Analysis and Synthesis of Extreme Events

Vahid Tarokh
Duke University

AFOSR MURI Kickoff

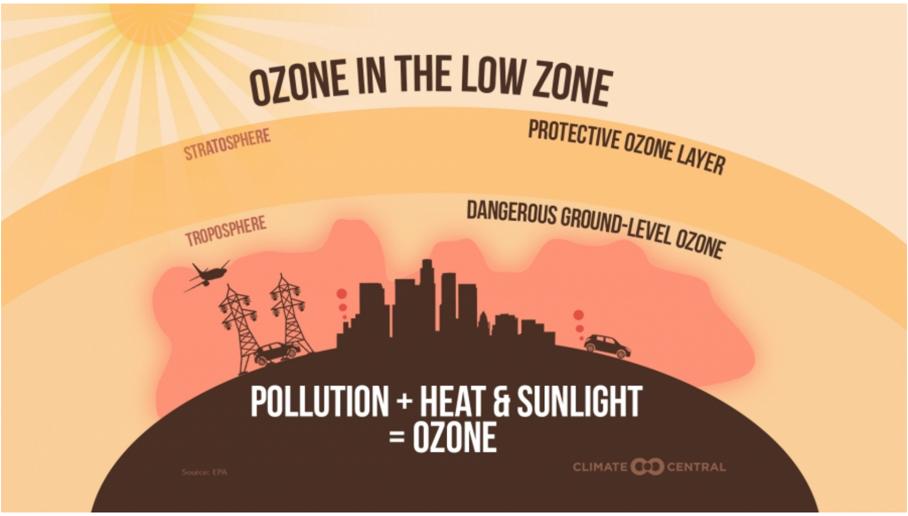
Grant No. FA9550-20-1-0397

Feb. 16, 2021

Introduction

- Extreme events are events that occur with low frequency but may have a great impact and very high associated costs.
  - Natural events: Floods, Hurricanes, Earthquakes, Tsunamis;
  - Human made events: Wars, Violent conflicts, Financial Crisis;
  - Complex interactions: Global warming, Epidemic disease spread
- Available Data at Duke
  - Data of CA Earthquakes
  - Data of Extreme Ozone Level Days (stations around Boston)
  - Seizure Data

Extreme Events



Extreme Events

- Having an idea of the probability of occurrence of rare events can give some insight on how to allocate resources more efficiently:
  - For instance, a good estimate of the probability of natural disasters (floods, hurricanes, earthquakes) can be used to define the budget for prevention and early detection of these.
  - Understanding the causality between rare events:
  - Do some earthquakes in Nevada cause earth-quakes in CA?
  - Can NY stock market crash lead to a crash in Hong Kong Stock exchange?

Research Agenda

- Multivariate Extreme Events
  - Modeling dependence between Extreme Events
  - Generation of Extreme Events samples with known dependence functions
  - Generation of new Extreme Events samples given training data
- Infinite dimensional case & Spatial Extreme Events
- Max-stable Processes
    - Max-stable processes governed by PDEs
  - Distribution of Extremes of solution to PDEs
- Applications to Scenarios of Interest

Extreme Events and PDEs

Distributions of Extremes of PDEs

- Suppose we wish to consider the mapping of distributions of initial conditions to solutions of PDEs over time
- Parameterize mapping $T_\theta(\cdot|t)$ from distribution of initial conditions to solutions while encoding properties of the PDE in the mapping
- Induces latent dynamics on latent distributions of initial conditions
- **Takeaway:** By learning diffeomorphisms between distributions of initial conditions and solutions, we can characterize the extreme events by their initial conditions

$$z_t \sim P_{0,t}(\cdot) \tag{7}$$

$$x_t \sim T_\theta(P_{0,t}(\cdot)|t) \quad \text{where } \mathcal{D}(x) = 0 \tag{8}$$

$\mathcal{D}$ are the differential operators assumed to be governing the data

Copulas, Extreme Value Theory, and Pickands Dependence Function

Copula

Sklar's Theorem: Every multivariate cumulative distribution function $F(x_1, \ldots, x_d) = \Pr[X_1 \leq x_1, \cdots, X_d \leq x_d]$ of a random vector $(X_1, X_2, \cdots, X_d)$ can be expressed in terms of its marginals $F_i(x_i) = \Pr[X_i \leq x_i]$ and a copula $C\left(F_1(x_1), \ldots, F_d(x_d)\right)$ defined as

$$\begin{aligned} F(x_1, \ldots, x_d) &= \mathbb{P}\left[X^{(1)} \leq x_1, \cdots, X^{(d)} \leq x_d\right] \\ &= \mathbb{P}\left[F_1(X^{(1)}) \leq F_1(x_1), \cdots, F_d(X^{(d)}) \leq F_d(x_d)\right] \\ &= C\left(u_1, \cdots, u_d\right), \end{aligned} \tag{1}$$

where $u_k = F_k(x_k)$.

- Clearly $C\left(u_1, \cdots, u_d\right)$ maps $[0,1]^d$ to $[0,1]$.

Extreme Value Theory

Let X_i be i.i.d random variables $\sim F$. We are interested in characterizing the behavior of maxima that's the distribution of $M_n = \max_{1 \leq i \leq n} X_i$. Then, similar to the central limit theorem, it is possible to find a location and scale b_n and a_n s.t. $\mathbb{P}[(M_n - b_n)/a_n \leq x] \to H(x)$ where $H(x)$ is non-degenerate. In this case, we say that $F \in \mathrm{MDA}(H)$.

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), & \text{if } \xi \neq 0 \\ \exp\left(-e^{-x}\right), & \text{if } \xi = 0 \end{cases} \tag{2}$$

- $\xi < 0$: the Weibull distribution (short-tailed).
- $\xi = 0$: The Gumbel distribution (tail decays exponentially).
- $\xi > 0$: The Fréchet distribution (heavy-tailed).

Extreme Value Copula

Let $\left(X_i^{(1)}, \ldots, X_i^{(d)}\right)_{i=1}^{n}$ be samples of i.i.d random vectors with common distribution F. We would like to examine the asymptotic distribution of componentwise maxima $M_n = \left(M_n^{(1)}, \ldots, M_n^{(d)}\right)$ where $M_n^{(k)} = \bigvee_{i=1}^{n} X_i^{(k)}$. It follows that the copula of M_n is given by:

$$
\begin{aligned}
C_n(u_1, \ldots, u_d) &= C_n(F_n(M_n^{(1)}) \leq u_1, \ldots, F_n(M_n^{(d)}) \leq u_d) \\
&= C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n
\end{aligned}
$$

- **C is an extreme value copula if $\exists$ C_F such that** $C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n \to C(u_1, \ldots, u_d)$.
- C is an extreme value copula iff it is max-stable: $C\left(u_1, \ldots, u_d\right) = C(u_1^{1/m}, \ldots, u_d^{1/m})^m$ for all $m \geq 1$.
- The extreme value copula coincides with the class of limit distributions with non-degenerate margins of $\left(\frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}}, \ldots, \frac{M_n^{(d)} - b_n^{(d)}}{a_n^{(d)}}\right)$.

Pickands Dependence Function

An extreme value copula C can be expressed in terms of the Pickands dependence function $A : \Delta_{d-1} \to [1/d, 1]$ as:

$$C(u_1, \dots, u_d) = \exp\left\{ \left(\sum_{k=1}^{d} \log u_k \right) A \left(\frac{\log u_1}{\sum_{k=1}^{d} \log u_k}, \dots, \frac{\log u_d}{\sum_{k=1}^{d} \log u_k} \right) \right\}$$

where

- A is convex and satisfies $\max_k w_k \leq A(w_1, \dots, w_d) \leq 1$ for all $(w_1, \dots, w_d) \in \Delta_{d-1}$ (d dimensional Simplex).

Modeling and Training Pickands Functions with Input Convex Neural Networks

Input Convex Neural Networks

B. Amos, L. Xu, J. Kolter. **Input Convex Neural Networks**, ICML 2017.

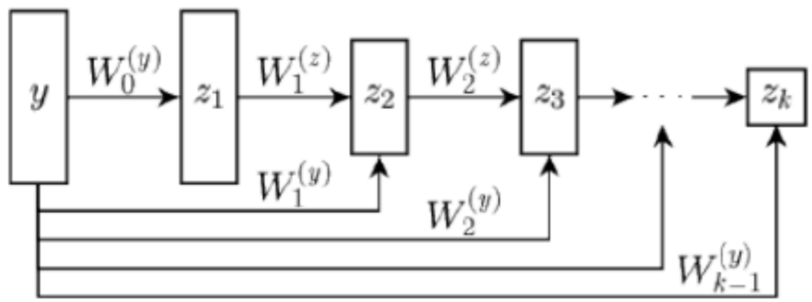


Figure 1. A fully input convex neural network (FICNN).

For $i \in \{0, \cdots, k-1\}$:

$$z_{i+1} = g_i \left[W_i^{(z)} z_i + W_i^{(y)} y + b_i \right]$$

where g_i are convex and non-decreasing, $W_{1:k-1}^{(z)} > 0$ and $W_0^{(z)} = 0$. Based on the following principles:

- $g \circ f$ is convex if f is convex and g is convex and non-decreasing.
- Non-negative sums of convex functions is also convex.

Training Pickands Copula with MLE

Let $\{w_k\}_k \in \Delta_{d-1}$ and Let $\tilde{M}_n^{(k)} = -\log F_k(\bar{M}_n{}^{(k)})$ and $Z_w = \bigwedge_{k=1}^{d} \frac{\tilde{M}_n{}^{(k)}}{w_k}$ Then, we have

$$\begin{aligned} \mathbb{P}[Z_w > z] &= \mathbb{P}[\tilde{M}_n{}^{(k)} > w_k z] \\ &= \mathbb{P}[F_k(\bar{M}_n{}^{(k)}) < \exp(-zw_k)] \\ &= \exp[\sum_{k=1}^{d}(-zw_k)A(\frac{-zw_i}{\sum_{k=1}^{d}(-zw_k)})] \\ &= \exp(-zA(w_1, ..., w_d)). \end{aligned}$$

Therefore $Z_w \sim \mathbf{Exp}(A(w))$.

- **We can model A by an input convex neural network (ICNN).**
- **We can learn the Pickands function A by MLE on samples $\{Z_{w,b}\}_{b=1}^{B}$ and train via SGD.**

Computing Survival Probabilities

We are interested in evaluating the following survival probability:

$$\mathbb{P}\left[M_n^{(1)} > \gamma_1, \cdots, M_n^{(d)} > \gamma_d\right] = \mathbb{P}\left[\bar{M}_n^{(1)} > \bar{\gamma}_1, \cdots, \bar{M}_n^{(d)} > \bar{\gamma}_d\right], \qquad (3)$$

This can not be directly evaluated with the E.V. copula!

Let $G_k(x) := F_k^{-1}(1 - F_k(x))$ for $k \in \{1, \ldots, d\}$, then the random variables $G_k(\bar{M}_n^{(k)})$ and $\bar{M}_n^{(k)}$ have the same marginal CDF F_k, for $k \in \{1, \ldots, d\}$, and

$$\begin{aligned} &\mathbb{P}\left[\bar{M}_n^{(1)} > \bar{\gamma}_1, \cdots, \bar{M}_n^{(d)} > \bar{\gamma}_d\right] \\ &= \mathbb{P}\left[G_1(\bar{M}_n^{(1)}) < G_1(\bar{\gamma}_1), \cdots, G_d(\bar{M}_n^{(d)}) < G_d(\bar{\gamma}_d)\right]. \end{aligned} \qquad (4)$$

$$\begin{aligned} &\mathbb{P}\left[\bar{M}_n^{(1)} > \bar{\gamma}_1, \cdots, \bar{M}_n^{(d)} > \bar{\gamma}_d\right] \\ &= \mathbb{P}\left[1 - F_1(\bar{M}_n^{(1)}) < 1 - F_1(\bar{\gamma}_1), \cdots, 1 - F_d(\bar{M}_n^{(d)}) < 1 - F_d(\bar{\gamma}_d)\right] \\ &= \mathbb{P}\left[G_1(\bar{M}_n^{(1)}) < G_1(\bar{\gamma}_1), \cdots, G_d\bar{M}_n^{(d)}) < G_d(\bar{\gamma}_d)\right], \end{aligned}$$

Recall that $G_k(\bar{M}_n^{(k)}) = F_k^{-1}(1 - F_k(\bar{M}_n^{(k)})) \sim F_k$

Computing Survival Probabilities

Algorithm 1 Estimating survival probabilities with the Pickands dependence function

1: **Input:** $\{\bar{M}_{n,b}^{(k)}\}_{b=1}^{B}$, thresholds: $(\gamma_1, \cdots, \gamma_d)$.

2: Train a model $A(\mathrm{w}; \theta)$ on $\{G_k(\bar{M}_{n,b}^{(k)})\}_{b=1}^{B}$ using MLE and obtain $A(\mathrm{w}; \theta_*)$.

3: Evaluate the Pickands copula:

$$C\left(1 - F_1(\bar{\gamma}_1), \cdots, 1 - F_d(\bar{\gamma}_d)\right).$$

Experiments (Synthetic Data)

- Symmetric logistic model:

$$A_{\mathrm{SL}}(\mathrm{w}) = \left(\sum_{k=1}^{d} w_k^{1/\alpha} \right)^{\alpha}, \ \mathrm{w} \in \Delta_{d-1},$$

- Asymmetric logistic model:

$$A_{\mathrm{ASL}}(\mathrm{w}) = \sum_{b \in \mathcal{P}_d} \left(\sum_{i \in b} (\lambda_{i,b} w_i)^{1/\alpha_b} \right)^{\alpha_b},$$

Experiments (Synthetic Data)

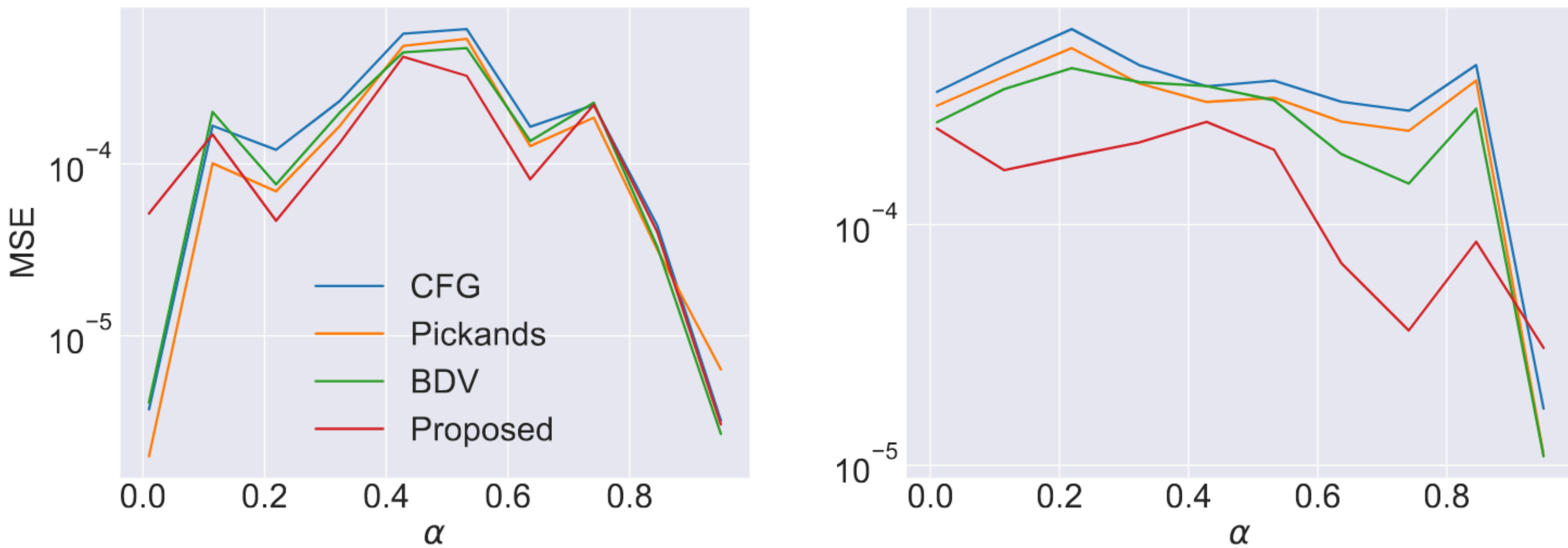


Figure 3.1: MSEs of survival probabilities for A_{SL} (left) and A_{ASL} (right) for the $2d$ case and and $\alpha \in (0, 1)$. CFG, Pickands, BDV, Proposed

Experiments (Synthetic Data)

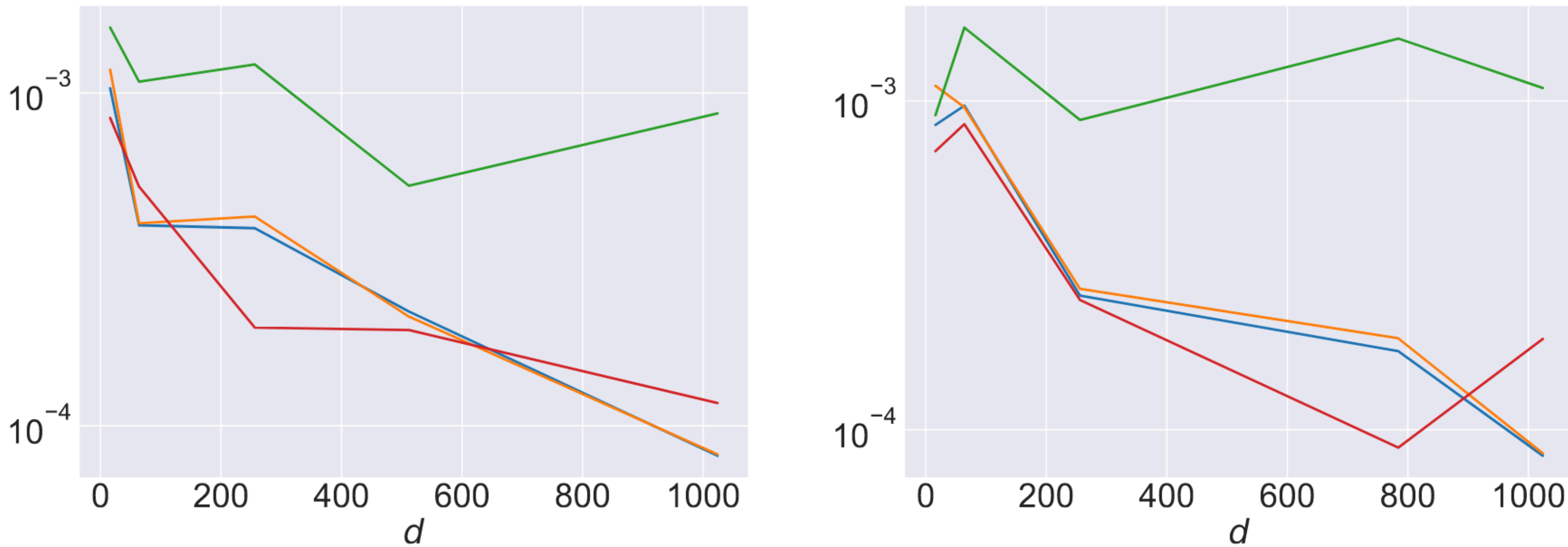


Figure 3.2: $\|\hat{A} - A\|_2^2$ of $A = A_{\mathrm{SL}}$ (left) and $A = A_{\mathrm{ASL}}$ (right) for different dimensions and $\alpha = 0.5$. CFG, Pickands, BDV, Proposed

Experiments (Real Data)

Ozone data

- 4 stations ($d = 4$).
- Daily maxima from Jan 1984 to Dec 1996.
- Train over a sinlge month and compute accuracy over the following one.

California wind data

- 10 locations ($d = 10$).
- Daily maxima from Dec 1989 to Dec 2020.
- Train over a sinlge month and compute accuracy over the following one.

Commodities prices

- 3 commodities (Nickel, Zinc and Copper).
- Monthly maxima of daily prices log returns from Jan 2010 to Dec 2020.
- Train over 2 years and compute accuracy over the following 3 years.

Experiments (Real Data)

	Pickands	CFG	BDV	Proposed
Ozone	$3.16(\pm 4.27)\times 10^{-3}$	$1.59(\pm 3.21)\times 10^{-1}$	$2.98(\pm 4.05)\times 10^{-3}$	$1.54(\pm 2.04)\times 10^{-3}$
Wind	$2.57(\pm 4.53)\times 10^{-3}$	$7.31(\pm 21.8)\times 10^{-2}$	$8.02(\pm 14.0)\times 10^{-4}$	$7.44(\pm 12.9)\times 10^{-4}$
Commodities	$3.12(\pm 4.68)\times 10^{-3}$	$3.12(\pm 4.64)\times 10^{-3}$	$2.97(\pm 4.26)\times 10^{-3}$	$3.03(\pm 3.92)\times 10^{-3}$

Table 1: MSE performance on the test set for different estimators and datasets. All data are for the 75th percentile and above.

Exact Conditional Sampling

Spectral Representation of Pickands Dependence Function

The tail dependence function (and thus Pickands dependence function) can be written with respect to the *spectral measure* H,

$$A(w_1, ..., w_d) = \int_{\Delta_{d-1}} \max_{1 \leq i \leq d} w_i s_i dH(s_1, ..., s_d) \qquad (6)$$

$$\text{where} \int_{\Delta_{d-1}} s_i dH(s_1, ..., s_d) = 1 \quad \forall i \in [1, \ldots, d].$$

H a Borel measure on Δ_{d-1}, which is a necessary and sufficient condition for A [Seg12].

- Integral is intractable, we propose modeling H with generative models.
- Gives us additional information on clustering behavior of covariates.
- When H is a density, we obtain characterizations useful for spatial extremes.

Spectral Measure: Sampling From The Copula

- Suppose we're given a dependence function, $A(w)$. How can we sample from high dimensions?
- Many implications in simulating rare events.
- We learn the spectral measure through a generative model $G(\cdot;\phi)$

$$\min_{\phi} \mathbb{E}_{\mathrm{w}\sim\Delta_{d-1}} \left\| A(\mathrm{w}) - \mathbb{E}_{\mathrm{y}}[\max_{0\le k\le d} w_k y_k] \right\|_2^2 + \|\mathbb{E}_{\mathrm{y}}[\mathrm{y}] - 1_d\|_2^2 ,$$

where $y \sim G(\cdot;\phi) \in \mathbb{R}_+^d$ is a generative model parameterized by ϕ.

- We provide a heuristic for sampling by taking the maximum of the product a sample from the spectral measure and unit Frechét sample.
- Efficient exact sampling is then permitted through techniques developed by Liu, Blanchet, Dieker, Mikosch [LBDM16].

Sampling From MEV Distributions

Algorithm 2 Heuristic for sampling from a Pickands copula

1: **Input:** $G(\cdot;\phi^*)$, p_z
2: **for** $i \in \{1,\ldots,N_{\text{gen}}\}$ **do**
3: Generate $\mathrm{y}^{(i)}$ where $\mathrm{y}^{(i)} = G(\mathrm{z}^{(i)};\phi_*), \mathrm{z}^{(i)} \sim p_z$.
4: Sample $\xi^{(i)}$ from a unit Fréchet distribution.
5: **end for**
6: Compute the component-wise maxima as:
$M = \max_{1\leq i\leq N_{\text{gen}}}\{\xi^{(i)} \odot \mathrm{y}^{(i)}\}$.
7: **Output:** M.

Sampling From MEV Distributions

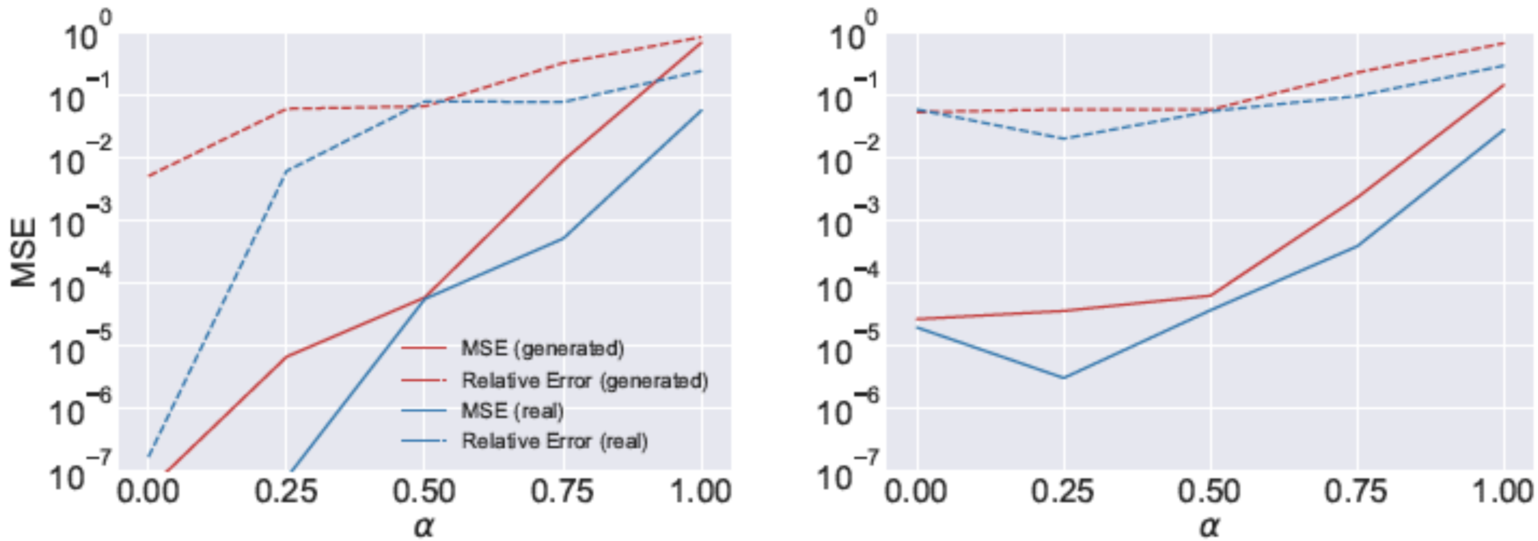


Figure 4.1: The differences in the state of the art Pickands dependence estimator (CFG) estimate of data sampled from our generator and exact samples for $d = 225$. Left is symmetric logistic and right is asymmetric logistic. Exact, Proposed

Sampling From MEV Distributions

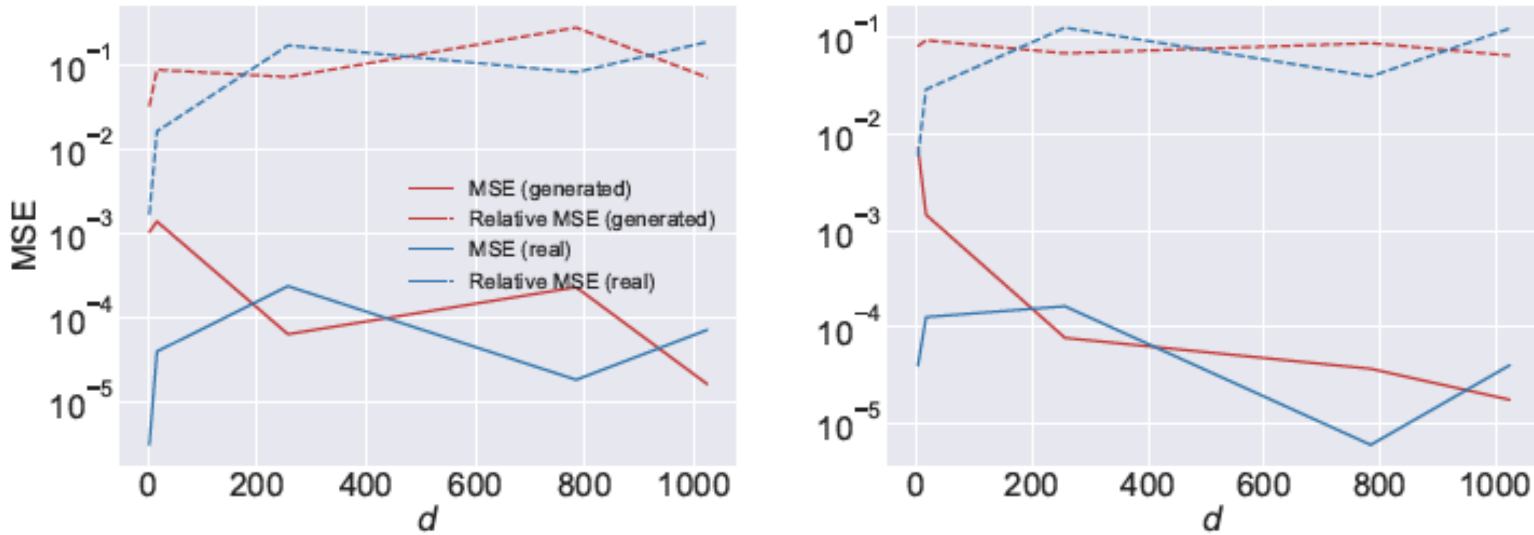


Figure 4.2: The differences in the state of the art Pickands dependence estimator (CFG) estimate of data sampled from our generator and exact samples for $\alpha = 0.5$. Left is symmetric logistic and right is asymmetric logistic. Exact, Proposed

Sampling Rare Events Using Adversarial Techniques

- The natural way to create a generative model is to estimate the Pickands dependence function from data using MLE, then use Algorithm 2 to generate samples from the estimated Pickands.
- To avoid direct estimation of the Pickands function, we can learn the generator through the method of moments where we design a generator parameterized by ϕ to match the moments of $Z_w \sim \mathrm{Exp}(A(\mathrm{w}))$:

$$\min_{\phi} \left| \mathbb{E}_{\mathrm{w} \sim \mathcal{U}(\Delta_{d-1})} \mathbb{E}_{Z_w} z^m - \mathbb{E}_{\mathrm{w} \sim \mathcal{U}(\Delta_{d-1})} \mathbb{E}_{Z_{w,\phi}} z^m \right|, \quad m = 1, 2, \cdots$$

Spatial Extremes via The Spectral Measure

Distributions of Extremes of PDEs

- Suppose we wish to consider the mapping of distributions of initial conditions to solutions of PDEs over time
- Parameterize mapping $T_\theta(\cdot|t)$ from distribution of initial conditions to solutions while encoding properties of the PDE in the mapping
- Induces latent dynamics on latent distributions of initial conditions
- **Takeaway:** By learning diffeomorphisms between distributions of initial conditions and solutions, we can characterize the extreme events by their initial conditions

$$z_t \sim P_{0,t}(\cdot) \tag{7}$$

$$x_t \sim T_\theta(P_{0,t}(\cdot)|t) \quad \text{where } \mathcal{D}(x) = 0 \tag{8}$$

$\mathcal{D}$ are the differential operators assumed to be governing the data

Max-Stable Processes

- Infinite dimensional extreme value distributions are max-stable processes
- Maxima of a stochastic process over a field, $\mathcal{T}$

$$\frac{\max_{1\leq i\leq n} X_i(t) - b_n(t)}{a_n(t)} \to M(t) \text{ as } n \to \infty, \ \ t \in \mathcal{T}$$

- When $M(t)$ is stationary, there is a spectral decomposition [DH+84]

$$M(t) = \max_{i\geq 1} \xi_i Y_i^+(t)$$

- $Y_i(t)$ is the i^{th} realization of a continuous stochastic process over $\mathcal{T}$ with $\mathbb{E}[Y] = 1$.
- ξ_i is a point from a Poisson process with intensity $\xi^{-2} d\xi$

Max-Stable Processes Governed by PDEs

- Suppose that we wish to consider maxima of processes driven by PDEs with random initial conditions
- From the spectral decomposition, we can restrict $Y_i(t)$ to be solutions to PDEs
- Learn the differential operators and the distribution of the max of the stochastic process from given data over field
- Consider estimators such as the Malliavin-based estimator proposed by Blanchet and Liu for optimizing parameters [BL16]

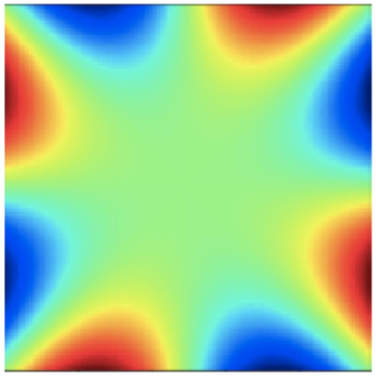


Figure 5.1: Solution to the Helmholtz equation for particular BC.

Max-Stable Processes Governed by PDEs

- Suppose we are interested in the conditional height of a wave at our current location given the height of the wave at another (nearby) location
- We inform our max-stable process with equations of wave mechanics
- Compute the conditional probability of the wave height occurring above a certain level

Figure 5.2: Image source: Unsplash

Learning The Intensity Measure

- Define a risk functional, $r(\cdot) : \mathbb{R}^d \to \mathbb{R}_+$ and a threshold vector $u \in \mathbb{R}^d_+$
- We can define the $r-$Pareto process with respect to the max-stable process as

$$M_r(t) = \max_{i \geq 1} \xi_i \frac{Y_i^+(t)}{r(Y_i^+)}$$

- The log-likelihood function for an $r-$Pareto process at d locations $\in \mathcal{T}$, x can be expressed with respect to the density of the intensity function, $\lambda(\cdot)$ (obtained by differentiating the exponent measure of the max-stable process, normally intractable) de Fondeville and Davison 2018 [dFD18]

$$\mathcal{L}(\cdot;\theta) = \sum_{n=1}^{N} \mathbb{1}\left\{ r\left(\frac{\mathrm{x}^{(n)}}{u}\right) \geq 1 \right\} \log\left(\lambda(\mathrm{x}^{(n)};\theta)\right)$$

- **Takeaway:** Using flexible generative models we can compute likelihoods for $r-$Pareto processes.

Latent Components via the Spectral Measure

- By understanding the properties of H on Δ_{d-1} we can understand latent structure of our data
- For high dimensional data, this involves understanding which covariates become large simultaneously
- Apply recent results from nonlinear ICA to guarantee recovery of latent variables

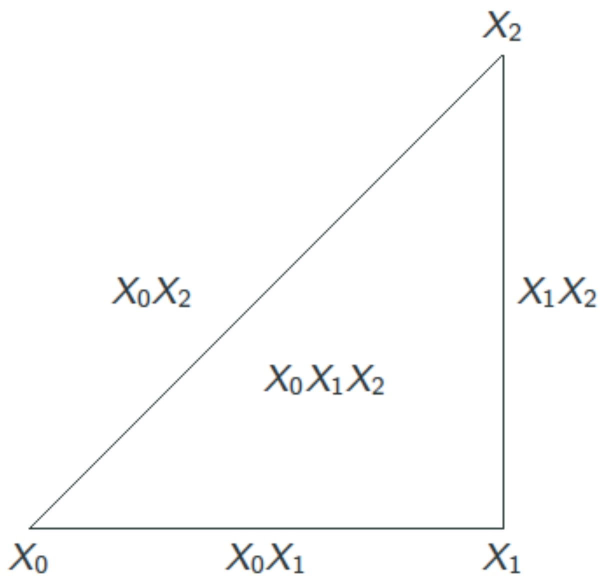


Figure 5.3: Different regions of Δ_2 where mass can be placed corresponding to different asymptotic dependence between covariates. E.g. delta functions at the vertices corresponds to independent components.

Latent Components via the Spectral Measure

- Related work proved that true latent space and dimension is recoverable up to affine transformation (details in [SRK20])

$$Z^* = A\hat{Z} + b$$

  where Z^* are the true latent variables and $\hat{Z}$ are the estimated latent variables

- Components that do not contribute to the data will be seen as noise terms in the learned latent distribution, giving an estimate of independent components
- Dirichlet(α) distributions can be represented through the Gamma distribution as vectors $\langle y_1/\sum y_i, \ldots, y_d/\sum y_i\rangle \quad y_i \sim \Gamma(\alpha_i, 1)$, sufficient statistics of estimated latent variables are related to the true distribution of latent variables

References

- Jose Blanchet and Zhipeng Liu, *Malliavin-based multilevel monte carlo estimators for densities of max-stable processes*, International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer, 2016, pp. 75–97.
- Raphaël de Fondeville and Anthony C Davison, *High-dimensional peaks-over-threshold inference*, Biometrika **105** (2018), no. 3, 575–592.
- Laurens De Haan et al., *A spectral representation for max-stable processes*, The annals of probability **12** (1984), no. 4, 1194–1204.
- Zhipeng Liu, Jose H Blanchet, AB Dieker, and Thomas Mikosch, *On optimal exact simulation of max-stable and related random fields*, arXiv preprint arXiv:1609.06001 (2016).
- Johan Segers, *Max-stable models for multivariate extremes*, arXiv preprint arXiv:1204.0332 (2012).

References (cont.)

Peter Sorrenson, Carsten Rother, and Ullrich Köthe, *Disentanglement by nonlinear ica with general incompressible-flow networks (gin)*, arXiv preprint arXiv:2001.04872 (2020).

Thank you!

Nonparametric estimators

- $$\widehat{A}_{\text{Pickands}}(\mathrm{w}) = \left(\frac{1}{B} \sum_{i=1}^{B} Z_{w,i} \right)^{-1}$$

- $$\widehat{A}_{\text{CFG}}(\mathrm{w}) = \exp \left[-\gamma - \frac{1}{B} \sum_{i=1}^{B} \log Z_{w,i} \right]$$

  where $\gamma = -\int_0^\infty \log x e^{-x} dx$ denotes Euler's constant.

- $$\widehat{A}_{\text{BDV},h}(\mathrm{w}) = B_h^{-1} \int_0^1 \frac{\log \tilde{C}(y^{w_1}, \ldots, y^{w_d})}{\log y} h^*(y)\, dy$$

  where $h : (0,1) \to \mathbb{R}_0^+$ is any positive weight function, $h^*(y) := h(y)(\log y)^2$, and $\tilde{C}$ is the empirical copula.