Compresión de datos sin pérdida
Práctico 5

Ejercicio 1 *¿Cuánto dice un prefijo de F sobre X?*
Sea $\{X_i\}_{i\geq 0}$ un proceso de Markov *estacionario* de orden 1, sobre un alfabeto binario, con la siguiente matriz de probabilidades de transición

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

1. Calcular los primeros tres dígitos binarios (después de la coma) de $F(X^n)$ sabiendo que X^n tiene a 1010111 como prefijo.
2. ¿Cuántos bits de X^n se pueden deducir a partir de estos tres dígitos binarios?

Ejercicio 2 *¿Cuánto dice un prefijo de X sobre F?*
Sea $\{X_i\}_{i\geq 0}$ un proceso de Markov *estacionario* de orden 1, sobre un alfabeto binario, con la siguiente matriz de probabilidades de transición

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

1. Calcular $F(0111)$.
2. ¿Cuántos bits de la representación binaria de $F(x^n)$ pueden determinarse sabiendo que x^n tiene a 0111 como prefijo?

Sugerencia: Para obtener fácilmente la representación binaria de algunos números puede convenir tener en cuenta que

$$\frac{1}{3} = \sum_{i>0} 2^{-2i}, \quad \frac{1}{9} = 7\sum_{i>0} 2^{-6i}, \quad \frac{1}{27} = 9709\sum_{i>0} 2^{-18i}.$$

Ejercicio 3 *Codificación / Decodificación con precisión finita*
Una fuente emite símbolos i.i.d. del alfabeto $\mathcal{A} = \{0, 1, 2, 3, 4, 5\}$ con probabilidades

$$Q = (0{,}01,\ 0{,}08,\ 0{,}01,\ 0{,}80,\ 0{,}01,\ 0{,}09).$$

1. Codificar la secuencia $x^n = 3\,3\,1\,3\,3\,3\,3\,5\,3\,3$ con los parámetros $D = 10$, $J = 2$, $K = 2$.
2. Suponga que el decodificador recibe la secuencia de dígitos $2\,1\,9\,8\,3\,1\,4\,1\,5\,9\ldots$. Decodificar 10 símbolos usando los mismos parámetros que en la parte anterior.

Ejercicio 4 *El largo máximo*
Probar que en el codificador aritmético con precisión finita visto en el curso, el largo de código $L(x^n)$ está acotado por

$$L \leq nJ + 1.$$

Ejercicio 5 *Retardo*
Mostrar que si en la iteración i del codificador el dígito en la posición r (correspondiente a la potencia D^{-r}) del valor calculado de F_i es menor que $D-1$, donde $r < \tau_i$, entonces todos los dígitos en posiciones menores que r (más significativos) coinciden con los de W y, por lo tanto, pueden transmitirse al decodificador.

Ejercicio 6 *El ancho de los registros*
Sea $\{X_i\}_{i\geq 0}$ una secuencia de variables i.i.d. sobre un alfabeto finito $\mathcal{A} = \{1, 2, \ldots M\}$. Sea $\hat{p}(x^n)$ la distribución de probabilidad empírica sobre $\mathcal{A}$ inducida por la secuencia x^n, esto es, $\hat{p}(x^n)(a) = \frac{n_a}{n}$ para todo $a \in \mathcal{A}$, donde $n_a = |\{i = 1 \ldots n : x_i = a\}|$. Se propone el siguiente esquema de codificación para secuencias de largo n:

- Paso 1: Se describe el vector de probabilidad $\hat{p}(x^n)$ codificando $(M-1)$ contadores n_a, $a = 1 \ldots M-1$, cada uno con su representación binaria como entero sin signo de $\lceil \log(n+1) \rceil$ bits.
- Paso 2: Se codifica x^n con un codificador aritmético usando la distribución $\hat{p}(x^n)$ aproximada con J dígitos de precisión.

1. Probar que trabajando con registros de ancho $C \log n$ bits, con C constante suficientemente grande, (J y K de orden $\Omega(\log n)$), el largo de código asociado al paso 2, L_2, satisface
$$L_2(x^n) \leq -\log P_{ML}(x^n) + O(1)\,.$$
2. Concluir que la tasa de compresión media del código propuesto (incluyendo los pasos 1 y 2), satisface
$$\frac{1}{n} E\left[L(X^n)\right] \leq \frac{1}{n} H(X^n) + (M-1)\frac{\log n}{n} + O\left(\frac{1}{n}\right)\,.$$