A new unified perspective on the problem of limited Monte Carlo for likelihood calculations

Based on 1712.01293
and "to be published"
(in a few days on arXiv)

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

PHYSTAT-ν CERN 2019

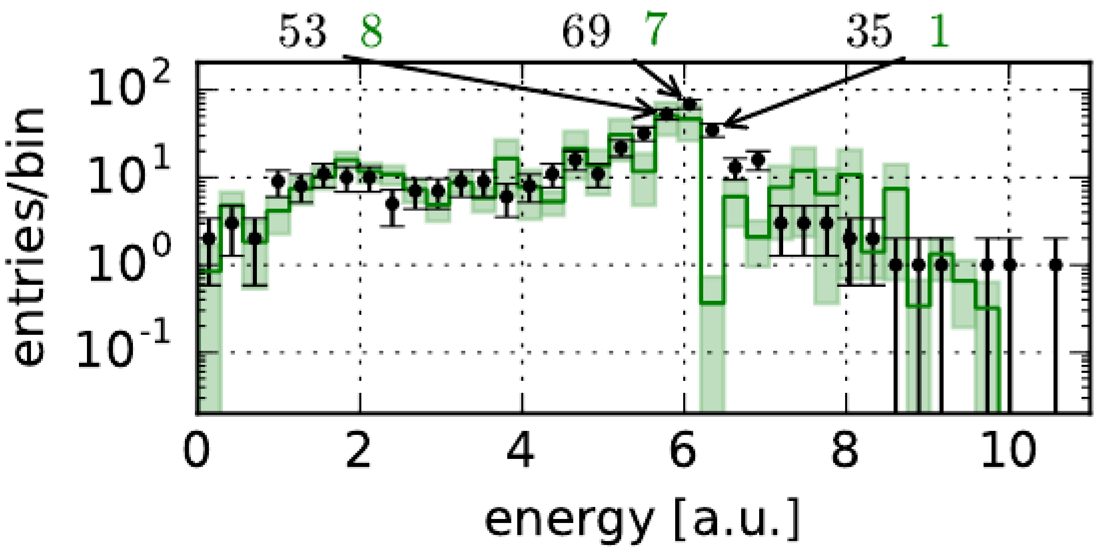
Thorsten Glüsenkamp,

Phystat-ν, CERN, Jan. 2019

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!}$$

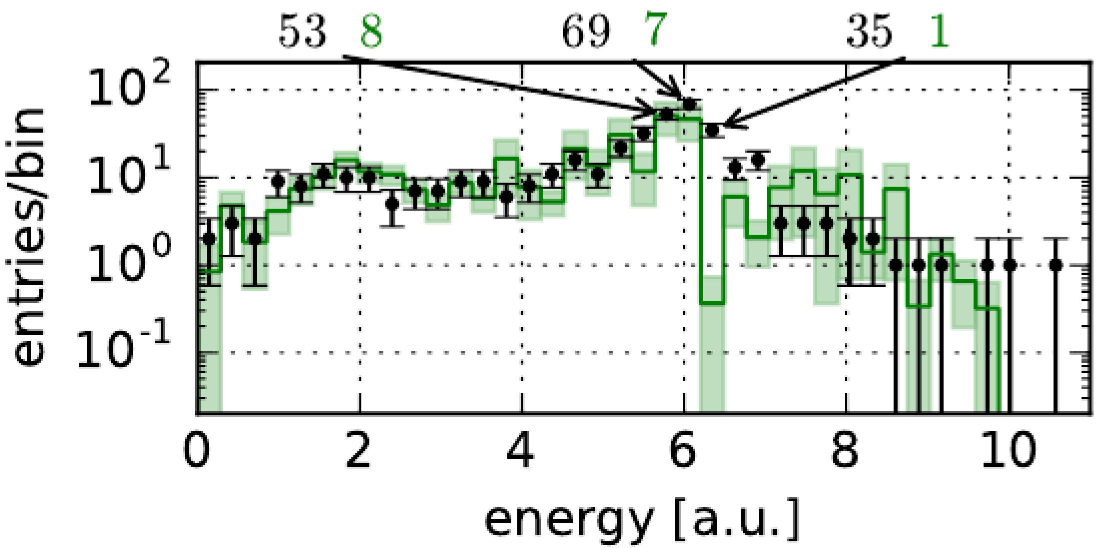


MC ≈ 0.1DATA

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum w_i^{\,k}}{k!}$$

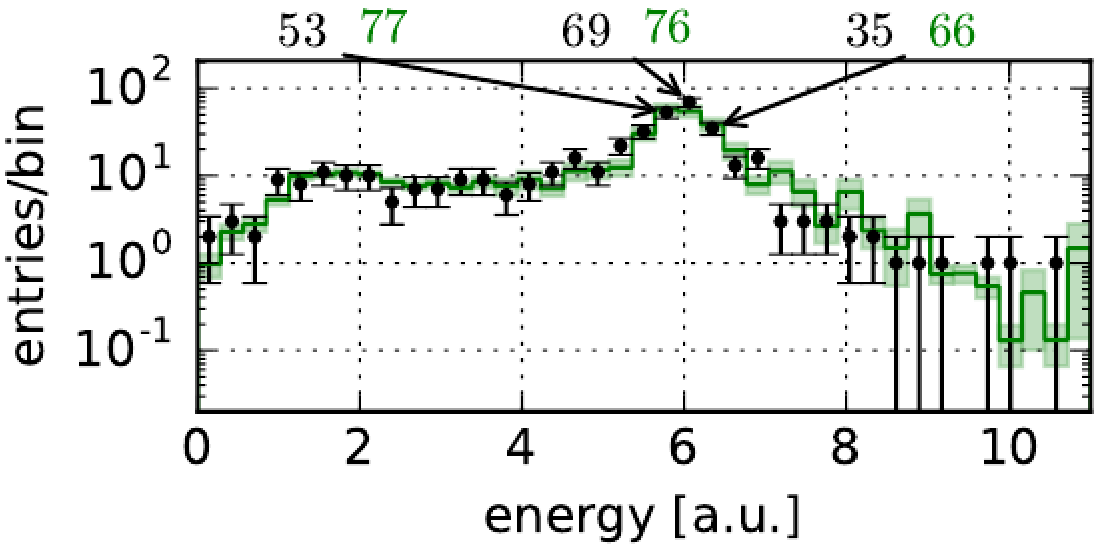


MC ≈ 0.1DATA

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum w_i{}^k}{k!}$$

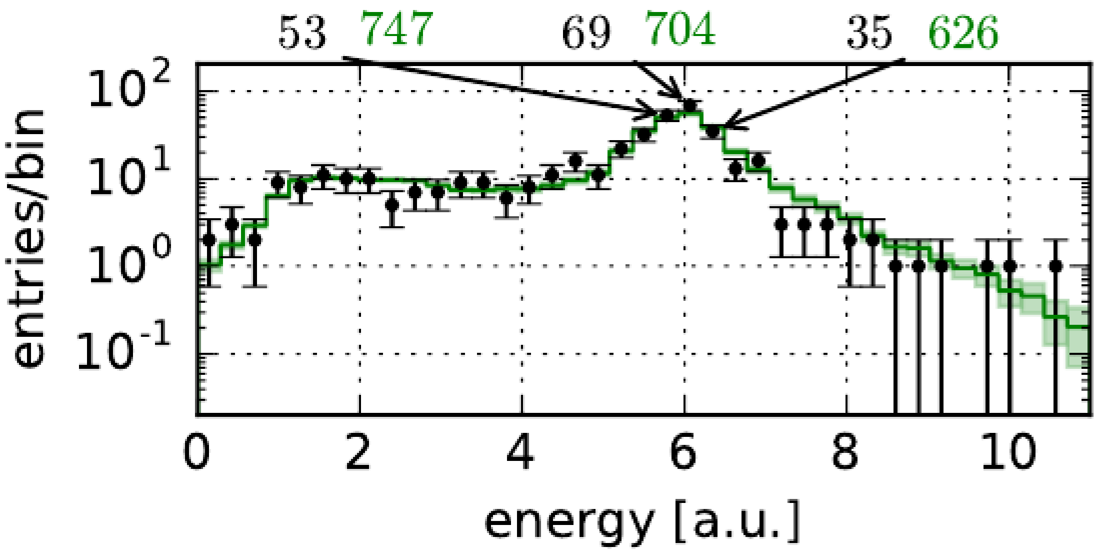


MC ≈ DATA

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!}$$



MC ≈ 10 DATA

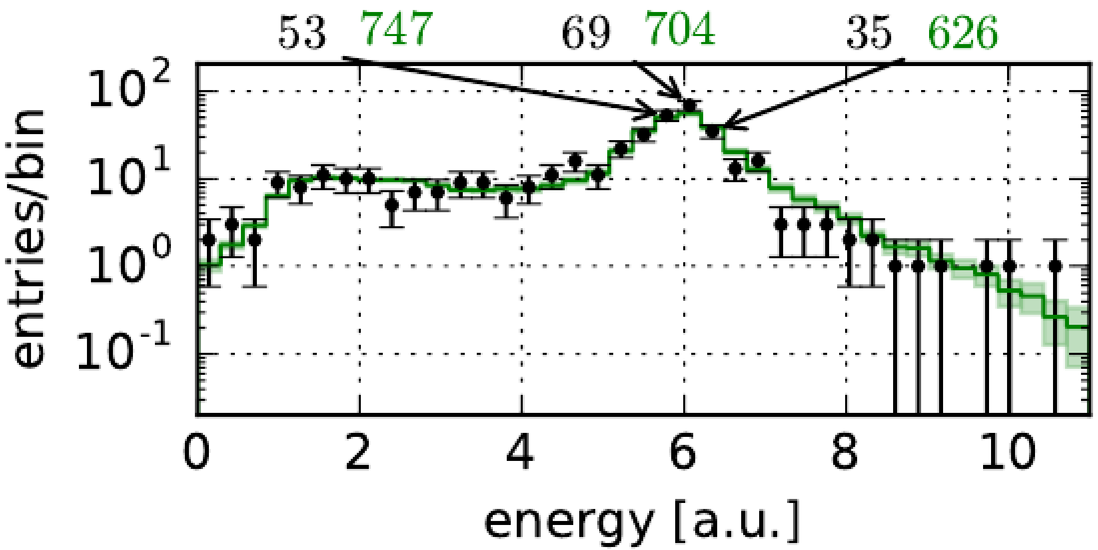
Barlow, Beeston (93)

that these are damped by a factor N_D/N_j, but we cannot hope that this is small. There is a general rule of thumb that the MC samples should be ten times larger than the data sample, so any effects of finite MC data size are relatively small. Unfortunately many

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!}$$



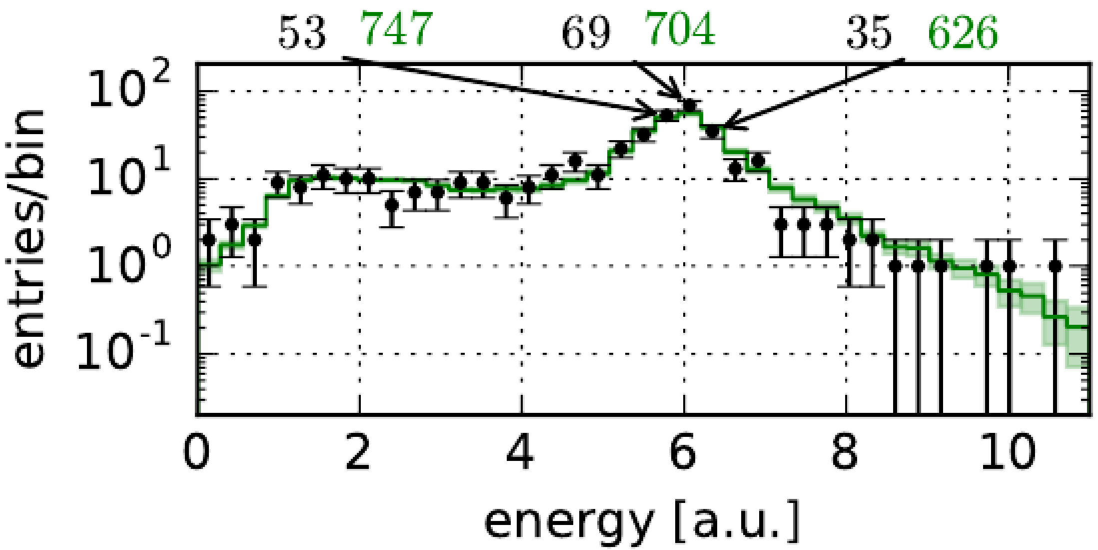
MC $\approx$ 10 DATA

Q: If we cannot get 10X MC, which procedures exist to handle the small MC samples?

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!}$$



MC ≈ 10 DATA

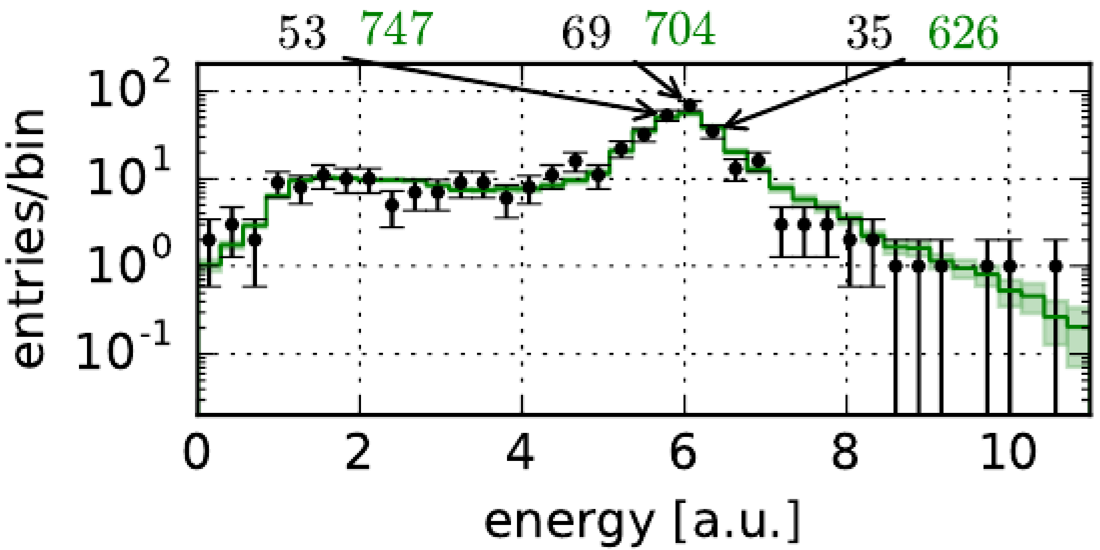
Q: If we cannot get 10X MC, which procedures exist to handle the small MC samples?

A: Barlow/Beeston (´93) or Bohm/Zech (´12/´14) or Chirkin (´13) or T.G. (´18) or Argüelles et al (19´)

The challenge: limited MC in binned likelihoods

Example Poisson (1 bin):

$$\frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!}$$



MC ≈ 10 DATA

Q: If we cannot get 10X MC, which procedures exist to handle the small MC samples?

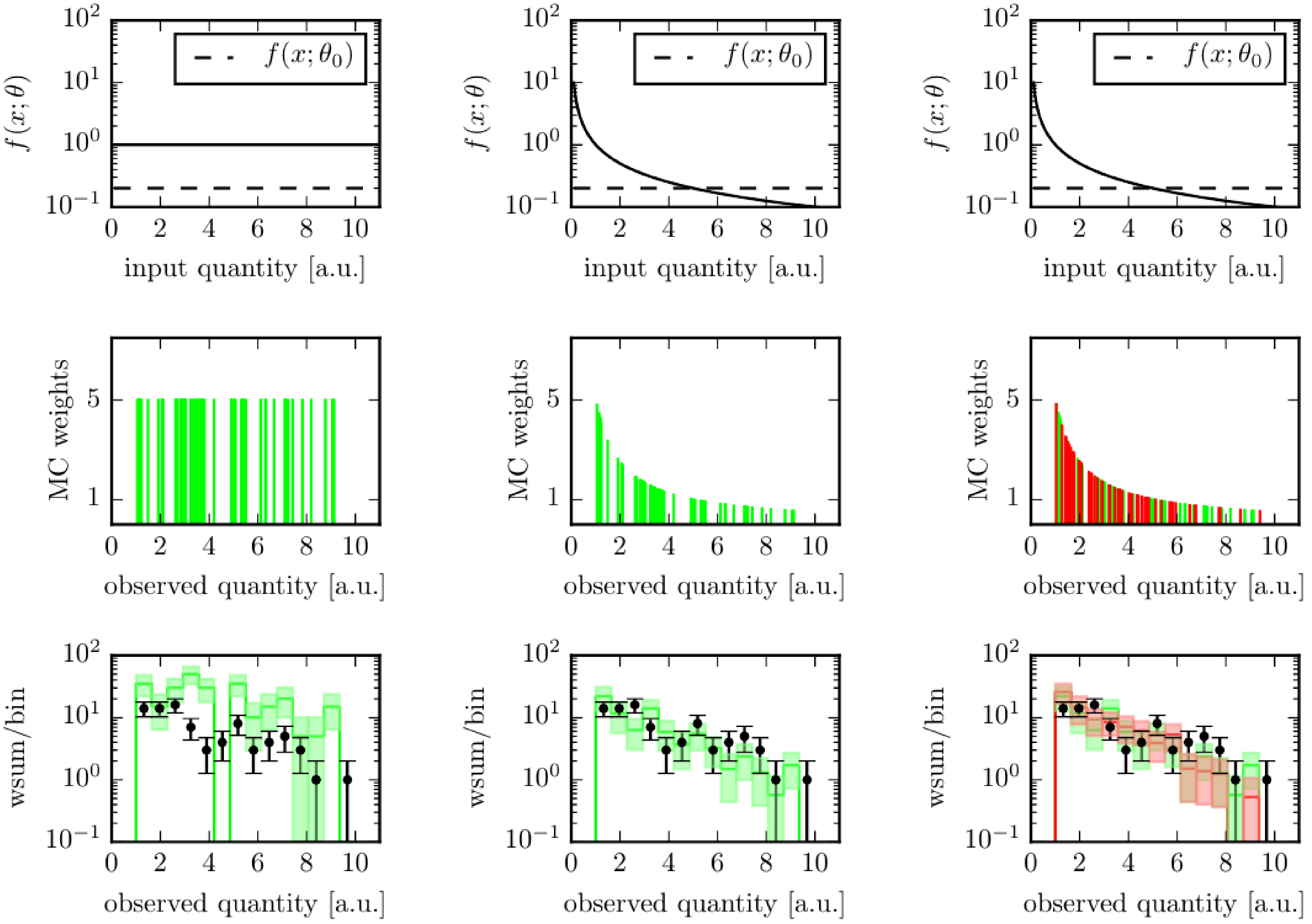
A: Barlow/Beeston (´93) or Bohm/Zech (´12/´14) or Chirkin (´13) or T.G. (´18) or Argüelles et al (19´)

This talk: All approaches fundamentally approximate the CPD – with pros n cons

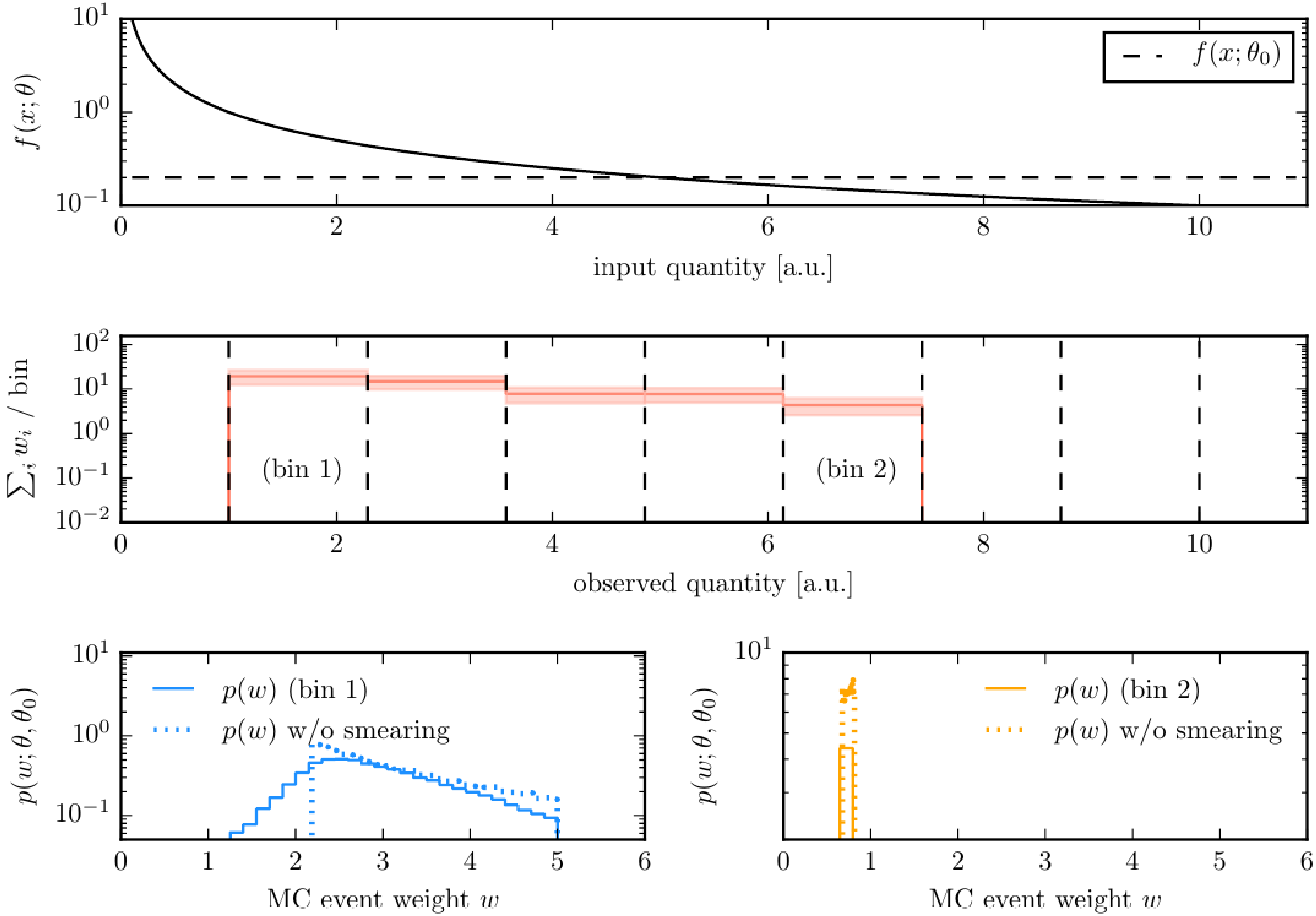
Overview

1. The probability distribution for the sum of weights
   Compound **P**oisson **D**istribution (CPD)
2. Approximations of the CPD in existing approaches
   - Probabilistic approaches have interesting connections to special functions, statistics, B-Splines
3. Some further not-yet discussed solutions
4. Performance Comparisons
5. Summary

3 steps to understand the CPD ..

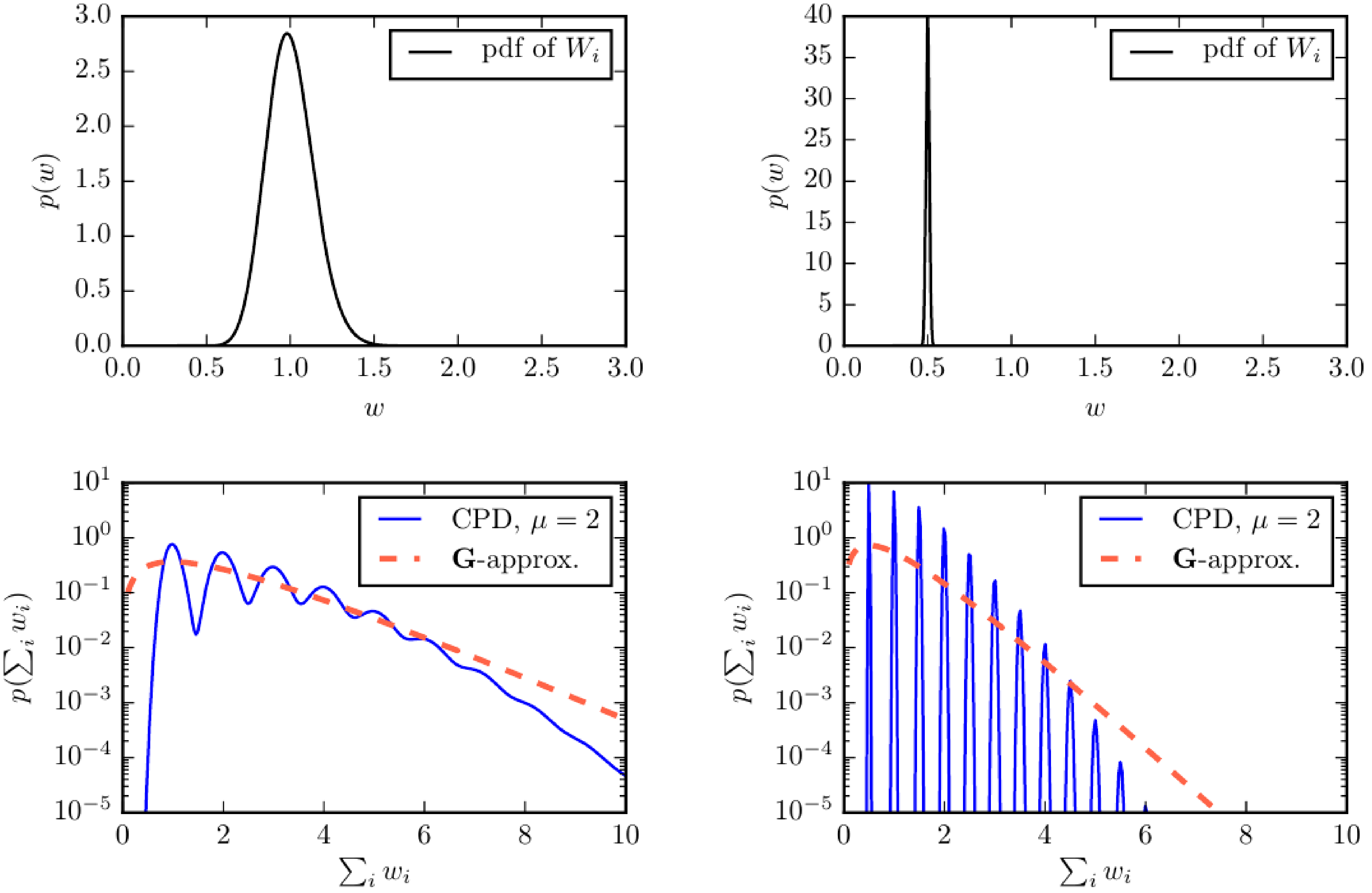


Step1: MC weighting

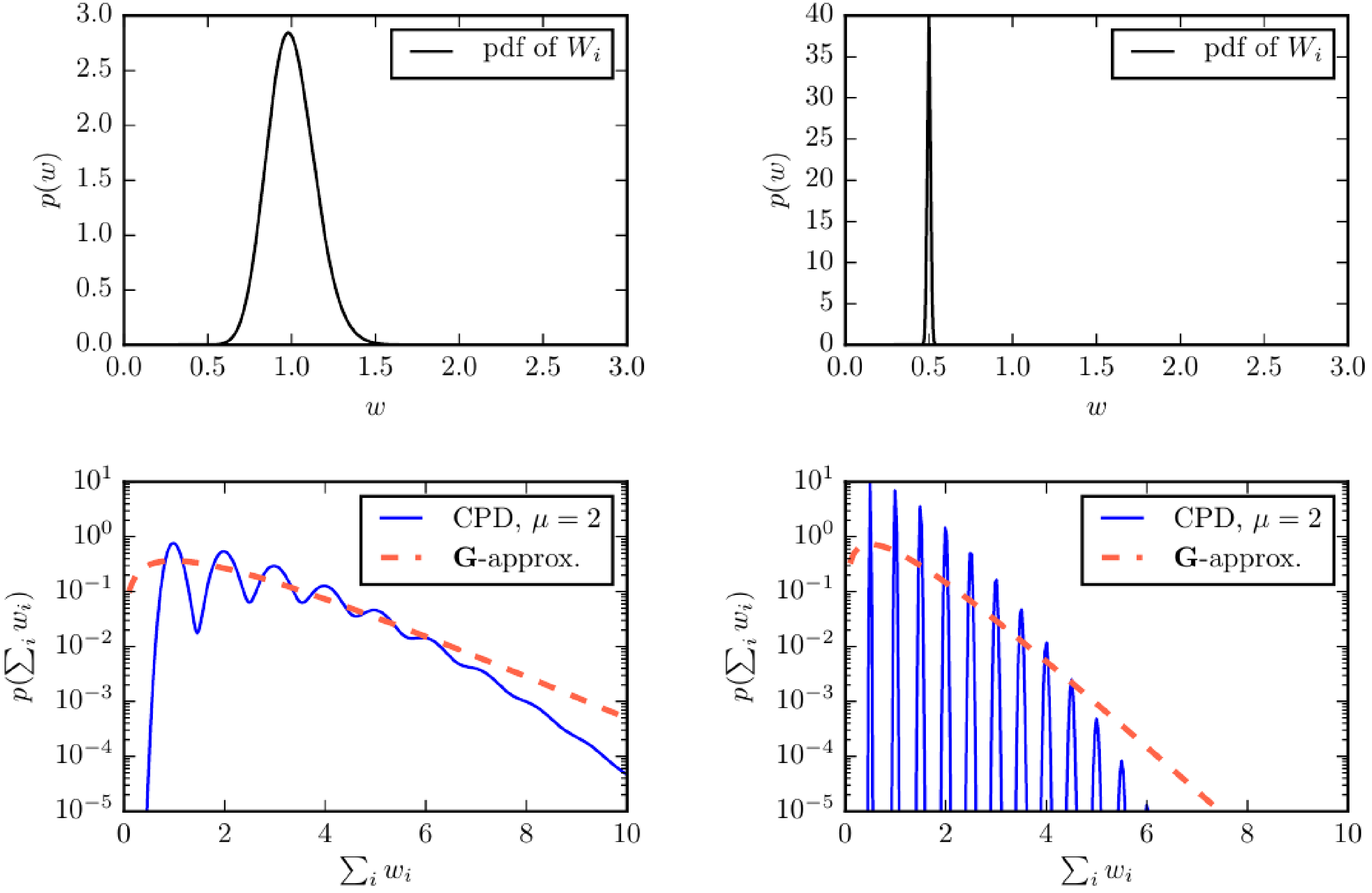


Step2: the distribution p(w)

Step3 – the CPD



Step3 – the CPD



Probability distribution for the sum of weights: $p(\sum w)$

Random variable of CPD: Z

$$Z = \sum_{i=1}^{N} W_i$$

$$N \sim \text{Poisson}$$

$$\widehat{\mu}(Z) = \sum_i w_i$$

$$\widehat{\text{var}}(Z) = \sum_i w_i^2$$

↑

$$\text{E}[Z] = \mu_z = \text{E}[N] \cdot \text{E}[W]$$

$$\text{Var}[Z] = \text{E}[N] \cdot \text{Var}[W] + (\text{E}[W])^2 \cdot \text{Var}[N]$$

What we would like to do:

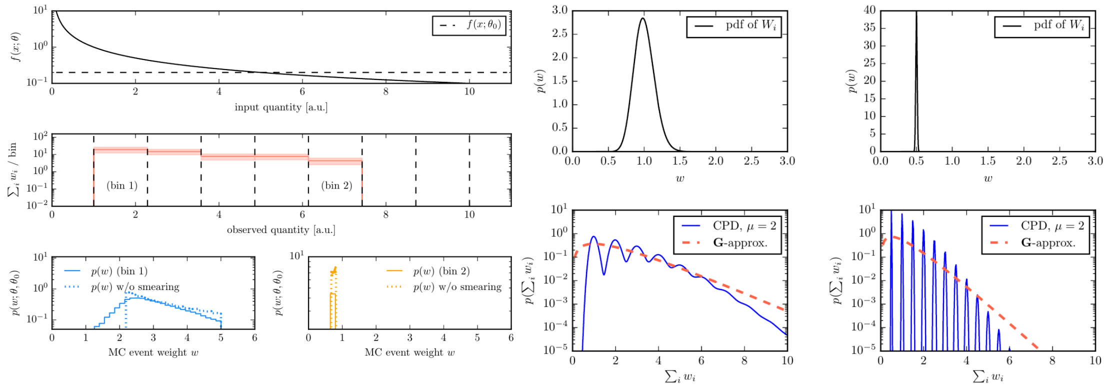
- Take PDF of the sum of weights (CPD) and integrate the Poisson mean

$$L_{bin,exact} = \int \frac{e^{-\lambda} \cdot \lambda^k}{k!} \cdot p_{CPD}(\lambda) d\lambda$$

What we would like to do:

- Take PDF of the sum of weights (CPD) and integrate the Poisson mean

$$L_{bin,exact} = \int \frac{e^{-\lambda} \cdot \lambda^k}{k!} \cdot p_{CPD}(\lambda) d\lambda$$



- However: CPD is <u>not tractable</u> (except for equal weights)

One possibility: Approximate and then integrate

Statistics of weighted MC

$$Z = \sum_{i=1}^{N} W_i$$
$$N \sim \text{Poisson}$$
$$\widehat{\mu}(Z) = \sum_i w_i$$
$$\widehat{\text{var}}(Z) = \sum_i w_i^2$$

Intead of

$$L_{bin,exact} = \int \frac{e^{-\lambda} \cdot \lambda^k}{k!} \cdot p(\lambda)_{CPD} d\lambda$$

Argüelles et al. (2019)

$$\int \mathbf{P}(k;\lambda) \cdot \mathbf{G}(\lambda;\alpha,\beta) d\lambda$$

$$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a \ , \ \beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$$

Statistics of weighted MC

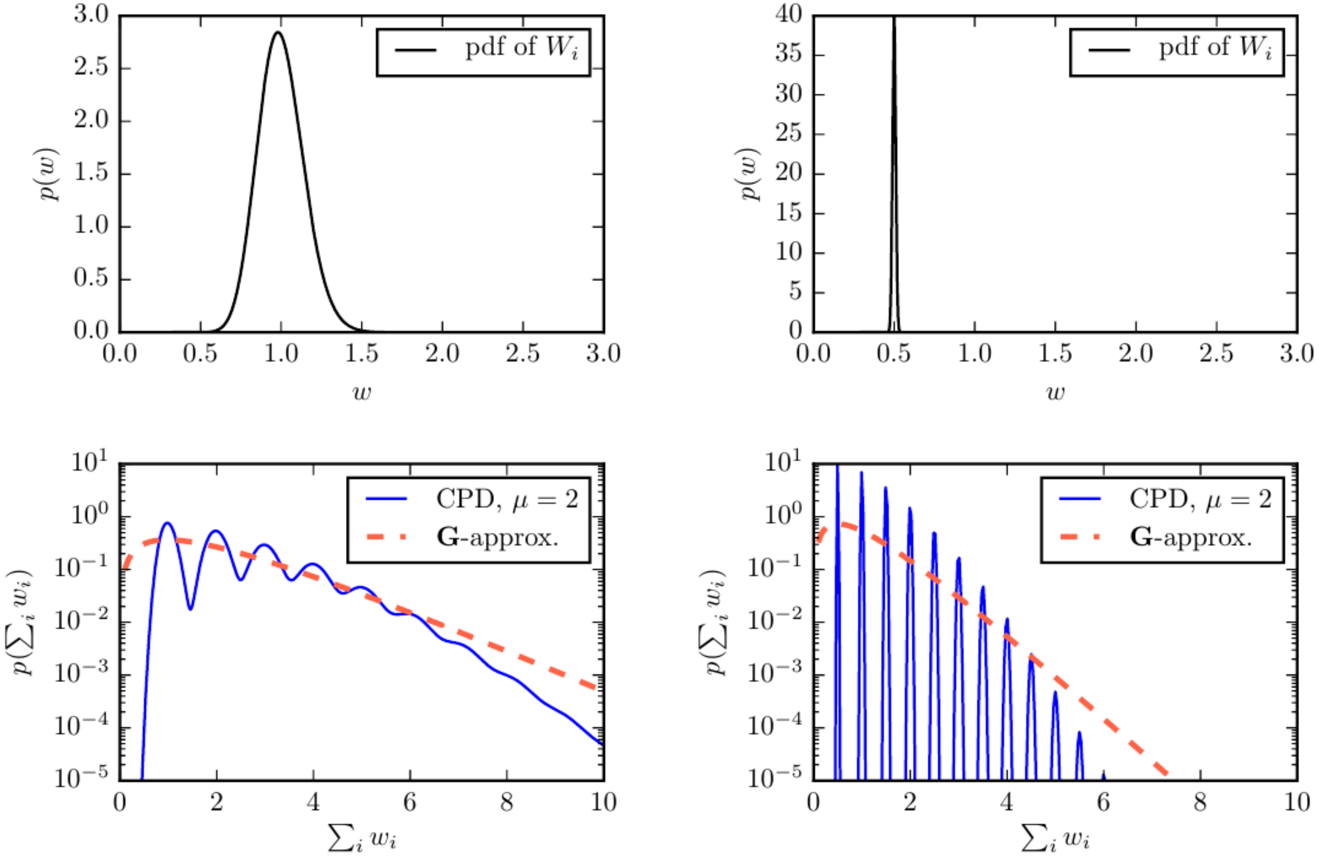
$Z = \sum_{i=1}^{N} W_i$

$N \sim$ Poisson

$\widehat{\mu}(Z) = \sum_i w_i$

$\widehat{\text{var}}(Z) = \sum_i w_i^2$

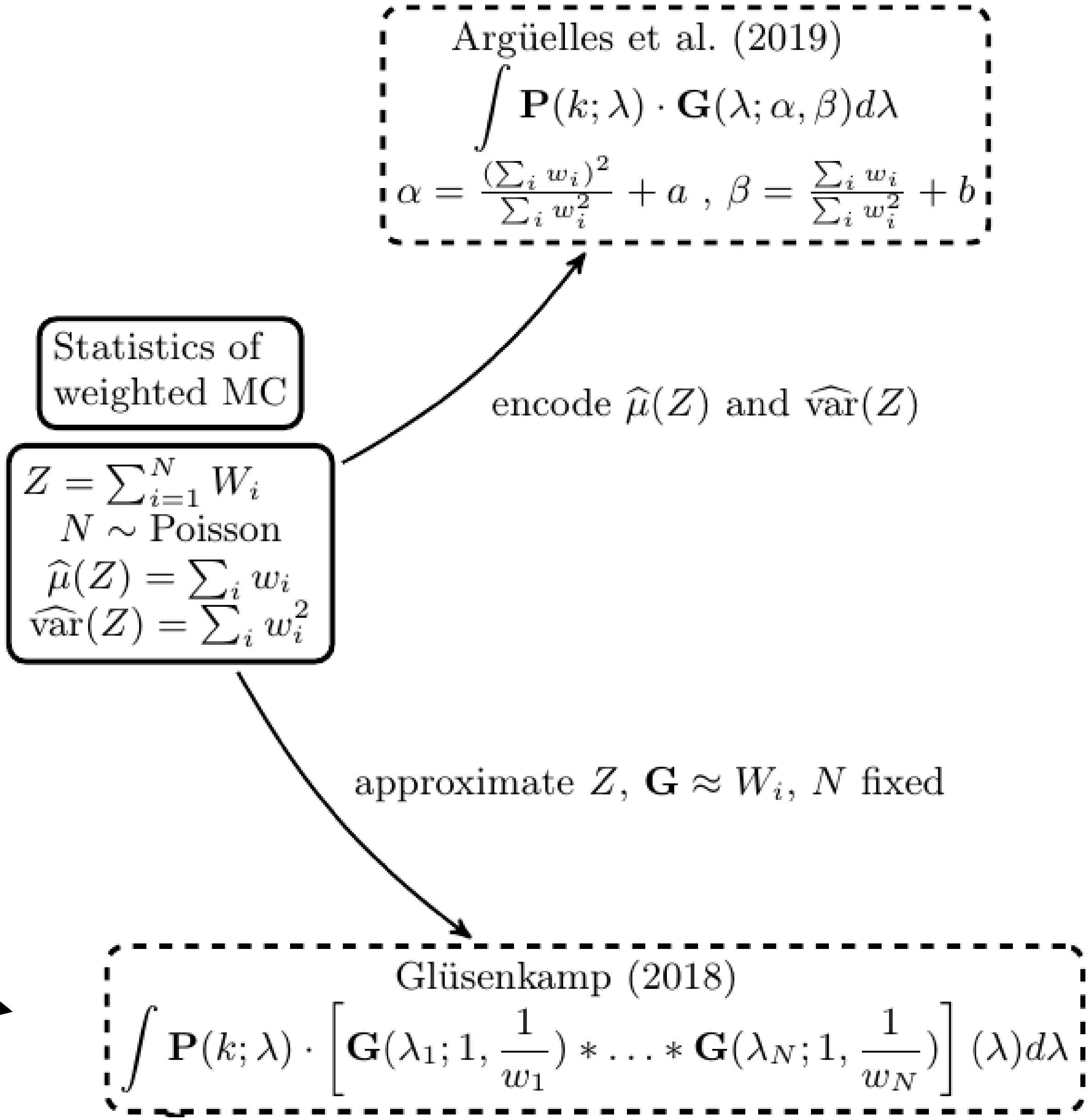
encode $\widehat{\mu}(Z)$ and $\widehat{\text{var}}(Z)$

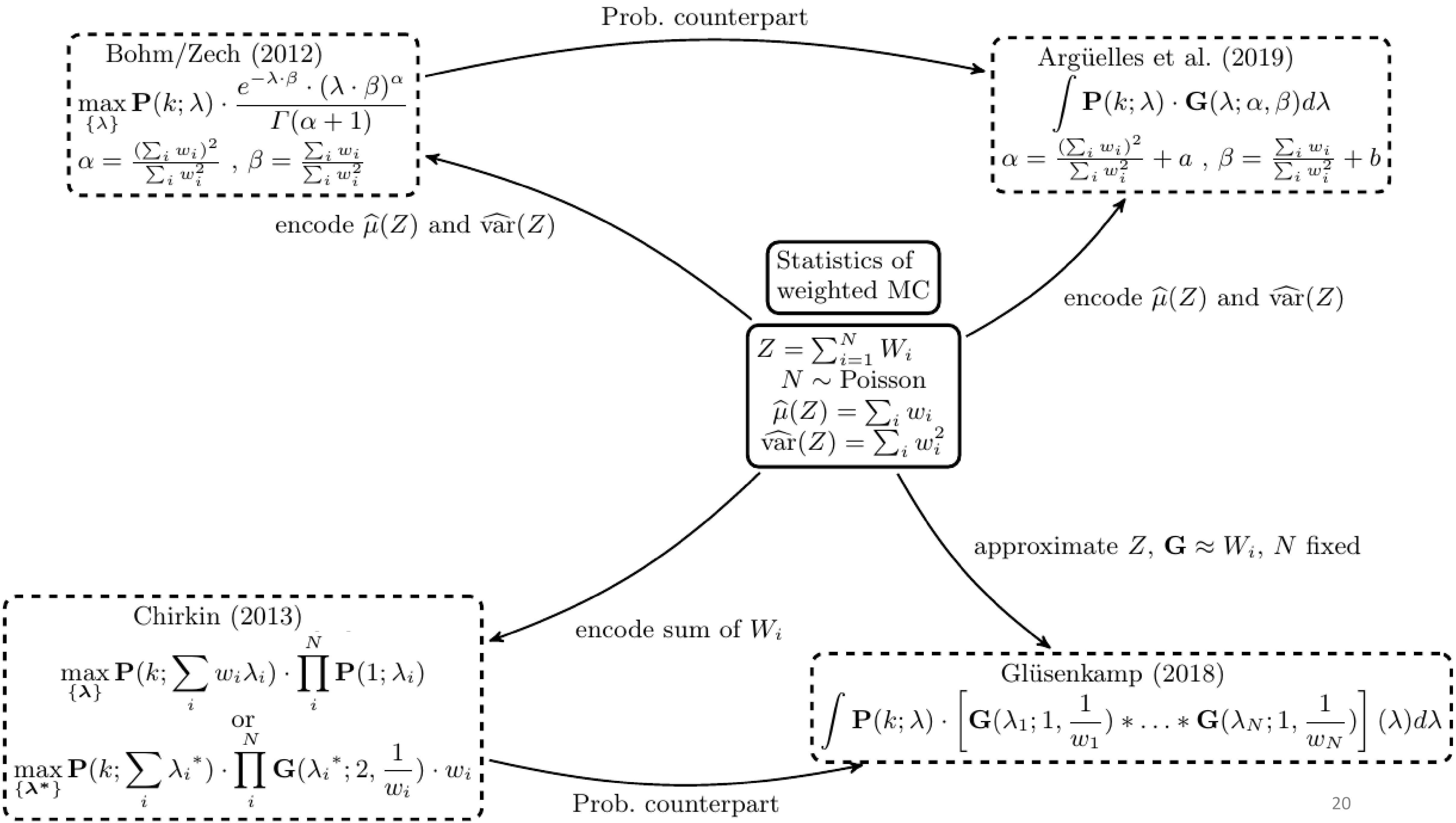


$$P(k; \Sigma\, w_i) = \frac{e^{-\sum w_i} \cdot \sum {w_i}^k}{k!} =$$

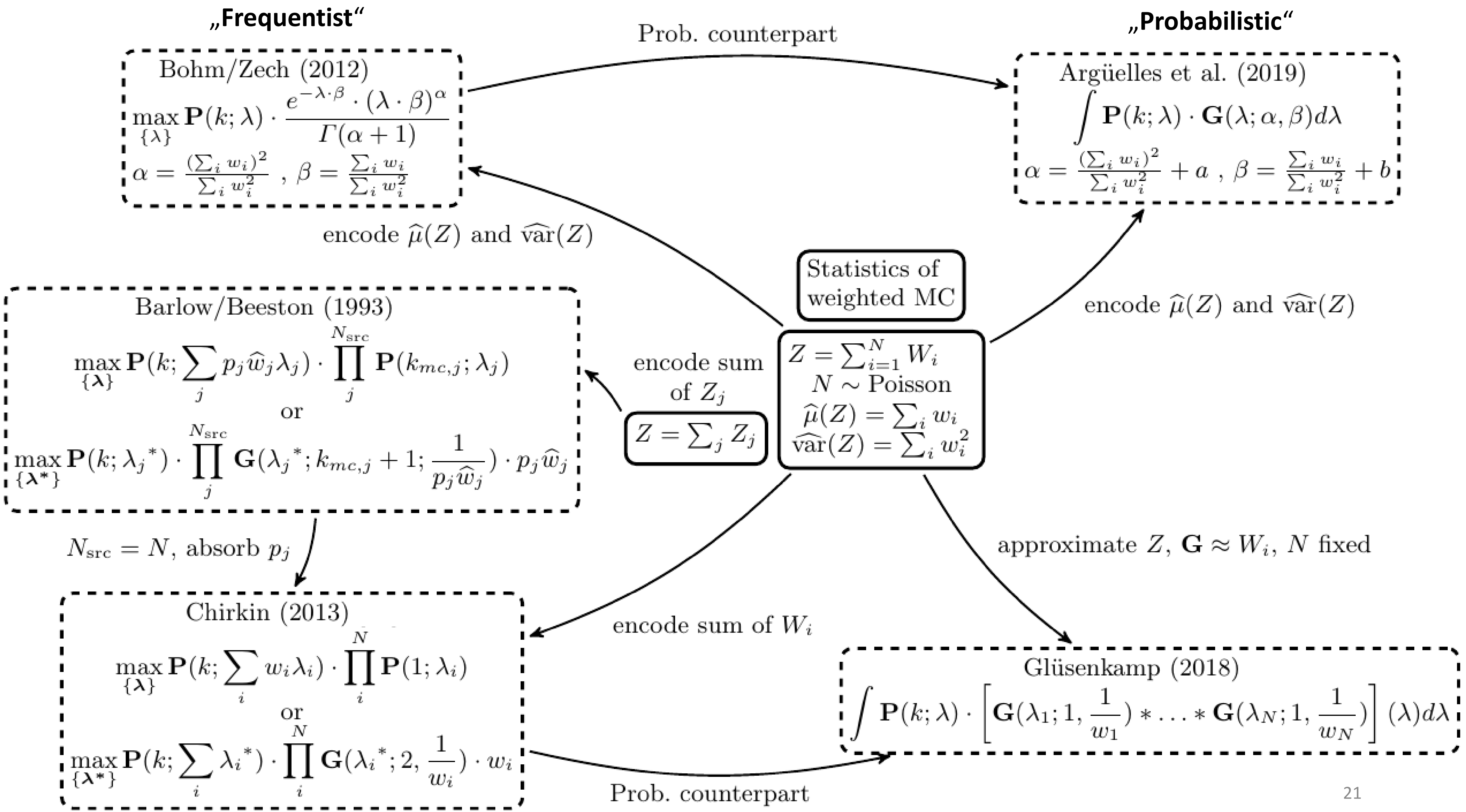
$$= \int \frac{e^{-\lambda} \cdot \lambda^k}{k!}\, \delta\left(\lambda - \sum w_i\right) d\lambda$$

$$= \int P(k; \lambda) \cdot [\delta(\lambda - w_1) * \cdots * \delta(\lambda - w_n)](\lambda)\, d\lambda$$





Prob. counterpart
Bohm/Zech (2012)
$\max_{\{\lambda\}} \mathbf{P}(k;\lambda) \cdot \frac{e^{-\lambda \cdot \beta} \cdot (\lambda \cdot \beta)^{\alpha}}{\Gamma(\alpha+1)}$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2}$
Argüelles et al. (2019)
$\int \mathbf{P}(k;\lambda) \cdot \mathbf{G}(\lambda;\alpha,\beta) d\lambda$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$
encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$
Statistics of weighted MC
encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$
$Z = \sum_{i=1}^{N} W_i$
$N \sim$ Poisson
$\widehat{\mu}(Z) = \sum_i w_i$
$\widehat{\mathrm{var}}(Z) = \sum_i w_i^2$
approximate Z, $\mathbf{G} \approx W_i$, N fixed
Chirkin (2013)
$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k; \sum_i w_i \lambda_i) \cdot \prod_i^N \mathbf{P}(1;\lambda_i)$
or
$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k; \sum_i \lambda_i{}^*) \cdot \prod_i^N \mathbf{G}(\lambda_i{}^*; 2, \frac{1}{w_i}) \cdot w_i$
encode sum of W_i
Glüsenkamp (2018)
$\int \mathbf{P}(k;\lambda) \cdot \left[\mathbf{G}(\lambda_1; 1, \frac{1}{w_1}) * \ldots * \mathbf{G}(\lambda_N; 1, \frac{1}{w_N})\right](\lambda) d\lambda$
Prob. counterpart


„Frequentist“
„Probabilistic“
Prob. counterpart
Bohm/Zech (2012)
$\max_{\{\lambda\}} \mathbf{P}(k;\lambda) \cdot \frac{e^{-\lambda\cdot\beta}\cdot(\lambda\cdot\beta)^{\alpha}}{\Gamma(\alpha+1)}$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2}$
Argüelles et al. (2019)
$\int \mathbf{P}(k;\lambda)\cdot\mathbf{G}(\lambda;\alpha,\beta)d\lambda$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$
encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$
Statistics of weighted MC
encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$
Barlow/Beeston (1993)
$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_j p_j \widehat{w}_j \lambda_j) \cdot \prod_j^{N_{\mathrm{src}}} \mathbf{P}(k_{mc,j};\lambda_j)$
or
$\max_{\{\boldsymbol{\lambda^*}\}} \mathbf{P}(k;\lambda_j{}^*) \cdot \prod_j^{N_{\mathrm{src}}} \mathbf{G}(\lambda_j{}^*; k_{mc,j}+1; \frac{1}{p_j\widehat{w}_j}) \cdot p_j\widehat{w}_j$
encode sum of Z_j
$Z = \sum_j Z_j$
$Z = \sum_{i=1}^{N} W_i$
$N \sim$ Poisson
$\widehat{\mu}(Z) = \sum_i w_i$
$\widehat{\mathrm{var}}(Z) = \sum_i w_i^2$
$N_{\mathrm{src}} = N$, absorb p_j
approximate Z, $\mathbf{G} \approx W_i$, N fixed
encode sum of W_i
Chirkin (2013)
$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_i w_i\lambda_i) \cdot \prod_i^{N} \mathbf{P}(1;\lambda_i)$
or
$\max_{\{\boldsymbol{\lambda^*}\}} \mathbf{P}(k;\sum_i \lambda_i{}^*) \cdot \prod_i^{N} \mathbf{G}(\lambda_i{}^*;2,\frac{1}{w_i}) \cdot w_i$
Glüsenkamp (2018)
$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{G}(\lambda_1;1,\frac{1}{w_1}) * \ldots * \mathbf{G}(\lambda_N;1,\frac{1}{w_N})\right](\lambda)d\lambda$
Prob. counterpart

Prob. counterpart

Bohm/Zech (2012)

$$\max_{\{\lambda\}} \mathbf{P}(k;\lambda) \cdot \frac{e^{-\lambda\cdot\beta}\cdot(\lambda \ldots}{\Gamma(\alpha + \ldots}$$

$$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}\ ,\ \beta = \frac{\sum_i w_i}{\sum_i w_i^2}$$

Argüelles et al. (2019)

$$\int \mathbf{P}(k;\lambda)\cdot\mathbf{G}(\lambda;\alpha,\beta)d\lambda$$

$$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a\ ,\ \beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$$

encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$

Statistics of weighted MC

$$Z = \sum_{i=1}^{N} W_i$$
$$N \sim \text{Poisson}$$
$$\widehat{\mu}(Z) = \sum_i w_i$$
$$\widehat{\mathrm{var}}(Z) = \sum_i w_i^2$$

encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$

Barlow/Beeston (1993)

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_j p_j\widehat{w}_j\lambda_j)\cdot\prod_j^{N_{\mathrm{src}}}\mathbf{P}(k_{mc,j};\lambda_j)$$

or

$$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\lambda_j{}^*)\cdot\prod_j^{N_{\mathrm{src}}}\mathbf{G}(\lambda_j{}^*;k_{mc,j}+1;\frac{1}{p_j\widehat{w}_j})\cdot p_j\widehat{w}_j$$

encode sum of Z_j

$$Z = \sum_j Z_j$$

Pros: - often faster
- Limit of large statistics equal to Poisson
- Interpretability
- simplicity

$N_{\mathrm{src}} = N$, absorb p_j

approximate Z, $\mathbf{G} \approx$ … fixed

Chirkin (2013)

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_i w_i\lambda_i)\cdot\prod_i^{N}\mathbf{P}(1;\lambda_i)$$

or

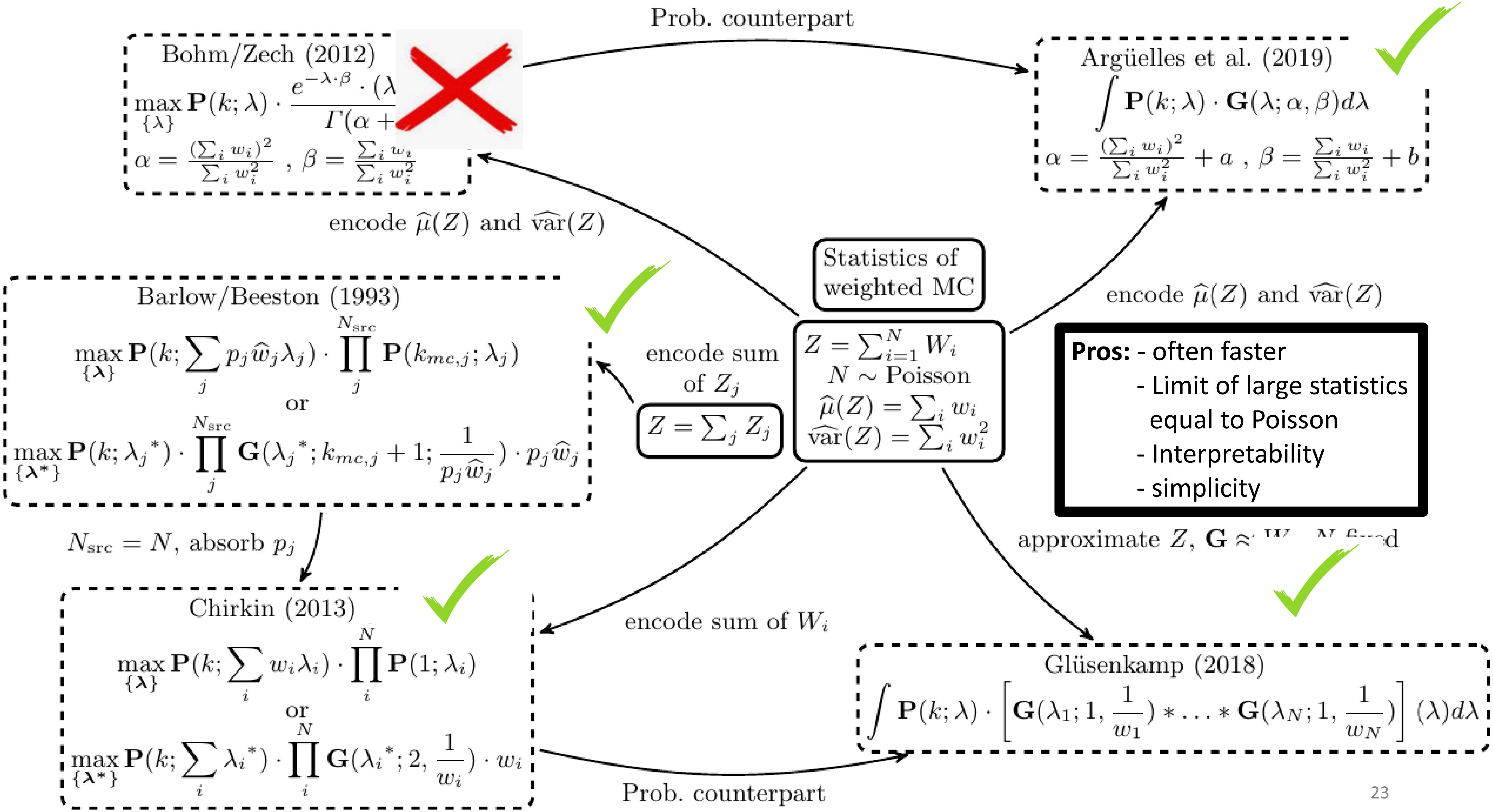
$$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\sum_i \lambda_i{}^*)\cdot\prod_i^{N}\mathbf{G}(\lambda_i{}^*;2,\frac{1}{w_i})\cdot w_i$$

encode sum of W_i

Glüsenkamp (2018)

$$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{G}(\lambda_1;1,\frac{1}{w_1}) * \ldots * \mathbf{G}(\lambda_N;1,\frac{1}{w_N})\right](\lambda)d\lambda$$

Prob. counterpart



Prob. counterpart
Bohm/Zech (2012)
$\max_{\{\lambda\}} \mathbf{P}(k;\lambda) \cdot \frac{e^{-\lambda\cdot\beta}\cdot(\lambda}{\Gamma(\alpha+}$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2}$
Argüelles et al. (2019)
$\int \mathbf{P}(k;\lambda)\cdot\mathbf{G}(\lambda;\alpha,\beta)d\lambda$
$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a$, $\beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$
encode $\widehat{\mu}(Z)$ and $\widehat{\text{var}}(Z)$
Statistics of weighted MC
Barlow/Beeston (1993)
$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_j p_j\widehat{w}_j\lambda_j)\cdot\prod_j^{N_{\text{src}}}\mathbf{P}(k_{mc,j};\lambda_j)$
or
$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\lambda_j{}^*)\cdot\prod_j^{N_{\text{src}}}\mathbf{G}(\lambda_j{}^*;k_{mc,j}+1;\frac{1}{p_j\widehat{w}_j})\cdot p_j\widehat{w}_j$
encode sum of Z_j
$Z = \sum_j Z_j$
$Z = \sum_{i=1}^N W_i$
$N \sim$ Poisson
$\widehat{\mu}(Z) = \sum_i w_i$
$\widehat{\text{var}}(Z) = \sum_i w_i^2$
encode $\widehat{\mu}(Z)$ and $\widehat{\text{var}}(Z)$
Pros: - often faster
- Limit of large statistics equal to Poisson
- Interpretability
- simplicity
$N_{\text{src}} = N$, absorb p_j
approximate Z, $\mathbf{G} \approx$
Chirkin (2013)
$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_i w_i\lambda_i)\cdot\prod_i^N\mathbf{P}(1;\lambda_i)$
or
$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\sum_i \lambda_i{}^*)\cdot\prod_i^N\mathbf{G}(\lambda_i{}^*;2,\frac{1}{w_i})\cdot w_i$
encode sum of W_i
Glüsenkamp (2018)
$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{G}(\lambda_1;1,\frac{1}{w_1}) * \ldots * \mathbf{G}(\lambda_N;1,\frac{1}{w_N})\right](\lambda)d\lambda$
Prob. counterpart

Prob. counterpart

Bohm/Zech (2012)

$$\max_{\{\lambda\}} \mathbf{P}(k;\lambda) \cdot \frac{e^{-\lambda\cdot\beta}\cdot(\lambda}{\Gamma(\alpha +}$$

$$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}\ ,\ \beta = \frac{\sum_i w_i}{\sum_i w_i^2}$$

Argüelles et al. (2019)

$$\int \mathbf{P}(k;\lambda)\cdot \mathbf{G}(\lambda;\alpha,\beta)d\lambda$$

$$\alpha = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} + a\ ,\ \beta = \frac{\sum_i w_i}{\sum_i w_i^2} + b$$

encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$

Statistics of weighted MC

Barlow/Beeston (1993)

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_j p_j \widehat{w}_j \lambda_j)\cdot \prod_j^{N_{\mathrm{src}}} \mathbf{P}(k_{mc,j};\lambda_j)$$

or

$$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\lambda_j{}^*)\cdot \prod_j^{N_{\mathrm{src}}} \mathbf{G}(\lambda_j{}^*; k_{mc,j}+1; \frac{1}{p_j\widehat{w}_j})\cdot p_j\widehat{w}_j$$

Pretty much indistinguishable results

Both allow for prior freedom

encode $\widehat{\mu}(Z)$ and $\widehat{\mathrm{var}}(Z)$

Pros: - often faster
- Limit of large statistics equal to Poisson
- Interpretability
- simplicity

$N_{\mathrm{src}} = N$, absorb p_j

approximate Z, $\mathbf{G} \approx$ ... fixed

Chirkin (2013)

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_i w_i\lambda_i)\cdot \prod_i^{N} \mathbf{P}(1;\lambda_i)$$

or

$$\max_{\{\boldsymbol{\lambda}^*\}} \mathbf{P}(k;\sum_i \lambda_i{}^*)\cdot \prod_i^{N} \mathbf{G}(\lambda_i{}^*;2,\frac{1}{w_i})\cdot w_i$$

encode sum of W_i

Glüsenkamp (2018)

$$\int \mathbf{P}(k;\lambda)\cdot \left[\mathbf{G}(\lambda_1;1,\frac{1}{w_1}) * \ldots * \mathbf{G}(\lambda_N;1,\frac{1}{w_N})\right](\lambda)d\lambda$$

Prob. counterpart

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda) \cdot \left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1) * \ldots * \mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\ d\lambda$$

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda)\cdot\left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1) * \ldots * \mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\ d\lambda \xrightarrow{\textit{Di Salvo´08}} \sim F_D$$

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda) \cdot \left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1) * \ldots * \mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\ d\lambda$$

Di Salvo´08 → $\sim F_D$

↓ *Carlson ´63*

$\sim$ Carlson R_n

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda) \cdot \left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1) * \ldots * \mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\ d\lambda$$

Di Salvo´08 → $\sim F_D$

↓ *Carlson ´63*

$\sim$ Carlson R_n

↓ *Dickey ´82*

$$\sum_{\sum_i k_i = k,\ k_i \geq 0} \prod_i \frac{\Gamma(k_i+\alpha_i)}{k_i! \cdot \Gamma(\alpha_i)} \cdot \beta_i^{\alpha_i} \cdot \left(\frac{1}{1+\beta_i}\right)^{k_i+\alpha_i}$$

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda)\cdot\left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1)*\ldots*\mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\;d\lambda$$

Di Salvo´08 →

$$\sim F_D$$

↓ *Carlson ´63*

$$\sim \text{Carlson } R_n$$

↓ *Dickey ´82*

$$\sum_{\sum_i k_i=k,\;k_i\geq 0}\prod_i \frac{\Gamma(k_i+\alpha_i)}{k_i!\cdot\Gamma(\alpha_i)}\cdot\beta_i^{\alpha_i}\cdot\left(\frac{1}{1+\beta_i}\right)^{k_i+\alpha_i}$$

← *Egorychev Rules ´80s*

$$\sim \frac{1}{2\pi i}\cdot\oint_{\rho=\epsilon}\frac{1}{t^{a-c+1}\cdot\prod_i^N (t-1/z_{+1,i})^{b_{+1,i}}}dt$$

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda)\cdot[\mathrm{G}(\lambda_1;\alpha_1,\beta_1)*\ldots*\mathrm{G}(\lambda_N;\alpha_N,\beta_N)](\lambda)\,d\lambda$$

Di Salvo´08 → $\sim F_D$

$$D_k(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{k}\sum_{j=1}^{k}\left[\left(\sum_{i=1}^{N}\alpha_i\cdot\frac{1}{1+\beta_i}^{j}\right)D_{k-j}\right] \quad \text{and } D_0 = 1$$

Carlson ´63

$\sim$ Carlson R_n

Dickey ´82

$$\sum_{\sum_i k_i=k,\ k_i\geq 0}\prod_i \frac{\Gamma(k_i+\alpha_i)}{k_i!\cdot\Gamma(\alpha_i)}\cdot\beta_i^{\alpha_i}\cdot\left(\frac{1}{1+\beta_i}\right)^{k_i+\alpha_i}$$

Egorychev Rules ´80s

$$\sim\frac{1}{2\pi i}\cdot\oint_{\rho=\epsilon}\frac{1}{t^{a-c+1}\cdot\prod_i^N(t-1/z_{+1,i})^{b_{+1,i}}}dt$$

How can we calculate this integral

$$\int_0^\infty \mathrm{P}(k;\lambda)\cdot\left[\mathrm{G}(\lambda_1;\alpha_1,\beta_1) * \ldots * \mathrm{G}(\lambda_N;\alpha_N,\beta_N)\right](\lambda)\; d\lambda$$

Di Salvo´08 → $\sim F_D$

$$D_k(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{k}\sum_{j=1}^{k}\left[\left(\sum_{i=1}^{N}\alpha_i\cdot\frac{1}{1+\beta_i}^{\,j}\right)D_{k-j}\right] \quad \text{and } D_0 = 1$$

$\sim F_D$ → *Carlson ´63* → $\sim$ Carlson R_n

Exact n-th B-Spline Moment w/ arbitrary knots

Legendre Polynomials and others

....

...

Dickey ´82

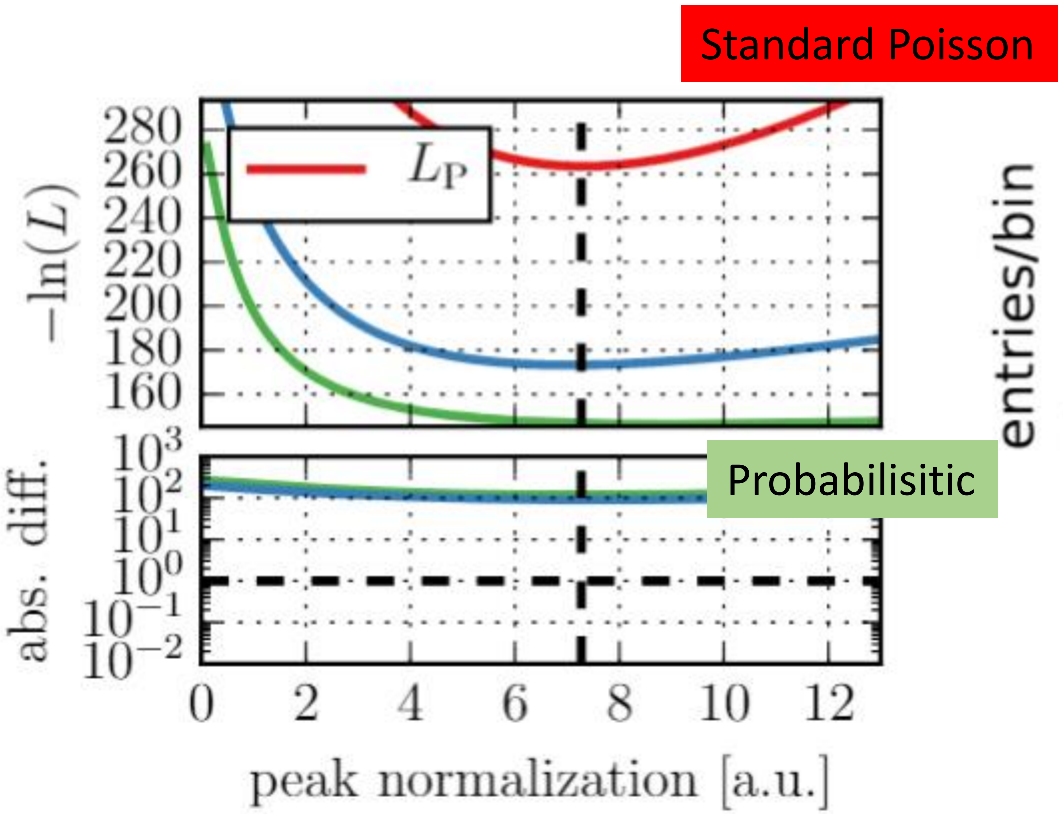
$\sim \frac{1}{2\pi}$ … dt

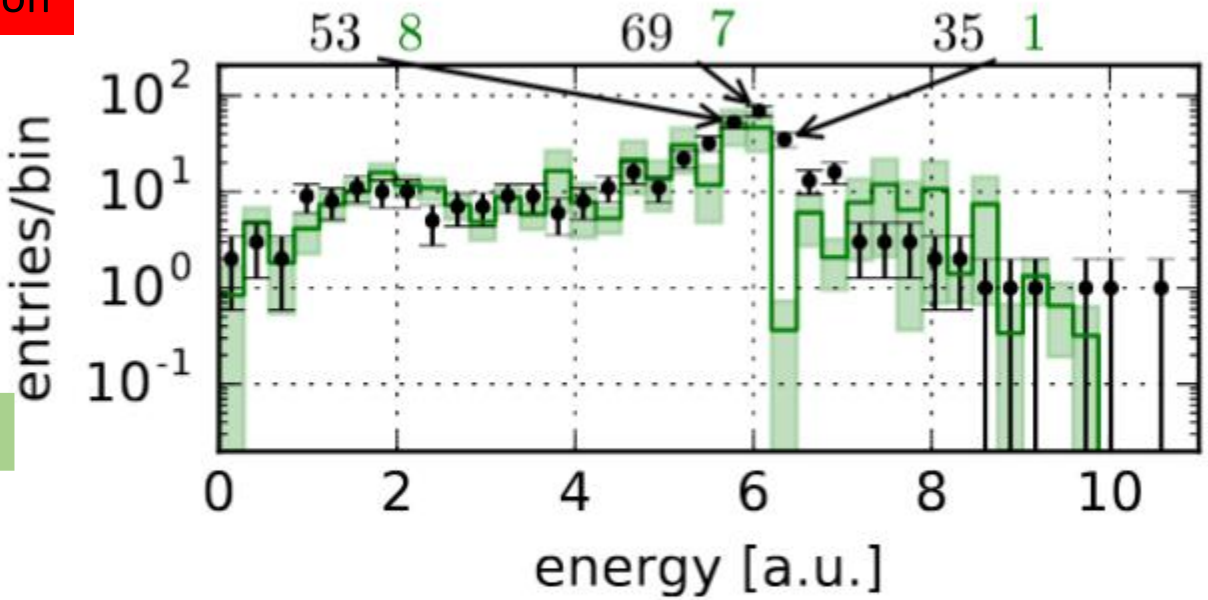
The problem of limited Monte Carlo is connected to these other areas in statistics and special functions

Egorychev Rules ´80s

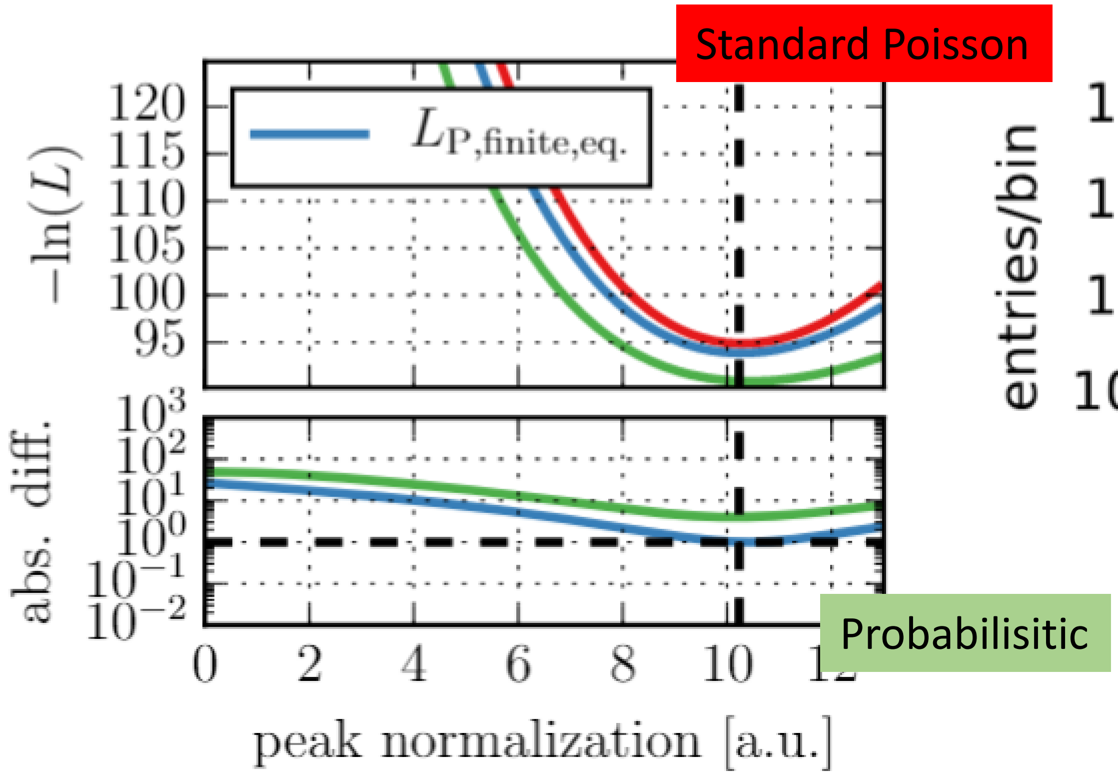
$$\sum_{\sum_i k_i = k,\; k_i \geq 0}\;\prod_i \frac{\Gamma(k_i+\alpha_i)}{k_i!\cdot\Gamma(\alpha_i)}\cdot\beta_i^{\alpha_i}\cdot\left(\frac{1}{1+\beta_i}\right)^{k_i+\alpha_i}$$

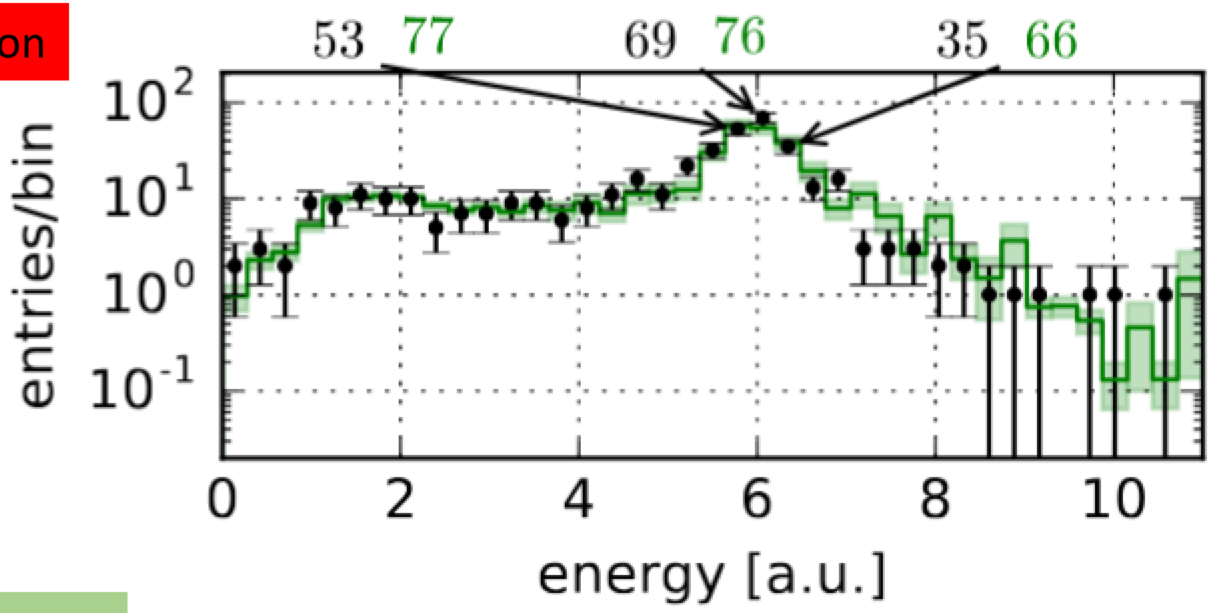
Comparisons



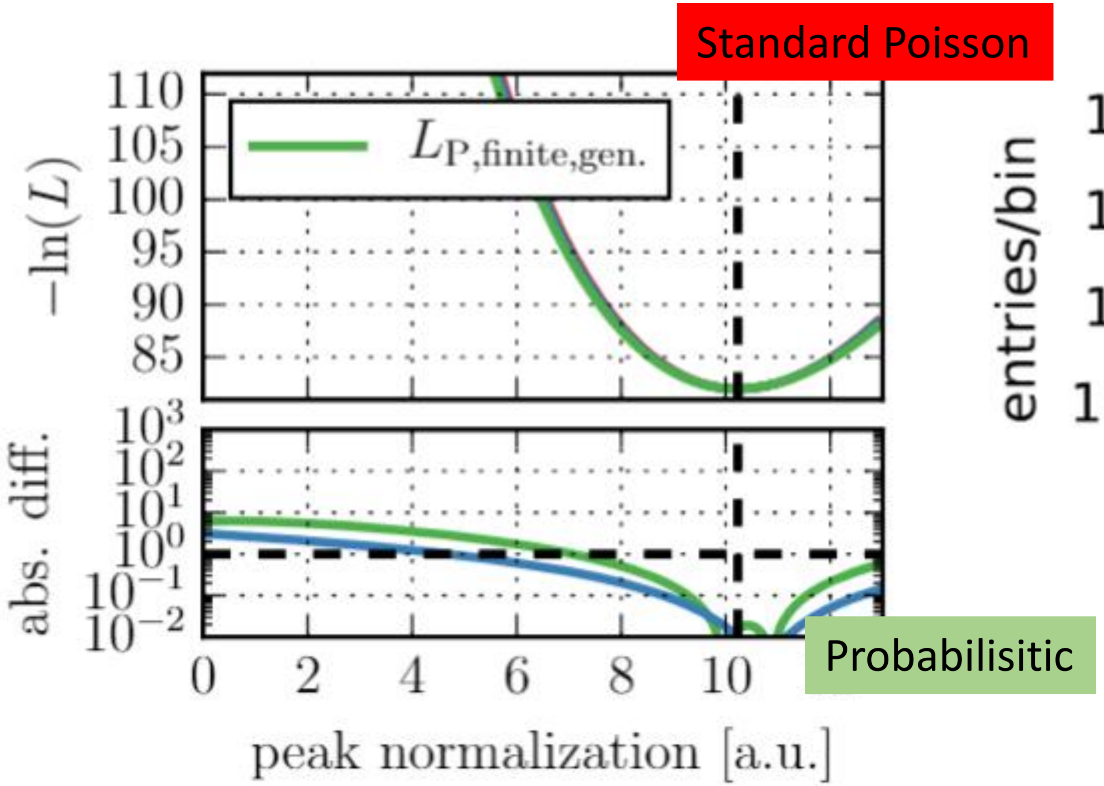


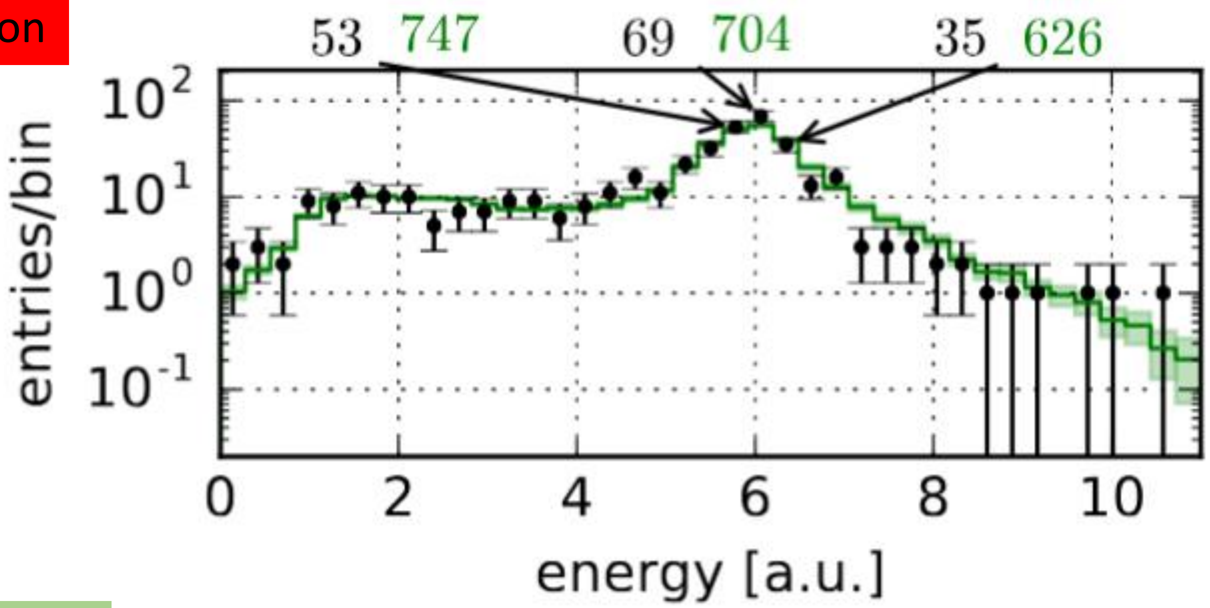
Comparisons



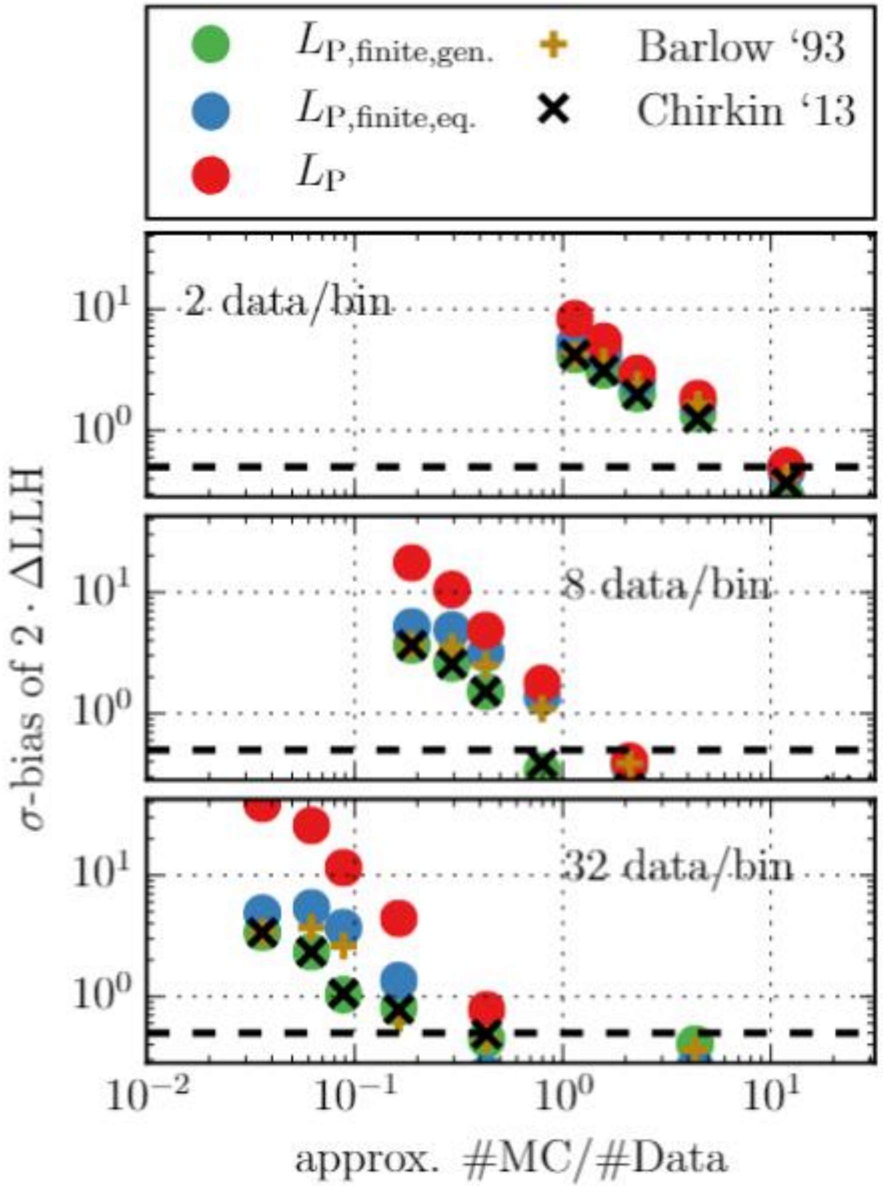


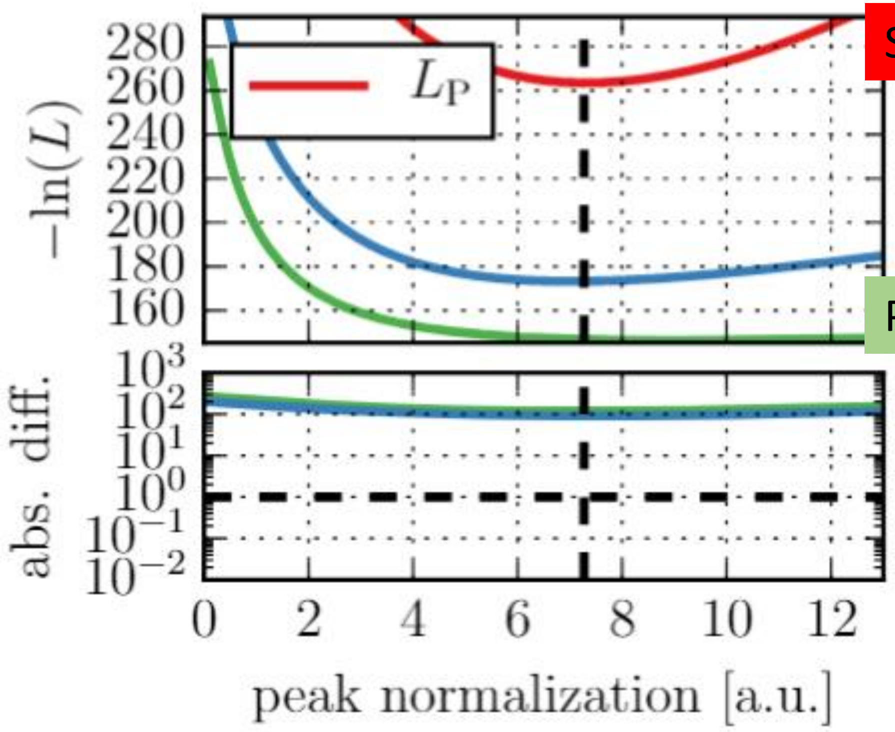
Comparisons





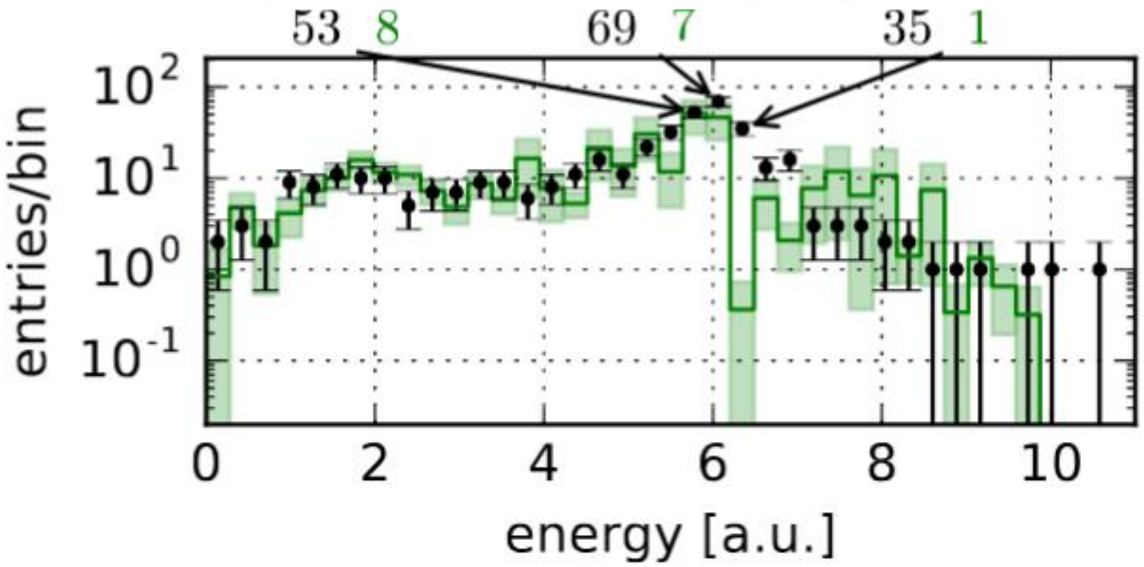
Comparisons



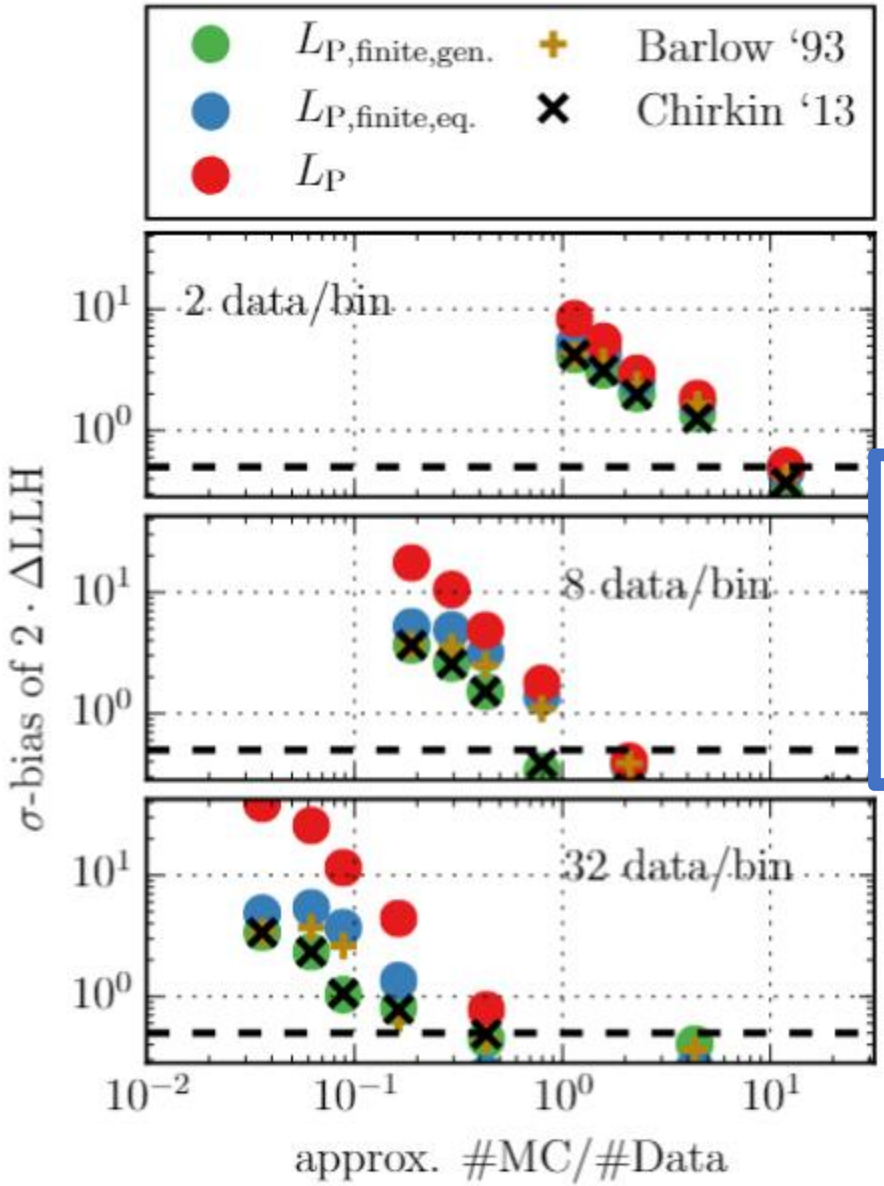


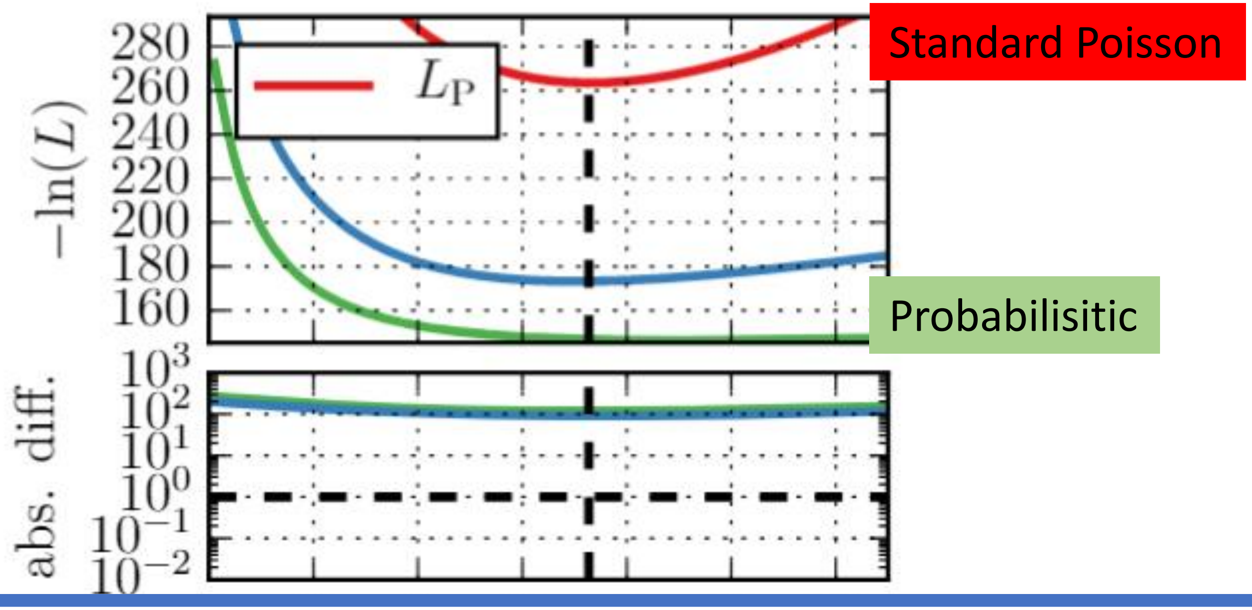
Standard Poisson

Probabilisitic



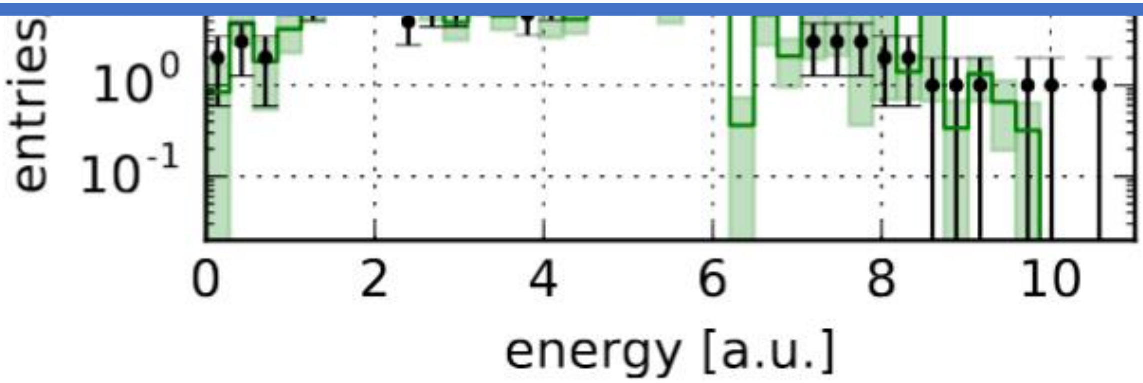
Comparisons



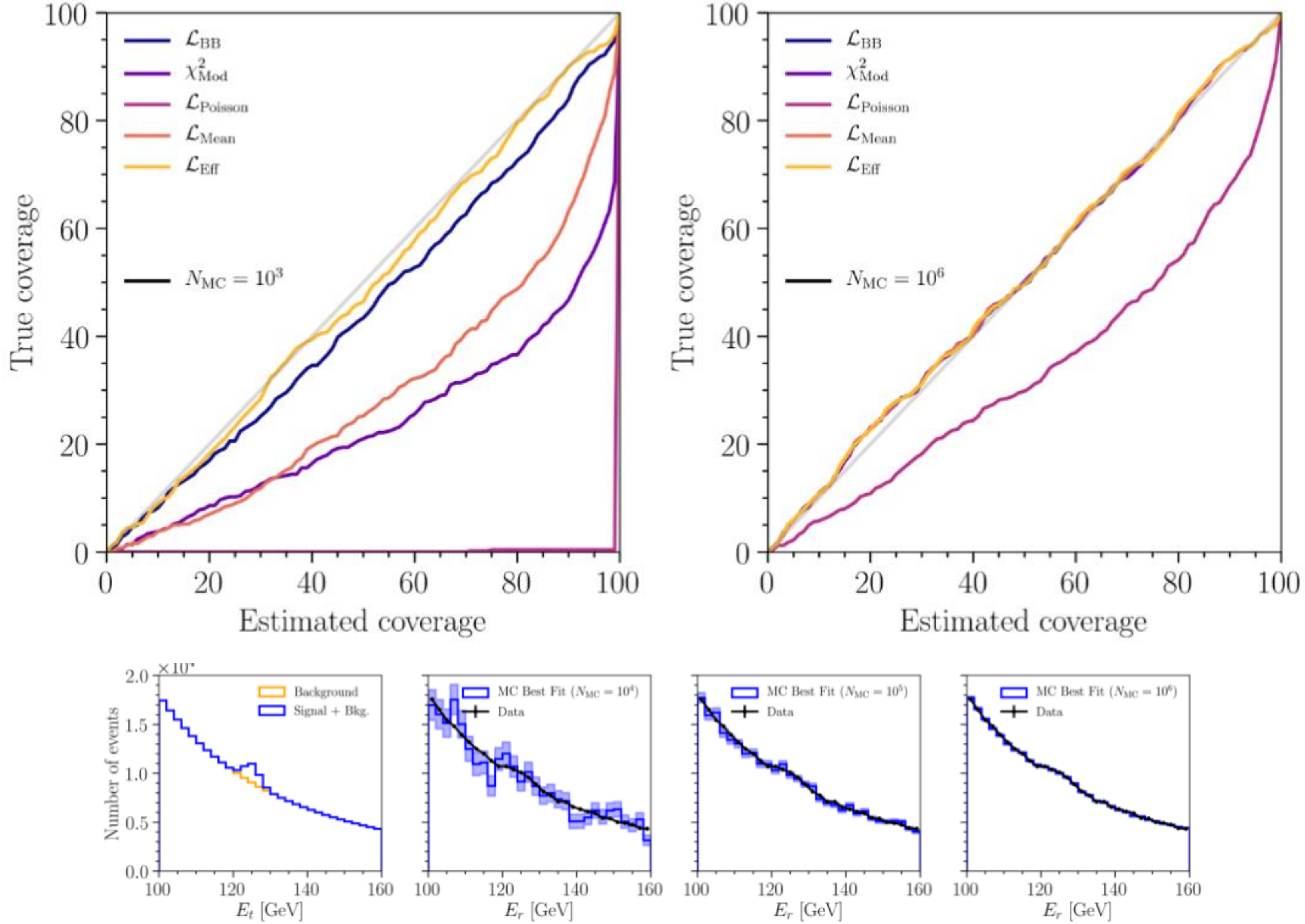


Not a surprise: We have 1 weighted source dataset, Barlow/Beeston averages the weights

Not a realistic situation

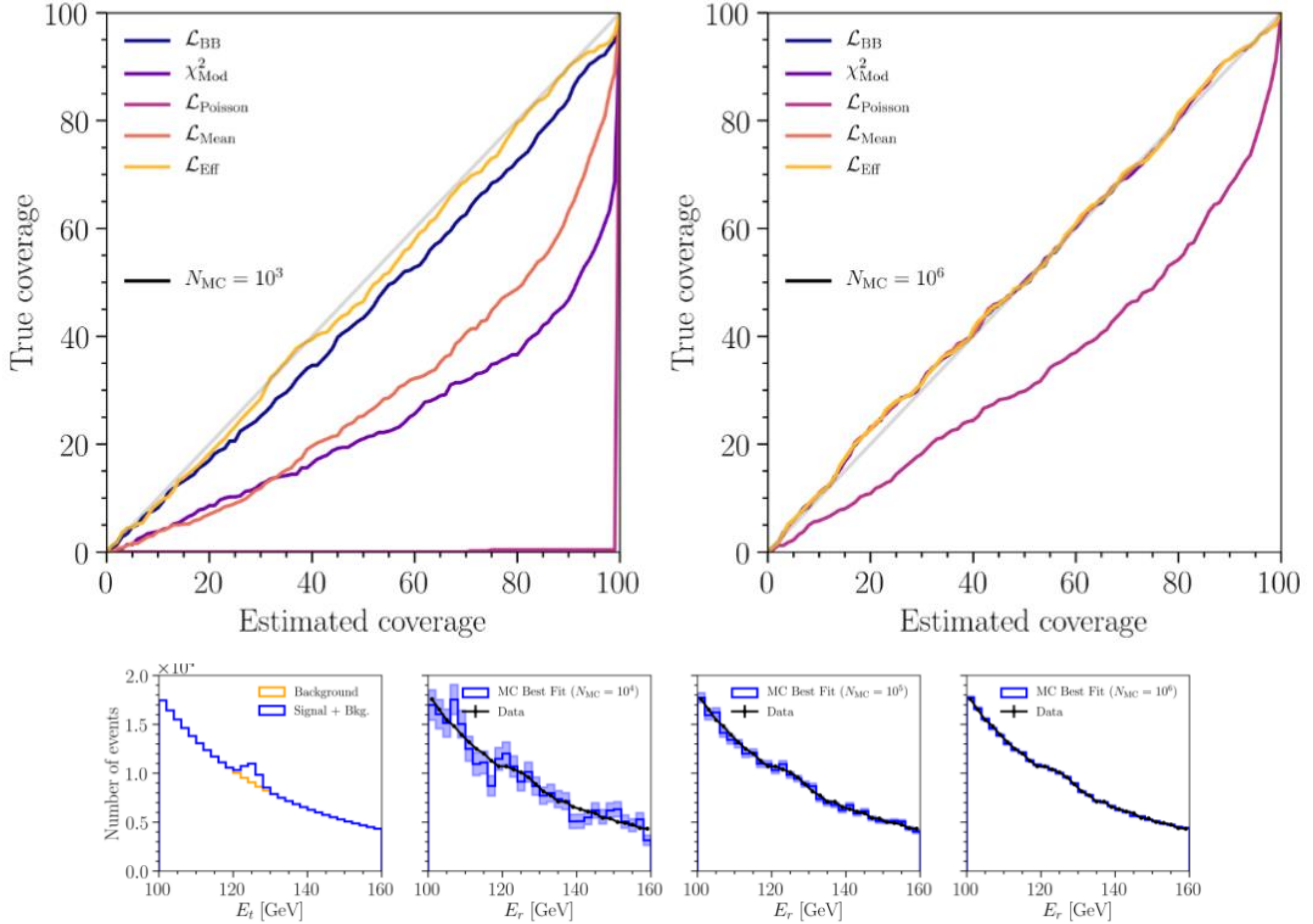


Comparisons



1901.04645

Comparisons



Also works well
with 2 sources (signal+bg)

BUT:
-Requires usage of different hyper parameters (Tuning) "$\mathcal{L}_{Eff}$"/"$\mathcal{L}_{Mean}$"

Signal+Background increase statistics simultaneously
-> also not realistic

At other statistical levels "$\mathcal{L}_{Mean}$" can be better than "$\mathcal{L}_{Eff}$"

1901.04645

We need further generalizations and more tests…

Generalization (1)

$$\int \mathbf{P}(k;\lambda) \cdot [\mathbf{GPG}_1 * \ldots * \mathbf{GPG}_N](\lambda) d\lambda$$

Generalization (2)

$$\int \mathbf{P}(k;\lambda) \cdot [\mathbf{G}_1 * \ldots * \mathbf{G}_{N_{src}}](\lambda) d\lambda$$

Generalization (3)

$$\int \mathbf{P}(k;\lambda) \cdot [\mathbf{GG}_1 * \ldots * \mathbf{GG}_{N_{src}}](\lambda) d\lambda$$

To appear

We need further generalizations and more tests…

Generalization (1)

$$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{GPG}_1 * \ldots * \mathbf{GPG}_N\right](\lambda)d\lambda$$

→ Interpretation 1: Apprixmate W_i

Tries to model the CPD better (as CPGD)

Generalization (2)

$$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{G}_1 * \ldots * \mathbf{G}_{N_{src}}\right](\lambda)d\lambda$$

Generalization (3)

$$\int \mathbf{P}(k;\lambda)\cdot\left[\mathbf{GG}_1 * \ldots * \mathbf{GG}_{N_{src}}\right](\lambda)d\lambda$$

→ Interpretation 2: Apprixmate CPDs directly

Direct Counterpart of Barlow/Beeston
Model individual source datasets

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k;\sum_j p_j \widehat{w}_j \lambda_j)\cdot\prod_j^{N_{\mathrm{src}}} \mathbf{P}(k_{mc,j};\lambda_j)$$

All of these can be exactly calculated!

To appear

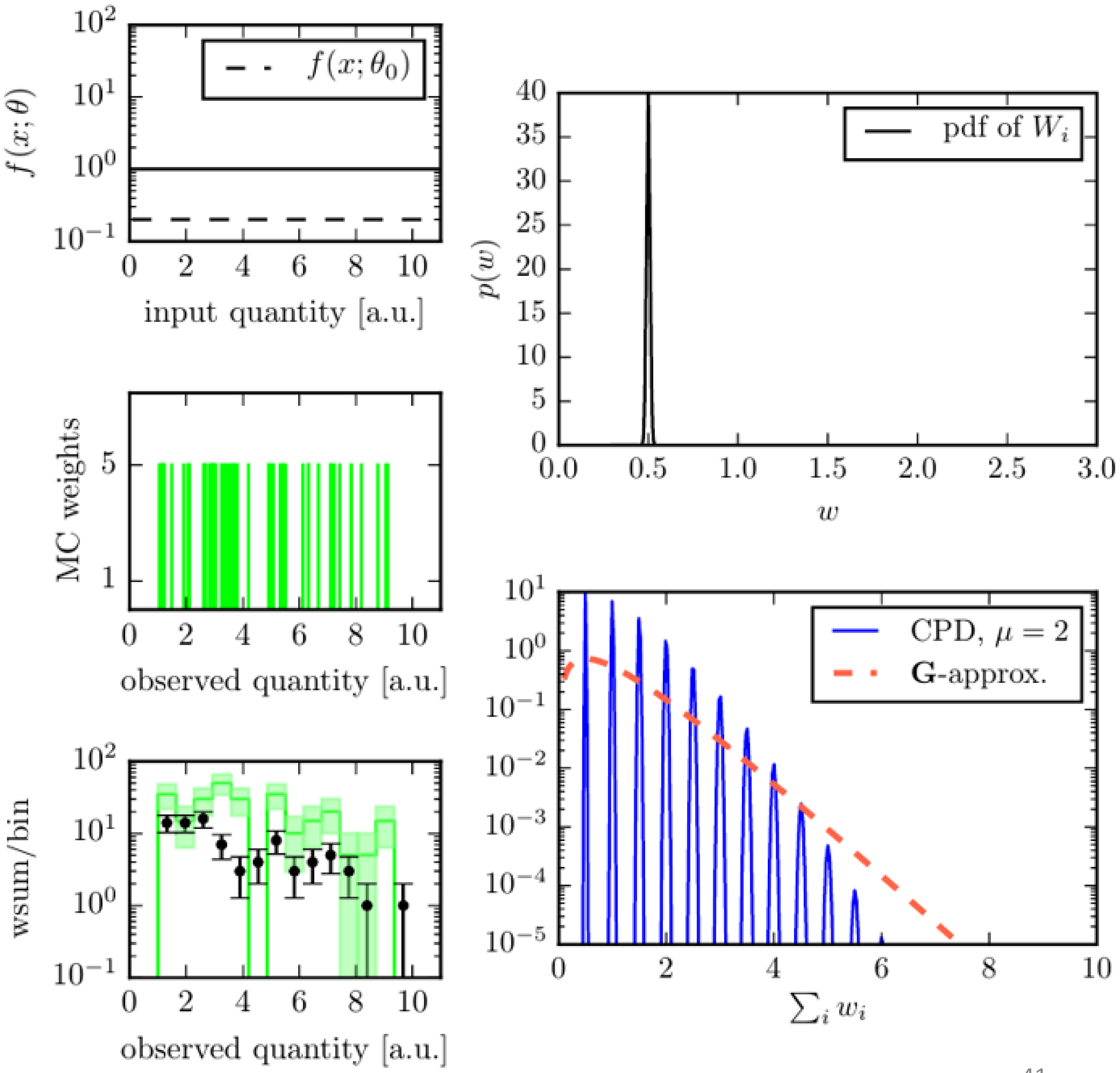
First test: Equal weights

$$L_{bin,exact} = \int \frac{e^{-\lambda} \cdot \lambda^k}{k!} \cdot p_{CPD}(\lambda) d\lambda$$

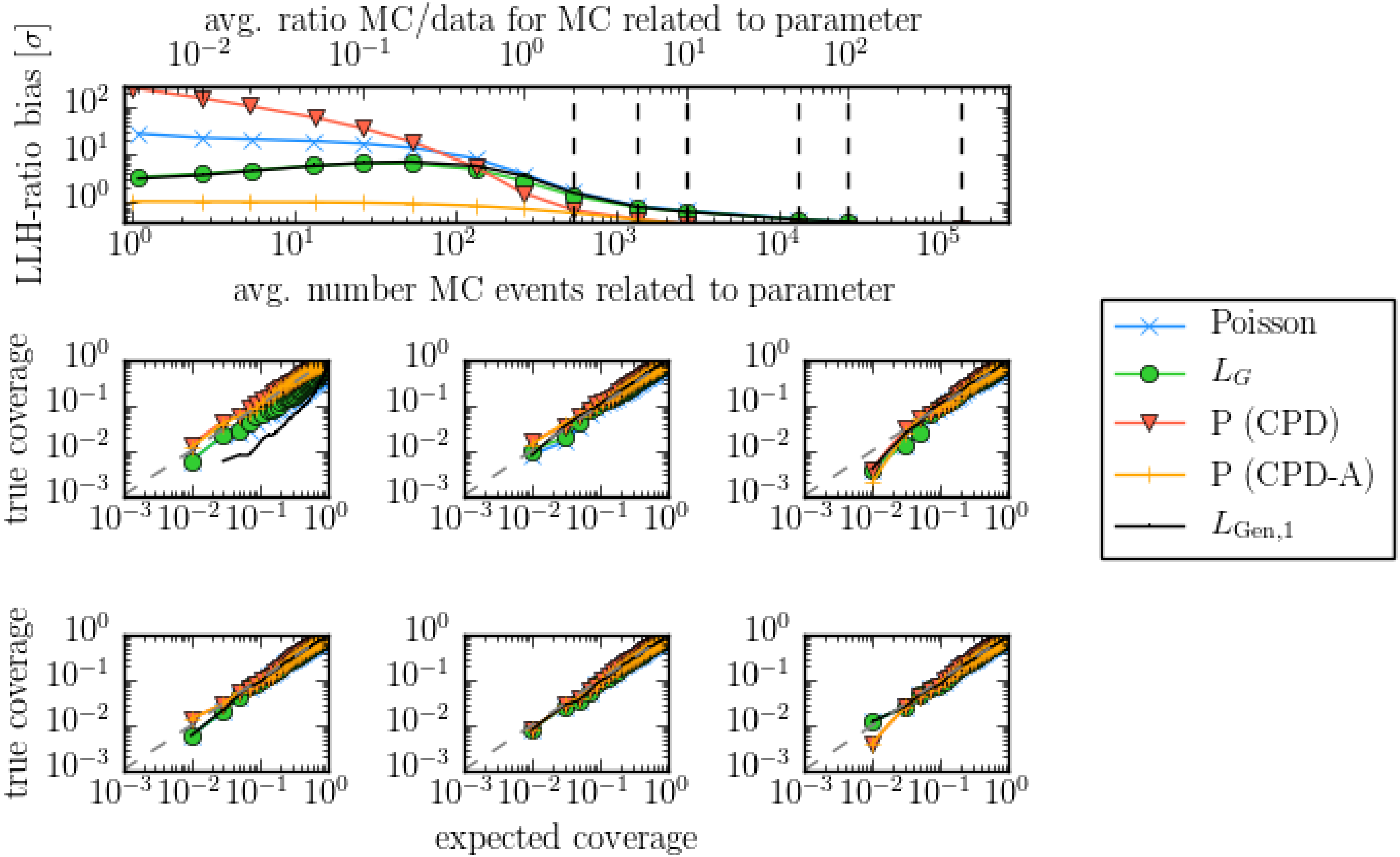
$$= \int \frac{e^{-\lambda} \cdot \lambda^k}{k!} \cdot \sum_{k_{mc}=0}^{\infty} \frac{e^{-\mu} \cdot \mu^{k_{mc}}}{k_{mc}!} \cdot \delta(\lambda - k_{mc} \cdot w) d\lambda$$

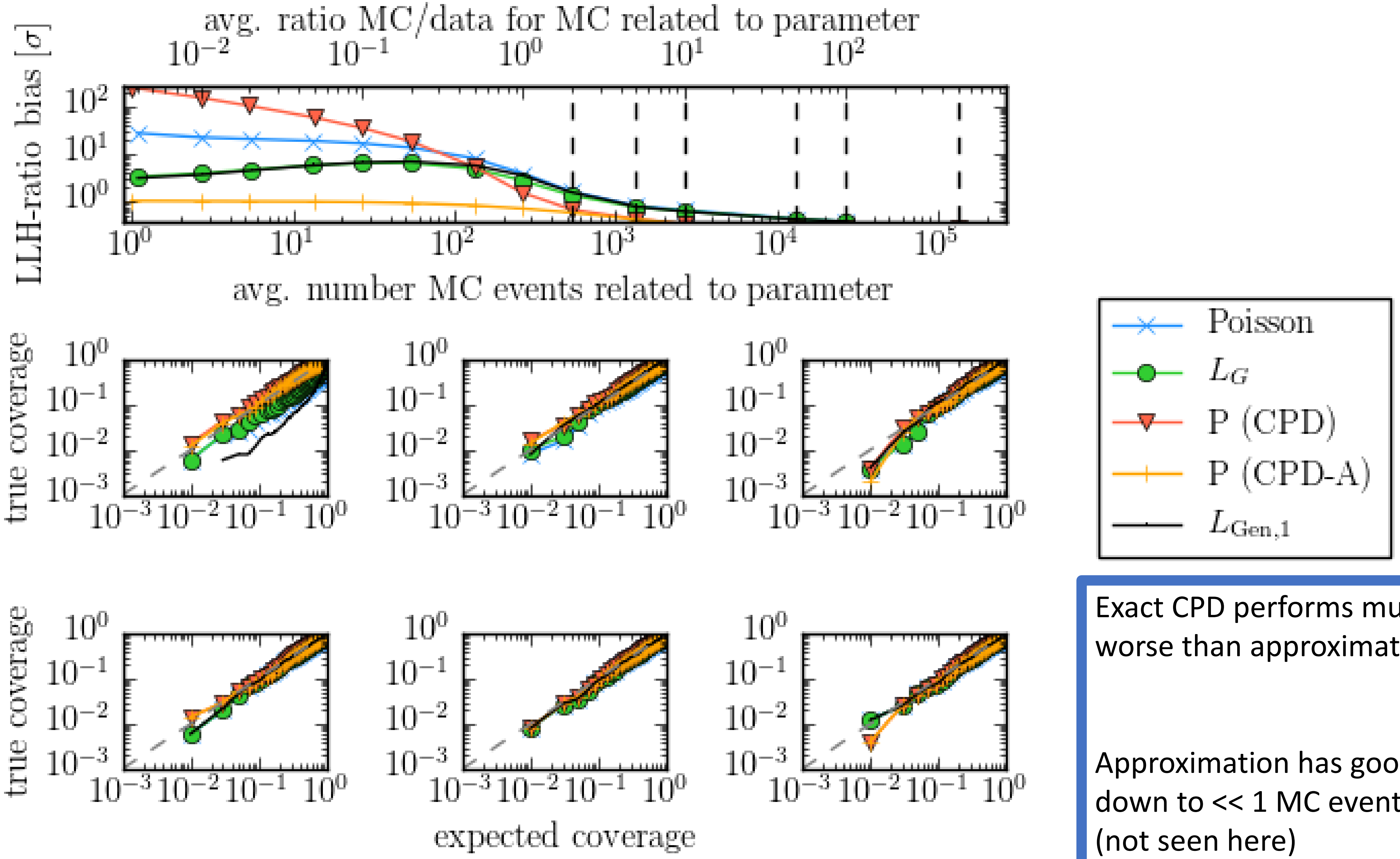
$$= \sum_{k_{mc}=0}^{\infty} \frac{e^{-k_{mc} w} \cdot (k_{mc} w)^k}{k!} \cdot \frac{e^{-\mu} \cdot \mu^{k_{mc}}}{k_{mc}!}$$

We can calculate the CPD exactly
For equal weights, including μ



To appear

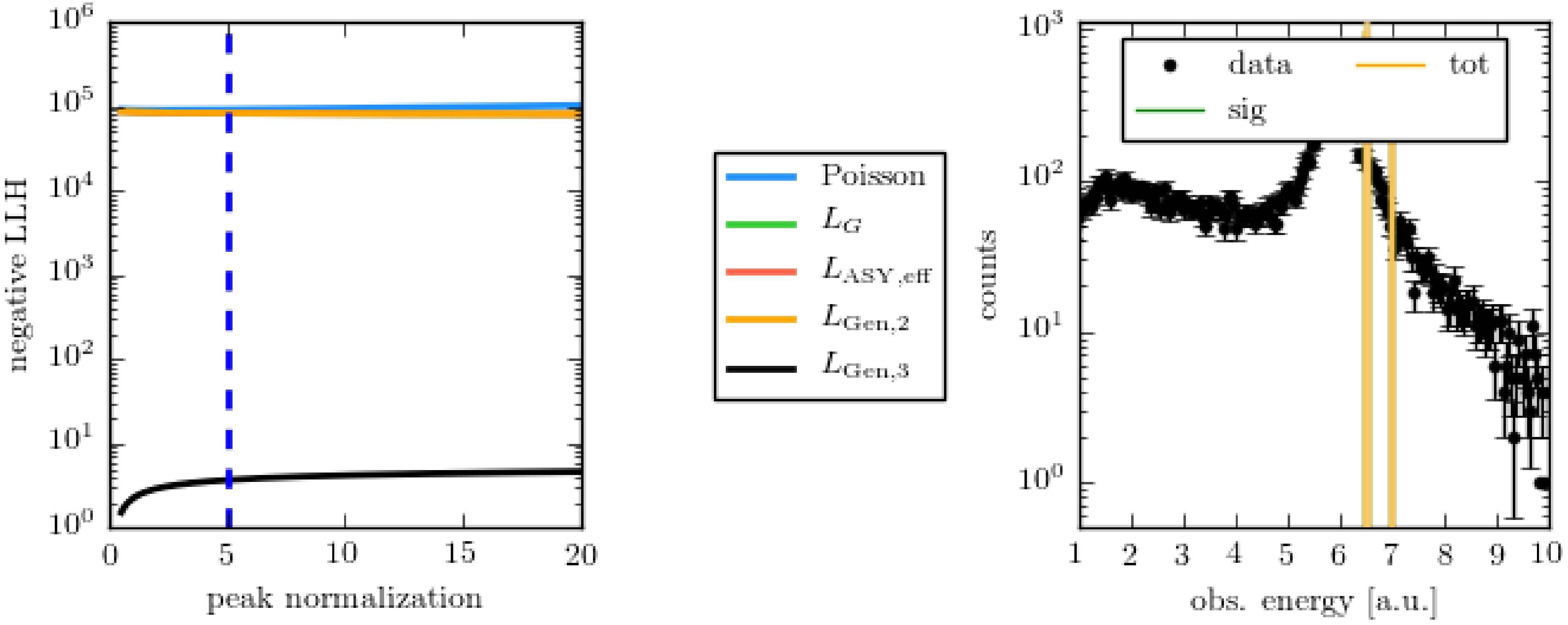

avg. ratio MC/data for MC related to parameter
LLH-ratio bias [σ]
avg. number MC events related to parameter
true coverage
expected coverage
Poisson
L_G
P (CPD)
P (CPD-A)
$L_{\mathrm{Gen},1}$



Exact CPD performs much worse than approximation

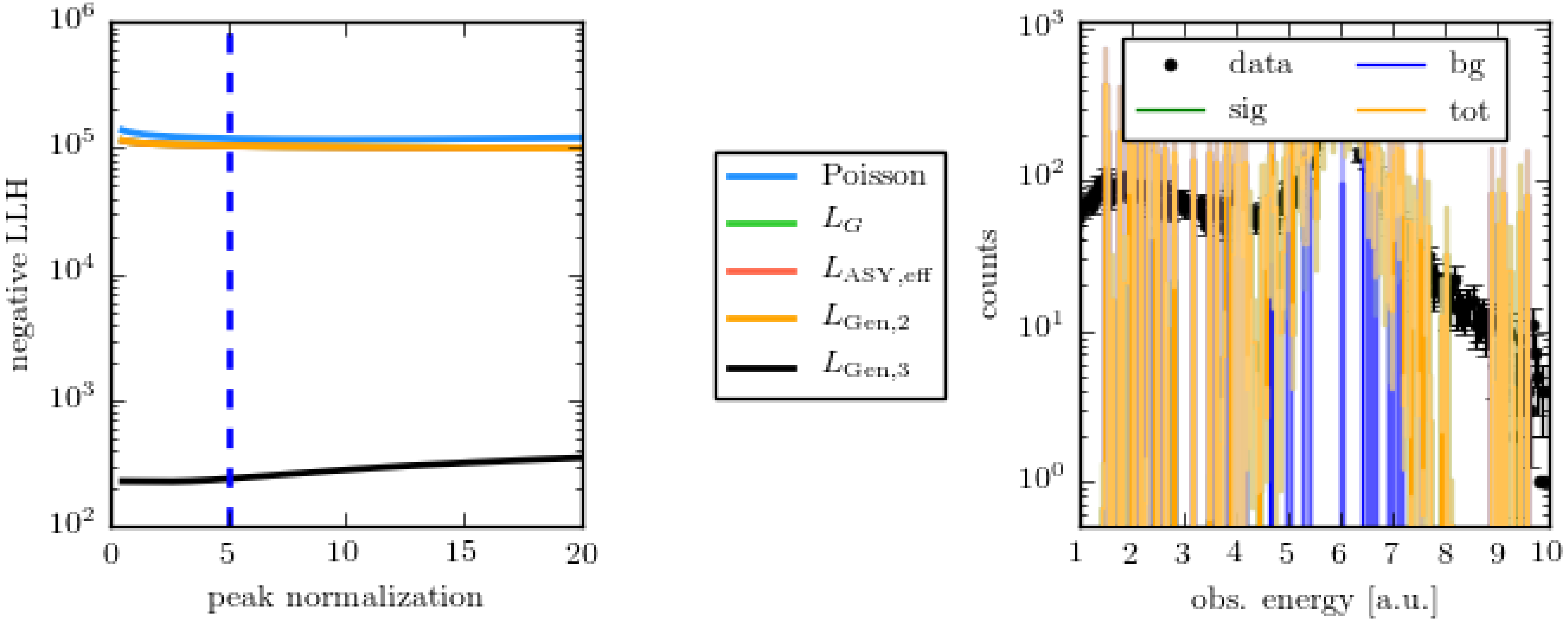
Approximation has good coverage down to << 1 MC event / bin (not seen here)

2nd Test: Increse statistics in both sig/bg



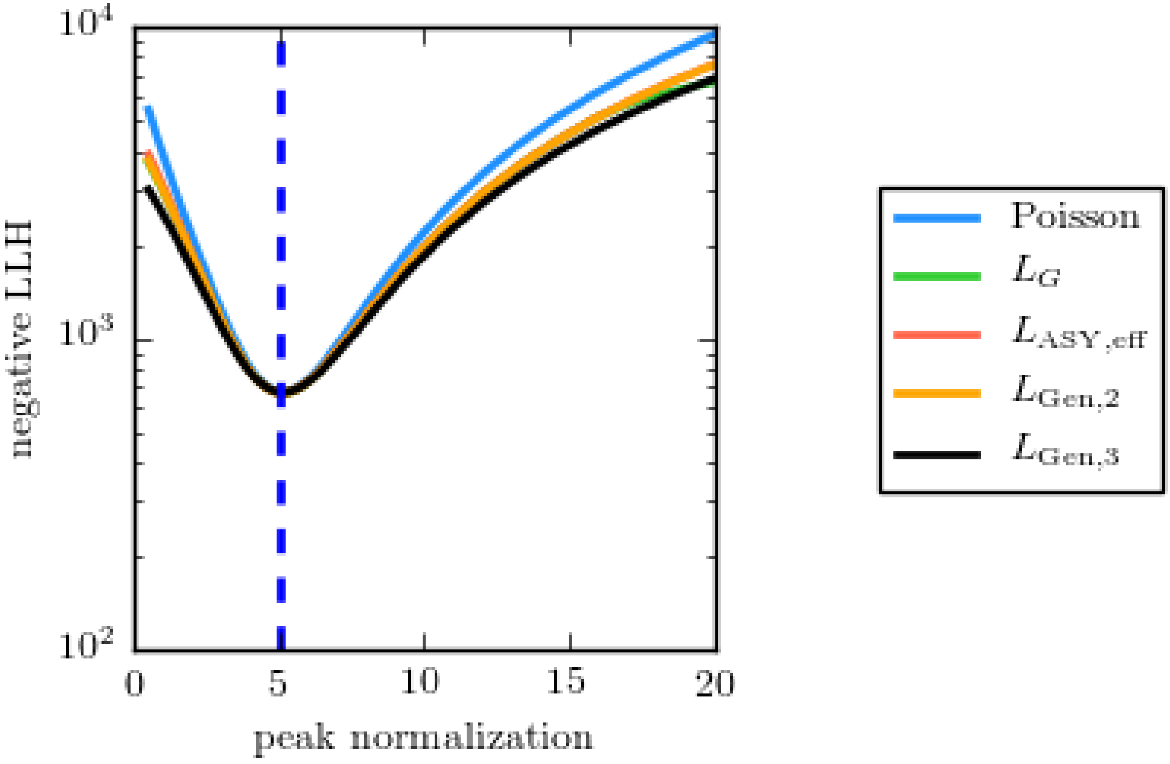
To appear

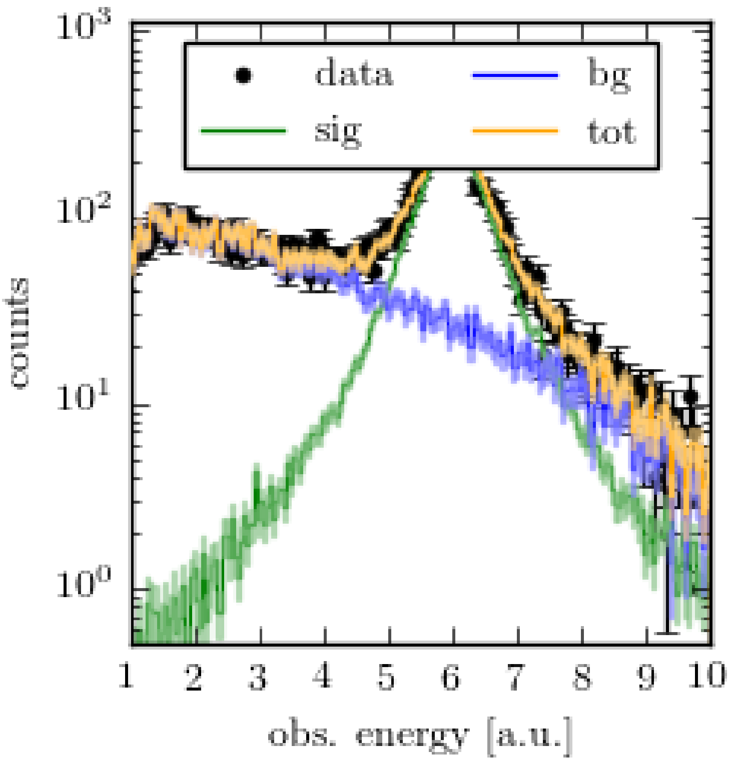
2nd Test: Increse statistics in both sig/bg



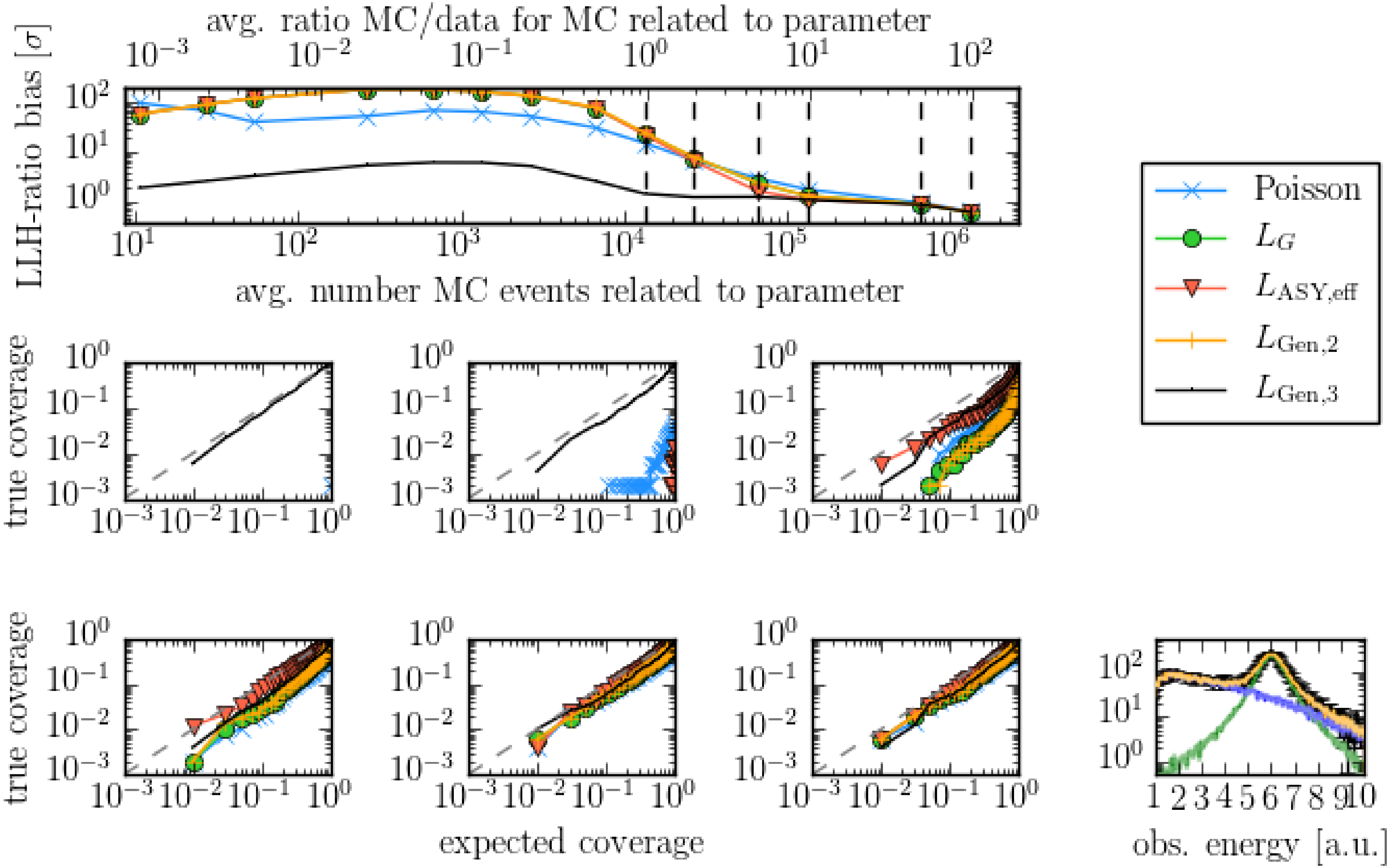
To appear

2nd Test: Increse statistics in both sig/bg

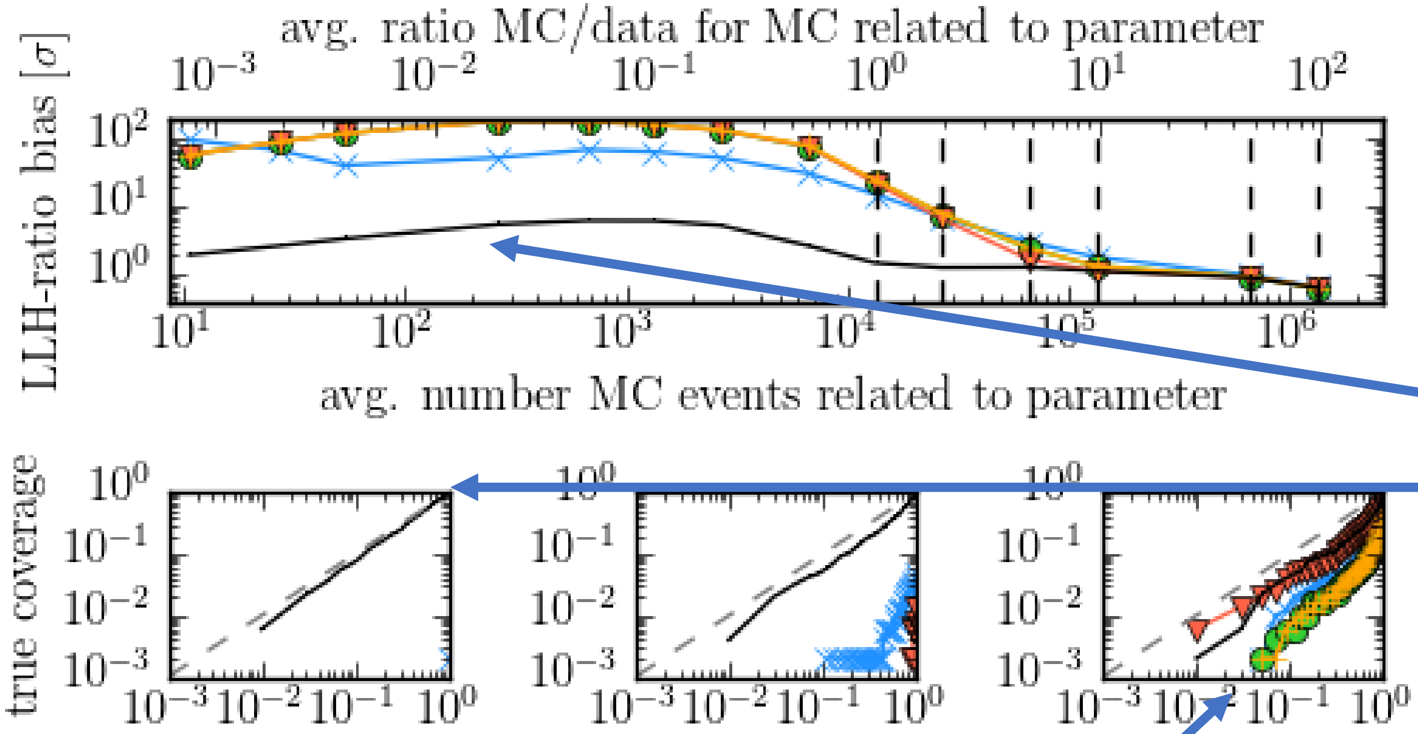


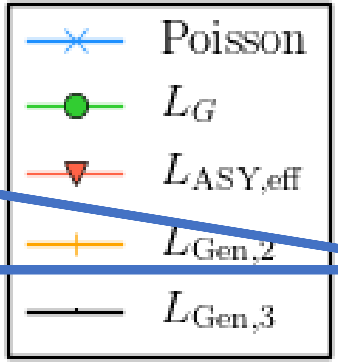


To appear


avg. ratio MC/data for MC related to parameter
LLH-ratio bias [σ]
avg. number MC events related to parameter
true coverage
expected coverage
obs. energy [a.u.]
Poisson
L_G
$L_{\text{ASY,eff}}$
$L_{\text{Gen},2}$
$L_{\text{Gen},3}$

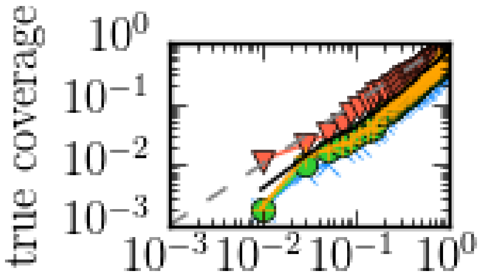
To appear

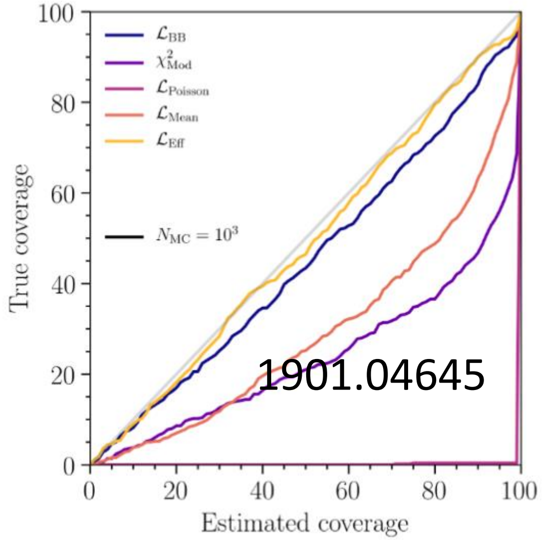




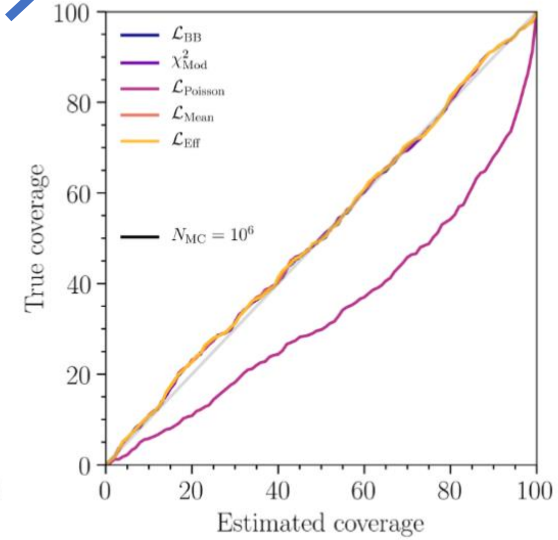
Including uncertainty about number of events greatly reduces bias

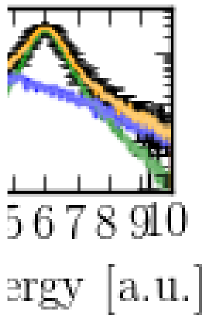
Similar behavior of "$\mathcal{L}_{Eff}$
(forgot Barlow here)





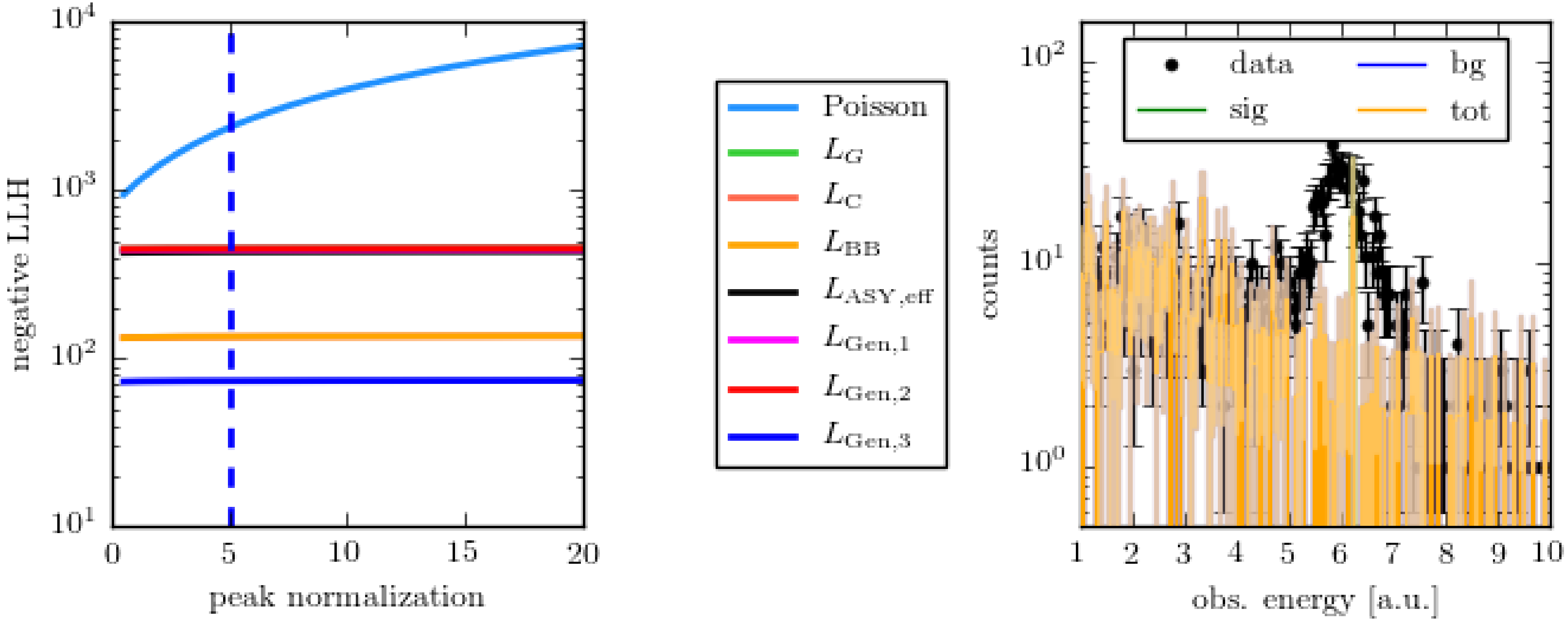
1901.04645



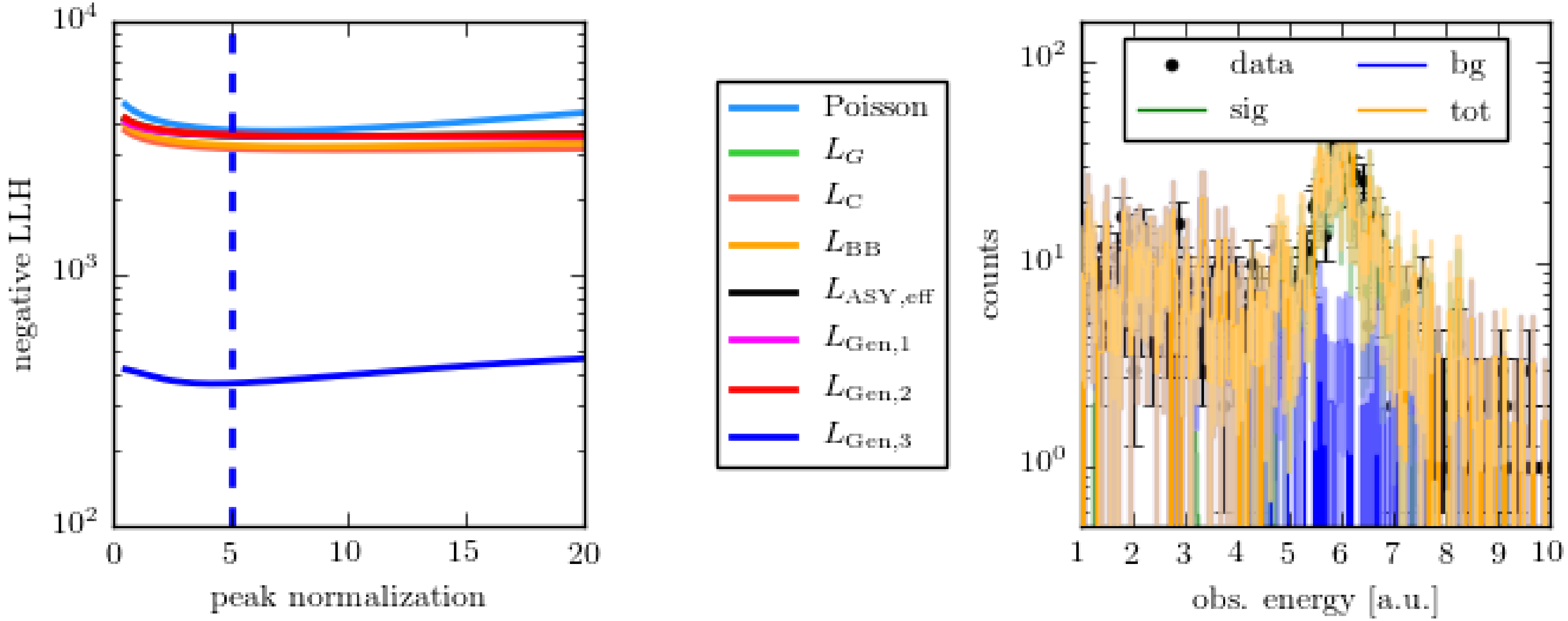


To appear

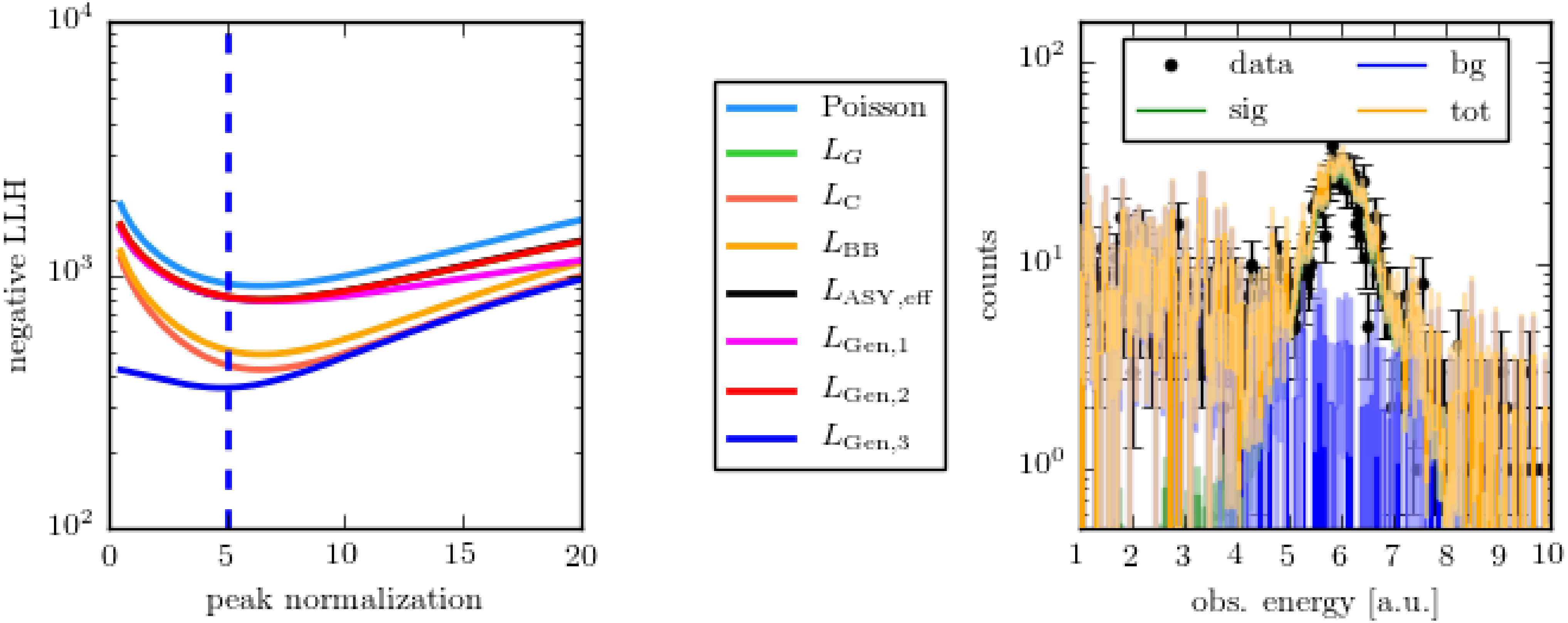
Test 3: Background statistics is limited

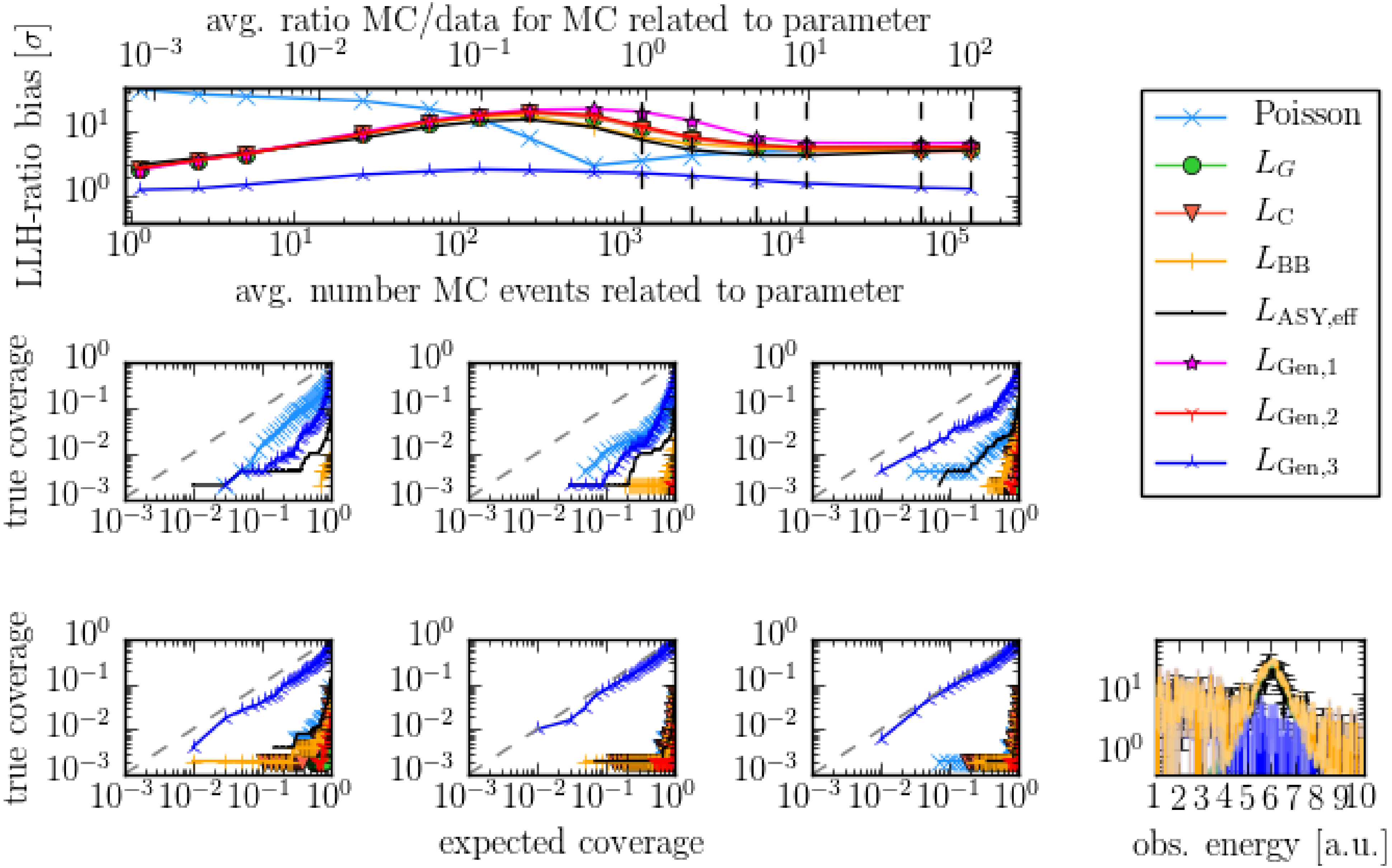


Test 3: Background statistics is limited

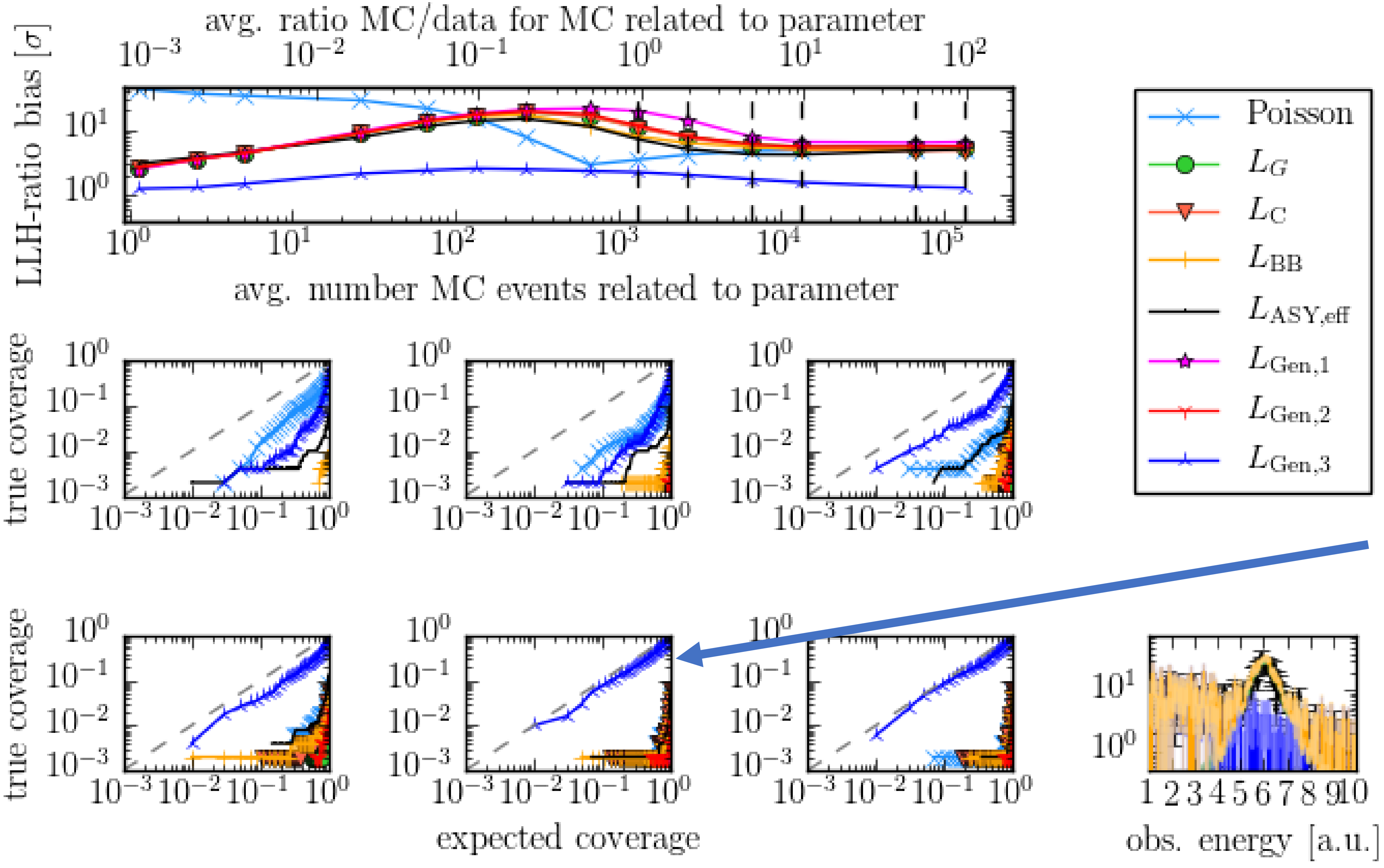


Test 3: Background statistics is limited





avg. ratio MC/data for MC related to parameter
LLH-ratio bias [σ]
avg. number MC events related to parameter
Poisson
L_G
L_C
L_{BB}
$L_{ASY,eff}$
$L_{Gen,1}$
$L_{Gen,2}$
$L_{Gen,3}$
true coverage
true coverage
expected coverage
obs. energy [a.u.]



Generalization 3
seems to be the
Only approach
to handle the
Limited background

Summary

- All approaches approximate the CPD + integrate over Poisson mean
  either with nuisance optimization **or** via integration
- exact CPD /equal weights (scaled Poisson) behaves badly in likelihood scans … probably because of multimodality?
- Some advantages of probabilistic approaches:
  Interpretability, simplicity, convergence to Poisson as $n_{MC} \rightarrow \infty$
- There is now a precise probabilistic counterpart of Barlow/Beeston

$$\max_{\{\boldsymbol{\lambda}\}} \mathbf{P}(k; \sum_j p_j \widehat{w}_j \lambda_j) \cdot \prod_j^{N_{\mathrm{src}}} \mathbf{P}(k_{mc,j}; \lambda_j) \longrightarrow \int \mathbf{P}(k; \lambda) \cdot \left[\mathbf{GG}_1 * \ldots * \mathbf{GG}_{N_{src}}\right](\lambda) d\lambda$$

(all of this will be on arXiv in a couple of days)

Useful links

- Barlow et al 93 https://www.sciencedirect.com/science/article/pii/001046559390005W
- Bohm/Zech 2012 https://www.sciencedirect.com/science/article/pii/S0168900212006705?via%3Dihub
- Chirkin 2013 https://arxiv.org/abs/1304.0735
- Glüsenkamp 2018 https://arxiv.org/abs/1712.01293
- Argüelles et al 2019 https://arxiv.org/abs/1901.04645

- Code for probabilistic likelihood implementations (c++/python):
  https://github.com/thoglu/mc_uncertainty
  (will be updated in next days with new formulas)