«“‘<HEAD ======= ”’»>b98200e65a396b5fb64504f663bb0fb97b9bd066

UNIVERSIDAD NACIONAL DE COLOMBIA

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
TEORIA DE CODIFICACIÓN

Edgar Santiago Ochoa Quiroga
María Alejandra Rodríguez Ríos ..........................................................

**Ejercicio 1**
Suponga que una fuente genera dígitos binarios con distribución uniforme, los mensajes son enviados a través de un canal cambia los símbolos que genera una fuente binaria con las probalidades que se muestran en la gráfica.
poner foto Responda cada una de las siguientes preguntas justificando su razonamiento:
**a).** ¿Cuál es el número más probable de errores que se puede encontrar en un mensaje de 5000 bits que pase por el canal?

**Solución.** Teniendo en cuenta que al evaluar si un bit fue bien codificado através del canal, el resultado que ponemos obtener es sí o no, podemos resolver este ejercicio utilizando la distribución binomial.
Sabemos que, los bits pueden tomar valores de 0 o 1, sin embargo, el enunciado nos dice que la manera en la que se distribuyen através de los 5000 bits es uniforme, es decir, que podemos tomar la probabilidad de que el bit sea 0 o 1 como $\frac{1}{2}$.
Ahora bien, para calcular cada uno de los errores tomamos en cuenta la información del gráfico, así obtenemos que

$$P(0|1) = 0{,}18$$
$$P(1|0) = 0{,}15$$

Por lo cual, el error total al codificar es

$$P(error) = \frac{1}{2} \cdot 0{,}18 + \frac{1}{2} \cdot 0{,}5 = 0{,}165$$

Luego, al calcular la esperanza tenemos que la cantidad de errores es

$$\mu = 5000 \cdot 0{,}165 = 825$$

Con lo cual podemos decir que el número más probable de errores es 825.

**b).** Si se codifica cada bit que genera la fuente triplicandolo, en qué porcentaje se reduce el número de errores? Suponga que para la decodificación se usa mayoría de bits por tripla.

**Solución.** Ahora, tenemos que calcular el error para una fuente que triplica los bits, sabemos que mediante esta codificación para que se de un error quiere decir que 2 o 3 bits al pasar por el canal se codificaron de manera incorrecta. Así que calculemos esa probabilidad

$$P(errortripla) = P(2errores) + P(3errores)$$

Al igual que en el numeral anterior, vamos a utilizar la distribución binomial para calcular la probabilidad de que cambien 2 o 3 bits

$$\begin{aligned} P(2\text{ errores}) &= \binom{3}{2}p^2(1-p) = 3 \cdot (0{,}165)^2(0{,}835) = 0{,}0681986 \\ P(3\text{ errores}) &= \binom{3}{3}p^2(1-p) = (0{,}165)^3 = 0{,}004492. \end{aligned}$$

Luego el error por tripla tenemos que es

$$\begin{aligned} P(errortripla) &= P(2errores) + P(3errores) \\ &= 0{,}0681986 + 0{,}004492 \\ &= 0{,}0726906. \end{aligned}$$

Por lo cual tenemos que la reducción porcentual del error es

$$Reducciónporcentual = \frac{0{,}165 - 0{,}0726906}{0{,}165} \cdot 100 = 55{,}9451\,\%.$$

**c).** Si se codifica cada bit que genera la fuente con una n-tupla y se usa mayoría de bit por n tupla en la decodificación, cuál sería la longitud mínima n - tupla para obtener por lo menos un 95 % de certeza en la decondificación?

**Solución.** Para realizar este ejercicio tenemos que tener en cuenta que $n$ para que el criterio se decidible debe ser impar y por el numeral anterior tenemos cuando $n = 3$ el error es de $7{,}269\,\%$, como estamos buscando que el error sea de $5\,\%$ o menor vamos a probar los siguientes casos

- **Caso** $n = 5$ Para este caso vamos a aplicar la distribución de binomial ahora tomando los casos donde se presenten por lo menos 3 errores, asi calculamos lo siguiente

$$\begin{aligned} P(3\text{ errores}) &= \binom{5}{3}p^3(1-p)^{5-3} = 10(0{,}165)^3(0{,}835)^{5-3} = 0{,}0313202, \\ P(4\text{ errores}) &= \binom{5}{4}p^4(1-p)^{5-4} = 5(0{,}165)^4(0{,}835)^{5-4} = 0{,}00309451, \\ P(5\text{ errores}) &= \binom{5}{5}p^5(1-p)^{5-5} = (0{,}165)^5(0{,}835) = 0{,}000122298 \end{aligned}$$

  Entonces,tenemos que el error del código que generan las $5-$tuplas es

$$\begin{aligned} P(5tuplaserror) &= P(3\text{ errores}) + P(4\text{ errores}) + P(5\text{ errores}) \\ &= 0{,}0313202 + 0{,}00309451 + 0{,}000122298 \\ &= 0{,}034537. \end{aligned}$$

Luego el error es de 3,4537 % y por lo tanto la longitud mínima para que la certeza sea de 95 % es de 5.

**Ejercicio 2**
Se codifican los bits de una fuente de acuerdo a la siguiente función

$$\mathrm{Cod}(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, x_3, x_4, x_5, x_1+x_2+x_4, x_1+x_3+x_4, x_2+x_3+x_4, x_1+x_3+x_5, x_3+x_4+x_5)$$

Si el mensaje recibido es 1101110101 determine si hay un error en el mensaje y si es posible corregir el error . Use diagramas de Ven o el método de Galager para llegar a la solución. Deje todos los cálculos que le llevan a la solución en su hoja de respuesta.

**Solución.** Hola :p

**Ejercicio 3**
Clasifique los siguientes códigos de tal manera que puedan codificar la misma fuente. Luego indique por fuente cuál sería el mejor código. Justifique todas sus afirmaciones.

- $C_1 = \{0, 1\}$
- $C_2 = \{00, 01, 10\}$
- $C_3 = \{000, 111\}$
- $C_4 = \{000, 011, 101, 110, 001, 100\}$
- $C_5 = \{00000, 11111, 22222\}$
- $C_6 = \{001, 010, 012, 021, 100, 221\}$

**Solución.** Hola :p

**Ejercicio 4**
Determine justificando, si los códigos $C_1 = \{1, 10, 100, 1000\}$ y $C_2 = \{1, 01, 001, 0001\}$ son univocamente decodificables.

**Solución.** Procederemos con cada código por separado

- Sea $C = \{1, 10, 100, 1000\}$., Donde nuestro alfabeto para codificar es $D = \{0, 1\}$ Procedamos por medio se Sardinas-Patterson. Primero definimos $C_0 = C$, por el algoritmo tenemos que

  $$C_1 = \{w \in D^+ | uw = v,\ \text{con}\ u \in C_0\ \text{y}\ v \in C\ \text{o}\ u \in C\ \text{y}\ v \in C_0\}.$$

  Note que como $C_0 = C$ Basta con resolver ecuaciones de concatenación de palabras donde solo revisamos los elementos de $C$. Procedemos por casos.

**Caso 1:** Si $u = 1000$, tenemos las siguientes cuatro ecuaciones

$$1000w = \begin{cases} 1 \\ 10 \\ 100 \\ 1000 \end{cases}$$

Note que $|1000w| \geq 5$ ya que $|w| \geq 1$. Pero las longitudes de las palabras al otro lado de la igualdad son a lo mas 4, por lo que estas ecuaciones no tienen solución.

**Caso 2:** Si $u = 100$, tenemos las siguientes cuatro ecuaciones

$$100w = \begin{cases} 1 \\ 10 \\ 100 \\ 1000 \end{cases}$$

Note que para las primeras tres al igual que en el caso anterior no hay solución simplemente viendo las longitudes, pero para la ultima tomando $w = 0$, tenemos nuestra primera solución.

**Caso 3:** Si $u = 10$, tenemos las siguientes cuatro ecuaciones

$$10w = \begin{cases} 1 \\ 10 \\ 100 \\ 1000 \end{cases}$$

Siguiendo la lógica de los anteriores casos nos damos cuenta que $w = 00$ es solución de la cuarta ecuación, mientras que $w = 0$ de la tercera y las otras dos no tienen.

**Caso 4:** Si $u = 1$, tenemos las siguientes cuatro ecuaciones

$$1w = \begin{cases} 1 \\ 10 \\ 100 \\ 1000 \end{cases}$$

Siguiendo la lógica de los anteriores casos nos damos cuenta que $w = 000$ es solución de la cuarta ecuación, mientras que $w = 00$ de la tercera y $w = 0$ de la segunda, mientras que la primera no tiene solución.

Así tenemos que $C_1 = \{0, 00, 000\}$. Siguiendo el algoritmo tenemos que

$$C_2 = \{w \in D^+ | \, uw = v, \text{ con } u \in C_1 \text{ y } v \in C \text{ o } u \in C \text{ y } v \in C_1\}.$$

Veamos los posibles casos.

**Caso 1:** Si $u \in C_1$ y $v \in C$, note que la palabra $uw$ siempre empieza en $0$, ya que $u$ es igual a $0, 00$ o $000$, mientras que todas las palabras $v$ en C empiezan en 1, así ninguna ecuación tiene solución.

**Caso 2:** Si $u \in C$ y $v \in C_1$, pasa igual que en el caso anterior. Todas la posibles palabras $uw$ empiezan por 1, ya que cualquier palabra en C empieza por uno, mientras que todas las palabras en $C_1$ empiezan por 0, así ninguna ecuación tiene solución. Así $C_2 = \varnothing$.

Por la construcción podemos concluir que $C_3 = C_4 = \cdots = \varnothing$. Por lo tanto

$$C_\infty = \bigcup_{i=1}^{\infty} C_i = C_1.$$

Así podemos notar que

$$C \cap C_\infty = \{1, 10, 100, 1000\} \cap \{0, 00, 000\} = \varnothing.$$

Concluyendo por el teorema de Sardinas-Patterson que C es unívocamente decodificable.

- Sea $C = \{1, 01, 001, 0001\}$, donde nuestro alfabeto para codificar es $D = \{0, 1\}$. Siguiendo el algoritmo, tomamos $C_0 = C$, siguiendo el algoritmo tenemos que

$$C_1 = \{w \in D^+ | \, uw = v, \text{ con } u \in C_0 \text{ y } v \in C \text{ o } u \in C \text{ y } v \in C_0\}.$$

Veamos los casos.

**Caso 1:** Sea $u = 0001$, Note que $|uw| \geq 5$, y las posibles longitudes para $v \in C$ son menores o iguales a 4, luego para este $u$, las ecuaciones no tienen solución.

**Caso 2:** Sea $u = 001$, note que por el argumento de las longitudes la única ecuación que se podría considerar es

$$001w = 0001.$$

Pero note que el tercer símbolo de izquierda a derecha de la palabra de la izquierda es 1, mientras que el de la derecha es 0, por lo que ninguna ecuación tiene solución.

**Caso 3:** Sea $u = 01$, nuevamente por el argumento de las longitudes las ecuaciones que valen la pena considerar son

$$01w = \begin{cases} 001 \\ 0001 \end{cases}$$

Pero nuevamente el segundo símbolo de izquierda a derecha de la palabra de la izquierda es 1, mientras que en ambas palabras de la derecha el símbolo es 0, así para estas ecuaciones nuevamente no tenemos soluciones.

**Caso 4:** Para este ultimo caso, siguiendo la misma idea de las longitudes, cuando $u = 1$ basta con considerar las siguientes ecuaciones

$$1w = \begin{cases} 01 \\ 001 \\ 0001 \end{cases}$$

Pero la palabra de la izquierda empieza en 1, mientras que las de la derecha en 0, por lo que nuevamente no ha soluciones.

Como en ningún caso existe $w$ que cumpla las ecuaciones, tenemos que $C_1 = \varnothing$. Luego tenemos que $C_2 = C_3 = \cdots = \varnothing$. De esta manera tenemos que

$$C_\infty = \bigcup_{i=1}^{\infty} C_i = \varnothing.$$

Luego es claro que

$$C \cap C_\infty = \varnothing.$$

Así por el teorema de Sardinas-Patterson podemos concluir que C es unívocamente decodificable.

De esta manera concluimos por medio de Sardinas-Patterson que los $C_1$ y $C_2$, presentados en el enunciado, son unívocamente decodificables.

ʊ̬ʊ

**Ejercicio 5**
Pruebe, diseñando un algoritmo (sin usar el algoritmo de SardinasPaterson), que el código triario $\{aa, b, ba, abc\}$ es univocamente decodificable

**Solución.** Hola :p

**Ejercicio 6**
Determine si el código triario $C = \{ab, cb, abbc, cbc, abb\}$ es univocamente decodificable. Existe un código binario instantaneo con las longitudes de palabras de C ? Si existe construyalo.

**Solución.**

**Ejercicio 7**
Una fuente genera símbolos con una distribución de probabilidad

$$\{0{,}729, 0{,}081, 0{,}081, 0{,}081, 0{,}009, 0{,}009, 0{,}009, 0{,}001\}.$$

Determine un código binario que permita calcular la codificación de los símbolos de la fuente con menor longitud promedio. Cálcule la eficiencia del código.

**Solución.** Hola :p

**Ejercicio 8**
Diseñe un código de Fano para una fuente con la siguiente distribución de probabilidad

$$\{0{,}20, 0{,}19, 0{,}18, 0{,}17, 0{,}15, 0, 10, 0{,}01\}$$

.

**Solución.** Hola :p

**Ejercicio 9**
Resuelva el ejercicio 1.5.1 de las notas de clase. Construya un código que no sea instantáneo cuyas longitudes de palabra cumplen la desigualdad de Kraft.

**Solución.** Sea $C = \{1, 10, 100, 1000\}$. Este codigo claramente no es instantaneo, ya que no es prefijo, por que la palabra 1 es segmento inicial de todas la demas palabras del codigo. Note que para este codigo tenemos que $D = 2$ ya que es un codigo binario, y tenemos longitudes de palabra $\ell_1 = 1, \ell_2 = 2, \ell_3 = 3$ y $\ell_4 = 4$. Asi

$$\begin{aligned}\sum_{i=1}^{4} D^{-\ell_i} &= 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} \\ &= 2^{-4}(2^3 + 2^2 + 2 + 1) \\ &= \frac{8+4+2+1}{16} \\ &= \frac{15}{16} \leq 1.\end{aligned}$$

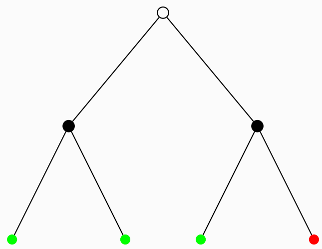
De esta forma C es un codigo no instantaneo que cumple la desigualdad de Kraft como queriamos. ʘ‿ʘ

**Ejercicio 10**
Resuelva el ejercicio 1.5.2 de la notas de clase, calcule la eficiencia. Construya, en caso de que exista, un código binario instantáneo constituido por 5 palabras de longitudes: 2, 2, 2, 3 y 4.

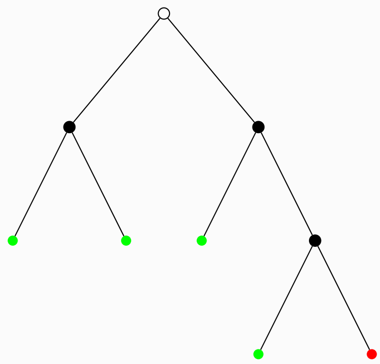
**Solución.** Primero verifiquemos que tal codigo exista. Tenemos que $D = 2$, por que es binario y tenemos longitudes de palabra $\ell_1 = 2, \ell_2 = 2, \ell_3 = 2, \ell_4 = 3$ y $\ell_5 = 4$. Asi

$$\begin{aligned}\sum_{i=1}^{5} D^{-\ell_i} &= 2^{-2} + 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} \\ &= 2^{-4}(2^2 + 2^2 + 2^2 + 2 + 1) \\ &= \frac{4+4+4+2+1}{16} \\ &= \frac{15}{16} \leq 1.\end{aligned}$$

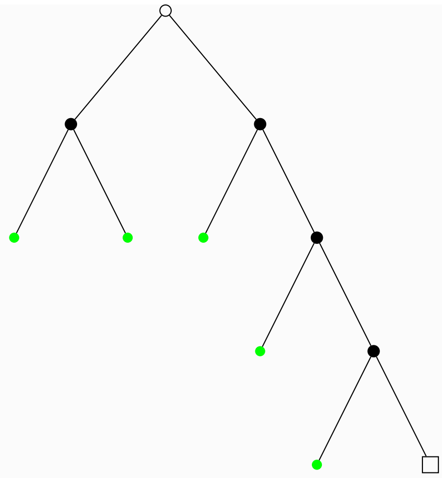
Como cumple la desigualdad de Kraft, sabemos que existe un codigo instantaneo (no completo ya que no se tiene la igualdad), con las longitudes de palabra indicadas. Siguiendo el algoritmo de construccion, tenemos para el paso uno el siguiente arbol de altura 2.



Donde los nodos verdes serán hojas, y el rojo es un nodo terminal. Para el segundo paso aumentamos a un árbol de altura 3 como se muestra a continuación



Para finalizar aumentamos el árbol a una altura 4 como se muestra a continuación



De esta forma si a las ramas que abren hacia la izquierda les asignamos el 0 y a las otras el 1, obtenemos el siguiente código instantáneo con las longitudes de palabra deseadas.

$$C = \{00, 01, 10, 110, 1110\}.$$

**Ejercicio 11**
Resuelva el ejercicio 2.3.1 de la página 36 de las notas de clase, calcule la eficiencia en cada caso. Considere el alfabeto S con la distribución de probabilidades que se muestra en la tabla:

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 0,02 | 0,03 | 0,04 | 0,04 | 0,12 | 0,20 | 0,20 | 0,35 |

Suponga que se ha usado un código de Huffman para codificar los mensajes sobre un alfabeto binario. Si en el árbol de codificación se le asigna 1 a las ramas sobre la izquierda y 0 a las ramas sobre la derecha, ¿qué palabra representa la secuencia 111011111101110? Determine la longitud promedio de palabra para este código. ¿Cuál sería la codificación de los símbolos sobre un código triario?

**Solución.** Hola :p

**Ejercicio 12**
Resuelva el ejercicio 3.5.1 de la página 61 de las notas de clase, calcule la eficiencia en cada caso. Considere una fuente que genera símbolos del alfabeto $S = \{0, 1\}$ con probabilidades $p(0) = 0{,}9$ y $p(1) = 0{,}1$. Diseñe códigos de Huffman y de Shannon-Fano para los alfabetos extendidos $S^2$, $S^3$ y $S^4$, en cada caso calcule la eficiencia.

**Solución.** Hola :p