Estimating extreme quantile regions for two dependent risks

John H.J. Einmahl

Tilburg University

Joint work *in progress* with

Laurens de Haan (Rotterdam and Lisbon)

ICSA, Piscataway, June 6, 2008

INTRODUCTION

– UNIVARIATE CASE

df F, with quantile function F^{-1}.

Sample size $n = 1000$, say.

Estimate $F^{-1}(0.6)$.

Estimate $F^{-1}(0.999)$. The interval

$$[F^{-1}(0.999), \infty)$$

contains hardly or no data and therefore statistical inference is difficult.

Estimate $F^{-1}(1-p)$.

In particular when we want to protect ourselves against a calamity that has not yet occurred, we consider the case where p is smaller than $1/n$.

Motivation:

SPC

finance (VaR)

flooding

. . .

– BIVARIATE CASE

We simultaneously monitor two dependent risks $X, Y > 0$.

Quantile? No natural definition.

Quadrant?

Region bounded by circle?

Triangle?

Which shape?

Let the pair (X, Y) have df F with density f on $(0, \infty)^2$. Denote the corresponding probability measure with P.

We have a random sample of size n from F.

We define quantile regions determined by the levels of f:

$$Q = \{(x, y) \in (0, \infty)^2 : f(x, y) \leq \varepsilon\}.$$

So, for (very small) p find Q of this type such that $PQ = p$.

Consider $Q^c = \{(x, y) \in (0, \infty)^2 : f(x, y) > \varepsilon\}$. This Q^c has the property that on Q^c, f is larger than on Q, i.e. the quantile region Q is the set of less likely points. As a consequence, Q^c is the region with smallest area such that $PQ^c = 1 - p$.

Now let $p = p_n$ be very small. For the asymptotics think of $np_n \to c \in [0, \infty)$, so $c = 0$ is possible.

How to estimate $Q = Q_n$ nonparametrically?

Since we are working on the boundary of the sample or beyond we ‘need’ Statistics of Extremes.

RESULTS

We assume that $F \in D(G)$, with extreme-value indices $\gamma_1, \gamma_2 > 0$, i.e. when $t \to \infty$

$$(1) \qquad t\,(1 - F(U_1(t)x^{\gamma_1}, U_2(t)y^{\gamma_2})) \to \iint_{u>x \text{ or } v>y} g(u,v)\,du\,dv,$$

on $(0,\infty]^2 \backslash \{(\infty,\infty)\}$, with

$$U_j(t) = F_j^{-1}(1 - 1/t)$$

and F_j, $j = 1, 2$, the marginals of F.

Here g is the density corresponding to $-\log G_0$; G_0 is obtained from G after standardization to standard Fréchet marginals.

We also need a density version of (1): when $t \to \infty$

(2) $$tU_1(t)U_2(t)f(U_1(t)x^{\gamma_1}, U_2(t)y^{\gamma_2}) \to \frac{1}{\gamma_1\gamma_2}x^{1-\gamma_1}y^{1-\gamma_2}g(x,y),$$

on $(0,\infty)^2$.

We have $g(ax, ay) = a^{-3}g(x,y), a > 0$.

We assume that f is non-increasing in each coordinate, outside $(0, M]^2$ for some $M > 0$ and on $(0, M]^2$, f is bounded away from zero. As a consequence the boundary of Q_n is non-increasing.

Recall

$$Q_n = \{(x, y) \in (0, \infty)^2 : f(x, y) \leq \varepsilon\}$$

with ε such that $PQ_n = p$.

Set

$$S = \{(x, y) : x^{1-\gamma_1} y^{1-\gamma_2} g(x, y) \leq \gamma_1 \gamma_2\},$$

see (2). S is a fixed (not depending on n) basis for our estimator of Q_n. We will estimate it later and then transform that estimator, using in particular p.

Throughout let k be a sequence of positive numbers such that $k \to \infty$ and $k/n \to 0$.

A first step is to replace ε by

$$\widetilde{\varepsilon} = \left(\frac{np}{k\nu(S)}\right)^{\gamma_1+\gamma_2+1} \frac{1}{(n/k)U_1(n/k)U_2(n/k)},$$

where ν is the so-called exponent measure, the measure corresponding to $-\log G_0$.

Let $z = (x, y)$ and define, in vector notation, the map T_n by

$$T_n(z) = U(n/k)z^{\gamma}.$$

Write

$$\widetilde{Q}_n = T_n\left(\frac{k\nu(S)}{np}S\right) = U\left(\frac{n}{k}\right)\left(\frac{k\nu(S)}{np}\right)^{\gamma} S^{\gamma}.$$

$\widetilde{Q}_n$ is obtained from

$$\{(x, y) \in (0, \infty)^2 : f(x, y) \leq \widetilde{\varepsilon}\}$$

by using the limiting relation between f and g, i.e. the domain of attraction condition for densities (2).

$\widetilde{Q}_n$ is a good approximation to Q_n: we have

$$\frac{P(Q_n \triangle \widetilde{Q}_n)}{p} \to 0.$$

Here $\triangle$ denotes 'symmetric difference': $A \triangle B = A \backslash B \cup B \backslash A$.

The obvious step to obtain an estimator of Q_n is now to estimate $\tilde{Q}_n$, which can be done by estimating $T_n, \nu(S)$ and in particular S.

We write S in polar coordinates:

$$\left\{(r,\theta) : r \geq \left(\frac{1}{\gamma_1\gamma_2}\psi(\theta)\cos^{1-\gamma_1}\theta\sin^{1-\gamma_2}\theta\right)^{\frac{1}{\gamma_1+\gamma_2+1}}\right\},$$

where

$$\psi(\theta) = g(\cos\theta, \sin\theta).$$

The function ψ is the density of Ψ, the so-called spectral measure. Just like the exponent measure, Ψ describes the dependence structure of the limit distribution G.

In order to estimate these quantities, we have to estimate $U_1(n/k), U_2(n/k), \gamma_1, \gamma_2$ and the spectral density ψ.

Estimation of the first four is well-known. We estimate the two extreme-value indices with the moment estimator.

The estimator for ψ will be obtained by smoothing the empirical likelihood estimator of the spectral measure Ψ in Einmahl and Segers (2008). Hence we obtain $\widehat{Q}_n$.

Indeed, since all these estimators have good asymptotic properties, $\widehat{Q}_n$ is close to $\widetilde{Q}_n$ and hence to Q_n:

THEOREM. Under appropriate conditions we have that

$$\frac{P(\widehat{Q}_n \triangle Q_n)}{p} \xrightarrow{p} 0.$$

This is a consistency result. It is presented in a ratio setting, since $p = p_n \to 0$.

In practice it is important that the k's for estimating γ_1, γ_2 and ψ can be different. The theorem holds true in this case.

SIMULATIONS

(Thanks to Andrea Krajina)

Bivariate Cauchy distribution on first quadrant.

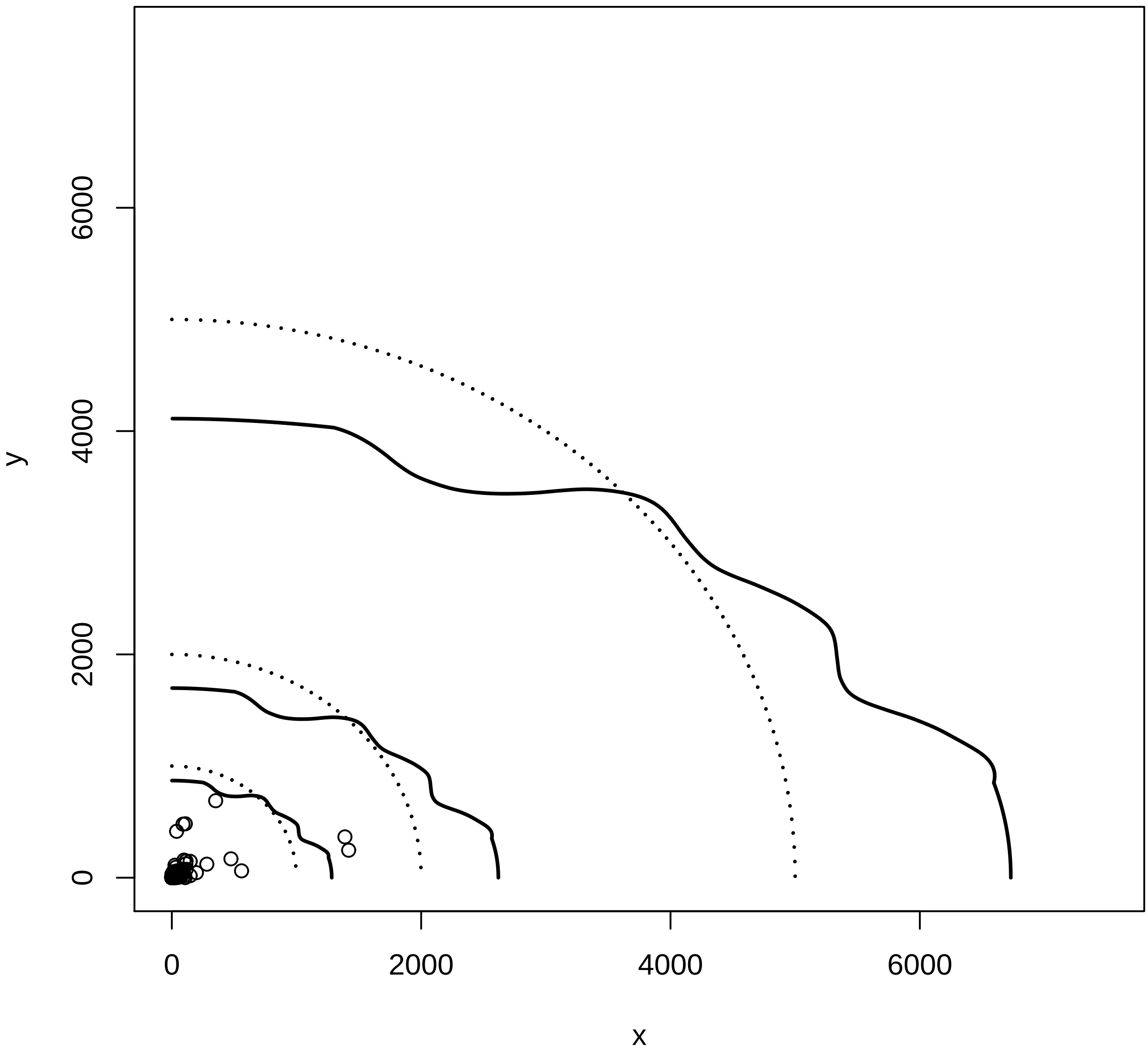
Very heavy tailed distribution.

Density

$$f(x,y) = \frac{2}{\pi(1+x^2+y^2)^{3/2}}, \quad x,y > 0.$$

Two samples of size $n = 2000$.

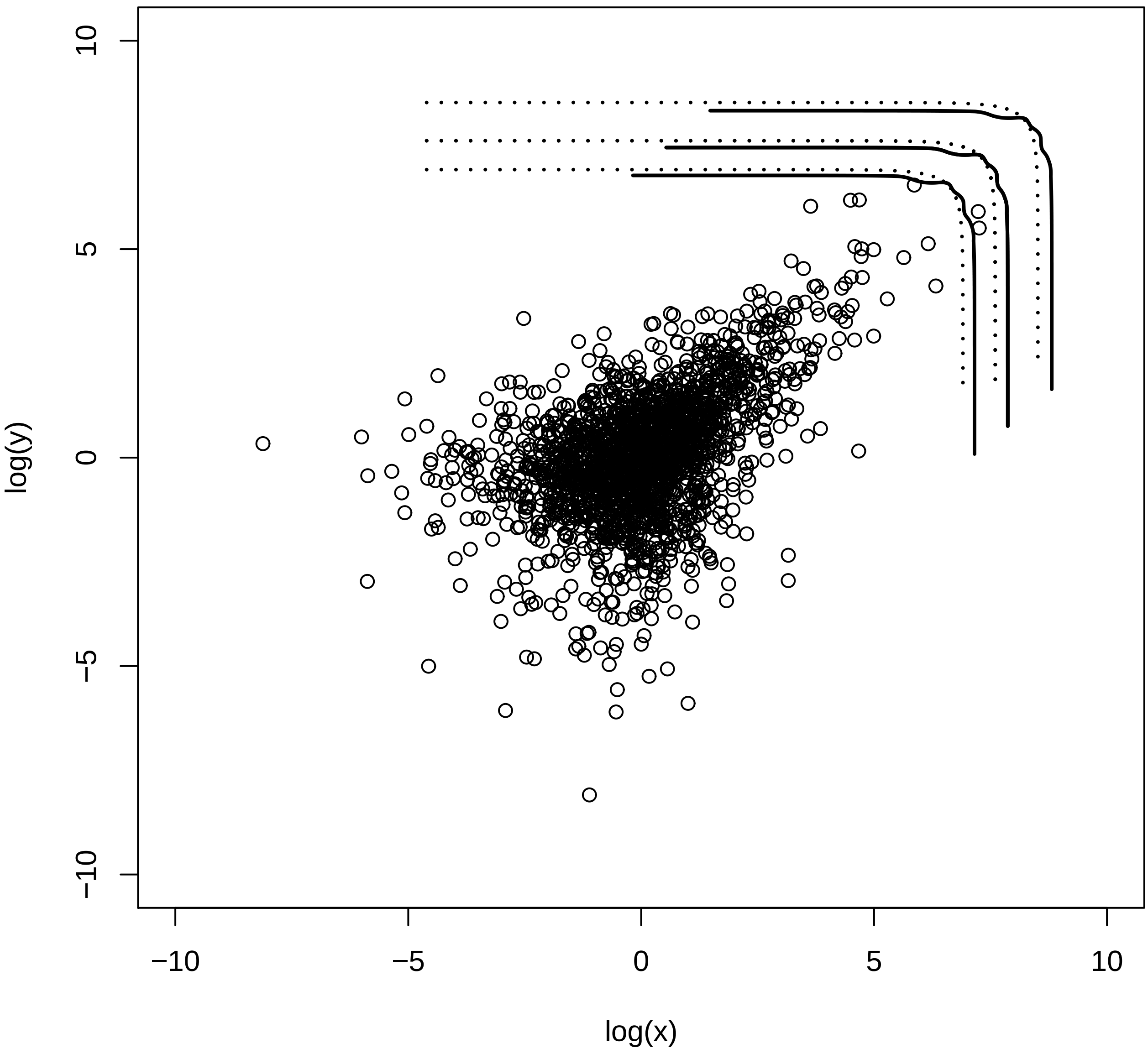
We take $p = 1/1000, 1/2000, 1/5000$. Note that for the latter p we have $np = 0.4$.



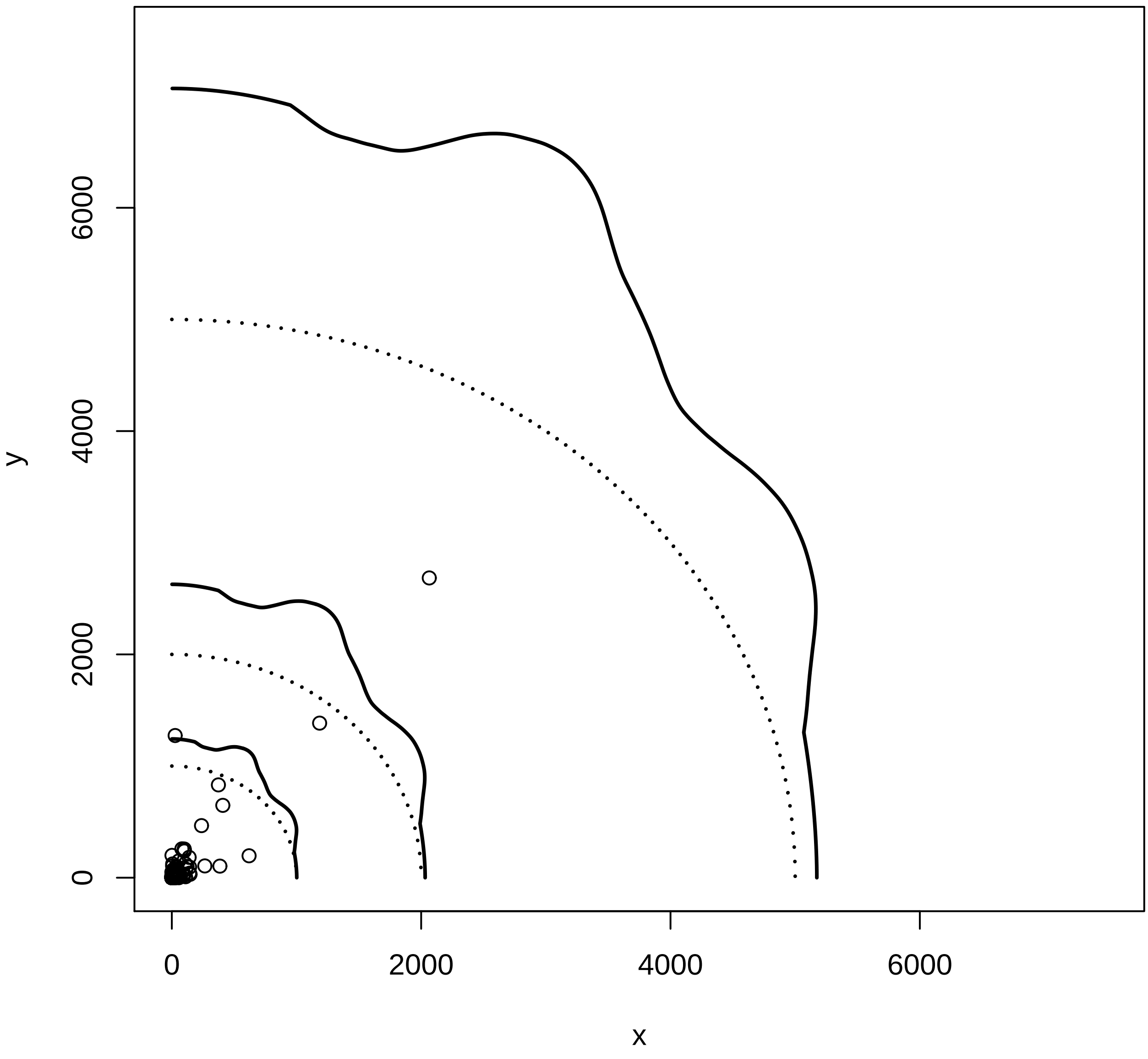
first sample of size 2000

$p = 1/1000, 1/2000, 1/5000$

· · · · · · true ——— estimator



on log scale for different viewing



second sample of size 2000

$p = 1/1000, 1/2000, 1/5000$

· · · · · · true ——— estimator