

Robust Prior for Simulation-based Inference

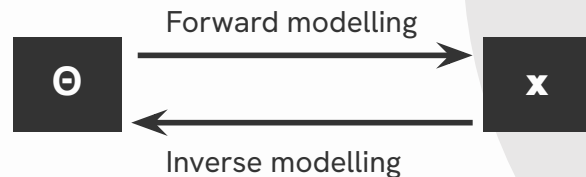
Chaipat (Tata) Tirapongprasert

$$p(\Theta | \mathbf{x}_1, \dots, \mathbf{x}_n) \sim \prod_i p(\mathbf{x}_i | \Theta) p(\Theta)$$

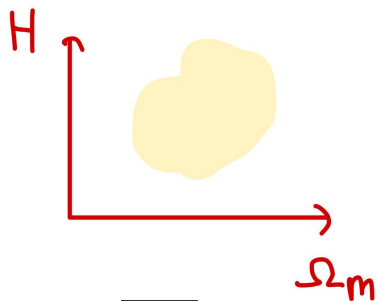
↓
posterior

↙
likelihood

↘
prior



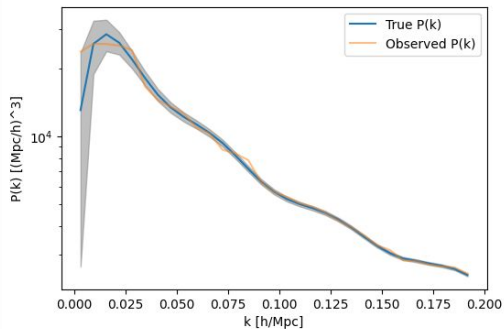
Goal: Do **Bayesian inference** on **new measured data** (x_0)



01

Sample from the **prior**

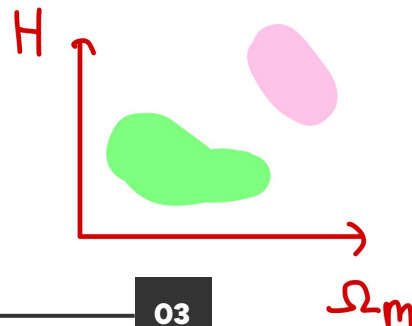
$\Theta \sim p(\Theta)$
prior distribution



02

Simulate from the **forward model**

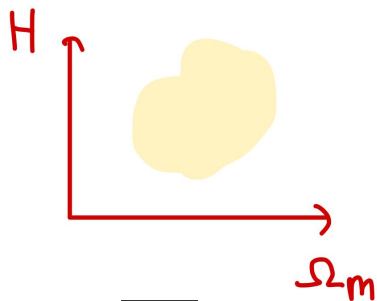
simulated (observable) data $\leftarrow x \sim p(x | \Theta)$
data-generating process (or likelihood)



03

Calculate the **posterior** $p(\Theta | x_0)$

Goal: Do **Bayesian inference** on **new measured data** (x_0)

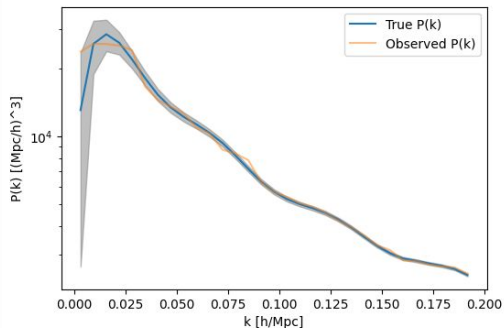


01

Sample from the
prior

$$\Theta \sim p(\Theta)$$

prior distribution



02

Simulate from the
forward model

simulated
(observable)
data

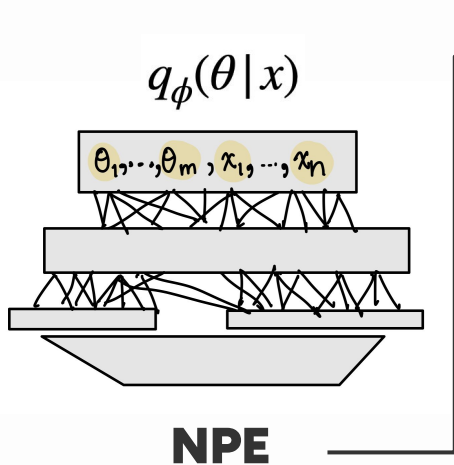
$$\mathbf{x} \sim p(\mathbf{x} | \Theta)$$

data-generating
process (or
likelihood)

Problem

Simulators can be very expensive! (usually the likelihood also cannot be analytically written down)

Neural Posterior Estimation (NPE)



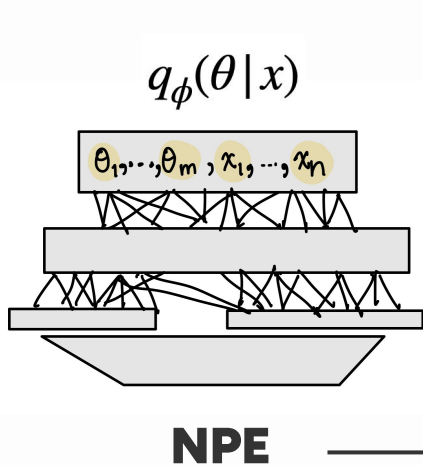
Goal

Train a neural density estimator $q_\phi(\theta|x)$ to approximate the true posterior $p(\Theta|x)$

$$\mathcal{L}_{\text{NPE}} := -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log q_\phi(\theta|x)$$

↓
expected value over simulations of the
negative log-probability estimator

Neural Posterior Estimation (NPE)



Goal

Train a neural density estimator $q_\phi(\theta | x)$ to approximate the true posterior $p(\Theta | x)$

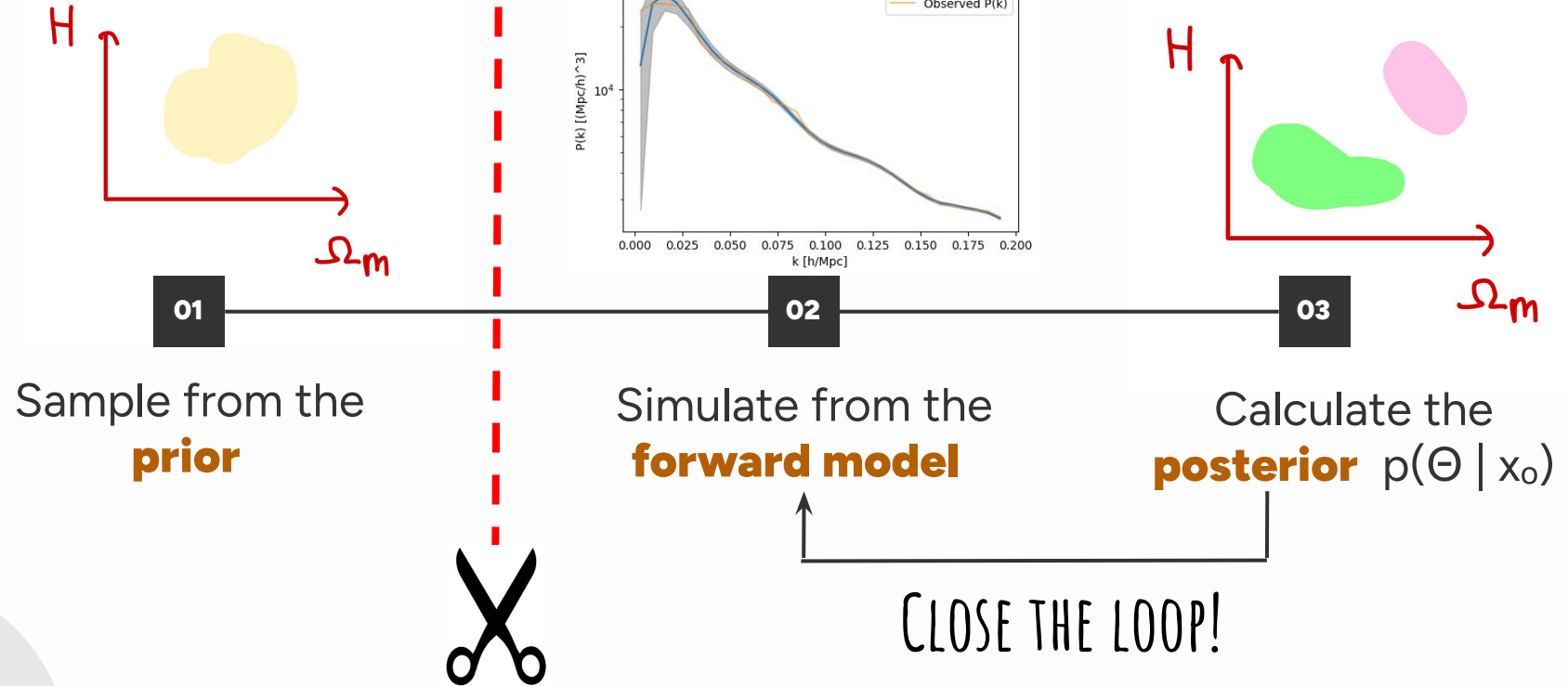
$$\mathcal{L}_{\text{NPE}} := -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log q_\phi(\theta | x)$$

↓
expected value over simulations of the negative log-probability estimator

Amortization

Plug empirical data x_0 into the density estimator to calculate the posterior $p(\Theta | x_0)$

Sequential NPE (SNPE) iteratively replace the prior $p(\Theta)$ with a suitably chosen proposal distribution



Problem

The network will learn to approximate a *proposal-biased* posterior

$$q_{\Phi}^{\text{NPE}}(\Theta | \mathbf{x}) \simeq \frac{1}{Z} p(\mathbf{x} | \Theta) \tilde{p}(\Theta)$$

Problem

The network will learn to approximate a *proposal-biased* posterior

$$q_{\Phi}^{\text{NPE}}(\Theta | \mathbf{x}) \simeq \frac{1}{Z} p(\mathbf{x} | \Theta) \tilde{p}(\Theta)$$

Fix:

Importance weights in the loss function

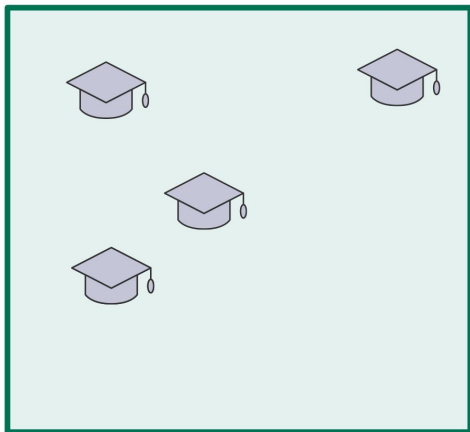
$$\mathcal{L}_{\text{NPE}} = -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log \left[\frac{\mathcal{P}(\theta)}{\tilde{\mathcal{P}}(\theta)} q_{\phi}(\theta | x) \right]$$

↓
Proposal prior

← typically concentrates simulation on parameter regions that are pertinent to the relevant observation x_0 .

So...how do we sample from a **prior**?

- Assumed prior
- Proposal prior

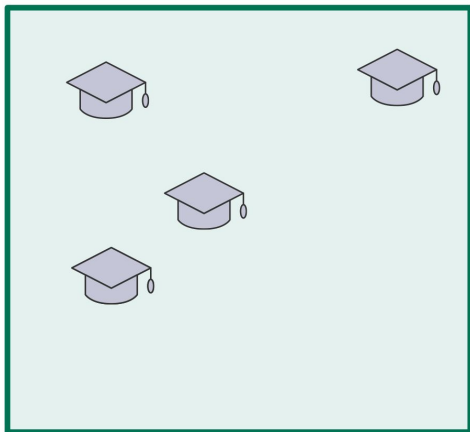


Monte Carlo Sampling

What's the problem with this?

So...how do we sample from a **prior**?

- Assumed prior
- Proposal prior



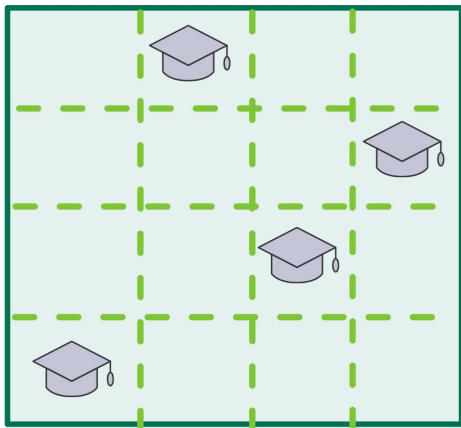
Monte Carlo Sampling

What's the problem with this?

The previously generated sample points are not taken into account when producing new samples.

So...how do we sample from a **prior**?

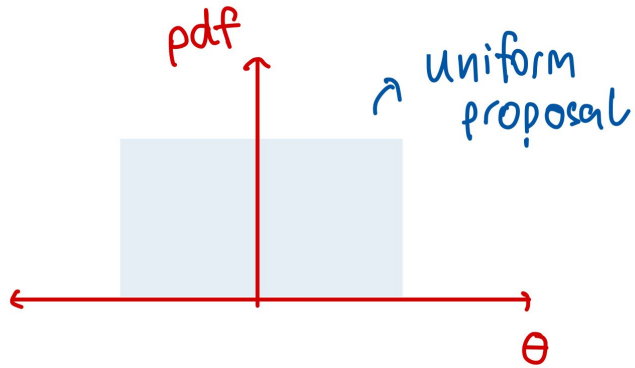
- Assumed prior
- Proposal prior



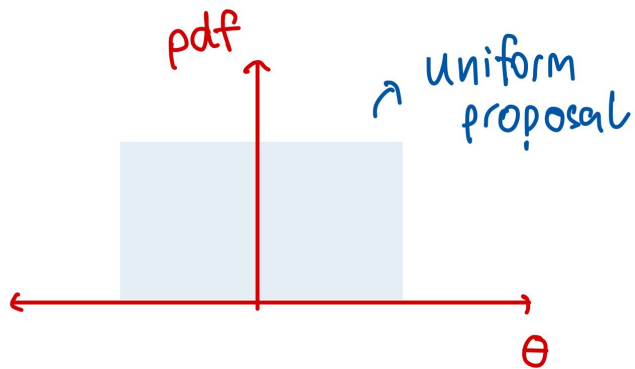
**Latin Hypercube
Sampling (LHS)**

- The input probability distribution should be stratified before a random sample is chosen from each interval.
- As a result, there is one sample in each row and column

Challenges with LHS



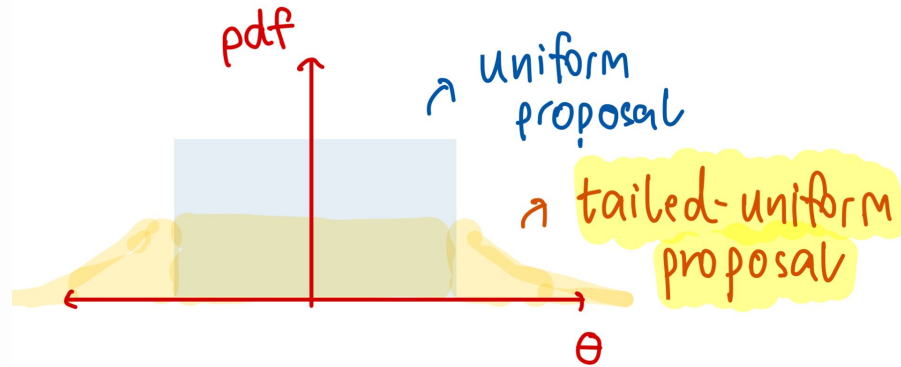
Challenges with LHS



Problem

Without gradient information at the boundaries, our flow-based density estimator cannot learn the correct density transitions near the cutoff.

Challenges with LHS



$$\tilde{\mathcal{P}}_{\text{TAILED-UNIFORM}}(x; a, b, \sigma) = \begin{cases} A \cdot \mathcal{N}(a, \sigma^2), & x \leq a \\ B \cdot \mathcal{U}(a, b), & x \in [a, b] \\ A \cdot \mathcal{N}(b, \sigma^2), & x \geq b \end{cases}$$



Tailed-Uniform Proposal

By feeding the network training samples from outside the uniform region (the extended Gaussian support beyond the original grid)

1. Tailed-Uniform captures the posterior density outside the assumed prior bounds, and
2. Tailed-Uniform learns a smooth, well-calibrated posterior that extends into boundary-adjacent test points.

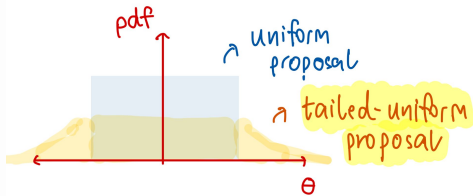
Toy Problem



Uniform proposal



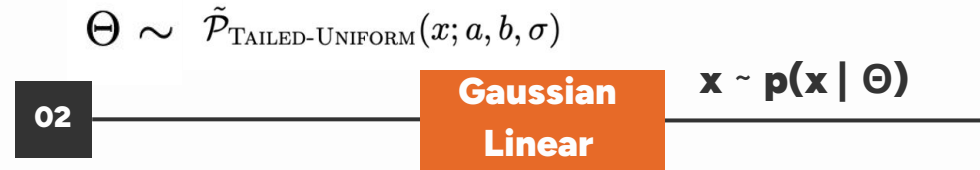
Tailed-Uniform proposal



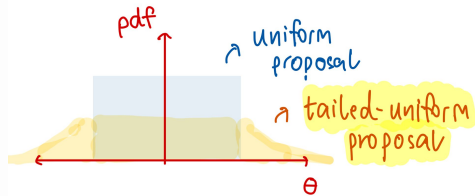
Toy Problem



Uniform proposal



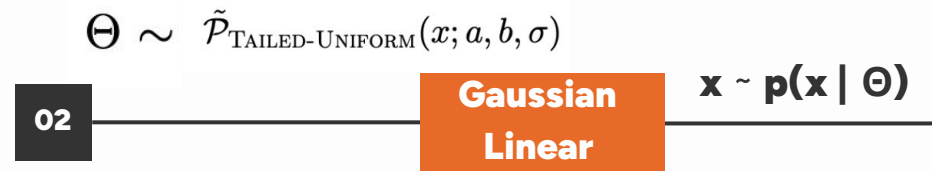
Tailed-Uniform proposal



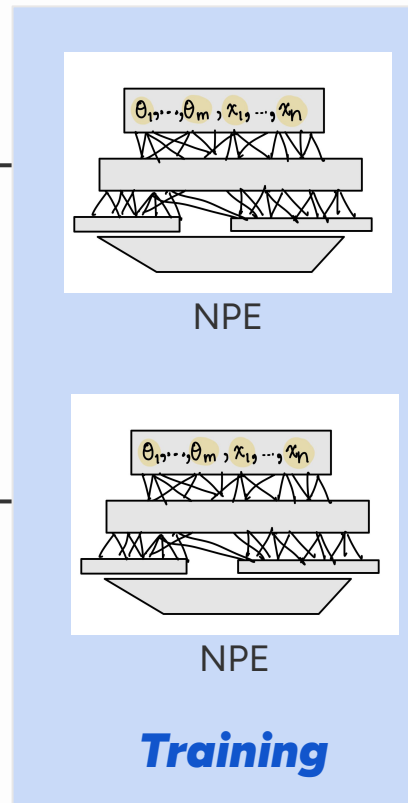
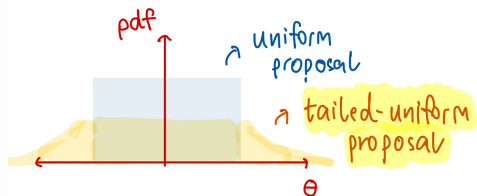
Toy Problem



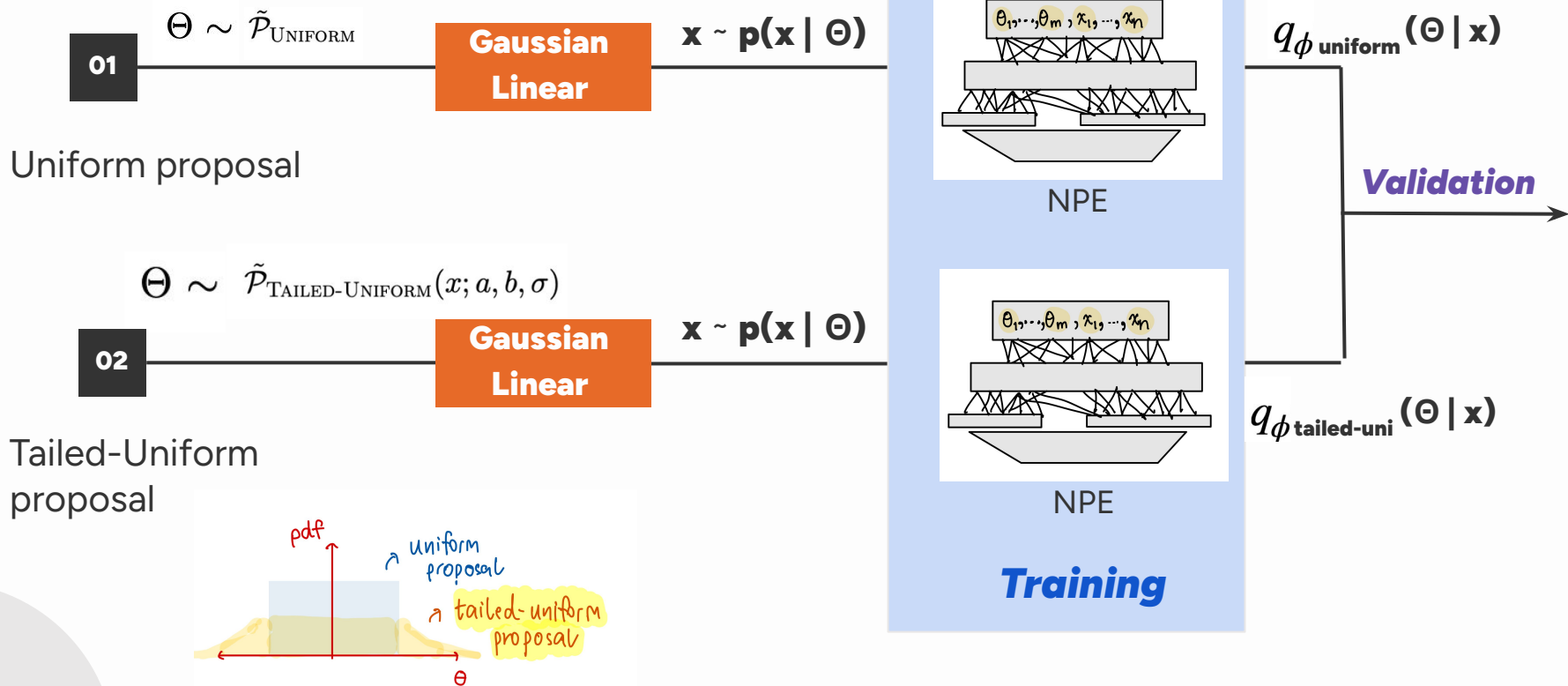
Uniform proposal



Tailed-Uniform proposal

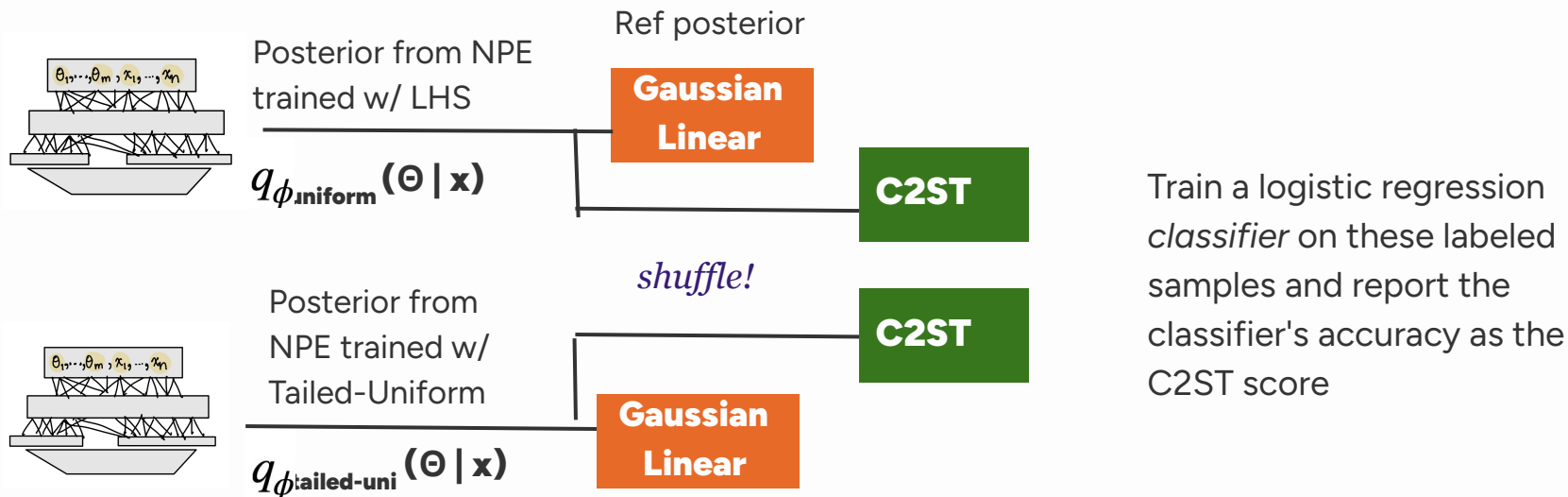


Toy Problem



Validation: Posterior Predictive Checks

Goal: If our NN is accurate, the produced data ought to "look close" to the observed data.



Result

5

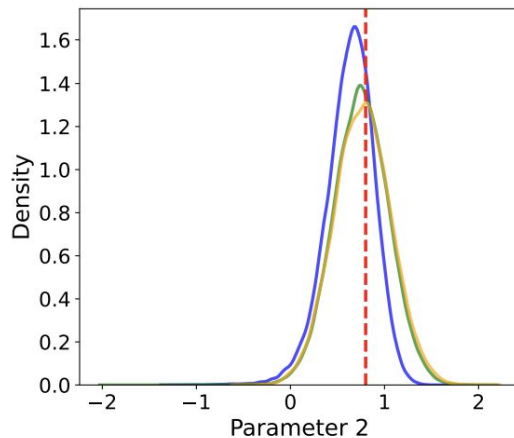
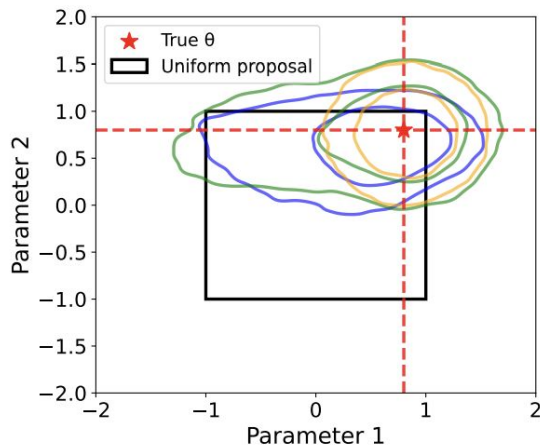
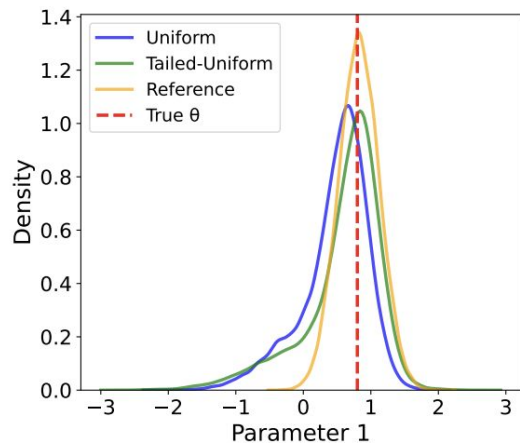


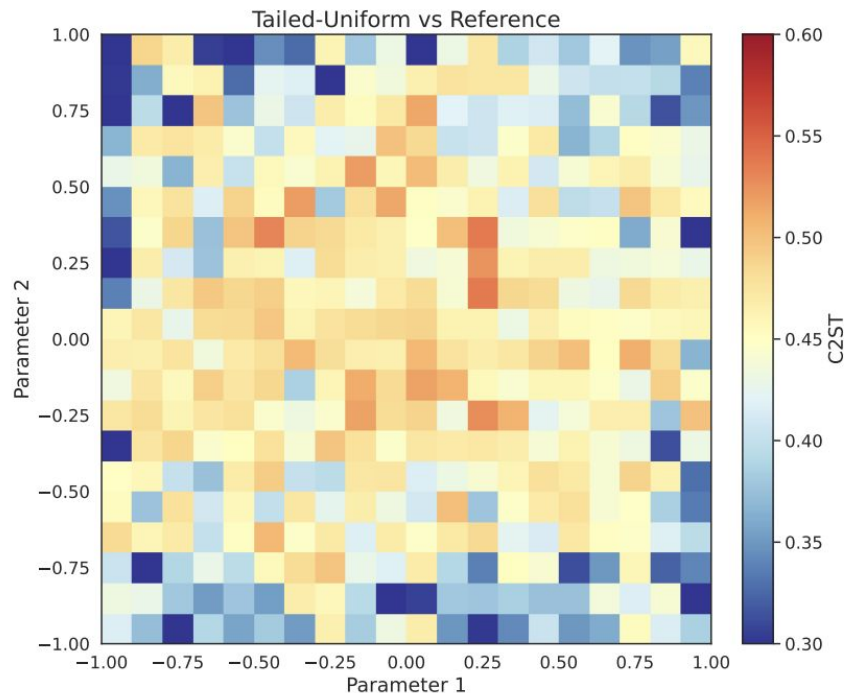
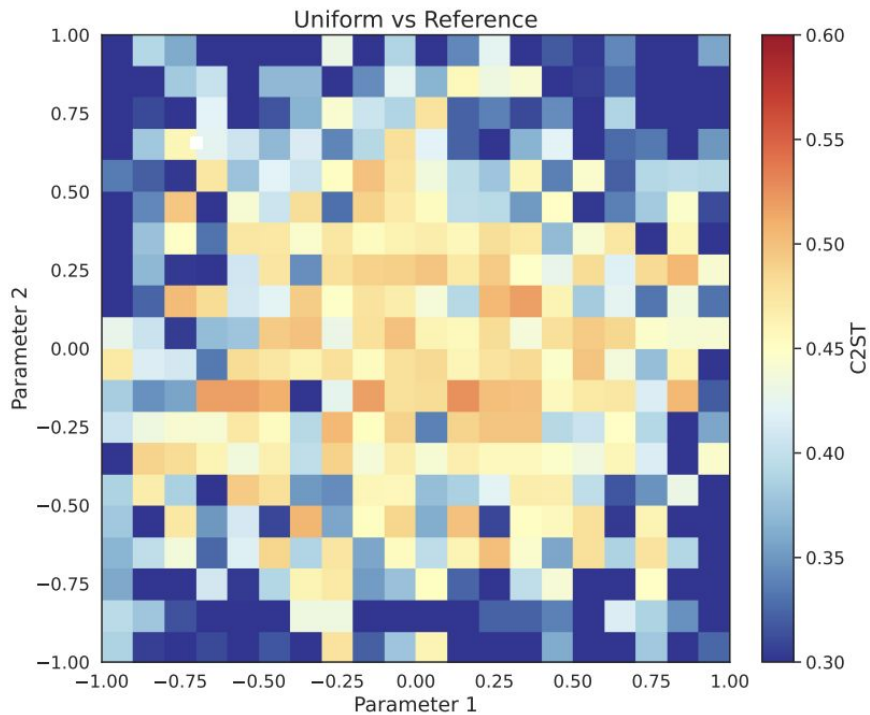
TABLE 1
POSTERIOR PERFORMANCE FOR $\theta_{\text{TRUE}} = (0.8, 0.8)$.

Method	Parameter Estimates		C2ST
	θ_1	θ_2	
Analytical Reference	0.825 ± 0.301	0.763 ± 0.299	0.500
TAILED-UNIFORM	0.568 ± 0.590	0.777 ± 0.279	0.417
UNIFORM	0.469 ± 0.519	0.644 ± 0.286	0.329
C2ST Improvement			+27%

Result

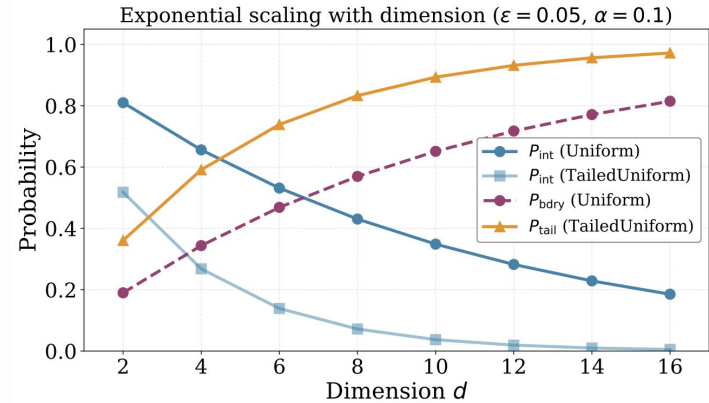
5

- Uniform performance deteriorates (becomes blue) as it drifts away from the central regions.
- Tailed-Uniform maintains more red and orange coloration across the parameter space.



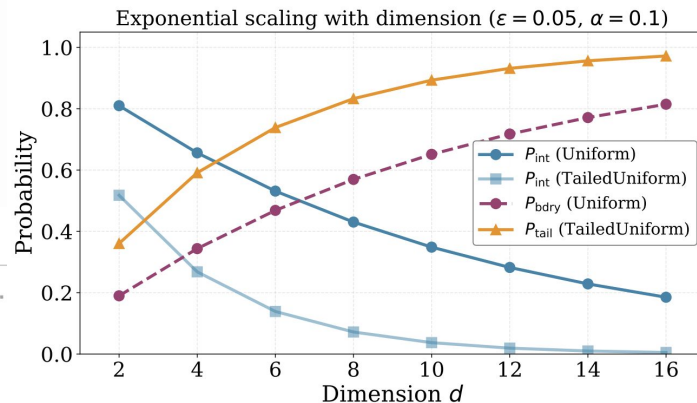
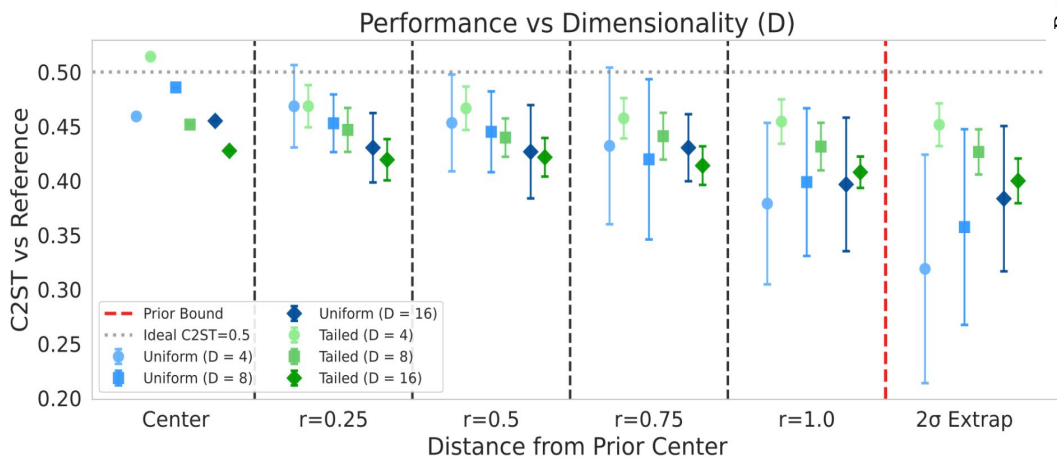
Boundary Dominance in Higher Dimensions

Curse of dimensionality: the volume of the parameter space is increasingly dominated by the boundary shell rather than the interior



Boundary Dominance in Higher Dimensions

Curse of dimensionality: the volume of the parameter space is increasingly dominated by the boundary shell rather than the interior



The ratio of information gained from exterior support to cost of interior sample dilution remains favorable or even improves with dimension.

Science Problem: Change

Gaussian
Linear

into

syren_new

+ cosmic noise



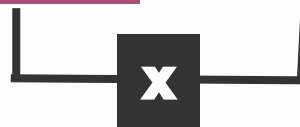
Observed power spectrum $P(k)$

→ Run MCMC to get reference posteriors

Science Problem: Change

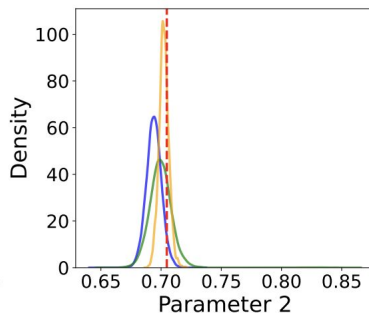
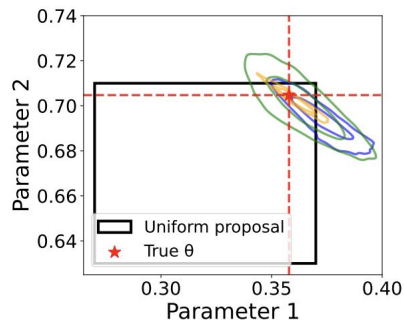
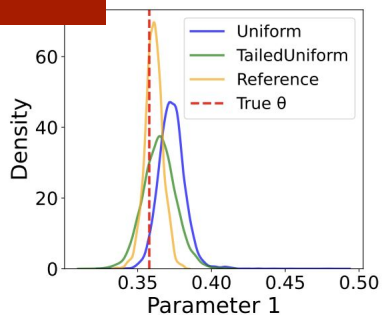
Gaussian
Linear

into **syren_new** + *cosmic noise*



Result

\mathcal{S}



Observed power spectrum $P(k)$

→ Run MCMC to get reference posteriors

Degeneracies: The information embedded in our posterior will be weak along certain directions in the parameter space, where Ω_m trades off against h such that $\Omega_m h$ remains about constant.

WORK IN PROGRESS: Degeneracy Detector

S-curve

Gaussian

x	y	z	a	b	c	d

Posterior

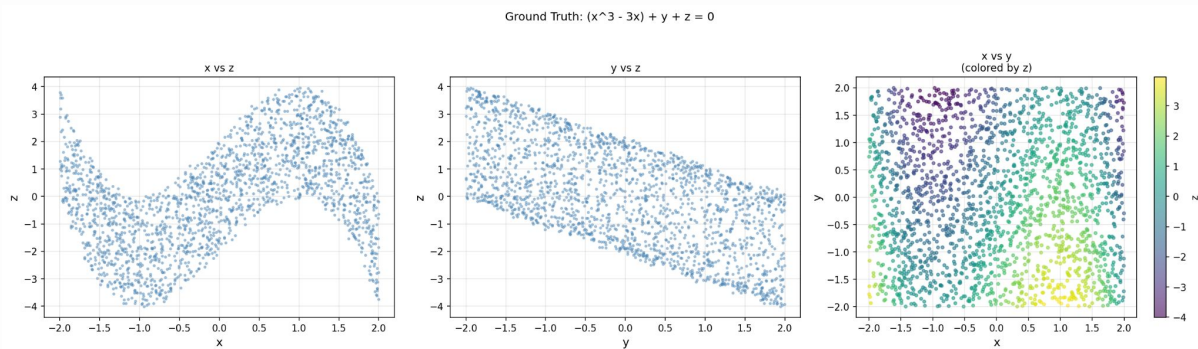
WORK IN PROGRESS: Degeneracy Detector

S-curve

Gaussian

x	y	z	a	b	c	d

Posterior



WORK IN PROGRESS: Degeneracy Detector

S-curve

Gaussian

x	y	z	a	b	c	d

Posterior

Goal:

- Detect parameter degeneracies in posterior samples.
- Discovers symbolic relationship

$$\mathbf{g}_1(\Theta_1) + \mathbf{g}_2(\Theta_2) + \dots + \mathbf{g}_k(\Theta_k) =$$

using **mutual information** and **symbolic regression**.

```
result = detector.search_couplings(  
    coupling_depth=3,      # tuple size (2=pairs, 3=triplets)  
    max_fits=2,           # limit number of fits (top N by MI)  
)
```

$$I(X; Y) \triangleq \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$



Mutual information (MI)

- the average information about X gained when observing Y.

$$I(X; Y) \triangleq \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

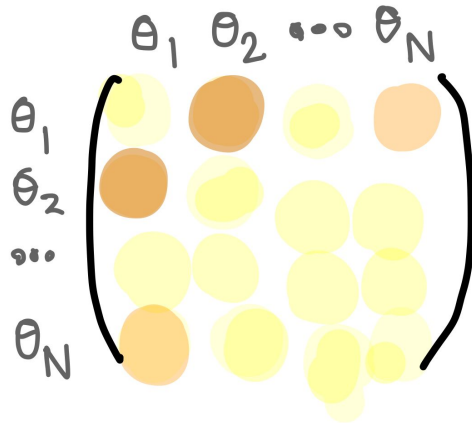


Mutual information (MI)

- the average information about X gained when observing Y.



calculate pairwise MI

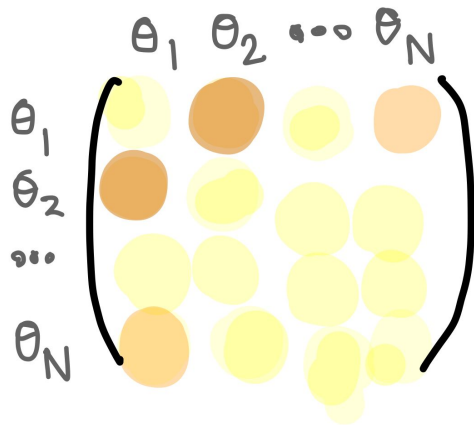


$$I(X; Y) \triangleq \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Mutual information (MI)

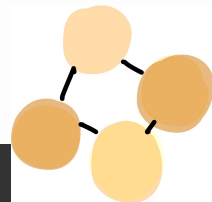
- the average information about X gained when observing Y.

calculate pairwise MI

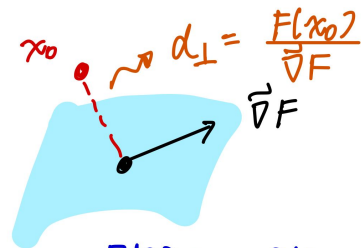


create k-tuples

```
coupling_depth=4
```



fit with symbolic regression



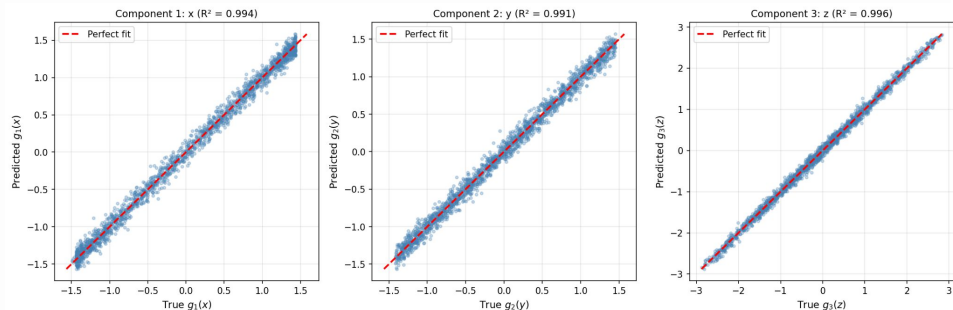
$$F(\theta) = y - f(\theta_1, \dots, \theta_k) = 0$$

Evaluate with R_{ortho}^2

COMING SOON (in ~ 1-2 months): A Python package for detecting parameter degeneracy in posterior samples

MI find these triplets!

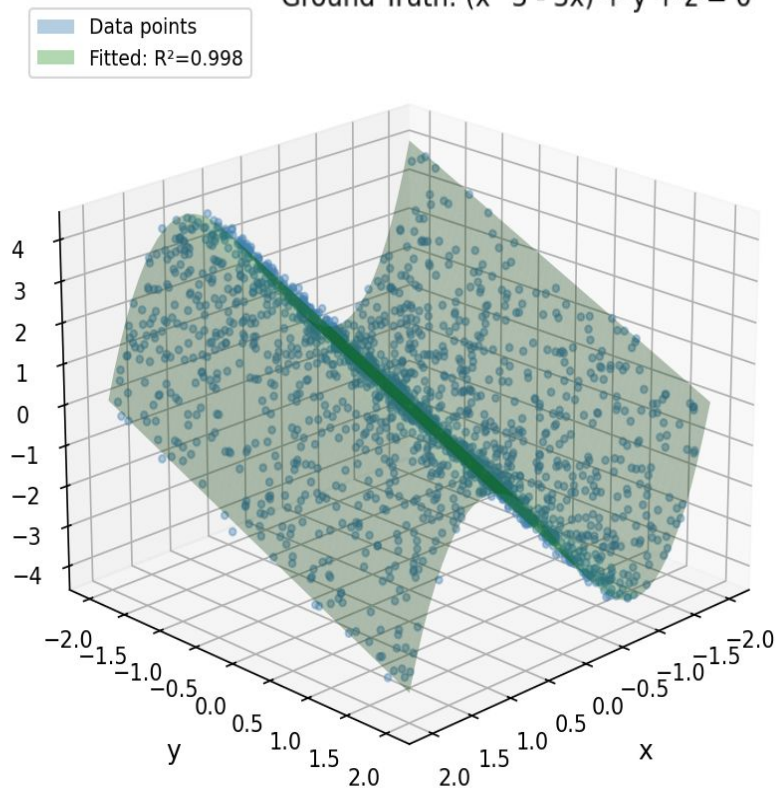
x	y	z	a	b	c	d



PySR fit the surface

Data + Fitted Surface
 $R^2_{\text{ortho}} = 0.9983$

Ground Truth: $(x^3 - 3x) + y + z = 0$

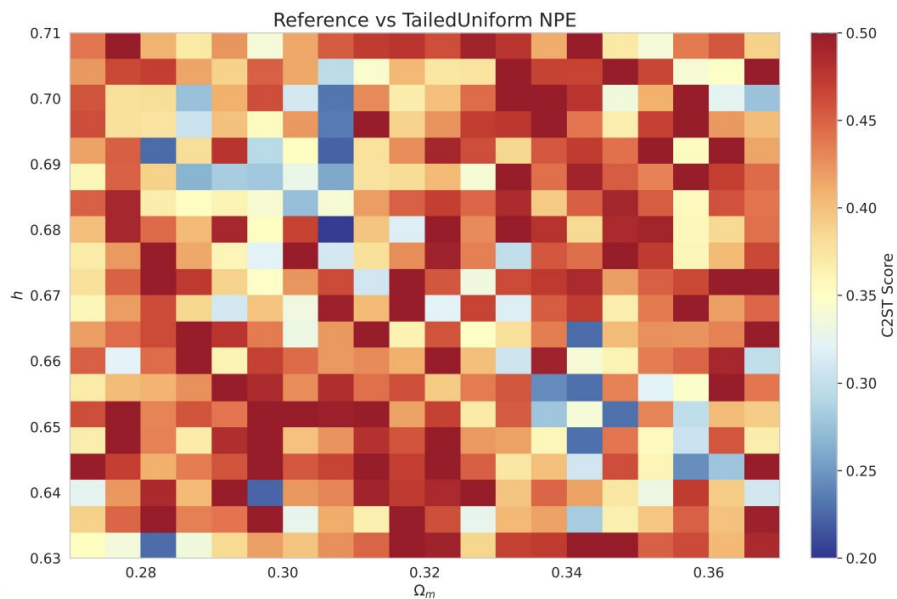
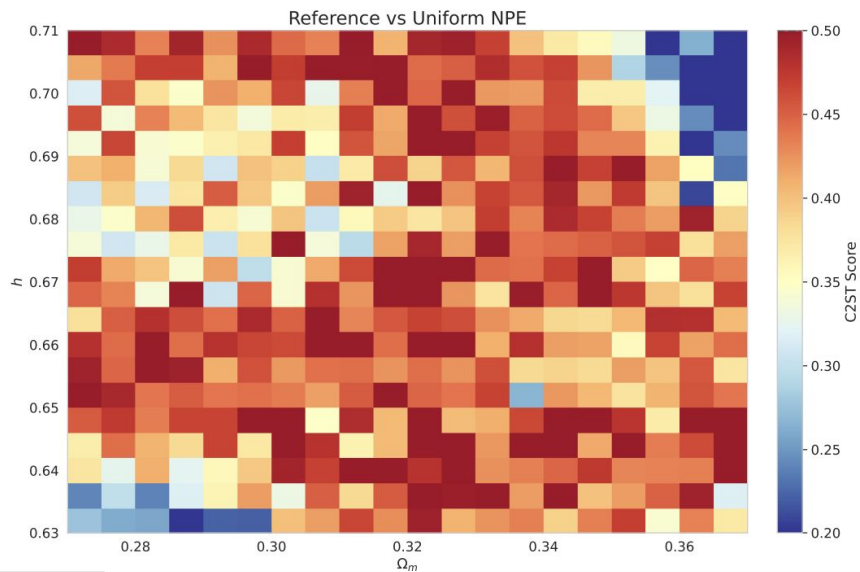


Back to Tailed-Uniform....

Result

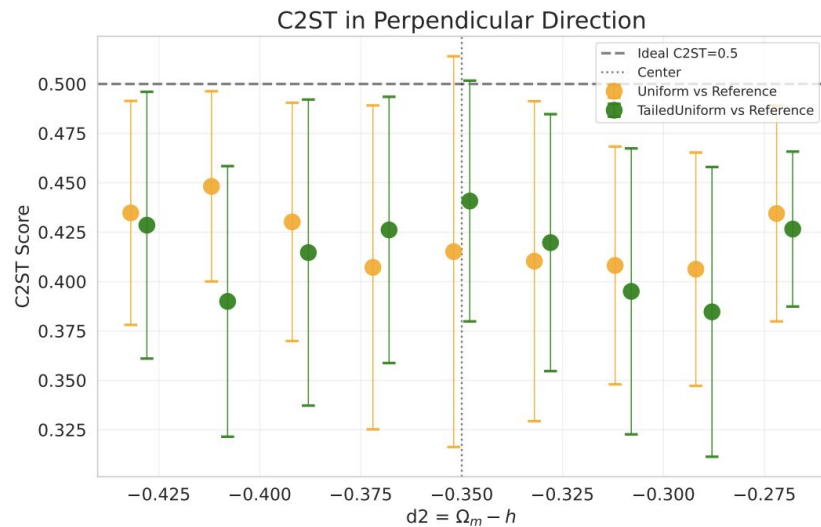
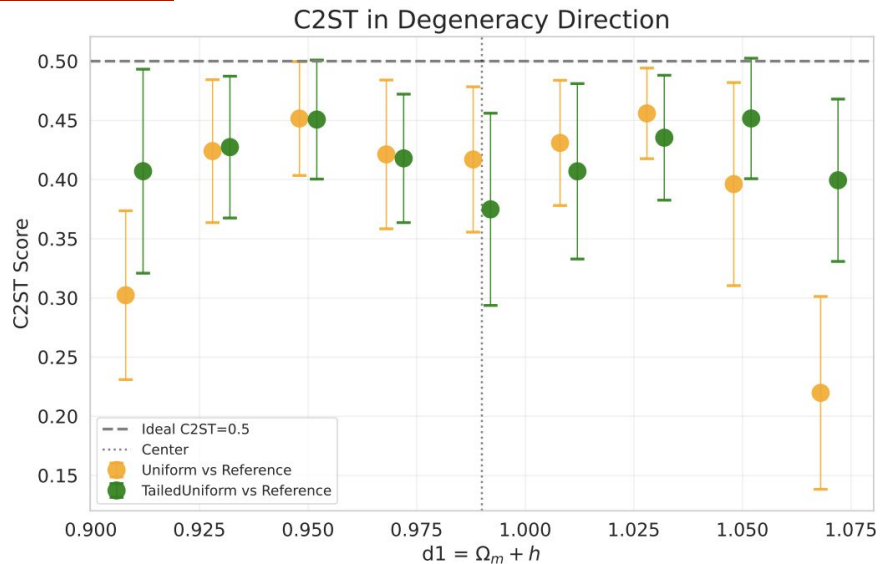
s

Boundary pathology is most severe along degenerate directions where posteriors are elongated into the prior boundaries.



Result

16



- Along the **degeneracy** direction, Uniform C2ST falls to 0.22–0.30 at boundaries; Tailed-Uniform stays robust.
- Along the **perpendicular** direction, posteriors don't elongate into the edges, so both methods perform similarly.

Conclusion

1. We present Tailed-Uniform, a novel parameter sampling technique for simulation-based inference. This hybrid proposal distribution combines Uniform cores with smooth Gaussian tails to sample the edges of our prior boundary.
2. We show that, empirically, neural nets trained with our proposal achieve superior performance near the boundary and maintain consistent posterior quality across the overall parameter space.
3. We also validate Tailed-Uniform on both synthetic benchmarks and realistic cosmological inference tasks, revealing that Tailed-Uniform enables robust inference with competitive simulation budgets.



Our Paper

arxiv.org/abs/2601.17120