Exceptional event analysis in the deep tail
or
Consistency of estimates of parameters in models with fixed covariates

Axel Bücher, Johan Segers, **Torben Staud**

No Pro(o)f-Seminar
April 1, 2025

Exceptional event analysis in the deep tail
Consistency of estimates of parameters in models with fixed covariates

---

Axel Bücher, Johan Segers, **Torben Staud**

No Pro(o)f-Seminar
April 1, 2025

Motivation

Extreme Value Statistics: State of the Art

- Limit results for (empirical) processes over Maxima.

Extreme Value Statistics: State of the Art

- Limit results for (empirical) processes over Maxima.
- CLT for maxima says: Class of distributional limits are mostly contained the GEV-family.

Extreme Value Statistics: State of the Art

- Limit results for (empirical) processes over Maxima.
- CLT for maxima says: Class of distributional limits are mostly contained the GEV-family.
- GEV assumption only in the limit (Domain of Attraction).

Extreme Value Statistics: State of the Art

- Limit results for (empirical) processes over Maxima.
- CLT for maxima says: Class of distributional limits are mostly contained the GEV-family.
- GEV assumption only in the limit (Domain of Attraction).
- Later: For i.i.d.; then extended to stationary (with some mixing).

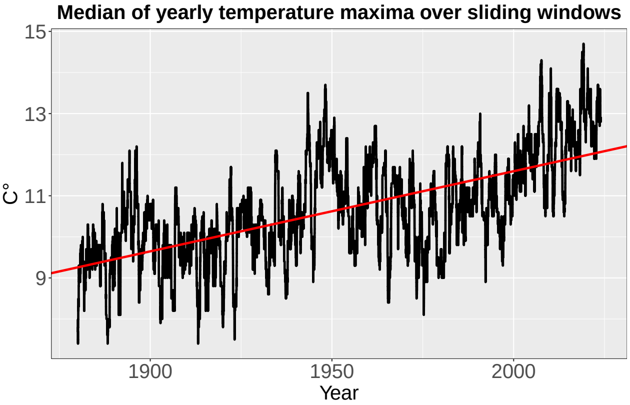
Extreme Value Statistics: State of the Art

- Limit results for (empirical) processes over Maxima.
- CLT for maxima says: Class of distributional limits are mostly contained the GEV-family.
- GEV assumption only in the limit (Domain of Attraction).
- Later: For i.i.d.; then extended to stationary (with some mixing).

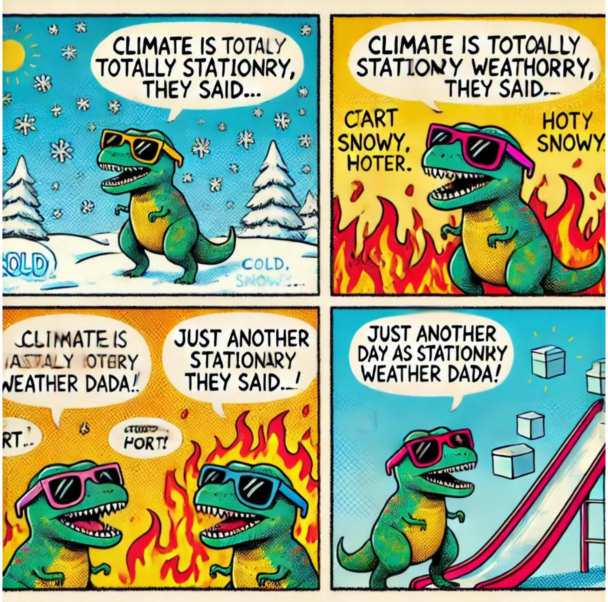
Where is the problem for many applications?

Problem

Where is the problem for many applications?



T-Rex says


CLIMATE IS TOTALY TOTALLY STATIONRY, THEY SAID...
COLD.
CLIMATE IS TOTOALLY STATIONY WEATHORRY, THEY SAID...
CTART SNOWY, HOTER.
HOTY SNOWY
CLIMATE IS
WEATHER DADA!
JUST ANOTHER STATIONARY THEY SAID...!
PORT!
JUST ANOTHER DAY AS STATIONRY WEATHER DADA!

State of the Art climate models

Temperature Events

- Trend modelled as *shift fit* with GMST
- $M_t = \mathrm{GEV}(\mu_t, \sigma_0, \xi_0)$

$$\mu_t = \mu_0 + \alpha_0 \times \mathrm{GMST}(t).$$

State of the Art climate models

Temperature Events

- Trend modelled as *shift fit* with GMST
- $M_t = \mathrm{GEV}(\mu_t, \sigma_0, \xi_0)$

$$\mu_t = \mu_0 + \alpha_0 \times \mathrm{GMST}(t).$$

Precipitation and Wind Events

- Distribution scales with GMST.
- $M_t = \mathrm{GEV}(\mu_t, \sigma_t, \xi_0)$

$$\mu_t = \mu_0 \exp\left[\alpha_0 \times \mathrm{GMST}(t)/\mu_0\right], \quad \sigma_t = \sigma_0 \exp\left[\alpha_0 \times \mathrm{GMST}(t)/\mu_0\right].$$

State of the Art climate models

Temperature Events

- Trend modelled as *shift fit* with GMST
- $M_t = \mathrm{GEV}(\mu_t, \sigma_0, \xi_0)$

$$\mu_t = \mu_0 + \alpha_0 \times \mathrm{GMST}(t).$$

Precipitation and Wind Events

- Distribution scales with GMST.
- $M_t = \mathrm{GEV}(\mu_t, \sigma_t, \xi_0)$

$$\mu_t = \mu_0 \exp\left[\alpha_0 \times \mathrm{GMST}(t)/\mu_0\right], \quad \sigma_t = \sigma_0 \exp\left[\alpha_0 \times \mathrm{GMST}(t)/\mu_0\right].$$

Combination (*Shift and Scale fit*) also possible (not yet standard though)

Our Solution

What are we doing? I

Tackle non-stationarity by treating it as a covariate (fixed setting)

What are we doing? I

Tackle non-stationarity by treating it as a covariate (fixed setting) Statistical Model

1. $\{Y_{n,i}\}_{i,n}$ rowwise $\mathbb{Y}$-valued independent triangular array.
2. $\{x_{n,i}\}_{i,n} \subset \mathbb{X}$ deterministic triangular array (**we observe those**).

What are we doing? I

Tackle non-stationarity by treating it as a covariate (fixed setting) Statistical Model

1. $\{Y_{n,i}\}_{i,n}$ rowwise $\mathbb{Y}$-valued independent triangular array.
2. $\{x_{n,i}\}_{i,n} \subset \mathbb{X}$ deterministic triangular array (**we observe those**).

$$Y_{n,i} \sim Q_n(x_{n,i}, \mathrm{d}y),$$

where $Q_n : \mathbb{X} \times \mathcal{B}(\mathbb{Y}) \to [0,1]$ is a *regular* sequence of Markov-Kernels.

Think of $Y_{n,t}$ = Maximal temperature of the $t-$th year (depends on t via the covariate $\mathrm{GMST}(t)$).
Toy example:

$$Y_{n,i} \sim \frac{i}{n} + \mathcal{N}(1/n, 1).$$

What are we doing? II

Tackle non-stationarity by treating it as a covariate (fixed setting)

Consider 'M-type estimators'

Tackle non-stationarity by treating it as a covariate (fixed setting)

Consider 'M-type estimators'

- $m\colon H \times \mathbb{X} \times \mathbb{Y} \to [-\infty, \infty]$, where H is a compact metric space. η is the parameter we want to estimate. We observe $x \in \mathbb{X}, y \in \mathbb{Y}$.

What are we doing? II

Tackle non-stationarity by treating it as a covariate (fixed setting)

Consider 'M-type estimators'

- $m\colon H \times \mathbb{X} \times \mathbb{Y} \to [-\infty, \infty]$, where H is a compact metric space. η is the parameter we want to estimate. We observe $x \in \mathbb{X}, y \in \mathbb{Y}$.
- Criterion functions:

$$M_n(\eta) = \mathbb{P}_n m_\eta = \frac{1}{n}\sum_{i=1}^{n} m_\eta(x_{n,i}, Y_{n,i}) \qquad \text{(subasymptotic)}$$

$$M(\eta) = \int_{\mathbb{X}}\int_{\mathbb{Y}} m_\eta(x, y)\, Q(x, \mathrm{d}y) P_X(\mathrm{d}x) \qquad \text{(tentative limit).}$$

What are we doing? II

Tackle non-stationarity by treating it as a covariate (fixed setting)

Consider 'M-type estimators'

- $m\colon H \times \mathbb{X} \times \mathbb{Y} \to [-\infty, \infty]$, where H is a compact metric space. η is the parameter we want to estimate. We observe $x \in \mathbb{X}, y \in \mathbb{Y}$.
- Criterion functions:

$$M_n(\eta) = \mathbb{P}_n m_\eta = \frac{1}{n}\sum_{i=1}^{n} m_\eta(x_{n,i}, Y_{n,i}) \qquad \text{(subasymptotic)}$$

$$M(\eta) = \int_{\mathbb{X}}\int_{\mathbb{Y}} m_\eta(x, y)\, Q(x, \mathrm{d}y) P_X(\mathrm{d}x) \qquad \text{(tentative limit).}$$

Now (kinda) maximize the arg of $M_n(\cdot)$ to estimate the argmax $\eta_0 \in H$ of $M(\cdot)$!

Considered estimators

All estimators which satisfy this *asymptotic maximizer condition*

$$M_n(\hat{\eta}_n) \geq M_n(\arg\max_{\tilde{\eta} \in H} M(\tilde{\eta})) - o_{a.s.}(1).$$

Considered estimators

All estimators which satisfy this *asymptotic maximizer condition*

$$M_n(\hat{\eta}_n) \geq M_n(\arg\max_{\tilde{\eta} \in H} M(\tilde{\eta})) - o_{a.s.}(1).$$

Note: The MLE falls in that category!

Considered estimators

All estimators which satisfy this *asymptotic maximizer condition*

$$M_n(\hat{\eta}_n) \geq M_n(\arg\max_{\tilde{\eta} \in H} M(\tilde{\eta})) - o_{a.s.}(1).$$

Note: The MLE falls in that category!
Essence: $\hat{\eta}_n$ should asymptotically maximize $M_n(\cdot)$.

Result

Theorem Axel, Johan, Torben (2024+): Under *some* conditions we have that $\hat{\eta}_n$ consistently estimates η_0 in the following sense

$$\mathbb{P}\big[\lim_{n\to\infty} \hat{\eta}_n = \eta_0\big] = 1,$$

where (again) $\hat{\eta}_n$ is an estimator satisfying $M_n(\hat{\eta}_n) \geq M_n(\eta_0) - o_{a.s.}(1)$ and $\eta_0 = \arg\max M(\cdot)$.

Result

Theorem Axel, Johan, Torben (2024+): Under *some* conditions we have that $\hat{\eta}_n$ consistently estimates η_0 in the following sense

$$\mathbb{P}\big[\lim_{n\to\infty} \hat{\eta}_n = \eta_0\big] = 1,$$

where (again) $\hat{\eta}_n$ is an estimator satisfying $M_n(\hat{\eta}_n) \geq M_n(\eta_0) - o_{a.s.}(1)$ and $\eta_0 = \arg\max M(\cdot)$.

That's a very general result:

- Dependence on covariate allowed via x.
- Non-stationarity of the $Y_{n,i}$ possible via dependence on the Feller-Markov kernels Q_n and $x_{n,i}$.
- Even non regular models like "support changes with η" (think of GEV model or uniform model $X \sim \mathcal{U}(0,\eta)$) allowed, since $m = -\infty$ allowed.
- We can handle even more general estimators than M-estimators.

Applications

Conditional MLE

- Model:

$$Y_{n,i} \sim P_{\theta(x_{n,i},\eta_0)},$$

- $(P_\theta : \theta)$ is a parametric family (think of GEV shift/scale fit) with densities wrt to a $\sigma-$finite measure ν.
- η_0 is unknown.
- $\theta \colon H \to \Theta$ is a known link function (like the scale/shift fit functions).
- We observe $Y_{n,i}$ and the covariates $x_{n,i}$.
- $Q_n \equiv Q = P_{\theta(\cdot,\eta_0)}$ is constant in n in this example.

Pairwise likelihood from multivariate extremes

- Now $\mathbb{Y} \subset \mathbb{R}^d, Y_{n,i} \sim P_{\theta(x_{n,i},\eta_0)}$ as before
- Sometimes calulcating the d-variate density is expensive. Solution:
- *Pairwise criterion function*

$$m_\eta(x,y) = \sum_{(j,l)\in D_2} \log p_{\theta(x,\eta)}(j,l),$$

where $D_2 = \{(j,l)\colon 1 \leq j \neq l \leq d\}$ and $p_\theta^{(j,l)}$ denotes the bivariate marginal (of d) density corresponding to j,l coordinates.

Optimum score estimation

- Fancy Data science
- Find a *scoring rule* S under which one calibrates a model.
- S takes two arguments P, y where P is a probability measure and y an observation. $S(P, Q) := \int S(P, y)\,\mathrm{d}Q(y)$.
- Identifiability of the model corresponds to assuming $S(P, P) > S(P, Q)$ for all $Q \neq P$ (unique maximum).
- We then have the estimator

$$\hat{\eta}_n = \arg\max_{\eta \in H} \frac{1}{n} \sum_{i=1}^{n} S(P_{\theta(x_{n,i}, \eta)}, y_i).$$

And quite some more examples (functional data . . .).

Outlook

- We have consistency (minimal property).
- Asymptotic normality next step (want confidence intervals and tests).
- Bootstraps?
- Dependence between the $Y_{n,i}$ (with mixing this shouldn't be to much of a problem).

Most important references

Bücher, A., Segers, J. and Staud., T. Consistency of estimates of parameters in models with fixed covariates (**WiP**).

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, Vol. 21, No. 1

Philip S., et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Clim. Meteorol. Oceanogr.*, 6, 177-203.

Thank you

Conditions

Weak convergence regularities

1. There exists a non-degenerate RV X on $\mathbb{X}$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{x_{n,i}} \rightsquigarrow P_X;$$

think of the uniform design in regression.

Weak convergence regularities

1. There exists a non-degenerate RV X on $\mathbb{X}$ such that
$$\frac{1}{n}\sum_{i=1}^{n}\delta_{x_{n,i}} \rightsquigarrow P_X;$$
think of the uniform design in regression.

2. There exists a Markov kernel $Q\colon \mathbb{X}\times\mathcal{B}(\mathbb{Y})\to[0,1]$ such that $x\mapsto Q(x,\,\mathrm{d}y)$ is weakly continuous and such that $(Q_n(\cdot,\,\mathrm{d}y)_n)$ converges continuously to $Q(\cdot,\,\mathrm{d}y)$ in the weak topology.

Identifiability

The asymptotic criterion function M has a unique point of maximum η_0, that is for $\eta \neq \eta_0$

$$M(\eta) = \int_{\mathbb{X}} \int_{\mathbb{Y}} m_\eta(x, y)\, Q(x,\, \mathrm{d}y) P_X(\,\mathrm{d}x) < \int_{\mathbb{X}} \int_{\mathbb{Y}} m_{\eta_0}(x, y)\, Q(x,\, \mathrm{d}y) P_X(\,\mathrm{d}x) = M(\eta_0).$$

Boundedness

1. The criterion function is bounded from above (may be unbounded below)
2. Q_n is L^2 sochastically dominated, that is, there exists $W \in L^2$ with

$$\forall t \geq 0, n \in \mathbb{N}, x \in \mathbb{X} \colon Q_n\big(x, \{y \in \mathbb{Y} \colon |m_{\eta_0}(x,y)| > t\}\big) \leq P(W > t).$$

Space and criterion function regularity

1. The metric spaces $(\mathbb{X}, d_{\mathbb{X}})$ and (H, d_H) are compact
2. The criterion function m is upper semicontinuous.
3. The function $(x, y) \mapsto m_{\eta_0}(x, y)$ is continuous $P_{X,Y}$ a.e.