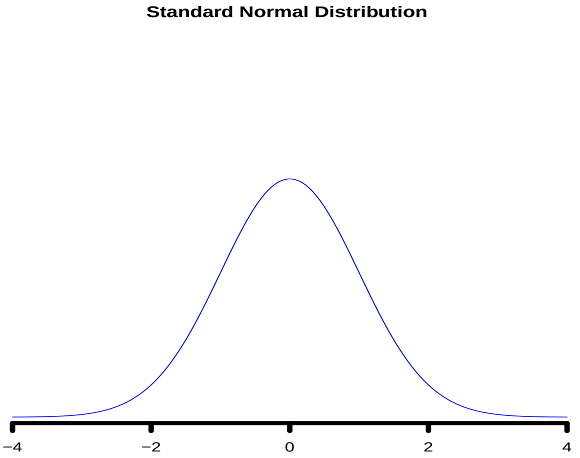
Quantiles and data-depth: the next generation

Snigdhansu Chatterjee

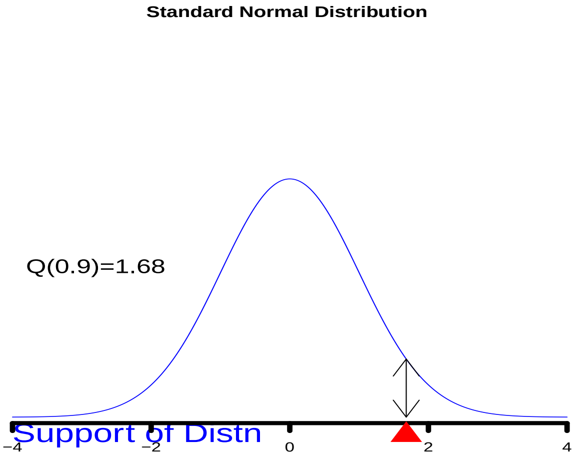
School of Statistics, University of Minnesota

August 15, 2013

Normal probability density function

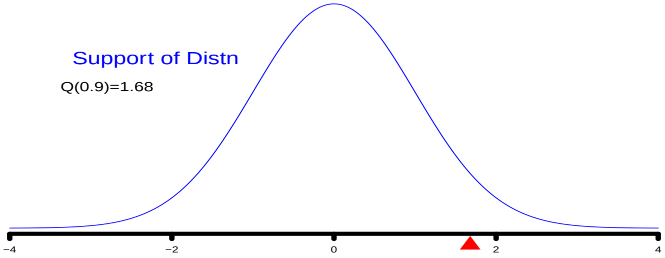


Normal quantile

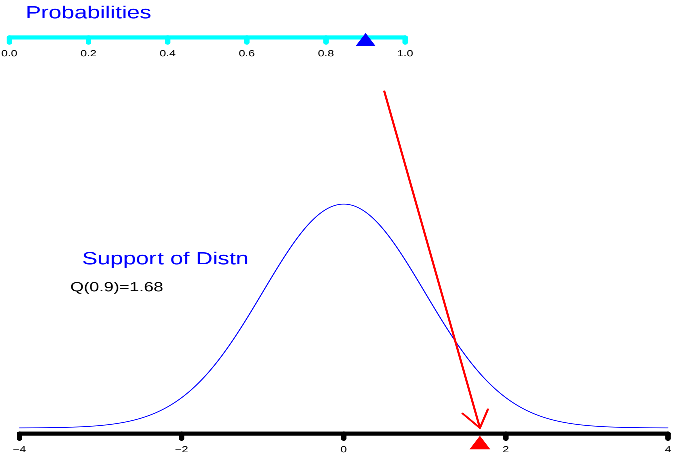


Probability range

Normal quantile



Univariate quantile mapping

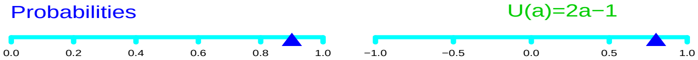


Univariate quantile mapping:formula

$$
\begin{aligned}
Q(a) &= \arg\min \mathbb{E}\left\{|X - q| + (2a - 1)(X - q)\right\}, \\
F_X(Q(a)) &= a.
\end{aligned}
$$

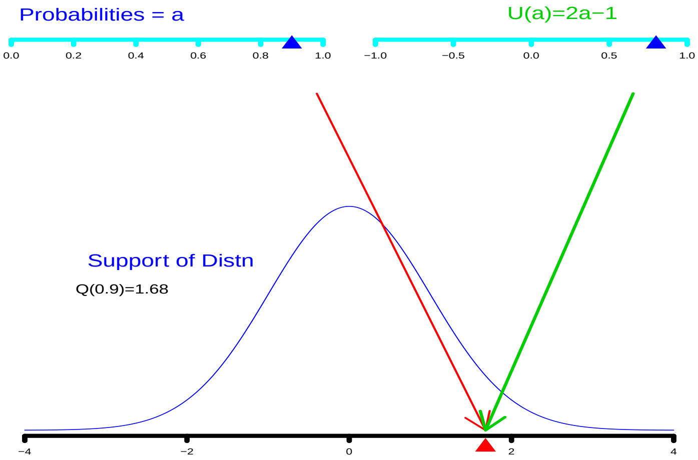
so $a \leftrightarrow Q(a)$ is a bijection. This is extremely important for doing Statistics.

A transformation



U(Probability)=2 Probability − 1

Univariate quantiles



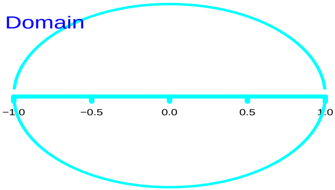
Univariate quantiles:formula

$$Q(a) = \arg\min \mathbb{E}\left\{|X - q| + (2a-1)(X - q)\right\}$$

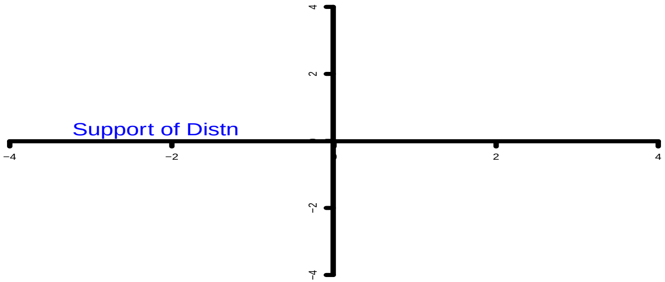
Define

$$\begin{aligned} u &= 2a - 1, \\ Q(u) &= \arg\min \mathbb{E}\left\{|X - q| + u(X - q)\right\} \end{aligned}$$

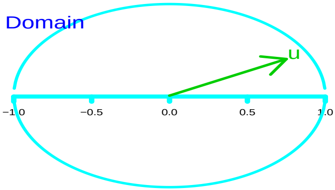
Bivariate quantiles: domain

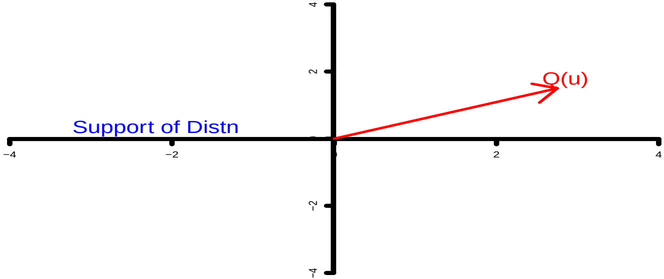


Bivariate quantiles: range



Bivariate quantiles





Generalized spatial quantiles

For every $\lambda \in \mathbb{R}$, the generalized spatial quantiles minimize:

$$\Psi_{u\lambda}(q) = \mathbb{E}\left[|X_U - q_U| \left\{ 1 + \lambda (X_U - q_U)^{-2} ||X_{U^\perp} - q_{U^\perp}||^2 \right\}^{1/2} + \beta(X_U - q_U) \right].$$

Further generalization (not pursued): For every $k \geq 1$, define

$$\Psi_{u\lambda k}(q) = \mathbb{E}\left[|X_U - q_U| \left\{ 1 + \lambda (X_U - q_U)^{-k} ||X_{U^\perp} - q_{U^\perp}||^k \right\}^{1/k} + \beta(X_U - q_U) \right].$$

Properties of generalized spatial quantiles

Generalized spatial quantiles minimize:

$$\Psi_{u\lambda}(q) \quad = \quad \mathbb{E}\left[|X_U - q_U| \left\{ 1 + \lambda (X_U - q_U)^{-2} ||X_{U^\perp} - q_{U^\perp}||^2 \right\}^{1/2} + \beta (X_U - q_U) \right].$$

Theorem

Write $\Psi_{u\lambda}(\cdot) = \mathbb{E}f(X, \cdot)$. Let $g(X, \cdot)$ be the subgradient of $f(X, \cdot)$, q^* the unique minimizer of $\Psi_{u\lambda}(\cdot)$, and q_n a minimizer of its sample version.

1. *$q_n \to q^*$ almost surely as $n \to \infty$.*
2. *If $\mathbb{E}||g(X, q^*)||^2 < \infty$ and if $\mathbb{E}f(X, q)$ is twice continuously differentiable at q^* with the second derivative H being positive definite, then as $n \to \infty$*

$$n^{1/2}(q_n - q^*) = -n^{-1/2} H^{-1} S_n + o_P(1),$$

where $S_n = \sum_{i=1}^{n} g(X_i, q^*)$.

Theorem

1. *Under the conditions of the previous item, the generalized bootstrap approximation for the distribution of $n^{1/2}(q_n - q^*)$ is consistent, and hence resampling may be used for statistical inference.*

Bahadur representation of generalized spatial quantiles

Theorem

1. *In addition to the conditions of the previous Theorem, assume that*

$$\begin{aligned} \|\frac{\partial}{\partial q}\mathbb{E}\Psi_{u,\lambda}(X,q) &- \frac{\partial^2}{\partial q^2}\mathbb{E}\Psi_{u,\lambda}(X,q^*)(q-q^*)\| \\ &= O(\|q-q^*\|^{(3+s)/2}) \text{ as } q \to q^*, \\ \mathbb{E}\|g(X,q)-g(X,q^*)\|^2 &= O(\|q-q^*\|^{1+s}) \text{ as } q \to q^*, \\ \mathbb{E}\|g(X,q)\|^r &< \infty \text{ as } q \to q^*, \end{aligned}$$

for some $s \in (0,1)$ *and* $r > (8 + p(1+s))/(1-s)$. *Then the following asymptotic Bahadur-type representation holds with probability 1:*

$$n^{1/2}(q_n - q^*) = -n^{-1/2}H^{-1}S_n + O(n^{-(1+s)/4}(\log\ n)^{1/2}(\log\log\ n)^{(1+s)/4})$$

as $n \to \infty$.

Projection quantiles

Generalized spatial quantiles minimize:

$$\Psi_{u\lambda}(q) = \mathbb{E}\left[|X_U - q_U| \left\{ 1 + \lambda (X_U - q_U)^{-2} ||X_{U^\perp} - q_{U^\perp}||^2 \right\}^{1/2} + \beta (X_U - q_U) \right].$$

Set $\lambda = 0$ to get projection quantiles.

Theorem

Projection quantiles have a one-to-one relationship with the unit ball, like univariate quantiles.

- Computationally extremely simple, no limitations from sample size and dimension (high p, low n allowed).
- Projection quantiles based confidence sets have exact coverage.
- For any λ, we now have asymptotic results as $\beta \to 1$. This provides a potentially new way of doing multivariate extreme values.

Applications: climate and beyond

- We have used some of these techniques to study *joint extreme value properties* of climate variables, eg, hurricane windspeed and pressure.
- We have used these to study *change in tail behavior* of climate characteristics.
- We are studying multivariate (extreme) quantile regression.
- Study relations between climate and economic or biological variables.
- Other application domains might be in finance, genomics, *ldots*.

Example

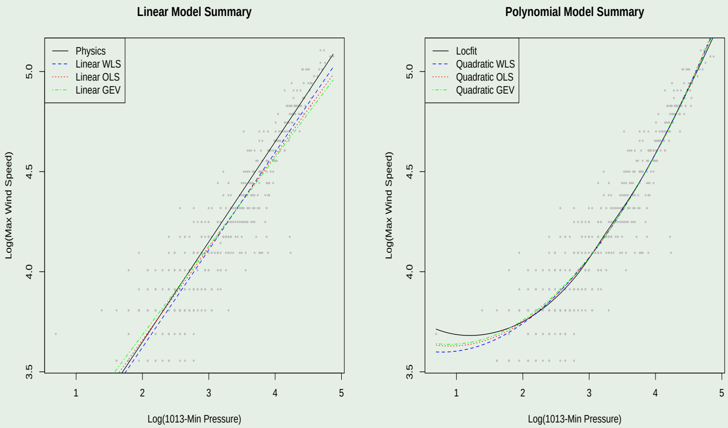


Figure : Physics, linear and quadratic statistical fits for bivariate extremes data

Example

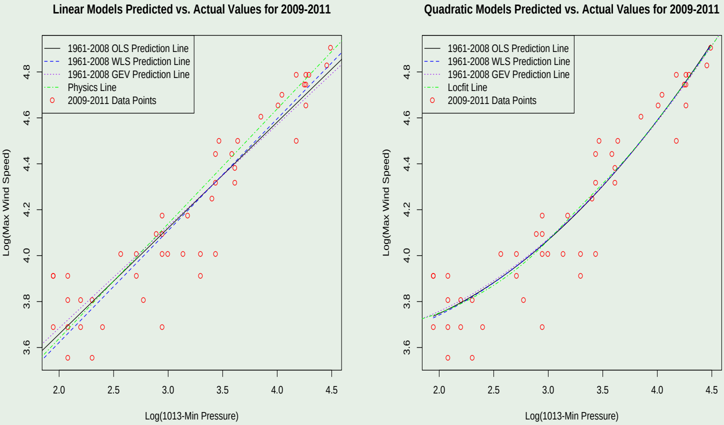


Figure : Bivariate extremes: projections

Acknowledgment:

- This research is partially supported by the National Science Foundation under grant # IIS-1029711 and # SES-0851705.
- This research is partially supported by the Institute on the Environment (IonE), U. Minnesota.
- This research is partially supported by a grant from the College of Liberal Arts (CLA), U. Minnesota.
- Thanks to colleagues Partha Lahiri, Karen Monsen, Nitai Mukhopadhyay, and students Emily Koenke, Hao Ren, Abhishek Nandy, Heng Zhang, Ying Lu, Sen Yuan, Megan Hyman, Lindsey Deitz.