Wasserstein goodness-of-fit test for mixtures

Eustasio del Barrio, Hélène Boistard
Universidad de Valladolid

Wasserstein normality test [del Barrio et al.,1999, 2000]

Wasserstein distance

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

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

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

Another expression with the quantile function :

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

with F_1^{-1} and F_2^{-1} the quantile functions associated to P_1 and P_2.

Distance to a location-scale family, given by the family of distribution functions (d.f.)

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

where the d.f. F is associated to a $(0,1)$ variable:
for $P \in \mathcal{P}_2(\mathbb{R})$ with d.f. F_0, mean μ_0 and standard deviation σ_0, the L_2-Wasserstein distance of P to $\mathcal{H}_F$ is given by

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

A measure of the distance to the family $\mathcal{H}_F$: the ratio $\frac{\mathcal{W}^2(P,\mathcal{H}_F)}{\sigma_0^2}$.

Test statistics, normality test

Let $X_1, \ldots, X_n$ be an i.i.d. sample with underlying d.f. F_0, empirical d.f. F_n and sample variance S_n^2.

Statistics to test the fit of the sample to the family $\mathcal{H}_F$:

$$\mathcal{R}_n := \frac{\mathcal{W}^2(F_n, \mathcal{H}_F)}{S_n^2} = 1 - \frac{\left(\int_0^1 F_n^{-1}(t) F^{-1}(t) dt\right)^2}{S_n^2}$$

Theorem: asymptotic distribution of $\mathcal{R}_n$. Let $X_1, \ldots, X_n$ be an i.i.d. sequence of r.v. with empirical d.f. F_n, underlying d.f. $F_0 \in \mathcal{H}_F$ and let
$a = \sup\{x : F(x) = 0\}$, $b = \inf\{x : F(x) = 1\}$. We assume that F is twice differentiable on (a, b) with $F'(x) = f(x) > 0$ on (a, b), and that for some $\gamma > 0$

$$\sup_{0<t<1} t(1-t)|f'(F^{-1}(t))|/f^2(F^{-1}(t)) \leq \gamma. \ (*)$$

Moreover, we suppose that

$$\int_0^1 \frac{t(1-t)}{(f(F^{-1}(t)))^2} dt < \infty$$

and that the following conditions on the behaviour of the extremes hold:

$$n \int_0^{\frac{1}{n}} (F_n^{-1}(t) - F_0^{-1}(t))^2 dt \to^P 0 \text{ and } n \int_{\frac{n-1}{n}}^1 (F_n^{-1}(t) - F_0^{-1}(t))^2 dt \to^P 0.$$

Then $\mathcal{R}_n$ converges in distribution to

$$\int_0^1 \left(\frac{B(t)}{f(F^{-1}(t))} \right)^2 dt - \left(\int_0^1 \frac{B(t)}{f(F^{-1}(t))} dt \right)^2 - \left(\int_0^1 \frac{B(t) F^{-1}(t)}{f(F^{-1}(t))} dt \right)^2.$$

Theorem for the normal case: let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of normal r.v. (Φ, ϕ d.f. and density function). Then

$$n(\mathcal{R}_n - a_n) \to^d \int_0^1 \frac{B^2(t) - EB^2(t)}{(\phi(\Phi^{-1}(t)))^2} dt - \left(\int_0^1 \frac{B(t)}{\phi(\Phi^{-1}(t))} dt\right)^2 - \left(\int_0^1 \frac{B(t)\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))} dt\right)^2$$

with $a_n = \frac{1}{n} \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{t(1-t)}{(\phi(\Phi^{-1}(t)))^2} dt$, and B a Brownian bridge.

This holds because the following properties are satisfied by the normal law:

$$\int_0^1 \int_0^1 \left(\frac{s \wedge t - st}{\phi(\Phi^{-1}(s))\phi(\Phi^{-1}(t))}\right)^2 dsdt < +\infty \text{ and}$$

$$\lim_{n \to +\infty} \frac{1}{\sqrt{n}} \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{\sqrt{t(1-t)}}{(\phi(\Phi^{-1}(t)))^2} dt = 0.$$

A principal component decomposition of the process $\frac{B(t)}{\phi(\phi^{-1}(t))}$ with covariance function

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

leads to:

<u>Theorem</u>: let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of normal r.v.. Then

$$n(\mathcal{R}_n - a_n) \to^d -\frac{3}{2} + \sum_{j=3}^{+\infty} \frac{Z_j^2 - 1}{j},$$

where $\{Z_j\}_{j=3}^{+\infty}$ is a sequence of independent N(0,1) normal r.v. and $a_n = \frac{1}{n}\int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{t(1-t)}{(\phi(\Phi^{-1}(t)))^2}dt$. With this expression can be calculated the limit distribution of $n(\mathcal{R}_n - a_n)$ via its characteristic function. [de Wet and Venter, 1972].

Remark: The asymptotic distribution of $\mathcal{R}_n$ is proved for $X_1, \ldots, X_n \sim N(0,1)$, owing to the following decomposition:

$$\mathcal{R}_n = \frac{1}{S_n^2} \left(\|F_n^{-1} - \Phi^{-1}\|_2^2 - < F_n^{-1} - \Phi^{-1}, 1 >^2 - < F_n^{-1} - \Phi^{-1}, \Phi^{-1} >^2 \right),$$

where $< . \, , . >$ denotes the inner product in $L_2([0,1])$.

Let us denote by $H_j(x)$ the j-th polynomial of Hermite (of degree j).
$H_0(x) = 1$, $H_1(x) = x$. These polynomials form a basis of $L_2(\mathbb{R})$ for the inner product $< f, g >= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) g(x) e^{-\frac{x^2}{2}} dx$.

The functions $H_j \circ \Phi^{-1}$ are then a basis of $L_2([0,1])$, which yields:

$$R_n = \frac{1}{S_n^2} \sum_{j=2}^{+\infty} < F_n^{-1} - \Phi^{-1}, H_j \circ \Phi^{-1} >^2 .$$

Power study under contiguous alternatives

Examples of contiguous mixtures

Let $X_1, \ldots, X_n$ be an i.i.d sample with underlying d.f.

$$\Phi_n(x) = \gamma N(\mu, \sigma^2) + (1-\gamma) N(\mu + \sigma a, \sigma^2),$$

in two situations:

- $\gamma = 1 - \frac{\delta}{\sqrt{n}}, a \in \mathbb{R}, \delta \in \mathbb{R}_+$
- $\gamma \in [0,1], a = \frac{\alpha}{\sqrt{n}}, \alpha \in \mathbb{R}$.

The purpose: the asymptotic distribution of $\mathcal{R}_n$ under these alternatives.

Remark: the contiguity hypothesis here is stated in terms of Wasserstein's distance and not of Hellinger's distance like in the classical contiguity requirement on the probability density function ϕ_n:

$$\int \left(\sqrt{n}(\phi_n^{1/2} - \phi^{1/2}) - \frac{1}{2} g \phi^{1/2} \right)^2 \to 0.$$

General hypothesis for a theoretical study of the power

Hypothesis (H): $X_1, \ldots, X_n$ an i.i.d. sample with underlying d.f. Φ_n such that there exists a d.f. $\tilde{\Phi}_n$, and $\mu \in \mathbb{R}, \sigma > 0$, with $\Phi_n(x) = \tilde{\Phi}_n(\frac{x-\mu}{\sigma}) \forall x \in \mathbb{R}$ and $\tilde{\Phi}_n$ satisfying $(\tilde{H})$: there exists a function h in $L_2([0,1])$ such that

$$\int_0^1 \left(\sqrt{n}(\tilde{\Phi}_n^{-1}(t) - F^{-1}(t)) - h(t)\right)^2 dt \to 0, n \to \infty$$

and $h_n(t) := \sqrt{n}(\tilde{\Phi}_n^{-1}(t) - F^{-1}(t))$ is dominated by a function in $L_4([0,1])$. We also suppose that if $Y_i = \frac{X_i - \mu}{\sigma}$, the sample variance associated to $\{Y_i\}$ tends to 1 in probability.

Theorem: suppose we are testing fit to the location-scale family $\mathcal{H}_F$, with F satisfaying the condition:

$$\int_0^1 \frac{t(1-t)}{(f(F^{-1}(t)))^2} dt < \infty$$

and the set of conditions (*). Let $X_1, \ldots, X_n$ be an i.i.d. sample satisfaying hypothesis (H) for a continuous function h on $[0,1]$, then $n\mathcal{R}_n$ converges in distribution to

$$\int_0^1 \left(h(t) + \frac{B(t)}{f(F^{-1}(t))}\right)^2 dt - \left(\int_0^1 (h(t) + \frac{B(t)}{f(F^{-1}(t))}) dt\right)^2 - \left(\int_0^1 (h(t) + \frac{B(t)}{f(F^{-1}(t))}) F^{-1}(t) dt\right)^2$$

Theorem in the normal case with a stronger hypothesis (H): we impose

$$\log\log n \int_0^1 \left(\sqrt{n}(\tilde{\Phi}_n^{-1}(t) - \Phi^{-1}(t)) - h(t)\right)^2 dt \to 0, n \to \infty.$$

Under the hypothesis (H), there exist a probability space, r.v. V_i with uniform law on $[0,1]$ and empirical d.f. G_n and a sequence of Brownian bridges B_n on this space, such that the Wasserstein statistics $\mathcal{R}_n = \frac{\mathcal{W}^2(F_n, \mathcal{H}_\Phi)}{S_n^2}$ satisfies:

$$\begin{aligned} n(\mathcal{R}_n & - & a_n) = \int_0^1 h^2(t)dt + \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n^2(t) - EB_n^2(t)}{(\phi(\Phi^{-1}(t)))^2} dt + 2\int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n(t) h(G_n^{-1}(t))}{\phi(\Phi^{-1}(t))} dt \\ & - & \left(\int_0^1 h(t)dt + \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n(t)}{\phi(\Phi^{-1}(t))} dt\right)^2 - \left(\int_0^1 h(t)\Phi^{-1}(t)dt + \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n(t)}{\phi(\Phi^{-1}(t))} \Phi^{-1}(t)dt\right)^2 \\ & + & o_P(1). \end{aligned}$$

Supplementary hypothesis:

if there exists $\lambda > 0$ such that $n^\lambda \int_{\frac{1}{n}}^{\frac{n-1}{n}} \left(h(G_n^{-1}(t)) - h(t)\right)^2 dt = O_P(1)$, then

$$\int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n(t)h(G_n^{-1}(t))}{\phi(\Phi^{-1}(t))}dt = \int_{\frac{1}{n}}^{\frac{n-1}{n}} \frac{B_n(t)h(t)}{\phi(\Phi^{-1}(t))}dt + o_P(1),$$

and the following decomposition is true:

Theorem: $n(\mathcal{R}_n - a_n) \to^d \sum_{j=3}^{+\infty} \left(\tilde{Z}_j^2 - E\tilde{Z}_j^2\right) - E\tilde{Z}_1^2 - E\tilde{Z}_2^2 + \int_0^1 h^2(t)dt$, where $\tilde{Z}_j$ is an sequence of normal r.v. with mean $E\tilde{Z}_j = \int_0^1 h(t)H_j(\Phi^{-1}(t))dt$ and covariance $E\tilde{Z}_i\tilde{Z}_j = \frac{1}{j}\delta_{ij}$, for $i, j \geq 1$.

The example of mixtures

For the first example of mixture: $\Phi_n(x) = (1 - \frac{\delta}{\sqrt{n}})\Phi(x) + \frac{\delta}{\sqrt{n}}\Phi(x - a)$, and

$$h(t) = \delta \frac{t - \Phi(\Phi^{-1}(t) - a)}{\phi(\Phi^{-1}(t))}.$$

For the second example: $\Phi_n(x) = \gamma\Phi(x) + (1-\gamma)\Phi(x - \frac{\alpha}{\sqrt{n}})$ and

$$h(t) = \alpha(1 - \gamma),$$

which permits to write the expansion of the asymptotic distribution of $\mathcal{R}_n$.

Power study with a weight

The next step of the study will be finding a specifically designed weight for contiguous mixtures. A weight w will be a positive measurable function on $(0,1)$. The $L_2(w)$-Wasserstein distance is

$$\mathcal{W}_w(F,G) = \left(\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 w(t)dt\right)^{1/2}$$

and the test statistics $\mathcal{R}_n$ is replaced by

$$\mathcal{R}_n^w = \frac{\mathcal{W}_w^2(F_n, \mathcal{H})}{\sigma_w^2(F_n)}$$

where $\mu_w(F) = \int_0^1 F^{-1}(t)w(t)dt$ and $\sigma_w^2(F) = \int_0^1 (F^{-1}(t))^2 w(t)dt - (\mu_w(F))^2$.

See [del Barrio, Giné and Utzet, 2003, preprint] for the asymptotic null distribution for some location-scale families.

We hope it can be proved under good conditions that

$$n\|F_n^{-1} - F^{-1}\|_{2,w}^2 \to \|\frac{B}{f \circ F^{-1}} + h\|_{2,w}^2$$

and thus, look for a weight w which maximizes the norm $\|h\|_{2,w}^2$.