Version January 2, 2026
Preprint typeset using LaTeX style openjournal v. 09/06/15

# LEARNING AT THE EDGE: TAILED-UNIFORM SAMPLING FOR ROBUST SIMULATION-BASED INFERENCE

Chaipat Tirapongprasert[1*] and Matthew Ho[1*]
[1] Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA

*Version January 2, 2026*

## ABSTRACT

We introduce the *Tailed-Uniform* proposal distribution for simulation-based inference. Instead of sampling parameters uniformly within bounded regions, we extend the distribution beyond prior boundaries with smooth Gaussian tails. This eliminates sharp discontinuities that cause neural posterior estimators to fail near parameter space boundaries. The method requires minimal hyperparameter tuning, with tail widths of 10–30% of the prior width proving robust across problems. We demonstrate these benefits on a synthetic Gaussian linear task and cosmological parameter inference from the matter power spectrum. We also demonstrate that boundary pathologies are systematic rather than data-starved: adding more uniform samples provides negligible benefit, while *Tailed-Uniform* outperforms uniform sampling even with 8× fewer simulations. This advantage grows in higher dimensions, where boundaries dominate parameter space volume. All code is publicly available at github.com/chaipattira/tailed-uniform-sbi.

*Subject headings:* astrophysics, machine learning, deep learning, simulation-based inference, neural posterior estimation, Bayesian methods

## 1. INTRODUCTION

Today, researchers—among them astronomers—face a growing challenge, as many systems of interest (e.g., exoplanets, black hole binaries, or galaxy clusters) are not amenable to direct probing (Feigelson and Babu 2012; Dodelson and Schmidt 2020). Thus emerges a new class of problems known as "inverse problems," in which the key goal is to identify the model parameters that generated some observed data. The key challenge is that almost all mathematical models of real-world phenomena are complex in that they demand high-dimensional parameter space $\{\boldsymbol{\theta}_i\}$. This makes it prohibitively expensive to compute the probability density for some observation $\mathbf{x}$ (also known as the likelihood $\mathcal{P}(\mathbf{x} \mid \boldsymbol{\theta})$ in the Bayesian parlance).

Simulation-based inference (SBI) has emerged as the gold standard framework for tackling these challenges (Cranmer et al. 2020). Instead of evaluating the explicit likelihood, which has proven to be intractable, we can now leverage model simulations to learn approximate posteriors $\widetilde{\mathcal{P}}(\boldsymbol{\theta} \mid \mathbf{x})$ straight from simulated training dataset $\mathcal{D}_{\text{train}} = \{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}_{i=1}^{N}$. Needless to say, most exciting SBI methods today rely on neural density estimation (Papamakarios and Murray 2016; Greenberg et al. 2019; Lueckmann et al. 2017), which incorporates some flavors of deep neural networks to learn the posterior distribution. Trained on pairs of parameters and simulations, these neural networks learn the mapping between observations and the corresponding parameters that generated them.

Despite these advances, current SBI methodologies face scalability issues for a myriad of reasons. The first and perhaps most prominent is that as the number of parameters increases, so does the computational cost of generating simulations and approximating the posterior. In high dimensions, the volume of parameter space expands so fast that achieving adequate coverage requires an unacceptably large number of simulations.

The second shortcoming lies in the method by which we generate model realizations. Often, the de facto technique involves a uniform sampling of some parameter combinations within some Latin hypercube (McKay et al. 1979)

$$\boldsymbol{\theta} \sim \mathcal{U}([\theta_{\min,1}, \theta_{\max,1}] \times \cdots \times [\theta_{\min,d}, \theta_{\max,d}]), \qquad (1)$$

which has the advantage of covering the whole region of interest. But as the old adage goes, there is no such thing as a free lunch: the training distribution, sampled with this cookie-cutter approach, will exhibit a sharp discontinuity at the boundary $\partial\Theta$, where the density collapses to zero. Driven by stochastic gradient descent, our neural network can interpolate the internal regions at ease but falls short near those boundaries. The absence of adequate support corrodes the resulting posterior estimates. We also believe that such pathology will be exacerbated in higher dimensions by the "curse of dimensionality," (Bellman 1961) according to which the volume fraction of points close to boundaries in $d$-dimension scales as $1 - (1 - \epsilon)^d$.

To address these problems, we present a novel parameter sampling technique for simulation-based inference. In particular,

- we propose *Tailed-Uniform*, a hybrid proposal that combines uniform cores with smooth Gaussian tails.

- empirically, *Tailed-Uniform* achieves superior performance *near* the boundary and maintains consistent posterior quality across the overall parameter space.
- We show that our mathematically motivating scheme allows for robust inference with competitive simulation budgets on representative cosmological inference tasks.

## 2. METHODS

### 2.1. *Simulation-based Inference*

For clarity, we review, very briefly first, the notion of simulation-based inference (SBI) (Cranmer et al. 2020). We define a simulator as a black box function $\mathcal{M}$ that aims to mimic a physical, possibly stochastic, generative process; that is, it maps some parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ to observable quantities $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$.

To generate each data-parameter pair $(\boldsymbol{\theta}_i, \mathbf{x}_i)$, we: (1) sample parameters $\boldsymbol{\theta}_i \sim \widetilde{\mathcal{P}}(\boldsymbol{\theta})$ from the proposal prior, (2) run the forward model $\mathbf{x}_i = \mathcal{M}(\boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_i$ where $\boldsymbol{\epsilon}_i$ represents some systematic noise, and (3) store the resulting data-parameter pairs. The upshot is that we can later feed these into an inverse model to make inferences on some real data $\mathbf{x} = \mathbf{x}_0$.

**Bayesian workflow** The goal of SBI—like any other inference problems—is to discern the posterior distribution, which encodes all the "relevant" information (that is, point estimates and error bars) about the parameters of interest given an observed dataset $\mathbf{x_0}$. In most cases, such posterior is not available in closed form, but because it is contingent upon the provided data, we can invoke Bayes' theorem to write

$$\mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x}_0) \;=\; \frac{\mathcal{P}(\mathbf{x}_0 \mid \boldsymbol{\theta})\, \mathcal{P}(\boldsymbol{\theta})}{\mathcal{P}(\mathbf{x}_0)}. \tag{2}$$

Here, $\mathcal{P}(\mathbf{x}_0 \mid \boldsymbol{\theta})$ is the likelihood (the probability of the data given certain parameters), $\mathcal{P}(\boldsymbol{\theta})$ is the prior (the proxy for our initial belief regarding the true distribution of the parameters), and $\mathcal{P}(\mathbf{x}_0) = \int \mathcal{P}(\mathbf{x}_0 \mid \boldsymbol{\theta})\, \mathcal{P}(\boldsymbol{\theta})\; \mathrm{d}\boldsymbol{\theta}$ is the evidence, which is merely a normalization constant.

**Neural posterior estimation (NPE)** Our objective then is to construct a neural architecture $q_w(\boldsymbol{\theta} \mid \mathbf{x})$ with weights $w$ that outputs a probability distribution over $\boldsymbol{\theta}$ which approximates mapping from observational data to full posterior distribution $\mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x})$ (Papamakarios and Murray 2016; Greenberg et al. 2019). To train this network to accurately represent the conditional probability, we maximize the joint likelihood of our training data $\mathcal{D}_{\text{train}}$. This is akin to minimizing the negative log-probability loss

$$\mathcal{L}_{\text{NPE}} := -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log q_w(\boldsymbol{\theta}|\mathbf{x}) \tag{3}$$

$$= -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log \left[ \frac{\mathcal{P}(\boldsymbol{\theta})}{\tilde{\mathcal{P}}(\boldsymbol{\theta})} q_w(\boldsymbol{\theta}|\mathbf{x}) \right], \tag{4}$$

where the expectation $\mathbb{E}_{\mathcal{D}_{\text{train}}}$ in the loss function is taken over all parameter-observation pairs $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}_{i=1}^N$ in the training dataset.

If such training data exhibits sufficient diversity and the neural architecture possesses adequate coverage of the parameter space $\Theta \subset \mathbb{R}^d$, then minimization of $\mathcal{L}_{\text{NPE}}$ over the network weights $w$ will drive $q_w(\boldsymbol{\theta} \mid \mathbf{x})$ toward the true posterior distribution $\mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x})$ (Hornik et al. 1989). As a result, we can sample at inference time by evaluating the log-density of simulated parameters beforehand during training.

**Priors** It is crucial to make clear the difference between proposal and assumed priors:

- The proposal prior $\tilde{\mathcal{P}}(\boldsymbol{\theta})$, by definition, is the empirical distribution of parameters $\boldsymbol{\theta}$ found in the training dataset $\mathcal{D}_{\text{train}}$. For one thing, it will depend on our sampling strategy (i.e., a Latin hypercube) used to create training simulations.
- On the other hand, our preconceived notion of the global parameter distribution is encoded in the assumed prior $\mathcal{P}(\boldsymbol{\theta})$, which, we should note, is an experimental design choice. It reflects scientific understanding, theoretical constraints, or previous empirical knowledge about plausible parameter values. In cosmological inference—for example—we may use theoretical arguments about structure formation processes to assume that certain parameters follow log-normal distributions.

Broadly speaking, we can choose the assumed and proposed priors to be identical. This alignment, however, is not always preferable. Take a scenario where we sample training simulations from a uniform distribution $\tilde{\mathcal{P}}(\theta) = \mathcal{U}([\theta_{\min}, \theta_{\max}])$ to ensure adequate coverage, while our scientific understanding may instead motivate a Gaussian prior $\mathcal{P}(\theta) = \mathcal{N}(\mu, \Sigma)$. In such a case, the neural network can only learn to assign reasonable probability densities to regions where it has observed training examples, that is, $\tilde{\mathcal{P}}(\theta) > 0$. What about regions that lack training data? Since we do not have any support there, the neural output will tend to zero $q_w(\boldsymbol{\theta}|\mathbf{x}) \approx 0$, leading to poor inference quality near the boundaries.

### 2.2. *The Tailed-Uniform Proposal*

Now for the heart of the matter, we will describe the theoretical foundations of the *Tailed-Uniform*, our suggested proposal that offers a smooth, continuously differentiable density across $\mathbb{R}^d$. We hope that by assigning weights beyond the primary region of interest, we can alleviate the sharp discontinuities at the boundary.

**One-dimension derivation** To kick off the discussion, we consider a one-dimensional ($d = 1$) case and define the probability density function as

$$\tilde{\mathcal{P}}_{\text{TailedUniform}}(x; a, b, \sigma) = \begin{cases} A \cdot \mathcal{N}(a, \sigma^2), & x \leq a \\ B \cdot \mathcal{U}(a, b), & x \in [a, b] \\ A \cdot \mathcal{N}(b, \sigma^2), & x \geq b \end{cases} \tag{5}$$

where $a$ and $b$ define the boundaries of the uniform core region, and $\sigma$ controls the width of the Gaussian tails. We also impose continuity at the boundary to guarantee that our distribution is well-defined, which in turn establishes the values of the normalization constants.

Putting things together, we have

$$A = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2} + (b-a)} \text{ and } B = \frac{b-a}{\sqrt{2\pi\sigma^2} + (b-a)}, \tag{6}$$

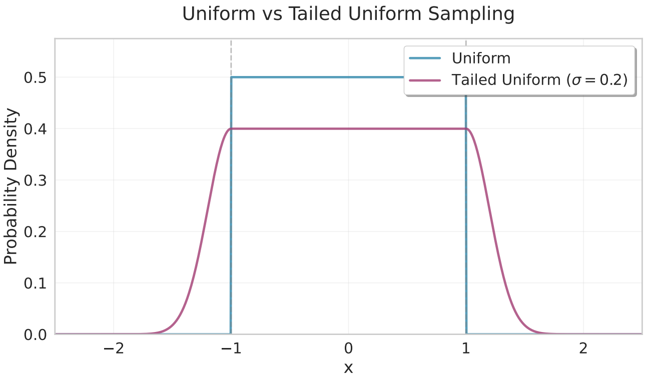

FIG. 1.— The standard uniform distribution (blue) has constant probability density between -1 and 1 with zero probability outside this range. The tailed uniform distribution (magenta) maintains a uniform density in the same central region $[-1, 1]$ but extends beyond these boundaries with Gaussian tails characterized by standard deviation $\sigma = 0.2$.

which ensures that the distribution integrates to unity across the entire domain, and the Gaussian tail matches the uniform core at the left and right boundaries.

**Generalizing to higher dimensions** To generalize to a multivariate parameter space $\boldsymbol{\theta} \in \mathbb{R}^d$, we write it as a product of independent uni-variate *Tailed-Uniform* distributions

$$\tilde{\mathcal{P}}_{\text{TailedUniform}}(\boldsymbol{\theta}; \mathbf{a}, \mathbf{b}, \boldsymbol{\sigma}) = \prod_{i=1}^{d} \tilde{\mathcal{P}}_{\text{TailedUniform}}(\theta_i; a_i, b_i, \sigma_i), \tag{7}$$

where $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$ define the hypercube boundaries, and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ controls the tail-widths in each dimension.

**Algorithm 1:** Tailed-Uniform Sampling

```
Input: Bounds $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, tail widths $\boldsymbol{\sigma} \in \mathbb{R}^d$
Output: Sample $\boldsymbol{\theta} \in \mathbb{R}^d$
for $i = 1$ to $d$ do
    Compute normalization constants $A_i, B_i$;
    Sample $r \sim \mathcal{U}(0, 1)$;
    if $r < A_i$ then
        Sample $z \sim \mathcal{N}(0, \sigma_i^2)$;
        if $z < 0$ then
            $\theta_i \leftarrow z + a_i$ ;          // Left tail
        else
            $\theta_i \leftarrow z + b_i$ ;          // Right tail
        end
    else
        $\theta_i \sim \mathcal{U}(a_i, b_i)$ ;          // Uniform core
    end
end
return $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$;
```

Here, the key hyperparameter is the tail width $\sigma_i$, which controls the balance between boundary smoothness and sampling efficiency. Smaller values of $\sigma_i$ concentrate more samples within the target region $[a_i, b_i]$ but provide less smoothing at the boundaries. Larger values, on the other hand, provide better boundary smoothness but allocate more resources to regions far from the true posterior support. In practice, we recommend setting $\sigma_i$ as about $10 - 40$ percent of the box width; that is,

$$\sigma_i = \alpha \cdot (b_i - a_i), \tag{8}$$

where $\alpha \in [0.1, 0.4]$ (see Section 3.3 for more detail).

## 3. TOY PROBLEM

### 3.1. *Gaussian Linear Task*

We formulate a toy problem using the Gaussian Linear task from the simulation-based inference benchmark (sbibm) (Lueckmann et al. 2021). A Gaussian distribution is ubiquitous in the astrophysical enterprise, as it is—to borrow the jargon of information theory—the *highest entropy* option (Cover and Thomas 2006). One may say that it emerges almost inherently when many independent processes aggregate, per the Central Limit Theorem.

**Model** As a proof-of-concept, we consider a simple, linear, 2-dimensional Gaussian simulator $\mathcal{M} : \mathbb{R}^2 \to \mathbb{R}^2$, which is defined as

$$\mathcal{M}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} = \mathbf{I}_2\boldsymbol{\theta} = \boldsymbol{\theta} \tag{9}$$

with a standard Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ centered in the bounded parameter space $\boldsymbol{\theta} \in [-1, 1]^2$. We hence have a likelihood of the form

$$\mathcal{P}(\mathbf{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\theta}, \mathbf{I}_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\|\mathbf{x} - \boldsymbol{\theta}\|^2\right), \tag{10}$$

which implies that both the likelihood and prior are Gaussian.

**Ground Truth** We can leverage the aforementioned conjugate Gaussian-Gaussian structure (Gelman et al. 2013) to obtain a closed-form analytical posterior $\mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x}) = \mathcal{N}\left(\frac{\mathbf{x}}{2}, \frac{\mathbf{I}_2}{2}\right)$, where the posterior covariance $\frac{\mathbf{I}_2}{2}$ results from $(\mathbf{I}_2^{-1} + \mathbf{I}_2^{-1})^{-1} = \frac{\mathbf{I}_2}{2}$. By running Markov Chain Monte Carlo (MCMC) chains (Brooks et al. 2011), we can generate *reference posterior* samples from which to evaluate performance.

**Training** Our goal is to investigate two distinct strategies for generating the training data (Ho et al. 2024):

1. Baseline (*Uniform* Proposal): Training data is generated by sampling parameters *uniform*ly across the 2-dimensional hypercube $\tilde{\mathcal{P}}_{Uniform} = \mathcal{U}([-1, 1]^2)$.

2. Proposed (*Tailed-Uniform* Proposal): Training data is generated using our hybrid distribution $\tilde{\mathcal{P}}_{\text{TailedUniform}}(\mathbf{a} = [-1, -1], \mathbf{b} = [1, 1], \boldsymbol{\sigma} = [0.2, 0.2])$, where the tail width $\sigma_i = 0.1 \times (b_i - a_i)$ represents 10% of each parameter range.

For both proposals, we generate $N = 4000$ simulation pairs $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}_{i=1}^N$ by: (1) sampling $\boldsymbol{\theta}_i$ from the respective proposal distribution, (2) running the forward simulator $\mathbf{x}_i = \mathcal{M}(\boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_i$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ and (3) training identical neural posterior estimators (NPE) on each dataset. Note again that the assumed prior in both cases is a Gaussian distribution centered at $(0, 0)$. In what follows, we will associate *Uniform* and *Tailed-Uniform* to the NPE trained with these proposals (rather than the proposal itself).

**Model Architectures** We experiment with ensembles of neural density estimators comprising Masked Autoregressive Flows (MAF) (Papamakarios et al. 2017) and Masked Autoencoder for Distribution Estimation

(MADE) (Germain et al. 2015) architectures, each featuring 16 hidden features and 5 coupling layers. Training uses a batch size of 64 and a learning rate of $5 \times 10^{-5}$. Our pipeline is built on the `LtU-ILI` (Learning the Universe—Implicit Likelihood Inference) framework (Ho et al. 2024), which offers standardized data handling and ensemble utilities for simulation-based inference. It is worth pointing out that there is nothing special about the above configuration. The Gaussian Linear task, with its analytically tractable posterior and low dimensionality, is simple enough for any garden-variety neural network to learn the target distribution.

### 3.2. *Validation*

Model validation aims to ascertain whether the approximate posteriors generated by our neural density estimators $q_w(\boldsymbol{\theta} \mid \mathbf{x}_0)$ will be "good enough" when faced with real unlabeled observational data $\mathbf{x}_0$. This is no trivial matter, as we do not a priori know the learned weights in the hidden layers, and the internal workings of neural networks, qua oracles, are more or less invisible.

**The bias-variance tradeoff** One useful proxy for gauging the quality of our learned posterior is the degree to which it can concentrate probability mass around the true parameter $\boldsymbol{\theta}_{\rm true}$. Practically worthless, for instance, will be a model that regurgitates the prior distribution, as it offers no additional insight whatsoever. After all, the goal of inference is to extract as much information as possible from the observed data $\mathbf{x}_0$ to narrow down the possible parameter values $\boldsymbol{\theta}$. Yet, if one only optimizes for the training set, the resulting model might end up overfitting and fail to generalize when exposed to new data.

Achieving the sweet spot between bias and variance (Gelman and Rubin 1992) is thus the characteristic of a "good" model, as we want a model that is versatile enough to extract genuine information from observations (low bias) while remaining generalizable across different realizations of the data (low variance).

TABLE 1
POSTERIOR PERFORMANCE FOR $\boldsymbol{\theta}_{\rm TRUE} = (0.6, 0.6)$.

| **Method** | **Parameter Estimates** | | **C2ST** |
|---|---|---|---|
| | $\theta_1$ | $\theta_2$ | |
| Reference | $0.446 \pm 0.300$ | $0.641 \pm 0.308$ | 0.500 |
| *TailedUniform* | $0.339 \pm 0.547$ | $0.708 \pm 0.286$ | **0.458** |
| *Uniform* | $0.259 \pm 0.446$ | $0.588 \pm 0.272$ | 0.409 |
| **Improvement** | **+31%** | **+10%** | **+12%** |

#### 3.2.1. *Tier One: Single-Point Analysis*

As a first pass, we select test points $\boldsymbol{\theta}_{\rm true}$ near the boundaries of parameter space, namely $\boldsymbol{\theta}_{\rm true} = (0.6, 0.6)$ (i.e., $3\sigma$ away from the center of the prior) and draw $M = 1000$ independent and identically distributed samples from the learned posterior. The goal is to assess their constraining power by investigating the distribution and shape around the true value.

**Quantitative Metrics** We also use the Classifier two-sample tests (abbreviated C2ST) (Lopez-Paz and Oquab 2017), which is recognized as one of the best interpretable metrics for comparing two distributions. It trains a logistic regression classifier to differentiate between samples from two distributions: a score of 0.5 indicates identical distributions (desirable), whereas diverging scores indicate easily distinguishable distributions (poor performance).

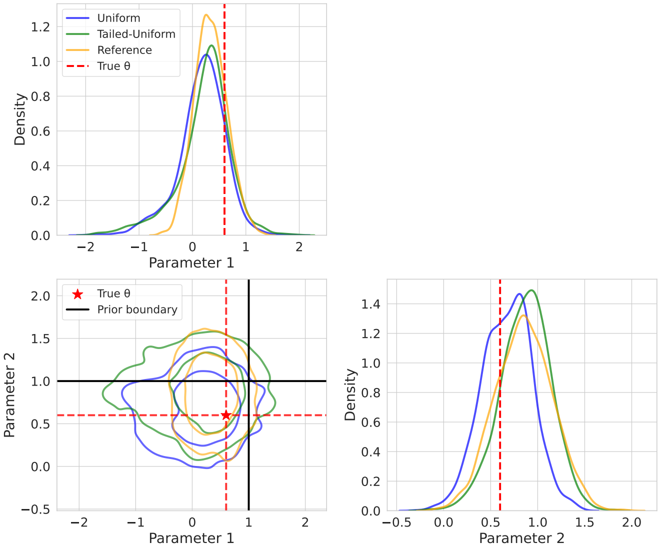

FIG. 2.— Corner plots comparing posterior estimation performance for the boundary test case. The *Tailed-Uniform* (green) demonstrates superior boundary behavior compared to the *Uniform* (blue), closely matching the MCMC reference (yellow). The red dashed line indicates the true parameter value.

As anticipated in Figure 2, the *Tailed-Uniform* (orange) keeps a tight concentration around the true parameter values, while the *Uniform* (blue) leaks into unusable parameter regions. Table 1 further corroborates *Tailed-Uniform*'s superior performance across all metrics over the *Uniform* baseline.

#### 3.2.2. *Tier Two: Spatial Performance Analysis*

Next we need a way to assess *Tailed-Uniform*'s performance across the entire parameter space. And so, we discretize a uniform rectangular grid with $n = 20$ test locations along each dimension of the parameter space $[-1, 1]^2$, yielding a total of 400 evaluation points. We want to exhaustively sample areas of varying distances, ranging from the center of the priors, where traditional techniques work well, to points near the boundaries. As before, we generate observations for each test point, draw $M = 1000$ posterior samples from all three methods, and calculate pairwise C2ST scores at each test location.

Figure 3 shows how *Uniform*'s performance deteriorates as it drifts away from the central regions, as evidenced by the proliferation of blue patches near the boundary. In contrast, our *Tailed-Uniform* maintains more red and orange coloration (with C2ST scores around 0.44–0.49) across the parameter space.

Furthermore, we observe that the resulting posterior is spherically symmetric (with equal variance in all directions), allowing us to visualize the improvements by binning the test points as a radial function from the center of our parameter space, as seen in Figure 4. It is, if nothing else, evident that *Tailed-Uniform* exhibits robustness and mitigates the undesired boundary pathology that

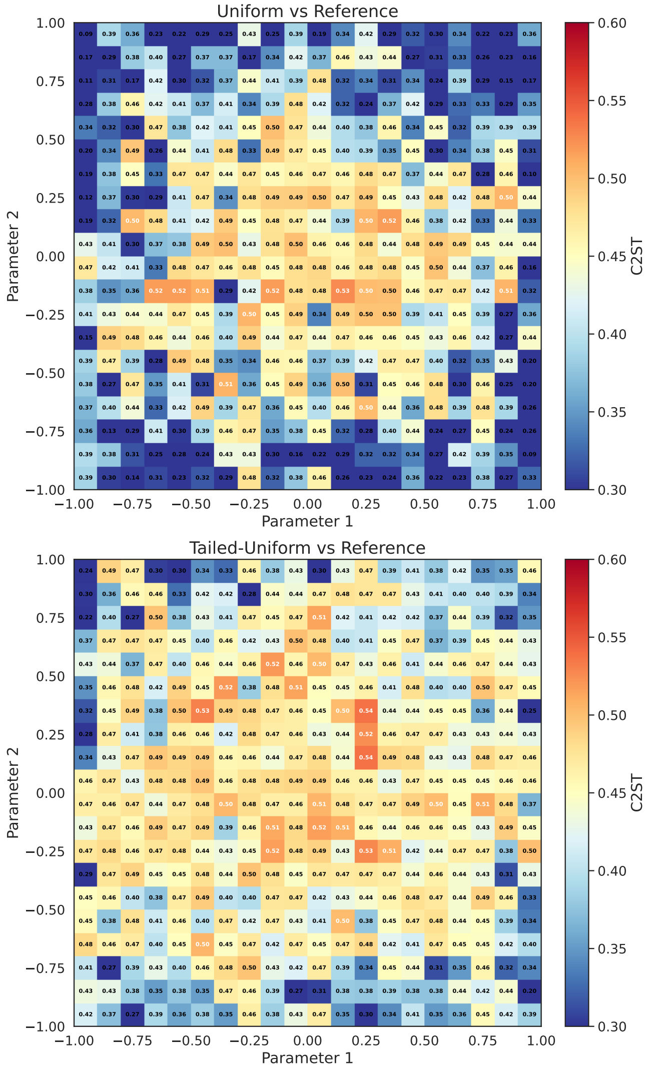

FIG. 3.— Spatial distribution of C2ST performance across the parameter space, with blue regions indicating poor distributional matching (C2ST $\ll 0.5$) and red/orange regions indicating good performance (C2ST $\approx 0.5$) **Top**: *Uniform* versus analytical reference, showing systematic boundary degradation with glaring blue regions near parameter space edges. **Bottom**: *TailedUniform* versus reference, demonstrating consistent performance across the entire parameter space.

characterizes the traditional sampling technique.

### 3.3. *Discussion*

Our subsequent evaluations aim to answer the following operational questions:

1. How large should the tails be?
2. Could we mitigate the boundary pathology by throwing more data at the NPE?
3. How does *Tailed-Uniform* perform in higher dimensions?

#### 3.3.1. *Ablation on the tail widths*

The tail width $\sigma$ represents a bias-variance tradeoff: narrower tails recover the boundary pathologies, while wider tails allocate more simulations outside the primary region of interest. To assess the importance of the tails, we train one *Uniform* baseline and several *Tailed-Uniform* models, each with different $\sigma$ values as fractions of prior width: $\alpha \in \{0.01, 0.05, 0.1, 0.2, 0.4\}$.

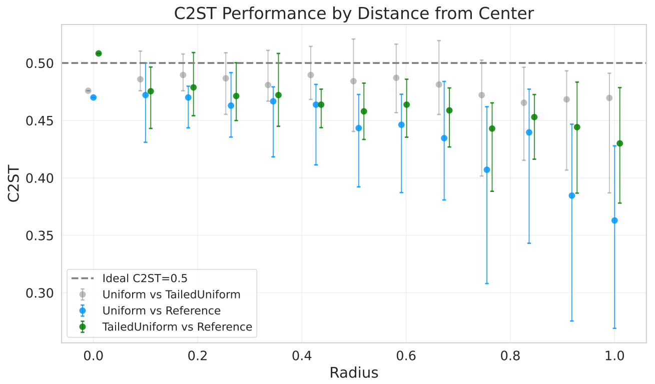

FIG. 4.— C2ST performance degradation as a function of distance from parameter space center. The blue curve reveals systematic deterioration of the Uniform NPE method near boundaries, while the green curve demonstrates that TailedUniform NPE maintains consistent performance across all radii. The gray curve quantifies the increasing divergence between methods, with boundary regions showing substantial differences in posterior approximation quality. Values closer to 0.5 indicate better distributional matching.

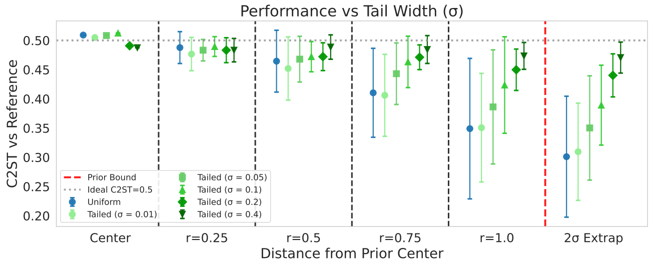

FIG. 5.— C2ST performance versus distance from center, stratified by tail width $\sigma$. The *Uniform* baseline (blue) exhibits systematic boundary degradation. The *Tailed-Uniform*'s with varying tail widths (shown in green; lighter shades for smaller $\sigma$ values) exhibit robustness across a broad range of tail widths. Notably, performance becomes quite good for $\sigma \geq 0.2$

Figure 5 reveals that performance improves monotonically with tail width. At the prior boundary (r=1.0), C2ST scores improve from 0.331 ($\sigma = 0.01$, equivalent to Uniform) to 0.465 ($\sigma = 0.4$, near-ideal), demonstrating the value of wider tails. The reassuring robustness indicates that even moderate tail samples suffice for learning smooth density transitions, which will prove insightful in higher dimensions. It also implies that practitioners do not need to hyperparameter tune $\sigma$, as any reasonable choice within this range will suffice. At $\sigma = 0.01$, *Tailed-Uniform* degenerates as anticipated toward *Uniform* behavior with characteristic boundary degradation, confirming that the tails themselves (not some arbitrary artifacts) are responsible for the observed improvements.

#### 3.3.2. *Ablation on the number of simulations*

If insufficient data in the near-boundary regions causes degradation, just increasing the training set size could very well solve the problem. After all, neural networks are universal function approximators, so given enough training data, they should be able to learn com-

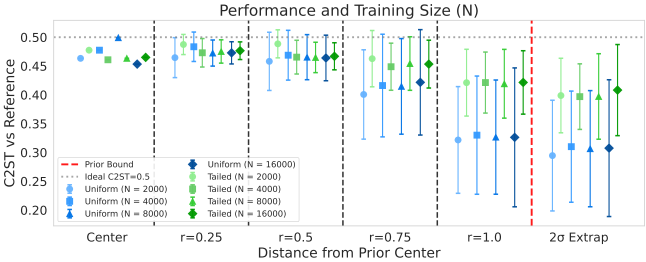

FIG. 6.— C2ST performance versus distance from parameter space center, stratified by training set size $N$ (circles: 2000, squares: 4000, triangles: 8000, diamonds: 16000). Increasing $N$ provides minimal benefit for both proposals.

plex mappings, including sharp transitions at parameter space boundaries. To interrogate this hypothesis, we test an ablation over the training set size, varying $N \in \{2000, 4000, 8000, 16000\}$ while holding all other hyperparameters fixed.

Empirically, we find in Figure 6 that increasing the number of simulations barely helps with the NPE's performance, suggesting that the boundary degradation stems not from absence of data, but from an inbuilt discontinuous mapping that cannot be learned from samples within the prior alone. Since no amount of supplementary data can smooth a non-smooth function, the hard truncation at prior boundaries poses a fundamental learning barrier. In contrast, we see that every single *Tailed-Uniform* ($N = 2000$) outperforms even the largest *Uniform* ($N = 16000$) with 8 times more training data, demonstrating the impact of different proposals. One can achieve better posterior quality at a fraction of the computational cost just by sampling in a judicious manner.

### 3.3.3. *Scaling to higher dimensions*

While we have shown *Tailed-Uniform* to be effective in two dimensions, a natural question arises: does this benefit transfer to higher-dimensional parameter spaces, which are necessary for the majority of astrophysical inference problems? This is indeed a critical concern. By the curse of dimensionality, high-dimensional spaces become ever more sparse, as the volume of a $d$-dimensional hypercube grows exponentially while the number of samples remains fixed. Corners, edges, and boundary regions come to dominate the landscape.

Having proffered stakes, let's examine the probability of obtaining samples near boundaries. For a $d$-dimensional hypercube $[0, 1]^d$, we define the interior region as the set of points at least a distance $\varepsilon$ away from any boundary, which occupies a fraction $P_{\text{int}}^{\text{Uniform}}(d, \varepsilon) = (1-2\varepsilon)^d$ of the total volume. Complementarily, the probability that a uniformly sampled point lies in the boundary shell (within $\varepsilon$ of any face) grows exponentially as $P_{\text{bdry}}^{\text{Uniform}}(d, \varepsilon) = 1-(1-2\varepsilon)^d$, which approaches unity at high dimensions.

**The Tail Allocation Budget** For *Tailed-Uniform*, samples extend beyond the hypercube boundaries into tail regions. With independent marginals, the probability that a sample lands inside the $d$-dimensional hypercube is the product

$$P_{\text{cube}}^{\text{TailedUniform}}(d) = \prod_{i=1}^{d} B_i = B^d = \left( \frac{1}{1 + \alpha\sqrt{2\pi}} \right)^d, \tag{11}$$

assuming identical $\alpha$ across dimensions. This means the probability of sampling in the tail regions grows as $P_{\text{tail}}^{\text{TailedUniform}}(d) = 1 - B^d$. By the same token, the probability of landing in the interior becomes $P_{\text{int}}^{\text{TailedUniform}}(d, \varepsilon) = B^d \cdot (1-2\varepsilon)^d$, while the boundary shell captures $P_{\text{bdry}}^{\text{TailedUniform}}(d, \varepsilon) = B^d \left[1 - (1-2\varepsilon)^d\right]$. Table 2 summarizes these quantities for representative dimensions.

TABLE 2
SAMPLING BUDGET ACROSS REGIONS FOR *Uniform* AND *Tailed-Uniform* PROPOSALS ($\varepsilon = 0.05$, $\alpha = 0.1$).

| | **Uniform** | | | **Tailed-Uniform** | | |
|---|---|---|---|---|---|---|
| $d$ | $P_{\text{int}}$ | $P_{\text{bdry}}$ | $P_{\text{tail}}$ | $P_{\text{int}}$ | $P_{\text{bdry}}$ | $P_{\text{tail}}$ |
| 2 | 0.810 | 0.190 | 0.000 | 0.617 | 0.145 | 0.238 |
| 4 | 0.656 | 0.344 | 0.000 | 0.381 | 0.200 | 0.419 |
| 8 | 0.430 | 0.570 | 0.000 | 0.145 | 0.192 | 0.663 |
| 16 | 0.185 | 0.815 | 0.000 | 0.021 | 0.115 | 0.864 |

At first glance, it might seem that by bleeding samples into the tails, too many samples will wind up outside the region of interest, thus diminishing the overall performance of *Tailed-Uniform*. To refute this suspicion, we evaluate NPE trained with *Uniform* and *Tailed-Uniform* proposals on a Gaussian linear inference task across $d \in \{4, 8, 16\}$ dimensions.

**Boundary Dominance in Higher Dimensions:** Figure 7 shows that while both methods degrade with increasing dimensionality—an expected consequence of the aforementioned "curse of dimensionality"—*Tailed-Uniform* consistently outperforms *Uniform*, with the performance gap widening at higher dimensions. This somewhat paradoxical finding suggests that the ratio of information gained from exterior support to cost of interior sample dilution remains favorable or even improves with dimension. Extra samples from *Tailed-Uniform* flow into the tail regions, surrounding the boundaries and offering regularization. Indeed, at $d = 16$, boundaries are so dominant that the marginal value of additional interior samples is low, as the network has ample interior coverage even when 99.99% of *Tailed-Uniform* samples lie outside the region of interest. Meanwhile, the marginal value of tail samples is high, as they are the only source of information near the boundary, which explains why *Tailed-Uniform* still achieves slightly better C2ST scores than *Uniform* at the boundary and in the extrapolation regime ($2\sigma$ beyond the prior boundary). We acknowledge that at very high dimensions, dilution effects will eventually dominate, but these regimes are beyond the scope of most practical simulation-based inference tasks.

## 4. SCIENCE EXPERIMENT

### 4.1. *Matter Power Spectrum*

We next demonstrate the versatility of the *Tailed-Uniform* proposal in a popular inference benchmark in cosmology: the inference of cosmological parameters from the matter power spectrum (Dodelson and Schmidt 2020). To recreate the problem, we aim to predict the

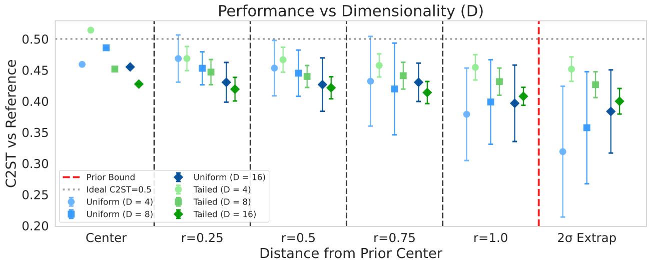

FIG. 7.— C2ST performance versus distance from parameter space center in different representative dimensions. *Tailed-Uniform* maintains superior performance with the advantage growing at higher dimensions, particularly near boundaries ($r \geq 0.75$) and in the extrapolation regime.

posterior distribution of two cosmological parameters $\boldsymbol{\theta} = (\Omega_m, h)$ using the forward simulator $\mathcal{M} : \mathbb{R}^2 \to \mathbb{R}^{64}$ that maps a two-dimensional parameter vector to the power spectrum $\mathbf{P} \in \mathbb{R}^{64}$. Here, $\Omega_m$ is the matter density parameter that governs the growth rate of perturbations and $h$ is the dimensionless Hubble parameter, which determines the physical scale of the horizon at matter-radiation equality.

In modern cosmology, the power spectrum $\mathbf{P}(k)$ measures the amplitude of density contrast fields $\delta(\mathbf{k})$ in Fourier space and is strictly dependent on the magnitude of the wave vector $k = |\mathbf{k}|$ (Dodelson and Schmidt 2020). This can be thought of as the ensemble average of all universes with the same cosmological parameters but different initial fluctuations. We also assume a flat $\Lambda$CDM cosmology with $w_0 = -1$, $w_a = 0$, and negligible neutrino masses $m_\nu = 0$, and fix the remaining cosmological parameters to Planck 2018 (Planck Collaboration et al. 2020) values (the primordial scalar amplitude $A_s = 2.105 \times 10^{-9}$, the baryon density parameter $\Omega_b h^2 = 0.02242$, and the scalar spectral index $n_s = 0.9665$).

**Model** To circumvent the need to solve the perturbation equations numerically, we use the `syren-new` emulator (Sui et al. 2024), which leverages symbolic regression to produce a closed functional form for the theoretical power spectrum $\mathbf{P}_{\text{theory}}$ at relatively negligible compute power. The power spectrum from `syren-new` is not the observed power spectrum per se, as our universe (armed with specific initial conditions) is merely one realization of some stochastic process. If we observe a sufficiently large volume—however—different regions within that volume will serve as independent samples of the same underlying process (Baumann 2022). Such a system with this nice property is called "ergodic." But since galaxy surveys cannot cover an infinite volume, the spatial average will not be *exactly* equal to the ensemble average.

For a comoving survey volume of $V = L^3$, the longest wavelength that fits in the box (i.e., the fundamental mode) is $k_f = |\mathbf{k}| = \frac{2\pi}{L}$, and by assuming the universe to be homogeneous and isotropic (i.e., rotationally invariant) on large scales (Dodelson and Schmidt 2020), we then find the number of modes inside a spherical shell at wavenumber $k$ with thickness $\Delta k = k_f$ to be

$$N_k = \frac{4\pi k^2 \Delta k}{k_f^3} = \frac{L^3 k^2 k_f}{2\pi^2}. \tag{12}$$

This is called "cosmic variance," which arises from the limited number of independent modes available in our finite observation volume (Dodelson and Schmidt 2020; Baumann 2022).

Together, we can construct an observation as

$$\mathbf{x}_i = \mathbf{P}(k \mid \boldsymbol{\theta}_i) = \underbrace{\mathbf{P}_{\text{theory}}(k \mid \boldsymbol{\theta}_i)}_{\texttt{syren\_new}\text{ emulator}} + \underbrace{\boldsymbol{\epsilon}_i}_{\text{cosmic noise}}, \tag{13}$$

where $\boldsymbol{\epsilon}_i \sim \mathcal{N}\left(\mathbf{0}, \text{diag}(\sigma_1^2, \ldots, \sigma_{N_k}^2)\right)$ represents the heteroskedastic (scale-dependent) cosmic noise with a variance $\sigma_k^2 = \frac{2\mathbf{P}^2_{\text{theory}}(k|\boldsymbol{\theta}_i)}{N_k}$.

**Ground Truth** As in Section 3, we have the luxury of direct sampling via MCMC (Brooks et al. 2011), as the forward model is explicit and the joint likelihood $\mathcal{P}(\mathbf{x} \mid \boldsymbol{\theta})$ can be evaluated point-wise. We choose a truncated Gaussian prior centered at the midpoint of $\Omega_m \in [0.24, 0.40]$ and $h \in [0.61, 0.73]$. Assuming independent Gaussian noise in each $k$-bin, we can write the log-likelihood as

$$\log \mathcal{P}(\mathbf{x} \mid \boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mathbf{x}_i - P_{\text{theory},i}(\boldsymbol{\theta})}{\sigma_i(\boldsymbol{\theta})}\right)^2 - \sum_{i=1}^{N}\log\sigma_i(\boldsymbol{\theta}), \tag{14}$$

where the first term quantifies the chi-squared misfit between theory and observation, and the second term accounts for the uncorrelated nature of the noise. Putting things together, the log posterior we aim to sample is $\log \mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x} = \mathbf{x}_0) \propto \log \mathcal{P}(\mathbf{x}_0 \mid \boldsymbol{\theta}) + \log \mathcal{P}(\boldsymbol{\theta})$.

To explore this posterior distribution, we run 4 MCMC walkers via the affine-invariant ensemble sampler in the `emcee` package (Foreman-Mackey et al. 2013), each generating 2000 samples after a burn-in period of 500 steps to ensure convergence. In total, we collect 4 × 2000 = 8000 posterior samples, which we treat as our ground truth.

**Training and Model Architectures** In this experiment, we choose a comoving box size of $L = 1000, h^{-1}$Mpc and measure the power spectrum at $N = 64$ logarithmically spaced wavenumbers spanning from the fundamental mode $k_f = 2\pi/L$ to the Nyquist frequency $k_{\text{nyq}} = \pi N/L$, evaluated at the present epoch ($a = 1.0$). Recall again that our raison d'être is to compare the performance of neural networks trained with two distinct proposals $\tilde{\mathcal{P}}_{\text{Uniform}}$ and $\tilde{\mathcal{P}}_{\text{TailedUniform}}$, which by the relative complexity of the problem, will most likely be affected by our architectural choices. One could now, like a fervent flâneur, proceed by trial and error: training dozens of models with different configurations and choosing the ones with the best log-probability. But in an effort to avoid unnecessary angst, we choose to perform Bayesian optimization via the `Optuna` framework (Akiba et al. 2019) with the search space consisting of the flow architecture, number of hidden features, number of transform layers, batch size, and learning rate. Once the hyperparameter search concludes, we rank all trained networks by their validation log-probability and ensemble the top 10 performing models.

### 4.2. *Validation*

To assess the performance of our competing proposals in this cosmological inference task, we adopt the same

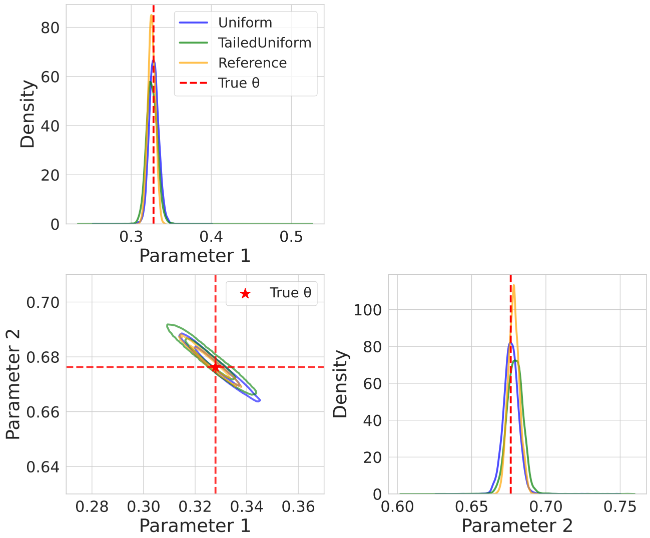

FIG. 8.— Corner plots comparing posterior estimation performance for test point $\boldsymbol{\theta}_{\rm true} = (0.328, 0.676)$ near the parameter space boundary. The *TailedUniform* (green) closely matches the MCMC reference (yellow), maintaining tight concentration along the banana-shaped posterior. In contrast, the *Uniform* (blue) exhibits excessive spread along the degenerate direction, leaking probability mass into regions inconsistent with the data. The characteristic anti-correlation between $\Omega_m$ and $h$ is visible in all methods, but only *TailedUniform* accurately captures the posterior near the boundary.

spatial evaluation framework as in Section 3. We should first discuss why $\Omega_m$ and $h$ are given special consideration, which led us to select them as the inference parameter. There is a variety of factors, the most notable of which is the strong "degeneracies" between them.

**Degeneracies** Degeneracy refers to situations in which multiple parameters replicate the same observable. In our case, the matter power spectrum at small and intermediate scales is sensitive to the physical matter density, which happens to scale as $\Omega_m h$, rather than by $\Omega_m$ and $h$ separately (Peacock and Dodds 1992). The two parameters are thereby anti-correlated.

As a result, the information embedded in our posterior will be weak along certain directions in the parameter space—that is, where $\Omega_m$ trades off against $h$ such that $\Omega_m h$ remains approximately constant—leading to the slanted banana-shaped contours (as also reflected in the posteriors' performance in Figure 9). Such geometric degeneracies in the power spectrum from different combinations of $\Omega_m$ and $h$, combined with the likelihood being painstakingly non-linear, make this a challenging inference problem.

### 4.2.1. *Tier One: Single-Point Analysis*

To study this phenomenon in detail, we first select a test point $\boldsymbol{\theta} = (0.328, 0.676)$ positioned 40% from the top-right corner. We generate observations and draw $M = 1000$ posterior samples from the MCMC reference, *Uniform*, and *Tailed-Uniform*.

Figure 8 shows that *Uniform* has notable dispersion along the degenerate direction. This leakage shows that when observations overlap boundary regions of the training distribution, *Uniform* has learned an excessively broad posterior and is unable to appropriately constrain parameters. The *Tailed-Uniform*, on the other hand, reproduces the banana-shaped contours of the MCMC reference, with samples tightly concentrated around the true parameter values. Table 3 quantifies these results.

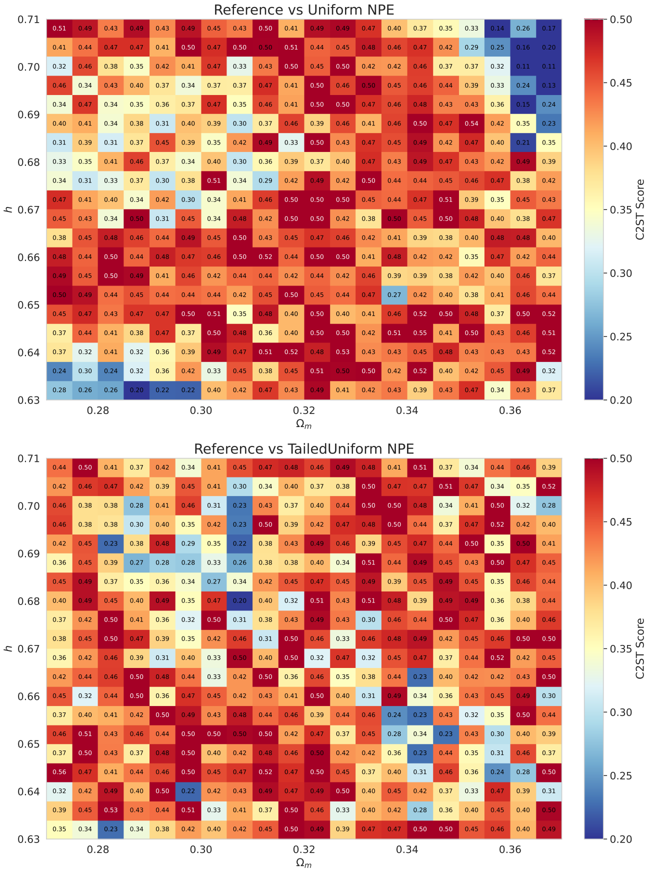

FIG. 9.— Spatial distribution of C2ST performance across the $(\Omega_m, h)$ parameter space, with darker blue regions indicating poor distributional matching (C2ST $\ll 0.5$) and orange-red regions indicating good performance (C2ST $\approx 0.5$). **Top**: *Uniform* versus MCMC reference, showing systematic degradation with blue patches near parameter space edges and corners. **Bottom**: *TailedUniform* NPE versus reference, demonstrating consistent orange coloration (C2ST $\approx$ 0.41–0.45) across the parameter space.

TABLE 3

POSTERIOR PERFORMANCE FOR TEST POINT $\boldsymbol{\theta}_{\rm TRUE} = (0.328, 0.676)$

| **Method** | **Parameter Estimates** | | **C2ST** |
|---|---|---|---|
| | $\Omega_m$ | $h$ | |
| Reference | $0.326 \pm 0.005$ | $0.678 \pm 0.004$ | 0.500 |
| *Tailed-Uniform* | $0.326 \pm 0.008$ | $0.679 \pm 0.006$ | **0.464** |
| *Uniform* | $0.329 \pm 0.006$ | $0.676 \pm 0.005$ | 0.395 |
| **Improvement** | – | – | **+17%** |

### 4.2.2. *Tier Two: Spatial Performance Analysis*

As before, we discretize a uniform rectangular grid with $n = 20$ test locations along each dimension, spanning $\Omega_m \in [0.27, 0.37]$ and $h \in [0.63, 0.71]$. For each of the 400 evaluation points, we generate observations, draw $M = 1000$ posterior samples from all three methods, and calculate pairwise C2ST scores.

Both methods achieve similar C2ST scores in Figure 9, but by eye inspection, *TailedUniform* appears to provide reliable inference everywhere, whereas *Uniform* is only

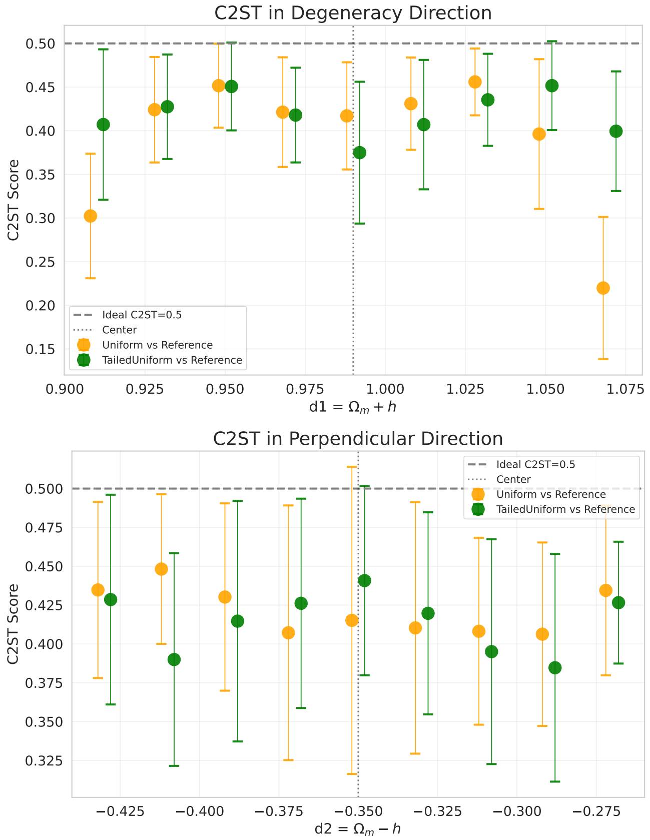

FIG. 10.— C2ST performance as a function of position along two orthogonal directions in parameter space. **Top**: Along the degeneracy direction, the *Uniform* (yellow) exhibits performance degradation at the boundaries. The *TailedUniform* (green) maintains robust performance throughout. **Bottom**: Along the perpendicular direction, both methods perform similarly.

productive near the center. To connect the performance degradation trend to the underlying physics, we propose a different binning strategy. We first define the degeneracy direction as $d_1 = \Omega_m + h$, which describes movement along the banana-shaped posterior where the physical matter density $\Omega_m h$ varies slowly. Perpendicular to this, we define $d_2 = \Omega_m - h$, which captures variation across the degeneracy.

Figure 10 shows that boundary pathology is anisotropic, with the most severe manifestation along degenerate directions where posteriors are elongated into the prior boundaries. As a result, both approaches perform similarly (excluding statistical fluctuations) along the perpendicular direction $d_2$. However, along the degeneracy direction $d_1$, the *Uniform* performance sinks at both boundaries, with C2ST scores falling to 0.22–0.30. The *Uniform* is unable to learn the posterior tails because the elongated banana-shaped posterior has spread into and beyond areas with sparse training coverage. By extending the simulation distribution beyond the prior boundaries, the *TailedUniform* enables robust, unbiased inference across the full parameter space.

## 5. CONCLUSION

We identified and resolved a fundamental limitation of neural posterior estimation: boundary pathology caused by sharp discontinuities in uniform proposal distributions, which is exacerbated in high-dimensional spaces and when parameter degeneracies extend posteriors toward prior boundaries. To address this issue, we proposed the *Tailed-Uniform* proposal, which makes use of smooth, continuously differentiable Gaussian tails to smooth the abrupt boundaries. We validated the *Tailed-Uniform* proposal on both synthetic and real-world inference problems, demonstrating substantial improvements over traditional uniform sampling. Perhaps most critically, our ablation studies revealed that the boundary pathology is fundamental rather than a consequence of insufficient training data. Our method also exhibited counterintuitive scaling behavior in higher dimensions: despite allocating increasingly more samples to tail regions, *Tailed-Uniform* maintains good performance at a reasonable number of dimensions. By extending the training distribution beyond the primary region of interest, our method provides neural networks with the support they require to learn posterior approximations near the boundaries, improving the quality of the posterior across the parameter space.

We appreciate Shy Genel's insightful comments on a previous draft. This project was developed as part of the Simons Collaboration on "Learning the Universe."

## REFERENCES

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019.

Daniel Baumann. *Cosmology.* Cambridge University Press, Cambridge, 2022. ISBN 978-1108838078.

Richard Bellman. Adaptive control processes. *Princeton University Press*, 1961.

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo.* CRC Press, 2011.

Thomas M Cover and Joy A Thomas. *Elements of information theory.* John Wiley & Sons, 2nd edition, 2006.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Scott Dodelson and Fabian Schmidt. *Modern Cosmology.* Academic Press, 2nd edition, 2020.

Eric D Feigelson and Jogesh G Babu. Modern statistical methods for astronomy: with R applications. *Cambridge University Press*, 2012.

Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4): 457–472, 1992.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis.* CRC Press, 3rd edition, 2013.

Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. MADE: Masked autoencoder for distribution estimation. In *International Conference on Machine Learning*, pages 881–889. PMLR, 2015.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019.

Matthew Ho, Gwendolyn M Farren, Helen Shao, Natalí Anau Montel, Chirag Modi, Bruno Régaldo-Saint Blancard, Pablo Lemos, Adam Coogan, Yashar Hezaveh, Laurence Perreault-Levasseur, Benjamin D Wandelt, and Natalí SM de Santi. Ltu-ili: An all-in-one framework for implicit inference in astrophysics and cosmology. *The Open Journal of Astrophysics*, 7, 2024. arXiv:2402.05137.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. In *International Conference on Learning Representations*, 2017. arXiv:1610.06545.

Jan-Matthis Lueckmann, Pedro J Gonçalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in Neural Information Processing Systems*, 30, 2017.

Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro J Gonçalves, and Jakob H Macke. Benchmarking simulation-based inference. In *International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, 2021.

M D McKay, R J Beckman, and W J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21 (2):239–245, 1979.

George Papamakarios and Iain Murray. Fast $\epsilon$-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, volume 29, pages 1028–1036, 2016.

George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in Neural Information Processing Systems*, 30, 2017.

J. A. Peacock and S. J. Dodds. Non-linear evolution of cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 258(1):1P–11P, sep 1992. doi:10.1093/mnras/258.1.1P.

Planck Collaboration, N Aghanim, Y Akrami, et al. Planck 2018 results. vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020. arXiv:1807.06209.

Ce Sui, Deaglan J Bartlett, Shivam Pandey, Harry Desmond, Pedro G Ferreira, and Benjamin D Wandelt. Syren-new: Precise formulae for the linear and nonlinear matter power spectra with massive neutrinos and dynamical dark energy. *arXiv preprint arXiv:2410.14623*, 2024.

This paper was built using the Open Journal of Astrophysics LaTeX template. The OJA is a journal which provides fast and easy peer review for new papers in the astro-ph section of the arXiv, making the reviewing process simpler for authors and referees alike. Learn more at http://astro.theoj.org.