Compresión de datos sin pérdida
Práctico 7

Ejercicio 1 *Esperanza de largo de código para Golomb PO2 con distribución geométrica.*
Sea P la distribución de probabilidad geométrica de parámetro γ sobre los naturales,

$$P(x) = (1-\gamma)\gamma^x, \quad 0 < \gamma < 1, \quad x \in \mathbb{N}\,.$$

Considere un código de Golomb PO2, G_k^*, de parámetro k. Sea L_k la esperanza del largo de código de G_k^* bajo la distribución $P(x)$,

$$L_k = \sum_{x=0}^{\infty} |G_k^*(x)| P(x)\,,$$

donde $|G_k^*(x)|$ es el largo de código que G_k^* asigna al entero x.

1. Demuestre que $|G_k^*(x)| \geq k+1$ para todo x.
2. Demuestre que para todo $j \geq 0$, el conjunto

   $$S_j = \left\{ x \in \mathbb{N} \;\middle|\; |G_k^*(x)| = k+1+j \right\}$$

   contiene exactamente 2^k enteros. ¿Cuáles son estos enteros?
3. Sea $P(S_j) = \sum_{x \in S_j} P(x)$. Demuestre que

   $$L_k = k + 1 + \sum_{j=0}^{\infty} j P(S_j)\,.$$

4. Definimos $z \triangleq \gamma^{2^k}$. Demuestre que $P(S_0) = 1 - z$.
5. Demuestre que

   $$P(S_{j+1}) = zP(S_j), \quad j \geq 0.$$

   Aplicando esta identidad y el inciso anterior, escriba $P(S_j)$ explícitamente en función de z.
6. Demuestre que

   $$L_k = k + 1 + \frac{z}{1-z}\,.$$

Ejercicio 2 *Códigos de Golomb elementales (run length coding 2.0).*
Sea k un entero no negativo, y $m = 2^k$. El código EG_k codifica una cadena binaria de la manera siguiente: se recorre la cadena desde el principio, y se identifica la primera aparición de una palabra del conjunto $W = \{1, 01, 001, \ldots, 0^{m-1}1, 0^m\}$ (donde 0^ℓ representa una cadena de ℓ ceros). Llamamos a los elementos del conjunto W *palabras elementales* de orden k. Se codifica la palabra elemental 0^m con un 0, y la palabra elemental $0^\ell 1$, $0 \leq \ell \leq m-1$ con un 1 seguido de la representación binaria de ℓ como entero sin signo de k bits. Luego de codificada una palabra elemental, se repite el proceso a partir de la palabra elemental siguiente, continuando hasta agotar la cadena de entrada (suponemos que la cadena de entrada termina en 1, para evitar efectos de borde). Notar que el ejercicio 6 del práctico 3 define un código EG_3.

1. Demuestre que codificar una cadena binaria x^n con EG_k es equivalente (en el sentido de producir exactamente la misma codificación) al procedimiento siguiente: se descompone x^n en una concatenación de palabras $u_1 u_2 \ldots u_s$, donde $u_i = 0^{\ell_i}1$, $\ell_i \geq 0$, y se codifica como $G_m(\ell_1), G_m(\ell_2), \ldots, G_m(\ell_s)$, donde G_m es el código de Golomb de orden $m = 2^k$, y suponemos que la parte unaria de G_m precede a la binaria. Note que en las cadenas u_i, el número de ceros, ℓ_i, no está limitado, y por lo tanto estas cadenas no son necesariamente palabras elementales.

2. Consideramos una fuente binaria de Bernoulli, P_q, donde el parámetro $q \in [0, 1]$ es la probabilidad del símbolo 0. Sea $N_1(k)$ el largo medio de una palabra elemental de orden k, y $N_2(k)$ el largo medio de la palabra de código correspondiente en EG_k. Calcule $N_1(k)$ y $N_2(k)$ en función de q y k. Notar que esto fue resuelto para valores particulares de q y k en los incisos 3 y 4 del ejercicio 6 del práctico 3.

   Sugerencia: Recuerde la siguiente identidad
   $$\sum_{i=1}^{m} i q^i = \frac{q - (m+1)q^{m+1} + m q^{m+2}}{(1-q)^2}\,. \tag{1}$$

3. Sea $\mathcal{L}(k) = N_2(k)/N_1(k)$ y sea $\phi = (\sqrt{5} - 1)/2$. Muestre que el valor de k que minimiza $\mathcal{L}(k)$ es el entero que satisface
   $$\begin{cases} k = 0, & q \leq \phi, \\ q^{2^k} < \phi \leq q^{2^{k-1}}, & q \geq \phi\,. \end{cases} \tag{2}$$

   Por lo tanto, los puntos de transición de un valor minimizante k al siguiente están en $q_k = \phi^{2^{-k}}$.

   Sugerencia: Estudie el punto de corte entre $\mathcal{L}(k)$ y $\mathcal{L}(k+1)$, $k \geq 0$.