## 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





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



Experiment

# How we learn



# 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  $A_1 \rightarrow C, A_2 \rightarrow C, \dots$ 

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



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

Then use 
$$P(\lambda|M, D) = \frac{P(D|\lambda, M)P_0(\lambda|M)}{P(D|M)}$$

$$Z = P(D|M) = \int d\lambda P(D|\lambda, M) P_0(\lambda|M)$$
 "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.,  $5\sigma$  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 ``threesigma 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:

#### ТНТННТНТН

We now repeat the process with a different coin and get

### ТТТТТТТТТТТ

Which outcome has higher probability?

Take a model where H, T are equally likely. Then,

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

And

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

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:

$$\left(\begin{array}{c}10\\n\end{array}\right)\left(\frac{1}{2}\right)^{10}$$

And we find the following result:

n	р
0	$1 \cdot 2^{-10}$
1	$10.2^{-10}$
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

### Summarizing a Distribution

• Mean • rms  $E[p] = \int_0^1 pP(p|N,r)dp \quad \sigma = \sqrt{E[p^2] - E[p]^2}$ 

## 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) • Shortest interval(s)  $p_{\text{mode}} = \max_{p} \left[ P(p|r, N) \right] \quad 1 - \alpha = \int_{P(p|r, N) > C} P(p|r, N) dp$ 

### 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

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

# **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  $N \rightarrow \infty$  and  $p \rightarrow 0$ , 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

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

## **Poisson Distribution - derivation**

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

 $N \to \infty$ 

$$\frac{N!}{(N-n)!} = N \cdot (N-1) \cdot \dots \cdot (N-n+1) \approx N^n$$
$$\left(1 - \frac{\nu}{N}\right)^{N-n} \to \left(1 - \frac{\nu}{N}\right)^N \to e^{-\nu}$$
$$P(n|\nu) = \frac{e^{-\nu}\nu^n}{n!}$$
Poisson Distribution

# **Poisson Example**



# Poisson Distribution-cont.





















Û

 $r^* = \lfloor \nu \rfloor$  $r^* = \lceil \nu \rceil - 1$  $E[n] = \nu$  $\sigma^2 = \nu$ 

#### Notes:

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

## Example for v=10/3

Example					
Ordinal cumulative probability			rank		Cumulative probability according to
0	$P(o \nu)$	$F(o \nu)$	$\mid R$	$F_R(o \nu)$	Tank
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**

$$r_{1} = \sup_{r \in 0, \dots, N} \{ \sum_{i=0}^{r} P(i|N, p) \le \alpha/2 \} + 1$$
$$P(r = 0|N, p) > \alpha/2 \to r_{1} = 0$$

$$r_{2} = \inf_{r \in 0, \dots, N} \{ \sum_{i=r}^{N} P(i|N, p) \le \alpha/2 \} - 1$$
$$P(r = N|N, p) > \alpha/2 \to r_{2} = N$$

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

# **Smallest Interval**

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

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

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

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

#### Exercise

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

0	$  P(o \nu)$	$\mid F(o \nu)$	$\mid R$	$F_R(o \nu)$
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 v are accepted with confidence level 1- $\alpha$ . For 1-  $\alpha$ =0.9, central intervals:



# 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

$$P(n) = \sum_{n_s=0}^{n} P(n_s | v_s) P(n - n_s | v_B)$$
  
=  $e^{-(v_B + v_s)} \sum_{n_s=0}^{n} \frac{v_s^{n_s} v_B^{n-n_s}}{n_s!(n - n_s)!}$   
Binomial formula with  $p = \left(\frac{v_s}{v_s + v_B}\right)$   
=  $e^{-(v_B + v_s)} \frac{(v_s + v_B)^n}{n!} \sum_{n_s=0}^{n} \frac{n!}{n_s!(n - n_s)!} \left(\frac{v_s}{v_s + v_B}\right)^{n_s} \left(\frac{v_B}{v_s + v_B}\right)^{n-n_s}$   
=  $e^{-(v_B + v_s)} \frac{(v_s + v_B)^n}{n!}$   
= 1 by normalization

## **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 v 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.



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

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

#### Poisson - cont.

The cumulative distribution function:

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

### Poisson – Examples

First, no background, measure zero counts.

With flat prior assumption

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

For a 95% upper limit

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



 $\nu$ 

#### Poisson – cont.

And now suppose we have background:

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

$$P(\nu \mid x, \lambda) = \frac{\left(e^{-(\lambda + \nu)}(\lambda + \nu)^{x}/x\right)}{\int_{0}^{\infty} \left(e^{-(\lambda + \nu)}(\lambda + \nu)^{x}/x\right)} P_{0}(\nu) d\nu$$

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

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

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

#### Poisson – cont.



Comment:

For x=0,  $P(v|x, \lambda)=e^{-v}$ . 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.

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

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

# 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



N=7



# With Background knowledge

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

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

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

Note: a likelihood is not a probability !

# With Background knowledge

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

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

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

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



#### That's it

#### Enjoy the school !