# Learning at the Edge: Tailed-Uniform Sampling for Robust Simulation-Based Inference

**Chaipat Tirapongprasert**[1*] **Matthew Ho**[1]
[1]Department of Astronomy, Columbia University, New York, NY, 10027

## Abstract

We introduce the TAILED-UNIFORM proposal distribution for generating training simulations in 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 transitions that cause neural posterior estimators to fail when the posterior distribution intersects or extends beyond the prior bounds. We show these benefits on a toy problem and cosmological parameter inference from the matter power spectrum. Such an advantage grows in high dimensions, where boundaries dominate parameter space volume.

## 1 Introduction

Modern astrophysics increasingly relies on complex forward simulations to model non-linear phenomena, ranging from the formation of large-scale structure [Pakmor et al., 2023] to binary black hole evolution [Siwek et al., 2023]. As the fidelity of these physical models improves, so does the dimensionality of their parameter space, creating a significant bottleneck for inverse problems where the goal is to recover governing parameters from observations. Traditionally, parameter estimation relied on Markov Chain Monte Carlo [MCMC; Metropolis et al., 1953, Hastings, 1970], which necessitates the computation of an explicit, closed-form likelihood. As the complexity and dimensionality of our models grow, this is typically not possible or too costly [Pretorius, 2005].

Simulation-based inference [SBI; Cranmer et al., 2020] has emerged as the standard framework for tackling intractable likelihoods and is now a cornerstone methodology in physics, spanning cosmology, gravitational-wave astronomy, and particle physics. Most SBI methods rely on neural density estimation [Papamakarios and Murray, 2018, Greenberg et al., 2019, Lueckmann et al., 2017]; we focus on Neural Posterior Estimation (NPE), which trains on simulated $(\boldsymbol{\theta}, \mathbf{x})$ pairs to directly approximate the posterior and amortizes the simulation cost so that one trained network can infer posteriors for any new observation.

However, the fidelity of the amortized posterior depends on having a diverse and representative training set, which begins with the choice of proposal distribution from which simulation parameters are drawn. Often, simulators preparing datasets for large-scale public campaigns, such as the CAMELS [Villaescusa-Navarro et al., 2021], choose to simulate at a uniform distribution of points in parameter space, typically via a Latin Hypercube bounded within a specific region of interest [McKay et al., 2000]. While efficient for volume coverage, this strategy introduces a sharp discontinuity in the probability density at the prior boundaries. Neural density estimators, driven by stochastic gradient descent, struggle to model this artificial step-function [Cornish et al., 2020]. Consequently, the estimators often fail to capture the posterior accurately near the edges of the parameter space. This boundary pathology is exacerbated in high-dimensional settings by the curse of dimensionality [Bellman and Kalaba, 2003], where the volume of the parameter space is increasingly dominated by the boundary shell rather than the interior.

To address these problems, we present TAILED-UNIFORM, a hybrid proposal distribution that combines UNIFORM cores with smooth Gaussian tails beyond the prior boundaries. Neural networks

trained with TAILED-UNIFORM achieve superior posterior quality near boundaries while maintaining consistent performance across the full parameter space, with advantages growing in higher dimensions. We validate TAILED-UNIFORM on a Gaussian linear benchmark and on cosmological parameter inference from the matter power spectrum. All code is publicly available on Github.[1]

## 2 Methods

We define a simulator as a function $\mathcal{M}$ that maps parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ to observables $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$ [Cranmer et al., 2020]. Training data consists of pairs $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}$ generated by sampling from a proposal distribution and running the forward model. We construct a neural architecture $q_w(\boldsymbol{\theta} \mid \mathbf{x})$ that approximates the posterior $\mathcal{P}(\boldsymbol{\theta} \mid \mathbf{x})$ by minimizing the negative log-probability loss $\mathcal{L}_{\text{NPE}} = -\mathbb{E}_{\mathcal{D}_{\text{train}}} \log \left[\frac{\mathcal{P}(\boldsymbol{\theta})}{\tilde{\mathcal{P}}(\boldsymbol{\theta})} q_w(\boldsymbol{\theta}|\mathbf{x})\right]$, where the expectation is over training pairs $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}_{i=1}^N$ [Papamakarios and Murray, 2018].

**Priors** It is crucial to distinguish between proposal and assumed priors. The proposal prior $\tilde{\mathcal{P}}(\boldsymbol{\theta})$ is the empirical distribution of training parameters, while the assumed prior $\mathcal{P}(\boldsymbol{\theta})$ encodes scientific beliefs or previous empirical knowledge about plausible parameter values [Cranmer et al., 2020]. Note that when we sample training simulations from a UNIFORM distribution $\tilde{\mathcal{P}} = \mathcal{U}([\theta_{\min}, \theta_{\max}]^d)$, the density drops sharply to zero at the prior boundaries. Because the network can only learn densities where training data exist, posteriors near boundaries are systematically unreliable.

### 2.1 TAILED-UNIFORM Proposal

By assigning weights beyond the primary region of interest, our proposal offers a smooth, continuously differentiable density across $\mathbb{R}^d$, alleviating the sharp discontinuities at the boundary. In a one-dimensional ($d = 1$) case, we define the probability density function as

$$\tilde{\mathcal{P}}_{\text{TAILED-UNIFORM}}(x; a, b, \sigma) = \begin{cases} A \cdot \mathcal{N}(a, \sigma^2), & x \leq a \\ B \cdot \mathcal{U}(a, b), & x \in [a, b] \\ A \cdot \mathcal{N}(b, \sigma^2), & x \geq b \end{cases} \tag{1}$$

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 as $A = \frac{\sigma\sqrt{2\pi}}{\sigma\sqrt{2\pi}+(b-a)}$ and $B = \frac{b-a}{\sigma\sqrt{2\pi}+(b-a)}$. To generalize to a multivariate parameter space $\boldsymbol{\theta} \in \mathbb{R}^d$, we write it as a product of independent univariate TAILED-UNIFORM distributions

$$\tilde{\mathcal{P}}_{\text{TAILED-UNIFORM}}(\boldsymbol{\theta}; \mathbf{a}, \mathbf{b}, \boldsymbol{\sigma}) = \prod_{i=1}^{d} \tilde{\mathcal{P}}_{\text{TAILED-UNIFORM}}(\theta_i; a_i, b_i, \sigma_i), \tag{2}$$

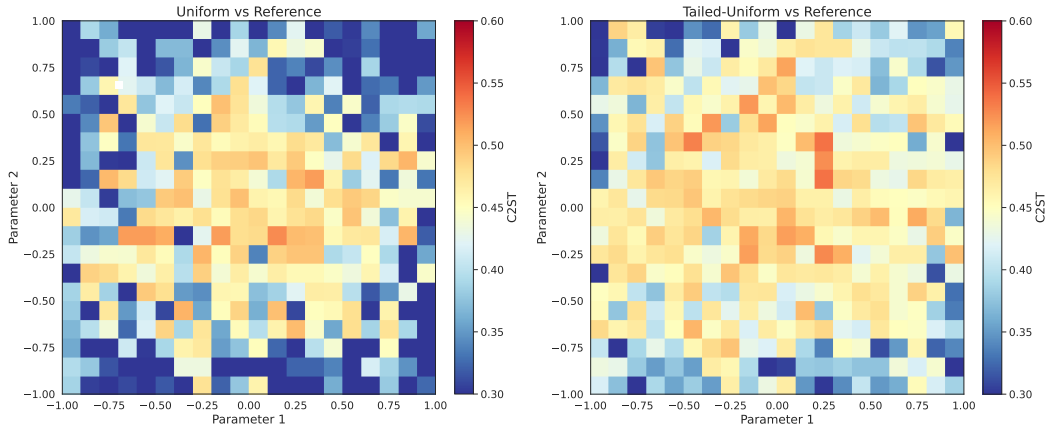
where $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$ define the hypercube boundaries, and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ controls the tail-widths in each dimension. Smaller values of $\sigma_i$ concentrate more samples within the target region $[a_i, b_i]$ but provide less smoothing at the boundaries. In practice, we recommend setting $\sigma_i$ as about $10 - 30$ percent of the box width.

## 3 Toy Problem

We first demonstrate TAILED-UNIFORM on a toy problem where the true posterior is analytically tractable. We use the Gaussian Linear task from the simulation-based inference benchmark [`sbibm`; Lueckmann et al., 2021] to construct a 2D Gaussian simulator $\mathcal{M}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ with a standard Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ on $\boldsymbol{\theta} \in [-1, 1]^2$, giving a Gaussian likelihood $\mathcal{P}(\mathbf{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\theta}, \mathbf{I}_2)$.

**Training** We generate $N = 6{,}000$ pairs $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}$ from two NPE models trained with distinct strategies (proposals): a UNIFORM baseline $\mathcal{U}([-1, 1]^2)$ and our TAILED-UNIFORM with tail width $\sigma_i = 0.1\,(b_i - a_i)$. Built on the `LtU-ILI` framework [Ho et al., 2024], both NPE models use ensembles of Masked Autoregressive Flows (MAF) [Papamakarios et al., 2018] and Masked Autoencoders (MADE) [Germain et al., 2015] with 16 hidden features and 5 coupling layers.

[1] https://github.com/chaipattira/tailed-uniform-sbi.




(a) UNIFORM: boundary degradation (blue near edges). (b) TAILED-UNIFORM: improved, more consistent performance across the full space.

Figure 1: C2ST scores across the parameter space. Blue $\ll 0.5$ = poor; orange/red $\approx 0.5$ = good.

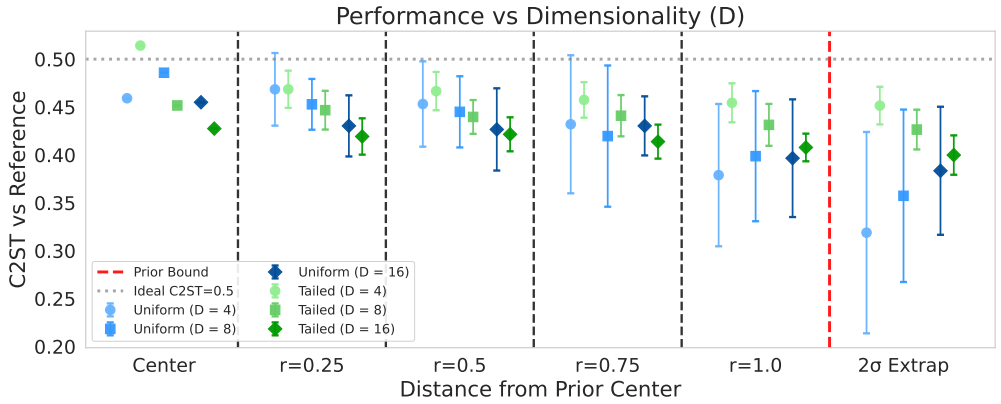
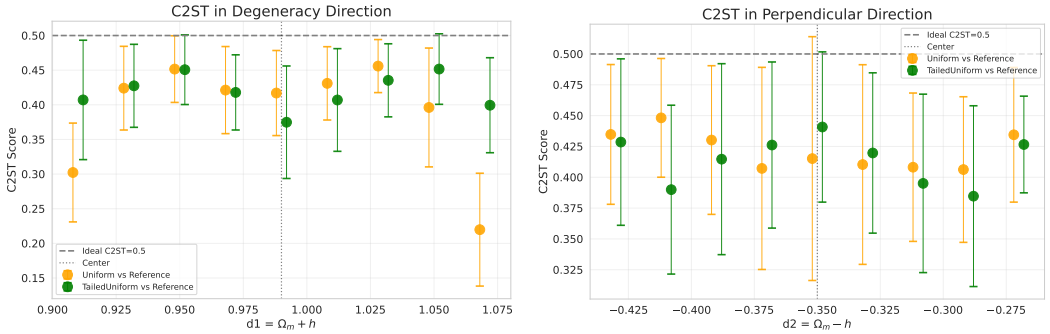



Figure 2: C2ST vs. distance from center across dimensions. TAILED-UNIFORM maintains superior performance over UNIFORM, with the gap generally widening at higher $d$, though the relative advantage varies across radial distances. Error bars: 16th–84th percentile.

**Validation** To measure TAILED-UNIFORM's performance across the entire parameter space, we discretize a uniform rectangular grid with $n = 20$ test locations along each dimension of $[-1, 1]^2$, yielding 400 evaluation points ranging from the center to the boundaries. For each point, we generate an observation, draw $M = 1000$ posterior samples from all three methods, and compute pairwise Classifier Two-Sample Test (C2ST) scores [Lopez-Paz and Oquab, 2018]. C2ST trains an independent classifier to distinguish samples from two distributions; a score of 0.5 indicates identical distributions (perfect inference).

Figure 1 shows how the C2ST degradation pattern is rotationally invariant, depending only on radial distance from the center: UNIFORM's performance deteriorates as it drifts away from the prior center, 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.

**Scaling to higher dimensions** By the curse of dimensionality, boundary regions of a $d$-dimensional hypercube dominate the total volume exponentially as $d$ grows, making boundary pathology increasingly severe for high-dimensional inference tasks common in physics. Although TAILED-UNIFORM allocates a growing fraction of samples to the Gaussian-tail regions beyond $[a_i, b_i]$ (24% at $d = 2$, 86% at $d = 16$), Figure 2 shows that TAILED-UNIFORM maintains superior performance over UNI-




(a) Degeneracy direction ($d_1 = \Omega_m + h$): UNIFORM degrades at both ends; TAILED-UNIFORM is stable.

(b) Perpendicular direction ($d_2 = \Omega_m - h$): both methods perform comparably.

Figure 3: C2ST vs. position along two orthogonal directions. Error bars: 16th–84th percentile.

FORM across all tested dimensions $d \in \{4, 8, 16\}$. At high dimensions, boundary regions are so dominant that the marginal value of tail samples outweighs the cost of diluting interior coverage.

## 4 Science Experiment

Following the same training and validation pipeline in Section 3, we now apply TAILED-UNIFORM to a popular inference benchmark in cosmology: the inference of cosmological parameters $\boldsymbol{\theta} = (\Omega_m, h)$ from the matter power spectrum $\mathbf{P} \in \mathbb{R}^{64}$, where $\Omega_m$ is the matter density parameter and $h$ is the dimensionless Hubble parameter. We generate observations using the `syren-new` emulator [Sui et al., 2025] with heteroskedastic cosmic variance noise [Dodelson and Schmidt, 2020]. As reference posteriors, we use MCMC to collect samples under a truncated Gaussian prior centered in $\Omega_m \in [0.24, 0.40]$, $h \in [0.61, 0.73]$. To ensure a fair comparison between the NPEs, we perform independent `Optuna` hyperparameter searches [Akiba et al., 2019] for each proposal over flow architecture (MAF or NSF), hidden features, transform layers, batch size, and learning rate.

**Degeneracy** The matter power spectrum is sensitive to the combination $\Omega_m h$ rather than to each parameter independently [Peacock and Dodds, 1996], creating an anti-correlation that manifests as slanted banana-shaped posterior contours. As a result, the information embedded in the posterior will be weak along the degenerate directions in the parameter space. To connect the performance degradation trend to the underlying physics, we propose a directional binning strategy along the degeneracy direction $d_1 = \Omega_m + h$ and the perpendicular direction $d_2 = \Omega_m - h$.

Figure 3 shows that boundary pathology is anisotropic, with the most severe manifestation along degenerate directions. While both approaches perform comparably (excluding statistical fluctuations) along the perpendicular direction $d_2$, the UNIFORM performance sinks at both boundaries across the degeneracy direction $d_1$.

## 5 Conclusion

We proposed TAILED-UNIFORM, a hybrid proposal distribution that replaces the sharp boundary of the standard UNIFORM prior with smooth Gaussian tails, mitigating the structural pathology that causes NPE posteriors to degrade near the edges of parameter space. With minimal hyperparameter tuning, TAILED-UNIFORM enables more robust inference across the full parameter space.

In the Gaussian Linear benchmark with a Gaussian (assumed) prior, TAILED-UNIFORM maintains more consistent C2ST scores throughout the domain while UNIFORM degrades near the boundary. The advantage also appears to widen in higher dimensions where boundary regions dominate the parameter volume. For cosmological parameter inference, TAILED-UNIFORM provides support along the degenerate direction where elongated posteriors extend toward prior edges. Future work should explore behavior beyond $d = 16$ and whether aligning the proposal tails with the posterior degeneracy direction rather than the physical parameter axes yields further gains.

## 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.

Richard Bellman and Robert Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9, 2003.

Rob Cornish, Anthony Caterini, George Deligiannidis, and Arnaud Doucet. Relaxing bijectivity constraints with continuously indexed normalising flows. In *International conference on machine learning*, pages 2133–2143. PMLR, 2020.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, May 2020. ISSN 1091-6490. doi: 10.1073/pnas.1912789117. URL http://dx.doi.org/10.1073/pnas.1912789117.

Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Academic press, 2020.

Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation, 2015. URL https://arxiv.org/abs/1502.03509.

David S. Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. Automatic posterior transformation for likelihood-free inference, 2019. URL https://arxiv.org/abs/1905.07488.

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

Matthew Ho, Deaglan J Bartlett, Nicolas Chartier, Carolina Cuesta-Lazaro, Simon Ding, Axel Lapel, Pablo Lemos, Christopher C Lovell, T Lucas Makinen, Chirag Modi, et al. Ltu-ili: An all-in-one framework for implicit inference in astrophysics and cosmology. *arXiv preprint arXiv:2402.05137*, 2024.

David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests, 2018. URL https://arxiv.org/abs/1610.06545.

Jan-Matthis Lueckmann, Pedro J. Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Macke. Flexible statistical inference for mechanistic models of neural dynamics, 2017. URL https://arxiv.org/abs/1711.01861.

Jan-Matthis Lueckmann, Jan Boelts, David S. Greenberg, Pedro J. Gonçalves, and Jakob H. Macke. Benchmarking simulation-based inference, 2021. URL https://arxiv.org/abs/2101.04653.

Michael D McKay, Richard J Beckman, and William J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

Rüdiger Pakmor, Volker Springel, Jonathan P Coles, Thomas Guillet, Christoph Pfrommer, Sownak Bose, Monica Barrera, Ana Maria Delgado, Fulvio Ferlito, Carlos Frenk, et al. The millenniumtng project: the hydrodynamical full physics simulation and a first look at its galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 524(2):2539–2555, 2023.

George Papamakarios and Iain Murray. Fast $\epsilon$-free inference of simulation models with bayesian conditional density estimation, 2018. URL https://arxiv.org/abs/1605.06376.

George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation, 2018. URL https://arxiv.org/abs/1705.07057.

JA Peacock and SJ Dodds. Non-linear evolution of cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 280(3):L19–L26, 1996.

Frans Pretorius. Evolution of binary black-hole spacetimes. *Physical review letters*, 95(12):121101, 2005.

Magdalena Siwek, Rainer Weinberger, and Lars Hernquist. Orbital evolution of binaries in circumbinary discs. *Monthly Notices of the Royal Astronomical Society*, 522(2):2707–2717, 2023.

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. *Astronomy &; Astrophysics*, 698:A1, May 2025. ISSN 1432-0746. doi: 10.1051/0004-6361/202452854. URL http://dx.doi.org/10.1051/0004-6361/202452854.

Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N Spergel, Rachel S Somerville, Romeel Dave, Annalisa Pillepich, Lars Hernquist, Dylan Nelson, Paul Torrey, et al. The camels project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, 915(1):71, 2021.