
PDF uncertainties. What is the meaning of this?

Wally Melnitchouk



Outline

- Why are PDF uncertainties important to know?
- Why the need for a new global QCD analysis paradigm?
 - Bayesian approach to fitting
 - single-fit (Hessian) vs. Monte Carlo approaches
- Incompatible data sets
 - “tolerance” factors
- Generalization to non-Gaussian likelihoods
 - disjoint probabilities, empirical Bayes, ...
- Outlook

Why are PDF uncertainties important?

- In searches for new physics beyond the Standard Model a major source of uncertainty on limits/discoveries is from calculation of QCD backgrounds → PDF errors!

“The PDF and α_s uncertainties were calculated using the PDF4LHC prescription [39] with the MSTW2008 68% CL NNLO [40, 41], CT10 NNLO [42, 43], and NNPDF2.3 5f FFN [44] PDF sets, and added in quadrature to the scale uncertainty.”

Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s} = 8$ TeV with the ATLAS detector PRD **94**, 032006 (2016)

→ drives a large part of the global PDF community (esp. LHC)

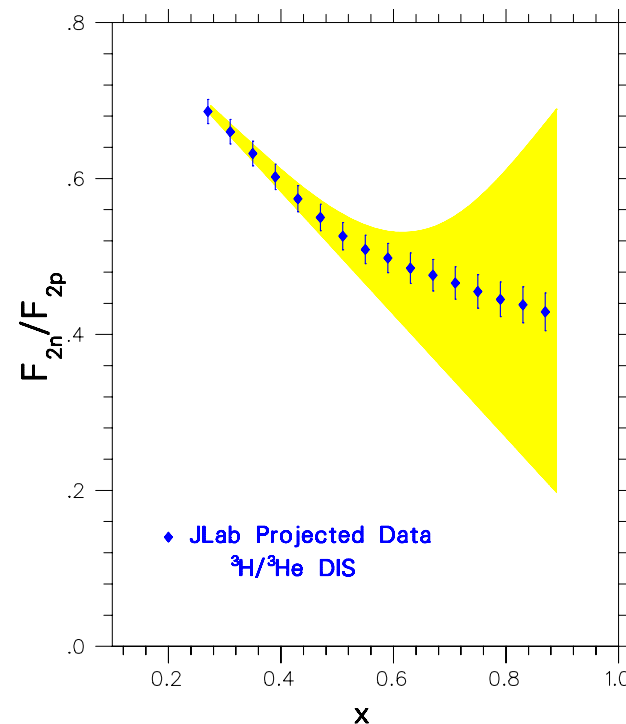
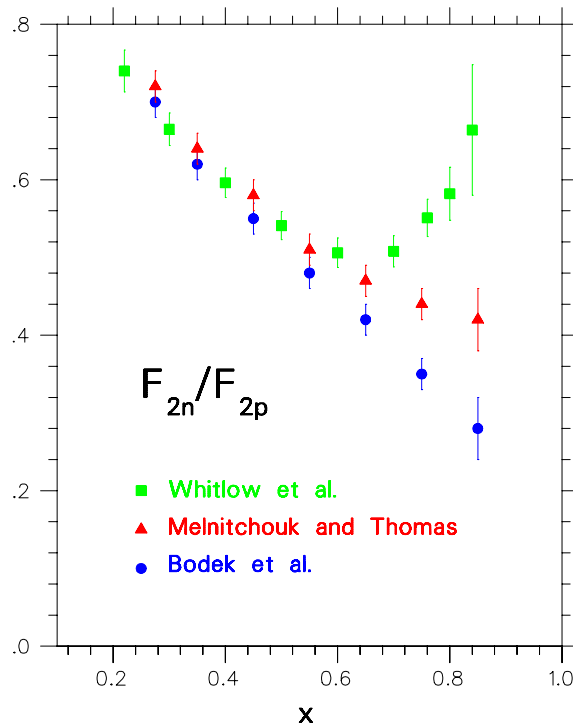
- Limits understanding of nucleon structure

→ *e.g.* momentum and spin distributions of d quarks at large x

→ motivation for several JLab12 experiments
(MARATHON, BONuS, SoLID, ...)

d/u ratio at large x

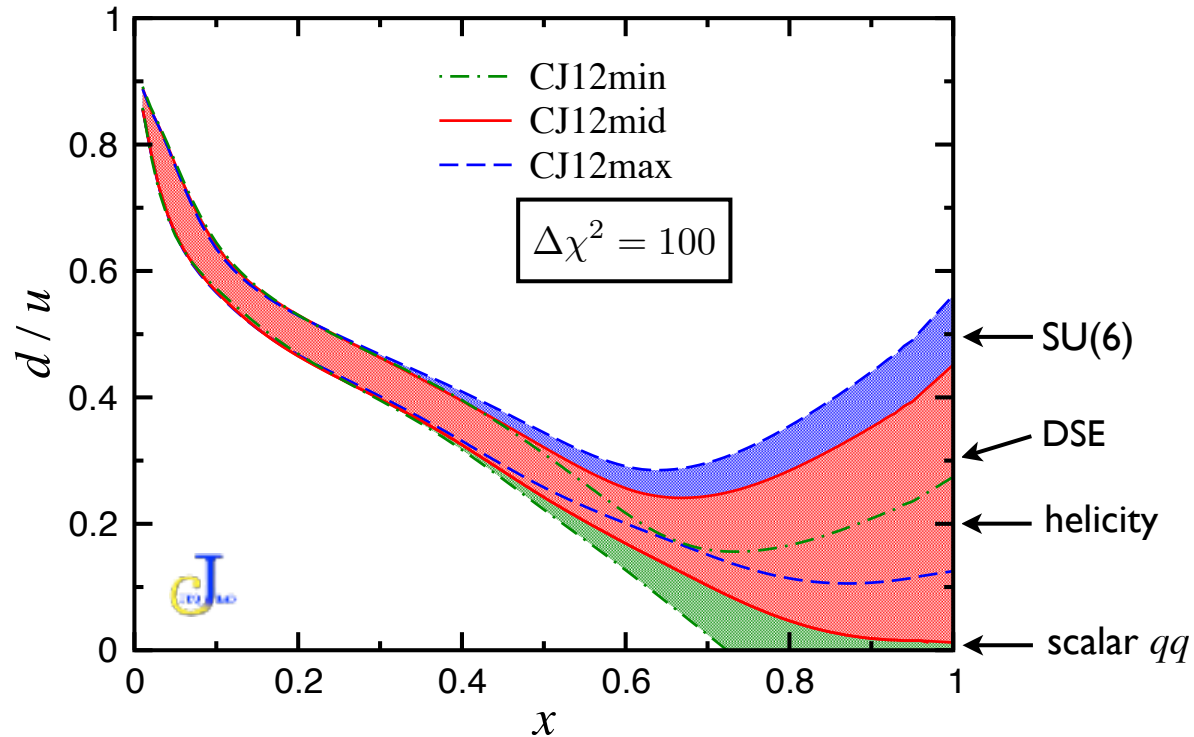
- Traditionally extracted from neutron / proton structure function ratio (where “neutron” $\sim$ deuteron – proton), but large nuclear uncertainties affect high- x region



MARATHON Collaboration
Petratos, Katramatou, Gomez et al.

d/u ratio at large x

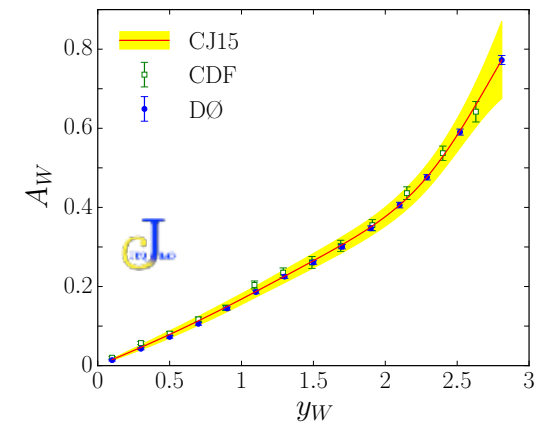
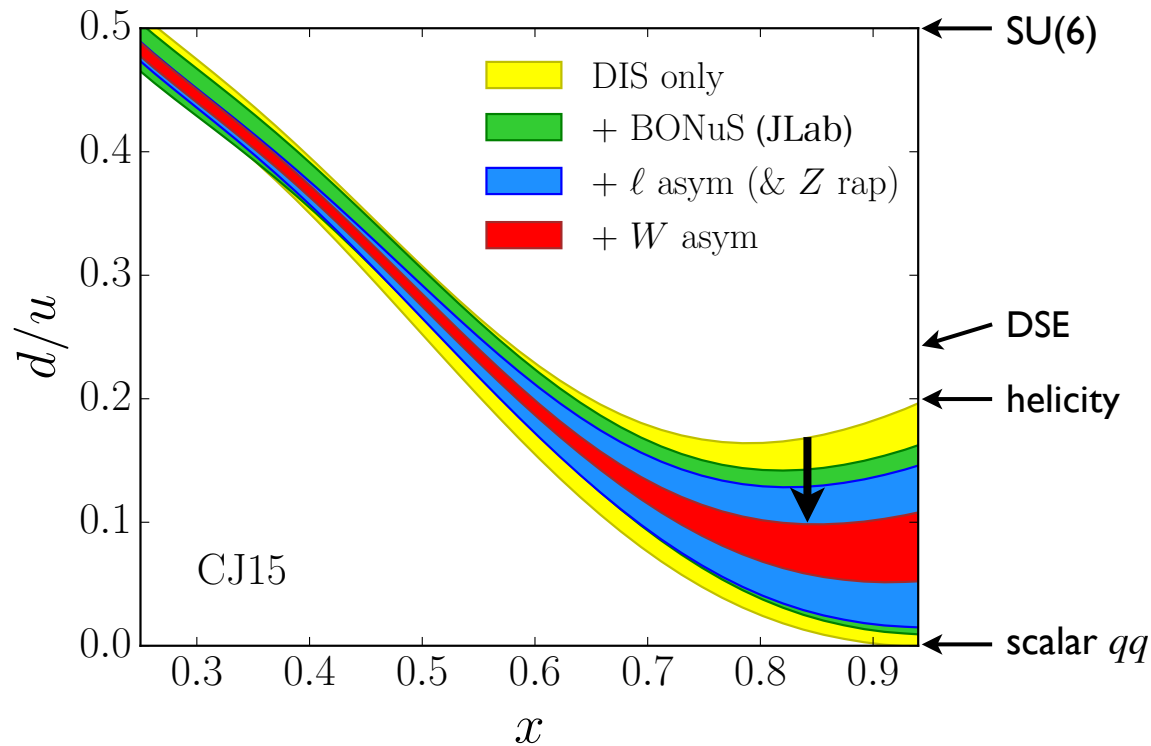
- Traditionally extracted from neutron / proton structure function ratio (where “neutron” $\sim$ deuteron – proton), but large nuclear uncertainties affect high- x region
 - cannot discriminate between predictions for d/u at $x \sim 1$



Owens, Accardi, WM (2013)

d/u ratio at large x

- More recently CJ15 analysis found significant reduction of PDF errors with inclusion of DØ W -asymmetry & BONuS data



Accardi, Brady, WM, Owens, Sato (2016)

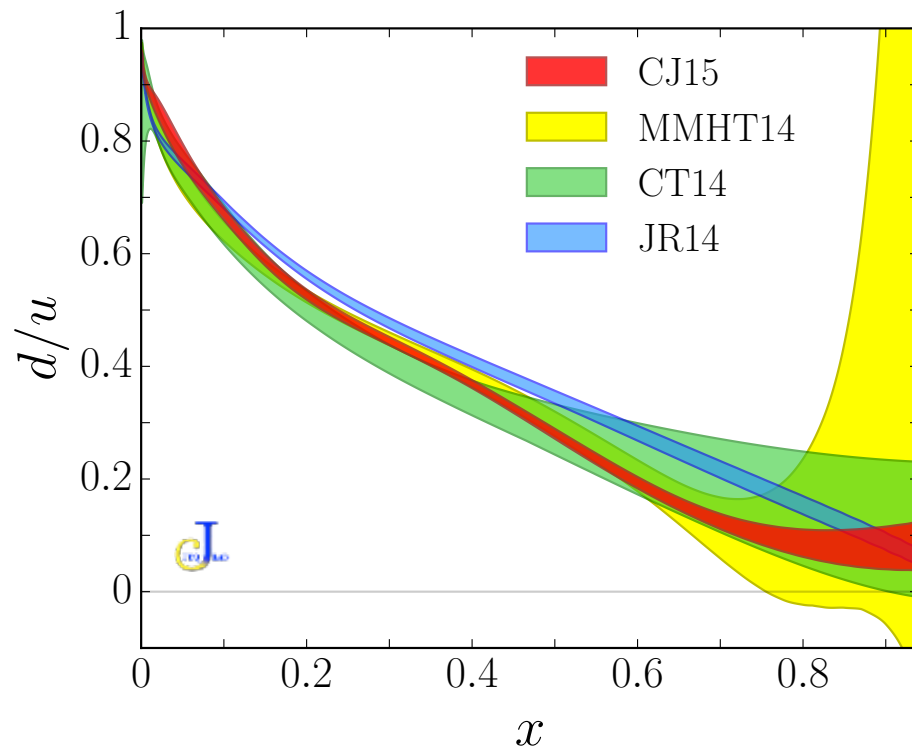
→ extrapolated ratio at $x = 1$: $d/u \rightarrow 0.09 \pm 0.03$

→ note: errors are 90% CL ($\Delta\chi^2 = 2.7$)

d/u ratio at large x

- Different groups use different definitions of PDF uncertainties to take into account *tensions* between data sets

→ multiply uncertainties by “tolerance” factor $T = \sqrt{\Delta\chi^2}$



→ CJ15: $\Delta\chi^2 = 2.7$

→ MMHT: $\Delta\chi^2 \approx 25 - 100$

→ CT14: $\Delta\chi^2 \approx 100$

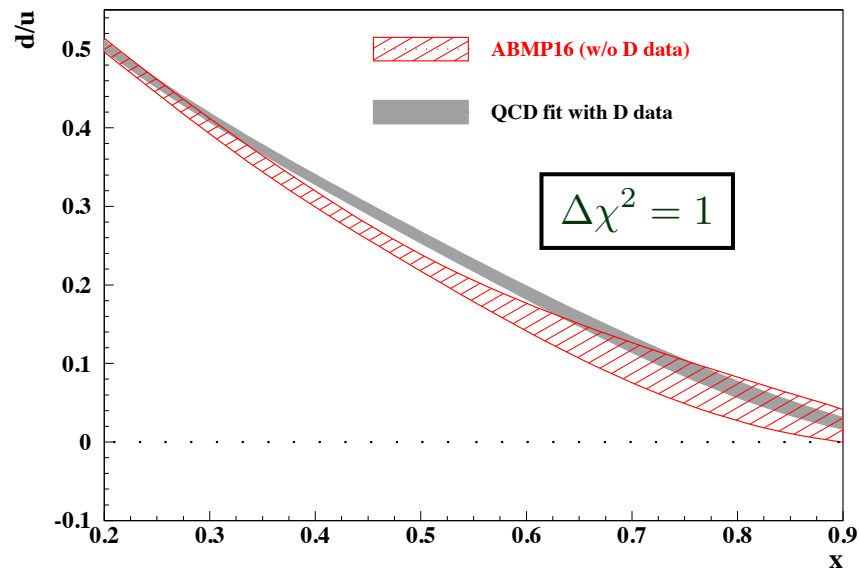
→ JR14: $\Delta\chi^2 = 1$

... is this a meaningful comparison?

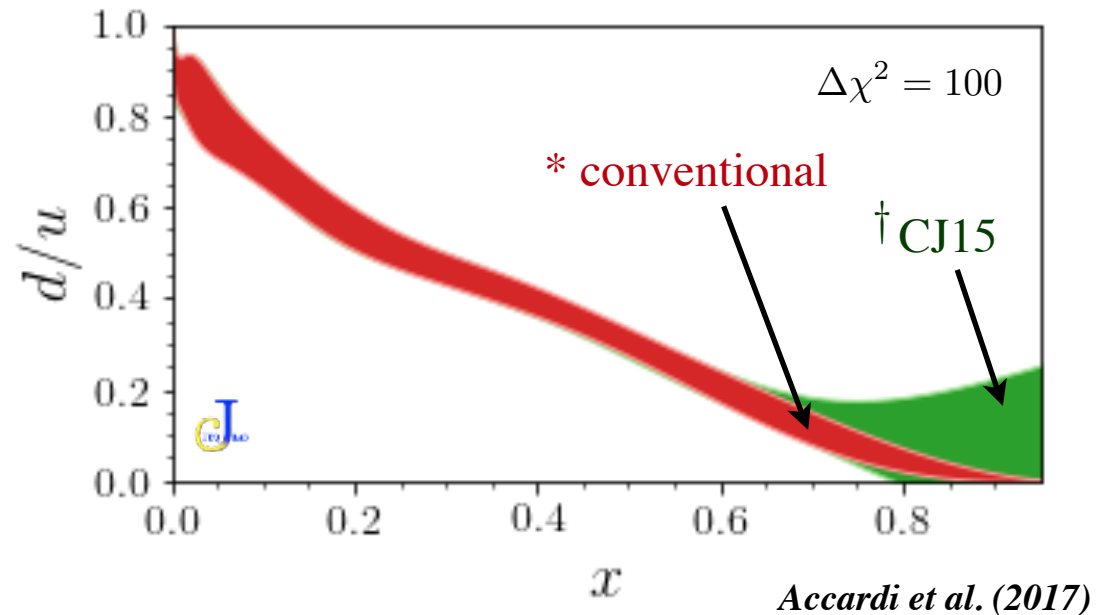
d/u ratio at large x

■ Dependence on PDF parametrization

→ recent analysis by AKP has tiny uncertainties, and $d/u \rightarrow 0$, which we (CJ) believe is simply parametrization bias!



Alekhin, Kulagin, Petti (2017)



Accardi et al. (2017)

* same functional form for u & $d \sim (1-x)^\beta$

† more flexible form $d \rightarrow d + a x^b u$

... is there a more robust analysis?

Need for new technology

- A major challenge has been to characterize PDF uncertainties — in a statistically meaningful way — in the presence of *tensions* among data sets
- Previous attempts sought to address tensions in data sets by introducing
 - “tolerance” factors (artificially inflating PDF errors)
 - “neural net” parametrization (instead of polynomial parametrization), together with MC techniques
- However, to address the problem in a more statistically rigorous way, one requires going *beyond* the standard χ^2 minimization paradigm
 - utilize modern techniques based on Bayesian statistics!

Need for new technology

- In the near future, standard χ^2 minimization techniques will be unsuitable — even in the absence of tensions — *e.g.* for
 - simultaneous analysis of collinear distributions (unpolarized & polarized PDFs, fragmentation functions)
 - new types of observables — TMDs or GPDs — that will involve $> \mathcal{O}(10^5)$ data points, with $\mathcal{O}(10^3)$ parameters

Need for new technology

- Typically PDF parametrizations are nonlinear functions of the PDF parameters, *e.g.*

$$xf(x, \mu) = Nx^\alpha(1-x)^\beta P(x)$$

where P is a polynomial *e.g.* $P(x) = 1 + \epsilon\sqrt{x} + \eta x$,
or Chebyshev, neural net, ...

→ have multiple local minima present in the χ^2 function

- Robust parameter estimation that thoroughly scans over a realistic parameter space, including multiple local minima, is only possible using MC methods

Bayesian approach to fitting

Nobuo Sato



main instigator

Bayesian approach to fitting

- Analysis of data requires estimating expectation values E and variances V of “observables” $\mathcal{O}$ (= PDFs, FFs) which are functions of parameters $\vec{a}$

$$E[\mathcal{O}] = \int d^n a \mathcal{P}(\vec{a}|\text{data}) \mathcal{O}(\vec{a})$$
$$V[\mathcal{O}] = \int d^n a \mathcal{P}(\vec{a}|\text{data}) [\mathcal{O}(\vec{a}) - E[\mathcal{O}]]^2$$

“Bayesian master formulas”

- Using Bayes’ theorem, probability distribution $\mathcal{P}$ given by

$$\mathcal{P}(\vec{a}|\text{data}) = \frac{1}{Z} \mathcal{L}(\text{data}|\vec{a}) \pi(\vec{a})$$

in terms of the likelihood function $\mathcal{L}$

Bayesian approach to fitting

■ Likelihood function

$$\mathcal{L}(\text{data}|\vec{a}) = \exp\left(-\frac{1}{2}\chi^2(\vec{a})\right)$$

is a Gaussian form in the data, with χ^2 function

$$\chi^2(\vec{a}) = \sum_i \left(\frac{\text{data}_i - \text{theory}_i(\vec{a})}{\delta(\text{data})} \right)^2$$

with priors $\pi(\vec{a})$ and “evidence” Z

$$Z = \int d^n a \mathcal{L}(\text{data}|\vec{a}) \pi(\vec{a})$$

→ Z tests if *e.g.* an n -parameter fit is statistically different from $(n+1)$ -parameter fit

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

Monte Carlo

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

- maximize probability distribution $\mathcal{P}$ by minimizing χ^2 for a set of best-fit parameters $\vec{a}_0$

$$E [\vec{a}] = \vec{a}_0$$

- if $\mathcal{O}$ is $\approx$ linear in the parameters, and if probability is symmetric in all parameters

$$E [\mathcal{O}(\vec{a})] \approx \mathcal{O}(\vec{a}_0)$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

→ variance computed by expanding $\mathcal{O}(\vec{a})$ about $\vec{a}_0$
e.g. in 1 dimension have “master formula”

$$V[\mathcal{O}] \approx \frac{1}{4} \left[\mathcal{O}(a + \delta a) - \mathcal{O}(a - \delta a) \right]^2$$

where

$$\delta a^2 = V[a]$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

→ generalization to multiple dimensions via Hessian approach:

find set of (orthogonal) contours in parameter space around $\vec{a}_0$ such that $\mathcal{L}$ along each contour is parametrized by statistically independent parameters — directions of contours given by eigenvectors $\hat{e}_k$ of Hessian matrix H , with elements

$$H_{ij} = \frac{1}{2} \left. \frac{\partial^2 \chi^2(\vec{a})}{\partial a_i \partial a_j} \right|_{\vec{a}=\vec{a}_0}$$

and contours parametrized as $\Delta a^{(k)} = a^{(k)} - a_0 = t_k \frac{\hat{e}_k}{\sqrt{v_k}}$,
with v_k eigenvalues of H

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

→ basic assumption: $\mathcal{P}$ factorizes along each eigendirection

$$\mathcal{P}(\Delta a) \approx \prod_k \mathcal{P}_k(t_k)$$

where

$$\mathcal{P}_k(t_k) = \mathcal{N}_k \exp \left[-\frac{1}{2} \chi^2 \left(a_0 + t_k \frac{\hat{e}_k}{\sqrt{v_k}} \right) \right]$$

note: in quadratic approximation for χ^2 , this becomes a normal distribution

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood

(χ^2 minimization)

→ uncertainties on $\mathcal{O}$ along each eigendirection
(assuming linear approximation)

$$(\Delta \mathcal{O}_k)^2 \approx \frac{1}{4} \left[\mathcal{O} \left(a_0 + T_k \frac{\hat{e}_k}{\sqrt{v_k}} \right) - \mathcal{O} \left(a_0 - T_k \frac{\hat{e}_k}{\sqrt{v_k}} \right) \right]^2$$

where T_k is finite step size in t_k , with total variance

$$V[\mathcal{O}] = \sum_k (\Delta \mathcal{O}_k)^2$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Monte Carlo

- in practice, generally one has $E[\mathcal{O}(\vec{a})] \neq \mathcal{O}(E[\vec{a}])$
so the maximal likelihood method will sometimes fail
- Monte Carlo approach samples parameter space and assigns weights w_k to each set of parameters a_k
- expectation value and variance are then weighted averages

$$E[\mathcal{O}(\vec{a})] = \sum_k w_k \mathcal{O}(\vec{a}_k), \quad V[\mathcal{O}(\vec{a})] = \sum_k w_k (\mathcal{O}(\vec{a}_k) - E[\mathcal{O}])^2$$

Bayesian approach to fitting

- Two methods generally used for computing Bayesian master formulas:

Maximum Likelihood (χ^2 minimization)

- fast
- assumes Gaussianity
- no guarantee that global minimum has been found
- errors only characterize local geometry of χ^2 function

Monte Carlo

- slow
- does not rely on Gaussian assumptions
- includes all possible solutions
- accurate

Incompatible data sets

N. Sato, M. Albright, WM, H. Prosper, M. White (2017)

A. Accardi, E. Nocera, N. Sato, WM (2018)

Incompatible data sets

- Incompatible data sets can arise because of errors in determining central values, or underestimation of systematic experimental uncertainties
→ requires some sort of modification to standard statistics
- Modify the master formula by introducing a “tolerance” factor T

$$V[\mathcal{O}] \rightarrow T^2 V[\mathcal{O}]$$

e.g. for one dimension

$$V[\mathcal{O}] = \frac{T^2}{4} \left[\mathcal{O}(a + \delta a) - \mathcal{O}(a - \delta a) \right]^2$$

→ effectively modifies the likelihood function

Incompatible data sets

- Simple example: consider observable m , and two measurements

$$(m_1, \delta m_1), \quad (m_2, \delta m_2)$$

→ compute exactly the χ^2 function

$$\chi^2 = \left(\frac{m - m_1}{\delta m_1} \right)^2 + \left(\frac{m - m_2}{\delta m_2} \right)^2$$

and, from Bayesian master formula, the mean value

$$E[m] = \frac{m_1 \delta m_2^2 + m_2 \delta m_1^2}{\delta m_1^2 + \delta m_2^2}$$

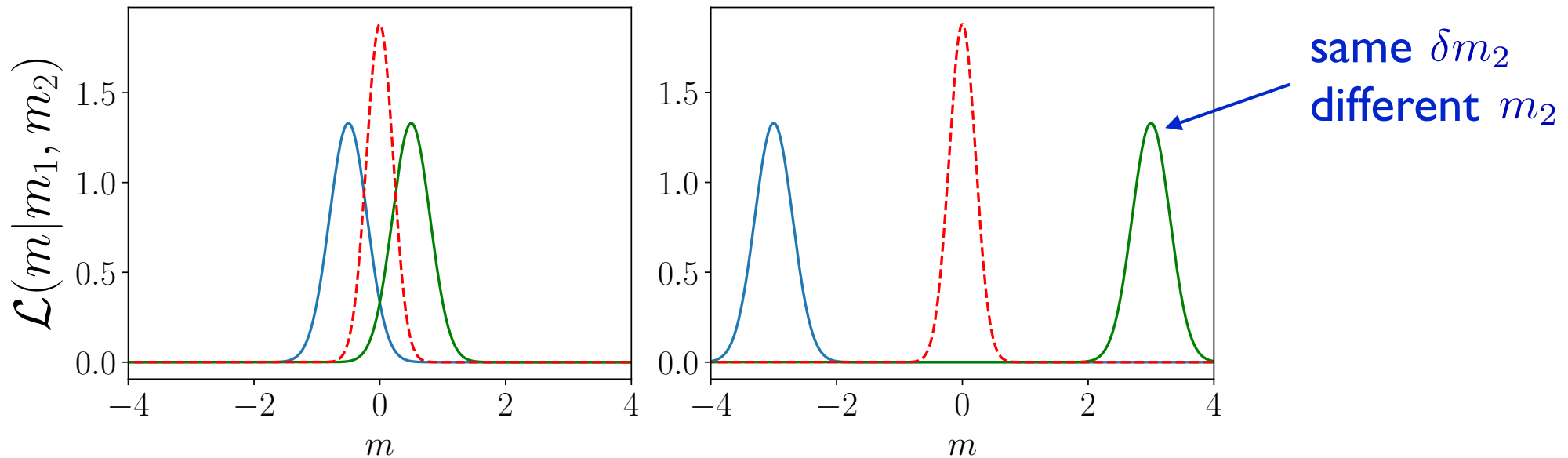
and variance

$$V[m] = H^{-1} = \frac{\delta m_1^2 \delta m_2^2}{\delta m_1^2 + \delta m_2^2}$$

does not
depend on
 $m_1 - m_2$!

Incompatible data sets

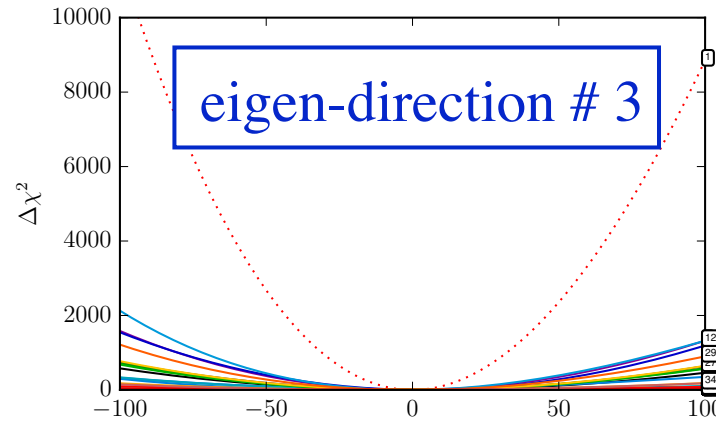
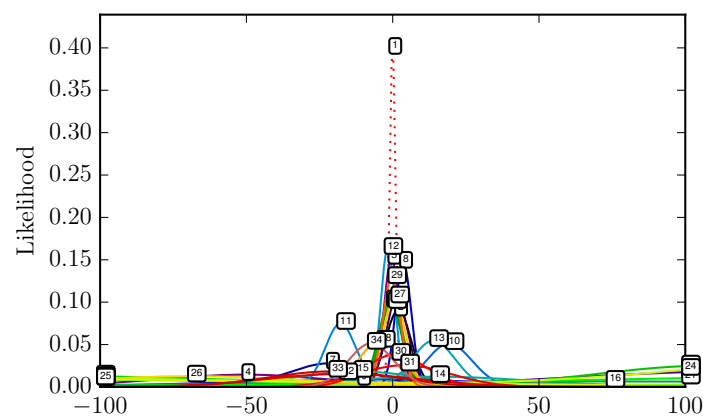
- Simple example: consider observable m , and two measurements $(m_1, \delta m_1)$, $(m_2, \delta m_2)$



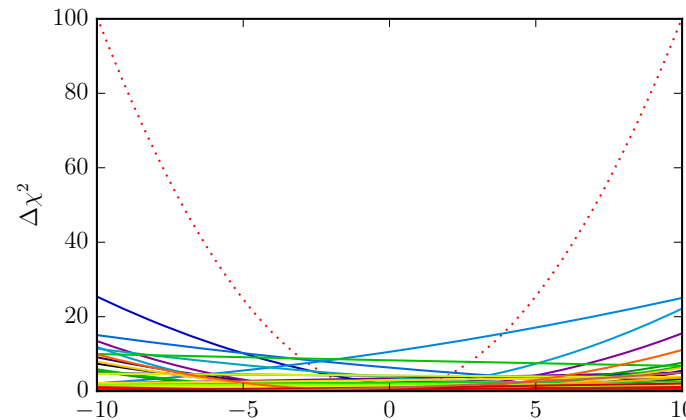
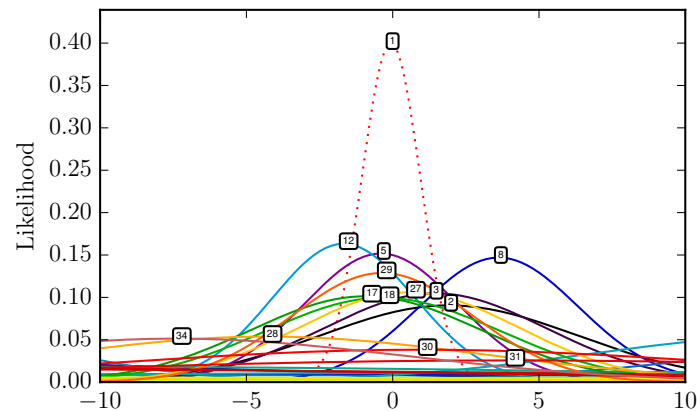
- total uncertainty remains independent of degree of (in)compatibility of data
- Gaussian likelihood gives unrealistic representation of true uncertainty

Incompatible data sets

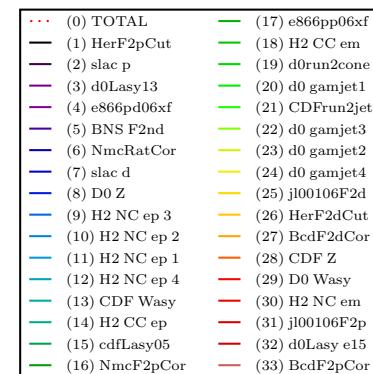
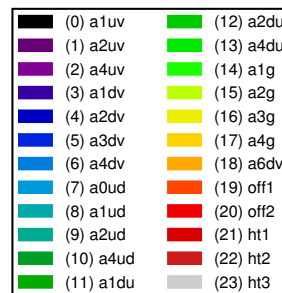
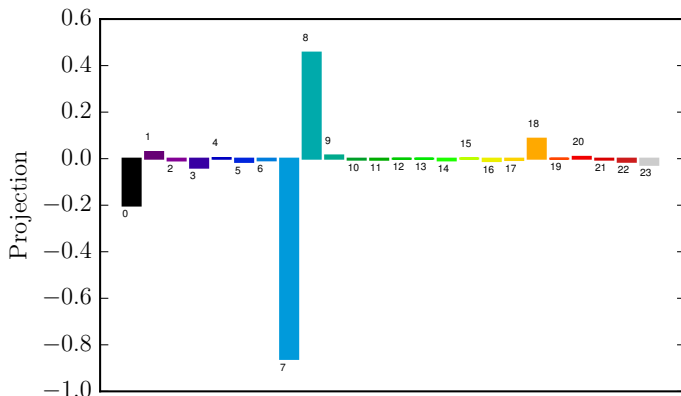
Realistic example: recent CJ (CTEQ-JLab) global PDF analysis



→ 24 parameters,
33 data sets

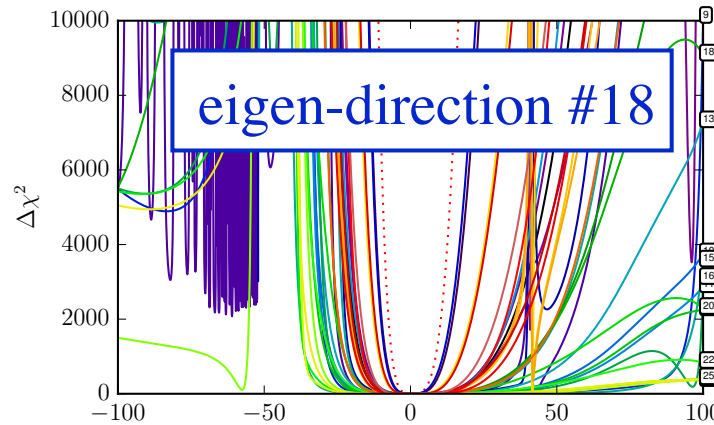
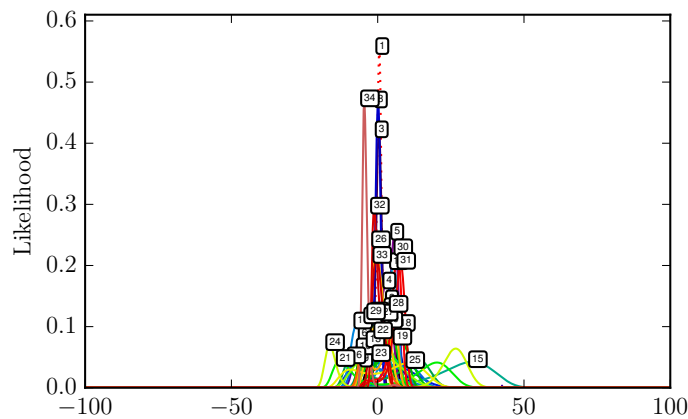


→ data sets
compatible
along this
e-direction

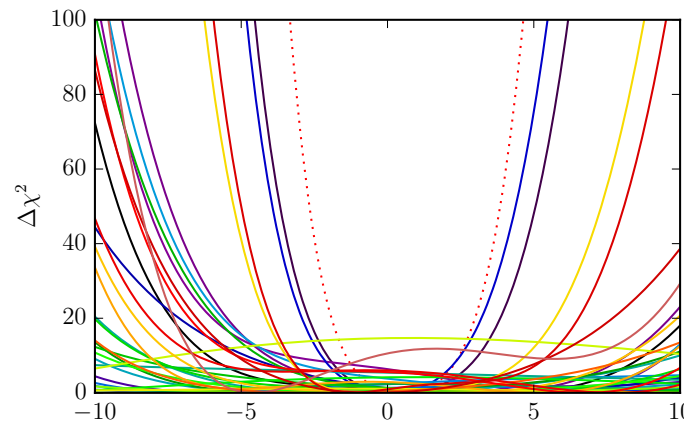
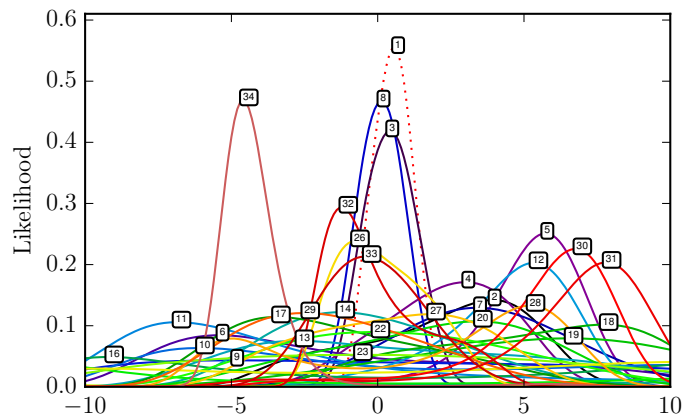


Incompatible data sets

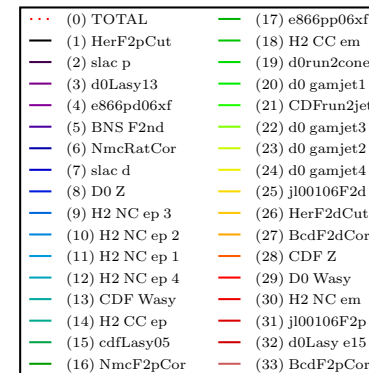
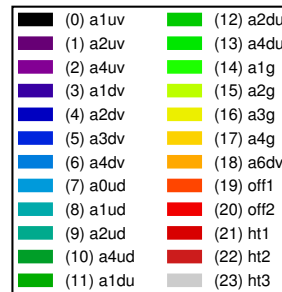
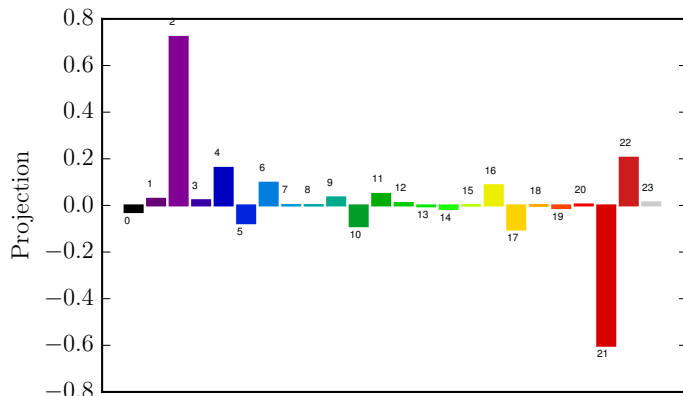
Realistic example: recent CJ (CTEQ-JLab) global PDF analysis



→ 24 parameters,
33 data sets

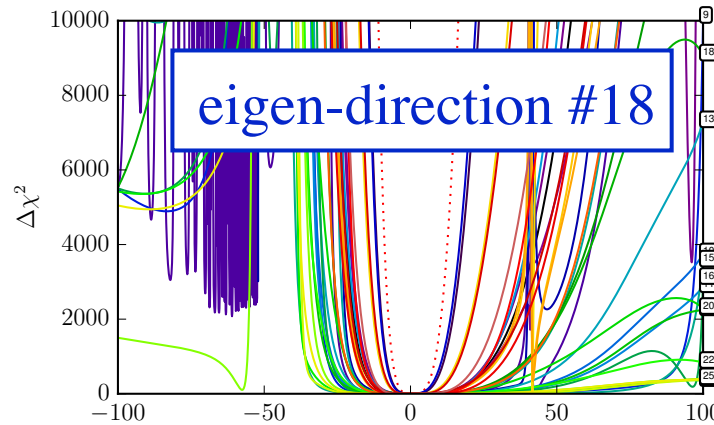
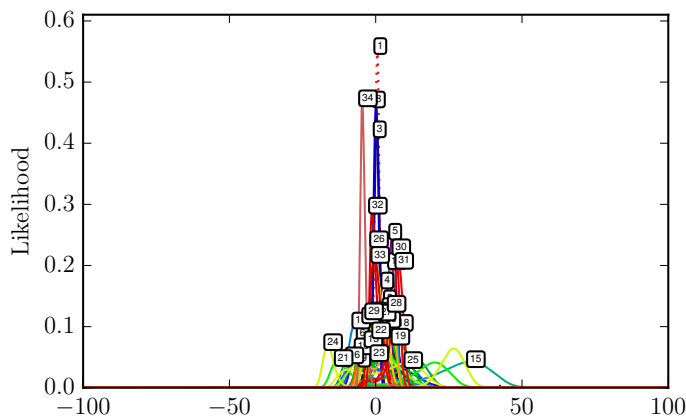


→ data sets not
compatible
along this
e-direction

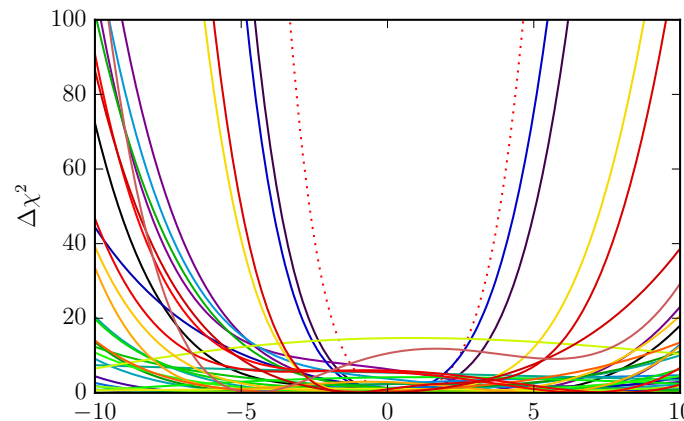
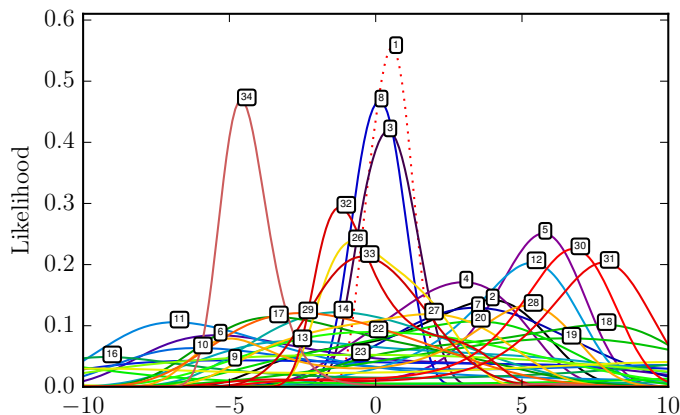


Incompatible data sets

■ Realistic example: recent CJ (CTEQ-JLab) global PDF analysis



→ 24 parameters,
33 data sets

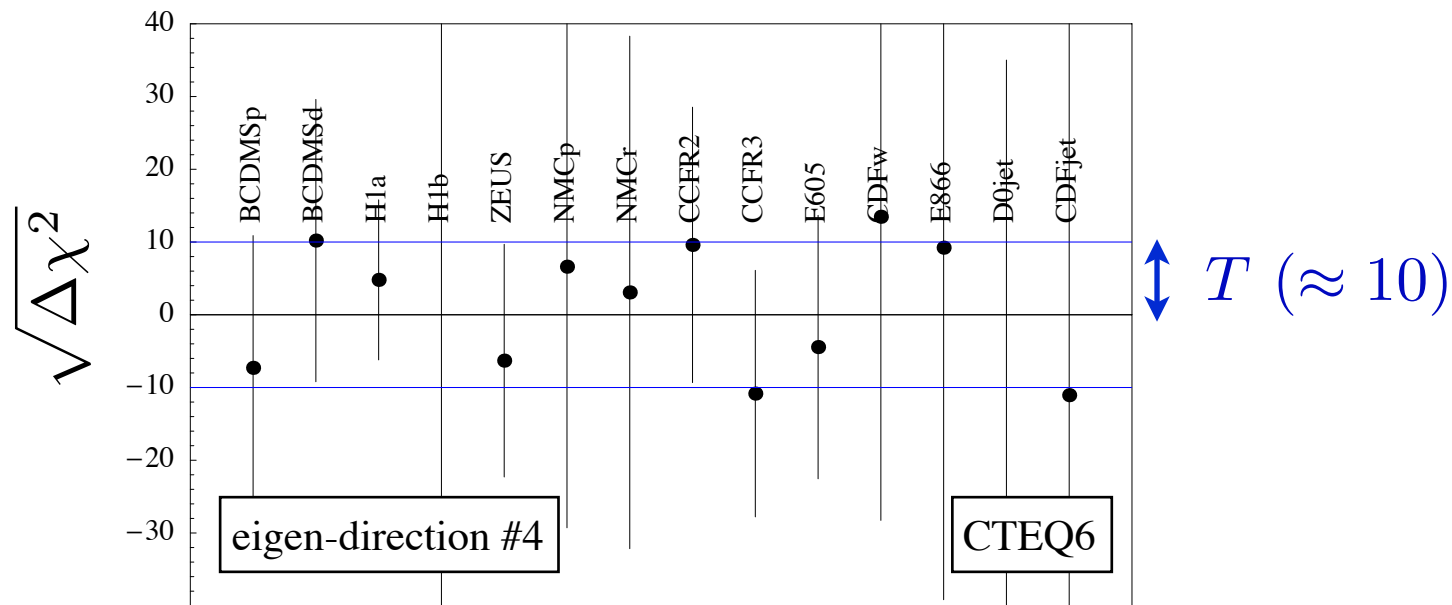


→ data sets not
compatible
along this
e-direction

- standard Gaussian likelihood incapable of accounting for underestimated individual errors (leading to incompatible data sets)
— not designed for such scenarios!

Incompatible data sets

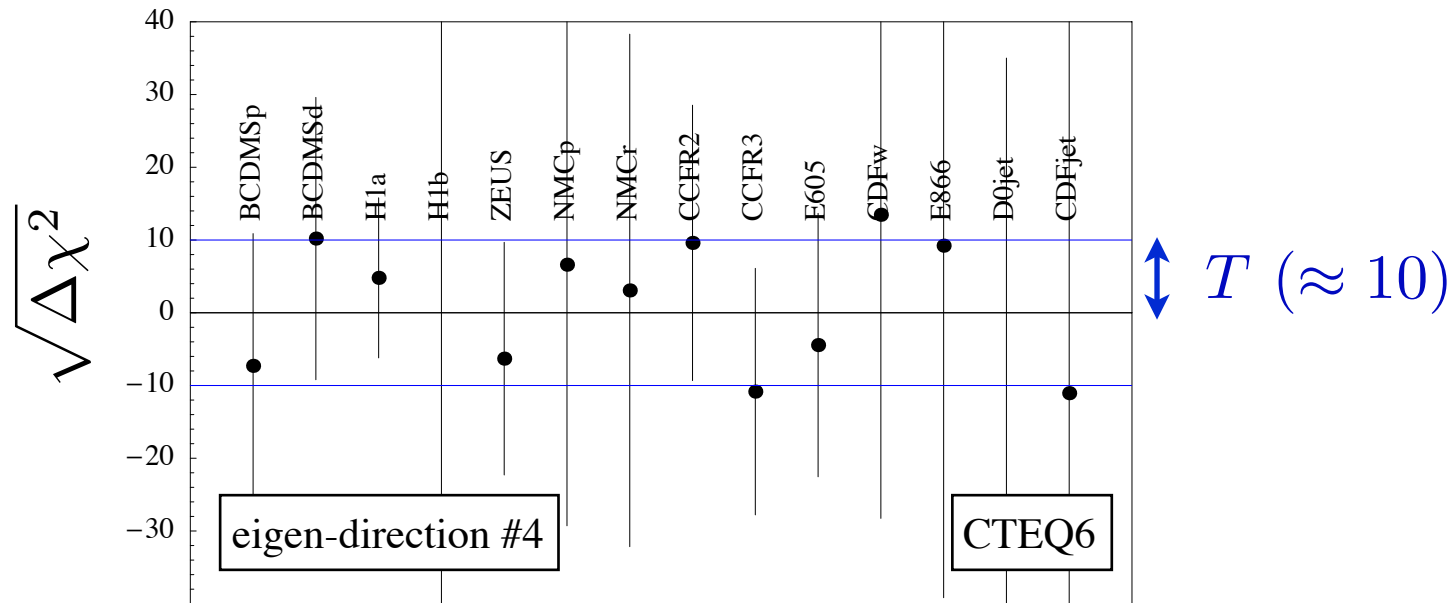
CTEQ tolerance criteria



- for each experiment, find minimum χ^2 along given e-direction
- from χ^2 distribution determine 90% CL for each experiment
- along each side of e-direction, determine maximum range $d_k^\pm$ allowed by the most constraining experiment
- T computed by averaging over all $d_k^\pm$ (typically $T \sim 5 - 10$)

Incompatible data sets

■ CTEQ tolerance criteria



■ This approach is *not consistent* with Gaussian likelihood

→ no clear Bayesian interpretation of uncertainties
(ultimately, a prescription...)

To summarize standard maximum likelihood method...

- Gradient search (in parameter space) depends how “good” the starting point is
 - for ~ 30 parameters trying different starting points is impractical, if do not have some information about shape
- Common to free parameters initially, then freeze those not sensitive to data (χ^2 flat locally)
 - introduces bias, does not guarantee that flat χ^2 globally
- Cannot guarantee solution is unique
- Error propagation characterized by quadratic χ^2 near minimum
 - no guarantee this is quadratic globally (*e.g.* Student t -distribution?)
- Introduction of tolerance modifies Gaussian statistics

Monte Carlo methods

Monte Carlo

- Designed to faithfully compute Bayesian master formulas
- Do not assume a single minimum, include all possible solutions (with appropriate weightings)
- Do not assume likelihood is Gaussian in parameters
- Allows likelihood analysis to be extended to address tensions among data sets via Bayesian inference
- More computationally demanding compared with Hessian method

Monte Carlo

- First group to use MC for global PDF analysis was NNPDF, using neural network to parametrize $P(x)$ in

Forte et al. (2002)

$$f(x) = N x^\alpha (1 - x)^\beta P(x)$$

— α, β are fitted “preprocessing coefficients”

- Iterative Monte Carlo (IMC), developed by JAM Collaboration, variant of NNPDF, tailored to non-neutral net parametrizations

- Markov Chain MC (MCMC) / Hybrid MC (HMC)

— recent “proof of principle” analysis, ideas from lattice QCD

Gbedo, Mangin-Brinet (2017)

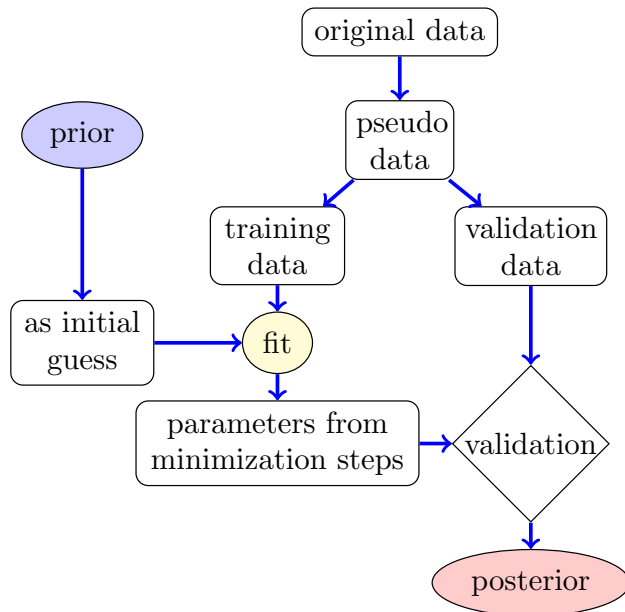
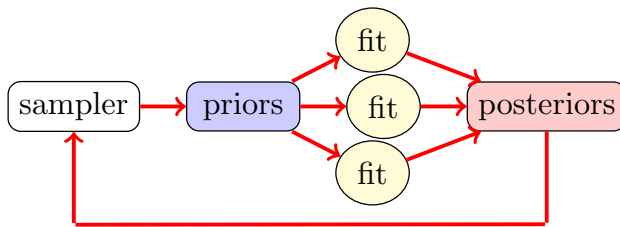
- Nested sampling (NS) — computes integrals in Bayesian master formulas (for E , V , Z) explicitly

Skilling (2004)

Iterative Monte Carlo (IMC)

- Use traditional functional form for input distribution shape, but sample significantly larger parameter space than possible in single-fit analyses

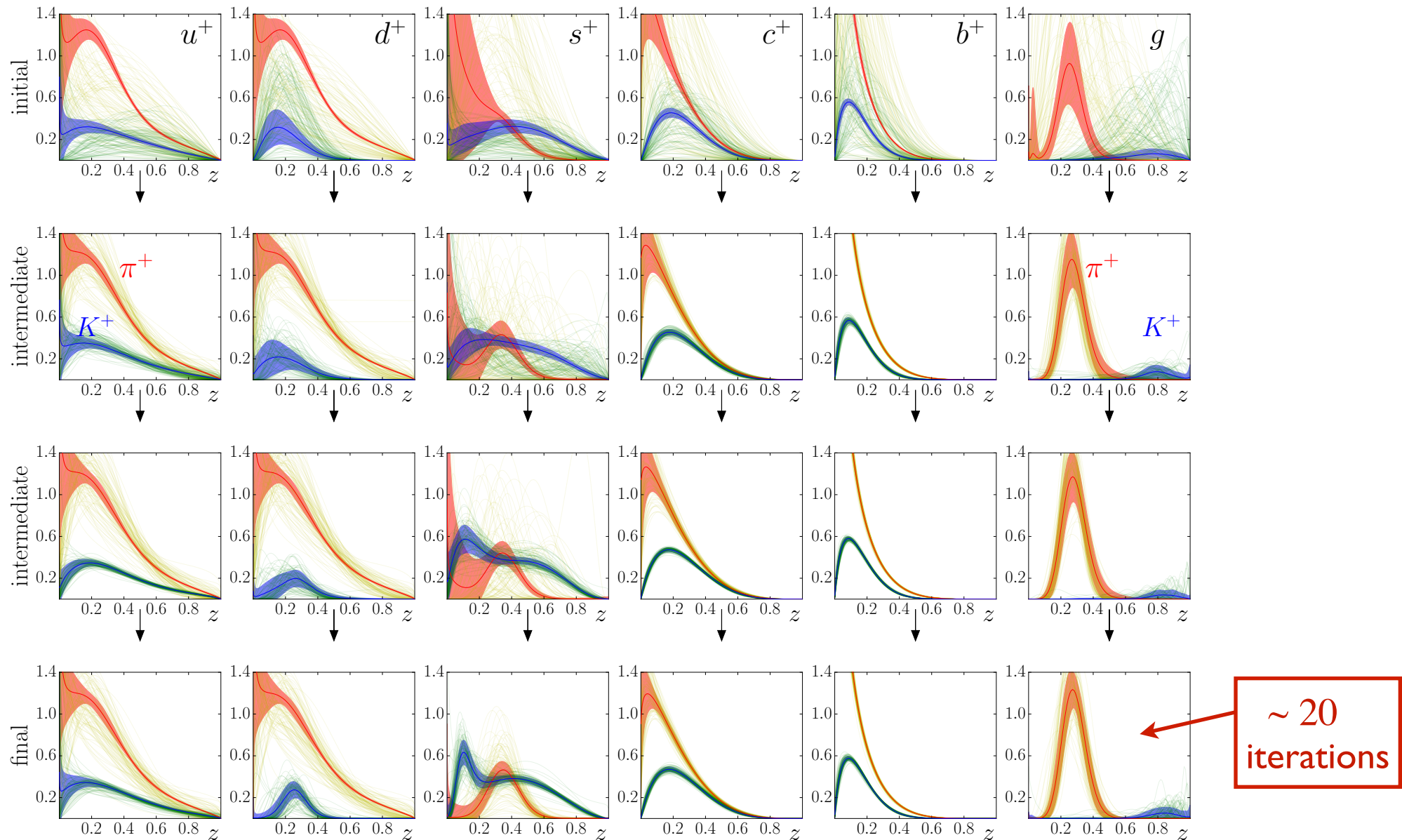
Iterative Monte Carlo (IMC)



- no assumptions for exponents
- cross-validation to avoid overfitting
- iterate until convergence criteria satisfied

Iterative Monte Carlo (IMC)

■ *e.g.* of convergence (for fragmentation functions) in IMC

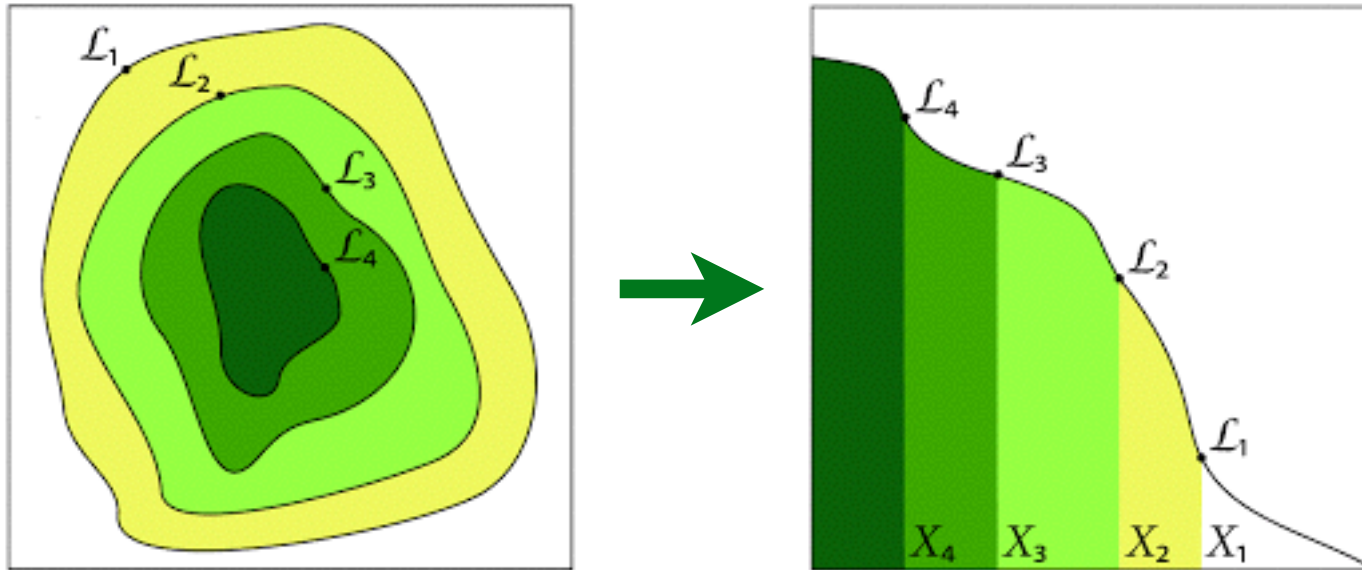


Nested Sampling

- Basic idea: transform n -dimensional integral to 1-D integral

$$Z = \int d^n a \mathcal{L}(\text{data}|\vec{a}) \pi(\vec{a}) = \int_0^1 dX \mathcal{L}(X)$$

where *prior volume* $dX = \pi(\vec{a}) d^n a$



such that $0 < \dots < X_2 < X_1 < X_0 = 1$

Feroz et al.
arXiv:1306.2144 [astro-ph]

Nested Sampling

- Approximate evidence by a weighted sum

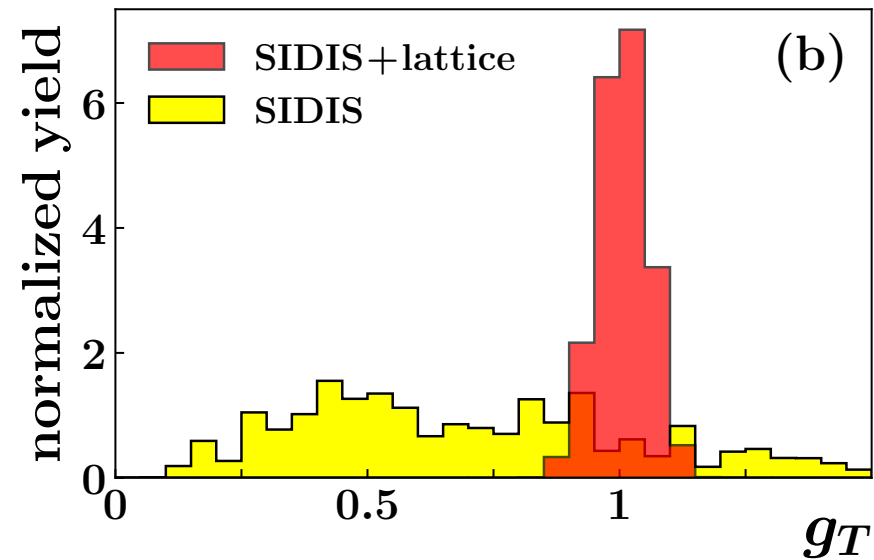
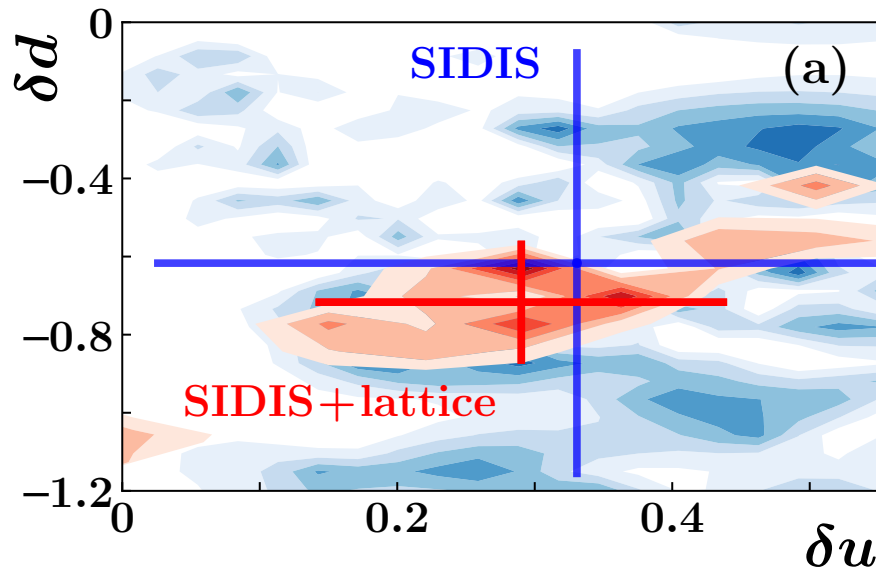
$$Z \approx \sum_i \mathcal{L}_i w_i \quad \text{with weights} \quad w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$$

- Algorithm:

- randomly select samples from full prior s.t. initial volume $X_0 = 1$
- for each iteration, remove point with lowest $\mathcal{L}_i$, replacing it with point from prior with constraint that its $\mathcal{L} > \mathcal{L}_i$
- repeat until entire prior volume has been traversed
- can be parallelized
- performs better than VEGAS for large dimensions
- increasingly used in fields outside of (nuclear) analysis

Nested Sampling

- Recent application in global analysis of transversity TMD PDF
(SIDIS data + lattice QCD constraint on isovector moment)



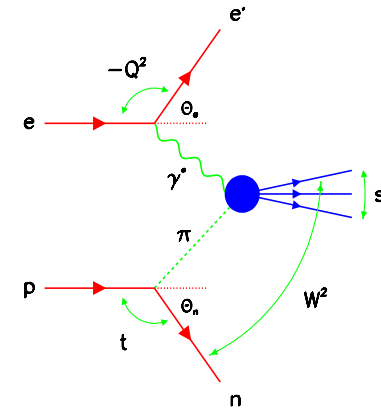
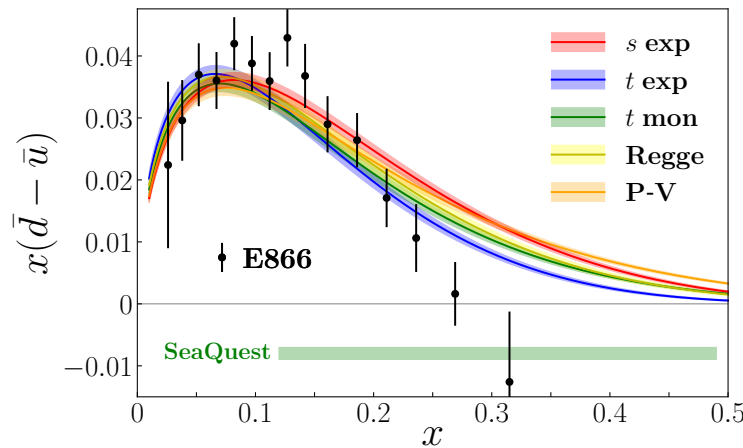
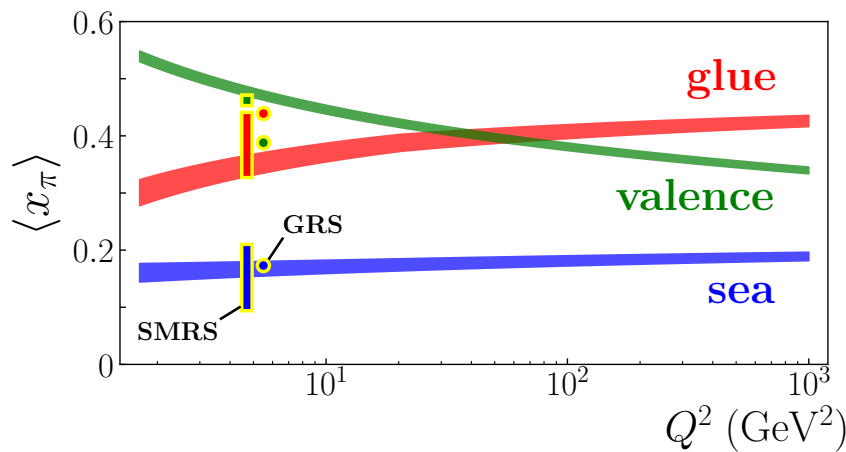
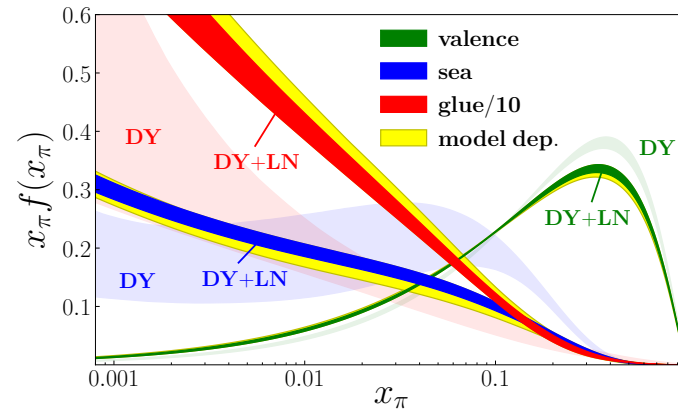
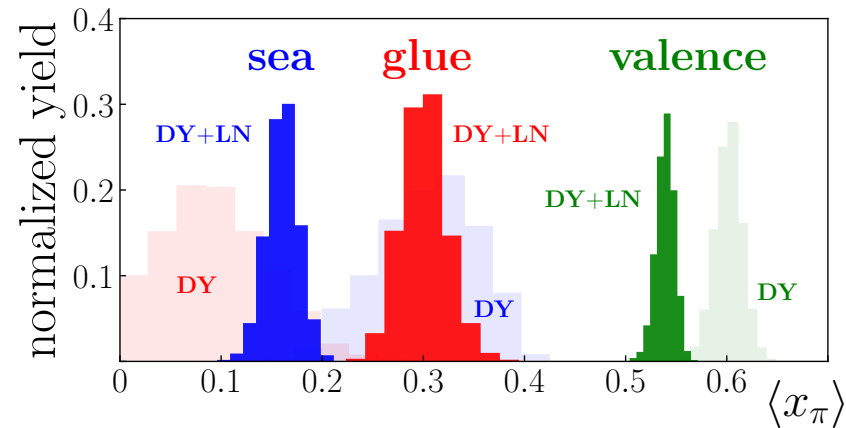
Lin, WM, Prokudin, Sato, Shows (PRL, 2018)

- distributions do not look very Gaussian!
- MC analysis gives $\delta u = 0.3 \pm 0.2$, $\delta d = -0.7 \pm 0.2$ → $g_T = 1.0 \pm 0.1$
- maximum likelihood analysis would have given $g_T \approx 0.5$

Nested Sampling

Most recently applied to global analysis of pion PDFs

(πA Drell-Yan data + leading neutron production at HERA)

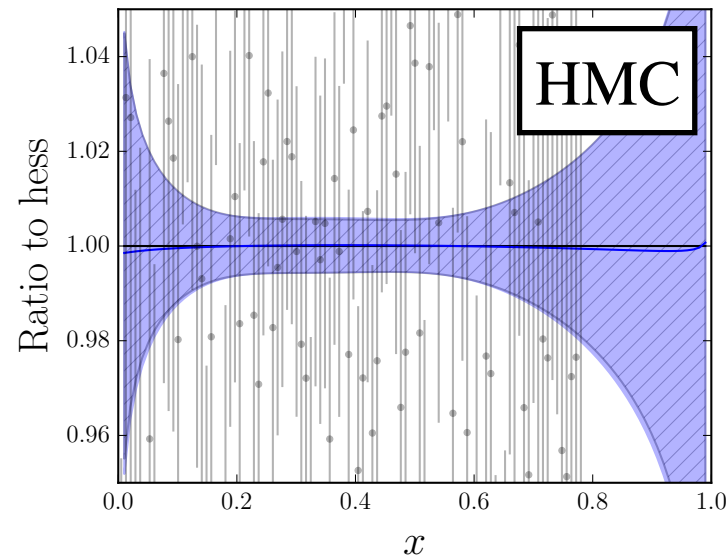
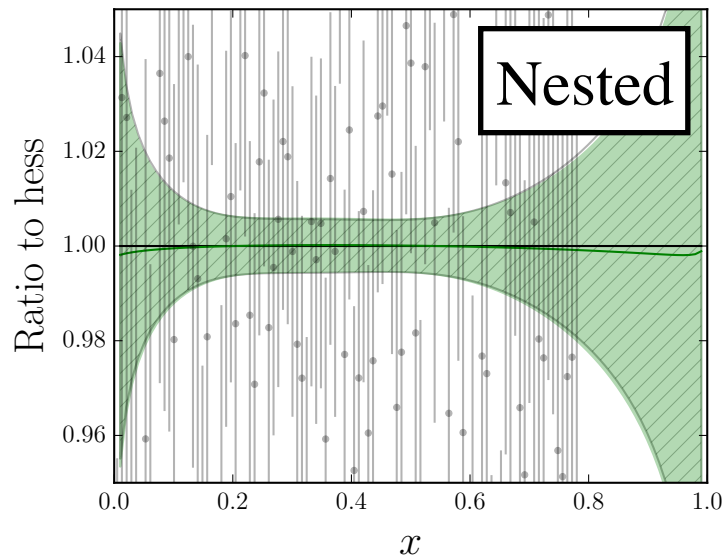
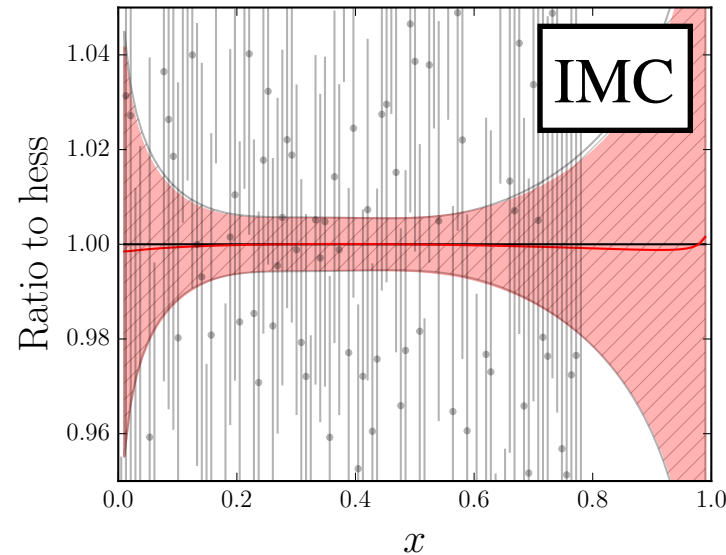
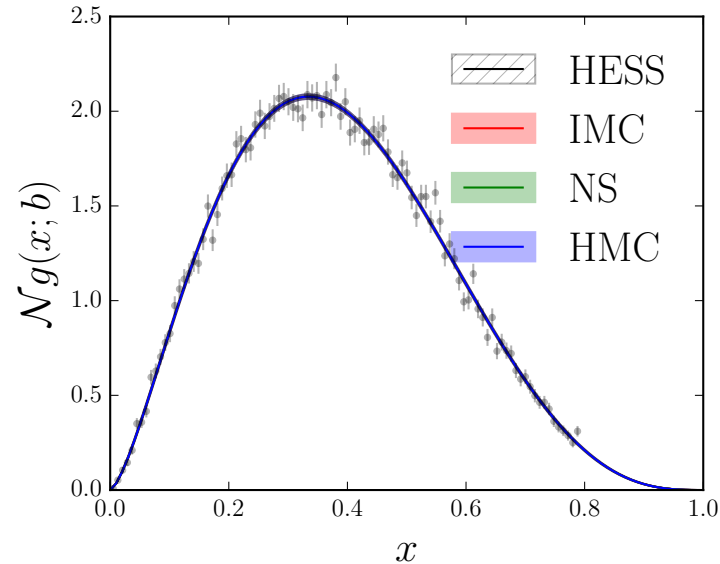


P. Barry, N. Sato, WM, C.-R. Ji (2018)

→ first constraints on sea quark PDFs in the pion

MC Error Analysis

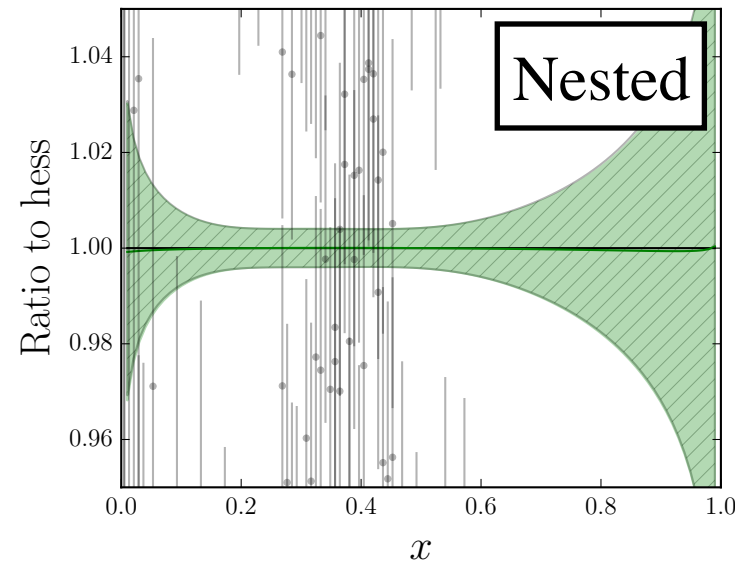
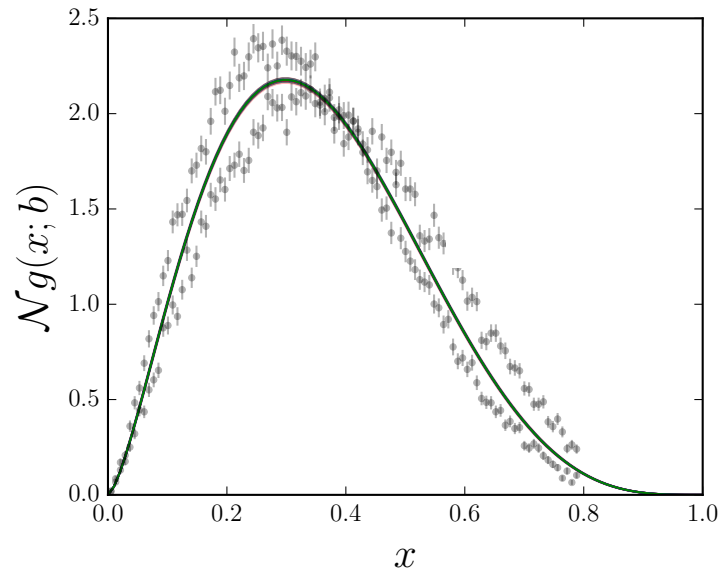
- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function
 - analysis of pseudodata, generated using Gaussian distribution



MC Error Analysis

- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function

→ also for discrepant data



→ almost identical uncertainty bands for Hessian and for MC!

MC Error Analysis

- Assuming a single minimum, a Hessian or MC analysis *must* give same results, if using same likelihood function
- Approaches that use Hessian + tolerance factor not consistent with Gaussian likelihood function
- NNPDF group claim that within their neural net MC methodology, no need for a tolerance factor, since uncertainties similar to other groups who use Hessian + tolerance
 - how can this be? *E. Nocera, A. Accardi, N. Sato, WM, work in progress*
- Assuming sufficient observables to determine PDFs, then PDF uncertainties cannot depend on parametrization!

Non-Gaussian likelihood

Incompatible data sets

- Rigorous (Bayesian) way to address incompatible data sets is to use generalization of Gaussian likelihood
 - joint vs. disjoint distributions
 - empirical Bayes
 - hierarchical Bayes
 - others, used in different fields

Disjoint distributions

- Instead of using total likelihood that is a product (“and”) of individual likelihoods, *e.g.* for simple example of two measurements

$$\mathcal{L}(m_1 m_2 | m; \delta m_1 \delta m_2) = \mathcal{L}(m_1 | m; \delta m_1) \times \mathcal{L}(m_2 | m; \delta m_2)$$

use instead sum (“or”) of individual likelihoods

$$\mathcal{L}(m_1 m_2 | m; \delta m_1 \delta m_2) = \frac{1}{2} \left[\mathcal{L}(m_1 | m; \delta m_1) + \mathcal{L}(m_2 | m; \delta m_2) \right]$$

→ gives rather different expectation value and variance

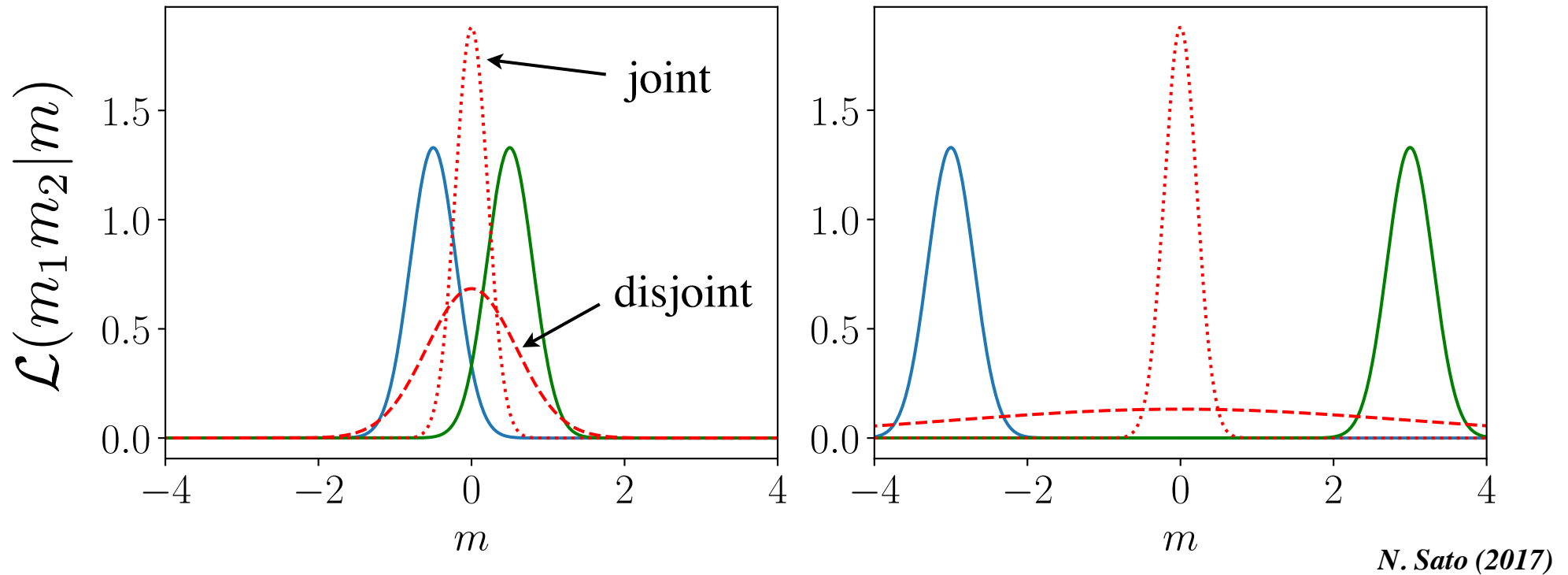
$$E[m] = \frac{1}{2} (m_1 + m_2)$$

$$V[m] = \frac{1}{2} (\delta m_1^2 + \delta m_2^2) + \left(\frac{m_1 - m_2}{2} \right)^2$$

depends on
separation!

Disjoint distributions

■ Symmetric uncertainties $\delta m_1 = \delta m_2$

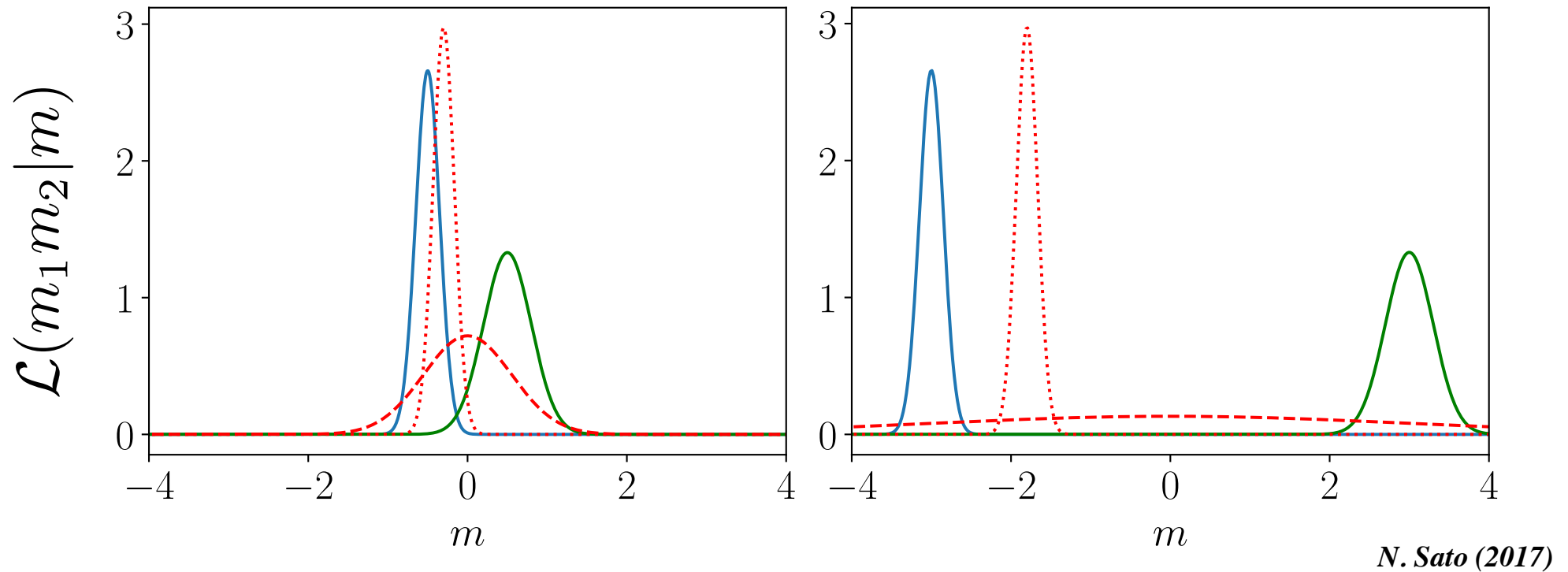


disjoint:
$$V[m] = \frac{1}{2}(\delta m_1^2 + \delta m_2^2) + \left(\frac{m_1 - m_2}{2}\right)^2$$

joint:
$$V[m] = \frac{\delta m_1^2 \delta m_2^2}{\delta m_1^2 + \delta m_2^2}$$

Disjoint distributions

- Asymmetric uncertainties $\delta m_1 \neq \delta m_2$



→ disjoint likelihood gives broader overall uncertainty, overlapping individual (discrepant) data

Empirical Bayes

- Shortcoming of conventional Bayesian — still assume prior distribution follows specific form (*e.g.* Gaussian)
- Extend approach to more fully represent prior uncertainties, with final uncertainties that do not depend on initial choices
- In generalized approach, data uncertainties modified by distortion parameters, whose probability distributions given in terms of “hyperparameters” (or “nuisance parameters”)
- Hyperparameters determined from data
 - give posteriors for both PDF and hyperparameters

Empirical Bayes

- Standard mean and variance that characterize data

$$\theta = \mu + \sigma \longrightarrow f(\mu) + g(\sigma)$$

where $f(\mu), g(\sigma)$ are unknown functions that account for faulty measurements

- Simple choice is

$$(\mu, \sigma) \rightarrow (\zeta_1 \mu + \zeta_2, \zeta_3 \sigma)$$

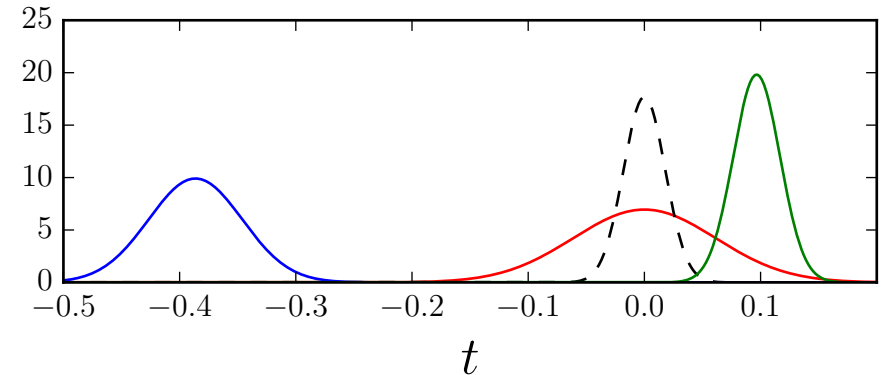
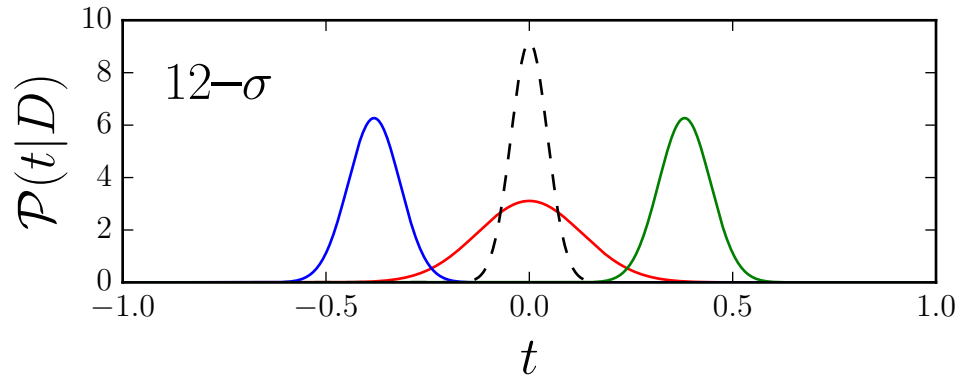
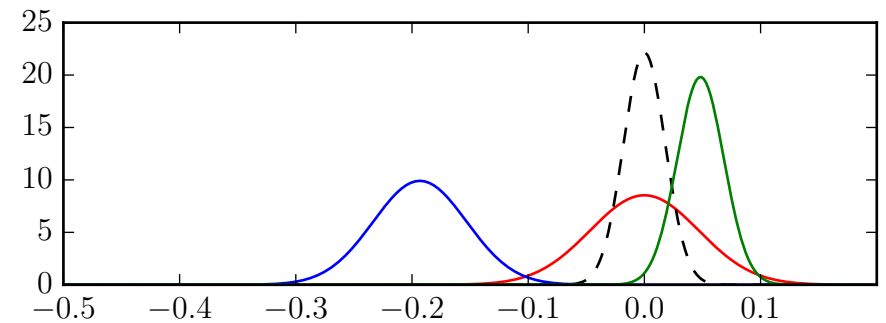
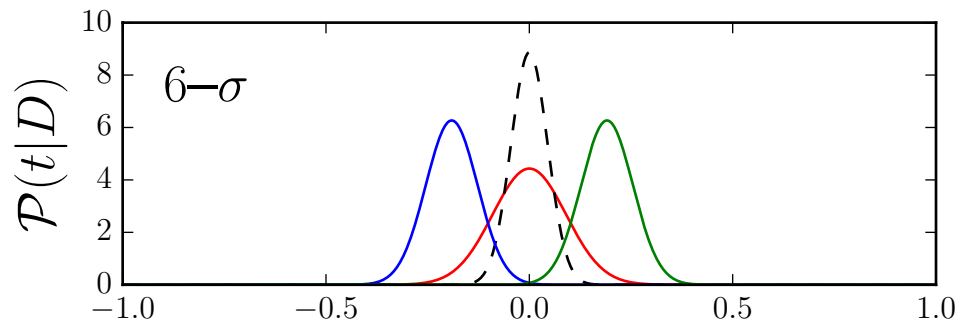
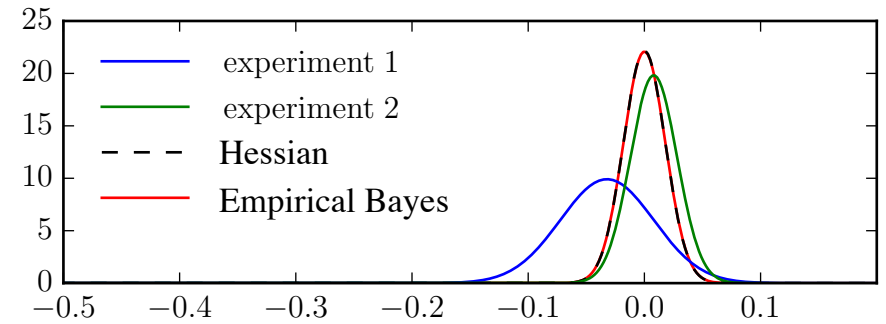
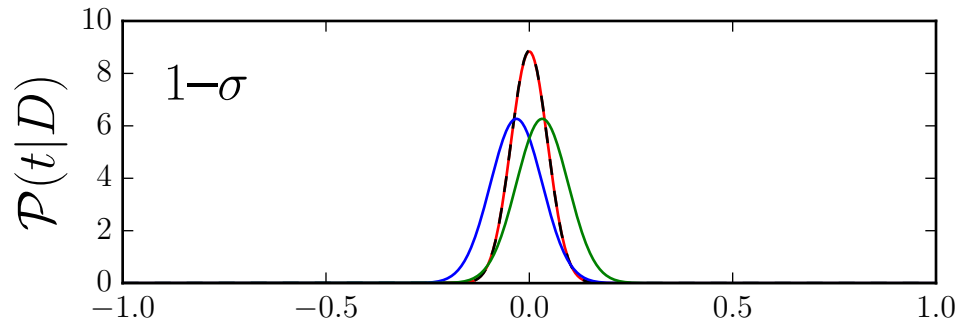
where $\zeta_{1,2,3}$ are distortion parameters, with prob. dists. described by hyperparameters $\phi_{1,2,3}$

- Likelihood function is then

$$\mathcal{L}(\text{data}|\vec{a}, \zeta_{1,2,3}) \sim \exp \left[-\frac{1}{2} \sum_i \left(\frac{d_i - f(\mu_i(\vec{a}, \zeta_{1,2}))}{g(\sigma, \zeta_3)} \right)^2 \right] \pi_1(\zeta_1|\phi_1) \pi_2(\zeta_2|\phi_2) \pi_3(\zeta_3|\phi_3)$$

Empirical Bayes

■ Simple example of EB for symmetric & asymmetric errors



N. Sato (2017)

Outlook

- New approaches being developed for global QCD analysis
 - simultaneous determination of parton distributions using Monte Carlo sampling of parameter space
- Treatment of discrepant data sets needs serious attention
 - Bayesian perspective has clear merits
- Near-term future: “universal” QCD analysis of all observables sensitive to collinear (unpolarized & polarized) PDFs and FFs
- Longer-term: apply MC technology to global QCD analysis of transverse momentum dependent (TMD) PDFs and FFs



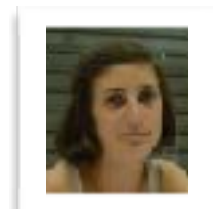
Nobuo



Jake



Alberto



Carlota

...

A close-up photograph of a triangular slice of chocolate cake resting on a white plate. The cake has multiple layers of dark brown chocolate cake and white frosting. A single, bright red chili pepper with a green stem is placed on top of the white frosting. The equation $E = mc^2$ is overlaid in orange text on the white frosting. The background shows a yellow plate and a brown surface.
$$E = mc^2$$