Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

MAX-PLANCK-GESELLSCHAFT

Frequentist Confidence Intervals and Bayesian Credibility Intervals

Allen Caldwell

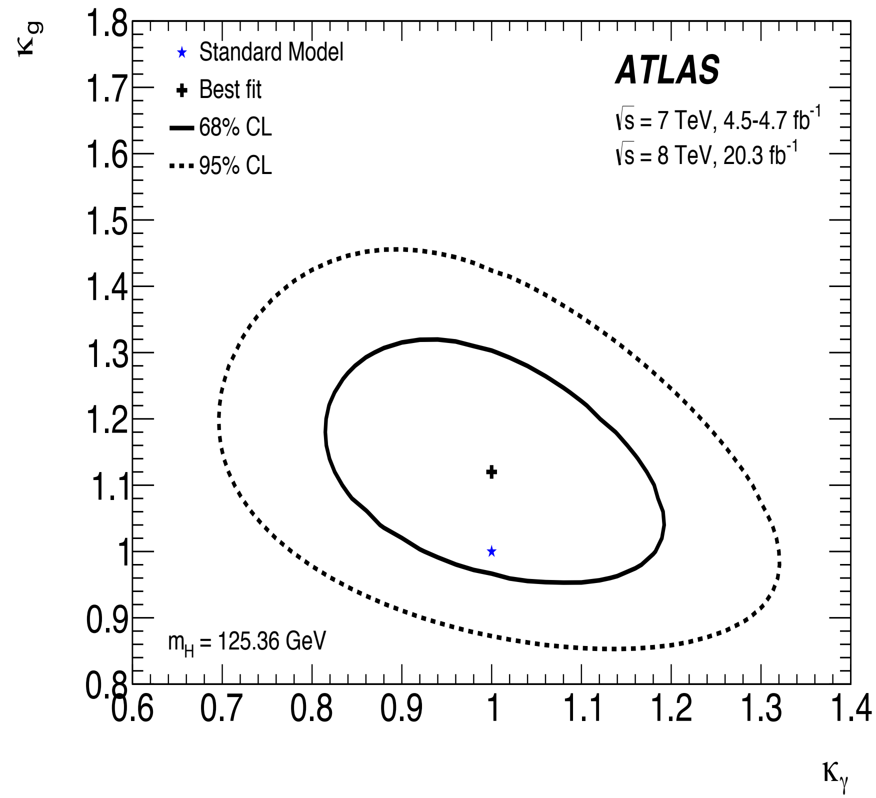
February 19, 2016

Observation of Gravitational Waves

B. P. Abb

(LIGO Scientific Collaborati

(Received 21 January 2016;



On September 14, 2015 at 09:50:45 UTC the two de Observatory simultaneously observed a transient grav frequency from 35 to 250 Hz with a peak gravitational predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than 5.1σ. The source lies at a luminosity distance of 410^{+160}_{-180} Mpc corresponding to a redshift $z = 0.09^{+0.03}_{-0.04}$. In the source frame, the initial black hole masses are $36^{+5}_{-4}M_\odot$ and $29^{+4}_{-4}M_\odot$, and the final black hole mass is $62^{+4}_{-4}M_\odot$, with $3.0^{+0.5}_{-0.5}M_\odot c^2$ radiated in gravitational waves. All uncertainties define 90% credible intervals. These observations demonstrate the existence of binary stellar-mass black hole systems. This is the first direct detection of gravitational waves and the first observation of a binary black hole merger.

$$P(\lambda|D,M) = \frac{P(D|\lambda,M)P(\lambda|M)}{P(D|M)}$$

Bayes Equation

In frequentist analysis, only use $P(D|\lambda,M)$

Statements are constrained to the form:

For parameter values in the range $\lambda \in [\lambda_1, \lambda_2]$

The **DATA** satisfies some probability criteria

Can lead to empty intervals

In Bayesian analysis, use $P(\lambda|D,M)$

Statements are of the form:

The **PARAMETER** values in the range $\lambda \in [\lambda_1, \lambda_2]$

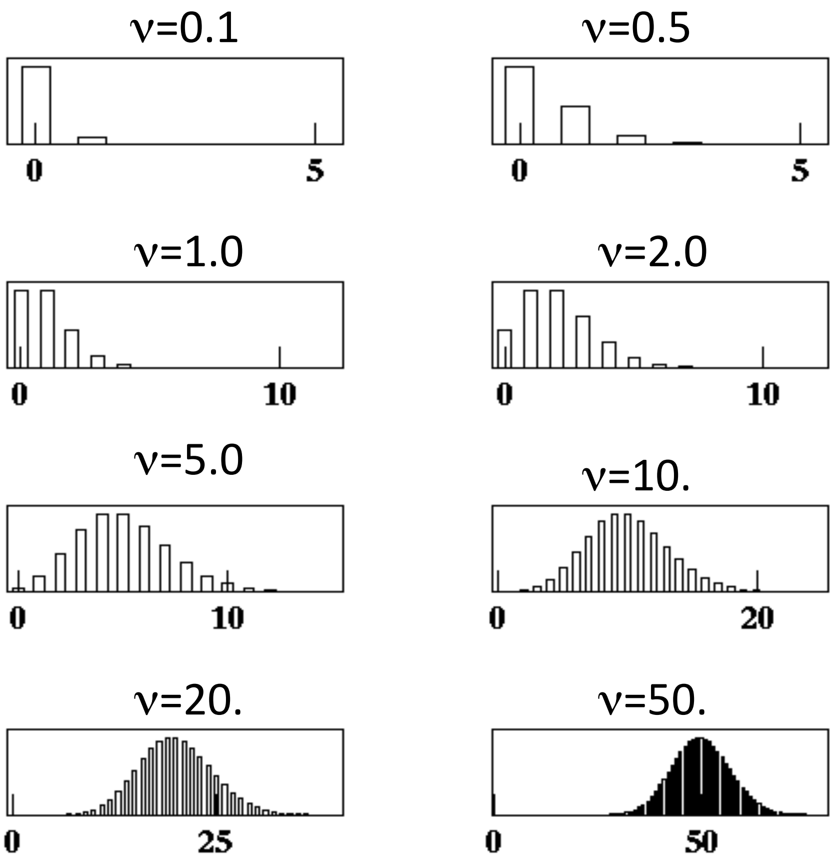
satisfy some probability criteria

Cannot lead to empty intervals (no goodness-of-fit)

Poisson Distribution

$$\boxed{P(n \mid \nu) = \frac{\nu^n e^{-\nu}}{n!}}$$

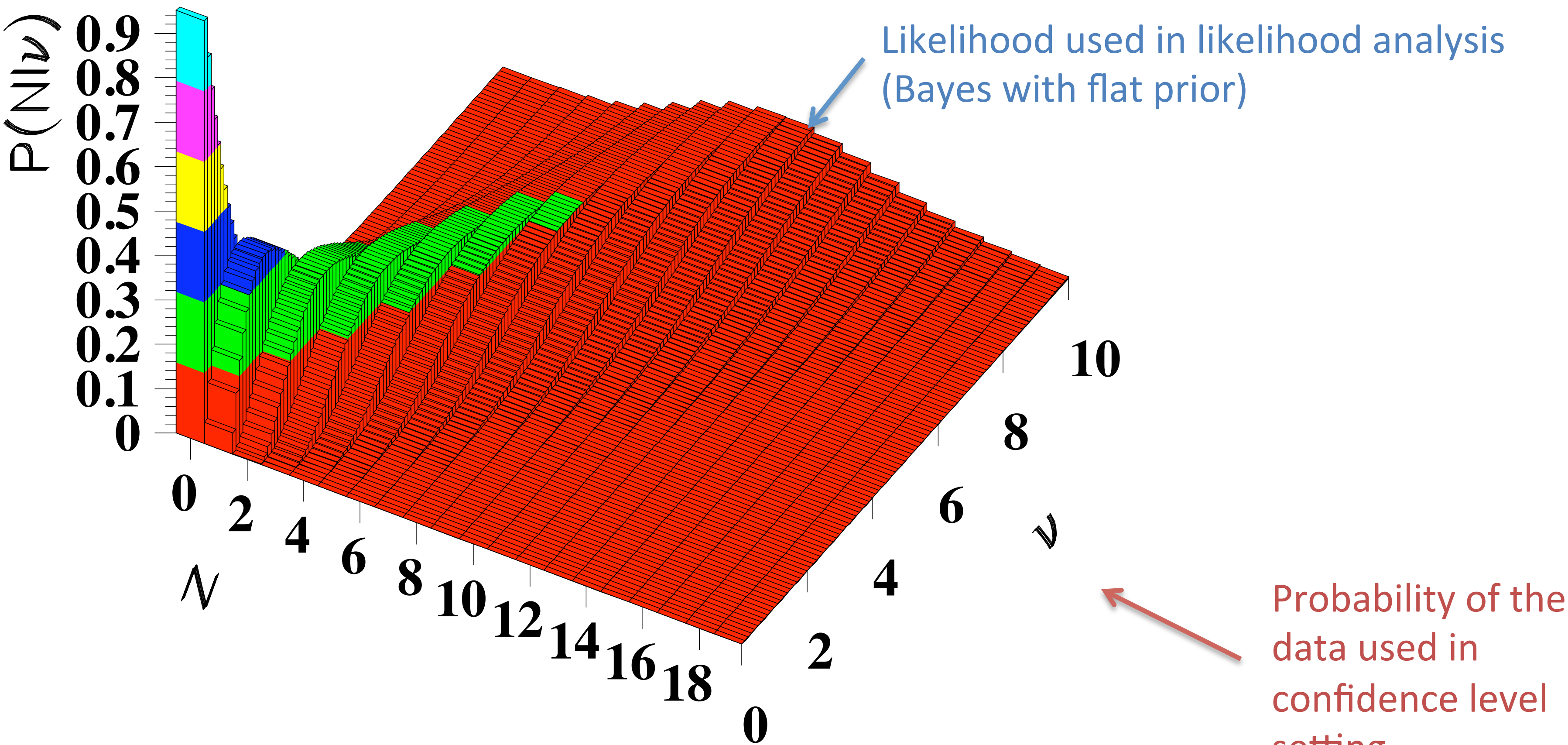
Applies to processes with constant rate, large number of trials & small probability.



Notes:

- As ν increases, the distribution becomes more symmetric
- Approximately Gaussian for large ν
- Poisson formula is much easier to use that the Binomial formula.

Poisson Example



Frequentist Confidence Level Intervals

Neyman construction – for each value of the parameter(s), find set of possible outcomes that contain at least 1-α probability. For the central interval and Poisson distribution:

$$n_1 = \sup_{n \in 0,...,\infty} \left\{ \sum_{i=0}^{n} P(i|\nu) \leq \alpha/2 \right\} + 1$$

$$P(n=0|\nu) > \alpha/2 \rightarrow n_1 = 0$$

$$n_2 = \inf_{n \in 0,...,\infty} \left\{ \sum_{i=n}^{\infty} P(i|\nu) \leq \alpha/2 \right\} - 1$$

$$\mathcal{O}^C_{1-\alpha} = \{n_1, \ldots, n_2\}$$

Set contains at least 1-α probability of DATA outcomes for a given expectation. Central interval – balance probability on either side as evenly as possible.

Poisson Example

Example for ν=10/3

E.g., Take 1-α =0.9   α /2=0.05

$$\mathcal{O}^C_{1-\alpha} = \{n_1, \ldots, n_2\}$$

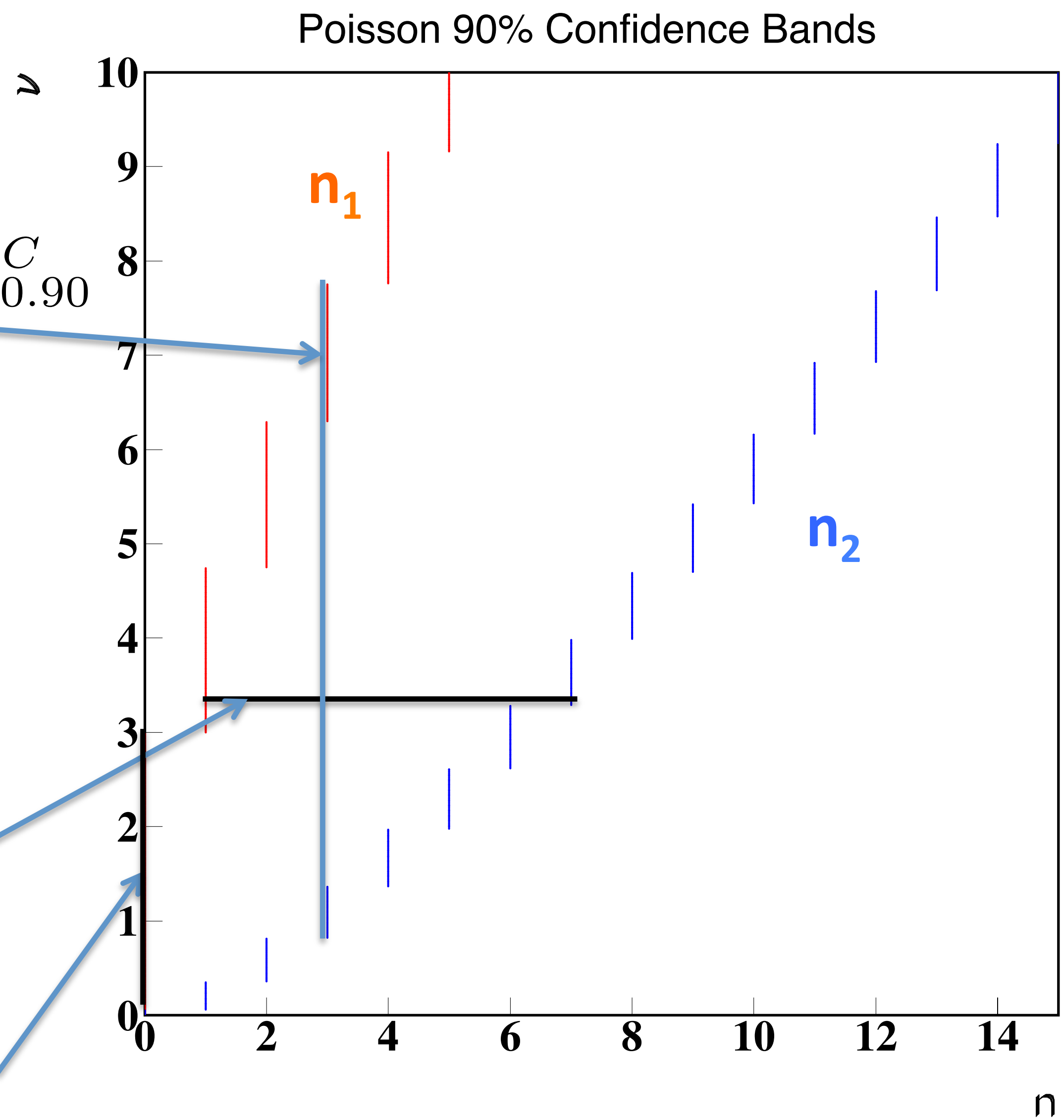
$$\mathcal{O}^C_{0.9} = \{1, 2, 3, 4, 5, 6, 7\}$$

$$\mathcal{P}^C_{0.9} = 0.9435$$

n	$P(n\|\nu)$	$F(n\|\nu)$	R	$F_R(n\|\nu)$
0	0.0357	0.0357	7	0.9468
1	0.1189	0.1546	5	0.8431
2	0.1982	0.3528	2	0.4184
3	0.2202	0.5730	1	0.2202
4	0.1835	0.7565	3	0.6019
5	0.1223	0.8788	4	0.7242
6	0.0680	0.9468	6	0.9111
7	0.0324	0.9792	8	0.9792
8	0.0135	0.9927	9	0.9927
9	0.0050	0.9976	10	0.9976
10	0.0017	0.9993	11	0.9993
11	0.0005	0.9998	12	0.9998
12	0.0001	1.0000	13	1.0000

Now repeat for all interesting values of parameter.

Neyman Construction – Central Interval



e.g., n=3

Values of ν for which $n \in \mathcal{O}^{C}_{0.90}$

$\nu \in [0.7, 7.7]$

Allowed n for $\nu = 3.\bar{3}$

$n = 0\ \ 95\ \%\ \mathrm{CL},\ \ \nu \in [0., 3]$

Central Probability Interval

Confidence Level Calculation

$$n \in \mathcal{O}^C_{1-\alpha}$$

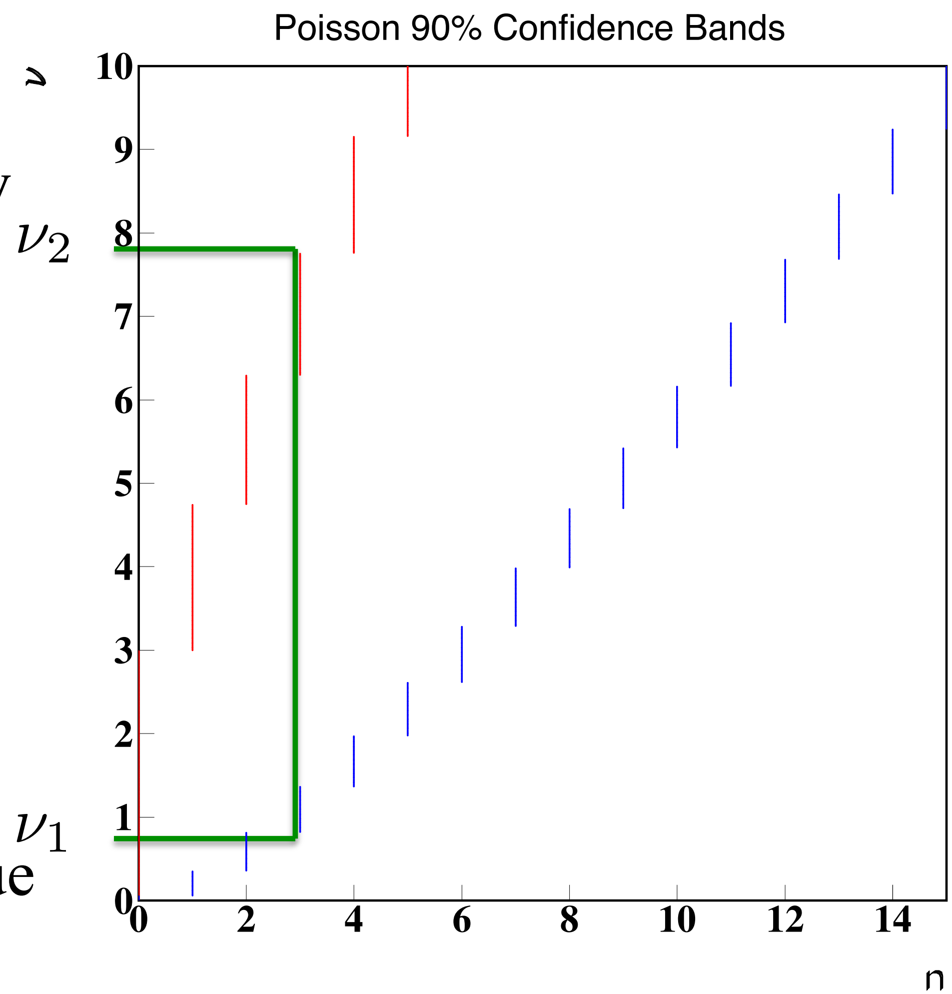
Coverage: Given the true value, the observed n will be included in $\mathcal{O}^C_{1-\alpha} = \{n_1, \ldots, n_2\}$ in at least 1-α of experiments by construction.

So, for given observed n, true value with be contained within

$$\{\nu_1, \nu_2\}$$

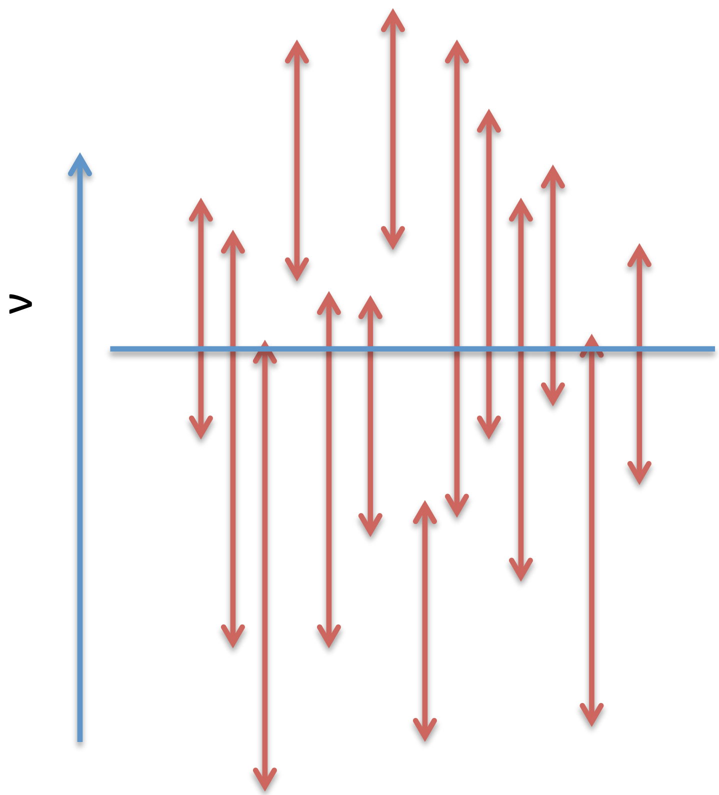
in at least 1-α of experiments (coverage).

No statement about the true value in any single experiment …



Use of Frequentist CL Intervals

Not intended to be used individually. Rather, collect a lot of intervals and use these to `find' the true value (not specified how).



True value should be the one that is in the set of intervals the right fraction of the time (e.g., 68% of the time for the 68% CL intervals).

Smallest Interval

The CL intervals will depend on the probability criteria chosen. E.g., we can choose the smallest interval containing a given probability:

Rank outcomes according to probability.

$$\mathcal{O}^S_{0.9} = [1, 2, 3, 4, 5, 6]$$

Recall:

$$\mathcal{O}^C_{0.9} = [1, 2, 3, 4, 5, 6, 7]$$

So – need to specify criterion of selection with %CL specification.

n	$P(n\|\nu)$	$F(n\|\nu)$	R	$F_R(n\|\nu)$
0	0.0357	0.0357	7	0.9468
1	0.1189	0.1546	5	0.8431
2	0.1982	0.3528	2	0.4184
3	0.2202	0.5730	1	0.2202
4	0.1835	0.7565	3	0.6019
5	0.1223	0.8788	4	0.7242
6	0.0680	0.9468	6	0.9111
7	0.0324	0.9792	8	0.9792
8	0.0135	0.9927	9	0.9927
9	0.0050	0.9976	10	0.9976
10	0.0017	0.9993	11	0.9993
11	0.0005	0.9998	12	0.9998
12	0.0001	1.0000	13	1.0000

Bayesian Data Analysis-Poisson Distribution

Result will depend on choice of prior. If we assume a flat prior starting at 0 and extending up to some maximum of ν much larger than *n*.

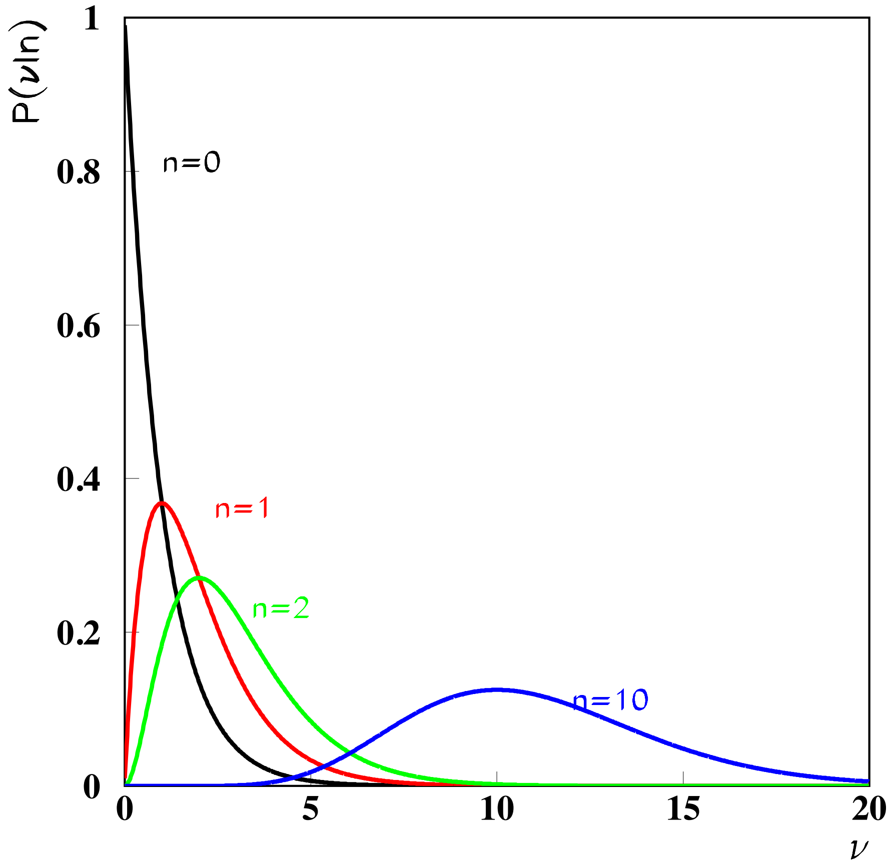
$$P(\nu|n) = \frac{\frac{\nu^n e^{-\nu}}{n!} P_0(\nu)}{\int_0^\infty \frac{\nu^n e^{-\nu}}{n!} P_0(\nu) d\nu} = \frac{\frac{\nu^n e^{-\nu}}{n!}}{\int_0^{\nu_{max}} \frac{\nu^n e^{-\nu}}{n!} d\nu}$$

$$\int_0^{\nu_{max}} \frac{\nu^n e^{-\nu}}{n!} d\nu \approx \frac{1}{n!} \int_0^\infty \nu^n e^{-\nu} d\nu = \frac{1}{n!} n! = 1$$

$$P(\nu|n) = \frac{e^{-\nu}\nu^n}{n!} \qquad \nu^* = n$$

Poisson – cont.

Some examples



Comments:

If you decide to quote the mode as your nominal result, you would use $\nu^*=n$. For large enough n, the 68% probability region is then approximately

$$n - \sqrt{n} \rightarrow n + \sqrt{n}$$

Poisson - cont.

The cumulative distribution function:

$$
\begin{aligned}
F(\nu|n) &= \int_0^{\nu} \frac{\nu'^n e^{-\nu'}}{n!} d\nu' \\
&= \frac{1}{n!}\left[-\nu'^n e^{-\nu'}|_0^{\nu} + n\int_0^{\nu} \nu'^{n-1} e^{-\nu'} d\nu'\right] \\
&= 1 - e^{-\nu}\sum_{i=0}^{n}\frac{\nu^i}{i!}
\end{aligned}
$$

e.g.,

$$n = 3 \;\; \nu \in [0.94, 6.94]$$
$$90\ \%\ CI \text{ assuming flat prior}$$

Smallest credible interval

$$n = 0 \;\; \nu \in [0, 3]$$
$$95\ \%\ CI \text{ assuming flat prior}$$

Numerically identical to 95% CL upper limit

Poisson – signal+background

If we have a background (which we always do ...), then we have a total expectation

$$\mu = \nu + \lambda$$

where λ is the background expectation. We initially assume it is known.

The number of events is expected to follow a Poisson probability distribution

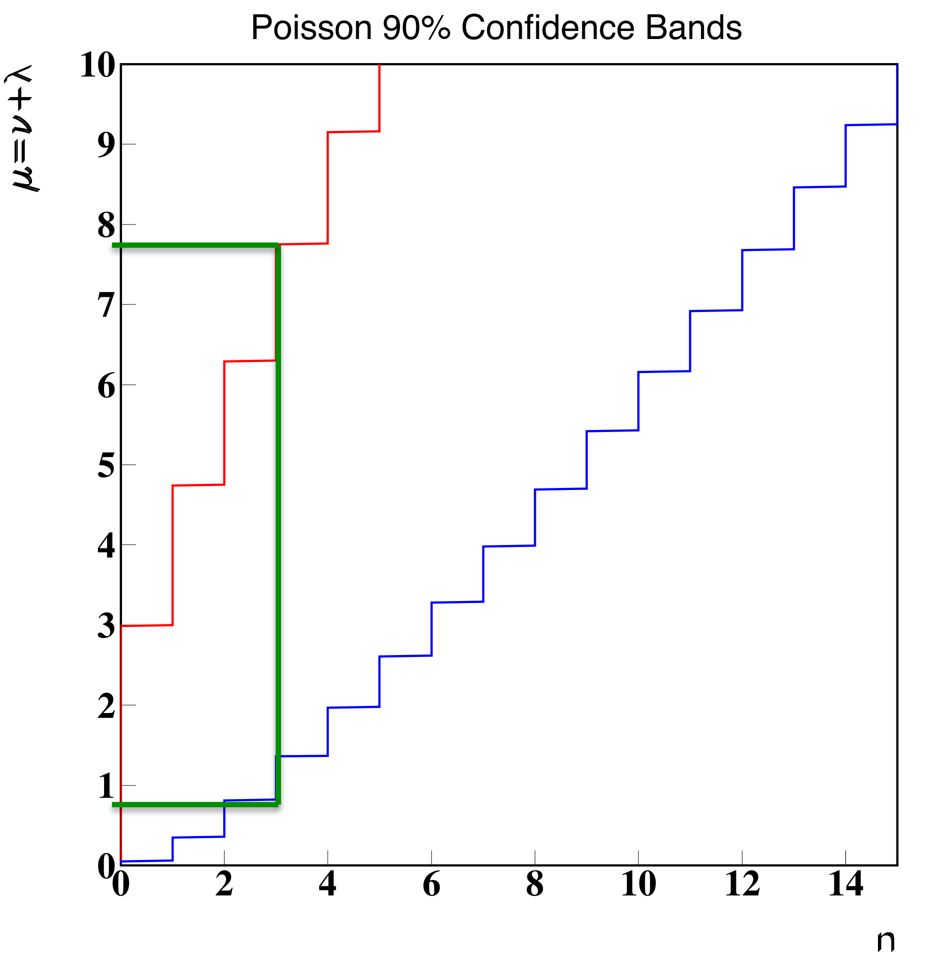
$$P(n|\nu, \lambda) = \frac{\mu^n}{n!} e^{-\mu}$$

The Neyman band plot is unchanged, and we find CL ranges for μ. We then translate to a range on ν with

$$\nu \in [\mu_1 - \lambda, \mu_2 - \lambda]$$

Frequentist Statistics

Poisson distribution in the presence of background, with mean λ. Then we have the same bands as for signal only, but replace ν with (ν+λ).



$$n = 3, \lambda = 3.0, \mu \in [0.7, 7.7]$$

$$\nu \in [0, 4.7]@90\ \%CL$$

For $n = 0$ and $\lambda > 3$
empty interval for ν

Discussion

- Procedure is perfectly well defined

- Can result in unphysical values for the signal parameter or empty intervals. There can be data outcomes for which no values of the signal parameter satisfy our **probability criterion on the data.**

- This is not a problem in itself, but given the tendency to misinterpret CL intervals as statements about parameter values, this can lead to confusing results. E.g., higher background levels can be advantageous for getting stronger limits. Many examples in S. Biller & S. Oser, Nucl.Instrum.Meth. A774 (2015) 103-119 .

- Has led to new probability criteria for the data that do not produce empty intervals.

Feldman-Cousins Confidence Levels

The most popular (at least in particle physics) is the Feldman-Cousins construction, where a rank is assigned to possible outcomes based on

$$r = \frac{P(n|\mu = \lambda + \nu)}{P(n|\hat{\mu})}$$

Where $\hat{\mu}$ is the value of μ that maximizes $P(n|\mu)$ given the constraints. Example constraint $\mu \geq \lambda$

Gary J. Feldman and Robert D. Cousins, A Unified Approach to the Classical Statistical Analysis of Small Signals, Phys.Rev.D57:3873-3889,1998

Concrete example: $\lambda = 3.0 \quad \nu = 0.\bar{3}$

n	$P(n\|\overset{\mu}{\cancel{\nu}})$	$\hat{\mu}$	$P(n\|\hat{\mu})$	r	Rank	$F_R(n\|\nu)$
0	0.0357	3.0	0.050	0.717	5	0.7565
1	0.1189	3.0	0.149	0.796	4	0.7208
2	0.1982	3.0	0.224	0.885	3	0.6091
3	0.2202	3.0	0.224	0.983	1	0.2202
4	0.1835	4.0	0.195	0.941	2	0.4037
5	0.1223	5.0	0.175	0.699	6	0.8788
6	0.0680	6.0	0.161	0.422	7	0.9468
7	0.0324	7.0	0.149	0.217	8	0.9792
8	0.0135	8.0	0.140	0.096	9	0.9927
9	0.0050	9.0	0.132	0.038	10	0.9976
10	0.0017	10.0	0.125	0.014	11	0.9993
11	0.0005	11.0	0.119	0.004	12	0.9998

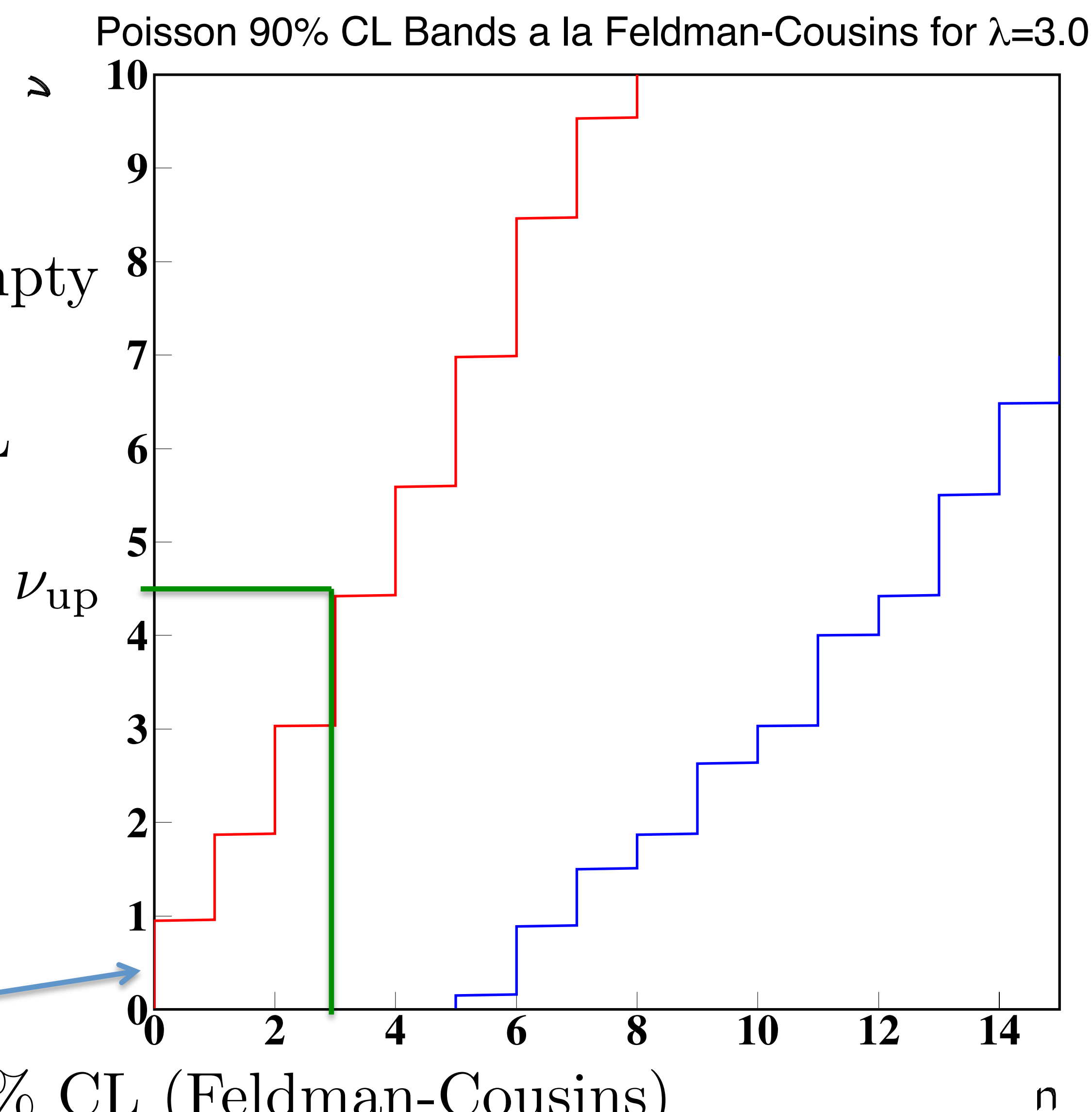
e.g., $\mathcal{O}^{FC}_{0.9} = \{0, \ldots, 6\}$

Similar to Neyman construction, but use Rank to add elements to set until reach at least 1-α probability content of possible results.

interval for ν not empty

$n = 3 \quad \lambda = 3.0$

$\nu \leq 4.4$ @ 90 % CL



$n = 0$, $\nu \leq 1$ @ 95 % CL (Feldman-Cousins)

Bayes Analysis for Poisson Data

$$\mu = \lambda + \nu \qquad P(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$$

Assuming that the background is perfectly known:

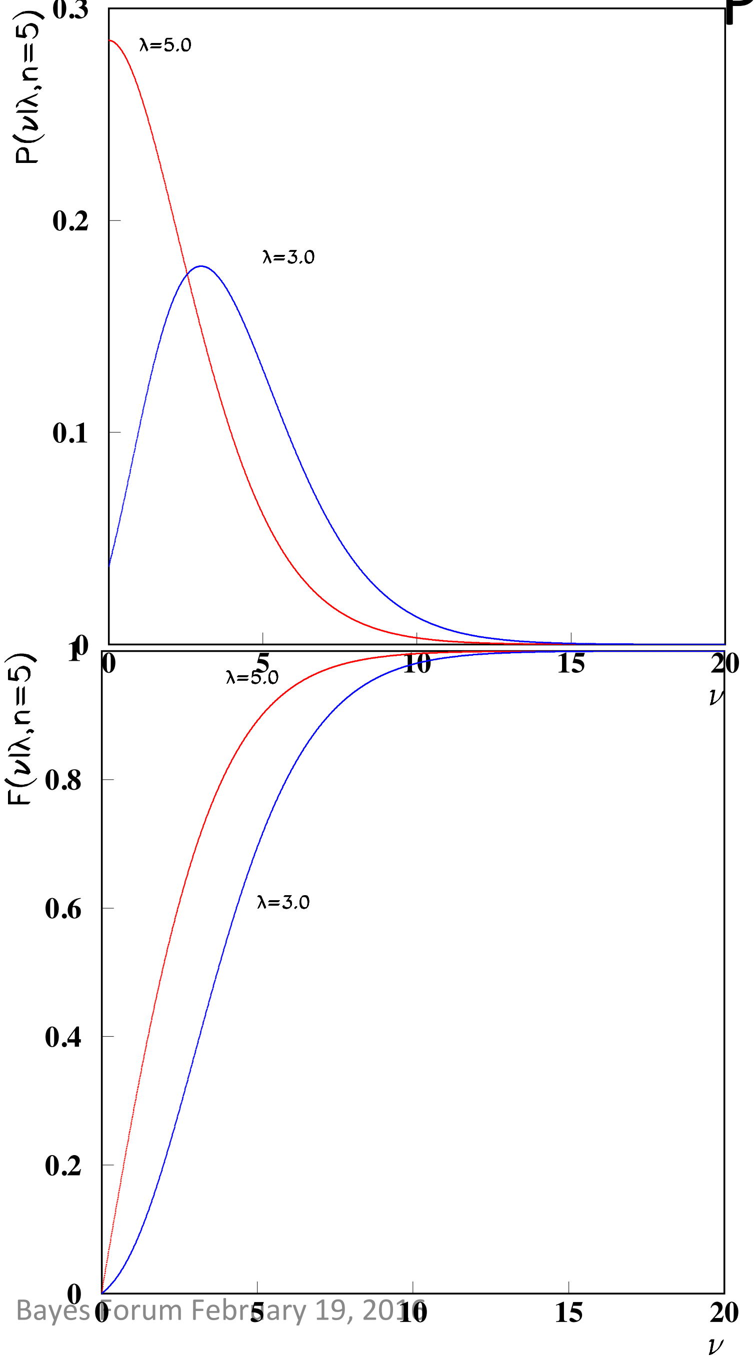
$$P(\nu|n,\lambda) = \frac{P(n|\nu,\lambda)P_0(\nu)}{\int P(n|\nu,\lambda)P_0(\nu)d\nu}$$

assuming a flat $P_0(\nu)$ and integrating by parts.

$$\boxed{P(\nu|n,\lambda) = \frac{e^{-\nu}(\lambda+\nu)^n}{n!\sum_{i=0}^{n}\frac{\lambda^i}{i!}}}$$

The cumulative pdf is $\boxed{F(\nu|n,\lambda) = 1 - \frac{e^{-\nu}\sum_{i=0}^{n}\frac{(\lambda+\nu)^i}{i!}}{\sum_{i=0}^{n}\frac{\lambda^i}{i!}}}$

Poisson – cont.



Posterior probability density for v with n = 5 events observed and two different known background expectations (top). The cumulative of the posterior probability density (bottom).

Comment:

For n=0, $P(\nu|n, \lambda)=e^{-\nu}$. It does not matter how much background you have, you get the same probability distribution for the signal.

Comparing Feldman-Cousins with Bayesian Analysis with same background $\lambda = 3.0$ and a flat prior.

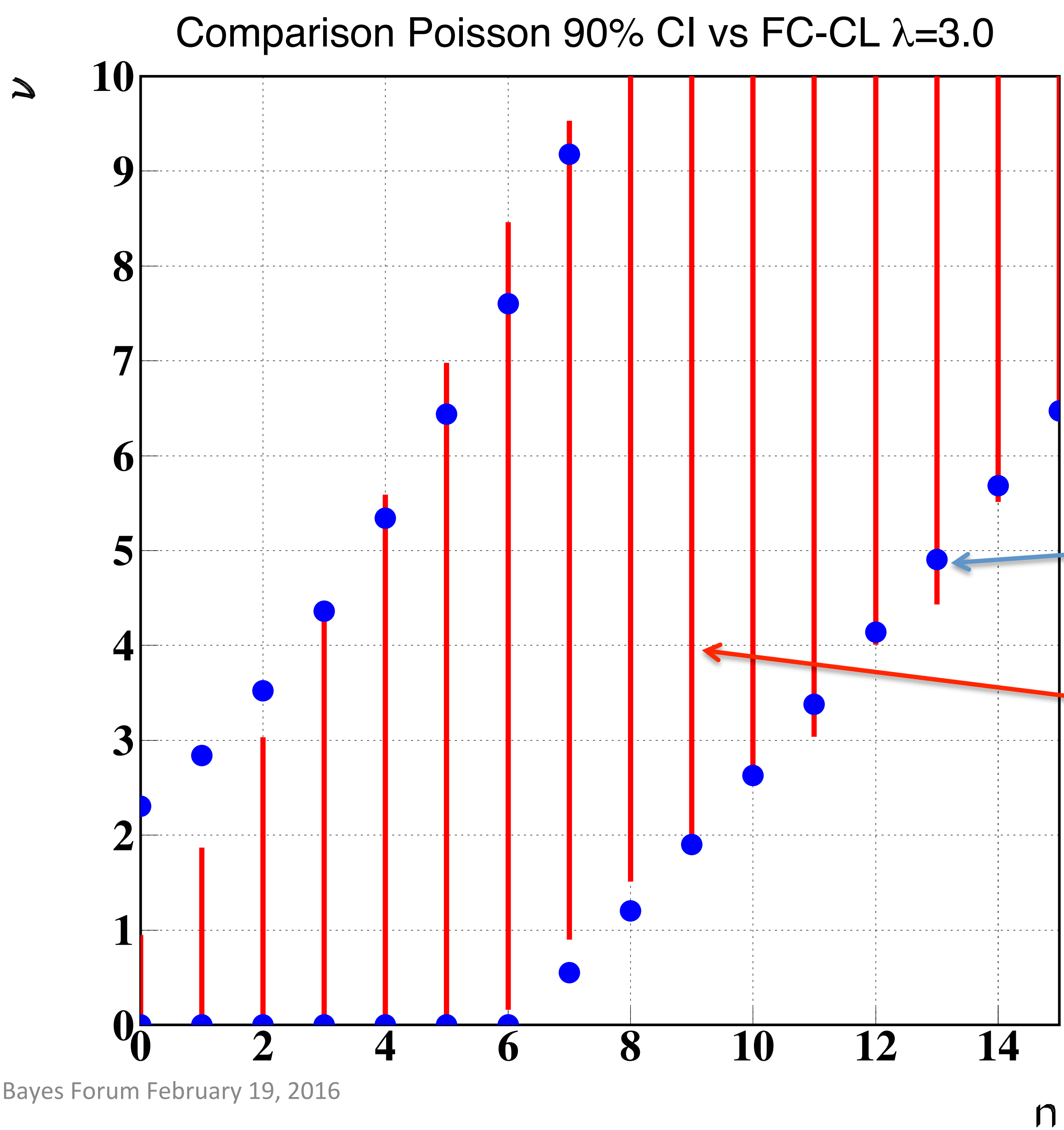
Recall: $P(\nu|n,\lambda) = \dfrac{e^{-\nu}(\lambda+\nu)^n}{n!\sum_{i=0}^{n}\frac{\lambda^i}{i!}}$

$$F(\nu|n,\lambda) = 1 - \frac{e^{-\nu}\sum_{i=0}^{n}\frac{(\lambda+\nu)^i}{i!}}{\sum_{i=0}^{n}\frac{\lambda^i}{i!}}$$

We will take the smallest interval with 90% credibility. I.e.,

$$\int_{P>C} P(\nu|n,\lambda)d\nu = 0.90$$

We find ν_{down} ν_{up} fulfilling this condition. Numerical integration.



Numerically similar large n, but big differences for $n < \lambda$

Bayes: flat prior, smallest interval

Feldman-Cousins

Bayesian Limit vs Feldman-Cousin

Main difference is for number of observed events < expected background. In the Bayesian approach, use the fact that the number of background events cannot be larger than the number of observed events:

$$F(\nu|n,\lambda) = 1 - \frac{e^{-\nu}\sum_{i=0}^{n}\frac{(\lambda+\nu)^i}{i!}}{\sum_{i=0}^{n}\frac{\lambda^i}{i!}}$$

Frequentist limits often `look stronger` because of the behavior for small numbers of events. Limits on parameter are produced in region where the experiment has no sensitivity.

S. Biller & S. Oser, Nucl.Instrum.Meth. A774 (2015) 103-119 .

The EXO collaboration published first results from the EXO-200 neutrinoless double beta-decay experiment in 2012 [13]. In the $\pm 1\sigma$ energy resolution window around the endpoint, 1 event was observed where a background of 4.1±0.3 counts was expected. Using a spectrum fit, the authors derived a bound of <2.8 total signal counts at the 90% CL, corresponding to a lower bound to the half-life for $0\nu\beta\beta$ of 1.6×10^{25} years. A Feldman-Cousins bound based on the $\pm 1\sigma$ bin would have yielded a limit of less than 2.0 signal counts at 90% CL (accounting for the 68% signal efficiency of the bin), correspond even more restrictive 90% CL lower bound for life due to the negative fluctuation of 2.2×10^{25} y the other hand, a Bayesian bound with a prior u counting rate based on the $\pm 1\sigma$ bin would hav a limit of less than 4.0 signal counts, correspon 90% CI lower bound to the half-life of only 1.1×1 — seemingly less restrictive than the Feldmar bound by a factor of two.

EXO: First run

$$\mathrm{FC}: \lambda = 4.1,\ n = 1,\ \nu < 2.$$

$$T_{1/2} > 2.2 \cdot 10^{25}\ yr$$

$$\text{Bayes: } T_{1/2} > 1.1 \cdot 10^{25}\ \text{yr}$$

centuates this effect [14]. Within the $\pm 1\sigma$ energy resolution window around the endpoint, 21 events were observed where a background of 16±2 counts was expected [15]. All approaches derive similar bounds for this case: the authors derived a 1-sided 90% CL lower bound to the $0\nu\beta\beta$ half-life of $t_{1/2} > 1.1 \times 10^{25}$ years by applying Wilks' Theorem to a likelihood analysis, which is a factor of 1.45 less restrictive than the initial result. A Feldman-Cousins analysis based on the $\pm 1\sigma$ bin would yield a bound of $t_{1/2} > 1.3 \times 10^{25}$ years, a factor of 1.7 less restrictive than the F-C bound from the initial result. However, a Bayesian bound, with a prior uniform in counting rate and based on the same bin yields a value of $t_{1/2} > 1.4 \times 10^{25}$ years or, using the appropriate integration of the posterior probability derived from the provided likelihood curve assuming a uniform prior, a value of $t_{1/2} > 1.2 \times 10^{25}$ years. Both Bayesian calculations are modestly *more* restrictive than the initial Bayesian

EXO: Second run (4x exposure)

$$\mathrm{FC}: \lambda = 16,\ n = 21,$$

$$T_{1/2} > 1.3 \cdot 10^{25}\ yr$$

$$\text{Bayes: } T_{1/2} > 1.2 \cdot 10^{25}\ \text{yr}$$

CL_s Method, Profile Likelihood

There is some unhappiness in the community because the F-C method can produce intervals/exclusions where the experiment has no sensitivity. As with Neyman, not really a problem, except that people interpret the CL Intervals as statements about the model/parameters.

A. L. Read, ***Presentation of search results: the CL_S technique***, J. Phys. G: Nucl. Part. Phys. **28** (2002) 2693-2704.
nearly all physicists tend to misinterpret frequentist results as statements about the theory given the data.

Frequentist statements are not statements about the model – only about the data in the context of the model. This is not what we wanted to know … at least not the ultimate statement. Leads naturally to misunderstandings.

Discussion similar for more complicated situations (multiple parameters, other types of distributions, …)

Summary:

- Frequentist CL intervals provide a range of parameter values for which the data has satisfied certain probability requirements. Range depends on criteria, so they should be specified. Not a statement about probable values of the parameters.

- Bayesian CI provide a range of parameter values for which the parameters have satisfied certain probability requirements. Range depends on priors as well as interval definition, so they should be specified.