Confidence Intervals

First ICFA Instrumentation School/Workshop
At Morelia, Mexico, November 18-29, 2002

Harrison B. Prosper
Florida State University

Outline

- Lecture 1
  - Introduction
  - Confidence Intervals - Frequency Interpretation
  - Poisson Example
  - Summary
- Lecture 2
  - Deductive and Inductive Reasoning
  - Confidence Intervals - Bayesian Interpretation
  - Poisson Example
  - Summary

Introduction

- We physicists often talk about calculating "errors", but what we really mean, of course, is
  - quantifying our uncertainty
- A measurement is not uncertain, but it has an error ε about which we are uncertain!

$$\varepsilon = \hat{m} - m$$

$$\hat{m} = \text{measurement}$$

$$m = \text{true value}$$

Introduction - i

- One way to quantify uncertainty is the standard deviation or, even better, the root mean square deviation of the distribution of measurements.

- In 1937 Jerzy Neyman invented another measure of uncertainty called a confidence interval.

$$\langle \varepsilon^2 \rangle = \langle (\hat{m} - m)^2 \rangle$$
$$= \langle \hat{m}^2 \rangle - \langle \hat{m} \rangle^2$$
$$+ \left(\langle \hat{m} \rangle - m \right)^2$$

$$\text{rms} = \sqrt{\langle \varepsilon \rangle^2}$$
$$\text{std. dev.} = \sqrt{\langle \hat{m}^2 \rangle - \langle \hat{m} \rangle^2}$$
$$\text{bias} = \langle \hat{m} \rangle - m$$

Introduction - ii

- Consider the following questions
  - What is the mass of the top quark?
  - What is the mass of the tau neutrino?
  - What is the mass of the Higgs boson?

- Here are possible answers
  - $m_t = 174.3 \pm 5.1$ GeV
  - $m_\nu < 18.2$ MeV
  - $m_H > 114.3$ GeV

Introduction – iii

- These answers are unsatisfactory
  - because they do not specify how much confidence we should place in them.

- Here are better answers
  - $m_t = 174.3 \pm 5.1$ GeV, with CL = 0.683
  - $m_\nu < 18.2$ MeV, with CL = 0.950
  - $m_H > 114.3$ GeV, with CL = 0.950

CL = Confidence Level

Introduction - iv

- Note that the statements
  - m_t = 174.3 ± 5.1 GeV, CL = 0.683
  - m_ν < 18.2 MeV, CL = 0.950
  - m_H > 114.3 GeV, CL = 0.950

are just an asymmetric way of writing

  - m_t lies in [169.2, 179.4] GeV, CL = 0.683
  - m_ν lies in [0, 18.2] MeV, CL = 0.950
  - m_H lies in [114.3, ∞) GeV, CL = 0.950

Introduction - v

- The goal of these lectures is to explain the precise meaning of statements of the form

  θ lies in [L, U], with CL = β
  - L = lower limit
  - U = upper limit

- For example

  m_t lies in [169.2, 179.4] GeV, with CL = 0.683

What is a Confidence Level?

- A confidence level is a probability that quantifies in some way the reliability of a given statement

- But, what exactly is probability?

  – Bayesian: The degree of belief in, or plausibility of, a statement (Bayes, Laplace, Jeffreys, Jaynes)

  – Frequentist: The relative frequency with which something happens (Boole, Venn, Fisher, Neyman)

Probability: An Example

- Consider the statement
  - S = "It will rain in Morelia on Monday"
- And the probability assignment
  - **Pr**{S} = 0.01
- Bayesian interpretation
  - The plausibility of the statement S is 0.01
- Frequentist interpretation
  - The relative frequency with which it rains on Mondays in Morelia is 0.01

Confidence Level: Interpretation

- Since probability can be interpreted in (at least) two different ways, the interpretation of statements such as

  – m_t lies in [169.2, 179.4] GeV, with CL = 0.683

  depends on which interpretation of probability is being used.

- A great deal of confusion arises in our field because of our tendency to forget this fact

Confidence Intervals

Frequency Interpretation

Confidence Intervals

- The basic idea
  - Imagine a set of ensembles of experiments, each member of which is associated with a fixed value of the parameter to be measured θ (for example, the top quark mass).
  - Each experiment E, within an ensemble, yields an interval $[l(E), u(E)]$, which either contains or does not contain θ.

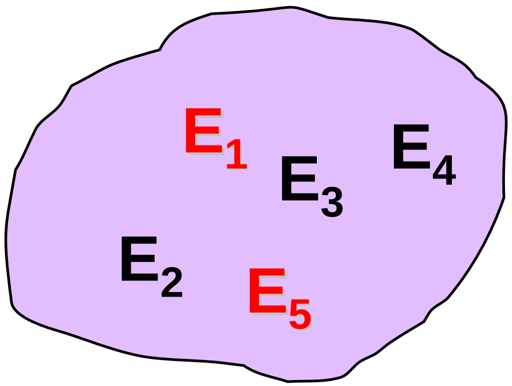
Coverage Probability

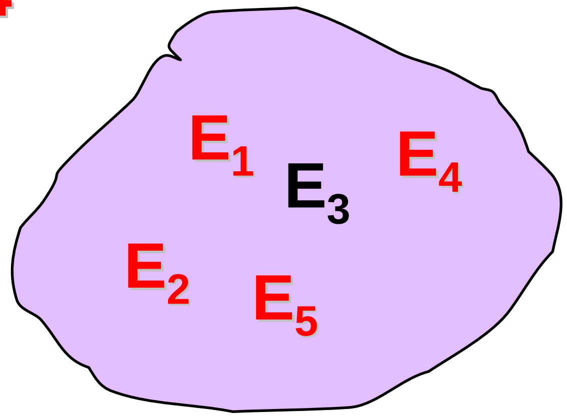
- For a given ensemble, the fraction of experiments with intervals containing the θ value associated with that ensemble is called the coverage probability of the ensemble.

- In general, the coverage probability will vary from one ensemble to another.

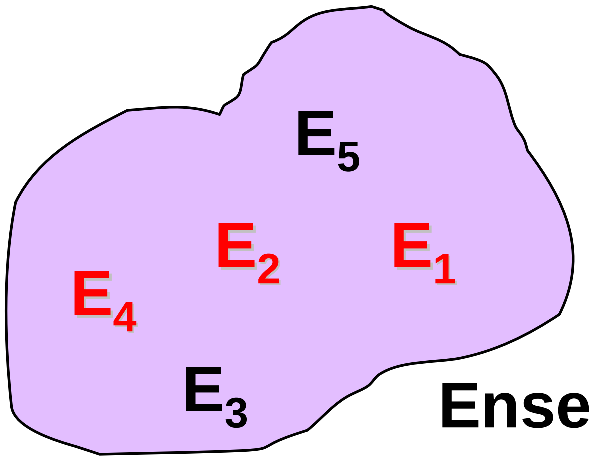
Example

Ensemble with $\theta = \theta_1$ with Pr = 0.4





Ensemble with $\theta = \theta_2$ with Pr = 0.8



Ensemble with $\theta = \theta_3$ with Pr = 0.6

Confidence Level – Frequency Interpretation

– If our experiment is selected at random from the ensemble to which it belongs (presumably the one associated with the true value of θ) then the probability that its interval $[l(E), u(E)]$ contains θ is equal to the coverage probability of that ensemble.

– The crucial point is this: We try to construct the set of ensembles so that the coverage probability over the set is never less than some pre-specified value β, called the confidence level.

Confidence Level - ii

- Points to Note
  - In the frequency interpretation, the confidence level is a property of the set of ensembles; In fact, it is the minimum coverage probability over the set.
  - Consequently, if the set of ensembles is unspecified or unknown the confidence level is undefined.

Confidence Intervals – Formal Definition

Any set of intervals

$$[\, l(E),\ u(E)\,]$$

E	Experiment
l(*E*)	Lower limit
u(*E*)	Upper limit

with a minimum coverage probability equal to β is a set of confidence intervals at 100 β % confidence level (CL). (Neyman,1937)

Confidence intervals are defined not by how they are constructed, but by their frequency properties.

Confidence Intervals: An Example

- Experiment:
  - To measure the mean rate θ of UHECRs above 10^{20} eV per unit solid angle.
- Assume the probability of N events to be a given by a Poisson distribution

$$P(n \mid \theta) = \frac{e^{-\theta}\theta^{n}}{n!}, \quad \theta = < n >, \quad \text{std.dev.} = \sqrt{\theta}$$

Confidence Intervals – Example - ii

- Goal: Compute a set of intervals

$$[l(n), u(n)]$$

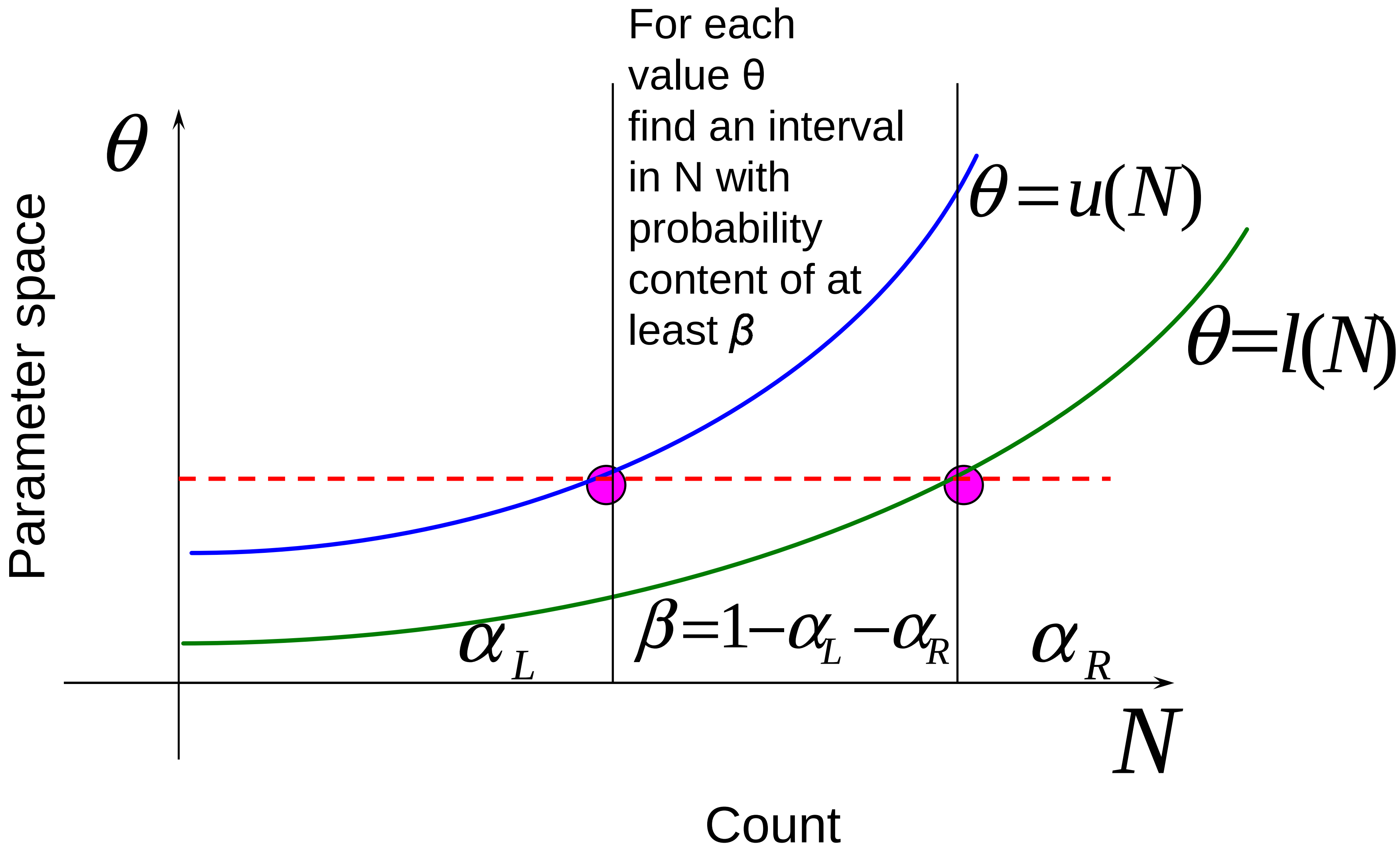
for N = 0, 1, 2, … with CL = 0.683 for a set of ensembles, each member of which is characterized by a different mean event count θ.

Why 68.3%?

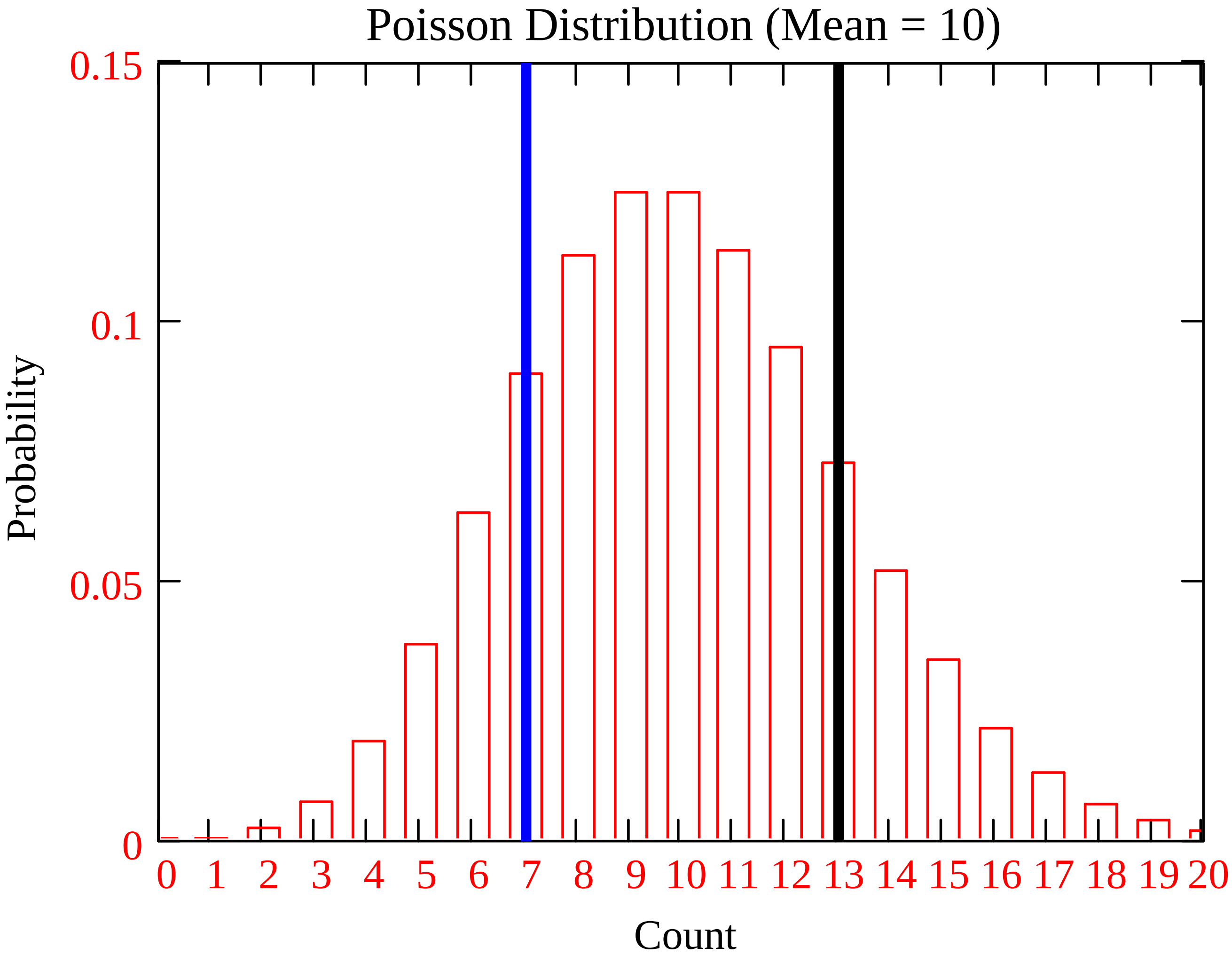
- It is just a useful convention!
  - It comes from the fact that for a Gaussian distribution the confidence intervals given by $[x-\sigma, x+\sigma]$ are associated with a set of ensembles whose confidence level is 0.683. (x = measurement, σ = std. dev.)

- The main reason for this convention is the Central Limit Theorem
  - Most sensible distributions become more and more Gaussian as the data increase.

Confidence Interval – General Algorithm



Example: Interval in N for θ = 10



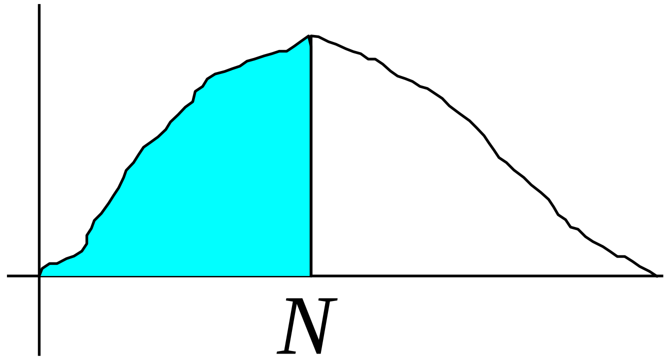
Confidence Intervals – Specific Algorithms

- Neyman
  - Region: fixed probabilities on either side
- Feldman – Cousins
  - Region: containing largest likelihood ratios $P(n|\theta)/P(n|n)$
- Mode – Centered
  - Region: containing largest probabilities $P(n|\theta)$

Neyman Construction

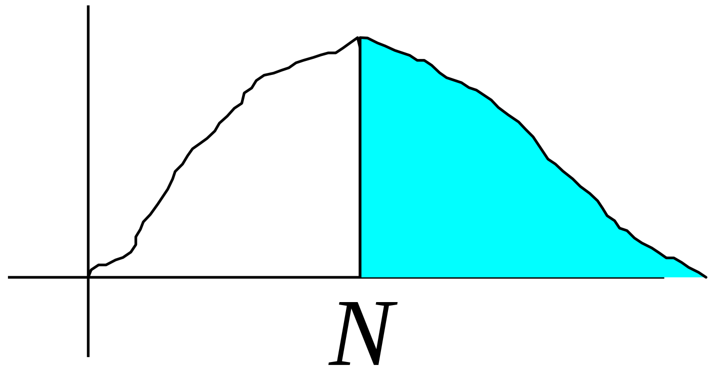
Define

$$C_L(N \mid \theta) = \int_{z \le N} P(z \mid \theta)$$



Left cumulative distribution function

$$C_R(N \mid \theta) = \int_{z \ge N} P(z \mid \theta)$$



Right cumulative distribution function

Valid for both continuous and discrete distributions.

Neyman Construction - ii

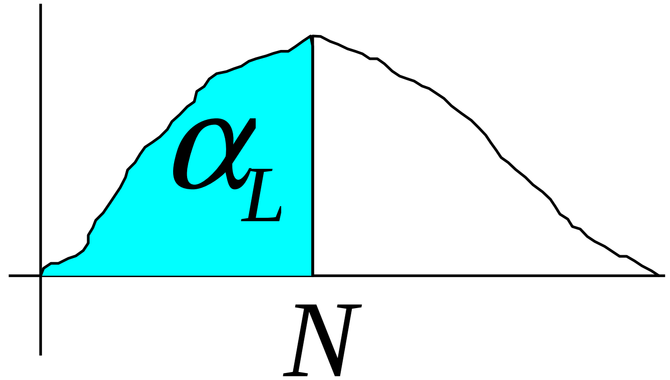
Solve

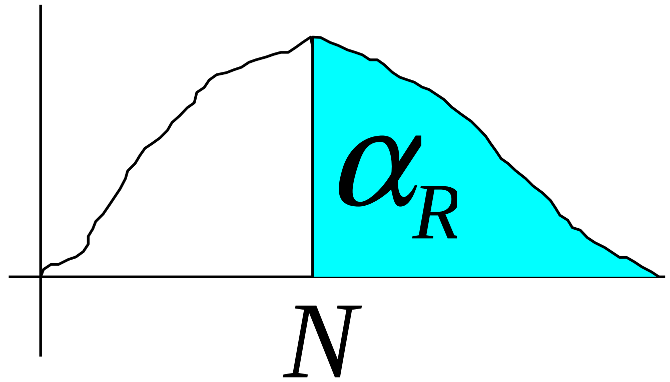
$$C_L(N\,|\,u) = \alpha_L$$

$$C_R(N\,|\,l) = \alpha_R$$

where

$$\beta = 1 - \alpha_L - \alpha_R$$





Remember: Left is UP and Right is LOW!

Central Confidence Intervals

Choose $\alpha_L = \alpha_R = (1 - 0.683) / 2$

and solve

$$\alpha_L = \sum_{i=0}^{n} P(i \mid u)$$

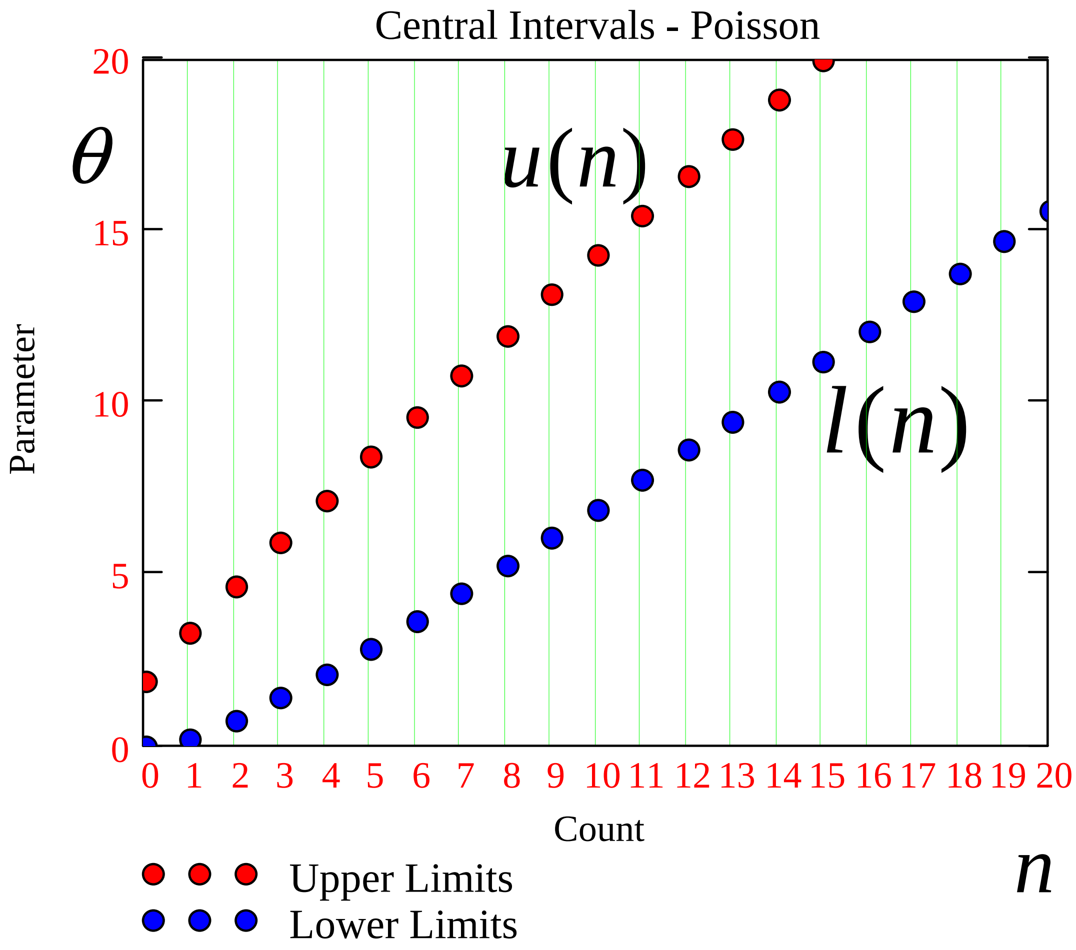
$$\alpha_R = \sum_{i=n}^{\infty} P(i \mid l) = 1 - \sum_{i=0}^{n-1} P(i \mid l)$$

for the interval $[l(n), u(n)]$

Central Confidence Intervals - ii

Poisson Distribution

$$P(n\,|\,\theta) = \frac{e^{-\theta}\theta^{n}}{n!}$$



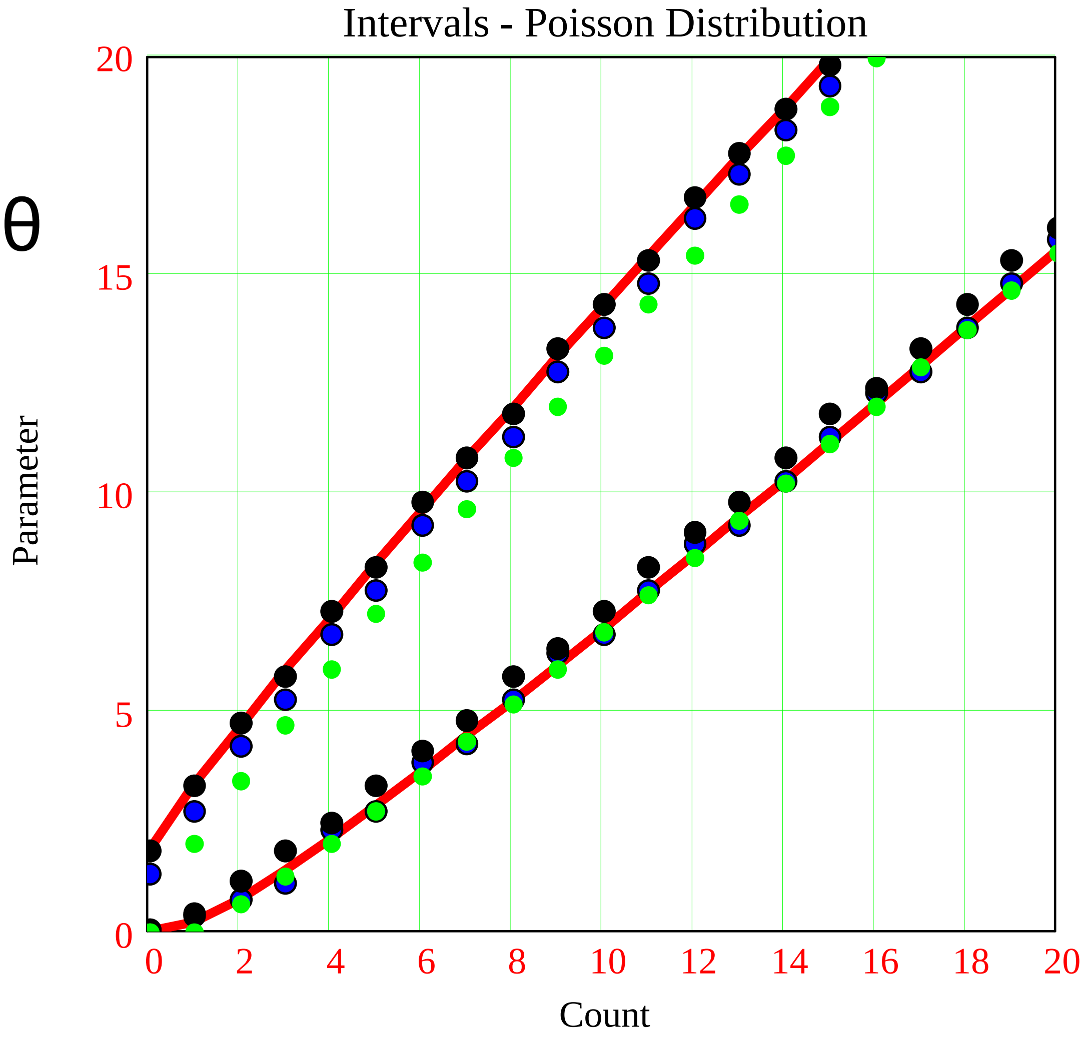
Comparison of Confidence Intervals

Central

Feldman-Cousins

Mode-Centered

N±√N



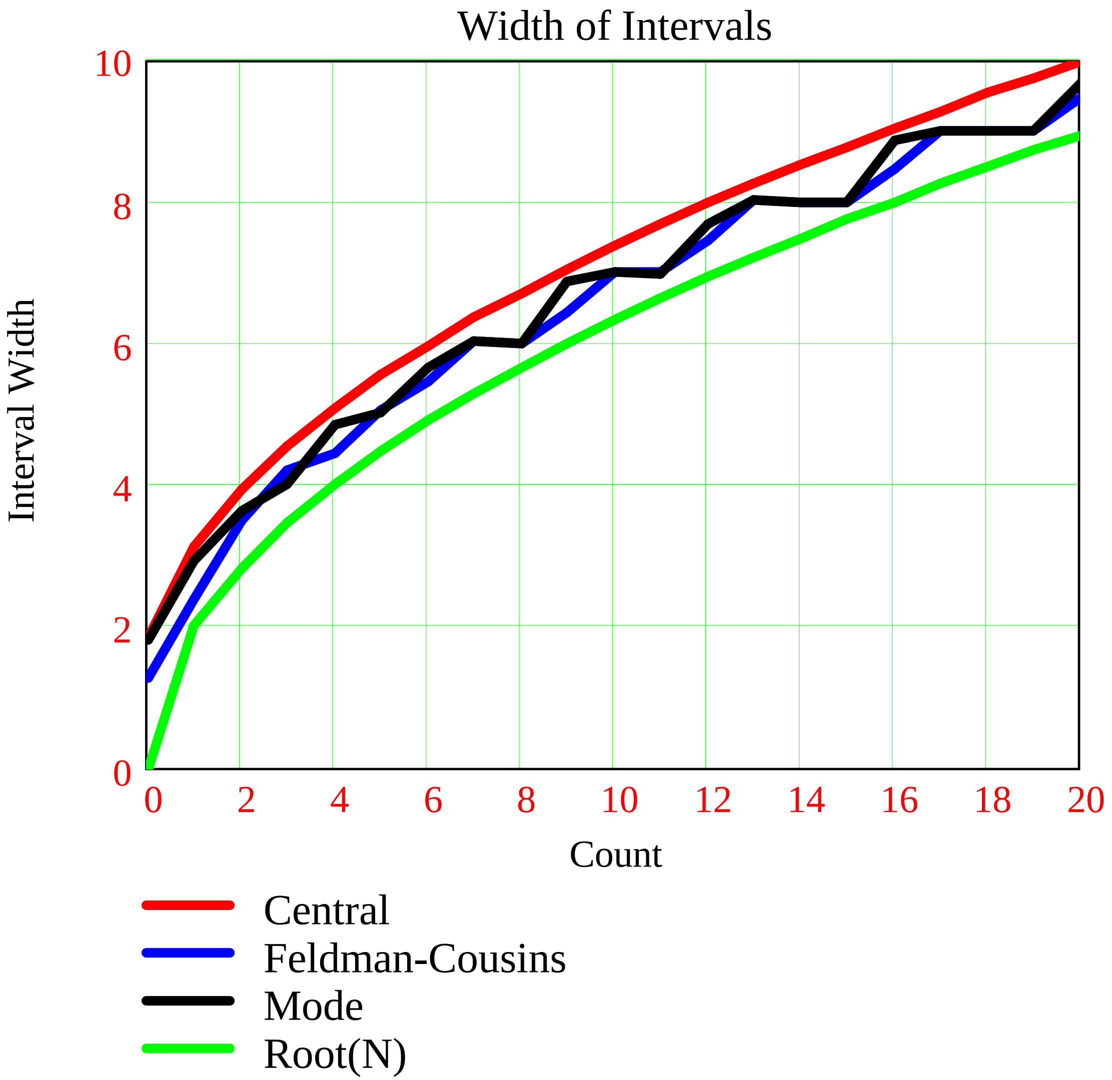
Comparison of Confidence Interval Widths

Central

Feldman-Cousins

Mode-Centered

N±√N



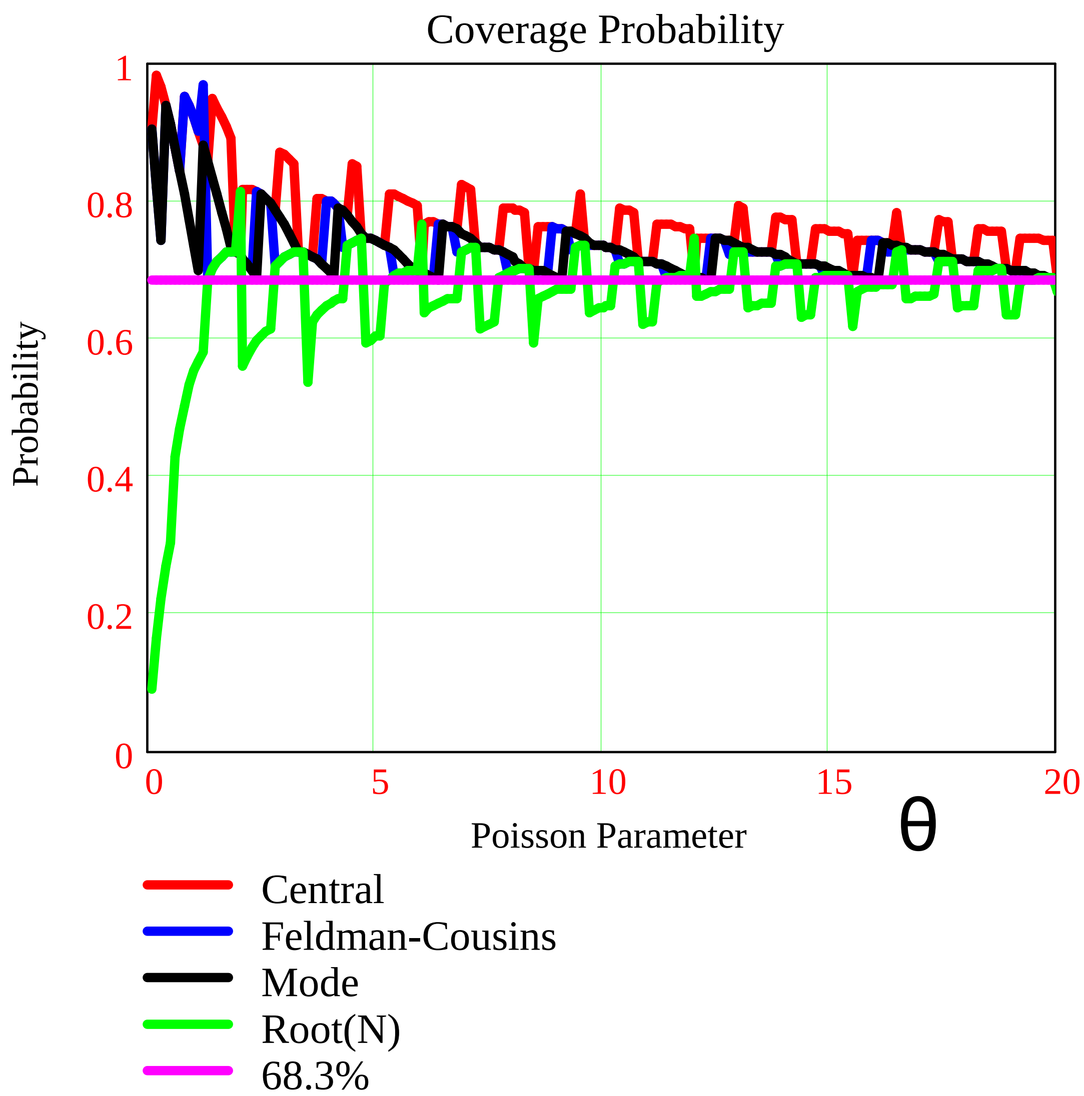
Comparison of Coverage Probabilities

Central

Feldman-Cousins

Mode-Centered

N±√N



Summary

- The interpretation of confidence intervals and confidence levels depends on which interpretation of probability one is using

- The coverage probability of an ensemble of experiments is the fraction of experiments that produce intervals containing the value of the parameter associated with that ensemble

- The confidence level is the minimum coverage probability over a set of ensembles.

- The confidence level is undefined if the set of ensembles is unspecified or unknown