Accounting for systematic uncertainties in unfolding uncertainty quantification

Mike Stanley | Systematic in Particle Physics Data Analysis | April 27, 2023
with Mikael Kuusela (CMU) and Pratik Patil (UC Berkeley) [Stanley, Kuusela, and Patil, 2022]

Carnegie Mellon University

Talk overview

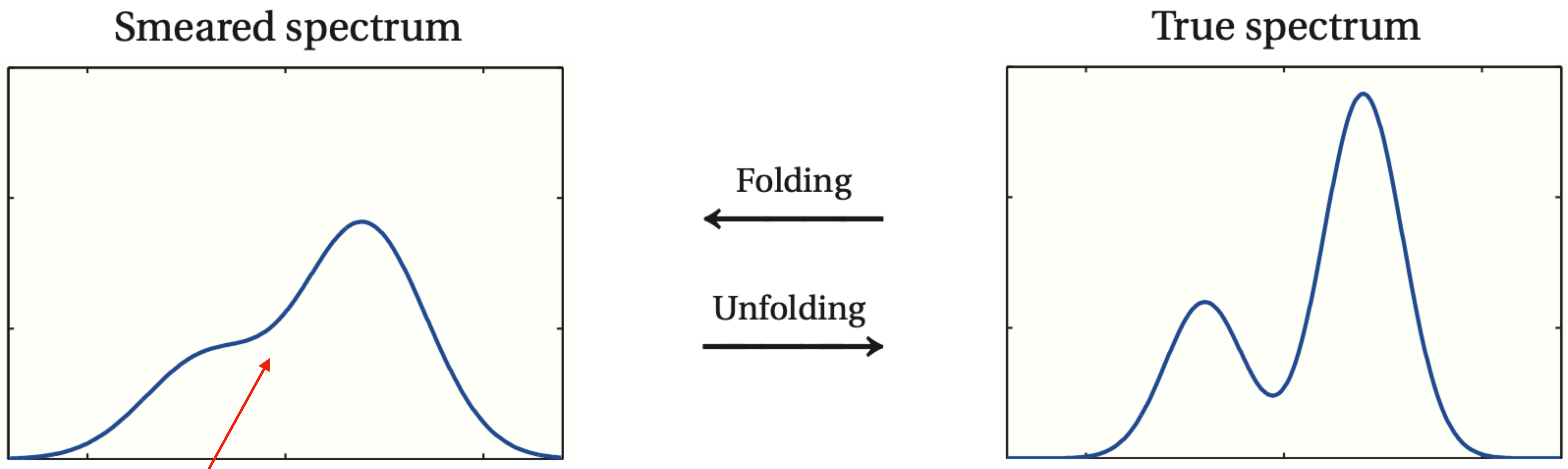
1. Provide a brief mathematical overview of the unfolding problem and our uncertainty quantification (UQ) objective,

2. characterize how two types of systematic uncertainties, regularization bias and wide-bin bias, affect our UQ objective, and

3. present a framework and methods that can address these challenges

The unfolding problem and density deconvolution

Our goal is to estimate a true (unknown) probability distribution for some variable of interest (e.g., energy) via a finite-resolution detector observations



Kuusela, 2016.

There are many true spectra that could result in this same smeared spectrum, making the problem ill-posed.

Mathematical formulation

Let f be the true, particle-level spectrum and g the smeared, detector-level spectrum (both intensity functions for the underlying Poisson point process).

Let $T \subset \mathbb{R}$ be the true space and $S \subset \mathbb{R}$ the smeared space

$$g(s) = \int_T k(s,t) f(t)\, dt$$

$k(s,t) = p(Y = s \mid X = t, X \text{ obs}) P(X \text{ obs} \mid X = t)$, X true event and Y smeared.

Goal: infer the true spectrum f given observations from g

We discretize into a histogram with uniform bins

Create binnings for T and S: $\{T_j\}_{j=1}^{n}$ and $\{S_i\}_{i=1}^{m}$

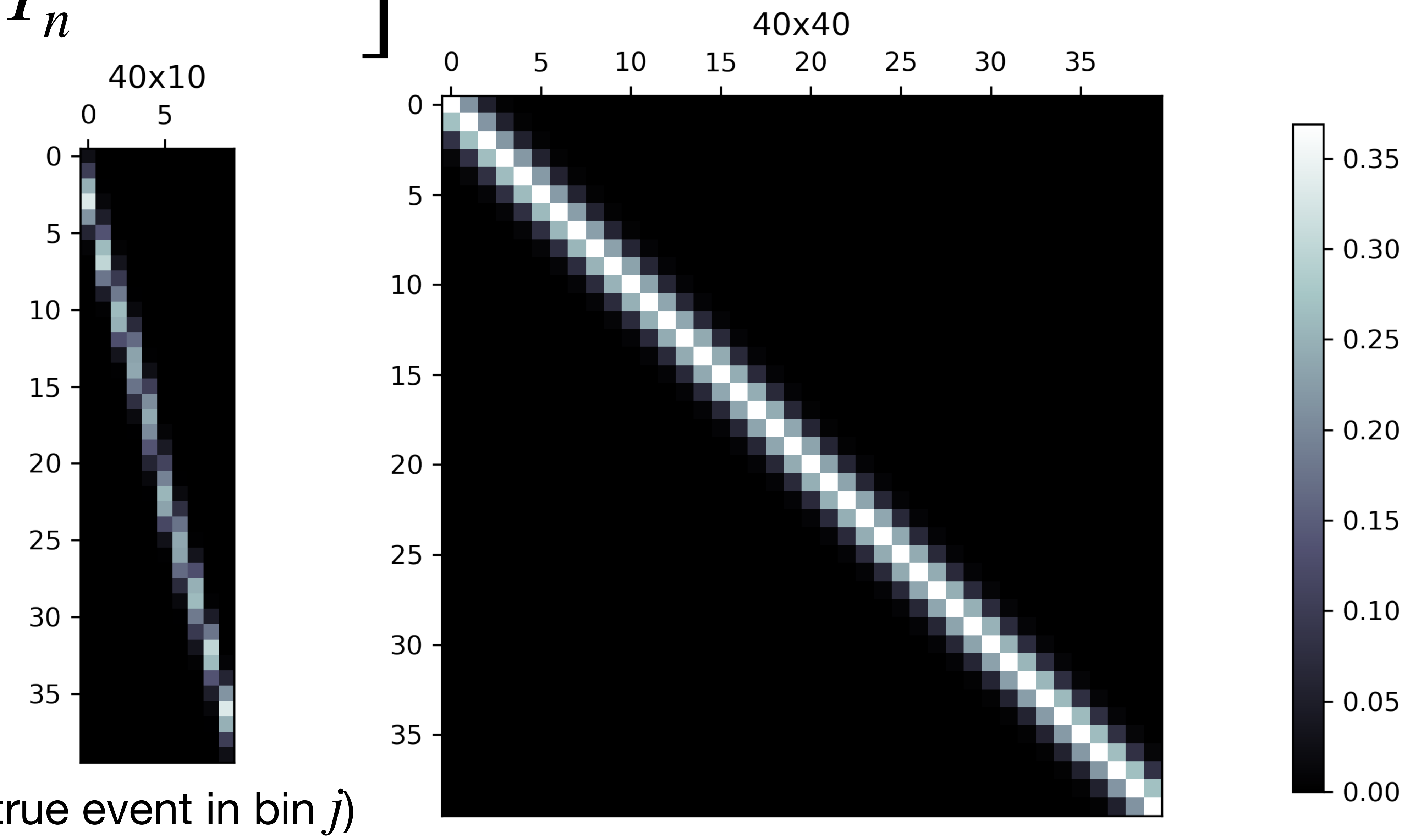
<u>The data</u>: $\boldsymbol{y} \in \mathbb{R}^m$ with $\mathbb{E}[\boldsymbol{y}] = \boldsymbol{\mu} = \left[\int_{S_1} g(s)\, ds \quad \dots \quad \int_{S_m} g(s)\, ds \right]^T$

<u>Parameters of interest</u>: $\boldsymbol{\lambda} = \left[\int_{T_1} f(t)\, dt \quad \dots \quad \int_{T_n} f(t)\, dt \right]^T$

Bin means are related via $\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\lambda}$ where,

$$\boldsymbol{K}_{ij} = \frac{\int_{S_i} \int_{T_j} k(s,t) f(t)\, dt\, ds}{\int_{T_j} f(t)\, dt}.$$

Can interpret as
Prob(smeared event is in bin i | true event in bin j)



Our discretized histogram model is approximately a linear-Gaussian model

The data generating process for our histogram is

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

which we approximate by

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

Our UQ goal is to compute confidence intervals (CIs) for particle-level bins

More precisely, we want to compute CIs for $\theta(\boldsymbol{\lambda}) = \boldsymbol{h}^T \boldsymbol{\lambda}$

⟶ For example, aggregated bins ($\theta(\boldsymbol{\lambda}) = \sum_{i=4k}^{4k+3} \boldsymbol{\lambda}_i$) or a single bin ($\theta(\boldsymbol{\lambda}) = \boldsymbol{\lambda}_k$)

Our UQ goal is to find a random interval with a *coverage guarantee*. I.e., for any $\alpha \in (0,1)$, an interval $I_\alpha(\boldsymbol{y}) = \left[\theta_l(\boldsymbol{y}), \theta_u(\boldsymbol{y})\right]$ such that

$$\mathbb{P}\left(\theta(\boldsymbol{\lambda}) \in I_\alpha(\boldsymbol{y})\right) \geq 1 - \alpha.$$

Four sources of systematic uncertainty in unfolding

1. Regularization bias
2. Wide-bin bias
3. Missing auxiliary variables
4. Uncertainty in the response kernel k

Covered in this talk (1–2)

Covered in Richard's talk (4)

A Monte Carlo ansatz is a source of systematic uncertainty

1. Computing $\boldsymbol{K}$ requires knowing f…

2. We do not know f, but we can use a Monte Carlo ansatz, f^{MC}, to approximate

Affects UQ with regularization

$$\boldsymbol{\lambda}^{MC} = \begin{bmatrix} \int_{T_1} f^{MC}(t)\, dt & \dots & \int_{T_1} f^{MC}(t)\, dt \end{bmatrix}^T$$

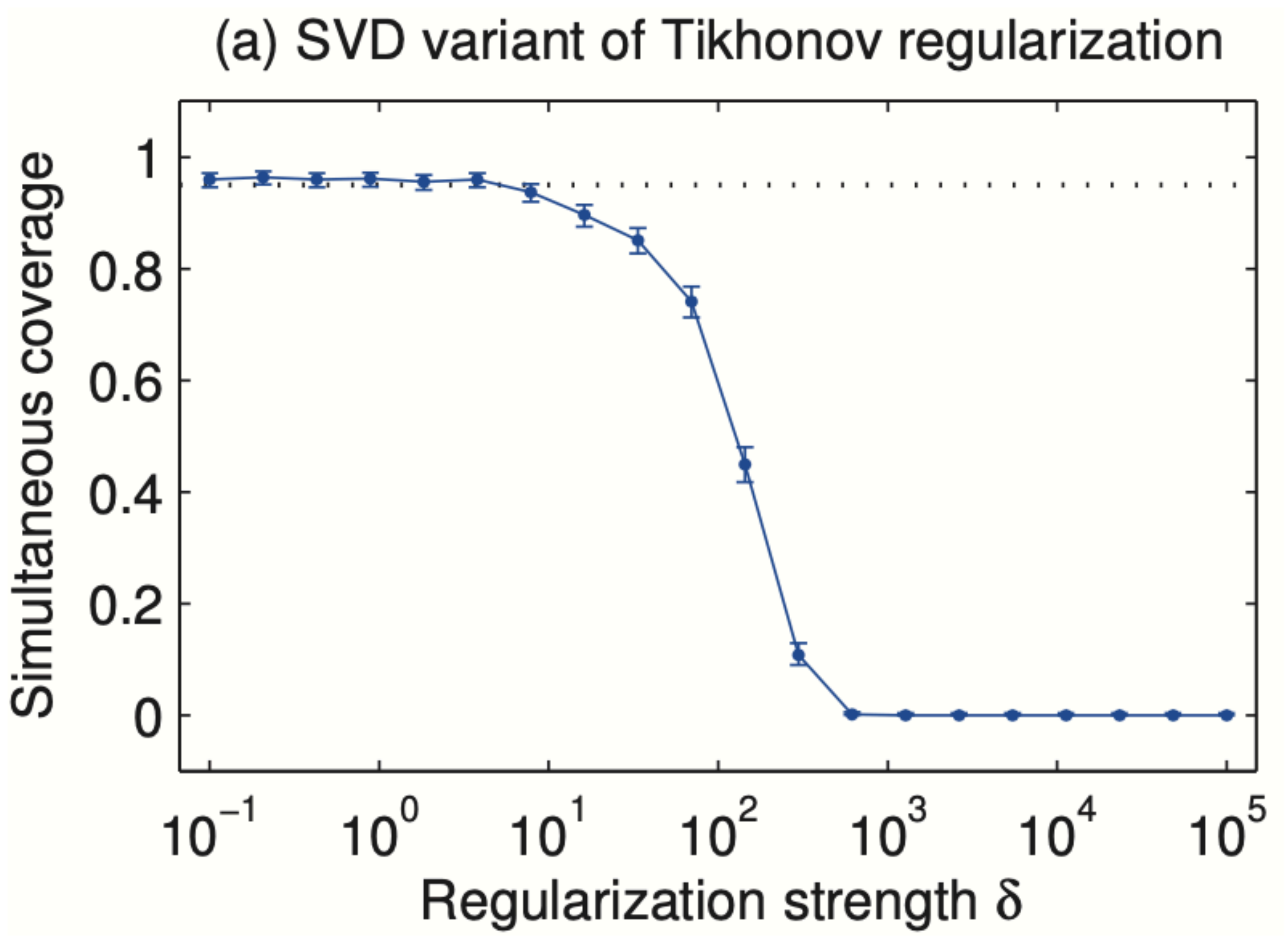
Affects UQ with wide-bins

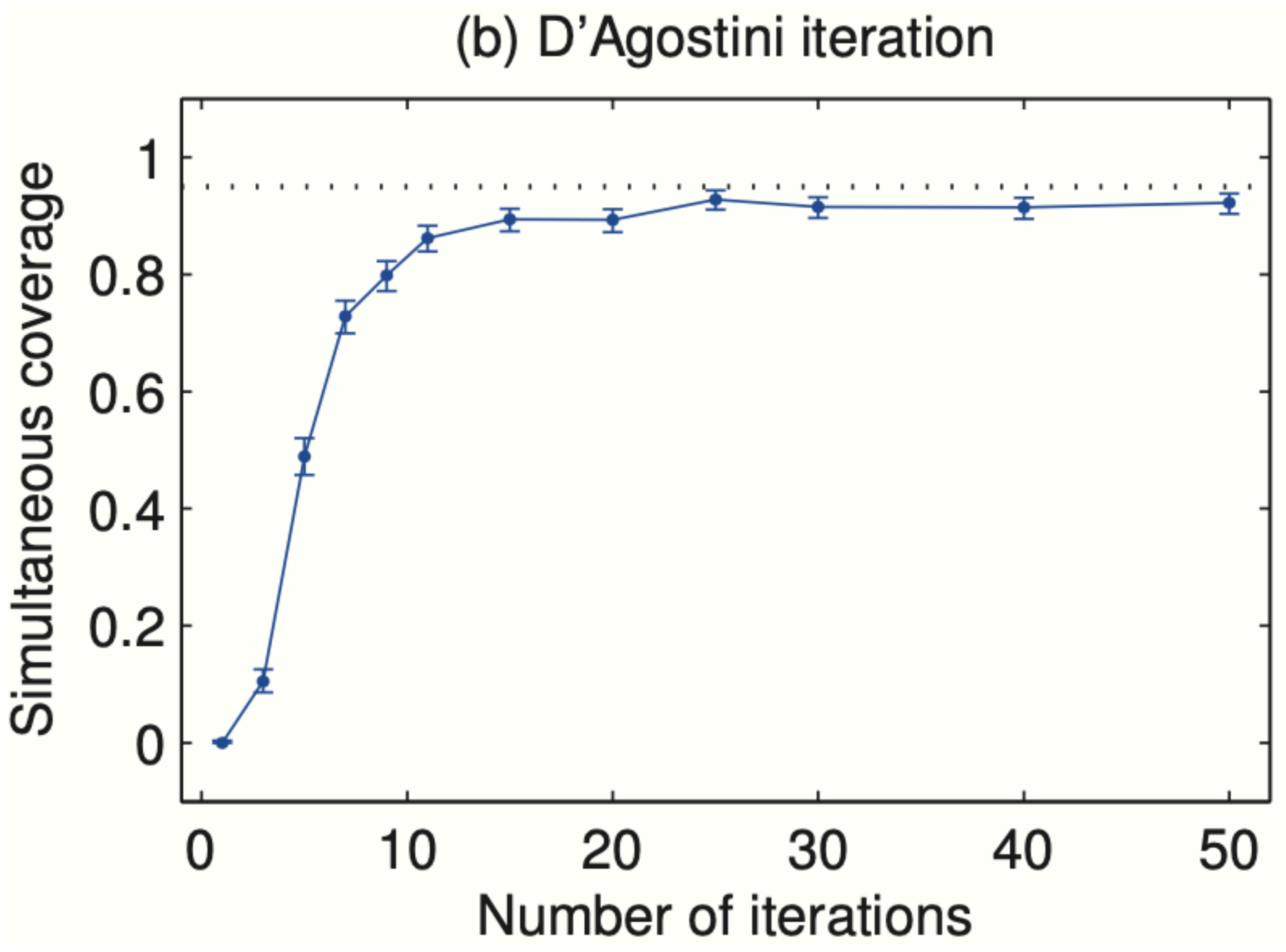
$$\boldsymbol{K}_{ij} = \frac{\int_{S_i} \int_{T_j} k(s,t) f(t)\, dt\, ds}{\int_{T_j} f(t)\, dt} \approx \boldsymbol{K}_{ij}^{MC} = \frac{\int_{S_i} \int_{T_j} k(s,t) f^{MC}(t)\, dt\, ds}{\int_{T_j} f^{MC}(t)\, dt}$$

Systematic 1 - Regularization bias

1. When the number of true bins (n) is large, the smearing matrix $\boldsymbol{K}$ is severely ill-conditioned -> the LS estimator, $\hat{\lambda}_{LS} = \text{argmin}_{\boldsymbol{\lambda}} \|\boldsymbol{y} - \boldsymbol{K}^{MC}\boldsymbol{\lambda}\|^2_{\boldsymbol{\Sigma}^{-1}}$ is very sensitive to noise
2. One solution to this sensitivity is to regularize using
    1. Tikhonov regularization - SVD [Höcker and Kartvelishvili, 1996] or TUnfold [Schmitt, 2012]
    2. EM Iterations with early stopping [D'Agostini, 1995]
3. Both approaches bias the solution towards a Monte Carlo prediction $\boldsymbol{\lambda}^{MC}$ of the true histogram bin mean, $\boldsymbol{\lambda}$
4. Extensively discussed in [Kuusela, 2016]

Regularization bias impacts coverage





[Kuusela, 2016]

<u>Take-away</u>: non-zero bias means coverage is not $1 - \alpha$. Can we not regularize?

Systematic 2 - Wide-bin bias

1. An alternative to *explicit* regularization is to *implicit* regularization by using fewer wider bins $\{T_j\}_{j=1}^{n}$
2. Intuitively, we should choose the bin width to be of the order of the detector resolution (as opposed to the fine-bin explicit regularization approach)
3. However, using wide-bins exposes us to the Monte Carlo misspecification of $\boldsymbol{K}^{MC}$, resulting in intervals that under-cover

We simulate using a GMM

Allows empirical study of interval coverage

True Intensity: $f(t) = \lambda_{tot}\left(\pi_1 \mathcal{N}(t; \mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(t; \mu_2, \sigma_2^2)\right)$

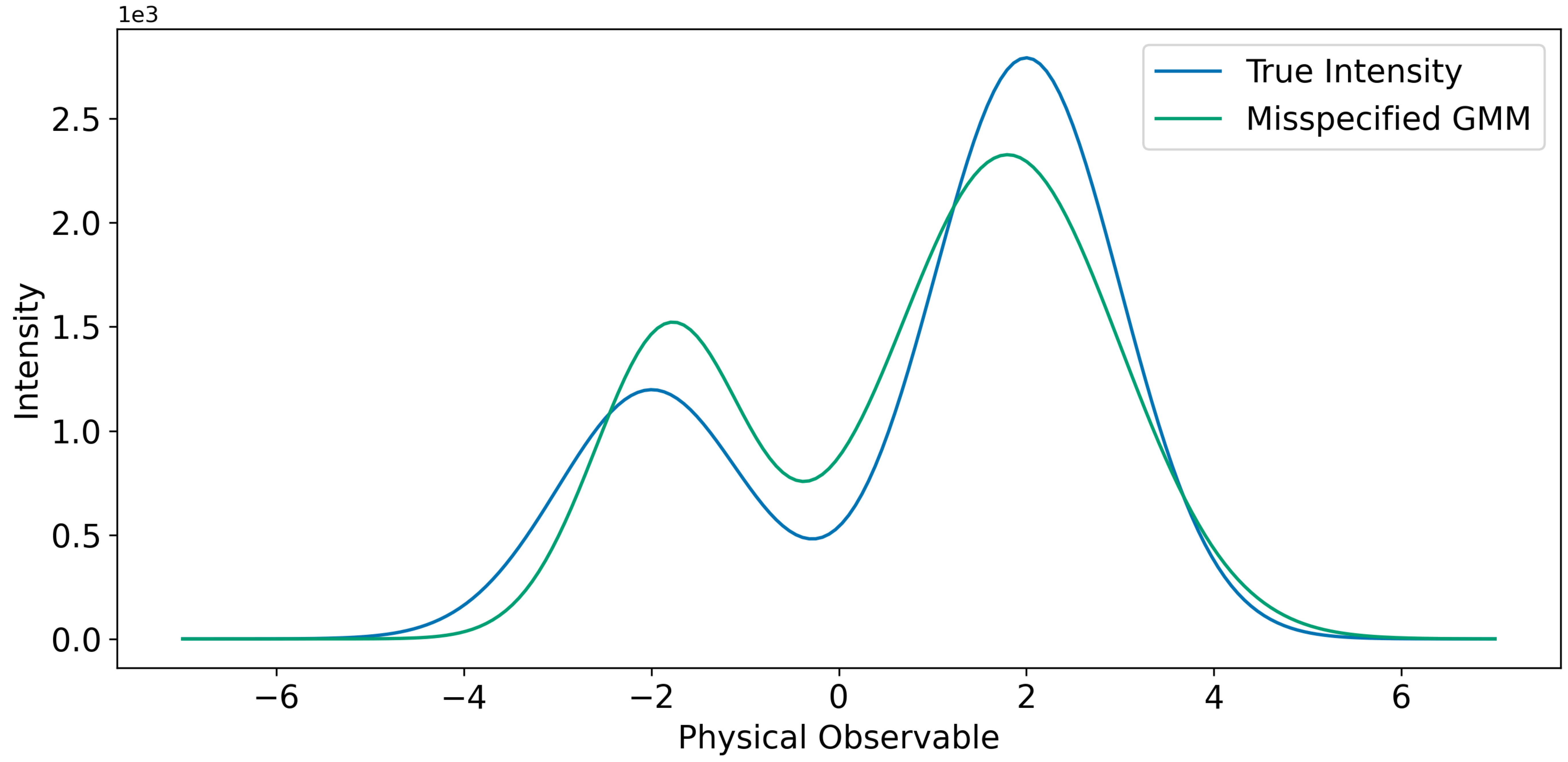
MC Intensity: $f^{MC}(t) = \lambda_{tot}\left(\pi_1 \mathcal{N}(t; \tilde{\mu}_1, \tilde{\sigma}_1^2) + \pi_2 \mathcal{N}(t; \tilde{\mu}_2, \tilde{\sigma}_2^2)\right)$

$(\pi_1, \pi_2) = (0.3, 0.7)$

$(\mu_1, \mu_2) = (-2, 2)$

$(\sigma_1^2, \sigma_2^2) = (1, 1)$

$\lambda_{tot} = 10^4$



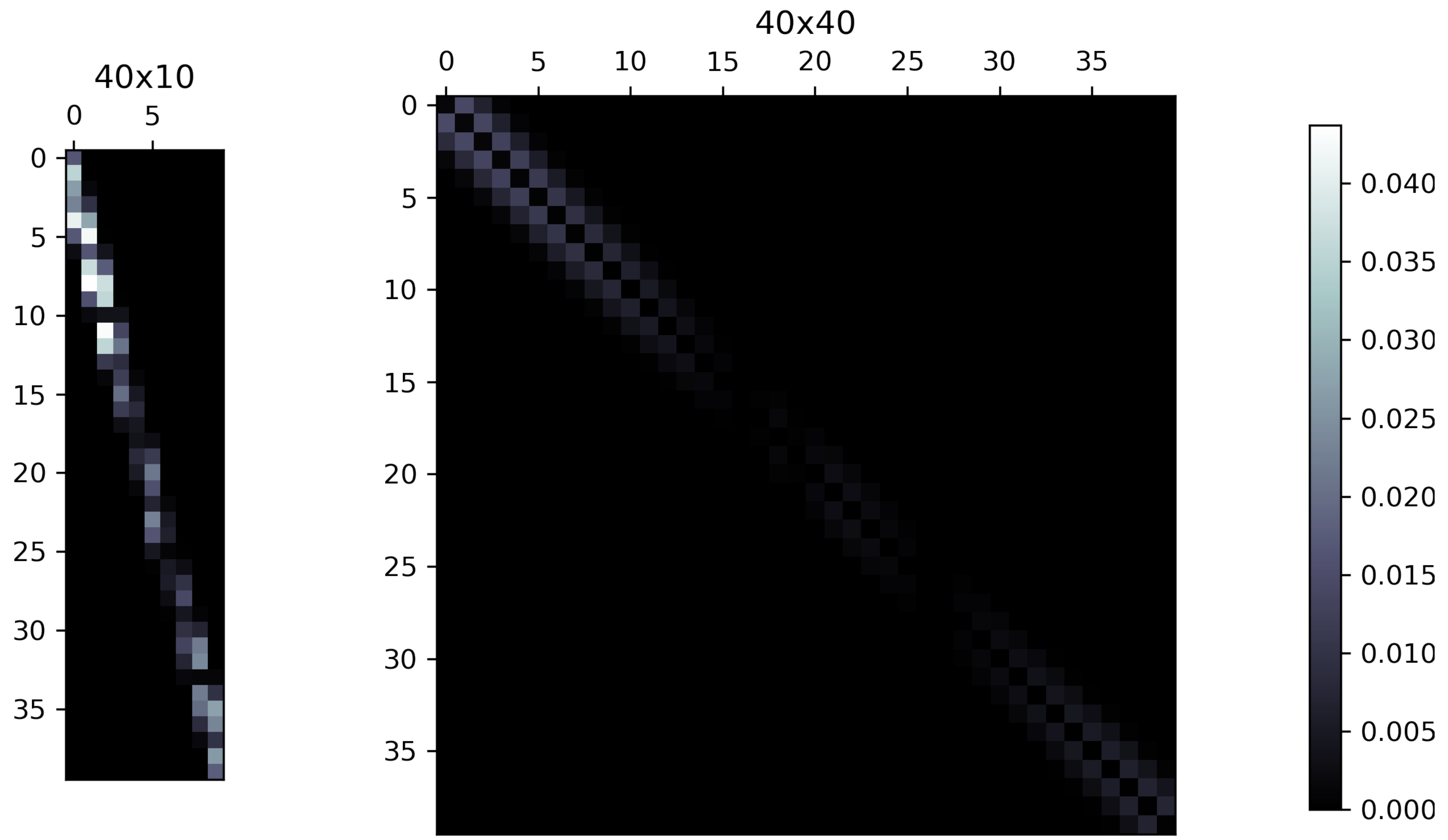
Addressing wide-bin bias using fine-bins

<u>High-level Idea</u>: reduce the dependence of $\boldsymbol{K}^{MC}$ on f^{MC} by unfolding with a higher number of fine bins followed by aggregating bins to wide-bin granularity

<u>General Recipe</u> [Stanley, Kuusela, Patil, 2022]:

1. Unfold with fine bins and no regularization

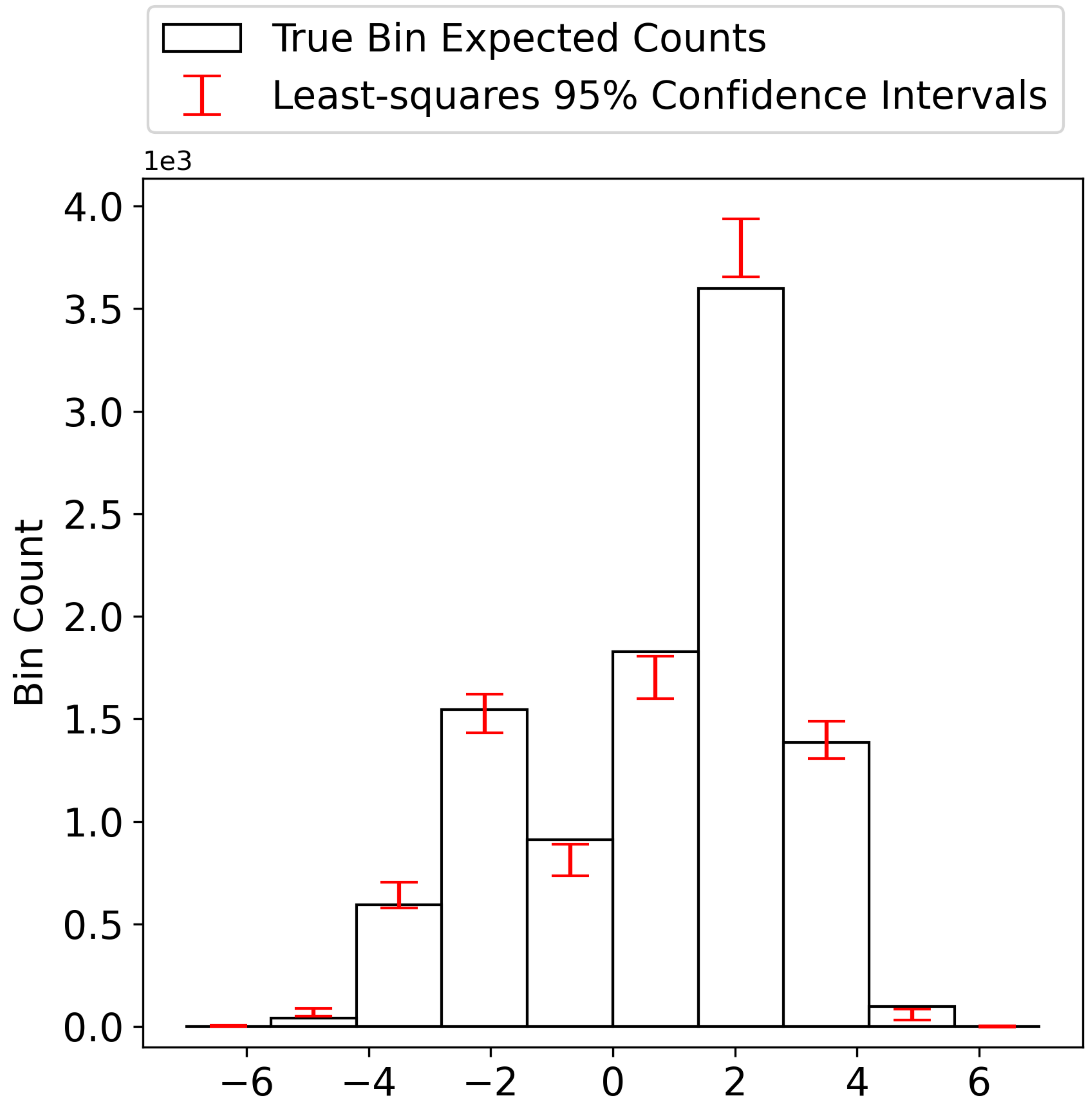
2. Aggregate into wide bins keeping track of correlation for error propagation

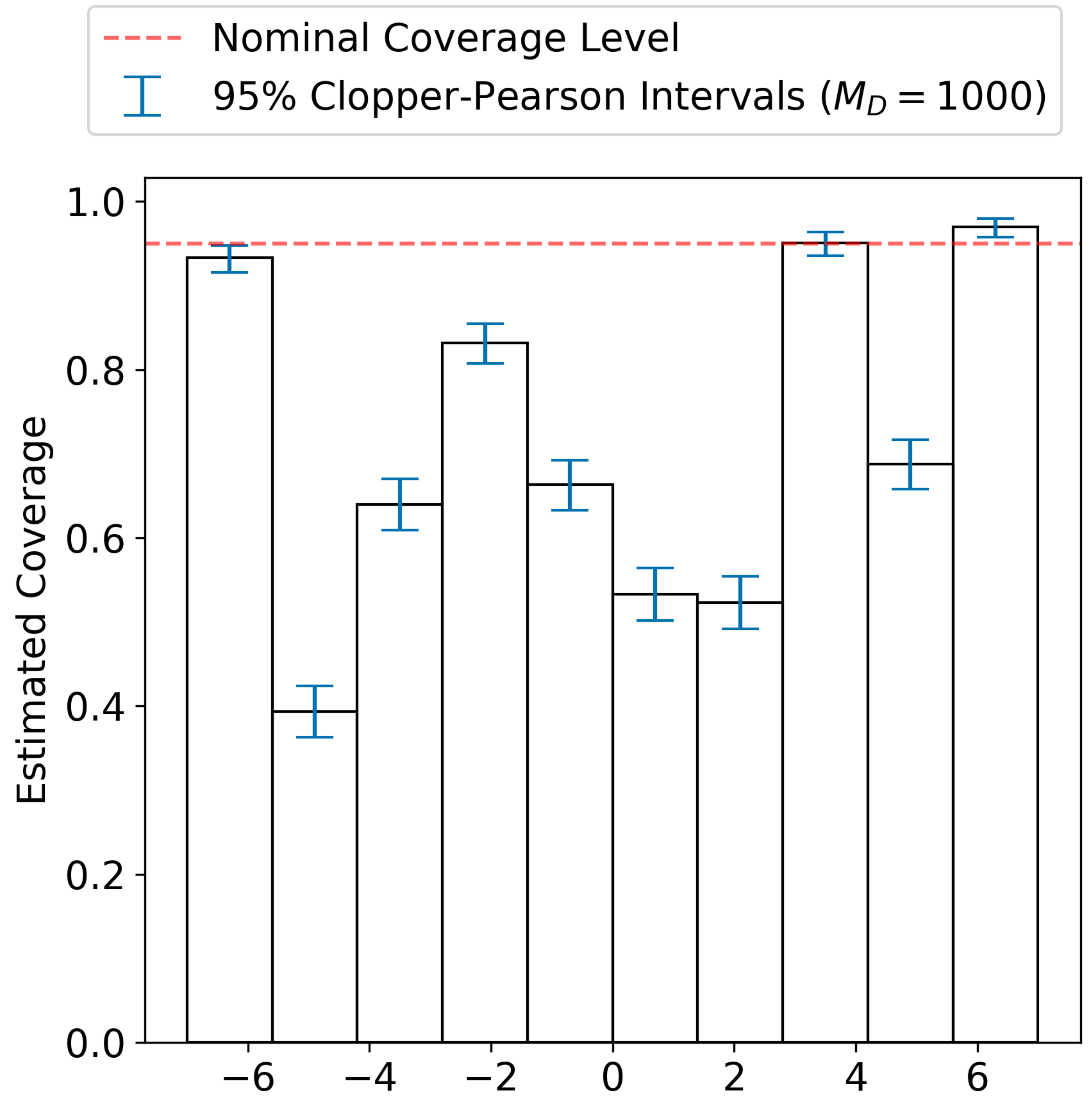


A view at misspecification via $|\boldsymbol{K}_{ij} - \boldsymbol{K}_{ij}^{MC}|$

Using Least-squares intervals with wide-bins suffers from MC misspecification

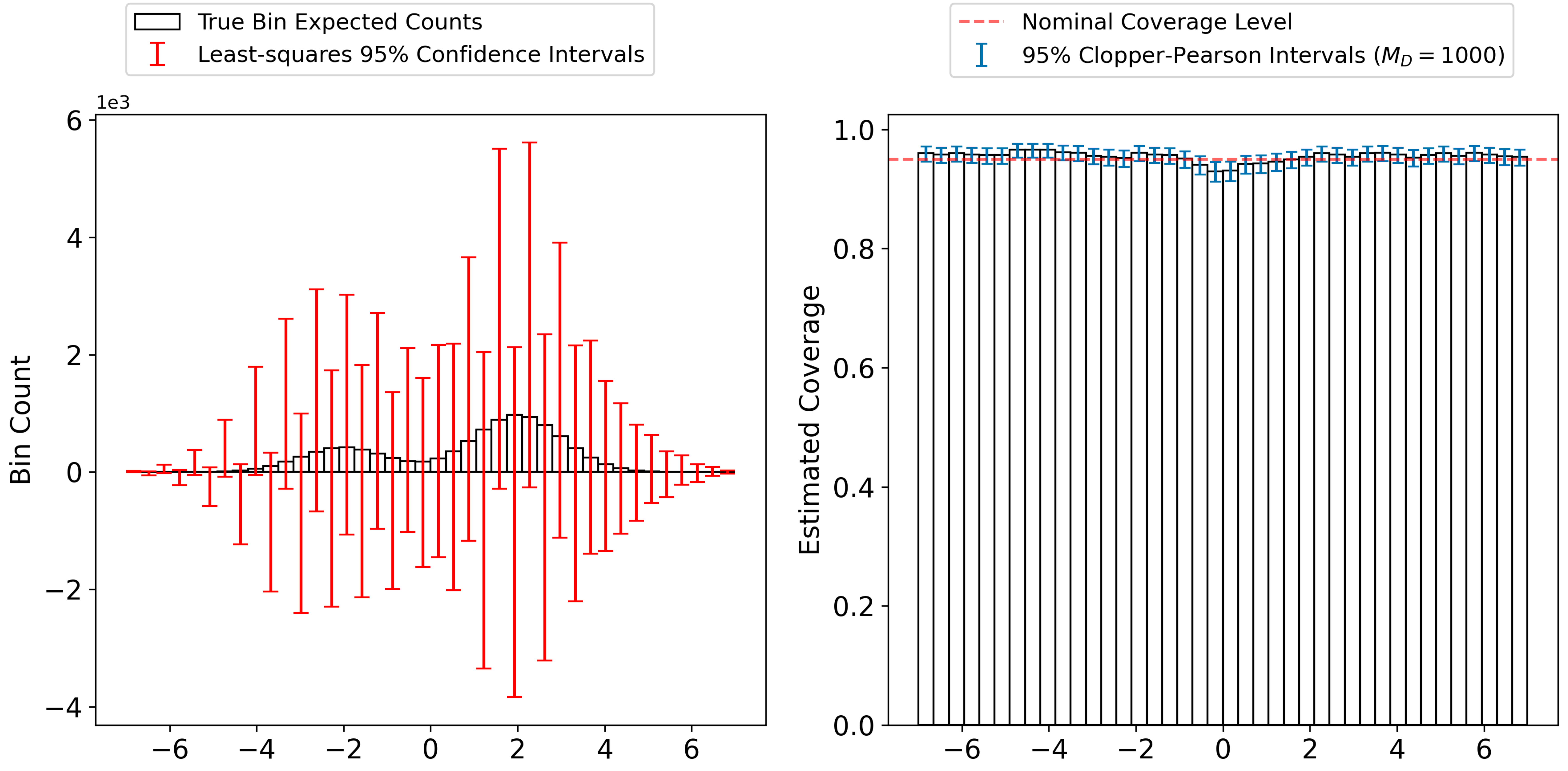
Create intervals with LS estimator, $\hat{\lambda}_{LS} = \text{argmin}_{\lambda} \|\boldsymbol{y} - \boldsymbol{K}^{MC}\boldsymbol{\lambda}\|^2_{\boldsymbol{\Sigma}^{-1}}$





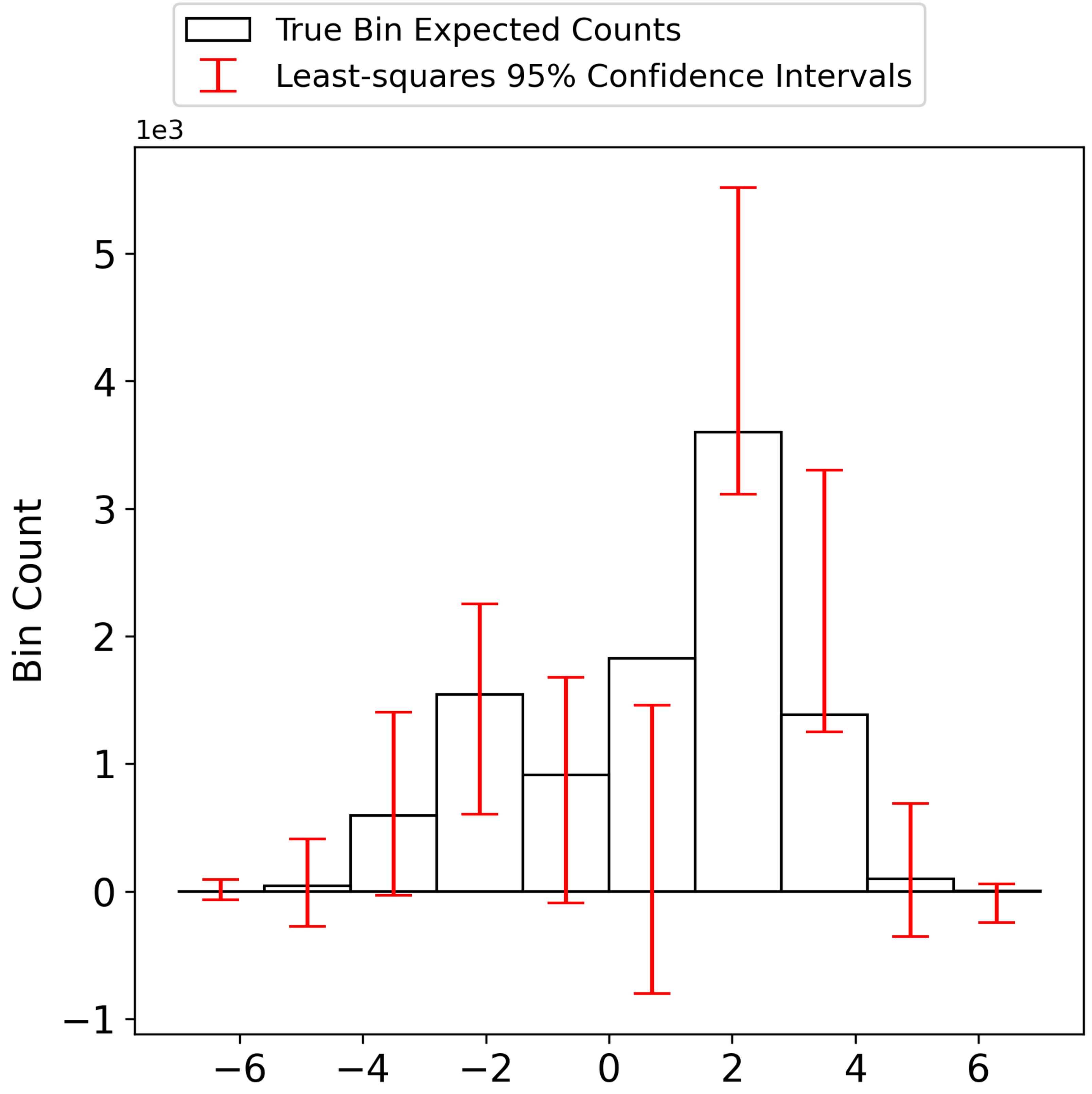
Unfolding to fine-binning with LS intervals fixes coverage

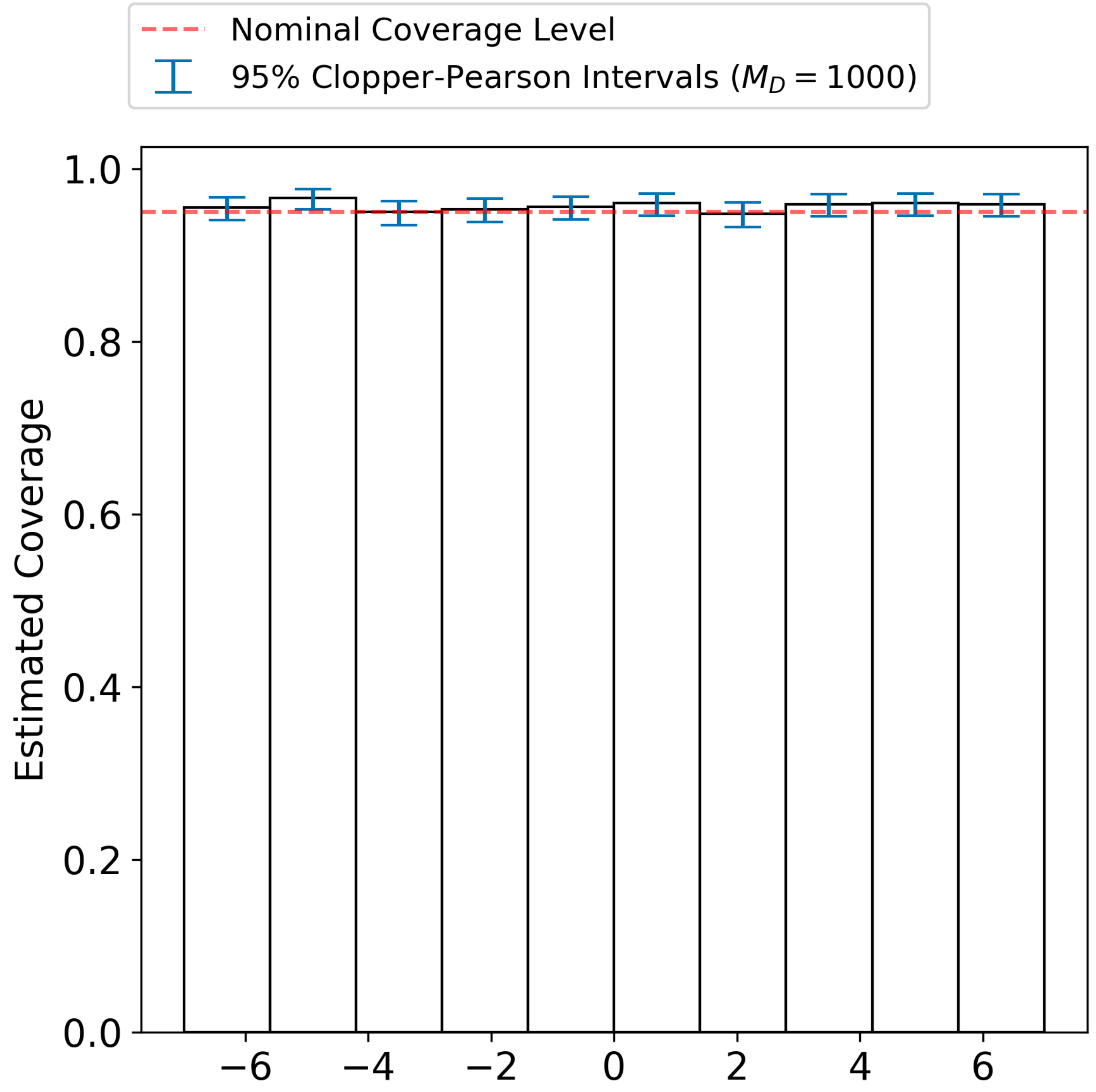
Narrower bins reduce the misspecification effect of f^{MC}, but at the cost of wide intervals.



Our proposed recipe produces intervals with correct coverage

Use the sampling distribution of $\hat{\lambda}_{LS}$ to create CIs for $\theta_k\left(\hat{\lambda}_{LS}\right) = \boldsymbol{h}_k^T\hat{\lambda}_{LS}$, i.e., aggregated fine bins

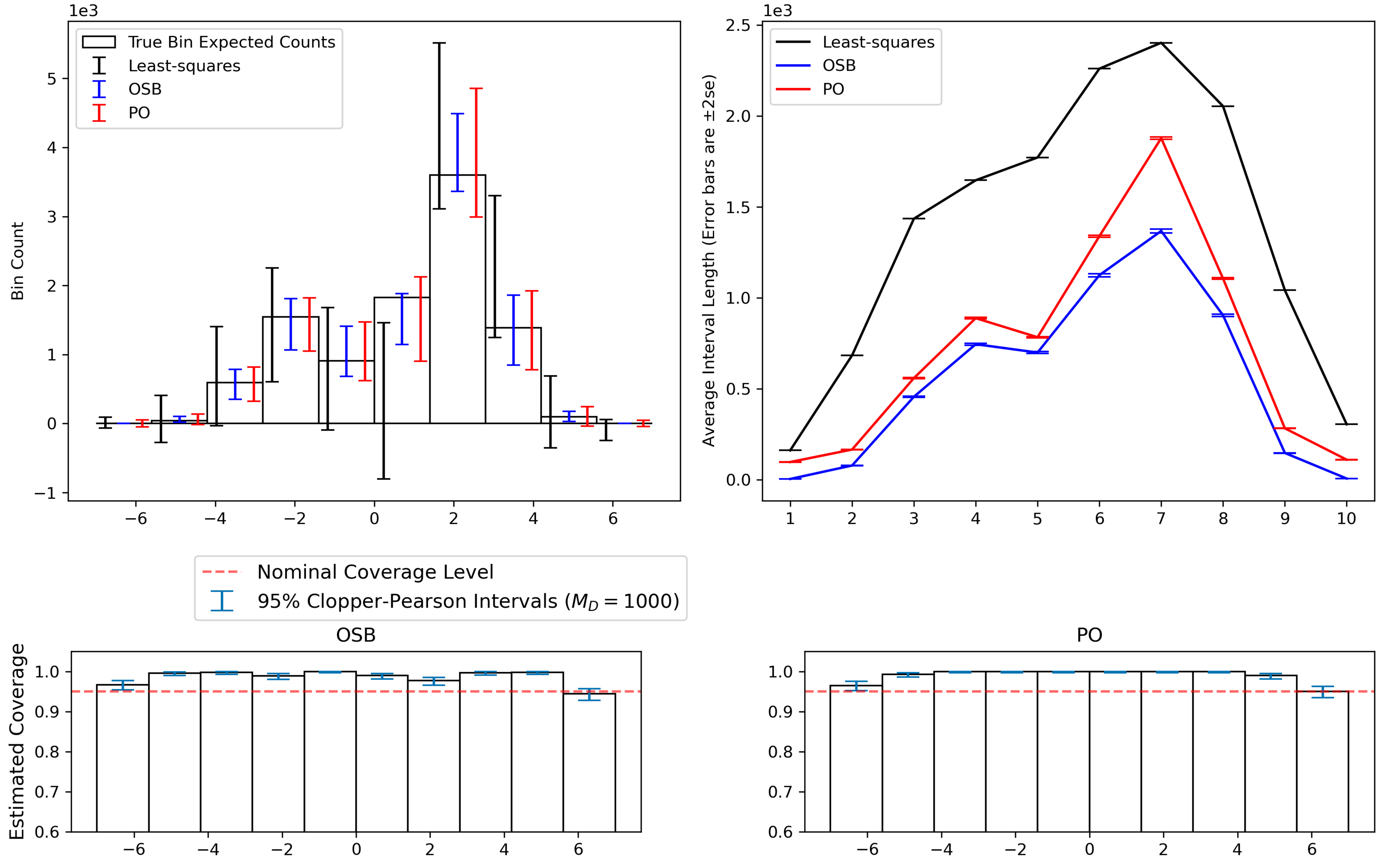




We can do even better by including physical constraints and allowing for rank-deficient $\boldsymbol{K}$

1. Some limitations of LS intervals — long, violate non-negativity, $\boldsymbol{K}^T\boldsymbol{K}$ is invertible only if $\boldsymbol{K}$ is full-column rank —> the misspecification from f^{MC} can only be reduced by so much
2. We proposed two solutions allowing for rank-deficient $\boldsymbol{K}$ and incorporation of non-negativity constraints [Stanley, Kuusela, Patil, 2022]:
    1. One-at-a-time strict bounds (OSB) intervals
        1. Modified simultaneous strict bounds (SSB) intervals ([Stark, 1992], [Rust & O'Leary, 1994]) to have bin-wise coverage
    2. Prior-Optimized (PO) Intervals
        1. decision-theoretic intervals where a prior is used to optimize expected interval length subject to constraints guaranteeing coverage
        2. **Important**: prior misspecification does not affect coverage, as we only use the prior to choose from a class of intervals with guaranteed coverage

OSB and PO intervals show significant length improvements while maintaining coverage



NOTE: this result is only showing the benefit of including a non-negativity constraint. We still used full-rank $\boldsymbol{K}$. See [Stanley, Kuusela, and Patil, 2022] for the rank-deficient case.

Conclusions and next steps

1. We have identified four sources of systematics in unfolding: regularization bias, wide-bin bias, missing auxiliary variables, and response kernel uncertainty

2. We can potentially address the first two by unfolding on fine bins, and then aggregating to wide bins. We provide two methods (OSB and PO intervals) allowing for both physical constraint inclusion and rank deficient $\boldsymbol{K}$

3. **Local next:** Richard will discuss how to address response kernel uncertainty

4. **Non-local next**: generalizations of strict bounds intervals as LR test inversion…stay tuned!

Thank you!

And special thanks to the **NSF AI Planning Institute for Data-Driven Discovery in Physics** for travel support.

References

G. D’Agostini. A multidimensional unfolding method based on bayes’ theorem. *Nuclear Instruments* and Methods in Physics Research A, 362:487–498, 1995.

Andreas Höcker and Vakhtang Kartvelishvili. Svd approach to data unfolding. *Nuclear Instruments and Methods in Physics Research A*, 372:469–481, 1996.

Mikael Kuusela. Uncertainty quantification in unfolding elementary particle spectra at the Large *Hadron Collider*. PhD thesis, École Polytechnique Fédérale de Lausanne, 7 2016.

Bert W. Rust and Dianne P. O’Leary. Confidence intervals for discrete approximations to ill-posed problems. Journal of Computational and Graphical Statistics, 3(1):67–96, 1994.

Michael Stanley, Mikael Kuusela, and Pratik Patil. Uncertainty quantification for wide-bin unfolding: one-at-a-time strict bounds and prior optimized confidence intervals. JINST. 2022.

Philip B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. Journal of Geophysical Research, 97(B10):14055–14082, 1992.

LS Interval construction

<u>Data generation process</u>: $\boldsymbol{y} = \boldsymbol{K\lambda} + \varepsilon, \quad \varepsilon \sim N(\boldsymbol{0}, \boldsymbol{I}_m)$

<u>LS Estimator</u>: $\hat{\lambda}_{LS} = \left(\boldsymbol{K}^T\boldsymbol{K}\right)^{-1}\boldsymbol{K}^T\boldsymbol{y}$

We can thus find the sampling distribution

$$\theta\left(\hat{\boldsymbol{\lambda}}_{LS}\right) = \boldsymbol{h}^T\hat{\boldsymbol{\lambda}}_{LS} \sim N\left(\boldsymbol{\lambda}, \boldsymbol{h}^T\left(\boldsymbol{K}^T\boldsymbol{K}\right)^{-1}\boldsymbol{h}\right),$$

and can construct the typical CIs.

OSB Interval Construction

<u>Data generation process</u>: $\boldsymbol{y} = \boldsymbol{K}\boldsymbol{\lambda} + \varepsilon, \quad \varepsilon \sim N(\boldsymbol{0}, \boldsymbol{I}_m), \ \boldsymbol{\lambda} \geq 0$

<u>Quantity of Interest</u>: $\theta(\boldsymbol{\lambda}) = \boldsymbol{h}^T \boldsymbol{\lambda}$

Intervals are constructed by solving the following endpoint optimizations

$$
\begin{aligned}
\text{min/max} \quad & \theta(\boldsymbol{\lambda}) \\
\text{subject to} \quad & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\lambda}\|_2^2 \leq z_{1-\alpha/2}^2 + s^2 \\
& \boldsymbol{\lambda} \geq 0
\end{aligned}
$$

PO Interval Construction (1)

We consider intervals of the form*

$$\delta_{PO}(\boldsymbol{y}) = \left[\underline{\boldsymbol{w}}^T \boldsymbol{y} - z_{1-\alpha/2}\|\underline{\boldsymbol{w}}\|_2, \bar{\boldsymbol{w}}^T \boldsymbol{y} + z_{1-\alpha/2}\|\bar{\boldsymbol{w}}\|_2\right] = \left[\underline{\theta}(\boldsymbol{y}; \underline{\boldsymbol{w}}), \bar{\theta}(\boldsymbol{y}; \bar{\boldsymbol{w}})\right]$$

We consider δ_{PO} to be a *decision rule* for intervals on the real line. As shown in [Stanley, Kuusela, Patil, 2022], any decision rule in

$$\mathscr{D}_c := \left\{\delta : \boldsymbol{h} - \boldsymbol{K}^T \underline{\boldsymbol{w}} \leq \boldsymbol{0}, \boldsymbol{h} - \boldsymbol{K}^T \bar{\boldsymbol{w}} \leq \boldsymbol{0}\right\}$$

produces a $1 - \alpha$ CI. Therefore, we seek a $\delta^* \in \mathscr{D}_c$ that is optimal, i.e., the Bayes-rule with respect to a prior on the expectation of the bin counts, $\boldsymbol{m}_\lambda$

*: This interval parameterization can take on a more general form to accommodate more general parameter constraints $\boldsymbol{A\lambda} \leq \boldsymbol{b}$. The above shows the case when $\boldsymbol{A} = -\boldsymbol{I}$ and $\boldsymbol{b} = \boldsymbol{0}$.

PO Interval Construction (2)

For each decision rule δ, we define a risk functional as the expected interval length

$$R(\delta) = \mathbb{E}_{\boldsymbol{y}}[L(\delta)] = \left(\bar{\boldsymbol{w}} - \underline{\boldsymbol{w}}\right)^T \boldsymbol{K\lambda} + z_{1-\alpha/2}\left(\|\bar{\boldsymbol{w}}\|_2 + \|\underline{\boldsymbol{w}}\|_2\right)$$

Hence, the *Bayes risk* is

$$r(\boldsymbol{m}_\lambda; \delta) := \left(\bar{\boldsymbol{w}} - \underline{\boldsymbol{w}}\right)^T \boldsymbol{K}\boldsymbol{m}_\lambda + z_{1-\alpha/2}\left(\|\bar{\boldsymbol{w}}\|_2 + \|\underline{\boldsymbol{w}}\|_2\right)$$

So, to find the Bayes rule we find δ^* such that

$$r(\boldsymbol{m}_\lambda; \delta^*) = \min\left\{r(\boldsymbol{m}_\lambda, \delta) : \delta \in \mathscr{D}_c\right\}$$

PO intervals are competitive with OSB on our GMM simulation example