Trimming methods in model checking

Eustasio del Barrio

Universidad de Valladolid (Spain)

joint work with P.C. Álvarez, J.A. Cuesta and C. Matrán

Journées Statistiques du Sud 2008
Toulouse, June 16-18 2008

Outline

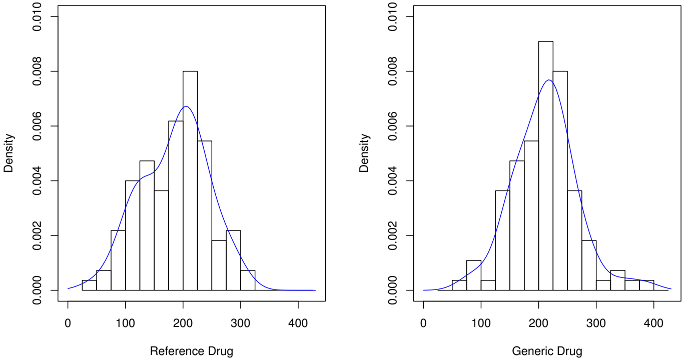
1. Introduction
2. Trimmed Distributions
3. Trimming with a common patern
4. Optimal incomplete transportation of mass
5. Conclusions

Introduction

- Trimming methods used in Statistics as a way to robustify procedures.
- Classical trimming: symmetric (trimmed mean, ...).
- Alternative trimming methods:
    - depth: introduced by Donoho and Gasko (1992)
    - peeling: convex peeling (Barnett, 1976; and Bebbington, 1978) or peeling based on ellipsoids (Titterington, 1978)
    - "impartial" trimming: Rousseeuw (1985) and Gordaliza (1991)
- Impartial trimming methodology: location estimation (Rousseeuw, 1985; and Gordaliza, 1991), regression problems (Rousseeuw, 1985), cluster analysis (Cuesta-Albertos et al. 1997, 1998, 2002, 2008; and García-Escudero et al. 1999, 1999a, 1999b, 2003, 2005) and principal component analysis (Maronna, 2005)
- Data-driven trimming methods: goodness-of-fit (Alvarez-Esteban et al, 2008)

Model Validation: Similar Distributions

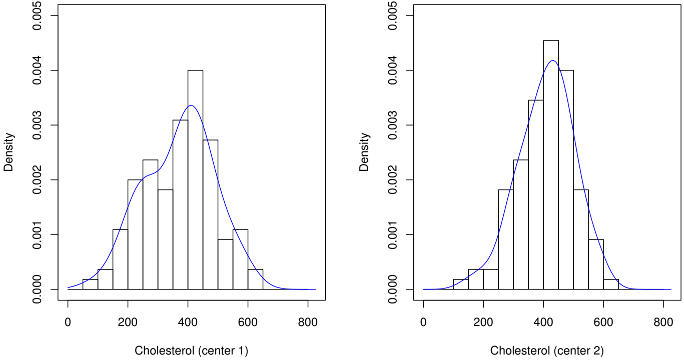
Does the generic drug have the same therapeutic effect as the reference drug?



Due to possible existence of subgroups with own peculiarities (ethnic, professional, ...) maybe enough if generic drug has the same effect for 95% of patients.

Model Validation: Goodness-of-Fit

Aim: Studying the effect of a new drug to control the cholesterol level.



Can this sample be obtained from a gaussian population?

Departures from the Model

One-sample problems: observe $X \sim P$, check $P = Q$ or $P \in \mathcal{F}$

Two-sample problems: observe $X \sim P$, $Y \sim Q$, check $P = Q$

Often $P = Q$ or $P \in \mathcal{F}$ not really important; instead $P \simeq Q$ or $P \simeq \mathcal{F}$

Usually we fix $\theta = \theta(P)$ and a metric, d. Rather than testing

$$H_0 : \theta(P) = \theta(Q) \quad \text{vs.} \quad H_a : \theta(P) \neq \theta(Q)$$

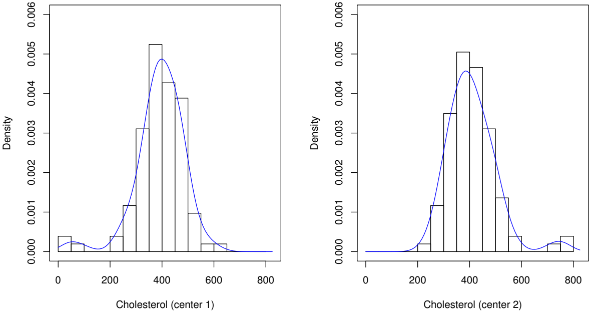
we consider

$$H_0 : d(\theta(P), \theta(Q)) \leq \Delta \quad \text{vs.} \quad H_a : d(\theta(P), \theta(Q)) > \Delta$$

$$H_0 : d(\theta(P), \theta(Q)) > \Delta \quad \text{vs.} \quad H_a : d(\theta(P), \theta(Q)) \leq \Delta$$

The ‘Core’ of a Distribution

Removing 5% of data in first sample and 5% in second the remaining data in both samples produce very similar histograms



The core of the underlying distributions are similar

Examples: trying to assess similarity of two human populations with different inmigration patterns; checking equality in different measurements of the same phisical magnitude

Trimming the Sample

Remove a fraction, of size at most α, of the data in the sample for a better comparison to a pattern/other sample:

$$\text{replace} \qquad \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \qquad \text{with } \frac{1}{n}\sum_{i=1}^{n}b_i\delta_{x_i}$$

$b_i = 0$ for observations in the bad set; $b_i/n = \frac{1}{n-k}$ others,

k number of trimmed observations; $k \leq n\alpha$ and $\frac{1}{n-k} \leq \frac{1}{n}\frac{1}{1-\alpha}$ Instead

keeping/removing we could increase weight in good ranges (by $\frac{1}{1-\alpha}$ at most); downplay in bad zones, not necessarily removing

$$\frac{1}{n}\sum_{i=1}^{n}b_i\delta_{x_i}, \text{ with } 0 \leq b_i \leq \frac{1}{(1-\alpha)}, \text{ and } \frac{1}{n}\sum_{i=1}^{n}b_i = 1.$$

Examples of Trimming Techniques

Nonparametric assessment of bioequivalence: α-trimmed version of the Mallows distance (Munk and Czado, 1998)

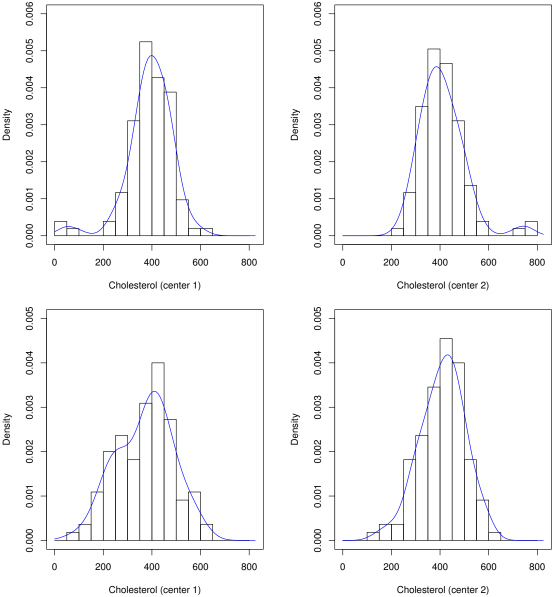
$P \sim F,\ Q \sim G$

$$\Gamma_{\alpha,p}(F,G) = \frac{1}{1-\alpha}\left\{\int_{\alpha/2}^{1-\alpha/2} |F^{-1}(s) - G^{-1}(s)|^p ds\right\}^{1/p}$$

$$H : \Gamma_{\alpha,p}(F,G) > \Delta_0 \qquad \text{vs} \qquad K : \Gamma_{\alpha,p}(F,G) \leq \Delta_0$$

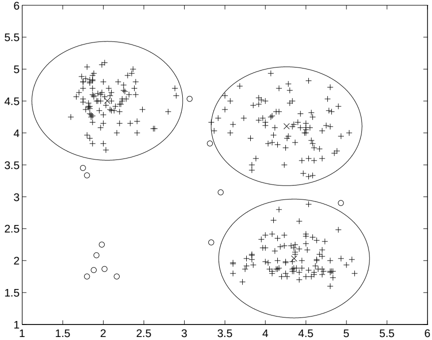
Trimming introduced to robustify the procedure

Trimming at tails may not work


Density
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0
200
400
600
800
Cholesterol (center 1)
Density
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0
200
400
600
800
Cholesterol (center 2)
Density
0.000
0.001
0.002
0.003
0.004
0.005
0
200
400
600
800
Cholesterol (center 1)
Density
0.000
0.001
0.002
0.003
0.004
0.005
0
200
400
600
800
Cholesterol (center 2)

Data-driven Trimming Methods

Old Faithfull Geyser example



Horizontal axis: Eruption time (min.)
Vertical axis: Previous eruption time

- A more flexible way to trim: data structure itself tell us which is the best way of removing data.
- Cuesta-Albertos, Gordaliza & Matrán (1997) proposed this trimming procedure to robustify k-means
- In the multivariate setting is more clear the inexistence of *a priori* directions to trim as well as the possible existence of "inliers".

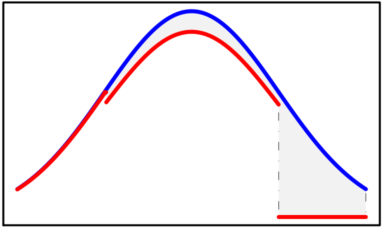
Trimmed Distributions

$(\mathcal{X}, \beta)$ measurable space; $\mathcal{P}(\mathcal{X}, \beta)$ prob. measures on $(\mathcal{X}, \beta)$, $P \in \mathcal{P}(\mathcal{X}, \beta)$

Definition

For $0 \leq \alpha \leq 1$

$$\mathcal{R}_\alpha(P) = \left\{ Q \in \mathcal{P}(\mathcal{X}, \beta) : \quad Q \ll P, \quad \frac{dQ}{dP} \leq \frac{1}{1-\alpha} \quad P\text{-a.s.} \right\}$$



Equivalently, $Q \in \mathcal{R}_\alpha(P)$ iff $Q \ll P$ and $\frac{dQ}{dP} = \frac{1}{1-\alpha} f$ with $0 \leq f \leq 1$

If $f \in \{0, 1\}$ then $f = I_A$ with $P(A) = 1 - \alpha$: trimming reduces to $P(\cdot|A)$.

Trimming allows to play down the weight of some regions of the measurable space without completely removing them from the feasible set

Trimmed Distributions II

Some basic properties:

Proposition

(a) $\alpha_1 \leq \alpha_2 \Rightarrow \mathcal{R}_{\alpha_1}(P) \subset \mathcal{R}_{\alpha_2}(P)$

(b) $\mathcal{R}_\alpha(P)$ *is a convex set.*

(c) *For* $\alpha < 1$, $Q \in \mathcal{R}_\alpha(P)$ *iff* $Q(A) \leq \frac{1}{1-\alpha} P(A)$ *for all* $A \in \beta$

(d) *If* $\alpha < 1$ *and* $(\mathcal{X}, \beta)$ *is separable metric space then* $\mathcal{R}_\alpha(P)$ *is closed for the topology of the weak convergence in* $\mathcal{P}(\mathcal{X}, \beta)$.

(e) *If* $\mathcal{X}$ *is also complete, then* $\mathcal{R}_\alpha(P)$ *is compact.*

Parametrizing Trimmed Distributions: $\mathcal{X} = \mathbb{R}$

Define

$$\mathcal{C}_\alpha := \left\{ h \in \mathcal{AC}[0,1] :\, h(0) = 0,\, h(1) = 1,\, 0 \leq h' \leq \frac{1}{1-\alpha} \right\}$$

$\mathcal{C}_\alpha$ is the set of distribution functions of probabilities in $\mathcal{R}_\alpha(U(0,1))$

Call $h \in \mathcal{C}_\alpha$ a trimming function

Take P with d.f. F. Let P_h the prob. with d.f. $h \circ F$: $P_h \in \mathcal{R}_\alpha(P)$; in fact

Proposition

$$\mathcal{R}_\alpha(P) = \{P_h :\, h \in \mathcal{C}_\alpha\}$$

The parametrization need not be unique (it is not if P is discrete)

A useful fact: $\mathcal{C}_\alpha$ is compact for the uniform topology

Parametrizing Trimmed Distributions: general $\mathcal{X}$

Proposition

If T transports Q to P, then

$$\mathcal{R}_\alpha(P) = \left\{ Q^* \circ T^{-1} : Q^* \in \mathcal{R}_\alpha(Q) \right\}.$$

If $Q = U(0,1)$, $P \sim F$, $T = F^{-1}$ we recover the $\mathcal{C}_\alpha$-parametrization

For separable, complete $\mathcal{X}$ we can take $Q = U(0,1)$; T Skorohod-Dudley-Wichura

For $\mathcal{X} = \mathbb{R}^k$, more interesting $Q \ll \ell^k$, T the Brenier-McCann map: the unique cyclically monotone map transporting Q to P.

Measuring dissimilarities through common trimming

d a metric on $\mathcal{F} \subset \mathcal{P}(\mathbb{R}^k, \beta)$; $P_0 \in \mathcal{P}(\mathbb{R}^k, \beta)$; $P_0 \ll \ell^k$

$$\mathcal{T}(P_1, P_2) = \min_{P_0^* \in \mathcal{R}_\alpha(P_0)} d(P_0^* \circ T_1^{-1}, P_0^* \circ T_2^{-1})$$

T_i Brenier-McCann map from P_0 to P_i

$$P_{0,\alpha} = \operatorname*{argmin}_{P_0^* \in \mathcal{R}_\alpha(P_0)} d(P_0^* \circ T_1^{-1}, P_0^* \circ T_2^{-1})$$

$P_{0,\alpha}$ is a best (P_0, α)-trimming for P_1 and P_2

On $\mathbb{R}$, taking $P_0 = U(0,1)$

$$\mathcal{T}(P, Q) = \min_{h \in \mathcal{C}_\alpha} d(P_h, Q_h)$$

$$h_\alpha = \operatorname*{argmin}_{h \in \mathcal{C}_\alpha} d(P_h, Q_h)$$

h_α is a best α-matching function for P and Q

$h \mapsto d(P_h, Q_h)$ continuous in $\| \cdot \|_\infty$ for $d_{BL}, \mathcal{W}_p, \ldots \Rightarrow$

a best α-matching function exists

Wasserstein distance

We consider the Wasserstein metric, $\mathcal{W}_p$, $p \geq 1$,

$$\mathcal{W}_p^p(P,Q) = \inf_{\pi \in \Pi(P,Q)} \{ \int \|x-y\|^p d\pi(x,y) \}$$

$\mathcal{W}_p$ a metric on $\mathcal{F}_p$, probabilities with finite p-th moment

Proposition

$P \in \mathcal{F}_p \Rightarrow \mathcal{R}_\alpha(P) \subset \mathcal{F}_p$ and $\mathcal{R}_\alpha(P)$ compact in the $\mathcal{W}_p$ topology

On the real line

$$\mathcal{W}_p^p(P,Q) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt, \quad P \sim F, Q \sim G, \quad P,Q \in \mathcal{F}_p(\mathbb{R})$$

For $\mathcal{W}_p$, h_α easy to compute: $P \sim F$, $Q \sim G$

$$\mathcal{W}_2^2(P_h, Q_h) = \int_0^1 \left(F^{-1} \circ h^{-1} - G^{-1} \circ h^{-1} \right)^2 = \int_0^1 (F^{-1} - G^{-1})^2 h'$$

Define $L_{F,G}(x) = \ell\{t \in (0,1) : |F^{-1}(t) - G^{-1}(t)| \le x\}$, $x \ge 0$

Then $\quad h'_\alpha(t) = \frac{1}{1-\alpha} I(|F^{-1}(t) - G^{-1}(t)| \le L^{-1}_{F,G}(1-\alpha))$

In general, (mild assumptions)

$$\mathcal{W}_2^2(P_0^* \circ T_1^{-1}, P_0^* \circ T_2^{-1}) = \int \|T_1(x) - T_2(x)\|^2 dP_0^*(x),$$

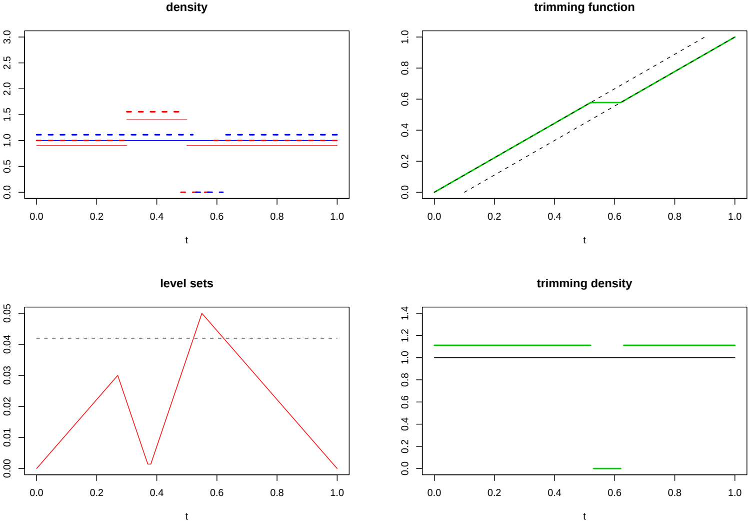
$$\frac{dP_{0,\alpha}}{dP_0} = \frac{1}{1-\alpha} I_{\{\|T_1 - T_2\| \le c_\alpha(P_{1,2})\}}$$

and

$$\mathcal{T}^2(P_1, P_2) = \int \|T_1(x) - T_2(x)\|^2 dP_{0,\alpha}(x)$$

Matching functions: $P=(1-\beta)U(0,1)+\beta U(\gamma,\gamma+\delta)$, $Q=U(0,1)$

$$h_\alpha = \operatorname*{argmin}_{h\in\mathcal{C}_\alpha} \mathcal{W}_2(P_h,Q_h)$$



Trimmed comparisons

Using trimmings for tests about the core of the distribution of the data

One sample problems:
Assume $X_1, \ldots, X_n$ i.i.d. P and fix Q. We are interested in testing

$$H_1 : \mathcal{T}^{(\alpha)}(P, Q) = 0 \text{ against } K_1 : \mathcal{T}^{(\alpha)}(P, Q) > 0$$

$$H_2 : \mathcal{T}^{(\alpha)}(P, Q) > \Delta \text{ against } K_2 : \mathcal{T}^{(\alpha)}(P, Q) \leq \Delta$$

Two sample problems:
Assume $X_1, \ldots, X_n$ i.i.d. P and $Y_1, \ldots, Y_m$ i.i.d. Q. Still interested in testing H_i against K_i, but here Q is unknown

In the one sample case we reject H_1/H_2 for large/small $T_n^{(\alpha)} = \mathcal{T}^{(\alpha)}(P_n, Q)$
In the two sample case we reject H_1/H_2 for large/small $T_{n,m}^{(\alpha)} = \mathcal{T}^{(\alpha)}(P_n, Q_m)$

P_n, Q_m empirical measures
In general, $T_n^{(\alpha)}$, $T_{n,m}^{(\alpha)}$ not distribution free; tests use asymptotics, bootstrap,. . .

Asymptotics for $T_n^{(\alpha)}$ $\quad (d = \mathcal{W}_2,\ \mathcal{T}^{(\alpha)}(P,Q) = 0)$

$h_{n,\alpha} = \operatorname*{argmin}_{h \in \mathcal{C}_\alpha} d((P_n)_h, Q_h)$ is the α-trimmed empirical matching function

$T_n^{(\alpha)} = d((P_n)_{h_{n,\alpha}}, Q_{h_{n,\alpha}})$

Define $\mathcal{C}_\alpha(P,Q) = \{h \in \mathcal{C}_\alpha :\ d(P_h, Q_h) = 0\}$

$\mathcal{C}_\alpha(F,F) = \mathcal{C}_\alpha; \quad F \neq G \Rightarrow \mathcal{C}_\alpha(F,G) \subsetneq \mathcal{C}_\alpha; \quad \mathcal{C}_\alpha(F,G) \neq \emptyset$ iff $\mathcal{T}^{(\alpha)}(P,Q) = 0$

The size of $\mathcal{C}_\alpha(F,G)$ depends on $\ell\{t \in (0,1) : F^{-1}(t) \neq G^{-1}(t)\}$

$\mathcal{C}_\alpha(F,G)$ compact for $\|\cdot\|_\infty$

Theorem

$$n(T_n^{(\alpha)})^2 \underset{w}{\to} \min_{h \in \mathcal{C}_\alpha(F,G)} \int_0^1 \frac{B(t)^2}{f^2(F^{-1}(t))} h'(t)\, dt = \int_0^1 \frac{B(t)^2}{f^2(F^{-1}(t))} h'_{\alpha,F,G}(t)\, dt$$

$$h_{n,\alpha} \underset{w}{\to} h_{\alpha,F,G}$$

Asymptotics for $T_n^{(\alpha)}$ $(d = \mathcal{W}_2,\ \mathcal{T}^{(\alpha)}(P,Q) > 0)$

Theorem

$\sqrt{n}((T_n^{(\alpha)})^2 - (\mathcal{T}^{(\alpha)}(P,Q))^2) \underset{w}{\to} N(0, \sigma_\alpha^2(P,Q))$

$$\sigma_\alpha^2(P,Q) = 4\left(\int_0^1 l^2(t)dt - \left(\int_0^1 l(t)dt\right)^2\right),$$

where

$$l(t) = \int_{F^{-1}(1/2)}^{F^{-1}(t)} (x - G^{-1}(F(x)))h'_\alpha(F(x))dx,$$

and $h'_\alpha(t) = \frac{1}{1-\alpha} I(|F^{-1}(t) - G^{-1}(t)| \leq L_{F,G}^{-1}(1-\alpha))$
$\sigma_\alpha^2(P,Q)$ *consistently estimated by*

$$s_{n,\alpha}^2(G) = \frac{4}{(1-\alpha)^2}\frac{1}{n}\sum_{i,j=1}^{n-1}(i \wedge j - \frac{ij}{n})a_{n,i}a_{n,j},$$

$$a_{n,i} = (X_{(i+1)} - X_{(i)})((X_{(i+1)} + X_{(i)})/2 - G^{-1}(i/n))I_{(|X_{(i)} - G^{-1}\left(\frac{i}{n}\right)| \leq \ell_{F_n,G}^{-1}(1-\alpha))}.$$

Asymptotics for $T_{n,m}^{(\alpha)}$ $(d = \mathcal{W}_2,\, \mathcal{T}^{(\alpha)}(P,Q) > 0)$

Theorem

If $\frac{n}{n+m} \to \lambda \in (0,\infty)$

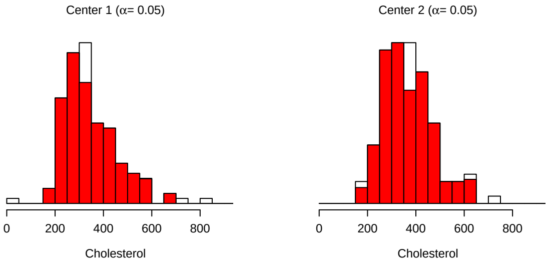
$$\sqrt{\frac{n}{n+m}}((T_{n,m}^{(\alpha)})^2 - (\mathcal{T}^{(\alpha)}(P,Q))^2) \underset{w}{\to} N(0,(1-\lambda)\sigma_\alpha^2(P,Q) + \lambda\sigma_\alpha^2(Q,P))$$

$(1-\lambda)\sigma_\alpha^2(P,Q) + \lambda\sigma_\alpha^2(Q,P))$ *consistently estimated by*

$$s_{n,m,\alpha}^2 = \tfrac{m}{n+m} s_{n,\alpha}^2(G_m) + \tfrac{n}{n+m} s_{m,\alpha}^2(F_n)$$

Data Example: Cholesterol and fibrinogen levels

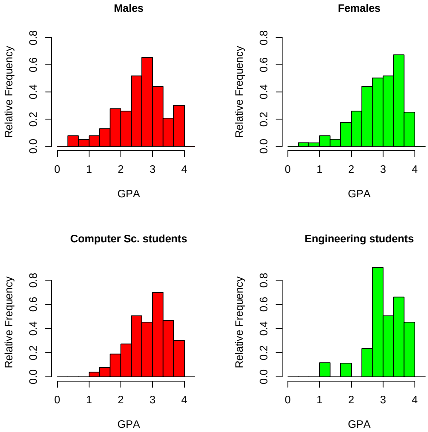
Data: Cholesterol and fibrinogen levels in two sets of patients (of sizes 116 and 141) in two clinical centers



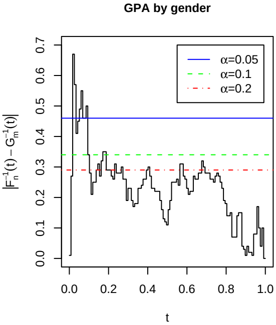
- For fibrinogen data, data-driven trimming $\simeq$ symmetric trimming ($\alpha = 0.05$)
- For cholesterol data, significant trimming also at central regions in both samples; this improves the level of similarity
- Impartial trimming useful as descriptive tool to detect ranges of dissimilarity

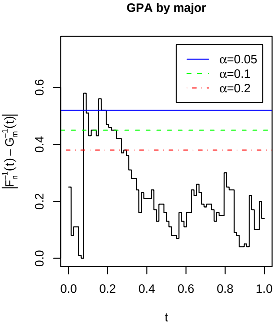
Data Example: GPA

Data: College Grade Point Average $\in [0, 4]$, 234 students
Students classified by the Gender and Major (1 = CS, 2 = Eng, 3 = Other)
Distributional similarity males ($n = 117$) vs females ($m = 117$); CS ($n = 78$) vs Eng ($m = 78$)



Test	**p-value**	
	GPA by gender	GPA by major
Shapiro-Wilks (Sample 1)	0.0176	0.0360
Shapiro-Wilks (Sample 2)	0.0217	0.0001
Kolmogorov-Smirnov	0.0028	0.0040
Wilcoxon-Mann-Whitney	0.0004	0.0175





p-value curves

Test $H_0 : \mathcal{T}^{(\alpha)}(F,G) > \Delta_0^2$ against $H_a : \mathcal{T}^{(\alpha)}(F,G) \leq \Delta_0^2$
We use the statistic

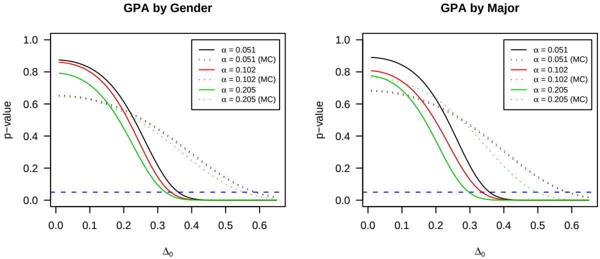
$$Z_{n,m,\alpha} = \sqrt{\frac{nm}{n+m}} \frac{(T_{n,m}^{(\alpha)} - \Delta_0^2)}{s_{n,m,\alpha}}.$$

Asymptotic p-value curve:

$$P(\Delta_0) := \sup_{\{(F,G):(F,G)\in H_0\}} \lim_{n,m\to\infty} P_{F,G}\left(Z_{n,m,\alpha} \leq z_0\right) = \Phi\left(\sqrt{\frac{nm}{n+m}} \frac{T_{n,m}^{(\alpha)} - \Delta_0^2}{s_{n,m,\alpha}}\right),$$

$z_0 =$ observed value of $Z_{n,m,\alpha}$ (sup attained when distance $= \Delta_0$)
Use of asymptotic p-value curves:

- For a fixed Δ_0 (controlling degree of dissimilarity), we can find the level of significance at which F and G cannot be assumed similar
- For a fixed test level (p-value), we can find the value of Δ_0 such that for every $\Delta \geq \Delta_0$ we should reject the hypothesis $H_0 : \mathcal{T}^{(\alpha)}(F,G) \geq \Delta^2$



p-value curves using impartial trimming and Munk & Czado (MC)

(For F and G different only in location, Wasserstein distance = absolute difference of locations)

- GPA points of males and females show similarity up to $\Delta_0 =$ 0.32 to 0.36 (between 11.4% to 12.8% of the average of the medians of the samples)
- using symmetrical trimmings cutpoint for $\Delta_0 =$ 0.56 to 0.59 (20% to 21% of the average of the medians of the samples)

- Data-driven common trimming is a useful tool for checking model adequacy in noisy sets of data
- It works for other applications (e.g., robust normality testing, Álvarez-Esteban et al. (2008c))
- It works for multivariate data, functional data (open problems)
- Other trimming patterns can be of interest (maybe much more interesting)

$$\begin{aligned}\mathcal{T}_1(P_1,P_2) &:= \min_{P_2^*\in\mathcal{R}_\alpha(P_2)} d(P_1,P_2^*),\\ \mathcal{T}_2(P_1,P_2) &:= \min_{P_1^*\in\mathcal{R}_\alpha(P_1),P_2^*\in\mathcal{R}_\alpha(P_2)} d(P_1^*,P_2^*),\end{aligned}$$

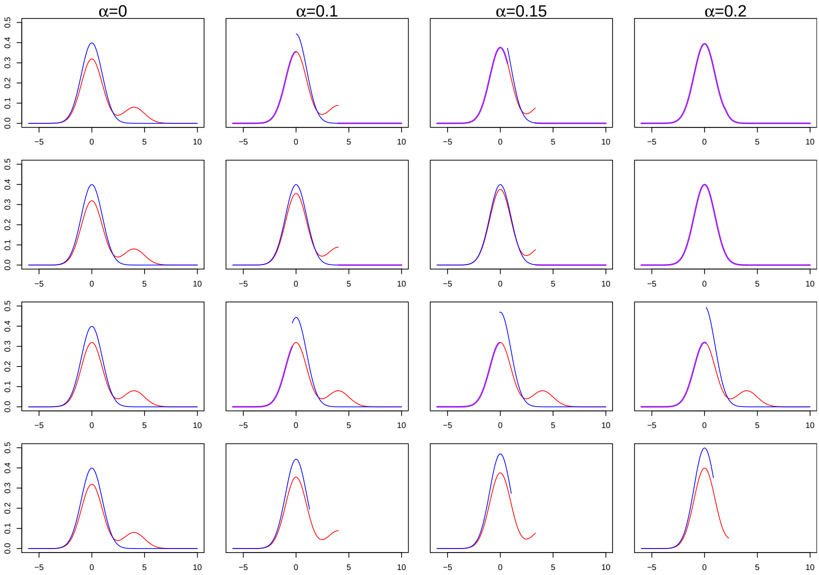
$\mathcal{T}_1$ removes contamination: $P_2=(1-\varepsilon)\,P_1+\varepsilon\,Q$, $\Rightarrow P_1\in\mathcal{R}_\alpha(P_2)$ $(\alpha\geq\varepsilon)$

$$(1-\alpha)\,P_1(A)\leq(1-\varepsilon)\,P_1(A)+\varepsilon\,Q(A)\qquad\forall A\in\beta$$

Hence,

$$\mathcal{T}_1(P_1,P_2)=0$$

One-sided/two-sided/common trimming



Optimal incomplete transportation of mass

Setup

Supply: Mass (pile of sand, some other good) located around X

Demand: Mass needed at several locations scattered around Y

Assume total supply exceeds total demand (demand$=(1-\alpha)\times$ supply, $\alpha \in (0,1)$)

We don't have to move all the initial mass; some α- fraction can be dismissed

Find a way to complete this task with a minimal cost.

Rescale to represent the *target distribution* by Q, p.m. on Y

Represent the *initial distribution* by $\frac{1}{1-\alpha}P$, P p.m. on X

$c(x,y)$ cost of moving a unit of mass from x to y

(Incomplete) transportation plan: a way to move part of the mass in $\frac{1}{1-\alpha}P$ to Q represented by π, a joint probability measure on $X \times Y$

Optimal incomplete transportation of mass

Target distribution $= Q \Leftrightarrow$

$$\pi(X \times B) = Q(B), \quad B \subset Q$$

Amount of mass taken from a location in X cannot exceed available mass:

$$\pi(A \times Y) \leq \frac{1}{1-\alpha} P(A), \quad A \subset X$$

π transportation plan $\Leftrightarrow \pi \in \Pi(\mathcal{R}_\alpha(P), Q)$

Now

$$\inf_{\pi \in \Pi(\mathcal{R}_\alpha(P), Q)} \int_{X \times Y} c(x,y) d\pi(x,y)$$

is the *optimal incomplete transportation problem*

If $X = Y$ Banach separable and $c(x,y) = \|x-y\|^2$ then

$$\mathcal{W}_2^2(\mathcal{R}_\alpha(P), Q) = \inf_{\pi \in \Pi(\mathcal{R}_\alpha(P), Q)} \int_{X \times Y} c(x,y) d\pi(x,y)$$

Dual problem

Write $I[\pi] = \int_{X\times Y} c(x,y)d\pi(x,y)$ and

$$J_\alpha(\varphi,\psi) = \frac{1}{1-\alpha}\int_X \varphi dP + \int_Y \psi dQ$$

$(\varphi,\psi) \in \mathcal{C}_b(X) \times \mathcal{C}_b(Y)$ such that

$$\varphi(x) \le 0 \quad \text{and} \quad \varphi(x) + \psi(y) \le c(x,y), \quad x \in X,\, y \in Y$$

For $\pi \in \Pi(\mathcal{R}_\alpha(P), Q)$

$$\frac{1}{1-\alpha}\int_X \varphi dP + \int_Y \psi dQ \le \int_{X\times Y} (\varphi(x)+\psi(y))d\pi(x,y) \le \int_{X\times Y} c(x,y)d\pi(x,y)$$

Therefore

$$\sup_{(\varphi,\psi)\in\Phi_c} J_\alpha(\varphi,\psi) \le \inf_{\pi\in\Pi(\mathcal{R}_\alpha(P),Q)} I[\pi]$$

Theorem

$$\sup_{(\varphi,\psi)\in\Phi_c} J_\alpha(\varphi,\psi) = \min_{\pi\in\Pi(\mathcal{R}_\alpha(P),Q)} I[\pi]$$

and the $\min$ *in the right-hand side is attained.*

X, Y complete, separable; c lower semicontinuous

For X and Y compact, c continuous, duality follows from general duality (Fenchel-Rockafellar): E normed space, E^* dual, A, B convex then

$$\inf_{x\in E}(A(x)+B(x)) = \max_{y\in E^*}(-A^*(-y)-B^*(y))$$

provided $A(x_0)<\infty$, $B(x_0)<\infty$ and A is continuous at x_0

$$A^*(y) = \sup_{x\in E}(\langle y,x\rangle - A(x)), \quad y\in E^* \quad \text{Legendre-Fenchel transform}$$

For c unif. continuous, bounded the sup is also attained in Φ_c; without boundedness enlarged Φ_c required

Incomplete transportation with quadratic cost

$X = Y = \mathbb{R}^k$, $c(x,y) = \|x-y\|^2$

$\tilde{\Phi}_c$ class of pairs $(\varphi,\psi) \in L^1(P) \times L^1(Q)$ such that

$$\varphi(x) \leq 0 \quad P-\text{a.s.} \quad \text{and} \quad \varphi(x)+\psi(y) \leq c(x,y), \quad P\times Q-\text{a.s.}.$$

Theorem

$$\max_{(\varphi,\psi)\in\tilde{\Phi}_c} J_\alpha(\varphi,\psi) = \min_{\pi\in\Pi(\mathcal{R}_\alpha(P),Q)} I[\pi].$$

max attained at (φ,ψ) with $\varphi(x) = \|x\|^2 - a_0(x)$ and $\psi(y) = \|y\|^2 - 2a_0^*(y)$

a_0 convex, lower semicontinuous, P-integrable with $a_0(x) \geq \|x\|^2/2$, $x \in \mathbb{R}^n$ such that

$$\frac{1}{1-\alpha}\int a_0 dP + \int a_0^* dQ = \min_a \left[\frac{1}{1-\alpha}\int a dP + \int a^* dQ\right],$$

a^* convex-conjugate of a

Characterization of optimal incomplete t.p.'s

P and Q p.m. on $\mathbb{R}^k$ with finite second moment

Theorem

(i) $\pi \in \Pi(\mathcal{R}_\alpha(P), Q)$ *optimal incomplete t.p. iff there exists* a *convex, lower semicontinuous satisfying* $a(x) \geq \frac{\|x\|^2}{2}$, $x \in \mathbb{R}^k$, *such that* $y \in \partial a(x)$ π*-a.s.*

$$\textit{and} \quad \tfrac{1}{1-\alpha} \int (\|x\|^2 - 2a(x)) dP(x) = \int (\|x\|^2 - 2a(x)) d\pi(x, y).$$

$\tilde{\varphi}(x) = \|x\|^2 - 2a(x)$, $\tilde{\psi}(y) = \|y\|^2 - 2a^*(y)$ *maximize* $J_\alpha(\varphi, \psi)$.

(ii) Q *absolutely continuous,* a *as in (i)* $\Rightarrow$ a^* Q*-a.s. diferentiable and there is a unique optimal incomplete t.p.:* $\pi = Q \circ (\nabla a^* \times \mathrm{Id})^{-1}$

(iii) P *with density* f, P_α *optimal* α*-trimming* $\Rightarrow$ P_α *has density* f_α, a *as* P*-a.s. differentiable and* $\pi = P_\alpha \circ (\mathrm{Id} \times \nabla a)^{-1}$ *is an optimal t.p.*

$$\textit{Further} \quad \left(a(x) - \tfrac{\|x\|^2}{2}\right)\left(\tfrac{1}{1-\alpha} f(x) - f_\alpha(x)\right) = 0, \textit{a.e.}$$

Characterization of optimal incomplete t.p.'s (II)

Corollary

If Q is absolutely continuous there is a unique $P_\alpha \in \mathcal{R}_\alpha(P)$ such that

$$\mathcal{W}_2^2(P_\alpha, Q) = \min_{R \in \mathcal{R}_\alpha(P)} \mathcal{W}_2^2(R, Q).$$

More precisely, $P_\alpha = Q \circ (\nabla a^*)^{-1}$ and

$$\min_{R \in \mathcal{R}_\alpha(P)} \mathcal{W}_2^2(R, Q) = \int_{\mathbb{R}^n} \|y - \nabla a^*(y)\|^2 dQ(y).$$

Corollary (Trim or move)

If P absolutely continuous, $P_\alpha \circ (\nabla a)^{-1} = Q$ and

$$\|x - \nabla a(x)\|^2 \left(\tfrac{1}{1-\alpha} f(x) - f_\alpha(x) \right) = 0, \quad \textit{a.e.}$$

$$\min_{R \in \mathcal{R}_\alpha(P)} \mathcal{W}_2^2(R, Q) = \frac{1}{1-\alpha} \int_{\mathbb{R}^n} \|x - \nabla a(x)\|^2 \, dP(x).$$

Monge-Ampère equation in incomplete transportation

Theorem

If P and Q absolutely continuous, $P_\alpha = \mathrm{argmin}_{R\in\mathcal{R}_\alpha}\mathcal{W}_2(R,Q)$ *then*

$$(f_\alpha(x) - \tfrac{1}{1-\alpha}f(x))(f_\alpha(x) - g(x)) = 0 \quad \textit{a.e.}.$$

If $\nabla a(x)$ *is the optimal incomplete transportation plan then*

$$\|x - \nabla a(x)\|^2 \left(\tfrac{1}{1-\alpha}f(x) - g(\nabla a(x))\det D^2 a(x)\right) = 0, \quad \textit{a.e.}.$$

Doubly incomplete transportation of mass

Assume now we only have to satisfy a fraction of the demand, $1 - \alpha_2$

Total amount of demand to be served only a fraction of the total supply, $1 - \alpha_1$

Try to minimize the transportation cost.

This is the *doubly incomplete transportation problem*:

$$\min_{\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))} I[\pi] = \min_{\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))} \int_{X \times Y} c(x, y) d\pi(x, y).$$

The min is attained if X, Y complete, separable

If $X = Y$ Banach separable, $c(x, y) = \|x - y\|^2$ then

$$\mathcal{W}_2^2(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q)) = \min_{\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))} \int_{X \times Y} c(x, y) d\pi(x, y)$$

Dual problem

$$J_{\alpha_1,\alpha_2}(\varphi,\psi)=\frac{1}{1-\alpha_1}\int\varphi dP+\frac{1}{1-\alpha_2}\int\psi dQ-\frac{\alpha_1}{1-\alpha_1}\bar{\varphi}-\frac{\alpha_2}{1-\alpha_2}\bar{\psi}$$

$(\varphi,\psi)\in\Psi$, class of pairs in $\mathcal{C}_b(\mathbb{R}^k)\times\mathcal{C}_b(\mathbb{R}^k)$ s.t. $\varphi(x)+\psi(y)\leq\|x-y\|^2$

$\bar{\varphi}=\sup_x\varphi(x)$, $\bar{\psi}=\sup_y\psi(y)$

Theorem

$$\max_{(\varphi,\psi)\in\Phi}J_{\alpha_1,\alpha_2}(\varphi,\psi)=\min_{\pi\in\Pi(\mathcal{R}_{\alpha_1}(P),\mathcal{R}_{\alpha_2}(Q))}I[\pi]$$

and the max in the left-hand is attained.

No need to enlarge Ψ to have the max attained

max attained at $\varphi(x)=\|x\|^2-2a(x)$, $\psi(y)=\|y\|^2-2a^*(y)$ with a finite convex s.t.

$$\frac{\|x\|^2}{2}-M\leq a(x)\leq\frac{\|x\|^2}{2}+M$$

as a consequence, a^* finite convex s.t. $\frac{\|y\|^2}{2}-M\leq a^*(y)\leq\frac{\|y\|^2}{2}+M$

Characterization of optimal doubly incomplete t.p.'s

Theorem

(i) $\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(P))$ *is an optimal incomplete t.p. iff there exists* a *finite, convex s.t.* $\varphi(x) = \|x\|^2 - 2a(x)$ *bounded for which*

$$\begin{aligned} a(x) + a^*(y) &= x \cdot y \quad \pi - \text{a.s.} \\ \frac{1}{1-\alpha_1}\int (\varphi(x) - \bar{\varphi}) dP(x) &= \int (\varphi(x) - \bar{\varphi}) d\pi(x,y) \\ \frac{1}{1-\alpha_2}\int (\psi(y) - \bar{\psi}) dQ(y) &= \int (\psi(y) - \bar{\psi}) d\pi(x,y) \end{aligned}$$

(ii) *If* P *absolutely continuous,* π*,* a *as in (i), then* $\pi = P_{\alpha_1} \circ (\mathrm{Id} \times \nabla a)^{-1}$*,* $Q_{\alpha_2} = P_{\alpha_1} \circ (\nabla a)^{-1}$ *and*

$$\|x - \nabla a(x)\|^2 (f_{\alpha_1}(x) - \tfrac{1}{1-\alpha_1} f(x)) = 0 \quad \text{a.s.}$$

Also, $$\mathcal{W}_2^2(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q)) = \frac{1}{1-\alpha_1}\int_{\mathbb{R}^n} \|x - \nabla a(x)\|^2 dP(x).$$

Uniqueness in doubly incomplete optimal transportation

$P \mapsto \mathcal{W}_2^2(P,Q)$ is stricly convex if Q abs. continuous

$(P,Q) \mapsto \mathcal{W}_2^2(P,Q)$ is not stricly convex in general, in fact

Theorem

If $Q_1 \neq Q_2$ and there is no common o.t.p. T such that $Q_1 = P_1 \circ T^{-1}$ and $Q_2 = P_2 \circ T^{-1}$, then

$$\mathcal{W}_2^2(\gamma P_1 + (1-\gamma)P_2, \gamma Q_1 + (1-\gamma)Q_2) < \gamma \mathcal{W}_2^2(P_1,Q_1) + (1-\gamma)\mathcal{W}_2^2(P_2,Q_2).$$

Strict convexity gives uniqueness of minimizer in $\mathcal{W}_2(\mathcal{R}_\alpha(P),Q)$; from duality:

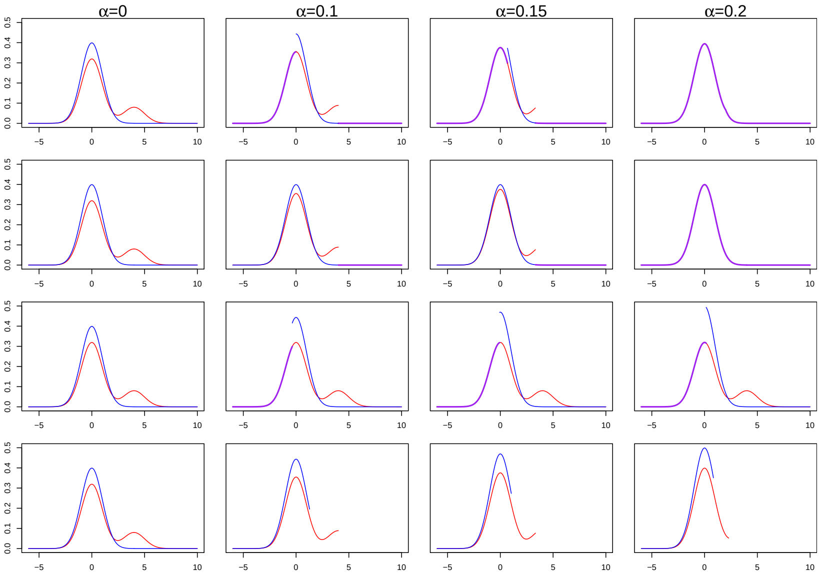
Theorem

There exists a unique $\pi_0 \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))$ such that

$$I[\pi_0] = \min_{\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))} I[\pi]$$

provided $\min_{\pi \in \Pi(\mathcal{R}_{\alpha_1}(P), \mathcal{R}_{\alpha_2}(Q))} I[\pi] > 0$ and P or Q is absolutely continuous.

Optimal incomplete transportation plans: Examples



Consistency of best trimmed approximations/matchings

$\{X_n\}_n$, $\{Y_n\}_n$ sequences of i.i.d. r.v.'s; $\mathcal{L}(X_n) = P$, $\mathcal{L}(Y_n) = Q$, $P, Q \in \mathcal{F}_2(R^k)$

P_n, Q_n empirical distributions

Theorem

(a) *If* $Q \ll \ell^k$ *and* $P_{n,\alpha} := \underset{P^* \in \mathcal{R}_\alpha(P_n)}{\text{arg min}} \mathcal{W}_2(P^*, Q)$, *then*

$$\mathcal{W}_2(P_{n,\alpha}, P_\alpha) \to 0 \text{ a.s., where } P_\alpha := \underset{P^* \in \mathcal{R}_\alpha(P)}{\text{arg min}} \mathcal{W}_2(P^*, Q).$$

(b) *If* $P \ll \ell^k$ *and* $Q_{n,\alpha} \in \mathcal{R}_\alpha(Q)$ *minimizes* $\mathcal{W}_2(P_n, \mathcal{R}_\alpha(Q))$, *then*

$$\mathcal{W}_2(Q_{n,\alpha}, Q_\alpha) \to 0 \text{ a.s., where } Q_\alpha := \underset{Q^* \in \mathcal{R}_\alpha(Q)}{\text{arg min}} \mathcal{W}_2(P, Q^*).$$

(c) *If* P *or* $Q \ll \ell^k$ *then* $\mathcal{W}_2(P_{n,\alpha}, P_\alpha) \to 0$ *and* $\mathcal{W}_2(Q_{n,\alpha}, Q_\alpha) \to 0$ *a.s.,*

where $(P_\alpha, Q_\alpha) := \text{arg min}\{\mathcal{W}_2(P^*, Q^*) : \ P^* \in \mathcal{R}_\alpha(P), Q^* \in \mathcal{R}_\alpha(Q)\}.$

Conclusions/Future work

- Data-driven trimming methods are a very powerful tool in Statistics
- Even for model checking
- Different trimming patterns can be of interest
- A correct understanding of them (and use of the related tools) leads to some probabilistic and computational challenges
- Optimal transportation metric is a very useful choice
- Lots of open problems!

References

Alvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2008a). Trimmed Comparison of Distributions. *To appear in J.A.S.A.*

Alvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2008b). Similarity of probability measures through trimming. Preprint.

Alvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2008c). Assessing when a sample is mostly normal. Preprint.

del Barrio, E. (2008). A duality based approach to optimal incomplete transportation of mass. Preprint.