Carnegie Mellon University
Statistics & Data Science

Adaptively Calibrated Optimization-Based Confidence Intervals for Inverse Problem Uncertainty Quantification

Mike Stanley (CMU) | February 28, 2024 | SIAM UQ 2024 | Trieste, Italy

Collaborators

Pau Batlle (Caltech)

Pratik Patil (UC Berkeley)

Mikael Kuusela (CMU)

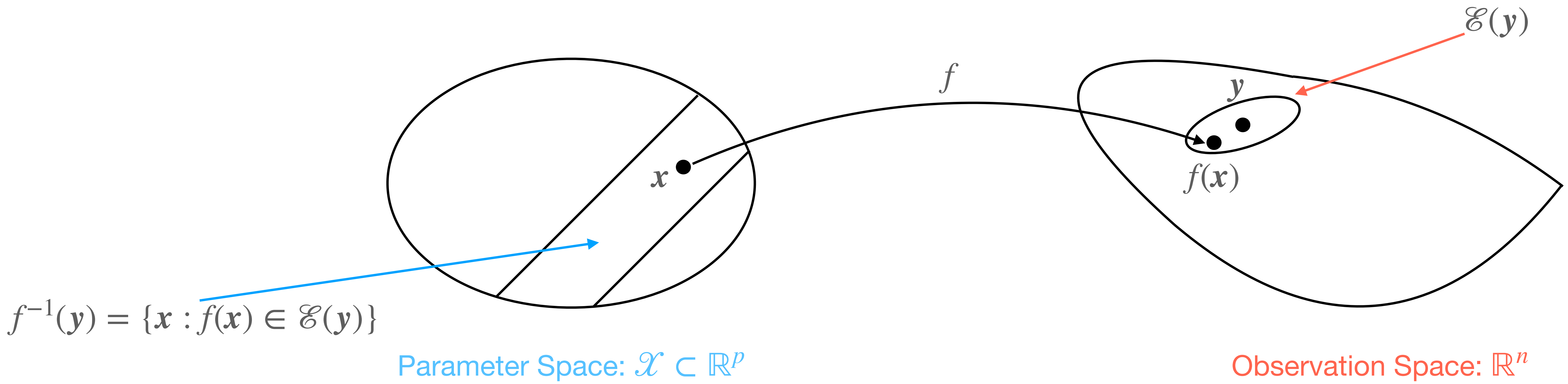
Houman Owhadi (Caltech)

Carnegie Mellon University
Statistics & Data Science

Caltech

Berkeley
UNIVERSITY OF CALIFORNIA

Setting of interest



- Deterministic forward model: $f : \mathbb{R}^p \to \mathbb{R}^n, \boldsymbol{x} \mapsto f(\boldsymbol{x})$, (e.g., $f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$, $\boldsymbol{K} \in \mathbb{R}^{n \times p}$)
- Forward model parameter constraints: $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ (e.g., $\boldsymbol{x} \geq \boldsymbol{0}$)
- Additive noise: $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, (more generally, $\boldsymbol{y} \sim P_{\boldsymbol{x}}$)
- Inferential object(s): parameter functionals, $\varphi(\boldsymbol{x}) \in \mathbb{R}$ (e.g., $\varphi(\boldsymbol{x}) = \boldsymbol{h}^T \boldsymbol{x}$)
- Applications where this setting arises: carbon flux inversion [Stanley et al., 2024b], remote sensing (XCO2) [Patil et al., 2022], and particle unfolding [Kuusela, 2016], [Stanley et al., 2022], [Han et al., 2023].

UQ in this setting and some challenges

Inverse Problem Uncertainty Quantification

=

Reporting statistically guaranteed uncertainty quantification of the inferred functional value following from the noisy observation and the forward model

Statistically guaranteed: a confidence interval, $I(\boldsymbol{y})$, with a ***coverage*** guarantee, i.e., $\forall \boldsymbol{x}^* \in \mathscr{X},\ \mathbb{P}\left(\varphi(\boldsymbol{x}^*) \in I(\boldsymbol{y})\right) \geq 1 - \alpha$ for a chosen level $\alpha \in [0,1]$.

- Ill-posed problems make $f^{-1}(\mathscr{E}(\boldsymbol{y}))$ difficult to work with (e.g., null$(\boldsymbol{K}) \neq \{\boldsymbol{0}\}$),
- Making $I(\boldsymbol{y})$ constraint-aware (e.g., $\boldsymbol{x} \geq \boldsymbol{0}$) while retaining the desired coverage guarantee is highly non-trivial.

Optimization-based confidence intervals provide a start to a solution

There is a way to frame the interval computation as two endpoint optimizations

$$I\left(\psi_\alpha^2, \boldsymbol{y}\right) := \left[\varphi^l(\boldsymbol{y}), \varphi^u(\boldsymbol{y})\right] = \left[\min_{\boldsymbol{x} \in D\left(\psi_\alpha^2, \boldsymbol{y}\right)} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x} \in D\left(\psi_\alpha^2, \boldsymbol{y}\right)} \varphi(\boldsymbol{x})\right]$$

such that

$$\forall \boldsymbol{x}^* \in \mathscr{X}, \; \mathbb{P}\left(\varphi(\boldsymbol{x}^*) \in I\left(\psi_\alpha^2, \boldsymbol{y}\right)\right) \geq 1 - \alpha$$

A key challenge: setting ψ_α^2 to obtain this coverage guarantee

where

$$D(\psi_\alpha^2, \boldsymbol{y}) := \left\{\boldsymbol{x} : \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq \psi_\alpha^2 \text{ and } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\right\}.$$

Related references: [Rust/Burrus, 1972], [Stark, 1992], [O'Leary/Rust, 1994], [Tenorio et al., 2007], [Patil et al., 2022], [Stanley et al., 2022], [Batlle et al., 2023]

Optimization-based confidence intervals provide a start to a solution (cont.)

- They provide a start to a solution because they,
  - reframe inference as optimization (good for computation),
  - elegantly handle the parameter constraints in the endpoint optimizations.
- However, setting ψ_α^2 to provide the coverage guarantee turns out to be non-trivial.
  - For ***simultaneous (SSB) coverage*** : $\psi_\alpha^2 := \chi^2_{n,\alpha}$ [Stark, 1992]
  - For ***one-at-a-time (OSB) coverage*** : $\psi_\alpha^2 := \chi^2_{1,\alpha} + s^2$, where $s^2 = \min_{\boldsymbol{x}:\, \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ [Patil et al., 2022], [Rust and O'Leary, 1994], [Stanley et al., 2022]
- However, the OSB setting does not hold in general [Tenorio et al. 2007, Batlle et al. 2023]

An **outline** of this talk and some main **take-aways**

1. Building on the work of [Batlle et al. 2023], we present a method to set ψ_α^2 in a data-dependent way to achieve interval coverage and improve interval length relative to OSB.

    - **Take-away**: our method is the first computationally feasible approach to properly calibrate these optimization-based intervals.
    - We call it **adaOSB** for "adaptive OSB"

2. We explore three numerical studies to demonstrate the method and its advantages.

    - **Take-away**: our method provides coverage in low dimensional ($p = 3$) example where OSB does not, and improves interval length in a scenario where OSB empirically over-covers ($p = 80$).

The optimized interval can be seen an inverted hypothesis test

- There is a particular hypothesis test and log-likelihood ratio test statistic recovering the interval

OSB interval — Inverted hypothesis test

$$\boxed{I\left(\psi_\alpha^2, \boldsymbol{y}\right)} = \boxed{\left\{\mu \in \mathbb{R} : \lambda(\mu, \boldsymbol{y}) \leq q_\alpha\right\}} \text{ where } \psi_\alpha^2 := q_\alpha + s^2$$

$\lambda(\mu, \boldsymbol{y})$ is an LLR test statistic and q_α is a critical value ensuring $\mathbb{P}(\text{type-1 error}) \leq \alpha$ for all $\boldsymbol{x}^* \in \mathcal{X}$

- **Theorem 2.4** [Batlle et al. 2023]: the critical value controlling type-1 error of the test can also be used to calibrate $I(\psi_\alpha^2, \boldsymbol{y})$

The previously mentioned q_α is both difficult to obtain and statistically conservative

- Difficult to obtain: finding q_α involves solving a chance constrained optimization which is known to be strong NP-hard and non-convex [Batlle et al. 2023].

- Statistically conservative: In order for q_α to control type-1 error for all $\boldsymbol{x}^* \in \mathscr{X}$, we are protecting against all potential true parameter states and therefore might be overly conservative.

We will demonstrate a method that avoids solving the complicated optimizations and provides length benefits in some cases

The hypothesis test connection can calibrate these optimization-based intervals

- First, let $Q_{\boldsymbol{x}} : [0,1] \to \mathbb{R}$ be the quantile function of $\lambda(\mu, \boldsymbol{y})$ at $\boldsymbol{x}$, i.e. $\mathbb{P}\left(\lambda(\mu, \boldsymbol{y}) \leq Q_{\boldsymbol{x}}(1-\alpha)\right) = 1 - \alpha$.
- Let $\boldsymbol{x}^* \in \mathcal{X}$ denote the true but unknown parameter. Clearly, if we knew $\boldsymbol{x}^*$, we could compute $Q_{\boldsymbol{x}^*}(1-\alpha)$ and calibrate our interval using $\psi_\alpha^2 := Q_{\boldsymbol{x}^*}(1-\alpha) + s^2$. But we don't!
- ***Core of the idea***: we can always obtain a $1-\eta$ confidence set for $\boldsymbol{x}^*$ by $f^{-1}(\Gamma_\eta(\boldsymbol{y})) := \left\{\boldsymbol{x} \in \mathcal{X} : \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq \chi^2_{n,\eta}\right\}$. We can then calibrate the interval by using $q_\gamma := \max_{\boldsymbol{x} \in f^{-1}(\Gamma_\eta(\boldsymbol{y}))} Q_{\boldsymbol{x}}(1-\gamma)$, such that $(1-\eta)(1-\gamma) = 1 - \alpha$.

  Uncertainty budget, trading off between confidence set (η) and quantile level (γ)
- This idea is similar to [Berger and Boos, 1994] and [Masserano et al., 2024], where tests involving nuisance parameters are controlled maximizing a p-value over a data-informed set.

We estimate q_γ using sampling and quantile regression

- Similar to the ideas present in [Dalmasso et al., 2020] [Dalmasso et al., 2022], [Masserano et al., 2023], Masserano et al., 2024]

adaOSB Algorithm

1. Let $\alpha, \gamma, \eta \in (0,1)$ such that $\gamma(1-\eta)+\eta=\alpha$, and $\Gamma_\eta(\boldsymbol{y})$ be a $1-\eta$ confidence set containing $f(\boldsymbol{x}^*)$. Then $f^{-1}\left(\Gamma_\eta(\boldsymbol{y})\right)$ is a $1-\eta$ confidence set for $\boldsymbol{x}^*$.
2. Generate samples $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_M \sim \mathscr{U}\left(f^{-1}\left(\Gamma_\eta(\boldsymbol{y})\right)\right)$. Task 1
3. Sample LLRs $\lambda_i \sim F_{\tilde{\boldsymbol{x}}_i}$ Task 2
4. Use generated data $\{(\tilde{\boldsymbol{x}}_i, \lambda_i)\}_{i=1}^M$ to fit a quantile regressor, $\hat{q}(\boldsymbol{x})$, and estimate $q_\gamma := \max_{\boldsymbol{x} \in f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathscr{X}} \hat{q}(\boldsymbol{x})$
5. Compute $I\left(q_\gamma + s^2, \boldsymbol{y}\right)$ with the guarantee that $\forall \boldsymbol{x}^* \in \mathscr{X},\ \mathbb{P}_{\boldsymbol{y}\sim P_{\boldsymbol{x}^*}}\left(\varphi(\boldsymbol{x}^*) \in I\left(q_\gamma + s^2, \boldsymbol{y}\right)\right) \geq 1-\alpha$.

The pros and cons of this approach

- By finding $q_\gamma := \max_{\boldsymbol{x} \in f^{-1}(\Gamma_\eta(\boldsymbol{y}))} Q_{\boldsymbol{x}}(1-\gamma)$, we avoid both

  1. Optimizing over a potentially unbounded space (e.g., $\mathcal{X} = \mathbb{R}^p_+$)

  2. Controlling for all $\boldsymbol{x} \in \mathcal{X}$ since we simply focus on the parameter values in $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$ and adequately adjust the quantile we use.

- We shift the complexity to estimating q_γ:

  1. Sample generation is non-trivial - we develop two approaches for this.

  2. Estimating the max quantile via quantile regression

Implementation practicalities

Sampling $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$

- For our examples, we focus on the scenario when $\mathcal{X} = \mathbb{R}^p_+, f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$, and $\varepsilon \sim N(\boldsymbol{0}, \Sigma)$, implying that we want to sample the intersection of the ellipsoid $\mathcal{E}(\boldsymbol{y}) := \{\boldsymbol{x} : \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq \chi^2_{n,\eta}\}$ and the non-negative orthant.
- **Two strategies**
  - Accept/Reject sampling uniformly from the pre-image ellipsoid is possible via [Voelker et al., 2017] but $\mathbb{P}(\boldsymbol{x}_i \in \mathbb{R}^p_+) \to 0$ as p gets large, and therefore becomes practically infeasible in higher dimensions (e.g., $p \geq 10$)
  - MCMC: Convex body [Smith, 1984] or polytope samplers [Chen et al., 2018]
    - We find a bounding polytope of $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$ with hyperplanes defined by both the principal axes of the pre-image ellipsoid, the non-negativity constraints, and H additional randomly chosen hyperplanes.

Implementation practicalities

Quantile Regression

- Once we sample $\{(\tilde{\boldsymbol{x}}_i, \lambda_i)\}_{i=1}^M$, we perform quantile regression to learn $\hat{q}(\boldsymbol{x})$
- In principle, this regression can be done with any supervised learning algorithm using the pin-ball loss (e.g., [Meinshausen, 2006], [Takeuchi et al., 2006], [Dalmasso et al., 2020], [Dalmasso et al., 2021], [Masserano et al., 2023])
    - we use gradient-boosted regression since it has a clean implementation in sklearn.
- **Estimation of** q_γ: we sample in independent MCMC chain, $\{\bar{\boldsymbol{x}}_i\}_{i=1}^M$, and use the maximum out-of-sample predicted γ-quantile: $\hat{q}_\gamma := \max_{i \in [M]} \hat{q}(\bar{\boldsymbol{x}}_i)$
    - **Lemma 3.3** [Stanley et al., 2024]: $\hat{q}_\gamma$ is a consistent estimator of q_γ.
    - **Theorem 1** [Dalmasso et al., 2021]: Quantile regression provides a consistent estimator of the quantile function.

Numerical Examples

Examples we consider

1. **Exposition** of method in simple 2d example [Tenorio et al., 2007] [Batlle et al. 2023]

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I}), \quad \varphi(\boldsymbol{x}) = x_1 - x_2, \quad \mathscr{X} = \mathbb{R}^2_+, \quad \boldsymbol{x}^* = (0.5 \quad 0.5)^T$$

2. **Valid Coverage** in a 3d scenario when OSB fails [Batlle et al. 2023]

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I}), \quad \varphi(\boldsymbol{x}) = x_1 + x_2 - x_3, \quad \mathscr{X} = \mathbb{R}^3_+, \quad \boldsymbol{x}^* = (0 \quad 0 \quad 1)^T$$

3. **Length Improvement** in a high dimensional ($p = 80$) scenario when OSB is empirically valid [Stanley et al. 2022] [Stanley et al. 2024a]

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \varphi(\boldsymbol{x}) = \boldsymbol{h}^T \boldsymbol{x}, \quad \mathscr{X} = \mathbb{R}^{80}_+, \quad \boldsymbol{x}^* \text{ defined mean bin counts}$$

We use $N = 1000$ samples to estimate interval coverage and length of OSB and adaOSB

Example 1: Exposition

2d

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I}), \quad \varphi(\boldsymbol{x}) = x_1 - x_2, \quad \mathcal{X} = \mathbb{R}^2_+, \quad \boldsymbol{x}^* = (0.5 \quad 0.5)^T$$

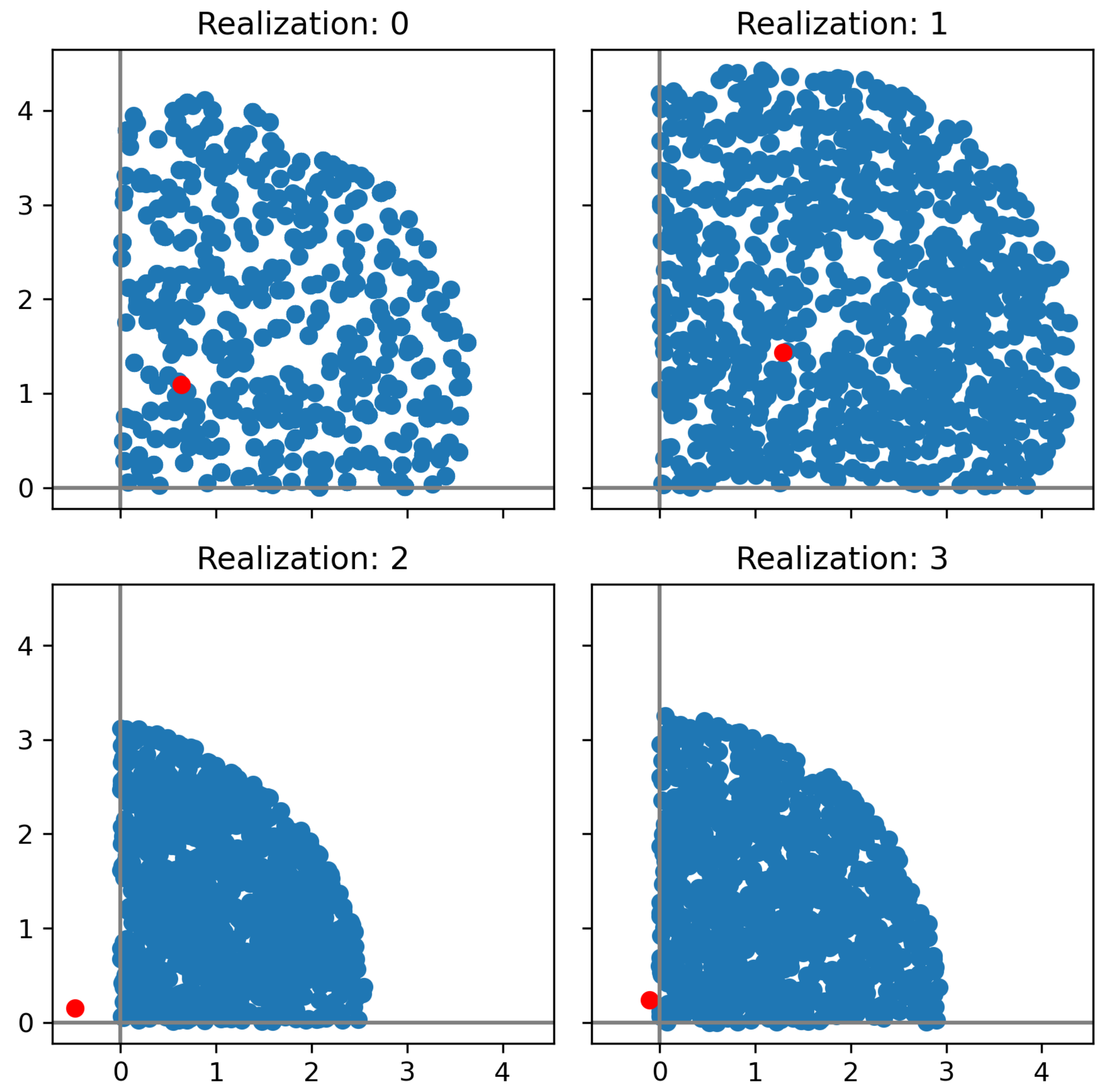
- Since for any $\boldsymbol{x}$, we can efficiently estimate $Q_{\boldsymbol{x}}(1-\gamma)$ in this example using Monte Carlo simulation, we do not use quantile regression, but rather use $\hat{q}_\gamma := \max_{i \in [M]} Q_{\tilde{\boldsymbol{x}}}(1-\gamma)$
- We look to optimize a 68% interval ($\alpha = 0.32$). With $\eta := 0.01$, this implies that $\gamma = 0.3131$.
- Since $p = 2$, our accept/reject ellipsoid sampler is effective for sampling $\tilde{\boldsymbol{x}}_i \sim \mathcal{U}(f^{-1}(\Gamma_\eta(\boldsymbol{y})))$

Example 1: Exposition

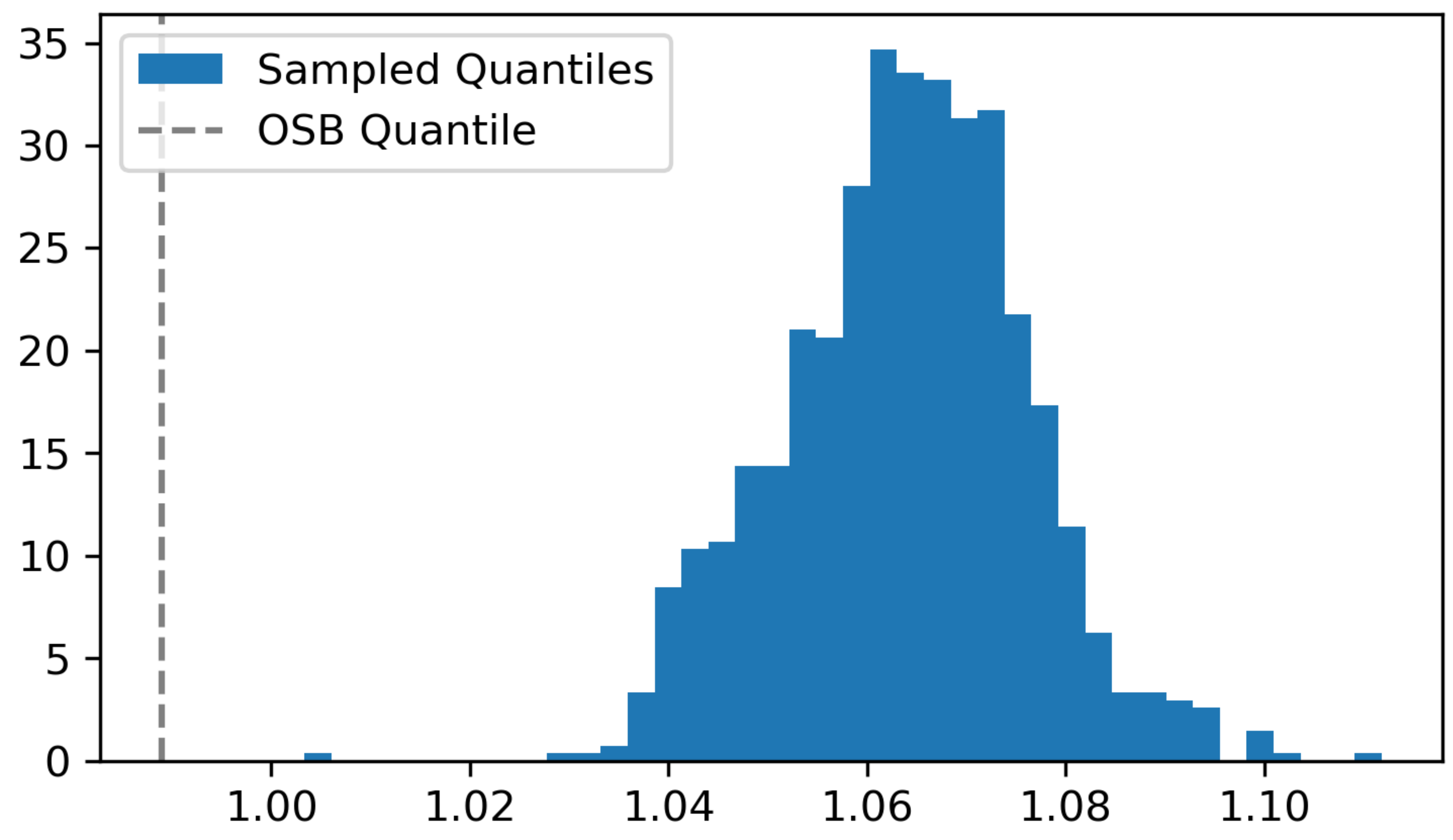
2d - We can see all the moving parts

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}), \quad \varphi(\boldsymbol{x}) = x_1 - x_2, \quad \mathcal{X} = \mathbb{R}^2_+, \quad \boldsymbol{x}^* = (0.5 \quad 0.5)^T$$

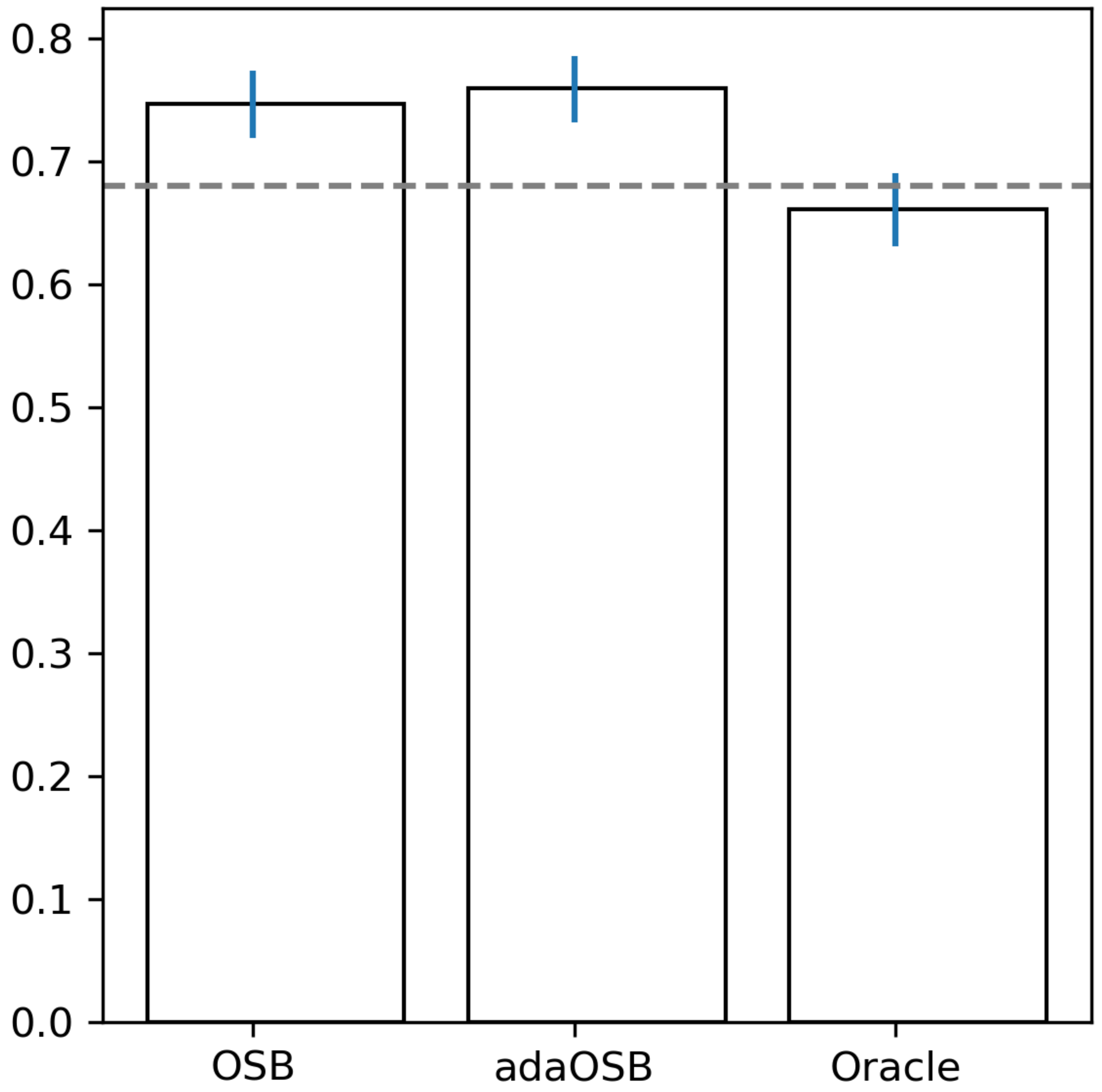
Sampling $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$



Observed distribution of $\hat{q}_\gamma$



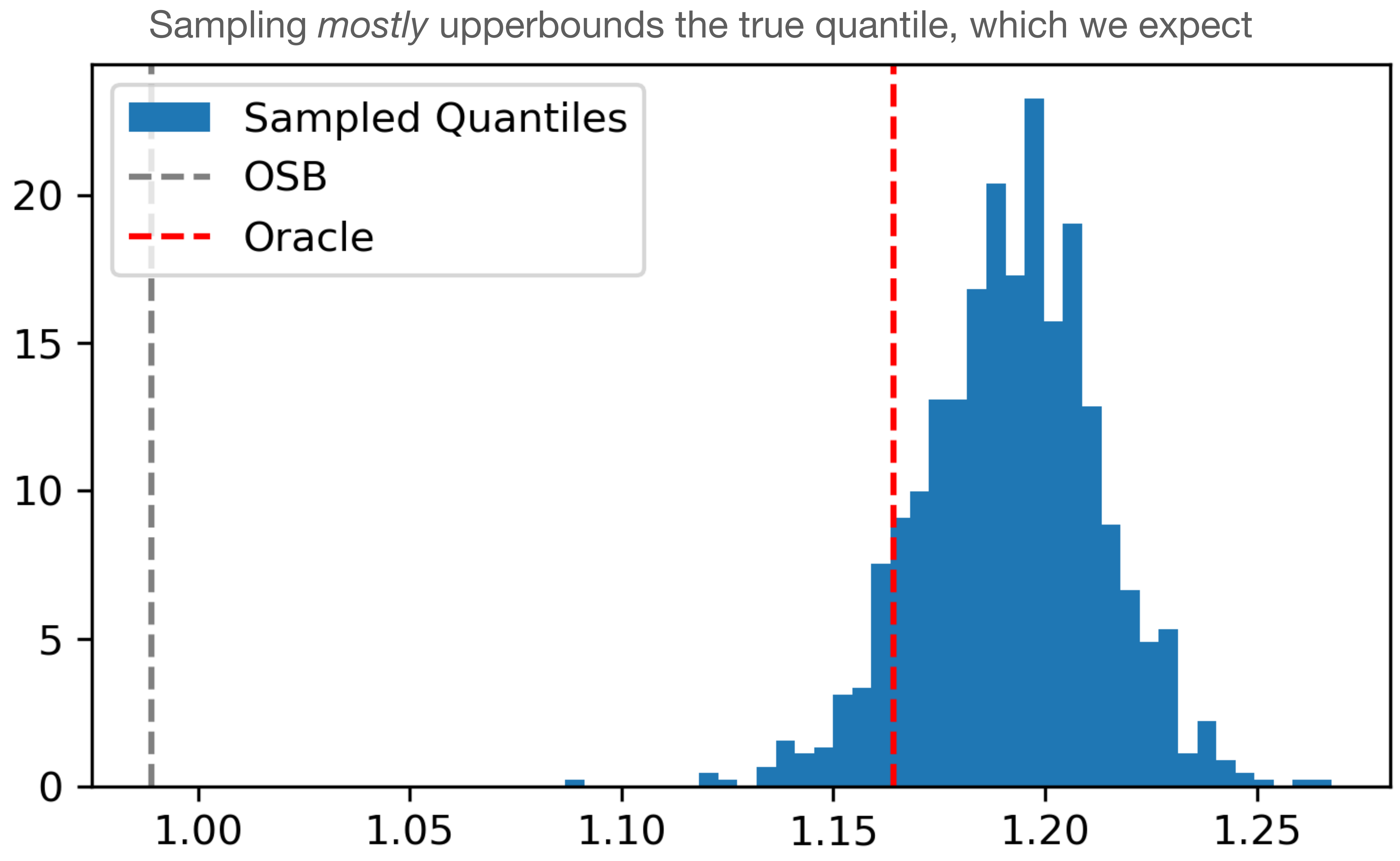
Estimated Interval Coverages

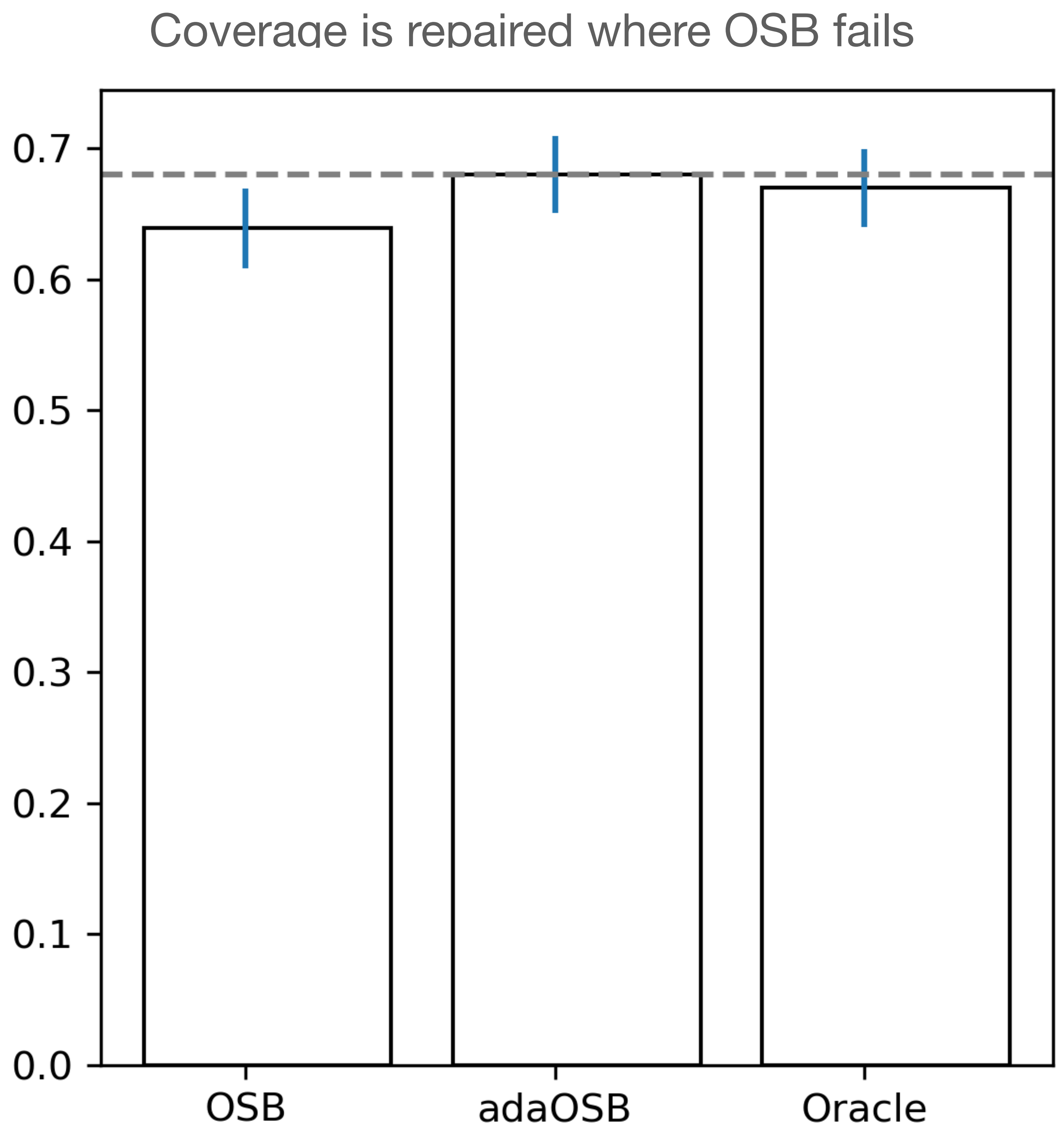


Example 2: Valid Coverage

3d - adaOSB adequately upper bounds true quantile and thus fixes coverage

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I}), \quad \varphi(\boldsymbol{x}) = x_1 + x_2 - x_3, \quad \mathcal{X} = \mathbb{R}^3_+, \quad \boldsymbol{x}^* = (0 \quad 0 \quad 1)^T$$





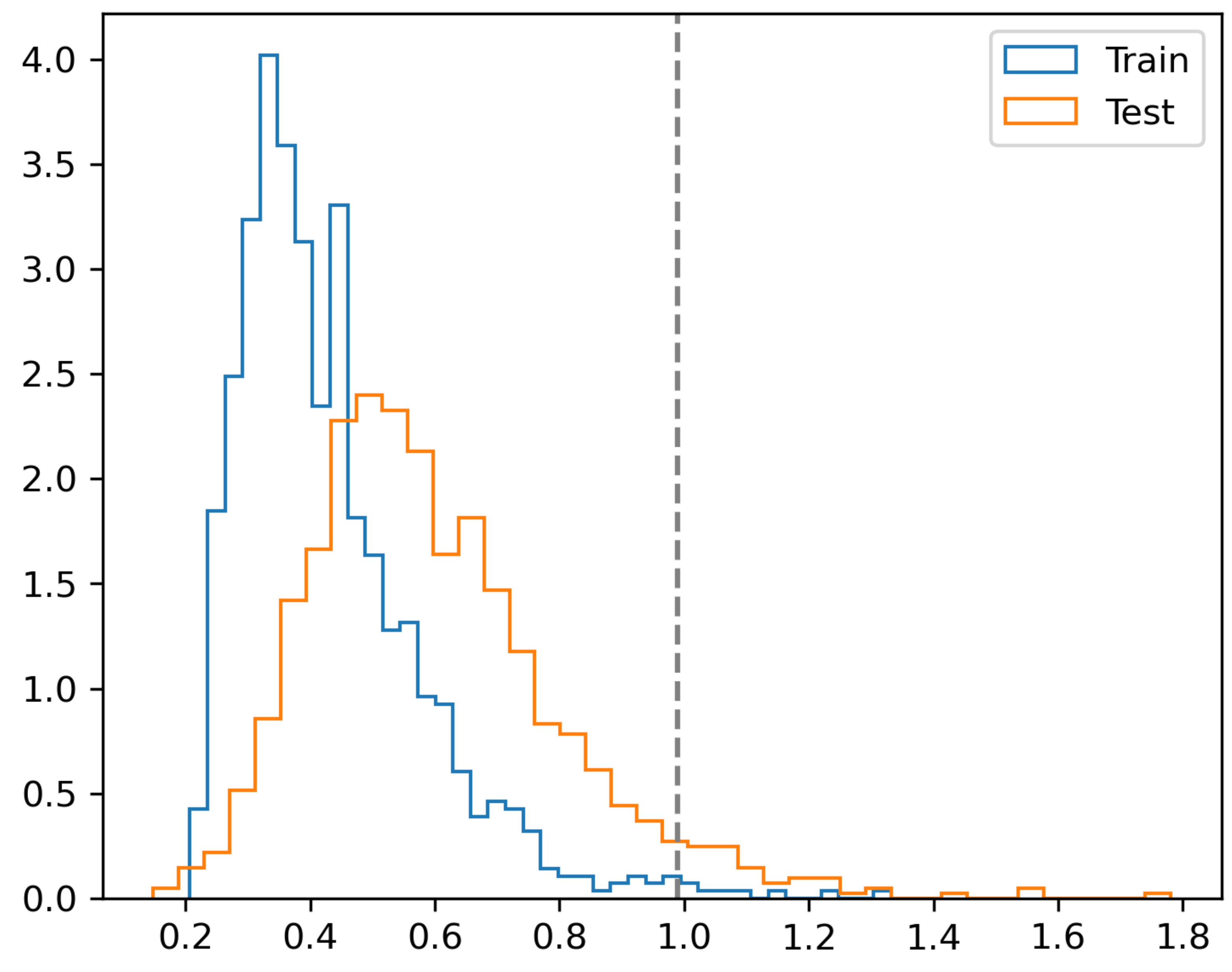
Example 3: Length Improvement

High dimension - Particle unfolding simulation where adaSB shows a dramatic length improvement

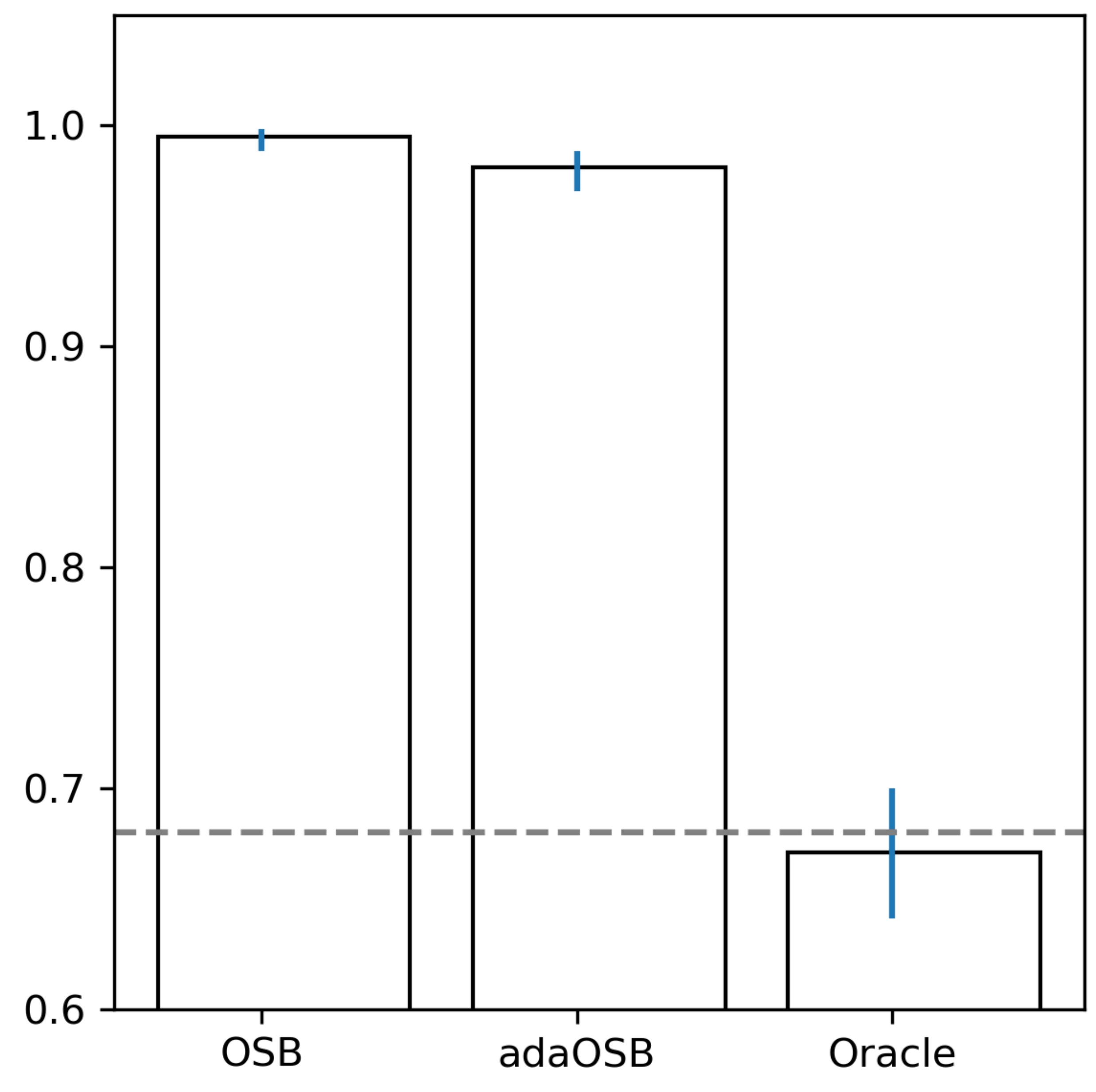
$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \varphi(\boldsymbol{x}) = \boldsymbol{h}^T\boldsymbol{x}, \quad \mathcal{X} = \mathbb{R}^{80}_{+}, \quad \boldsymbol{x}^* \text{ defined mean bin counts}$$

- High dimensions necessitate MCMC polytope sampler and quantile regression

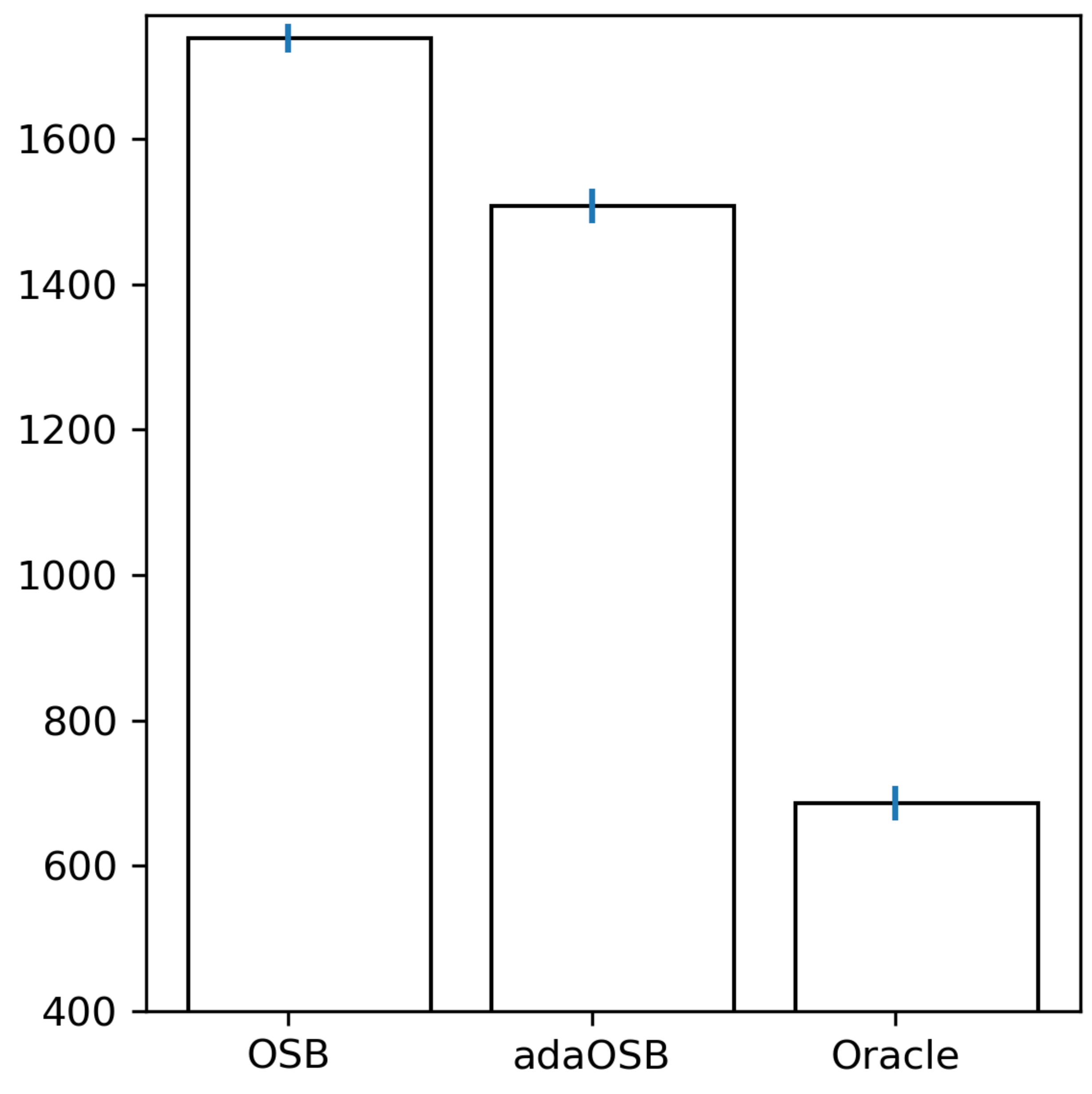
Distributions of MCMC/QR derived $\hat{q}_\gamma$



adaOSB has a **coverage** advantage



adaOSB has a clear **length** advantage



Recap and conclusions

1. Building on the work of [Batlle et al. 2023], we presented a method to set ψ_α^2 in a data-dependent way to achieve interval coverage and improve interval length relative to OSB.

   - **Take-away**: our method is the first computationally feasible approach to properly calibrate these optimization-based intervals.

   - **Key Steps:** using an uncertainty budget to bound the set of feasible parameter values, sampling the pre-image confidence set, estimating the max quantile.

2. We explored three numerical studies to demonstrate the method and its advantages.

   - **Take-away**: our method provides coverage in low dimensional ($p = 3$) example where OSB does not, and improves interval length in a scenario where OSB empirically over-covers ($p = 80$).

Thank You!

Please let me know if you have any follow up questions:

mcstanle@andrew.cmu.edu

References

(1) Batlle, P., Stanley, M., Patil, P., Kuusela, M., Owhadi, H. (2023). Confidence intervals for functionals in constrained inverse problems: An optimization-centric perspective with frequentist guarantees. *Arxiv*.

(2) Roger L. Berger and Dennis D. Boos. (1994). *P* Values Maximized Over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association.*

(3) Chen, Y., Dwivedi, R., Wainwright, M., Yu, B. (2018). Fast MCMC Sampling Algorithms on Polytopes. *Journal of Machine Learning Research*.

(4) Dalmasso, N., Izbiki, R., Lee, A. (2020) Confidence and Hypothesis Testing in a Likeilhood-Free Inference Setting. *International Conference on Machine Learning.*

(5) Dalmasso, N., Masserano, L., Zhao, D., Izbiki, R., Lee, A. (2021). Likelihood-Free Frequentist Inference: Bridging Classical Statistics and Machine Learning For Reliable Simulator-Based Inference. *Arxiv*.

(6) Han, T., Low, M., Wu, T. A. (2023). Quantum Entanglement and Bell Inequality Violation in Semi-Leptonic Top Decays. *Arxiv*.

(7) Kuusela, M. (2016). Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider. *PhD Thesis.*

(8) Masserano, L., Dorigo, T., Izbiki, R., Kuusela, M., Lee, A. (2023). Simulator-based Inference with WALDO: Confidence Regions by Leveraging Prediction Algorithms and Posterior Estimators for Inverse Problems. *International Conference on Artificial Intelligence and Statistics.*

(9) Masserano, L., Shen, A., Doro, M., Dorigo, T., Izbiki, T., Lee, A. (2024). Classification Under Nuisance Parameters and Generalized Label Shift in Likelihood-Free Inference. *Arxiv*.

(10)Meinshausen, N. (2006). Quantile Regression Forests. *Journal of Machine Learning Research*.

(11)Patil, P., Kuusela, M., Hobbs, J. (2022). Objective frequentist uncertainty quantification for atmospheric CO2 retrievals. *SIAM / ASA Journal on Uncertainty Quantification*.

(12)B. W. Rust and W. R. Burrus (1972). Mathematical programming and the numerical solution of linear equations. *American Elsevier Publishing Company*.

(13)B. W. Rust and D. P. O'Leary (1994). Confidence intervals for discrete approximations to ill-posed problems. *Journal of Computational and Graphical Statistics. 3*.

(14)Smith, R.L. (1984). Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed Over Bounded Regions. *Operations Research*.

(15)Stanley, M., Patil, P., Kuusela, M. (2022). Uncertainty quantification for wide-bin unfolding: one-at-a-time strict bounds and prior-optimized confidence intervals. *Journal of Instrumentation*.

(16)Stanley, M., Batlle, P., Kuusela, M., Owhadi, H. (2024a). AdaOSB: adaptive one-at-a-time strict bounds for functionals in constrained inverse problems. *In progress*.

(17)Stanley, M., Kuusela, M., Byrne, B., Liu, J. (2024b). Optimization-based confidence intervals for regional carbon flux inversion uncertainty quantification. *In progress*.

(18)Stark, P. (1992). Inference in Infinite-Dimensional Inverse Problems: Discretization and Duality. *Journal of Geophysical Research.*

(19)Takeuchi, I., Q. V., Sears, T. D., and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning*, 7:1231-1264.

Appendix

Developing the test inversion formalism in this setting provides a new perspective

Key set definitions $\mathcal{X} \subset \mathbb{R}^p \quad \Phi_\mu := \{\boldsymbol{x} : \varphi(\boldsymbol{x}) = \mu\} \subset \mathbb{R}^p$

Fundamental HT $H_0 : \boldsymbol{x}^* \in \Phi_\mu \cap \mathcal{X}$ versus $H_1 : \boldsymbol{x}^* \in \mathcal{X} \setminus \Phi_\mu$

Test Statistic (LLR)
$$\lambda(\mu, \boldsymbol{y}) := -2 \log \Lambda(\mu, \boldsymbol{y}) = -2 \left(\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \ell_{\boldsymbol{x}}(\boldsymbol{y}) - \sup_{\boldsymbol{x} \in \mathcal{X}} \ell_{\boldsymbol{x}}(\boldsymbol{y}) \right)$$
$$= \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y})$$

Level α test
$$\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \mathbb{P}_{\lambda \sim F_{\boldsymbol{x}}} \left(\lambda > q_\alpha\right) \leq \alpha$$

Test T_μ is a *level-α* test

Let $Q_{\boldsymbol{x}} : [0,1] \rightarrow \mathbb{R}$ be the quantile function of $\lambda(\mu, \boldsymbol{y})$ at $\boldsymbol{x}$. Using $Q_{\boldsymbol{x}}(1-\alpha)$ produces a *level*-α test.

Ellipsoid Sampler

Uniform sampling in p-ball + Accept/reject

- [Voelker et al., 2017] presented and proved an interesting and efficient algorithm to sample uniformly at random from the p-ball. First, sample uniformly from the $(p+1)$-sphere (possible with Gaussian RNG) followed by dropping any two coordinates.
  - We refer to a sample drawn from the p-ball via "Voelker-Gosmann-Stewart" (VGS) by $\boldsymbol{x} \sim VGS(p)$
- Consider an ellipsoid defined by $\mathscr{E}(r) := \{\boldsymbol{x} : \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \leq r\}$ and let $\boldsymbol{P}\boldsymbol{\Omega}^2\boldsymbol{P}^T$ be the eigendecomposition of PSD $\boldsymbol{A}$.
- If $\boldsymbol{x} \sim VGS(p)$, then $\boldsymbol{y} := \sqrt{\chi^2_{n,\eta}}\boldsymbol{P}\boldsymbol{\Omega}\boldsymbol{x}$ is sampled uniformly at random from $\mathscr{E}(\chi^2_{n,\eta})$
- To incorporate constraints, simple reject $\boldsymbol{y}$ if $\boldsymbol{y} \notin \mathscr{X}$
- NOTE: this approach works well in low dimensions and when $f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$, where $\boldsymbol{K}$ is full column rank.

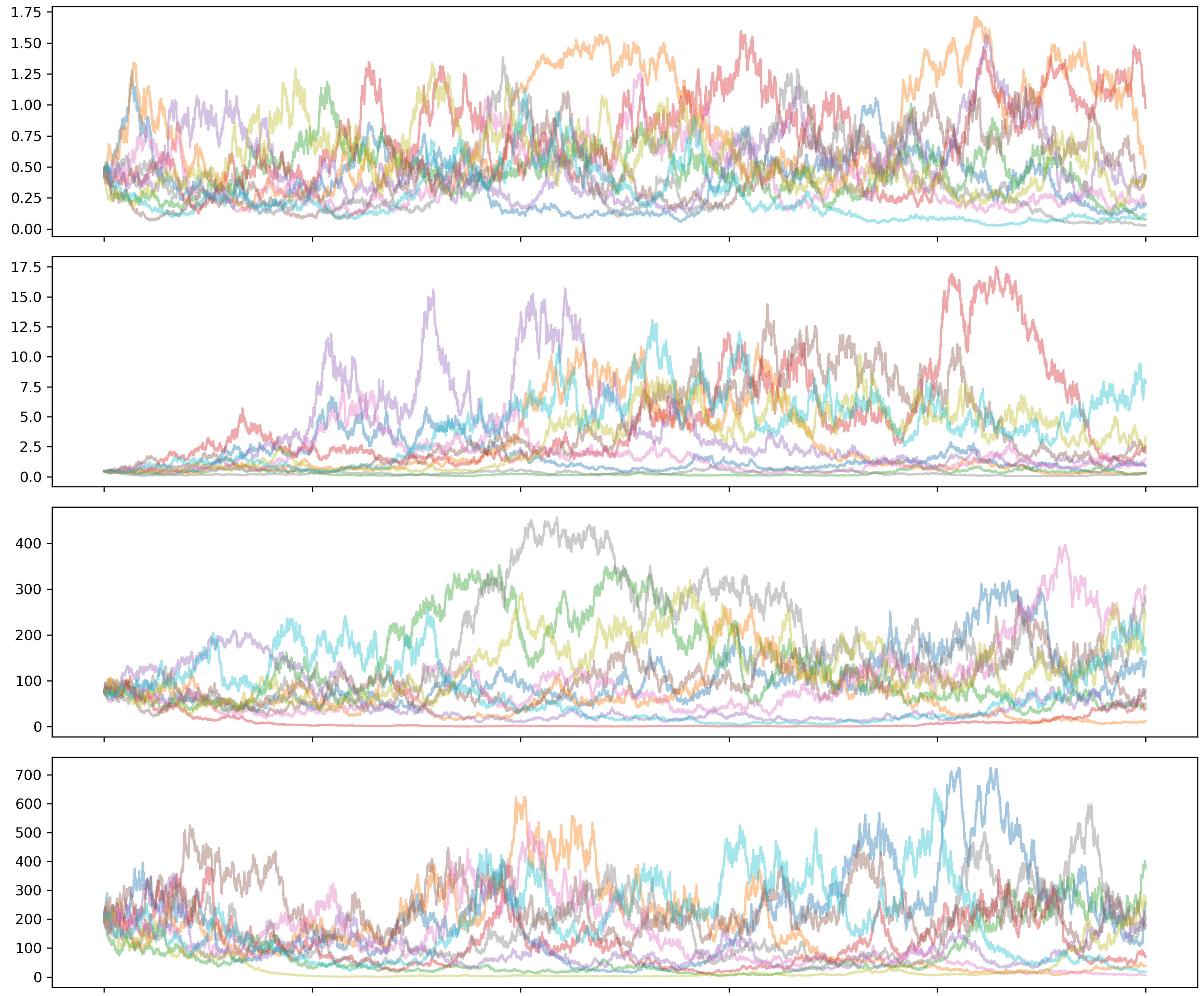
MCMC Polytope Sampler

Implementation details and considerations

- We construct a bounding polytope for $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$ using the principal axes of the confidence set ellipsoid ($2p$), the hyper-rectangle defined by the non-negativity constraints ($2p$) and 200 additional randomly chosen hyperplanes.
- We use the Vaidya sampler detailed in [Chen et al., 2018], where the uniform distribution over the defined polytope is the Markov chain's stationary distribution
- Since this sampling is an MCMC algorithm, we consider a few different convergence plots to assess sufficient mixing:
    - Trace plots of individual parameters
    - Ensembles of max predicted quantiles for both **fixed data set** size and **cumulative**
        - **Fixed** allows for us to get a sense of the Markov chain convergence
        - **Cumulative** allows us to assess the stability the max predicted quantile

MCMC Polytope Sampler (con't)

Parameter Trace Plots

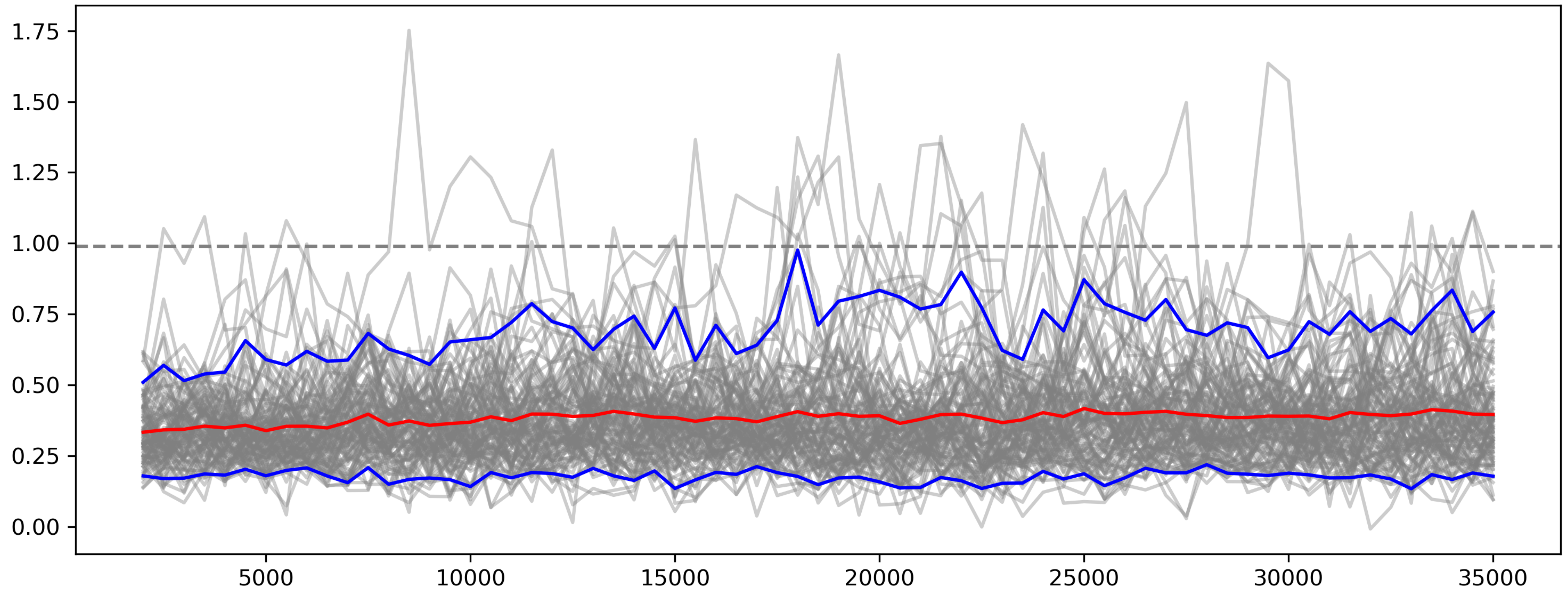


Four arbitrarily chosen parameter trace plots show nice mixing

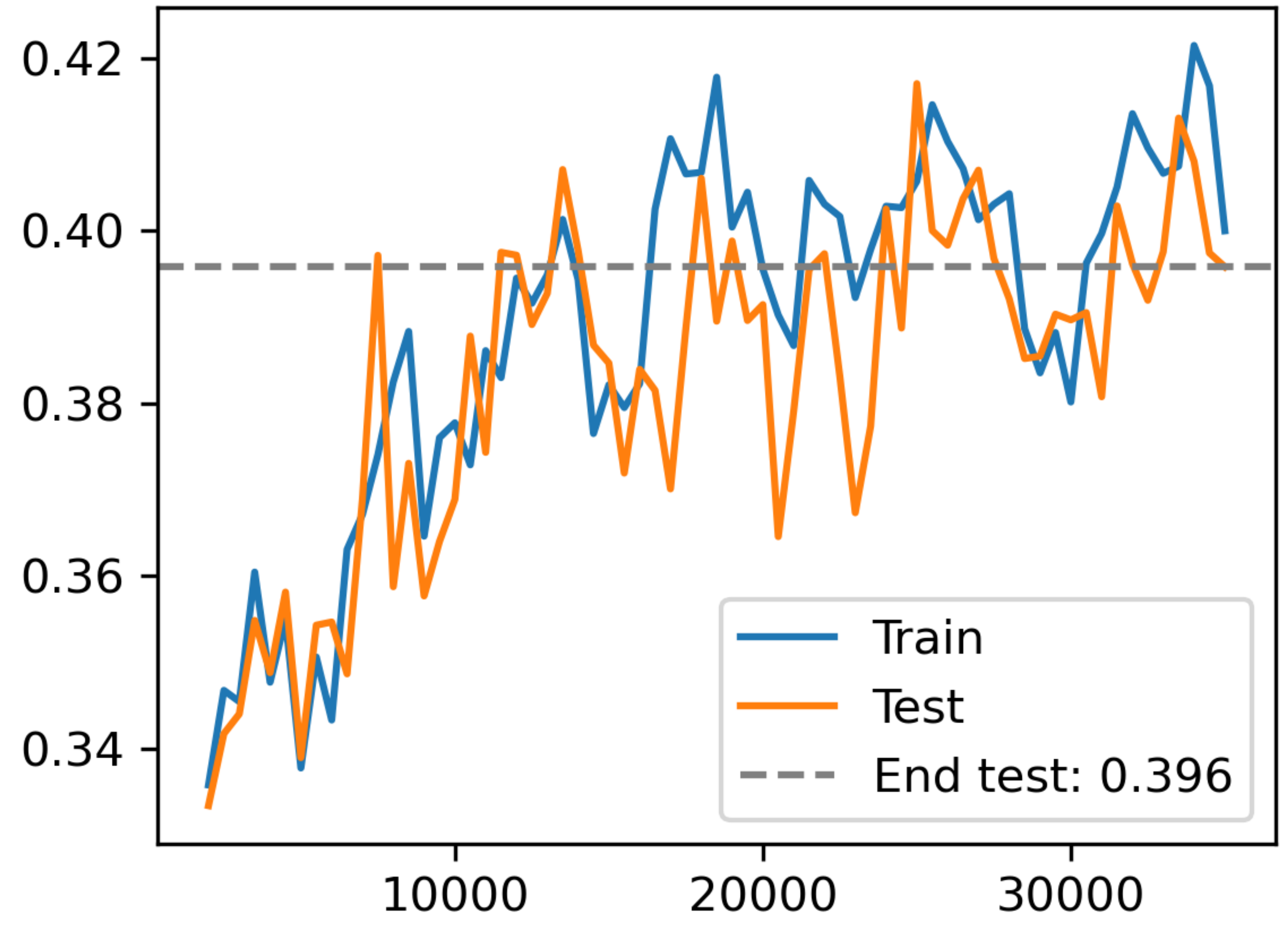
MCMC Polytope Sampler (con’t)

Fixed Max-q trace plots

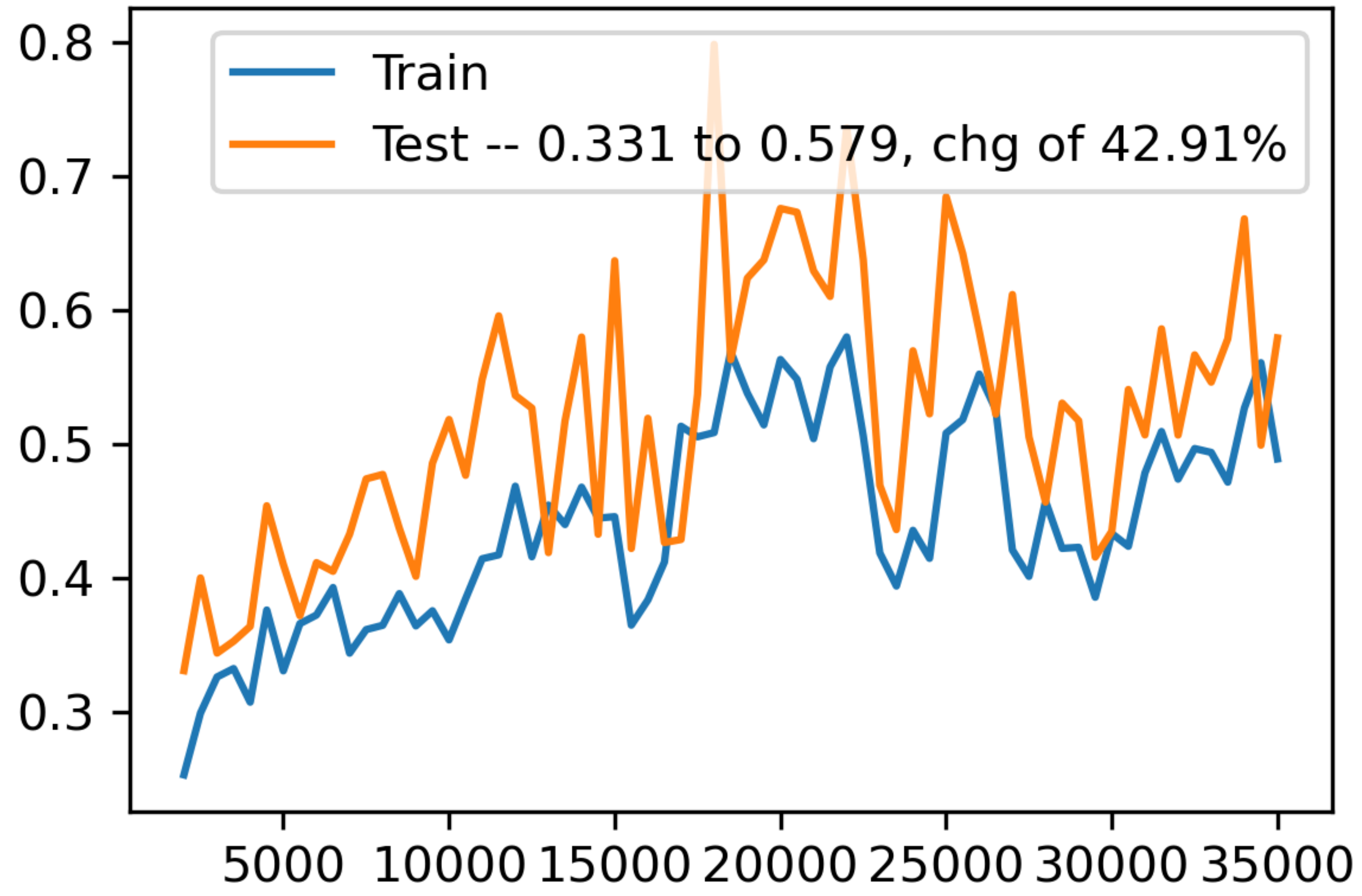
The ensemble of Max predicted quantiles with dataset size fixed at 2k



Ensemble mean stabilizes after ~15k iterations



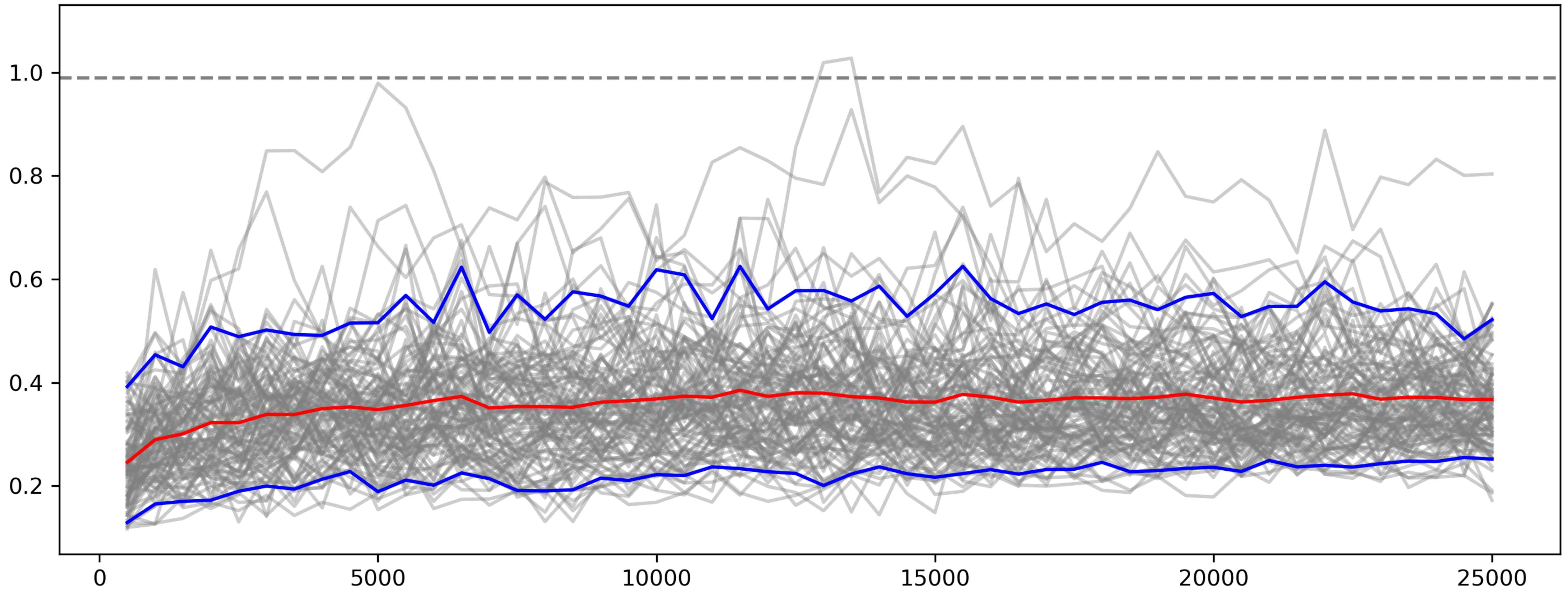
Ensemble width stabilizes after ~15k iterations



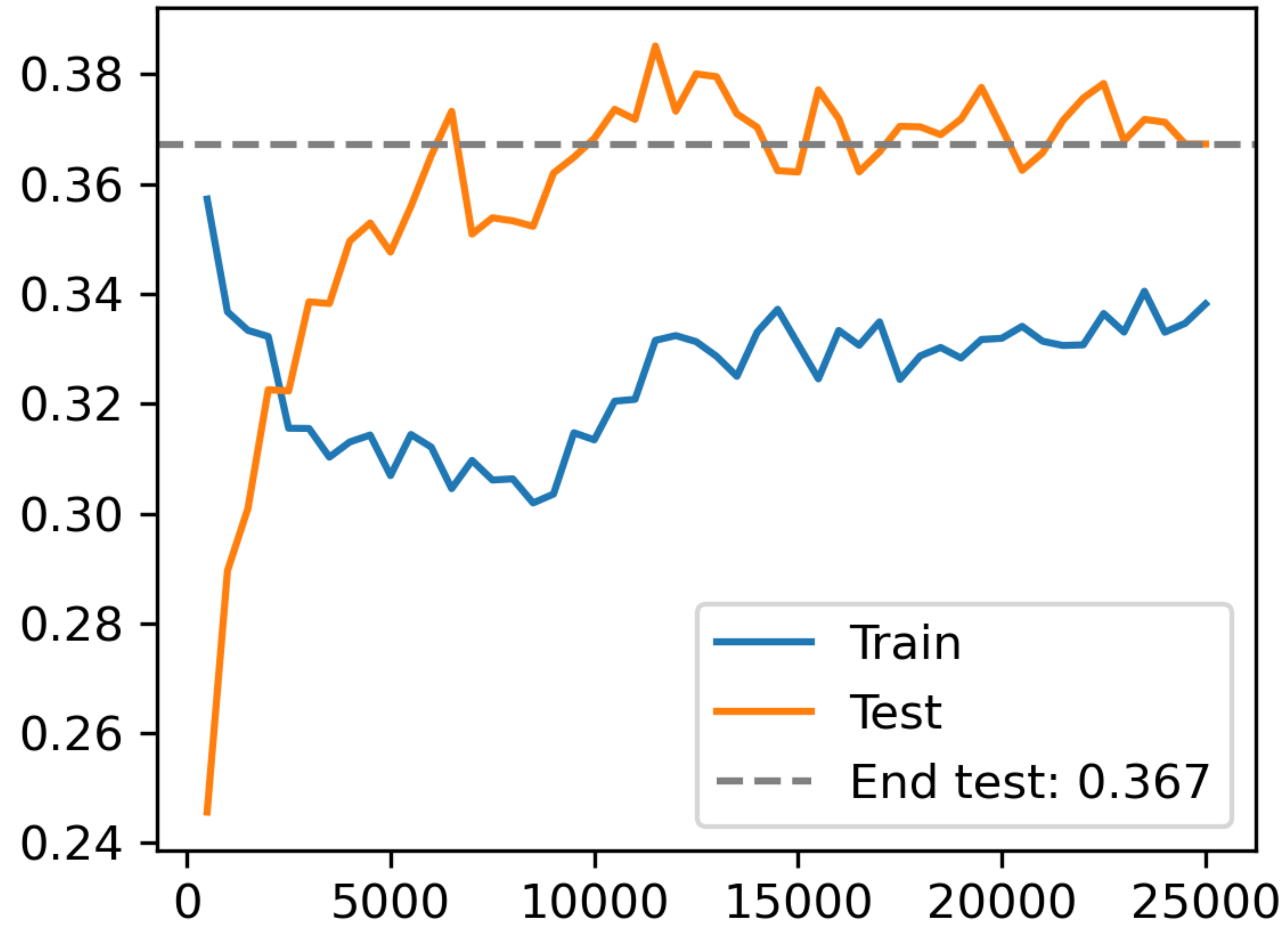
MCMC Polytope Sampler (con’t)

Cumulative Max-q trace plots

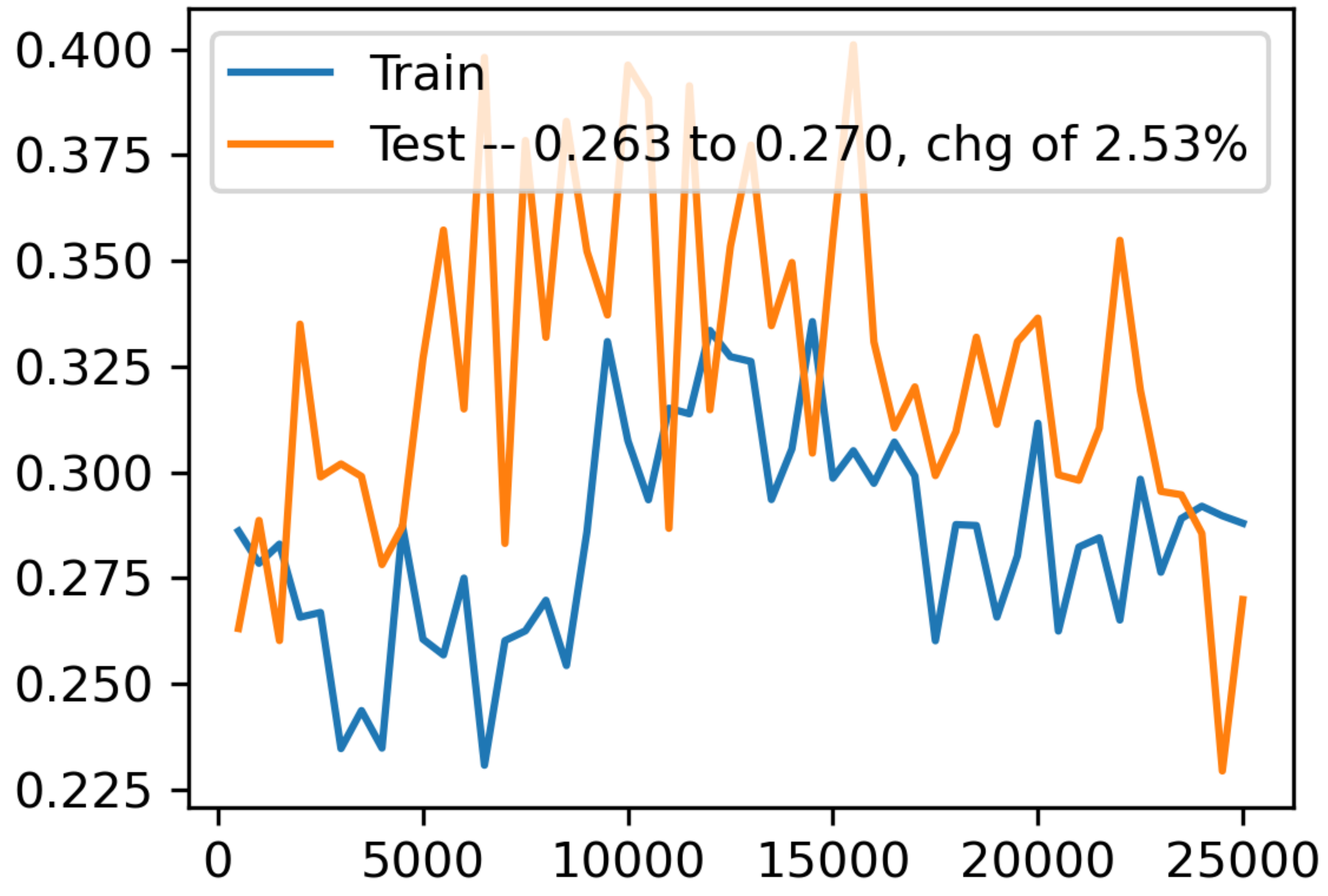
The ensemble of Max predicted quantiles with dataset size fixed at 2k



Ensemble mean stabilizes after ~10k iterations

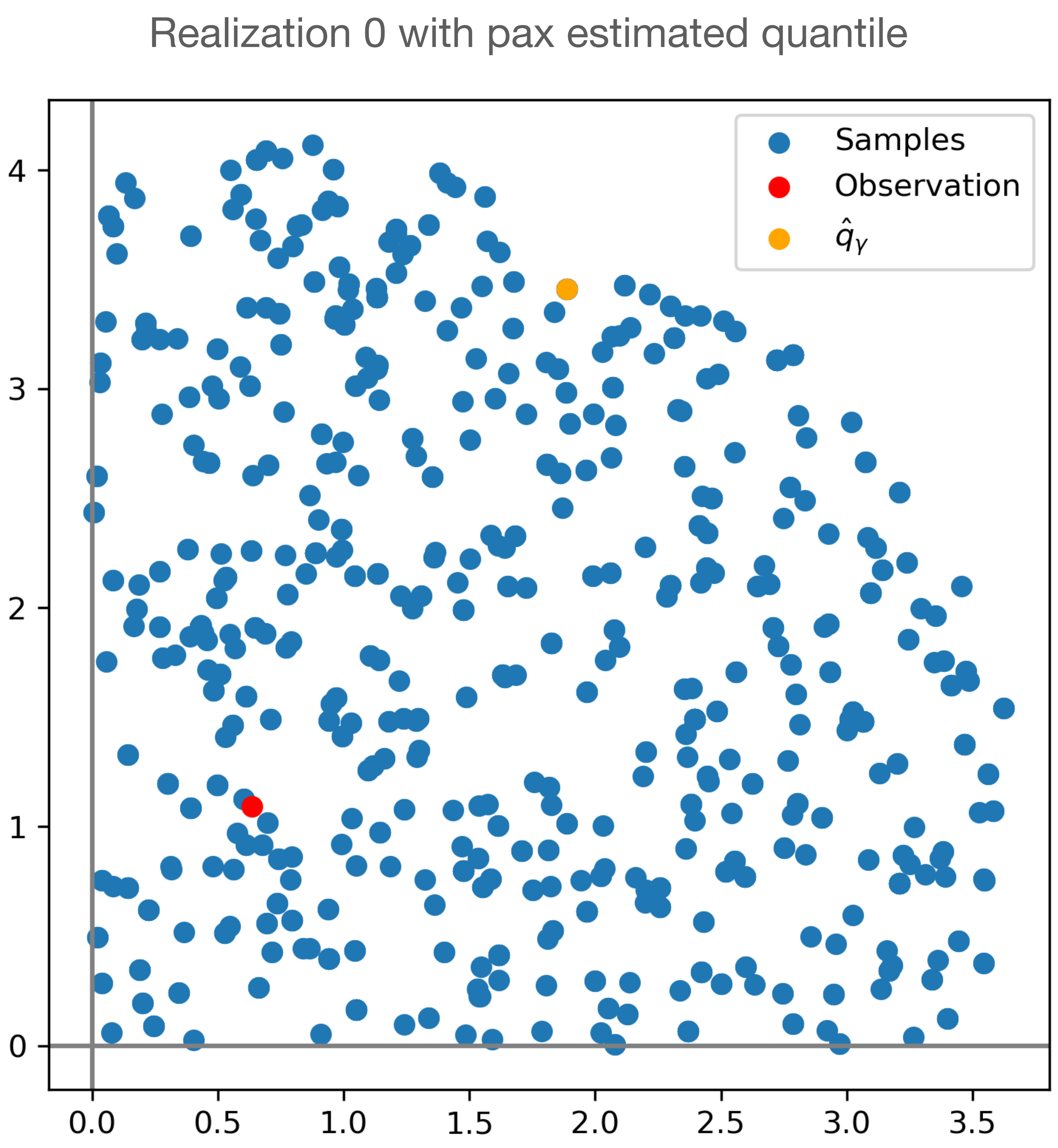


Ensemble width decreases



2d Exposition example

Additional figures and details



Monte Carlo Sampling to estimate $Q_{\boldsymbol{x}}(1-\gamma)$

1. Generate an ensemble of samples, $\boldsymbol{y}_i = \boldsymbol{x} + \boldsymbol{\varepsilon}_i$ and therefore LLR samples, $\lambda(\boldsymbol{h}^T\boldsymbol{x}, \boldsymbol{y}_i)$.
2. From our generated ensemble, we can simply use the $(1-\gamma)$ percentile estimator.

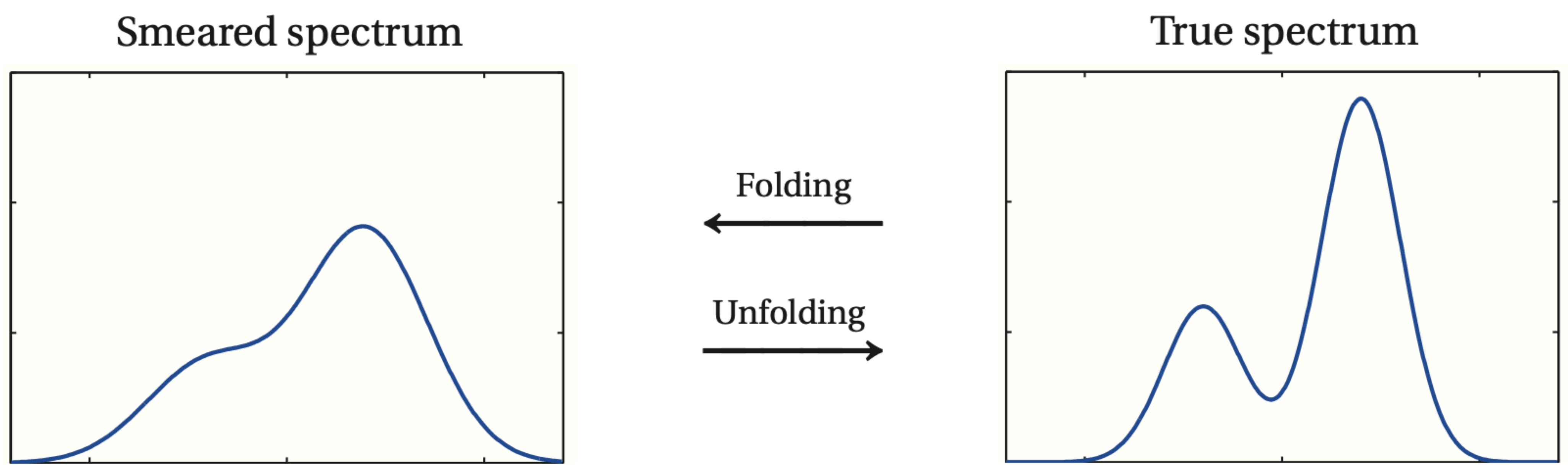
More on particle unfolding

The data generating process for our histogram is

$$\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{K\lambda}),$$

which we approximate by

$$\boldsymbol{y} \sim N(\boldsymbol{K\lambda}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma}_{ii} = (\boldsymbol{K\lambda})_i, \forall i.$$



For more information, see [Kuusela, 2016] and [Stanley et al., 2022]