PRÁCTICO 1 TEORÍA DE LA INFORMACIÓN

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?

Calculamos la cantidad de información para la primera imagen:

Total de píxeles: 2560 pixeles x 1440 pixeles = 3.686.400 píxeles

1 pixel → 30 bits

3.686.400 píxeles → $\frac{3.686.400 \; píxeles \; x \; 30 \; bits}{1 \; pixel} = \; 110.592.000 \; bits \; de \; información.$

En bytes = $\frac{110.592.000 \; bits}{8 \; Bytes} = 13.824.000 \; bytes \; de \; información$

Luego para la otra imagen:

Total de píxeles: 7680 píxeles x 4320 píxeles = 33.177.600 píxeles
Bits por píxel: 12 bits/canal × 3 canales = 36 bits/píxel.

1 píxel → 36 bits

33.177.600 píxeles → $\frac{33.177.600 \; píxeles \; x \; 36 \; bits}{1 \; pixel}$ = $1.194.393.600 \; bits \; de \; información.$

En bytes = $\frac{1.194.393.600 \; bits}{8 \; Bytes} = \; 149.299.200 \; Bytes \; de \; información.$

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.

100 palabras tomadas
vocabulario de 20000 palabras

Calculamos la entropía del vocabulario: $log_2(20000)$ = 14,2877 bits de información

Para 1000 palabras independientes

1000 x 14,2877 bits = 14287,7 bits de información

Para la imagen de información
Total de píxeles: 640 píxeles x 480 píxeles = 307.200 píxeles

1 píxel → 24 bits
307.200 píxeles → $\frac{307.200\ píxeles\ x\ 24\ bits}{1\ píxel}$= 7.372.800 bits de información.

Luego para saber si una imagen vale más que 1000 palabras hacemos:

$$\frac{H_{imagen}}{H_{palabra}} = \frac{7.372.800\ bits\ de\ información}{14287,7\ bits\ de\ información} = 516.$$

$$\frac{H_{palabra}}{H_{imagen}} = \frac{14287,7\ bits\ de\ información}{7.372.800\ bits\ de\ información} = 0,0019.$$

Por lo tanto la imagen contiene 516 veces más de información.

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.

Para 200 caracteres tomados al azar de un alfabeto de 32 símbolos y de 64 símbolos

$H_{32} = Log_2(32) = 5\ bits$
$H_{64} = Log_2(64) = 6\ bits$

$I_{200,32} = 5\ x\ 200 = 1000\ bits\ de\ información$

$I_{200,64} = 6\ x\ 200 = 1200\ bits\ de\ información$

Para 400 caracteres tomados al azar de un alfabeto de 32 símbolos y de 64 símbolos

$I_{400,32} = 5\ x\ 400 = 2000\ bits\ de\ información$

$I_{400,64} = 6\ x\ 400 = 2400\ bits\ de\ información$

Duplicar el tamaño del alfabeto no duplica la información, sino que la aumenta en 1 bit por carácter. Duplicar la longitud del mensaje sí duplica la información total.

4. Demostrar las siguientes igualdades:

Para $a,b > 0$ con $a \neq 1$, $b \neq 1$ y $x > 0$.

Sea y = log_a(x). Por definición, a^y = x.

Tomamos logaritmo en base b a ambos lados:

log_b(a^y) = log_b(x)

Usando la propiedad log_b(u^k) = k * log_b(u):

y * log_b(a) = log_b(x)

Despejando y:

y = log_b(x) / log_b(a)

Pero y = log_a(x). Luego

log_a(x) = log_b(x) / log_b(a)

Además,

log_a(b) = 1 / log_b(a)

Entonces

log_b(x) / log_b(a) = log_b(x) * log_a(b)

Queda demostrada la igualdad:

log_a(x) = 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.

I(pi) = -log_2(pi)

I(p1) = 1,32
I(p2) = 2,32
I(p3) = 2,74
I(p4) = 3,32

I(p5) = 3,32
I(p6) = 4,32

H(F) = 0,4 * 1,32 + 0,2 * 2,32 + 0,15 * 2,74 + 0,1 * 2,32 * 2 + 0,05 * 4,32 = 2,083

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.

La entropía mide la incertidumbre de una variable aleatoria. Una distribución uniforme representa el estado de máxima ignorancia, todas las opciones son igualmente probables y no existe ninguna pista o sesgo que ayude a predecir el resultado. Esta es la situación más incierta posible.

La justificación formal se obtiene usando la Desigualdad de Gibbs, que establece que para dos distribuciones de probabilidad p_i y q_i :

$$\sum_{i=1}^{q} p_i \, log \frac{p_i}{q_i} >= 0$$

con igualdad sólo si $p_i = q_i$ para todo i .

Si tomamos como distribución de referencia la uniforme $q_i = \frac{1}{q}$, y sustituimos :

$$\sum_{i=1}^{q} p_i \, log \frac{p_i}{1/q} >= 0$$

$$\sum_{i=1}^{q} p_i \, log\,(q * p_i) >= 0$$

$$\sum_{i=1}^{q} p_i \, log\,(q) + \sum_{i=1}^{q} p_i \, log\,(p_i) >= 0$$

Luego de aquí se deduce que:

$$\sum_{i=1}^{q} p_i \, log\,(p_i) = -H(S) \text{ y } \sum_{i=1}^{q} p_i \, log\,(q) = log\,(q) * \sum_{i=1}^{q} p_i = log(q)$$

Por lo tanto:

$log(q) - H(S)$ >= 0

$H(S)$ <= $log(q)$

Es decir, la entropía nunca puede superar $log(q)$, y este valor máximo se alcanza únicamente cuando $p_i = \frac{1}{q}$ para todo i, es decir, cuando la distribución es uniforme.

Comparación de dados
$H = -\Sigma\, p(x) \log_2 p(x)$

Para un dado justo:

- Cada cara tiene probabilidad p(x) = 1/6
- Hay 6 posibles resultados (x = 1, 2, 3, 4, 5, 6)

Sustituyendo en la fórmula:

$$H = -6 \times (1/6) \times \log_2(1/6)$$

Calculando $\log_2(1/6)$:

- $\log_2(1/6) = \log_2(1) - \log_2(6) = 0 - \log_2(6)$
- $\log_2(6) \approx 2.585$

Por lo tanto:

- $\log_2(1/6) \approx -2.585$

Sustituyendo: **$H = -6 \times (1/6) \times (-2.585)$ $H = 6 \times (1/6) \times 2.585$ $H = 2.585$ bits**

La entropía de un dado justo de 6 caras es **2.585 bits**.

Para un dado sesgado con las siguientes probabilidades:

- $p_1 = 0.30$
- $p_2 = 0.25$
- $p_3 = 0.15$
- $p_4 = 0.15$
- $p_5 = 0.10$
- $p_6 = 0.05$

Calculando cada término:

1. **$p_1 \log_2 p_1 = -0.30 \times \log_2(0.30)$**
    - $\log_2(0.30) \approx -1.737$
    - Término 1: $-0.30 \times (-1.737) = 0.521$

2. **$p_2 \log_2 p_2 = -0.25 \times \log_2(0.25)$**
   - $\log_2(0.25) = -2$
   - Término 2: $-0.25 \times (-2) = 0.500$
3. **$p_3 \log_2 p_3 = -0.15 \times \log_2(0.15)$**
   - $\log_2(0.15) \approx -2.737$
   - Término 3: $-0.15 \times (-2.737) = 0.411$
4. **$p_4 \log_2 p_4 = -0.15 \times \log_2(0.15)$**
   - Término 4: 0.411 (igual que el anterior)
5. **$p_5 \log_2 p_5 = -0.10 \times \log_2(0.10)$**
   - $\log_2(0.10) \approx -3.322$
   - Término 5: $-0.10 \times (-3.322) = 0.332$
6. **$p_6 \log_2 p_6 = -0.05 \times \log_2(0.05)$**
   - $\log_2(0.05) \approx -4.322$
   - Término 6: $-0.05 \times (-4.322) = 0.216$

Sumando todos los términos: **$H = 0.521 + 0.500 + 0.411 + 0.411 + 0.332 + 0.216 = 2.391$ bits**

La entropía del dado sesgado es **2.391 bits**.

El cálculo muestra que la entropía del dado justo ($H \approx 2.585$ bits) es mayor que la del dado sesgado ($H \approx 2.391$ bits). Esto confirma lo que predice la teoría: la distribución uniforme maximiza la incertidumbre. En el caso del dado justo, todos los resultados son igualmente probables, por lo que cada lanzamiento es completamente impredecible y aporta la máxima cantidad de información posible. En cambio, el dado sesgado introduce conocimiento previo sobre el sistema (algunas caras son más probables que otras), lo que reduce la incertidumbre y, en consecuencia, la entropía. Esta diferencia refleja que la entropía mide el grado de impredecibilidad: cuanto más uniforme es la distribución, mayor es la “sorpresa” promedio que obtenemos de cada observación.

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

La moneda defectuosa puede ser cualquiera de las 12.Además, esta moneda defectuosa puede ser más pesada o más liviana.Por lo tanto, el número total de estados posibles es:

$$12(\text{monedas}) \times 2(\text{tipos de defecto}) = 24 \text{ estados posibles}$$

Cada pesada con una balanza de dos platillos tiene 3 resultados posibles:

1. El platillo izquierdo es más pesado.
2. El platillo derecho es más pesado.
3. Equilibrio.

La cantidad máxima de información (en bits) que puede proporcionar una pesada es:

$$\log_2(3) \approx 1.58 \text{bits}$$

La incertidumbre inicial (entropía) es el logaritmo de la cantidad de estados posibles:

$$H=\log_2(24)\approx 4.58 \text{bits}$$

El número mínimo de pesadas necesarias teóricamente es:

$$\frac{log_2(24)}{log_2(3)} = \frac{4.58}{1.58} = 2.92$$

Como el número de pesadas debe ser un entero, se necesitan 3 pesadas para distinguir entre los 24 estados posibles.

Cada pesada tiene 3 resultados posibles: izquierda más pesada, derecha más pesada o equilibrio.

Primera Pesada:Dividir las 12 monedas en 3 grupos de 4.
Grupo A: Monedas 1, 2, 3, 4
Grupo B: Monedas 5, 6, 7, 8
Grupo C: Monedas 9, 10, 11, 12

Primera Pesada: Pesar Grupo A vs. Grupo B (1,2,3,4 vs. 5,6,7,8). Hay 3 posibles resultados:
Caso 1: A = B (Equilibrio)

- La moneda diferente está en el Grupo C (9,10,11,12).
- Las monedas de A y B son todas normales.

Caso 2: A > B (A más pesado que B)

- La moneda diferente está en A o B.
- Si está en A, es más pesada; si está en B, es más liviana.

Caso 3: A < B (A más liviano que B)

- La moneda diferente está en A o B.
- Si está en A, es más liviana; si está en B, es más pesada.

Segunda Pesada (Depende del resultado de la primera):

- Si Primera Pesada fue A = B (la falsa está en C):Ya sabemos que A y B son normales.
- Pesar 3 monedas de C (9,10,11) vs. 3 normales (ej. 1,2,3).
    - Si (9,10,11) = (1,2,3):La moneda 12 es diferente. Ir a Tercera Pesada para saber si es más pesada o liviana (pesar 12 vs. cualquier normal).
    - Si (9,10,11) > (1,2,3):Una de 9,10,11 es más pesada.
    - Si (9,10,11) < (1,2,3):Una de 9,10,11 es más liviana.
- Si Primera Pesada fue A > B (la falsa está en A o B, y es más pesada en A o más liviana en B):
    - Pesar: (1,2,5) vs. (3,6,9): Donde 9 es una moneda normal (del grupo C, que ya sabemos que es normal porque A y B no equilibraron).
        - Si (1,2,5) = (3,6,9):La falsa es 4 (más pesada) o 7 u 8 (más livianas).

        - Si (1,2,5) > (3,6,9):Puede ser: 1 o 2 más pesadas, o 6 más liviana.
        - Si (1,2,5) < (3,6,9):Puede ser: 3 más pesada, o 5 más liviana.

- Si Primera Pesada fue A < B (simétrico a A > B):La falsa es más liviana en A o más pesada en B.
    - Pesar: (1,2,5) vs. (3,6,9) (igual que arriba, pero interpretando inversamente).
        - Si (1,2,5) = (3,6,9):La falsa es 4 (más liviana) o 7 u 8 (más pesadas).
        - Si (1,2,5) > (3,6,9):Puede ser 3 más liviana, o 5 más pesada.
        - Si (1,2,5) < (3,6,9):Puede ser 1 o 2 más livianas, o 6 más pesada.

Tercera Pesada: En todos los casos, la tercera pesada es una comparación simple entre 2 monedas (o una contra una normal) que permite identificar la moneda falsa y su tipo.

Ejemplo para el caso A = B y (9,10,11) > (1,2,3):Sabemos que una de 9,10,11 es más pesada.

- Pesar 9 vs. 10:
    - Si 9 = 10 → 11 es más pesada.
    - Si 9 > 10 → 9 es más pesada.
    - Si 10 > 9 → 10 es más pesada.

Ejemplo para el caso A > B y (1,2,5) = (3,6,9):La falsa es 4 (más pesada) o 7 u 8 (más livianas).

- Pesar 7 vs. 8:
    - Si 7 = 8 → 4 es más pesada.
    - Si 7 < 8 → 7 es más liviana.
    - Si 7 > 8 → 8 es más liviana.

Ejemplo para el caso A > B y (1,2,5) > (3,6,9):Puede ser 1 o 2 más pesadas, o 6 más liviana.
- Pesar 1 vs. 2:
    - Si 1 = 2 → 6 es más liviana.
    - Si 1 > 2 → 1 es más pesada.
    - Si 2 > 1 → 2 es más pesada.

8. Calcula la información de una letra al azar de un alfabeto de 27 símbolos (incluyendo ñ). Luego, de pares y ternas de letras.

1. Información de una letra (alfabeto de 27 símbolos equiprobables):

Si cada letra es igualmente probable: $p = 1/27$
La información de un suceso (según Shannon) es:

$$I(1) = -log_2(p) = -log_2(1/27) = log_2(27) = 4.75488\, bits$$

Por lo tanto, una letra al azar aporta aproximadamente **4.755 bits de información.**

2. Información de pares de letras

Un par de letras tiene:

$27^2 = 729$ posibilidades igualmente probables

La probabilidad de cada par es: $p = 1/729$

La información de cada par de letras es:

$I(2) = -log_2(p) = -log_2(1/729) = log_2(729) = 9.50977\, bits$

Un par de letras aporta aproximadamente **9.51 bits de información.**

3. Información de ternas de letras

Una terna de letras tiene:

$27^3 = 19683$ posibilidades igualmente probables

La probabilidad de cada terna es: $p = 1/19683$

La información de cada terna es:

$I(3) = -log_2(p) = -log_2(1/19683) = log_2(19683) = 14.26466\, bits$

Una terna de letras aporta aproximadamente **14.265 bits de información.**

En conclusión podemos ver que:

1. Una letra: $log_2(27)$
2. Dos letras: $2 * log_2(27)$
3. Tres letras: $3 * log_2(27)$

En general, para n letras equiprobables:

$$I(n) = n * log_2(27)$$

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 escritura cotidiana del castellano los valores de información serían menores que los calculados para el caso equiprobable. Esto se debe a que las letras no aparecen con la misma frecuencia: por ejemplo, la “e” y la “a” son mucho más comunes que la “k” o la “w”. Además, ciertas combinaciones (pares o ternas) son muy probables (“que”, “de”, “la”), mientras que otras prácticamente nunca ocurren.

Como la entropía se maximiza cuando todos los símbolos son igualmente probables, la presencia de estas desigualdades de frecuencia reduce la incertidumbre promedio. Por eso, en la práctica, la cantidad de bits necesaria para codificar letras, pares o ternas en castellano es menor que la obtenida bajo la hipótesis de equiprobabilidad.

Por lo tanto, sí existe diferencia, y se debe a la estructura estadística propia del idioma (frecuencia desigual de letras y restricciones en las combinaciones posibles).

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

a. La cantidad de información (I) es siempre positiva (mayor o igual a cero).

Nunca obtenemos “información negativa”. El mínimo de información es 0, lo que ocurre cuando un evento es seguro es decir probabilidad = 1.

Ejemplo: Si tiramos un dado trucado que siempre sale “6”, no hay sorpresa (probabilidad = 1), entonces la información de que salga “6” es 0 bits.

b. La cantidad de información (I) de un evento, es inversamente proporcional a la probabilidad de la ocurrencia del mismo.

   Cuanto menos probable es un evento, más información obtenemos al observarlo.

   Ejemplo: Si lanzamos una moneda, la probabilidad de obtener “cara” es de 0,5. Como es un resultado bastante probable, la información que nos aporta no es muy grande. En cambio, si tiramos un dado y aparece un “6”, ese suceso es menos probable (1 de cada 6 intentos). Por eso, recibir la noticia de que salió “6” nos aporta más información que enterarnos de que salió “cara”.

c. La cantidad de información (I) aumenta con la cantidad de posibles eventos o mensajes.

   Si tenemos más resultados posibles, la incertidumbre inicial es mayor, y por lo tanto la información que aporta conocer el resultado también lo es.

   Ejemplo: Si lanzamos una moneda, solo hay dos posibilidades (cara o cruz), así que la cantidad de información es pequeña. En cambio, si tiramos un dado, ya tenemos seis posibles resultados, lo que implica más incertidumbre y, por lo tanto, más información al conocer el resultado. Si pensamos en una ruleta con muchas casillas, la cantidad de opciones es todavía mayor y la información que aporta saber en cuál cayó se incrementa aún más.

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

   Si dos sucesos son independientes, la información total es la suma de las informaciones de cada uno.

   Ejemplo: Si tiramos una moneda y también un dado, son dos sucesos que no dependen uno del otro. Saber el resultado de la moneda nos da cierta información, y saber el del dado nos da otra. La información total de hacer ambas cosas es la suma de las dos, porque cada resultado aporta por separado.

e. La cantidad de información promedio, (que recibe el nombre de entropía, H) se maximiza ante sucesos equiprobables.

La incertidumbre es mayor cuando todos los eventos son igual de probables. Si las probabilidades están desbalanceadas, la entropía baja.

Ejemplo: Al jugar piedra, papel o tijera con un rival que elige al azar, no podemos anticipar qué va a sacar: las tres opciones son igual de probables y la incertidumbre es máxima. En cambio, sí sabemos que siempre juega “piedra”, ya no hay sorpresa y la información que obtenemos al ver su jugada es mínima.

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).

Problema:

- Número de sucesos: V = 2000
- Información (en base b) = $I = log_b(V) = log_b(2000)$
- Debemos minimizar la función: $f(b) = I * b = b * log_b(2000)$

Expresamos $log_b(2000)$ en términos de logaritmo natural (o base 10)

Usamos fórmula de cambio de base:

$$log_b(2000) = \frac{ln(2000)}{ln(b)}$$

Entonces, la función a minimizar es:

$$f(b) = b * \frac{ln(2000)}{ln(b)} = ln(2000) * \frac{b}{ln(b)}$$

Como ln(2000) es una constante, minimizar f(b) es equivalente a minimizar:

$$g(b) = \frac{b}{ln(b)}$$

Ahora debemos encontrar el mínimo de g(b):

Calculamos la derivada de g(b) con respecto a b, y la igualamos a cero.

Usamos la regla del cociente:

$$g'(b) = \frac{1*ln(b)-b*\frac{1}{b}}{[ln(b)]^2} = \frac{ln(b)-1}{[ln(b)]^2}$$

Igualamos a cero para encontrar puntos críticos:

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

Esto implica:

$$ln(b) - 1 = 0 \Rightarrow ln(b) = 1 \Rightarrow b = e$$

Donde e≈2.71828 es la base de los logaritmos naturales.

Verificamos que es un mínimo utilizando la segunda derivada de g(b):

$$g''(b) = \frac{-ln(b)+2}{b*ln^3(b)}$$

En b=e se tiene ln(b)=1, por lo que:

$$g''(b) = \frac{-1+2}{e*1} = 1/e > 0$$

Por lo tanto b es mínimo.
El valor mínimo es: $f(e) = I * e = e * log_e(2000) = e * ln(2000) = 20.66139$

La base óptima que minimiza $f(b) = I * b$ es $b = 2.71828$
Sin embargo, como las bases deben ser enteras, debemos probar con las bases enteras más cercanas a e: b=2 y b=3.
Calculamos $f(b) = b * log_b(2000)$ para b=2 y b=3.

- Para b=2:
$$f(2) = 2 * log_2(2000) = 21.93156$$
- Para b=3:
$$f(3) = 3 * log_3(2000) = 20.75591$$

Comparamos y llegamos a la conclusión que la base entera óptima es b = 3, ya que da un valor menor para f(b).

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?

```
import math
from collections import Counter

def calcular_entropia(texto):
   # Contar la frecuencia de cada carácter
   contador = Counter(texto)
   longitud = len(texto)
   entropia = 0

   for freq in contador.values():
       p = freq / longitud
       entropia -= p * math.log2(p)

   return entropia

# Texto contenido en el código QR
texto = "Sgeun un etsduio de una uivenrsdiad ignlsea, no ipmotra el odren
en el que las ltears etsan ersciats, la uicna csoa ipormtnate es que la
pmrirea y la utlima ltera esten ecsritas en la psiocion cocrrtea. El rsteo
peuden estar ttaolmntee mal y aun pordas lerelo sin pobrleams. Etso es
pquore no lemeos cada ltera por si msima preo la paalbra es un tdoo.
Pesornamelnte me preace icrneilbe..."
```

```python
# Texto correcto
texto1 = "Segun un estudio de una universidad inglesa, no importa el orden en el que las letras estan escritas, la unica cosa importante es que la primera y la ultima letra esten escritas en la posicion correcta. El resto pueden estar totalmente mal y aun podras leerlo sin problemas. Esto es porque no leemos cada letra por si misma pero la palabra es un todo. Personalmente me parece increible..."

print(f"Entropía del Texto contenido en el código QR: {calcular_entropia(texto):.4f} bits/caracter")

print(f"Entropía del Texto correcto: {calcular_entropia(texto1):.4f} bits/caracter")
```

Al calcular la entropía de orden cero del texto del QR y del texto correcto, vemos que dan el mismo valor. Esto pasa porque esta medida solo se fija en la frecuencia de cada letra y símbolo, no en el orden de las palabras o la legibilidad. Es decir, aunque el texto del QR esté desordenado la entropía no “nota” esa diferencia.

La entropía de orden cero sirve para medir la diversidad de caracteres en un texto, pero no para medir si un texto tiene sentido o es fácil de leer. Los textos que contienen las mismas letras y símbolos, aunque estén mezcladas, tendrán la misma cantidad de información a nivel de caracteres.

<u>**PRÁCTICO 2 CANAL DE INFORMACIÓN**</u>

1. Sea el siguiente canal:

	b_1	b_2	b_3
a_1	0,6	0,3	0,1
a_2	0,2	0,5	0,3
a_3	0,4	0,2	0,4

Calcular los valores de p(ai/bj) y las probabilidades de salida para el caso particular de p(a1)=0.5, p(a2)=0.25, p(a3)=0.25

Probabilidades de salida:

$$p(b_j) = \sum_i p(a_i) \cdot p(b_j|a_i)$$

$$p(b_1) = \sum_i p(a_i) \cdot p(b_1|a_i) = 0.5 * 0.6 + 0.25 * 0.2 + 0.25 * 0.4 = 0.45$$

$$p(b_2) = \sum_i p(a_i) \cdot p(b_2|a_i) = 0.5 * 0.3 + 0.25 * 0.5 + 0.25 * 0.2 = 0.325$$

$$p(b_3) = \sum_i p(a_i) \cdot p(b_3|a_i) = 0.5 * 0.1 + 0.25 * 0.3 + 0.25 * 0.4 = 0.225$$

Probabilidades inversas (Teorema de Bayes):

$$p(a_i|b_j) = \frac{p(a_i)\cdot p(b_j|a_i)}{p(b_j)}$$

Para b_1 donde $p(b_1) = 0.45$

$$p(a_1|b_1) = \frac{p(a_1)\cdot p(b_1|a_1)}{p(b_1)} = \frac{0.5*0.6}{0.45} = 0.6667$$

$$p(a_2|b_1) = \frac{p(a_2)\cdot p(b_1|a_2)}{p(b_1)} = \frac{0.25*0.2}{0.45} = 0.1111$$

$$p(a_3|b_1) = \frac{p(a_3)\cdot p(b_1|a_3)}{p(b_1)} = \frac{0.25*0.4}{0.45} = 0.2222$$

Para b_2 donde $p(b_2) = 0.325$

$$p(a_1|b_2) = \frac{p(a_1)\cdot p(b_2|a_1)}{p(b_2)} = \frac{0.5*0.3}{0.325} = 0.46153$$

$$p(a_2|b_2) = \frac{p(a_2)\cdot p(b_2|a_2)}{p(b_2)} = \frac{0.25*0.5}{0.325} = 0.38461$$

$$p(a_3|b_2) = \frac{p(a_3)\cdot p(b_2|a_3)}{p(b_2)} = \frac{0.25*0.2}{0.325} = 0.15384$$

Para b_3 donde $p(b_3) = 0.225$

$$p(a_1|b_3) = \frac{p(a_1)\cdot p(b_3|a_1)}{p(b_3)} = \frac{0.5*0.1}{0.225} = 0.2222$$

$$p(a_2|b_3) = \frac{p(a_2)\cdot p(b_3|a_2)}{p(b_3)} = \frac{0.25*0.3}{0.225} = 0.3333$$

$$p(a_3|b_3) = \frac{p(a_3)\cdot p(b_3|a_3)}{p(b_3)} = \frac{0.25*0.4}{0.225} = 0.4444$$

2. Considera un Canal Binario Asimétrico con las siguientes probabilidades:

- **p(a1)=0,4**
- **p(b1/a1)=4/5**
- **p(b1/a2)=1/4**

p(a1) = 0,4 => p(a2) = 0,6

p(b1/a1) = 0,8 => p(b2/a1) = 0,2

p(b1/a2)=0,25 => p(b2/a2) = 0,75

	b1	b2
a1 0,4	0,8	0,2
a2 0,6	0,25	0,75

a) Calcula las probabilidades condicionales hacia atrás y las probabilidades conjuntas.

Probabilidades conjuntas

- p(a1,b1) = p(a1) * p(b1|a1) = 0,4 * 0,8 = 0,32
- p(a1,b2) = p(a1) * p(b2|a1) = 0,4 * 0,2 = 0,08
- p(a2,b1) = p(a2) * p(b1|a2) = 0,6 * 0,25 = 0,15
- p(a2,b2) = p(a2) * p(b2|a2) = 0,6 * 0,75 = 0,45

Probabilidades de salida

p(b1) = p(a1,b1) + p(a2,b1) = 0,32 + 0,15 = 0,47

p(b2) = 1 − p(b1) = 0.53

Probabilidades condicionales hacia atrás

- p(a1|b1) = p(a1,b1) / p(b1) = 0,32 / 0,47 = 0,68
- p(a1|b2) = p(a1,b2) / p(b2) = 0,08 / 0,53 = 0,15
- p(a2|b1) = p(a2,b1) / p(b1) = 0,15 / 0,47 = 0,32
- p(a2|b2) = p(a2,b2) / p(b2) = 0,45 / 0,53 = 0,85

b) Obtén la entropía del emisor, y las entropías condicionales de la fuente para cualquier símbolo de salida.

$H(A) = -[p(a_1) \cdot log_2(p(a_1)) + p(a_2) \cdot log_2(p(a_2))] = 0,97$

$H(A|b_1) = -[p(a_1|b_1) \cdot log_2(p(a_1|b_1)) + p(a_2|b_1) \cdot log_2(p(a_2|b_1))] = 0,9$

$H(A|b_2) = -[p(a_1|b_2) \cdot log_2(p(a_1|b_2)) + p(a_2|b_2) \cdot log_2(p(a_2|b_2))] = 0,61$

$H(A|B) = [p(b_1) \cdot H(A|b_1)) + p(b_2) \cdot H(A|b_2))] = 0,75$

c) Calcula la información mutua y la capacidad del canal.

$I(A,B) = H(A) - H(A|B) = 0,97 - 0,75 = 0,22$

3. Considera un canal determinista con cuatro símbolos de entrada (a_1 ,a_2 ,a_3 ,a_4) y cuatro símbolos de salida (b_1 ,b_2 ,b_3 ,b_4), donde cada símbolo de entrada se corresponde exclusivamente con un símbolo de salida. Define las probabilidades para cada símbolo de entrada y calcula la información mutua.

Definimos las probabilidades para los símbolos de entrada:

$P(a_1)$ = 0.4

$P(a_2)$ = 0.3

$P(a_3)$ = 0.2

$P(a_4)$ = 0.1

Verificamos que la sumatoria de las probabilidades: $\sum P(a_i)$ = 0.4 + 0.3 + 0.2 + 0.1 = 1

Como es un canal determinista $P(b_j)$ = $P(a_j)$:

$P(b_1)$ = $P(a_1)$ = 0.4

$P(b_2)$ = $P(a_2)$ = 0.3

$P(b_3)$ = $P(a_3)$ = 0.2

$P(b_4) = P(a_4)$ = 0.1

Ahora realizamos el cálculo de la entropía de la fuente.

H(X) = $-\sum_{i=1}^{4} P(a_i) * log_2(P(a_i))$

H(X)= $-[P(a_1) * log_2 P(a_1) + P(a_2) * log_2 P(a_2) + P(a_3) * log_2 P(a_3) + P(a_4) * log_2 P(a_4)]$

H(X) = $-[0.4 * log_2(0.4) + 0.3 * log_2(0.3) + 0.2 * log_2(0.2) + 0.1 * log_2(0.1)]$

H(X) = $-[0.4 * (-1.322) + 0.3 * (-1.737) + 0.2 * (-2.322) + 0.1 * (-3.322)]$

H(X) = $-[(-0.529) + (-0.521) + (-0.464) + (-0.332)]$

H(X) = $-\ [-\ 1.846]\ =\ 1.846$

En un canal determinista, al conocer la entrada X, la salida Y está completamente determinada. Por tanto, para cada a_1:

H(Y|X =a_1) = $-\sum_{i=1}^{4} P(b_j|a_i)\ log_2(P(b_j|a_i))\ =\ 0$

Ya que la $P(b_i|a_i)$ = 1 y $log_2(1)\ =\ 0$ y $P(b_j|a_i)$ = 0 para $j\ \neq\ i$ y 0 * $log_2(0)\ =\ 0$ por convención.

Luego H(Y|X =a_1) = $-\sum_{i=1}^{4}(a_i)\ H(Y|X\ \ = a_1)$

= 0.4 * 0 + 0.3 * 0 + 0.2 * 0 + 0.1 * 0 = 0

Luego utilizando la fórmula:

I(X;Y) = H(Y) - H(Y|X)

H(Y) = H(X) = 1.846 bits por que $P(b_j)$ = $P(a_j)$

H(Y|X) = 0 bits.

Entonces: I(X;Y) = H(Y) - H(Y|X) = 1.846 - 0 = 1.846 bits

La información mutua I(X;Y)=1.846 bits indica que, en promedio, cada símbolo transmitido lleva 1.846 bits de información útil. Dado que el canal es determinista, no hay pérdida de información, por lo que la información mutua coincide con la entropía de la fuente H(X). Esto confirma que el canal opera de manera ideal para la distribución de entrada dada.

4. Dada la siguiente matriz de canal, obtén las respectivas codificaciones de la fuente. p(a)=1/3, p(b)=1/6, p(c)=1/2

	a	b	c
a	2/5	2/5	1/5
b	2/5	1/5	2/5
c	1/4	1/2	1/4

Primero calculamos las probabilidades de salida $p(y_j)$.

Tenemos tres símbolos de salida: a, b, c. Usamos la ley de la probabilidad total:

$$p(y_j) = \sum_i p(x_i) \cdot p(y_j|x_i)$$

$p(a) = p(a|x = a) \cdot p(a) + p(a|x = b) \cdot p(b) + p(a|x = c) \cdot p(c) = 2/5 \cdot 1/3 + 2/5 \cdot 1/6 + 1/4 \cdot 1/2$

$$p(a) = 2/15 + 1/15 + 1/8 = 13/40$$

$p(b) = 2/5 \cdot 1/3 + 1/5 \cdot 1/6 + 1/2 \cdot 1/2 = 5/12$

$p(c) = 1/5 \cdot 1/3 + 2/5 \cdot 1/6 + 1/4 \cdot 1/2 = 31/120$

Verificamos que 13/40 + 5/12 + 31/120 = 1

Ahora calculamos probabilidades a posteriori $p(x_i|y_j)$.

$$p(x_i|y_j) = \frac{p(x_i)\cdot p(y_j|x_i)}{p(y_j)}$$

Para $y = a$

$p(a|a) = \frac{p(a)\cdot p(a|a)}{p(a)} = \frac{(1/3)\cdot(2/5)}{13/40} = 16/39$

$p(b|a) = \frac{p(b)\cdot p(a|b)}{p(a)} = \frac{(1/6)\cdot(2/5)}{13/40} = 8/39$

$p(c|a) = \frac{p(c)\cdot p(a|c)}{p(a)} = \frac{(1/2)\cdot(1/4)}{13/40} = 5/13$

Para b

$p(a|b) = \frac{p(a)\cdot p(b|a)}{p(b)} = \frac{(1/3)\cdot(2/5)}{5/12} = 8/25$

$p(b|b) = \frac{p(b)\cdot p(b|b)}{p(b)} = \frac{(1/6)\cdot(1/5)}{5/12} = 2/25$

$p(c|b) = \frac{p(c)\cdot p(b|c)}{p(b)} = \frac{(1/2)\cdot(1/2)}{5/12} = 3/5$

Para c

$p(a|c) = \frac{p(a)\cdot p(c|a)}{p(c)} = \frac{(1/3)\cdot(1/5)}{31/120} = 8/31$

$p(b|c) = \frac{p(b)\cdot p(c|b)}{p(c)} = \frac{(1/6)\cdot(2/5)}{31/120} = 8/31$

$p(c|c) = \frac{p(c)\cdot p(c|c)}{p(c)} = \frac{(1/2)\cdot(1/4)}{31/120} = 15/31$

Estas probabilidades a posteriori son cruciales para la decodificación. Por ejemplo, para decodificar un símbolo recibido, se elige el x_i con la mayor $p(x_i|y_j)$.

Para y = a: el máximo $p(x_i|y_j)$ es $p(a|a) = 0.41$. Por lo tanto, si se recibe el símbolo ‘a’, se debe decodificar como el símbolo ‘a’. Cuando el receptor obtiene el símbolo 'a' la mejor decisión es interpretar que el símbolo original que se transmitió fue también una 'a'.

Para y = b: el máximo $p(x_i|y_j)$ es $p(c|b) = 0.6$. Cuando el receptor obtiene el símbolo 'b' la mejor decisión es interpretar que el símbolo original que se transmitió fue una 'c'.

Para y = c: el máximo $p(x_i|y_j)$ es $p(c|c) = 0.484$. Si se recibe el símbolo ‘c’, se debe decodificar como ‘c’.