Asymptotics for dissimilarity measures based on trimming

Eustasio del Barrio

Universidad de Valladolid (Spain)

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

5èmes Rencontre Statistiques Mathématiques
Bordeaux-Santander-Toulouse-Valladolid, Parc du Teich, 3-5 June, 2009

Model validation

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

Model validation

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 $P = Q$

Model validation

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 $P \simeq \mathcal{F}$

Model validation

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

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

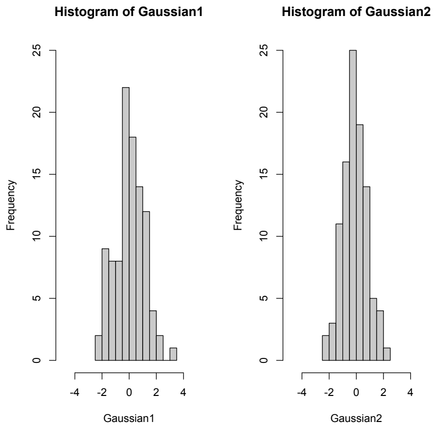
we consider

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

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

Example

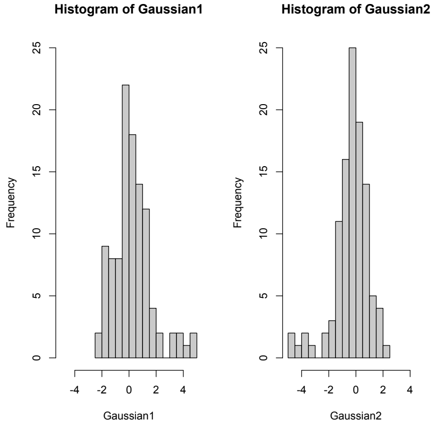
Generate 2 samples of size 100 from N(0,1)



Two-sample K-S test: p-`value = .2106`

Example

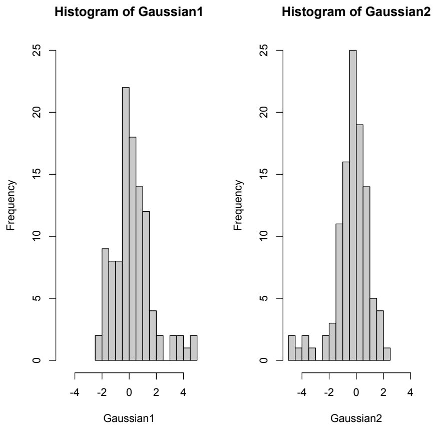
Add six anomalous points



Two-sample K-S test: p-value = .2106

Example

Add six anomalous points



Two-sample K-S test: p-value = ~~.2106~~ .0312 < .05 ⇒ Reject!

Similarity

Even checking $H_0 : d(P, Q) \leq \Delta$ vs. $H_a : d(P, Q) > \Delta$ can be badly affected by a few outliers

Similarity

Even checking $H_0 : d(P,Q) \leq \Delta$ vs. $H_a : d(P,Q) > \Delta$ can be badly affected by a few outliers

The probabilities P and Q are similar at level $\alpha \in [0,1]$ if

there exists a probability R such that $\begin{cases} P &= (1-\alpha)R + \alpha\tilde{P} \\ Q &= (1-\alpha)R + \alpha\tilde{Q} \end{cases}$

Similarity

Even checking $H_0 : d(P,Q) \leq \Delta$ vs. $H_a : d(P,Q) > \Delta$ can be badly affected by a few outliers

The probabilities P and Q are similar at level $\alpha \in [0,1]$ if

there exists a probability R such that $\begin{cases} P = (1-\alpha)R + \alpha\tilde{P} \\ Q = (1-\alpha)R + \alpha\tilde{Q} \end{cases}$

(equivalently, $d_{TV}(P,Q) \leq \alpha$).

Other null models also of interest:

Similarity

Even checking $H_0 : d(P, Q) \leq \Delta$ vs. $H_a : d(P, Q) > \Delta$ can be badly affected by a few outliers

The probabilities P and Q are similar at level $\alpha \in [0, 1]$ if

there exists a probability R such that $\begin{cases} P = (1-\alpha)R + \alpha\tilde{P} \\ Q = (1-\alpha)R + \alpha\tilde{Q} \end{cases}$

(equivalently, $d_{TV}(P, Q) \leq \alpha$).

Other null models also of interest:

$$H_0 : P = \mathcal{L}(\varphi_1(Z)), Q = \mathcal{L}(\varphi_2(Z)) \text{ and } \mathbb{P}(\varphi_1(Z) \neq \varphi_2(Z)) \leq \alpha$$

φ_i in some restricted class

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

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

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 $0 \leq \alpha \leq 1$

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

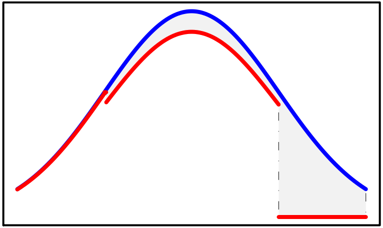
Trimmed Distributions

$(\mathcal{X}, \beta)$ 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 $0 \leq f \leq 1$

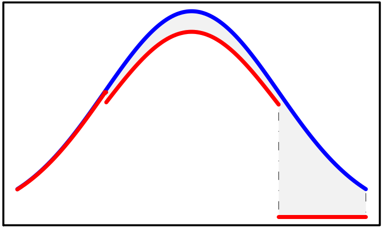
Trimmed Distributions

$(\mathcal{X}, \beta)$ 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 $P(\cdot|A)$.

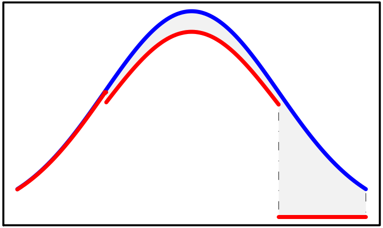
Trimmed Distributions

$(\mathcal{X}, \beta)$ 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 P_0 to P, then

$$\mathcal{R}_\alpha(P) = \left\{ R \circ T^{-1} :\ R \in \mathcal{R}_\alpha(P_0) \right\}.$$

If $P_0 = U(0,1)$, $P \sim F$, $T = F^{-1}$ we recover the $\mathcal{C}_\alpha$-parametrization

For separable, complete $\mathcal{X}$ we can take $P_0 = U(0,1)$; T Skorohod-Dudley-Wichura

For $\mathcal{X} = \mathbb{R}^k$, more interesting $P_0 \ll \ell^k$, T the Brenier-McCann map: the unique cyclically monotone map transporting P_0 to P.

With this choice $\mathcal{R}_\alpha(P) = \{P_R :\ R \in \mathcal{R}_\alpha(P_0)\}$, $P_R = R \circ T^{-1}$

Common trimming

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

$$\mathcal{T}_0(P, Q) = \min_{R \in \mathcal{R}_\alpha(P_0)} d(P_R, Q_R)$$

$$P_{0,\alpha} = \underset{R \in \mathcal{R}_\alpha(P_0)}{\text{argmin}}\ d(P_R, Q_R)$$

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

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

$$\mathcal{T}_0(P, Q) = \min_{h \in \mathcal{C}_\alpha} d(P_h, Q_h)$$

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

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

Independent trimming

$$\begin{aligned}\mathcal{T}_1(P,Q) &:= \min_{R\in\mathcal{R}_\alpha(P)} d(R,Q),\\ \mathcal{T}_2(P,Q) &:= \min_{R_1\in\mathcal{R}_\alpha(P),R_2\in\mathcal{R}_\alpha(Q)} d(R_1,R_2),\end{aligned}$$

$$P_\alpha = \underset{R\in\mathcal{R}_\alpha(P)}{\text{argmin}}\ d(R,Q) \quad \text{best } \alpha\text{-trimming of } P \text{ for } Q$$

$$(P_\alpha,Q_\alpha) = \underset{(R_1,R_2)\in\mathcal{R}_\alpha(P)\times\mathcal{R}_\alpha(Q)}{\text{argmin}} d(R_1,R_2) \quad \text{best } \alpha\text{-matching of } P \text{ and } Q$$

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

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

Hence,

$$\mathcal{T}_1(P,Q)=0$$

Independent trimming

$$\mathcal{T}_1(P,Q) := \min_{R\in\mathcal{R}_\alpha(P)} d(R,Q),$$
$$\mathcal{T}_2(P,Q) := \min_{R_1\in\mathcal{R}_\alpha(P),R_2\in\mathcal{R}_\alpha(Q)} d(R_1,R_2),$$

$P_\alpha = \underset{R\in\mathcal{R}_\alpha(P)}{\text{argmin}}\ d(R,Q)$ best α-trimming of P for Q

$(P_\alpha,Q_\alpha) = \underset{(R_1,R_2)\in\mathcal{R}_\alpha(P)\times\mathcal{R}_\alpha(Q)}{\text{argmin}}\ d(R_1,R_2)$ best α-matching of P and Q

If d makes $\mathcal{R}_\alpha(P)$ closed

$$\mathcal{T}_2(P,Q) = 0 \quad \Leftrightarrow \quad d_{TV}(P,Q) \le \alpha$$

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

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

In general, (mild assumptions)

$$\mathcal{W}_2^2(P_R, Q_R) = \int \|T_P(x) - T_Q(x)\|^2 dR(x),$$

$$\frac{dP_{0,\alpha}}{dP_0} = \frac{1}{1-\alpha} I_{\{\|T_1 - T_2\| \leq c_\alpha(P,Q)\}}$$

and

$$\mathcal{T}^2(P,Q) = \int \|T_P(x) - T_Q(x)\|^2 dP_{0,\alpha}(x)$$

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

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

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

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

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

Incomplete transportation: $c(x,y) = \|x-y\|$

$X = Y = \mathbb{R}^k$

Theorem

$$\mathcal{W}_1(\mathcal{R}_\alpha(P), Q) = \sup_{f \leq 0; \|f\|_{Lip} \leq 1} \left(\frac{1}{1-\alpha} \int f dP - \int f dQ \right)$$

Incomplete transportation: $c(x,y) = \|x-y\|$

$X = Y = \mathbb{R}^k$

Theorem

$$\mathcal{W}_1(\mathcal{R}_\alpha(P), Q) = \sup_{f\leq 0; \|f\|_{Lip}\leq 1} \left(\frac{1}{1-\alpha} \int f dP - \int f dQ \right)$$

A simple consequence:

Corollary

$X_1, \ldots, X_n$ i.i.d. P, $D = diam(supp(P))$; P_n empirical measure.

$$\mathbb{P}\left(|\mathcal{W}_1(\mathcal{R}_\alpha(P_n), P) - E(\mathcal{W}_1(\mathcal{R}_\alpha(P_n), P))| > t \right) \leq 2e^{-\frac{2n(1-\alpha)^2 t^2}{D^2}}, t > 0$$

Dimension free concentration

Incomplete transportation: $c(x,y)=\|x-y\|^2$

$X=Y=\mathbb{R}^k$

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

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

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 $a_0(x)\ge\|x\|^2/2$, $x\in\mathbb{R}^n$ such that

$$\frac{1}{1-\alpha}\int a_0dP+\int a_0^*dQ=\min_a\left[\frac{1}{1-\alpha}\int adP+\int a^*dQ\right],$$

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

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

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

Theorem (Trim or move)

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

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

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

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: uniqueness

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

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

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.

Dual problem: uniqueness

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

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

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.

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

Theorem

If P or Q is absolutely continuous there exists a unique pair $(P_{\alpha_1},Q_{\alpha_2})\in\mathcal{R}_{\alpha_1}(P)\times\mathcal{R}_{\alpha_2}(Q)$ such that

$$\mathcal{W}_2(P_{\alpha_1},Q_{\alpha_2})=\mathcal{W}_2(\mathcal{R}_{\alpha_1}(P),\mathcal{R}_{\alpha_2}(Q))$$

provided $\mathcal{W}_2(\mathcal{R}_{\alpha_1}(P),\mathcal{R}_{\alpha_2}(Q))>0$

Trimmed comparisons

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

One sample problems:
Assume $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$ i.i.d. P and $Y_1, \dots, Y_m$ 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 $\mathcal{C}_\alpha(P,Q) = \{h \in \mathcal{C}_\alpha :\ d(P_h, Q_h) = 0\}$ (compact for $\|\cdot\|_\infty$)

Theorem

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

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

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 $\mathcal{C}_\alpha(P,Q) = \{h \in \mathcal{C}_\alpha : d(P_h, Q_h) = 0\}$ (compact for $\|\cdot\|_\infty$)

Theorem

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

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

Testing $\mathcal{T}^{(\alpha)}(P,Q) = 0$ equivalent to testing $\mathbb{P}(\varphi_1(Z) = \varphi_2(Z)) \geq 1 - \alpha$
$P = \mathcal{L}(\varphi_1(Z))$, $Q = \mathcal{L}(\varphi_2(Z))$

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

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

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

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

$\sigma_\alpha^2(P,Q)$ *consistently estimated by*

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

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

Test $H_0 : \mathcal{T}^{(\alpha)}(F,G) > \Delta_0^2$ against $H_a : \mathcal{T}^{(\alpha)}(F,G) \le \Delta_0^2$ (AE et al. 2008)

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

Asymptotics for $\mathcal{W}_2(\mathcal{R}_\alpha(P_n), Q)$, $(\mathcal{W}_2(\mathcal{R}_\alpha(P), Q) > 0)$

Theorem

$$\sqrt{n}(\mathcal{W}_2(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_2(\mathcal{R}_\alpha(P), Q)) \underset{w}{\to} \frac{1}{1-\alpha}\mathbb{G}_P(\varphi_\alpha)$$

Asymptotics for $\mathcal{W}_2(\mathcal{R}_\alpha(P_n),Q)$, $\quad (\mathcal{W}_2(\mathcal{R}_\alpha(P),Q)>0)$

Theorem

$$\sqrt{n}(\mathcal{W}_2(\mathcal{R}_\alpha(P_n),Q)-\mathcal{W}_2(\mathcal{R}_\alpha(P),Q)) \underset{w}{\rightarrow} \frac{1}{1-\alpha}\mathbb{G}_P(\varphi_\alpha)$$

Similarly, for $\mathcal{W}_2(\mathcal{R}_\alpha(P_n),\mathcal{R}_\alpha(Q_m))$, $\quad (\mathcal{W}_2(\mathcal{R}_\alpha(P),\mathcal{R}_\alpha(Q))>0)$

Theorem

$$\sqrt{n}(\mathcal{W}_2(\mathcal{R}_\alpha(P_n),\mathcal{R}_\alpha(Q_n))-\mathcal{W}_2(\mathcal{R}_\alpha(P),\mathcal{R}_\alpha(Q)) \underset{w}{\rightarrow} \frac{1}{1-\alpha}\left(\mathbb{G}_P(\varphi_\alpha)+\mathbb{G}_Q(\psi_\alpha)\right)$$

φ_α, ψ_α optimizers of dual problem
$\mathbb{G}_P$, $\mathbb{G}_Q$ independent P, Q-Brownian bridges

Asymptotics for $\mathcal{W}_2(\mathcal{R}_\alpha(P_n), Q)$, $\quad (\mathcal{W}_2(\mathcal{R}_\alpha(P), Q) > 0)$

Theorem

$$\sqrt{n}(\mathcal{W}_2(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_2(\mathcal{R}_\alpha(P), Q)) \underset{w}{\to} \frac{1}{1-\alpha} \mathbb{G}_P(\varphi_\alpha)$$

Similarly, for $\mathcal{W}_2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_m))$, $\quad (\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) > 0)$

Theorem

$$\sqrt{n}(\mathcal{W}_2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_n)) - \mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) \underset{w}{\to} \frac{1}{1-\alpha} (\mathbb{G}_P(\varphi_\alpha) + \mathbb{G}_Q(\psi_\alpha))$$

φ_α, ψ_α optimizers of dual problem
$\mathbb{G}_P$, $\mathbb{G}_Q$ independent P, Q-Brownian bridges

Usable for testing $H_0 : \mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) \leq \Delta_0 \quad$ (for fixed $\Delta_0 > 0$)

(in general dimension)

Sketch of proof ($k = 1$)

A *trimming process*: $\mathbb{V}_n(h) = \sqrt{n}\left(\mathcal{W}_2^2((P_n)_h, Q) - \mathcal{W}_2^2(P_h, Q)\right), \quad h \in \mathcal{C}_\alpha$

Define $\mathbb{V}(h) = 2\int_0^1 \frac{B(t)}{f(F^{-1}(t))}(F^{-1}(t) - G^{-1}(h(t)))h'(t)\,dt, \quad h \in \mathcal{C}_\alpha,$

$B(t)$ Brownian bridge on $(0,1)$; $\quad \mathbb{V}$ centered Gaussian process.

Sketch of proof ($k = 1$)

A *trimming process*: $\mathbb{V}_n(h) = \sqrt{n}\left(\mathcal{W}_2^2((P_n)_h, Q) - \mathcal{W}_2^2(P_h, Q)\right), \quad h \in \mathcal{C}_\alpha$

Define $\mathbb{V}(h) = 2\int_0^1 \frac{B(t)}{f(F^{-1}(t))}(F^{-1}(t) - G^{-1}(h(t)))h'(t)\,dt, \quad h \in \mathcal{C}_\alpha,$

$B(t)$ Brownian bridge on $(0,1)$; $\quad \mathbb{V}$ centered Gaussian process.

Theorem

Under mild assumptions $\mathbb{V}$ *is a tight, Borel measurable map into* $\ell^\infty(\mathcal{C}_\alpha)$ *and* $\mathbb{V}_n$ *converges weakly to* $\mathbb{V}$ *in* $\ell^\infty(\mathcal{C}_\alpha)$.

Sketch of proof ($k = 1$)

A *trimming process*: $\mathbb{V}_n(h) = \sqrt{n}\left(\mathcal{W}_2^2((P_n)_h, Q) - \mathcal{W}_2^2(P_h, Q)\right), \quad h \in \mathcal{C}_\alpha$

Define $\mathbb{V}(h) = 2\int_0^1 \frac{B(t)}{f(F^{-1}(t))}(F^{-1}(t) - G^{-1}(h(t)))h'(t)\,dt, \quad h \in \mathcal{C}_\alpha,$

$B(t)$ Brownian bridge on $(0,1)$; $\mathbb{V}$ centered Gaussian process.

Theorem

Under mild assumptions $\mathbb{V}$ is a tight, Borel measurable map into $\ell^\infty(\mathcal{C}_\alpha)$ and $\mathbb{V}_n$ converges weakly to $\mathbb{V}$ in $\ell^\infty(\mathcal{C}_\alpha)$.

$$
\begin{aligned}
\sqrt{n}(\mathcal{W}_2^2(P_{n,\alpha}, Q) - \mathcal{W}_2^2(P_\alpha, Q)) &= \sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2(P_{h_\alpha}, Q)) \\
&= \mathbb{V}_n(h_\alpha) + \sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2((P_n)_{h_\alpha}, Q)).
\end{aligned}
$$

Sketch of proof ($k = 1$)

A *trimming process*: $\mathbb{V}_n(h) = \sqrt{n}\left(\mathcal{W}_2^2((P_n)_h, Q) - \mathcal{W}_2^2(P_h, Q)\right), \quad h \in \mathcal{C}_\alpha$

Define $\mathbb{V}(h) = 2\int_0^1 \frac{B(t)}{f(F^{-1}(t))}(F^{-1}(t) - G^{-1}(h(t)))h'(t)\,dt, \quad h \in \mathcal{C}_\alpha$,

$B(t)$ Brownian bridge on $(0,1)$; $\mathbb{V}$ centered Gaussian process.

Theorem

Under mild assumptions $\mathbb{V}$ is a tight, Borel measurable map into $\ell^\infty(\mathcal{C}_\alpha)$ and $\mathbb{V}_n$ converges weakly to $\mathbb{V}$ in $\ell^\infty(\mathcal{C}_\alpha)$.

$$\begin{aligned}\sqrt{n}(\mathcal{W}_2^2(P_{n,\alpha}, Q) - \mathcal{W}_2^2(P_\alpha, Q)) &= \sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2(P_{h_\alpha}, Q))\\ &= \mathbb{V}_n(h_\alpha) + \sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2((P_n)_{h_\alpha}, Q)).\end{aligned}$$

$$\begin{aligned}\sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2((P_n)_{h_\alpha}, Q)) - \sqrt{n}(\mathcal{W}_2^2(P_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2(P_{h_\alpha}, Q))\\ = \mathbb{V}_n(h_{n,\alpha}) - \mathbb{V}_n(h_\alpha) \to 0.\end{aligned}$$

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi, \psi; P_n, Q) - J_\alpha(\varphi, \psi; P, Q))$$

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi, \psi; P_n, Q) - J_\alpha(\varphi, \psi; P, Q)) = \frac{1}{1-\alpha}\mathbb{G}_n(\varphi), \quad \varphi \in \Phi_c$$

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi,\psi;P_n,Q) - J_\alpha(\varphi,\psi;P,Q)) = \frac{1}{1-\alpha}\mathbb{G}_n(\varphi), \quad \varphi \in \Phi_c$$

If $\varphi_{n,\alpha}$, φ_α maximizers, for some $r_{n,i} \geq 0$

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n),Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P),Q)) = M_n(\varphi_\alpha) + r_{n,1} = M_n(\varphi_{n,\alpha}) - r_{n,2},$$

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi, \psi; P_n, Q) - J_\alpha(\varphi, \psi; P, Q)) = \frac{1}{1-\alpha}\mathbb{G}_n(\varphi), \quad \varphi \in \Phi_c$$

If $\varphi_{n,\alpha}$, φ_α maximizers, for some $r_{n,i} \geq 0$

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) = M_n(\varphi_\alpha) + r_{n,1} = M_n(\varphi_{n,\alpha}) - r_{n,2},$$

If Φ_c is Donsker and φ_α is unique

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) \underset{w}{\rightarrow} \frac{1}{1-\alpha}\mathbb{G}_P(\varphi_\alpha)$$

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi, \psi; P_n, Q) - J_\alpha(\varphi, \psi; P, Q)) = \frac{1}{1-\alpha}\mathbb{G}_n(\varphi), \quad \varphi \in \Phi_c$$

If $\varphi_{n,\alpha}$, φ_α maximizers, for some $r_{n,i} \geq 0$

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) = M_n(\varphi_\alpha) + r_{n,1} = M_n(\varphi_{n,\alpha}) - r_{n,2},$$

If Φ_c is Donsker and φ_α is unique

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) \underset{w}{\to} \frac{1}{1-\alpha}\mathbb{G}_P(\varphi_\alpha)$$

Φ_c usually not Donsker for large k (even if is not too large)

Sketch of proof (general k; general cost)

Dual *trimming process*:

$$\mathbb{M}_n(\varphi) = \sqrt{n}(J_\alpha(\varphi, \psi; P_n, Q) - J_\alpha(\varphi, \psi; P, Q)) = \frac{1}{1-\alpha}\mathbb{G}_n(\varphi), \quad \varphi \in \Phi_c$$

If $\varphi_{n,\alpha}$, φ_α maximizers, for some $r_{n,i} \geq 0$

$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) = M_n(\varphi_\alpha) + r_{n,1} = M_n(\varphi_{n,\alpha}) - r_{n,2},$$

If Φ_c is Donsker and φ_α is unique

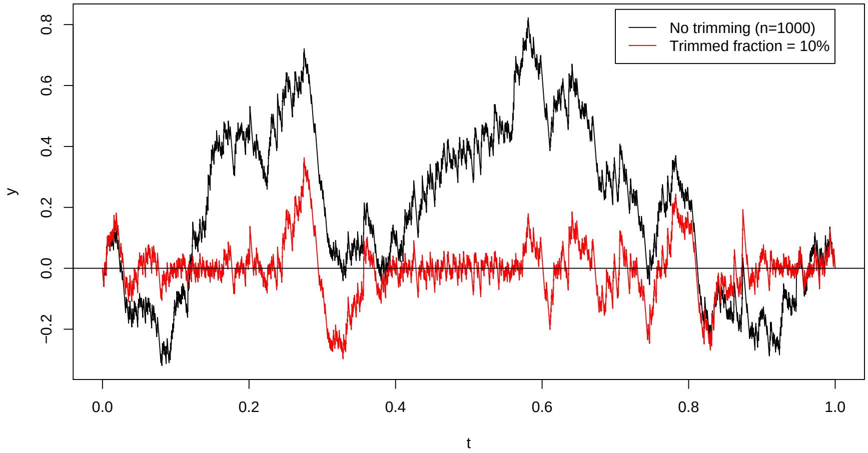
$$\sqrt{n}(\mathcal{W}_c(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_c(\mathcal{R}_\alpha(P), Q)) \underset{w}{\rightarrow} \frac{1}{1-\alpha}\mathbb{G}_P(\varphi_\alpha)$$

Φ_c usually not Donsker for large k (even if is not too large)

But, under smoothness, Φ_c can be replaced by a smaller class! Work in progress.

Overfitting effects of independent trimming.

Trajectories of uniform empirical process: $\sqrt{n}(G_n(t) - t)$ and α-trimmed uniform empirical process: $\sqrt{n}(G_{n,\alpha}(t) - t)$ $(n = 1000, \alpha = 0.1)$



Trimming & overfitting

Trimming increases the rate of convergence of $P_{n,\alpha}$ to P

$X_1, \ldots, X_n$ i.i.d. $P \quad (X_i \in \mathbb{R}^k)$

$$\mathcal{W}_2(P_n, P) \leq \mathcal{W}_2(P_{n,\alpha}, P) \leq \mathcal{W}_2(P_{n,1}, P)$$

Theorem If $k = 1$ $n\mathcal{W}_2^2(P_n, P) = O_P(1)$.

$n\mathcal{W}_2^2(P_{n,\alpha}, P) = o_P(1)$, $0 < \alpha \leq 1$

Trimming & overfitting

Trimming increases the rate of convergence of $P_{n,\alpha}$ to P

$X_1, \ldots, X_n$ i.i.d. $P \quad (X_i \in \mathbb{R}^k)$

$$\mathcal{W}_2(P_n, P) \le \mathcal{W}_2(P_{n,\alpha}, P) \le \mathcal{W}_2(P_{n,1}, P)$$

Theorem If $k = 1$ $n\mathcal{W}_2^2(P_n, P) = O_P(1)$.

$$n\mathcal{W}_2^2(P_{n,\alpha}, P) = o_P(1),\ 0 < \alpha \le 1$$

Theorem

$$n^{2/k}\ \mathsf{E}(\mathcal{W}_2^2(P_{n,1}, P)) \to c_k \int f(x)^{1-2/k} dx.$$

For $k \ge 3$, $n^{2/k}\ \mathsf{E}(\mathcal{W}_2^2(P_n, P)) = O(1)$.

Overfitting occurs only in low dimension!

Trimming & overfitting

Trimming increases the rate of convergence of $P_{n,\alpha}$ to P

$X_1, \ldots, X_n$ i.i.d. $P \quad (X_i \in \mathbb{R}^k)$

$$\mathcal{W}_2(P_n, P) \leq \mathcal{W}_2(P_{n,\alpha}, P) \leq \mathcal{W}_2(P_{n,1}, P)$$

Theorem If $k = 1$ $n\mathcal{W}_2^2(P_n, P) = O_P(1)$.

$$n\mathcal{W}_2^2(P_{n,\alpha}, P) = o_P(1),\ 0 < \alpha \leq 1$$

Theorem

$$n^{2/k}\ \mathsf{E}(\mathcal{W}_2^2(P_{n,1}, P)) \to c_k \int f(x)^{1-2/k} dx.$$

For $k \geq 3$, $n^{2/k}\ \mathsf{E}(\mathcal{W}_2^2(P_n, P)) = O(1)$.

Overfitting occurs only in low dimension!

But it is very significant: for $k = 1$ and $\nu > 1$

$$\frac{n^2}{(\log n)^{2\nu}} \mathcal{W}_2^2(P_{n,\alpha}, P) \underset{\mathsf{Pr}}{\to} 0$$

A random allocation problem

P_N uniform distribution on $\{a_1, \ldots, a_N\}$; $X_1, \ldots, X_n$ i.i.d. P_N; P_n empirical m.

A random allocation problem

P_N uniform distribution on $\{a_1, \ldots, a_N\}$; $X_1, \ldots, X_n$ i.i.d. P_N; P_n empirical m.

$P_n = \frac{1}{n}\sum_{i=1}^{N} B_i \delta_{a_i}$; $(B_1, \ldots, B_N) \sim \mathcal{M}(n; \frac{1}{N}, \ldots, \frac{1}{N})$

A random allocation problem

P_N uniform distribution on $\{a_1,\ldots,a_N\}$; $X_1,\ldots,X_n$ i.i.d. P_N; P_n empirical m.

$P_n=\frac{1}{n}\sum_{i=1}^N B_i\delta_{a_i};\quad (B_1,\ldots,B_N)\sim\mathcal{M}(n;\frac{1}{N},\ldots,\frac{1}{N})$

$P_N\in\mathcal{R}_\alpha(P_n)\Leftrightarrow\frac{1}{N}\leq\frac{1}{1-\alpha}\frac{B_i}{n},\ i=1,\ldots,N\Leftrightarrow\min_i B_i\geq(1-\alpha)\frac{n}{N}$

A random allocation problem

P_N uniform distribution on $\{a_1, \ldots, a_N\}$; $X_1, \ldots, X_n$ i.i.d. P_N; P_n empirical m.

$P_n = \frac{1}{n} \sum_{i=1}^N B_i \delta_{a_i}$; $(B_1, \ldots, B_N) \sim \mathcal{M}(n; \frac{1}{N}, \ldots, \frac{1}{N})$

$P_N \in \mathcal{R}_\alpha(P_n) \Leftrightarrow \frac{1}{N} \leq \frac{1}{1-\alpha} \frac{B_i}{n}$, $i = 1, \ldots, N \Leftrightarrow \min_i B_i \geq (1-\alpha) \frac{n}{N}$

For fixed N: a.s. $\mathcal{W}_2(\mathcal{R}_\alpha(P), P_N) = 0$ for large n

A random allocation problem

P_N uniform distribution on $\{a_1, \ldots, a_N\}$; $X_1, \ldots, X_n$ i.i.d. P_N; P_n empirical m.

$P_n = \frac{1}{n}\sum_{i=1}^N B_i \delta_{a_i}$; $(B_1, \ldots, B_N) \sim \mathcal{M}(n; \frac{1}{N}, \ldots, \frac{1}{N})$

$P_N \in \mathcal{R}_\alpha(P_n) \Leftrightarrow \frac{1}{N} \leq \frac{1}{1-\alpha}\frac{B_i}{n}$, $i = 1, \ldots, N \Leftrightarrow \min_i B_i \geq (1-\alpha)\frac{n}{N}$

For fixed N: a.s. $\mathcal{W}_2(\mathcal{R}_\alpha(P), P_N) = 0$ for large n

Random allocation + Discretization: For any $\nu > 1/k$

$$\mathcal{W}_2(\mathcal{R}_\alpha(P_n), P) = o_P\left(\frac{(\log n)^\nu}{n^{1/k}}\right).$$

Works for other metrics!