit in the Classification Problem.

4. Another alternative: consider posterior expected distance of s from 0.