# A comparison of generative deep learning methods for multivariate angular simulation

Jakob Wessel (Exeter), Callum Murphy-Barltrop (TU Dresden)
**Emma Simpson (UCL)**

Workshop on "Generative AI Modelling for Extreme Events"
University of Edinburgh
13 June 2025

# Motivation

Angular-radial representations are reasonably common in the study of multivariate extremes.

For a $d$-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$, it may be convenient to consider

$$R = \|\boldsymbol{X}\|_A, \qquad \boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|_B,$$

i.e., a *radial* component and a *(pseudo-)angular* vector.

# Motivation - multivariate regular variation

Take $\boldsymbol{X}$ to have some common, heavy-tailed (e.g., Fréchet) margins, and use the $L_1$ norm in both cases.

Then, $R > 0$ and $\boldsymbol{W}$ takes values on an angular simplex

$$\left\{ \boldsymbol{w} \in [0,1]^d : \sum_{i=1}^{d} w_i = 1 \right\}.$$

# Motivation - multivariate regular variation

Take $\boldsymbol{X}$ to have some common, heavy-tailed (e.g., Fréchet) margins, and use the $L_1$ norm in both cases.

Then, $R > 0$ and $\boldsymbol{W}$ takes values on an angular simplex

$$\left\{ \boldsymbol{w} \in [0,1]^d : \sum_{i=1}^{d} w_i = 1 \right\}.$$

Under the assumption of multivariate regular variation,

$$\lim_{t \to \infty} \Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = \frac{H(B)}{r}.$$

## Motivation - multivariate regular variation

Take $\boldsymbol{X}$ to have some common, heavy-tailed (e.g., Fréchet) margins, and use the $L_1$ norm in both cases.

Then, $R > 0$ and $\boldsymbol{W}$ takes values on an angular simplex

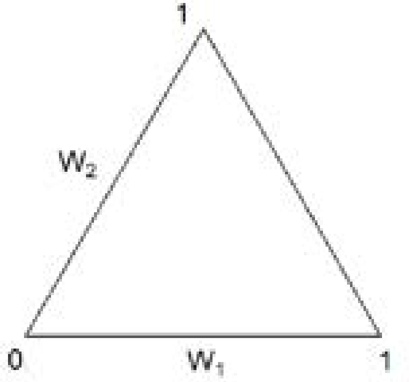
$$\left\{ \boldsymbol{w} \in [0,1]^d : \sum_{i=1}^{d} w_i = 1 \right\}.$$

Under the assumption of multivariate regular variation,

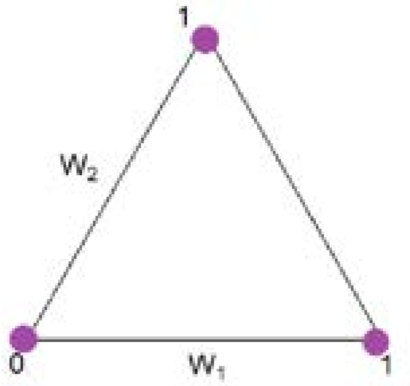
$$\lim_{t \to \infty} \Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = \frac{H(B)}{r}.$$

- So, $R$ and $\boldsymbol{W}$ become independent in the limit.
- The limiting spectral measure $H$ gives information about the extremal dependence structure of $\boldsymbol{X}$.

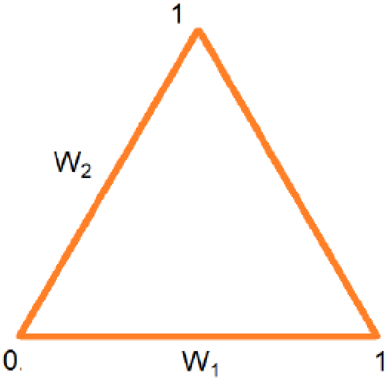
# Mass on the angular simplex ($d = 3$)



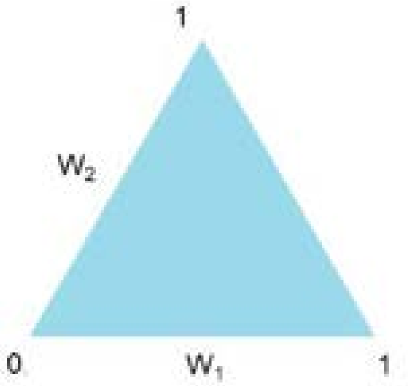
# Mass on the angular simplex ($d = 3$)



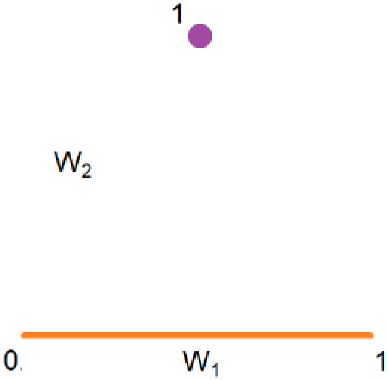
# Mass on the angular simplex ($d = 3$)



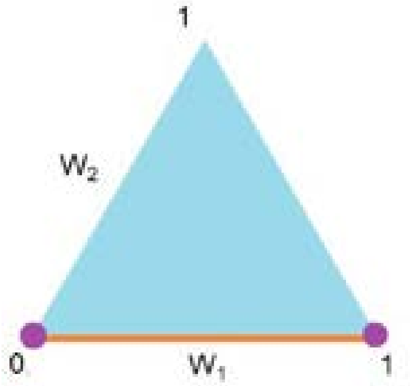
# Mass on the angular simplex ($d = 3$)



# Mass on the angular simplex ($d = 3$)



# Mass on the angular simplex ($d = 3$)



# Motivation - radial-angular multivariate extremes

More recently, there has been growing interest in radial-angular approaches for multivariate extremes, e.g.:

- geometric extremes;
- SPAR models.

# Motivation - radial-angular multivariate extremes

More recently, there has been growing interest in radial-angular approaches for multivariate extremes, e.g.:

- geometric extremes;
- SPAR models.

For inference and extrapolation, both require modelling or simulation of an angular component.

# Motivation - SPAR models (Mackay and Jonathan, 2023)

Consider a decomposition of the density of $(R, \boldsymbol{W})$ via

$$f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w})\, f_{R|\boldsymbol{W}}(r|\boldsymbol{w}).$$

Then, the multivariate extremes problem relies on estimation of an angular density $f_{\boldsymbol{W}}$ and a model for the tail of $R|\boldsymbol{W}$, i.e., for $r > u(\boldsymbol{w})$

$$f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = \tau f_{\boldsymbol{W}}(\boldsymbol{w}) f_{GPD}(r - u(\boldsymbol{w}); \sigma(\boldsymbol{w}), \xi(\boldsymbol{w})),$$

where $\tau = \Pr(R > u(\boldsymbol{w}) | \boldsymbol{W} = \boldsymbol{w})$ is close to 0.

# Motivation - SPAR models (Mackay and Jonathan, 2023)

Consider a decomposition of the density of $(R, \boldsymbol{W})$ via

$$f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w})\, f_{R|\boldsymbol{W}}(r|\boldsymbol{w}).$$

Then, the multivariate extremes problem relies on estimation of an angular density $f_{\boldsymbol{W}}$ and a model for the tail of $R|\boldsymbol{W}$, i.e., for $r > u(\boldsymbol{w})$

$$f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = \tau f_{\boldsymbol{W}}(\boldsymbol{w}) f_{GPD}(r - u(\boldsymbol{w}); \sigma(\boldsymbol{w}), \xi(\boldsymbol{w})),$$

where $\tau = \Pr(R > u(\boldsymbol{w})|\boldsymbol{W} = \boldsymbol{w})$ is close to 0.

- Extrapolation can be achieved by drawing from the angular distribution, then simulating from the conditional GPD.

# The role of deep learning

- Empirical approaches could be used to simulate from the angular distribution.
  - This becomes infeasible in higher dimensions.
- Parametric models could be used instead.
  - These may lack the flexibility required to cover the range of possible structures.

# The role of deep learning

- Empirical approaches could be used to simulate from the angular distribution.
  - This becomes infeasible in higher dimensions.
- Parametric models could be used instead.
  - These may lack the flexibility required to cover the range of possible structures.

- Simulation using generative deep learning approaches could be used.
  - This can estimate the angular distribution without specifying a density.
  - Is less restrictive than relying on exact observations.

# Our focus - simulating angular variables

- So, modelling and simulating angular data can play an important role in multivariate extreme value analysis.
- This may also be useful in other settings where angular data arise.

# Our focus - simulating angular variables

- So, modelling and simulating angular data can play an important role in multivariate extreme value analysis.
- This may also be useful in other settings where angular data arise.

**Our question:**
**Can generative deep learning methods be helpful for this task?**

## Spherical coordinates

We focus on the spherical coordinates $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_{d-1})$

$$\begin{aligned}
\Theta_1 &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2 + \cdots + {X_2}^2}, X_1\right), \\
\Theta_2 &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2 + \cdots + {X_3}^2}, X_2\right), \\
&\vdots \\
\Theta_{d-2} &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2}, X_{d-2}\right), \\
\Theta_{d-1} &= \operatorname{atan2}\left(X_d, X_{d-1}\right),
\end{aligned}$$

where $\Theta_1, \ldots \Theta_{d-2} \in [0, \pi]$ and $\Theta_{d-1} \in (-\pi, \pi]$. (Blumenson, 1960)

# Deep learning approaches

- Generative adversarial networks (GANs) - Goodfellow et al. (2014)
- Normalizing flows
  - Neural spline flows (NFNSFs) - Durkan et al. (2019)
  - Masked autoregressive flows (NFMAFs) - Papamakarios et al. (2017)
- Flow matching (FM) - Lipman et al. (2022)

# Deep learning approaches

- Generative adversarial networks (GANs) - Goodfellow et al. (2014)
- Normalizing flows
  - Neural spline flows (NFNSFs) - Durkan et al. (2019)
  - Masked autoregressive flows (NFMAFs) - Papamakarios et al. (2017)
- Flow matching (FM) - Lipman et al. (2022)
- Maximum mean discrepancy networks/Energy score networks

# Adaptations for angular variables - GANs

- Feedforward neural networks for the generator and the discriminator.

- To ensure the spherical coordinates lie in the correct range, we specify activation functions after the last generator layer as

$$\theta_i \leftarrow \pi \cdot \frac{1}{1+e^{-x_i}} \text{ for } i = 1, \ldots, d-2,$$
$$\theta_{d-1} \leftarrow \pi \cdot \tanh x_{d-1}.$$

- Note: circular wrapping resulted in some training instabilities.

# Adaptations for angular variables - NFs

- For both the NFNSF and NFMAF approaches, we use a transformation of the form

$$\theta_i \leftarrow \pi \cdot \Phi(x_i) \text{ for } i = 1, \ldots, d-2,$$
$$\theta_{d-1} \leftarrow (x_{d-1} \bmod 2\pi) - \pi.$$

- In these cases, the sigmoid has slower convergence at 0 and 1 than the Gaussian CDF, which led to poor results near the endpoints of $[0, \pi]$.
- Circular wrapping for the final component worked well in this case.

## Adaptations for angular variables - FM

- For flow matching, we don't need to map onto the angular space, since flows can be defined directly on the sphere.

# Baseline approach - mixture of vMF distributions

- The von Mises-Fisher (vMF) distribution has density

$$f_{vMF}(\boldsymbol{x} \mid \boldsymbol{\mu}, \kappa) = c_d(\kappa) e^{\kappa \boldsymbol{\mu}^T \boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{S}^{d-1},$$

for some concentration $\kappa \geq 0$, mean direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and normalising constant $c_d(\kappa)$.

- This is an angular extension of a multivariate Gaussian distribution.

# Baseline approach - mixture of vMF distributions

- The von Mises-Fisher (vMF) distribution has density

$$f_{vMF}(\boldsymbol{x} \mid \boldsymbol{\mu}, \kappa) = c_d(\kappa) e^{\kappa \boldsymbol{\mu}^T \boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{S}^{d-1},$$

  for some concentration $\kappa \geq 0$, mean direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and normalising constant $c_d(\kappa)$.

- This is an angular extension of a multivariate Gaussian distribution.

- A mixture of vMF distributions provides a flexible, parametric approach to modelling angular data. This provides a baseline method for comparing our deep learning approaches.

# Simulation study - overview

To cover a wide range of examples, we consider the following options:

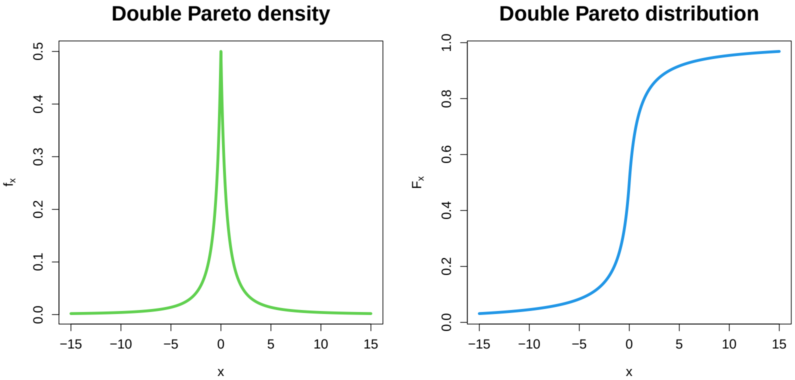
- dimensions $d = 5$ and $d = 10$;
- datasets of size $n = 1,000$, $n = 10,000$ and $n = 100,000$;
- light and heavy-tailed marginal distributions;
- five different dependence (copula) structures.

# Simulation study - marginal models

To allow the potential for mass in all orthants, we choose common marginal models with support on $(-\infty, \infty)$.

- A standard Laplace distribution, having light upper and lower tails.
- A 'double Pareto' distribution, with heavy upper and lower tails.

# Simulation study - marginal models



## Simulation study - copula models

We consider copulas with different types of dependence structure, taking into account AD/AI and sparsity:

- a Gaussian copula (AI);
- a mixture of a Gaussian copula and a student-$t$ copula (AI);
- a logistic copula (AD);
- a mixture distribution of a logistic copula and an independence copula (more complex EDS);
- a 'sparse' Gaussian copula specified such that the variables are clustered into groups of dependent variables with independence between groups (AI).

# Simulation study - hyperparameter optimisation

- Bespoke hyperparameter tuning for each individual setting is very time consuming and therefore infeasible.
- We instead optimised the architectures for one specific case - the sparse Gaussian copula with double Pareto margins - and using a data example.

# Simulation study - hyperparameter optimisation

- Bespoke hyperparameter tuning for each individual setting is very time consuming and therefore infeasible.
- We instead optimised the architectures for one specific case - the sparse Gaussian copula with double Pareto margins - and using a data example.
- This seemed reasonable as long as the hyperparameters that were selected allowed for some flexibility.
- Details of our suggested hyperparameters and architectures are in the paper.

# Simulation study - evaluation strategies

We consider a range of evaluation metrics to assess both marginal and joint structure.

- Numerical metrics - circular CRPS
- Histograms and QQ-plots for spherical margins
- Scatterplots between pairs of spherical angles
- Orthant probability plots to assess full structure

# Simulation study - general takeaways

- Using the cCRPS, the baseline approach gave the best results in the majority of cases, but the results are very close.
  - Seems difficult to distinguish between methods using cCRPS.
  - Possibility that cCRPS focuses too much on the marginal fits.

# Simulation study - general takeaways

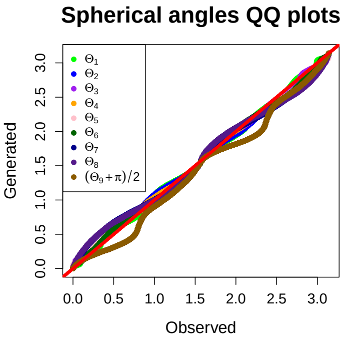
- Using the cCRPS, the baseline approach gave the best results in the majority of cases, but the results are very close.
  - Seems difficult to distinguish between methods using cCRPS.
  - Possibility that cCRPS focuses too much on the marginal fits.
- The deep learning methods generally do well at recreating marginal *and* dependence structures.
  - There are some cases where the GAN doesn't do so well.
  - But there is no clear overall "winner".

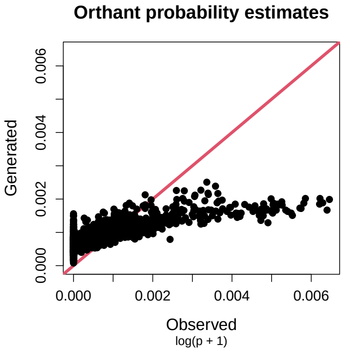
## Simulation study - general takeaways

- Using the cCRPS, the baseline approach gave the best results in the majority of cases, but the results are very close.
  - Seems difficult to distinguish between methods using cCRPS.
  - Possibility that cCRPS focuses too much on the marginal fits.
- The deep learning methods generally do well at recreating marginal *and* dependence structures.
  - There are some cases where the GAN doesn't do so well.
  - But there is no clear overall "winner".
- Marginal distributions do have an effect.
  - In some cases, we do well for Laplace margins but poorly for the equivalent double Pareto case.

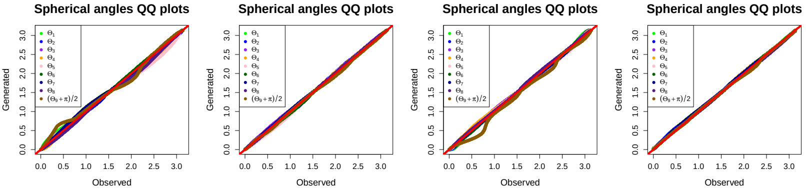
# Simulation study - some example results

Sparse Gaussian copula, Double Pareto margins, $d = 10$ and $n = 100,000$

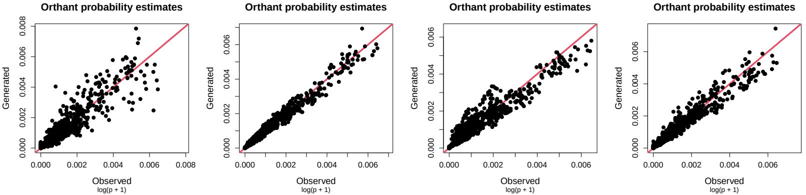


# Simulation study - some example results



# Simulation study - some example results



# Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.

# Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.
- The baseline approach was generally good at capturing marginal distributions, but struggled with some complex dependence structures.

# Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.
- The baseline approach was generally good at capturing marginal distributions, but struggled with some complex dependence structures.
- No methods can be completely discounted, and there is no individual 'best' approach. So model checks and validation are key.

# Avenues for further work

- Validation and model comparison is somewhat difficult.
  - It would be useful to have more metrics for evaluation in angular/spherical contexts.
  - Model rankings depend on the metric used.

# Avenues for further work

- Validation and model comparison is somewhat difficult.
  - It would be useful to have more metrics for evaluation in angular/spherical contexts.
  - Model rankings depend on the metric used.
- Hyperparameter tuning is time consuming.
  - Theoretical developments or general guidelines related to this would be helpful.

# Avenues for further work

- Validation and model comparison is somewhat difficult.
  - It would be useful to have more metrics for evaluation in angular/spherical contexts.
  - Model rankings depend on the metric used.
- Hyperparameter tuning is time consuming.
  - Theoretical developments or general guidelines related to this would be helpful.
- Improvements in computational efficiency could improve the feasibility of hyperparameter optimisation and uncertainty assessment.

# Thank you!

## Based on the paper:

- Wessel, J. B., Murphy-Barltrop, C. J. R. and Simpson, E. S. (2025). A comparison of generative deep learning methods for multivariate angular simulation. *arXiv:2504.21505.*
- Accompanying code available from `https://github.com/callumbarltrop/DeGeMoH`

---

## Other references:


- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2019). Neural spline flows. *NeurIPS.*
- Goodfellow, I. J., et al. (2014). Generative adversarial nets. *NeurIPS.*
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. (2022). Flow matching for generative modeling. *ICLR.*
- Mackay, E. and Jonathan, P. (2023). Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv:2310.12711.*
- Papamakarios, G., Pavlakou, T., and Murray, I. (2017). Masked autoregressive flow for density estimation. *NeurIPS.*


**Contact:** emma.simpson@ucl.ac.uk