PRÁCTICO 1 – TEORÍA DE INFORMACIÓN

CANTIDAD DE INFORMACIÓN - ENTROPÍA

Grupo 2:

- ***Caballero, Ariel***
- ***Espejo, Luciano***
- ***Giuliani, Melissa Ann***

1) Supón una imagen de alta definición dividida en cuadros de 1920 por 1080 píxeles cada uno. Indica la cantidad de información que proporciona un cuadro de imagen si está codificada en 24 bits por píxel. Luego, considera una imagen con una resolución de 3840 x 2160 píxeles, ¿cuánta información se genera si se codifica en HDR (High Dynamic Range) con 10 bits por canal de color?

RTA:

Datos:

- Resolución: **1920 x 1080 píxeles = 2.037.600 píxeles**
- Bits por píxel: **24 bits**

Con estos datos, el número total de píxeles en la imagen es de:

- Total de píxeles en imagen: (2.037.600) x 24 = **49.766.400 bits (6.220.800 Bytes = 6.22 MB)**

Entonces la cantidad de información que proporciona un cuadro de imagen de 1920 x 1080 píxeles codificada en 24 bits por píxel es **6.22MB**

Luego, consideramos los siguientes datos:

- Resolución 2: **3840 x 2160 píxeles = 8.294.400 píxeles**
- Bits por canal: **10 bits**
- Canales de Color: **3 (por ser RGB)**

A partir de esto, hacemos los siguientes cálculos:

- Cantidad de bits por píxel: **10 x 3 = 30 bits por píxel**

- ♣ Cantidad de Información en bits: **8,294,400 x 30 = 248,832,000 bits (31,104,000 B = 31.10 MB)**

Entonces, la cantidad de información que proporciona un cuadro de imagen de 3840 x 2160 píxeles codificada en HDR con 10 bits por canal de color es **31.10 MB.**

2) ¿Qué información genera un locutor de radio a través de 1000 palabras tomadas al azar de un vocabulario de 10000? ¿Qué relación existe entre las dos cantidades anteriores?.

RTA:

La cantidad de información que una palabra transmite puede ser calculada como:

$$I(w) = \log_2(N)$$

Siendo **I(w)** la cantidad de información en bits, y **N** el tamaño del vocabulario. Como nuestro tamaño de vocabulario es 10.000, calculamos:

$$I(w) = \log_2(10{,}000) \approx \log_2(10^4) = 4 \times 3.32 = 13.28 \text{ bits/píxel}$$

Ahora, para calcular la información generada al utilizar 1000 palabras es:

Cantidad Total de Información = Número de bits por palabra x Número de palabras

I(1000) = 13.28 x 1000 ≈ 13,280 bits

Por lo tanto, la información generada por el locutor al pronunciar 1000 palabras de un vocabulario de 10000 palabras es aproximadamente 13280 bits.

Si nos estamos refiriendo a las siguientes cantidades:

- ♠ Información generada por el locutor: **13280 bits**
- ♠ Información de la imagen (alta definición): **49,766,400 bits** (el de imagen de 1920 x 1080)

Al comparar estas cantidades, podemos observar cuán alta es la diferencia en cantidad de información que ofrece cada tipo de datos. Podemos pensar que en 1000 palabras podemos recibir mucho contenido de información, pero no es una cantidad cercana a la que se obtiene a partir de la imagen. Concluimos que cantidad de información que brinda una imagen es extremadamente superior a la cantidad que brindan mil palabras. Podríamos aplicar la famosa frase: "*Una imagen dice más que mil palabras*".

3) ¿Qué cantidad de información genera un mensaje de texto compuesto por 150 caracteres tomados al azar de un alfabeto de 100 símbolos? ¿Cómo cambiaría si el mensaje es de 250 caracteres con el mismo alfabeto?

RTA:

Si vamos a calcular la cantidad de información que genera un mensaje de texto, podemos usar la fórmula de Entropía de Shannon. La cantidad de Información I en bits de un mensaje compuesto por **n** caracteres elegios al azar de un alfabeto de tamaño **N** se calcula como:

$$\mathbf{I = n \times \log_2(N)}$$

Donde **n** es el número de caracteres en el mensaje, **N** es el tamaño del alfabeto (el número de símbolos), y $\mathbf{\log_2(N)}$ es la cantidad de información por caracter. Vemos ahora los datos que nos proveen:

Datos para mensaje de 150 caracteres:

- Número de caracteres: **n = 150**
- Tamaño de Alfabeto: **N = 100**
- Cantidad de información por carácter: $\mathbf{\log_2(100)} \approx \log_2(10^2) = 2 \times \log_2(10) \approx 2 \times 3.32 =$ **6.64 bits**

Entonces, la cantidad total de información para el mensaje de 150 caracteres es:

$$I_{150} = 150 \times 6.64 = \mathbf{996\ bits}$$

Datos para mensaje de 250 caracteres:

- Número de caracteres: **n = 250**
- Tamaño de Alfabeto: **N = 100**
- Cantidad de información por carácter: $\mathbf{\log_2(100)}$ = **6,64 bits**

Entonces, la cantidad total de información para el mensaje de 250 caracteres es:

$$I_{250} = 250 \times 6{,}64 = 1660 \text{ bits}$$

Por lo tanto, un mensaje de 150 caracteres elegidos al azar genera aproximadamente 996 bits de información, y uno de 250 caracteres tomados al azar genera aproximadamente 1660 bits de información.

4) Demostrar las siguientes igualdades:

$$\log_a x = \frac{\log_b x}{\log_b a} = \log_a b \cdot \log_b x$$

RTA:

$log_a x = log_b x / log_b a$ **(1)**	$log_a x = log_a b * log_b x$
Demostración $a^y = x \quad (por\ def\ de\ log)$ $log_b a^y = log_b x$ $y * log_b a = log_b x$ $y = log_b x / log_b a$ c.q.d.	**Demostración** $log_a b = y$ $a^y = b\ (por\ def\ de\ log)$ $log_b a^y = log_b b$ $y * log_b a = 1$ $y = 1/ log_b a$ $log_a b = 1/log_b a$ **(2)** $de\ (1)\ log_a x = log_b x/ log_b a$ $de\ (2)\ log_a x = log_a b * log_b x$ c.q.d.

5) Considera una fuente F con cinco símbolos (s1, s2, s3, s4, s5) con las probabilidades p1=1/3, p2=1/6, p3=1/6, p4=1/6, p5=1/6. Calcula la cantidad de información individual y la entropía de esta fuente.

RTA:

Sabemos que para calcular la información individual para un símbolo **x**, del cual conocemos su probabilidad $\mathbf{p_x}$, usamos la Fórmula de Shannon.

$$\mathbf{I(x) = - \log_2(p_x) = \log_2(1/p_x)}$$

Tenemos los siguientes datos:

Datos:

- **p_{s1} = 1/3**
- **p_{s2} = 1/6**
- **p_{s3} = 1/6**
- **p_{s4} = 1/6**
- **p_{s5} = 1/6**

Aplicamos entonces la formula a cada símbolo (s1, s2, s3, s4, s5) para encontrar la **Cantidad de Información Individual**:

- **I(s1)** = $\log_2(3)$ = **1.5849 bits**
- **I(s2)** = $\log_2(6)$ = **2.5849 bits**
- **I(s3)** = $\log_2(6)$ = **2.5849 bits**
- **I(s4)** = $\log_2(6)$ = **2.5849 bits**
- **I(s5)** = $\log_2(6)$ = **2.5849 bits**

(Recordar: para calcular estos logaritmos con la calculadora, sería: $\log_a b$ = log b / log a)

Ahora podemos calcular la **entropía H(S)**, y se calcula como la esperanza matemática de la cantidad de información:

$$H(S) = \sum_{x \in S} p(x)\, I(x)$$

H(S) = (1/3 x 1.5849) + 4 x (1/6 x 2.5849)

= 0.5283 + 4 x (0.4308)

= 0.5283 + 1.7232

H(S) = 2.2515 bits

6) Comprobar que, al tirar un dado, si las probabilidades individuales son iguales se logra el máximo valor de la información promedio. Luego, compara este resultado con una

distribución de probabilidades en el mismo dado donde p1=0.5, p2=0.2, p3=0.1, p4=0.08, p5=0.06, p6=0.04, p7=0.015, p8=0.005.

RTA:

Datos al tirar un dado:

- $\mathbf{p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6}$

Entonces, podemos calcular la cantidad de información individual de la probabilidad de tirar alguna de las caras (tienen el mismo valor de probabilidad para cada caso):

- $\mathbf{I_{p1}} = \log_2 (1/p_1) = \log_2 (6) =$ **2.5849 bits**

Luego, calculamos su entropía:

- **H** = 6 x ((1/6) x 2.5849) = **2.5849 bits**

Datos provistos:

- $\mathbf{p_1 = 1/3}$
- $\mathbf{p_2 = 1/4}$
- $\mathbf{p_3 = 1/6}$
- $\mathbf{p_4 = 1/9}$
- $\mathbf{p_5 = 1/12}$
- $\mathbf{p_6 = 1/18}$

Ahora calculamos las $I_{px:}$

- $\mathbf{I_{p1}} = \log_2 (1/p_1) = \log_2 (3) =$ **1.5849**
- $\mathbf{I_{p2}} = \log_2 (1/p_2) = \log_2 (1/4) =$ **2**
- $\mathbf{I_{p3}} = \log_2 (1/p_3) = \log_2 (1/6) =$ **2.5849**
- $\mathbf{I_{p4}} = \log_2 (1/p_4) = \log_2 (9) =$ **3.1699**
- $\mathbf{I_{p5}} = \log_2 (1/p_5) = \log_2 (12) =$ **3.5849**
- $\mathbf{I_{p6}} = \log_2 (1/p_6) = \log_2 (18) =$ **4.1699**

Luego, calculamos su entropía:

H = (1/3) x 1.584962500721156 + (1/4) x 2 + (1/6) x 2.584962500721156 + (1/9) x 3.1699250014423126 + (1/12) x 3.5849625007211565 + (1/18) x 4.169925001442312

H = **2.3417**

Observando las dos entropías calculadas, concluimos que el valor de la entropía es mayor cuando todas las probabilidades son iguales.

7) Sean 12 monedas una de las cuales tiene peso diferente, indicar cuántas pesadas son necesarias para encontrarla, especifique la metodología de peso

RTA:

Para encontrar una moneda diferente en peso (ya sea más ligera o más pesada) entre 12 monedas usando una balanza de dos platillos, podemos hacerlo en un máximo de 3 pesadas. Aquí se explica el método paso a paso:

Paso 1: Primera Pesada

Divide las 12 monedas en 3 grupos de 4 monedas cada uno.

- **Pesada 1:** Coloca 4 monedas en un platillo y 4 monedas en el otro platillo.
    - **Caso A:** Si los platillos están equilibrados, entonces la moneda diferente está entre las 4 monedas que no se pesaron.
    - **Caso B:** Si los platillos no están equilibrados, la moneda diferente está en el platillo más ligero o más pesado.

Paso 2: Segunda Pesada

Dependiendo del resultado de la primera pesada:

- **Caso A:** (Moneda diferente entre las 4 que no se pesaron):
    - Toma 3 de las 4 monedas que no se pesaron y compara 2 de ellas en la balanza.
        - **Caso A1:** Si las dos monedas pesadas son iguales, la moneda diferente es la que no se pesó.
        - **Caso A2:** Si las dos monedas pesadas son diferentes, puedes identificar la moneda diferente por la diferencia de peso (más ligera o más pesada).

- **Caso B:** (Moneda diferente entre las 8 monedas pesadas en la primera pesada):
    - Toma las 4 monedas del platillo más ligero o más pesado (dependiendo de si la moneda diferente es más ligera o más pesada).
    - Divide estas 4 monedas en 2 grupos de 2 monedas cada uno y colócalas en la balanza.
        - **Caso B1:** Si los platillos están equilibrados, la moneda diferente está entre las 2 monedas que no se pesaron.
        - **Caso B2:** Si los platillos no están equilibrados, la moneda diferente está en el platillo más ligero o más pesado.

Paso 3: Tercera Pesada

- **Caso A1/B1:** (Moneda diferente entre 2 restantes):
    - Pesa las 2 monedas restantes. La más ligera o más pesada es la moneda diferente.
- **Caso A2/B2:** (Moneda diferente ya identificada entre 2 monedas):
    - La moneda diferente es la que desequilibra la balanza.

8) ¿Cuál es el contenido de información de una letra del abecedario de 27 letras? ¿Cuál es el de las letras tomadas de a dos y de a tres?, suponiendo a priori, que no existe preferencia por ninguna letra.

RTA:

Para 27 letras:

- **letra = a, b, c, d, e, f, g, h, i, j, k, l, m, n, ñ, o, p, q, r, s, t, u, v, w, x, y, z**
- **p(letra) = 1/27**

$$\log_2 \frac{1}{\left(\frac{1}{27}\right)} = \log_2 27 = \mathbf{4.7548}$$

La cantidad de información de una letra del abecedario de 27 letras es de **4.7548 bits.**

De a 2 letras:

$$2 * 4.7548 = \mathbf{9.5097}$$

La cantidad de información de las letras tomadas de a 2 es de **9.5097 bits.**

De a 3 letras:

$$3 * 4.7548 = \mathbf{14.2644}$$

La cantidad de información de las letras tomadas de a 3 es de **14.2644 bits.**

9) ¿Cómo serán las cifras anteriores respecto de la forma castellana de escritura cotidiana? Si a su criterio existiera diferencia ¿a qué se debería?

RTA:

La cantidad de información en la escritura cotidiana en castellano es **significativamente menor** que en el caso de 27 letras. Pues la probabilidad de elegir una letra se basa en la dependencia, es decir, la aparición de una letra depende de otra anterior. Esto está ligado a las reglas del idioma castellano. Diríamos que los símbolos de nuestro idioma no son equiprobables.

10) Explica las propiedades de la cantidad de información. Da un ejemplo para cada una de estas propiedades.

RTA:

1) La **cantidad de información (I)** es **siempre positiva**.
   a. Si un evento tiene una probabilidad de 1 (es decir, es seguro que ocurrirá), la cantidad de información es $I(x) = -\log_2(1) = 0$. Si un evento tiene una probabilidad menor, por ejemplo, 0.5, entonces la cantidad de información es $I(x) = -\log_2(0.5) = 1$ bit. Esto ilustra que la cantidad de información es siempre no negativa.
2) La **cantidad de información de un evento**, es **inversamente proporcional a la probabilidad de ocurrencia** del mismo. Si $\mathbf{p_{x1} > p_{x2}}$, entonces $\mathbf{I(x_1) < I(x_2)}$
   a. Si lanzamos un dado justo, la probabilidad de obtener cualquier número es 1/6, lo que equivale a $I(x) = -\log_2(1/6) \approx 2.58$ bits de información. Sin embargo, si usamos un dado sesgado donde la probabilidad de obtener un "1" es 0.9, la cantidad de información asociada a obtener un "1" será menor: $I(x) = -\log_2(0.9) \approx 0.15$ bits. Cuanto más probable sea el resultado, menor será la cantidad de información asociada.
3) La cantidad de información aumenta con la cantidad de posibles eventos o

mensajes.

a. Imaginemos dos situaciones diferentes:

Lanzar una moneda justa:

Aquí, solo hay **2 posibles resultados**: cara o cruz. La probabilidad de cada resultado es p(x)= ½. La cantidad de información es $I(x) = \log_2(2) = 1$ bit.

Esto significa que conocer el resultado de un lanzamiento de moneda justa proporciona 1 bit de información.

Lanzar un dado justo:

En este caso, hay **6 posibles resultados**: 1, 2, 3, 4, 5, 6. La probabilidad de cada resultado es p(x)= 1/6. La cantidad de información es $I(x) = \log_2(6) \approx 2.58$ bits.

Esto significa que conocer el resultado de un lanzamiento de dado proporciona más información que un lanzamiento de moneda, ya que hay más posibles resultados y, por lo tanto, más incertidumbre.

Comparación:

En el caso del lanzamiento de la moneda, donde solo hay 2 posibles eventos, la cantidad de información es **1 bit**.

En el caso del lanzamiento del dado, donde hay 6 posibles eventos, la cantidad de información es **2.58 bits**.

4) La **cantidad de información** de **eventos** o mensajes **independientes** A, B, ..., C; es **aditiva**: **I(A, B, ..., C) = I(A) + I(B) + ... + I(C).**

a. Si lanzamos dos monedas justas, la probabilidad de obtener "cara" en ambas es **p(x** y **y)** = ½ x ½ = ¼ . La cantidad de información es:

I(**x** y **y**) = -log2(¼) = 2 bits.

Esto es igual a la suma de la cantidad de información de cada evento individual (1 bit por cada lanzamiento de moneda).

5) La cantidad de información promedio (H) se maximiza ante sucesos equiprobables.

a.

11) Dada una variedad V = 1000 sucesos encontrar la base óptima de representación que hace mínima la siguiente relación número más corto, es decir que minimice I . b (cantidad de información por base).

RTA:

Datos:

- V = 1000

Para calcular la Cantidad de Información. El cálculo es: $\mathbf{I = \log_b(V)}$ donde **b** es la base del sistema de representación. Luego para el producto I.b:

I.b = $\log_b$ 1000 *b

Ahora debemos lograr que el valor de b minimice la expresión.

Paso 1: Expresar el problema en términos más simples

Podemos reescribir $\log_b(V)$ en términos de logaritmos en base 2:

$$(V) \ = \frac{(V)}{(b)}$$

Por lo tanto, el producto a minimizar es:

$$I \ b = \frac{(V)}{(b)} \ \ b$$

Simplificamos la expresión:

$$I \ b = (V) \ \ \frac{b}{(b)}$$

Aquí, $\log_2(V)$ es constante, así que necesitamos minimizar la función:

$$f(b) \ = \frac{b}{(b)}$$

Paso 2: Encontrar el valor óptimo de b:

Para minimizar f(b), tomamos la derivada de f(b) respecto a b y se iguala a 0.

$$f(b) \ = \frac{(b) - b^* \ \frac{1}{b^* \mathrm{ln} ln\,(2)}}{(b)^2}$$

Simplificando:

$$(b) \ \ - \frac{1}{ln(2)} = \ 0$$

Entonces:

$$(b) \ \ = \ \frac{1}{ln(2)}$$

Tomando logaritmo base 2 de ambos lados:

$$b \ = 2^{1/\mathrm{ln} ln\,(2)}$$

Calculando *b*: **b ≈ 2.718**

Finalmente, la base óptima b para minimizar el producto entre la cantidad de información y la base de representación es **e** (aproximadamente 2.718). Esto significa que la base que minimiza la cantidad de información por base en este contexto es la base **e**.

En la práctica, si solo se permiten bases enteras, la base óptima más cercana es **2** o **3**, dependiendo del contexto. En la mayoría de los casos, **base 2** (sistema binario) es una elección óptima para sistemas digitales.