Wasserstein's goodness of fit test under some local alternatives

Hélène Boistard, Eustasio del Barrio
Universidad de Valladolid, Spain

1.Wasserstein's normality test

Wasserstein's distance

$$\mathcal{P}_2(\mathbb{R}) = \{\text{Probability measures on } \mathbb{R} \text{ with finite second moment}\}$$

For $P_1, P_2 \in \mathcal{P}_2(\mathbb{R})$, the L_2 Wasserstein distance between P_1 and P_2 is

$$\mathcal{W}(P_1, P_2) = \inf \left\{ \left[E(X_1 - X_2)^2 \right]^{1/2} : \mathcal{L}(X_1) = P_1, \mathcal{L}(X_2) = P_2 \right\}.$$

Expression with the quantile functions F_1^{-1} and F_2^{-1} :

$$\mathcal{W}(P_1, P_2) = \left(\int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t) \right)^2 dt \right)^{1/2}.$$

Distance to a location-scale family

A location-scale family is obtained from a law with mean 0, variance 1 and distribution function F as follows :

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

If $P \in \mathcal{P}_2(\mathbb{R})$ with distribution function F_0, standard deviation σ_0 :

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

Wasserstein's test for normality

We denote by $\mathcal{N}$ the normal family, Φ the standard normal distribution function, Φ^{-1} the standard normal quantile function, ϕ the standard normal density function.

$X_1, \ldots, X_n$ an i.i.d. sample, F_n the empirical distribution function and F_n^{-1} the empirical quantile function. The test statistics is a normalized empirical version of the distance to the normal family.

$$\mathcal{R}_n := \frac{\mathcal{W}^2(F_n, \mathcal{N})}{S_n^2} = 1 - \frac{(\int_0^1 F_n^{-1}(t)\Phi^{-1}(t)dt)^2}{S_n^2}$$

Remark : to test fit to the standard normal law, the statistics is

$$\int_0^1 \left(F_n^{-1}(t) - \Phi^{-1}(t)\right)^2 dt.$$

2.Asymptotic distribution for the normality test

Asymptotic distribution under the null hypothesis

Theorem 1 (del Barrio et. al., 1999) *Suppose that* $(X_i)_{i=1}^n$ *is an i.i.d. sample with underlying normal law. Then*

$$n\mathcal{R}_n - \int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{t(1-t)}{\phi^2(\Phi^{-1}(t))}dt$$

converges in distribution to

$$\int_0^1 \frac{B^2(t) - EB^2(t)}{\phi^2(\Phi^{-1}(t))}dt - \left(\int_0^1 \frac{B(t)}{\phi(\Phi^{-1}(t))}dt\right)^2 - \left(\int_0^1 \frac{B(t)\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))}dt\right)^2.$$

where B *is a Brownian bridge on* $[0,1]$, *and the last integrals have to be understood as* L_2 *limits of the same integrals on* $[\frac{1}{n}, 1-\frac{1}{n}]$.

Principal component decomposition for the limit law

Call K the covariance kernel of the centered Gaussian process $\frac{B}{\phi \circ \Phi^{-1}}$.
For $s, t \in [0,1]$:

$$K(s,t) = \frac{s \wedge t - st}{\phi(\Phi^{-1}(s))\phi(\Phi^{-1}(t))}.$$

It satisfies $\int_0^1 \int_0^1 K^2(s,t) ds dt < +\infty$ and in particular, it is continuous on $L^2(0,1)$.

Decomposition in terms of eigenfunctions and eigenvalues :

$$K(s,t) = \sum_{j=0}^{+\infty} \frac{1}{j+1} f_j(s) f_j(t).$$

The eigenfunctions f_j form an orthonormal base of $L^2(0,1)$ related to normalized Hermite polynomials h_j by :

$$f_j(t) = h_j(\Phi^{-1}(t)), t \in [0,1].$$

The first normalized Hermite polynomials are :

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

Projecting $\frac{B}{\phi\circ\Phi^{-1}}1_{[\frac{1}{n},1-\frac{1}{n}]}$ on this orthonormal base provides the following expression for the limit law

$$n\mathcal{R}_n - \int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{t(1-t)}{\phi^2(\Phi^{-1}(t))}dt \xrightarrow{w} -\frac{3}{2} + \sum_{j=3}^{+\infty} \frac{Y_j^2-1}{j},$$

where $\{Y_j\}_j$ is a sequence of independent $N(0,1)$ random variables.

Asymptotic distribution under alternative

Theorem 2 *Suppose that after some possible change in the location or scale, the variables* $(X_i)_{i=1}^n$ *have distribution function* Φ_n *such that :*

$$h_n := \sqrt{n}\left(\Phi_n^{-1} - \Phi^{-1}\right) \overset{L_2(0,1)}{\longrightarrow} h \in L_2(0,1).$$

Moreover, we suppose that two additional conditions are satisfied, namely (a) and (b) above.

Then $n\mathcal{R}_n - \int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{t(1-t)}{\phi^2(\Phi^{-1}(t))}dt$ *converges in distribution to*

$$\int_0^1 \frac{B_h^2(t) - EB^2(t)}{\phi^2(\Phi^{-1}(t))}dt - \left(\int_0^1 \frac{B_h(t)}{\phi(\Phi^{-1}(t))}dt\right)^2 - \left(\int_0^1 \frac{B_h(t)\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))}dt\right)^2,$$

where $B_h(t) = B(t) + h.\phi \circ \Phi^{-1}$ *and* B *is a Brownian bridge.*

The conditions for the theorem are the following :

• (a) $\log\log n \int (h_n - h)^2 \to 0$ (additional speed condition)

• (b) $\log\log n \int (h \circ G_n^{-1} - h)^2 \to 0$ (additional regularity condition), where G_n^{-1} is the empirical quantile function associated with uniform i.i.d. random variables on $[0, 1]$.

3.Representation as a Gaussian shift

Model of Gaussian shift

We use the tools of the theory of Le Cam for statistical experiments, as exposed in Strasser, 1985. $(H, \langle , \rangle)$ denotes a real Hilbert space. An experiment $(\Omega, \mathcal{A}, \{P_h, h \in H\})$ on H is a Gaussian shift experiment, if and only if for all h, $P_h \ll P_0$ and the process $(L(h))_{h \in H}$ defined by the log-likelihood ratio

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

is a standard Gaussian process under P_0, i.e. : it is centered and for any $h_1, h_2 \in H$

$$Cov(L(h_1), L(h_2)) = \langle h_1, h_2, \rangle.$$

Spectral decomposition

If $\varphi : \Omega \to [0,1]$ is a test function for testing the hypothesis $H_0 \subset H$ (for a linear subspace H_0, for instance $H_0 = \{0\}$), then its power function can be studied under straight lines of alternatives directed by some $h \in H \backslash \{0\}$ (Janssen, 1995). Suppose that φ is α H_0-similar. The function $t \mapsto E_{th}\varphi$ admits the following Taylor expansion :

$$E_{th}\varphi = \alpha + b(h)t + a(h)\frac{t^2}{2} + o(t^2), t \to 0. \tag{1}$$

Theorem 3 *1. There exists a* gradient h_0 *such that :*

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

2. There exists a self-adjoint Hilbert-Schmidt operator $T : H \to H$*, an ortonormal system* (h_i) *and eigenvectors* (λ_i) *such that :*

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

Efficiency

Theorem 4 *We sum up some important properties from Janssen, 1995, based on the generalized Neyman Pearson lemma.*
1.$\|h_0\| \leq \phi(\Phi^{-1}(1-\alpha))$.
The equality holds iff

$$\varphi = 1_{\{L(h_0) > \Phi^{-1}(1-\alpha)\|h_0\|\}}.$$

2.The largest eigenvalue of T *satisfies the inequality*

$$|\lambda_1| \leq 2\phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)\Phi^{-1}\left(1-\frac{\alpha}{2}\right).$$

The equality holds iff

$$\varphi = 1_{\{|L(h_1)| > \Phi^{-1}(1-\frac{\alpha}{2})\}}.$$

Janssen (1995) introduces a concept of local asymptotic relative efficiency :

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

The equality holds iff φ is the one-sided Neyman Pearson test in direction h.

$$ARE_L^{(2)}(\varphi, h) = \frac{\langle h, T(h) \rangle}{2\|h\|^2 \phi(\Phi^{-1}(1-\frac{\alpha}{2}))\Phi^{-1}(1-\frac{\alpha}{2})}.$$

The equality holds iff φ is the two-sided Neyman Pearson test in direction h.

Gaussian shift for Wasserstein's test

$(W(t)_{t\in[0,1]})$ a Brownian motion on a probability space with probability measure P, B the Brownian bridge defined by $B(t) = W(t) - tW(1)$. For $g \in H = \{f \in L^2(0,1) : \int_0^1 f(t)dt = 0\}$, define

$$W_g : t \to W(t) + \int_0^t g(s)ds \text{ and } h : t \to \frac{\int_0^t g(s)ds}{\phi(\Phi^{-1}(t))}.$$

Girsanov's theorem : W_g is a Brownian motion under the probability measure P_g such that :

$$\frac{dP_g}{dP} = \exp\left(-\int_0^1 g(s)dW(s) - \frac{1}{2}\|g\|^2\right)$$

Define the kernel :

$$\tilde{K}(s,t) = \sum_{j=3}^{+\infty} \frac{1}{j} f_j(s) f_j(t)$$

and for a Brownian motion W form the multiple integral :

$$\mathcal{K}(W) = \int\int \tilde{K}(s,t) dW(s) dW(t).$$

Theorem 5 *With P-probability 1,*

$$\mathcal{K}(W_g) = \int\int \tilde{K}(s,t) dW_g(s) dW_g(t)$$

is equal to

$$\int_0^1 \frac{B_h^2(t) - EB^2(t)}{\phi^2(\Phi^{-1}(t))} dt - \left(\int_0^1 \frac{B_h(t)}{\phi(\Phi^{-1}(t))} dt\right)^2 - \left(\int_0^1 \frac{B_h(t)\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))} dt\right)^2,$$

where as before $h : t \to \frac{\int_0^t g(s)ds}{\phi(\Phi^{-1}(t))}$.

4.Study of the efficiency for Wasserstein's test

Gradient and eigenfunctions

The asymptotic α-level test function for Wasserstein's test is hence :

$$\phi_\alpha(W_g) = 1_{\{\mathcal{K}(W_g) > c_\alpha\}},$$

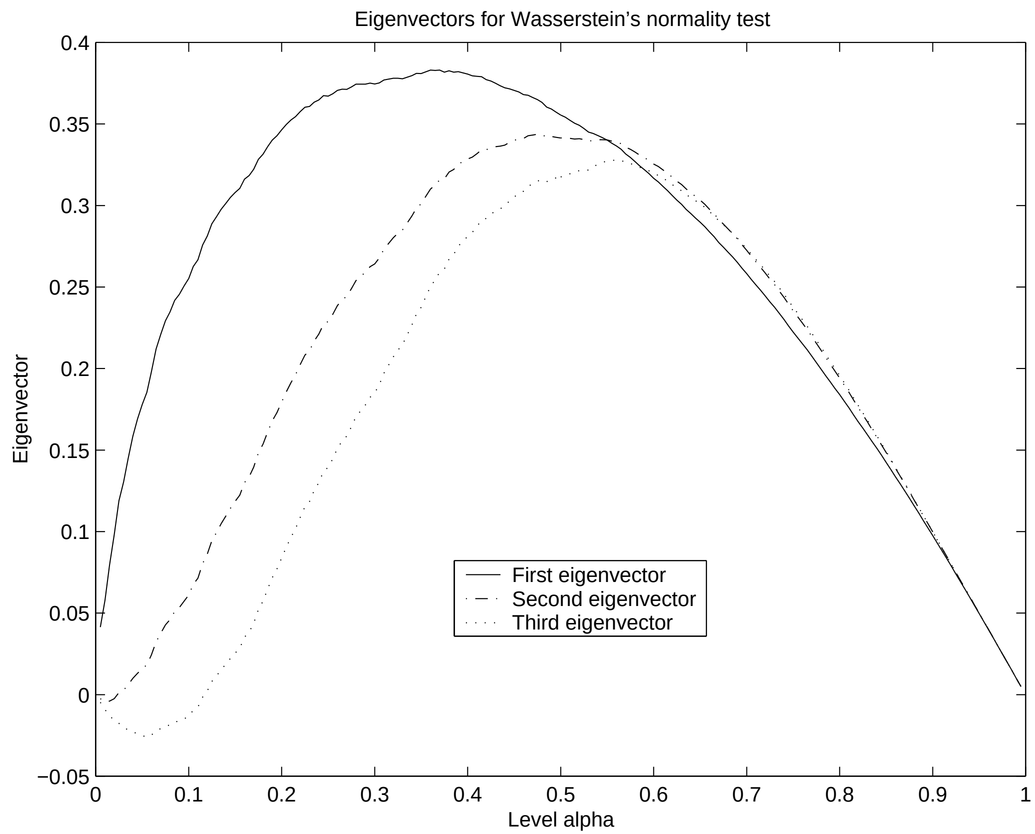
where c_α is chosen to obtain level α. We denote by b_α and a_α the functions that appear in the Taylor expansion (1)

Theorem 6 *For all* $\alpha \in [0,1], g \in H$,
(i) $b_\alpha(g) = 0$,
(ii) $a_\alpha(g) = \langle g, T_\alpha g \rangle$, T_α *has decomposition* $T_\alpha(g) = \sum_{i=1}^{+\infty} \mu_{\alpha,i} \langle f_i, g \rangle$. *If we denote* $Z_i = \sum_{k=3, k \neq i}^{+\infty} \frac{Z_k^2 - 1}{k}$, *the eigenvectors are for* $i \geq 1$ *:*

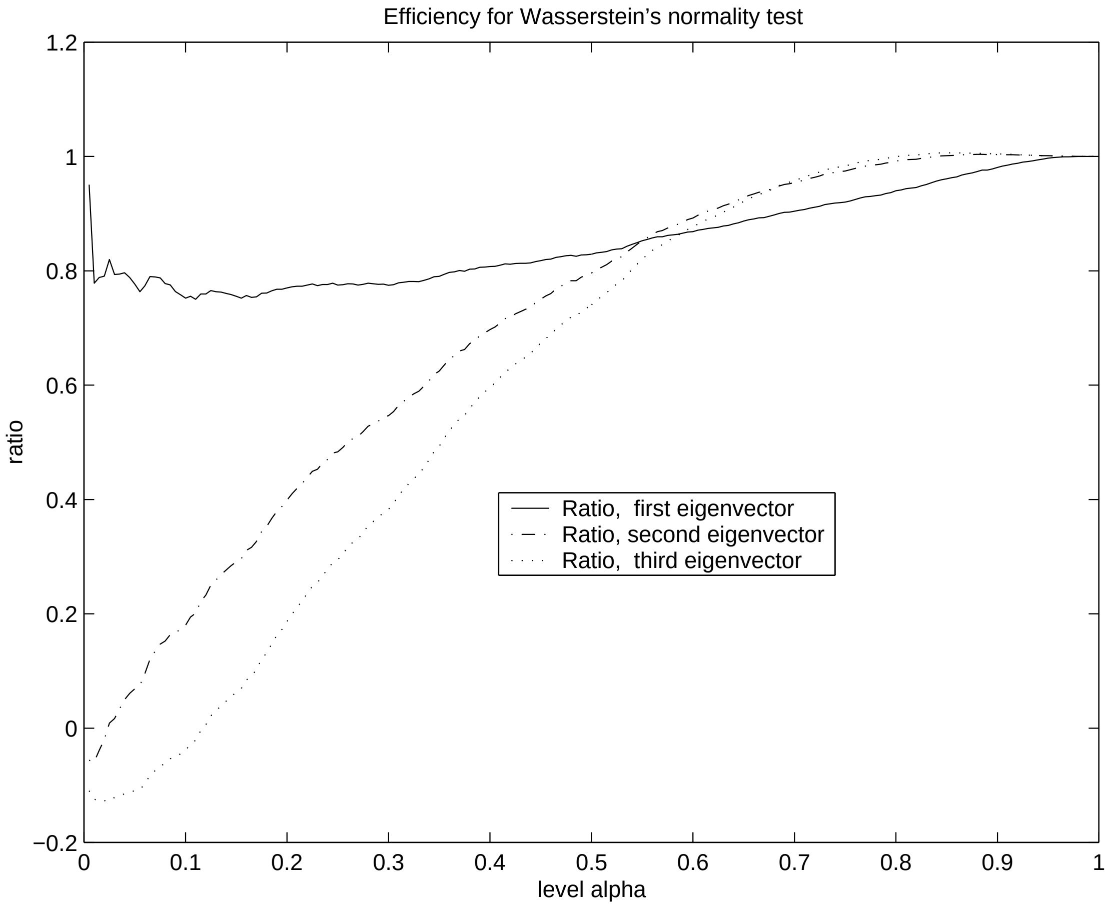
$$\mu_{\alpha,i} = 1 - \alpha - \int_{\mathbb{R}} \int_{\mathbb{R}} y^2 1_{\{\frac{y^2-1}{i} + z \leq c_\alpha\}} \phi(y) dy dP_{Z_i}(z).$$

Numerical approximation for the eigunvectors.

Simulating variables Z_j, we obtain by a Monte-Carlo method an approximation for the eigenvectors.



Approximate asymptotic relative efficiency



References

Del Barrio, E., J.A. Cuesta-Albertos, C. Matrán and J. Rodríguez-Rodríguez (1999). Tests of goodness of fit based on the L_2 Wasserstein distance, *Annals of Statistics*, **27**, 1230-1239.

Janssen, A. (1995). Principal component decomposition of non-parametric tests, *Probability Theory Related Fields*, **101(2)**, 193-209.

Strasser, H. (1985). Mathematical Theory of Statistics, *De Gruyter*, Berlin.