ALGUNAS APLICACIONES DE LAS DESIGUALDADES NO-SHANNON A LA TEORÍA DE GRUPOS.

Daniel Felipe Castro Ovalle.
Matemático

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Matemáticas
Bogotá, D.C.
Noviembre de 2011

ALGUNAS APLICACIONES DE LAS DESIGUALDADES NO-SHANNON A LA TEORÍA DE GRUPOS.

Daniel Felipe Castro Ovalle.
Matemático

Trabajo de tesis para optar al título de
Magister en Ciencias - Matemáticas

Director
Humberto Sarria Zapata, Ph.D.
Profesor Asociado, Universidad Nacional de Colombia.

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Matemáticas
Bogotá, D.C.
Noviembre de 2011

Título en español

Algunas aplicaciones de las desigualdades no-Shannon a la Teoría de Grupos.

Title in English

Some applications of non-Shannon inequalities to the Theory of Groups.

Resumen: En este trabajo se hace un estudio de una relación encontrada en el 2002 por T. H. Chan y R. W. Yeung [7] entre la Teoría de Grupos y la Teoría de la Información, la cual relaciona el espacio de las funciones de entropía con el espacio de las funciones de entropía grupo-caracterizables. Una consecuencia de esta relación es que a cada desigualdad de la información le corresponde cierta desigualdad de la Teoría de grupos y viceversa. Además, se presenta un estudio de algunas desigualdades no-Shannon encontradas por Z. Zhang y R. W. Yeung [8], R. Dougherty, C. Freiling y K. Zeger [4] y F. Matus [2], las cuales producen ciertas desigualdades de la Teoría de Grupos que eran desconocidas. Se estudian también algunas propiedades del espacio de las funciones de entropía para n variables aleatorias Γ_n^*, hecho por Raymond W. Yeung [9], [10] y F. Matus [2].

Abstract: This paper makes a study of a relationship found in 2002 by T. H. Chan and R. W. Yeung [7] between Group Theory and Information Theory, which relates the space of entropy functions with space of entropy group-characterizable functions. A consequence of this relationship is that each information inequality can correspond to some inequality of the Group Theory and vice versa. We also present a study of some non-Shannon inequalities found by Z. Zhang and R. W. Yeung [8], R. Dougherty, C. Freiling, and K. Zeger [4] and F.Matus [2], which produces certain inequalities of the Group Theory that were unknown. Also studied some properties of space of entropy functions for n random variables Γ_n^*, made by Raymond W. Yeung [9], [10] and F. Matus [2].

Palabras clave: Medidas de información, entropía, vector entrópico, función de entropía, espacio de las funciones de entropía, desigualdades tipo Shannon, desigualdades tipo no Shannon, Teoría de grupos .

Keywords: Information measures, entropy, entropic vector, entropy function, space of entropy functions, Shannon type inequalities, non-Shannon type inequalities, Group Theory.

Dedicado a

Mis Padres, quienes me inculcaron la importancia del estudio para mi desarrollo personal.

Agradecimientos

Al profesor Humberto Sarria Zapata por guiarme a través de esté proceso y motivarme en la búsqueda minuciosa no solo de los conocimientos propios de la disciplina si no también en el perfeccionamiento del estilo escrito.
A Gabriela Estevez por la paciencia y apoyo en el transcurso del desarrollo de este proyecto.

Índice general

Índice general **I**

Introducción **III**

1. Medidas de la información **1**

1.1. Medidas de la información Shannon 1

1.2. La I-Medida 12

1.2.1. Preliminares 12

1.2.2. La I-Medida para dos variables aleatorias 13

1.2.3. Construcción de la I-medida μ^* 15

1.3. Desigualdades de la Información 19

1.3.1. La región Γ_n^* 21

1.3.2. Expresiones de información en forma canónica 23

1.3.3. Marco geométrico 25

1.4. Desigualdades tipo Shannon 26

1.4.1. Las desigualdades elementales 26

1.4.2. Enfoque de la Programación Lineal 28

2. Más allá de las desigualdades tipo Shannon **32**

2.1. Caracterizaciones de Γ_2^*, Γ_3^* y Γ_n^* 32

2.2. Desigualdades tipo no-Shannon sin restricciones 37

3. Entropía y Grupos **47**

3.1. Nociones Básicas de Teoría de Grupos 48

3.2. Funciones de entropía grupo-caracterizables. 49

3.3. Una caracterización con grupos de $\overline{\Gamma}_n^*$. 52

3.4. Desigualdades de información y desigualdades de grupo. 59
3.5. Otros resultados. 62

4. Conclusiones **68**

5. Trabajo futuro **69**

6. Bibliografía **70**

Introducción

En este trabajo se hace un estudio de una relación encontrada en el 2002 por T. H. Chan y R. W. Yeung [7] entre la Teoría de Grupos y la Teoría de la Información, la cual relaciona el espacio de las funciones de entropía con el espacio de las funciones de entropía grupo-caracterizables. Como una consecuencia de esta relación se tiene que a cada desigualdad de la información le corresponde cierta desigualdad de la Teoría de grupos y viceversa.

La Teoría de la información fue creada en 1948 por Shannon [6] con el fin de determinar la cantidad de código mínima necesaria para representar un mensaje de manera óptima, de tal forma que, éste se pueda enviar por un canal en forma eficiente. Shannon introduce algunas medidas de información, las cuales se relacionan de cierta forma, produciendo las desigualdades e igualdades de la información. La búsqueda de éstas ha sido impulsada por la necesidad de resolver diversos problemas de comunicación. Actualmente se conocen una gran cantidad de desigualdades e igualdades presentes en la Teoría de la información.

En este trabajo, se presenta un estudio de algunas desigualdades no-Shannon encontradas por Z. Zhang y R. W. Yeung [8], R. Dougherty, C. Freiling y K. Zeger [4] y F. Matus [2], las cuales producen ciertas desigualdades de la Teoría de Grupos que eran desconocidas. Además, se estudian algunas propiedades del espacio de las funciones de entropía para n variables aleatorias Γ_n^*, basándose en los trabajos de Raymond W. Yeung [9], [10] y F. Matus [2].

El presente trabajo se encuentra organizado de la siguiente manera: en el Capítulo 1, se presentan algunas nociones básicas de la Teoría de la Información, tales como: medidas de la información, desigualdades e igualdades de la información y espacio de las funciones de entropía para n variables aleatorias. En el Capítulo 2, se presentan algunas desigualdades no-Shannon y se da una caracterización del espacio de las funciones de entropía para n variables, mediante las desigualdades de la información. En el Capítulo 3, se introducen algunos conceptos de la Teoría de Grupos, se muestra la relación entre la Teoría de Grupos y la Teoría de la Información, y se presentan algunas aplicaciones de las desigualdades de la información a la Teoría de Grupos.

CAPÍTULO 1

Medidas de la información

En este Capítulo se estudiarán cinco medidas de información básicas, éstas son: la entropía, la entropía condicional, la información mutua y la información mutua condicional. Para el desarrollo de casi todo este Capítulo, se tomó como referencia el trabajo de Raymond W. Yeung en [9] y [10].

1.1. Medidas de la información Shannon

Sea X una variable aleatoria que toma valores en un alfabeto $\mathcal{X}$, enumerable. La distribución de probabilidad de X se denota por $\{p_X(x) : x \in \mathcal{X}\}$, donde $p_X(x) := Pr\{X = x\}$. Cuando no exista ambigüedad, $p_X(x)$ será abreviado como $p(x)$. El soporte de X, denotado por $\mathcal{S}_X$, corresponde al conjunto de todos los $x \in \mathcal{X}$ tales que $p(x) > 0$. Si $\mathcal{S}_X = \mathcal{X}$, diremos que p es estrictamente positivo.

Nota 1. *De aquí en adelante, supondremos, sin pérdida de generalidad, que el alfabeto es infinito enumerable, a menos que se diga lo contrario.*

Definición 2. *Sean X, Y, Z variables aleatorias y $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ sus respectivos alfabetos, con distribución de probabilidad*

$$\{p(x,y,z) : (x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}\}, \quad \textit{con } p(x,y,z) = Pr(X = x, Y = y, Z = z).$$

$$(\sum_{x,y,z} p(x,y,z) = 1).$$

Se define

$$p(x,y) := \sum_{z} p(x,y,z),$$

$$p(x) := \sum_{y,z} p(x,y,z),$$

$$p(x|y) := \frac{p(x,y)}{p(y)}; \quad \textit{si } p(y) \neq 0,$$

$$p(x,y|z) := \frac{p(x,y,z)}{p(z)}; \quad si\ p(z) \neq 0,$$

y

$$p(x|y,z) := \frac{p(x,y,z)}{p(y,z)}; \quad si\ p(y,z) \neq 0.$$

Definición 3. *Dos variables aleatorias X y Y son independientes, si*

$$p(x,y) = p(x)p(y),$$

para todo x y y (e.d. para todo $(x,y) \in \mathcal{X} \times \mathcal{Y}$).

Para más de dos variable aleatorias, aparecen dos tipos de independencia

Definición 4. *(Independencia Mutua) Para $n \geq 2$, las variables aleatorias $X_1, X_2, \ldots, X_n$ son mutuamente independientes, si*

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n),$$

para todo $x_1, x_2, \ldots, x_n$.

Definición 5. *(Independencia dos a dos) Para $n \geq 2$, el conjunto de las variables aleatorias $X_1, X_2, \ldots, X_n$ es independiente dos a dos, si cada par de variables X_i y X_j son independientes, para todo $1 \leq i < j \leq n$.*

Definición 6. *La entropía $H(X)$ de una variable aleatoria X que toma valores en un alfabeto $\mathcal{X}$, es el valor*

$$H(X) := -\sum_{x \in \mathcal{S}_X} p(x) \log p(x). \tag{1.1}$$

Para simplificar la notación, (1.1) se escribe como

$$H(X) := -\sum_{x} p(x) \log p(x). \tag{1.2}$$

En todas las definiciones de medida de información, se asumirá que la suma se toma sobre el soporte correspondiente, ya que $p(x)\log p(x)$ en (1.1) no está definido, si $p(x) = 0$. La base del logaritmo en (1.1) se escoge como cualquier número real mayor que 1 conveniente. Se escribe $H(X)$ como $H_\alpha(X)$, cuando la base del logaritmo es α. Cuando la base del logaritmo es 2, la unidad para la entropía se llama un bit. Si la base del logaritmo es un entero $D > 2$, la unidad de la entropía se llama un $D-it$. En el contexto del código fuente, la base se interpreta usualmente como el tamaño del código del alfabeto.

Ejemplo 7. *Considérese el experimento de lanzar una moneda legal hasta que ésta caiga en cara. La variable aleatoria X, representa el número de lanzamientos necesarios para que la moneda caiga en cara, por lo tanto X toma valores sobre el alfabeto $\mathcal{X} = \{1, 2, \ldots\}$ y la distribución de probabilidad de X es $\{p(n) := Pr\{X = n\} = \frac{1}{2^n} : n \in \mathcal{X}\}$. Si se calcula la entropía de X en base 2, ésta es igual a:*

$$H(X) = -\sum_{n=1}^{\infty} p(n) \log_2 p(n) = -\sum_{n=1}^{\infty} \frac{1}{2^n} \log_2 \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{n}{2^n} = 2$$

Definición 8. *La entropía mutua $H(X,Y)$ de un par de variables aleatorias X y Y está dada por*

$$H(X,Y) := -\sum_{x,y} p(x,y)\log p(x,y). \tag{1.3}$$

Definición 9. *Para variables aleatorias X y Y, la entropía condicional de Y dado X es*

$$H(Y|X) := -\sum_{x,y} p(x,y)\log p(y|x). \tag{1.4}$$

(1.4) puede escribirse como

$$H(Y|X) = \sum_{x} p(x)\Big[-\sum_{y} p(y|x)\log p(y|x)\Big]. \tag{1.5}$$

Por lo tanto,

$$H(Y|X) = \sum_{x} p(x)H(Y|X=x), \tag{1.6}$$

donde

$$H(Y|X=x) := -\sum_{y} p(y|x)\log p(y|x), \tag{1.7}$$

es la entropía de Y dado que $X=x$.

Similarmente se definen

Definición 10. *Para las variables aleatorias X, Y y Z se define la entropía mutua condicional de (X,Y) dado Z como*

$$H(X,Y|Z) := \sum_{x,y,z} p(x,y,z)\log p(x,y|z). \tag{1.8}$$

(1.8) puede escribirse como

$$H(X,Y|Z) = \sum_{z} p(z)\Big[-\sum_{x,y} p(x,y|z)\log p(x,y|z)\Big]. \tag{1.9}$$

Por lo tanto,

$$H(X,Y|Z) = \sum_{z} p(z)H(X,Y|Z=z), \tag{1.10}$$

donde

$$H(X,Y|Z=z) := -\sum_{x,y} p(x,y|z)\log p(x,y|z), \tag{1.11}$$

es la entropía de (X,Y) dado que $Z=z$.
La entropía condicional de X dado Y,Z se define como

$$H(X|Y,Z) := \sum_{x,y,z} p(x,y,z)\log p(x|y,z). \tag{1.12}$$

(1.12) puede escribirse como

$$H(X,Y|Z) = \sum_{y,z} p(y,z)\Big[-\sum_{x} p(x|y,z)\log p(x|y,z)\Big]. \tag{1.13}$$

Por lo tanto,

$$H(X,Y|Z) = \sum_{y,z} p(y,z)H(X|Y=y,Z=z), \tag{1.14}$$

donde

$$H(X|Y=y,Z=z) := -\sum_{x} p(x|y,z)\log p(x|y,z). \tag{1.15}$$

es la entropía de X dado que $Y=y$ y $Z=z$.

Definición 11. *Si X y Y son variables aleatorias, la información mutua condicional de X y Y es el valor dado por*

$$I(X;Y) := \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \tag{1.16}$$

Definición 12. *Si X, Y y Z son variables aleatorias, la información mutua de X y Y dado Z es el valor*

$$I(X;Y|Z) := \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}. \tag{1.17}$$

La entropía condicional (1.17) también puede escribirse como

$$I(X;Y|Z) := \sum_{z} p(z)I(X;Y|Z=z), \tag{1.18}$$

donde

$$I(X;Y|Z=z) := \sum_{x,y} p(x,y|z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}. \tag{1.19}$$

Observación 13. *Para simplificar la notación, para variables aleatorias conjuntamente distribuidas, $X_1, X_2, \ldots, X_n$, $Y_1, Y_2, \ldots, Y_m$, y $Z_1, Z_2, \ldots, Z_l$, tomando valores en los alfabetos $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$, $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_m$, y $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_l$, respectivamente. Se define*

$$\begin{aligned} I(X_1, X_2, \ldots X_n; Y_1, Y_2, \ldots Y_m) :=& I((X_1, X_2, \ldots X_n); (Y_1, Y_2, \ldots Y_m)), \\ I(X_1, \ldots X_n; Y_1, \ldots Y_m | Z_1, \ldots Z_l) :=& I((X_1, \ldots X_n); (Y_1, \ldots Y_m) | (Z_1, \ldots Z_l)). \end{aligned}$$

donde la variable aleatoria $(X_1, X_2, \ldots X_n)$ toma valores en el alfabeto $\mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_n$, etc.

El siguiente resultado es una consecuencia de las definiciones anteriores.

Proposición 14. *Las funciones de entropía $H(X), H(X,Y), H(X,Y,Z), H(X|Y)$, $H(X|Y,Z)$ y $H(X,Y|Z)$ son no negativas.*

Ahora se mostrará, algunas formas de relacionar los diferentes tipos de entropía.

Proposición 15. *Sean X, Y y Z variables aleatorias arbitrarias. Entonces, las siguientes igualdades se satisfacen:*

(a)
$$H(X,Y) = H(X) + H(Y|X). \tag{1.20}$$

(b)
$$H(X,Y) = H(Y) + H(X|Y). \tag{1.21}$$

(c)
$$H(X,Y,Z) = H(X) + H(Y,Z|X). \tag{1.22}$$

(d)
$$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y). \tag{1.23}$$

(e)
$$H(X,Y,Z) = H(X,Y) + H(Z|X,Y). \tag{1.24}$$

Demostración. **(a)**

$$\begin{aligned}
H(X,Y) &= -\sum_{x,y} p(x,y)\log p(x,y) \\
&= -\sum_{x,y} p(x,y)\log(p(x)p(y|x)) \\
&= -\sum_{x,y} p(x,y)\log p(x) - \sum_{x,y} p(x,y)\log p(y|x) \\
&= -\sum_{x} p(x)\log p(x) - \sum_{x,y} p(x,y)\log p(y|x) \\
&= H(X) + H(Y|X).
\end{aligned}$$

(b) La prueba es similar a la anterior.

(c)

$$\begin{aligned}
H(X,Y,Z) &= -\sum_{x,y,z} p(x,y,z)\log p(x,y,z) \\
&= -\sum_{x,y,z} p(x,y,z)\log(p(x)p(y,z|x)) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x) - \sum_{x,y,z} p(x,y,z)\log p(y,z|x) \\
&= H(X) + H(Y,Z|X).
\end{aligned}$$

(d)

$$\begin{aligned}
H(X,Y,Z) &= -\sum_{x,y,z} p(x,y,z)\log p(x,y,z) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(p(x)\frac{p(x,y,z)}{p(x,y)}\frac{p(x,y)}{p(x)}\right) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x) - \sum_{x,y,z} p(x,y,z)\log p(z|x,y) \\
&- \sum_{x,y,z} p(x,y,z)\log p(y|x) \\
&= -\sum_{x} p(x)\log p(x) - \sum_{x,y,z} p(x,y,z)\log p(z|x,y) \\
&- \sum_{x,y} p(x,y)\log p(y|x) \\
&= H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y).
\end{aligned}$$

(e)

$$\begin{aligned}
H(X,Y,Z) &= H(X) + H(Y|X) + H(Z|X,Y) && \text{por} \quad (1.23) \\
&= H(X,Y) + H(Z|X,Y). && \text{por} \quad (1.20)
\end{aligned}$$

□

Proposición 16. *Sean X, Y y Z variables aleatorias arbitrarias. Entonces, las siguientes desigualdades se satisfacen:*

(a)

$$H(X|Y) \leq H(X). \tag{1.25}$$

(b)

$$H(X|Y,Z) \leq H(X|Z). \tag{1.26}$$

(c)

$$H(X|Y,Z) \leq H(X,Y). \tag{1.27}$$

(d)

$$H(X,Y|Z) \leq H(X|Z) + H(Y|Z). \tag{1.28}$$

Demostración. Para la demostración de esta proposición, se usará la convexidad de la función logaritmo con base mayor que 1, es decir.

$$\log(\sum_{i=1}^{n} \alpha_i x_i) \geq \sum_{i=1}^{n} \alpha_i \log(x_i), \quad \textit{donde los } \alpha_i > 0 \textit{ y } \sum_{i=1}^{n} \alpha_i = 1.$$

(a)

$$\begin{aligned}
H(X) - H(X|Y) &= -\sum_{x,y} p(x,y)\log p(x) + \sum_{x,y} p(x,y)\log p(x|y) \\
&= -\sum_{x,y} p(x,y)\log\left(\frac{p(x)}{p(x|y)}\right) \\
&\geq -\log\left(\sum_{x,y}\left(p(x,y)\frac{p(x)}{p(x|y)}\right)\right) \\
&\geq -\log\left(\sum_{x,y} p(x)p(y)\right) \\
&= -\log(1) = 0.
\end{aligned}$$

(b)

$$\begin{aligned}
&H(X|Z) - H(X|Y,Z) \\
&= H(X,Z) - H(Z) - H(X|Y,Z) \qquad \text{por} \quad (1.20) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x,z) + \sum_{x,y,z} p(x,y,z)\log p(z) \\
&+ \sum_{x,y,z} p(x,y,z)\log p(x|y,z) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,z)}{p(z)p(x|y,z)}\right) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,z)p(y,z)}{p(z)p(x,y,z)}\right) \\
&\geq -\log\left(\sum_{x,y,z}\left(p(x,y,z)\left(\frac{p(x,z)p(y,z)}{p(z)p(x,y,z)}\right)\right)\right) \\
&= -\log\left(\sum_{x,z}\frac{p(x,z)}{p(z)}\left(\sum_{y} p(y,z)\right)\right) \\
&= -\log\left(\sum_{x,z}\frac{p(x,z)}{p(z)}p(z)\right) \\
&= -\log(1) = 0.
\end{aligned}$$

(c)

$$
\begin{aligned}
&H(X,Y) - H(X|Y,Z) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x,y) + \sum_{x,y,z} p(x,y,z)\log p(x|y,z) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,y)}{p(x|y,z)}\right) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,y)p(y,z)}{p(x,y,z)}\right) \\
&\geq -\log\left(\sum_{x,y,z}\left(p(x,y,z)\left(\frac{p(x,y)p(y,z)}{p(x,y,z)}\right)\right)\right) \\
&= -\log\left(\sum_{x,y,z} p(x,y)p(y,z)\right) \\
&\geq -\log\left(\sum_{x,y,z} p(x,y)p(z)\right) \qquad\qquad (\text{pues } \ p(y,z) \leq p(z)) \\
&= -\log(1) = 0.
\end{aligned}
$$

(d)

$$
\begin{aligned}
&H(X|Z) + H(Y|Z) - H(X,Y|Z) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x|z) - \sum_{x,y,z} p(x,y,z)\log p(y|z) \\
&+ \sum_{x,y,z} p(x,y,z)\log p(x,y|z) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x|z)p(y|z)}{p(x,y|z)}\right) \\
&= -\sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,z)p(y,z)p(z)}{p(z)p(z)p(x,y,z)}\right) \\
&\geq -\log\left(\sum_{x,y,z}\left(p(x,y,z)\left(\frac{p(x,z)p(y,z)}{p(z)p(x,y,z)}\right)\right)\right) \\
&= -\log\left(\sum_{y,z}\frac{p(y,z)}{p(z)}(\sum_{x} p(x,z))\right) \\
&= -\log\left(\sum_{y,z} p(y,z)\right) \\
&= -\log(1) = 0.
\end{aligned}
$$

□

Se verá a continuación, otras caracterizaciones de la información mutua, en términos de las funciones de entropía.

Proposición 17. *Sean X, Y variables aleatorias arbitrarias, las siguientes igualdades se tienen:*

(a)

$$I(X;Y) = H(X) - H(X|Y). \tag{1.29}$$

(b)

$$I(X;Y) = H(Y) - H(Y|X). \tag{1.30}$$

(c)

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{1.31}$$

Demostración. **(a)**

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y) \left(\log \frac{p(x,y)}{p(y)} - \log p(x) \right) \\ &= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x|y) \\ &= H(X) - H(X|Y). \end{aligned}$$

(b) La demostración es similar a la de (a)

(c)

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && \text{por } (1.29) \\ &= H(X) - (H(X,Y) - H(Y)) && \text{por } (1.21) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

□

Proposición 18. *Sean X y Y variables aleatorias arbitrarias. Entonces, $I(X;Y) \geq 0$. La igualdad se alcanza, si y sólo si, X y Y son independientes.*

Demostración. Resulta de (1.29) y de (1.25). □

En el siguiente resultado se muestran otras formas de caracterizar la información mutua condicional.

Proposición 19. *Sean X, Y y Z variables aleatorias arbitrarias. Entonces, las siguientes igualdades se satisfacen:*

(a)

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z). \tag{1.32}$$

(b)

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z). \tag{1.33}$$

(c)

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z). \tag{1.34}$$

(d)

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(Z) - H(X,Y,Z). \tag{1.35}$$

(e)

$$I(X,Y;Z) = I(X;Z) + I(X;Y|Z). \tag{1.36}$$

Demostración. **(a)**

$$\begin{aligned}
I(X;Y|Z) &= -\sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \\
&= -\sum_{x,y,z} p(x,y,z) \log p(x|z) + \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(y|z)} \\
&= H(X|Z) + \sum_{x,y,z} p(x,y,z) \log \left(\frac{p(x,y,z)}{p(z)} \frac{p(z)}{p(y,z)} \right) \\
&= H(X|Z) + \sum_{x,y,z} p(x,y,z) \log p(x|y,z) \\
&= H(X|Z) - H(X|Y,Z).
\end{aligned}$$

(b) La prueba es similar a la anterior.

(c)

$$\begin{aligned}
I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) && \text{por } (1.32) \\
&= H(X|Z) - (H(X,Y,Z) - H(Y,Z)) && \text{por } (1.24) \\
&= H(X|Z) - (H(X,Y|Z) + H(Z)) + H(Y,Z) && \text{por } (1.22) \\
&= H(X|Z) + H(Y|Z) - H(X,Y|Z). && \text{por } (1.21)
\end{aligned}$$

(d)

$$\begin{aligned}
I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) && \text{por } (1.32) \\
&= (H(X,Z) - H(Z)) - (H(X,Y,Z) - H(Y,Z)) && \text{por } (1.20) \text{ y } (1.24) \\
&= H(X,Z) + H(Y,Z) - H(Z) - (H(X,Y,Z).
\end{aligned}$$

(e)

$$\begin{aligned}
I(X,Y;Z) &= H(X,Y) + H(Z) - H(X,Y,Z) && \text{por } (1.31)\\
&= H(X,Y) + H(Z) - H(X,Y,Z)\\
&+ H(X) - H(X) + H(X,Z) - H(X,Z)\\
&= (H(X) + H(Z) - H(X,Z))\\
&+ (H(X,Y) + H(X,Z) - H(X) - H(X,Y,Z))\\
&= I(X;Z) + I(X;Y|Z). && \text{por } (1.31) \text{ y } (1.35)
\end{aligned}$$

□

Proposición 20. *Sean X, Y y Z variables aleatorias arbitrarias. Entonces, $I(X;Y|Z) \geq 0$.*

Demostración. Resulta de (1.32) y de (1.26). □

Gracias a algunas proposiciones anteriores, se puede afirmar que las medidas de la información son no negativas.

Ejemplo 21. *Sean X_1 y X_2 variables aleatorias tomando valores en $\{0,1\}$ independientes con*

$$Pr\{X_i = 0\} = Pr\{X_i = 1\} = 0.5,$$

$i = 1, 2$. Sea

$$X_3 = (X_1 + X_2) mod\ 2, \tag{1.37}$$

es fácil verificar que, X_3 tiene la misma distribución marginal de X_1 y X_2. Así, si todas las entropías se toman en base 2, se tiene que

$$H(X_i) = 1$$

para $i = 1, 2, 3$. Además, X_1, X_2 y X_3 son independientes dos a dos. Por lo tanto,

$$H(X_i, X_j) = 2$$

y

$$I(X_i, X_j) = 0$$

para $1 \leq i < j \leq 3$. Por otra parte, por (1.37) se tiene que cada variable aleatoria es una función de las otras dos variables aleatorias. Entonces, por (1.24)

$$\begin{aligned}
H(X_1, X_2, X_3) &= H(X_1, X_2) + H(X_3|X_1, X_2)\\
&= 2 + 0 = 2.
\end{aligned}$$

Ahora, para $i, j, k \in \{1, 2, 3\}$ diferentes, por (1.35)

$$\begin{aligned}
I(X_i, X_j|X_k) &= H(X_i, X_k) + H(X_j, X_k) - H(X_k) - H(X_i, X_j, X_k)\\
&= 2 + 2 - 1 - 2 = 1.
\end{aligned}$$

1.2. La I-Medida

En esta sección, se desarrollará una teoría que establece una correspondencia entre las medidas de la información Shannon y la Teoría de conjuntos. Con esta correspondencia, toda medida de la información Shannon se pueden interpretar como operaciones entre conjuntos.

1.2.1. Preliminares

Se introducen a continuación, algunos conceptos básicos de la Teoría de la medida.

Definición 22. *El campo $\mathcal{F}_n$ generado por los conjuntos $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$, es la colección de todos los conjuntos que se pueden obtener al aplicar un número finito de operaciones $(\cup, \cap, -, ^c)$ entre los elementos de la colección $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$.*

Definición 23. *Los átomos de $\mathcal{F}_n$, son conjuntos de la forma $\bigcap_{i=1}^n Y_i$ donde Y_i es o $\widetilde{X}_i$ ó $\widetilde{X}_i^c$, el complemento de $\widetilde{X}_i$.*

Hay 2^n átomos y 2^{2^n} conjuntos en $\mathcal{F}_n$. Evidentemente, todos los átomos en $\mathcal{F}_n$ son disjuntos y cada conjunto en $\mathcal{F}_n$ puede expresarse unívocamente como unión de un subconjunto de átomos de $\mathcal{F}_n$. Se adopta la convención de que la unión del subconjunto vacío de átomos de $\mathcal{F}_n$ es el conjunto vacío. Se asume también que, los conjuntos $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ se intersecan unos con otros genéricamente, es decir, los átomos de $\mathcal{F}_n$ son no vacíos a menos que se especifique lo contrario.

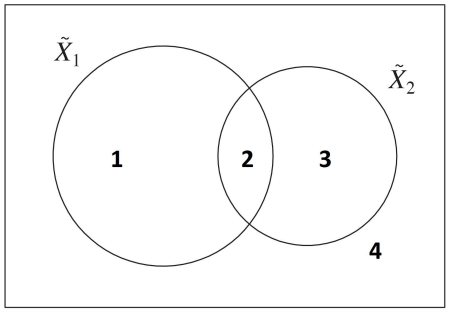

Figura 1.1: Diagrama de Venn para $\widetilde{X_1}$ y $\widetilde{X_2}$.

Ejemplo 24. *Los conjuntos $\widetilde{X}_1$ y $\widetilde{X}_2$ generan el campo $\mathcal{F}_2$. Los átomos de $\mathcal{F}_2$ son:*

$$\widetilde{X}_1 \cap \widetilde{X}_2^c, \quad \widetilde{X}_1 \cap \widetilde{X}_2, \quad \widetilde{X}_1^c \cap \widetilde{X}_2 \text{ y } \widetilde{X}_1^c \cap \widetilde{X}_2^c. \tag{1.38}$$

Estos conjuntos tienen una representación gráfica en términos del diagrama de Venn tal como se muestra en la Figura 1.1. El campo $\mathcal{F}_2$, consiste de las uniones de subconjuntos de átomos en (1.38). Hay un total de 16 conjuntos en $\mathcal{F}_2$, que son precisamente todos los conjuntos que pueden obtenerse de $\widetilde{X}_1$ y $\widetilde{X}_2$ mediante operaciones usuales entre conjuntos.

Definición 25. *Una función real μ definida sobre $\mathcal{F}_n$, se denomina una medida con signo, si es aditiva sobre conjuntos, es decir, para cualquier par de conjuntos disjuntos $A, B \in \mathcal{F}_n$,*

$$\mu(A \cup B) = \mu(A) + \mu(B). \tag{1.39}$$

Para una medida μ con signo se tiene que $\mu(\emptyset) = 0$, lo cual resulta de lo siguiente: para cualquier $A \in \mathcal{F}_n$, $\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset)$.

Una medida con signo μ sobre $\mathcal{F}_n$, está completamente determinada por sus valores en los átomos de $\mathcal{F}_n$. Los valores de μ en los otros conjuntos de $\mathcal{F}_n$ pueden obtenerse por medio de la aditividad.

Observación 26. *Una medida con signo μ sobre $\mathcal{F}_2$ está completamente determinada por los valores*

$$\mu(\widetilde{X}_1 \cap \widetilde{X}_2), \quad \mu(\widetilde{X}_1 \cap \widetilde{X}_2^c), \quad \mu(\widetilde{X}_1^c \cap \widetilde{X}_2), \quad \mu(\widetilde{X}_1^c \cap \widetilde{X}_2^c). \tag{1.40}$$

Por ejemplo, el valor de μ sobre $\widetilde{X}_1$, se puede obtener mediante.

$$\mu(\widetilde{X}_1) = \mu((\widetilde{X}_1 \cap \widetilde{X}_2) \cup (\widetilde{X}_1 \cap \widetilde{X}_2^c)) = \mu(\widetilde{X}_1 \cap \widetilde{X}_2) + \mu(\widetilde{X}_1 \cap \widetilde{X}_2^c). \tag{1.41}$$

1.2.2. La I-Medida para dos variables aleatorias

Se formulará en esta sección, una correspondencia uno a uno entre medidas de la información Shannon y la Teoría de conjuntos para dos variables aleatorias. Sean, X_1 y X_2, variables aleatorias, y, $\widetilde{X}_1$ y $\widetilde{X}_2$ conjuntos, diferentes y no disjuntos. Los conjuntos $\widetilde{X}_1$ y $\widetilde{X}_2$ generan el campo $\mathcal{F}_2$ cuyos átomos se muestran en (1.38). En la formulación que se planteó, el conjunto universal Ω, es el conjunto $\widetilde{X}_1 \cup \widetilde{X}_2$, por razones que serán más claras después. Nótese que con la elección del conjunto universal, el átomo $\widetilde{X}_1^c \cap \widetilde{X}_2^c = \emptyset$ porque $\widetilde{X}_1^c \cap \widetilde{X}_2^c = (\widetilde{X}_1 \cup \widetilde{X}_2)^c = \Omega^c = \emptyset$.
Las medidas de la información Shannon para las variables aleatorias X_1 y X_2 son:

$$H(X_1), \quad H(X_2), \quad H(X_1|X_2), \quad H(X_2|X_1), \quad H(X_1, X_2) \;\; y \;\; I(X_1; X_2). \tag{1.42}$$

Estas inducen una medida con signo μ^* sobre $\widetilde{X}_1 \cup \widetilde{X}_2$ definida por las siguiente igualdades:

$$\mu^*(\widetilde{X}_1 - \widetilde{X}_2) := H(X_1|X_2), \tag{1.43}$$

$$\mu^*(\widetilde{X}_2 - \widetilde{X}_1) := H(X_2|X_1), \tag{1.44}$$

$$\mu^*(\widetilde{X}_1 \cap \widetilde{X}_2) := I(X_1; X_2). \tag{1.45}$$

Estos son los valores de μ^* en los átomos no vacíos de $\mathcal{F}_2$. Los valores de μ^* en los otros conjuntos de $\mathcal{F}_2$ pueden obtenerse vía aditividad:

$$\begin{aligned} \mu^*(\widetilde{X}_1 \cup \widetilde{X}_2) &= \mu^*(\widetilde{X}_1 - \widetilde{X}_2) + \mu^*(\widetilde{X}_2 - \widetilde{X}_1) + \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2) \\ &= H(X_1|X_2) + H(X_2|X_1) + I(X_1; X_2) = H(X_1, X_2). \end{aligned} \tag{1.46}$$

$$\begin{aligned} \mu^*(\widetilde{X}_1) &= \mu^*(\widetilde{X}_1 - \widetilde{X}_2) + \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2) \\ &= H(X_1|X_2) + I(X_1; X_2) = H(X_1). \end{aligned} \tag{1.47}$$

$$\begin{aligned}\mu^*(\widetilde{X}_2) &= \mu^*(\widetilde{X}_2 - \widetilde{X}_1) + \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2) \\ &= H(X_2|X_1) + I(X_1;X_2) = H(X_2).\end{aligned} \tag{1.48}$$

Esto corresponde a las seis medidas de información de Shannon para 2 variables aleatorias. En cada una de las ecuaciones anteriores, el lado derecho y el izquierdo corresponden uno al otro según la siguiente correspondencia de símbolos:

$$\begin{aligned} X_i &\longleftrightarrow \widetilde{X}_i, \\ H, I &\longleftrightarrow \mu^*, \\ ; &\longleftrightarrow \cap, \\ , &\longleftrightarrow \cup, \\ | &\longleftrightarrow -. \end{aligned} \tag{1.49}$$

Nótese que no se hizo distinción entre los símbolos H e I en esta correspondencia. Así, para dos variables aleatorias X_1 y X_2, las medidas de información Shannon pueden considerarse formalmente como una medida con signo sobre $\mathcal{F}_2$. Nos referiremos a μ^* como la I-medida para las variables aleatorias X_1 y X_2.

Como las medidas de la información Shannon, pueden verse como una medida con signo,

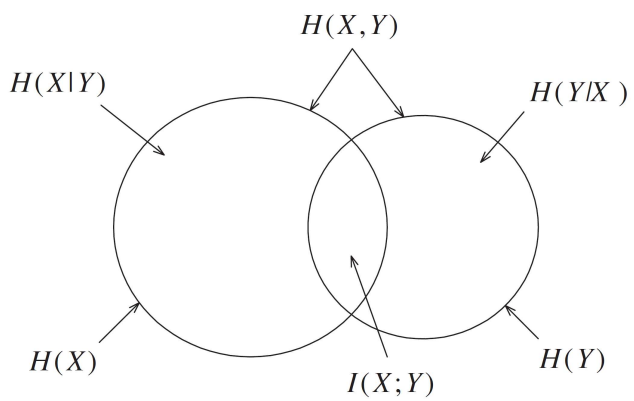

Figura 1.2: Diagrama de información para dos variables aleatorias.

se pueden aplicar las diferentes operaciones de la Teoría de conjuntos a la Teoría de la información. Esto explica porque la Figura 1.2, representa las relaciones entre todas las medidas de información Shannon para dos variables aleatorias correctamente. Como un ejemplo, si se considera la siguiente identidad en conjuntos:

$$\mu^*(\widetilde{X}_1 \cup \widetilde{X}_2) = \mu^*(\widetilde{X}_1) + \mu^*(\widetilde{X}_2) - \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2).$$

Esta identidad, es un caso especial de la fórmula de inclusión-exclusión en la Teoría de conjuntos. Mediante la sustitución de símbolos dada por (1.49), se obtiene la siguiente identidad de información

$$H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1;X_2).$$

Se terminará esta sección con el siguiente comentario: el valor de μ^* en el átomo $\widetilde{X}_1^c \cap \widetilde{X}_2^c$ aparentemente no tiene ningún significado en la Teoría de la información. En esta formulación, se tomó $\Omega = \widetilde{X}_1 \cup \widetilde{X}_2$ de tal forma que el átomo $\widetilde{X}_1^c \cap \widetilde{X}_2^c$ represente el conjunto vacío. Luego $\mu^*(\widetilde{X}_1^c \cap \widetilde{X}_2^c) = 0$, pues μ^* es una medida. Así, μ^* está completamente determinada por todas las medidas de la información que involucran las variables aleatorias X_1 y X_2.

1.2.3. Construcción de la I-medida μ^*

Para demostrar el resultado central de esta sección es necesario demostrar la siguiente lema.

Lema 27. *Para una función aditiva μ,*

$$\begin{aligned}\mu\left(\bigcap_{k=1}^{n} A_k - B\right) = \sum_{1\le i\le n} \mu(A_i - B) - \sum_{1\le i<j\le n} \mu(A_i \cup A_j - B) \\ + \cdots + (-1)^{n+1}\mu(A_1 \cup \cdots \cup A_n - B).\end{aligned} \tag{1.50}$$

Demostración. (Por inducción sobre n). Para $n = 1$ se tiene trivialmente. Ahora, se supone que (1.50) es válida para n y se demostrará para n+1.

$$\begin{aligned}\mu\left(\bigcap_{k=1}^{n+1} A_k - B\right) &= \mu\left(\left(\bigcap_{k=1}^{n} A_k\right) \cap A_{n+1} - B\right) \\ &= \mu\left(\bigcap_{k=1}^{n} A_k - B\right) + \mu(A_{n+1} - B) - \mu\left(\left(\bigcap_{k=1}^{n} A_k\right) \cup A_{n+1} - B\right) \\ &= \Bigg(\sum_{1\le i\le n} \mu(A_i - B) - \sum_{1\le i<j\le n} \mu(A_i \cup A_j - B) + \cdots \\ &+ (-1)^{n+1}\mu(A_1 \cup \cdots \cup A_n - B)\Bigg) \\ &+ \mu(A_{n+1} - B) - \mu\left(\bigcap_{k=1}^{n} (A_k \cup A_{n+1}) - B\right) \\ &= \Bigg(\sum_{1\le i\le n} \mu(A_i - B) - \sum_{1\le i<j\le n} \mu(A_i \cup A_j - B) + \cdots \\ &+ (-1)^{n+1}\mu(A_1 \cup \cdots \cup A_n - B)\Bigg) + \mu(A_{n+1} - B) \\ &- \Bigg(\sum_{1\le i\le n} \mu(A_i \cup A_{n+1} - B) - \sum_{1\le i<j\le n} \mu(A_i \cup A_j \cup A_{n+1} - B) \\ &+ \cdots + (-1)^{n+1}\mu(A_1 \cup \cdots \cup A_n \cup A_{n+1} - B)\Bigg) \\ &= \sum_{1\le i\le n+1} \mu(A_i - B) - \sum_{1\le i<j\le n+1} \mu(A_i \cup A_j - B) + \cdots \\ &+ (-1)^{n+2}\mu(A_1 \cup \cdots \cup A_{n+1} - B).\end{aligned}$$

Con lo cual queda demostrado el Teorema. □

En la sección anterior se construyó la I-medida para dos variables aleatorias. Ahora se construirá la I-medida para cualquier número n, $n \geq 2$ de variables aleatorias.

Se considerarán n variables aleatorias $X_1, X_2, \ldots, X_n$. Sean $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$, subconjuntos de los números naturales, tales que los conjuntos de la forma $\bigcap_{i=1}^{n} Y_i$ donde Y_i es o $\widetilde{X}_i$ ó $\widetilde{X}_i^c$, son no vacíos. Sea $\mathcal{N}_n = \{1, 2, \ldots, n\}$, se define el conjunto universal Ω como

$$\Omega = \bigcup_{i \in \mathcal{N}_n} \widetilde{X}_i. \tag{1.51}$$

Se usa $\mathcal{F}_n$ para denotar el campo generado por $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$. El conjunto $A_0 = \bigcap_{i \in \mathcal{N}_n} \widetilde{X}_i^c$ se denomina el átomo vacío de $\mathcal{F}_n$, ya que

$$\bigcap_{i \in \mathcal{N}_n} \widetilde{X}_i^c = \left(\bigcup_{i \in \mathcal{N}_n} \widetilde{X}_i \right)^c = \Omega^c = \emptyset \tag{1.52}$$

Todos los átomos de $\mathcal{F}_n$ distintos de A_0 son no vacíos (por la forma como se escogieron los $\widetilde{X}_i$). Sea $\mathcal{A}$ el conjunto de todos los átomos no vacíos de $\mathcal{F}_n$, entonces $|\mathcal{A}| = 2^n - 1$.

Una medida con signo μ sobre $\mathcal{F}_n$, está completamente determinada por los valores de μ en los átomos no vacíos de $\mathcal{F}_n$.

Para simplificar la notación, X_α denotará $(X_i : i \in \alpha)$ y $\widetilde{X}_\alpha$ denotará $\bigcup_{i \in \alpha} \widetilde{X}_i$ para cualquier subconjunto no vacío α de $\mathcal{N}_n$.

Teorema 28. *Sea*

$$\mathcal{B} = \{\widetilde{X}_\alpha : \alpha \text{ es un conjunto no vacío de } \mathcal{N}_n\}. \tag{1.53}$$

Entonces, una medida con signo μ sobre $\mathcal{F}_n$, está completamente determinada por $\{\mu(B) : B \in \mathcal{B}\}$ el cual puede ser cualquier conjunto de números reales.

Demostración. El número de elementos en $\mathcal{B}$ es igual al número de subconjuntos no vacíos de $\mathcal{N}_n$, el cual es $2^n - 1$. Así $|\mathcal{B}| = |\mathcal{A}| = 2^n - 1$, sea $k = 2^n - 1$. Sean, $\mathbf{u}$ el k-vector columna cuyas componentes son los $\mu(A)$, $A \in \mathcal{A}$, y $\mathbf{h}$ el k-vector columna cuyas componentes son los $\mu(B)$, $B \in \mathcal{B}$.
Como todos los conjuntos en $\mathcal{B}$ pueden expresarse unívocamente como unión de algunos átomos no vacíos en $\mathcal{A}$, entonces por la aditividad de μ, para cada $B \in \mathcal{B}$, $\mu(B)$ puede expresarse unívocamente como la suma de ciertas componentes de $\mathbf{u}$. Así,

$$\mathbf{h} = \mathbf{C}_n \mathbf{u}, \tag{1.54}$$

donde $\mathbf{C}_n$ es una matriz $k \times k$ (única).

Por otro lado, un átomo no vacío de $\mathcal{F}_n$ tiene la forma $\bigcap_{i=1}^n Y_i$, donde Y_i es $\widetilde{X}_i$ ó $\widetilde{X}_i^c$ y existe al menos un i tal que $Y_i = \widetilde{X}_i$ para evitar el vacío. Luego,

$$\bigcap_{i=1}^n Y_i = \bigcap_{i:Y_i=\widetilde{X}_i} \widetilde{X}_i - \left(\bigcup_{j:Y_j=\widetilde{X}_j^c} \widetilde{X}_j^c \right) = \bigcap_{k=1}^m \widetilde{X}_k - \bigcup_{l=1}^p \widetilde{X}_l. \tag{1.55}$$

Ahora, se probará que para cada $A \in \mathcal{A},\ \mu(A)$ puede expresarse como combinación lineal de los $\mu(B),\ B \in \mathcal{B}$. Aplicando el Lema (27) y las siguientes dos identidades a (1.55).

$$\mu(A \cap B - C) = \mu(A - C) + \mu(B - C) - \mu(A \cup B - C) \tag{1.56}$$

$$\mu(A - B) = \mu(A \cup B) - \mu(B). \tag{1.57}$$

Se obtiene.

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

La cual es una combinación lineal de los $\mu(B),\ B \in \mathcal{B}$.
No obstante, la existencia de dicha expresión, no implica su unicidad. Sin embargo, se puede escribir

$$\mathbf{u} = \mathbf{D}_n \mathbf{h} \tag{1.58}$$

para alguna matriz $\mathbf{D}_n$, de tamaño $k \times k$. Sustituyendo (1.54) en (1.58) se obtiene

$$\mathbf{u} = (\mathbf{D}_n \mathbf{C}_n)\mathbf{u} \tag{1.59}$$

lo cual implica que $\mathbf{D}_n$ es la inversa de $\mathbf{C}_n$ y como $\mathbf{C}_n$ es única, entonces $\mathbf{D}_n$ es única. Así, $\mu(A),\ A \in \mathcal{A}$ está unívocamente determinado una vez $\mu(B),\ B \in \mathcal{B}$ se especifica. Por lo tanto una medida con signo μ en $\mathcal{F}_n$ está completamente determinada por $\{\mu(B) : B \in \mathcal{B}\}$, el cual puede ser cualquier conjunto de números reales. □

Ahora se probarán los siguientes dos resultados que están relacionados por la sustitución de símbolos en (1.49).

Lema 29.

$$\mu(A \cap B - C) = \mu(A \cup C) + \mu(B \cup C) - \mu(A \cup B \cup C) - \mu(C). \tag{1.60}$$

Demostración. De (1.56) y (1.57), se tiene que

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

□

Lema 30.

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z). \tag{1.61}$$

Demostración. Esta identidad corresponde a la identidad (1.35) que ya se probó. □

Ahora se construirá la I-medida μ^* sobre $\mathcal{F}_n$, usando el teorema 28. Se define.

$$\mu^*(\widetilde{X}_\alpha) := H(X_\alpha)$$

para todo subconjunto no vacío α de $\mathcal{N}_n$. Para que μ^* tenga sentido, tiene que ser consistente con todas las medidas de la información Shannon (vía sustitución de símbolos).

En este caso, lo siguiente debe tenerse para todos los subconjuntos (no necesariamente disjuntos) $\alpha, \alpha', \alpha''$ de $\mathcal{N}_n$, donde α y α' son no vacíos:

$$\mu^*(\widetilde{X}_\alpha \cap \widetilde{X}_{\alpha'} - \widetilde{X}_{\alpha''}) = I(X_\alpha; X_{\alpha'}|X_{\alpha''}) \tag{1.62}$$

Cuando $\alpha'' = \emptyset$, (1.62) se convierte en

$$\mu^*(\widetilde{X}_\alpha \cap \widetilde{X}_{\alpha'}) = I(X_\alpha; X_{\alpha'}). \tag{1.63}$$

Cuando $\alpha = \alpha'$ (1.62) se convierte en

$$\mu^*(\widetilde{X}_\alpha - \widetilde{X}_{\alpha''}) = H(X_\alpha|X_{\alpha''}). \tag{1.64}$$

Cuando $\alpha = \alpha'$ y $\alpha'' = \emptyset$, (1.62) se convierte en

$$\mu^*(\widetilde{X}_\alpha) = H(X_\alpha). \tag{1.65}$$

Así (1.62) cubre los cuatro casos de medidas de la información Shannon y ésta es una condición necesaria y suficiente para que μ^* sea consistente con todas las medidas de la información Shannon.

Teorema 31. *μ^* es una medida con signo sobre $\mathcal{F}_n$ que es consistente con todas las medidas de información Shannon.*

Demostración.

$$\begin{aligned}\mu^*(\widetilde{X}_\alpha \cap \widetilde{X}_{\alpha'} - \widetilde{X}_{\alpha''}) &= \mu^*(\widetilde{X}_{\alpha\cup\alpha''}) + \mu^*(\widetilde{X}_{\alpha'\cup\alpha''}) - \mu^*(\widetilde{X}_{\alpha\cup\alpha'\cup\alpha''}) - \mu^*(\widetilde{X}_{\alpha''}) \\ &= H(X_{\alpha\cup\alpha''}) + H(X_{\alpha'\cup\alpha''}) - H(X_{\alpha\cup\alpha'\cup\alpha''}) - H(X_{\alpha''}) \\ &= I(X_\alpha; X_{\alpha'}|X_{\alpha''}) \qquad \text{por lema 30.}\end{aligned}$$

□

Teorema 32. *Si no hay restricciones sobre $X_1, X_2, \ldots, X_n$, entonces μ^* puede tomar cualquier conjunto de valores no negativos en los átomos no vacíos de $\mathcal{F}_n$.*

Demostración. Se probará el teorema, construyendo una μ^* que pueda tomar cualquier conjunto de valores no negativos sobre los átomos no vacíos de $\mathcal{F}_n$. Recordemos que $\mathcal{A}$ es el conjunto de todos los átomos no vacíos de $\mathcal{F}_n$.
Se considera el conjunto de variables aleatorias mutuamente independientes $\{Y_A : A \in \mathcal{A}\}$. Ahora, sea $X_i,\ i = 1, \ldots, n$ la variable aleatoria definida por

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

Se define la I-medida μ^* para $X_1, X_2, \ldots, X_n$ como sigue. Como las Y_A son mutuamente independientes para todo subconjunto no vacío α de $\mathcal{N}_n$ se tiene que

$$H(X_\alpha) = \sum_{A\in\mathcal{A}:A\subseteq\widetilde{X}_\alpha} H(Y_A).$$

Por otro lado,

$$H(X_\alpha) = \mu^*(\widetilde{X}_\alpha) = \sum_{A\in\mathcal{A}:A\subseteq\widetilde{X}_\alpha} \mu^*(A)$$

$$\Rightarrow \sum_{A\in\mathcal{A}:A\subseteq\widetilde{X}_\alpha} H(Y_A) = \sum_{A\in\mathcal{A}:A\subseteq\widetilde{X}_\alpha} \mu^*(A)$$

Esta igualdad se cumple para todo subconjunto no vacío α de $\mathcal{N}_n$ tomando $\mu^*(A) = H(Y_A)$, para todo $A \in \mathcal{A}$. Por la unicidad de μ^*, ésta es también la única posibilidad para μ^*.

Como $H(Y_A)$ puede tomar cualquier valor no negativo, entonces μ^* puede tomar cualquier conjunto de valores no negativos en los átomos no vacíos de $\mathcal{F}_n$. □

1.3. Desigualdades de la Información

Una expresión de información f se refiere a una combinación lineal de medidas de información Shannon que involucren un número finito de variables aleatorias. Por ejemplo,

$$H(X,Y) + 2I(X;Z) - I(X;Y|Z) \tag{1.66}$$

es una expresión de información. Una desigualdad de información tiene la forma

$$f \geq c \tag{1.67}$$

donde c usualmente es igual a cero. Se consideran sólo las desigualdades no-estrictas por que éstas son usualmente la forma que toman las desigualdades en la Teoría de la Información. Igualmente, una identidad de información tiene la forma

$$f = c. \tag{1.68}$$

Definición 33. *Una desigualdad o identidad de información, se dice que siempre se cumple, si esta se cumple para cualquier distribución conjunta de las variables aleatorias involucradas.*

Por ejemplo, de dice que la desigualdad de información

$$I(X;Y) \geq 0, \tag{1.69}$$

siempre se cumple porque ésta se cumple para cualquier distribución conjunta $p(x,y)$. Por otro lado, se dice que una desigualdad de información no siempre se cumple, si existe una distribución de variables aleatorias para las cuales la desigualdad es falsa.
Por ejemplo la desigualdad $I(X;Y) \leq 0 \Leftrightarrow I(X;Y) = 0$ (pues $I(X;Y) \geq 0$ siempre se cumple.) Pero $I(X;Y) = 0$ se cumple, si y sólo si, X y Y son independientes, es decir, $I(X;Y) = 0$ no se cumple, si X y Y no son independientes. Entonces $I(X;Y) \leq 0$ no siempre se cumple.
Las desigualdades de la información gobiernan las imposibilidades en la Teoría de la información. Más precisamente, las desigualdades de la información implican que ciertas cosas no pueden suceder. Por eso son llamadas las leyes de la Teoría de la información.

Definición 34. *Sean X, Y y Z variables aleatorias. Las siguientes desigualdades se denominan desigualdades básicas,*

$$H(X) \geq 0, \tag{1.70}$$
$$H(X|Y) \geq 0, \tag{1.71}$$
$$I(X,Y) \geq 0, \tag{1.72}$$
$$I(X;Y|Z) \geq 0. \tag{1.73}$$

Las desigualdades básicas forman el conjunto más importante de desigualdades de la información.

Definición 35. *Una desigualdad de la información que siempre se cumple y que es implicada por las desigualdades básicas se denomina desigualdad tipo Shannon.*

Definición 36. *Una desigualdad de la información que siempre se cumple pero que no es implicada por las desigualdades básicas se llama desigualdad tipo no Shannon.*

Teorema 37. *Todas las medidas de la información Shannon son casos especiales de informaciones mutuas condicionales.*

Demostración. Sean X, Y y Z variable aleatorias, se verá que todas las medidas de información Shannon son casos especiales de $I(X;Y|Z)$. Sea Φ una variable aleatoria degenerada, es decir, Φ toma un valor constante con probabilidad 1, entonces

$$\begin{aligned}
&I(X;Y) = I(X;Y|Z), \ si\ Z = \Phi,\\
&H(X|Z) = I(X,Y|Z), \ si\ X = Y,\\
&H(X) = I(X,Y|Z), \ si\ X = Y\ y\ Z = \Phi.
\end{aligned}$$

□

Por lo tanto, las desigualdades tipo Shannon pueden definirse nuevamente como sigue.

Definición 38. *Una desigualdad de información tipo Shannon es una desigualdad de información que es (o que puede arreglarse para que sea) una suma finita de la forma*

$$\sum_i \alpha_i I(X_i; Y_i|Z_i) \geq 0, \tag{1.74}$$

donde cada α_i es un real no negativo.

1.3.1. La región Γ_n^*

Sea $\mathcal{N}_n := \{1, 2, \ldots, n\}$, donde $n \geq 2$, y sea $\Theta := \{X_i : i \in \mathcal{N}_n\}$ cualquier colección de n variables aleatorias. Asociadas con Θ, hay $k = 2^n - 1$ entropías mutuas. Por ejemplo, para $n = 3$, las 7 entropías mutuas asociadas con las variables aleatorias X_1, X_2 y X_3 son:

$$H(X_1),\ H(X_2), H(X_3),\ H(X_1, X_2),\ H(X_1, X_3),\ H(X_2, X_3),\ H(X_1, X_2, X_3). \tag{1.75}$$

Para cualquier α subconjunto no vacío de $\mathcal{N}_n$, se define.

$$X_\alpha := (X_i : i \in \alpha) \tag{1.76}$$

y

$$H_\Theta(\alpha) := H(X_\alpha). \tag{1.77}$$

Para un Θ fijo, H_Θ puede verse como una función de $2^{\mathcal{N}_n}$ en $\mathbb{R}$ con $H(\emptyset) = 0$. Por esta razón, H_Θ se denomina la función de entropía de Θ.
Sean, $\mathcal{H}_n$ el espacio euclidiano k-dimensional con coordenadas notadas por h_α, donde $\alpha \in 2^{\mathcal{N}_n} \setminus \{\emptyset\}$, y h_α corresponde al valor de $H_\Theta(\alpha)$ para cualquier colección Θ de n variables aleatorias. $\mathcal{H}_n$ se denomina el espacio de entropía para n variables aleatorias. Entonces, una función de entropía puede representarse como un vector columna en $\mathcal{H}_n$. Por otro lado, se tiene la siguiente definición:

Definición 39. *Un vector columna $\mathbf{h} \in \mathcal{H}_n$, se denomina entrópico, si $\mathbf{h}$ es igual a la función de entropía H_Θ de alguna colección Θ de n variables aleatorias.*

Estamos motivados a definir la siguiente región en $\mathcal{H}_n$:

Definición 40.

$$\Gamma_n^* := \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \textit{ es entrópico}\} \tag{1.78}$$

Por conveniencia, los vectores en Γ_n^* también serán referidos como funciones de entropía.

Ejemplo 41. *Si $n = 3$, las coordenadas de un vector en $\mathcal{H}_3$ son notadas por: h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123}. Γ_3^* es la región de $\mathcal{H}_3$ de todas las funciones de entropía para 3 variables aleatorias.*

Ejemplo 42. *Sean X_1 y X_2 variables aleatorias independientes tomando valores en $\{0,1\}$ con distribución de probabilidades: $Pr\{X_1 = 0\} = 2p$, $Pr\{X_1 = 1\} = 1 - 2p$, donde $0 \leq p \leq \frac{1}{2}$, $Pr\{X_2 = 0\} = Pr\{X_2 = 1\} = \frac{1}{2}$, y sean $X_3 = X_1 \cdot X_2$ y $X_4 = (1-X_1)(1-X_2)$. Así, si todas las entropías se toman en base e, se tiene que:*

$$\begin{aligned}
H(X_1) &= -2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_2) &= \ln 2,\\
H(X_3) &= \ln 2 - (\frac{1}{2}+p)\ln(1+2p) - (\frac{1}{2}-p)\ln(1-2p),\\
H(X_4) &= -(1-p)\ln(1-p) - p\ln(p),
\end{aligned}$$

$$\begin{aligned}
H(X_1,X_2) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_1,X_3) &= -2p\ln(2p) - (1-2p)\ln(1-2p) + (1-2p)\ln 2,\\
H(X_1,X_4) &= -2p\ln p - (1-2p)\ln(1-2p),\\
H(X_2,X_3) &= \ln 2 - p\ln p - p\ln 2 - (\frac{1}{2}-p)\ln(1-2p),\\
H(X_2,X_4) &= \ln 2 - p\ln p - p\ln 2 - (\frac{1}{2}-p)\ln(1-2p),\\
H(X_3,X_4) &= \ln 2 - p\ln p - p\ln 2 - (\frac{1}{2}-p)\ln(1-2p),\\
H(X_1,X_2,X_3) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_1,X_2,X_4) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_1,X_3,X_4) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_2,X_3,X_4) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p),\\
H(X_1,X_2,X_3,X_4) &= \ln 2 - 2p\ln(2p) - (1-2p)\ln(1-2p).
\end{aligned}$$

Si se denota el vector entrópico generado por las variables aleatorias X_1, X_2, X_3, X_4 por $h_{X,p}$, entonces $(h_{X,p})_\alpha = H(X_\alpha)$.

La región Γ_n^* satisface las siguientes propiedades:

1. Γ_n^* contiene el origen: El origen del espacio de entropía corresponde a la función de entropía de n variables aleatorias degeneradas que toman valores constantes.

2. Γ_n^* está en el ortante no negativo del espacio de entropía $\mathcal{H}_n$: las coordenadas de un vector entrópico en el espacio de entropía $\mathcal{H}_n$, corresponden a entropías mutuas que son no negativas.

3. $\overline{\Gamma_n^*}$, es convexo

Nota 43. *Las propiedades 1 y 3 implican que $\overline{\Gamma_n^*}$ es un cono convexo.*

De las propiedades anteriores sólo falta probar 3. Lo cual se hará a continuación.

Teorema 44. $\overline{\Gamma}_n^*$ *es un cono convexo.*

Demostración. Por la Nota anterior sólo falta mostrar que $\overline{\Gamma_n^*}$ es convexo. Sean $\mathbf{h}$ y $\mathbf{h}'$ en Γ_n^* funciones de entropía para los conjuntos arbitrarios de variables aleatorias $Y_1, Y_2, \ldots, Y_n$ y $Z_1, Z_2, \ldots, Z_n$ respectivamente. Para demostrar que $\overline{\Gamma}_n^*$, es convexo, es suficiente mostrar que, si $\mathbf{h}$ y $\mathbf{h}'$ están en Γ_n^* entonces $b\mathbf{h}+\overline{b}\mathbf{h}'$ está en $\overline{\Gamma}_n^*$ para todo $0 < b < 1$, donde $\overline{b} = 1-b$.

Sean $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)$ k copias independientes de $(Y_1, Y_2, \ldots, Y_n)$ y $(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n)$ k copias independientes de $(Z_1, Z_2, \ldots, Z_n)$. Sea U una variable aleatoria ternaria independiente de todas las demás variables aleatorias tal que

$$Pr(U = 0) = 1 - \delta - \mu,\ Pr(U = 1) = \delta \ y\ Pr(U = 2) = \mu.$$

Ahora, se construyen variables aleatorias $X_1, X_2, \ldots, X_n$ de la siguiente forma:

$$X_i = \begin{cases} 0, & si\ \ U = 0 \\ \mathbf{Y}_i, & si\ \ U = 1 \\ \mathbf{Z}_i, & si\ \ U = 2 \end{cases}$$

Note que $H(U) \to 0$ cuando $\delta, \mu \to 0$. Entonces, para todo α subconjunto no vacío de $\mathcal{N}_n$,

$$\begin{aligned} H(X_\alpha) &\leq H(X_\alpha, U) && \text{por (1.20)} \\ &= H(U) + H(X_\alpha|U) && \text{por (1.20)} \\ &= H(U) + (Pr\{U = 0\}H(X_\alpha|U = 0) \\ &+ Pr\{U = 1\}H(X_\alpha|U = 1) + Pr\{U = 2\}H(X_\alpha|U = 2)) \\ &= H(U) + \delta H(\mathbf{Y}_\alpha) + \mu H(\mathbf{Z}_\alpha) \\ &= H(U) + \delta k H(Y_\alpha) + \mu k H(Z_\alpha) \end{aligned}$$

Por otro lado,

$$H(X_\alpha) \geq H(X_\alpha|U) = \delta k H(Y_\alpha) + \mu k H(Z_\alpha)$$

pues $(0 \leq I(X, Y) = H(X) - H(X|Y) \Rightarrow H(X|Y) \leq H(X))$.
Combinando las anteriores desigualdades, se obtiene

$$0 \leq H(X_\alpha) - (\delta k H(Y_\alpha) + \mu k H(Z_\alpha)) \leq H(U).$$

Ahora tomando $\delta = \frac{b}{k}$ y $\mu = \frac{\overline{b}}{k}$ obtiene

$$0 \leq H(X_\alpha) - (bH(Y_\alpha) + \overline{b}H(Z_\alpha)) \leq H(U)$$

Si $k \to \infty$, $\delta, \mu \to 0$ entonces $H(U) \to 0$. Por lo tanto, $H(X_\alpha) \to bH(Y_\alpha) + \overline{b}H(Z_\alpha)$ cuando $k \to \infty$. Esto muestra que $b\mathbf{h} + \overline{b}\mathbf{h}' \in \overline{\Gamma}_n^*$. □

1.3.2. Expresiones de información en forma canónica

Cualquier medida de información de Shannon, puede expresarse como combinación lineal de entropías mutuas aplicando las identidades de la información (1.20), (1.31) y (1.35):

$$H(X|Y) = H(X, Y) - H(Y) \tag{1.79}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{1.80}$$

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z) \tag{1.81}$$

Así cualquier expresión de la información que involucre n variables aleatorias puede expresarse como combinación lineal de las k entropías mutuas asociadas. Esta es la forma canónica de una expresión de información. Cuando se escribe una expresión de información f como $f(\mathbf{h})$, se dice que f está en forma canónica. Como una expresión de información en forma canónica es una combinación de entropías mutuas, ésta tiene la forma

$$\mathbf{b}^T\mathbf{h}$$

donde $\mathbf{b}^T$ representa la transpuesta de un vector columna constante $\mathbf{b}$ en $\mathbb{R}^k$.
Las identidades de la (1.79) a (1.81) proveen una forma de expresar toda expresión de la información en una forma canónica. Sin embargo, no es claro si tal forma canónica es única, se resolverá esta duda a continuación.
En la prueba del Teorema (28), el vector $\mathbf{h}$ representaba los valores de la I-medida μ^* sobre las uniones en $\mathcal{F}_n$. Además, $\mathbf{h}$ está relacionado con los valores de μ^* sobre los átomos de $\mathcal{F}_n$, representados como $\mathbf{u}$ por

$$\mathbf{h} = \mathbf{C}_n\mathbf{u}$$

donde $\mathbf{C}_n$ es una matriz $k \times k$ única. El siguiente resultado es una consecuencia del Teorema (32).

Lema 45. *Sea*

$$\Psi_n^* := \{\mathbf{u} \in \mathbb{R}^k : \mathbf{C}_n\mathbf{u} \in \Gamma_n^*\}. \tag{1.82}$$

Entonces, el ortante no negativo de $\mathbb{R}^k$ es un subconjunto de Ψ_n^*.

Teorema 46. *Sea f una expresión de información. Entonces, la identidad de información sin restricciones $f = 0$ siempre se cumple, si y sólo si, f es la función cero.*

Demostración. Sin perdida de generalidad, se asume que f está en su forma canónica y sea $f(\mathbf{h}) = \mathbf{b}^T\mathbf{h}$.
($\Rightarrow$) Supóngase que f no es la función cero, es decir, $\mathbf{b} \neq 0$, entonces el conjunto $A = \{\mathbf{h} : \mathbf{b}^T\mathbf{h} = 0\}$ es un hiperplano en el espacio de entropías que tiene medida de Lebesgue cero, (ya que $dim(A) < dim(\mathcal{H}_n)$). Como la identidad de información sin restricciones $f = 0$ siempre se cumple, es decir, se cumple para todas las distribuciones conjuntas, entonces Γ_n^* debe estar contenido en el hiperplano A, de otra manera, existiría $\mathbf{h}_0 \in \Gamma_n^*$ tal que $f(\mathbf{h}_0) = \mathbf{b}^T\mathbf{h}_0 \neq 0$, y esto no puede suceder.
Veremos que Γ_n^* tiene medida de Lebesgue positiva, y en consecuencia no puede estar contenida en el hiperplano A que tiene medida de Lebesgue igual a cero, lo cual es una contradicción (porque ya se mostró que Γ_n^* debe estar contenido en el hiperplano A), y por lo tanto, f debe ser la función cero. Se vio en el Lema 45 que el ortante no negativo de $\mathcal{H}_n$, el cual tiene medida de Lebesgue positiva, es un subconjunto de Ψ_n^*. Así Ψ_n^* tiene medida de Lebesgue positiva. Como Γ_n^* es una transformación invertible de Ψ_n^* su medida de Lebesgue también es positiva.

($\Leftarrow$) Si f es la función cero, entonces es evidente que la identidad de información $f = 0$ siempre se cumple. $\square$

Corolario 47. *La forma canónica de una expresión de información es única.*

Demostración. Sean f_1 y f_2 formas canónicas de una expresión de información g. Como $g = f_1$ y $g = f_2$ son identidades de información que siempre se cumplen, entonces $f_1 - f_2 = 0$ es una identidad de información que siempre se cumple. Por lo tanto, $f_1 - f_2$ es la función cero, así $f_1 = f_2$. □

Debido a la unicidad de la forma canónica de una expresión de información, una forma fácil de verificar, si una identidad de información $f_1 = f_2$ siempre se cumple, es expresar $f_1 - f_2$ en forma canónica, si todos los coeficientes son nulos, entonces $f_1 = f_2$ siempre se cumple, en caso contrario, $f_1 = f_2$ no siempre se cumple.

1.3.3. Marco geométrico

En esta sección se explicará el significado geométrico de las desigualdades de información sin restricciones, las desigualdades de información restringidas y las identidades de información restringidas en términos de Γ_n^*. Sin perdida de generalidad, se asumirá que, todas las expresiones de información están en forma canónica.

Desigualdades sin restricciones

Sea $f \geq 0$, una desigualdad de información sin restricciones donde $f(\mathbf{h}) = \mathbf{b}^T\mathbf{h}$. Entonces, $f \geq 0$ corresponde al conjunto $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^T\mathbf{h} \geq 0\}$ el cual es un semi-espacio en el espacio de entropías $\mathcal{H}_n$ que contiene al origen. Específicamente, para cualquier $\mathbf{h} \in \mathcal{H}_n$, $f(\mathbf{h}) \geq 0$, si y sólo si, $\mathbf{h}$ pertenece a ese conjunto. Para simplificar, ese conjunto se referirá al semi espacio $f \geq 0$.

Ejemplo 48. *Para $n = 2$ la desigualdad de la información*

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \geq 0, \tag{1.83}$$

escrita como $h_1 + h_2 - h_{12} \geq 0$ corresponde al semi-espacio $\{h \in \mathcal{H}_2 : h_1 + h_2 - h_{12} \geq 0\}$ en el espacio de entropía $\mathcal{H}_2$.

Como una desigualdad de la información siempre se cumple, si y sólo si, se cumple para la función de entropía de cualquier distribución conjunta de las variables aleatorias involucradas, entonces se tiene la siguiente interpretación geométrica de una desigualdad de información:

$$f \geq 0 \text{ siempre se cumple, si y sólo si, } \Gamma_n^* \subseteq \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}. \tag{1.84}$$

Esto caracteriza todas las desigualdades sin restricciones en términos de Γ_n^*.

Desigualdades restringidas

En Teoría de la información, a menudo se trabaja con desigualdades (identidades) con ciertas restricciones sobre la distribución conjunta de las variables aleatorias involucradas. Estas son llamadas desigualdades (identidades)de información restringidas y las restricciones sobre la distribución usualmente puede expresarse como restricciones lineales sobre las entropías.

Ejemplo 49. *Los siguientes son ejemplos de restricciones sobre la distribución conjunta.*

1. X_1, X_2, *y* X_3 *son mutuamente independientes, si y sólo si,* $H(X_1, X_2, X_3) = H(X_1) + H(X_2) + H(X_3)$.
2. X_1, X_2, *y* X_3 *son dos a dos independientes, si y sólo si,* $I(X_1; X_2) = I(X_2; X_3) = I(X_1; X_3) = 0$.
3. X_1 *es una función de* X_2, *si y sólo si,* $H(X_1|X_2) = 0$.

Sean q restricciones lineales sobre las entropías, dadas por $\mathbf{Qh} = 0$, donde $\mathbf{Q}$ es una matriz $q \times k$. No se asume que las restricciones sean linealmente independientes, así que $\mathbf{Q}$ no necesariamente tiene rango k. Sea

$$\Phi = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{Qh} = 0\}.$$

En otras palabras, las q restricciones confinan $\mathbf{h}$ a un subespacio lineal Φ en el espacio de entropía. Paralelamente, se hace la siguiente interpretación geométrica de una desigualdad restringida:

Bajo la restricción Φ, $f \geq 0$ siempre se cumple, si y sólo si, $(\Gamma_n^* \cap \Phi) \subset \{\mathbf{h} : f(\mathbf{h}) \geq 0\}$.

$\Phi = \mathcal{H}_n$ si no hay ninguna restricción sobre las entropías. (En este sentido, una desigualdad sin restricciones es un caso particular de una desigualdad restringida).

Identidades restringidas

$f = 0$ siempre se cumple, si y sólo si, $f \geq 0$ y $f \leq 0$ siempre se cumplen. Entonces, se tiene que, bajo la restricción Φ, $f = 0$ siempre se cumple, si y sólo si,

$$\Gamma_n^* \cap \Phi \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\} \cap \{\mathbf{h} : f(\mathbf{h}) \leq 0\}$$

o bajo la restricción Φ, siempre se cumple, si y sólo si,

$$\Gamma_n^* \cap \Phi \subseteq \{\mathbf{h} : f(\mathbf{h}) = 0\}$$

es decir, $\Gamma_n^* \cap \Phi$ está contenido en el hiperplano $f = 0$.

1.4. Desigualdades tipo Shannon

Las desigualdades básicas forman uno de los conjuntos más importante de desigualdades de la información. Estas son llamadas desigualdades tipo Shannon. En esta sección, se mostrará que la verificación de una desigualdad tipo Shannon puede formularse como un problema de programación lineal.

1.4.1. Las desigualdades elementales

Definición 50. *Una medida de información Shannon, se dice que es reducible, si existe una variable aleatoria que aparece más de una vez en la medida de información. En caso contrario la medida de información se dice irreducible.*

Sin perdida de generalidad, se consideran sólo medidas de la información Shannon irreducibles, porque una medida de información Shannon reducible puede ser siempre escrita como la suma de medidas de información Shannon irreducibles.

La no negatividad de todas las medidas de información Shannon forman el conjunto de las desigualdades básicas. El conjunto de desigualdades básicas no es minimal en el sentido que algunas desigualdades básicas son implicadas por otras.

Por ejemplo, $H(X|Y) \geq 0$ y $I(X;Y) \geq 0$ son desigualdades básicas que implican que $H(X) = H(X|Y) + I(X;Y) \geq 0$, la cual es una desigualdad básica.

Considérense las siguientes identidades:

$$H(X) = H(X|Y) + I(X;Y) \tag{1.85}$$
$$H(X,Y) = H(X) + H(Y|X) \tag{1.86}$$
$$I(X;Y,Z) = I(X;Y) + I(X;Z|Y) \tag{1.87}$$
$$H(X|Z) = H(X|Y,Z) + I(X;Y|Z) \tag{1.88}$$
$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z) \tag{1.89}$$
$$I(X;Y,Z|T) = I(X;Y|T) + I(X;Z|Y,T) \tag{1.90}$$

Sea $\mathcal{N}_n = \{1, 2, \ldots, n\}$, donde $n \geq 2$, y sean $X_1, X_2, \ldots, X_n$ variables aleatorias. Aplicando las identidades anteriores, cualquier medida de información puede expresarse como la suma de medidas de información Shannon de las siguientes dos formas elementales:

1. $H(X_i|X_{\mathcal{N}_n - \{i\}}),\ i \in \mathcal{N}_n$.
2. $I(X_i; X_j|X_k),\ i \neq j,\ k \subseteq \mathcal{N}_n - \{i,j\}$.

El número de formas elementales de medidas de información Shannon para n variables aleatorias es:

$$m = n + \binom{n}{2} 2^{n-2}. \tag{1.91}$$

Ejemplo 51. *Se puede expandir $H(X_1, X_2)$ en una suma de formas elementales de medidas de la información Shannon para $n = 3$ mediante la aplicación de las identidades de (1.85) a (1.90) de la siguiente manera*

$$\begin{aligned}
H(X_1, X_2) &= H(X_1) + H(X_2|X_1) && \textit{por(1.86)} \\
&= H(X_1|X_2, X_3) + I(X_1; X_2, X_3) \\
&\quad + H(X_2|X_1, X_3) + I(X_2; X_3|X_1) && \textit{por (1.85) y (1.88)} \\
&= H(X_1|X_2, X_3) + I(X_1; X_2) + I(X_1; X_3|X_2) \\
&\quad + H(X_2|X_1, X_3) + I(X_2; X_3|X_1) && \textit{por(1.87).}
\end{aligned}$$

La no negatividad de las dos formas elementales de medidas de la información Shannon forman un subconjunto propio del conjunto de las desigualdades básicas. En este conjunto más pequeño, las m desigualdades se llamarán desigualdades elementales. Estas son equivalentes a las desigualdades básicas porque cada desigualdad básica se obtiene como la suma de un conjunto de desigualdades elementales utilizando (1.85) a (1.90).

Ejemplo 52. *En el último ejemplo, se expreso $H(X_1, X_2)$ como*

$$H(X_1|X_2, X_3) + I(X_1; X_2) + I(X_1; X_3|X_2) + H(X_2|X_1, X_3) + I(X_2; X_3|X_1).$$

Las cinco medidas de información Shannon en la expresión anterior están en forma elemental para $n = 3$. Entonces, la desigualdad básica $H(X_1, X_2) \geq 0$ se obtiene como la suma de las siguientes desigualdades elementales:

$$\begin{aligned} H(X_1|X_2, X_3) &\geq 0 \\ I(X_1; X_2) &\geq 0 \\ I(X_1; X_3|X_2) &\geq 0 \\ H(X_2|X_1, X_3) &\geq 0 \\ I(X_2; X_3|X_1) &\geq 0. \end{aligned}$$

1.4.2. Enfoque de la Programación Lineal

Se vio que, cualquier expresión de información puede expresarse en forma canónica (esta expresión es única), es decir, se expresa como combinación lineal de las $k = 2^n - 1$ entropías mutuas que involucran algunas o todas las variables aleatorias $X_1, X_2, \ldots, X_n$.

Si las desigualdades elementales se expresan en forma canónica, entonces se convierten en desigualdades en el espacio de entropías $\mathcal{H}_n$. Se denota este conjunto de desigualdades por $\mathbf{Gh} \geq 0$, donde $\mathbf{G}$ es una matriz $m \times k$, y definimos

$$\Gamma_n := \{\mathbf{h} \in \mathcal{H}_n : \mathbf{Gh} \geq 0\} \tag{1.92}$$

Se tiene la siguiente propiedad para Γ_n

Lema 53. *Γ_n es un cono poliedral en el ortante no negativo de $\mathcal{H}_n$.*

Demostración. Nótese que Γ_n contiene el origen. Considérese la base canónica $\{\mathbf{e}_j\}_{1 \leq j \leq k}$ de $\mathbb{R}^k$, entonces la desigualdad $\mathbf{e}_j^T \mathbf{h} \geq 0$ corresponde a la no negatividad de una entropía mutua, la cual es una desigualdad básica. Como el conjunto de las desigualdades elementales es equivalente al conjunto de las desigualdades básicas, si $\mathbf{h} \in \Gamma_n$, es decir, $\mathbf{h}$ satisface todas las desigualdades elementales, entonces $\mathbf{h}$ satisface las desigualdades básicas dadas en $\mathbf{e}_j^T \mathbf{h} \geq 0$. En otras palabras,

$$\Gamma_n \subseteq \{\mathbf{h} : \mathbf{e}_j^T \mathbf{h} \geq 0\} \quad \text{para todo } 1 \leq j \leq k. \tag{1.93}$$

Esto implica que Γ_n está en el ortante no negativo de $\mathcal{H}_n$. Como Γ_n contiene el origen y las restricciones $\mathbf{Gh} \geq 0$ son lineales, entonces Γ_n es una pirámide en el ortante no negativo de $\mathcal{H}_n$. □

Como las desigualdades elementales son satisfechas por la función de entropía de cualquier colección de n variables aleatorias $X_1, X_2, \ldots, X_n$, para cualquier $\mathbf{h}$ en Γ_n^*; $\mathbf{h}$ también está en Γ_n. Es decir,

$$\Gamma_n^* \subseteq \Gamma_n. \tag{1.94}$$

Por lo tanto, para cualquier desigualdad sin restricciones $f \geq 0$, si

$$\Gamma_n \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\}, \tag{1.95}$$

entonces

$$\Gamma_n^* \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\}. \tag{1.96}$$

Es decir, $f \geq 0$ siempre se cumple. En otras palabras, (1.95) es condición suficiente para que $f \geq 0$ siempre se cumpla. Además, una desigualdad $f \geq 0$ tal que (1.95) se cumple, es implicada por las desigualdades básicas, porque si $\mathbf{h}$ satisface las desigualdades básicas ($\mathbf{h} \in \Gamma_n$), entonces $\mathbf{h}$ satisface $f(\mathbf{h}) \geq 0$.

Para desigualdades restringidas, imponiendo la restricción $\mathbf{Qh} = 0$ y sea $\Phi := \{\mathbf{h} : \mathbf{Qh} = 0\}$. Para una desigualdad $f \geq 0$, si

$$(\Gamma_n \cap \Phi) \subseteq \{h : f(h) \geq 0\}, \tag{1.97}$$

entonces por (1.94)

$$(\Gamma_n^* \cap \Phi) \subseteq \{h : f(h) \geq 0\}, \tag{1.98}$$

es decir, $f \geq 0$ siempre se cumple bajo la restricción Φ. En otras palabras, (1.97) es condición suficiente para que $f \geq 0$ siempre se cumpla bajo la restricción Φ.

Además, una desigualdad $f \geq 0$ bajo la restricción Φ tal que (1.97) se satisface, es implicada por las desigualdades básicas y la restricción Φ pues, si $\mathbf{h} \in \Phi$ y $\mathbf{h}$ satisface las desigualdades básicas ($\mathbf{h} \in \Gamma_n \cap \Phi$), entonces $\mathbf{h}$ satisface $f(\mathbf{h}) \geq 0$.

Desigualdades sin restricciones

Para verificar si una desigualdad sin restricciones $\mathbf{b}^T\mathbf{h} \geq 0$ es una desigualdad tipo Shannon, se debe verificar si Γ_n es un subconjunto de $\{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$. El siguiente teorema induce un proceso computacional para este propósito.

Teorema 54. $\mathbf{b}^T\mathbf{h} \geq 0$ *es una desigualdad tipo Shannon, si y sólo si, la solución del problema*

$$\textit{Minimizar } \mathbf{b}^T\mathbf{h}, \textit{sujeto a } \mathbf{Gh} \geq 0 \tag{1.99}$$

es cero. En este caso, el mínimo ocurre en el origen.

Demostración. Como $0 \in \Gamma_n$ y $\mathbf{b}^T 0 = 0$ para todo $\mathbf{b}$, el mínimo del problema (1.99) es a lo sumo cero.

$\Rightarrow$) Si se supone que $\Gamma_n \subseteq \{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$ y que el mínimo del problema (1.99) es negativo. Entonces, existe $\mathbf{h} \in \Gamma_n$ tal que $\mathbf{b}^T\mathbf{h} < 0$, por lo tanto, $\Gamma_n \nsubseteq \{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$, que es una contradicción. Luego, si $\Gamma_n \subseteq \{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$, entonces el mínimo del problema (1.99) es cero.

$\Leftarrow$) Si se supone que el mínimo del problema (1.99) es 0, pero $\Gamma_n \nsubseteq \{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$. Entonces, existe $\mathbf{h} \in \Gamma_n$ tal que $\mathbf{b}^T\mathbf{h} < 0$, esto implica que el mínimo del problema (1.99) es negativo, es decir, no es cero, lo cual contradice la hipótesis.

Finalmente, si el mínimo del problema (1.99) es cero, como Γ_n contiene el origen y $\mathbf{b}^T 0 = 0$, el mínimo ocurre en el origen. □

En virtud de este teorema, para verificar si $\mathbf{b}^T\mathbf{h} \geq 0$ es una desigualdad sin restricciones tipo Shannon, lo único que hay que hacer es aplicar el test de optimalidad del método simplex para verificar si el punto $\mathbf{h} = 0$ es óptimo para el problema de minimización (1.99). Si $\mathbf{h} = 0$ es óptimo, entonces $\mathbf{b}^T\mathbf{h} \geq 0$ es una desigualdad sin restricciones tipo Shannon, en caso contrario no lo es.

Desigualdades e identidades restringidas

Para verificar si una desigualdad $\mathbf{b}^T\mathbf{h} \geq 0$ bajo la restricción Φ es una desigualdad tipo Shannon, se debe verificar si $\Gamma_n \cap \Phi \subseteq \{\mathbf{h} : \mathbf{b}^T\mathbf{h} \geq 0\}$.

Teorema 55. $\mathbf{b}^T\mathbf{h} \geq 0$ *es una desigualdad tipo Shannon bajo la restricción* Φ, *si y sólo si, la solución del problema*

$$\textit{Minimizar } \mathbf{b}^T\mathbf{h}, \textit{sujeto a } \mathbf{G}\mathbf{h} \geq 0 \textit{ y } \mathbf{Q}\mathbf{h} = 0 \tag{1.100}$$

es cero. En este caso, el mínimo ocurre en el origen.

La prueba de este teorema es similar a la del anterior teorema, por lo tanto se omite. Tomando ventaja de la estructura lineal de la restricción Φ, se puede reformular el problema de minimización en (1.100) de la siguiente forma. Sea r el rango de $\mathbf{Q}$. Como $\mathbf{h}$ está en el espacio nulo de $\mathbf{Q}$, se puede escribir

$$\mathbf{h} = \widetilde{\mathbf{Q}}\mathbf{h}',$$

donde $\widetilde{\mathbf{Q}}$ es una matriz $k \times (k-r)$ tal que las columnas de $\widetilde{\mathbf{Q}}^T$ forman una base del complemento ortogonal del espacio fila de $\mathbf{Q}$, y $\mathbf{h}'$ es un $(k-r)$-vector columna. Entonces, las desigualdades elementales pueden expresarse como

$$\mathbf{G}\widetilde{\mathbf{Q}}\mathbf{h}' \geq 0.$$

En términos de $\mathbf{h}'$, Γ_n se convierte en

$$\Gamma'_n = \{\mathbf{h}' : \mathbf{G}\widetilde{\mathbf{Q}}\mathbf{h}' \geq 0\}$$

que es un cono poliedral en $\mathbb{R}^{k-r}$ (pero no necesariamente en el ortante no negativo).

De igual manera, $\mathbf{b}^T\mathbf{h}$ puede escribirse como $\mathbf{b}^T\widetilde{\mathbf{Q}}\mathbf{h}'$.

Con todas las expresiones de información en términos de $\mathbf{h}'$, el problema (1.100) se convierte en

$$\text{Minimizar } \mathbf{b}^T\widetilde{\mathbf{Q}}\mathbf{h}', \text{sujeto a } \mathbf{G}\widetilde{\mathbf{Q}}\mathbf{h}' \geq 0 \tag{1.101}$$

Luego para verificar si $\mathbf{b}^T\mathbf{h} \geq 0$ es una desigualdad tipo Shannon bajo la restricción Φ, sólo se necesita aplicar el test de optimalidad del método simplex, para verificar cuando el punto $\mathbf{h}' = 0$ es óptimo para el problema (1.101). Si $\mathbf{h}' = 0$ es óptimo, entonces $\mathbf{b}^T\mathbf{h} \geq 0$ es una desigualdad tipo Shannon bajo la restricción Φ, en otro caso, no lo es.

Al imponer la restricción Φ, el número de desigualdades elementales se mantiene, mientras que la dimensión del problema se reduce de k a $k - r$.

Finalmente, para verificar que $\mathbf{b}^T\mathbf{h} = 0$ es una identidad tipo Shannon bajo la restricción Φ, es decir, $\mathbf{b}^T\mathbf{h} = 0$ es implicado por las desigualdades básicas, se debe verificar que $\mathbf{b}^T\mathbf{h} \geq 0$ y $\mathbf{b}^T\mathbf{h} \leq 0$ son desigualdades tipo Shannon bajo la restricción Φ.

En la página `http://xitip.epfl.ch/` se encuentra disponible el software $XITIP$, que permite verificar si una desigualdad de información es tipo Shannon, el algoritmo que se usa en este sofware fue propuesto por Raymond Yeung y Ying-On Yan en su sofware ITIP, el cual utiliza el solucionador de programación lineal, una herramienta de optimización de MATLAB. XITIP es una adaptación de este sofware, que utiliza un solucionador de programación lineal basado en C.

CAPÍTULO 2

Más allá de las desigualdades tipo Shannon

En el Capítulo anterior se introdujo las regiones Γ_n^* y Γ_n en el espacio de entropías $\mathcal{H}_n$ para n variables aleatorias. De Γ_n^* se puede determinar si cualquier desigualdad de la información siempre se cumple. La región Γ_n, definida por el conjunto de todas las desigualdades básicas (equivalentemente todas las desigualdades elementales) involucrando n variables aleatorias, es una cota superior de Γ_n^*. De Γ_n, se puede determinar si dada una desigualdad de la información ésta es implicada por las desigualdades básicas. De ser así, está es llamada una desigualdad tipo Shannon.

Si las dos regiones Γ_n^* y Γ_n son iguales, entonces todas las desigualdades de la información que siempre se cumple son desigualdades tipo Shannon, y por lo tanto todas las desigualdades de información están completamente caracterizadas.
Sin embargo, si Γ_n^* es un subconjunto propio de Γ_n, entonces existen restricciones sobre una función de entropía que no son implicadas por las desigualdades básicas. Tal restricción, si estuviera en la forma de una desigualdad, se refiere a una desigualdad tipo no-Shannon.

El hecho de que $\Gamma_n^* \neq \Gamma_n$ no necesariamente implica la existencia de una desigualdad tipo no Shannon, como se verá para el caso $n = 3$.
En esta Capítulo se presentarán caracterizaciones de Γ_n^* las cuales son más refinadas que Γ_n. Estas caracterizaciones producen la existencia de desigualdades tipo no-Shannon para $n \geq 4$

2.1. Caracterizaciones de Γ_2^*, Γ_3^* y Γ_n^*

En la prueba del Teorema 28, el vector $\mathbf{h}$ representa los valores de la I-medida μ^* en las uniones en $\mathcal{F}_n$. Además, $\mathbf{h}$ está relacionado con los valores de μ^* sobre los átomos de $\mathcal{F}_n$, representados como $\mathbf{u}$, por

$$\mathbf{h} = \mathbf{C}_n \mathbf{u},$$

donde $\mathbf{C}_n$ es una matriz $k \times k$ única, con $k = 2^n - 1$. Sea $\mathcal{I}_n$ el espacio euclidiano k dimensional con las coordenadas notadas por las componentes de $\mathbf{u}$. Nótese que cada

coordenada en $\mathcal{I}_n$ corresponde al valor de μ^* en un átomo no vacío de $\mathcal{F}_n$. En el Lema 45 se dio la definición de la región

$$\Psi_n^* = \{\mathbf{u} \in \mathcal{I}_n : \mathbf{C}_n\mathbf{u} \in \Gamma_n^*\}, \tag{2.1}$$

que se obtiene de la región Γ_n^* vía la transformación lineal $\mathbf{C}_n^{-1}$. Análogamente, se definió la región

$$\Psi_n = \{\mathbf{u} \in \mathcal{I}_n : \mathbf{C}_n\mathbf{u} \in \Gamma_n\}. \tag{2.2}$$

La región Γ_n^*, como se verá, es difícil de caracterizar para un n general. Por lo tanto, se empezará la discusión con el caso simple, de $n = 2$.

Teorema 56. $\Gamma_2^* = \Gamma_2$.

Demostración. Para $n = 2$ las desigualdades elementales son:

$$H(X_1|X_2) = \mu^*(\widetilde{X}_1 - \widetilde{X}_2) \geq 0. \tag{2.3}$$

$$H(X_2|X_1) = \mu^*(\widetilde{X}_2 - \widetilde{X}_1) \geq 0. \tag{2.4}$$

$$I(X_1; X_2) = \mu^*(\widetilde{X}_1 \cap \widetilde{X}_1) \geq 0. \tag{2.5}$$

Nótese que las cantidades en el lado derecho son precisamente los valores de μ^* sobre los átomos de $\mathcal{F}_2$. Entonces,

$$\Psi_2 = \{\mathbf{u} \in \mathcal{I}_2 : \mathbf{u} \geq 0\}, \tag{2.6}$$

por lo tanto, Ψ_2 es el ortante no negativo de $\mathcal{I}_2$. Como $\Gamma_2^* \subseteq \Gamma_2$, entonces $\Psi_2^* \subseteq \Psi_2$. Por otro lado, por el Lema 45, el ortante no negativo de $\mathbb{R}^k$ es un subconjunto de Ψ_n^*, luego $\Psi_2 \subseteq \Psi_2^*$. Así, $\Psi_2^* = \Psi_2$, que implica que $\Gamma_2^* = \Gamma_2$. □

Ahora, se verá que el teorema anterior no puede generalizarse.

Teorema 57. $\Gamma_3^* \neq \Gamma_3$.

Demostración. Para $n = 3$ las desigualdades elementales son

$$H(X_i|X_j, X_k) = \mu^*(\widetilde{X}_i - (\widetilde{X}_j \cup \widetilde{X}_k)) \geq 0. \tag{2.7}$$

$$I(X_i; X_j|X_k) = \mu^*(\widetilde{X}_i \cap \widetilde{X}_j - \widetilde{X}_k) \geq 0. \tag{2.8}$$

$$I(X_i; X_j) = \mu^*(\widetilde{X}_i \cap \widetilde{X}_j) = \mu^*(\widetilde{X}_i \cap \widetilde{X}_j \cap \widetilde{X}_k) + \mu^*(\widetilde{X}_i \cap \widetilde{X}_j - \widetilde{X}_k) \geq 0. \tag{2.9}$$

Para $i, j, k \in \{1, 2, 3\}$ diferentes. Para $\mathbf{u} \in \mathcal{I}_3$, sea $\mathbf{u} = (u_1, u_2, u_3, u_4, u_5, u_6, u_7)$ donde $u_i,\ 1 \leq i \leq 7$, corresponde a los valores

$$\begin{aligned} &\mu^*(\widetilde{X}_1 - \widetilde{X}_2 - \widetilde{X}_3),\ \mu^*(\widetilde{X}_2 - \widetilde{X}_1 - \widetilde{X}_3),\ \mu^*(\widetilde{X}_3 - \widetilde{X}_1 - \widetilde{X}_2), \\ &\mu^*(\widetilde{X}_1 \cap \widetilde{X}_2 - \widetilde{X}_3),\ \mu^*(\widetilde{X}_1 \cap \widetilde{X}_3 - \widetilde{X}_2),\ \mu^*(\widetilde{X}_2 \cap \widetilde{X}_3 - \widetilde{X}_1), \\ &\mu^*(\widetilde{X}_1 \cap \widetilde{X}_2 \cap \widetilde{X}_3), \end{aligned} \tag{2.10}$$

respectivamente. Estos son los valores de μ^* sobre los átomos no vacíos de $\mathcal{F}_3$. Entonces, de (2.7), (2.8), (2.9) se tiene que

$$\Psi_3 = \{\mathbf{u} \in \mathcal{I}_3 : u_i \geq 0,\quad 1 \leq i \leq 6;\quad u_j + u_7 \geq 0,\quad 4 \leq j \leq 6\}. \tag{2.11}$$

Es fácil verificar que el punto $(0,0,0,a,a,a,-a)$ está en Ψ_3 para cualquier $a \geq 0$, y se puede ver que para este punto las relaciones

$$H(X_i|X_j,X_k) = 0 \tag{2.12}$$

$$I(X_i;X_j) = 0 \tag{2.13}$$

para $i,j,k \in \{1,2,3\}$ diferentes, se satisfacen, es decir, cada variable aleatoria es función de las otras dos, y las tres variables aleatorias son independientes dos a dos.

Sea $\mathcal{S}_{X_i}$ el soporte de $X_i,\ i=1,2,3$. Para todo $x_1 \in \mathcal{S}_{X_1}$ y $x_2 \in \mathcal{S}_{X_2}$, como X_1 y X_2 son independientes, se tiene que

$$p(x_1,x_2) = p(x_1)p(x_2) > 0.$$

Como X_3 es función de X_1 y X_2, existe un único $x_3 \in \mathcal{S}_{X_3}$ tal que

$$p(x_1,x_2,x_3) = p(x_1,x_2) = p(x_1)p(x_2) > 0. \tag{2.14}$$

Ahora, como X_2 es función de X_1 y X_3, y X_1 y X_3 son independientes, podemos escribir

$$p(x_1,x_2,x_3) = p(x_1,x_3) = p(x_1)p(x_3) > 0. \tag{2.15}$$

igualando (2.14) y (2.15), se obtiene que

$$p(x_2) = p(x_3). \tag{2.16}$$

Ahora, considérese cualquier $x_2' \in \mathcal{S}_{X_2}$ tal que $x_2' \neq x_2$. Como X_2 y X_3 son independientes, tenemos

$$p(x_2,x_3) = p(x_2)p(x_3) > 0.$$

Como X_1 es función de X_2 y X_3, existe un único $x_1' \in \mathcal{S}_{X_1}$ tal que

$$p(x_1',x_2',x_3) = p(x_2',x_3) = p(x_2')p(x_3) > 0.$$

Ahora, como X_2 es función de X_1 y X_3, y X_1 y X_3 son independientes, entonces

$$p(x_1',x_2',x_3) = p(x_1',x_3) = p(x_1')p(x_3) > 0. \tag{2.17}$$

Similarmente, como X_3 es función de X_1 y X_2, y X_1 y X_2 son independientes, se puede escribir

$$p(x_1',x_2',x_3) = p(x_1',x_2') = p(x_1')p(x_2') > 0. \tag{2.18}$$

Igualando (2.17) y (2.18), se tiene

$$p(x_2') = p(x_3). \tag{2.19}$$

Y de (2.16), se tiene que, $p(x_2') = p(x_2)$. Por lo tanto X_2 debe tener distribución uniforme sobre su soporte. Lo mismo puede probarse para X_1 y X_3.

Ahora,

$$\begin{aligned} H(X_1) &= H(X_1|X_2,X_3) + I(X_1;X_2|X_3) + I(X_1;X_3|X_2) + I(X_1;X_2;X_3) \\ &= 0 + a + a + (-a) = a. \end{aligned}$$

De manera similar se obtiene que, $H(X_2) = H(X_3) = a$.

Entonces, los únicos valores que a puede tomar son $\log M$ donde M es la suma de los cardinales de los soportes de X_1, X_2 y X_3. En otras palabras, si a no es igual a $\log M$ para algún entero positivo M, entonces el punto $(0,0,0,a,a,a,-a)$ no está en Ψ_3^*. Esto prueba que $\Psi_3^* \neq \Psi_3$, que implica $\Gamma_3^* \neq \Gamma_3$. □

Aunque la región $\overline{\Gamma}_n^*$ no es suficiente para caracterizar todas las desigualdades de la información, si es suficiente para caracterizar todas las desigualdades de la información sin restricciones. Esto puede verse de lo siguiente. De la sección 1.3.3, una desigualdad sin restricciones $f \geq 0$ involucrando n variables aleatorias siempre se cumple, si y sólo si,

$$\Gamma_n^* \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\}. \tag{2.20}$$

Como $\{\mathbf{h} : f(\mathbf{h}) \geq 0\}$ es cerrado, tomando clausura a ambos lados, se obtiene

$$\overline{\Gamma}_n^* \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\}. \tag{2.21}$$

Por otro lado, si $f \geq 0$ satisface (2.21) entonces

$$\Gamma_n^* \subseteq \overline{\Gamma}_n^* \subseteq \{\mathbf{h} : f(\mathbf{h}) \geq 0\}. \tag{2.22}$$

Así, (2.20) y (2.21) son equivalentes, y por lo tanto $\overline{\Gamma}_n^*$ es suficiente para caracterizar todas las desigualdades sin restricciones.

Teniendo que $\Gamma_3^* \neq \Gamma_3$, es natural preguntarse si el espacio entre Γ_3^* y Γ_3 tiene medida de Lebesgue 0. En otras palabras, $\overline{\Gamma}_3^* = \Gamma_3$ lo cual se probará en el siguiente teorema. Para lo cual, se demostrará primero el siguiente resultado.

Lema 58. *Si* $\mathbf{h}$ *y* $\mathbf{h}'$ *están en* Γ_n^*, *entonces* $\mathbf{h} + \mathbf{h}'$ *esta en* Γ_n^*.

Demostración. Sean $\mathbf{h}$ y $\mathbf{h}'$ en Γ_n^*. y se supone que $\mathbf{h}$ representa la función de entropía para las variables aleatorias $X_1, X_2, \ldots, X_n$ y $\mathbf{h}'$ representa la función de entropía para las variables aleatorias $X_1', X_2', \ldots, X_n'$. Sean $(X_1, \ldots, X_n)$ y $(X_1', \ldots, X_n')$ independientes; se definen las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ de la siguiente manera: $Y_i := (X_i, X_i')$ para todo $i \in \mathcal{N}_n$. Entonces, para todo subconjunto α de $\mathcal{N}_n$:

$$H(Y_\alpha) = H(X_\alpha) + H(X_\alpha') = h_\alpha + h_\alpha'$$

Por lo tanto, $\mathbf{h} + \mathbf{h}'$ que representa la función de entropía para $Y_1, Y_2, \ldots, Y_n$, está en Γ_n^*. □

Corolario 59. *Si* $\mathbf{h} \in \Gamma_n^*$, *entonces* $k\mathbf{h} \in \Gamma_n^*$ *para todo* $k \in \mathbb{Z}^+$.

Teorema 60. $\overline{\Gamma}_3^* = \Gamma_3$.

Demostración. Primero nótese que $\overline{\Gamma}_3^* = \Gamma_3$, si y sólo si, $\overline{\Psi}_3^* = \Psi_3$. Como $\Gamma_3^* \subseteq \Gamma_3$ y Γ_3 es cerrado, tomando clausura a ambos lados en la anterior contenencia, se obtiene que $\overline{\Gamma}_3^* \subseteq \Gamma_3$, esto implica que $\overline{\Psi}_3^* \subseteq \Psi_3$. Por lo tanto para probar el teorema, es suficiente con mostrar que $\Psi_3 \subseteq \overline{\Psi}_3^*$.

Primero se demostrará que el punto $(0,0,0,a,a,a,-a) \in \overline{\Psi}_3^*$ para todo $a > 0$. Tomando las entropías con el logaritmo en base 2. Sean X_1 y X_2 variables aleatorias binarias con $Pr\{X_i = 0\} = Pr\{X_i = 1\} = 0.5$, $i = 1, 2$. Sea

$$X_3 = (X_1 + X_2) mod\ 2$$

Es fácil chequear que X_3 tiene la misma distribución marginal que X_1 y X_2, así $H(X_i) = 1$, $i = 1, 2, 3$. Además X_1, X_2 y X_3 son independientes dos a dos y cada variable aleatoria es una función de las otras dos, por lo tanto

$$\mu^*(\widetilde{X}_i - (\widetilde{X}_j \cup \widetilde{X}_k)) = H(X_i|X_j, X_k) = 0 \tag{2.23}$$

$$\begin{aligned} \mu^*(\widetilde{X}_i \cap \widetilde{X}_j - \widetilde{X}_k) &= I(X_i; X_j|X_k) \\ &= H(X_i|X_k) - H(X_i|X_j, X_k) = 1 \end{aligned} \tag{2.24}$$

$$\begin{aligned} \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2 \cap \widetilde{X}_3) &= \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2) - \mu^*(\widetilde{X}_1 \cap \widetilde{X}_2 - \widetilde{X}_3) \\ &= I(X_1; X_2) - I(X_1; X_2|X_3) = 0 - 1 = -1 \end{aligned} \tag{2.25}$$

para todo $i, j, k \in \{1, 2, 3\}$ diferentes.

Sea $\mathbf{h} \in \Gamma_3^*$ la función de entropía para X_1, X_2 y X_3 y sea

$$\mathbf{u} = \mathbf{C}_3^{-1}\mathbf{h}.$$

Como en la prueba del Teorema 57, sean $u_i,\ 1 \leq i \leq 7$ las coordenadas de $\mathcal{I}_3$ que corresponden a los valores de las cantidades en (2.10) respectivamente. Entonces de (2.23), (2.24) y (2.25), se tiene que

$$u_i = \begin{cases} 0, & \text{para } i = 1, 2, 3 \\ 1, & \text{para } i = 4, 5, 6 \\ -1, & \text{para } i = 7 \end{cases}$$

Así, el punto $(0,0,0,1,1,1,-1)$ está en Ψ_3^*. Entonces por el corolario 59 para todo entero positivo k, $(0,0,0,k,k,k,-k)$ esta en Ψ_3^* y por lo tanto, en $\overline{\Psi}_3^*$. Como $\overline{\Gamma}_3^*$ contiene el origen entonces $\overline{\Psi}_3^*$ también contiene el origen. Por teorema 44, $\overline{\Gamma}_3^*$ es convexo esto implica que $\overline{\Psi}_3^*$ es convexo. Por lo tanto, $(0,0,0,a,a,a,-a) \in \overline{\Psi}_3^*$ para todo $a > 0$.

Considérese cualquier $\mathbf{u} \in \Psi_3$. Por (2.11) se tiene que $u_i \geq 0$ $1 \leq i \leq 6$. Así, u_7 es la única componente de u que puede ser negativa. Primero se considera el caso en que $u_7 \geq 0$,

entonces $\mathbf{u}$ está en el ortante no negativo de $\mathcal{I}_3$ y por el Lema 45, $\mathbf{u} \in \Psi_3^*$. Ahora, se considera el caso donde $u_7 < 0$. Sea $\mathbf{t} = (0, 0, 0, -u_7, -u_7, -u_7, u_7)$ entonces $\mathbf{u} = \mathbf{w} + \mathbf{t}$ donde $\mathbf{w} = (u_1, u_2, u_3, u_4 + u_7, u_5 + u_7, u_6 + u_7, 0)$. como $u_7 < 0$ entonces $\mathbf{t} \in \overline{\Psi}_3^*$. De (2.11) se tiene que $u_i + u_7 \geq 0$ para $i = 4, 5, 6$. Así, $\mathbf{w}$ está en el ortante no negativo de $\mathcal{I}_3$ y por lo tanto, $\mathbf{w}$ está en Ψ_3^* por Lema 45. Ahora, dado cualquier $\varepsilon > 0$ sea $\mathbf{t}' \in \Psi_3^*$ tal que

$$\|\mathbf{t} - \mathbf{t}'\| < \varepsilon$$

y sea $\mathbf{u}' = \mathbf{w} + \mathbf{t}'$. Como $\mathbf{w}$ y $\mathbf{t}'$ están en Ψ_3^*, por lema 58, u' también está en Ψ_3^*, y

$$\|\mathbf{u} - \mathbf{u}'\| = \|\mathbf{t} - \mathbf{t}'\| < \varepsilon$$

Por lo tanto, $\mathbf{u} \in \overline{\Psi_3^*}$, entonces $\Psi_3 \subseteq \overline{\Psi}_3^*$. □

Nota 61. *Se mostró que la región $\overline{\Gamma_n^*}$ caracteriza completamente todas las desigualdades de la información sin restricciones involucrando n variables aleatorias. Como $\overline{\Gamma}_3^* = \Gamma_3$, se sigue que no existen desigualdades de la información sin restricciones involucrando tres variables aleatorias diferentes de las desigualdades tipo Shannon.*

2.2. Desigualdades tipo no-Shannon sin restricciones

En esta sección se mostrará que para el caso $n = 4$ aparecen desigualdades tipo no-Shannon, y por lo tanto $\overline{\Gamma}_4^* \neq \Gamma_4$.

La primera desigualdad tipo no-Shannon la encontraron Zhang y Yeung [8], se presentará la demostración que hicieron R. Dougherty, C. Freiling y K. Zeger [4], de esta desigualdad.

Lema 62. *Sean X_1, X_2, X_3, X_4 variable aleatorias con distribución conjunta. Entonces, existe otra variable aleatoria Y, conjuntamente distribuida con X_1, X_2, X_3, X_4, con las siguientes propiedades.*

C1. *La distribución marginal de (X_1, X_2, X_3) y (X_1, X_2, Y) son las mismas, con Y reemplazando a X_3.*

C2. $I(X_3, X_4; Y | X_1, X_2)$.

En este caso se dice que Y es una X_4-Copia de X_3 sobre (X_1, X_2).

Demostración. Sean X_1, X_2, X_3 y X_4 variables aleatorias, con sus respectivos alfabetos $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4$. Sean x_1, x_2, x_3, x_4 denotando elementos arbitrarios de X_1, X_2, X_3, X_4 respectivamente, con probabilidad $p(x_1, x_2, x_3, x_4)$. Sea Y una nueva variable y sea y elemento arbitrio de su alfabeto, el cual es $\mathcal{X}_3$. Se define la distribución conjunta de X_1, X_2, X_3, X_4, Y por

$$p'(x_1, x_2, x_3, x_4, y) = \frac{p(x_1, x_2, x_3, x_4) p(x_1, x_2, y)}{p(x_1, x_2)}. \tag{2.26}$$

Es claro que ésta es no negativo. Sumando sobre los y se obtiene.

$$\sum_y p'(x_1, x_2, x_3, x_4, y) = \frac{p(x_1, x_2, x_3, x_4) \sum_y p(x_1, x_2, y)}{p(x_1, x_2)} = p(x_1, x_2, x_3, x_4), \tag{2.27}$$

así, p' es una extensión de p, lo cual implica que la suma de todas las probabilidades p' es 1. Similarmente, la distribución marginal de (X_1, X_2, Y) esta dada por

$$\sum_{x_3,x_4} p'(x_1,x_2,x_3,x_4,y) = \sum_{x_3,x_4} p(x_1,x_2,x_3,x_4)\frac{p(x_1,x_2,y)}{p(x_1,x_2)} = p(x_1,x_2,y). \tag{2.28}$$

Mientras la distribución marginal de (X_1, X_2, X_3) esta dado por $p(x_1, x_2, x_3)$, demostrándose (C1).

Si se escribe (C2) en términos de entropías, se obtiene

$$H(X_1,X_2,X_3,X_4) + H(X_1,X_2,Y) - H(X_1,X_2) - H(X_1,X_2,X_3,X_4,Y) = 0,$$

pero

$$H(X_1,X_2,X_3) = H(X_1,X_2,Y) \textit{ por } (C1).$$

Así, resta mostrar que

$$H(X_1,X_2,X_3,X_4,Y) = H(X_1,X_2,X_3,X_4) + H(X_1,X_2,X_3) - H(X_1,X_2).$$

$$\begin{aligned}
H(X_1,X_2,X_3,X_4,Y) &= -\sum_{x_1,x_2,x_3,x_4,y} p'(x_1,x_2,x_3,x_4,y)\log(p'(x_1,x_2,x_3,x_4,y)) \\
&= -\sum_{x_1,x_2,x_3,x_4,y} p'(x_1,x_2,x_3,x_4,y)\log\left(\frac{p(x_1,x_2,x_3,x_4)p(x_1,x_2,y)}{p(x_1,x_2)}\right) \\
&= -\sum_{x_1,x_2,x_3,x_4}\left(\sum_y p'(x_1,x_2,x_3,x_4,y)\right)\log(p(x_1,x_2,x_3,x_4)) \\
&- \sum_{x_1,x_2,y}\left(\sum_{x_3,x_4} p'(x_1,x_2,x_3,x_4,y)\right)\log(p(x_1,x_2,y)) \\
&+ \sum_{x_1,x_2}\left(\sum_{x_3,x_4,y} p'(x_1,x_2,x_3,x_4,y)\right)\log(p(x_1,x_2)) \\
&= -\sum_{x_1,x_2,x_3,x_4} p(x_1,x_2,x_3,x_4)\log(p(x_1,x_2,x_3,x_4)) \\
&- \sum_{x_1,x_2,y} p(x_1,x_2,y)\log(p(x_1,x_2,y)) \\
&+ \sum_{x_1,x_2} p(x_1,x_2)\log(p(x_1,x_2)) \\
&= H(X_1,X_2,X_3,X_4) + H(X_1,X_2,X_3) - H(X_1,X_2)
\end{aligned}$$

□

Teorema 63. *(Desigualdad de Zhang-Yeung) Sean X_1, X_2, X_3 y X_4 variables aleatorias arbitrarias. Entonces,*

$$\begin{aligned}
I(X_1;X_2) \leq\ & 2I(X_1;X_2|X_3) + I(X_1;X_3|X_2) + I(X_2;X_3|X_1) \\
& + I(X_1;X_2|X_4) + I(X_3;X_4)
\end{aligned} \tag{2.29}$$

Demostración. Expandiendo en entropías mutuas y cancelando términos se obtiene la siguiente identidad para 5 variables

$$
\begin{aligned}
& I(X_1;X_2) \\
& + I(X_3;Y|X_1) && \text{(S)} \\
& + I(X_3;X_4|Y) && \text{(S)} \\
& + I(X_1,X_2;Y|X_3,X_4) && \text{(S)} \\
& + I(X_4;Y|X_2) && \text{(S)} \\
& + I(X_1;X_2|Y,X_4) && \text{(S)} \\
& + I(X_4;Y|X_1) && \text{(S)} \\
& + I(X_3;Y|X_2) && \text{(S)} \\
& + I(X_1;X_2|X_3,Y) && \text{(S)} \\
& + I(X_3;Y|X_1,X_2,X_4) && \text{(S)} \\
& + I(X_4;Y|X_1,X_2,X_4) && \text{(S)} \\
& = 2I(X_1;X_2|X_3) + I(X_1;X_3|X_2) + I(X_2;X_3|X_1) \\
& + I(X_1;X_2|X_4) + I(X_3;X_4) \\
& + 3I(X_3,X_4;Y|X_1,X_2) && \text{(C2)} \\
& + I(X_1;X_2|Y) - I(X_1;X_2|X_3) && \text{(C1)} \\
& + I(X_1;Y|X_2) - I(X_1;X_3|X_2) && \text{(C1)} \\
& + I(X_2;Y|X_2) - I(X_1;X_3|X_1) && \text{(C1)}
\end{aligned}
$$

Cada uno de los términos marcados con (S) es una información mutua condicional y son por lo tanto no negativos. Así, si los términos marcados con (S) son borrados y " $=$ " es reemplazada por " $\leq$ ", se obtiene una desigualdad tipo Shannon para 5 variables. Por el Lema 62, escogiendo Y como una X_4-copia de X_3 sobre (X_1, X_2), entonces el término marcado con (C2) es cero por la condición (C2) y cada uno de los términos marcados con (C1) son cero por la condición (C1). Luego se obtiene la desigualdad (2.29). □

Teorema 64. *La desigualdad (2.29) es una desigualdad tipo no-Shannon, y por lo tanto* $\overline{\Gamma}_4^* \neq \Gamma_4$.

Demostración. Para cualquier $a > 0$ se considera el punto $\widetilde{\mathbf{h}}(a) \in \mathcal{H}_4$, donde

$$
\begin{aligned}
& \widetilde{h}_1(a) = \widetilde{h}_2(a) = \widetilde{h}_3(a) = \widetilde{h}_4(a) = 2a, \\
& \widetilde{h}_{12}(a) = \widetilde{h}_{13}(a) = \widetilde{h}_{14}(a) = 3a, \\
& \widetilde{h}_{23}(a) = \widetilde{h}_{24}(a) = 3a, \quad \widetilde{h}_{34}(a) = 4a, \\
& \widetilde{h}_{123}(a) = \widetilde{h}_{124}(a) = \widetilde{h}_{134}(a) = \widetilde{h}_{234}(a) = \widetilde{h}_{1234}(a) = 4a,
\end{aligned}
$$

Es fácil verificar que $\widetilde{\mathbf{h}}(a) \in \Gamma_4$, La desigualdad (2.29) en términos de entropías mutuas es

$$
\begin{aligned}
& 2H(X_1) + 2H(X_2) + H(X_3) + H(X_3,X_4) + 4H(X_1,X_2,X_3) + H(X_1,X_2,X_4) \\
& \leq 3H(X_1,X_2) + 3H(X_1,X_3) + H(X_1,X_4) + 3H(X_2,X_3) + H(X_2,X_4)
\end{aligned} \tag{2.30}
$$

y si se reemplaza las respectivas componentes del vector $\widetilde{\mathbf{h}}(a)$ en la desigualdad (2.30) se obtiene que

$$34a \leq 33a$$

lo cual es una contradicción porque $a > 0$. En otras palabras, $\widetilde{\mathbf{h}}(a)$ no satisface (2.29). Equivalentemente,

$$\widetilde{\mathbf{h}}(a) \notin \{\mathbf{h} \in \mathcal{H}_4 \ : \ \mathbf{h} \text{ satisface } (2.29)\}.$$

Como $\widetilde{\mathbf{h}}(a) \in \Gamma_4$, entonces

$$\Gamma_4 \nsubseteq \{\mathbf{h} \in \mathcal{H}_4 \ : \ \mathbf{h} \text{ satisface } (2.29)\},$$

es decir, (2.29) no es implicada por las desigualdades básicas para cuatro variables aleatorias. Por lo tanto, (2.29) es una desigualdad tipo no-Shannon.

Como (2.29) se cumple para todas las funciones de entropía para cuatro variables aleatorias, se tiene que

$$\Gamma_4^* \subseteq \{\mathbf{h} \in \mathcal{H}_4 \ : \ \mathbf{h} \text{ satisface } (2.29)\}.$$

Tomando la clausura a ambos lados, se tiene que

$$\overline{\Gamma}_4^* \subseteq \{\mathbf{h} \in \mathcal{H}_4 \ : \ \mathbf{h} \text{ satisface } (2.29)\}.$$

Entonces, $\widetilde{\mathbf{h}}(a) \notin \overline{\Gamma}_4^*$ y como $\widetilde{\mathbf{h}}(a) \in \Gamma_4$, se concluye que $\overline{\Gamma}_4^* \neq \Gamma_4$.

□

Ahora se mostrará un método para encontrar desigualdades de la información a partir de otras, el cual es debido a R. Dougherty, C. Freiling y K. Zeger [4]. Se empezará demostrando las siguientes desigualdades de la información restringidas.

Lema 65. *Sean X_1, X_2, X_3, X_4, Y variables aleatorias tal que "Y es una X_4-copia de X_1 sobre (X_2, X_3)". Entonces, las siguiente desigualdades se cumplen:*

(a)

$$\begin{aligned} I(X_1, X_2) \leq\ & I(X_1; X_3|X_2) + I(X_1; X_2|X_4) + I(X_1, Y; X_2, Y) \\ & - I(X_1, Y; X_2, Y|X_4) \end{aligned} \tag{2.31}$$

(b)

$$0 \leq I(X_1; X_2|X_3) - I(X_1, Y; X_2, Y|X_3, Y) \tag{2.32}$$

(c)

$$0 \leq I(X_1; X_3|X_2) - I(X_1, Y; X_3, Y|X_2, Y) \tag{2.33}$$

(d)

$$\begin{aligned} I(X_1, X_2) \leq\ & 2I(X_1; X_2|X_3) + I(X_1; X_3|X_2) + I(X_2; X_3|X_1) \\ & + I(X_1; X_2|X_4) + I(X_3; X_4) - I(X_2, Y; X_3, Y|X_1, Y) \end{aligned} \tag{2.34}$$

(e)

$$0 \leq I(X_1; X_2|X_3) + I(X_3; X_4) - I(X_3, Y; X_4) \tag{2.35}$$

Demostración. **(a)** Expandiendo en entropías mutuas y cancelando términos se obtiene la siguiente identidad para 5 variables

$$
\begin{aligned}
& I(X_1;X_2) \\
& + I(X_4;Y|X_1) && \text{(S)} \\
& + I(X_4;Y|X_2) && \text{(S)} \\
& + I(X_3;Y|X_1,X_2,X_4) && \text{(S)} \\
& + I(X_1;Y|X_2) && \text{(S)} \\
& = I(X_1;X_3|X_2) + I(X_1;X_2|X_4) + I(X_1,Y;X_2,Y) - I(X_1,Y;X_2,Y|X_4) \\
& + I(X_1,X_4;Y|X_2,X_3) && \text{(C2)} \\
& + H(X_1,X_2,X_3) - H(Y,X_2,X_3) && \text{(C1)} \\
& - H(X_1,X_2) + H(Y,X_2) && \text{(C1)}
\end{aligned}
$$

Cada uno de los términos marcados con (S) es una información mutua condicional y son por lo tanto no negativos. Así, si los términos marcados con (S) son borrados y " $=$ " es reemplazada por " $\leq$ ", entonces se obtiene una desigualdad tipo Shannon para 5 variables. Como "Y es una X_4-copia de X_1 sobre (X_2, X_3)" entonces por el Lema 62 el término marcado con (C2) es cero por la condición (C2) y cada uno de los términos marcados con (C1) son cero por la condición (C1). De aquí se obtiene la desigualdad (2.31).

(b) Expandiendo en entropías mutuas y cancelando términos se obtiene la siguiente identidad para 5 variables

$$
\begin{aligned}
& I(X_1;X_2|X_3) - I(X_1,Y;X_2,Y|X_3,Y) \\
& = I(X_1;Y|X_3) && \text{(S)} \\
& + I(X_4;Y|X_1,X_2,X_3) && \text{(S)} \\
& - I(X_1,X_4;Y|X_2,X_3) && \text{(C2)}
\end{aligned}
$$

Como "Y es una X_4-copia de X_1 sobre (X_2, X_3)," entonces por el Lema 62 el término marcado con (C2) es cero por la condición (C2), y como cada uno de los términos marcados con (S) es una información mutua condicional, estos son por lo tanto no negativos. Así, la expresión de información de arriba es no negativa. Luego se obtiene la desigualdad (2.32).

(c) La demostración es similar a la anterior

(d) Expandiendo en entropías mutuas y cancelando términos obtenemos la siguiente iden-

tidad para 5 variables

$$
\begin{aligned}
& I(X_1;X_2) \\
& + I(X_4;Y|X_1) && \text{(S)} \\
& + I(X_4;Y|X_2) && \text{(S)} \\
& + I(X_1,X_2;Y|X_3,X_4) && \text{(S)} \\
& + I(X_3;Y|X_1,X_2,X_4) && \text{(S)} \\
& + I(X_4;Y|X_1,X_2,X_3) && \text{(S)} \\
& + I(X_1;Y|X_2) && \text{(S)} \\
& + I(X_1;Y|X_3) && \text{(S)} \\
& + I(X_3;X_4|Y) && \text{(S)} \\
& + I(X_1;X_2|Y,X_4) && \text{(S)} \\
& = 2I(X_1;X_2|X_3) + I(X_1;X_3|X_2) + I(X_2;X_3|X_1) \\
& + I(X_1;X_2|X_4) + I(X_3;X_4) - I(X_2,Y;X_3,Y|X_1,Y) \\
& + 3I(X_1,X_4;Y|X_2,X_3) && \text{(C2)} \\
& + H(Y) - H(X_1) && \text{(C1)} \\
& - 2(H(Y,X_2) - H(X_1,X_2)) && \text{(C1)} \\
& - 2(H(Y,X_3) - H(X_1,X_3)) && \text{(C1)} \\
& + 3(H(Y,X_2,X_3) - H(X_1,X_2,X_3)) && \text{(C1)}
\end{aligned}
$$

Cada uno de los términos marcados con (S) es una información mutua condicional y son por lo tanto no negativos. Así, si los términos marcados con (S) son borrados y " $=$ " es reemplazada por " $\leq$ ", entonces se obtiene una desigualdad tipo Shannon para 5 variables. Como "Y es una X_4-copia de X_1 sobre (X_2, X_3)," entonces por el Lema 62 el término marcado con (C2) es cero por la condición (C2) y cada uno de los términos marcados con (C1) son cero por la condición (C1). Luego se obtiene la desigualdad (2.34).

(e) Expandiendo en entropías mutuas y cancelando términos se obtiene la siguiente identidad para 5 variables

$$
\begin{aligned}
& I(X_1;X_2|X_3) + I(X_3;X_4) - I(X_3,Y;X_4) \\
& = I(X_1,X_2;Y|X_3,X_4) && \text{(S)} \\
& - I(X_1,X_4;Y|X_2,X_3) && \text{(C2)} \\
& + H(Y,X_2,X_3) - H(X_1,X_2,X_3) && \text{(C1)}
\end{aligned}
$$

Como "Y es una X_4-copia de X_1 sobre (X_2, X_3)," entonces por el Lema 62 el término marcado con (C2) es cero por la condición (C2), y cada uno de los términos marcados con (C1) son cero por la condición (C1), y como cada uno de los términos marcados con (S) es una información mutua condicional, estos son por lo tanto no negativos. Así, la expresión de información de arriba es no negativa. Luego se obtiene la desigualdad (2.35).

□

Lema 66. *Sean X_1, X_2, X_3, X_4 variables aleatorias arbitrarias. Si*

$$\begin{aligned} aI(X_1;X_2) \leq\ & bI(X_1;X_2|X_3) + cI(X_1;X_3|X_2) + dI(X_2;X_3|X_1) \\ & + aI(X_1;X_2|X_4) + hI(X_3;X_4) \end{aligned} \tag{2.36}$$

es una desigualdad de la información con coeficientes no negativos, entonces

$$\begin{aligned} (a+d)I(X_1;X_2) \leq\ & (b+2d+h)I(X_1;X_2|X_3) + (a+c+d)I(X_1;X_3|X_2) + dI(X_2;X_3|X_1) \\ & + (a+d)I(X_1;X_2|X_4) + (d+h)I(X_3;X_4) \end{aligned} \tag{2.37}$$

es también una desigualdad de la información.

Demostración. Como a, b, c, d y h son reales no negativos, de las desigualdades (2.31), (2.32), (2.33), (2.34), (2.35) se obtiene la siguiente desigualdad restringida

$$\begin{aligned} (a+d)I(X_1;X_2) \leq\ & (b+2d+h)I(X_1;X_2|X_3) + (a+c+d)I(X_1;X_3|X_2) + dI(X_2;X_3|X_1) \\ & + (a+d)I(X_1;X_2|X_4) + (d+h)I(X_3;X_4) \\ & + [aI(X_1,Y;X_2,Y) - bI(X_1,Y;X_2,Y|X_3,Y) \\ & - cI(X_1,Y;X_3,Y|X_2,Y) - dI(X_2,Y;X_3,Y|X_1,Y) \\ & - aI(X_1,Y;X_2,Y|X_4) - hI(X_3,Y;X_4)] \end{aligned} \tag{2.38}$$

donde "Y es una X_4-copia de X_1 sobre (X_2, X_3)." Como por hipótesis

$$\begin{aligned} aI(X_1;X_2) \leq\ & bI(X_1;X_2|X_3) + cI(X_1;X_3|X_2) + dI(X_2;X_3|X_1) \\ & + aI(X_1;X_2|X_4) + hI(X_3;X_4) \end{aligned}$$

es una desigualdad de la información (es decir, siempre se cumple, luego no hay problema si se sustituye X_1, X_2, X_3, X_4 por $(X_1, Y), (X_2, Y), (X_3, Y)(X_4, Y)$ entonces la expresión de información encerrada entre corchetes cuadrados en (2.38) es menor o igual a cero. Por lo tanto

$$\begin{aligned} (a+d)I(X_1;X_2) \leq\ & (b+2d+h)I(X_1;X_2|X_3) + (a+c+d)I(X_1;X_3|X_2) + dI(X_2;X_3|X_1) \\ & + (a+d)I(X_1;X_2|X_4) + (d+h)I(X_3;X_4), \end{aligned}$$

también es una desigualdad de la información. Con lo cual queda probado el lema. □

Ahora se mostrará una colección infinita de desigualdades de información, debidas a F. Matus [2]

Corolario 67. *Sean X_1, X_2, X_3, X_4 variables aleatorias arbitrarias, entonces se tienen la siguientes desigualdades de la información sin restricciones.*

$$\begin{aligned} sI(X_1;X_2) \leq\ & \frac{s(s+3)}{2} I(X_1;X_2|X_3) + \frac{s(s+1)}{2} I(X_1;X_3|X_2) \\ & + I(X_2;X_3|X_1) + sI(X_1;X_2|X_4) + sI(X_3;X_4), \end{aligned} \tag{2.39}$$

donde s es un entero no negativo.

Demostración. (Por inducción sobre s.) Si $s = 0$, entonces (2.39), se convierte en

$$0 \leq I(X_2; X_3|X_1), \tag{2.40}$$

la cual es una desigualdad básica, que fue probada en la Proposición 20.
Si $s = 1$ entonces (2.39), se convierte en

$$\begin{aligned} I(X_1; X_2) \leq\ & 2I(X_1; X_2|X_3) + I(X_1; X_3|X_2) + I(X_2; X_3|X_1) \\ & + I(X_1; X_2|X_4) + I(X_3; X_4) \end{aligned}$$

La cual es la desigualdad de Zhang-Yeung (2.29) que ya se probó que es una desigualdad de la información no-Shannon.
Ahora, se supone válida la afirmación para s y se demostrará para $s + 1$. Por hipótesis

$$\begin{aligned} sI(X_1; X_2) \leq\ & \frac{s(s+3)}{2} I(X_1; X_2|X_3) + \frac{s(s+1)}{2} I(X_1; X_3|X_2) \\ & + I(X_2; X_3|X_1) + sI(X_1; X_2|X_4) + sI(X_3; X_4) \end{aligned}$$

es una desigualdad de la información, entonces aplicando el Lema 66, se obtiene que:

$$\begin{aligned} (s+1)I(X_1; X_2) \leq\ & \frac{(s+1)(s+4)}{2} I(X_1; X_2|X_3) + \frac{(s+1)(s+2)}{2} I(X_1; X_3|X_2) \\ & + I(X_2; X_3|X_1) + (s+1)I(X_1; X_2|X_4) + (s+1)I(X_3; X_4) \end{aligned}$$

es también una desigualdad de la información. Con lo cual queda demostrado el Corolario. □

Para terminar la Sección, se mostrará que el cono $\overline{\Gamma}_4^*$ no es poliédrico, es decir, el cono $\overline{\Gamma}_4^*$ no se puede expresar como la intersección de un número finito de semiespacios cerrados. Este hecho es debido a F. Matus [2]. Para esto se necesitará de algunos resultados previos.

Proposición 68. *Si* $\mathbf{h}$ *y* $\mathbf{h}'$ *están en* Γ_n^*, *entonces* $a\mathbf{h}+a'\mathbf{h}'$ *esta en* $\overline{\Gamma}_n^*$, *para todo* $a, a' \in \mathbb{R}^+$.

Demostración. Como para toda $k \in \mathbb{Z}^+$, $k\mathbf{h} \in \Gamma_n^*$ y $k\mathbf{h}' \in \Gamma_n^*$, por el Corolario 59 y como $\overline{\Gamma}_n^*$, contiene al origen y es convexo entonces $a\mathbf{h}, a'\mathbf{h}' \in \overline{\Gamma}_n^*$ para todo $a, a' \in \mathbb{R}^+$. Por la observación anterior y la convexidad de $\overline{\Gamma}_n^*$, se tiene que $a\mathbf{h} + a'\mathbf{h}'$ está en $\overline{\Gamma}_n^*$, para todo $a, a' \in \mathbb{R}^+$. □

Lema 69. *Si una curva* $g : [0, a] \mapsto A : p \to g(p)$, *donde* a *es un número real positivo y* A *es un conjunto poliédrico, tiene tangente lateral derecha* $g'_+(0) = \lim_{p\to 0^+} \frac{1}{p}[g(p) - g(0)]$ *en el punto* $g(0)$, *entonces* A *contiene al punto* $g(0) + \varepsilon g'_+(0)$ *para algún* $\varepsilon > 0$.

Demostración. Se asume que A es el conjunto poliédrico de los $\mathbf{x} \in \mathbb{R}^d$ tal que satisfacen $\mathbf{v}_k\mathbf{x} \leq t_k$, $1 \leq k \leq l$, donde l es un entero positivo, $t_k \in \mathbb{R}$ y $\mathbf{v}_k \in \mathbb{R}^d$. El punto $g(0)$ en A satisface las desigualdades $\mathbf{v}_k g(0) \leq t_k$, $1 \leq k \leq l$, de lo cual, después de reenumerar si es necesario, existe un entero l' tal que $\mathbf{v}_k g(0) = t_k$, para $k \leq l'$ y $\mathbf{v}_k g(0) < t_k$, para $l' < k \leq l$. Entonces, $\mathbf{v}_k g'_+(0) \leq 0$ para $k \leq l'$, y si $\varepsilon > 0$ es suficientemente pequeño $\mathbf{v}_k(g(0) + \varepsilon g'_+(0)) < t_k$, para $l' < k \leq l$. Esta observación implica que las desigualdades $\mathbf{v}_k(g(0)+\varepsilon g'_+(0)) < t_k$, $1 \leq k \leq l$, se cumplen para algún $\varepsilon > 0$, con lo cual queda probada la afirmación. □

Ahora se construirán ciertas funciones de entropía que se necesitarán para lograr el objetivo.

Proposición 70. *Sea $\mathcal{N}_4 = \{1,2,3,4\}$. Para $I \subseteq \mathcal{N}_4$, $I \neq \emptyset$ y $t \in \{0,1,\ldots,|\mathcal{N}_4 \setminus I|\}$ sea*

$$r_t^I(J) = \min\{t, |J \setminus I|\}, \quad J \subseteq \mathcal{N}_4, \tag{2.41}$$

entonces el vector

$$\begin{aligned} r_t^I :=& (r_t^I(1), r_t^I(2), r_t^I(3), r_t^I(4), r_t^I(12), r_t^I(13), r_t^I(14), r_t^I(23), r_t^I(24), \\ & r_t^I(34), r_t^I(\{123), r_t^I(124), r_t^I(\{134), r_t^I(234), r_t^I(1234)), \end{aligned} \tag{2.42}$$

(donde $r_t^I(1234)) := r_t^I(\{1,2,3,4\}))$, etc) es entrópico.

Demostración. Como $I \neq \emptyset$ entonces $1 \leq t \leq 3$. Para la demostración de está afirmación, se deben encontrar variables aleatorias X_1, X_2, X_3, X_4, tal que r_t^I es la función de entropía de esas variables. Se tomarán las funciones de entropía en base 2 y se hará la demostración por casos.
Si $t = 0$, $r_0^I = 0$ entonces r_0^I es la función de entropía de las variables X_1, X_2, X_3, X_4, donde todas estás variables son degeneradas.
Si $t = 1$, para todo $i \in I$ se toma X_i como una variable degenerada y para todo $i \in \mathcal{N}_4 \setminus I$ se toma $X_i = X$, donde X es la variable aleatoria que toma valores en $\{0,1\}$ con $Pr\{X = 0\} = Pr\{X = 1\} = 0.5$. Es fácil verificar que r_t^I es la función de entropía de las variables aleatorias X_1, X_2, X_3, X_4.
Si $t = 2$, para todo $i \in I$ se toma X_i como una variable degenerada y para todo $i \in \mathcal{N}_4 \setminus I$ se toma X_i, como una variable aleatoria que toma valores en $\{0,1\}$ con $Pr\{X = 0\} = Pr\{X = 1\} = 0.5$, y el conjunto de las variables $A = \{X_i : i \in \mathcal{N}_4 \setminus I\}$ se toma de tal forma que las variables son independientes dos a dos y si $|I| = 1$, el conjunto de las variables A es dependiente . Es fácil verificar que r_t^I es la función de entropía de las variables aleatorias X_1, X_2, X_3, X_4.
Si $t = 3$, para todo $i \in I$ se toma X_i como una variable degenerada y para todo $i \in \mathcal{N}_4 \setminus I$ se toma X_i, como una variable aleatoria que toma valores en $\{0,1\}$ con $Pr\{X = 0\} = Pr\{X = 1\} = 0.5$, y las variables $\{X_i : i \in \mathcal{N}_4 \setminus I\}$ se toman mutuamente independientes. Es fácil verificar que r_t^I es la función de entropía de las variables aleatorias X_1, X_2, X_3, X_4. □

Lema 71. *Sea $p \mapsto g(p)$ con $0 \leq p \leq \frac{1}{2}$ la curva dada por*

$$(\ln 2) g(p) = h_{X,p} + \beta(p) r_1^{14} + [\ln 2 + 2p \ln 2 - \frac{1}{2}\beta(2p)](r_1^{23} + r_2^4), \tag{2.43}$$

donde β denota la función $p \mapsto -p \ln p - (1-p)\ln(1-p)$ y $h_{X,p}$ es el vector entrópico dado en el Ejemplo 42.

Entonces, $g(0) = r_1^{14} + r_1^{23} + r_2^4$, y la tangente lateral derecha $g'_+(0) = \lim_{p \to 0^+} \frac{1}{p}[g(p) - g(0)] = f_{12} + r_1^{24} + r_2^3$ donde f_{12} es el vector dado por

$$\begin{aligned} f_{12} :=& (f_{12}(1), f_{12}(2), f_{12}(3), f_{12}(4), f_{12}(12), f_{12}(13), f_{12}(14), f_{12}(23), f_{12}(24), \\ & f_{12}(34), f_{12}(123), f_{12}(124), f_{12}(134), f_{12}(234), f_{12}(1234)), \end{aligned} \tag{2.44}$$

donde para $J \subseteq \mathcal{N}_4$

$$f_{12}(J) = \begin{cases} 3, & si \quad J \in \{13, 23, 14, 24, 34\} \\ \min\{4, 2|J|\}, & en \quad otro \quad caso. \end{cases}$$

Para simplificar notación se toma $f_{12}(1234) := f_{12}(1,2,3,4)$*, etc.*

Demostración. $g(0) = r_1^{14} + r_1^{23} + r_2^4$, ya que $\beta(0) = 0$ y $h_{X,0} = (\ln 2) r_1^{14}$. Por otro lado es fácil verificar que

$$\begin{aligned} g'_+(0) &= \lim_{p\to 0^+} \frac{1}{p}[g(p) - g(0)] \\ &= (4,3,3,3,7,5,6,5,5,5,7,7,7,7,7) \\ &= f_{12} + r_1^{24} + r_2^3 \end{aligned}$$

□

Teorema 72. *El cono* $\overline{\Gamma}_4^*$ *no es poliédrico*

Demostración. La curva $p \mapsto g(p)$ dada en el Lema 71 tiene tangente lateral derecha $g'_+(0)$ y como $h_{X,p}, r_1^{14}, r_1^{23}$ y r_2^4 son entrópicos, por la proposición 68, la curva $p \mapsto g(p)$ tiene rango en el cono convexo $\overline{\Gamma}_4^*$.
Si en la desigualdad de información (2.39) cambiamos X_1 por X_4, X_2 por X_3, X_3 por X_2, y X_4 por X_1, se obtiene la siguiente desigualdad de información

$$\begin{aligned} sI(X_3;X_4) \leq \frac{s(s+3)}{2} I(X_3;X_4|X_2) + \frac{s(s+1)}{2} I(X_2;X_4|X_3) \\ + I(X_2;X_3|X_4) + sI(X_3;X_4|X_1) + sI(X_1;X_2). \end{aligned} \tag{2.45}$$

Si se escribe la desigualdad de la información (2.45) en su forma canónica, está es igual a

$$\begin{aligned} &-\frac{s(s+1)}{2}H(X_2) - \frac{s(s+3)}{2}H(X_3) - (s+1)H(X_4) - sH(X_1,X_2) \\ &+ sH(X_1,X_3) + sH(X_1,X_4) + s(s+2)H(X_2,X_3) + \frac{(s+1)(s+2)}{2}H(X_2,X_4) \\ &+ \frac{(s+1)(s+2)}{2}H(X_3,X_4) - sH(X_1,X_3,X_4) - (s+1)^2 H(X_2,X_3,X_4) \geq 0. \end{aligned} \tag{2.46}$$

Entonces, la desigualdad de información (2.39), puede escribirse como $f(\mathbf{h}) = \mathbf{b}^T\mathbf{h}$ donde

$$\begin{aligned} \mathbf{b}^T =& (0, -\frac{s(s+1)}{2}, -\frac{s(s+3)}{2}, -(s+1), -s, s, s, s(s+2), \\ & \frac{(s+1)(s+2)}{2}, \frac{(s+1)(s+2)}{2}, 0, 0, -s, -(s+1)^2, 0). \end{aligned}$$

Es fácil verificar que $f(g(0) + g'_+(0)) = 1 - \varepsilon s$. Por lo tanto, para un s suficientemente grande $g(0)+g'_+(0)$ no satisface la desigualdad de información (2.46) y así este no pertenece a $\overline{\Gamma}_4^*$. Esta observación y el Lema 69 implica que $\overline{\Gamma}_4^*$ no es poliédrico. □

CAPÍTULO 3

Entropía y Grupos

Sean X_1 y X_2 variables aleatorias arbitrarias. Entonces,

$$H(X_1) + H(X_2) \geq H(X_1, X_2) \tag{3.1}$$

es equivalente a la desigualdad básica

$$I(X_1; X_2) \geq 0. \tag{3.2}$$

Sean G cualquier grupo finito y G_1 y G_2 subgrupos de G. Se mostrará en la Sección 3.4 que

$$\mid G \mid\mid G_1 \cap G_2 \mid \geq \mid G_1 \mid\mid G_2 \mid, \tag{3.3}$$

donde $\mid G \mid$ denota el orden de G ($G_1 \cap G_2$ es también subgrupo de G). Reagrupando términos, la anterior desigualdad se puede escribir como

$$\log \frac{\mid G \mid}{\mid G_1 \mid} + \log \frac{\mid G \mid}{\mid G_2 \mid} \geq \log \frac{\mid G \mid}{\mid G_1 \cap G_2 \mid}. \tag{3.4}$$

Comparando (3.1) y (3.4), se puede identificar fácilmente una correspondencia uno a uno entre esas dos desigualdades, a saber que X_i corresponde a G_i, $i = 1, 2$, y (X_1, X_2) corresponde a $G_1 \cap G2$. Mientras (3.1) es verdadero para cualquier par de variables aleatorias X_1 y X_2, (3.4) es verdadero para cualquier grupo finito G y subgrupos G_1 y G_2.
En el Capítulo 1, se mostró que la región Γ_n^* caracteriza todas las desigualdades de información (involucrando n variables aleatorias). En particular, se vio en la Sección 2.2, que la región $\overline{\Gamma_n^*}$ es suficiente para caracterizar todas las desigualdades de información sin restricciones, es decir, conociendo $\overline{\Gamma_n^*}$ uno puede determinar si alguna desigualdad de la información sin restricción siempre se cumple. El objetivo principal de este Capítulo es obtener una caracterización de $\overline{\Gamma_n^*}$ en términos de grupos finitos. Una consecuencia importante de este resultado es una correspondencia uno a uno entre desigualdades de información sin restricciones y desigualdades en grupos. Específicamente, para cualquier desigualdad de información sin restricciones, existe una correspondiente desigualdad de grupos, y vice versa. Un caso especial de esta correspondencia fue dada en (3.1) y (3.4). Por medio de este resultado, las desigualdades de la información sin restricciones pueden probarse mediante la Teoría de grupos, y unas ciertas formas de desigualdades en Teoría

de grupos pueden probarse mediante técnicas de la Teoría de información. En particular, la desigualdad sin restricción tipo no-Shannon (2.29) corresponde a la desigualdad en grupos

$$\begin{aligned}&|G_1 \cap G_2|^3|G_1 \cap G_3|^3|G_2 \cap G_3|^3|G_1 \cap G_4||G_2 \cap G_4| \\ &\leq |G_1|^2|G_2|^2|G_3||G_3 \cap G_4||G_1 \cap G_2 \cap G_3|^4|G_1 \cap G_2 \cap G_4|,\end{aligned} \tag{3.5}$$

donde los G_i son subgrupos de un grupo finito G, $i = 1, 2, 3, 4$. El significado de esta desigualdad y sus implicaciones en la Teoría de grupos están todavía estudiándose.

3.1. Nociones Básicas de Teoría de Grupos

Definición 73. *Un grupo $\langle G, *\rangle$ es un conjunto G junto con una operación binaria $*$ que satisface los siguientes axiomas:*

1. $(a * b) \in G$ *Para cada* $a, b \in G$ *(Clausurativa).*
2. *Para cada* $a, b, c \in G$, $a * (b * c) = (a * b) * c$ *(Asociatividad).*
3. *Existe un elemento* $e \in G$ *tal que* $e * a = a * e = a$ *para todas las* $a \in G$. *(Este elemento* e *es un elemento identidad para* $*$ *en* G*).*
4. *Para cada* $a \in G$ *existe un elemento* $a' \in G$ *tal que* $a * a' = a' * a = e$. *(El elemento* a' *es un inverso de* a *respecto* a $*$*).*

Definición 74. *El número de elementos de un grupo G se llama el orden de G. Si $|G| < \infty$, G se llama un grupo finito, si no, esté se llama un grupo infinito.*

Ejemplo 75. *(GRUPO DE PERMUTACIONES)*
Considere una permutación de las componentes de un vector $\mathbf{x} = (x_1, x_2, \ldots, x_r)$ *dada por* $\sigma[\mathbf{x}] = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(r)})$, *donde* $\sigma : \{1, 2, \ldots, r\} \mapsto \{1, 2, \ldots, r\}$ *es una función uno a uno. La función uno a uno* σ *se llama una permutación sobre* $\{1, 2, \ldots, r\}$ *que es representada por* $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(r))$.
Para dos permutaciones σ_1 *y* σ_2, *se define* $\sigma_1 \circ \sigma_2$ *como la función compuesta de* σ_1 *y* σ_2. *El conjunto de todas las permutaciones sobre* $\{1, 2, \ldots, r\}$ *y la operación "* $\circ$ *" forman un grupo (llamado el grupo de permutaciones). El orden de este grupo es* $(r!)$.

Definición 76. *Sea $\langle G, *\rangle$ un grupo, y H un subconjunto de G. Si $\langle H, *\rangle$ es un grupo, entonces H se llama un subgrupo de G.*

Definición 77. *Sean $\langle G, *\rangle$ un grupo, H un subgrupo de G, y a un elemento de G, la clase lateral a izquierda de H con respecto a a, es el conjunto $a * H = \{a * h : h \in H\}$. Similarmente, la clase lateral a derecha de H con respecto a a, es el conjunto $H * a = \{h * a : h \in H\}$.*

En lo que sigue, sólo se usarán clases laterales a izquierda. Para simplificar, $a * H$ se nota por aH.

Proposición 78. *Para a_1, a_2 en G, a_1H y a_2H son o idénticos o disjuntos. Además, a_1H y a_2H son idénticos, si y sólo si, a_1 y a_2 pertenecen a la misma clase lateral izquierda de H*

Demostración. Si a_1H y a_2H no son disjuntos, entonces existe un elemento $b \in a_1H \cap a_2H$, luego $b = a_1 \circ h_1 = a_2 \circ h_2$, para algún h_i en H , $i = 1, 2$. Entonces, $a_1 = (a_2 \circ h_2) \circ h_1^{-1} = a_2 \circ (h_2 \circ h_1^{-1}) = a_2 \circ t$, donde $t = h_2 \circ h_1^{-1}$ esta en H. Se tiene que $a_1H \subseteq a_2H$, ya que si $a_1 \circ h \in a_1H$, donde $h \in H$, $a_1 \circ h = (a_2 \circ t) \circ h = a_2 \circ (t \circ h) = a_2 \circ u$, donde $u = t \circ h$ esta en H. Esto implica que, $a_1 \circ h$ esta en a_2H. Así, $a_1H \subseteq a_2H$. Análogamente, se muestra que $a_2H \subseteq a_1H$, luego $a_1H = a_2H$. Por lo tanto, si a_1H y a_2H no son disjuntos, entonces son idénticos.
Ahora se probará la segunda parte de la proposición. Como H es un grupo, contiene al elemento identidad e. Entonces, para cualquier elemento a del grupo G, $a = a \circ e$ está en aH. Si a_1H y a_2H son idénticos, entonces como $a_1 \in a_1H = a_2H$ y $a_2 \in a_2H = a_1H$, por lo tanto a_1 y a_2 están en la misma clase lateral izquierda. Por otro lado, en la primera parte de la proposición, se vio que un elemento del grupo pertenece a sólo una clase lateral a derecha de H. Como a_1 está en a_1H y a_2 está en a_2H, entonces si a_1 y a_2 pertenecen a la misma clase lateral izquierda de H, tenemos que a_1H y a_2H son idénticos. La proposición está probada. □

Proposición 79. *Sean, H un subgrupo de un grupo G y a un elemento de G, entonces $|aH| = |H|$, es decir, todas las clases laterales a izquierdas de H tienen el mismo número de elementos y esté es igual al orden de H.*

Demostración. Considere dos elementos $a \circ h_1$ y $a \circ h_2$ en aH, donde h_1 y h_2 están en H tal que $a \circ h_1 = a \circ h_2$. Entonces, $a^{-1} \circ (a \circ h_1) = a^{-1} \circ (a \circ h_2)$, por lo tanto $h_1 = h_2$.
Así, a cada elemento en H le corresponde un único elemento en aH, luego, $|aH| = |H|$ para todo $a \in G$. □

Teorema 80. *(TEOREMA DE LAGRANGE) Si H es un grupo de G, entonces $|H|$ divide a $|G|$.*

Demostración. Como $a \in aH$ para todo $a \in G$, todo elemento de G pertenece a una clase lateral a izquierda de H. Entonces, de la Proposición 78, se obtiene que las distintas clases laterales izquierdas de H generan una partición de G. Por lo tanto, $|G|$ el número de elementos en G, es igual al número de las distintas clases laterales a izquierda de H multiplicado por el número de elementos en cada clase lateral a izquierda, el cual es igual a $|H|$ (por la Proposición 79). Esto implica que $|H|$ divide a $|G|$. □

Corolario 81. *Sea H un subgrupo de un grupo G. El número de las distintas clases laterales izquierdas de H es igual a $\frac{|G|}{|H|}$.*

3.2. Funciones de entropía grupo-caracterizables.

Proposición 82. *Sean G_1 y G_2 subgrupos de un grupo G. Entonces, $G_1 \cap G_2$ también es un subgrupo de G.*

Corolario 83. *Sean $G_1, G_2, \ldots, G_n$ subgrupos de un grupo G. Entonces $\cap_{i=1}^{n} G_i$ también es un subgrupo de G.*

En el resto de este Capítulo, se tomará $\mathcal{N}_n = \{1, 2, \ldots, n\}$ y se denotará $\cap_{i\in\alpha} G_i$ por G_α, donde α es un subconjunto no vacío de $\mathcal{N}_n$.

Lema 84. *Sea G_i un subgrupo de un grupo G y a_i un elemento de G, para cada $i \in \alpha$, y α es un subconjunto no vacío de $\mathcal{N}_n$. Entonces,*

$$| \cap_{i\in\alpha} a_i G_i | = \begin{cases} | G_\alpha |, & si \quad \cap_{i\in\alpha} a_i G_i \neq \phi. \\ 0, & en \quad otro \quad caso. \end{cases} \tag{3.6}$$

Demostración. Para el caso especial donde α es un singleton, es decir, $\alpha = \{i\}$ para algún $i \in \mathcal{N}_n$, la igualdad se reduce a $|a_i G_i| = |G_i|$, la cual ya se probó.
Sea α cualquier subconjunto no vacío de $\mathcal{N}_n$. Si $\cap_{i\in\alpha} a_i G_i = \phi$, entonces (3.6) se tiene trivialmente. Si $\cap_{i\in\alpha} a_i G_i \neq \phi$, entonces existe $x \in \cap_{i\in\alpha} a_i G_i$ tal que para todo $i \in \alpha$

$$x = a_i \circ s_i,$$

donde $s_i \in G_i$. Para cualquier $i \in \alpha$ y $y \in G_\alpha$, considere

$$x \circ y = (a_i \circ s_i) \circ y = a_i \circ (s_i \circ y).$$

Como s_i y y están en G_i, $s_i \circ y$ está en G_i, entonces $x \circ y$ está en $a_i G_i$, para todo $i \in \alpha$, luego $x \circ y$ está en $\cap_{i\in\alpha} a_i G_i$. Además, para $y, y' \in G_\alpha$, si $x \circ y = x \circ y'$, entonces $y = y'$. Por lo tanto, a cada elemento en G_α le corresponde un único elemento en $\cap_{i\in\alpha} a_i G_i$. Así,

$$|G_\alpha| \leq | \cap_{i\in\alpha} a_i G_i|.$$

Por otro lado, si $z \in \cap_{i\in\alpha} a_i G_i \neq \phi$, entonces $z = a_i \circ t_i$ para todo $i \in \alpha$, con $t_i \in G_i$, luego $z = (x \circ s_i^{-1}) \circ t_i = x \circ (s_i^{-1} \circ t_i)$, para todo $i \in \alpha$, por lo tanto $s_i^{-1} \circ t_i = s_j^{-1} \circ t_j$, para todo $i, j \in \alpha$, entonces $s_i^{-1} \circ t_i \in G_\alpha$, para todo $i \in \alpha$. De lo anterior se concluye que, para todo $z \in \cap_{i\in\alpha} a_i G_i$, existe un único $y \in G_\alpha$ tal que $z = x \circ y$. Así,

$$| \cap_{i\in\alpha} a_i G_i| \leq |G_\alpha|.$$

Con lo cual queda probado que $|G_\alpha| = | \cap_{i\in\alpha} a_i G_i|$. □

Teorema 85. *Sea G_i, para cada $i \in \mathcal{N}_n$ un subgrupo de un grupo G. Entonces, $\mathbf{h} \in \mathcal{H}_n$ definido por:*

$$h_\alpha = \log \frac{| G |}{| G_\alpha |} \tag{3.7}$$

para todo subconjunto no vacío α de $\mathcal{N}_n$ es entrópico, es decir, $\mathbf{h} \in \Gamma_n^*$.

Demostración. Es suficiente con mostrar que existe una colección de variables aleatorias $X_1, X_2, \cdots, X_n$ tal que

$$H(X_\alpha) = \log \frac{| G |}{| G_\alpha |} \tag{3.8}$$

para todo subconjunto no vacío α de $\mathcal{N}_n$. Primero se introducirá una variable aleatoria uniforme Λ definida sobre el espacio muestral G con función de probabilidad de masa

$$Pr\{\Lambda = a\} = \frac{1}{| G |}, \tag{3.9}$$

para todo $a \in G$. Para cada $i \in \mathcal{N}_n$, sea la variable aleatoria X_i una función de Λ tal que $X_i = aG_i$ si $\Lambda = a$. Sea α un subconjunto no vacío de $\mathcal{N}_n$. Como $X_i = a_iG_i$ para todo $i \in \alpha$, si y sólo si, Λ es igual a algún $b \in \cap_{i\in\alpha} a_iG_i$.

$$Pr\{X_i = a_iG_i\} = \frac{\mid \cap_{i\in\alpha} a_iG_i \mid}{\mid G \mid} = \begin{cases} \frac{|G_\alpha|}{|G|}, & \text{si} \quad \cap_{i\in\alpha} a_iG_i \neq \emptyset. \\ 0, & en \quad otro \quad caso. \end{cases}$$

por el Lema 3.6. En otras palabras, $(X_i, i \in \alpha)$ se distribuye uniformemente sobre su soporte cuya cardinalidad es $\frac{|G|}{|G_\alpha|}$. Entonces, (3.8) se obtiene y el teorema queda probado. □

Definición 86. *Sean G un grupo finito, $G_1, G_2, \cdots, G_n$ subgrupos de G, y $\boldsymbol{h}$ un vector en $\mathcal{H}_n$. Si $h_\alpha = \log \frac{|G|}{|G_\alpha|}$ para todo subconjunto no vacío α de $\mathcal{N}_n$, entonces $(G, G_1, \cdots, G_n)$ en un grupo de caracterización de $\boldsymbol{h}$.*

El teorema anterior afirma que ciertas funciones de entropía en Γ_n^* tiene un grupo de caracterización. Estas son llamadas funciones de entropía grupo-caracterizables, los siguientes son ejemplos de tales funciones de entropía.

Ejemplo 87. *Fijando cualquier subconjunto β de $\mathcal{N}_3 = \{1, 2, 3\}$ y definiendo un vector $\boldsymbol{h} \in \mathcal{H}_3$ por*

$$h_\alpha = \begin{cases} \log 2, & si \quad \alpha \cap \beta \neq \emptyset. \\ 0, & en \quad otro \quad caso. \end{cases}$$

Se mostrará que $\boldsymbol{h}$ tiene un grupo de caracterización. Sea $G = \{0, 1\}$ el grupo de adición módulo 2 y para $i = 1, 2, 3.$, sea

$$G_i = \begin{cases} \{0\}, & si \quad i \in \beta. \\ G, & en \quad otro \quad caso. \end{cases}$$

Entonces, para un subconjunto no vacío α de $\mathcal{N}_3$, si $\alpha \cap \beta \neq \emptyset$, existe un i en α tal que i está también en β, y por lo tanto por definición $G_i = \{0\}$. Así,

$$G_\alpha = \bigcap_{i\in\alpha} G_i = \{0\}. \tag{3.10}$$

Luego,

$$\log \frac{\mid G \mid}{\mid G_\alpha \mid} = \log \frac{\mid G \mid}{\mid \{0\} \mid} = \log 2. \tag{3.11}$$

Si $\alpha \cap \beta = \emptyset$, entonces $G_i = G$ para todo $i \in \alpha$, y $G_\alpha = \bigcap_{i\in\alpha} G_i = G$. Así,

$$\log \frac{\mid G \mid}{\mid G_\alpha \mid} = \log \frac{\mid G \mid}{\mid G \mid} = \log 1 = 0. \tag{3.12}$$

Entonces, se tiene que $h_\alpha = \log \frac{|G|}{|G_\alpha|}$ para todo subconjunto no vacío α de $\mathcal{N}_3$ por lo tanto, (G, G_1, G_2, G_3) es un grupo de caracterización de $\boldsymbol{h}$.

Ejemplo 88. *Esta es una generalización del ejemplo anterior. Fijando algún subconjunto no vacío β de $\mathcal{N}_n$ y definiendo un vector $\boldsymbol{h} \in \mathcal{H}_n$ por*

$$h_\alpha = \begin{cases} \log 2, & si \quad \alpha \cap \beta \neq \emptyset. \\ 0, & en \quad otro \quad caso. \end{cases}$$

entonces, $(G, G_1, \cdots, G_n)$ *es un grupo de caracterización de* $\boldsymbol{h}$*, donde* G *es el grupo de adición módulo 2, y*

$$G_i = \begin{cases} \{0\}, & si \quad i \in \beta. \\ G, & en \quad otro \quad caso. \end{cases}$$

tomando $\beta = \emptyset$*,* $\boldsymbol{h} = 0$*. Así se ve que* $(G, G_1, G_2, \cdots, G_n)$ *es un grupo de caracterización del origen de* $\mathcal{H}_n$*, con* $G = G_1 = G_2 = \ldots = G_n$*.*

Ejemplo 89. *Definiendo un vector* $\boldsymbol{h}$ *en* $\mathcal{H}_3$ *como sigue:*

$$h_\alpha = \min\{\mid \alpha \mid, 2\} \tag{3.13}$$

Sea $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ *y* $G_1 = \{(0,0),(1,0)\}, G_2 = \{(0,0),(0,1)\}, G_3 = \{(0,0),(1,1)\}$ *entonces* (G, G_1, G_2, G_3) *es un grupo de caracterización de* $\boldsymbol{h}$*.*

3.3. Una caracterización con grupos de $\overline{\Gamma}_n^*$.

Se introdujo en la Sección anterior la clase de funciones de entropías en Γ_n^* que tienen un grupo de caracterización. Sin embargo, una función de entropía $\mathbf{h} \in \Gamma_n^*$ puede no tener un grupo de caracterización, debido a la siguiente observación. Suponga $\mathbf{h} \in \Gamma_n^*$, entonces existe una colección de variables aleatorias $X_1, \cdots, X_n$ tal que $h_\alpha = H(X_\alpha)$ para todo subconjunto no vacío α de $\mathcal{N}_n$. Si $(G, G_1, \cdots, G_n)$ es un grupo de caracterización de $\mathbf{h}$, entonces

$$H(X_\alpha) = \log \frac{\mid G \mid}{\mid G_\alpha \mid} \tag{3.14}$$

para todo subconjunto no vacío de $\mathcal{N}_n$. Donde ambos $\mid G \mid$ y $\mid G_\alpha \mid$ son enteros, luego $H(X_\alpha)$ debe ser el logaritmo de un número racional. Sin embargo, la entropía conjunta de un conjunto de variables aleatorias no es necesariamente el logaritmo de un número racional. Por lo tanto, es posible construir una función de entropía $\mathbf{h} \in \Gamma_n^*$, que no tiene un grupo de caracterización.

Aunque $\mathbf{h} \in \Gamma_n^*$ no implica que $\mathbf{h}$ tenga un grupo de caracterización, se tiene que el conjunto de todos los $\mathbf{h} \in \Gamma_n^*$ que tienen un grupo de caracterización son casi suficientes para caracterizar la región Γ_n^*, como se verá a continuación.

Definición 90. *Se define la siguiente región en* $\mathcal{H}_n$

$$\Upsilon_n = \{\boldsymbol{h} \in \mathcal{H}_n : \boldsymbol{h} \textit{ tiene un grupo de caracterización}\}. \tag{3.15}$$

Por el Teorema 85, si $\mathbf{h} \in \mathcal{H}_n$ tiene un grupo de caracterización, entonces $\mathbf{h} \in \Gamma_n^*$. Por lo tanto, $\Upsilon_n \subseteq \Gamma_n^*$.

Lema 91. *Para cualquier* $n \in \mathbb{Z}^+$*,*

$$n \ln n - n \leq \ln n! \leq (n+1)\ln(n+1) - n. \tag{3.16}$$

Demostración. Primero, si se escribe $\ln n! = \ln 1 + \ln 2 + \cdots + \ln n$, como $\ln x$ es una función creciente, se tiene que,

$$\int_{k-1}^{k} \ln x \; dx < \ln k < \int_{k}^{k+1} \ln x \; dx.$$

Entonces,

$$\sum_{k=1}^{n} \left(\int_{k-1}^{k} \ln x \; dx \right) < \sum_{k=1}^{n} (\ln k) < \sum_{k=1}^{n} \left(\int_{k}^{k+1} \ln x \; dx \right).$$

De donde,

$$\int_{0}^{n} \ln x dx < \ln n! < \int_{1}^{n+1} \ln x dx.$$

Se sigue que,

$$(x \ln x - x \mid_0^n) < \ln n! < \left(x \ln x - x \mid_1^{n+1} \right).$$

En consecuencia,

$$n \ln n - n < \ln n! < (n+1) \ln(n+1) - n.$$

□

Lema 92. *Sea X una variable aleatoria tal que, $\mid \mathcal{X} \mid < \infty$ y su distribución $\{p(x)\}$ es racional, es decir, $p(x)$ es un número racional para todo $x \in \mathcal{X}$. Sin pérdida de generalidad, se asume que $p(x)$ es un número racional con denominador q para todo $x \in \mathcal{X}$. Entonces, para $r = q, 2q, 3q, \cdots$,*

$$\lim_{r \to \infty} \frac{1}{r} \log \frac{r!}{\prod_x (rp(x))!} = H(X) \tag{3.17}$$

Demostración. Aplicando el lema anterior, se obtiene

$$\begin{aligned}
\frac{1}{r} \ln \frac{r!}{\prod_x (rp(x))!} &= \frac{1}{r} \left[\ln(r!) - \ln(\prod_x (rp(x))!) \right] \\
&= \frac{1}{r} \left[\ln(r!) - \sum_x \ln(rp(x))! \right] \\
&\leq \frac{1}{r} \left[((r+1)\ln(r+1) - r) - \sum_x (rp(x) \ln(rp(x)) - rp(x)) \right] \\
&\leq \left(\frac{r+1}{r} \right) \ln(r+1) - 1 - \sum_x p(x) \ln(rp(x)) + \sum_x p(x) \\
&\leq \left(\frac{r+1}{r} \right) \ln(r+1) - \sum_x p(x)(\ln r + \ln p(x)) \\
&\leq \left(\frac{r+1}{r} \right) \ln(r+1) - \ln r \left(\sum_x p(x) \right) - \sum_x p(x) \ln p(x) \\
&\leq \left(\frac{r+1}{r} \right) \ln(r+1) - \ln r - \frac{1}{r} \ln r + \frac{1}{r} \ln r - H_e(x) \\
&\leq \left(\frac{r+1}{r} \right) \ln(r+1) - \left(\frac{r+1}{r} \right) \ln r + \frac{1}{r} \ln r + H_e(x) \\
&\leq \left(1 + \frac{1}{r} \right) \ln \left(1 + \frac{1}{r} \right) + \frac{1}{r} \ln r + H_e(x)
\end{aligned}$$

Esta cota superior tiende a $H_e(x)$ cuando $r \to \infty$. Por otro lado, se obtiene

$$\begin{aligned}
\frac{1}{r}\ln\frac{r!}{\prod_x (rp(x))!} &= \frac{1}{r}\left[\ln(r!) - \ln(\prod_x (rp(x))!)\right] \\
&= \frac{1}{r}\left[\ln(r!) - \sum_x \ln(rp(x))!\right] \\
&\geq \frac{1}{r}\left[r\ln r - r - \sum_x \left((rp(x)+1)\ln(rp(x)+1) - rp(x)\right)\right] \\
&\geq \ln r - 1 - \sum_x \left(p(x)+\frac{1}{r}\right)\left(\ln\left(p(x)+\frac{1}{r}\right) + \ln r\right) - \sum_x p(x) \\
&\geq \ln r - \sum_x \left(p(x)+\frac{1}{r}\right)\ln\left(p(x)+\frac{1}{r}\right) - \ln r \sum_x \left(p(x)+\frac{1}{r}\right) \\
&\geq -\sum_x \left(p(x)+\frac{1}{r}\right)\ln\left(p(x)+\frac{1}{r}\right) - \frac{\ln r}{r}\left(\sum_x 1\right) \\
&\geq -\sum_x \left(p(x)+\frac{1}{r}\right)\ln\left(p(x)+\frac{1}{r}\right) - \frac{\ln r}{r} \mid \mathcal{X} \mid
\end{aligned}$$

Esta cota inferior tiende a $H_e(x)$ cuando $r \to \infty$ (ya que por hipótesis $\mid \mathcal{X} \mid < \infty$). Entonces, la demostración se completa haciendo un cambio de base del logaritmo si fuese necesario. □

Proposición 93. *Para cualquier $\boldsymbol{h} \in \Gamma_n^*$, es siempre posible construir una sucesión $\{\boldsymbol{h}^k\}$ en Γ_n^* tal que $\lim_{k\to\infty} \boldsymbol{h}^k = \boldsymbol{h}$ donde $\boldsymbol{h}^k$ es una función de entropía de una colección de variables aleatorias $X_1^k, X_2^k, \cdots, X_n^k$ con alfabetos finitos.*

Demostración. Sea $\mathbf{h} \in \Gamma_n^*$ una función de entropía de una colección de variables aleatorias $X_1, X_2, \cdots, X_n$. Si las variables aleatorias tienen alfabetos finitos el resultado ya se tiene. Se verá que ocurre cuando las variables aleatorias tienen alfabetos infinitos. Por ejemplo, si

$$\begin{aligned}
\mathcal{X}_1 &= \{x_{11}, x_{12}, x_{13}, \cdots\}, \\
\mathcal{X}_2 &= \{x_{21}, x_{22}, x_{23}, \cdots\}, \\
&\vdots \\
\mathcal{X}_n &= \{x_{n1}, x_{n2}, x_{n3}, \cdots\}.
\end{aligned}$$

Se supone que $H(X_i) < \infty$ para todo $i \in \{1, 2, \cdots, n\}$.
Para todo $m \geq 1$, se definen las variables aleatorias binarias:

$$B(i,m) = \begin{cases} 1, & si \quad X_i \in \{x_{i1}, x_{i2}, \cdots, x_{im}\}. \\ 0, & en \quad otro \quad caso. \end{cases}$$

Se considera,

$$H(X_i) = -\sum_{k=1}^{m} Pr\{X_i = x_{ik}\}\log Pr\{X_i = x_{ik}\} \; - \sum_{k=m+1}^{\infty} Pr\{X_i = x_{ik}\}\log Pr\{X_i = x_{ik}\}$$

Cuando $m \to \infty$,

$$-\sum_{k=1}^{m} Pr\{X_i = x_{ik}\} \log Pr\{X_i = x_{ik}\} \to H(X_i) \tag{3.18}$$

ya que $H(X_i) < \infty$ y

$$\sum_{k=m+1}^{\infty} Pr\{X_i = x_{ik}\} \log Pr\{X_i = x_{ik}\} \to 0, \tag{3.19}$$

cuando $k \to \infty$. Ahora, se considera

$$\begin{aligned} H(X_i) &= H(X_i \mid B(i,m)) + I(X_i; B(i,m)) \\ &= H(X_i \mid B(i,m) = 1) Pr\{B(i,m) = 1\} \\ &+ H(X_i \mid B(i,m) = 0) Pr\{B(i,m) = 0\} + I(X_i; B(i,m)) \end{aligned} \tag{3.20}$$

Se definen las variables aleatorias X_i^m que toman valores en $\mathcal{X}_i^m = \{x_{i1}, x_{i2}, \cdots, x_{im}\}$ tal que

$$Pr\{X_i^m = x_{ik}\} = \frac{Pr\{X_i = x_{ik}\}}{Pr\{X_i \in \mathcal{X}_i^m\}}$$

para todo $k \in \{1, 2, \cdots, m\}$.
Como

$$\begin{aligned} Pr\{X_i = x_{ik} \mid B(i,m) = 1\} &= \frac{Pr\{X_i = x_{ik}, B(i,m) = 1\}}{Pr\{B(i,m) = 1\}} \\ &= \begin{cases} \dfrac{Pr\{X_i = x_{ik}\}}{Pr\{X_i \in \mathcal{X}_i^m\}}, & \text{si} \quad k \leq m. \\ 0, & \text{si} \quad k > m. \end{cases} \end{aligned} \tag{3.21}$$

Entonces,

$$\begin{aligned} H(X_i \mid B(i,m) = 1) &= -\sum_{k=1}^{\infty} Pr\{X_i = x_{ik} \mid B(i,m) = 1\} \log Pr\{X_i = x_{ik} \mid B(i,m) = 1\} \\ &= -\sum_{k=1}^{m} \frac{Pr\{X_i = x_{ik}\}}{Pr\{X_i \in \mathcal{X}_i^m\}} \log \frac{Pr\{X_i = x_{ik}\}}{Pr\{X_i \in \mathcal{X}_i^m\}} \\ &= -\sum_{k=1}^{m} Pr\{X_i^m = x_{ik}\} \log Pr\{X_i^m = x_{ik}\} \\ &= H(X_i^m) \end{aligned}$$

Reemplazando la ecuación anterior en (3.20) se tiene que

$$\begin{aligned} H(X_i) &= H(X_i^m) Pr\{B(i,m) = 1\} \\ &+ H(X_i \mid B(i,m) = 0) Pr\{B(i,m) = 0\} + I(X_i; B(i,m)) \end{aligned} \tag{3.22}$$

cuando $m \to \infty$, $H(B(i,m)) \to 0$ ya que $Pr\{B(i,m) = 1\} \to 1$. Esto implica que $I(X_i; B(i,m)) \to 0$, por que

$$I(X_i; B(i,m)) \leq H(B(i,m)) \quad (H(B(i,m)) = H(B(i,m)|X_i) + I(X_i; B(i,m))). \tag{3.23}$$

De (3.22) adicionalmente se considera:

$$
\begin{aligned}
&H(X_i|B(i,m)=0)Pr\{B(i,m)=0\}\\
&\quad=\left(-\sum_{k=m+1}^{\infty}\frac{Pr\{X_i=x_{ik}\}}{Pr\{B(i,m)=0\}}\log\frac{Pr\{X_i=x_{ik}\}}{Pr\{B(i,m)=0\}}\right)Pr\{B(i,m)=0\}\\
&\quad=-\sum_{k=m+1}^{\infty}Pr\{X_i=x_{ik}\}\left(\log Pr\{X_i=x_{ik}\}-\log Pr\{B(i,m)=0\}\right)\\
&\quad=-\sum_{k=m+1}^{\infty}Pr\{X_i=x_{ik}\}\log Pr\{X_i=x_{ik}\}\\
&\quad+\left(\sum_{k=m+1}^{\infty}Pr\{X_i=x_{ik}\}\right)\log Pr\{B(i,m)=0\}\\
&\quad=-\sum_{k=m+1}^{\infty}Pr\{X_i=x_{ik}\}\log Pr\{X_i=x_{ik}\}+Pr\{B(i,m)=0\}\log Pr\{B(i,m)=0\}
\end{aligned}
$$

Cuando $m\to\infty$, la suma anterior tiende a 0 por (3.19) y debido a que $Pr\{B(i,m)=0\}\log\{B(i,m)\}\to 0$, (ya que $Pr\{B(i,m)=0\}\to 0$).
Por lo tanto,

$$H(X_i|B(i,m))Pr\{B(i,m)=0\}\to 0$$

y se obtiene de (3.22) que $H(X_i^m)\to H(X_i)$ cuando $m\to\infty$, luego, se ha probado la proposición. □

Proposición 94. *Sean $a_1,a_2,\cdots,a_k$ números reales no negativos tal que $\sum_{i=1}^k a_i=1$, entonces dado $\epsilon>0$, existen $b_1,b_2,\cdots,b_k$ números racionales no negativos tales que $|a_i-b_i|<\epsilon$ y $\sum_{i=1}^k b_i=1$.*

Demostración. Por la densidad de los racionales en los reales, para cada $1\le i\le k-1$ existen $b_i\in\mathbb{Q}$ tal que $0\le b_i\le a_i$ y $|a_i-b_i|<\frac{\epsilon}{k-1}<\epsilon$. Sea $b_k=1-(b_1+\cdots+b_{k-1})\ge 0$, entonces

$$
\begin{aligned}
|a_k-b_k|&=|1-(a_1+\cdots+a_{k-1})-1+(b_1+\cdots+b_{k-1})|\\
&=|(b_1-a_1)+(b_2-a_2)+\cdots+(b_{k-1}-a_{k-1})|\\
&\le|b_1-a_1|+|b_2-a_2|+\cdots+|b_{k-1}-a_{k-1}|\\
&<\sum_{i=1}^{k-1}\frac{\epsilon}{k-1}\\
&=\epsilon
\end{aligned}
$$

Con lo cual queda probada la proposición □

Se probará como un corolario del siguiente teorema que $\overline{con}(\Upsilon_n)$, la clausura convexa de Υ_n es igual a $\overline{\Gamma}_n^*$.

Teorema 95. *Dado cualquier $\boldsymbol{h}\in\Gamma_n^*$, existe una sucesión $\{\boldsymbol{f}^{(r)}\}$ en Υ_n tal que*

$$\lim_{r\to\infty}\frac{1}{r}\boldsymbol{f}^{(r)}=\boldsymbol{h}. \tag{3.24}$$

Demostración. Para todo $\mathbf{h} \in \Gamma_n^*$, existe una colección de variables aleatorias $X_1, X_2, \cdots, X_n$ tal que:

$$h_\alpha = H(X_\alpha), \tag{3.25}$$

para todo subconjunto no vacío α de $\mathcal{N}_n$. Primero se considera el caso especial que $|\mathcal{X}_i| < \infty$, para todo $i \in \mathcal{N}_n$ y la distribución conjunta de $X_1, X_2, \cdots, X_n$ es racional. Se busca mostrar que existe una sucesión $\{\mathbf{f}^{(r)}\}$ en Υ_n tal que $\lim_{r\to\infty} \frac{1}{r}\mathbf{f}^{(r)} = \mathbf{h}$.
Se denotará $\prod_{i\in\alpha} \mathcal{X}_i$ por $\mathcal{X}_\alpha$. Para todo subconjunto no vacío α de $\mathcal{N}_n$, sea Q_α la distribución marginal de $\mathcal{X}_\alpha$. Se asume, sin pérdida de generalidad, que para cualquier subconjunto no vacío α de $\mathcal{N}_n$, y para todo $a \in \mathcal{X}_\alpha$, $Q_\alpha(a)$ es un número racional con denominador q. Para cada $r = q, 2q, 3q, \cdots$, fijamos una sucesión

$$\mathbf{x}_{\mathcal{N}_n} = (x_{\mathcal{N}_n,1}, x_{\mathcal{N}_n,2}, \cdots, x_{\mathcal{N}_n,r}),$$

donde para todo $j = 1, 2, \cdots, r$, $x_{\mathcal{N}_n,j} = (x_{i,j} : i \in \mathcal{N}_n) \in \mathcal{X}_{\mathcal{N}_n}$ tal que, $N(a|\mathbf{x}_{\mathcal{N}_n})$, el número de ocurrencias de a en la sucesión $\mathbf{x}_{\mathcal{N}_n}$, es igual a $rQ_{\mathcal{N}_n}(a)$ para todo $a \in \mathcal{X}_{\mathcal{N}_n}$. La existencia de tal sucesión es garantizada por que todos los valores de la distribución conjunta de $X_{\mathcal{N}_n}$ son números racionales con denominador q.
Además, se denotará la sucesión de r elementos de $\mathcal{X}_\alpha$, $(x_{\alpha,1}, x_{\alpha,2}, \cdots, x_{\alpha,r})$ donde $x_{\alpha,j} = (x_{i,j} : i \in \alpha)$, por $\mathbf{x}_\alpha$. Sea $a \in \mathcal{X}_\alpha$, $N(a|\mathbf{x}_\alpha)$ el número de ocurrencias de a en la sucesión $\mathbf{x}_\alpha$ es igual a $rQ_\alpha(a)$ para todo $a \in \mathcal{X}_\alpha$, ya que como $N(b|\mathbf{x}_{\mathcal{N}_n}) = rQ_{\mathcal{N}_n}(b)$ para todo $b \in \mathcal{X}_{\mathcal{N}_n}$, entonces, $N(a|\mathbf{x}_\alpha) = \sum_{a\subseteq b} N(b|\mathbf{x}_{\mathcal{N}_n}) = \sum_{a\subseteq b} rQ_{\mathcal{N}_n}(b) = rQ_\alpha(a)$.
Sea G el grupo de permutaciones sobre $\{1, 2, \cdots, r\}$. El grupo G depende de r, para simplificar, no se presentará esta dependencia explícitamente. Para todo $i \in \mathcal{N}_n$, se define $G_i = \{\sigma \in G : \sigma[x_i] = x_i\}$, donde $\sigma[x_i] = (x_{i,\sigma(1)}, x_{i,\sigma(2)}, \cdots, x_{i,\sigma(r)})$. Se verá que G_i es un subgrupo de G. Sea σ_1 y $\sigma_2 \in G_i$, entonces $\sigma_1 \circ \sigma_2 \in G_i$, ya que

$$\begin{aligned}
\sigma_1 \circ \sigma_2[\mathbf{x}_i] &= (x_{i,\sigma_1(\sigma_2(1))}, x_{i,\sigma_1(\sigma_2(2))}, \cdots, x_{i,\sigma_1(\sigma_2(r))}) \\
&= (x_{i,\sigma_2(1)}, x_{i,\sigma_2(2)}, \cdots, x_{i,\sigma_2(r)}) \quad \text{porque } \sigma_1 \in G_i \\
&= (x_{i,1}, x_{i,2}, \cdots, x_{i,r}) \quad \text{porque } \sigma_2 \in G_i \\
&= \mathbf{x}_i.
\end{aligned}$$

Es claro que $id \in G_i$, por lo tanto, G_i tiene elemento identidad.
Sea $\sigma \in G_i$, como σ es uno a uno, posee inversa σ^{-1}, veamos que $\sigma^{-1} \in G_i$.

$$\begin{aligned}
\sigma^{-1}[\mathbf{x}_i] &= (x_{i,\sigma^{-1}(1)}, x_{i,\sigma^{-1}(2)}, \cdots, x_{i,\sigma^{-1}(r)}) \\
&= (x_{i,\sigma(\sigma^{-1}(1))}, x_{i,\sigma(\sigma^{-1}(2))}, \cdots, x_{i,\sigma(\sigma^{-1}(r))}) \quad \text{porque } \sigma \in G_i \\
&= (x_{i,1}, x_{i,2}, \cdots, x_{i,r}) = \mathbf{x}_i.
\end{aligned}$$

Observe que la asociatividad de los G_i, es heredada de G. Se concluye que G_i es un grupo, para cada $i = 1, 2, \ldots, n$.
Sea α un subconjunto no vacío de $\mathcal{N}_n$, entonces,

$$\begin{aligned}
G_\alpha &= \bigcap_{i\in\alpha} G_i \\
&= \bigcap_{i\in\alpha} \{\sigma \in G : \sigma[\mathbf{x}_i] = \mathbf{x}_i\} \\
&= \{\sigma \in G : \sigma[\mathbf{x}_i] = \mathbf{x}_i \text{ para todo } i \in \alpha\} \\
&= \{\sigma \in G : \sigma[\mathbf{x}_\alpha] = \mathbf{x}_\alpha\}
\end{aligned}$$

donde

$$\sigma[\mathbf{x}_\alpha] = (x_{\alpha,\sigma(1)}, x_{\alpha,\sigma(2)}, \cdots, x_{\alpha,\sigma(r)}).$$

Para todo $a \in \mathcal{X}_\alpha$ se define el conjunto:

$$L_{\mathbf{x}_\alpha}(a) = \{j \in \{1, 2, \cdots, r\} : x_{\alpha,j} = a\}$$

$L_{\mathbf{x}_\alpha}(a)$ contiene las "locaciones"de a en $\mathbf{x}_\alpha$. Entonces, $\sigma[\mathbf{x}_\alpha] = \mathbf{x}_\alpha$, si y sólo si, para todo $a \in \mathcal{X}_\alpha$, $j \in L_{\mathbf{x}_\alpha}(a)$, implica que $\sigma_j \in L_{\mathbf{x}_\alpha}(a)$. Puesto que,

$$|L_{\mathbf{x}_\alpha}(a)| = N(a|\mathbf{x}_\alpha) = rQ_\alpha(a),$$

$$|G_\alpha| = \prod_{a \in \mathcal{X}_\alpha} (rQ_\alpha(a))!.$$

Por lo tanto,

$$\frac{|G|}{|G_\alpha|} = \frac{r!}{\prod_{a \in \mathcal{X}_\alpha}(rQ_\alpha(a))!}.$$

Por el Lema 92,

$$\lim_{r\to\infty} \frac{1}{r} \log \frac{|G|}{|G_\alpha|} = \lim_{r\to\infty} \frac{1}{r} \log \frac{r!}{\prod_{a \in \chi_\alpha}(rQ_\alpha(a))!} = H(\mathbf{x}_\alpha) = h_\alpha.$$

Recuerde que G, y por lo tanto, sus subgrupos dependen de r. Se define $\mathbf{f}^{(r)}$ por

$$f_\alpha^{(r)} = \log \frac{|G|}{|G_\alpha|},$$

para todo subconjunto no vacío α de $\mathcal{N}_n$. Entonces, $\mathbf{f}^{(r)} \in \Upsilon_n$ y

$$\lim_{r\to\infty} \frac{1}{r}\mathbf{f}^{(r)} = \mathbf{h}$$

se ha probado el teorema para el caso especial en el que $\mathbf{h}$ es la función de entropía de una colección de variables aleatorias $X_1, X_2, \cdots, X_n$ con alfabetos finitos y distribución conjunta racional. Para completar la prueba sólo se tiene que notar que para cualquier $\mathbf{h} \in \Gamma_n^*$, siempre es posible construir una sucesión $\{\mathbf{h}^{(k)}\}$ en Γ_n^* tal que, $\lim_{k\to\infty} \mathbf{h}^{(k)} = \mathbf{h}$, donde $\mathbf{h}^k$ es una función de entropía de una colección de variables aleatorias $X_1^{(k)}, X_2^{(k)}, \cdots, X_n^{(k)}$ con alfabetos finitos y distribución de probabilidades conjunta racional. Esto, debido a las Proposiciones 93 y 94, ya que la Proposición 93 generaliza el teorema para alfabetos infinitos y la Proposición 94 generaliza el teorema para cualquier distribución que tengan las variables. □

Corolario 96. $\overline{con}(\Upsilon_n) = \overline{\Gamma}_n^*$.

Demostración. Primero que todo, $\Upsilon_n \subseteq \Gamma_n^*$, tomando la clausura convexa, se tiene que $\overline{con}(\Upsilon_n) \subseteq \overline{con}\Gamma_n^*$. Por Teorema 44 , $\overline{\Gamma}_n^*$ es convexo. Por lo tanto, $\overline{con}(\Gamma_n^*) = \overline{\Gamma}_n^*$, y se obtiene que $\overline{con}(\Upsilon_n) \subseteq \overline{\Gamma}_n^*$. Por otro lado, se mostró en el ejemplo 88 que el origen de $\mathcal{H}_n$ tiene un grupo de caracterización y por lo tanto, está en Υ_n. Se sigue del Teorema 95, que $\overline{\Gamma}_n^* \subseteq \overline{con}(\Upsilon_n)$, se concluye que $\overline{\Gamma}_n^* = \overline{con}(\Upsilon_n)$, completando la prueba. □

3.4. Desigualdades de información y desigualdades de grupo.

Se probó en el Capítulo 1 que una desigualdad de información sin restricciones

$$\mathbf{b}^T\mathbf{h} \geq 0 \tag{3.26}$$

se cumple, si y sólo si,

$$\overline{\Gamma}_n^* \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^T\mathbf{h} \geq 0\}. \tag{3.27}$$

En otras palabras, toda desigualdad de la información es totalmente caracterizada por $\overline{\Gamma}_n^*$. Además, se probó al final de la Sección anterior que $\overline{con}(\Upsilon_n) = \overline{\Gamma}_n^*$. Puesto que, $\Upsilon_n \subseteq \Gamma_n^* \subseteq \overline{\Gamma}_n^*$, si (3.27) se tiene, entonces

$$\Upsilon_n \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^T\mathbf{h} \geq 0\}. \tag{3.28}$$

Por otro lado, si (3.28) se cumple, como $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^T\mathbf{h} \geq 0\}$ es cerrado y convexo, tomando la clausura convexa en (3.28) se obtiene que,

$$\overline{\Gamma}_n^* = \overline{con}(\Upsilon_n) \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^T\mathbf{h} \geq 0\}. \tag{3.29}$$

Por lo tanto, (3.27) y (3.28) son equivalentes.
Ahora, (3.28) es equivalente a

$$\mathbf{b}^T\mathbf{h} \geq 0 \quad \text{para todo } \mathbf{h} \in \Upsilon_n. \tag{3.30}$$

Puesto que $\mathbf{h} \in \Upsilon_n$, si y sólo si, existen un grupo finito G y subgrupos $G_1, G_2, \cdots, G_n$ de G, tal que

$$h_\alpha = \log \frac{|G|}{|G_\alpha|},$$

para todo subconjunto no vacío α de $\mathcal{N}_n$. Se ve que la desigualdad (3.26), se tiene para cualquier conjunto de variables aleatorias $X_1, X_2, \cdots, X_n$, si y sólo si, la desigualdad obtenida en (3.26) al reemplazar h_α por $\log \frac{|G|}{|G_\alpha|}$ para todo subconjunto no vacío α de $\mathcal{N}_n$ se tiene para todo grupo finito G y subgrupos $G_1, G_2, \cdots, G_n$. En otras palabras, para toda desigualdad de la información sin restricciones, existe una correspondiente desigualdad en grupos y viceversa. Por lo tanto, desigualdades de la Teoría de la información se pueden probar por métodos de la Teoría de grupos , y ciertas desigualdades en Teoría de grupos pueden probarse por métodos de la Teoría de la información.
En el resto de la Sección, se explora esta correspondencia, entre la Teoría de la información y la Teoría de grupos. Se dará una prueba con ayuda de la Teoría de la información, para la desigualdad de la Teoría de grupos (3.5).

Definición 97. *Sean G_1 y G_2 subgrupos de un grupo finito G. Se define*

$$G_1 \circ G_2 = \{a \circ b : a \in G_1 \quad y \quad b \in G_2\}.$$

$G_1 \circ G_2$ no es, en general, un subgrupo de G. Sin embargo, se puede mostrar que $G_1 \circ G_2$ es un subgrupo de G, si G es abeliano.

Proposición 98. *Sean G_1 y G_2 subgrupos de un grupo finito G. Entonces,*

$$|G_1 \circ G_2| = \frac{|G_1||G_2|}{|G_1 \cap G_2|}. \tag{3.31}$$

Demostración. Se fija $(a_1, a_2) \in G_1 \times G_2$, entonces $a_1 \circ a_2$ está en $G_1 \circ G_2$. Sea algún $(b_1, b_2) \in G_1 \times G_2$, tal que

$$b_1 \circ b_2 = a_1 \circ a_2. \tag{3.32}$$

Se determinará el número de (b_1, b_2) en $G_1 \times G_2$ que satisfacen esta relación. De (3.32), se tiene que

$$b_2 \circ a_2^{-1} = b_1^{-1} \circ a_1. \tag{3.33}$$

Sea k este elemento común en G, es decir,

$$k = b_2 \circ a_2^{-1} = b_1^{-1} \circ a_1. \tag{3.34}$$

Como $b_1^{-1} \circ a_1 \in G_1$ y $b_2^{-1} \circ a_2 \in G_2$, entonces $k \in G_1 \cap G_2$. En otras palabras, dado $(a_1, a_2) \in G_1 \times G_2$, $(b_1, b_2) \in G_1 \times G_2$ satisface (3.32), si y sólo si, (b_1, b_2) satisface (3.34), para algún $k \in G_1 \cap G_2$.
Ahora, de (3.34), se obtiene

$$b_1(k) = (k \circ a_1^{-1})^{-1} \tag{3.35}$$

y

$$b_2(k) = k \circ a_2. \tag{3.36}$$

Donde se escribe b_1 y b_2 como $b_1(k)$ y $b_2(k)$ para enfatizar su dependencia de k. Ahora, se consideran $k, k' \in G_1 \cap G_2$ tal que

$$(b_1(k), b_2(k)) = (b_1(k'), b_2(k')). \tag{3.37}$$

Ya que $b_1(k) = b_1(k')$, de (3.35), se obtiene

$$(k \circ a_1^{-1}) = (k' \circ a_1^{-1})^{-1}, \tag{3.38}$$

que implica que $k = k'$.
Por lo tanto, a cada $k \in G_1 \cap G_2$ le corresponde un único par $(b_1, b_2) \in G_1 \times G_2$ que satisface (3.32). Por lo tanto, el número de elementos distintos en $G_1 \circ G_2$ esta dado por

$$|G_1 \circ G_2| = \frac{|G_1 \times G_2|}{|G_1 \cap G_2|} = \frac{|G_1||G_2|}{|G_1 \cap G_2|},$$

completando la prueba. □

Teorema 99. *Sean G_1, G_2 y G_3 subgrupos de un grupo finito G. Entonces,*

$$|G_3||G_{123}| \geq |G_{13}||G_{23}|.$$

Demostración. Primero que todo,

$$G_{13} \cap G_{23} = (G_1 \cap G_3) \cap (G_2 \cap G_3) = G_1 \cap G_2 \cap G_3 = G_{123}. \tag{3.39}$$

Por la proposición anterior, se tiene que

$$|G_{13} \circ G_{23}| = \frac{|G_{13}||G_{23}|}{|G_{123}|}.$$

Es claro que $G_{13} \circ G_{23}$ es un subconjunto de G_3, por lo tanto,

$$|G_{13} \circ G_{23}| = \frac{|G_{13}||G_{23}|}{|G_{123}|} \leq |G_3|.$$

El teorema está probado. □

Corolario 100. *Para variables aleatorias X_1, X_2 y X_3,*

$$I(X_1; X_2|X_3) \geq 0 \tag{3.40}$$

Demostración. Sean G_1, G_2 y G_3 subgrupos de un grupo finito G. Entonces,

$$|G_3||G_{123}| \geq |G_{13}||G_{23}| \tag{3.41}$$

por Teorema 99, o,

$$\frac{|G|^2}{|G_{13}||G_{23}|} \geq \frac{|G|^2}{|G_3||G_{123}|}. \tag{3.42}$$

Esto es equivalente a

$$\log \frac{|G|}{|G_{13}|} + \log \frac{|G|}{|G_{23}|} \geq \log \frac{|G|}{|G_3|} + \log \frac{|G|}{|G_{123}|}. \tag{3.43}$$

Esta desigualdad de grupos, corresponde a la desigualdad de información

$$H(X_1, X_3) + H(X_2, X_3) \geq H(X_3) + H(X_1, X_2, X_3), \tag{3.44}$$

que es equivalente a $I(X_1; X_2|X_3) \geq 0$. □

El anterior Corolario muestra que todas las desigualdades básicas en la Teoría de la información tienen una prueba grupo-teórica. Naturalmente, el Teorema 99, es también implicado por las desigualdades básicas. Como comentario, la desigualdad en (3.3) es vista como un caso especial de la desigualdad (2.29) tomando $G_3 = G$. ($|G||G_{12}| \geq |G_1||G_2|$).
La desigualdad tipo no-Shannon (2.29), puede expresarse en forma canónica como

$$\begin{aligned} &2H(X_1) + 2H(X_2) + H(X_3) + H(X_3, X_4) + 4H(X_1, X_2, X_3) + H(X_1, X_2, X_4) \\ &\leq 3H(X_1, X_2) + 3H(X_1, X_3) + H(X_1, X_4) + 3H(X_2, X_3) + H(X_2, X_4), \end{aligned} \tag{3.45}$$

que corresponde a la desigualdad de grupos

$$\begin{aligned} &2\log \frac{|G|}{|G_1|} + 2\log \frac{|G|}{|G_2|} + \log \frac{|G|}{|G_3|} + \log \frac{|G|}{|G_{34}|} + 4\log \frac{|G|}{|G_{123}|} + \log \frac{|G|}{|G_{124}|} \\ &\leq 3\log \frac{|G|}{|G_{12}|} + 3\log \frac{|G|}{|G_{13}|} + \log \frac{|G|}{|G_{14}|} + 3\log \frac{|G|}{|G_{23}|} + \log \frac{|G|}{|G_{24}|}. \end{aligned} \tag{3.46}$$

Reagrupando los términos, se obtiene

$$\begin{aligned} &|G_1 \cap G_2|^3 |G_1 \cap G_3|^3 |G_2 \cap G_3|^3 |G_1 \cap G_4||G_2 \cap G_4| \\ &\leq |G_1|^2 |G_2|^2 |G_3||G_3 \cap G_4||G_1 \cap G_2 \cap G_3|^4 |G_1 \cap G_2 \cap G_4| \end{aligned} \tag{3.47}$$

que es la desigualdad de grupos (3.5).

Ahora se probarán las desigualdades de la Teoría de Grupos que producen las desigualdades (2.39). La desigualdad (2.39) puede expresarse en forma canónica como:

$$\begin{aligned}
&(s+1)H(X_1) + \frac{s(s+3)}{2}H(X_2) + \frac{s(s+1)}{2}H(X_3) \\
&+ sH(X_3, X_4) + (s^2+2s+1)H(X_1, X_2, X_3) + sH(X_1, X_2, X_4) \\
&\leq \frac{(s+2)(s+1)}{2}H(X_1, X_2) + \frac{(s+2)(s+1)}{2}H(X_1, X_3) + sH(X_1, X_4) \\
&+ s(s+2)H(X_2, X_3) + sH(X_2, X_4)
\end{aligned}$$

que corresponde a la desigualdad de grupos

$$\begin{aligned}
&(s+1)\log\frac{|G|}{|G_1|} + \frac{s(s+3)}{2}\log\frac{|G|}{|G_2|} + \frac{s(s+1)}{2}\log\frac{|G|}{|G_3|} \\
&+ s\log\frac{|G|}{|G_{34}|} + (s^2+2s+1)\log\frac{|G|}{|G_{123}|} + s\log\frac{|G|}{|G_{124}|} \\
&\leq \frac{(s+2)(s+1)}{2}\log\frac{|G|}{|G_{12}|} + \frac{(s+2)(s+1)}{2}\log\frac{|G|}{|G_{13}|} + s\log\frac{|G|}{|G_{14}|} \\
&+ s(s+2)\log\frac{|G|}{|G_{23}|} + s\log\frac{|G|}{|G_{24}|}.
\end{aligned}$$

Reagrupando los términos, se obtiene

$$\begin{aligned}
&|G_1 \cap G_2|^{\frac{(s+2)(s+1)}{2}}|G_1 \cap G_3|^{\frac{(s+2)(s+1)}{2}}|G_1 \cap G_4|^s|G_2 \cap G_3|^{s(s+2)}|G_2 \cap G_4|^s \\
&\leq |G_1|^{(s+1)}|G_2|^{\frac{s(s+3)}{2}}|G_3|^{\frac{s(s+1)}{2}}|G_3 \cap G_4|^s|G_1 \cap G_2 \cap G_3|^{(s^2+2s+1)}|G_1 \cap G_2 \cap G_4|^s.
\end{aligned}$$

3.5. Otros resultados.

Proposición 101. *Sean G_1 y G_2 subgrupos de un grupo finito G. Entonces, $G_1 \circ G_2$ es un subgrupo, si G es abeliano.*

Demostración. Sean $a_1, a_2 \in G_1$ y $b_1, b_2 \in G_2$. entonces

$$\begin{aligned}
(a_1 \circ b_1) \circ (a_2 \circ b_2) &= a_1 \circ (b_1 \circ a_2) \circ b_2 \\
&= a_1 \circ (a_2 \circ b_1) \circ b_2 \\
&= (a_1 \circ a_2) \circ (b_1 \circ b_2).
\end{aligned}$$

Este último término pertenece a $G_1 \circ G_2$, ya que, $a_1 \circ a_2 \in G_1$ y $b_1 \circ b_2 \in G_2$. Luego, $G_1 \circ G_2$ es cerrado para " $\circ$ " si G es abeliano.
Además, $e \in G_1 \circ G_2$, ya que $e = e \circ e$.
Sea $a_1 \circ b_1 \in G_1 \circ G_2$, entonces $a_1^{-1} \circ b_1^{-1} \in G_1 \circ G_2$ es su inverso ya que

$$\begin{aligned}
(a_1 \circ b_1) \circ (a_1^{-1} \circ b_1^{-1}) &= a_1 \circ (b_1 \circ a_1^{-1}) \circ b_1^{-1} \\
&= (a_1 \circ a_1^{-1}) \circ (b_1 \circ b_1^{-1}) \\
&= e.
\end{aligned}$$

□

Proposición 102. *Sean* $\mathbf{g}_1$ *y* $\mathbf{g}_2$ *funciones de entropía grupo caracterizables, entonces* $m_1\mathbf{g}_1 + m_2\mathbf{g}_2$ *es grupo caracterizable, donde* m_1 *y* m_2 *son enteros positivos cualesquiera.*

Demostración. Como $\mathbf{g}_1$ y $\mathbf{g}_2$ son grupo caracterizables, existen $(G, G_1, \cdots, G_n)$ y $(H, H_1, \cdots, H_n)$ donde, G y H son grupos finitos, $G_1, \cdots, G_n$ son subgrupos de G y $H_1, \cdots, H_n$ son subgrupos de H, tales que

$$g_{1\alpha} = \log \frac{\mid G \mid}{\mid G_\alpha \mid} \quad \text{y} \quad g_{2\alpha} = \log \frac{\mid H \mid}{\mid H_\alpha \mid}.$$

entonces,

$$\begin{aligned}(m_1\mathbf{g}_1 + m_2\mathbf{g}_2)_\alpha &= m_1 g_{1\alpha} + m_2 g_{2\alpha} \\ &= m_1 \log \frac{\mid G \mid}{\mid G_\alpha \mid} + m_2 \log \frac{\mid H \mid}{\mid H_\alpha \mid} \\ &= \log \frac{\mid G \mid^{m_1} \mid H \mid^{m_2}}{\mid G_\alpha \mid^{m_1} \mid H_\alpha^{m_2}}\end{aligned}$$

Si se define, $K = G^{m_1} \times H^{m_2}$ el cual es un grupo finito y $K_1 = G_1^{m_1} \times H_1^{m_2}, K_2 = G_2^{m_1} \times H_2^{m_2}, \cdots, K_n = G_n^{m_1} \times H_n^{m_2}$ subgrupos de K entonces $(K, K_1, K_2, \cdots, K_n)$ es un grupo de caracterización de $m_1\mathbf{g}_1 + m_2\mathbf{g}_2$. □

Proposición 103. *Sean* $\mathbf{g}_1$ *y* $\mathbf{g}_2$ *funciones de entropía grupo caracterizables. Para reales positivos* a_1, a_2 *arbitrarios se puede construir una sucesión de funciones de entropía* $\mathbf{f}^{(k)}$ *grupo caracterizables, para* $k = 1, 2, \cdots$ *tal que,* $\lim_{k\to\infty} \frac{\mathbf{f}^{(k)}}{\parallel \mathbf{f}^{(k)} \parallel} = \frac{\mathbf{h}}{\parallel \mathbf{h} \parallel}$, *donde* $\mathbf{h} = a_1\mathbf{g}_1 + a_2\mathbf{g}_2$.

Demostración. Primero se verá el caso cuando a_1 y a_2 son números racionales positivos. Entonces, $a_1 = \frac{p_1}{q_1}$ y $a_2 = \frac{p_2}{q_2}$, luego $a_1 = \frac{p_1 q_2}{q}$ y $a_2 = \frac{p_2 q_1}{q}$ donde $q = q_1 q_2$, por lo tanto $\mathbf{h}$, se puede escribir $\mathbf{h} = \frac{1}{q}(b_1 g_1 + b_2 g_2)$ donde $b_1, b_2, q \in \mathbb{Z}^+$. Por la proposición anterior, $\mathbf{f} = b_1\mathbf{g}_1 + b_2\mathbf{g}_2$ es grupo caracterizable, y se tiene que:

$$\begin{aligned}\frac{\mathbf{h}}{\parallel \mathbf{h} \parallel} &= \frac{\frac{1}{q}(b_1\mathbf{g}_1 + b_2\mathbf{g}_2)}{\parallel \frac{1}{q}(b_1\mathbf{g}_1 + b_2\mathbf{g}_2) \parallel} \\ &= \frac{\frac{1}{q}(b_1\mathbf{g}_1 + b_2\mathbf{g}_2)}{\frac{1}{q} \parallel (b_1\mathbf{g}_1 + b_2\mathbf{g}_2) \parallel} \\ &= \frac{\mathbf{f}}{\parallel \mathbf{f} \parallel}\end{aligned}$$

Luego, queda probada la proposición para este caso.
Se vera ahora, el caso cuando a_1, a_2 son reales positivos. Por la densidad de los números racionales en los números reales, existen sucesiones (b_k) y (c_k) de números racionales positivos tal que $b_k \to a_1$ y $c_k \to a_2$, cuando $k \to \infty$. Si se escribe $b_k = \frac{\widetilde{b_k}}{q_k}$ y $c_k = \frac{\widetilde{c_k}}{q_k}$

donde, $\widetilde{c_k}, \widetilde{b_k}, q_k \in (Z)^+$, entonces $f^{(k)} = \widetilde{b_k} g_1 + \widetilde{c_k} g_2$ (la cual es grupo caracterizable por la Proposición anterior) cumple que

$$\begin{aligned}
\lim_{k\to\infty} \frac{f^{(k)}}{\| f^{(k)} \|} &= \lim_{k\to\infty} \frac{(\widetilde{b_k} g_1 + \widetilde{c_k} g_2)(\frac{1}{q_k})}{\| \widetilde{b_k} g_1 + \widetilde{c_k} g_2 \| (\frac{1}{q_k})} \\
&= \lim_{k\to\infty} \frac{b_k g_1 + c_k g_2}{\| b_k g_1 + c_k g_2 \|} \\
&= \frac{a_1 g_1 + a_2 g_2}{\| a_1 g_1 + a_2 g_2 \|} \\
&= \frac{h}{\| h \|}
\end{aligned}$$

□

Proposición 104. *Sea $(G, G_1, G_2, \cdots, G_n)$ un grupo de caracterización de $\mathbf{g} \in \Gamma_n^*$, donde $\mathbf{g}$ es la función de entropía de las variables aleatorias $X_1, X_2, \cdots, X_n$. Fijando cualquier subconjunto no vacío α de $\mathcal{N}_n$ y definiendo $\mathbf{h}$ por $h_\beta = g_{\alpha\cup\beta} - g_\alpha$ para todo subconjunto no vacío β de $\mathcal{N}_n$. Es fácil verificar que $h_\beta = H(X_\beta \mid X_\alpha)$. Entonces, $(K, K_1, \cdots, K_n)$ es un grupo de caracterización de $\mathbf{h}$, donde $K = G_\alpha$ y $K_i = G_i \cap G_\alpha$.*

Demostración.

$$\begin{aligned}
h_\beta &= g_{\alpha\cup\beta} - g_\alpha \\
&= \log \frac{\mid G \mid}{\mid G_{\alpha\cup\beta} \mid} - \log \frac{\mid G \mid}{\mid G_\alpha \mid} \\
&= H(X_{\alpha\cup\beta}) - H(X_\alpha) \\
&= H(X_\beta \mid X_\alpha).
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
h_\beta &= g_{\alpha\cup\beta} - g_\alpha \\
&= \log \frac{\mid G \mid}{\mid G_{\alpha\cup\beta} \mid} - \log \frac{\mid G \mid}{\mid G_\alpha \mid} \\
&= \log \left(\frac{\mid G \mid}{\mid G_{\alpha\cup\beta} \mid} \cdot \frac{\mid G_\alpha \mid}{\mid G \mid} \right) \\
&= \log \frac{\mid G_\alpha \mid}{\mid G_{\alpha\cup\beta} \mid} \\
&= \log \frac{\mid K \mid}{\mid K_\beta \mid}.
\end{aligned}$$

Entonces, $(K, K_1, K_2, \cdots, K_n)$ es un grupo de caracterización de H. □

Proposición 105. *Sea $(G, G_1, G_2, \cdots, G_n)$ un grupo de caracterización de $\mathbf{g} \in \Gamma_n^*$ donde $\mathbf{g}$ es la función de entropía para las variables aleatorias $X_1, X_2, \cdots, X_n$. Si X_i es una función de $(X_j : j \in \alpha)$, entonces G_α es un subgrupo de G_i.*

Demostración. Como X_i es una función de $(X_j : j \in \alpha)$, $H(X_i|X_\alpha) = 0$, entonces por la Proposición anterior, $h_i = H(X_i|X_\alpha) = 0$, es decir, $\log \left(\frac{|G_\alpha|}{|G_{\alpha\cup\{i\}}|} \right) = 0$, esto implica que,

$|G_\alpha| = |G_{\alpha\cup\{i\}}|$. Así,

$$|\bigcap_{j\in\alpha} G_j| = |\bigcap_{j\in\alpha} G_j \cap G_i|.$$

Por lo tanto,

$$\bigcap_{j\in\alpha} G_j \cap G_i = \bigcap_{j\in\alpha} G_j,$$

entonces $G_\alpha \subseteq G_i$. □

Proposición 106. *Sean G_1, G_2, G_3 subgrupos de un grupo finito G entonces,*

$$|G||G_1 \cap G_2 \cap G_3|^2 \geq |G_1 \cap G_2||G_2 \cap G_3||G_1 \cap G_3|.$$

Demostración.

$$\begin{aligned}
&I(X_1; X_3|X_2) + I(X_2; X_1X_3) \geq 0\\
&\Leftrightarrow (H(X_1, X_2) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_2))\\
&+ (H(X_2) + H(X_1, X_3) - H(X_1, X_2, X_3)) \geq 0\\
&\Leftrightarrow H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3) \geq 2H(X_1, X_2, X_3)\\
&\Leftrightarrow \log\frac{|G|}{|G_{12}|} + \log\frac{|G|}{|G_{23}|} + \log\frac{|G|}{|G_{13}|} \geq 2\log\frac{|G|}{|G_{123}|}\\
&\Leftrightarrow \frac{|G|}{|G_1 \cap G_2|} + \frac{|G|}{|G_2 \cap G_3|} + \frac{|G|}{|G_1 \cap G_3|} \geq \frac{|G|^2}{|G_1 \cap G_2 \cap G_3|^2}\\
&\Leftrightarrow |G||G_1 \cap G_2 \cap G_3|^2 \geq |G_1 \cap G_2||G_2 \cap G_3||G_1 \cap G_3|.
\end{aligned}$$

□

Proposición 107. *Sea $\mathbf{h} \in \Gamma_2^*$, la función de entropía para las variables aleatorias X_1 y X_2 tal que $h_1 + h_2 = h_{1,2}$, es decir, X_1 y X_2 son independientes. Sea (G, G_1, G_2) un grupo de caracterización de $\mathbf{h}$. Entonces, $G = G_1 \circ G_2$.*

Demostración.

$$h_1 = \log\frac{|G|}{|G_1|}, \quad h_2 = \log\frac{|G|}{|G_2|}, \quad h_{12} = \log\frac{|G|}{|G_{12}|}$$

$$h_1 + h_2 = h_{12} \Rightarrow \log\left(\frac{|G|^2|G_{12}|}{|G_1||G_2||G|}\right) = 0 \Rightarrow \frac{|G||G_{12}|}{|G_1||G_2|} = 1 \Rightarrow |G| = \frac{|G_1||G_2|}{|G_{12}|}$$

entonces por (3.31)

$$|G| = \frac{|G_1||G_2|}{|G_1 \cap G_2|} = |G_1 \circ G_2|.$$

Luego, $G_1 \circ G_2 = G$. □

Proposición 108. *(Desigualdad de Ingleton.) Sea G un grupo abeliano finito y G_1, G_2, G_3, G_4 subgrupos de G. Sea (G, G_1, G_2, G_3, G_4) un grupo de caracterización de $\mathbf{g}$, donde $\mathbf{g}$ es una función de entropía para las variables aleatorias X_1, X_2, X_3 y X_4. Entonces, se tienen las siguientes afirmaciones:*

1.
$$|(G_1 \cap G_3) \circ (G_1 \cap G_4)| \leq |G_1 \cap (G_3 \circ G_4)|. \tag{3.48}$$

2.
$$|G_1 \circ G_3 \circ G_4| \leq \frac{|G_1||G_3 \circ G_4||G_1 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4|}. \tag{3.49}$$

3.
$$|G_1 \circ G_2 \circ G_3 \circ G_4| \leq \frac{|G_1 \circ G_3 \circ G_4||G_2 \circ G_3 \circ G_4|}{|G_3 \circ G_4|}. \tag{3.50}$$

4.
$$|G_1 \circ G_2 \circ G_3 \circ G_4| \leq \frac{|G_1||G_2||G_3||G_4||G_1 \cap G_3 \cap G_4||G_2 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4||G_2 \cap G_3||G_2 \cap G_4||G_3 \cap G_4|}. \tag{3.51}$$

5.
$$\begin{aligned} &|G_1 \cap G_3||G_1 \cap G_4||G_2 \cap G_3||G_2 \cap G_4||G_3 \cap G_4| \\ &\leq |G_3||G_4||G_1 \cap G_2||G_1 \cap G_3 \cap G_4||G_2 \cap G_3 \cap G_4|. \end{aligned} \tag{3.52}$$

6.
$$\begin{aligned} &H(X_{13}) + H(X_{14}) + H(X_{23}) + H(X_{24}) + H(X_{34}) \\ &\geq H(X_3) + H(X_4) + H(X_{12}) + H(X_{134}) + H(X_{234}). \end{aligned} \tag{3.53}$$

7. (3.53) es una desigualdad que no siempre se cumple.

Demostración. 1. Primero se verá que $(G_1 \cap G_3) \circ (G_1 \cap G_4) \subseteq G_1 \cap (G_3 \circ G_4)$. Sea $a \in (G_1 \cap G_3) \circ (G_1 \cap G_4)$. Entonces, $a = a_1 \circ a_2$ donde $a_1 \in G_1 \cap G_3$ y $a_2 \in G_1 \cap G_4$, de donde $a_1, a_2 \in G_1$ luego $a_1 \circ a_2 \in G_1$ y como $a_1 \in G_3$ y $a_2 \in G_4$ se sigue que $a_1 \circ a_2 \in G_3 \circ G_4$, por lo tanto, $a = a_1 \circ a_2 \in G_1 \cap (G_3 \circ G_4)$. Así, se tiene que $(G_1 \cap G_3) \circ (G_1 \cap G_4) \subseteq G_1 \cap (G_3 \circ G_4)$ por lo tanto

$$|(G_1 \cap G_3) \circ (G_1 \cap G_4)| \leq |G_1 \cap (G_3 \circ G_4)|.$$

2.
$$\begin{aligned} |G_1 \circ (G_3 \circ G_4)| &= \frac{|G_1||G_3 \circ G_4|}{|G_1 \cap (G_3 \circ G_4)|} && \text{por (3.31)} \\ &\leq \frac{|G_1||G_3 \circ G_4|}{|(G_1 \cap G_3) \circ (G_1 \cap G_4)|} && \text{por (3.48)} \\ &= \frac{|G_1||G_3 \circ G_4||G_1 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4|} && \text{por (3.31).} \end{aligned}$$

3.
$$\begin{aligned} |G_1 \circ G_2 \circ G_3 \circ G_4| &= |(G_1 \circ G_3 \circ G_4) \circ (G_2 \circ G_3 \circ G_4)| \\ &\quad \text{(pues } G \text{ es abeliano y } G_i \circ G_i = G_i) \\ &= \frac{|G_1 \circ G_3 \circ G_4||G_2 \circ G_3 \circ G_4|}{|(G_1 \circ G_3 \circ G_4) \cap (G_2 \circ G_3 \circ G_4)|} && \text{por (3.31)} \end{aligned}$$

Ahora, como $(G_3 \circ G_4) \subseteq (G_1 \circ G_3 \circ G_4) \cap (G_2 \circ G_3 \circ G_4)$, entonces

$$|G_3 \circ G_4| \leq |(G_1 \circ G_3 \circ G_4) \cap (G_2 \circ G_3 \circ G_4)|$$

por lo tanto,

$$|G_1 \circ G_2 \circ G_3 \circ G_4| \leq \frac{|G_1 \circ G_3 \circ G_4||G_2 \circ G_3 \circ G_4|}{|G_3 \circ G_4|}.$$

4.

$$\begin{aligned}
|G_1 \circ G_2 \circ G_3 \circ G_4| &\leq \frac{|G_1 \circ G_3 \circ G_4||G_2 \circ G_3 \circ G_4|}{|G_3 \circ G_4|} && \text{por (3.50)} \\
&\leq \frac{|G_1||G_1 \cap G_3 \cap G_4||G_2||G_3 \circ G_4||G_2 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4||G_2 \cap G_3||G_2 \cap G_4|} && \text{por (3.50)} \\
&= \frac{|G_1||G_2||G_1 \cap G_3 \cap G_4||G_2 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4||G_2 \cap G_3||G_2 \cap G_4|} && \text{por (3.31).}
\end{aligned}$$

5. Como $G_1 \circ G_2 \subseteq G_1 \circ G_2 \circ G_3 \circ G_4$, entonces

$$|G_1 \circ G_2| \leq |G_1 \circ G_2 \circ G_3 \circ G_4|$$

luego,

$$\begin{aligned}
\frac{|G_1||G_2|}{|G_1 \cap G_2|} &\leq |G_1 \circ G_2 \circ G_3 \circ G_4| && \text{por (3.31)} \\
&\leq \frac{|G_1||G_2||G_3||G_4||G_1 \cap G_3 \cap G_4||G_2 \cap G_3 \cap G_4|}{|G_1 \cap G_3||G_1 \cap G_4||G_2 \cap G_3||G_2 \cap G_4||G_3 \cap G_4|} && \text{por (3.51).}
\end{aligned}$$

De aquí se sigue (3.52).

6. Se tiene directamente de (3.52).

7. La desigualdad (3.53) no siempre se cumple, ya que si las variable aleatorias X_1, X_2, X_3, X_4 cumplen las siguientes propiedades:

   - X_1 independiente de (X_2, X_3)
   - X_2 independiente de (X_3, X_4)
   - $X_1 = X_4$

   la desigualdad no es válida, sin embargo se puede probar que (3.53) es una desigualdad de la información Shannon restringida, bajo las restricciones $I(X_1; X_2) = I(X_1; X_2|X_3) = 0$

□

4

Conclusiones

Gracias a la relación mostrada, se podría pensar en encontrar desigualdades de la información tipo no-Shannon mediante métodos de la Teoría de Grupos.

Las desigualdades de la información tipo Shannon tienen una prueba grupo-teórica.

La clausura del espacio de las funciones de entropía para $n \geq 4$ no es poliédrico, es decir, este no se puede expresar como la intersección de un número finito de semiespacios cerrados.

5

Trabajo futuro

Encontrarle aplicación en la Teoría de Grupos a las desigualdades mostradas en este trabajo.

Encontrar alguna desigualdad no-Shannon utilizando la Teoría de Grupos y la relación que se mostró entre la Teoría de la Información y la Teoría de Grupos.

6

Bibliografía

[1] John B. Fraleigh. *Algebra abstracta. Primer curso.* 1982.

[2] F. Matus. Infinitely many information inequalities. *IEEE International Symposium on Information Theory, Nice, France*, pages 2101–2105, 2007.

[3] C. Freiling y K. Zeger R. Dougherty. Six new non-shannon information inequatilies. *IEEE International Symposium on Information Theory, Seattle, Washington*, pages 233–236, 2006.

[4] C. Freiling y K. Zeger R. Dougherty. Non-shannon information inequatilies in four ramdom variables. 2011.

[5] S. Diggavi R. Pulikkoonattu, E. Perron. *Xitip (Information theoretic inequality prover)* `http://xitip.epfl.ch/`.

[6] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, Vol. 27*, pages 379–423, 1948.

[7] T. H. Chan y R. W. Yeung. On a relation between information inequalities and group theory. *IEEE Transactions on information theory. Vol. 48. No. 7*, pages 1992–1995, 2002.

[8] Z. Zhang y R. W. Yeung. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory, Vol.44*, pages 1440–1452, 1998.

[9] Raymond W. Yeung. *A First Course in Information Theory.* Springer, 2002.

[10] Raymond W. Yeung. *Information Theory and Network Coding.* Springer, 2008.

[11] Humberto Sarria Zapata. *An introduction to Network Coding.* 2009.