Karlsruher Institut für Technologie

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Calibrating Bayesian Generative Machine Learning for Bayesiamplification

Sebastian Bieringer, Sascha Diefenbacher, Gregor Kasieczka, Mathias Trabs

Universität Hamburg, Germany

sebastian.guido.bieringer@uni-hamburg.de

07.11.2024 - ML4jets

GANplify

What is data amplification?

- set of data $D_n = \{x_1, \ldots, x_n\}$ with $x_i \sim p_X$
- we are interested in some property $h(X)$

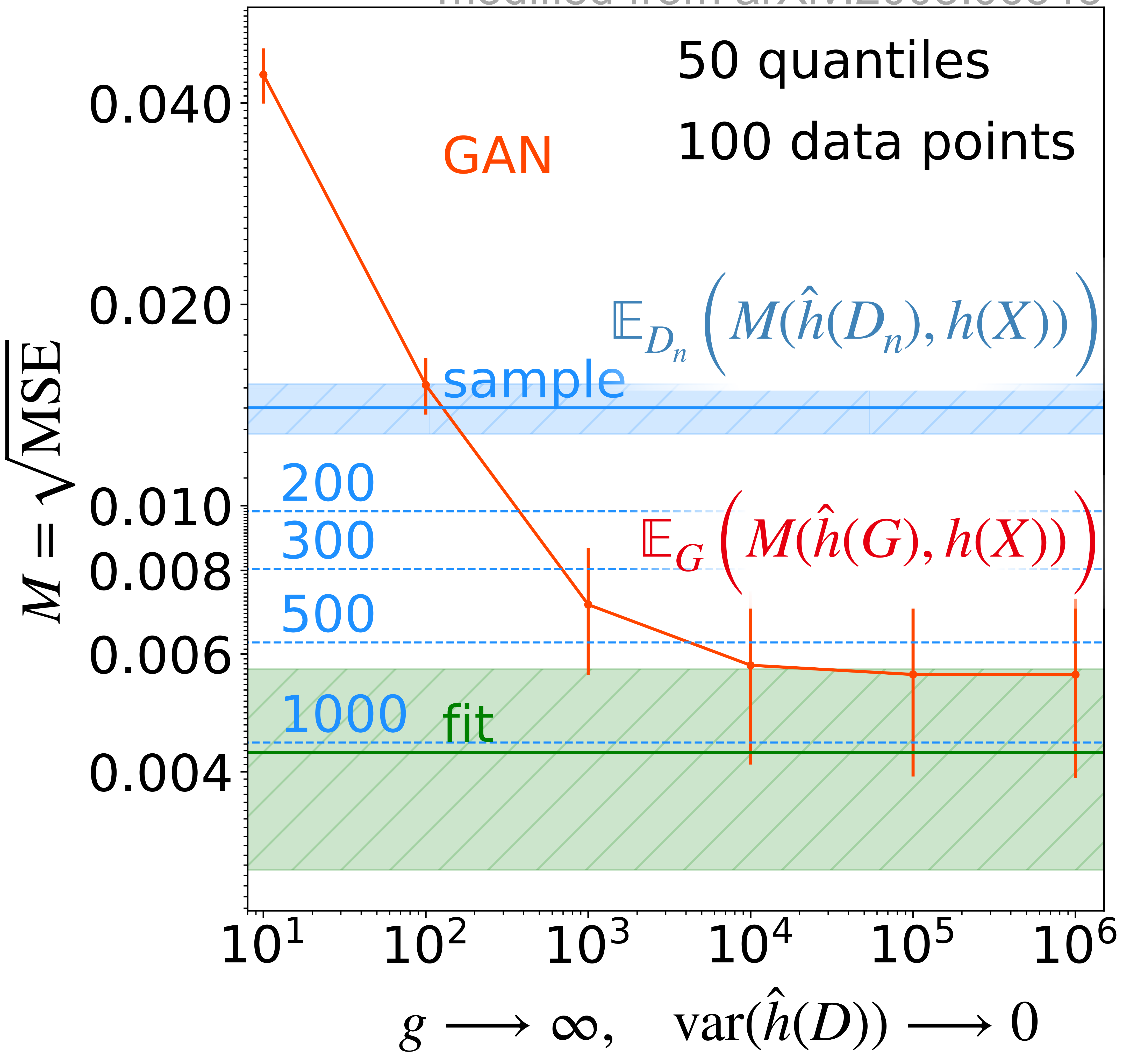
  $\Rightarrow$ use stochastic estimator $\hat{h}(D)$ for some set D
- set of generative Neural Network samples $G = \{x'_1, \ldots, x'_g\}$

  with $x'_i \sim \hat{p}_{X,\theta} \approx p_X$

Is $\hat{h}(G)$ better than $\hat{h}(D_n)$?

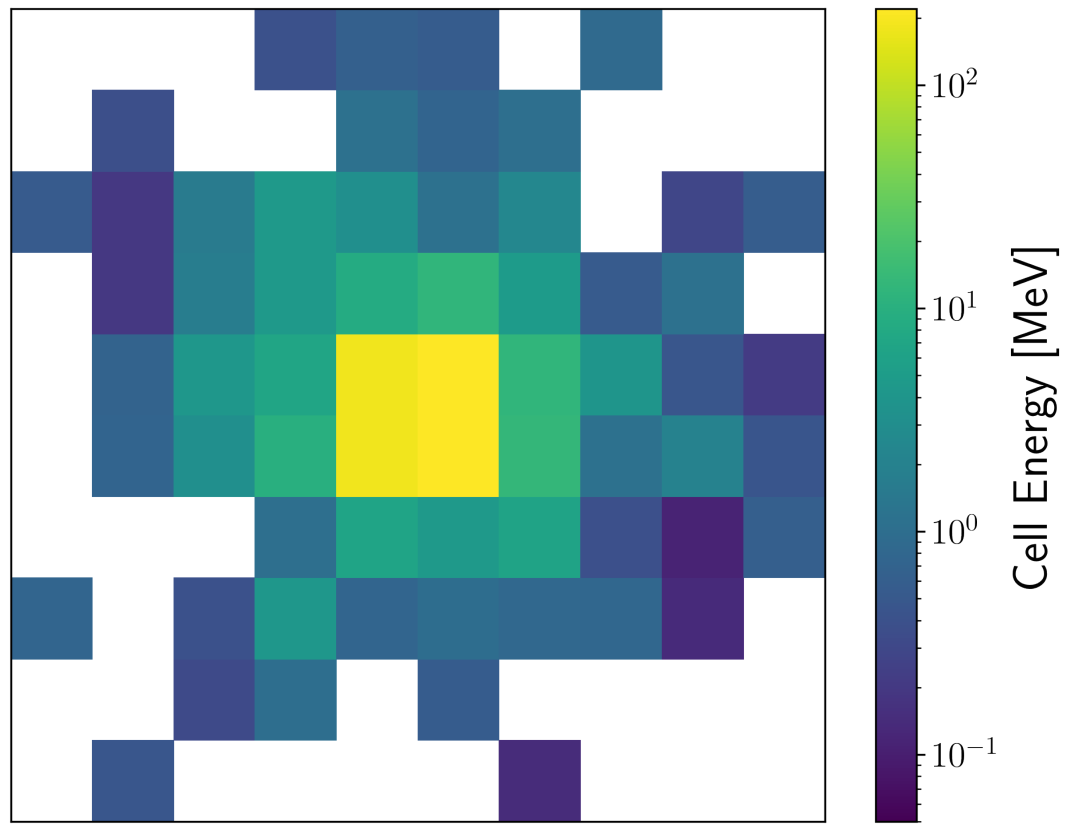
- measure discrepancy to $h(X)$ with a distance $M(\hat{h}(D), h(X))$

modified from arXiv:2008.06545

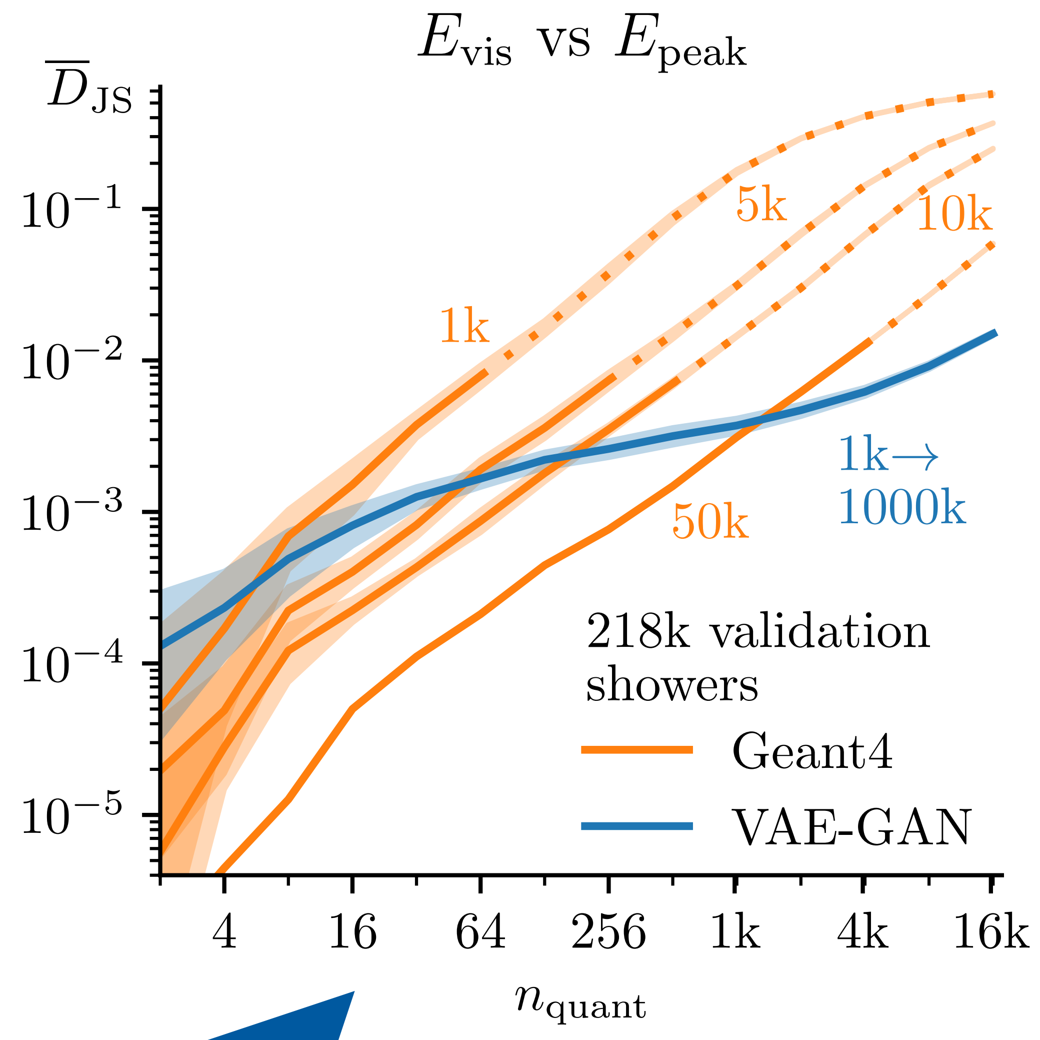


Calomplify

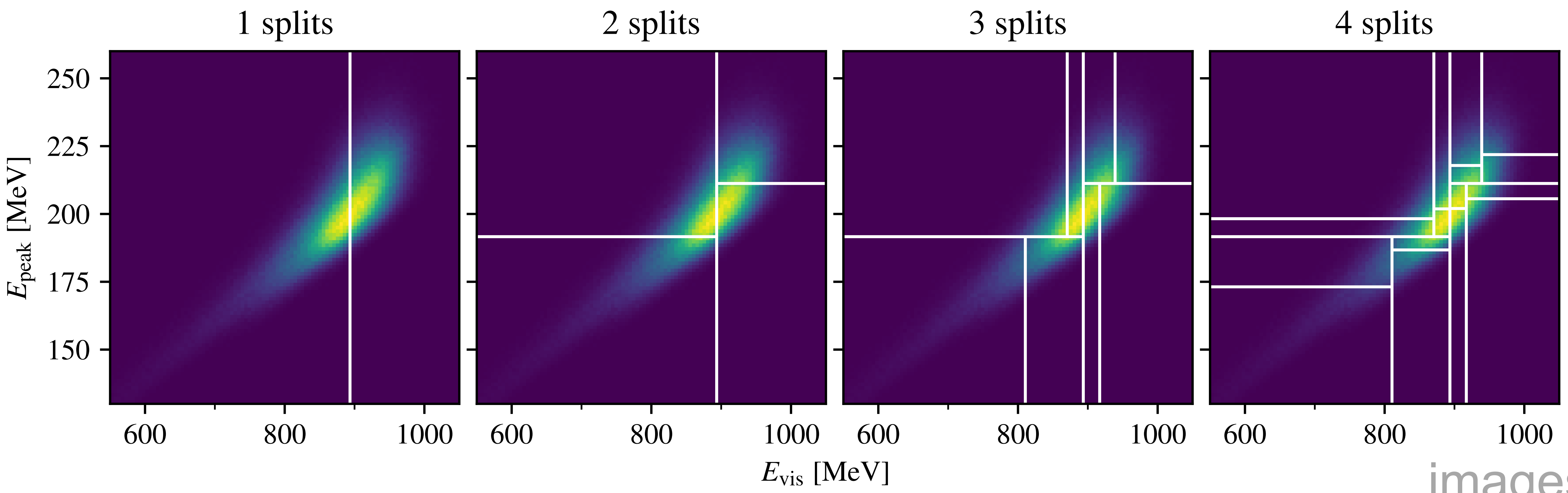
p_X: 10x10 pixel detector images
of 50 GeV photon showers



M: Jensen-Shannon divergence



$\hat{h}$: histogram values of high-level observables



images taken from arXiv:2202.07352

Bayesian Uncertainties

Bayesian Neural Networks use a set

$\Theta = \{\theta^{(1)}, \dots, \theta^{(s)}\}$, with $\theta^{(i)} \sim \pi(\theta \mid D_n)$

$$\{\hat{p}_{X,\theta^{(1)}}, \dots, \hat{p}_{X,\theta^{(s)}}\}$$

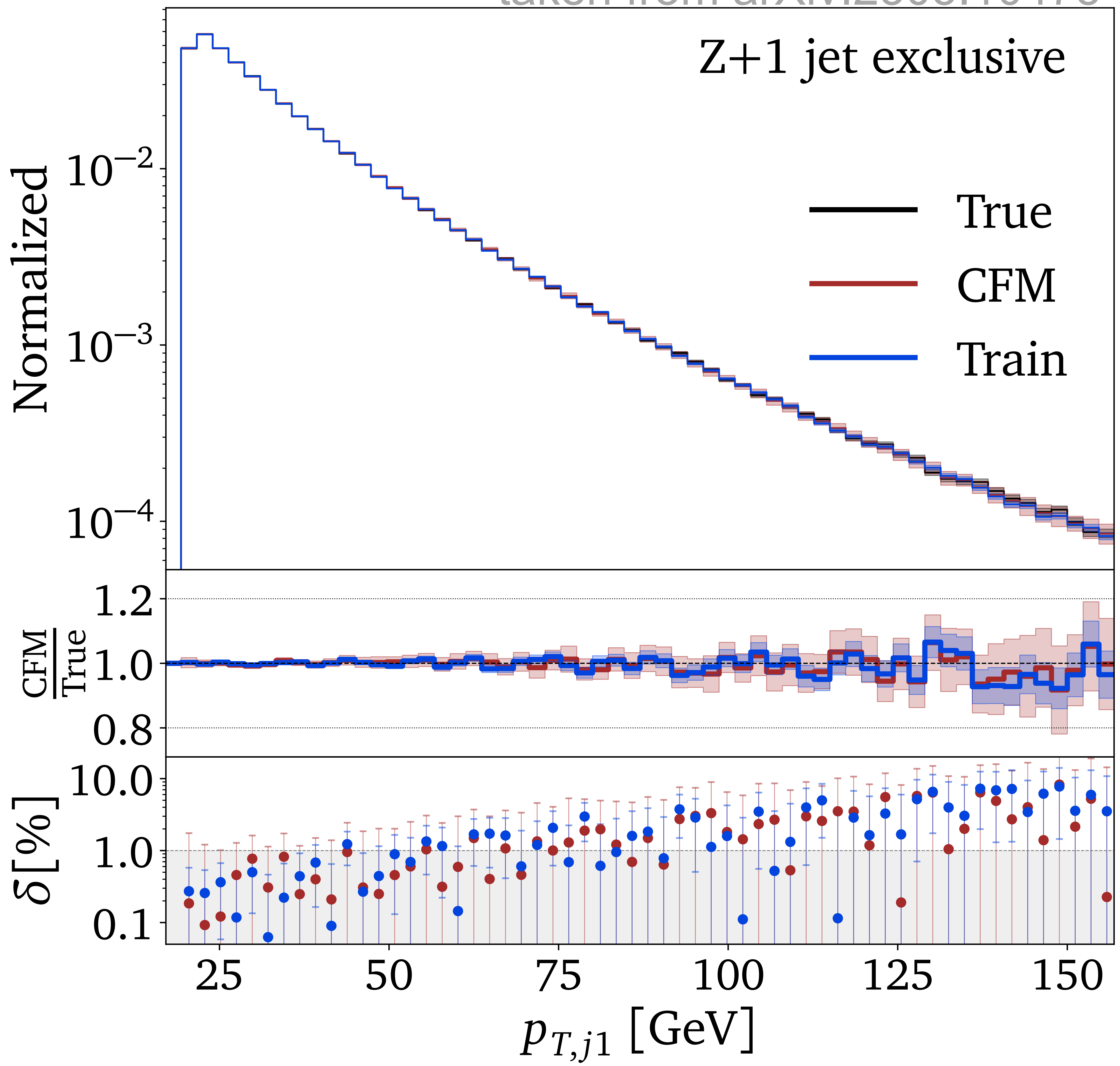
$$\{G^{(1)}, \dots, G^{(s)}\}$$

$$\{\hat{h}(G^{(1)}), \dots, \hat{h}(G^{(s)})\}$$

$\mathrm{std}_\Theta(\hat{h}(G))$ estimates the bias of $\mathrm{mean}_\Theta(\hat{h}(G))$ due to the limitations of D_n

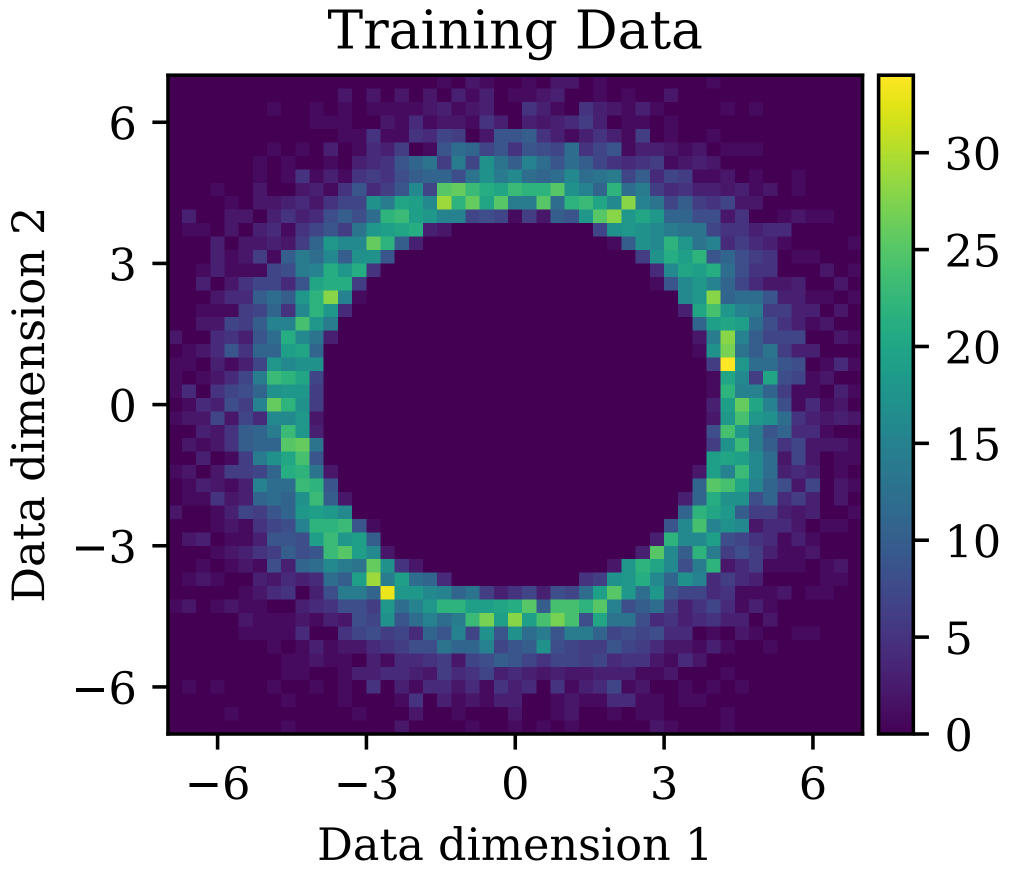
How is this connected to data amplification?

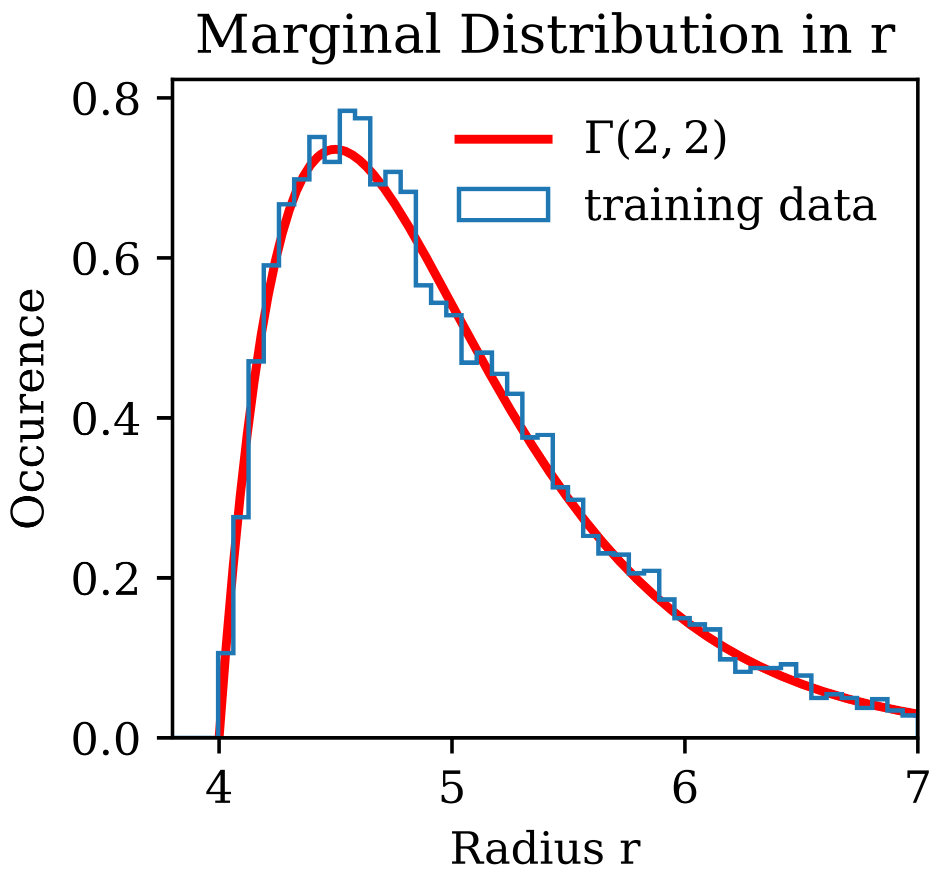
taken from arXiv:2305.10475



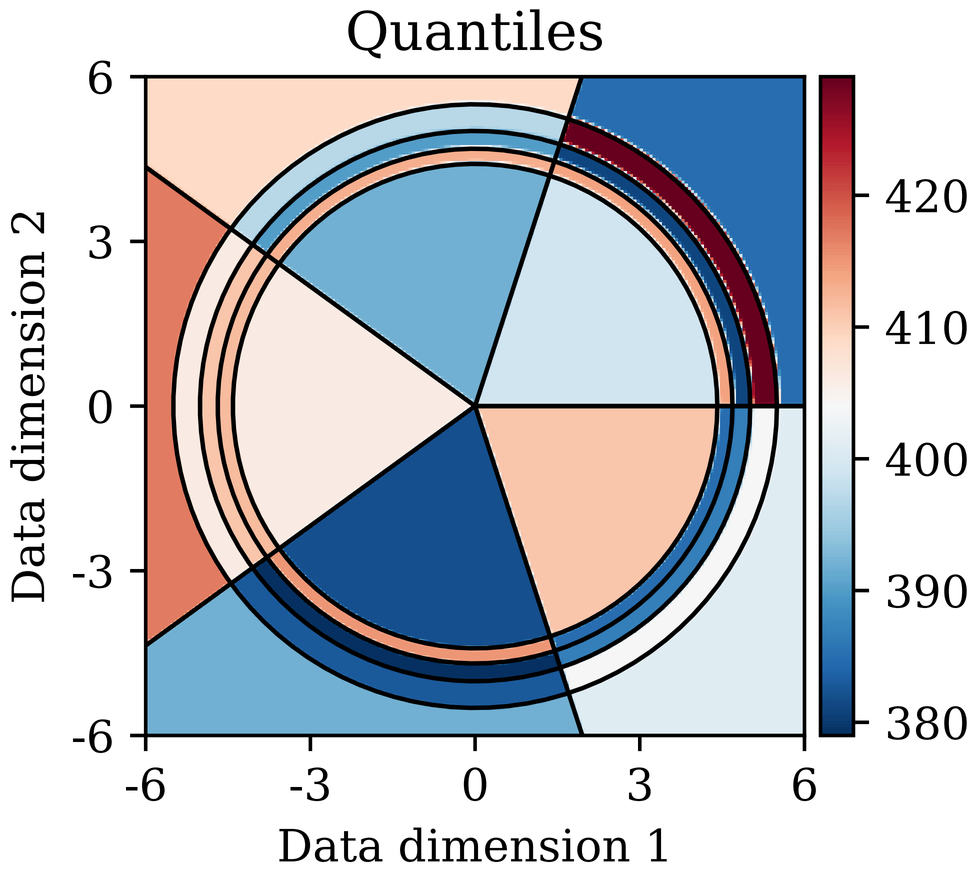
Bayesian Amplification Estimate

p_X: toy data from a 2D ring distribution





$\hat{h}$: histogram counts in quantiles

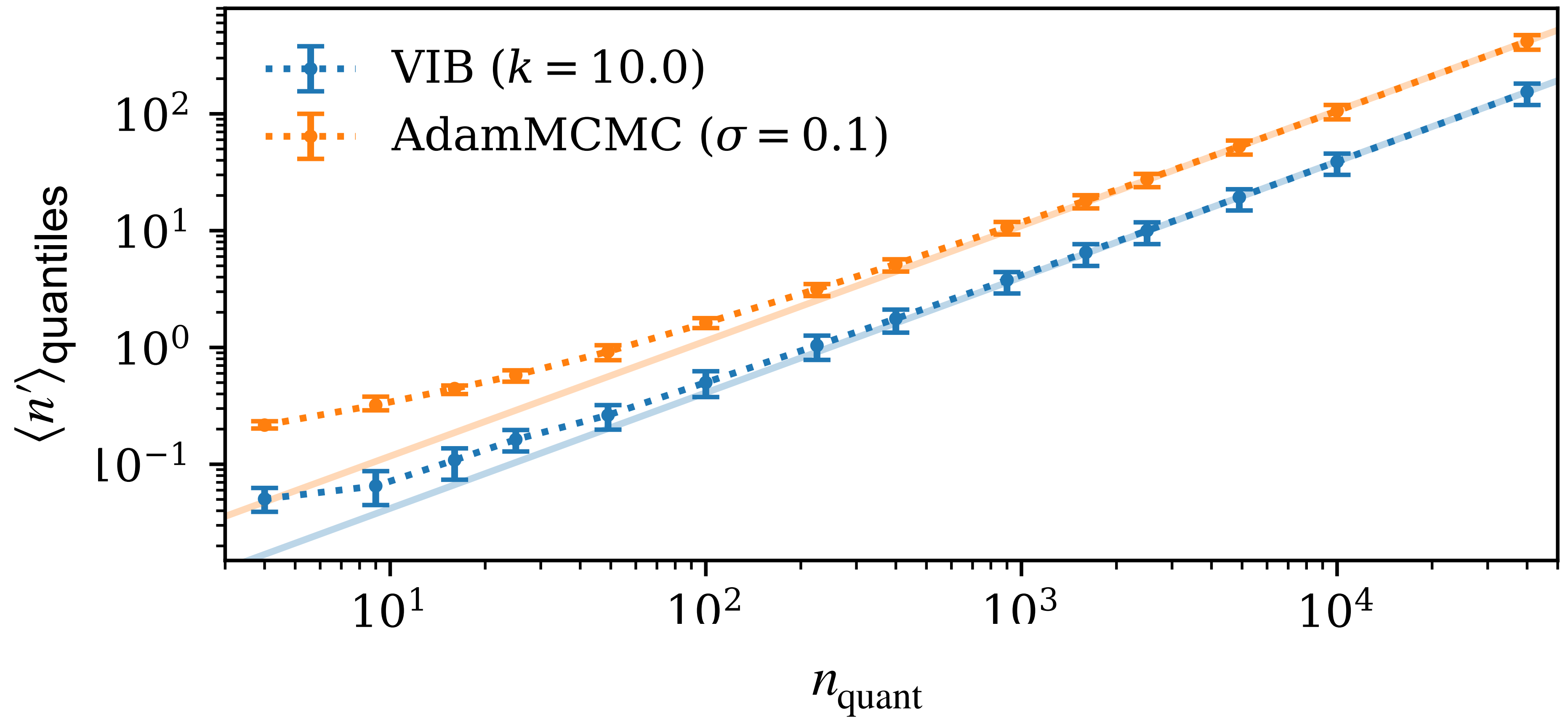


by construction

$$\text{mean}_{D_{n'}}(h_j(D_{n'})) = \text{std}^2_{D_{n'}}\left(h_j(D_{n'})\right) = \frac{n'}{n_\text{quant}}$$

$$\frac{\text{std}_{D_{n'}}\left(h_j(D_{n'})\right)}{\text{mean}_{D_{n'}}(h_j(D_{n'}))} = \sqrt{\frac{n_\text{quant}}{n'}} \stackrel{!}{=} \frac{\text{std}_\Theta(\hat{h}_j(G))}{\text{mean}_\Theta(\hat{h}_j(G))}$$

$$\Longleftrightarrow \quad n' = n_\text{quant} \frac{\text{mean}^2_\Theta(\hat{h}_j(G))}{\text{std}^2_\Theta(\hat{h}_j(G))}$$

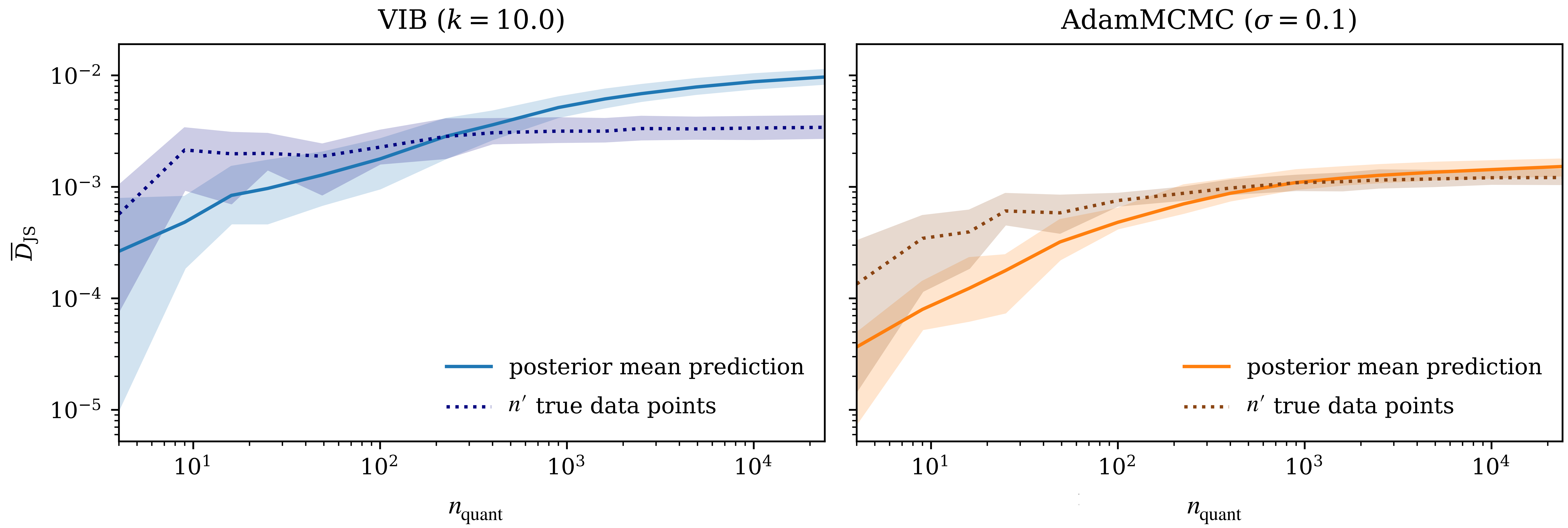


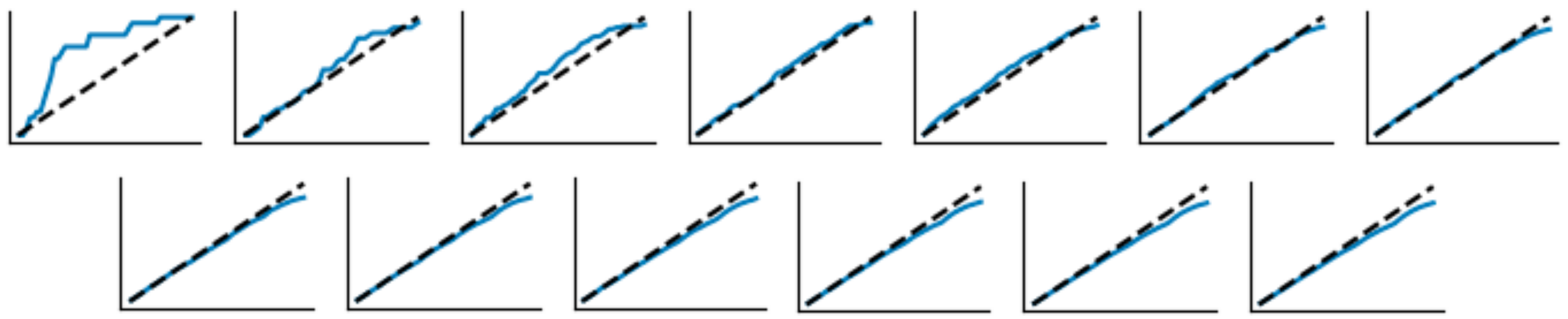
Bayesian Amplification Estimate

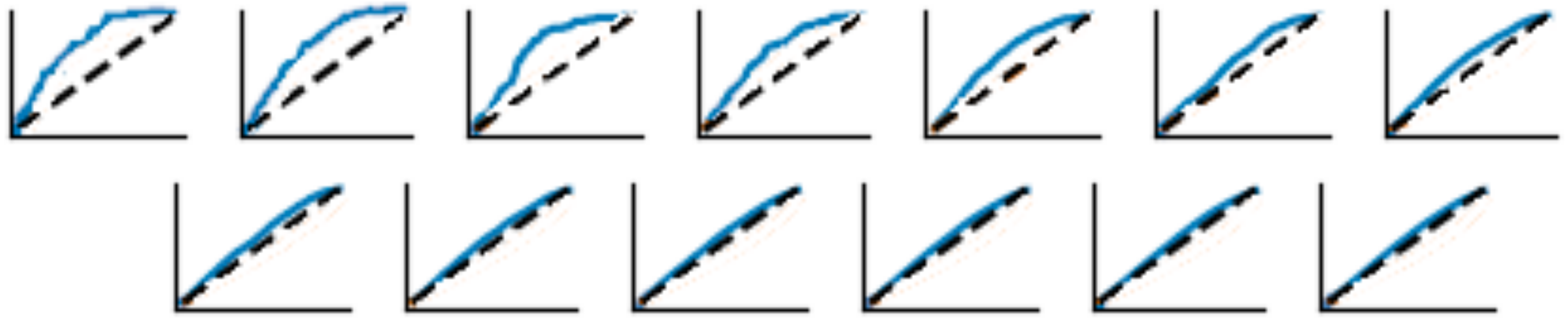
$$n' = n_{\text{quant}} \frac{\text{mean}^2_\Theta(\hat{h}_j(G))}{\text{std}^2_\Theta(\hat{h}_j(G))}$$

but does it work?

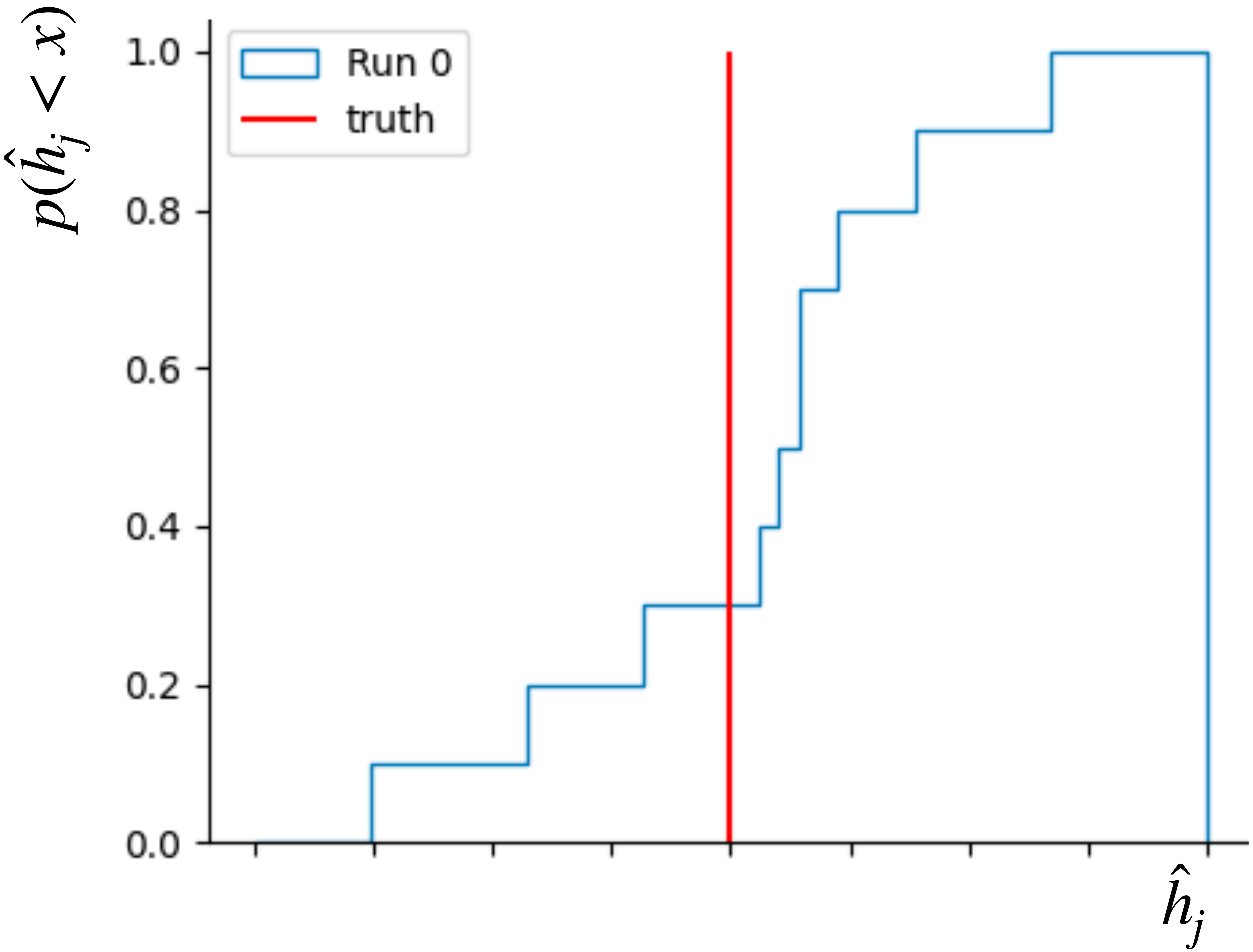
$\Longrightarrow$ **only when the uncertainties are well calibrated**





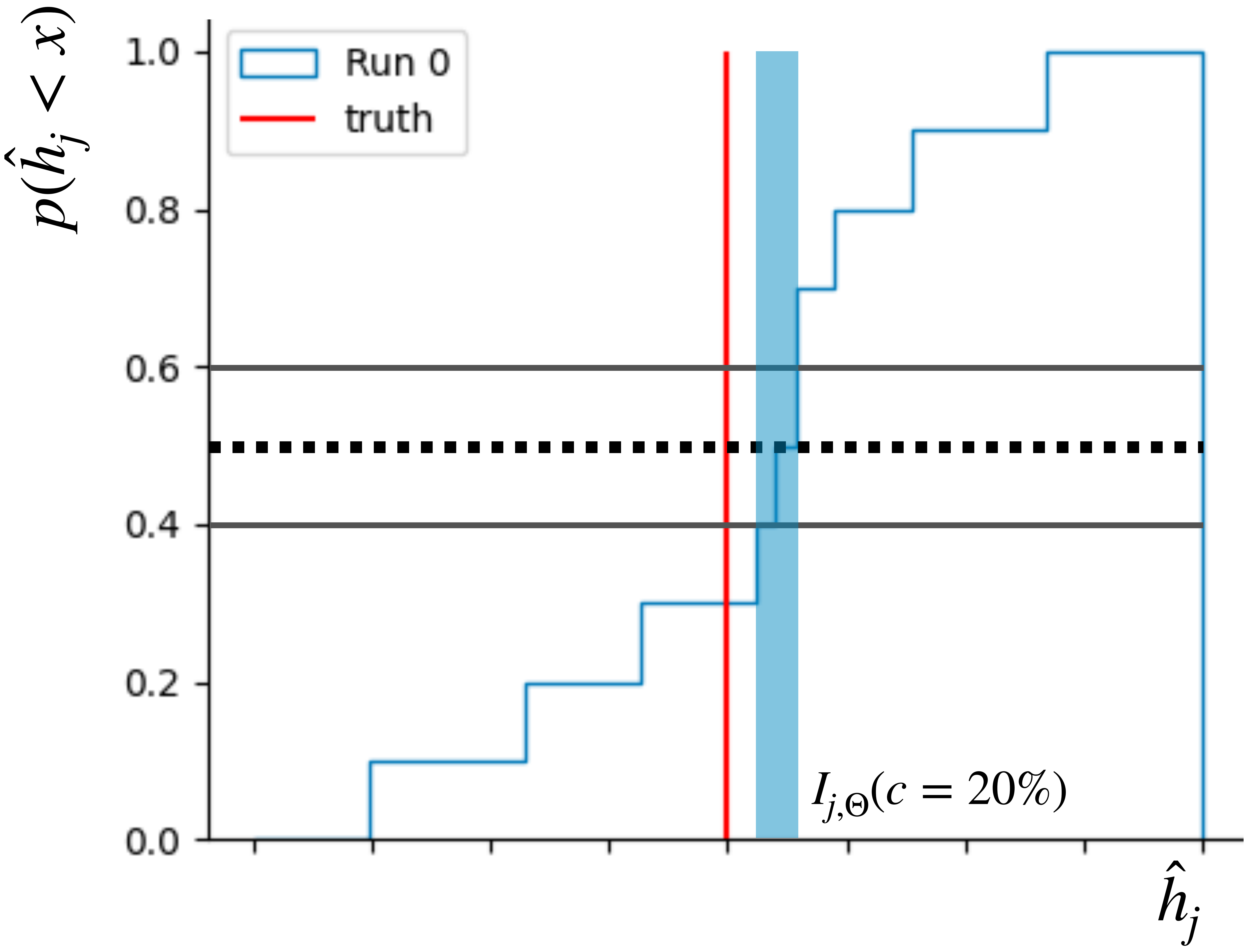


Calibration



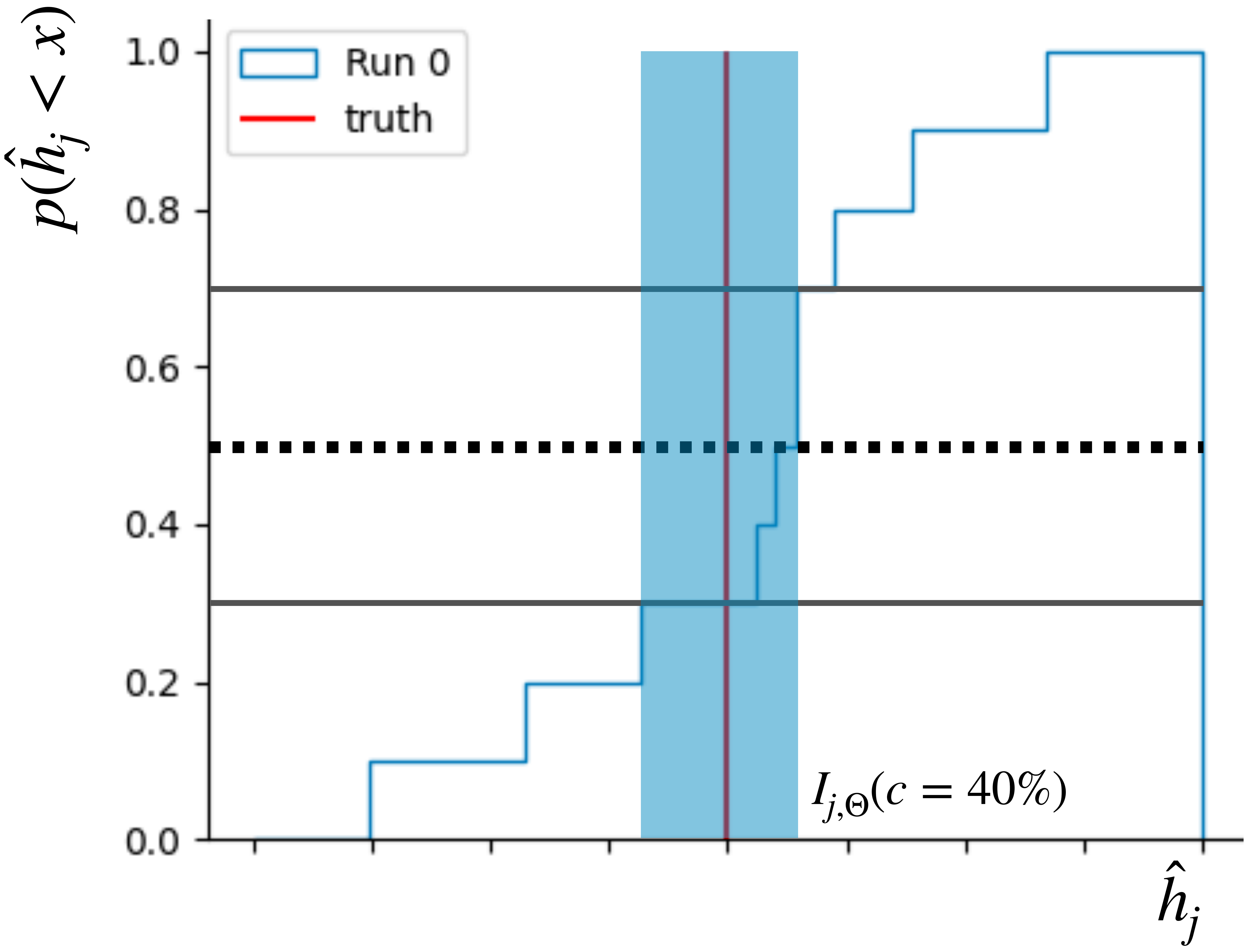
- cumulative histogram of $\{\hat{h}_j(G^{(1)}), \ldots, \hat{h}_j(G^{(10)})\}$

Calibration



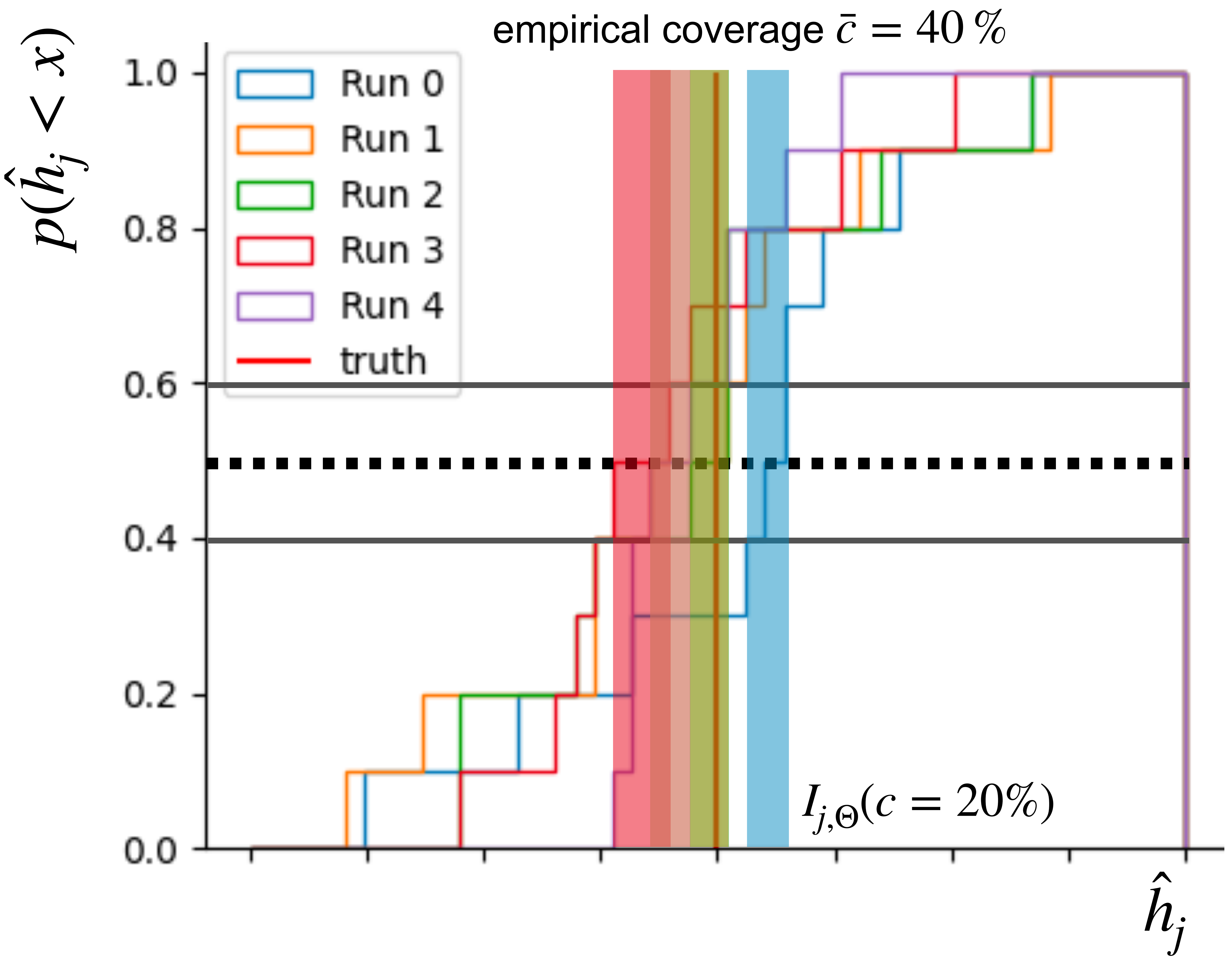
- cumulative histogram of $\{\hat{h}_j(G^{(1)}), \ldots, \hat{h}_j(G^{(10)})\}$
- construct credible set $I_{j,\Theta}(c)$

Calibration



- cumulative histogram of $\{\hat{h}_j(G^{(1)}), \ldots, \hat{h}_j(G^{(10)})\}$
- construct credible set $I_{j,\Theta}(c)$

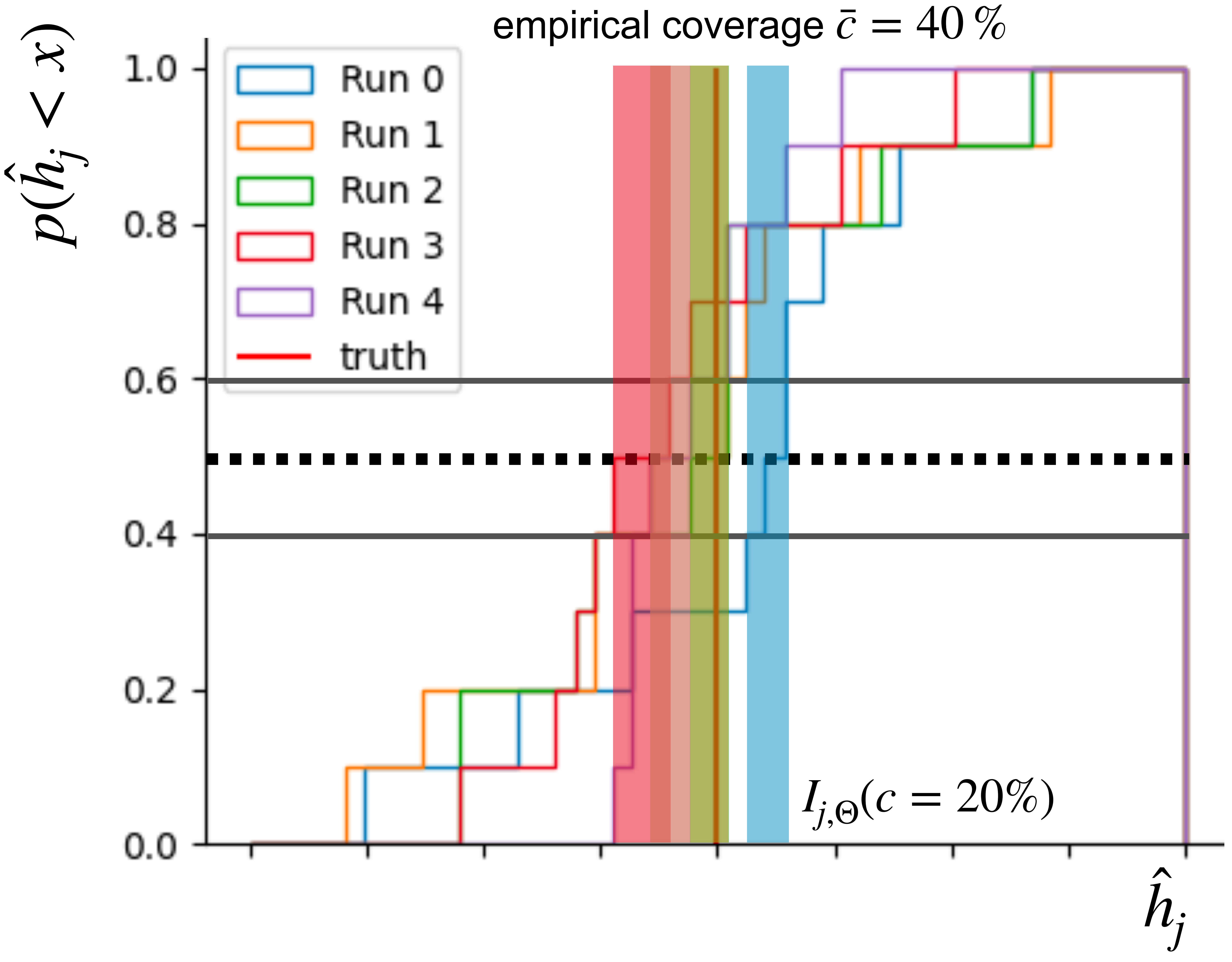
Calibration



- cumulative histogram of $\{\hat{h}_j(G^{(1)}), \dots, \hat{h}_j(G^{(10)})\}$
- construct credible set $I_{j,\Theta}(c)$
- repeat for 5 independent BNNs $\Theta_1, \dots, \Theta_5$
- check how often truth $\in I_{j,\Theta_i}(c)$

  $\Rightarrow$ empirical coverage $\bar{c}$

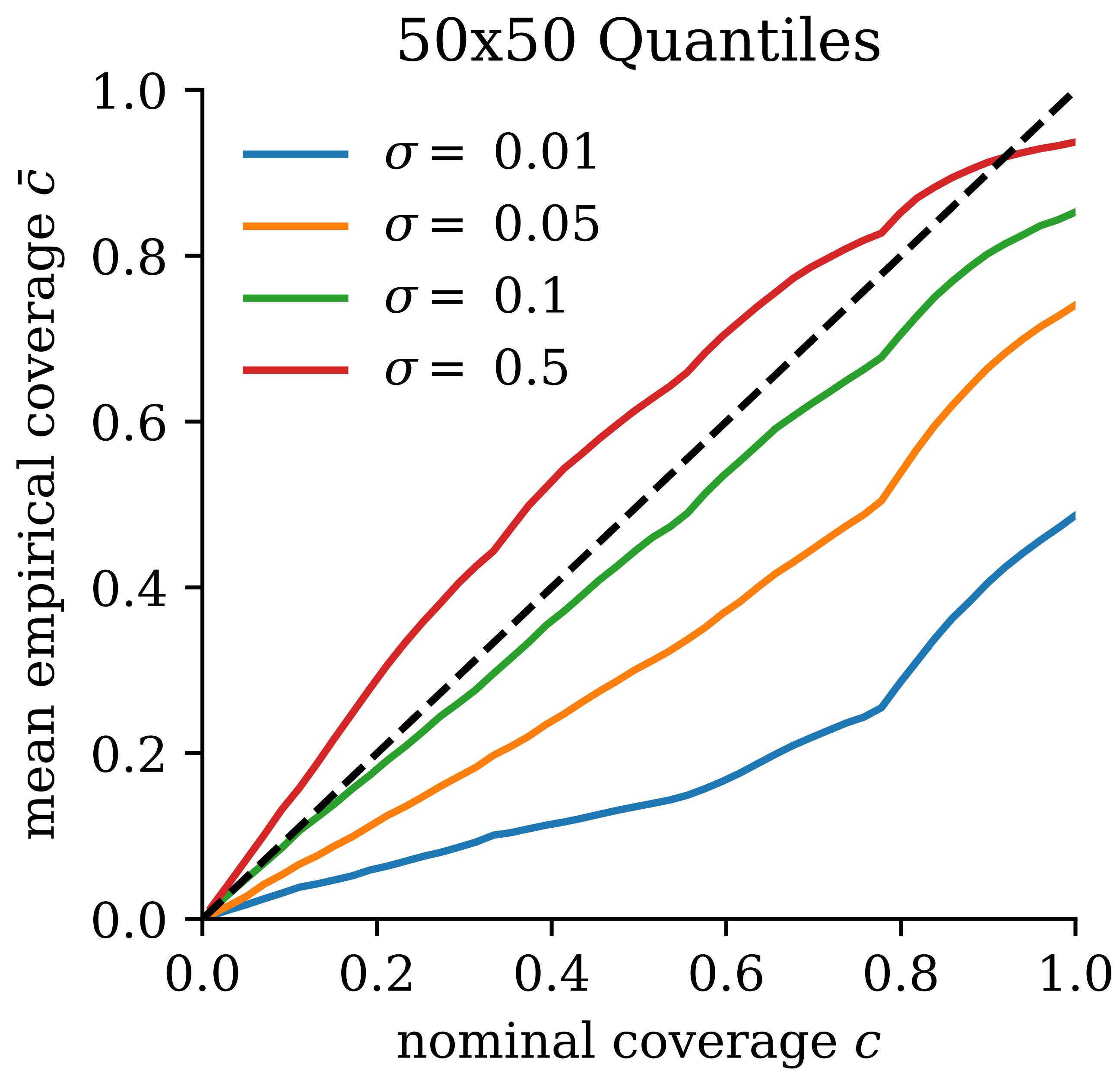
Calibration



- cumulative histogram of $\{\hat{h}_j(G^{(1)}), \ldots, \hat{h}_j(G^{(10)})\}$
- construct credible set $I_{j,\Theta}(c)$
- repeat for 5 independent BNNs $\Theta_1, \ldots, \Theta_5$
- check how often truth $\in I_{j,\Theta_i}(c)$

  $\Rightarrow$ empirical coverage $\bar{c}$

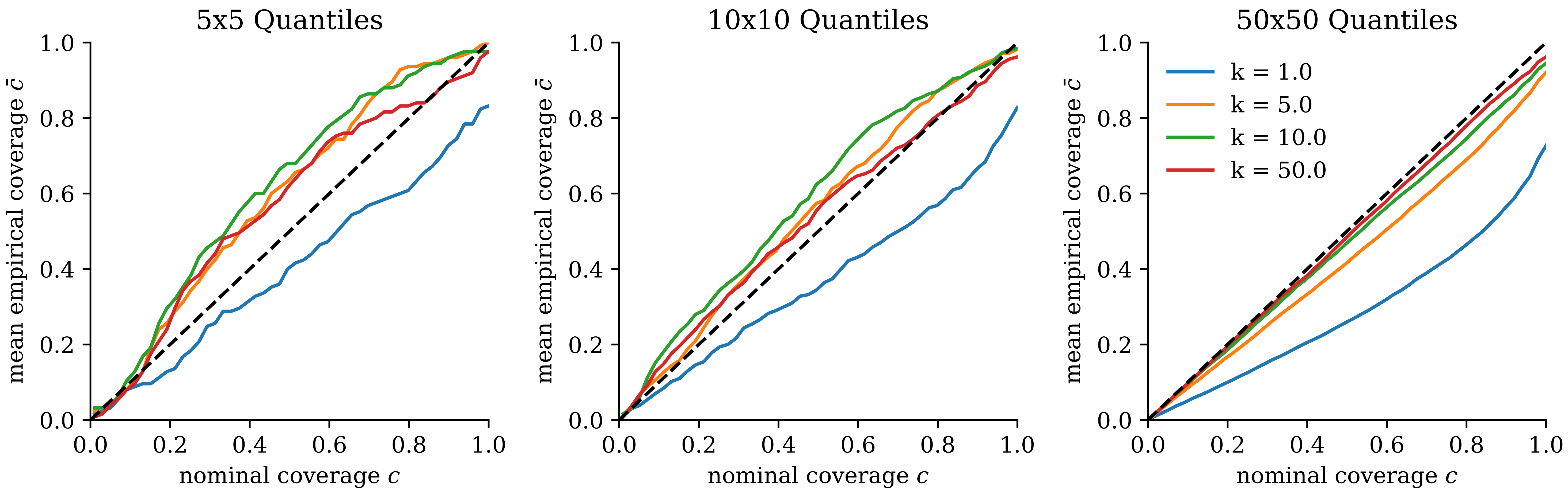
plot $\bar{c}$ against c



Different Bayesian Methods

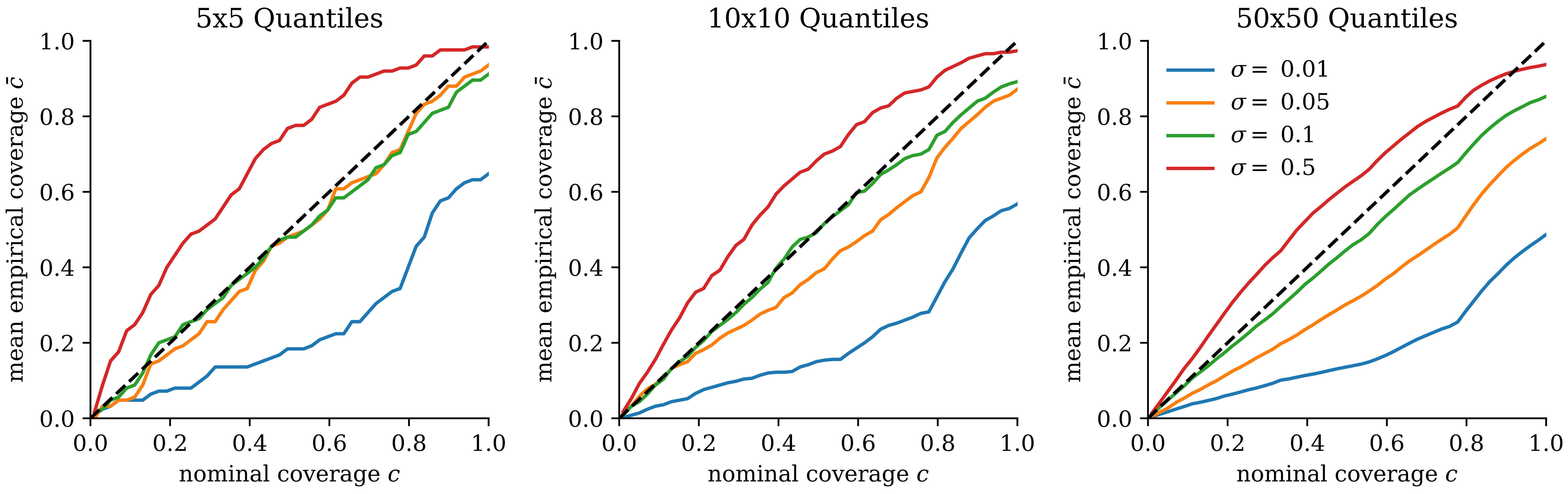
Variational Inference

- $\pi(\theta \mid D_n) \approx \mathcal{N}\left(\theta; \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_P \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & & \\ & \cdots & \\ & & \sigma_P^2 \end{pmatrix}\right)$
- train with stochastic gradient descent



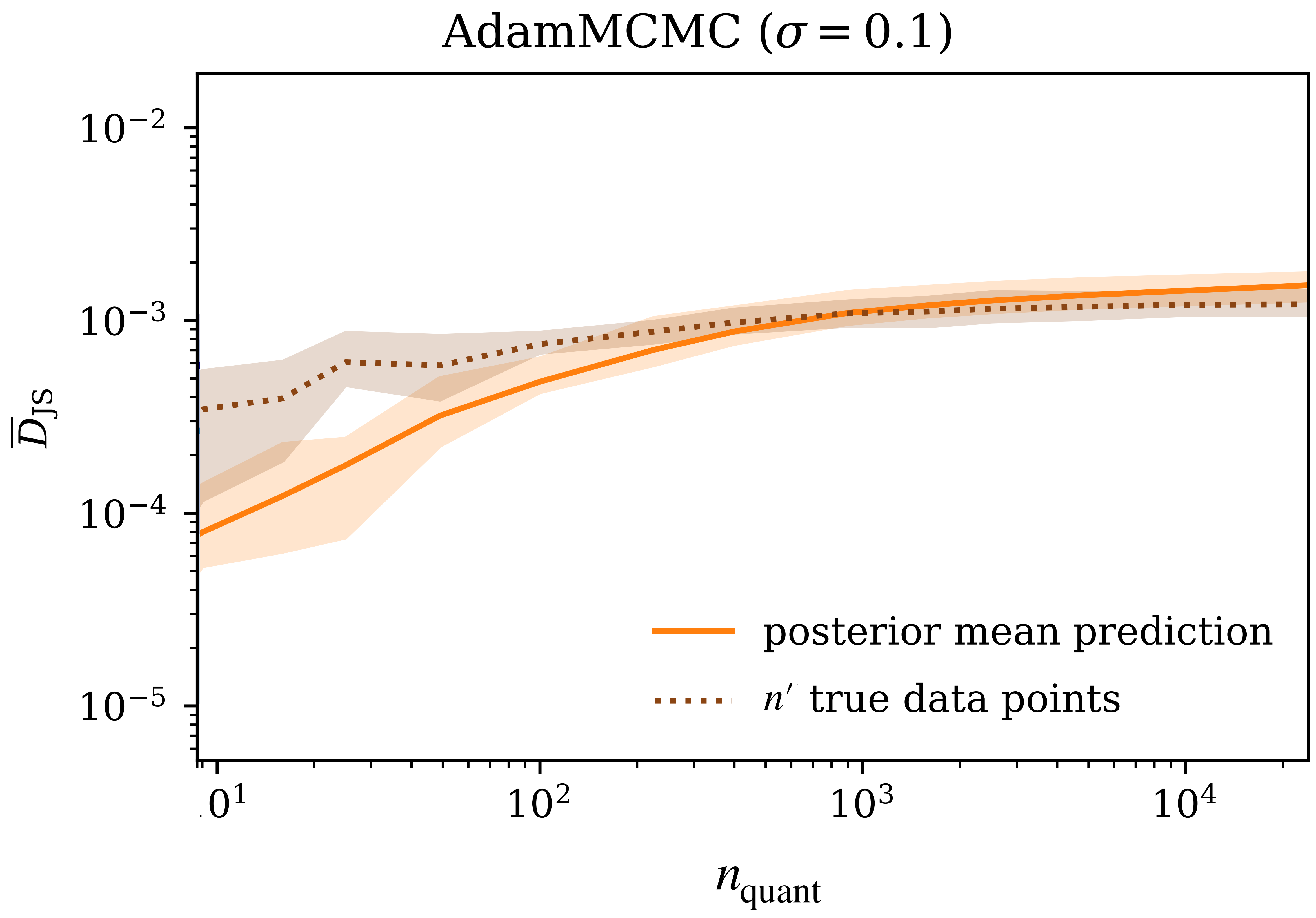
stochastic gradient MCMC

- replace stochastic gradient descent with a Markov chain $\theta_{t+1} \sim p(\theta_{t+1} \mid \theta_t)$
- after some burn-in b and with some gap-length c we can draw $\theta_b, \ldots, \theta_{b+cn} \sim \pi(\theta \mid D_n)$



Conclusion

- The amplification effect can be estimated from the uncertainty estimate of a generative BNN
- It is only truthful when the network is well calibrated



Read the paper: [arXiv:2408.00838](https://arxiv.org/abs/2408.00838)