LA AGREGACIÓN SECTORIAL Y LA PÉRDIDA DE INFORMACIÓN CONTENIDA EN UNA TABLA INPUT-OUTPUT

Carmen Ramos

Regiolab
Departamento de Economía Aplicada
Universidad de Oviedo

MOTIVACIÓN DEL TRABAJO

¿POR QUÉ ES NECESARIO AGREGAR SECTORES EN UNA TIO?

CONSIDERACIÓN DE DISTINTAS FUENTES ESTADÍSTICAS

UN EJEMPLO

TIO D España	TIO D Asturias	Cuentas de Emisiones a la atmósfera	Innovación tecnológica	Encuesta de consumos energéticos
74 Ramas	66 Ramas	63 Ramas	50 Ramas	96 Ramas de Industria Extractiva y Manufacturera

DISTINTAS FUENTES ESTADÍSTICAS → DISTINTAS AGREGACIONES

MOTIVACIÓN DEL TRABAJO

UN EJEMPLO

CONSTRUCCIÓN MATRIZ SIMÉTRICA A PARTIR DE TABLAS DE ORIGEN Y DESTINO

Ramas de actividad

Productos

	1	2	…	74
1				
2				
3				
4				
5				
…				
108				
109				

Ramas de actividad

	1	…	74
1			
…			
74			

Productos

Agregación

OBJETIVO DEL TRABAJO

Si agregar es necesario

¿QUÉ "COSTES" TIENE?

- ✓ Pérdida de información
- ✓ Distintos resultados en el análisis estructural

TEORÍA ESTADÍSTICA DE LA INFORMACIÓN

Shanon (1948) $\Rightarrow$ ENTROPÍA

Sea X una v.a. discreta cuya distribución de probabilidad asociada es P= (p1, p2,…pn)
Se denomina entropía de la variable aleatoria X o de la distribución P= (p1, p2,…pn) a la expresión

$$H(X) = H(p_1, p_2 \ldots . p_n) = -\sum_{i=1}^{n} p_i \log p_i$$

Si la distribución es uniforme

$$p_i = \frac{1}{n}$$

H(X)=logn $\Leftarrow$ Máxima entropía

Si la distribución es degenerada

H(X)=0 $\Leftarrow$ Mínima entropía

TEORÍA ESTADÍSTICA DE LA INFORMACIÓN

Sea una v.a. bidimensional (X,Y), donde X representa los sectores que realizan compras e Y a los sectores vendedores.
La entropía de Shannon se formulará

$$H(X,Y) = -\sum_{i}\sum_{j} p_{ij} \log p_{ij}$$

Si todos los sectores tienen una capacidad de compra/venta similar , H(X,Y)=log r

Si el mercado se concentra alrededor de las compras /ventas que realiza un sector H(X,Y)=0

CANTIDAD DE INFORMACIÓN

¿POR QUÉ?

Demanda intermedia

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

$$\boxed{f_{ij} = \frac{x_{ij}}{\sum_i \sum_j x_{ij}}}$$

$$\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$$

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

AGREGACIÓN, TIOs E INFORMACIÓN

Antes de la agregación

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

$$H_0(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\log p_{ij}$$

Se agregan las filas y columnas de 1 a m

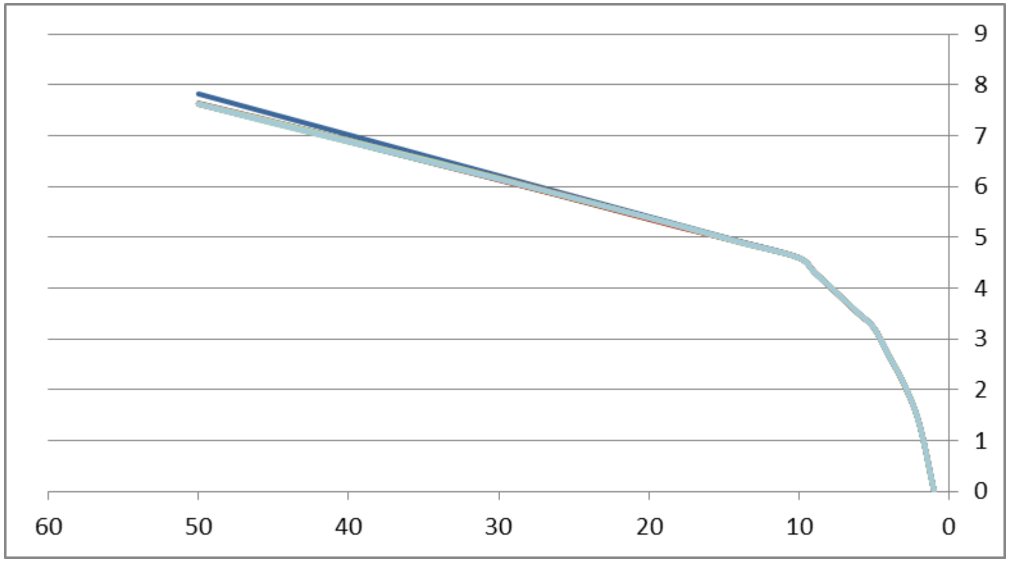
$$\begin{bmatrix} \sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij} & \sum_{i=1}^{m} p_{i,im+1} & \cdots & \sum_{i=1}^{m} p_{in} \\ \sum_{j=1}^{m} p_{m+1,j} & p_{m+1,m+1} & \cdots & p_{m+1,m+1} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{j=1}^{m} p_{nj} & p_{n,m+1} & \cdots & p_{nm} \end{bmatrix}$$

$$H_1(X,Y) = -\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right) + \left(\sum_{i=1}^{m} p_{i,m+1}\right)\log\left(\sum_{i=1}^{m} p_{i,m+1}\right) + \ldots \left(\sum_{i=1}^{m} p_{i,n}\right)\log\left(\sum_{i=1}^{m} p_{i,n}\right) + \left(\sum_{j=1}^{m} p_{m+1,j}\right)\log\left(\sum_{j=1}^{m} p_{m+1,j}\right) + \right.$$

$$\left. \ldots + \left(\sum_{j=1}^{m} p_{n\,j}\right)\log\left(\sum_{j=1}^{m} p_{nj}\right) + \sum_{i=m+1}^{n}\sum_{j=m+1}^{n} p_{ij}\log(p_{ij}) \right]$$

Gráfico N 1. Información asociada a una TIO según la agregación realizada

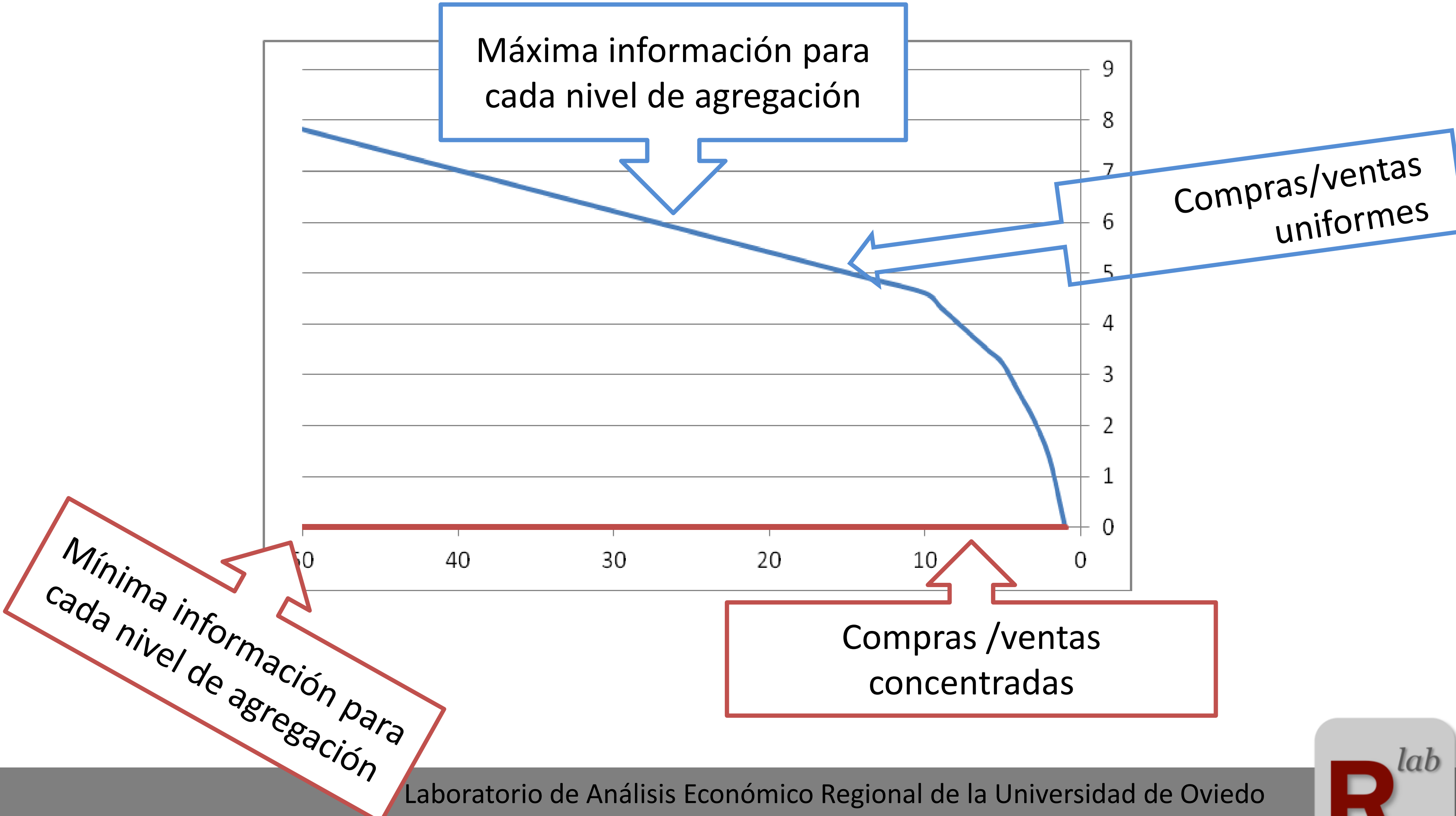
Situaciones intermedias



Muestras aleatorias n=2500

RESULTADOS EMPÍRICOS

Gráfico N 2. Cantidad de información asociada a una TIO. Situaciones extremas



PÉRDIDA DE INFORMACIÓN

$$\Delta(H) = H_0(X,Y) - H_1(X,Y) =$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} \log p_{ij} +$$

$$\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right) \log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right) + \left(\sum_{i=1}^{m} p_{i,m+1}\right) \log\left(\sum_{i=1}^{m} p_{i,m+1}\right) + .. + . \left(\sum_{i=1}^{m} p_{i,n}\right) \log\left(\sum_{i=1}^{m} p_{i,n}\right) + \left(\sum_{j=1}^{m} p_{m+1,j}\right) \log\left(\sum_{j=1}^{m} p_{m+1,j}\right) + \ldots . \right.$$

$$+\left(\sum_{j=1}^{m} p_{nj}\right) \log\left(\sum_{j=1}^{m} p_{nj}\right) + \sum_{i=m+1}^{n}\sum_{j=m+1}^{n} p_{ij} \log$$

AGREGACIÓN, TIOs E INFORMACIÓN

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\log p$$

Agregación en filas y columnas	No se agregan	Agregación en columnas	Agregación en filas

$$\Delta(H) = \left[\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\right]^{T} + \left[\sum_{j=m+1}^{n}\left(\sum_{i=1}^{m} p_{i,m+1}\right)\log\left(\sum_{i=1}^{m} p_{i,m+1}\right)\right]^{T} + \left[\sum_{i=m+1}^{n}\left(\sum_{j=1}^{m} p_{i,j}\right)\log\left(\sum_{j=1}^{m} p_{i,j}\right)\right]^{T}$$

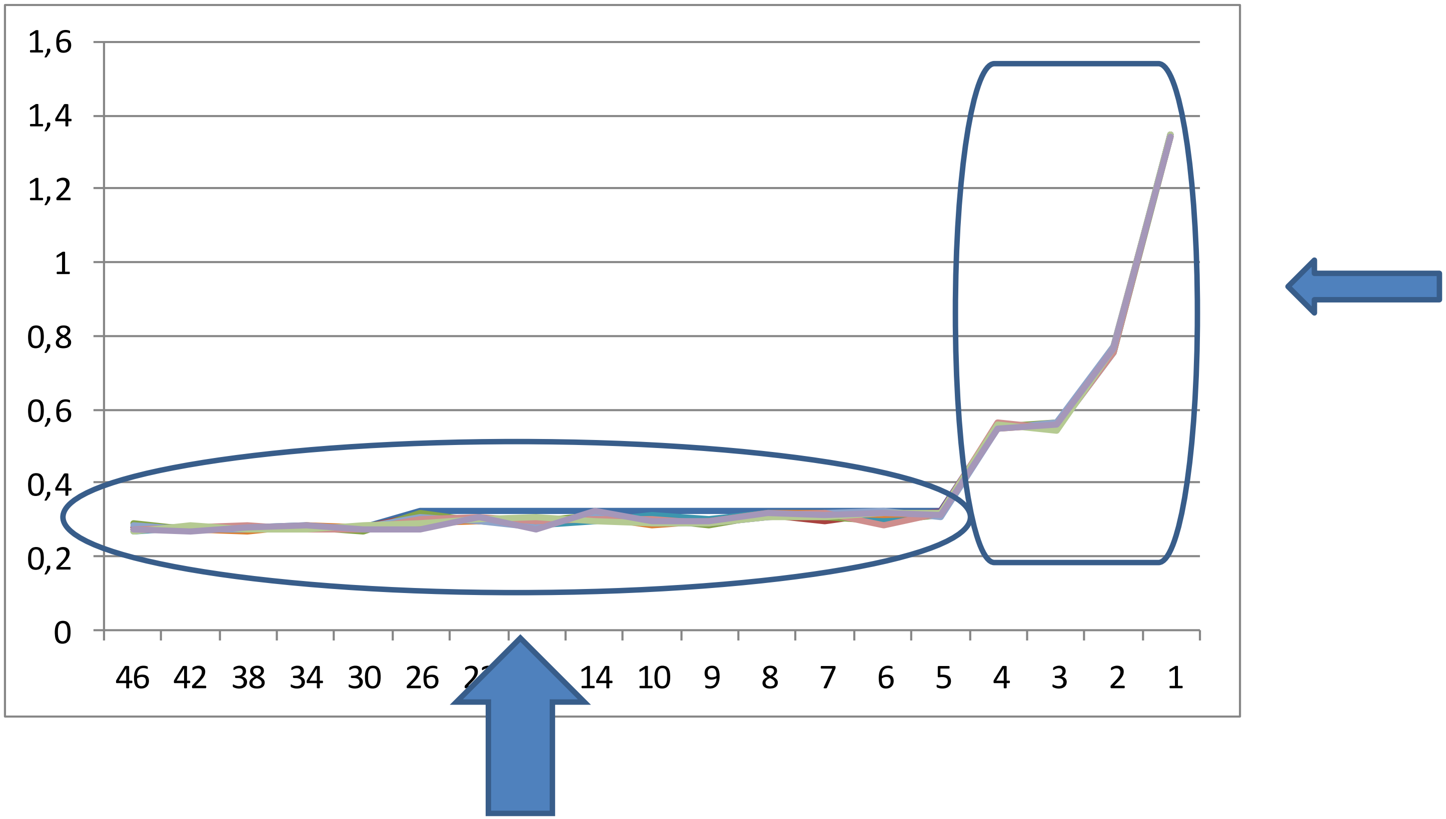
$$\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\right]^{T} = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right) - \sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\log p_{ij}$$

$$\left[\sum_{j=m+1}^{n}\left(\sum_{i=1}^{m} p_{i,m+1}\log\left(\sum_{i=1}^{m} p_{i,m+1}\right)\right)\right]^{T} = \left[\sum_{jm+1}^{n}\left(\sum_{i=1}^{m} p_{i,m+1}\log\left(\sum_{i=1}^{m} p_{i,m+1}\right)\right)\right] - \sum_{i=m+1}^{n}\sum_{j=1}^{m} p_{ij}\log p_{ij}$$

$$\left[\sum_{i=m+1}^{n}\left(\sum_{j=1}^{m} p_{i,j}\log\left(\sum_{j=1}^{m} p_{i,j}\right)\right)\right]^{T} = \left[\sum_{i=m+1}^{n}\left(\sum_{j=1}^{m} p_{i,j}\log\left(\sum_{j=1}^{m} p_{i,j}\right)\right)\right] - \sum_{i=m+1}^{n}\sum_{j=1}^{m} p_{ij}\log p_{ij}$$

Gráfico N. 3. Pérdida de información. Distribuciones intermedias

ΔH(X,Y)



Sectores

PÉRDIDA RELATIVA DE INFORMACIÓN

$$\Delta^R(H) = \boxed{\frac{H_0(X,Y)}{\log r}} - \boxed{\frac{H_1(X,Y)}{\log s}}$$

r: Número de sectores antes de agregar
s: Número de sectores después de agregar

$$\Delta^R(H) = \frac{-\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij} \log p_{ij}}{\log r} +$$

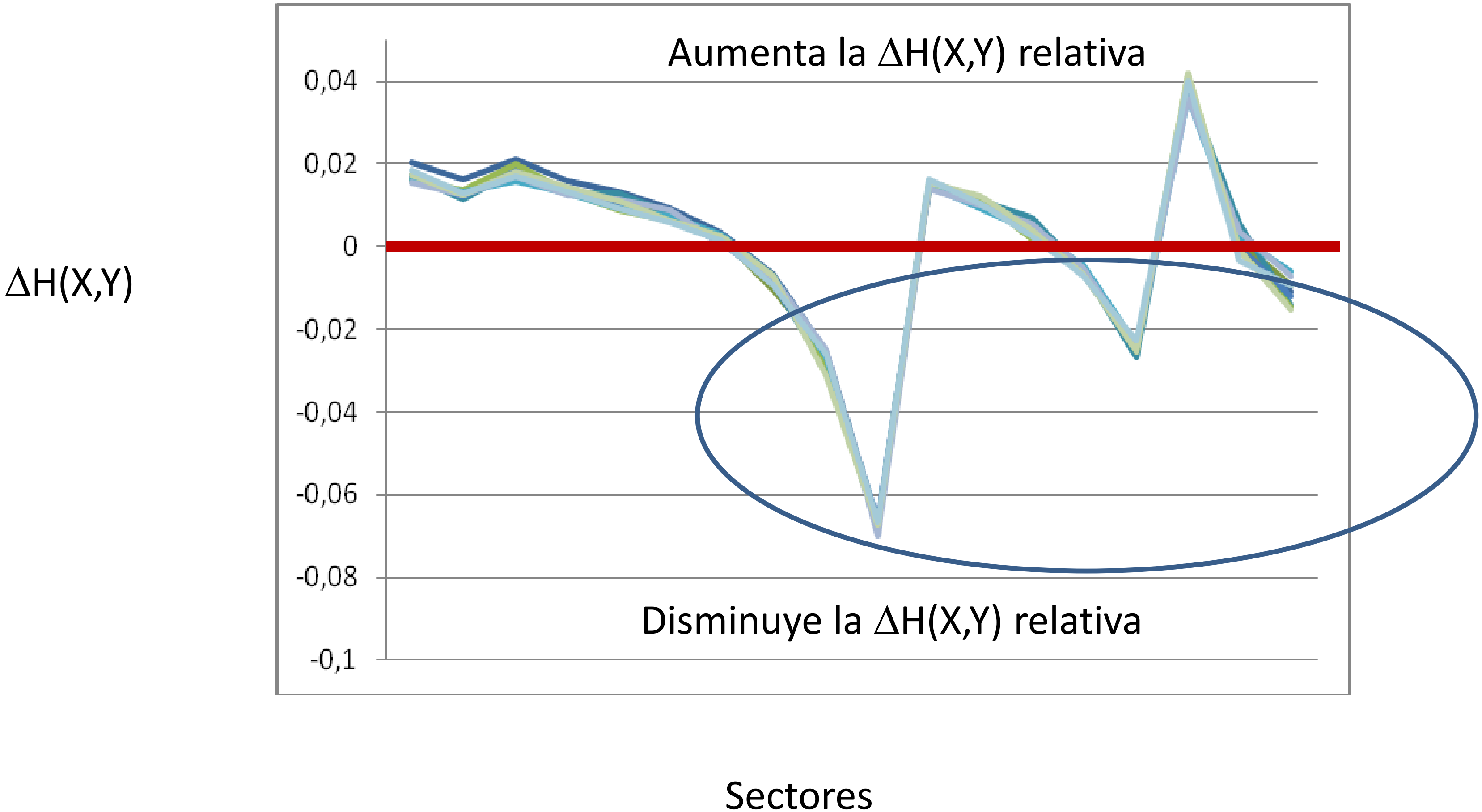
$$\frac{\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right)\log\left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\right) + \sum_{j=1m+1}^{n}\left(\sum_{i=1}^{m} p_{i,m+1}\right)\log\left(\sum_{i=1}^{m} p_{i,m+1}\right) + \sum_{i=m+1}^{n}\left(\sum_{j=1}^{m} p_{i,j}\right)\log\left(\sum_{j=1}^{m} p_{i,j}\right) + \sum_{i=m+1}^{n}\sum_{j=m+1}^{n} p_{ij}\log(p_{ij}) \right]}{\log s}$$

$$\boxed{\frac{H_0(X,Y)}{\log r} > \frac{H_1(X,Y)}{\log s}}$$

$$\boxed{\frac{H_0(X,Y)}{\log r} < \frac{H_1(X,Y)}{\log s}}$$

PÉRDIDA RELATIVA DE INFORMACIÓN

Gráfico N. 5. Pérdida de información relativa. Distribuciones intermedias



CANTIDAD DE INFORMACIÓN EN LA TIO DE ASTURIAS 2010

·

Agregaciones empleadas

FUENTE	Ramas
SADEI	66
EUROSTAT	60 (59)
WIOD	32
CNAE	21 (20)
CIIU	16
EUROSTAT	10
SADEI	4

CANTIDAD DE INFORMACIÓN EN LA TIO DE ASTURIAS 2010

Cuadro N.1. Cantidad de información asociada a la TIO de Asturias 2010

FUENTE	Ramas	H(X,Y)	Cota Superior	Porcentaje respecto cota
SADEI	66	5,8537	8,3793	69,859
EUROSTAT	59	5,6938	8,1551	69,819
WIOD	32	5,1830	6,9315	74,775
CNAE	20	3,9775	5,9915	66,386
CIIU	16	3,4658	5,5452	62,501
EUROSTAT	10	2,9605	4,6052	64,286
SADEI	4	1,8819	2,7726	67,874

CANTIDAD DE INFORMACIÓN EN LA TIO DE ASTURIAS 2010

Cuadro N. 2. Pérdida relativa debida a la agregación

Antes agregación

	H66	H59	H32	H20	H16	H10
H59	0,040					
H32	-4,915	-4,956				
H20	3,473	3,433	8,389			
H16	7,358	7,318	12,273	3,885		
H10	5,574	5,533	10,489	2,100	-1,785	
H4	1,985	1,945	6,900	-1,488	-5,373	-3,589

Después agregación

CONCLUSIONES

1. La agregación de sectores en una tabla input-output puede ser necesaria en ciertas situaciones
2. Una tabla contiene información que puede ser cuantificada mediante medidas de información estadística, entre las que se encuentra la entropía de Shanon.
3. A medida que una tabla se agrega su contenido de información se reduce, esta reducción es más fuerte cuando las tablas están muy agregadas.
4. La pérdida de información presenta una forma de L invertida, es decir, la mayor pérdida se presenta en tablas muy agregadas
5. Al determinar la pérdida relativa se aprecia que hay situaciones en los que una agregación conduce a una "ganancia" relativa de información.
6. Hemos aplicado los anteriores conceptos a la tabla de Asturias de 2010, la cual está inicialmente agregada a 66 ramas de actividad. Sobre dicha tabla se han aplicado diferentes agregaciones; la de EUROSTAT referida a tablas input-output, la de la base WIOD, la CNAE a 20 sectores, la CIIU, la de EUROSTAT a 10 ramas y la de SADEI a 4. Aunque la cantidad de información absoluta disminuye con la agregación, si consideramos la pérdida relativa, ésta es negativa (ganancia relativa) en el caso de las agregaciones propuestas por WIOD y por SADEI a 4 ramas.