Información y Computación Cuántica

José Caballero Béjar

21 de septiembre de 2009

EJERCICIO: Calcular la información de una secuencia de 10 letras y de una secuencia de 26 dígitos

Suponiendo que las 26 letras son equiprobables, se trata de averiguar qué tiene más información, si una secuencia de 10 símbolos pertenecientes a un grupo de 26, o una secuencia de 26 símbolos pertenecientes a un grupo de 10.

$$\mathrm{I}(10 \text{ letras}) = 10\ \mathrm{I}(1 \text{ letra}) \ = \ 10 \ \log_2 \frac{1}{\frac{1}{26}} \ = \ 10 \ \log_2 26 \ \sim \ 47$$

$$\mathrm{I}(26 \text{ digitos}) = 26\ \mathrm{I}(1 \text{ digito}) \ = \ 26 \ \log_2 \frac{1}{\frac{1}{10}} \ = \ 26 \ \log_2 10 \ \sim \ 86{,}4$$

EJERCICIO: Calcular la información de un discurso de 1000 palabras y de una imagen de 500 x 600 pixels con 16 niveles de brillo cada uno

Si suponemos que las 10000 palabras son equiprobables, la información de una secuencia de 1000 es:

$$\mathrm{I}(1000 \text{ palabras}) = 1000\ \mathrm{I}(1 \text{ palabra}) \ = \ 1000 \ \log_2 \frac{1}{\frac{1}{10000}} \ = \ 1000 \ \log_2 10000 \ \sim \ 13287{,}7$$

La información de una imagen es:

$$\mathrm{I}(500\text{x}600 \text{ pixels}) = 300000\ \mathrm{I}(1 \text{ pixel}) \ = \ 300000 \ \log_2 \frac{1}{\frac{1}{16}} \ = \ 300000 \ \log_2 16 \ \sim \ 1200000$$

Realmente vale más una imagen que mil palabras.

EJERCICIO: Dado un grupo de nueve bolas, encontrar una estrategia que permita, en tres pesadas, encontrar la de diferente masa y decir si es más pesada o ligera que las demás

En general, si el grupo consta de n bolas, el número total de posibilidades es $2n$, pues cada una de las n bolas puede ser más pesada o más ligera que las demás. La mejor estrategia en cada pesada es intentar dividir las bolas que se estén considerando en 3 grupos, lo más parecidos en número que se pueda. En este caso, hay tres posibilidad:

- la balanza se inclina hacia la izquierda: entonces la bola es más pesada y está en el platillo izquierdo o más ligera y está en el derecho
- la balanza se inclina hacia la derecha: entonces la bola es más pesada y está en el platillo derecho o más ligera y está en el izquierdo
- la balanza no se inclina: entonces la bola, sea más pesada o más ligera, está entre las restantes

Es fácil ver que la mejor estrategia es dividir las bolas en grupos iguales. En efecto, sabemos que la información que aporta un experimento *random* es máxima si la distribución de probabilidad es uniforme. Dicho de otro modo, la entropía de un conjunto es máxima si todas las posibilidad son equiprobables. Podemos verlo haciendo uso de los multiplicadores de Lagrange para optimizar la función $H(X) = \sum p_i \ln \frac{1}{p_i}$ con la restricción de normalización $\sum p_i = 1$

$$L(p_1, \ldots, p_n, \lambda) \;=\; \sum p_i \ln \frac{1}{pi} + (\lambda - 1)\left(1 - \sum p_i\right)$$

(NOTA: se hace uso del factor $\lambda - 1$ para simplificar los cálculos)

Derivando...

$$-\frac{\partial L}{\partial p_i} = \ln p_i + 1 + (\lambda - 1) \;=\; \ln p_i + \lambda \;=\; 0 \;\Rightarrow\; \ln p_i = -\lambda$$

Con lo que vemos que todos los p_i han de valer igual para que la entropía sea máxima. En estas condiciones, el algoritmo que se encuentre será, por tanto, el óptimo. Por otro lado, como cada pesada tiene tres lecturas (listadas anterioremente), tras k pesadas el número de conclusiones que se obtiene es 3^k, luego el número de pesadas necesarias para resolver el problema ha de satisfacer que $3^k > 2n$

En el ejemplo de nueve bolas es, por tanto, suficiente con 3 pesadas ($3^3 = 27 > 18 = 2 \cdot 9$). Veamos una posible estrategia:

PASO 1

Hay nueve bolas (numeradas de 1 a 9), y una de ellas puede ser más pesada (lo que denotaremos con una P) o más ligera (L). Las 18 posibilidades son, por tanto:

1P	2P	3P	4P	5P	6P	7P	8P	9P
1L	2L	3L	4L	5L	6L	7L	8L	9L

Dividimos las 9 nueve bolas en 3 grupos de 3. En el plato izquierdo ponemos el grupo $\{1, 2, 3\}$, en el derecho $\{4, 5, 6\}$ y dejamos el resto. Los tres posibles resultados son:

- **1A** la balanza se inclina hacia la izquierda. Las posibilidad son entonces

1P	2P	3P
4L	5L	6L

- **1B** la balanza se inclina hacia la derecha, en cuyo caso las posibilidades son

4P	5P	6P
1L	2L	3L

- **1C** la balanza no se inclina, por tanto las posibilidades son

7P	8P	9P
7L	8L	9L

Vemos que con una sola pesada hemos reducido el número de posibilidades de 18 a 6.

PASO 2

De nuevo volvemos a dividir cada grupo en 3 partes iguales, en este caso de 2 bolas, pero procurando que ninguna pareja sea un subgrupo de las ternas usadas en el primer paso puesto que, si así fuese y la balanza repitiese el movimiento de la primera pesada, esa pareja no aportaría información nueva.

- **en el caso A** En el plato izquierdo ponemos las bolas $\{1,4\}$ y en el derecho $\{2,5\}$
  - si se balancea a la izquierda las posibilidades son 1P o 5L.
  - si se balancea a la derecha las posibilidades son 2P o 4L.
  - si no se balancea las posibilidades son 3P o 6L.
- **en el caso B** Hacemos la misma pesada.
  - si se balancea a la izquierda las posibilidades son 4P o 2L.
  - si se balancea a la derecha las posibilidades son 5P o 1L.
  - si no se balancea las posibilidades son 6P o 3L.
- **en el caso C** En el plato izquierdo ponemos las bolas $\{7,8\}$ y en el derecho cualquier grupo de dos de entre las ya descartadas
  - si se balancea a la izquierda las posibilidades son 7P o 8P.
  - si se balancea a la derecha las posibilidades son 7L o 8L.
  - si no se balancea es la bola 9 la que pesa diferente.

PASO 3

En cualquiera de los casos quedan dos posibilidades. Es suficiente con hacer una tercera pesada de una de ellas comparándola con cualquiera de las ya descartadas. Si la balanza se mueve es que hemos escogido la bola correcta, y si se queda estable es que era la otra.

EJERCICIO: ¿Qué podemos decir sobre el tamaño del mayor de N cubos de lado medio 10 y volumen medio 1000?

Para plantear el ejercicio podemos hacer uso de la *desigualdad de Jensen*:

Una función **f** es convexa en el intervalo (a,b) si para todo $x_1, x_2 \in (a,b)$ y $0 \leq \lambda \leq 1$ se verifica que:

$$f\left(\lambda x_1 \;+\; (1-\lambda)x_2\right) \;\leq\; \lambda f(x_1) \;+\; (1-\lambda)f\left(x_2\right)$$

y estrictamente convexa si la igualdad se obtiene sólo para $\lambda = 0$ o $\lambda = 1$.

La desigualdad dice que dada una función **f** y una variable aleatoria **x**, si **f** es convexa, entonces $E[f(x)] \geq f(E[x])$. Es decir:

$$si \;\sum_{i=1}^{l} p_i = 1 \;con\; p_i \geq 0 \;\Rightarrow\; \sum_{i=1}^{l} p_i f\left(x_i\right) \geq f\left(\sum_{i=1}^{l} p_i x_i\right) \tag{1}$$

Si **f** es estrictamente convexa y $E[f(x)] = f(E[x])$, entonces la variable **x** es constante.

La Eq. 1 es fácil de probar por recursión:

$$f\left(\sum_{i=1}^{l} p_i x_i\right) = f\left(p_1 x_1 + \sum_{i=2}^{l} p_i x_i\right)$$

si tomamos $p_1 = \frac{p_i}{\sum_{i=1}^{l} p_i} = \lambda$ y $\frac{\sum_{i=2}^{l} p_i}{\sum_{i=1}^{l} p_i} = 1-\lambda$ en la expresión anterior, y hacemos uso de la definición de convexidad, queda

$$f\left(\sum_{i=1}^{l} p_i x_i\right) \;\leq\; p_1 f\left(x_1\right) + \left[\sum_{i=2}^{l} p_i\right]\left[f\left(\frac{\sum_{i=2}^{l} p_i x_i}{\sum_{i=2}^{l} p_i}\right)\right]$$

volvemos ahora a aplicar el mismo razonamiento tomando $\lambda = \frac{p_2}{\sum_{i=2}^{l} p_i}$ y $1-\lambda = \frac{\sum_{i=3}^{l} p_i}{\sum_{i=2}^{l} p_i}$

$$f\left(\sum_{i=1}^{l} p_i x_i\right) \;\leq\; p_1 f\left(x_1\right) + \left[\sum_{i=2}^{l} p_i\right]\left[\frac{p_2}{\sum_{i=2}^{l} p_i} f\left(x_2\right) + \frac{\sum_{i=3}^{l} p_i}{\sum_{i=2}^{l} p_i} f\left(\frac{\sum_{i=3}^{l} p_i x_i}{\sum_{i=3}^{l} p_i}\right)\right]$$

y así sucesivamente. En el caso de que $E[f(x)] = f(E[x])$, todas las desigualdades anteriores se deben cambiar por igualdades. Si la función **f** es, además, estrictamente convexa, entonces en cada paso hay que aplicar que $\lambda = 0$ o $\lambda = 1$. Así, por ejemplo, en el primer paso de la demostración por recurrencia diríamos que p_1 sólo puede valor 0 o 1. Si se da el segundo caso, entonces p_2 es 0 o 1; y así sucesivamente. Sólo uno de los p_i vale 1, y el resto 0, con lo que se concluye que la variable **x** sólo admite un único valor posible.

Éstas son precisamente las condiciones del ejercicio. La función x^3 es estrictamente convexa, pues para que se verifique que

$$\left[\lambda x_1 + (1-\lambda)x_2\right]^3 \;=\; \lambda^3 x_1^3 + (1-\lambda)^3 x_2^3 + 3\lambda(1-\lambda)x_1 x_2\left[\lambda x_1 + (1-\lambda)x_2\right] \;=\; \lambda x_1^3 + (1-\lambda)x_2^3 \;\; \forall x_1, x_2 \in (0,\infty)$$

ha de cumplirse necesariamente que λ valga 0 o 1.

Por tanto, tenemos una función, el volumen, que es estrictamente convexa, y una muestra para la que se verifica que $E[V(l)] = V(E[l])$. Según hemos visto, eso indica que sólo hay un valor posible para la variable l, y por tanto todos los cubos son iguales, con l=10 y V=1000.

EJERCICIO: Calcular H(X), H(Y), H(X|Y) y H(Y|X) si X es el resultado de lanzar un dado e Y vale 1 cuando X es impar y 0 en caso contrario

La variable X toma valores $x_i = \{1, 2, 3, 4, 5, 6\}$, todos con probabilidad $p_i = \frac{1}{6}$.
La variable Y toma valores $y_i = \{0, 1\}$, con probabilidades $p_j = \frac{1}{2}$

$$H(X) \ = \ \sum_{i=1}^{6} \frac{1}{6} \log_2 6 \ = \ \log_2 6 \ \sim \ 2{,}58$$

$$H(Y) \ = \ \sum_{i=1}^{2} \frac{1}{2} \log_2 2 \ = \ \log_2 2 \ = \ 1$$

No sería necesario calcular $H(Y|X)$ para saber su valor, pues si interpretamos esta cantidad como la incertidumbre en la variable Y conocido el valor de la variable X, esta incertidumbre es 0, pues dado un valor de X, el valor de Y está determinado por su propia definición. En efecto, podemos aplicar la expresión:

$$H(X,Y) \ = \ \sum_{x,y} p(x,y) log_2 \frac{1}{p(x,y)} \qquad , \ p(x,y) \ = \ p(y)p(x|y)$$

y como

$$\begin{array}{lclclcl}
p(x=1,y=0) & = & p(y=0)p(x=1|y=0) & = & \frac{1}{2}\,0 & = & 0 \\
p(x=2,y=0) & = & p(y=0)p(x=2|y=0) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
p(x=3,y=0) & = & p(y=0)p(x=3|y=0) & = & \frac{1}{2}\,0 & = & 0 \\
p(x=4,y=0) & = & p(y=0)p(x=4|y=0) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
p(x=5,y=0) & = & p(y=0)p(x=5|y=0) & = & \frac{1}{2}\,0 & = & 0 \\
p(x=6,y=0) & = & p(y=0)p(x=6|y=0) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
 & & & & & & \\
p(x=1,y=1) & = & p(y=1)p(x=1|y=1) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
p(x=2,y=1) & = & p(y=1)p(x=2|y=1) & = & \frac{1}{2}\,0 & = & 0 \\
p(x=3,y=1) & = & p(y=1)p(x=3|y=1) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
p(x=4,y=1) & = & p(y=1)p(x=4|y=1) & = & \frac{1}{2}\,0 & = & 0 \\
p(x=5,y=1) & = & p(y=1)p(x=5|y=1) & = & \frac{1}{2}\,\frac{1}{3} & = & \frac{1}{6} \\
p(x=6,y=1) & = & p(y=1)p(x=6|y=1) & = & \frac{1}{2}\,0 & = & 0
\end{array}$$

queda que

$$H(X,Y) \ = \ 0 + \frac{1}{6}\log_2 6 + 0 + \frac{1}{6}\log_2 6 + 0 + \frac{1}{6}\log_2 6 + 0 + \frac{1}{6}\log_2 6 + 0 + \frac{1}{6}\log_2 6 + 0 + \frac{1}{6}\log_2 6$$

$$H(X,Y) \ = \ \log_2 6 \ \sim \ 2{,}58$$

y como sabemos que $H(X,Y) \ = \ H(X) \ + \ H(Y|X) \ = \ H(Y) \ + \ H(X|Y)$ queda que

$$H(Y|X) \ = \ 0$$

$$H(X|Y) \ = \ H(X) - H(Y) \ \sim \ 1{,}58$$

EJERCICIO: Probar el principio de máxima entropía

Haremos uso del método de los multiplicadores de Lagrange. Debemos hallar la expresión de N probabilidades p_j que maximicen la función entropía,

$$S = \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i} = \frac{1}{\ln 2} \sum_{i=1}^{N} p_i \ln \frac{1}{p_i}$$

con las siguientes restricciones:

$$\sum_{i=1}^{N} p_i x_i = \mu \qquad\qquad \sum_{i=1}^{N} p_i = 1$$

La función Lagrangiana se puede expresar, por tanto, de la siguiente manera:

$$L\left(p_1, \ldots, p_N, \eta, \lambda\right) = \sum_{i=1}^{N} p_i \ln \frac{1}{p_i} + \eta \left(\mu - \sum_{i=1}^{N} p_i x_i \right) + (\lambda - 1) \left(1 - \sum_{i=1}^{N} p_i \right)$$

(NOTA: se hace uso del factor $\lambda - 1$ para simplificar los cálculos)

Derivando respecto de cada variable, e igualando a 0, queda:

$$-\frac{\partial L}{\partial p_j} = 0 \;\Rightarrow\; \ln p_j + 1 + \eta x_j + (\lambda - 1) \;\Rightarrow\; p_j = e^{-\lambda} e^{-\eta x_j}$$

o sea, que podemos expresar las probabilidades como funciones exponenciales de los valores de x_i:

$$p_i = A e^{\alpha x_i} \qquad \textit{con} \;\; A = e^{-\lambda} \;\; \textit{y} \;\; \alpha = -\eta$$

Para obtener los valores de λ y η hacemos uso de las dos restricciones:

$$\sum_{i=1}^{N} p_i = 1 \;\Leftrightarrow\; e^{-\lambda} \sum_{i=1}^{N} e^{-\eta x_i} = 1$$

$$\sum_{i=1}^{N} p_i x_i = \mu \;\Leftrightarrow\; e^{-\lambda} \sum_{i=1}^{N} e^{-\eta x_i} x_i = \mu$$

De la primera ecuación podemos despejar λ como una función de η:

$$\lambda = \ln \sum_{i=1}^{N} e^{-\eta x_i}$$

y el valor de η se obtendría resolviendo la ecuación que se obtiene tras dividir las dos expresiones anteriores:

$$\frac{\sum\limits_{i=1}^{N} e^{-\eta x_i} x_i}{\sum\limits_{i=1}^{N} e^{-\eta x_i}} = \mu \;\Rightarrow\; \sum_{i=1}^{N} e^{-\eta x_i} (x_i - \mu) = 0$$