# QCD theory and machine learning for global analysis

**Nobuo Sato**
ODU

**Supported by CNF19-06**

Femtography 2019
Washington DC

# Personnel

- **QCD theorists**
  - Nobuo Sato **PI** (ODU)
  - Wally Melnitchouk **Co-PI** (JLab)
  - Andreas Metz (Temple University)
  - Ian Cloët (ANL)

- **Machine learning scientists**
  - Yaohang Li (ODU)
  - Yasir Awadh Alanazi (ODU)
  - Manal Almaeen (ODU)

  - Michelle Kuchera (Davidson College)
  - Raghu Ramanujan (Davidson College)
  - Meg Houck (Davidson College)
  - Eleni Tsitinidi (Davidson College)

# Motivations

To build the next generation of **global QCD analysis** tools using **machine learning** (ML) techniques to study the **quantum probability distributions** (QPD) characterizing the internal structure of the nucleon.

# List of Proposed Milestones and Deliverables

- Prototypes of networks that map collinear PDFs into inclusive electron–proton scattering observables

- Web interface for user access (visualization)

- TMDs and GPDs from nonperturbative models

# Progress (May 10 – Aug 12, 2019)

- **Machine learning prototypes**
  - Prototypes for the NN inverse mappings
  - Validation tests to quantify quality of the mappings
- **QCD models**
  - Numerical implementation of DDY cross section using model GTMD (under development)

## Universality $\rightarrow$ the predictive power of QCD

- lepton-hadron reactions (COMPASS, JLab, **EIC**)

$$\sigma^{\mathrm{EXP}}_{l+P\to l+H+X} = \boxed{C_{l+k\to l+k+X}} \otimes \boxed{\mathrm{PDF}_P} \otimes \boxed{\mathrm{FF}_H}$$

## Universality $\rightarrow$ the predictive power of QCD

- lepton-hadron reactions (COMPASS, JLab, **EIC**)

$$\sigma^{\mathrm{EXP}}_{l+P\to l+H+X} = \boxed{C_{l+k\to l+k+X}} \otimes \boxed{\mathrm{PDF}_P} \otimes \boxed{\mathrm{FF}_H}$$

- hadron-hadron reactions (LHC)

$$\sigma^{\mathrm{EXP}}_{P+P\to l+\bar{l}+X} = \boxed{C_{k+k\to l+\bar{l}+X}} \otimes \boxed{\mathrm{PDF}_P} \otimes \boxed{\mathrm{PDF}_P}$$

# Universality $\rightarrow$ the predictive power of QCD

- lepton-hadron reactions (COMPASS, JLab, **EIC**)

$$\sigma^{\mathrm{EXP}}_{l+P\to l+H+X} = \boxed{C_{l+k\to l+k+X}} \otimes \boxed{\mathrm{PDF}_P} \otimes \boxed{\mathrm{FF}_H}$$

- hadron-hadron reactions (LHC)

$$\sigma^{\mathrm{EXP}}_{P+P\to l+\bar{l}+X} = \boxed{C_{k+k\to l+\bar{l}+X}} \otimes \boxed{\mathrm{PDF}_P} \otimes \boxed{\mathrm{PDF}_P}$$

- lepton-lepton reactions (Belle)

$$\sigma^{\mathrm{EXP}}_{l+\bar{l}\to H+X} = \boxed{C_{l+\bar{l}\to k+X}} \otimes \boxed{\mathrm{FF}_H}$$

# Global analysis in a nutshell

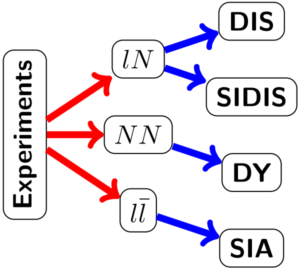
Experiments

$lN$

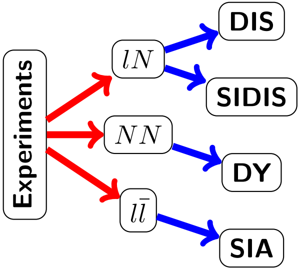
$NN$

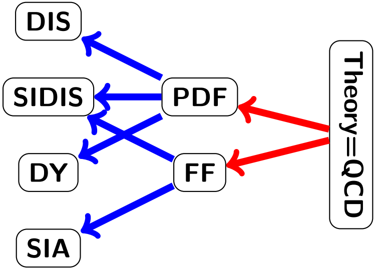
$l\bar{l}$

# Global analysis in a nutshell

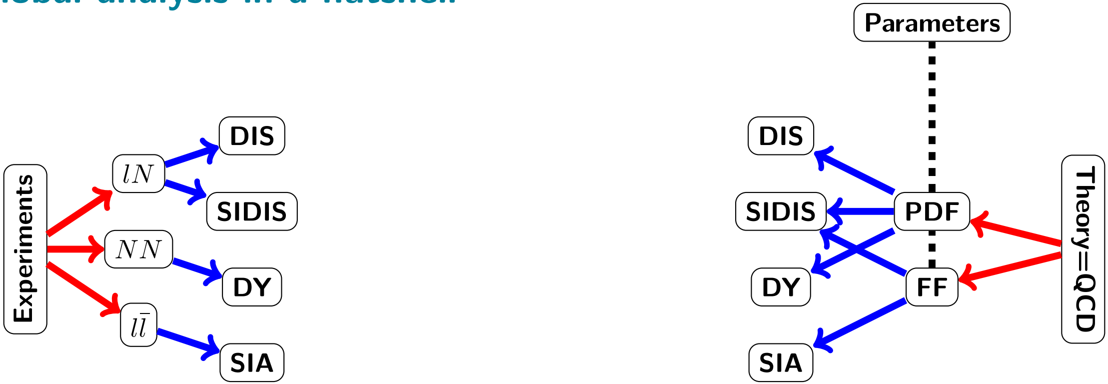


# Global analysis in a nutshell

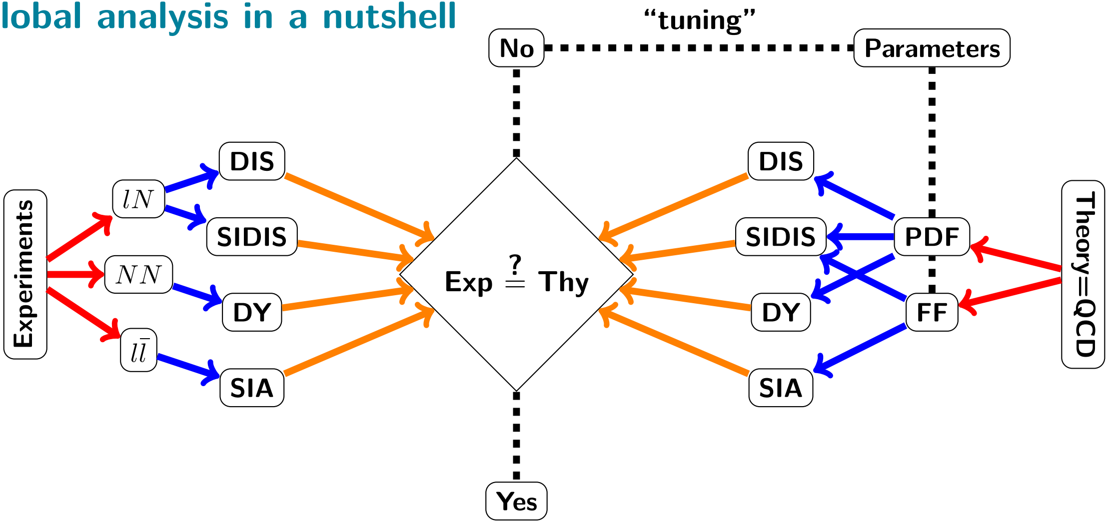


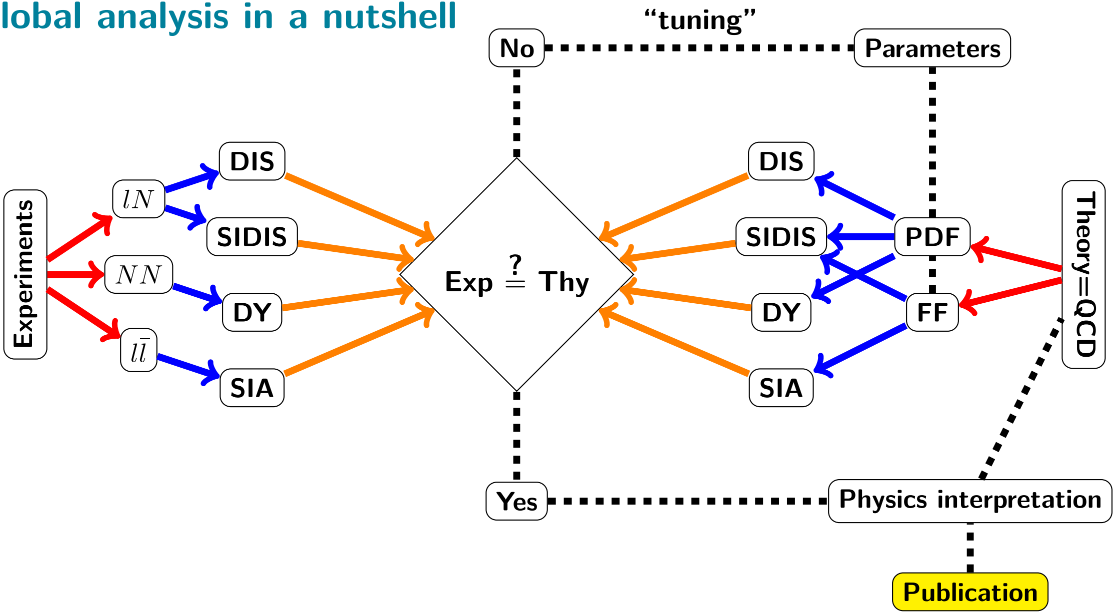
# Global analysis in a nutshell



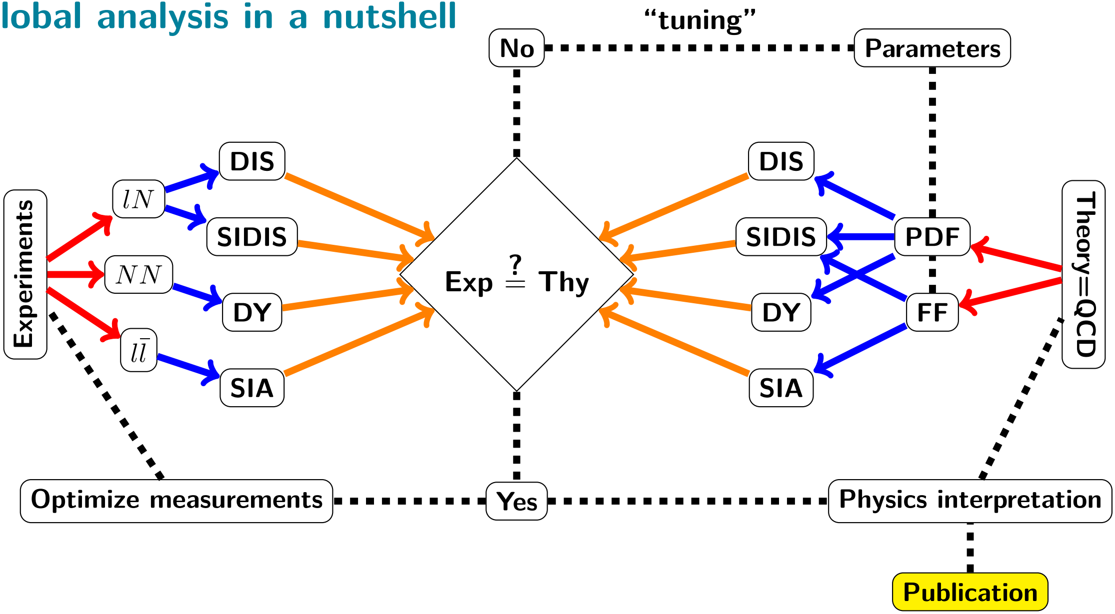
# Global analysis in a nutshell



# Global analysis in a nutshell



## Global analysis in a nutshell



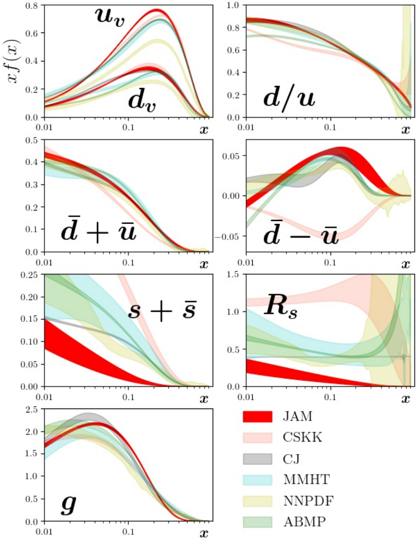
# **JAM19:** *Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions*

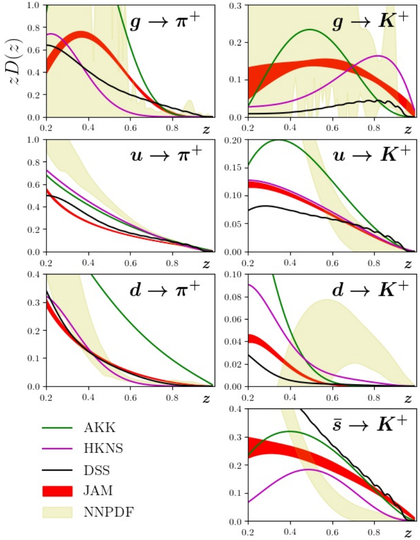
NS, Carlota Andres, Jake Ethier,
Wally Melnitchouk

arXiv:1905.03788

Jefferson Lab Angular Momentum Collaboration

# PDFs and FFs


$u_v$
$d_v$
$xf(x)$
$d/u$
$\bar{d}+\bar{u}$
$\bar{d}-\bar{u}$
$s+\bar{s}$
$R_s$
$g$
JAM
CSKK
CJ
MMHT
NNPDF
ABMP


$zD(z)$
$g \to \pi^+$
$g \to K^+$
$u \to \pi^+$
$u \to K^+$
$d \to \pi^+$
$d \to K^+$
$\bar{s} \to K^+$
AKK
HKNS
DSS
JAM
NNPDF

# What are the challenges?

# What are the challenges?

**Cross sections** → **Link** → **QPD**

# What are the challenges?

Cross sections $\rightarrow$ Link $\rightarrow$ QPD

- **The link is not unique**

# What are the challenges?

**Cross sections** → **Link** → **QPD**

- **The link is not unique**
  - choice of parametrization

## What are the challenges?

**Cross sections** → **Link** → **QPD**

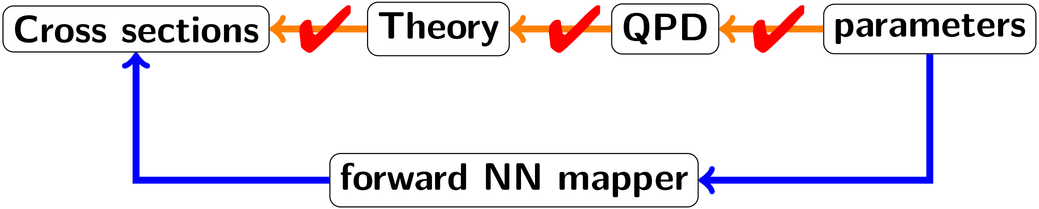
- **The link is not unique**
  - choice of parametrization
  - factorization accuracy

# What are the challenges?

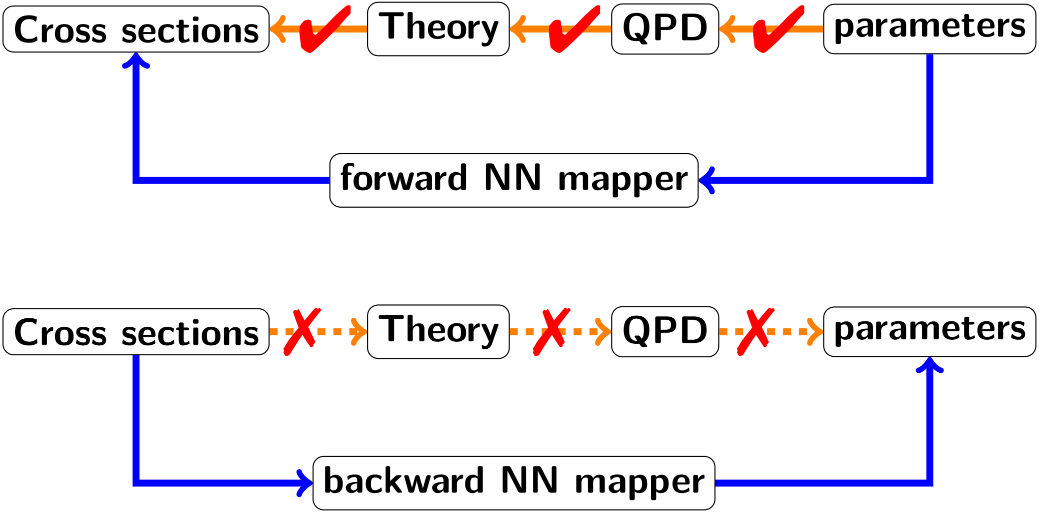
**Cross sections** → **Link** → **QPD**

- **The link is not unique**
  - choice of parametrization
  - factorization accuracy
  - treatment of experimental uncertainties

# ML for global analysis



# ML for global analysis



# Why is this useful?

# Why is this useful?

- **Backward mapper**

# Why is this useful?

- **Backward mapper**
  - What data sets or data points constrain specific QPD?

# Why is this useful?

- **Backward mapper**
  - What data sets or data points constrain specific QPD?
  - How do QPDs depend on the choice of parametrization?

# Why is this useful?

- **Backward mapper**
  - What data sets or data points constrain specific QPD?
  - How do QPDs depend on the choice of parametrization?
  - What additional measurements are needed to resolve QPDs?

## Why is this useful?

- **Backward mapper**
  - What data sets or data points constrain specific QPD?
  - How do QPDs depend on the choice of parametrization?
  - What additional measurements are needed to resolve QPDs?
- **Forward mapper**

# Why is this useful?

- **Backward mapper**
  - What data sets or data points constrain specific QPD?
  - How do QPDs depend on the choice of parametrization?
  - What additional measurements are needed to resolve QPDs?
- **Forward mapper**
  - Make predictions for future experiments (cross section rates)

# Progress

- First prototypes of backward mappers

# Progress

- First prototypes of backward mappers

- Validation in a controlled test model

# Progress

- First prototypes of backward mappers
- Validation in a controlled test model
- Blind control test gives accuracy under $1$–$\sigma$ CL

# Progress

- First prototypes of backward mappers
- Validation in a controlled test model
- Blind control test gives accuracy under $1$–$\sigma$ CL
- Code development of Wigner distributions $\rightarrow$ double DY observable

# Test model 1D QPD

- Parametrization of QPD

$$u(x,Q^2) = N_u(Q^2)\, x^{\alpha_u(Q^2)}(1-x)^{\beta_u(Q^2)}(1+\gamma_u(Q^2)\sqrt{x}+\delta_u(Q^2)\, x),$$
$$d(x,Q^2) = N_d(Q^2)\, x^{\alpha_d(Q^2)}(1-x)^{\beta_d(Q^2)}\,(1+\gamma_d(Q^2)\sqrt{x}+\delta_d(Q^2)\, x),$$

# Test model 1D QPD

- Parametrization of QPD

$$
\begin{aligned}
u(x,Q^2) &= N_u(Q^2)\, x^{\alpha_u(Q^2)}(1-x)^{\beta_u(Q^2)}(1+\gamma_u(Q^2)\sqrt{x}+\delta_u(Q^2)\,x),\\
d(x,Q^2) &= N_d(Q^2)\, x^{\alpha_d(Q^2)}(1-x)^{\beta_d(Q^2)}\,(1+\gamma_d(Q^2)\sqrt{x}+\delta_d(Q^2)\,x),\\
p^{(i)} &= \{N^i_{u,d},\, \alpha^i_{u,d},\, \beta^i_{u,d},\, \gamma^i_{u,d},\, \delta^i_{u,d}\}\\
p(Q^2) &= p^{(0)} + p^{(1)} s(Q^2)\\
s(Q^2) &= \log\left(\frac{\log(Q^2/\Lambda^2_{\mathrm{QCD}})}{\log(Q_0^2/\Lambda^2_{\mathrm{QCD}})}\right).
\end{aligned}
$$

# Test model 1D QPD

- Parametrization of QPD

$$
\begin{aligned}
u(x,Q^2) &= N_u(Q^2)\, x^{\alpha_u(Q^2)}(1-x)^{\beta_u(Q^2)}(1+\gamma_u(Q^2)\sqrt{x}+\delta_u(Q^2)\,x),\\
d(x,Q^2) &= N_d(Q^2)\, x^{\alpha_d(Q^2)}(1-x)^{\beta_d(Q^2)}\,(1+\gamma_d(Q^2)\sqrt{x}+\delta_d(Q^2)\,x),\\
p^{(i)} &= \{N^i_{u,d},\ \alpha^i_{u,d},\ \beta^i_{u,d},\ \gamma^i_{u,d},\ \delta^i_{u,d}\}\\
p(Q^2) &= p^{(0)} + p^{(1)} s(Q^2)\\
s(Q^2) &= \log\left(\frac{\log(Q^2/\Lambda^2_{\mathrm{QCD}})}{\log(Q_0^2/\Lambda^2_{\mathrm{QCD}})}\right).
\end{aligned}
$$

- The observable

$$
F_{p,d} = F_{p,d}(x,Q^2;u,d)
$$

# The training samples

Cross sections $\leftarrow$ Theory $\leftarrow$ QPD $\leftarrow$ parameters

# The training samples

Cross sections $\leftarrow$ Theory $\leftarrow$ QPD $\leftarrow$ parameters

Cross sections 1

parameters 1

# The training samples

Cross sections ← Theory ← QPD ← parameters

Cross sections 1

parameters 1

Cross sections 2

parameters 2

# The training samples

Cross sections ← Theory ← QPD ← parameters

Cross sections 1

Cross sections 2

Cross sections 3

parameters 1

parameters 2

parameters 3

# The training samples

Cross sections ← Theory ← QPD ← parameters

Cross sections 1

Cross sections 2

Cross sections 3

Cross sections N

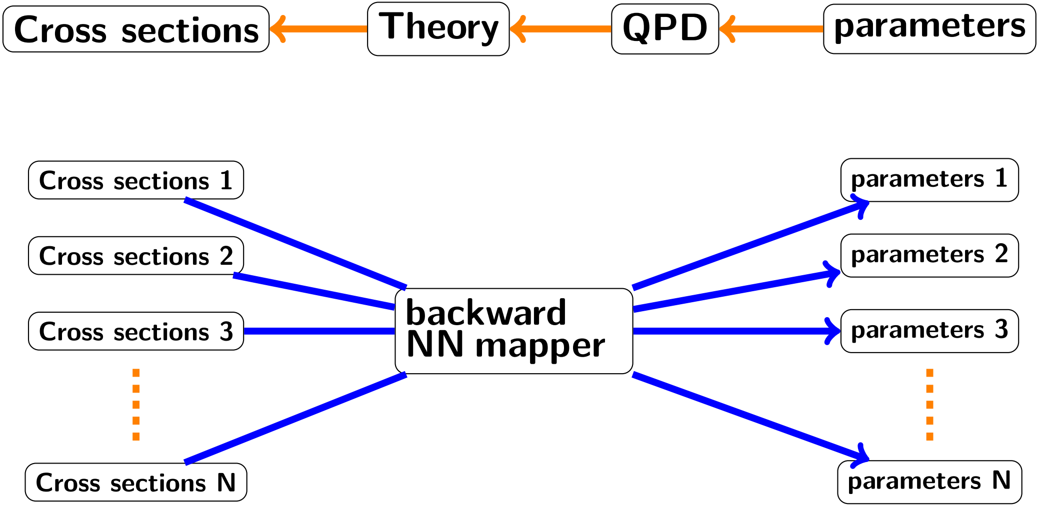
parameters 1

parameters 2

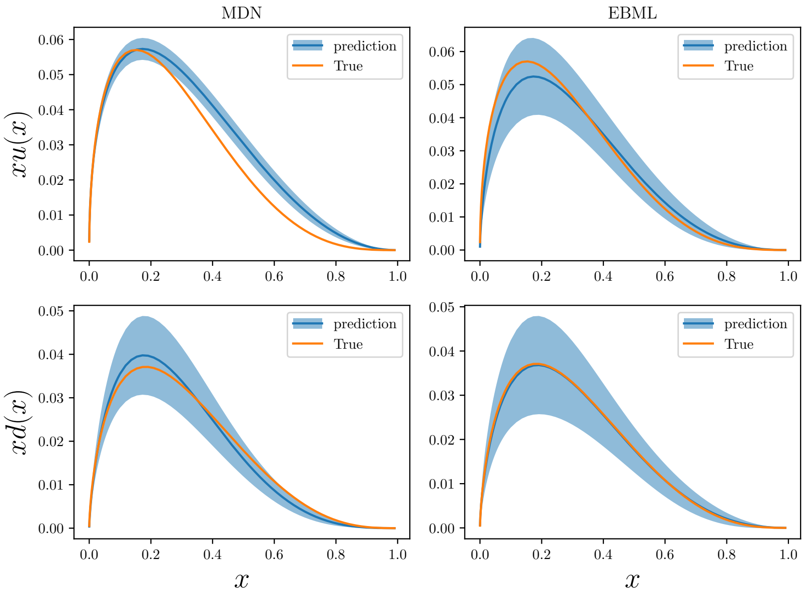
parameters 3

parameters N

# The training samples



# Blind control test




- Error bands from data resampling
- EBML give accurate predictions within $1\sigma$ CL

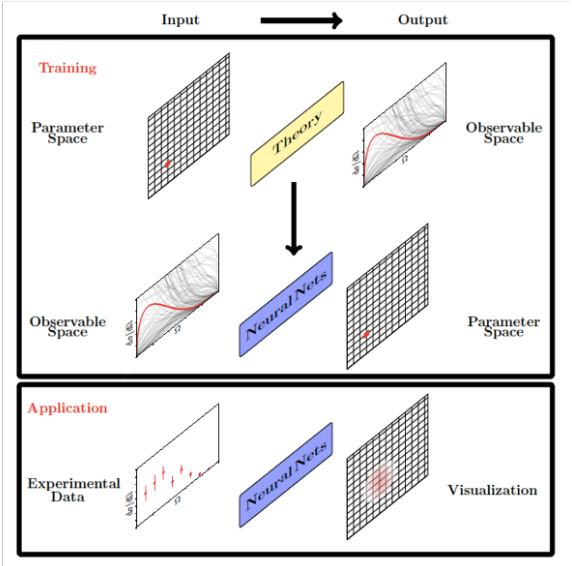
*Mapping experimental observables to quantum probability distributions*

# Machine Learning Architectures for predicting QPDs

CNF
JLab Theory Center
Old Dominion University
Davidson College

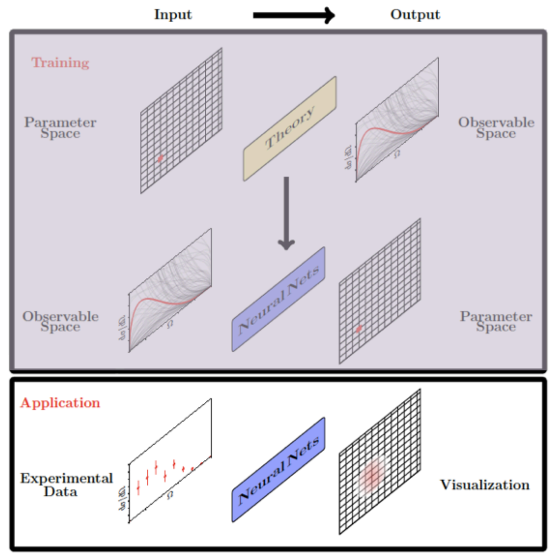
# High-level overview

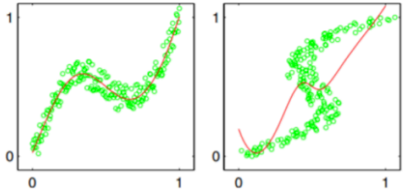
- Forward:
  - theory → simulation → observation
  - Backward:
  - observation → theory



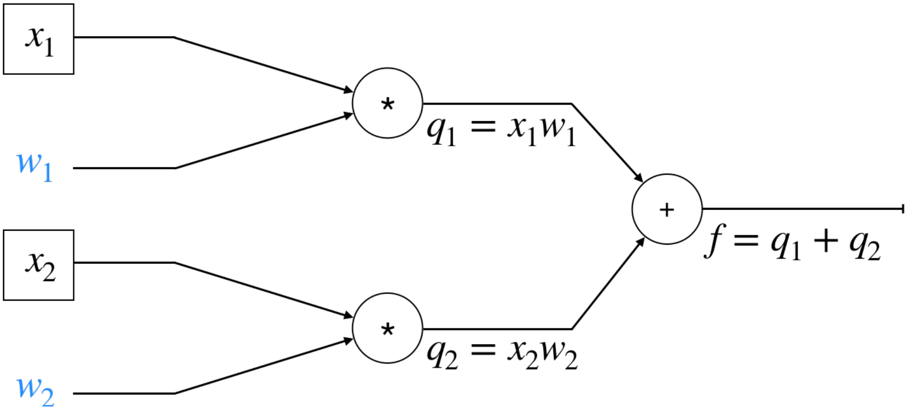
# How do experimental observations constrain theoretical models?

- Potential for multiple predictions in the parameter space

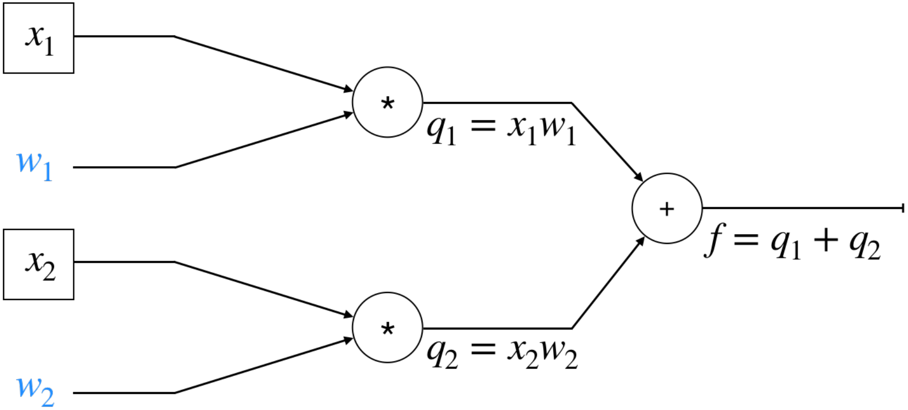


# NETWORK GRAPH



# SUPERVISED LEARNING




Loss function

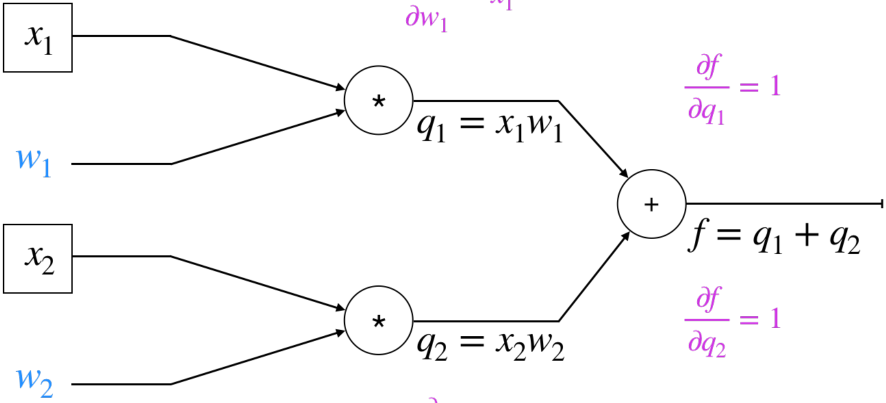
$$J(w) = f - \hat{f}$$

# BACKPROPAGATION

$$w_1 = w_1 + \eta * \frac{\partial f}{\partial q_1}\frac{\partial q_1}{\partial w_1}$$

$$\frac{\partial q_1}{\partial w_1} = x_1$$

$$\frac{\partial f}{\partial q_1} = 1$$




$$\frac{\partial J}{\partial f} = 1$$

$$\frac{\partial f}{\partial q_2} = 1$$

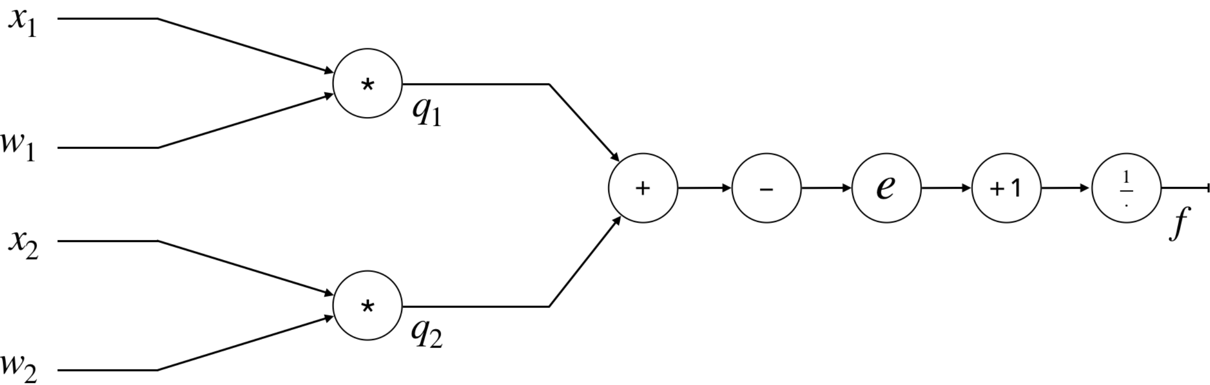
$$\frac{\partial q_2}{\partial w_2} = x_2$$

$$w_2 = w_2 + \eta * \frac{\partial f}{\partial q_1}\frac{\partial q_1}{\partial w_2}$$

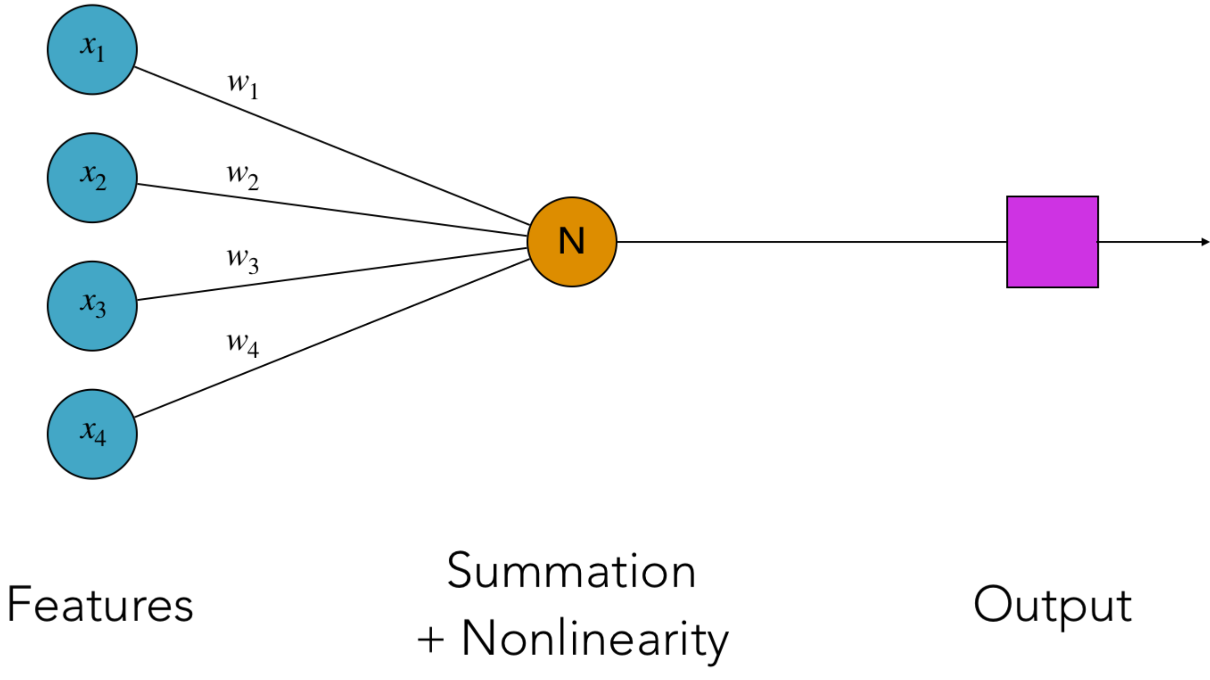
Loss function

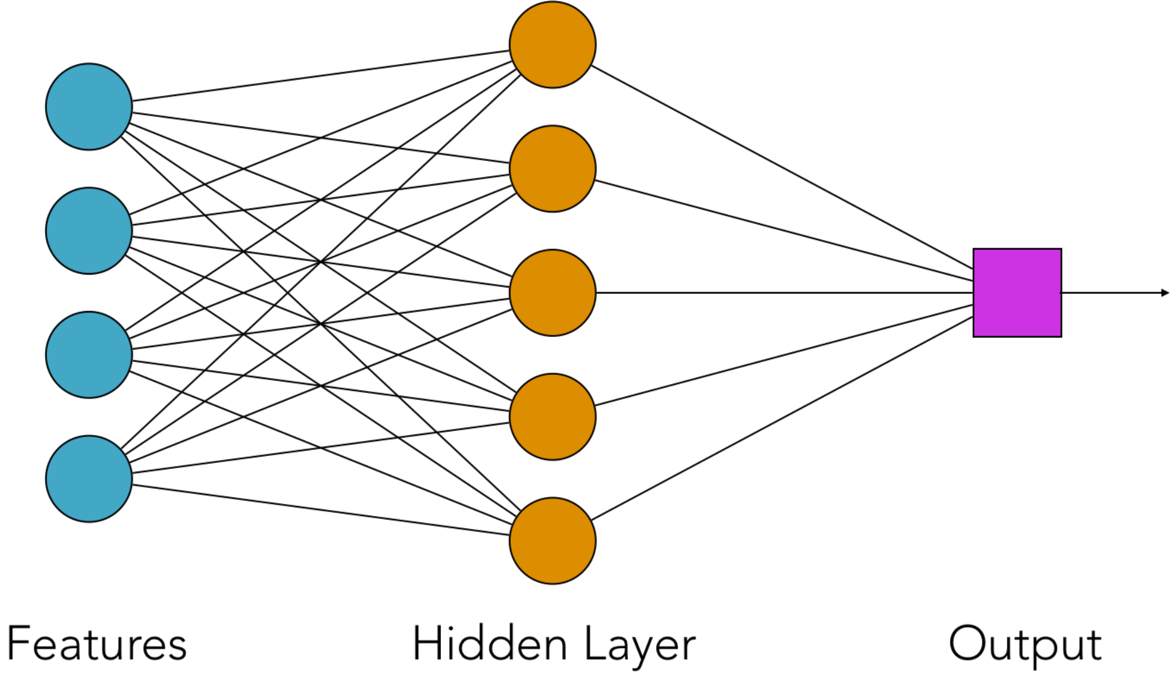
$$J(w) = f - \hat{f}$$

# LOGISTIC REGRESSION

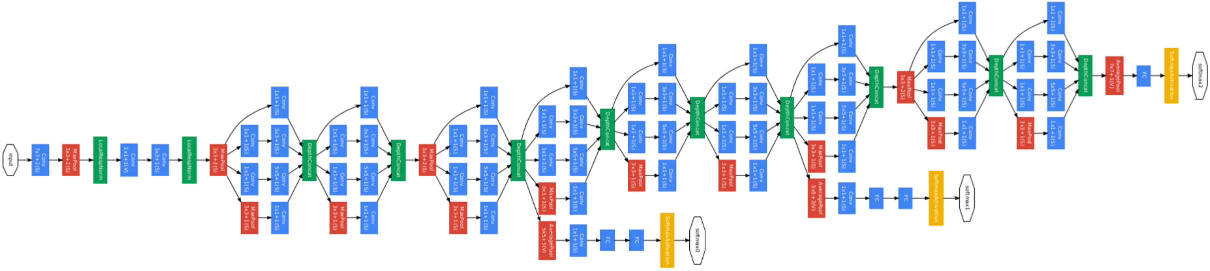



$$f = \frac{1}{1 + e^{-(x_1 w_1 + x_2 w_2)}}$$


$x_1$
$x_2$
$x_3$
$x_4$
$w_1$
$w_2$
$w_3$
$w_4$
N
Features
Summation
+ Nonlinearity
Output


Features
Hidden Layer
Output

“GoogLeNet network with all the bells and whistles”



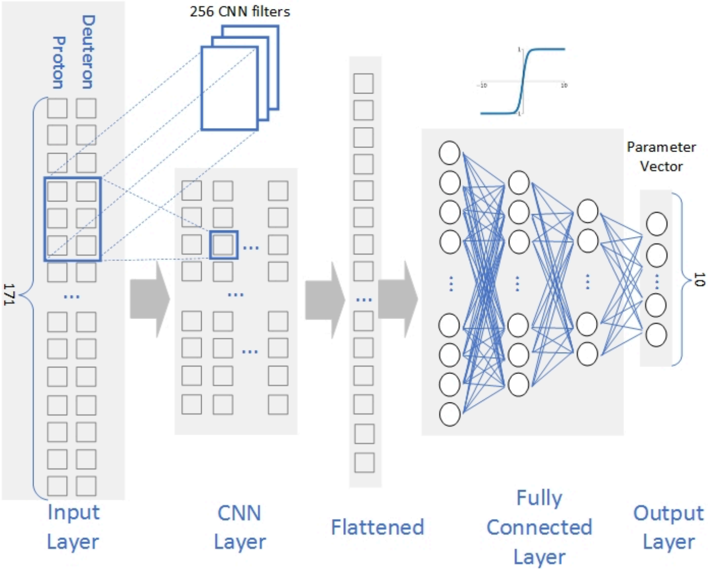
# Two models:

- 80000 training examples
- 80/20 train/validation split
- Blind test

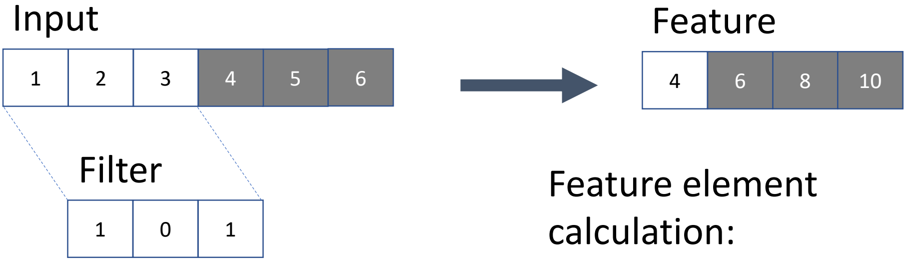
- 80000 training examples
- 80/20 train/validation split
- Blind test

## EBML:

- **INPUT: xsec**: 126x2 inputs
- 256 10x2 filters
- 3 256 node FCNN layers
- **OUTPUT: PDF parameters:** 10 outputs, 5 u 5d



# 1D Convolutions

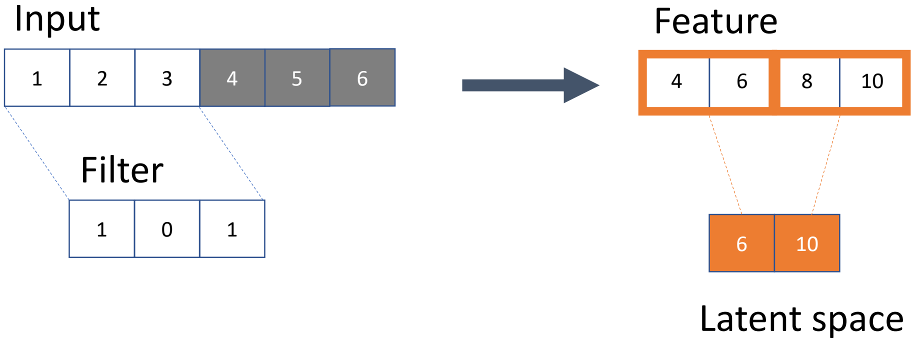



Feature element calculation:

1*1 + 2*0 + 3*1 = 4

Stride: 1

# Max pooling

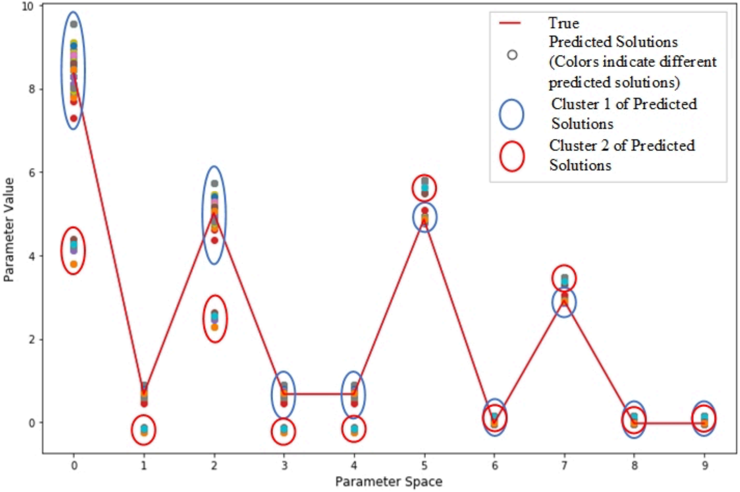


# Training data

- 80000 training examples
- 80/20 train/validation split
- Blind test

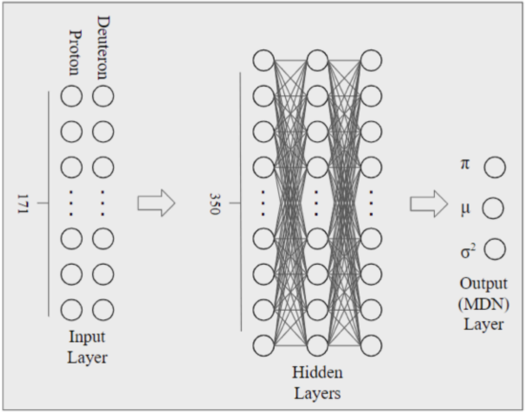
- EBML: Bagging
  - Trained many independent models
  - Each on subset of data with replacement

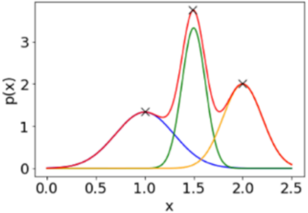
# Ensemble-based learning



# MDN:

- **INPUT: xsec:** 171x2 inputs
- 3 256 node FCNN layers
- **OUTPUT: PDF parameters, std, and probabilities:** 10x3 outputs, 5 u 5d




Output Layer Interpretation:

$$p(\mathbf{t}|\mathbf{x}) = \sum_{k=1}^{K} \pi_k(\mathbf{x}) \mathcal{N}\left(\mathbf{t}|\boldsymbol{\mu}_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\right)$$

# Mixture Density Network

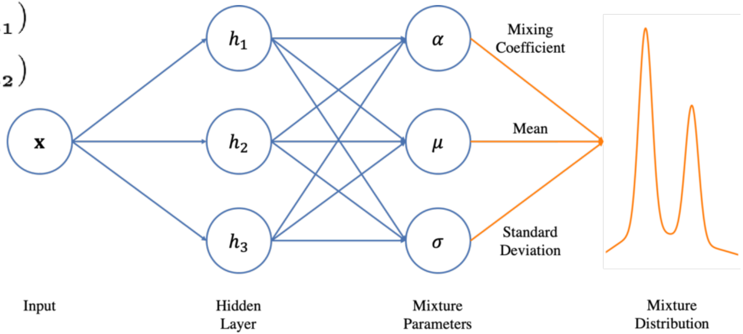
$$h_1(\mathbf{x}) = max(\mathbf{W}_1^\top \mathbf{x} + \mathbf{b}_1, 0)$$

$$\alpha(\mathbf{x}) = softmax(\mathbf{W}_\alpha^\top h_1(\mathbf{x}) + \mathbf{b}_\alpha)$$

$$\lambda_1(\mathbf{x}) = (\mathbf{W}_{\lambda_1}^\top h_1(\mathbf{x}) + \mathbf{b}_{\lambda_1})$$

$$\lambda_2(\mathbf{x}) = (\mathbf{W}_{\lambda_2}^\top h_1(\mathbf{x}) + \mathbf{b}_{\lambda_2})$$

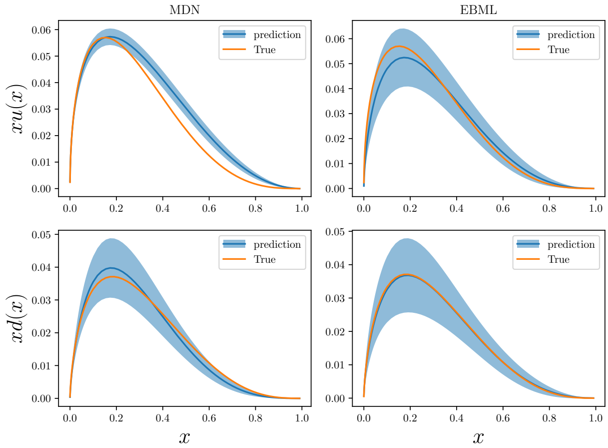
softmax = Boltzmann/Gibbs distribution!




$$p(y|\mathbf{x}) = \sum_{c=1}^{C} \alpha_c(\mathbf{x}) \mathcal{D}\left(y|\lambda_{1,c}(\mathbf{x}), \lambda_{2,c}(\mathbf{x}), \ldots\right)$$

$$\arg\min_{\boldsymbol{\Theta}} l(\boldsymbol{\Theta}) = -\frac{1}{|\mathbb{D}|} \sum_{(\mathbf{x},y)\in\mathbb{D}} \log p(y|\mathbf{x})$$

# Blind test results




- Blind test:
  - Random cross-section inputs with known outputs
  - Resampling cross-section with added noise
  - Predict parameters
  - Average, std for QPD predictions

- **EBML within 1 std**

# Summary and outlook

- **Progress**
  - First prototypes for backward mappers has been designed and validated
- **Projected work to completion (until September 30, 2019)**
  - Develop the forward mapper and its validation
  - Train the mappers on inclusive DIS and bench mark the results with recent JAM results
- **Future**
  - Train NNs to additional observables e.g. SIDIS, SIA, DY, DDY, and TMDs, GPDs observables
  - Build a web space where all the mappers are available for users