



UNIVERSIDAD NACIONAL DE COLOMBIA

Una mirada a la Teoría de la Información a través de la Teoría de Conjuntos.

Sherly Paola Alfonso Sánchez.

Universidad Nacional de Colombia.
Facultad de Ciencias, Departamento de Matemáticas.
Bogotá D.C. Colombia.
2013

Una mirada a la Teoría de la Información a través de la Teoría de Conjuntos.

Sherly Paola Alfonso Sánchez.

Tesis de grado presentada como requisito parcial para optar al título de:
Magister en Ciencias - Matemáticas.

Director:

Doctor en Matemáticas. Humberto Sarria Zapata.

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Matemáticas.
Bogotá D.C. Colombia.

2013

A mis padres

Alvaro Alfonso y Stella Sánchez Tafurt

Agradecimientos

Gracias al Padre Eterno y a la madre Natura.

Gracias a la Universidad Nacional de Colombia por brindarme valiosas condiciones durante mi estudio en la Maestría.

Gracias al profesor Humberto Sarria por motivarme al estudio de un tema novedoso, brindarme su conocimiento y luz, los cuales me guiaron y fortalecieron como profesional y ser humano.

Gracias a mi familia, amigos y seres que me acompañaron durante este caminar.

Resumen

En este trabajo se introduce el concepto de medidas de información generadas por una función polimatroide, dando algunos ejemplos e interpretaciones. Posteriormente, se reconocen, bajo ciertas condiciones, dichas medidas de información como una f -medida, tomando como fundamento la construcción de la I -medida para las medidas de la información Shannon hecha en [12]. Es así como se realiza una mirada de estas medidas de información a través de la Teoría de Conjuntos.

Además, se estudian los elementos del campo F_n y su información, dada cualquier medida con signo. Por último, al existir una conexión entre las medidas de la información Shannon y la Teoría de Conjuntos se determina el concepto de I -diagramas.

Abstract

In this document, we introduce the concept of information measures generated by a polimatroid function giving some examples and interpretations. Then, we recognize under certain conditions the information measures as a f -measure, taking as a basis the I -measure construction in case of Shannon's information measures made in [12]. This is how is made a look of this information measures through the Set Theory.

In addition, we study the F_n field's elements and their information given by any signed measure. Finally, with the connection between Shannon information measures given and the Set Theory. the I -diagrams concept should be resolved at the end.

Contenido

| | |
|---|------------|
| Agradecimientos | vii |
| Resumen | ix |
| Introducción | 1 |
| 1. Preliminares | 4 |
| 1.1. Medidas de la información Shannon | 5 |
| 2. Funciones Polimatroides. | 14 |
| 2.1. Medidas de información generadas por una función polimatroide. | 15 |
| 3. La f-medida | 27 |
| 3.1. Nociones de Teoría de la Medida | 29 |
| 3.2. La f -medida para dos conjuntos | 31 |
| 3.3. Construcción general de la f -medida | 38 |
| 3.4. Elementos del campo F_n y su información | 50 |
| 3.4.1. Conjunto de Información | 50 |
| 3.4.2. Algoritmo para obtener conjuntos de información generados. | 58 |
| 4. Los I-diagramas | 67 |
| 4.1. I - diagramas para cadenas de Markov | 75 |
| 5. Conclusiones y recomendaciones | 87 |
| 5.1. Conclusiones | 87 |
| 5.2. Recomendaciones | 88 |
| A. Anexo: Una variación de la fórmula de inclusión-exclusión. | 89 |
| Bibliografía | 91 |

Introducción

Ya que el ser humano es social por naturaleza, toda sociedad humana siempre ha estado basada en procesos informativos; sin embargo, la forma y los medios de comunicación permiten diferenciar una época de otra. Tanto es así, que la información es vista en el ámbito social como algo elemental de forma similar al capital y al trabajo.

Después de la Segunda Guerra Mundial, al surgir la aplicación de la computación en procesos bibliográficos, fue necesario el uso de la noción técnica de información, es decir, no se hizo énfasis en el significado de los mensajes enviados de un emisor a un receptor, sino en la codificación y transmisión de éstos. Dicha noción técnica de información proviene desde finales del siglo XIX y comienzos del siglo XX por parte de físicos e ingenieros como Ludwig Boltzmann, John Von Neumann y Ralph Vinton Lyon Hartley. No obstante, el origen de la ciencia de la información actual es el artículo de Claude E. Shannon, “*A mathematical theory of communication*” en 1948. Shannon abre las perspectivas para el uso objetivo y formal del concepto de información, preguntas como: ¿De qué manera medir la información de forma objetiva?, tuvieron respuesta por medio de la introducción del concepto de entropía (Ver [2]).

La entropía mide la incertidumbre de una variable aleatoria X . Sea X una variable aleatoria discreta en un alfabeto \mathfrak{A} , cuyos posibles eventos son x_1, x_2, \dots, x_n con probabilidades de ocurrencia $p(x_1), p(x_2), \dots, p(x_n)$. Se define la entropía de X denotada por $H(X)$:

$$H(X) := - \sum_{x \in S_X} p(x) \log_{|\mathfrak{A}|} p(x),$$

donde $S_X = \{x_i | p(x_i) > 0\}$. Además de la entropía, se definen otras medidas de la información Shannon como lo son la entropía conjunta, la entropía condicional, la información mutua y la información mutua condicional.

Existe una conexión entre las medidas de la información Shannon y la Teoría de Conjuntos, ya que a cada identidad de Teoría de la información Shannon le corresponde una identidad en la Teoría de Conjuntos vía sustitución formal de símbolos. Este hecho, fue demostrado en 1991 por Yeung en su artículo “*A new outlook on Shannon’s information measures*”. Sin embargo, el uso de diagramas de Venn para representar la estructura de las medidas de la información Shannon para dos y tres variables aleatorias, fue sugerida por varios autores anteriores como lo fueron: Reza, Abramson, Paupoulis y Hu Guo Ding, pero sin justificación formal (Ver [9], [1], [8], [3]).

La conexión establecida entre estas dos teorías, permitió a Yeung demostrar ciertos hechos de la función de entropía, por medio de los I -diagramas desligándose de su definición. Adicionalmente, Yeung utiliza la construcción de la I -medida para demostrar la existencia de desigualdades no Shannon (Ver [12]).

Dado que la función de entropía pertenece a un conjunto general de funciones llamadas funciones polimatroides, con el objetivo de generalizar el estudio de Yeung, en este trabajo se definen las medidas de información generadas por cualquier función polimatroide. Posteriormente, se cuestiona si es posible establecer de forma análoga a lo hecho por Yeung, una conexión entre las medidas de información generadas por una función polimatroide y la Teoría de Conjuntos. Con el fin de responder a este cuestionamiento el trabajo se encuentra en el siguiente orden:

En el Capítulo 1, se presenta una introducción a las medidas de la información Shannon, definiciones y ciertos teoremas preliminares.

En el Capítulo 2, se definen las medidas de información generadas por una función polimatroide, se dan ejemplos e interpretaciones en Teoría de Matroides y en Topología. Uno de los ejemplos presentados se basa en el trabajo “*Una nueva construcción de los espacios topológicos Finitos desde las funciones submodulares*”.

En el Capítulo 3, se establece la conexión entre las medidas de información generadas por una función polimatroide y la Teoría de Conjuntos, mediante la construcción de la f -medida. De esta manera, la I -medida se puede considerar como un ejemplo de f -medida. Debido a esta construcción, se motiva el estudio de los elementos del campo F_n y su información, se crean algunas definiciones y demuestran afirmaciones acerca de estos conjuntos.

En el Capítulo 4, se hace un estudio de los diagramas relacionados con las medidas de la información Shannon, los cuales reciben el nombre de I -diagramas. Se presenta la estructura de los I -diagramas para tres y cuatro variables aleatorias que forman una cadena de Markov.

Los aportes del trabajo son:

- Se definen las medidas de información generadas por una función polimatroide, luego se estudian e interpretan en algunas teorías particulares.
- Se generaliza la construcción hecha por Yeung, por medio de la f -medida.
- Se establece una conexión entre las medidas de información generadas por una función polimatroide y la Teoría de Conjuntos. Haciendo explícita la construcción de esta relación entre ambas teorías y algunas limitaciones.

- Se crean definiciones relativas a la información determinada por un subconjunto del campo F_n , esto en términos de cualquier medida con signo. Se llega a la introducción de conceptos y resultados nuevos.
- Se retoman aspectos subrayados por Yeung en su artículo “*A new outlook on Shannon’s information measures*”, los cuales no se hacen explícitos en su libro más reconocido ([12]).

1. Preliminares

El nacimiento de la ciencia de la información actual, es sin lugar a dudas, el artículo de Claude E. Shannon (1916-2001) “*A Mathematical Theory of Communication*” publicado en 1948. Sin embargo, la noción técnica de información proviene desde finales del siglo XIX y comienzos del siglo XX de físicos e ingenieros como Ludwig Boltzmann, John Von Neumann y Ralph Vinton Lyon Hartley. Acerca de esto, Lyon Hartley en 1928 escribe:

“...dado que los sistemas de transmisión eléctrica no tienen nada que ver con seres humanos sino con máquinas es mejor eliminar todos los factores psicológicos involucrados en dicha noción”¹

Para Shannon, no es información sino mensajes lo que se envía de un emisor a un receptor. Es decir, Shannon se desliga del significado de los mensajes y se enfoca en la codificación y transmisión de estos

Así, Shannon abre las perspectivas para el uso objetivo y formal del concepto de información. Dado que el mundo en el cual vivimos es probabilístico para medir la información se tendrá en cuenta un cierto experimento X y se denotará:

- por $\Omega = \{x_1, x_2, \dots, x_n\}$ el conjunto de posibles resultados del experimento.
- y por $p(x_i)$ La posibilidad de ocurrencia del evento x_i .

La entropía mide la incertidumbre de una variable aleatoria X , por lo tanto debe tener las siguientes propiedades:

- La entropía es máxima cuando todos los eventos tienen la misma probabilidad.
- La entropía es cero cuando sucede solo un evento.
- La entropía es simétrica respecto a sus argumentos.
- La entropía es una función continua.

¹R.V.L.Hartley. Transmission of Information. *Bell System Technical Journal*, 7, p.536: “*it is desirable therefore to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities*”

1.1. Medidas de la información Shannon

En primer lugar se dará la definición de función de entropía, la cual fue motivada por la noción de entropía ya conocida en física.

Sea X una variable aleatoria discreta en un alfabeto \mathfrak{A} , cuyos posibles eventos son x_1, x_2, \dots, x_n con probabilidades de ocurrencia $p(x_1), p(x_2), \dots, p(x_n)$

Definición 1.1.1. Sean X, Y, Z variables aleatorias en el alfabeto \mathfrak{A} , con eventos $\{x_1, x_2, \dots, x_n\}$, $\{y_1, y_2, \dots, y_m\}$ y $\{z_1, z_2, \dots, z_r\}$ respectivamente. Se definen:

- $p(x_i, y_j, z_k) := p(X = x_i, Y = y_j, Z = z_k)$
- $p(x_i, y_j) := \sum_{k=1}^r p(x_i, y_j, z_k)$
- $p(x_i) := \sum_{j,k} p(x_i, y_j, z_k)$
- $p(x_i|y_j) := \frac{p(x_i, y_j)}{p(y_j)}$, si $p(y_j) \neq 0$
- $p(x_i, y_j|z_k) := \frac{p(x_i, y_j, z_k)}{p(z_k)}$, si $p(z_k) \neq 0$

Definición 1.1.2. La entropía se define como:

$$H(X) := - \sum_{x \in S_X} p(x) \log_{|\mathfrak{A}|} p(x),$$

donde $S_X = \{x_i | p(x_i) > 0\}$.

La base del logaritmo de la definición anterior puede ser cualquier $\beta > 1$. Y por notación se escribirá de ahora en adelante la entropía de la siguiente forma

$$H(X) := - \sum_{x \in S_X} p(x) \log p(x).$$

Definición 1.1.3. Para dos variables aleatorias X y Y se define la entropía conjunta como:

$$H(X, Y) := - \sum_{i,j} p(x_i, y_j) \log p(x_i, y_j).$$

Definición 1.1.4. Se definen las entropías condicionales. La entropía condicional de X dado que $Y = y_i$ es

$$H(X|Y = y_i) := - \sum_j p(x_j|y_i) \log p(x_j|y_i),$$

y la entropía condicional de X dado Y

$$H(X|Y) := \sum_i p(y_i) H(X|Y = y_i).$$

De manera similar:

$$H(X, Y|Z = z_k) := - \sum_{i,j} p(x_i, y_j|z_k) \log p(x_i, y_j|z_k),$$

$$H(X, Y|Z) := \sum_k p(z_k) H(X, Y|Z = z_k),$$

$$H(X|Y = y_j, Z = z_k) := - \sum_i p(x_i|y_j, z_k) \log p(x_i|y_j, z_k),$$

$$H(X|Y, Z) := \sum_{j,k} p(y_j, z_k) H(X|Y = y_j, Z = z_k).$$

Ahora se definirán la información mutua y la información mutua condicional.

Definición 1.1.5. Se define la información mutua de las variables X y Y :

$$I(X \wedge Y) := \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}.$$

Proposición 1.1.6. Si X y Y son variables aleatorias independientes, entonces

$$I(X \wedge Y) = 0.$$

Demostración. Al ser X y Y variables independientes, entonces $p(x_i, y_j) = p(x_i)p(y_j)$. Por lo tanto

$$\begin{aligned} I(X \wedge Y) &= \sum_{i,j} p(x_i)p(y_j) \log \frac{p(x_i)p(y_j)}{p(x_i)p(y_j)} \\ &= 0. \end{aligned}$$

□

Definición 1.1.7. Se define la información mutua condicional de las variables X y Y dado que $Z = z_k$:

$$I(X \wedge Y|Z = z_k) := \sum_{i,j} p(x_i, y_j|z_k) \log \frac{p(x_i, y_j|z_k)}{p(x_i|z_k)p(y_j|z_k)},$$

y la información mutua condicional de X y Y dado Z

$$I(X \wedge Y|Z) := \sum_k p(z_k)I(X \wedge Y|Z = z_k).$$

En lo que resta de esta sección con base en las definiciones dadas, se probarán ciertas propiedades que satisfacen la entropía, la entropía condicional, la información mutua y la información mutua condicional.

Proposición 1.1.8.

- a) $H(X) = - \sum_{i,j} p(x_i, y_j) \log p(x_i)$
- b) $H(X) = - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i)$
- c) $H(X, Y) = - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i, y_j)$

Demostración. Se tiene debido a la definición 1.1.1. □

Proposición 1.1.9.

- a) $H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j)$
- b) $H(X, Y|Z) = - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i, y_j|z_k)$
- c) $H(X|Y, Z) = - \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)}$

Demostración.

a)

$$\begin{aligned} H(X|Y) &= - \sum_j p(y_j)H(X|Y = y_j) \\ &= \sum_j p(y_j) \left(- \sum_i p(x_i|y_j) \log p(x_i|y_j) \right) \\ &= - \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j). \end{aligned}$$

b)

$$\begin{aligned}
H(X, Y|Z) &= \sum_k p(z_k) H(X, Y|Z = z_k) \\
&= \sum_k p(z_k) \left(- \sum_{i,j} p(x_i, y_j|z_k) \log p(x_i, y_j|z_k) \right) \\
&= - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i, y_j|z_k).
\end{aligned}$$

c)

$$\begin{aligned}
H(X|Y, Z) &= \sum_{j,k} p(y_j, z_k) H(X|Y = y_j, Z = z_k) \\
&= \sum_{j,k} p(y_j, z_k) \left(- \sum_i p(x_i|y_j, z_k) \log p(x_i|y_j, z_k) \right) \\
&= - \sum_{i,j,k} p(y_j, z_k) p(x_i|y_j, z_k) \log p(x_i|y_j, z_k) \\
&= - \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)}.
\end{aligned}$$

□

Se verán ciertas propiedades para la entropía condicional.

Teorema 1.1.10. *Para X y Y variables aleatorias,*

$$H(X|Y) = H(X, Y) - H(Y).$$

Demostración.

$$\begin{aligned}
H(X, Y) &= - \sum_{i,j} p(x_i, y_j) \log p(x_i, y_j) \\
&= - \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j) p(y_j) \\
&= - \sum_{i,j} p(x_i, y_j) (\log p(x_i|y_j) + \log p(y_j)) \\
&= - \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j) - \sum_{i,j} p(x_i, y_j) \log p(y_j) \\
&= H(X|Y) + H(Y), \text{ por las proposiciones 1.1.8 y 1.1.9}
\end{aligned}$$

□

Teorema 1.1.11. $I(X \wedge Y) = H(X) - H(X|Y)$

Demostración.

$$H(X) - H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log p(x_i) + \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j),$$

Por proposiciones 1.1.8 y 1.1.9

$$\begin{aligned} &= \sum_{i,j} p(x_i, y_j) (-\log p(x_i) + \log p(x_i|y_j)) \\ &= \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i|y_j)}{p(x_i)} = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \\ &= I(X \wedge Y). \end{aligned}$$

□

Se mostrará que la información mutua se puede escribir en términos de las entropías conjuntas.

Corolario 1.1.12. $I(X \wedge Y) = H(X) + H(Y) - H(X, Y)$

Demostración. Se tiene por el Teorema 1.1.10 y el Teorema 1.1.11.

□

En el siguiente teorema se probará que la entropía condicional, también se puede escribir en términos de las entropías conjuntas.

Teorema 1.1.13. $H(X|Y, Z) = H(X, Y, Z) - H(Y, Z)$.

Demostración.

$$\begin{aligned}
H(X, Y, Z) &= - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i, y_j, z_k) \text{ por definición 1.1.3} \\
&= - \sum_{i,j,k} p(x_i, y_j, z_k) \log(p(y_j, z_k)p(x_i|y_j, z_k)) \\
&= - \sum_{i,j,k} p(x_i, y_j, z_k) \log(p(z_k)p(y_j|z_k)p(x_i|y_j, z_k)) \\
&= - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(z_k) - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(y_j|z_k) \\
&\quad - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i|y_j, z_k) \\
&= - \sum_k \left(\sum_{i,j} p(x_i, y_j, z_k) \right) \log p(z_k) - \sum_{j,k} \left(- \sum_i p(x_i, y_j, z_k) \right) \log p(y_j|z_k) \\
&\quad - \sum_{i,j,k} p(y_j, z_k)p(x_i|y_j, z_k) \log p(x_i|y_j, z_k) \\
&= - \sum_k p(z_k) \log p(z_k) - \sum_{j,k} p(y_j, z_k) \log p(y_j|z_k) \\
&\quad + \sum_{j,k} p(y_j, z_k) \left(- \sum_i p(x_i|y_j, z_k) \log p(x_i|y_j, z_k) \right) \\
&= H(Z) + H(Y|Z) + \sum_{j,k} p(y_j, z_k) H(X|Y = y_j, Z = z_k) \\
&= H(Y, Z) + H(X|Y, Z) \text{ por teorema 1.1.10}
\end{aligned}$$

□

Teorema 1.1.14. $I(X \wedge Y|Z) = H(X|Z) - H(X|Y, Z)$.

Demostración.

$$H(X|Z) - H(X|Y, Z) = - \sum_{i,k} p(x_i, z_k) \log p(x_i|z_k) + \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)},$$

por proposición 1.1.9

$$\begin{aligned} &= - \sum_{i,k} p(x_i, z_k) \log \frac{p(x_i, z_k)}{p(z_k)} + \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} \\ &= - \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, z_k)}{p(z_k)} + \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \left(\log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} - \log \frac{p(x_i, z_k)}{p(z_k)} \right) \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \log \left(\frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} \frac{p(z_k)}{p(x_i, z_k)} \right) \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)p(z_k)}{p(y_j, z_k)p(x_i, z_k)} \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \log \left(\frac{p(x_i, y_j, z_k)p(z_k)}{p(z_k)p(z_k)} \frac{p(z_k)}{p(y_j, z_k)p(x_i, z_k)} \right) \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \log \left(\frac{p(x_i, y_j, z_k)}{p(z_k)} \frac{p(z_k)}{p(y_j, z_k)p(x_i, z_k)} \right) \\ &= \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j|z_k)}{p(y_j|z_k)p(x_i|z_k)} \\ &= \sum_{i,j,k} p(z_k)p(x_i, y_j|z_k) \log \frac{p(x_i, y_j|z_k)}{p(y_j|z_k)p(x_i|z_k)} \\ &= \sum_k p(z_k) \sum_{i,j} p(x_i, y_j|z_k) \log \frac{p(x_i, y_j|z_k)}{p(y_j|z_k)p(x_i|z_k)} \\ &= \sum_k p(z_k) I(X \wedge Y|Z = z_k) \\ &= I(X \wedge Y|Z). \end{aligned}$$

□

Debido a los teoremas anteriores se tiene que la información mutua condicional se escribe en términos de las entropías conjuntas.

Corolario 1.1.15.

$$I(X \wedge Y|Z) = H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z).$$

Demostración.

$$\begin{aligned} I(X \wedge Y|Z) &= H(X|Z) - H(X|Y, Z), \text{ por Teorema 1.1.14} \\ &= H(X, Z) - H(Z) - (H(X, Y, Z) - H(Y, Z)), \text{ por Teoremas 1.1.10 y 1.1.13} \\ &= H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z). \end{aligned}$$

□

Por último, se mostrará que la entropía, la entropía condicional y la información mutua son no negativas.

Teorema 1.1.16.

$$\begin{aligned} H(X) &\geq 0 \\ H(X, Y) &\geq 0 \\ H(X, Y, Z) &\geq 0 \\ H(X|Y) &\geq 0 \\ H(X, Y|Z) &\geq 0 \\ H(X|Y, Z) &\geq 0 \end{aligned}$$

Demostración. Esto se tiene fácilmente, debido a las definiciones dadas de cada una de estas medidas. □

Ahora, se mostrará la no negatividad de la información mutua condicional y por último de la información mutua.

Teorema 1.1.17. $I(X \wedge Y|Z) \geq 0$

Demostración.

$I(X \wedge Y|Z) = H(X|Z) - H(X|Y, Z)$ por teorema 1.1.14

$$\begin{aligned}
 &= - \sum_{i,k} p(x_i, z_k) \log p(x_i|z_k) - \sum_{i,j,k} p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} \text{ por proposición 1.1.9} \\
 &= - \sum_{i,k} \left(\sum_j p(x_i, y_j, z_k) \right) \log p(x_i|z_k) - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i|y_j, z_k) \\
 &= - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i|z_k) - \sum_{i,j,k} p(x_i, y_j, z_k) \log p(x_i|y_j, z_k) \\
 &= - \sum_{i,j,k} p(x_i, y_j, z_k) \log(p(x_i|z_k)p(x_i|y_j, z_k)) \geq 0.
 \end{aligned}$$

□

Teorema 1.1.18. $I(X \wedge Y) \geq 0$

Demostración.

$I(X \wedge Y) = H(X|Z) - H(X|Y, Z)$

$$\begin{aligned}
 &= - \sum_{i,j} p(x_i, y_j) \log p(x_i) + \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j) \\
 &= - \sum_{i,j} p(x_i, y_j) \log \left(\frac{p(x_i)}{p(x_i|y_j)} \right) \\
 &\geq - \log \left(\sum_{i,j} p(x_i, y_j) \frac{p(x_i)}{p(x_i|y_j)} \right) \\
 &= - \log \left(\sum_{i,j} p(x_i)p(y_j) \right) = 0
 \end{aligned}$$

□

2. Funciones Polimatroides.

En el capítulo anterior se dio la definición de función de entropía y con ésta, la de medidas de la información Shannon. En este capítulo, se define un conjunto más general de funciones llamado funciones polimatroides. Además, se introduce el concepto de medidas de información generadas por una función polimatroide f .

Si X es una variable aleatoria discreta, la entropía de X , $H(X)$ mide la cantidad de incertidumbre de X , a partir de esta función, se generan las medidas de la información Shannon. De manera análoga, para una función polimatroide f pueden generarse ciertas medidas de información, hecho que motiva la siguiente pregunta: ¿Qué clase de información brindan estas medidas?.

Para dar respuesta a la pregunta anterior, se dan dos ejemplos de funciones polimatroides asociadas con matroides y topologías. En cada caso, se interpreta el significado de algunas medidas de la información generadas por dichas funciones polimatroides. Con base en estos dos ejemplos y en lo ya estudiado de las medidas de la información Shannon, se concluye que las medidas de información generadas por una función polimatroide brindan información de la teoría en la cual se este trabajando. Por ejemplo, las medidas de información generadas por las funciones polimatroides asociadas a matroides y a topologías, brindan cierta información acerca de los conjuntos independientes del matroide y de los conjuntos cerrados de la topología.

Es importante señalar que, el segundo ejemplo presentado fue expuesto por Leonardo Roa en su Tesis de Maestría, *Una nueva construcción de los espacios topológicos Finitos desde las funciones submodulares* realizada en el año 2012.

2.1. Medidas de información generadas por una función polimatroide.

Definición 2.1.1. Sea S un conjunto finito. Una función $f : 2^S \rightarrow \mathbb{R}$ es una función polimatroide, si:

$$(P.1) \quad f(\emptyset) = 0$$

$$(P.2) \quad (\text{No decreciente}). \text{ Si } A \subseteq B \subseteq S, \text{ entonces } f(A) \leq f(B).$$

$$(P.3) \quad (\text{Submodularidad}). \text{ Si } A, B, C \subseteq S, \text{ entonces}$$

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B).$$

Proposición 2.1.2. Los axiomas polimatroides son equivalentes a (P.1) y la siguiente propiedad:

$$(P.4) \quad \text{Si } A, B, C \subseteq S, \text{ entonces}$$

$$f(A \cup C) + f(B \cup C) \geq f(C) + f(A \cup B \cup C).$$

Demostración. Consultar [13]. □

Proposición 2.1.3. El conjunto de todas las funciones polimatroides definidas sobre S forman un cono convexo.

Demostración. Sean f, g funciones polimatroides sobre S , y α, θ números reales positivos, se vera que $\alpha f + \theta g$ es una función polimatroide.

- $(\alpha f + \theta g)(\emptyset) = (\alpha f)(\emptyset) + (\theta g)(\emptyset) = 0 + 0 = 0$, al ser f y g funciones polimatroides.
- Al ser α, θ números reales positivos, $\alpha f + \theta g$ es una función no decreciente al ser suma de funciones no decrecientes.
- Sean $A, B \subseteq S$,

$$\begin{aligned} (\alpha f + \theta g)(A) + (\alpha f + \theta g)(B) &= \alpha(f(A) + f(B)) + \theta(g(A) + g(B)) \\ &\geq \alpha(f(A \cup B) + f(A \cap B)) + \theta(g(A \cup B) + g(A \cap B)) \\ &\geq (\alpha f + \theta g)(A \cup B) + (\alpha f + \theta g)(A \cap B) \end{aligned}$$

Por lo tanto $\alpha f + \theta g$ es una función polimatroide. □

Como ejemplos de funciones polimatroides se encuentran: la dimensión de subespacios vectoriales, la función de probabilidad y la función de cardinalidad; note que para estos ejemplos, en el axioma (P.3) se satisface la igualdad.

Además de los ejemplos ya citados, existen funciones polimatroides asociadas en diferentes tipos de teorías, como lo son: *los polimatroides lineales* asociados con matrices (Ver [10]) y la función rango de un matroide (Ver [5]). Más aún, existen funciones polimatroides en Teoría de Juegos (función característica) (Ver [10]), en Teoría de Redes y en Topologías finitas (Ver [6]).

Observación:

Posteriormente se introducirá la definición de medidas de información generadas por una función polimatroide, estas son motivadas por las identidades que se satisfacen para las medidas de la información Shannon, como se mostró en los preliminares:

$$\begin{aligned} H(X, Y) &= H(X \cup Y) \\ H(X|Y) &= H(X, Y) - H(Y) \\ I(X \wedge Y) &= H(X) - H(X|Y) \\ I(X \wedge Y|Z) &= H(X|Z) - H(X|Y, Z) \end{aligned}$$

Definición 2.1.4. Si $f : 2^S \rightarrow \mathbb{R}$ es una función polimatroide y $X, Y, Z \subseteq S$, se definen las siguientes medidas de información generadas por f :

- (i) $f(X, Y) := f(X \cup Y)$ Información conjunta
- (ii) $f(X|Y) := f(X, Y) - f(Y)$ Información condicional
- (iii) $I(X \wedge Y) := f(X) - f(X|Y)$ Información mutua
- (iv) $I(X \wedge Y|Z) := f(X|Z) - f(X|Y, Z)$ Información mutua condicional

Con base en las anteriores definiciones se probarán ciertas identidades para funciones polimatroides (*identidades polimatroides*).

Teorema 2.1.5. Si $f : 2^S \rightarrow \mathbb{R}$ es una función polimatroide y $X, Y, Z \subseteq S$ se tienen las siguientes igualdades:

$$f(\emptyset) = 0 \quad (2-1)$$

$$f(X) \geq 0 \quad (2-2)$$

$$f(X|Y) \geq 0 \quad (2-3)$$

$$I(X \wedge Y) \geq 0 \quad (2-4)$$

$$f(X|Z) + f(Y|Z) \geq f(X, Y|Z) \quad (2-5)$$

$$f(X, Z|Y) \geq f(X|Y) \geq f(X|Y, Z) \quad (2-6)$$

$$I(X \wedge Y) = f(X) + f(Y) - f(X, Y) \quad (2-7)$$

$$I(X \wedge Y|Z) = f(X, Z) + f(Y, Z) - f(Z) - f(X, Y, Z) \quad (2-8)$$

$$I(X \wedge Y, Z) = I(X \wedge Z) + I(X \wedge Y|Z) \quad (2-9)$$

$$I(X \wedge Y|Z) \geq 0 \quad (2-10)$$

Demostración. (2-1) Se tiene por la propiedad (P.1).

(2-2) Ya que para todo X , $\emptyset \subseteq X$ por (P.2) se tiene que $f(X) \geq f(\emptyset)$. Y de (P.1), $f(X) \geq 0$.

(2-3) Como $Y \subseteq X \cup Y$ por (P.2):

$$\begin{aligned} f(Y) &\leq f(X, Y) \\ 0 &\leq f(X, Y) - f(Y) =: f(X|Y) \end{aligned}$$

(2-4)

$$\begin{aligned} I(X \wedge Y) &:= f(X) - f(X|Y) \\ &= f(X) - (f(X, Y) - f(Y)) \\ &= f(X) + f(Y) - f(X, Y) \\ &\geq f(X \cap Y), \text{ por (P.3)} \\ &\geq 0, \text{ por (2-2)} \end{aligned}$$

(2-5)

$$\begin{aligned} f(X|Z) + f(Y|Z) &= f(X, Z) - f(Z) + f(Y, Z) - f(Z) \\ &= (f(X, Z) + f(Y, Z) - f(Z)) - f(Z) \\ &\geq f(X, Y, Z) - f(Z), \text{ por (P.4)} \\ &= f(X, Y|Z) \end{aligned}$$

(2-6) Las desigualdades se probarán en dos pasos:

i) Como $X \cup Y \subseteq X \cup Y \cup Z$ entonces por (P.2), $f(X \cup Y) \leq f(X \cup Y \cup Z)$. Teniendo en cuenta esto:

$$f(X, Z|Y) = f(X, Z, Y) - f(Y) \geq f(X, Y) - f(Y) = f(X|Y).$$

ii)

$$\begin{aligned} f(X|Y) &= f(X, Y) - f(Y) \\ &\geq f(X, Y, Z) - f(Y), \text{ por (P.4)} \\ &= f(X, Z|Y). \end{aligned}$$

(2-7)

$$\begin{aligned} f(X) + f(Y) - f(X, Y) &= f(X) - (f(X, Y) - f(Y)) \\ &= f(X) - f(X|Y) = I(X \wedge Y). \end{aligned}$$

(2-8)

$$\begin{aligned} I(X \wedge Y|Z) &= f(X|Z) - f(X|Y, Z) \\ &= f(X, Z) - f(Z) - (f(X, Y, Z) - f(Y, Z)) \\ &= f(X, Z) + f(Y, Z) - f(Z) - f(X, Y, Z). \end{aligned}$$

(2-9) Por un lado se tiene que:

$$I(X \wedge Y, Z) = f(X) - f(X|Y, Z) = f(X) - f(X, Y, Z) + f(Y, Z)$$

Y

$$\begin{aligned} I(X \wedge Z) + I(X \wedge Y|Z) &= f(X) - f(X|Z) + f(X|Z) - f(X|Y, Z) \\ &= f(X) - f(X, Y, Z) + f(Y, Z) \end{aligned}$$

$$\text{Así } I(X \wedge Y, Z) = I(X \wedge Z) + I(X \wedge Y|Z)$$

(2-10) Por (2-8) y (P.4) se tiene que $I(X \wedge Y|Z) \geq 0$.

□

Las funciones polimatroides se presentan en varias estructuras matemáticas. A continuación, se presentan dos ejemplos el primero relativo a matroides y el segundo a topologías.

Ejemplo 2.1.6. *Considerese el grafo $G = (V, E)$ que se muestra a continuación*

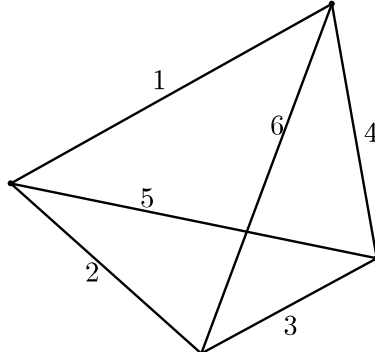


Figura 2-1.: Grafo $G = (V, E)$

Donde V y E son los conjuntos de vértices y aristas del grafo, respectivamente.

Sea $M(G) = (E, I)$ el matroide generado por el grafo G . Se recuerda que en este caso, el conjunto de independientes I , está dado por:

$$I = \{X \subseteq E \mid X \text{ no contiene un ciclo}\}.$$

La función rango del matroide, es una función $r : 2^E \rightarrow \mathbb{R}$ tal que, si $X \subseteq E$, $r(X)$ es el cardinal del independiente más grande contenido en X . Además, la función rango de un matroide resulta ser una función polimatroide (Ver [5]).

Si $X = \{1, 2, 3, 4\}$, $Y = \{2, 3, 5\}$ y $Z = \{1, 2, 4, 5\}$, entonces para este caso particular se pueden realizar las siguientes interpretaciones:

i) $r(X) = r(Z) = 3$ y $r(Y) = 2$.

$r(X)$ es el número de elementos de un subconjunto de X con el mayor cardinal tal que no contiene ciclos, por ejemplo $\{1, 2, 3\} \subseteq X$.

ii) $r(X, Y) = r(Y, Z) = r(X, Z) = 3$.

iii)

$$r(X|Y) = r(X, Y) - r(Y) = 3 - 2 = 1$$

$$r(X|Z) = r(X, Z) - r(Z) = 3 - 3 = 0$$

$$r(Y|X) = r(X, Y) - r(X) = 3 - 3 = 0$$

Se observa que en este caso $r(X|Y)$ se puede interpretar como el número mínimo de aristas de $X - Y$ tales que al unirlos con Y el conjunto resultante tiene rango igual al de $X \cup Y$. Note que Y se puede unir con la arista 1 ó 4 de $X - Y$, de manera que $Y \cup \{1\}$ o $Y \cup \{4\}$ tiene rango igual a 3, el cual es el rango de $X \cup Y$.

iv)

$$I(X \wedge Y) = r(X) - r(X|Y) = 3 - 1 = 2$$

$$I(X \wedge Z) = r(X) - r(X|Z) = 3 - 0 = 3$$

$$I(Y \wedge Z) = r(Y) - r(Y|Z) = 2 - 0 = 2$$

Se observa que en este caso $I(X \wedge Y)$ se puede interpretar como el rango de $X \cap Y$. De manera similar, se pueden interpretar las demás informaciones mutuas.

v) Para calcular la informaciones mutuas condicionales primero, se observará lo siguiente:

$$r(X|Y, Z) = r(X, Y, Z) - r(Y, Z) = 3 - 3 = 0,$$

$$r(Y|X, Z) = r(X, Y, Z) - r(X, Z) = 3 - 3 = 0,$$

$$r(Z|X, Y) = r(X, Y, Z) - r(X, Y) = 3 - 3 = 0.$$

Por lo tanto, en este caso todas las medidas de la información mutua condicional coinciden con alguna información condicional, por ejemplo:

$$I(X \wedge Y|Z) := r(X|Z) - r(X|Y, Z) = r(X|Z).$$

Algunas de las interpretaciones presentadas en el ejemplo anterior se pueden generalizar, con este objetivo surgen las siguientes proposiciones.

Proposición 2.1.7.

$$r(X|Y) = r(X, Y) - r(Y),$$

determina el número mínimo de aristas de $X - Y$ que se deben agregar a Y , para que el rango del conjunto resultante sea igual al de $X \cup Y$.

Demostración. Como la función rango de un matroide es una función polimatroide y $Y \subseteq X \cup Y$, entonces $r(X, Y) \geq r(Y)$. La demostración se hará teniendo en cuenta dos casos posibles:

(a) $r(X, Y) = r(Y)$.

Si se tiene esta condición el número mínimo de aristas de $X - Y$ que se deben agregar a Y para que el conjunto resultante tenga rango igual al de $X \cup Y$, es cero.

Lo anterior, coincide con $r(X|Y) = 0$, ya que

$$\begin{aligned} r(X|Y) &= r(X, Y) - r(Y) \\ &= r(Y) - r(Y) = 0. \end{aligned}$$

(b) $r(X, Y) > r(Y)$.

Si se tiene lo anterior, entonces $X - Y \neq \emptyset$, ya que de lo contrario, $X \subseteq Y$ y $r(X, Y) = r(Y)$, que corresponde al caso contemplado en (a).

Como la imagen de la función rango de un matroide son números enteros no negativos, y $r(X, Y) > r(Y)$, existe $s \in \mathbb{Z}^+$ tal que

$$r(X, Y) = r(Y) + s. \quad (1)$$

Por lo tanto, existen mínimo s aristas de $X - Y$ tal que al unir las con Y , el rango del conjunto resultante sea igual al rango de $X \cup Y$.

Se verá que $r(X|Y) = s$:

$$\begin{aligned} r(X|Y) &= r(X, Y) - r(Y) \\ &= r(Y) + s - r(Y), \text{ por (1)} \\ &= s. \end{aligned}$$

Así de (a) y (b), $r(X|Y)$ tiene la interpretación dada en la proposición. \square

La siguiente proposición establece cierta relación entre la información mutua de X y Y y el rango de $X \cap Y$.

Proposición 2.1.8.

$$r(X \cap Y) \leq I(X \wedge Y).$$

Demostración. Dado que la función rango de un matroide es una función polimatroide, se tiene:

$$0 \leq r(X, Y) + r(X \cap Y) \leq r(X) + r(Y).$$

Por lo tanto,

$$\begin{aligned} r(X \cap Y) &\leq r(X) + r(Y) - r(X, Y) \\ &\leq r(X) - r(X|Y) = I(X \wedge Y) \end{aligned}$$

\square

Corolario 2.1.9. *Existe un único número entero no negativo z tal que*

$$I(X \wedge Y) = r(X \cap Y) + z.$$

Demostración. Se tiene por la proposición 2.1.8. □

Después de haber dado el primer ejemplo de función polimatroide asociada a matroides, se dará un ejemplo de función polimatroide asociada a topología. Este ejemplo y las proposiciones presentadas acerca de las medidas de información generadas por dicha función polimatroide, son expuestos en la tesis de Maestría “Una nueva construcción de los espacios topológicos finitos desde las funciones submodulares”, realizada por Leonardo Roa.

Antes de dar dicho ejemplo, es necesario recordar algunos conceptos preliminares acerca de espacios topológicos.

Definición 2.1.10. *Una topología τ sobre un conjunto E , es un subconjunto de 2^E que satisface:*

1. Si $\{O_\lambda\}_{\lambda \in \Lambda} \subseteq \tau$, entonces $\bigcup_{\lambda \in \Lambda} O_\lambda \in \tau$.
2. Si $\{O_\lambda\}_{i=1}^n \subseteq \tau$, entonces $\bigcap_{i=1}^n O_i \in \tau$.
3. $E, \emptyset \in \tau$.

Los elementos de τ se denominan conjuntos abiertos. Los complementos de los conjuntos abiertos se denominan conjuntos cerrados. La pareja (E, τ) se denomina espacio topológico.

Definición 2.1.11. *Sea E un conjunto. Diremos que el operador $c : 2^E \rightarrow 2^E$ es de clausura, si dados $I, J \subseteq E$:*

1. $I \subseteq c(I)$ (*Expansión*)
2. $c(I) = c(c(I))$ (*Idempotente*)
3. $c(\emptyset) = \emptyset$.
4. $c(I \cup J) = c(I) \cup c(J)$

Teorema 2.1.12. *Cualquier operador de clausura $c : 2^E \rightarrow 2^E$, determina una única topología sobre E , en la cual los conjuntos cerrados son los puntos fijos del operador, es decir $A \subseteq E$ es cerrado si y solo si $c(A) = A$.*

Demostración. Ver [7]. □

Con base en la definición de operador de clausura, se construirá a continuación, una función polimatroide:

Definición 2.1.13. *Sea \mathcal{C} el conjunto de cerrados del espacio topológico sobre E . Para todo $I \subseteq E$, y cada $J \in \mathcal{C}$. se define la función*

$$\begin{aligned} q_J : 2^E &\rightarrow \mathbb{Z} \\ I &\mapsto q_J(I) = q_\phi(I - J), \end{aligned} \tag{2-11}$$

donde

$$\begin{aligned} q_\phi : 2^E &\rightarrow \mathbb{Z} \\ I &\mapsto q_\phi(I) = \begin{cases} 0, & \text{si } I = \phi, \\ 1, & \text{si } I \neq \phi. \end{cases} \end{aligned} \tag{2-12}$$

Proposición 2.1.14. *La función $q_J : 2^E \rightarrow \mathbb{Z}$, definida anteriormente, es una función polimatroide.*

Demostración. Se mostrará que q_J satisface los tres axiomas polimatroides.

(P.1) $q_J(\emptyset) = q_\phi(\emptyset - J) = q_\phi(\emptyset) = 0$.

(P.2) Sean $A \subseteq B \subseteq E$. Entonces, $A - J \subseteq B - J$. Se consideran los siguientes dos casos:

- Si $B - J = \emptyset$, entonces $A - J = \emptyset$. Así:

$$\begin{aligned} q_\phi(A - J) &= q_\phi(B - J) = 0 \\ q_J(A) &= q_J(B) = 0. \end{aligned}$$

- Si $B - J \neq \emptyset$, entonces $q_\phi(B - J) = 1$. Por otro lado, $q_\phi(A - J)$ es a lo más 1, así:

$$\begin{aligned} q_\phi(A - J) &\leq q_\phi(B - J) \\ q_J(A) &\leq q_J(B) \end{aligned}$$

(P.3) q_J es submodular, pues si $A, B \subseteq E$, los posibles casos son:

- a) Si $A - J = \phi$ y $B - J = \phi$, entonces $(A \cup B) - J = \phi$, $(A \cap B) - J = \phi$ luego
 $q_J(A) + q_J(B) = q_J(A \cup B) + q_J(A \cap B) = 0$.
- b) Si $A - J = \phi$ y $B - J \neq \phi$, entonces $(A \cup B) - J \neq \phi$, $(A \cap B) - J = \phi$ luego
 $q_J(A) + q_J(B) = q_J(A \cup B) + q_J(A \cap B) = 1$.
- c) Si $A - J \neq \phi$ y $B - J \neq \phi$, entonces $(A \cup B) - J \neq \phi$ y
 $2 = q_J(A) + q_J(B) \geq q_J(A \cup B) + q_J(A \cap B)$,
ya que $q_J(A \cap B) \leq 1$.

□

Teniendo en cuenta que q_J es una función polimatroide, se define la función $f_{\mathbb{Z}} : 2^E \rightarrow \mathbb{Z}$ de la siguiente forma:

$$f_{\mathbb{Z}}(I) = \sum_{J \in \mathcal{C}} q_J(I).$$

La función $f_{\mathbb{Z}}$ es polimatroide, debido a la Proposición 2.1.3. Además, la función $f_{\mathbb{Z}}(I)$ es el número de cerrados que no contienen a I .

Ejemplo 2.1.15. Sea $E = \{a, b, c\}$. Consideremos el operador de clausura ¹

$$\begin{aligned} c_{\tau} : 2^E &\rightarrow 2^E \\ \phi &\mapsto \phi \\ \{a\} &\mapsto \{a\} \\ \{b\} &\mapsto \{b, c\} \\ \{c\} &\mapsto \{b, c\} \\ \{a, b\} &\mapsto \{a, b, c\} \\ \{a, c\} &\mapsto \{a, b, c\} \\ \{b, c\} &\mapsto \{b, c\} \\ \{a, b, c\} &\mapsto \{a, b, c\} \end{aligned}$$

Dado este operador de clausura, el conjunto de cerrados de la topología está determinado por $\mathcal{C} = \{\phi, \{a\}, \{b, c\}, \{a, b, c\}\}$.

¹Este ejemplo es dado en [6], pág 40

A continuación se muestra la tabla con los valores de la función $f_{\mathbb{Z}}(X|Y) = f_{\mathbb{Z}}(X, Y) - f_{\mathbb{Z}}(Y)$. El conjunto X corresponde a la fila, y el conjunto Y corresponde a la columna.:

| $f_{\mathbb{Z}}(X Y)$ | ϕ | {a} | {b} | {c} | {a,b} | {a,c} | {b,c} | {a,b,c} |
|-----------------------|--------|-----|-----|-----|-------|-------|-------|---------|
| ϕ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| {a} | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| {b} | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| {c} | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| {a,b} | 3 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| {a,c} | 3 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| {b,c} | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| {a,b,c} | 3 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |

Ahora se muestra la tabla con los valores $I(X \wedge Y) = f_{\mathbb{Z}}(X) - f_{\mathbb{Z}}(X|Y)$:

| $I(X \wedge Y)$ | ϕ | {a} | {b} | {c} | {a,b} | {a,c} | {b,c} | {a,b,c} |
|-----------------|--------|-----|-----|-----|-------|-------|-------|---------|
| ϕ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| {a} | 0 | 2 | 1 | 1 | 2 | 2 | 1 | 2 |
| {b} | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| {c} | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| {a,b} | 0 | 2 | 2 | 2 | 3 | 3 | 2 | 3 |
| {a,c} | 0 | 2 | 2 | 2 | 3 | 3 | 2 | 3 |
| {b,c} | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| {a,b,c} | 0 | 2 | 2 | 2 | 3 | 3 | 2 | 3 |

Ahora, se presentan de forma general, las interpretaciones de algunas medidas de información generadas por $f_{\mathbb{Z}}$, en el ámbito de la topología ².

Proposición 2.1.16. Sean (E, τ) un espacio topológico y $f_{\mathbb{Z}}$ la función submodular no decreciente asociada. Entonces,

$$I(X \wedge Y) = f_{\mathbb{Z}}(X) - f_{\mathbb{Z}}(X|Y)$$

determina el número de conjuntos cerrados, que no contienen a X y a Y .

²Resultado demostrado en [6], páginas 41-43.

Demostración. Fijando $J \in \mathcal{C}$, se tiene que:

1. Si $q_J(X) = 0$ y $q_J(Y) = 0$, entonces $X \subseteq J$ y $Y \subseteq J$, así $X \cup Y \subseteq J$, es decir, $q_J(X \cup Y) = 0$, de donde

$$q_J(X) + q_J(Y) - q_J(X, Y) = 0.$$

2. Si $q_J(X) = 0$ y $q_J(Y) = 1$, entonces $X \subseteq J$ y $Y \not\subseteq J$, así $X \cup Y \not\subseteq J$, es decir, $q_J(X \cup Y) = 1$, de donde

$$q_J(X) + q_J(Y) - q_J(X, Y) = 0.$$

3. Si $q_J(X) = 1$ y $q_J(Y) = 0$, de forma similar a 2 se tiene

$$q_J(X) + q_J(Y) - q_J(X, Y) = 0.$$

4. Si $q_J(X) = 1$ y $q_J(Y) = 1$, entonces $X \not\subseteq J$ y $Y \not\subseteq J$, así $X \cup Y \not\subseteq J$, es decir, $q_J(X \cup Y) = 1$, de donde

$$q_J(X) + q_J(Y) - q_J(X, Y) = 1.$$

De este modo, el valor $q_J(X) + q_J(Y) - q_J(X, Y) = 1$ solo cuando $X \not\subseteq J$ y $Y \not\subseteq J$. Entonces

$$I(X \wedge Y) = \sum_{J \in \mathcal{C}} q_J(X) + q_J(Y) - q_J(X, Y),$$

determina el número de conjuntos cerrados, que no tienen a X ni a Y como subconjuntos. \square

Proposición 2.1.17. Sean (E, τ) un espacio topológico y $f_{\mathbb{Z}}$ la función submodular no decreciente asociada. Entonces,

$$f_{\mathbb{Z}}(X|Y) = f_{\mathbb{Z}}(X, Y) - f_{\mathbb{Z}}(Y)$$

determina el número de conjuntos cerrados, que no contienen a X ; menos el número de conjuntos cerrados, que no tienen a X ni a Y .

Demostración. Se tiene debido a la Proposición 2.1.15. \square

3. La f -medida

En el capítulo anterior, se introdujo el concepto de medidas de información generadas por cualquier función polimatroide. En este capítulo, se mostrará que la función de entropía es polimatroide. Así, las medidas de la información Shannon son las medidas de información generadas por la función polimatroide de entropía.

Existe un puente entre la Teoría de Conjuntos y la Teoría de la Información Shannon; el cual fue formalizado en 1991, por Yeung en *A new outlook on Shannon's information measures* ([11]), por medio del concepto de I -medida. De lo anterior, surge la siguiente pregunta, ¿Existe, siguiendo de forma análoga la construcción de Yeung, un puente entre la Teoría de Conjuntos y las medidas de información generadas por cualquier función polimatroide f ? Esta pregunta se responde afirmativamente, bajo ciertas condiciones. El puente entre estas teorías es posible mediante la construcción de una medida con signo, la f -medida; así, bajo ciertas condiciones sobre los conjuntos del dominio de la función polimatroide f , las medidas de información generadas por f se pueden ver como una medida con signo sobre cierto campo.

Esta mirada de la Teoría de la Información Shannon a través de la Teoría de Conjuntos, permite establecer por cada identidad en Teoría de Conjuntos, una identidad en Teoría de la Información y viceversa. En este capítulo, se aclarará qué se entiende por identidad en Teoría de Conjuntos y se verán algunas limitaciones de este puente entre ambas teorías.

En la última sección, se desarrolla un tema motivado por el teorema principal para la construcción de la f -medida, en el que se estudia la información generada por los elementos de un campo F_n en cuanto a cualquier medida con signo.

Se recuerda que, si X es una variable aleatoria discreta,

$$H(X) := \sum_{x \in S_X} p(x) \log p(x),$$

donde $S_X = \{x_i | p(x_i) > 0\}$.

Definición 3.0.18.

$$H(\emptyset) := 0.$$

Teorema 3.0.19. *La función de entropía es una función polimatroide.*

Demostración. Sean n variables aleatorias discretas X_1, X_2, \dots, X_n . Observe que la función de entropía H se puede interpretar como una función tal que ¹

$$H : 2^E \longrightarrow \mathbb{R},$$

donde $E = \{X_1, X_2, \dots, X_n\}$ y si $J = \{X_{i_1}, X_{i_2}, \dots, X_{i_k}\} \in 2^E$, entonces:

$$H(J) := H(X_{i_1}, \dots, X_{i_k}).$$

Por la proposición 2.1.2 los axiomas polimatroides son equivalentes a (P.1) y (P.4). Se mostrará que la entropía satisface las propiedades (P.1) y (P.4).

(P.1) Se tiene por definición.

(P.4) Sean $A, B, C \subseteq E$, $A = \{Y_1, \dots, Y_p\}$, $B = \{Z_1, \dots, Z_q\}$ y $C = \{W_1, \dots, W_r\}$. Se mostrará que se tiene la propiedad (P.4), es decir:

$$H(A \cup C) + H(B \cup C) \geq H(C) + H(A \cup B \cup C).$$

Observe que la identidad anterior es equivalente a

$$H(A \cup C) + H(B \cup C) - H(C) - H(A \cup B \cup C) \geq 0$$

Realizando las sumas sobre el soporte de las variables aleatorias, es decir:

- $y_i \in S_{Y_i}, 1 \leq i \leq p.$
- $z_j \in S_{Z_j}, 1 \leq j \leq q.$
- $w_k \in S_{W_k}, 1 \leq k \leq r.$

¹Esta definición es dada en [4]

$$H(A \cup C) + H(B \cup C) - H(C) - H(A \cup B \cup C) =$$

$$- \sum p(y_1, \dots, y_p, z_1, \dots, z_q, w_1, \dots, w_r) \log \left(\frac{p(y_1, \dots, y_p, w_1, \dots, w_r)p(z_1, \dots, z_q, w_1, \dots, w_r)}{p(w_1, \dots, w_r)p(y_1, \dots, y_p, z_1, \dots, z_q, w_1, \dots, w_r)} \right).$$

Por la convexidad de la función logaritmo y de lo anterior

$$H(A \cup C) + H(B \cup C) - H(C) - H(A \cup B \cup C) \geq$$

$$- \log \left(\sum p(y_1, \dots, y_p, z_1, \dots, z_q, w_1, \dots, w_r) \frac{p(y_1, \dots, y_p, w_1, \dots, w_r)p(z_1, \dots, z_q, w_1, \dots, w_r)}{p(w_1, \dots, w_r)p(y_1, \dots, y_p, z_1, \dots, z_q, w_1, \dots, w_r)} \right)$$

$$\geq - \log (\sum p(y_1, \dots, y_p | w_1, \dots, w_r) p(z_1, \dots, z_q, w_1, \dots, w_r))$$

$$\geq - \log [(\sum p(y_1, \dots, y_p | w_1, \dots, w_r)) (\sum p(z_1, \dots, z_q, w_1, \dots, w_r))] = 0$$

Así, se tiene la propiedad (P.4) y (P.1); por lo tanto, la función de entropía es una función polimatroide.

□

Con el fin de construir una conexión entre las medidas de la información generadas por una función polimatroide y la Teoría de Conjuntos se introducen algunos conceptos de Teoría de la Medida.

3.1. Nociones de Teoría de la Medida

Definición 3.1.1. *El campo F_n generado por los conjuntos $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$, es la colección de todos los conjuntos que se obtienen al aplicar un número finito de operaciones: $\cup, \cap, -,^c$, entre los conjuntos $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$.*

Definición 3.1.2. *Los átomos de F_n son los conjuntos de la forma $\bigcap_{i=1}^n Y_i$, donde Y_i es \tilde{X}_i ó \tilde{X}_i^c .*

El número de átomos de F_n es 2^n . Además, nótese que los átomos diferentes son disjuntos. El número de conjuntos de F_n es 2^{2^n} , ya que cada conjunto de F_n puede expresarse unívocamente como unión de un subconjunto de átomos.

Se asumirá que la unión de un conjunto vacío de átomos es el conjunto vacío, y que los conjuntos $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ se intersecan unos con otros genéricamente, por lo tanto los átomos de F_n son no vacíos mientras no se indique lo contrario.

Ejemplo 3.1.3. Los conjuntos \tilde{X}_1 y \tilde{X}_2 generan el campo F_2 . Los átomos de F_2 son:

$$\tilde{X}_1 \cap \tilde{X}_2^c \text{ (1), } \tilde{X}_1 \cap \tilde{X}_2 \text{ (2), } \tilde{X}_1^c \cap \tilde{X}_2 \text{ (3), } \tilde{X}_1^c \cap \tilde{X}_2^c \text{ (4)}$$

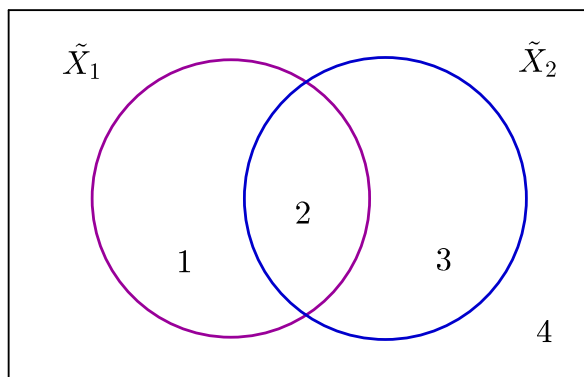


Figura 3-1.: Campo F_2

Hay 16 conjuntos en el campo F_2 , los cuales son uniones de subconjuntos de los átomos descritos anteriormente.

Definición 3.1.4. Una función real μ definida sobre F_n , se denomina una medida con signo, si es aditiva sobre conjuntos disjuntos, es decir, para cualquier par de conjuntos disjuntos $A, B \in F_n$

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Si μ es una medida con signo, $\mu(\emptyset) = 0$. Puesto que para $A \in F_n - \{\emptyset\}$:

$$\begin{aligned} \mu(A \cup \emptyset) &= \mu(A) \\ \mu(A) + \mu(\emptyset) &= \mu(A), \text{ por aditividad de } \mu. \\ \mu(\emptyset) &= 0. \end{aligned}$$

Cualquier medida con signo sobre F_n , está completamente determinada por sus valores sobre los átomos de F_n , debido a la aditividad de μ y a que todo elemento del campo se puede escribir de manera única como una unión de átomos no vacíos.

Ejemplo 3.1.5. Una medida con signo μ sobre F_2 , está completamente determinada por los valores:

$$\mu(\tilde{X}_1 \cap \tilde{X}_2^c), \mu(\tilde{X}_1 \cap \tilde{X}_2), \mu(\tilde{X}_1^c \cap \tilde{X}_2), \mu(\tilde{X}_1^c \cap \tilde{X}_2^c).$$

En particular, para obtener el valor de μ sobre $\tilde{X}_1 \cup \tilde{X}_2$ se tiene lo siguiente:

$$\begin{aligned} \mu(\tilde{X}_1 \cup \tilde{X}_2) &= \mu((\tilde{X}_1 \cap \tilde{X}_2^c) \cup (\tilde{X}_1^c \cap \tilde{X}_2) \cup (\tilde{X}_1 \cap \tilde{X}_2)) \\ &= \mu(\tilde{X}_1 \cap \tilde{X}_2^c) + \mu(\tilde{X}_1^c \cap \tilde{X}_2) + \mu(\tilde{X}_1 \cap \tilde{X}_2) \end{aligned}$$

Ahora, se aclarará que se entenderá por identidad en Teoría de Conjuntos.

Observación:

Según el trabajo de Yeung, una identidad en Teoría de Conjuntos es una igualdad válida para cualquier medida con signo. Por ejemplo, si μ es una medida con signo

$$\mu(A \cup B) = \mu(A - B) + \mu(B - A) + \mu(A \cap B),$$

es una identidad en Teoría de Conjuntos.

3.2. La f -medida para dos conjuntos

En esta sección se establecerá una correspondencia entre las medidas de información generadas por una función polimatroide f y la Teoría de Conjuntos, en el caso de dos conjuntos; bajo ciertas condiciones.

Sean f una función polimatroide, $f : 2^S \rightarrow \mathbb{R}$, y \tilde{X}_1 y \tilde{X}_2 ² conjuntos diferentes no disjuntos, los cuales generan el campo F_2 . El conjunto universal Ω se define como el conjunto $\tilde{X}_1 \cup \tilde{X}_2$ y por lo tanto, el átomo $\tilde{X}_1^c \cap \tilde{X}_2^c$ es el átomo vacío, ya que $\tilde{X}_1^c \cap \tilde{X}_2^c = (\tilde{X}_1 \cup \tilde{X}_2)^c = \Omega^c = \emptyset$.

Las medidas de información generadas por una función polimatroide f , para dos conjuntos \tilde{X}_1 y \tilde{X}_2 , son:

$$f(\tilde{X}_1), f(\tilde{X}_2), f(\tilde{X}_1, \tilde{X}_2), f(\tilde{X}_1 | \tilde{X}_2), f(\tilde{X}_2 | \tilde{X}_1), I(\tilde{X}_1 \wedge \tilde{X}_2).$$

Estas medidas de información inducen una medida con signo μ_f^* sobre F_2 , definida por:

$$\mu_f^*(\tilde{X}_1 - \tilde{X}_2) := f(\tilde{X}_1 | \tilde{X}_2) \tag{3-1}$$

$$\mu_f^*(\tilde{X}_2 - \tilde{X}_1) := f(\tilde{X}_2 | \tilde{X}_1) \tag{3-2}$$

$$\mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) := I(\tilde{X}_1 \wedge \tilde{X}_2) \tag{3-3}$$

²Si $f : 2^S \rightarrow \mathbb{R}$ es de entropía, \tilde{X}_1, \tilde{X}_2 se interpretan como el conjunto de valores que pueden tomar las variables aleatorias discretas X_1 y X_2 , respectivamente. De lo contrario $\tilde{X}_1, \tilde{X}_2 \subseteq S$

Para hallar los valores de μ_f^* en los demás conjuntos de F_2 , se utilizará la aditividad de μ_f^* y las propiedades de una función polimatroide vistas en el Teorema 2.1.5:

Para hallar la medida de $\tilde{X}_1 \cup \tilde{X}_2$

$$\begin{aligned}\mu_f^*(\tilde{X}_1 \cup \tilde{X}_2) &= \mu_f^*(\tilde{X}_1 - \tilde{X}_2) + \mu_f^*(\tilde{X}_2 - \tilde{X}_1) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \\ &= f(\tilde{X}_1|\tilde{X}_2) + f(\tilde{X}_2|\tilde{X}_1) + I(\tilde{X}_1 \wedge \tilde{X}_2)\end{aligned}$$

De lo anterior y (2-7)

$$\mu_f^*(\tilde{X}_1 \cup \tilde{X}_2) = f(\tilde{X}_1, \tilde{X}_2). \quad (3-4)$$

Ahora para calcular la medida en \tilde{X}_1 :

$$\begin{aligned}\mu_f^*(\tilde{X}_1) &= \mu_f^*(\tilde{X}_1 - \tilde{X}_2) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \\ &= f(\tilde{X}_1|\tilde{X}_2) + I(\tilde{X}_1 \wedge \tilde{X}_2)\end{aligned}$$

De lo anterior y (2-7)

$$\mu_f^*(\tilde{X}_1) = f(\tilde{X}_1). \quad (3-5)$$

Y para calcular la medida en \tilde{X}_2

$$\begin{aligned}\mu_f^*(\tilde{X}_2) &= \mu_f^*(\tilde{X}_2 - \tilde{X}_1) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \\ &= f(\tilde{X}_2|\tilde{X}_1) + I(\tilde{X}_1 \wedge \tilde{X}_2)\end{aligned}$$

De lo anterior y (2-7)

$$\mu_f^*(\tilde{X}_2) = f(\tilde{X}_2) \quad (3-6)$$

Así (3-1),(3-2),(3-3),(3-4),(3-5) y (3-6) corresponden a las seis medidas de información generadas por la función polimatroide f para \tilde{X}_1 y \tilde{X}_2 .

En cada una de las igualdades anteriores, el lado derecho y el izquierdo corresponden uno al otro según la siguiente correspondencia de símbolos.

$$\begin{aligned}\tilde{X}_i &\leftrightarrow \tilde{X}_i, \\ f, I &\leftrightarrow \mu_f^*, \\ \wedge &\leftrightarrow \cap, \\ , &\leftrightarrow \cup, \\ | &\leftrightarrow -.\end{aligned}$$

Así, para dos conjuntos \tilde{X}_1, \tilde{X}_2 , las medidas de información generadas por una función polimatroide pueden considerarse como una medida con signo sobre F_2 . Se denominará μ_f^* , la f -medida para los conjuntos \tilde{X}_1 y \tilde{X}_2 .

Observación:

Si la función polimatroide es la función de entropía, es decir, si $f = H$, entonces las medidas de la información generadas por H (medidas de la información Shannon) para dos variables aleatorias X_1 y X_2 son:

$$H(X_1), H(X_2), H(X_1, X_2), H(X_1|X_2), H(X_2|X_1), I(X_1 \wedge X_2).$$

Se denota por \tilde{X}_1 y \tilde{X}_2 los conjuntos generados por las variables aleatorias X_1 y X_2 , respectivamente. Dichos conjuntos deberán satisfacer las condiciones de la construcción realizada, es decir deben ser diferentes y no disjuntos.

De esta manera, la correspondencia establecida entre símbolos es:

$$\begin{aligned} X_i &\leftrightarrow \tilde{X}_i, \\ H, I &\leftrightarrow \mu_H^*, \\ \wedge &\leftrightarrow \cap, \\ , &\leftrightarrow \cup, \\ | &\leftrightarrow -. \end{aligned}$$

Se notará $\mu_H^* := \mu^*$ la I-medida para las variables aleatorias X_1 y X_2 .

Por lo tanto, las medidas de la información Shannon para dos variables aleatorias se pueden ver como una medida con signo sobre F_2 .

La construcción de la f -medida establece un puente entre las medidas de información generadas por funciones polimatroides y la Teoría de Conjuntos. Más aún, la Figura 3-2 representa correctamente las medidas de la información generadas por una función polimatroide f para los conjuntos \tilde{X}_1 y \tilde{X}_2 .

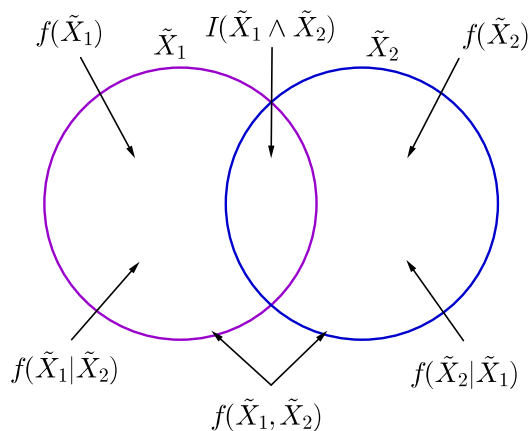


Figura 3-2.: Medidas de la información generadas por f .

La construcción de la f -medida para dos conjuntos lleva a la formulación de las siguientes preguntas:

1. ¿A cualquier identidad de las medidas de información generadas por f , le corresponde una identidad en la Teoría de Conjuntos vía sustitución formal de símbolos?
2. ¿A cualquier identidad de la Teoría de Conjuntos le corresponde una identidad de las medidas de información generadas por f ? Si lo es así, ¿En qué sentido?

Estos cuestionamientos que aquí se hacen para medidas generadas por cualquier función polimatroide, en primer lugar, se hicieron para las medidas de información Shannon. En este contexto, Hu Guo Ding respondió la primera pregunta afirmativamente; aunque dicho resultado es fundamental no resulta ser útil al plantear el descubrimiento de una identidad en Teoría de Conjuntos mediante una identidad en Teoría de la Información Shannon. Debido a que la Teoría de Conjuntos es una teoría más desarrollada y que cuenta con más herramientas, lo que se deseaba era encontrar identidades de Teoría de la Información (Shannon) mediante identidades de Teoría de Conjuntos, es decir, dar respuesta a la segunda pregunta. En *A New Outlook on Shannon's Information Measures*, Yeung logra responder las preguntas planteadas, mediante un solo paso: la construcción de la I -medida. Más aún, mediante esta construcción se encuentra un método para demostrar identidades de la Teoría de la Información, sin necesidad de recurrir a la definición de entropía.

Siguiendo los pasos de Yeung se puede dar respuesta a las dos preguntas realizadas de manera general, mediante la construcción de la f -medida. Así, se visualiza el puente que hay entre las dos teorías.

Por ejemplo, la identidad en Teoría de Conjuntos:

$$\mu(\tilde{X}_1 \cup \tilde{X}_2) = \mu(\tilde{X}_1 - \tilde{X}_2) + \mu(\tilde{X}_2 - \tilde{X}_1) + \mu(\tilde{X}_1 \cap \tilde{X}_2)$$

implica, en particular para μ_f^* :

$$\begin{aligned} \mu_f^*(\tilde{X}_1 \cup \tilde{X}_2) &= \mu_f^*(\tilde{X}_1 - \tilde{X}_2) + \mu_f^*(\tilde{X}_2 - \tilde{X}_1) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \\ &= \left(\mu_f^*(\tilde{X}_1 - \tilde{X}_2) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \right) + \left(\mu_f^*(\tilde{X}_2 - \tilde{X}_1) + \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \right) - \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \\ &= \mu_f^*(\tilde{X}_1) + \mu_f^*(\tilde{X}_2) - \mu_f^*(\tilde{X}_1 \cap \tilde{X}_2) \end{aligned}$$

De la anterior igualdad interpretando μ_f^* se tiene la siguiente identidad de medidas de información generadas por f :

$$f(\tilde{X}_1, \tilde{X}_2) = f(\tilde{X}_1) + f(\tilde{X}_2) - I(\tilde{X}_1 \wedge \tilde{X}_2).$$

Sin embargo, la construcción de la f -medida no permite establecer una correspondencia entre una desigualdad de la Teoría de Información y algún concepto de la Teoría de Conjuntos; a pesar de la existencia de un puente entre ambas teorías, en cuanto a identidades. Esto se debe a que una desigualdad válida para alguna medida con signo, no necesariamente se tiene para toda medida con signo. Lo anterior se ilustrará mediante un ejemplo.

Ejemplo 3.2.1. *Se considera el campo F_2 , generado por un par de conjuntos diferentes A y B , no disjuntos, tales que, el único átomo vacío de F_2 es $A^c \cap B^c$.*

La función cardinal de un conjunto es una medida con signo que satisface la siguiente desigualdad:

$$|A \cup B| \leq |A| + |B| \tag{3-7}$$

Sin embargo, si se considera la medida con signo μ_1 definida sobre los átomos de F_2 de la siguiente forma

$$\begin{aligned} \mu_1(A - B) &= \mu_1(B - A) = 1, \\ \mu_1(A \cap B) &= -1, \end{aligned}$$

por aditividad de μ_1 se tiene que

$$\begin{aligned} \mu_1(A \cup B) &= 1 + (-1) + 1 = 1 > \mu_1(A) + \mu_1(B) \\ &= (1 + (-1)) + (1 + (-1)) \\ &= 0 \end{aligned}$$

Es decir, para μ_1 la desigualdad (3-7) no es válida.

Como conclusión de lo anterior, por medio de la Teoría de Conjuntos es posible estudiar ciertos temas de Teoría de la Información, pero no todos los resultados que se tienen en esta última teoría. Es por ésto, que lo que se realiza en este trabajo es *una mirada a la Teoría de la Información a través de la Teoría de Conjuntos*, más no una equivalencia entre ambas teorías.

A continuación se presentan dos ejemplos de f -medidas, basados en las funciones polimatroides dadas en el capítulo anterior, asociadas a matroides y topologías.

Ejemplo 3.2.2. Se recuerda el grafo $G = (V, E)$ dado en la Figura 3-3, el cual genera el matroide $M(G)$. Y $r : 2^E \rightarrow \mathbb{R}$ la función rango del matroide. Se considera $\tilde{X}_1 = \{1, 2, 3\}$,

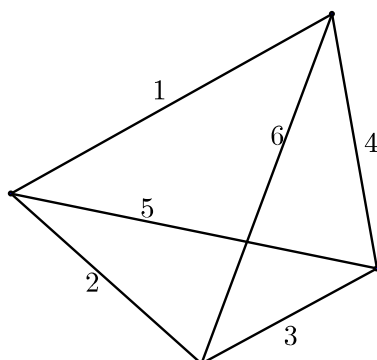


Figura 3-3.: Grafo $G = (V, E)$

$\tilde{X}_2 = \{2, 3, 5\}$ conjuntos que satisfacen las condiciones para la construcción de la r -medida para dos conjuntos. A continuación se grafican los átomos del campo F_2 generado por \tilde{X}_1 y \tilde{X}_2 :

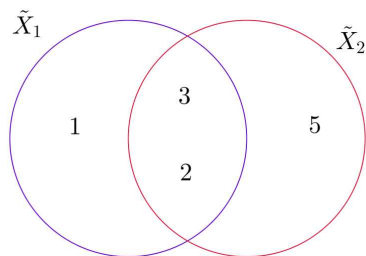


Figura 3-4.:

Para este caso las medidas de información generadas por la función rango del matroide para los conjuntos \tilde{X}_1 y \tilde{X}_2 son:

$$r(\tilde{X}_1), r(\tilde{X}_2), r(\tilde{X}_1, \tilde{X}_2), r(\tilde{X}_1|\tilde{X}_2), r(\tilde{X}_2|\tilde{X}_1), I(\tilde{X}_1 \wedge \tilde{X}_2).$$

Estas medidas se pueden observar como una medida con signo sobre el campo F_2 , para este caso dicha medida se denota como μ_r^* y está dada sobre los átomos de F_2 de la siguiente forma:

$$\begin{aligned}\mu_r^*(\tilde{X}_1 - \tilde{X}_2) &= r(\tilde{X}_1|\tilde{X}_2) \\ \mu_r^*(\tilde{X}_2 - \tilde{X}_1) &= r(\tilde{X}_2|\tilde{X}_1) \\ \mu_r^*(\tilde{X}_2 \cap \tilde{X}_1) &= I(\tilde{X}_1 \wedge \tilde{X}_2).\end{aligned}$$

Ahora, se dará un ejemplo basado en la función polimatroide $f_{\mathbb{Z}}$ definida en el capítulo anterior.

Ejemplo 3.2.3. Se recuerda que en el ejemplo 2.1.14, se estableció un espacio topológico (E, τ) , sobre $E = \{a, b, c\}$. Además, se definió la función polimatroide $f_{\mathbb{Z}}$, de manera que, si $I \subseteq E$, $f_{\mathbb{Z}}(I)$ es el número de cerrados que no contienen a I .

Sean $\tilde{X}_1 = \{a, b\}$, $\tilde{X}_2 = \{b, c\}$, subconjuntos de E que satisfacen las condiciones para la construcción de la $f_{\mathbb{Z}}$ -medida. Los átomos de F_2 generado por estos dos conjuntos se muestran a continuación:

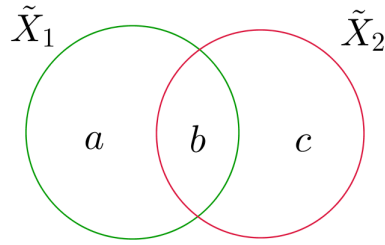


Figura 3-5.:

Las medidas generadas por la función $f_{\mathbb{Z}}$ para los conjuntos \tilde{X}_1 y \tilde{X}_2 son:

$$f_{\mathbb{Z}}(\tilde{X}_1), f_{\mathbb{Z}}(\tilde{X}_2), f_{\mathbb{Z}}(\tilde{X}_1, \tilde{X}_2), f_{\mathbb{Z}}(\tilde{X}_1|\tilde{X}_2), f_{\mathbb{Z}}(\tilde{X}_2|\tilde{X}_1), I(\tilde{X}_1 \wedge \tilde{X}_2).$$

Estas medidas se pueden observar como los valores de una medida con signo sobre algunos elementos del campo F_2 . Dicha medida se denota como $\mu_{f_{\mathbb{Z}}}^*$ y está definida sobre los átomos de la siguiente forma:

$$\begin{aligned}\mu_{f_{\mathbb{Z}}}^*(\tilde{X}_1 - \tilde{X}_2) &= f_{\mathbb{Z}}(\tilde{X}_1|\tilde{X}_2) \\ \mu_{f_{\mathbb{Z}}}^*(\tilde{X}_2 - \tilde{X}_1) &= f_{\mathbb{Z}}(\tilde{X}_2|\tilde{X}_1) \\ \mu_{f_{\mathbb{Z}}}^*(\tilde{X}_2 \cap \tilde{X}_1) &= I(\tilde{X}_1 \wedge \tilde{X}_2).\end{aligned}$$

3.3. Construcción general de la f -medida

En la sección anterior se construyó la f -medida para dos conjuntos ³. Ahora se construirá la f medida para cualquier número n de conjuntos ($n > 2$), siendo f una función polimatroide.

Sean $f : 2^S \rightarrow \mathbb{R}$ una función polimatroide y $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ ⁴ conjuntos tales que los conjuntos de la forma $\bigcap_{i=1}^n Y_i$ donde Y_i es \tilde{X}_i ó \tilde{X}_i^c sean no vacíos. Sea $N_n = \{1, 2, \dots, n\}$ se define el conjunto universal

$$\Omega = \bigcup_{i \in N_n} \tilde{X}_i.$$

Se usará F_n para denotar el campo generado por $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$. Así, el conjunto $A_0 = \bigcap_{i \in N_n} \tilde{X}_i^c$ es el átomo vacío de F_n , debido a las leyes de De Morgan:

$$A_0 = \bigcap_{i \in N_n} \tilde{X}_i^c = \left(\bigcup_{i \in N_n} \tilde{X}_i \right)^c = \Omega^c = \emptyset.$$

Observación:

Si $f = H$, entonces se obtiene la I -medida para n variables aleatorias X_1, X_2, \dots, X_n . Para este caso los conjuntos $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ de la construcción anterior son los subconjuntos de los naturales determinados por X_1, X_2, \dots, X_n , respectivamente.

Por construcción, todos los átomos de F_n diferentes de A_0 son no vacíos. Si \mathfrak{A} es el conjunto de átomos no vacíos de F_n , entonces $|\mathfrak{A}| = 2^n - 1$. Además, cualquier medida con signo μ sobre F_n , está completamente determinada por los valores de μ sobre los átomos no vacíos.

Con el fin de construir la f -medida se presentarán los siguientes resultados.

Lema 3.3.1. *Si μ es una función aditiva, entonces*

$$\mu \left(\bigcap_{k=1}^n A_k - B \right) = \sum_{1 \leq i \leq n} \mu(A_i - B) - \sum_{1 \leq i < j \leq n} \mu((A_i \cup A_j) - B) + \dots + (-1)^{n+1} \mu((A_1 \cup \dots \cup A_n) - B)$$

Demostración. Ver Anexo A. □

³Recuerde que si $f = H$ se estableció la I -medida para dos variables aleatorias.

⁴Si f es una función de entropía, $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$, denotarán los conjuntos generados por las variables aleatorias X_1, X_2, \dots, X_n , respectivamente. En caso contrario, $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n \subseteq S$.

Lema 3.3.2.

$$\mu(A \cap B - C) = \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C)$$

Demostración. Por el lema anterior:

$$\begin{aligned} \mu(A \cap B - C) &= \mu(A - C) + \mu(B - C) - \mu(A \cup B - C) \\ &= (\mu(A \cup C) - \mu(C)) + (\mu(B \cup C) - \mu(C)) - (\mu(A \cup B \cup C) - \mu(C)) \\ &= \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C) \end{aligned}$$

□

Lema 3.3.3.

$$I(X \wedge Y|Z) = f(X, Z) + f(Y, Z) - f(X, Y, Z) - f(Z)$$

Demostración. Esta igualdad corresponde a la igualdad (2-8) ya demostrada. □

Para simplificar la escritura, si α es un conjunto no vacío de N_n , se notará $\tilde{X}_\alpha := \bigcup_{i \in \alpha} \tilde{X}_i$ y si $\alpha = \emptyset$, $\tilde{X}_\alpha := \emptyset$.

Observación:

Si $f = H$, X_α denota el vector aleatorio $(X_i|i \in \alpha)$, para $\alpha \neq \emptyset$.

Teorema 3.3.4. *Sea*

$$\mathfrak{B} = \{\tilde{X}_\alpha : \alpha \text{ es un conjunto no vacío de } N_n\}.$$

Entonces, una medida con signo μ sobre F_n , está completamente determinada por $\{\mu(B) : B \in \mathfrak{B}\}$.

Demostración. El número de elementos en \mathfrak{B} , es el número de subconjuntos no vacíos de N_n , el cual es $2^n - 1$. Si \mathfrak{A} es el conjunto de átomos no vacíos del campo F_n , por construcción $|\mathfrak{A}| = 2^n - 1$. Así, $|\mathfrak{B}| = |\mathfrak{A}| = 2^n - 1 =: k$ y $\mathfrak{A} = \{A_1, \dots, A_k\}$, $\mathfrak{B} = \{B_1, \dots, B_k\}$.

Se definen los vectores $u := [\mu(A_1), \dots, \mu(A_k)]^T$ y $h := [\mu(B_1), \dots, \mu(B_k)]^T$.

Por otro lado como $B \in \mathfrak{B}$, entonces $B \in F_n$ y por la aditividad de μ , existen $C_{ij} \in \mathbb{R}$ para $1 \leq i, j \leq k$, tales que:

$$\mu(B_i) = \sum_{j=1}^k C_{ij} \mu(A_j),$$

así,

$$h = C_n u \quad (3-8)$$

donde C_n es una matriz de tamaño $k \times k$ y es única.

Ahora, sea $\bigcap_{i=1}^n Y_i$ un átomo no vacío, por lo tanto existe i tal que $Y_i = \tilde{X}_i$, por lo tanto:

$$\begin{aligned} \bigcap_{i=1}^n Y_i &= \bigcap_{i:Y_i=\tilde{X}_i} \tilde{X}_i \cap \bigcap_{j:Y_j=\tilde{X}_j^c} \tilde{X}_j^c \\ &= \bigcap_{i:Y_i=\tilde{X}_i} \tilde{X}_i \cap \left(\bigcup_{j:Y_j=\tilde{X}_j^c} \tilde{X}_j \right)^c \\ &= \bigcap_{k=1}^m \tilde{X}_k - \bigcup_{l=1}^p \tilde{X}_l \end{aligned}$$

Aplicando el Lema 3.3.1 y que $\mu(A - B) = \mu(A \cup B) - \mu(B)$, se tiene:

$$\begin{aligned} \mu \left(\bigcap_{i=1}^n Y_i \right) &= \mu \left(\bigcap_{k=1}^m \tilde{X}_k - \bigcup_{l=1}^p \tilde{X}_l \right) \\ &= \sum_{1 \leq i \leq m} \mu \left(\tilde{X}_i \cup \bigcup_{l=1}^p \tilde{X}_l \right) - \mu \left(\bigcup_{l=1}^p \tilde{X}_l \right) \\ &\quad \sum_{1 \leq i < j \leq m} \mu \left(\tilde{X}_i \cup \tilde{X}_j \cup \bigcup_{l=1}^p \tilde{X}_l \right) - \mu \left(\bigcup_{l=1}^p \tilde{X}_l \right) \\ &\quad + \dots + (-1)^{m+1} \left(\mu \left(\tilde{X}_1 \cup \dots \cup \tilde{X}_m \cup \bigcup_{l=1}^p \tilde{X}_l \right) - \mu \left(\bigcup_{l=1}^p \tilde{X}_l \right) \right) \end{aligned} \quad (3-9)$$

Por (3-9), existe una matriz D_n de tamaño $k \times k$ tal que:

$$u = D_n h.$$

De lo anterior y de (3-8):

$$u = D_n h = D_n C_n u.$$

Si se da el valor para $u = e_i$ para $1 \leq i \leq k$, entonces $D_n C_n = I$. Debido a que C_n es única, D_n es única al ser la inversa de C_n .

Entonces, $\{\mu(A_i), A_i \in \mathfrak{A}\}$ está unívocamente determinado por $\{\mu(B), B \in \mathfrak{B}\}$. \square

Definición 3.3.5. Se define la f -medida μ_f^* , sobre F_n , de la siguiente forma:

$$\mu_f^*(\tilde{X}_\alpha) := f(\tilde{X}_\alpha), \quad \alpha \neq \emptyset.$$

La medida μ_f^* está bien definida debido al Teorema 3.3.4.

Observación:

Si $f = H$, $\mu_H^* =: \mu^*$ se conoce como la I -medida sobre F_n y estará dada por:

$$\mu^*(\tilde{X}_\alpha) := H(X_\alpha), \quad \alpha \neq \emptyset.$$

Por otro lado, para que μ_f^* tenga sentido debe ser consistente con las medidas de información generadas por f , es decir, dichas medidas deben obtenerse a partir de μ_f^* vía sustitución de símbolos.

Para que lo anterior se satisfaga, se debe cumplir:

$$\mu_f^*(\tilde{X}_\alpha \cap \tilde{X}_{\alpha'} - \tilde{X}_{\alpha''}) = I(\tilde{X}_\alpha \wedge \tilde{X}_{\alpha'} | \tilde{X}_{\alpha''}) \quad (3-10)$$

siendo $\alpha, \alpha', \alpha''$ subconjuntos de N_n , con α y α' no vacíos.⁵

Se verá que (3-10) cubre los cuatro casos de medidas de información y por lo tanto es condición necesaria y suficiente para que μ_f^* sea consistente con todas las medidas de información generadas por f .

- Cuando $\alpha'' = \emptyset$, (3-10) se convierte en

$$\mu_f^*(\tilde{X}_\alpha \cap \tilde{X}_{\alpha'}) = I(\tilde{X}_\alpha \wedge \tilde{X}_{\alpha'}).$$

- Cuando $\alpha = \alpha'$, (3-10) se convierte en

$$\begin{aligned} \mu_f^*(\tilde{X}_\alpha - \tilde{X}_{\alpha''}) &= I(\tilde{X}_\alpha \wedge \tilde{X}_\alpha | \tilde{X}_{\alpha''}) \\ &= f(\tilde{X}_\alpha, \tilde{X}_{\alpha''}) + f(\tilde{X}_\alpha, \tilde{X}_{\alpha''}) - f(\tilde{X}_\alpha, \tilde{X}_\alpha, \tilde{X}_{\alpha''}) - f(\tilde{X}_{\alpha''}) \text{ por lema 3.3.3} \\ &= f(\tilde{X}_\alpha | \tilde{X}_{\alpha''}) \end{aligned}$$

⁵Observe que si f es una función de entropía, se debe tener: $\mu_f^*(\tilde{X}_\alpha \cap \tilde{X}_{\alpha'} - \tilde{X}_{\alpha''}) = I(X_\alpha \wedge X_{\alpha'} | X_{\alpha''})$

- Cuando $\alpha = \alpha'$ y $\alpha'' = \emptyset$, (3-10) se convierte en

$$\mu_f^*(\tilde{X}_\alpha) = f(\tilde{X}_\alpha).$$

El siguiente teorema muestra que la condición (3-10) se cumple.

Teorema 3.3.6. μ_f^* es una medida con signo sobre F_n , consistente con todas las medidas de la información generadas por la función polimatroide f .

Demostración. Se verá que se tiene (3-10).

$$\begin{aligned} \mu_f^*(\tilde{X}_\alpha \cap \tilde{X}_{\alpha'} - \tilde{X}_{\alpha''}) &= \mu_f^*(\tilde{X}_\alpha \cup \tilde{X}_{\alpha''}) + \mu_f^*(\tilde{X}_{\alpha'} \cup \tilde{X}_{\alpha''}) \\ &\quad - \mu_f^*(\tilde{X}_\alpha \cup \tilde{X}_{\alpha'} \cup \tilde{X}_{\alpha''}) - \mu_f^*(\tilde{X}_{\alpha''}) \quad \text{lema 3.3.2} \\ &= \mu_f^*(\tilde{X}_{\alpha \cup \alpha''}) + \mu_f^*(\tilde{X}_{\alpha' \cup \alpha''}) - \mu_f^*(\tilde{X}_{\alpha \cup \alpha' \cup \alpha''}) - \mu_f^*(\tilde{X}_{\alpha''}) \\ &= f(\tilde{X}_\alpha, \tilde{X}_{\alpha'}) + f(\tilde{X}_{\alpha'}, \tilde{X}_{\alpha''}) - f(\tilde{X}_\alpha, \tilde{X}_{\alpha'}, \tilde{X}_{\alpha''}) \\ &\quad - f(\tilde{X}_{\alpha''}) \\ &= I(\tilde{X}_\alpha \wedge \tilde{X}_{\alpha'} | \tilde{X}_{\alpha''}) \quad \text{Lema 3.3.3.} \end{aligned}$$

□

En el desarrollo anterior se creó el concepto de f -medida para cualquier función polimatroide dadas ciertas condiciones, es decir, se puede generalizar la construcción de la I-medida hecha por Yeung. Y así, se tiene un puente entre las medidas de información generadas por una función polimatroide y la Teoría de Conjuntos.

Resulta ser una construcción más general debido a lo mostrado en “*A First Course in Information Theory*”. Allí se definen los conjuntos Γ_n^* y Γ_n ($n \geq 2$), el primer conjunto describe los vectores entrópicos y el segundo los vectores que satisfacen las desigualdades básicas de la información Shannon; luego, se demuestra que en general, estos dos conjuntos no son iguales.

A continuación se describirán brevemente dichos conjuntos.

- Sean $N_n = \{1, 2, \dots, n\}$ y $\theta = \{X_i | i \in N_n\}$ una colección de n variables aleatorias. Asociadas a θ , hay $2^n - 1$ entropías conjuntas.

Si $\alpha \subseteq N_n$ no vacío, entonces

$$H_\theta(\alpha) := H(X_\alpha),$$

para un θ fijo. Es por esto que H_θ recibe el nombre de función de entropía de θ .

Se dirá que $\mathbf{h} \in \mathbb{R}^{2^n - 1}$ es un vector entrópico, si es igual a la función H_θ , para una

colección de n variables aleatorias.

Por ejemplo, para $n = 3$, un vector entrópico es de la forma

$$h = [h_1 \ h_2 \ h_3 \ h_{12} \ h_{13} \ h_{23} \ h_{123}]^T,$$

donde

$$h_1 = H(X_1)$$

$$h_2 = H(X_2)$$

$$h_3 = H(X_3)$$

$$h_{12} = H(X_1, X_2)$$

$$h_{13} = H(X_1, X_3)$$

$$h_{23} = H(X_2, X_3)$$

$$h_{123} = H(X_1, X_2, X_3)$$

para una colección $\theta = \{X_1, X_2, X_3\}$ de variables aleatorias.

Con base en lo anterior, se define el conjunto:

$$\Gamma_n^* = \{\mathbf{h} \in \mathbb{R}^{2^n-1} | \mathbf{h} \text{ es entrópico}\}$$

- Ahora, se reconocerán ciertos hechos para introducir la definición de Γ_n .

Primero se recuerda que cualquier medida de información Shannon de las variables aleatorias X_1, X_2, \dots, X_n se pueden expresar en términos de las siguientes formas elementales:

$$H(X_i | X_{N_n - \{i\}}), i \in N_n$$

$$I(X_i \wedge X_j | X_K), i \neq j, K \subseteq N_n - \{i, j\}$$

Se llamarán desigualdades básicas de la información Shannon a las $m = n + \binom{n}{2} 2^{n-2}$ desigualdades:

$$H(X_i | X_{N_n - \{i\}}) \geq 0, i \in N_n,$$

$$I(X_i \wedge X_j | X_K) \geq 0, i \neq j, K \subseteq N_n - \{i, j\}.$$

Debido a que cualquier medida de la información Shannon se puede escribir como combinación lineal de las $k = 2^n - 1$ entropías conjuntas, el conjunto de las m desigualdades básicas se pueden denotar por

$$Gh \geq 0,$$

donde G es una matriz de tamaño $m \times k$.

Así se define el conjunto

$$\Gamma_n = \{h \in \mathbb{R}^{2^n-1} | Gh \geq 0\}.$$

En [12] se demuestra que $\Gamma_2^* = \Gamma_2$. Sin embargo, para $n \geq 3$, $\Gamma_n^* \subsetneq \Gamma_n$, lo cual significa que existen vectores que satisfacen las desigualdades básicas pero que no son entrópicos. Es decir, existen vectores que satisfacen las desigualdades polimatroides, pero no son entrópicos. Por lo cual la construcción de la f -medida es más general que la de la I -medida.

Ejemplo 3.3.7. Se considera el grafo $G_1 = (V_1, E_1)$, donde V_1 es el conjunto de vértices y E_1 es el conjunto de aristas, mostrado a continuación:

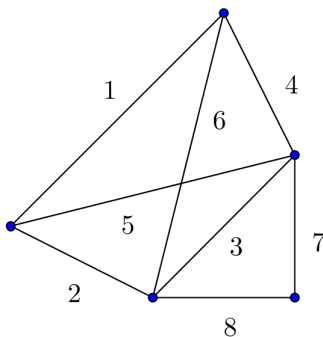


Figura 3-6.: Grafo $G_1 = (V_1, E_1)$

Sean:

$$\tilde{X}_1 = \{2, 3, 4, 7, 8\}$$

$$\tilde{X}_2 = \{2, 3, 5, 6, 7\}$$

$$\tilde{X}_3 = \{1, 2, 4, 5, 7\}$$

Los cuales generan el campo F_3 y cuyos átomos se muestra a continuación

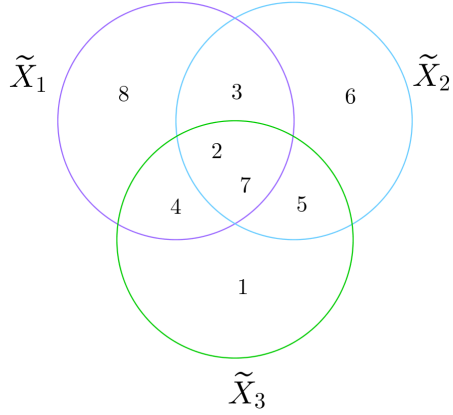


Figura 3-7.:

Debido a que ningún átomo diferente del vacío es vacío, las medidas de información generadas por la función rango del matroide, son los valores de la medida con signo sobre algunos conjuntos de F_3 . Así, por el Teorema 3.3.6, para los átomos de F_3 , la r -medida esta dada a continuación

$$\begin{aligned}
 \mu_r^*({8}) &= r(\tilde{X}_1 | \tilde{X}_2, \tilde{X}_3) \\
 \mu_r^*({3}) &= I(\tilde{X}_1 \wedge \tilde{X}_2 | \tilde{X}_3) \\
 \mu_r^*({6}) &= r(\tilde{X}_2 | \tilde{X}_1, \tilde{X}_3) \\
 \mu_r^*({2, 7}) &= I(\tilde{X}_1 \wedge \tilde{X}_2 \wedge \tilde{X}_3) \\
 \mu_r^*({5}) &= I(\tilde{X}_2 \wedge \tilde{X}_3 | \tilde{X}_1) \\
 \mu_r^*({1}) &= r(\tilde{X}_3 | \tilde{X}_1, \tilde{X}_2) \\
 \mu_r^*({4}) &= I(\tilde{X}_1 \wedge \tilde{X}_3 | \tilde{X}_2)
 \end{aligned}$$

El último objetivo de esta sección, es probar una condición necesaria para que ciertas funciones (las cuales resultan ser funciones polimatroides), pueda ser interpretadas como funciones entrópicas sobre las uniones. Con el fin de cumplir este propósito, se mostrarán los Teoremas 3.3.8 y 3.3.9.

Teorema 3.3.8. *La entropía de una variable aleatoria puede tomar cualquier valor no negativo.* ⁶

Demostración. Sea X una variable aleatoria con soporte $\tilde{X} = \{x_1, x_2, \dots, x_n\}$. Si X tiene una distribución uniforme sobre \tilde{X} , entonces:

$$\begin{aligned} H(X) &= - \sum_{i=1}^n p(x_i) \log p(x_i) \\ &= - \sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = - \log \frac{1}{n} = \log n. \end{aligned}$$

Si la variable aleatoria X tiene la distribución de probabilidad:

$$p(x_1) = 1 \quad p(x_j) = 0 \quad j \neq 1,$$

entonces

$$H(X) = 0.$$

Al ser la función de entropía una función continua, entonces por el teorema del valor intermedio para un valor a , $0 < a < \log n$, existe una distribución de X tal que:

$$H(X) = a.$$

Así $H(X)$ puede tomar cualquier valor haciendo $|\tilde{X}|$ suficientemente grande. \square

Teorema 3.3.9. *Si no hay restricción sobre las variables aleatorias X_1, \dots, X_n , entonces μ^* puede tomar cualquier valor no negativo sobre los átomos no vacíos de F_n .* ⁷

Demostración. Se recuerda que el campo F_n es generado por los conjuntos $\tilde{X}_1, \dots, \tilde{X}_n$.

Se construirá una I -medida μ^* que pueda tomar cualquier conjunto de valores no negativos sobre los átomos no vacíos \mathfrak{A} de F_n .

Sea Y_A , $A \in \mathfrak{A}$ variables mutuamente independientes. Se definen las variables aleatorias

$$X_i = (Y_A : A \in \mathfrak{A}, A \subseteq \tilde{X}_i).$$

Se determina la I -medida μ^* , para X_1, \dots, X_n . Al ser las variables Y_A mutuamente independientes, si $\emptyset \neq G \subseteq N_n$ se tiene:

$$H(X_G) = \sum_{A \in \mathfrak{A}: A \subseteq \tilde{X}_G} H(Y_A). \quad (3-9)$$

⁶Resultado demostrado en [12], pág 29.

⁷Resultado probado en [12], pág 107

Por otro lado,

$$H(X_G) = \mu^*(\tilde{X}_G) = \sum_{A \in \mathfrak{A}: A \subseteq \tilde{X}_G} \mu^*(A). \quad (3-10)$$

Igualando (3-9) y (3-10),

$$\sum_{A \in \mathfrak{A}: A \subseteq \tilde{X}_G} H(Y_A) = \sum_{A \in \mathfrak{A}: A \subseteq \tilde{X}_G} \mu^*(A). \quad (3-11)$$

La identidad (3-11) resulta ser válida para todos los subconjuntos no vacíos G de N_n , tomando la asignación:

$$\mu^*(A) = H(Y_A),$$

para todo $A \in \mathfrak{A}$. Por la unicidad de μ^* , la anterior asignación es la única posibilidad y por el Teorema 3.3.8, $H(Y_A)$ puede tomar cualquier valor no negativo. Así se tiene lo pedido. \square

Teorema 3.3.10. *Sea $f : 2^{\mathfrak{D}} \rightarrow \mathbb{R}$, donde \mathfrak{D} es el conjunto de átomos no vacíos del campo generado por los conjuntos $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n$ y $|\mathfrak{D}| = 2^n - 1$.*

Si f es positiva sobre los elementos de \mathfrak{D} y aditiva sobre conjuntos disyuntos, entonces existe una colección de n variables aleatorias X_1, \dots, X_n tales que

$$f(\tilde{Z}_\alpha) = H(X_\alpha).$$

Demostración. Sea la función aditiva $f : 2^{\mathfrak{D}} \rightarrow \mathbb{R}$. Por el teorema anterior, si no hay restricción sobre las variables aleatorias X_1, \dots, X_n , entonces la I -medida puede tomar cualquier valor no negativo sobre los átomos de F_n . Para cada $A_i \in \mathfrak{A}$, se puede asignar de manera conveniente un $B_i \in \mathfrak{D}$, de modo que:

$$\mu^*(A_i) = H(Y_{A_i}) = f(B_i), \quad 1 \leq i \leq 2^n - 1. \quad (3-12)$$

Si

$$\tilde{Z}_\alpha = \cup_{B \in \mathfrak{D}: B \subseteq \tilde{Z}_\alpha} B,$$

al ser f aditiva y de (3-12)

$$\begin{aligned} f(\tilde{Z}_\alpha) &= f(\cup_{B \in \mathfrak{D}: B \subseteq \tilde{Z}_\alpha} B) \\ &= \sum_{B \in \mathfrak{D}: B \subseteq \tilde{Z}_\alpha} f(B) \\ &= \sum_{A \in \mathfrak{A}: B \subseteq \tilde{X}_\alpha} \mu^*(A) \\ &= \sum_{A \in \mathfrak{A}: B \subseteq \tilde{X}_\alpha} H(Y_A) = H(X_\alpha). \end{aligned}$$

\square

Observe que si la función satisface las hipótesis del teorema anterior, entonces es una función polimatroide.

Si bien, la función de entropía es una función polimatroide, del teorema anterior surge la siguiente pregunta, ¿Toda función de entropía es una función polimatroide aditiva?. La anterior pregunta, se responde negativamente, ya que existen funciones de entropía que no son aditivas. Esto se mostrará en el siguiente ejemplo.

Ejemplo 3.3.11. Sean X_1 y X_2 variables aleatorias, de manera que sus distribuciones de probabilidades conjuntas están dadas en la siguiente tabla:

| | | |
|--------------|-----|-----|
| (x_i, y_i) | 0 | 1 |
| 0 | 1/4 | 0 |
| 1 | 1/4 | 1/4 |
| 2 | 0 | 1/4 |

Además, de las anteriores probabilidades conjuntas se deducen las distribuciones tanto de X_1 como de X_2 :

| | | | |
|----------------|-----|-----|-----|
| x_i | 0 | 1 | 2 |
| $P(X_1 = x_i)$ | 1/4 | 1/2 | 1/4 |

| | | |
|----------------|-----|-----|
| y_i | 0 | 1 |
| $P(X_2 = y_i)$ | 1/2 | 1/2 |

Se recuerda que la función de entropía para las variables aleatorias X_1 y X_2 es una función $H : 2^{\{X_1, X_2\}} \rightarrow \mathbb{R}$. Veremos que a pesar de que $\{X_1\} \cap \{X_2\} = \emptyset$

$$H(\{X_1\} \uplus \{X_2\}) \neq H(\{X_1\}) + H(\{X_2\}).$$

Para ésto, observemos lo siguiente:

$$\begin{aligned} H(\{X_1\} \uplus \{X_2\}) &:= H(X_1, X_2) = - \sum p(x_i, y_j) \log_2 p(x_i, y_j) \\ &= - \frac{4}{4} \log_2 \left(\frac{1}{4} \right) = - \log_2 \left(\frac{1}{4} \right) = 2 \end{aligned}$$

por otro lado

$$\begin{aligned} H(\{X_1\}) &:= H(X_1) \\ &= - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \\ &= \frac{2}{4} + \frac{1}{2} + \frac{2}{4} = \frac{3}{2} \end{aligned}$$

y

$$\begin{aligned} H(\{X_2\}) &:= H(X_2) \\ &= - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \\ &= \frac{2}{4} + \frac{2}{4} = 1 \end{aligned}$$

Así

$$H(\{X_1\}) + H(\{X_2\}) = H(X_1) + H(X_2) = \frac{5}{2},$$

por lo tanto, $H(\{X_1\} \uplus \{X_2\}) \neq H(\{X_1\}) + H(\{X_2\})$ es decir esta función de entropía no es aditiva.

3.4. Elementos del campo F_n y su información

El Teorema 3.3.4 muestra que para cualquier elemento del campo F_n , el valor $\mu(B)$ (siendo μ cualquier medida con signo) se puede determinar mediante las medidas sobre los átomos no vacíos o mediante las medidas sobre los elementos del conjunto

$$\mathfrak{B} = \{\tilde{X}_\alpha | \alpha \subseteq N_n, \alpha \neq \emptyset\}.$$

Este hecho motiva la siguiente pregunta, ¿Si $\mathfrak{C} \subseteq F_n$ que información se puede obtener a partir de \mathfrak{C} ?, en este contexto, se entenderá información como la imagen de cualquier medida con signo sobre el campo F_n . El objetivo de esta sección es responder a la pregunta anterior, para esto se introducirán ciertas

3.4.1. Conjunto de Información

Definición 3.4.1. Sea $\mathfrak{C} \subseteq F_n$. Se definen los elementos de información de \mathfrak{C} :

1. Todo elemento de \mathfrak{C} es un elemento de información de \mathfrak{C} .
2. Si A y B son elementos de información de \mathfrak{C} y $A \cap B = \emptyset$, entonces $A \cup B$ es un elemento de información de \mathfrak{C} .
3. Si A y B son elementos de información de \mathfrak{C} y $A \subseteq B$ entonces $B - A$ es un elemento de información de \mathfrak{C} .

Ejemplo 3.4.2. Sean A y B conjuntos diferentes no disjuntos y $\Omega = A \cup B$. A, B generan el campo F_2 . Si $B_1 = \{A \cup B, A \triangle B\}$, $A \cap B$ es un elemento de información de B_1 ya que:

- i) $A \cup B, A \triangle B$ son elementos de información de B_1 , por 1.
- ii) $A \cap B = (A \cup B) - (A \triangle B)$ es un elemento de información de B_1 por i) y 3.

Mientras que A no es un elemento de información de B_1 , puesto que:

1. $A \notin B_1$.
2. A no se puede expresar como unión disjunta de elementos de B_1 .
3. A no se puede expresar como diferencia de dos conjuntos de B_1 .

Definición 3.4.3.

- Dado $\mathfrak{C} \subseteq F_n$ no vacío. El conjunto de información generado por \mathfrak{C} es $\langle \mathfrak{C} \rangle$ donde:

$$\langle \mathfrak{C} \rangle = \{C \in F_n \mid C \text{ es un elemento de información de } \mathfrak{C}\}.$$

- $\langle \emptyset \rangle := \{\emptyset\}$

Observación:

El conjunto \emptyset es un elemento de información para todo $\mathfrak{C} \subseteq F_n$.

Ejemplo 3.4.4. Para B_1 del ejemplo 3.4.2, $\langle B_1 \rangle = \{\emptyset, A \cup B, A \triangle B, A \cap B\}$.

Definición 3.4.5. Sea $\mathfrak{C} \subseteq F_n$. Un conjunto de información básico C_b de \mathfrak{C} , es un subconjunto de $\langle \mathfrak{C} \rangle$ tal que

1. $\langle C_b \rangle = \langle \mathfrak{C} \rangle$.
2. Para todo $C \subsetneq C_b$, $\langle C \rangle \subsetneq \langle C_b \rangle$.

Observación:

1. No todo conjunto es un conjunto de información básico de sí mismo.

$B_1 = \{A \cup B, A \triangle B\}$ es un conjunto de información básico de B_1 , sin embargo, este hecho no se tiene en general. Por ejemplo, sobre el campo F_2 , generado por A y B , sea

$$B_2 = \{\emptyset, A \cup B, A \triangle B, A \cap B\}.$$

Se tiene que $\langle B_2 \rangle = B_2$, sin embargo B_2 no es un conjunto de información básico de B_2 puesto que existe $B_3 = \{A \triangle B, A \cap B\} \subsetneq B_2$, el cual satisface $\langle B_3 \rangle = B_2 = \langle B_2 \rangle$.

2. El conjunto de información básico de un conjunto dado no es único. Consideremos A, B, C conjuntos que se intersecan y los cuales generan el campo F_3 . Sea

$$B_4 = \{A \Delta B, A \cup B, A \Delta C, A \cap C\}.$$

$\langle B_4 \rangle = \{\emptyset, A \Delta B, A \cup B, A \cap B, A \Delta C, A \cup C, A \cap C\}$. B_4 es un conjunto de información básico de B_4 , pero este no es único; nótese que $B_5 = \{A \Delta B, A \cap B, A \Delta C, A \cap C\}$ es un conjunto de información básico de B_4 y $B_4 \neq B_5$.

A pesar de que $B_4 \neq B_5$ se tiene que $|B_4| = |B_5|$, hecho que más adelante se generalizará.

Definición 3.4.6. Una medida con signo μ sobre $\mathfrak{D} \subseteq F_n$ es una función real $\mu : \mathfrak{D} \rightarrow \mathbb{R}$ tal que para todo par $A, B \in \mathfrak{D}$ con $A \cap B = \emptyset$ se satisface

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Proposición 3.4.7. Sea $\mathfrak{C} \subseteq F_n$. Si C_b es un conjunto de información básico de \mathfrak{C} , entonces una medida con signo μ sobre $\langle \mathfrak{C} \rangle$ está determinada por $\{\mu(X) | X \in C_b\}$. Es decir, se conoce la medida de cualquier elemento de $\langle \mathfrak{C} \rangle$ a partir de $\{\mu(X) | X \in C_b\}$.

Demostración. Sea $B \in \langle \mathfrak{C} \rangle$ y μ una medida con signo sobre $\langle \mathfrak{C} \rangle$. Como C_b es un conjunto de información básico de \mathfrak{C} , $\langle C_b \rangle = \langle \mathfrak{C} \rangle$. Así, $B \in \langle C_b \rangle$, es decir B es un elemento de información de C_b .

Se consideran inicialmente tres casos básicos

i) Si $B \in C_b$, entonces $\mu(B) \in \{\mu(X) | X \in C_b\}$.

ii) Si $B = A \cup C$ con $A, C \in C_b$ y $A \cap C = \emptyset$, entonces

$$\mu(B) = \mu(A) + \mu(C).$$

Es decir, $\mu(B)$ se obtiene a partir de $\{\mu(X) | X \in C_b\}$.

iii) Si $B = A - C$ con $A, C \in C_b$ y $C \subseteq A$, entonces $B \cap C = \emptyset$ y $B \cup C = A$. Por lo tanto

$$\begin{aligned} \mu(B \cup C) &= \mu(A) \\ \mu(B) + \mu(C) &= \mu(A) \\ \mu(B) &= \mu(A) - \mu(C). \end{aligned}$$

Así, $\mu(B)$ se obtiene a partir de $\{\mu(X) | X \in C_b\}$.

iv) Supongamos ahora que $B \in \langle C_b \rangle$ y no cumple i), ii) o iii). Al ser B un elemento de información de C_b se debe tener alguno de los siguientes casos:

a) B es unión disjunta de elementos que satisfacen alguna de las condiciones i), ii), iii), es decir

$$B = \uplus_{j=1}^r B_j,$$

donde B_j cumple i), ii) o iii), $1 \leq j \leq r$. Así,

$$\mu(B) = \mu(\uplus_{j=1}^r B_j) = \sum_{j=1}^r \mu(B_j). \quad (\diamond)$$

Por i), ii) y iii), cada $\mu(B_j)$ se puede obtener a partir de $\{\mu(X) | X \in C_b\}$. De aquí y por (\diamond) , $\mu(B)$ se puede obtener a partir de $\{\mu(X) | X \in C_b\}$.

b) B es una diferencia de conjuntos que satisfacen i), ii), iii) o a), es decir:

$$B = C - C',$$

con $C' \subseteq C$ y C', C satisfaciendo i), ii), iii) o a). Por lo tanto $\mu(B) = \mu(C) - \mu(C')$ se puede obtener a partir de $\{\mu(C) | C \in C_b\}$.

v) Si B no satisface i), ii), iii) o iv), entonces de forma similar que en iv) se establecen dos casos posibles, para los cuales se muestra que $\mu(B)$ se puede obtener a partir de $\{\mu(X) | X \in C_b\}$.

Al ser $\langle \mathfrak{C} \rangle$ finito este proceso se detiene y por lo tanto de forma recursiva se tiene lo pedido. \square

Observación:

1. De forma similar a como se demostró la proposición anterior se puede mostrar que una medida con signo μ sobre $\langle \mathfrak{C} \rangle$ está determinada por $\{\mu(X) | X \in \mathfrak{C}\}$.
2. Si $|\langle \mathfrak{C} \rangle| = r$ y $|C_b| = s$, entonces $\langle \mathfrak{C} \rangle = \{C_1, C_2, \dots, C_r\}$ y $C_b = \{D_1, D_2, \dots, D_s\}$. Sean:

$$p = \begin{bmatrix} \mu(C_1) \\ \cdot \\ \cdot \\ \cdot \\ \mu(C_r) \end{bmatrix} \quad \text{y} \quad q = \begin{bmatrix} \mu(D_1) \\ \cdot \\ \cdot \\ \cdot \\ \mu(D_s) \end{bmatrix}.$$

Por la proposición anterior existe una matriz $E \in M_{r \times s}(\mathbb{R})$ tal que:

$$p = Eq.$$

La forma de encontrar la matriz E no es arbitraria, esta se obtiene a partir de la escritura de los elementos de información de $\langle \mathfrak{C} \rangle$ en términos de los elementos del conjunto de información básico C_b .

3. La escritura de todos los elementos de $\langle \mathfrak{C} \rangle$ no es única en términos de los elementos del conjunto de información básico C_b .

Para ilustrar 3 de la observación anterior se darán los siguientes ejemplos.

Ejemplo 3.4.8. Sea $C = \{A \cup B \cup C, A \cup B, A \triangle B, A \triangle B \triangle C\} \subseteq F_3$ ⁸. Veamos que C es un conjunto de información básico de si mismo. Se mostrará que si $D \subsetneq C$, entonces $\langle D \rangle \subsetneq \langle C \rangle$.

1. Si $|D| = 1$ claramente $\langle D \rangle \neq \langle C \rangle$.

2. Si $|D| = 2$, entonces las opciones son:

- $D = \{A \cup B, A \triangle B\}$ entonces $A \cup B \cup C \notin \langle D \rangle$.
- $D = \{A \cup B, A \triangle B \triangle C\}$ entonces $A \cup B \cup C \notin \langle D \rangle$.
- $D = \{A \cup B, A \cup B \cup C\}$ entonces $A \triangle B \notin \langle D \rangle$.
- $D = \{A \triangle B, A \cup B \cup C\}$ entonces $A \cup B \notin \langle D \rangle$.
- $D = \{A \triangle B, A \triangle B \triangle C\}$ entonces $A \cup B \notin \langle D \rangle$.
- $D = \{A \cup B \cup C, A \triangle B \triangle C\}$ entonces $A \cup B \notin \langle D \rangle$.

3. Si $|D| = 3$ se tiene que $\langle D \rangle \subsetneq \langle C \rangle$. Ya que para $E \in C - D$, E no es un elemento de información de D , al no poderse escribir como uniones disyuntas o diferencias de elementos de D .

Así C es un conjunto de información básico de si mismo.

Ahora se mostrará que $A \cap B \in \langle C \rangle$ no se escribe de forma única en términos de los elementos de C .

Sean

$$D_1 = A \cup B \cup C - (A \triangle B) = [C - (A \cup B)] \cup (A \cap B).$$

$$D_2 = A \cup B \cup C - (A \cup B) = C - (A \cup B)$$

⁸En los ejemplos dados en esta sección se supondrá que el número de átomos no vacíos de F_n es $2^n - 1$

Así $A \cap B$ tiene dos escrituras diferentes en términos de C , $A \cap B = D_1 - D_2 = [A \cup B \cup C - (A \Delta B)] - [A \cup B \cup C - (A \cup B)]$, ó $A \cap B = (A \cup B) - (A \Delta B)$.

Sin embargo, note que para cualquier medida con signo μ sobre $\langle \mathfrak{C} \rangle$ se tiene que

$$\mu(A \cap B) = \mu(A \cup B \cup C) - \mu(A \Delta B) - [\mu(A \cup B \cup C) - \mu(A \cup B)] = \mu(A \cup B) - \mu(A \Delta B)$$

Observe que a pesar de que $A \cap B$ tenga dos escrituras diferentes mediante términos de C , su medida se escribe en forma única como combinación lineal de los elementos $\{\mu(X) | X \in C\}$.

Ejemplo 3.4.9. Sea $G = \{A \cup B, A \cap B, A - B\} \subseteq F_2$. Veamos que G es un conjunto de información básico de sí mismo, para esto se mostrará que si $H \subsetneq G$, entonces $\langle H \rangle \subsetneq \langle G \rangle$.

1. Si $|H| = 1$, claramente $\langle H \rangle \neq \langle G \rangle$.
2. Si $|H| = 2$, entonces se tienen las siguientes opciones:
 - $H = \{A \cup B, A \cap B\}$, entonces $A - B \notin \langle H \rangle$.
 - $H = \{A \cup B, A - B\}$, entonces $A \cap B \notin \langle H \rangle$.
 - $H = \{A \cap B, A - B\}$, entonces $A \cup B \notin \langle H \rangle$

Así G es un conjunto de información básico de sí mismo. Por otro lado, $A \Delta B \in \langle G \rangle$ ya que

$$A \Delta B = (A \cup B) - (A \cap B).$$

$A \Delta B$ no tiene escritura única en términos de los elementos de G . Primero se observará que B y $B - A \in \langle G \rangle$

$$\begin{aligned} B &= (A \cup B) - (A - B) \\ B - A &= ((A \cup B) - (A - B)) - (A \cap B), \end{aligned}$$

Teniendo en cuenta lo anterior se concluye que

$$A \Delta B = (A - B) \cup (B - A) = (A - B) \cup [((A \cup B) - (A - B)) - (A \cap B)],$$

así, $A \Delta B$ tiene dos escrituras diferentes en términos de los elementos de C . Sin embargo,

$$\begin{aligned} \mu(A \Delta B) &= \mu((A - B) \cup [((A \cup B) - (A - B)) - (A \cap B)]) \\ &= \mu(A - B) + \mu([((A \cup B) - (A - B)) - (A \cap B)]) \\ &= \mu(A - B) + \mu(A \cup B) - \mu(A - B) - \mu(A \cap B) \\ &= \mu(A \cup B) - \mu(A \cap B) \end{aligned}$$

es decir, la escritura de $\mu(A \Delta B)$ es única en términos de las medidas de los elementos de C , a pesar de que $A \Delta B$ tenga escrituras diferentes en términos de los elementos de C .

En la siguiente proposición se probará que la matriz E de la observación 2 es única.

Proposición 3.4.10. *Sea $\mathfrak{C} \subseteq F_n$. Si C_b es un conjunto de información básico de \mathfrak{C} , entonces una medida con signo μ sobre $\langle \mathfrak{C} \rangle$ está unívocamente determinada por $\{\mu(X) | X \in C_b\}$.*

Demostración. Sea $B \in \langle \mathfrak{C} \rangle$. Hay dos posibilidades:

i) B se escribe de forma única mediante los elementos de C_b .

Así según la construcción de la matriz E , $\mu(B)$ se escribe de forma única como combinación lineal de $\{\mu(X) | X \in C_b\}$.

ii) B tiene diferentes escrituras en términos de los elementos de C_b . Puede suceder alguno de los siguientes casos:

- a) $\mu(B)$ se obtiene de forma única como combinación lineal de los valores $\{\mu(X) | X \in C_b\}$.
- b) Existen dos escrituras de B en términos de C_b que generan diferentes combinaciones lineales de $\{\mu(X) | X \in C_b\}$ las cuales expresan el valor $\mu(B)$.

Es decir, existen $\alpha_i, \alpha'_i \in \mathbb{R}$, para $1 \leq i \leq s$ tales que:

$$\alpha_1\mu(D_1) + \alpha_2\mu(D_2) + \dots + \alpha_s\mu(D_s) = \alpha'_1\mu(D_1) + \alpha'_2\mu(D_2) + \dots + \alpha'_s\mu(D_s) \quad (*)$$

siendo $C_b = \{D_1, D_2, \dots, D_s\}$ y con $\alpha_j \neq \alpha'_j$, para algún $1 \leq j \leq s$.

De (*) existen $\beta_i \in \mathbb{R}$, $1 \leq i \leq s$ no todos nulos tale que:

$$\beta_1\mu(D_1) + \dots + \beta_s\mu(D_s) = 0 \quad (\Delta)$$

(Δ) se tiene para toda medida con signo μ sobre $\langle \mathfrak{C} \rangle$. Por la proposición 3.4.7 se pueden dar los siguientes valores a las medidas de elementos de C_b :

Si $\beta_i \neq 0$ y $\beta_i > 0$, entonces $\mu(D_i) = 1$. Si $\beta_i \neq 0$ y $\beta_i < 0$, entonces $\mu(D_i) = -1$. los cuales generan una medida con signo sobre $\langle \mathfrak{C} \rangle$.

Dada la anterior medida se tiene:

$$\beta_1\mu(D_1) + \dots + \beta_s\mu(D_s) > 0, \quad (\rightarrow\leftarrow)$$

con (Δ). Por lo tanto este caso no puede suceder.

Así, $\mu(B)$ se escribe de manera única como combinación lineal de $\{\mu(X)|X \in C_b\}$, a pesar de que la escritura de B en términos de los elementos de C_b no resulte ser única. Así, la matriz E de la observación anteriormente hecha es única. \square

Ejemplo 3.4.11. *Se recuerda que si $B_4 = \{A \triangle B, A \cup B, A \triangle C, A \cap C\}$, cualquier medida con signo μ sobre $\langle B_4 \rangle = \{\emptyset, A \triangle B, C \triangle A, A \cup B, C \cap A, A \cap B, C \cup A\}$ está determinada por $\{\mu(X)|X \in B_4\}$. En particular*

$$\mu(A \cap B) = \mu(A \cup B) - \mu(A \triangle B) \quad (3-13)$$

$$\mu(A \cup C) = \mu(C \triangle A) - \mu(A \cap C). \quad (3-14)$$

Note que (3-13) y (3-14) se satisfacen para cualquier medida con signo, ya que estas medidas solo dependen de las expresiones de $A \cap B$ y $A \cup C$ en términos de los elementos de B_4 .

Observación:

Note que, si $B \in \mathfrak{C}$, para cualquier medida con signo μ , $\mu(B)$ depende de la expresión de B en términos de los elementos de C_b .

Ahora, se probará una propiedad de los conjuntos de información básicos.

Proposición 3.4.12. *Sea $\mathfrak{C} \subseteq F_n$. Si B_1 y B_2 son conjuntos de información básicos de \mathfrak{C} entonces $|B_1| = |B_2|$.*

Demostración. Se hará la prueba por contradicción. Supongamos $r = |B_1|$, $k = |B_2|$ y $r < k$. Al ser B_1 un conjunto de información básico de \mathfrak{C} , por la proposición anterior cualquier medida con signo sobre $\langle \mathfrak{C} \rangle$ esta completamente determinada por $\{\mu(B)|B \in B_1\}$.

Al ser B_2 un conjunto de información básico de \mathfrak{C} , entonces $B_2 \subseteq \langle \mathfrak{C} \rangle$. Así por las dos afirmaciones anteriores si $B_1 = \{B_1^1, \dots, B_r^1\}$ y $B_2 = \{B_1^2, \dots, B_k^2\}$, entonces

$$\begin{bmatrix} \mu(B_1^2) \\ \cdot \\ \cdot \\ \cdot \\ \mu(B_k^2) \end{bmatrix} = D \begin{bmatrix} \mu(B_1^1) \\ \cdot \\ \cdot \\ \cdot \\ \mu(B_r^1) \end{bmatrix}$$

donde $D \in M_{k \times r}(\mathbb{R})$. Como B_2 es un conjunto de información básico de \mathfrak{C} , debido a la proposición 3.4.7, se pueden hacer la siguiente asignación

$$e_i = \begin{bmatrix} \mu(B_1^2) \\ \cdot \\ \cdot \\ \cdot \\ \mu(B_k^2) \end{bmatrix}$$

para $1 \leq i \leq k$. Ahora como D es válida para cualquier medida con signo sobre $\langle \mathfrak{C} \rangle$, por la asignación anterior la dimensión del espacio columna de D es k pero el número de columnas de D es r y $r < k$ ($\rightarrow \leftarrow$).

Por lo tanto $|B_1| \geq |B_2|$. Análogamente si $|B_2| < |B_1|$ se llega a una contradicción.

En consecuencia $|B_1| = |B_2|$ □

3.4.2. Algoritmo para obtener conjuntos de información generados.

Se puede crear un algoritmo que genere el conjunto de información $\langle \mathfrak{C} \rangle$ de un conjunto $\mathfrak{C} \subseteq F_n$ dado. Para esto se da la siguiente definición.

Definición 3.4.13. Sea $\mathfrak{C} \subseteq F_n$. Se definen los siguientes conjuntos:

1. $\mathfrak{C}_\oplus \subseteq F_n$, donde

$$\mathfrak{C}_\oplus = \left\{ \bigoplus_{i=1}^r A_i \mid A_i \in \mathfrak{C} \right\}.$$

2. $\mathfrak{C}_- \subseteq F_n$, donde

$$\mathfrak{C}_- = \{C' - C \mid C \subseteq C', C, C' \in \mathfrak{C}\}.$$

Algoritmo para generar $\langle \mathfrak{C} \rangle$ *Entrada:* $\mathfrak{C} \subseteq F_n$.

Se construyen:

1. $\mathfrak{C}_0 := \mathfrak{C} \cup \{\emptyset\}$
2. $\mathfrak{C}_1 := \mathfrak{C}_0 \cup (\mathfrak{C}_0)_{\boxplus}$
3. $\mathfrak{C}_2 := \mathfrak{C}_1 \cup (\mathfrak{C}_1)_{-}$

En general si $i > 1$

$$\mathfrak{C}_i := \mathfrak{C}_{i-1} \cup (\mathfrak{C}_{i-1})_{\boxplus} \quad \text{si } i \equiv 1 \pmod{2}$$

$$\mathfrak{C}_i := \mathfrak{C}_{i-1} \cup (\mathfrak{C}_{i-1})_{-} \quad \text{si } i \equiv 0 \pmod{2}$$

Hasta: $\mathfrak{C}_i = \mathfrak{C}_{i+2}$.*Salida:* \mathfrak{C}_i tal que $\mathfrak{C}_i = \mathfrak{C}_{i+2}$.**Observación:**

- Dado $\mathfrak{C} \subseteq F_n$ el proceso de generar $\langle \mathfrak{C} \rangle$ se detiene ya que F_n es finito. Por otro lado observe que si $\mathfrak{C}_i = \mathfrak{C}_{i+2}$ entonces $\mathfrak{C}_i = \mathfrak{C}_{i+1}$, ya que por construcción:

$$\mathfrak{C}_i \subseteq \mathfrak{C}_{i+1} \subseteq \mathfrak{C}_{i+2} \subseteq \mathfrak{C}_i.$$

Es decir $\mathfrak{C}_m = \mathfrak{C}_i$ para $m \geq i$.

- Puede suceder que $C_i = C_{i+1}$ pero $c_{i+1} = c_{i+2}$. Por ejemplo, si $C_0 = \{A \cup B, A \cap B, \emptyset\}$, entonces $C_0 = C_1$; pero $C_0 \neq C_2$, porque $C_2 = \{A \cup B, A \cap B, A \triangle B, \emptyset\}$

Proposición 3.4.14. Sean $\mathfrak{C} \subseteq F_n$ y \mathfrak{C}_i tal que $\mathfrak{C}_i = \mathfrak{C}_{i+2}$ (del algoritmo anterior) entonces

$$\mathfrak{C}_i = \langle \mathfrak{C} \rangle.$$

Demostración. i) Se verá que $\mathfrak{C}_i \subseteq \langle \mathfrak{C} \rangle$.

En primer lugar se mostrará que si $\mathfrak{C}_j \subseteq \langle \mathfrak{C} \rangle$ y $j \equiv 0 \pmod{2}$ entonces $\mathfrak{C}_{j+1}, \mathfrak{C}_{j+2} \subseteq \langle \mathfrak{C} \rangle$.

Si $\mathfrak{C}_j \subseteq \langle \mathfrak{C} \rangle$, por definición de elemento de información de \mathfrak{C} se tiene que $(\mathfrak{C}_j)_\sqcup \subseteq \langle \mathfrak{C} \rangle$. De lo anterior y puesto que $\mathfrak{C}_{j+1} = \mathfrak{C}_j \cup (\mathfrak{C}_j)_\sqcup$, $\mathfrak{C}_{j+1} \subseteq \langle \mathfrak{C} \rangle$. De manera similar, $\mathfrak{C}_{j+2} \subseteq \langle \mathfrak{C} \rangle$ porque $\mathfrak{C}_{j+2} = \mathfrak{C}_{j+1} \cup (\mathfrak{C}_{j+1})_-$ y $(\mathfrak{C}_{j+1})_- \subseteq \langle \mathfrak{C} \rangle$.

Debido a la anterior observación y dado que $\mathfrak{C}_0 = \mathfrak{C} \cup \{\emptyset\} \subseteq \langle \mathfrak{C} \rangle$ entonces $\mathfrak{C}_1, \mathfrak{C}_2 \subseteq \langle \mathfrak{C} \rangle$. Y razonando inductivamente $\mathfrak{C}_j \subseteq \langle \mathfrak{C} \rangle$ para todo $j = 0, 1, \dots, i$ en particular $\mathfrak{C}_i \subseteq \langle \mathfrak{C} \rangle$ tal que $\mathfrak{C}_i = \mathfrak{C}_{i+2}$.

ii) Se verá que $\langle \mathfrak{C} \rangle \subseteq \mathfrak{C}_i$.

Sea $C \in \langle \mathfrak{C} \rangle$ entonces C debe satisfacer alguna de las condiciones i), ii), iii), iv), v) de la proposición 3.4.7. Por lo tanto, $C \in \mathfrak{C}_j$ para algún j . Como $\mathfrak{C}_j \subseteq \mathfrak{C}_i$ para todo j , entonces $C \in \mathfrak{C}_i$.

□

Ejemplo 3.4.15. Se considera de nuevo el conjunto $B_4 = \mathfrak{D} = \{A \triangle B, C \triangle A, A \cup B, A \cap B\}$, siguiendo el algoritmo descrito:

$$\mathfrak{D}_0 = \{\emptyset, A \triangle B, C \triangle A, A \cup B, A \cap B, \}$$

$$\mathfrak{D}_1 = \{\emptyset, A \triangle B, C \triangle A, A \cup B, A \cap B, A \cup C\}$$

$$\mathfrak{D}_2 = \{\emptyset, A \triangle B, C \triangle A, A \cup B, A \cap B, A \cup C, A \cap B\}$$

$$\mathfrak{D}_3 = \mathfrak{D}_2$$

$$\mathfrak{D}_4 = \mathfrak{D}_3.$$

Así $\mathfrak{D}_4 = \mathfrak{D}_2$ por lo tanto $\langle B_4 \rangle = \mathfrak{D}_2$.

Ejemplo 3.4.16. Sea $\mathfrak{C} = \{A \cup B, A \triangle B, B\}$ sobre F_2 . Para calcular $\langle \mathfrak{C} \rangle$ lo hacemos mediante el algoritmo descrito:

$$\mathfrak{C}_0 = \mathfrak{C} \cup \{\emptyset\}$$

$$\mathfrak{C}_1 = \mathfrak{C}_0$$

$$\mathfrak{C}_2 = \{\emptyset, A \cup B, A \triangle B, B, A \cap B, A - B\}$$

$$\mathfrak{C}_3 = \{\emptyset, A \cup B, A \triangle B, B, A \cap B, A - B, A\}$$

$$\mathfrak{C}_4 = \{\emptyset, A \cup B, A \triangle B, B, A \cap B, A - B, A, B - A\}$$

$$\mathfrak{C}_5 = F_2 = \mathfrak{C}_6 = \mathfrak{C}_7.$$

Así $\langle \mathfrak{C} \rangle = F_2$.

Observación:

En el ejemplo anterior se observó que los elementos de información de \mathfrak{C} son todos los elementos del campo F_2 , este hecho motiva la siguiente definición de conjunto de información completa.

Definición 3.4.17. Un conjunto $\mathfrak{C} \subseteq F_n$ es un conjunto de información completa si

$$\langle \mathfrak{C} \rangle = F_n.$$

Proposición 3.4.18. Un conjunto $\mathfrak{C} \subseteq F_n$ es un conjunto de información completa si y solo si \mathfrak{A} (conjunto de átomos no vacíos de F_n) es un conjunto de información básico de \mathfrak{C} .

Demostración. (\Rightarrow) Si \mathfrak{C} es un conjunto de información completa entonces:

$$\langle \mathfrak{C} \rangle = F_n.$$

Así $\mathfrak{A} \subseteq F_n = \langle \mathfrak{C} \rangle$. Ahora como $\langle \mathfrak{A} \rangle = F_n = \langle \mathfrak{C} \rangle$. Y si $C \subsetneq \mathfrak{A}$, $\langle C \rangle \subsetneq \mathfrak{A} = F_n$. De lo anterior y por definición \mathfrak{A} es un conjunto de información básico de \mathfrak{C} .

(\Leftarrow) Si \mathfrak{A} es un conjunto de información básico de \mathfrak{C} entonces

$$\langle \mathfrak{A} \rangle = \langle \mathfrak{C} \rangle \Leftrightarrow F_n = \langle \mathfrak{C} \rangle,$$

por lo tanto \mathfrak{C} es un conjunto de información completa. □

Proposición 3.4.19. *Si \mathfrak{C} es un conjunto de información completa entonces $|\mathfrak{C}| \geq |\mathfrak{A}|$*

Demostración. Por contradicción, supongamos que $s = |\mathfrak{C}| < |\mathfrak{A}| = k$.

Existe una matriz $D \in M_{k \times s}(\mathbb{R})$ tal que para toda medida con signo μ :

$$\begin{bmatrix} \mu(A_1) \\ \cdot \\ \cdot \\ \cdot \\ \mu(A_k) \end{bmatrix} = D \begin{bmatrix} \mu(C_1) \\ \cdot \\ \cdot \\ \cdot \\ \mu(C_s) \end{bmatrix}$$

donde $\mathfrak{A} = \{A_1, A_2, \dots, A_k\}$ y $\mathfrak{C} = \{C_1, C_2, \dots, C_s\}$. Como lo anterior se tiene para toda medida entonces la dimensión del espacio columna de D es k sin embargo el número de columnas de D es s ($\rightarrow \leftarrow$).

Por lo tanto se tiene lo pedido, es decir $|\mathfrak{C}| \geq |\mathfrak{A}|$. \square

Debido a la proposición anterior se puede pensar en conjuntos con el mínimo número de elementos que generen toda la información del campo, pensando en esto se da la siguiente definición.

Definición 3.4.20. $\mathfrak{C} \subseteq F_n$ es un conjunto mínimo de información completa si:

1. \mathfrak{C} es un conjunto de información completa.
2. $|\mathfrak{C}| = |\mathfrak{A}|$.

Ejemplo 3.4.21. *El conjunto de átomos no vacíos es un conjunto mínimo de información completa. Ya que*

1. \mathfrak{A} es un conjunto de información completa

$$\mathfrak{A}_0 = \mathfrak{A} \cup \{\emptyset\}$$

$$\mathfrak{A}_1 = \mathfrak{A}_0 \cup \mathfrak{A}_{\circ\sqcup} = \mathfrak{A}_0 \cup F_n = F_n$$

$$\mathfrak{A}_2 = \mathfrak{A}_1 \cup (\mathfrak{A}_1)_- = F_n$$

$$\mathfrak{A}_3 = \mathfrak{A}_2 \cup \mathfrak{A}_{2-} = F_n$$

Así como $\mathfrak{A}_1 = \mathfrak{A}_3 = F_n$, entonces $\langle \mathfrak{A} \rangle = F_n$.

2. $|\mathfrak{A}| = |\mathfrak{A}|$

Por lo anterior, \mathfrak{A} es un conjunto de información completa.

Ejemplo 3.4.22. Se mostrará que \mathfrak{B} es un conjunto mínimo de información completa. Donde \mathfrak{B} es el conjunto⁹

$$\mathfrak{B} = \{\tilde{X}_\alpha : \alpha \text{ es un conjunto no vacío de } N_n\}.$$

1. En primer lugar se mostrará que \mathfrak{B} es un conjunto de información completa. Se mostrará que $\langle \mathfrak{B} \rangle = F_n$, equivalentemente que $\mathfrak{A} \subseteq \langle \mathfrak{B} \rangle$.

Si $A_0 \in \mathfrak{A}$ entonces

$$A_0 = \bigcap_{i=1}^n Y_i,$$

donde $Y_i = \tilde{X}_i$ ó $Y_i = \tilde{X}_i^c$.

Sea $J = \{i \in N_n : Y_i = \tilde{X}_i\}$, entonces $J \neq \emptyset$ ya que $A_0 \in \mathfrak{A}$, A_0 no puede ser el átomo vacío.

Si $|J| = m > 0$, se escribirán de la siguiente forma los conjuntos:

- $\{\tilde{X}_i \mid i \in J\} = \{\tilde{X}'_i \mid 1 \leq i \leq m\}$.
- $\{\tilde{X}_i \mid i \notin J\} = \{\tilde{X}'_i \mid m+1 \leq i \leq n\}$.

Así

$$A_0 = \bigcap_{i=1}^m \tilde{X}'_i - \bigcup_{i=m+1}^n \tilde{X}'_i.$$

Se probará que $A_0 \in \langle \mathfrak{B} \rangle$, lo cual se hará por inducción sobre $|J|$.

i) Si $|J| = 1$, entonces

$$A_0 = \tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j.$$

Veamos que $A_0 = \Omega - \bigcup_{j \neq i} \tilde{X}'_j$:

(\subseteq) Si $x \in \tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j$, entonces $x \in \tilde{X}'_i$ y $x \notin \bigcup_{j \neq i} \tilde{X}'_j$.

Por lo tanto, $x \in \Omega - \bigcup_{j \neq i} \tilde{X}'_j$.

(\supseteq) Se tiene por la definición de Ω .

⁹Donde $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ conjuntos que se intersecan unos con otros.

Como Ω y $\bigcup_{j \neq i} \tilde{X}'_j \in \mathfrak{B}$, entonces por definición de elementos de información generados, $A_0 = \Omega - \bigcup_{j \neq i} \tilde{X}'_j \in \langle \mathfrak{B} \rangle$.

ii) Si $|J| = 2$, entonces

$$A_0 = (\tilde{X}'_i \cap \tilde{X}'_k) - \bigcup_{j \neq i, k} \tilde{X}'_j.$$

Veamos que $A_0 = C$, donde

$$C = \Omega - \left[\left(\tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j \right) \uplus \left(\tilde{X}'_k - \bigcup_{j \neq k} \tilde{X}'_j \right) \uplus \bigcup_{j \neq i, k} \tilde{X}'_j \right].$$

(\subseteq) Se verá que $C^c \subseteq A_0^c$.

Si $x \in C^c$, entonces $x \in \tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j$ ó $x \in \tilde{X}'_k - \bigcup_{j \neq k} \tilde{X}'_j$ ó $x \in \bigcup_{j \neq i, k} \tilde{X}'_j$.

Dadas las tres opciones anteriores se implican respectivamente:

$$x \in \tilde{X}'_i - \tilde{X}'_k \text{ ó } x \in \tilde{X}'_k - \tilde{X}'_i \text{ ó } x \notin A_0.$$

Así, para cualquier de las tres condiciones anteriores se tiene que $x \notin A_0$.

(\supseteq) Se verá que $C \subseteq A_0$.

Si $x \in C$, entonces $x \notin \bigcup_{j \neq i, k} \tilde{X}'_j$. Ahora, por la definición del conjunto universal Ω , $x \in \tilde{X}'_i \cup \tilde{X}'_k$ y como $x \notin \tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j$; se debe tener que $x \in (\tilde{X}'_i \cap \tilde{X}'_k) - \bigcup_{j \neq i, k} \tilde{X}'_j$.

Así $A_0 = C$ y por lo tanto $A_0 \in \langle \mathfrak{B} \rangle$, ya que:

- $\tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j, \tilde{X}'_k - \bigcup_{j \neq k} \tilde{X}'_j \in \langle \mathfrak{B} \rangle$ por i).
- $\Omega, \bigcup_{j \neq i, k} \tilde{X}'_j \in \mathfrak{B} \subseteq \langle \mathfrak{B} \rangle$.

De (i) y (ii), se observa que para probar el caso $|J| = 2$ es fundamental tener el caso $|J| = 1$. Se mostrará que en general, si para $|J| < m$ ($m < n$) los átomos de la forma

$$\bigcap_{i \in J} \tilde{X}'_i - \bigcap_{i \notin J} \tilde{X}'_i,$$

pertenecen a $\langle \mathfrak{B} \rangle$, entonces para $|J_1| = m$, los átomos de la forma

$$\bigcap_{i \in J_1} \tilde{X}'_i - \bigcap_{i \notin J_1} \tilde{X}'_i,$$

pertenecen a $\langle \mathfrak{B} \rangle$.

Demostración. Sea un átomo de la forma:

$$A_1 = \bigcap_{i=1}^m \tilde{X}'_i - \bigcup_{i=m+1}^n \tilde{X}'_i,$$

Veremos que $A_1 = D$, donde

$$D = \Omega - \left[\bigcup_{i=1}^m \left(\tilde{X}'_i - \bigcup_{j \neq i} \tilde{X}'_j \right) \uplus \bigcup_{i_1, i_2=1}^m \left(\left(\tilde{X}'_{i_1} \cap \tilde{X}'_{i_2} \right) - \bigcup_{j \neq i_1, i_2} \tilde{X}'_j \right) \uplus \dots \right. \\ \left. \uplus \bigcup_{i_1, \dots, i_{m-1}=1}^m \left(\left(\tilde{X}'_{i_1} \cap \dots \cap \tilde{X}'_{i_{m-1}} \right) - \bigcup_{j \neq i_1, \dots, i_{m-1}} \tilde{X}'_j \right) \uplus \bigcup_{i=m+1}^n \tilde{X}'_i \right]$$

(\subseteq) Se verá que $A_1 \subseteq D$, equivalentemente $D^c \subseteq A_1^c$.

Si $x \in D^c$, entonces $x \in \bigcup_{i=m+1}^n \tilde{X}'_i$ ó $x \in \left(\tilde{X}'_{i_1} \cap \dots \cap \tilde{X}'_{i_k} \right) - \bigcup_{j \neq i_1, \dots, i_k} \tilde{X}'_j$ con $k < m$.

- Si se tiene la primera condición se tiene que $x \notin A_1$.
- Si se tiene la segunda condición entonces $x \notin \bigcap_{i=1}^m \tilde{X}'_i$, ya que esta máximo en $m - 1$ conjuntos \tilde{X}'_i ; por lo tanto, $x \notin A_1$.

(\supseteq) Se verá que $D \subseteq A_1$.

Si $x \in D$, entonces $x \notin \bigcup_{i=m+1}^n \tilde{X}'_i$. Así $x \in \bigcap_{i=1}^m \tilde{X}'_i$.

Como x no pertenece a ningún conjunto de la forma

$$\left(\tilde{X}'_{i_1} \cap \dots \cap \tilde{X}'_{i_k} \right) - \bigcup_{j \neq i_1, \dots, i_k} \tilde{X}'_j \text{ con } k < m,$$

entonces

$$x \in \bigcap_{i=1}^m \tilde{X}'_i - \bigcup_{i=m+1}^n \tilde{X}'_i = A_1.$$

Como $A_1 = D$ y por hipótesis todos los átomos de la forma $\bigcap_{i \in J} \tilde{X}'_i - \bigcap_{i \notin J} \tilde{X}'_i$ pertenecen a $\langle \mathfrak{B} \rangle$, entonces $D \in \langle \mathfrak{B} \rangle$. Por lo tanto $A_1 \in \langle \mathfrak{B} \rangle$. Si $|J| = n$, entonces

$$A_1 = \bigcap_{i=1}^n \tilde{X}'_i = \Omega - \bigcup_{A \in \mathfrak{A}: A \neq A_1} A,$$

de forma análoga se demuestra que $A_1 \in \langle \mathfrak{B} \rangle$. □

Debido al resultado anterior y que todos los átomos de la forma $\bigcap_{i \in J_1} \tilde{X}'_i - \bigcap_{i \notin J_1} \tilde{X}'_i$ (para $|J| < 2$), pertenecen a $\langle \mathfrak{B} \rangle$, entonces por recursividad se muestra que $\mathfrak{A} \subseteq \langle \mathfrak{B} \rangle$ y con esto, que \mathfrak{B} es un conjunto de información completa.

2. Veamos que $|\mathfrak{B}| = |\mathfrak{A}|$.

Observemos que $|\mathfrak{B}|$ es el número de subconjuntos de $N_n = \{1, 2, \dots, n\}$ no vacíos, es decir, es $2^n - 1$ que es el mismo que el número de átomos no vacíos de F_n , es decir, $|\mathfrak{B}| = |\mathfrak{A}|$.

Así por 1 y 2, \mathfrak{B} es un conjunto de información mínimo de información completa.

En la última sección de este capítulo se han presentado ciertos resultados propios, los cuales fueron motivados por la manera en que se definió la f -medida. En un trabajo futuro es posible, profundizar en la búsqueda de propiedades de estos conjuntos.

4. Los I -diagramas

En el capítulo anterior se estableció que las medidas de la información Shannon son los valores de una medida con signo sobre algunos elementos del campo F_n . Es decir, se puede tratar algunos aspectos de la Teoría de la Información Shannon a través de la Teoría de Conjuntos, así, es válido usar diagramas similares a los diagramas de Venn para representar las relaciones entre las medidas de la información Shannon.¹ Dichos diagramas, introducidos por Yeung recibieron el nombre de I -diagramas.²

Primero, se recuerda la siguiente definición:

Definición 4.0.23. *Sea $C = \{C_1, C_2, \dots, C_n\}$ una colección de curvas simples cerradas en el plano. La colección C se dice una familia independiente, si la región formada por la intersección de Y_1, Y_2, \dots, Y_n es no vacía, donde cada Y_i es $\text{int}(C_i)$ o $\text{ext}(C_i)$. Además, si cada región es conexa y sólo hay finitos puntos de intersección entre curvas, C es un diagrama de Venn.*

Observación:

Según la definición anterior, si se piensa en un diagrama de Venn para los conjuntos $\tilde{X}_1, \dots, \tilde{X}_n$ que generan el campo F_n , se tendrá que en dicho diagrama los átomos no vacíos deben ser representados mediante regiones conexas.

En un I -diagrama, el *área* de una región representa el valor de μ^* sobre un subconjunto de Ω en el campo F_n . Análogamente en un f -diagrama, el *área* representa el valor de μ_f^* .

De ésta manera, el I -diagrama para dos variables aleatorias X_1, X_2 se muestra en la **figura 4-1**. Se señalan las áreas de cada uno de los átomos no vacíos de F_2 .

¹Más aún se estableció este hecho de manera general para las medidas de información generadas por una función polimatroide

²De forma similar se puede pensar en los f -diagramas, para una función polimatroide f . Sin embargo, este capítulo se enfocará en el estudio de los I -diagramas.

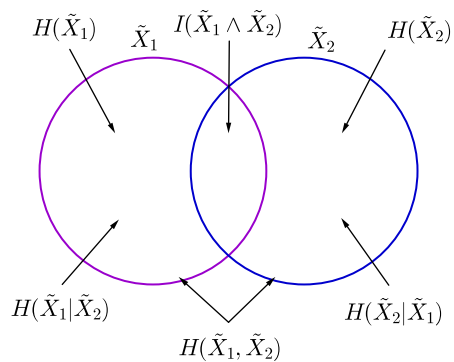


Figura 4-1.:

Y el I -diagrama para tres variables X_1, X_2 y X_3 se muestra a continuación

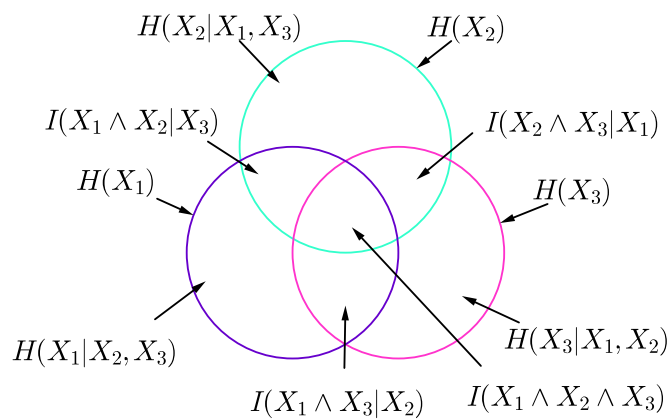


Figura 4-2.:

Observe que para los casos $n = 2$, μ^* es no negativa. Sin embargo, en general no se tiene que μ^* sea no negativa.

Ejemplo 4.0.24. Sean X_1 y X_2 ³ variables aleatorias independientes con valores binarios y tales que:

$$P(X_i = 0) = P(X_i = 1) = \frac{1}{2}.$$

Sea $X_3 = (X_1 + X_2)(\text{mod}2)$

$$P(X_3 = 0) = P(X_3 = 1) = \frac{1}{2}.$$

Por lo tanto, para $i = 1, 2, 3$:

$$\begin{aligned} H(X_i) &= - \left[\frac{1}{2} \log_2 \left(\frac{1}{2} \right) + \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right] \\ &= - \log_2 \left(\frac{1}{2} \right) = 1. \end{aligned}$$

Como X_i, X_j son variables aleatorias independientes para $i \neq j, i, j = 1, 2, 3$.

$$\begin{aligned} H(X_i, X_j) &= - \sum_{i,j} p(x_i, x_j) \log_2 p(x_i, x_j) \\ &= - \sum_{i,j} p(x_i)p(x_j) (\log_2 p(x_i) + \log_2 p(x_j)) \\ &= H(X_i) + H(X_j) = 1 + 1 = 2 \end{aligned}$$

Además $I(X_i \wedge X_j) = 0$ y :

$$H(X_1, X_2, X_3) = H(X_1, X_2) + H(X_3|X_1, X_2) = 2 + 0 = 2.$$

Ahora:

$$\begin{aligned} I(X_1 \wedge X_2|X_3) &= H(X_1, X_3) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_3) \\ &= 2 + 2 - 2 - 1 = 1 \end{aligned}$$

Como $\tilde{X}_1 \cap \tilde{X}_2 = (\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) \cup (\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c)$,

$$\begin{aligned} \mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) &= \mu^*(\tilde{X}_1 \cap \tilde{X}_2) - \mu^*(\tilde{X}_1 \cap \tilde{X}_2 - \tilde{X}_3) \\ &= I(X_1 \wedge X_2) - I(X_1 \wedge X_2|X_3) = 0 - 1 = -1. \end{aligned}$$

Por lo tanto, $I(X_1 \wedge X_2 \wedge X_3) = I(X_1 \wedge X_2) - I(X_1 \wedge X_2|X_3) = -1 < 0$.

³Ejemplo mostrado en R.W.Yeung, A First Course in Information Theory, Springer, pag 103.

Al ser $I(X_1 \wedge X_2 \wedge X_3)$ simétrica en X_1 , X_2 y X_3 , que $I(X_1 \wedge X_2 \wedge X_3)$ sea negativa significa que la información mutua entre cualquier dos variables se incrementa cuando la otra variable es dada.

El I-diagrama para este ejemplo se presenta a continuación:

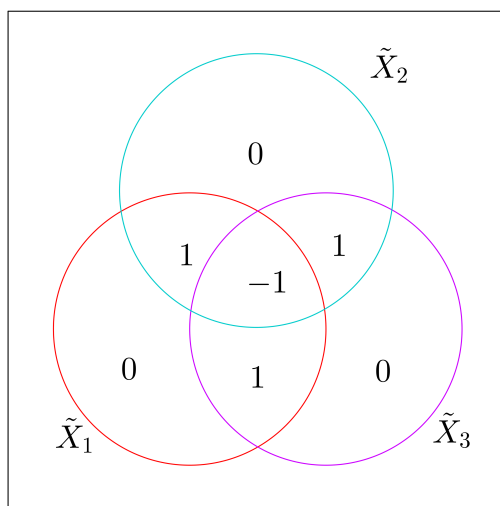


Figura 4-3.:

Por lo tanto, el área de una región en un I-diagrama puede ser representada por un valor negativo.

El uso de los I-diagramas permitió mostrar ciertas identidades de la información de manera sencilla e intuitiva, así como corregir afirmaciones incorrectas hechas por autores anteriores. Por ejemplo, de la siguiente forma se refirió Yeung acerca de algunas afirmaciones de Reza en “*An Introduction to Information Theory*”

“Como consecuencia, cuando dos variables aleatorias X , Y son independientes, esto sólo implica que la suma de las áreas que representan a $I(X \wedge Y)$ en el I-diagrama es cero; es incorrecto indicar en [2] que cuando dos variables aleatorias son independientes, los conjuntos correspondientes son disjuntos.”⁴

⁴R.W.Yeung, A new outlook on Shannon’s information measures, IEEE.Trans.Inform.Theory, pag 469.

Observando el I -diagrama para tres variables aleatorias

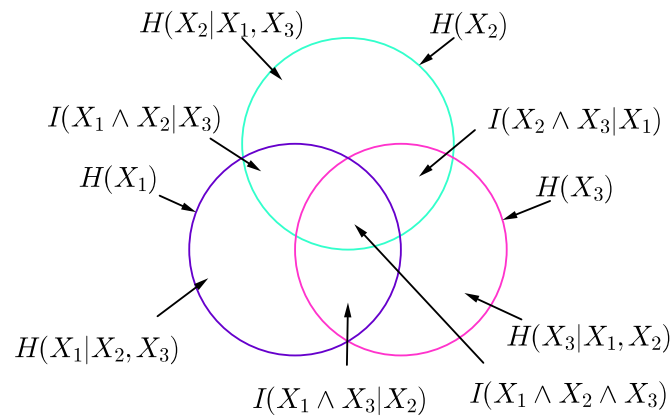


Figura 4-4.:

Es posible verificar los siguientes hechos.

- * $I(X_1 \wedge X_2 \wedge X_3)$ es simétrica en X_1, X_2 y X_3 , ya que del I -diagrama:

$$\begin{aligned} I(X_1 \wedge X_2 \wedge X_3) &= I(X_1 \wedge X_2) - I(X_1 \wedge X_2|X_3) \\ &= I(X_2 \wedge X_3) - I(X_2 \wedge X_3|X_1) \\ &= I(X_1 \wedge X_3) - I(X_1 \wedge X_3|X_2) \end{aligned}$$

- * Si X_1 y X_3 son independientes, $I(X_1 \wedge X_3) = 0$.

Del I -diagrama:

$$I(X_1 \wedge X_3) = I(X_1 \wedge X_3|X_2) + I(X_1 \wedge X_2 \wedge X_3) = 0.$$

Al ser $I(X_1 \wedge X_3|X_2) \geq 0$, de lo anterior $I(X_1 \wedge X_2 \wedge X_3) \leq 0$.

Por lo tanto

$$I(X_1 \wedge X_2) = I(X_1 \wedge X_2|X_3) + I(X_1 \wedge X_2 \wedge X_3) \leq I(X_1 \wedge X_2|X_3).$$

Lo cual significa que si X_1 y X_3 son variables independientes, entonces la información mutua de X_1 y X_2 se incrementa cuando la variable X_3 es dada.

* Sean X_1, X_2 y X_3 variables aleatorias independientes dos a dos. Así

$$I(X_1 \wedge X_2) = I(X_2 \wedge X_3) = I(X_1 \wedge X_3) = 0.$$

Como

$$\begin{aligned} I(X_1 \wedge X_2 \wedge X_3) &= I(X_1 \wedge X_2) - I(X_1 \wedge X_2|X_3) \\ &= I(X_2 \wedge X_3) - I(X_2 \wedge X_3|X_1) \\ &= I(X_1 \wedge X_3) - I(X_1 \wedge X_3|X_2) \end{aligned}$$

Entonces

$$\begin{aligned} -I(X_1 \wedge X_2|X_3) &= -I(X_2 \wedge X_3|X_1) = -I(X_1 \wedge X_3|X_2) \\ I(X_1 \wedge X_2|X_3) &= I(X_2 \wedge X_3|X_1) = I(X_1 \wedge X_3|X_2) \end{aligned}$$

Lo anterior, significa que en este caso la información mutua condicional es simétrica respecto a los argumentos.

Como se observó, existen I -diagramas para dos y tres variables aleatorias, de este hecho surge la pregunta: ¿Existen los I -diagramas para cualquier número n de variables aleatorias?, nótese que esta pregunta es equivalente a ¿Existen los diagramas de Venn para un número n de curvas?.

Según Yeung la respuesta a esta pregunta es negativa, es decir para $n \geq 4$ no es posible mostrar un diagrama de información en dos dimensiones. Sin embargo, plantea que para $n = 4$ un diagrama de información en dos dimensiones se tiene casi perfectamente, el cual se muestra a continuación:

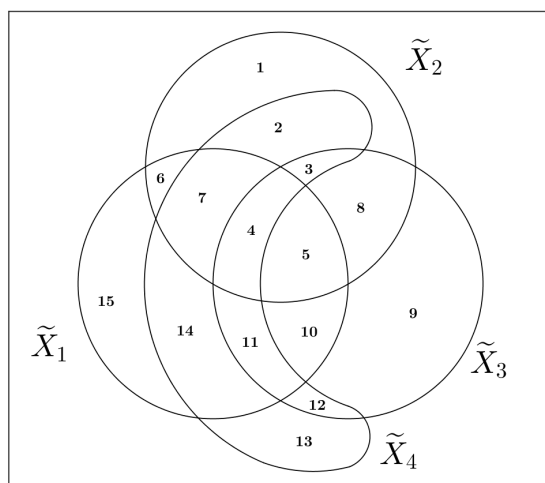


Figura 4-5.:

¿ Por qué el diagrama anterior no es perfecto?, según Yeung:

“...la proximidad de ciertos átomos no es mostrada correctamente. Por ejemplo, el conjunto $\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_4^c$, el cual consiste de los átomos $\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4^c$ y $\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4^c$, no es representado por una región conexa porque los dos átomos no son adyacentes el uno al otro.”⁵

Sin embargo, el razonamiento hecho por Yeung no es válido dada la definición de diagrama de Venn, ya que en la figura 4-5 se representan los 15 átomos no vacíos del campo F_4 mediante regiones conexas, además las curvas que determinan los conjuntos son simples y cerradas, y, entre ellas, se intersectan en finitos puntos. Por lo tanto, la figura 4-5 se puede ver como un I-diagrama, en el cual el área de las regiones enumeradas representarán la I-medida en ciertos átomos no vacíos, como se muestra a continuación:

⁵R.W.Yeung, A First Course in Information Theory, Springer, pag 110.

| Para el átomo enumerado con | El valor de la I-medida es |
|--------------------------------|---|
| 1 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4^c) = H(X_2 X_1, X_3, X_4)$ |
| 2 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4) = I(X_2 \wedge X_4 X_1, X_3)$ |
| 3 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4) = I(X_2 \wedge X_3 \wedge X_4 X_1)$ |
| 4 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4) = I(X_1 \wedge X_2 \wedge X_3 \wedge X_4)$ |
| 5 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4^c) = I(X_1 \wedge X_2 \wedge X_3 X_4)$ |
| 6 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4^c) = I(X_1 \wedge X_2 X_3, X_4)$ |
| 7 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4) = I(X_1 \wedge X_2 \wedge X_4 X_3)$ |
| 8 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4^c) = I(X_2 \wedge X_3 X_1, X_4)$ |
| 9 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4^c) = H(X_3 X_1, X_2, X_4)$ |
| 10 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4^c) = I(X_1 \wedge X_3 X_2, X_4)$ |
| 11 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4) = I(X_1 \wedge X_3 \wedge X_4 X_2)$ |
| 12 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4) = I(X_3 \wedge X_4 X_1, X_2)$ |
| 13 | $\mu^*(\tilde{X}_1^c \cap \tilde{X}_2^c \cap \tilde{X}_3^c \cap \tilde{X}_4) = H(X_4 X_1, X_2, X_3)$ |
| 14 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3^c \cap \tilde{X}_4) = I(X_1 \wedge X_4 X_2, X_3)$ |
| 15 | $\mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3^c \cap \tilde{X}_4^c) = H(X_1 X_2, X_3, X_4)$ |

Más aún, los diagramas de Venn que en primer lugar fueron usados por Euler ⁶, existen para cualquier número finito de curvas, hecho demostrado por John Venn en *On the Diagrammatic and Mechanical Representation of Propositions and Reasonings* (1880). A pesar de la existencia de dichos diagramas, en general no son simétricos, por lo tanto se necesita un estudio más profundo para responder si estos diagramas resultan útiles, es decir, si permiten la visualización de identidades de Teoría de la Información. Por lo pronto, es un hecho, que al aumentar el número de curvas los diagramas son cada vez más complejos, por ejemplo, el diagrama de Venn para 6 curvas se presenta a continuación:

⁶A.W.F.Edwards, Ian Stewart, Osmo Pekonen, Peter Hamburger. Cogwheels of the Mind: The Story of Venn Diagrams. The Mathematical Intelligencer, Vol 27, pag 36-38.

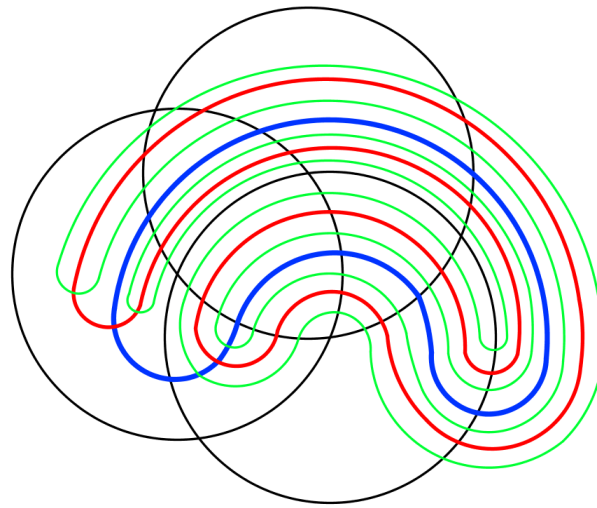


Figura 4-6.:

4.1. I- diagramas para cadenas de Markov

Sin embargo, si las variables aleatorias tienen ciertas características el *I*-diagrama resulta ser menos complejo, es decir en lugar de que el diagrama de información sea un diagrama de Venn, será un diagrama de Venn Euler, lo cual significa que no deben ser mostrados los átomos del campo cuya medida es cero.

En esta sección, se mostrará el caso de los *I*-diagramas para tres y cuatro variables aleatorias que forman una cadena de Markov; la estructura general de los *I*-diagramas para n variables aleatorias fue mostrada por Yeung y puede ser consultada en el capítulo 7 de [12].

En primer lugar, se definirá lo que significa una cadena de Markov.

Definición 4.1.1. Sean X_1, X_2, \dots, X_n ($n \geq 3$) variables aleatorias. $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$ es una cadena de Markov si:

$$p(x_1, x_2, \dots, x_n) = p(x_1, x_2)p(x_3|x_2)\dots p(x_n|x_{n-1})$$

cuando $p(x_2), p(x_3), \dots, p(x_{n-1}) > 0$. En otro caso,

$$p(x_1, x_2, \dots, x_n) = 0.$$

Dada la anterior definición, ¿Cómo pueden interpretarse las cadenas de Markov?

Las cadenas de Markov pueden interpretarse como cadenas con *memoria*, si se conoce la historia del sistema hasta su instante actual, su estado presente resume toda la información para describir en probabilidad su estado futuro, este hecho se probará con base en el siguiente teorema enunciado.

Proposición 4.1.2. *Dado $N_n = \{1, 2, \dots, n\}$ y dada $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \rightarrow X_{n+1}$ una cadena de Markov. Para cualquier subconjunto α de N_n , denote $(X_i, i \in \alpha)$ como X_α . Entonces para cualquier colección de subconjuntos disjuntos $\alpha_1, \alpha_2, \dots, \alpha_m$ de N_n tales que*

$$k_1 < k_2 < \dots < k_m$$

para todo $k_j \in \alpha_j, j = 1, 2, \dots, m$,

$$X_{\alpha_1} \rightarrow X_{\alpha_2} \rightarrow \dots \rightarrow X_{\alpha_m},$$

forma una cadena de Markov. Esto es, una subcadena de $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \rightarrow X_{n+1}$ es también una cadena de Markov.

Proposición 4.1.3. *Si $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \rightarrow X_{n+1}$ es una cadena de Markov, entonces:*

$$p(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1) = p(X_{n+1} = x_{n+1} | X_n = x_n)$$

Demostración. Se hará por inducción sobre n

i) Supongamos $X_1 \rightarrow X_2 \rightarrow X_3$ es una cadena de Markov.

Si $p(x_2) > 0$, entonces por definición

$$p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_2)$$

$$\frac{p(x_1, x_2, x_3)}{p(x_1, x_2)} = p(x_3|x_2)$$

$$p(x_3|x_2, x_1) = p(x_3|x_2).$$

ii) Supongamos es válido hasta n y veamos que se tiene para $n + 1$. Sea $X_1 \rightarrow \dots \rightarrow X_{n+1}$ una cadena de Markov.

$$\begin{aligned} p(x_{n+1}|x_1, \dots, x_n) &= \frac{p(x_1, \dots, x_{n+1})}{p(x_1, \dots, x_n)} \\ &= \frac{p(x_1, x_2)p(x_2, x_3) \cdots p(x_{n-1}, x_n)p(x_n, x_{n+1})}{p(x_1, \dots, x_n)p(x_2) \cdots p(x_{n-1})p(x_n)}. \end{aligned} \quad (1)$$

Observe que al ser $X_1 \rightarrow X_2 \cdots \rightarrow X_{n-2}$ una cadena de Markov:

$$\frac{p(x_1, x_2)p(x_2, x_3) \cdots p(x_{n-2}, x_{n-1})}{p(x_2) \cdots p(x_{n-2})} = p(x_1, \dots, x_{n-1}). \quad (2)$$

Ahora reemplazando (2) en (1):

$$\begin{aligned} p(x_{n+1}|x_1, \dots, x_n) &= \frac{p(x_1, \dots, x_{n-1})p(x_n|x_{n-1})p(x_{n+1}|x_n)}{p(x_1, \dots, x_n)} \\ &= \left[\frac{p(x_n|x_{n-1})}{p(x_n|x_1, \dots, x_{n-1})} \right] p(x_{n+1}|x_n) \end{aligned} \quad (3)$$

Por hipótesis de inducción al ser $X_1 \rightarrow \dots \rightarrow X_{n-1}$ una cadena de Markov se tiene que $p(x_n|x_{n-1}) = p(x_n|x_1, \dots, x_{n-1})$, así de (3):

$$p(x_{n+1}|x_1, \dots, x_n) = p(x_{n+1}|x_n).$$

Así queda demostrada la proposición. □

La proposición anterior describe lo que es una cadena de Markov. Ahora se mostrará el análisis para construir los I-diagramas para tres variables y cuatro variables que forman una cadena de Markov.

Proposición 4.1.4. *Si $X \rightarrow Y \rightarrow Z$ una cadena de Markov, entonces:*

$$I(X \wedge Z|Y) = 0$$

Demostración. Como

$$\begin{aligned} I(X \wedge Z|Y) &= \sum_j p(y_j) I(X \wedge Z|Y = y_j) \\ &= \sum_j p(y_j) \sum_{i,k} p(x_i, z_k|y_j) \log \frac{p(x_i, z_k|y_j)}{p(x_i|y_j)p(z_k|y_j)} \end{aligned} \quad (1)$$

y

$$\begin{aligned} \frac{p(x_i, z_k|y_j)}{p(x_i|y_j)p(z_k|y_j)} &= \frac{\frac{p(x_i, y_j, z_k)}{p(y_j)}}{\frac{p(x_i, y_j)}{p(y_j)} \frac{p(z_k, y_j)}{p(y_j)}} \\ &= \frac{p(x_i, y_j, z_k)p(y_j)}{p(x_i, y_j)p(z_k, y_j)} \\ &= \frac{p(x_i, y_j)p(z_k|y_j)p(y_j)}{p(x_i, y_j)p(z_k, y_j)}, \text{ por hipótesis.} \\ &= 1 \end{aligned} \quad (2)$$

sustituyendo (2) en (1) se tiene que:

$$I(X \wedge Z|Y) = 0.$$

□

Debido a la proposición anterior, si $X \rightarrow Y \rightarrow Z$ es una cadena de Markov, la región en el I -diagrama correspondiente a $I(X \wedge Z|Y)$ se puede contraer. Así el I -diagrama para la cadena de Markov de tres variables aleatorias queda representado como se muestra a continuación:

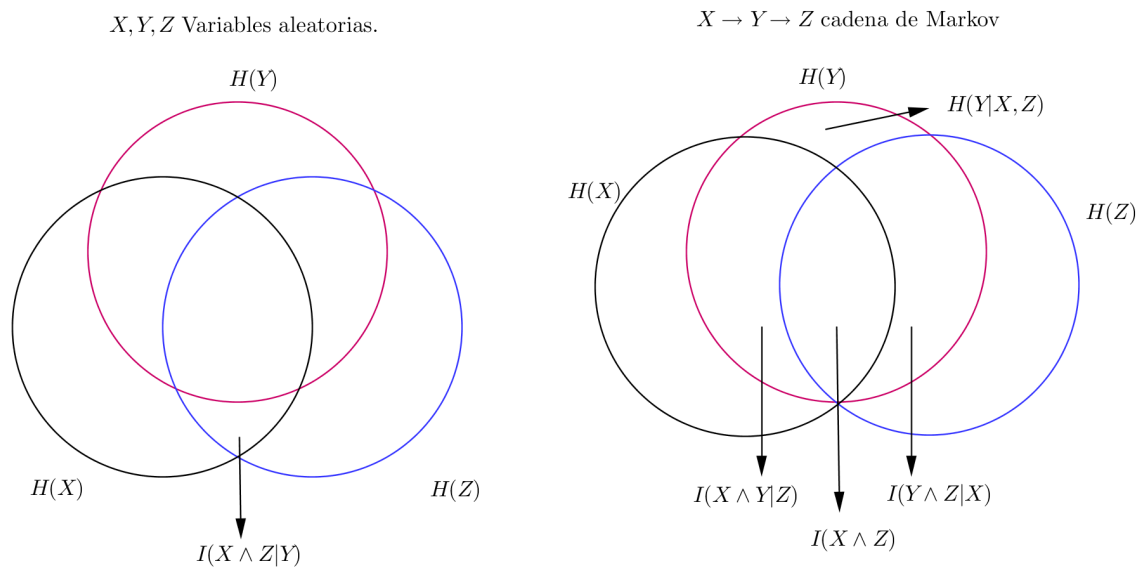


Figura 4-7.:

En la siguiente proposición se establecerán ciertos hechos que se implican debido al I-diagrama para las cadenas de Markov de tres variables aleatorias. Note que en el I-diagrama para la cadena de Markov $X \rightarrow Y \rightarrow Z$, todas las áreas de los elementos del campo F_3 son no negativas.

Proposición 4.1.5. Si $X \rightarrow Y \rightarrow Z$ es una cadena de Markov, entonces:

- a) $I(X \wedge Y) \geq I(X \wedge Z)$.
- b) $H(X|Y) \geq H(X|Z)$.
- c) $I(X \wedge Y) \geq I(Y \wedge Z)$ si y sólo si $I(X \wedge Y|Z) \geq I(Y \wedge Z|X)$.
- d) $I(X \wedge Y) \leq I(Y \wedge Z)$ si y sólo si $I(X \wedge Y|Z) \leq I(Y \wedge Z|X)$.

Demostración. Dado el I-diagrama para $X \rightarrow Y \rightarrow Z$ cadena de Markov:

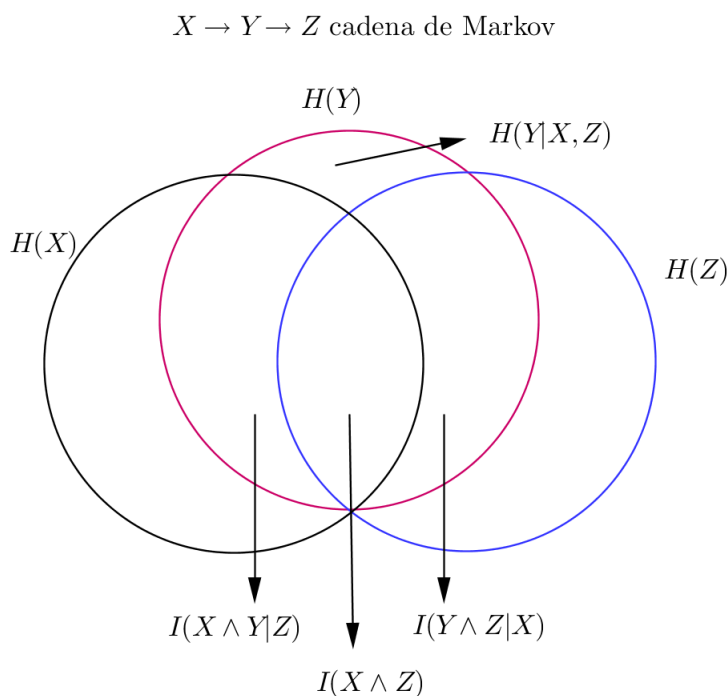


Figura 4-8.:

Como se dijo antes las áreas sobre los conjuntos representados en el I -diagrama son no negativas, al ser entropías condicionales e informaciones mutuas condicionales.

- i) Dado que $\mu^*(\tilde{X} \cap \tilde{Y}) = \mu^*(\tilde{X} \cap \tilde{Y} \cap \tilde{Z}) + \mu^*(\tilde{X} \cap \tilde{Y} - \tilde{Z})$, interpretando el significado de la I -medida:

$$\begin{aligned} I(X \wedge Y) &= I(X \wedge Y \wedge Z) + I(X \wedge Y|Z) \\ &= I(X \wedge Z) + I(X \wedge Y|Z), \text{ por la proposición 4.1.4} \\ &\geq I(X \wedge Z), \text{ ya que } I(X \wedge Y|Z) \geq 0 \end{aligned}$$

- ii) Del I -diagrama:

$$\tilde{X} - \tilde{Z} = (\tilde{X} - \tilde{Y}) \cup (\tilde{X} \cap \tilde{Y} - \tilde{Z})$$

Interpretando μ^* , de lo anterior:

$$H(X|Z) = H(X|Y) + I(X \wedge Y|Z) \geq H(X|Y),$$

ya que $I(X \wedge Y|Z) \geq 0$.

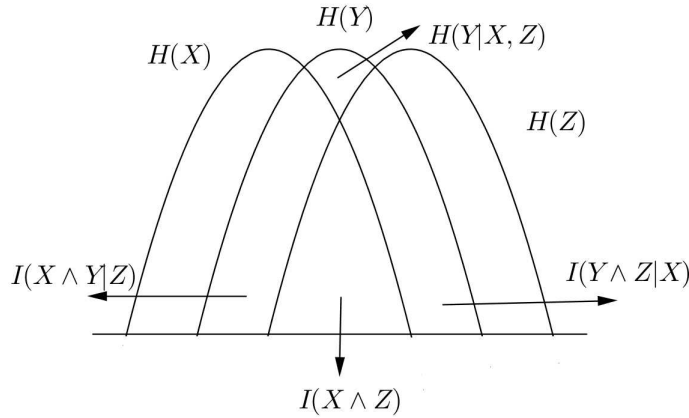
- iii) Las proposiciones $c)$ y $d)$, se tienen del I -diagrama dado que:

$$\begin{aligned} \tilde{X} \cap \tilde{Y} &= (\tilde{X} \cap \tilde{Y} - \tilde{Z}) \cup (\tilde{Y} \cap \tilde{Z}) \\ \tilde{Y} \cap \tilde{Z} &= (\tilde{Y} \cap \tilde{Z} - \tilde{X}) \cup (\tilde{Y} \cap \tilde{Z}) \end{aligned}$$

□

Observación:

La figura 4-8 presentada por Yeung en 1991, fue presentada de diferente forma en su libro *A First Course in Information Theory* (2008), con el fin de llegar a un esquema general para un número n de variables aleatorias. Así el I-diagrama para 3 variables aleatorias es:

**Figura 4-9.:**

Así, hemos visto el caso de tres variables aleatorias que forman una cadena de Markov. A continuación se verá el caso de cuatro variables aleatorias que forman una cadena de Markov, de forma similar se verá cuales conjuntos tienen I -medida nula, y por lo cual se pueden eliminar de la representación en el I -diagrama.

Proposición 4.1.6. Si $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$ una cadena de Markov, entonces:

- a) $I(X_1 \wedge X_3 | X_2) = 0$.
- b) $I(X_1 \wedge X_4 | X_2) = 0$.
- c) $I(X_1 \wedge X_4 | X_3) = 0$.
- d) $I(X_2 \wedge X_4 | X_3) = 0$.
- e) $I(X_1, X_2 \wedge X_4 | X_3) = 0$.
- f) $I(X_1 \wedge X_3 | X_2, X_4) = I(X_1 \wedge X_4 | X_2, X_3) = I(X_2 \wedge X_4 | X_1, X_3) = I(X_2 \wedge X_3 \wedge X_4 | X_1) = I(X_1 \wedge X_2 \wedge X_4 | X_3) = 0$.

Demostración. En la demostración de esta proposición, se utilizará reiteradamente la proposición 4.1.2 enunciada acerca de las subcadenas de Markov.

- a) Al ser $X_1 \rightarrow X_2 \rightarrow X_3$ una cadena de Markov, $I(X_1 \wedge X_3|X_2) = 0$, por proposición 4.1.4. Por lo tanto:

$$0 = I(X_1 \wedge X_3|X_2) = I(X_1 \wedge X_3 \wedge X_4|X_2) + I(X_1 \wedge X_3|X_2, X_4) \quad (1)$$

- b) Al ser $X_1 \rightarrow X_2 \rightarrow X_4$ una cadena de Markov, $I(X_1 \wedge X_4|X_2) = 0$, por proposición 4.1.4. Por lo tanto:

$$0 = I(X_1 \wedge X_4|X_2) = I(X_1 \wedge X_3 \wedge X_4|X_2) + I(X_1 \wedge X_4|X_2, X_3) \quad (2)$$

- c) Al ser $X_1 \rightarrow X_3 \rightarrow X_4$ una cadena de Markov, $I(X_1 \wedge X_4|X_3) = 0$, por proposición 4.1.4. Por lo tanto:

$$0 = I(X_1 \wedge X_4|X_3) = I(X_1 \wedge X_2 \wedge X_4|X_3) + I(X_1 \wedge X_4|X_2, X_3) \quad (3)$$

- d) Al ser $X_2 \rightarrow X_3 \rightarrow X_4$ una cadena de Markov, $I(X_2 \wedge X_4|X_3) = 0$. Por lo tanto:

$$0 = I(X_2 \wedge X_4|X_3) = I(X_1 \wedge X_2 \wedge X_4|X_3) + I(X_2 \wedge X_4|X_1, X_3) \quad (4)$$

- e) Al ser $(X_1, X_2) \rightarrow X_3 \rightarrow X_4$ una cadena de Markov, $I(X_1, X_2 \wedge X_4|X_3) = 0$. Como:

$$\begin{aligned} ((\tilde{X}_1 \cup \tilde{X}_2) \cap \tilde{X}_4) - \tilde{X}_3 &= (\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_4) - \tilde{X}_3 \uplus (\tilde{X}_1 \cap \tilde{X}_4) - (\tilde{X}_2 \cup \tilde{X}_3) \\ &\quad \uplus (\tilde{X}_2 \cap \tilde{X}_4) - (\tilde{X}_1 \cup \tilde{X}_3) \end{aligned}$$

Interpretando la anterior igualdad en términos de la I -medida μ^* :

$$\begin{aligned} 0 &= I(X_1, X_2 \wedge X_4|X_3) = \\ &\quad I(X_1 \wedge X_2 \wedge X_4|X_3) + I(X_1 \wedge X_4|X_2, X_3) + I(X_2 \wedge X_4|X_1, X_3) \end{aligned} \quad (5)$$

- f) ■ De (1) y (2):

$$\begin{aligned} I(X_1 \wedge X_3|X_2, X_4) &= -I(X_1 \wedge X_3 \wedge X_4|X_2) \\ I(X_1 \wedge X_4|X_2, X_3) &= -I(X_1 \wedge X_3 \wedge X_4|X_2) \end{aligned}$$

Es decir:

$$I(X_1 \wedge X_3|X_2, X_4) = I(X_1 \wedge X_4|X_2, X_3) \quad (6)$$

- Ahora de (3) y (6) se tiene:

$$I(X_1 \wedge X_2 \wedge X_4 | X_3) = -I(X_1 \wedge X_3 | X_2, X_4), \quad (7)$$

Y de (4) y (7):

$$I(X_2 \wedge X_4 | X_1, X_3) = I(X_1 \wedge X_3 | X_2, X_4), \quad (8)$$

- Si se reemplaza (6), (7), (8) en (5)

$$0 = -I(X_1 \wedge X_3 | X_2, X_4) + I(X_1 \wedge X_3 | X_2, X_4) + I(X_1 \wedge X_3 | X_2, X_4),$$

es decir $I(X_1 \wedge X_3 | X_2, X_4) = 0$, de lo anterior y (1), (6), (7), (8):

$$\begin{aligned} I(X_1 \wedge X_3 | X_2, X_4) &= I(X_1 \wedge X_4 | X_2, X_3) = I(X_2 \wedge X_4 | X_1, X_3) \\ &= I(X_2 \wedge X_3 \wedge X_4 | X_1) = I(X_1 \wedge X_2 \wedge X_4 | X_3) = 0 \end{aligned}$$

□

Probando la proposición anterior se ha mostrado que los átomos donde la I - medida es cero, para el caso de que $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$ sea una cadena de Markov, son:

$$\begin{aligned} \mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4^c) &= I(X_1 \wedge X_3 | X_2, X_4) = 0 \\ \mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4) &= I(X_1 \wedge X_3 \wedge X_4 | X_2) = 0 \\ \mu^*(\tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3^c \cap \tilde{X}_4) &= I(X_1 \wedge X_4 | X_2, X_3) = 0 \\ \mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4) &= I(X_1 \wedge X_2 \wedge X_4 | X_3) = 0 \\ \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4) &= I(X_2 \wedge X_4 | X_1, X_3) = 0 \end{aligned}$$

Teniendo en cuenta esto, el I -diagrama para las cuatro variables aleatorias que forman una cadena de Markov $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$ es:

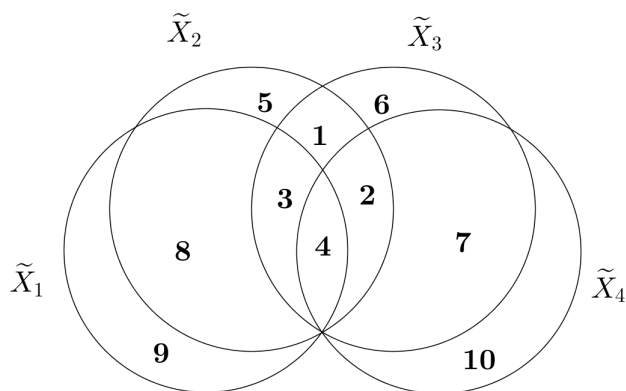


Figura 4-10.:

| Para el átomo enumerado con | El valor de la I -medida es |
|-----------------------------|--|
| 1 | $I(X_2 \wedge X_3 X_1, X_4)$ |
| 2 | $I(X_2 \wedge X_3 \wedge X_4 X_1) = I(X_2 \wedge X_4 X_1)$ |
| 3 | $I(X_1 \wedge X_2 \wedge X_3 X_4) = I(X_1 \wedge X_3 X_4)$ |
| 4 | $I(X_1 \wedge X_2 \wedge X_3 \wedge X_4) = I(X_1 \wedge X_4)$ |
| 5 | $H(X_2 X_1, X_3, X_4) = H(X_2 X_1, X_3)$ |
| 6 | $H(X_3 X_1, X_2, X_4) = H(X_3 X_2, X_4)$ |
| 7 | $I(X_3 \wedge X_4 X_1, X_2) = I(X_3 \wedge X_4 X_2)$ |
| 8 | $I(X_1 \wedge X_2 X_3, X_4) = I(X_1 \wedge X_2 X_3)$ |
| 9 | $H(X_1 X_2, X_3, X_4) = H(X_1 X_2)$ |
| 10 | $H(X_4 X_1, X_2, X_3) = H(X_4 X_3)$ |

Así, la I -medida sobre los átomos no vacíos es no negativa. Además, se tiene el siguiente teorema, el cual se mostrará a partir de la enumeración dada a los átomos en la tabla anterior.

Proposición 4.1.7. 1. $I(X_2 \wedge X_3) \geq I(X_1 \wedge X_4)$

$$2. I(X_2 \wedge X_3) = I(X_1 \wedge X_4) + I(X_1 \wedge X_3|X_4) + I(X_2 \wedge X_4|X_1) + I(X_2 \wedge X_3|X_1, X_4)$$

$$3. H(X_2, X_4|X_1, X_3) = H(X_2|X_1, X_3) + H(X_4|X_3)$$

$$4. H(X_2|X_1, X_4) = H(X_2|X_1, X_3) + I(X_2 \wedge X_3|X_1, X_4)$$

Demostración. 1. Esto se debe a que:

$$\mu^*(\tilde{X}_2 \cap \tilde{X}_3) = \mu^*(1) + \mu^*(2) + \mu^*(3) + \mu^*(4)$$

y

$$\mu^*(4) = I(X_1 \cap X_4).$$

Así, $\mu^*(\tilde{X}_2 \cap \tilde{X}_3) \geq \mu^*(\tilde{X}_1 \cap \tilde{X}_4)$, e interpretando el significado de la I-medida se tiene la igualdad.

2. Del I-diagrama:

$$\mu^*(\tilde{X}_2 \cap \tilde{X}_3) = \mu^*(4) + \mu^*(3) + \mu^*(2) + \mu^*(1)$$

Y como:

$$\mu^*(4) = I(X_1 \wedge X_4)$$

$$\mu^*(3) = I(X_1 \wedge X_3|X_4)$$

$$\mu^*(2) = I(X_2 \wedge X_4|X_1)$$

$$\mu^*(1) = I(X_2 \wedge X_3|X_1, X_4),$$

se tiene la segunda igualdad.

3.

$$\begin{aligned} H(X_2, X_4|X_1, X_3) &= \mu^*((\tilde{X}_2 \cup \tilde{X}_4) - (\tilde{X}_1 \cup \tilde{X}_3)) \\ &= \mu^*((\tilde{X}_2 \cup \tilde{X}_4) \cap \tilde{X}_1^c \cap \tilde{X}_3^c) \\ &= \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4) + \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \tilde{X}_3^c \cap \tilde{X}_4) + \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \tilde{X}_3^c \cap \tilde{X}_4) \\ &= H(X_2|X_1, X_3) + H(X_4|X_3) + I(X_2 \wedge X_4|X_1, X_3) \\ &= H(X_2|X_1, X_3) + H(X_4|X_3), \text{ por Proposición 4.1.6} \end{aligned}$$

4.

$$\begin{aligned}
H(X_2|X_1, X_4) &= \mu^*(\tilde{X}_2 - (\tilde{X}_1 \cup \tilde{X}_4)) \\
&= \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_4^c) \\
&= \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3^c \cap \tilde{X}_4^c) + \mu^*(\tilde{X}_1^c \cap \tilde{X}_2 \cap \tilde{X}_3 \cap \tilde{X}_4^c) \\
&= H(X_2|X_1, X_4) = H(X_2|X_1, X_3) + I(X_2 \wedge X_3|X_1, X_4)
\end{aligned}$$

□

Como se observó en este capítulo, no es sencillo encontrar los átomos en los cuales la I -medida se anula, sin embargo, el I -diagrama para una cadena de Markov $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$ tendrá la siguiente forma

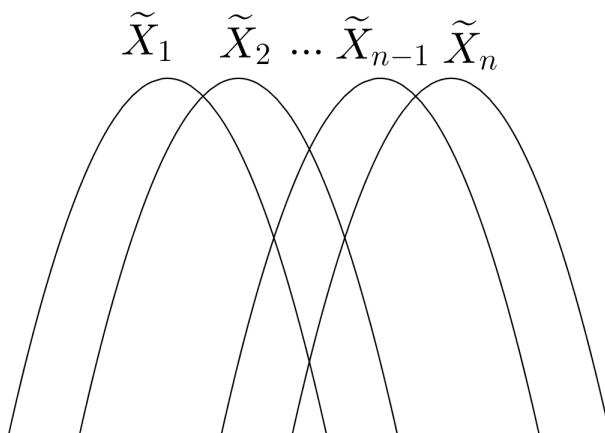


Figura 4-11.:

Esto hecho general es demostrado en el capítulo 7 de [12]

5. Conclusiones y recomendaciones

5.1. Conclusiones

Existe un puente entre las medidas de la información Shannon y la Teoría de Conjuntos, ¿ Se puede de forma análoga realizar una mirada de las medidas de la información generadas por una función polimatroide a través de la Teoría de Conjuntos?. Con el objetivo de dar respuesta a esta pregunta, surge este trabajo. Se llega a la conclusión que bajo ciertas condiciones sobre el conjunto del dominio esta mirada y su construcción siguen siendo validas. Por lo tanto, la I -medida es un caso particular de la f -medida.

Aunque existe dicha conexión entre ambas teorías, en este trabajo se muestra que existen algunas limitaciones, por ejemplo mediante la construcción de la f -medida por cada desigualdad de información no se puede implicar alguna desigualdad en Teoría de Conjuntos.

El estudio anterior motivó la introducción de las definiciones de conjuntos de información generados, conjuntos de información completa y conjuntos mínimos de información completa. A partir de estos conceptos se construyen ciertos resultados similares a los mostrados por Yeung para el caso de los átomos sobre un campo.

El desarrollo de este trabajo, se fundamenta en lo presentado por Yeung en [11] y [12]. Sin embargo, se profundiza en [11], ya que allí se plantean de manera explícita las preguntas que motivan el uso de la Teoría de conjuntos como una herramienta de la Teoría de la Información.

Ya que la entropía hace parte de un conjunto más grande de funciones conocidas como funciones polimatroides; siguiendo ciertas identidades de las medidas de la información Shannon, se definen las medidas de información generadas por una función polimatroide. Se estudian dos ejemplos relativos a matroides y topologías, interpretando las medidas de información generadas por la función rango y f_Z .

5.2. Recomendaciones

Durante la realización de este trabajo surgen algunos cuestionamientos que podrían considerarse en el futuro.

- Estudiar y generalizar las medidas de información generadas por funciones polimatroides en otras teorías.
- Profundizar en el estudio de los conjuntos de información y poder establecer otros resultados.
- Debido a que el I -diagrama para una cadena de Markov tiene una estructura definida, ¿Para qué otras clases de variables aleatorias este hecho es posible?.
- Dada la existencia de los I -diagramas para cualquier número de variables, ¿Qué tan útiles son?.

A. Anexo: Una variación de la fórmula de inclusión-exclusión.

En este anexo, se mostrará que para cada $A \in \mathfrak{A}$, $\mu(A)$ puede expresarse como combinación lineal de $\mu(B)$, $B \in \mathfrak{B}$.

Lema A.0.1. *Si μ es una función aditiva sobre conjuntos disyuntos*

$$\mu\left(\bigcap_{k=1}^n A_k - B\right) = \sum_{1 \leq i \leq n} \mu(A_i - B) - \sum_{1 \leq i < j \leq n} \mu((A_i \cup A_j) - B) + \dots + (-1)^{n+1} \mu((A_1 \cup \dots \cup A_n) - B)$$

Demostración. (Por inducción sobre n)

Para $n = 1$ es inmediato. Se verá que se tiene para $n = 2$, Se tiene que:

$$A_1 - B = (A_1 - (A_2 \cup B)) \cup ((A_1 \cap A_2) - B) \quad (\text{A-1})$$

$$A_2 - B = (A_2 - (A_1 \cup B)) \cup ((A_1 \cap A_2) - B) \quad (\text{A-2})$$

$$(A_1 \cup A_2) - B = (A_1 - (A_2 \cup B)) \cup ((A_1 \cap A_2) - B) \cup (A_2 - (A_1 \cup B)) \quad (\text{A-3})$$

De (A-1), (A-2) y (A-3)

$$\mu((A_1 \cup A_2) - B) = (\mu(A_1 - B) - \mu((A_1 \cap A_2) - B)) + \mu((A_1 \cap A_2) - B) \quad (\text{A-4})$$

$$+ \mu(A_2 - B) - \mu((A_1 \cap A_2) - B) \quad (\text{A-5})$$

$$= \mu(A_1 - B) + \mu(A_2 - B) - \mu((A_1 \cap A_2) - B) \quad (\text{A-6})$$

Así de (A-5) despejando $\mu((A_1 \cap A_2) - B)$

$$\mu((A_1 \cap A_2) - B) = \mu(A_1 - B) + \mu(A_2 - B), \quad (\text{A-7})$$

es decir se tiene lo pedido para $n = 2$.

Ahora, supongamos que el lema es válido hasta n ($n > 2$) y se demostrará para $n + 1$.

$$\begin{aligned} \mu \left(\bigcap_{k=1}^{n+1} A_k - B \right) &= \mu \left(\left(\bigcap_{k=1}^n A_k \right) \cap A_{n+1} - B \right) \\ &= \mu \left(\bigcap_{k=1}^n A_k - B \right) + \mu(A_{n+1} - B) \\ &\quad - \mu \left(\left(\bigcap_{k=1}^n A_k \right) \cup A_{n+1} - B \right), \end{aligned}$$

Lo anterior se tiene por (A-8) y (A-7)

$$\begin{aligned} &= \left\{ \sum_{1 \leq i \leq n} \mu(A_i - B) - \sum_{1 \leq i < j \leq n} \mu(A_i \cup A_j - B) + \dots \right. \\ &\quad \left. (-1)^{n+1} \mu(A_1 \cup \dots \cup A_n - B) \right\} + \mu(A_{n+1} - B) \\ &\quad - \mu \left(\bigcap_{k=1}^n (A_k \cup A_{n+1}) - B \right), \end{aligned}$$

por hipótesis de inducción

$$\begin{aligned} &= \left\{ \sum_{1 \leq i \leq n} \mu(A_i - B) - \sum_{1 \leq i < j \leq n} \mu(A_i \cup A_j - B) + \dots \right. \\ &\quad \left. (-1)^{n+1} \mu(A_1 \cup \dots \cup A_n - B) \right\} + \mu(A_{n+1} - B) \\ &\quad - \left\{ \sum_{1 \leq i \leq n} \mu(A_i \cup A_{n+1} - B) - \sum_{1 \leq i < j \leq n} \mu(A_i \cup A_j \cup A_{n+1} - B) + \dots \right. \\ &\quad \left. (-1)^{n+1} \mu(A_1 \cup \dots \cup A_n \cup A_{n+1} - B) \right\}, \quad (\text{A-8}) \end{aligned}$$

Note que si $C \in \mathfrak{B}$, entonces

$$\mu(C - B) = \mu(C \cup B) - \mu(B),$$

de esto y (A-8):

$$\begin{aligned} \mu \left(\bigcap_{k=1}^{n+1} A_k - B \right) &= \sum_{1 \leq i \leq n+1} \mu(A_i - B) - \sum_{1 \leq i < j \leq n+1} \mu((A_i \cup A_j) - B) \\ &\quad + \dots + (-1)^{n+2} \mu((A_1 \cup \dots \cup A_{n+1}) - B) \end{aligned}$$

Con esta igualdad se concluye la prueba para $n + 1$. □

Bibliografía

- [1] ABRAMSON, N.M.: *Information Theory and Coding*. 1963
- [2] DÍAZ, José M.: *Actas del Primer Encuentro Internacional de Expertos en Teorías de La Información . Un enfoque interdisciplinar*. Nov 6-8, 2008. – 1–19. p.
- [3] DING, Hu G.: *On the amount of Information*. 1962. – 447–455. p.
- [4] FUJISHIGE, Satoru: *Polymatroidal Dependence Structure of a Set of Random Variables..* Vol. 39. 1978. – 55–72. p.
- [5] JAMES, Oxley: *Matroid Theory*. Oxford Mathematics, 1992.
- [6] LEGUIZAMÓN, Leonardo R.: *Una nueva construcción de los espacios topológicos Finitos desde las funciones submodulares*. Colombia, 2012.
- [7] N, Rubiano G.: *Topología General*. Universidad Nacional de Colombia, 2002.
- [8] PAPOULIS, A.: *Probability Random Variables and Stochastic Process*. 1984
- [9] REZA, F.M.: *An Introduction to Information Theory*. 1961
- [10] RIOS, Arley Ramses G.: *Funciones Submodulares y algunas aplicaciones*. Colombia, 2011.
- [11] R.W.YEUNG: *A new outlook on Shannon's information measures*. Vol 37, pg 687-691 : IEE.Trans.Inform.Theory, 1991.
- [12] R.W.YEUNG: *A First Course in Information Theory*. Springer, 2008.
- [13] ZAPATA, Humberto S.: *An Introduction to Network Coding*. Colombia, 2009.