Summary of 'Random threshold driven tail dependence measures with application to precipitation data analysis'

Liujun Chen

March 11, 2020

Tail dependence definition

Two identically distributed random variables X and Y with distribution F are called tail independent, if

$$\lambda = \lim_{u \to x_F} P(Y > u | X > u)$$

exists and equal to 0, where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. The quantity λ, if exists, is called the bivariate tail dependence index.

Some Examples

- The tail dependence index of a bivariate normal (Gaussian) random vector is zero as long as the corresponding correlation coefficient is less than one.
- The tail dependence index of a bivariate t random vector with a positive correlation is greater than zero.

Dependent random variables are not necessarily tail dependent.

Focus of the Paper

Two fundamental questions:

- How to identify tail dependencies between variables.
- How to develop statistical models dealing with tail dependencies.

This paper is focus on the first issue.

Hypotheses of tail independence

$$\mathrm{H}_0 : X \text{ and } Y \text{ are tail independent,}$$
$$\leftrightarrow H_1 : X \text{ and } Y \text{ are tail dependent,}$$

which can also be written as

$$H_0 : \lambda = 0 \longleftrightarrow H_1 : \lambda > 0.$$

The remaining of the paper mainly discuss how to test the null and how to estimate λ under the null hypothesis.

Remarks of hypotheses of tail independence

- The null and alternative hypotheses in Ledford and Tawn (1996, 1997) are reversed in this paper.

$$\mathrm{H}_0 : X \text{ and } Y \text{ are tail dependent,}$$
$$\leftrightarrow H_1 : X \text{ and } Y \text{ are tail independent.}$$

- Other significant tests include Falk and Michel(2006),Hüsler and Li (2009), Bacro, Bel, and Lantuéjoul (2010) etc.

Tail independence test in Falk and Michel (2006)

Let $\epsilon > 0$ and $t \in [0, 1]$. When ϵ tends to 0, the conditional distribution function

$$K_\varepsilon(t) \equiv P\left(X^{-1} + Y^{-1} < \varepsilon t | X^{-1} + Y^{-1} < \varepsilon\right)$$

tends to t^2 if X and Y are asymptotically independent, and t otherwise. This result can be used to test for the asymptotic independence of X and Y using classical goodnessof- fit tests such as the Kolmogorov–Smirnov or the likelihood ratio as well as the chi-square test.

Tail independence test in Hüsler and Li (2009)

Assume some conditions, under H_0,

$$\mathrm{H}_0 : l(x,y) = x + y \quad \text{for all } x, y > 0,$$

it can be proved

$$\{T_n(x,y), x, y \in [0,1]\} \to \{W_1((1+\theta)x) + W_2((1+\theta)y), x, y \in [0,1]\}$$

where

$$T_n(x,y) := \sqrt{k}\left(l_n(x,y) - (x+y)\right), \quad x, y \in [0,1],$$

and W_1, W_2 are independent Brownian motions.

Tail independence test in Hüsler and Li (2009)-Continue

And

$$\int_0^1 \int_0^1 k \left(l_n(x,y) - (x+y)\right)^2 \mathrm{d}x\mathrm{d}y$$

$$\xrightarrow{\mathrm{d}} \int_0^1 \int_0^1 \left(W_1((1+\theta)x) + W_2((1+\theta)y)\right)^2 \mathrm{d}x\mathrm{d}y.$$

Tail independence test in Bacro, Bel, and Lantuéjoul (2010)

Consider the random variable

$$W = \frac{1}{2}|F(X) - F(Y)|,$$

it can be shown

$$\sqrt{n}\frac{\widehat{v_W} - \frac{1}{6}}{\widehat{\sigma_W}} \to N(0,1),$$

where $\widehat{V_W}$ is an estimator for $E(W)$, and $\widehat{\sigma_W}$ is an estimator for σ_W.

TQCC and its properties

- From the definition of λ, the estimation of λ mainly relies on choice of thresholds u the dependence mechanism between upper tails of two random variables X and Y.
- This suggests that one can ignore values of X and Y below u and construct sample-based measures of tail dependence.

Definition of TQCC

If $\{(X_i, Y_i)\}_{i=1}^n$ is a random sample of unit Fréchet random variables (X, Y),

$$\mathrm{q}_{n,t} = \frac{\max_{i\leq n} \frac{\max(X_i, u_n)}{\max(Y_i, u_n)} + \max_{i\leq n} \frac{\max(Y_i, u_n)}{\max(X_i, u_n)} - 2}{\max_{i\leq n} \frac{\max(X_i, u_n)}{\max(Y_i, u_n)} \times \max_{i\leq n} \frac{\max(Y_i, u_n)}{\max(X_i, u_n)} - 1},$$

is the tail quotient correlation coefficient(TQCC).
If $u_n \equiv u$, the TQCC reduces to the one in Zhang (2008). In this paper, u_n is allowed to be random.

Quotient correlation

The quotient correlation, which can be used as an alternative to Pearson's correlation,is defined as

$$\mathrm{q}_n = \frac{\max_{i\le n}\{Y_i/X_i\} + \max_{i\le n}\{X_i/Y_i\} - 2}{\max_{i\le n}\{Y_i/X_i\} \times \max_{i\le n}\{X_i/Y_i\} - 1}.$$

Zhang(2008) proved that if X and Y are independent, then

$$nq_n \xrightarrow{\mathcal{L}} \zeta,$$

where ζ is a gamma(2, 1) random variable.

Some properties of TQCC

- Take $\zeta_n^{\mathrm{I}} = \max_{1\le i\le n}\left\{\max\left(X_i, u_n\right)/\max\left(Y_i, u_n\right)\right\}$ and $\zeta_n^{\mathrm{II}} = \max_{1\le i\le n}\left\{\max\left(Y_i, u_n\right)/\max\left(X_i, u_n\right)\right\}$.
- Now $\mathrm{q}_{u_n} = f\left(\zeta_n^{\mathrm{I}}, \zeta_n^{\mathrm{II}}\right)$, where

$$f(x,y) = \frac{x+y-2}{xy-1}, \quad forx \ge 1, y \ge 1, x+y > 2.$$

- It is easy to see that $0 < f(x,y) < 1, f(1,y) = 1, f(x,1) = 1$, and $\partial f/\partial x < 0, \partial f/\partial y < 0$.

Some properties of TQCC-Continue

The properties of $\mathrm{q}_{u_n} = f\left(\zeta_n^{\mathrm{I}}, \zeta_n^{\mathrm{II}}\right)$ can be summarised as follows.

- q_{u_n} takes values between 0 and 1 for a sufficiently large n.
- For a fixed sample size n, the larger the ζ_n^{I} or the ζ_n^{II},the smaller the q_{u_n},hence the less the agreement of changing magnitudes at tails.
- As long as one of ζ_n^{I} and ζ_n^{II} is very large, q_{u_n} is largely determined by the smaller one.

Proposition 1

The tail dependence index $\lambda_{X,Y}$ of X and Y is equal to the tail dependence index $\lambda^*_{X,Y}$ of $\max(X, u_n)$ and $\max(Y, u_n)$ using a random threshold, $u_n = T_{n,t}$, where $T_{n,t}$ is a Fréchet variable with the distribution function $\exp(-n/x^t)$ for $x > 0, n \geq 0$ and $t > 1$.

$$\lim_{u\to\infty} \frac{P(X > u, Y > u)}{P(X > u)} = \lim_{u\to\infty} \frac{P\left\{\max\left(X, T_{n,t}\right) > u, \max\left(Y, T_{n,t}\right) > u\right\}}{P\left\{\max\left(X, T_{n,t}\right) > u\right\}}$$

Main Assumption T1

For $1 < t < 1+\delta, \delta > 0$, paired tail independent random variables (X_i, Y_i) satisfy

$$\frac{\max_{1\leq i\leq n} \max\left(X_i, T_{n,t}\right) / \max\left(Y_i, T_{n,t}\right)}{\max_{1\leq i\leq n} \max\left(X_i, T_{n,t}\right) / T_{n,t}} = 1 + o_p(1),$$

$$\frac{\max_{1\leq i\leq n} \max\left(Y_i, T_{n,t}\right) / \max\left(X_i, T_{n,t}\right)}{\max_{1\leq i\leq n} \max\left(Y_i, T_{n,t}\right) / T_{n,t}} = 1 + o_p(1).$$

Corollary 1

If random variables $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ satisfy T1, where X_i and Y_i -are unit Fréchet random variables, under the null hypothesis,

$$2n\left\{1-\exp\left(-\frac{1}{T_{n,t}}\right)\right\}\mathrm{q}_{T_{n,t}} \xrightarrow{\mathcal{L}} \chi_4^2,$$

where χ_4^2 is a chi-squared random variable with four degrees of freedom.

Corollary 2

If random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ satisfy T1,and $u_n = u_n^* a_n$ satisfies $u_n^* \xrightarrow{\mathrm{P}} u^*, a_n \to \infty$, and $a_n/n \to 0$ as $n \to \infty$, where X_i and Y_i -are unit Fréchet random variables, under the null hypothesis,

$$2n\left\{1 - \exp\left(-1/u_n\right)\right\} \mathrm{q}_{u_n} \xrightarrow{\mathcal{L}} \chi_4^2.$$

Condition A1

Let $\{\xi_{i,j} : i \geq 1, -\infty < j < \infty\}$ be mutually independent unit Fréchet random variables, and suppose that tail dependent unit Fréchet random variables X_i and Y_i have the approximation representations

$$X_i = \max_{-\infty<j<\infty} \alpha_j \xi_{i,j}, \quad Y_i = \max_{-\infty<j<\infty} \beta_j \xi_{i,j},$$

where $\alpha_j \geq 0, \beta_j \geq 0, \sum_{j=-\infty}^{\infty} \alpha_j = 1, \sum_{j=-\infty}^{\infty} \beta_j = 1$, and $\alpha_j > 0$ iff $\beta_j > 0$ for all j. Then there exist $0 < c_1 < c_2 < \infty$ such that $c_1 \leq \alpha_j/\beta_j \leq c_2$.

Theorem 5

Suppose that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ satisfy A1. Then for $u_n = T_{n,t} \xrightarrow{P} \infty$ in Theorem 3, and $u_n = u_n^* a_n$ with $u_n^* \xrightarrow{\mathrm{P}} u^* \in (0, \infty), a_n \to \infty$ and $a_n/n \to 0$ in Theorem 4, $2n \{1 - \exp(-1/u_n)\} \mathrm{q}_{u_n} \xrightarrow{\mathrm{P}} \infty$, and the test is consistent.

Proposition 3

Suppose $\alpha_1 + \alpha_2 = 1$ and $\beta_1 + \beta_2 = 1$. Then $\mathrm{q}_{u_n} \xrightarrow{\mathrm{P}} \lambda$.

Marginal Transformation

The GEV has the form

$$H(x;\xi,\mu,\psi) = \exp\left[-\left\{1+\frac{\xi(x-\mu)}{\psi}\right\}_+^{-1/\xi}\right].$$

In principle, unit Fréchet scales can be obtained using the transformation $-1/\log\{H(x;\xi,\mu,\psi)\}$.

Theorem 4

Let $\{(X_i, Y_i)\}_{i=1}^{n}$ be a sample of $T1$ type tail independent random variables (X, Y) whose marginal domains of attraction are GEV random variables with shape parameters $\xi_X = \xi_{0;X}$ and $\xi_Y = \xi_{0;Y}$, respectively. Suppose estimators of ξ_X and ξ_Y, satisfy $n^{\alpha_X}\left(\hat{\xi}_X - \xi_{0;X}\right) \xrightarrow{\mathcal{L}} W_X$ and $n^{\alpha_Y}\left(\hat{\xi}_Y - \xi_{0;Y}\right) \xrightarrow{\mathcal{L}} W_Y$, where $\alpha_X > 0$, $\alpha_Y > 0$ and W_X and W_Y are random variables. If $\hat{X}_i = -1/\log\left\{H\left(X_i; \hat{\xi}_X\right)\right\}$, $\hat{Y}_i = -1/\log\left\{H\left(Y_i; \hat{\xi}_Y\right)\right\}$ and
$\hat{\mathrm{q}}_{u_n} = \frac{\max_{1\le i\le n}\left\{\max\left(\hat{X}_i,u_n\right)/\max\left(\hat{Y}_i,u_n\right)\right\}+\max_{1\le i\le n}\left\{\max\left(\hat{Y}_i,u_n\right)/\max\left(\hat{X}_i,u_n\right)\right\}-2}{\max_{1\le i\le n}\left\{\max\left(\hat{X}_i,u_n\right)/\max\left(\hat{Y}_i,u_n\right)\right\}\times\max_{1<i\le n}\left\{\max\left(\hat{Y}_i,u_n\right)/\max\left(\hat{X}_i,u_n\right)\right\}-1}$,
then

$$2n\left\{1-\exp\left(-1/u_n\right)\right\}\hat{\mathrm{q}}_{u_n} \xrightarrow{\mathcal{L}} \chi_4^2.$$

Data Application

- The data are daily precipitation totals covering the period 1950-1999 over 5873 stations in the continental USA.
- Fit GEV to each of 5,873 series, performed marginal transformations, conduct TQCC-based tail independence tests, and report the tail dependence measure (TQCC) after controlling the false discover rate (FDR) at level 0.05.

GEV fitting and extreme precipitation comparison

Table 1. Information of six selected stations. The top three stations have smallest $\hat{\epsilon}$, and the bottom three have the largest.

Station ID	Shape $\hat{\xi}$ (s.e.)	Latitude	Longitude	Elevation	City name
(I)	−0.2316 (0.0604)	40.45	−111.70	1,720	Timpanogos, UT
(II)	−0.2300 (0.0711)	40.15	−79.03	558	Boswell, PA
(III)	−0.2248 (0.0703)	42.02	−86.25	265	Eau Claire, MI
(IV)	0.4383 (0.1065)	29.15	−95.45	8	Angleton, TX
(V)	0.4409 (0.0960)	31.30	−86.52	76	Andalusia, AL
(VI)	0.4977 (0.0970)	47.55	−116.17	680	Kellogg, ID

GEV fitting and extreme precipitation comparison

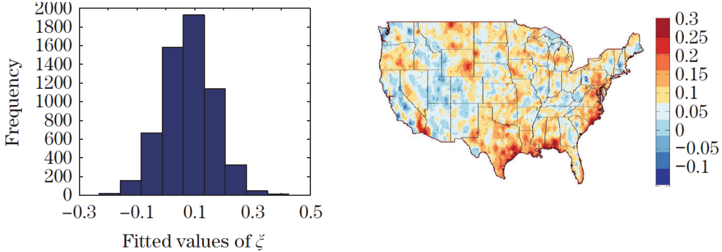

Figure 3. Fitted values of ξ from time series at each of 5,873 stations. The left panel shows the distribution of $\hat{\xi}$. The right panel plots $\hat{\xi}$ to US map using the inverse distance method. Note that the values larger than 0.3 are truncated to 0.3 in order to display an overall visual impression.

Overall tail dependency across all stations

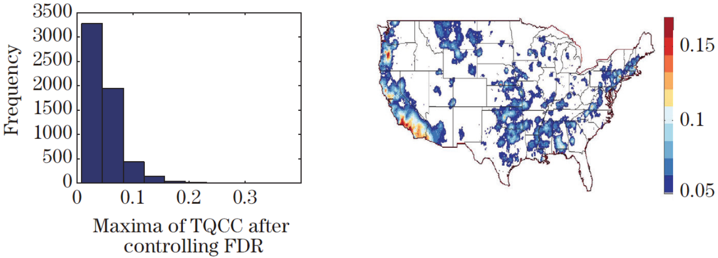

Figure 5. Maximal precipitation tail dependence between one station and the remaining stations on the same day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.

Overall tail dependency across all stations

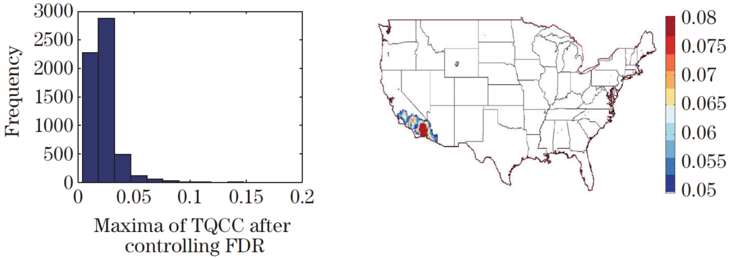

Figure 6. Maximal precipitation tail dependence between one station and the remaining stations on the lagged-1 day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.

Overall tail dependency across all stations

to the colorbar upper limit as they are just a few points.

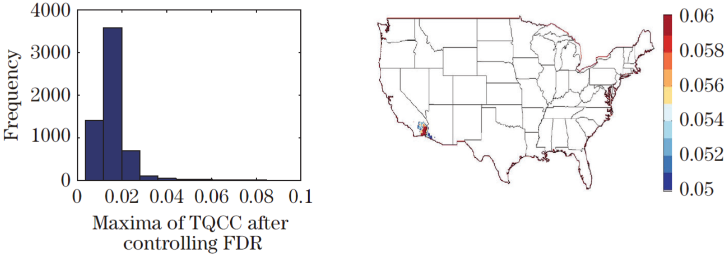

Figure 7. Maximal precipitation tail dependence between one station and the remaining stations on the lagged-7 day. The colorbar on the right panel has been adjusted to reflect the left panel with TQCCs being larger than 0.05, but the very large values are truncated to the colorbar upper limit as they are just a few points.