Predicciones usando distorsiones bivariates y regresión cuantílica

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.

Referencias

- La conferencia se basa en el artículo:

Referencias

- La conferencia se basa en el artículo:
- Navarro J, Calì C, Longobardi M, Durante F (2022). Distortion Representations of Multivariate Distributions. Statistical Methods & Applications. Published online first Jan. 2022. DOI: 10.1007/s10260-021-00613-2.

Outline

Distribuciones distorsionadas

Notación

- X variable aleatoria (tiempo de vida) sobre $(\Omega, \mathcal{S}, \Pr)$.

Notación

- X variable aleatoria (tiempo de vida) sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución $F(t) = \Pr(X \leq t)$.

Notación

- X variable aleatoria (tiempo de vida) sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución $F(t) = \Pr(X \leq t)$.
- Función de fiabilidad o supervivencia
  $\bar{F}(t) = \Pr(X > t) = 1 - F(t)$.

Notación

- X variable aleatoria (tiempo de vida) sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución $F(t) = \Pr(X \leq t)$.
- Función de fiabilidad o supervivencia $\bar{F}(t) = \Pr(X > t) = 1 - F(t)$.
- Función de densidad (PDF) $f(t) = F'(t) = -\bar{F}'(t)$.

Distribuciones distorsionadas univariantes

- Las distribuciones distorsionadas univariantes fueron introducidas por Wang (1996) y Yaari (1987) en el contexto de la teoría de decisión bajo riesgos. Mencionar también la referencia española (INE) Miras (1991).

Distribuciones distorsionadas univariantes

- Las distribuciones distorsionadas univariantes fueron introducidas por Wang (1996) y Yaari (1987) en el contexto de la teoría de decisión bajo riesgos. Mencionar también la referencia española (INE) Miras (1991).
- El proposito es permitir un cambio (distorsión) en la distribución inicial (histórica) del riesgo.

Distribuciones distorsionadas univariantes

- Las distribuciones distorsionadas univariantes fueron introducidas por Wang (1996) y Yaari (1987) en el contexto de la teoría de decisión bajo riesgos. Mencionar también la referencia española (INE) Miras (1991).
- El proposito es permitir un cambio (distorsión) en la distribución inicial (histórica) del riesgo.
- La **distribución distorsionada** (DD) asociada a F y a una función de distorsión creciente y continua $q:[0,1]\to[0,1]$ verificando $q(0)=0$ y $q(1)=1$ es

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

Propiedades

- Si q es una distorsión, entonces F_q es una función de distrobución para toda distribución F.

Propiedades

- Si q es una distorsión, entonces F_q es una función de distrobución para toda distribución F.
- De (1.1), las funciones de fiabilidad $\bar{F} = 1 - F$ y $\bar{F}_q = 1 - F_q$ verifican
$$\bar{F}_q(t) = \bar{q}(\bar{F}(t)), \text{ para todo } t \in \mathbb{R}, \tag{1.2}$$
donde $\bar{q}(u) := 1 - q(1 - u)$ se llama *distorsión dual*.

Propiedades

- Si q es una distorsión, entonces F_q es una función de distrobución para toda distribución F.
- De (1.1), las funciones de fiabilidad $\bar{F} = 1 - F$ y $\bar{F}_q = 1 - F_q$ verifican

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

donde $\bar{q}(u) := 1 - q(1-u)$ se llama *distorsión dual*.
- Las representaciones (1.1) son (1.2) equivalentes.

Propiedades

- Si q es una distorsión, entonces F_q es una función de distrobución para toda distribución F.
- De (1.1), las funciones de fiabilidad $\bar{F} = 1 - F$ y $\bar{F}_q = 1 - F_q$ verifican
$$\bar{F}_q(t) = \bar{q}(\bar{F}(t)), \text{ para todo } t \in \mathbb{R}, \tag{1.2}$$
donde $\bar{q}(u) := 1 - q(1 - u)$ se llama *distorsión dual*.
- Las representaciones (1.1) son (1.2) equivalentes.
- La densidad de F_q es
$$f_q(t) = q'(F(t))f(t) = \bar{q}'(\bar{F}(t))f(t).$$

Generalized distorted distributions

- Este concepto se extendió en Navarro, del Águila, Sordo and Suárez-Llorens (2016) mediante:

Generalized distorted distributions

- Este concepto se extendió en Navarro, del Águila, Sordo and Suárez-Llorens (2016) mediante:
- Llamamos **distribución distorsionada generalizada** asociada a n distribuciones $F_1, \ldots, F_n$ y a una distorsión creciente y continua $Q : [0,1]^n \to [0,1]$ tal que $Q(0,\ldots,0) = 0$ y $Q(1,\ldots,1) = 1$ a

$$F_Q(t) = Q(F_1(t), \ldots, F_n(t)), \text{ para todo } t \in \mathbb{R}. \qquad (1.3)$$

Generalized distorted distributions

- Este concepto se extendió en Navarro, del Águila, Sordo and Suárez-Llorens (2016) mediante:
- Llamamos **distribución distorsionada generalizada** asociada a n distribuciones $F_1, \ldots, F_n$ y a una distorsión creciente y continua $Q : [0,1]^n \to [0,1]$ tal que $Q(0,\ldots,0) = 0$ y $Q(1,\ldots,1) = 1$ a

$$F_Q(t) = Q(F_1(t), \ldots, F_n(t)), \text{ para todo } t \in \mathbb{R}. \qquad (1.3)$$

- Aplicaciones: Modelo de Riesgo Proporcional (PHR) de Cox, Modelo de Riesgo Reverso Proporcional (PRHR), Estadísticos Ordenados, Valores Record, Mixturas, Sistemas Coherentes, ver Navarro (2022).

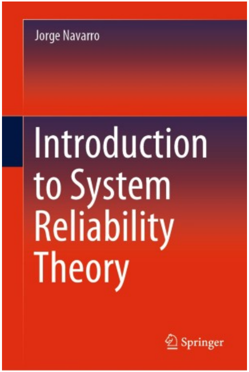

Figura: Publicidad de mi nuevo libro :-)

Notación

- $(X_1, \dots, X_n)$ vector aleatorio sobre $(\Omega, \mathcal{S}, \Pr)$.

Notación

- $(X_1, \ldots, X_n)$ vector aleatorio sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución conjunta

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

Notación

- $(X_1, \ldots, X_n)$ vector aleatorio sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución conjunta

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

- Representación con cópulas (teorema de Sklar)

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

donde $F_1, \ldots, F_n$ son las distribuciones marginales y C es una cópula.

Notación

- $(X_1, \dots, X_n)$ vector aleatorio sobre $(\Omega, \mathcal{S}, \Pr)$.
- Función de distribución conjunta

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

- Representación con cópulas (teorema de Sklar)

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

donde $F_1, \dots, F_n$ son las distribuciones marginales y C es una cópula.

- Hay una representación similar para la función de supervivencia conjunta

$$\overline{\mathsf{F}}(x_1, \dots, x_n) = \Pr(X_1 > x_1, \dots, X_n > x_n) = \hat{C}(\bar{F}_1(x_1), \dots, \bar{F}_n(x_n)),$$

donde $\hat{C}$ es la copula de supervivencia.

Definición de distorsiones multivariantes

Definición (Navarro, Calì, Longobardi and Durante (2021))

Una función de distribución multivariante F *es una* **distribución distorsionada multivariante** *(MDD) de las distribuciones univariantes* $G_1, \ldots, G_n$ *si existe una* **distorsión** $D : [0,1]^n \to [0,1]$ *tal que*

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

Usaremos la notación $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$*, cuando* F *es una MDD de* $G_1, \ldots, G_n$.

Definition

Definición

Una función continua $D : [0,1]^n \to [0,1]$ *es una* **distorsión** n–**dimensional** *(*$D \in \mathcal{D}_n$*) si*

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

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

(iii) D *es* n*–creciente, es decir, para todo* $\mathrm{x} = (x_1, \dots, x_n)$ *y* $\mathrm{y} = (y_1, \dots, y_n)$ *con* $x_i \le y_i$*, se verifica* $\triangle_{\mathrm{x}}^{\mathrm{y}} D \ge 0$*, donde*

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

con $1(z_1, \dots, z_n) = \sum_{i=1}^n 1(z_i = x_i)$ *y* $1(A) = 1$ *(respectivamente, 0) si* A *es cierta (respectivamente, falsa).*

Propiedades principales

- El teorema de Sklar asegura que la representación con cópulas siempre existe.

Propiedades principales

- El teorema de Sklar asegura que la representación con cópulas siempre existe.
- Si las marginales son continuas, la representación (la cópula) es única.

Propiedades principales

- El teorema de Sklar asegura que la representación con cópulas siempre existe.
- Si las marginales son continuas, la representación (la cópula) es única.
- Se tiene un resultado similar para distorsiones.

Teorema similar al teorema de Sklar

Proposición

Sea $(X_1, \ldots, X_n)$ *con distribución* F *y sean* $G_1, \ldots, G_n$ *distribuciones continuas. Supongamos que* G_i *es estrictamente creciente en el soporte de* X_i *para* $i = 1, \ldots, n$*. Entonces existe una única distorsión* $D \in \mathcal{D}_n$ *tal que*

$$\mathsf{F}(x_1, \ldots, x_n) = D(G_1(x_1), \ldots, G_n(x_n))$$

para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Construcción de modelos multivariantes

Proposición

Si $D \in \mathcal{D}_n$, entonces

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

es una distribución multivariante para todas las distribuciones multivariantes $G_1, \ldots, G_n$.

Relación con la cópula

Proposición

Sea $(X_1, \dots, X_n)$ *con una distribución continua* F. *Sean* $G_1, \dots, G_n$ *distribuciones continuas tales que* $\mathsf{F} \equiv MDD(G_1, \dots, G_n)$ *con distorsión* D. *Entonces*

$$D(u_1, \dots, u_n) = C(F_1(G_1^{-1}(u_1)), \dots, F_n(G_n^{-1}(u_n)))$$

para todo $(u_1, \dots, u_n) \in [0, 1]^n$, *donde* G_i^{-1} *es la cuasi-inversa de* G_i *y* F_i *es la marginal i-ésima de* F *para* $i = 1, \dots, n$.

Representación para supervivencias.

Proposición

Sea $(X_1, \ldots, X_n)$ *con distribución* F*. Si se cumple* (1.4) *para* $G_1, \ldots, G_n$ *y* $D \in \mathcal{D}_n$*, entonces*

$$\overline{\mathsf{F}}(x_1, \ldots, x_n) = \hat{D}(\bar{G}_1(x_1), \ldots, \bar{G}_n(x_n)) \qquad (1.5)$$

para todo $x_1, \ldots, x_n$*, donde* $\bar{G}_i = 1 - G_i$ *y* $\hat{D} \in \mathcal{D}_n$ *(distorsión dual o de supervivencia).*

Distribuciones marginales

- Una propiedad importante de $F \equiv MDD(G_1, \dots, G_n)$ es que todas sus marginales también son MDD de $G_1, \dots, G_n$.

Distribuciones marginales

- Una propiedad importante de $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$ es que todas sus marginales también son MDD de $G_1, \ldots, G_n$.
- Por ejemplo, si $F_{1,\ldots,m}$ es la distribución de $(X_1, \ldots, X_m)$, entonces

$$F_{1,\ldots,m}(x_1, \ldots, x_m) = D_{1,\ldots,m}(G_1(x_1), \ldots, G_m(x_m)) \quad (1.6)$$

para todo $(x_1, \ldots, x_m) \in \mathbb{R}^m$, donde

$$D_{1,\ldots,m}(u_1, \ldots, u_m) := D(u_1, \ldots, u_m, 1, \ldots, 1)$$

para $(u_1, \ldots, u_m) \in [0, 1]^m$ y $D_{1,\ldots,m} \in \mathcal{D}_m$.

Distribuciones marginales univariantes.

- In particular, la marginal de X_i es

$$F_i(x_i) = D(1, \dots, 1, G_i(x_i), 1, \dots, 1) = D_i(G_i(x_i)) \qquad (1.7)$$

para todo $x_i \in \mathbb{R}$, donde

$$D_i(u) := D(1, \dots, 1, u, 1, \dots, 1)$$

y u se pone en la posición i-ésimna.

Distribuciones marginales univariantes.

- In particular, la marginal de X_i es

$$F_i(x_i) = D(1,\dots,1,G_i(x_i),1,\dots,1) = D_i(G_i(x_i)) \qquad (1.7)$$

para todo $x_i \in \mathbb{R}$, donde

$$D_i(u) := D(1,\dots,1,u,1,\dots,1)$$

y u se pone en la posición i-ésimna.

- Se tiene $G_i = F_i$ para un $i \in \{1,\dots,n\}$ si y solo si $D_i(u) = u$ para todo $u \in [0,1]$.

Función de densidad conjunta

Si F es abs, continua con densidad f, tenemos

$$\mathrm{f}(x_1, \ldots, x_n) = \partial_{1,\ldots,n}\mathrm{F}(x_1, \ldots, x_n) \ (a.e.).$$

Proposición

Si $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$*,* $G_1, \ldots, G_n$ *tienen densidades* $g_1, \ldots, g_n$*, y la distorsión* D *tiene derivadas de orden* n*, entonces*

$$\mathrm{f}(x_1, \ldots, x_n) = g_1(x_1) \ldots g_n(x_n)\, \partial_{1,\ldots,n} D(G_1(x_1), \ldots, G_n(x_n)) \ (a.e.).$$

Distribuciones condicionadas

- Todas las distribuciones condicionadas de $\mathrm{F} \equiv MDD(G_1, \dots, G_n)$ tienen representaciones MDD.

Distribuciones condicionadas

- Todas las distribuciones condicionadas de $F \equiv MDD(G_1, \ldots, G_n)$ tienen representaciones MDD.
- Solo consideramos la distribución $F_{2|1}$ de $(X_2|X_1 = x_1)$.

Distribuciones condicionadas

- Todas las distribuciones condicionadas de $\mathrm{F} \equiv MDD(G_1, \ldots, G_n)$ tienen representaciones MDD.
- Solo consideramos la distribución $F_{2|1}$ de $(X_2|X_1 = x_1)$.

Proposición

Sea (X_1, X_2) *con* $\mathrm{F} \equiv MDD(G_1, G_2)$ *para* $D \in \mathcal{D}_2$ *con derivadas continuas de orden* 2, *entonces*

$$F_{2|1}(x_2|x_1) = D_{2|1}(G_2(x_2)|G_1(x_1)) \tag{1.8}$$

siempre que $\lim_{v\to 0^+} \partial_1 D(G_1(x_1), v) = 0$, *donde*

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

para $0 < v < 1$ *y* x_1 *tales que* $\partial_1 D(G_1(x_1), 1) > 0$.

Regresión cuantílica (QR) teórica

- La **curva de regresión (media)** para predecir X_2 a partir de X_1 es

$$\widetilde{m}_{2|1}(x_1) = E(X_2|X_1 = x_1)$$

Regresión cuantílica (QR) teórica

- La **curva de regresión (media)** para predecir X_2 a partir de X_1 es
$$\widetilde{m}_{2|1}(x_1) = E(X_2|X_1 = x_1)$$
- Otra opción es usar la **curva de regresión mediana**
$$m_{2|1}(x_1) := F_{2|1}^{-1}(0{,}5|x_1)$$

(ver Koenker (2005) o Nelsen (2006), p. 217).

Regresión cuantílica (QR) teórica

- La **curva de regresión (media)** para predecir X_2 a partir de X_1 es
$$\widetilde{m}_{2|1}(x_1) = E(X_2|X_1 = x_1)$$
- Otra opción es usar la **curva de regresión mediana**
$$m_{2|1}(x_1) := F^{-1}_{2|1}(0{,}5|x_1)$$
(ver Koenker (2005) o Nelsen (2006), p. 217).
- La función cuantílica $F^{-1}_{2|1}$ también se puede usar para obtener bandas de confianza α para esas predicciones con
$$\left[F^{-1}_{2|1}(\beta_1|x_1), F^{-1}_{2|1}(\beta_2|x_1)\right]$$
tomando $0 \leq \beta_1 < \beta_2 \leq 1$ tales que $\beta_2 - \beta_1 = \alpha$.

Bandas de confianza teóricas

- La función cuantílica $F_{2|1}^{-1}$ se puede calcular de (1.8) como

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

Bandas de confianza teóricas

- La función cuantílica $F_{2|1}^{-1}$ se puede calcular de (1.8) como

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

- Esta función se puede usar para obtener la curva de regresión mediana como

$$m_{2|1}(x_1) = G_2^{-1}(D_{2|1}^{-1}(0{,}5|G_1(x_1))).$$

Bandas de confianza teóricas

- La función cuantílica $F_{2|1}^{-1}$ se puede calcular de (1.8) como

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

- Esta función se puede usar para obtener la curva de regresión mediana como

$$m_{2|1}(x_1) = G_2^{-1}(D_{2|1}^{-1}(0{,}5|G_1(x_1))).$$

- También se puede usar para obtener las bandas de confianza. Por ejemplo, las bandas de confianza 50 % y 90 % centradas para $(X_2|X_1 = x_1)$ son

$$\left[G_2^{-1}(D_{2|1}^{-1}(0{,}25|G_1(x_1))), G_2^{-1}(D_{2|1}^{-1}(0{,}75|G_1(x_1)))\right]$$

y

$$\left[G_2^{-1}(D_{2|1}^{-1}(0{,}05|G_1(x_1))), G_2^{-1}(D_{2|1}^{-1}(0{,}95|G_1(x_1)))\right].$$

Regresión cuantílica (QR) práctica

- En la práctica debemos estimar los parámetros del modelo o estimar esas curvas de forma no-paramétrica usando QR a partir de una muestra de datos de (X_1, X_2): $(x_1(i), x_2(i)),\ i = 1, \dots, n.$

Regresión cuantílica (QR) práctica

- En la práctica debemos estimar los parámetros del modelo o estimar esas curvas de forma no-paramétrica usando QR a partir de una muestra de datos de (X_1, X_2): $(x_1(i), x_2(i)),\ i = 1, \ldots, n$.
- La mediana de X_1 se obtiene minimizando:

$$J(a) = \sum_i |x_1(i) - a| = \sum_{i:x_1(i)>a} (x_1(i) - a) + \sum_{i:x_1(i)<a} (a - x_1(i)).$$

Regresión cuantílica (QR) práctica

- En la práctica debemos estimar los parámetros del modelo o estimar esas curvas de forma no-paramétrica usando QR a partir de una muestra de datos de (X_1, X_2): $(x_1(i), x_2(i)),\ i = 1, \dots, n$.
- La mediana de X_1 se obtiene minimizando:

$$J(a) = \sum_i |x_1(i) - a| = \sum_{i:x_1(i)>a} (x_1(i) - a) + \sum_{i:x_1(i)<a} (a - x_1(i)).$$

- La curva mediana lineal $m(x) = a + bx$ se obtiene minimizando

$$J(a,b) = \sum_{i:x_2(i)>a+bx_1(i)} (x_2(i) - a - bx_1(i)) + \sum_{i:x_2(i)<a+bx_1(i)} (a + bx_1(i) - x_2(i))$$

usando programación lineal (ver Koenker (2005)).

Regresión cuantílica (QR) práctica

- Análogamente el cuantil $\widehat{x}_q = F_{X_1}^{-1}(q)$ para $0 < q < 1$ de X_1 se obtiene minimizando:

$$J_q(a) = q \sum_{i:x_1(i)>a} (x_1(i) - a) + (1-q) \sum_{i:x_1(i)<a} (a - x_1(i)).$$

Regresión cuantílica (QR) práctica

- Análogamente el cuantil $\widehat{x}_q = F_{X_1}^{-1}(q)$ para $0 < q < 1$ de X_1 se obtiene minimizando:

$$J_q(a) = q \sum_{i:x_1(i)>a} (x_1(i) - a) + (1 - q) \sum_{i:x_1(i)<a} (a - x_1(i)).$$

- La curva q-cuantil lineal $m_q(x) = a_q + b_q x$ se obtiene minimizando en a y b:

$$q \sum_{i:x_2(i)>a+bx_1(i)} (x_2(i) - a - bx_1(i)) + (1 - q) \sum_{i:x_2(i)<a+bx_1(i)} (a + bx_1(i) - x_2(i))$$

usando programación lineal (ver Koenker (2005)).

Regresión cuantílica (QR) práctica

- El paquete `quantreg` de R (Rstudio) permite obtener estas soluciones.

Regresión cuantílica (QR) práctica

- El paquete `quantreg` de R (Rstudio) permite obtener estas soluciones.
- Si los datos están en `x1` y `x2` basta hacer:

```
rq(x2~x1)
plot(x1,x2)
abline(rq(x2~x1),col='red')
```

Regresión cuantílica (QR) práctica

- El paquete `quantreg` de R (Rstudio) permite obtener estas soluciones.
- Si los datos están en `x1` y `x2` basta hacer:

```
rq(x2~x1)
plot(x1,x2)
abline(rq(x2~x1),col='red')
```

- La banda de confianza centrada al 90 % se obtiene con `rq(x2~x1,0.05)` y `rq(x2~x1,0.95)`

Regresión cuantílica (QR) práctica

- El paquete `quantreg` de R (Rstudio) permite obtener estas soluciones.
- Si los datos están en `x1` y `x2` basta hacer:

```
rq(x2~x1)
plot(x1,x2)
abline(rq(x2~x1),col='red')
```

- La banda de confianza centrada al 90 % se obtiene con `rq(x2~x1,0.05)` y `rq(x2~x1,0.95)`
- Y la banda de confianza centrada al 50 % se obtiene con `rq(x2~x1,0.25)` y `rq(x2~x1,0.75)`

Ejemplos

Tiempos de vida residuales

- Tiempos de vida residuales. Si $X_1, \ldots, X_n$ son tiempos de vida, entonces

$$\mathrm{X}_t = (X_1 - t, \ldots, X_n - t | X_1 > t, \ldots, X_n > t)$$

tienen una distribución distorsionada a partir de las de $(X_i - t | X_i > t)$.

Tiempos de vida residuales

- Tiempos de vida residuales. Si $X_1, \dots, X_n$ son tiempos de vida, entonces

$$\mathrm{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$$

tienen una distribución distorsionada a partir de las de $(X_i - t | X_i > t)$.

- Estadísticos ordenados $IID \sim F$ (datos censurados)

Tiempos de vida residuales

- Tiempos de vida residuales. Si $X_1, \dots, X_n$ son tiempos de vida, entonces

$$\mathrm{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$$

tienen una distribución distorsionada a partir de las de $(X_i - t | X_i > t)$.
- Estadísticos ordenados $IID \sim F$ (datos censurados)
- Valores record $IID \sim F$, ver Navarro (2021).

Tiempos de vida residuales

- Tiempos de vida residuales. Si $X_1, \ldots, X_n$ son tiempos de vida, entonces

$$\mathrm{X}_t = (X_1 - t, \ldots, X_n - t | X_1 > t, \ldots, X_n > t)$$

tienen una distribución distorsionada a partir de las de $(X_i - t | X_i > t)$.
- Estadísticos ordenados $IID \sim F$ (datos censurados)
- Valores record $IID \sim F$, ver Navarro (2021).
- Sistemas coherentes.

Tiempos de vida residuales

- Tiempos de vida residuales. Si $X_1, \dots, X_n$ son tiempos de vida, entonces

$$\mathbf{X}_t = (X_1 - t, \dots, X_n - t | X_1 > t, \dots, X_n > t)$$

  tienen una distribución distorsionada a partir de las de $(X_i - t | X_i > t)$.
- Estadísticos ordenados $IID \sim F$ (datos censurados)
- Valores record $IID \sim F$, ver Navarro (2021).
- Sistemas coherentes.
- Datos pareados

Tiempos de vida de sistemas

- Tiempos de vida de sistemas coherentes (conectividad en redes) $T_1 < T_2$ basados en los mismos componentes $(X_1, \ldots, X_n)$.

Tiempos de vida de sistemas

- Tiempos de vida de sistemas coherentes (conectividad en redes) $T_1 < T_2$ basados en los mismos componentes $(X_1, \ldots, X_n)$.
- Prededecir T_2 a partir de $T_1 = t$ para $t > 0$.

Tiempos de vida de sistemas

- Tiempos de vida de sistemas coherentes (conectividad en redes) $T_1 < T_2$ basados en los mismos componentes $(X_1, \ldots, X_n)$.
- Prededecir T_2 a partir de $T_1 = t$ para $t > 0$.
- El caso más usual es $T_1 = \min(X_1, \ldots, X_n)$.

Tiempos de vida de sistemas

- Tiempos de vida de sistemas coherentes (conectividad en redes) $T_1 < T_2$ basados en los mismos componentes $(X_1, \dots, X_n)$.
- Prededecir T_2 a partir de $T_1 = t$ para $t > 0$.
- El caso más usual es $T_1 = \text{mín}(X_1, \dots, X_n)$.
- Cuando los componentes tienen la misma función distribución G, (T_1, T_2) tienen una distribución distorsionada basada en G.

Tiempos de vida de sistemas

- Tiempos de vida de sistemas coherentes (conectividad en redes) $T_1 < T_2$ basados en los mismos componentes $(X_1, \ldots, X_n)$.
- Prededecir T_2 a partir de $T_1 = t$ para $t > 0$.
- El caso más usual es $T_1 = \min(X_1, \ldots, X_n)$.
- Cuando los componentes tienen la misma función distribución G, (T_1, T_2) tienen una distribución distorsionada basada en G.
- Esto también se puede aplicar a los datos ordenados (censurados) en una muestra (son sitemas k-out-of-n).

Datos pareados

- Supongamos que X e Y son ID y tienen una distribución continua F. Entonces

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

Datos pareados

- Supongamos que X e Y son ID y tienen una distribución continua F. Entonces

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

- Incluso podemos suponer que C es simétrica, es decir, (X, Y) is intercambiable (EXC).

Datos pareados

- Supongamos que X e Y son ID y tienen una distribución continua F. Entonces

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

- Incluso podemos suponer que C es simétrica, es decir, (X, Y) is intercambiable (EXC).
- Supongamos que $L = \min(X, Y)$ es conocido y que queremos predecir $U = \max(X, Y)$.

Datos pareados

- Supongamos que X e Y son ID y tienen una distribución continua F. Entonces

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

- Incluso podemos suponer que C es simétrica, es decir, (X, Y) is intercambiable (EXC).
- Supongamos que $L = \min(X, Y)$ es conocido y que queremos predecir $U = \max(X, Y)$.
- Para eso necesitamos la distribución condicionada

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

- Queremos obtener una representación MDD para (L, U) usando F y C.

Datos pareados

- La distribución $\mathrm{G}(x,y)=\Pr(L\le x, U\le y)$ de (L,U) es

$$\mathrm{G}(x,y)=\begin{cases} C(F(y),F(y)) & y\le x;\\ C(F(x),F(y))+C(F(y),F(x))-C(F(x),F(x)) & y>x.\end{cases}$$

Datos pareados

- La distribución $G(x, y) = \Pr(L \leq x, U \leq y)$ de (L, U) es

$$G(x, y) = \begin{cases} C(F(y), F(y)) & y \leq x; \\ C(F(x), F(y)) + C(F(y), F(x)) - C(F(x), F(x)) & y > x. \end{cases}$$

- Entonces $G(x, y) = D(F(x), F(y))$ con

$$D(u, v) = \begin{cases} C(v, v) & \text{for} \quad v \leq u; \\ C(u, v) + C(v, u) - C(u, u) & \text{for} \quad u < v. \end{cases} \tag{2.1}$$

Datos pareados

- Las marginales de (L, U) son

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

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

donde

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

y

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

para $u, v \in [0, 1]$.

Datos pareados, caso IID

- Por ejemplo, si X e Y son IID, entonces

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

y

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

para todo $u \in (0,1)$.

Datos pareados, caso IID

- Por ejemplo, si X e Y son IID, entonces

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

y

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

para todo $u \in (0, 1)$.

- La distorsión es

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

Datos pareados, caso IID

- Por ejemplo, si X e Y son IID, entonces

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

y

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

para todo $u \in (0,1)$.

- La distorsión es

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

- La función D no es una cópula y las marginales G_1 y G_2 de (L,U) no se usan (usamos F).

Datos pareados, distribución condicionada

- De (1.8), la distribución de $(U|L = x)$ se puede poner como

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.3}$$

para $y \geq x$, donde

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

$$\partial_1 D(u, v) = \partial_1 C(u, v) + \partial_2 C(v, u) - \partial_1 C(u, u) - \partial_2 C(u, u), \text{ si } v > u.$$

Datos pareados, distribución condicionada

- De (1.8), la distribución de $(U|L = x)$ se puede poner como

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.3}$$

para $y \geq x$, donde

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

$$\partial_1 D(u, v) = \partial_1 C(u, v) + \partial_2 C(v, u) - \partial_1 C(u, u) - \partial_2 C(u, u), \text{ si } v > u.$$

- En el caso EXC, $\partial_1 D(u, v) = 2\partial_1 C(u, v) - 2\partial_1 C(u, u)$ para $u \leq v \leq 1$.

Datos pareados, distribución condicionada

- De (1.8), la distribución de $(U|L = x)$ se puede poner como

$$G_{2|1}(y|x) = D_{2|1}(F(y)|F(x)) \tag{2.3}$$

para $y \geq x$, donde

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

$$\partial_1 D(u, v) = \partial_1 C(u, v) + \partial_2 C(v, u) - \partial_1 C(u, u) - \partial_2 C(u, u), \text{ si } v > u.$$

- En el caso EXC, $\partial_1 D(u, v) = 2\partial_1 C(u, v) - 2\partial_1 C(u, u)$ para $u \leq v \leq 1$.
- En el caso IID, $\partial_1 D(u, v) = 2(v - u)$ para $u \leq v \leq 1$.

Predicciones

SOCIEDAD

Fernando Simón: "España no va a tener, como mucho, más allá de algún caso diagnosticado"

Curvas QR exactas. Caso IID.

- Sea (X_i, Y_i) una muestra de(X, Y) donde X, Y son IID$\sim F$.

Curvas QR exactas. Caso IID.

- Sea (X_i, Y_i) una muestra de(X, Y) donde X, Y son IID$\sim F$.
- Sean $L_i = \min(X_i, Y_i)$ y $U_i = \max(X_i, Y_i)$.

Curvas QR exactas. Caso IID.

- Sea (X_i, Y_i) una muestra de(X, Y) donde X, Y son IID$\sim F$.
- Sean $L_i = \min(X_i, Y_i)$ y $U_i = \max(X_i, Y_i)$.
- L_i y U_i son dependientes (pareados).

Curvas QR exactas. Caso IID.

- Sea (X_i, Y_i) una muestra de(X, Y) donde X, Y son IID$\sim F$.
- Sean $L_i = \min(X_i, Y_i)$ y $U_i = \max(X_i, Y_i)$.
- L_i y U_i son dependientes (pareados).
- De (2.3), la distribución de $(U|L = x)$ es

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

para $y \geq x$, donde

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

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

Datos pareados. Caso IID exponencial.

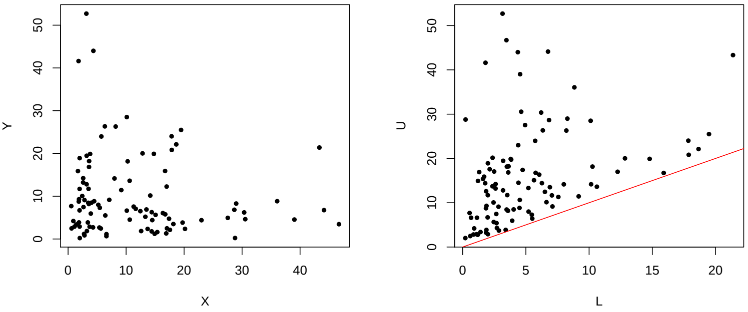

Figura: Datos de dos distribuciones exponenciales independientes de media $\mu = 10$ (izquierda) y datos pareados asociados (derecha).

Datos pareados. Caso IID exponencial.

- La función $F_{2|1}^{-1}$ se puede calcular como

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

para $0 < v < 1$, donde $D_{2|1}^{-1}(q|F(x)) = F(x) + q\bar{F}(x)$.

Datos pareados. Caso IID exponencial.

- La función $F_{2|1}^{-1}$ se puede calcular como

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

para $0 < v < 1$, donde $D_{2|1}^{-1}(q|F(x)) = F(x) + q\bar{F}(x)$.

- Si $\bar{F}(x) = \exp(-x/\mu)$, entonces $F^{-1}(y) = -\mu \log(1-y)$ y

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

Datos pareados. Caso IID exponencial.

- La función $F_{2|1}^{-1}$ se puede calcular como

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

para $0 < v < 1$, donde $D_{2|1}^{-1}(q|F(x)) = F(x) + q\bar{F}(x)$.

- Si $\bar{F}(x) = \exp(-x/\mu)$, entonces $F^{-1}(y) = -\mu \log(1-y)$ y

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

- Entonces la curva QR es

$$m(x) = x - \mu \log(0{,}5).$$

Datos pareados. Caso IID exponencial.

- La banda de confianza 90 % centrada es

$$[x - \mu \log(0{,}05), x - \mu \log(0{,}95)] \,.$$

Datos pareados. Caso IID exponencial.

- La banda de confianza 90 % centrada es

$$[x - \mu \log(0{,}05), x - \mu \log(0{,}95)] .$$

- La de confianza 50 % se obtiene de forma similar.

Datos pareados. Caso IID exponencial.

- La banda de confianza 90 % centrada es

$$[x - \mu \log(0{,}05), x - \mu \log(0{,}95)] \, .$$

- La de confianza 50 % se obtiene de forma similar.
- Aquí podemos preferir la banda inferior al 90 % dada por

$$[x, x - \mu \log(0{,}90)] \, .$$

Datos pareados. Caso IID exponencial.

- La banda de confianza 90 % centrada es

$$[x - \mu \log(0{,}05), x - \mu \log(0{,}95)] .$$

- La de confianza 50 % se obtiene de forma similar.
- Aquí podemos preferir la banda inferior al 90 % dada por

$$[x, x - \mu \log(0{,}90)] .$$

- Las podemos ver en las gráficas siguientes.

Datos pareados. Caso IID exponencial.

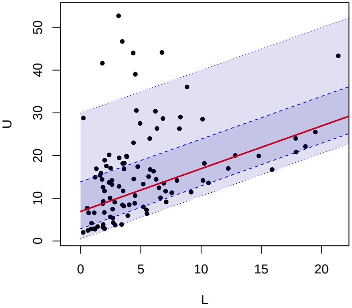

Figura: Curvas QR para los datos pareados (L, U) asociados a datos independientes (X, Y) con exponenciales de media $\mu = 10$ junto con las bandas de confianza 50 % y 90 % centradas (izquierda) e inferiores (derecha).

Predicciones

- La primera pareja de valores en nuestra muestra es $L_1 = 10{,}15771$ y $U_1 = 14{,}17195$.

Predicciones

- La primera pareja de valores en nuestra muestra es $L_1 = 10{,}15771$ y $U_1 = 14{,}17195$.
- La predicción de U_1 en L_1 es

$$m(L_1) = m(10{,}15771) = 10{,}15771 - \mu \log(0{,}5) = 17{,}08918.$$

Predicciones

- La primera pareja de valores en nuestra muestra es $L_1 = 10{,}15771$ y $U_1 = 14{,}17195$.
- La predicción de U_1 en L_1 es

$$m(L_1) = m(10{,}15771) = 10{,}15771 - \mu \log(0{,}5) = 17{,}08918.$$

- El intervalo de confianza 90 % centrado para esta predicción es $[10{,}67064, 40{,}11503]$.

Predicciones

- La primera pareja de valores en nuestra muestra es $L_1 = 10{,}15771$ y $U_1 = 14{,}17195$.
- La predicción de U_1 en L_1 es

$$m(L_1) = m(10{,}15771) = 10{,}15771 - \mu \log(0{,}5) = 17{,}08918.$$

- El intervalo de confianza 90 % centrado para esta predicción es $[10{,}67064, 40{,}11503]$.
- El del 50 % es $[13{,}03453, 24{,}02065]$.

Predicciones

- La primera pareja de valores en nuestra muestra es $L_1 = 10{,}15771$ y $U_1 = 14{,}17195$.
- La predicción de U_1 en L_1 es

$$m(L_1) = m(10{,}15771) = 10{,}15771 - \mu \log(0{,}5) = 17{,}08918.$$

- El intervalo de confianza 90 % centrado para esta predicción es $[10{,}67064, 40{,}11503]$.
- El del 50 % es $[13{,}03453, 24{,}02065]$.
- Ambos contienen al verdadero valor de U_1.

QR para datos dependientes EXC exponenciales

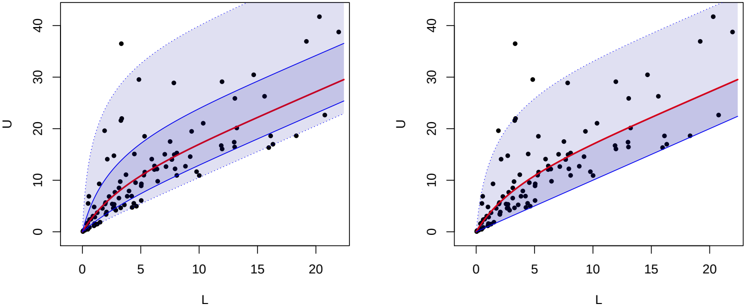

Figura: Curvas QR para los datos pareados (L, U) cuando (X, Y) son exponenciales y dependientes con una copula tipo Clayton con bandas de confianza centradas (izquierda) o inferiores (derecha).

Curvas QR estimadas, caso IID exponencial

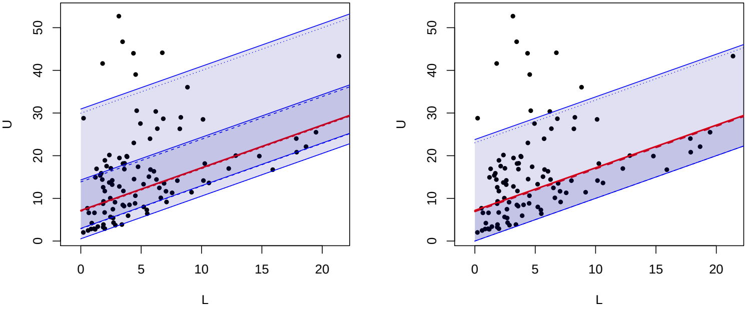

Figura: Curvas QR cuando se estiman las medias de las distribuciones exponenciales. Las líneas de puntos representan las curvas exactas.

Curvas QR estimadas, caso EXC-Clayton

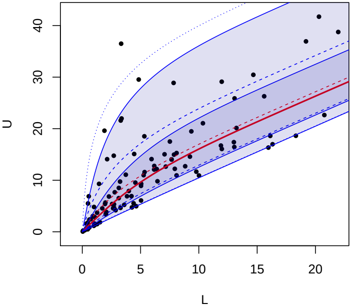

Figura: Curvas QR cuando se estiman las medias de las distribuciones exponenciales y el parámetro de la copula Clayton. Las líneas de puntos representan las curvas exactas.

Curvas QR no paramétricas, caso IID

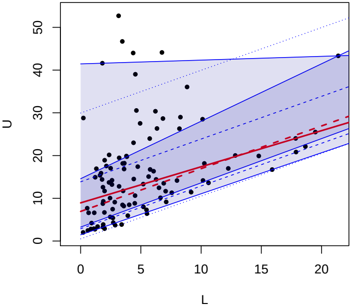

Figura: Curvas QR empíricas para los datos pareados (L, U) asociados a datos independientes (X, Y) exponenciales con medias $\mu = 10$ y grado 1 (izquierda) y 2 (derecha).

Curvas QR no paramétricas, caso EXC-Clayton

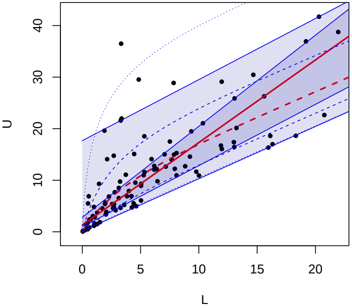

Figura: Curvas QR empíricas para los datos pareados (L, U) asociados a datos dependientes (X, Y) Clayton y exponenciales con medias $\mu = 10$ y grado 1 (izquierda) y 2 (derecha).

References

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

Koenker R, Bassett Jr G (1978). Regression quantiles. Econometrica 46, 33–50.

Miras, J. (1991). Deformación de funciones de distribución: una técnica estadística. Estadística Española Vol. 33, Núm. 127, 257–284.

Navarro, J. (2021). Prediction of record values by using quantile regression curves and distortion functions. orge Navarro. To appear in Metrika. Published online first Nov. 2021. DOI: 10.1007/s00184-021-00847-w.

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

Navarro J, Calì C, Longobardi M, Durante F (2021). Distortion Representations of Multivariate Distributions. Statistical Methods & Applications. Published online first Jan. 2022. DOI: 10.1007/s10260-021-00613-2.

Navarro, J. and del Águila, Y. (2017). Stochastic comparisons of distorted distributions, coherent systems and mixtures with ordered components. *Metrika* 80, 627-648

Navarro, J., del Águila, Y., Sordo, M.A. and Suárez-Llorens, A. (2013). Stochastic ordering properties for systems with dependent identically distributed components. *Applied Stochastic Models in Business and Industry* 29, 264–278.

Navarro, J., del Águila, Y., Sordo, M.A. and Suárez-Llorens, A. (2014). Preservation of reliability classes under the formation of coherent systems. *Applied Stochastic Models in Business and Industry* 30, 444–454.

Navarro, J., del Águila, Y., Sordo, M.A. and Suárez-Llorens, A. (2016). Preservation of stochastic orders under the formation of generalized distorted distributions. Applications to coherent systems. *Methodology and Computing in Applied Probability* 18, 529–545.

Navarro, J. and Gomis, M.C. (2016). Comparisons in the mean residual life order of coherent systems with identically distributed components. *Applied Stochastic Models in Business and Industry* 32, 33–47.

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

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

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

Fin

- **Las transparencias y más referencias se pueden ver en mi página web:**

$$https://webs.um.es/jorgenav$$

Fin

- **Las transparencias y más referencias se pueden ver en mi página web:**

$$https://webs.um.es/jorgenav$$

- **Gracias por su asistencias y atención!!**

Fin

- **Las transparencias y más referencias se pueden ver en mi página web:**

$$https://webs.um.es/jorgenav$$

- **Gracias por su asistencias y atención!!**
- **Preguntas?**