Estimadores basados en proyecciones para posición multivariada: su comportamiento en sesgo asintótico

Jorge G. Adrover[1,2], Victor J. Yohai[2,3]

[1]FaMAF, UNC, [2]CONICET, [3]Inst. de Calculo, FCEyN, UBA

Santa Fe, 4 de diciembre de 2009

Objetivo

Objetivo

Qué queremos?

Obtener estimadores para posición multivariada con el menor sesgo posible

Objetivo

Qué queremos?

Obtener estimadores para posición multivariada con el menor sesgo posible

Objetivo

Qué queremos?

Obtener estimadores para posición multivariada con el menor sesgo posible

Objetivo

La mediana tiene máximo sesgo asintotico (MSA) en vecindades de contaminacion de una distribucion simétrica unimodal que es menor que el MSA para cualquier otro estimador equivariante por traslación (Huber, 1964). Se conocen cotas inferiores para el MSA de cualquier estimador equivariante por transformaciones afines (He and Simpson, 1993). El estimador de proyecciones (Tyler, 1994) tiene MSA conocido (Adrover and Yohai, 2002, Zuo, Cui and Young, 2003) y permite una modificacion que conduce a un comportamiento casi minimax sobre contaminaciones puntuales.

Objetivo

Qué queremos?

Obtener estimadores para posición multivariada con el menor sesgo posible

Objetivo

La mediana tiene máximo sesgo asintotico (MSA) en vecindades de contaminacion de una distribucion simétrica unimodal que es menor que el MSA para cualquier otro estimador equivariante por traslación (Huber, 1964). Se conocen cotas inferiores para el MSA de cualquier estimador equivariante por transformaciones afines (He and Simpson, 1993). El estimador de proyecciones (Tyler, 1994) tiene MSA conocido (Adrover and Yohai, 2002, Zuo, Cui and Young, 2003) y permite una modificacion que conduce a un comportamiento casi minimax sobre contaminaciones puntuales.

Objetivo

Qué queremos?

Obtener estimadores para posición multivariada con el menor sesgo posible

Objetivo

La mediana tiene máximo sesgo asintotico (MSA) en vecindades de contaminacion de una distribucion simétrica unimodal que es menor que el MSA para cualquier otro estimador equivariante por traslación (Huber, 1964). Se conocen cotas inferiores para el MSA de cualquier estimador equivariante por transformaciones afines (He and Simpson, 1993). El estimador de proyecciones (Tyler, 1994) tiene MSA conocido (Adrover and Yohai, 2002, Zuo, Cui and Young, 2003) y permite una modificacion que conduce a un comportamiento casi minimax sobre contaminaciones puntuales.

Porqué?

Porque el sesgo domina el ECM del estimador en muestras grandes cuando nos apartamos del modelo central paramétrico

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = \mathit{varianza} + (\mathit{sesgo})^2$.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = \mathit{varianza} + (\mathit{sesgo})^2$.
- Estimación robusta para posición multivariada.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  - ☞ error cuadrático medio,
  - ☞ varianza,
  - ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = \mathit{varianza} + (\mathit{sesgo})^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T - \theta)^2 = \mathit{varianza} + (\mathit{sesgo})^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.
- Estimador de proyecciones multivariado (P-estimador)

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.
- Estimador de proyecciones multivariado (P-estimador)
- Propiedades de sesgo de este estimador.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  ☞ error cuadrático medio,
  ☞ varianza,
  ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.
- Estimador de proyecciones multivariado (P-estimador)
- Propiedades de sesgo de este estimador.
- Propuesta alternativa: mejorando el sesgo asintótico.

Lineamientos generales

- Tres ingredientes importantes mencionados:
  - ☞ error cuadrático medio,
  - ☞ varianza,
  - ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.
- Estimador de proyecciones multivariado (P-estimador)
- Propiedades de sesgo de este estimador.
- Propuesta alternativa: mejorando el sesgo asintótico.
- Comparación con otros estimadores..

Lineamientos generales

- Tres ingredientes importantes mencionados:
  - ☞ error cuadrático medio,
  - ☞ varianza,
  - ☞ sesgo.
- ☞ $ECM(T) = E(T-\theta)^2 = varianza + (sesgo)^2$.
- Estimación robusta para posición multivariada.
- Midiendo la robustez de un estimador: Sesgo asintótico.
- Mediana univariada: el mejor estimador respecto del sesgo asintótico.
- Estimando posición multivariada: distancia de Mahalanobis vs. seguimiento de proyecciones.
- Estimador de proyecciones multivariado (P-estimador)
- Propiedades de sesgo de este estimador.
- Propuesta alternativa: mejorando el sesgo asintótico.
- Comparación con otros estimadores..
- Conclusiones

Modelo de posición multivariado

- **Muestra aleatoria de una población normal multivariada en R^p,**

$$\boxed{\mathbf{X}_i = \boldsymbol{\theta} + \mathbf{U}_i, \quad \mathbf{U}_i \sim N(\mathbf{0}, \Sigma), i = 1, \ldots, n}$$

$$\boxed{\mathbf{X}_i \sim N(\boldsymbol{\theta}, \Sigma)}, i = 1, \ldots, n$$

donde la densidad de $\mathbf{X}_i$ está dada por

$$\boxed{f(\mathbf{x}, \boldsymbol{\theta}, \Sigma) = \frac{1}{\det(\Sigma)^{1/2}} \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta})\right)}$$

Modelo de posición multivariado

- **Muestra aleatoria de una población normal multivariada en R^p,**

$$\boxed{\mathbf{X}_i = \boldsymbol{\theta} + \mathbf{U}_i, \quad \mathbf{U}_i \sim N(\mathbf{0}, \Sigma), i = 1, \ldots, n}$$

$$\boxed{\mathbf{X}_i \sim N(\boldsymbol{\theta}, \Sigma)}, i = 1, \ldots, n$$

donde la densidad de $\mathbf{X}_i$ está dada por

$$\boxed{f(\mathbf{x}, \boldsymbol{\theta}, \Sigma) = \frac{1}{\det(\Sigma)^{1/2}} \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta})\right)}$$

- Problema: Estimar $(\boldsymbol{\theta}, \Sigma)$ desconocidos.

Modelo de posición multivariado

- **Muestra aleatoria de una población normal multivariada en** R^p,

$$\boxed{\mathbf{X}_i = \boldsymbol{\theta} + \mathbf{U}_i, \quad \mathbf{U}_i \sim N(\mathbf{0}, \Sigma), i = 1, ..., n}$$

$$\boxed{\mathbf{X}_i \sim N(\boldsymbol{\theta}, \Sigma)}, i = 1, ..., n$$

donde la densidad de $\mathbf{X}_i$ está dada por

$$\boxed{f(\mathbf{x}, \boldsymbol{\theta}, \Sigma) = \frac{1}{\det(\Sigma)^{1/2}} \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta})\right)}$$

- Problema: Estimar $(\boldsymbol{\theta}, \Sigma)$ desconocidos.
- **Estimadores clásicos (MLE)**

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i, \widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$$

Modelo de posición multivariado

- **Muestra aleatoria de una población normal multivariada en** R^p,

$$\boxed{\mathbf{X}_i = \boldsymbol{\theta} + \mathbf{U}_i, \quad \mathbf{U}_i \sim N(\mathbf{0}, \Sigma), i = 1, ..., n}$$

$$\boxed{\mathbf{X}_i \sim N(\boldsymbol{\theta}, \Sigma)}, i = 1, ..., n$$

donde la densidad de $\mathbf{X}_i$ está dada por

$$\boxed{f(\mathbf{x}, \boldsymbol{\theta}, \Sigma) = \frac{1}{\det(\Sigma)^{1/2}} \frac{1}{(2\pi)^{p/2}} \exp(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta}))}$$

- Problema: Estimar $(\boldsymbol{\theta}, \Sigma)$ desconocidos.
- **Estimadores clásicos (MLE)**

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i, \widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$$

- Ampliamos un poco la familia de errores,

$$\boxed{\mathbf{X}_i = \boldsymbol{\theta} + \mathbf{U}_i, \quad \mathbf{U}_i \sim E(\mathbf{0}, \Sigma), i = 1, ..., n}$$

donde la densidad de $\mathbf{X}_i$ está dada por

$$\boxed{f(\mathbf{x}, \boldsymbol{\theta}, \Sigma) = \frac{1}{\det(\Sigma)^{1/2}} g((\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta}))}$$

Modelo de posición multivariado - continuación

- ☞☞**PROBLEMAS**:

Modelo de posición multivariado - continuación

- ☞☞**PROBLEMAS**:
- ☞ SI LA DISTRIBUCION ELIPTICA NO TIENE MOMENTOS, LA MEDIA MUESTRAL Y LA COVARIANZA MUESTRAL NO PUEDEN APROXIMAR A $\boldsymbol{\theta}$ O A Σ!!!!!!!!!!!

Modelo de posición multivariado - continuación

- ☞☞**PROBLEMAS**:
- ☞ SI LA DISTRIBUCION ELIPTICA NO TIENE MOMENTOS, LA MEDIA MUESTRAL Y LA COVARIANZA MUESTRAL NO PUEDEN APROXIMAR A $\boldsymbol{\theta}$ O A Σ!!!!!!!!!!!
- ☞ AUNQUE LOS DATOS PROVINIERAN EN SU MAYORIA DE UNA POBLACION ELIPTICA, SI HUBIERA DATOS QUE NO SIGUEN EL MODELO ELIPTICO TAMPOCO PUEDO ESTIMAR $\boldsymbol{\theta}$ CONSISTENTEMENTE!!!!!

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??
- Qué queremos decir por **una distribución** F **diferente de** $F_{\boldsymbol{\theta},\Sigma}$?

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??
- Qué queremos decir por **una distribución** F **diferente de** $F_{\boldsymbol{\theta},\Sigma}$?
- $F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma}) = \{F = (1-\varepsilon)F_{\boldsymbol{\theta},\Sigma} + \varepsilon F^*$, F^* distribución en $R^p\}$

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??
- Qué queremos decir por **una distribución** F **diferente de** $F_{\boldsymbol{\theta},\Sigma}$?
- $F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma}) = \{F = (1-\varepsilon)F_{\boldsymbol{\theta},\Sigma} + \varepsilon F^*, F^*$ distribución en $R^p\}$
- Equivale a imaginar que lo que observamos es $\mathbf{Y}_i = (1-Z_i)\mathbf{X}_i + Z_i\mathbf{W}_i$ donde $Z_i \sim Bi(1,\epsilon)$, $\mathbf{X}_i \sim F_{\boldsymbol{\theta},\Sigma}$ y $\mathbf{W}_i \sim F^*$ (Z_i NO OBSERVADA), o sea

$$\boxed{\mathbf{Y}_i \sim F, F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma})}, i = 1, \ldots, n$$

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??
- Qué queremos decir por **una distribución** F **diferente de** $F_{\boldsymbol{\theta},\Sigma}$?
- $F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma}) = \{F = (1-\varepsilon)F_{\boldsymbol{\theta},\Sigma} + \varepsilon F^*, F^*$ distribución en $R^p\}$
- Equivale a imaginar que lo que observamos es $\mathbf{Y}_i = (1 - Z_i)\mathbf{X}_i + Z_i\mathbf{W}_i$ donde $Z_i \sim Bi(1,\epsilon)$, $\mathbf{X}_i \sim F_{\boldsymbol{\theta},\Sigma}$ y $\mathbf{W}_i \sim F^*$ (Z_i NO OBSERVADA), o sea

$$\boxed{\mathbf{Y}_i \sim F, F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma})}, i = 1, ..., n$$

- $\mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\circ})$ es lo que se llama $\varepsilon-$**vecindad de contaminación.**

Planteando el problema - continuación

- **Entonces** si el modelo paramétrico vale solo **aproximadamente,** esto es la distribución subyacente F está cercana a $F_{\boldsymbol{\theta},\Sigma}$, como estimamos a $\boldsymbol{\theta}$ y a Σ??
- Qué queremos decir por **una distribución** F **diferente de** $F_{\boldsymbol{\theta},\Sigma}$?
- $F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma}) = \{F = (1-\varepsilon)F_{\boldsymbol{\theta},\Sigma} + \varepsilon F^*, F^*$ distribución en $R^p\}$
- Equivale a imaginar que lo que observamos es $\mathbf{Y}_i = (1 - Z_i)\mathbf{X}_i + Z_i\mathbf{W}_i$ donde $Z_i \sim Bi(1, \epsilon)$, $\mathbf{X}_i \sim F_{\boldsymbol{\theta},\Sigma}$ y $\mathbf{W}_i \sim F^*$ (Z_i NO OBSERVADA), o sea

$$\boxed{\mathbf{Y}_i \sim F, F \in \mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\Sigma})}, i = 1, ..., n$$

- $\mathcal{F}_\epsilon(F_{\boldsymbol{\theta},\circ})$ es lo que se llama $\varepsilon-$**vecindad de contaminación.**
- Una pregunta clave para cualquier estimador $\mathbf{T}(.)$

Cuan lejos se ´´ubica´´ el estimador $\mathbf{T}(.)$

del desconocido parámetro $\boldsymbol{\theta}$

cuando la verdadera distribución subyacente F

es diferente de la distribución del modelo

paramétrico????

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE ``GRANDE'' } \mathcal{F} \text{ DE DISTRIBUCIONES}}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE “GRANDE” } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, \ldots, t_p) = \frac{1}{n} \sum I_{(-\infty, t_1] \times \cdots \times (-\infty, t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, \ldots \mathbf{Y}_n) = \mathbf{T}(F_n)}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE "GRANDE" } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, ..., t_p) = \frac{1}{n} \sum I_{(-\infty,t_1]\times\cdots\times(-\infty,t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, ...\mathbf{Y}_n) = \mathbf{T}(F_n)}$
- En general $\boxed{F_n \approx F}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE “GRANDE” } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, \ldots, t_p) = \frac{1}{n} \sum I_{(-\infty, t_1] \times \cdots \times (-\infty, t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, \ldots \mathbf{Y}_n) = \mathbf{T}(F_n)}$
- En general $\boxed{F_n \approx F}$
- $\boxed{\mathbf{T}(F_n) \approx \mathbf{T}(F) \quad \text{si la m.a. tiene distribucion F (CONTINUIDAD)}}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE "GRANDE" } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, ..., t_p) = \frac{1}{n}\sum I_{(-\infty,t_1]\times\cdots\times(-\infty,t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, ...\mathbf{Y}_n) = \mathbf{T}(F_n)}$
- En general $\boxed{F_n \approx F}$
- $\boxed{\mathbf{T}(F_n) \approx \mathbf{T}(F) \quad \text{si la m.a. tiene distribucion F (CONTINUIDAD)}}$
- $\boxed{\text{PERO } \mathbf{T}(F) \neq\neq \mathbf{T}(F_{\theta,\Sigma}) \text{ (SESGO!!!!!!)}}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE “GRANDE” } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, ..., t_p) = \frac{1}{n} \sum I_{(-\infty,t_1]\times\cdots\times(-\infty,t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, ...\mathbf{Y}_n) = \mathbf{T}(F_n)}$
- En general $\boxed{F_n \approx F}$
- $\boxed{\mathbf{T}(F_n) \approx \mathbf{T}(F) \quad \text{si la m.a. tiene distribucion F (CONTINUIDAD)}}$
- $\boxed{\text{PERO } \mathbf{T}(F) \neq\neq \mathbf{T}(F_{\boldsymbol{\theta},\Sigma}) \text{ (SESGO!!!!!!)}}$
- $\boxed{\mathbf{T}(F_{\boldsymbol{\theta},\Sigma}) = \boldsymbol{\theta} \text{ (CONSISTENCIA FISHER)}}$

Qué buscamos??

- $\boxed{\text{QUE ESTE DEFINIDO EN UNA CLASE “GRANDE” } \mathcal{F} \text{ DE DISTRIBUCIONES}}$
- Y que la clase $\mathcal{F}$ contenga a la función de distribución empírica: $\boxed{F_n(t_1, \ldots, t_p) = \frac{1}{n}\sum I_{(-\infty,t_1]\times\cdots\times(-\infty,t_p]}(\mathbf{Y}_i) : \mathbf{T}(\mathbf{Y}_1, \ldots \mathbf{Y}_n) = \mathbf{T}(F_n)}$
- En general $\boxed{F_n \approx F}$
- $\boxed{\mathbf{T}(F_n) \approx \mathbf{T}(F) \quad \text{si la m.a. tiene distribucion F (CONTINUIDAD)}}$
- $\boxed{\text{PERO } \mathbf{T}(F) \neq\neq \mathbf{T}(F_{\boldsymbol{\theta},\Sigma}) \text{ (SESGO!!!!!!)}}$
- $\boxed{\mathbf{T}(F_{\boldsymbol{\theta},\Sigma}) = \boldsymbol{\theta} \text{ (CONSISTENCIA FISHER)}}$
- $\boxed{\mathbf{T}(F_n) \neq\neq \boldsymbol{\theta} \text{ (SE PODRA “ACHICAR”???)}}$

Qué queremos de nuestro estimador T?

- En el modelo elíptico hay *equivariancia afín*, esto es, si $\mathbf{X} \sim E(\boldsymbol{\theta},\Sigma)$, y el vector cambia por una transformación afín, $\mathbf{Y} = A(X - \zeta)$, entonces

$$\boxed{\mathbf{Y} \sim E(A(\boldsymbol{\theta} - \zeta), A\Sigma A')}$$

Qué queremos de nuestro estimador T?

- En el modelo elíptico hay *equivariancia afín*, esto es, si $\mathbf{X} \sim E(\boldsymbol{\theta},\Sigma)$, y el vector cambia por una transformación afín, $\mathbf{Y} = A(X - \zeta)$, entonces

$$\boxed{\mathbf{Y} \sim E(A(\boldsymbol{\theta} - \zeta), A\Sigma A')}$$

- Entonces en general le pediremos a los estimadores que cumplan la propiedad de equivariancia afín, si A es una matriz no singular, y $\zeta \in \mathbf{R}^p$,

$$\boxed{\mathbf{T}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A(\mathbf{T}(\mathcal{D}(X)) - \zeta)}$$

$$\boxed{\widehat{\Sigma}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A\widehat{\Sigma}(\mathcal{D}(X))A'}$$

Qué queremos de nuestro estimador T?

- En el modelo elíptico hay *equivariancia afín*, esto es, si $\mathbf{X} \sim E(\boldsymbol{\theta},\Sigma)$, y el vector cambia por una transformación afín, $\mathbf{Y} = A(X - \zeta)$, entonces

$$\boxed{\mathbf{Y} \sim E(A(\boldsymbol{\theta} - \zeta), A\Sigma A')}$$

- Entonces en general le pediremos a los estimadores que cumplan la propiedad de equivariancia afín, si A es una matriz no singular, y $\zeta \in \mathbf{R}^p$,

$$\boxed{\mathbf{T}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A(\mathbf{T}(\mathcal{D}(X)) - \zeta)}$$

$$\boxed{\widehat{\Sigma}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A\widehat{\Sigma}(\mathcal{D}(X))A'}$$

- Si $\mathbf{X} \sim F_{\boldsymbol{\theta},\Sigma}$ entonces $\mathbf{X} \longrightarrow \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\theta})$ y $\mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma}) \longrightarrow \mathcal{F}_\varepsilon(F_{\mathbf{0},I})$

Qué queremos de nuestro estimador T?

- En el modelo elíptico hay *equivariancia afín*, esto es, si $\mathbf{X} \sim E(\boldsymbol{\theta},\Sigma)$, y el vector cambia por una transformación afín, $\mathbf{Y} = A(X - \boldsymbol{\zeta})$, entonces

$$\boxed{\mathbf{Y} \sim E(A(\boldsymbol{\theta} - \boldsymbol{\zeta}), A\Sigma A')}$$

- Entonces en general le pediremos a los estimadores que cumplan la propiedad de equivariancia afín, si A es una matriz no singular, y $\boldsymbol{\zeta} \in \mathbf{R}^p$,

$$\boxed{\mathbf{T}(\mathcal{D}(A(\mathbf{X} - \boldsymbol{\zeta})) = A(\mathbf{T}(\mathcal{D}(X)) - \boldsymbol{\zeta})}$$

$$\boxed{\widehat{\Sigma}(\mathcal{D}(A(\mathbf{X} - \boldsymbol{\zeta})) = A\widehat{\Sigma}(\mathcal{D}(X))A'}$$

- Si $\mathbf{X} \sim F_{\boldsymbol{\theta},\Sigma}$ entonces $\mathbf{X} \longrightarrow \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\theta})$ y $\mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma}) \longrightarrow \mathcal{F}_\varepsilon(F_{\mathbf{0},I})$
- QUE MAS? **La clase de distribuciones $\mathcal{F}$ donde vive F**

Qué queremos de nuestro estimador T?

- En el modelo elíptico hay *equivariancia afín*, esto es, si $\mathbf{X} \sim E(\boldsymbol{\theta},\Sigma)$, y el vector cambia por una transformación afín, $\mathbf{Y} = A(X - \zeta)$, entonces

$$\boxed{\mathbf{Y} \sim E(A(\boldsymbol{\theta} - \zeta), A\Sigma A')}$$

- Entonces en general le pediremos a los estimadores que cumplan la propiedad de equivariancia afín, si A es una matriz no singular, y $\boldsymbol{\zeta} \in \mathbf{R}^p$,

$$\boxed{\mathbf{T}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A(\mathbf{T}(\mathcal{D}(X)) - \zeta)}$$

$$\boxed{\widehat{\Sigma}(\mathcal{D}(A(\mathbf{X} - \zeta)) = A\widehat{\Sigma}(\mathcal{D}(X))A'}$$

- Si $\mathbf{X} \sim F_{\boldsymbol{\theta},\Sigma}$ entonces $\mathbf{X} \longrightarrow \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\theta})$ y $\mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma}) \longrightarrow \mathcal{F}_\varepsilon(F_{\mathbf{0},I})$
- QUE MAS? **La clase de distribuciones $\mathcal{F}$ donde vive F**
- QUE MAS? **La medida de lejanía para cuantificar la diferencia entre el estimador y el parámetro**

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES
- $\boxed{b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) = (\mathbf{T}(F) - \boldsymbol{\theta})' \left(\Lambda(F_{\boldsymbol{\theta},\Sigma})\right)^{-1} (\mathbf{T}(F) - \boldsymbol{\theta})}$

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES
- $\boxed{b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) = (\mathbf{T}(F) - \boldsymbol{\theta})' \left(\Lambda(F_{\boldsymbol{\theta},\Sigma})\right)^{-1} (\mathbf{T}(F) - \boldsymbol{\theta})}$
- donde $\Lambda(.)$ es una matriz de dispersion afinmente equivariante, y $\Lambda(F_{\mathbf{0},I}) = cI$.

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES
- $\boxed{b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) = (\mathbf{T}(F) - \boldsymbol{\theta})' \left(\Lambda(F_{\boldsymbol{\theta},\Sigma})\right)^{-1} (\mathbf{T}(F) - \boldsymbol{\theta})}$
- donde $\Lambda(.)$ es una matriz de dispersion afinmente equivariante, y $\Lambda(F_{\mathbf{0},I}) = cI$.
- Qué ocurre ahora

$$\begin{aligned}
\boxed{B^2(\mathbf{T},\varepsilon, F_{\boldsymbol{\theta}.\Sigma})} &= \sup_{F\in\mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma})} b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) \\
&= \sup_{F\in\mathcal{F}_\varepsilon(F_{0,I})} b^2(\mathbf{T},\varepsilon, F, F_{\mathbf{0},I}) = \boxed{B^2(\mathbf{T},\varepsilon, F_{0.I})}
\end{aligned}$$

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES
- $\boxed{b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) = (\mathbf{T}(F) - \boldsymbol{\theta})' (\Lambda(F_{\boldsymbol{\theta},\Sigma}))^{-1} (\mathbf{T}(F) - \boldsymbol{\theta})}$
- donde $\Lambda(.)$ es una matriz de dispersion afinmente equivariante, y $\Lambda(F_{\mathbf{0},I}) = cI$.
- Qué ocurre ahora

$$\begin{aligned}\boxed{B^2(\mathbf{T},\varepsilon, F_{\boldsymbol{\theta}.\Sigma})} &= \sup_{F\in\mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma})} b^2(\mathbf{T},\varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) \\ &= \sup_{F\in\mathcal{F}_\varepsilon(F_{0,I})} b^2(\mathbf{T},\varepsilon, F, F_{\mathbf{0},I}) = \boxed{B^2(\mathbf{T},\varepsilon, F_{0.I})}\end{aligned}$$

- *B^2 no es facil de calcular*, hay cantidades relacionadas que dicen acerca de la robustez del estimador

$\boxed{\varepsilon^*(\mathbf{T},F_{0.I}) = \inf\{\varepsilon > 0 : B^2(\mathbf{T},\varepsilon, F_{0.I}) = \infty\}}$ PUNTO DE RUPTURA

$\boxed{\gamma^*(\mathbf{T},F_{\mathbf{0}.I}) = \left.\frac{dB(\mathbf{T},\varepsilon,F_{\mathbf{0},I})}{d\varepsilon}\right|_{\varepsilon=0}}$ SENSIBILIDAD DE CONTAMINACION

Sesgo asintótico máximo

- Como medirlo?? INVARIANTE POR TRANSFORMACIONES AFINES
- $\boxed{b^2(\mathbf{T}, \varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) = (\mathbf{T}(F) - \boldsymbol{\theta})' \left(\Lambda(F_{\boldsymbol{\theta},\Sigma})\right)^{-1} (\mathbf{T}(F) - \boldsymbol{\theta})}$
- donde $\Lambda(.)$ es una matriz de dispersion afinmente equivariante, y $\Lambda(F_{\mathbf{0},I}) = cI$.
- Qué ocurre ahora

$$
\begin{aligned}
\boxed{B^2(\mathbf{T}, \varepsilon, F_{\boldsymbol{\theta}.\Sigma})} &= \sup_{F \in \mathcal{F}_\varepsilon(F_{\boldsymbol{\theta},\Sigma})} b^2(\mathbf{T}, \varepsilon, F, F_{\boldsymbol{\theta},\Sigma}) \\
&= \sup_{F \in \mathcal{F}_\varepsilon(F_{0,I})} b^2(\mathbf{T}, \varepsilon, F, F_{\mathbf{0},I}) = \boxed{B^2(\mathbf{T}, \varepsilon, F_{0.I})}
\end{aligned}
$$

- B^2 *no es facil de calcular*, hay cantidades relacionadas que dicen acerca de la robustez del estimador

$$\boxed{\varepsilon^*(\mathbf{T}, F_{0.I}) = \inf\{\varepsilon > 0 : B^2(\mathbf{T}, \varepsilon, F_{0.I}) = \infty\}} \text{ PUNTO DE RUPTURA}$$

$$\boxed{\gamma^*(\mathbf{T}, F_{\mathbf{0}.I}) = \left.\frac{dB(\mathbf{T}, \varepsilon, F_{\mathbf{0},I})}{d\varepsilon}\right|_{\varepsilon=0}} \text{SENSIBILIDAD DE CONTAMINACION}$$

- Estos dos conceptos miran el comportamiento de la curva de sesgo en los extremos de su intervalo de definición.

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$

Estimadores basados en distancia de Mahalanobis

- ESTIMADOR DE MAXIMA VEROSIMILITUD
- $(\mathbf{T}, \widehat{\Sigma}) = \arg\min_{\mathbf{t} \in \mathbf{R}^p, V>0} \frac{1}{n} \sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i, \mathbf{t}, V)) + \ln \det(V)$

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.

Estimadores basados en distancia de Mahalanobis

- ESTIMADOR DE MAXIMA VEROSIMILITUD
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.
- M-ESTIMADORES (Maronna, 1974)

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} - \ln f(d^2(\mathbf{x}_i, \mathbf{t}, V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.
- $\boxed{\text{M-ESTIMADORES (Maronna, 1974)}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} \rho(d^2(\mathbf{x}_i, \mathbf{t}, V)) + \ln\det(V)$

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.
- $\boxed{\text{M-ESTIMADORES (Maronna, 1974)}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} \rho(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T}, \widehat{\Sigma}) = \arg\min_{\mathbf{t} \in \mathbf{R}^p, V>0} \frac{1}{n} \sum_{i=1}^{n} - \ln f(d^2(\mathbf{x}_i, \mathbf{t}, V)) + \ln \det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.
- $\boxed{\text{M-ESTIMADORES (Maronna, 1974)}}$
- $(\mathbf{T}, \widehat{\Sigma}) = \arg\min_{\mathbf{t} \in \mathbf{R}^p, V>0} \frac{1}{n} \sum_{i=1}^{n} \rho(d^2(\mathbf{x}_i, \mathbf{t}, V)) + \ln \det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$

Estimadores basados en distancia de Mahalanobis

- $\boxed{\text{ESTIMADOR DE MAXIMA VEROSIMILITUD}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} -\ln f(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$.
- ▶ PUNTO DE RUPTURA: 0.
- $\boxed{\text{M-ESTIMADORES (Maronna, 1974)}}$
- $(\mathbf{T},\widehat{\Sigma}) = \arg\min_{\mathbf{t}\in\mathbf{R}^p, V>0} \frac{1}{n}\sum_{i=1}^{n} \rho(d^2(\mathbf{x}_i,\mathbf{t},V)) + \ln\det(V)$
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: menor o igual que $1/(p+1)$.

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}\left(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V)\right)$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*)) V^*$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), ..., d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), ..., d(\mathbf{x}_n, \mathbf{t}, V^*)) V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*)) V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}\left(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V)\right)$
- $\widehat{V} = C_p q_{0.5}\left(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*)\right) V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), ..., d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), ..., d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in \mathcal{C}_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), ..., d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), ..., d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in \mathcal{C}_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$
- $\mathcal{C}_{0.5}$ contiene a los subconjuntos con 50% de las observaciones,

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), ..., d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), ..., d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in C_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$
- $C_{0.5}$ contiene a los subconjuntos con 50% de las observaciones,
- $\overline{\mathbf{x}}(C)$ es el promedio de las observaciones en el conjunto C.

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*)) V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in \mathcal{C}_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$
- $\mathcal{C}_{0.5}$ contiene a los subconjuntos con 50% de las observaciones,
- $\overline{\mathbf{x}}(C)$ es el promedio de las observaciones en el conjunto C.
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), \ldots, d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), \ldots, d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in \mathcal{C}_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$
- $\mathcal{C}_{0.5}$ contiene a los subconjuntos con 50% de las observaciones,
- $\overline{\mathbf{x}}(C)$ es el promedio de las observaciones en el conjunto C.
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$

De varios estimadores - continuacion

- $\boxed{\text{Elipsoide de Minimo Volumen (Rousseeuw, 1983)}}$
- $(\mathbf{T}, V^*) = \arg\min_{\mathbf{t}, V>0, |V|=1} q_{0.5}(d^2(\mathbf{x}_1, \mathbf{t}, V), ..., d^2(\mathbf{x}_n, \mathbf{t}, V))$
- $\widehat{V} = C_p q_{0.5}(d(\mathbf{x}_1, \mathbf{t}, V^*), ..., d(\mathbf{x}_n, \mathbf{t}, V^*))V^*$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Kim and Pollard, 1990)
- ▶ TASA DE CONVERGENCIA: $n^{1/3}$
- ▶ PUNTO DE RUPTURA: 1/2.
- $\boxed{\text{Estimador de determinante de covarianza minimo (Rousseeuw, 1984)}}$
- $\widehat{C} = \arg\min_{C \in \mathcal{C}_{0.5}} \left| \frac{1}{n} \sum_{i \in C} (\mathbf{x}_i - \overline{\mathbf{x}}(C))(\mathbf{x}_i - \overline{\mathbf{x}}(C))' \right|$
- $\mathcal{C}_{0.5}$ contiene a los subconjuntos con 50% de las observaciones,
- $\overline{\mathbf{x}}(C)$ es el promedio de las observaciones en el conjunto C.
- ▶ ESTIMADOR ASINTOTICAMENTE NORMAL
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.

Estimadores basados en proyecciones

- IDEA: LLEVAR EL PROBLEMA MULTIVARIADO A MUCHOS PROBLEMAS UNIVARIADOS

Estimadores basados en proyecciones

- IDEA: LLEVAR EL PROBLEMA MULTIVARIADO A MUCHOS PROBLEMAS UNIVARIADOS
- *Porqué la mediana tiene un comportamiento relevante*? En el modelo $y = \theta + \varepsilon$, si T es cualquier estimador equivariante por traslaciones,

$$\min_{T} B(T, \varepsilon, F_0) = B(med, \varepsilon, F_0).$$

Estimadores basados en proyecciones

- IDEA: LLEVAR EL PROBLEMA MULTIVARIADO A MUCHOS PROBLEMAS UNIVARIADOS
- *Porqué la mediana tiene un comportamiento relevante*? En el modelo $y = \theta + \varepsilon$, si T es cualquier estimador equivariante por traslaciones,

$$\min_{T} B(T, \varepsilon, F_0) = B(med, \varepsilon, F_0).$$

- Se proyectan las observaciones centradas sobre toda direccion y se toma una medida de posición estandarizada para medir la lejania causada por $\boldsymbol{\zeta}$,

$$v(\boldsymbol{\zeta}, F_n) = \sup_{\mathbf{a} \in R^p - \{\mathbf{0}\}} \left| \frac{T_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), \ldots, \mathbf{a}'(\mathbf{x}_n - \zeta))}{S_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), \ldots, \mathbf{a}'(\mathbf{x}_n - \zeta))} \right|$$

donde $T_n(z_1, \ldots, z_n) = med(z_1, \ldots, z_n)$ $(mean(z_1, \ldots, z_n))$y
$S_n^2(z_1, \ldots, z_n) = med(|z_i - med(z_1, \ldots, z_n)|^2)$, $(mean(|z_i - mean(z_1, \ldots, z_n)|^2)$

Estimadores basados en proyecciones (continuación)

- Medias ponderadas con pesos basados en proyecciones (Stahel (1981), Donoho(1982

Estimadores basados en proyecciones (continuación)

- Medias ponderadas con pesos basados en proyecciones (Stahel (1981), Donoho(1982
- Si $w_{k,j} = W_k(v(\mathbf{x}_j, F_n))$, $k = 1, 2$, $W_k \geq 0$ tal que $\lim_{k\to\infty} W_k(t) = 0$

$$\widehat{\mathbf{T}} = \frac{\sum_{i=1}^{n} w_{1,j}\mathbf{x}_j}{\sum_{i=1}^{n} w_{1,j}} \qquad \widehat{V} = \frac{\sum_{i=1}^{n} w_{2,j}(\mathbf{x}_j - \widehat{\mathbf{T}})(\mathbf{x}_j - \widehat{\mathbf{T}})'}{\sum_{i=1}^{n} w_{2,j}}$$

Estimadores basados en proyecciones (continuación)

- Medias ponderadas con pesos basados en proyecciones (Stahel (1981), Donoho(1982
- Si $w_{k,j} = W_k(v(\mathbf{x}_j, F_n))$, $k = 1, 2$, $W_k \geq 0$ tal que $\lim_{k\to\infty} W_k(t) = 0$

$$\widehat{\mathbf{T}} = \frac{\sum_{i=1}^{n} w_{1,j}\mathbf{x}_j}{\sum_{i=1}^{n} w_{1,j}} \qquad \widehat{V} = \frac{\sum_{i=1}^{n} w_{2,j}(\mathbf{x}_j - \widehat{\mathbf{T}})(\mathbf{x}_j - \widehat{\mathbf{T}})'}{\sum_{i=1}^{n} w_{2,j}}$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999 and Zuo,Cui and He, 2004)

Estimadores basados en proyecciones (continuación)

- $\boxed{\text{Medias ponderadas con pesos basados en proyecciones (Stahel (1981), Donoho(1982}}$
- Si $w_{k,j} = W_k(v(\mathbf{x}_j, F_n))$, $k = 1, 2$, $W_k \geq 0$ tal que $\lim_{k\to\infty} W_k(t) = 0$

$$\widehat{\mathbf{T}} = \frac{\sum_{i=1}^{n} w_{1,j}\mathbf{x}_j}{\sum_{i=1}^{n} w_{1,j}} \quad \widehat{V} = \frac{\sum_{i=1}^{n} w_{2,j}(\mathbf{x}_j - \widehat{\mathbf{T}})(\mathbf{x}_j - \widehat{\mathbf{T}})'}{\sum_{i=1}^{n} w_{2,j}}$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999 and Zuo,Cui and He, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$

Estimadores basados en proyecciones (continuación)

- Medias ponderadas con pesos basados en proyecciones (Stahel (1981), Donoho(1982
- Si $w_{k,j} = W_k(v(\mathbf{x}_j, F_n))$, $k = 1, 2$, $W_k \geq 0$ tal que $\lim_{k\to\infty} W_k(t) = 0$

$$\widehat{\mathbf{T}} = \frac{\sum_{i=1}^{n} w_{1,j}\mathbf{x}_j}{\sum_{i=1}^{n} w_{1,j}} \qquad \widehat{V} = \frac{\sum_{i=1}^{n} w_{2,j}(\mathbf{x}_j - \widehat{\mathbf{T}})(\mathbf{x}_j - \widehat{\mathbf{T}})'}{\sum_{i=1}^{n} w_{2,j}}$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999 and Zuo,Cui and He, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.
- Hay P-estimadores para dispersion multivariada que usan otra medida de lejanía para las matrices (Maronna, Stahel and Yohai, 1992)

Estimadores basados en proyecciones - continuacion

- P-estimadores para posicion multivariada (Tyler, 1994)
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.
- Hay P-estimadores para dispersion multivariada que usan otra medida de lejanía para las matrices (Maronna, Stahel and Yohai, 1992)
- Mediana de Tukey (Tukey, 1974)

$$
\begin{aligned}
v(\zeta, F_n) &= \sup_{\mathbf{a} \neq \mathbf{0}, \eta \in R^p} F_n(|\mathbf{a}'(\mathbf{X} - \zeta)| > |\mathbf{a}'(\mathbf{X} - \eta)|) \\
\widehat{\mathbf{T}} &= \arg\min_{\zeta} v(\zeta, F_n)
\end{aligned}
$$

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.
- Hay P-estimadores para dispersion multivariada que usan otra medida de lejanía para las matrices (Maronna, Stahel and Yohai, 1992)
- $\boxed{\text{Mediana de Tukey (Tukey, 1974)}}$

$$\begin{aligned} v(\zeta, F_n) &= \sup_{\mathbf{a} \neq \mathbf{0}, \eta \in R^p} F_n(|\mathbf{a}'(\mathbf{X} - \zeta)| > |\mathbf{a}'(\mathbf{X} - \eta)|) \\ \widehat{\mathbf{T}} &= \arg\min_{\zeta} v(\zeta, F_n) \end{aligned}$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999, Nolan, 1998, Masse, 2002)

Estimadores basados en proyecciones - continuacion

- $\boxed{\text{P-estimadores para posicion multivariada (Tyler, 1994)}}$
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.
- Hay P-estimadores para dispersion multivariada que usan otra medida de lejanía para las matrices (Maronna, Stahel and Yohai, 1992)
- $\boxed{\text{Mediana de Tukey (Tukey, 1974)}}$

$$
\begin{aligned}
v(\zeta, F_n) &= \sup_{\mathbf{a} \neq \mathbf{0}, \eta \in R^p} F_n(|\mathbf{a}'(\mathbf{X} - \zeta)| > |\mathbf{a}'(\mathbf{X} - \eta)|) \\
\widehat{\mathbf{T}} &= \arg\min_{\zeta} v(\zeta, F_n)
\end{aligned}
$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999, Nolan, 1998, Masse, 2002)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$

Estimadores basados en proyecciones - continuacion

- P-estimadores para posicion multivariada (Tyler, 1994)
- $\widehat{\mathbf{T}}_{MP} = \arg\min_{\zeta \in R^p} v(\zeta, F_n)$
- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Zuo, 2004)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/2.
- Hay P-estimadores para dispersion multivariada que usan otra medida de lejanía para las matrices (Maronna, Stahel and Yohai, 1992)
- Mediana de Tukey (Tukey, 1974)

$$v(\zeta, F_n) = \sup_{\mathbf{a} \neq \mathbf{0}, \eta \in R^p} F_n(|\mathbf{a}'(\mathbf{X} - \zeta)| > |\mathbf{a}'(\mathbf{X} - \eta)|)$$

$$\widehat{\mathbf{T}} = \arg\min_{\zeta} v(\zeta, F_n)$$

- ▶ ESTIMADOR ASINTOTICAMENTE NO-NORMAL (Bai and He, 1999, Nolan, 1998, Masse, 2002)
- ▶ TASA DE CONVERGENCIA: $n^{1/2}$
- ▶ PUNTO DE RUPTURA: 1/3.

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.
- El sesgo máximo de la mediana en el entorno de contaminacion es $d_1 = B(med, \varepsilon, L_o) = L_0^{-1}(1/(2(1-\varepsilon))$

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.
- El sesgo máximo de la mediana en el entorno de contaminacion es $d_1 = B(med, \varepsilon, L_o) = L_0^{-1}(1/(2(1-\varepsilon))$
- Mediana de Tukey (Chen and Tyler, 2002)

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.
- El sesgo máximo de la mediana en el entorno de contaminacion es $d_1 = B(med, \varepsilon, L_o) = L_0^{-1}(1/(2(1-\varepsilon))$
- $\boxed{\text{Mediana de Tukey (Chen and Tyler, 2002)}}$
- $\boxed{\text{P-estimadores para posicion multivariada (Adrover and Yohai, 2002)}}$

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.
- El sesgo máximo de la mediana en el entorno de contaminacion es $d_1 = B(med, \varepsilon, L_o) = L_0^{-1}(1/(2(1-\varepsilon))$
- Mediana de Tukey (Chen and Tyler, 2002)
- P-estimadores para posicion multivariada (Adrover and Yohai, 2002)
- Medias con pesos basados en proyecciones (Zuo, Cui and Young, 2004)

Y el sesgo??

- No se conoce en general demasiado sobre máximo sesgo de estimadores.
- El sesgo máximo de la mediana en el entorno de contaminacion es $d_1 = B(med, \varepsilon, L_o) = L_0^{-1}(1/(2(1-\varepsilon))$
- $\boxed{\text{Mediana de Tukey (Chen and Tyler, 2002)}}$
- $\boxed{\text{P-estimadores para posicion multivariada (Adrover and Yohai, 2002)}}$
- $\boxed{\text{Medias con pesos basados en proyecciones (Zuo, Cui and Young, 2004)}}$
- Sean d_2 y d_3 el mayor y menor valor que toma $med(|\mathbf{a}'\mathbf{X} - d_1|)$ en el entorno de contaminación. Más precisamente satisfacen las ecuaciones (puedo suponer $\mathbf{X}$ esfericamente simetrica por la invariancia del sesgo)

$$
\begin{aligned}
P(c - m_1(c) &\leq \mathbf{a}'\mathbf{X} \leq c + m_1(c)) = \frac{1-2\varepsilon}{2(1-\varepsilon)} \\
P(c - m_2(c) &\leq \mathbf{a}'\mathbf{X} \leq c + m_2(c)) = \frac{1}{2(1-\varepsilon)} \\
d_2 &= m_2(d_1) \\
d_3 &= m_1(d_1) \\
d_0 &= \sup_{c \in [0, d_1]} \frac{c}{m_1(c)}
\end{aligned}
$$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2.$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^p\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^R(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2 / d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2 / d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.
- ▶ $\gamma^{*}(T_{MP}, F_{0,I}) = 2 \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2 / d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.
- ▶ $\gamma^{*}(T_{MP}, F_{0,I}) = 2 \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}$
- *Sabemos cotas inferiores para el maximo sesgo?*

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^{p}\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^{R}(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2 / d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.
- ▶ $\gamma^{*}(T_{MP}, F_{0,I}) = 2 \left. \frac{dd_1(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0}$
- *Sabemos cotas inferiores para el maximo sesgo?*
- Si **T** es cualquier estimador afinmente equivariate para posicion multivariada,

$$\boxed{B(\mathbf{T}, \varepsilon, F_{0,I}) \geq d_1(\varepsilon)} \qquad \boxed{\gamma^{*}(T, \varepsilon, F_{0,I}) \geq \left. \frac{dd_1(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0}}$$

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^p\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^R(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2 / d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.
- ▶ $\gamma^*(T_{MP}, F_{0,I}) = 2\left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}$
- *Sabemos cotas inferiores para el maximo sesgo?*
- Si $\mathbf{T}$ es cualquier estimador afinmente equivariate para posicion multivariada,

$$\boxed{B(\mathbf{T},\varepsilon, F_{0,I}) \geq d_1(\varepsilon)} \qquad \boxed{\gamma^*(T, \varepsilon, F_{0,I}) \geq \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}}$$

- En regresión simple a traves del origen, $(y = \beta x + \varepsilon)$, el estimador $\widehat{\beta} = med(y/x)$ tiene sesgo minimax (Martin, Yohai and Zamar, 1993).

Y el sesgo???

- $\varepsilon-$vecindad de contaminación restringida:

$$\mathcal{F}_{\varepsilon}^{R}(F_{0,I}) = \{(1-\varepsilon)F_{0,I} + \varepsilon\delta_{\mathbf{c}}, \mathbf{c} \in R^p\}.$$

- $B(T_{MP}, \varepsilon, F_{0,I}) = B^R(T_{MP}, \varepsilon, F_{0,I})$
- Cuando g en la definición de densidad elíptica es decreciente, resulta que ([AY, 2002])
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_0 d_2$.
- ▶ $B(T_{MP}, \varepsilon, F_{0,I}) = d_1 + d_1 d_2/d_3$ si $k(c) = c/m_1(c)$ es no decreciente, $c \leq d_1$.
- ▶ $\gamma^*(T_{MP}, F_{0,I}) = 2 \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}$
- *Sabemos cotas inferiores para el maximo sesgo?*
- Si $\mathbf{T}$ es cualquier estimador afinmente equivariate para posicion multivariada,

$$\boxed{B(\mathbf{T}, \varepsilon, F_{0,I}) \geq d_1(\varepsilon)} \qquad \boxed{\gamma^*(T, \varepsilon, F_{0,I}) \geq \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}}$$

- En regresión simple a traves del origen, $(y = \beta x + \varepsilon)$, el estimador $\widehat{\beta} = med(y/x)$ tiene sesgo minimax (Martin, Yohai and Zamar, 1993).
- P-estimadores para regresión múltiple (Maronna and Yohai, 1993) se conjeturan que tiene sesgo minimax.

Modificando el estimador de proyecciones

- Al medir la distorsion que provocaba un centro ζ usamos

$$h(\zeta, \mathbf{a}, F_n) = \frac{T_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}{S_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}$$

Modificando el estimador de proyecciones

- Al medir la distorsion que provocaba un centro $\boldsymbol{\zeta}$ usamos

$$h(\boldsymbol{\zeta}, \mathbf{a}, F_n) = \frac{T_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}{S_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}$$

- Como tomamos *el valor absoluto de esta cantidad*, perdemos información sobre el signo de la mediana de los datos proyectados. Entonces podríamos considerar medir el rango de la mediana estandarizada,

$$\sup_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\boldsymbol{\zeta}, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\boldsymbol{\zeta}, \mathbf{a}, F_n).$$

Modificando el estimador de proyecciones

- Al medir la distorsion que provocaba un centro ζ usamos

$$h(\zeta, \mathbf{a}, F_n) = \frac{T_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}{S_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), ..., \mathbf{a}'(\mathbf{x}_n - \zeta))}$$

- Como tomamos *el valor absoluto de esta cantidad*, perdemos información sobre el signo de la mediana de los datos proyectados. Entonces podríamos considerar medir el rango de la mediana estandarizada,

$$\sup_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n).$$

- Sin embargo esta cantidad nos vuelve a dar $v(\zeta, F_n)$ porque $h(\zeta, -\mathbf{a}, F_n) = -h(\zeta, \mathbf{a}, F_n)$.

Modificando el estimador de proyecciones

- Al medir la distorsion que provocaba un centro ζ usamos

$$h(\zeta, \mathbf{a}, F_n) = \frac{T_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), \ldots, \mathbf{a}'(\mathbf{x}_n - \zeta))}{S_n(\mathbf{a}'(\mathbf{x}_1 - \zeta), \ldots, \mathbf{a}'(\mathbf{x}_n - \zeta))}$$

- Como tomamos *el valor absoluto de esta cantidad*, perdemos información sobre el signo de la mediana de los datos proyectados. Entonces podríamos considerar medir el rango de la mediana estandarizada,

$$\sup_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in R^p - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n).$$

- Sin embargo esta cantidad nos vuelve a dar $v(\zeta, F_n)$ porque $h(\zeta, -\mathbf{a}, F_n) = -h(\zeta, \mathbf{a}, F_n)$.
- De alguna manera hay información redundante al mirar todas las direcciones porque se repite informacion cambiada de signo al mirar todas las direcciones. Por eso vamos a restringir la busqueda a hiperplanos de direcciones, porque todas las direcciones del espacio estan representadas alli.

Modificando el estimador de proyecciones

- *Pero con cual quedarnos de entre todos los posibles hiperplanos?*

Modificando el estimador de proyecciones

- *Pero con cual quedarnos de entre todos los posibles hiperplanos?*
- *Con el que produce el mínimo del rango porque es el que menos informacion redundante tiene.*

Modificando el estimador de proyecciones

- *Pero con cual quedarnos de entre todos los posibles hiperplanos?*
- *Con el que produce el mínimo del rango porque es el que menos informacion redundante tiene.*
- O sea si $L = L(\mathbf{b}) = \{\mathbf{x} : \mathbf{b}'\mathbf{x} \geq 0\}, \mathbf{b} \neq \mathbf{0}$ y $\mathcal{L}$ es el conjunto de tales hiperplanos $L's$ entonces

$$\boxed{V(\zeta, L, F_n) = \sup_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n)}$$

$$\boxed{M(\zeta, F_n) = \inf_{L \in \mathcal{L}} \ V(\zeta, L, F_n)}$$

$$\boxed{\mathbf{T}_{MP}^{M}(F_n) = \arg\inf_{\zeta \in R^p} M(\zeta, F_n)}$$

Modificando el estimador de proyecciones

- *Pero con cual quedarnos de entre todos los posibles hiperplanos?*
- *Con el que produce el mínimo del rango porque es el que menos informacion redundante tiene.*
- O sea si $L = L(\mathbf{b}) = \{\mathbf{x} : \mathbf{b}'\mathbf{x} \geq 0\}, \mathbf{b} \neq \mathbf{0}$ y $\mathcal{L}$ es el conjunto de tales hiperplanos $L's$ entonces

$$\boxed{V(\zeta, L, F_n) = \sup_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n)}$$

$$\boxed{M(\zeta, F_n) = \inf_{L \in \mathcal{L}} \ V(\zeta, L, F_n)}$$

$$\boxed{\mathbf{T}_{MP}^{M}(F_n) = \arg\inf_{\zeta \in R^p} M(\zeta, F_n)}$$

- $\mathbf{T}_{MP}^{M}$ es consistente en el sentido de Fisher y tiene punto de ruptura 0.5

Modificando el estimador de proyecciones

- *Pero con cual quedarnos de entre todos los posibles hiperplanos?*
- *Con el que produce el mínimo del rango porque es el que menos informacion redundante tiene.*
- O sea si $L = L(\mathbf{b}) = \{\mathbf{x} : \mathbf{b}'\mathbf{x} \geq 0\}, \mathbf{b} \neq \mathbf{0}$ y $\mathcal{L}$ es el conjunto de tales hiperplanos $L's$ entonces

$$\boxed{V(\zeta, L, F_n) = \sup_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n) - \inf_{\mathbf{a} \in L - \{\mathbf{0}\}} h(\zeta, \mathbf{a}, F_n)}$$

$$\boxed{M(\zeta, F_n) = \inf_{L \in \mathcal{L}} \; V(\zeta, L, F_n)}$$

$$\boxed{\mathbf{T}_{MP}^{M}(F_n) = \arg\inf_{\zeta \in R^p} M(\zeta, F_n)}$$

- $\mathbf{T}_{MP}^{M}$ es consistente en el sentido de Fisher y tiene punto de ruptura 0.5
- Y el maximo sesgo asintotico en la $\varepsilon-$vecindad restringida resulta ser ...

Y otra vez el sesgo!!!!

- Si g en la definicion de densidad elíptica es decreciente, entonces el maximo sesgo asintótico en la $\varepsilon-$vecindad restringida resulta ser (AZ, 2009, enviado)

$$\boxed{B^R(\mathbf{T}^M_{MP}, \varepsilon, F_{\mathbf{0},I}) = \begin{cases} d_1 & \text{si} \quad d_2(d_0 - 1) < d_1 \\ d_0 d_2 & \text{si} \quad d_2(d_0 - 1) \geq d_1 . \end{cases}}$$

Y otra vez el sesgo!!!!

- Si g en la definicion de densidad elíptica es decreciente, entonces el maximo sesgo asintótico en la $\varepsilon-$vecindad restringida resulta ser (AZ, 2009, enviado)

$$\boxed{B^R(\mathbf{T}^M_{MP}, \varepsilon, F_{\mathbf{0},I}) = \begin{cases} d_1 & \text{si} \quad d_2(d_0-1) < d_1 \\ d_0 d_2 & \text{si} \quad d_2(d_0-1) \geq d_1 . \end{cases}}$$

- Existe un valor $\varepsilon_0 > 0$ tal que $d_2(d_0-1) < d_1$ si $\varepsilon < \varepsilon_0$. Entonces en el intervalo $[0, \varepsilon_0)$ el estimador es *minimax en sesgo para contaminaciones puntuales*.

Y otra vez el sesgo!!!!

- Si g en la definicion de densidad elíptica es decreciente, entonces el maximo sesgo asintótico en la $\varepsilon-$vecindad restringida resulta ser (AZ, 2009, enviado)

$$\boxed{B^R(\mathbf{T}_{MP}^M, \varepsilon, F_{\mathbf{0},I}) = \begin{cases} d_1 & \text{si} \quad d_2(d_0 - 1) < d_1 \\ d_0 d_2 & \text{si} \quad d_2(d_0 - 1) \geq d_1 . \end{cases}}$$

- Existe un valor $\varepsilon_0 > 0$ tal que $d_2(d_0 - 1) < d_1$ si $\varepsilon < \varepsilon_0$. Entonces en el intervalo $[0, \varepsilon_0)$ el estimador es *minimax en sesgo para contaminaciones puntuales.*
- Y por lo tanto, la sensibilidad a contaminaciones alcanza su cota mínima,

$$\boxed{\gamma^*(\mathbf{T}_{MP}^M, F_{\mathbf{0},I}) = \left.\frac{dd_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}}$$

Y otra vez el sesgo!!!!

- Si g en la definicion de densidad elíptica es decreciente, entonces el maximo sesgo asintótico en la $\varepsilon-$vecindad restringida resulta ser (AZ, 2009, enviado)

$$\boxed{B^R(\mathbf{T}_{MP}^M, \varepsilon, F_{\mathbf{0},I}) = \begin{cases} d_1 & \text{si} \quad d_2(d_0-1) < d_1 \\ d_0 d_2 & \text{si} \quad d_2(d_0-1) \geq d_1 . \end{cases}}$$

- Existe un valor $\varepsilon_0 > 0$ tal que $d_2(d_0-1) < d_1$ si $\varepsilon < \varepsilon_0$. Entonces en el intervalo $[0, \varepsilon_0)$ el estimador es *minimax en sesgo para contaminaciones puntuales*.
- Y por lo tanto, la sensibilidad a contaminaciones alcanza su cota mínima,

$$\boxed{\gamma^*(\mathbf{T}_{MP}^M, F_{\mathbf{0},I}) = \left.\frac{d d_1(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}}$$

- Cuando $\mathbf{X}$ es normal multivariada para el modelo central, resulta que la funcion $k(c)$ es creciente en $[0, d_1)$ si $\varepsilon < 0.4088$. Entonces $d_0 = d_1/d_3$. Luego la condicion que controla la formula del sesgo de $\mathbf{T}_{MP}^M$ resulta $d_2(d_1 - d_3) < d_1 d_3$ que es valida para $\varepsilon < 0.3144$.

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).
- Se genera un conjunto de candidatos posibles del siguiente modo. Se obtienen M submuestras aleatorias J con $p+1$ elementos cada una.

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).
- Se genera un conjunto de candidatos posibles del siguiente modo. Se obtienen M submuestras aleatorias J con $p+1$ elementos cada una.
- Para cada uno de esos conjuntos se obtiene la media y matriz de covarianza muestral $\mu_J = \text{ave}_{i \in J}(\mathbf{x}_i)$ y $\Sigma_J = \text{ave}_{i \in J}(\mathbf{x}_i - \boldsymbol{\mu}_J)'(\mathbf{x}_i - \boldsymbol{\mu}_J)$.

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).
- Se genera un conjunto de candidatos posibles del siguiente modo. Se obtienen M submuestras aleatorias J con $p+1$ elementos cada una.
- Para cada uno de esos conjuntos se obtiene la media y matriz de covarianza muestral $\mu_J = \text{ave}_{i\in J}(\mathbf{x}_i)$ y $\Sigma_J = \text{ave}_{i\in J}(\mathbf{x}_i - \boldsymbol{\mu}_J)'(\mathbf{x}_i - \boldsymbol{\mu}_J)$.
- Sea $h = [n/2]$. Se realizan dos pasos de concentracion segun el mecanismo propuesto por Rousseeuw and Van Driessen (1999) para computar el MCD. Dadas las distancias de Mahalanobis $d_i(\boldsymbol{\mu}_J, \Sigma_J) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_J)\Sigma_J^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_J)'}, i = 1, ..., n$ se construye un $h-$subconjunto ordenando

$$d_{1:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq d_{2:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq ... \leq d_{h:n}(\boldsymbol{\mu}_J, \Sigma_J)$$

  y conservando los indices $H_1 = H_1(J) = \{\pi_1(1), ..., \pi_1(h)\}$, con π_1 la permutacion que da la muestra ordenada $d_{i:n}(\boldsymbol{\mu}_J, \Sigma_J) = d_{\pi(i)}(\boldsymbol{\mu}_J, \Sigma_J)$.

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).
- Se genera un conjunto de candidatos posibles del siguiente modo. Se obtienen M submuestras aleatorias J con $p+1$ elementos cada una.
- Para cada uno de esos conjuntos se obtiene la media y matriz de covarianza muestral $\mu_J = \text{ave}_{i\in J}(\mathbf{x}_i)$ y $\Sigma_J = \text{ave}_{i\in J}(\mathbf{x}_i - \boldsymbol{\mu}_J)'(\mathbf{x}_i - \boldsymbol{\mu}_J)$.
- Sea $h = [n/2]$. Se realizan dos pasos de concentracion segun el mecanismo propuesto por Rousseeuw and Van Driessen (1999) para computar el MCD. Dadas las distancias de Mahalanobis $d_i(\boldsymbol{\mu}_J, \Sigma_J) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_J)\Sigma_J^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_J)'}, i = 1, ..., n$ se construye un $h-$subconjunto ordenando

$$d_{1:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq d_{2:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq ... \leq d_{h:n}(\boldsymbol{\mu}_J, \Sigma_J)$$

  y conservando los indices $H_1 = H_1(J) = \{\pi_1(1), ..., \pi_1(h)\}$, con π_1 la permutacion que da la muestra ordenada $d_{i:n}(\boldsymbol{\mu}_J, \Sigma_J) = d_{\pi(i)}(\boldsymbol{\mu}_J, \Sigma_J)$.
- Computar $\boldsymbol{\mu}_{H_1} = \text{ave}_{i\in H_1}(\mathbf{x}_i)$ y $\Sigma_{H_1} = \text{ave}_{i\in H_1}(\mathbf{x}_i - \boldsymbol{\mu}_{H_1})'(\mathbf{x}_i - \boldsymbol{\mu}_{H_1})$.

Algoritmo de calculo

- Para cada observacion $\mathbf{x}_i$ se computa la lejania aproximada segun Maronna and Yohai (1995).
- Se genera un conjunto de candidatos posibles del siguiente modo. Se obtienen M submuestras aleatorias J con $p+1$ elementos cada una.
- Para cada uno de esos conjuntos se obtiene la media y matriz de covarianza muestral $\mu_J = \text{ave}_{i\in J}(\mathbf{x}_i)$ y $\Sigma_J = \text{ave}_{i\in J}(\mathbf{x}_i - \boldsymbol{\mu}_J)'(\mathbf{x}_i - \boldsymbol{\mu}_J)$.
- Sea $h = [n/2]$. Se realizan dos pasos de concentracion segun el mecanismo propuesto por Rousseeuw and Van Driessen (1999) para computar el MCD. Dadas las distancias de Mahalanobis $d_i(\boldsymbol{\mu}_J, \Sigma_J) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_J)\Sigma_J^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_J)'}, i = 1, \ldots, n$ se construye un $h-$subconjunto ordenando

$$d_{1:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq d_{2:n}(\boldsymbol{\mu}_J, \Sigma_J) \leq \ldots \leq d_{h:n}(\boldsymbol{\mu}_J, \Sigma_J)$$

  y conservando los indices $H_1 = H_1(J) = \{\pi_1(1), \ldots, \pi_1(h)\}$, con π_1 la permutacion que da la muestra ordenada $d_{i:n}(\boldsymbol{\mu}_J, \Sigma_J) = d_{\pi(i)}(\boldsymbol{\mu}_J, \Sigma_J)$.
- Computar $\boldsymbol{\mu}_{H_1} = \text{ave}_{i\in H_1}(\mathbf{x}_i)$ y $\Sigma_{H_1} = \text{ave}_{i\in H_1}(\mathbf{x}_i - \boldsymbol{\mu}_{H_1})'(\mathbf{x}_i - \boldsymbol{\mu}_{H_1})$.
- Realizar otro paso de concentracion ordenando $d_i(\boldsymbol{\mu}_{H_1}, \Sigma_{H_1}), i = 1, \ldots, n$ y se consigue otro conjunto $H_2 = H_2(J) = \{\pi_2(1), \ldots, \pi_2(h)\}$ a traves de la muestra ordenada $d_{i:n}(\boldsymbol{\mu}_{H_1}, \Sigma_{H_1}) = d_{\pi(i)}(\boldsymbol{\mu}_{H_1}, \Sigma_{H_1})$, $i = 1, \ldots, n$, (π_2 es la permutacion que da la muestra ordenada).

- La media resultante es $\boldsymbol{\mu}_{H_2}$ $=\text{ave}_{i\in H_2}(\mathbf{x}_i)$.

- La media resultante es $\boldsymbol{\mu}_{H_2}$ $=\text{ave}_{i\in H_2}(\mathbf{x}_i)$.
- El conjunto de M candidatos a minimizar la lejania modificada es $U=\{\boldsymbol{\mu}_{H_2(J)}: J$ submuestra aleatoria$\}$.

- La media resultante es $\boldsymbol{\mu}_{H_2} = \text{ave}_{i \in H_2}(\mathbf{x}_i)$.
- El conjunto de M candidatos a minimizar la lejania modificada es $U = \{\boldsymbol{\mu}_{H_2(J)} : J$ submuestra aleatoria$\}$.
- Considerar $M(M-1)/2$ semiespacios generados como $L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \{\mathbf{x} : (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\mathbf{x} \geq 0\}$, como $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$ in U.

- La media resultante es μ_{H_2} $=\mathrm{ave}_{i\in H_2}(\mathbf{x}_i)$.
- El conjunto de M candidatos a minimizar la lejania modificada es $U=\{\mu_{H_2(J)}: J$ submuestra aleatoria$\}$.
- Considerar $M(M-1)/2$ semiespacios generados como $L(\mu_1,\mu_2)=\{\mathbf{x}:(\mu_1-\mu_2)'\mathbf{x}\geq 0\}$, como $\mu_1\neq\mu_2$, μ_1 y μ_2 in U.
- El conjunto de direcciones $A=\{\mathbf{a}_1,\ldots,\mathbf{a_N}\}$ esta generado a través de una muestra aleatoria de una $N_p(\mathbf{0},I)$.

- La media resultante es μ_{H_2} $=\text{ave}_{i\in H_2}(\mathbf{x}_i)$.
- El conjunto de M candidatos a minimizar la lejania modificada es $U=\{\mu_{H_2(J)}:J$ submuestra aleatoria$\}$.
- Considerar $M(M-1)/2$ semiespacios generados como $L(\mu_1,\mu_2)=\{\mathbf{x}:(\mu_1-\mu_2)'\mathbf{x}\geq 0\}$, como $\mu_1\neq\mu_2$, μ_1 y μ_2 in U.
- El conjunto de direcciones $A=\{\mathbf{a}_1,...,\mathbf{a_N}\}$ esta generado a través de una muestra aleatoria de una $N_p(\mathbf{0},I)$.
- Entonces un MP aproximado $\widehat{\mu}_n$ y un MP modificado aproximado $\widehat{\mu}_n^M$ surgen del siguiente esquema:

$$
\begin{aligned}
h_n(\mu,\mathbf{a}) &= \frac{\text{med}_{1\leq i\leq n}(\mathbf{a}'(\mathbf{x}_i-\mu))}{\text{med}_{1\leq i\leq n}|\mathbf{a}'\mathbf{x}_i-\text{med}_{1\leq j\leq n}(\mathbf{a}'\mathbf{x}_j)|}, \mathbf{a}\in A, \mu\in U\\
\nu_n(\mu) &= \max_{\mathbf{a}\in A}|h_n(\mu,\mathbf{a})|, \mu\in U\\
V_n(\mu,\mu_1,\mu_2) &= \max_{\mathbf{a}\in L(\mu_1,\mu_2)\cap A} h_n(\mu,\mathbf{a}) - \min_{\mathbf{a}\in L(\mu_1,\mu_2)\cap A} h_n(\mu,\mathbf{a})\\
\widehat{\mu}_n &= \arg\min_{\mu\in U}\nu_n(\mu)\\
\widehat{\mu}_n^M &= \arg\min_{\mu\in U,\mu_1\in U,\mu_2\in U} V_n(\mu,\mu_1,\mu_2).
\end{aligned}
$$

- La media resultante es μ_{H_2} =ave$_{i \in H_2}(\mathbf{x}_i)$.
- El conjunto de M candidatos a minimizar la lejania modificada es $U = \{\mu_{H_2(J)} : J$ submuestra aleatoria$\}$.
- Considerar $M(M-1)/2$ semiespacios generados como $L(\mu_1, \mu_2) = \{\mathbf{x} : (\mu_1 - \mu_2)'\mathbf{x} \geq 0\}$, como $\mu_1 \neq \mu_2$, μ_1 y μ_2 in U.
- El conjunto de direcciones $A = \{\mathbf{a}_1, ..., \mathbf{a}_\mathbf{N}\}$ esta generado a través de una muestra aleatoria de una $N_p(\mathbf{0}, I)$.
- Entonces un MP aproximado $\widehat{\mu}_n$ y un MP modificado aproximado $\widehat{\mu}_n^M$ surgen del siguiente esquema:

$$h_n(\mu, \mathbf{a}) = \frac{\mathrm{med}_{1 \leq i \leq n}(\mathbf{a}'(\mathbf{x}_i - \mu))}{\mathrm{med}_{1 \leq i \leq n}|\mathbf{a}'\mathbf{x}_i - \mathrm{med}_{1 \leq j \leq n}(\mathbf{a}'\mathbf{x}_j)|}, \mathbf{a} \in A, \mu \in U$$

$$\nu_n(\mu) = \max_{\mathbf{a} \in A} |h_n(\mu, \mathbf{a})|, \mu \in U$$

$$V_n(\mu, \mu_1, \mu_2) = \max_{\mathbf{a} \in L(\mu_1, \mu_2) \cap A} h_n(\mu, \mathbf{a}) - \min_{\mathbf{a} \in L(\mu_1, \mu_2) \cap A} h_n(\mu, \mathbf{a})$$

$$\widehat{\mu}_n = \arg\min_{\mu \in U} \nu_n(\mu)$$

$$\widehat{\mu}_n^M = \arg \min_{\mu \in U, \mu_1 \in U, \mu_2 \in U} V_n(\mu, \mu_1, \mu_2).$$

- La media muestral fue incluida en el conjunto de candidatos para mejorar la eficiencia del estimador

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}^{M}_{MP}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.
- Para cada estimador $\mathbf{T}$ se computó el ECM calculado como

$$\frac{1}{500} \sum_{i=1}^{500} ||\mathbf{T}_i||^2,$$

donde $\mathbf{T}_i$ es el valor del estimador para la i-ésima muestra.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.
- Para cada estimador $\mathbf{T}$ se computó el ECM calculado como

$$\frac{1}{500}\sum_{i=1}^{500} ||\mathbf{T}_i||^2,$$

  donde $\mathbf{T}_i$ es el valor del estimador para la i-ésima muestra.
- Se computaron $\mathbf{T}_{MP}$ y $\mathbf{T}_{MP}^{M}$ segun el algoritmo anterior $M = 500$ and $N = 500$.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.
- Para cada estimador $\mathbf{T}$ se computó el ECM calculado como

$$\frac{1}{500}\sum_{i=1}^{500}||\mathbf{T}_i||^2,$$

donde $\mathbf{T}_i$ es el valor del estimador para la i-ésima muestra.
- Se computaron $\mathbf{T}_{MP}$ y $\mathbf{T}_{MP}^{M}$ segun el algoritmo anterior $M = 500$ and $N = 500$.
- Se dan los ECM de la media y eficiencias relativas respecto de la media de SD_0, SD_{90}, MVE, MCD, MP y MPM.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.
- Para cada estimador $\mathbf{T}$ se computó el ECM calculado como

$$\frac{1}{500} \sum_{i=1}^{500} ||\mathbf{T}_i||^2,$$

  donde $\mathbf{T}_i$ es el valor del estimador para la i-ésima muestra.
- Se computaron $\mathbf{T}_{MP}$ y $\mathbf{T}_{MP}^{M}$ segun el algoritmo anterior $M = 500$ and $N = 500$.
- Se dan los ECM de la media y eficiencias relativas respecto de la media de SD_0, SD_{90}, MVE, MCD, MP y MPM.
- Los estimadores mas eficientes son los SD, seguidos por el MP.

Comparando eficiencias de estimadores

- Monte Carlo para comparar las eficiencia bajo distribucion normal multivariada para muestras finitas de los estimadores $\mathbf{T}_{MP}$, $\mathbf{T}_{MP}^{M}$ y varios de los estimadores anteriormente presentados (MVE,MCD y SD con dos funciones de pesos determinadas).
- Se incluye la media muestral que tiene eficiencia óptima.
- Las dimensiones consideradas fueron $p = 2 - 10, 15$ y 20.
- La cantidad de observaciones fueron 50 y 100.
- El número de replicaciones fue 500.
- Para cada estimador $\mathbf{T}$ se computó el ECM calculado como

$$\frac{1}{500}\sum_{i=1}^{500}||\mathbf{T}_i||^2,$$

donde $\mathbf{T}_i$ es el valor del estimador para la i-ésima muestra.
- Se computaron $\mathbf{T}_{MP}$ y $\mathbf{T}_{MP}^{M}$ segun el algoritmo anterior $M = 500$ and $N = 500$.
- Se dan los ECM de la media y eficiencias relativas respecto de la media de SD_0, SD_{90}, MVE, MCD, MP y MPM.
- Los estimadores mas eficientes son los SD, seguidos por el MP.
- El MPM es menos eficiente que el MP pero mas eficiente que MVE y MCD.

- El MVE fue computado usando submuestreo (Seccion 6.7.3 de Maronna,Martin and Yohai (2006))

- El MVE fue computado usando submuestreo (Seccion 6.7.3 de Maronna,Martin and Yohai (2006))
- SD_0 y SD_{90}, usan una lejanía aproximada $\nu(\mathbf{x}_i, F_n)$ usando 500 direcciones. Cada una de estas direcciones es ortogonal a la del hiperplano determinado por una submuestra aleatoria con p elementos.

- El MVE fue computado usando submuestreo (Seccion 6.7.3 de Maronna,Martin and Yohai (2006))
- SD_0 y SD_{90}, usan una lejanía aproximada $\nu(\mathbf{x}_i, F_n)$ usando 500 direcciones. Cada una de estas direcciones es ortogonal a la del hiperplano determinado por una submuestra aleatoria con p elementos.
- Para MCD se usó el algoritmo rápido de Rousseeuw and Van Driessen (1999) con 500 submuestras y dos pasos de concentración.

- El MVE fue computado usando submuestreo (Seccion 6.7.3 de Maronna,Martin and Yohai (2006))
- SD_0 y SD_{90}, usan una lejanía aproximada $\nu(\mathbf{x}_i, F_n)$ usando 500 direcciones. Cada una de estas direcciones es ortogonal a la del hiperplano determinado por una submuestra aleatoria con p elementos.
- Para MCD se usó el algoritmo rápido de Rousseeuw and Van Driessen (1999) con 500 submuestras y dos pasos de concentración.
- Para SD_0 la función de peso w es $w(u) = 1/u$ y para SD_{90}

$$w(u) = \min\left(\frac{1}{u}, \frac{1}{c}\right)$$

con $c = \left(\chi^2_{0.90,p}\right)^{1/2}$ ($\chi^2_{\alpha,p}$ es el α-cuantil de la χ^1-distribución con p grados de libertad).

ECM para la media y eficiencias relativas

- **de estimadores robustos para la distribución normal**, $n = 50$.

p	MEDIA	SD0	SD90	MVE	MCD	MP	MPM
	ECM	ER	ER	ER	ER	ER	ER
2	0.041	0.81	0.97	0.21	0.24	0.79	0.69
3	0.064	0.82	0.97	0.22	0.29	0.82	0.78
4	0.082	0.87	0.97	0.21	0.33	0.76	0.75
5	0.101	0.87	0.96	0.21	0.35	0.77	0.78
6	0.118	0.87	0.95	0.20	0.35	0.75	0.72
7	0.138	0.89	0.95	0.22	0.38	0.80	0.75
8	0.157	0.91	0.95	0.22	0.39	0.80	0.73
9	0.178	0.91	0.95	0.25	0.42	0.83	0.78
10	0.199	0.92	0.95	0.26	0.43	0.84	0.75
15	0.302	0.92	0.93	0.34	0.45	0.94	0.80
20	0.398	0.92	0.92	0.44	0.47	0.98	0.78

ECM para la media y eficiencias relativas

- **de estimadores robustos para la distribucion normal**, $n = 100$.

p	MEDIA	SD0	SD90	MVE	MCD	MP	MPM
	ECM	ER	ER	ER	ER	ER	ER
2	0.019	0.80	0.97	0.17	0.20	0.72	0.64
3	0.030	0.83	0.98	0.14	0.22	0.77	0.75
4	0.039	0.89	0.98	0.13	0.28	0.76	0.75
5	0.049	0.89	0.98	0.13	0.30	0.77	0.78
6	0.058	0.91	0.98	0.13	0.32	0.76	0.78
7	0.069	0.91	0.98	0.13	0.34	0.79	0.77
8	0.081	0.92	0.99	0.13	0.35	0.83	0.78
9	0.090	0.93	0.99	0.13	0.38	0.82	0.79
10	0.099	0.94	0.99	0.14	0.39	0.86	0.80
15	0.148	0.95	0.99	0.18	0.43	0.94	0.81
20	0.203	0.95	0.99	0.22	0.45	0.97	0.84

Sesgos maximos

- **para niveles de contaminacion** $\varepsilon = 0.05$ **y** $\varepsilon = 0.10$

p	ε=0.05						ε =0.10					
	ESTIMADOR						ESTIMADOR					
	SD0	SD90	MVE	MCD	MP	MPM	SD0	SD90	MVE	MCD	MP	MPM
2	0.077	0.131	0.395	0.277	0.141	0.066	0.159	0.292	0.619	0.539	0.321	0.140
3	0.095	0.149	0.395	0.307	0.141	0.066	0.200	0.333	0.690	0.655	0.321	0.140
4	0.123	0.192	0.395	0.336	0.141	0.066	0.267	0.436	0.720	0.786	0.321	0.140
5	0.160	0.226	0.395	0.369	0.141	0.066	0.361	0.523	0.730	0.935	0.321	0.140
6	0.197	0.255	0.395	0.406	0.141	0.066	0.483	0.613	0.740	1.094	0.321	0.140
7	0.246	0.298	0.395	0.440	0.141	0.066	0.617	0.717	0.750	1.277	0.321	0.140
8	0.310	0.347	0.395	0.480	0.141	0.066	0.762	0.837	0.750	1.481	0.321	0.140
9	0.368	0.394	0.395	0.520	0.141	0.066	0.898	0.951	0.750	1.703	0.321	0.140
10	0.418	0.443	0.395	0.560	0.141	0.066	1.021	1.066	0.750	1.965	0.321	0.140
15	0.715	0.719	0.395	0.781	0.141	0.066	1.746	1.751	0.765	3.846	0.321	0.140
20	1.013	1.013	0.395	1.117	0.141	0.066	2.471	2.472	0.775	7.005	0.321	0.140

Sesgos maximos

- **para niveles de contaminacion** $\varepsilon = 0.15$ **y** $\varepsilon = 0.20$

p	ε=0.15						ε =0.20					
	ESTIMADOR						ESTIMADOR					
	SD0	SD90	MVE	MCD	MP	MPM	SD0	SD90	MVE	MCD	MP	MPM
2	0.27	0.51	0.91	0.90	0.56	0.23	0.40	0.81	1.29	1.49	0.90	0.32
3	0.36	0.56	1.03	1.20	0.56	0.23	0.57	0.98	1.48	2.22	0.90	0.32
4	0.48	0.79	1.08	1.58	0.56	0.23	0.84	1.31	1.57	3.19	0.90	0.32
5	0.67	0.95	1.12	2.04	0.56	0.23	1.18	1.59	1.65	4.46	0.90	0.32
6	0.90	1.14	1.14	2.58	0.56	0.23	1.58	1.91	1.71	6.17	0.90	0.32
7	1.13	1.33	1.16	3.23	0.56	0.23	1.98	2.25	1.78	8.53	0.90	0.32
8	1.41	1.54	1.18	4.07	0.56	0.23	2.44	2.62	1.84	11.69	0.90	0.32
9	1.67	1.77	1.20	5.05	0.56	0.23	2.91	3.02	1.90	15.92	0.90	0.32
10	1.91	1.99	1.22	6.28	0.56	0.23	3.41	3.48	1.97	21.68	0.90	0.32
15	3.27	3.28	1.31	18.00	0.56	0.23	5.76	5.77	2.30	116.8	0.90	0.32
20	4.63	4.63	1.41	49.92	0.56	0.23	8.12	8.11	2.65	412.0	0.90	0.32

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.
2. ε_0 depende de la distribución central y es igual a 0.3044 en el caso normal.

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.
2. ε_0 depende de la distribución central y es igual a 0.3044 en el caso normal.
3. En el subconjunto $[\varepsilon_0, 0.5)$, el sesgo resulta menor que el de otros estimados robusto conocidos, incluyendo el estimador de proyección original.

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.
2. ε_0 depende de la distribución central y es igual a 0.3044 en el caso normal.
3. En el subconjunto $[\varepsilon_0, 0.5)$, el sesgo resulta menor que el de otros estimados robusto conocidos, incluyendo el estimador de proyección original.
4. La complejidad computacional del estimador es un problema que hasta el momento podemos solucionar aproximadamente con el algoritmo descripto. En realidad el mismo estimador de proyeccion presenta complicaciones computacionales. Aparentemente hay un algoritmo exacto para $p = 2$ (Zuo, 2009, to be published).

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.
2. ε_0 depende de la distribución central y es igual a 0.3044 en el caso normal.
3. En el subconjunto $[\varepsilon_0, 0.5)$, el sesgo resulta menor que el de otros estimados robusto conocidos, incluyendo el estimador de proyección original.
4. La complejidad computacional del estimador es un problema que hasta el momento podemos solucionar aproximadamente con el algoritmo descripto. En realidad el mismo estimador de proyeccion presenta complicaciones computacionales. Aparentemente hay un algoritmo exacto para $p = 2$ (Zuo, 2009, to be published).
5. Los estimadores de proyeccion se estiman sin necesidad de una matriz de dispersion. Luego pueden utilizarse en la estimacion de cualquier matriz de dispersion robusta que lo use de centro de los datos. El tener un mejor estimador para centrar las observaciones permite una mejor estimacion de la matriz de dispersion.

Conclusiones

1. Se ha introducido una modificacion del estimador de proyeccion (Tyler, 1994). Este estimador resulta *minimax en sesgo* en una vecindad restringida de contaminaciones de masas puntuales para un rango de niveles de contaminacion $[0, \varepsilon_0)$, $\varepsilon_0 > 0$.
2. ε_0 depende de la distribución central y es igual a 0.3044 en el caso normal.
3. En el subconjunto $[\varepsilon_0, 0.5)$, el sesgo resulta menor que el de otros estimados robusto conocidos, incluyendo el estimador de proyección original.
4. La complejidad computacional del estimador es un problema que hasta el momento podemos solucionar aproximadamente con el algoritmo descripto. En realidad el mismo estimador de proyeccion presenta complicaciones computacionales. Aparentemente hay un algoritmo exacto para $p = 2$ (Zuo, 2009, to be published).
5. Los estimadores de proyeccion se estiman sin necesidad de una matriz de dispersion. Luego pueden utilizarse en la estimacion de cualquier matriz de dispersion robusta que lo use de centro de los datos. El tener un mejor estimador para centrar las observaciones permite una mejor estimacion de la matriz de dispersion.
6. Muchas gracias!