Some negative results on extreme multivariate quantiles defined through convex optimisation

Gilles Stupfler (University of Nottingham)

Joint work with Stéphane Girard (INRIA Rhône-Alpes)

CMStatistics 2017 Conference
London, UK, 16th December 2017

University of Nottingham
UK | CHINA | MALAYSIA

Outline

- Geometric quantiles
- Asymptotic behaviour of extreme geometric quantiles
- Real data illustration
- What about extreme geometric expectiles?
- Discussion

Multivariate quantiles?

- The natural order on $\mathbb{R}$ induces a universal definition of quantiles for univariate distribution functions.
- This is not true in $\mathbb{R}^d$, $d \geq 2$: no natural order exists in this case.

Many definitions of multivariate quantiles have been suggested:

- Depth-based quantiles: Liu *et al.* (1999), Zuo and Serfling (2000);
- Convex optimisation: Abdous and Theodorescu (1992), Chaudhuri (1996), Koltchinskii (1997).

These are generalisations of univariate quantiles (see Serfling, 2002).

Recent developments include the DOQR paradigm of Serfling (2010), the directional quantiles of Kong and Mizera (2012) linked to the Tukey depth, and level sets-based quantiles (Cousin and Di Bernardino, 2013).

Geometric quantiles

The univariate τ–th quantile of a real-valued random variable X is

$$q(\tau) = \inf\{t \in \mathbb{R} \,|\, \mathbb{P}(X \leq t) \geq \tau\}.$$

This can also be obtained by solving the L^1–optimisation problem

$$\arg\min_{q \in \mathbb{R}} \mathbb{E}(\varphi_\tau(X - q) - \varphi_\tau(X))$$

where φ_τ is the quantile check function (Koenker and Bassett, 1978):

$$\varphi_\tau(x) = |\tau - \mathbb{1}\{x \leq 0\}|\,|x|.$$

A technical trick shows that this optimisation problem is exactly

$$\arg\min_{q \in \mathbb{R}} \mathbb{E}(|X - q| - |X|) - (2\tau - 1)q.$$

In $\mathbb{R}^d$, $d \geq 2$, analogues of the absolute value and product are given by the Euclidean norm $\| \cdot \|$ and Euclidean inner product $\langle \cdot, \cdot \rangle$.

Note also that $\tau \in (0,1) \Rightarrow 2\tau - 1 \in (-1,1)$, the unit ball in $\mathbb{R}$.

This leads to the following notion of geometric quantiles of X:

Definition (Chaudhuri 1996)

If $u \in \mathbb{R}^d$ is an arbitrary vector then a geometric u–th quantile of X, if it exists, is a solution of the optimisation problem

$$\arg\min_{q \in \mathbb{R}^d} \mathbb{E}(\|X - q\| - \|X\|) - \langle u, q \rangle. \qquad (P_u)$$

Geometric quantiles have good, known central properties (uniqueness, orthogonal equivariance, characterisation of the underlying distribution...)

Our focus here is rather to investigate extreme geometric quantiles.

A first step

From now on, assume that the distribution of X is not concentrated on a single straight line in $\mathbb{R}^d$, and (for simplicity) that it is non-atomic.

Proposition (Chaudhuri 1996; Koltchinskii 1997; Girard and S. 2017)

The optimisation problem (P_u) has a solution if and only if $\|u\| < 1$.

Interesting asymptotics are therefore those of a geometric quantile $q(u)$ when $\|u\| \uparrow 1$.

This is exactly an “extreme geometric quantile” as in Chaudhuri (1996).

Theorem (Girard and S. 2017)

Let S^{d-1} be the unit sphere of $\mathbb{R}^d$.

(i) *The magnitude of extreme geometric quantiles diverges to infinity:*

$$\|q(u)\| \to \infty \quad \text{as} \quad \|u\| \uparrow 1.$$

(ii) *The extreme geometric quantile in the direction $u \in S^{d-1}$ has asymptotic direction u:*

$$\frac{q(\alpha u)}{\|q(\alpha u)\|} \to u \quad \text{as} \quad \alpha \uparrow 1.$$

- A consequence of (i) is that the norm of extreme geometric quantiles tends to infinity even if X has a compact support!
- Related point: sample geometric quantiles do not necessarily lie within the convex hull of the sample, see Breckling *et al.* (2001).

Asymptotic behaviour of extreme geometric quantiles

Our next result examines rates of convergence in the previous theorem.

Theorem (Girard and S. 2017)

Let $u \in S^{d-1}$*. Define* $\Pi_u(x) = x - \langle x, u \rangle u$ *to be the projection on* $u^\top$*.*

(i) *If* $\mathbb{E}\|X\| < \infty$ *then*

$$\|q(\alpha u)\| \left(\frac{q(\alpha u)}{\|q(\alpha u)\|} - u \right) \to \mathbb{E}(\Pi_u(X)) \quad \text{as} \quad \alpha \uparrow 1.$$

(ii) *If* $\mathbb{E}\|X\|^2 < \infty$ *and* Σ *denotes the covariance matrix of* X *then*

$$\|q(\alpha u)\|^2 (1 - \alpha) \to \frac{1}{2} \left(\operatorname{tr} \Sigma - u' \Sigma u \right) > 0 \quad \text{as} \quad \alpha \uparrow 1.$$

Consequences of Theorem 2

If $\|X\|$ has a finite second moment, then the magnitude of an extreme geometric quantile in the direction u is, asymptotically, fully determined by u and the covariance matrix Σ, which is a central parameter.

Moreover, the global maximum of the function $u \mapsto \operatorname{tr}\Sigma - u'\Sigma u$ on S^{d-1} is reached at a unit eigenvector of Σ for its smallest eigenvalue. Thus:

- ◊ The norm of an extreme geometric quantile is the largest in the direction where the variance is the smallest;
- ◊ For elliptically contoured distributions, the shapes of extreme geometric quantile contours and iso-density surfaces are orthogonal.

In general, no reliable information about the extremes of a multivariate distribution can be obtained from its extreme geometric quantiles.

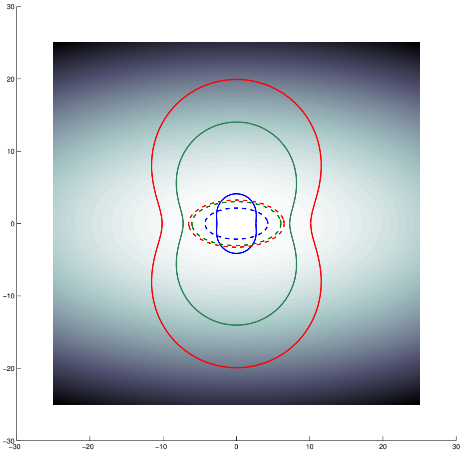

Figure: Full line: iso-quantile curve of level α, dashed line: iso-density curve $\mathcal{C}f_\alpha = \{x \in \mathbb{R}^d \mid f/\|f\|_\infty = (1-\alpha)\}$, for the Gaussian $\mathcal{N}_2(\mathbf{0}_2, \mathrm{diag}(2,1))$. The level α is 0.9 (blue), 0.99 (green), 0.995 (red).

Illustration on the Pima Indians Diabetes dataset

- Two-dimensional data set extracted from the Pima Indians Diabetes Database, downloadable at

  `ftp.ics.uci.edu/pub/machine-learning-databases`

- The data set is $n = 392$ pairs (X_i, Y_i), where X_i is the body mass index of the i-th individual and Y_i is its diastolic blood pressure.
- Already considered in Chaouch and Goga (2010) in the context of outlier detection using precisely geometric quantiles.
- We center the data and, recalling Theorem 2, we estimate extreme geometric quantile contours via

$$\widehat{q}_n(\alpha u) = (1-\alpha)^{-1/2} \left[\frac{1}{2} \left(\operatorname{tr} \widehat{\Sigma}_n - u' \widehat{\Sigma}_n u \right) \right]^{1/2} u,$$

for any $u \in S^{d-1}$, where $\widehat{\Sigma}_n$ is the sample covariance matrix.

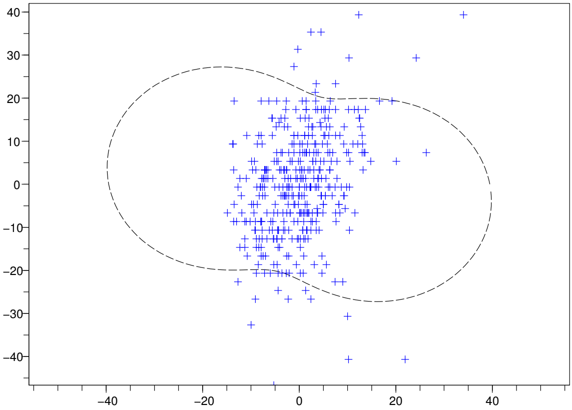

Figure: Pima Indians Diabetes data set: Estimated geometric iso-quantile curve at level $\|u\| = 0.95$. Estimator based on the sample covariance matrix.

What about extreme geometric expectiles?

The τ–expectile of a real-valued random variable X is obtained via an L^2–version of the optimisation problem for quantiles:

$$\arg\min_{q\in\mathbb{R}} \mathbb{E}(\eta_\tau(X-q)-\eta_\tau(X))$$

where η_τ is the expectile check function (Newey and Powell, 1987):

$$\eta_\tau(x) = |\tau - \mathbb{1}\{x \leq 0\}|\, x^2.$$

The technical trick used to rewrite the quantile optimisation problem still applies, and shows that the expectile optimisation problem is exactly

$$\arg\min_{q\in\mathbb{R}} \mathbb{E}(\psi_\tau(X-q)-\psi_\tau(X))$$

with $\psi_\tau(x) = |x|\,(|x| + (2\tau-1)x)$.

Replace again absolute value and product by Euclidean norm $\|\cdot\|$ and Euclidean inner product $\langle\cdot,\cdot\rangle$ to generate the loss function

$$\Psi_u(x) = \|x\|(\|x\| + \langle u, x\rangle).$$

This leads to the following notion of geometric expectiles:

Definition (Herrmann, Hofert and Mailhot 2017)

If $u \in \mathbb{R}^d$ is an arbitrary vector then a geometric u–th expectile of X, if it exists, is a solution of the optimisation problem

$$\arg\min_{q\in\mathbb{R}^d} \mathbb{E}(\Psi_u(X-q) - \Psi_u(X)). \tag{S_u}$$

Our results so far on extreme geometric expectiles

Proposition (Herrmann, Hofert and Mailhot 2017; Girard and S. 2017)

The optimisation problem (S_u) has a solution if and only if $\|u\| < 1$.

This means that, as for geometric quantiles $q(u)$, interesting asymptotics are those of a geometric expectile $e(u)$ when $\|u\| \uparrow 1$.

Theorem (Girard and S. 2017)

The magnitude of extreme geometric expectiles diverges to infinity:

$$\|e(u)\| \to \infty \quad \text{as} \quad \|u\| \uparrow 1.$$

In addition: for any $u \in S^{d-1}$,

$$\frac{e(\alpha u)}{\|e(\alpha u)\|} \to u \quad \text{as} \quad \alpha \uparrow 1.$$

Discussion

- Extreme geometric quantiles in the direction u are asymptotically equal for two distributions which have the same finite covariance matrix. This is not satisfying from the extreme value perspective.
- The shape of the iso-geometric quantile curves may be totally different from the shape of the density contour plots. Outlier detection with this notion should be conducted with great care.
- Our work so far shows that extreme geometric expectiles potentially suffer from the same kind of issues.

The part on geometric quantiles is based on the following paper:

Girard, S. and Stupfler, G. (2017). Intriguing properties of extreme geometric quantiles, *REVSTAT: Statistical Journal* **15**(1): 107–139.

The references for this part of the presentation can be found therein.

Results in the case of an undefined covariance matrix can be found in

Girard, S., Stupfler, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework, *Extremes* **18**(4): 629–663.

The part on geometric expectiles is recent ongoing work.

Thank you for your attention!