A Hitchhiker's Guide to Poisson Gradient Estimation

Reliable Pathwise Gradients for

Discrete Latent Variable Models

Michael Ibrahim*, Hanqi Zhao*, Eli Sennesh, Zhi Li, Anqi Wu, Jacob Yates, Chengrui Li† , Hadi Vafaii†

*Equal Contribution †Co-Senior Authors

Introduction & The Core Problem

Context

Poisson models are essential for capturing **noisy spike counts** in NeuroAI.

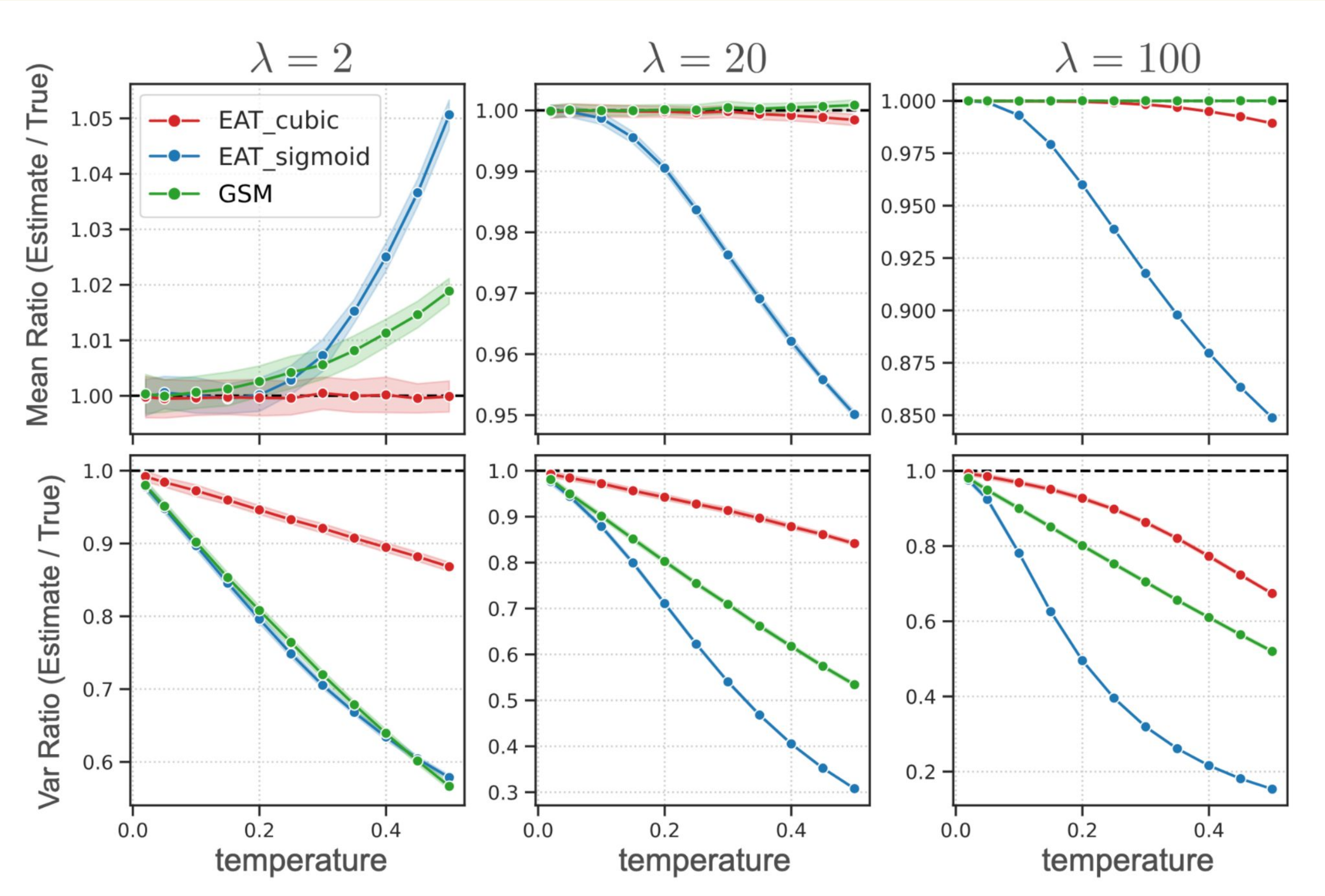
Roadblock

Backpropagation is blocked by the **discrete nature** of integer-valued samples.

Hyper-Sensitivity

Current relaxations (EAT/GSM) are dangerously **sensitive to temperature tuning**.

The Solution: $\mathrm{EAT}_{\mathrm{cubic}}$



Smoothstep Solution

We replace sigmoid with a **Cubic Hermite Interpolant** for compact support [0, 1].

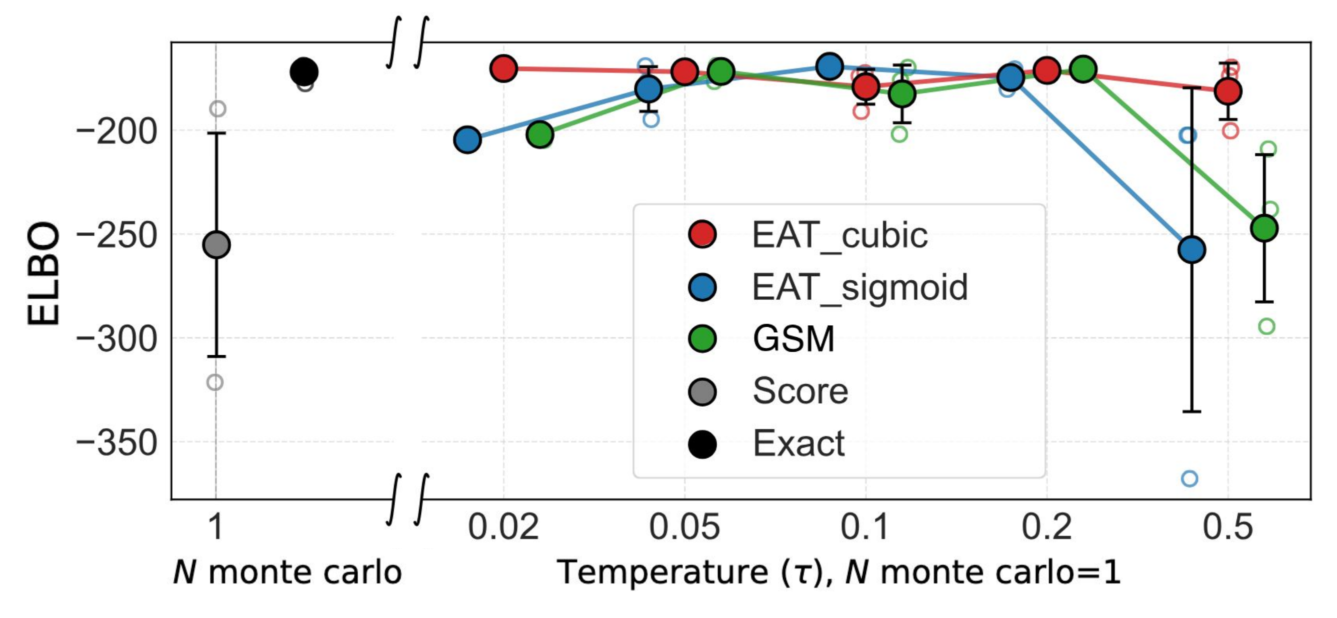
$$f(x) = 3x^2 - 2x^3$$

Result: Unbiased first moment matching true firing rate.

Fig: Empirical validation of moment biases across relaxation methods. We compare across temperatures $\tau \in [0.02, 0.5]$ and firing rates $\lambda \in \{2, 20, 100\}$ (columns). Top row: Ratio of empirical mean to true Poisson mean (λ). Bottom row: Ratio of empirical variance to true Poisson variance (λ)

P-VAE Results: Robust Optimization

Training linear P-VAEs on natural image patches provides a closed-form ELBO, giving us an **exact gradient baseline**.



Scaling to 40x40 ImageNet

At larger scale, $\mathrm{EAT}_{\mathrm{cubic}}$ was the **only method** able to bridge the performance gap to exact gradients, while baselines suffered a persistent gap regardless of tuning.

Fig: Validation ELBO across relaxation methods and temperatures for linear P-VAE trained on whitened natural image patches.
Left columns: Score function baseline across MC sample sizes.
Right columns: Pathwise methods across temperatures $\tau \in \{0.02, 0.05, 0.1, 0.2, 0.5\}$.

POGLM Results: Neural Data Analysis

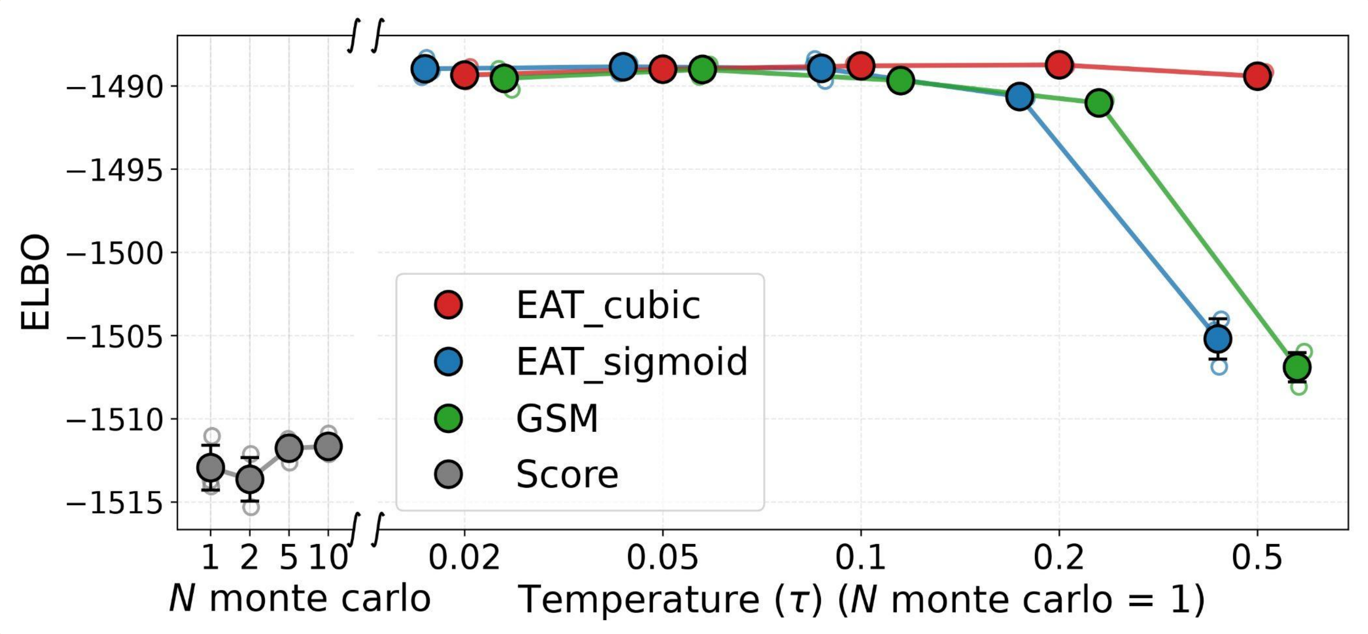
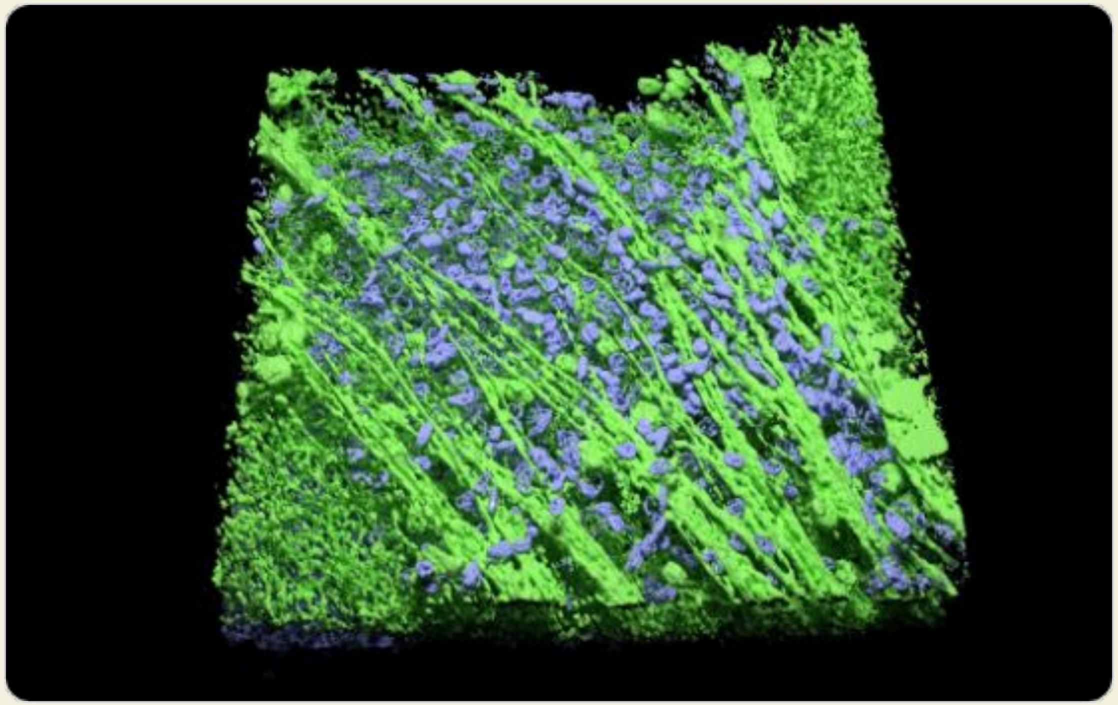


Fig: POGLM validation ELBO on a retinal ganglion cell (RGC) dataset (Pillow & Scott, 2012), assuming 3 hidden neurons.
Left columns: Score function baseline across MC sample sizes.
Right columns: Pathwise methods across temperatures τ in $\{0.02, 0.05, 0.1, 0.2, 0.5\}$.

- **Stability:** $\mathrm{EAT}_{\mathrm{cubic}}$ held steady across temperatures.
- **Practical Impact:** Practitioners no longer need to perform costly, exhaustive grid searches to find a working temperature.

POGLM Results: Neural Data Analysis



Inferring Latent Connectivity

We applied Partially Observable GLMs to real recordings from **27 retinal ganglion cells** to infer spiking neural network connectivity.

- **Stability:** $\mathrm{EAT}_{\mathrm{cubic}}$ held steady across 1, 2, or 3 hidden neurons at a wide range of temperatures.
- **Practical Impact:** Practitioners no longer need to perform costly, exhaustive grid searches to find a working temperature.

Conclusion & Recommendations

Method	Dist. Fidelity	Gradient Bias	Gradient Variance	Temp. Robustness	General-izability	Recommendation
$\text{EAT}_{\text{cubic}}$ (Ours)	★★★★★ *Unbiased*	★★★★★ *Low*	★★★★☆ *Moderate*	★★★★★ *Excellent*	★☆☆☆☆ *Poisson only*	Default Choice
$\text{EAT}_{\text{sigmoid}}$ [10]	★☆☆☆☆ *High bias*	★★★★☆ *High at high temp.*	★★★★★ *Lowest*	★★☆☆☆ *Sensitive*	★☆☆☆☆ *Poisson only*	Max Smoothness
GSM [20]	★★★☆☆ *Moderate*	★★★☆☆ *High at low temp.*	★★☆☆☆ *Poor*	★★☆☆☆ *Sensitive*	★★★★★ *Universal*	Non-Poisson Data

Open Source Code Available on https://github.com/hadivafaii/PoissonGradientEstimation

Check out our Poster for the Full Gradient Analysis!

Conclusion & Recommendations

Gumbel-Softmax (GSM)

The most versatile choice for arbitrary discrete PMFs (like Negative Binomial) or under-dispersed alternatives (COM-Poisson).

$\text{EAT}_{\text{cubic}}$ (Our Method)

The default choice for **Poisson Latent Variable Models**. Requires practically no tuning and guarantees an unbiased mean.

Open Source Code Available on https://github.com/hadivafaii/PoissonGradientEstimation

Check out our Poster for the Full Gradient Analysis!