Statistical Methods used in Neutrino Oscillation Experiments

Kamil Skwarczyński

NCBJ PhD Seminar

30.03.2023

NATIONAL CENTRE FOR NUCLEAR RESEARCH ŚWIERK

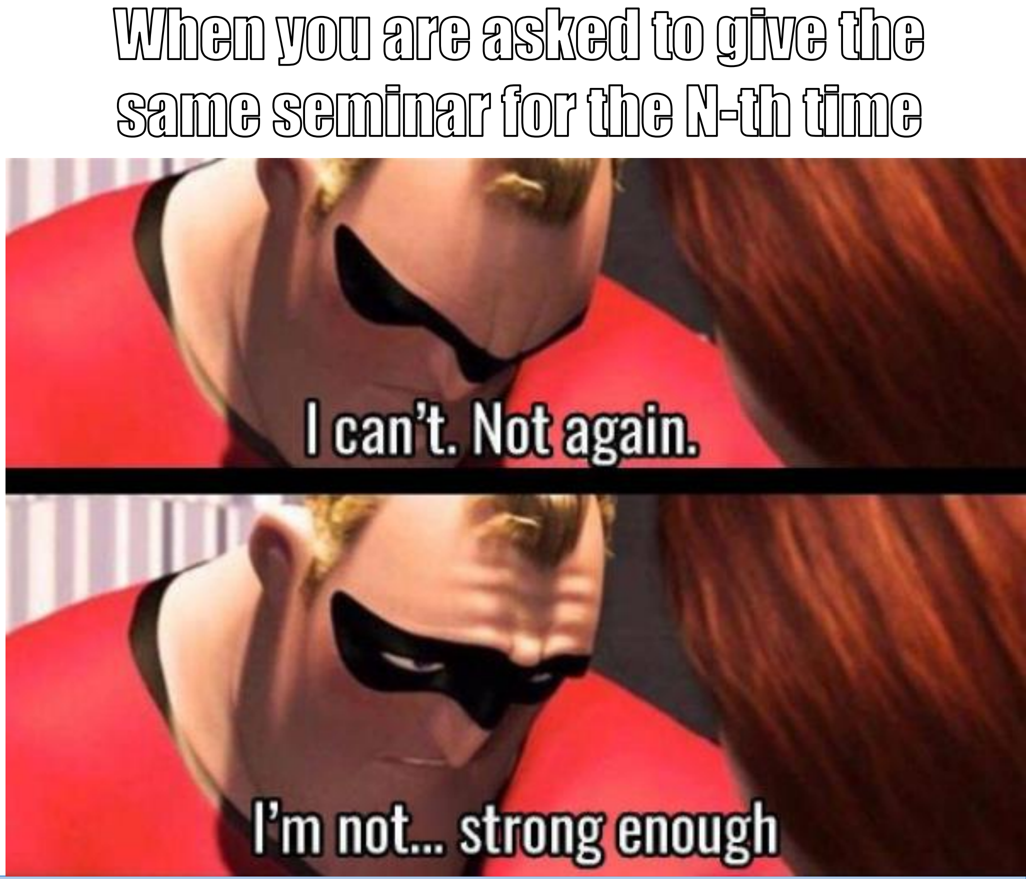
T2K

Introduction

This seminar will be partially related to my thesis.
I had to give several seminars/presentation about my results. I am tired showing exactly the same talk over and over again.

This talk will be about statistical method used in T2K experiment.

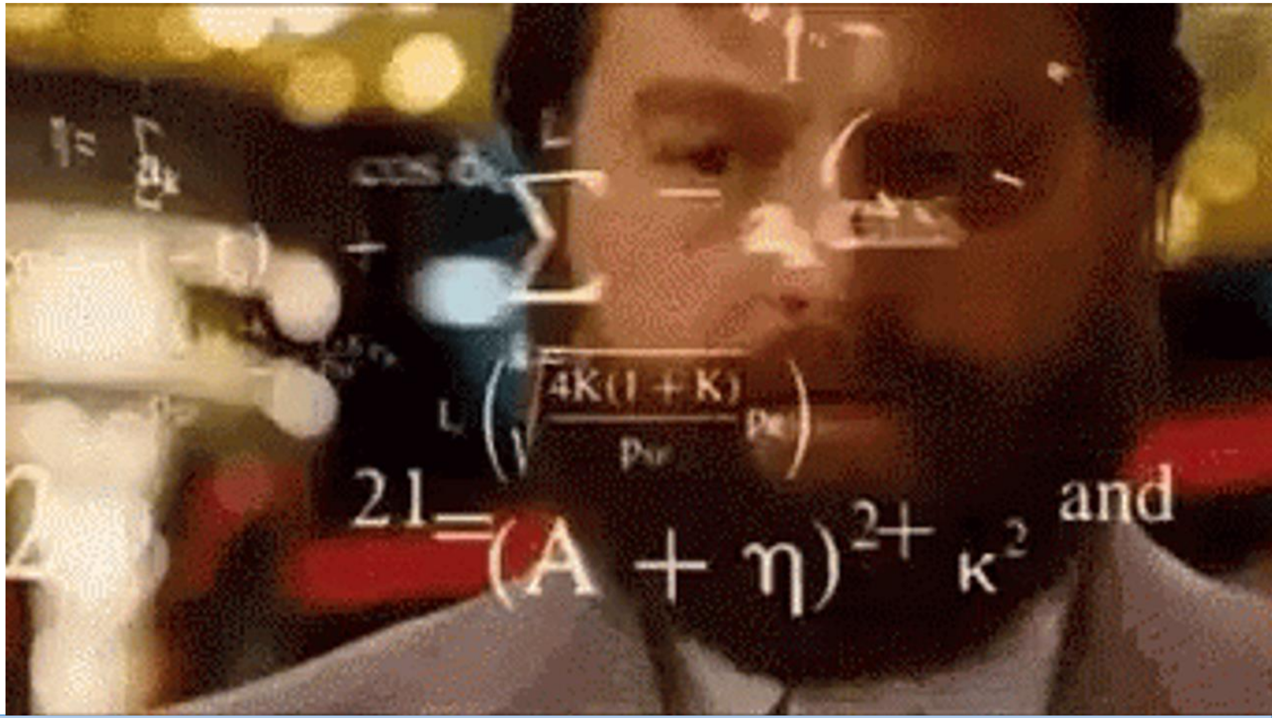
Thesis has been already submitted.



Outline

Outline

- Basics – Poisson
- MC Statistic – Barlow Beeston
- Another Approach
  Dembinski and Abdelmotteleb
- Ice Cube – The Holy Grail
- Pearson – Test Meant to Fail



Basics - Poisson

Introduction

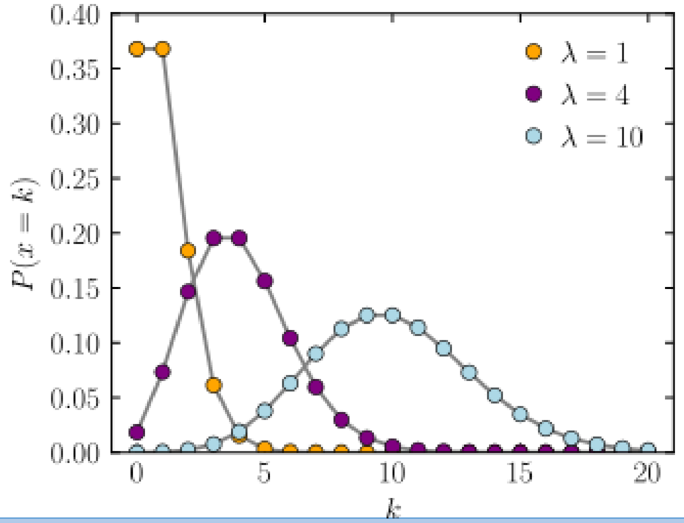
In particle physics experiment we are comparing data and parametrized MC prediction.

MC depends on some physical parameters like Highs mass, neutrino mixing angles. On top of that we include many nuisance parameters related to modeling of cross-section detector etc.

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Such analysis is basically a counting experiment.

The probability distribution describing the counting experiment is given by discrete Poisson distribution:



Poisson

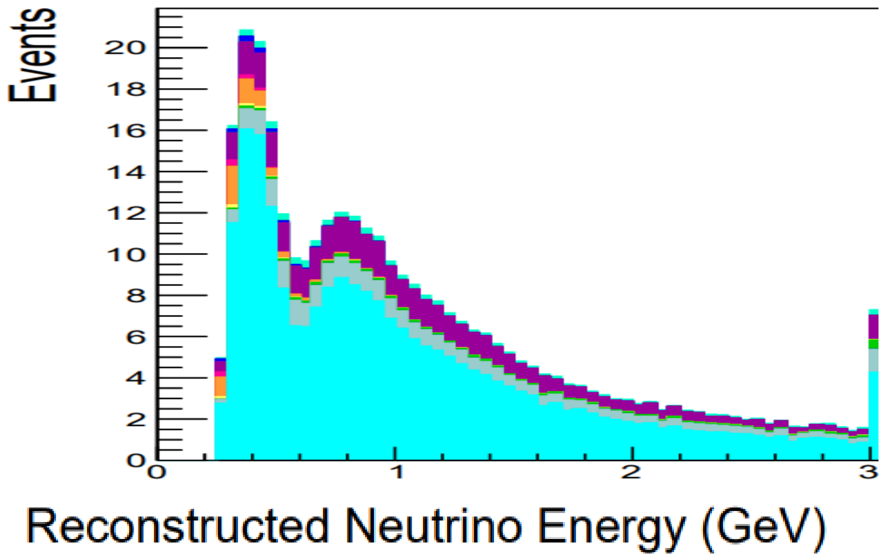
We usually don't count the global number of events but a number in a particular kinematic region often called bin.

For example, neutrino oscillations have characteristic shape as a function of neutrino momentum

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\mathcal{L}(Z|\vec{\theta}) = \prod_i \frac{\left(N_i^{\mathrm{MC}}(\vec{\theta})\right)^{N_i^{\mathrm{data}}} e^{-N_i^{\mathrm{MC}}(\vec{\theta})}}{N_i^{\mathrm{data}}!}$$

where i enumerates bins of p_μ and $\cos\theta_\mu$, N_i^{data} is the number of data events in i-th bin, while $N_i^{\mathrm{MC}}(\theta)$ is the number of events predicted by MC, and $\vec{\theta}$ is vector of parameters describing our model.



Wilks Theorem

According to Wilks' theorem, the logarithm of likelihood (LLH) ratio approaches asymptotically Chi2. This works if you have big number of data points or bins.

$$\Delta\chi^2 = -2\ln\left(\frac{\mathcal{L}(\vec{\theta})}{\mathcal{L}_0(\vec{\theta}_0)}\right)$$

$$\mathcal{L}(Z|\vec{\theta}) = \prod_i \frac{\left(N_i^{\text{MC}}(\vec{\theta})\right)^{N_i^{\text{data}}} e^{-N_i^{\text{MC}}(\vec{\theta})}}{N_i^{\text{data}}!}$$

where $\mathcal{L}_0(\vec{\theta}_0)$ is the likelihood of a null hypothesis (in this case, an ideal situation when expected values are in perfect agreement with data, which can be written as $N_i^{MC}(\vec{\theta}_0) = N_i^{data}$).

$$\Delta\chi^2 = -2\ln\left(\prod_i \left(N_i^{\text{MC}}(\vec{\theta})\right)^{N_i^{\text{data}}} e^{-N_i^{\text{MC}}(\vec{\theta})}\right) + 2\ln\left(\prod_i \left(N_i^{\text{MC}}(\vec{\theta}_0)\right)^{N_i^{\text{data}}} e^{-N_i^{\text{MC}}(\theta_0)}\right)$$

$$\xrightarrow{N_i^{MC}(\vec{\theta}_0)=N_i^{data}} 2\sum_i\left[-\ln\left(\left(N_i^{\text{MC}}(\vec{\theta})\right)^{N_i^{\text{data}}} e^{-N_i^{\text{MC}}(\vec{\theta})}\right) + \ln\left(\left(N_i^{\text{data}}\right)^{N_i^{\text{data}}} e^{-N_i^{\text{data}}}\right)\right]$$

$$\Delta\chi^2 = -2\log\mathcal{L}_{\text{Stat}} = 2\sum_i\left[N_i^{\text{MC}}(\vec{\theta}) - N_i^{\text{data}} + N_i^{\text{data}}\ln\left(\frac{N_i^{\text{data}}}{N_i^{\text{MC}}(\vec{\theta})}\right)\right]$$

LLH Scan

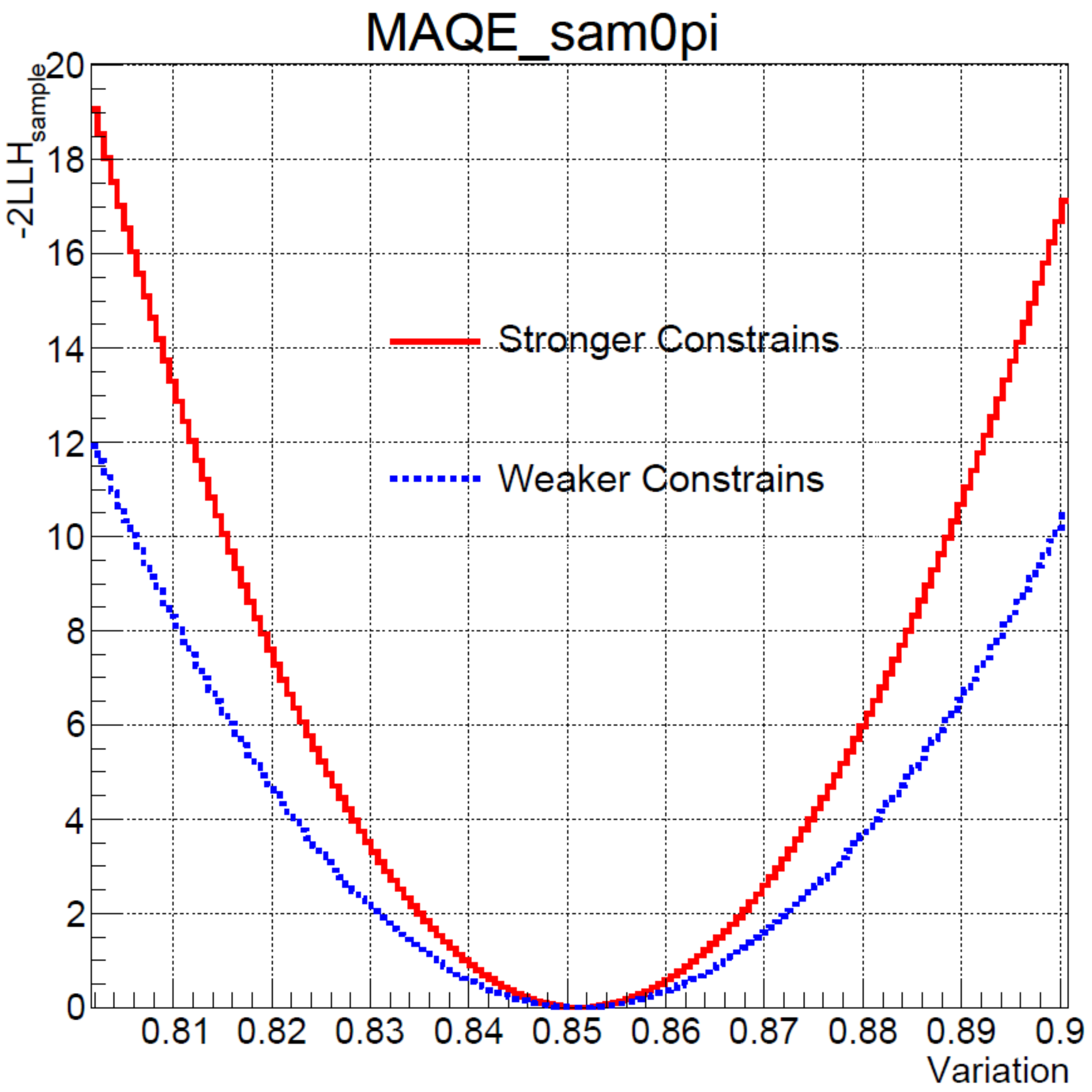
$$\Delta\chi^2 = -2\log\mathcal{L}_{\text{Stat}} = 2\sum_i \left[N_i^{\text{MC}}(\vec{\theta}) - N_i^{\text{data}} + N_i^{\text{data}} \ln\left(\frac{N_i^{\text{data}}}{N_i^{\text{MC}}(\vec{\theta})}\right)\right]$$

LLH scan helps us to determine sensitivity to a given model parameter.

We take two MC and set both to default values of parameters, one replaces data in LLH formula. If we compare such MCs, then for default values LLH is 0.

Then we start to change value of a chosen parameter in one MC, while the other one is fixed. Our MCs are then no longer in agreement which results in increase of LLH.

Rapidly changing LLH means we are greatly sensitive to given parameter. If LLH isn't changing at all then we lack sensitivity to given parameter.

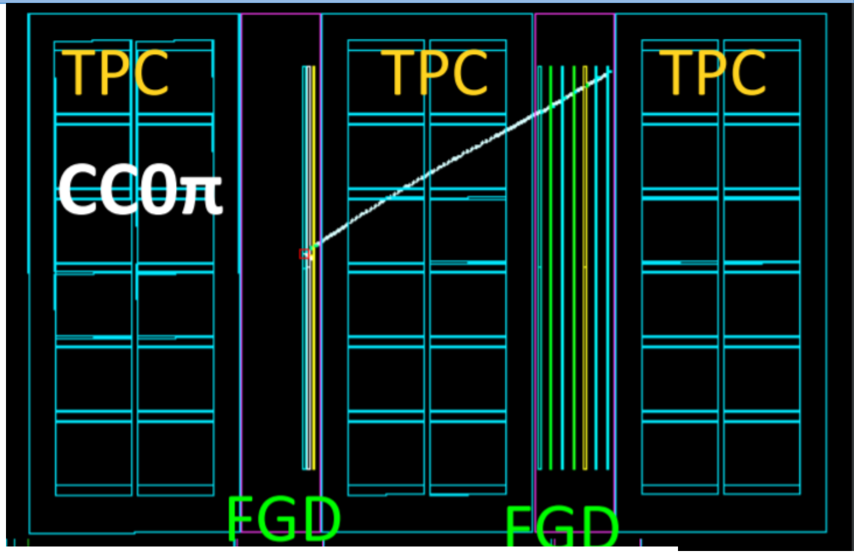


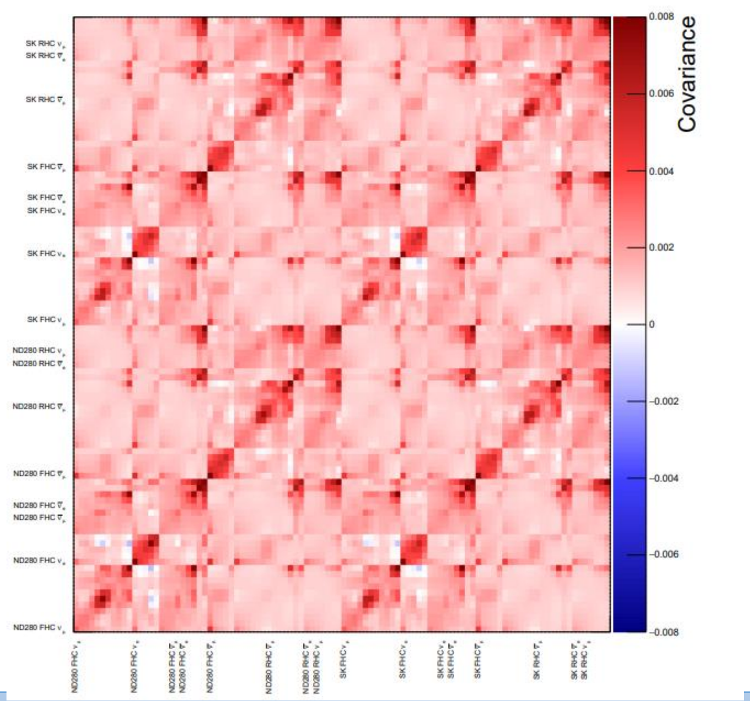
Systematic Parameters

MC depends on many nuisance parameters, we have to include detector effects like PID efficiencies, matchings. As our detector are very complex.

Such parameters are often correlated. Thus we use correlation matrix between different effects.

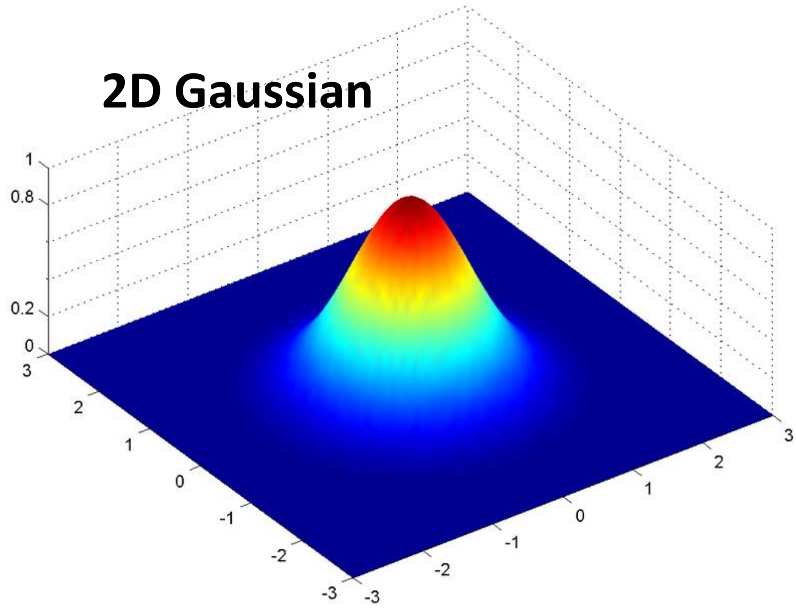
Each systematic parameter also has prior error. We usually take prior based on external measurements.





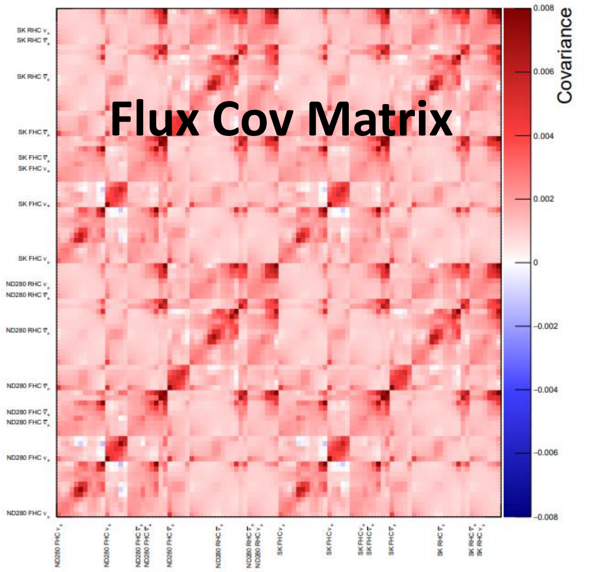
Penalty Term

The derivation presented here is not yet complete as prior uncertainties and correlations of parameters are not included in $\Delta\chi^2$. Model parameters are described by multivariate normal distributions with covariances of V_θ



$$\pi(\vec{\theta}) = \prod_{ij} \frac{1}{(2\pi)^{k/2}|(V_\theta)_{ij}|^{1/2}} e^{-\frac{1}{2}\Delta\vec{\theta}_i \left(V_\theta^{-1}\right)_{ij} \Delta\vec{\theta}_j^T},$$

where k is dimension of $\vec{\theta}$ parameter vector and $\Delta\vec{\theta} = \vec{\theta} - \vec{\theta}_0$, with $\vec{\theta}_0$ being prior value.



Likelihood

$$\pi(\vec{\theta}) = \prod_{ij} \frac{1}{(2\pi)^{k/2}|(V_\theta)_{ij}|^{1/2}} e^{-\frac{1}{2}\Delta\vec{\theta}_i \left(V_\theta^{-1}\right)_{ij} \Delta\vec{\theta}_j^T},$$

$$\Delta\chi^2 = -2\ln\left(\frac{\pi(\vec{\theta})\mathcal{L}(\vec{\theta})}{\pi_0(\vec{\theta}_0)\mathcal{L}_0(\vec{\theta}_0)}\right)$$
$$= -2\ln\left(\frac{\mathcal{L}(\vec{\theta})}{\mathcal{L}_0(\vec{\theta}_0)}\right) - 2\ln\left(\frac{\pi(\vec{\theta})}{\pi_0(\vec{\theta}_0)}\right) = -2\log\mathcal{L}_{\mathrm{Stat}} - 2\log\mathcal{L}_{\mathrm{Sys}}.$$

$$\Delta\chi^2 = 2\sum_i^{\mathrm{Nbins}} N_i^{\mathrm{MC}}(\vec{f},\vec{x},\vec{d}) - N_i^{\mathrm{data}} + N_i^{\mathrm{data}} \ln\left(\frac{N_i^{\mathrm{data}}}{N_i^{\mathrm{MC}}(\vec{f},\vec{x},\vec{d})}\right)$$

$$\left.\begin{aligned} &+ \sum_i^{E_\nu bins} \sum_j^{E_\nu bins} \Delta\vec{f}_i \left(V_f^{-1}\right)_{i,j} \Delta\vec{f}_j \\ &+ \sum_i^{\mathrm{xsecpars}} \sum_j^{\mathrm{xsecpars}} \Delta\vec{x}_i \left(V_x^{-1}\right)_{i,j} \Delta\vec{x}_j \\ &+ \sum_i^{\mathrm{ND280det}} \sum_j^{\mathrm{ND280det}} \Delta d\vec{N}D_i \left(V_d^{-1}\right)_{i,j} \Delta d\vec{N}D_j \end{aligned}\right\} -2\log\mathcal{L}_{\mathrm{Sys}}$$

MC Statistic – BarlowBeeston

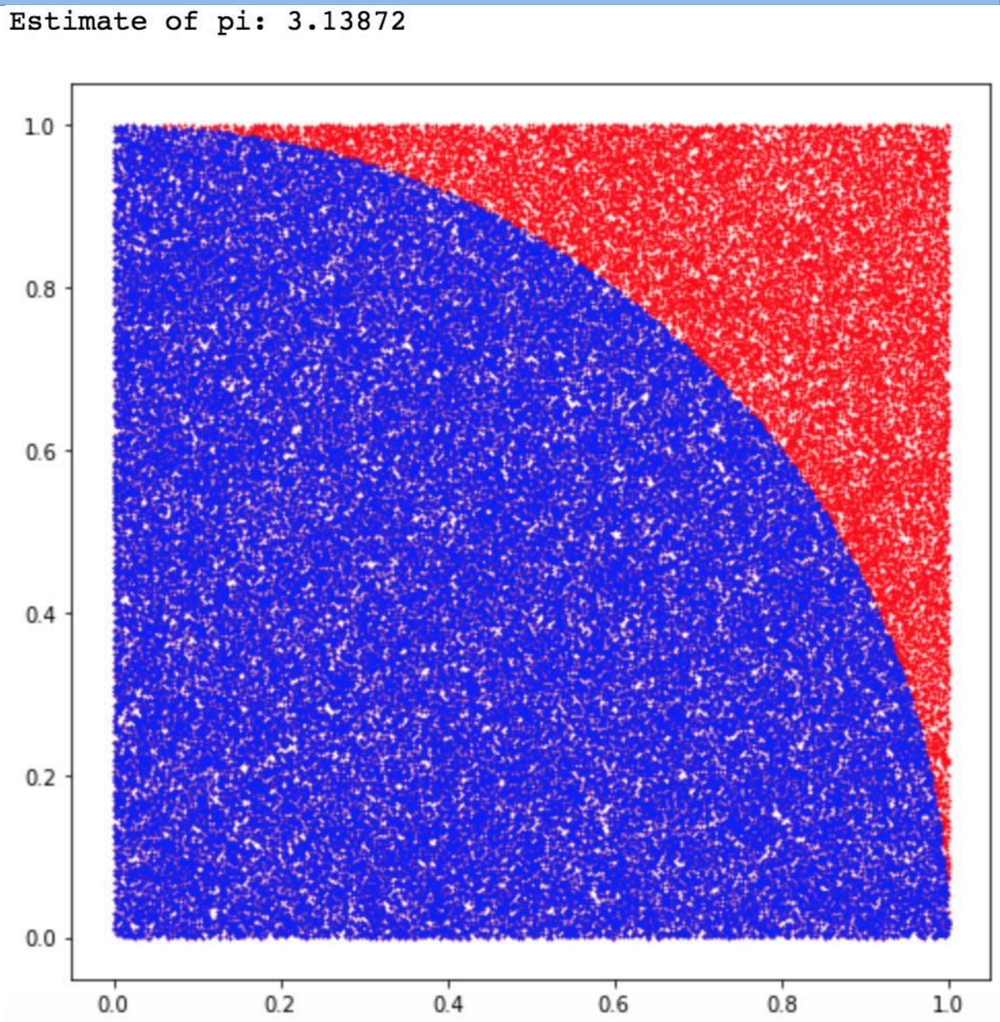
MC Statistic Uncertainty

Generation of MC is a stochastic process, so even identical settings can lead to different outputs. This fact introduces an uncertainty in the MC distributions.

In example if you don't have high MC stat you will wrongly calculate value of π.

This is one of the reason why we generate much more MC than collected data.

Run	Data (10^{19})	MC (10^{19})
2a FHC	3.60	168.02
2w FHC	4.34	120.38
3 FHC	15.93	307.80



Reweighting basics

MC generators include several models, each model is described by several parameters.
Very often each interaction has different set of parameters for example Quasi Elastic will have different description than Deep Inelastic Scattering.

Imagine you generate MC, you made mistake, and you produced it with twice larger xsec then it should be. What to do: ½ = New Xsec / Old Xsec.

You can assign weight ½ to each event.

N_Events = weight * N_{MC}

Imagine you made different mistake: you produced MC with twice larger QE xsec and two times lower for DIS

N_Event = $weigth_{QE} * N_{MC}^{QE} + weight_{DIS} * N_{MC}^{DIS}$ + …

Those were easy examples, but you can see that using reweighting we can modify MC by indirectly modifying xsec.

Splines and Reweighting

Imagine you can calculate xsec precisely. Although in general xsec depends on many parameters, let's focus on Axial Mass (MAQE) which can modify form factor for QE.

$$F_A(Q^2) = \frac{g_A}{\left(1 + \frac{Q^2}{M_A^{QE}}\right)^2}$$

When you change **MAQE** you change xsec, so you can calculate weight as a ratio of new xsec and xsec of produced MC.

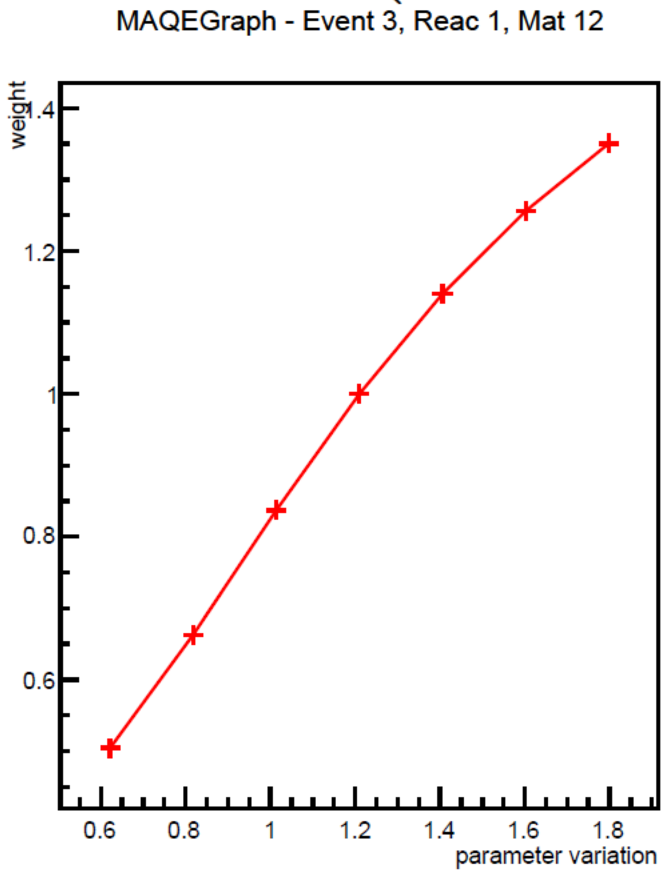
$$weight = \frac{New\ Xsec}{Old\ Xsec}$$

Then you can create spline which holds weights as values of **MAQE** variation. Making a spline allows to use those precalculated weights during the fit. We produce spline for each event.

Producing new MC can take days, weeks month. We can reweight full MC in less than 0.04 s !!!

Weights are multiplicative!!!

$$w_i = w_i^{POT} \cdot w_i^{flux} \cdot w_i^{beam} \cdot w_i^{NDdet} \cdot w_i^{NDcorr} \cdot w_i^{xsec}$$



Barlow-Beeston

To account for this we introduced MC statistic to the likelihood which account for this.

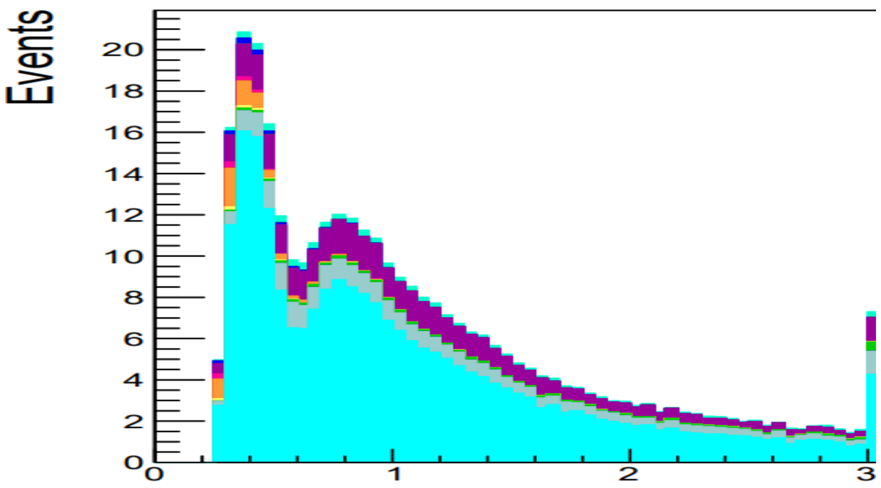
$$N_{\mathrm{MC}}^{true} = \beta \times N_{\mathrm{MC}}^{gen},$$

Beta parameters in scaling factor between true MC and one generated.

There is one beta per each fitted bin.

In the ideal situation beta = 1

In Barlow-Beeston implementation beta is treated as a fittable parameter. If your analysis has thousand of bins you might run into a problem.



Barlow-Beeston -> Conway

Using full Barlow-Beeston is problematic, thus we use Conway implementation of Barlow-Beeston based on 10.5170/CERN-2011-006.115

Barlow-Beeston Conway

$$-2\log\mathcal{L}_{\text{Stat}} = -2\log\mathcal{L}_{\text{Poisson}} - 2\log\mathcal{L}_{\text{MC}_{\text{stat}}} =$$

$$= 2\sum_i \left[\underbrace{N_i^{\text{MC}}(\vec{\theta}) - N_i^{\text{data}} + N_i^{\text{data}} \ln\left(\frac{N_i^{\text{data}}}{N_i^{\text{MC}}(\vec{\theta})}\right)}_{\text{Poisson}} + \frac{(\beta_i - 1)^2}{2\sigma_{\beta_i}^2} \right]$$

$$\sigma_{\beta_i}^2 = \sqrt{\sum w_i^2} / N_{MC,i'}^{gen}$$

w2 is as sum of weights in a particular bin.

$$w_i = w_i^{POT} \cdot w_i^{flux} \cdot w_i^{beam} \cdot w_i^{NDdet} \cdot w_i^{NDcorr} \cdot w_i^{xsec}$$

Beta Parameter

$$N_{\mathrm{MC}}^{true} = \beta \times N_{\mathrm{MC}}^{gen},$$

There are two ways how to calculate Betas.

By assuming that *Beta* has Gaussian distribution we can find it solution by solving this equation:

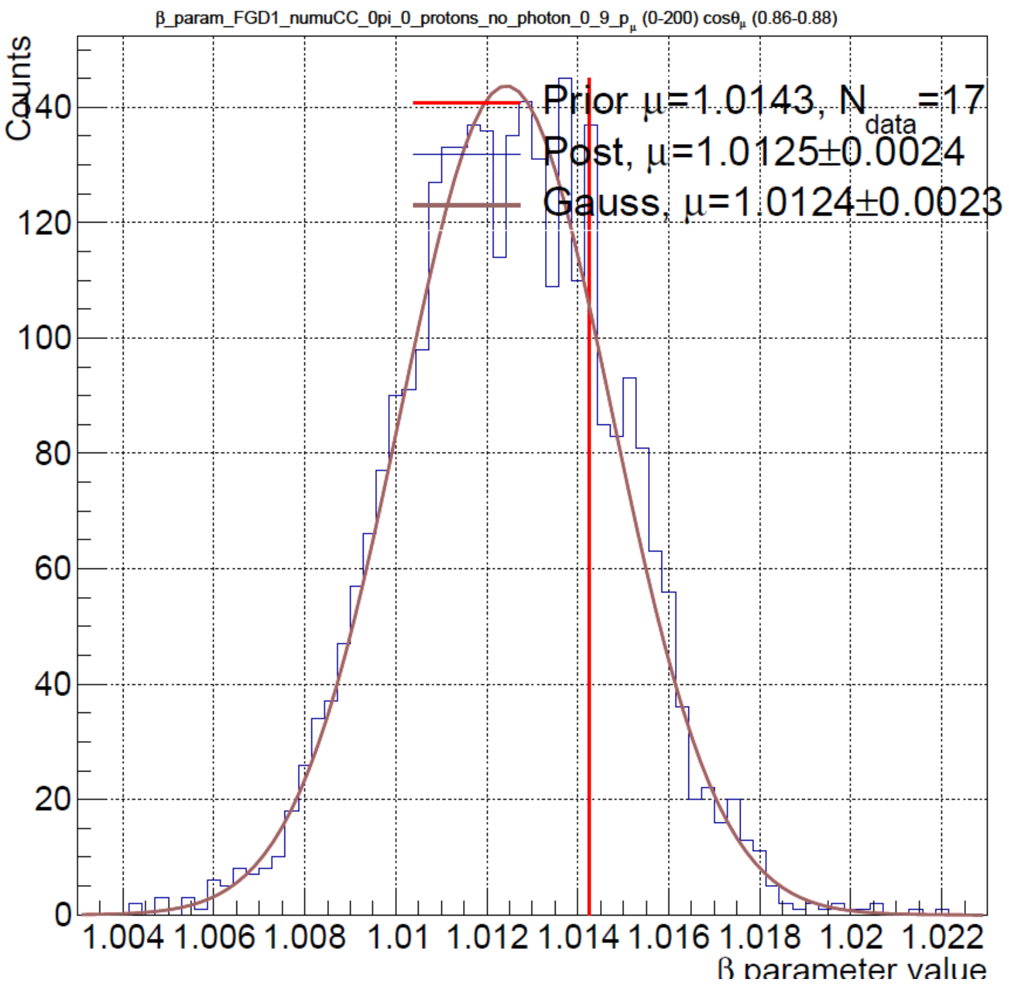
$$\beta_i^2 + (N_{MC,i}^{gen}\sigma_{\beta_i}^2 - 1)\beta_i - N_i^{\mathrm{data}}\sigma_{\beta_i}^2 = 0$$

$$\sigma_{\beta_i}^2 = \sqrt{\sum w_i^2} / N_{MC,i}^{gen},$$

$$w_i = w_i^{POT} \cdot w_i^{flux} \cdot w_i^{beam} \cdot w_i^{NDdet} \cdot w_i^{NDcorr} \cdot w_i^{xsec}$$

Beta Parameters

$$N_{\mathrm{MC}}^{true} = \beta \times N_{\mathrm{MC}}^{gen},$$



This is example of Beta parameter after the fit.

Majority of them have Gaussian distribution.

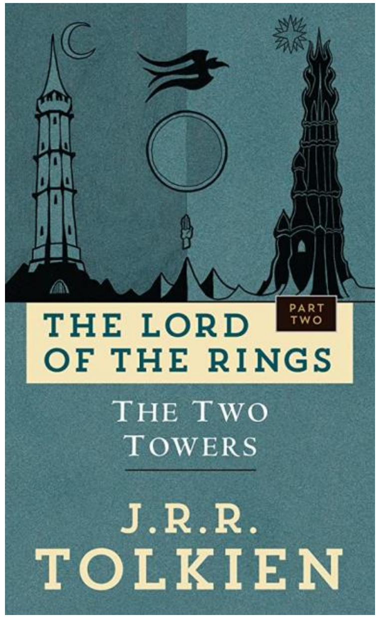
You can values are close to 1. Which is what we would expect.

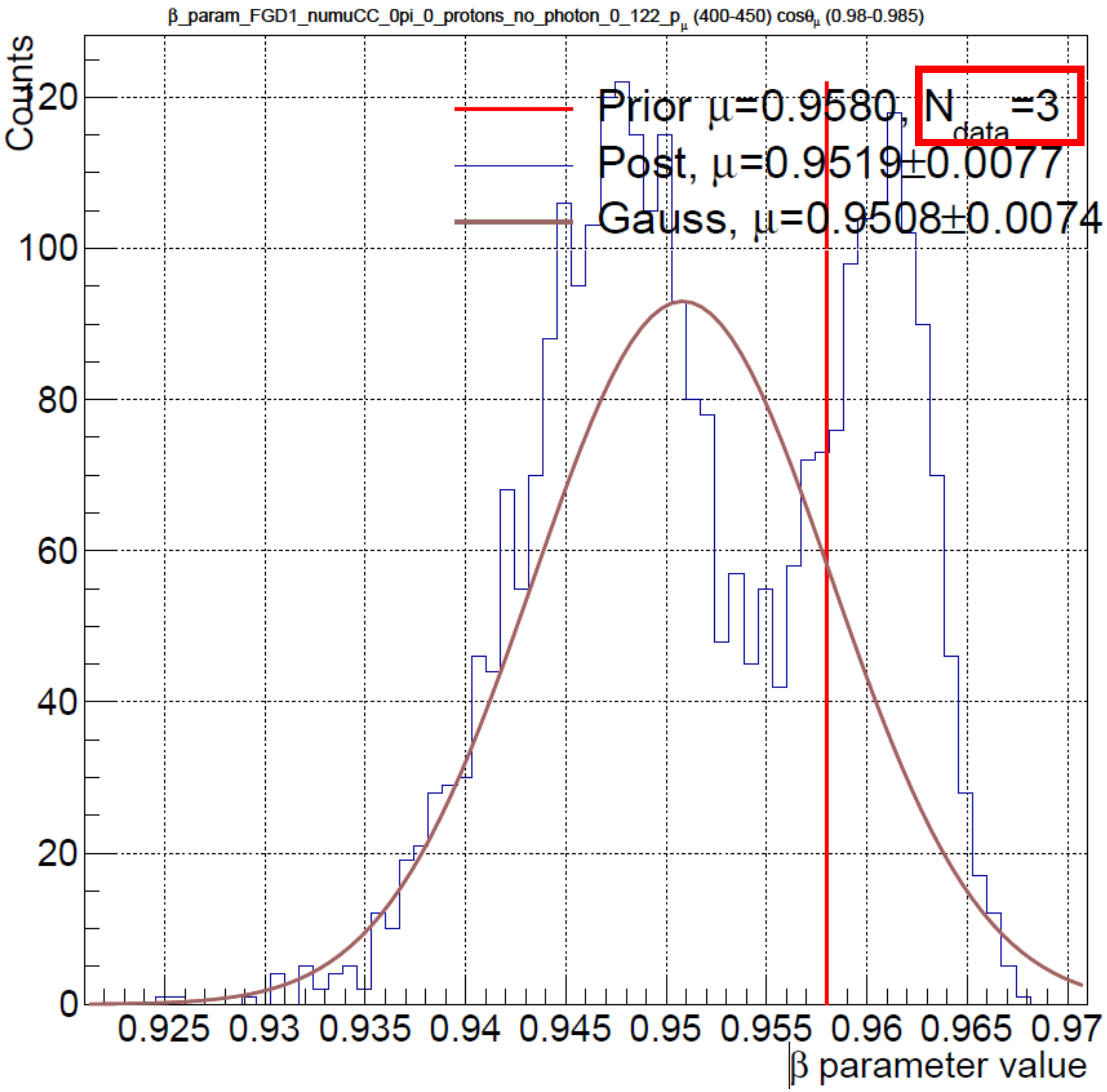
Non Gaussian Examples

$$N_{MC}^{true} = \beta \times N_{MC}^{gen},$$

Sometimes beta are non gaussian, although this happen when number of events is very low.

99% of beta are Gaussian.





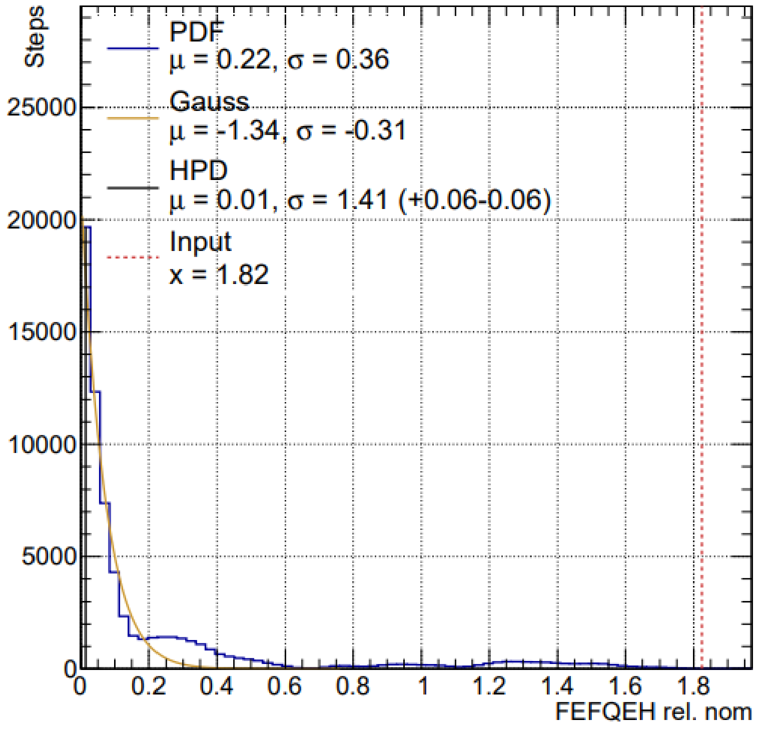
Problem with Conway

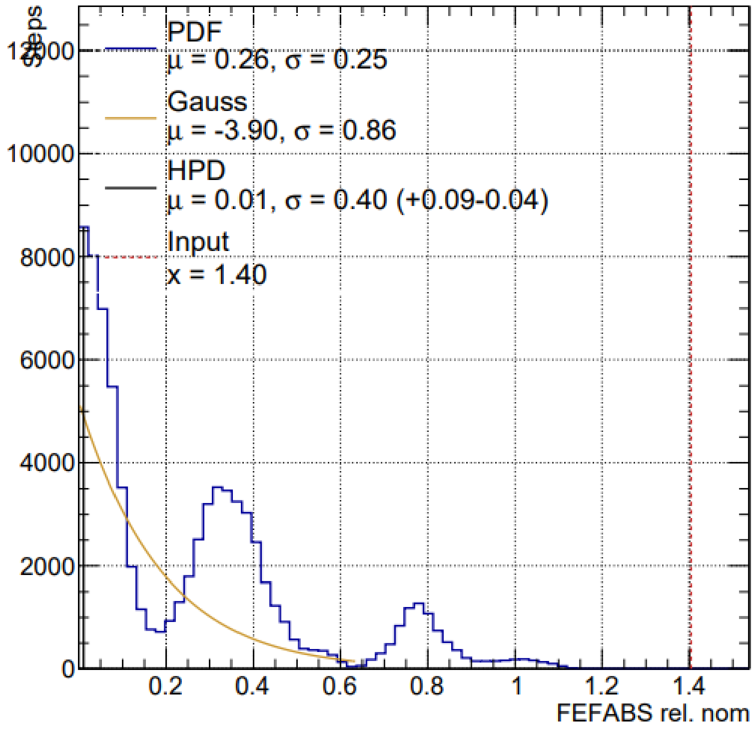
There is big problem with the Barlow-Beeston.

When running data fit we observed terrible problem where pion FSI went to 0

Posterior Distribution for a selected cross-section parameters.

Those describe probability of secondary interactions inside nucleus. We know they are not 0!





Solution

case of identical weights. Because both the first and second moments are matched, this approximation accounts for the variance of the CPD unlike $\mathcal{L}_{\mathrm{BB}}$, which only accounts for the mean. Thus, while $\mathcal{L}_{\mathrm{BB}}$ is valid only for the case of narrow weight distributions, our approximation remains valid for broader distributions.

arXiv:1901.04645v2

We know our weights can hugely differ, especially FSI!

We use weights to calculate sigma.

$$\sigma^2_{\beta_i} = \sqrt{\sum w_i^2} / N^{gen}_{MC,i},$$

Solution to this problem was simple. **Save values of w2 at prior values.**

Another Approach
Dembinski and Abdelmotteleb

Roasting Contest

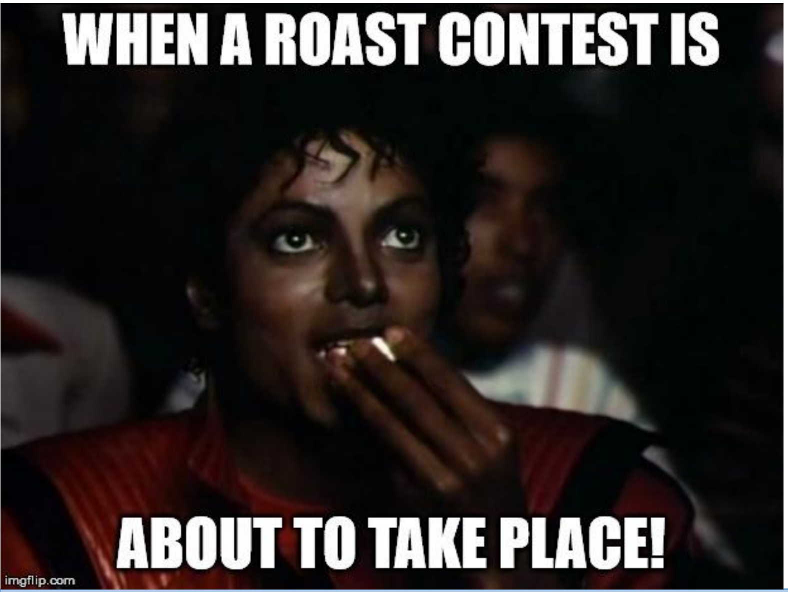
Dembinski and Abdelmotteleb suggest another implementation of Barlow-Beeston. arXiv:2206.12346v2

Authors are effectively bashing Conways through whole paper ;)

This formula for β is even simpler than the one derived by Conway, and can be easily interpreted. If $a \ll n$, β adjusts μ_0 to n, irrespective of the actual value of μ_0. The bin provides no informa-

Conway did not derive the simplified likelihood rigorously from the exact likelihood of Barlow and Beeston. Such a derivation is therefore attempted here, motivated by the wish to gain insight in which limit Conway's likelihood is an adequate proxy of the exact likelihood. This

metrically. Because of these properties, the new approximate likelihood is expected to perform better than Conway's in cases where the simulated sample is small.



Comparison

Conway

10.5170/CERN-2011-006.115

$$\beta_i^2 + (N_{\mathrm{MC},i}^{\mathrm{gen}}\sigma_{\beta_i}^2 - 1)\beta_i - N_i^{\mathrm{data}}\sigma_{\beta_i}^2 = 0$$

Dembinski and Abdelmotteleb

arXiv:2206.12346v2

$$k = \frac{MC_{nom}^2}{w^2} \qquad \beta = \frac{data + k}{MC_{nom} + k}$$

Value of Beta has much simpler expression

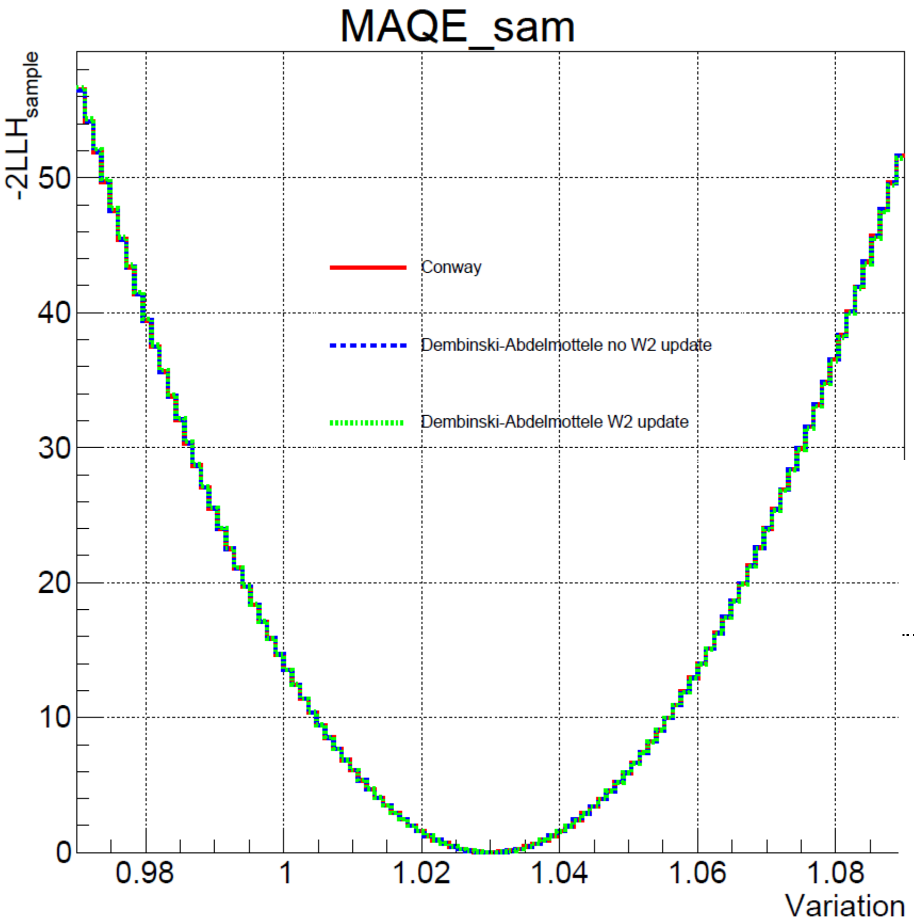
$$\frac{(\beta_i - 1)^2}{2\sigma_{\beta_i}^2}$$

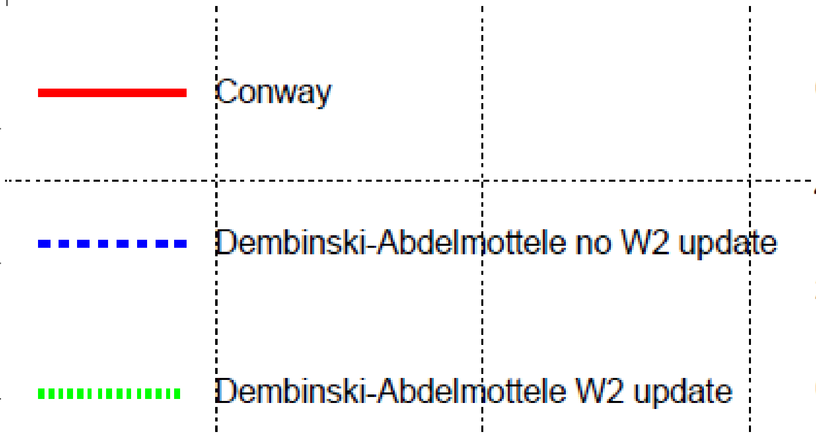
$$k * \beta - k - k * ln(\beta)$$

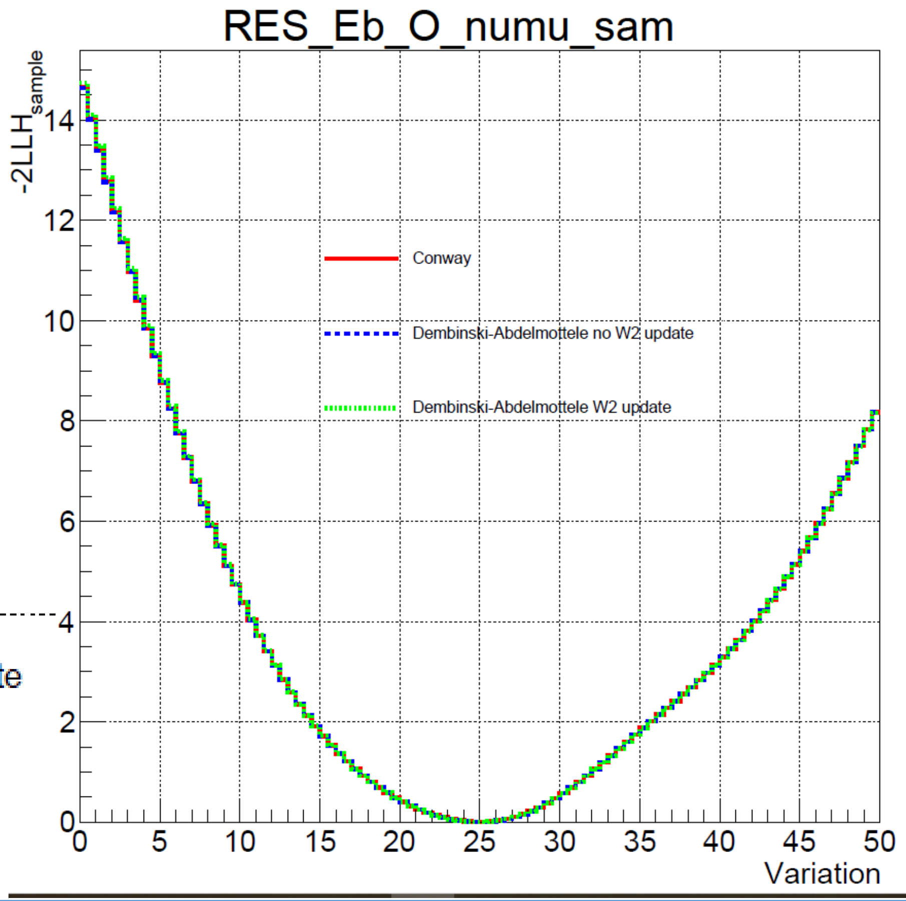
Penalty term strongly resembles Poisson making interpretation simpler

LLH scans

In spite of different equation
LLH scan are almost identical



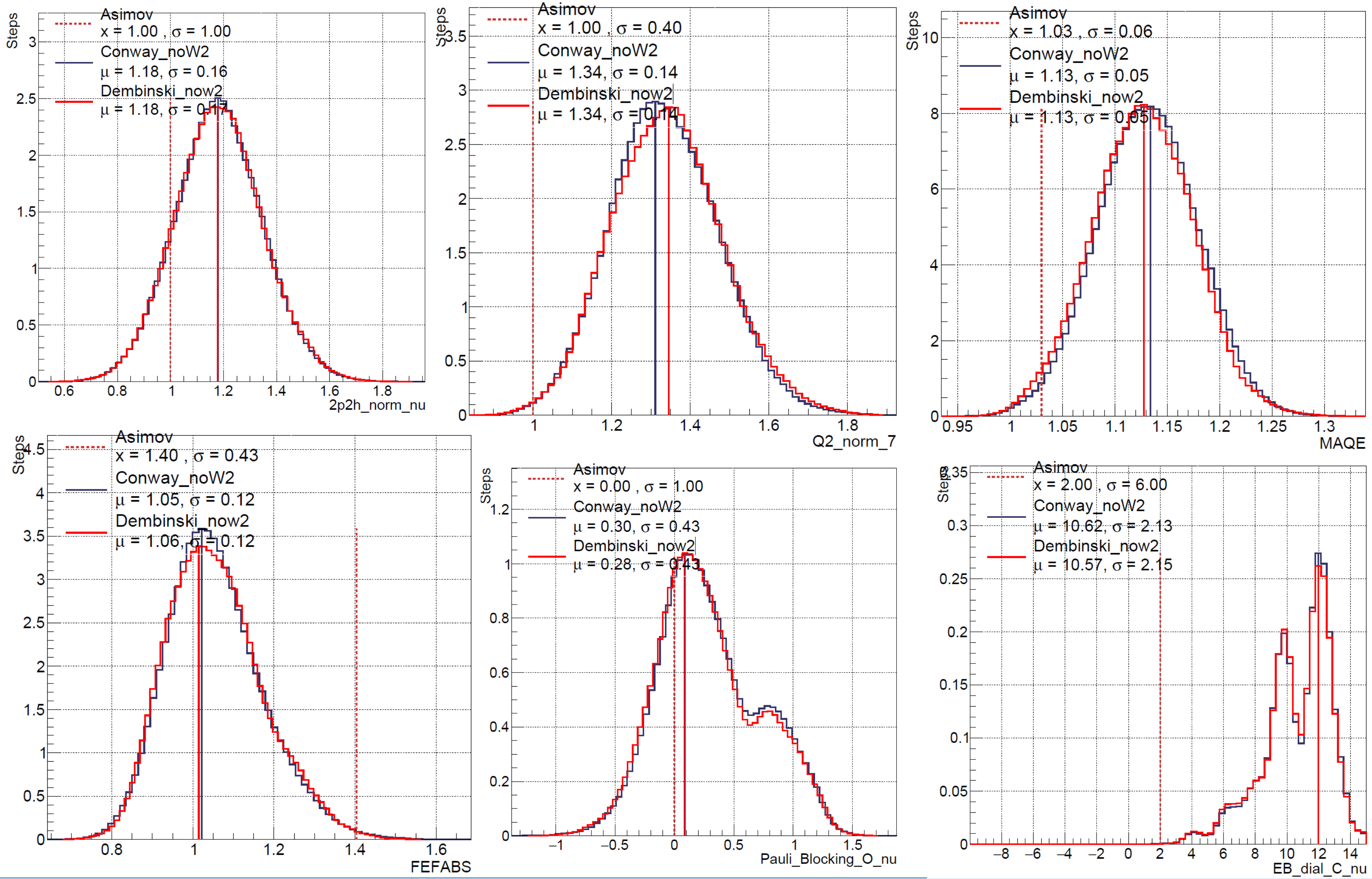




Conway vs Dembinski-Abdelmotteleb

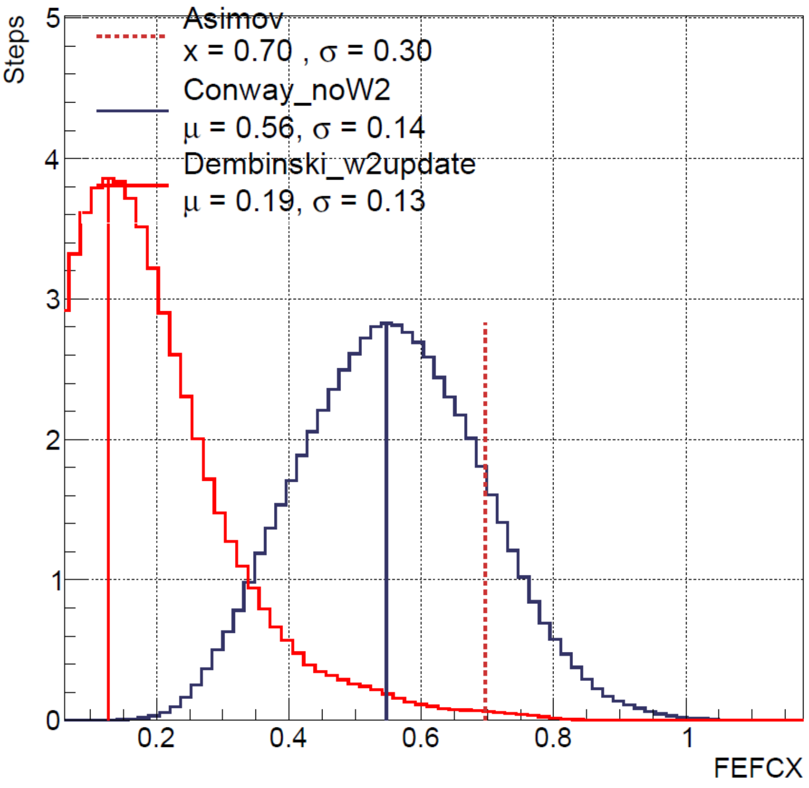
Plots presents posterior distribution for several systematic parameters.

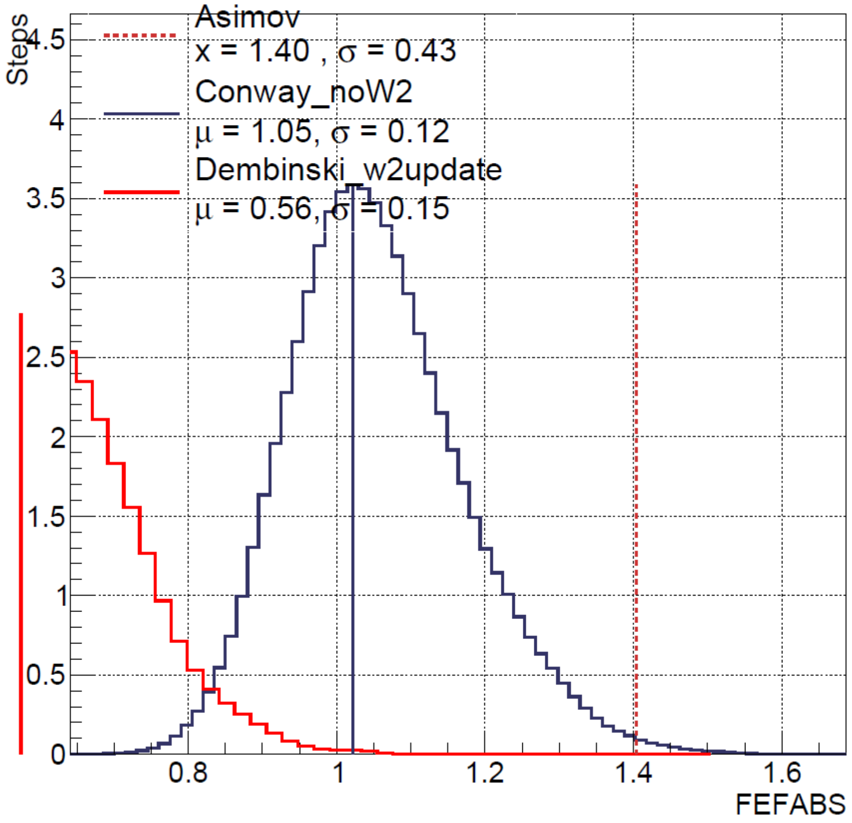
Results without updating w2 agree very well between both likelihoods.



Updating w2 ☹

When updating W2 it doesn't work as intented ☹





Ice Cube – The Holy Grail

Ice Cube

Another likelihood proposed by Ice Cube collaboration might solve all our problems. Full derivation can be found in **arXiv:1901.04645v2**

Assumptions:

- Marginalize over nuisance paramreters rather then fit them directly. This is done to not assume perfect knowledge of the nuisance parameters from a finite number of realizations.

$$\mathcal{P}\left(\lambda|\vec{w}(\vec{\theta})\right) = \frac{\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))\mathcal{P}(\lambda)}{\int_0^\infty \mathcal{L}(\lambda'|\vec{w}(\vec{\theta}))\mathcal{P}(\lambda')\, d\lambda'},$$

- Weights are sampled from a Compound Poisson Distribution (CPD), which can be approximated by a Scaled Poisson Distribution (SPD). They calculated SPD, under the maximum likelihood solution for the given MC realization, has **first** and **second** moment meaning weight variance is included.

$$\mathcal{L}(\lambda|\vec{w}(\vec{\theta})) = \mathcal{L}(\lambda|\mu,\sigma) = \frac{e^{-\lambda\mu/\sigma^2}\left(\lambda\mu/\sigma^2\right)^{\mu^2/\sigma^2}}{\Gamma(\mu^2/\sigma^2+1)},$$

Ice Cube

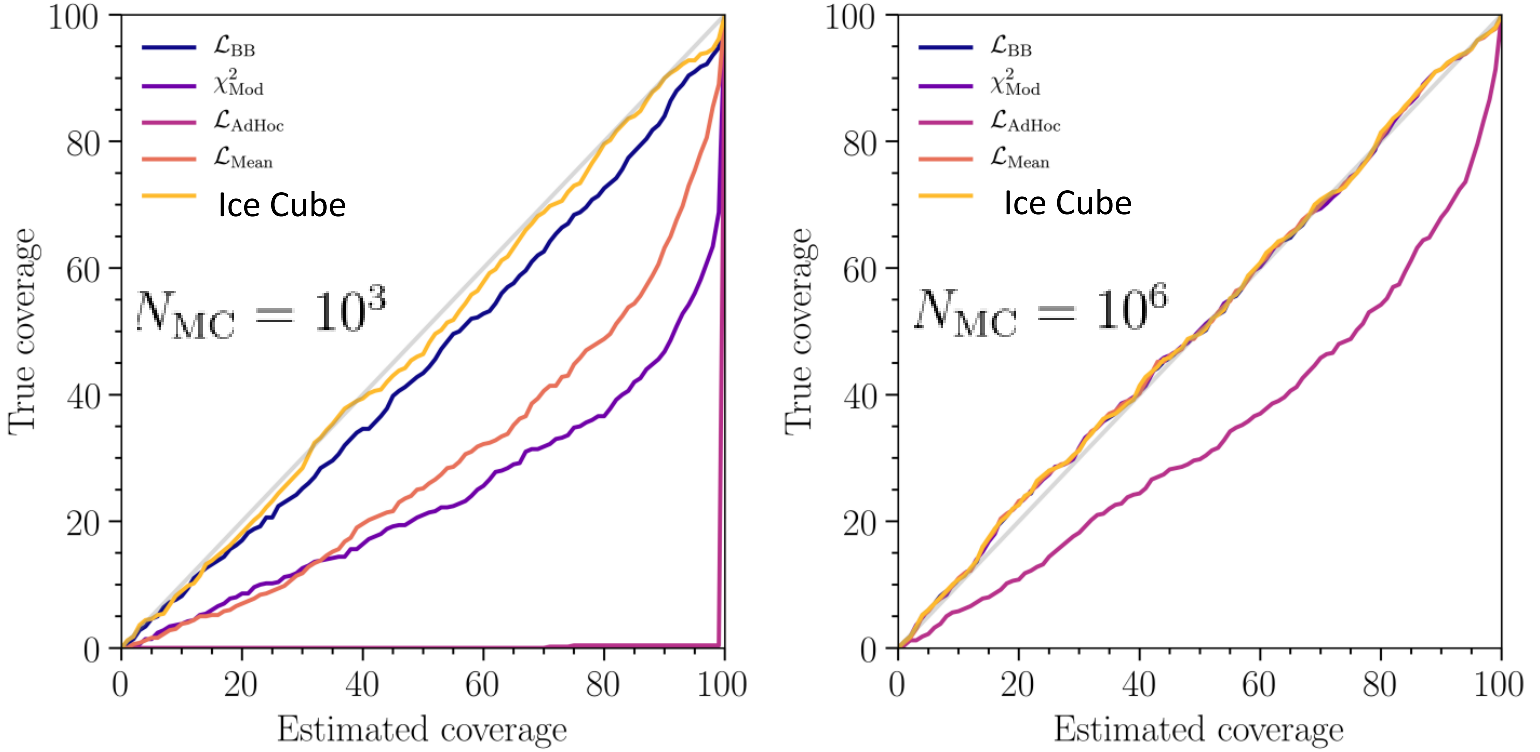
$$-2log\mathcal{L} = -2\sum_i \left(a_i \log(b_i) + \log[\Gamma(N_i^{\rm data} + a)] - (N_i^{\rm data} + a)\log(b_i + 1) - \log[\Gamma(a_i)]\right) \tag{A.1}$$

where the auxiliary variables $a_i = N^{\rm gen}_{{\rm MC},i} b_i + 1$ and $b_i = N^{\rm gen}_{{\rm MC},i} / \sum w_i^2$. An example of LLH

Advantages

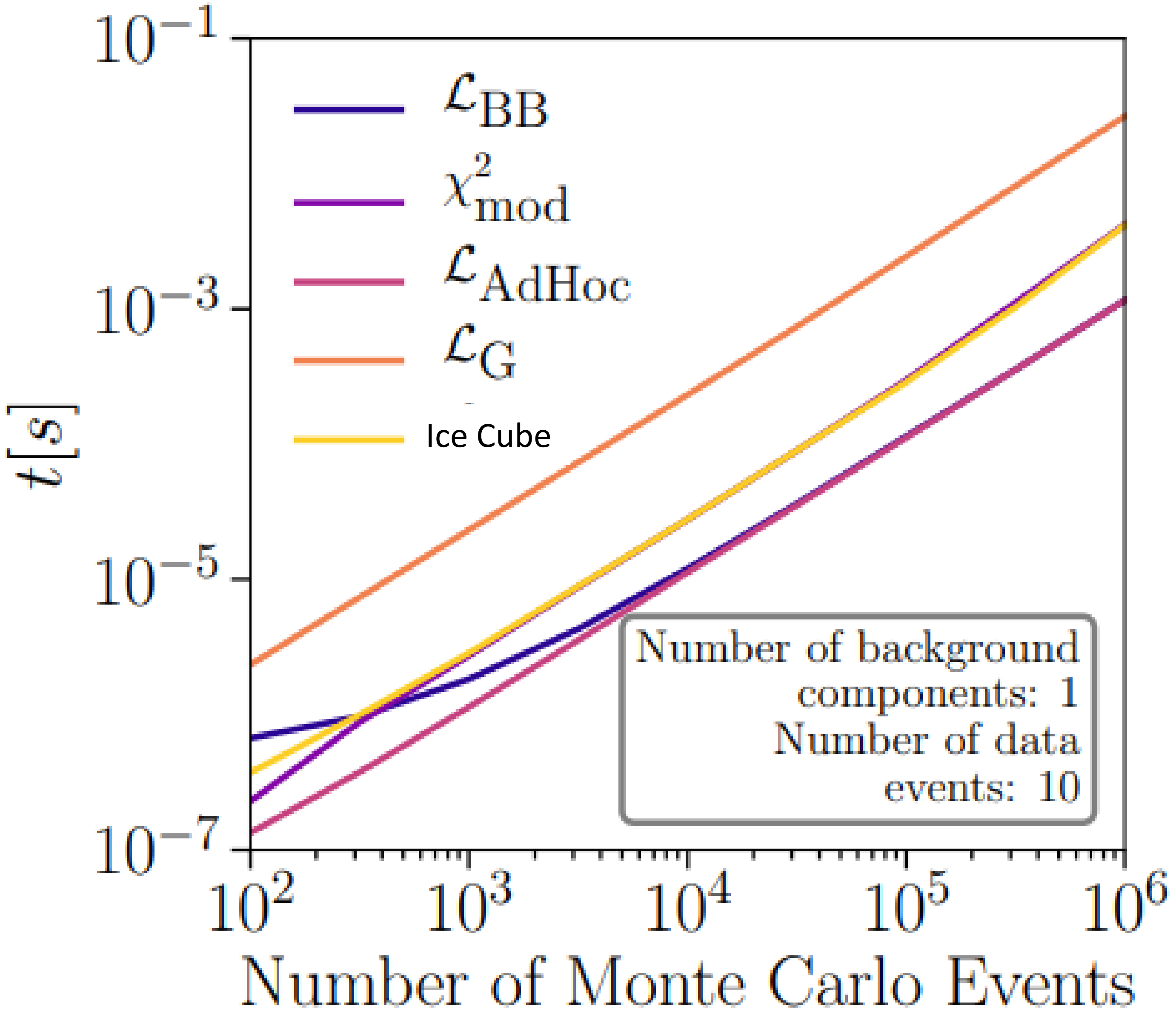
- Include weight variance
- Works very well in low statistic environment

$$\Gamma(n) = (n-1)!\,.$$



Performance

IceCube is minimally slower than Barlow-Beeston mostly due to gamma functions.



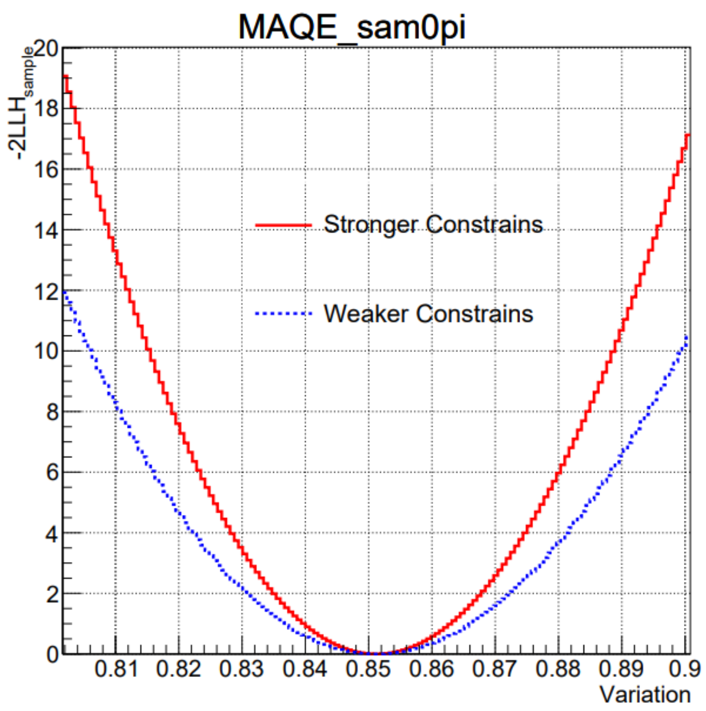
Ice Cube

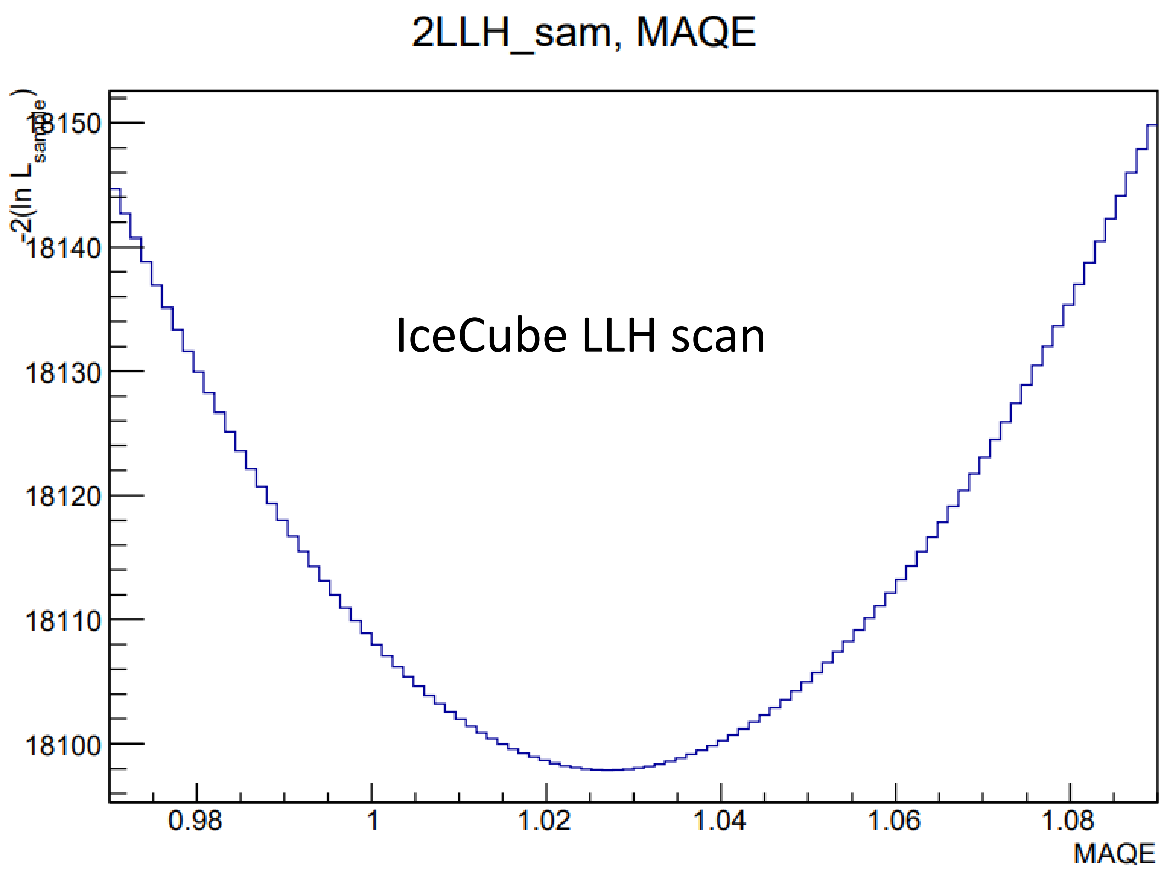
For technical reasons we don't normalize Ice Cube, this doesn't impact the analysis as we use delta chi-square

$$-2log\mathcal{L} = -2\sum_i \left(a_i \log(b_i) + \log[\Gamma(N_i^{\text{data}} + a)] - (N_i^{\text{data}} + a)\log(b_i + 1) - \log[\Gamma(a_i)]\right) \quad \text{(A.1)}$$

where the auxiliary variables $a_i = N_{\text{MC},i}^{\text{gen}} b_i + 1$ and $b_i = N_{\text{MC},i}^{\text{gen}} / \sum w_i^2$. An example of LLH

Barlow-Beeston LLH scan





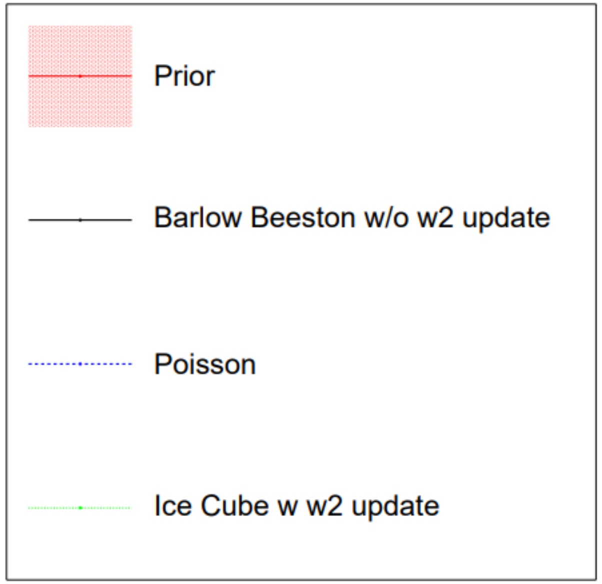
Fit Results

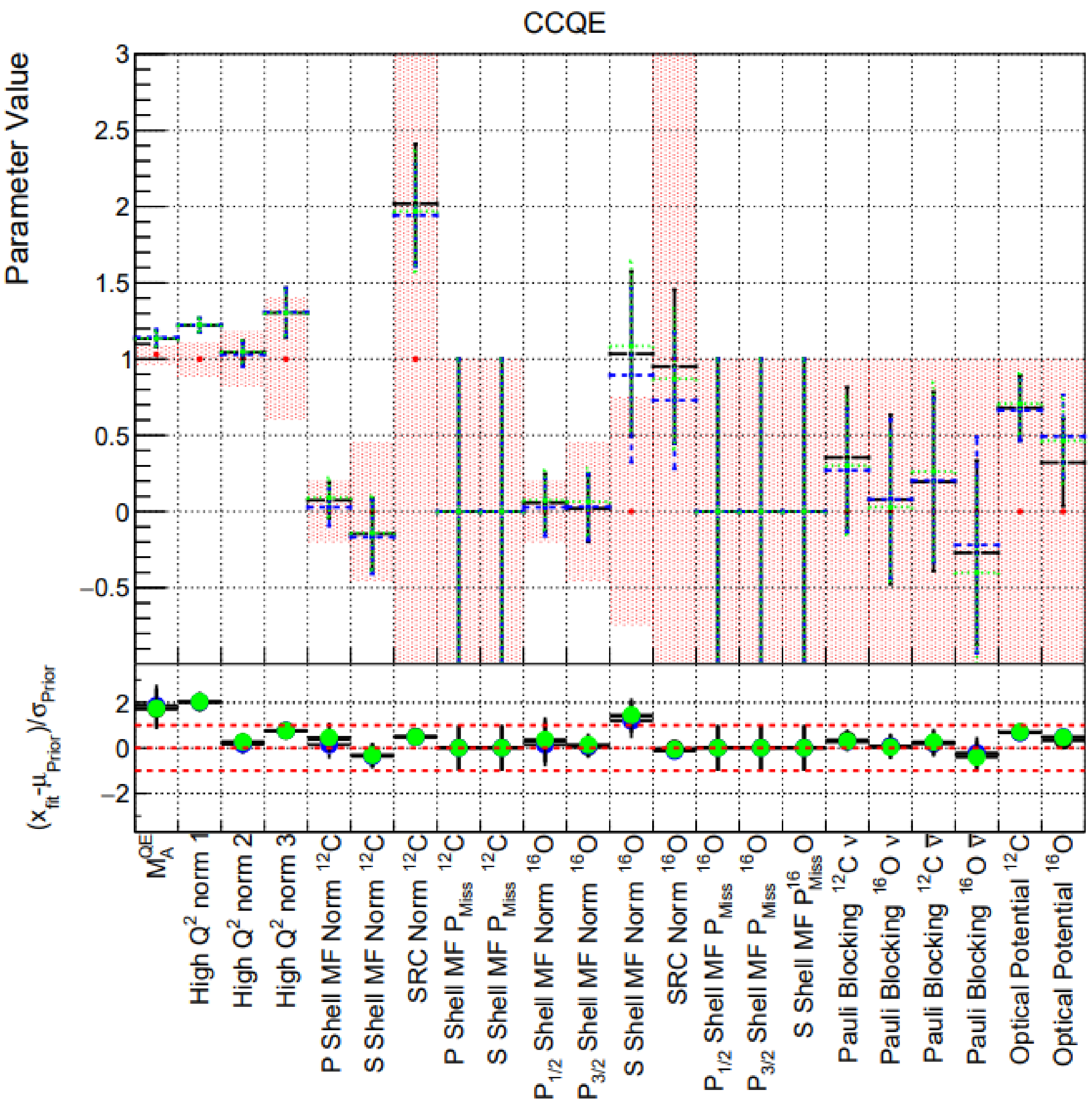
Best-fit value with error after analysis obtained using different likelihoods.

Results are almost identical.

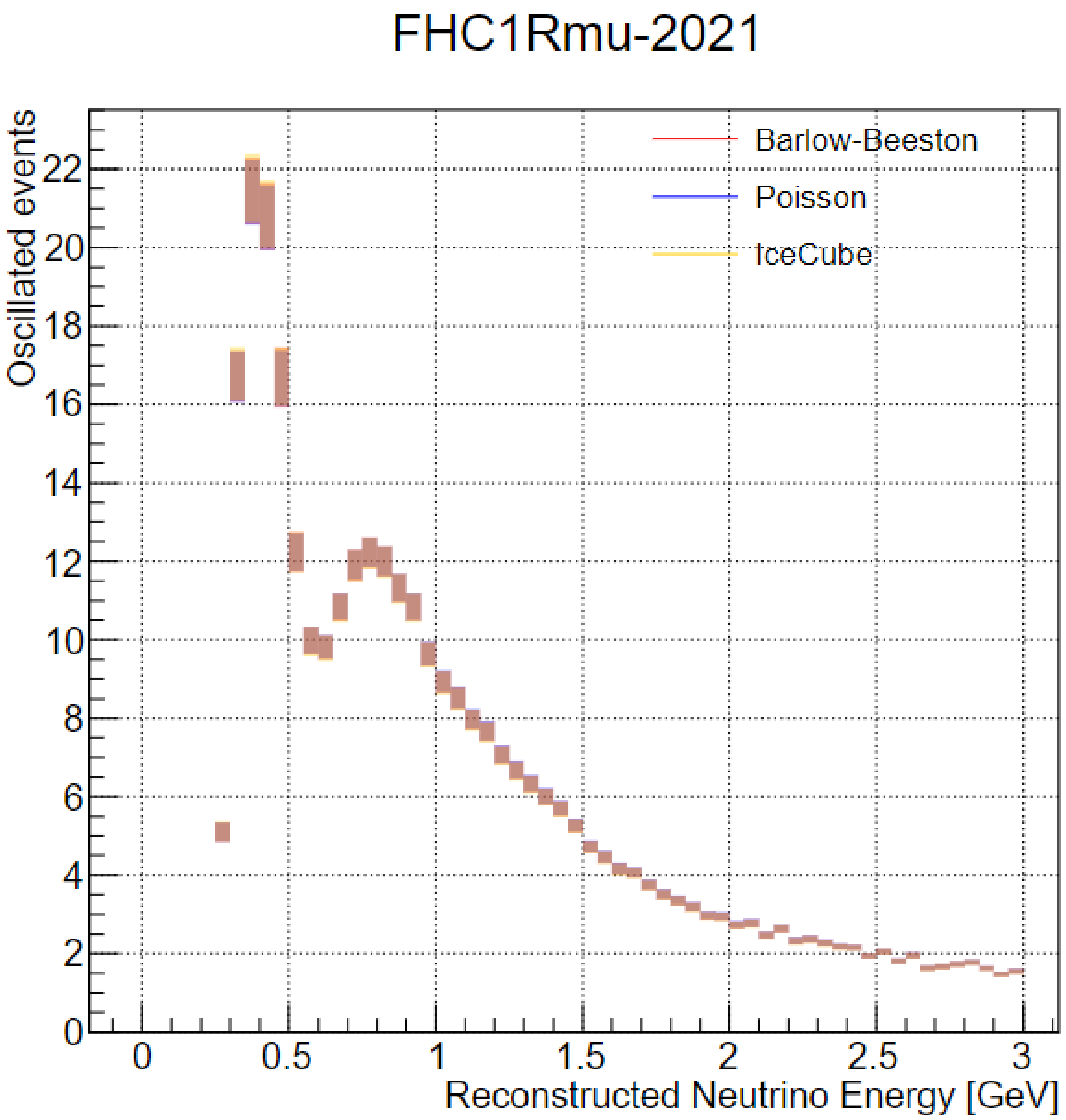
Ice Cube was updating weights!!!

At least in T2K this choice doesn't Impact results.





Spectra



Poisson

$$2\sum_i \left[N_i^{MC}(\vec{\theta}) - N_i^{data} + N_i^{data} \ln\left(\frac{N_i^{data}}{N_i^{MC}(\vec{\theta})} \right) \right]$$

Barlow-Beeston (Conway)

$$2\sum_i \left[N_i^{MC}(\vec{\theta}) - N_i^{data} + N_i^{data} \ln\left(\frac{N_i^{data}}{N_i^{MC}(\vec{\theta})} \right) + \frac{(\beta_i - 1)^2}{2\sigma_{\beta_i}^2} \right]$$

Ice-Cube

$$-2\sum_i \left(a_i \log(b_i) + \log[\Gamma(N_i^{data} + a)] - (N_i^{data} + a)\log(b_i + 1) - \log[\Gamma(a_i)] \right)$$

Predictions obtained after running a fit with different likelihoods.

Bonus!
Pearson – Test Meant to Fail

Pearson LLH

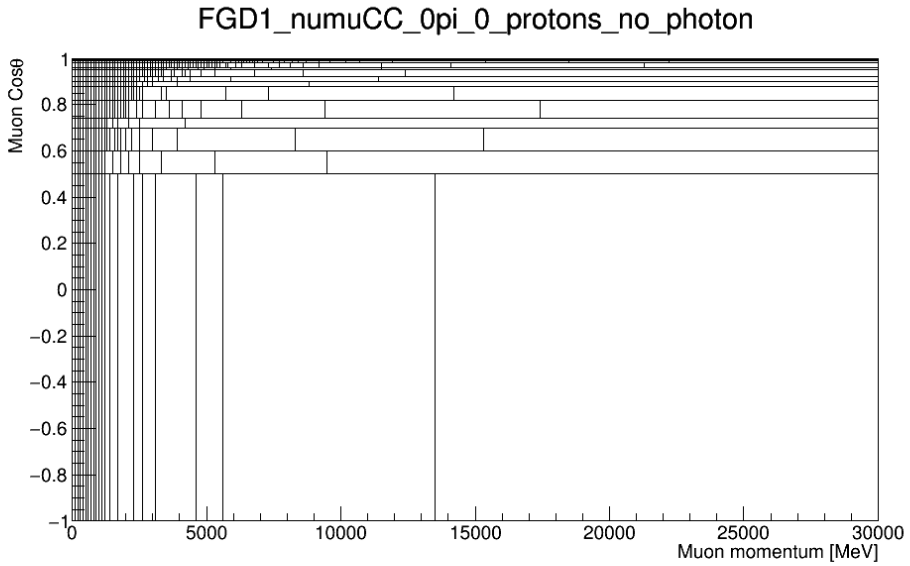
2.2 Uncertainties in the large-sample limit

In the large-sample regime, the Gaussian distribution is an appropriate description of the observed data. In this limit, the use of Pearson's χ^2 as a test-statistic [9] is common practice. For a single analysis bin, Pearson's χ^2 is defined as

$$\chi^2(\vec{\theta}) = \frac{(k - \lambda(\vec{\theta}))^2}{\lambda(\vec{\theta})}, \quad (2.6)$$

arXiv:1901.04645v2

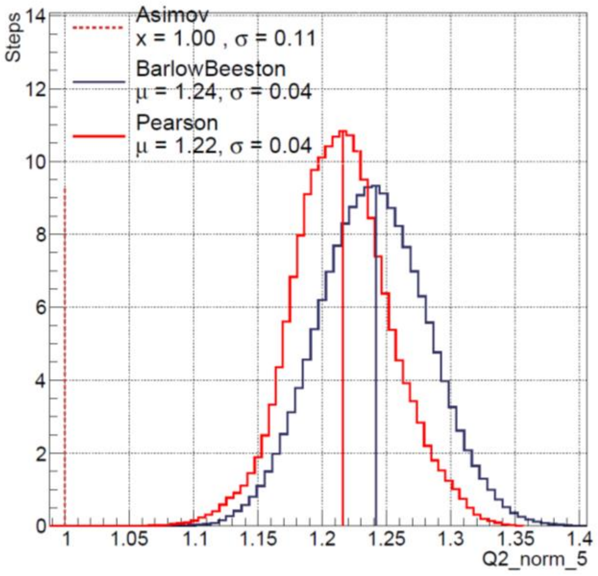
In my analysis we have bins with a very few number of events. Hence, we expect that Gaussian assumption will not work for every bin.

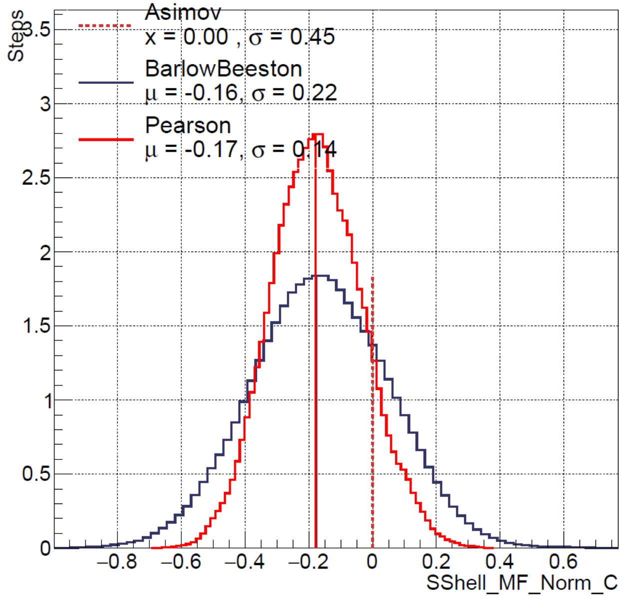


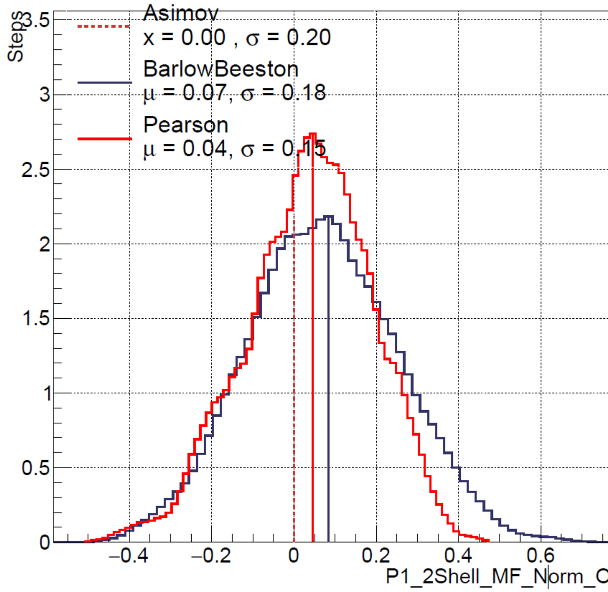
ND280 fit results

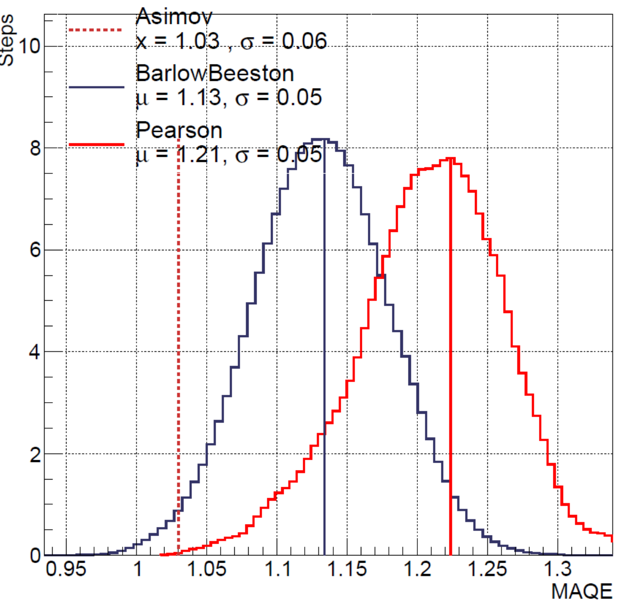
Results are for many cases close.

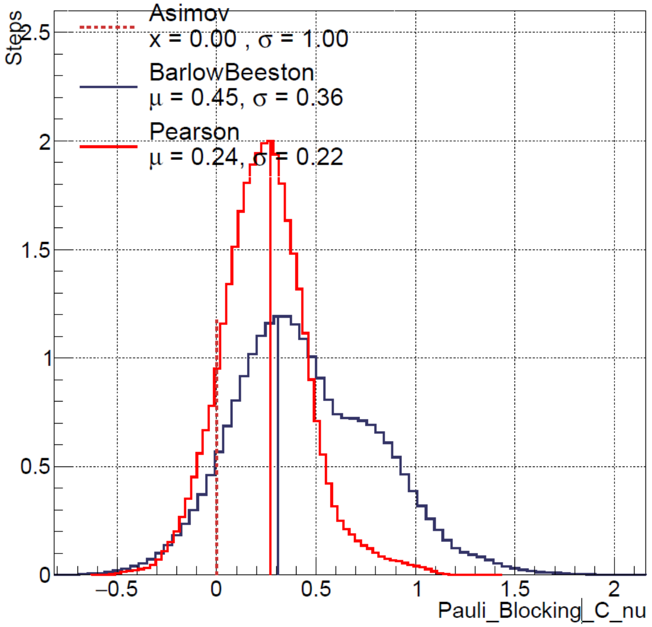
Sometimes are very much wrong.

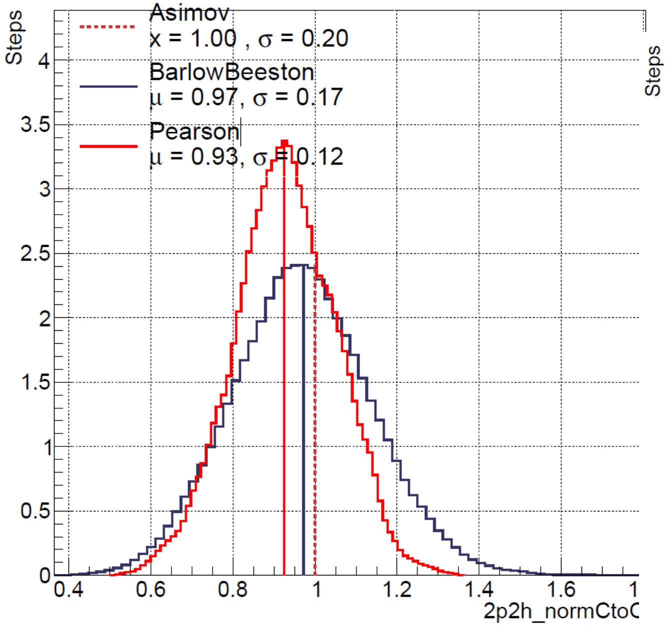


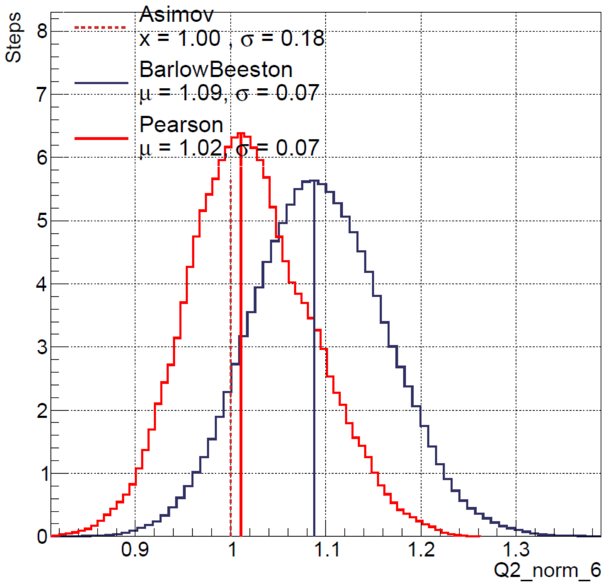












Ice Cube Summary

For counting experiments, we use Poisson likelihood to estimate data/MC agreement.

There are several ways to account for MC statistic uncertainty.

Many implementation has problem when we start updating weights.

Ice Cube confirms that using BB w/o w2 update didn’t bias our results.

Ice Cube is superior, might be slightly slower but no a big problem.

Choice of likelihood matters as demonstrated by Pearson.

$$-2log\mathcal{L} = -\Big(a\log(b) + \log[\Gamma(N^{\text{data}}+a)] - (N^{\text{data}}+a)\log(b+1) - \log[\Gamma(a)]\Big)$$

Backup

Abstract

Many analyses in particle physics are trying to determine for which set of systematic parameters MC predictions are in the best agreement with the collected data. To describe this agreement we use the likelihood function. There are several likelihoods suggested by statisticians each with different assumptions. Another important issue is the treatment of MC statistical uncertainty, which can be incorporated into the likelihood. The seminar will discuss the impact of several likelihood functions, like Conway's or Dembinski-Abdelmotteleb in T2K near detector analysis.

Perofmnce

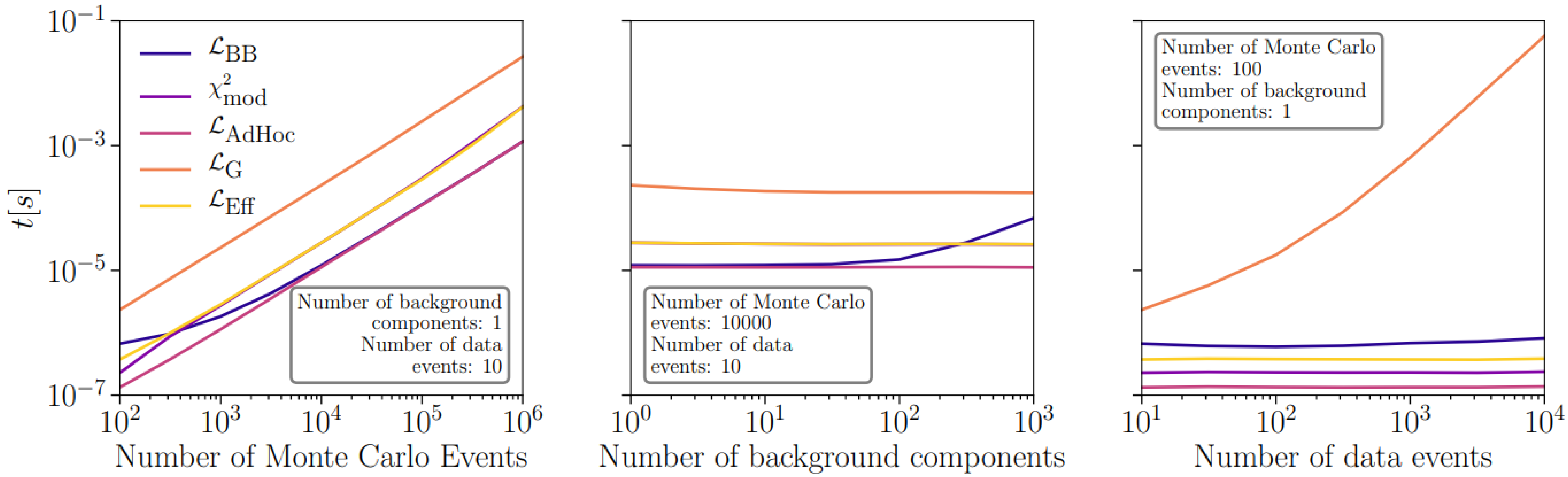


Figure 7. ***Likelihood function performance.*** Average single likelihood evaluation time is shown in the vertical axis in seconds. Different line colors show different likelihoods. Leftmost panel: the number of MC events used is shown on the horizontal axis. Center panel: the number of background components is shown on the horizontal axis. Rightmost panel: the number of data events is show on the horizontal axis.

LLH

Parameters	$\mu \equiv \sum_{i=1}^{m} w_i, \ \sigma^2 \equiv \sum_{i=1}^{m} w_i^2$
$\mathcal{L}_{\text{AdHoc}}$	$\frac{\mu^k e^{-\mu}}{k!}$
χ^2_{mod}	$\frac{(k-\mu)^2}{\mu+\sigma^2}$
$\mathcal{L}_{\text{BB}}^{s=1}$	$\max_{\bar{m}} \left\{ \frac{1}{k!m!} \left(\frac{\mu\bar{m}}{m}\right)^k \bar{m}^m e^{-\frac{\mu\bar{m}}{m}-\bar{m}} \right\}$
$\mathcal{L}_{\text{Mean}}$	$\left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}} \Gamma\left(k+\frac{\mu^2}{\sigma^2}\right) \left[k! \left(1+\frac{\mu}{\sigma^2}\right)^{k+\frac{\mu^2}{\sigma^2}} \Gamma\left(\frac{\mu^2}{\sigma^2}\right)\right]^{-1}$
$\mathcal{L}_{\text{Eff}}$	$\left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}+1} \Gamma\left(k+\frac{\mu^2}{\sigma^2}+1\right) \left[k! \left(1+\frac{\mu}{\sigma^2}\right)^{k+\frac{\mu^2}{\sigma^2}+1} \Gamma\left(\frac{\mu^2}{\sigma^2}+1\right)\right]^{-1}$

SUPPL. TABLE 1. ***Table of likelihood formulas.*** The likelihood functions discussed in this paper are given in each row. They are written in terms of μ and σ, whose explicit formulas are given in the top row, and the number of observed events, k, in the bin. In the case of $\mathcal{L}_{\text{BB}}$ we write the likelihood for the single-process case. Our main result and recommended likelihood, $\mathcal{L}_{\text{Eff}}$, is given in the last row.

Full LLH

$$
\begin{aligned}
-\ln \mathcal{L} = & \sum_{i}^{\text{ND280bins}} N_i^{\text{ND,MC}}(\vec{f},\vec{x},d\vec{ND}) - N_i^{\text{ND,data}} + N_i^{\text{ND,data}} \ln\left(\frac{N_i^{\text{ND,data}}}{N_i^{\text{ND,MC}}(\vec{f},\vec{x},d\vec{ND})}\right) + \frac{(\beta_i - 1)^2}{2\sigma_{\beta_i}^2} \\
& + \sum_{i}^{\text{SKbins}} N_i^{\text{SK,MC}}(\vec{f},\vec{x},d\vec{SK}) - N_i^{\text{SK,data}} + N_i^{\text{SK,data}} \ln\left(\frac{N_i^{\text{SK,data}}}{N_i^{\text{SK,MC}}(\vec{f},\vec{x},d\vec{SK})}\right) \\
& + \frac{1}{2}\sum_{i}^{E_\nu bins}\sum_{j}^{E_\nu bins} \Delta\vec{f}_i \left(V_f^{-1}\right)_{i,j} \Delta\vec{f}_j \\
& + \frac{1}{2}\sum_{i}^{\text{xsecpars}}\sum_{j}^{\text{xsecpars}} \Delta\vec{x}_i \left(V_x^{-1}\right)_{i,j} \Delta\vec{x}_j \\
& + \frac{1}{2}\sum_{i}^{\text{ND280det}}\sum_{j}^{\text{ND280det}} \Delta d\vec{ND}_i \left(V_{d^{ND}}^{-1}\right)_{i,j} \Delta d\vec{ND}_j \\
& + \frac{1}{2}\sum_{i}^{\text{SKdet}}\sum_{j}^{\text{SKdet}} \Delta d\vec{SK}_i \left(V_{d^{SK}}^{-1}\right)_{i,j} \Delta d\vec{SK}_j \\
& + \frac{1}{2}\sum_{i}^{\text{oscpars}}\sum_{j}^{\text{oscpars}} \Delta\vec{o}_i \left(V_o^{-1}\right)_{i,j} \Delta\vec{o}_j
\end{aligned}
$$

FD specific

Only ND280 uses Barlow-Beeston.

Since SK has much lower statistic that Impact of MC statistic should be negligible.

Problem with Results

In LLH calculation we include beta which work is a normalization of each ND280 bin.

However, this is done only when calculating LLH, post fit spectra are without this correction.

I talked with Clarence and he agree this should be fixed.

This doesn't affect results only ND280 best fit spectra plots.

Since beta are close to 1 it is not a big deal

$$N_{\mathrm{MC}}^{true} = \beta \times N_{\mathrm{MC}}^{gen},$$

How LLH is calculated

```
// Solve for the positive beta
double beta = (-1*temp+sqrt(temp2))/2.;
newmc = mc*beta;
```