Introducción a la Teoría de la Información

Primer parcial

23 de marzo de 2023

Problema 1

Se tiene un conjunto de probabilidad $\mathbf{p} = (p_1, p_2, ..., p_i, .., p_j, ...p_n)$.

(i) Comparar su entropía con la del conjunto $\mathbf{p'} = (p_1, ..., \frac{p_i+p_j}{2}, ..., \frac{p_i+p_j}{2}, ..., p_n)$.

(ii) Justificar la desigualdad.

(iii) Más en general, comparar $H(\mathbf{p})$ con $H(\mathbf{p''})$, siendo
$\mathbf{p''} = (p_1, ..., \lambda p_i + (1-\lambda)p_j, ..., (1-\lambda)p_i + \lambda p_j, ..., p_n)$ con $\lambda \in (0, 1)$.

(iv) ¿La entropía es cóncava o convexa? Vincular esta propiedad con el resultado de (iii).

Problema 2

Sea $\{X_n\}_{n\geq 1}$ un proceso de Markov estacionario definido sobre un alfabeto finito con distribución estacionaria π y matriz de probabilidades de transición P. Mostrar que la tasa de entropía del proceso es
$\mathcal{H} = -\sum_{i,j} \pi_i P_{ij} log P_{ij}$.

Problema 3

Hay un grupo de 16 personas y se sabe que una de ellas (y una sola) tiene COVID. Se quiere ubicar quién es.

(i) Dar un modelo de la fuente de información. ¿Cuántos bits de información implica saber quién es el infectado?

(ii) La primera estrategia que se le ocurre a alguien es ir haciendo los test de PCR uno por uno. ¿Cuántos análisis en media requiere encontrar al infectado de esa manera?

(iii) Proponer una estrategia mejor a partir de la idea de mezclar las muestras. ¿Cuál es la estrategia óptima y por qué se puede afirmar que lo es?

(iv) ¿Cómo cambia el problema si también pudiera no haber ninguna persona infectada? ¿Cuántos análisis son el mínimo ahora?

Solución problema 1.

$H(\mathbf{p})$ y $H(\mathbf{p'})$ difieren sólo en dos sumandos, así como $H(\mathbf{p})$ y $H(\mathbf{p''})$. Notar que (i) es un caso particular de (iii).

Se puede hacer (i) usando la desigualdad LogSum e (iii) a partir de que $x \log x$ es una función convexa.

(ii) Las mezclas convexas tienden a igualar probabilidades, por lo que aumentan la incertidumbre.

(iv) La entropía es cóncava. $H(\lambda\mathbf{p} + (1-\lambda)\mathbf{q}) \geq \lambda H(\mathbf{p}) + (1-\lambda)H(\mathbf{q})$

A partir de esto se puede responder todo el problema en forma alternativa, considerando $\mathbf{q} = (p_1, p_2, ...p_j, ..., p_i, ..., p_n)$, es decir que $\mathbf{q}$ es $\mathbf{p}$ con dos elementos permutados entre sí.

Solución problema 2. Está incluido el el teórico.

Solución problema 3.

Se tienen 16 frascos o tubos con líquido para analizar. (i) Los mensajes posibles son:

- 1 es quien da positivo
- 2 es quien da positivo
- 3 es quien da positivo
- ......

Cada uno tiene probabilidad $\frac{1}{16}$. $H(X) = 4bits$

(ii) Analizo uno por uno.

Llamo k a la cantidad de análisis hasta llegar a uno positivo.

- $P(k = 1) = \frac{1}{16}$
- $P(k = 2) = \frac{15}{16}\frac{1}{15} = \frac{1}{16}$ (que el primero sea negativo y el segundo positivo)

- $P(k=3) = \frac{15}{16}\frac{14}{15}\frac{1}{14} = \frac{1}{16}$ (que el primero y el segundo sean negativos y el tercero positivo)
- .... (la estructura es telescópica y siempre da $P(k=i) = \frac{1}{16}$

$E(k) = \frac{1}{16}\sum_{i=1}^{15} k = \frac{1}{16}\frac{15\times 16}{2} = \frac{15}{2}$

Es un problema igual al siguiente, más general: en una urna tenemos N-1 bolillas blancas y una roja. Vamos sacando SIN reposición. Queremos estimar en cuántas extracciones, en media, se saca la bolilla roja. Sea M un número entre 1 y N.

$P(k=M) = \Pi_{i=1}^{i=M-1}[\frac{N-i}{N-i+1}] \times \frac{1}{N-M+1} = \frac{1}{N}$

$E(k) = \frac{1}{N}\sum_{i=1}^{N} k = \frac{1}{N}\frac{(N+1)\times N}{2} = \frac{N+1}{2}$

En el caso de las bolillas se saca también la última; en el caso de los análisis no, por el planteo de las preguntas.

(iii) Se divide el conjunto a la mitad. Con un análisis se sabe en qué mitad está el infectado. Ese conjunto se divide nuevamente a la mitad. Cada paso tiene dos resultados posibes, equiprobables. Es equivalente a preguntas binarias. Cuando quedan dos personas, con un análisis se termina el proceso. Hay 4 pasos, correspondientes a los 4 bits.

(iv)En el caso de que puede haber ninguna persona infectada, es necesario hacer un análisis adicional.