An Introduction to Data Analysis

Allen Caldwell
Max Planck Institute for Physics

Terascale Alliance School "It's measurement time!"
March 31, 2014

1. Conceptual framework for data analysis
2. Probability of the data ?
3. Summarizing a probability distribution
4. Poisson Process
5. Poisson process with background
6. An example

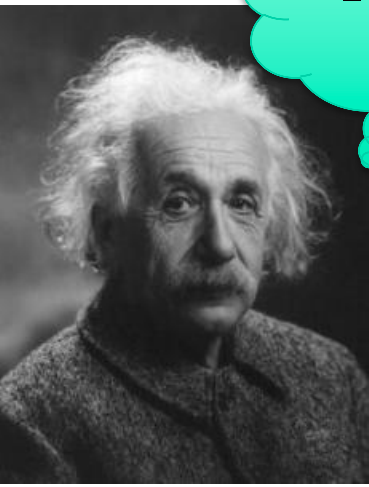
Write

MAX-PLANCK-GESELLSCHAFT

$\Delta p \cdot \Delta q \geq \frac{1}{2}\hbar$

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

Theory



$\vec{y}$ Are the theoretical observables

$\vec{\lambda}$ Are the parameters of the theory

M Is the model or theory

$$g(\vec{y}|\vec{\lambda}, M)$$

Modeling of experiment

$$f(\vec{x}|\vec{\lambda}, M)$$

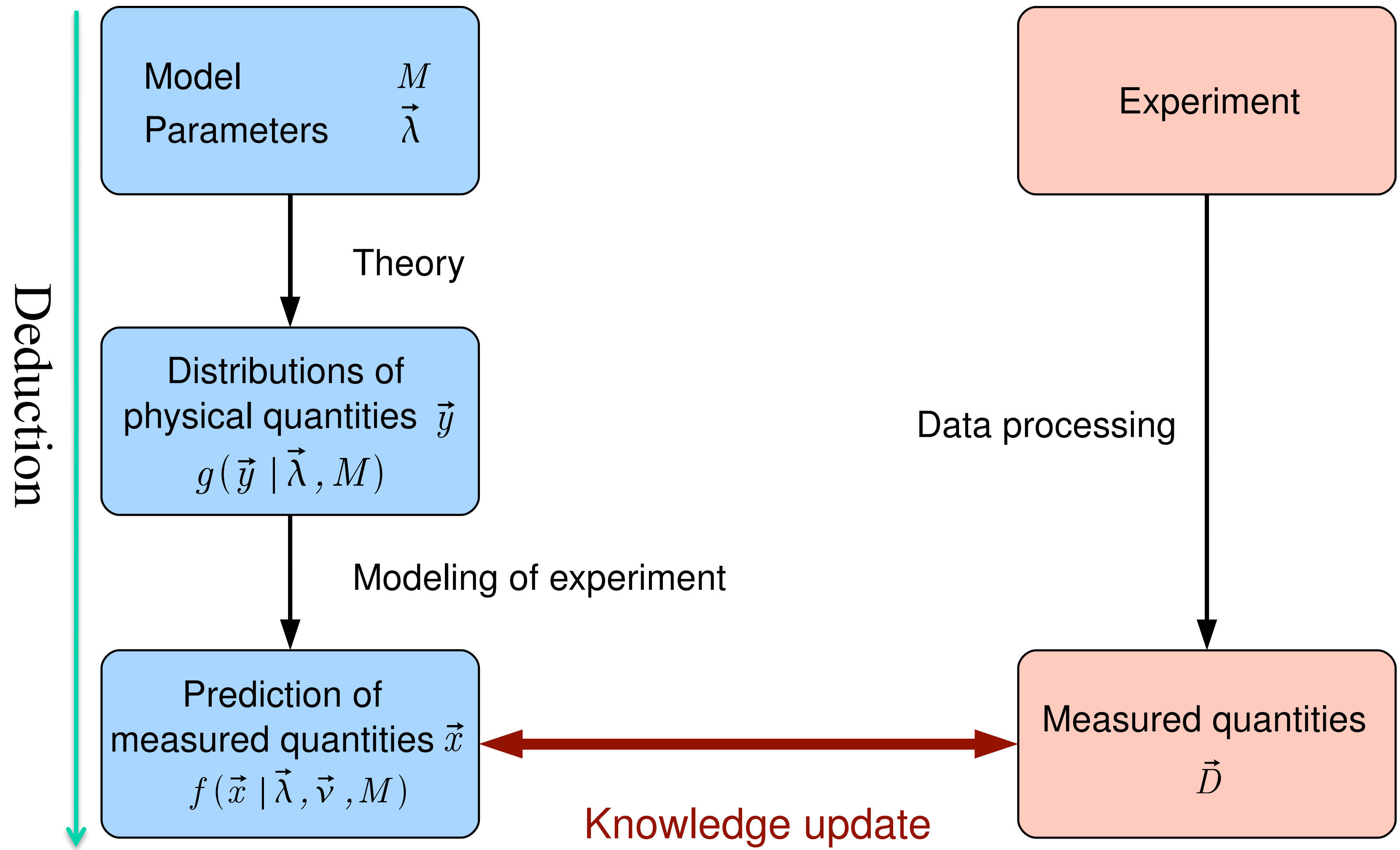
compare

$\vec{D}$

Experiment

$\vec{x}$ Is a possible data outcome

How we learn


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

How we Learn

We learn by comparing measured data with frequency distributions for possible results resulting from a theory, parameters, and a modeling of the experimental process.

What we typically want to know:

• Is the theory reasonable ? I.e., is the observed data a `likely result' from this theory.

• If we have more than one potential explanation, then we want to be able to quantify which theory is more likely to be correct given the observations

• Assuming we have a reasonable theory, we want to estimate the most probable values of the parameters, and their uncertainties. This includes setting limits (>< some value at XX% probability).

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 $A \Rightarrow C$
Measure C, what can we say about A ?

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

We now need inductive logic. We can never say anything absolutely conclusive about A unless we can guarantee a complete set of alternatives A_i and only one of them can give outcome C. This does not happen in science, so we can never say we found the true model.

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

Do the experiment

Data analysis gives updated knowledge. Experimental results in line with model predictions give justification for believing our model.

Elements of Data Analysis

Bayes

Probability of data given model M, par, λ

Probability of the data (Likelihood) $P(D|\lambda, M) = \mathcal{L}(\lambda|D)$

Likelihood ≠ probability

data fixed

λ variable

e.g., Poisson process $P(n|\lambda) = \dfrac{e^{-\lambda}\lambda^n}{n!}$ λ fixed

prob of observing n events given expectation λ

$\mathcal{L}(\lambda|n) = \dfrac{e^{-\lambda}\lambda^n}{n!}$ n fixed

All that is used in frequentist or classical approach

In a Bayesian analysis, also need the prior probability $P_0(\lambda|M)$

model assumed

Then use $$\underbrace{P(\lambda|M, D)}_{\text{posterior pdf}} = \frac{P(D|\lambda, M)P_0(\lambda|M)}{P(D|M)}$$

Bayes' Theorem

$$Z = P(D|M) = \int d\lambda P(D|\lambda, M)P_0(\lambda|M)$$ "evidence" often not needed

Bayesians and Frequentists

It is possible (Frequentism) to make statements of the kind:

'Assuming the model is correct, this result will occur in XX% of the experiments'

In the 'classical' approach, this is then converted to 'assuming the model, the bounds [a,b] will contain the true value in XX% of experiments performed' (confidence levels). Does not imply that the true value is in the range [a,b] with probability XX !

Only use deductive reasoning and the probability of the data assuming the model. *The inductive part of the reasoning is left out of the analysis – up to the user to decide what to believe. Often proceed with a community consensus (e.g., 5σ tail for background only hypothesis) (but only when convenient, e.g., Higgs but not superluminal neutrinos).*

Bayesians and Frequentists

It is also possible (Bayesianism) to make statements of the kind:

'the degree-of-belief in model A is XX (between 0,1)'

Given the new data, the degree-of-belief is updated using the frequencies of possible outcomes in the context of the models

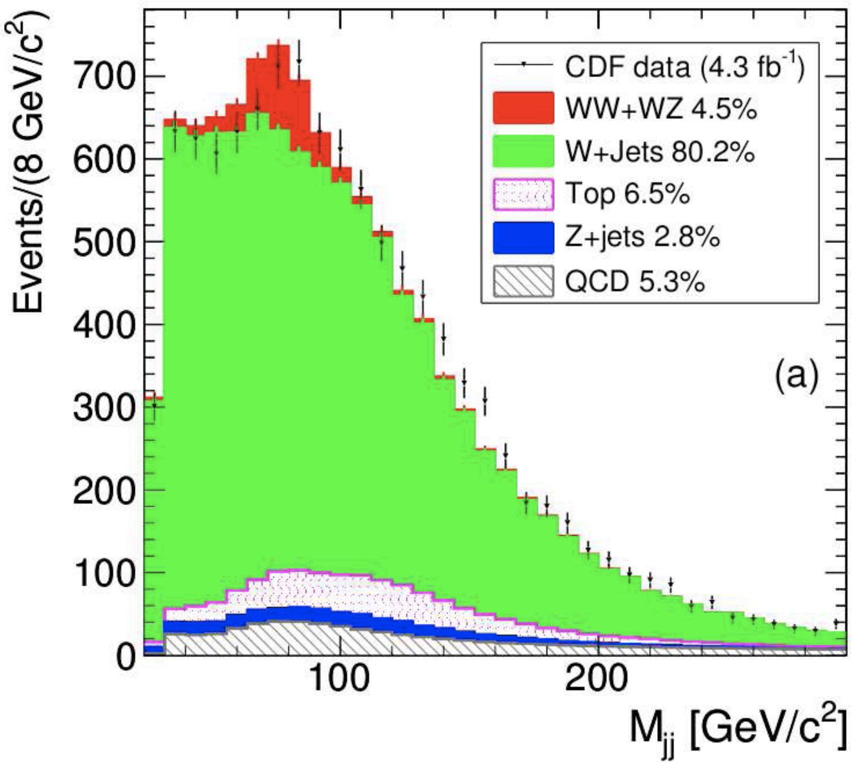
Credible regions are then defined: with XX% credibility, the parameter is in the interval [a,b]. **Note – very different from a CL.**

The inductive part of the reasoning is built into the analysis, and the connection between prior beliefs and posterior beliefs is made clear.

Subjective, but the subjective element is made explicit.

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 logical nonsense - confusion is very widespread !

Probability of the data

The expected distribution (density) of the data assuming a model M and parameters $\vec{\lambda}$ is written as $P(\vec{x}|\vec{\lambda}, M)$ where $\vec{x}$ is a possible realization of the data. There is usually **no unique definition** of the 'probability of the data.' Different choices incorporate different information.

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

$S_1 =$ T H T H H T H T T H

We now repeat the process with a different coin and get

$S_2 =$ 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,

outcome 1 $\quad prob = (1/2)^{10}$

And

outcome 2 $\quad prob = (1/2)^{10}$

Something seem wrong with this result ? This is because (in our head) we evaluate many probabilities at once. The result above is the probability for any sequence of ten flips of a fair coin. Given a fair coin, we could also calculate the chance of getting n times H:

trials

$P(n \mid N, p)$ $\qquad$ prob of success

successes

$$\binom{10}{n}\left(\frac{1}{2}\right)^{10}$$

$$P(n \mid N, p) = \binom{N}{n} p^{n} (1-p)^{N-n}$$

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 many more ways to get 5 H than 0, so this is why the first result somehow looks more probable, even if each sequence has exactly the same probability in the model.

Maybe the model is wrong and one coin is not fair ? How would we test this ?

Exercise – think up two more possible probabilities of the data for the heads & tails experiment

e.g. L = length of longest H sequence

P(L | N), S₁: L = 2, S₂: L = 0

or F = # transition H → T or T → H

S₁: F = 7, S₂: F = 0

P(F | N)

Summarizing a Distribution

moments

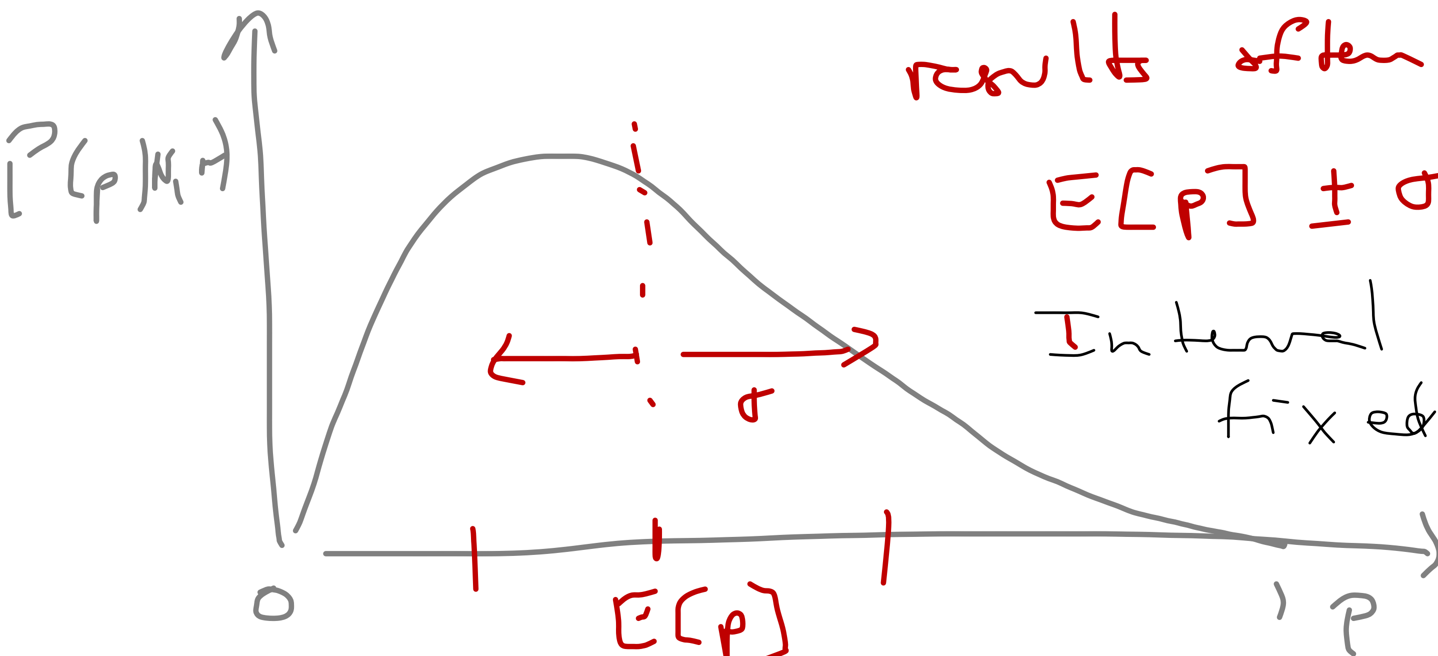
- Mean

Binomial

- rms

$$E[p] = \int_0^1 \underbrace{pP(p|N,r)}_{\text{Posterior pdf}} dp \qquad \sigma = \sqrt{\underbrace{E[p^2] - E[p]^2}_{2^{nd}\text{ "central" moment}}}$$

first moment



results often quoted

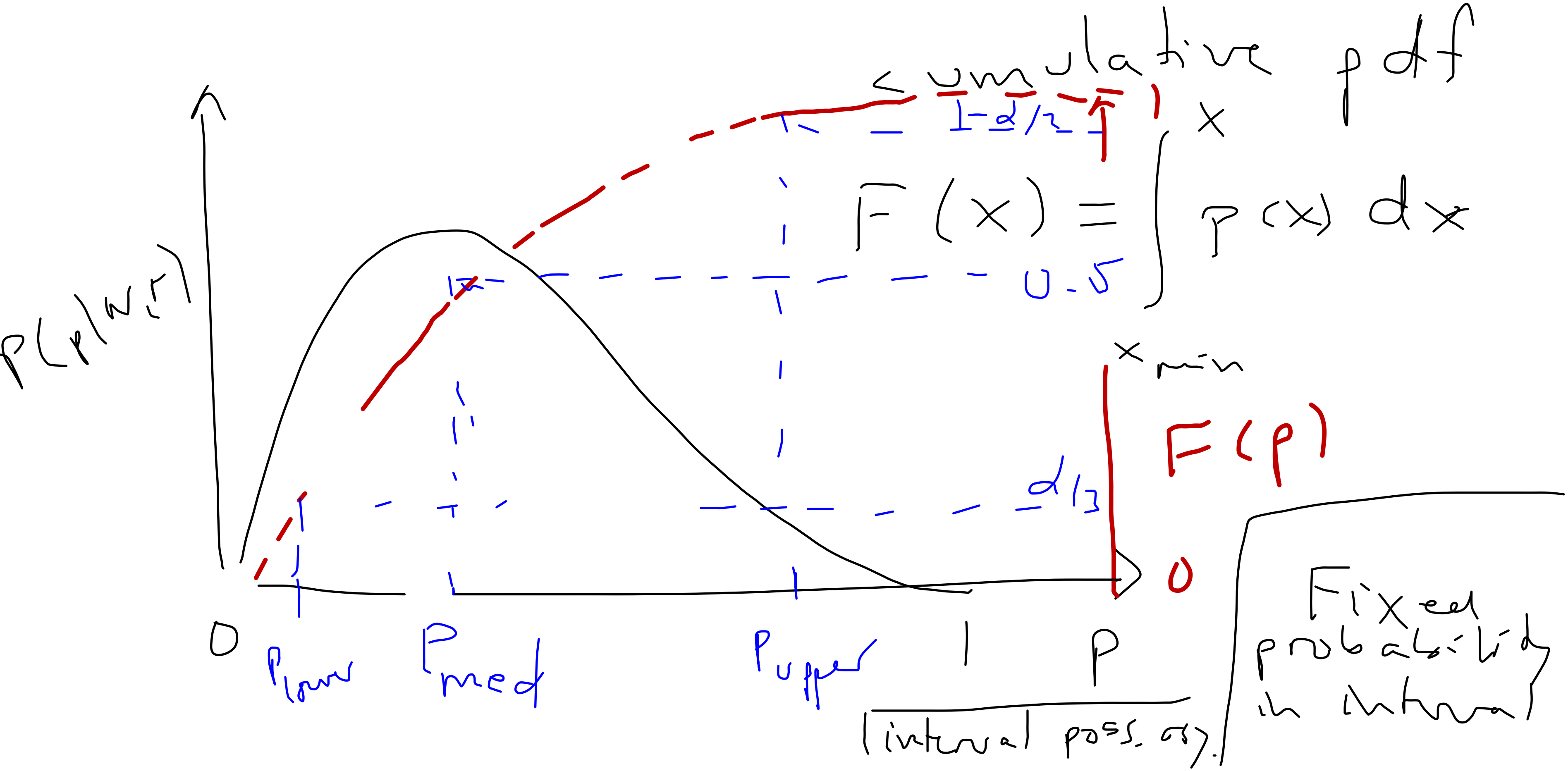
$E[p] \pm \sigma$

Interval has no fixed probability content

Summarizing a Distribution

- Median
- Central interval

$$F(p_{\text{median}}) = 0.5 \quad F(p_{\text{lower}}) = \frac{\alpha}{2} \quad F(p_{\text{upper}}) = 1 - \frac{\alpha}{2}$$



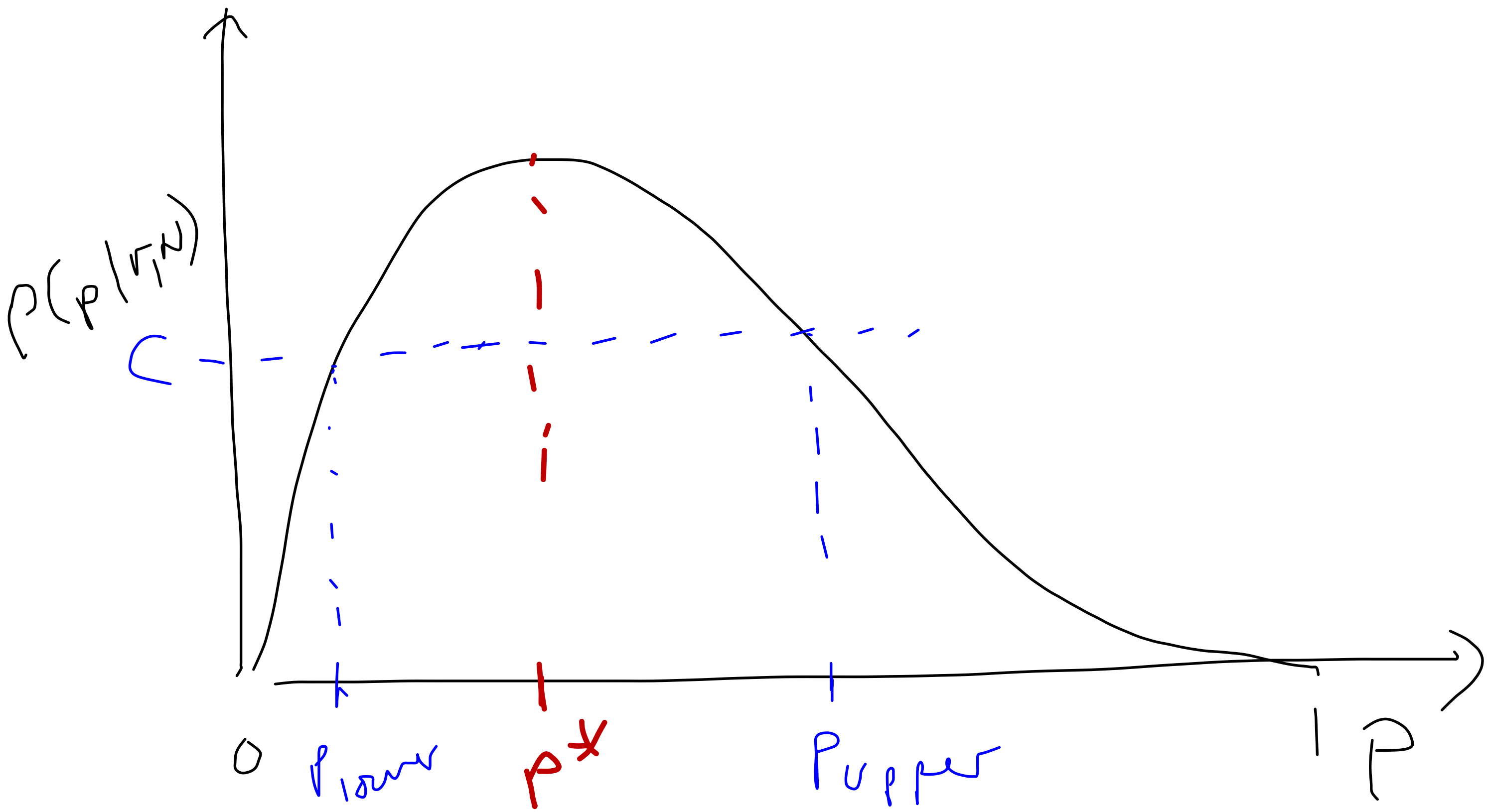
Summarizing a Distribution

- Most-probable value (mode)

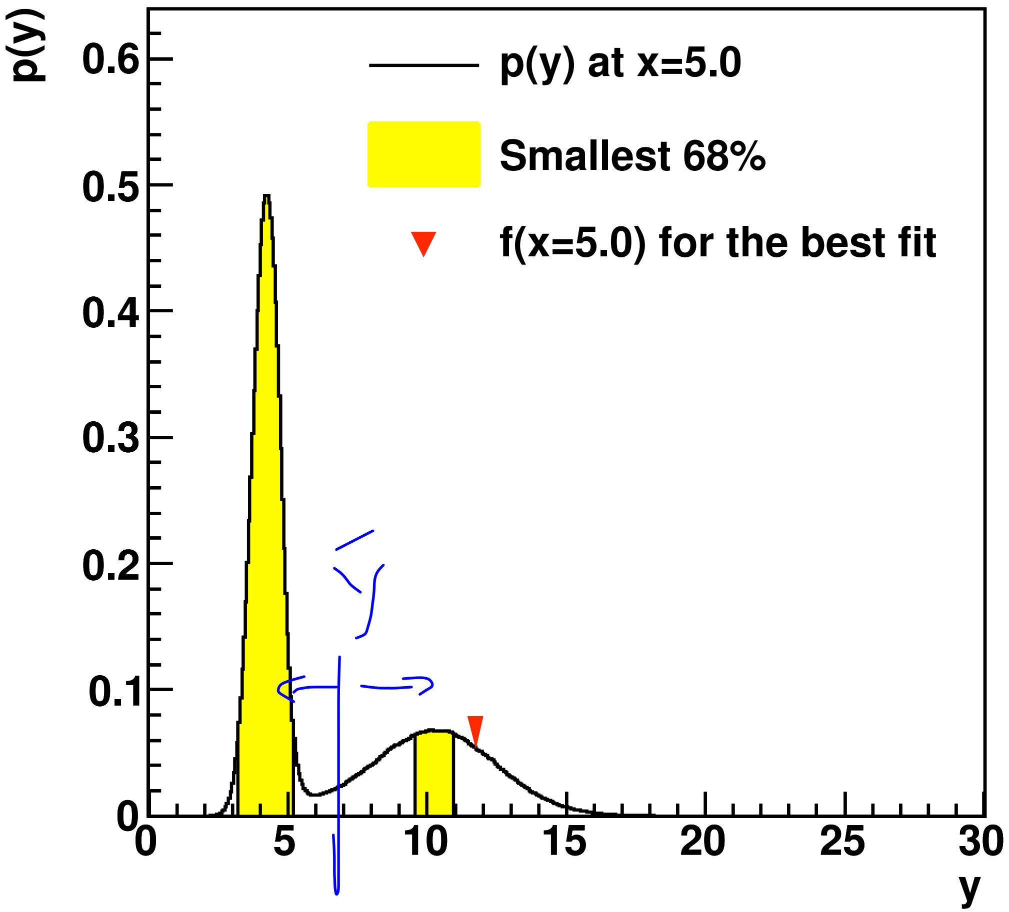
$$p_{\rm mode} = \max_p \left[P(p|r,N) \right]$$

- Shortest interval(s)

$$1-\alpha = \int_{P(p|r,N)>C} P(p|r,N)\,dp$$



Example – multimodal distribution

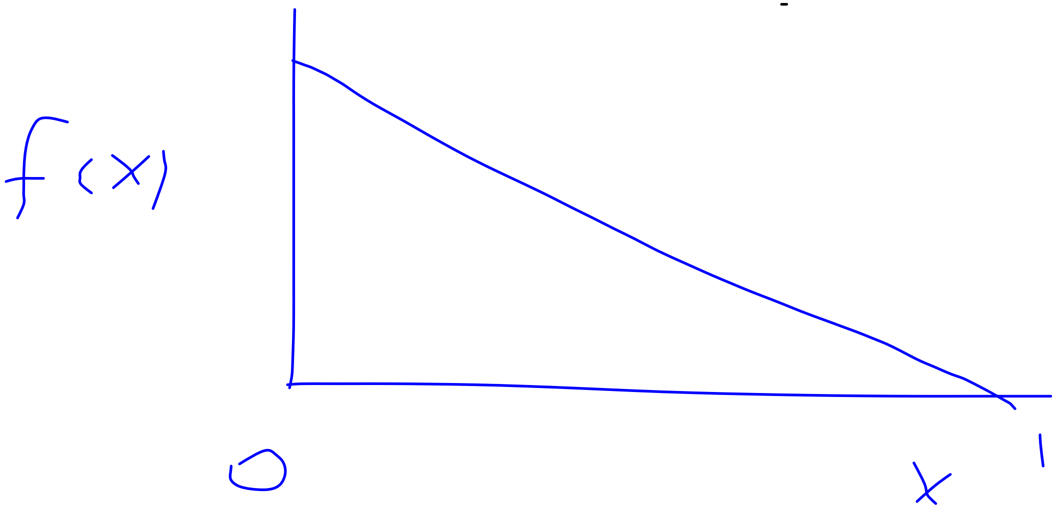


Exercise – compare the mode & smallest 68% probability interval; the mean and rms; and the median and central 68% probability interval for the function

$$f(x) = 1 - x \qquad 0 \le x \le 1$$

warning → not normalized!

i.e., normalize it!



ANSWERS:
mode $x^* = 0$, shortest int $[0, 0.43]$
mean $\langle x \rangle = 0.3\bar{3}$, $\sigma = 0.1\bar{6}$
$x_{med} = 0.29$, central int $[0.083, 0.60]$

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 →∞ and p →0, but Np fixed and finite. Then

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

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

luminosity

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

cross section

n

Poisson Distribution - derivation

$\nu = Np$

$p = \frac{\nu}{N}$

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

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

$N \longrightarrow \infty$

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

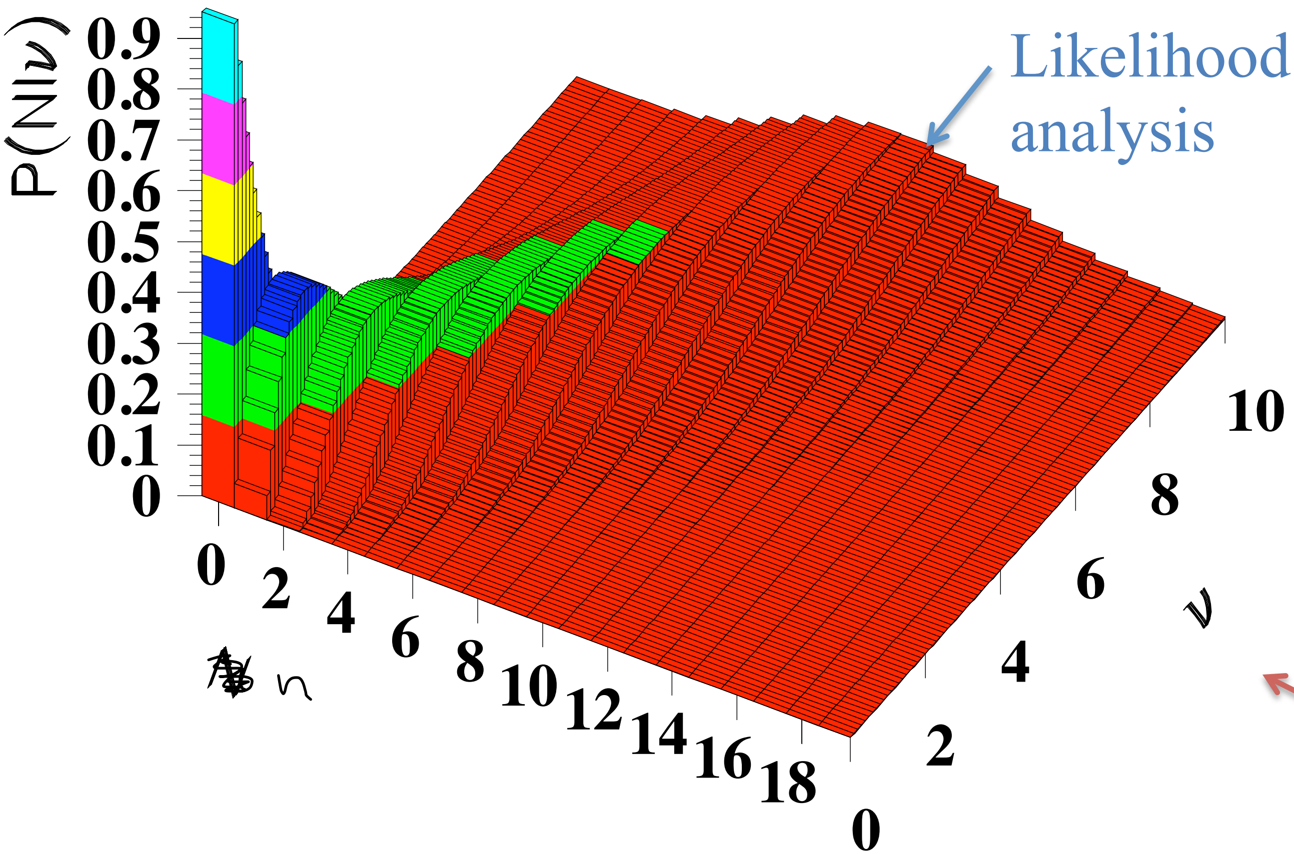
$$\left(1 - \frac{\nu}{N}\right)^{N-n} \rightarrow \left(1 - \frac{\nu}{N}\right)^{N} \longrightarrow e^{-\nu}$$

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

Poisson Distribution

Poisson Example

fixed ν
→ discrete probs
for n



Likelihood used in likelihood analysis

probability density
n fixed,
fun chang
ν

Probability of the data used in confidence level setting

Poisson Distribution-cont.

e.g.,

$\nu = 3.3$

$\lfloor \nu \rfloor = 3$

$\lceil \nu \rceil = 4$

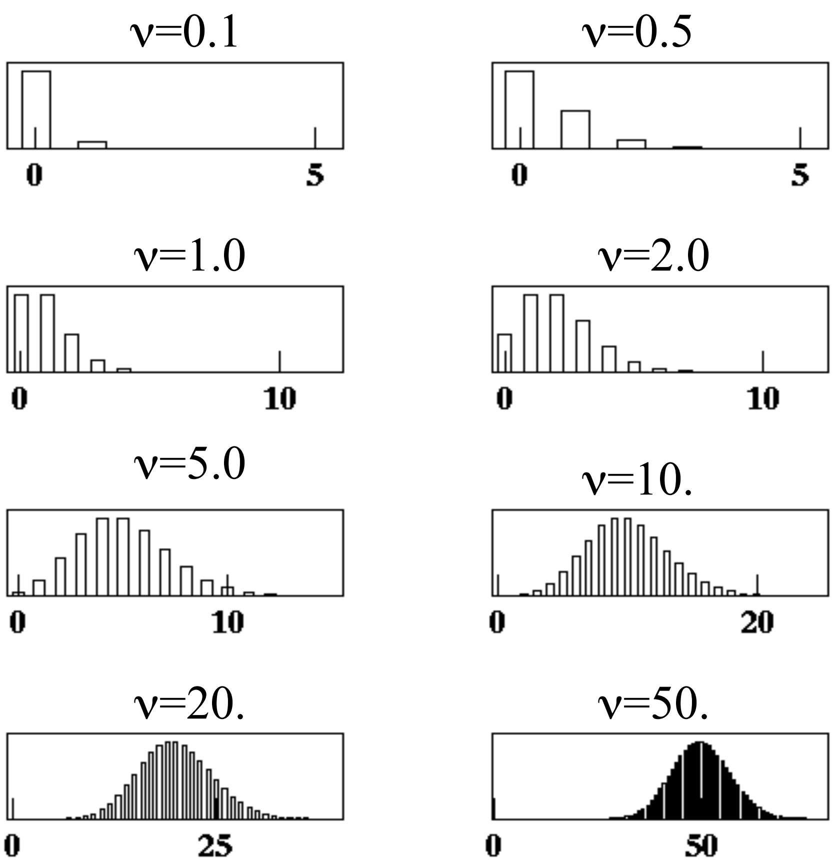
$$P(n \mid \nu) = \frac{\nu^n e^{-\nu}}{n!}$$

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

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

$$E[n] = \nu$$

$$\sigma^2 = \nu$$



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.

Example

Example for ν=10/3

Ordinal cumulative probability

rank

Cumulative probability according to rank

ϕ	$P(\phi\|\nu)$	$F(\phi\|\nu)$	R	$F_R(\phi\|\nu)$
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

Central Interval

1 − α probability interval

ν for Poisson case

$$r_1 = \sup_{r \in 0,...,N} \{ \sum_{i=0}^{r} P(i|N,p) \leq \alpha/2 \} + 1$$

First member of set for central interval

$$P(r=0|N,p) > \alpha/2 \rightarrow r_1 = 0$$

$$r_2 = \inf_{r \in 0,...,N} \{ \sum_{i=r}^{N} P(i|N,p) \leq \alpha/2 \} - 1$$

$$P(r=N|N,p) > \alpha/2 \rightarrow r_2 = N$$

$$\mathcal{O}^{C}_{1-\alpha} = \{ r_1, r_1+1, ..., r_2 \}$$

central interval @ 1 − α probability

Smallest Interval

$$\mathcal{O}^{\mathrm{S}}_{1-\alpha} = \{\cancel{r}n^*\} \qquad P(\mathcal{O}^{\mathrm{S}}_{1-\alpha}|\cancel{N,}p) \overset{?}{\geq} 1-\alpha$$

$$P(\cancel{r}n^* + 1|N,p) \overset{?}{>} P(\cancel{r}n^* - 1|N,p)$$

$$\mathcal{O}^{\mathrm{S}}_{1-\alpha} = \{\cancel{r}n^*, \cancel{r}n^* + 1\} \qquad \mathcal{O}^{\mathrm{S}}_{1-\alpha} = \{\cancel{r}n^*, \cancel{r}n^* - 1\}$$

$$P(\mathcal{O}^{\mathrm{S}}_{1-\alpha}|N,p) \overset{?}{\geq} 1-\alpha$$

Exercise

What is the $1-\alpha=0.9$ central interval; smallest interval ?

ANSWER!

Central Interval

$\{1,2,3,4,5,6,7\}$

Shortest Int.

$\{1,2,3,4,5,6\}$

o	$P(o\|\nu)$	$F(o\|\nu)$	R	$F_R(o\|\nu)$
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

Confidence Interval Calculation

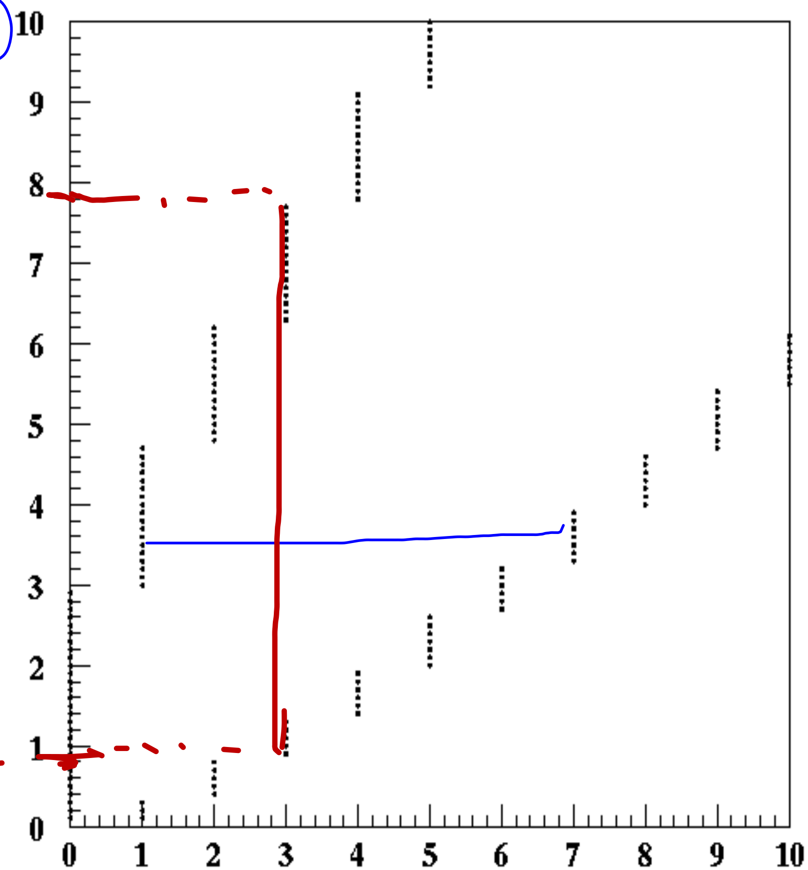
We observe x events, and ask which values of ν are accepted with confidence level $1\text{-}\alpha$. For $1\text{-}\alpha=0.9$, central intervals:

Poisson mean

Neymann

ν_{up}

Poisson case



$P(n=0) < 5\%$

expt returns $x=3$

ν_{down}

$1-\alpha$ interval $[\nu_d, \nu_u]$

events

Poisson Distribution-cont.

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

$P(n_s|\nu_s) = \frac{e^{-\nu_s}\nu_s^{n_s}}{n_s!}$

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

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

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

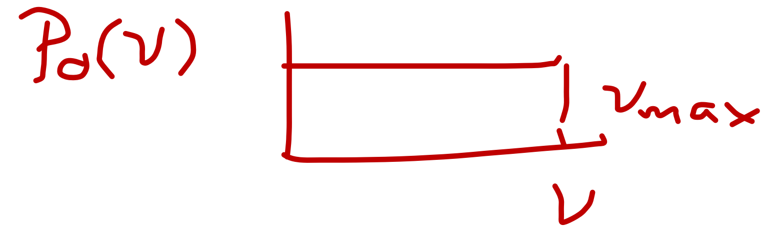
$$= e^{-(\nu_B+\nu_s)} \frac{(\nu_s+\nu_B)^n}{n!} \underbrace{\overbrace{\sum_{n_s=0}^{n} \frac{n!}{n_s!(n-n_s)!}\left(\frac{\nu_s}{\nu_s+\nu_B}\right)^{n_s}\left(\frac{\nu_B}{\nu_s+\nu_B}\right)^{n-n_s}}^{\text{Binomial formula with } p=\left(\frac{\nu_s}{\nu_s+\nu_B}\right)}}_{=1 \text{ by normalization}}$$

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

$\nu = \nu_B + \nu_s$

Bayesian Data Analysis-Poisson Distribution

Typical examples – counting experiments such as source activity, 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.

Poisson - cont.

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

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

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

Bayes with flat priors $\to \mathcal{L}$

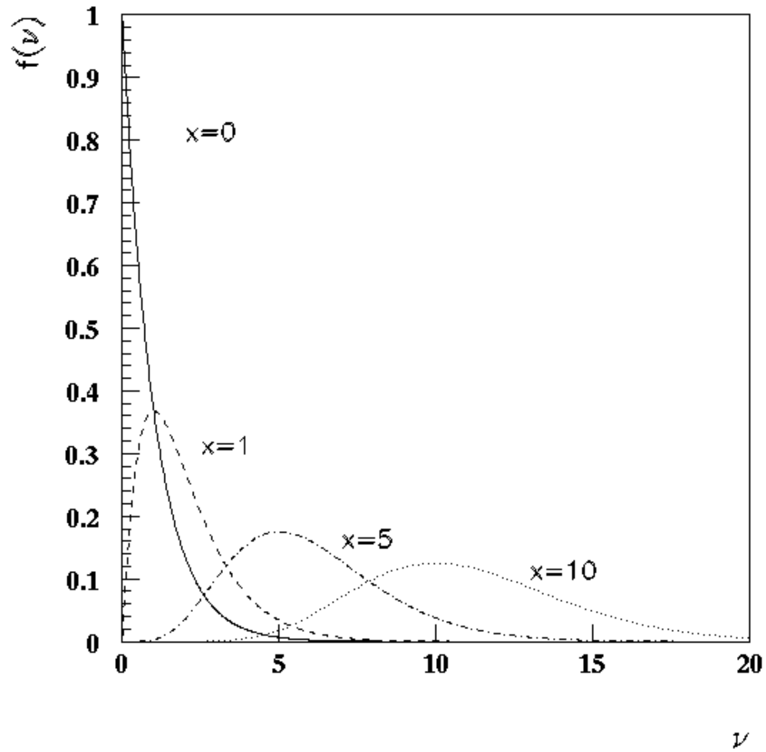
$$\underline{P(\nu|n) = \frac{e^{-\nu}\nu^n}{n!}}$$

$$\boxed{\nu^* = n}$$

same form as $P(n|\nu)$ $\mathcal{L}(\nu)$

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

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

Poisson - cont.

The cumulative distribution function:

Needed for credibility intervals

$P(\nu|n)$

$$F(\nu|n) = \int_0^{\nu} \frac{\nu'^n e^{-\nu'}}{n!} d\nu'$$

$$= \frac{1}{n!}\left[-\nu'^n e^{-\nu'}|_0^{\nu} + n\int_0^{\nu} \nu'^{n-1} e^{-\nu'} d\nu'\right]$$

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

Poisson – Examples

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

First, no background, measure zero counts.

With flat prior assumption

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

$0! = 1$

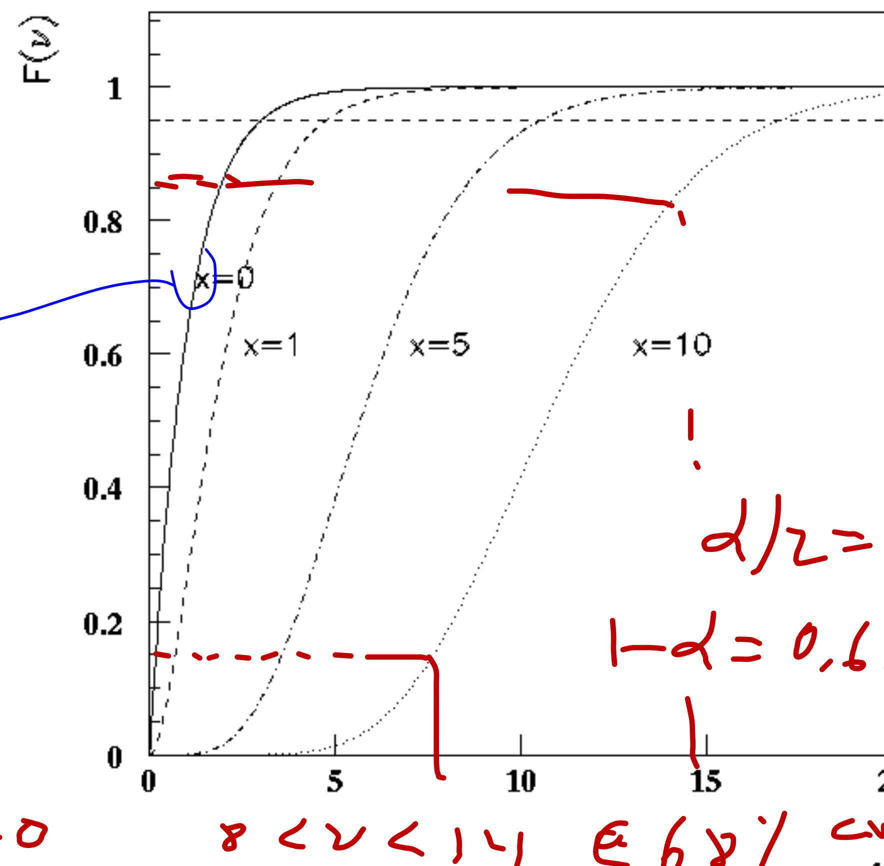
$\nu^0 = 1$

For a 95% upper limit

$$0.95 = 1 - e^{-\nu}$$
$$\nu \approx 3$$

$\nu \leq 3$ @ 95% credibility

for $n = 0$



n

$\alpha/2 = 0.16$

$1-\alpha = 0.68$

$n=0$ $8 < \nu < 14.1 \in 68\%$ cred

Poisson – cont.

And now suppose we have background:

background

$$\mu = \lambda + \nu, \quad P(x \mid \mu) = \frac{e^{-\mu}\mu^x}{x!}$$

signal

background known exactly

$$P(\nu \mid x, \lambda) = \frac{\left(e^{-(\lambda+\nu)}(\lambda+\nu)^x \big/ x!\right) P_0(\nu)}{\int_0^\infty \left(e^{-(\lambda+\nu)}(\lambda+\nu)^x \big/ x!\right) P_0(\nu)\,d\nu}$$

$P_1(\lambda) = \delta(\lambda' - \lambda)$

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

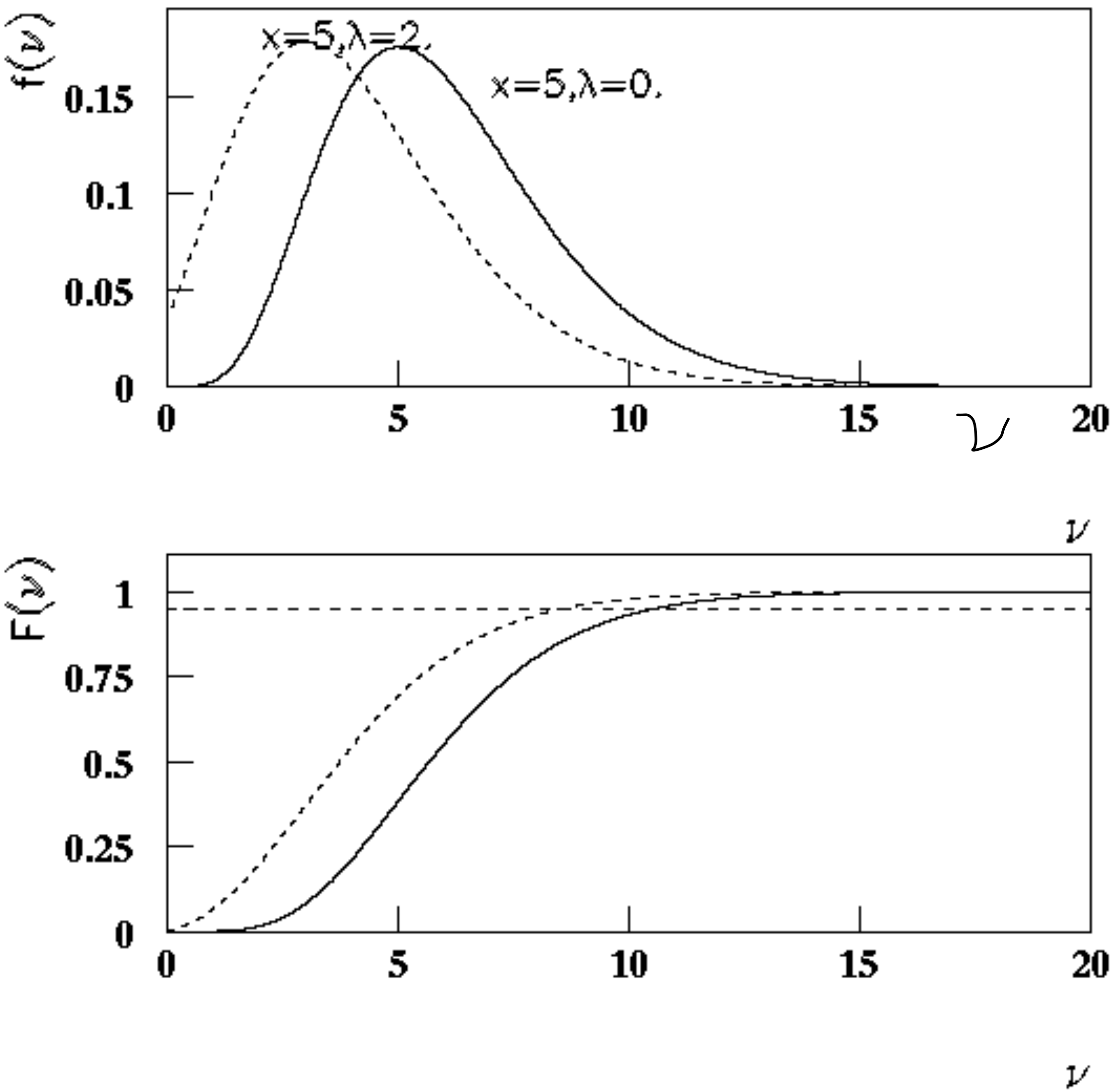
assuming a flat $P_0(\nu)$ and integrating by parts.

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

For setting intervals on signal when have known backgrounds

Poisson – cont.

$P(\nu|x,\lambda)$



Comment:

For $\boxed{x=0,\ P(\nu|x,\lambda)=e^{-\nu}}$. It does not matter how much background you have, you get the same probability distribution for the signal. Source of much confusion & discussion (very different for Confidence Level calculation).

Exercise

Imagine two measurements are performed where the same Poisson mean, v, is expected. The measurements yield n_1 and n_2 events. Starting with a flat prior for the Poisson mean, find the resulting posterior pdf.

$P(\nu|n_1)$ for flat prior

→ use as prior for 2^{nd} meas.

How does it compare to running the experiment twice as long (expectation 2v) and measuring n_1+n_2 events ? (starting with flat prior)

(you will not have enough time now for these calculations – set up the formulas and work them out when you have time)

$$y = y(x)$$

$$P(y)\,dy = P(x)\,dx$$

$$P(y) = P(x)\left(\frac{dy}{dx}\right)^{-1}$$

ANSWER:

$$P(\nu) = \frac{2e^{-2\nu}(2\nu)^{n_1+n_2}}{(n_1+n_2)!}$$

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:

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

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

λ is not fixed → uncertain

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

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

background signal

Example

Compare two situations:

1) no knowledge on the background

2) Separate data help us constrain the background

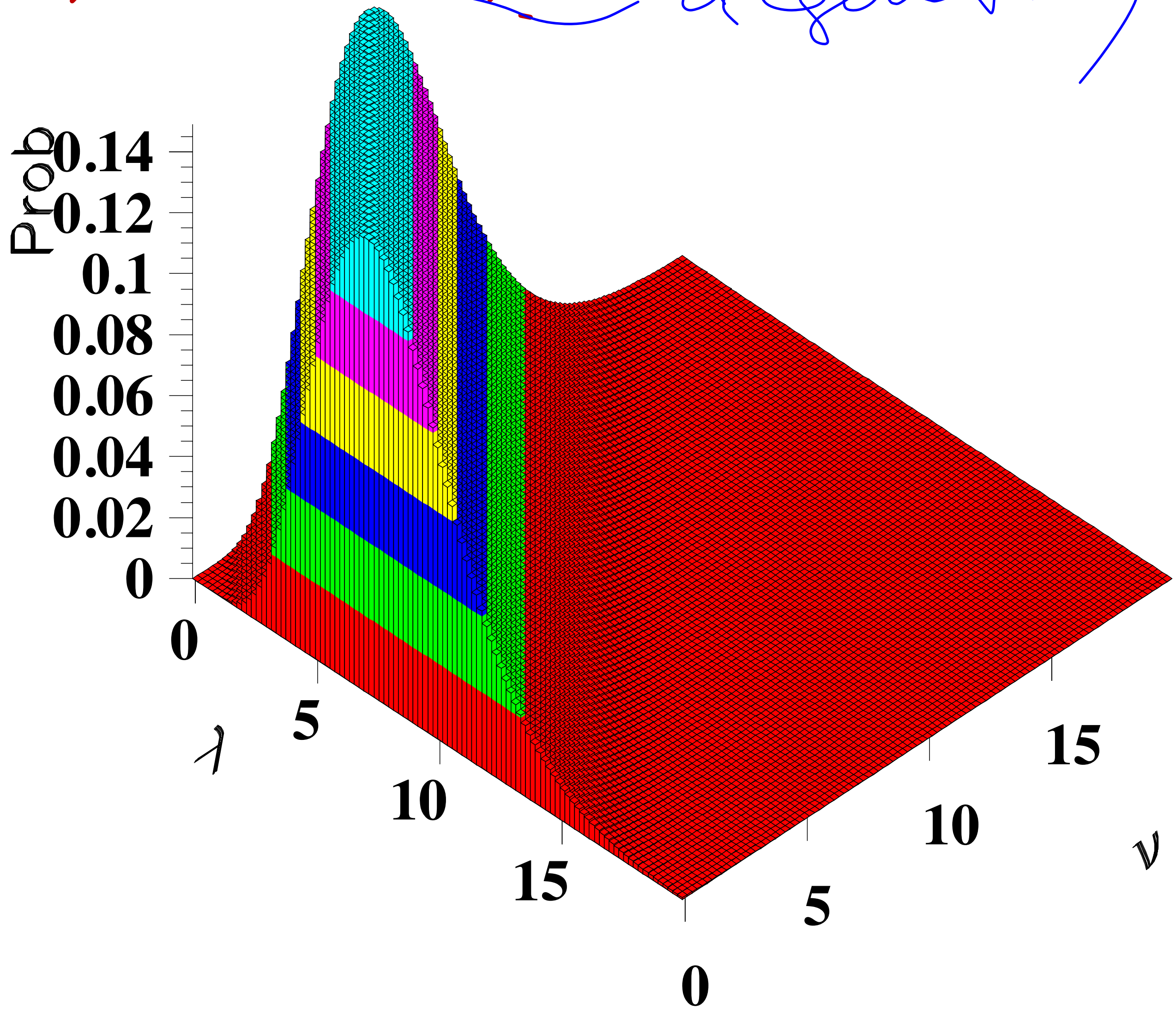
Suppose we measure n=7 events, what can we say ?

N=7 Poisson

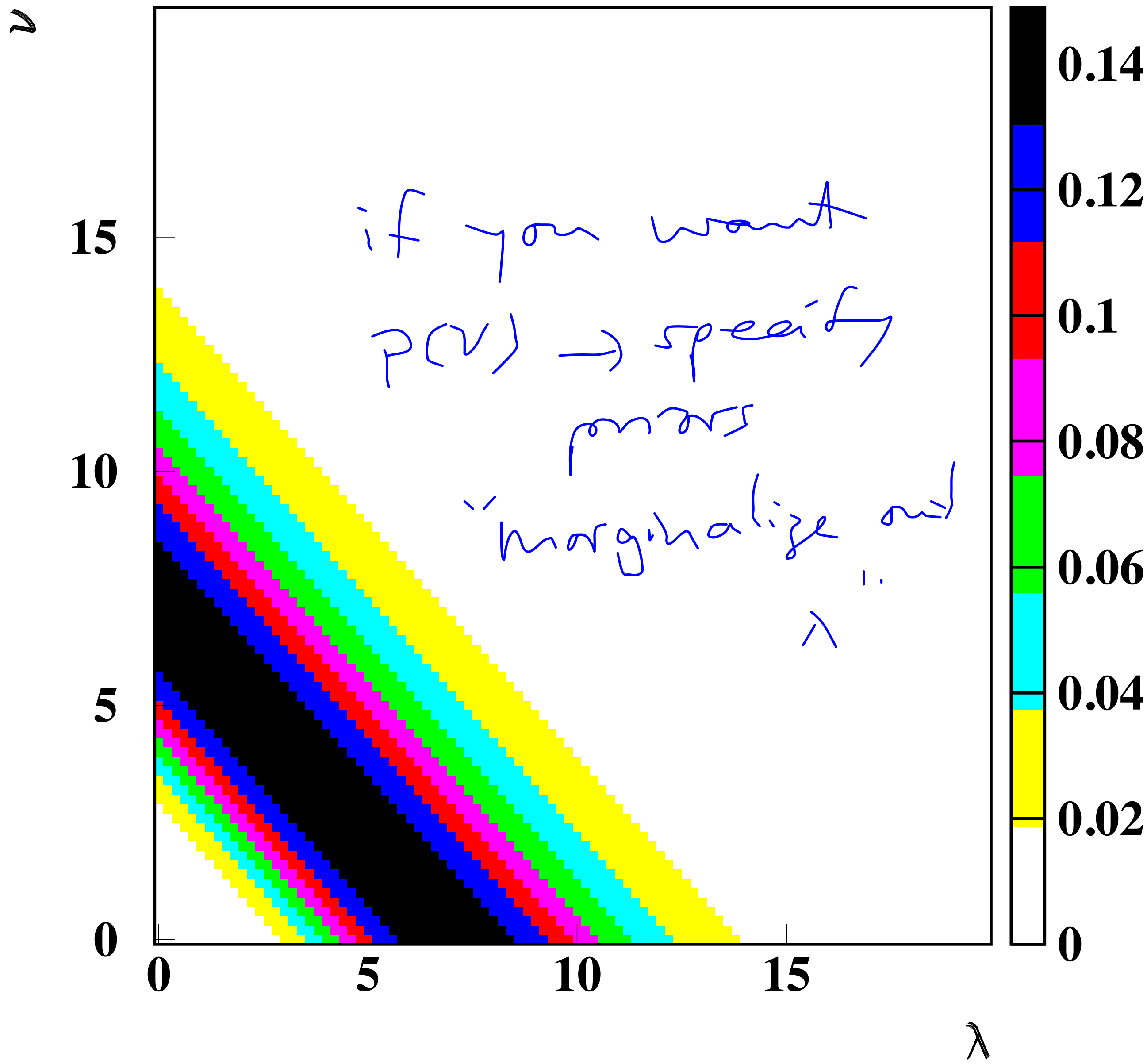
$$P(7 \mid \lambda, \nu) = \frac{e^{-(\lambda+\nu)} (\lambda+\nu)^7}{7!}$$

degeneracy

$\nu + \lambda = 7$



N=7



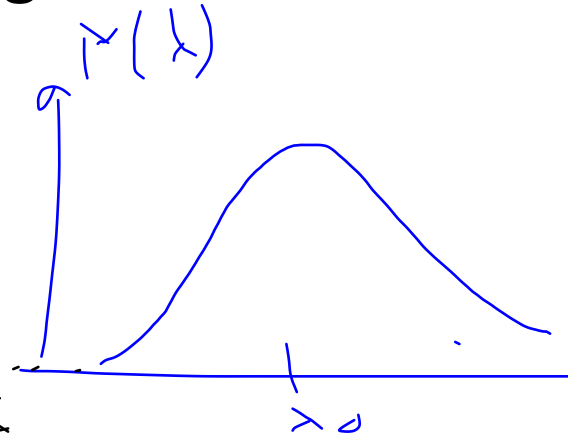
$P(7|\nu,\lambda)$

With Background knowledge

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

$\lambda_0 = 3, \quad \sigma_\lambda = 1$

$\int \frac{1}{\sqrt{2\pi}\sigma_\lambda} \cdots$



Can build this into the likelihood (e.g., frequentist analysis) or call it prior knowledge (either way for Bayes)

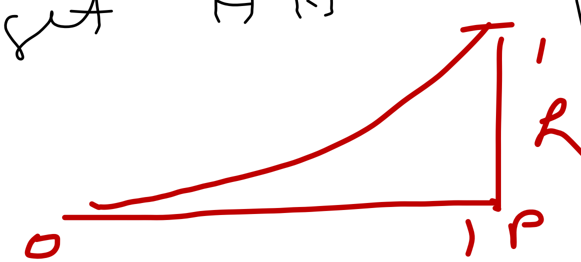
$$\mathcal{L}(\nu,\lambda) = P(N|\nu,\lambda)P(\lambda)$$

Digression

Note: a likelihood is not a probability !

Toss coin, get HH $\quad p(HH|2,p) = p^2$

$= \mathcal{L}(p)$



With Background knowledge

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

ν "physics par"

λ "nuisance par"

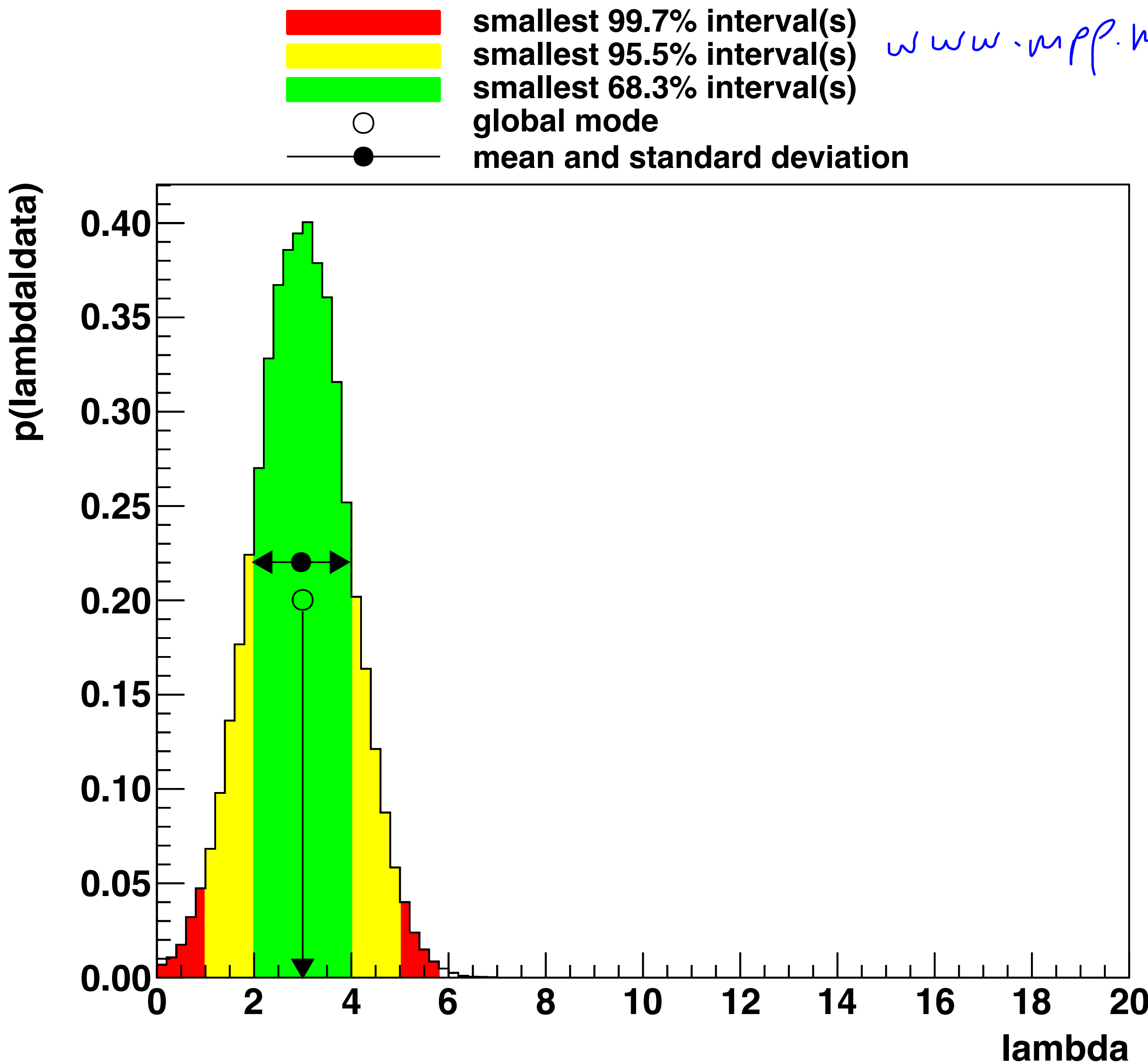
Can build this into the likelihood (Frequentist Analysis) or call it prior knowledge (either way for Bayes)

$$P(\nu, \lambda | \underset{7}{N}) = \frac{P(N|\nu,\lambda)P(\lambda)P(\nu)}{\int P(N|\nu,\lambda)P(\lambda)P(\nu)d\lambda d\nu}$$

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

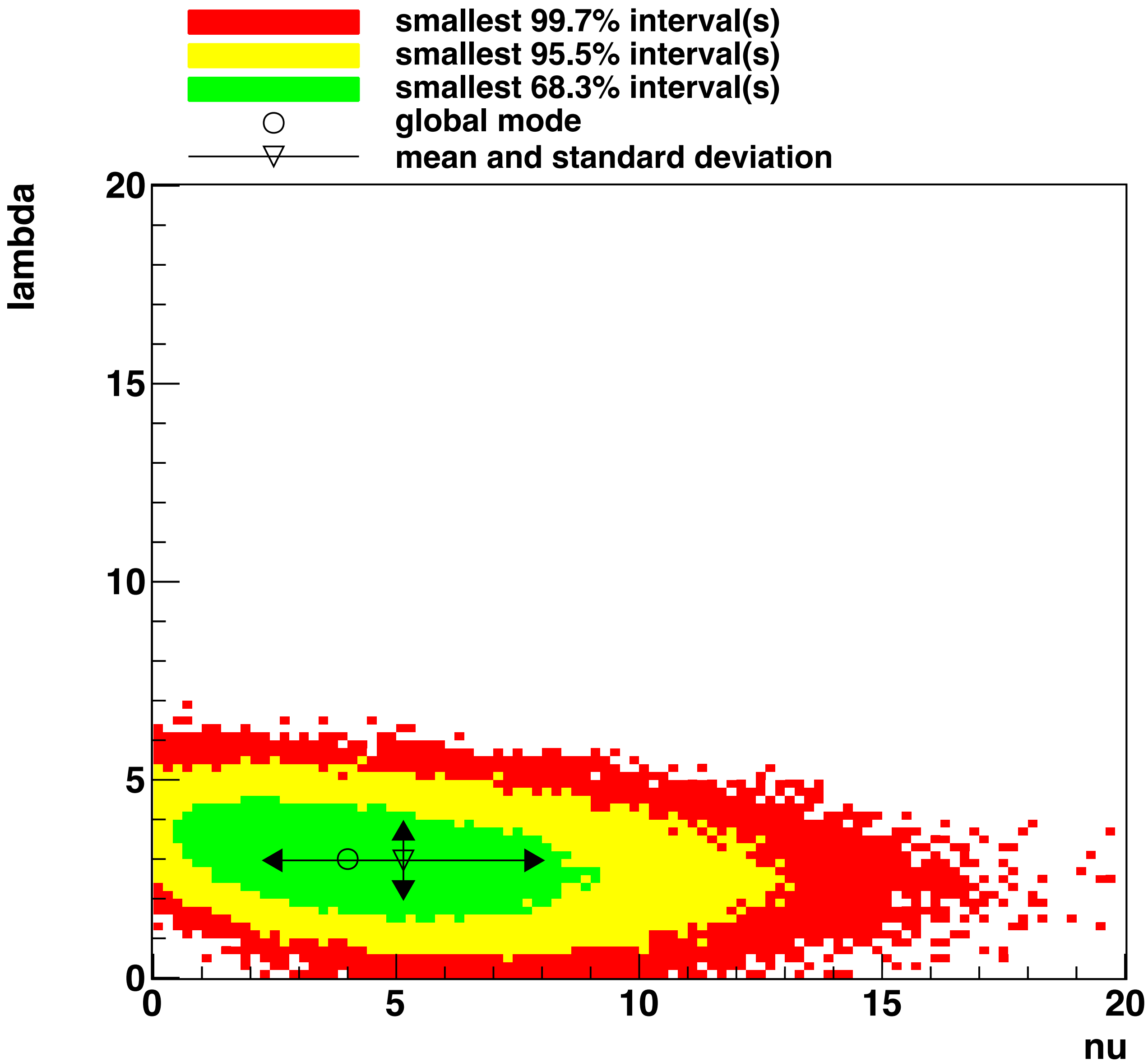


www.mpp.mpg.de/bat

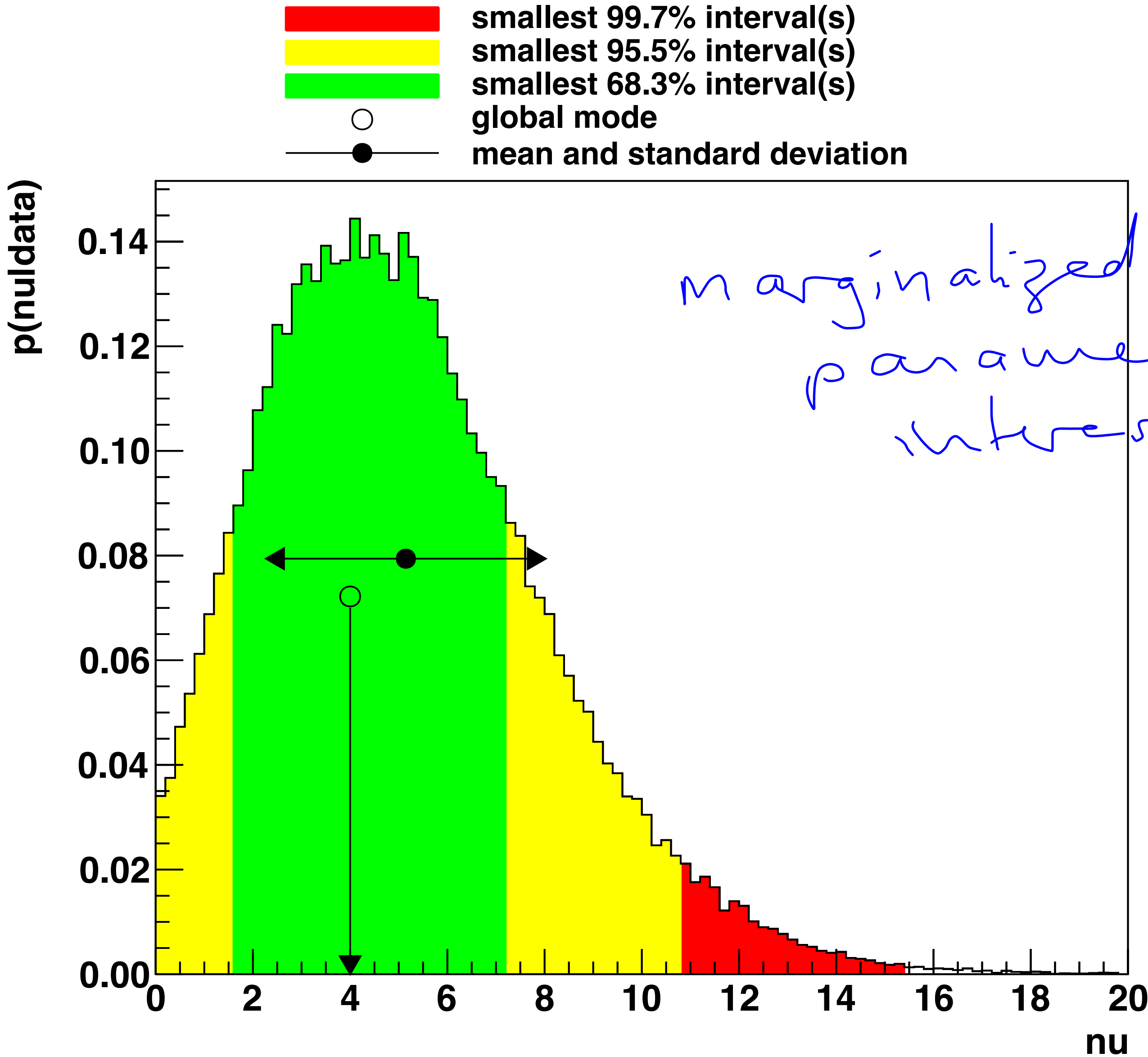
Markov Chain

⋮

N=7 Constrained Background



N=7 Constrained Background



marginalized for parameter of interest

That’s it

Enjoy the school !