A Projection-based Central Symmetry Test

Fraiman Ricardo - Moreno Leonardo - Vallejo Sebastian
rfraiman@cmat.edu.uy - mrleo@iesta.edu.uy - sebastian.vallejo@gmail.com

UNIVERSIDAD DE LA REPUBLICA URUGUAY

IESTA

Abstract

We introduce a non parametric central symmetry test for multivariate data based on random projections supported by the results proposed in [5] and applying an univariate Kolmogorov-Smirnov type test developed by [1]. We show that the test is distribution free (under the null hypothesis) and universally consistent. It is possible a similar development for Cramer Von Mises or Range type symmetry tests. We compare in a simulation study the efficiency of the proposed test with respect to other multivariate symmetry tests. The test is valid for high dimensional data, including Hilbert spaces.

Basic Concepts

Central Symmetry A random vector $\boldsymbol{X}$ has a distribution centrally symmetric about the origin in $\mathbb{R}^d$ iif $\boldsymbol{X} \stackrel{d}{=} -\boldsymbol{X}$.

Orthogonal Projection. If π_h denotes the orthogonal projection of $\mathbb{R}^d$ on the subspace spanned by the vector $\boldsymbol{h}$ (with norm one), and B is a Borel set in this subspace, then induced measure on the subspace is,

$P_{\langle h \rangle}(B) = P\left[\pi_h^{-1}(B)\right]$

We denote $\langle X, \boldsymbol{h} \rangle = X^{\boldsymbol{h}}$

A simple characterization of symmetry

Cramér Wold Theorem(1936)
$\mathscr{E}(P,Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$, then

$$\mathscr{E}(P,Q) = \mathbb{R}^n \Rightarrow P = Q \qquad (1)$$

Corollary.
$X \in \mathbb{R}^d$ centrally symmetric iif

$$\langle \boldsymbol{X}, \boldsymbol{h} \rangle \stackrel{d}{=} -\langle \boldsymbol{X}, \boldsymbol{h} \rangle, \qquad (2)$$

for any vector $\boldsymbol{h} \in \mathbb{R}^d$ with norm 1.

Cuesta-Albertos, Fraiman, Ransford (2007)

Theorem. Given two Borel measures P and Q on $\mathbb{R}^d$ where $d \geq 2$. If it holds:

- P is determined by its moments,
- $\mathscr{E}(P,Q)$ is not contained in any projective hypersurface in $\mathbb{R}^d$.

Then $P = Q$
Corollary. Given P and Q Borel measures on $\mathbb{R}^d$, $(d \geq 2)$. If it holds:

- P is determined by its moments.
- $\mathscr{E}(P,Q)$ has positive H-measure on $\mathbb{R}^d$, H an absolutely continuous measure about the Lebesgue measure.

Then $P = Q$
Remark: The authors also extended this result on Hilbert spaces. See [5]

Application to symmetry

Theorem. Let $X \in \mathbb{R}^d$ having a distribution determined by its moments. If the set of directions $\boldsymbol{h}$ where $X^{\boldsymbol{h}}$ is symmetric in $\mathbb{R}$ has positive H-measure then, X is symmetric.

The Test

Let $\{X_1, X_2, \ldots, X_n\}$ a set of random vectors i.i.d, which distribution is determined by its moments. We want to make a central symmetry test in $\mathbb{R}^d$, i.e.

$$H_0) \boldsymbol{X} \stackrel{d}{=} -\boldsymbol{X} \qquad H_1) \boldsymbol{X} \stackrel{d}{\neq} -\boldsymbol{X}. \qquad (3)$$

Methodology

If we denote by $F^{\boldsymbol{h}}$ the cumulative distribution of $X_1^{\boldsymbol{h}}$, the test in $\mathbb{R}$ is given by,

$$H_0) F^{\boldsymbol{h}}(x) + F^{\boldsymbol{h}}(-x) - 1 = 0 \quad \forall x$$

$$H_1) |F^{\boldsymbol{h}}(x) + F^{\boldsymbol{h}}(-x) - 1| > 0 \quad \text{for some} \quad x \in \mathbb{R}$$

If $F_n^{\boldsymbol{h}}$ stands for the empirical distribution of the projected data, the proposed statistic is,

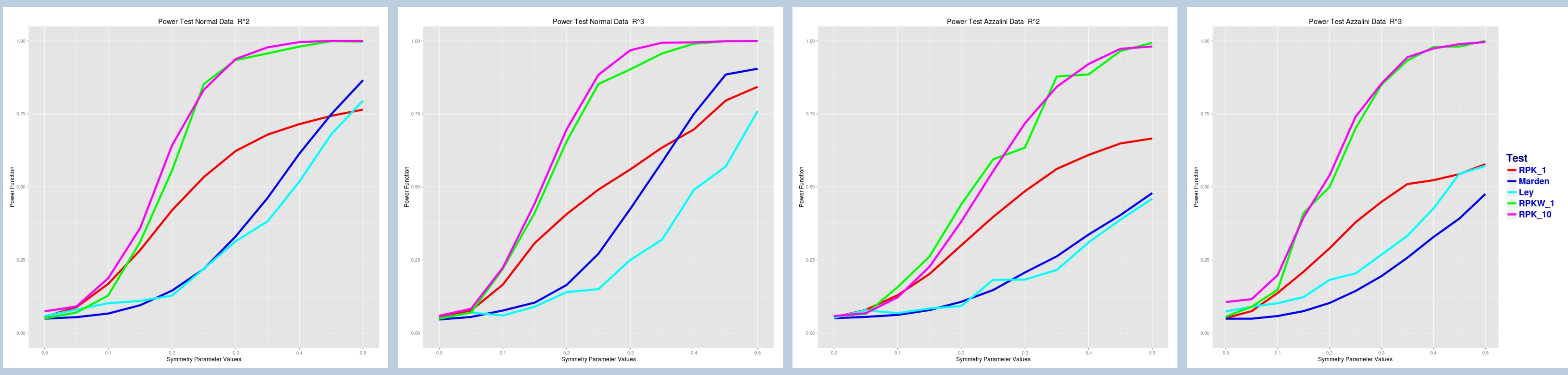
$$D^{\boldsymbol{h}}(n) = \sup_{x \geq 0} |F_n^{\boldsymbol{h}}(x) + F_n^{\boldsymbol{h}}(-x^-) - 1|$$

H_0 is rejected for large values of the statistic. It will be denoted RPK_1 test.

Main properties of the test

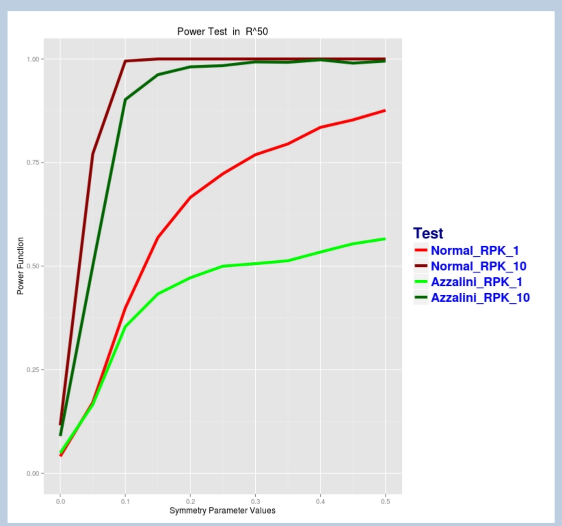
- An explicit expression for the distribution of the statistic under H_0 and its asymptotic distribution is obtained.
- It is, under H_0, **distribution-free**, i.e. it doesn't depend on the distribution H, neither the data distribution.
- The proposed test is **universally consistent.** (The power converges to one under any non symmetrical alternative.)
- To improve the power of the test we compute a finite number of projections (RPK_j) or weighted projections ($RPKW$)

Simulations in $\mathbb{R}^d$, $d = 2, 3$, n=100

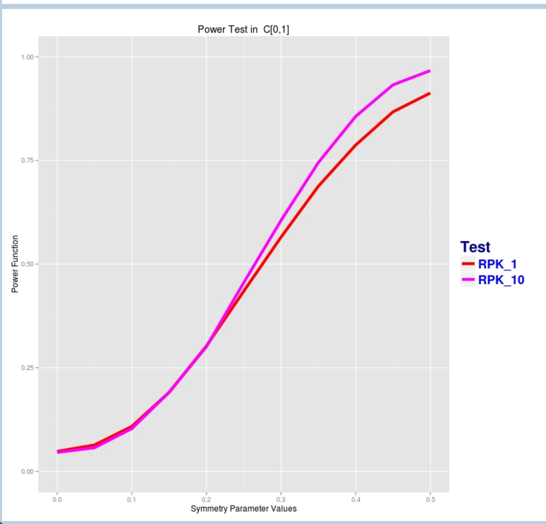


Power functions for the different tests for a Normal distribution and for a Skew-Normal for different values of the asymmetry parameter. If we compute ten random projections we take $\alpha = 0{,}01$ for each test to control the probability of Type-I error. To compute the weighted projections test ($RPKW$) we used part of the sample to weight the directions depending on the value of the median applying cubic splines.

Simulations in $\mathbb{R}^{50}$ and $\mathbb{R}^{\infty}$



- In $\mathbb{R}^{50}$, $n = 1000$ for $N(\mu\boldsymbol{1}, \boldsymbol{I}_{3x3})$ and Azzalini's Skew-Normal with $\mu \in [0, 1/2]$.



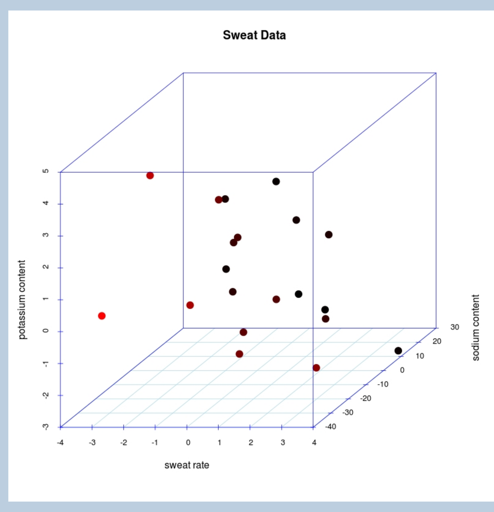
- In $\mathbb{R}^{\infty}$, $n = 100$ and

  $X(t) = W(t) + mt,$

  W(t) is a standard Brownian on $[0, 1]$ and $m \in [0, 1/2]$.

Real Data

The data: Perspiration from twenty healthy females was analyzed with three components, X_1 = sweat rate, X_2 = sodium content, and X_3 = potassium content. The data are centered to compute the central symmetry test.



- **Results:** We get a p-value= 0.12. So we don't reject H_0. This is consistent with the conclusions in [9].

Source: Johnson A.R, Wichern D.W (Fifth Edition), page: 214

References

[1] Chatterjee S.K, Sen P.K., *"On Kolmogorov-Smirnov type test for symmetry"* Annals of the Institute of Statistical Mathematics 25, 288-300, (1973).
[2] H. Cramér, H. Wold , *"Some theorems on distribution functions"* J. London Math. Soc. 11 290-295 (1936)
[3] R. Serfling *"Multivariate Symmetry and Asymmetry"* (2000).
[4] Ley Ch. *"Univariate and Multivariate Symmetry: Statistical Inference and Distributional Aspects"* (2011).
[5] Cuesta-Albertos, J.A., Fraiman, R. and Ransford, T. *"Random projections and goodness of fit tests in infinite-dimensional spaces"* Bull. Braz. Math. Soc., New Series 37, 4, 477-501, (2006).
[6] Cuesta-Albertos, J.A., Fraiman, R. and Ransford, T. *"A Sharp Form of the Cramér-Wold Theorem"* Journal of Theoretical Probability, Volume: 20, Issue: 2, Pages: 201-209, (2007)
[7] Marden J. *"Multivariate rank tests"* (1999).
[8] Azzalini, A. *Statistical applications of the multivariate skew-normal distribution,* (2002)
[9] Johnson A.R, Wichern D.W *"Applied Multivariate Statistical analysis"* (1988).