Marginal Tail-Adaptive Normalizing Flows

Mike Laszkiewicz[1,2], Johannes Lederer[1], Asja Fischer[2]

1 Faculty of Mathematics, Ruhr University Bochum
2 Center of Computer Science, Ruhr University Bochum

International Conference on Machine Learning 2022 – July 17-23, 2022

Overview

Research Question

How good are **Normalizing Flows** at modeling **heavy-tailed distributions?**

Autoregressive Lipschitz Flows

$$\mathbf{z} \sim \begin{cases} \mathcal{N}(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are light-tailed} \\ t_\nu(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are heavy-tailed} \end{cases}$$

Can we do better? Yes!

Marginal Tail-Adaptive Flows can model distributions with both, **light-tailed**, as well as **heavy-tailed** marginals!

Overview

Research Question

How good are **Normalizing Flows** at modeling **heavy-tailed distributions?**

Autoregressive Lipschitz Flows

$$\mathbf{z} \sim \begin{cases} \mathcal{N}(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are light-tailed} \\ t_\nu(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are heavy-tailed} \end{cases}$$

Can we do better? Yes!

Marginal Tail-Adaptive Flows can model distributions with both, **light-tailed**, as well as **heavy-tailed** marginals!

RUB

Overview

Research Question

How good are **Normalizing Flows** at modeling **heavy-tailed distributions?**

Autoregressive Lipschitz Flows

$$\mathbf{z} \sim \begin{cases} \mathcal{N}(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are light-tailed} \\ t_\nu(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are heavy-tailed} \end{cases}$$

Can we do better? Yes!

Marginal Tail-Adaptive Flows can model distributions with both, **light-tailed**, as well as **heavy-tailed** marginals!

RUB

Overview

Research Question

How good are **Normalizing Flows** at modeling **heavy-tailed distributions?**

Autoregressive Lipschitz Flows

$$\mathbf{z} \sim \begin{cases} \mathcal{N}(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are light-tailed} \\ t_\nu(0, I) & \Rightarrow \text{All marginals of } T_\theta(\mathbf{z}) \text{ are heavy-tailed} \end{cases}$$

Can we do better? Yes!

Marginal Tail-Adaptive Flows can model distributions with both, **light-tailed**, as well as **heavy-tailed** marginals!

RUB

Normalizing Flows and Heavy Tails

Change of Variables

Let $\mathbf{z}$ and $\mathbf{x} := T(\mathbf{z})$ be real random variables with PDF p and q, respectively, and T a diffeomorphism. Then,

$$p(x) = q\left(T^{-1}(x)\right)\left|\det J_{T^{-1}}(x)\right| \quad \forall x \in \mathbb{R}^D.$$

Idea of Normalizing Flows

Fix $\mathbf{z}$ and learn T_θ such that $\mathbf{x} \approx T_\theta(\mathbf{z})$.

Informally:

Heavy-tailed distributions decay slower, which allows the generation of **extreme samples**.

RUB

Normalizing Flows and Heavy Tails

Change of Variables

Let $\mathbf{z}$ and $\mathbf{x} := T(\mathbf{z})$ be real random variables with PDF p and q, respectively, and T a diffeomorphism. Then,

$$p(x) = q\big(T^{-1}(x)\big)\big|\det J_{T^{-1}}(x)\big| \quad \forall x \in \mathbb{R}^D.$$

Idea of Normalizing Flows

Fix $\mathbf{z}$ and learn T_θ such that $\mathbf{x} \approx T_\theta(\mathbf{z})$.

Informally:

Heavy-tailed distributions decay slower, which allows the generation of **extreme samples**.

Normalizing Flows and Heavy Tails

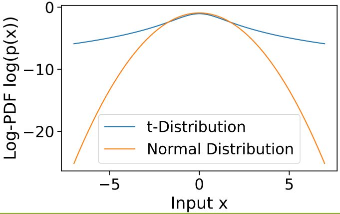
Change of Variables

Let $\mathbf{z}$ and $\mathbf{x} := T(\mathbf{z})$ be real random variables with PDF p and q, respectively, and T a diffeomorphism. Then,

$$p(x) = q\big(T^{-1}(x)\big)\big|\det J_{T^{-1}}(x)\big| \quad \forall x \in \mathbb{R}^D.$$

Idea of Normalizing Flows

Fix $\mathbf{z}$ and learn T_θ such that $\mathbf{x} \approx T_\theta(\mathbf{z})$.



Informally:

Heavy-tailed distributions decay slower, which allows the generation of **extreme samples**.

Normalizing Flows and Heavy Tails

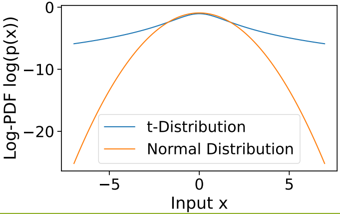
Change of Variables

Let $\mathbf{z}$ and $\mathbf{x} := T(\mathbf{z})$ be real random variables with PDF p and q, respectively, and T a diffeomorphism. Then,

$$p(x) = q\big(T^{-1}(x)\big)\big|\det J_{T^{-1}}(x)\big| \quad \forall x \in \mathbb{R}^D.$$

Idea of Normalizing Flows

Fix $\mathbf{z}$ and learn T_θ such that $\mathbf{x} \approx T_\theta(\mathbf{z})$.



Informally:

Heavy-tailed distributions decay slower, which allows the generation of **extreme samples**.

Tail Behavior of autoregressive NFs

Tail-Adaptive Flows, Jaini et al., 2020; Informal Take-Away:

If $\|\mathbf{z}\|$ is **light-tailed**, then $\|T_\theta(\mathbf{z})\|$ is also **light-tailed**.

Tail-Adaptive Flows (TAF)

Replace $\mathbf{z} \sim \mathcal{N}(0, I)$ by a multivariate t_ν**-distribution** with learnable degree of freedom ν.

Proposition 3.3 from our paper; Informal

The marginals of **TAF** are **all heavy-tailed**.

Tail Behavior of autoregressive NFs

Tail-Adaptive Flows, Jaini et al., 2020; Informal Take-Away:

If $\|\mathbf{z}\|$ is **light-tailed**, then $\|T_\theta(\mathbf{z})\|$ is also **light-tailed**.

Tail-Adaptive Flows (TAF)

Replace $\mathbf{z} \sim \mathcal{N}(0, I)$ by a multivariate $\mathbf{t}_\nu$**-distribution** with learnable degree of freedom ν.

Proposition 3.3 from our paper; Informal

The marginals of **TAF** are **all heavy-tailed**.

Tail Behavior of autoregressive NFs

Tail-Adaptive Flows, Jaini et al., 2020; Informal Take-Away:

If $\|\mathbf{z}\|$ is **light-tailed**, then $\|T_\theta(\mathbf{z})\|$ is also **light-tailed**.

Tail-Adaptive Flows (TAF)

Replace $\mathbf{z} \sim \mathcal{N}(0, I)$ by a multivariate t_ν**-distribution** with learnable degree of freedom ν.

Proposition 3.3 from our paper; Informal

The marginals of **TAF** are **all heavy-tailed**.

Marginal Tail-Adaptive Flows

Let $\mathbf{z}$ be a random variable that is j-light-tailed for $j \in \{1, \ldots, d_l\}$ and j-heavy-tailed for $j \in \{d_l + 1, \ldots, D\}$ Under certain conditions, we obtain:

Theorem (Tail Behavior after applying Autoregressive Transformations)

If T is an autoregressive map, then $\mathbf{z}$ and $T(\mathbf{z})$ have the **same marginal tail behavior**.

Theorem (Tail Behavior after applying Block-Triangular Linear Transformations)

Consider a block-diagonal invertible matrix

$$W = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} \tag{1}$$

with $A \in \mathbb{R}^{d_l \times d_l}$, $B \in \mathbb{R}^{(D-d_l) \times d_l}$, $C \in \mathbb{R}^{(D-d_l) \times (D-d_l)}$ and 0 is a zero matrix of size $d \times (D - d_l)$. Then, it follows that $\mathbf{z}$ and $W\mathbf{z}$ have the **same marginal tail behavior**.

Marginal Tail-Adaptive Flows

Let $\mathbf{z}$ be a random variable that is j-light-tailed for $j \in \{1, \dots, d_l\}$ and j-heavy-tailed for $j \in \{d_l + 1, \dots, D\}$ Under certain conditions, we obtain:

Theorem (Tail Behavior after applying Autoregressive Transformations)

If T is an autoregressive map, then $\mathbf{z}$ and $T(\mathbf{z})$ have the **same marginal tail behavior**.

Theorem (Tail Behavior after applying Block-Triangular Linear Transformations)

Consider a block-diagonal invertible matrix

$$W = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} \quad (1)$$

with $A \in \mathbb{R}^{d_l \times d_l}$, $B \in \mathbb{R}^{(D-d_l) \times d_l}$, $C \in \mathbb{R}^{(D-d_l) \times (D-d_l)}$ and 0 is a zero matrix of size $d \times (D - d_l)$. Then, it follows that $\mathbf{z}$ and $W\mathbf{z}$ have the **same marginal tail behavior**.

Marginal Tail-Adaptive Flows

Let $\mathbf{z}$ be a random variable that is j-light-tailed for $j \in \{1, \ldots, d_l\}$ and j-heavy-tailed for $j \in \{d_l + 1, \ldots, D\}$ Under certain conditions, we obtain:

Theorem (Tail Behavior after applying Autoregressive Transformations)

If T is an autoregressive map, then $\mathbf{z}$ and $T(\mathbf{z})$ have the **same marginal tail behavior***.*

Theorem (Tail Behavior after applying Block-Triangular Linear Transformations)

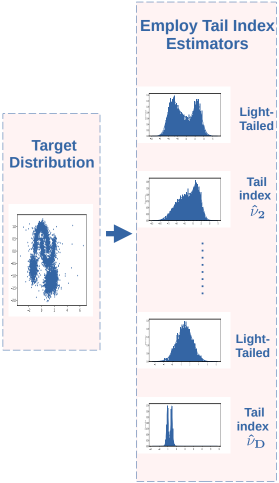
Consider a block-diagonal invertible matrix

$$W = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} \tag{1}$$

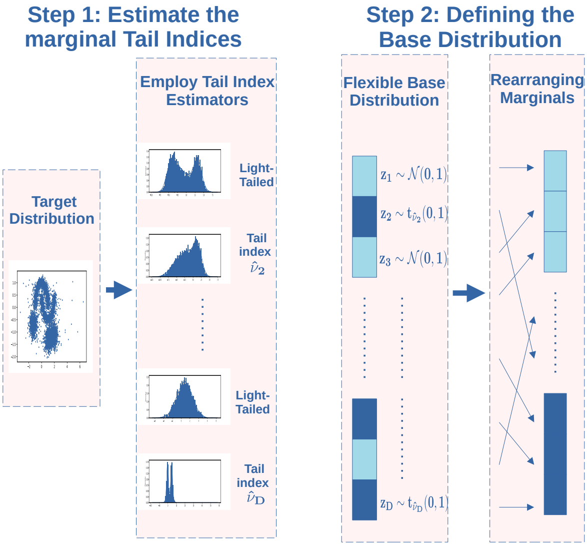
with $A \in \mathbb{R}^{d_l \times d_l}$, $B \in \mathbb{R}^{(D-d_l) \times d_l}$, $C \in \mathbb{R}^{(D-d_l) \times (D-d_l)}$ and 0 is a zero matrix of size $d \times (D - d_l)$. Then, it follows that $\mathbf{z}$ and $W\mathbf{z}$ have the **same marginal tail behavior***.*

Marginal Tail-Adaptive Flows

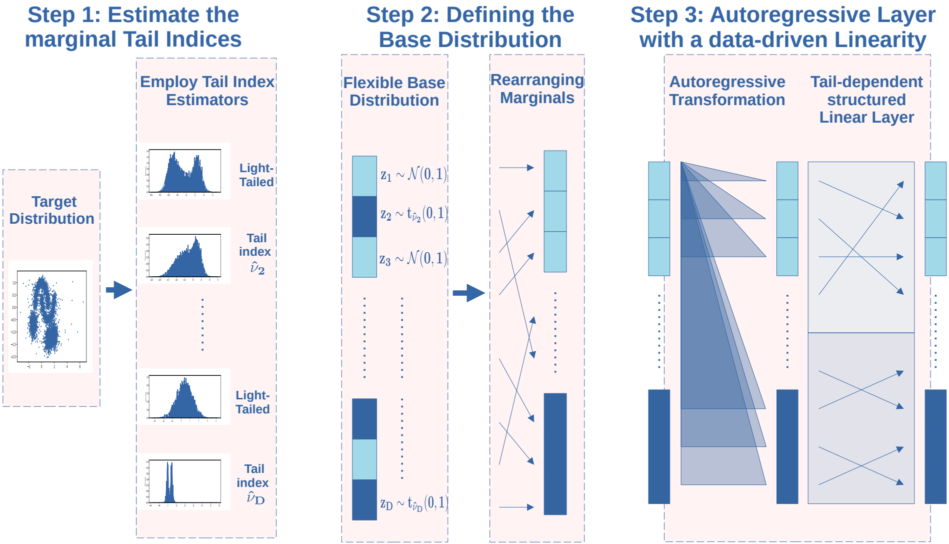
Step 1: Estimate the marginal Tail Indices



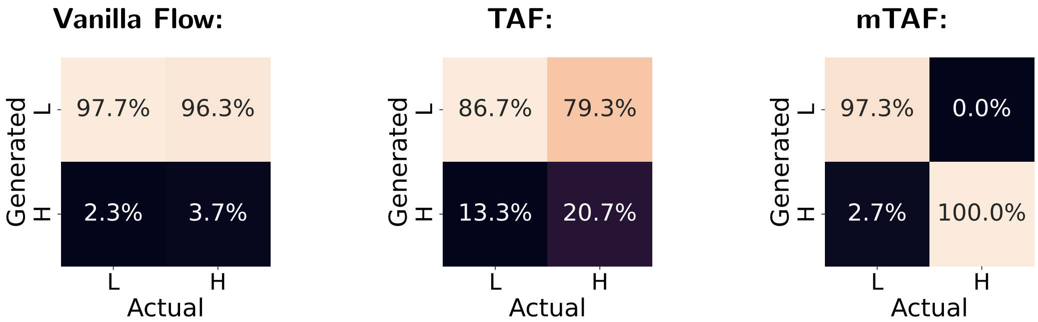
Marginal Tail-Adaptive Flows



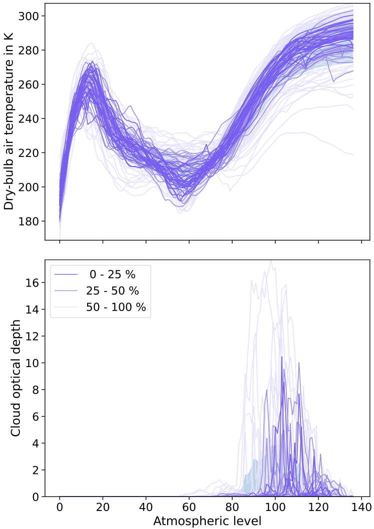
Marginal Tail-Adaptive Flows



Experiments: Generating mixed-tailed Samples



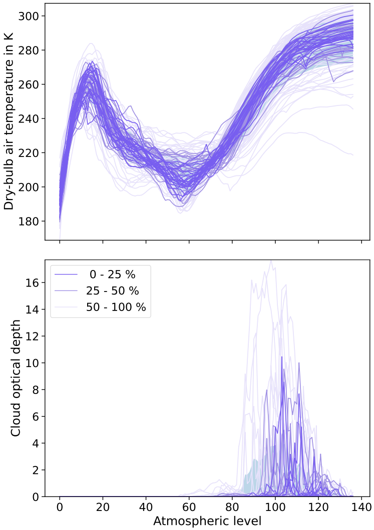
Experiments: Learning from Weather and Climate Data



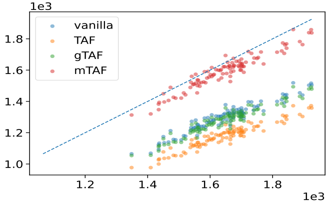
Randomly projected Standard deviation:

Randomly projected 1%-Quantiles

Experiments: Learning from Weather and Climate Data

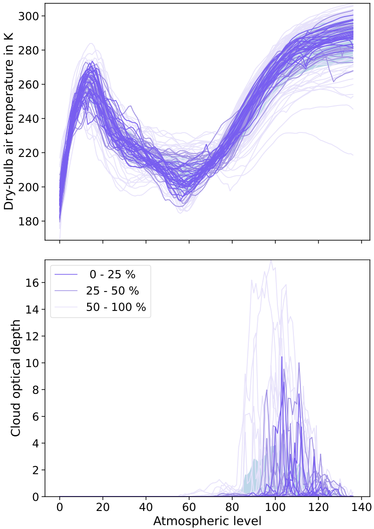


Randomly projected Standard deviation:

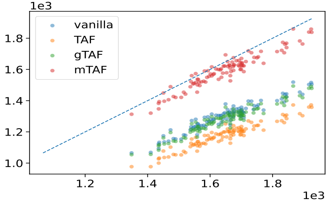


Randomly projected 1%-Quantiles

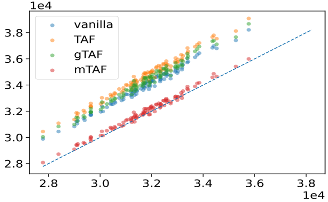
Experiments: Learning from Weather and Climate Data



Randomly projected Standard deviation:



Randomly projected 1%-Quantiles



Summary

If you want to learn a **generative model** for data with **heavy**- as well as **light-tailed marginals**, you should contemplate to use **mTAF**

arXiv:

github:

RUB