

Test de esfericidad parcial en alta dimensión

¿Cuántas componentes importan?

Antonella Gieco

Facultad de Ingeniería Química, UNL

Liliana, Forzani; Carlos, Tolmasky

FIQ

UNL

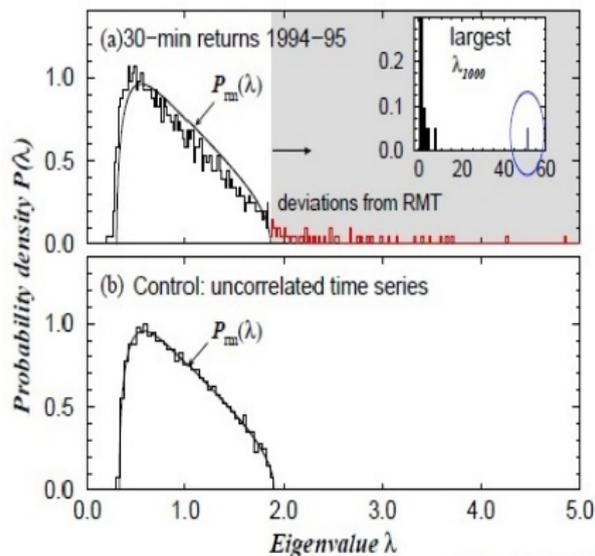
- 1 Matrices de covarianza con sólo unos pocos autovalores importantes.

Motivación

- 1 Matrices de covarianza con sólo unos pocos autovalores importantes.
- 2 El resto es ruido.

Motivación

- 1 Matrices de covarianza con sólo unos pocos autovalores importantes.
 - 2 El resto es ruido.
- Ejemplo: stock market.
 - El primer (mayor) autovalor es mucho más grande que el resto.
 - Después de unos pocos primeros, el resto de ellos son pequeños.
 - Plerou et al (2002).



Plerou et al (2002).

Modelos de covarianza Spiked (Johnstone, 2001)

- Este es un fenómeno común (no solo en finanzas).
- Uno o unos pocos ($d \ll p$) autovalores grandes y bien separados del resto ($\lambda_1 \geq \dots \lambda_d \gg \lambda_{d+1} = \dots = \lambda_p = \sigma^2$).

$$\begin{pmatrix} \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_d & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

- Estadística multivariada “clásica”:
 - Una cantidad fija variables (p) y muchas observaciones (n).
 - Dadas p variables, se pueden tener tantas observaciones como se quieran.
 - Resultados límites para $n \rightarrow \infty$.
- Estadística multivariada en “Alta-dimensión”:
 - el número de variables (p) es comparable el número de observaciones (n).
 - Incrementar en el número de observaciones incrementa el número de variables.
 - Resultados límites cuando $p/n \rightarrow y$ (una constante fija).

Algunos resultados en alta dimensión

Distribución de los autovalores

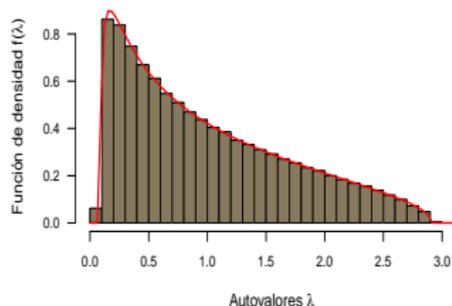
- p variables independientes con varianza σ^2 , n observaciones (p y n grandes). Si estimamos los autovalores de la matriz de covarianzas, no nos da cerca de $(\sigma^2, \sigma^2, \dots, \sigma^2)$.

Algunos resultados en alta dimensión

Distribución de los autovalores

- p variables independientes con varianza σ^2 , n observaciones (p y n grandes). Si estimamos los autovalores de la matriz de covarianzas, no nos da cerca de $(\sigma^2, \sigma^2, \dots, \sigma^2)$.

Distribución Marchenko–Pastur, $p/n=0.5$, $\sigma^2=1$

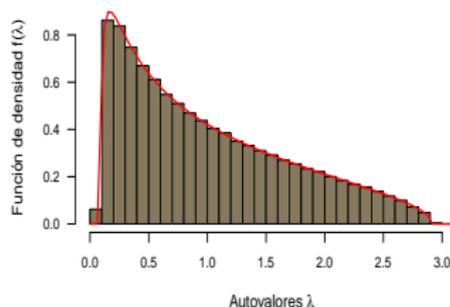


Algunos resultados en alta dimensión

Distribución de los autovalores

- p variables independientes con varianza σ^2 , n observaciones (p y n grandes). Si estimamos los autovalores de la matriz de covarianzas, no nos da cerca de $(\sigma^2, \sigma^2, \dots, \sigma^2)$.

Distribución Marchenko–Pastur, $p/n=0.5$, $\sigma^2=1$



La distribución límite (línea roja) es la M-P(y, σ^2).

El soporte es $\sigma^2 (1 \pm \sqrt{y})^2$ ($\approx (0.09 ; 2.9)$).

- Más resultados para estadística HD para modelos spiked.
 - Los estimadores de los λ 's en el modelo de Johnstone tienen dos estados (Baik, J., G. B. Arous, and S. Péché (2005)):
 - $\hat{\lambda} \rightarrow \lambda (1 + \frac{y\sigma^2}{\lambda - \sigma^2})$ if $\lambda > \sigma^2 (1 + \sqrt{y})$.
 - $\hat{\lambda} \rightarrow \sigma^2 (1 + \sqrt{y})^2$ if $\lambda \in (\sigma^2, \sigma^2 (1 + \sqrt{y})]$.
 - Por lo tanto, si $\sigma^2 < \lambda \leq \sigma^2 (1 + \sqrt{y})$ el estimador de λ no se separa del “montón”.
 - Esos λ 's son, en consecuencia, difíciles (o imposible) de detectar.

¿Cómo estimar la dimensión que importa?

Uno puede (Plerou et al, Bouchaud etc.):

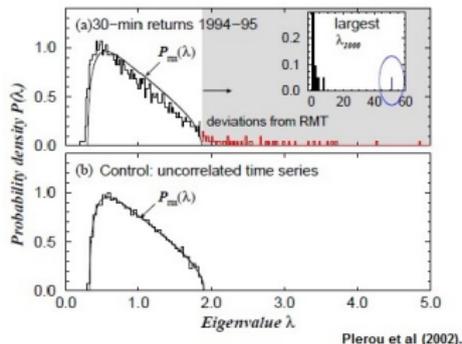
- Sacar el autovalor más grande y ver si el resto de los autovalores se ajustan a la distribución MP (gráfico inferior).
- Si esto sucede, la dimensión es igual a 1.
- En otro caso, sacamos el segundo autovalor más grande y repetimos.

¿Cómo estimar la dimensión que importa?

Uno puede (Plerou et al, Bouchaud etc.):

- Sacar el autovalor más grande y ver si el resto de los autovalores se ajustan a la distribución MP (gráfico inferior).
- Si esto sucede, la dimensión es igual a 1.
- En otro caso, sacamos el segundo autovalor más grande y repetimos.

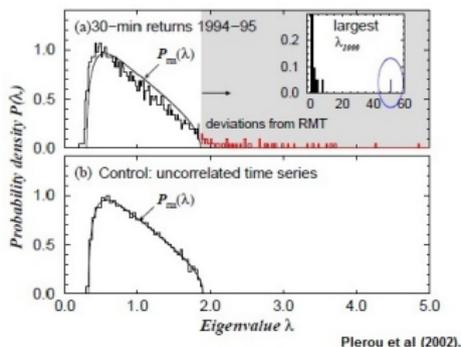
Observemos nuevamente este gráfico:



¿Cómo estimar la dimensión que importa?

Otro enfoque (Passemier, D. y Yao, J.F. (2012)):

- Separación entre los autovalores es menor en la parte correspondiente al ruido.
- La parte más estrecha corresponde a ruido, la más dispersa es la información.
- Chequear cuando el espaciado pasa de “ancho” a “estrecho”.



Hallar un test de cociente de máxima verosimilitud para determinar cuántas componentes importan.

Test de cociente de máxima verosimilitud

Trabajos previos:

$$H_0 : d = 0 \quad \text{vs} \quad H_a : d > 0$$

Para $p < n$, el LRT es:

$$\text{LRT} = \frac{\hat{\lambda}_1 \dots \hat{\lambda}_p}{\left(\frac{1}{p} \sum_{i=1}^p \hat{\lambda}_i\right)^p}.$$

Test de cociente de máxima verosimilitud

Trabajos previos:

$$H_0 : d = 0 \quad \text{vs} \quad H_a : d > 0$$

Para $p < n$, el LRT es:

$$LRT = \frac{\hat{\lambda}_1 \dots \hat{\lambda}_p}{\left(\frac{1}{p} \sum_{i=1}^p \hat{\lambda}_i\right)^p}.$$

- Un resultado asintótico clásico (Muirhead, 1982) para p fijo:

$$-(n-1)\rho \log LRT \xrightarrow{d} \chi^2(f) \quad \text{cuando } n \rightarrow \infty$$

El factor $\rho = \rho_n \rightarrow 1$ es un término de corrección para mejorar la tasa de convergencia.

Test de cociente de máxima verosimilitud

Trabajos previos:

$$H_0 : d = 0 \quad \text{vs} \quad H_a : d > 0$$

Para $p < n$, el LRT es:

$$LRT = \frac{\hat{\lambda}_1 \dots \hat{\lambda}_p}{\left(\frac{1}{p} \sum_{i=1}^p \hat{\lambda}_i\right)^p}.$$

- Para el caso $p \rightarrow \infty$ con $p/n \rightarrow y \leq 1$ Jiang and Yang (2013) prueban que

$$\frac{\log(LRT) - \mu^*}{\sigma^*} \rightarrow N(0, 1)$$

y a partir de la distribución encuentran un test asintótico.

Test de cociente de máxima verosimilitud

H_d : dimensión = d vs H_a : dimensión $> d$

Para $p < n$, el LRT_d es:

$$LRT_d = \frac{\hat{\lambda}_{d+1} \cdots \hat{\lambda}_p}{\left(\frac{1}{p-d} \sum_{i=d+1}^p \hat{\lambda}_i \right)^{p-d}}.$$

Test de cociente de máxima verosimilitud

H_d : dimensión = d vs H_a : dimensión $> d$

Para $p < n$, el LRT_d es:

$$LRT_d = \frac{\hat{\lambda}_{d+1} \cdots \hat{\lambda}_p}{\left(\frac{1}{p-d} \sum_{i=d+1}^p \hat{\lambda}_i\right)^{p-d}}.$$

- De nuevo, si p es fijo y $n \rightarrow \infty$, bajo H_d

$$-\rho \log LRT_d \rightarrow \chi_{(p-d+2)(p-d-1)/2}^2,$$

(Lawley (1956), James (1969))

Test de cociente de máxima verosimilitud

H_d : dimensión = d vs H_a : dimensión $> d$

Para $p < n$, el LRT_d es:

$$LRT_d = \frac{\hat{\lambda}_{d+1} \cdots \hat{\lambda}_p}{\left(\frac{1}{p-d} \sum_{i=d+1}^p \hat{\lambda}_i\right)^{p-d}}.$$

- En nuestro trabajo consideramos el caso $p \rightarrow \infty$ con $p/n \rightarrow y < 1$.
- Probamos que

$$\frac{\log(LRT_d) - \mu_{p,n,d}}{\sigma_{p,n,d}} \rightarrow N(0, 1)$$

- Test secuencial para estimar d .

¿Que hacemos si tenemos más variables que observaciones?

¿Que hacemos si tenemos más variables que observaciones?

LRT no existe ni siquiera para el caso $\sigma^2 I$ (hay muchos autovalores nulos).

¿Que hacemos si tenemos más variables que observaciones?

LRT no existe ni siquiera para el caso $\sigma^2 I$ (hay muchos autovalores nulos).

Sin embargo podemos:

¿Que hacemos si tenemos más variables que observaciones?

LRT no existe ni siquiera para el caso $\sigma^2 I$ (hay muchos autovalores nulos).

Sin embargo podemos:

- Invertir los roles de p y n .
- Definimos

$$\text{LRT}_0^* = \frac{\hat{\lambda}_1 \dots \hat{\lambda}_n}{\left(\frac{1}{n} \sum_{i=1}^n \hat{\lambda}_i\right)^n}.$$

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.
- Sea $\widehat{\Sigma}$ la covarianza muestral. Entonces, bajo H_0 ,
 $\mathbf{W} = m\widehat{\Sigma} = \mathbf{Y}^T \mathbf{Y} \sim W_p(m, \Sigma)$ donde
 $\mathbf{Y} \in \mathbb{R}^{m \times p} \sim N(0, \mathbf{I}_m \otimes \Sigma)$ (Muirhead, 1982).

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.
- Sea $\widehat{\Sigma}$ la covarianza muestral. Entonces, bajo H_0 ,
 $\mathbf{W} = m\widehat{\Sigma} = \mathbf{Y}^T\mathbf{Y} \sim W_p(m, \Sigma)$ donde
 $\mathbf{Y} \in \mathbb{R}^{m \times p} \sim N(0, \mathbf{I}_m \otimes \Sigma)$ (Muirhead, 1982).
- Definimos $\widetilde{\mathbf{W}} = \mathbf{Y}\mathbf{Y}^T \sim W_m(p, \sigma^2\mathbf{I}_m)$ con $p > m$.

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.
- Sea $\widehat{\Sigma}$ la covarianza muestral. Entonces, bajo H_0 ,
 $\mathbf{W} = m\widehat{\Sigma} = \mathbf{Y}^T\mathbf{Y} \sim W_p(m, \Sigma)$ donde
 $\mathbf{Y} \in \mathbb{R}^{m \times p} \sim N(0, \mathbf{I}_m \otimes \Sigma)$ (Muirhead, 1982).
- Definimos $\widetilde{\mathbf{W}} = \mathbf{Y}\mathbf{Y}^T \sim W_m(p, \sigma^2\mathbf{I}_m)$ con $p > m$.
- Los autovalores no nulos de \mathbf{W} coinciden con los de $\widetilde{\mathbf{W}}$.

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.
- Sea $\widehat{\Sigma}$ la covarianza muestral. Entonces, bajo H_0 ,
 $\mathbf{W} = m\widehat{\Sigma} = \mathbf{Y}^T\mathbf{Y} \sim W_p(m, \Sigma)$ donde
 $\mathbf{Y} \in \mathbb{R}^{m \times p} \sim N(0, \mathbf{I}_m \otimes \Sigma)$ (Muirhead, 1982).
- Definimos $\widetilde{\mathbf{W}} = \mathbf{Y}\mathbf{Y}^T \sim W_m(p, \sigma^2\mathbf{I}_m)$ con $p > m$.
- Los autovalores no nulos de \mathbf{W} coinciden con los de $\widetilde{\mathbf{W}}$.
- Por tanto, LRT_0^* es el cociente de verosimilitud para $\widetilde{\mathbf{W}}$

¿Qué significa invertir los roles de p y n ?

- Sin pérdida de generalidad supongamos que la covarianza poblacional es Σ es diagonal.
- Sea $\widehat{\Sigma}$ la covarianza muestral. Entonces, bajo H_0 ,
 $\mathbf{W} = m\widehat{\Sigma} = \mathbf{Y}^T \mathbf{Y} \sim W_p(m, \Sigma)$ donde
 $\mathbf{Y} \in \mathbb{R}^{m \times p} \sim N(0, \mathbf{I}_m \otimes \Sigma)$ (Muirhead, 1982).
- Definimos $\widetilde{\mathbf{W}} = \mathbf{Y}\mathbf{Y}^T \sim W_m(p, \sigma^2 \mathbf{I}_m)$ con $p > m$.
- Los autovalores no nulos de \mathbf{W} coinciden con los de $\widetilde{\mathbf{W}}$.
- Por tanto, LRT_0^* es el cociente de verosimilitud para $\widetilde{\mathbf{W}}$
- La distribución de LRT_0^* se sigue de Jiang and Yang (2013), invirtiendo los roles de p y m .

Esto nos motiva a definir LRT_d^* para esfericidad parcial, en el caso $p > n$ como:

$$LRT_d^* = \frac{\hat{\lambda}_{d+1} \cdots \hat{\lambda}_n}{\left(\frac{1}{n-d} \sum_{i=d+1}^n \hat{\lambda}_i\right)^{n-d}}.$$

- Para obtener un test de nivel asintótico α , necesitamos su distribución bajo H_d .
- Pero $\widetilde{\mathbf{W}}$ ya no es más una Wishart, sin embargo,

$$\widetilde{\mathbf{W}} = \sum_{i=1}^d \lambda_i \mathbf{Z}_{(i)} \mathbf{Z}_{(i)}^T + \sigma^2 \widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^T$$

donde $\widetilde{\mathbf{Z}} \widetilde{\mathbf{Z}}^T \sim W_m(p-d, \mathbf{I}_m)$ independiente de $\mathbf{Z}_d = (\mathbf{Z}_{(1)}, \mathbf{Z}_{(2)}, \dots, \mathbf{Z}_{(d)}) \sim N(0, \mathbf{I}_m \otimes \mathbf{I}_d)$

Para el caso $p < n$:

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.
- 2 Se prueba que $\log \mathbb{E} (LRT_d^t) \sim a_{p,n}t + b_{p,n} \frac{t^2}{2} + o(1)$

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.
- 2 Se prueba que $\log \mathbb{E} (LRT_d^t) \sim a_{p,n}t + b_{p,n} \frac{t^2}{2} + o(1)$
- 3 Para p fijo, $a_{p,n}$ y $b_{p,n}$ fácil.

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.
- 2 Se prueba que $\log \mathbb{E} (LRT_d^t) \sim a_{p,n}t + b_{p,n} \frac{t^2}{2} + o(1)$
- 3 Para p fijo, $a_{p,n}$ y $b_{p,n}$ fácil.
- 4 Para $p, n \rightarrow \infty$ necesitamos acotaciones de la función $\Gamma_p(\cdot)$

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.
- 2 Se prueba que $\log \mathbb{E} (LRT_d^t) \sim a_{p,n}t + b_{p,n} \frac{t^2}{2} + o(1)$
- 3 Para p fijo, $a_{p,n}$ y $b_{p,n}$ fácil.
- 4 Para $p, n \rightarrow \infty$ necesitamos acotaciones de la función $\Gamma_p(\cdot)$
- 5 Luego $\frac{\log LRT_d^{-a_{n,p}}}{b_{n,p}} \rightarrow N(0, 1)$

Para el caso $p < n$:

- 1 Se estudian los momentos de LRT_d^t , t en un entorno de 0.
- 2 Se prueba que $\log \mathbb{E} (LRT_d^t) \sim a_{p,n}t + b_{p,n} \frac{t^2}{2} + o(1)$
- 3 Para p fijo, $a_{p,n}$ y $b_{p,n}$ fácil.
- 4 Para $p, n \rightarrow \infty$ necesitamos acotaciones de la función $\Gamma_p(\cdot)$
- 5 Luego $\frac{\log LRT_d^{-a_{n,p}}}{b_{n,p}} \rightarrow N(0, 1)$
- 6 En ambos caso, se usa fuertemente la densidad Wishart (cálculo de esperanza).

El caso $p > n$ (y $d > 0$)

El caso $p > n$ (y $d > 0$)

1 \widetilde{W} ya no es Wishart.

El caso $p > n$ (y $d > 0$)

1 \widetilde{W} ya no es Wishart.

2 Sin embargo $LRT_d^* = \frac{|\widetilde{\Sigma}|}{\left(\frac{1}{m-d} \sum_{i=d+1}^m \hat{\lambda}_i\right)^{m-d}} \frac{1}{\hat{\lambda}_1 \cdots \hat{\lambda}_d}$

El caso $p > n$ (y $d > 0$)

1 \widetilde{W} ya no es Wishart.

2 Sin embargo $LRT_d^* = \frac{|\widetilde{\Sigma}|}{\left(\frac{1}{m-d} \sum_{i=d+1}^m \hat{\lambda}_i\right)^{m-d}} \frac{1}{\hat{\lambda}_1 \cdots \hat{\lambda}_d}$

3 $|\widetilde{\Sigma}| = \left| \frac{\sigma^2}{m} \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^T \right| \left| \mathbf{I}_d + \frac{1}{\sigma^2} \mathbf{\Lambda}_d^{1/2} \mathbf{Z}_d^T (\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^T)^{-1} \mathbf{Z}_d \mathbf{\Lambda}_d^{1/2} \right|$

- 1 A partir de la distribución asintótica podemos construir un test que tenga nivel (asintótico) igual a α considerando la región de rechazo $\{LRT_d < c_\alpha\}$ donde $c_\alpha = \exp(C_\alpha)$ con $C_\alpha = \mu_{m,p,d} + z_\alpha \sigma_{m,p,d}$ y z_α es el cuantil α de la normal estándar.

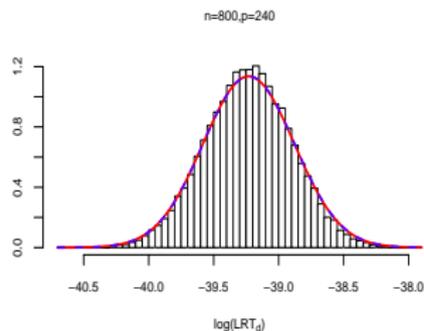
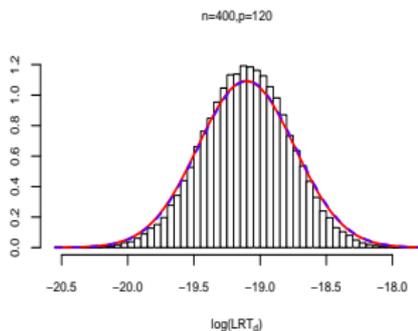
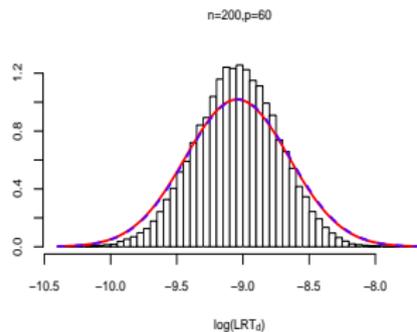
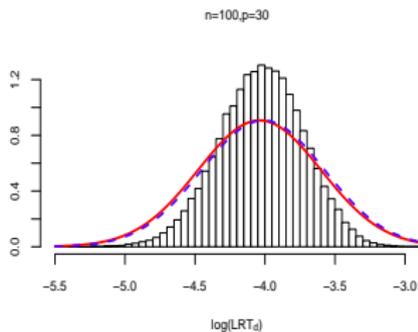
- 1 A partir de la distribución asintótica podemos construir un test que tenga nivel (asintótico) igual a α considerando la región de rechazo $\{LRT_d < c_\alpha\}$ donde $c_\alpha = \exp(C_\alpha)$ con $C_\alpha = \mu_{m,p,d} + z_\alpha \sigma_{m,p,d}$ y z_α es el cuantil α de la normal estándar.
- 2 Pero C_α depende de los verdaderos parámetros y, por lo tanto, debemos reemplazarlos por funciones de los valores muestrales.

- 1 A partir de la distribución asintótica podemos construir un test que tenga nivel (asintótico) igual a α considerando la región de rechazo $\{LRT_d < c_\alpha\}$ donde $c_\alpha = \exp(C_\alpha)$ con $C_\alpha = \mu_{m,p,d} + z_\alpha \sigma_{m,p,d}$ y z_α es el cuantil α de la normal estándar.
- 2 Pero C_α depende de los verdaderos parámetros y, por lo tanto, debemos reemplazarlos por funciones de los valores muestrales.
- 3 σ^2 se puede reemplazar por su estimador consistente $\hat{\sigma}^2 = \sum_{i=d+1}^p \hat{\lambda}_i / (p - d)$.

- 1 A partir de la distribución asintótica podemos construir un test que tenga nivel (asintótico) igual a α considerando la región de rechazo $\{LRT_d < c_\alpha\}$ donde $c_\alpha = \exp(C_\alpha)$ con $C_\alpha = \mu_{m,p,d} + z_\alpha \sigma_{m,p,d}$ y z_α es el cuantil α de la normal estándar.
- 2 Pero C_α depende de los verdaderos parámetros y, por lo tanto, debemos reemplazarlos por funciones de los valores muestrales.
- 3 σ^2 se puede reemplazar por su estimador consistente $\hat{\sigma}^2 = \sum_{i=d+1}^p \hat{\lambda}_i / (p - d)$.
- 4 Con los λ_i , debemos considerar la transición de fase y su sesgo.

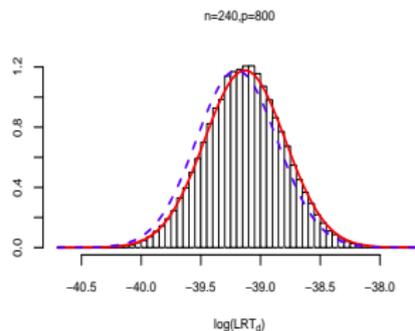
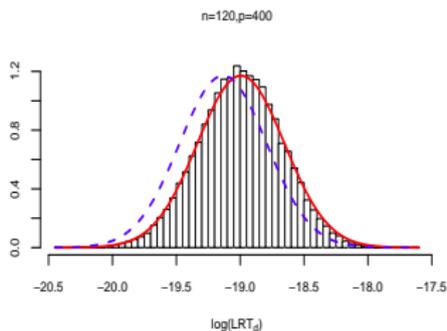
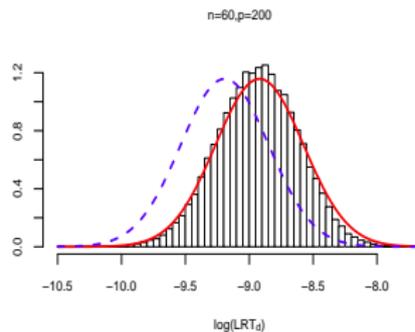
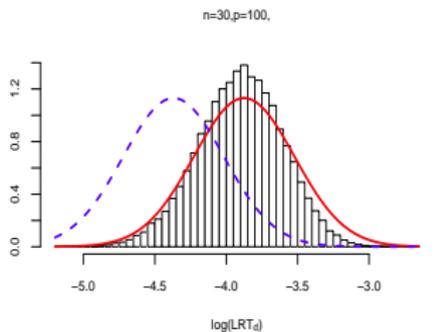
Simulación: Distribución asintótica

$$d = 4, \lambda = [7, 6, 5, 4], \sigma^2 = 1, p/n = 0.3.$$



Simulación: Distribución asintótica

$$d = 4, \lambda = [7, 6, 5, 4], \sigma^2 = 1, p/n = 1/0.3.$$



Simulación: detección de la dimensión d

(p,n)	1	2	3	4	5	6	7
(30,100)	0.000	0.000	0.039	0.901	0.036	0.014	0.002
(60,200)	0.000	0.000	0.014	0.932	0.041	0.009	0.004
(120,400)	0.000	0.000	0.006	0.935	0.042	0.011	0.004
(240,800)	0.000	0.000	0.006	0.949	0.035	0.009	0.001

Table: Valores de d estimado para spikes = [7, 6, 5, 4] (verdadera dimensión = 4) y $\frac{p}{m} = .3$

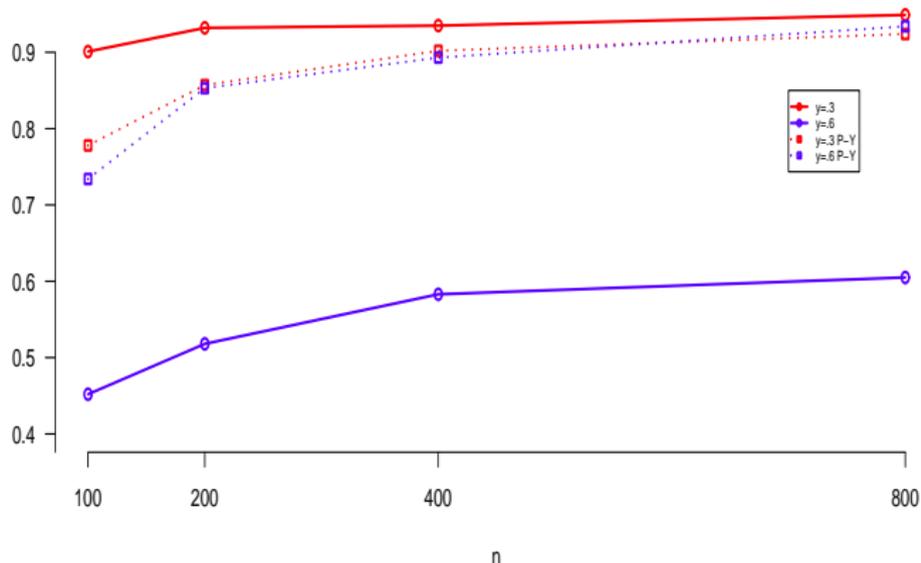
Simulación: detección de la dimensión d

(p,n)	1	2	3	4	5	6	7
(60,100)	0.000	0.018	0.482	0.452	0.028	0.011	0.004
(120,200)	0.000	0.011	0.427	0.518	0.028	0.008	0.006
(240,400)	0.000	0.002	0.368	0.583	0.027	0.012	0.007
(480,800)	0.000	0.002	0.352	0.605	0.024	0.013	0.003

Table: Valores de d estimado para spikes = [7, 6, 5, 4] (verdadera dimensión= 4) y $\frac{p}{m} = .6$

Simulación: Estimación de la dimensión

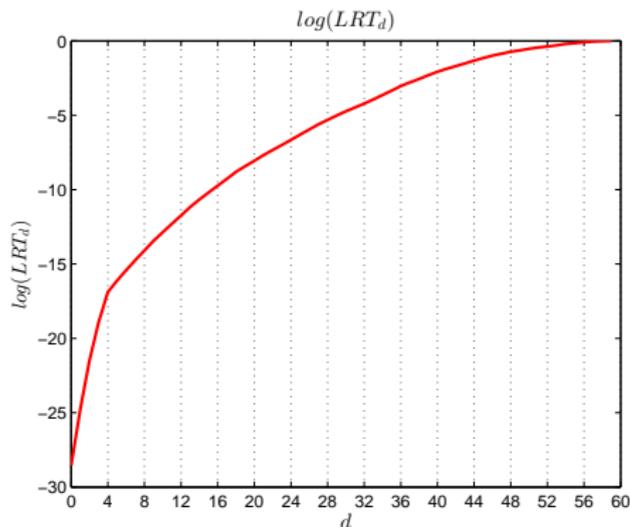
Frecuencia de estimación correcta



Hacia la versión penalizada

Consideremos nuevamente $d = 4$, $\lambda = [7, 6, 5, 4]$, $\sigma^2 = 1$

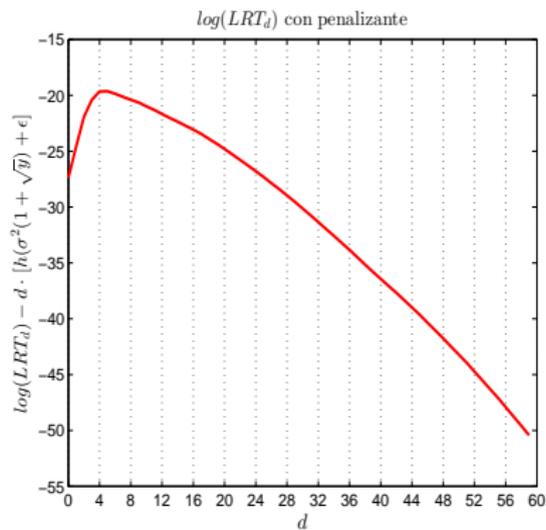
Graficamos la función $\log(LRT_d)$:



Estudio de los incrementos:

- Después de $d = 4$, (en el límite cuando $p, n \rightarrow \infty$) sólo dependen de p/n .
- Antes de $d = 4$, dependen de las estimaciones $\hat{\lambda}_i$.
- Es difícil detectar el quiebre.
- La penalización tiene en cuenta este cambio.
- Idea para penalizar: en cada paso restar el menor incremento.

Hacia la versión penalizada



Simulaciones: Estimación de d (penalizado)

(p,n)	1	2	3	4	5	6	7
(30,100)	0.000	0.000	0.000	0.992	0.008	0.000	0.000
(60,200)	0.000	0.000	0.000	1.000	0.000	0.000	0.000
(120,400)	0.000	0.000	0.000	1.000	0.000	0.000	0.000
(240,800)	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Table: Valores de d estimado via HD-MDL para spikes = [7, 6, 5, 4]
(verdadera dimensión = 4) y $\frac{p}{m} = .3$

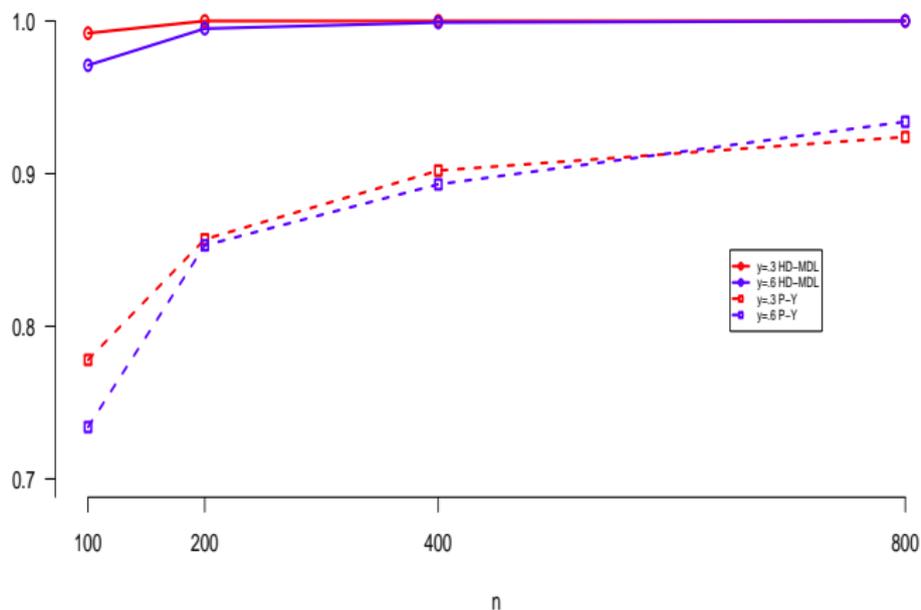
Simulaciones: Estimación de d (penalizado)

(p,n)	1	2	3	4	5	6	7
(60,100)	0.000	0.000	0.002	0.971	0.027	0.000	0.000
(120,200)	0.000	0.000	0.000	0.995	0.005	0.000	0.000
(240,400)	0.000	0.000	0.000	0.999	0.001	0.000	0.000
(480,800)	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Table: Valores de d estimado via HD-MDL para spikes = [7, 6, 5, 4]
(verdadera dimensión= 4) y $\frac{p}{m} = .6$

Simulación: Estimación de la dimensión

Frecuencia de estimación correcta (penalizado)



Conclusión

- Para el caso $p < n$ and $p, n \rightarrow \infty$, estudiamos la distribución asintótica del LRT_d para esfericidad parcial para el caso de modelos de covarianza spiked introducidos por Johnstone (2001).

Conclusión

- Para el caso $p < n$ and $p, n \rightarrow \infty$, estudiamos la distribución asintótica del LRT_d para esfericidad parcial para el caso de modelos de covarianza spiked introducidos por Johnstone (2001).
- Para $p > n$ and $p, n \rightarrow \infty$ obtenemos también la distribución asintótica del estadístico LRT_d invirtiendo los roles de p y n .

Conclusión

- Para el caso $p < n$ and $p, n \rightarrow \infty$, estudiamos la distribución asintótica del LRT_d para esfericidad parcial para el caso de modelos de covarianza spiked introducidos por Johnstone (2001).
- Para $p > n$ and $p, n \rightarrow \infty$ obtenemos también la distribución asintótica del estadístico LRT_d invirtiendo los roles de p y n .
- Conocer estas distribuciones nos permite desarrollar un test para elegir la dimensión de la parte spike .

Conclusión

- Para el caso $p < n$ and $p, n \rightarrow \infty$, estudiamos la distribución asintótica del LRT_d para esfericidad parcial para el caso de modelos de covarianza spiked introducidos por Johnstone (2001).
- Para $p > n$ and $p, n \rightarrow \infty$ obtenemos también la distribución asintótica del estadístico LRT_d invirtiendo los roles de p y n .
- Conocer estas distribuciones nos permite desarrollar un test para elegir la dimensión de la parte spike .
- Mediante el estudio del comportamiento de la distribución del estadístico LRT_d para valores por debajo y por encima de la verdadera dimensión del spike, podemos modificar el estimador de la dimensión y probar la consistencia del mismo.

GRACIAS!



UNIVERSIDAD NACIONAL DEL LITORAL
FACULTAD DE INGENIERÍA QUÍMICA

| f | | | | **FIQUNL**

www.fiq.unl.edu.ar

email: estadistica.sfe@gmail.com