Códigos

Estímulo del Talento Matemático

Real Academia de Ciencias

8 de mayo de 2004

1. Transmitiendo información

La comunicación es tan vieja como el hombre. . .

Hoy nos encontramos en la era de la información o, más bien, de la transmisión de la información: millones de ordenadores están conectados a Internet, millones de correos electrónicos circulan cada día, millones de mensajes de texto se envían de un móvil a otro. . .

Todo esto es muy reciente, hace menos de 150 años que se inventó el teléfono y la revolución de Internet no empezó hasta los años 70.

¿Qué es lo que ha hecho posible esta revolución?

Una *Teoría matemática de la comunicación*, justo el título de un trabajo de C. Shannon en 1948, un ingeniero y matemático norteamericano que puso los cimientos de esta teoría.

Entonces, ¿qué es la información? Algo que se transmite por algún medio, o algo que se almacena, por ejemplo, en el disco duro de un ordenador, pero . . . ¿en qué unidades se mide? Shannon propuso medirla en *bits*. Un bit (viene del inglés: *binary unit*) puede tomar dos valores 0 y 1 (como un interruptor, que puede estar encendido o apagado). En nuestros días, todo (o casi todo) es *digital*, es decir, representado por símbolos separados, los bits, por ejemplo.

Si queremos responder a un mensaje con sí o no, nos basta con un bit que puede ser un 0 ó un 1; con dos bits, ya podemos dar cuatro respuestas diferentes 00, 01, 10 y 11; tres bits nos proporcionan ocho: 000, 001, 010, 100, 011, 110, 101 y 111; así sucesivamente. ¿Cuántas pueden representar n bits?

Los ruidos

Pero en cuanto una información se transmite por cualquier tipo de $canal$, aparece el $ruido$, las interferencias: en las transmisiones por cable o en las de radio, el medio está plagado, por ejemplo, de partículas cargadas que alteran las señales.

Esta actividad, que es totalmente $aleatoria$, o sea, depende del azar, puede introducir errores graves en la información que llega al receptor (un sí en vez de un no: ¡glub!, "debes ir al oeste" en lugar de "debes ir al este": ¡te pierdes!, etc.)

Un modelo sencillo de canal

Pensemos en la transmisión de un bit: con cierta probabilidad, digamos p, el mensaje llegará correctamente y con probabilidad $1 - p$ se recibirá incorrectamente (un 1 cuando se envió un 0 y un 0 cuando fue enviado un 1). Este modelo se conoce como $canal\ binario\ simétrico$.

Además, siempre podemos suponer que $p < 0{,}5$ (¿por qué?).

¿Qué valores de p te parece que será razonable considerar?

¿Te parece éste un modelo realista?

Luego experimentaremos con Excel "enviando" una lista de ceros y unos y estudiando los errores que se producen con diversas probabilidades p.

Códigos

Lógicamente queremos detectar (y, si es posible, corregir) cuándo se ha producido un error en la transmisión. ¿Cómo?

¿Qué hacemos en la vida diaria cuando alguien no entiende bien nuestro apellido?. . . se lo repetimos. Pues ésa puede ser una solución, repetir cada símbolo del mensaje, es decir, añadir $redundancia$ enviando, por ejemplo, 000 en lugar de 0 y 111 en vez de 1.

Por ejemplo, si el mensaje que tenemos que enviar es 01101, lo codificaríamos como

$$000\,111\,111\,000\,111$$

Una vez codificado, lo enviamos por el canal.

A este proceso de transformar el mensaje original para tratar de aumentar la fiabilidad de la transmisión lo llamamos $codificación$. En el ejemplo anterior, codificamos con un $código\ de\ repetición\ 3$, porque repetimos 3 veces cada elemento del mensaje original.

La lista de ceros y unos que recibamos (probablemente, no idéntica a la enviada), habrá que interpretarla de alguna manera ($descodificación$) para, si todo funciona, recuperar el mensaje original, 01101.

Luego, con ayuda del ordenador, comprobaremos que codificando el mensaje de esta manera, la transmisión es (muuuucho) más fiable.

Pero observemos también que el mensaje será el triple de largo.

Ejemplo 1 Para comunicarnos con nuestros semejantes "codificamos" nuestros mensajes (pensamientos) en algún idioma, por ejemplo, el castellano. Recibimos el siguiente mensaje:

La bisidleta es verde.

¿Por qué sabemos que nos llega alterado?, es decir, ¿cómo detectamos que hay errores?

En este caso, cualquiera diría que el mensaje que realmente fue emitido era

La bicicleta es verde.

¿Cuál es la razón por la que hemos podido *corregir* el error?

Pensemos ahora en el mensaje:

La cafa es grande.

¿Ha habido algún error?

¿Podemos corregirlo?

¿Cuál es la diferencia con el caso anterior?

El castellano (y cualquier otro idioma "natural") tiene incluido, casi como un accesorio de serie, un montón de redundancia.

Seguro que eso lo sabes: ¿cómo escribes, en un mensaje en el móvil, las palabras

- porque

- también
- para
- . . .

Entonces, ¿por qué escribimos (generalmente) "porque" en lugar de "xq"?

Ejemplo 2 Un robot se mueve por la superficie de Marte dirigido desde una estación espacial. Es necesario poder enviar cuatro mensajes diferentes: norte, sur, este y oeste. Sabemos ya que bastan dos bits para ello. Por ejemplo:

$$N = 00, \quad S = 11, \quad E = 10 \quad \text{y} \quad O = 01$$

Si se envían tal cual pero llegan alterados por culpa de las interferencias, el desastre puede ser mayúsculo.

Supongamos que se le quiere ordenar "vete al sur" (se envía 11), pero el mensaje se altera, al robot le llega 10 y se va hacia el este tan tranquilo, sin que pueda detectar el error . . .

Está claro que conviene codificar el mensaje. Hagamos lo siguiente, añadimos redundancia al mensaje de la siguiente manera: si los dos bits del mensaje son iguales, añadimos un 0 y si son distintos, un 1. Por ejemplo, codificamos 11 como 110, etc. Al bit que añadimos se le llama *comprobador de paridad*, ¿te imaginas por qué?

¿Puede el robot detectar ahora que se ha producido un error en el mensaje? Y si lo detecta, ¿puede corregirlo?

Si se produjeran dos errores, ¿qué ocurriría? ¿Por qué?

Distancia

Ya intuimos que un código puede corregir errores y detectarlos en función de lo separadas que estén las $palabras$ del código entre sí.

En el ejemplo anterior, las palabras del código que hemos usado son $000, 110, 101, 011$. Cada una se diferencia de cualquiera de las restantes al menos en dos dígitos o lugares, por eso, si en la transmisión se produce un único error (sólo se altera un bit), lo que nos llega no es una palabra del código y entonces detectamos el error. ¿Y si se producen dos errores? Haz un dibujo para convencerte.

En esta situación diremos que el código tiene distancia 2 (la mínima entre dos palabras cualesquiera del código).

¿Cuántos errores detectaríamos con un código de distancia 3? ¿Y si la distancia es 4? ¿Cuál es la regla en general?

Volvamos a nuestro código de distancia 2, con él podemos detectar un error, y ahora nos gustaría corregirlo. Si se envía, por ejemplo, 000 y se recibe 010, sabemos que hay un error. Para corregirlo bucamos la palabra del código más cercana a 010 (como hacíamos con "bisidleta" en el ejemplo del castellano). ¿Qué ocurre?

Ejemplo 3 En el ejemplo anterior, el problema (para corregir) es que no hay suficiente distancia entre las palabras del código (como ocurría con casa, cama y cala en el ejemplo del castellano).

Pensemos, por ejemplo, en el código de repetición de orden 3: los mensajes 0 y 1 se codifican a 000 y a 111 respectivamente. La distancia (mínima, y en este caso única) del código es 3. ¿Cuántos errores se pueden detectar?

¿Y corregir? (Observa que esto requiere establecer un criterio de corrección). Supongamos que se recibe 100. ¿Cómo "descodificarías"? ¿Por qué?

Escribe cómo descodificarías

- 000
- 010
- 001
- 110
- 101
- 011
- 111

Intenta ahora construir el resultado general. Un cierto código tiene distancia mínima d. Supongamos, por comodidad, que es un número impar, $d = 2e + 1$.

¿Cuántos errores es capaz de detectar?

¿Cuántos será capaz de corregir?

¿Y si se cometen más errores?

Ejemplo 4. Códigos a nuestro alrededor

La mayor parte de los productos que consumimos a diario están identificados por medio de un número que aparece impreso en el exterior del envase y acompañado de un **código de barras** que hace posible su lectura por un escáner.

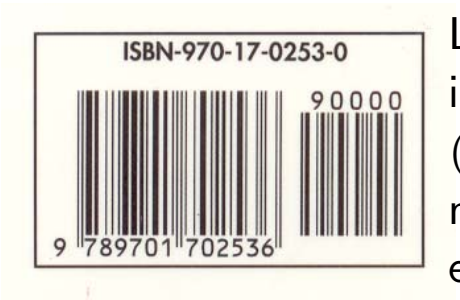
El sistema EAN-13 identifica cada producto con trece cifras,

$$a_1\, a_2\, a_3\, a_4\, a_5\, a_6\, a_7\, a_8\, a_9\, a_{10}\, a_{11}\, a_{12}\, a_{13}$$

Los primeros siete números contienen la identificación del fabricante, en tanto que el número de referencia del producto está formado por las cinco siguientes. El último dígito, a_{13} se calcula de forma que

$$(a_1+a_3+a_5+a_7+a_9+a_{11})+3\,(a_2+a_4+a_6+a_8+a_{10}+a_{12})+\mathbf{a_{13}}$$

sea 0 (pero en la aritmética módulo 10).



La mayoría de los libros modernos están identificados por un número de diez cifras $(a_1\,,\ldots,a_{10})$ conocido como ISBN. Las nueve primeras cifras dan información sobre el libro (editorial, título, edición, etc.) y la última, a_{10}, es un carácter de control que se calcula para que

$$a_1 + 2a_2 + 3a_3 + 4a_4 + 5a_5 + 6a_6 + 7a_7 + 8a_8 + 9a_9 = a_{10}$$

pero ahora en la aritmética módulo 11. Uno de los posibles resultados es que a_{10} sea 10; en ese caso, se utiliza la letra X.

Por último, el NIF se forma añadiendo al DNI una letra final. Se calcula el valor del DNI módulo 23 y luego se busca el resultado en la siguiente (extraña) tabla:

0	1	2	3	4	5	6	7	8	9	10	11
T	R	W	A	G	M	Y	F	P	D	X	B

12	13	14	15	16	17	18	19	20	21	22
N	J	Z	S	Q	V	H	L	C	K	E

¡Comprueba el tuyo!

Práctica de simulación con EXCEL