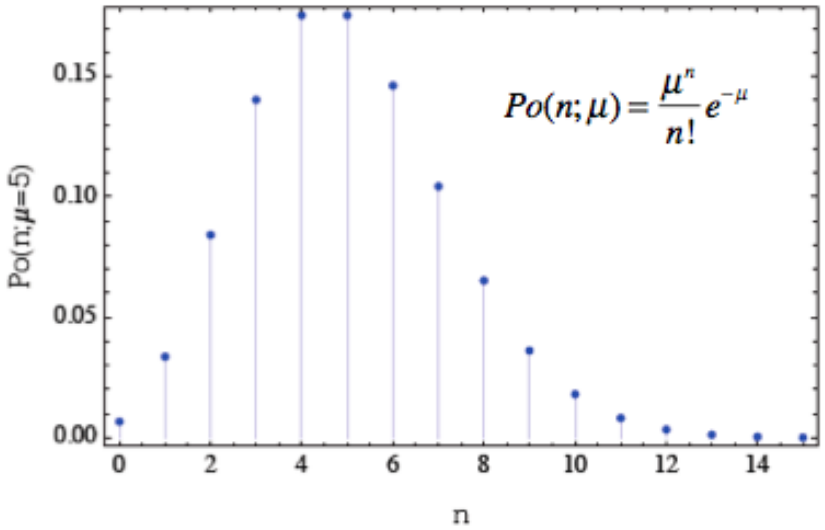
**The FREQUENTIST example, by J.J.Gomes.Cadenas**

http://www.pit.physik.uni-tuebingen.de/grabmayr/workshop/talks/gomez-statistics.pdf

# Poisson statistics

$N \sim 5$ for $T_{1/2}^{0\nu} \sim 10^{27}\ y$




17.5% of the experiments will observe 5 events

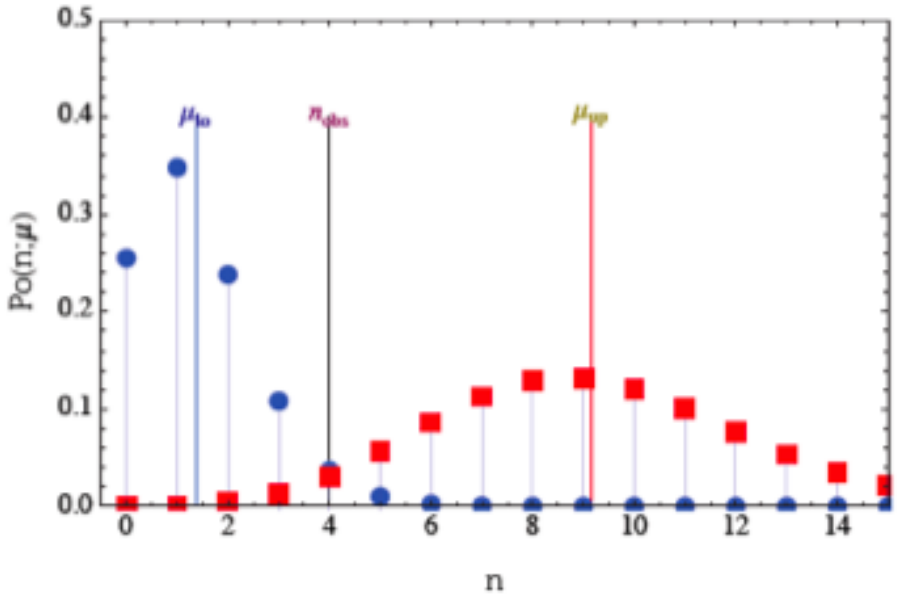
17.5% of the experiments will observe 4 events

0.7% of the experiments will observe 0 events

Different identical experiments running for the same total exposure will observe different number of events

# Upper and lower limit

Suppose that a given experiment observes n=4 events. What can inferred about the true value, $\mu$?




$Po(n;\mu_{lo}=1.37)$

90% of the blue dots are below n=4 (lower limit at 90% CL)

$Po(n;\mu_{up}=9.15)$

90% of the red dots are above n=4 (upper limit at 90% CL)

$$\sum_{nobs}^{\infty} Po(n;\mu_{up}) \rightarrow Po(4;\mu_{up}) + Po(5;\mu_{up}) + ... \geq \alpha(0.9)$$

$$\sum_{0}^{nobs} Po(n;\mu_{lo}) \rightarrow Po(0;\mu_{lo}) + ... + Po(4;\mu_{lo}) \geq \beta(0.9)$$

*Ambiguity in the interval definition: < or ≤ ?*

| | μ=4 | μ=3 | μ=1.74 | μ=5 | μ=8.0 | μ=9.15 |
|---|---|---|---|---|---|---|
| x=0 | 1.8% | 5.0% | 17.6% | 0.7% | 0.0% | 0.0% |
| x=1 | 7.3% | 14.9% | 30.5% | 3.4% | 0.3% | 0.1% |
| x=2 | 14.7% | 22.4% | 26.6% | 8.4% | 1.1% | 0.4% |
| x=3 | 19.5% | 22.4% | 15.4% | 14.0% | 2.9% | 1.4% |
| x>3 | 56.7% | 35.3% | 9.9% | | | |
| x=4 | 19.5% | | | 17.5% | 5.7% | 3.1% |
| Σ(x=0,1,2,3,4) | 62.9% | | | 44.0% | 10.0% | 5.0% |

Observation

**1.74 < μ < 8.0 at 90% C.L.**

# Experiments with background

$$Po(n;\mu+b)=\frac{(\mu+b)^n}{n!}e^{-(\mu+b)}$$

Where mu is the true value of the signal (unknown) and b the expected background (known)

Suppose that we run an experiment in which we expect 5 background events, and observe 5 events. What can we infer about mu?

$$\sum_{nobs}^{\infty}\frac{(\mu_{up}+b)^n}{n!}e^{-(\mu_{up}+b)}\geq\alpha$$

$$\sum_{0}^{nobs}\frac{(\mu_{lo}+b)^n}{n!}e^{-(\mu_{lo}+b)}\geq\beta$$

In principle we can try the same trick that for the case of zero background... right?

# The case n=b

$$\alpha = \sum_{5}^{\infty} \frac{(\mu_{up}+5)^n}{5!} e^{-(\mu_{up}+5)} \rightarrow \mu_{up} = 5.51$$

$$\beta = \sum_{0}^{nobs} \frac{(\mu_{lo}+b)^n}{n!} e^{-(\mu_{lo}+b)} \rightarrow \mu_{lo} = -3.03$$

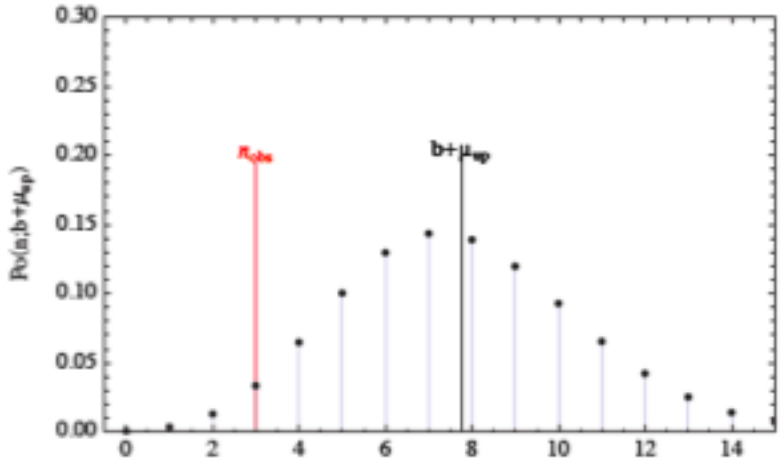
A lower limit cannot be set therefore mu_lo = 0

The upper limit at 95% Cl is 5.51. No problem!

But what happens if n<b?

# The colapse of the classical limit

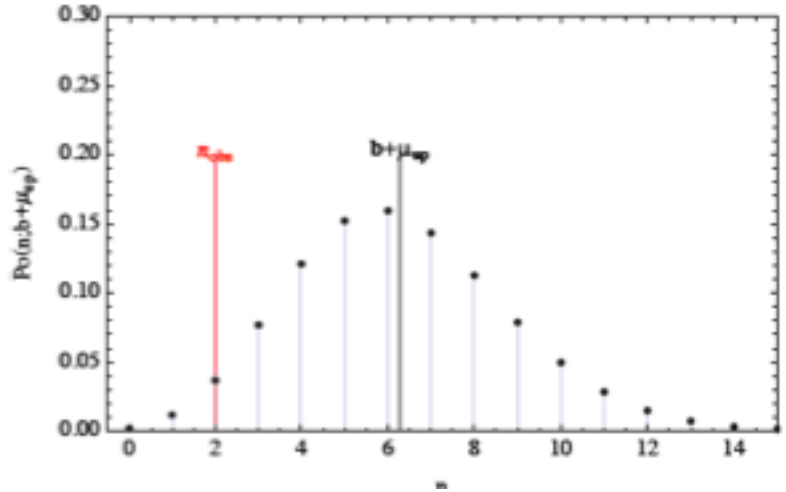
what happens if n<b?




$n_{obs} = 3$

$b + \mu_{up} = 7.75 \rightarrow 95\%$ of black dots above $n_{obs}$

$\mu_{up} = 2.75$



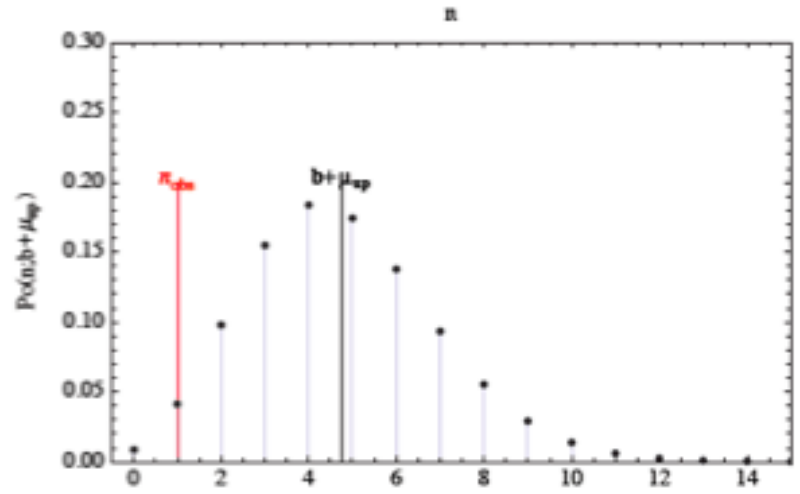

$n_{obs} = 2$

$b + \mu_{up} = 6.29 \rightarrow 95\%$ of black dots above $n_{obs}$

$\mu_{up} = 1.29$

# The colapse of the classical limit

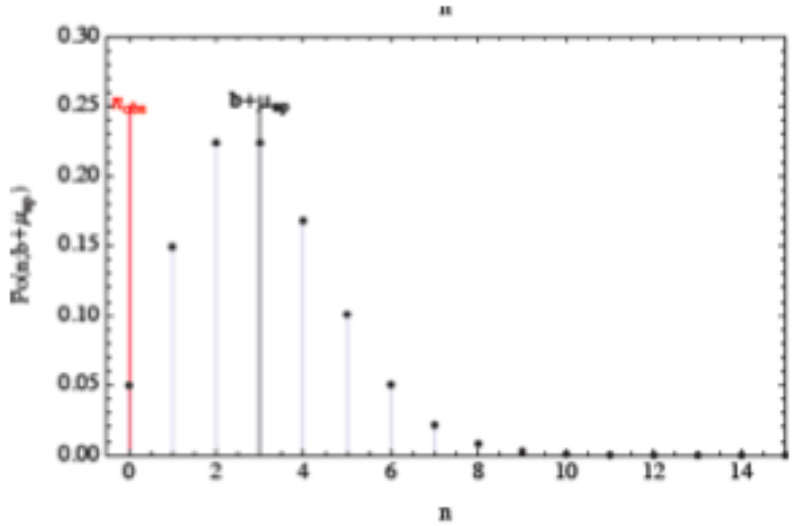
what happens if n<b?




$n_{obs} = 1$

$b + \mu_{up} = 4.75 \rightarrow 95\%$ of black dots above $n_{obs}$

$\mu_{up} = -0.25$



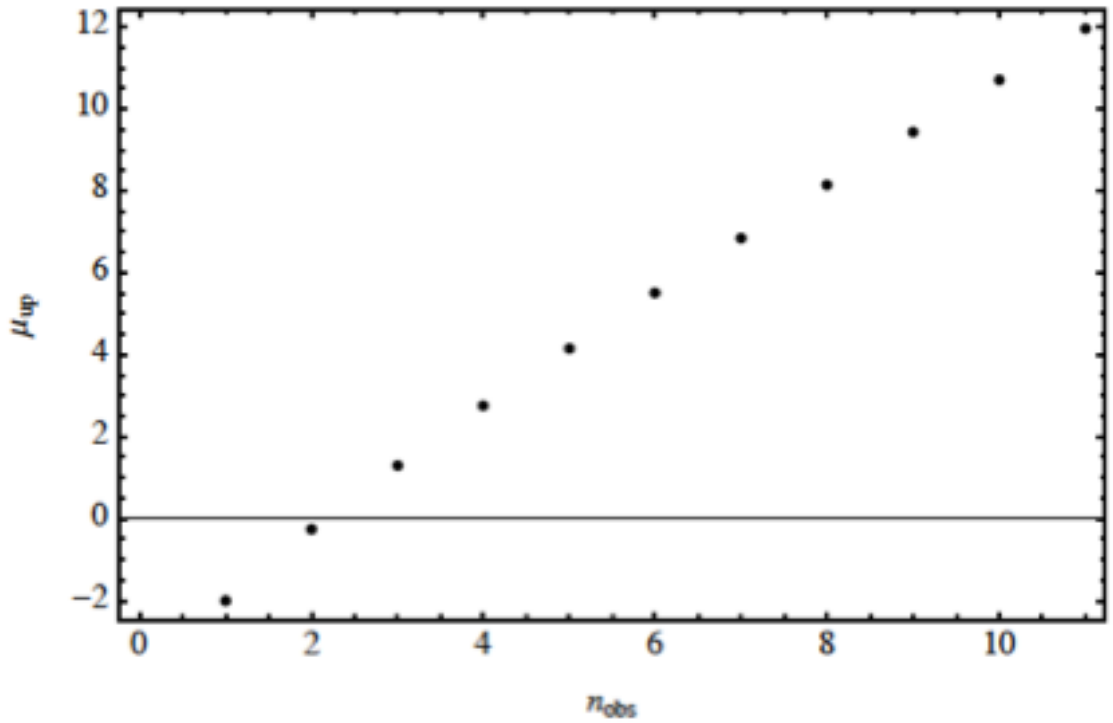

$n_{obs} = 0$

$b + \mu_{up} = 3 \rightarrow 95\%$ of black dots above $n_{obs}$

$\mu_{up} = -2$

# The colapse of the classical limit

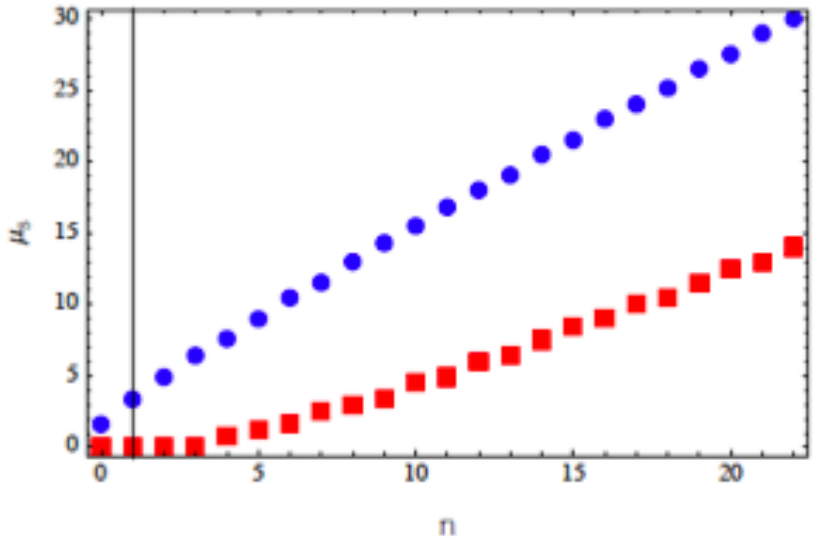
what happens if n<b?

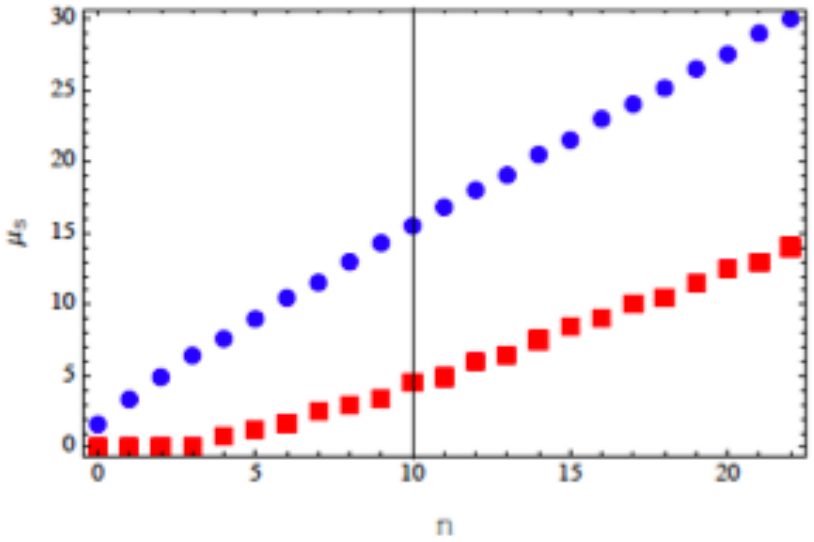


# CONFIDENCE BELTS

$$\sum_{n=n_1}^{n_2} \frac{(\mu_{lo}+b)^n}{n!} e^{-(\mu_{lo}+b)} \geq CL$$

For a given value of b construct belts for all possible values of mu




The famous FELDMAN & COUSINS technique (for a Physicist):

G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," Phys. Rev. D 57, 3873 (1998) http://prola.aps. org/pdf/PRD/v57/i7/p3873 1.

# The Unified Approach

$$\sum_{n=n_1}^{n_2} \frac{(\mu_{lo}+b)^n}{n!} e^{-(\mu_{lo}+b)} \geq CL$$

Acceptance interval constrained by equation but not fully specified by it.

Unified approach (a.k.a Feldman & Cousins)

ordering principle based on likelihood ratio

1) Compute the mean background expectation b (e.g, b=1)
2) Given b, compute the best estimator for the true value of the mean signal, mu_best for each possible measurement outcome, n. If mu is unconstrained, mu_best is found by maximizing the Poisson probability for any given b and n.

$$\left.\frac{dPo(n;\mu+b)}{d\mu}\right|_{n,b} = \frac{d}{d\mu}\frac{(\mu+b)^n}{n!}e^{-(\mu+b)}$$ =0; μ=n-b

$$\mu_{best} = \max(0, n-b)$$

Since only non-negative values for mu are allowed

# The Unified Approach

3) Given b, then compute, mu_best and Po(n; mu_best+b) for each possible n

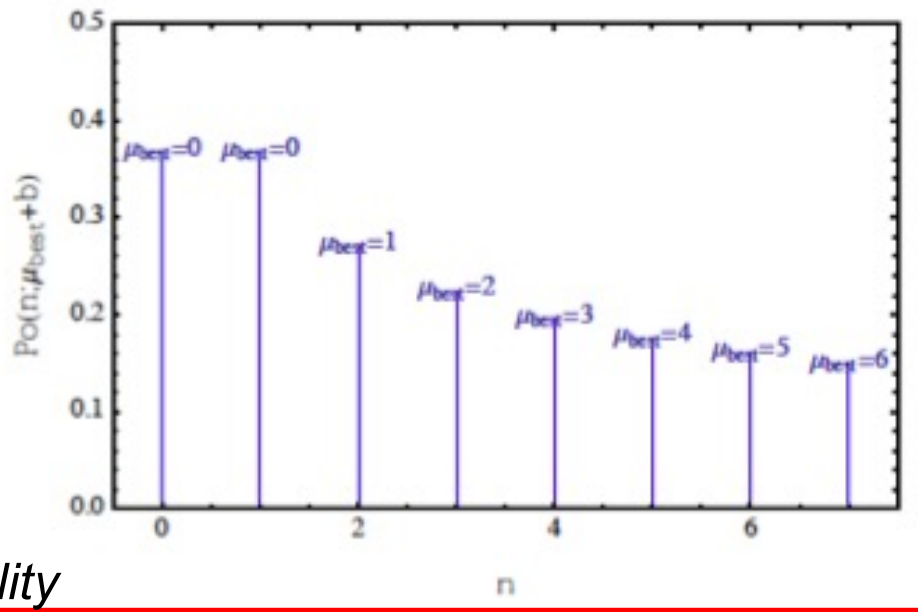
$$Po(n;\mu_{best}+b)=\begin{cases}\frac{b}{n!}e^{-b} & \text{if } n\le b\ (\Rightarrow \mu_{best}=0)\\ \frac{n^n}{n!}e^{-n} & \text{if } n>b\ (\Rightarrow \mu_{best}>0)\end{cases}$$

$$\mu_{best}=\max(0,n-b)$$

Example b=1

| $n$ | $\mu_{\text{best}}$ | $\text{Po}(n;\mu_{\text{best}}+b)$ |
|---|---|---|
| 0 | 0.0 | 0.368 |
| 1 | 0.0 | 0.368 |
| 2 | 1.0 | 0.271 |
| 3 | 2.0 | 0.224 |
| 4 | 3.0 | 0.195 |
| 5 | 4.0 | 0.175 |

*max probability*



# The Unified Approach

4) Scan all possible values of the unknown true mean signal mu
5) Given b and mu order all possible measurements outcomes n from most to least likely. This ordering is done according to the values of the Likelihood Ratio:

$$L_R = \frac{Po(n;\mu+b)}{Po(n;\mu_{best}+b)}$$

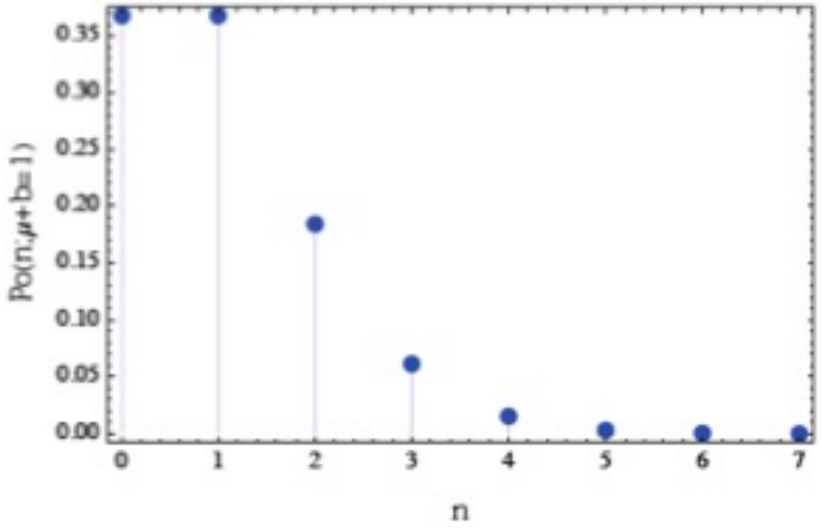
$$L_R = \begin{cases} \left(\frac{n+b}{b}\right)^n e^{-\mu} & \text{if } n \le b \ (\Rightarrow \mu_{best}=0) \\ \left(\frac{\mu+b}{n}\right)^n e^{-(\mu+b-n)} & \text{if } n > b \ (\Rightarrow \mu_{best}>0) \end{cases}$$

The highest rank (R=1) is assigned to the n value having the highest value of L_R

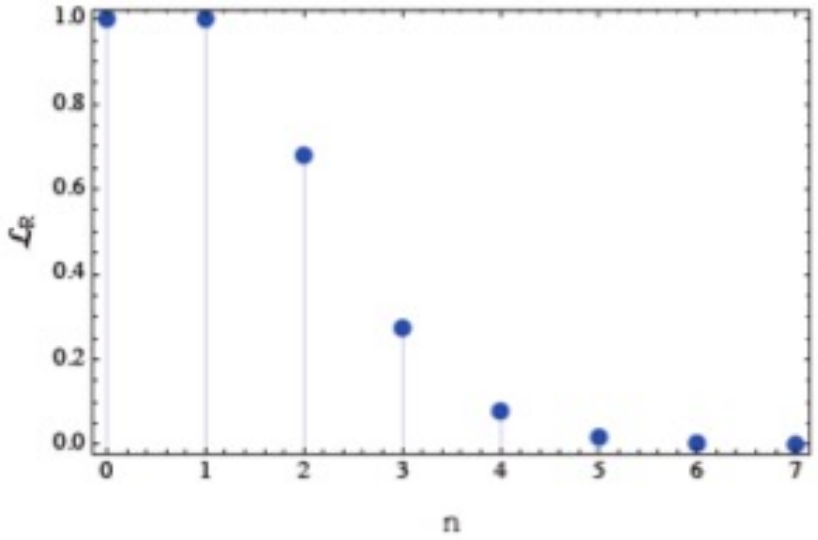
# The Unified Approach

b=1, mu =0

| $n$ | $\mathrm{Po}(n;\mu+b)$ | $\mathcal{L}_R$ | $\mathcal{R}$ | $\mathcal{P}$ |
|---|---|---|---|---|
| 0 | 0.368 | 1.000 | 1 | 0.368 |
| 1 | 0.368 | 1.000 | 2 | 0.736 |
| 2 | 0.184 | 0.680 | 3 | 0.920 |
| 3 | 0.061 | 0.274 | | |
| 4 | 0.015 | 0.078 | | |
| 5 | 0.003 | 0.017 | | |




| $n$ | $\mu_{\mathrm{best}}$ | $\mathrm{Po}(n;\mu_{\mathrm{best}}+b)$ |
|---|---|---|
| 0 | 0.0 | 0.368 |
| 1 | 0.0 | 0.368 |
| 2 | 1.0 | 0.271 |
| 3 | 2.0 | 0.224 |
| 4 | 3.0 | 0.195 |
| 5 | 4.0 | 0.175 |



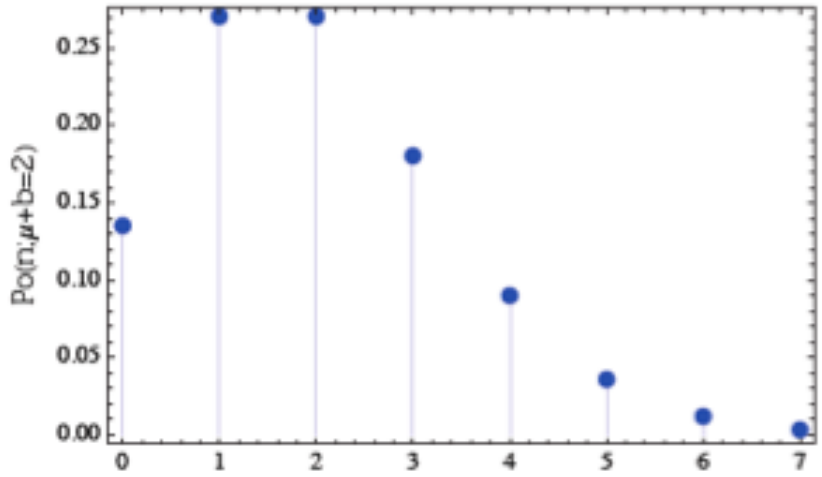

NOTE: $P_0$ is a DENSITY, $\mathcal{P}$ is a PROBABILITy

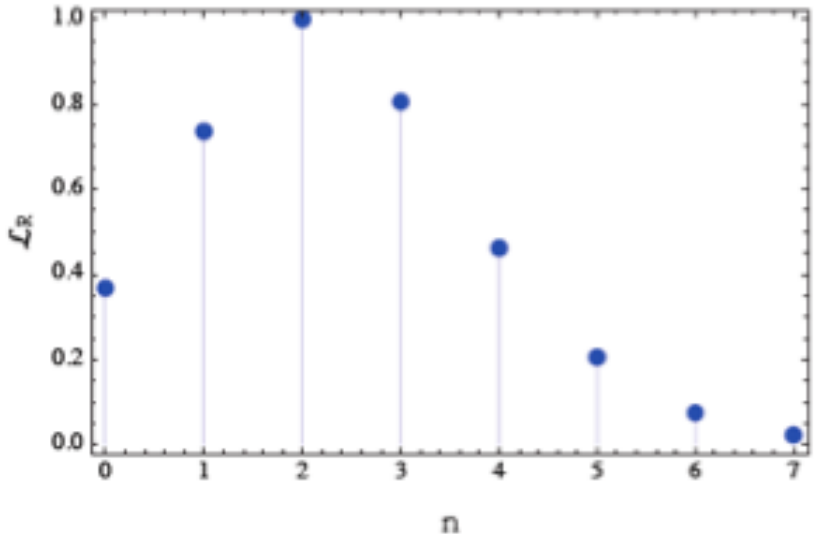
# The Unified Approach

b=1, mu=1

| $n$ | $\mathrm{Po}(n;\mu+b)$ | $\mathcal{L}_R$ | $\mathcal{R}$ | $\mathcal{P}$ |
|---|---|---|---|---|
| 0 | 0.135 | 0.368 | 5 | 0.947 |
| 1 | 0.271 | 0.736 | 3 | 0.722 |
| 2 | 0.271 | 1.000 | 1 | 0.271 |
| 3 | 0.180 | 0.805 | 2 | 0.451 |
| 4 | 0.090 | 0.462 | 4 | 0.812 |
| 5 | 0.036 | 0.206 | | |

| $n$ | $\mu_{\mathrm{best}}$ | $\mathrm{Po}(n;\mu_{\mathrm{best}}+b)$ |
|---|---|---|
| 0 | 0.0 | 0.368 |
| 1 | 0.0 | 0.368 |
| 2 | 1.0 | 0.271 |
| 3 | 2.0 | 0.224 |
| 4 | 3.0 | 0.195 |
| 5 | 4.0 | 0.175 |



# The Unified Approach

6) For a given b and mu construct a CI in n by adding the Po(n; mu+b) values until the cumulative probability is at least as large as the desired CL. The values are added according to their ranking, starting by R=1

b=1, mu =0

| $n$ | $\mathrm{Po}(n;\mu+b)$ | $\mathcal{L}_R$ | $\mathcal{R}$ | $\mathcal{P}$ |
|---|---|---|---|---|
| 0 | 0.368 | 1.000 | 1 | 0.368 |
| 1 | 0.368 | 1.000 | 2 | 0.736 |
| 2 | 0.184 | 0.680 | 3 | 0.920 |
| 3 | 0.061 | 0.274 | | |
| 4 | 0.015 | 0.078 | | |
| 5 | 0.003 | 0.017 | | |

[n1,n2]-->[0,2] at 90% CL

b=1, mu =1

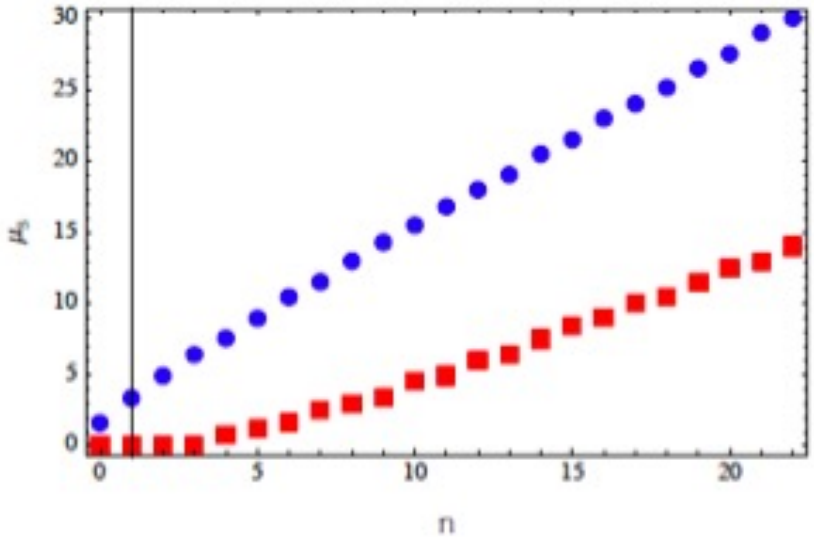
| $n$ | $\mathrm{Po}(n;\mu+b)$ | $\mathcal{L}_R$ | $\mathcal{R}$ | $\mathcal{P}$ |
|---|---|---|---|---|
| 0 | 0.135 | 0.368 | 5 | 0.947 |
| 1 | 0.271 | 0.736 | 3 | 0.722 |
| 2 | 0.271 | 1.000 | 1 | 0.271 |
| 3 | 0.180 | 0.805 | 2 | 0.451 |
| 4 | 0.090 | 0.462 | 4 | 0.812 |
| 5 | 0.036 | 0.206 | | |

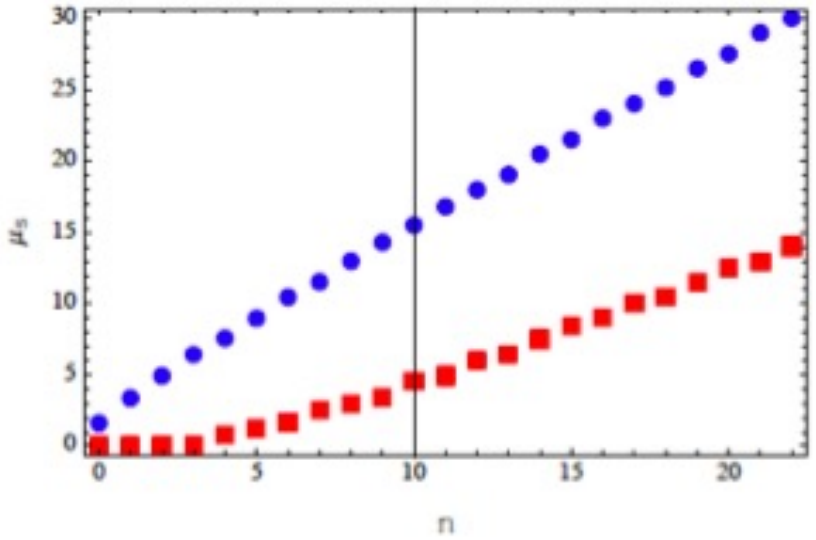
[n1,n2]-->[0,4] at 90% CL

# The Unified Approach

7) Repeat for different values of mu.
8) Now perform the measurement. To obtain [mu_low, mu_up] read vertically

b=1,Confidence belt
at 95% CL



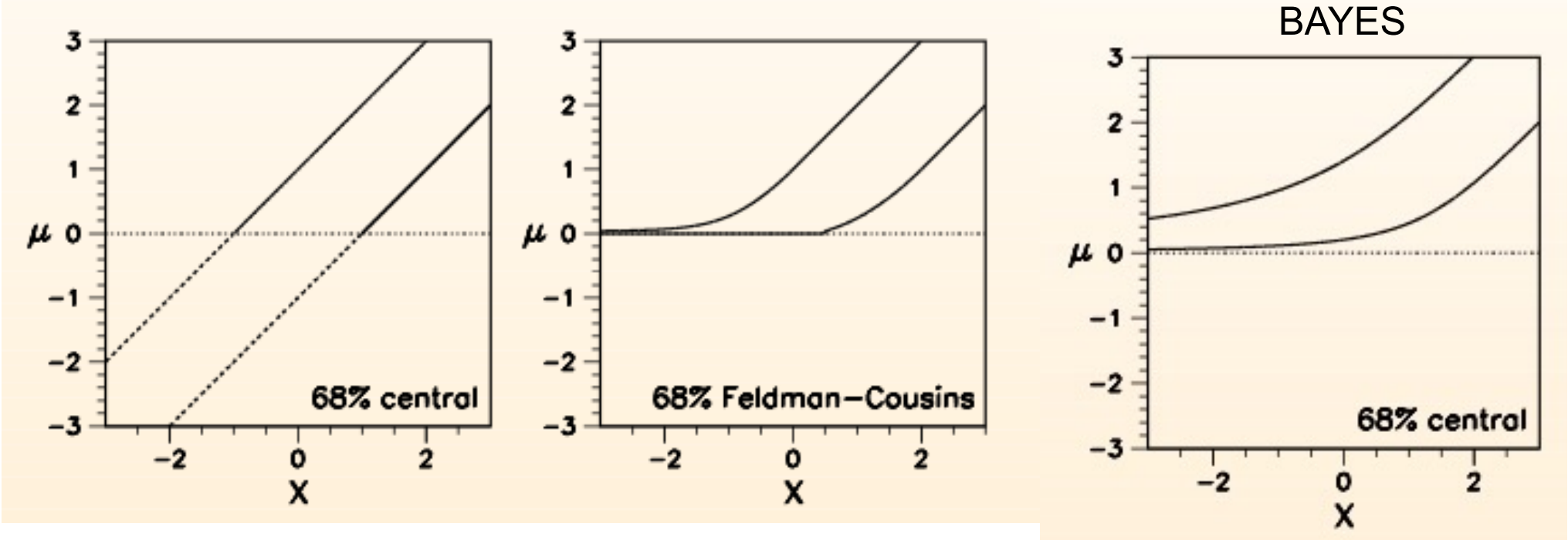

$$n_{obs} = 1 \rightarrow CI \Rightarrow [0,3.4]$$

$$n_{obs} = 10 \rightarrow CI \Rightarrow [4.5,15.5]$$

PROBLEMS with Feldman & Cousins

- For small n and n<b the UPPER LIMIT depends on b !
  (it happens that better limits are obtained for larger values of b)
- When dealing with more than 1 parameter
- Uncertainties on the estimation of background
- Coverage

1 observation for $\mu \geq 1$ gaussian with $\sigma=1$



Instead of computing β, classically, as:

$$\beta = P(\hat{\nu}_s \leq \hat{\nu}_s^{obs}; \nu_s^{up}) = \sum_{n \leq n_{obs}} \frac{(\nu_s^{up} + \nu_b)^n \, e^{-(\nu_s^{up}+\nu_b)}}{n!}$$

Variable is $\nu_s^{up}$

**We can use the Bayes Theorem** *(Cowan, section 9.9)*

*(better solution, but assuming a uniform Prior …)*

$$p(\nu_s|n_{obs}) = \frac{L(n_{obs}|\nu_s)\,\pi(\nu_s)}{\int_{-\infty}^{\infty} L(n_{obs}|\nu_s')\,\pi(\nu_s')\,d\nu_s'}$$

$\pi(\nu_s)$ = costant for $\nu_s \geq 0$
= 0 for $\nu_s < 0$

$$L(n_{obs}|\nu_s) = \frac{(\nu_s + \nu_b)^{n_{obs}}}{n_{obs}!} e^{-(\nu_s+\nu_b)} \quad \text{and} \quad 1 - \beta = \frac{\int_0^{\nu_s^{up}} L(n_{obs}|\nu_s)\,d\nu_s}{\int_0^{\infty} L(n_{obs}|\nu_s)\,d\nu_s}$$

Integration over $\nu_s^{up}$

Integrating for parts $n_{obs}$ times β is obtained as:

*(otherwise, one can integrate via Monte Carlo)*

$$\beta = \frac{e^{-(\nu_s^{up}+\nu_b)} \sum_{n=0}^{n_{obs}} \frac{(\nu_s^{up}+\nu_b)^n}{n!}}{e^{-\nu_b} \sum_{n=0}^{n_{obs}} \frac{\nu_b^n}{n!}}$$

## Carefulness

Sensitivity of an experiment:
either the averaged expected central value from an ensemble of experiment
or averaged expected upper limit from an ensemble of experiment with expected background and no true signal.

***What to do if the expected C.V. or U.L. are rather distant from the expectation ?***

Critical question when the observed background is less than what expected

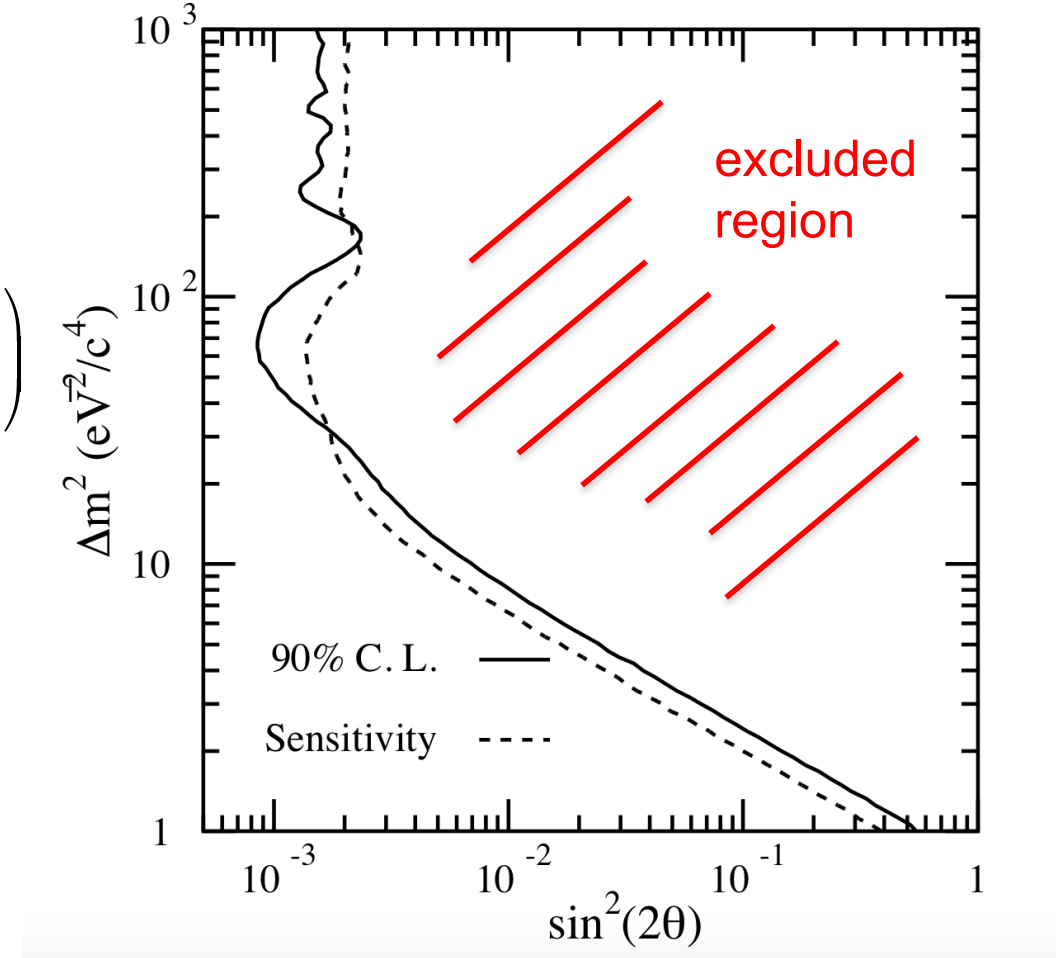
Answer: report both, e.g. U.L. from data and U.L. from sensititvity !

Simple toy model, mixing of two neutrino flavor states:

$$\left|\upsilon_\mu\right\rangle = \left|\upsilon_2\right\rangle \cos\theta + \left|\upsilon_4\right\rangle \sin\theta$$

$$\left|\upsilon_s\right\rangle = \left|\upsilon_4\right\rangle \cos\theta - \left|\upsilon_2\right\rangle \sin\theta$$

$$P\left(\upsilon_\mu \rightarrow \upsilon_s\right) = \sin^2(2\theta)\sin^2\left(\frac{1.27\Delta m^2 L(km)}{E(GeV)}\right)$$




*(from Feldman and R. D. Cousins, Phys. Rev. D 57, 3873)*

# Statistical vs. systematic errors

**Statistical errors:**

How much would the result fluctuate upon repetition of the measurement?

Implies some set of assumptions to define the probability of outcome of the measurement.

**Systematic errors:**

What is the uncertainty in my result due to uncertainty in my assumptions, e.g.,

model (theoretical) uncertainty;
modeling of measurement apparatus

The sources of error do not vary upon repetition of the measurement. Often result from uncertain value of, e.g., calibration constants, efficiencies, etc.

***(we are not talking here of systematical, fixed "errors" that shift the expectation)***

Include sytematic errors in the STATISTICAL INFERENCE

**Limits including systematics error**

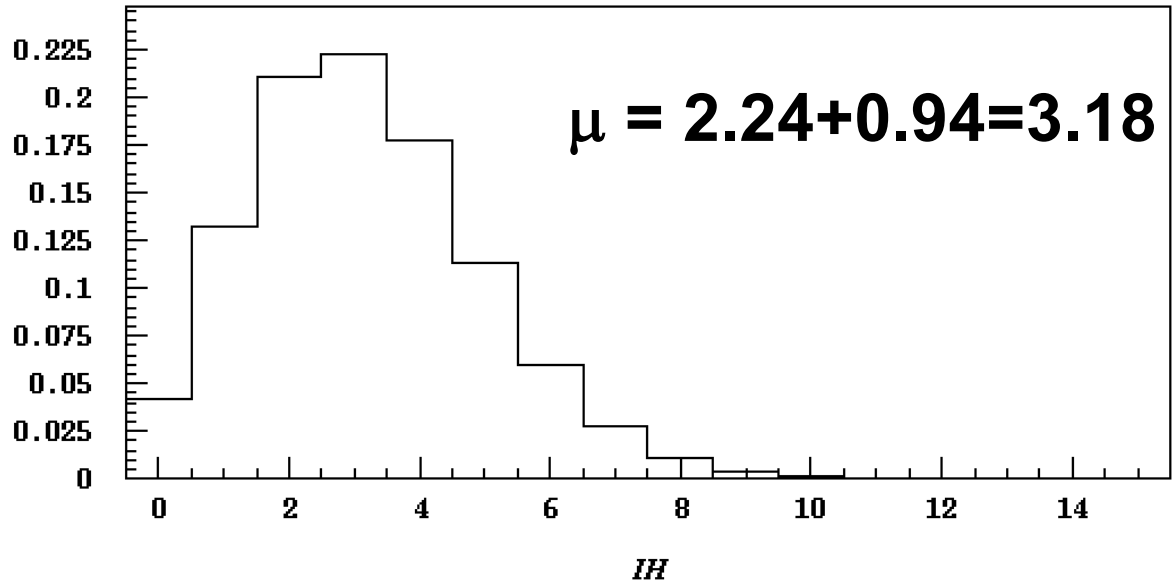
The usual way to incorporate the effects of systematic uncertainties (also called nuisance parameters) is to convolute the Poisson probability with a function representing the prior probability density in each parameter (usually one uses Gaussian but more complicated dependences may be treated via toy Monte Carlo's). Then one "marginalize" the nussance parameters by integrating over them.

Integration is equivalent to

- simulate an adequate large number of pseudo-experiments and
- extract from the obtained distributions the interesting statistical parameters (e.g. Higgs mass)

Neutrino mass ordering for IH case in NOvA experiment (2016)

Signal: 2.24±0.29
Background: 0.94±0.09
Observation: 6 events




$$p-value=\sum_{i=6}^{\infty} Poi\left(x_i \middle| \mu=3.18\right)=10.3\%$$

More sophisticated way: make convolution of Poisson with Gaussian's

$$PDF(x) = \int Poi\left(x \middle| \mu_S + \mu_B\right) \otimes Gau\left(\mu_S \middle| \hat{\mu}_S, \hat{\sigma}_S\right) \otimes Gau\left(\mu_B \middle| \hat{\mu}_B, \hat{\sigma}_B\right) d\mu_S d\mu_B$$

Practically, the convolution can be made in the following way:

$$PDF(x) = \frac{\sum_{i=1}^{i=13}\sum_{j=1}^{j=13} \frac{1}{\sigma_B\sqrt{2\pi}} e^{-\frac{\left(\hat{\mu}_B-\hat{\mu}_B+a_i\sigma_B\right)^2}{2\sigma_B^2}} \frac{1}{\sigma_S\sqrt{2\pi}} e^{-\frac{\left(\hat{\mu}_S-\hat{\mu}_S+a_j\sigma_S\right)^2}{2\sigma_S^2}} \bullet \frac{e^{-\left(\hat{\mu}_S+a_j\sigma_S+\hat{\mu}_B+a_i\sigma_B\right)}\left(\hat{\mu}_S+a_j\sigma_S+\hat{\mu}_B+a_i\sigma_B\right)^x}{x!}}{\sum_{x=0}^{x=\infty}\sum_{i=1}^{i=13}\sum_{j=1}^{j=13} \frac{1}{\sigma_B\sqrt{2\pi}} e^{-\frac{\left(\hat{\mu}_B-\hat{\mu}_B+a_i\sigma_B\right)^2}{2\sigma_B^2}} \frac{1}{\sigma_S\sqrt{2\pi}} e^{-\frac{\left(\hat{\mu}_S-\hat{\mu}_S+a_j\sigma_S\right)^2}{2\sigma_S^2}} \bullet \frac{e^{-\left(\hat{\mu}_S+a_j\sigma_S+\hat{\mu}_B+a_i\sigma_B\right)}\left(\hat{\mu}_S+a_j\sigma_B+\hat{\mu}_B+a_i\sigma_B\right)^x}{x!}}$$

with

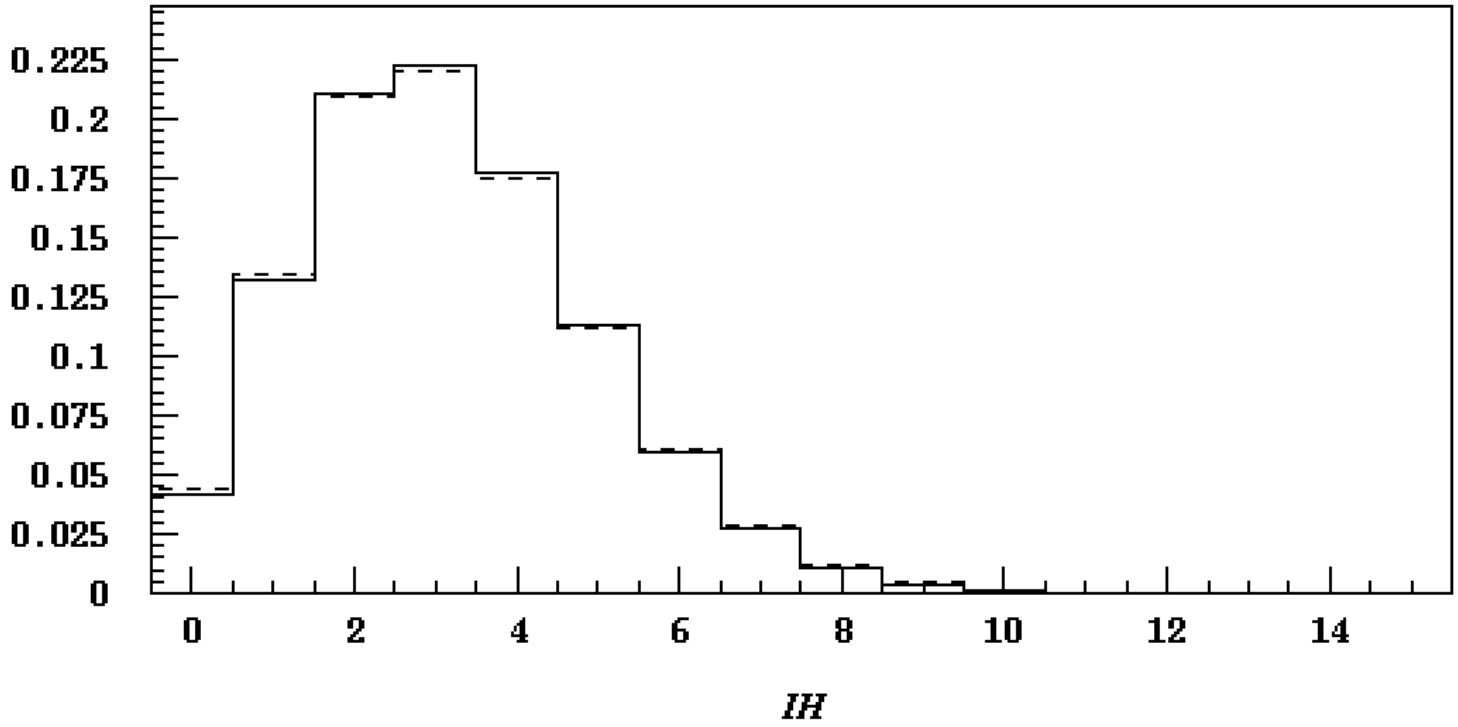
| | |
|---|---|
| $a_1$ | -3 |
| $a_2$ | -2.5 |
| $a_3$ | -2 |
| … | |
| $a_7$ | 0 |
| … | |
| $a_{13}$ | 3 |

$$\mu_B = \hat{\mu}_B + a_i\sigma_B$$

$$\mu_S = \hat{\mu}_S + a_j\sigma_S$$

$$PDF(x) = \frac{\sum_{i=1}^{i=13}\sum_{j=1}^{j=13} \frac{1}{\sigma_B\sqrt{2\pi}} e^{-\frac{a_i^2}{2}} \frac{1}{\sigma_S\sqrt{2\pi}} e^{-\frac{a_j^2}{2}} \bullet \frac{e^{-(\mu_B+\mu_S)}\left(\mu_B+\mu_S\right)^x}{x!}}{\sum_{x=0}^{x=\infty}\sum_{i=1}^{i=13}\sum_{j=1}^{j=13} \frac{1}{\sigma_B\sqrt{2\pi}} e^{-\frac{a_i^2}{2}} \frac{1}{\sigma_S\sqrt{2\pi}} e^{-\frac{a_j^2}{2}} \bullet \frac{e^{-(\mu_B+\mu_S)}\left(\mu_B+\mu_S\right)^x}{x!}}$$

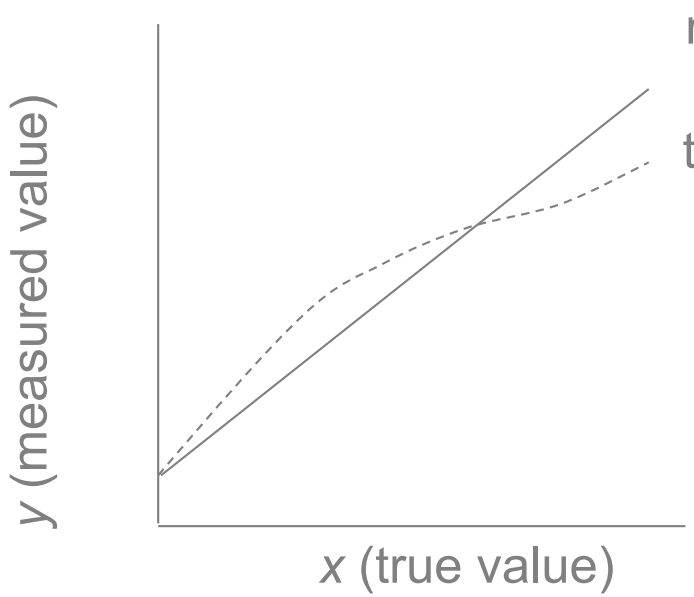
## Convoluted Poisson for IH




$$p-value=\sum_{i=6}^{\infty}PDF\left(x_i\middle|\tilde{\hat{\mu}}_S+\tilde{\hat{\mu}}_B\right)=10.6\%$$

# Systematic errors and nuisance parameters

Response of measurement apparatus is never modelled perfectly:




model: $y = \alpha + \beta x$

truth: $y = \alpha + \beta x + \gamma x^2 + \varepsilon x^3 + \ldots$

Model can be made to approximate better the truth by including more free parameters.

systematic uncertainty ↔ nuisance parameters

# Nuisance parameters

Suppose the outcome of the experiment is some set of data values $x$ (here shorthand for e.g. $x_1$, ..., $x_n$).

We want to determine a parameter $\theta$,
(could be a vector of parameters $\theta_1$, ..., $\theta_n$).

The probability law for the data $x$ depends on $\theta$:

$L(x|\,\theta)$ (the likelihood function)

E.g. maximize $L$ to find estimator $\hat{\theta}$ .

Now suppose, however, that the vector of parameters:
contains some that are of interest, $\psi_1, \ldots, \psi_n$
and others that are not of interest: $\lambda_1, \ldots, \lambda_m$ .
Symbolically: $\theta = (\psi, \lambda)$

The $\lambda_1, \ldots, \lambda_m$ are called nuisance parameters.

# Digression: marginalization with MCMC

Bayesian computations involve integrals like

$$p(\theta_0|x) = \int p(\theta_0, \theta_1|x)\, d\theta_1 \ .$$

often high dimensionality and impossible in closed form,
also impossible with 'normal' acceptance-rejection Monte Carlo.

Markov Chain Monte Carlo (MCMC) has revolutionized Bayesian computation.
Google for 'MCMC', 'Metropolis', 'Bayesian computation', ...

MCMC generates correlated sequence of random numbers:

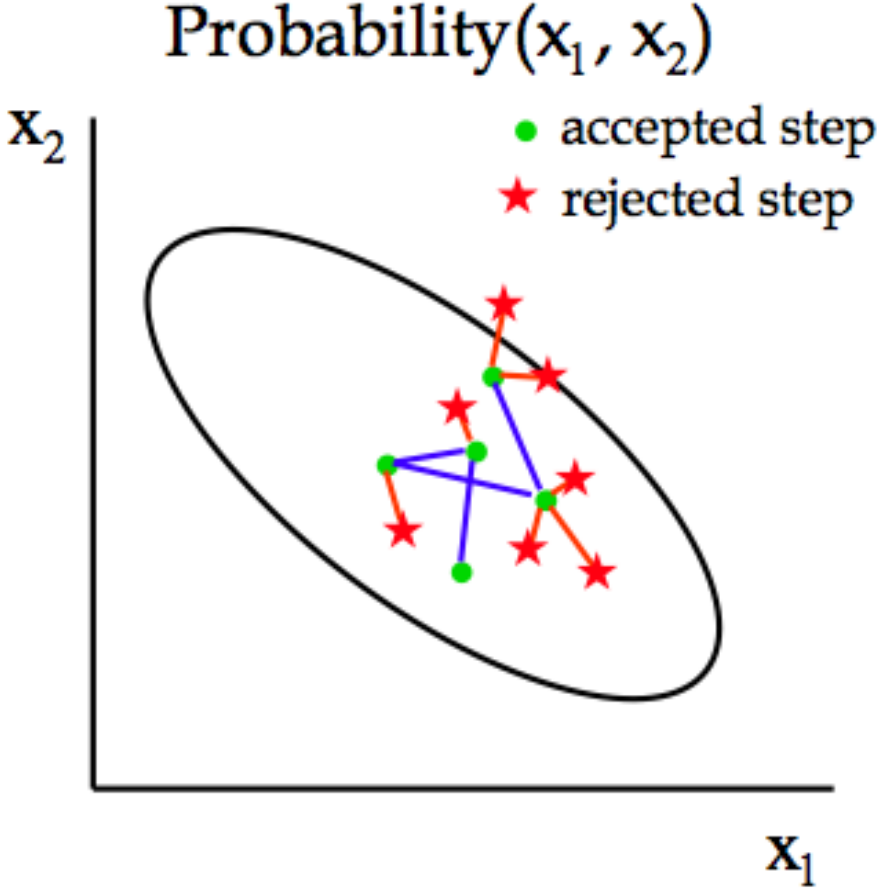
cannot use for many applications, e.g., detector MC;
effective stat. error greater than $\sqrt{n}$ .

Basic idea: sample multidimensional $\vec{\theta}$ ,
look, e.g., only at distribution of parameters of interest.

# MARKOV CHAIN MONTE CARLO: MCMC

Generates sequence of random samples from an arbitrary probability density function

- Metropolis algorithm:
  - draw trial step from symmetric pdf, i.e., $t(\Delta \mathbf{x}) = t(-\Delta \mathbf{x})$
  - accept or reject trial step
  - simple and generally applicable
  - relies only on calculation of target pdf for any $\mathbf{x}$



**A famous Markov chain is the so-called "drunkard's walk", a random walk on a number line, where, at each step, the position may change by +1 or −1 with equal probability. From any position there are two possible transitions, to the next or previous integer. The transition probabilities depend only on the current position, not on the manner in which the position was reached. For example, the transition probabilities from 5 to 4 and 5 to 6 are both 0.5, and all other transition probabilities from 5 are 0.**
**These probabilities are independent of whether the system was previously in 4 or 6.**

In alternative to the full integral approach one may use the NEXT STEP after the "Unified Approach": **profile likelihood function**

*(from 2013 G. Cowan et al, arXiv:1007.1727v3)*

Example for testing New Physics in Particle Physics

Let consider ***x*** random variables, and two set of parameters, $\mu$ and $\theta$.
Only $\mu$ are interesting parameters, while $\theta$ are e.g. nuisances.
If it not possible to marginalize on $\theta$, then the MAX LIKELIHHOD is given by:

$$(\hat{\mu},\hat{\theta}) = \text{argmax}_{\mu,\theta}\, L(\mu,\theta) \qquad \text{where} \qquad L(\mu,\theta) = \sum \log f(x;\mu,\theta)$$

However, it may be quite difficult to perform the full integration. Then, let assume that $\mu$ is known (and independent of $\theta$) and write: $L(\mu,\theta) = L_\mu(\theta)$

For each $\mu$ we therefore evaluate:

$$\hat{\theta}_\mu = \text{argmax}_\theta\, L_\mu(\theta) \quad \text{and} \quad \hat{\mu} = \text{argmax}_\mu\, L_\mu(\hat{\theta}) = \text{argmax}_\mu\, L(\mu,\hat{\theta})$$

Tautology, but we have ***profiled*** out the $\theta$ parameters $L(\mu,\hat{\theta}) = \sum \log f(x;\mu,\hat{\theta})$

Further proceed with computation of C.I. for $\mu$

One possibility for searching of new signals:
$H_0$ = known process (background)
$H_1$= background + sought signal

*p-value: error type I*

One possibility for setting limit on new signals:
$H_0$ = signal + background
$H_1$= background

*power: 1 - error type II*

In either case compute the *p*-value (of $H_0$), i.e. probability of finding data of equal or greater incompatibility with the prediction of $H_0$.
Hypothesis $H_0$ is excluded if *p*-value < predefined-threshold (C.L.)
Convert *p*-value into a significance $Z=\Phi^{1}(1-p)$, where $\Phi$ is the Gaussian Cum.Function.

Quantify the (median) sensitivity on $H_1$ expected from an experiment by computing Z (formally this is NOT the probability for $H_1$ !)

Work out an histogram $\vec{n} = (n_1, \ldots, n_N)$ with expectation $E(n_i) = \mu \cdot s_i + b_i$ and

$$s_i = s_{TOT} \int_{bin\,i} f_s\left(x; \vec{\theta}_s\right) dx, \quad b_i = b_{TOT} \int_{bin\,i} f_b\left(x; \vec{\theta}_b\right) dx$$

$\mu$ be the strength of the signal process *(sorry for the confusion about $\mu$ !)*
$\mu$=0 corresponds to background only, while $\mu$=1 is the nominal signal hypothesis,
i.e. $s_{TOT}$ is given by the nominal signal model,
$b_{TOT}$ is computed aside from data without signal, $\vec{m} = (m_1, \ldots, m_M)$, $E(m_i) = u_i\left(\vec{\theta}\right)$

The likelihood function is $L\left(\mu, \vec{\theta}\right) = \prod_{j=1}^{N} \frac{\left(\mu \cdot s_j + b_j\right)^{n_j}}{n_j!} e^{-\left(\mu \cdot s_j + b_j\right)} \prod_{k=1}^{M} \frac{u_k^{m_j}}{m_j!} e^{-u_k}$

The test function is the profile likelihood ratio: $\lambda(\mu) = \frac{L\left(\mu, \hat{\hat{\vec{\theta}}}\right)}{L\left(\hat{\mu}, \hat{\hat{\vec{\theta}}}\right)}$, $\hat{\hat{\vec{\theta}}}$ maximizes L for a specific $\mu$

And for the test-statistic takes $t = -2 \ln \lambda(\mu)$ since 0≤λ≤1

Then $p_\mu = \int_{t_{\mu,obs}}^{\infty} f\left(t_\mu \middle| \mu\right) dt_\mu$ for a specific $\mu$

For the case of a single parameter, Wilks(1938) and Wald (1943) showed that:

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathrm{O}\left(1/\sqrt{N}\right)$$

where N is the size of the data sample,
$\hat{\mu}$ follows a Gaussian with $\sigma$ standard dev.

*t* follows a $\chi^2$ distribution as far as $\mu$ is approaching the true value

**DISCOVERY**

Test $\mu$=0 against the observed $\hat{\mu} > 0$ i.e. compute the statistic

$$q_0 = \begin{cases} -2\ln\lambda(0) & for\ \hat{\mu} \geq 0 \\ 0 & for\ \hat{\mu} < 0 \end{cases}$$

with $f\left(q_0 \middle| \mu = 0\right) = p.d.f.\left(\chi^2_1\right)$
for Wilks-Wald

From Wald approximation the cumulative distribution of $q_0$ is $F\left(q_0 \middle| \mu\right) = \Phi\left(\sqrt{q_0}\right)$

And therefore $Z_0 = \Phi^{-1}\left(1-p_0\right) = \sqrt{q_0}$

Some time, we see also $Z_0 = \Phi^{-1}\left(1-p_0\right) = \sqrt{\chi^2}$

## SET UPPER LIMIT

Test U.L. on $\mu$ against the observed $\hat{\mu}$ i.e. compute the statistic

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & for\ \hat{\mu} \le \mu \\ 0 & for\ \hat{\mu} > \mu \end{cases}$$

with $f\left(q_\mu \middle| \mu > 0\right) \approx p.d.f.\left(\chi^2_1\right)$ for Wilks-Wald (non-central $\chi^2$)

From Wald approximation the cumulative distribution of $q_0$ is $F\left(q_\mu \middle| \mu\right) = \Phi\left(\sqrt{q_\mu}\right)$

And therefore $Z_\mu = \Phi^{-1}\left(1 - p_\mu\right) = \sqrt{q_\mu}$

Sometime we see also $Z_\mu = \Phi^{-1}\left(1 - p_\mu\right) = \sqrt{\chi^2}$

**Note: try to use Monte Carlo simulation as much as possible, and surely for looking at the $q_0$ or $q_\mu$ distributions !**