Estimation and uncertainty quantification for extreme quantile regions

Boris Beranger
Simone A. Padoan & Scott A. Sisson

UNSW & ACEMS

ANZSC 2021, July 5th

Motivation

World Health Organisation:

† Air pollution kills 7 million people worldwide each year

Regulated emissions of pollutants (short-term concentrations):

- particulate matter (PM_{10}): daily average of $50\mu\mathrm{g}/\mathrm{m}^3$
- nitrogen dioxide (NO_2): daily average of $200\mu\mathrm{g}/\mathrm{m}^3$

Alert notice can be twice as high

Estimating extreme pollutant concentrations conditional on meteorological variables to understand the evolution of air pollution at high levels.

Motivation (2)

Assess high quantiles with low probabilities p with $p < 1/n$

Focus on $X \in \mathbb{R}_+^d$ with $d = 1, 2$.

- $d = 1$: Extreme quantile $Q(p) := F^{\leftarrow}(1 - p)$
- $d = 2$: Extreme quantile region $\mathcal{Q} = \{x \in \mathbb{R}_+^2 : f(x) \leq \alpha\}$, $\alpha > 0$
  $\mathbb{P}(\mathcal{Q}) = p$ for some very small p
- EV approach assumes $n \to \infty$ for asymptotic models to be used
  Approximate Q by Q_n.
  Approximate $\mathcal{Q}$ by $\mathcal{Q}_n$, $\mathbb{P}(\mathcal{Q}_n) = p_n \to 0, n \to \infty$.
  Assumption: $np \to c \in [0, \infty)$

Quantify the uncertainty around the estimate of extreme quantiles

Talk Outline

1. Estimating extreme quantiles
2. Inference
3. Simulation experiments
4. Analysis of extreme air pollution levels in Milan

Estimating univariate extreme quantiles

- Let $U(t) = F^{\leftarrow}(1 - 1/t), t > 1 \Longrightarrow Q(p) = U(1/p)$.

- Using dHF(2006, Chap. 1) $\frac{U(tx) - U(t)}{a(t)} \overset{t \to \infty}{\longrightarrow} \frac{x^{\gamma} - 1}{\gamma}$, we get

$$Q(p) \approx \mu + \sigma \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \quad \text{as } n \to \infty. \tag{1}$$

where $a(t) \approx \sigma$ and $U(t) \approx \mu$ as $t \to \infty$.

Estimating bivariate extreme quantiles

- From dHF(2006, Ch 6.1.2) we have that

$$t\left(1-F(tU_1(x_1),tU_2(x_2))\right)\overset{t\to\infty}{\longrightarrow}\nu(\{\mathrm{v}\in\mathbb{R}_+^2: v_1>x_1 \text{ or } v_2>x_2\})$$
$$=\iint_{\{v_1>x_1 \text{ or } v_2>x_2\}} g(\mathrm{v})\mathrm{dv}.$$

- We define the basic density function q by

$$tU_1(t)U_2(t)f\left(tU_1(x_1),tU_2(x_2)\right)\overset{t\to\infty}{\longrightarrow}(\gamma_1\gamma_2)^{-1}x_1^{1-\gamma_1}x_2^{1-\gamma_2}g(\mathrm{x})=:q(\mathrm{x}).$$

- Following EdHK(2013) we focus on

$$\mathcal{S}=\{\mathrm{x}\in\mathbb{R}_+^2: q(\mathrm{x})\le 1\}=\left\{\mathrm{x}\in\mathbb{R}_+^2: r\ge q_*^{-1}(w), w\in[0,1]\right\}$$

where $q_*(w)=q(w,1-w)^{-\frac{1}{1+\gamma_1+\gamma_2}}$ is called the angular basic density function.

- Inflate the basic set $\mathcal{S}$ into an extreme set:

$$\widetilde{\mathcal{Q}}_n\approx\left\{\left(\mu_1+\sigma_1\frac{\left(\frac{k\nu(\mathcal{S})x_1}{np}\right)^{\gamma_1}-1}{\gamma_1},\mu_2+\sigma_2\frac{\left(\frac{k\nu(\mathcal{S})x_2}{np}\right)^{\gamma_2}-1}{\gamma_2}\right):(x_1,x_2)\in\mathcal{S}\right\}.$$

Inference: Univariate case

- Use a **censored likelihood** approach:

$$\mathcal{L}(\mathrm{y}_{1:n};\theta) = \prod_{i=1}^{n} \mathcal{L}(y_i;\theta), \quad \mathcal{L}(y_i;\theta) \propto \begin{cases} G^{k/n}(t;\theta), & \text{if } y_i \leq t, \\ \frac{\partial}{\partial y} G^{k/n}(y;\theta)|_{y=y_i}, & \text{if } y_i > t, \end{cases}$$

where $G^{k/n}(y;\theta) \equiv G^{k/n}((y-\mu)/\sigma;\gamma)$ with $\theta = (\mu,\sigma,\gamma)^\top$.

- Apply the adaptive (Gaussian) random-walk Metropolis-Hastings (RWMH) algorithm of GFS(2016)

Inference: Bivariate case (1)

- For some large threshold $\mathrm{t} = (t_1, t_2)$, we have the approximation

$$F(\mathrm{y}) \approx \exp\left(-(z_1 + z_2)A(v)\right), \qquad \mathrm{y} \geq \mathrm{t},$$

where $v = z_2/(z_1 + z_2)$ with $z_i = \frac{k}{n}\left(1 + \gamma_i \frac{\cdot - \mu_i}{\sigma_i}\right)_+^{-1/\gamma_i}$, $i = 1, 2$.

- Model the angular density using **Bernstein polynomials** as in MPAV(2016).

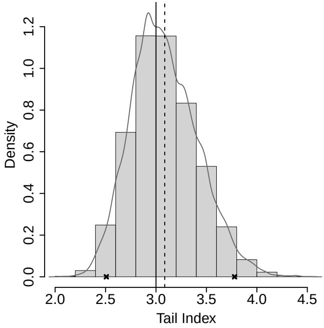
Inference: Bivariate case (2)

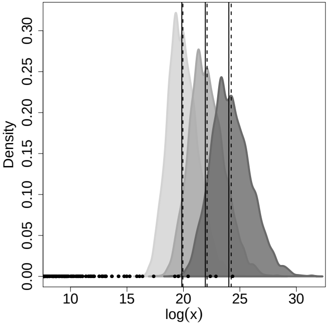
- Use a **censored likelihood** approach.
- **Priors**: $\boldsymbol{\eta}$ (function of polynomial coefficients $\boldsymbol{\beta}$), p_0, p_1 and κ
- **MCMC algorithm**:
  - **Step 1:** Margin 1
  - **Step 2:** Margin 2
  - **Step 3:** Dependence
    - Draw proposal $\kappa' \sim q_\kappa(\kappa|\kappa^{(j)})$ and $\boldsymbol{\eta}'_{\kappa'} \sim q_\eta(\boldsymbol{\eta}_\kappa|\kappa')$, and compute $\boldsymbol{\beta}'_{\kappa'}$
    - Compute acceptance probability

$$\pi_3 = \min\left(c\frac{\Pi(\kappa')}{\Pi(\kappa^{(j)})}\frac{\mathcal{L}\left(\theta_1^{(j+1)}, \theta_2^{(j+1)}, \kappa', \boldsymbol{\beta}'_{\kappa'}\right)}{\mathcal{L}\left(\theta_1^{(j+1)}, \theta_2^{(j+1)}, \kappa^{(j)}, \boldsymbol{\beta}^{(j)}_{\kappa^{(j)}}\right)}, 1\right).$$

Experiment: Univariate

- Generate $n = 1500$ obs from the **Half**-t with $\sigma_0 = 1$ and $\nu_0 = 1/3$
- Censoring at the 90-th empirical quantile
- **Prior:** $\Pi(\theta) := \Pi(\mu)\Pi(\log(\sigma))\Pi(\gamma) \propto 1/\sigma$ with $\sigma > 0$
- Run the MCMC for 50,000 iterations $\Rightarrow$ burn-in 20,000 iterations
- Posterior densities of quantiles corresponding to the exceedance probabilities $p = 1/750$ (light grey), $1/1500$ and $1/3000$ (dark grey)

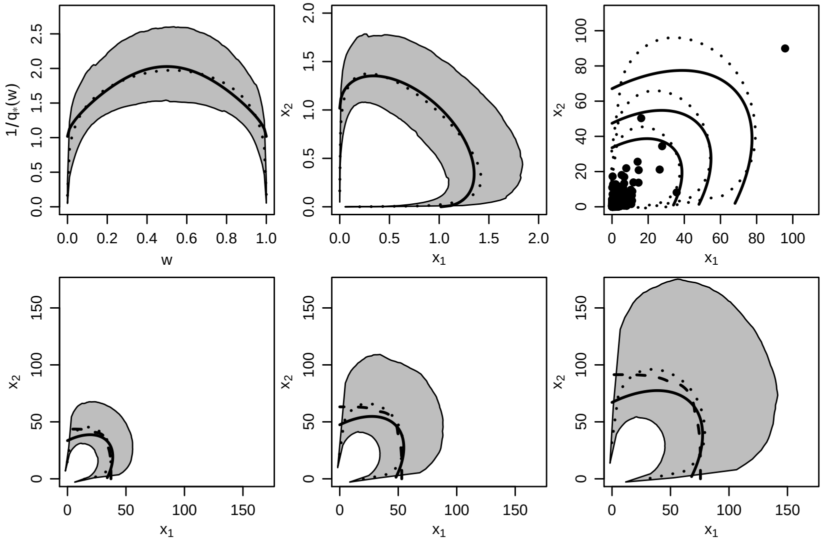




Experiment: Bivariate (1)

- Generate $n = 1500$ obs from the **bivariate truncated**-t with $\rho = 0.5$ and $\nu_0 = 2$
- Same thresholds and priors on the margins
- For each posterior sample: compute $q_*(w)$
- Estimate the basic set $\mathcal{S}$ through the points $\left(wq_*^{-1}(w), (1-w)q_*^{-1}(w)\right)$
- Posterior densities of quantiles corresponding to the exceedance probabilities $p = 1/750$, $1/1500$ and $1/3000$
- Comparison with EdHK (2013) – dashed lines.

Experiment: Bivariate (2)



Analysis of extreme air pollution

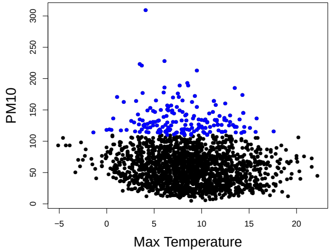
• **Air pollution levels in Milan over winter period** (end Oct - end Feb) between Dec 31st 2001 and Dec 30th 2017.

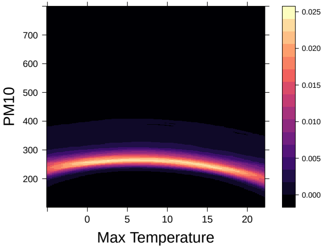
• Daily mean level of PM_{10} and daily maximum levels of NO_2

• Interactions between pollution and temperature: $\mu_i = \beta_{0,i} + \beta_{1,i} z + \beta_{2,i} z^2$

• **Threshold:** marginal 90-th empirical quantile (Mean Residual Life plots)

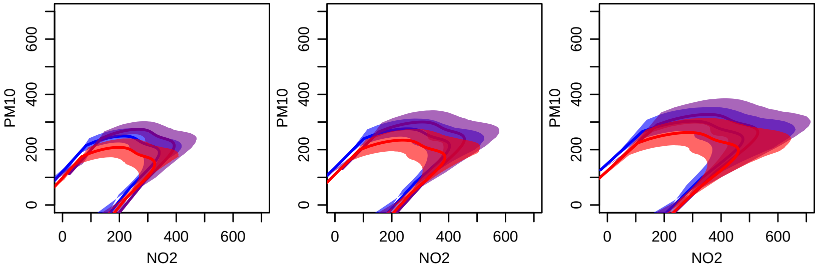
• Quantiles associated with the probabilities $p = 1/1200$ (event once every 10 winters)





Analysis of extreme air pollution (2)

- **Bivariate quantile regions** for $p = 1/600, 1/1200$ and $1/2400$ at minimum (blue), median (purple) and maximum (red) temperatures.



Conclusion:

- Methodology to estimate extreme quantile regions
- Quantification of the uncertainty under Bayesian paradigm
- Routines available in `R` package `ExtremalDep`
- Limited to the positive reals

Manuscript:

Beranger B., S. A. Padoan & S. A. Sisson (2020). Estimation and uncertainty quantification for extreme quantile regions. Extremes, In press.

Contact:

B.Beranger@unsw.edu.au
www.borisberanger.com

THANK YOU