FACULTAD DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES

UNIVERSIDAD NACIONAL DE SAN JUAN

DEPARTAMENTO DE INFORMÁTICA

Teoría de la Información

TRABAJO PRÁCTICO N° 1

DOCENTES:

- Klenzi Raul
- Manuel Ortega
- Amaya Fabricio

ALUMNOS:

- Aubone Franco
- Morandi, Jeremías
- Santillan, Joaquin
- Zumel, Candela

2025

Curiosidades de la información que pueden extraer las personas.......

ÁREA TEMÁTICA No1

PRÁCTICO 1 TEORÍA DE LA INFORMACIÓN

CANTIDAD DE INFORMACIÓN - ENTROPÍA

1. Supón una imagen de resolución 2560 x 1440 píxeles, codificada en 30 bits por píxel. Calcula la cantidad de información que proporciona un cuadro de imagen. Luego, considera una imagen de 7680 x 4320 píxeles (resolución 8K) codificada en HDR con 12 bits por canal de color, ¿cuánta información se genera?
2. Un narrador utiliza 1000 palabras tomadas al azar de un vocabulario de 20.000 palabras. Calcula la información generada. Luego, compárala con la información contenida en una única imagen de resolución 640x480 píxeles codificada en 24 bits por píxel (color verdadero). Analiza cuál contiene más información y verifica si se cumple el dicho popular 'una imagen vale más que mil palabras' incluso con baja resolución..
3. Calcula la información generada por un mensaje de 200 caracteres tomados al azar de un alfabeto de 32 símbolos. ¿Y por un mensaje de 200 caracteres con un alfabeto de 64 símbolos? Ahora repite el cálculo para un mensaje de 400 caracteres con alfabetos de 32 y 64 símbolos. Analiza cómo cambia la cantidad de información y verifica si duplicar el tamaño del alfabeto duplica la información total..
4. Demostrar las siguientes igualdades:

$$\log_a x = \frac{\log_b x}{\log_b a} = \log_a b \,.\, \log_b x$$

5. Una fuente F tiene 6 símbolos con probabilidades: p1=0.4, p2=0.2, p3=0.15, p4=0.1, p5=0.1, p6=0.05. Calcula la información individual y la entropía de la fuente.
6. Justifica por qué una distribución uniforme maximiza la entropía. Compara un dado justo (6 caras, p=1/6) con un dado sesgado: p1=0.3, p2=0.25, p3=0.15, p4=0.15, p5=0.1, p6=0.05..
7. Sean 12 monedas una de las cuales tiene peso diferente, indicar cuántas pesadas son necesarias para encontrarla, especifique la metodología de peso
8. Calcula la información de una letra al azar de un alfabeto de 28 símbolos (incluyendo ñ). Luego, de pares y ternas de letras.
9. ¿Cómo serán las cifras anteriores respecto de la forma castellana de escritura cotidiana? Si a su criterio existiera diferencia ¿a qué se debería?
10. Explica las propiedades de la cantidad de información. Da un ejemplo para cada una de estas propiedades.
11. Dada una variedad V = 2000 sucesos encontrar la base *b* óptima de representación que hace mínima la siguiente relación número más corto, es decir que minimice I . b (cantidad de información por base).
12. Para el texto contenido en el código QR de la parte superior de la página encuentre la entropía de orden cero o independiente tanto de lo relevado en el QR como en el texto correcto. Que conclusiones saca?

CANTIDAD DE INFORMACIÓN - ENTROPÍA

1. Supón una imagen de resolución 2560 x 1440 píxeles, codificada en 30 bits por píxel. Calcula la cantidad de información que proporciona un cuadro de imagen. Luego, considera una imagen de 7680 x 4320 píxeles (resolución 8K) codificada en HDR con 12 bits por canal de color, ¿cuánta información se genera?

Primero, obtenemos el total de píxeles en la imágen, con:
2560 x 1440 = 3.686.400 píxeles en total
Y sabiendo que está codificado en 30 bits por píxel, hacemos:
3.686.400 píxeles x 30 bits/pixel = **110.592.000 bits**
Lo hacemos más legible pasándolo a MegaBits (dividiendo en 1.000.000)
110.592.000 / 1.000.000 = 110,59 Mb
Se genera 110,59 Mb de información

Pasamos a la segunda imágen de 7690 x 4320 píxeles, los pasos son similares al anterior, pero antes debemos calcular los bits por pixels, sabiendo que se codifica con 12 bits por canal de color (siendo RGB 3 colores), nos da 12 x 3 = **36 bits por píxel**, a partir de ahora los pasos son similares al anterior:
7690 x 4320 = 33.177.600 píxeles en total
y ahora sabiendo que son 36 bits por píxel:
33.177.600 pixeles x 36 bits/pixel = **1.194.393.600 bits**
Lo convertimos a GigaBits (dividiendo en 1.000.000.000)
1.194.393.600 / 1.000.000.000 = 1,19 Gb
Se genera 1,19 Gb de información.

2. Un narrador utiliza 1000 palabras tomadas al azar de un vocabulario de 20.000 palabras. Calcula la información generada. Luego, compárala con la información contenida en una única imagen de resolución 640x480 píxeles codificada en 24 bits por píxel (color verdadero). Analiza cuál contiene más información y verifica si se cumple el dicho popular 'una imagen vale más que mil palabras' incluso con baja resolución..

Primero, calculamos a través de la fórmula de Shannon la información por palabra:
Información por palabra = $\log_2$(número de opciones disponibles)
= $\log_2(20.000)$
= 14,29 bits por palabra

Por ende, 1000 palabras x 14,29 bits = **14.290 bits por palabra**

Ahora calculamos la información de la imágen
640 x 480 = 307.200 píxeles
Bits por píxel = 24 bits
307.200 pixels x 24 bits/pixel = **7.372.800 bits**

A la vista salta que se transmiten más información a través de la imagen
7.372.800 / 14.290 = 515
Hasta 515 veces más de información transmite la imágen con respecto a las palabras.

3. Calcula la información generada por un mensaje de 200 caracteres tomados al azar de un alfabeto de 32 símbolos. ¿Y por un mensaje de 200 caracteres con un alfabeto de 64 símbolos? Ahora repite el cálculo para un mensaje de 400 caracteres con alfabetos de 32 y 64 símbolos. Analiza cómo cambia la cantidad de información y verifica si duplicar el tamaño del alfabeto duplica la información total..

Iniciamos con el primero, un mensaje de 200 caracteres de un alfabeto de 32 símbolos
Usaremos: Información total = númeroDeCaracteres x $\log_2$(TamañoAlfabeto)
númeroDeCaracteres = 200
$\log_2(32)$ = 5 bits/caracter
200 caracteres x 5 bits/carácter = 1000 bits
Es decir, se generó 1000 bits de información

Seguimos con el de 200 caracteres con un alfabeto de 64 símbolos, siguiendo la fórmula anterior
númeroDeCaracteres = 200
$\log_2(64) = 6$
200 caracteres x 6 bits/carácter = 1200 bits
Se generaron 1200 bits de información

Ahora continuamos con los demás casos, 400 caracteres con un alfabeto de 32 símbolos:
númeroDeCaracteres = 400
$\log_2(32) = 5$
400 caracteres x 5 bits/carácter = 2000 bits

Último caso, 400 caracteres con un alfabeto de 64 símbolos
númeroDeCaracteres = 400
$\log_2(64) = 6$
400 caracteres x 6 bits/carácter = 2400 bits

Podemos notar que duplicar el tamaño del alfabeto NO duplica la cantidad de información generada

4. Demostrar las siguientes igualdades:

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

Demostración:
Por definición de logaritmo se tiene que:
(1): $y = log_a x$ Que también se puede escribir como $x = a^y$.
Entonces teniendo esto en cuenta se puede escribir que:
(2): $log_b a^y = log_b x$
Lo que por propiedad de logaritmos se va a tener:
(3): $log_b a^y = y * log_b a$
Teniendo (2) y (3) en cuenta:
$y * log_b a = log_b x$
Si se despeja la Y entonces:
$y = log_b x / log_b a$
Y por (1) se tiene que:
$log_a x = log_b x / log_b a$
Como se quería demostrar.

Dado que $log_a x = log_b x / log_b a$ y teniendo en cuenta la propiedad de logaritmo $1/log_b a = log_a b$

$$log_b x / log_b a = 1 / log_b a * log_b x$$
$$log_b x / log_b a = log_a b * log_b x$$

Por lo tanto $log_a x = log_a b * log_b x$

5. Una fuente F tiene 6 símbolos con probabilidades: p1=0.4, p2=0.2, p3=0.15, p4=0.1, p5=0.1, p6=0.05. Calcula la información individual y la entropía de la fuente.

$I_{p1} = log_2(\frac{1}{0,4})$ = 1,321

$I_{p2} = log_2(\frac{1}{0,2})$ = 2,321

$I_{p3} = log_2(\frac{1}{0,15})$ = 2,736

$I_{p4} = I_{p5} = log_2(\frac{1}{0,1})$ = 3,321

$I_{p6} = log_2(\frac{1}{0,05})$ = 4,321

$\sum_{i=1}^{q} p_i * I_i = 0,4 * 1,321 + 0,2 * 2,321 + 0,15 * 2,736 +$
$2 * (0,1 * 3,321) + 0,05 * 4,321 =$ <u>2.28325</u>

6. Justifica por qué una distribución uniforme maximiza la entropía. Compara un dado justo (6 caras, p=1/6) con un dado sesgado: p1=0.3, p2=0.25, p3=0.15, p4=0.15, p5=0.1, p6=0.05..

En el caso de que exista una baja entropía eso quiere decir que hay poca incertidumbre, y si hay mucha incertidumbre entonces existe una entropía alta. Entonces se dice que una distribución uniforme maximiza la entropía porque se está en el estado de máxima ignorancia.

Al tirar un dado:
p1=p2=p3=p4=p5=p6= ⅙

Además se tiene que:
I(x)=$- log_2 p(x)$

H(S)=∑p(x)*I(x)

Entonces teniendo en cuenta esto:
I(X)=$-\ log_2 1/6$=$log_2 6 \simeq\ 2,585$

H(S)= $\sum_{i=1}^{6}$ 1/6*$log_2 6$= 6*⅙ * $log_2 6$ = $log_2 6 \simeq\ 2,585$

Cara	p(x)	I(x)	p(x)*I(x)
1	0.3	1.737	0.521
2	0.25	2	0.5
3	0.15	2.737	0.411
4	0.15	2.737	0.411
5	0.1	3.322	0.332
6	0.05	4.322	0.216

Si se hace la sumatoria entonces el resultado de H(S)=2.391 comparándola con el resultado de 2.585 inicial, se puede decir que el sesgo baja la incertidumbre.

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

1. Se hacen tres grupos de cuatro monedas cada uno.
2. Se balancean dos grupos (primer balanceada).

Existen dos posibilidades: a) que pesen igual, b) que pesen diferente.
a) Si pesan igual, entonces la moneda que buscamos está en el tercer grupo; el que se apartó, por lo que de este último grupo se toman 3 monedas y se balancean contra 3 monedas de cualquiera de los 2 primeros grupos (segunda balanceada).
De nuevo existen dos posibilidades: I) que pesen igual, II) que pesen diferente.
I) Si pesan igual, entonces la moneda que buscamos es la que se apartó del tercer grupo y lo único que hay que determinar es si es liviana o pesada, para ello se balancea contra cualquiera de las 11 monedas (tercer balanceada)
II) Si pesan diferente, sabremos dos cosas, que la que buscamos está entre las tres del tercer grupo y además si es liviana o pesada ya que si el plato subió, entonces la que buscamos es liviana, pero si la balanza bajó, entonces la que buscamos es pesada, por lo tanto se toman 2 de las tres sospechosas y se balancean entre sí

(tercer balanceada), a partir de este resultado se deducirá cuál de las tres es la que se busca porque ya sabemos si es liviana o pesada.
b) Si en la primer balanceada el grupo 1 y 2 pesan diferente, entonces deduciremos que hay 4 monedas que pesan igual; las del grupo 3, llamémosles “pesadas”, “livianas” e “iguales”, así que la siguiente comparación se formará de la siguiente manera: Plato 1: Tres “pesadas” y una “liviana” y Plato 2: Tres “iguales” y una “pesada”, de tal forma que tres “livianas” quedarán fuera de la comparación (segunda balanceada).
Ahora existen tres probabilidades: I) que pesen igual, II) que el plato 1 suba, III) que el plato 1 baje.
I) Si pesan igual, entonces la que buscamos es liviana y está en las tres que se apartaron, por lo que se procede a balancear 2 de ellas (tercer balanceada), a partir de este resultado se deducirá cuál de las tres es la que se busca.
II) Si el plato 1 sube, solamente tendremos dos monedas sospechosas, la “liviana” del plato 1 y la “pesada” del plato 2, por lo que se procede a balancear cualquiera de estas dos sospechosas con cualquiera de las otras 10 (tercer balanceada), a partir de este resultado se deducirá cuál de las dos es la que se busca.
III) Si el plato 1 baja, entonces deduciremos que la que buscamos es pesada y se encuentra entre las tres “pesadas” del plato 1, así que se toman 2 de ellas y se balancean entre sí (tercer balanceada), a partir de este resultado se deducirá cuál de las tres es la que se busca.

8. Calcula la información de una letra al azar de un alfabeto de 28 símbolos (incluyendo ñ). Luego, de pares y ternas de letras.

Recordando que I(x)=$-\ log_2 p(x)$ se tiene que:

x=28 -> p(x)=1/28 -> I(x)=$log_2(28)$ =4.807

y=28^2 -> p(y)=1/784 -> I(y)=$log_2(784)$ =9.615

z=28^3 -> p(z)=1/21952 -> I(z)=$log_2(21952)$ =14.422

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

En la forma castellana de escritura cotidiana las cifras serán menores, ya que en el idioma, las letras no son equiprobables, algunas aparecen más que otras como la “e” en comparación a la “k”, y además existen dependencias entre letras como la “q” con la “u”.

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

Propiedades:

1- Un evento seguro no aporta información.

Por ejemplo: El sol siempre sale por el este, siendo este evento de probabilidad 1, lo que no nos da información (I_i = 0).

2- Cuanto menor es la probabilidad de un evento, mayor es la cantidad de información que aporta.

Por ejemplo: Si saco una carta al azar de una baraja española:

Probabilidad de sacar un as de espadas P = 1/40

I = $log_2(\frac{1}{\frac{1}{40}})$ = 5,32 bits

Probabilidad de sacar cualquier carta P = 1

I = 0 bits

Saber que salió una carta cualquiera no informa nada; saber que fue un as de espadas informa mucho más.

3- La cantidad de información siempre es mayor o igual que cero.

I(x) ≥ 0

11. Dada una variedad V = 2000 sucesos encontrar la base b óptima de representación que hace mínima la siguiente relación número más corto, es decir que minimice I . b (cantidad de información por base).

V=2000

I=$log_b(V)$

Luego para I*b= $log_b(V)$* b =$\frac{ln(V)}{ln(b)}$* b=$\frac{b}{ln(b)}$* $ln(V)$

Como $ln(V)$ es constante, entonces se minimiza f(b)= $\frac{b}{ln(b)}$

Para minimizar se hace f'(b)=0

f'(b)=$\frac{ln(b)-1}{ln^2(b)}$ y para qué esto dé 0 entonces ln(b)=1.

Lo que significa que la base óptima que mínima I*b es b=e≃2.718, pero como b tiene que ser un entero, la base óptima más cercana es 2 o 3, entonces calculando:

b=2, I= $log_2(2000)$≃10.963 → 2*10.963 ≃ 21.93

b=3, I=$log_2(2000)$≃6.918 → 3*6.918 ≃ 20.75

Por lo tanto, la base óptima es b=3.

12. Para el texto contenido en el código QR de la parte superior de la página encuentre la entropía de orden cero o independiente tanto de lo relevado en el QR como en el texto correcto. ¿Qué conclusiones saca?
En ambos casos su entropía es H = 3,9601
La conclusión es que por más que los símbolos estén desordenados, las probabilidades de ocurrencia de todos los símbolos se mantienen igual, por ende la incertidumbre es la misma en ambos, por lo cual la entropía es igual en ambos casos.

Código en python para calcular la entropía:

```
import math
from collections import Counter

def calcular_entropia(texto):
    # Contar frecuencia de cada símbolo
    frecuencias = Counter(texto)
    total = len(texto)
    # Calcular probabilidades
    probabilidades = [freq / total for freq in frecuencias.values()]
    # Calcular entropía
    entropia = -sum(p * math.log2(p) for p in probabilidades)
    return entropia
texto = "texto del qr aca"
print("Texto:", texto)
print("Entropía:", calcular_entropia(texto))
```