Limit theorems for multiple integrals with respect to the empirical process

Hélène Boistard

Joint work with
Eustasio del Barrio (Universidad de Valladolid)

SANTOUVAL 2008 - Castro-Urdiales - 30-01-2008

The Wasserstein test

The Wasserstein distance

$$\mathcal{P}_2(\mathbb{R}) = \{\text{prob. measures on } \mathbb{R} \text{ with finite 2nd moment}\}$$

$$\begin{aligned}
\mathcal{W}(P_1, P_2) &= \inf\left\{\left[E(X_1 - X_2)^2\right]^{1/2} : \mathcal{L}(X_1) = P_1, \mathcal{L}(X_2) = P_2\right\} \\
&= \left(\int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t)\right)^2 dt\right)^{1/2}.
\end{aligned}$$

Distance to a location-scale family

F a distribution function (d.f.) with mean 0 and variance 1.

$$\mathcal{H}_F = \left\{ H : H(x) = F\left(\frac{x-\mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma > 0 \right\}.$$

$P \in \mathcal{P}_2(\mathbb{R})$ with d.f. F_0, and standard deviation σ_0:

$$\begin{aligned} \mathcal{W}^2(P, \mathcal{H}_F) &= \inf \left\{ \mathcal{W}^2(P, H) : H \in \mathcal{H}_F \right\} \\ &= \sigma_0^2 - \left(\int_0^1 F_0^{-1}(t) F^{-1}(t) dt \right)^2. \end{aligned}$$

Empirical version

- $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$,
- S_n^2 the empirical variance,
- $F_n^{-1}(t) = X_{(i)},$ si $\frac{i-1}{n} < t \leq \frac{i}{n}$ the empirical quantile function
- $X_{(1)} \leq \cdots \leq X_{(n)}$ the order statistics.

$$
\begin{aligned}
\mathcal{R}_n &= \frac{\mathcal{W}^2(\mathbb{P}_n, \mathcal{H}_F)}{S_n^2} \\
&= \frac{1}{S_n^2}\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right)^2 dt - \frac{1}{S_n^2}\left(\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right) dt\right)^2 \\
&\quad -\frac{1}{S_n^2}\left(\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right) F^{-1}(t)dt\right)^2
\end{aligned}
$$

The Wasserstein normality test (cf. del Barrio et al. (1999))

φ the standard normal density, Φ^{-1} the standard normal quantile function, $\rho_n(t) = \sqrt{n}\varphi(\Phi^{-1}(t))\left(F_n^{-1}(t) - \Phi^{-1}(t)\right)$

$$nS_n^2\mathcal{R}_n = \int_0^1 \left(\frac{\rho_n(t)}{\varphi(\Phi^{-1}(t))}\right)^2 dt - \left(\int_0^1 \frac{\rho_n(t)}{\varphi(\Phi^{-1}(t))}dt\right)^2 - \left(\int_0^1 \frac{\rho_n(t)\Phi^{-1}(t)}{\varphi(\Phi^{-1}(t))}dt\right)^2$$

$$nS_n^2\mathcal{R}_n - a_n \xrightarrow{w} \int_0^1 \frac{B^2(t) - EB^2(t)}{\varphi^2(\Phi^{-1}(t))}dt - \left(\int_0^1 \frac{B(t)}{\varphi(\Phi^{-1}(t))}dt\right)^2 - \left(\int_0^1 \frac{B(t)\Phi^{-1}(t)}{\varphi(\Phi^{-1}(t))}dt\right)^2,$$

where $a_n = \int_{1/n}^{1-1/n} \frac{t(1-t)}{\varphi^2(\Phi^{-1}(t))}dt$.

Equivalence with a U-statistic

- $u_n(t) = \sqrt{n}\left(\Phi(F_n^{-1}(t)) - t\right)$ the uniform quantile process, $\beta_n(t) = \sqrt{n}\left(F_n(\Phi^{-1}(t)) - t\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\Phi(X_i)\le t} - t\right)$ the uniform empirical process

- Kernel for the normality test:

$$\begin{aligned} h(x,y) &= \int_0^1 \frac{\left(\mathbf{1}_{(\Phi(x)\le t)} - t\right)\left(\mathbf{1}_{(\Phi(y)\le t)} - t\right)}{\varphi^2(\Phi^{-1}(t))} dt, \\ \bar{h}(x,y) &= h(x,y) - xy - \frac{1}{4}(x^2-1)(y^2-1). \end{aligned}$$

Theorem 1. *For a $N(0,1)$ i.i.d. sample :*

$$n\mathcal{R}_n - a_n = \frac{1}{n}\sum_{i\neq j}\bar{h}(X_i, X_j) - \frac{3}{2} + o_P(1).$$

Multiple integrals with respect to the empirical process

Definition (Cf. Major (2006, 2007).)

$$J_{n,m}(h) = \int' h(x_1, \ldots, x_m)\, d\mathbb{G}_n(x_1) \cdots d\mathbb{G}_n(x_m), \text{ where}$$

- $h(x_1, ..., x_m)$ real-valued symmetric function on $\mathcal{X}^m$,
- $X_1, \ldots, X_n$ v.a. i.i.d. sur $(X, \mathcal{X})$ with distribution P,
- $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ the normalized empirical process,
- $\int'$ the integral out of the diagonal $x_r = x_s$, $1 \leq r < s \leq m$.

Link with U-statistics

$$U_n(h) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}).$$

For completely degenerate h :

$$J_{n,m}(h) = \frac{n!}{(n-m)!n^{m/2}} U_n(h).$$

CLT for U-statistics:

Rubin and Vitale (1980) : CLT for completely degenerate U-statistics.

$$J_{n,m}(h) = \frac{n!}{(n-m)!n^{m/2}} U_n(h) \to I_m(h),$$

with $I_m(h)$: the multiple integral of h with respect to the Brownian motion.

For non completely degenerate U-statistics : CLT obtained via the Hoeffding decomposition (cf. Arcones and Giné (1992) or de la Peña and Giné (1999)).

CLT for multiple integrals with respect to the empirical process

$\pi_j h$ the Hoeffding projection of h, defined as:

$$\pi_j h(x_1,\dots,x_j) = \begin{cases} (\delta_{x_1}-P)\times\cdots\times(\delta_{x_j}-P)\times P^{m-j}h & \text{for} \quad j=1,\dots,m \\ P^m h & \text{for } j=0. \end{cases}$$

Results by Major (2006,2007)

- $J_{n,m}(h) = \sum_{j=0}^{m} K_{n,j,m} J_{n,j}(\pi_j h)$ with $K_{n,j,m} \to K_{j,m}, n \to \infty$,

- Consequence :

$$J_{n,m}(h) \xrightarrow{w} \sum_{j=0}^{m} K_{j,m} I_j(\pi_j h).$$

Integrals with respect to the Brownian bridge

For $m = 2$ and P the uniform distribution on $(0, 1)$:

$$\int'_{[0,1]^2} h(x_1, x_2) d\mathbb{G}_n(x_1) d\mathbb{G}_n(x_2) \underset{w}{\rightarrow} \int_{[0,1]^2} h(x_1, x_2) dB(x_1) dB(x_2),$$

with $\{B(x)\}_{x\in(0,1)}$ a Brownian bridge.

Expansion using $B(x) = W(x) - xW(1)$

with $\{W(x), x \in (0, 1)\}$ a Brownian bridge:

$$\begin{aligned}\int_{[0,1]^m} h(x_1, x_2) dB(x_1) dB(x_2) \quad = \quad & \int_{[0,1]^2} h(x_1, x_2) dW(x_1) dW(x_2) \\ & -2 \int_{[0,1]} \left(\int h(x_1, t) dt \right) dW(x_1) W(1) \\ & + \left(\int h(s, t) ds dt \right) W(1)^2\end{aligned}$$

Multiple integrals with respect to the Brownian bridge

We define:

$$\rho_j h(x_1, \ldots, x_j) = \delta_{x_1} \times \cdots \times \delta_{x_j} \times P^{m-j} h$$

Definition 1. *For every symmetric function* $h \in L^2(P^m)$*,*

$$J_m(h) = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{m-j} I_j(\rho_j h)(I_1(1))^{m-j}.$$

Proposition 1. *For symmetric* $h \in L^2(P^m)$*,*

$$J_m(h) = \sum_{j=0}^{m} \frac{m!}{j!} H_{m-j}(0) I_j(\pi_j h),$$

where H_{m-j} *is the Hermite polynomial of degree* $m-j$*.*

Theorem 2. $(X_1, \ldots, X_n)$ *i.i.d. sample with distribution* P *without atoms and* $h_1, \ldots, h_k$ *square-integrable functions in* $m_1, \ldots, m_k$ *variables. Then:*

$$(J_{n,m_1}(h_1), \ldots, J_{n,m_k}(h_k)) \xrightarrow{w} (J_{m_1}(h_1), \ldots, J_{m_k}(h_k))\,.$$

Bootstrap of double integrals

$(X_1^*, \ldots, X_n^*)$ i.i.d. bootstrap sample, underlying distribution: $\mathbb{P}_n$.
New empirical measure: $\mathbb{P}_n^*$.

Bootstrap double integral:

$$J_{n,2}^*(h) = \int' h(x,y) d\mathbb{G}_n^*(x) d\mathbb{G}_n^*(y),$$

with $\mathbb{G}_n^* = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ the bootstrap empirical process.

Theorem 3. $h \in L_2(P^2)$, *P without atoms. Then:*

$$d_{LB}\left(\mathcal{L}^*\left(J_{n,2}^*(h)\right), \mathcal{L}\left(J_2(h)\right)\right) \to 0 \textit{ a.s. when } n \to \infty.$$

Limit under contiguous alternative

Gaussian shift experiment

$$(\Omega, \mathcal{A}, \{P_g, g \in H\}) \text{ with}$$

- $(H, \langle , \rangle)$ a Hilbert space,
- for all g, $P_g \ll P_0$, and $(L(g))_{g \in H}$ is defined as

$$\log \frac{dP_g}{dP_0} = L(g) - \frac{\|g\|^2}{2},$$

  with L standard Gaussian process under P_0, i.e.: centered and

$$\forall g_1, g_2 \in H,\ Cov(L(g_1), L(g_2)) = \langle g_1, g_2 \rangle.$$

Spectral decomposition

$H_0 = \{0\}$, $\eta : \Omega \to [0,1]$ a test with level α.

Taylor expansion, for $g \in H\backslash\{0\}$:

$$E_{tg}\eta = \alpha + b(g)t + a(g)\frac{t^2}{2} + o(t^2), t \to 0.$$

- $\exists h_0 \in H$ gradient, such that:

$$b(g) = \langle g, h_0 \rangle, \forall g \in H.$$

- $\exists K_\alpha : H \to H$ a Hilbert-Schmidt self-adjoint operator, a system of orthonormal eigenfunctions (h_i) and eigenvalues (λ_i) such that:

$$a(g) = \langle g, K_\alpha g \rangle \forall g \in H, \text{ and } K_\alpha = \sum_{i=1}^{+\infty} \lambda_i h_i \otimes h_i.$$

Local asymptotic efficiency

- One-sided test : η test with level α in the Gaussian shift experiment. Alternative : $K \setminus \{0\}$, with K a cone.

$$ARE_L^{(1)}(\eta, g) = \left(\frac{\langle g, h_0 \rangle}{\|g\| \varphi \left(\Phi^{-1} (1 - \alpha) \right)} \right)^2 .$$

- Two-sided test: η unbiased test with level α for $H_0 = \{0\}$ versus $H \setminus \{0\}$, i.e.:

$$E_0 \eta = \alpha \text{ and } E_g \eta \geq \alpha \text{ for } g \in H \setminus \{0\}.$$

Then $h_0 = 0$ and $\lambda_i \geq 0 \; \forall i \geq 1$.

$$ARE_L^{(2)}(\eta, g) = \frac{\langle g, K_\alpha g \rangle}{2\|g\|^2 \varphi \left(\Phi^{-1} (1 - \alpha/2) \right) \Phi^{-1} (1 - \alpha/2)}.$$

Contiguous alternative

$(X_1, \ldots, X_n)$ with joint distribution P_n such that $\exists g \in L_2^0(P)$,

$$\log \frac{dP_n}{dP^n}(x_1, \ldots, x_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(x_i) - \frac{1}{2} P g^2 + \epsilon_n(x_1, \ldots, x_n),$$

where $\epsilon_n \to 0$ under P_n, $n \to \infty$.

$L_2^0(P) = \{g \in L_2(P), Pg = 0\}$.

Theorem 4. *$h \in L_2(P^m)$ and P_n a P-contiguous distribution. Then:*

$$J_{n,m}(h) \xrightarrow{w} \begin{cases} J_m(h) & \text{under } P, \\ J_m^g(h) & \text{under } P_n, \end{cases}$$

with $J_m^g(h)$ the multiple integral with respect to the shifted Brownian bridge

$$B^g(x) = B(x) + \int_{-\infty}^{x} g(u)dP(u).$$

Example

- $m = 2$, P uniform on $(0,1)$, $h = \sum_{j\geq 1} \lambda_j h_j \otimes h_j \in L_2^0(P^2)$, $g \in L_2^0(P)$.

- $I_2(h) = \sum_{j\geq 1} \lambda_j \left(\left(\int h_j(s) dW(s) \right)^2 - 1 \right)$

- $I_2^g(h) = \sum_{j\geq 1} \lambda_j \left(\left(\int h_j(s) dW(s) + \langle h_j, g \rangle \right)^2 - 1 \right)$

Test in the limit experiment

Limit Gaussian experiment (P the uniform distribution on $(0,1)$):

$$\left(\mathcal{C}([0,1]), \mathcal{B}(\mathcal{C}([0,1])), \{\mathcal{L}(W^g) : g \in L_2^0(0,1)\}\right).$$

Asymptotic test with level α based on the double stochastic integral of the kernel:

$$\eta_\alpha(x) = \mathbf{1}_{\int\int h(s,t)dx(s)dx(t) > C_\alpha},$$

with C_α such that $P\left(\int\int h(s,t)dW(s)dW(t) > C_\alpha\right) = \alpha$.

Curvature of double integrals

$h(s,t) = \sum_{j=1}^{+\infty} \lambda_j f_j(s) f_j(t)$, for an orthonormal basis $\{f_j\}_{j=1}^{+\infty}$ of $L_2^0(0,1) = \{1\}^\perp$ and decreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$.

- Gradient of η_α: 0.

- Curvature operator a_α of η_α:

$$a_\alpha(g_1, g_2) = \langle g_1, K_\alpha g_2 \rangle$$

with

$$K_\alpha(s,t) = \sum_{j=1}^{+\infty} \lambda_{j,\alpha} f_j(s) f_j(t),$$

Result by Dencker and Liese (2004):

- Several results of approximation of the power by the expansion at order 2.

- When $\sum \lambda_j^2 < \infty$, there exist d_1, d_2: $d_1\lambda_j \leq \lambda_{j,\alpha} \leq d_2\lambda_j \ \forall j$.

Theorem 5. *When $\sum \lambda_j^2 < \infty$, the eigenvalues of the curvature are sorted in the same order as the eigenvalues $\{\lambda_j\}$ of h, i.e.*

$$\lambda_{1,\alpha} \geq \lambda_{2,\alpha} \geq \cdots .$$

The Wasserstein test for normal mixtures

Theorem 6. *P the standard normal distribution and P_n a contiguous distribution.*

$$n\mathcal{R}_n - a_n \xrightarrow{w} \begin{cases} I_2(\bar{h}) - \frac{3}{2} & \text{under } P, \\ I_2^g(\bar{h}) - \frac{3}{2} & \text{under } P_n. \end{cases}$$

Asymptotic test

$$\bar{\eta}_\alpha(W^g) \;=\; \mathbf{1}_{\left\{I_2^g(\bar{h}) > \bar{c}_\alpha\right\}},$$

with $\bar{c}_\alpha$ chosen such that:

$$P\left(I_2(\bar{h}) > \bar{c}_\alpha\right) \;=\; \alpha.$$

Study of the kernel $\bar{h}$

$h_j = \sqrt{j!}H_j,\ j \geq 0.$

$$\begin{aligned} h_0(x) &= 1, h_1(x) = x \\ h_2(x) &= \frac{1}{\sqrt{2}}(x^2 - 1), h_3(x) = \frac{1}{\sqrt{6}}(x^3 - 3x). \end{aligned}$$

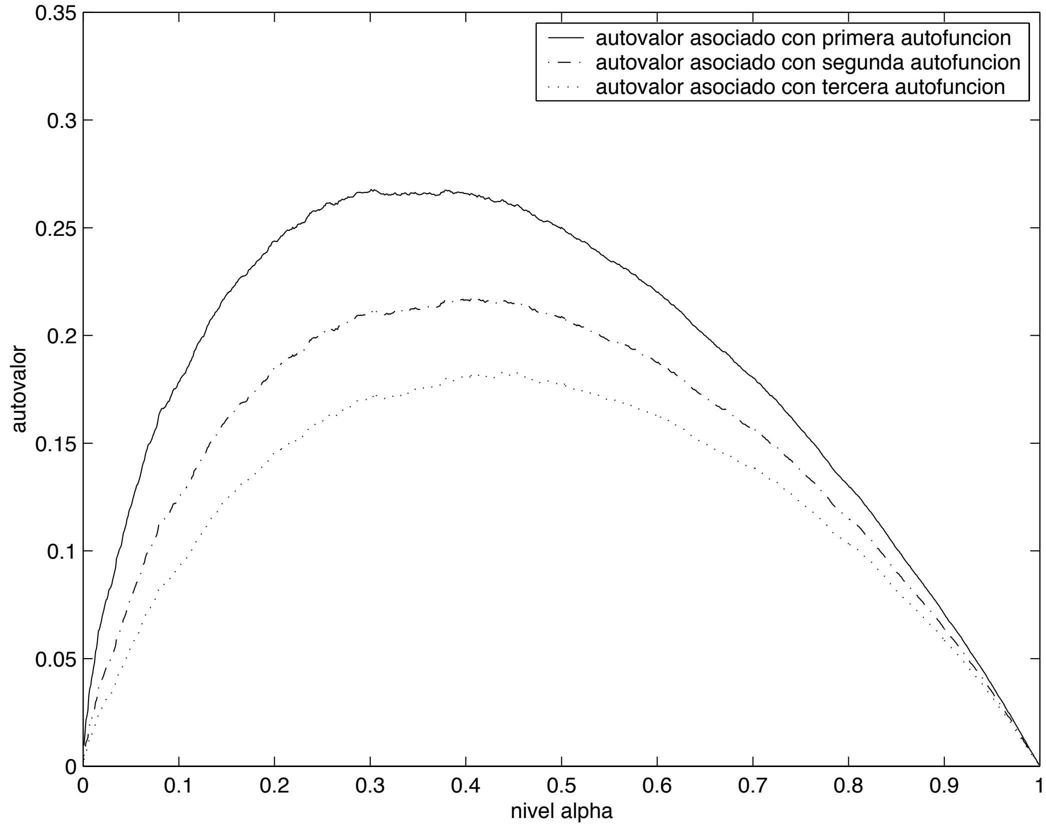
Proposition 2.

$$\bar{h}(x, y) = \sum_{j=3}^{+\infty} \frac{1}{j} h_j(x) h_j(y).$$

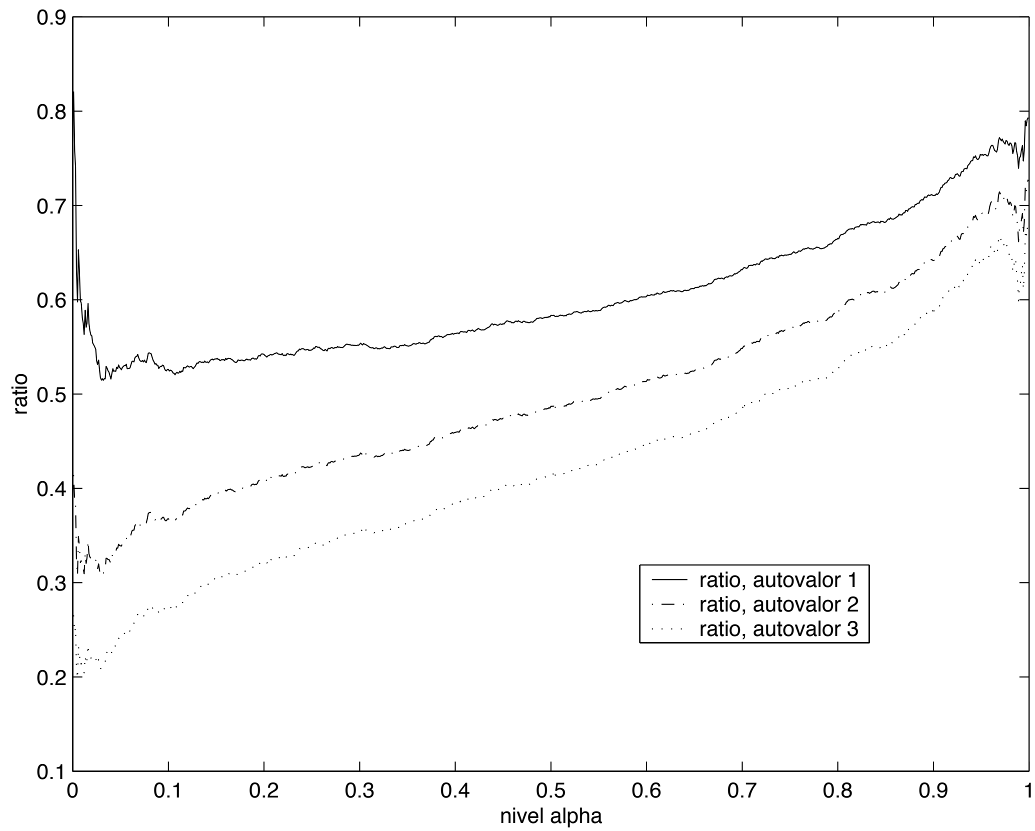
Theorem 7. $\bar{K}_\alpha$ *the curvature operator of* $\bar{\eta}_\alpha$.

$$\begin{aligned} \bar{K}_\alpha(s, t) &= \sum_{j=3}^{+\infty} \bar{\lambda}_{j,\alpha} h_j(s) h_j(t), \text{ where} \\ \bar{\lambda}_{j,\alpha} &= E\left((\bar{\eta}_\alpha(W) - \alpha) \left(\int h_j dW \right)^2 \right). \end{aligned}$$

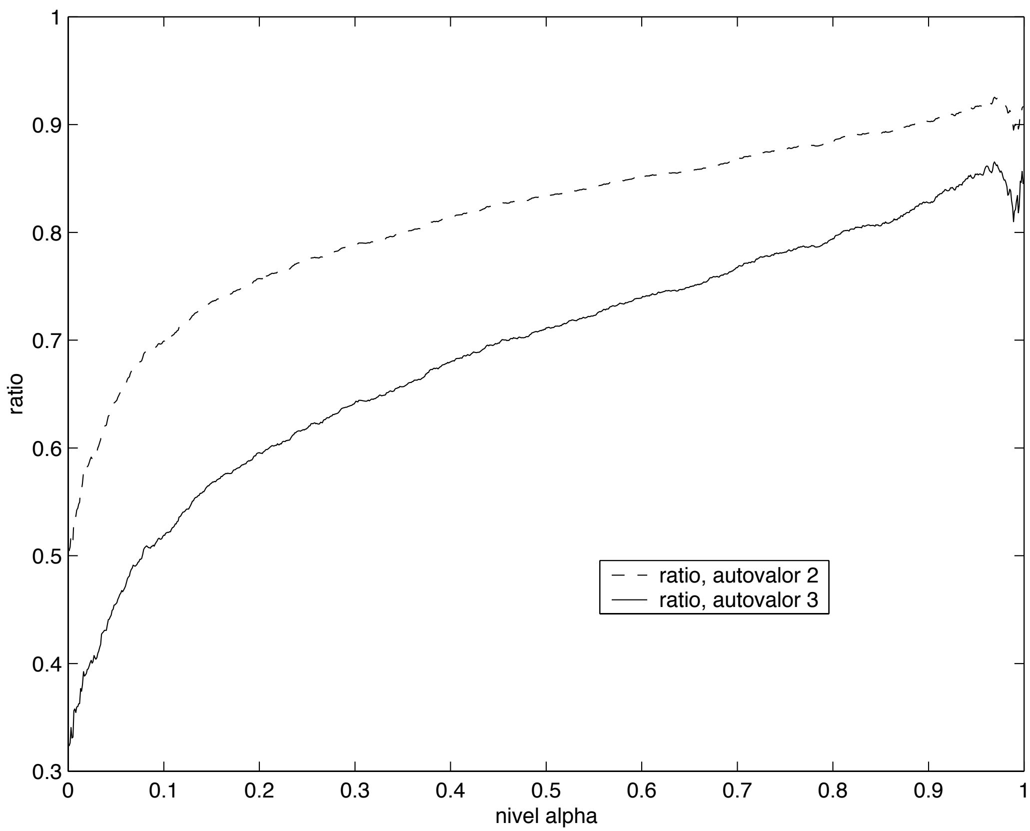
Eigenvalues of the curvature



Ratio by the Neyman-Pearson bound



Ratio by the greatest eigenvalue



Normal mixture $M_{p,\mu,\sigma}$: distribution function

$$\Phi_{p,\mu,\sigma}(x) = (1-p)\Phi(x) + p\Phi\left(\frac{x-\mu}{\sigma}\right).$$

- $M_{p,\mu/\sqrt{n},1}$ contiguous, with asymptotic direction:

$$g : x \mapsto p\mu x.$$

- $M_{p/\sqrt{n},\mu,1}$ contiguous, with asymptotic direction:

$$g : x \mapsto p\frac{\varphi(x-\mu)-\varphi(x)}{\varphi(x)}.$$

Example of the skewness test

$$l_n = \frac{n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^3}{\left(S_n^2\right)^{3/2}},$$

with $\bar{X}_n$ the empirical mean, S_n^2 the empirical variance.

$$\sqrt{n} l_n \to \begin{cases} \int \left(x^3 - 3x\right) dW(x) & \text{under the null hypothesis (normality),} \\ \int \left(x^3 - 3x\right) dW^g(x) & \text{under contiguous alternative } P_n. \end{cases}$$

Asymptotic test with level α :

$$\eta_\alpha^S\left(W^{P,g}\right) = \mathbf{1}_{\left|\int (x^3 - 3x) dW^{P,g}(x)\right| > C_\alpha}.$$

Null gradient, curvature a_α^S : eigenfunction and eigenvalue

$$h_3(x) = (x^3 - 3x)/\sqrt{6},\ \lambda_\alpha^S = 2\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right).$$

Comparison of powers

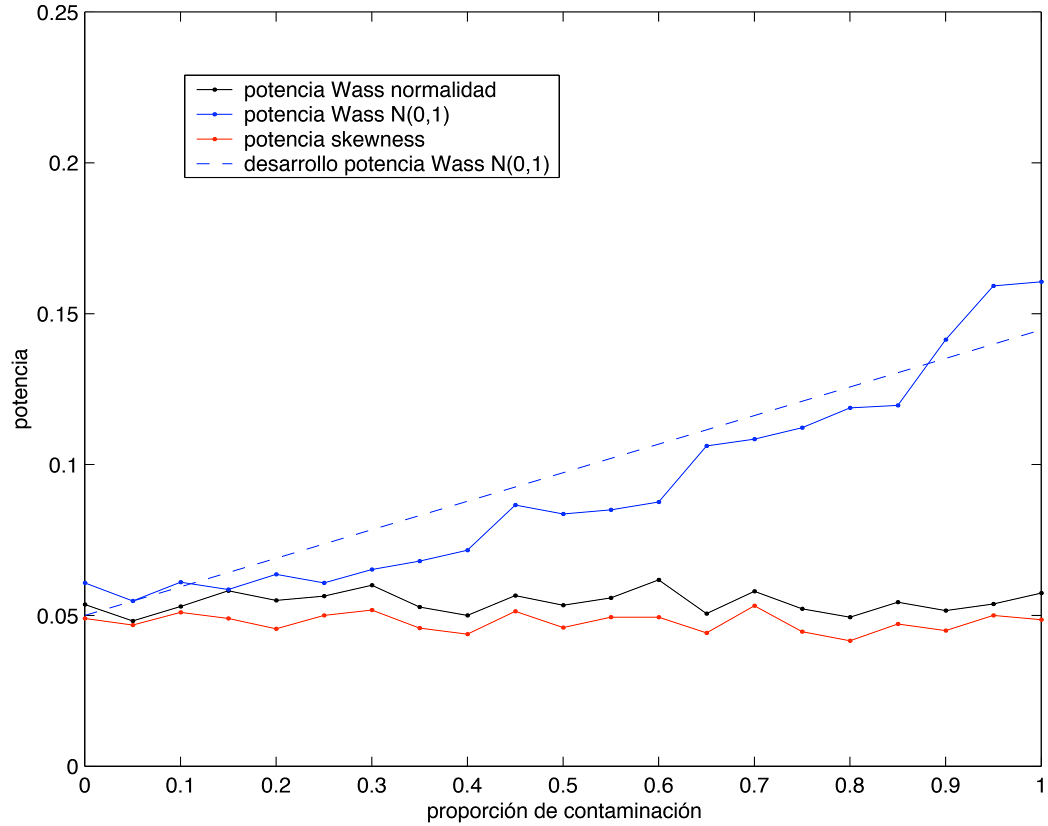
- Monte Carlo method for the power of the Wasserstein goodness of fit test for $N(0,1)$ and the Wasserstein normality test

- Monte Carlo method for the skewness test

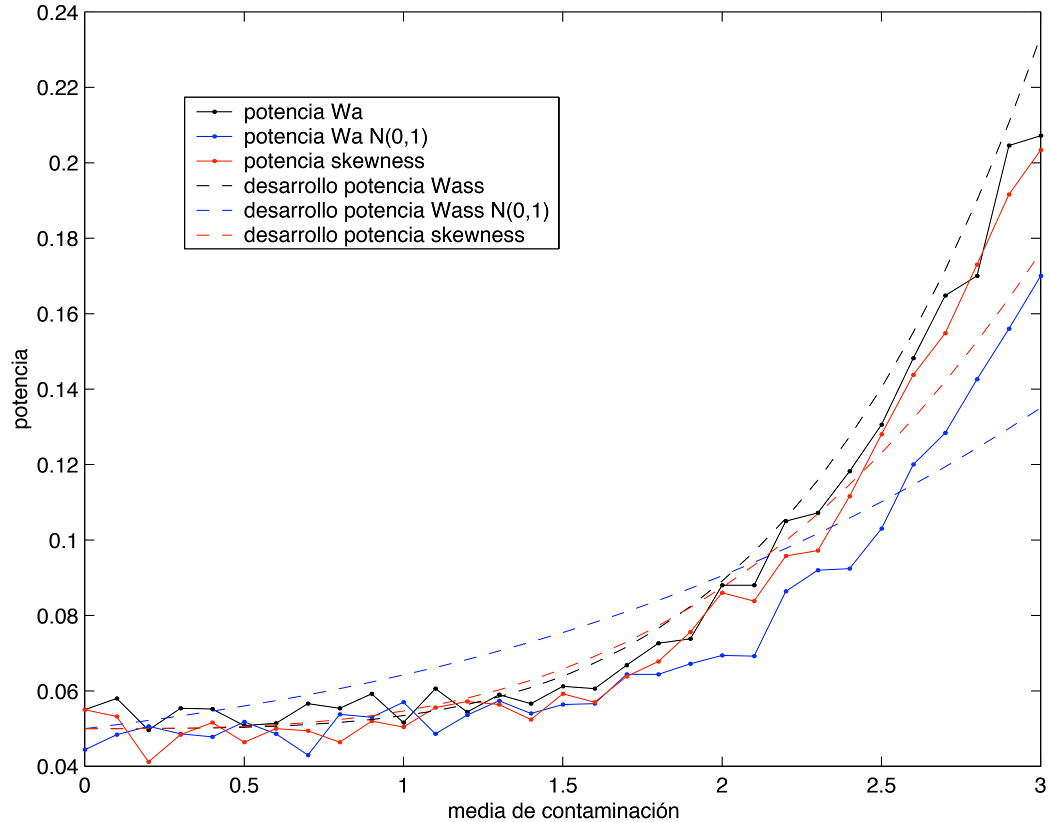
- Expansion at order 2 of the power owing to the curvature:

$$E_{tg}\eta = E_0\eta + tb(g) + \frac{t^2}{2}a(g) + o(t^2)$$

Mixture $M_{p,\mu/\sqrt{n},1}$, $\mu = 1$, $n = 100$.



Mixture $M_{p/\sqrt{n},\mu,1}$, $p = 0.1$.



To go further

- Multiple integral with respect to the empirical process: extension to non i.i.d. samples (for instance: dependance)

- Study of the multiple integral with respect to the Brownian bridge

- Bounds for the approximation of the power by the expansion of the curvature

- Tests designed for particular alternatives