Confidence Intervals and Nuisance Parameters

Common Example

Interval Estimation in Poisson Sampling with Scale Factor and Background Subtraction

The Problem (eg): A "Cut and Count" analysis for a branching fraction B finds n events.

- The background estimate is $\hat{b} \pm \sigma_b$ events.
- The efficiency and parent sample are estimated to give a scaling factor $\hat{f} \pm \sigma_f$.

How do we determine a (frequency) Confidence Interval?

- Assume n is sampled from Poisson, $\mu = \langle n \rangle = fB + b$.
- Assume $\hat{b}$ is sampled from normal $N(b, \sigma_b)$.
- Assume $\hat{f}$ is sampled from normal $N(f, \sigma_f)$.

Example, continued

The likelihood function is:

$$\mathcal{L}(n, \hat{b}, \hat{f}; B, b, f) = \frac{\mu^n e^{-\mu}}{n!} \frac{1}{2\pi\sigma_b\sigma_f} e^{-\frac{1}{2}\left(\frac{\hat{b}-b}{\sigma_b}\right)^2 - \frac{1}{2}\left(\frac{\hat{f}-f}{\sigma_f}\right)^2}.$$

$$\mu = \langle n \rangle = fB + b$$

We are interested in the branching fraction B. In particular, would like to summarize data relevant to B, for example, in the form of a confidence interval, without dependence on the uninteresting quantities b and f.

b and f are "nuisance parameters".

Interval Estimation in Poisson Sampling (continued)

Variety of Approaches – Dealing With the Nuisance Parameters

- Just give n, $\hat{b} \pm \sigma_b$, and $\hat{f} \pm \sigma_f$.
  - Provides "complete" summary.
  - Should be done anyway.
  - But it isn't a confidence interval...
- Integrate over $N(\hat{f}, \sigma_f)$ "PDF" for f, $N(\hat{b}, \sigma_b)$ "PDF" for b. (variant: normal assumption in $1/f$).
  - Quasi-Bayesian (uniform prior for f, b (or, eg, for $1/f$)).
- Ad hoc: eg, Upper limit – Poisson statistics for n, but with scale, background shifted by uncertainty.
  - Easy
  - makeshift; extension to two-sided intervals?
- Fix f and b at maximum likelihood estimates; include uncertainty in systematics.
- Approximate evaluation with change in likelihood as in "MINOS".

The "MINOS" Method

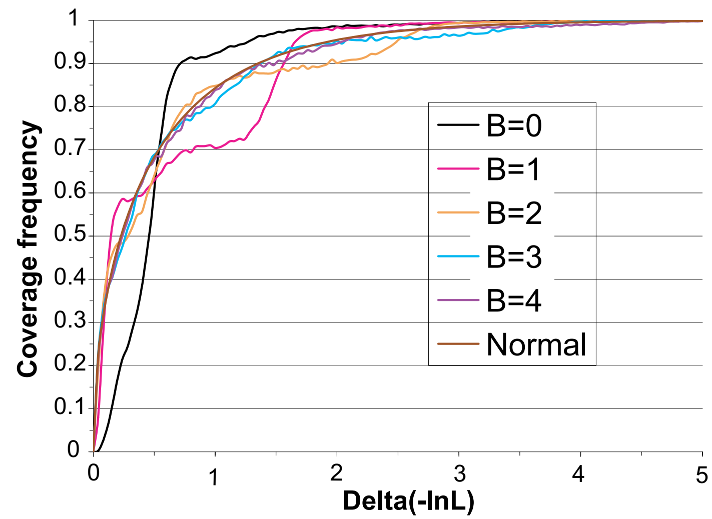
1. Write down the likelihood function in all parameters.
2. Find the global maximum.
3. Search in B parameter for where $-\ln L$ increases from minimum by specified amount (e.g., $\Delta = 1/2$), re-optimizing with respect to f and b.

Does it work? Investigate the frequency behavior of this algorithm.

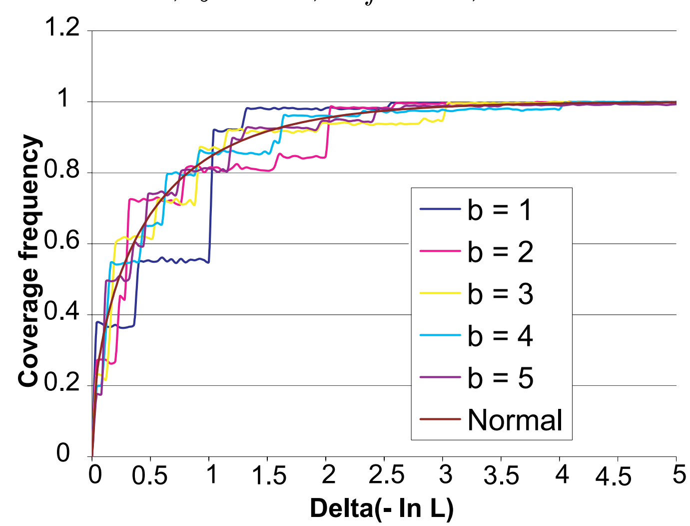
- For large statistics (normal distribution), we know that for $\Delta = 1/2$ this produces a 68% confidence interval on B.
- How far can we trust it into the small statistics regime?

Method also applicable to unbinned analysis.

$f = 1.0, \sigma_f = 0.1,\ b = 0.5, \sigma_b = 0.1$



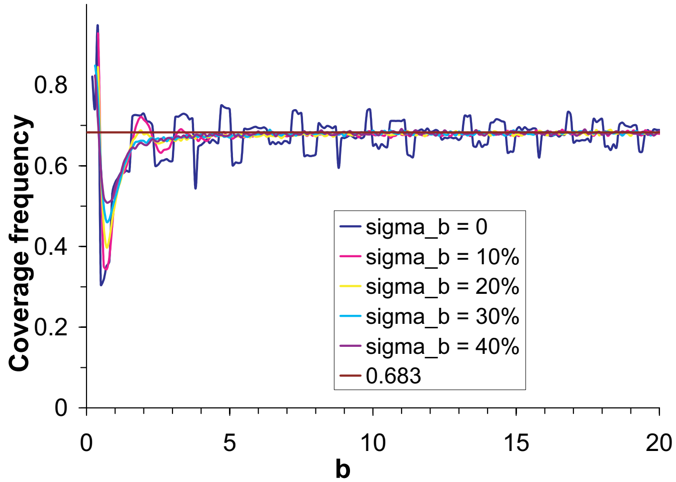
$B = 0,\ f = 1,\ \sigma_f = 0,\ \sigma_b = 0$



Study of coverage (continued)

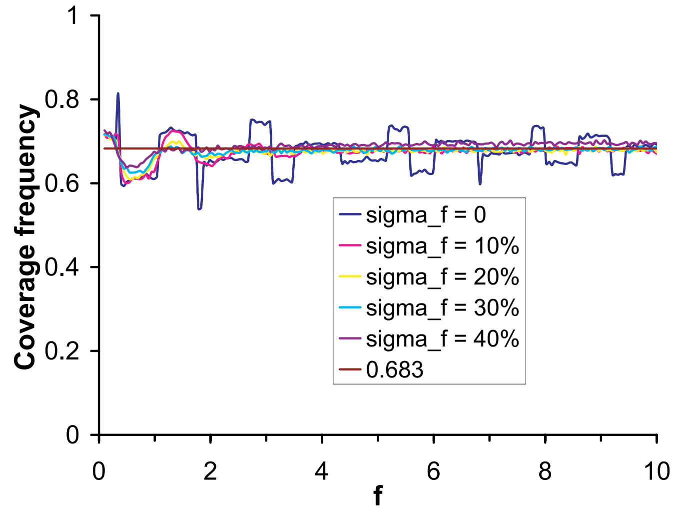
Dependence on b and σ_b

$B = 0,\ f = 1,\ \sigma_f = 0,\ \Delta = 1/2$



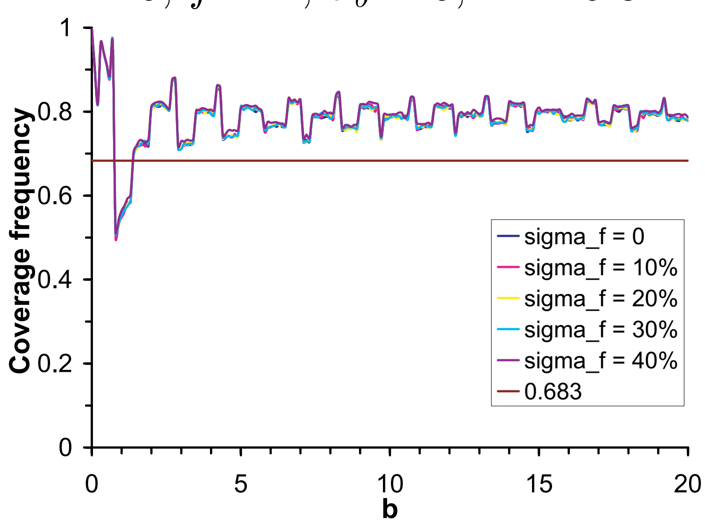
Dependence on f and σ_f for $B = 1$

$B = 1,\ b = 2,\ \sigma_b = 0,\ \Delta = 1/2$



Changing Δ

$B = 0,\ f = 1,\ \sigma_b = 0,\ \Delta = 0.8$



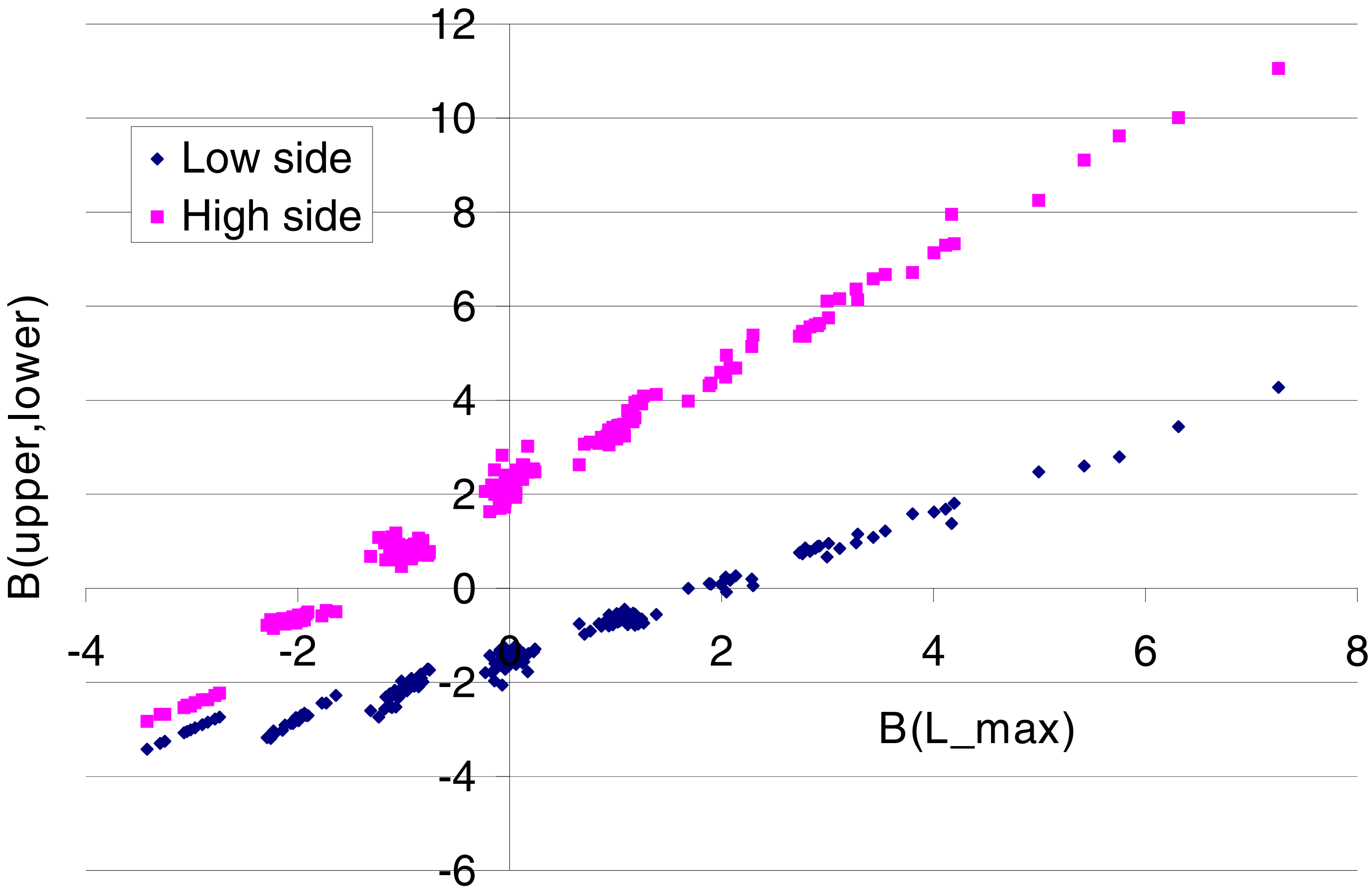
– Uncertainty in background and scale helps.

– Can increase Δ if want to put a floor on coverage.

What the intervals look like

200 experiments

$$\Delta = 1/2,\ B = 0,\ f = 1.0, \sigma_f = 0.1,\ b = 3.0, \sigma_b = 0.1.$$



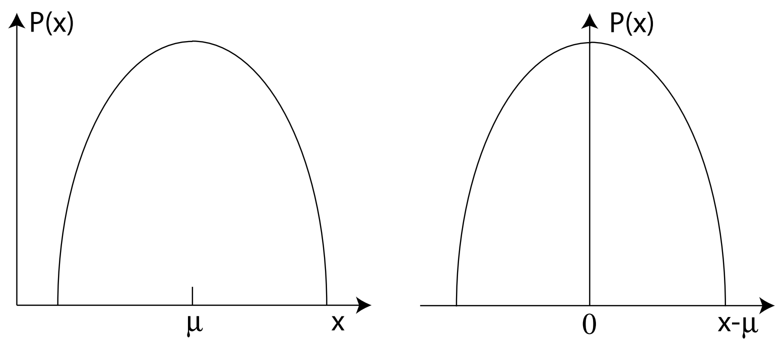
Summary: Confidence Intervals with Low Statistics

- Always give n, $\hat{b} \pm \sigma_b$, and $\hat{f} \pm \sigma_f$.
- Justify chosen approach with computation of frequency.
- Likelihood method considered here works pretty well (Well enough?) even for rather low expected counts, for 68% confidence intervals. Uncertainty in b, f improves coverage.
- If $\sigma_b \approx b$ or $\sigma_f \approx f$, enter a regime not studied here. Normal assumption probably invalid.
- Could choose larger $\Delta(-\ln L)$ if want to insure at least 68%, or push to very low statistics.
- Good enough for 68% confidence interval doesn't mean good enough for significance test. If statistics is such that Gaussian intuition is misleading, should ensure this is understood.

What is in the statistics books? (without all the rigor...)

Pivotal Quantity: Consider a sample $X = (X_1, X_2, \ldots, X_n)$ from population P, governed by parameters θ. A function $R(X, \theta)$ is called pivotal iff the distribution of R does not depend on θ.

Generalization of the feature of a "Location parameter": If θ is a location parameter for X, then the distribution of $X - \theta$ is independent of θ.



Confidence Intervals from Pivotal Quantities

Let $R(X, \theta)$ be a pivotal quantity, and α be a desired confidence level.

Find (constants!) c_1, c_2 such that:

$$P\left[c_1 \leq R(X, \theta) \leq c_2\right] \geq \alpha.$$

[We'll use "$= \alpha$" henceforth, presuming a continuous distribution.]

Now define:

$$C(X) \equiv \{\theta : c_1 \leq R(X, \theta) \leq c_2\}.$$

$C(X)$ is a confidence region with α confidence level, since

$$P[\theta \in C(X)] = P[c_1 \leq R(X, \theta) \leq c_2] = \alpha.$$

Pivotal Quantities: Example

Consider sampling (i.i.d.) $X = X_1, \ldots, X_n$ from pdf of form (eg, Gaussian):

$$p(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right).$$

- Case I: σ known. Then $X_i - \mu$, for any i, is pivotal. Also, the quantity $\bar{X} - \mu$ is pivotal, where $\bar{X}$ is the sample mean, $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$. As a sufficient statistic, $\bar{X}$ is a better choice for forming a confidence set for μ.

- Case II: Both μ and σ unknown. Let s^2 be the sample variance:

$$s^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

  – s/σ is a pivotal quantity, and can be used to derive a confidence set (interval) for σ (since μ does not appear).

Case II, continued

– Another pivotal quantity is:

$$t(X) \equiv \frac{\bar{X} - \mu}{(s/\sqrt{n})}.$$

This permits confidence intervals for μ:

$$\{\mu : c_1 \leq t(X) \leq c_2\} = \left(\bar{X} - c_2 s/\sqrt{n}, \bar{X} - c_1 s/\sqrt{n}\right)$$

at the α confidence level, where

$$P\left(c_1 \leq t(X) \leq c_2\right) = \alpha.$$

Remark: $t(X)$ is often called a "Studentized† statistic" (though it isn't a statistic, since it depends also on unknown μ). In the case of a normal sampling, the distribution of t is Student's t_{n-1}.

†Student was a pseudonym for William Gosset.

Confidence Intervals from Inverting Test Acceptance Regions

- For any test T (of hypothesis H_0 versus H_1) we define statistic ("decision rule") $T(X)$ with values 0 or 1 (for a "non-randomized test").
- $T(X) = 0$ corresponds to acceptance of H_0, and $T(X) = 1$ to rejection.
- The set $A = \{x : T(x) \neq 1\}$ is called the "acceptance region". We call $1 - \alpha$ the "significance level" of the test if

$$1 - \alpha = P\left[T(X) = 1\right], \quad H_0 \text{ is true.}$$

  That is, the significance level is the probability of rejecting H_0 when H_0 is true ("Type I error").

Confidence Intervals from Inverting Test Acceptance Regions

Let T_{θ_0} be a test for $H_0 : \theta = \theta_0$ with significance level $1 - \alpha$ and acceptance region $A(\theta_0)$. Let, for each x,

$$C(x) = \{\theta : x \in A(\theta)\}.$$

Now, if $\theta = \theta_0$,

$$P\left(X \notin A(\theta_0)\right) = P(T_{\theta_0} = 1) = 1 - \alpha.$$

That is, again for $\theta = \theta_0$,

$$\alpha = P\left[X \in A(\theta_0)\right] = P\left[\theta_0 \in C(X)\right].$$

This holds for all θ_0, hence, for any $\theta_0 = \theta$,

$$P\left[\theta \in C(X)\right] = \alpha.$$

That is, $C(X)$ is a confidence region for θ, at the α confidence level.

Confidence Intervals from Inverting Test Acceptance Regions

We often use ordering on the likelihood ratio to determine our acceptance region. Hence, the likelihood ordering may be used to construct confidence sets.

That is, we define the "Likelihood Ratio":

$$\lambda(\theta; x) \equiv \frac{L(\theta; x)}{\max_{\theta'} L(\theta'; x)}.$$

For any $\theta = \theta_0$, we build acceptance region according to:

$$A(\theta_0) = \{x : T_{\theta_0}(x) \neq 1\},$$

where

$$T_{\theta_0}(x) = \begin{cases} 0 & \lambda(x; \theta_0) > \lambda_\alpha(\theta_0) \\ 1 & \lambda(x; \theta_0) < \lambda_\alpha(\theta_0), \end{cases}$$

and $\lambda_\alpha(\theta_0)$ is determined by requiring, for $\theta = \theta_0$,

$$P\left[\lambda(X; \theta_0) > \lambda_\alpha(\theta_0)\right] = \alpha.$$

What about nuisance parameters?

- Suppose we are interested in some parameters $\mu \subset \theta$, where $\dim(\mu) < \dim(\theta)$. Let $\eta \subset \theta$ stand for the remaining "nuisance" parameters.
- If you can find pivotal parameters (e.g., normal distribution), great! But not always possible.
- Test acceptance region approach also problematic: H_0 becomes "composite"†, since nuisance parameters are unspecified. In general, we don't know how to construct the acceptance region with specified significance level for

$$H_0 : \mu = \mu_0; \eta \text{ unspecified.}$$

† A "simple hypothesis" is one in which the population is completely specified.

Asymptotic Inference

- When can't (or won't) do exact solution, can base approximate treatment on asymptotic crieria.

Let $X = (X_1, ..., X_n)$ be a sample from population $P \in \mathcal{P}$. Let θ be a parameter vector for P, and let $C(X)$ be a confidence set for θ. If $\liminf_n P\left[\theta \in C(X)\right] \geq \alpha$ for any $P \in \mathcal{P}$, then α is called an "Asymptotic Significance Level" of $C(X)$.

If $\lim_{n\to\infty} P\left[\theta \in C(X)\right] = \alpha$ for any $P \in \mathcal{P}$, then $C(X)$ is an "α Asymptotically Correct" confidence set.

- Many possible approaches, for example, can look for "Asymptotically Pivotal" quantities; or invert acceptance regions of "Asymptotic Tests".

Profile Likelihood

Consider likelihood $L(\mu, \eta)$, based on observation $X = x$. Let

$$L_P(\mu) = \sup_\eta L(\mu, \eta).$$

$L_P(\mu) = L\left(\mu, \eta(\mu)\right)$ is called the "Profile Likelihood" for μ.

The "MINOS" method of error estimation makes use of the profile likelihood.

Let $\dim(\mu) = r$. Consider the likelihood ratio test for $H_0 : \mu = \mu_0$ with

$$\lambda(\mu_0) = \frac{L_P(\mu_0)}{\max_{\theta'} L(\theta')},$$

where $\theta = \{\mu, \eta\}$. The set

$$C(X) = \{\mu : -2\ln\lambda(\mu) \geq c_\alpha\},$$

where c_α is the χ^2 corresponding to the α probablity point of a χ^2 with r degrees of freedom, is an α asymptotically correct confidence set.

Conditional Likelihood

Consider likelihood $L(\mu, \eta)$. Suppose $T_\eta(X)$ is a sufficient statistic for η for any given μ. Then conditional distribution $f(X|T_\eta)$ does not depend on η. The likelihood function corresponding to this conditional distribution is called the "Conditional Likelihood".

- Note that estimates (e.g., MLE for μ) based on conditional likelihood may be different than for those based on full likelihood.
- This eliminates the nuisance parameter problem, if it can be done without too high a price.

More someday. . .

Bootstrap?