STATISTICAL ANALYSIS IN HIGH-ENERGY PHYSICS

Elizabeth Worcester

Belle II Summer School

July 30, 2019

Introduction

- Like most physicists, I'm not a statistician!
  - This is not (primarily) a math talk
  - My background is kaons and neutrinos
- Useful expertise exists in the world of mathematics, statistics, computer science…often no need to reinvent the wheel
  - Most important thing as a physicist is to understand and communicate the question we are trying to answer and the assumptions we are making
- The goal of this talk is to introduce basics of statistical analysis and survey some of the ways we use statistical analysis in physics
  - Many details and caveats not included!
  - Will try to provide both proper statistics terms and physics examples/jargon
  - Google is your friend!

Frequentist or Bayes?

- Two “schools of thought” in statistical analysis
- Difference is philosophical – different questions
- Frequentist:
  - Frequency of outcome of repeatable experiment (eg: coin flips)
  - Probability for hypothesis is not defined
  - “Confidence intervals”, “p-values”
- Bayesian:
  - Probability for hypothesis is defined
  - Allows incorporation of additional information (prior) which may, in principle, be subjective (“belief”)
  - “Prior” and “posterior” distributions
- In practice, it’s a bit muddier
  - Frequentist analysis can include previous measurements (prior) via nuisance parameters
  - Physicists doing Bayesian analysis try to avoid subjective priors

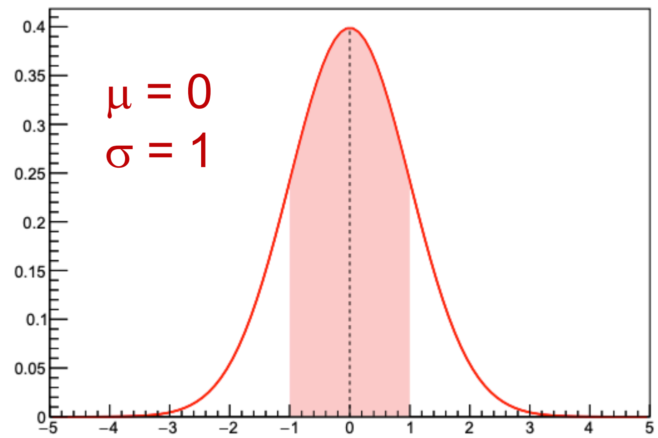
Gaussian Probability Distribution

$$g(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

μ = mean
σ = standard deviation

- Commonly used in physics because many random variables in real experiments can be approximated by a Gaussian distribution

"Normal distribution"



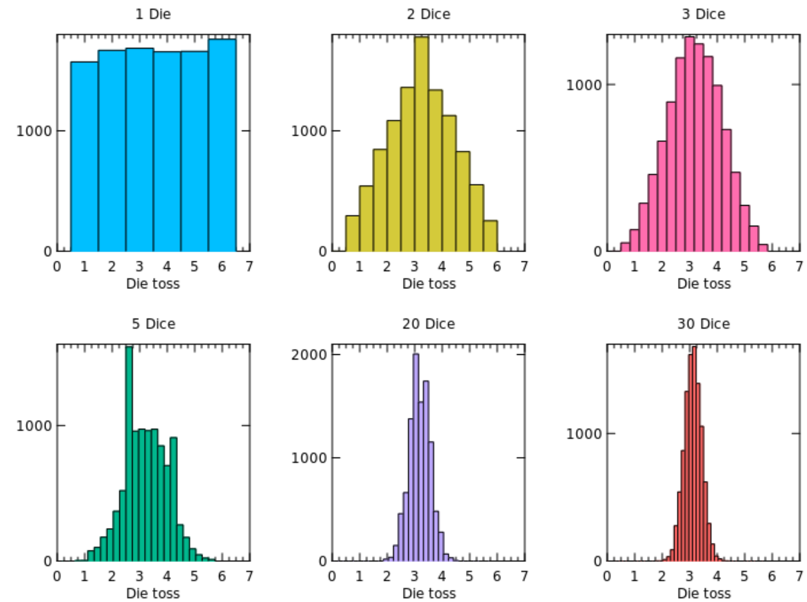
Gaussian Probability Distribution

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

μ = mean
σ = standard deviation

- Commonly used in physics because many random variables in real experiments can be approximated by a Gaussian distribution
- Central Limit Theorem → when independent random variables are added, the sum will tend towards a normal distribution, even if the original variables are not normally distributed

n=1, not normal at all



n>>1, pretty normal

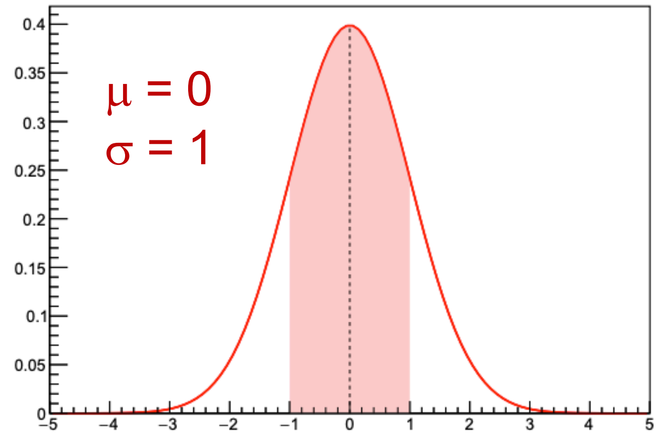
Gaussian Probability Distribution

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

μ = mean
σ = standard deviation

- Commonly used in physics because many random variables in real experiments can be approximated by a Gaussian distribution
- Central Limit Theorem → when independent random variables are added, the sum will tend towards a normal distribution, even if the original variables are not normally distributed
- Probability corresponding to values in the range [μ-nσ, μ+nσ] often used to quote significance

"Normal distribution"

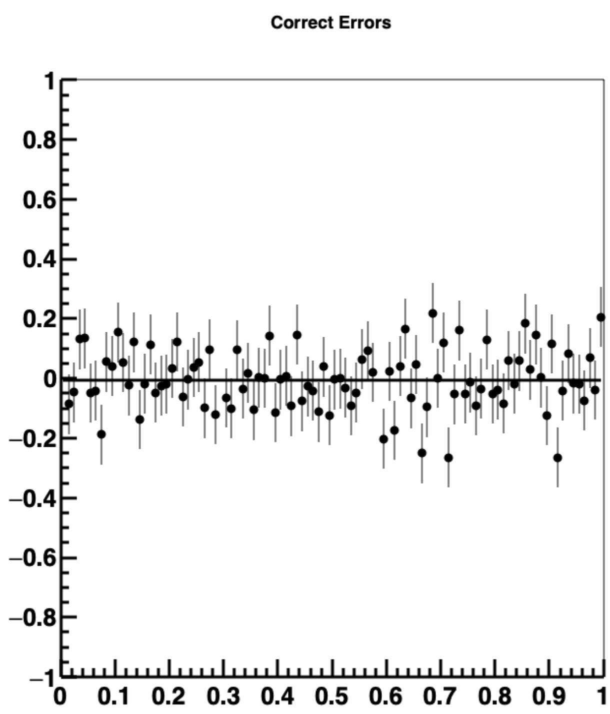


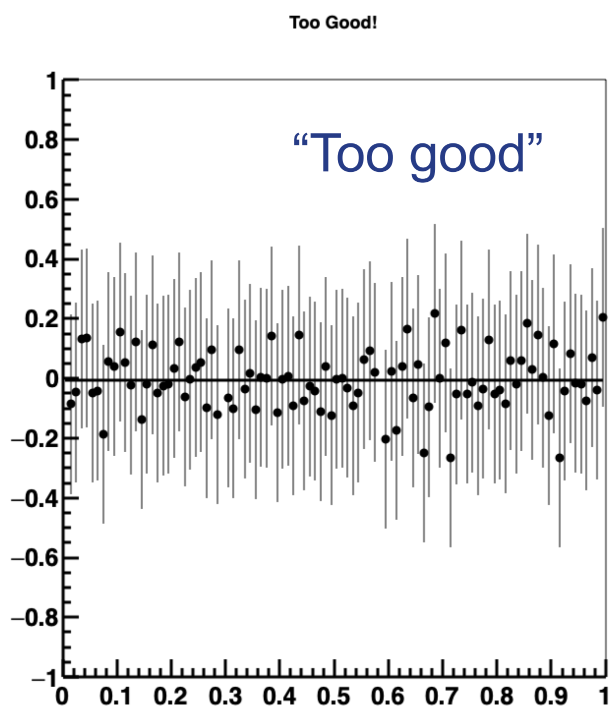
n	Prob.
1	0.683
2	0.954
3	0.997
4	$1-6.5\times10^{-5}$
5	$1-5.7\times10^{-7}$

$\chi^2(1)$

Too Good?

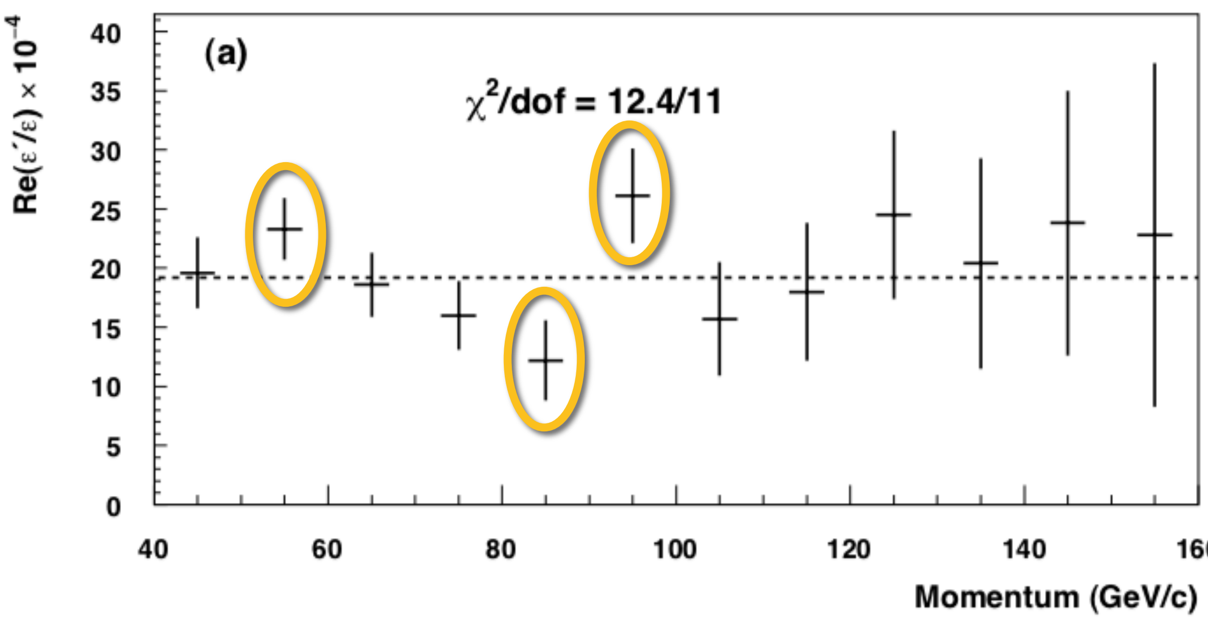
- If uncertainties are correct and uncorrelated we expect to see ~68% of measurements within 1σ of the mean → we expect to see ~32% of measurements >1σ from the mean





Too Good?

- If uncertainties are correct and uncorrelated we expect to see ~68% of measurements within 1σ of the mean → we expect to see ~32% of measurements >1σ from the mean

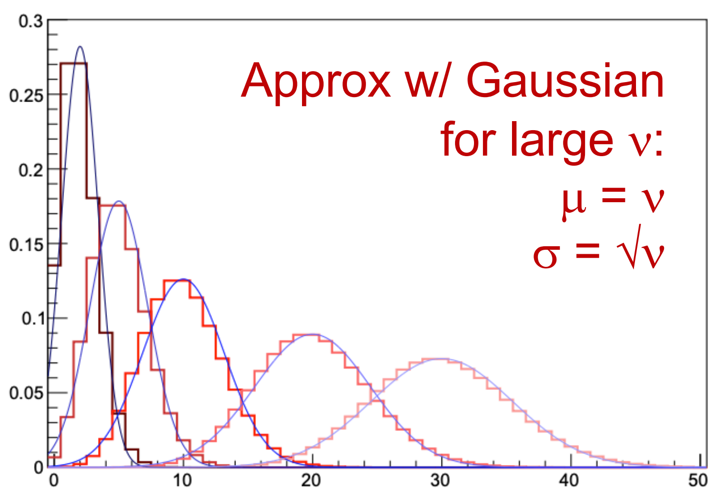


"Seems reasonable"

Other Commonly Used Probability Distributions

Poisson: $$P(n;\nu) = \frac{\nu^n}{n!}e^{-\nu}$$

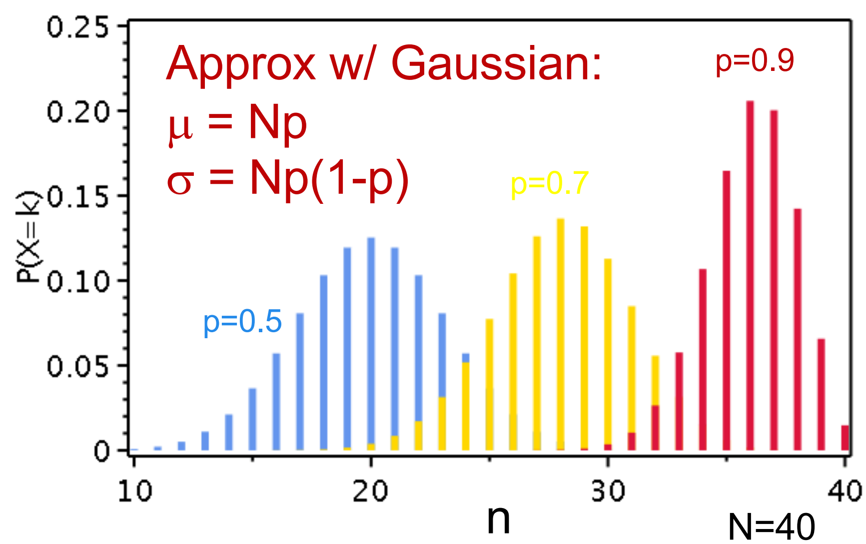
- ν is the average number of independent events in a time interval
- P(n) is the distribution of the observed integral number of events in a time interval
- eg: counting radioactive decays or cosmic ray flux through detector



Binomial:

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

- P(n) is the probability of n “successes” given N trials, where the probability of success is p
- eg: n is the number of detected particles for a detector with efficiency p

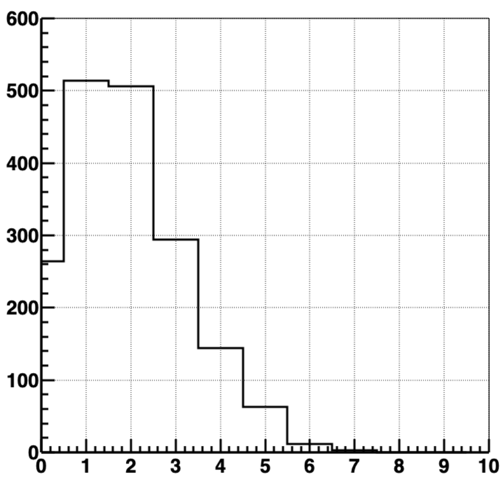


Simple Example: Cosmic Ray Detector

- 10 cm x 10 cm scintillator detector
- Simple DAQ reads out once per second and reports number of hits in that second
- Google tells you the average muon flux at sea level is 1 muon/cm^2/minute → expect 1.7 muons/s in your detector at sea level

Simulation of data collected over 30 minute period:

- n = "true" muon rate = flux through 100 cm^2 at your location = 1.9 Hz (just a choice)
- Number of trials = 1800

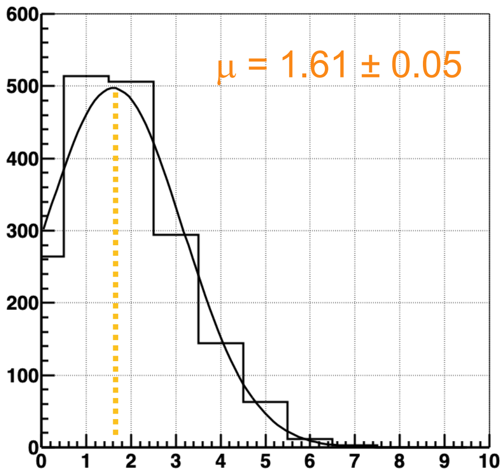



```
nu = 1.9 #Hz = number of cosmics in detector per second
array_poisson = np.random.poisson(nu,1800)
hp = ROOT.TH1I("hp","hp",10,-0.5,9.5)

for n in array_poisson:
    hp.Fill(n)
```

Simple Example: Cosmic Ray Detector

- 10 cm x 10 cm scintillator detector
- Simple readout system reads out once per second and tells you number of hits in that second
- Google tells you the average muon flux at sea level is 1 muon/cm^2/minute → expect 1.7 muons/s in your detector at sea level



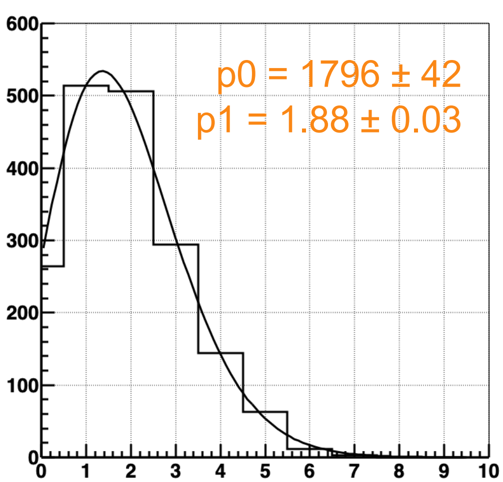
Fit to a Gaussian distribution:

```
hg.Fit("gaus")
```

Fit doesn't look terrible by eye, but we didn't recover our "true" rate of 1.9 Hz

Simple Example: Cosmic Ray Detector

- 10 cm x 10 cm scintillator detector
- Simple readout system reads out once per second and tells you number of hits in that second
- Google tells you the average muon flux at sea level is 1 muon/cm^2/minute → expect 1.7 muons/s in your detector at sea level



Fit to a Poisson distribution:

```
fp = ROOT.TF1("poi","[0]* TMath::Poisson(x,[1])", 0, 10)
fp.SetParameter(0,1800)
fp.SetParameter(1,1.7)
fp.SetParName(0,"p0")
fp.SetParName(1,"p1")
hp.Fit(fp)
```

Set fit parameters to reasonable starting values:
- We know number of trials
- Guess flux is sea level flux

Now we get back 1.9 Hz!

Measurement ("Inference")

- I took some data, now I want to determine the underlying parameters with an associated uncertainty
  - eg: what is the true flux of cosmic muons through my detector?
- We have to be careful about what we are asking:
  - Bayesian: "Given the data I have taken, what is the probability that the true rate is ν?" = $P(\nu|n)$

Likelihood Prior

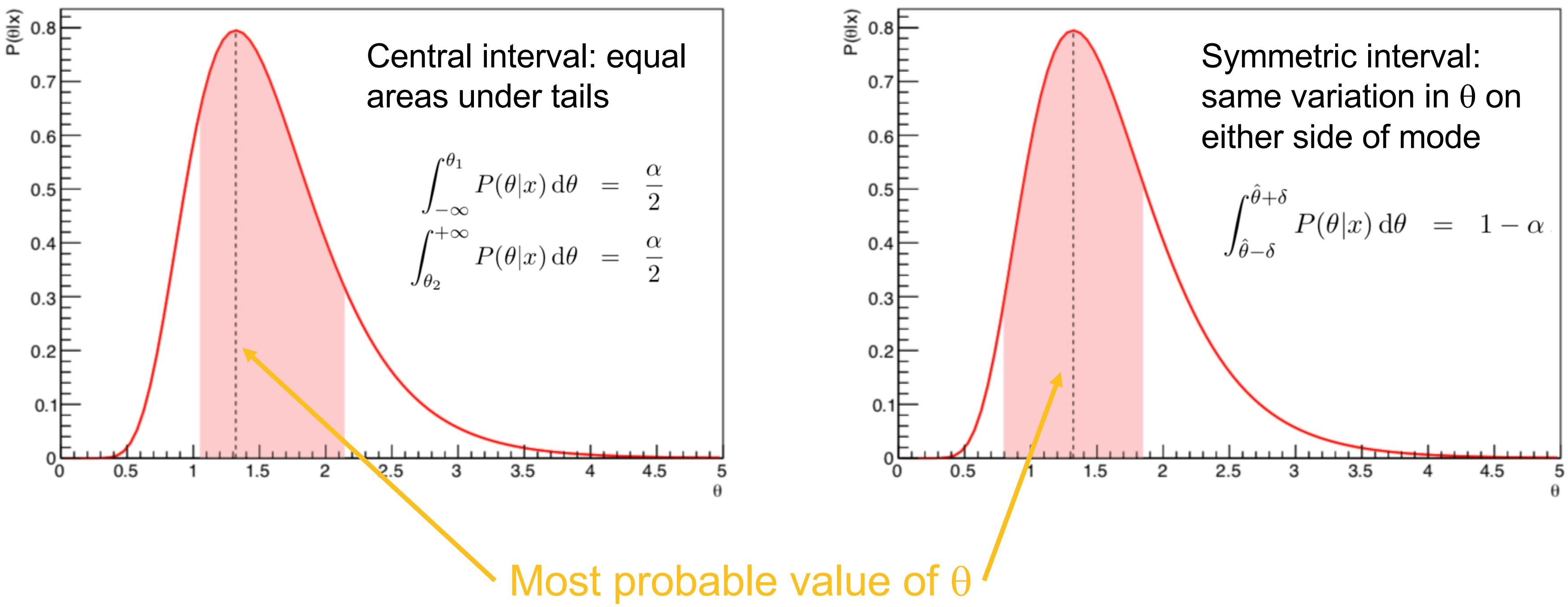
Posterior

$$P(\theta|x) = \frac{L(x;\theta)\pi(\theta)}{\int L(x;\theta)\pi(\theta)\,\mathrm{d}\theta}$$

Posterior: all the information we have about θ given our experimental data
Likelihood: all the information about the experiment (probability function evaluated at the observed data)
Prior: all the pre-existing information about θ (choice very important!)

Bayesian Inference

Two different definitions of 68% probability interval



"Given my data and my prior, there is a 68% probability that the true value of θ is in the range $[\theta_1:\theta_2]$."

Measurement ("Inference")

- I took some data, now I want to determine the underlying parameters with an associated uncertainty
  - eg: what is the true flux of cosmic muons through my detector?
- We have to be careful about what we are asking:
  - Frequentist: "What is the best-fit value of ν? Given the data I have taken, what is a range of possible values of ν, such that if I repeat the procedure many times, 90% of intervals will contain the true value?"

Frequentist Confidence Interval

https://seeing-theory.brown.edu/frequentist-inference/

Measurement ("Inference")

- I took some data, now I want to determine the underlying parameters with an associated uncertainty
  - eg: what is the true flux of cosmic muons through my detector?
- We have to be careful about what we are asking:
  - Frequentist: "What is the best-fit value of ν? Given the data I have taken, what is a range of possible values of ν, such that if I repeat the procedure many times, 90% of intervals will contain the true value?"

Most common "estimator" is the maximum likelihood estimator, in which the best fit value is one that maximizes the maximum likelihood function:

x_i are random variables (data)

$$L = f(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m) \Rightarrow L = \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

N independent measurements

θ_i are parameters describing experiment

Maximum Likelihood Estimates

- Maximum likelihood estimate is the value of the parameter θ that maximizes $L(\theta)$, which is equivalent to minimizing -ln(L)
  - Significance (χ^2) corresponds to -2ln(L)
  - Usually minimization must be done numerically

Gaussian distribution:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Log likelihood:
(likelihood is probability distribution evaluated at the data points)

$$\ln L(\mu,\sigma^2) = \sum_{i=1}^{n} \ln f(x_i;\mu,\sigma^2) = \sum_{i=1}^{n} \left(\ln \frac{1}{\sqrt{2\pi}} - \ln \sigma - \frac{(x_i-\mu)^2}{2\sigma^2} \right)$$

Maximum Likelihood Estimates

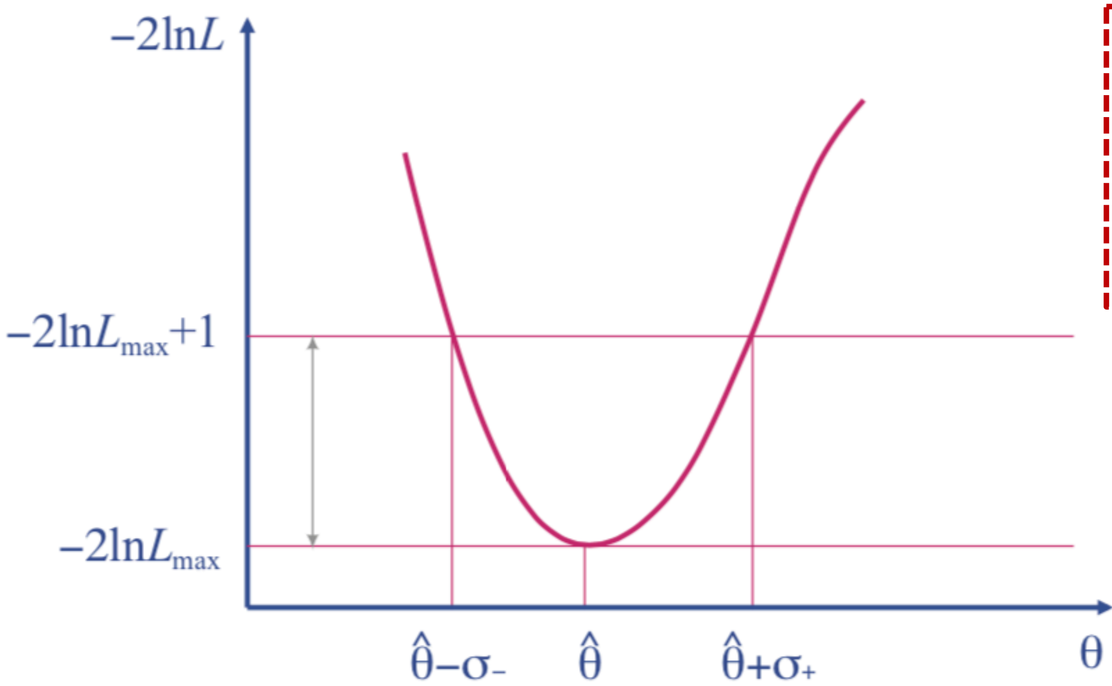
- Maximum likelihood estimate is the value of the parameter θ that maximizes L(θ), which is equivalent to minimizing -2ln(L)
  - Significance (χ^2) corresponds to -2ln(L)
  - Usually minimization must be done numerically

Minimize:

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Uncertainty for Maximum Likelihood Estimates



Alternatively, covariance matrix

$$V_{ij}^{-1} = - \frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\bigg|_{\theta_k=\hat{\theta}_k,\, \forall k}$$

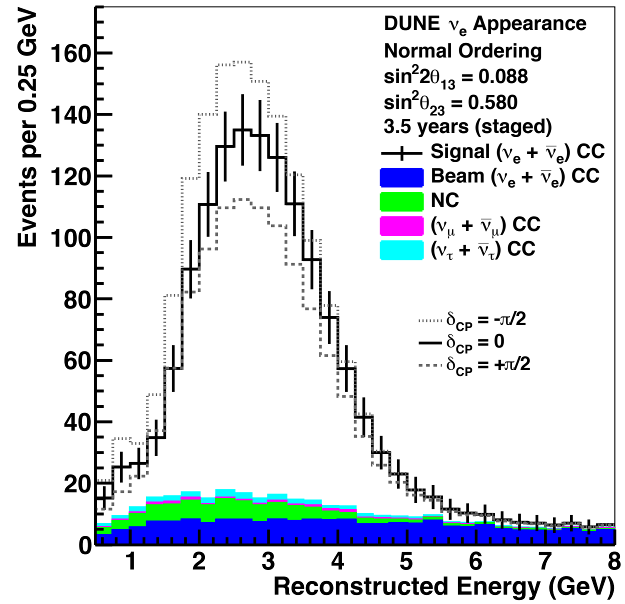
1σ uncertainty is range around the minimum of -2ln(L) for which -2ln(L) increases by 1
(can be asymmetric)

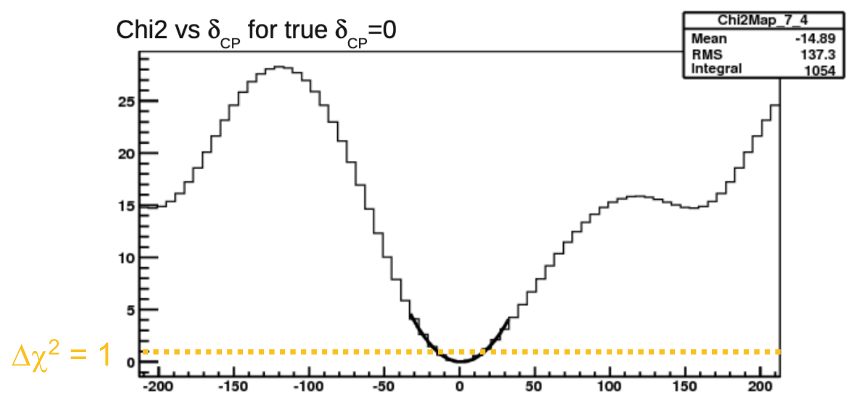
MLE Physics Example

- Example: DUNE prediction of resolution on δ_{CP}
- Use log likelihood function for binned, Poisson-distributed data:

$$\chi^2 = -2\log\mathcal{L} = 2\sum_{i}^{N_{\text{bins}}}\left[M_i - D_i + D_i \ln\left(\frac{D_i}{M_i}\right)\right]$$

  - M_i is expected number of events in bin (from MC)
  - D_i is observed number of events in bin
- Generate spectra at δ_{CP}=0 (expected value) and for a range of δ_{CP} values (observed value)
- Since -2ln(L) corresponds to χ^2 define 1σ resolution as the range of d_{CP} values for which $\chi^2 < 1$





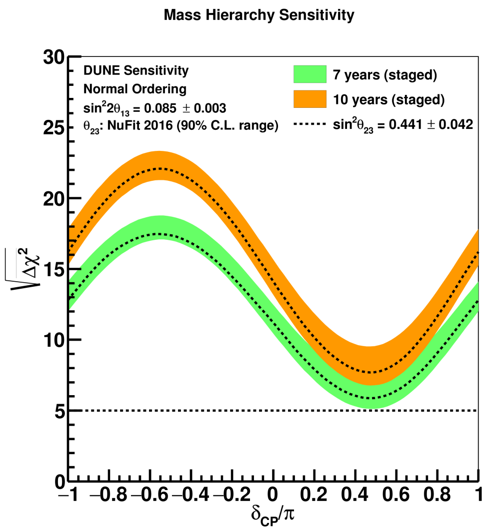
Hypothesis Testing

- Is the data more compatible with one or the other hypothesis?
  - H_0: data is more compatible with null hypothesis (eg: all bg, no signal)
  - H_1: data is more compatible with alternate hypothesis (eg: bg + signal)
- Use a test statistic (eg: χ^2) to choose between the hypotheses and quantify confidence in that choice
- "Significance level" or "type I error" (α): probability to reject H_0 if H_0 is true
  - Probability to mistakenly claim discovery
  - Chosen a priori
- "Misidentification probability" or "type II error" (β): probability to reject H_1 if H_1 is true. $1-\beta$ = "power"
  - Probability to miss a discovery
  - Chosen a priori
- p-value: probability if H_0 is true of getting a test statistic at least as extreme as the observed test statistic
  - Comes from the data – each test-statistic has associated p-value
  - Compare against the threshold set by $\alpha \rightarrow$ if p-value $\leq \alpha$, reject the null hypothesis

Hypothesis Testing Summary

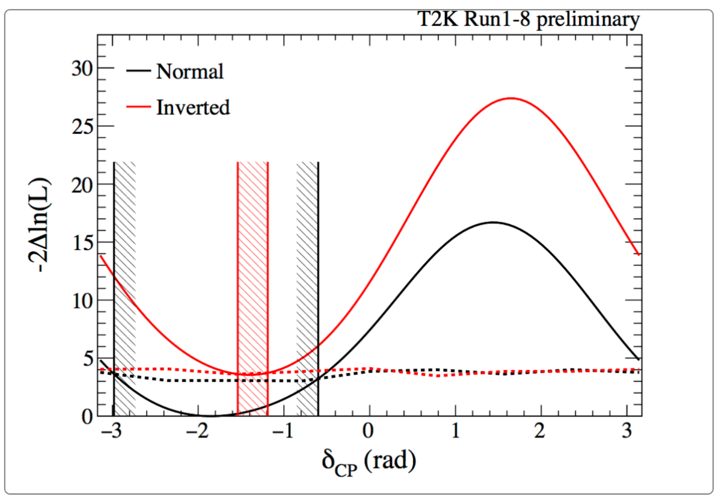
	Accept H_0	Reject H_0
H_0 is true	Correct $1-\alpha$	Error (type I) α (significance)
H_0 is false	Error (type II) β	Correct $1-\beta$ (power)

Physics Hypothesis Test Examples



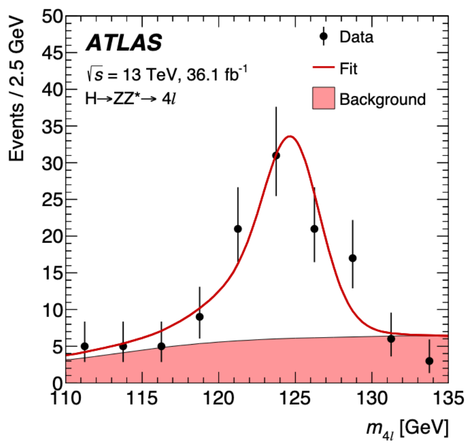
DUNE sensitivity to determination of the neutrino mass ordering, as a function of the true value of δ_{CP}

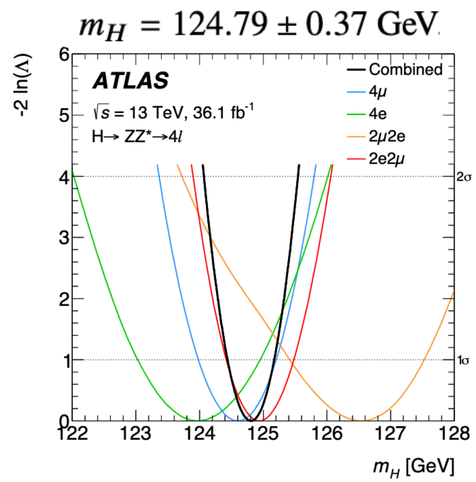
$\sqrt{\Delta\chi^2} > 5$ corresponds to $\alpha = 5.7 \times 10^{-7}$
($\sigma > 5$ is typical "discovery" threshold in HEP)



"T2K excludes CP conservation in neutrino oscillation at 95% CL"

Physics Observation





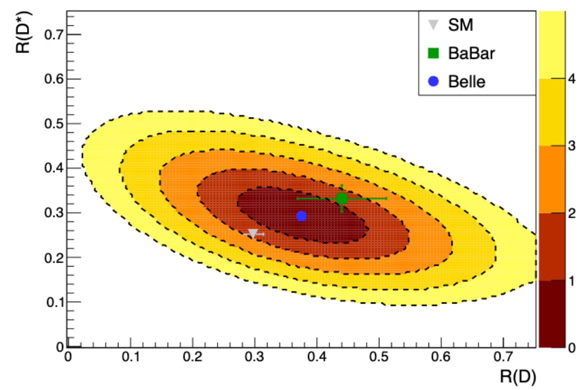


FIG. 6. Exclusion level of $R(D)$-$R(D^*)$ value assumptions in standard deviations, systematic uncertainties included.

ATLAS measurement of Higgs mass in H→ZZ→4ℓ channel

Belle measurement of branching ratios

$$R(D) = \frac{\mathcal{B}(\bar{B} \to D\tau^-\bar{\nu}_\tau)}{\mathcal{B}(\bar{B} \to D\ell^-\bar{\nu}_\ell)}$$

$$R(D^*) = \frac{\mathcal{B}(\bar{B} \to D^*\tau^-\bar{\nu}_\tau)}{\mathcal{B}(\bar{B} \to D^*\ell^-\bar{\nu}_\ell)}$$

Notes on Observations

- Typically quote 1D measurements with 1σ uncertainties, (often) separated into statistical and systematic uncertainties

$$\frac{B(\pi^0 \to e^+e^-\gamma)}{B(\pi^0 \to \gamma\gamma)} = [1.1559 \pm 0.0047(stat) \pm 0.0106(syst)]\%$$

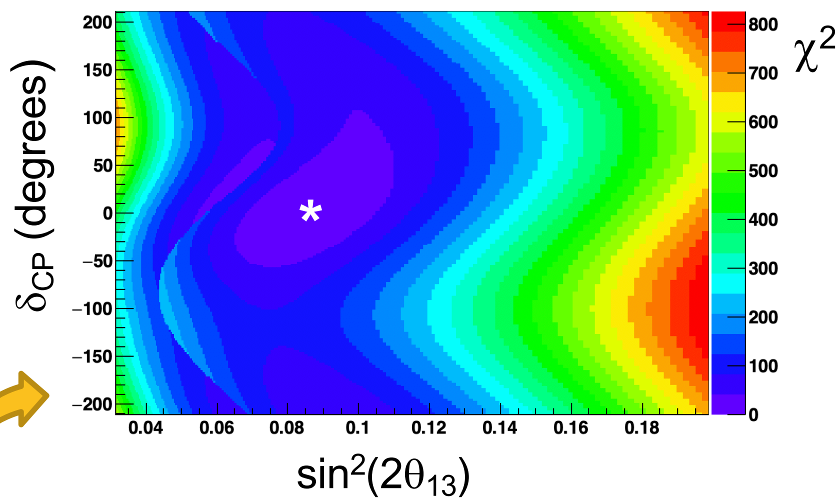
$\Rightarrow$

$$\frac{B(\pi^0 \to e^+e^-\gamma)}{B(\pi^0 \to \gamma\gamma)} = (1.1559 \pm 0.0116)\%$$

  - More on systematics later
- For 2D allowed regions:
  - Scan over parameter space and calculate χ^2 = -2ln(L) for each pair of values
  - Draw contours by selecting regions with χ^2 less than a particular critical value (hypothesis test)
  - Note that critical values correspond to different probabilities depending on the number of dimensions

Allowed Region Example

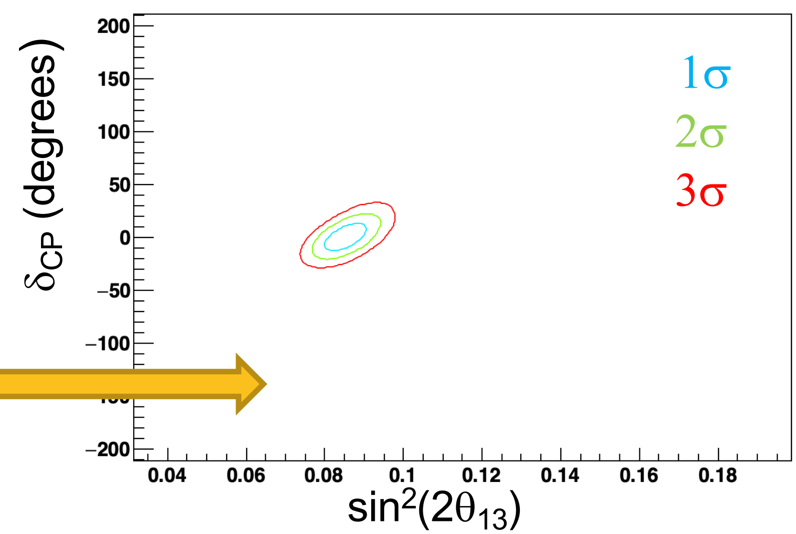
DUNE measurement of δ_{CP} and θ_{13}



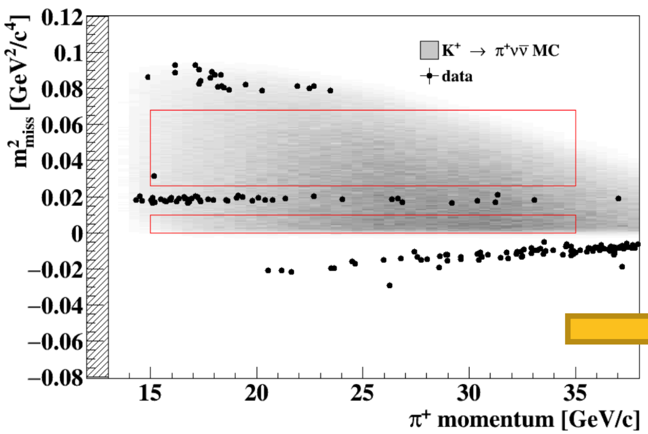
Critical values:

$(1-\alpha)$ (%)		$m=1$	$m=2$	$m=3$
68.27	1σ	1.00	2.30	3.53
90.		2.71	4.61	6.25
95.		3.84	5.99	7.82
95.45	2σ	4.00	6.18	8.03
99.		6.63	9.21	11.34
99.73	3σ	9.00	11.83	14.16

```
root [8] twodscan->Draw("colz")
root [9] twodscan->SetContour(3)
root [10] twodscan->SetContourLevel(0,2.30)
root [11] twodscan->SetContourLevel(1,6.18)
root [12] twodscan->SetContourLevel(2,11.83)
root [13] twodscan->Draw("cont1")
```

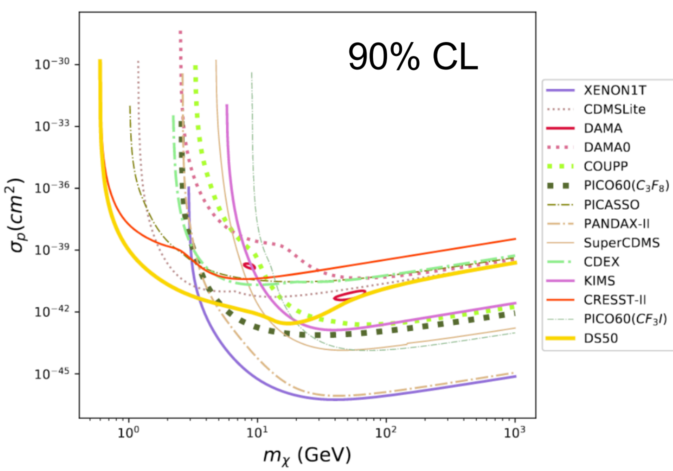


Physics Limit



NA62 branching ratio upper limit for $K^+ \rightarrow \pi^+\nu\nu$

$$\mathcal{B}\left(K^+ \rightarrow \pi^+\nu\bar{\nu}\right) < 14 \times 10^{-10}\ 95\,\%\ \text{C.L.}$$

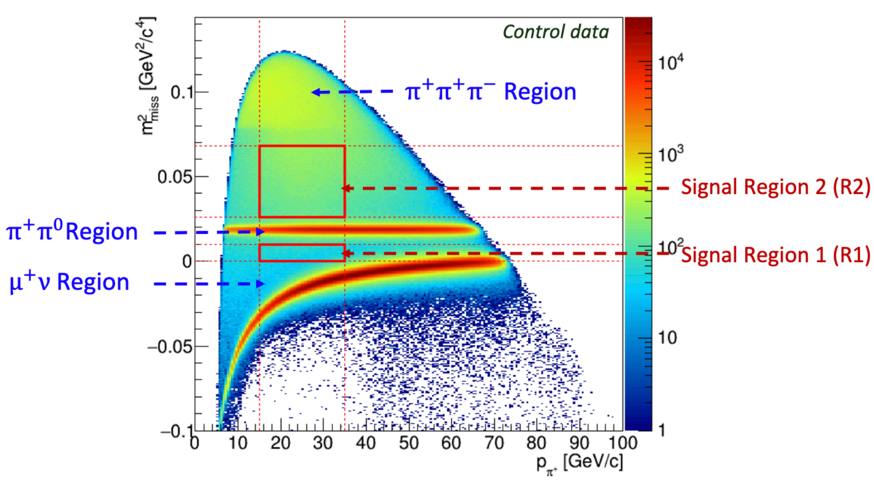


Summary of experimental limits on dark matter from direct detection experiments

Notes on Limits

- Single event sensitivity: true signal rate at which an experiment would expect to observe one signal event
  - SES = 1/Nε, where N is the total number of events ε is overall efficiency to observe a signal event
  - Single metric that includes intensity, run time, detector acceptance, analysis acceptance, etc.

Example: NA62 $K^+ \rightarrow \pi^+ \nu\nu$



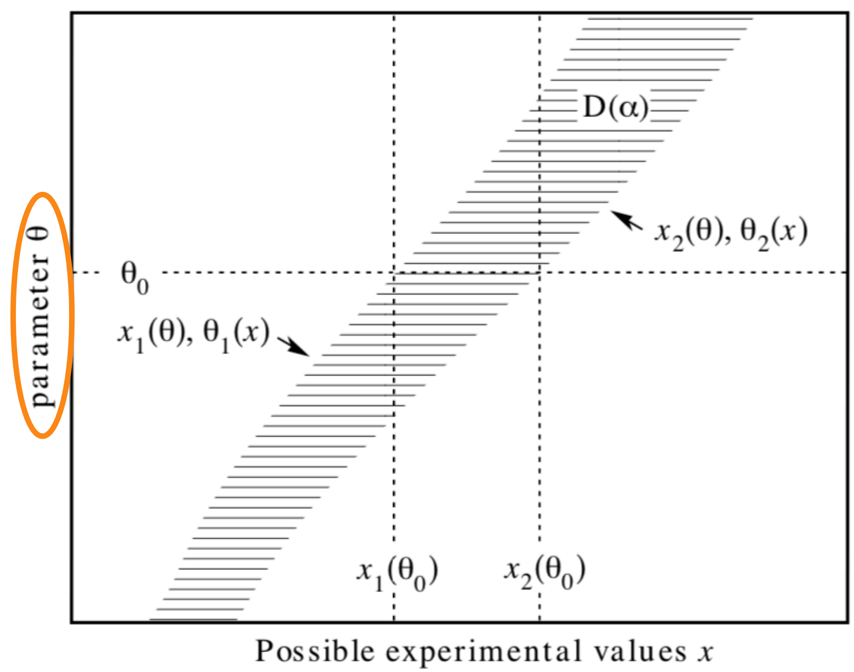
N_{K+} = (1.21 ± 0.02) x 10^{11}
ε = overall efficiency ~ 0.026
SES = (3.15 ± 0.24) x 10^{-10}
SM BR = (0.84 ± 0.10) x 10^{-10}
SM Expected Signal: 0.267 ± 0.038

Notes on Limits

- Upper limit vs discovery:
  - Hypothesis test where the null hypothesis (H_0) is background only and the alternative hypothesis (H_1) is signal + background
  - In HEP convention, 3σ significance is considered "evidence for" and 5σ significance is considered the threshold for discovery(*)
- Upper limit s^{up} is the upper extreme of a fully asymmetric confidence interval $[0,s^{up}]$
- Neyman confidence interval procedure:
  - Scan allowed range of parameter space
  - Given a value θ_ι of θ, compute the interval $[x_1(\theta_\iota),x_2(\theta_\iota)]$ that contains x with the desired confidence level
  - For the observed value of x, invert the confidence belt to find the corresponding interval $[\theta_1(x),\theta_2(x)]$

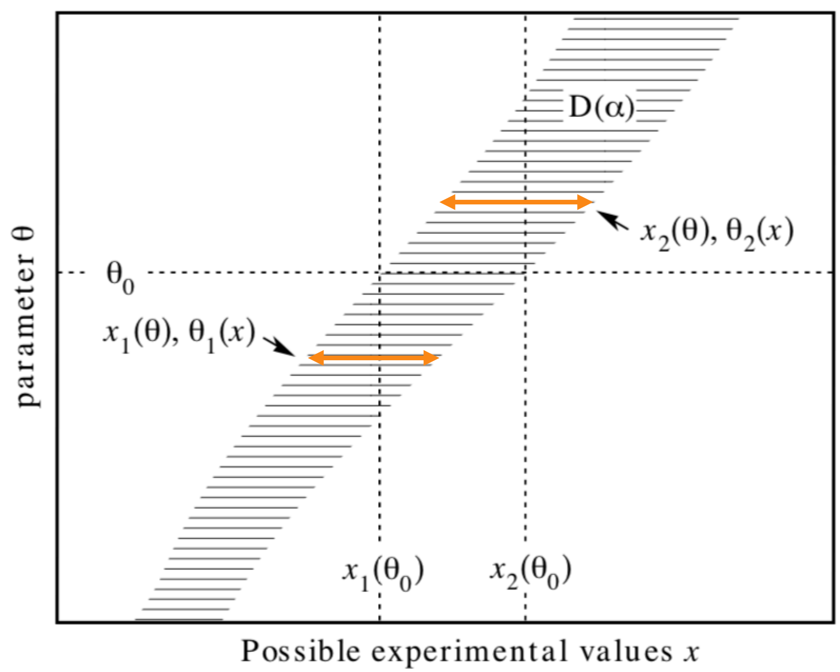
Notes on Limits

- Upper limit vs discovery:
  - Hypothesis test where the null hypothesis (H_0) is background only and the alternative hypothesis (H_1) is signal + background
  - In HEP convention, 3σ significance is considered "evidence for" and 5σ significance is considered the threshold for discovery(*)
- Upper limit s^{up} is the upper extreme of a fully asymmetric confidence interval $[0,s^{up}]$
- Neyman confidence interval procedure:
  - Scan allowed range of parameter space
  - Given a value θ_ι of θ, compute the interval $[x_1(\theta_\iota),x_2(\theta_\iota)]$ that contains x with the desired confidence level
  - For the observed value of x, invert the confidence belt to find the corresponding interval $[\theta_1(x),\theta_2(x)]$



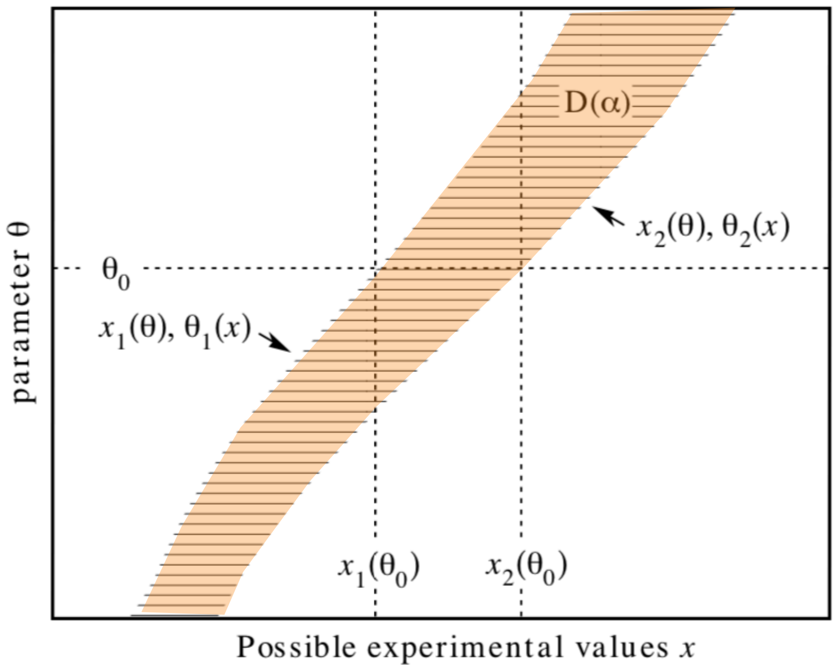
Notes on Limits

- Upper limit vs discovery:
  - Hypothesis test where the null hypothesis (H_0) is background only and the alternative hypothesis (H_1) is signal + background
  - In HEP convention, 3σ significance is considered "evidence for" and 5σ significance is considered the threshold for discovery(*)
- Upper limit s^{up} is the upper extreme of a fully asymmetric confidence interval $[0,s^{up}]$
- Neyman confidence interval procedure:
  - Scan allowed range of parameter space
  - Given a value θ_ι of θ, compute the interval $[x_1(\theta_\iota),x_2(\theta_\iota)]$ that contains x with the desired confidence level
  - For the observed value of x, invert the confidence belt to find the corresponding interval $[\theta_1(x),\theta_2(x)]$



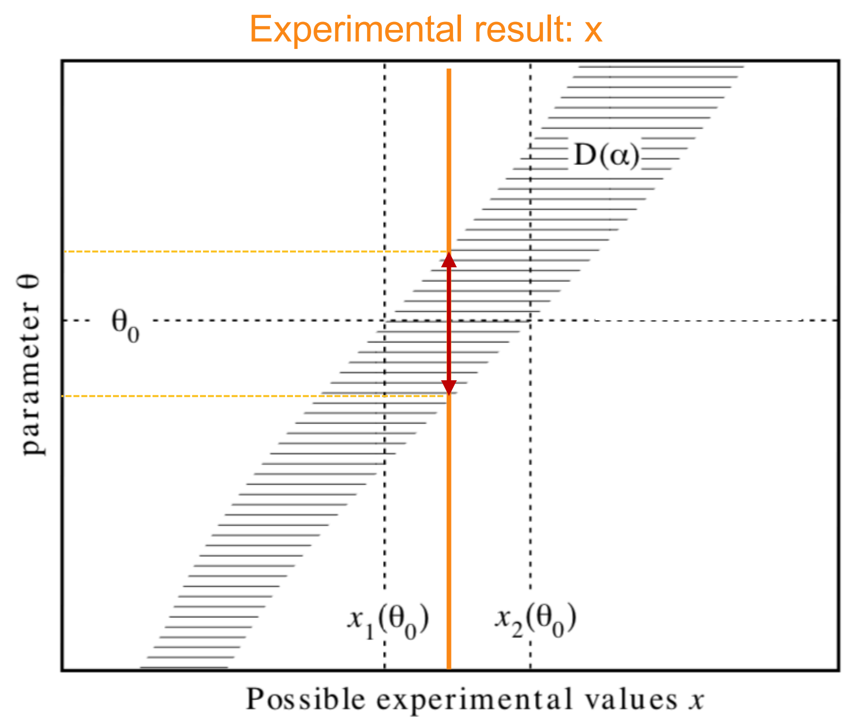
Notes on Limits

- Upper limit vs discovery:
  - Hypothesis test where the null hypothesis (H_0) is background only and the alternative hypothesis (H_1) is signal + background
  - In HEP convention, 3σ significance is considered "evidence for" and 5σ significance is considered the threshold for discovery(*)
- Upper limit s^{up} is the upper extreme of a fully asymmetric confidence interval $[0,s^{up}]$
- Neyman confidence interval procedure:
  - Scan allowed range of parameter space
  - Given a value θ_ι of θ, compute the interval $[x_1(\theta_\iota),x_2(\theta_\iota)]$ that contains x with the desired confidence level
  - For the observed value of x, invert the confidence belt to find the corresponding interval $[\theta_1(x),\theta_2(x)]$



Notes on Limits

- Upper limit vs discovery:
  - Hypothesis test where the null hypothesis (H_0) is background only and the alternative hypothesis (H_1) is signal + background
  - In HEP convention, 3σ significance is considered "evidence for" and 5σ significance is considered the threshold for discovery(*)
- Upper limit s^{up} is the upper extreme of a fully asymmetric confidence interval $[0,s^{up}]$
- Neyman confidence interval procedure:
  - Scan allowed range of parameter space
  - Given a value θ_0 of θ, compute the interval $[x_1(\theta_\iota),x_2(\theta_\iota)]$ that contains x with the desired confidence level
  - For the observed value of x, invert the confidence belt to find the corresponding interval $[\theta_1(x),\theta_2(x)]$

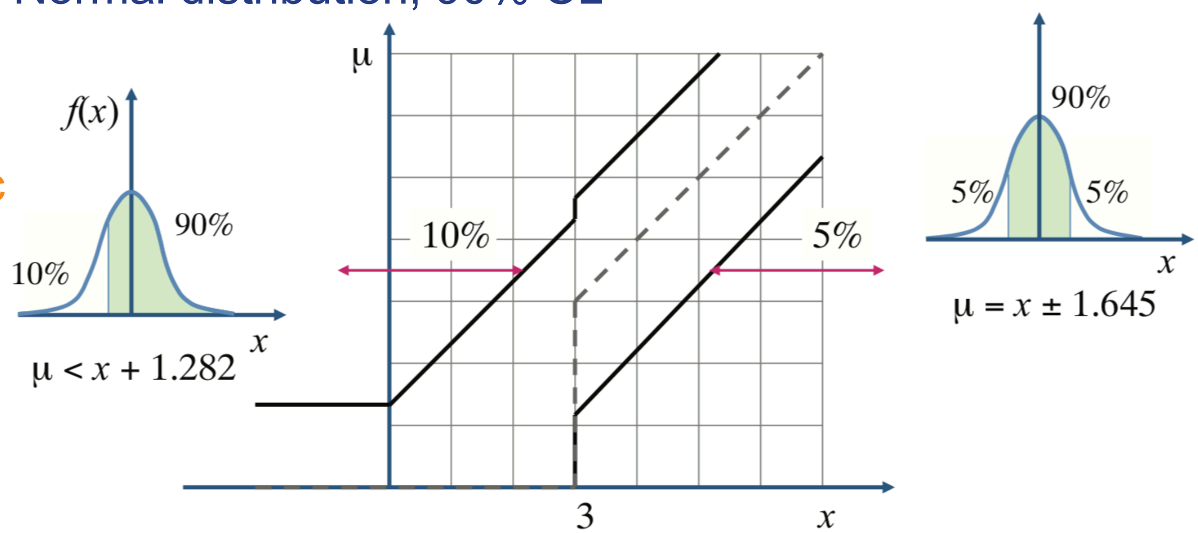


Feldman-Cousins Procedure

- A problem:
  - For a limit, use fully asymmetric confidence interval
  - For a measurement, quote CV and symmetric confidence interval
  - Typically the threshold for measurement is 3σ
  - “Coverage” of confidence belt is incorrect

Example: Normal distribution, 90% CL

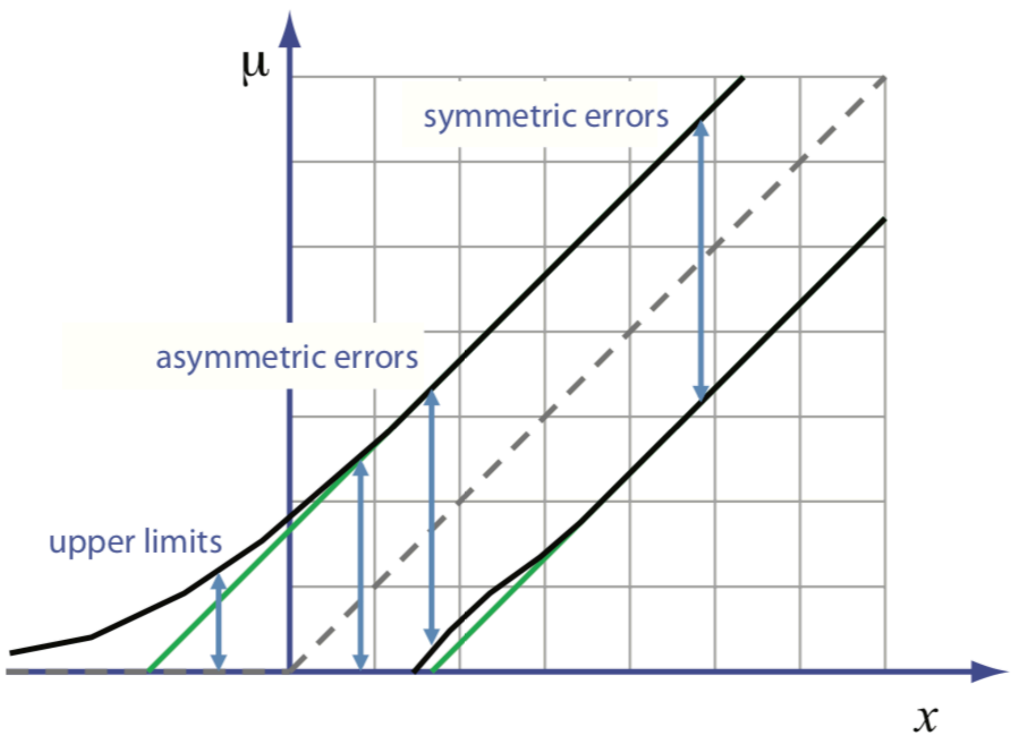
Below 3σ,
Asymmetric
interval



Above 3σ,
symmetric
interval

Feldman-Cousins Procedure for Limits

- A solution (Feldman-Cousins procedure) is designed to create a confidence belt that transitions smoothly from a limit to asymmetric errors to fully symmetric errors



Feldman-Cousins Confidence Interval:

$$R_\mu = \left\{ x : \frac{L(x;\theta_0)}{L(x;\hat{\theta})} > k_\alpha \right\}$$

Procedure must be done numerically and can be computationally expensive!

CL_S is an alternative procedure for setting limits.

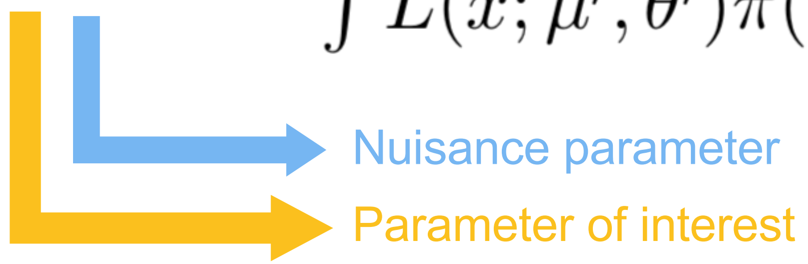
Systematic Uncertainties

- Any uncertainty not resulting from data statistics is a systematic uncertainty. A few examples:
  - Detector calibration uncertainty (energy scale, momentum resolution, vertex position, etc)
  - Trigger efficiency uncertainty
  - Beam luminosity uncertainty
  - PDG uncertainty on branching ratio of normalization mode
  - Neutrino interaction cross-section uncertainty
- May have a statistical component, eg: with more data, I may be able to make a more precise energy scale measurement
- Become more important for high-precision experiments where statistical uncertainty is small
  - "Systematics limited" when systematic uncertainties become >> than statistical uncertainties
- It is not "conservative" to inflate systematics!!!
- Systematics are often handled in fits as "nuisance parameters"

Systematic Uncertainty in Bayesian Analysis

- No "special" treatment required – if the systematic uncertainties are built into the posterior, they will be properly incorporated into resulting inference

$$P(\mu,\theta|\vec{x}) = \frac{L(\vec{x};\mu,\theta)\pi(\mu,\theta)}{\int L(\vec{x};\mu',\theta')\pi(\mu',\theta')\,\mathrm{d}\mu'\,\mathrm{d}\theta'}$$

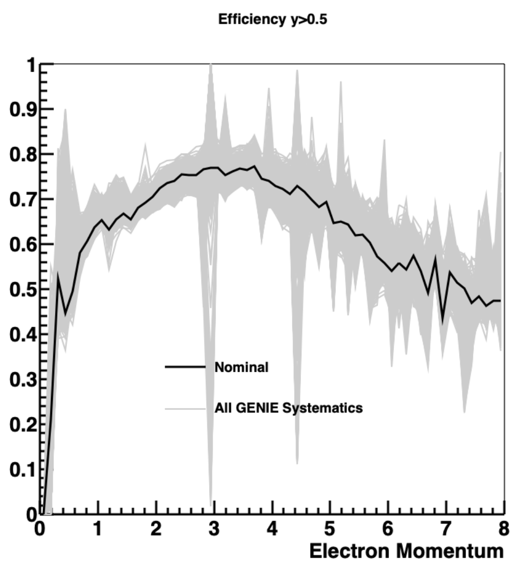


Find "marginal" PDF by integrating over nuisance parameter ("marginalization")

$$P(\mu|\vec{x}) = \int P(\mu,\theta|\vec{x}),\mathrm{d}\theta = \frac{\int L(\vec{x};\mu,\theta)\pi(\mu,\theta)\,\mathrm{d}\theta}{\int L(\vec{x};\mu',\theta)\pi(\mu',\theta)\,\mathrm{d}\mu'\,\mathrm{d}\theta}$$

Systematic Uncertainty in Frequentist Analysis

- Monte Carlo "throws":
  - Randomly choose values for each nuisance parameter according to its PDF, creating a "thrown universe"
  - Perform nominal analysis in this universe
  - Repeat many times
  - Plot distribution of results
  - Can assign 1σ error band as central 68% of resulting measurements



Example: Selection efficiency for Monte Carlo with neutrino interaction systematics

- Each uncertain parameter is assumed to be Gaussian distributed
- Each curve is a "thrown universe" where each uncertain parameter has been chosen randomly

```
for (uint j = 0; j < 1000; ++j) {
  SystShifts shift;
  for(const ISyst* s: systlist) shift.SetShift(s, gRandom->Gaus());
```

Systematic Uncertainty in Frequentist Analysis

- Δ method:
  - If systematic uncertainties are uncorrelated, can combine uncertainties by adding in quadrature
- Pull method or Profile Likelihood method
  - Similar to Bayesian approach – fit for both the parameters of interest and nuisance parameters simultaneously, with nuisance parameters constrained by uncertainty in their values
  - ”Profile” over nuisance parameters (minimize -2ln(L)) with respect to these parameters)

Systematic Uncertainty in Frequentist Analysis

- Δ method:
  - If systematic uncertainties are uncorrelated, can combine uncertainties by adding in quadrature
- Pull method or Profile Likelihood method
  - Similar to Bayesian approach – fit for both the parameters of interest and nuisance parameters simultaneously, with nuisance parameters constrained by uncertainty in their values
  - ”Profile” over nuisance parameters (minimize -2ln(L)) with respect to these parameters)

Example: Binned log likelihood function for neutrino oscillation measurements

$$\Delta\chi^2 = 2\sum_i^N \left[N_i^{exp}(\boldsymbol{\theta}, \boldsymbol{f}) - N_i^{true} + N_i^{true} \ln\left[\frac{N_i^{true}}{N_i^{exp}(\boldsymbol{\theta}, \boldsymbol{f})}\right]\right]$$

Likelihood function (θ are oscillation parameters we want to measure)

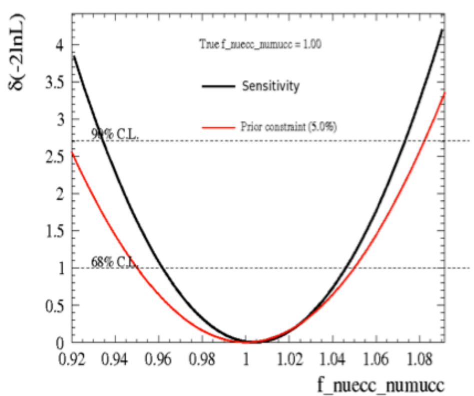
$$+ \sum_j^{N_{systs}} \frac{f_j^2}{\sigma_{f_j}^2} + \sum_k^{N_{oscs}} \frac{\left(\theta_k^{nominal} - \theta_k\right)^2}{\sigma_{\theta_k}^2}$$

“Penalty term” (f,θ are the nuisance parameters)

Systematic Uncertainty in Frequentist Analysis

- ∆ method:
  - If systematic uncertainties are uncorrelated, can combine uncertainties by adding in quadrature
- Pull method or Profile Likelihood method
  - Similar to Bayesian approach – fit for both the parameters of interest and nuisance parameters simultaneously, with nuisance parameters constrained by uncertainty in their values
  - ”Profile” over nuisance parameters (minimize -2ln(L) with respect to these parameters)

Example: Binned log likelihood function for neutrino oscillation measurements



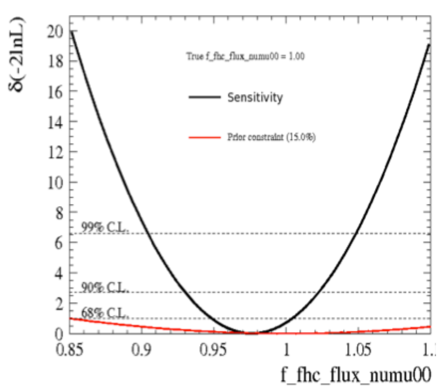
Constraint on nuisance parameter after fit

Constraint on nuisance parameter before fit

Systematic Uncertainty in Frequentist Analysis

- Δ method:
  - If systematic uncertainties are uncorrelated, can combine uncertainties by adding in quadrature
- Pull method or Profile Likelihood method
  - Similar to Bayesian approach – fit for both the parameters of interest and nuisance parameters simultaneously, with nuisance parameters constrained by uncertainty in their values
  - "Profile" over nuisance parameters (minimize -2ln(L) with respect to these parameters)

Example: Binned log likelihood function for neutrino oscillation measurements



Constraint on nuisance parameter after fit

Constraint on nuisance parameter before fit

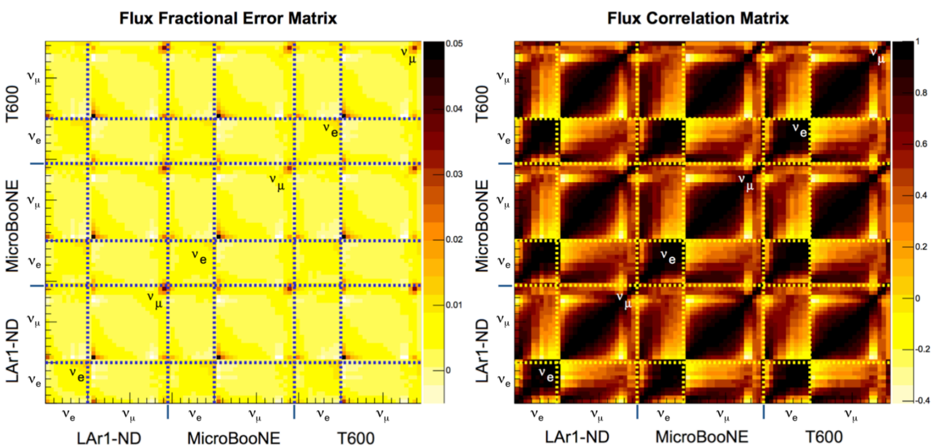
Systematic Uncertainty in Frequentist Analysis

- Covariance matrix
  - Includes correlations among parameters
  - In principle is equivalent to the pull method

$$\chi^2 = \sum_{i,j=1}^{n} (m_i - M_i(\vec{\theta}))\, C_{ij}^{-1}\, (m_j - M_j(\vec{\theta}))$$

Covariance matrix

Example: Flux uncertainty for three detectors in SBN program

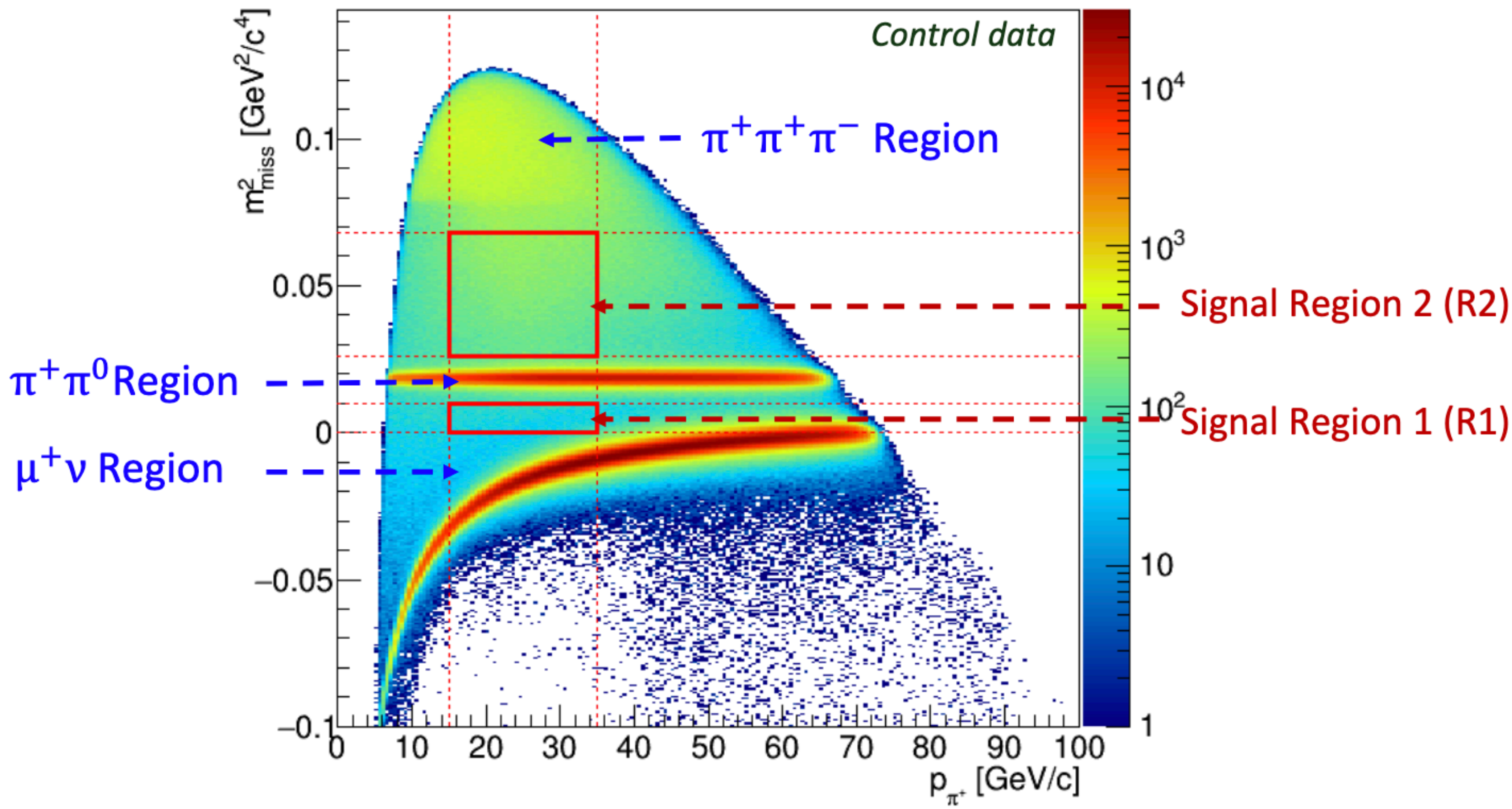


Matrix constructed using thrown universes:

$$E_{ij} = \frac{1}{\mathcal{N}} \sum_{m=1}^{\mathcal{N}} [N^i_{\rm CV} - N^i_m] \times [N^j_{\rm CV} - N^j_m]$$

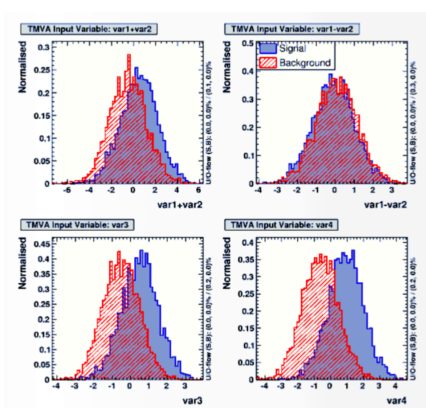
"Cut and Count" Analysis

Example: NA62

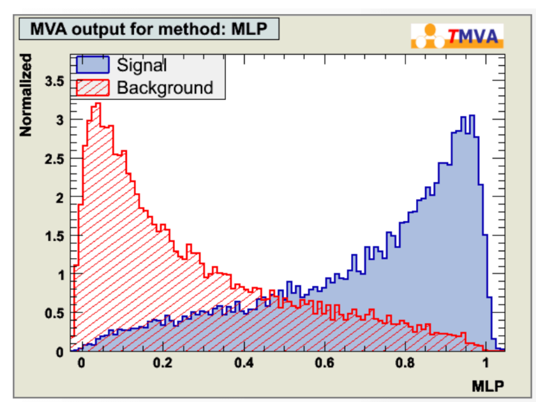


Statistical Analysis for PID/Event Selection

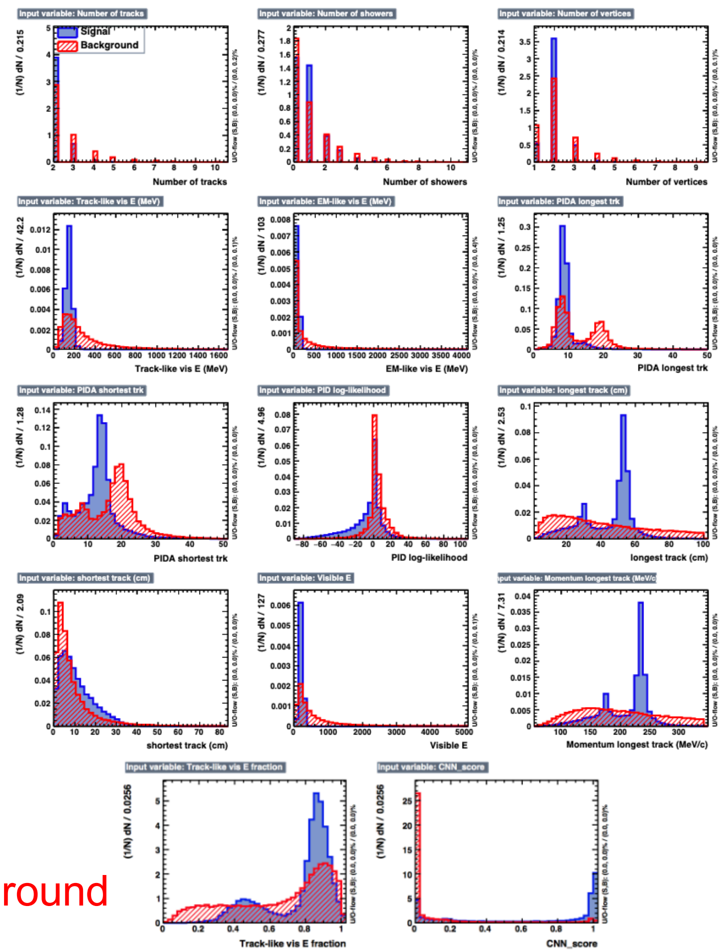
- Multivariate discriminator: test statistic for sample with multiple variables
  - Instead of 5 individual cuts on kinematic quantities, combine this information into a single discriminant and cut on that
  - Is fundamentally a hypothesis test "trained" on MC or other sample where either H_0 or H_1 is known to be true
- Examples: boosted decision tree, neural network, machine learning, deep learning
  - Algorithm "learns" features of signal and background by iteratively applying weights from previous layer or iteration
  - Details in next talk



Magic?

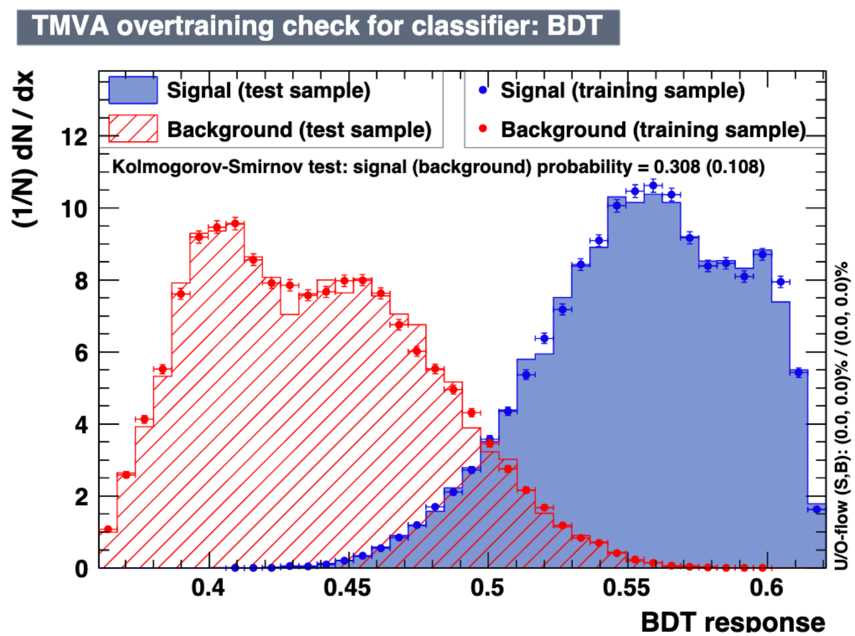


Boosted Decision Tree Real World Example

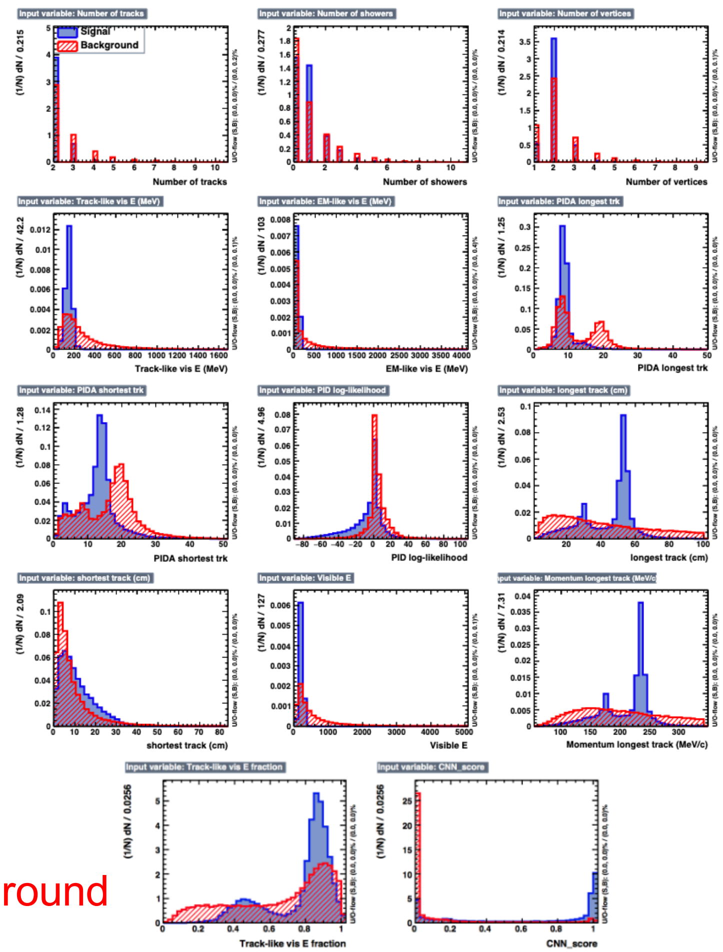


Signal
Background

- 14 input variables; uses TMVA (Toolkit for multivariate analysis with ROOT)

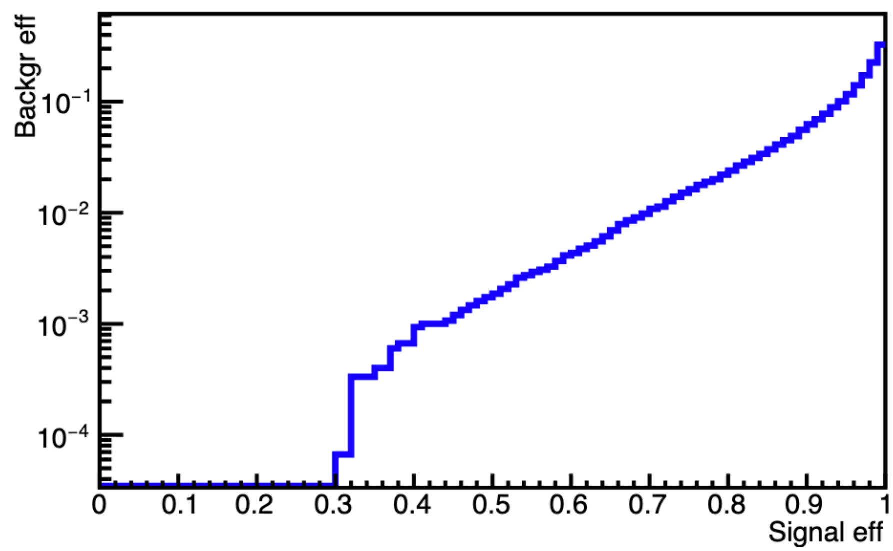


Boosted Decision Tree Real World Example



Signal
Background

- 14 input variables; uses BDT method in TMVA (Toolkit for multivariate analysis with ROOT)



Suggested Simple Exercises

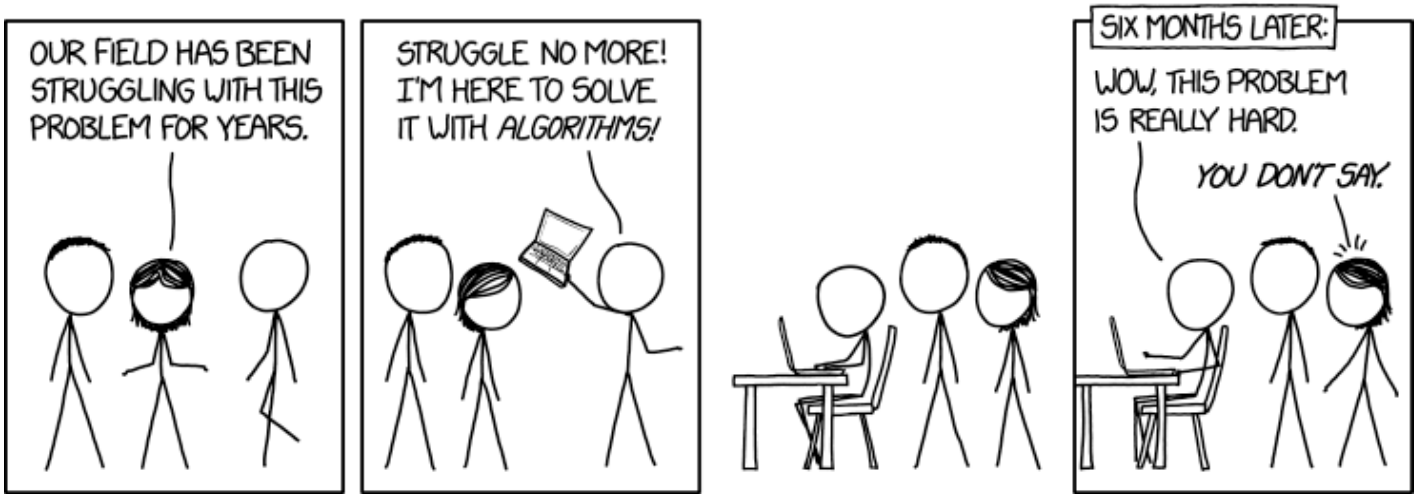
- Repeat my cosmic ray example from slides 10-12 using your favorite programming/scripting/plotting tools
- For the same example, do a scan of the parameter ν calculating -2ln(L) for a Poisson distributed variable for each value of ν. Find the uncertainty in your estimate by determining where the difference in likelihoods crosses +1
  - (An easy way to do this is a "pol2" fit to a histogram filled with -2ln(L)-1)
  - Try changing the number of data points. How much data do you really need to take?
  - If you repeat the exercise over and over again, how often does your range not include the true value of ν?
  - Try making a 2σ or 3σ confidence interval instead. Now how often does your range not include the true value of ν?
- Complicate things by introducing a nuisance parameter – say the efficiency of your detector – that is Gaussian distributed ($\varepsilon \pm \sigma_\varepsilon$)
  - Easiest way to do this (IMO) is a pull term added to your log likelihood
  - What happens if σ_ε is tiny? Does it matter what value ε has?
  - At what value of σ_ε does your sensitivity start to suffer? When do you become systematics limited?

References

- Much of this talk is following and using figures from arXiv:1609.04150 (Lista, "Practical Statistics for Particle Physicists", 2017)
- Cool visualizations: https://seeing-theory.brown.edu/
- Other useful resources:
  - PDG statistics review: http://pdg.lbl.gov/2015/reviews/rpp2015-rev-statistics.pdf
  - PhyStat/PhyStat-nu workshops
    - eg: https://indico.ipmu.jp/indico/event/82
  - Scott Oser lecture notes: https://www.phas.ubc.ca/~oser/p509
  - RooFit/RooStats:
    - https://root.cern.ch/roofit
    - https://twiki.cern.ch/twiki/bin/view/RooStats/RooStats
  - Stan: https://mc-stan.org/

Physicists & Statisticians

Goes both ways!



https://xkcd.com/1831/