Multivariate Distorted Distributions

Jorge Navarro[1]
Universidad de Murcia, Murcia, Spain.

[1]Supported by Ministerio de Ciencia e Innovación of Spain under grant PID2019-108079GB-C22/AEI/10.13039/501100011033.

References

The conference is based on the following references:

- Navarro, Calì, Longobardi and Durante (2022). Distortion representations of multivariate distributions. Statistical Methods & Applications 31, 925–954.
  DOI: 10.1007/s10260-021-00613-2.

References

The conference is based on the following references:

- Navarro, Calì, Longobardi and Durante (2022). Distortion representations of multivariate distributions. Statistical Methods & Applications 31, 925–954.
  DOI: 10.1007/s10260-021-00613-2.
- Navarro (2022). Prediction of record values by using quantile regression curves and distortion functions. Metrika 85, 675–706.
  DOI: 10.1007/s00184-021-00847-w.

References

The conference is based on the following references:

- Navarro, Calì, Longobardi and Durante (2022). Distortion representations of multivariate distributions. Statistical Methods & Applications 31, 925–954.
  DOI: 10.1007/s10260-021-00613-2.
- Navarro (2022). Prediction of record values by using quantile regression curves and distortion functions. Metrika 85, 675–706.
  DOI: 10.1007/s00184-021-00847-w.
- Navarro and Buono (2022). Predicting future failure times by using quantile regression. To appear in Metrika. Published online first Sept 2022.
  DOI: https://doi.org/10.1007/s00184-022-00884-z

Outline

Distorted distributions

- Definitions
- Main properties
- Examples

Prediction of ordered paired data

- Representations
- Exact QR curves
- Estimated QR curves

Distorted distributions

Univariate distorted distributions

- The distorted distributions were introduced by Wang (1996) and Yaari (1987) in the context of theory of choice under risk.

Univariate distorted distributions

- The distorted distributions were introduced by Wang (1996) and Yaari (1987) in the context of theory of choice under risk.
- The purpose was to allow a "distortion" (a change) of the initial (or past) risk distribution function.

Univariate distorted distributions

- The distorted distributions were introduced by Wang (1996) and Yaari (1987) in the context of theory of choice under risk.
- The purpose was to allow a "distortion" (a change) of the initial (or past) risk distribution function.

Definition

The **distorted distribution** (DD) associated to a distribution function (DF) F and to an increasing continuous *distortion function* $q : [0,1] \to [0,1]$ such that $q(0) = 0$ and $q(1) = 1$, is given by

$$F_q(t) = q(F(t)), \text{ for all } t \in \mathbb{R}. \tag{1.1}$$

Properties

- If q is a distortion function, then F_q is a proper distribution function for all distribution functions F.

Properties

- If q is a distortion function, then F_q is a proper distribution function for all distribution functions F.
- From (1.1), $\bar{F} = 1 - F$ and $\bar{F}_q = 1 - F_q$ satisfy

$$\bar{F}_q(t) = \bar{q}(\bar{F}(t)), \text{ for all } t \in \mathbb{R}, \tag{1.2}$$

where $\bar{q}(u) := 1 - q(1 - u)$ is called the *dual distortion function*.

Properties

- If q is a distortion function, then F_q is a proper distribution function for all distribution functions F.
- From (1.1), $\bar{F} = 1 - F$ and $\bar{F}_q = 1 - F_q$ satisfy

$$\bar{F}_q(t) = \bar{q}(\bar{F}(t)), \text{ for all } t \in \mathbb{R}, \tag{1.2}$$

  where $\bar{q}(u) := 1 - q(1-u)$ is called the *dual distortion function*.
- (1.1) and (1.2) are equivalent.

Examples of univariate distorted distributions

- Proportional Hazard Rate (PHR) Cox model $\bar{F}_\theta(t) = \bar{F}^\theta(t)$, where $\bar{q}(u) = u^\theta$ and $\theta > 0$.

Examples of univariate distorted distributions

- Proportional Hazard Rate (PHR) Cox model $\bar{F}_\theta(t) = \bar{F}^\theta(t)$, where $\bar{q}(u) = u^\theta$ and $\theta > 0$.
- Proportional Reversed Hazard Rate (PRHR) model $F_\theta(t) = F^\theta(t)$, where $q(u) = u^\theta$ and $\theta > 0$.

Examples of univariate distorted distributions

- Proportional Hazard Rate (PHR) Cox model $\bar{F}_\theta(t) = \bar{F}^\theta(t)$, where $\bar{q}(u) = u^\theta$ and $\theta > 0$.
- Proportional Reversed Hazard Rate (PRHR) model $F_\theta(t) = F^\theta(t)$, where $q(u) = u^\theta$ and $\theta > 0$.
- Order statistics $X_{1:n}, \dots X_{n:n}$ from IID $X_1, \dots, X_n$:

$$\bar{F}_{i:n}(t) = \sum_{j=0}^{i-1} \binom{n}{j} F^j(t)\bar{F}^{n-j}(t) = \bar{q}_{i:n}(\bar{F}(t)),$$

where $\bar{q}_{i:n}(u) = \sum_{j=0}^{i-1} \binom{n}{j}(1-u)^j u^{n-j}$.

Examples of univariate distorted distributions

- Proportional Hazard Rate (PHR) Cox model $\bar{F}_\theta(t) = \bar{F}^\theta(t)$, where $\bar{q}(u) = u^\theta$ and $\theta > 0$.
- Proportional Reversed Hazard Rate (PRHR) model $F_\theta(t) = F^\theta(t)$, where $q(u) = u^\theta$ and $\theta > 0$.
- Order statistics $X_{1:n}, \dots X_{n:n}$ from IID $X_1, \dots, X_n$:

$$\bar{F}_{i:n}(t) = \sum_{j=0}^{i-1} \binom{n}{j} F^j(t) \bar{F}^{n-j}(t) = \bar{q}_{i:n}(\bar{F}(t)),$$

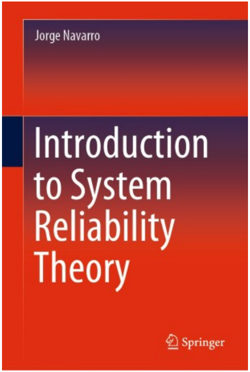
  where $\bar{q}_{i:n}(u) = \sum_{j=0}^{i-1} \binom{n}{j} (1-u)^j u^{n-j}$.
- Coherent system lifetimes T:

$$\bar{F}_T(t) = \bar{Q}(\bar{F}_1(t), \dots, \bar{F}_n(t)), \tag{1.3}$$

  where $\bar{Q} : [0,1]^n \to [0,1]$ is a generalized distortion function, see e.g. Navarro (2022a).

Navarro (2022a)

- This is my new book on System Reliability Theory.



Multivariate distorted distributions

- $(X_1, \ldots, X_n)$ random vector over $(\Omega, \mathcal{S}, \Pr)$.

Multivariate distorted distributions

- $(X_1, \ldots, X_n)$ random vector over $(\Omega, \mathcal{S}, \Pr)$.
- Joint distribution function (DF)

$$\mathrm{F}(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Multivariate distorted distributions

- $(X_1, \dots, X_n)$ random vector over $(\Omega, \mathcal{S}, \Pr)$.
- Joint distribution function (DF)

$$\mathsf{F}(x_1, \dots, x_n) = \Pr(X_1 \le x_1, \dots, X_n \le x_n).$$

- Copula representation

$$\mathsf{F}(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)),$$

where $F_1, \dots, F_n$ are the marginals, $F_i(x_i) = \Pr(X_i \le x_i)$.

Multivariate distorted distributions

- $(X_1, \ldots, X_n)$ random vector over $(\Omega, \mathcal{S}, \Pr)$.
- Joint distribution function (DF)

$$\mathrm{F}(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

- Copula representation

$$\mathrm{F}(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)),$$

where $F_1, \ldots, F_n$ are the marginals, $F_i(x_i) = \Pr(X_i \leq x_i)$.

- A similar representation holds for the joint survival function

$$\overline{\mathrm{F}}(x_1, \ldots, x_n) = \Pr(X_1 > x_1, \ldots, X_n > x_n) = \widehat{C}(\bar{F}_1(x_1), \ldots, \bar{F}_n(x_n)).$$

Definition

Definition (Navarro, Calì, Longobardi and Durante (2022))

A multivariate distribution function F is said to be a *multivariate distorted distribution* (MDD) of the univariate distribution functions $G_1, \ldots, G_n$ if there exists a *distortion* function D such that

$$\mathrm{F}(x_1, \ldots, x_n) = D(G_1(x_1), \ldots, G_n(x_n)), \ \forall x_1, \ldots, x_n \in \mathbb{R}. \quad (1.4)$$

We write $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$, when F is a MDD of $G_1, \ldots, G_n$.

Definition

Definition

A continuous function $D : [0,1]^n \to [0,1]$ is called *(n–dimensional) distortion function* (shortly written as $D \in \mathcal{D}_n$) if:

(i) $D(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_n) = 0$ for all $u_1, \dots, u_n \in [0,1]$.

(ii) $D(1, \dots, 1) = 1$.

(iii) D is n–increasing, i.e. for all $\mathrm{x} = (x_1, \dots, x_n)$ and $\mathrm{y} = (y_1, \dots, y_n)$ with $x_i \le y_i$, it holds $\triangle_{\mathrm{x}}^{\mathrm{y}} D \ge 0$, where

$$\triangle_{(x_1,\dots,x_n)}^{(y_1,\dots,y_n)} D := \sum_{z_i \in \{x_i, y_i\}} (-1)^{1(z_1,\dots,z_n)} D(z_1, \dots, z_n),$$

with $1(z_1, \dots, z_n) = \sum_{i=1}^{n} 1(z_i = x_i)$ and $1(A) = 1$ (respectively, 0) if A is true (respectively, false).

Main properties

- As in Sklar's theorem for copulas, the MDD representation is unique for fixed continuous DF $G_1, \dots, G_n$.

Main properties

- As in Sklar's theorem for copulas, the MDD representation is unique for fixed continuous DF $G_1, \dots, G_n$.
- If $D \in \mathcal{D}_n$, then

$$D(G_1(x_1), \dots, G_n(x_n))$$

is a multivariate distribution function for all DF $G_1, \dots, G_n$.

Main properties

- As in Sklar's theorem for copulas, the MDD representation is unique for fixed continuous DF $G_1, \dots, G_n$.
- If $D \in \mathcal{D}_n$, then

$$D(G_1(x_1), \dots, G_n(x_n))$$

is a multivariate distribution function for all DF $G_1, \dots, G_n$.
- If $\mathrm{F} \equiv MDD(G_1, \dots, G_n)$, then

$$\overline{\mathrm{F}}(x_1, \dots, x_n) = \widehat{D}(\bar{G}_1(x_1), \dots, \bar{G}_n(x_n)), \qquad (1.5)$$

where $\bar{G}_i = 1 - G_i$ and $\widehat{D} \in \mathcal{D}_n$.

Marginal distributions

- If F $\equiv MDD(G_1,\dots,G_n)$, then all the marginal distributions of F are also MDD from $G_1,\dots,G_n$.

Marginal distributions

- If F $\equiv MDD(G_1, \ldots, G_n)$, then all the marginal distributions of F are also MDD from $G_1, \ldots, G_n$.
- In particular, the ith marginal is

$$F_i(x_i) = D(1, \ldots, 1, G_i(x_i), 1, \ldots, 1) = D_i(G_i(x_i)), \qquad (1.6)$$

where $D_i(u) := D(1, \ldots, 1, u, 1, \ldots, 1)$ and the value u is placed at the ith position.

Marginal distributions

- If F $\equiv MDD(G_1, \dots, G_n)$, then all the marginal distributions of F are also MDD from $G_1, \dots, G_n$.
- In particular, the ith marginal is

$$F_i(x_i) = D(1, \dots, 1, G_i(x_i), 1, \dots, 1) = D_i(G_i(x_i)), \quad (1.6)$$

  where $D_i(u) := D(1, \dots, 1, u, 1, \dots, 1)$ and the value u is placed at the ith position.
- Clearly, we have $G_i = F_i$ for a fixed $i \in \{1, \dots, n\}$ when $D_i(u) = u$ for all $u \in [0, 1]$.

Conditional distributions

- All the conditional distributions of $F \equiv MDD(G_1, \dots, G_n)$ have MDD representations.

Conditional distributions

- All the conditional distributions of $F \equiv MDD(G_1, \dots, G_n)$ have MDD representations.
- We just consider the DF $F_{2|1}$ of $(X_2|X_1 = x_1)$.

Conditional distributions

- All the conditional distributions of $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$ have MDD representations.
- We just consider the DF $F_{2|1}$ of $(X_2|X_1 = x_1)$.

Proposition (Navarro, Calì, Longobardi and Durante (2022))

Let (X_1, X_2) *with* $\mathrm{F} \equiv MDD(G_1, G_2)$ *for* $D \in \mathcal{D}_2$*, then*

$$F_{2|1}(x_2|x_1) = D_{2|1}(G_2(x_2)|G_1(x_1)) \tag{1.7}$$

whenever $\lim_{v\to 0^+} \partial_1 D(G_1(x_1), v) = 0$*, where*

$$D_{2|1}(v|G_1(x_1)) = \frac{\partial_1 D(G_1(x_1), v)}{\partial_1 D(G_1(x_1), 1)}$$

for $0 < v < 1$ *and* x_1 *such that* $\partial_1 D(G_1(x_1), 1) > 0$*.*

Theoretical Quantile Regression

- The (mean) regression curve to predict X_2 from X_1 is $E(X_2|X_1 = x_1)$ can be obtained from (1.7).

Theoretical Quantile Regression

- The (mean) regression curve to predict X_2 from X_1 is $E(X_2|X_1 = x_1)$ can be obtained from (1.7).
- Another option to predict X_2 from X_1 is the *conditional median regression curve* $m_{2|1}(x_1) := F_{2|1}^{-1}(0.5|x_1)$ (see Koenker (2005) or Nelsen (2006), p. 217).

Theoretical Quantile Regression

- The (mean) regression curve to predict X_2 from X_1 is $E(X_2|X_1 = x_1)$ can be obtained from (1.7).
- Another option to predict X_2 from X_1 is the *conditional median regression curve* $m_{2|1}(x_1) := F_{2|1}^{-1}(0.5|x_1)$ (see Koenker (2005) or Nelsen (2006), p. 217).
- This quantile function $F_{2|1}^{-1}$ can be computed from (1.7) as

$$F_{2|1}^{-1}(q|x_1) = G_2^{-1}(D_{2|1}^{-1}(q|G_1(x_1))),\ 0 < q < 1.$$

Theoretical Quantile Regression

- The (mean) regression curve to predict X_2 from X_1 is $E(X_2|X_1 = x_1)$ can be obtained from (1.7).
- Another option to predict X_2 from X_1 is the *conditional median regression curve* $m_{2|1}(x_1) := F_{2|1}^{-1}(0.5|x_1)$ (see Koenker (2005) or Nelsen (2006), p. 217).
- This quantile function $F_{2|1}^{-1}$ can be computed from (1.7) as

$$F_{2|1}^{-1}(q|x_1) = G_2^{-1}(D_{2|1}^{-1}(q|G_1(x_1))),\ 0 < q < 1.$$

- Moreover, it can be used to obtain α-prediction bands for X_2

$$\left[F_{2|1}^{-1}(\beta_1|x_1), F_{2|1}^{-1}(\beta_2|x_1)\right]$$

taking $0 \le \beta_1 < \beta_2 \le 1$ such that $\beta_2 - \beta_1 = \alpha \in (0, 1)$.

Examples of MDD

- **Multivariate residual lifetimes**
  $\mathrm{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$, see Navarro, Calì, Longobardi and Durante (2022).

Examples of MDD

- **Multivariate residual lifetimes**
  $\mathbf{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$, see Navarro, Calì, Longobardi and Durante (2022).
- **Ordered paired data**, see next section.

Examples of MDD

- **Multivariate residual lifetimes**
  $\mathbf{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$, see Navarro, Calì, Longobardi and Durante (2022).
- **Ordered paired data**, see next section.
- **Coherent systems with ID components**, see Navarro, Calì, Longobardi and Durante (2022); Navarro et al. (2023).

Examples of MDD

- **Multivariate residual lifetimes** $\mathbf{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$, see Navarro, Calì, Longobardi and Durante (2022).
- **Ordered paired data**, see next section.
- **Coherent systems with ID components**, see Navarro, Calì, Longobardi and Durante (2022); Navarro et al. (2023).
- **Order statistics** (k-out-of-n systems with ID components), see Navarro and Buono (2022).

Examples of MDD

- **Multivariate residual lifetimes** $\mathbf{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$, see Navarro, Calì, Longobardi and Durante (2022).
- **Ordered paired data**, see next section.
- **Coherent systems with ID components**, see Navarro, Calì, Longobardi and Durante (2022); Navarro et al. (2023).
- **Order statistics** (k-out-of-n systems with ID components), see Navarro and Buono (2022).
- **Record values**, see Navarro (2022b).

Prediction of ordered paired data

Representations of ordered paired data

- Let us assume that X and Y are ID, that is, they have a common absolutely continuous distribution function F. Then

$$F_{X,Y}(x, y) = C(F(x), F(y)).$$

Representations of ordered paired data

- Let us assume that X and Y are ID, that is, they have a common absolutely continuous distribution function F. Then

$$F_{X,Y}(x,y) = C(F(x), F(y)).$$

- We also assume that C is abs. cont. and permutation symmetric, i.e., (X, Y) is exchangeable (EXC).

Representations of ordered paired data

- Let us assume that X and Y are ID, that is, they have a common absolutely continuous distribution function F. Then

$$F_{X,Y}(x,y) = C(F(x), F(y)).$$

- We also assume that C is abs. cont. and permutation symmetric, i.e., (X, Y) is exchangeable (EXC).
- We want to predict $U = \max(X, Y)$ from $L = \min(X, Y)$.

Representations of ordered paired data

- Let us assume that X and Y are ID, that is, they have a common absolutely continuous distribution function F. Then

$$F_{X,Y}(x,y) = C(F(x), F(y)).$$

- We also assume that C is abs. cont. and permutation symmetric, i.e., (X, Y) is exchangeable (EXC).
- We want to predict $U = \max(X, Y)$ from $L = \min(X, Y)$.
- To this purpose we need the conditional distribution function

$$G_{2|1}(s|t) := \Pr(U \leq s|L = t), \ s \geq t.$$

Representations of ordered paired data

- We assume that we have a training sample $(X_1, Y_1), \ldots, (X_m, Y_m)$ from (X, Y) to estimate C and F (or the unknown parameters in them).

Representations of ordered paired data

- We assume that we have a training sample $(X_1, Y_1), \dots, (X_m, Y_m)$ from (X, Y) to estimate C and F (or the unknown parameters in them).
- For other individuals, we just know $L = \min(X, Y)$ and we want to estimate $U = \max(X, Y)$.

Representations of ordered paired data

- We assume that we have a training sample $(X_1, Y_1), \dots, (X_m, Y_m)$ from (X, Y) to estimate C and F (or the unknown parameters in them).
- For other individuals, we just know $L = \min(X, Y)$ and we want to estimate $U = \max(X, Y)$.
- So, we want to obtain a MDD representation for the random vector (L, U) in terms of F and C.

Representations of ordered paired data

- We assume that we have a training sample $(X_1, Y_1), \ldots, (X_m, Y_m)$ from (X, Y) to estimate C and F (or the unknown parameters in them).
- For other individuals, we just know $L = \min(X, Y)$ and we want to estimate $U = \max(X, Y)$.
- So, we want to obtain a MDD representation for the random vector (L, U) in terms of F and C.
- It can be used to compute $G^{-1}_{2|1}(q|t)$, the median regression curve $m(t)$ and the associated prediction bands.

Representations of ordered paired data

- The joint distribution function $\mathrm{G}(x, y) = \Pr(L \leq x, U \leq y)$ of (L, U) can be written as

$$\mathrm{G}(x, y) = \begin{cases} C(F(y), F(y)) & \text{for} \quad y \leq x; \\ 2C(F(x), F(y)) - C(F(x), F(x)) & \text{for} \quad y > x. \end{cases}$$

Representations of ordered paired data

- The joint distribution function $\mathrm{G}(x,y)=\Pr(L\leq x, U\leq y)$ of (L,U) can be written as

$$\mathrm{G}(x,y)=\begin{cases} C(F(y),F(y)) & \text{for} \quad y\leq x; \\ 2C(F(x),F(y))-C(F(x),F(x)) & \text{for} \quad y>x. \end{cases}$$

- Therefore, $\mathrm{G}\equiv MDD(F,F)$, that is,

$$\mathrm{G}(x,y)=D(F(x),F(y)) \tag{2.1}$$

with the following distortion function

$$D(u,v)=\begin{cases} C(v,v) & \text{for} \quad v\leq u; \\ 2C(u,v)-C(u,u) & \text{for} \quad u<v. \end{cases} \tag{2.2}$$

Representations of ordered paired data

- Then the marginal distributions of (L, U) can be written as

$$G_1(x) := \Pr(L \leq x) = D(F(x), 1) = D_1(F(x)),$$

$$G_2(y) := \Pr(U \leq y) = D(1, F(y)) = D_2(F(y)),$$

where

$$D_1(u) = D(u, 1) = 2u - C(u, u)$$

and

$$D_2(v) = D(1, v) = C(v, v)$$

for all $u, v \in [0, 1]$.

Representations of ordered paired data, IID case

- If X and Y are independent, then the distortion function is

$$D(u,v)=\begin{cases} v^2 & \text{for} \quad v\leq u; \\ 2uv-u^2 & \text{for} \quad u<v. \end{cases} \tag{2.3}$$

Representations of ordered paired data, IID case

- If X and Y are independent, then the distortion function is

$$D(u,v)=\begin{cases} v^2 & \text{for} \quad v\le u; \\ 2uv-u^2 & \text{for} \quad u<v. \end{cases} \tag{2.3}$$

- Then, for all $u\in(0,1)$,

$$D_1(u)=D(u,1)=2u-u^2\neq u$$

and

$$D_2(u)=D(1,u)=u^2\neq u.$$

Representations of ordered paired data, IID case

- If X and Y are independent, then the distortion function is

$$D(u,v) = \begin{cases} v^2 & \text{for} \quad v \le u; \\ 2uv - u^2 & \text{for} \quad u < v. \end{cases} \tag{2.3}$$

- Then, for all $u \in (0,1)$,

$$D_1(u) = D(u,1) = 2u - u^2 \neq u$$

and

$$D_2(u) = D(1,u) = u^2 \neq u.$$

- Note that D is not a copula and that the marginals G_1 and G_2 of G do not appear in (2.1) (we use F instead).

Representations of ordered paired data

- From (1.7) and (2.1), the distribution function of $(U|L = x)$ is

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.4}$$

for $y \geq x$, where

$$D_{2|1}(v|F(x)) := \frac{\partial_1 D(F(x), v)}{\partial_1 D(F(x), 1)}$$

and $\partial_1 D(u, v) = 2\partial_1 C(u, v) - 2\partial_1 C(u, u)$, for $0 < u \leq v \leq 1$.

Representations of ordered paired data

- From (1.7) and (2.1), the distribution function of $(U|L=x)$ is

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.4}$$

for $y \geq x$, where

$$D_{2|1}(v|F(x)) := \frac{\partial_1 D(F(x), v)}{\partial_1 D(F(x), 1)}$$

and $\partial_1 D(u,v) = 2\partial_1 C(u,v) - 2\partial_1 C(u,u)$, for $0 < u \leq v \leq 1$.

- In the IID case, $\partial_1 D(u,v) = 2(v-u)$ for $u \leq v \leq 1$ and

$$D_{2|1}(v|F(x)) = \frac{v - F(x)}{\bar{F}(x)}$$

for $F(x) \leq v \leq 1$.

Example 1: IID exponential case

- Let (X_i, Y_i) be a sample from (X, Y) where X, Y are IID$\sim F$.

Example 1: IID exponential case

- Let (X_i, Y_i) be a sample from (X, Y) where X, Y are IID$\sim F$.
- Let $L_i = \min(X_i, Y_i)$ and $U_i = \max(X_i, Y_i)$.

Example 1: IID exponential case

- Let (X_i, Y_i) be a sample from (X, Y) where X, Y are IID$\sim F$.
- Let $L_i = \min(X_i, Y_i)$ and $U_i = \max(X_i, Y_i)$.
- Note that L_i and U_i are dependent.

Example 1: IID exponential case

- Let (X_i, Y_i) be a sample from (X, Y) where X, Y are IID$\sim F$.
- Let $L_i = \min(X_i, Y_i)$ and $U_i = \max(X_i, Y_i)$.
- Note that L_i and U_i are dependent.
- From (2.4), the distribution function of $(U|L = x)$ is

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.5}$$

for $y \geq x$, where

$$D_{2|1}(v|F(x)) = \frac{v - F(x)}{\bar{F}(x)}$$

for $F(x) \leq v \leq 1$.

Example 1: IID exponential case

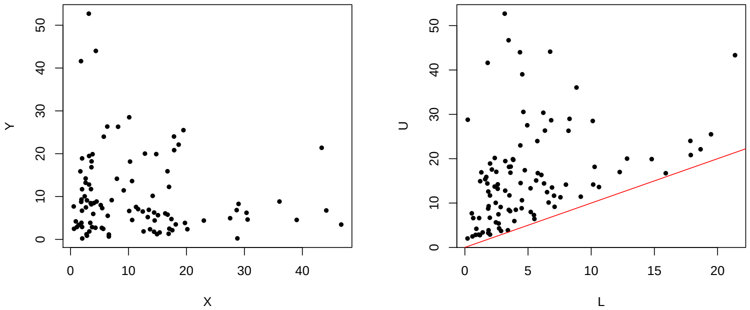


Figure: Independent data from two exponential distributions with mean $\mu = 10$ (left) and the associated ordered paired data (right).

Example 1: IID exponential case

- The quantile function $F_{2|1}^{-1}$ can be computed as

$$F_{2|1}^{-1}(q|x) = F^{-1}(D_{2|1}^{-1}(q|F(x)))$$

for $0 < v < 1$, where $D_{2|1}^{-1}(q|F(x)) = F(x) + q\bar{F}(x)$, when $\bar{F}(x) = \exp(-x/\mu)$ and $F^{-1}(y) = -\mu \log(1-y)$. Then

$$F_{2|1}^{-1}(q|x) = -\mu \log\left((1-q)e^{-x/\mu}\right) = x - \mu \log(1-q).$$

Example 1: IID exponential case

- The quantile function $F_{2|1}^{-1}$ can be computed as

$$F_{2|1}^{-1}(q|x) = F^{-1}(D_{2|1}^{-1}(q|F(x)))$$

for $0 < v < 1$, where $D_{2|1}^{-1}(q|F(x)) = F(x) + q\bar{F}(x)$, when $\bar{F}(x) = \exp(-x/\mu)$ and $F^{-1}(y) = -\mu \log(1-y)$. Then

$$F_{2|1}^{-1}(q|x) = -\mu \log\left((1-q)e^{-x/\mu}\right) = x - \mu \log(1-q).$$

- Therefore, the exact QR curve is

$$m(x) = x - \mu \log(0.5).$$

Example 1: IID exponential case

- Analogously, the exact QR centered 90% prediction band is

$$[x - \mu \log(0.95), x - \mu \log(0.05)] .$$

Example 1: IID exponential case

- Analogously, the exact QR centered 90% prediction band is

$$[x - \mu \log(0.95), x - \mu \log(0.05)] .$$

- The 50% centered prediction band is obtained in a similar way.

Example 1: IID exponential case

- Analogously, the exact QR centered 90% prediction band is

$$[x - \mu \log(0.95), x - \mu \log(0.05)] .$$

- The 50% centered prediction band is obtained in a similar way.
- The exact QR lower 90% prediction band is

$$[x, x - \mu \log(0.10)] .$$

Example 1: IID exponential case

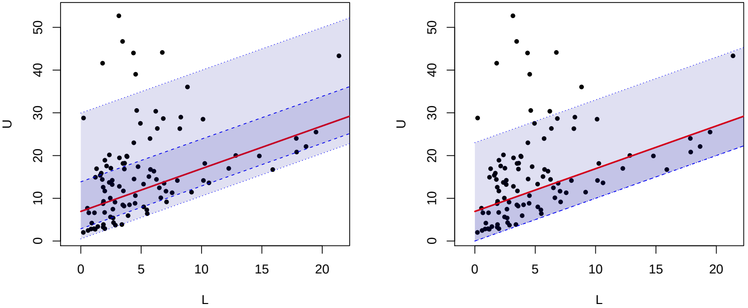


Figure: QR for the ordered paired data (L, U) associated to independent data (X, Y) from two exponential distributions with mean $\mu = 10$ jointly with 50% and 90% centered (left) or bottom (right) prediction bands.

Predictions

- The first ordered pair in our simulated sample is $L_1 = 10.15771$ and $U_1 = 14.17195$.

Predictions

- The first ordered pair in our simulated sample is $L_1 = 10.15771$ and $U_1 = 14.17195$.
- The prediction for U_1 from L_1 is

$$m(L_1) = m(10.15771) = 10.15771 - \mu \log(0.5) = 17.08918.$$

Predictions

- The first ordered pair in our simulated sample is $L_1 = 10.15771$ and $U_1 = 14.17195$.
- The prediction for U_1 from L_1 is

$$m(L_1) = m(10.15771) = 10.15771 - \mu \log(0.5) = 17.08918.$$

- The centered 90% prediction interval for this prediction is $[10.67064, 40.11503]$.

Predictions

- The first ordered pair in our simulated sample is $L_1 = 10.15771$ and $U_1 = 14.17195$.
- The prediction for U_1 from L_1 is

$$m(L_1) = m(10.15771) = 10.15771 - \mu \log(0.5) = 17.08918.$$

- The centered 90% prediction interval for this prediction is $[10.67064, 40.11503]$.
- The centered 50% prediction interval for this prediction is $[13.03453, 24.02065]$.

Example 2: Dependent EXC exponential case

- Let us consider now that (X, Y) are DID with a copula C and a common marginal distribution F.

Example 2: Dependent EXC exponential case

- Let us consider now that (X, Y) are DID with a copula C and a common marginal distribution F.
- We consider again the exponential model

$$\bar{F}(t) = \exp(-t/\mu), \ \ t \geq 0$$

and the Clayton EXC copula

$$C(u, v) = \frac{uv}{u + v - uv}, \ \ (u, v) \in [0, 1]^2. \qquad (2.6)$$

Example 2: Dependent EXC exponential case

- To get the QR curves we need the distribution $G_{2|1}(y|x)$ of $(U|L = x)$. From (2.4) we need

$$\partial_1 D(u, v) = 2\partial_1 C(u, v) - 2\partial_1 C(u, u) = \frac{2v^2}{(u + v - uv)^2} - \frac{2}{(2 - u)^2}$$

and

$$\partial_1 D(u, 1) = \frac{2}{(u + 1 - u)^2} - \frac{2}{(2 - u)^2} = 2 - \frac{2}{(2 - u)^2}.$$

Example 2: Dependent EXC exponential case

- To get the QR curves we need the distribution $G_{2|1}(y|x)$ of $(U|L=x)$. From (2.4) we need

$$\partial_1 D(u,v) = 2\partial_1 C(u,v) - 2\partial_1 C(u,u) = \frac{2v^2}{(u+v-uv)^2} - \frac{2}{(2-u)^2}$$

and

$$\partial_1 D(u,1) = \frac{2}{(u+1-u)^2} - \frac{2}{(2-u)^2} = 2 - \frac{2}{(2-u)^2}.$$

- Hence, for $v \geq u$, we get

$$D_{2|1}(v|u) = \frac{\frac{v^2}{(u+v-uv)^2} - \frac{1}{(2-u)^2}}{1 - \frac{1}{(2-u)^2}}$$

Example 2: Dependent EXC exponential case

- To get the QR curves we need the distribution $G_{2|1}(y|x)$ of $(U|L = x)$. From (2.4) we need

$$\partial_1 D(u,v) = 2\partial_1 C(u,v) - 2\partial_1 C(u,u) = \frac{2v^2}{(u+v-uv)^2} - \frac{2}{(2-u)^2}$$

and

$$\partial_1 D(u,1) = \frac{2}{(u+1-u)^2} - \frac{2}{(2-u)^2} = 2 - \frac{2}{(2-u)^2}.$$

- Hence, for $v \geq u$, we get

$$D_{2|1}(v|u) = \frac{\frac{v^2}{(u+v-uv)^2} - \frac{1}{(2-u)^2}}{1 - \frac{1}{(2-u)^2}}$$

- To compute the inverse, we need to solve in y the equation $G_{2|1}(y|x) = q$ for $q \in (0,1)$.

Example 2: Dependent EXC exponential case

- This equation leads to

$$G_{2|1}^{-1}(q|x) = F^{-1}\left(\frac{F(x)}{F(x) - 1 + \frac{2-F(x)}{\sqrt{1-q+q(2-F(x))^2}}}\right).$$

Example 2: Dependent EXC exponential case

- This equation leads to

$$G_{2|1}^{-1}(q|x) = F^{-1}\left(\frac{F(x)}{F(x) - 1 + \frac{2-F(x)}{\sqrt{1-q+q(2-F(x))^2}}}\right).$$

- Therefore, the exact median regression curve to predict U from $L = x$ is

$$m(x) = G_{2|1}^{-1}(0.5|x).$$

Example 2: Dependent EXC exponential case

- This equation leads to

$$G_{2|1}^{-1}(q|x) = F^{-1}\left(\frac{F(x)}{F(x)-1+\frac{2-F(x)}{\sqrt{1-q+q(2-F(x))^2}}}\right).$$

- Therefore, the exact median regression curve to predict U from $L = x$ is

$$m(x) = G_{2|1}^{-1}(0.5|x).$$

- Analogously, the 90% centered prediction band is

$$\left[G_{2|1}^{-1}(0.05|x), G_{2|1}^{-1}(0.95|x)\right].$$

Example 2: Dependent EXC exponential case

- This equation leads to

$$G_{2|1}^{-1}(q|x) = F^{-1}\left(\frac{F(x)}{F(x)-1+\frac{2-F(x)}{\sqrt{1-q+q(2-F(x))^2}}}\right).$$

- Therefore, the exact median regression curve to predict U from $L = x$ is

$$m(x) = G_{2|1}^{-1}(0.5|x).$$

- Analogously, the 90% centered prediction band is

$$\left[G_{2|1}^{-1}(0.05|x), G_{2|1}^{-1}(0.95|x)\right].$$

- The other prediction bands can be obtained in a similar way.

Example 2: Dependent EXC exponential case

- This equation leads to

$$G_{2|1}^{-1}(q|x) = F^{-1}\left(\frac{F(x)}{F(x) - 1 + \frac{2-F(x)}{\sqrt{1-q+q(2-F(x))^2}}}\right).$$

- Therefore, the exact median regression curve to predict U from $L = x$ is

$$m(x) = G_{2|1}^{-1}(0.5|x).$$

- Analogously, the 90% centered prediction band is

$$\left[G_{2|1}^{-1}(0.05|x), G_{2|1}^{-1}(0.95|x)\right].$$

- The other prediction bands can be obtained in a similar way.
- For an exponential distribution with $\mu = 10$ we get

Example 2: Dependent EXC exponential case

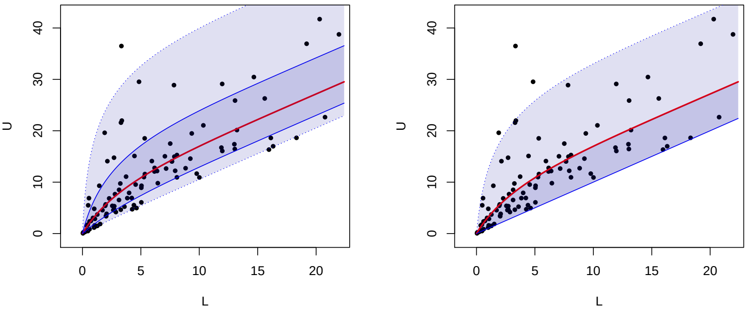


Figure: QR curves for paired ordered data (L, U) associated to dependent data (X, Y) from two exponential distributions with centered (left) and bottom (right) prediction bands.

Parametric QR curves

- Our model can contain some unknown parameters.

Parametric QR curves

- Our model can contain some unknown parameters.
- They can be both in F and/or in C.

Parametric QR curves

- Our model can contain some unknown parameters.
- They can be both in F and/or in C.
- We can use the training sample (X_i, Y_i) from (X, Y) to estimate the unknown parameters.

Parametric QR curves

- Our model can contain some unknown parameters.
- They can be both in F and/or in C.
- We can use the training sample (X_i, Y_i) from (X, Y) to estimate the unknown parameters.
- Then we can use the MDD representation with the estimated parameters to get the estimated QR curves.

Example 1: IID exponential case

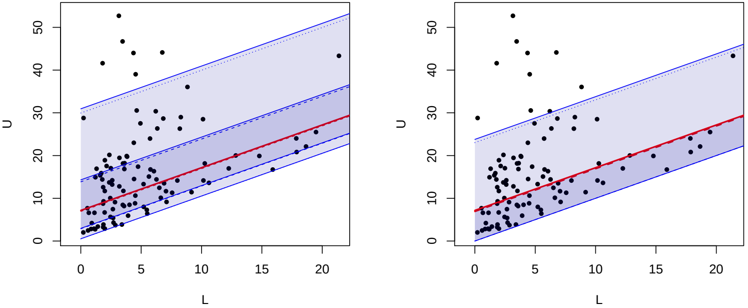
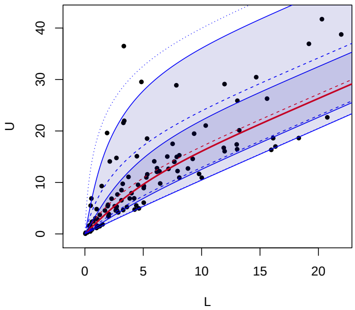


Figure: Parametric QR curves for (L, U) associated to IID data (X, Y) from an exponential distribution jointly with centered (left) and bottom (right) prediction bands. The dashed lines are the exact curves.

Example 2: Dependent EXC exponential case



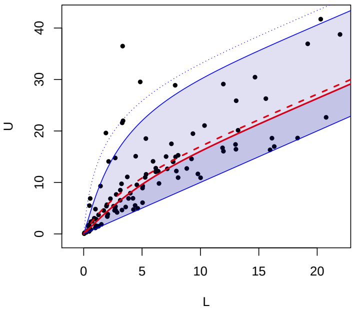


Figure: Parametric QR curves for (L, U) associated to data (X, Y) from an exponential distribution with unknown mean μ and a Clayton copula with unknown parameter θ. The dashed lines are the exact curves.

Non-parametric QR curves.

- If we do not have a parametric model, we can use non-parametric estimators for F and C.

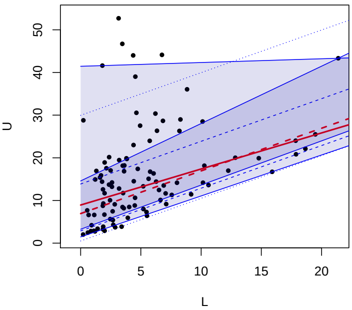
Non-parametric QR curves.

- If we do not have a parametric model, we can use non-parametric estimators for F and C.
- We can also use the statistical program R with `library('quantreg')` to estimate the exact curves from the training sample (see Koenker, 2005).

Non-parametric QR curves.

- If we do not have a parametric model, we can use non-parametric estimators for F and C.
- We can also use the statistical program R with `library('quantreg')` to estimate the exact curves from the training sample (see Koenker, 2005).
- Here k denotes the degree of the polynomial used to estimate the QR curves.

Example 1: Non-parametric QR, IID exponential case



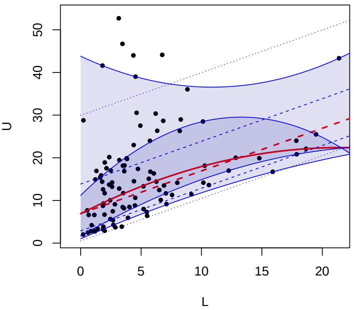


Figure: Non-parametric QR curves for paired ordered data (L, U) associated to IID data (X, Y) from an exponential distribution with $\mu = 10$ and $k = 1$ (left) or $k = 2$ (right).

Example 2: Non-parametric QR, DID exponential case

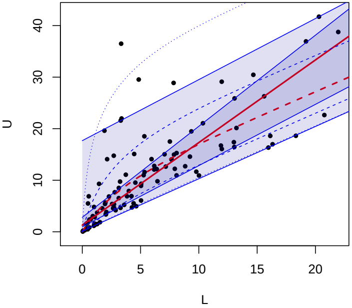


Figure: Non-parametric QR curves for (L, U) associated to data (X, Y) from an exponential distribution and a Clayton copula with $\theta = 1$ and $k = 1$ (left) or $k = 2$ (right). The dashed lines are the exact curves.

Conclusions

- Similar results for future order statistics (ordered data) with or without dependency can be seen in Navarro and Buono (2022) and Buono and Navarro (2022).

Conclusions

- Similar results for future order statistics (ordered data) with or without dependency can be seen in Navarro and Buono (2022) and Buono and Navarro (2022).
- Similar results for record values can be seen in Navarro (2022b).

Conclusions

- Similar results for future order statistics (ordered data) with or without dependency can be seen in Navarro and Buono (2022) and Buono and Navarro (2022).
- Similar results for record values can be seen in Navarro (2022b).
- Predictions for sums of dependent random lifetimes are included in Navarro, Pellerey and Mulero (2022).

Conclusions

- Similar results for future order statistics (ordered data) with or without dependency can be seen in Navarro and Buono (2022) and Buono and Navarro (2022).
- Similar results for record values can be seen in Navarro (2022b).
- Predictions for sums of dependent random lifetimes are included in Navarro, Pellerey and Mulero (2022).
- Prediction for system failure times will be included in Navarro et al. (2023).

Conclusions

- Similar results for future order statistics (ordered data) with or without dependency can be seen in Navarro and Buono (2022) and Buono and Navarro (2022).
- Similar results for record values can be seen in Navarro (2022b).
- Predictions for sums of dependent random lifetimes are included in Navarro, Pellerey and Mulero (2022).
- Prediction for system failure times will be included in Navarro et al. (2023).
- The representations based on distortions are very useful as an alternative to the classic copula representations !!

References

Awad AM, Raqab MZ (2000) Prediction intervals for the future record values from exponential distribution: comparative study. Journal of Statistical Computation and Simulation 65, 325–340.

Buono F, Navarro J (2022) Simulations and predictions of future values in the time homogeneous load-sharing model. Submitted.

Koenker R (2005). Quantile Regression. Cambridge University Press.

Navarro J (2022a). Introduction to System Reliability Theory. Springer.

Navarro, J (2022b). Prediction of record values by using quantile regression curves and distortion functions. Metrika 85, 675–706. DOI: 10.1007/s00184-021-00847-w.

Navarro J (2023). Predicting system failure times. In preparation.

Navarro J, Buono F (2022). Predicting future failure times by using quantile regression. To appear in Metrika. Published online first Sept 2022. DOI https://doi.org/10.1007/s00184-022-00884-z.

References

Navarro J, Calì C, Longobardi M, Durante F (2022). Distortion representations of multivariate distributions. Statistical Methods & Applications 31, 925–954. DOI: 10.1007/s10260-021-00613-2.

Navarro J, Pellerey F, Mulero J (2022). On sums of dependent random lifetimes under the time-transformed exponential model. To appear in Test. Published online first March 2022. DOI: 10.1016/j.physa.2022.127214.

Nelsen RB (2006). An introduction to copulas. Second edition. Springer, New York.

Nevzorov VB (2001) Records: Mathematical Theory. American Mathematical Society, Providence, RI.

Wang S (1996). Premium calculation by transforming the layer premium density. Astin Bulletin 26, 71–92.

Yaari ME (1987). The dual theory of choice under risk. Econometrica 55, 95–115.

Final slide

- For more references visit: https://webs.um.es/jorgenav.

Final slide

- For more references visit: https://webs.um.es/jorgenav.
- Publicity: The next year we will organize the

12th International Conference on Mathematical Methods in Reliability (MMR2023)
May 30 to June 2, 2023, in Murcia, Spain.

Final slide

- For more references visit: https://webs.um.es/jorgenav.
- Publicity: The next year we will organize the

  12th International Conference on Mathematical Methods in Reliability (MMR2023)
  May 30 to June 2, 2023, in Murcia, Spain.

- That's all.

Final slide

- For more references visit: https://webs.um.es/jorgenav.
- Publicity: The next year we will organize the

12th International Conference on Mathematical Methods in Reliability (MMR2023)
May 30 to June 2, 2023, in Murcia, Spain.

- That's all.
- Thank you for your attention!!

Final slide

- For more references visit: https://webs.um.es/jorgenav.
- Publicity: The next year we will organize the

12th International Conference on Mathematical Methods in Reliability (MMR2023)
May 30 to June 2, 2023, in Murcia, Spain.

- That's all.
- Thank you for your attention!!
- Questions?