Spatial extremes and M-estimation for max-stable models

R.A. Yuen† and Stilian Stoev, Department of Statistics, University of Michigan

Introduction

- Numerous events display the human and financial costs associated with extreme physical and environmental phenomena

6/08 Cedar Rapids, IA (David Greedy/Getty Images)

1/13 Beijing, China (Diego Azubel/EPA)

9/11 Manor, TX (Jay Janner/Austin American Statesman/AP)

2/13 Boston, MA (Gene J. Puskar/Associated Press)

- We are interested in characterizing the probability distribution of such extreme events over a spatial region $T \subset \mathbb{R}^d$.
- Data consist of measurements recorded at fixed locations $\mathbf{t}_j \in T$, $j = 1, \ldots, m$ resulting in vector observations

$$\mathbf{Y}^{(i)} = \left(Y^{(i)}_{\mathbf{t}_1}, Y^{(i)}_{\mathbf{t}_2}, \ldots, Y^{(i)}_{\mathbf{t}_m}\right)^\top \subset \mathbb{R}^m$$

where $\mathbf{Y}^{(i)}, i = 1, 2, \ldots$ are independent and identically distributed.

- To characterize extremes we consider the limit of point-wise maximums

$$\left\{\frac{1}{a_n(\mathbf{t})} \max_{i=1,\ldots,n} Y^{(n)}_{\mathbf{t}} - b_n(\mathbf{t})\right\}_{\mathbf{t}\in T} \xrightarrow{d} \{X_\mathbf{t}\}_{\mathbf{t}\in T}, \text{ as } n \to \infty \qquad (1)$$

where $a_n(\mathbf{t})$ and $b_n(\mathbf{t})$ are normalization functions.

- The limiting process $\{X_\mathbf{t}\}_{\mathbf{t}\in T}$ models **worst case scenaria** and must be max-stable (Resnick 1987):

Max-stable process

For independent copies $X^{(i)}_\mathbf{t}, i = 1, \ldots, n$ of $X_\mathbf{t}$, there exists functions $c_n(\mathbf{t})$ and $d_n(\mathbf{t})$ such that

$$\left\{\frac{1}{c_n(\mathbf{t})} \max_{i=1,\ldots,n} X^{(n)}_\mathbf{t} - d_n(\mathbf{t})\right\}_{\mathbf{t}\in T} \stackrel{d}{=} \{X_\mathbf{t}\}_{\mathbf{t}\in T}$$

- All max-stable processes have a spectral representation for their finite dimensional distribution functions (de Haan 1984):

de Haan's spectral representation

Let λ be a measure on S and $g_\mathbf{t}(\mathbf{s}) : T \times S \mapsto \mathbb{R}_+$ such that for all $\mathbf{t} \in T$, $\int_S g_\mathbf{t}(\mathbf{s})\,\lambda(d\mathbf{s}) < \infty$. Then for every $\mathbf{x} = (x_{\mathbf{t}_1}, \ldots, x_{\mathbf{t}_m})^\top \in \mathbb{R}^m_+$

$$F(\mathbf{x}) := \mathbb{P}\left(X_{\mathbf{t}_j} \le x_{\mathbf{t}_j}, j = 1, \ldots, m\right) = \exp\left\{-\int_S \max_{j=1,\ldots,m} \frac{g_{\mathbf{t}_j}(\mathbf{s})}{x_{\mathbf{t}_j}} \lambda(d\mathbf{s})\right\}$$

Max-stable models

- By specifying the measure λ and parametric family of spectral functions $\{g_\mathbf{t}(\mathbf{s}|\boldsymbol\theta), \boldsymbol\theta \in \Theta \subset \mathbb{R}^p\}$, one can construct flexible parametric models for multivariate extremes.
- Examples
  - **Max-linear model**

$$F(\mathbf{x}|\boldsymbol\theta) = \exp\left\{-\sum_{k=1}^{L} \max_{j=1,\ldots,m} \frac{\theta_{jk}}{x_{\mathbf{t}_j}}\right\}, \ \theta_{jk} \ge 0. \qquad (2)$$

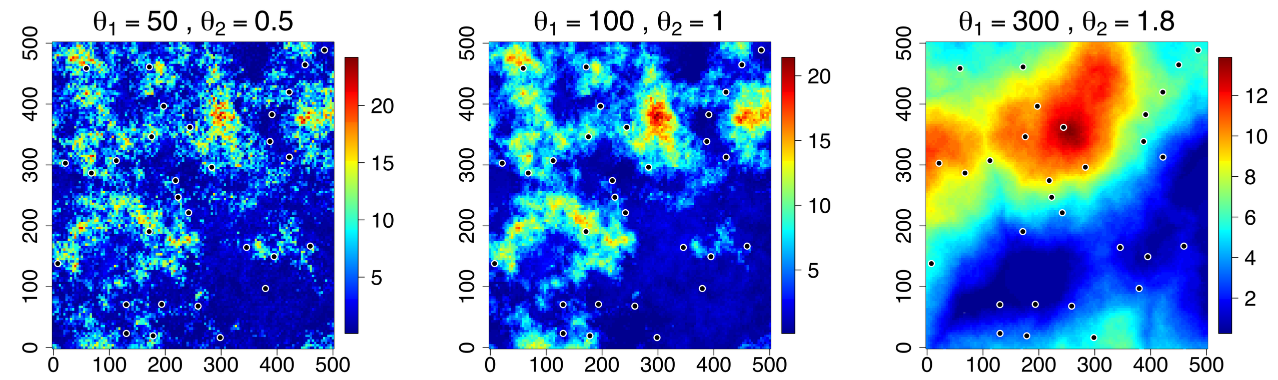
  - **Extremal Gaussian model** (Schlather 2002)

$$F(\mathbf{x}|\boldsymbol\theta) = \exp\left\{-\int_{\mathbb{R}^m}\left(\max_{j=1,\ldots,m}\left[\sum_{k=1}^{m}\tilde\rho(\mathbf{t}_j,\mathbf{t}_k|\boldsymbol\theta)\frac{s_k}{x_{\mathbf{t}_j}}\right]_+\right)\frac{e^{-\frac{1}{2}\mathbf{s}^\top\mathbf{s}}}{(\sqrt{2\pi})^{m-1}}d\mathbf{s}\right\}. \qquad (3)$$

where $\rho(\mathbf{t}_i,\mathbf{t}_j|\boldsymbol\theta) = \sum_{k=1}^m \tilde\rho(\mathbf{t}_i,\mathbf{t}_k|\boldsymbol\theta)\,\tilde\rho(\mathbf{t}_k,\mathbf{t}_j|\boldsymbol\theta)$ is the correlation function of a Gaussian random field on T.

- Realizations from the extremal Gaussian model with stable correlation function

$$\rho(\mathbf{t},\mathbf{s}|\boldsymbol\theta) = \exp\left[-\left(\|\mathbf{t}-\mathbf{s}\|/\theta_1\right)^{\theta_2}\right], \theta_1 > 0, \theta_2 \in (0,2] \qquad (4)$$



Problem Formulation

- Many useful max-stable models including (2) and (3) have **no tractable likelihood**

$$\frac{\partial}{\partial x_{\mathbf{t}_1}\cdots\partial x_{\mathbf{t}_m}} F(\mathbf{x}|\boldsymbol\theta) = ?$$

- Standard inferential methods unavailable.

MLE Bayesian Inference

- Bivariate maximum composite likelihood estimator (MCLE) exists (Padoan et. al 2010) for some models including (3)

$$\hat{\boldsymbol\theta}_{MCLE} = \underset{\boldsymbol\theta\in\Theta}{\arg\min} \sum_{i=1}^n \sum_{1\le j<k\le m} \ell\left(X^{(i)}_{\mathbf{t}_j}, X^{(i)}_{\mathbf{t}_k}|\boldsymbol\theta\right)$$

  - Not available for max-linear models (2).
  - Some models are unidentifiable through pairwise marginals.

Solution

- Minimum distance method (Wolfowitz 1957).

$$\underset{\boldsymbol\theta\in\Theta}{\arg\min} \int_{\mathbb{R}^m} \left(F_n(\mathbf{z}) - F(\mathbf{z}|\boldsymbol\theta)\right)^2 \mu(d\mathbf{z})$$

where F_n is the empirical distribution function.

  - Equivalent to minimizing the continuous ranked probability score (CRPS).

CRPS M-estimation

- Let $F(\mathbf{x}|\boldsymbol\theta) = \mathbb{P}_{\boldsymbol\theta}(\mathbf{X} \le \mathbf{x})$ be a multivariate CDF. Define the *continuous ranked probability score* (CRPS) as

$$\mathcal{E}_{\boldsymbol\theta}(\mathbf{x}) = \int_{\mathbb{R}^m_+} \left(F(\mathbf{z}|\boldsymbol\theta) - \mathbb{I}\{\mathbf{x} \le \mathbf{z}\}\right)^2 \mu(d\mathbf{z}) \qquad (5)$$

where μ is a tuning measure.

- For $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)} \stackrel{i.i.d.}{\sim} F_{\boldsymbol\theta_0}$, define the minimum CRPS estimate of $\boldsymbol\theta_0$ as

$$\hat{\boldsymbol\theta}_n = \underset{\boldsymbol\theta\in\Theta}{\arg\min} \sum_{i=1}^n \mathcal{E}_{\boldsymbol\theta}\left(\mathbf{X}^{(i)}\right) \qquad (6)$$

Consistency and asymptotic normality

Subject to mild regularity conditions the following results hold as $n \to \infty$

- *(Consistency)* $\hat{\boldsymbol\theta}_n \xrightarrow{p} \boldsymbol\theta_0$.
- *(Asymptotic normality)* $\sqrt{n}\left(\hat{\boldsymbol\theta}_n - \boldsymbol\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}_{\boldsymbol\theta_0} J_{\boldsymbol\theta_0} H^{-1}_{\boldsymbol\theta_0}\right)$

where elements of the $p \times p$ matrices $H_{\boldsymbol\theta_0}$ and $J_{\boldsymbol\theta_0}$ are

$$\left(H_{\boldsymbol\theta_0}\right)_{ij} = \int_{\mathbb{R}^m_+} \frac{\partial}{\partial\theta_i} F(\mathbf{z}|\boldsymbol\theta) \frac{\partial}{\partial\theta_j} F(\mathbf{z}|\boldsymbol\theta)\,\mu(d\mathbf{z}) \qquad (7)$$

and

$$\left(J_{\boldsymbol\theta_0}\right)_{ij} = \int_{\mathbb{R}^m_+}\int_{\mathbb{R}^m_+} \beta_{\boldsymbol\theta_0}(\mathbf{z}_1,\mathbf{z}_2) \frac{\partial}{\partial\theta_i} F(\mathbf{z}_1|\boldsymbol\theta_0) \frac{\partial}{\partial\theta_j} F(\mathbf{z}_2|\boldsymbol\theta_0)\,\mu(d\mathbf{z}_1)\,\mu(d\mathbf{z}_2) \qquad (8)$$

with $\beta_{\boldsymbol\theta_0}(\mathbf{z}_1,\mathbf{z}_2) = F(\mathbf{z}_1 \wedge \mathbf{z}_2|\boldsymbol\theta_0) - F(\mathbf{z}_1|\boldsymbol\theta_0)\,F(\mathbf{z}_2|\boldsymbol\theta_0)$.

Remark: We have identified a concrete specification of the tuning measure μ that allows accurate numerical evaluation of the expressions (5) - (8) in a wide variety of max-stable models.

Simulation Study

- Using the model (3) with correlation function (4), we set $\boldsymbol\theta_0 = (100, 1)$ and simulated 100 replications at $m = 30$ uniformly sampled locations over a 500×500 grid. Realizations were generated using the **R** package **`SpatialExtremes`** (Ribatet 2012).
- For sample sizes $n = 100$ and $n = 1000$ we numerically optimize CRPS criterion (6).

Table: *Empirical mean and standard deviation from 100 replications of the CRPS M-estimator. Coverage rates are calculated numerically using plug-in estimates of expressions (7) and (8).*

	θ_1 (100)		θ_2 (1)	
n	100	1000	100	1000
mean	110.56	97.00	1.25	1.10
sd	113.73	32.89	0.63	0.34
.95 coverage	0.98	0.96	0.90	0.92

†Partially supported by Univ. of Mich. Rackham Merit Fellowship and NSF-AGEP grant DMS 1106695.