# On the Determination of Uncertainties in Parton Densitites

Nicholas Hunt-Smith
arXiv:2206.10782
Supervisors: Tony Thomas and Martin White
Collaborators: Wally Melnitchouk, Nobuo Sato and Alberto Accardi

University of Adelaide

December 13, 2022

# Summary

# Bayesian Methods

## Bayes' Theorem:

$$p(\boldsymbol{a}|\boldsymbol{m}) = \frac{1}{\mathcal{Z}}\, p(\boldsymbol{m}|\boldsymbol{a})\, p(\boldsymbol{a})\,, \text{ with evidence: } \mathcal{Z} = \int \mathrm{d}\boldsymbol{a}\; p(\boldsymbol{m}|\boldsymbol{a})\, p(\boldsymbol{a})$$

$$\text{and likelihood: } p(\boldsymbol{m}|\boldsymbol{a}) = \mathcal{N} \exp\left[ -\frac{1}{2}\chi^2(\boldsymbol{a}, \boldsymbol{m}) \right].$$

# Bayesian Methods: MCMC

Expectation value and variance of any observable $\mathcal{O}$ can be written as:

$$E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n}\sum_{k=1}^{n} \mathcal{O}(\boldsymbol{a}_k)\,,$$

$$V_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n}\sum_{k=1}^{n} \left[\mathcal{O}(\boldsymbol{a}_k) - E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\}\right]^2.$$

Metropolis-Hastings and Hamiltonian Monte Carlo (HMC) are examples of MCMC algorithms that obtain samples $\boldsymbol{a}_k$. Nested Sampling is an alternative Bayesian technique that primarily aims to evaluate the evidence $\mathcal{Z}$, also produces samples as a byproduct.

# Approximations to Bayesian Posterior: Hessian

Change of variables $p(\boldsymbol{a}|\boldsymbol{m}) \to p(\boldsymbol{t}|\boldsymbol{m})$: $\boldsymbol{a}(\boldsymbol{t}) = \boldsymbol{a}_0 + \sum_{k=1}^{n_{\text{par}}} t_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}$ ,

$$E_{\text{Hess}}\{\mathcal{O}(\boldsymbol{a})\} = \int \mathrm{d}^n t \; p(\boldsymbol{t}|\boldsymbol{m})\, \mathcal{O}(\boldsymbol{a}(\boldsymbol{t})) \approx \mathcal{O}(\boldsymbol{a}_0) \; .$$

$V_{\text{Hess}}\{\mathcal{O}(\boldsymbol{a})\} \approx \sum_k T_k^2 \left( \left. \frac{\partial \mathcal{O}(\boldsymbol{a}(\boldsymbol{t}))}{\partial t_k} \right|_{\boldsymbol{a}_0} \right)^2$, where: $T_k^2 = \int \mathrm{d}t_k \; p_k(t_k|\boldsymbol{m})\, t_k^2$.

$T_k^2$ is set between $5 - 10$, inflating uncertainties to 68% coverage in global PDF fits according to ad hoc "tolerance criterion". This is motivated by statistical inconsistencies in data.

# Approximations to Bayesian Posterior: Data Resampling

Data resampling uses frequentist logic to approximate Bayesian posterior with distribution in maximum likelihood estimators:

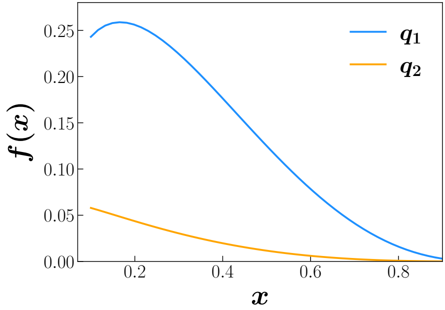
$$E_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n_{\text{rep}}} \sum^{n_{\text{rep}}} \mathcal{O}(\boldsymbol{a}_{\text{rep}})\,,$$

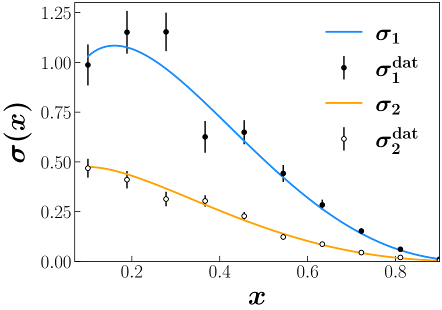
$$V_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n_{\text{rep}}} \sum^{n_{\text{rep}}} \left[\mathcal{O}(\boldsymbol{a}_{\text{rep}}) - E_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\}\right]^2.$$

# Toy 4D Quark model

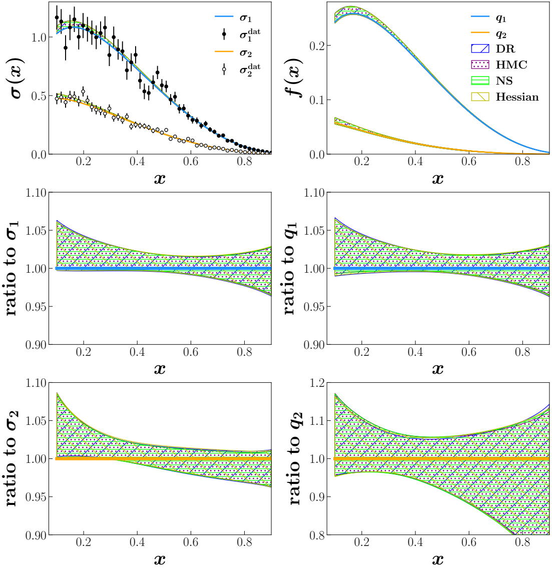
$$q_i(x) = x^{\alpha_i}(1-x)^{\beta_i},$$
$$i = 1, 2.$$

$$\sigma_j = \sum_{i=1,2} c_{ji}\, q_i,$$
$$c_{11} = 4c_{12} = 4c_{21} = c_{22}.$$

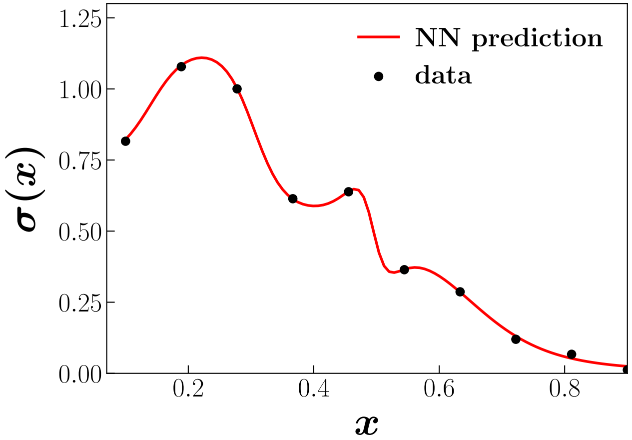


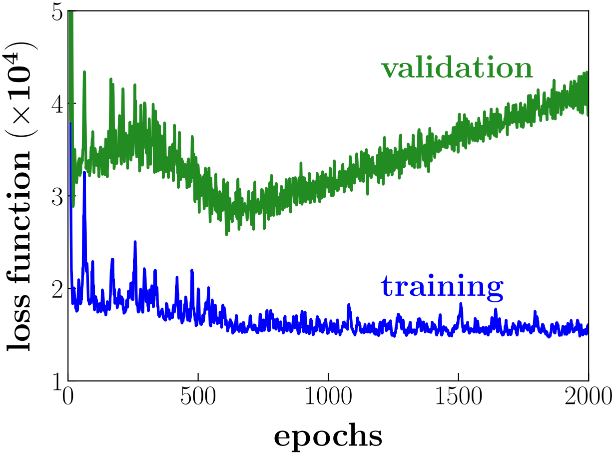
# Equivalency of Parametric Methods



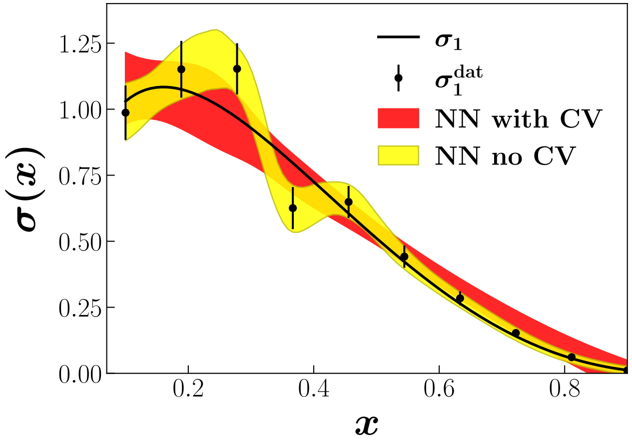
# Overfitting with NNs



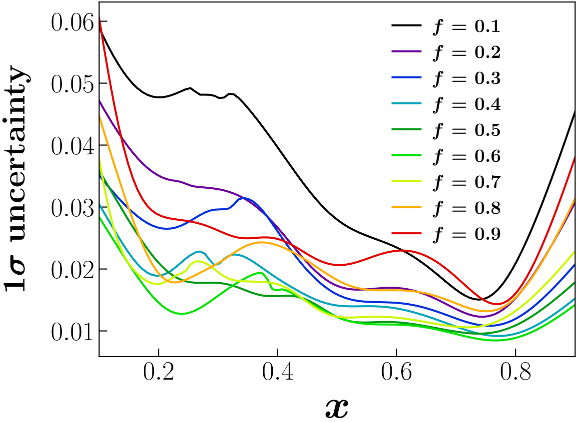
# Overfitting with NNs



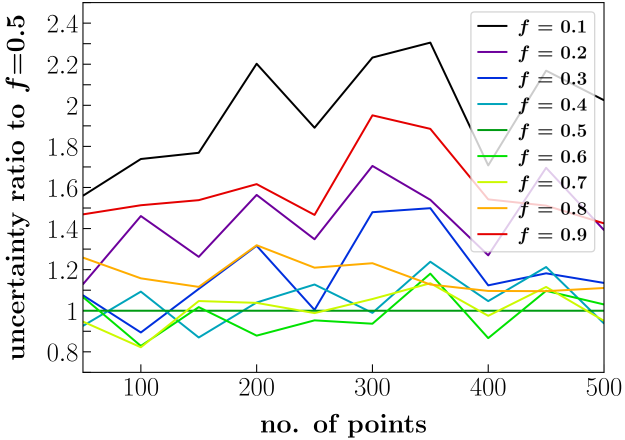
# Overfitting with NNs



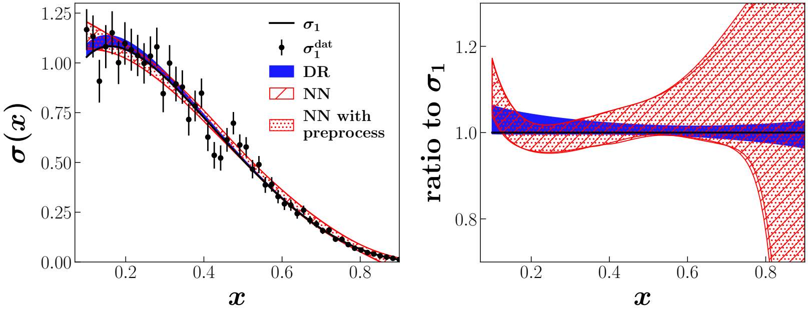
# Uncertainty Dependence on Partition Fraction



# Uncertainty Dependence on Partition Fraction



# Comparison of Neural Nets to Parametric Methods



# Closure Test of Neural Nets