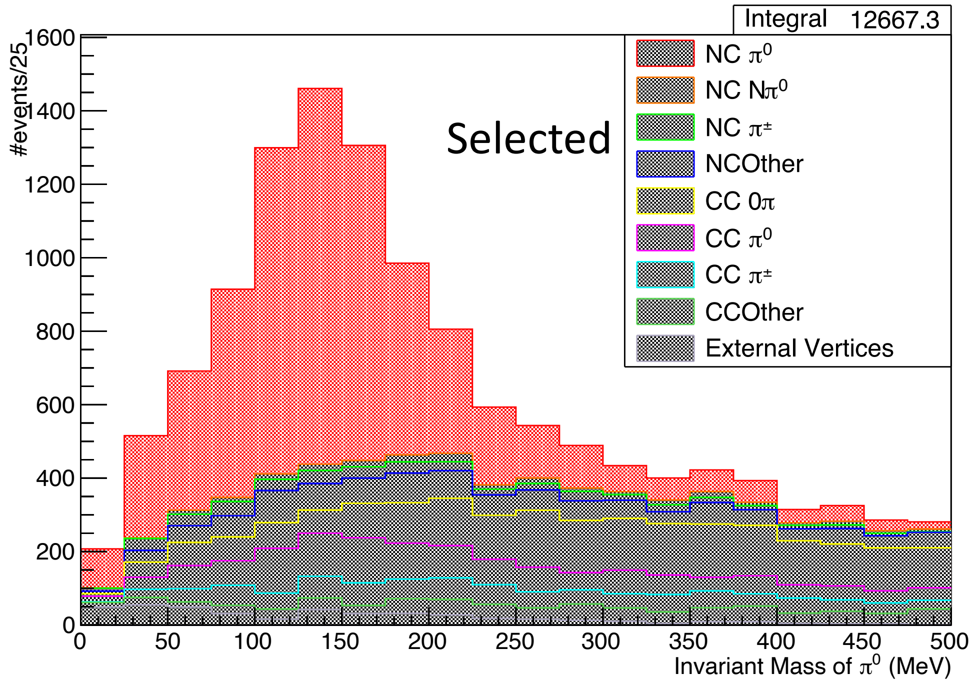
Extended Binned Likelihood including Monte Carlo Statistical Uncertainty in Bayesian Inference

NN Group Meeting

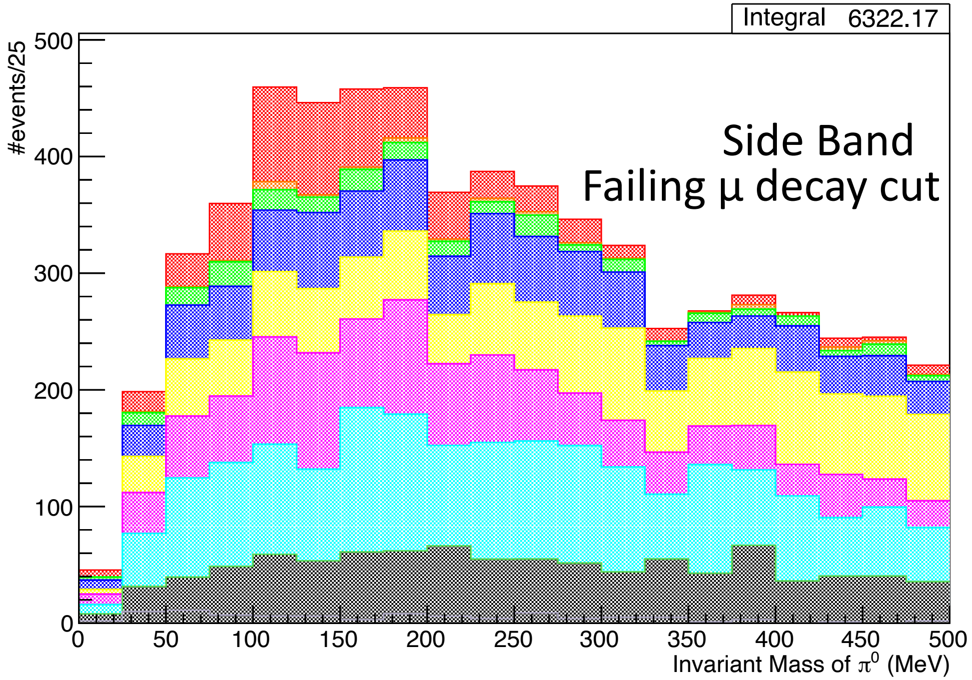
03/02/2023

Shilin Liu

Selected Sample and Side Band Sample



P6T P0D water-in MC Events



P6T P0D water-in MC Events

NC $1\pi^0$	**46.0% (5814 events)**
NC $N\pi^0$	**1.8% (230 events)**
NC $\pi^{+/-}$	**3.2% (409 events)**
NC Other	**8.0%(1026 events)**
CC 0π	**16.4% (2084 events)**
CC π^0	**9.8% (1252 events)**
CC $\pi^{+/-}$	**6.4% (817 events)**
CC Other	**4.9% (627 events)**
External to P0D	**3.4% (427 events)**

NC $1\pi^0$	**8.4% (529 events)**
NC $N\pi^0$	**0.6% (41 events)**
NC $\pi^{+/-}$	**3.4% (218 events)**
NC Other	**13.7%(867 events)**
CC 0π	**17.9% (1131 events)**
CC π^0	**16.3% (1030 events)**
CC $\pi^{+/-}$	**24.7% (1561 events)**
CC Other	**13.6% (859 events)**
External to P0D	**1.5% (93 events)**

Bayes' Theorem and Model Parameters

- The data we observe (denoted by x) can constrain our model parameters (denoted by θ) by Bayes' Theorem:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

Our prior knowledge on the model parameters (e.g. from previous measurement, other literature)

Given the data we measured, the probability density distribution of our model parameters

Constant, from law of total probability

$$P(x) = \int P(x|\theta)\,P(\theta)d\theta$$

Bayes' Theorem and Extended Binned Likelihood

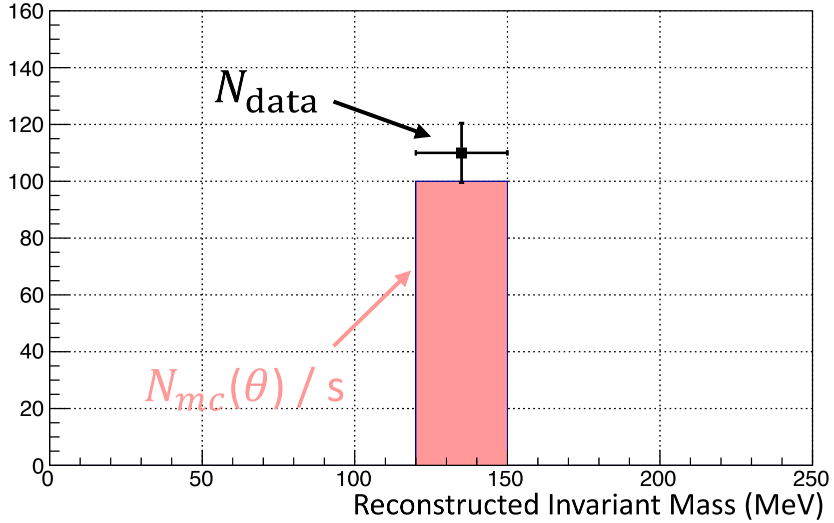
- The data we observe (denoted by x) can constrain our model parameters (denoted by θ) by Bayes' Theorem:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

What we want to discuss today.
Likelihood:
1. Given that our model parameters are true
2. The probability density distribution of observing the data we obtained

Standard treatment: treat data as an incident from a Poisson distribution with expected rate N_{mc}/s:

$$P(x|\theta) = \frac{(N_{mc}/s)^{N_{\text{data}}} e^{-N_{mc}/s}}{N_{\text{data}}!}$$



- Observed data and Monte Carlo prediction MC as a function of θ
- s is the POT scaling factor between Mote Carlo and Data
- Only 1 bin shown here as an example for likelihood

Defect of Extended Binned Likelihood

- The data we observe (denoted by x) can constrain our model parameters (denoted by θ) by Bayes' Theorem:

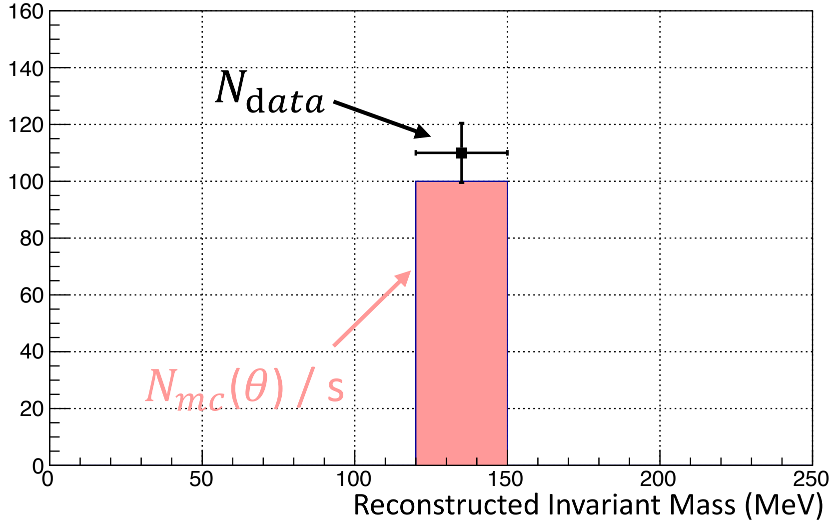
$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

Monte Carlo sample is a finite sample, N_{mc} does not strictly represent the population mean expectation N_θ

- There's a statistical fluctuation between N_{mc} and N_θ

Standard treatment: treat data as an incident from a Poisson distribution with expected rate N_{mc}/s:

$$P(x|\theta) = \frac{(N_{mc}/s)^{N_{data}} e^{-N_{mc}/s}}{N_{data}!}$$



- Observed data and Monte Carlo prediction MC as a function of θ
- s is the POT scaling factor between Mote Carlo and Data
- Only 1 bin shown here as an example for likelihood

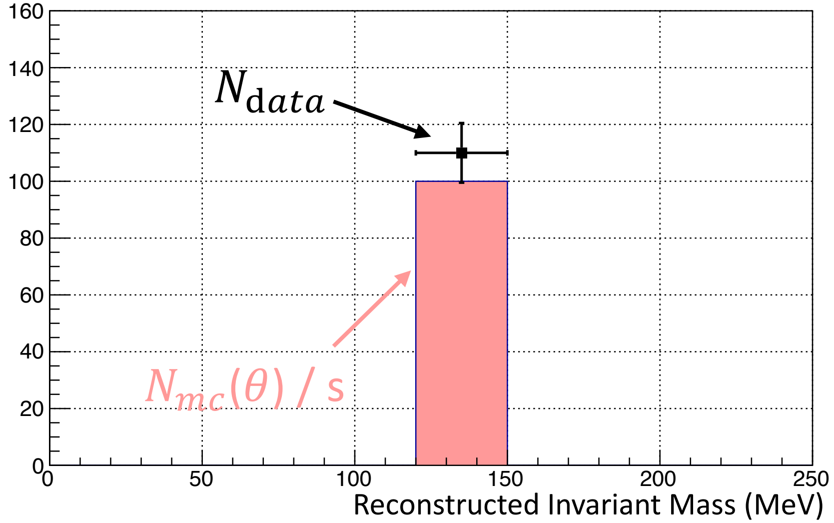
Solution: Derive it with Bayes' Theorem

- From the law of total probability

$$P(x|\theta) = \int P(x|N_{true})\left[\int P(N_{true}|N_\theta)P(N_\theta|\theta)dN_\theta\right]dN_{true}$$

$$\downarrow$$

$$P(N_{true}|\theta)$$



- Observed data and Monte Carlo prediction MC as a function of θ
- s is the POT scaling factor between Mote Carlo and Data
- Only 1 bin shown here as an example for likelihood

Solution: Derive it with Bayes' Theorem

- From the law of total probability

$$P(x|\theta) = \int P(x|N_{true})\left[\int P(N_{true}|N_\theta)P(N_\theta|\theta)dN_\theta\right]dN_{true}$$

N_{true}: Data Truth Expectation
Poisson distribution:

$$P(x|N_{true}) = \frac{{N_{true}}^{N_{data}}}{N_{data}!}e^{-N_{true}}$$

N_θ: Monte Carlo Truth Expectation

Recall the condition: Given that our model parameters are true.

We expect N_θ to equal N_{true}

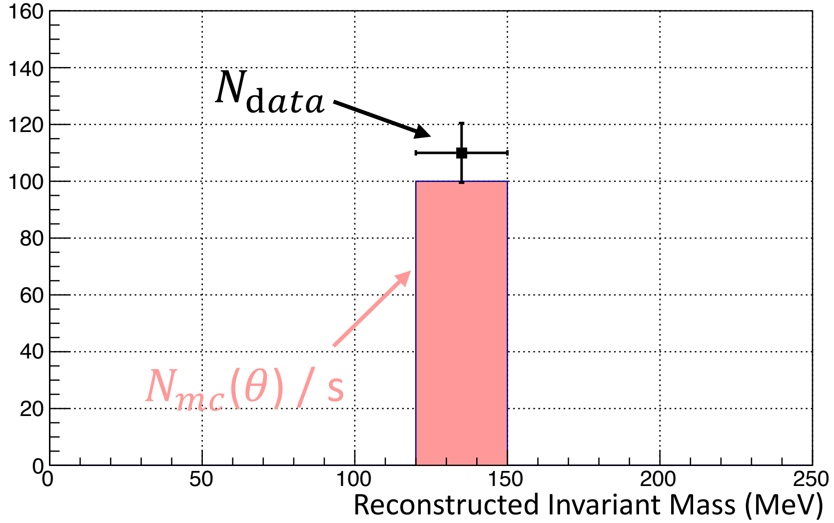
$$P(N_{true}|N_\theta) = \delta(N_{true} - N_\theta/s)$$

Apply Bayes' Theorem again:

$$P(N_\theta|\theta) = \frac{P(\theta|N_\theta)P(N_\theta)}{P(\theta)}$$

1. $P(N_\theta)$: only information we have is that N_θ is finite, assume a uniform distribution between [0, large number]
2. $P(\theta)$: constant
3. $P(\theta|N_\theta)$: Poisson distribution

$$P(\theta|N_\theta) = \frac{{N_\theta}^{N_{mc}}}{N_{mc}!}e^{-N_\theta}$$



- Observed data and Monte Carlo prediction MC as a function of θ
- s is the POT scaling factor between Mote Carlo and Data
- Only 1 bin shown here as an example for likelihood

Coincidence with Binominal Distribution

- From the law of total probability

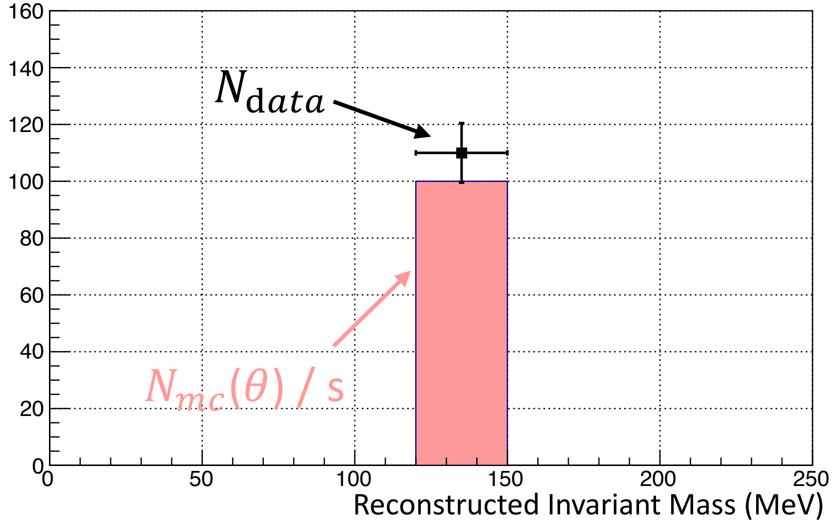
$$P(x|\theta) = \int P(x|N_{true})\left[\int P(N_{true}|N_\theta)P(N_\theta|\theta)dN_\theta\right]dN_{true}$$

$$\frac{1}{ub}\frac{1}{P(\theta)}\frac{s}{s+1}\frac{(N_{data}+N_{mc})!}{N_{data}!\,N_{mc}!}\left(\frac{1}{s+1}\right)^{N_{data}}\left(\frac{s}{s+1}\right)^{N_{mc}}$$

Binominal distribution with total number of events $N_{data} + N_{mc}$

Two out comes:

1. Data
   - $P(data) = \frac{1}{s+1}$
   - Number of outcome: N_{data}
2. MC
   - $P(mc) = \frac{s}{s+1}$
   - Number of outcome: N_{mc}



- Observed data and Monte Carlo prediction MC as a function of θ
- s is the POT scaling factor between Mote Carlo and Data
- Only 1 bin shown here as an example for likelihood

Bayesian Likelihood with Monte Carlo Stat Err

Extended binned likelihood: $-\ln \mathcal{L}_{stat} = N_{mc}/s - N_{data} + N_{data} \ln \frac{N_{data}}{N_{mc}/s}$

Approximated Barlow-Beeston (from TN-395): $-\ln \mathcal{L}_{stat} = N_{MC}^{true} - N_{data} + N_{data} \ln \frac{N_{data}}{N_{MC}^{true}} + \frac{(\beta-1)^2}{2\sigma_\beta^2}$

- $\begin{cases} \beta^2 + (N_{mc}\sigma_\beta^2 - 1)\beta - N_{data}\sigma_\beta^2 = 0 \\ (\sigma_\beta = \sqrt{\sum w^2/N_{mc}}) \end{cases} \rightarrow \beta \rightarrow N_{MC}^{true} = \beta N_{mc}$

Bayesian likelihood: $-\ln \mathcal{L}_{stat} = (N_{mc} + N_{data} + 1)\ln\left(1 + \frac{1}{s}\right) - N_{mc}\ln\left(1 + \frac{N_{data}}{N_{mc}}\right) + N_{data} \ln \frac{N_{data}}{N_\theta/s + N_{data}/s}$

- When s is large: $-\ln \mathcal{L}_{\mathrm{stat}} \approx \frac{1}{s} + N_{mc}/s + N_{data}/s - N_{data} + N_{data} \ln \frac{N_{data}}{N_{mc}/s + N_{data}/s}$
- Similar idea as Barlow-Beeston, account for likelihood data vs. truth & mc vs. truth

Comparison

Approximated Barlow-Beeston likelihood is derived in frequentist inference, and with assumptions

- β follows normal distribution, σ_β approximated as a constant (assumed in implementation)
- In frequentist inference, maximum likelihood is wanted, solve for β

These are valid and reasonable in frequentist inference, and even an almost acceptable approximation in Bayesian inference.

Bayesian Likelihood (the Bayesian and accurate way)

- Entirely derived from Bayes' theorem
- Assumes N_θ follows a uniform distribution (no prior information on it except the upper bound)
- Can be thought of as data constrains N_θ, and N_θ constrains model

Bayesian Likelihood (TN under finalization)

- A detailed TN has been written describing and deriving the likelihood
- T2K-TN-454
- Next slides will show some example of MCMC sampling results from this likelihood

Binned Likelihood with Monte Carlo Statistical Uncertainty in Baysian Inference

Shilin Liu, Clark McGrew

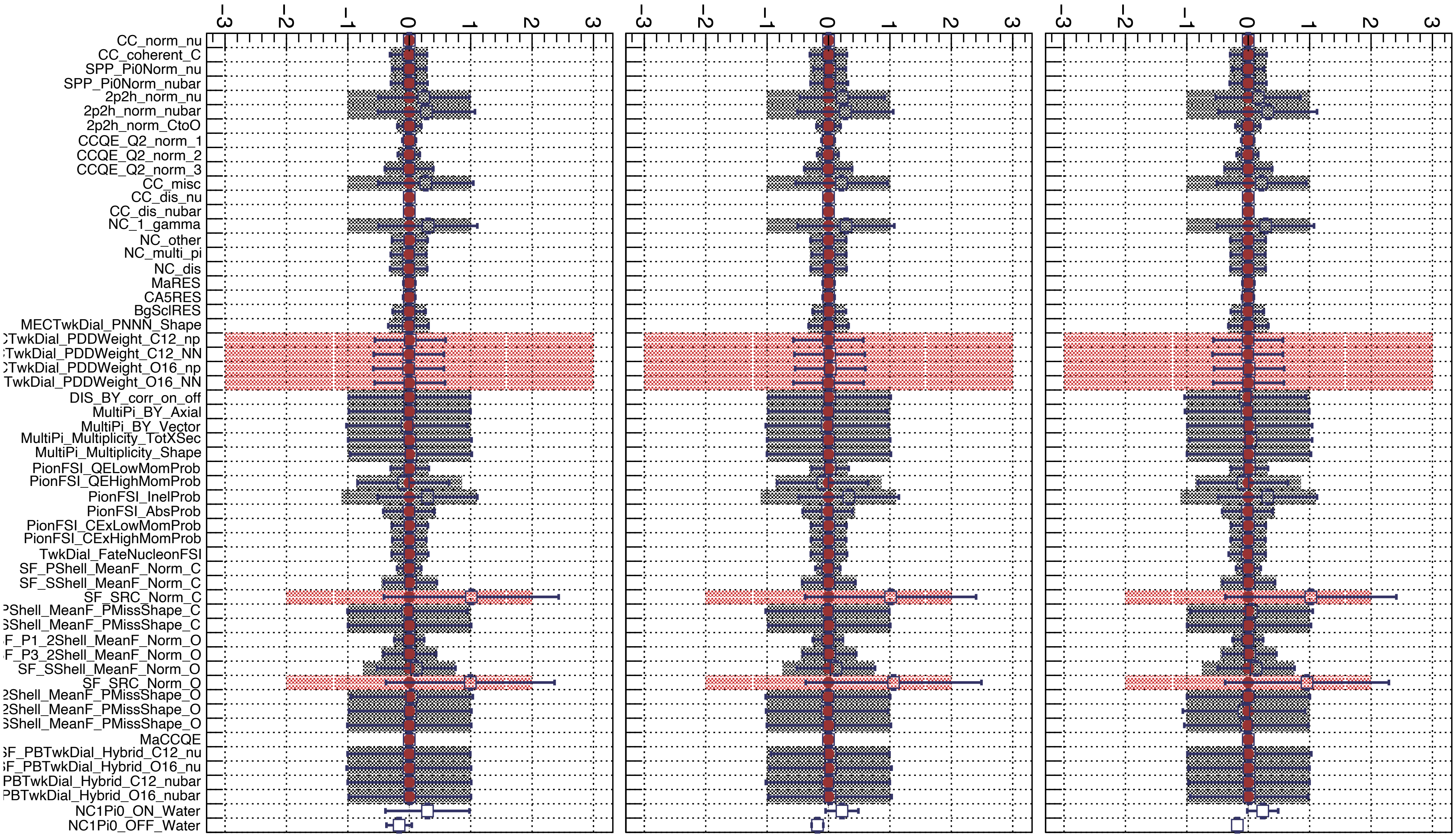
January 2023

1 Introduction

Bayes's theorem is used to obtain the data evidenced theory parameters distribution, as shown in (1) :

$$P(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{x})} \quad (1)$$

MCMC Sampling Results for My Bayesian Likelihood



S = 0.1 | S = 10 | Extended Binned Likelihood

MCMC Sampling Results for My Bayesian Likelihood

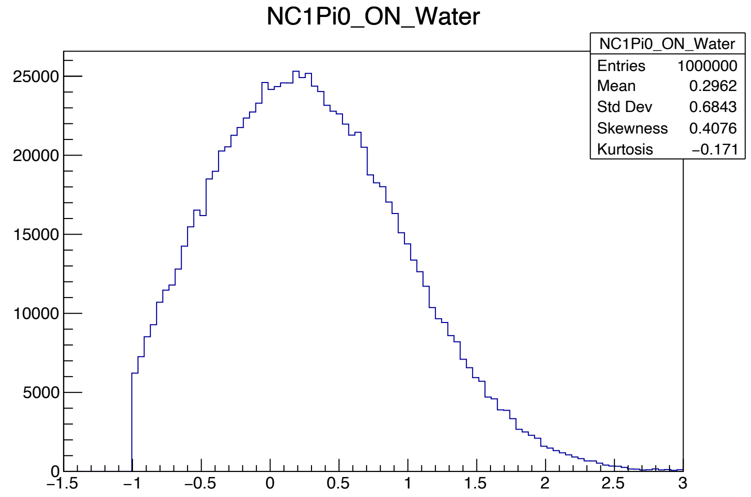
As expected, no constraint to xsec parameters, MCMC is sampling the prior distribution of them

Thinking about following cases separately

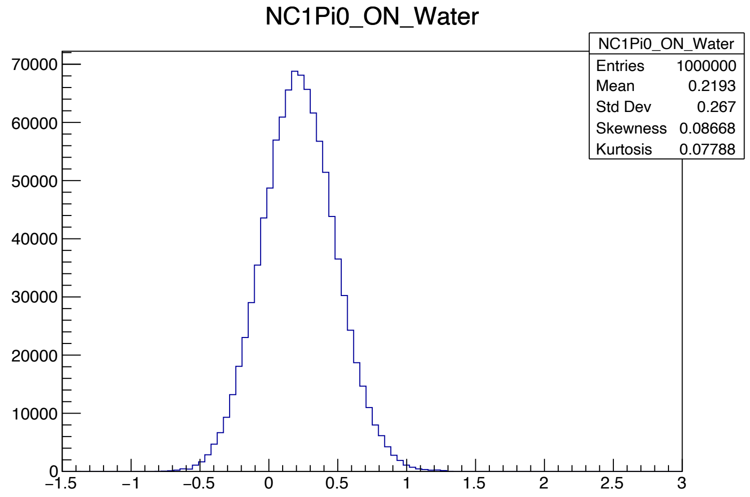
- Extended binned likelihood: Data constrains the generated MC, assuming no MC stat fluctuation

My Bayesian Likelihood: data constrains the truth histogram → the truth histogram constrains model (parameter)

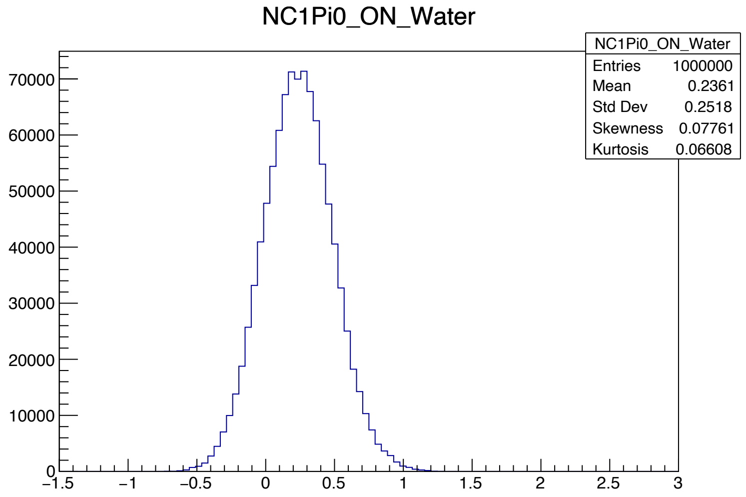
- S = 0.1, Monte Carlo sample is small, we know less about where the parameter is
- S = 10, Monte Carlo sample is large, we know better about where the parameter is
- S = 10, still a little wider (std dev) comparing to extended binned likelihood, which is from MC stat fluctuation



S = 0.1



S = 10



Extended Binned Likelihood

Backup