Statistics II

Frank Porter

September 19, 2013

Plan for the Statistics Lectures

- Lecture I (Wednesday, September 18, 11:45-12:30)
  1. Important probability concepts
  2. Point estimation
- Lecture II (Thursday, September 19, 10:45-12:30)
  1. Frequency and Bayes interpretations
  2. Interval estimation
  3. Systematic uncertainties
- Lecture III (Friday, September 20, 10:45-12:30)
  1. Hypothesis tests
  2. Resampling methods
- Lecture IV (Saturday, September 21, 10:45-12:30)
  1. Density estimation

Lecture boundaries won't be so crisp!!

Frequency and Bayes interpretations

- A frequentist makes a statement about the data
  - Descriptive
  - Make probability statements that are correct in the frequency sense
    [Imagine repeating the experiment many times]
  - Sampling distribution is essential for frequentist
- A Bayesian makes a statement about the truth
  - Inferential
  - Make probability statements about degree-of-belief
    [Inherently subjective; different prejudices]
  - Uses Bayes theorem for inferences about truth
  - Just needs likelihood function
- Objective Bayes approach is different
  - Still uses Bayes theorem, but follow a prescription ("objective")
  - Give up degree-of-belief interpretation
  - Also not frequentist
- We'll see how differences arise in interval estimation

Interval estimation

1. Basics
2. Nuisance parameters
3. Pivotal quantities
4. Inversion of test acceptance region
5. Profile likelihood
6. Asymptotic intervals (supplemental slide)
7. Bootstrap (when we get to resampling)

Neyman's definition of Confidence Interval

"If the functions θ_ℓ and θ_u possess the property that, whatever be the possible value ϑ_1 of θ_1 and whatever be the values of the unknown parameters $\theta_2, \theta_3, \ldots, \theta_s$, the probability

$$P\{\theta_\ell \le \vartheta_1 \le \theta_u | \vartheta_1, \theta_2, \ldots, \theta_s\} \equiv \alpha,$$

then we will say that the functions θ_ℓ and θ_u are the lower and upper confidence limits of θ_1, corresponding to the confidence coefficient α."

J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability", *Phil. Trans. Royal Soc. London* **236** (1937) 333-380

- The interval (θ_ℓ, θ_u) is called the Confidence Interval (CI) for θ_1
- We'll use $1-\alpha$ instead of α for the confidence coefficient (or Confidence Level (CL))

Example: Mean of a Normal Distribution

- Suppose we sample a value x from a $\mathrm{N}(\theta, 1)$ distribution
- We could form a 68% confidence interval for θ as follows:
  - Throw a random number, u, uniform in $(0, 1)$
  - If $u < 0.68$, quote the interval $(-\infty, \infty)$
  - Otherwise, quote the null interval
- This is clearly a valid confidence interval, including the exact value of θ, independent of θ, with a probability (frequency) of precisely 68%
- However, it is a useless exercise – it has nothing to do with the measurement!
- To be useful, we must require more from our interval estimation; we should strive for sufficiency!
- Could also ask for other properties, such as equal distances on each side of a point estimator or for the smallest interval

Computing a confidence set

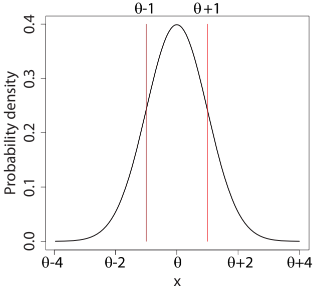
- In general can compute a confidence set in a multi-dimensional parameter space
- Construct a confidence set as follows:
  - For any possible (including perhaps "non-physical") parameter vector θ, construct a set of observations x such that θ will be included in the $1-\alpha$ confidence region. Call this set $S_\alpha(\theta)$
  - Many possible ways to construct such sets. Often, choose to construct the smallest such set
  - In this case, $S_\alpha(\theta)$ is the smallest set for which

$$\int_{S_\alpha(\theta)} f(x;\theta)\mu(dS) \geq 1-\alpha,$$

    where $x \in S_\alpha(\theta)$ for all x such that $f(x;\theta) \geq \min_{x\in S_\alpha(\theta)} f(x;\theta)$
  - That is, the set is constructed by ordering probabilities, and including x values for which the probabilities are greatest
  - Finally, given an observation x, the confidence set $C_\alpha(x)$ at the $1-\alpha$ confidence level is the set of all θ for which $x \in S_\alpha(\theta)$
  - By construction, $C_\alpha(X)$ has a probability of at least $1-\alpha$ to include θ. Note: $C_\alpha(X)$ is a RV.

Example Continued; Try Again

- Notice that 68% of the time we sample x from an $\mathrm{N}(\theta, 1)$ distribution, the value of x will be within 1 unit of θ
- Thus, if we quote interval $(x-1, x+1)$, we have a valid 68% confidence interval for θ – The quoted interval will include θ with probability (frequency) of precisely 68%



  - For any given sample, the quoted interval either includes θ or it doesn't. Might even know that it doesn't, e.g., if the interval is outside some "physical" boundary on allowed values of θ. This is irrelevant!
  - This is precisely the construction of the smallest interval as on the previous slide. These statistics are also sufficient

The Mind Set Problem

Neyman goes on to say:

"In spite of the complete simplicity of the above definition, certain persons have difficulty in following it. These difficulties seem to be due to what Karl Pearson (1938) used to call routine of thought. In the present case the routine was established by a century and a half of continuous work with Bayes's theorem..."

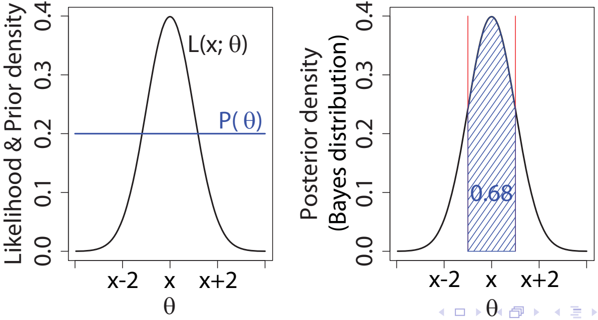
Physicists have the same difficulty

Bayesian Intervals

- A Bayesian Interval, at the $1-\alpha$ confidence level, for a population parameter θ is an interval which contains a fraction $1-\alpha$ of the area under a Bayes' distribution
- A Bayes' Distribution is a function of θ [Bayes thm!]:

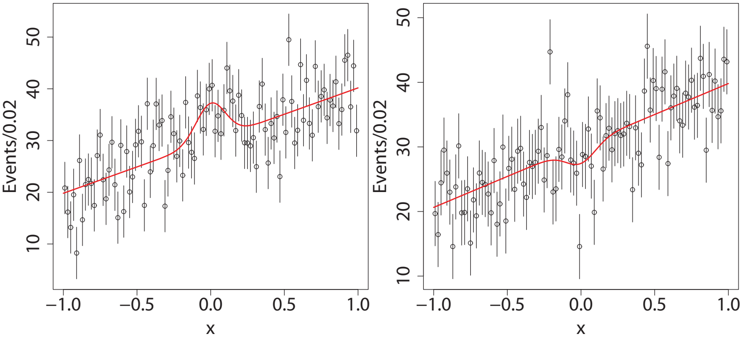
$$P(\theta; x) = f(x;\theta)P(\theta) \Big/ \int_{-\infty}^{\infty} f(x;\theta)P(\theta)d\theta,$$

where $P(\theta)$ is called the prior distribution, and $f(x;\theta)$ is the pdf evaluated at the observed value x (likelihood)



Compare frequentist and Bayes intervals – Example

Meaure signal strength in a histogram. Large statistics (bin contents distributed according to Gaussians to good approximation). Look at two examples, measuring event yields:

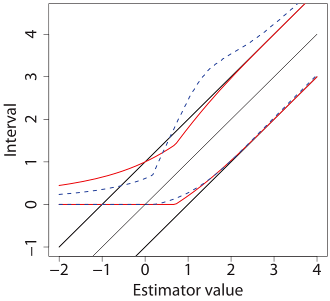


Frequentist:	90.4 ± 16.4	-34.8 ± 16.4
Bayesian:	90.4 ± 16.4	$0^{+7.1}_{-0.0}$

(Uniform prior assumed for Bayes intervals, negative signals unphysical)

Two kinds of statistics – comparison of intervals

Eg, Sampling from $N(\theta, 1)$



Black: 68% confidence interval (frequentist)

Red: 68% Bayesian interval (uniform prior)

Dashed blue: 68% Bayesian interval (square root prior)

Negative peaks are assumed to be unphysical

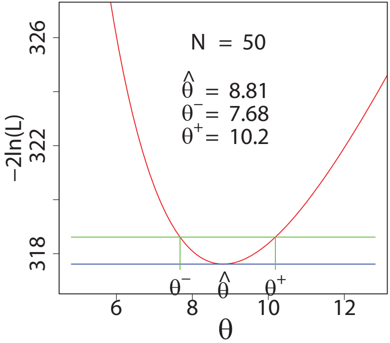
- Irrelevant for frequentist interval, describing the measurement
- Bayesian prior excludes unphysical regions
- Choice of prior matters (we tend to use ignorance priors)
- Bayes intervals may undercover or overcover in terms of frequency, but this is irrelevant

Error estimation for earlier exponential example

In the asymptotic (central limit theorem) limit, a 68% confidence interval is obtained by finding those points where $-2\ln L$ increases by 1 unit from the value at the minimum. Example, $N = 50$, $\theta = 10$:

Note the asymmetry, $\hat{\theta} = 8.8^{+1.4}_{-1.1}$

Exercise: Check coverage.



For Gaussian sampling, $-2\ln L(\theta; x) = [x - g(\theta)]^{\mathsf{T}}\, \Sigma^{-1}\, [x - g(\theta)]$

Advice on intervals

- At least describe observation
  - Two-sided confidence intervals (frequentist) recommended – facilitates combining results
  - This recommendation is independent of "significance" – Note that deciding how to quote result based on what the result is introduces biases
  - In low-statistics regime, give raw numbers
- Optionally, provide an interpretation
  - E.g., Bayesian upper limit if desired
  - State prior used
  - Check sensitivity to possible priors

The problem of 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
- We can use our ordering method, or other methods below, to construct confidence regions in θ
- But this fails if we want to construct a confidence set in the proper subspace for parameters μ, since we cannot construct set $S_\alpha(\theta)$ without specifying θ completely
- This is a hard problem in general in frequency statistics, but we'll develop several approaches
- It is an easy problem (up to choosing priors) in Bayesian statistics

Nuisance parameters in Bayesian statistics

- Write down the full likelihood function in all parameters, $L(\mu, \eta; x)$ [or, more properly, the posterior]
- Integrate over the nuisance parameters to eliminate them, yielding the marginal likelihood:

$$L_M(\mu; x) = \int L(\mu, \eta; x) d\eta$$

- In principle, need a prior in nuisance parameters, but typically taken to be uniform, as here

Profile likelihood approach to nuisance parameters

- Consider likelihood $L(\mu, \eta; x)$. Define the Profile Likelihood for μ (suppressing x):
$$L_P(\mu) = \sup_\eta L(\mu, \eta)$$
- Minuit's MINOS method uses the profile likelihood idea
- For Gaussian sampling, intervals obtained with the profile likelihood have correct coverage
- More generally, obtain a lower bound on coverage
- Good asymptotic behavior: Let $\dim(\mu) = k$. Consider likelihood ratio:
$$\lambda(\mu) = \frac{L_P(\mu)}{\max_\theta L(\theta)}$$
The set
$$C_\alpha(x) = \{\mu : -2 \ln \lambda(\mu) \le c_\alpha\},$$
where c_α is the χ^2 corresponding to the $1-\alpha$ probablity point of a χ^2 with k degrees of freedom, is a $1-\alpha$ "asymptotically correct" confidence set

Case study for a Common Situation

Interval Estimation in Poisson Sampling with Scale Factor and Background Subtraction

The Problem: 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 for B?
Must know the sampling distribution

- Assume n is sampled from Poisson, $\mu = \langle n \rangle = fB + b$
- Assume $\hat{b}$ is sampled from normal $\mathrm{N}(b, \sigma_b^2)$
- Assume $\hat{f}$ is sampled from normal $\mathrm{N}(f, \sigma_f^2)$
  [Variant:$1/\hat{f}$ sampled from a normal]

Note that since we don't know b, f, we have some missing information: nuisance parameters

Example, continued

The likelihood function is:

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

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

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 nuisance parameters b and f

*Example, continued

Variety of Approaches – Dealing With the Nuisance Parameters
Just a few of the things people do, not an endorsement!!!

- 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 for B...
- Integrate over N$(\hat{f}, \sigma_f)$ "pdf" for f, N$(\hat{b}, \sigma_b)$ "pdf" for b
  - 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? Not recommended!
- Fix f and b at ML estimates; include uncertainty in systematics
- Approximate evaluation with profile likelihood ratio. Let's investigate this further

The profile likelihood method: Case study

- Write down the likelihood function in all parameters
- Find the global maximum
- 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 (ie, use profile likelihood)

Does it work? Investigate the frequency behavior of this algorithm

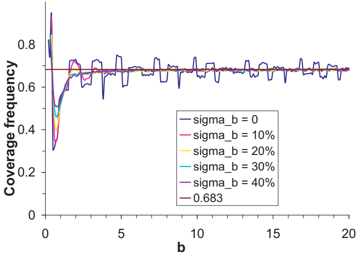
- For large statistics (normal distribution), we know that for $\Delta = 1/2$ this produces a 68% confidence interval on B
- How far can we trust it into the small statistics regime?

Remark: Method also applicable to unbinned analysis

Case study – Coverage

Dependence on b and σ_b

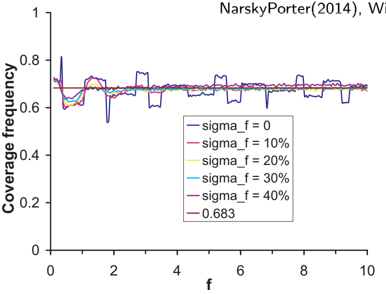
$B = 0,\ f = 1,\ \sigma_f = 0,\ \Delta = 1/2$



Dependence on f and σ_f for $B = 1$

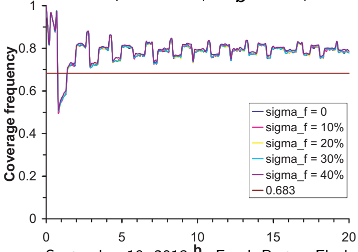
$B = 1,\ b = 2,\ \sigma_b = 0,\ \Delta = 1/2$

NarskyPorter(2014), Wiley



Changing Δ

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

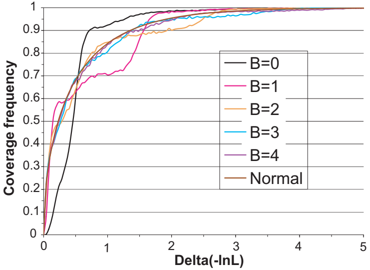


- Uncertainty in background and scale helps.
- Can increase Δ if want to put a floor on coverage (probably can do better though)

*The profile likelihood method: Case study
Study of coverage

Coverage as function of $\Delta(-\ln L)$ for:

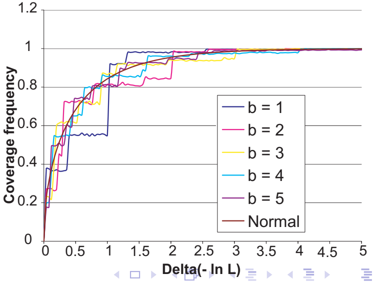
- $f = 1.0$
- $\sigma_f = 0.1$
- $b = 0.5$
- $\sigma_b = 0.1$



NarskyPorter(2014), Wiley

Coverage for:

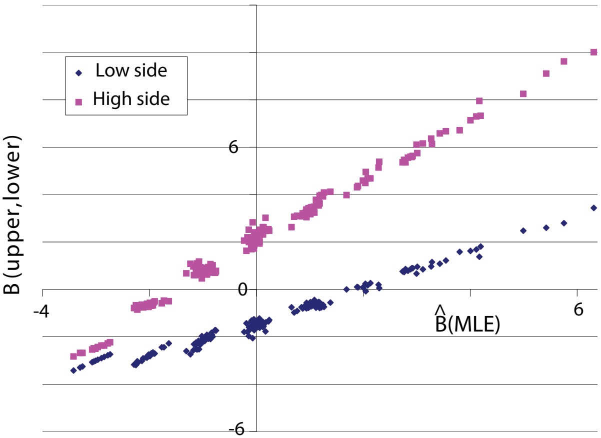
- $B = 0$
- $f = 1$
- $\sigma_f = 0$
- $\sigma_b = 0$



*What the intervals look like

200 experiments

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



NarskyPorter(2014), Wiley

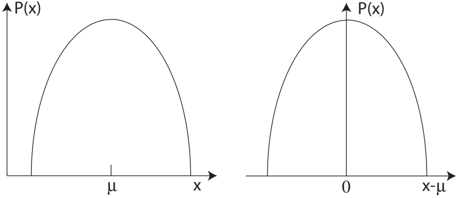
Summary of case study

Confidence Intervals with Low Statistics and Nuisance Parameters

- 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
- Note that our case study has two approximations:
  - Use of profile likelihood
  - Use of Gaussian approximation ($\Delta = 1/2$)
- If $\sigma_b \approx b$ or $\sigma_f \approx f$, enter a regime not studied here
- 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

Pivotal Quantities

- 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 μ is a location parameter for X, then the distribution of $R = X - \mu$ is independent of μ



NarskyPorter(2014), Wiley

If a suitable pivotal quantity can be found, it may be used to eliminae nuisance parameters

Confidence Intervals from Pivotal Quantities

Let $R(X,\theta)$ be a pivotal quantity, and $1-\alpha$ be the desired CL. Find (constants!) c_1, c_2 such that:

$$P\left[c_1 \le R(X,\theta) \le c_2\right] \ge 1-\alpha$$

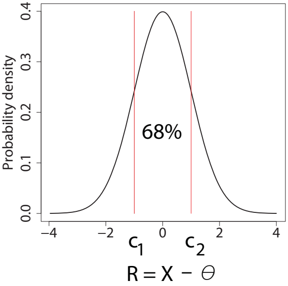
[We'll use equality henceforth, for a continuous distribution]

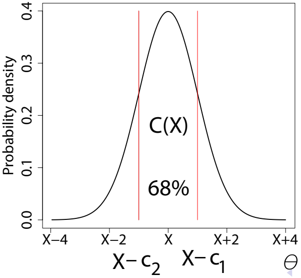
Now define:

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

$C(X)$ is a confidence region with $1-\alpha$ confidence level, since

$$P[\theta \in C(X)] = P[c_1 \le R(X,\theta) \le c_2] = 1-\alpha$$





Already used this method, in example on slide 8

NarskyPorter(2014), Wiley

Pivotal Quantities: Example

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

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

- Case I: σ known. Then $X_n - \mu$, for any n, is pivotal. Also, the quantity $\bar{X} - \mu$ is pivotal, where $\bar{X}$ is the sample mean, $\bar{X} \equiv \frac{1}{N}\sum_{n=1}^{N} X_n$. 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:

$$s^2 \equiv \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2$$

  - 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 $1 - \alpha$ CL, where

$$P\left[c_1 \leq t(X) \leq c_2\right] = 1 - \alpha$$

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 normal sampling, the distribution of t is Student's t_{n-1}

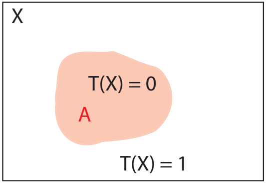
†Student is a pseudonym for William Gosset

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
- $T(X) = 0$ corresponds to acceptance of H_0, and $T(X) = 1$ to rejection
- The set $A = \{x : T(x) = 0\}$ is called the acceptance region. We call α the significance level of the test if

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

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



CIs from inverting test acceptance regions

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

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

Now, if $\theta = \theta_0$,

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

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

$$1 - \alpha = P\left[X \in A(\theta_0)\right] = P\left[\theta_0 \in C(X)\right]$$

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

$$P\left[\theta \in C(X)\right] = 1 - \alpha$$

That is, $C(X)$ is a confidence region for θ, at the $1 - \alpha$ CL

CIs from Inverting Test Acceptance Regions – Likelihood ratio test

We often use ordering on the likelihood ratio to determine our acceptance region. Hence, 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:

$$A_\alpha(\theta_0) = \{x : T_{\theta_0}(x) = 0\},$$

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 $\theta = \theta_0$,

$$P\left[\lambda(X; \theta_0) > \lambda_\alpha(\theta_0)\right] = 1 - \alpha$$

CIs from Inverting Test Acceptance Regions – Likelihood ratio test

- Supposing sufficiency, rewrite the likelihood ratio in the form:

$$\lambda(\theta, \hat{\theta}) = \frac{L(\theta; \hat{\theta})}{L(\hat{\theta}; \hat{\theta})}$$

- Suppose we observe a result $\hat{\theta}$. We go through our table of sets $A_\alpha(\theta)$ looking for $\hat{\theta}$
- Everytime we find it, we include that value of θ in our confidence region
- This gives a confidence region for θ at the $1 - \alpha$ confidence level. That is, the true value of θ will be included in the interval with probability $1 - \alpha$

CIs from Inverting Test Acceptance Regions – Likelihood ratio

Repeat this procedure as an algorithm:

1. Find $\hat{\theta}$, the value of θ for which the likelihood is maximized
2. For any point θ^* in parameter space, form the statistic

$$\lambda(\theta^*, \hat{\theta}) \equiv \frac{L(\theta^*; \hat{\theta})}{L(\hat{\theta}; \hat{\theta})}$$

3. Evaluate the probability distribution for λ (considering all possible experimental outcomes), under hypothesis that $\theta = \theta^*$. Using this distribution, determine critical value $\lambda_\alpha(\theta^*)$
4. If $\lambda(\theta^*, \hat{\theta}) \geq \lambda_\alpha(\theta^*)$, then θ^* is inside the confidence region; otherwise it is outside
5. Consider all possible θ^* to construct the entire confidence region

CIs from Inverting Test Acceptance Regions – Likelihood ratio

Analytic evaluation of probability in step (3) is often intractable. We may use MC to compute this probability. Steps (3)–(5) become:

3. Simulate many experiments using θ^* as the value(s) of the parameter(s), obtaining for each experiment the MLE $\hat{\theta}_{\rm MC}$
4. For each MC experiment, form the statistic:
$$\lambda_{\rm MC} \equiv \frac{L(\theta^*; \hat{\theta}_{\rm MC})}{L(\hat{\theta}_{\rm MC}; \hat{\theta}_{\rm MC})}$$
Critical value $\lambda_\alpha(\theta^*)$ is the number for which fraction α of MC experiments have a larger value of $\lambda_{\rm MC}$
5. If $\lambda(\theta^*, \hat{\theta}) \geq \lambda_\alpha(\theta^*)$, then θ^* is inside the confidence region; otherwise it is outside. In other words, if $\lambda(\theta^*, \hat{\theta})$ is larger than at least a fraction α of the MC experiments, then θ^* is inside the confidence region
6. Procedure repeated for many choices of θ^* to map out the confidence region

Confidence Intervals from Inverting Test Acceptance Regions

Example in 2 Dimensions: D mixing and DCSD
Using ordering on the likelihood ratio

Two mixing parameters to be determined:

$$x' \equiv \frac{\Delta m}{\Gamma}\cos\delta + \frac{\Delta\Gamma}{2\Gamma}\sin\delta,$$

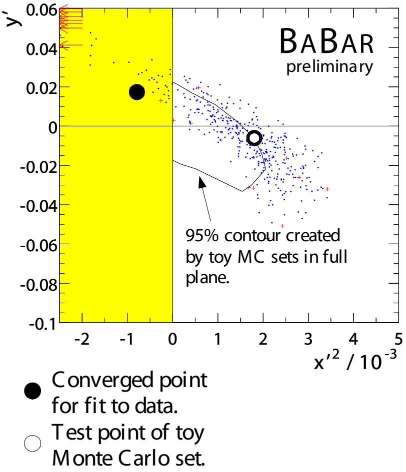
$$y' \equiv \frac{\Delta\Gamma}{2\Gamma}\cos\delta - \frac{\Delta m}{\Gamma}\sin\delta,$$

where δ is an unknown strong phase (between Cabibbo-favored and doubly Cabibbo-suppressed amplitudes). Only sensitive to x'^2, y'; ML may occur at $x'^2 < 0$ ("unphysical" region)

Example in 2 dimensions: D mixing and DCSD

Construction of 2-D confidence region (95% CL) using likelihood ratio test

Blue: $\lambda_{\mathrm{Data}} > \lambda_{\mathrm{MC}}$
Red: $\lambda_{\mathrm{Data}} < \lambda_{\mathrm{MC}}$



(U. Egede, International Workshop on Frontier Science, Frascati, October 6-11, 2002)

Confidence Intervals: "Valid" doesn't mean "Good"

Example using χ^2 statistic

- We do an experiment to measure θ by sampling 10 times from normal distribution $\mathrm{N}(\theta, 1)$
- The sum of the squared-deviations from the mean is

$$\chi^2 = \sum_{i=1}^{10} (x_i - \theta)^2$$

- Our estimator, $\hat{\theta}$, for θ, is the value of θ that minimizes this, namely the sample mean
- The minimum value is:

$$\chi^2_{\min} = \sum_{i=1}^{10} (x_i - \hat{\theta})^2$$

- This statistic is distributed according to a chi-square distribution with 9 DOF

Wish to find a 68% CI for θ

χ^2 Example – Method 1

- Let $\Delta\chi^2(\theta)$ be the difference between the χ^2 estimated at θ and the minimum value:

$$\begin{aligned}\Delta\chi^2 &\equiv \chi^2 - \chi^2_{\min} \\ &= \sum_{i=1}^{10}\left[(x_i-\theta)^2 - (x_i-\hat\theta)^2\right] \\ &= \sum_{i=1}^{10}\left[2x_i(\hat\theta-\theta) + \theta^2 - \hat\theta^2\right] \\ &= 2\times 10\hat\theta(\hat\theta-\theta) + 10\theta^2 - 10\hat\theta^2 \\ &= 10(\hat\theta-\theta)^2\end{aligned}$$

- $\hat\theta$ is normally distributed with mean θ and variance $1/10$. Finding the points where $\Delta\chi^2 = 1$ corresponds to our familiar method for finding the 68% confidence interval:

$$\left(\hat\theta - 1/\sqrt{10},\quad \hat\theta + 1/\sqrt{10}\right)$$

χ^2 Example – Method 2

- Consider the chi-square goodness-of-fit test for:

$$H_0 : \theta = \theta_0,$$
$$H_1 : \theta \neq \theta_0$$

- At the 68% significance level, we accept H_0 if

$$\chi^2(\theta_0) < \chi^2_{\rm crit}$$

where

$$F(\chi^2_{\rm crit}, 10) \equiv P(\chi^2 < \chi^2_{\rm crit}, 10) = 68\%$$

(Note: 10 DOF, since θ_0 is specified)

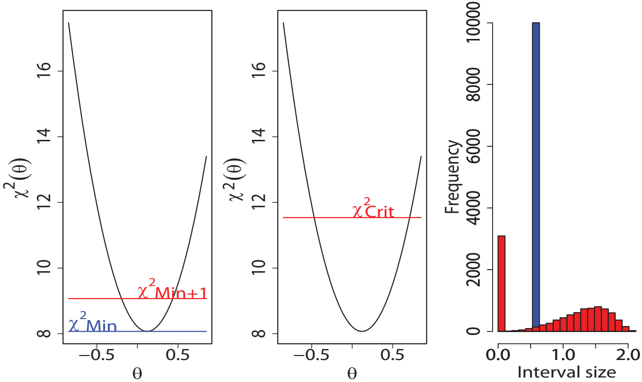
- If $\chi^2_{\rm crit} > \chi^2_{\rm min}$, we have confidence interval

$$\hat{\theta} \pm \sqrt{\left(\chi^2_{\rm crit} - \chi^2_{\rm min}\right)/10}$$

Otherwise, we have a null confidence interval

- In present example, $\chi^2_{\rm crit} = 11.54$

χ^2 Example – Comparison of Methods



NarskyPorter(2014), Wiley

Blue: Method 1 ($\Delta\chi^2$)
Red: Method 2 (χ^2 test)

Astrophysics example: Mueller & Madejski, *Ap. J.* **700** (2009) 243

*Conditional Likelihood

Consider likelihood $L(\mu, \eta)$, with interesting paramer(s) μ and nuisance parameter(s) η. 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

- 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
- We'll see an example later of the use of conditional likelihoods

Confidence intervals - Recommendations

- Use frequency statistics for summarizing information
- Goal is to decribe what we observe, with properties:
  - Simple, coherent interpretation
  - Facilitate combination with other results
  - Can be counter-productive to impose "physical" constraints. No reason to obscure observation of an "unlikely" result. Imposing constraint may complicate combination.
- Generally, recommendation is to quote two-sided 68% confidence intervals as primary result
- Check for frequency validity (coverage)
- If you want to provide an interpretation, domain is Bayesian statistics
  - Upper limits have this flavor – they are treated as implying "how big θ could be"

Comparison: Feldman-Cousins, "Usual", and "Bayes" Intervals

Assume our experiment is a sampling from an N$(\theta, 1)$ distribution.
Assume we have "physical" knowledge that $\theta > 1$.

Comparison of some 68% "Confidence" Intervals

x	"usual" interval $(x - 1\sigma, x + 1\sigma)$	Feldman-Cousins	Bayes upper limit with uniform prior
-2	(-3,-1)	(1,1.04)	1.33
-1	(-2,0)	(1,1.07)	1.44
0	(-1, 1)	(1, 1.27)	1.64
1	(0,2)	(1, 2)	1.99
2	(1,3)	(1.24,3)	2.6
3	(2,4)	(2,4)	3.5

[Feldman-Cousins is inversion of a test acceptance region, using ordering on likelihood ratio, with additional feature of stopping at physical boundary.]

Systematic uncertainties

- An error is the difference between a measured value and a true value
- It is a RV
- An uncertainty is a source of error, or a characterization of its expected size
- Error may arise from
  - Random fluctuations
  - Bias (incorrect model assumptions, including "mis-calibration"; biased procedure)
  - Mistakes (philosophical question: are mistakes RVs?), eg, mis-reading a scale
- Distinction between mistake and bias: A bias reliably repeats with repeated measurements; a mistake shouldn't
- Bias is systematic

See also Barlow, arXiv:hep-ex/0207026v1 (2002)

Systematic uncertainties

$$B(e^+e^- \to \text{Nobel prize}) = 10 \pm 1 \pm 5$$

- They may be important!
  - Maybe the ± 5 is a systematic uncertainty in the estimate of the background expectation. A "10σ" statistical significance is really only a "2σ" effect
- They may be not quite so important
  - Maybe the ± 5 is a systematic uncertainty on the efficiency, entering as a multiplicative factor. It makes no difference to the significance whether the result is 10 ± 1 or 5 ± 0.5

Systematic uncertainties – estimation

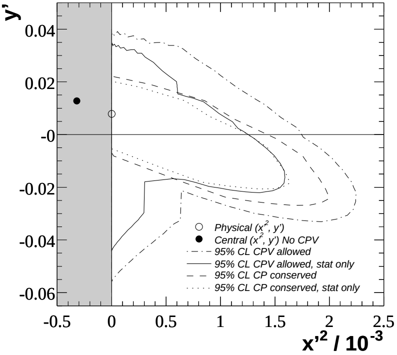
- If possible, correct for biases
- But maybe don't know sign or magnitude
  - Happens when source is a (nuisance) parameter determined in some other measurement
  - That is, bias may be another RV
  - Eg, branching fractions may be measured relative to some "standard" decay; then absolute BFs have a systematic uncertainty
- Then we quote a systematic uncertainty
- Should be quoted separately, since may be possible to correct for later (eg, when standard BF is better measured)

Systematic uncertainties – not

- Not everything is a systematic uncertainty
  - If it were,we would have endless sources of uncertainty
- We check many possibilities for mistakes or unanticipated problems
  - If find a problem, fix it, or discard result
  - Otherwise, do nothing
  - Results are quoted in context of “no mistakes”

Systematic uncertainties example: D mixing and DCSD revisited

- Want simple procedure
- Willing to accept approximation
- Scale statistical contour uniformly along ray from best-fit value. Factor is $\sqrt{1+\sum m_i^2}$, where m_i is estimated systematic uncertainty i measured in units of statistical uncertainty. Estimate obtained by determining effect of the systematic uncertainty on $\hat{x}'^2, \hat{y}'$
  - Method conservative in sense that scaling for a given systematic in one direction is applied uniformly in all directions. On the other hand, a linear approximation is being made



PRL **91** (2003) 171801

Next: Hypothesis tests; Resampling methods

Supplemental Material

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".