A Heavy-Tailed Algebra for Probabilistic Programming

Feynman Liang, Liam Hodgkinson & Michael W. Mahoney

Problem: *Normalizing flows do not accurately reflect tail probabilities unless the tail of the base distribution matches the target.*

Solution: *Exploit the PPL compiler to choose the correct base distribution!*

Generalized Gamma Algebra (GGA)

Operations acting on a three parameter family of densities with generalized Gamma tails

$$X \equiv (\nu, \sigma, \rho) \iff p_{|X|}(x) \sim cx^{\nu} e^{-\sigma x^{\rho}}$$

Ordering	$(\nu_1,\sigma_1,\rho_1) \leq (\nu_2,\sigma_2,\rho_2) \iff \limsup_{x\to\infty} \frac{x^{\nu_1} e^{-\sigma_1 x^{\rho_1}}}{x^{\nu_2} e^{-\sigma_2 x^{\rho_2}}} < +\infty.$
Addition	$(\nu_1,\sigma_1,\rho_1) \oplus (\nu_2,\sigma_2,\rho_2) \equiv \begin{cases} \max\{(\nu_1,\sigma_1,\rho_1),(\nu_2,\sigma_2,\rho_2)\} & \text{if } \rho_1 \neq \rho_2 \text{ or } \rho_1,\rho_2 < 1 \\ (\nu_1+\nu_2+1, \min\{\sigma_1,\sigma_2\}, 1) & \text{if } \rho_1 = \rho_2 = 1 \\ \left(\nu_1+\nu_2+\frac{2-\rho}{2}, (\sigma_1^{-\frac{1}{\rho-1}}+\sigma_2^{-\frac{1}{\rho-1}})^{1-\rho}, \rho\right) & \text{if } \rho = \rho_1 = \rho_2 > 1. \end{cases}$
Powers	$(\nu,\sigma,\rho)^{\beta} \equiv \left(\frac{\nu+1}{\beta} - 1, \sigma, \frac{\rho}{\beta}\right)$ for $\beta > 0$
Reciprocal	$(\nu,\sigma,\rho)^{-1} \equiv \begin{cases} (-\nu-2,\sigma,-\rho) & \text{if } (\nu+1)/\rho > 0 \text{ and } \rho \neq 0 \\ (-2,1,0) & \text{otherwise} \end{cases}$
Scalar Multiplication	$c(\nu,\sigma,\rho) \equiv (\nu, \sigma\lvert c\rvert^{-\rho}, \rho)$
Multiplication	$(\nu_1,\sigma_1,\rho_1) \otimes (\nu_2,\sigma_2,\rho_2) \equiv \begin{cases} \left(\frac{1}{\mu}\left(\frac{\nu_1}{\lvert\rho_1\rvert}+\frac{\nu_2}{\lvert\rho_2\rvert}+\frac{1}{2}\right), \sigma, -\frac{1}{\mu}\right) & \text{if } \rho_1,\rho_2 < 0 \\ \left(\frac{1}{\mu}\left(\frac{\nu_1}{\rho_1}+\frac{\nu_2}{\rho_2}-\frac{1}{2}\right), \sigma, \frac{1}{\mu}\right) & \text{if } \rho_1,\rho_2 > 0 \\ (-\lvert\nu_1\rvert, 1, 0) & \text{if } \rho_1 \leq 0, \rho_2 > 0 \\ (-\min\{\lvert\nu_1\rvert,\lvert\nu_2\rvert\}, 1, 0) & \text{if } \rho_1 = 0, \rho_2 = 0 \end{cases}$ where $\mu = \frac{1}{\lvert\rho_1\rvert} + \frac{1}{\lvert\rho_2\rvert} = \frac{\lvert\rho_1\rvert + \lvert\rho_2\rvert}{\lvert\rho_1\rho_2\rvert}$, $\sigma = \mu(\sigma_1\lvert\rho_1\rvert)^{\frac{1}{\mu\lvert\rho_1\rvert}}(\sigma_2\lvert\rho_2\rvert)^{\frac{1}{\mu\lvert\rho_2\rvert}}$.
Product of Densities	$(\nu_1,\sigma_1,\rho_1) \,\&\, (\nu_2,\sigma_2,\rho_2) \equiv \begin{cases} (\nu_1+\nu_2, \sigma_1, \rho_1) & \text{if } \rho_1 < \rho_2 \\ (\nu_1+\nu_2, \sigma_1+\sigma_2, \rho) & \text{if } \rho = \rho_1 = \rho_2 \\ (\nu_1+\nu_2, \sigma_2, \rho_2) & \text{otherwise.} \end{cases}$
Exponentials	$\exp(\nu,\sigma,\rho) \equiv \begin{cases} (\sigma+1,1,0) & \text{if } \rho \geq 1 \\ (-1,1,0) & \text{otherwise.} \end{cases}$
Logarithms	$\log(\nu,\sigma,\rho) \equiv \begin{cases} (0,\lvert\nu\rvert-1,1) & \text{if } \nu < -1 \\ (\tilde{\infty},1,2\tilde{\infty}) & \text{otherwise.} \end{cases}$
Functions (L-Lipschitz)	$f(X_1,\ldots,X_n) \equiv L\max\{X_1,\ldots,X_n\}$

Posterior distributions can be evaluated by performing inverse operations in reverse.

Procedure

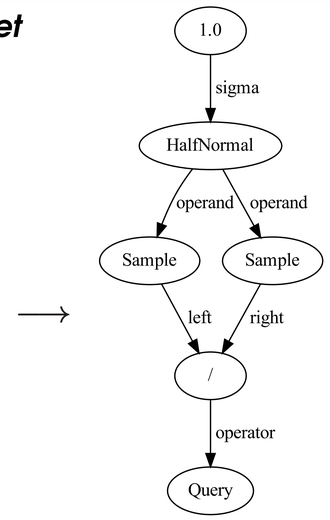
Analyze Target

(probabilistic program)

```
@bm.random_variable
def n():
    return HalfNormal(1.0)

@bm.random_variable
def x():
    return HalfNormal(1.0)

@bm.functional
def x2():
    return x() / n()
```



GGA

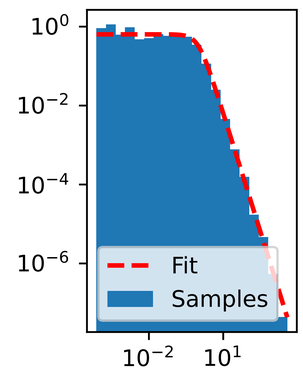
$(-2, 1, 0)$

Calibrate Tails

Operations are traced using GGA to identify the target's tail

Refine Bulk

(normalizing flow)



Density Estimation

GGA base distributions yield accurate tails without compromising on the bulk!

Target	α	Estimated α			Log-likelihood		
		Cauchy	GGA	Normal	Cauchy	GGA	Normal
Cauchy	2	**2.1 (0.03)**	**2.1 (0.07)**	7.7 (2.5)	**−2.53 (0.05)**	−3.22 (0.06)	-1.2×10^3 (6×10^3)
Inverse Gaussian	2	**1.9 (0.03)**	**1.9 (0.092)**	7.3 (1.7)	−3.55 (0.08)	**−3.26 (0.05)**	-2.6×10^4 (6×10^3)
Student T	3	2.0 (0.06)	**3.3 (0.45)**	7.7 (2.3)	**−2.12 (0.03)**	−2.75 (0.04)	−2.92 (0.47)
Chi-squared	∞	2.1 (0.07)	**5.2 (1.6)**	**6.8 (2.4)**	−2.30 (0.05)	**−2.03 (0.04)**	−2.24 (0.04)
Normal	∞	2.9 (0.6)	**8.2 (4.0)**	**8.4 (3.5)**	−1.53 (0.03)	**−1.41 (0.02)**	**−1.42 (0.02)**

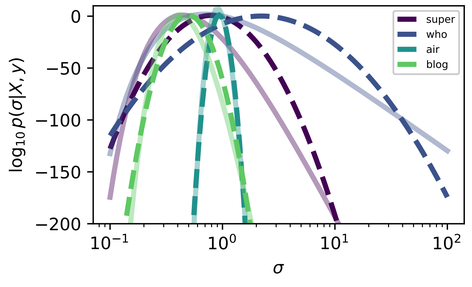
Table 1: Log-likelihoods and estimated tail exponents from obtained samples vs. actual tail exponent (α) for normalizing flows with Cauchy, GGA, and normal base distributions on a variety of basic target densities.

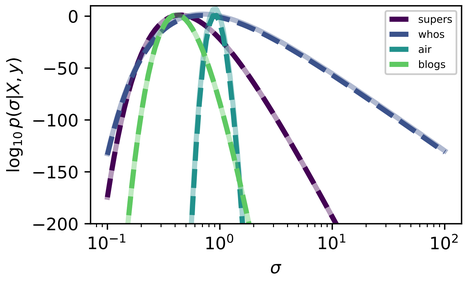
Bayesian Linear Regression

The Bayesian posterior for linear regression has power law tails.

(**left**) Normalizing flows inevitably fail to capture these tails with a Gaussian base distribution.

(**right**) The GGA guarantees perfect fits, even with an affine flow!





The University of Melbourne · Berkeley University of California · ICSI · Neural Information Processing Systems