Stabilizing the Kumaraswamy Distribution

Max Wasserman Gonzalo Mateos

University of Rochester, Rochester, NY

Paper

Latent Variable Models (LVMs)

Generative model of observations $\boldsymbol{x}$ with latent variables $\boldsymbol{z}$ and model parameters $\boldsymbol{\theta}$:

$$p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z}) p_\theta(\boldsymbol{z}) d\boldsymbol{z}.$$

- Typically $\mathbf{z} \in \mathbb{R}^n$, but often desire *interval-supported* $\boldsymbol{z}$. Ex: prob. of click $\in [0,1]$, joint angle $\in [0, 2\pi]$, Pearson correlation $\in [-1,1]$.

Amortized Variational Inference

- Challenge: Posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ often intractable.
- **Amortized variational inference**: approximate $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ with tractable distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$, parameterized via shared NN encoder $e_{\boldsymbol{\phi}}(\mathbf{x})$.
- Optimize the Evidence Lower BOund (ELBO):

$$\mathcal{L} = \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x})} [\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] - D_{\mathrm{KL}} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}) \,\|\, p(\boldsymbol{z}) \right)$$

- **Reparameterization Trick**: Enables efficient ELBO optimization by unblocking gradient path through samples $\nabla_{\boldsymbol{\phi}} \tilde{\boldsymbol{z}}$, where $\tilde{\boldsymbol{z}} \sim q_{\boldsymbol{\phi}}(\boldsymbol{z})$.

Interval-Supported Distributions

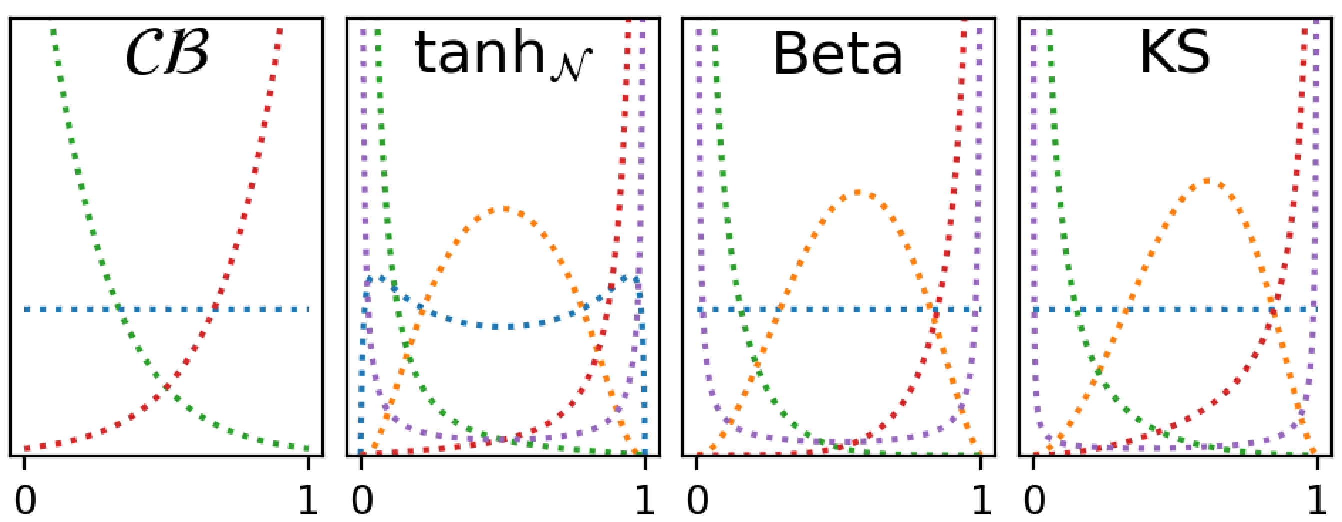

Figure 1. Relevant interval-supported distributions.

Criteria for LVM use: 1. Reparameterization trick; 2. Expressive; 3. Simple expressions.

Kumaraswamy fulfills all! Why not widely used?

Stabilizing the Kumaraswamy

The Kumaraswamy (KS) distribution, with parameters $a, b > 0$ and support $x \in (0,1)$, has

- PDF: $f(x) = abx^{a-1}(1-x^a)^{b-1}$
- inverse CDF: $F^{-1}(u) = (1-u^{\frac{1}{b}})^{\frac{1}{a}}, u \in (0,1)$

Identifying the instability: $\log(1-\exp(x))$.

- Log-pdf, inverse CDF, and their gradients contain $\log(1\text{-}\exp(x))$ terms, stabilized via

$$\texttt{log1mexp}(x) = \begin{cases} \texttt{log(-expm1(x))}, & -\log 2 \le x < 0 \\ \texttt{log1p(-exp(x))}, & x < -\log 2. \end{cases}$$

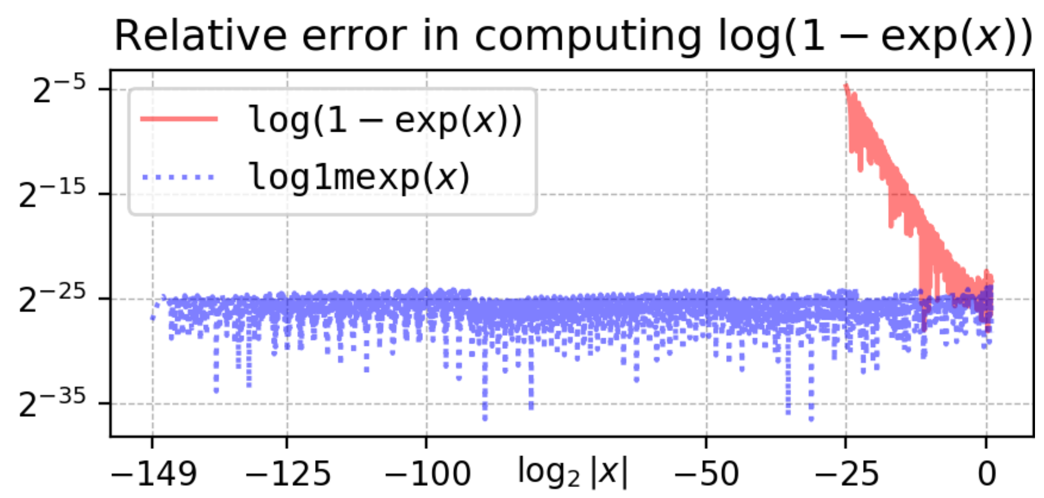

Figure 2. `log1mexp(x)` maintains accuracy.

Contribution. Stable, unconstrained parameterization of KS log-pdf, inverse CDF, and gradients.

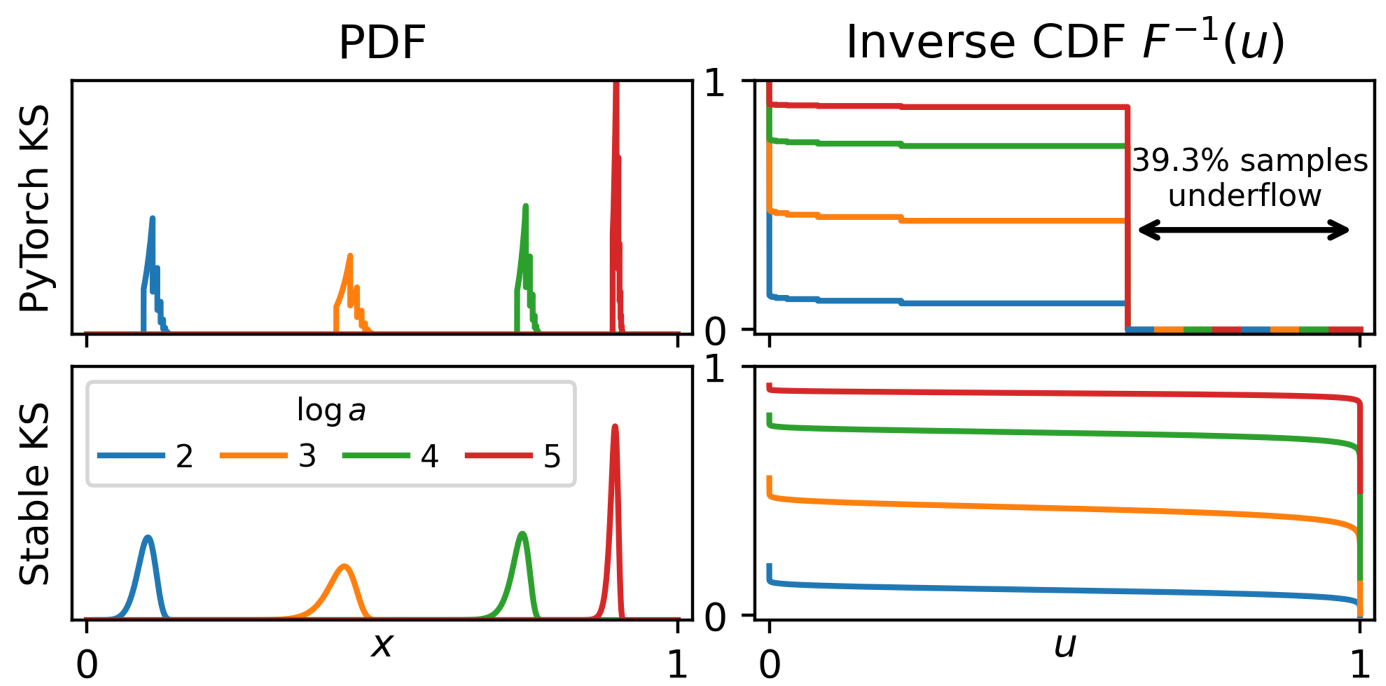

Figure 3. Stabilizing $\log(1-\exp(x))$ terms eliminates numerical instabilities in the KS log-pdf and inverse CDF.

Novel, Scalable LVMs

Variational Bandit Encoder (VBE)

- Task: contextual Bernoulli multi-armed bandits (MABs)
- Model: KS variational posterior $\prod_k q_{\boldsymbol{\phi}}(z_k|\boldsymbol{x}_k)$, prior $p(\boldsymbol{z}) = U^K_{(0,1)}$, and Bernoulli reward likelihood $p(r|z_k)$. $\tilde{z}_k \sim q_{\boldsymbol{\phi}}(z_k|\boldsymbol{x}_k)$ parameterizes likelihood. Maximize ELBO.

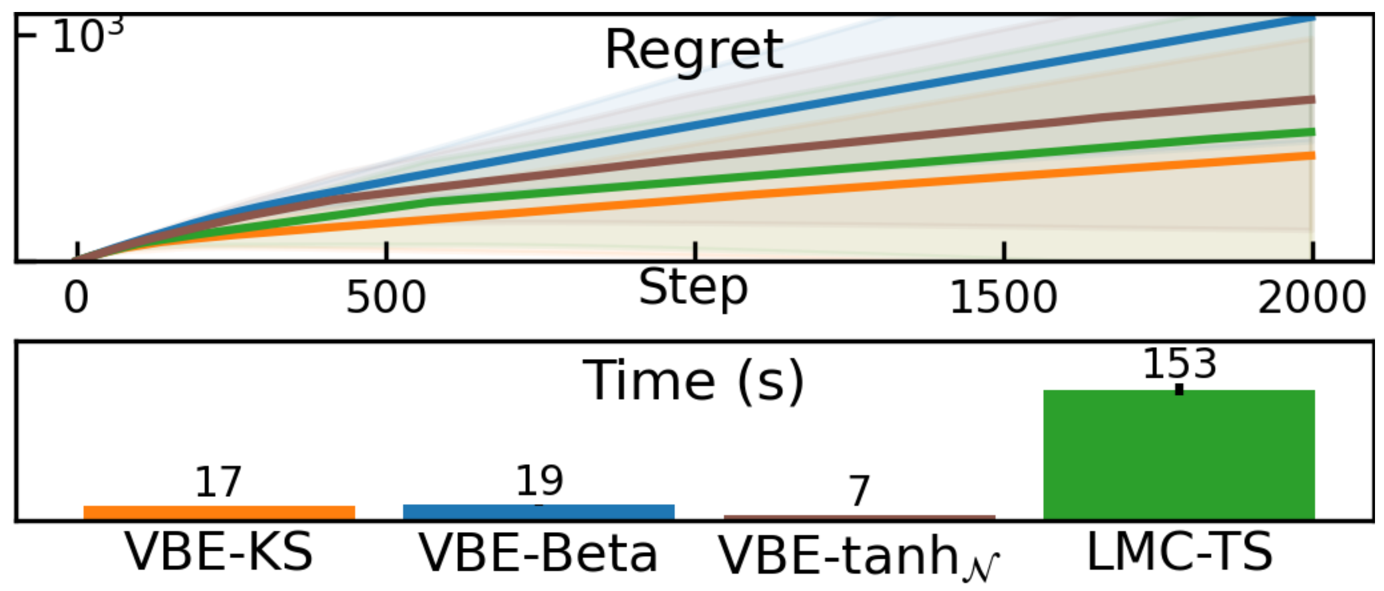

Figure 4. VBEs are scalable and most performant with KS.

Variational Edge Encoder (VEE)

- Task: link prediction with graph neural networks (GNN).
- Model: Analogous to VBE, but GNN encoder parameterizes $q_{\boldsymbol{\phi}}(z_{u,v}|\boldsymbol{X}, \mathcal{D}_{tr})$ for each possible edge.

Conclusions

- The KS distribution is uniquely suitable for LVMs, but no stable implementation exists.
- Introduce stable, unconstrained parameterization of KS.
- Introduce scalable LVMs, VBE and VEE, which are most performant with the KS.

https://github.com/maxwass Twitter: @MaxWasserman1 maxw14k@gmail.com