Taller de Econofísica

Andrés García Medina

Investigador Cátedras CONACyT, CIMAT-Monterrey
andres.garcia@cimat.mx

8 y 9 de Noviembre

Introducción

Teoría de Matrices aleatorias (RMT)

- RMT desde la Física
- RMT desde la estadística
- RMT desde la econometría

Transferencia de Entropía (TE)

- TE desde la Física
- TE desde la estadística
- TE desde la econometría

Correlaciones y Causalidad en el contexto de la Economía Conductual.

Motivación

- ¿Cuáles son las leyes estadísticas que obedecen los cambios en los precios? ¿Que tan suave son estos cambios? ¿Que tan frecuente son los saltos?
- ¿Pueden ser predichos estos cambios? ¿Pueden ser predichas las caídas de la bolsa?
- Estas son preguntas tratadas por matemáticos, econometristas, y cada vez más por físicos.

¿Por qué los físicos? ¿Por qué modelos de la física?

- Los procesos estocásticos son bien conocidos en física.
- La física estadística describe el comportamiento complejo observado en muchos sistemas físicos en términos de las interacciones de sus constituyentes.
- Las decisiones de inversión influyen en el precio, y a su vez estos cambios influyen a la hora de tomar decisiones.

¿Es posible mejorar nuestra comprensión económica y financiera mediante la física?

¿Es posible mejorar nuestra comprensión económica y financiera mediante la física?

- Mediante el entendimiento de fenómenos paralelos en la naturaleza:
  - Difusión
  - terremotos
  - fenómenos de transición
  - fluidos turbulentos
  - gases de electrones
  - etc.

¿Es posible mejorar nuestra comprensión económica y financiera mediante la física?

- Mediante el entendimiento de fenómenos paralelos en la naturaleza:
  - Difusión
  - terremotos
  - fenómenos de transición
  - fluidos turbulentos
  - gases de electrones
  - etc.
- Mediante el método matemático asociado

¿Es posible mejorar nuestra comprensión económica y financiera mediante la física?

- Mediante el entendimiento de fenómenos paralelos en la naturaleza:
  - Difusión
  - terremotos
  - fenómenos de transición
  - fluidos turbulentos
  - gases de electrones
  - etc.
- Mediante el método matemático asociado
- La experiencia en modelación de fenómenos en física
  - Identificación de parámetros, factores de causalidad, y estimación de ordenes de magnitud.
  - Simplicidad en modelos cualitativos
  - Corroboración empírica utilizando datos reales.
  - Aproximación progresiva a la realidad incorporando nuevos elementos.

Nota: Existen excelentes especialistas.

Antecedentes Históricos

- *Isaac Newton (1642–1726):* Perdió gran parte de su fortuna en la *South Sea Bubble*
- *Carl Friedrich Gauss (1777–1855):* Dejo una fortuna de 170,000 táleros, mientras que su salario era de 1000 taleros (rumores de haber derivado la distribución normal para calcular el riesgo predeterminado al dar prestamos a sus vecinos.)
- *Louis Bachelier (1870-1946):* Tesis doctoral *Theorie de la Speculation* (1900)

Surgimiento de la caminata aleatoria

Surgimiento de la caminata aleatoria

- Consideremos dos eventos mutuamente excluyentes A y B, con probabilidad p, $q = 1 - p$; respectivamente.

Surgimiento de la caminata aleatoria

- Consideremos dos eventos mutuamente excluyentes A y B, con probabilidad p, $q = 1 - p$; respectivamente.
- La probabilidad de observar en m eventos, a realizaciones de A y $b = m - a$ realizaciones de B está dado por la distribución binomial

$$P_{A,B}(a, m-a) = \binom{m}{a} p^a (1-p)^{m-a} = \frac{m!}{a!(m-a)!} p^a (1-p)^{m-a} \quad (1)$$

Surgimiento de la caminata aleatoria

- Consideremos dos eventos mutuamente excluyentes A y B, con probabilidad p, $q = 1 - p$; respectivamente.
- La probabilidad de observar en m eventos, a realizaciones de A y $b = m - a$ realizaciones de B está dado por la distribución binomial

$$P_{A,B}(a, m-a) = \binom{m}{a} p^a (1-p)^{m-a} = \frac{m!}{a!(m-a)!} p^a (1-p)^{m-a} \quad (1)$$

- ¿Cual es el valor de p que maximiza P? *Respuesta*: $p = a/m \Rightarrow a = pm, b = m - a = (1-p)m = qm$

Surgimiento de la caminata aleatoria

- Consideremos dos eventos mutuamente excluyentes A y B, con probabilidad p, $q = 1 - p$; respectivamente.
- La probabilidad de observar en m eventos, a realizaciones de A y $b = m - a$ realizaciones de B está dado por la distribución binomial

$$P_{A,B}(a, m-a) = \binom{m}{a} p^a (1-p)^{m-a} = \frac{m!}{a!(m-a)!} p^a (1-p)^{m-a} \quad (1)$$

- ¿Cual es el valor de p que maximiza P? *Respuesta*: $p = a/m \Rightarrow a = pm, b = m - a = (1-p)m = qm$
- Interpretación financiera: El cambio de precio más probable al paso m esta dado por $x_m = a - b = m(p-q)x_0$

- ¿Cuál es la función de distribución de los cambios en el precio?

[1] L. Bachelier. Théorie de la Spéculation. Ann. Sci. Ecole Norm. Super., Sér. 3, 17, 21 (1900)
https://en.wikipedia.org/wiki/De_Moivre-Laplace_theorem

- ¿Cuál es la función de distribución de los cambios en el precio?
  Respuesta[1] Para $m \to \infty, a \to \infty$, con $h = a - mp$ finita

$$P(h) = \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{h^2}{2mpq}} \tag{2}$$

[1] L. Bachelier. Théorie de la Spéculation. Ann. Sci. Ecole Norm. Super., Sér. 3, 17, 21 (1900)
https://en.wikipedia.org/wiki/De_Moivre-Laplace_theorem

- ¿Cuál es la función de distribución de los cambios en el precio?
  Respuesta[1] Para $m \to \infty, a \to \infty$, con $h = a - mp$ finita

$$P(h) = \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{h^2}{2mpq}} \tag{2}$$

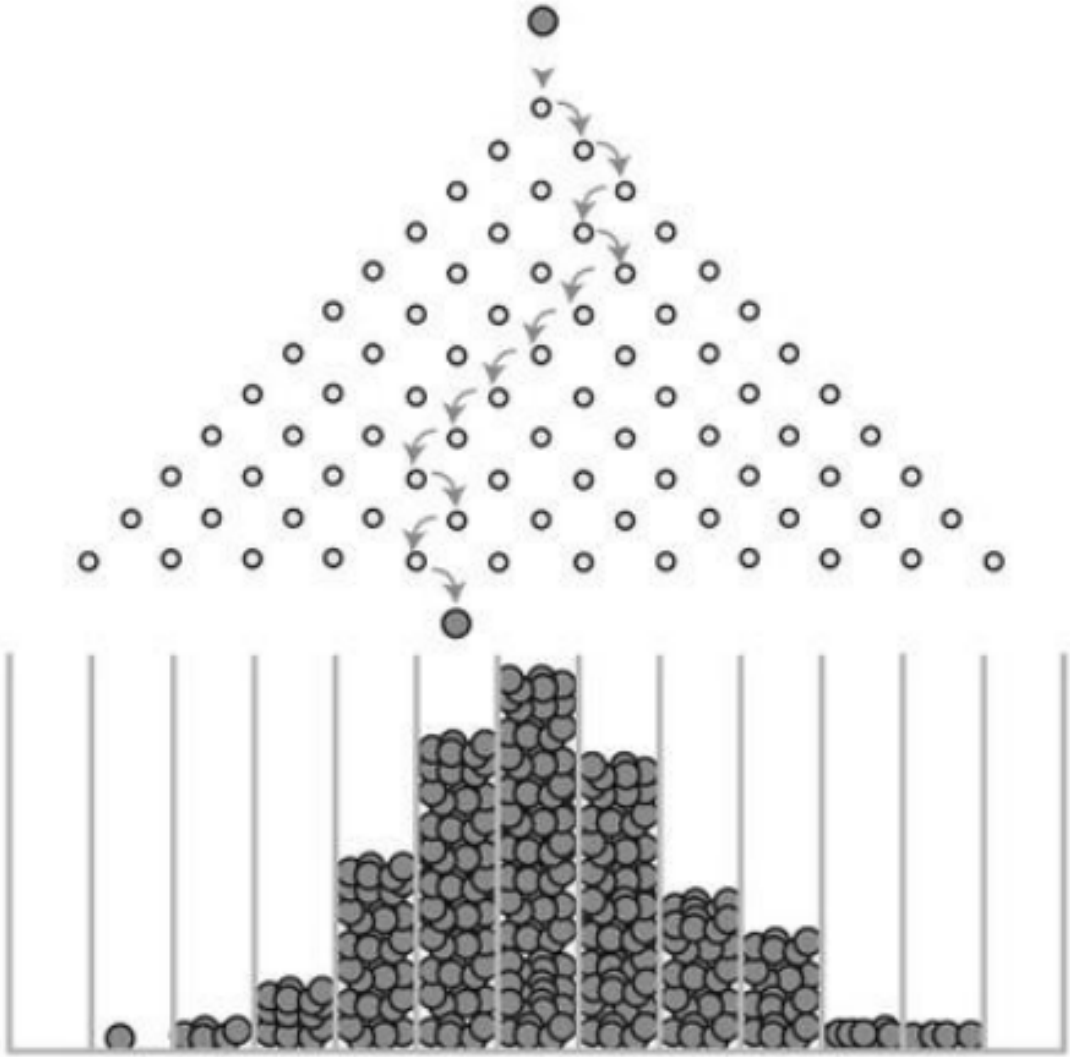
- Asumiendo $p = q = 1/2$, con $h \to x$, y definiendo $m = t/\Delta t$, $H = \sqrt{2\Delta t/\pi}$

$$P(x) = \frac{H}{\sqrt{t}} \exp\left(-\frac{\pi H x^2}{t}\right) \tag{3}$$

  Nota: El tiempo ahora es continuo

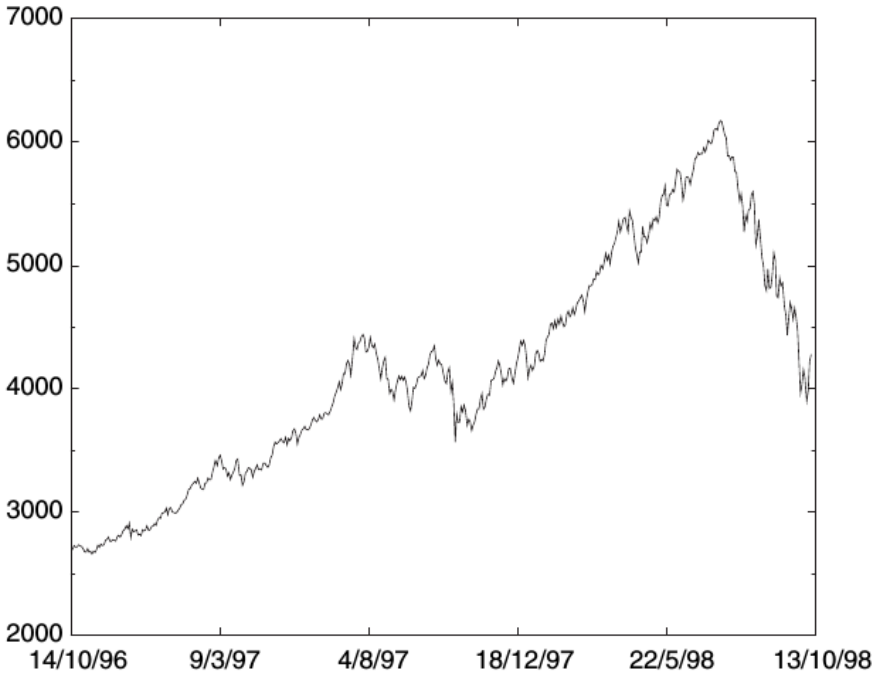
[1] L. Bachelier. Théorie de la Spéculation. Ann. Sci. Ecole Norm. Super., Sér. 3, 17, 21 (1900)
https://en.wikipedia.org/wiki/De_Moivre-Laplace_theorem

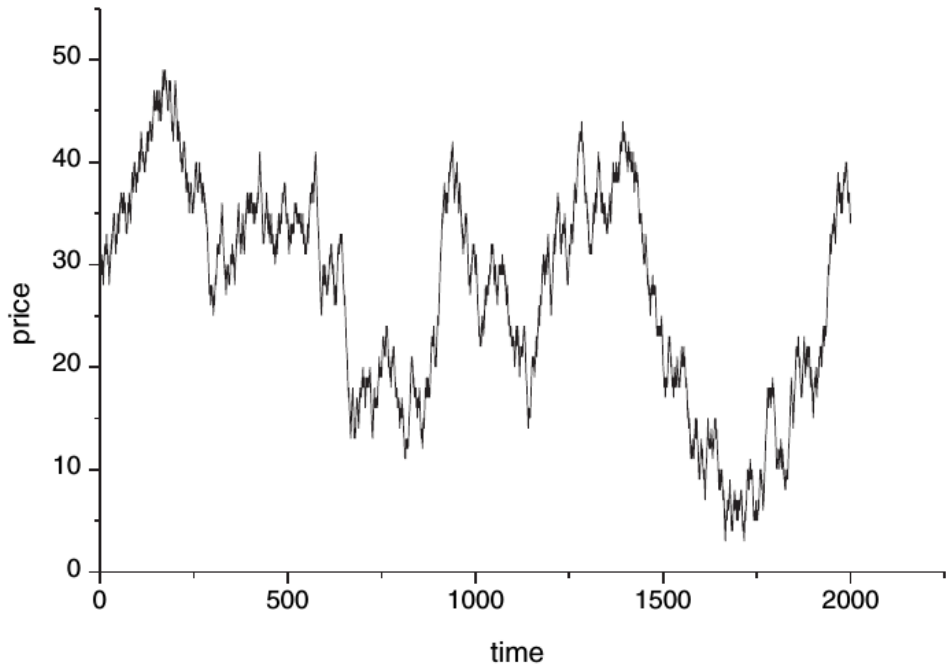
Esta es la primera formulación de la **caminata aleatoria** = **movimiento Browniano** = **proceso estocástico de Einstein-Wiener.**



DAX vs. Lanzar una moneda

"A Random Walk down Wall Street" by B. G. Malkiel, a professor of economics at Princeton





Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos $\rightarrow$ surgimiento de la computadora en 1960.

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos $\rightarrow$ surgimiento de la computadora en 1960.
  - Se integró el marco teórico de la economía Hipótesis, conceptos, métodos, etc.) junto con las herramientas probabilistas en el análisis de los mercados financieros $\rightarrow$ cursos en la universidad, congresos, libros, etc.

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos $\rightarrow$ surgimiento de la computadora en 1960.
  - Se integró el marco teórico de la economía Hipótesis, conceptos, métodos, etc.) junto con las herramientas probabilistas en el análisis de los mercados financieros $\rightarrow$ cursos en la universidad, congresos, libros, etc.

- **Econofísica**: Surgío en la decada de 1990

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos $\rightarrow$ surgimiento de la computadora en 1960.
  - Se integró el marco teórico de la economía Hipótesis, conceptos, métodos, etc.) junto con las herramientas probabilistas en el análisis de los mercados financieros $\rightarrow$ cursos en la universidad, congresos, libros, etc.

- **Econofísica**: Surgío en la decada de 1990
  - Nuevos resultados en la teoría moderna de la probabilidad $\rightarrow$ Aplicaciones de los procesos de Lévy para caracterizar datos financieros con distribuciones no-gaussianas (Mantegna, 1991).

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad $\rightarrow$ axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos $\rightarrow$ surgimiento de la computadora en 1960.
  - Se integró el marco teórico de la economía Hipótesis, conceptos, métodos, etc.) junto con las herramientas probabilistas en el análisis de los mercados financieros $\rightarrow$ cursos en la universidad, congresos, libros, etc.

- **Econofísica**: Surgío en la decada de 1990
  - Nuevos resultados en la teoría moderna de la probabilidad $\rightarrow$ Aplicaciones de los procesos de Lévy para caracterizar datos financieros con distribuciones no-gaussianas (Mantegna, 1991).
  - Nuevos datos empíricos $\rightarrow$ automatización de centros financieros con datos en tiempo real de alta frecuencia (1990)

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

Economia financiera y econofísica[2]

- **Economía financiera**: Se institucionalizó en la década de 1960
  - Se volvieron más accesibles las herramientas de la teoría moderna de la probabilidad → axiomas de Kolmogorov (1930s); Teoría de portafolios de Markowitz (1952); Teoría de mercado eficiente (Fama, 1965)
  - Acceso a datos empíricos → surgimiento de la computadora en 1960.
  - Se integró el marco teórico de la economía Hipótesis, conceptos, métodos, etc.) junto con las herramientas probabilistas en el análisis de los mercados financieros → cursos en la universidad, congresos, libros, etc.

- **Econofísica**: Surgío en la decada de 1990
  - Nuevos resultados en la teoría moderna de la probabilidad → Aplicaciones de los procesos de Lévy para caracterizar datos financieros con distribuciones no-gaussianas (Mantegna, 1991).
  - Nuevos datos empíricos → automatización de centros financieros con datos en tiempo real de alta frecuencia (1990)
  - Comienza a institucionalizarse desde la publicación de Bouchaud and Stanley (1999) → Physica A, Quantitative finance, Econophysics colloquium (desde 2005), Programa de doctorado en la Universidad de Houston (desde 2006)...

[2] F. Jovanovic, C. Schinckus, J. Hist. Econ. Thought 35 (2013) 319

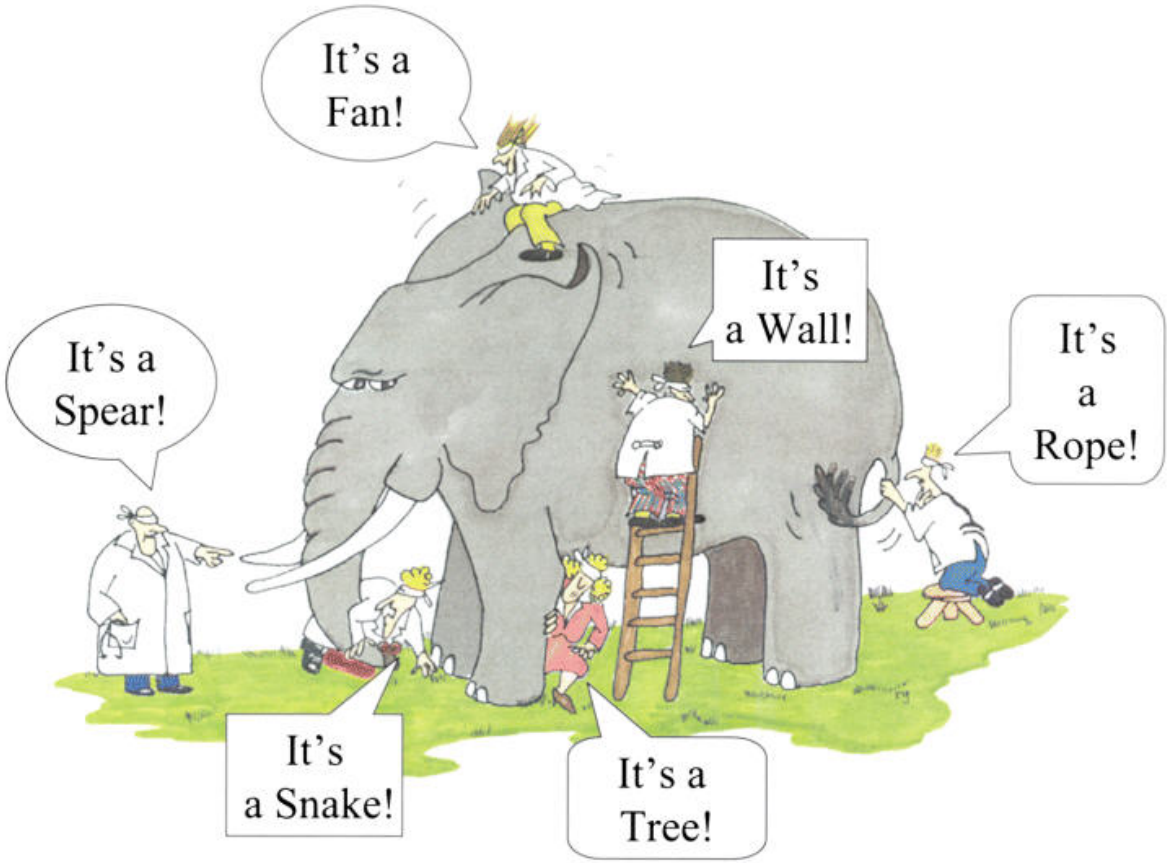
Limitaciones

Limitaciones

- La ausencia de discusión entre los economistas financieros y los econofísicos.

Limitaciones

- La ausencia de discusión entre los economistas financieros y los econofísicos.
- La ausencia de hipótesis económicas.



Correlaciones entre Mercados Financieros

Los movimientos de conjuntos de acciones en el mercado están correlacionados.

Correlaciones entre Mercados Financieros

Los movimientos de conjuntos de acciones en el mercado están correlacionados.

- Es común observar periodos en la que todas las acciones se mueven en conjunto a la alza o la baja.
- En algunos periodos los sectores financieros se mueven unos respecto a otros como resultado de cambios estructurales: en la posesión de la acciones o debido a factores psicológicos.
- ¿Pueden cuantificarse las correlaciones entre diferentes acciones, entre acciones y el índice del mercado, etc.?
- Conocer estás correlaciones es un prerrequisito para la optimización del riesgo en un portafolio de inversión.
- Estas correlaciones son difíciles de obtener.

Matriz de Correlación

Las correlaciones entre los precios o retornos de dos activos financieros i, j de horizonte temporal T, se miden a través de la matriz de correlación

$$C_{i,j} = \frac{\langle [\delta S_i(t) - \langle \delta S_i(t) \rangle][\delta S_j(t) - \langle \delta S_j(t) \rangle] \rangle}{\sigma_i \sigma_j} = \frac{1}{T} \sum_{t=1}^{T} \delta s_i(t) \delta_j(t), \qquad (4)$$

Matriz de Correlación

Las correlaciones entre los precios o retornos de dos activos financieros i, j de horizonte temporal T, se miden a través de la matriz de correlación

$$C_{i,j} = \frac{\langle [\delta S_i(t) - \langle \delta S_i(t) \rangle][\delta S_j(t) - \langle \delta S_j(t) \rangle] \rangle}{\sigma_i \sigma_j} = \frac{1}{T} \sum_{t=1}^{T} \delta s_i(t) \delta_j(t), \qquad (4)$$

donde

(i) Serie de tiempo de los retornos (logarítmicos) del precio de un activo $S(t)$ sobre la escala temporal τ

$$\delta S(t) = \log \left(\frac{S(t)}{S(t-\tau)} \right) \approx \frac{S(t) - S(t-\tau)}{S(t-\tau)} \qquad (5)$$

(ii) Serie de tiempo de los retornos estandarizados ($\mu = 0$, $\sigma^2 = 1$)

$$\delta s(t) = \frac{\delta S(t) - \langle \delta S(t) \rangle}{\sqrt{\langle [\delta S(t)]^2 \rangle - \langle \delta S(t) \rangle^2}} \qquad (6)$$

Efecto de tamaño de muestra

Para el caso de dos series de tiempo no correlacionadas δs_1 y δs_2, cada una de longitud T

$$C_{1,2} = \frac{1}{T} \sum_{t=1}^{T} \delta s_1(t) \delta s_2(t). \tag{7}$$

Efecto de tamaño de muestra

Para el caso de dos series de tiempo no correlacionadas δs_1 y δs_2, cada una de longitud T

$$C_{1,2} = \frac{1}{T} \sum_{t=1}^{T} \delta s_1(t) \delta s_2(t). \tag{7}$$

(i) Asumiendo (de teoría de la medida) que la suma y producto de variables aleatorias es una variable aleatoria, $C_{i,j}$ es la suma de T variables aleatorias con media cero.

Efecto de tamaño de muestra

Para el caso de dos series de tiempo no correlacionadas δs_1 y δs_2, cada una de longitud T

$$C_{1,2} = \frac{1}{T} \sum_{t=1}^{T} \delta s_1(t) \delta s_2(t). \tag{7}$$

(i) Asumiendo (de teoría de la medida) que la suma y producto de variables aleatorias es una variable aleatoria, $C_{i,j}$ es la suma de T variables aleatorias con media cero.

(ii) Más allá de la ausencia de correlaciones (por construcción) entre las dos series de tiempo, para T finita, $C_{1,2}$ es una variable aleatoria diferente de cero.

Efecto de tamaño de muestra

Para el caso de dos series de tiempo no correlacionadas δs_1 y δs_2, cada una de longitud T

$$C_{1,2} = \frac{1}{T} \sum_{t=1}^{T} \delta s_1(t) \delta s_2(t). \tag{7}$$

(i) Asumiendo (de teoría de la medida) que la suma y producto de variables aleatorias es una variable aleatoria, $C_{i,j}$ es la suma de T variables aleatorias con media cero.

(ii) Más allá de la ausencia de correlaciones (por construcción) entre las dos series de tiempo, para T finita, $C_{1,2}$ es una variable aleatoria diferente de cero.

(iii) $C_{1,2}$ viene de una distribución con $\mu = 0$ y $\sigma = 1/\sqrt{T} \Rightarrow C_{1,2} \to 0$, cuando $T \to \infty$.

Efecto de tamaño de muestra

Para el caso de dos series de tiempo no correlacionadas δs_1 y δs_2, cada una de longitud T

$$C_{1,2} = \frac{1}{T} \sum_{t=1}^{T} \delta s_1(t) \delta s_2(t). \tag{7}$$

(i) Asumiendo (de teoría de la medida) que la suma y producto de variables aleatorias es una variable aleatoria, $C_{i,j}$ es la suma de T variables aleatorias con media cero.

(ii) Más allá de la ausencia de correlaciones (por construcción) entre las dos series de tiempo, para T finita, $C_{1,2}$ es una variable aleatoria diferente de cero.

(iii) $C_{1,2}$ viene de una distribución con $\mu = 0$ y $\sigma = 1/\sqrt{T} \Rightarrow C_{1,2} \to 0$, cuando $T \to \infty$.

(iv) El valor finito de T produce un *noise dressing* del coeficiente de correlación.

Caso multivariado

Para las matrices de correlaciones el efecto *noise dressing* se vuelve muy severo.

- Considere la matriz M, compuesta de N series de tiempo de longitud T.

Caso multivariado

Para las matrices de correlaciones el efecto *noise dressing* se vuelve muy severo.

- Considere la matriz M, compuesta de N series de tiempo de longitud T.
- La matriz de correlación se define como $C = \frac{1}{T} MM^T$

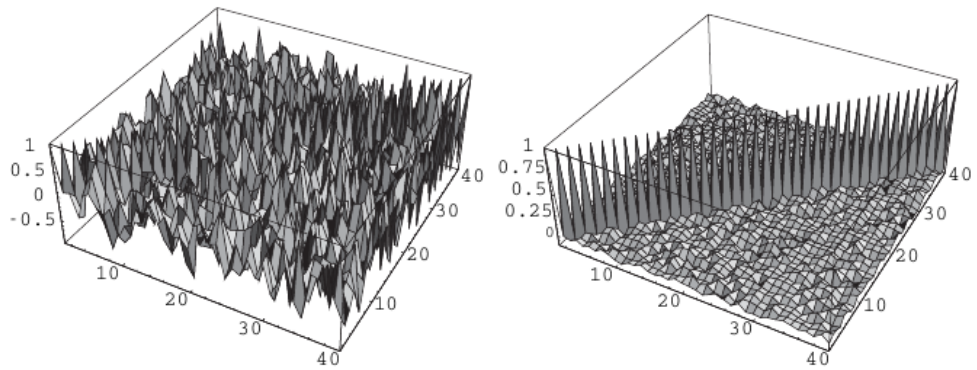


Fig. 5.26. Noise dressing of a correlation matrix. The correlation matrix of 40 uncorrelated time series is shown for a length of 10 steps (*left panel*) and 1000 steps (*right panel*)

Caso multivariado

Para las matrices de correlaciones el efecto *noise dressing* se vuelve muy severo.

- Considere la matriz M, compuesta de N series de tiempo de longitud T.
- La matriz de correlación se define como $C = \frac{1}{T} MM^T$

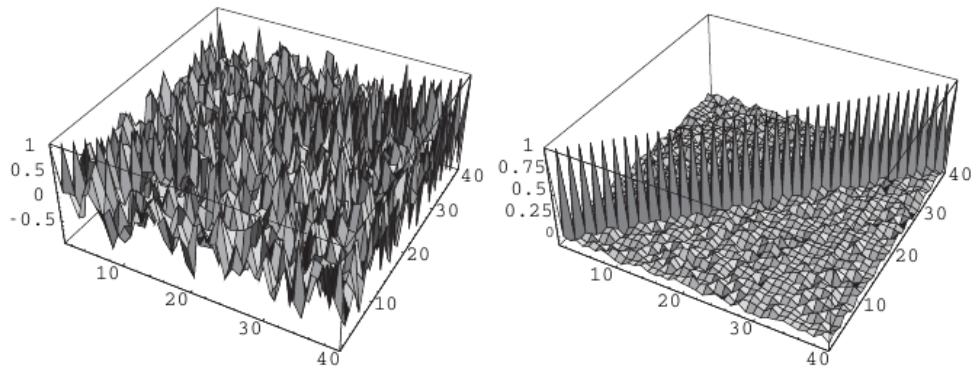


Fig. 5.26. Noise dressing of a correlation matrix. The correlation matrix of 40 uncorrelated time series is shown for a length of 10 steps (*left panel*) and 1000 steps (*right panel*)

- Para $T = 10$, $C_{ii} \sim N(1, 0.48)$ y $C_{i,j}$ $(i \neq j) \sim N(0, .32)$

Caso multivariado

Para las matrices de correlaciones el efecto *noise dressing* se vuelve muy severo.

- Considere la matriz M, compuesta de N series de tiempo de longitud T.
- La matriz de correlación se define como $C = \frac{1}{T} MM^T$

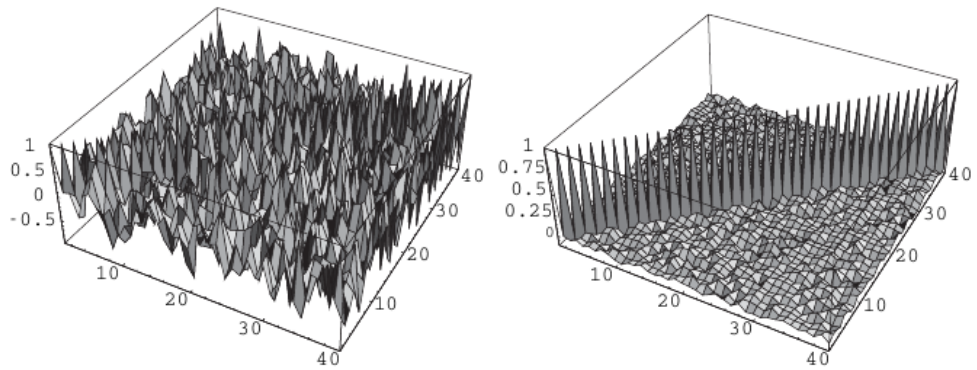


Fig. 5.26. Noise dressing of a correlation matrix. The correlation matrix of 40 uncorrelated time series is shown for a length of 10 steps (*left panel*) and 1000 steps (*right panel*)

- Para $T = 10$, $C_{ii} \sim N(1, 0.48)$ y $C_{i,j}$ $(i \neq j) \sim N(0, .32)$
- Para $T = 1000$, la media es la misma, pero las desviaciones estándar han decrecido por un orden de magnitud.

Caso multivariado

Para las matrices de correlaciones el efecto *noise dressing* se vuelve muy severo.

- Considere la matriz M, compuesta de N series de tiempo de longitud T.
- La matriz de correlación se define como $C = \frac{1}{T} MM^T$

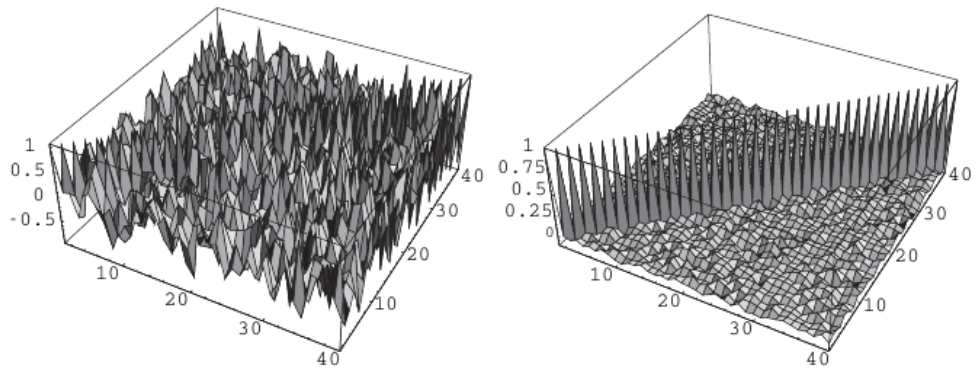


Fig. 5.26. Noise dressing of a correlation matrix. The correlation matrix of 40 uncorrelated time series is shown for a length of 10 steps (*left panel*) and 1000 steps (*right panel*)

- Para $T = 10$, $C_{ii} \sim N(1, 0.48)$ y $C_{i,j}$ $(i \neq j) \sim N(0, .32)$
- Para $T = 1000$, la media es la misma, pero las desviaciones estándar han decrecido por un orden de magnitud.
- Moraleja: se requiere $T \gg N$ para producir matrices de correlación con significancia estadística

¿y las matrices aleatorias?

¿y las matrices aleatorias?

La teoría de matrices aleatorias (RMT), predice que la distribución de probabilidad conjunta de los eigenvalores λ de una matriz aleatoria (con las mismas características que C) sigue la densidad

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \tag{8}$$

siendo

$$\lambda_{min}^{max} = \sigma^2(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}), \tag{9}$$

donde $Q = T/N \geq 1$.

Trabajo de referencia en Econofísica

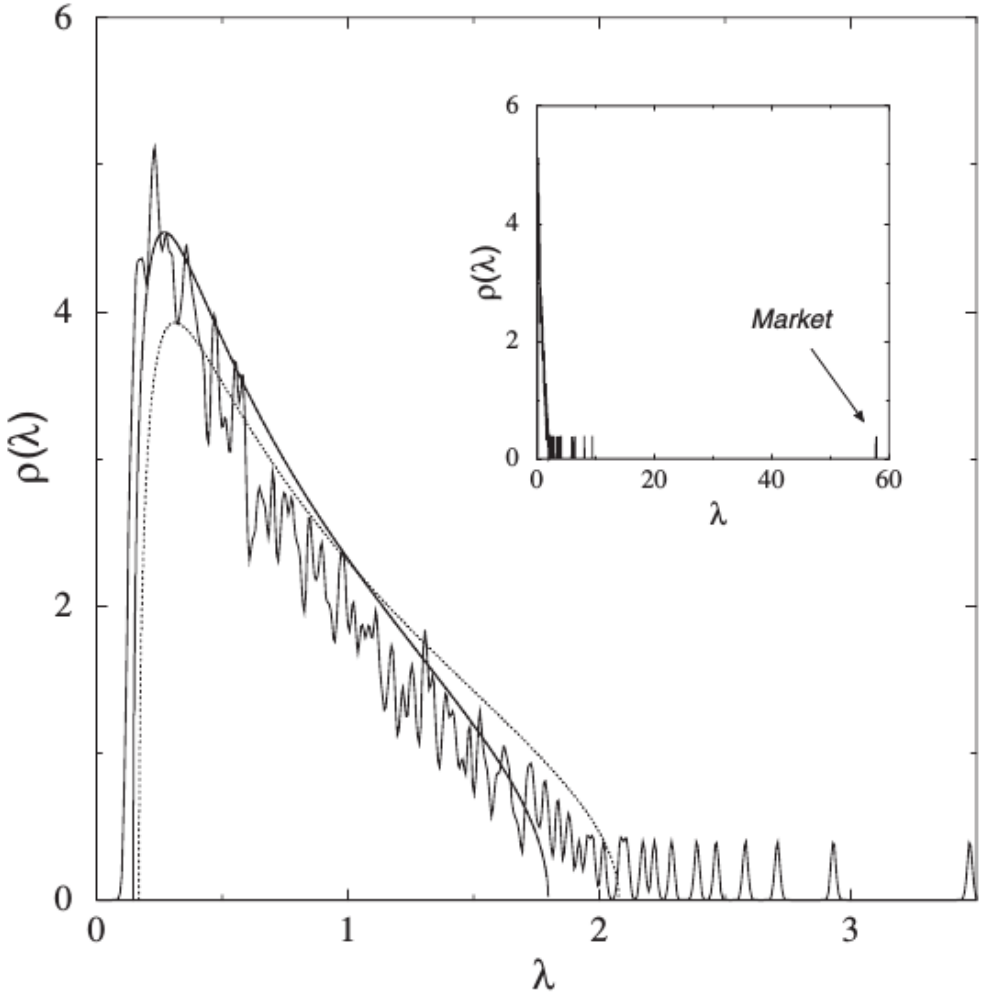


Fig. 5.27. Density of eigenvalues of the correlation matrix of 406 firms out of the S&P500 index. Daily closing prices from 1991 to 1996 were used. The *dotted line* is the prediction of random matrix theory. The *solid line* is a best fit with a variance smaller than the total sample variance. The inset shows the complete spectrum including the largest eigenvalue which lies about 25 times higher than the body of the spectrum. By courtesy of J.-P. Bouchaud. Reprinted from L. Laloux et al.: Phys. Rev. Lett. **83**, 1467 (1999), ©1999 by the American Physical Society

Antecedente desde la Física Matemática

La teoría de matrices aleatorias es un nuevo tipo de mecánica estadística, donde en lugar de tener un ensemble de estados gobernados por el mismo hamiltoniano, se tiene un ensemble de hamiltonianos gobernados por la misma simetría.

Antecedente desde la Física Matemática

La teoría de matrices aleatorias es un nuevo tipo de mecánica estadística, donde en lugar de tener un ensemble de estados gobernados por el mismo hamiltoniano, se tiene un ensemble de hamiltonianos gobernados por la misma simetría.

- Introducida en estadística matemática por Wishart en 1928.
- Muchos matemáticos trabajaron después por interés puramente teórico.
- En 1935 Elie Cartan clasifica los ensembles según la simetría que se conserva (en francés).
- El primer libro que contiene los resultados matemáticos más relevantes lo publicó L. K. Hua en 1958 (en chino).
- En la década de 1950 Wigner la utiliza para tratar con los eigenvalores y eigenvectores de sistemas cuánticos complejos de muchos cuerpos.
- F.J. Dyson y M.L. Mehta hicieron cálculos analıticos (1960).
- Mehta publicó en 1967 un libro con las principales técnicas desarrolladas hasta ese momento.
- Conjetura de caos cuántico, se publicó en un artículo de Bohigas, Giannoni y Schmidt en 1984.

Ensembles Clásicos

- Para sistemas con invariancia ante inversión temporal y simetría rotacional, la matriz hamiltoniana H se puede elegir real y simétrica

$$H = H^T. \tag{10}$$

Ensembles Clásicos

- Para sistemas con invariancia ante inversión temporal y simetría rotacional, la matriz hamiltoniana H se puede elegir real y simétrica

$$H = H^T. \tag{10}$$

- Para los sistemas sin invariancia ante inversión temporal, las matriz H es hermitiana

$$H = H^\dagger. \tag{11}$$

Ensembles Clásicos

- Para sistemas con invariancia ante inversión temporal y simetría rotacional, la matriz hamiltoniana H se puede elegir real y simétrica

$$H = H^T. \tag{10}$$

- Para los sistemas sin invariancia ante inversión temporal, las matriz H es hermitiana

$$H = H^\dagger. \tag{11}$$

- Para los sistemas con invariancia ante inversión temporal con espín $1/2$ y sin simetría rotacional, el hamiltoniano se escribe en términos de las matrices de Pauli σ_γ

$$H^0_{nm} I_2 - i \sum_{\gamma=1}^{3} H^{(\gamma)}_{nm} \sigma_\gamma, \tag{12}$$

donde H^0 es simétrica y H^γ son antisimétricas.

Propiedades

- La densidad de probabilidad de encontrar una matriz particular dentro de uno de estos ensembles está dada por

$$P_{N\beta}(H) \propto e^{-\frac{\beta N}{2} Tr(H^2)} \tag{13}$$

Propiedades

- La densidad de probabilidad de encontrar una matriz particular dentro de uno de estos ensembles está dada por

$$P_{N\beta}(H) \propto e^{-\frac{\beta N}{2} Tr(H^2)} \tag{13}$$

- Las propiedades de simetría y las funciones de peso $P_{N\beta}(H)$ son invariantes bajo transformaciones ortogonales ($\beta = 1$), unitarias ($\beta = 2$), y simplécticas ($\beta = 4$) del hamiltoniano.

Propiedades

- La densidad de probabilidad de encontrar una matriz particular dentro de uno de estos ensembles está dada por

$$P_{N\beta}(H) \propto e^{-\frac{\beta N}{2} Tr(H^2)} \tag{13}$$

- Las propiedades de simetría y las funciones de peso $P_{N\beta}(H)$ son invariantes bajo transformaciones ortogonales ($\beta = 1$), unitarias ($\beta = 2$), y simplécticas ($\beta = 4$) del hamiltoniano.
- A estos ensembles se les conoce como *GOE*, *GUE* y *GSE*.

Propiedades

- La densidad de probabilidad de encontrar una matriz particular dentro de uno de estos ensembles está dada por

$$P_{N\beta}(H) \propto e^{-\frac{\beta N}{2} Tr(H^2)} \tag{13}$$

- Las propiedades de simetría y las funciones de peso $P_{N\beta}(H)$ son invariantes bajo transformaciones ortogonales ($\beta = 1$), unitarias ($\beta = 2$), y simplécticas ($\beta = 4$) del hamiltoniano.
- A estos ensembles se les conoce como *GOE*, *GUE* y *GSE*.
- Cada miembro H_β puede representarse en la forma

$$H_\beta = VEV^{-1}, \tag{14}$$

donde E matriz de eigenvalores y V matrices ortogonales, unitarias o simplécticas (dependiendo de si $\beta = 1$, 2 o 4).

$\Rightarrow$ Al medir propiedades estadísticas de los ensembles la distribución de los eigenvalores es independiente de los eigenvectores.

Ejemplo: Derivación GOE

Teorema

Suponga que $H = (A + A^T)/2$ *es una matrix* $N \times N$*, donde los elementos* $A_{i,j}$ *son variables aleatorias reales i.i.d.* $\sim N(0, 1)$*. Entonces, cuando* $N \to \infty$ *la función de densidad espectral de H converge (a.s.) a la ley del semicírculo de Wigner:*

$$\rho(x) = \frac{1}{\pi}\sqrt{2 - x^2} \tag{15}$$

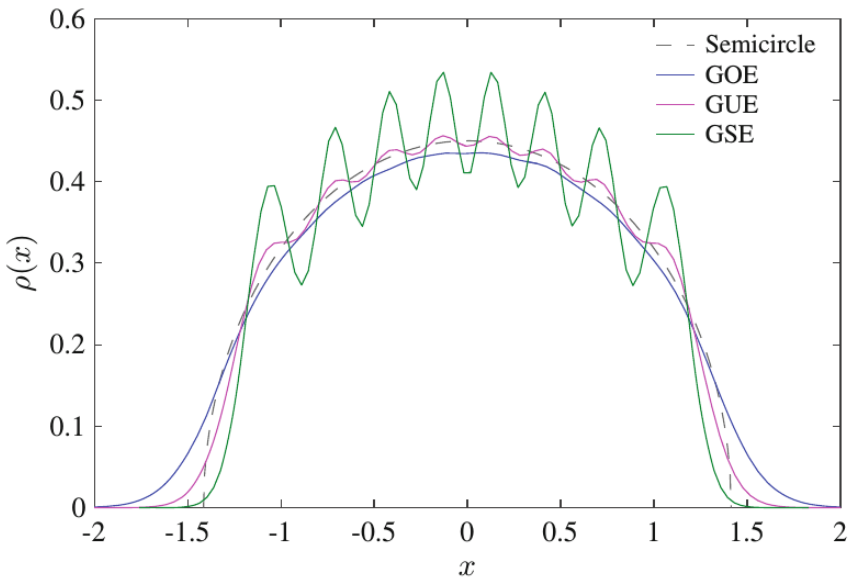


Figura: $N = 8$

Demostración (1ra parte)

Demostración (1ra parte)

- $\rho(H_{11}, \ldots, H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$

Demostración (1ra parte)

- $\rho(H_{11}, \ldots, H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$

Demostración (1ra parte)

- $\rho(H_{11}, \ldots, H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11}, \ldots, H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^2\right)$

Demostración (1ra parte)

- $\rho(H_{11}, \ldots, H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11}, \ldots, H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^2\right)$
- $|J(H \to X, O)| = \prod_{j<k} |x_j - x_k|$ *(Determinante de Vandermonde)*

Demostración (1ra parte)

- $\rho(H_{11}, \ldots, H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11}, \ldots, H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^2\right)$
- $|J(H \to X, O)| = \prod_{j<k} |x_j - x_k|$ *(Determinante de Vandermonde)*
- $\rho(H_{11}, \ldots, H_{NN}) \prod_{i \leq j} dH_{ij} =$
  $\rho(H_{11}(X, O), \ldots, H_{NN}(X, O)) \, |J(H \to X, O)| \, dO \prod_{i=1}^{N} dx_i =$
  $C_N \exp\left(\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| dO \prod_{i=1}^{N} dx_i$

Demostración (1ra parte)

- $\rho(H_{11},\ldots,H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11},\ldots,H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right)$
- $|J(H \to X, O)| = \prod_{j<k} |x_j - x_k|$ *(Determinante de Vandermonde)*
- $\rho(H_{11},\ldots,H_{NN}) \prod_{i\leq j} dH_{ij} =$
  $\rho(H_{11}(X,O),\ldots,H_{NN}(X,O)) \, |J(H \to X, O)| \, dO \prod_{i=1}^{N} dx_i =$
  $C_N \exp\left(\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| dO \prod_{i=1}^{N} dx_i$
- $\rho(x_1,\ldots,x_N) = C_N \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| \int_{V_N} dO$

Demostración (1ra parte)

- $\rho(H_{11},\ldots,H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11},\ldots,H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right)$
- $|J(H \to X, O)| = \prod_{j<k} |x_j - x_k|$ *(Determinante de Vandermonde)*
- $\rho(H_{11},\ldots,H_{NN}) \prod_{i\leq j} dH_{ij} =$
  $\rho(H_{11}(X,O),\ldots,H_{NN}(X,O)) \, |J(H \to X, O)| \, dO \prod_{i=1}^{N} dx_i =$
  $C_N \exp\left(\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| dO \prod_{i=1}^{N} dx_i$
- $\rho(x_1,\ldots,x_N) = C_N \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| \int_{V_N} dO$
- $Vol(V_N) = \int_{V_N} dO = \frac{2^N \pi^{N^2/2}}{\Gamma_N(N/2)}$ *(Medida de Haar)*

Demostración (1ra parte)

- $\rho(H_{11},\ldots,H_{NN}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(H)_{ii}^2}{2}} \prod_{i<j} \frac{1}{\sqrt{\pi}} e^{-(H)_{ij}^2}$
- $H \to \{X, O\} : H = OXO^T$
- $\rho(H_{11},\ldots,H_{NN}) \propto \exp\left(-\frac{1}{2} Tr(H^2)\right) = \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right)$
- $|J(H \to X, O)| = \prod_{j<k} |x_j - x_k|$ *(Determinante de Vandermonde)*
- $\rho(H_{11},\ldots,H_{NN}) \prod_{i\leq j} dH_{ij} =$
  $\rho(H_{11}(X,O),\ldots,H_{NN}(X,O)) \, |J(H \to X, O)| \, dO \prod_{i=1}^{N} dx_i =$
  $C_N \exp\left(\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| dO \prod_{i=1}^{N} dx_i$
- $\rho(x_1,\ldots,x_N) = C_N \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| \int_{V_N} dO$
- $Vol(V_N) = \int_{V_N} dO = \frac{2^N \pi^{N^2/2}}{\Gamma_N(N/2)}$ *(Medida de Haar)*
- $\rho(x_1,\ldots,x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|.$

Demostración (2da parte)

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$
  $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$
- $\langle G_N(z)\rangle \to G_\infty^{(av)}(z) = \int dx' \frac{\rho(x')}{z-x'}$, *para* $N \to \infty$ *(Resolvente, Stieltjes)*

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$
- $\langle G_N(z)\rangle \to G_\infty^{(av)}(z) = \int dx' \frac{\rho(x')}{z-x'}$, *para* $N \to \infty$ *(Resolvente, Stieltjes)*
- $\rho(x) = \frac{1}{\pi} \lim_{\varepsilon\to 0^+} \mathrm{Im}\, G_\infty^{(av)}(x - i\varepsilon)$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$
- $\langle G_N(z)\rangle \to G_\infty^{(av)}(z) = \int dx' \frac{\rho(x')}{z-x'}$, *para* $N \to \infty$ *(Resolvente, Stieltjes)*
- $\rho(x) = \frac{1}{\pi} \lim_{\varepsilon\to 0^+} \mathrm{Im}\, G_\infty^{(av)}(x - i\varepsilon)$
- $G_\infty^{(av)}2(z) - 2zG_\infty^{(av)}(z) + 2 = 0$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$
- $\langle G_N(z)\rangle \to G_\infty^{(av)}(z) = \int dx' \frac{\rho(x')}{z-x'}$, *para* $N \to \infty$ *(Resolvente, Stieltjes)*
- $\rho(x) = \frac{1}{\pi}\,\text{lím}_{\varepsilon\to 0^+}\, \text{Im}\, G_\infty^{(av)}(x - \text{i}\varepsilon)$
- $G_\infty^{(av)} 2(z) - 2zG_\infty^{(av)}(z) + 2 = 0$
- $G_\infty^{(av)}(z) = z \pm \sqrt{z^2 - 2}$

Demostración (2da parte)

- $\rho(x_1, \ldots, x_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k|$.
- $x_i \to x_i\sqrt{N}$
- $Z_N \propto \int_R \prod_{j=1}^{N} dx_j \exp\left(-\frac{1}{2}\sum_{i=1}^{N} x_i^2\right) \prod_{j<k} |x_j - x_k| =$ $\int_R \prod_{j=1}^{N} dx_j e^{-N\mathcal{V}[x]}$, *donde*
- $\mathcal{V}[x] = \frac{1}{2}\sum_i x_i^2 - \frac{1}{2N}\sum_{i\neq j} ln|x_i - x_j|$
- $\frac{\partial \mathcal{V}[x]}{\partial x_i} = 0 \Rightarrow x_i = \frac{1}{N}\sum_{j\neq i} \frac{1}{x_i - x_j}$ *(Método de punto de silla)*
- $G_N(z) = \frac{1}{N} Tr \frac{1}{z-H} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{z-x_i}$
- $\langle G_N(z)\rangle \to G_\infty^{(av)}(z) = \int dx' \frac{\rho(x')}{z-x'}$, *para* $N \to \infty$ *(Resolvente, Stieltjes)*
- $\rho(x) = \frac{1}{\pi} \lim_{\varepsilon\to 0^+} \mathrm{Im}\, G_\infty^{(av)}(x - \mathrm{i}\varepsilon)$
- $G_\infty^{(av)}2(z) - 2zG_\infty^{(av)}(z) + 2 = 0$
- $G_\infty^{(av)}(z) = z \pm \sqrt{z^2 - 2}$
- $\rho(x) = \frac{1}{\pi}\sqrt{2 - x^2}$

Ensemble de Wishart-Laguerre

Teorema

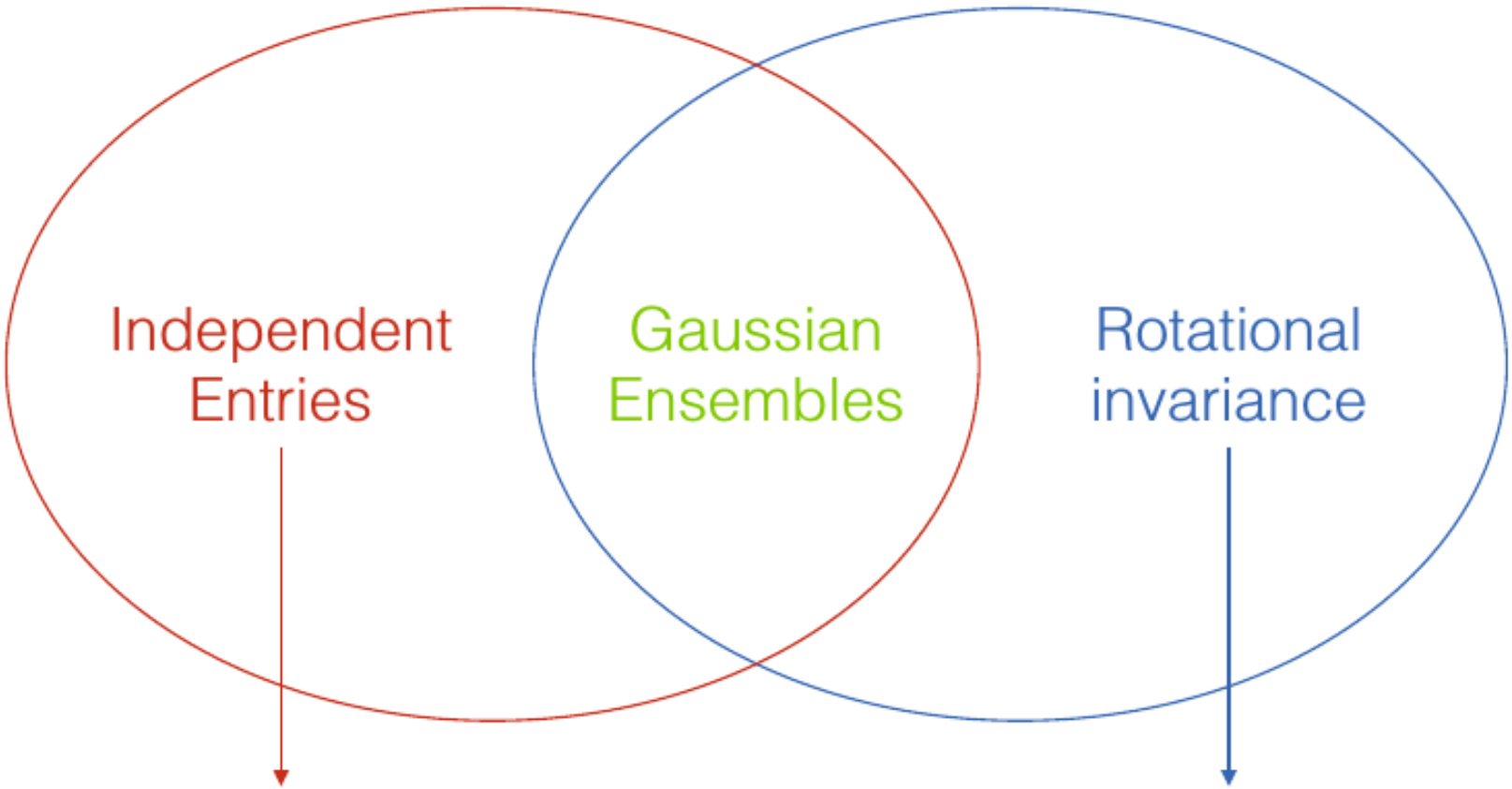
Suponga que A es una matrix $N \times T$, donde los elementos $A_{i,j}$ son variables aleatorias reales i.i.d. $\sim N(0,1)$. Entonces, cuando $N, T \to \infty$, tal que $\frac{T}{N} \to \gamma \in (0,1]$ la función de densidad espectral de $W = T^{-1}AA^T$ converge (a.s.) a la ley Marcenko-Pastur:

$$\rho(x) = \frac{\sqrt{(x_{max} - x)(x - x_{min})}}{2\pi\gamma x}, \tag{16}$$

donde

$$x_{min}^{max} = (1 \pm \sqrt{\gamma})^2. \tag{17}$$

Clasificación de ensembles de matrices aleatorias



$$\rho[H] \propto \prod_{i=1}^{N} f_i(H_{ii}) \prod_{i<j} f_{ij}(H_{ij}) \qquad \rho[H] = \rho[UHU^{-1}]$$

Test para el valor propio más grande

Test para el valor propio más grande

- Suponga que en una muestra de $n = 10$ observaciones con distribución $N_{10}(0, \Sigma)$, encontramos que $\lambda_1 = 4.25$

Test para el valor propio más grande

- Suponga que en una muestra de $n = 10$ observaciones con distribución $N_{10}(0, \Sigma)$, encontramos que $\lambda_1 = 4.25$
- ¿El valor observado es consistente con $\Sigma = \mathrm{I}$, aún y cuando cae fuera del suporte [0,4]?

Test para el valor propio más grande

- Suponga que en una muestra de $n = 10$ observaciones con distribución $N_{10}(0, \Sigma)$, encontramos que $\lambda_1 = 4.25$
- ¿El valor observado es consistente con $\Sigma = \mathrm{I}$, aún y cuando cae fuera del suporte [0,4]?
- En estadística, estamos probando la hipótesis nula $H_0 : \Sigma = \mathrm{I}$ contra $H_A : \Sigma \neq \mathrm{I}$.

Test para el valor propio más grande

- Suponga que en una muestra de $n = 10$ observaciones con distribución $N_{10}(0, \Sigma)$, encontramos que $\lambda_1 = 4.25$
- ¿El valor observado es consistente con $\Sigma = \mathrm{I}$, aún y cuando cae fuera del suporte [0,4]?
- En estadística, estamos probando la hipótesis nula $H_0 : \Sigma = \mathrm{I}$ contra $H_A : \Sigma \neq \mathrm{I}$.
- Se necesita una hipótesis nula para la distribución muestral del valor propio más grande:

$$P\{\hat{\lambda}_1 > t : H_0 \sim W_p(n, \mathrm{I})\} \tag{18}$$

Distribución Tracy-Widow

Resultado

Asuma que $H_0 \sim W_p(n, \mathrm{I})$, *que* $p/n \to \gamma \in (0, \infty)$, *y que* $\hat{\lambda}_1$ *es el valor propio más grande de* $H_0 u = \lambda u$. *Entonces, la distribución del valor propio más grande se aproxima a una de las distribuciones Tracy-Widow* F_β:

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{19}$$

donde $\mu_{np} = (\sqrt{n} + \sqrt{p})^2$, $\sigma_{np} = \mu_{np} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{1/3}$.

Distribución Tracy-Widow

Resultado

Asuma que $H_0 \sim W_p(n, \mathrm{I})$, *que* $p/n \to \gamma \in (0, \infty)$, *y que* $\hat{\lambda}_1$ *es el valor propio más grande de* $H_0 u = \lambda u$. *Entonces, la distribución del valor propio más grande se aproxima a una de las distribuciones Tracy-Widow* F_β:

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{19}$$

donde $\mu_{np} = (\sqrt{n} + \sqrt{p})^2$, $\sigma_{np} = \mu_{np} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{1/3}$.

- Existen formulas elegantes para las funciones de distribución:
  $F_2(s) = \exp\left(-\int_s^\infty (x-s)^2 q(x) dx\right)$, $F_1^2(s) = F_2(s) \exp\left(-\int_s^\infty q(x) dx\right)$

Distribución Tracy-Widow

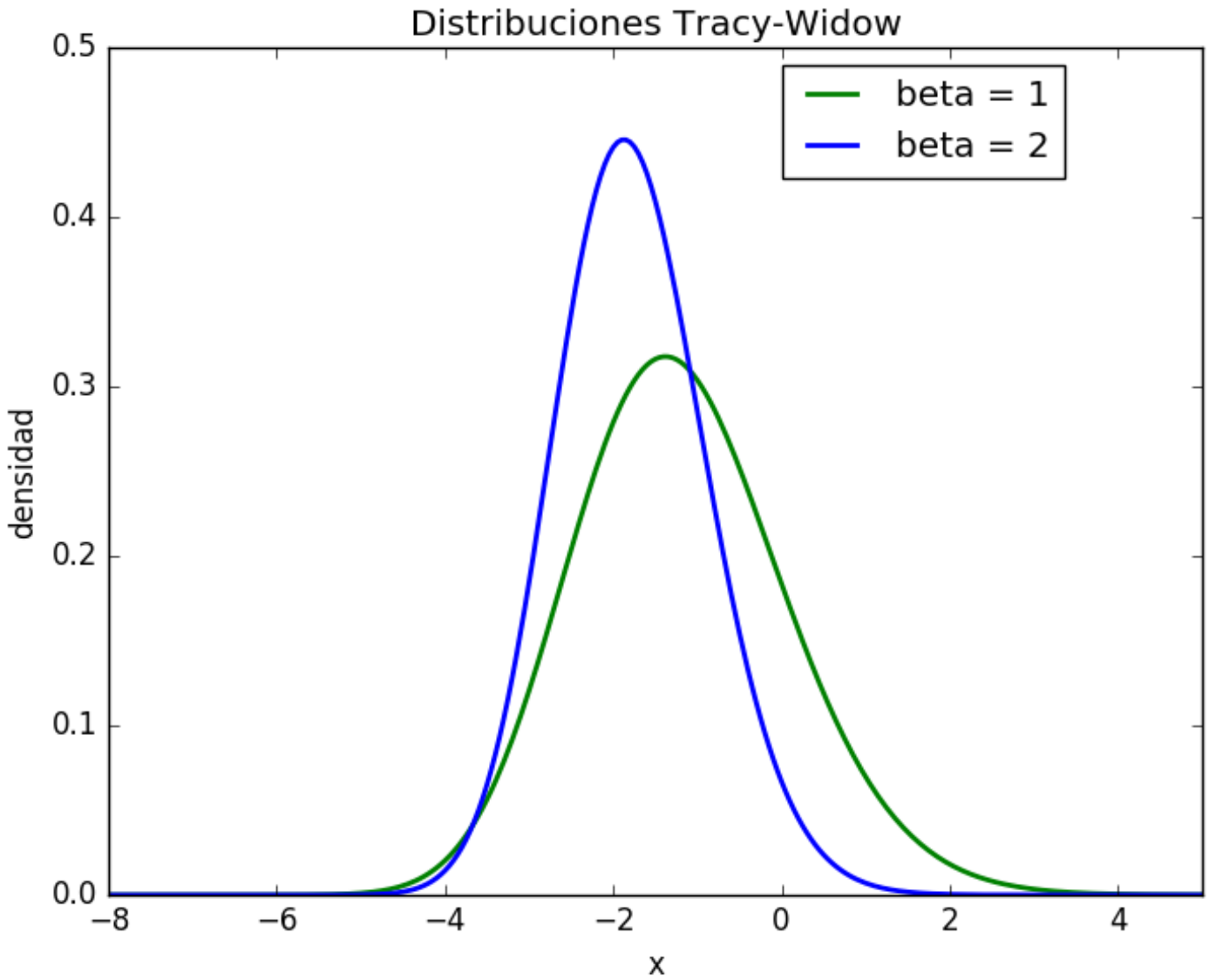
Resultado

Asuma que $H_0 \sim W_p(n, \mathrm{I})$, *que* $p/n \to \gamma \in (0, \infty)$, *y que* $\hat{\lambda}_1$ *es el valor propio más grande de* $H_0 u = \lambda u$. *Entonces, la distribución del valor propio más grande se aproxima a una de las distribuciones Tracy-Widow* F_β:

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{19}$$

donde $\mu_{np} = (\sqrt{n} + \sqrt{p})^2$, $\sigma_{np} = \mu_{np} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{1/3}$.

- Existen formulas elegantes para las funciones de distribución:
  $F_2(s) = \exp\left(-\int_s^\infty (x-s)^2 q(x) dx\right)$, $F_1^2(s) = F_2(s) \exp\left(-\int_s^\infty q(x) dx\right)$
- En términos de la solución para q de la ecuación diferencial no-lineal de segundo orden (Painlevé II):
  $q'' = sq + 2q^3$, $q(s) \sim A_i(s)$ cuando $s \to \infty$.



En el caso de análisis de datos es una función especial como la gaussiana.

Aproximación de 2do orden [3]

Para aplicaciones reales en estadística es importante considerar el caso en que p, n no son muy grandes.

La convergencia $\mathcal{O}(p^{-1/3}) \to \mathcal{O}(p^{-2/3})$

$$|P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} - F_\beta(s)| \to Ce^{-Cs} p^{-2/3} \tag{20}$$

donde $\mu_{np} = (\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}})^2, \sigma_{np} = \mu_{np} \left(\frac{1}{\sqrt{n-\frac{1}{2}}} + \frac{1}{\sqrt{p-\frac{1}{2}}} \right)^{1/3}$.

[3] El Karoui, N., A rate of convergence result for the largest eigenvalue of complex white Wishart Matrices. Ann. Probab. 34 (2006), 2077–2117.

Regresando a nuestro ejemplo

¿$\lambda_1 = 4.25$ es consistente con $H_0 : \Sigma = \mathrm{I}$, cuando $n = p = 10$?

Regresando a nuestro ejemplo

¿$\lambda_1 = 4.25$ es consistente con $H_0 : \Sigma = \mathrm{I}$, cuando $n = p = 10$?

Utilizando la aproximación de 2do orden, la distribución de Tracy-Widow nos dice que existe un 6 % de probabilidad de observar un valor $\lambda_1 > 4.25$ (aunque no se presente ninguna estructura).

Regresando a nuestro ejemplo

¿$\lambda_1 = 4.25$ es consistente con $H_0 : \Sigma = \mathrm{I}$, cuando $n = p = 10$?

Utilizando la aproximación de 2do orden, la distribución de Tracy-Widow nos dice que existe un 6 % de probabilidad de observar un valor $\lambda_1 > 4.25$ (aunque no se presente ninguna estructura).

Comparando con el *benchmark* del 5 %, no existe evidencia suficiente para rechazar la hipótesis nula.

Regresando a nuestro ejemplo

¿$\lambda_1 = 4.25$ es consistente con $H_0 : \Sigma = \mathrm{I}$, cuando $n = p = 10$?

Utilizando la aproximación de 2do orden, la distribución de Tracy-Widow nos dice que existe un 6 % de probabilidad de observar un valor $\lambda_1 > 4.25$ (aunque no se presente ninguna estructura).

Comparando con el *benchmark* del 5 %, no existe evidencia suficiente para rechazar la hipótesis nula.

Surge una pregunta inmediata acerca del test:
¿$P\{\hat{\lambda}_1 > t : W_p(n, \Sigma)\}$, cuando $\Sigma \neq \mathrm{I}$?

Regresando a nuestro ejemplo

¿$\lambda_1 = 4.25$ es consistente con $H_0 : \Sigma = \mathrm{I}$, cuando $n = p = 10$?

Utilizando la aproximación de 2do orden, la distribución de Tracy-Widow nos dice que existe un 6 % de probabilidad de observar un valor $\lambda_1 > 4.25$ (aunque no se presente ninguna estructura).

Comparando con el *benchmark* del 5 %, no existe evidencia suficiente para rechazar la hipótesis nula.

Surge una pregunta inmediata acerca del test:
¿$P\{\hat{\lambda}_1 > t : W_p(n, \Sigma)\}$, cuando $\Sigma \neq \mathrm{I}$?

¿Que tan alejado de 1 debe estar $\lambda_{max}(\Sigma)$ antes de que H_0 sea rechazada?

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

- Desde la teoría de la estadística multivariada apenas hemos tocado la superficie con los ensembles clásicos de RMT.

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

- Desde la teoría de la estadística multivariada apenas hemos tocado la superficie con los ensembles clásicos de RMT.
- Sólo hemos analizado la situación en que Σ no tiene estructura.

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

- Desde la teoría de la estadística multivariada apenas hemos tocado la superficie con los ensembles clásicos de RMT.
- Sólo hemos analizado la situación en que Σ no tiene estructura.
- Las aplicaciones (econometría) en general necesitan una estructura especifica de Σ.

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

- Desde la teoría de la estadística multivariada apenas hemos tocado la superficie con los ensembles clásicos de RMT.
- Sólo hemos analizado la situación en que Σ no tiene estructura.
- Las aplicaciones (econometría) en general necesitan una estructura especifica de Σ.
- En estadística de alta dimensionalidad la variación de los datos se modela al suponer una *señal* de baja dimensionalidad *oculta* en ruido de alta dimensionalidad.

Más allá de la hipótesis nula

¿Bajo que condiciones sobre Σ la aproximación Tracy-Widow sigue siendo válida?

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np} s | H_0\} \to F_\beta(s) \tag{21}$$

¿Modificando μ_{np} y σ_{np} se puede reflejar la estructura de Σ?

- Desde la teoría de la estadística multivariada apenas hemos tocado la superficie con los ensembles clásicos de RMT.
- Sólo hemos analizado la situación en que Σ no tiene estructura.
- Las aplicaciones (econometría) en general necesitan una estructura especifica de Σ.
- En estadística de alta dimensionalidad la variación de los datos se modela al suponer una *señal* de baja dimensionalidad *oculta* en ruido de alta dimensionalidad.
- Si además se asume una estructura aditiva, entonces se busca describir los datos en términos de un modelo de factores.

Modelo de covarianza *spiked* [4]

Asuma que

$$\Sigma = diag(\lambda_1, \ldots, \lambda_M, \sigma_e^2, \ldots, \sigma_e^2) \tag{22}$$

donde $\lambda_1 \geq \cdots \geq \lambda_M \geq \sigma_e^2 > 0$, y $p/n \to \gamma \in (0, \infty)$.

[4] I. Johnstone, On the distribution of the largest principal component, Ann. Statist. 29 (2001) 295–327.

Modelo de covarianza *spiked* [4]

Asuma que

$$\Sigma = diag(\lambda_1, \ldots, \lambda_M, \sigma_e^2, \ldots, \sigma_e^2) \tag{22}$$

donde $\lambda_1 \geq \cdots \geq \lambda_M \geq \sigma_e^2 > 0$, y $p/n \to \gamma \in (0, \infty)$.

Por simplicidad sea $M = 1$, y $\sigma_e^2 = 1$, y consideremos datos complejos $(a + \mathrm{i}b)$.

[4] I. Johnstone, On the distribution of the largest principal component, Ann. Statist. 29 (2001) 295–327.

Modelo de covarianza *spiked* [4]

Asuma que

$$\Sigma = diag(\lambda_1, \ldots, \lambda_M, \sigma_e^2, \ldots, \sigma_e^2) \tag{22}$$

donde $\lambda_1 \geq \cdots \geq \lambda_M \geq \sigma_e^2 > 0$, y $p/n \to \gamma \in (0, \infty)$.

Por simplicidad sea $M = 1$, y $\sigma_e^2 = 1$, y consideremos datos complejos $(a + \mathrm{i}b)$.

Se ha encontrado una *transición de fase* dependiendo del valor que tome λ_1 (Baik et. al., 2006):

[4] I. Johnstone, On the distribution of the largest principal component, Ann. Statist. 29 (2001) 295–327.

Modelo de covarianza *spiked* [4]

Asuma que

$$\Sigma = diag(\lambda_1, \ldots, \lambda_M, \sigma_e^2, \ldots, \sigma_e^2) \tag{22}$$

donde $\lambda_1 \geq \cdots \geq \lambda_M \geq \sigma_e^2 > 0$, y $p/n \to \gamma \in (0, \infty)$.

Por simplicidad sea $M = 1$, y $\sigma_e^2 = 1$, y consideremos datos complejos $(a + \mathrm{i}b)$.

Se ha encontrado una *transición de fase* dependiendo del valor que tome λ_1 (Baik et. al., 2006):

(i) Si $\lambda_1 \leq 1 + \sqrt{\gamma}$, entonces

$$n^{2/3}(\hat{\lambda}_1 - \mu)/\sigma \xrightarrow{d} \begin{cases} F_2 & \text{si } \lambda_1 < 1 + \sqrt{\gamma} \\ \tilde{F}_2 & \text{si } \lambda_1 = 1 + \sqrt{\gamma} \end{cases} \tag{23}$$

donde $\mu = (1 + \sqrt{\gamma})^2$, $\sigma = (1 + \sqrt{\gamma}) \left(1 + \sqrt{\gamma^{-1}}\right)^{1/3}$

[4] I. Johnstone, On the distribution of the largest principal component, Ann. Statist. 29 (2001) 295–327.

(ii) Si $\lambda_1 > 1 + \sqrt{\gamma}$, entonces

$$n^{1/2}(\hat{\lambda}_1 - \mu(\lambda_1))/\sigma(\lambda_1) \xrightarrow{d} N(0,1) \tag{24}$$

donde $\mu(\lambda_1) = \lambda_1 \left(1 + \frac{\gamma}{\lambda_1 - 1}\right)$, $\sigma^2(\lambda_1) = \lambda_1^2(1 - \frac{\gamma}{(\lambda_1-1)^2})$.

(ii) Si $\lambda_1 > 1 + \sqrt{\gamma}$, entonces

$$n^{1/2}(\hat{\lambda}_1 - \mu(\lambda_1))/\sigma(\lambda_1) \xrightarrow{d} N(0,1) \tag{24}$$

donde $\mu(\lambda_1) = \lambda_1 \left(1 + \frac{\gamma}{\lambda_1 - 1}\right)$, $\sigma^2(\lambda_1) = \lambda_1^2(1 - \frac{\gamma}{(\lambda_1-1)^2})$.

Consecuencias:

1. En el valor de $\lambda_1 = 1 + \sqrt{\gamma}$ existe una transición de la distribución Tracy-Widow a la Normal para $\hat{\lambda}_1$.
2. El punto crítico cae dentro de soporte de la distribución de Marcenko-Pastur, cuya cota superior es $(1 + \sqrt{\gamma})^2$.

Teoría del Arbitrage (APT)

APT modela los valores esperados de los retornos como una función lineal de un número determinado de factores.

Teoría del Arbitrage (APT)

APT modela los valores esperados de los retornos como una función lineal de un número determinado de factores.

Es práctica usual determinar el número y magnitud de estos factores mediante PCA.

Teoría del Arbitrage (APT)

APT modela los valores esperados de los retornos como una función lineal de un número determinado de factores.

Es práctica usual determinar el número y magnitud de estos factores mediante PCA.

En 1989 S. J. Brown calibró datos historicos de NYSE con 4 factores independientes.

Teoría del Arbitrage (APT)

APT modela los valores esperados de los retornos como una función lineal de un número determinado de factores.

Es práctica usual determinar el número y magnitud de estos factores mediante PCA.

En 1989 S. J. Brown calibró datos historicos de NYSE con 4 factores independientes.

En este modelo el retorno al tiempo t de la acción k esta dado por:

$$R_{kt} = \sum_{i=1}^{4} b_{ki} f_{it} + e_{kt}, \qquad k = 1, \ldots, K, t = 1, \ldots, T, \tag{25}$$

donde se asume $b \sim N(\beta, \sigma_b^2)$, $f \sim N(0, \sigma_f^2)$, y $e \sim N(0, \sigma_e^2)$, independientes unos de otros.

La matriz de covarianza del proceso tiene la forma dada por el modelo *Spiked* con $M = 4$. En este caso especifico

$$\lambda_j = p\sigma_f^2(\sigma_b^2 + 4\beta\delta_{j1}) + \sigma_e^2, \qquad j = 1, \ldots, 4. \tag{26}$$

La matriz de covarianza del proceso tiene la forma dada por el modelo *Spiked* con $M = 4$. En este caso especifico

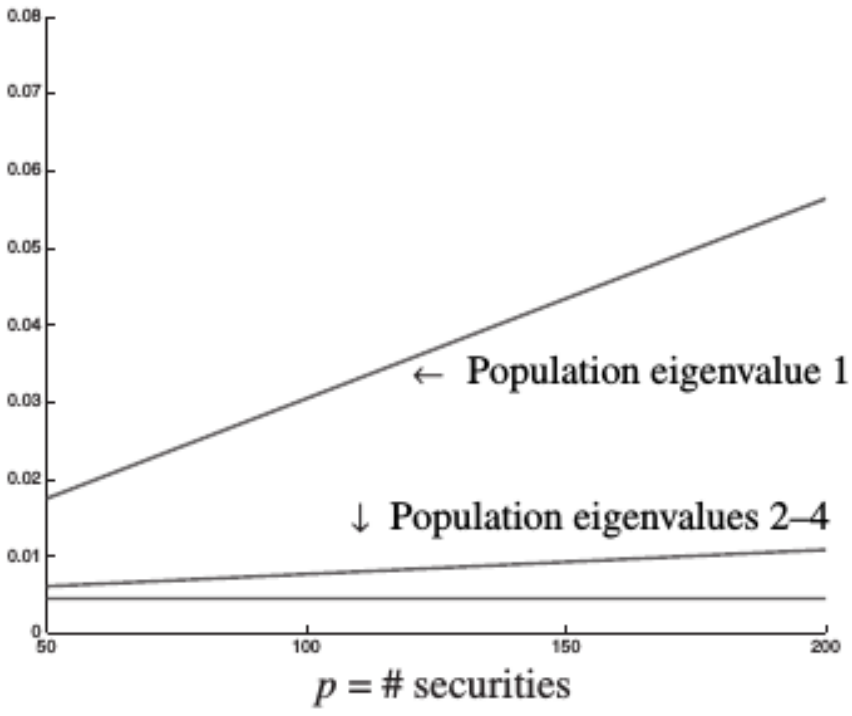
$$\lambda_j = p\sigma_f^2(\sigma_b^2 + 4\beta\delta_{j1}) + \sigma_e^2, \qquad j = 1, \ldots, 4. \tag{26}$$

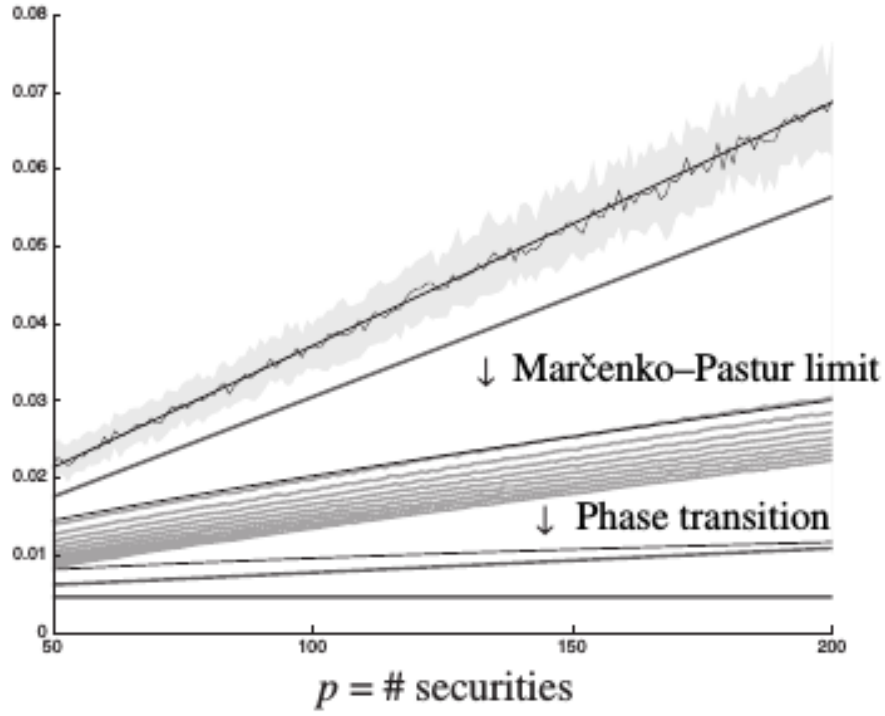
- λ_1 es el valor dominante
- $\lambda_2 = \lambda_3 = \lambda_4$ son valores comunes
- $\lambda_5 = \sigma_e^2$ es el valor base

Inconsistencia PCA

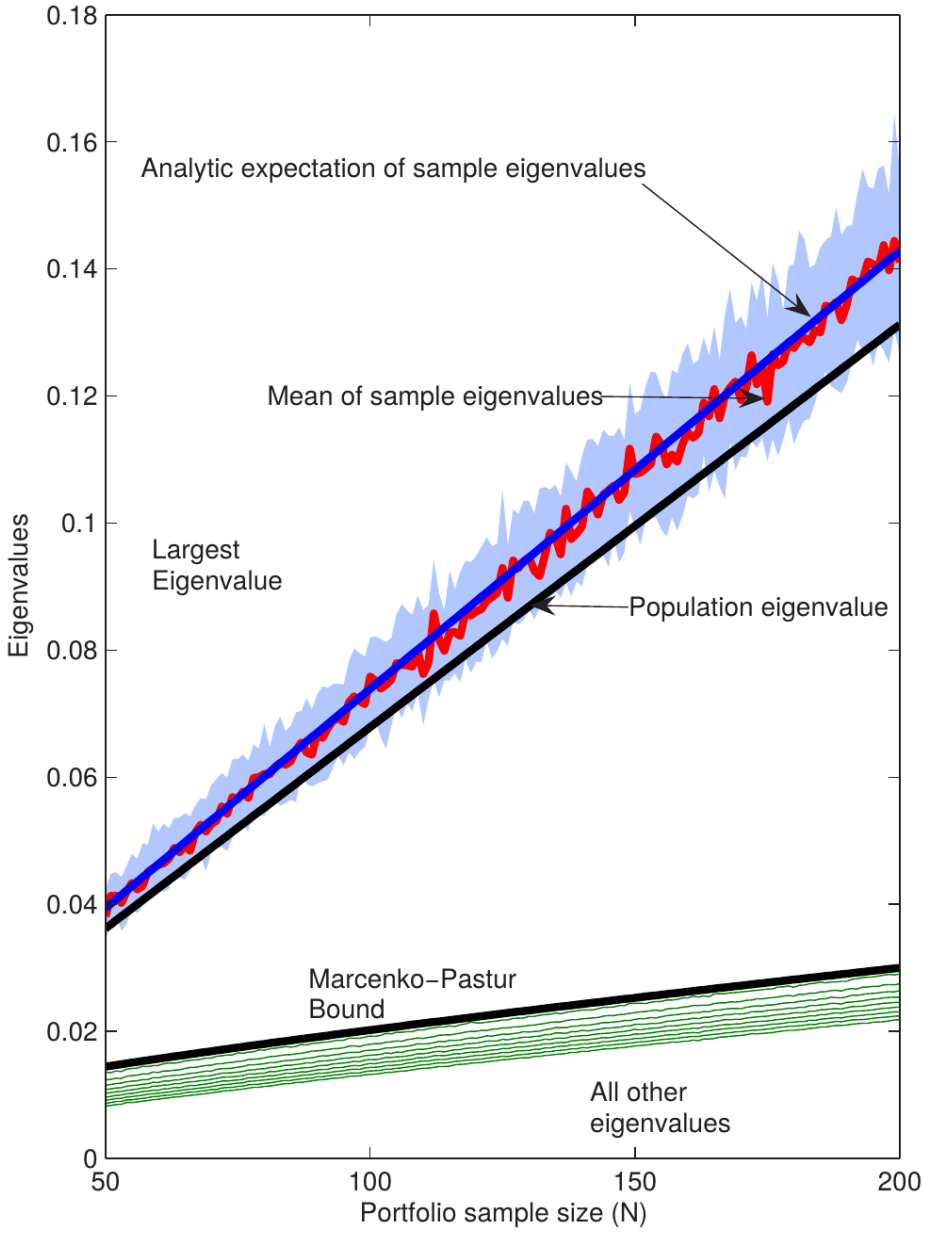
Uno esperaría recuperar la estimación de las λ's

Para $p \in [50, 200]$ cuando $T = 80$ (simulado 300 veces)





Explicación de Harding(2008):



- No es posible identificar los factores $K = 2, 3, 4$ ya que se encuentran en el régimen *equivocado* de $\sigma_e^2(1 + \sqrt{p/T})$.
- Estos valores caen dentro del régimen de ruido idiosincrático (Marcenko-Pastur).
- La estimación por PCA sólo revela el factor del mercado.
- Para este ejemplo $T_{min}(p = 50) \approx 40{,}000$ (120 años de datos diarios).
- se necesita el uso de correcciones sesgadas debido a la finitud de las muestras.

El potencial de la Probabilidad libre en la modelación econométrica

Veamos una metodología para extraer las correlaciones entre dos conjuntos de datos de diferente tamaño (alta dimensionalidad)

El potencial de la Probabilidad libre en la modelación econométrica

Veamos una metodología para extraer las correlaciones entre dos conjuntos de datos de diferente tamaño (alta dimensionalidad)

- Considere las matrices de datos X y Y construidas de M y N series de tiempo, respectivamente

El potencial de la Probabilidad libre en la modelación econométrica

Veamos una metodología para extraer las correlaciones entre dos conjuntos de datos de diferente tamaño (alta dimensionalidad)

- Considere las matrices de datos X y Y construidas de M y N series de tiempo, respectivamente
- Considere el par $\{\lambda, V\}$ de cada matriz de datos para definir el nuevo conjunto de series no-correlacionadas: $\hat{X}$ y $\hat{Y}$

El potencial de la Probabilidad libre en la modelación econométrica

Veamos una metodología para extraer las correlaciones entre dos conjuntos de datos de diferente tamaño (alta dimensionalidad)

- Considere las matrices de datos X y Y construidas de M y N series de tiempo, respectivamente
- Considere el par $\{\lambda, V\}$ de cada matriz de datos para definir el nuevo conjunto de series no-correlacionadas: $\hat{X}$ y $\hat{Y}$
- La matriz de correlación $G_{M\times N}$ entre estas variables está dada por

$$(G)_{jk} = \sum_{i=1}^{T} \hat{Y}_{jt}\hat{X}_{kt} = \hat{Y}\hat{X}^{'}, \tag{27}$$

donde $\hat{X}_{at} = \frac{1}{\sqrt{T\lambda_a}} \sum_b V_{ab}X_{bt}$, y de manera similar para $\hat{Y}$.

El potencial de la Probabilidad libre en la modelación econométrica

Veamos una metodología para extraer las correlaciones entre dos conjuntos de datos de diferente tamaño (alta dimensionalidad)

- Considere las matrices de datos X y Y construidas de M y N series de tiempo, respectivamente
- Considere el par $\{\lambda, V\}$ de cada matriz de datos para definir el nuevo conjunto de series no-correlacionadas: $\hat{X}$ y $\hat{Y}$
- La matriz de correlación $G_{M\times N}$ entre estas variables está dada por

$$(G)_{jk} = \sum_{i=1}^{T} \hat{Y}_{jt}\hat{X}_{kt} = \hat{Y}\hat{X}^{'}, \tag{27}$$

  donde $\hat{X}_{at} = \frac{1}{\sqrt{T\lambda_a}} \sum_b V_{ab}X_{bt}$, y de manera similar para $\hat{Y}$.
- El valor singular más grande s_{max} y sus correspondientes vectores L y R nos dicen como construir el mejor par de variables predictoras-predichas (dados los datos).

En el límite $N, M, T \to \infty$ la función de densidad de s está dada por [5]:

$$\rho(s) = max(1-n, 1-m)\delta(s) + max(m+n-1, 0)\delta(s-1) + \frac{\Re\sqrt{(s^2-\gamma_-)(\gamma_+-s^2)}}{\pi s(1-s^2)}, \tag{28}$$

donde $\gamma_\pm = n + m - 2mn \pm 2\sqrt{mn(1-n)(1-m)}$, y $n = N/T$, $m = M/T$.

[5] J.P. Bouchaud, L. Laloux, M. Miceli, M. Potters,The European Phys. Journ. B 55, 201 (2007)

En el límite $N, M, T \to \infty$ la función de densidad de s está dada por [5]:

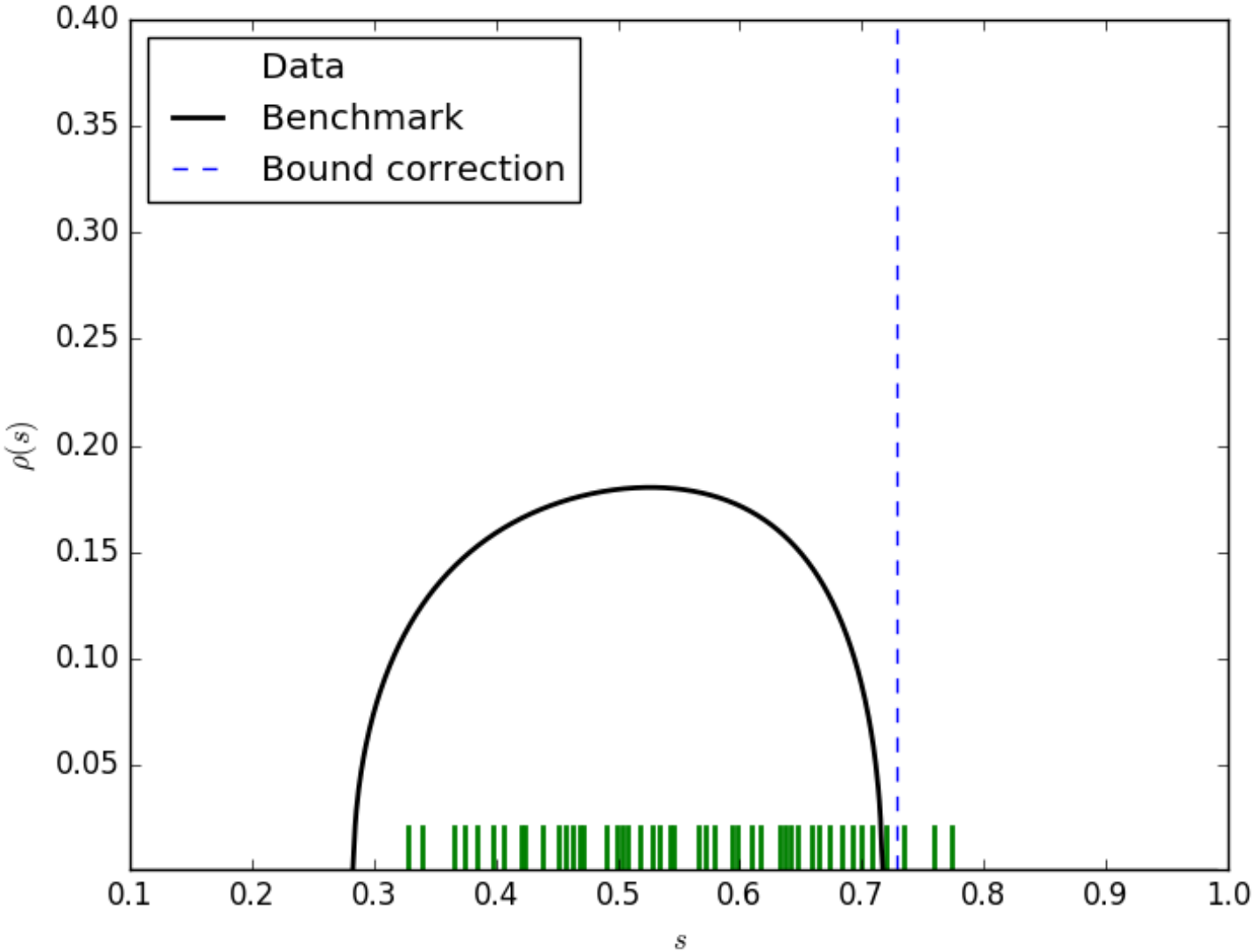
$$\rho(s) = max(1-n, 1-m)\delta(s) + max(m+n-1, 0)\delta(s-1) + \frac{\Re\sqrt{(s^2-\gamma_-)(\gamma_+-s^2)}}{\pi s(1-s^2)}, \tag{28}$$

donde $\gamma_\pm = n + m - 2mn \pm 2\sqrt{mn(1-n)(1-m)}$, y $n = N/T$, $m = M/T$.

Esta densidad es el *benchmark* de ruido debido a la finitud de los datos (sin estructura), más allá de esta cota existen correlaciones genuinas (considerando la corrección muestral de orden $T^{-2/3}$).

[5] J.P. Bouchaud, L. Laloux, M. Miceli, M. Potters,The European Phys. Journ. B 55, 201 (2007)

Conjunto de cryptmonedas predictor-predicho: $N = 200,\ M = 50$



¿Pero como se derivan estos resultados?[6]

Probabilidad Clásica	Probabilidad no-conmutativa (FRV)
Función característica $g_x(z) \equiv \left\langle e^{izx} \right\rangle$	Función de Green: $G_H(z) = \frac{1}{N}\left\langle Tr\frac{1}{z1-H}\right\rangle$ Transformación M: $M(z) = zG_H(z) - 1$
Independencia	Libertad (freeness)
Suma de r.v. El logaritmo es aditivo $r_x(z) \equiv \log g_x(z)$ $r_{x_1+x_2}(z) = r_{x_1}(z) + r_{x_2}(z)$	Suma de f.r.v. La función de Blue es aditiva $G_H(B_H(z)) = B_H(G_H(z)) = z$ $B_{H_1+H_2}(z) = B_{H_1} + B_{H_2}(z) - \frac{1}{z}$
Multiplicación de r.v. Se reduce al problema aditivo Mapeo exponencial $e^{x_1}e^{x_2} = e^{x_1+x_2}$	Multiplicación de f.r.v. La transformación N $M_H(N_H(z)) = N_H(M_H(z)) = z$ $N_{H_1H_2}(z) = \frac{z}{1+z}N_{H_1}(z)N_{H_2}(z)$

[6] D.V. Voiculescu, J. Oper. Theory 18, 223 (1987)

¿Pero como se derivan estos resultados?[6]

Probabilidad Clásica	Probabilidad no-conmutativa (FRV)
Función característica $g_x(z) \equiv \left\langle e^{izx} \right\rangle$	Función de Green: $G_H(z) = \frac{1}{N}\left\langle Tr\frac{1}{z1-H}\right\rangle$ Transformación M: $M(z) = zG_H(z) - 1$
Independencia	Libertad (freeness)
Suma de r.v. El logaritmo es aditivo $r_x(z) \equiv \log g_x(z)$ $r_{x_1+x_2}(z) = r_{x_1}(z) + r_{x_2}(z)$	Suma de f.r.v. La función de Blue es aditiva $G_H(B_H(z)) = B_H(G_H(z)) = z$ $B_{H_1+H_2}(z) = B_{H_1} + B_{H_2}(z) - \frac{1}{z}$
Multiplicación de r.v. Se reduce al problema aditivo Mapeo exponencial $e^{x_1}e^{x_2} = e^{x_1+x_2}$	Multiplicación de f.r.v. La transformación N $M_H(N_H(z)) = N_H(M_H(z)) = z$ $N_{H_1H_2}(z) = \frac{z}{1+z}N_{H_1}(z)N_{H_2}(z)$

Nota: La independencia mutua de todas las entradas de dos matrices no es suficiente para destruir todas las posibles correlaciones angulares entre las dos bases.

[6] D.V. Voiculescu, J. Oper. Theory 18, 223 (1987)

Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuáles son los mercados que se comportan más similares. Sin embargo, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica.

Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuáles son los mercados que se comportan más similares. Sin embargo, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica.

- La transferencia de entropía es una medida dinámica y no simétrica, la cual fue desarrollada inicialmente por Schreiber (2000), y está basada en conceptos relacionados con la entropía de Shannon (1948).

Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuáles son los mercados que se comportan más similares. Sin embargo, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica.

- La transferencia de entropía es una medida dinámica y no simétrica, la cual fue desarrollada inicialmente por Schreiber (2000), y está basada en conceptos relacionados con la entropía de Shannon (1948).
- Esta medida fue diseñada para determinar la direccionalidad de la transferencia de información entre dos procesos, al detectar la asimetría entre sus interacciones.

Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuáles son los mercados que se comportan más similares. Sin embargo, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica.

- La transferencia de entropía es una medida dinámica y no simétrica, la cual fue desarrollada inicialmente por Schreiber (2000), y está basada en conceptos relacionados con la entropía de Shannon (1948).
- Esta medida fue diseñada para determinar la direccionalidad de la transferencia de información entre dos procesos, al detectar la asimetría entre sus interacciones.
- Existen estudios que la relaciona con la flecha del tiempo en termodinámica.

Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuáles son los mercados que se comportan más similares. Sin embargo, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica.

- La transferencia de entropía es una medida dinámica y no simétrica, la cual fue desarrollada inicialmente por Schreiber (2000), y está basada en conceptos relacionados con la entropía de Shannon (1948).
- Esta medida fue diseñada para determinar la direccionalidad de la transferencia de información entre dos procesos, al detectar la asimetría entre sus interacciones.
- Existen estudios que la relaciona con la flecha del tiempo en termodinámica.
- La transferencia de entropía se reduce al test de causalidad de Granger para un modelo auto-regresivo.

Fundamentos desde la Física Estadística

La física estadística consiste en el estudio de las leyes que gobiernan el comportamiento y propiedades de objetos macroscópicos.

Fundamentos desde la Física Estadística

La física estadística consiste en el estudio de las leyes que gobiernan el comportamiento y propiedades de objetos macroscópicos.

- Supongamos que deseamos acomodar N objetos distintos en n cajas. Si denotamos como n_i al número de partículas en la caja i, el factor de peso W de la configuración $N_1, N_2, \ldots, N_n$, con $\sum_{i=1}^{n} N_i = N$, esta dado por $W = \frac{N!}{N_1! N_2! \ldots N_n!}$

Fundamentos desde la Física Estadística

La física estadística consiste en el estudio de las leyes que gobiernan el comportamiento y propiedades de objetos macroscópicos.

- Supongamos que deseamos acomodar N objetos distintos en n cajas. Si denotamos como n_i al número de partículas en la caja i, el factor de peso W de la configuración $N_1, N_2, \ldots, N_n$, con $\sum_{i=1}^{n} N_i = N$, esta dado por $W = \frac{N!}{N_1!N_2!\ldots N_n!}$
- En ausencia de cualquier otra restricción física, la configuración más probable es aquella que maximiza W, lo cual sucede cuando $N_1 = N_2 = \cdots = N/n$.

Fundamentos desde la Física Estadística

La física estadística consiste en el estudio de las leyes que gobiernan el comportamiento y propiedades de objetos macroscópicos.

- Supongamos que deseamos acomodar N objetos distintos en n cajas. Si denotamos como n_i al número de partículas en la caja i, el factor de peso W de la configuración $N_1, N_2, \ldots, N_n$, con $\sum_{i=1}^{n} N_i = N$, esta dado por $W = \frac{N!}{N_1! N_2! \ldots N_n!}$
- En ausencia de cualquier otra restricción física, la configuración más probable es aquella que maximiza W, lo cual sucede cuando $N_1 = N_2 = \cdots = N/n$.
- En la mecánica estadística clásica los objetos son partículas y las cajas representan elementos de igual volumen, con energía ϵ_i asignada a cada caja i.

Ensemble Microcanónico

Si consideramos los sistemas en donde el número total de partículas N y de energía E son constantes:

$$\begin{aligned} \sum_{i=1}^{n} N(i) &= N \\ \sum_{i=1}^{n} \epsilon(i) N(i) &= E, \end{aligned} \tag{29}$$

Ensemble Microcanónico

Si consideramos los sistemas en donde el número total de partículas N y de energía E son constantes:

$$\begin{aligned} \sum_{i=1}^{n} N(i) &= N \\ \sum_{i=1}^{n} \epsilon(i) N(i) &= E, \end{aligned} \tag{29}$$

- Cuando $N \gg 1$, las únicas cantidades físicas relevantes para una descripción termodinámica son N, E y el volumen total.

Ensemble Microcanónico

Si consideramos los sistemas en donde el número total de partículas N y de energía E son constantes:

$$\begin{aligned} \sum_{i=1}^{n} N(i) &= N \\ \sum_{i=1}^{n} \epsilon(i) N(i) &= E, \end{aligned} \tag{29}$$

- Cuando $N \gg 1$, las únicas cantidades físicas relevantes para una descripción termodinámica son N, E y el volumen total.
- La configuración con el máximo factor de peso W esta dada por la expresión de Maxwell-Boltzmann

$$N_i = e^{-\alpha - \beta \epsilon_i} \tag{30}$$

donde α, β, son los multiplicadores de Lagrange dadas las constricciones del sistema.

Entropía de Boltzmann

Este resultado puede ser reescrito en términos de la entropía de Boltzmann

$$s = -k_\beta \sum_{i=1}^{n} p(i) \log p(i) \tag{31}$$

donde $p(i) = N_i/N$ es la fracción de partículas (o la probabilidad de encontrar a una partícula) en la caja i, con las condiciones

$$\begin{aligned} \sum_i p(i) &= 1 \\ \sum_i \epsilon(i) p(i) &= \bar{\epsilon} \end{aligned} \tag{32}$$

donde $\bar{\epsilon}$ es el promedio de la energía sobre todas las cajas o estados i, y $p(i) = e^{-\alpha - \beta\epsilon(i)}$ maximiza la *entropía*.

Entropía de Shannon

Esta definición de entropía termodinámica puede ser generalizada más allá del contexto de la física estadística, y ser asignada a una distribución de probabilidad arbitraria $p(i)$, $i = 1, \ldots, n$, en cuyo caso se le denomina entropía de Shannon o entropía de la información $(= -S)$

$$S = -\sum_{i=1}^{n} p(i) \log p(i) \tag{33}$$

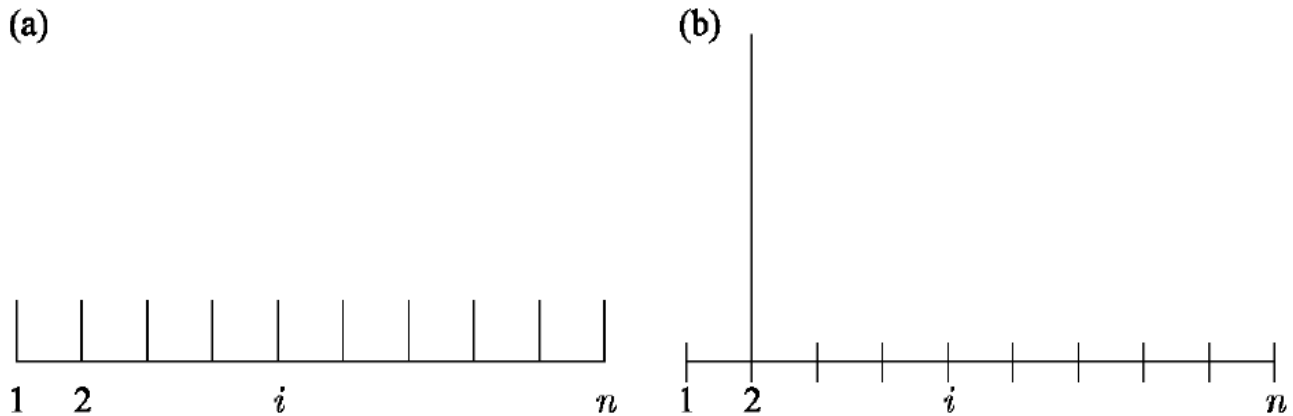


Figura: (a) distribución de mínima información. (b) distribución de máxima información.

Midiendo el flujo de información

Midiendo el flujo de información

- Para una variable aleatoria discreta I con distribución de probabilidad $p(i)$, donde i denota los diferentes valores que la variable I puede tomar, el número promedio de bits requeridos para codificar de manera óptima las realizaciones independientes de la distribución de I esta dado por la fórmula de Shannon.

Midiendo el flujo de información

- Para una variable aleatoria discreta I con distribución de probabilidad $p(i)$, donde i denota los diferentes valores que la variable I puede tomar, el número promedio de bits requeridos para codificar de manera óptima las realizaciones independientes de la distribución de I esta dado por la fórmula de Shannon.
- Para medir el flujo de información entre dos procesos, la entropía de Shannon combina con los conceptos de la distancia de Kullback-Leibler (1951) asumiendo que los procesos subyacentes evolucionan en el tiempo como un proceso de Markov.

Midiendo el flujo de información

- Para una variable aleatoria discreta I con distribución de probabilidad $p(i)$, donde i denota los diferentes valores que la variable I puede tomar, el número promedio de bits requeridos para codificar de manera óptima las realizaciones independientes de la distribución de I esta dado por la fórmula de Shannon.
- Para medir el flujo de información entre dos procesos, la entropía de Shannon combina con los conceptos de la distancia de Kullback-Leibler (1951) asumiendo que los procesos subyacentes evolucionan en el tiempo como un proceso de Markov.
- La propiedad de Markov implica que la probabilidad de observar I al tiempo $t+1$ en el estado i, condicionado sobre las k previas observaciones esta dado por

$$p(i_{t+1}|i_t, \ldots, i_{t-k+1}) = p(i_{t+1}|i_t, \ldots, i_{t-k}). \tag{34}$$

- El número promedio de bits necesarios para codificar un estado adicional del sistema si todos los estados previos son conocidos, está dado por la *tasa de entropía*

$$h_I = S_{I^{(k+1)}} - S_{I^{(k)}} = \sum p(i_{t+1}, i_t^{(k)}) \log p(i_{t+1}|i_t^{(k)}), \qquad (35)$$

donde $i_t^{(k)} \equiv (i_t, \ldots, i_{t-k+1})$

- El número promedio de bits necesarios para codificar un estado adicional del sistema si todos los estados previos son conocidos, está dado por la *tasa de entropía*

$$h_I = S_{I^{(k+1)}} - S_{I^{(k)}} = \sum p(i_{t+1}, i_t^{(k)}) \log p(i_{t+1}|i_t^{(k)}), \tag{35}$$

donde $i_t^{(k)} \equiv (i_t, \ldots, i_{t-k+1})$

- En el caso bivariado, el flujo de información del proceso J al proceso I es medido al cuantificar la desviación del proceso generalizado de Markov

$$p(i_{t+1}|i_t^{(k)}, j_t^{(l)}) = p(i_{t+1}|i_t^{(k)}) \tag{36}$$

- El número promedio de bits necesarios para codificar un estado adicional del sistema si todos los estados previos son conocidos, está dado por la *tasa de entropía*

$$h_I = S_{I^{(k+1)}} - S_{I^{(k)}} = \sum p(i_{t+1}, i_t^{(k)}) \log p(i_{t+1}|i_t^{(k)}), \tag{35}$$

donde $i_t^{(k)} \equiv (i_t, \ldots, i_{t-k+1})$

- En el caso bivariado, el flujo de información del proceso J al proceso I es medido al cuantificar la desviación del proceso generalizado de Markov

$$p(i_{t+1}|i_t^{(k)}, j_t^{(l)}) = p(i_{t+1}|i_t^{(k)}) \tag{36}$$

- Por otro lado, dada la la distribución de probabilidad $p(i)$, el valor excedente de bits que son codificados al utilizar una distribución diferente $q(i)$, esta dado por entropía de Kullback-Leibler:

$$K_I = \sum_i p(i) \log \frac{p(i)}{q(i)}. \tag{37}$$

Combinando los elementos anteriores, obtenemos la *Transferencia de Entropía* (Schreiber, PRL, 2000)

$$T_{J\to I}(k,l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{38}$$

Combinando los elementos anteriores, obtenemos la *Transferencia de Entropía* (Schreiber, PRL, 2000)

$$T_{J\to I}(k,l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{38}$$

- La estimación de TE está comúnmente sesgada debido a efectos de la finitud de la muestra.

Combinando los elementos anteriores, obtenemos la *Transferencia de Entropía* (Schreiber, PRL, 2000)

$$T_{J\to I}(k,l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{38}$$

- La estimación de TE está comúnmente sesgada debido a efectos de la finitud de la muestra.
- Marschinski y Kantz (2002) proponen la transferencia de entropía efectiva para lidiar con este problema

$$ET_{J\to I}(k,l) = T_{J\to I}(k,l) - T_{J_{shuffled}\to I}(k,l) \tag{39}$$

Combinando los elementos anteriores, obtenemos la *Transferencia de Entropía* (Schreiber, PRL, 2000)

$$T_{J\to I}(k,l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{38}$$

- La estimación de TE está comúnmente sesgada debido a efectos de la finitud de la muestra.
- Marschinski y Kantz (2002) proponen la transferencia de entropía efectiva para lidiar con este problema

$$ET_{J\to I}(k,l) = T_{J\to I}(k,l) - T_{J_{shuffled}\to I}(k,l) \tag{39}$$

- $T_{J_{shuffled}\to I}(k,l) \to 0$ cuando $t \to \infty$, por lo que cualquier valor diferente de cero es debido a efectos finitos de la muestra.

Combinando los elementos anteriores, obtenemos la *Transferencia de Entropía* (Schreiber, PRL, 2000)

$$T_{J\to I}(k,l) = \sum_{i,j} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{38}$$

- La estimación de TE está comúnmente sesgada debido a efectos de la finitud de la muestra.
- Marschinski y Kantz (2002) proponen la transferencia de entropía efectiva para lidiar con este problema

$$ET_{J\to I}(k,l) = T_{J\to I}(k,l) - T_{J_{shuffled}\to I}(k,l) \tag{39}$$

- $T_{J_{shuffled}\to I}(k,l) \to 0$ cuando $t \to \infty$, por lo que cualquier valor diferente de cero es debido a efectos finitos de la muestra.
- Para obtener un estimador consistente, este procedimiento se repite muchas veces y se resta el valor promedio de $T_{J_{shuffled}\to I}(k,l)$ a $T_{J\to I}(k,l)$.

Significancia estadística

Una manera de medir la significancia estadística es a través de *Markov block bootstrap* (Dimptfl y Peter, 2013).

Significancia estadística

Una manera de medir la significancia estadística es a través de *Markov block bootstrap* (Dimptfl y Peter, 2013).

- Al contrario del *shuffling*, esta técnica preserva las dependencias de los procesos originales.

Significancia estadística

Una manera de medir la significancia estadística es a través de *Markov block bootstrap* (Dimptfl y Peter, 2013).

- Al contrario del *shuffling*, esta técnica preserva las dependencias de los procesos originales.
- Los bloques del proceso J son reordenados aleatoriamente para formar series de tiempo sintéticas, las cuales preservan las dependencias de J, pero eliminan las dependencias estadísticas entre los procesos J y I.

Significancia estadística

Una manera de medir la significancia estadística es a través de *Markov block bootstrap* (Dimptfl y Peter, 2013).

- Al contrario del *shuffling*, esta técnica preserva las dependencias de los procesos originales.
- Los bloques del proceso J son reordenados aleatoriamente para formar series de tiempo sintéticas, las cuales preservan las dependencias de J, pero eliminan las dependencias estadísticas entre los procesos J y I.
- Al repetir este procedimiento se produce la distribución de TE, estimada bajo la hipótesis nula de ausencia de flujo de información de J a I.

Significancia estadística

Una manera de medir la significancia estadística es a través de *Markov block bootstrap* (Dimptfl y Peter, 2013).

- Al contrario del *shuffling*, esta técnica preserva las dependencias de los procesos originales.
- Los bloques del proceso J son reordenados aleatoriamente para formar series de tiempo sintéticas, las cuales preservan las dependencias de J, pero eliminan las dependencias estadísticas entre los procesos J y I.
- Al repetir este procedimiento se produce la distribución de TE, estimada bajo la hipótesis nula de ausencia de flujo de información de J a I.
- El *p-value* asociado a $H_0 : T_{J \not\to I}$ esta dado por $1 - \hat{q}_{TE}$, donde $\hat{q}_{TE}$ denota el cuantil de la distribución simulada que corresponde a la estimación original de TE.

Estimación discreta

La estimación de TE está basada en datos discretos.

Estimación discreta

La estimación de TE está basada en datos discretos.

- Cuando los datos no tienen estructura discreta, una manera de usar este estimador es recodificar los datos simbólicamente.

Estimación discreta

La estimación de TE está basada en datos discretos.

- Cuando los datos no tienen estructura discreta, una manera de usar este estimador es recodificar los datos simbólicamente.
- Denote los bordes de las n cajas $q_1, q_2, \ldots, q_n$, donde $q_1 < q_2 < \cdots < q_n$, y considere la serie de tiempo y_t. Los datos son recodificados como

$$S_t = \begin{cases} 1 & \text{si} \quad y_t \leq q_1 \\ 2 & \text{si} \quad q_1 \leq y_t \leq q_2 \\ \vdots & \vdots \\ n-1 & \text{si} \quad q_{n-1} \leq y_t \leq q_n \\ n & \text{si} \quad y_t \geq q_n \end{cases} \tag{40}$$

Estimación discreta

La estimación de TE está basada en datos discretos.

- Cuando los datos no tienen estructura discreta, una manera de usar este estimador es recodificar los datos simbólicamente.
- Denote los bordes de las n cajas $q_1, q_2, \ldots, q_n$, donde $q_1 < q_2 < \cdots < q_n$, y considere la serie de tiempo y_t. Los datos son recodificados como

$$S_t = \begin{cases} 1 & \text{si} \quad y_t \leq q_1 \\ 2 & \text{si} \quad q_1 \leq y_t \leq q_2 \\ \vdots & \vdots \\ n-1 & \text{si} \quad q_{n-1} \leq y_t \leq q_n \\ n & \text{si} \quad y_t \geq q_n \end{cases} \tag{40}$$

- Los valores de y_t son remplazados por un entero $(1, 2, \ldots, n)$ de acuerdo a la relación que guarda S_t con el intervalo especificado por los bordes inferior y superior q_1, q_n.

Ejemplo en R

Consideremos una relación lineal entre dos variables aleatorias X y Y:

$$\begin{aligned} x_t &= 0.2x_{t-1} + \varepsilon_{x,t} \\ y_t &= x_{t-1} + \varepsilon_{y,t} \end{aligned} \tag{41}$$

donde $\varepsilon_{x,t}, \varepsilon_{y,t} \sim N(0,1)$.

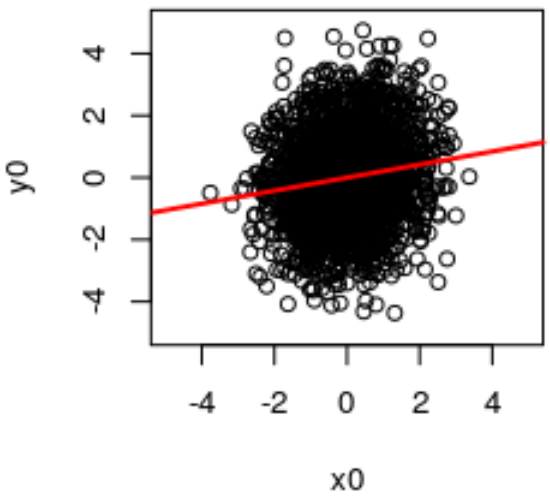
Ejemplo en R

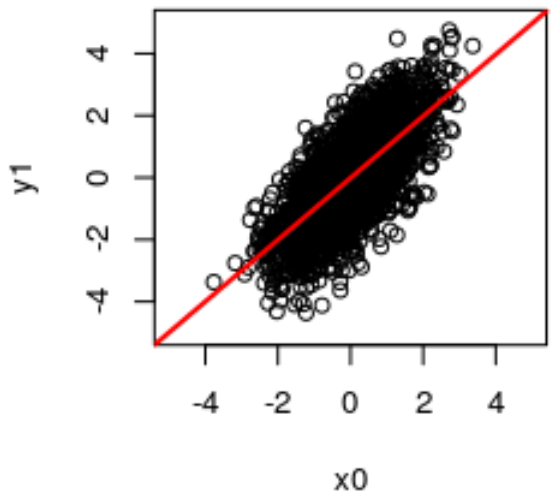
Consideremos una relación lineal entre dos variables aleatorias X y Y:

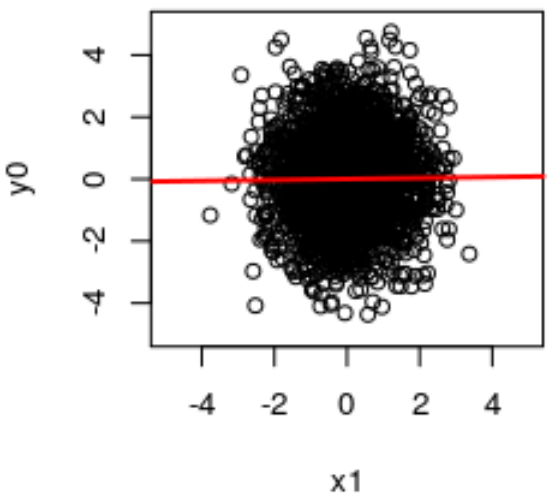
$$\begin{aligned} x_t &= 0.2x_{t-1} + \varepsilon_{x,t} \\ y_t &= x_{t-1} + \varepsilon_{y,t} \end{aligned} \tag{41}$$

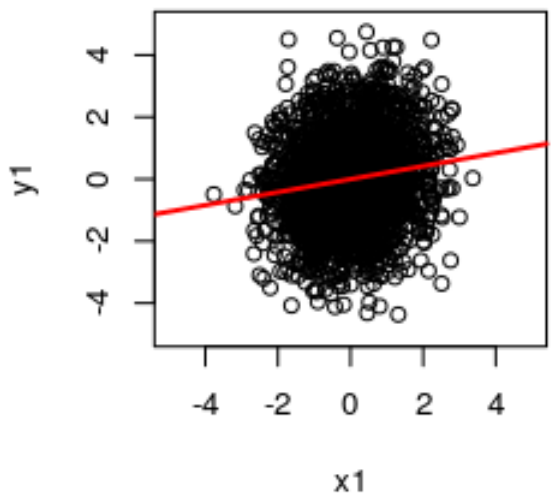
donde $\varepsilon_{x,t}, \varepsilon_{y,t} \sim N(0,1)$.

```
>library(RTransferEntropy)

set.seed(12345)
n <- 2500
x <- rep(0, n + 1)
y <- rep(0, n + 1)
for (i in 2:(n + 1)) {
  x[i] <- 0.2 * x[i - 1] + rnorm(1, 0, 1)
  y[i] <- x[i - 1] + rnorm(1, 0, 1)
}

x0 <- x[1:n-1]
y0 <- y[1:n-1]
x1 <- x[2:n]
y1 <- y[2:n]
```


y0
x0
-4
-2
0
2
4


y1
x0
-4
-2
0
2
4


y0
x1
-4
-2
0
2
4


y1
x1
-4
-2
0
2
4

```
>library(future)
>plan(multiprocess)
>shannon_te <- transfer_entropy(x, y)

Shannon's entropy on 8 cores with 100 shuffles.
x and y have length 2501 (0 NAs removed)
[calculate] X->Y transfer entropy
[calculate] Y->X transfer entropy
[bootstrap] 300 times
Done - Total time 4.43 seconds
```

```
>summary(shannon_te)

Shannon's Transfer Entropy

Coefficients:
             te        ete      se p-value
X->Y 0.0932985 0.0900503 0.0013  <2e-16 ***
Y->X 0.0024552 0.0000000 0.0012    0.69
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bootstrapped TE Quantiles (300 replications):
Direction      0%     25%     50%     75%    100%
     X->Y  0.0007  0.0023  0.0031  0.0039  0.0079
     Y->X  0.0008  0.0022  0.0031  0.0038  0.0094

Number of Observations: 2501
```

Relación con Causalidad Granger

La causalidad de Granger está basada en la premisa de que la *causa* precede el *efecto*, y que la causa contiene información única acerca del efecto, y por lo tanto no se encuentra en otra variable.

Relación con Causalidad Granger

La causalidad de Granger está basada en la premisa de que la *causa* precede el *efecto*, y que la causa contiene información única acerca del efecto, y por lo tanto no se encuentra en otra variable.

- Considere el caso bivariado de dos procesos estocásticos conjuntamente estacionarios X_t, Y_t

$$\begin{aligned}(X_t, \ldots, X_{t+h}, Y_t, \ldots, Y_{t+h}) &\stackrel{d}{=} (X_0, \ldots, X_h, Y_0, \ldots, Y_h) \\ C_{XY}(h) \equiv Cov(X_t, Y_{t+h}) &= Cov(X_0, Y_h)\end{aligned} \tag{42}$$

Relación con Causalidad Granger

La causalidad de Granger está basada en la premisa de que la *causa* precede el *efecto*, y que la causa contiene información única acerca del efecto, y por lo tanto no se encuentra en otra variable.

- Considere el caso bivariado de dos procesos estocásticos conjuntamente estacionarios X_t, Y_t

$$\begin{aligned}(X_t, \ldots, X_{t+h}, Y_t, \ldots, Y_{t+h}) &\stackrel{d}{=} (X_0, \ldots, X_h, Y_0, \ldots, Y_h) \\ C_{XY}(h) \equiv Cov(X_t, Y_{t+h}) &= Cov(X_0, Y_h)\end{aligned} \tag{42}$$

- Sea $F\left(x_t | x_{t-1}^{(k)}, y_{t-1}^{(l)}\right)$ la función de distribución de la variable objetivo X condicionada en la historia conjunta (k,l) de $X_{t-1}^{(k)}$, $Y_{t-1}^{(l)}$.

Relación con Causalidad Granger

La causalidad de Granger está basada en la premisa de que la *causa* precede el *efecto*, y que la causa contiene información única acerca del efecto, y por lo tanto no se encuentra en otra variable.

- Considere el caso bivariado de dos procesos estocásticos conjuntamente estacionarios X_t, Y_t

$$\begin{aligned}(X_t, \ldots, X_{t+h}, Y_t, \ldots, Y_{t+h}) &\stackrel{d}{=} (X_0, \ldots, X_h, Y_0, \ldots, Y_h) \\ C_{XY}(h) \equiv Cov(X_t, Y_{t+h}) &= Cov(X_0, Y_h)\end{aligned} \tag{42}$$

- Sea $F\left(x_t | x_{t-1}^{(k)}, y_{t-1}^{(l)}\right)$ la función de distribución de la variable objetivo X condicionada en la historia conjunta (k,l) de $X_{t-1}^{(k)}$, $Y_{t-1}^{(l)}$.
- Sea $F\left(x_t | x_{t-1}^{(k)}\right)$ la función de distribución de X condicionada sólamente a su propia historia (k).

Relación con Causalidad Granger

La causalidad de Granger está basada en la premisa de que la *causa* precede el *efecto*, y que la causa contiene información única acerca del efecto, y por lo tanto no se encuentra en otra variable.

- Considere el caso bivariado de dos procesos estocásticos conjuntamente estacionarios X_t, Y_t

$$\begin{aligned}(X_t, \ldots, X_{t+h}, Y_t, \ldots, Y_{t+h}) &\stackrel{d}{=} (X_0, \ldots, X_h, Y_0, \ldots, Y_h) \\ C_{XY}(h) \equiv Cov(X_t, Y_{t+h}) &= Cov(X_0, Y_h)\end{aligned} \tag{42}$$

- Sea $F\left(x_t | x_{t-1}^{(k)}, y_{t-1}^{(l)}\right)$ la función de distribución de la variable objetivo X condicionada en la historia conjunta (k,l) de $X_{t-1}^{(k)}$, $Y_{t-1}^{(l)}$.
- Sea $F\left(x_t | x_{t-1}^{(k)}\right)$ la función de distribución de X condicionada sólamente a su propia historia (k).
- Entonces, se dice que la variable Y causa en el sentido de Granger a la variable X (con rezagos k, l) si y sólo si:

$$F\left(x_t | x_{t-1}^{(k)}, y_{t-1}^{(l)}\right) \neq F\left(x_t | x_{t-1}^{(k)}\right). \tag{43}$$

En otras palabras:

Y causa en el sentido de Granger a X, si y sólo si X (condicionada en su propia historia) no es independiente de la historia de Y.

En otras palabras:

Y causa en el sentido de Granger a X, si y sólo si X (condicionada en su propia historia) no es independiente de la historia de Y.

- La conexión con Transferencia de Entropía: $T_{Y \to X}^{(k,l)} \neq 0$.

En otras palabras:

Y causa en el sentido de Granger a X, si y sólo si X (condicionada en su propia historia) no es independiente de la historia de Y.

- La conexión con Transferencia de Entropía: $T_{Y \to X}^{(k,l)} \neq 0$.
- La transferencia de entropía se puede considerar un test estadístico no paramétrico de la idea general de la causalidad de Granger.

En otras palabras:

Y causa en el sentido de Granger a X, si y sólo si X (condicionada en su propia historia) no es independiente de la historia de Y.

- La conexión con Transferencia de Entropía: $T_{Y \to X}^{(k,l)} \neq 0$.
- La transferencia de entropía se puede considerar un test estadístico no paramétrico de la idea general de la causalidad de Granger.
- La idea operacional de la formulación de Granger está basada en la modelación VAR

En otras palabras:

Y causa en el sentido de Granger a X, si y sólo si X (condicionada en su propia historia) no es independiente de la historia de Y.

- La conexión con Transferencia de Entropía: $T_{Y\to X}^{(k,l)} \neq 0$.
- La transferencia de entropía se puede considerar un test estadístico no paramétrico de la idea general de la causalidad de Granger.
- La idea operacional de la formulación de Granger está basada en la modelación VAR
- Consideremos los siguientes modelos VAR (para procesos estacionarios $X_t,\ Y_t$)

$$\begin{aligned} X_t &= A_1 X_{t-1} + \cdots + A_k X_{t-k} + B_1 Y_{t-1} + \cdots + B_l Y_{t-l} + \varepsilon_t \text{(completo)} \\ X_t &= A'_1 X_{t-1} + \cdots + A'_k X_{t-k} + \varepsilon'_t \qquad \text{(reducido)} \end{aligned} \tag{44}$$

donde los parámetros del modelo son las matrices de coeficientes A_i, B_j, A'_i; $\varepsilon_t, \varepsilon'_t$ son los residuos serialmente no correlacionados, y se tienen las matrices de covarianza $\Sigma \equiv c(\varepsilon_t), \Sigma' \equiv c(\varepsilon'_t)$.

Estadístico de Granger

El estadístico de Granger cuantifica el grado en el que modelo completo da una mejor predicción de la variable objetivo en relación al modelo reducido.

[7] J. Geweke. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc., 77(378):304–313, 1982 (Mejor resultado que R^2)

Estadístico de Granger

El estadístico de Granger cuantifica el grado en el que modelo completo da una mejor predicción de la variable objetivo en relación al modelo reducido.

- Una forma conveniente de medir este estadístico es a través de la definición[7]:

$$F_{Y\to X}^{(k,l)} \equiv log\frac{|\Sigma'|}{|\Sigma|}, \tag{45}$$

donde $|\cdot|$ denota el determinante de la matriz.

[7] J. Geweke. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc., 77(378):304–313, 1982 (Mejor resultado que R^2)

Estadístico de Granger

El estadístico de Granger cuantifica el grado en el que modelo completo da una mejor predicción de la variable objetivo en relación al modelo reducido.

- Una forma conveniente de medir este estadístico es a través de la definición[7]:

$$F_{Y\to X}^{(k,l)} \equiv log\frac{|\Sigma'|}{|\Sigma|}, \tag{45}$$

  donde $|\cdot|$ denota el determinante de la matriz.

- La otra aproximación es a través de la *máxima verosimilitud*:

[7] J. Geweke. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc., 77(378):304–313, 1982 (Mejor resultado que R^2)

Estadístico de Granger

El estadístico de Granger cuantifica el grado en el que modelo completo da una mejor predicción de la variable objetivo en relación al modelo reducido.

- Una forma conveniente de medir este estadístico es a través de la definición[7]:

$$F_{Y\to X}^{(k,l)} \equiv log\frac{|\Sigma'|}{|\Sigma|}, \tag{45}$$

  donde $|\cdot|$ denota el determinante de la matriz.
- La otra aproximación es a través de la *máxima verosimilitud*:
- $F_{Y\to X}$ es el estadístico de la relación de verosimilitud logarítmica del modelo completo bajo la hipótesis nula

$$H_0 : B_1 = B_2 = \cdots = B_l = 0. \tag{46}$$

  H_0 se interpreta como *no-causalidad de Granger*.

[7] J. Geweke. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc., 77(378):304–313, 1982 (Mejor resultado que R^2)

Equivalencia con Transferencia de Entropía[8]

Teorema

Si los procesos conjuntos X_t, Y_t son gaussianos (si cualquier subconjunto finito $\{X_{t_1}, Y_{t_2} : (t_1, t_2) \in S\}$ de las variables $\sim N_2(\mu, \sigma^2)$), entonces existe una equivalencia exacta entre la causalidad de Granger y el estadístico de transferencia de entropía:

$$T_{Y \to X}^{(k,l)} = \frac{1}{2} F_{Y \to X}^{(k,l)}. \tag{47}$$

[8] L. Barnett, A. B. Barrett, and A. K. Seth. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett., 103(23):238701, 2009.

Equivalencia con Transferencia de Entropía[8]

Teorema

Si los procesos conjuntos X_t, Y_t son gaussianos (si cualquier subconjunto finito $\{X_{t_1}, Y_{t_2} : (t_1, t_2) \in S\}$ de las variables $\sim N_2(\mu, \sigma^2)$), entonces existe una equivalencia exacta entre la causalidad de Granger y el estadístico de transferencia de entropía:

$$T_{Y \to X}^{(k,l)} = \frac{1}{2} F_{Y \to X}^{(k,l)}. \tag{47}$$

- Existen generalizaciones para una clase muy general de modelos predictivos en el marco de ML (Barnett, 2012).

Puente entre la inferencia causal de datos bajo enfoques autoregresivos y de teoría de la información.

[8] L. Barnett, A. B. Barrett, and A. K. Seth. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett., 103(23):238701, 2009.

Causalidad de Granger vs. Transeferencia de Entropía

Causalidad de Granger vs. Transeferencia de Entropía

- Granger
  1. Se ajusta a los datos
  2. Se extiende al análisis de datos funcionales.
  3. Dentro de VAR es más fácil y eficiente la estimación de los parámetros
  4. Existe una colección conocida de distribuciones muestrales teóricas para realizar la inferencia estadística.

Causalidad de Granger vs. Transeferencia de Entropía

- Granger
  1. Se ajusta a los datos
  2. Se extiende al análisis de datos funcionales.
  3. Dentro de VAR es más fácil y eficiente la estimación de los parámetros
  4. Existe una colección conocida de distribuciones muestrales teóricas para realizar la inferencia estadística.
- ¿Por que usar entonces la transferencia de entropía
  1. Para datos altamente no-lineales y no-gaussianos.
  2. Cuando se tenga un componente grande de MA.
  3. Datos integrados fraccionariamente
  4. Cuando se presente alta heterocedasticidad

Causalidad de Granger vs. Transeferencia de Entropía

- Granger
  1. Se ajusta a los datos
  2. Se extiende al análisis de datos funcionales.
  3. Dentro de VAR es más fácil y eficiente la estimación de los parámetros
  4. Existe una colección conocida de distribuciones muestrales teóricas para realizar la inferencia estadística.
- ¿Por que usar entonces la transferencia de entropía
  1. Para datos altamente no-lineales y no-gaussianos.
  2. Cuando se tenga un componente grande de MA.
  3. Datos integrados fraccionariamente
  4. Cuando se presente alta heterocedasticidad
- Síntesis:
  1. Para procesos no-gaussianos, la transferencia de entropía y la causalidad de Granger simplemente no están midiendo lo mismo.
  2. Si la intención es explicitamente medir el flujo de información, en lugar de la *causalidad a la Granger*, debemos usar la transferencia de entropía.

Ejemplo en R

Consideremos una relación no-lineal entre las variables aleatorias X y Y:

$$\begin{aligned} x_t &= 0.2x_{t-1} + \varepsilon_{x,t} \\ y_t &= \sqrt{|x_{t-1}|} + \varepsilon_{y,t} \end{aligned} \tag{48}$$

donde $\varepsilon_{x,t}, \varepsilon_{y,t} \sim N(0,1)$.

Ejemplo en R

Consideremos una relación no-lineal entre las variables aleatorias X y Y:

$$\begin{aligned} x_t &= 0.2x_{t-1} + \varepsilon_{x,t} \\ y_t &= \sqrt{|x_{t-1}|} + \varepsilon_{y,t} \end{aligned} \tag{48}$$

donde $\varepsilon_{x,t}, \varepsilon_{y,t} \sim N(0,1)$.
Se simulan estos procesos, descartando las primeras 200 observaciones:

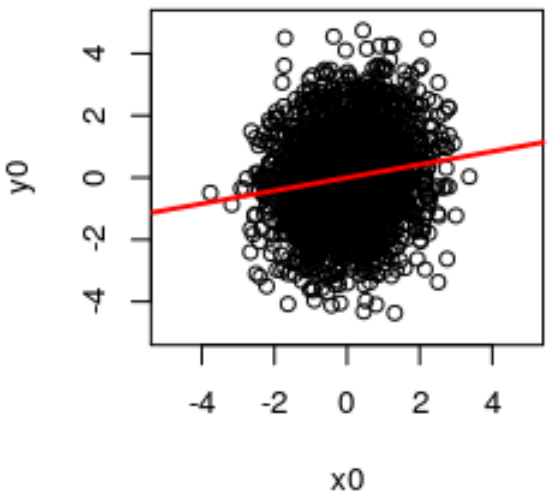
Ejemplo en R

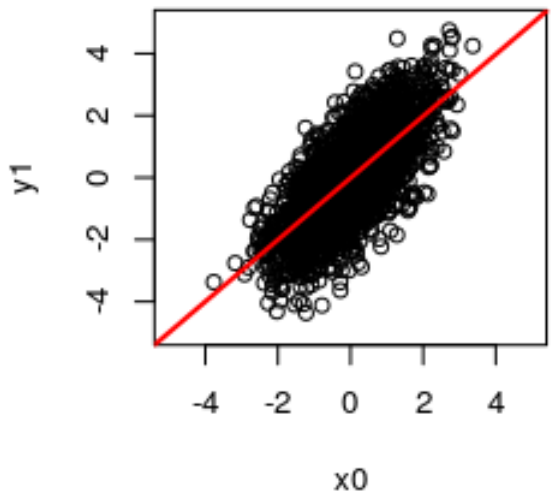
Consideremos una relación no-lineal entre las variables aleatorias X y Y:

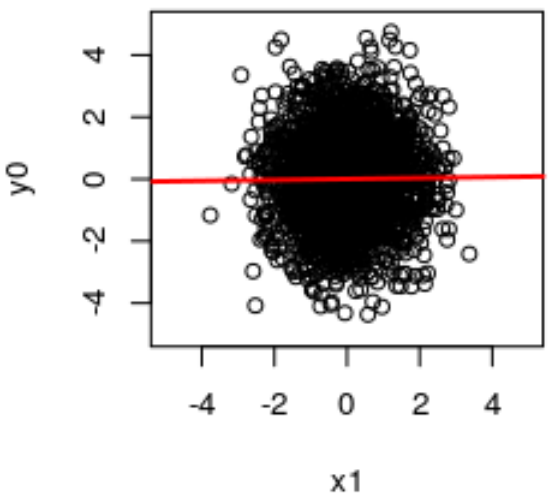
$$\begin{aligned} x_t &= 0.2x_{t-1} + \varepsilon_{x,t} \\ y_t &= \sqrt{|x_{t-1}|} + \varepsilon_{y,t} \end{aligned} \tag{48}$$

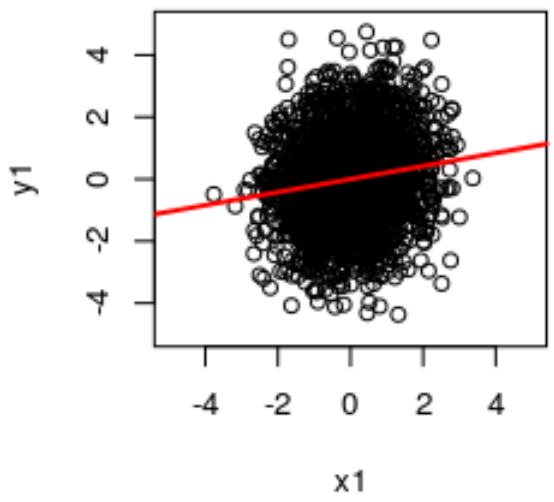
donde $\varepsilon_{x,t}, \varepsilon_{y,t} \sim N(0,1)$.
Se simulan estos procesos, descartando las primeras 200 observaciones:

```
>library(RTransferEntropy)

set.seed(12345)
n <- 2500
x <- rep(0, n + 200)
y <- rep(0, n + 200)
x[1] <- rnorm(1, 0, 1)
y[1] <- rnorm(1, 0, 1)
for (i in 2:(n + 200)) {
x[i] <- 0.2 * x[i - 1] + rnorm(1, 0, 1)
y[i] <- sqrt(abs(x[i - 1])) + rnorm(1, 0, 1)
}
x <- x[-(1:200)]
y <- y[-(1:200)]
```


y0
x0
-4
-2
0
2
4


y1
x0
-4
-2
0
2
4


y0
x1
-4
-2
0
2
4


y1
x1
-4
-2
0
2
4

```
>library(future)
>plan(multiprocess)
>shannon_te <- transfer_entropy(x, y)

Shannon's entropy on 8 cores with 100 shuffles.
x and y have length 2500 (0 NAs removed)
[calculate] X->Y transfer entropy
[calculate] Y->X transfer entropy
[bootstrap] 300 times
Done - Total time 5.32 seconds
```

```
>summary(shannon_te)

Shannon's Transfer Entropy

Coefficients:
           te      ete     se p-value
X->Y 0.014734 0.011276 0.0011  <2e-16 ***
Y->X 0.003250 0.000000 0.0013  0.4367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bootstrapped TE Quantiles (300 replications):
Direction      0%     25%     50%     75%    100%
     X->Y  0.0007  0.0023  0.0031  0.0039  0.0084
     Y->X  0.0010  0.0024  0.0031  0.0038  0.0074

Number of Observations: 2500
```

9

[9] bootstrap: with enough data the empirical distribution will be a good approximation of the true distribution

Ajustando un model VAR

```
>library(vars)
>varfit <- VAR(cbind(x, y), p = 1, type = "const")
>svf <- summary(varfit)
>svf$varresult$y
```

VAR no revela la dependencia no lineal de Y sobre X

```
Call:
lm(formula = y ~ -1 + ., data = datamat)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6282 -0.6811  0.0170  0.6941  3.5409

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
x.l1  -0.004120   0.020498  -0.201    0.841
y.l1   0.006781   0.020016   0.339    0.735
const  0.795104   0.026262  30.276   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.041 on 2496 degrees of freedom
Multiple R-squared:  6.218e-05,Adjusted R-squared:  -0.000739
F-statistic: 0.07761 on 2 and 2496 DF,  p-value: 0.9253
```

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)
- 18 países (G20)

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)
- 18 países (G20)
- objetivo: establecer una red de influencia económica internacional.

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)
- 18 países (G20)
- objetivo: establecer una red de influencia económica internacional.
- Justificación: Entender que tan estable y sustentable es nuestra sistema socioeconómico.

Economía interna y su lugar en la economía global

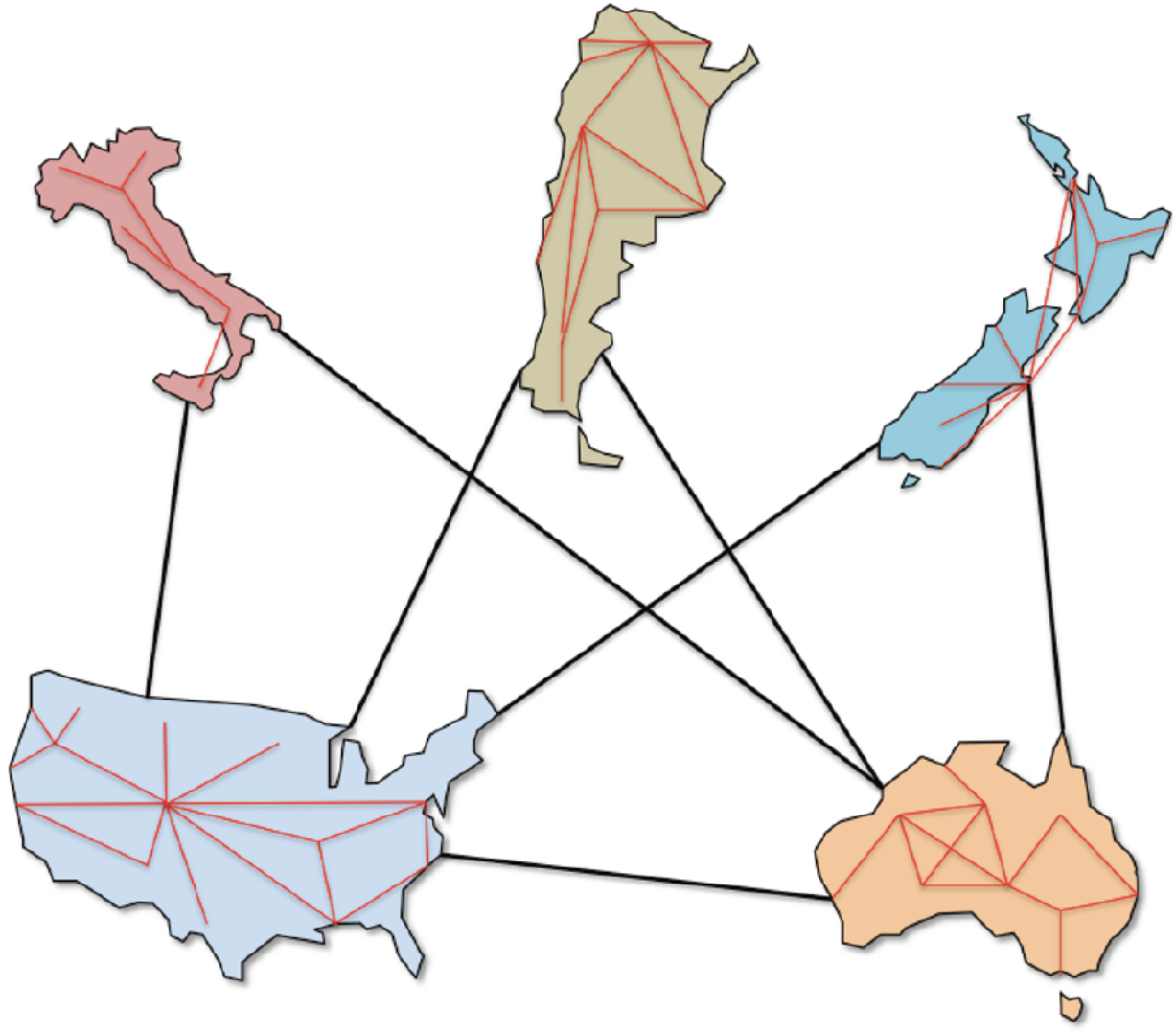
¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)
- 18 países (G20)
- objetivo: establecer una red de influencia económica internacional.
- Justificación: Entender que tan estable y sustentable es nuestra sistema socioeconómico.
- Resultado 1: Los países occidentales son globalmente los más influyentes. Japon ha perdido influencia después de la crisis financiera de asia en 1997.

Economía interna y su lugar en la economía global

¿Hasta qué punto se puede utilizar la transferencia de entropía para medir las señales económicas y financieras tanto dentro como entre los países?

- 192 observaciones mensuales (1990s, 2000s)
- 5 variables macroeconómicas:
  1. Consumer price index (CPI)
  2. Industrial production index (IPI)
  3. Exchange rate (XR)
  4. Stock market index (SMI)
  5. Trade balance (TB)
- 18 países (G20)
- objetivo: establecer una red de influencia económica internacional.
- Justificación: Entender que tan estable y sustentable es nuestra sistema socioeconómico.
- Resultado 1: Los países occidentales son globalmente los más influyentes. Japon ha perdido influencia después de la crisis financiera de asia en 1997.
- Resultado 2: La influencia es transmitida más significativamente entre los países europeos en comparación con Asia y America.



10

[10] J. Kim, G. Kim, S. An, Y.-K. Kwon, and S. Yoon. Entropy-based analysis and bioinformatics-inspired integration of global economic information transfer. PloS One, 8(1):e51986, 2013.

Moraleja

Moraleja

- Entender tanto la fuerza como la dirección de indicadores macroeconómicos en un contexto internacional proporciona una visión importante de los efectos inducidos que otros países sienten como resultado de la crisis económica interna de un país.

Moraleja

- Entender tanto la fuerza como la dirección de indicadores macroeconómicos en un contexto internacional proporciona una visión importante de los efectos inducidos que otros países sienten como resultado de la crisis económica interna de un país.
- Nos promete el uso de una herramienta no-paramétrica (model-free) que colabore en la siguiente evolución de la teoría económica, en su análisis y modelacíón.

Economía conductual

La economía conductual argumenta que las estrategias de inversión en la bolsa han olvidado el importante factor humano por enfocar su análisis desde el marco de referencia de HME.

Economía conductual

La economía conductual argumenta que las estrategias de inversión en la bolsa han olvidado el importante factor humano por enfocar su análisis desde el marco de referencia de HME.

- Si los precios son correctos $\Rightarrow$ no se puede generar ganancias del comercio en la bolsa.

pero en este caso, la implicación inversa no es verdadera:

- Si no se pueden generar ganancias del comercio en la bolsa $\not\Rightarrow$ los precios son correctos.

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}
  3. stemming $\rightarrow$ {`sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx`}

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}
  3. stemming $\rightarrow$ {`sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx`}
  4. scoring $\rightarrow$ Harvard IV dictionary: 8,000 palabras, 182 categorías.

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}
  3. stemming $\rightarrow$ {`sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx`}
  4. scoring $\rightarrow$ Harvard IV dictionary: 8,000 palabras, 182 categorías.
  5. categorize $\rightarrow$ {`0, 0, 0, 0, negative, 0, 0, 0, 0, 0, negative, 0, 0`}

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}
  3. stemming $\rightarrow$ {`sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx`}
  4. scoring $\rightarrow$ Harvard IV dictionary: 8,000 palabras, 182 categorías.
  5. categorize $\rightarrow$ {`0, 0, 0, 0, negative, 0, 0, 0, 0, 0, negative, 0, 0`}
  6. $polarity = \frac{positive - negative}{positive + negative} = -1$

Datos Twitter

- N = 20 países
- T = 166 días (2014)
- Procedimiento para obtener series de tiempo:
  1. `"#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX"`
  2. tokenization $\rightarrow$ {`SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX`}
  3. stemming $\rightarrow$ {`sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx`}
  4. scoring $\rightarrow$ Harvard IV dictionary: 8,000 palabras, 182 categorías.
  5. categorize $\rightarrow$ {`0, 0, 0, 0, negative, 0, 0, 0, 0, 0, negative, 0, 0`}
  6. $polarity = \frac{positive - negative}{positive + negative} = -1$
  7. $P_k(t) \rightarrow$ promedio diario de polaridad para cada país k al día t

Datos NYT

- N = 40 países; T = 217 días (2015-2016).

Datos NYT

- N = 40 países; T = 217 días (2015-2016).

The New York Times | http://nyti.ms/24NJ0xs

AMERICAS

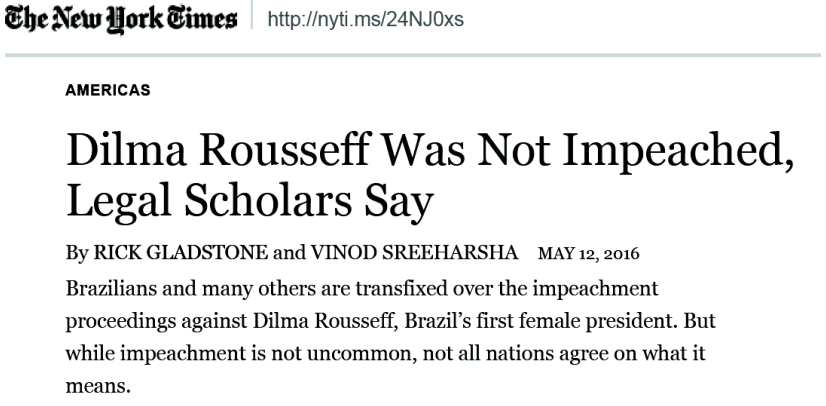
Dilma Rousseff Was Not Impeached, Legal Scholars Say

By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016

Brazilians and many others are transfixed over the impeachment proceedings against Dilma Rousseff, Brazil's first female president. But while impeachment is not uncommon, not all nations agree on what it means.

Datos NYT

- N = 40 países; T = 217 días (2015-2016).



The New York Times | http://nyti.ms/24NJ0xs

AMERICAS

Dilma Rousseff Was Not Impeached, Legal Scholars Say

By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016

Brazilians and many others are transfixed over the impeachment proceedings against Dilma Rousseff, Brazil's first female president. But while impeachment is not uncommon, not all nations agree on what it means.

```
data-title="Dilma Rousseff Was Not Impeached, Legal
Scholars Say"data-author="By RICK GLADSTONE and VINOD
SREEHARSHA"data-publish-date="May 12, 2016"...
<p class="story-body-text
story-content"data-para-count="212"data-total-count="212»Brazilians
and many others are transfixed over the impeachment
proceedings against Dilma Rousseff, Brazil's
first female president. But while impeachment
is not uncommon, not all nations agree on
what it means.</p><p class="story-body-text
story-content"data-para-count="213"data-total-count="425»
```

Datos NYT

- N = 40 países; T = 217 días (2015-2016).

The New York Times | http://nyti.ms/24NJ0xs

AMERICAS

Dilma Rousseff Was Not Impeached, Legal Scholars Say

By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016

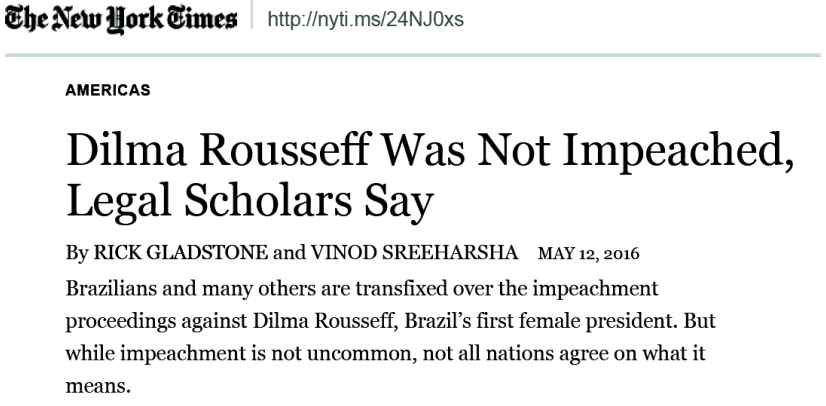
Brazilians and many others are transfixed over the impeachment proceedings against Dilma Rousseff, Brazil's first female president. But while impeachment is not uncommon, not all nations agree on what it means.

```
data-title="Dilma Rousseff Was Not Impeached, Legal
Scholars Say"data-author="By RICK GLADSTONE and VINOD
SREEHARSHA"data-publish-date="May 12, 2016"...
<p class="story-body-text
story-content"data-para-count="212"data-total-count="212»Brazilians
and many others are transfixed over the impeachment
proceedings against Dilma Rousseff, Brazil's
first female president. But while impeachment
is not uncommon, not all nations agree on
what it means.</p><p class="story-body-text
story-content"data-para-count="213"data-total-count="425»
```

```
Americas
Dilma Rousseff Was Not Impeached, Legal Scholars Say
By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016
Brazilians and many others are transfixed over the
impeachment proceedings against Dilma Rousseff, Brazil's
first female president. But while impeachment is not
uncommon, not all nations agree on what it means.
```

Datos NYT

- N = 40 países; T = 217 días (2015-2016).


The New York Times | http://nyti.ms/24NJ0xs

AMERICAS

Dilma Rousseff Was Not Impeached, Legal Scholars Say

By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016

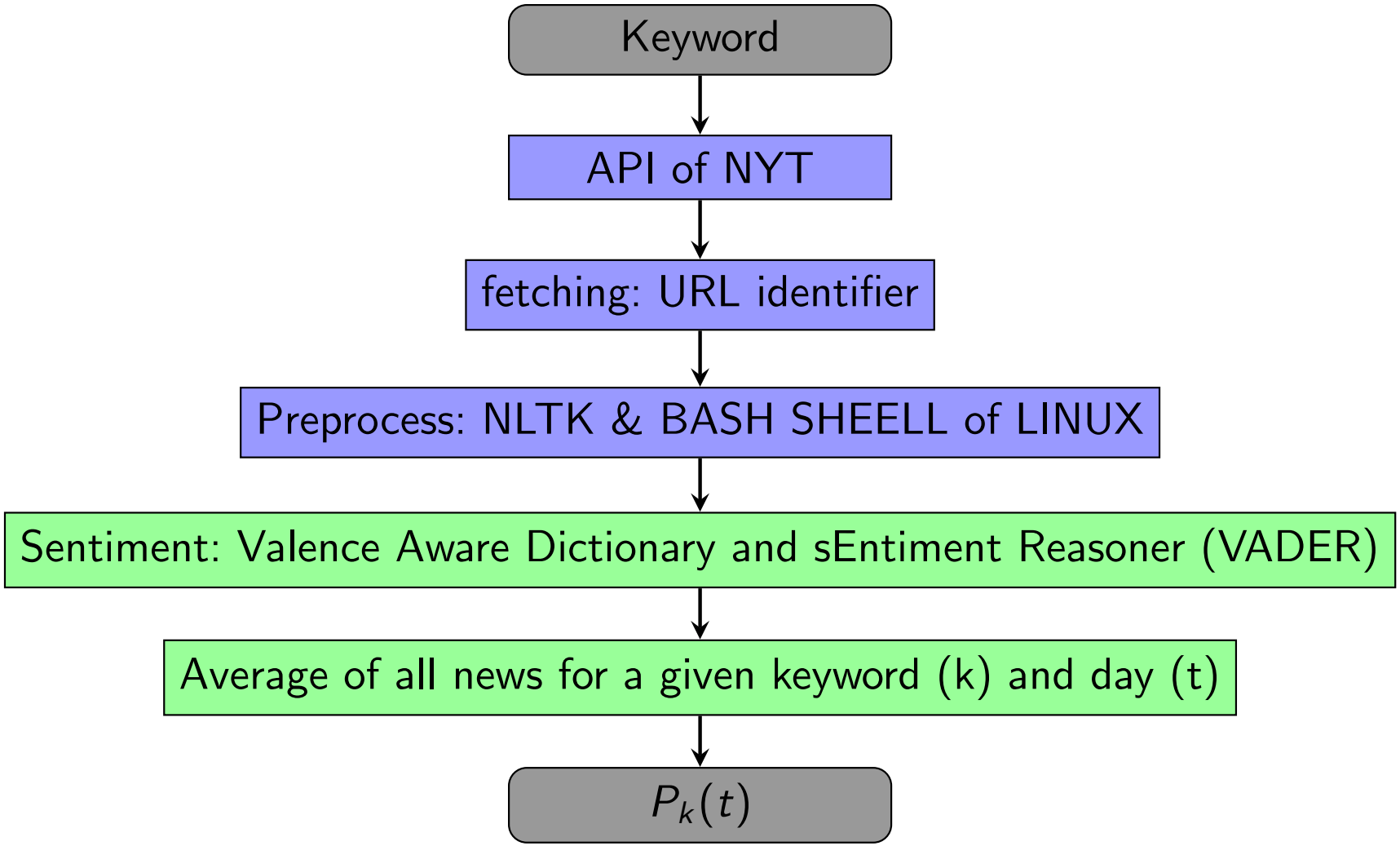
Brazilians and many others are transfixed over the impeachment proceedings against Dilma Rousseff, Brazil's first female president. But while impeachment is not uncommon, not all nations agree on what it means.

```
data-title="Dilma Rousseff Was Not Impeached, Legal
Scholars Say"data-author="By RICK GLADSTONE and VINOD
SREEHARSHA"data-publish-date="May 12, 2016"...
<p class="story-body-text
story-content"data-para-count="212"data-total-count="212»Brazilians
and many others are transfixed over the impeachment
proceedings against Dilma Rousseff, Brazil's
first female president. But while impeachment
is not uncommon, not all nations agree on
what it means.</p><p class="story-body-text
story-content"data-para-count="213"data-total-count="425»
```

```
Americas
Dilma Rousseff Was Not Impeached, Legal Scholars Say
By RICK GLADSTONE and VINOD SREEHARSHA MAY 12, 2016
Brazilians and many others are transfixed over the
impeachment proceedings against Dilma Rousseff, Brazil's
first female president. But while impeachment is not
uncommon, not all nations agree on what it means.
```

$$\text{VADER} \Rightarrow P_k(t) = -0.0369$$

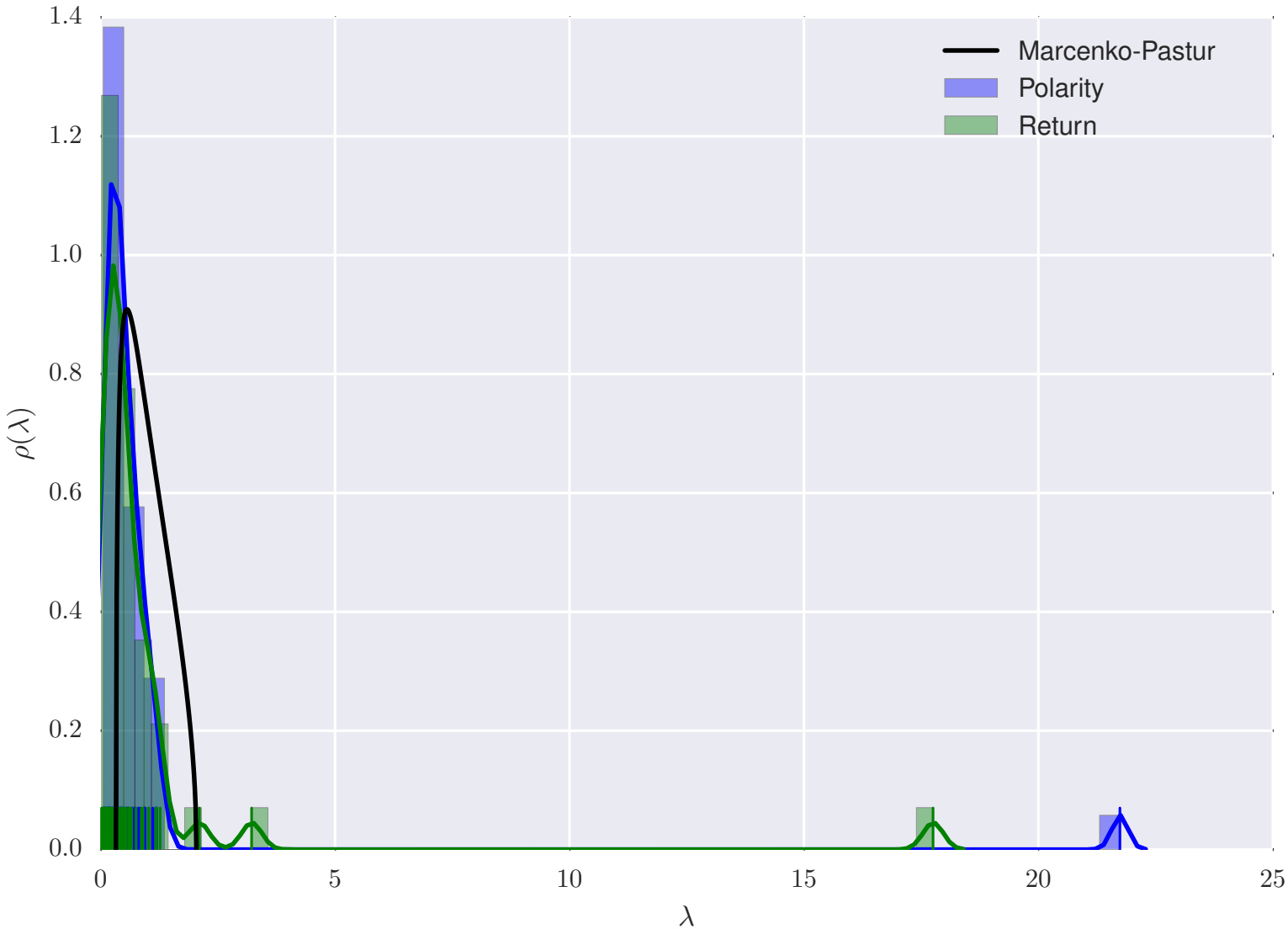
Diagrama


Keyword
API of NYT
fetching: URL identifier
Preprocess: NLTK & BASH SHEELL of LINUX
Sentiment: Valence Aware Dictionary and sEntiment Reasoner (VADER)
Average of all news for a given keyword (k) and day (t)
$P_k(t)$

índices

Country	Bloomberg Ticker	NYT Keyword
United States	SPX	United States
Canada	SPTSX	Canada
Mexico	MEXBOL	Mexico
Colombia	IGBC	Colombia
Venezuela	IBVC	Venezuela
Chile	IPSA	Chile
Argentina	MERVAL	Argentina
Brazil	IBOV	Brazil
Nigeria	NGSEINDX	Nigeria
United Kingdom	UKX	England
France	CAC	France
Belgium	BEL20	Belgium
Italy	FTSEMIB	Italy
Switzerland	SMI	Switzerland
Netherlands	AEX	Netherlands
Denmark	KFX	Denmark
Norway	OBX	Norway
Sweden	OMX	Sweden
Germany	DAX	Germany
Poland	WIG	Poland
⋮	⋮	⋮
Poland	WIG	Poland
Austria	ATX	Austria
Greece	ASE	Greece
Hungary	BUX	Hungary
Ukraine	PFTS	Ukraine
Russia	INDEXCM	Russia
Turkey	XU100	Turkey
Egypt	CASE	Egypt
Israel	TA-25	Israel
Arabia	SASEIDX	Arabia
Pakistan	KSE100	Pakistan
India	SENSEX	India
Indonesia	JCI	Indonesia
Malaysia	FBMKLCI	Malaysia
Singapore	FSSTI	Singapore
China	SHCOMP	China
Hong Kong	HSI	Hong Kong
Taiwan	TWSE	Taiwan
South Korea	KOSPI	South Korea
Japan	NKY	Japan
Australia	AS51	Australia

Resultados para NYT



Modelo para matrices no-simétricas[11]

[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)

Modelo para matrices no-simétricas[11]

- Sea un ensemble del tipo $\mathcal{C} = \mathcal{W}\mathcal{W}^t/T$, donde $\mathcal{W} = \xi^{1/2}W$, ξ es una matriz definida positiva, y $W_{ij} \sim N(0, \sigma^2)$ i.i.d. $\Rightarrow \bar{\mathcal{C}} = \xi$

[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)

Modelo para matrices no-simétricas[11]

- Sea un ensemble del tipo $\mathcal{C} = \mathcal{W}\mathcal{W}^t/T$, donde $\mathcal{W} = \xi^{1/2}W$, ξ es una matriz definida positiva, y $W_{ij} \sim N(0, \sigma^2)$ i.i.d. $\Rightarrow \bar{\mathcal{C}} = \xi$
- Suponga que la matriz $\mathcal{W}$ se construye mediante dos matrices aleatorias $\mathcal{A}$ and B

$$\mathcal{W} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \Rightarrow \mathcal{C} = \begin{pmatrix} \mathcal{A}\mathcal{A}^t & \mathcal{A}\mathcal{B}^t \\ \mathcal{B}\mathcal{A}^t & \mathcal{B}\mathcal{B}^t \end{pmatrix} \Rightarrow \xi = \begin{pmatrix} \xi_{AA} & \xi_{AB} \\ \xi_{BA} & \xi_{BB} \end{pmatrix} \quad (49)$$

[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)

Modelo para matrices no-simétricas[11]

- Sea un ensemble del tipo $\mathcal{C} = \mathcal{W}\mathcal{W}^t/T$, donde $\mathcal{W} = \xi^{1/2}W$, ξ es una matriz definida positiva, y $W_{ij} \sim N(0, \sigma^2)$ i.i.d. $\Rightarrow \bar{\mathcal{C}} = \xi$
- Suponga que la matriz $\mathcal{W}$ se construye mediante dos matrices aleatorias $\mathcal{A}$ and B

$$\mathcal{W} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \Rightarrow \mathcal{C} = \begin{pmatrix} \mathcal{A}\mathcal{A}^t & \mathcal{A}\mathcal{B}^t \\ \mathcal{B}\mathcal{A}^t & \mathcal{B}\mathcal{B}^t \end{pmatrix} \Rightarrow \xi = \begin{pmatrix} \xi_{AA} & \xi_{AB} \\ \xi_{BA} & \xi_{BB} \end{pmatrix} \quad (49)$$

- Se desea comparar la densidad espectral bajo la hipótesis nula: $\xi_{AA} = \mathbb{1}_{N\times N}, \xi_{BB} = \mathbb{1}_{M\times M}$, and $\xi_{AB} = \mathbb{0}$.

[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)

Modelo para matrices no-simétricas[11]

- Sea un ensemble del tipo $\mathcal{C} = \mathcal{W}\mathcal{W}^t/T$, donde $\mathcal{W} = \xi^{1/2}W$, ξ es una matriz definida positiva, y $W_{ij} \sim N(0, \sigma^2)$ i.i.d. $\Rightarrow \bar{\mathcal{C}} = \xi$
- Suponga que la matriz $\mathcal{W}$ se construye mediante dos matrices aleatorias $\mathcal{A}$ and B

$$\mathcal{W} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \Rightarrow \mathcal{C} = \begin{pmatrix} \mathcal{A}\mathcal{A}^t & \mathcal{A}\mathcal{B}^t \\ \mathcal{B}\mathcal{A}^t & \mathcal{B}\mathcal{B}^t \end{pmatrix} \Rightarrow \xi = \begin{pmatrix} \xi_{AA} & \xi_{AB} \\ \xi_{BA} & \xi_{BB} \end{pmatrix} \quad (49)$$

- Se desea comparar la densidad espectral bajo la hipótesis nula: $\xi_{AA} = \mathbb{1}_{N\times N}, \xi_{BB} = \mathbb{1}_{M\times M}$, and $\xi_{AB} = \mathbb{0}$.
- Para esto, se decorrelacionan las matrices individualmente:

$$A = \xi_{AA}^{-1/2}\mathcal{A}, \quad B = \xi_{BB}^{-1/2}\mathcal{B} \Rightarrow \eta = \overline{AB^t}/T = \quad \eta = \xi_{AA}^{-1/2}\xi_{AB}\xi_{BB}^{-1/2}$$

[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)

Modelo para matrices no-simétricas[11]

- Sea un ensemble del tipo $\mathcal{C} = \mathcal{W}\mathcal{W}^t/T$, donde $\mathcal{W} = \xi^{1/2}W$, ξ es una matriz definida positiva, y $W_{ij} \sim N(0, \sigma^2)$ i.i.d. $\Rightarrow \bar{\mathcal{C}} = \xi$
- Suponga que la matriz $\mathcal{W}$ se construye mediante dos matrices aleatorias $\mathcal{A}$ and B

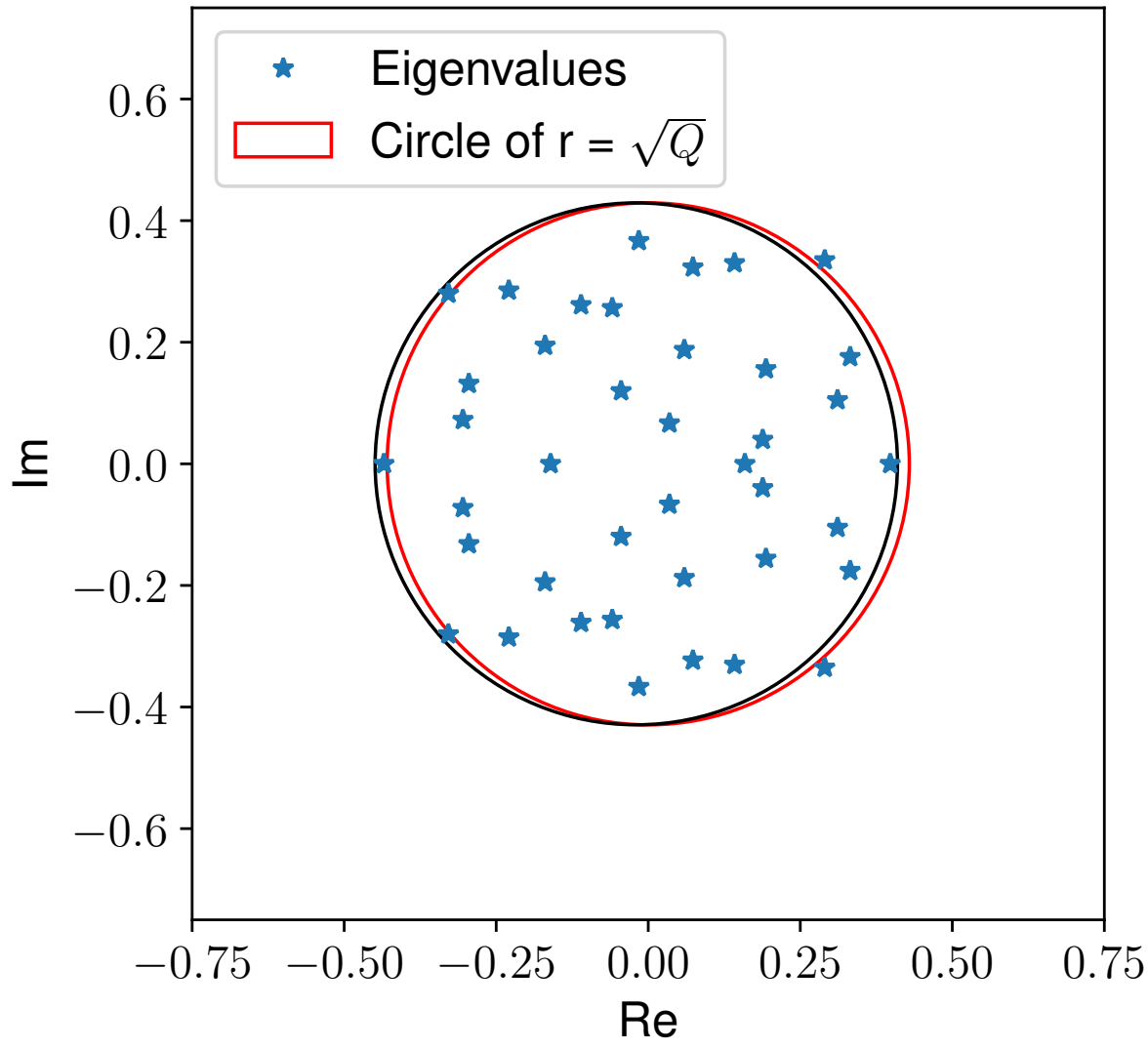
$$\mathcal{W} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \Rightarrow \mathcal{C} = \begin{pmatrix} \mathcal{A}\mathcal{A}^t & \mathcal{A}\mathcal{B}^t \\ \mathcal{B}\mathcal{A}^t & \mathcal{B}\mathcal{B}^t \end{pmatrix} \Rightarrow \xi = \begin{pmatrix} \xi_{AA} & \xi_{AB} \\ \xi_{BA} & \xi_{BB} \end{pmatrix} \quad (49)$$

- Se desea comparar la densidad espectral bajo la hipótesis nula: $\xi_{AA} = \mathbb{1}_{N\times N}, \xi_{BB} = \mathbb{1}_{M\times M}$, and $\xi_{AB} = \mathbb{0}$.
- Para esto, se decorrelacionan las matrices individualmente:

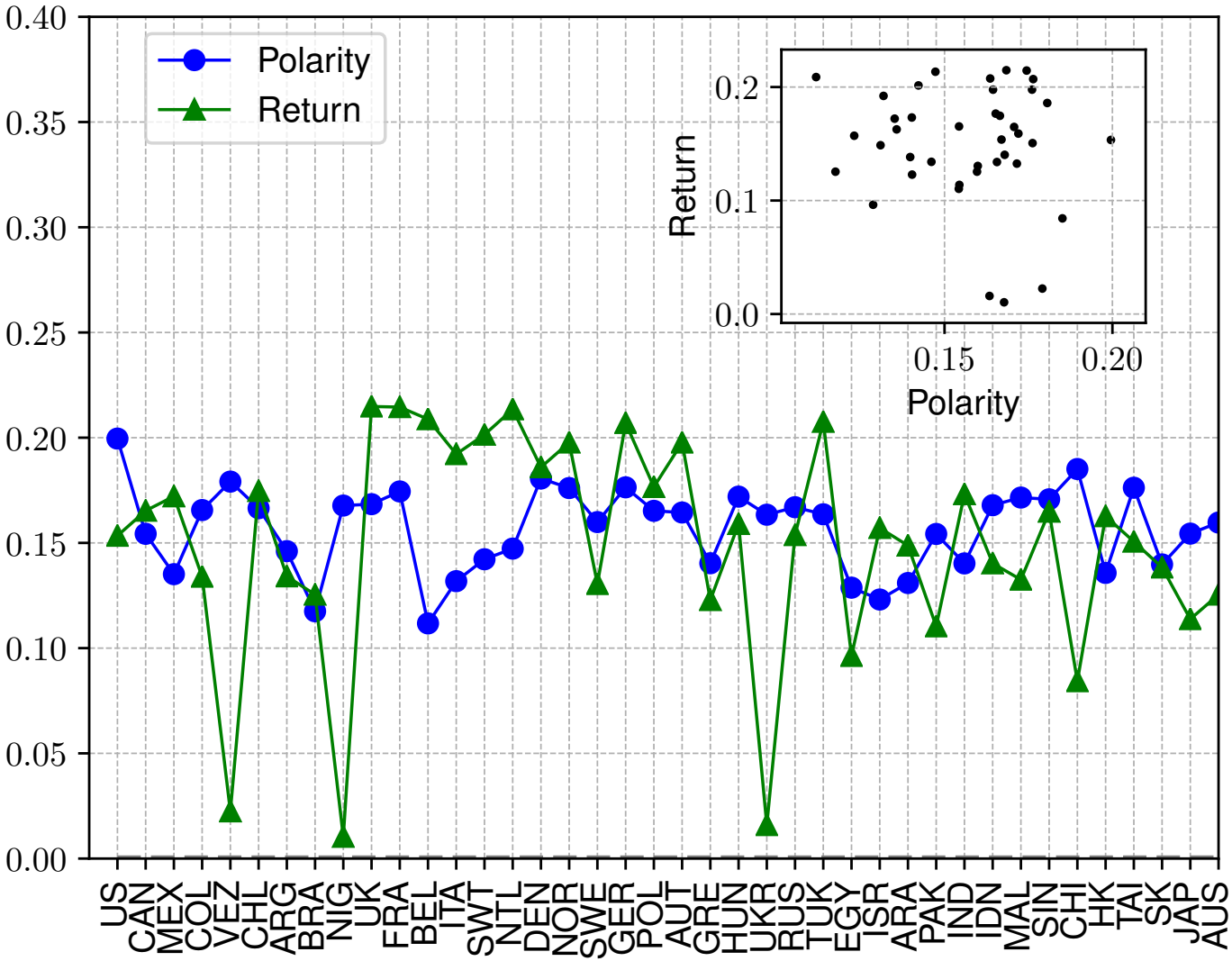
$$A = \xi_{AA}^{-1/2}\mathcal{A}, \quad B = \xi_{BB}^{-1/2}\mathcal{B} \Rightarrow \eta = \overline{AB^t}/T = \quad \eta = \xi_{AA}^{-1/2}\xi_{AB}\xi_{BB}^{-1/2}$$

- En el caso que η es diagonal con el elemento $\eta_{ii} = c, (i = 1, ..., N)$ (constante en promedio), el contorno del dominio del espectro esta dado por una elipse desplazada: $\frac{[x-c(1+Q)]^2}{Q(1+c^2)^2} + \frac{y^2}{Q(1-c^2)^2} = 1$, donde $Q = T/N$.

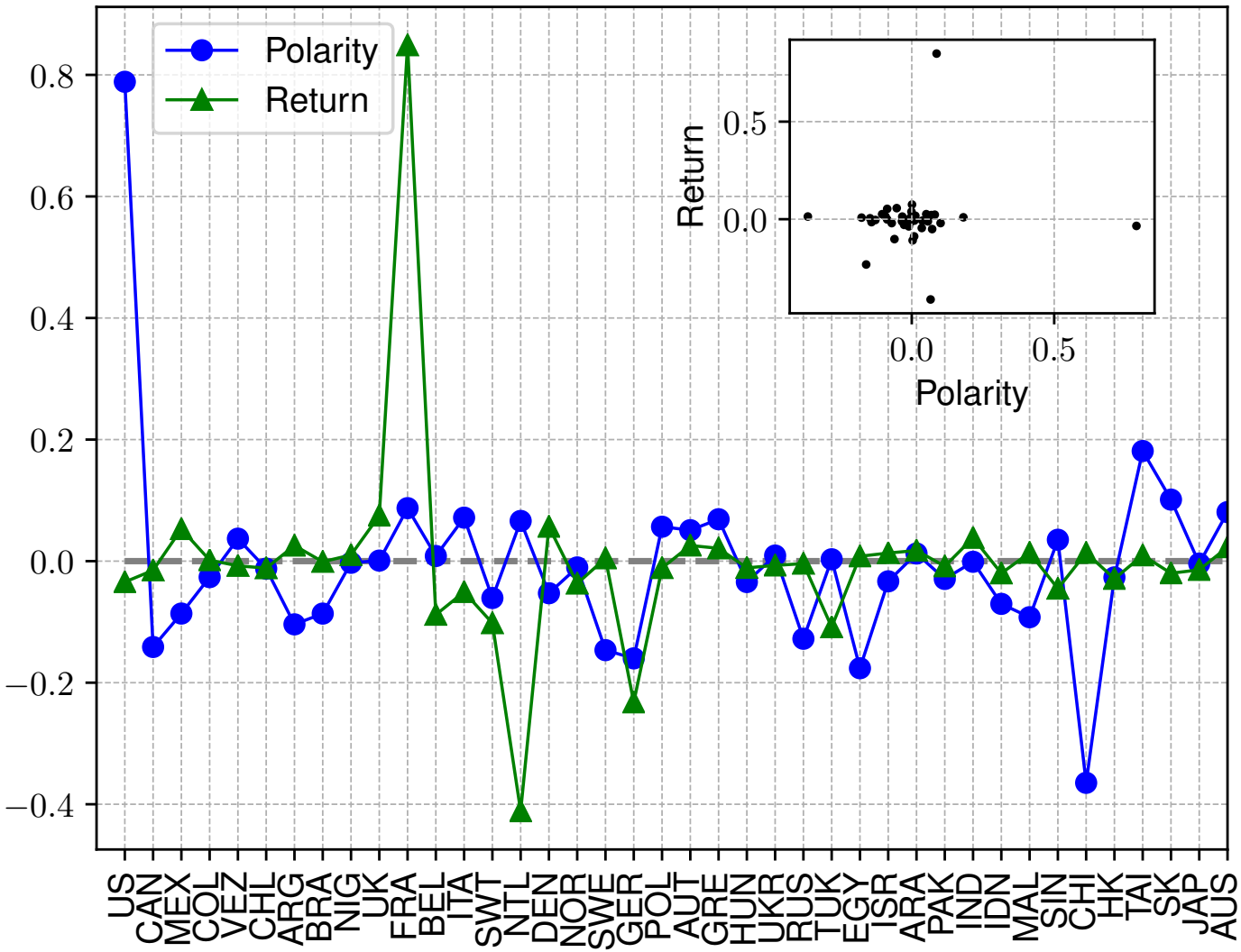
[11] Vinayak and L. Benet, Phys. Rev. E 90, 042109 (2014)


Eigenvalues
Circle of r = $\sqrt{Q}$
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
Im
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
Re

Vector propio asociado a λ_1

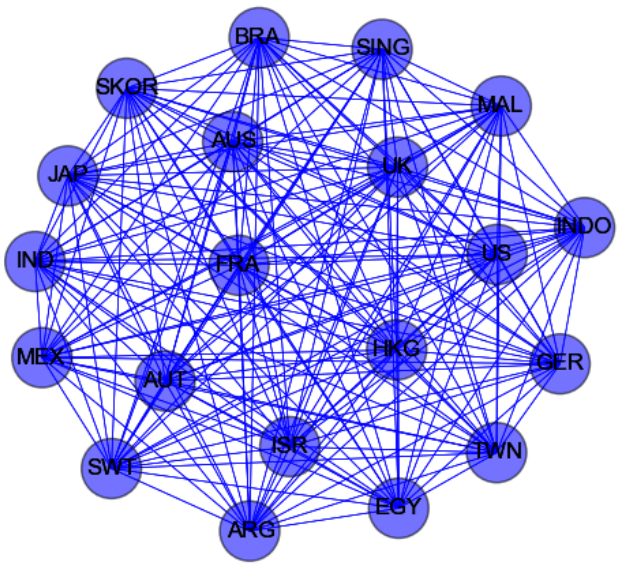


Vector propio asociado a λ_{40}



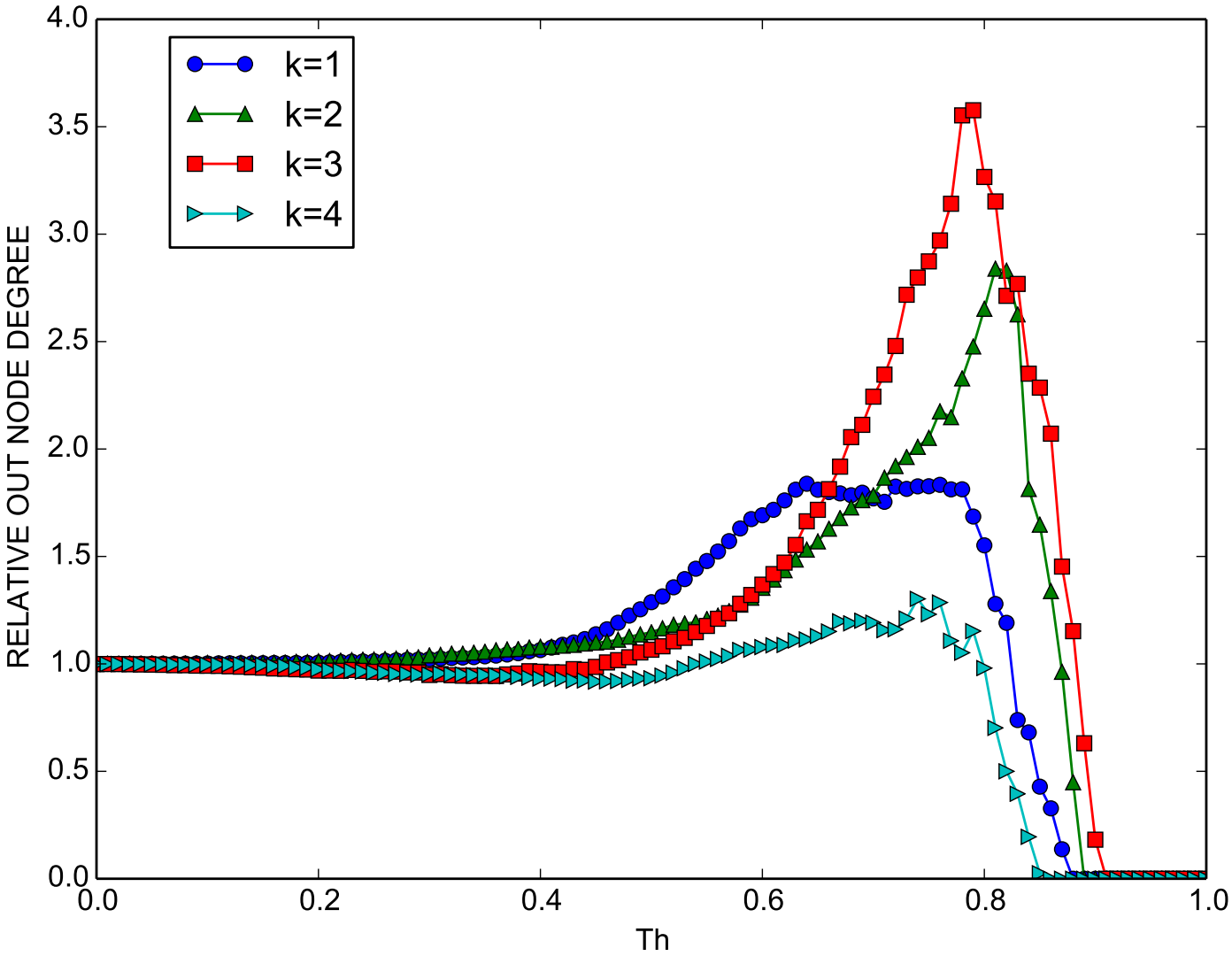
Transferencia de Entropía como red dirigida

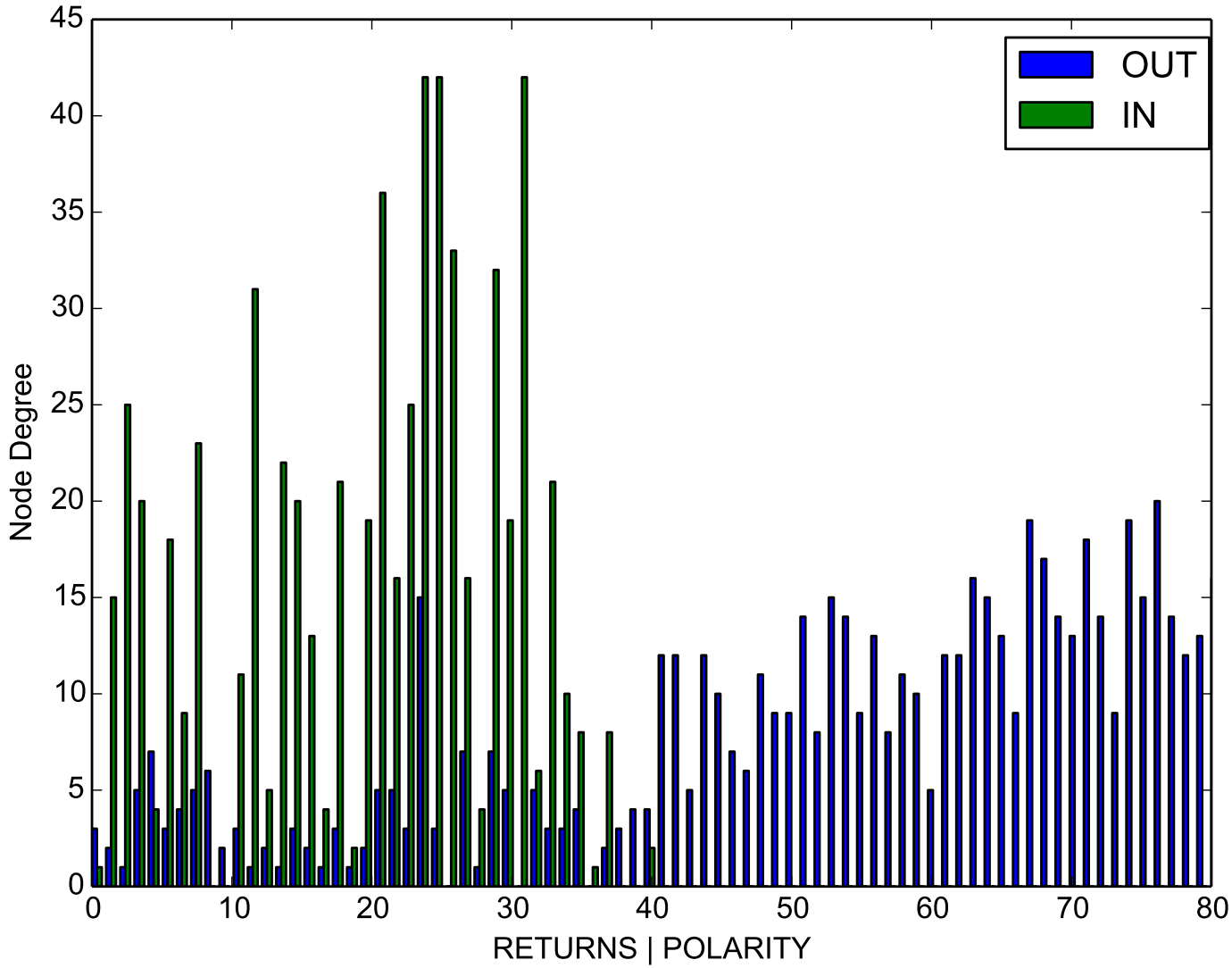
Un grafo es un par ordenado $G(V, E)$, donde V representa nodos, y E aristas.



Se definió *relative out node degree* como la razón de las aristas que salen de los nodos de polaridad respecto a los de retorno:

$$ND_{out}(polarity)/ND_{out}(returns)$$


k=1
k=2
k=3
k=4
RELATIVE OUT NODE DEGREE
Th
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0


OUT
IN
Node Degree
RETURNS | POLARITY
0
5
10
15
20
25
30
35
40
45
0
10
20
30
40
50
60
70
80

Referencias

1. R.N. Mantenga, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, (2000).
2. J.P. Bouchaud, M. Potters, Theory of Financial Risks: from Statistical Physics to Risk Management, Cambridge University Press, Cambridge, (2004).
3. J. Voit, The Statistical Mechanics of Financial Markets, Springer-Verlag, Berlin, (2005).
4. Introduction to Random Matrices: Theory and Practice by Giacomo Livan Marcel Novaes y Pierpaolo Vivo, Springer (2018).
5. An Introduction to Transfer Entropy: Information Flow in Complex Systems by Terry Bossomaier, Lionel Barnett, Michael Harré and Joseph T. Lizier. Springer, (2016).
6. Global financial indices and twitter sentiment: a random matrix theory approach. A. García. Physica A. 461 (2016) 509.
7. Correlations and flow of information between The New York Times and Stock Markets. Andrés García-Medina, Leonidas Sandoval Junior, Efraín Urrutia Bañuelos y A. M. Martínez-Argüello. Physica A. 502 (2018) 403.

Gracias