Uncertainty Metrics

Sascha Diefenbacher

AI and the Uncertainty Challenge in Fundamental Physics 2023

Uncertainty Quantification

- ML methods lead to great improvements in in High Energy Physics
  - Tagging, fast simulation, analyses, …
  - Uncertainty quantification only sometimes take into consideration
- In line with other ML fields
  - Many industry applications do not require UQ to perform well
- Problem in HEP
  - Physics needs error bars

➡ Need to find and benchmark UQ methods

➡ Need comparison metric

ML Performance Metrics

- Standard Classification:
  - Simple goal: Maximize correct prediction
  - Simple metrics: Accuracy, ROC curves, AUC scores

- Classification with Uncertainty
  - Complex goals: maximize correct predictions, give accurate confidence interval, minimize confidence interval
  - Complex interactions: **accurate** confidence interval vs. **minimal** confidence interval

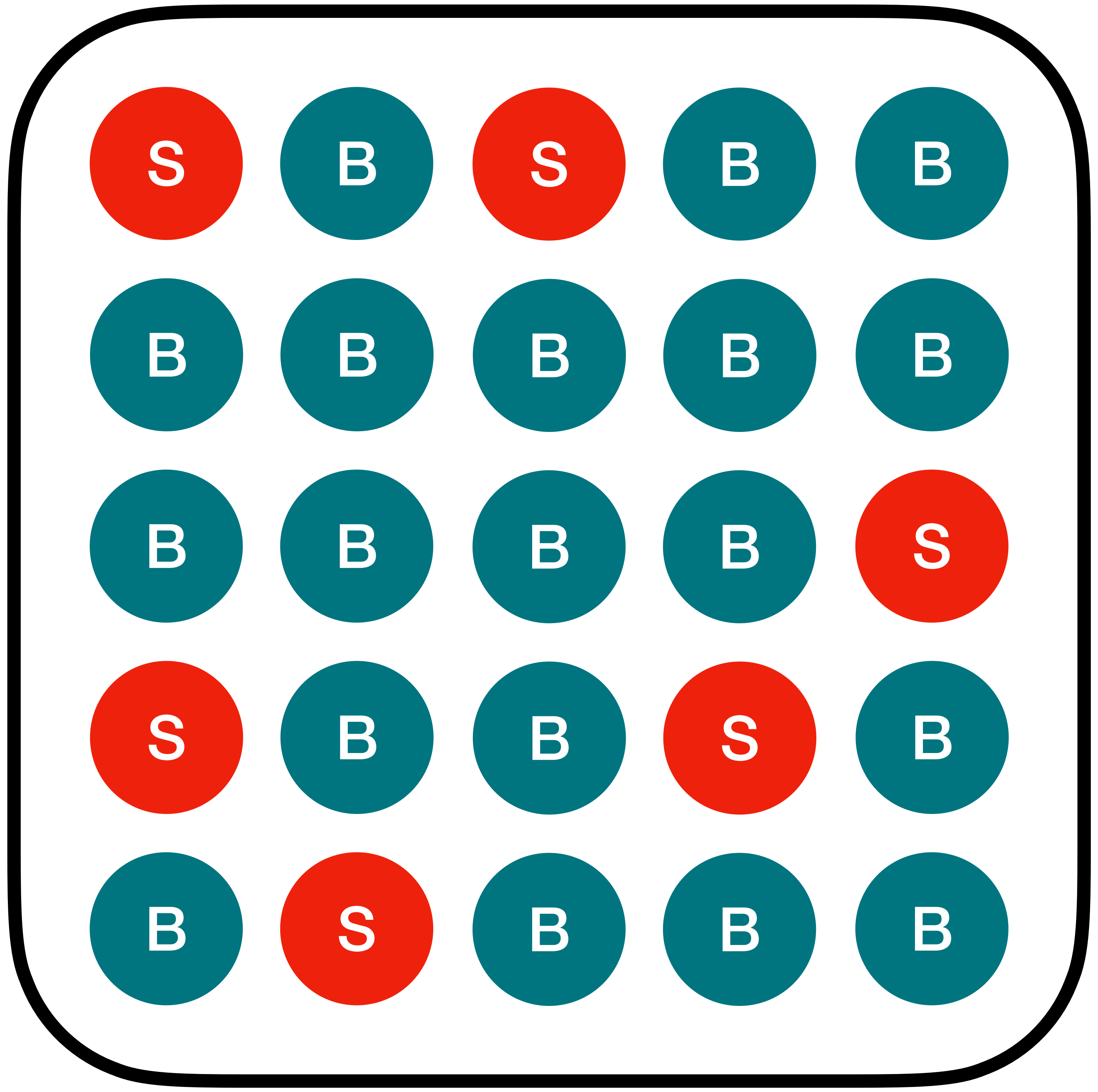
➡ Complex metrics

Uncertainty Quantification Metrics

- Quantities we care about:
  - Simple example

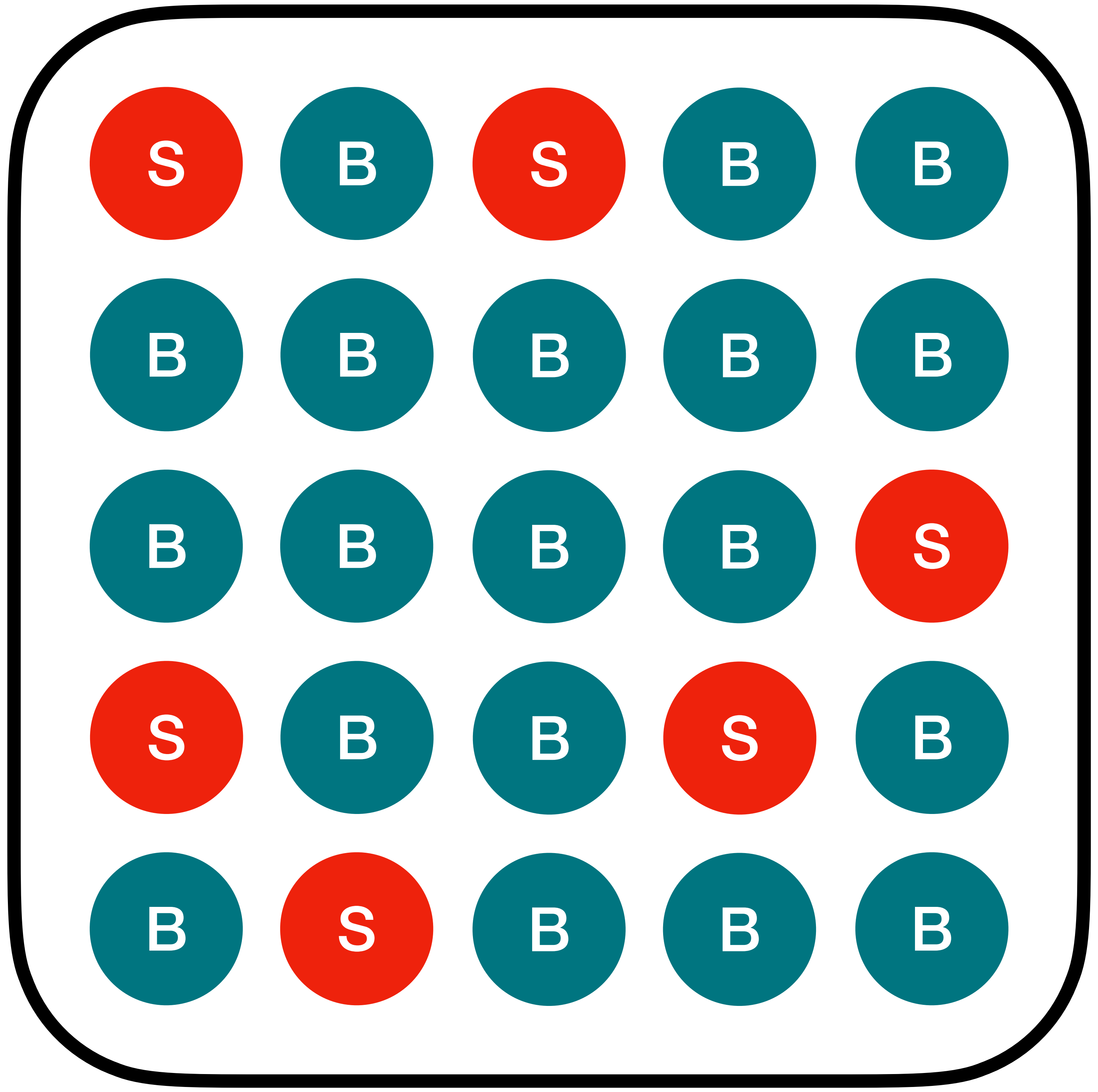
Uncertainty Quantification Metrics

- Quantities we care about:
  - Simple example
  - Sample containing
    - S signal events
    - B background events



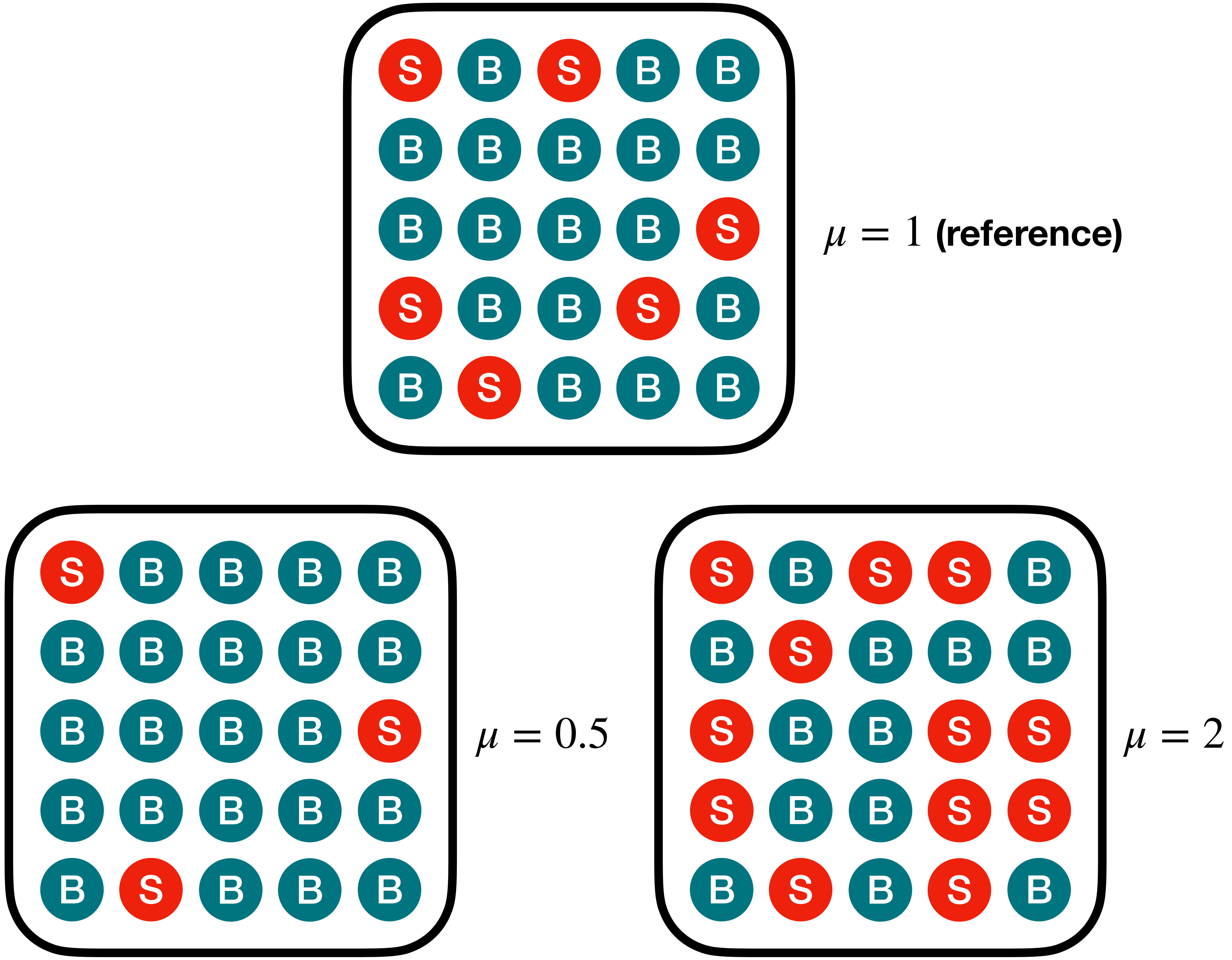
Uncertainty Quantification Metrics

- Quantities we care about:
  - Simple example
  - Sample containing
    - S signal events
    - B background events
  - Determine signal rate μ



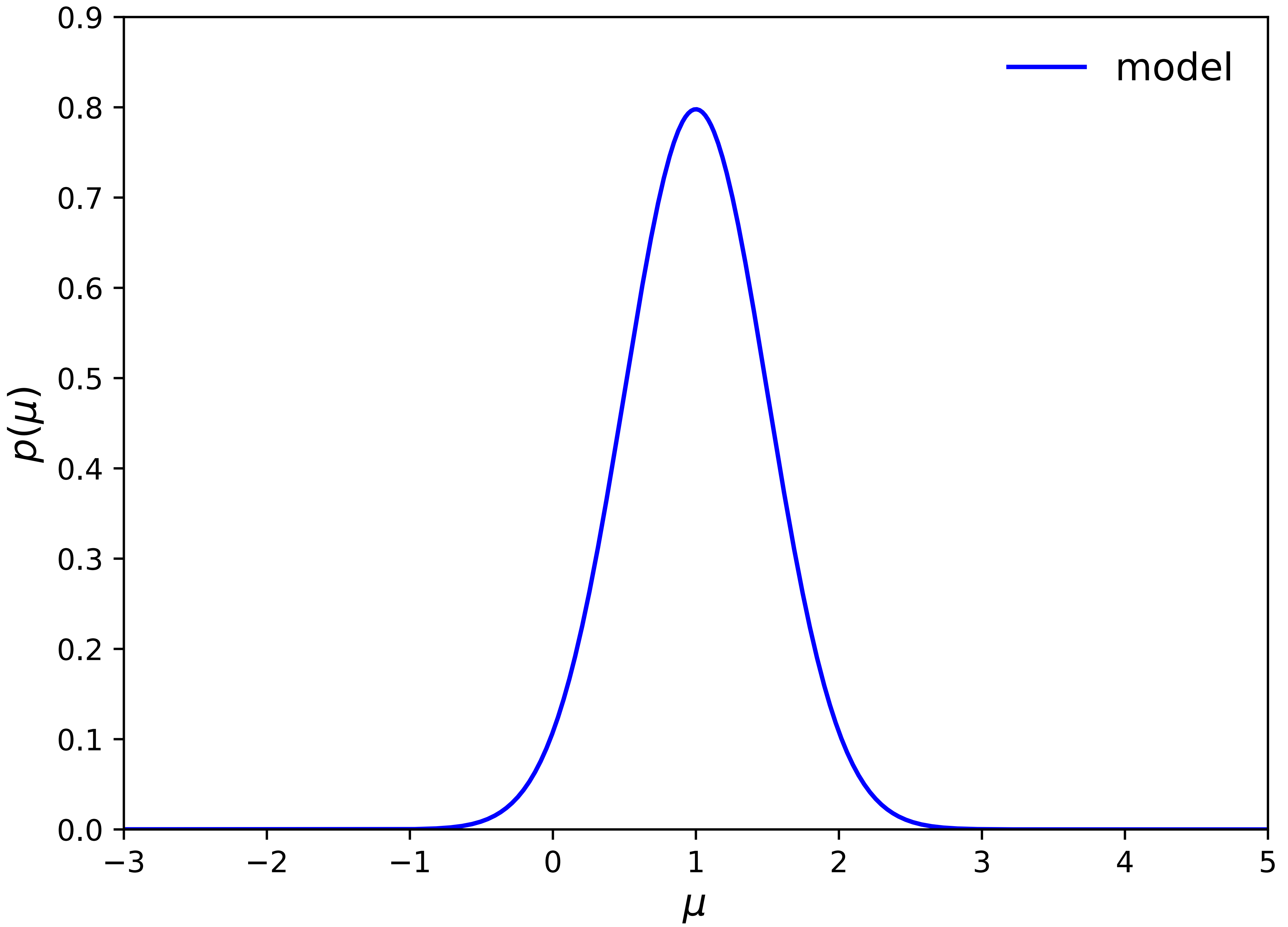
Uncertainty Quantification Metrics

- Quantities we care about:
  - Simple example
  - Sample containing
    - S signal events
    - B background events
  - Determine signal rate μ
  - Determine signal rate μ relative to reference sample



Uncertainty Quantification Metrics

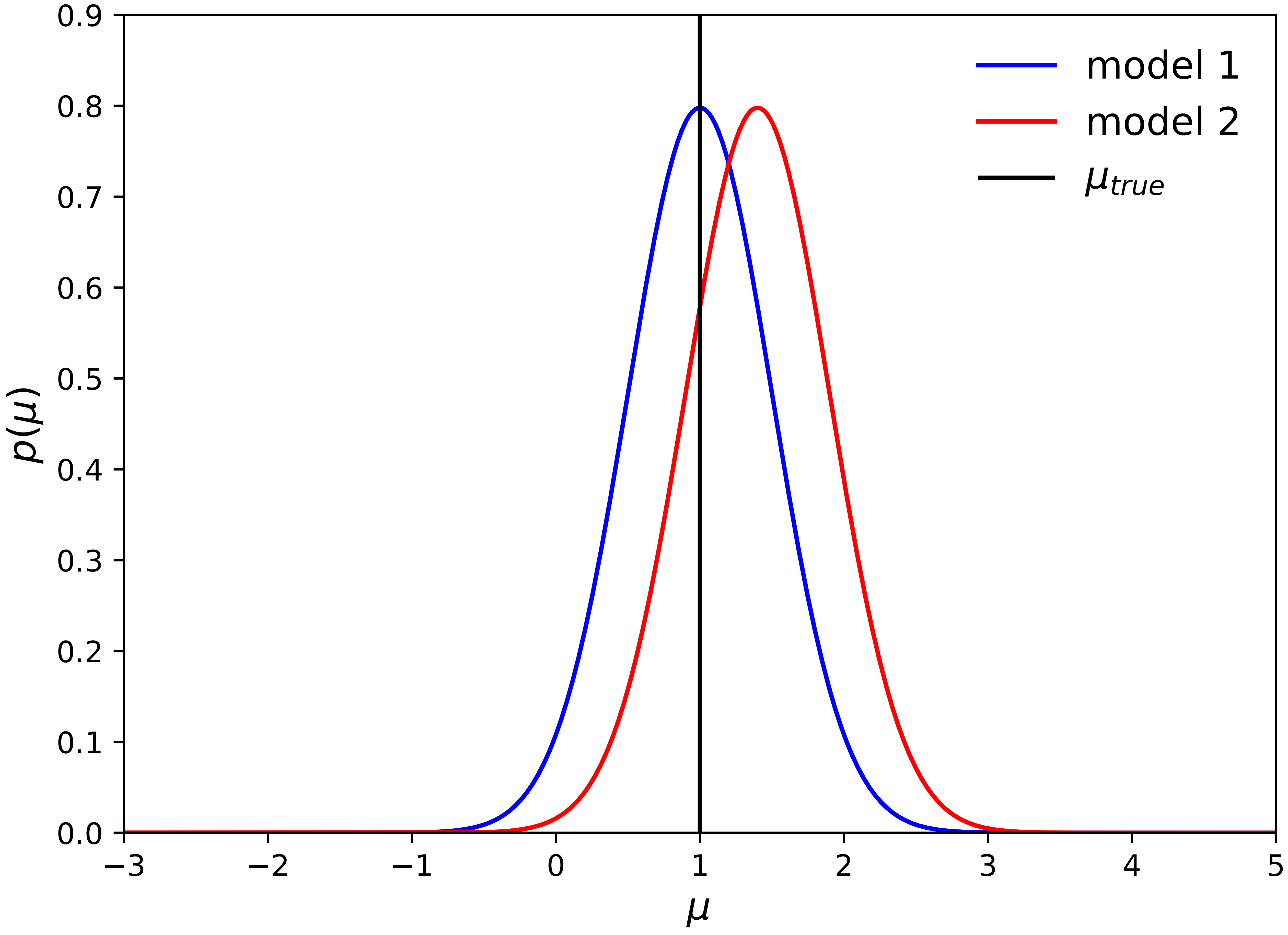
- Determine signal rate μ relative to reference sample
  - UQ method returns likelihood $p(\mu)$



Uncertainty Quantification Metrics

- Quantities we care about:

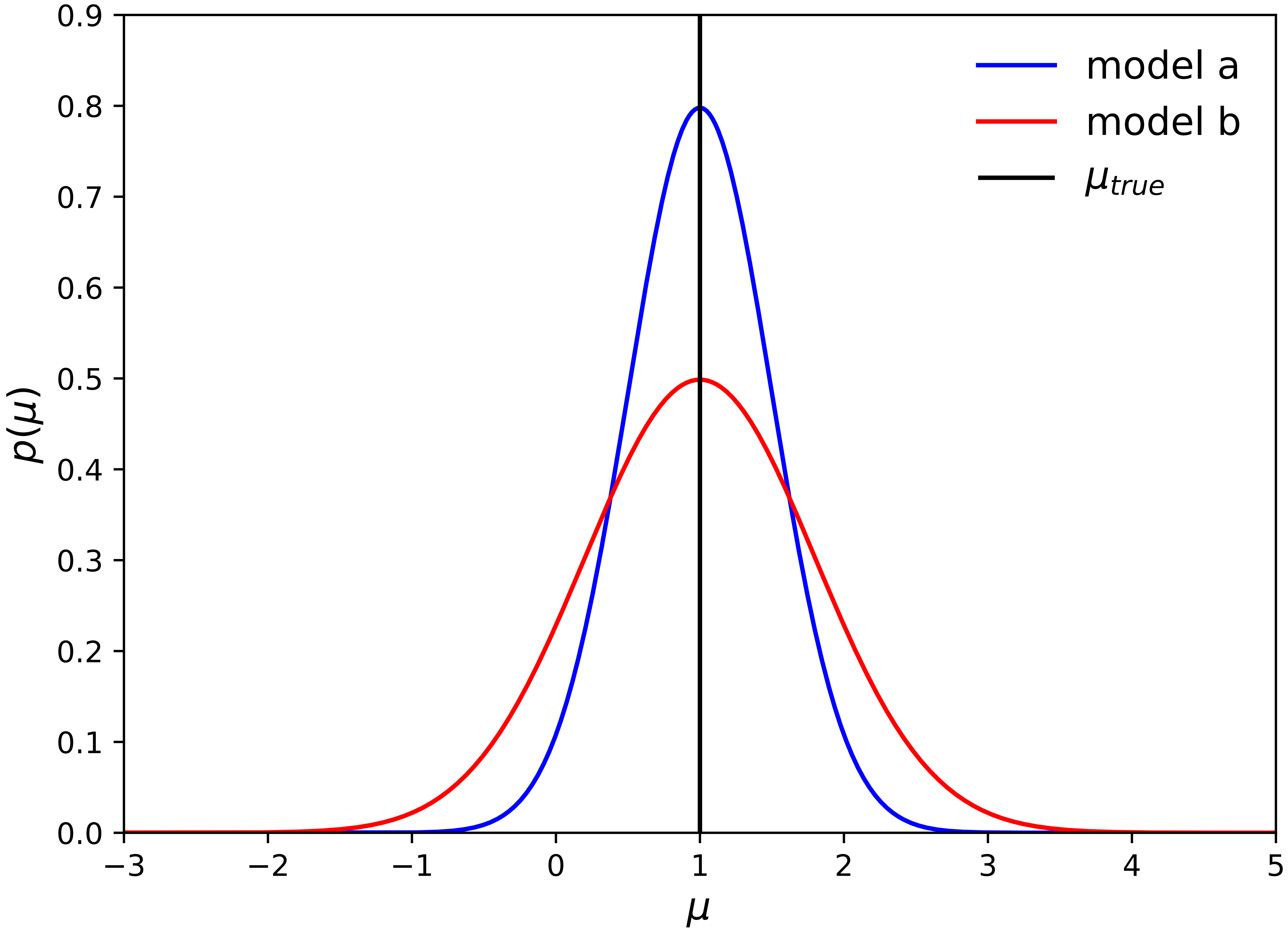
1. Prediction accuracy
    - How close is the predicted μ to μ_{true}



Uncertainty Quantification Metrics

- Quantities we care about:

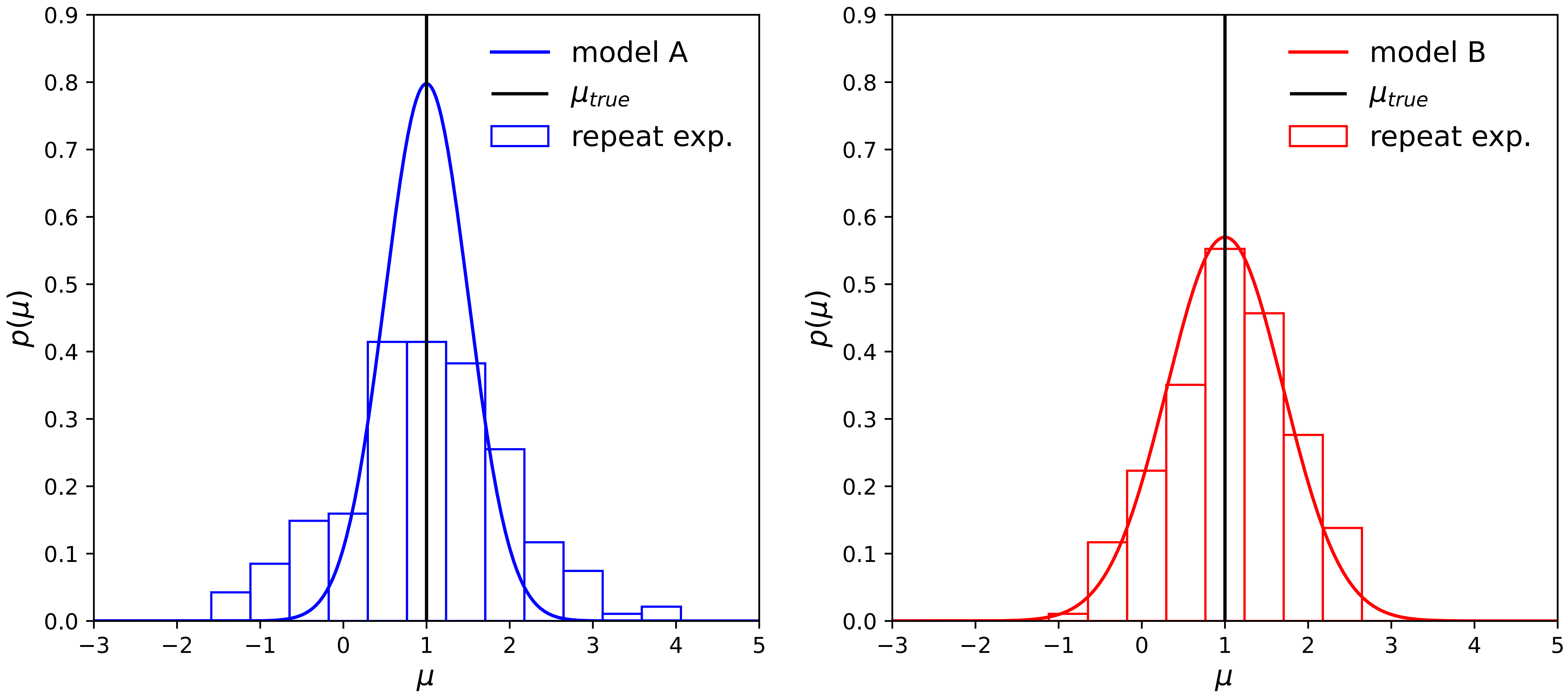
1. Prediction accuracy
2. Prediction uncertainty
   - How large is the uncertainty on the predicted μ?



Uncertainty Quantification Metrics

- Quantities we care about:

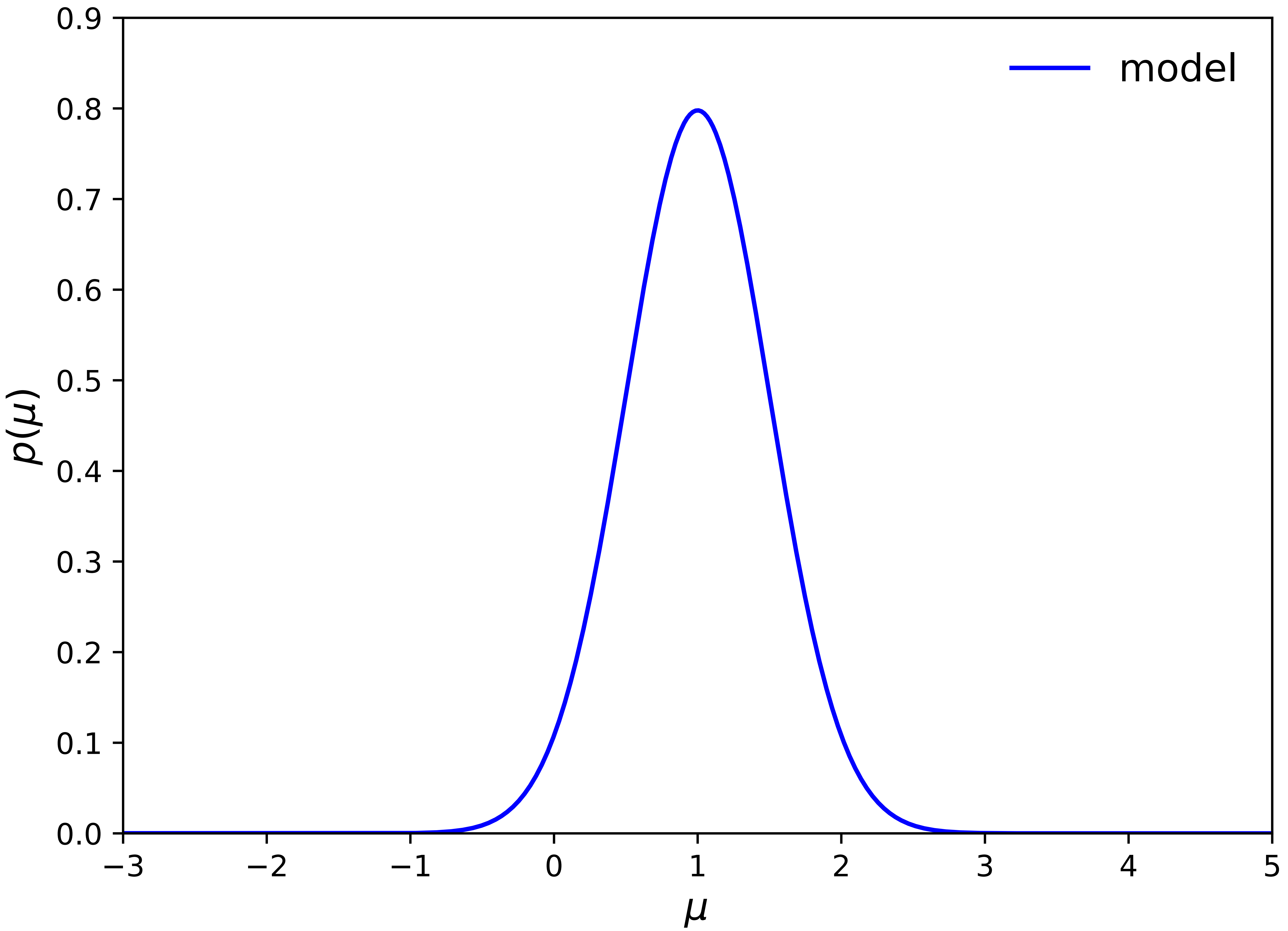
1. Prediction accuracy
2. Prediction uncertainty
3. Uncertainty coverage
    - Does the predicted uncertainty match the observed uncertainty?



Uncertainty Quantification Metrics

- Quantities we care about:

1. Prediction accuracy
2. Prediction uncertainty
3. Uncertainty coverage
4. Uncertainty Quantification
    - Our metric should only work on methods that do quantify an uncertainty



Uncertainty Quantification Metrics

- First Idea: MSE/MAE of relevant quantities
- For N test sets and predicted $\mu_i, \Delta\mu_i, i \in [0,N]$, calculate:

$$\text{MAE}_\mu = \frac{1}{N}\sum_{i=0}^{N} |\mu_i - \mu_{\text{true},i}|$$

$$\text{MSE}_\mu = \frac{1}{N}\sum_{i=0}^{N} (\mu_i - \mu_{\text{true},i})^2$$

$$\text{MAE}_{\Delta\mu} = \frac{1}{N}\sum_{i=0}^{N} |(\mu_i - \mu_{\text{true},i}) - \Delta\mu_i|$$

$$\text{MSE}_{\Delta\mu} = \frac{1}{N}\sum_{i=0}^{N} ((\mu_i - \mu_{\text{true},i}) - \Delta\mu_i)^2$$

$$\text{Score}_\mu = \text{MAE}_\mu + \text{MAE}_{\Delta\mu}$$

$$\text{Score}_\mu = \sqrt{\text{MSE}_\mu + \text{MSE}_{\Delta\mu}}$$

Uncertainty Quantification Metrics

- Alternative Idea: Quantile score:
  - Method should return interval $[\mu_{16}, \mu_{84}]$
  - Corresponds to central 68% quantile of likelihood function
  - Also corresponds to interval defined by 1 standard deviation (under Gaussian uncertainty assumption)
  - Interval can also be defined with Bayesian methods that output a posterior

Uncertainty Quantification Metrics

- For N test sets and predicted $[\mu_{16}, \mu_{84}]_i, i \in [0,N]$
  - Calculate fraction of times interval contains μ_{true} to get coverage c:

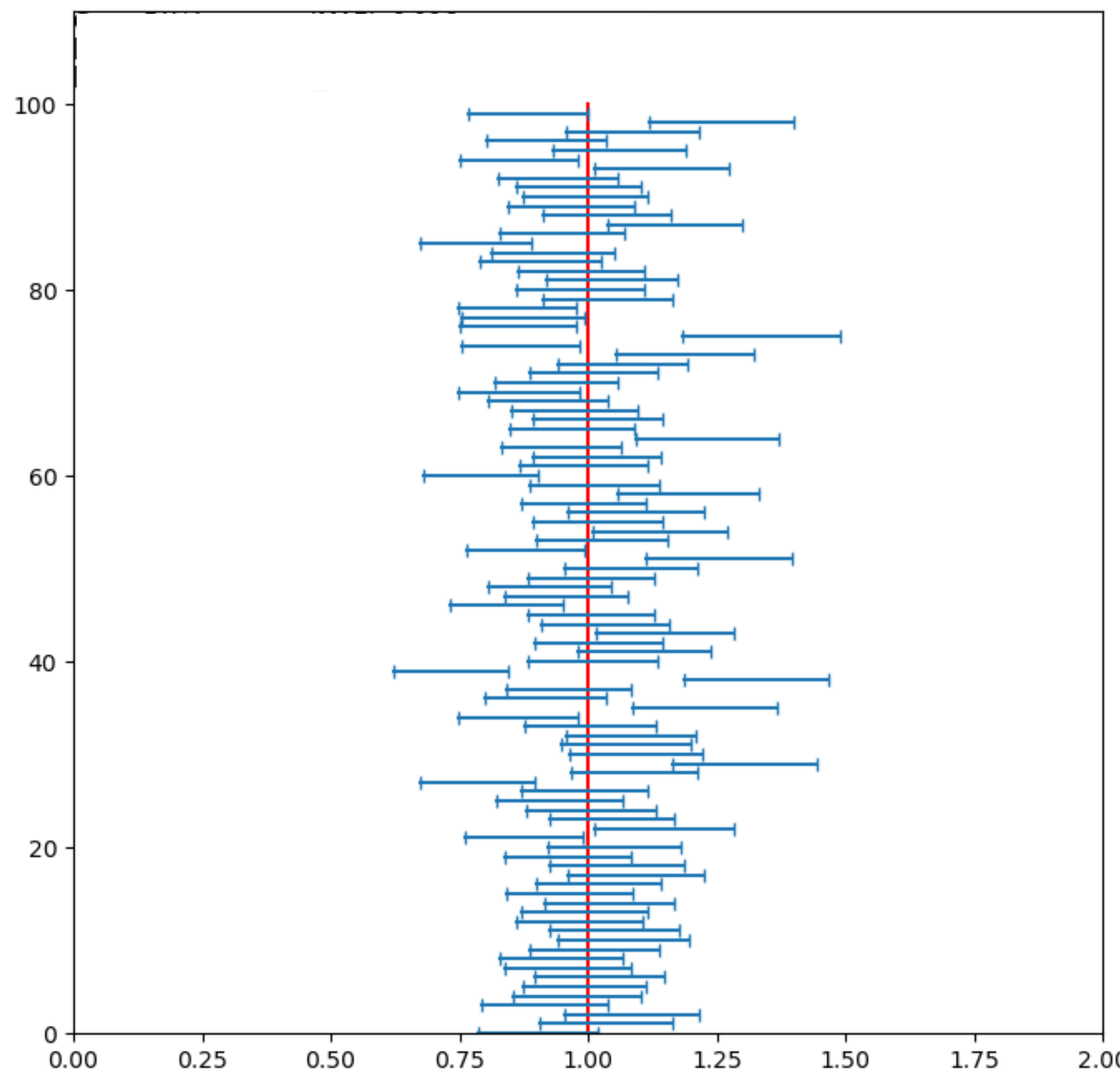
$$c = \frac{1}{N}\sum_{i=0}^{N} 1 \text{ if}(\mu_{\text{true},i} \in [\mu_{84} - \mu_{16}]_i)$$

  - Calculate average interval width w:

$$w = \frac{1}{N}\sum_{i=0}^{N} \mu_{84,i} - \mu_{16,i}$$

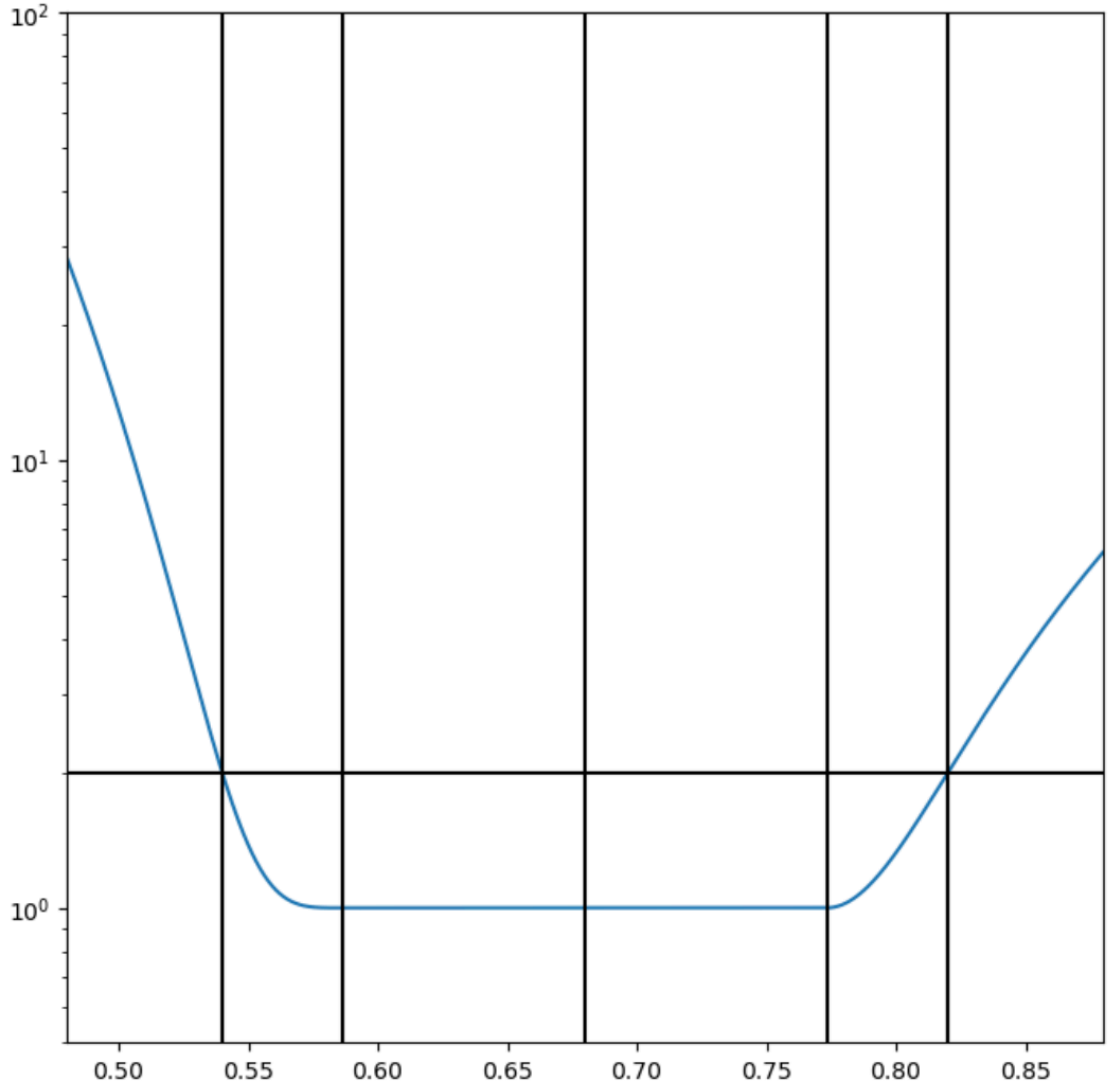
  - Combine both values for score s:

$$s = w\, f(c)$$



Uncertainty Quantification Metrics

- Combine both values for score $s = w\, f(c)$
- Scaling function f:
  - Ideal coverage: 0.68 (68% interval)
  - $f = 1$ around $c = 0.68$
  - Power scaling outside of $c = 0.68$
  - Stricter penalty for undershooting



Uncertainty Quantification Metrics

- 3 remaining problems with $s = w\, f(c)$:
  1. Scores can become very large
  2. Lower scores winning is unintuitive

  ➡ $s = -\ln[w\, f(c)]$

1	Ihsan Ullah	2	2023-11-02	[redacted]
2	ragansu	5	2023-11-03	515061194.28
3	ragansu	5	2023-11-03	[redacted]

Uncertainty Quantification Metrics

- 3 remaining problems with $s = w\, f(c)$:
  1. Scores can become very large
  2. Lower scores winning is unintuitive

     ➡ $s = -\ln[w\, f(c)]$
  3. Methods that return $\mu_{16} = \mu_{84}$ always win,

     since $w = 0$ ➡ $s = \inf$

     ➡ $s = -\ln[(w + \epsilon)\, f(c)]$

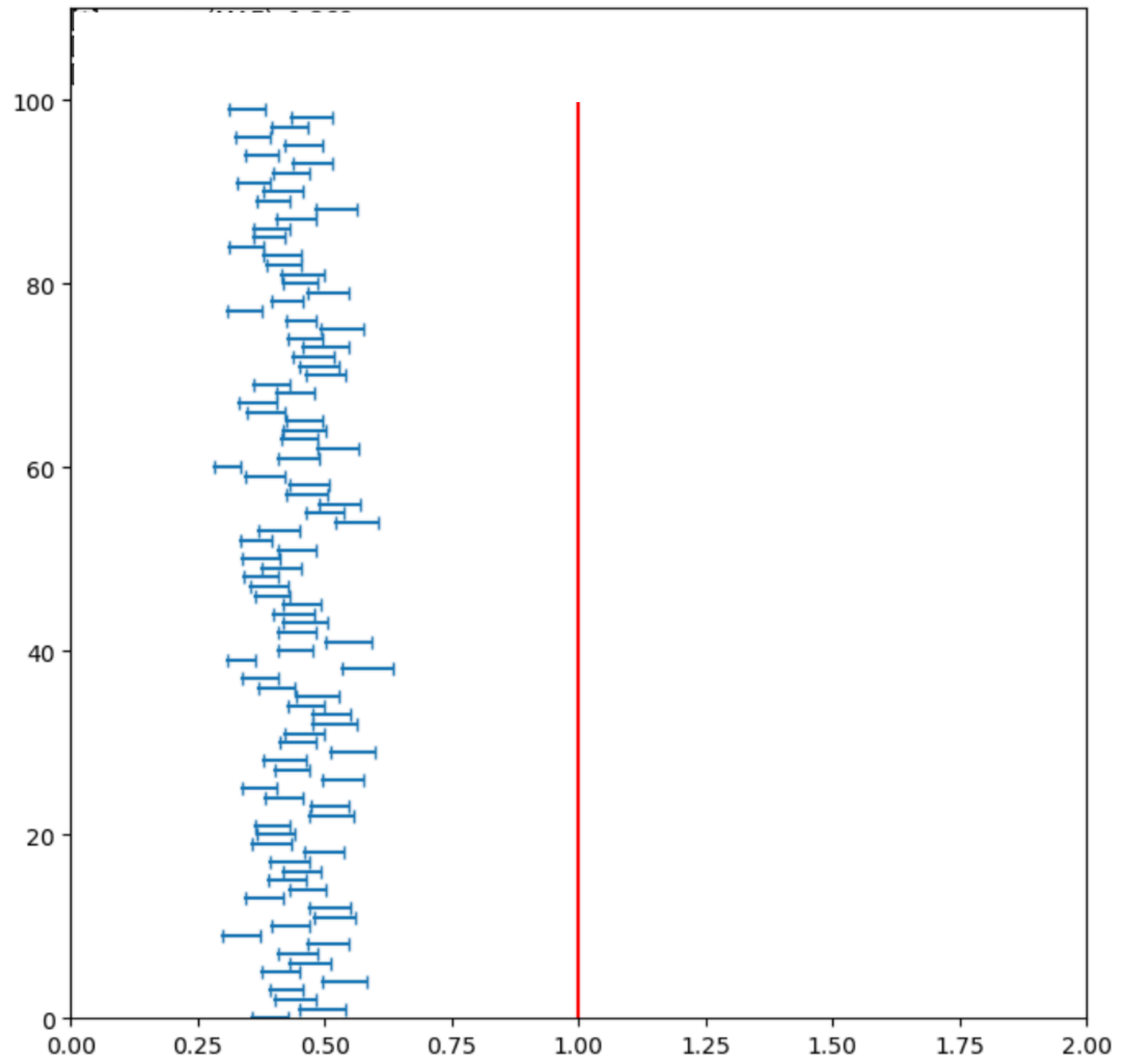
     choose ϵ significantly smaller than minimal width

1	Ihsan Ullah	2	2023-11-02	[redacted]
2	ragansu	5	2023-11-03	515061194.28
3	ragansu	5	2023-11-03	[redacted]

Uncertainty Quantification Metrics

- Quantities we care about: $s = -\ln[(w + \epsilon)\, f(c)]$

1. Prediction accuracy ✓ covered by c
2. Prediction uncertainty
3. Uncertainty coverage
4. Uncertainty Quantification



Uncertainty Quantification Metrics

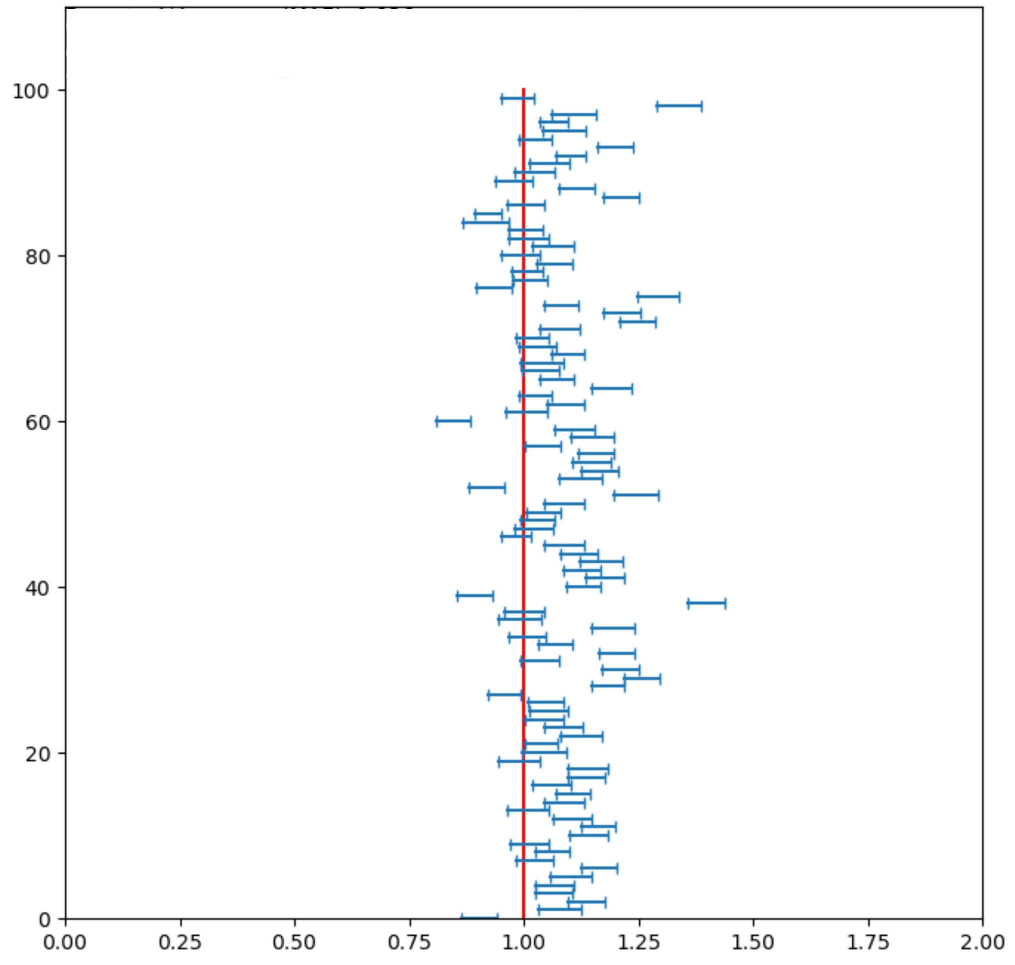
- Quantities we care about: $s = -\ln[(w + \epsilon)\, f(c)]$

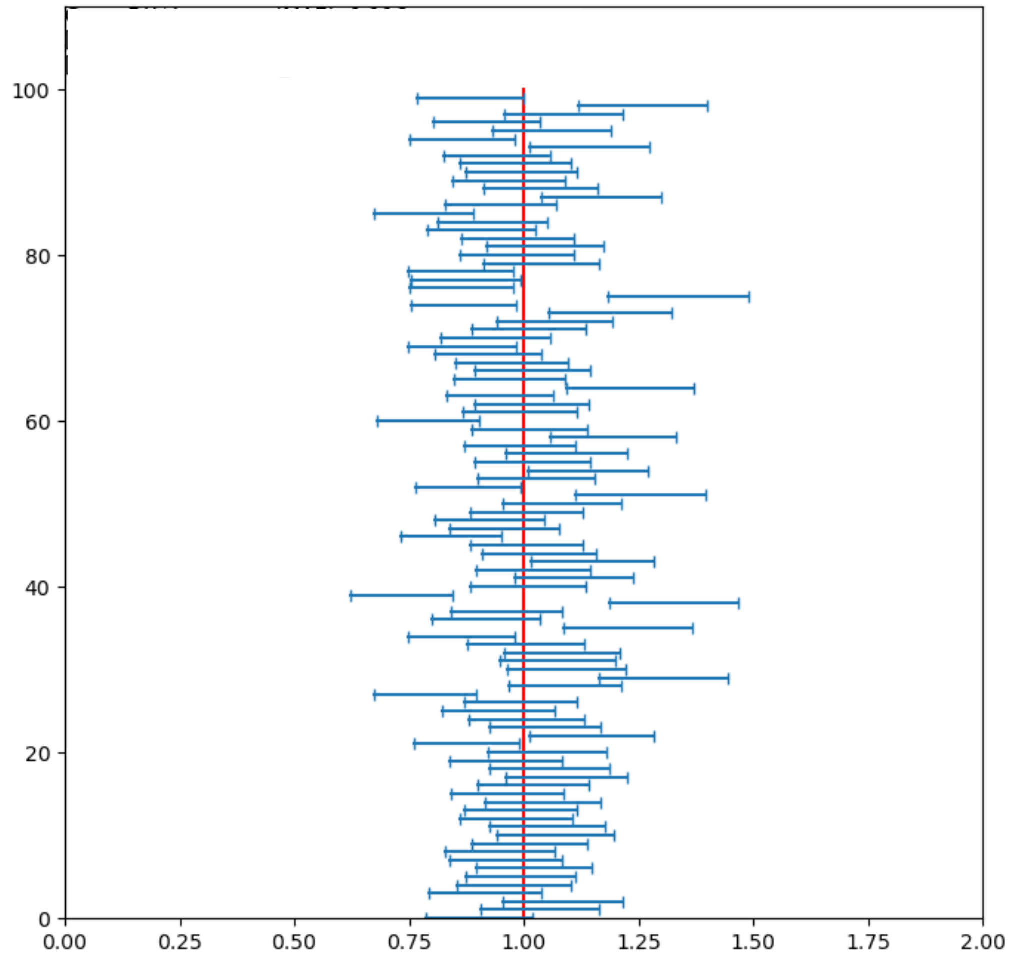
1. Prediction accuracy ✓ covered by c
2. Prediction uncertainty ✓ covered by w
3. Uncertainty coverage
4. Uncertainty Quantification

Uncertainty Quantification Metrics

• Quantities we care about: $s = -\ln[(w+\epsilon)\, f(c)]$

1. Prediction accuracy ✔ covered by c
2. Prediction uncertainty ✔ covered by w
3. Uncertainty coverage ✔ covered by c
4. Uncertainty Quantification





Uncertainty Quantification Metrics

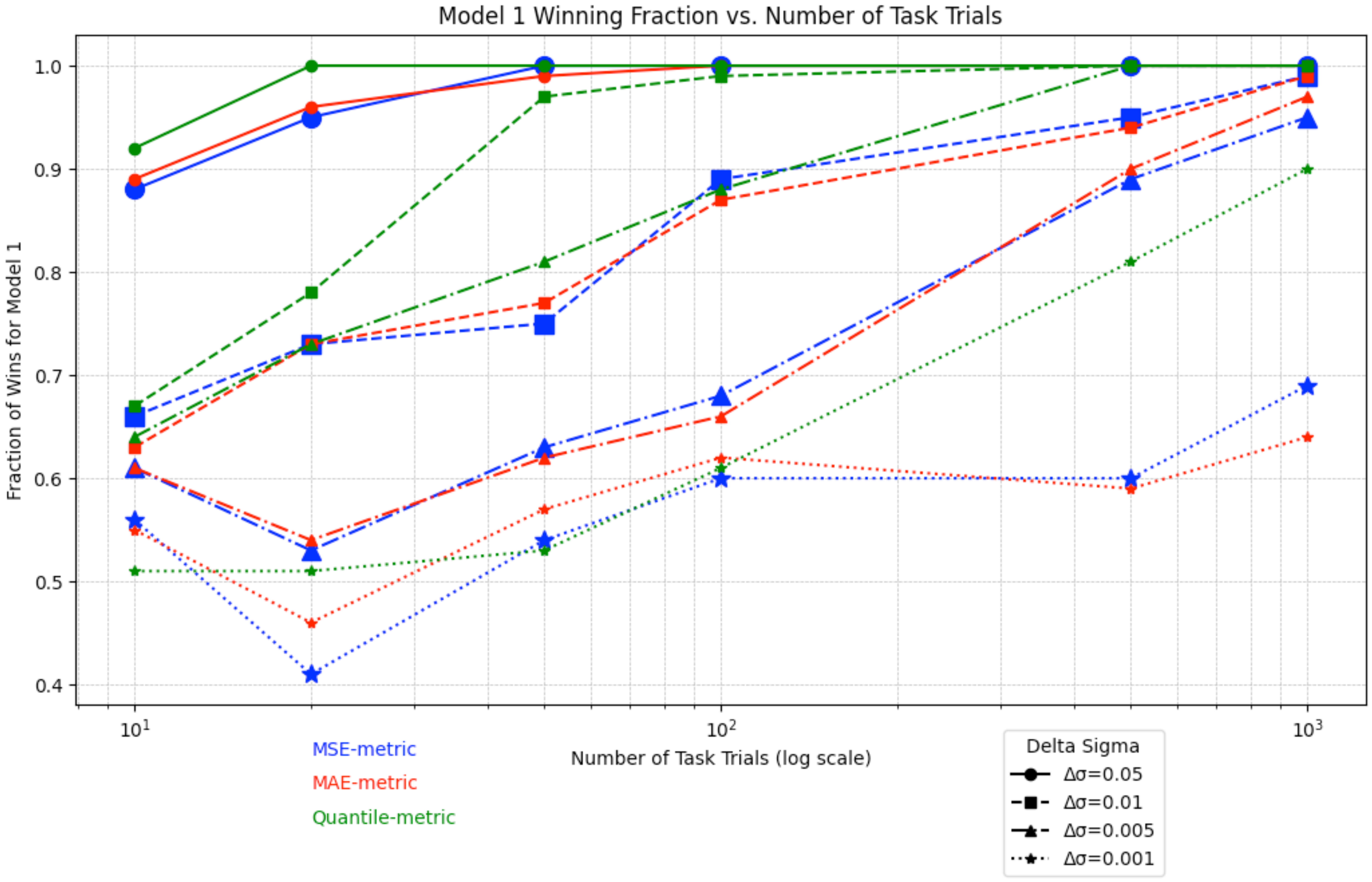
- Quantities we care about: $s = -\ln[(w + \epsilon)\, f(c)]$

1. Prediction accuracy ✓ covered by c
2. Prediction uncertainty ✓ covered by w
3. Uncertainty coverage ✓ covered by c
4. Uncertainty Quantification ✓ per interval definition

UQ Metrics Metric

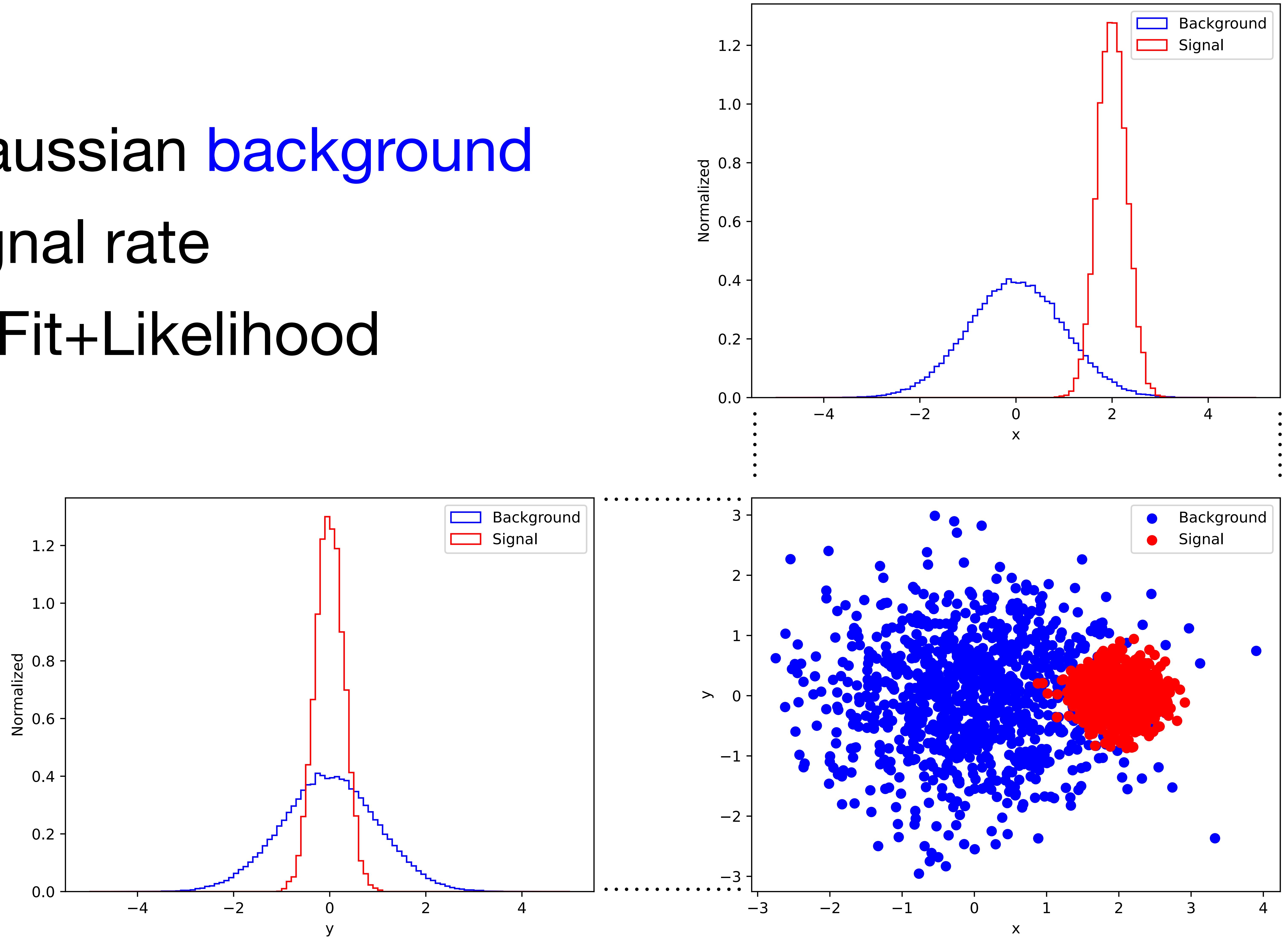
UQ Metrics Metric

- Define simple Poisson UQ task
  - Model 1 nominal solution
  - Model 2 disturbed nominal solution
- Model 1 objectively better than Model 2
- How well can metrics differentiate the two



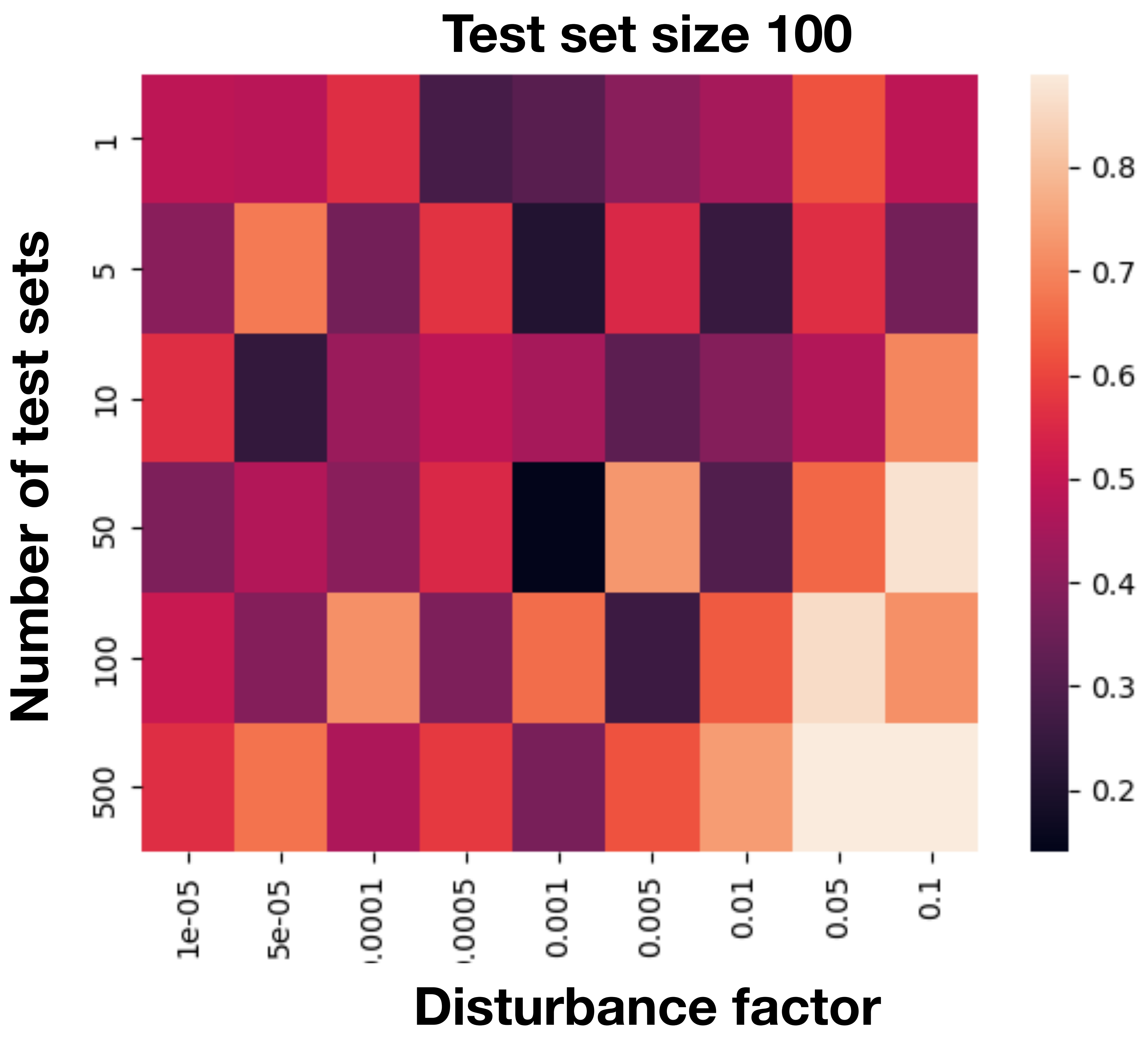
UQ Metrics Metric

- Example 2:
  - Gaussian signal, gaussian background
  - Task: determine signal rate
  - Method: Gaussian Fit+Likelihood evaluation



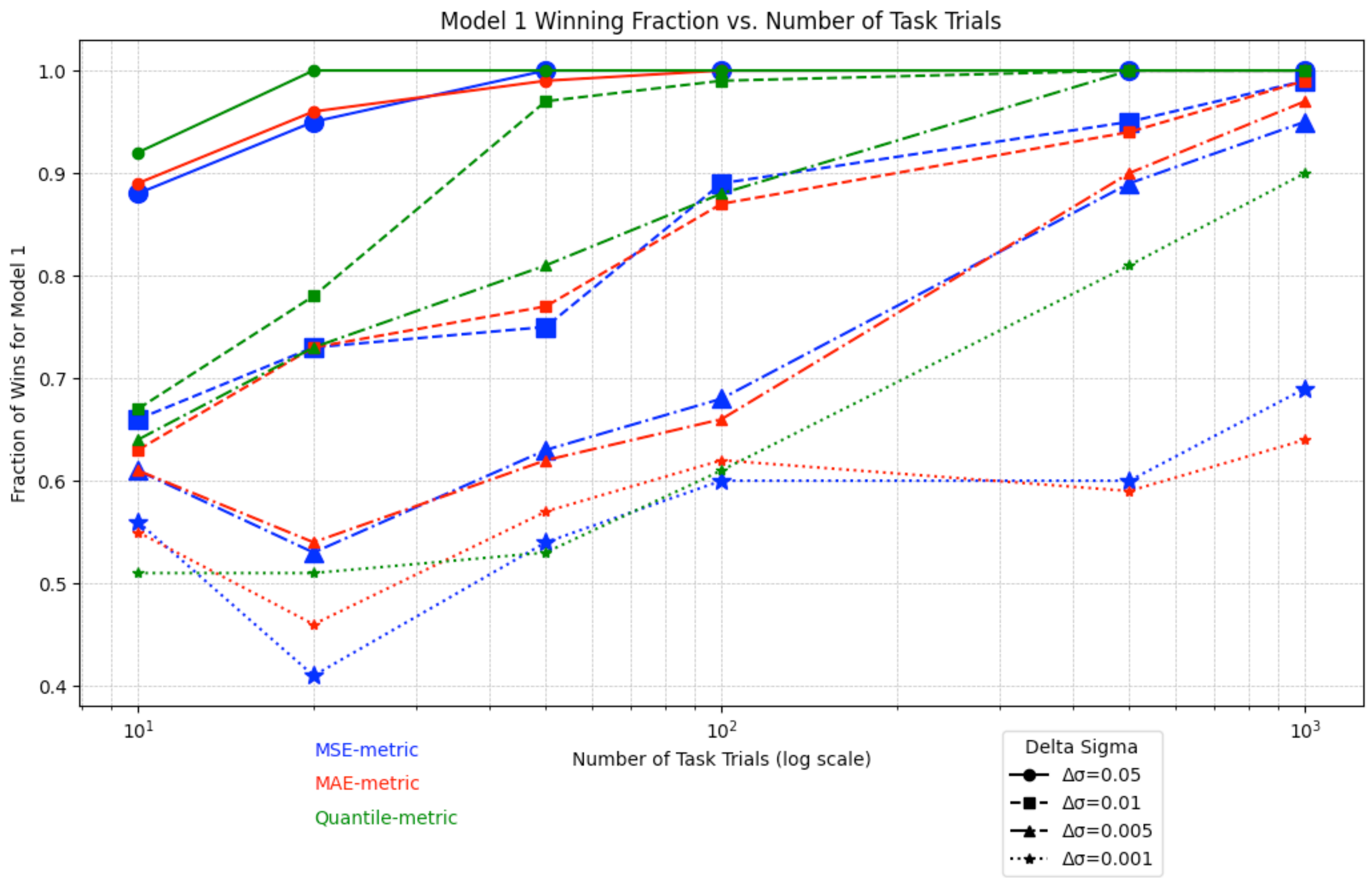
UQ Metrics Metric

- Model 1: near perfect fit, Model 2: disturbed by factor
- Heatmap: rate of quantile metric correctly identifying model 1 as better
- Check stability of metric under hyper-parameter changes



Conclusion

- Proposed quantile metric appears stable
  - Differentiate models of different qualities
  - Current go-to approach for Fair Universe HEP challenge (hackathon, more details on challenge and score Wednesday)
- Topic open for discussion, interested about everyone's input



Comments