High-Dimensional Extremes and Copulas

Haijun Li

lih@math.wsu.edu
Washington State University

CIAS-CUFE, Beijing, January 3, 2014
Outline

1. From Tukey to High-Dimensional Data Analysis
2. Univariate Extreme Value Theory: A Primer
3. Multivariate Extremes
4. Copula Representation of Multivariate Regular Variation
5. Tail Densities
6. Tail Densities of Vines
7. Concluding Remark
8. References
John Wilder Tukey, 1915-2000

- A chemist first and then a pure mathematician by training (PhD thesis “On Denumerability in Topology” in 1939)
- Statistician, information scientist, Presidential advisor...

Figure: From Paul Halmos’ “I Have a Photographic Memory”
Tukey’s Achievements

- Teichmüller-Tukey Lemma, equivalent to the Axiom of Choice in axiomatic set theory
- Tukey theory of analytic ideals, Tukey reducibility, ... (A competitor to Bourbaki’s approach to topology)
- Computation: Fast Fourier Transform
- Statistics: Box plot, Quenouille-Tukey Jackknife...
- High-dimensional data analysis: Projection Pursuit (seeking non-Gaussianity)
Tukey’s Collaborators and Awards

- Focused on obtaining results rather than collecting credits.
- More than 100 collaborators and 55 PhD students. Worked well with Samuel Wilks, Walter Shewhart, Claude Shannon, John von Neumann, Richard Feynman, etc.
- National Medal of Science, National Academy of Science, Foreign Member of the Royal Society, IEEE Medal of Honor, Samuel S. Wilks Memorial Award...
Tukey’s Belief

- Much statistical methodology placed too great an emphasis on confirmatory data analysis rather than exploratory data analysis. (Also see Leo Breiman, “Statistical Modeling: The Two Cultures”, Statistical Science, 2001)
- People should often start with their data and then look for a theorem, rather than vice versa.

Future of Data Analysis (Tukey, 1962)

Data Analysis VS Mathematical Statistics? ... “All in all, I have come to feel that my central interest is in data analysis ...”
A respected mathematician turned his back on mathematical proof, focusing on analyzing data instead.

“He legitimized that, because he wasn’t doing it because he wasn’t good at math,’ Mr. Wainer (a Tukey’s former student) said. “He was doing it because it was the right thing to do.”

Univariate Extremes

- $X_1, \ldots, X_n$ are iid with df $F$.
- $M_n := \max\{X_1, \ldots, X_n\} =: \vee_{i=1}^n X_i$.
- $\land_{i=1}^n X_i := \min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$, and so we focus on large extremes only.

**Definition**

If there exist a non-degenerate df $H$, two normalizing sequences $\{c_n\}$ and $\{d_n\}$ with $d_n > 0$, such that

$$\lim_{n \to \infty} P \left( \frac{M_n - c_n}{d_n} \leq x \right) = \lim_{n \to \infty} F^n(c_n + d_nx) = H(x), \ x \in \mathbb{R},$$

for all continuity points $x$ of $H$, then $F$ is said to be in the maximum domain of attraction of $H$, and this is denoted as $F \in \text{MDA}(H)$ or $X_n \in \text{MDA}(H)$. 
Generalized Extreme Value Distribution (EV)

**Fisher-Tippett-Gnedenko Theorem**

If \( F \in \text{MDA}(H) \), then

\[
H(x; \gamma) = \exp\{- (1 + \gamma x)^{-1/\gamma}\}, \forall x \in \mathbb{R},
\]

where \( \gamma \in \mathbb{R} \) is called the extreme value index.

After some location-scale transforms, we have

- Fréchet distribution: \( H_+(x; \theta) = \exp\{-x^{-\theta}\}, x > 0, \theta > 0 \).
- Gumbel or double-exponential distribution: \( H_0(x) = \exp\{-e^{-x}\}, x \in \mathbb{R} \).
- (Reverse) Weibull distribution: \( H_-(x; \theta) = \exp\{-(x)^{\theta}\}, x < 0, \theta > 0 \).
Precise Tail Variations

- $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}$, the upper boundary of the support.
- $F^{←}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$, $0 < u < 1$, the left-continuous inverse.

Gnedenko-de Haan Theorem: Fréchet MDA Case

$F \in \text{MDA}(H(\cdot; \gamma))$ with $\gamma > 0$ if and only if $x_F = \infty$ and

$$
\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad x > 0,
$$

where $c_n = 0$ and $d_n = F^{←}(1 - n^{-1})$. 
Precise Tail Variations (cont’d)

Gnedenko-de Haan Theorem: Gumbel and Weibull MDA Cases

- $F \in \text{MDA}(H_0(\cdot))$ if and only if $\int_{t_0}^{x_F} (1 - F(x)) dx < \infty$ for some $t_0 < x_F$ and

$\lim_{t \to x_F} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp\{-x\}$, $x \in \mathbb{R}$

where $R(t) := \int_t^{x_F} (1 - F(x)) dx / (1 - F(t))$, $t < x_F$ and $c_n = F^{-}(1 - n^{-1})$ and $d_n = R(c_n)$.

- $F \in \text{MDA}(H(\cdot; \gamma))$ with $\gamma < 0$ if and only if $x_F < \infty$ and

$\lim_{t \to \infty} \frac{1 - F(x_F - (tx)^{-1})}{1 - F(x_F - t^{-1})} = x^{1/\gamma}$, $x > 0$,

where $c_n = x_F$ and $d_n = x_F - F^{-}(1 - n^{-1})$. 
Consider the Pareto distribution

\[ F(x) = 1 - x^{-1/\gamma}, \quad x > 1, \quad \gamma > 0. \]

- Since

\[ \lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \]

\( F \in \text{MDA}(H(\cdot; \gamma)) \) with Fréchet distribution \( H(\cdot; \gamma), \gamma > 0. \)

- Set normalizing constants \( c_n = 0 \) and \( d_n = F^{-1}(1 - n^{-1}) = n^\gamma, \)

and clearly, for \( x > 0, \)

\[ \lim_{n \to \infty} F^n(c_n + d_n x) = \lim_{n \to \infty} (1 - x^{-1/\gamma}/n)^n = \exp \{-x^{-1/\gamma}\}. \]
Without loss of generality, assume that $X > 0$.

$X \sim F \in \text{MDA}(H) \iff \lim_{n \to \infty} F^n(c_n + d_n x) = H(x)$.

Since $-\log x \approx 1 - x$ as $x \uparrow 1$, it can be rephrased as follows

$$\lim_{n \to \infty} n(1 - F(c_n + d_n x)) = -\log H(x).$$

That is, $X \sim F \in \text{MDA}(H) \iff$ the tail dispersion stability:

$$\lim_{n \to \infty} n P \left( \frac{X - c_n}{d_n} > x \right) = -\log H(x), \ \forall x > 0.$$
A Dimension-Free Tail Property of MDA

- Define a (Radon) exponent measure
  \[ \mu(x, \infty) := -\log H(x), \quad \forall \ x \in \mathbb{R}_+. \]

- Rewrite: \( \forall \) “good” sets \( B \subset \mathbb{R}_+ \setminus \{0\} \), \( \mu(\partial B) = 0 \),

\[
\lim_{n \to \infty} n P \left( \frac{X - c_n}{d_n} > x \right) = -\log H(x)
\]

\[
\iff \lim_{n \to \infty} n P \left( \frac{X - c_n}{d_n} \in B \right) = \mu(B).
\]

- Tail Dispersion

- Note that the exponent measure \( \mu(\cdot) \) only has an asymptotic scaling property:

\[
\mu(t B) \approx t^{-1/\gamma} \mu(B), \quad \forall \ “good” \ sets \ B \subset \mathbb{R}_+ \setminus \{0\}, \ t \to \infty.
\]
We want: $\mu(tB) = t^{-1/\gamma} \mu(B)$, $\forall$ “good” sets $B \subset \mathbb{R}_+ \setminus \{0\}$, $\forall t > 0$.

- The tail stability and scaling allow us to estimate a tail region containing fewer observations by a tail region containing more observations, whose distribution has the same shape.
- Extreme Value Theory = Mechanism for relating the probabilistic structure within the range of the observed data to regions of greater extremity (Anderson, 1994).
Definition

A random variable $Z$ is regularly varying in the standard form if

$$\lim_{t \to \infty} t P \left( \frac{Z}{t} \in B \right) = \nu(B), \quad \forall \text{ "good" sets } B \subset \mathbb{R}_+ \setminus \{0\}, \quad \mu(\partial B) = 0.$$

Note that $\nu(tB) = t^{-1} \nu(B)$ for any $t > 0$.

- Consider two functions $c(\cdot), d(\cdot)$ such that
  $$\frac{c(tx) - c(t)}{d(t)} \to \psi(x) \neq 0, \quad \forall x > 0, \quad t \to \infty.$$

- Observe that
  $$\left\{ \frac{c(Z)}{d(t)} > x \right\} = \left\{ \frac{Z}{t} > \psi^{-1}(x) \right\}$$
Consider two functions $c(\cdot), d(\cdot)$ such that

$$
\frac{c(tx) - c(t)}{d(t)} \to \psi(x) \neq 0, \quad \forall x > 0, \quad t \to \infty.
$$

1. If $Z$ is standard regularly varying, then

   $$X := c(Z) \in \text{MDA}(H)$$

   for some EV distribution $H$.

2. If $X \in \text{MDA}(H)$ for an EV distribution $H(x)$, then there exists a monotone transformation $c(t)$ such that

   $$Z := c^{-1}(X)$$

   is standard regularly varying.
Remark

- Regular variation is also the key ingredient in weak convergence to stable laws (Meerschaert and Scheffler, 2001).
- Regular variation can be extended to general metric spaces to deal with functional data (Hult and Lindskog, 2006; Bingham and Ostaszewski, 2008).
- “Extreme value distributions” for data clouds in high-dimensional or functional spaces are difficult to define. Extremes emerging from high-dimensional data clouds may be best studied by exploring stability patterns of tail probability decays near data boundaries (Ledoux and Talagrand, 1991; Balkema and Embrechts, 2007).
- Tail risk measures (e.g., Conditional Tail Expectation) can be derived directly from regular variation properties (see, e.g., Joe and Li, 2011).
von Mises Condition

Let $dF$ have a positive derivative on $[x_0, x_F)$ for some $0 < x_0 < x_F$, and $h_F(x) = F'(x)/(1 - F(x))$ be the hazard rate of $F$ in a left neighborhood of $x_F$.

1. If there exist $\gamma \in \mathbb{R}$ and $c > 0$, such that
   
   $$x_F = \begin{cases} 
   \infty & \text{if } \gamma \geq 0 \\
   -\gamma^{-1} & \text{if } \gamma < 0
   \end{cases}$$

   $$\lim_{x \to x_F} (1 + \gamma x) h_F(x) = c,$$

   then $F \in \text{MDA}(H(\cdot; \gamma/c))$.

2. If the derivative of $F$ is monotone in a left neighborhood of $x_F$ for some $\gamma \neq 0$, and if $F \in \text{MDA}(H(\cdot; \gamma/c))$ for some $c > 0$, then $F$ satisfies (1).
Consider the df

\[ F(x) = 1 - \frac{1}{\log x}, \quad x > e, \]

that satisfies that

\[ \lim_{x \to \infty} (x \log x) h_F(x) = 1. \]

Since the reciprocal hazard rate of this distribution is not asymptotically linear, there is no extreme value limit for \( F \), with affine thresholds.

Note, however, that for the case where \( \gamma = 0 \), the von Mises condition is not necessary. For example, it is well known that the standard normal df \( \Phi \in \text{MDA}(H_0(\cdot)) \), but asymptotically

\[ \lim_{x \to \infty} x / h_\Phi(x) = 1. \]
Laws of Small Numbers

- Let \( \{Z_1, \ldots, Z_n\} \) be iid and standard regular varying:

\[
n P (Z_1 \in n B) \to \nu(B), \quad \forall \text{ "good" sets } B \subset \mathbb{R}_+ \setminus \{0\}, \quad \mu(\partial B) = 0,
\]

with scaling \( \nu(t B) = t^{-1} \nu(B) \) for any \( t > 0 \).

- In particular, take \( B = (x, \infty) \) for \( x > 0 \), we have

\[
\lim_{n \to \infty} n P (Z_1 > nx) = \nu((x, \infty)) =: \nu.
\]

- Construct a counting process of exceedances of scaled observations before time \( t \):

\[
N_n(t) = \sum_{i=1}^{n} I_{\{\frac{i}{n} \leq t, \ Z_i > nx\}},
\]

where \( I_A \) denotes the indicator function of set \( A \).
The count of exceedences before $t$:

$$N_n(t) = \sum_{i=1}^{n} I\left\{ \frac{i}{n} \leq t, Z_i > nx \right\},$$

- Take the expectation, we have

$$E(N_n(t)) = t \times nP(Z_1 > nx) \to t\nu, \ n \to \infty.$$

- The counting processes $(N_n(t), t \geq 0)$ converge to a Poisson process with rate $\nu$, as $n \to \infty$.

- The rate $\nu$, or exponent measure $\nu(\cdot)$ (also called the intensity measure for a good reason) encodes all the information of extremes.
Multivariate Extremes

- $X_n = (X_{1,n}, \ldots, X_{d,n}), n = 1, 2, \ldots$, are iid with df $F(x_1, \ldots, x_d)$.
- $M_n := (M_{1,n}, \ldots, M_{d,n})$, where $M_{i,n} := \vee_{j=1}^{n} X_{i,j}, 1 \leq i \leq d$.

**Notation:** For any $x, y \in \mathbb{R}^d$, the sum $x + y$, quotient $x/y$, and the vector inequalities are all operated component-wise.

**Definition: Multivariate Extreme Value Distribution (MEV)**

If there exist a df $G$ with non-degenerate margins, two normalizing sequences of real vectors $\{c_n\}$ and $\{d_n\}$ with $d_n > 0$, such that

$$\lim_{n \to \infty} P \left( \frac{M_n - c_n}{d_n} \leq x \right) = \lim_{n \to \infty} F^n(c_n + d_n x) = G(x), \ x \in \mathbb{R}^d,$$

for all continuity points $x$ of $G$, then $F$ is said to be in the maximum domain of attraction of $G$, and this is denoted as $F \in \text{MDA}(G)$ or $X_n \in \text{MDA}(G)$. 
Let $X = (X_1, \ldots, X_d) \sim F$

- $X_i \sim F_i$: the $i$-th marginal df of $F$.
- $G_i$: the $i$-th marginal df of $G$.
- $F_i \in \text{MDA}(G_i)$, where $G_i$ is the univariate, parametric EV distribution.
- We assume that $F_i$ is standard regularly varying.

### Scaling Property Emerged from Equalizing Margins

1. $G_i$ is the standard Fréchet distribution with
   \[
   \log G_i(t x_i) = t^{-1} \log G_i(x_i), \quad \forall x_i \in \mathbb{R}_+.
   \]

2. Jointly, the multivariate scaling emerges from the limiting process:
   \[
   \log G(t x) = t^{-1} \log G(x), \quad \forall x \in \mathbb{R}_+^d.
   \]
A Short Proof for Multivariate Scaling:

The normalizing vectors: \( c_n = (0, \ldots, 0), \ d_n = (n, \ldots, n) \).

1. Let \( k \) be any positive integer. We have

\[
\lim_{n \to \infty} P \left( \frac{M_{kn}}{n} \leq x \right) = \lim_{n \to \infty} \left( F^n(nx) \right)^k = G^k(x), \quad x \in \mathbb{R}^d.
\]

2. On the other hand,

\[
\lim_{n \to \infty} P \left( \frac{M_{kn}}{n} \leq x \right) = \lim_{n \to \infty} P \left( \frac{M_{kn}}{kn} \leq \frac{x}{k} \right) = G \left( \frac{x}{k} \right), \quad x \in \mathbb{R}^d.
\]

3. (1)+(2) yields that \( G^k(x) = G \left( \frac{1}{k} x \right) \) for any positive integer \( k \).

4. Rewrite (3) as \( G(kx) = G^{1/k}(x) \) for \( x \in \mathbb{R}^d \).

5. (3)+(4) yields that \( G^{k/r}(x) = G \left( \frac{r}{k} x \right) \) for any two positive integers \( k, r \).

6. By taking limits, we have \( G^{t^{-1}}(x) = G(tx) \) for any real \( t > 0 \). \( \square \)
Remark on scaling $G(t \, x) = G^{t^{-1}}(x), \ t > 0$

- Since margins $G_i$s are standard Fréchet,

$$G_i^{-1}(u) = (-\log u)^{-1}, \ 0 \leq u \leq 1.$$  

The copula $C_{EV}$ of MEV distribution $G$ enjoys the multiplicative scaling property:

$$C_{EV}(u_1^t, \ldots, u_d^t) = C_{EV}^t(u_1, \ldots, u_d), \ \forall \ (u_1, \ldots, u_d) \in [0, 1]^d, \ t > 0.$$  

This is known as the extreme value (EV) copula.

- If we don’t standardize margins, we can decompose $G$ into the EV copula and univariate EV distributions. We can still establish the scaling property of the EV copula (Joe, 1997, page 173).

- If we are not allowed to standardize margins nor to use copulas, we can establish the operator-scaling property (Meerschaert and Scheffler, 2001). But such a scaling property is not explicit.
Multivariate Scaling ⇒ Pickands Representation

The multivariate scaling property allows us to establish a semi-parametric representation for MEV $G$.

Let $S^{d-1}_+ = \{ a : a = (a_1, \ldots, a_d) \in \mathbb{R}^d, ||a|| = 1 \}$, where $|| \cdot ||$ is a norm defined on $\mathbb{R}^d$. Using the polar coordinates, $G$ and its copula can be expressed as follows:

$$G(x_1, \ldots, x_d) = \exp \left\{ -b \int_{S^{d-1}_+} \bigvee_{1 \leq i \leq d} \left( \frac{a_i}{x_i} \right) Q(da) \right\},$$

$$C_{EV}(u_1, \ldots, u_d) = \exp \left\{ b \int_{S^{d-1}_+} \bigwedge_{1 \leq i \leq d} (a_i \ln u_i) Q(da) \right\},$$

where $b > 0$ and $Q$ is a probability measure defined on $S^{d-1}_+$ such that

$$b \int_{S^{d-1}_+} a_i Q(da) = 1, \ 1 \leq i \leq d.$$
The probability measure $Q$ is known as the spectral or angular measure. Its estimation and asymptotic properties can be found in Resnick (2007).

- Any probability measure $Q$ on $S^{d-1}_+$ can be approximated via discrete probability measures on $S^{d-1}_+$.

- The discretization of $Q$ leads to a rich parametric family of max-stable multivariate Fréchet distributions (min-stable multivariate exponential distributions, including the well-known Marshall-Olkin distribution).

- The discretization of $Q$ leads to a rich parametric family of EV copulas with singularity components (e.g., Lévy-frailty copulas, Mai and Scherer, 2009).
Y = (Y_1, \ldots, Y_d) is said to positively associated if

\[ E[f(Y)g(Y)] \geq Ef(Y)Eg(Y), \quad \forall f, g : \mathbb{R}^d \to \mathbb{R}, \text{ non-decreasing.} \]

The positive association is a strong positive dependence and has been extended to random elements in partially ordered, complete separable metric space (Lindqvist, 1988).

The positive association, equivalent to FKG inequality widely used in statistical physics (Fortuin, Kastelyn, and Ginibre, 1971), is one of several basic inequalities in analyzing concentration phenomena (Ledoux and Talagrand, 1991).

**Theorem (Marshall and Olkin, 1983)**
The MEV distribution G is positively associated.
Multivariate Regular Variation (MRV)

- Assume that df $F \in \text{MDA}(G)$.
- $X = (X_1, \ldots, X_d)$ denote a random vector with distribution $F$ and continuous, univariate margins $F_1, \ldots, F_d$.
- Without loss of generality, margins $F_1, \ldots, F_d$ are identical.

**Definition (Resnick, 1987 and 2007)**

The df $F$ or $X$ is said to be multivariate regularly varying at $\infty$ with intensity measure $\nu$ if there exists a scaling function $b(t) \to \infty$ and a non-zero Radon measure $\nu(\cdot)$ such that as $t \to \infty$,

$$t \mathbb{P} \left( \frac{X}{b(t)} \in B \right) \to \nu(B), \ \forall \ \text{“good” sets } B \subset \mathbb{R}_+^d \setminus \{0\}, \ \text{with } \nu(\partial B) = 0.$$

- $X$ is standard regularly varying if $b(t) = t$. 
Multivariate Regular Variation (MRV)

- Assume that $\text{df } F \in \text{MDA}(G)$.
- $X = (X_1, \ldots, X_d)$ denote a random vector with distribution $F$ and continuous, univariate margins $F_1, \ldots, F_d$.
- Without loss of generality, margins $F_1, \ldots, F_d$ are identical.

**Definition (Resnick, 1987 and 2007)**

The df $F$ or $X$ is said to be multivariate regularly varying at $\infty$ with intensity measure $\nu$ if there exists a scaling function $b(t) \to \infty$ and a non-zero Radon measure $\nu(\cdot)$ such that as $t \to \infty$,

$$t \mathbb{P} \left( \frac{X}{b(t)} \in B \right) \to \nu(B), \ \forall \text{"good" sets } B \subset \mathbb{R}_+^d \setminus \{0\}, \ \text{with } \nu(\partial B) = 0.$$

- $X$ is standard regularly varying if $b(t) = t$. 
Equivalently,

$$\lim_{t \to \infty} \frac{\mathbb{P}(X \in tB)}{\mathbb{P}(X_1 > t)} = \nu(B), \ \forall \text{ “good” sets } B \subset \mathbb{R}^d_+ \setminus \{0\},$$

satisfying that \(\nu(\partial B) = 0\).

The intensity measure \(\nu\) is a Radon measure with scaling property

$$\nu(tB) = t^{-\alpha} \nu(B), \ \forall \text{ “good” sets } B, \text{ bounded away from } 0,$$

where \(\alpha > 0\) is known as the tail index.

\(F_j, 1 \leq j \leq d\), is regularly varying with tail index \(\alpha > 0\).

**Example:** Multivariate t distribution, multivariate Pareto distributions, certain elliptical distributions, ... and many many more if we use copulas and vines.
MDA ⇔ Standard RV (Klüppelberg and Resnick, 2008)

Consider two functions $c(t) = (c_1(t), \ldots, c_d(t))$, and $d(t) = (d_1(t), \ldots, d_d(t))$ such that component-wise

$$\frac{c_i(tx_i) - c_i(t)}{d_i(t)} \to \psi_i(x) \neq 0, \quad \forall x_i > 0, \ t \to \infty.$$ 

1. If $Z = (Z_1, \ldots, Z_d)$ is standard regularly varying, then

$$X = (c_1(Z_1), \ldots, c_d(Z_d)) \in \text{MDA}(G)$$

for some MEV distribution $G$.

2. If $X \in \text{MDA}(G)$ for an MEV distribution $G(x)$, then there exists a component-wise monotone transformation $c(t)$ such that

$$Z = (c_1^-(X_1), \ldots, c_d^-(X_d))$$

is standard regularly varying.
Example: Heavy-Tail Case

- \((Z_1, \ldots, Z_d)\) is non-negative.
- For any \(\alpha > 0\), \(c_i(z) := z^{1/\alpha}\) is increasing on \(\mathbb{R}_+\), \(1 \leq i \leq d\).
- \((X_1, \cdots, X_d) = (c_1(Z_1), \ldots, c_d(Z))\) has the df \(F(x_1, \ldots, x_d)\).

Theorem (de Haan & Resnick, 1977; Marshall & Olkin, 1983)

The following three statements are equivalent:

1. \((Z_1, \ldots, Z_d)\) is standard regular varying.
2. \((X_1, \cdots, X_d)\) is MRV with intensity measure \(\nu(\cdot)\).
3. \(F \in \text{MDA}(G)\), where \(G(x) = \exp\{-\nu([0, x]^c)\}\), \(x \in \mathbb{R}_+^d\), is a \(d\)-dimensional distribution with Fréchet margins

\[
G_i(x_i) = \exp\{-x_i^{-\alpha}\}, \quad 1 \leq i \leq d.
\]

Note that \(\nu(tB) = t^{-\alpha}\nu(B)\) for any \(t > 0\).
The Copula Approach

- $X_n = (X_{1,n}, \ldots, X_{d,n}), \ n = 1, 2, \ldots$, are iid with df $F(x_1, \ldots, x_d)$ that has continuous margins $F_1, \ldots, F_d$.
- The copula of $F$:

$$C(u_1, \ldots, u_d) := F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_1)), \ (u_1, \ldots, u_d) \in [0, 1]^d.$$ 

Figure: Two “squares” are topologically equivalent. Since limiting properties are topological, the copula method should be equivalent to the standard RV method.
Exponent and Tail Dependence Functions

\((U_1, \ldots, U_d) := (F_1(X_{n,1}), \ldots, F_d(X_{n,d}))\).

Define the exponent function: for all \((w_1, \ldots, w_d) \in \mathbb{R}_+^d \setminus \{0\},

\[ a(w_1, \ldots, w_d) := \lim_{u \downarrow 0} \frac{\mathbb{P}(\bigcup_{j=1}^d \{U_j > 1 - uw_j\})}{u}, \]

with scaling \(a(tw_1, \ldots, tw_d) = t a(w_1, \ldots, w_d)\) for \(t > 0\).

Define the tail dependence function:

\[ b(w_1, \ldots, w_d) := \lim_{u \downarrow 0} \frac{\mathbb{P}(\bigcap_{j=1}^d \{U_j > 1 - uw_j\})}{u}, \]

with scaling \(b(tw_1, \ldots, tw_d) = tb(w_1, \ldots, w_d)\) for \(t > 0\).

The existence of the exponent function ensures the existence of the exponent and tail dependence functions of all multivariate margins of \(F\).
Various tail dependence parameters are associated with the function $b(\cdot)$; for example, $b(1, \ldots, 1)$ is known as the tail dependence coefficient.

Due to the scaling property, $b(\cdot) = 0$ iff $b(1, \ldots, 1) = 0$, and in this case, the copula $C$ is said to be (upper) tail independent.

The notion of the exponent function $a(\cdot)$ can be traced back to Gumbel (1960) and Pickands (1981). It describes the full extremal dependence structure of $C$, including the dependence of any multivariate marginal tails.

The tail dependence function $b(\cdot)$ is studied in Jaworski (2004, 2006), Klüppelberg, Kuhn & Peng (2008), Nikoloulopoulos, Joe & Li (2009). Note that $b(w_1, \ldots, w_d)$ does not necessarily cover the extremal dependence structure of a multivariate margin.
Example: Archimedean Copulas

- Let $C(u; \phi) = \phi(\sum_{i=1}^{d} \phi^{-1}(u_i))$ be an Archimedean copula where the generator $\phi^{-1}$ is regularly varying at 1 with tail index $\beta > 1$.

  \[
a(w_1, \ldots, w_d) = \left(\sum_{j=1}^{d} w_j^\beta\right)^{1/\beta}, \quad (w_1, \ldots, w_d) \in \mathbb{R}_+^d \setminus \{0\}.
\]

  (Genest & Rivest, 1989; Barbe, Fougères & Genest, 2006)

- Let $C$ have the survival copula $\hat{C}(u; \phi) = \phi(\sum_{i=1}^{d} \phi^{-1}(u_i))$ with strict generator $\phi^{-1}$, where $\phi$ is regularly varying at $\infty$ with tail index $\theta > 0$.

  \[
b(w_1, \ldots, w_d) = \left(\sum_{j=1}^{d} w_j^{-1/\theta}\right)^{-\theta}, \quad (w_1, \ldots, w_d) \in \mathbb{R}_+^d \setminus \{0\}.
\]

  (Charpentier & Segers, 2009)
Tail Dependence Function “=” Intensity Measure

Theorem (Li and Sun, 2009)

Let $X = (X_1, \ldots, X_d)$ be a random vector with df $F$ and copula $C$.

1. If $F$ is MRV with intensity measure $\nu$ satisfying the scaling property that $\nu(tB) = t^{-\alpha} \nu(B)$, then for $w = (w_1, \ldots, w_d)$,

   $$a(w) = \nu \left( \left( \prod_{i=1}^{d} [0, w_i^{-1/\alpha}] \right)^c \right), \quad b(w) = \nu \left( \prod_{i=1}^{d} [w_i^{-1/\alpha}, \infty] \right).$$

2. If the exponent function exists and marginal dfs $F_1, \ldots, F_d$ are (univariate) regularly varying, then

   $$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$ is MRV.

- The Radon measure generated by the exponent function $a(\cdot)$ is a rescaled version of the intensity measure $\nu(\cdot)$. 
Ingredients for High-Dimensional MRV

1. Univariate dfs $F_1, \ldots, F_d$ are regularly varying with tail index $\alpha_1, \ldots, \alpha_d$ respectively.
2. $C$ is a copula with (upper) tail dependence limits.

Then $F(x_1, \ldots, x_d) := C(F_1(x_1), \ldots, F_d(x_d))$ is MRV.

Remark:

- Tail dependence properties for vine copulas are obtained using recursive constructions according to underlying graph structures (Joe, Li and Nikoloulopoulos, 2010).
Let $C$ be a copula with (upper) exponent function $a(\cdot)$.

- The (upper) extreme-value copula of $C$:
  \[ C_{EV}(u_1, \ldots, u_d) = \exp\{ -a(-\log u_1, \ldots, -\log u_d) \}, \]

- where $a(w_1, \ldots, w_d) = \int_{S^{d-1}} \max_{1 \leq i \leq d} \{ a_i w_i \} \mathbb{U}(da)$, or
  \[ b(w_1, \ldots, w_d) = \int_{S^{d-1}} \min_{1 \leq i \leq d} \{ a_i w_i \} \mathbb{U}(da). \]

- The spectral measure $\mathbb{U}(\cdot)$ defined on $S^{d-1}$ is a finite measure.

- In contrast to the Pickands representation for intensity measures, the spectral measure $\mathbb{U}(\cdot)$ does not depend on margins.
Scaling Property, Again

Consider the $d$-dimensional copula $C$ of a random vector $(U_1, \ldots, U_d)$ with standard uniform margins.

- Recall that
  
  \[ a(tw) = ta(w), \quad b(tw) = tb(w), \quad \forall \ w \in \mathbb{R}_+^d, \ t > 0. \]

- The well-known Euler’s homogeneous theorem implies that
  
  \[ a(w) = \sum_{j=1}^{d} \frac{\partial a}{\partial w_j} w_j, \quad b(w) = \sum_{j=1}^{d} \frac{\partial b}{\partial w_j} w_j, \quad \forall w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d, \]

- The partial derivatives can be interpreted as conditional limiting distributions of the underlying copula $C$. For example,
  \[ \frac{\partial b}{\partial w_j} = \lim_{u \downarrow 0} \mathbb{P}(U_i > 1 - uw_i, \forall i \neq j \mid U_j = 1 - uw_j). \]
Let $X = (X_1, \ldots, X_d)$ have the t distribution $T_{d, \nu, \Sigma}$ with $\nu$ degrees of freedom and dispersion matrix $\Sigma$.

The margins $F_i = T_\nu$ for all $1 \leq i \leq d$, where $T_\nu$ is the t distribution function with $\nu$ degrees of freedom, and that $\Sigma = (\rho_{ij})$ satisfies $\rho_{ii} = 1$ for all $1 \leq i \leq d$.

The t-copula is defined as

$$C_t(u_1, \ldots, u_d) = P(T_\nu(X_1) \leq u_1, \ldots, T_\nu(X_d) \leq u_d).$$

Derive the extreme value copula $C_{t-EV}$. 
Partial Correlation Matrix

For any $i \neq j, k \neq j$, let

$$
\rho_{i,k;j} = \frac{\rho_{ik} - \rho_{ij}\rho_{kj}}{\sqrt{1 - \rho_{ij}^2} \sqrt{1 - \rho_{kj}^2}}
$$

denote the partial correlations, and

$$
R_j = \begin{pmatrix}
1 & \cdots & \rho_{1,j-1;j} & \rho_{1,j+1;j} & \cdots & \rho_{1,d;j} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\rho_{1,j-1;j} & \cdots & 1 & \rho_{j-1,j+1;j} & \cdots & \rho_{j-1,d;j} \\
\rho_{1,j+1;j} & \cdots & \rho_{j-1,j+1;j} & 1 & \cdots & \rho_{j+1,d;j} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{1,d;j} & \cdots & \rho_{j-1,d;j} & \rho_{j+1,d;j} & \cdots & 1
\end{pmatrix}.
$$
1. The tail dependence function of $C$ is given by

$$b(w) = \sum_{j=1}^{d} w_j T_{d-1, \nu + 1, R_j} \left( \frac{\sqrt{\nu + 1}}{\sqrt{1 - \rho_{ij}^2}} \left[ \left( \frac{w_i}{w_j} \right)^{-1/\nu} - \rho_{ij} \right], i \neq j \right),$$

for all $w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d$.

2. The $t$-EV copula is given by

$$C_{t-EV}(u_1, \ldots, u_d) = \exp\{-a(w_1, \ldots, w_d)\}, \quad w_j = -\log u_j, \quad j = 1, \ldots, d,$$

with exponent

$$a(w) = \sum_{j=1}^{d} w_j T_{d-1, \nu + 1, R_j} \left( \frac{\sqrt{\nu + 1}}{\sqrt{1 - \rho_{ij}^2}} \left[ \left( \frac{w_i}{w_j} \right)^{-1/\nu} - \rho_{ij} \right], i \neq j \right).$$
Two Limiting Distributions

- Under some scaling conditions, $C_{t-EV}(\cdot)$ converges weakly to the Hüsler-Reiss copula as $\nu \to \infty$.
- As $\nu \to 0$, $C_{t-EV}(\cdot)$ converges weakly to a Marshall-Olkin distribution with some linear constraints.
The Tail Density Approach

- The notion of tail density is local and geometric (Balkema and Embrechts, 2007).
- Asymptotic analysis of tail risk measures often involves directly integral functionals of tail densities (Joe and Li, 2011).

Geometric Risk Analysis via Tail Densities

- A data cloud is a realization of a random set (e.g., random vectors, spatial point processes, high-dimensional Brownian motion, ...).
- “Central limit theorems” are concerned with clustering behaviors at the “center” of a data cloud.
- “Extreme value theorems” are concerned with dispersive stability patterns at the boundaries of a data cloud.
- Tail risk lives along the boundaries; it could be on the northeast boundary (max domain of attraction) or could be on the southwest boundary (min domain of attraction), or could be in other parts of boundaries.
MRV Tail Densities

- $X = (X_1, \ldots, X_d) \sim F$ that has identical continuous margins $F_1, \ldots, F_d$.
- $\overline{F}_1(t) := 1 - F_1(t)$ denotes the survival function of $F_1$.

**Theorem (de Haan and Resnick, 1987)**

Assume the density $f$ of $F$ exists. If \( \frac{f(tx)}{t^{-d} \overline{F}_1(t)} \rightarrow \lambda(x) > 0 \) on $\mathbb{R}_+^d \setminus \{0\}$ and uniformly on $\{x > 0 \mid \|x\| = 1\}$, as $t \rightarrow \infty$, then

$$
\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{\overline{F}_1(t)} = \int_{[0,x]^c} \lambda(y) dy =: \mu([0, x]^c), \ x \in \mathbb{R}_+^d \setminus \{0\},
$$

that is, $F$ is MRV with intensity measure $\mu(\cdot)$. 

Haijun Li
High-Dimensional Extremes and Copulas
CIAS-CUFE, Beijing, January 3, 2014
Perspective Shot of MTCJ Copula Density

Consider a bivariate MTCJ copula \( C(x, y) = (x^{-\theta} + y^{-\theta} - 1)^{-1/\theta} \) with density \( c(x, y) = (1 + \theta)(xy)^{-\theta-1}(x^{-\theta} + y^{-\theta} - 1)^{-2-1/\theta}, \theta > 0 \). Approaching the origin along the ray passing though \((w_1, w_2), w_1 > 0 \) and \( w_2 > 0 \), we have, as \( u \to 0 \),

\[
c(uw_1, uw_2) \approx u^{-1}[(1 + \theta)(w_1 w_2)^{-\theta-1}(w_1^{-\theta} + w_2^{-\theta})^{-2-1/\theta}]
\]
Idea: Let \((U_1, \ldots, U_d) \overset{d}{\sim} C\). When \(u\) is sufficiently small,

tail density \(\approx \frac{\mathbb{P}(1 - uw_i \leq U_i \leq 1 - u(w_i - dw_i), 1 \leq i \leq d)}{u^\kappa dw_1 \cdots dw_d }
\approx u^{d-\kappa} \frac{\mathbb{P}(1 - uw_i \leq U_i \leq 1 - uw_i + d(uw_i), 1 \leq i \leq d)}{d(uw_1) \cdots d(uw_d)}
\)

In what follows, \(\kappa = 1\).
Assume that all the necessary regularity conditions hold.

**Definition**

The upper and lower tail densities of $C$ are defined as

$$
\lambda^U(w) := \lim_{u \to 0} \frac{D_w \bar{C}(1 - uw_1, \ldots, 1 - uw_d)}{u},
$$

$$
\lambda^L(w) := \lim_{u \to 0} \frac{D_w C(uw_1, \ldots, uw_d)}{u}, \forall \ w = (w_1, \ldots, w_d) \in \mathbb{R}^d_+
$$

where $D_w$ denotes the $d$-order partial differentiation operator and $\bar{C}$ is the survival function of $C$.

- $C$ is upper (lower) tail independent if and only if $\lambda^U = 0 \ (\lambda^L = 0)$.  


Properties

Let $c$ denote the density of $C$, then

$$
\lambda^U(w) = \lim_{u \to 0} u^{d-1} c(1 - uw_i, 1 \leq i \leq d),
$$

$$
\lambda^L(w) = \lim_{u \to 0} u^{d-1} c(uw_i, 1 \leq i \leq d), \ \forall \ w = (w_1, \ldots, w_d) \in \mathbb{R}^d_+
$$

For a $d$-dimensional copula $C$, the tail densities are homogeneous of order $1 - d$. That is, $\lambda^U(tw) = t^{1-d} \lambda^U(w)$ for any $w \in \mathbb{R}^d_+$ and $t > 0$.

The Euler’s homogeneous theorem implies that the tail densities are directionally decreasing and convex, and go down to zero at infinity.
Relation with Tail Dependence Functions

Recall that exponent and tail dependence functions:

\[ a(w) := \lim_{u \to 0} \frac{P(U_i > 1 - uw_i, \text{ for some } i)}{u}, \quad \forall \ w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d. \]

\[ b(w) := \lim_{u \to 0} \frac{P(U_i > 1 - uw_i, \text{ for all } i)}{u}, \quad \forall \ w = (w_1, \ldots, w_d) \in \mathbb{R}_+^d. \]

**Theorem**

Recall that \( D_w \) is the \( d \)-order partial differentiation operator.

\[ \lambda^U(w) = D_w b(w) = (-1)^{d-1} D_w a(w) \]

for all \( w \in \mathbb{R}_+^d. \)
Relation with Tail Density of MRV

Recall that \( \lambda(\cdot) \) denotes the tail density of an MRV df \( F \) such that

\[
\lim_{t \to \infty} \frac{1 - F(tx)}{F_1(t)} = \int_{[0,x]^c} \lambda(y)dy, \quad x \in \mathbb{R}^d_+.
\]

Once again, \( \lambda^U \) denotes the upper tail density of the copula \( C \) of \( F \).

**Theorem (Li and Wu, 2013)**

Let \( \alpha \) denote the tail index of \( F \), then

\[
\lambda(w_1, \ldots, w_d) = \alpha^d (w_1 \cdots w_d)^{-\alpha - 1} \lambda^U(w_1^{-\alpha}, \ldots, w_d^{-\alpha})
\]

\[
= |J(w_1^{-\alpha}, \ldots, w_d^{-\alpha})| \lambda^U(w_1^{-\alpha}, \ldots, w_d^{-\alpha}),
\]

where \( |J(w_1^{-\alpha}, \ldots, w_d^{-\alpha})| \) is the Jacobian determinant of the homeomorphism \( y_i = w_i^{-\alpha}, 1 \leq i \leq d \).
Archimedean Tail Densities

Let \( C(u; \phi) = \phi(\sum_{i=1}^{d} \phi^{-1}(u_i)) \) be an Archimedean copula where the Laplace transform \( \phi \).

**Lower Tail Density**

If \( \phi \) is regularly varying at \( \infty \) with tail index \( \theta > 0 \), then

\[
\lambda^L(w) = \prod_{i=1}^{d} \left( 1 + \frac{i-1}{\theta} \right) \left( \prod_{i=1}^{d} w_i \right)^{-1-1/\theta} \left( \sum_{i=1}^{d} w_i^{-1/\theta} \right)^{-\theta-d}.
\]

**Upper Tail Density**

If \( \phi^{-1} \) is regularly varying at \( 1 \) with tail index \( \beta > 1 \), then

\[
\lambda^U(w) = \prod_{i=1}^{d} ((i-1)\beta - 1) \left( \prod_{i=1}^{d} w_i \right)^{\beta-1} \left( \sum_{i=1}^{d} w_i^{\beta} \right)^{-d-1/\beta}.
\]
Lower Archimedean Tail Density

Figure 2: graphs for lower tail case

Its graph is shown below ($\theta = 1$).

**Theorem 2.5.** Let $X = (X_1, \ldots, X_d)$ be a non-negative MRV random vector with intensity measure $\mu$, copula $C$ and continuous margins $F_1, \ldots, F_d$. If the margins are tail equivalent (i.e. $\bar{F}_i(t)/\bar{F}_i(t) \to 1$ as $t \to \infty$ for any $i \neq j$) with heavy-tail index $\beta > 0$, then the upper tail dependence function $\lambda^*(\cdot)$ exists and

1. $\lambda^*(w) = \frac{\partial}{\partial w} \mu([w, \infty))^\mu([0, 1] \times \bar{R}^{d-1})$

2. $\frac{\partial}{\partial w_1} \cdots \frac{\partial}{\partial w_d} \mu([w, \infty)^\mu([0, 1] \times \bar{R}^{d-1}) = (-1)^d a^*(1) \cdot \beta^d \cdot \prod_{i=1}^d w_i^{-\beta} \lambda^*(w)$

3 Tail approximation via tail density

In this section, we derive the tail asymptotics for $\operatorname{VaR}_p(||X||)$, as $p \to 1$, for any fixed norm $||\cdot||$ on $\mathbb{R}^d$. The results discussed in [6] can be obtained by taking the $l_1$-norm and the tail dependence function of Archimedean copulas.
Upper Archimedean Tail Density

Figure 1: Graphs for upper tail case

4. $\lambda(\infty,\infty) = 0.$

The graph of this bivariate Gumbel copula is as follows ($\delta = 2$).

Example 2.4. Consider a bivariate Clayton copula $C(u,v; \theta) = (u - \theta + v - \theta - 1)^{-1} \theta, \theta > 0$.

We know that Clayton copula only has lower tail dependence, no upper tail dependence. Its lower tail density function is given by the following steps:

The bivariate copula density function with parameter $\theta$ is

$\lambda(u) = \lim_{u \to 0} u \cdot c(u_i, 1 \leq i \leq 2)$

$= \lim_{u \to 0} u \cdot (1 + \theta)u^{\theta - 1}(w_1 + w_2 - u \theta - 1)\theta - 1\theta^{-2}(w_1w_2)^{-\theta - 1}, \theta > 0.$

The expression of $\lambda(w)$ implies some properties:

1. $\lambda(0,w_2) = d(w_1,0) = \infty$
2. $\lambda(0,0) = \infty$
3. $\lambda(\infty,w_2) = d(w_1,\infty) = 0$
4. $\lambda(\infty,\infty) = 0.$
t Tail Density

Consider a $d$-dimensional symmetric $t$ distribution with cdf

$$f(x; \nu, \Sigma) = \frac{\Gamma\left(\frac{\nu + d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu \pi)^{d/2}} |\Sigma|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu} (x^T \Sigma^{-1} x)\right]^{-\frac{\nu + d}{2}}, \ x \in \mathbb{R}^d$$

where $\nu > 0$ is the degree of freedom, and $\Sigma$ is a $d \times d$ dispersion matrix.

Tail Density of t-Copula

$$\lambda^U(w) = \lambda^L(w) = \pi^{-\frac{d-1}{2}} |\Sigma|^{-\frac{1}{2}} \nu^{1-d} \frac{\Gamma\left(\frac{\nu + d}{2}\right)}{\Gamma\left(\frac{\nu + 1}{2}\right)} \left[(\frac{1}{\nu})^T \Sigma^{-1} (\frac{1}{\nu})\right]^{-\frac{\nu + d}{2}} \prod_{i=1}^{d} w_i^{\nu + 1}. $$

for any $w \in \mathbb{R}_+^d$. 
Rewrite It as a Norm-Based Tail Density

- Define a norm \( \| w \| := (w^T \Sigma^{-1} w)^{1/2}, \ w \in \mathbb{R}_+^d \).
- Let \( |J(w_1^{-1/\nu}, \ldots, w_d^{-1/\nu})| \) be the Jacobian determinant of topologically invariant transform \( y_i = w_i^{-1/\nu}, 1 \leq i \leq d \).
- Rewrite the upper tail density of a t copula:

\[
\lambda^U(w_1, \ldots, w_d) = \zeta \prod_{i=1}^{d} w_i^{-\nu+1/\nu} \left( (w^{-1/\nu})^T \Sigma^{-1} w^{-1/\nu} \right)^{-\nu+d/2} = \zeta \nu^d |J(w_1^{-1/\nu}, \ldots, w_d^{-1/\nu})| \times \| w^{-1/\nu} \|^{-\nu-d}
\]

- For the symmetric t distribution itself,

\[
\lambda(w_1, \ldots, w_d) = \zeta \nu^d \| w \|^{-\nu-d}
\]

(de Haan and Resnick, 1987).
Vine Copula $C$ of $(U_1, \ldots, U_d)$

For any $u = (u_1, \ldots, u_d) \in [0, 1]^d$, define

$$u_S = (u_i, i \in S), \ \forall \ S \subseteq \{1, \ldots, d\}.$$ 

Ingredients for Vines

- $\{c_{i,j}, 1 \leq i < j \leq d\} = A$ set of the densities of bivariate linking copulas.
- For any $S \subseteq \{1, \ldots, d\}$, define the $S$-marginal density $c_S := c_S(u_S)$.
- Define the conditional distribution of $U_k$ given $U_S = u_S$

$$C_{k|S} := C_{k|S}(u_k|u_S), \ k \notin S.$$
The density \( c_{\{1,...,d\}} \) of \( C \) is constructed recursively as follows.

**D-Vine Construction (Bedford and Cooke, 2001 and 2002)**

1. **Baseline:** For any \( 1 \leq i \leq d - 1 \), the density of the \( \{i, i + 1\} \) margin is \( c_{i,i+1} \).
2. **Recursion:**

\[
\frac{c_{\{1,...,d\}}}{c_{\{2,...,d-1\}}} = c_{1,d} \left( \frac{c_{\{1,\ldots,d-1\}}}{c_{\{2,\ldots,d-1\}}} \cdot \frac{c_{\{2,\ldots,d\}}}{c_{\{2,\ldots,d-1\}}} \right)
\]
The Recursion of D-Vine Lower Tail Densities

- For any $S \subseteq \{1, \ldots, d\}$, $\lambda^L_S(w_S)$ = lower tail density of the $S$-margin $C_S$.
- For any $S \subseteq \{1, \ldots, d\}$ and $k \notin S$, define the lower tail conditional distribution of $U_k$ given $U_S = u_S$:

$$t_{k|S}(w_k|w_S) := \lim_{u \to 0} C_{k|S}(uw_k|uw_S).$$

**Theorem**

Fix $S = \{2, \ldots, d - 1\}$. If all the bivariate baseline linking copulas have lower tail dependence, then

$$\frac{\lambda^L(w)}{\lambda^L(S)(w_S)} = c_{1,d} \left( t_{1|S}(w_1|w_S), t_{d|S}(w_d|w_S) \right) \frac{\lambda^L_{\{1\} \cup S}(w)}{\lambda^L_S(w_S)} \frac{\lambda^L_{\{d\} \cup S}(w)}{\lambda^L_S(w_S)}.$$
Examples

The 3-dimensional D-vine:

\[ \lambda^L(w_1, w_2, w_3) = \lambda^L_{12}(w_1, w_2) \cdot \lambda^L_{23}(w_2, w_3) \cdot c_{13}(t_{1|2}(w_1|w_2), t_{3|2}(w_3|w_2)). \]

The 4-dimensional D-vine:

\[ \lambda^L(w_1, w_2, w_3, w_4) = \lambda^L_{12}(w_1, w_2) \cdot \lambda^L_{23}(w_2, w_3) \cdot \lambda^L_{34}(w_3, w_4) \]
\[ \cdot c_{13}(t_{1|2}(w_1|w_2), t_{3|2}(w_3|w_2)) \cdot c_{24}(t_{2|3}(w_2|w_3), t_{4|3}(w_4|w_3)) \]
\[ \cdot c_{14}(t_{1|23}(w_1|w_2, w_3), t_{4|23}(w_4|w_2, w_3)). \]
Cautionary Remark

What happens if some or all bivariate baseline linking copulas of a D-vine $C$ are tail independent? Joe et al (2010) has a partial answer:

- $C$ must be tail independent, e.g.,

$$C(uw_i, 1 \leq i \leq d) \sim u^\kappa h(w), \text{ as } u \to 0, \kappa > 1.$$  

- Some margins of $C$ can still be tail dependent.

But can we quantify the order of scaling, e.g., $\kappa$, in the case of tail independence of a vine copula?
Seeking Recursions of RV Near Data Boundaries

Consider the 3-dimensional D-vine

\[
c(uw_1, uw_2, uw_3) = c_{12}(uw_1, uw_2) \cdot c_{23}(uw_2, uw_3) \\
\cdot c_{13}(C_{1|2}(uw_1|uw_2), C_{3|2}(uw_3|uw_2)), \text{ when } u \text{ is small.}
\]

As functions of \( u \), the regular variations of constructs \( c_{12}, c_{23}, C_{1|2}, C_{3|2}, \) and \( c_{13} \) should yield the regular varying property of \( c \).

**Example:** If the baseline linking copulas \( C_{12} \) and \( C_{23} \) are Morgenstern copulas, and the linking copula \( C_{13} \) is a t copula, then \( c_{12}(uw_1, uw_2) \) and \( c_{23}(uw_2, uw_3) \) are asymptotically constant as \( u \to 0 \) and

\[
C_{1|2}(uw_1|uw_2) \sim uh_{1|2}(w_1, w_2), C_{3|2}(uw_3|uw_2) \sim uh_{3|2}(w_2, w_3), \text{ as } u \to 0.
\]

Thus, as \( u \to 0 \),

\[
c(uw_1, uw_2, uw_3) \sim u^{-1} h(w_1, w_2, w_3).
\]

That is, \( \kappa = 2 \).
In analyzing extremes, the goal is to develop accurate prediction rather than to provide good interpretability.

In analyzing high-dimensional data clouds, margins are of multiple scales and it is entirely possible that only some multivariate margins converge under possibly different surface/hyperplane thresholdings (Balkema and Embrechts, 2007).

“Occam’s Razor” is often of little use in high-dimensional dependence analysis. Simple and interpretable models do not make the most accurate predictions. Sophisticated and flexible models are called for.
If all a man has is a hammer, ...


- An old saying: “If all a man has is a hammer, then every problem looks like a nail.”
- But the trouble for statisticians lately is that some of the new problems have stopped looking like nails.
- Complex big data pose significant challenges.
References


References


