A critical study of the Monte Carlo replica method

Mark N. Costantini

DIS 2024

arXiv:2404.xxxxx, MNC, M. Madigan, L. Mantani, J. Moore

8-12 April 2024, Grenoble, France

UNIVERSITY OF CAMBRIDGE

Funded by the European Union

European Research Council
Established by the European Commission

PBSP

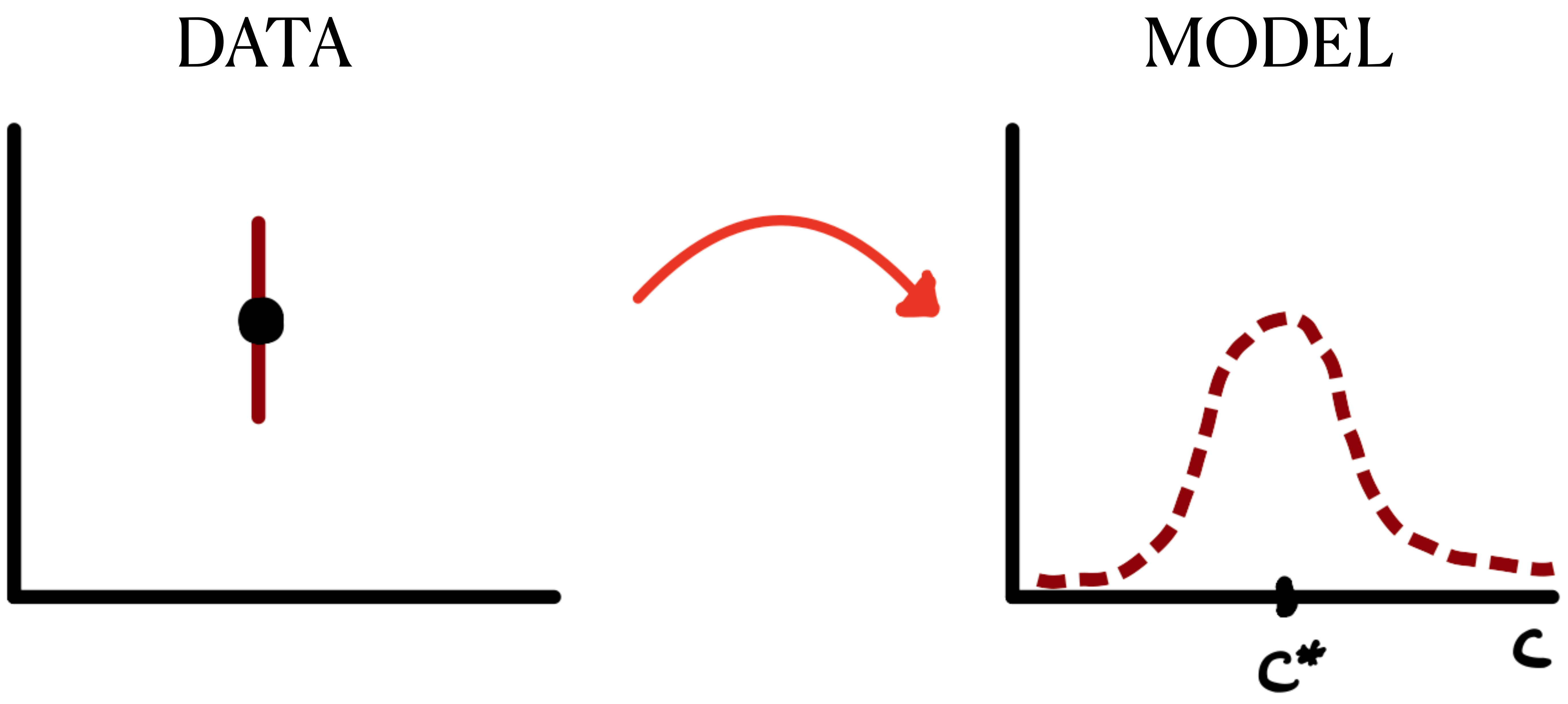
Outline

- Introduction: estimation of uncertainties on parameters
  - Bayesian credible regions
  - Monte Carlo replica method
- The Monte Carlo replica posterior
  - Mathematical derivation
  - Toy models
- Applications in high energy physics
  - SMEFT Fits
  - PDF Fits
- Summary and outlook

Introduction

Introduction

Setup of the problem



X : observed data

$X \sim \mathbb{P}$: Data generating process

$t[\boldsymbol{c}]$: Modelling of the data

What is a plausible range of values for the parameter(s) $\boldsymbol{c}$?

Introduction

Bayesian approach: credible regions

$\boldsymbol{c}$: Random variable

$$\pi(\boldsymbol{c} \,|\, X) \propto \pi(\boldsymbol{c})\; L(X \,|\, \boldsymbol{c})$$

Posterior distribution

Prior distribution

Likelihood

Credible region:

A $(100 \cdot \gamma)\,\%$ credible region $(\boldsymbol{A}, \boldsymbol{B})$ has $\pi(\boldsymbol{A} \leq \boldsymbol{c} \leq \boldsymbol{B} \,|\, X) = \gamma$

Introduction

Monte Carlo replica method approach

use the knowledge of $\mathbb{P}$ to generate N (pseudo) data samples

$$X_1, \ldots, X_N \sim \mathbb{P}$$

for each of the samples we compute an estimate of the parameter

$$\{\boldsymbol{c}_1^*, \ldots, \boldsymbol{c}_N^*\}$$

e.g. $\boldsymbol{c}_i^* = \operatorname{argmin}_{\boldsymbol{c}} \log(L(X_i \mid \boldsymbol{c}))$

use this collection of samples as an approximation of the posterior

The Monte Carlo Posterior

Monte Carlo replica method

Suppose that experimental data is distributed according to a multivariate normal

$$\boldsymbol{d} \sim \mathcal{N}(\boldsymbol{t}(\boldsymbol{c}), \Sigma), \;\; \Sigma \in \mathbb{R}^{N_{\text{dat}} \times N_{\text{dat}}}, \;\; \boldsymbol{t} : \mathbb{R}^{N_{\text{param}}} \to \mathbb{R}^{N_{\text{dat}}}$$

Bayesian posterior

$$P_B(\boldsymbol{c} \,|\, \boldsymbol{d}_0) \propto \pi(\boldsymbol{c}) \;\exp\left(-\frac{1}{2}\chi^2(\boldsymbol{t}(\boldsymbol{c}), \boldsymbol{d}_0)\right)$$

Monte Carlo replica method:

- Pseudodata distribution $\boldsymbol{d}_p \sim \mathcal{N}(\boldsymbol{d}_0, \Sigma)$
- “Best fit parameter"
  $\boldsymbol{c}_p(\boldsymbol{d}_p) = \text{argmin}_{\boldsymbol{c}}(\boldsymbol{d}_p - \boldsymbol{t}(\boldsymbol{c}))^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{t}(\boldsymbol{c}))$

How is $\boldsymbol{c}_p(\boldsymbol{d}_p)$ distributed?

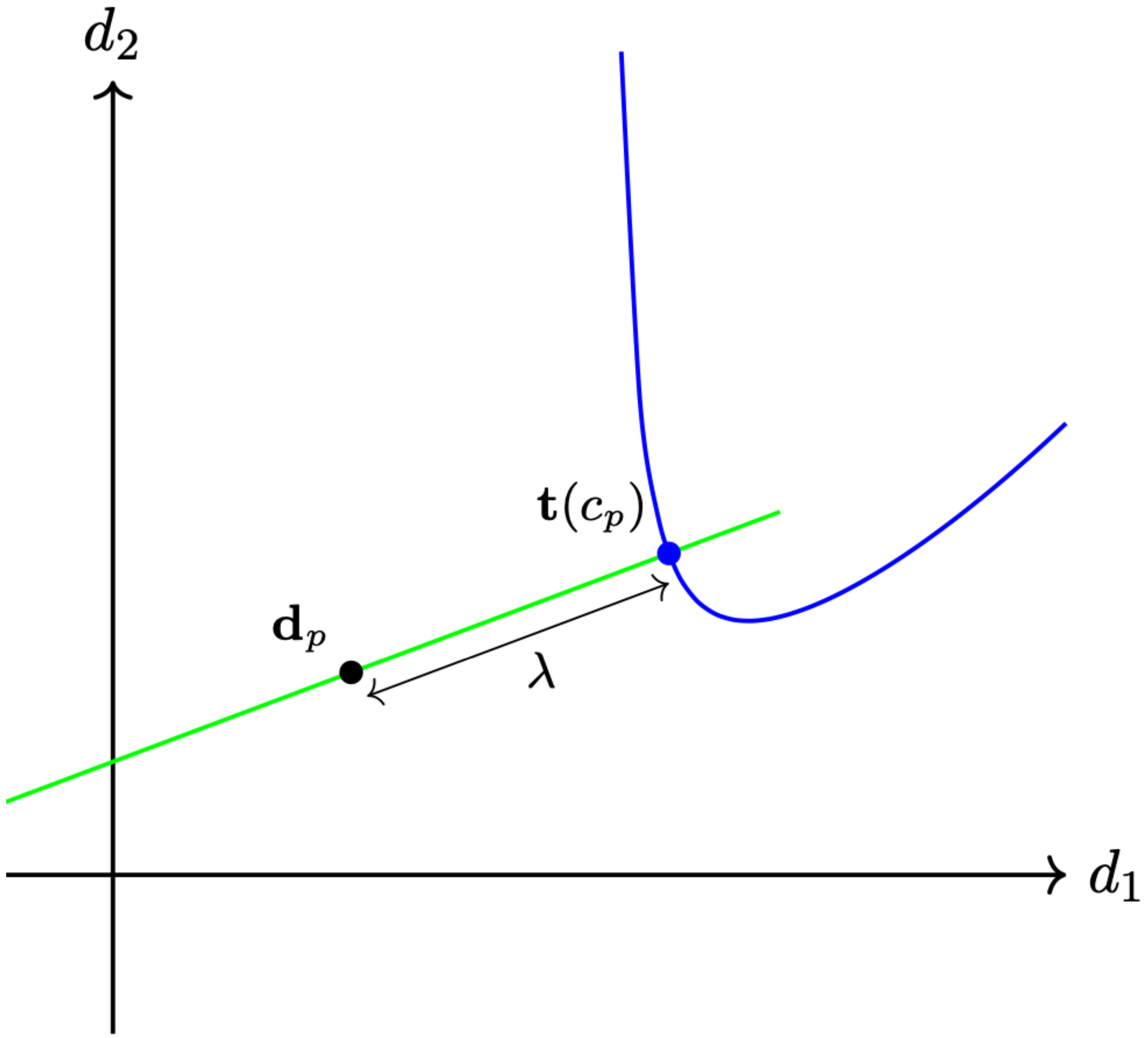
Monte Carlo Posterior

$\boldsymbol{c}_p(\boldsymbol{d}_p)$ Is a random function of $\boldsymbol{d}_p$

$$P_{\mathrm{MC}}(\boldsymbol{c}) \propto \int d^{N_{\mathrm{dat}}}\boldsymbol{d}_p \;\delta\left(\boldsymbol{c} - \boldsymbol{c}_p(\boldsymbol{d}_p)\right) \exp\left(-\frac{1}{2}(\boldsymbol{d}_p - \boldsymbol{d}_0)^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{d}_0)\right)$$

A general form for $P_{\mathrm{MC}}(\boldsymbol{c})$ can be found by performing a coordinate transformation

$$\boldsymbol{d}_p \to (\boldsymbol{c}_p, \boldsymbol{\lambda})$$



Monte Carlo Posterior

Linear theory

See also L. D. Debbio, T. Giani, M. Wilson [2111.05787]

$$\boldsymbol{t}(\boldsymbol{c}) = \boldsymbol{t}_0 + \boldsymbol{t}_{\text{lin}}\boldsymbol{c}$$

The Monte Carlo posterior

$$P_{\text{MC}}(\boldsymbol{c}) \propto \int d^{N_{\text{dat}}}\boldsymbol{d}_p \delta\left(\boldsymbol{c} - \boldsymbol{c}_p(\boldsymbol{d}_p)\right) \exp\left(-\frac{1}{2}(\boldsymbol{d}_p - \boldsymbol{d}_0)^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{d}_0)\right)$$

reduces to

$$P_{\text{MC}}(\boldsymbol{c}) \propto \exp\left(-\frac{1}{2}\chi^2_{\mathbf{d}_0}(\mathbf{c})\right)$$

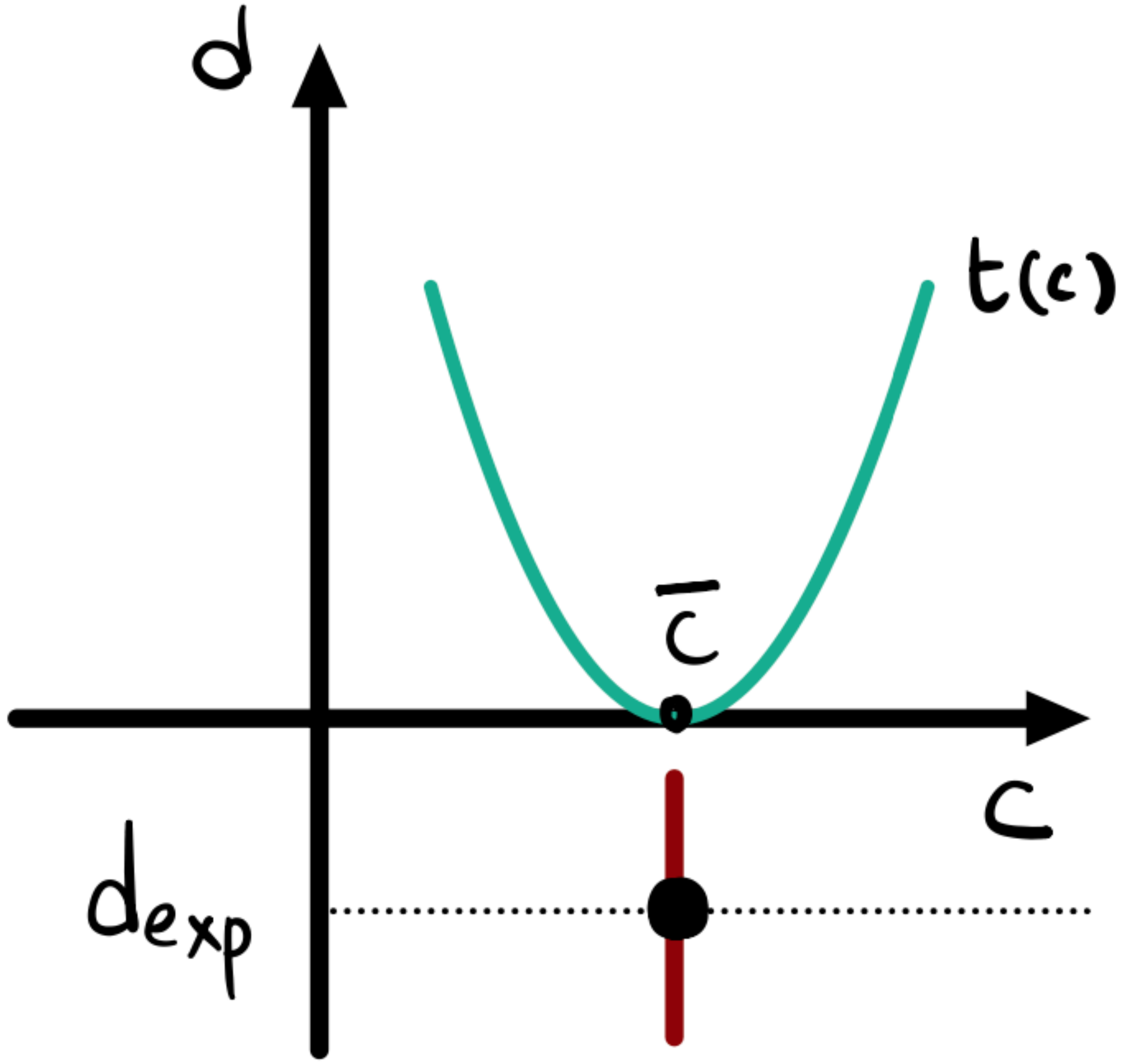
For a linear theory, the Bayesian and Monte Carlo approaches coincide (wide flat prior)

Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

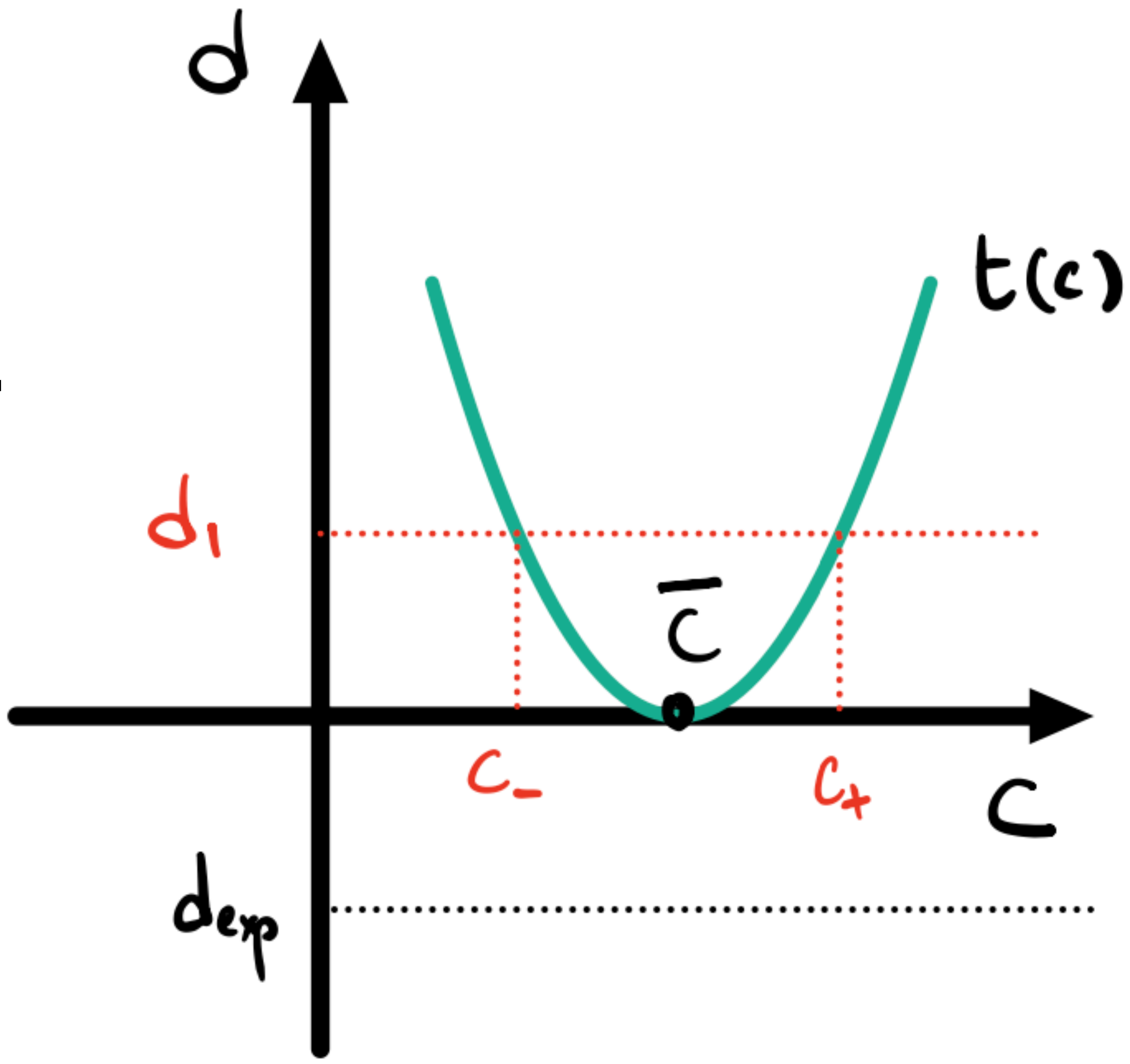
$$t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2,\ t_{\text{quad}} > 0$$



Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter



$$t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2,\ t_{\text{quad}} > 0$$

$$\chi^2 = \frac{(t(c) - d_{\text{exp}})^2}{\sigma^2_{\text{exp}}},\ \Delta\chi^2 = \chi^2 - \chi^2_{\text{min}} = 1 \quad \to [c_-, c_+]$$

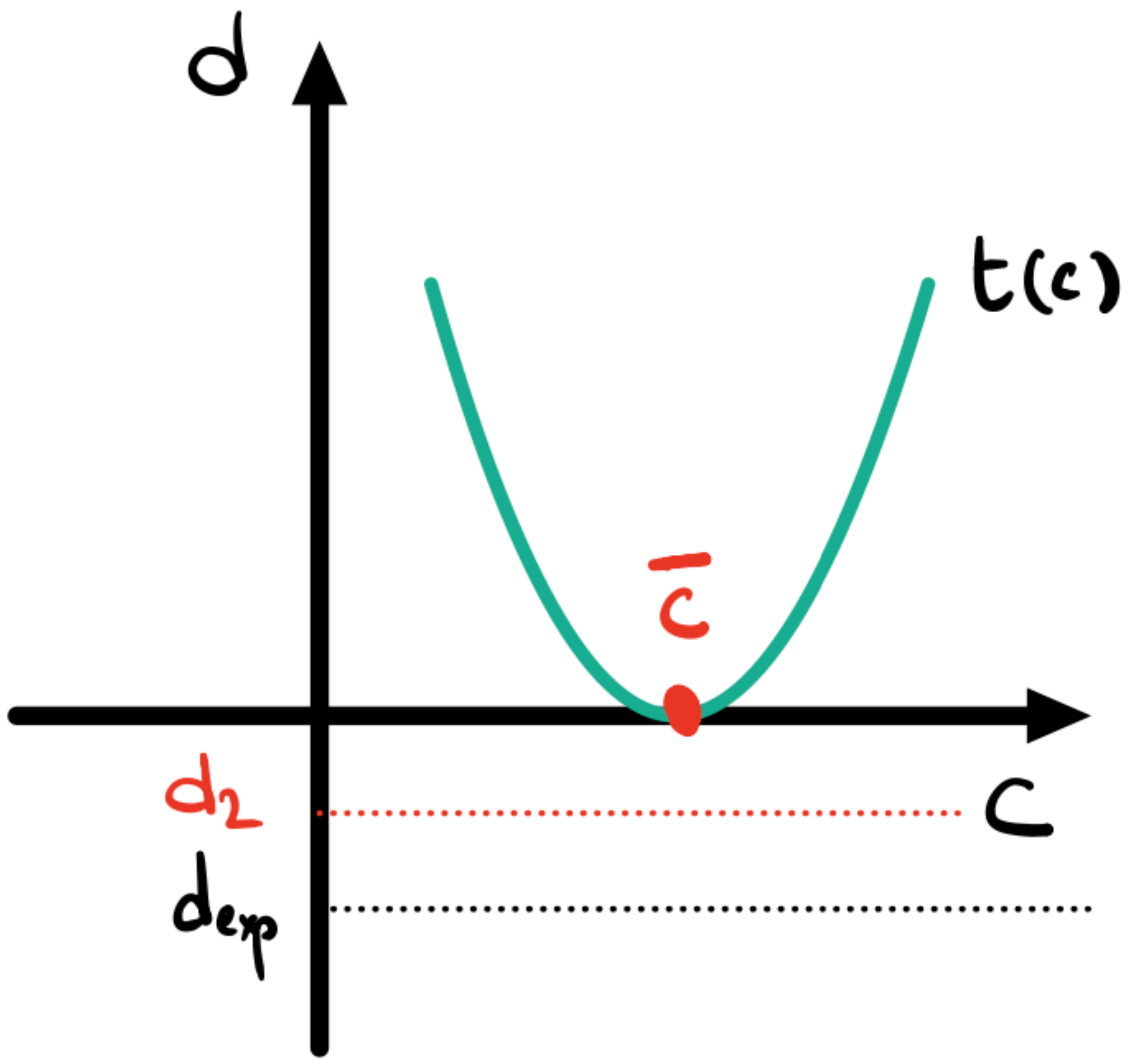
Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2, \; t_{\text{quad}} > 0$$

$$c = -\frac{t_{\text{lin}}}{2 t_{\text{quad}}}$$

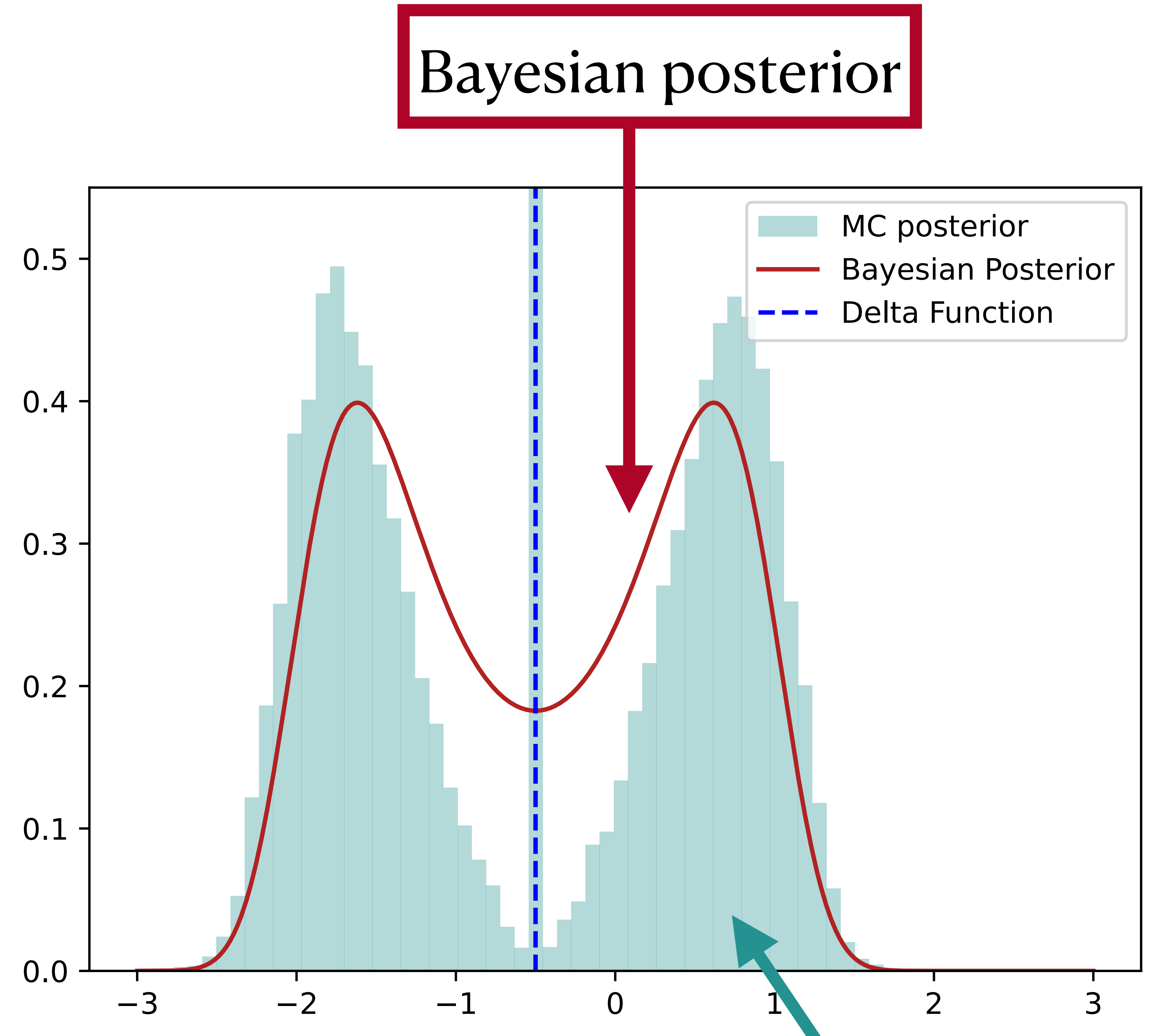


Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2,\ t_{\text{quad}} > 0$$



$$P_{\text{MC}}(c) \propto \delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{tquad}}}\right)\int_{-\infty}^{t_{\min}} dd_p \exp\left(-\frac{1}{2}\frac{(d_0 - d_p)^2}{\sigma^2}\right) + 2\,|\,t_{\text{lin}} + 2ct_{\text{quad}}\,|\exp\left(-\frac{1}{2}\frac{(d_0 - t(c))^2}{\sigma^2}\right)$$

Independent on d_p: $\frac{\partial t(c)}{\partial c}$ has not full rank

2 roots

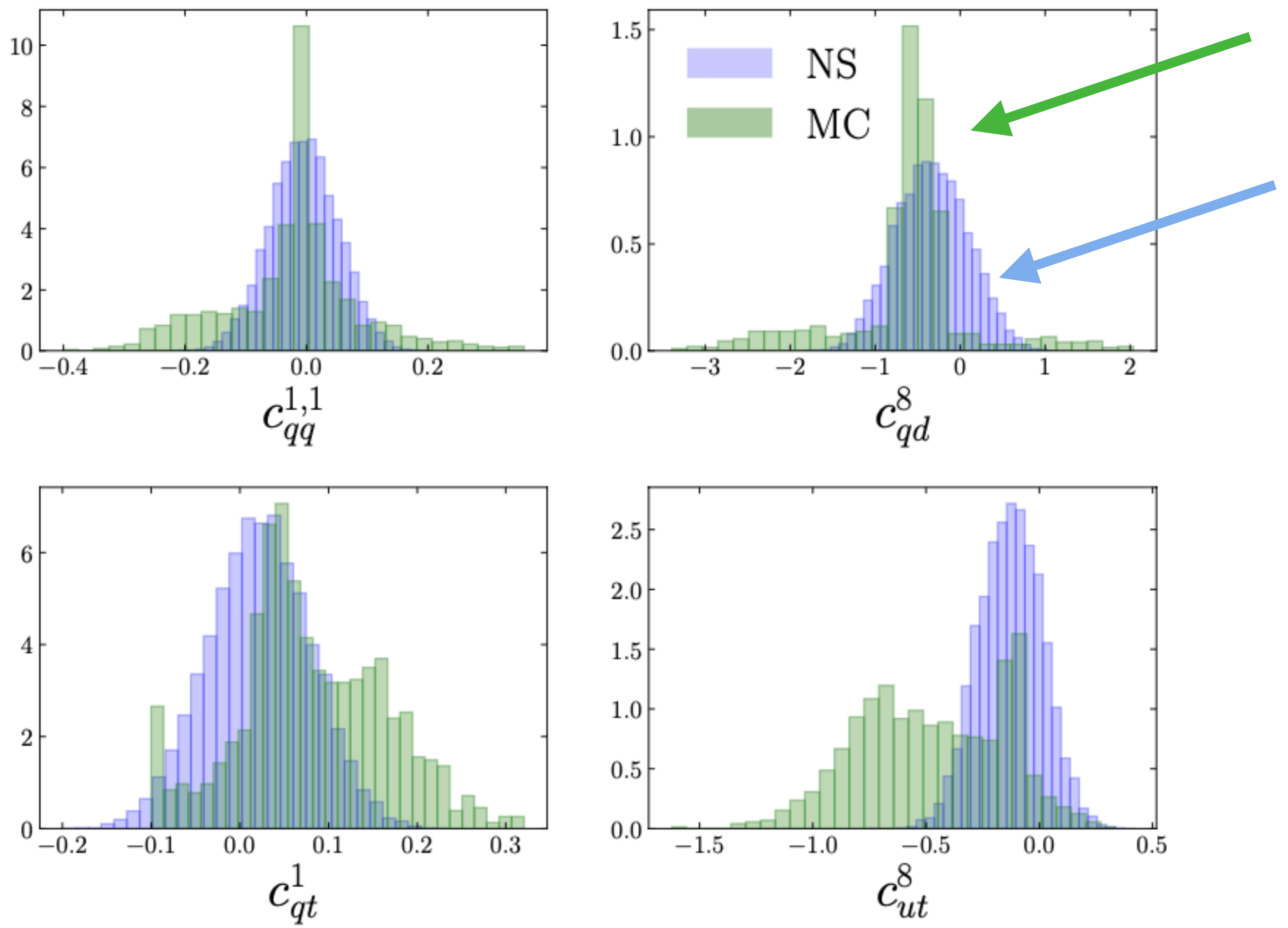
"Determinant" factor

Applications in HEP

Applications in HEP

SMEFT fit in TOP sector

SMEFiT [2302.06660]



Monte Carlo

Nested Sampling

4 fermions operators are dominated by quadratic Wilson coefficients

$$\mathcal{O} = \mathcal{O}_{SM} + \frac{C_i}{\Lambda^2}\,\mathcal{O}_i^{INT} + \frac{C_i\,C_j}{\Lambda^4}\,\mathcal{O}_{ij}^{SQ}$$

Applications in HEP

PDF fits

three flavours in the evolution basis: Σ, g, V

for each flavour we choose a grid

$$(x_1^f, \ldots, x_N^f)$$

reduced grid is treated as an interpolation grid

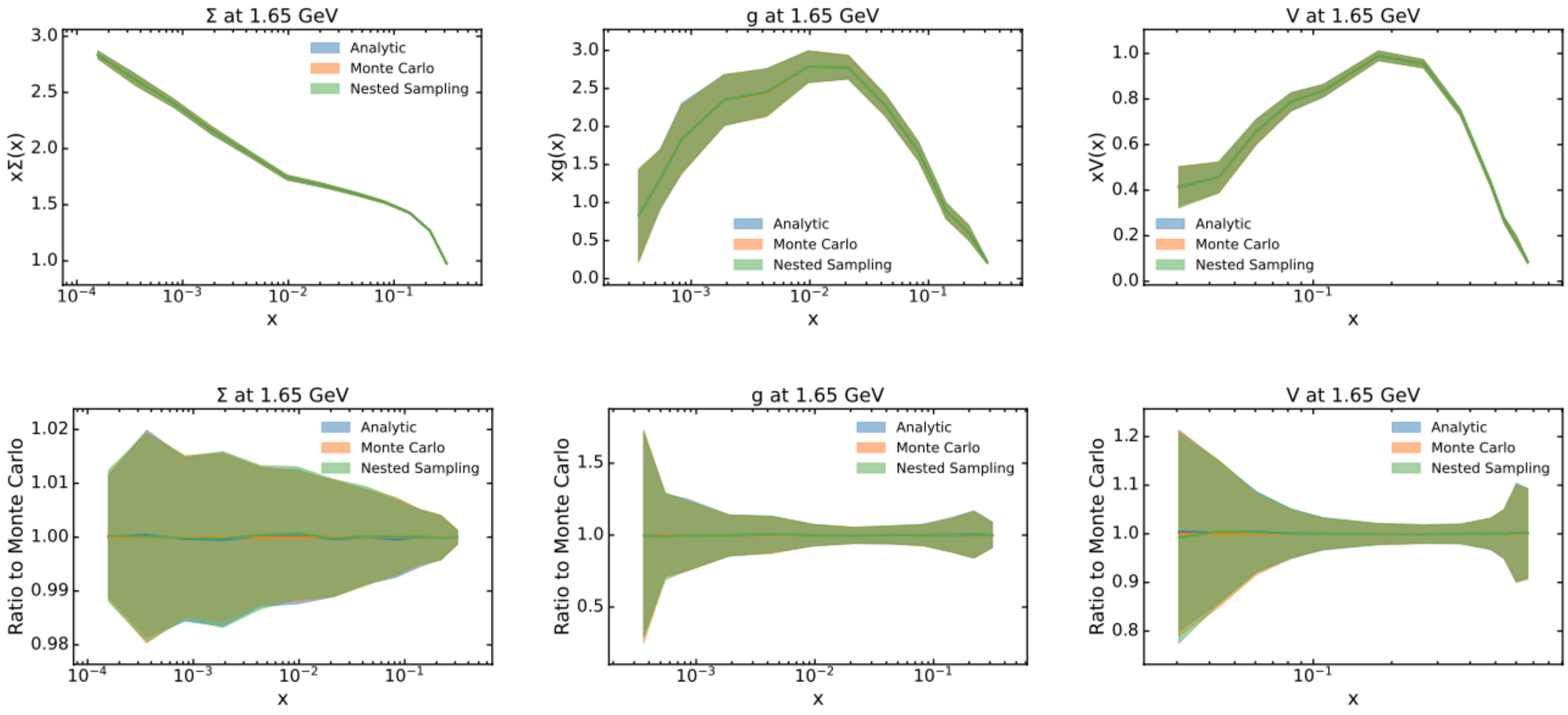
Level 1Closure test:
data = theory + noise
(noise sampled from
covariance matrix)

Linear Interpolation!

$$f(x, Q_0^2) = \begin{cases} f(x_1^f, Q_0^2) & \text{if } x \leq x_1^f; \\ \left(\frac{x_{i+1}^f - x}{x_{i+1}^f - x_i^f}\right) f(x_i^f, Q_0^2) + \left(\frac{x - x_i^f}{x_{i+1}^f - x_i^f}\right) f(x_{i+1}, Q_0^2) & \text{if } x \in [x_i, x_{i+1}], \text{ for } i = 1, \ldots, N_{\text{grid}}(f); \\ f(x_{N_{\text{grid}}(f)}^f, Q_0^2) & \text{if } x > x_{N_{\text{grid}}(f)}^f . \end{cases}$$

Applications in HEP

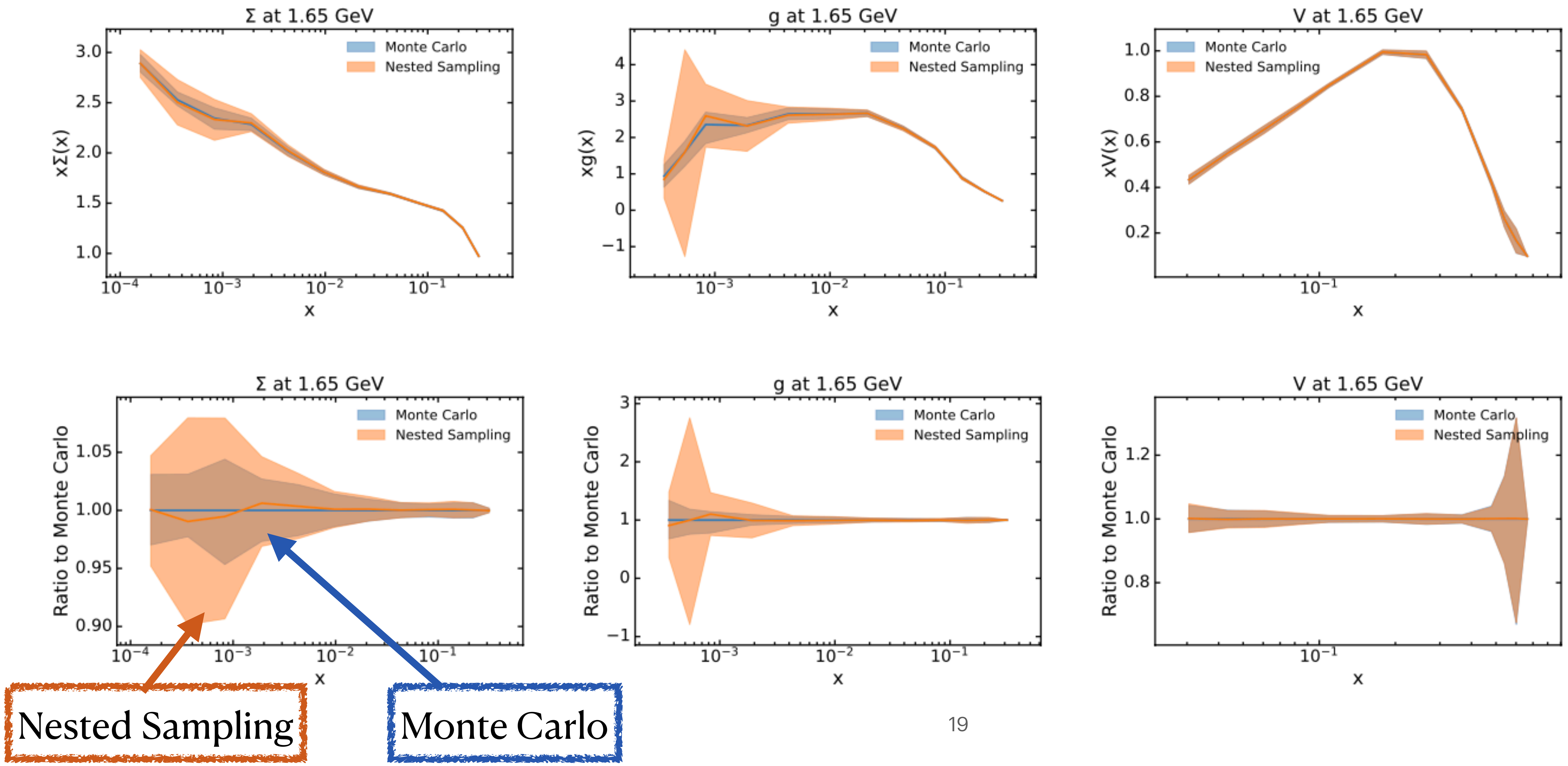
PDF fits: DIS benchmark



~ 3000 datapoints

Applications in HEP

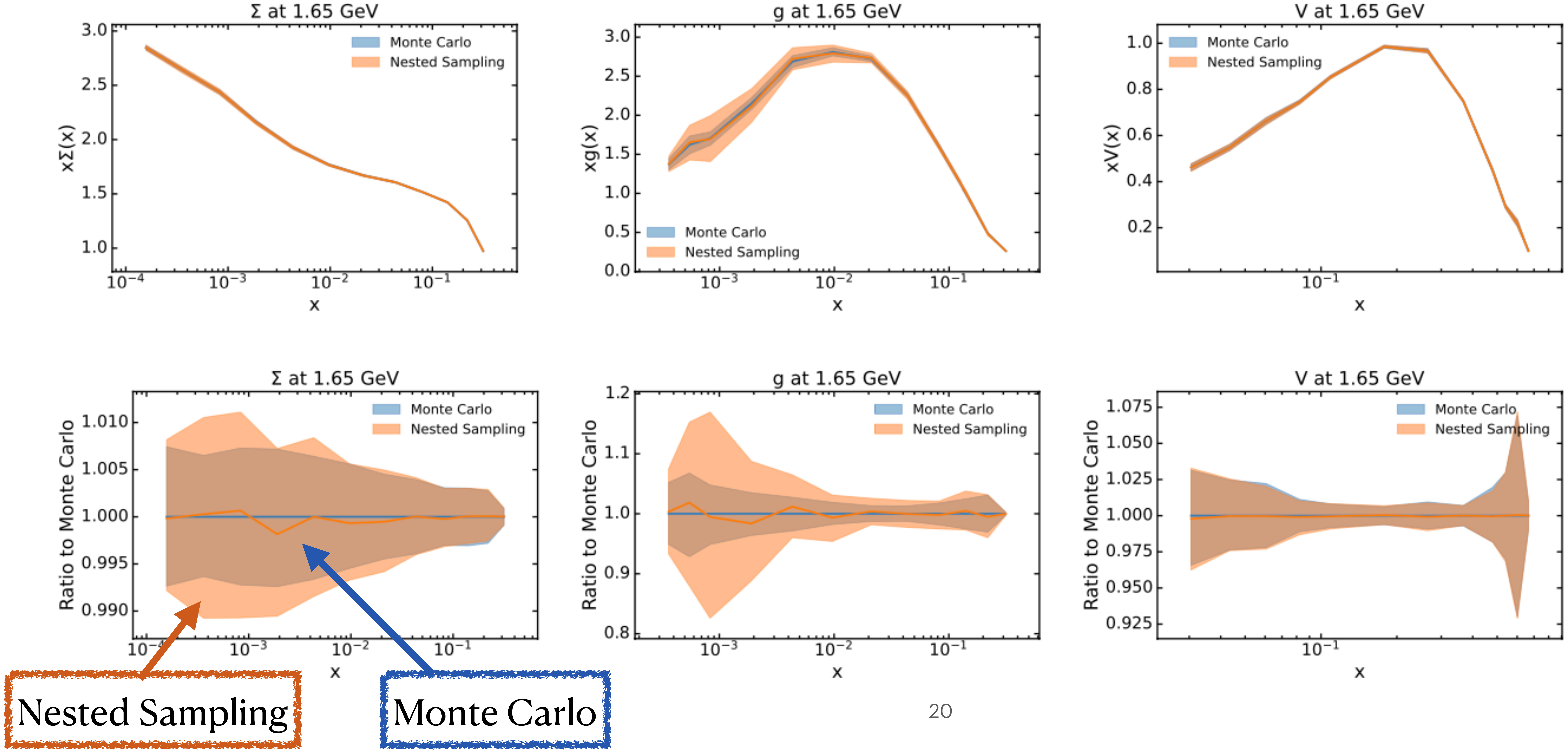
PDF fits: Hadronic-only fit



~ 1000 datapoints

Applications in HEP

PDF fits: Global fit



~ 4000 datapoints

Summary

Summary / Outlook

→ **First rigorous mathematical formulation of the Monte Carlo replica method**

→ **Compare Monte Carlo replica approach against Bayesian approach**

- Agreement for Linear models
- Examples for non linear models

→ **Applications to phenomenologically relevant scenarios: SMEFT**

- SMEFT Fit of the top sector
- Bayesian and Monte Carlo replica method do not agree for quadratic fits

Summary / Outlook

→ Applications to phenomenologically relevant scenarios: PDF fits

- Truly linear parameterisation of PDF
- DIS benchmark: Bayesian and Monte Carlo agree
- For purely hadronic and global fits disagreement is observed

→ Outlook

- Understanding potential limitations of uncertainty quantification methods is really important for the coming years and this sort of study is the starting point
- Repeat exercise with other fitting methodologies (such as Hessian with a tolerance)

Backup

Parametrical Bootstrap

Given data $X_1, \ldots, X_N \sim \mathbb{P}$

And a statistic $\theta_N = \theta_N(X_1, \ldots, X_N)$ that depends on it

make parametrical assumption on distribution $\mathbb{P} = \mathbb{P}_\theta$ and find parameter(s) θ using (e.g.) MLE

use $\mathbb{P}_\theta$ to generate bootstrap sample $\{X_1^{*,1}, \ldots, X_N^{*,1}\}$ and compute the bootstrapped statistic $\theta_N^{*,1} = \theta_N^{*,1}(X_1^{*,1}, \ldots, X_N^{*,1})$

repeat procedure B times so as to get B bootstrapped estimators $\{\theta_N^{*,1}, \ldots, \theta_N^{*,B}\}$ and compute mean, variance, and quantiles on these.

Monte Carlo Posterior

$\boldsymbol{c}_p(\boldsymbol{d}_p)$ Is a random function of $\boldsymbol{d}_p$

$$P_{MC}(\boldsymbol{c}) \propto \int d^{N_{\text{dat}}}\boldsymbol{d}_p \delta\left(\boldsymbol{c} - \boldsymbol{c}_p(\boldsymbol{d}_p)\right) \exp\left(-\frac{1}{2}(\boldsymbol{d}_p - \boldsymbol{d})^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{d})\right)$$

Jacobian factor

Change of coordinates to absorb delta function: $\boldsymbol{d}_p \rightarrow (\boldsymbol{c}_p, \boldsymbol{\lambda})$

$$P_{MC}(\boldsymbol{c}) \propto \exp\left(-\frac{1}{2}\chi^2_{\boldsymbol{d}}(\boldsymbol{c})\right) \int d^{N_{\text{dat}}-N_{\text{param}}}\boldsymbol{\lambda} \left|\det\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{c}} + \frac{\partial(\Sigma M \boldsymbol{\lambda})}{\partial \boldsymbol{c}} \middle| \Sigma M(\boldsymbol{c})\right)\right|$$

$$\times \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M(\boldsymbol{c})^T \Sigma M(\boldsymbol{c})\boldsymbol{\lambda} + \boldsymbol{\lambda}^T M(\boldsymbol{c})^T(\boldsymbol{d} - \boldsymbol{t}(\boldsymbol{c}))\right)$$

Same as for Bayesian posterior

M: basis for the kernel of $\left(\frac{\partial \boldsymbol{t}(\boldsymbol{c})}{\partial \boldsymbol{c}}\right)^T (\boldsymbol{c}_p(\boldsymbol{d}_p))$

Monte Carlo Posterior

Linear Toy Model

one datapoint , one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c, \; t_{\text{lin}} \neq 0$$

$$c_p(d_p) = \text{argmin}_c \left(\frac{(t(c) - d_p)^2}{\sigma^2} \right) = \frac{d_p - t_0}{t_{\text{lin}}}$$

Same as for Bayesian posterior

$$P_{\text{MC}}(c) = \int dd_p \delta(c - c_p(d_p)) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(d_0 - d_p)^2}{\sigma^2} \right) = \frac{t_{\text{lin}}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(d_0 - t(c))^2}{\sigma^2} \right)$$

MC and Bayesian posterior agree in the linear case (General)

Monte Carlo Posterior

Quadratic Toy Model

one datapoint , one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2, \ t_{\text{quad}} > 0$$

$$c_p(d_p) = \text{argmin}_c \left(\frac{(t(c) - d_p)^2}{\sigma^2} \right)$$ can be found analytically

$$P_{\text{MC}}(c) \propto \delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{tquad}}} \right) \int_{-\infty}^{t_{\min}} dd_p \exp\left(-\frac{1}{2} \frac{(d_0 - d_p)^2}{\sigma^2} \right) + 2 \,|\, t_{\text{lin}} + 2ct_{\text{quad}} \,|\, \exp\left(-\frac{1}{2} \frac{(d_0 - t(c))^2}{\sigma^2} \right)$$

Independent on d_p: $\frac{\partial t(c)}{\partial c}$ has not full rank

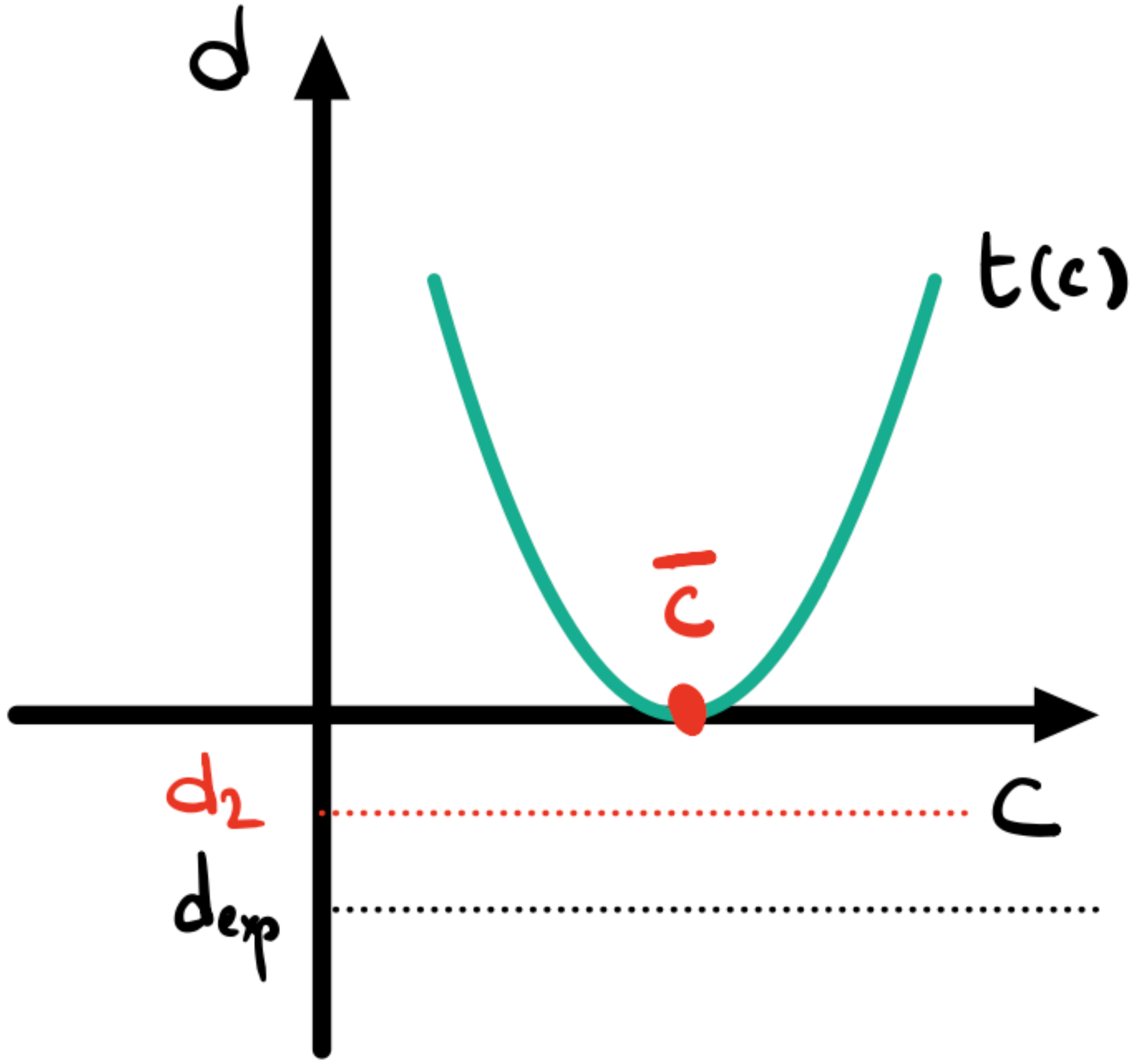
2 roots

“Determinant” factor

Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter



$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2,\ t_{\text{quad}} > 0$$

$$P_{\text{MC}}(c) \propto \delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{tquad}}}\right) \int_{-\infty}^{t_{\min}} dd_p \exp\left(-\frac{1}{2}\frac{(d_0 - d_p)^2}{\sigma^2}\right) + 2\,|\,t_{\text{lin}} + 2ct_{\text{quad}}\,|\exp\left(-\frac{1}{2}\frac{(d_0 - t(c))^2}{\sigma^2}\right)$$

Independent on d_p: $\dfrac{\partial t(c)}{\partial c}$ has not full rank

2 roots

"Determinant" factor

Monte Carlo Posterior

Quadratic Toy Model

one datapoint , one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2, \; t_{\text{quad}} > 0$$

MC and Bayesian posterior do not agree in the quadratic case $P_{\text{MC}}(c) \neq P_B(c \,|\, d_0)$

Note: In general it is unclear whether MC method under- or over-estimates the Bayesian uncertainties. Interestingly both scenarios can happen

→ A numerical simulation is needed in order to assess the effect of the "determinant" factor

Nested Sampling

General Idea

- Monte Carlo algorithm for computing an integral over a model parameter space
- Nested Sampling provides both the posterior samples as well as the marginalised likelihood Z

Bayes Rule

$$P(\Theta|D) = \frac{L(D|\Theta)\pi(\Theta)}{Z}$$

Marginalised Likelihood

$$Z = \int L(D|\Theta)\pi(\Theta)d\Theta$$

Nested Sampling

Algorithm

1. Initialisation: sample randomly from the prior N live points and compute the Likelihood at each point
2. Shrinkage: remove point with the lowest likelihood L_1
3. Likelihood Restricted Prior Sampling: sample new point from prior with Likelihood $> L_1$
4. Iterate

Iteration i reduces integration volume by a factor $\delta V_i \approx \left(1 - \frac{1}{N}\right)^i \frac{1}{N}$,

The integral Z is simply $Z \approx \sum_i \delta V_i \times L_i$

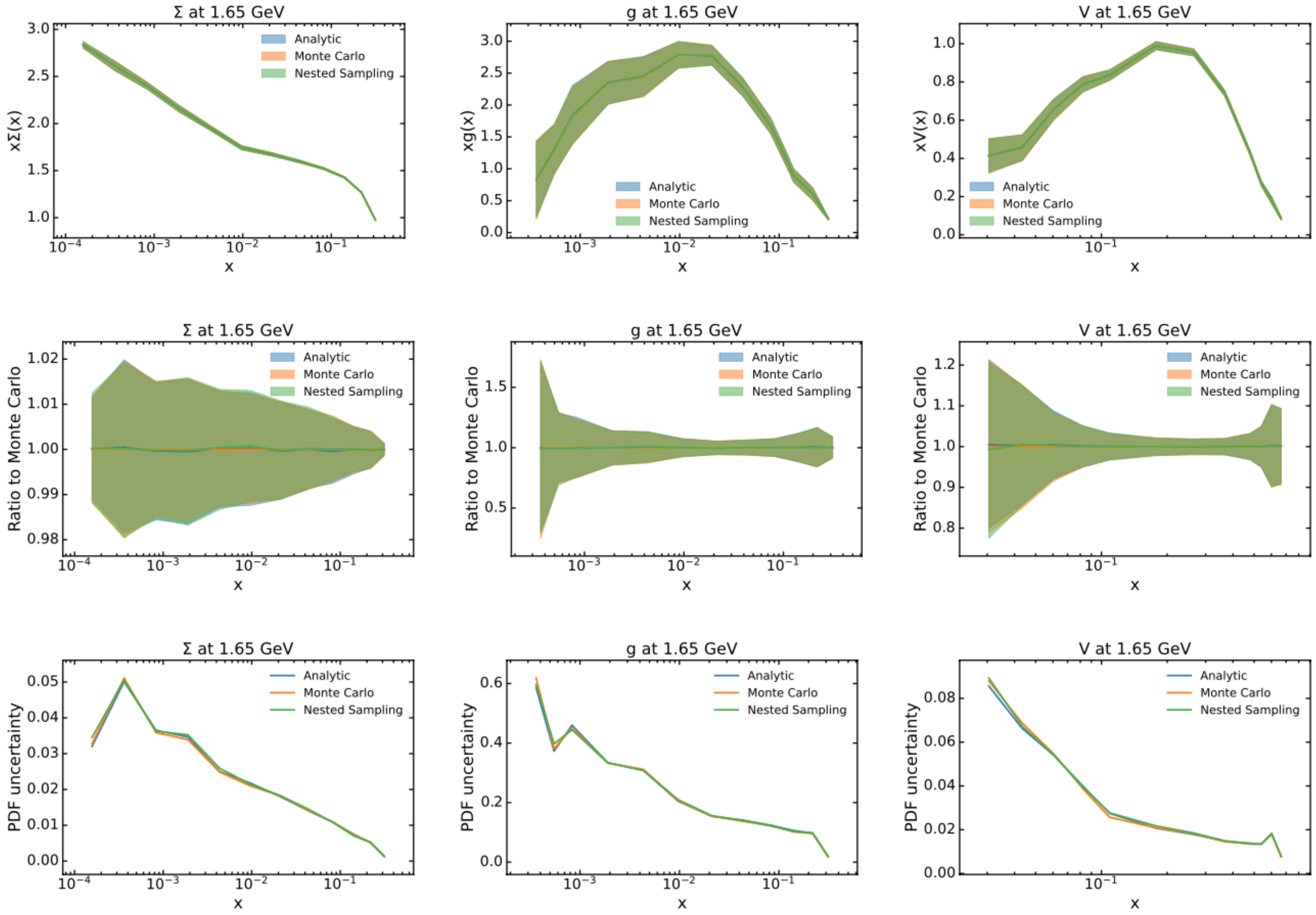
Termination: when $\delta V_i \times L_i$ contributions to Z are negligible

Nested Sampling

Summary

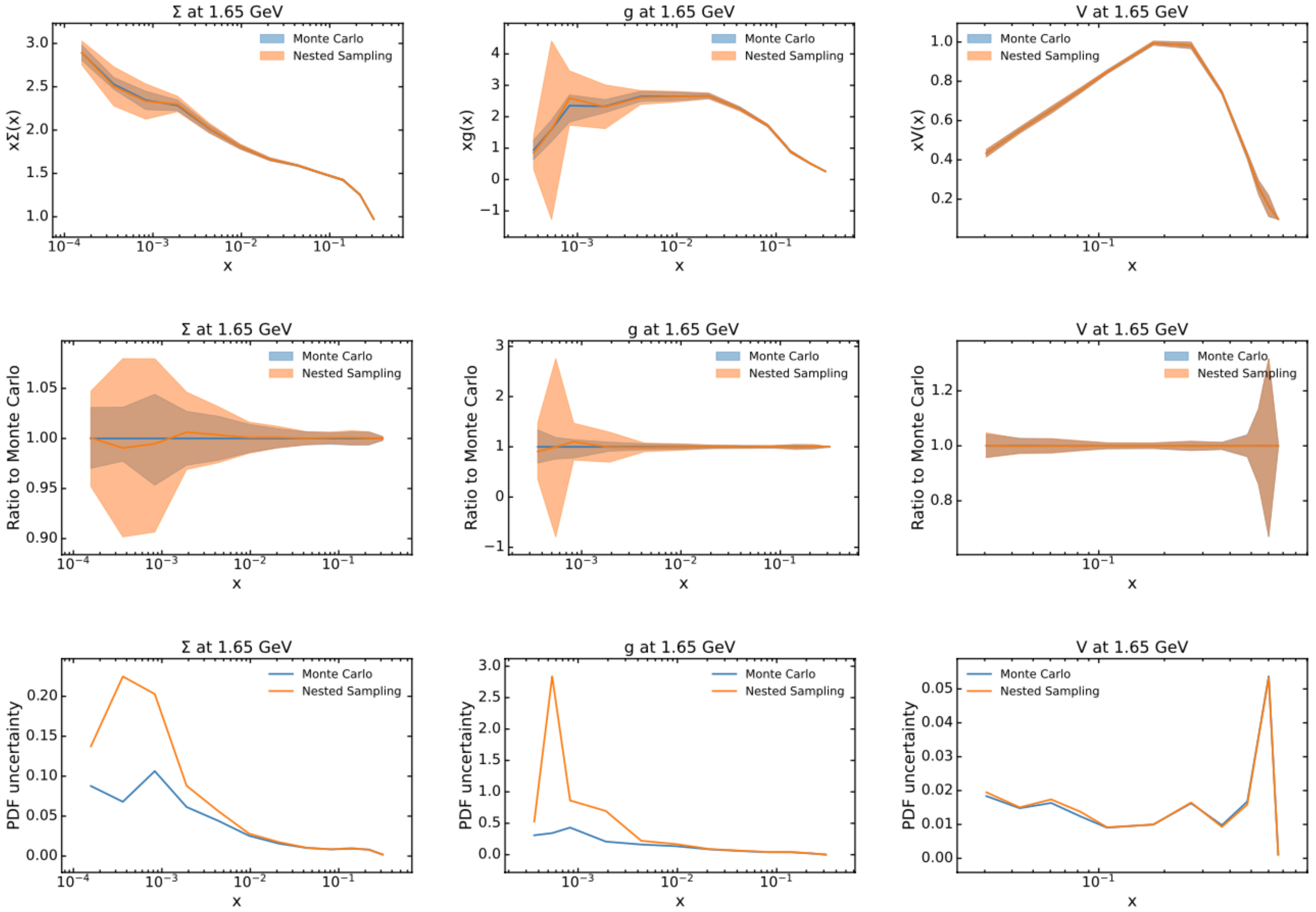
1. It explores the parameter space globally;
2. it handles multi-modal distributions well;
3. it initialises and terminates at a well defined point -> no supervision;
4. it provides both marginal likelihood and posterior samples;

DIS benchmark



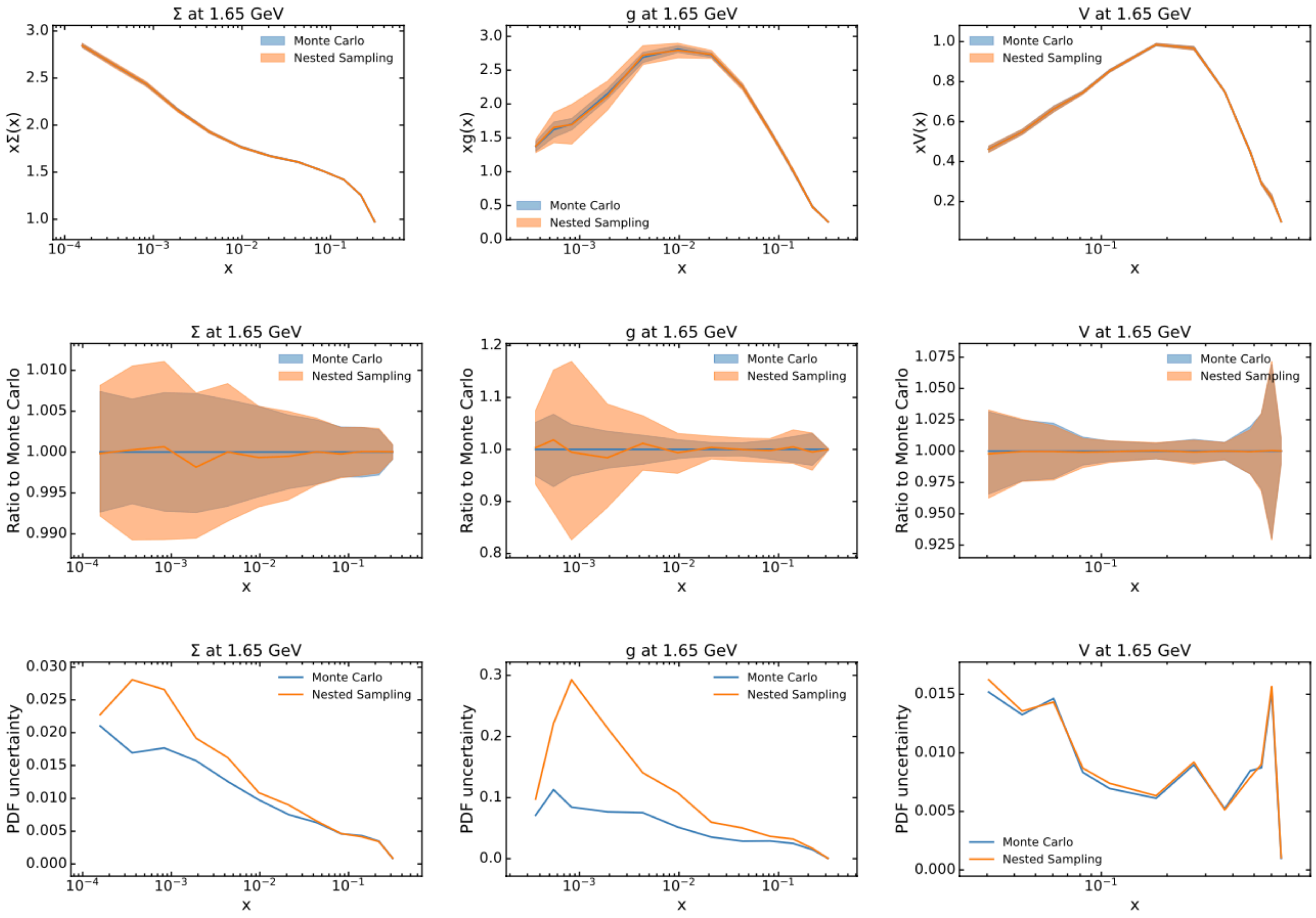
~ 3000 datapoints

Hadronic-only



~ 1000 datapoints

Global fit



~ 4000 datapoints