"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

Departamento de Matemáticas

Programa de Licenciatura en Matemáticas

Desigualdad de Jensen y sus aplicaciones

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

José Ítalo Sánchez Bermúdez

Director de tesis: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora, México, Septiembre 2023

SINODALES

Dra. Carmen Geraldi Higuera Chan

Dra. Alejandra Fonseca Morales

Dr. Yofre Hernan García Gómez

Dr. Jesús Adolfo Minjárez Sosa

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo recibido para la realización de este trabajo por medio del proyecto Ciencia Frontera 87787 "Juegos estocásticos de grandes poblaciones"

A mi madre, mi padre, mis hermanos y mi abuela...

Agradecimientos

Agradezco a mis padres, Mayra e Ítalo, que todo lo que he logrado es gracias a ustedes y para ustedes.

A mi padre le agradezco mostrarme el mundo de las matemáticas y ser el mejor maestro que he tenido que tendré. A mi madre el apoyo incondicional, las palabras y el sacrificio que me ha llevado a estar donde estoy. A mi abuela, Flora, por ser una madre para mí, por ser la persona que me ha cuidado toda mi vida, cada paso que daré será gracias a usted. A mis hermanos, Braulio y Bella, por apoyarme todo el tiempo y por estar siempre, no lo podría haber logrado sin ustedes. A mi abuela, Nelly, por los cuidados y atenciones de los cuales siempre estaré agradecido. A mi tía Marycruz, por todos esos apoyos antes y durante mis estudios.

Les agradezco a mis compañeros de licenciatura por ser grandes personas, excelentes amigos y grandes matemáticos. En especial a Rafael, Valeria, Abigail, Keren, Samai, Abraham, que hicieron de estos últimos semestres más disfrutables y divertidos.

También agradecer a todos los profesores del Departamento de Matemáticas por su trabajo en la formación de matemáticos. Especialmente a los profesores Adrián Acuña, Carmen Higuera, Genaro Hernández, Carolina Villalba, quienes fueron muy importantes durante mi formación con sus clases y constantes apoyos.

Por último, agradecer a mi director de tesis, Adolfo Minjárez, por guiarme durante la realización de este trabajo, por los nuevos conocimientos y herramientas que encontré en el camino, además de agradecer el tiempo y disposición mostrada para completar esta tesis. De igual forma, agradecer a mis sinodales Carmen Higuera, Alejandra Fonseca, Yofre García, por el tiempo empleado para la revisión de este trabajo y por los consejos durante la misma.

Índice general

Agradecimientos III

Índice general III

Introducción 1

Preliminares 4

1. La desigualdad de Jensen y su relación con otras desigualdades 11
- 1.1. Desigualdad de Jensen . . . 11
  - 1.1.1. Historia . . . 11
  - 1.1.2. Desigualdad y demostración . . . 12
  - 1.1.3. Ejemplos . . . 14
  - 1.1.4. Variantes . . . 15
- 1.2. Otras desigualdades . . . 16

2. Cotas para la desigualdad de Jensen 24
- 2.1. Introducción . . . 24
- 2.2. Cotas globales para la desigualdad de Jensen . . . 25
- 2.3. La mejor cota global posible . . . 30

3. Aplicaciones de la Desigualdad de Jensen 37
- 3.1. Introducción . . . 37
- 3.2. Desigualdades Clásicas . . . 37
- 3.3. Teoría de la Información . . . 44
- 3.4. Geometría . . . 48
- 3.5. Probabilidad . . . 50
- 3.6. Teoría de números . . . 51

Introducción

La desigualdad de Jensen ha sido uno de los resultados más importantes en el desarrollo de las matemáticas. Propuesta por el matemático danés Johan Jensen a principios del siglo XX en el artículo [8], esta desigualdad nos brinda una relación entre los valores de una función convexa y sus respectivos promedios ponderados. Su forma más simple la podemos establecer de la siguiente manera:

Sea $f : I \to \mathbb{R}$ una función convexa definida en un intervalo I. Entonces para cualquier $n \in \mathbb{N}$, $x_1, x_2, .., x_n \in I$ y $p_1, p_2, ..., p_n \geq 0$ tal que $\sum_{i=1}^{n} p_i = 1$, tenemos

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i).$$

Aun cuando su enunciado es sencillo y fácil de entender, sus aplicaciones han logrado resolver problemas importantes y trascendentes en áreas tan diversas de las matemáticas como Optimización, Probabilidad, Estadística, Procesos Estocásticos, Teoría de la Información, Análisis Matemáticos, Geometría, entre otras. Con su aplicación se ha logrado analizar de forma más sencilla importantes problemas añejos, como por ejemplo el obtener propiedades del *centroide*, concepto introducido por Arquímedes y después estudiado por Gauss. Además la Desigualdad de Jensen fue incluida en el libro *Inequalities* [6] de Hardy, Littlewood y Polya.

El objetivo de este trabajo es analizar la desigualdad de Jensen desde varios aspectos. En particular presentamos algunas de sus variantes en diferentes ramas de las matemáticas, así como generalizaciones al considerar que algunos pesos pueden ser negativos. Además, estudiamos condiciones para que se cumpla la desigualdad inversa aun para funciones convexas. Por otro lado, planteamos y resolvemos el problema de encontrar cotas para la desigualdad de Jensen por medio del llamado operador de Jensen. Específicamente, sean

$\mathbf{p} = \{p_i\}_{i=1}^n$ y $\mathbf{x} = \{x_i\}_{i=1}^n$. Definimos el operador de Jensen como

$$J_f(\mathbf{x}, \mathbf{p}) := \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right).$$

En este sentido, en términos del operador J_f, la desigualdad de Jensen se puede expresar como $0 \leq J_f(\mathbf{x}, \mathbf{p})$. Entonces el objetivo es acotar dicho operador.

El trabajo está estructurado de las siguiente manera.

En el primer capítulo nos centramos en la presentación y demostración de la desigualdad tomando como referencia el artículo original de Jensen [8], y basándonos también en el libro [2]. Además introducimos otras condiciones que garantizan que se sigue cumpliendo la desigualdad con las cuales obtenemos algunas generalizaciones. Esta parte está basada principalmente en los resultados de Steffensen en [20] y de otros resultados obtenidos de [9],[12],[13],[14].

En el segundo capítulo estudiamos el problema de las cotas para la desigualdad de Jensen vía el operador de Jensen. En particular introducimos condiciones para obtener cotas globales en el sentido que solo dependen de la función a considerar y su dominio; es decir dichas cotas serán independientes de los valores a considerar y de sus respectivos pesos. De igual forma daremos ciertos resultados que nos muestren las relaciones entre las cotas obtenidas, lo cual nos permitirá obtener la mejor cota global posible. Estos resultados están basados en los artículos [1],[4],[5],[7],[17],[18],[19].

Por último, en el capítulo final mostramos algunas de las aplicaciones de la teoría desarrollada en los capítulos anteriores a otras áreas de las matemáticas. Primeramente estableceremos algunas relaciones entre los diferentes promedios usuales como las medias aritmética, geométrica, logarítmica, armónica, entre otras. Muchas de estas relaciones vienen dadas por desigualdades bien conocidas en la literatura. Sin embargo, haciendo uso de los resultados del Capítulo 2 podremos establecer condiciones para que se cumplan las desigualdades inversas. Otra de las aplicaciones importantes que presentaremos es a la *Teoría de Información*. Es importante señalar que aun cuando esta teoría data del año 1948 cuando Claude Shannon publica el artículo [15], con las cotas obtenidas en el Capítulo 2 se obtienen nuevos resultados que sirven para estimar la *entropía*.
Concluimos el capítulo dando algunas aplicaciones de la Desigualdad de Jensen a otras áreas de las matemáticas como la Geometría, la Probabilidad y a la Teoría de Números.

Para finalizar, comentamos que el desarrollo del trabajo está basado principalmente en las referencias antes citadas. El resto de la bibliografía [3],[10],[16],[21],[22], sirvió como apoyo para la mejor exposición de los resultados.

Preliminares

En este capítulo presentaremos algunos de los resultados que nos serán de gran utilidad durante todo el trabajo.

Denotamos una sucesión finita de elementos $\{x_i\}$ en un conjunto arbitrario D como

$$\mathbf{x} := \{x_i\}_{i=1}^n,$$

y $\mathbf{p} := \{p_i\}_{i=1}^n$ una sucesión finita de pesos p_i asociada a $\mathbf{x}$, es decir, p_i será el peso asociado a $x_i, i = 1, 2, ..., n$.

Además, denotamos $P_n = \sum_{i=1}^{n} p_i$, $\overline{P_k} = P_n - P_k = \sum_{i=k+1}^{n} p_i$

Definición 0.0.1. La sucesión $\mathbf{x}$ es *constante* si $x_1 = x_2 = \cdots = x_n$.

Definición 0.0.2. La sucesión $\mathbf{x}$ es *creciente* si $x_i \leq x_{i+1}$, $\forall i = 1, ..., n$.

Definición 0.0.3. La sucesión $\mathbf{x}$ es *decreciente* si $x_{i+1} \leq x_i$, $\forall i = 1, ..., n$.

Definición 0.0.4. La sucesión $\mathbf{x}$ es *mónotona* si es *creciente* o *decreciente*.

Definición 0.0.5. Sea $\mathbf{x}$ una sucesión finita y $\mathbf{p}$ la sucesión de pesos asociada a $\mathbf{x}$, diremos que $\mathbf{x}$ es *esencialmente constante* si $x_i = x$ para algún x y $i = 1, 2, ..., n$; tal que $p_i \neq 0$. Es decir, todos los elementos de $\mathbf{x}$ que tienen pesos distintos de cero son iguales: si $p_i, p_j \neq 0$ entonces $x_i = x_j$.

Esta definición lo que nos dice es que no es necesario que todos los términos de la sucesión $\mathbf{x}$ sean iguales, solo aquellos que tengan pesos asociados distintos de cero.

Definición 0.0.6. Decimos que f es *convexa* en el intervalo $I \subseteq \mathbb{R}$ si para cada $x, y \in I$ y $\lambda \in [0, 1]$, se cumple que

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Definición 0.0.7. Decimos que f es *cóncava* cuando $-f$ es *convexa*.

Proposición 0.0.1. La definición 0.0.6 es equivalente al siguiente hecho:
si $x_1, x_2, x_3 \in I$ con $x_1 < x_2 < x_3$, entonces

1. $$f(x_2) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3).$$

2. $$\frac{f(x_1)}{(x_1 - x_2)(x_1 - x_3)} + \frac{f(x_2)}{(x_2 - x_3)(x_2 - x_1)} + \frac{f(x_3)}{(x_3 - x_1)(x_3 - x_2)} \geq 0.$$

3. $$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}.$$

Demostración.
(1) Sea $\lambda = \dfrac{x_3 - x_2}{x_3 - x_1}$ y $1 - \lambda = \dfrac{x_2 - x_1}{x_3 - x_1}$, ahora por la Definición 0.0.6 para tenemos que

$$f\left(\left(\frac{x_3 - x_2}{x_3 - x_1}\right) x_1 + \left(\frac{x_2 - x_1}{x_3 - x_1}\right) x_3\right) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3),$$

$$f\left(\frac{x_1 x_3 - x_1 x_2 + x_3 x_2 - x_1 x_3}{x_3 - x_1}\right) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3),$$

$$f\left(\frac{x_3 x_2 - x_1 x_2}{x_3 - x_1}\right) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3),$$

$$f\left(\frac{x_2(x_3 - x_1)}{x_3 - x_1}\right) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3),$$

$$f(x_2) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3).$$

(2) Sea otra vez $\lambda = \dfrac{x_3 - x_2}{x_3 - x_1}$ y $1 - \lambda = \dfrac{x_2 - x_1}{x_3 - x_1}$, por (1) tenemos que

$$f(x_2) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3),$$

$$(x_3 - x_1) f(x_2) \leq (x_3 - x_2) f(x_1) + (x_2 - x_1) f(x_3),$$

$$0 \leq -(x_3 - x_1)f(x_2) + (x_3 - x_2)f(x_1) + (x_2 - x_1)f(x_3),$$

$$0 \leq (x_3 - x_2)f(x_1) + (x_1 - x_3)f(x_2) + (x_2 - x_1)f(x_3),$$

$$0 \leq f(x_1) + \frac{(x_1 - x_3)}{x_3 - x_2}f(x_2) + \frac{(x_2 - x_1)}{x_3 - x_2}f(x_3),$$

$$0 \leq \frac{f(x_1)}{x_3 - x_1} - \frac{f(x_2)}{x_3 - x_2} + \frac{(x_2 - x_1)}{(x_3 - x_2)(x_3 - x_1)}f(x_3),$$

$$0 \leq \frac{f(x_1)}{(x_3 - x_1)(x_2 - x_1)} - \frac{f(x_2)}{(x_3 - x_2)(x_2 - x_1)} + \frac{f(x_3)}{(x_3 - x_2)(x_3 - x_1)},$$

$$0 \leq \frac{f(x_1)}{(x_1 - x_2)(x_1 - x_3)} + \frac{f(x_2)}{(x_2 - x_3)(x_2 - x_1)} + \frac{f(x_3)}{(x_3 - x_1)(x_3 - x_2)}.$$

(3) Por la convexidad de f tenemos que

$$f(x_2) \leq (1 - \lambda)f(x_1) + \lambda f(x_3) \quad \text{para} \quad \lambda \in [0, 1].$$

De aquí obtenemos

$$f(x_2) - f(x_1) \leq \lambda\,(f(x_3) - f(x_1)) \quad \text{y} \quad f(x_3) - f(x_2) \geq (1 - \lambda)\,(f(x_3) - f(x_1))\,.$$

Sustituyendo $1 - \lambda = \dfrac{x_3 - x_2}{x_3 - x_1}$ y $\lambda = \dfrac{x_2 - x_1}{x_3 - x_1}$ entonces uniendo ambas desigualdades

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \quad \text{y} \quad \frac{f(x_3) - f(x_2)}{x_3 - x_2} \geq \frac{f(x_3) - f(x_1)}{x_3 - x_1}.$$

Uniendo ambas desigualdades

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}. \quad \blacksquare$$

Proposición 0.0.2. Sea I un intervalo en $\mathbb{R}$, $f : I \to \mathbb{R}$ una función convexa. Para cada $a \in I$ consideramos la función $f_a : I - \{a\} \to \mathbb{R}$,

$$f_a(x) = \frac{f(x) - f(a)}{x - a}, \quad \forall x \in I - \{a\}.$$

Entonces f_a es creciente $\forall a \in I$.

Demostración. Sean $a \in I,\ x, y \in I - \{a\}$ con $x < y$. Consideremos tres casos.
Si $a < x < y$ por la primera desigualdad de la Proposición 0.0.1 (3) con $x_1 = a, x_2 = x, x_3 = y$ tenemos que:

$$f_a(x) = \frac{f(x) - f(a)}{x - a} \leq \frac{f(y) - f(a)}{y - a} = f_a(y),$$

de esta manera $f_a(x) \leq f_a(y)$.
Si $x < y < a$ por la segunda desigualdad de la Proposición 0.0.1 (3) con $x_1 = x, x_2 = y, x_3 = a$ tenemos que:

$$f_a(x) = \frac{f(x) - f(a)}{x - a} = \frac{f(a) - f(x)}{a - x} \leq \frac{f(a) - f(y)}{a - y} = \frac{f(y) - f(a)}{y - a} = f_a(y),$$

por tanto $f_a(x) \leq f_a(y)$.
Si $x < a < y$ por la Proposición 0.0.1 (3) con $x_1 = x, x_2 = a, x_3 = y$ tenemos que:

$$f_a(x) = \frac{f(a) - f(x)}{a - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(a) - f(y)}{a - y} = f_a(y). \quad \blacksquare$$

Observación 0.0.1. Podemos hacer un cambio de notación para las funciones de la proposición anterior

$$f_a(c) = \frac{f(c) - f(a)}{c - a} = \frac{f(a) - f(c)}{a - c} = f_c(a).$$

Otras caracterizaciones de las funciones convexas son las siguientes:

Teorema 0.0.1. Sea I un intervalo, $f : I \to \mathbb{R}$ cuando la derivada f' existe. Las siguientes afirmaciones son equivalentes:

(I). f es convexa.

(II). f' es creciente.

(III). Para cualesquiera $a, x \in I$ se tiene que $f(x) \geq f(a) + f'(a)(x - a)$.

Demostración.
(I) $\Longrightarrow$ (II). Sean $a, b \in I$ con $a < b, c \in [a, b]$. Para $x, y \in [a, b], a < x < c < y < b$. Como

f_a y f_b son crecientes (por la Proposición 0.0.2) y donde $f_c(a) \leq f'(c) \leq f_c(b)$ pues son las derivadas laterales de f. Así que tenemos que

$$f_a(x) \leq f_a(c) = f_c(a) \leq f'(c) \leq f_c(b) = f_b(c) \leq f_b(y),$$

de donde deducimos que

$$f'(a) = \lim_{x \to a} f_a(x) \leq f'(c) \leq \lim_{y \to b} f_b(y) = f'(b).$$

Por lo tanto f' es creciente.

(II) $\Longrightarrow$ (III). Sean $a, x \in I$ con $a < x$. Por el Teorema del Valor Medio existe $c \in [a, x]$ tal que

$$f(x) = f(a) + f'(c)(x - a).$$

Por hipótesis f' es creciente, por esto tenemos que $f'(c) \geq f'(a)$, entonces

$$f(x) = f(a) + f'(c)(x - a) \geq f(a) + f'(a)(x - a),$$

y así obtenemos (III).

(III) $\Longrightarrow$ (I). Dados $x, y \in I$ fijos con $x < y$ y $\lambda \in [0, 1]$, escribimos $a = (1 - \lambda)x + \lambda y$. Es claro que $f(a) \leq f(x)(1 - \lambda) + \lambda f(y)$ cuando $\lambda = 0$ y $\lambda = 1$.
Ahora consideremos

$$\lambda = \frac{a - x}{y - x} \quad , \quad 1 - \lambda = \frac{y - a}{y - x}$$

por (III) $f(x) \geq f(a) + f'(a)(x - a)$ y $f(y) \geq f(a) + f'(a)(y - a)$ como $y - a > 0$ y $x - a < 0$ entonces

$$\frac{f(x) - f(a)}{x - a} \leq f'(a) \quad , \quad \frac{f(y) - f(a)}{y - a} \geq f'(a)$$

uniendo ambas desigualdades

$$\frac{f(x) - f(a)}{x - a} \leq f'(a) \leq \frac{f(y) - f(a)}{y - a} \Longrightarrow \frac{f(x) - f(a)}{x - a} \leq \frac{f(y) - f(a)}{y - a} \Longrightarrow$$
$$\frac{y - a}{x - a}(f(x) - f(a)) \leq f(y) - f(a),$$

$$\frac{y - a}{x - a} f(x) - \frac{y - a}{x - a} f(a) \leq f(y) - f(a),$$

$$\frac{y-a}{x-a}f(x) - \frac{y-a}{x-a}f(a) \leq f(y) - f(a),$$

$$f(a) - \frac{y-a}{x-a}f(a) \leq f(y) - \frac{y-a}{x-a}f(x),$$

$$\left(1 - \frac{y-a}{x-a}\right) f(a) \leq f(y) - \frac{y-a}{x-a}f(x),$$

$$\frac{x-a-y+a}{x-a}f(a) \leq f(y) - \frac{y-a}{x-a}f(x),$$

$$\frac{x-y}{x-a}f(a) \leq f(y) - \frac{y-a}{x-a}f(x),$$

$$f(a) \leq \frac{x-a}{x-y}f(y) - \frac{y-a}{x-y}f(x) = \lambda f(y) + (1-\lambda)f(x).$$

Por lo tanto f es convexa. ■

Como consecuencia del Teorema anterior obtenemos el siguiente Corolario

Corolario 0.0.1. Sea I un intervalo, $f : I \to \mathbb{R}$ con segunda derivada f''. Las siguientes afirmaciones son equivalentes:

(I). f es convexa.

(II). $f''(x) \geq 0$ para todo $x \in I$.

Demostración.
(I) $\Longrightarrow$ (II). Por el Teorema anterior tenemos que si f es convexa entonces f' es creciente y como f' es creciente entonces $f''(x) \geq 0, \ \forall x \in I$.
(II) $\Longrightarrow$ (I). Si $f''(x) \geq 0$ entonces f' es creciente y por el Teorema anterior, si f' es creciente entonces f es convexa.

Corolario 0.0.2. f es convexa si y solo si αf también es convexa, para todo $\alpha > 0$.

Demostración. Sea $f : I \to \mathbb{R}$ y $x, y \in I, \lambda \in [0,1]$.
($\Longrightarrow$) Llamamos $g(x) = \alpha f(x), \alpha \in \mathbb{R}^+$. Si f es convexa entonces

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y),$$

así multiplicando por α

$$\alpha f(\lambda x + (1-\lambda)y) \leq \alpha(\lambda f(x) + (1-\lambda)f(y)) = \alpha\lambda f(x) + \alpha(1-\lambda)f(y)$$

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$

por tanto g es convexa.
($\Longleftarrow$) Si αf es convexa entonces $\alpha f(\lambda x+(1-\lambda)y) \leq \lambda\alpha f(x)+(1-\lambda)\alpha f(y)$ multiplicando por $\frac{1}{\alpha}$ tenemos que $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ por tanto f es convexa.

Definición 0.0.8. El *centroide* de los puntos $x_1, x_2, ..., x_n$ se define como

$$\tilde{x} = \frac{x_1 + x_2 + ... + x_n}{n}.$$

El *centroide con pesos asociados*, tales que $p_i \geq 0,\ 1 \leq i \leq n$ es:

$$\overline{x} = \frac{p_1x_1 + ... + p_nx_n}{P_n}.$$

El *centroide de la función con pesos asociados* es:

$$\overline{f} = \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i).$$

Definición 0.0.9. Una función es de *rápido crecimiento* si

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = \begin{cases} 0 & \text{si} \quad 0 < \lambda < 1 \\ 1 & \text{si} \quad \lambda = 1 \\ \infty & \text{si} \quad \lambda > 1 \end{cases}.$$

Capítulo 1

La desigualdad de Jensen y su relación con otras desigualdades

1.1. Desigualdad de Jensen

1.1.1. Historia

Johan Ludwig William Valdemar Jensen fue un matemático nacido el 8 de mayo de 1859 en Nakskov, Dinamarca. Durante su etapa como estudiante Jensen descubriría una afición a las matemáticas que mantendría el resto de su vida. Sin embargo a pesar de su gusto por estas no se dedicó profesionalmente a ellas, sino que ingresó a trabajar en la División de Copenhague de la International Bell Telephone Company donde destacaría y ocuparía un puesto importante que mantendría por el resto de su vida.

Aunque las matemáticas no eran más que una afición para Jensen, lograría dos resultados importantes dentro de ellas. El primero fue en el Análisis Complejo con *"La fórmula de Jensen"* y el segundo, y motivación para esta tesis, fue un artículo publicado en 1906 en la revista *Acta Mathematica* del matemático Mittag-Leffler, *"Sur les fonctions convexes et les inègalitès entre les valeurs moyennes"* (Sobre las funciones convexas y las desigualdades entre valores promedio)[8], artículo en el cual presenta la *Desigualdad de Jensen* (J).

Jensen fallecería el 5 de marzo de 1925 en Copenhague.

Johan Ludwig William Valdemar Jensen (1859 - 1925)

1.1.2. Desigualdad y demostración

Jensen presenta su desigualdad en la página 6 de *"Sur les fonctions convexes et les inègalitès entre les valeurs moyennes"*[8], que enunciamos de la siguiente manera:

Teorema 1.1.1 (Desigualdad de Jensen)**.** Sea f una función convexa definida en un dominio $D_f \subseteq \mathbb{R}$, $\mathbf{x} := \{x_i\}_{i=1}^n$ una sucesión finita en D_f y $\mathbf{p}:= \{p_i\}$ una sucesión positiva de pesos asociada a $\mathbf{x}$, entonces

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i), \tag{J}$$

donde $P_n := \sum_{i=1}^{n} p_i$.

Demostración. Para esta demostración procedemos por inducción.

Primero comprobamos que la desigualdad se cumple para $n = 2$, es decir,

$$f\left(\frac{1}{P_2}\sum_{i=1}^{2} p_i x_i\right) = f\left(\frac{1}{P_2}p_1x_1 + \frac{1}{P_2}p_2x_2\right).$$

Sean $q_1 = \dfrac{p_1}{P_2}$ y $q_2 = \dfrac{p_2}{P_2}$. De esta forma $q_1 + q_2 = 1$, y por convexidad de f tenemos

$$f\left(\frac{1}{P_2}p_1x_1 + \frac{1}{P_2}p_2x_2\right) = f(q_1x_1 + q_2x_2) \leq q_1f(x_1) + q_2f(x_2) = \frac{p_1}{P_2}f(x_1) + \frac{p_2}{P_2}f(x_2)$$
$$= \frac{1}{P_2}\sum_{i=1}^{2} p_i f(x_i).$$

Ahora suponemos que la desigualdad se cumple para $2 \leq i \leq n-1$ y demostraremos que también se cumple para n.

Observemos que se cumple

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) = f\left(\frac{1}{P_n}p_nx_n + \frac{1}{P_n}\sum_{i=1}^{n-1} p_i x_i\right) = f\left(\frac{p_n}{P_n}x_n + \frac{P_{n-1}}{P_{n-1}}\frac{1}{P_n}\sum_{i=1}^{n-1} p_i x_i\right).$$

Sea $q_1 = \dfrac{p_n}{P_n}$ y $q_2 = \dfrac{P_{n-1}}{P_n}$. De esta manera $q_1 + q_2 = 1$, y por el caso cuando $n = 2$, tenemos que

$$f\left(\frac{p_n}{P_n}x_n + \frac{P_{n-1}}{P_nP_{n-1}}\sum_{i=1}^{n-1} p_i x_i\right) = f\left(q_1x_n + q_2\frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i x_i\right)$$
$$\leq q_1f(x_n) + q_2f\left(\frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i x_i\right).$$

Además, por hipótesis de inducción sabemos que

$$f\left(\frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i x_i\right) \leq \frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i f\left(x_i\right),$$

por lo tanto

$$q_1 f(x_n) + q_2 f\left(\frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i x_i\right) \leq q_1 f(x_n) + q_2 \frac{1}{P_{n-1}}\sum_{i=1}^{n-1} p_i f\left(x_i\right)$$

$$= \frac{p_n}{P_n} f(x_n) + \frac{P_{n-1}}{P_n P_{n-1}}\sum_{i=1}^{n-1} p_i f\left(x_i\right) = \frac{1}{P_n}\left(p_n f(x_n) + \sum_{i=1}^{n-1} p_i f\left(x_i\right)\right) = \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i). \quad \blacksquare$$

Corolario 1.1.1. Si $P_n = 1$, entonces tenemos

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i). \tag{J_1}$$

La desigualdad (J_1) es la forma clásica o más de común de la desigualdad de Jensen.

De aquí en adelante, la desigualdad de Jensen será referenciada como (J). En este sentido, la desigualdad inversa será (¬J). Este último caso, en principio se cumple para funciones cóncavas. Sin embargo, es importante estudiar casos cuando se cumple (¬J) para funciones convexas, lo cual será analizado más adelante.

1.1.3. Ejemplos

Algunos de los ejemplos de funciones convexas para las cuales se cumple la desigualdad:

- $f(x) = x^2$

$$\sum_{i=1}^{n} \left(p_i x_i\right)^2 \leq \sum_{i=1}^{n} p_i (x_i)^2$$

- $f(x) = e^x$

$$e^{\sum_{i=1}^{n} p_i x_i} \leq \sum_{i=1}^{n} p_i e^{x_i}$$

- $f(x) = -\ln x$

$$-\ln\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i(-\ln(x_i))$$

- La igualdad se da para las ecuaciones lineales $f(x) = mx + b$

$$\sum_{i=1}^{n} p_i f(x_i) = \sum_{i=1}^{n} p_i(mx_i + b) = \sum_{i=1}^{n} p_i m x_i + p_i b = \sum_{i=1}^{n} m p_i x_i + \sum_{i=1}^{n} b p_i$$
$$= m\left(\sum_{i=1}^{n} p_i x_i\right) + b\sum_{i=1}^{n} p_i = m\left(\sum_{i=1}^{n} p_i x_i\right) + b = f\left(\sum_{i=1}^{n} p_i x_i\right)$$

1.1.4. Variantes

A continuación introducimos algunas variantes de la Desigualdad de Jensen que se presentan en diferentes contextos las cuales retomaremos con más detalle en el Capítulo 3.

Probabilidad

Sea X una variable aleatoria y f una función convexa entonces

$$f(E(X)) \leq E(f(X)),$$

donde $E(\cdot)$ representa el operador esperanza.

Teoría de la medida

Sea $(X, \mathcal{F}, \mu)$ un espacio de medida, sea g una función real $\mu-$integrable y f una función convexa en el eje real entonces

$$f\left(\int_X g \ d\mu\right) \leq \int_X f \circ g \ d\mu.$$

Teoría de la información

Haciendo uso de (J) obtenemos la *Desigualdad de Gibbs*, un resultado muy conocido en la Teoría de la Información.

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i,$$

donde $P = \{p_1, p_2, ..., p_n\}$ y $Q = \{q_1, q_2, ..., q_n\}$ son distribuciones de probabilidad discretas.

1.2. Otras desigualdades

La desigualdad de Jensen nos otorga la posibilidad de encontrar otras desigualdades mediante distintas hipótesis a las planteadas originalmente por Jensen en [8]. Algunas de estas las podemos considerar como extensión de (J).

En 1919 el matemático danés Johan Frederik Steffensen publica *"On Certain Inequalities and Methods of Approximation"* [20] en el Journal of the Institute of Actuaries. En este artículo aparece por primera vez la *Desigualdad de Jensen-Steffensen*, que es una generalización de (J) y consecuencia de un trabajo previo de Steffensen.

Teorema 1.2.1 (Desigualdad Jensen - Steffensen)**.** Si I es un intervalo de $\mathbb{R}$ donde f es convexa y $\mathbf{x} \in I$, monótona con $n > 2$, y si $\mathbf{p}$ una sucesión real que satisface

$$P_n \neq 0, \quad 0 \leq \frac{P_i}{P_n} \leq 1, \quad 1 \leq i \leq n, \tag{$*$}$$

entonces (J) se cumple.

Observación 1.2.1. La condición ($*$) es equivalente a

$$P_n \neq 0, \quad 0 \leq \frac{P_n - P_i}{P_n} \leq 1, \quad 1 \leq i \leq n,$$

y si suponemos que $P_n > 0$, entonces

$$0 \leq P_n - P_i \leq P_n, \quad 1 \leq i \leq n.$$

Demostración. Como en este caso los pesos pueden ser negativos o ceros, entonces veremos que $\overline{x} = \dfrac{1}{P_n}\displaystyle\sum_{i=1}^{n} p_i x_i$ se encuentra en el intervalo $I = [a, b]$. Como $\mathbf{x}$ es monótona, sin pérdida de generalidad suponemos que $\mathbf{x}$ es creciente.

Observemos que

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i = x_n + \sum_{i=1}^{n-1} \frac{P_i}{P_n}(x_i - x_{i+1})$$

pues

$$x_n + \sum_{i=1}^{n-1} \frac{P_i}{P_n}(x_i - x_{i+1})$$

$$= x_n + \frac{P_1}{P_n}x_1 - \frac{P_1}{P_n}x_2 + \frac{P_2}{P_n}x_2 - \frac{P_2}{P_n}x_3 + \cdots + + \frac{P_{n-1}}{P_n}x_{n-1} - \frac{P_{n-1}}{P_n}x_n$$

$$= \left(\frac{P_n}{P_n} - \frac{P_{n-1}}{P_n}\right)x_n + \left(\frac{P_{n-1}}{P_n} - \frac{P_{n-2}}{P_n}\right)x_{n-1} + \cdots + \left(\frac{P_2}{P_n} - \frac{P_1}{P_n}\right)x_2 + \frac{P_1}{P_n}x_1$$

$$= \frac{p_n}{P_n}x_n + \frac{p_{n-1}}{P_n}x_{n-1} + \cdots + \frac{p_1}{P_n}x_1 = \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i.$$

Como supusimos que $\mathbf{x}$ es creciente, entonces para toda $i = 1, ..., n-1$ sabemos que $x_i - x_{i+1} < 0$, por tanto $x_n + \sum_{i=1}^{n-1} \frac{P_i}{P_n}(x_i - x_{i+1}) < x_n$ con $\frac{P_i}{P_n} \geq 0$.

De igual forma, usando que

$$\sum_{i=1}^{n} p_i x_i = p_1x_1 + p_2x_2 + \cdots + p_nx_n$$

y $x_1 \leq x_2 \leq \cdots \leq x_n$, tenemos

$$p_1x_1 + p_2x_1 + \cdots + p_nx_1 \leq p_1x_1 + p_2x_2 + \cdots + p_nx_n,$$
$$x_1(p_1 + p_2 + \cdots + p_n) \leq \sum_{i=1}^{n} p_i x_i,$$
$$P_n x_1 \leq \sum_{i=1}^{n} p_i x_i,$$
$$x_1 \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i = \overline{x}.$$

Con lo cual $\overline{x} \in I$.

Ahora por la Teorema 0.0.1 podemos caracterizar una función convexa de la siguiente forma

$$f(x) - f(y) \geq M(x - y), \tag{1}$$

para toda $x, y \in [a, b]$. Donde $M = f'(y)$ cuando $f'(y)$ existe. Entonces, por (1) tenemos que

$$f(x) - f(z) \geq M(x - z) \quad \text{si} \quad y \leq z \leq x, \tag{2}$$

$$f(x) - f(z) \leq M(x - z) \quad \text{si} \quad z \leq x \leq y. \tag{3}$$

Llamamos $\overline{x} = \dfrac{1}{P_n}\sum_{i=1}^{n} p_i x_i \in I$ y sea m tal que $\overline{x} \in [x_{m+1}, x_m]$. Escribimos

$$f\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i f(x_i) = \sum_{i=1}^{m-1} \frac{P_i}{P_n}(M(x_i - x_{i+1}) - f(x_i) + f(x_{i+1}))$$
$$\frac{P_m}{P_n}(M(x_m - \overline{x}) - f(x_m) + f(\overline{x})) + \frac{\overline{P_{m+1}}}{P_n}(f(\overline{x}) - f(x_{m+1}) - M(\overline{x} - x_{m+1}))$$
$$+ \sum_{i=m+1}^{n-1} \frac{\overline{P_{i+1}}}{P_n}(f(x_i) - f(x_{i+1}) - M(x_i - x_{i+1}))).$$

Por (2) tenemos que

$$M(x_i - x_{i+1}) - f(x_i) + f(x_{i+1}) \leq 0, \quad \forall i = 1, 2, .., m - 1,$$

y usando (3)

$$f(x_i) - f(x_{i+1}) - M(x_i - x_{i+1}) \leq 0, \quad \forall i = m + 1, ..., n - 1.$$

Además, por hipótesis $\dfrac{P_i}{P_n} \geq 0$ y $\dfrac{\overline{P_{i+1}}}{P_n} \geq 0, \quad \forall i = 1, ..., n$, por lo que

$$f\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i f(x_i) \leq 0,$$
$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i). \quad \blacksquare$$

Vale la pena observar que la presentación y la demostración del Teorema 1.2.1 no son de la forma en que lo presenta Steffenson en su artículo. Esta demostración es más simple y fue tomada de [13].

El siguiente resultado establece condiciones bajo las cuales (¬J) se cumple.

Teorema 1.2.2. Sea $I \subseteq \mathbb{R}$ donde f es convexa. Si $n \geq 2, \mathbf{p}$ una sucesión real con $P_n > 0, p_1 > 0, p_i < 0$ para $2 \leq i \leq n$ y $\mathbf{x}$ una sucesión en I con $\overline{x} = \dfrac{1}{P_n}\sum_{i=1}^{n} p_i x_i \in I$,

entonces (¬J) se cumple.

Demostración. Sean las sucesiones $\mathbf{x}:=(\overline{x}, x_2, ..., x_n), \mathbf{p} := (p_1, p_2, ..., p_n)$ y $\mathbf{p}_0 := (P_n, -p_2, -p_3, ..., -p_n)$ donde $P_n = \sum_{i=1}^{n} p_i$. Notemos que $\sum \mathbf{p}_0 = p_1$, entonces por (J) tenemos que

$$\frac{Pn}{p_1} f(\overline{x}) - \frac{1}{p_1} p_2 f(x_2) - \cdots - \frac{1}{p_1} p_n f(x_n) \geq f\left(\frac{1}{p_1} P_n \overline{x} - \cdots - \frac{1}{p_1} p_n x_n\right),$$

Ahora

$$\frac{P_n}{p_1} f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) - \frac{p_2}{p_1} f(x_2) - \cdots - \frac{p_n}{p_1} f(x_n) \geq f\left(\frac{P_n}{p_1}\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i - \frac{p_2}{p_1} x_2 - \cdots - \frac{p_n}{p_1} x_n\right)$$
$$= f(x_1).$$

De aquí,

$$\frac{P_n}{p_1} f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \geq f(x_1) + \frac{p_2}{p_1} f(x_2) + \cdots + \frac{p_n}{p_1} f(x_n),$$

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \geq \frac{1}{P_n}\left(p_1 f(x_1) + p_2 f(x_2) + \cdots + p_n f(x_n)\right) = \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i). \quad \blacksquare$$

Otro punto que será analizado en el Capítulo 2, es el obtener cotas para la Desigualdad de Jensen por medio del operador de Jensen $J_f(\mathbf{x}\ , \mathbf{p}\ )$. Una cota que se pueda obtener fácilmente con la teoría que hemos desarrollado hasta ahora es dada en el siguiente resultado.

Teorema 1.2.3. Sea f convexa en un intervalo $I = [a, b]$ y $\mathbf{x}$ una sucesión con elementos en I y $\mathbf{p}$ una sucesión positiva. Si $\overline{x} = \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i$ entonces

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) \leq \frac{b-\overline{x}}{b-a} f(a) + \frac{\overline{x}-a}{b-a} f(b). \qquad (**)$$

Demostración. La desigualdad (∗∗) se obtiene aplicando la Proposición 0.0.1. En efecto, observe que

$$f(x_i) \leq \frac{b-x_i}{b-a}f(a) + \frac{x_i-a}{b-a}f(b),$$

lo cual implica

$$\frac{p_i}{P_n}f(x_i) \leq \frac{p_i}{P_n}\left(\frac{b-x_i}{b-a}f(a) + \frac{x_i-a}{b-a}f(b)\right) = \frac{1}{P_n}\left(\frac{b-x_i}{b-a}p_if(a) + \frac{x_i-a}{b-a}p_if(b)\right).$$

De aquí

$$\frac{1}{P_n}\left[p_1f(x_1) + p_2f(x_2) + \cdots + p_nf(x_n)\right] = \frac{1}{P_n}\sum_{i=1}^{n} p_if(x_i)$$

$$\leq \frac{1}{P_n}\left(\frac{b-x_1}{b-a}p_1f(a) + \frac{x_1-a}{b-a}p_1f(b) + \cdots + \frac{b-x_n}{b-a}p_nf(a) + \frac{x_n-a}{b-a}p_nf(b)\right)$$

$$= \frac{bp_1f(a) - p_1x_1f(a) + p_1x_1f(b) - p_1af(b) + \cdots + p_nbf(a) - p_nx_nf(a) + p_nx_nf(b) - p_naf(b)}{P_n(b-a)}$$

$$= \frac{bf(a)[p_1 + \cdots + p_n] - f(a)[p_1x_1 + \cdots + p_nx_n] + f(b)[p_1x_1 + \cdots + p_nx_n] - af(b)[p_1 + \cdots + p_n]}{P_n(b-a)}$$

$$= \frac{P_n(bf(a)) - P_n\overline{x}f(a)}{P_n(b-a)} + \frac{P_n\overline{x}f(b) - P_n(af(b))}{P_n(b-a)} = \frac{b-\overline{x}}{b-a}f(a) + \frac{\overline{x}-a}{b-a}f(b),$$

donde la última igualdad se sigue de la relación $p_1x_1 + \cdots + p_nx_n = P_n\overline{x}$. ∎

El hecho de tratar con funciones convexas nos permite dar interpretaciones geométricas y obtener resultados de la geometría. Concluimos este capítulo presentando dos resultados geométricos relacionados con el centroide los cuales son consecuencia de la desigualdad de Jensen. Para una fácil referencia (ver Definición 0.0.8) recordemos las siguientes definiciones

El *centroide* de los puntos $x_1, x_2, ..., x_n$ se define como

$$\tilde{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

El *centroide con pesos asociados*, tales que $p_i \geq 0,\ \ 1 \leq i \leq n$ es:

$$\overline{x} = \frac{p_1 x_1 + \cdots + p_n x_n}{P_n}.$$

El *centroide de la función con pesos asociados* es:

$$\overline{f} = \frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i).$$

Teorema 1.2.4. Si f es convexa en el intervalo $I = [a, b], \mathbf{x}$ una sucesión en I y $\mathbf{p}$ una sucesión positiva, entonces el centroide $G = (\overline{x}, \overline{f})$ de los puntos $(x_i, f(x_i))$ con pesos $p_i,\ \ 1 \leq i \leq n$, se encuentra sobre la gráfica de f y debajo de la cuerda AB donde $A = (a, f(a)), B = (b, f(b))$.

Demostración. G se encuentra por encima de la gráfica ya que por (J), $\overline{f} \geq f(\overline{x})$, entonces el punto $(\overline{x}, \overline{f})$ está por encima del punto $(\overline{x}, f(\overline{x}))$.

Ahora, llamemos $R(x)$ a la recta que pasa por A y B. Por la ecuación punto-pendiente

$$R(x) = \frac{f(b) - f(a)}{b - a}(x - a) + f(a) = \frac{f(b)(x - a)}{b - a} - \frac{f(a)(x - a)}{b - a} + \frac{f(a)(b - a)}{b - a}$$
$$= \frac{f(b)(x - a)}{b - a} + \frac{-xf(a) + af(a) + bf(a) - af(a)}{b - a} = \frac{x - a}{b - a} f(b) + \frac{b - x}{b - a} f(a)$$

Evaluando en $\overline{x}$ y por el Teorema 1.2.3 tenemos que

$$\overline{f} = \frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i) \leq \frac{b - \overline{x}}{b - a} f(a) + \frac{\overline{x} - a}{b - a} f(b) = R(\overline{x})$$

de esta forma $G = (\overline{x}, \overline{f})$ está por debajo del segmento AB. ∎

Teorema 1.2.5. Sea $I = [a, b]$ y f una función positiva, estrictamente convexa y dos veces diferenciable en $I, \mathbf{x}$ una sucesión con elementos en $I, n \geq 2$ y $\mathbf{p}$ una sucesión positiva. Entonces

$$\frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i) \leq \lambda f\left(\frac{1}{P_n} \sum_{i=1}^{n} p_i x_i\right)$$

donde, si

$$\phi = (f')^{-1},\ \ \mu = \frac{f(b) - f(a)}{b - a},\ \ \nu = \frac{bf(a) - af(b)}{b - a}$$

entonces λ es la única solución de la ecuación

$$f \circ \phi \left(\frac{\mu}{\lambda}\right) = \frac{\mu}{\lambda}\phi\left(\frac{\mu}{\lambda}\right) + \frac{\nu}{\lambda}. \tag{3}$$

Además

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) \leq \gamma + f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right)$$

donde $\gamma = \mu\phi(\mu) + \nu - f \circ \phi(\mu)$.

Demostración. Por el Corolario 0.0.2, sabemos que λf es convexa, con $\lambda > 0$. Elegimos un λ tal que $\lambda f(x)$ sea tangente a la cuerda AB definida como el Corolario 1.2.4, de esta forma $\lambda f(x) = \mu x + \nu$ para algún x y $\mu = \dfrac{f(b) - f(a)}{b - a}$ y $\nu = \dfrac{bf(a) - af(b)}{b - a}$. Derivando obtenemos $\lambda f'(x) = \mu$ y sustituyendo $\lambda = \dfrac{\mu x + \nu}{f(x)}$, tenemos que

$$\frac{\mu x + \nu}{f(x)} f'(x) = \mu$$
$$(\mu x + \nu) f'(x) = \mu f(x)$$

Definimos $F(x) = \mu f(x) - (\mu x + \nu) f'(x)$. Para concluir tenemos que ver que λ es efectivamente la única solución posible a la ecuación (3).

Primero veremos que F tiene una única solución en I, esto es claro pues

$$F(a) = f(a)\left(\frac{f(b) - f(a)}{b - a} - f'(a)\right) \quad \text{y} \quad F(b) = f(b)\left(\frac{f(b) - f(a)}{b - a} - f'(b)\right),$$

pero por convexidad de f, tenemos que $\dfrac{f(b) - f(a)}{b - a} - f'(a) \geq 0$ y $\dfrac{f(b) - f(a)}{b - a} - f'(b) \leq 0$.

y $f(a), f(b) \geq 0$, por lo tanto $F(a)F(b) \leq 0$, así sabemos que hay al menos una solución en $[a, b]$.

Ahora demostramos que $F(x)$ es estrictamente decreciente para que esta solución sea única. Sea

$$F'(x) = -f''(x)\left[\frac{f(b) - f(a)}{b - a}x + \frac{bf(a) - af(b)}{b - a}\right].$$

entonces $F'(a) = -f''(a)f(a) < 0$ y $F'(b) = -f''(b)f(b) < 0$, pues $f''(x) > 0$.

Por lo anterior solo hace resta probar que $\dfrac{f(b)-f(a)}{b-a}x+\dfrac{bf(a)-af(b)}{b-a}>0$.

$$\frac{f(b)-f(a)}{b-a}x+\frac{bf(a)-af(b)}{b-a}=\frac{xf(b)-xf(a)+bf(a)-af(b)}{b-a}$$

$$=\frac{f(a)(b-x)+f(b)(x-a)}{b-a}=\frac{b-x}{b-a}f(a)+\frac{x-a}{b-a}f(b)>f(x).$$

Esto último por la Proposición 0.0.1 y el hecho que f es positiva y estrictamente convexa. De esta forma, $F(x)=0,\ \forall x\in I$ y por lo tanto F es decreciente y tiene una única solución en $[a,b]$.

De la misma forma si consideramos $g(x)=\gamma+f(x)$ con γ de tal forma que $g(x)$ sea tangente a AB, tendremos que $\gamma+f(x)=\mu x+\nu$, derivando $f'(x)=\mu$, que es equivalente a $x=(f')^{-1}(\mu)$, por tanto

$$\gamma+f(x)=\mu x+\nu$$
$$\gamma=\mu x+\nu-f(x)=\mu(f')^{-1}(\mu)+\nu-f\left((f')^{-1}(\mu)\right)=\mu\phi(\mu)+\nu-f\circ\phi(\mu) \quad \blacksquare$$

Capítulo 2

Cotas para la desigualdad de Jensen

2.1. Introducción

El problema de calcular cotas en varias áreas de las matemáticas es importante. Por ejemplo, en métodos de aproximación por medio de desigualdades, el conocer tanto la cota superior como la inferior, nos permite tener una idea del error máximo y mínimo que se puede cometer.

En el caso de la Desigualdad de Jensen, el problema de las cotas se puede plantear de la siguiente manera.

Sea f una función convexa sobre un dominio $I = [a, b] \subset \mathbb{R}, \mathbf{x} = \{x_i\}$ una sucesión finita en I y $\mathbf{p} = \{p_i\}$ una sucesión de pesos asociada a $\{x_i\}$ donde $\sum_{i=1}^{n} p_i = 1$. Definimos el funcional de Jensen J_f como

$$J_f(\mathbf{x}\ , \mathbf{p}\ ) := \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right).$$

Observemos que la clásica Desigualdad de Jensen (J_1) se puede expresar como

$$0 \leq J_f(\mathbf{p},\mathbf{x}).$$

Este hecho motiva el planteamiento del problema. En efecto, observemos que el cero es una cota inferior global del operador de Jensen J_f.

Entonces, nuestro problema a estudiar en este capítulo es encontrar cotas superiores y otras cotas inferiores de J_f que sean globales, es decir, que solo dependa de la función f y el intervalo $[a, b]$.

Además, analizaremos las condiciones bajo las cuales podemos obtener las mejores cotas globales.

2.2. Cotas globales para la desigualdad de Jensen

Antes de presentar nuestros resultados, establecemos el siguiente lema.

Lema 2.2.1. Para una función convexa f en I, $x, y \in I$, y $0 < p, q$ tales que $p + q = 1$, tenemos que

$$\text{mín}(p, q)\mathcal{B}_f(x, y) \leq pf(x) + qf(y) - f(px + qy) \leq \text{máx}(p, q)\mathcal{B}_f(x, y),$$

donde $\mathcal{B}_f(x, y) := f(x) + f(y) - 2f\left(\dfrac{x+y}{2}\right)$.

Demostración. Si $p = 1$ y $q = 0$ entonces tenemos que

$$0 = 0 \cdot \mathcal{B}_f(x, y) \leq f(x) - f(x) = 0 \leq f(x) + f(y) - 2f\left(\frac{x+y}{2}\right),$$

lo cual es cierto, ya que por la convexidad de f tenemos

$$2f\left(\frac{x+y}{2}\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) \leq 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) = f(x) + f(y).$$

Análogamente para $p = 0$ y $q = 1$.

Si $p = q = \frac{1}{2}$, entonces

$$\text{mín}(p, q)\mathcal{B}_f(x, y) = \text{máx}(p, q)\mathcal{B}_f(x, y).$$

Por lo tanto, para este caso también se cumple la igualdad.

Ahora, sin pérdida de generalidad suponemos que $0 < p < q < 1$.
Para la primera parte, sean $\tilde{p} = q - p$ y $\tilde{q} = 2p$. Observemos que $\tilde{p} + \tilde{q} = 1$, $0 < \tilde{p}, \tilde{q} < 1$.

Por la convexidad de f tenemos que

$$-f(\tilde{p}x + \tilde{q}y) \geq -\tilde{p}f(x) - \tilde{q}f(y).$$

Particularmente tenemos que

$$-f\left(\tilde{p}y + \tilde{q}\left(\frac{x+y}{2}\right)\right) \geq -\tilde{p}f(y) - \tilde{q}f\left(\frac{x+y}{2}\right),$$

$$-f\left((q-p)y + 2p\left(\frac{x+y}{2}\right)\right) \geq -(q-p)f(y) - 2pf\left(\frac{x+y}{2}\right),$$

$$-f(qy - py + px + py) \geq -qf(y) + pf(y) - 2pf\left(\frac{x+y}{2}\right),$$

$$-f(px + qy) \geq pf(y) - qf(y) - 2pf\left(\frac{x+y}{2}\right),$$

$$pf(x) + qf(y) - f(px + qy) \geq pf(x) + pf(y) - 2pf\left(\frac{x+y}{2}\right)$$

$$= p\left(f(x) + f(y) - 2f\left(\frac{x+y}{2}\right)\right),$$

$$pf(x) + qf(y) - f(px + qy) \geq \text{mín}\{p, q\}\mathcal{B}(x, y).$$

Para la segunda parte de la desigualdad tomamos $\tilde{p} = \dfrac{q-p}{2q}$ y $\tilde{q} = \dfrac{1}{2q}$. De igual manera $\tilde{p} + \tilde{q} = 1,\ 0 < \tilde{p}, \tilde{q} < 1$. Puesto que, como f es convexa entonces para $x, y \in I$, tenemos que

$$\tilde{p}f(x) + \tilde{q}f(px + qy) \geq f(\tilde{p}x + \tilde{q}(px + qy)) = f\left(\frac{q-p}{2q}x + \frac{1}{2q}(px + qy)\right)$$

$$= f\left(\frac{qx - px + px + qy}{2q}\right) = f\left(\frac{x+y}{2}\right).$$

Ahora observe que se cumplen las siguientes relaciones:

$$\frac{q-p}{2q}f(x)+\frac{1}{2q}f(px+qy)\geq f\left(\frac{x+y}{2}\right),$$

$$(q-p)f(x)+f(px+qy)\geq 2qf\left(\frac{x+y}{2}\right),$$

$$qf(x)-pf(x)+f(px+qy)\geq 2qf\left(\frac{x+y}{2}\right),$$

$$qf(x)-2qf\left(\frac{x+y}{2}\right)\geq pf(x)-f(px+qy),$$

$$qf(y)+qf(x)-2qf\left(\frac{x+y}{2}\right)\geq pf(x)+qf(y)-f(px+qy).$$

Por lo tanto concluimos

$$q\left(f(x)+f(y)-2f\left(\frac{x+y}{2}\right)\right)=\text{máx}(p,q)\mathcal{B}_f(x,y)\geq pf(x)+qf(y)-f(px+qy). \quad \blacksquare$$

Teorema 2.2.1. Sea f una función convexa definida en un intervalo cerrado $I:=[a,b]$, $\mathbf{x}:=\{x_i\}$ una sucesión de n números en I, y $\mathbf{p}:=\{p_i\}, \sum_{i=1}^{n} p_i=1$ una sucesión positiva de pesos asociada a $\mathbf{x}$. Entonces tenemos que

$$J_f(\mathbf{p},\mathbf{x})\leq f(a)+f(b)-2f\left(\frac{a+b}{2}\right)=\mathcal{B}_f(a,b) \qquad (I)$$

Demostración. Como $x_i\in I$, tenemos que $x_i=\lambda_i a+(1-\lambda_i)b$ para algún $\lambda_i\in[0,1]$. De esta forma

$$\sum_{i=1}^{n}p_i f(x_i)-f\left(\sum_{i=1}^{n}p_i x_i\right)=\sum_{i=1}^{n}p_i f(\lambda_i a+(1-\lambda_i)b)-f\left(\sum_{i=1}^{n}p_i(\lambda_i a+(1-\lambda_i)b)\right)$$

$$\leq\sum_{i=1}^{n}p_i(\lambda_i f(a)+(1-\lambda_i)f(b))-f\left(a\sum_{i=1}^{n}p_i\lambda_i+b\sum_{i=1}^{n}p_i(1-\lambda_i)\right)$$

$$= f(a)\sum_{i=1}^{n} p_i\lambda_i + f(b)\left(1-\sum_{i=1}^{n} p_i\lambda_i\right) - f\left(a\sum_{i=1}^{n} p_i\lambda_i + b\left(1-\sum_{i=1}^{n} p_i\lambda_i\right)\right),$$

donde la desigualdad se obtiene de la convexidad de f.

Llamamos $p := \sum_{i=1}^{n} p_i\lambda_i$ y $q := 1 - \sum_{i=1}^{n} p_i\lambda_i$, donde $0 \leq p, q \leq 1$ y $p + q = 1$. Por tanto

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \leq pf(a) + qf(b) - f(pa + qb), \tag{2.1}$$

y por el Lema 2.2.1, finalmente llegamos a que se cumple

$$J_f(\mathbf{p},\mathbf{x}) = \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \text{máx}(p, q)\mathcal{B}_f(a, b) \leq \mathcal{B}_f(a, b). \quad \blacksquare$$

Observación 2.2.1. Asumiendo la diferenciabilidad de la función convexa f, en el trabajo [4] se muestra la desigualdad

$$J_f(\mathbf{p},\mathbf{x}) \leq \frac{1}{4}(b-a)(f'(b) - f'(a)) := \mathcal{A}_f(a, b).$$

Por lo tanto, el no asumir que f es diferenciable en el Teorema 2.2.1 hace que la desigualdad (I) sea mejor. A continuación, mostraremos un refinamiento de la desigualdad (I). Para esto introducimos la siguiente definición.

Definición 2.2.1. Sea $D_f \subset \mathbb{R}$ el dominio de una función convexa f y $I = [a, b] \subset D_f$. Definimos

$$C(f) := \sup_{\mathbf{x}\in[a,b], a,b\in D_f}\left[\frac{J_f(\mathbf{p},\mathbf{x})}{\mathcal{B}_f(a, b)}\right].$$

Teorema 2.2.2. Para, $p, q > 0, p + q = 1$, si f es una función convexa, entonces

$$C(f) = \sup_{p,a,b}\left[\frac{pf(a) + qf(b) - f(pa + qb)}{f(a) + f(b) - 2f\left(\frac{a+b}{2}\right)}\right],$$

donde $a < b$, $a, b \in D_f$.

Demostración. Por la relación (2.1) tenemos que

$$\frac{\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right)}{\mathcal{B}_f(a,b)} \leq \frac{pf(a)+qf(b)-f(pa+qb)}{\mathcal{B}_f(a,b)} \leq \sup_{p,a,b}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{\mathcal{B}_f(a,b)}\right]$$

Por tanto

$$C(f) = \sup\left[\frac{J_f(\mathbf{p},\mathbf{x})}{\mathcal{B}_f(a,b)}\right] = \sup_{p,a,b}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{\mathcal{B}_f(a,b)}\right]. \quad \blacksquare$$

Teorema 2.2.3. Para cualquier función convexa tenemos que

$$\frac{1}{2} \leq C(f) \leq 1.$$

Demostración. Por el Teorema 2.2.1 tenemos que

$$C(f) \leq \sup[\text{máx}(p,q)] = 1.$$

Por otro lado, por el Lema 2.2.1 y para $p_1, p_2 > 0, p_1 + p_2 = 1$.

$$C(f) = \sup_{\mathbf{x},\mathbf{p}}\left[\frac{J_f(\mathbf{p},\mathbf{x})}{\mathcal{B}_f(a,b)}\right] \geq \sup\left[\frac{p_1 f(a)+p_2 f(b)-f(p_1 a+p_2 b)}{\mathcal{B}_f(a,b)}\right] \geq \sup[\text{mín}(p_1,p_2)] = \frac{1}{2}. \quad \blacksquare$$

Observación 2.2.2. Es claro que

$$J_f(\mathbf{p},\mathbf{x}) \leq C(f)\mathcal{B}_f(a,b) := \mathcal{C}_f(a,b). \qquad (II)$$

Entonces, como $\frac{1}{2} \leq C(f) \leq 1$, tenemos que (II) mejora la desigualdad (I).

A continuación presentamos casos particulares de funciones convexas con la correspondiente constante $C(f)$.

Teorema 2.2.4. Sea f una función convexa. Si el dominio de f es $D_f = (a,\infty), a \in \mathbb{R}$ y si se cumple alguna de las siguientes condiciones

i) $\lim_{x\to\infty} f(x) = x_0$ con $|x_0| < \infty$.

ii) f es una función de *rápido crecimiento*,

entonces $C(f) = 1$.

Demostración.

Si se cumple (I), vemos que

$$\lim_{b\to\infty}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{\mathcal{B}_f(a,b)}\right] = \frac{pf(a)+qd-d}{f(a)+d-2d} = \frac{pf(a)+(1-p)d-d}{f(a)-d} = p$$

por tanto $C(f) = \sup(p) = 1$.

Ahora, si se cumple (II), utilizando la definición de rápido crecimiento, tenemos

$$\lim_{x\to\infty} f(x) = \infty \qquad \lim_{x\to\infty}\frac{f(tx)}{f(x)} = 0$$

para $0 < t < 1$. Aún más por lo presentado por Heiberg en [7] la relación anterior se cumple uniformemente en t. Por tanto, para a fijo

$$\lim_{b\to\infty}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{f(a)+f(b)-2f\left(\dfrac{a+b}{2}\right)}\right] = \lim_{b\to\infty}\left[\frac{p\frac{f(a)}{f(b)}+q-\frac{f(pa+qb)}{f(b)}}{\frac{f(a)}{f(b)}+1-2\frac{f\left(\frac{a+b}{2}\right)}{f(b)}}\right] = q.$$

Entonces, $C(f) = \sup(q) = 1$. ∎

Otros valores de $C(f)$ para casos específicos son los siguientes:

$$C(x^2) = \frac{1}{2}; \quad C(\sqrt{x}) = \frac{\sqrt{2}+1}{4}; \quad C(x\ln x) = (e\ln 2)^{-1}; \quad C(e^x) = 1$$

2.3. La mejor cota global posible

En esta sección estableceremos una desigualdad que mejora las anteriores. De hecho, la cota que obtenemos es la mejor cota global posible en el sentido que es menor a las cotas correspondientes a las desigualdades anteriores. Primero presentaremos la desigualdad, y después mostraremos, que en efecto, es la menor posible.

Teorema 2.3.1. Sea f una función convexa, $\mathbf{x} \in D_f$ y $\mathbf{p}$ una sucesión positiva de pesos asociada a $\mathbf{x}$, $P_n = 1$, además $p, q > 0, p + q = 1$, entonces

$$J_f(\mathbf{p},\mathbf{x}) \leq \max_p [pf(a) + qf(b) - f(pa + qb)] := \mathcal{D}_f(a, b)$$

Demostración. Como $x_i \in [a, b],\ \forall i = 1, 2, ..., n$, sea $\{\lambda_i\}, \lambda_i \in [0, 1]$ una sucesión tal que $x_i = \lambda_i a + (1 - \lambda_i)b, \quad i = 1, 2, ...$

De esta forma, como f es convexa, las siguientes relaciones se cumplen

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) = \sum_{i=1}^{n} p_i f\left(\lambda_i a + (1 - \lambda_i)b\right) - f\left(\sum_{i=1}^{n} p_i(\lambda_i a + (1 - \lambda_i)b)\right)$$

$$\leq \sum_{i=1}^{n} p_i(\lambda_i f(a) + (1 - \lambda_i)f(b)) - f\left(a\sum_{i=1}^{n} p_i\lambda_i + b\sum_{i=1}^{n} p_i(1 - \lambda_i)\right)$$

$$= f(a)\left(\sum_{i=1}^{n} p_i\lambda_i\right) + f(b)\left(1 - \sum_{i=1}^{n} p_i\lambda_i\right) - f\left(a\left(\sum_{i=1}^{n} p_i\lambda_i\right) + b\left(1 - \sum_{i=1}^{n} p_i\lambda_i\right)\right).$$

Denotamos $\sum_{i=1}^{n} p_i\lambda_i := p; \quad 1 - \sum_{i=1}^{n} p_i\lambda_i := q$, tenemos que $0 \leq p, q \leq 1; \quad p + q = 1.$

Por consecuencia

$$J_f(\mathbf{p},\mathbf{x}) = \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \leq pf(a) + qf(b) - f(pa + qb)$$
$$\leq \max_p [pf(a) + qf(b) - f(pa + qb)] = \mathcal{D}_f(a, b). \quad \blacksquare$$

Teorema 2.3.2. Sea f una función convexa y diferenciable con dominio D_f y $I := [a, b] \subset D_f$. Entonces se cumplen las siguientes desigualdades para cada $I \subset D_f$

(I) $\mathcal{D}_f(a, b) \leq \mathcal{A}_f(a, b)$,

(II) $\mathcal{D}_f(a, b) \leq \mathcal{C}_f(a, b)$.

Demostración.

(I) Como f es convexa y diferenciable, tenemos que

$$f(x) \geq f(t) + (x-t)f'(t), \forall x, t \in I$$

En particular,

$$f(pa+qb) \leq f(a) + q(b-a)f'(a) \quad ; \quad f(pa+qb) \geq f(b) + p(a-b)f'(b)$$

con $p+q=1$.
Por lo tanto

$$pf(a)+qf(b)-f(pa+qb) = pf(a)+qf(b)-f(pa+qb)+pf(pa+qb)-pf(pa+qb)$$

$$= pf(a)+qf(b)-f(pa+qb)+(1-q)f(pa+qb)-pf(pa+qb)$$

$$= pf(a)-pf(pa+qb)+qf(b)-f(pa+qb)-qf(pa+qb)+f(pa+qb)$$

$$= p(f(a)-f(pa+qb))+q(f(b)-f(pa+qb)) \leq p(q(a-b)f'(a))+q(p(b-a)f'(b))$$

$$= pq(b-a)(f'(b)-f'(a))$$

Con lo que

$$\mathcal{D}_f(a,b) = \max_p[pf(a)+qf(b)-f(pa+qb)] \leq \max_p[pq(b-a)(f'(b)-f'(a))]$$
$$= \frac{1}{4}(b-a)(f'(b)-f'(a)) = \mathcal{A}_f(a,b)$$

(II) Por el Teorema 2.2.2 y la Definición 2.2.1 sabemos que

$$\frac{pf(a)+qf(b)-f(pa+qb)}{\mathcal{B}_f(a,b)} \leq \sup_{p;a,b\in D_f}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{f(a)+f(b)-2f\left(\frac{a+b}{2}\right)}\right] = C(f).$$

Por lo tanto, concluimos

$$\mathcal{D}_f(a,b) = \max_p\left[pf(a)+qf(b)-f(pa+qb)\right] \leq C(f)\mathcal{B}_f(a,b) = \mathcal{C}_f(a,b). \blacksquare$$

Observación 2.3.1. Observemos que

(a) Para $a, b \in D_f, a \neq b$ la función $g(p) := pf(a) + (1-p)f(b) - f(pa + (1-p)b)$ es cóncava para $0 \leq p \leq 1$ con $g(0) = g(1) = 0$. Por tanto, existe un único $p_0 \geq 0$ tal que $\max_p g(p) = g(p_0) = \mathcal{D}_f(a, b)$

(b) De los resultados anteriores obtenemos la Desigualdad *pre-Grüss*

$$\sum_{i=1}^{n} p_i x_i^2 - \left(\sum_{i=1}^{n} p_i x_i \right)^2 \leq \frac{1}{4}(b-a)^2$$

En el mismo sentido de la Observación 2.3.1 (a), podemos ver que la cota $D_f(a, b)$ es alcanzable, y más aun fuerte, como lo establece el siguiente resultado.

Teorema 2.3.3. Para cualquier función convexa f arbitraria con $I = [a, b] \subseteq D_f$, existe una sucesión $\mathbf{y} \in I$ y una sucesión de pesos asociada $\mathbf{q}$ tal que

$$J_f(\mathbf{y}, \mathbf{q}) = \mathcal{D}_f(a, b)$$

Demostración. Construimos la sucesión $\mathbf{y}$ como $x_1 = a; x_2 = x_3 = \cdots = b$ y la sucesión $\mathbf{q}$ como $p_1 = p_0$. Entonces por la Observación 2.3.1 (a)

$$\begin{aligned} J_f(\mathbf{y}, \mathbf{q}) = p_1 f(x_1) + p_2 f(x_2) - f(p_1 x_1 + p_2 x_2) &= p_0 f(a) + (1-p_0) f(b) - f(p_0 a + (1-p_0) b) \\ &= g(p_0) = \mathcal{D}_f(a, b). \quad \blacksquare \end{aligned}$$

Además, podemos dar la forma explicita de $D_f(a, b)$ en el caso de funciones diferenciables.

Teorema 2.3.4. Para cualquier función f convexa y diferenciable, tenemos que

$$\mathcal{D}_f(a, b) = \frac{f(b) - f(a)}{b - a} \Theta_f(a, b) + \frac{b f(a) - a f(b)}{b - a} - f(\Theta_f(a, b)),$$

donde $\Theta_f(a, b)$ es el valor medio de Lagrange de a, b definido como

$$\Theta_f(a, b) := (f')^{-1} \left(\frac{f(b) - f(a)}{b - a} \right)$$

Demostración. Sea $\mathcal{D}_f(a, b) := \max_p [pf(a) + qf(b) - f(pa + qb)]$. Aplicaremos el criterio de la primer derivada para encontrar el máximo de una función.

Sea $g(p) = pf(a) + (1-p)f(b) - f(pa + (1-p)b)$ entonces $g'(p) = f(a) - f(b) - f'(pa + (1-p)b)(a-b)$. Igualando a 0, tenemos

$$g'(p) = f(a) - f(b) - f'(pa + (1-p)b)(a-b) = 0,$$

de donde se obtiene

$$p = \frac{(f')^{-1}\left(\dfrac{f(b)-f(a)}{b-a}\right) - b}{a-b} = \frac{\Theta_f(a,b)}{a-b} - \frac{b}{a-b} = p_0.$$

Ahora, evaluando en g obtenemos

$$\mathcal{D}_f(a,b) = g(p_0) = \frac{(\Theta_f(a,b)-b)f(a)}{a-b} + f(b) - \frac{(\Theta_f(a,b)-b)f(b)}{a-b}$$

$$-f\left(\frac{\Theta_f(a,b)-b}{a-b}a + \left(1 - \frac{\Theta_f(a,b)-b}{a-b}\right)b\right) = \frac{f(b)-f(a)}{b-a}\Theta_f(a,b) +$$

$$\frac{-bf(a) + bf(b) + af(b) - bf(b)}{a-b} - f\left(\frac{\Theta_f(a,b)a - ba}{a-b} + b - \frac{\Theta_f(a,b)b - b^2}{a-b}\right)$$

$$= \frac{f(b)-f(a)}{b-a}\Theta_f(a,b) + \frac{bf(a)-af(b)}{b-a} - f\left(\frac{\Theta_f(a,b)(a-b) + ba - ba - b^2 + b^2}{a-b}\right)$$

$$= \frac{f(b)-f(a)}{b-a}\Theta_f(a,b) + \frac{bf(a)-af(b)}{b-a} - f\left(\Theta_f(a,b)\right). \quad \blacksquare$$

Teorema 2.3.5. Si f es convexa en $I := [a,b]$, entonces

$$0 \leq \max_{1\leq i<j\leq n}\left[p_i f(x_i) + p_j f(x_j) - (p_i+p_j)f\left(\frac{p_i x_i + p_j x_j}{p_i + p_j}\right)\right] \leq J_f(\mathbf{p},\mathbf{x}) \qquad (*)$$

Demostración. Suponemos sin pérdida de generalidad $x_r < x_s$, con $x_r, x_s \in [a,b]$. Escribimos $x_s = b - \varepsilon_1, x_r = b - \varepsilon_1 - \varepsilon_2$, con $\varepsilon_i \geq 0, i = 1,2$, entonces

$$\frac{p_s(b-\varepsilon_1) + p_r(b-\varepsilon_1-\varepsilon_2)}{p_s+p_r} = b - \varepsilon_1 - \varepsilon_2\frac{p_r}{p_s+p_r} < b - \varepsilon_1 - \varepsilon_2 = x_r \leq b$$

pues $\dfrac{p_r}{p_s+p_r} < 1$. Ahora $x_s = a+\varepsilon_1+\varepsilon_2, x_r = a+\varepsilon_1$

$$\frac{p_r(a+\varepsilon_1)+p_s(a+\varepsilon_1+\varepsilon_2)}{p_r+p_s} = a+\varepsilon_1+\varepsilon_2\frac{p_s}{p_s+p_r} \geq a$$

como $\varepsilon_1 \geq 0$ y $\varepsilon_2\dfrac{p_s}{p_s+p_r} \leq 0$.

Por (J),

$$\begin{aligned} f\left(\sum_1^n p_i x_i\right) &= f\left(\sum_{i\neq r,s} p_i x_i + (p_r+p_s)\left(\frac{p_r x_r + p_s x_s}{p_r+p_s}\right)\right) \\ &\leq \sum_{i\neq r,s} p_i f(x_i) + (p_r+p_s) f\left(\frac{p_r x_r + p_s x_s}{p_r+p_s}\right) \end{aligned}$$

por eso

$$\sum_1^n p_i f(x_i) - f\left(\sum_1^n p_i x_i\right) \geq p_r f(x_r) + p_s f(x_s) - (p_r+p_s) f\left(\frac{p_r x_r + p_s x_s}{p_r+p_s}\right)$$

como x_r, x_s fueron arbitrarios, entonces completamos la demostración.

Corolario 2.3.1. Si f es convexa en $I = [a,b]$ y $\kappa := \min\limits_{1\leq i\leq n} x_i, \tau := \max\limits_{1\leq i\leq n} x_i,\ [\kappa,\tau] \subseteq I$, entonces

$$\frac{1}{n}\mathcal{B}_f(\kappa,\tau) \leq \frac{1}{n}\sum_{i=1}^n f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \leq \mathcal{B}_f(\kappa,\tau).$$

Demostración. Como consecuencia del Teorema 2.3.5 sabemos que

$$\sum_{i=1}^n p_i f(x_i) - f\left(\sum_{i=1}^n p_i x_i\right) \geq p_r f(x_r) + p_s f(x_s) - (p_r+p_s) f\left(\frac{p_r x_r + p_s x_s}{p_r+p_s}\right)$$

donde x_r y x_s son arbitrarios. Entonces sean $x_r = \kappa = a,\ \ x_s = \tau = b,\ \ p_i = \frac{1}{n}, i = 1,...,n$, así

$$\sum_{i=1}^n \frac{1}{n} f(x_i) - f\left(\sum_{i=1}^n \frac{1}{n} x_i\right) \geq \frac{1}{n} f(\kappa) + \frac{1}{n} f(\tau) - \left(\frac{1}{n}+\frac{1}{n}\right) f\left(\frac{\frac{1}{n}\kappa + \frac{1}{n}\tau}{\frac{1}{n}+\frac{1}{n}}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}x_i\right) \geq \frac{1}{n}\left[f(\kappa) + f(\tau)\right] - \frac{2}{n} f\left(\frac{\frac{1}{n}\left(\kappa+\tau\right)}{\frac{2}{n}}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}x_i\right) \geq \frac{1}{n}\left[f(\kappa) + f(\tau) - 2f\left(\frac{\kappa+\tau}{2}\right)\right]$$

y por el Teorema 2.2.1, entonces

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \leq f(\kappa) + f(\tau) - 2f\left(\frac{\kappa+\tau}{2}\right). \quad \blacksquare$$

Capítulo 3

Aplicaciones de la Desigualdad de Jensen

3.1. Introducción

En este capítulo presentaremos algunas aplicaciones de la teoría desarrollada los capítulos anteriores. Nos centraremos en dos tipos de aplicaciones, aquellas donde la Desigualdad de Jensen juega un papel importante de forma directa, y otras por medio de las cotas obtenidas en el Capítulo 2 podemos obtener otros resultados interesantes. Dentro de esto último están las aplicaciones a la Teoría de la Información al poder establecer cotas para la entropía, y también están las aplicaciones para obtener nuevas relaciones entre diferentes promedios.

3.2. Desigualdades Clásicas

En esta sección estudiamos aplicaciones a desigualdades que involucran promedios bien conocidos, enlistados a continuación. Dados $a, b \in \mathbb{R}$, definimos las siguientes medias:

Media Aritmética	Media Geométrica	Media Logarítmica
$A(a,b) := \dfrac{a+b}{2}$	$G(a,b) := \sqrt{ab}$	$L(a,b) := \dfrac{b-a}{\ln b - \ln a}$

Media Idéntica	Media Armónica	Media de Gini
$I(a,b) := \dfrac{\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}}{e}$	$H(a,b) := \dfrac{2ab}{a+b}$	$Gi(a,b) := (a^a b^b)^{\frac{1}{a+b}}$

Para comenzar introducimos un resultado bien conocido que establece las relaciones entre estas medias.

Teorema 3.2.1. Para $0 < a < b$ tenemos que

$$a < H(a,b) < G(a,b) < L(a,b) < A(a,b) < b.$$

Demostración. Primero veremos que $a < H(a,b) < G(a,b)$. Como $a < b$ entonces $a(a-b) < 0$, de aquí

$$a(a-b) < 0 \implies a^2 - ab < 0 \implies a^2 - ab + 2ab < 2ab \implies a^2 + ab < 2ab$$
$$a(a+b) < 2ab \implies a < \frac{2ab}{a+b} = H(a,b).$$

Para $H(a,b) < G(a,b)$ utilizamos que $0 < (a-b)^2$, así

$$0 < (a-b)^2 \implies 0 < a^2 - 2ab + b^2 \implies 4ab < a^2 - 2ab + b^2 + 4ab$$
$$\implies 4ab < a^2 + 2ab + b^2 \implies 4a^2b^2 < ab(a+b)^2 \implies \frac{4a^2b^2}{(a+b)^2} < ab$$
$$\implies H(a,b) = \frac{2ab}{a+b} < \sqrt{ab} = G(a,b).$$

Ahora demostraremos que $G(a,b) < A(a,b)$. A partir de aquí, obtendremos las otras relaciones.

Supongamos que $G(a,b) > A(a,b)$. Entonces

$$\sqrt{ab} > \frac{a+b}{2} \implies ab > \frac{(a+b)^2}{4} \implies 4ab > a^2 + 2ab + b^2 \implies 0 > a^2 - 2ab + b^2$$
$$\implies 0 > (a-b)^2,$$

lo que claramente es una contradicción. Por tanto $A(a,b) > G(a,b)$.

Ahora, por lo demostrado anteriormente podemos afirmar lo siguiente:

$$x^2+(a+b)x+\left(\frac{a+b}{2}\right)^2 > x^2+(a+b)x+ab > x^2+2\sqrt{ab}x+ab$$
$$\left(x+\frac{a+b}{2}\right)^2 > (x+a)(x+b) > \left(x+\sqrt{ab}\right)^2.$$

Integramos sobre el recíproco de cada término

$$\int_0^\infty \frac{dx}{\left(x+\frac{a+b}{2}\right)^2} < \int_0^\infty \frac{dx}{(x+a)(x+b)} < \int_0^\infty \frac{dx}{\left(x+\sqrt{ab}\right)^2}.$$

Resolviendo cada integral

$$\int_0^\infty \frac{dx}{\left(x+\frac{a+b}{2}\right)^2} = \frac{2}{a+b} \qquad \int_0^\infty \frac{dx}{\left(x+\sqrt{ab}\right)^2} = \frac{1}{\sqrt{ab}}.$$

Por fracciones parciales resolvemos

$$\int_0^\infty \frac{dx}{(x+a)(x+b)} = \int_0^\infty \frac{\frac{1}{b-a}}{x+a} + \int_0^\infty \frac{\frac{1}{a-b}}{x+b} = \frac{1}{b-a}\int_0^\infty \frac{dx}{x+a} - \frac{1}{b-a}\int_0^\infty \frac{dx}{x+b}$$

$$= \frac{1}{b-a}\left[\ln(x+a)-\ln(x+b)\right]_0^\infty = \frac{1}{b-a}\lim_{R\to\infty}\left[\ln(x+a)-\ln(x+b)\right]_0^R = \frac{\ln b-\ln a}{b-a}.$$

De esta forma tenemos que

$$\frac{2}{a+b} < \frac{\ln b-\ln a}{b-a} < \frac{1}{\sqrt{ab}} \implies \frac{a+b}{2} > \frac{b-a}{\ln b-\ln a} > \sqrt{ab}$$
$$\implies A(a,b) > L(a,b) > G(a,b). \quad \blacksquare$$

La desigualdad que relaciona la media aritmética con la media geométrica se le conoce como Desigualdad A-G. Esta desigualdad, en su forma general la establece el siguiente resultado.

Teorema 3.2.2 (Desigualdad A-G)**.** Para $\mathbf{x} := \{x_i\}, \mathbf{x} \in \mathbb{R}^+$ y $\mathbf{p} := \{p_i\}$ tales que $\sum_{i=1}^{n} p_i = 1$, definimos las media aritmética y geométrica generalizada como

$$A(\mathbf{x},\mathbf{p}) := \sum_{i=1}^{n} p_i x_i \quad \text{y} \quad G(\mathbf{x},\mathbf{p}) := \prod_{i=1}^{n} x_i^{p_i},$$

Entonces $G(\mathbf{x},\mathbf{p}) \leq A(\mathbf{x},\mathbf{p})$.

Demostración. Por (J) sabemos que si $f(x) = -\ln x$ entonces

$$-\ln\left(\sum_{i=1}^{n} p_i x_i\right) \leq -\sum_{i=1}^{n} p_i \ln x_i \implies \ln\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{i=1}^{n} p_i \ln x_i.$$

Por las propiedades del logaritmo sabemos que $p_i \ln x_i = \ln x_i^{p_i}$ y $\sum_{i=1}^{n} \ln x_i^{p_i} = \ln\left(\prod_{i=1}^{n} x_i^{p_i}\right)$.
De esta manera

$$\ln\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{i=1}^{n} p_i \ln x_i = \sum_{i=1}^{n} \ln x_i^{p_i} = \ln\left(\prod_{i=1}^{n} x_i^{p_i}\right).$$

Entonces

$$\ln\left(\sum_{i=1}^{n} p_i x_i\right) \geq \ln\left(\prod_{i=1}^{n} x_i^{p_i}\right).$$

Como la función $f(x) = \ln x$ es estrictamente creciente concluimos

$$\sum_{i=1}^{n} p_i x_i \geq \prod_{i=1}^{n} x_i^{p_i} \implies A(\mathbf{x},\mathbf{p}\,) \geq G(\mathbf{x},\mathbf{p}\,). \quad \blacksquare$$

El punto importante ahora, es establecer condiciones bajo las cuales podemos obtener la inversa de la desigualdad A-G. Es decir, encontrar una constante Υ_1 tal que se cumpla

$$A(\mathbf{x}\,,\mathbf{p}\,) \leq \Upsilon_1 G(\mathbf{x}\,,\mathbf{p}\,).$$

La respuesta a este punto la proporciona el siguiente teorema.

Teorema 3.2.3. Para $\mathbf{x} := \{x_i\}, \mathbf{x} \in [a,b]$ con $0 < a < b$ y $\mathbf{p} := \{p_i\}, \sum_{i=1}^{n} p_i = 1$ entonces

$$1 \leq \frac{A(\mathbf{x},\mathbf{p})}{G(\mathbf{x},\mathbf{p})} \leq \frac{L(a,b)I(a,b)}{G^2(a,b)} := \Upsilon_1(a,b).$$

Demostración. Aplicando el Teorema 2.3.1, a la función $f(x) = -\ln x$, obtenemos que

$$0 \leq \sum_{i=1}^{n} p_i(-\ln x_i) + \ln\left(\sum_{i=1}^{n} p_i x_i\right) = -\ln\left(\prod_{i=1}^{n} x_i^{p_i}\right) + \ln\left(\sum_{i=1}^{n} p_i x_i\right) = \ln\left(\frac{\sum_{i=1}^{n} p_i x_i}{\prod_{i=1}^{n} x_i^{p_i}}\right)$$

$$= \ln \frac{A(\mathbf{x},\mathbf{p})}{G(\mathbf{x},\mathbf{p})} \leq \max_p[-p\ln a - q\ln b + \ln(pa+qb)].$$

Con el criterio de la primer derivada y sabiendo que $q = 1-p$ encontramos el máximo de la expresión, de la siguiente manera:

$$-\ln a + \ln b + \frac{a-b}{pa+(1-p)b} = 0$$

$$(pa+(1-p)b)(\ln a - \ln b) = a-b$$

$$pa + b - pb = \frac{a-b}{\ln a - \ln b}$$

$$p(a-b) = \frac{a-b}{\ln a - \ln b} - b$$

$$p = \frac{1}{\ln a - \ln b} - \frac{b}{a-b} = \frac{b}{b-a} - \frac{1}{\ln b - \ln a} = \frac{b-L(a,b)}{b-a}.$$

Entonces el valor máximo, es:

$$\max_p[-p\ln a - q\ln b + \ln(pa+qb)] = \max_p\left[p\ln\left(\frac{b}{a}\right) + \ln(p(a-b)+b) - \ln b\right]$$

$$= \ln\left(\frac{b-L(a,b)}{b-a}(a-b)+b\right) + \frac{b-L(a,b)}{b-a}\ln\left(\frac{b}{a}\right) - \ln b$$

$$= \ln(L(a,b)) + \frac{b - \dfrac{b-a}{\ln b - \ln a}}{b-a}\ln\left(\frac{b}{a}\right) - \ln b$$

$$= \ln(L(a,b)) + \frac{b\ln b - b\ln a - b + a - b\ln b + a\ln b}{b-a}$$

$$= \ln(L(a,b)) + \frac{a\ln b - b\ln a}{b-a} - \frac{b-a}{b-a}$$

$$= \ln(L(a,b)) + \frac{a\ln b - b\ln a + a\ln a - a\ln a + b\ln b - b\ln b}{b-a} - 1$$

$$= \ln(L(a,b)) + \frac{b\ln b - a\ln a}{b-a} + \frac{a\ln b - b\ln a + a\ln a - b\ln b}{b-a} - 1$$

$$= \ln(L(a,b)) + \frac{b\ln b - a\ln a}{b-a} + \frac{a(\ln b + \ln a) - b(\ln a + \ln b)}{b-a} - 1$$

$$= \ln(L(a,b)) + \frac{b\ln b - a\ln a}{b-a} - \ln(ab) - 1.$$

Por tanto tenemos

$$0 \leq \ln\frac{A(\mathbf{x},\mathbf{p})}{G(\mathbf{x},\mathbf{p})} \leq \ln(L(a,b)) + \frac{b\ln b - a\ln a}{b-a} - \ln(ab) - 1,$$

lo cual a su vez implica que se cumplen las siguientes relaciones:

$$e^0 \leq e^{\frac{A(\mathbf{x},\mathbf{p})}{G(\mathbf{x},\mathbf{p})}} \leq e^{\ln(L(a,b)) + \frac{b\ln b - a\ln a}{b-a} - \ln(ab) - 1}$$

$$1 \leq \frac{A(\mathbf{x},\mathbf{p})}{G(\mathbf{x},\mathbf{p})} \leq e^{\ln(L(a,b))} \cdot e^{\frac{b\ln b - a\ln a}{b-a}} \cdot e^{\ln\frac{1}{ab}} \cdot e^{-1} = L(a,b) \cdot \frac{1}{ab} \cdot \left[e^{b\ln b - a\ln a}\right]^{\frac{1}{b-a}} \cdot \frac{1}{e}$$

$$= \frac{L(a,b)I(a,b)}{G^2(a,b)} = \Upsilon_1(a,b).$$

Esto demuestra el resultado. ■

Similarmente podemos obtener la inversa de la desigualdad entre las medias Geométrica y Armónica generalizada.

Proposición 3.2.1. Sea $H(\mathbf{x},\mathbf{p}) := \left(\sum_{i=1}^{n} \frac{p_i}{x_i}\right)^{-1}$, la media armónica generalizada. Entonces

$$1 \leq \frac{G(\mathbf{x},\mathbf{p})}{H(\mathbf{x},\mathbf{p})} \leq \Upsilon_1(a,b).$$

Demostración. Realizando un cambio de variable $x_i \to \frac{1}{x_i}, i = 1,2,...,n$. Entonces

$$A\left(\mathbf{p},\frac{1}{\mathbf{x}}\right) = \sum_{i=1}^{n} p_i \frac{1}{x_i} = \sum_{i=1}^{n} \frac{p_i}{x_i} = \frac{1}{H(\mathbf{p},\mathbf{x})}$$

y

$$G\left(\mathbf{p},\frac{1}{\mathbf{x}}\right)=\prod_{i=1}^{n}\left(\frac{1}{x_i}\right)^{p_i}=\frac{1}{\prod\limits_{i=1}^{n}x_i^{p_i}}=\frac{1}{G(\mathbf{p},\mathbf{x})}.$$

Para concluir, veremos que

$$\Upsilon_1\left(\frac{1}{b},\frac{1}{a}\right)=\Upsilon_1(a,b).$$

Esto se deduce de las siguientes relaciones:

$$\Upsilon_1\left(\frac{1}{b},\frac{1}{a}\right)=\frac{\dfrac{\dfrac{1}{a}-\dfrac{1}{b}}{\ln\left(\frac{1}{a}\right)-\ln\left(\frac{1}{b}\right)}\cdot\dfrac{\left(\dfrac{\left(\frac{1}{a}\right)^{\frac{1}{a}}}{\left(\frac{1}{b}\right)^{\frac{1}{b}}}\right)^{\frac{1}{\frac{1}{a}-\frac{1}{b}}}}{e}}{\left(\sqrt{\frac{1}{ab}}\right)^2}$$

$$=\frac{\dfrac{\dfrac{b-a}{ab}}{\ln 1-\ln a-\ln 1+\ln b}\cdot\dfrac{\left(\dfrac{\left(\frac{1}{a}\right)^{\frac{1}{a}}}{\left(\frac{1}{b}\right)^{\frac{1}{b}}}\right)^{\frac{ab}{b-a}}}{e}}{\dfrac{1}{ab}}=\frac{\dfrac{b-a}{ab(\ln b-\ln a)}\cdot\dfrac{\left(\frac{b^a}{a^b}\right)^{\frac{1}{b-a}}}{e}}{\dfrac{1}{ab}}$$

$$=\frac{\dfrac{b-a}{\ln b-\ln a}\cdot ab\dfrac{\left(\frac{b^a}{a^b}\right)^{\frac{1}{b-a}}}{e}}{ab}=\frac{\dfrac{b-a}{\ln b-\ln a}\cdot a^{\frac{b-a}{b-a}}b^{\frac{b-a}{b-a}}\dfrac{\left(\frac{b^{\frac{a}{b-a}}}{a^{\frac{b}{b-a}}}\right)}{e}}{ab}$$

$$\frac{\dfrac{b-a}{\ln b-\ln a}\cdot\dfrac{\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}}{e}}{ab}=\Upsilon_1(a,b).\quad\blacksquare$$

Concluimos esta sección presentada otras relaciones entre las medias.

Teorema 3.2.4. Para $\mathbf{x}:=\{x_i\},\mathbf{x}\in[a,b]$ con $0<a<b$ y $\mathbf{p}:=\{p_i\},\sum\limits_{i=1}^{n}p_i=1$ tenemos que

$$0\leq\sum_{i=1}^{n}p_ix_i\ln x_i-\left(\sum_{i=1}^{n}p_ix_i\right)\ln\left(\sum_{i=1}^{n}p_ix_i\right)\leq I(a,b)-\frac{G^2(a,b)}{L(a,b)}:=\Upsilon_3(a,b)\qquad(3.1)$$

Demostración.

Por el Teorema 2.3.4, para $f(t) = t \ln t$, tenemos

$$\Theta_f(a,b) = e^{\frac{b\ln b - a\ln a}{b-a} - 1} = e^{\frac{\ln\left(\frac{b^b}{a^a}\right)}{b-a}} e^{-1} = \frac{\left(e^{\ln\left(\frac{b^b}{a^a}\right)}\right)^{\frac{1}{b-a}}}{e} = \frac{\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}}{e} = I(a,b)$$

pues $f'(t) = \ln t + 1$, por lo tanto $(f')^{-1} = e^{t-1}$. Así

$$\mathcal{D}_f(a,b) = \frac{b\ln b - a\ln a}{b-a} I(a,b) + \frac{ba\ln a - ab\ln b}{b-a} - I(a,b)\ln I(a,b)$$

Ahora, por el Teorema 2.3.1

$$\sum_{i=1}^{n} p_i x_i \ln x_i - \left(\sum_{i=1}^{n} p_i x_i\right) \ln\left(\sum_{i=1}^{n} p_i x_i\right) \leq \mathcal{D}_f(a,b)$$

$$= \frac{b\ln b - a\ln a}{b-a} I(a,b) + \frac{ba\ln a - ab\ln b}{b-a} - I(a,b)\ln I(a,b)$$

$$= I(a,b)\left[\frac{\ln\left(\frac{b^b}{a^a}\right)}{b-a} - \ln\left(\frac{\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}}{e}\right)\right] - ab\frac{\ln b - \ln a}{b-a}$$

$$= I(a,b)\left[\frac{\ln\left(\frac{b^b}{a^a}\right)}{b-a} - \ln\left(\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}\right) + \ln e\right] - \frac{G^2(a,b)}{L(a,b)}$$

$$= I(a,b)\left[\frac{\ln\frac{b^b}{a^a}}{b-a} - \frac{\ln\frac{b^b}{a^a}}{b-a} + 1\right] - \frac{G^2(a,b)}{L(a,b)} = I(a,b) - \frac{G^2(a,b)}{L(a,b)}. \quad \blacksquare$$

3.3. Teoría de la Información

El concepto principal que es clave en el desarrollo de la Teoría de la Información es la llamada *entropía* o entropía de Shannon. En términos generales, la entropía representa la cantidad de información promedio que contiene los símbolos que se transmiten, es decir, es una medida de la incertidumbre de los mensajes enviados [15]. En términos matemáticos la entropía la definimos de la siguiente manera:

Definición 3.3.1. Sea X una variable aleatoria con distribución

$$P(X=i)=p_i, \quad p_i>0, \;\; i=1,2,...,r, \quad \sum_{i=1}^{r} p_i = 1.$$

La entropía se define como

$$\mathcal{H}(X) := \sum_{i=1}^{r} p_i \log \frac{1}{p_i}.$$

Observe que si p_i es la probabilidad de transmitir el símbolo i, entonces, los símbolos con menor probabilidad son las que aportan mayor información $\mathcal{H}(X)$. Entonces es importante tener una estimación de $\mathcal{H}(X)$ mediante desigualdades.

Proposición 3.3.1. Si $\mu := \min\limits_{1\le i\le r} p_i, \nu := \max\limits_{1\le 1\le r} p_i$, entonces

$$\mathcal{E}(\mu,\nu) := \mu \log\left(\frac{2\mu}{\mu+\nu}\right) + \nu \log\left(\frac{2\nu}{\mu+\nu}\right) \le \log r - \mathcal{H}(X) \le \log\left(\frac{(\mu+\nu)^2}{4\mu\nu}\right) := \mathcal{F}(\mu,\nu).$$

Además

$$\mathcal{E}(\mu,\nu) \le \log r - \mathcal{H}(X) \le \min\{\mathcal{F}(\mu,\nu), r\mathcal{E}(\mu,\nu)\}$$

Demostración. Para la parte izquierda hacemos uso del Teorema 2.3.5 con la función $f(x) = -\log x$, la sucesión $x_i = \frac{1}{p_i}, i = 1,...,r, \;\; p_r = \mu = a, p_s = \nu = b$. Entonces

$$\max_{1\le i<j\le n}\left[-p_i\log(x_i) - p_j\log(x_j) + (p_i+p_j)\log\left(\frac{p_ix_i+p_jx_j}{p_i+p_j}\right)\right] \le -\sum_{i=1}^{r} p_i\log x_i + \log\sum_{i=1}^{r} p_ix_i,$$

$$-p_r\log\frac{1}{p_r} - p_s\log\frac{1}{p_s} + (p_r+p_s)\log\left(\frac{p_r\frac{1}{p_r}+p_s\frac{1}{p_s}}{p_r+p_s}\right) \le \log\sum_{i=1}^{r} p_i\frac{1}{p_i} - \sum_{i=1}^{r} p_i\log\frac{1}{p_i},$$

$$p_r\log p_r + p_s\log p_s + (p_r+p_s)\log\left(\frac{2}{p_r+p_s}\right) \le \log r - \mathcal{H}(X),$$

$$p_r\left[\log p_r + \log\left(\frac{2}{p_r+p_s}\right)\right] + p_s\left[\log p_s + \log\left(\frac{2}{p_r+p_s}\right)\right] \le \log r - \mathcal{H}(X).$$

De aquí,

$$p_r\log\left(\frac{2p_r}{p_r+p_s}\right) + p_s\log\left(\frac{2p_s}{p_r+p_s}\right) \le \log r - \mathcal{H}(X)$$

y

$$\mu\log\left(\frac{2\mu}{\mu+\nu}\right) + \nu\log\left(\frac{2\nu}{\mu+\nu}\right) \le \log r - \mathcal{H}(X).$$

Para la parte derecha, por el Teorema 2.2.1 tenemos

$$J_f(\mathbf{p},\mathbf{x}) = \log r - \mathcal{H}(X) \leq -\log\mu - \log\nu + 2\log\left(\frac{\mu+\nu}{2}\right) =$$
$$\log\frac{1}{\mu} + \log\frac{1}{\nu} + \log\left(\frac{(\mu+\nu)^2}{4}\right) = \log\left(\frac{(\mu+\nu)^2}{4\mu\nu}\right). \quad \blacksquare$$

Teorema 3.3.1. Dados $x_i \in (0,\infty), p_i \geq 0$ con $P_n = 1$ y $a > 1$. Entonces

$$0 \leq \log_a\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i \log_a x_i \leq \frac{1}{2\ln a}\sum_{i,j=1}^{n}\frac{p_i p_j}{x_i x_j}(x_i - x_j)^2.$$

Demostración. Sea $f(x) = \log_a x$, es claro que f es diferenciable y cóncava en $(0,\infty)$ entonces tenemos que $f(x) - f(y) \leq f'(y)(x-y),\ \ \forall x, y > 0$, es decir

$$\log_a x - \log_a y \leq \frac{1}{\ln b}\frac{x-y}{y},\ \ \forall x, y > 0.$$

Sea $x = \sum_{j=1}^{n} p_j x_j$ y $y = x_i, i = 1, 2, ..., n$. Entonces,

$$\log_a\left(\sum_{j=1}^{n} p_j x_j\right) - \log_a x_i \leq \frac{1}{\ln}\frac{1}{x_i}\left(\sum_{j=1}^{n} p_j x_j - x_i\right),\ \ \forall i = 1, 2, ..., n.$$

Multiplicando por p_i y sumando sobre i de 1 a n, obtenemos

$$0 \leq \log_a\left(\sum_{j=1}^{n} p_j x_j\right) - \sum_{i=1}^{n} p_i \log x_i \leq \frac{1}{\ln b}\sum_{i=1}^{n}\frac{p_i}{x_i}\left(\sum_{j=1}^{n} p_j x_j - x_i\right)$$
$$= \frac{1}{2\ln b}\sum_{j=1}^{n}\sum_{i=1}^{n} p_i p_j\left(\frac{x_i}{x_j} + \frac{x_j}{x_i} - 2\right) = \frac{2}{\ln b}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{p_i p_j}{x_i x_j}(x_i - x_j)^2. \quad \blacksquare$$

Definición 3.3.2 (Divergencia Kullback-Leibler)**.** Sean P y Q distribuciones de una variable aleatoria discreta X tales que

$$P(X=i)=p_i>0,\ \ Q(X=i)=q_i>0,\ \ i=1,2,...,r,\ \ \sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i=1.$$

Definimos *Divergencia Kullback-Leibler* como

$$D_{KL}(P,Q):=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right)$$

Teorema 3.3.2. Sea $m:=\text{mín}\left(\frac{q_i}{p_i}\right)$; $M:=\text{máx}\left(\frac{q_i}{p_i}\right)$, $i=1,2,...,n$. Tenemos

$$0\leq D_{KL}(P,Q)\leq\log\Upsilon_1(m,M)$$

Demostración. Por el Teorema (3.2.3), aplicando log en ambos lados de la desigualdad

$$\log\left(\frac{A(\mathbf{x}\ ,\mathbf{p}\ )}{G(\mathbf{x}\ ,\mathbf{p}\ )}\right)\leq\log\left(\Upsilon_1(a,b)\right)\implies\log(A(\mathbf{x}\ ,\mathbf{p}\ ))-\log(G(\mathbf{x}\ ,\mathbf{p}\ ))\leq\log\left(\Upsilon_1(a,b)\right)$$
$$\implies\log\left(\sum_{i=1}^{n}p_ix_i\right)-\log\left(\prod x_i^{p_i}\right)\leq\log\left(\Upsilon_1(a,b)\right),$$

entonces

$$\log\left(\sum_{i=1}^{n}p_ix_i\right)-\sum_{i=1}^{n}p_i\log\left(x_i\right)\leq\log\left(\Upsilon_1(a,b)\right).$$

Ahora, sea $x_i=\frac{q_i}{p_i}$, $\forall i=1,2,..,n$ y $a=m:=\text{mín}\left(\frac{q_i}{p_i}\right), b=M:=\text{máx}\left(\frac{q_i}{p_i}\right)$, por lo tanto

$$\log\left(\sum_{i=1}^{n}p_i\frac{q_i}{p_i}\right)-\sum_{i=1}^{n}p_i\log\left(\frac{q_i}{p_i}\right)\leq\log\left(\Upsilon_1(m,M)\right)$$
$$\log\sum_{i=1}^{n}q_i-\sum_{i=1}^{n}p_i\log\left(\frac{q_i}{p_i}\right)\leq\log\left(\Upsilon_1(m,M)\right)$$
$$D_{KL}(P,Q)=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right)\leq\log\left(\Upsilon_1(m,M)\right)$$

3.4. Geometría

En esta sección presentamos una aplicación de la Desigualdad de Jensen relacionada con los polígonos inscritos en una circunferencia. Específicamente mostramos que el polígono de mayor área que se puede inscribir en una circunferencia es el regular.

Teorema 3.4.1. De todos los polígonos convexos de n lados inscritos en un círculo de radio r, el polígono de área máxima es el polígono regular.

Demostración. Sea V_α el conjunto de vértices del polígono α de n lados, es decir,

$$V_\alpha = \{A_1^\alpha, A_2^\alpha, ..., A_n^\alpha\},$$

donde A_i^α son los vértices del polígono α.
Si $A_k^\alpha, A_{k+1}^\alpha$ son vértices del polígono α y C es el centro de la circunferencia de radio r, es claro que el triángulo $\triangle A_k C A_{k+1} = A_{\Delta_k}$ es isósceles, pues $\overline{A_k C} = \overline{C A_{k+1}} = r$.

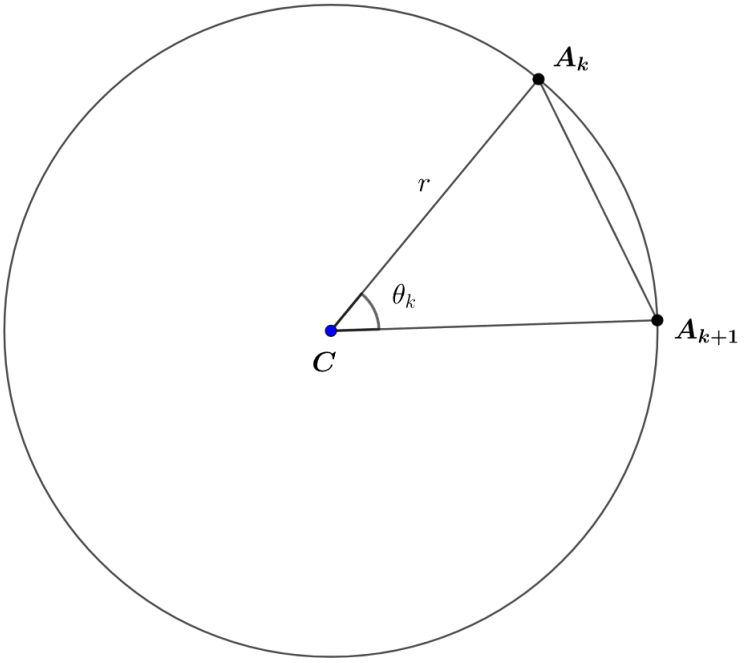

Figura 3.1

Sea θ_k el ángulo central del triángulo formado por $A_k C A_{k+1}$, y consideremos la altura del vértice A_k al segmento $A_{k+1}C$, formando el segmento $A_k B_k$.

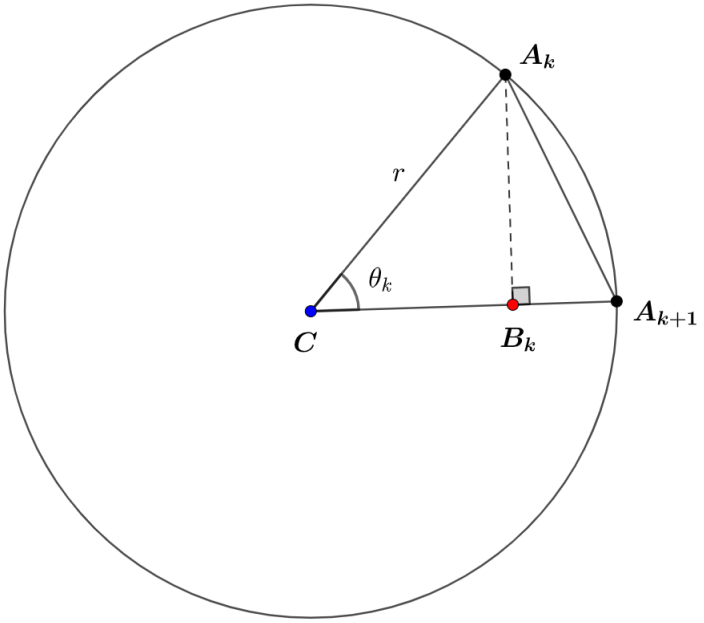

FIGURA 3.2

Como la circunferencia es de radio r, tenemos que el triángulo $\triangle A_k B_k C$ es rectángulo y las longitudes de sus lados son:

$$\overline{A_k C} = r, \quad \overline{A_k B_k} = \overline{A_k C} \sin\theta_k = r \sin\theta_k.$$

De esta forma el área del triángulo $A_{\Delta_k} = \dfrac{\overline{A_{k+1}C} \cdot \sin(\theta_k)}{2} = \frac{r}{2}\sin(\theta_k)$.
Así, al dividir al polígono α en n triángulos donde cada triángulo tiene área $\frac{r}{2}\sin(\theta_k)$ para $k = 1, ..., n$, el área del polígono α es

$$A = \frac{r}{2}\sum_{k=1}^{n} \sin(\theta_k), \quad 0 < \theta_k < \pi, \ \sum_{k=1}^{n} \theta_k = 2\pi.$$

Por otra parte, si consideramos el polígono regular α' tenemos que al dividirlo en n triángulos $\triangle A_k C A_{k+1}$, todos serán iguales y en particular todos tienen ángulo central $\theta = \dfrac{2\pi}{n}$. Por lo tanto el área del polígono α' es

$$A' = \frac{r}{2} n \sin\left(\frac{2\pi}{n}\right).$$

Como la función $\sin(x)$ es cóncava, por (J) tenemos que

$$\sin\left(\sum_{k=1}^{n}\frac{1}{n}\theta_k\right) \geq \sum_{k=1}^{n}\frac{1}{n}\sin\theta_k.$$

Finalmente multiplicando por nr y simplificando obtenemos

$$\frac{r}{2}n\sin\left(\frac{1}{n}\sum_{k=1}^{n}\theta_k\right) \geq \frac{r}{2}n\frac{1}{n}\sum_{k=1}^{n}\sin\theta_k,$$

lo cual implica

$$A' = \frac{r}{2}n\sin\left(\frac{2\pi}{n}\right) \geq \frac{r}{2}\sum_{k=1}^{n}\sin\theta_k = A.$$

Por tanto, concluimos que el polígono regular inscrito en la circunferencia de radio r es el de área máxima. ■

3.5. Probabilidad

En esta sección analizamos la versión probabilista de la Desigualdad de Jensen. Esta desigualdad relaciona el valor esperado de funciones de variables aleatorias con funciones del valor esperado, cuando las funciones son convexas.

Sea X una variable aleatoria arbitraria con valor esperado $E(X) = \mu_X < \infty$, y sea $f : \mathbb{R} \to \mathbb{R}$ una función. Entonces, $Y = f(X)$ define una variable aleatoria, y supondremos que su valor esperado es $\mu_Y = E(f(X)) = E(Y) < \infty$.

Por otro lado, recordemos que si f es una función convexa en $I = (a, b)$ entonces existe $m \in (a, b)$ tal que para cada $t \in (a, b)$,

$$f(t) + m(t - x) \leq f(x) \quad \forall x \in (a, b). \tag{3.5}$$

Teorema 3.5.1. Sea X una variable aleatoria que toma valores en $I = (a, b)$ y $f : I \to \mathbb{R}$ una función convexa. Entonces

$$f\left(E(X)\right) \leq E\left(f(X)\right).$$

Demostración. Por (3.5), tomando $t = \mu = E(X) \in (a, b)$ tenemos

$$f(\mu) + m(\mu - X) \leq f(X).$$

Calculado esperanza en ambos lados,

$$E(f(\mu) + m(\mu - X)) \leq E(f(X)).$$

Por las propiedades de la esperanza obtenemos

$$E(f(\mu)) + mE(\mu - X) \leq E(f(X)).$$

y

$$E(f(\mu)) + m(E(\mu) - E(X)) \leq E(f(X)).$$

Ahora, como $E(X) = \mu$, $E(f(\mu)) = f(\mu)$ y $E(\mu) = \mu$, concluimos que

$$f\left(E(X)\right) \leq E(f(X)). \quad \blacksquare$$

3.6. Teoría de números

El siguiente resultado establece una desigualdad importante en Teoría de Números la cual fue publicada por Nesbitt en 1903 en el artículo [11]. La demostración que presentamos está basada en la Desigualdad de Jensen la cual es más sencilla que la original.

Teorema 3.6.1. *(Desigualdad de Nesbitt).* Si a, b, c son números positivos, entonces

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \geq \frac{3}{2}.$$

Demostración. Veamos que esta desigualdad es homogénea, es decir que dada cualquier terna (a, b, c) que cumplen la desigualdad, entonces para cualquier $t > 0$ la terna (ta, tb, tc) también la cumple.

$$\frac{ta}{tb+tc} + \frac{tb}{tc+ta} + \frac{tc}{ta+tb} = \frac{ta}{t(b+c)} + \frac{tb}{t(c+a)} + \frac{tc}{t(a+b)} = \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \geq \frac{3}{2}.$$

por tanto es homogénea.

Así, sin pérdida de generalidad suponemos que $a+b+c=1$, entonces

$$\frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b}=\frac{a}{1-a}+\frac{b}{1-b}+\frac{c}{1-c}=D$$

Consideremos la función $f(x)=\dfrac{3x}{1-x}$, $0<x<1$, esta función es convexa, pues $f''(x)=\dfrac{6}{(1-x)^3}>0, \forall x\in(0,1)$.

Utilizando (J)

$$\begin{aligned} D=\frac{f(a)+f(b)+f(c)}{3}&=\frac{1}{3}f(a)+\frac{1}{3}f(b)+\frac{1}{3}f(c)\geq f\left(\frac{1}{3}a+\frac{1}{3}b+\frac{1}{3}c\right)\\ &=f\left(\frac{1}{3}(a+b+c)\right)=f\left(\frac{1}{3}\right)=\frac{3}{2}. \quad \blacksquare \end{aligned}$$

Bibliografía

[1] N. Bingham, C. Goldie, and J. Teugels. *Regular variation*. Number 27. Cambridge university press, 1989. 2

[2] P. Bullen. *Handbook of Means and Their Inequalities*. Mathematics and Its Applications. Springer Netherlands, 2013. ISBN 9789401703994. 2

[3] B. Carlson. The logarithmic mean. *The American Mathematical Monthly*, 79(6): 615–618, 1972. 3

[4] S. Dragomir. A converse result for jensen's discrete inequality via grüss' inequality and applications in information theory. *An. Univ. Oradea Fasc. Mat*, 7(4):178–189, 1999. 2, 28

[5] S. Dragomir and C. Goh. Some bounds on entropy measures in information theory. *Applied Mathematics Letters*, 10(3):23–28, 1997. 2

[6] G.H. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. Cambridge university press, 1952. 1

[7] C. Heiberg. Functions with asymptotically infinite differences. *Publ. Inst. Math. Belgrade*, 12(26):45–49, 1971. 2, 30

[8] J. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, 30(1):175–193, 1906. 1, 2, 11, 12, 16

[9] M. Khan, N. Latif, and J. Pečarić. Generalization of majorization theorem. *J. Math. Inequal*, 9(3):847–872, 2015. 2

[10] A. Marshall, I. Olkin, and B. Arnold. Inequalities: theory of majorization and its applications. 1979. 3

[11] A. Nesbitt. Problem 15114. *Educational Times*, 3(2):37–38, 1903. 51

[12] J. Pečarić. Inverse of jensen-steffensen's inequality. *Glasnik Mat.*, 16(36):229–233, 1981. 2

[13] J. Pečarić. A simple proof of the jensen-steffensen inequality. *The American Mathematical Monthly*, 91(3):195–196, 1984. 2, 18

[14] J. Pečarić and Y. Tong. *Convex functions, partial orderings, and statistical applications*. Academic Press, 1992. 2

[15] C. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 2, 44

[16] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Number parte 11 in The Mathematical Theory of Communication. University of Illinois Press, 1962. 3

[17] S. Simic. On a global upper bound for jensen's inequality. *Journal of mathematical analysis and applications*, 343(1):414–419, 2008. 2

[18] S. Simic. Best possible global bounds for jensen's inequality. *Applied mathematics and computation*, 215(6):2224–2228, 2009. 2

[19] S. Simic. Jensen's inequality and new entropy bounds. *Applied Mathematics Letters*, 22(8):1262–1265, 2009. 2

[20] J. Steffensen. On certain inequalities and methods of approximation. *Journal of the Institute of Actuaries*, 51(3):274–297, 1919. 2, 16

[21] K. Stolarsky. Generalizations of the logarithmic mean. *Mathematics Magazine*, 48 (2):87–92, 1975. 3

[22] J. Sándor. On the identric and logarithmic means. *Aequationes Mathematicae*, 40 (1):261–270, 1990. 3