Teoría de la información

Problemas de Criptología - Hoja 4

Curso 07-08

1 **Problemas del libro de Manuel Lucena.** Haced los problemas 2, 5, 6, y 7 del capítulo 3, páginas 51 y siguientes, del libro de Manuel Lucena.

Solución: Véase su libro.

2 **Entropía de bolas.** Se tiene una bolsa con cinco bolas rojas, tres bolas blancas y dos bolas negras. Las bolas del mismo color se consideran idénticas entre sí.

(a) Se escogen dos bolas de las bolsa con reemplazamiento. ¿Cuál es la entropía de este experimento?

(b) ¿Cuál es la entropía si ahora se escogen sin reemplazamiento?

Solución:

(a) Con reemplazamiento las probabilidades de sacar 2 bolas son:

Suceso	(R,R)	(R, B)	(R, N)	(B, B)	(B, N)	(N, N)
Probabilidad	$\frac{25}{100}$	$\frac{30}{100}$	$\frac{20}{100}$	$\frac{9}{100}$	$\frac{12}{100}$	$\frac{4}{100}$

Luego la entropía es:

$$H(X) = -\left(\frac{25}{100}\log_2\frac{25}{100}+\frac{30}{100}\log_2\frac{30}{100}+\frac{20}{100}\log_2\frac{20}{100}+ \frac{9}{100}\log_2\frac{9}{100}+\frac{12}{100}\log_2\frac{12}{100}+\frac{4}{100}\log_2\frac{4}{100}\right) = 2'3509 \text{ bits.}$$

(b) Sin reemplazamiento las probabilidades de sacar 2 bolas son:

Suceso	(R,R)	(R, B)	(R, N)	(B, B)	(B, N)	(N, N)
Probabilidad	$\frac{5\cdot4}{90}$	$\frac{30}{90}$	$\frac{20}{90}$	$\frac{3\cdot2}{90}$	$\frac{12}{90}$	$\frac{2\cdot1}{90}$

Luego la entropía es:

$$H(X) = -\left(\frac{20}{90}\log_2\frac{20}{90}+\frac{30}{90}\log_2\frac{30}{90}+\frac{20}{90}\log_2\frac{20}{90}+ \frac{6}{90}\log_2\frac{6}{90}+\frac{12}{90}\log_2\frac{12}{100}+\frac{2}{90}\log_2\frac{2}{100}\right) = 2'1733 \text{ bits.}$$

3 **Cálculo de entropías.** Se tira una moneda equilibrada tres veces. Sea X la variable aleatoria que da el número de cruces e Y la que proporciona el valor absoluto entre el número de caras y cruces.

(a) Hallar la distribución de probabilidad de X, Y y (X, Y).

(b) Hallar la distribución de $X|Y = 1$.

(c) Hallar las entropías $H(X), H(Y)$, $H(X, Y)$, $H(X|Y)$ y $H(Y|X)$.

(d) Hallar la cantidad de información de Y sobre X.

Solución:

(a) La variable X toma los valores $\{0, 1, 2, 3\}$ y la variable Y, los valores $\{1, 3\}$. La distribución conjunta se muestra en la siguiente tabla:

	Valores de X				
Valores de Y	0	1	2	3	$P(Y = y)$
1	0	3/8	3/8	0	6/8
3	1/8	0	0	1/8	2/8
$P(X = x)$	1/8	3/8	3/8	1/8	1

Las distribuciones marginales se calculan a partir de esta tabla sumando por columnas para X y por filas para la Y.

(b) La variable $X|Y = 1$ se obtiene usando la fórmula $P(X|Y = 1) = \dfrac{P(X, \{Y = 1\})}{P(Y = 1)}$.

Tenemos la siguiente tabla:

Valores de $X\|Y = 1$	0	1	2	3
Probabilidad	0	1/2	1/2	0

Obsérvese que:

$$P(X = 0|Y = 1) = P(X = 3|Y = 1) = 0,$$
$$P(X = 1|Y = 1) = \frac{P(X = 2, Y = 1)}{P(Y = 1)} = \frac{3/8}{6/8} = \frac{1}{2},$$
$$P(X = 3|Y = 1) = \frac{P(X = 3, Y = 1)}{P(Y = 1)} = \frac{3/8}{6/8} = \frac{1}{2}$$

(c) Usando la tabla anterior tenemos:

$$
\begin{aligned}
H(X) &= -\left(\frac{1}{8}\log_2\frac{1}{8}+\frac{3}{8}\log_2\frac{3}{8}+\frac{3}{8}\log_2\frac{3}{8}+\frac{1}{8}\log_2\frac{1}{8}\right) = 1'2806 \text{ bits.}\\
H(Y) &= -\left(\frac{6}{8}\log_2\frac{6}{8}+\frac{2}{8}\log_2\frac{2}{8}\right) = 0'8112 \text{ bits.}\\
H(X,Y) &= -\left(0\log_2 0+\frac{3}{8}\log_2\frac{3}{8}+\frac{3}{8}\log_2\frac{3}{8}+0\log_2 0+\right.\\
&\quad \left.\frac{1}{8}\log_2\frac{1}{8}+0\log_2 0+0\log_2 0+\frac{1}{8}\log_2\frac{1}{8}\right) = 1'8112 \text{ bits.}\\
H(X|Y) &= -\sum_{i=1}^{4}\sum_{j=1}^{2} P(y_j)P(x_i|y_j)\log_2 P(x_i|y_j) =\\
&\quad -\left(\frac{6}{8}\cdot 0\log_2 0+\frac{6}{8}\cdot\frac{1}{2}\log_2\frac{1}{2}+\frac{6}{8}\cdot\frac{1}{2}\log_2\frac{1}{2}+\frac{6}{8}\cdot 0\log_2 0+\right.\\
&\quad \left.\frac{2}{8}\cdot\frac{1}{2}\log_2\frac{1}{2}+\frac{2}{8}\cdot 0\log_2 0+\frac{2}{8}\cdot\frac{1}{2}\log_2\frac{1}{2}+\frac{2}{8}\cdot 0\log_2 0\right) = 1 \text{ bit.}
\end{aligned}
$$

También se puede calcular $H(X|Y)$ como $H(X|Y) = H(X,Y) - H(Y) = 1'8112 - 0'8112 = 1$ bit. Para $H(Y|X)$ usamos que $H(Y|X) = H(X,Y) - H(X) = 1'8112 - 1'2806 = 0'5306$ bits.

(d) La cantidad de información de Y sobre X es $I(Y,X) = H(X) - H(X|Y) = 1'2806 - 0'8112 = 0'4693$ bits.

4 **Entropía de funciones de variables aleatorias.** Sea X una variable aleatoria que toma valores en $\{1, 2, \cdots, 36\}$ con igual probabilidad en cada valor.

(a) ¿Cuánto vale $H(X)$?

(b) Si hacemos $Y = X^{36} \bmod 37$, ¿cuánto vale $H(Y)$? (Pista: construye la tabla de valores de Y con Maple.) ¿Y si hacemos ahora $Z = X^2 \bmod 36$?

Solución:

(a) Dado que los estados son equiprobables, $H(X) = \log_2 36 = 2\log_2 6 = 5'1699$ bits.

(b) Se tiene que $i^{36} \bmod 37 = 1$ para todo $i = 1, \cdots, 36$. En consecuencia, solo hay un estado y la entropía $H(Y) = 0$. Para i^2 se tiene la siguiente tabla:

$$
[1, 4, 9, 16, 25, 36, 12, 27, 7, 26, 10, 33, 21, 11, 3, 34, 30,\\
28, 28, 30, 34, 3, 11, 21, 33, 10, 26, 7, 27, 12, 36, 25, 16, 9, 4, 1]
$$

Se puede apreciar que cada número aparece exactamente 2 veces. Hay 18 estados con probabilidad $\frac{2}{36}$. La entropía $H(Z)$ es $\log_2 18 = 4'1699$.

5 **Seguridad perfecta.** Sea $M = \{a, b, c\}$ el espacio de mensajes, $C = \{U, V, W\}$ el espacio de criptogramas. Como espacio de claves K se toma las permutaciones de 3 símbolos. Por ejemplo, la siguiente permutación

$$\pi = \begin{pmatrix} a & b & c \\ V & W & U \end{pmatrix}$$

lleva a a V, b a W y c a U y es una de las 6 claves posibles. Las claves son equiprobables. Las probabilidades en M son $P(a) = \frac{1}{2}, P(b) = \frac{3}{10}$ y $P(c) = \frac{2}{10}$.

(a) Hallar la entropía de C.

(b) Hallar la entropía de M.

(c) Estudiar si es un criptosistema de seguridad perfecta.

(d) Hallar el índice del lenguaje en M así como su índice absoluto. Hallar la redundancia.

(e) Hallar la distancia de unicidad.

Solución: Para facilitar la notación llamamos a los elementos de M por $m_i, i = 1, 2, 3$, por $c_j, j = 1, 2, 3$ a los elementos de C y por $k_i, i = 1, \cdots, 6$ a los de K.

(a) Dado que la aparición de un símbolo del criptograma depende de la clave, usando el teorema de la probabilidad total, tenemos:

$$\begin{aligned} P(U) &= P(k_1)P(m_1) + P(k_2)P(m_1) + P(k_3)P(m_2) + P(k_4)P(m_3) \\ &\quad + P(k_5)P(m_2) + P(k_6)P(m_3) \\ &= \frac{1}{6}(P(m_1) + P(m_1) + P(m_2) + P(m_3) + P(m_2) + P(m_3)) = \\ &= \frac{1}{6}(1 + 1) = \frac{1}{3} \end{aligned}$$

Con un cálculo similar se obtiene que $P(V) = P(W) = \frac{1}{3}$. En conclusión, el espacio de mensajes cifrados es equiprobable.

(b) La entropía $H(M)$ es:

$$\begin{aligned} H(M) &= -\sum_{i=1}^{3} P(m_i)\log_2(P(m_i)) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{3}{10}\log_2\frac{3}{10} + \frac{2}{10}\log_2\frac{2}{10}\right) \\ &= 1'0296 \text{ bits.} \end{aligned}$$

Clave	Permutación	Transformación
k_1	(1 2 3)	U V W
k_2	(1 3 2)	U W V
k_3	(2 1 3)	V U W
k_4	(2 3 1)	V W U
k_5	(3 1 2)	W U V
k_6	(3 2 1)	W V U

(c) Las claves se pueden resumir en la siguiente tabla:

Tenemos:

$$\begin{aligned} P(a|U) &= \frac{P(a,k_1)+P(a,k_2)}{P(U)} = \frac{\frac{1}{2}\cdot\frac{1}{6}+\frac{1}{2}\cdot\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2} \\ P(b|U) &= \frac{P(b,k_3)+P(b,k_5)}{P(U)} = \frac{\frac{3}{10}\cdot\frac{1}{6}+\frac{3}{10}\cdot\frac{1}{6}}{\frac{3}{10}} = \frac{3}{10} \\ P(c|U) &= \frac{P(c,k_4)+P(b,k_6)}{P(U)} = \frac{\frac{2}{10}\cdot\frac{1}{6}+\frac{2}{10}\cdot\frac{1}{6}}{\frac{2}{10}} = \frac{2}{10} \end{aligned}$$

Análogos cálculos muestran que $P(a|V)=\frac{1}{2}, P(b|V)=\frac{3}{10}, P(c|V)=\frac{2}{10}$ y que $P(a|W)=\frac{1}{2}, P(b|W)=\frac{3}{10}, P(c|W)=\frac{2}{10}$. Observamos a partir de las probabilidades que M y C son independientes.

Calculamos $H(M|C)$:

$$\begin{aligned} H(M|C) &= -\sum_{i=1}^{3}\sum_{j=1}^{3} P(c_j)P(m_i|c_j)\log_2(P(m_i|c_j)) \\ &= -\sum_{i=1}^{3}\sum_{j=1}^{3} P(c_j)P(m_i)\log_2(P(m_i)) \\ &= -\sum_{i=1}^{3}\left(\sum_{j=1}^{3} P(c_j)\right)P(m_i)\log_2(P(m_i)) \\ &= -\sum_{i=1}^{3} 1\cdot P(m_i)\log_2(P(m_i)) = H(M) \end{aligned}$$

En efecto, este criptosistema es de seguridad perfecta.

(d) Hace falta calcular el número de mensajes de longitud k. En general, este problema es complicado. Por ejemplo, para mensajes de longitud 2, tenemos $3^2 = 9$ mensajes con probabilidades no uniformes. La probabilidad de aa es $1/4$, la de ab es $1/20$, etc.

La expresión $H_k(M)$ es muy complicada y a menudo solo se obtienen aproximaciones numéricas.

El índice absoluto del lenguaje R es meramente $\log_2 3$ y coincide con el número anterior. La redundancia del lenguaje es $R - r = 0$.

(e) La distancia de unicidad está dada por $H(K|C)$:

$$H(K|C) = -\sum_{i=1}^{6}\sum_{j=1}^{3} P(c_j)P(k_i|c_j)\log_2 P(k_i|c_j)$$

Las probabilidades $P(c_j)$ se calcularon más arriba. Respecto a las probabilidades $P(k_i|c_j)$, pensemos que la aparición de c_j no nos dice nada sobre la aparición de k_i, es decir, que las claves y los criptogramas son independientes y $P(k_i|c_j) = \frac{1}{6}$. Luego:

$$H(K|C) = -\sum_{i=1}^{6}\sum_{j=1}^{3} P(c_j)P(k_i|c_j)\log_2 P(k_i|c_j) = -\sum_{i=1}^{6}\sum_{j=1}^{3} \frac{1}{3}\cdot\frac{1}{6}\log_2\frac{1}{6} = \log_2 6$$

Esta cantidad corresponde a la incertidumbre máxima para un sistema de 6 estados.

6 **Cifrado afín.** Supongamos que M es el conjunto de mensajes formado por letras del alfabeto español. Analiza si un cifrado afín sobre M tiene seguridad perfecta.

Solución: El conjunto de claves para el cifrado afín sobre el alfabeto español es $\{(a,b) \mid (a,27) = 1\}$, donde a, b son números de $\{1, 2, \cdots, 26\}$. El número de claves posible es $(27-9)\cdot 27 = 486$. Dados una letra $m \in M$ y un criptograma $c \in C$, si el cifrado afín fuese de seguridad perfecta, debería haber una única clave (a,b) tal que $E_{(a,b)}(m) = c$. Sin embargo, la ecuación $ax + b = c \bmod 27$ tiene más de una solución en a y en b. Por ejemplo, si $m = 5$ y $c = 7$, se tiene que $1\cdot 5 + 2 = 7 \bmod 27$ y $5\cdot 5 + 9 = 34 = 7 \bmod 27$. Por tanto, el cifrado afín no tiene seguridad perfecta.

7 **Cifrado de Vernam o de un solo uso.** Cifra la cadena de bits 1101100101 mediante un cifrado de un solo uso. Construye el generador cuadrático para la cadena aleatoria de bits (los primos que uses deberían tener 10 dígitos al menos). Usa Maple para los cálculos.

Solución: Primero hallamos 2 primos grandes. Usamos la función nextprime.

```
> nextprime(157894652785);
157894652809

>nextprime(157894652785*157894652785);
24930721378095708256273
```

A continuación, usando la función seq localizamos un primo que sea congruente con 3 módulo 4. En nuestro caso escogimos 157894653019. Después generamos otro primo que sea congruente 3 módulo 4.

```
> seq([nextprime(5+i), nextprime(i+5) mod 4],
i=157894652809..157894653000);

> 1578946530192;
2493072145199040581436 1

> nextprime(2493072145199040581436 1);
2493072145199040581438 3

> \% mod 4;
3
```

Calculamos n como el producto de los 2 primos anteriores:

```
> n:=157894653019*2493072145199040581438 3;
n := 3936427613175364992979003904572277
```

La siguiente función nos calcula una matriz que contiene los números y el bit menos significativo:

```
> vernam:= proc(x, n, m)
> local j, z;
> z:=Matrix(1..m,1..2,0);
> z[1,1]:=x;
> z[1,2]:= x mod 2;
> for j from 2 to m do
>    z[j,1]:= z[j-1,1]^2 mod n;
>    z[j,2]:= (z[j-1,1]^2 mod n) mod 2;
> end do;
> return(z);
> end:
```

Como valor inicial de x usamos el primo 965432178546213577. Cuando se ejecuta sobre los x, n y m=10, que es la longitud de la cadena de bits que codificar, tenemos:

```
vernam(965432178546213577,39364276131753649929790039045722 77 ,10);
[                965432178546213577  1]
[3062374663101872547865762220077557  1]
[1572923238129638370241074309947625  1]
[1890052347342472892300737441876105  1]
[1413200681585781544579034606393990  0]
[1179776936114049902396701938444972  0]
[3276047822429412244108301339507789  1]
[1512218404013605550090747886584261  1]
[3735603606031274765213803892691245  1]
[2895172330653263817283422099535476  0]
```

Luego la cadena de bits resultante es 1111001110. Cuando hacemos o exclusivo con la cadena original se obtiene 1101010100.