

Un Test de Simetría Central Mediante Proyecciones al Azar



Ricardo Fraiman
Leonardo Moreno
Sebastián Vallejo

I Jornadas de Estadística Aplicada LPE
Ariel Roche in memoriam

La Paloma, Octubre 2013

Introducción

La idea de "simetría" ha servido desde la antigüedad como una punto de referencia conceptual en el arte, las matemáticas y sus aplicaciones. En estética, es un principio de orden, en matemáticas un artefacto de estructura geométrica, en filosofía una abstracción de balance, armonía y perfección, en poesía una esencia intuitiva de naturaleza y divinidad.

Robert J. Serfling,
Multivariate Symmetry and Asymmetry
Encyclopedia of Statistical Sciences
2004

Introducción

En este trabajo nos enfocamos en el concepto de simetría en el contexto de las distribuciones multivariadas de probabilidad.

Distintos tipos de Simetría

Esférica

Diremos que un vector aleatorio $\mathbf{X} \in \mathbb{R}^d$ tiene una distribución esférica simétrica respecto a un centro $\mathbf{a}$ sí y sólo sí

$$\mathbf{X} - \mathbf{a} \stackrel{d}{=} A(\mathbf{X} - \mathbf{a}), \quad (1)$$

para cualquier matriz $A_{d \times d}$ ortogonal.

Elíptica

Diremos que un vector aleatorio $\mathbf{X} \in \mathbb{R}^d$ tiene una distribución elíptica simétrica respecto a un centro $\mathbf{a}$ sí y sólo sí

$$\mathbf{X} - \mathbf{a} \stackrel{d}{=} A'(\mathbf{Y} - \mathbf{a}), \quad (2)$$

para cualquier matriz $A_{k \times d}$ tal que $A'A = \Sigma$ con $\text{rank}(\Sigma) = k \leq d$ y $\mathbf{Y} \in \mathbb{R}^d$ tiene distribución simétrica esférica respecto de $\mathbf{a}$.

Distintos tipos de Simetría

Central

Diremos que un vector aleatorio $\mathbf{X} \in \mathbb{R}^d$ tiene una distribución central simétrica respecto al centro $\mathbf{a}$ sí y sólo sí

$$\mathbf{X} - \mathbf{a} \stackrel{d}{=} -(\mathbf{X} - \mathbf{a}), \quad (3)$$

Obsevación

Simetría central es la más débil de las tres definidas.

Simetría bajo proyecciones

Distribución de la Proyección Ortogonal

Si anotamos $\pi_{\mathbf{h}}$ a la proyección ortogonal de $\mathbb{R}^d$ en el subespacio generado por el vector $\mathbf{h}$ (que supondremos de norma 1), y B un boreliano de este subespacio, entonces la medida inducida en el subespacio es,

$$P_{\langle \mathbf{h} \rangle}(B) = P[\pi_{\mathbf{h}}^{-1}(B)] \quad (4)$$

Anotaremos $\langle X, \mathbf{h} \rangle = X^{\mathbf{h}}$

Observación

X es simétrica entonces $X^{\mathbf{h}}$ es simétrica

Algunas preguntas

- ¿Quién es el centro de simetría?

Algunas preguntas

- Quién es el centro de simetría?

En este trabajo evitamos este problema y lo consideramos dado.
Sin perder generalidad $\mathbf{a} = \mathbf{0}$

Algunas preguntas

- Quién es el centro de simetría?

En este trabajo evitamos este problema y lo consideramos dado.
Sin perder generalidad $\mathbf{a} = \mathbf{0}$

- Si X_1 y X_2 son v.a simétricas en $\mathbb{R}$, entonces $X = (X_1, X_2)$ es siempre una v.a simétrica en $\mathbb{R}^2$?

Algunas preguntas

- Quién es el centro de simetría?

En este trabajo evitamos este problema y lo consideramos dado.
Sin perder generalidad $\mathbf{a} = \mathbf{0}$

- Si X_1 y X_2 son v.a simétricas en $\mathbb{R}$, entonces $X = (X_1, X_2)$ es siempre una v.a simétrica en $\mathbb{R}^2$?

NO

Algunas preguntas

- Quién es el centro de simetría?

En este trabajo evitamos este problema y lo consideramos dado.
Sin perder generalidad $\mathbf{a} = \mathbf{0}$

- Si X_1 y X_2 son v.a simétricas en $\mathbb{R}$, entonces $X = (X_1, X_2)$ es siempre una v.a simétrica en $\mathbb{R}^2$?

NO

- Sea $X \in \mathbb{R}$, si $X^{\mathbf{h}}$ es simétrica para infinitas direcciones $\mathbf{h}$ entonces X es simétrica?

Algunas preguntas

- Quién es el centro de simetría?

En este trabajo evitamos este problema y lo consideramos dado.
Sin perder generalidad $\mathbf{a} = \mathbf{0}$

- Si X_1 y X_2 son v.a simétricas en $\mathbb{R}$, entonces $X = (X_1, X_2)$ es siempre una v.a simétrica en $\mathbb{R}^2$?

NO

- Sea $X \in \mathbb{R}$, si $X^{\mathbf{h}}$ es simétrica para infinitas direcciones $\mathbf{h}$ entonces X es simétrica?

NO

Caracterización de la simetría por proyecciones

Teorema de Cramér Wold (1936)

Si anotamos $\mathcal{E}(P, Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$,

$$\mathcal{E}(P, Q) = \mathbb{R}^n \Rightarrow P = Q \quad (5)$$

Corolario: Caracterización de la simetría central

$\mathbf{X} \in \mathbb{R}^d$ simétrica centralmente sí y sólo sí

$$\langle \mathbf{X}, \mathbf{h} \rangle \stackrel{d}{=} -\langle \mathbf{X}, \mathbf{h} \rangle, \quad (6)$$

para cualquier vector $\mathbf{h} \in \mathbb{R}^d$ de norma 1.

Por tanto es necesario y suficiente para que haya simetría central en el espacio original que todas las distribuciones de la proyecciones univariadas sean simétricas.

Proyecciones al Azar: Definiciones Previas

Medida determinada por sus momentos

Sea P una medida de Borel en $\mathbb{R}^d$. Se dice que P está determinada por sus momentos si para cada $n \in \mathbb{N}$ se cumple que $\int \|x\|^n P(dx) < \infty$, y si Q es otra medida de Borel en $\mathbb{R}^d$ se cumple que si

$$\int \langle x, y \rangle^n P(dy) = \int \langle x, y \rangle^n Q(dy) \quad \text{para todo } x \in \mathbb{R}^d \text{ y } n \in \mathbb{N}, \quad (7)$$

Entonces $P = Q$

Proposición. (Condición de Carleman)

Sea P una medida de Borel en $\mathbb{R}^d$. Si sus momentos absolutos $m_n = \int \|x\|^n P(dx)$ son finitos y satisfacen la condición $\sum_{n \geq 1} m_n^{-1/n} = \infty$. Entonces P es determinada por sus momentos.

Proyecciones al Azar: Definiciones Previas

Hipersuperficie proyectiva

Diremos que S es una Hipersuperficie proyectiva de $\mathbb{R}^d$ sí y sólo sí existe un polinomio homogéneo $p(x)$ de $\mathbb{R}^d$ tal que

$$S = \{x \in \mathbb{R}^d / p(x) = 0\} \quad (8)$$

Proposición.

Toda hipersuperficie proyectiva tiene medida de Lebesgue 0

Una forma precisa del Teorema de Cramér Wold

Teorema. Cuesta-Albertos, Fraiman, Ransford (2007)

Sean P y Q medidas de Borel en $\mathbb{R}^d$ donde $d \geq 2$. Si se cumple que

- P está determinada por sus momentos,

Entonces $P = Q$

Corolario.

Siendo $\mathcal{E}(P, Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$, P y Q medidas de Borel en $\mathbb{R}^d$, ($d \geq 2$). Si se cumple que

Entonces $P = Q$

Una forma precisa del Teorema de Cramér Wold

Teorema. Cuesta-Albertos, Fraiman, Ransford (2007)

Sean P y Q medidas de Borel en $\mathbb{R}^d$ donde $d \geq 2$. Si se cumple que

- P está determinada por sus momentos,
- $\mathcal{E}(P, Q)$ no está contenido en alguna hipersuperficie proyectiva en $\mathbb{R}^d$.

Entonces $P = Q$

Corolario.

Siendo $\mathcal{E}(P, Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$, P y Q medidas de Borel en $\mathbb{R}^d$, ($d \geq 2$). Si se cumple que

Entonces $P = Q$

Una forma precisa del Teorema de Cramér Wold

Teorema. Cuesta-Albertos, Fraiman, Ransford (2007)

Sean P y Q medidas de Borel en $\mathbb{R}^d$ donde $d \geq 2$. Si se cumple que

- P está determinada por sus momentos,
- $\mathcal{E}(P, Q)$ no está contenido en alguna hipersuperficie proyectiva en $\mathbb{R}^d$.

Entonces $P = Q$

Corolario.

Siendo $\mathcal{E}(P, Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$, P y Q medidas de Borel en $\mathbb{R}^d$, ($d \geq 2$). Si se cumple que

- P está determinada por sus momentos.

Entonces $P = Q$

Una forma precisa del Teorema de Cramér Wold

Teorema. Cuesta-Albertos, Fraiman, Ransford (2007)

Sean P y Q medidas de Borel en $\mathbb{R}^d$ donde $d \geq 2$. Si se cumple que

- P está determinada por sus momentos,
- $\mathcal{E}(P, Q)$ no está contenido en alguna hipersuperficie proyectiva en $\mathbb{R}^d$.

Entonces $P = Q$

Corolario.

Siendo $\mathcal{E}(P, Q) = \{h \in \mathbb{R}^d / P_{\langle h \rangle} = Q_{\langle h \rangle}\}$, P y Q medidas de Borel en $\mathbb{R}^d$, ($d \geq 2$). Si se cumple que

- P está determinada por sus momentos.
- El conjunto $\mathcal{E}(P, Q)$ tiene H -medida positiva en $\mathbb{R}^d$, siendo H una medida absolutamente continua respecto de la medida de Lebesgue.

Entonces $P = Q$

Una aplicación a variables aleatorias simétricas

A partir del corolario anterior podemos afirmar que,

Sea X una v.a en $\mathbb{R}^d$ determinada por sus momentos y el conjunto de direcciones $\mathbf{h}$ donde $X^{\mathbf{h}}$ es simétrica en $\mathbb{R}$ tiene H -medida positiva entonces, X es simétrica.

El Test

Las Hipótesis

Sean $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ un conjunto de vectores aleatorios i.i.d, determinados por sus momentos. Se quiere realizar una prueba de simetría central en $\mathbb{R}^d$, o sea

$$H_0) \mathbf{X} \stackrel{d}{=} -\mathbf{X} \quad H_1) \mathbf{X} \stackrel{d}{\neq} -\mathbf{X}. \quad (9)$$

Metodología

- Se sortea al azar con una medida de probabilidad H en $\mathbb{R}^d$ (H absolutamente continua respecto a la medida de Lebesgue) una dirección $\mathbf{h}$

Metodología

- Se sortea al azar con una medida de probabilidad H en $\mathbb{R}^d$ (H absolutamente continua respecto a la medida de Lebesgue) una dirección $\mathbf{h}$
- Fijada esa dirección se proyecta ortogonalmente la muestra i.i.d $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ sobre el espacio unidimensional generado por $\mathbf{h}$ obteniendo una nueva muestra i.i.d en $\mathbb{R}$, $\{X_1^{\mathbf{h}}, X_2^{\mathbf{h}}, \dots, X_n^{\mathbf{h}}\}$

Metodología

- Se realiza sobre estos datos proyectados con cierto nivel de significación un test de simetría en $\mathbb{R}$ del tipo Kolmogorov-Smirnov desarrollado por K. Sen y K.S Chatterje. Si llamamos F^h a la distribución acumulada de X_1^h , la prueba en $\mathbb{R}$ a realizar es,

$$H_0) F^h(x) + F^h(-x) - 1 = 0 \quad \forall x \quad H_1) |F^h(x) + F^h(-x) - 1| > 0 \quad \text{para algún } x \quad (10)$$

Si se denota F_n^h a la distribución empírica de los datos proyectados el estadístico propuesto es

$$D^h(n) = \sup_{x \geq 0} |F_n^h(x) + F_n^h(-x^-) - 1| \quad (11)$$

A valores “grandes” del estadístico se rechaza H_0 de las hipótesis originales expresadas en (9).

Propiedades del test

- Se conoce la distribución exacta del estadístico y también su distribución asintótica.

Propiedades del test

- Se conoce la distribución exacta del estadístico y también su distribución asintótica.
- Tiene **distribución libre**, es decir, no depende de la distribución H y tampoco de la distribución $F \in \mathcal{F}_0$, siendo $\mathcal{F}_0$ el conjunto de todas las distribuciones simétricas en $\mathbb{R}^d$.

Propiedades del test

- Se conoce la distribución exacta del estadístico y también su distribución asintótica.
- Tiene **distribución libre**, es decir, no depende de la distribución H y tampoco de la distribución $F \in \mathcal{F}_0$, siendo $\mathcal{F}_0$ el conjunto de todas las distribuciones simétricas en $\mathbb{R}^d$.
- el test propuesto es **asintóticamente consistente** bajo cualquier alternativa no simétrica. Es decir,

$$H \left\{ h \in \mathbb{R}^d : P \left(\liminf_{n \rightarrow +\infty} D_n^h > 0 \right) = 1 \right\} = 1.$$

Potencia del Test

En general se presentan problemas con la potencia,

A nivel univariado Los test no paramétricos univariados del tipo Kolmogorov-Smirnov si bien son universalmente consistentes, no presentan una alta potencia para cualquier alternativa.(Posibles alternativa: Mason [1983], Cabaña [2002])

Potencia del Test

En general se presentan problemas con la potencia,

A nivel univariado Los test no paramétricos univariados del tipo Kolmogorov-Smirnov si bien son universalmente consistentes, no presentan una alta potencia para cualquier alternativa. (Posibles alternativa: Mason [1983], Cabaña [2002])

A nivel multivariado Si bien bajo H_1 el conjunto de direcciones donde la proyección tiene distribución simétrica tiene H-medida 0, hay un “entorno angular” de estas direcciones con H-medida positiva donde, al ser la muestra finita, no se rechaza H_0 cuando se debería rechazar. (Posibles alternativas: Número finito de proyecciones, proyecciones ponderadas)

Simulaciones

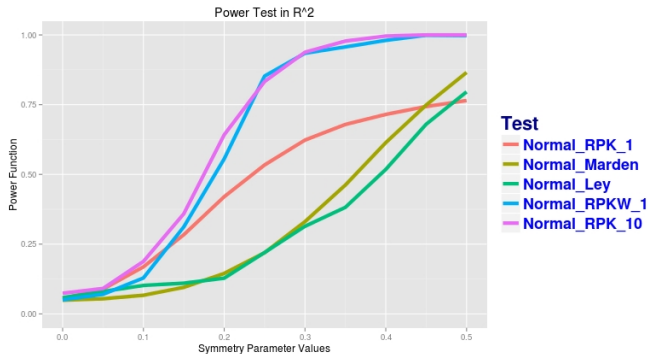
Escenarios

- Muestra de tamaño $n = 100$ con distribución $N(\mu \mathbf{1}_d, \mathbf{I}_d)$ donde μ es número real que va a tomar valores entre el intervalo $[0; 0,5]$, y d es la dimensión del espacio.
- Muestra de tamaño $n = 100$ con distribución “asimétrica-normal multivariada”, Azzalini [2002]. $SN_d(\mathbf{I}_d, \mu)$ donde μ es parámetro de simetría que va a tomar valores entre 0 y 0,5.

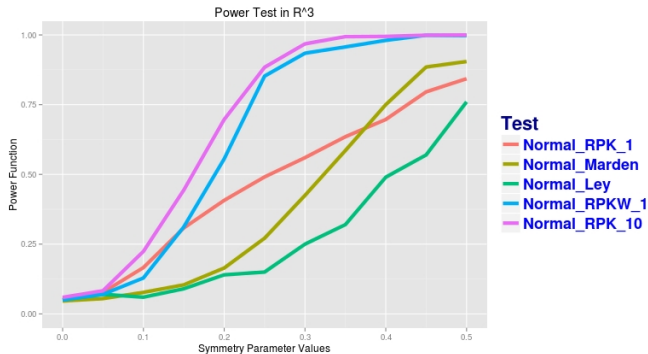
Otros Test simulados

- Test de Marden [1999].
- Test de Ley [2010].

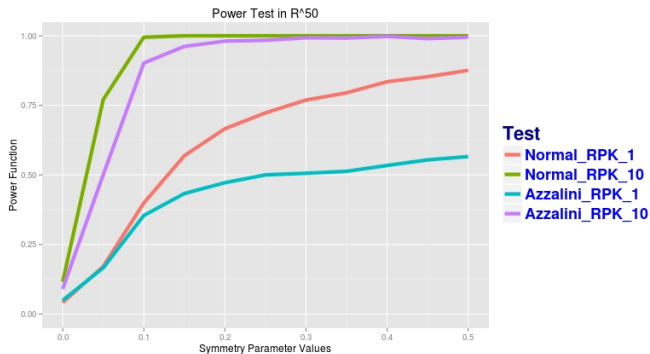
Funciones de Potencia en dimensión 2: alternativa Normal desplazada.



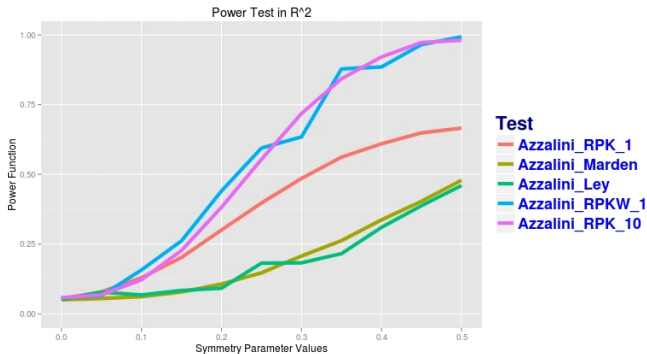
Funciones de Potencia en dimensión 3: alternativa Normal desplazada.



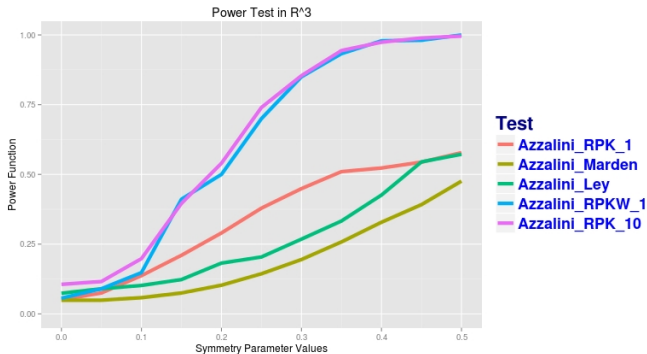
Funciones de Potencia en dimensión 50



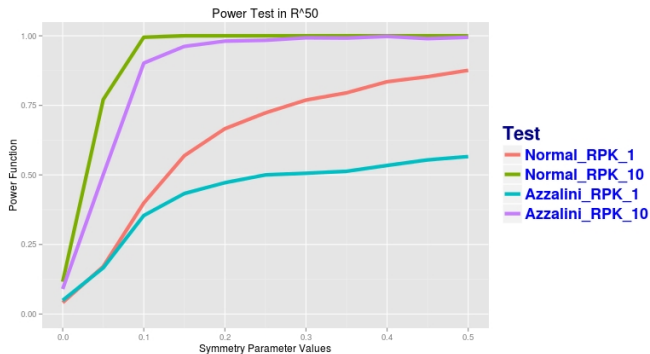
Funciones de Potencia en dimensión 2: alternativa Azzalini.



Funciones de Potencia en dimensión 3: alternativa Azzalini.

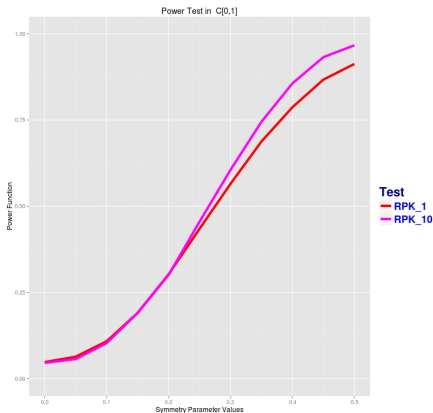


Funciones de Potencia en dimensión 50.



En dimensión infinita.

En $\mathbb{R}^\infty$ tomamos $n = 100$ con $X(t) = W(t) + mt$, siendo $W(t)$ un movimiento Browniano en $[0, 1]$ y $m \in [0, 1/2]$.



Conclusiones

- El test es extensible a dimensión infinita.
- Es posible modificar el estadístico univariado por otro que conserve sus cualidades (Test de simetría del tipo Cramer Von Mises, Test de Rangos).
- Mejora su eficiencia respecto a otros test clásicos a medida que aumenta la dimensión del espacio.
- Para ser reproducible por otro investigador es necesario fijar la semilla. El test es de resultado aleatorio, como todo test.

GRACIAS !!