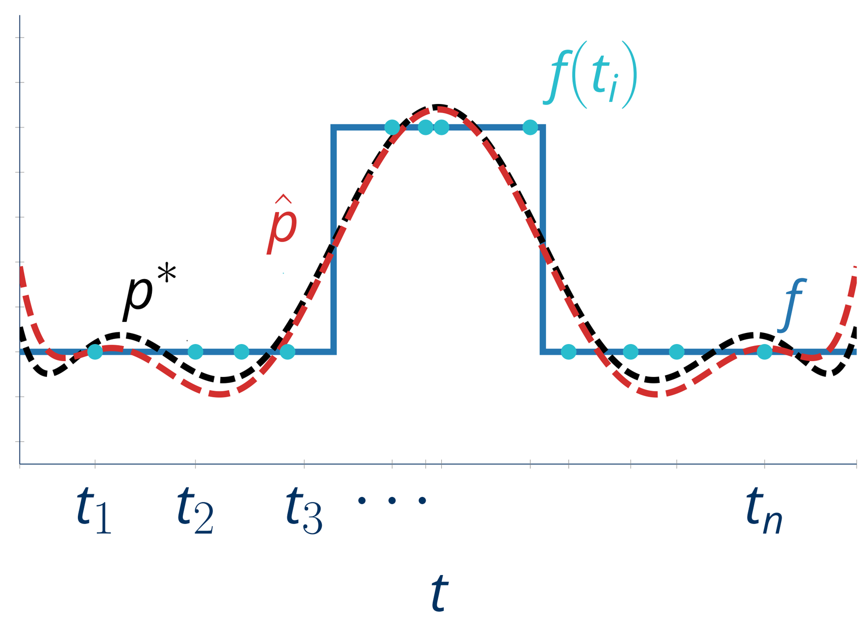
Debiasing Polynomial and Fourier Regression

Chris Camaño Raphael A. Meyer Kevin Shu

Polynomial Regression



We want to approximate a real-valued function f using a degree d polynomial $\hat{p}$. We can evaluate $f(t)$ for any $t \in \mathbb{R}$, and want near-minimal residual:

$$\|f - \hat{p}\|_\mu^2 \leq (1+\varepsilon) \min_{\deg(p) \leq d} \|f - p\|_\mu^2$$

where $\|f\|_\mu^2 = \int |f(t)|^2 \mu(t) dt$.

Prior Work

If you sample $n = \tilde{\mathcal{O}}(d/\varepsilon)$ samples iid from the *μ-leverage score distribution* τ_μ, then we can recover a near-optimal polynomial.

(+) Near-optimal sample complexity!

(–) Resulting polynomial is ***BIASED***:

$$\mathbb{E}[\hat{p}] \neq p^* := \operatorname*{argmin}_{\deg(p) \leq d} \|f - p\|_\mu$$

Our Contribution

For many measures μ, we can easily debias leverage score sampling.

(+) Near-optimal sample complexity!

(+) Unbiased!

(+) Much lower error for small sample sizes!

Algorithm

The μ-Unitary Ensemble (μ-UE) is the distribution over *complex* Hermitian matrices $\boldsymbol{X} \in \mathbb{C}^{d+1 \times d+1}$ with pdf

$$q(\boldsymbol{X}) \propto \prod_{i=1}^{d+1} \mu(\lambda_i(\boldsymbol{X})).$$

Debiased Regression

1. **Sample** Hermitian $\boldsymbol{X} \in \mathbb{C}^{d+1 \times d+1}$ from μ-UE
2. **Compute** eigenvalues $t_1, \ldots, t_{d+1}$ of $\boldsymbol{X}$
3. **Sample** $t_{d+1}, \ldots, t_n$ iid from distribution τ_μ
4. **Return** interpolating polynomial with

$$\hat{p} = \operatorname*{argmin}_{\deg(p) \leq d} \sum_{i=1}^{n} \frac{1}{\tau_\mu(t_i)} |f(t_i) - p(t_i)|^2$$

Results

Debiased regression is fast, data-efficient, and works for many μ distributions!

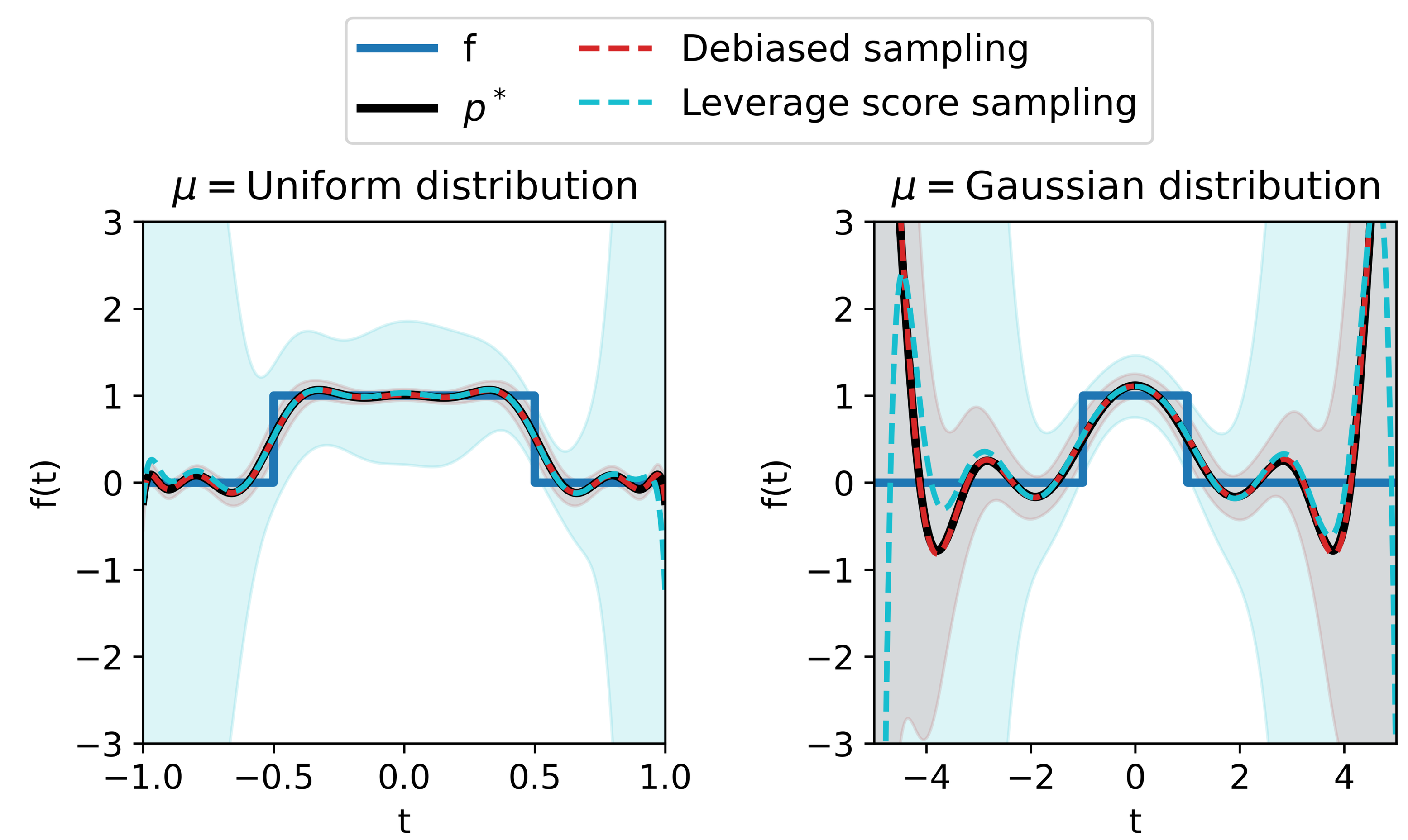


Figure 1: For each method, dashed lines shows empirical mean of $\hat{p}(t)$ across $100{,}000$ trials, with shaded intervals showing 1 standard deviation. Leveraged sampling is visibly biased while our method is not.

Fast & Easy Sampling

Sampling $\boldsymbol{X} \in \mathbb{C}^{d+1 \times d+1}$ from the μ-UE distribution seems hard. But sometimes it's easy:

Random Matrix Models [Dumitriu Edelman' 03]

Let $\boldsymbol{T} \in \mathbb{R}^{d+1 \times d+1}$ be symmetric tridiagonal with

- *Diagonal entries $\alpha_i \sim \mathcal{N}(0,1)$*
- *Off-diagonal entries $\beta_i \sim \chi_{2i}$*

Then the eigenvalues of $\boldsymbol{T}$ are distributed as the eigenvalues of $\boldsymbol{X}$ if $\mu = \mathcal{N}(0,1)$.

(+) Computing eigenvalues of $\boldsymbol{T}$ takes $\mathcal{O}(d \log d)$ time!

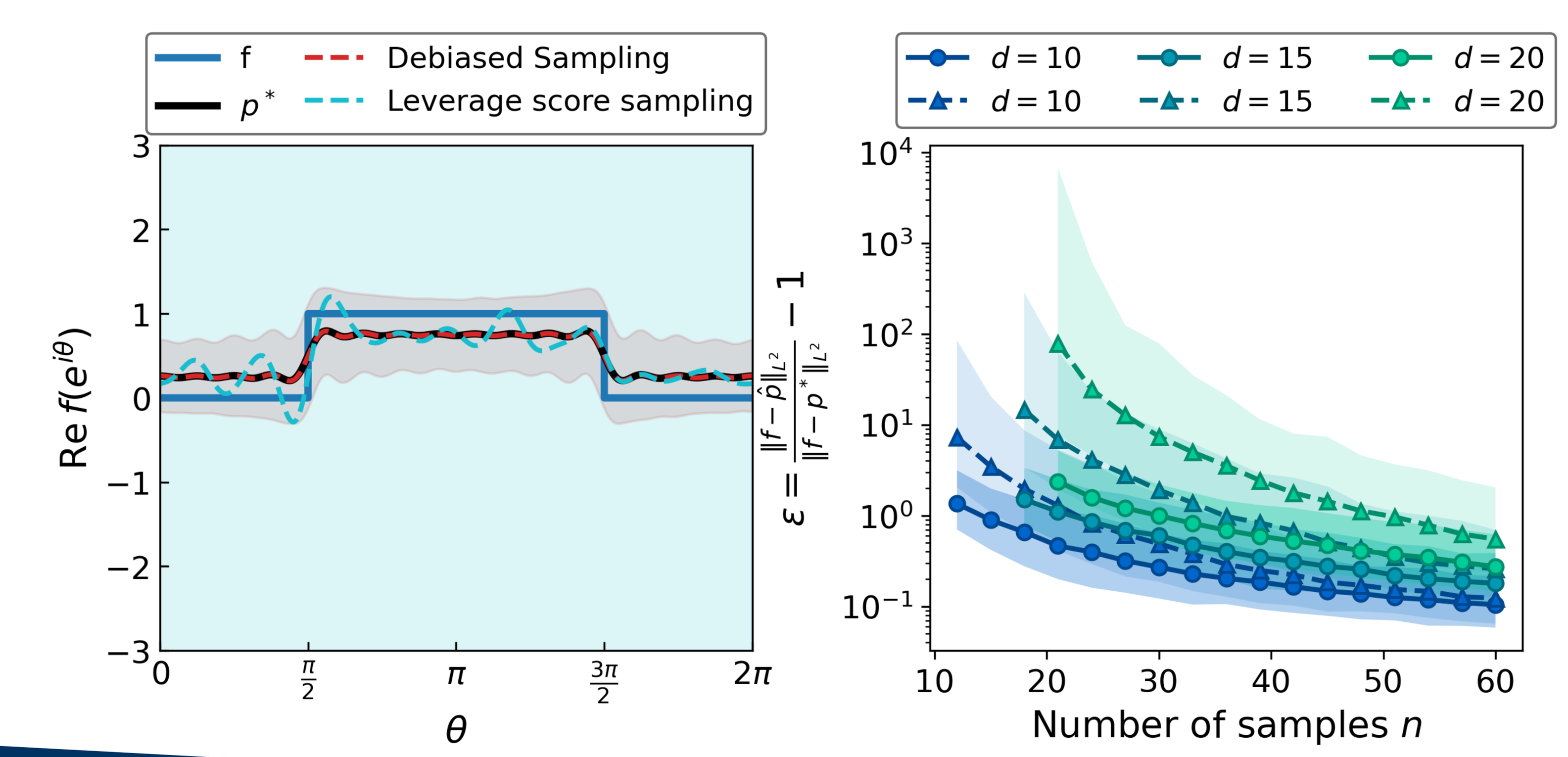
(+) Similar results hold for uniform distribution, beta distribution, chi-squared distribution, and more!

Why does it work?

To subsample-then-solve any least-squares problems without bias, [Derezinksi et al.] show that sampling from the *P-DPP* and leverage scores of the rows of $\boldsymbol{A}$ suffices. We observe that the P-DPP for polynomial regression is exactly distributed as eigenvalues of a μ-UE.

Fourier Regression

Consider a periodic function $f : \{e^{i\theta} : \theta \in [0, 2\pi)\} \to \mathbb{C}$ we want to approximate with a Fourier series $\hat{p}$ of degree d. We use similar ideas to also debias this problem.



Berkeley UNIVERSITY OF CALIFORNIA Caltech