M-estimation for max-stable models

Robert A. Yuen

with Stilian Stoev

University of Michigan
Department of Statistics

August 5th, 2013
Joint Statistical Meetings
Montreal, QC

Outline

1. Introduction to max-stable processes.
2. Inference problem and solutions.
3. Swiss rainfall.

Preliminaries

- We are interested in estimating the spatial dependence structure of extreme phenomena over a region $S \subset \mathbb{R}^2$.
- We assume an underlying stochastic process $X \equiv \{X(s)\}_{s \in S}$ that characterizes extreme spatial dependence.

Preliminaries

- We are interested in estimating the spatial dependence structure of extreme phenomena over a region $S \subset \mathbb{R}^2$.
- We assume an underlying stochastic process $X \equiv \{X(s)\}_{s \in S}$ that characterizes extreme spatial dependence.

- Claim: Max-stable processes are a canonical class of models for spatial extremes.

Definition

- Let $X_1, X_2, \ldots$ be iid copies of a stochastic process $X \equiv \{X(s)\}_{s \in S}$.
- If for every n, there exists $c_n(s) > 0$ and $d_n(s)$ such that

$$\left\{ \frac{1}{c_n(s)} \bigvee_{i=1}^{n} X_i(s) - d_n(s) \right\}_{s \in S} \stackrel{fdd}{=} \{X(s)\}_{s \in S},$$

then X is *max-stable.*

Why max-stable?

Prop 5.9 Resnick (1987)

- Let $Y_1, Y_2, \ldots$ be iid copies of a stochastic process $Y \equiv \{Y(s)\}_{s \in S}$.
- If there exists sequence of functions $a_m(s) > 0$ and $b_m(s)$ such that

$$\left\{ \frac{1}{a_m(s)} \bigvee_{i=1}^{m} Y_i(s) - b_m(s) \right\}_{s \in S} \xrightarrow{fdd} \{X(s)\}_{s \in S} \quad \text{as } m \to \infty,$$

then X is *max-stable.*

Why max-stable?

Prop 5.9 Resnick (1987)

- Let $Y_1, Y_2, \ldots$ be iid copies of a stochastic process $Y \equiv \{Y(s)\}_{s \in S}$.
- If there exists sequence of functions $a_m(s) > 0$ and $b_m(s)$ such that

$$\left\{ \frac{1}{a_m(s)} \bigvee_{i=1}^{m} Y_i(s) - b_m(s) \right\}_{s \in S} \xrightarrow{fdd} \{X(s)\}_{s \in S} \quad \text{as } m \to \infty,$$

then X is *max-stable.*

Prop 5.10 Resnick (1987)

Concerning spatial dependence, WLOG: $\mathbb{P}[X(s) \leq x] = \exp(-1/x)$ for all $s \in S$. (standard Fréchet margins)

Characterizations of max-stable models

Finite dimensional distribution functions

Spectral representation (de Haan, 1984)

For every $D \subset S$ with $|D| = d < \infty$, there exists a *unique* finite measure H_D on $\mathbb{S}_+^{d-1} = \left\{u \in \mathbb{R}_+^d : \|w\| = 1\right\}$ such that

$$F(x) := \mathbb{P}\left[X(s_j) \le x_j, j = 1, \ldots, d\right] = \exp\left[-\int_{\mathbb{S}_+^{d-1}} \left\{\bigvee_{j=1}^{d} \frac{u_j}{x_j}\right\} H_D(du)\right].$$

Characterizations of max-stable models

Finite dimensional distribution functions

Spectral representation (de Haan, 1984)

For every $D \subset S$ with $|D| = d < \infty$, there exists a *unique* finite measure H_D on $\mathbb{S}_+^{d-1} = \left\{u \in \mathbb{R}_+^d : \|w\| = 1\right\}$ such that

$$F(x) := \mathbb{P}\left[X(s_j) \leq x_j, j = 1, \ldots, d\right] = \exp\left[-\int_{\mathbb{S}_+^{d-1}} \left\{\bigvee_{j=1}^{d} \frac{u_j}{x_j}\right\} H_D(du)\right].$$

- By a change of variables, a variety of flexible parametric max-stable models can be specified through finite demensional distribution functions

$$F_\theta(x) := \exp\left[-\int_E \left\{\bigvee_{j=1}^{d} \frac{g_{s_j}(w)}{x_j}\right\} \nu_\theta(dw)\right], \quad \theta \in \Theta \subset \mathbb{R}^p,$$

where $g_s \in \mathcal{L}_+^1(E, \mathcal{B}(E), \nu)$ for all $s \in S$.

Characterizations of max-stable models

Example: spectrally discrete

- Max-linear model:

$$F_{\theta}(x) = \exp\left[-\sum_{k=1}^{K}\left\{\bigvee_{j=1}^{d}\frac{\theta_{jk}}{x_j}\right\}\right], \quad \theta_{jk} \geq 0.$$

- Generated by $X(s_j) = \bigvee_{k=1}^{K} \theta_{jk} Z_k$, where $\{Z_k\}_{k=1}^{K}$ iid 1-Fréchet.
- Arise naturally in economics and finance, as models of extreme losses (Einmahl et al., 2012).

Characterizations of max-stable models

Example: spectrally Gaussian

- Extremal Gaussian (Schlather, 2002):

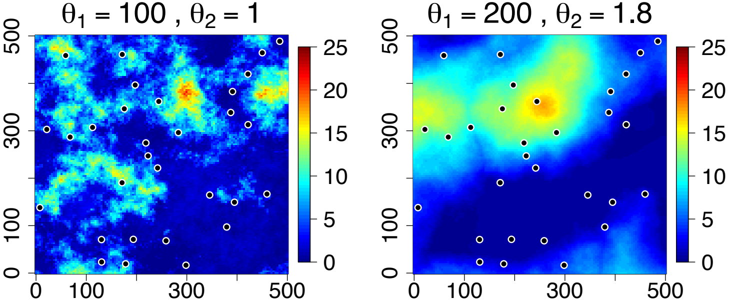
$$F_\theta(x) = \exp\left[-\int_{\mathbb{R}^d}\left\{\bigvee_{j=1}^{d}\frac{\sqrt{2\pi}w_j \vee 0}{x_j}\right\}\phi_d(w|\theta)\,dw\right]$$

- $\phi_d(w|\theta)$ is a d-dim density of stationary standard Gaussian process with correlation function $\rho(\|s_i - s_j\|\,|\theta)$.
- Generated by $X(s_j) = \bigvee_{k=1}^{\infty}\left(\sqrt{2\pi}w_j^{(k)} \vee 0\right)/\varepsilon_k$, $\left\{\left(w^{(k)}, \varepsilon_k\right)\right\}_{k=1}^{\infty}$ points of a Poisson point process on $\mathbb{R}^d \times \mathbb{R}_+$ with intensity measure $\phi_d(w|\theta)\,dw \times d\varepsilon$.

Others include well known Smith (Smith, 1990) and Brown-Resnick (Brown and Resnick, 1977) processes.

Characterizations of max-stable models

- Two realizations from extremal Gaussian with $\rho\left(\|s_i - s_j\| | \theta\right) = \exp\left[-\left(\|s_i - s_j\|/\theta_1\right)^{\theta_2}\right]$



- generated using `R` package `SpatialExtremes` (Ribatet, 2011).

Inference problem

- When $d > 2$, full joint likelihoods for max-stable processes are not available except in the special case of trivariate Smith models (Genton et al., 2011).

$$L_d(\theta, x) = \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} \exp\left[-\int_E \left\{ \bigvee_{j=1}^{d} \frac{g_{s_j}(w)}{x_j} \right\} \nu_\theta(dw) \right] = ?$$

- Standard MLE and Bayesian inference is infeasible.
- Spawned many recent works (Padoan et al., 2010; Erhardt and Smith, 2011; Reich and Shaby, 2012; Einmahl et al., 2012).

Inference problem

- When $d > 2$, full joint likelihoods for max-stable processes are not available except in the special case of trivariate Smith models (Genton et al., 2011).

$$L_d(\theta, x) = \frac{\partial^d}{\partial x_1 \partial x_2 \cdots \partial x_d} \exp\left[- \int_E \left\{ \bigvee_{j=1}^{d} \frac{g_{s_j}(w)}{x_j} \right\} \nu_\theta(dw) \right] = ?$$

- Standard MLE and Bayesian inference is infeasible.
- Spawned many recent works (Padoan et al., 2010; Erhardt and Smith, 2011; Reich and Shaby, 2012; Einmahl et al., 2012).

- (Yuen and Stoev 2013) arXiv:1307.7209 - "CDF appears natural to work with".

Minimum distance estimation

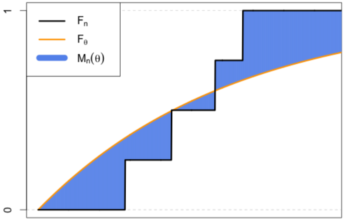
M-estimator

- Let $F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{X_i(s_j) \le x_j, j=1,\dots,d\}}$ - d-dim empirical CDF.
- Define

$$M_n(\theta) := \int_{\mathbb{R}_+^d} (F_n(x) - F_\theta(x))^2 \mu(dx), \; \hat{\theta}_n := \arg\min_{\theta \in \Theta} M_n(\theta),$$

where μ is tuning measure.

- Equivalent to minimum CRPS.
- Under mild regularity, $\hat{\theta}_n$ is consistent and asymptotically normal.



Minimum distance estimation

- M-estimator is applicable to broad variety of max-stable models.
- Identifies complex tail dependence structure beyond pairwise.
- Can estimate marginal and dependence parameters together.
- Computationally expensive but dimension in 100's are possible.

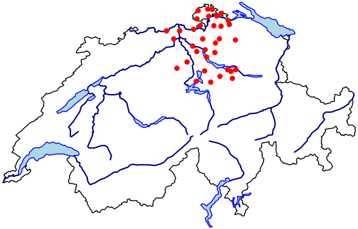
Swiss rainfall

- We model maximum single day rainfall during summer months using the extremal Gaussian model with Matérn correlation function

$$\rho\left(s_i, s_j|\theta\right) = \frac{\left(\|s_i - s_j\|/\theta_1\right)^{\theta_2}}{2^{\theta_2-1}\Gamma(\theta_2)} K_{\theta_2}\left(\|s_i - s_j\|/\theta_1\right), \quad \theta_1 > 0, \theta_2 > 0.$$

K_{θ_2} denotes the modified Bessel function of order θ_2.

Data are recorded from 1962-2008 at 35 observation locations near Zurich, Switzerland. (Data courtesy of R package `SpatialExtremes`)

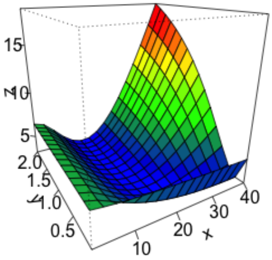


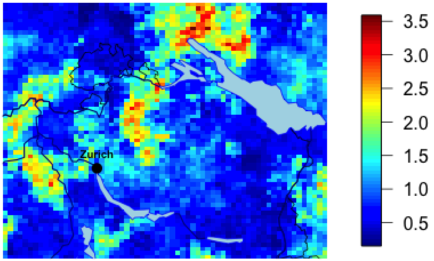
Swiss rainfall

- Summer maxima are transformed to unit Fréchet margins.
- Specifying an empirical measure for μ, we optimize the criterion $M_n(\theta)$ resulting in estimates (standard errors).

$$\hat{\theta}_1 = 18.46(25.58), \quad \hat{\theta}_2 = 0.5038(1.133).$$

- Optimization surface and one realization from the fitted model





References

Brown, B. and Resnick, S. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14:732–739.

de Haan, L. (1984). A spectral representation for max-stable processes. *Annals of Probability*, 12(4):1194–1204.

Einmahl, J. H., Krajina, A., and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimension. *Annals of Statistics*, 40(3):1764–1793.

Erhardt, R. and Smith, R. (2011). Approximate bayesian computing for spatial extremes. *arXiv:1109.4166*.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 28:180.

Genton, M., Ma, Y., and Sang, H. (2011). On the likelihood function of gaussian max-stable processes. *Biometrika*, 98(2):481–488.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Reich, B. and Shaby, B. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, 6(4).

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*, volume 4. Springer, New York, Berlin.

Ribatet, M. (2011). *SpatialExtremes: Modelling Spatial Extremes*. R package version 1.8-1.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5:33–44.

Smith, R. (1990). Max-stable processes and spatial extremes. *Unpublished Manuscript*.

A.I Consistency of M-estimator

Suppose the following regularity conditions hold

1. (Identifiability)
$$\theta_1 \neq \theta_2 \implies F_{\theta_1} \neq F_{\theta_2} \ a.e.\ \mu.$$
2. (Integrability) For $B(\theta_0) \subset \mathbb{R}^p$, an open neighborhood of θ_0
$$\int_{\mathbb{R}_+^d} \sup_{\theta \in B(\theta_0)} (1 - F_\theta(x))\, \mu(dx) < \infty.$$
3. (Continuity) The function $\theta \mapsto \int_{\mathbb{R}_+^d} (F_\theta(x) - F_{\theta_0}(x))^2 \mu(dx)$ is continuous in the compact parameter space $\Theta \subset \mathbb{R}^p$.

Then $\hat{\theta}_n \xrightarrow{p} \theta_0,$ as $n \to \infty$.

A.II Asymptotic normality of M-estimator

Assume the conditions above such that $\hat{\theta}_n \xrightarrow{p} \theta_0$. Suppose

1. The function $\theta \mapsto F_\theta(y)$ is twice continuously differentiable in an open neighborhood $B(\theta_0) \subset \mathbb{R}^p$ of θ_0.
2. $\int_{\mathbb{R}_+^d} \sup_{\theta \in B(\theta_0)} \|\dot{F}_\theta(y)\| + \|\dot{F}_\theta(y)\|^2 + \|\ddot{F}_\theta(y)\| \mu(dy) < \infty$
3. The matrix
$$H_{\theta_0} := \int_{\mathbb{R}_+^d} \dot{F}_{\theta_0}(y) \dot{F}_{\theta_0}(y)^\top \mu(dy)$$
is nonsingular.

Then

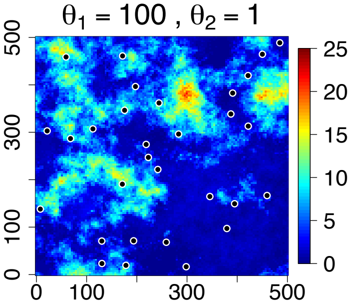
$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H_{\theta_0}^{-1} J_{\theta_0} H_{\theta_0}^{-1}\right), \quad \text{as n} \to \infty,$$

$J_{\theta_0} = \int_{\mathbb{R}_+^d} \int_{\mathbb{R}_+^d} \beta_{\theta_0}(y_1, y_2) \dot{F}_{\theta_0}(y_1) \dot{F}_{\theta_0}(y_2)^\top \mu(dy_1) \mu(dy_2)$ with $\beta_{\theta_0}(y_1, y_2) = F_{\theta_0}(y_1 \wedge y_2) - F_{\theta_0}(y_1) F_{\theta_0}(y_2)$.

A.III Simulation:

Extremal Gaussian model

- 500×500 domain.
- 30 observation locations.
- 100 replications
- CI's using plug-in estimates of $H_{\theta_0}, J_{\theta_0}$.



	$\theta_1\,(100)$		$\theta_2\,(1)$	
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

A.IV Marginal transformations

GEV (Fisher and Tippett, 1928; Gnedenko, 1943)

- For every $s \in S$

$$\mathbb{P}[Z(s) \leq z] = \exp\left\{ -\left[1 + \xi(s) \left(\frac{z - \mu(s)}{\sigma(s)} \right) \right]_{+}^{-1/\xi(s)} \right\} \quad \sigma(s) > 0.$$

- The cases $\xi(s) > 0, \xi(s) < 0$ and $\xi \to 0$ correspond to Fréchet, reverse Weibull and Gumbel distributions.

- Specify trend surface for parameters $\mu(s) = \beta_0^{\mu} + \beta^{\mu} s$, $\sigma(s) = \beta_0^{\sigma} + \beta^{\sigma} s$, $\xi(s) = \beta_0^{\xi}$.
- Let $X(s) = [1 + \xi(s)(Z(s) - \mu(s))/\sigma(s)]^{1/\xi(s)}$
- Optimize $\int_{\mathbb{R}_+^d} (F_n(x) - F_\theta(x))^2 \mu(dx)$ over θ, β.