Bayesian Data Analysis

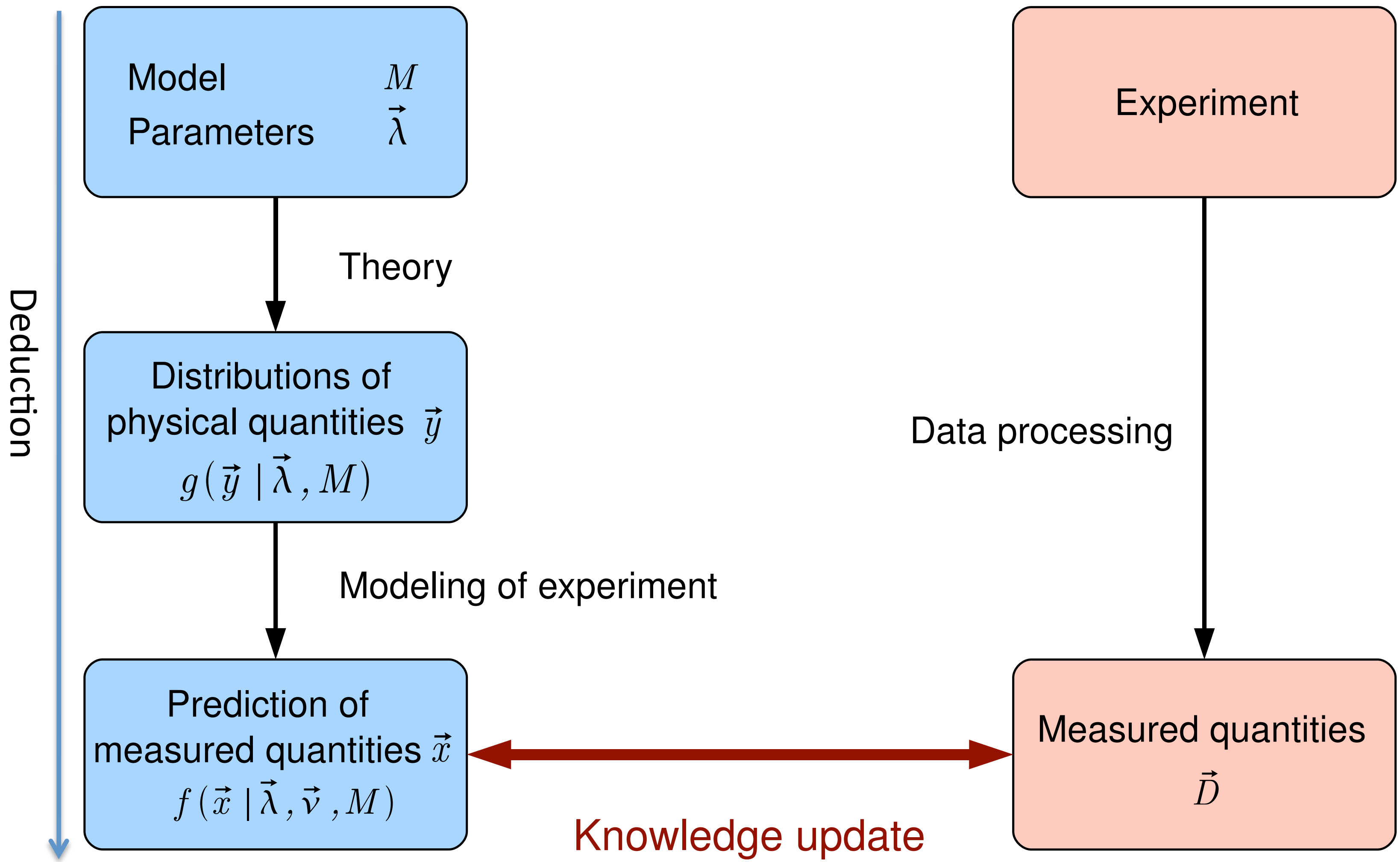
and some other things
A. Caldwell
Max Planck Institute for Physics

1. Some fundamentals
2. The Poisson Distribution
3. Bayesian Analysis with the Poisson distribution
4. Poisson distribution for signal and background
5. Examples

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

MAX-PLANCK-GESELLSCHAFT

How we learn


Model M
Parameters $\vec{\lambda}$
Theory
Distributions of physical quantities $\vec{y}$
$g(\vec{y}\,|\,\vec{\lambda}, M)$
Modeling of experiment
Prediction of measured quantities $\vec{x}$
$f(\vec{x}\,|\,\vec{\lambda}, \vec{\nu}, M)$
Deduction
Experiment
Data processing
Measured quantities
$\vec{D}$
Knowledge update

Logical Basis

Model building and making predictions from models follows deductive reasoning:

Given A→B (major premise)
Given B→C (major premise)
Then, given A you can conclude that C is true

etc.

Everything is clear, we can make frequency distributions of possible outcomes within the model, etc. This is math, so it is correct …

Logical Basis

However, in physics what we want to know is the validity of the model given the data. i.e., logic of the form:

Given A$\Rightarrow$C with some ‘probability’
Measure C, what can we say about A ?

Well, maybe $A_1 \Rightarrow C$, $A_2 \Rightarrow C$, …

We can only disprove (C not possible in A, then A invalid).

We are only capable of expressing a ‘degree of belief’ in A. And since we can never say anything is true, the question is – is it good enough ? Are we willing to bet on A providing the right answer to the next question ? Under what odds ?

Logical basis

Instead of truth, we consider knowledge

Knowledge = **justified ~~true~~ belief**

Justification comes from the data.

Start with some knowledge or maybe plain belief

Build a model

Make some predictions

Do the experiment

Data analysis gives updated knowledge (belief in possible parameter values)

Which probability ?

Data analysis is based on building a ‘probability’ for the data. But is this well defined ?

Imagine we flip a coin 10 times, and get the following result:

T H T H H T H T T H

We now repeat the process with and get

T T T T T T T T T T

Which outcome has higher probability ?

Take a model where H, T are equally likely. Then, probability of the sequence is

outcome 1

And

outcome 2

Something seem wrong with this result ?

Given a fair coin, we could also calculate the chance of getting n times H:

And we find the following result:

n	p
0	$1 \cdot 2^{-10}$
1	$10 \cdot 2^{-10}$
2	$45 \cdot 2^{-10}$
3	$120 \cdot 2^{-10}$
4	$210 \cdot 2^{-10}$
5	$252 \cdot 2^{-10}$
6	$210 \cdot 2^{-10}$
7	$120 \cdot 2^{-10}$
8	$45 \cdot 2^{-10}$
9	$10 \cdot 2^{-10}$
10	$1 \cdot 2^{-10}$

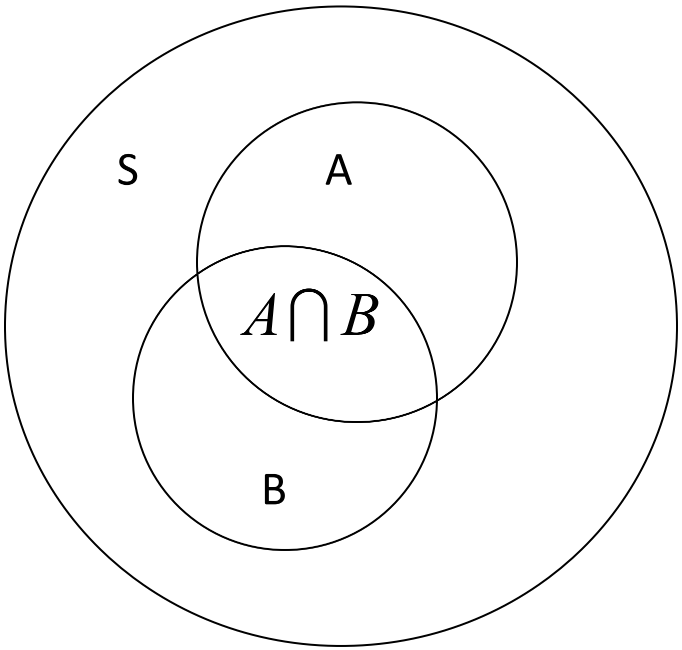
There are typically an infinite number of choices you can make for the 'probability of the data' or likelihood.

If someone claims to have an optimal definition, ask them 'based on what criterion ?' There is no one best answer !

Choosing a probability of your data is a critical component of the analysis process. Get the most out of your data !

Mathematical Definitions

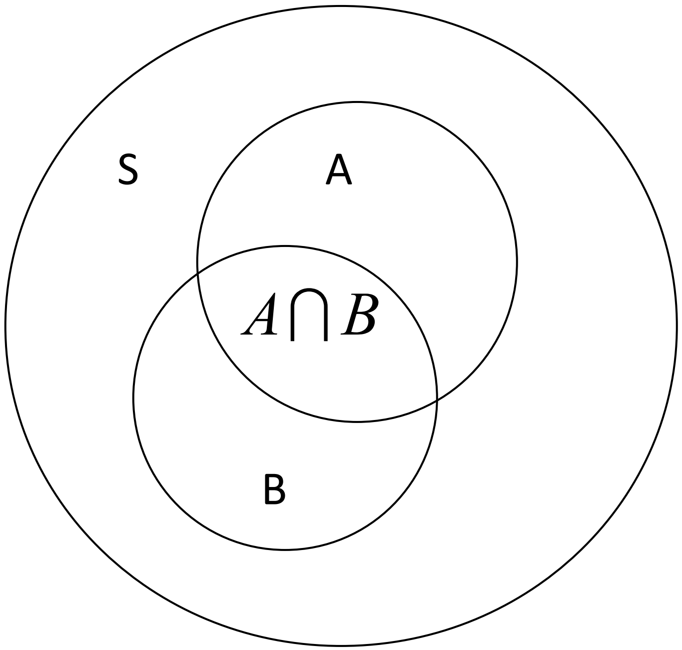
Consider a set, S, the sample space, which can be divided into subsets.



Probability is a real-valued function defined by the Axioms of Probability (Kolmogorov):

1. For every subset A in S, P(A)≥0.
2. For disjoint subsets
   $$A \cap B = \phi,$$
   $$P(A \cup B) = P(A) + P(B)$$
3. P(S)=1

Mathematical Definitions



Definition of conditional probability:

$$P(A \mid B) = \frac{P(A \bigcap B)}{P(B)}$$

$$\text{Since } P(A \bigcap B) = P(B \bigcap A),\ \text{Bayes' Theorem follows}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Law of Total Probability

$$P(B) = \sum_i P(B \mid A_i) P(A_i)$$

for any subset B and for disjoint A_i such that $\bigcup_i A_i = S$

Combining with Bayes' Theorem gives

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}}$$

If you want to make a statement about how much ‘probability’ to assign to A, there is only one way – Bayes’ Theorem.

Why isn't everyone a Bayesian ?

My suspicion: it is because most people do not understand the frequentist approach. Frequentist statements and Bayesian statements are thought to be about the same logical concept, and the frequentist statement does not require a prior, so …

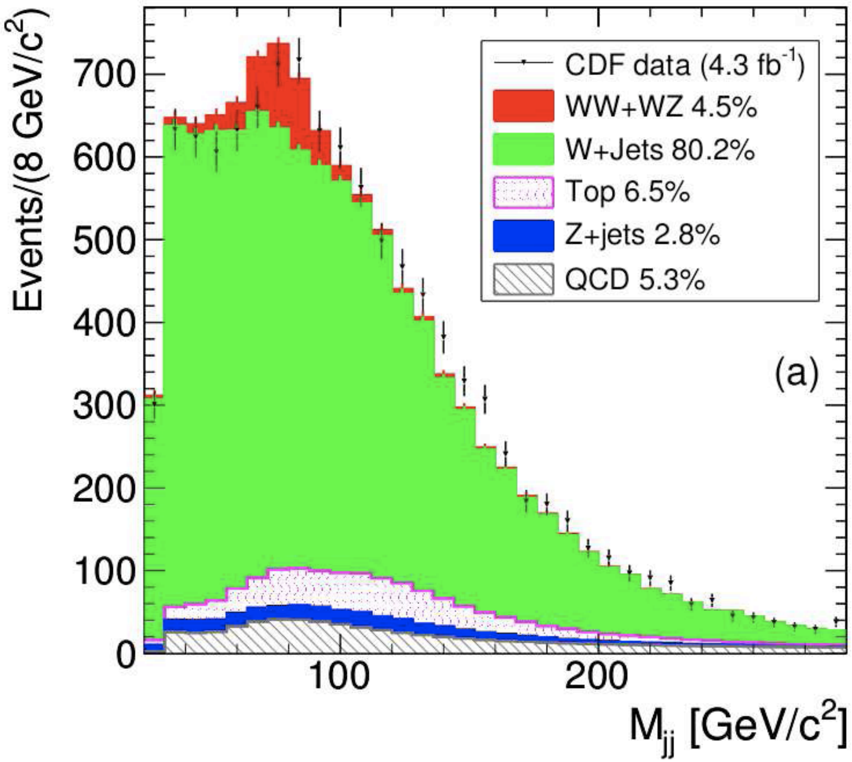
A. L. Read, *Presentation of search results: the CL_S technique,* J. Phys. G: Nucl. Part. Phys. **28** (2002) 2693-2704.

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.

Why isn't everyone a Bayesian ?

G. D'Agostini, Probably a discovery: Bad mathematics means rough scientific communication, arXiv:1112.3620v2 [physics.data-an]



Quoting a Discovery article:
It is what is known as a ``three-sigma event,'' and this refers to the statistical certainty of a given result. In this case, this result has a 99.7 percent chance of being correct (and a 0.3 percent chance of being wrong).''

$$1 - P(D|H_0) = P(H_1|D)$$

This is nonsense !

The Higgs announcement

Gemeinsame Presseerklärung des
Komitee für Elementarteilchenphysik KET
Forschungsschwerpunkt ATLAS (BMBF-FSP 101 ATLAS)
Forschungsschwerpunkt CMS (BMBF-FSP 102 CMS)
Deutsches Elektronen-Synchrotron DESY
Max-Planck-Institut für Physik
Helmholtz-Allianz „Physik an der Teraskala“

Der Nachweis eines neuen Teilchens wird in der Teilchenphysik klassischerweise auf zwei Stufen gestellt: Die Messungen, die die Wissenschaftler an ihren Experimenten durchführen, beruhen auf Statistik. Sie geben daher zu jedem ihrer Ergebnisse die Sicherheit als so genannte Signifikanz an. Die Einheit, die sie dafür verwenden ist sigma, dargestellt durch den griechischen Buchstaben σ. Die erste Stufe eines Teilchenfunds („evidence“) ist erreicht, wenn sich das Signal des Teilchens mit einer Deutlichkeit zeigt, dass die Physiker mit 99,75 Prozent Sicherheit von seiner Echtheit ausgehen. Dies entspricht einer Signifikanz von 3σ. Von einer „Entdeckung“ und damit der zweiten Stufe sprechen die Forscher bei einer Signifikanz von 5σ, das entspricht einer Fehlerwahrscheinlichkeit von 0,000057%.

Translation - Probability of error is 0,000057%

Error on what ????? That the Higgs is found - not correct

What happened

equated $$1 - P(D|H_0) = P(H_1|D)$$

Probability of observing the data or something more extreme given the background only hypothesis

Probability that the Higgs exists

This is logical nonsense …

Who's fault is this confusion ? I would say – physicists should know better ! In the Bayesian approach, we state our prior assumptions and show how they lead to the conclusions.

Poisson Distribution

A Poisson distribution applies when we do not know the number of trials (it is a large number), but we know that there is a fixed probability of 'success' per trial, and the trials occur independently of each other.

Alternatively – a continuous time process with a constant rate will produce a Poisson distributed number of events in a fixed time interval.

High energy physics example: beams collide at a high frequency (10 MHz, say), and the chance of a 'good event' is very small. The resulting number of events in a fixed time will follow a Poisson distribution. A single trial is one crossing of the beams.

Nuclear physics example: a large sample of radioactive atoms will produce a Poisson distributed number of events in a fixed time interval (assuming a $\tau >> T$)

Poisson Distribution

The Poisson distribution can be derived from the Binomial distribution in the limit when N $\to\infty$ and p $\to 0$, but Np fixed and finite. Then

$$P(r|N,p) \to P(n|\nu)$$

The expected number of events is calculated from a rate, or from a luminosity and cross section or some other way

$$\nu = R \cdot T \text{ or } \nu = \mathcal{L} \cdot \sigma \text{ or...}$$

Poisson Distribution - derivation

$$P(n|N,p) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n}$$

$$P(n|N,\frac{\nu}{N}) = \frac{N!}{n!(N-n)!}\frac{\nu^n}{N^n}\left(1-\frac{\nu}{N}\right)^{N-n}$$

$$N \longrightarrow \infty$$

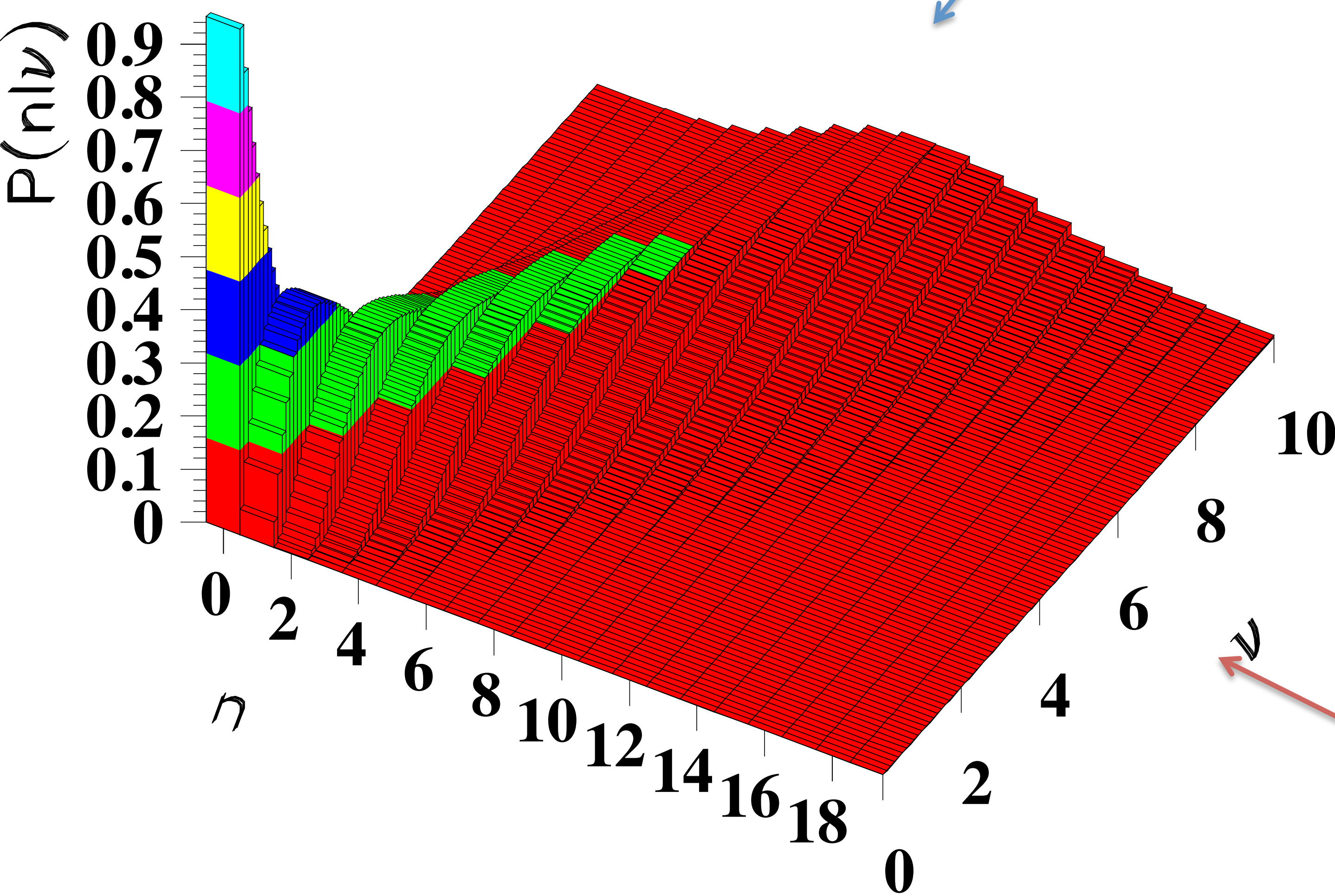
$$\frac{N!}{(N-n)!} = N\cdot(N-1)\cdot\ldots\cdot(N-n+1) \approx N^n$$

$$\left(1-\frac{\nu}{N}\right)^{N-n} \rightarrow \left(1-\frac{\nu}{N}\right)^{N} \longrightarrow e^{-\nu}$$

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

Poisson Distribution

Poisson Example



Quantity used in likelihood analysis

Probability of the data used in confidence level setting

Poisson Distribution-cont.

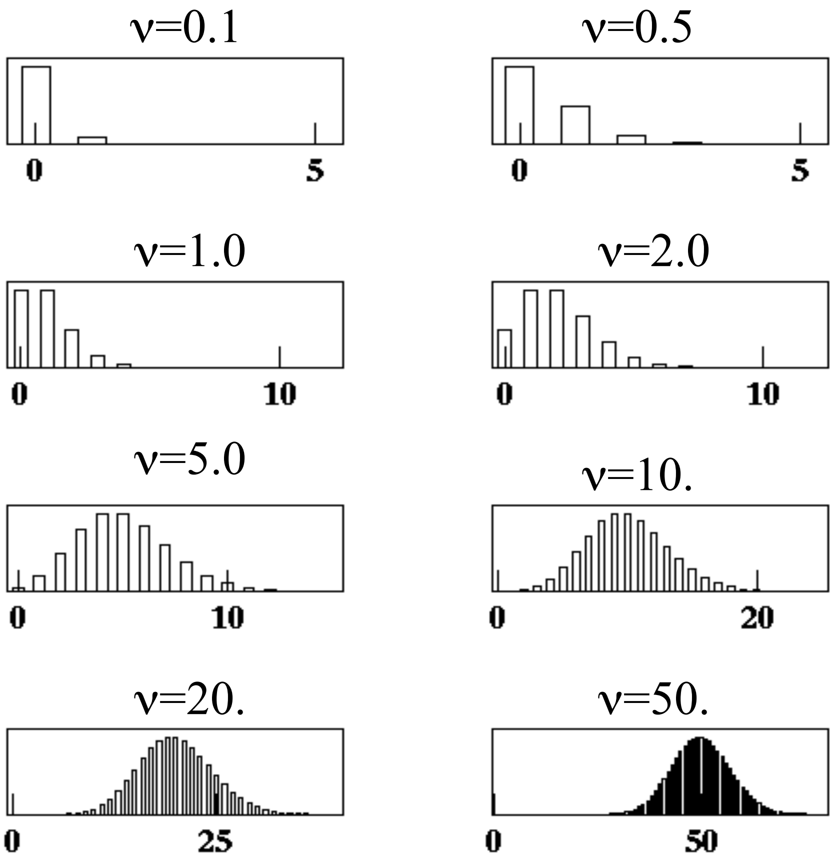
$$P(n \mid \nu) = \frac{\nu^n e^{-\nu}}{n!}$$

$$n^* = \lfloor \nu \rfloor$$

$$n^* = \lceil \nu \rceil - 1$$

$$E[n] = \nu$$

$$\mathrm{V}[\mathrm{n}] = \nu$$



Notes:

- As ν increases, the distribution becomes more symmetric
- Approximately Gaussian for large ν

Bayesian Data Analysis-Poisson Distribution

Typical examples – counting experiments, failure rates, cross sections,…

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

This is our master formula. Result in general will depend on choice of prior. *In general, we need to go straight to the numerical solution.*

Why not – computing is cheap today. Avoid the simplifications and approximations. You can do the full calculation given the right tool.

Poisson - cont.

This is a lecture, so you expect some formulae.

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.

The expectation value:

$$< \nu >= \int_0^\infty P(\nu|n)\nu d\nu = \int_0^\infty \frac{\nu^n e^{-\nu}}{n!}\nu d\nu = \frac{(n+1)!}{n!} = n+1$$

The variance:

$$\begin{aligned}\sigma^2 &= \int_0^\infty P(\nu|n)(\nu - < \nu >)^2 d\nu \\ &= \int_0^\infty \frac{\nu^n e^{-\nu}}{n!}\nu^2 d\nu - < \nu >^2 \overbrace{\int_0^\infty \frac{\nu^n e^{-\nu}}{n!} d\nu}^{1} \\ &= \frac{(n+2)!}{n!} - (n+1)^2 = n+1\end{aligned}$$

Poisson - cont.

Note: $n=0 \quad <\nu>=1 \quad ???$

From prior, expect
$$< \nu > = \int_0^{\nu_{max}} P_0(\nu)\nu d\nu = \int_0^{\nu_{max}} \frac{\nu}{\nu_{max}} d\nu$$
$$= \left[\frac{\nu^2}{2(\nu_{max})}\right]_0^{\nu_{max}}$$
$$= \frac{\nu_{max}}{2}$$

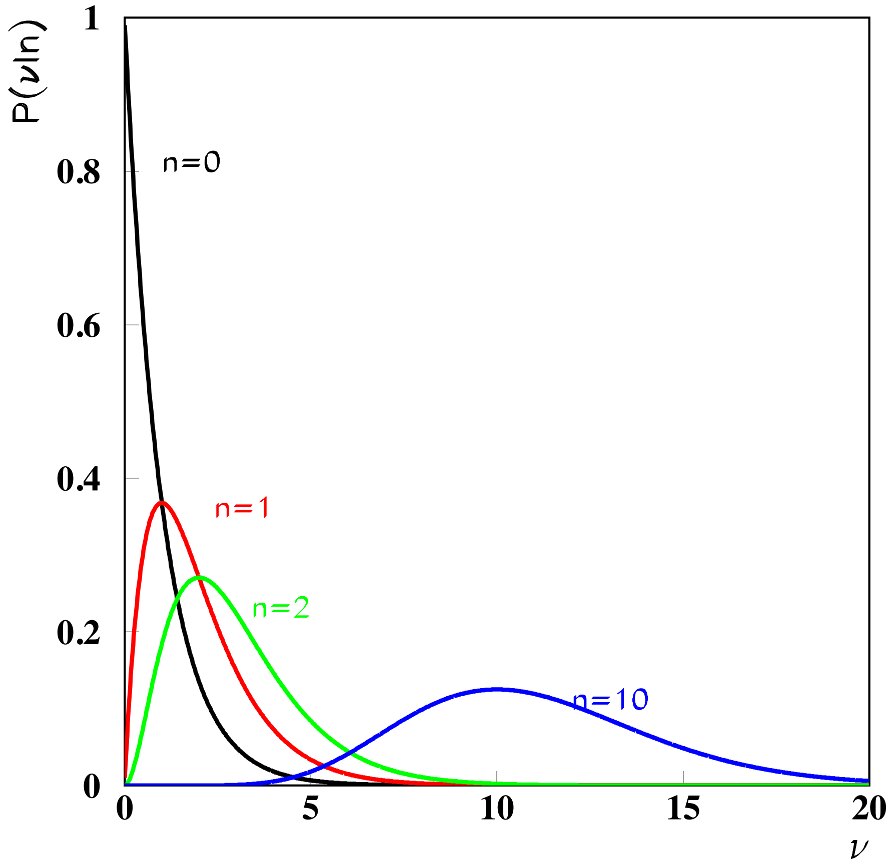
What happened ?

n=0 is a measurement !

$$P(\nu|0) = e^{-\nu}$$

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:

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

Poisson – Examples

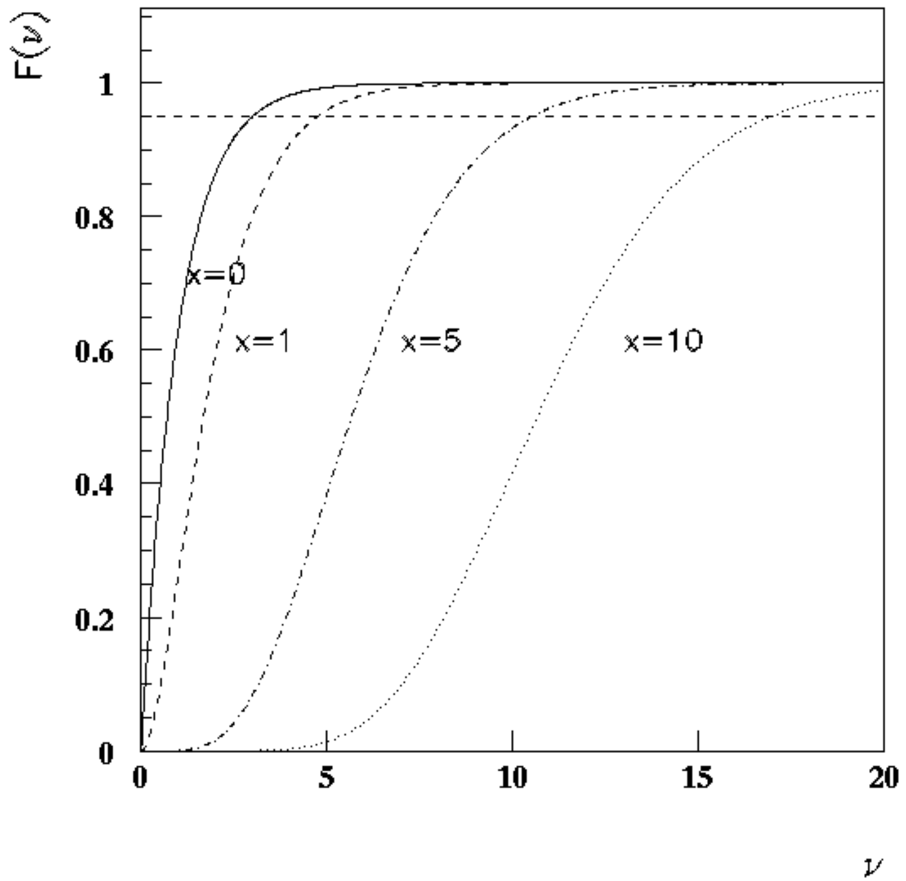
Assume measure zero counts.

With flat prior assumption

$$
\begin{aligned}
P(\nu|n=0) &= e^{-\nu} \\
F(\nu|n=0) &= 1-e^{-\nu}
\end{aligned}
$$

For a 95% credibility upper limit

$$
\begin{aligned}
0.95 &= 1-e^{-\nu} \\
\nu &\approx 3
\end{aligned}
$$



Poisson – cont.

What if we cannot (or do not want to) take a flat prior

Suppose we can model the prior belief as $P_0(\nu) = \dfrac{1}{10} e^{-\nu/10}$

Now Bayes tells us $P(\nu \mid x=0) = \dfrac{P(0 \mid \nu) P_0(\nu)}{\int_0^\infty P(0 \mid \nu) P_0(\nu) d\nu} = \dfrac{e^{-\nu} \frac{1}{10} e^{-\nu/10}}{\int_0^\infty \frac{1}{10} e^{-11\nu/10} d\nu} = \dfrac{11}{10} e^{-11\nu/10}$

$$< \nu > = \int_0^\infty \frac{11}{10} e^{-11\nu/10} \nu d\nu = 0.91$$

$P(\nu \le 2.7) = 95\%$, i.e., $\nu \le 2.7$ with 95% probability

Exercises

1. paper and pencil
   a) Find the distribution of the waiting time for the k^{th} event in a process with a constant rate λ.
   b) For a Poisson with mean 1.5, what is the probability to see 6 or more events ? What is the probability to see exactly 0 events ?
   c) Prove that for a Poisson distribution

$$n^* = \lfloor \nu \rfloor = \lceil \nu \rceil - 1$$

Poisson Distribution-cont.

We often have to deal with a superposition of two Poisson processes – the signal and the background, which are indistinguishable in the experiment. Usually we know the background expectations and want to know the probability of a signal in addition.

Example, the signal for large extra dimensions may be the observation of events where momentum balance is (apparently) strongly violated. However this can be mimicked by neutrinos, energy leakage from the detector, etc.

Use the subscripts B for background, s for signal, and assume n events are observed

$$P(n) = \sum_{n_s=0}^{n} P(n_s \mid \nu_s) P(n - n_s \mid \nu_B)$$

$$= e^{-(\nu_B + \nu_s)} \sum_{n_s=0}^{n} \frac{\nu_s^{\,n_s} \nu_B^{\,n-n_s}}{n_s!(n-n_s)!}$$

$$= e^{-(\nu_B + \nu_s)} \frac{(\nu_s + \nu_B)^n}{n!} \overbrace{\underbrace{\sum_{n_s=0}^{n} \frac{n!}{n_s!(n-n_s)!} \left(\frac{\nu_s}{\nu_s + \nu_B} \right)^{n_s} \left(\frac{\nu_B}{\nu_s + \nu_B} \right)^{n-n_s}}_{=1 \text{ by normalization}}}^{\text{Binomial formula with } p = \left(\frac{\nu_s}{\nu_s + \nu_B} \right)}$$

$$= e^{-(\nu_B + \nu_s)} \frac{(\nu_s + \nu_B)^n}{n!}$$

The Bayesian Way

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

Assuming that the background is perfectly known:

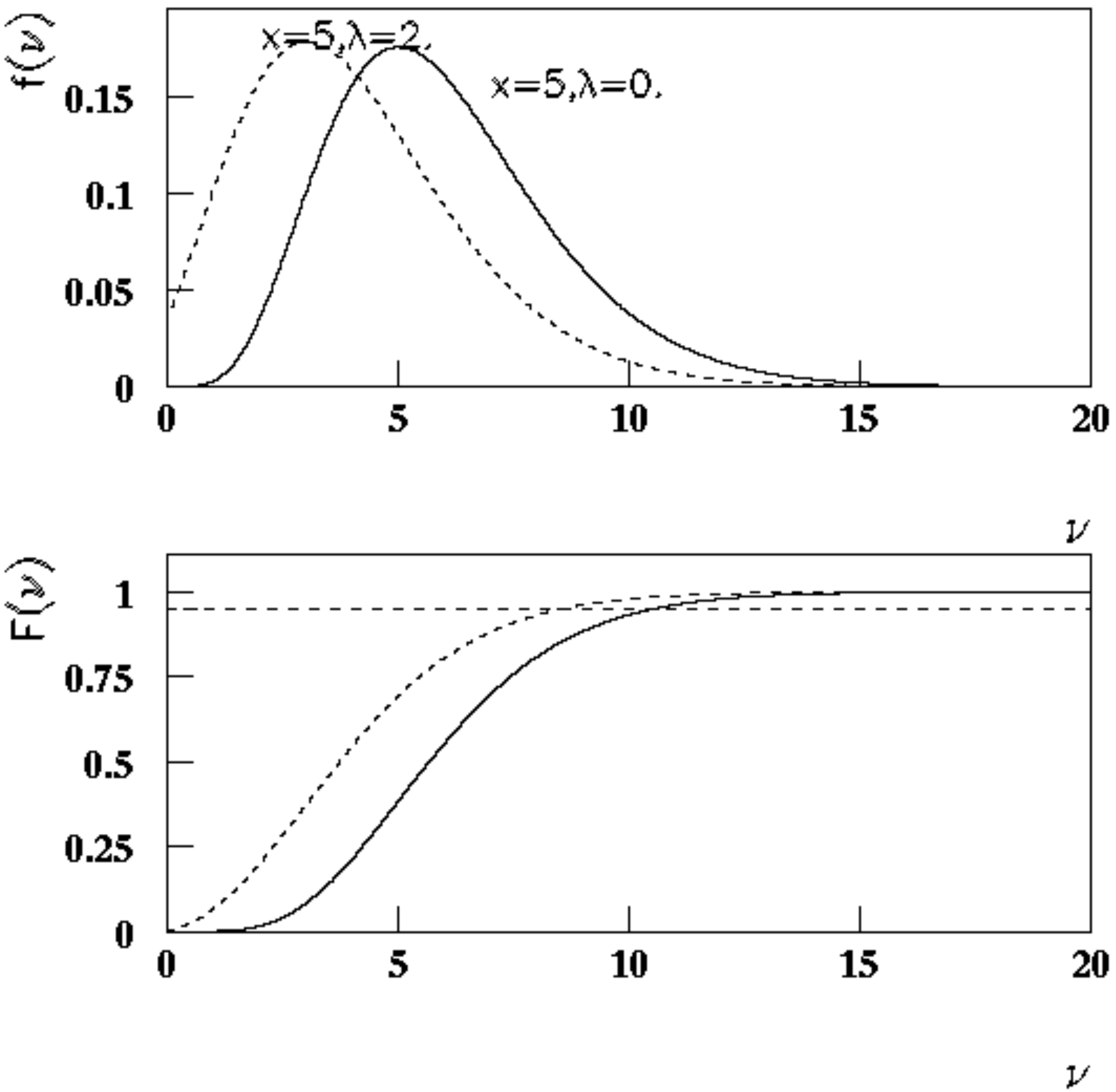
$$P(\nu|n,\lambda) = \frac{P(n|\nu,\lambda)P_0(\nu)}{\int P(n|\nu,\lambda)P_0(\nu)d\nu}$$

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.



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.

Example

Want to test a new theory – Large Extra Dimensions. If this hypothesis is correct, we expect events with certain characteristics in (let's say) proton-proton collisions. We design an experiment to look for this process.

There will also be indistinguishable events from 'known' physics. The analysis has been designed to reduce these, but there will be some background left.

Background expectation: $\lambda = \sigma_{SM} \cdot \mathcal{L} \cdot a_{SM}$

Signal expectation: $\nu = \sigma_{LED} \cdot \mathcal{L} \cdot a_{LED}$

Have a nearly infinite number of collisions of protons with very small probability to generate an event per bunch crossing: Poisson process

Example

Probabilistic model:

$$P(n_B|\lambda) = \frac{e^{-\lambda}\lambda^{n_B}}{n_B!}$$

$$P(n_S|\nu) = \frac{e^{-\nu}\nu^{n_S}}{n_S!}$$

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

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

Example

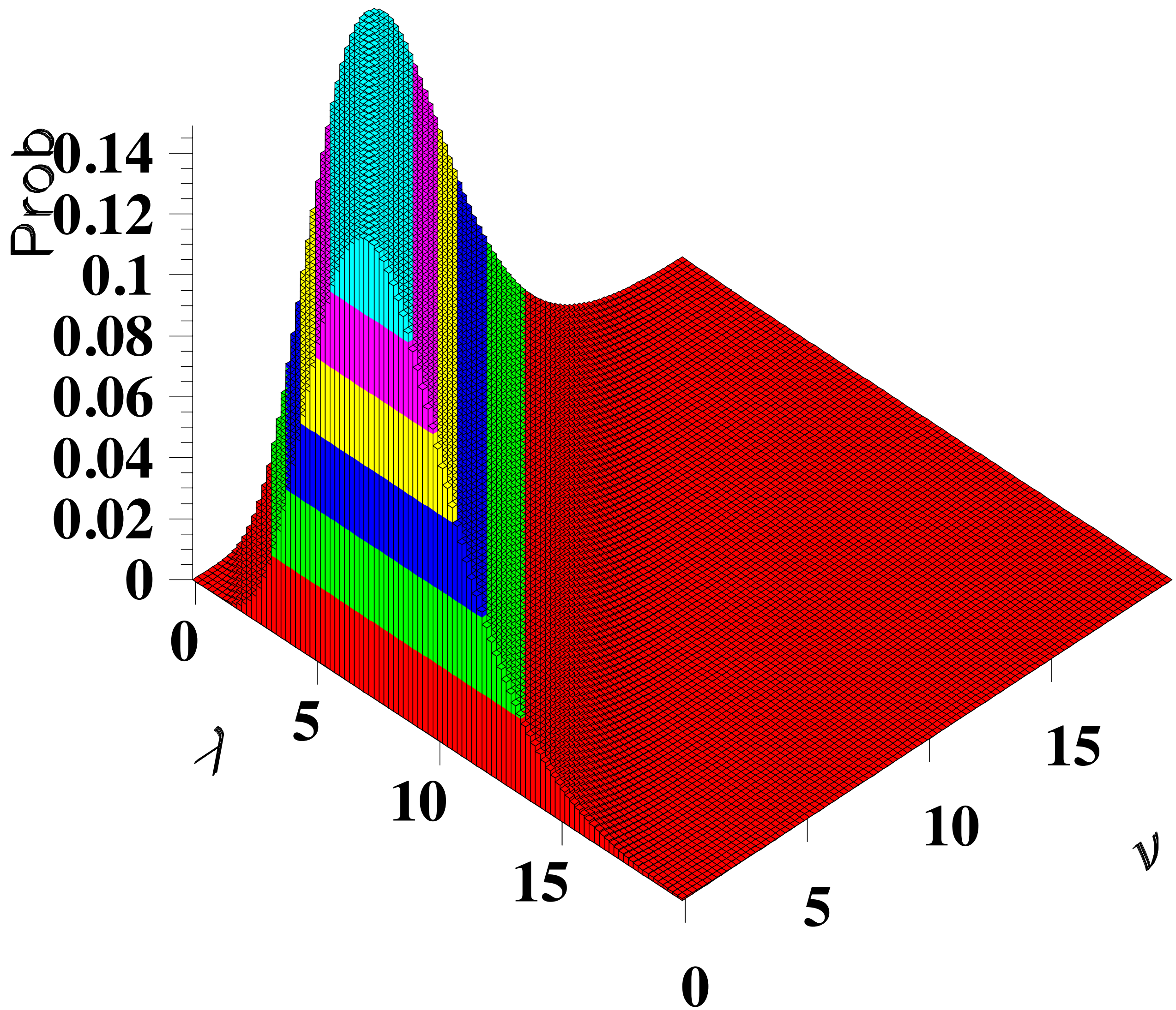
Compare two situations:

1) no knowledge on the background

2) Separate data help us constrain the background

Suppose we measure n=7 events, what can we say ?

n=7 Poisson



With Background knowledge - Bayes

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

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

$$P(\lambda) = \frac{1}{\sqrt{2\pi}\sigma_\lambda} e^{-\frac{1}{2}\frac{(\lambda-\lambda_0)^2}{\sigma_\lambda^2}}$$

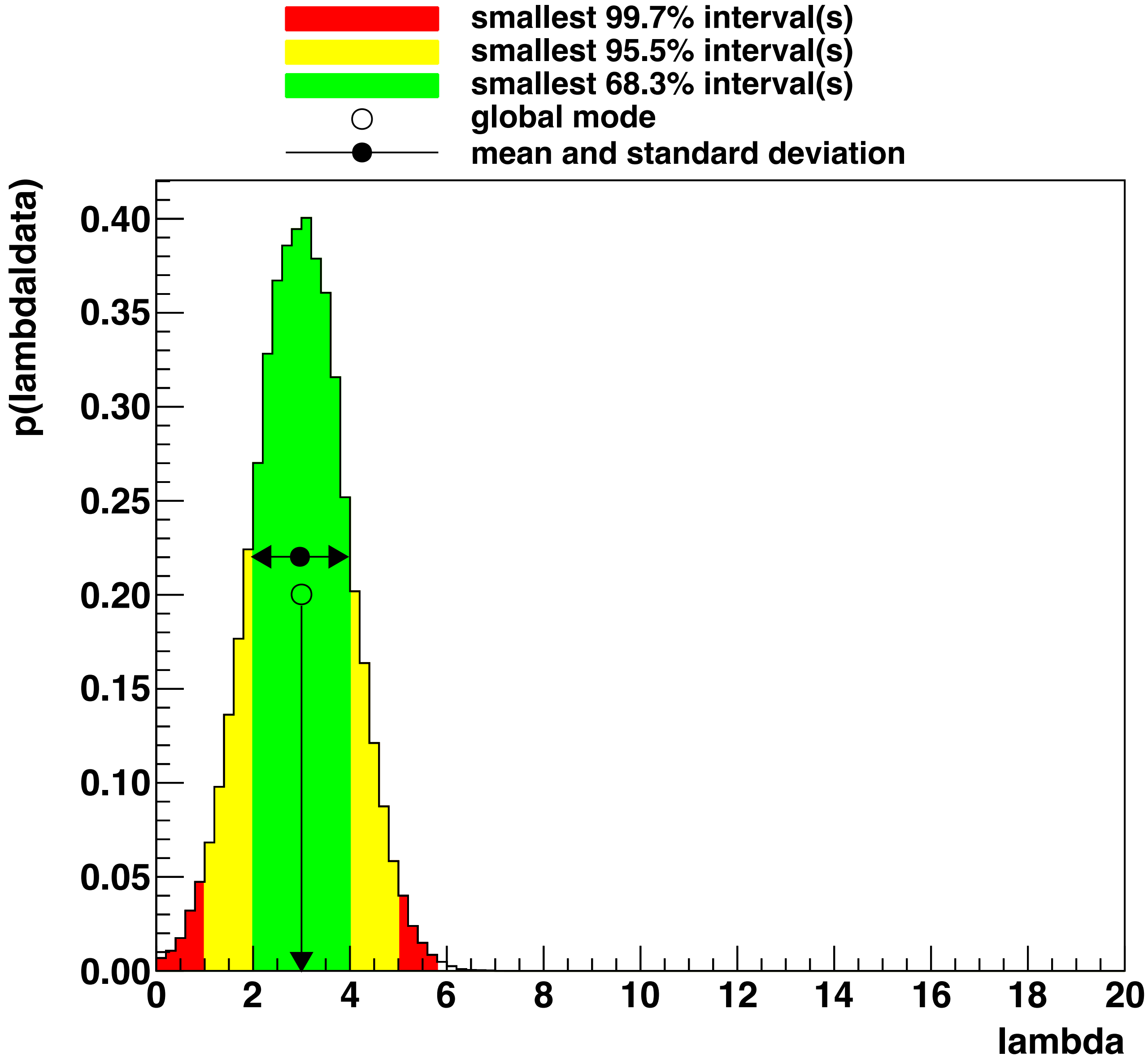
$$P_0(\nu) = \text{constant}$$

We solve this numerically (here with the BAT package) https://www.mpp.mpg.de/bat/

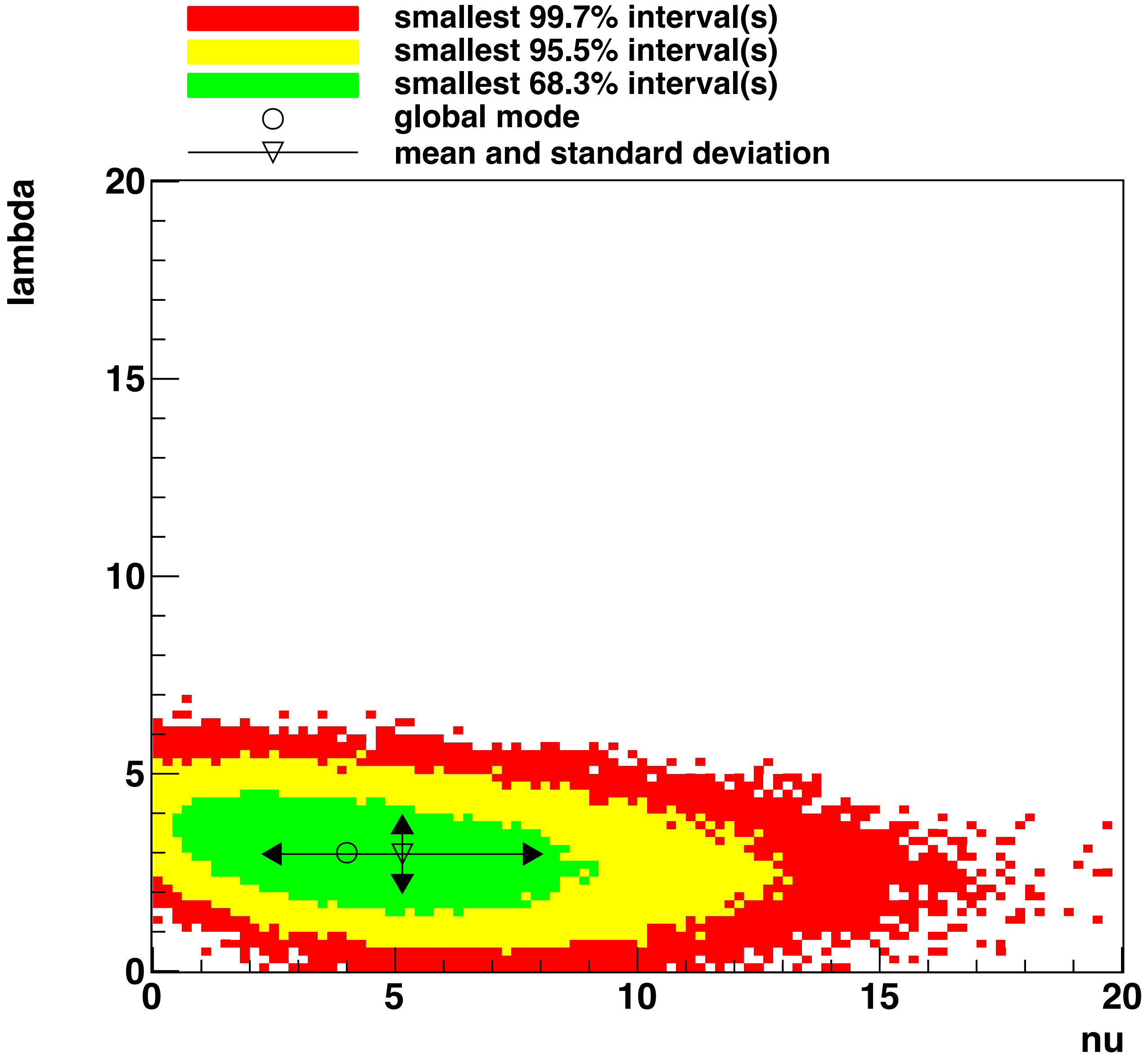
To get a probability distribution for the physics parameter, we marginalize

$$P(\nu|n) = \int P(\nu, \lambda|n) d\lambda$$

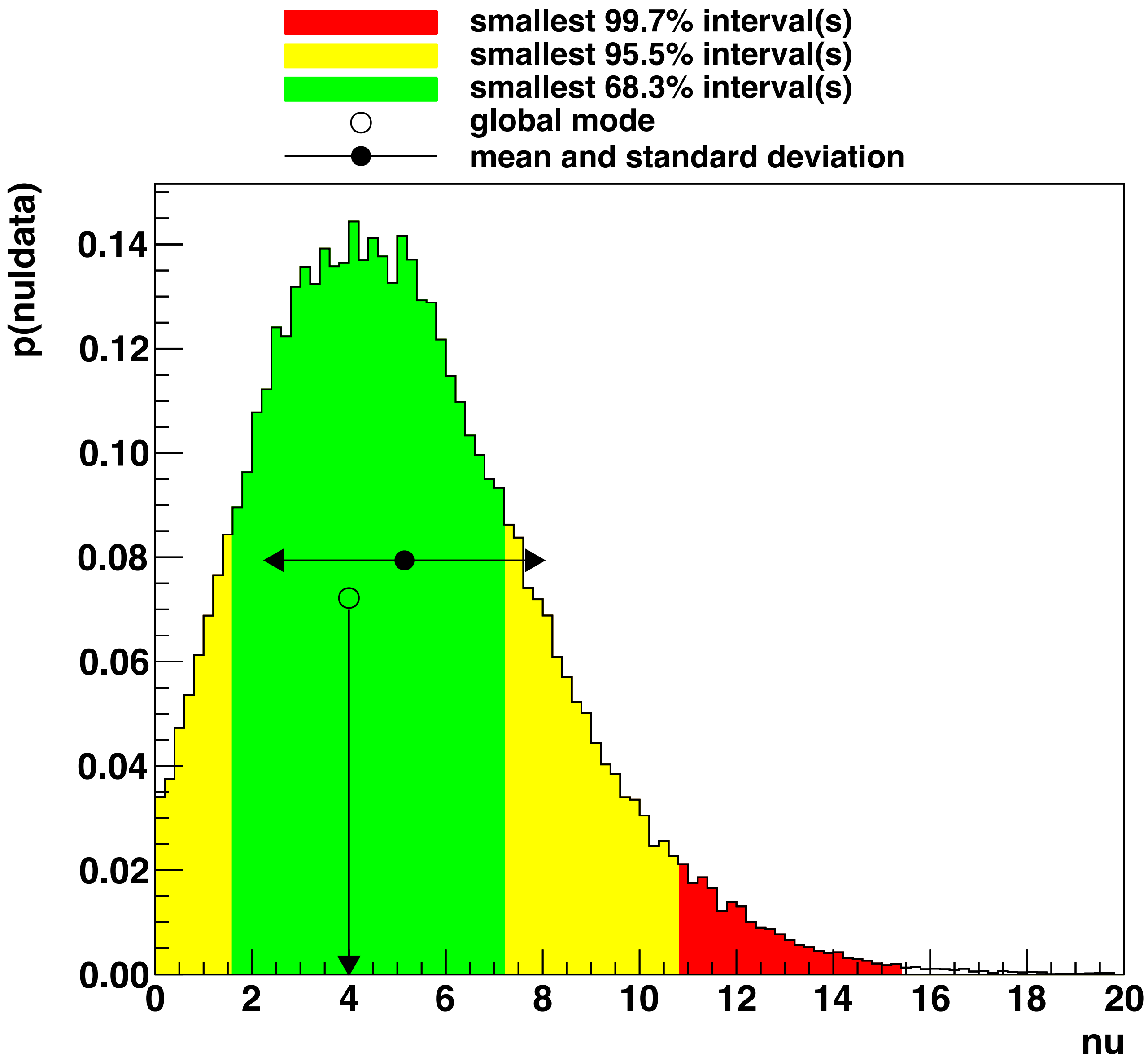
n=7 Constrained Background



n=7 Constrained Background



n=7 Constrained Background



Example: Double Beta Decay

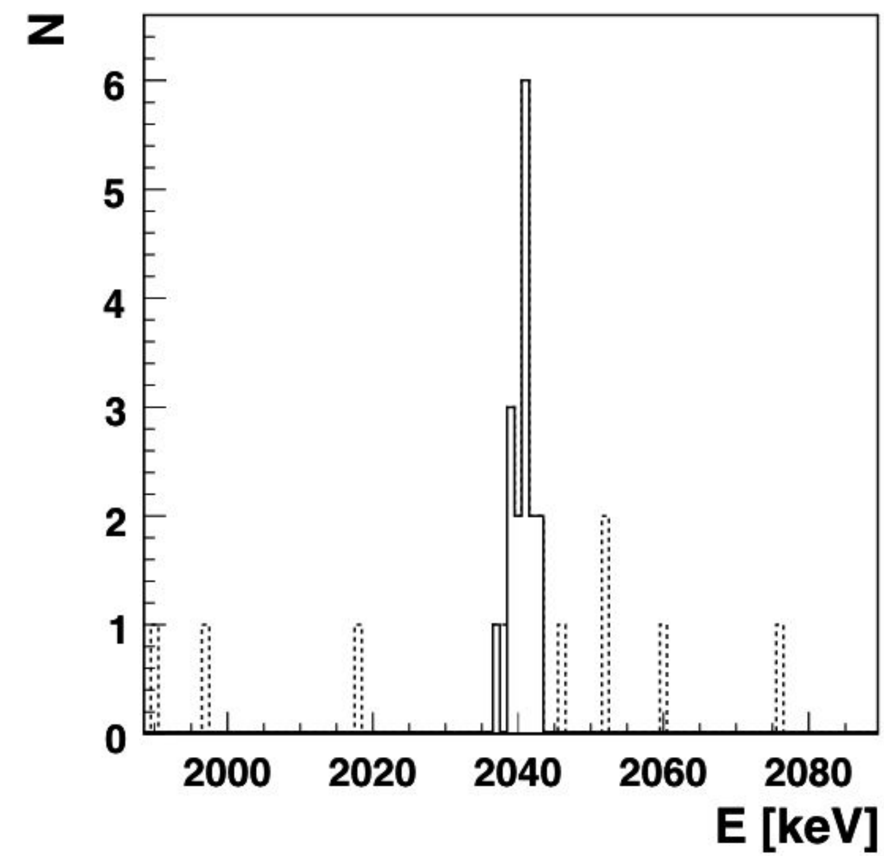
One of the outstanding questions in Particle Physics is whether the neutrino is its own antiparticle (so-called Majorana particle).

The only practical way which has been found to search for the Majorana nature of neutrinos (particle same as antiparticle) is double beta decay (because of the light mass of neutrinos, helicity flip is very unlikely unless the neutrinos have very low energy).

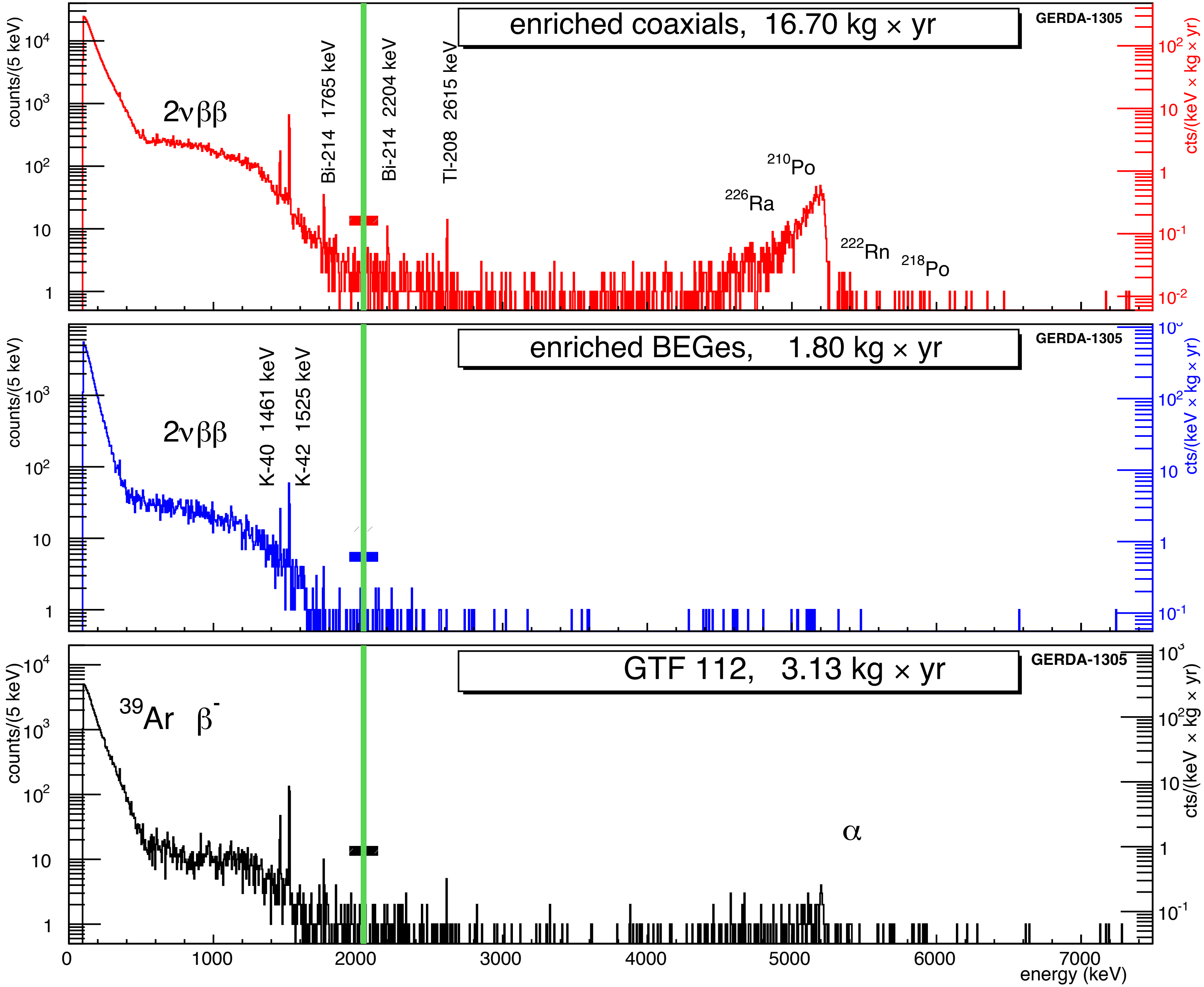
For us, what is interesting is that we are looking for a peak at a well-defined energy in a sparse spectrum.

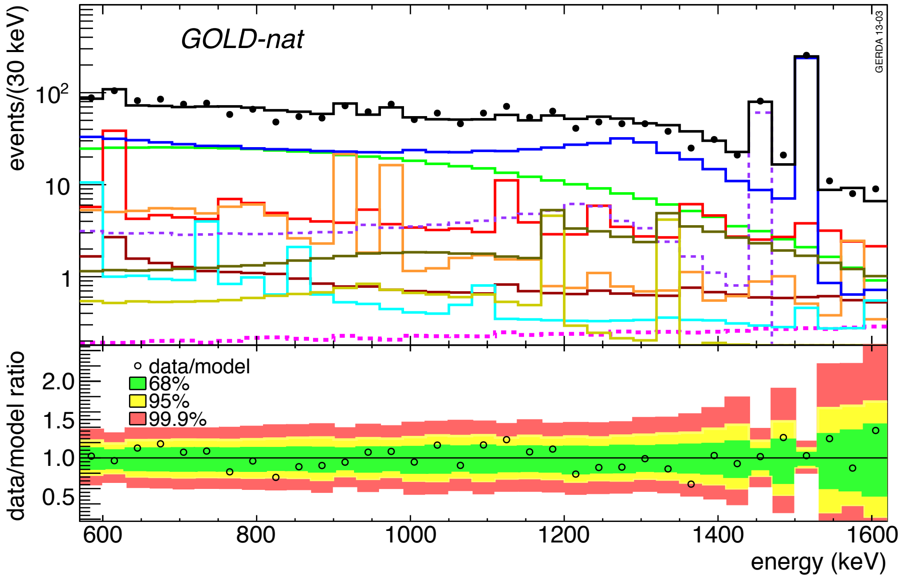
A. Caldwell, K. Kröninger, Phys. Rev. D 74 (2006) 092003

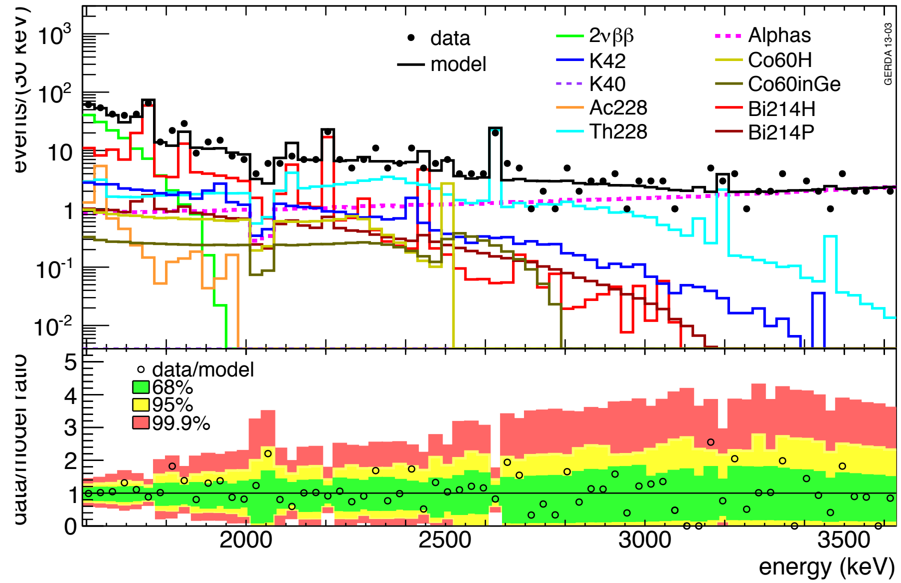
Discovery or not ?



Analyze energy spectrum and decide if there is evidence for a signal. Counting experiment – Poisson statistics.

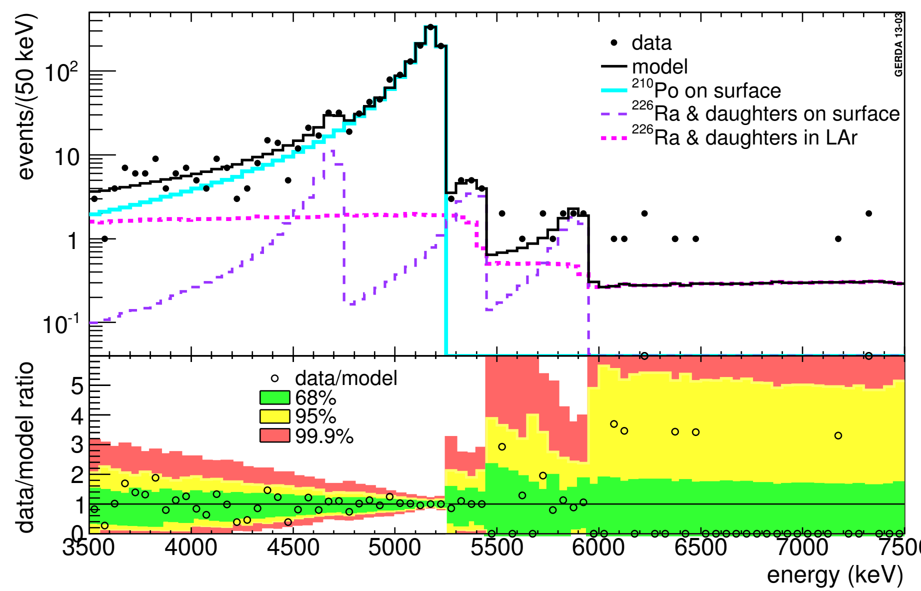

enriched coaxials, 16.70 kg × yr
GERDA-1305
2νββ
Bi-214 1765 keV
Bi-214 2204 keV
Tl-208 2615 keV
226Ra
210Po
222Rn
218Po
counts/(5 keV)
cts/(keV × kg × yr)
enriched BEGes, 1.80 kg × yr
GERDA-1305
K-40 1461 keV
K-42 1525 keV
GTF 112, 3.13 kg × yr
GERDA-1305
39Ar β-
α
energy (keV)





$$P(\vec{D}|\vec{N}) = \prod_i \frac{e^{-\nu_i}\nu_i^{n_i}}{n_i!}$$

$$\nu_i(\vec{N}) = \sum_j \epsilon_j N_j \int_{\Delta E_i} f_j(E)dE$$



Error Bars for Distributions of Numbers of Events

[Ritu Aggarwal, Allen Caldwell](#)

[European Physical Journal Plus](#)

Do not put error bars on event counts !

Exercises

1. For the following data set:

Data Set	Source in/out	Run Time	Events
1	Out	1000 s	100
2	In	2000 s	250

a) Plot the probability distribution for the background rate from Data set 1 only
b) Analyze the two data sets simultaneously; plot the 2D probability density for the background and signal rates.
c) Find the 68% central credibility interval for the decay rate. If your sample had a mass of one gram, and the isotope in the sample has an atomic mass of m_A=110 gm/mole, what is the lifetime of the isotope (value with uncertainty) ?