Flexible Tails for Normalizing Flows: Modelling Financial Return Data

Tennessee Hickling
Supervised by Dennis Prangle

University of Bristol

Outline of Talk

- Motivations
- What is a **Normalising Flow (NF)**?
- The problem – **tails** of distributions
- Our solution
- Experimental evidence

Motivations: Challenges

Characteristics of financial data (asset returns)

- Temporal structure
  - volatility clustering
  - time varying correlations
- Complex structure in high dimensions
- Heavy tails

Motivation: Aims

- Use ML methods to get better models for observed phenomena
- Sample new synthetic data sets with realistic characteristics

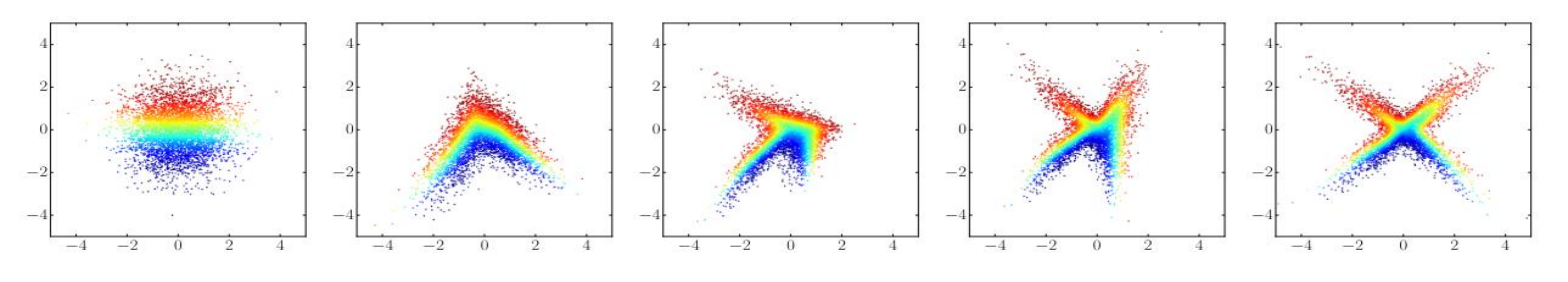
Motivation: Focus

Characteristics of financial data

- Temporal structure
  - volatility clustering
  - time varying correlations
- **Complex structure in high dimensions**
- **Heavy tails**

Normalising Flows: Setup

- Assume a base (latent) distribution over $Z \in \mathbb{R}^d$
- Sample $X = T(Z; \boldsymbol{\theta}) \in \mathbb{R}^d$
- Parameterisation $\boldsymbol{\theta}$ provided by neural networks
- Compose multiple T for more flexibility



[Papamakarios, 2021]

Normalising Flows: Key Ideas

Key idea - Construct T such that:

- We have analytic inverse T^{-1}
- The determinant of the Jacobian of T is tractable

Then:

- Analytic approximate density for X, $q(x; \boldsymbol{\theta})$ (transformation of a random variable)
- Able to sample new X

Normalising Flows: Training

Fit by maximum likelihood:

Observed data $\{x_i\}_{i=0,\dots,n}$

$$\text{Maximise } L(\theta) = \sum_i \log q(x_i;\, \theta)$$

The gradient of $L(\theta)$ can be numerically evaluated with automatic differentiation.

We can optimise with SGD methods.

Good choice for Financial Data?

Pros:

- Evidence of working well for high dimensional complex data (e.g images) [Kingma 2018]
- Exact density (may be useful for risk)

Cons:

- May be less flexible than e.g. GAN, VAE
- No dimension reduction

The Problem

Lipshitz (≈bounded derivative) transformations cannot alter the tails of distributions

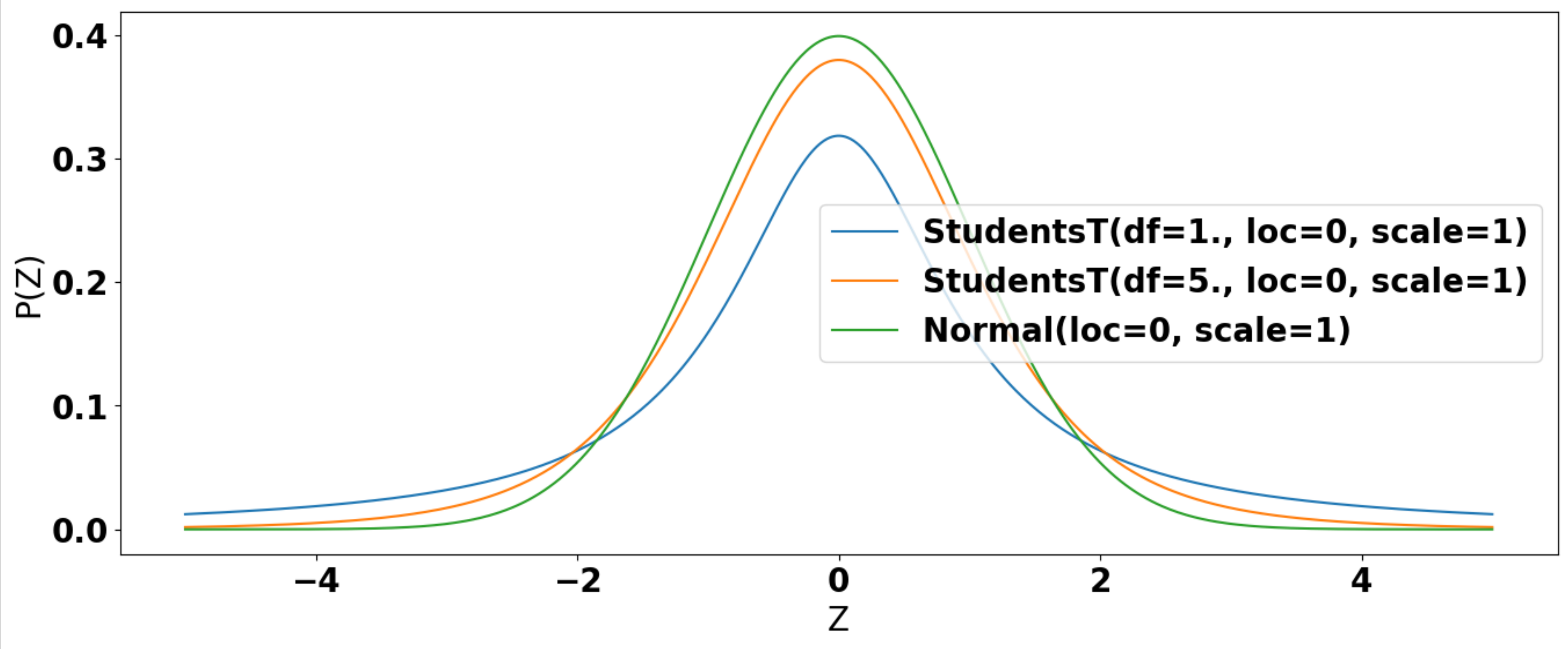
[Jaini 2020]

->Many NF transformations are Lipshitz

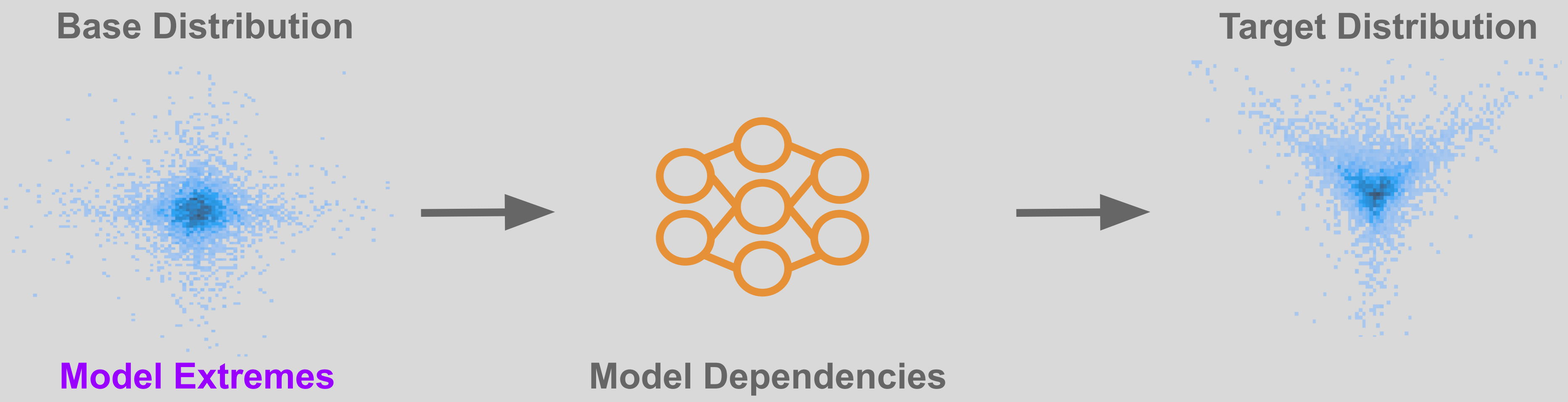
->Very important for simulating financial data

Solutions: Current

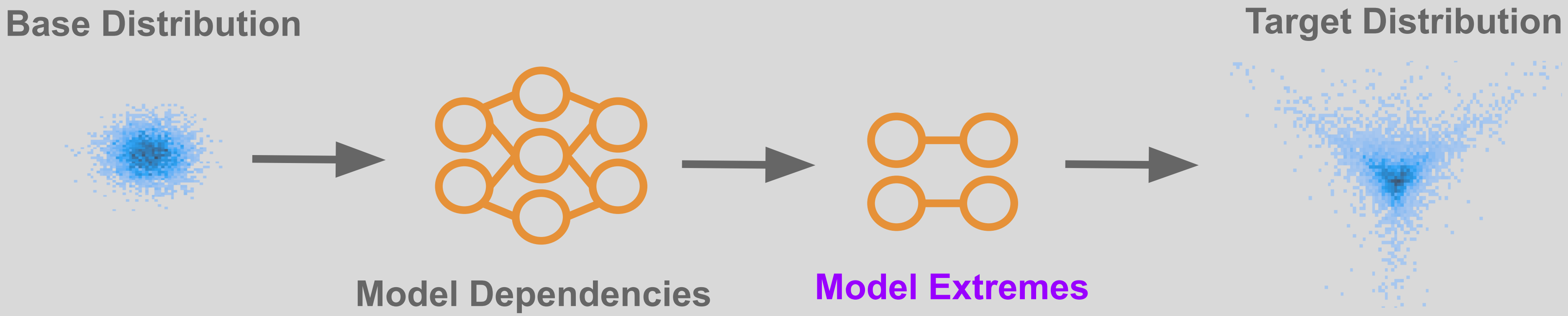
Introduce a base distribution with trainable tails.
[Jaini 2020, Laszkiewicz 2022, Liang 2022]



CURRENT PROPOSALS



OUR PROPOSAL (also [**McDonald 2022**])



Details of Our Solution: Tail Transform

Introduce transformation based on the functional form of the **Generalised Pareto Distribution (GPD).**

GPD – Well theoretically justified model for tails.

[Coles 2001, Pickands 1975]

Details of Our Solution: Tail Transform

Inverse CDF of the GPD:

$$Q(u; \lambda) = [(1 - u)^{-\lambda} - 1] / \lambda$$

Tail parameter **$\lambda > 0$ is the GEV tail parameter.**

Extend to reals via standard Guass error function

$$T(z; \lambda) = Q(\mathrm{erf}(z); \lambda)$$

Details of Our Solution: Tail Transform

Other modifications:

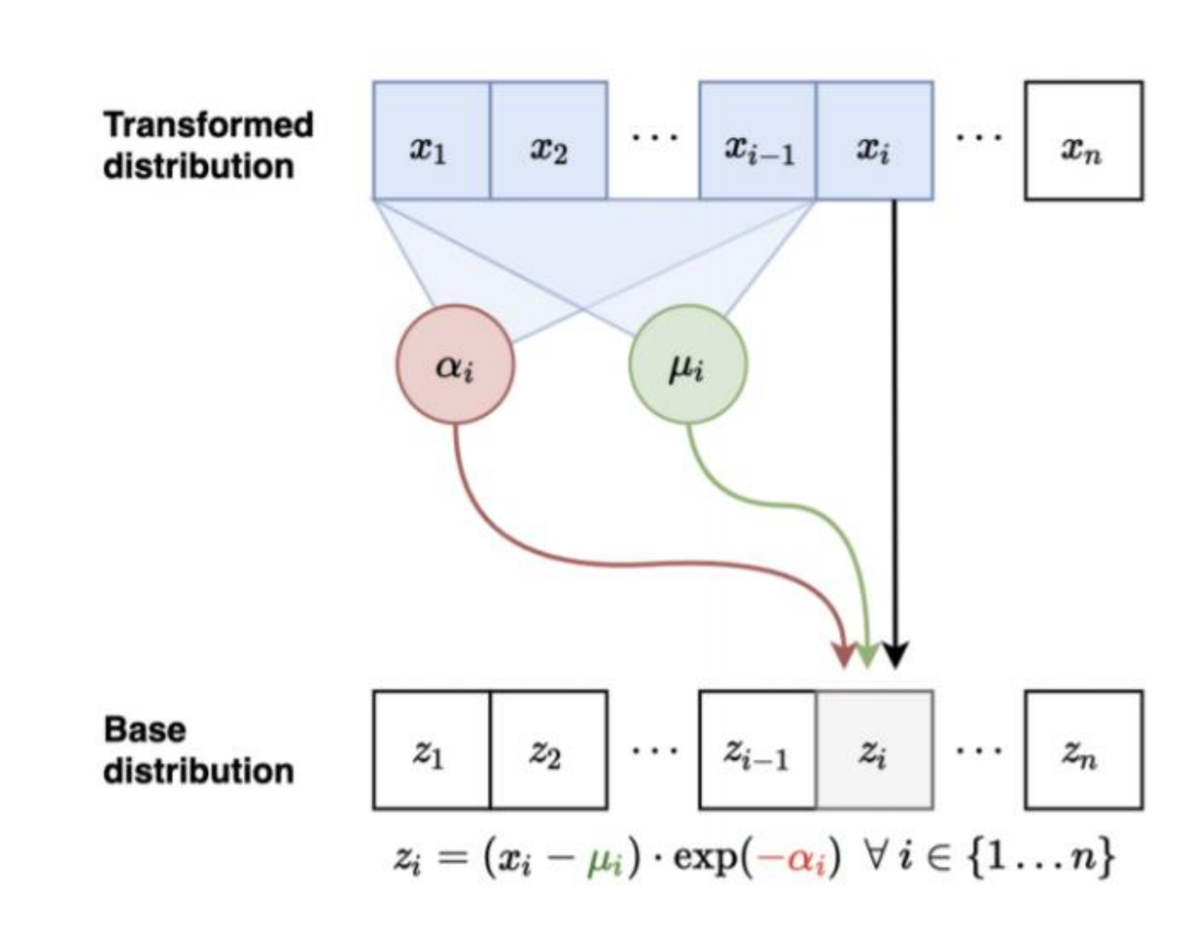
- **Tail asymmetry-** allow for different positive and negative tail parameter
- **Numerical stability** – use of erfc
- **Guassian tails** - Switch to power transform (including identity) at transition for $-1 < \lambda < 0$
- Incorporate joint **shift and scale**

Parameters for each dimension:

$$h_i = (\lambda_-, \lambda_+, \mu, \sigma)$$

Details: Masked Autoregression

Multivariate problem: Use standard masked autoregressive approach **[Papamakarios 2022]**



Key Ideas:

- Parameters of transformation are a function of inputs
- We can evaluate in a single pass of NN

[Anadan, Dalmia]

Details: Marginal Transform

- Want to avoid passing any extremes to NN if possible
- Also consider marginal transformations

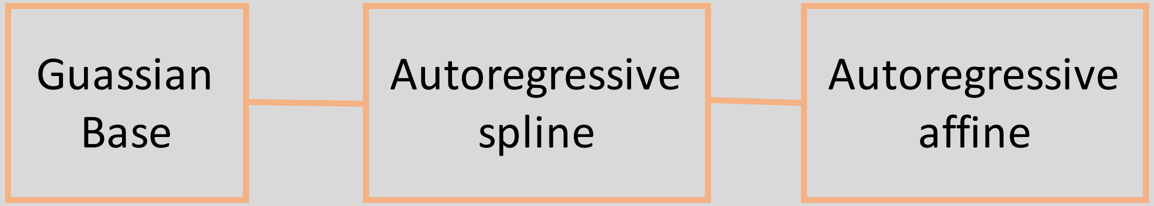
Experiments: Experimental Details

We test the approaches on S&P 500 daily returns 2010-2022 – treat as IID

- Consider top d most traded stocks
- 10 repeats for each model
- 40/20/40 train/validation/test split
- Adam optimiser, Early stopping

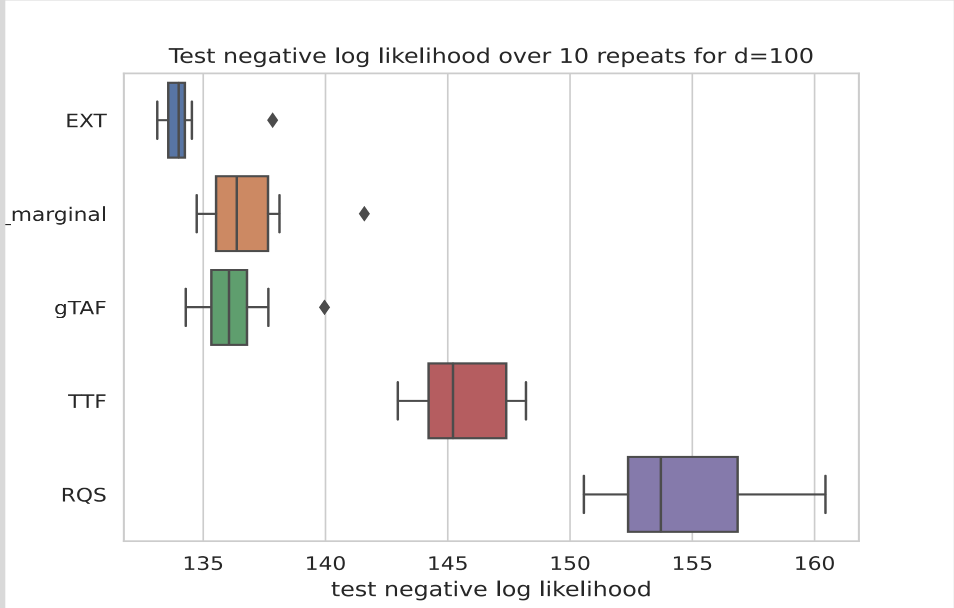
Experiments: Models

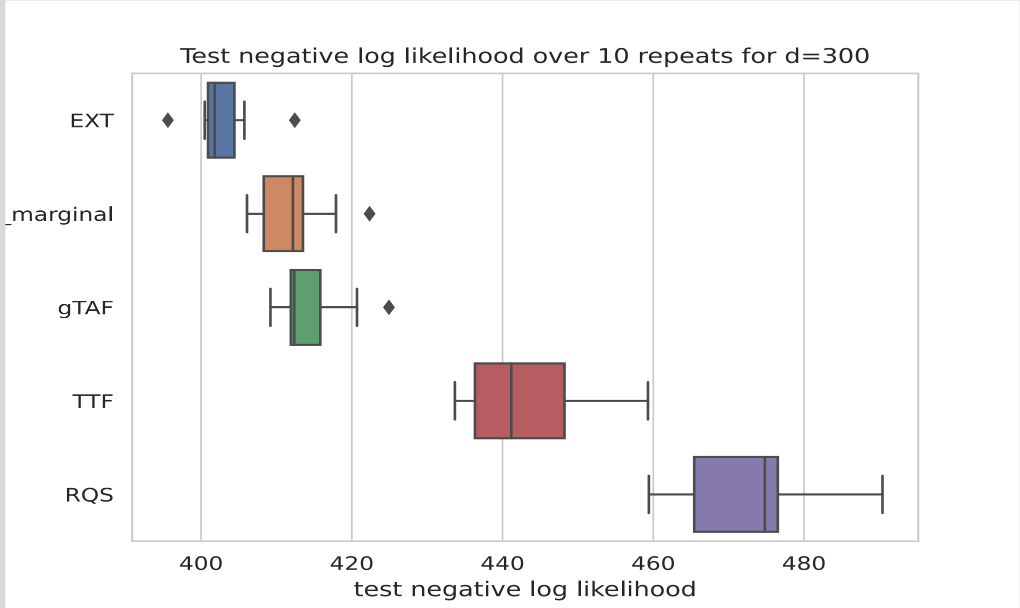
Base architecture **RQS**:



- **gTAF**: -Gaussian Base, +Students T base
  - **[Laszkiewicz 2022]** (generalised tail adaptive flow)
- **TTF**: -affine, +full tail transformation
- **EXT**: -affine, +full tail transformation with $\lambda > 0$
- **TTF_m**: -affine, +marginal tail transformation

Experiments: Results





Conclusion + Future Work

- Modelling tails provides far superior fit relative to naïve approach
- Our experiments provide some evidence that capturing extremes in final transformation is good
- More investigations required
- Opportunities to incorporate temporal information (conditioning on hidden state)

References

[Papamakarios 2021] Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B.: Normalizing flows for probabilistic modeling and inference. Journal of Machine Learning Research 22(57), 1–64 (2021)

[Kingma 2018] Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In: Advances in Neural Information Processing Systems. vol. 31 (2018)

[Jaini 2020]P. JAINI, I. KOBYZEV, Y. YU, AND M. BRUBAKER, “TAILS OF LIPSCHITZ TRIANGULAR FLOWS,” IN PROCEEDINGS OF THE 37TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, VOL. 119, PP. 4673–4681, PMLR, 2020

References

[Laszkiewicz 2022] Laszkiewicz, M., Lederer, J., Fischer, A.: Marginal tail-adaptive normalizing flows. In: International Conference on Machine Learning. vol. 162, pp. 12020–12048.

PMLR (2022)

[Liang 2022] Liang, F.T., Hodgkinson, L., Mahoney, M.W.: Fat-tailed variational inference with anisotropic tail adaptive flows. In: International Conference on Machine Learning. vol. 162, pp. 13257–13270. PMLR (2022)

[**Pickands 1975**] **Pickands** III, J.: Statistical inference using extreme order statistics. the Annals of

Statistics pp. 119–131 (1975)

References

[Coles 2001] Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer London, London (2001)

[Anadan, Dalmia] Kriti Anandan, Swaraj Dalmia Fig 5, https://tech.skit.ai/normalizing-flows-part-2/

[McDonald 2022] McDonald, A., Tan, P.-N., and Luo, L. Comet flows: Towards generative modeling of multivariate extremes and tail dependence. In International Joint Conference on Artificial Intelligence, 2022

Normalising Flows: Training

Analytic Density:

$$q(x;\ \theta) = q_z(T^{-1}(x;\ \theta))\,|\det J_{T^{-1}}|$$

Implied Density

Inverse Transformation

Base Density

Jacobian Determinant

Fit by maximum likelihood:

Observed data $\{x_i\}_{i=0,\dots,n}$

Maximise $L(\theta) = \sum_i \log q(x_i;\ \theta)$