Universidad de Valladolid

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Matemáticas

Información, Codificación y Malabares:

Claude E. Shannon

Autor: Pablo Lorenzo Vaquero

Tutor: Antonio Campillo López

Índice general

Introducción

El año pasado, se conmemoró el centenario del nacimiento del matemático **Claude Elwood Shannon** (1916-2001), conocido por ser considerado el "*Padre de la Teoría de la Información*", siendo de esta forma una de las mentes más influyentes del siglo XX. Con este Trabajo, se plantea sumarnos a la conmemoración hacia esta mente brillante, siendo el objetivo del trabajo la revisión e investigación en algunos de los modelos y resultados de Shannon, así como proporcionar una versión actual de las teorías y resultados matemáticos que expondremos.

Hemos de entender a Shannon como un hombre completamente adelantado a su tiempo, que desarrolló los resultados más importantes de su teoría en su trabajo en *Bell System Laboratories*. Cabe destacar, para remarcar la importancia del momento, que en el mismo laboratorio y momento se inventó también el transistor. Shannon se centra en el estudió de la información, creando la teoría básica que subyace en la actualidad en todas las comunicaciones digitales y los métodos de almacenamiento y procesamiento de la información. El impacto de esta investigación en el mundo práctico es infinitamente más importante en la actualidad que en su momento, hace 60 años. Cuando nadie se cuestionaba acerca de su utilidad, Shannon desarrolló fórmulas con las que medir la información y su transmisión; con que velocidad y que cantidad se puede transmitir sobre diferentes medios.

Todas estas ideas aparecen recogidas en 1948 en el artículo *A Mathematical Theory of Communication*, con el que revoluciona la forma en que ingenieros y científicos del momento conciben las comunicaciones digitales. En este texto, considerado la "*Carta Magna de la Era de La Información*", sienta las bases de la Teoría de la Información fijando los conceptos de mensaje, emisor, transmisor, canal, receptor y destino tal como los conocemos ahora.

En la primera parte de este trabajo se ha realizado una revisión de este artículo, complementado con otro de sus artículos cercanos en temática y tiempo: *Communication in the Presence of Noise*, donde amplía y desarrolla la teoría en el caso concreto de canales ruidosos.
En este **primer capítulo** introduciremos el sistema de comunicaciones estándar, distinguiendo y estudiando por separado los casos en los que los mensajes a comunicar sean secuencias discretas o continuas, así como la presencia o no de ruido que altere las señales durante la comunicación. En este sentido, desarrollaremos conceptos de gran importancia en toda la Teoría de la Información, definiendo y demostrando conceptos que son desde ese momento universales manteniendo su fórmula y definición, destacando la *Capacidad* de un canal,

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T},$$

la *Entropía* (definida también como Fórmula de Shannon para medir la información)

$$H = -\sum p(x) \log p(x),$$

los ratios de información o el *bit* como unidad básica de información.

A través del desarrollo de estos conceptos, podremos concluir con los *Teoremas Fundamentales* respecto a cada caso, donde Shannon estableció en cada uno de ellos los límites en los que se puede realizar la comunicación. En este sentido, el matemático ya comprendió que independientemente del desarrollo tecnológico, la información no puede ser transmitida más allá de cierto ratio, un límite fundamental conocido como *Límite o Capacidad de Shannon.*

Precisamente, uno de los grandes problemas a resolver que Shannon enuncia con esta Teoría es la posibilidad de alcanzar una transmisión con la máxima efectividad; situándonos en el límite posible. En su momento, el intercambio de información era mucho más lento que los límites posibles que él mismo enunció, por lo que crea un desafío que los propios ingenieros no comenzaron de resolver hasta 50 años más tarde, siendo en la actualidad un motivo constante de estudio la construcción de formas de comunicarse que alcancen el límite de la capacidad establecida. Ciertamente, la Teoría de Shannon enuncia en que situaciones se pueden construir estas formas de comunicación, pero no habla del cómo. Sobre este aspecto tratará el **segundo capítulo** de este trabajo.

En este, basado principalmente en el libro *A Course in Error-Correcting Codes* de Jørn Justesen y Tom Høholdt, nos introduciremos en cuestiones algebraicas sobre la forma en comunicar mensajes, su codificación, procesamiento y corrección de errores. Para ello, se tratará de dar una versión actualizada y bastante general de estas formas, desarrollando los conceptos de códigos correctores lineales, estructuras de comprobación de paridad, peso Hamming, descodificaciones, etc. Así mismo, se buscarán las cotas de probabilidad de error para errores y fallos de descodificación, según si si se produce una alteración del mensaje y si esta alteración no es corregida respectivamente.

Posteriormente, se tratará de dar una visión más actual con ciertas anotaciones referidas a la Teoría de la Información, donde se redefine los conceptos de Entropía, Canal o se incluye la Información mutua.

Para terminar, se hablará de nuevas formas de comunicación que se están desarrollando en los últimos años para las mejoras de la comunicación. En este sentido, introduciremos el concepto de *Estructura de Datos* como una forma de almacenar y transmitir la información de manera conjunta; así como finalizaremos con un breve acercamiento a los códigos convolucionales, que se caracterizan por ser códigos en bloques de longitud no fija. Por ejemplo, la NASA o la Agencia Europea del Espacio utilizan estos códigos.

Para finalizar de dar una visión más completa a todas las aportaciones de Shannon en las matemáticas, es necesario destacar que realizo importantes aportaciones en distintas materias como el desarrollo de circuitos integrados, ordenadores, criptografía, inteligencia artificial o incluso genómica. No en vano, es considerado una de las últimas figuras renacentistas, pues su interés se ocupó de diversas áreas del saber y las ciencias. También creo modelos matemáticos y máquinas en el ámbito lúdico, como máquinas de ajedrez, uniciclos, robots capaces de aprender por sí solos (precursores de la Inteligencia Artificial como técnica) o malabares.

Precisamente, para hacer honor a estas otras aportaciones también destacables fuera de la Teoría de la Información, el **tercer capítulo** se dedica al último tema que hemos referido, el estudio de los Juegos Malabares desde un punto de vista matemático. En esta sección, se desarrollan diferentes modelizaciones de un malabar en función de la cantidad de manos y objetos en el aire, así como el caso particular de los malabares uniformes (objeto principal de estudio de Shannon). Tomando como referencia el libro *The Mathematics of Juggling* de Burkald Polster,

así como el propio artículo de Shannon *Scientific Aspects of Juggling* se introducirá la notación "SITESWAP" como modelo matemático para definir los lanzamientos de diferentes alturas, a partir de las funciones y secuencias o matrices malabares respectivamente para los casos de malabar simple (una persona, dos manos) o múltiples manos. Posteriormente, se enunciarán y demostrarán resultados principales como el *Teorema de la Media* para el número de bolas o el *Test de Permutación* para comprobar si una secuencia de números es o no malabar. Para la demostración de estos resultados nos hemos servido del artículo *Juggling drops and descents* escrito por Joe Buhler, David Eisenbud, Ron Graham y Colin Wright

A continuación, se estudiará la forma de crear todas las secuencias malabares posibles, así como las nociones de estados malabares y grafos de estado para establecer las conexiones entre todas las secuencias de un número de bolas concreto. Para finalizar, se introducirán los malabares uniformes, característicos por su constancia en los tiempos de vuelo , permanencia y vacante. En estos malabares precisamente desarrollaremos los tres *Teoremas Malabares de Shannon*, que relacionan el número de manos y bolas disponibles con los tiempos indicados (en el caso del primer Teorema) y con las esencialmente diferentes maneras posibles de realizar los malabares (Segundo y Tercer Teorema).

Capítulo 1

Teoría de la Información

En este capítulo vamos a introducir los principales conceptos incluidos en la comunicación de mensajes, que ya anticipó Shannon en 1948, creando una teoría básica para entender las comunicaciones digitales, así como el almacenamiento y procesamiento de la información.

El principal objetivo de una comunicación es reproducir en un lugar dado el mensaje exactamente igual al que se ha producido en otro punto. Para ello, se entiende que el mensaje tiene (en la mayoría de los casos) un *significado*, es decir, se refiere o está relacionado con un cierto sistema de entidades físicas o conceptuales, no consiste en la mera transmisión de símbolos. La importancia del significado, a pesar de que va a ser irrelevante desde el punto de vista matemático, es notoria: El objetivo principal de la transmisión del mensaje se basa en la transmisión de dicho significado, que podría ser alterado por una comunicación inexacta o incompleta.

Además, suponemos que el mensaje es seleccionado entre un conjunto de mensajes de posibles. No obstante, el sistema de comunicación debe estar preparado para operar con cualquier posible selección, pues el o los mensajes que se enviarán son desconocidos al momento de diseñar el sistema.

Un concepto de gran importancia será nuestra capacidad de medir la información, de cara a poder determinar la cantidad de información que contiene un mensaje, cuanta puede ser enviada o recibida en un momento de tiempo, etc. En este sentido, la opción más natural para medir la información está relacionada con la función logaritmo, en tanto que una medida logarítmica es la opción más práctica matemáticamente (valórese la linealidad del logaritmo frente a parámetros como el tiempo, ancho de banda, número de transmisiones, etc.) y encaja con nuestra intuición de la medida (parece lógico pedir que dos canales idénticos dupliquen la capacidad de uno de transmitir la información, por ejemplo).

De la elección de la base del logaritmo dependerá nuestra unidad de medida de la información. Así, dado que en la mayoría de los casos la transmisión realiza a través de un canal digital mediante un sistema binario (base 2), nuestra unidad de información será el *bit*. En este aspecto, cabe destacar la gran aportación que realizó Shannon a este concepto, cambiando la definición que se tenía hasta el momento, perteneciente a Tukey. A partir de Shannon, un bit ya no se refiere a una unidad de almacenamiento que guarda un dígito binario 0 ó 1 (aunque se correspondería con ellos en el sistema binario); sino que se trata de la unidad básica de información que nos permitirá su medida. Si partimos de un sistema decimal, por ejemplo, dado que

$$\log_2 M = \frac{\log_{10} M}{\log_{10} 2} \approx 3{,}32 \log_{10} M$$

y entonces una cifra digital decimal equivale a 3,32 bits aproximadamente.

Un sistema de comunicación se refiere a un sistema del tipo indicado en la Figura 1.1 en

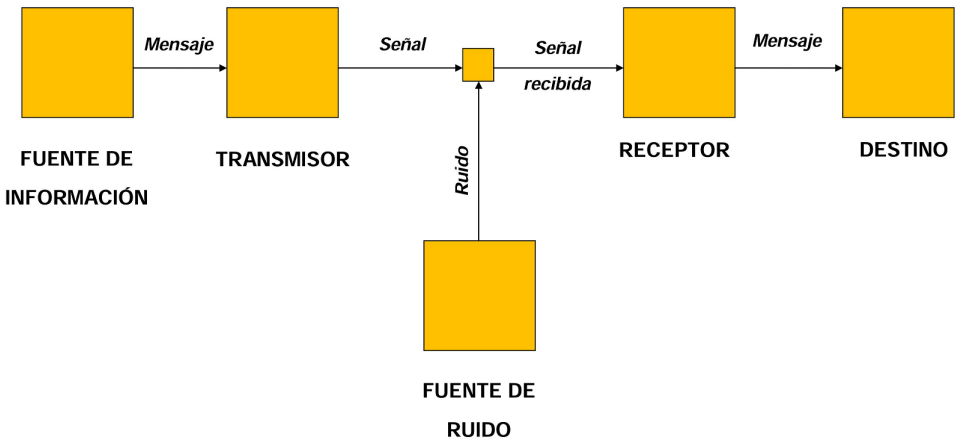

Figura 1.1: Diagrama de un sistema de comunicación general

forma de esquema. De manera general, consiste esencialmente en cinco partes:

1. Una *fuente de información*, que produce el mensaje que se quiere comunicar. Este mensaje puede ser de varios tipos

   a) Una secuencia de letras o símbolos, como por ejemplo en un telégrafo.

   b) Una función $f(t)$ respecto al tiempo, como en una radio o un teléfono.

   c) Una función de varias variables. Por ejemplo, en un televisor en blanco y negro, la función $f(x, y, t)$ indicaría la intensidad de la luz en el tiempo t en una posición (x, y).

   d) Varias funciones de una (tiempo) o varias variables. Por ejemplo, en una televisión a color el mensaje consiste en tres funciones $f(x, y, t)$, $g(x, y, t)$ y $h(x, y, t)$, que indican la intensidad de cada color de un sistema RGB en el tiempo y posición dadas.

   e) Cualquier combinación de las anteriores, como se daría en en una televisión con audio, por ejemplo.

2. Un *transmisor*, que trabajará con el mensaje produciendo una señal capaz de ser enviada a través del canal. Este hecho (transformar la información de entrada en información para el canal) lo conoceremos como *codificación*; siendo de gran importancia, como veremos más adelante el tipo de código que se utilice.

3. El *canal*, el medio que se usa para transmitir la señal desde el transmisor hasta el receptor. Sus cualidades, como el ancho de banda o la capacidad del mismo, son determinantes en la comunicación.

4. El *receptor*, que recibe la señal y la transforma de nuevo en el mensaje, *descodificandola*.

5. El *destino* del mensaje, hacia donde se ha enviado.

6. En algunos casos, puede ocurrir que la señal enviada sea perturbada a través del canal o en alguna de las terminales. Nos referimos a este hecho como *ruido*. En este caso, la señal recibida puede ser diferente de la enviada, por lo que serán necesarios diferentes mecanismos para recuperar el mensaje inicial, los cuales también trataremos más adelante.

Podemos clasificar los sistemas de comunicación en tres categorías: Discretos, en los que el mensaje y la señal estarán formados por una secuencia discreta de símbolos; continuos, si se tratarán de funciones continuas; o mixtos, con variables continuas y discretas. Empecemos con el caso discreto.

1.1. Sistemas discretos sin ruido

En un caso discreto, el mensaje que se envía a través del canal es una secuencia de símbolos escogidos entre el conjunto finito de símbolos $S_1,\ S_2,\ \ldots, S_n$ que pueden ser transmitidos de un punto a otro. Cada símbolo S_i requiere de un tiempo t_i para ser transmitido.

Definición 1.1.1. *La* capacidad *de un canal viene dada por*

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T} \tag{1.1}$$

donde $N(T)$ representa el número de señales permitidas de duración T.

La capacidad de un canal indica la cantidad de información que puede ser transmitida a través de dicho canal simultáneamente. Si suponemos que todas las señales de símbolos $S_1,\ S_2,\ \ldots, S_n$ están permitidas, tenemos que $N(T)$ será la suma de de los números de secuencias terminadas en cada símbolo respectivamente, es decir,

$$N(T) = N(T - t_1) + N(T - t_2) + \ldots + N(T - t_n)$$

Resolviendo la ecuación en diferencias, $N(T)$ se comporta asintóticamente como X_0^t, siendo X_0 la solución real de mayor valor de la ecuación característica

$$X^{-t_1} + X^{-t_2} + \ldots + X^{-t_n} = 1$$

y por lo tanto

$$C = \log X_0$$

De cara a las secuencias posibles, se supone que hay un número de posibles estados $a_1,\ a_2,\ \ldots,\ a_m$ para los cuales sólo hay un subconjunto del conjunto de símbolos que pueden ser transmitidos desde ese estado. Cada vez que un símbolo es transmitido, el estado cambia en función del anterior estado y el símbolo transmitido. En estas condiciones, podemos calcular la capacidad:

Teorema 1.1.1. *Sea $b_{ij}^{(s)}$ la duración del s-ésimo símbolo, partiendo del estado i para pasar al estado j. Entonces la capacidad C del canal es igual a* $\log W$*, el logaritmo de la mayor solución real de la ecuación*

$$\left| \sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \right| = 0$$

Donde $\delta_{ij} = 1$ si $i = j$ y cero en otro caso.

Demostración. Con la misma notación, sea $N_i(T)$ el número de secuencias de símbolos de duración T que terminan en el estado i. Entonces

$$N_j(T) = \sum_{i,s} N_i(T - b_{ij}^{(s)})$$

Al ser una ecuación lineal en diferencias, el comportamiento cuando $T \to \infty$ es de la forma $N_j = A_j W^T$, para ciertos A_j y W reales. Sustituyendo,

$$A_j W^T = \sum_{i,s} A_i W^{L-b_{ij}^{(s)}} \;\Rightarrow\; A_j = \sum_{i,s} A_i W^{-b_{ij}^{(s)}} \;\Rightarrow\; \sum_i \left(\sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \right) A_i = 0$$

Pero para que este último se cumpla, el determinante $D(W) = |a_{ij}| = \left| \sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \right|$ debe ser nulo, lo que nos determina W, que será la mayor raíz real de $D = 0$.
Entonces C viene dado por

$$C = \lim_{T \to \infty} \frac{\log \left(\sum_j A_j W^T \right)}{T} = \log W$$

□

Una vez conocida la capacidad del canal, parece lógico estudiar la fuente de información, de forma que nuestro objetivo es describirla de forma matemática en la manera que la información estadística que podamos obtener de ella nos ayude a reducir la capacidad requerida por el canal para transmitir dicha información. Una opción eficaz es codificar la información de forma que las letras o conjuntos de letras más usados tengan una codificación tal que la señal de envío sea más corta o de menor información, lo que produce un gran ahorro de tiempo y capacidad. En el idioma castellano, por ejemplo, la letra *E* es mucho más usada que la *X*, o una estructura (por ejemplo, sílabas) del tipo *TE* es más frecuente que *IU*. En este sentido podemos considerar que una fuente discreta genera el mensaje, símbolo por símbolo, teniendo cada uno una probabilidad que depende de las elecciones anteriores, el estado del canal o los propios símbolos del canal. Podemos interpretar entonces la generación como un proceso estocástico, donde cada símbolo es producido en función de una serie de probabilidades.

El caso bastante general puede ser descrito como un proceso de Markov: Existe un número finito de estados del sistema: $a_1, a_2, \ldots, a_m$ con un conjunto de probabilidades, $p_i(j)$, que indica la probabilidad de ir del estado a_i al estado a_j. En este sentido, en el proceso de comunicación supondremos que un símbolo es producido en cada transición de un estado a otro.

Entre los procesos de Markov, hay un grupo con propiedades especiales que merece la pena destacar. Se trata de los procesos *ergódicos*, en donde nos referiremos a las fuentes como ergódicas. En un proceso ergódico, cada secuencia producida tiene las mismas propiedades estadísticas. Por lo tanto, estas propiedades estadísticas pueden ser deducidas a partir de un único y simple ejemplo aleatorio suficientemente largo del proceso, pues la probabilidad del conjunto donde estas propiedades no se cumplen es cero, con lo que tenemos una homogeneidad estadística. Esta propiedad está relacionada con el grafo del proceso, pues será ergódico si:

1. El grafo es conexo, es decir, no pueden ser dividido en dos o más partes de forma que dichos subgrafos no tengan ninguna arista que los conecte.

2. El máximo común divisor de la longitud de todos los circuitos integrados en el grafo es uno.

Si la segunda condición no se cumple, teniendo un máximo común divisor $d > 1$, entonces las secuencias tendrán una cierta estructura periódica, de forma que derivarán en d diferentes clases con propiedades estadísticas similares, salvo por alguna variación en su origen.
Si la primera condición y el grafo se puede dividir en subgrafos desconectados que sí cumplen las

dos condiciones, decimos que la fuente es *mixta*, hecha de varios componentes puros, $L_1, L_2, \ldots$, correspondiendo cada componente con uno de los subgrafos. Entonces podemso escribir la fuente como

$$L = p_1 L_1 + p_2 L_2 + p_3 L_3 + \ldots$$

donde p_i es la probabilidad del componente L_i respectivamente. Esta situación representa la existencia de diferentes fuentes, cada una con una estructura estadística homogénea propia (son ergódicas). A pesar de no saber cuál será escogida, una vez que una secuencia empieza en una fuente L_i, continuará indefinidamente de acuerdo a la estructura estadística de tal componente. A menos que se especifique lo contrario, en adelante supondremos que la fuente es ergódica.

El siguiente paso se trata de medir la cantidad de información producida por la fuente en este proceso; o el ratio de tal. Supongamos que tenemos una serie de eventos con probabilidades $p_1, p_2, \ldots, p_n$. Se quiere medir precisamente la cantidad de incertidumbre respecto a la salida de la fuente. Para ello, se busca una medida $H(p_1, p_2, \ldots, p_n)$ que cumpla:

1. H sea continua respecto a p_i.

2. Si todas las probabilidades son iguales, $p_i = \frac{1}{n}$ entonces H debe ser creciente monótona respecto a n (más eventos posibles, mayor será la incertidumbre).

3. Si una decisión puede ser dividida en dos decisiones consecutivas, el total de H deberá ser la suma de los valores para cada H particular. Esta propiedad es fácil de ver con un ejemplo: supongamos que tenemos tres eventos de posibilidades $\frac{1}{2}$, $\frac{1}{3}$ y $\frac{1}{6}$ respectivamente. Este proceso se podría descomponer en dos elecciones, la primera con dos posibilidades de probabilidad $\frac{1}{2}$, y en el caso de que la segunda ocurra, hacer otra elección con probabilidades $\frac{2}{3}$ y $\frac{1}{3}$. Las probabilidades finales son las mismas en ambos casos, por lo que se requiere que

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right)$$

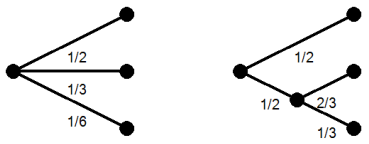

Figura 1.2: Descomposición de las posibilidades

Pidiendo estas condiciones, podemos establecer:

Teorema 1.1.2. *La única función H que cumple las condiciones establecidas es de la forma*

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{1.2}$$

donde K es una constante positiva.

Demostración. Sea $H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = A(n)$. De la tercera condición, se puede descomponer s^m posibilidades de igual probabilidad en series de m elecciones, cada una con s posibilidades de

igual probabilidad:

$$A(s^m) = H\left(\frac{1}{s^m}, \frac{1}{s^m}, \ldots, \frac{1}{s^m}\right) = H\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}\right) + s\left(\frac{1}{s}H\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}\right)\right) +$$
$$+s^2\left(\frac{1}{s^2}H\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}\right)\right) + \ldots + s^m\left(\frac{1}{s^m}H\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}\right)\right) = mH\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}\right) = mA(s)$$

De la misma manera, $A(t^n) = nA(t)$.
Escogiendo n arbitrariamente grande, tomamos m tal que

$$s^m \leq t^n \leq s^{(m+1)}$$

Tomando logaritmos y dividiendo entre $n \log s$

$$m \log s \leq n \log t \leq (m+1) \log s$$

$$\frac{m}{n} \leq \frac{\log t}{\log s} \leq \frac{m}{n} + \frac{1}{n}$$

Luego

$$\left|\frac{m}{n} - \frac{\log t}{\log s}\right| < \epsilon$$

Con ϵ arbitrariamente pequeño (dependerá de n). Análogamente, debido por la segunda condición (crecimiento monótono de $A(n)$) tenemos que:

$$A(s^m) \leq A(t^n) \leq A(s^{m+1})$$

$$mA(s) \leq nA(t) \leq (m+1)A(s)$$

Dividiendo por $nA(s)$,

$$\frac{m}{n} \leq \frac{A(t)}{A(n)} \leq \frac{m}{n} + \frac{1}{n} \quad \text{o} \quad \left|\frac{m}{n} - \frac{A(t)}{A(n)}\right| < \epsilon$$

Y del anterior obtenemos que

$$\left|\frac{A(t)}{A(n)} - \frac{\log t}{\log s}\right| < 2\epsilon$$

concluyendo

$$A(t) = K \log t$$

con K constante y positiva para satisfacer la segunda condición.
Ahora supongamos que tenemos n posibilidades de probabilidades $p_i = \frac{n_i}{\sum_{i=1}^{n} n_i}$, con n_i entero. Se puede descomponer una elección entre las $\sum n_i$ posibilidades en una elección de n posibilidades con probabilidades $p_1, p_2, \ldots, p_n$; tal que si la i-ésima es escogida, nos queda una elección con n_i posibilidades de igual probabilidad $\frac{1}{n_i}$ (nótese, que, en probabilidad final, todas las decisiones tienen la misma probabilidad, $\frac{1}{\sum n_i}$).
Usando de nuevo la tercera condición, se puede calcular H respecto del número total $\sum n_i$ de dos maneras diferentes:

$$H\left(\frac{1}{\sum n_i}, \frac{1}{\sum n_i}, \ldots, \frac{1}{\sum n_i}\right) = K \log \sum n_i$$

$$H\left(\frac{1}{\sum n_i}, \frac{1}{\sum n_i}, \ldots, \frac{1}{\sum n_i}\right) = H(p_1, p_2, \ldots, p_n) + K \sum p_i \log n_i$$

Por lo que

$$H = K\left[\sum p_i \log\left(\sum n_i\right) - \sum p_i \log n_i\right] = -K\sum p_i \log \frac{n_i}{\sum n_i} = -K\sum p_i \log p_i$$

□

En el caso de que p_i no sea medible, se puede aproximar mediante racionales y la misma expresión es válida debido a nuestra hipótesis de continuidad, por lo que la expresión (1.2) es válida en general. Por otro lado, el coeficiente K equivale a la unidad de medida, luego si usamos el logaritmo en base 2 supondremos que $K = 1$ para medir en bits.

Denominamos entonces H como la *entropía* de un sistema de probabilidades $p_1, p_2, \ldots, p_n$. Si tenemos una variable aleatoria x, denominaremos $H(x)$ a su entropía, a pesar de que x no es propiamente una variable de la función H. Conviene destacar de manera breve algunas propiedades:

1. $H \geq 0$ y $H = 0$ sí y solo sí todos los p_i son nulos, excepto uno con el valor unidad (es decir, no hay más de una elección)

2. Para un n dado, el máximo de H se alcanza si todas las probabilidades son iguales, $p_i = \frac{1}{n}$.

3. Si tenemos dos variables, x e y, con una probabilidad conjunta $p(i,j)$ (probabilidad de que ocurra el suceso i en el primer caso y j en el segundo), la entropía del evento conjunto será
$$H(x,y) = -\sum_{i,j} p(i,j) \log p(i,j)$$
con $H(x,y) \leq H(x) + H(y)$, obteniéndose la igualdad sólo en el caso de que las variables sean independientes.

4. Cualquier cambio de las probabilidades hacía la equidad aumentará la entropía.

5. De nuevo en un sistema de dos variables, siendo $p_i(j)$ la probabilidad condicionada de que y tenga valor j si x es i, podemos hablar de la entropía condicionada $H_x(y)$ aplicando tal probabilidad a (1.2). Además, $H(x,y) = H(x) - H_x(y)$.

Con todas estas propiedades, podemos definir finalmente la entropía de una fuente como el valor esperado de acuerdo a la probabilidad de que el sistema se encuentre en cada estado, es decir, la entropía por símbolo es

$$H = \sum_i P_i H_i = -\sum_{i,j} P_i p_i(j) \log p_i(j) \tag{1.3}$$

con P_i probabilidad de encontrarse en el estado i y $p_i(j)$ la probabilidad condicionada de pasar del estado i a j. Si el proceso se produce con unos tiempos definidos, podemos tenemos que la entropía por segundo corresponde a

$$H' = \sum_i f_i H_i$$

con f_i la frecuencia media de cada estado. Como es de esperar, se cumple que

$$H' = mH$$

siendo m el número de símbolos por segundo.

Sí los símbolos son independientes entre ellos, entonces simplemente $H = -\sum_i p_i \log p_i$, con p_i la probabilidad de un símbolo i. Para un mensaje lo suficientemente largo (longitud N), la probabilidad de este mensaje será cercana a $p = p_1^{p_1 N} p_2^{p_2 N} \cdots p_n^{p_n N}$; y podemos aproximar este valor mediante $H = \dfrac{\log p^{-1}}{N}$:

Teorema 1.1.3. *Dado cualquier $\epsilon > 0$ y $\delta > 0$, existe un valor N_0 tal que para cualquier secuencia de longitud $N \geq N_0$ pueden ocurrir dos sucesos:*

1. *La secuencia pertenece a un conjunto de de probabilidad menor que ϵ.*
2. *Siendo p la probabilidad de la secuencia, se cumple que*

$$\left| \frac{\log p^{-1}}{N} - H \right| < \delta$$

Una vez descrito matemáticamente la producción del mensaje en la fuente, abordemos el proceso de codificación y descodificación producido en el transmisor y receptor. En el transmisor, denominaremos *símbolos de entrada* (o simplemente entrada) a los producidos por la fuente, que llegan al transmisor; y *símbolos de salida* los producidos en el transmisor como resultado de la codificación, que pueden ser enviados a través del canal. El proceso de codificación puede tener o no *memoria*, de forma que la salida no esté condicionada sólo por el símbolo de entrada actual, sino por un número concreto de los últimos símbolos de entrada. Si tenemos una memoria finita, entonces habrá un número m de estados del transmisor en los que la salida dependerá del símbolo de entrada y el estado del transmisor en ese momento.
Representaremos la codificación como

$$\begin{aligned} y_n &= f(x_n, a_n) \\ a_{n+1} &= g(x_n, a_n) \end{aligned}$$

Siendo x_n el n-ésimo símbolo de entrada, a_n el estado del transmisor cuando el n-ésimo símbolo es introducido e y_n es el símbolo o secuencia de símbolos de salida correspondiente.

Teorema 1.1.4. Teorema Fundamental para un Canal Discreto sin Ruido
Sean H la entropía de una fuente (en bits por símbolo) y C la capacidad de un canal (en bits por segundo) dados. Entonces es posible codificar la salida de la fuente de manera que se trasmita una media de $\dfrac{C}{H} - \epsilon$ símbolos por segundo a través del canal, donde ϵ es arbitrariamente pequeño. No es posible transmitir a un ratio medio mayor que $\dfrac{C}{H}$.

Demostración. El límite es fácil de definir: basta con notar que la entropía de la entrada en el canal será igual que la de la fuente por segundo, siendo esta necesariamente menor que la capacidad del canal para que se pueda transmitir la información. Entonces $H' \leq C$ y en número de símbolos por segundo es $\dfrac{H'}{H} \leq \dfrac{C}{H}$. Veamos que esta cota puede ser prácticamente alcanzada: Si consideramos todas las secuencias de N símbolos producidas por la fuente, estas las podemos dividir en dos grupos: un grupo de alta probabilidad, que contendrá alrededor de 2^{HN} secuencias (dado que la media son H bits por símbolo, con N símbolos habrá una media de NH bits por secuencia, con lo que se pueden formar alrededor del número indicado). Para contener este grupo, tomamos un valor pequeño η de forma que el grupo de alta probabilidad tenga menos de $2^{(H+\eta)N}$ secuencias; y un segundo grupo de baja probabilidad, formado por menos de 2^{RN} (siendo R el logaritmo del número total de distintos símbolos existentes), de forma que la probabilidad de este segundo grupo sea menor que una cota μ. A medida que N crece, η y μ

tenderán a cero.
Si fijamos un tiempo T, tenemos que en el canal hay al menos $2^{(C-\theta)T}$ señales posibles con esa duración, con θ pequeño cuando T es lo suficientemente grande (pues nos acercamos al límite de capacidad del canal). Tomemos

$$T = \left(\frac{H}{C} + \lambda\right) N$$

entonces el número de señales a través del canal será mayor que

$$2^{(C-\theta)T} = 2^{(C-\theta)\left(\frac{H}{C}+\lambda\right)N} = 2^{(H+\eta(n))N}$$

siendo $\eta(N)$ una función que tiende a cero cuando N y T son lo suficientemente grandes. En este caso, habrá suficientes secuencias del canal de duración T de forma que se pueden asociar todos las secuencias del primer grupo con las del canal en una relación individual (una secuencia del canal para cada mensaje). El resto del mensajes (los de baja probabilidad) para los que no quede una señal de duración T, se representan mediante secuencias más largas, que empezarán y terminarán con una señal especial especificada. Para estos mensajes de baja probabilidad, se requerirá un tiempo

$$T_1 = \left(\frac{R}{C} + \varphi\right) N$$

con φ pequeño.
El ratio medio de transmisión en símbolos por segundo será mayor que

$$\left[(1-\delta)\frac{T}{N} + \delta\frac{T_1}{N}\right]^{-1} = \left[(1-\delta)\left(\frac{H}{C}+\lambda\right) + \delta\left(\frac{R}{C}+\varphi\right)\right]^{-1}$$

A medida que N aumenta, δ, λ y φ tienden hacia cero, por lo que este ratio se aproxima hacia $\frac{C}{H}$. □

A pesar de haber realizado algunos pasos para la construcción de un código durante la demostración, existen otras posibles codificaciones más efectivas. Una de las más usuales en este tipo de casos es, ordenando todos los mensajes de longitud N en orden decreciente de su probabilidad $p_1 \geq p_2 \geq \ldots \geq p_n$, se puede determinar para cada uno la probabilidad acumulada excluyéndose a sí mismo, $P_s = \sum_{i=1}^{s-1} p_i$ y codificando al sistema binario de forma que la transmisión a través del canal del mensaje sea la expansión de P_s en sistema binario.
Este método es fácil de ilustrar con un ejemplo sencillo: supongamos que tenemos 4 símbolos A, B, C, D con probabilidades $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ y $\frac{1}{8}$ respectivamente, que son elegidos de manera independiente. La entropía de este sistema es

$$H = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + 2\frac{1}{8}\log\frac{1}{8}\right) = \frac{7}{4} \text{ bits por símbolo.}$$

Si usamos el sistema de codificación descrito anteriormente, la codificación que obtenemos es

A	0
B	10
C	110
D	111

Si tomamos todos los mensajes de N símbolos, el promedio de bits usados en su codificación será

$$N\left(\frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{2}{8}\times 3\right) = \frac{7}{4}N$$

Y por lo tanto se alcanza la cota de H bits por símbolo.

1.2. Sistemas discretos con ruido

Como ya se ha mencionado previamente, entendemos como ruido la alteración que se produce en algún momento del proceso de comunicación de forma que la señal es modificada en tanto que los mensajes enviados y recibidos a través del canal no son necesariamente iguales. Hay dos casos principales:

1. La alteración es siempre la misma. Es decir, cada señal diferente se recibe siempre de la misma forma, que puede ser perturbada o no, pero esta perturbación es una función definida respecto a las señales enviadas. Conocemos este caso como *distorsión*. Si esta función tiene una inversa definida, entonces la distorsión es fácil de corregir: basta aplicar la función inversa a la señal recibida antes de comenzar con la descodificación.

2. La alteración no es constante. Este es el caso de interés pues es el más común y su solución no es a priori tan sencilla como en la distorsión.

La señal recibida E es entonces una función que depende de la enviada S y el ruido N:

$$E = f(S, N)$$

En general, el ruido es representado como un proceso estocástico, tomando una probabilidad condicionada $p_{\alpha,i}(\beta, j)$ que representa la probabilidad de, si el canal está en un estado α y se transmite el símbolo i, se recibirá el símbolo j quedando el canal en un estado β.
Dado que la fuente era también un proceso estocástico, tenemos diferentes entropías a tener en cuenta: La entropía de entrada al canal $H(x)$ que se corresponde con la de la fuente; y la entropía de salida, $H(y)$. En el caso del canal sin ruido se cumplía que $H(x) = H(y)$. A tener en cuenta serán también la entropía conjunta $H(x, y)$ y las condicionales, $H_x(y)$ y $H_y(x)$, en los casos en los que se conoce la entrada o salida respectivamente. En relación a todas estas medidas, conviene recordar las igualdades y desigualdad:

$$H(x, y) = H(x) + H_x(y) = H(y) + H_y(x) \leq H(x) + H(y)$$

Definición 1.2.1. *El* ratio de transmisión*, R, es el promedio de información verdadera que se trasmite mediante el canal, calculado*

$$R = H(x) - H_y(x) \tag{1.4}$$

donde $H(x)$ representa la entropía de la fuente, el ratio de producción de información; y $H_y(x)$ mide la ambigüedad o incertidumbre media de la información recibida, que denominaremos equivocación.

Otras formas de calcular R son

$$R = H(x) - H_y(x) = H(y) - H_x(y) = H(x) + H(y) - H(x, y)$$

En esta definición se ha tomado la entropía condicionada $H_y(x)$ como una medida de la información perdida. A partir de conocer el mensaje enviado, esta medida nos indicará la cantidad de información que no se corresponde. Para justificar esta medida, nos valemos del siguiente teorema:

Teorema 1.2.1. *Sea un sistema de comunicación del tipo de la Figura 1.3, con un observador (dispositivo exterior) que conoce tanto el mensaje enviado como el recibido, anotando los errores y transmitiendoselos a un dispositivo corrector a través de un canal corrector.*
En esta situación, si el canal corrector tiene una capacidad igual o mayor que $H_y(x)$, entonces es posible codificar la información del observador de manera que se envíe por el canal corrector y corregir todos los errores salvo una pequeña fracción ϵ de estos. Si la capacidad del canal es menor, no es posible la corrección.

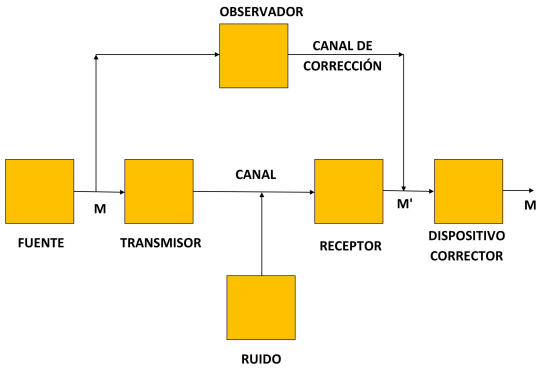

Figura 1.3: Sistema de comunicación con sistema de corrección

Demostración. En la primera parte, si consideramos secuencias de mensajes M' recibidos lo suficientemente largos, habiendo enviado mensajes M; tenemos que en un tiempo T se habrán producido logarítmicamente una cantidad de $TH_y(x)$ de mensajes M que pueden haber producido M'. Entonces debemos enviar $TH_y(x)$ bits en T segundos, lo que puede ser hecho con capacidad $H_y(x)$ salvo por una pequeña frecuencia ϵ de errores en el canal.
Para la segunda parte, si tomamos además la variable z como la señal enviada a través del canal de corrección, tenemos que (por las propiedades de la función H):

$$H_y(x,z) \geq H_y(x)$$

$$H_y(z) + H_{y,z}(x) \geq H_y(x)$$

$$H_{y,z}(x) \geq H_y(x) - H_y(z) \geq H_y(x) - H(z)$$

Si la capacidad del canal es menor que $H_y(x)$ entonces el lado derecho de la desigualdad será positivo y $H_{y,z}(x) > 0$, que simboliza la incertidumbre de lo enviado conociendo la señal recibida y la corrección. Si este dato es positivo, entonces los errores no pueden ser arbitrariamente pequeños. □

Podemos entender entonces $H_y(x)$ como la cantidad de información adicional que se debería incluir en el mensaje para garantizar la corrección de errores. Sobre este aspecto nos detallaremos en mayor longitud en la segunda parte de este trabajo, en cuanto a códigos correctores. En estas condiciones, podemos definir la capacidad de un canal con ruido:

Definición 1.2.2. *La* capacidad de un canal con ruido, *C, es el máximo posible del ratio de transmisión*

$$C = \text{máx}\,(H(x) - H_y(x)) \tag{1.5}$$

Donde el máximo es con respecto a todas las posibles fuentes de entrada al canal.

Nótese que en un canal sin ruido $H_y(x) = 0$, y por lo tanto la máxima entropía es la capacidad del canal, como ya demostramos; por lo que esta definición es consistente con la de capacidad que ya dimos en la primera sección. Continuando de manera análoga, podemos enunciar el Teorema Fundamental:

Teorema 1.2.2. Teorema Fundamental de un Canal Discreto con Ruido
Sea un canal discreto con capacidad C y una fuente discreta de entropía por segundo H. Si $H \leq C$ entonces existe un sistema de codificación tal que la salida de la fuente puede ser transmitida a través del canal con una frecuencia de error (equivocación) arbitrariamente pequeña. Si $H > C$ es posible codificar la fuente de manera que la equivocación sea menor que $H - C + \epsilon$, con ϵ arbitrariamente pequeño, pero no es posible una equivocación menor que $H - C$ con ningún método de codificación.

Demostración. Consideremos una fuente S_0 con la que se alcanza la máxima capacidad C, según (1.5), como la entrada del canal. Para un tiempo fijo T, todos los posibles mensajes recibidos y transmitidos de duración T se pueden diferenciar en varios grupos:

1. Los mensajes enviados se dividen en dos grupos, un grupo de alta probabilidad con alrededor de $2^{TH(x)}$ miembros y un segundo grupo de baja probabilidad con los restantes.

2. Análogamente, los recibidos se dividirán en un grupo de alta probabilidad de tamaño cercano a $2^{TH(y)}$ y otro de baja probabilidad.

3. Por su parte, si tomamos un mensaje recibido concreto del primer grupo anterior, este mensaje ha sido producido con alta probabilidad por alrededor de $2^{TH_y(x)}$ entradas concretas, mientras que la probabilidad de que provenga del resto es relativamente pequeña.

En todas estas separaciones, la probabilidad del segundo grupo (el menos probable) tiende a cero a medida que T aumenta.
Sea una segunda fuente que produce información en un promedio R, con $R < C$. Como en el caso anterior, en un tiempo T habrá un grupo de alta probabilidad con una media de 2^{TR} mensajes. Dichos mensajes se asociarán aleatoriamente con alguno los $2^{TH(x)}$ secuencias del canal descritas anteriormente. La probabilidad de que una de estas secuencias sea un mensaje es entonces

$$\frac{2^{TR}}{2^{TH(x)}} = 2^{T(R-H(x))}$$

Calculemos ahora la probabilidad de un error: Si fijamos un mensaje de salida concreto, vamos a calcular la probabilidad de que este mensaje haya sido producido por más de un mensaje de entrada. La probabilidad de que ninguna otra de las secuencias del canal que llevarían al mensaje de salida concreto no tengan un mensaje asociado es:

$$P = \left[1 - 2^{T(R-H(x))}\right]^{2^{TH_y(x)}}$$

Y dado que $R < C = H(x) - H_y(x)$, entonces existirá $\eta > 0$ tal que $R - H(x) = -H_y(x) - \eta$,

$$P = \left[1 - 2^{-TH_y(x)-T\eta}\right]^{2^{TH_y(x)}}$$

cuyo límite cuando $T \to \infty$ es 1 (basta aplicar logaritmo del límite e infinitésimos en el logaritmo), por lo que la probabilidad de de error tenderá a 0.
Para la segunda parte, si el canal sólo puede transmitir una cantidad C de bits por segundo, entonces no se ve capaz de transmitir el resto de $H(x) - C$, lo que provoca un error mínimo de la misma cantidad añadido. Veamos que este error no se puede reducir: si tuviésemos una codificación de una fuente de entropía $H(x) = C + a$ de forma que equivocación sea $H_y(x) = a - \epsilon$ con $\epsilon > 0$ (es decir, $H_y(x) < a = H(x) - C$). Entonces $H(x) - H_y(x) = C + \epsilon$, lo que contradice la definición de C como el máximo en (1.5). □

Es importante destacar que, al contrario que en el caso sin ruido, en esta demostración no hemos podido dar un método propiamente constructivo para elaborar un código con el que alcanzar las cotas descritas. En la práctica, este hecho resulta mucho más complicado y en ocasiones puede requerir de gran cantidad de información de redundancia para evitar los errores, lo cual limita la capacidad del canal y la velocidad de la transmisión de datos (en tanto que parte de esta capacidad estará 'ocupada' en la transmisión de la información de redundancia

Para finalizar con esta parte, veamos un ejemplo de código eficiente con el que corregir: En un canal binario, tenemos un ruido que puede modificar o no un único dígito de cada secuencia de 7 dígitos. Las ocho posibilidades (no error o cambio de dígito en cada posición) son igualmente distribuidas. De cada bloque de 7 dígitos $(X_1, X_2, \ldots, X_7)$ definimos X_3, X_5, X_6, X_7 como símbolos del mensaje y el resto de redundancia, obtenidos mediante:

X_4 de forma que $\alpha = X_4 + X_5 + X_6 + X_7$ sea par
X_2 de forma que $\beta = X_2 + X_3 + X_6 + X_7$ sea par
X_1 de forma que $\gamma = X_1 + X_3 + X_5 + X_7$ sea par

Al recibir un bloque, se calculan α, β y γ, dándoles un valor 0 si es par o 1 si es impar. En caso de que todos sean nulos, no hay error, si alguno es 1, entonces la posición indicada en binario por $\alpha\beta\gamma$ es en la que ha ocurrido el error.

En un caso concreto, si tenemos el mensaje $(X_3, X_5, X_6, X_7) = (1, 0, 0, 1)$, entonces el bloque completo será $(0, 0, 1, 1, 0, 0, 1)$. Si se recibe el mensaje $(0, 0, 1, 1, 0, 1, 1)$, tenemos que $\alpha = 1$, $\beta = 1$, $\gamma = 0$, luego se ha cometido un error en la posición $110 \to 6$; X_6 es incorrecto.

1.3. Sistemas continuos

1.3.1. Conceptos previos

Antes de entrar de lleno con el caso continuo, vamos a definir una serie de conceptos con los que poder describir matemáticamente las cualidades un sistema de comunicación continuo que posteriormente estudiaremos. En este caso, trabajaremos con grupos y conjuntos de funciones:

Definición 1.3.1. *Un* grupo de funciones *es una clase o colección de funciones (generalmente respecto de una sola variable, el tiempo). Se puede referir tanto a una colección definida explícitamente como a una agrupación de funciones que cumplan una condición determinada.*

Por ejemplo, dado un conjunto Θ, podemos definir el grupo de funciones

$$f_\theta(t) = \sin(t + \theta)$$

en donde para cada $\theta \in \Theta$ tenemos una función diferente.

Definición 1.3.2. *Denominaremos un* conjunto de funciones *a un grupo de funciones con una probabilidad inducida, es decir, cada función del grupo tiene una probabilidad determinada; o se puede determinar la probabilidad de las funciones que cumplan una determinada característica dentro del conjunto, etc.*

Con el mismo ejemplo, al grupo de funciones anterior definido le podemos otorgar una probabilidad de distribución para θ, $P(\theta)$, convirtiendo el grupo en un conjunto.

Definición 1.3.3. *Se dice que un conjunto es* estacionario *si al desplazarse una cantidad fija de tiempo en todas las funciones, el conjunto en sí permanece siendo el mismo a pesar de los cambios puntuales en las funciones.*

En nuestro ejemplo, supongamos que θ tiene una distribución uniforme en $[0, 2\pi]$. Si desplazamos todas las funciones un tiempo t_0, tenemos que

$$f_\theta(t + t_0) = \sin(t + \theta + t_0) = \sin(t + \varphi) = f_\varphi(t)$$

con φ distribuido uniformemente $[0, 2\pi]$. A pesar de que cada función ha cambiado, el conjunto sigue siendo el mismo.

Definición 1.3.4. *Se dice que un conjunto es* ergódico *si es estacionario y no hay ningún subconjunto de probabilidad diferente de 0 ó 1 que sea estacionario.*

Siguiendo con las propiedades ergódicas que ya definimos, el significado de esta propiedad quiere decir que en un conjunto ergódico cada función es típica, contiene todas las propiedades estadísticas que se le pueden atribuir al conjunto.

Podemos operar en los conjuntos para obtener otros nuevos (lo que se correspondería físicamente al paso por algún dispositivo como filtros, moduladores, etc.). Si tenemos un conjunto $f_\alpha(t)$ y un operador T, entonces a cada función del conjunto le corresponde una nueva función del conjunto $g_\alpha(t)$ tal que

$$g_\alpha(t) = T f_\alpha(t)$$

donde la medida de probabilidad es la misma en ambos conjuntos, así como los correspondientes subconjuntos relacionados por el operador también conservan la misma probabilidad.

Un operador T es denominado *invariante* si cualquier desplazamiento en la variable se corresponde también en la salida; es decir,

$$g_\alpha(t) = T f_\alpha(t) \qquad \text{implica} \qquad g_\alpha(t + t_0) = T f_\alpha(t + t_0)$$

para todo $f_\alpha(t)$ y t_0 posibles.

Lema 1.3.1. *Si un operador T es invariante y el conjunto de entrada es estacionario, entonces el conjunto de salida es también estacionario.*

Corolario 1.3.1. *Si un operador T es invariante y el conjunto de entrada es ergódico, entonces el conjunto de salida es también ergódico.*

Para la entropía en los sistemas continuos, tiene sentido realizar una fórmula análoga al caso discreto (1.2) de forma que la entropía de una distribución continua $p(x)$ sea

$$H = -\int p(x) \log p(x) dx \tag{1.6}$$

A través de esta fórmula, podemos describir en el caso de tener dos argumentos x e y las entropías conjuntas y condicionadas:

$$H(x,y) = -\iint p(x,y) \log p(x,y) dx dy$$

$$H_x(y) = -\iint p(x,y) \log \frac{p(x,y)}{p(x)} dx dy$$

$$H_y(x) = -\iint p(x,y) \log \frac{p(x,y)}{p(y)} dx dy$$

con

$$p(x) = \int p(x,y) dy\,; \qquad p(y) = \int p(x,y) dx$$

La entropía en el caso continuo conserva alguna de las propiedades del caso discreto, aunque también hay otras nuevas que merece la pena destacar. Las principales propiedades son:

1. Siendo V el volumen en el que está limitado el argumento x, el máximo se alcanza cuando x tiene una distribución uniforme en todo el volumen (mayor desorden posible), es decir, $p(x) = \dfrac{1}{V}$. El mayor valor posible de H es entonces

$$H = -\int_V \frac{1}{V} \log\left(\frac{1}{V}\right) dx = \log V$$

2. Para cualesquiera dos variables x, y:

$$H(x,y) \leq H(x) + H(y)$$

dándose la igualdad sí y sólo sí x e y son independientes $p(x,y) = p(x)p(y)$.

3. Al igual que en el caso discreto

$$H(x,y) = H(x) + H_x(y) = H(y) + H_y(x)$$

$$H_x(y) \leq H(y), \qquad H_y(x) \leq H(x)$$

4. Siendo $p(x)$ unidimensional, el máximo de entropía bajo la condición de tener una desviación estándar fija σ se da en la distribución normal. En este caso, el valor de la entropía es

$$H(x) = \log \sqrt{2\pi e}\sigma$$

5. Una diferencia fundamental entre la entropía para el caso discreto y el continuo es que en el caso continuo la medida es relativa al sistema de coordenadas. Si este sistema es cambiado, entonces la medida también cambia de manera que, si cambiamos a coordenadas y:

$$H(y) = -\int p(x)J(x,y)\log\left(p(x)J(x,y)\right)dy$$

donde $J(x,y)$ es el Jacobiano del cambio de coordenadas. Expandiendo el logaritmo, podemos obtener que entonces

$$H(y) = H(x) - \int p(x)\log\left(J(x,y)\right)dx$$

Debido a estos posibles cambios de escala, en el caso continuo la entropía puede ser negativa.

A pesar de esta dependencia, en la mayoría de conceptos que vamos a trabajar posteriormente dependerán de una diferencia entre dos entropías, la cual no se verá afectada por el sistema de coordenadas, pues cualquier cambio cargaría igual a ambas entropías.

Vamos ahora a suponer que el canal que se usa en la comunicación tiene un cierto ancho de banda W en ciclos por segundo (en la actualidad, nos referimos a esta medida como *hertz*, Hz) empezando en la frecuencia nula *(nótese que el ancho de banda es la diferencia entre la frecuencia superior y la inferior. Al denotar un ancho de banda W comenzando en 0, entonces estamos limitando la señal del canal en frecuencias entre 0 y W)*; podemos determinar el siguiente principio:

Teorema 1.3.1. *Si una función $f(t)$ no contiene frecuencias superiores de W Hz, entonces está completamente determinada dando sus valores en una serie discreta de puntos espaciados $\frac{1}{2W}$ segundos entre ellos.*
De hecho, $f(t)$ se puede representar como

$$f(t) = \sum_{n=-\infty}^{\infty} X_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)} \tag{1.7}$$

donde X_n es el n-ésimo valor de los puntos;

$$X_n = f\left(\frac{n}{2W}\right)$$

Si la función está limitada a un tiempo de intervalo T y los puntos son espaciados por $\frac{1}{2W}$ segundos como hemos indicado, entonces hay un total de $2TW$ puntos usados como referencia. Más precisos, podemos definir la función como limitada a un cierto intervalo T sí y sólo sí todas las muestras X_n fuera del intervalo T son nulas.
En un conjunto de funciones limitadas por el ancho de banda y la duración, estas pueden ser representadas por la distribución de probabilidad $p(x)$ en dicho intervalo. A partir de este momento, podemos estudiar las propiedades de las funciones mediante el estudio de como se comportan las funciones en las muestras.

Si consideramos un conjunto de funciones limitadas por un ancho de banda de W Hz, podemos definir la *entropía del conjunto por grado de libertad* como

$$H' = -\lim_{n\to\infty} \frac{1}{n} \int \cdots \int p(x_1, x_2, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n \tag{1.8}$$

Podemos definir también la entropía por segundo dividiendo entre el intervalo T para n muestras. Puesto que $n = 2TW$,

$$H = 2WH' \tag{1.9}$$

La entropía está relacionada con el logaritmo de una distribución lo suficientemente larga (suficientemente alto número de muestras). Así, si asumimos $p(x_1, x_2, \ldots, x_n)$ continuo para todas las variables y todo n, entonces para n lo suficientemente largo tenemos un resultado similar al caso discreto:

$$\left| \frac{\log p}{n} - H' \right| < \epsilon$$

para todos los posibles $(x_1, x_2, \ldots, x_n)$ excepto por un conjunto cuya probabilidad total es menor que δ; con δ y ϵ arbitrariamente pequeños.

Un caso típico de estudio y de gran importancia por nuestro estudio posterior se trata del caso en el que el ruido en el sistema es *ruido térmico blanco*. Este ruido tiene la propiedad de que cada muestra es perturbada independientemente respecto a las otras; y que la distribución de cada amplitud de la perturbación (distancia entre la muestra perturbada y la original) sigue una distribución normal con desviación estándar $\sigma = \sqrt{N}$, con N siendo la potencia media del ruido.
En este caso su función de distribución es

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi N)^{\frac{n}{2}}} e^{-\frac{1}{2N}\sum x_i^2}$$

y obtenemos una entropía

$$H' = \log \sqrt{2\pi e N}$$

$$H = W \log 2\pi e N$$

Nótese que, por las propiedades de la distribución normal que ya describimos, para una potencia media N determinada el ruido blanco tiene la máxima entropía posible.

Este ruido adquiere toda su importancia para medir la entropía en los casos continuos, donde es conveniente no medir directamente la entropía de un conjunto de funciones; si no trabajar con una comparación con el ruido blanco, calculando la potencia de un ruido térmico blanco con la misma entropía. A este concepto lo denominaremos *potencia de entropía*:

Definición 1.3.5. *La* potencia de entropía *de un conjunto de funciones es la potencia de un ruido blanco limitado por el mismo ancho de banda que las funciones del conjunto y teniendo la misma entropía. Si H' es la entropía del conjunto, entonces la potencia de entropía es*

$$N_1 = \frac{1}{2\pi e} e^{2H'} \tag{1.10}$$

Por último, antes de comenzar el estudio de capacidades y ratios del caso continuo, introduzcamos brevemente el caso de suma de conjuntos.

Si tenemos dos conjuntos de funciones $f_\alpha(t)$ y $g_\beta(t)$ con funciones de densidad de la probabilidad respectivas $p(x_1, x_2, \ldots, x_n)$ y $p(x_1, x_2, \ldots, x_n)$; podemos formar un nuevo conjunto que representa la suma de ambos. Este nuevo conjunto supone el resultado final de la superposición de los dos ruidos y/o señales representados en los conjuntos descritos. La función de densidad de la suma viene dada por la convolución

$$r(x_1, x_2, \ldots, x_n) = \int \cdots \int p(y_1, y_2, \ldots, y_n) q(x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n) dy_1 dy_2 \cdots dy_n$$

y podemos obtener el siguiente resultado:

Teorema 1.3.2. *Sean $\overline{N}_1$ y $\overline{N}_2$ la potencia media de dos conjuntos de funciones y sean N_1 y N_2 sus potencias de entropía respectivas. Entonces la potencia de entropía de la suma, $\overline{N}_3$, está acotada por*

$$\overline{N}_1 + \overline{N}_2 \leq \overline{N}_3 \leq N_1 + N_2$$

1.3.2. Canales continuos

En un canal continuo la entrada y señales transmitidas serán una función respecto del tiempo $f(t)$, mientras que la salida será una versión perturbada de la misma debido al ruido. Consideremos solo el caso en el que las señales transmitidas y recibidas están limitadas por un cierto ancho de banda W. Entonces para un periodo de tiempo T la estructura estadística viene determinada por la función de distribución finita de $n = 2TW$ números, siendo $P(x) = P(x_1, x_2, \ldots, x_n)$ la probabilidad de la señal transmitida; mientras que el ruido se puede representar por la probabilidad condicional $P_x(y) = P_{x_1, x_2, \ldots, x_n}(y_1, y_2, \ldots, y_n)$.

El ratio de transmisión tiene la misma definición que en el caso discreto (1.4):

$$R = H(x) - H_y(x)$$

con $H(x)$ la entropía de la entrada y $H_y(x)$ la equivocación. La capacidad del canal C es entonces el máximo ratio de transmisión posible para los posibles conjuntos de funciones que actúen como entrada; es decir, el máximo respecto a los posibles $P(x) = P(x_1, x_2, \ldots, x_n)$, maximizando

$$R = H(x) - H_y(x) = -\int P(x) \log P(x) dx + \iint P(x, y) \log \frac{P(x, y)}{P(y)} dx dy$$

Dado que $P(x) = \int P(x, y) dy$, podemos reescribir como

$$\iint P(x, y) \log \frac{P(x, y)}{P(y) P(x)} dx dy$$

y por lo tanto la capacidad del canal es

$$C = \lim_{T\to\infty} \max_{P(x)} \frac{1}{T} \iint P(x,y) \log \frac{P(x,y)}{P(y)P(x)} dxdy \tag{1.11}$$

Si el logaritmo se encuentra en base 2, entonces la capacidad del canal describe el número de bits por segundo que se pueden enviar a través del canal con una equivocación arbitrariamente pequeña, al igual que el caso discreto. Además, tenemos que para un mensaje u, siendo x, y las señales enviada y recibida respectivamente; y v el mensaje descodificado desde y (y ha sido afectada por el ruido), se cumple que

$$H(u) - H_v(u) \leq H(x) - H_y(x)$$

sin importar los métodos de codificación/descodificación que se puedan usar para obtener x o v. Es decir, sea cual sea el tipo de codificación, el ratio no superará la capacidad del canal. En cambio, bajo condiciones generales se puede conseguir un método de codificación con el que trasmitir a ratio C con una arbitrariamente pequeña frecuencia de error.

Un caso de especial interés se encuentra si el ruido es independiente (estadísticamente) de la señal transmitida. Entonces $P_x(y)$ es una función que depende solo de la diferencia $n = x - y$, $Q(x - y)$. Denominemos como $H(n)$ a la entropía de la distribución $Q(n)$.

Teorema 1.3.3. *Si el ruido es independiente de la señal transmitida y la señal recibida es la suma de la enviada junto con el ruido, entonces el ratio de transmisión es*

$$R = H(y) - H(n)$$

y por lo tanto la capacidad del canal es

$$C = \max_{P(x)} H(y) - H(n)$$

Demostración. Dado que $y = x + n$, y se trata de una transformación lineal, entonces

$$H(x,y) = H(x,n)$$

que se puede expandir en ambos lados

$$H(y) + H_y(x) = H(x) + H_x(n) = H(x) + H(n)$$

dado que x y n (señal y ruido) son independientes. Finalmente se deduce que

$$R = H(x) - H_y(x) = H(y) - H(n)$$

□

Supongamos ahora que nos encontramos en el caso en el que el ruido es térmico blanco

Teorema 1.3.4. *La capacidad de un canal de ancho de banda W perturbado por ruido térmico blanco de potencia N cuando la potencia media del transmisor es P es*

$$C = W \log \frac{P+N}{P} \tag{1.12}$$

Demostración. Si el ruido tiene una potencia media N y las señales transmitidas están limitadas por una potencia media P entonces la señal perturbada tiene una potencia media $P + N$. La máxima entropía de la señal recibida se da cuando la señal transmitida sigue una distribución normal (máximo caso al estar la potencia limitada) y entonces

$$H(y) = W \log 2\pi e(P + N)$$

y la del ruido blanco, como ya vimos,

$$H(n) = W \log 2\pi eN$$

Aplicando el anterior Teorema 1.3.3, nos queda

$$C = H(y) - H(n) = W \log \frac{P + N}{P}$$

□

Con un sistema de codificación adecuado, podemos transmitir información a través del canal a razón de $\log_2 \frac{P+N}{P}$. En este sistema "ideal", la señal transmitida se aproxima al ruido blanco en propiedades estadísticas.

En el caso en que el ruido siga cualquier otra distribución, el hecho de calcular la capacidad se vuelvo un proceso mucho más complicado o imposible de resolver. No obstante, esta capacidad se puede acotar en función de la potencia del ruido y la entropía de potencia.

Teorema 1.3.5. *Sea un canal con un ancho de banda W perturbado por un ruido arbitrario de potencia media N y potencia de entropía N_1, siendo P la potencia media de transmisión. Entonces la capacidad del canal, C está acotada de la siguiente manera:*

$$W \log \frac{P + N_1}{N_1} \leq C \leq W \log \frac{P + N}{N_1} \tag{1.13}$$

Demostración. La señal recibida tiene una potencia media $P + N$. La máxima entropía dada está potencia ocurre si la señal recibida tiene una distribución normal y en ese caso $H(y) = W \log 2\pi e(P + N)$. Entonces

$$C = \max_{P(x)} H(y) - H(n) \leq W \log 2\pi e(P + N) - W \log 2\pi eN_1 = W \log \frac{P + N}{N_1}$$

Para la cota inferior, supongamos que transmitimos como señal ruido blanco de potencia P. Entonces la potencia de entropía de la señal recibida ha de ser mayor que $P + N_1$, dada la desigualdad en la suma de conjuntos de funciones del Teorema 1.3.2.
Entonces $\max H(y) \geq W \log 2\pi e(P + N_1)$ y

$$C \geq W \log 2\pi e(P + N_1) - W \log 2\pi eN_1 = W \log \frac{P + N_1}{N_1}$$

□

1.3.3. Fuentes continuas

En el caso de una fuente continua, sin ninguna información previa se podría considerar una frecuencia infinita de generación de la información, puesto que especificar de manera exacta una cantidad que tiene un rango continuo de posibilidades requiere de un número infinito de bits. De esta forma, para poder transmitir dicha información a través de un canal sería necesario una capacidad del canal infinita. No obstante, dado que los canales suelen tener cierto ruido asociado (y por lo tanto la capacidad es finita), una transmisión exacta es imposible.

Afortunadamente, mensajes continuos no requieren de ser enviados de una manera exacta, si no que admitiremos transmisiones con cierta tolerancia, con lo que se puede definir un ratio finito de información por segundo asignado a la fuente continua. Este ratio depende de la magnitud del error permitido entre los mensajes inicial y final; por lo que lo denominamos *relativo al criterio de fidelidad*.

Para poder dar una formulación matemática a esta fidelidad, trabajando en el conjunto de mensajes de una duración dada T suficientemente larga; vamos a denotar x al mensaje producido e y el mensaje recibido. Todo el sistema de comunicación puede ser descrito a partir de la función de probabilidad $P(x, y)$, por lo que representamos la fidelidad como una función $\nu\left(P(x, y)\right)$ sobre las posibles funciones de probabilidad $P(x, y)$. De manera precisa, podemos medir esta fidelidad como un promedio de la función $\rho(x, y)$ sobre los posibles x e y:

$$\nu\left(P(x, y)\right) = \iint P(x, y)\rho(x, y)dxdy \tag{1.14}$$

Para que la fórmula (1.14) sea válida es preciso asumir que tanto la fuente como el sistema son ergódicos; así como que esta evaluación es posible.
La función $\rho(x, y)$ mide como de improbable sería recibir y a partir de x de acuerdo al criterio de fidelidad escogido, por lo que se denomina *función de evaluación o distancia* (pues actúa como una "distancia" entre x e y, a pesar de no definir un métrica). Algunas de estas funciones más comunes son:

1. El criterio RMS, que utiliza la distancia euclídea

$$\rho(x, y) = \frac{1}{T}\int_0^T \left[x(t) - y(t)\right]^2 dt$$

   de forma que

$$\nu = \overline{\left(x(t) - y(t)\right)^2}$$

2. El criterio de error absoluto,

$$\rho(x, y) = \frac{1}{T}\int_0^T \left|x(t) - y(t)\right| dt$$

Ahora estamos en condiciones de definir el ratio de generación de información:

Definición 1.3.6. *Dado un sistema particular con una probabilidad conjunta $P(x, y)$ y una fidelidad ν, el* ratio de información generada relativo al criterio de fidelidad ν *se define como el mínimo ratio R de transmisión al variar la probabilidad condicionada $P_x(y)$ manteniendo constante la fidelidad, esto es,*

$$R = \min_{P_x(y)} \iint P(x, y) \log \frac{P(x, y)}{P(x)P(y)} dxdy \tag{1.15}$$

sujeto a

$$\nu = \iint P(x,y)\rho(x,y)dxdy$$

Esta definición nos lleva directamente al siguiente teorema:

Teorema 1.3.6. *Si una fuente tiene un ratio R para una fidelidad ν, es posible codificar la salida de la fuente y transmitir a través de un canal de capacidad C con fidelidad tan cercana a ν como se quiera sí y sólo sí $R \leq C$.*

Por último, veamos por su interés práctico el caso concreto en el que aplicamos el criterio RMS y el conjunto de funciones del mensaje es ruido blanco de potencia Q, pues el ratio puede ser calculado explícitamente. En este caso

$$R = \text{mín}\,[H(x) - H_y(x)] = H(x) - \text{máx}\,H_y(x) = W\log 2\pi eQ - W\log 2\pi eN = W\log\frac{Q}{N}$$

siendo $N = \overline{(x-y)^2}$ el cuadrado del error medio permitido, W el ancho de banda. Tengasé en cuenta que $\text{máx}\,H_y(x)$ se alcanza cuando $y-x$ es ruido blanco, por lo que es igual a $W\log 2\pi eN$. A partir de este caso, para otros tipos de fuentes de mensajes podemos acotar de la siguiente manera:

Teorema 1.3.7. *El ratio de cualquier fuente con un ancho de banda W está acotado por*

$$W\log\frac{Q_1}{N} \leq R \leq W\log\frac{Q}{N} \tag{1.16}$$

siendo Q_1 la potencia media de la fuente, Q_1 la potencia de entropía de la misma y N el cuadrado del error medio permitido.

Capítulo 2

Códigos correctores

La implementación en la práctica de las nociones y los resultados del capítulo 1 no es posible a partir de los métodos basados en probabilidad sobre los que se fundamentó inicialmente la teoría de Shannon. Por ello,durante las tres primeras décadas de la segunda mitad del siglo XX se consideraron los modelos algebraicos de los códigos correctores, principalmente lineales, de interés matemático como de los códigos convolucionales de mayor interés en ingeniería.

Dichos códigos dan lugar a modelos para corregir errores producidos por la distorsión de la transmisión en presencia de ruido, que frecuentemente es grande. Se fundamentan en la capacidad de corregir un número significativo de errores cuando se realiza la codificación por un procedimiento inteligente que permita decodificar mediante procedimientos matemáticos adecuados.

Los modelos algebraicos fueron los únicos modelos prácticos para el sistema discreto hasta 1980, cuando V. Goppa convirtió en geométricos sus últimos modelos algebraicos de códigos lineales, y ello permitió la extensión de dichos modelos a casos generales de la geometría algebraica. Sólo poco después ya se lograron construir familias excelentes de códigos lineales definidas sobre cuerpos finitos F_q de cardinal $q \geq 49$ y q cuadrado, lo que supuso un hito histórico ya que dichos códigos optimizan a la vez la proporción de errores que se corrigen y el coste de la codificación.

En los últimos 25 años se han logrado avances sistemáticos muy significativos en la decodificación y en particular en la evolución y creación de nuevos códigos convolucionales y de los llamados turbo-códigos, que permiten respaldar también en la práctica la teoría de Shannon. No se han logrado todavía familias de códigos excelentes para $q < 49$, en particular se está muy lejos de lograrlo para el caso $q = 2$ que es el más importante para las aplicaciones. Precisamente por esa circunstancia, en este capítulo vamos a considerar principalmente códigos correctores sobre el cuerpo F_2, donde trataremos de introducir sus principales propiedades.

2.1. Definiciones

Un código bloque C es un conjunto de M *palabras* $C = \{c_1, c_2, \ldots, c_M\}$ donde cada palabra es una n-tupla $c_i = (c_{i0}, c_{i1}, \ldots, c_{in-1})$ formada por elementos pertenecientes a un *alfabeto* finito de g elementos. Nos referiremos a n como la longitud del código.

En la mayoría de los casos, vamos a trabajar con códigos lineales, trabajando en espacios vectoriales. Si partimos de un cuerpo finito $\mathbb{F}$ (nuestro alfabeto), consideremos $\mathbb{F}^n$ como el espacio vectorial $V = \{\mathbb{F}^n, +, \mathbb{F}\}$. En dicho espacio, definimos el concepto de código lineal:

Definición 2.1.1. *Un* (n, k) código lineal C *es un subespacio de dimensión* k *del espacio vectorial* V*.*

En tales códigos, el número total de palabras es $M = q^k$. Al igual que cualquier espacio lineal, la ventaja de estos códigos es que nos permitirá trabajar exclusivamente con la base lineal.

Definición 2.1.2. *La* matriz generadora G *de un* (n, k) *código lineal* C *es una matriz de dimensión* $k \times n$ *cuyas filas son linealmente independientes (es decir, sus filas están formadas por una base del código).*

Si partimos de una información en forma de un vector u de dimensión k, se puede establecer la una codificación tal que $c = uG$. A pesar de tratarse del mismo código, deferentes elecciones de las filas de G daría lugar a diferentes codificaciones, por lo que establecemos una forma sistemática tal que las primeras k columnas forman la matriz identidad de dimensión k, quedando

$$G = (I, A)$$

Definimos también como *comprobación de paridad* a cualquier vector h del espacio vectorial V tal que

$$Gh^T = 0$$

Estos vectores forman a su vez un subespacio de V de dimensión $n - k$.

Definición 2.1.3. *La* matriz de comprobación de paridad H *de un* (n, k) *código lineal* C *es una matriz de dimensión* $(n-k) \times k$ *cuyas filas son vectores de comprobación de paridad linealmente independientes.*

De esta forma, $GH^T = 0$, y podemos escribir H a partir de la forma sistemática de G como

$$H = (-A^T, I)$$

siendo I la matriz identidad de dimensión $(n - k)$ en este caso.

Definición 2.1.4. *Dada la matriz de comprobación de paridad* H *de un* (n, k) *código lineal* C*, el* síndrome *de un vector* $r \in \mathbb{F}^n$*,* $s = \text{syn}(r)$ *es tal que*

$$s = Hr^T$$

El síndrome nos servirá para poder cuantificar los errores cometidos al recibir el mensaje. Así, si r es la palabra recibida al enviar la palabra c, tenemos que $r = c + e$, siendo e el error en la palabra recibida y por lo tanto

$$s = Hr^T = H(c + e)^T = He^T$$

con lo que s *refleja* el error.

Definición 2.1.5. *El* peso Hamming *de un vector* x*,* $w_H(x)$*, es igual al número de coordenadas no nulas de dicho vector.*

En una palabra recibida $r = c + e$, el numero de errores será $w_H(e)$.

Definición 2.1.6. *Un código es* t-corrector *si para cualesquiera dos palabras diferentes* c_i*,* c_j *y dos errores* e_1*,* e_2 *de peso menor o igual que* t*, entonces* $c_i + e_1 \neq c_j + e_2$

Es decir, en un código t-corrector es imposible recibir la misma palabra desde dos palabras diferentes si los errores son menores o iguales que t.

Definición 2.1.7. *La* distancia Hamming *entre dos vectores x, y, denotado $d_H(x, y)$, es el número de coordenadas en el que difieren.*

Con esta última definición, podemos dar a V la estructura de espacio métrico.

Definición 2.1.8. *La* mínima distancia *de un código, d, es la mínima distancia Hamming entre cualquier par de palabras diferentes del código.*

En un código lineal, es directo que esta mínima distancia es igual al mínimo peso posible entre todas la palabras no nulas del código.
En posteriores notaciones, podremos referirnos a un (n, k, d) código lineal como un (n, k) código lineal de distancia mínima d.

Definición 2.1.9. *La* distribución de peso *de un código es un vector $A = (A_1, A_2, \ldots, A_n)$, donde cada coordenada A_i representan el número de palabras del código con peso i. Así mismo, definimos como enumerador de peso al polinomio*

$$A(z) = \sum_{w=0}^{n} A_w z^w$$

Por otro lado, es muy importante en los códigos correctores la figura del descodificador, que podemos considerar como una aplicación de $\mathbb{F}_q^n$ en el código C tal que a cada palabra recibida r le corresponde una palabra del código (es decir, descodifica la información recibida).

Definición 2.1.10. *Un* descodificador de distancia mínima *es un descodificador tal que, dada una palabra recibida r, devuelve la palabra c del código de forma que $d(r, c) < \frac{d}{2}$ si esta palabra existe, o declara fallo en otro caso.*

2.2. Cotas en la probabilidad de error para códigos correctores

Durante esta sección, trataremos de establecer la efectividad de un código corrector en términos de su probabilidad de cometer un error.

Además, a menos que se especifique lo contrario, trabajaremos en un alfabeto binario, es decir de sólo dos elementos: $\{0, 1\}$; $\mathbb{F} = \mathbb{F}_2$
Asumimos que dichos símbolos tienen probabilidades $P[1] = p, \quad P[0] = 1 - p$. Se supone que los símbolos son mutuamente independientes.

Lema 2.2.1. *La probabilidad de que una palabra de longitud n tenga un número j de* 1 *viene dada por la distribución binomial*

$$P[n, j] = \binom{n}{j} p^j (1 - p)^{n-j}$$

No obstante, cuando n es suficientemente grande parece razonable aproximar mediante una distribución de Poisson, de forma que

$$P[j] = e^{-\mu} \frac{\mu^j}{j!}$$

Normalmente, en un (n, k) código lineal se asume que cada palabra es usada con la misma

probabilidad, con lo que $P[c_j] = 2^{-k}$ (recuérdese que el número total de palabras es 2^k para un (n, k) código lineal binario). Si los errores ocurren con probabilidad p, independientemente del símbolo transmitido y entre ellos, entonces se trata de un canal binario simétrico (*Binary Symmetric Channel,* BSC).

Además, trabajaremos con una descodificación de distancia acotada, tal que todos los patrones de máximo t errores son corregidos, mientras que aquellos que lo superan no.
En ese caso, hablaremos de *error de descodificación* si la palabra obtenida por el descodificador es diferente a la transmitida, mientras que se producirá un *fallo de descodificación* si la palabra correcta no es corregida, es decir, no se obtiene la palabra correcta como salida final. Obviamente, un fallo de descodificación conlleva necesariamente un error de descodificación; por lo que su probabilidad, P_{fallo} será mayor que la del error, P_{error}.

Teorema 2.2.1. *La probabilidad de un fallo de descodificación en un descodificador de distancia acotada es*

$$P_{fallo} = 1 - \sum_{j=0}^{t} \binom{n}{j} p^j (1-p)^{n-j} = \sum_{j=t+1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \tag{2.1}$$

Busquemos una forma de calcular la probabilidad de error:

Lema 2.2.2. *Sea una transmisión de la palabra nula, recibiendo una palabra descodificada de peso w. Si el error tiene peso j y la distancia entre la palabra recibida y la descodificada es l, entonces*

$$j + l - w = 2i \geq 0$$

Lema 2.2.3. *Supongamos que el de la palabra descodificada es w. Entonces el número de vectores a distancia j de la palabra transmitida y a distancia l de la descodificada es*

$$T(j, l, w) = \binom{w}{j-i} + \binom{n-w}{i} \tag{2.2}$$

Con $i = \frac{j+l-w}{2}$, para $(j + l - w)$ par y $w - l \leq j \leq w + l$. En otro caso, $T = 0$.

Ahora, podemos calcular la probabilidad de error a partir de sumar (2.2) sobre todas las palabras y distancias posibles:

Teorema 2.2.2. *La probabilidad de un error de descodificación en un descodificador de distancia acotada en un BSC de probabilidad de error p es*

$$P_{error} = \sum_{w>0} \sum_{j=w-t}^{w+t} \sum_{l=0}^{t} A_w T(j, l, w) p^j (1-p)^{n-j} \tag{2.3}$$

Siendo A_w la w-ésima coordenada de la distribución de peso del código.

Como en la mayoría de algoritmos calcular el valor exacto es muy costoso y/o poco efectivo, por lo que conviene también poder acotar dicha probabilidad de manera más efectiva.

En este sentido, toman importancia los descodificadores de máxima probabilidad:

Definición 2.2.1. *Un* descodificador de máxima probabilidad *es un descodificador que, dada una palabra recibida r, devuelve la palabra c del código tal que la distancia $d(c, r)$ es mínima.*

En este caso, el término de máxima probabilidad se refiere a la probabilidad condicionada de recibir la palabra r cuando se ha transmitido c.

Teorema 2.2.3. *Para un código con enumerador de peso $A(z)$ en un* BSC *con un error en cada símbolo de probabilidad p, se da que*

$$P_{error} \leq \sum_{w>0} \sum_{j>\frac{w}{2}} A_w \binom{w}{j} p^j (1-p)^{w-j} + \frac{1}{2} \sum_{j>0} A_{2j} \binom{2j}{j} p^j (1-p)^j \tag{2.4}$$

Un mayor número de errores ocurrirán cuando j es cercano a $\frac{w}{2}$, por lo que se puede usar este valor para j en todos los casos, simplificando:

$$P_{error} < \sum_{w>0} A_w 2^{w-1} (p-p^2)^{\frac{w}{2}} = \frac{1}{2} \sum_{w>0} A_w Z^w \tag{2.5}$$

siendo Z la función $Z = \sqrt{4p(1-p)}$.

De estas últimas ecuaciones se puede deducir que el número de palabras de diferentes pesos tiene una gran implicación en la probabilidad de error. Así, si p aumenta, la palabras de mayor peso contribuirán más a dicha probabilidad, dado que su número es mucho mayor.

Por otro lado, también podemos asumir que si j es lo suficientemente grande, todos los vectores de peso j causarán error. Esta aplicación se convierte en la *cota de Poltyrev*:

Teorema 2.2.4. *Para un (n,k,d) código con enumerador de peso $A(z)$ en un* BSC *con un error en cada símbolo de probabilidad p, se da que*

$$P_{error} \leq \sum_{w>0} A_w \sum_{\frac{w}{2}<j\leq J} \sum_{l<J} T(j,l,w) p^j (1-p)^{n-j} + \sum_{j>J} \binom{n}{j} p^j (1-p)^{n-j} + \\ + \frac{1}{2} \sum_{l>0} \sum_{l\leq j\leq J} A_{2l} T(j,j,2l) p^j (1-p)^{n-j}$$

2.3. Canales de comunicación y Teoría de la Información

Como ya mencionamos en el capítulo anterior, un canal de información es el medio por el que se transmite un mensaje de forma que la entrada es dicho mensaje y la salida una reproducción (en la mayoría de casos imperfecta) de tal mensaje. En esta sección, veremos que tipos de canales podemos encontrar y la cantidad de información que se puede transmitir en cada uno. En ella, repetiremos algunas nociones del primer capítulo con el objeto de adaptarlas y actualizarlas para el uso de códigos correctores.

Para ello, asumimos que queremos enviar una cantidad ilimitada de información, que estará dividida en mensajes (cadenas de símbolos binarios). En una *fuente discreta sin memoria*, la salida X será una secuencia de variables aleatorias independientes, con valores en un alfabeto finito $\{x_1, x_2, \ldots, x_r\}$, siendo la probabilidad de cada uno $P(x_j) = p_j$ respectivamente. Así, hablaremos del vector de distribución de X al vector $Q(X) = \{p_1, p_2, \ldots p_r\}$. De nuevo, para medir la cantidad de información, recordemos el concepto de *entropía*:

Definición 2.3.1. *La* entropía *de una fuente discreta sin memoria X es*

$$H(X) = E[-\log P(X)] = -\sum_j p_j \log p_j \tag{2.6}$$

donde E representa el valor esperado. Normalmente, se toma el logaritmo en base 2 (al trabajar en código binario) y la unidad de información es denominada bit.

Recuérdese que en un alfabeto de r símbolos, el valor máximo de la entropía es es $\log r$, el cual se consigue si todos los símbolos tienen la misma probabilidad, $\frac{1}{r}$.

Un canal es discreto y sin memoria si la información de salida depende únicamente de la respectiva entrada. El canal conecta un par de variables aleatorias:

$$X = \{x_1, x_2, \ldots, x_r\}; \quad Y = \{y_1, y_2, \ldots, y_s\}$$

El canal es descrito por las probabilidades condicionadas de obtener y_i dado x_j como entrada, $P(y_i|x_j) = p_{ji}$. El vector de distribución de Y es entonces

$$Q(Y) = Q(X)Q(Y|X)$$

siendo $Q(Y|X) = [p_{ji}]$ la matriz de transición.

Definición 2.3.2. *La* información mutua *del par* $(X|Y)$ *es*

$$I(X;Y) = E\left[\log \frac{P(y|x)}{P(y)}\right] = \sum_j P(x_j) \sum_i p_{ji} \left[\log p_{ji} - \log P(y_i)\right] \tag{2.7}$$

Téngase en cuenta que, dado $P(Y|X) = \dfrac{P(Y,X)}{P(X)}$; I es simétrico, $I(Y;X) = I(X;Y)$

Lema 2.3.1.

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$$

Por lo que $H(X)$ *es el máximo valor de* I*, que se alcanza cuando* $Q(Y|X)$ *es la matriz unidad.*

Definición 2.3.3. *La* capacidad de un canal discreto*,* $C(Y|X)$ *es el máximo de* $I(Y;X)$ *con respecto de* $P(X)$

La capacidad nos indica que mensajes de k bits pueden ser comunicados de forma segura usando ese canal al menos más de $n = \frac{k}{C}$ veces, es decir, mediante n símbolos codificados usando un código de 2^k vectores.
Si definimos el ratio de un (n, k) código como $R = \frac{k}{n}$, entonces para longitud N del código y ratio R tenemos que existe una constante positiva $E(R)$ tal que algunos errores satisfacen

$$P(e) < 2^{-NE(R)}$$

De esta forma, si queremos transmitir k bits de información en un canal sin memoria deberíamos usarlo al menos $n = \frac{k}{C}$ para obtener suficiente información. Un código bloque sería la regla en que seleccionamos un vector de n símbolos de forma que podamos obtener la información enviada desde la recibida con la probabilidad más alta posible. Si las entradas no son independientes, entonces menos información podrá ser enviada a través del canal.

También podemos derivar cierta información del Lema 2.3.1: Si queremos que la información mutua sea cercana a $k = nC$, una versión nos dice que $H(X) = k$, con lo que $H(X|Y)$ debería ser cero, con lo que la salida sería siempre el mismo mensaje. Por otra parte, se podría que

$$H(Y_1, Y_2, \ldots, Y_n) \approx nC + nH(Y|X) = nH(Y)$$

con lo que la salida sería cercana a n símbolos independientes, con lo que el canal debería extender los vectores de salida lo suficiente para eliminar el efecto de los de entrada.

2.4. Estructuras de datos

Hasta ahora, se ha considerado que la transmisión de información se realizaba mediante códigos y símbolos independientes. No obstante, la información es normalmente ordenada en ciertas estructuras de forma que el receptor también conoce dichas estructuras para una descodificación más efectiva.

Definición 2.4.1. *Una* estructura de datos *es una entidad de datos que puede ser independientemente almacenada, comunicada y transmitida. Consiste en un* encabezado*, un* cuerpo de datos *y un* cuerpo de paridad*.*

Se asume que la estructura se compone de un número fijo de símbolos. Cuando esta información es transmitida a través del canal, es codificada (normalmente en otro alfabeto) a través de un código corrector de errores lo más adecuado posible. La estructura transmitida consistirá en un número fijo de símbolos del alfabeto usado en el canal (En general, trabajaremos con un canal binario).

Definición 2.4.2. *Una* estructura de transferencia *es una entidad de información codificada que se transmite a través del canal. Consiste en un* encabezado *(posiblemente codificado por separado) y un* cuerpo*.*

El encabezado en la estructura de transferencia funciona como indicador del comienzo de la estructura, proceso que se conoce como sincronización estructural. También incluye información sobre la ruta a seguir, así como la dirección del receptor. En cambio, el encabezado de la estructura de datos contiene información relevante para la interpretación del mensaje, como el número de estructuras, identificación de los contenidos, etc.

El cuerpo de datos consiste en un número K_f de símbolos binarios, que en general está protegido por un cuerpo de paridad de B_f símbolos, que actúan como comprobaciones de paridad para poder verificar la integridad de la estructura recibida. Esta estructura de datos es codificada mediante un código de ratio R, aunque el encabezado puede ser codificado de manera diferente para simplificar el proceso. Denominamos entonces H_f al número de símbolos usados para transmitir el encabezado.

Definición 2.4.3. *La* eficiencia de la transmisión*, η es el ratio del número de símbolos del canal que representan los datos y la paridad respecto del número total de símbolos transmitidos. Los* gastos de la transmisión *son* $1-\eta$*.*

De esta definición:

Lema 2.4.1.

$$1-\eta = \frac{H_f}{K_f + B_f + H_f} \tag{2.8}$$

Al igual que hablamos en el anterior capítulo, podemos enviar de manera efectiva K símbolos transmitiendo N a través del canal si $\frac{K}{N}$ es menor que la capacidad del canal, C.
Además, de la misma forma que las definiciones anteriores tenemos:

Definición 2.4.4. *La* eficiencia de la información*, γ, es el ratio $\gamma = \dfrac{R}{C}$. La* información perdida *es* $1-\gamma$*.*

Lema 2.4.2.

$$\gamma = \frac{K_f}{(K_f + B_f)C} \tag{2.9}$$

De la anterior fórmula, observamos que la mayor eficiencia se obtiene cuanto mayor es la longitud de la estructura.

Para medir la calidad de las estructuras, nos basaremos en dos números:

Definición 2.4.5. *La* probabilidad de un error indetectado*, $P(ue)$ es la probabilidad de que la estructura sea declarada libre de errores cuando en realidad partes de la estructura no sean correctas.*

Definición 2.4.6. *La* probabilidad de un error de estructura*, $P(fe)$ es la probabilidad de que el receptor distinga un error en la estructura, pero no sea capaz de corregirlo (al menos con la suficiente confianza).*

Obviamente, la introducción de comprobaciones de paridad se realiza con la intención de solucionar los errores en las estructuras. Un tipo de comprobación bastante usual es usar un *código cíclico binario*, generado por el polinomio

$$g(x) = p(x)(x+1)$$

donde $p(x)$ es un polinomio primitivo de grado m, teniendo el código parámetros $(n, k, d) = (2^m - 1, 2^m - m - 2, 4)$.
Normalmente este código no ayuda a la corrección de errores, pero puede detectar errores en la transmisión. Así, si más de dos errores ocurren, la probabilidad de que la estructura sea aceptada se estima con la probabilidad de que una secuencia aleatoria genere la secuencia nula, es decir,

$$P(ue) \approx P[t \geq 2] \cdot 2^{-m-1} \tag{2.10}$$

Una estructura de longitud N puede ser dividida en m palabras de un (n, k) código, tomando k tal que divida a N. Además, podemos establecer cotas para dichas probabilidades a partir de las probabilidades de error y fallo que vimos anteriormente:

Lema 2.4.3. *Si la estructura consiste en $\frac{N}{n}$ bloques, y el ruido es independiente, la probabilidad de una estructura correcta es*

$$1 - P(ue) = (1 - P_{error})^{\frac{N}{n}} \tag{2.11}$$

Mientras que la probabilidad de un fallo de estructura será

$$P(fe) = 1 - (1 - P_{fallo})^{\frac{N}{n}} \tag{2.12}$$

En este sentido, se consiguen mejores resultados si la información es transmitida en pequeñas estructuras. Por todo ello, para conseguir la mayor eficiencia es conveniente combinar la corrección de errores con las comprobaciones de las estructuras. Usando 16 o 32 bits para este fin es posible hacer la probabilidad de fallo indetectado muy pequeña y seguir teniendo estructuras relativamente pequeñas para aprovechar al máximo el código corrector.

2.5. Códigos convolucionales

Los códigos convolucionales destacan por tratarse de unos códigos en los que la longitud es variable, por lo que se definen reglas de codificación locales. Para esta sección, seguiremos suponiendo que todos los códigos son binarios, así como denotaremos N a la longitud de la estructura codificada y R al ratio del código.
Sea $R = \frac{k}{n}$, k y n enteros positivos, y N un múltiplo de n.

Definición 2.5.1. *Una* estructura de entrada*, u, es un vector binario de RN símbolos de información. Un* vector de información*, $v(u)$, es un vector binario de longitud N tal que*

$$v_j = \begin{cases} u_i & \text{para } j = \lfloor \frac{i}{R} \rfloor \\ 0 & \text{en otro caso} \end{cases}$$

Definición 2.5.2. *La* codificación convolucional *del vector de información $v(u)$ mediante un* vector generador*, g, con $g_0 = g_m = 1$, transforma $v(u)$ en el vector de N-dimensional*

$$y_j = \sum_{i=0}^{m} g_i v_{j-i} \tag{2.13}$$

donde la suma es en módulo 2, $N > m$ y los índices se interpretan en módulo N.

La fórmula de (2.13) se denomina *convolución cíclica* de g y v, dando nombre a los códigos que estamos describiendo en esta sección. También se puede realizar de forma simple una convolución no cíclica, añadiendo mR ceros al final de la estructura de entrada.

Si el ratio es la división de enteros muy pequeños, $R = \frac{k}{n}$, la secuencia codificada se puede dividir en bloques de longitud n, teniendo cada uno k símbolos de información y $n-k$ símbolos de paridad. Para simplificar la lectura, las secuencias se separan mediante periodos.

Lema 2.5.1. *Para un vector generador dado, g, la codificación convolucional define un (N, K) código lineal de dimensión $K \leq RN$.*

La codificación de un código convolucional puede ser también expresada mediante una matriz que refleje la norma de codificación.

Definición 2.5.3. *La* matriz generadora *$K \times N$ dimensional para un código convolucional, $G(N)$, se define mediante*

$$y = uG(N)$$

donde y es el vector codificado mediante (2.13).

Definición 2.5.4. *Una codificación con secuencia generadora g y ratio R es* no catastrófica *si todos los códigos bloque obtenidos por la codificación convolucional (2.13) con N lo suficientemente largo tienen dimensión $K = RN$.*

Lema 2.5.2. *Una secuencia generadora es no catastrófica sí y solo sí hay una longitud L tal que una secuencia codificada puede contener como máximo un número L de ceros consecutivos cuando el vector de información tiene al menos un 1 por cada m bits.*

Este hecho hace que las *codificaciones catastróficas* tengan la propiedad de que una entrada no nula periódica podría producir una salida con solo un segmento no nulo finito. También, dos entradas diferentes en la mayoría de las posiciones podrían tener la misma salida salvo un número finito de símbolos, lo que provocaría demasiados errores en estructuras lo suficientemente largas (proceso que se conoce como *propagación catastrófica del error*).

Definición 2.5.5. *La* memoria *de un código es el entero $M = \lceil R(m+1) \rceil - 1$.*

Los parámetros de un código convolucional se dan como (n, k, M). Si $R = \frac{k}{n}$ y la longitud de g, $m+1$, es $(M+1)n$, entonces el codificador almacena kM símbolos de entrada. En general,

Lema 2.5.3. *Cada símbolo codificado depende como máximo de los $M+1$ símbolos de información previos (tomando los índices en módulo RN).*

Con toda esta información, es posible definir el código mediante una manera alternativa:

Definición 2.5.6. *Un (n, k, M) código convolucional es la secuencia de de códigos de longitud $N = jn$ obtenidos mediante la codificación (2.13).*

Si se quiere enfatizar en en esta codificación cuando la estructura es relativamente pequeña, hablamos de un código de *cola mordida* (*tail-bitting*). Se trata de una ligera transición entre códigos bloque y convolucionales, bastante útiles si es pequeño como descripción alternativa de un código bloque con propiedades similares; o si es grande como definición alternativa de un código convolucional.

A la hora de analizar los códigos, las matrices de comprobación de paridad resultan de gran importancia. Por ello, desarrollemos este concepto dentro de los códigos convolucionales. Para un código de cola mordida de longitud fija N, se puede encontrar la matriz de forma similar a como hemos visto para códigos lineales, de forma que

$$G(N)H(N)^T = 0$$

No obstante, si queremos que H tenga una estructura similar a G (en particular, teniendo las partes nulas de las filas independientes de N), el problema de hallar la matriz de paridad se complica.

Para códigos no catastróficos de ratio $R = \frac{1}{2}$, la matriz de paridad tiene la misma forma que G, pues se puede obtener la secuencia no nula h mediante la inversa de g. El símbolo de paridad en la posición j se asocia con una fila de H tal que su última entrada no nula sea en la posición j.

Para códigos de ratio diferente, se puede obtener h resolviendo un sistema de ecuaciones lineales. Si tomamos la fila de H con la última entrada no nula en posición j, la última fila de G que debemos considerar es aquella con el primer símbolo no nulo en el mismo bloque. Si M es la memoria del código, las M filas previas tendrán símbolos no nulos en el mismo bloque, por lo que h debe ser ortogonal a tales filas. No queda claro entonces la longitud exacta de h, pues podría ocurrir que haya más variables que ecuaciones. En ese caso, podemos escoger h como el vector no nulo más corto ortogonal al segmento de G que se requiere. Para simplificar, sólo consideraremos los casos en los que hay una única solución del siguiente modo:

Definición 2.5.7. *Una codificación convolucional de ratio R se denomina* regular *si la matriz generadora tiene coeficientes no nulos $g_{i,j}$ en la fila i solo para $\left\lfloor \frac{i}{R} \right\rfloor \leq j \leq \left\lfloor \frac{i+M+1}{R} \right\rfloor - 1$. De manera similar, los coeficientes de la matriz de paridad son no nulos solo para $\left\lceil \frac{i}{1-R} \right\rceil \leq j \leq \left\lceil \frac{i+M+1}{1-R} \right\rceil - 1$, y la parte no nula de la i-ésima fila de cada matriz es el único vector ortogonal no nulo al correspondiente segmento de la otra matriz.*

Es decir, la parte no nula de una fila de H será $h = (h_{m'}, h_{m'-1}, \ldots, h_0)$, donde $M + 1 = \lceil (1 - R)(m' + 1) \rceil$.

Definición 2.5.8. *El* síndrome *de una secuencia recibida se define como*

$$s_j = \sum_{i=0}^{m'} r_{j-i} h_i \tag{2.14}$$

donde $s_j = 0$ si r es una palabra codificada.

Lema 2.5.4. *Un error individual afecta como máximo a $M+1$ bits del síndrome.*

Por otro lado, es importante de cara a la corrección de errores establecer términos similares a la distancia en códigos convolucionales. Si eliminamos los M últimos bits de la estructura de entrada u, o los hacemos nulos, la codificación (1.11) se convierte en una codificación no cíclica. Esta forma de adaptar los códigos se conoce como *terminación*.

Definición 2.5.9. *Un (n,k,M) código convolucional* terminado *es una secuencia de códigos bloque de longitud $N=jn$ obtenidos mediante la codificación (2.13) con los últimos M bits de entrada fijados nulos.*

El ratio del código terminado será menor, pero la diferencia es escasa en caso estructuras largas. Si consideramos j filas consecutivas de la matriz generadora y el segmento de la secuencia codificada donde dichas filas no son nulas, la longitud N' de la secuencia satisface $j=\lfloor R(N'-m)\rfloor$. El conjunto de de secuencias codificadas forma un (N',j) código lineal denominado j-ésimo código terminado.

Definición 2.5.10. *La* j-ésima distancia de fila *del código convolucional es el mínimo peso de un vector no nulo generado por j filas consecutivas.*

Esta es la mínima distancia del j-ésimo código terminado. Dado que cada código contiene todos los códigos más cortos que él,

Lema 2.5.5. *La distancia de fila es una función decreciente respecto a j.*

Definición 2.5.11. *La* distancia libre, *d_f es el mínimo valor de las distancias filas.*

En la mayoría de los códigos, la distancia de fila es constante desde $j=1$ o se alcanza la distancia libre con j muy pequeño. Se trata de una medida muy importante para la capacidad de corrección de errores de un código convolucional.

Lema 2.5.6. *Dado un código convolucional no catastrófico, todos los códigos de mordedura de cola de longitud mayor que cierta constante N' tienen mínima distancia d_f.*

Teorema 2.5.1. *Un código convolucional con distancia libre d_f puede corregir cualquier error de patrón de peso menor que $\frac{d_f}{2}$ si N es lo suficientemente largo.*

Teorema 2.5.2. *La distancia libre de un código convolucional está superiormente acotada por la distancia mínima del mejor (N',j) código bloque, donde $j=\lfloor R(N'-m)\rfloor$.*

Por otro lado, podemos escribir la regla de codificación lineal de forma matricial:

$$y_j=\sum_{i\geq 0} G_i u_{j-i} \tag{2.15}$$

donde el bloque de entrada, u tiene longitud k y el de salida, y, longitud n, siedno las matrices G_i de dimensiones $k\times n$. Esta ecuación permite métodos de sistemas lineales o series de potencias , en particular mediante funciones generadoras. Esto nos permite expresar el código como una función polinomial generadora

$$y(D)=u(D)G(D) \tag{2.16}$$

Dado que se asume una memoria finita, las entradas en $G(D)$ serán polinomios de grado m como máximo. Además, se puede transformar mediante escalares y operaciones de fila en la forma sistemática, con la matriz identidad de dimensión $k\times k$ seguida de funciones racionales de D. Esta transformación es particularmente conveniente para $(n,1)$, donde se puede escribir la matriz generadora como

$$G(D)=(G_1(D),G_2(D),\ldots,G_n(D))$$

Teorema 2.5.3. *Un $(n,1)$ código es no catastrófico si y solo si los G_i no tienen ningún factor común no trivial.*

Códigos de memoria unidad

Un *código de memoria unidad* (UMC) es generado por una regla de codificación del tipo

$$y_j = u_j G_0 + u_{j-1} G_1 \tag{2.17}$$

Siendo G_0 una matriz generadora de un (n, k) código y G_1 una matriz $k \times n$, de rango $m \leq k$. Si $m < k$, es conveniente escribir las matrices generadoras como

$$G = \begin{bmatrix} G_0 & G_1 \end{bmatrix} = \begin{bmatrix} G_{00} & 0 \\ G_{01} & G_{11} \end{bmatrix}$$

Definición 2.5.12. *El código de memoria unidad generado por G es* regular *si los subespacios lineales generados por G_{00}, G_{01} y G_{11} solo comparten el vector nulo.*

Podemos establecer un algoritmo para descodificar un UMC:
Entrada: Palabra recibida r.

1. Asumiendo que los errores han sido corregidos hasta el bloque $i-1$, se calculan los siguientes bits del síndrome de r, $s_i = H_0 r_i^T$. Dado que H_0 es además la matriz de comprobación de paridad de un código Hamming extendido, se puede corregir inmediatamente errores simples y detectar errores dobles.

2. Si un error simple ha sido corregido, podemos eliminar el termino $H_1 r_i^T$ del siguiente síndrome.

3. Si hay dos errores en el bloque i, serán detectados como una paridad par y síndrome no nulo. En esta situación, la descodificación es retrasada hasta encontrar un bloque con paridad par.

4. En los siguientes bloques se detecta sólo la paridad. Bloques con paridad impar contiene un único error que será corregido más tarde.

5. Dado que se asume que j bloques contienen como máximo $j+1$ errores, y el primero dos, el siguiente bloque con paridad par será corregido.

6. Podemos ahora corregir los bloques previos en la dirección inversa usando H_{11}, dado que todas las columnas de esta matriz son distintas.

7. Finalmente los dos errores del bloque i pueden ser descodificados usando un método de descodificación de doble error, dado que ambos síndromes han sido calculados.

Salida: Palabra descodificada.

Capítulo 3

Juegos Malabares

3.1. Malabares simples

En esta última parte del trabajo, vamos a tratar algunos de los aspectos más destacados de otras investigaciones de Claude Shannon que, a pesar de no haber tenido una trascendencia vital en la historia como sí ha ocurrido con la teoría de la información; conviene destacar, para completar la imagen de Shannon como un matemático multidisciplinar que trabajó en muchas y variadas áreas de las matemáticas. Uno de estos casos es la relación existente entre las matemáticas y los juegos malabares, o simplemente malabares.

La historia de los malabares es antigua. Desde sus primeras apariciones en grabados egipcios encontrados en las tumbas egipcias de Beni Hassan, se conoce que dichos juegos han existido en numerosas civilizaciones tanto occidentales como orientales a lo largo de la historia. Como no podía ser de otra manera, era cuestión de tiempo que las matemáticas, presentes en todos los conceptos del universo, terminaran estudiando este movimiento que trata sobre ideas tan lógicas como patrones y secuencias. Precisamente, uno de los primeros matemáticos en estudiar tales movimientos y enunciar sus propiedades (los famosos teoremas de Shannon que veremos más adelante) fue Claude Shannon en su publicación *Scientific Aspects of Juggling* que data de 1970.

A pesar de ello, no seguiremos un razonamiento histórico sino que comenzaremos desarrollando las ideas principales de los malabares desde la simplicidad. La relación entre matemáticas y malabares comienza también por la pura necesidad: los propios malabaristas, ante la imposibilidad de poder transmitir sus ideas y trucos mediante vídeos (bien por la inexistencia de Internet, o incluso con este, como ya hemos visto a lo largo de todo el trabajo, supone un gran ahorro de tiempo y capacidad enviar información en un código matemático en vez de un vídeo, de considerable mayor cantidad de información, se crea un *lenguaje de malabares*, siendo el más común, y con el que trabajaremos, la *notación transposicional* o *SITESWAP* que representa el movimiento de un malabar mediante patrones numéricos. Para mayor simplificación, en todo el trabajo supondremos que estamos trabajando con bolas individuales que son lanzadas y recibidas mediante el mismo movimiento de mano respectivamente. Comencemos con los malabares simples:

Supongamos una única persona se encuentra realizando malabares con un número b de bolas y dos manos. Diremos que estos malabares son *simples* si se cumplen tres condiciones:

1. El ritmo es constante, es decir, los lanzamientos ocurren en momentos equidistantes de tiempo. Si denominamos T a este tiempo, empezando los malabares en un tiempo t_0 entonces los siguientes lanzamientos ocurrirán en los tiempos t_0, $t_0 + T$, $t_0 + 2T, \ldots$ De manera general, se supondrá que $t_0 = 0$ y el ritmo T es la unidad, por lo que se hará referencia a instantes o lanzamientos (entendido como el instante de lanzamiento) de

manera análoga.

2. Los patrones son periódicos. Cada periodo tiene un número p de lanzamientos, tras los cuales se vuelven a repetir los mismos lanzamientos. Se asume que siguiendo dicho periodo, sin ningún cambio, el juego no tiene principio ni fin (el malabarista lleva haciendo lo mismo desde siempre y seguirá así).

3. Sólo una bola es recogida y lanzada en cada momento, siendo además la misma bola recibida la que es lanzada. Por ahora, se asumirá que el tiempo en el que una bola permanece en la mano es nulo.

Si la tercera condición no se cumple, más de una bola es lanzada o recibida a la vez, entonces se trata de *malabares múltiples* o *multiplex*, los cuales omitiremos en este trabajo pues su estudio es similar.

Cada bola lanzada sigue un patrón que se define por la altura k del lanzamiento. Esta altura no representa una medida real de longitud, sino que se trata de un número entero no negativo que indica, si se empieza a contar en ese momento, en que recepción o instante llegará esa bola o, equivalentemente, si se realiza un lanzamiento de altura k entonces ocurrirán otros $k-1$ lanzamientos antes de que esa bola llegue a una mano.

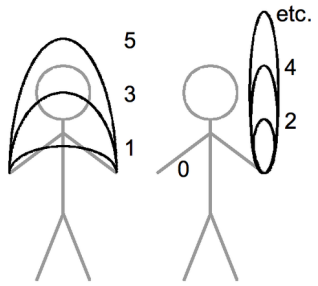

Figura 3.1: Lanzamientos a diferentes alturas

De esta forma, en tanto que la mano con la que se lanza/recibe va alternando, es fácil observar que los lanzamientos pares vuelven a la misma mano mientras que los impares cambian de mano. Para alturas $k \geq 3$ esta definición es fácil de comprender. No obstante conviene matizarlo para los otros casos:

- Si $k = 0$, no se está realizando ningún lanzamiento real. En este caso, se trata de un turno de espera en la que la mano se mantiene vacía.
- $k = 1$ equivale a un cambio horizontal rápido de mano de la bola, siendo esta inmediatamente lanzada de nuevo.
- Aunque en aspectos matemáticos $k = 2$ no supone ningún problema, en la práctica (que equivaldría a un lanzamiento sobre la misma mano muy corto), se entiende como un turno de espera con la bola en la mano, el equivalente al caso $k = 0$ pero con la mano llena.

Como ya hemos descrito, estos lanzamientos se asocian en periodos. En un periodo de p lanzamientos, el malabarista repite indefinidamente los lanzamientos descritos en el periodo. Si esto es posible de hacer, se denomina *secuencia malabar*.

Definición 3.1.1. *Una secuencia* $s = \{a_i\}_{i=0}^{p-1}$ *de enteros no negativos es* malabar *si se puede realizar un juego malabar indefinido ejecutando de manera periódica lanzamientos de altura* a_i *respectivamente.*

Para visualizar estas secuencia, podemos recurrir a los *diagramas malabares*. Se trata de grafos cuyos vértices representan las manos (que se van alternando) mientras las aristas supondrían la trayectoria de las bolas a través del tiempo. En la Figura 3.2 podemos observar algunos disgramas de diferentes secuencias malabares. Precisamente, un diagrama malabar es una forma directa de comprobar dos cosas:La secuencia es un malabar simple si al dibujar la

secuencia solo hay un máximo de dos aristas por vértice (es decir, sólo una bola pasa cada momento por cada mano); y el número de bolas necesarias para realizar esa secuencia es, si trazamos una línea vertical que no se cruce con ningún vértice ni ninguna intersección entre las aristas, el número de intersecciones de la línea vertical con las aristas (que supone el número de bolas en el aire en ese preciso momento). De esta forma, es fácil observar que en nuestro ejemplo las secuencias 3 o 414 se realizan con 3 bolas mientras que la secuencia 255 requiere de 4 bolas.

Entre otras ideas, una forma rápida de comprobar si una secuencia es malabar (y un concepto

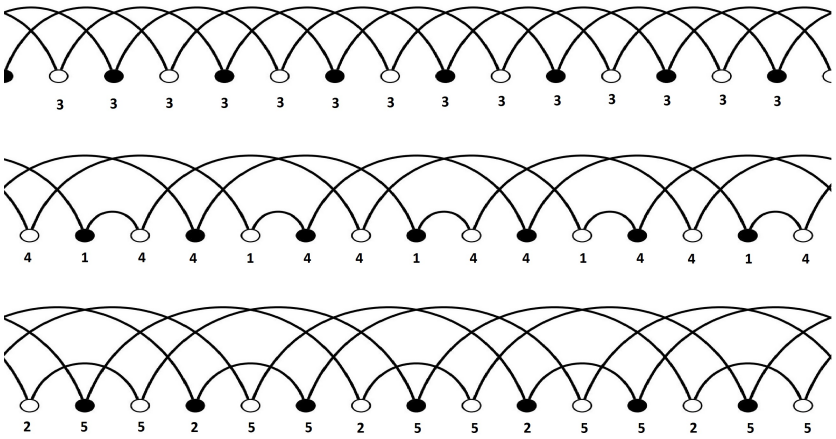

Figura 3.2: Diferentes diagramas para las secuencias 3, 414 y 255 respectivamente

que utilizaremos más adelante) es que si la secuencia no tiene dos números consecutivos siendo el primero una unidad mayor, es decir, de la forma $\ldots n(n-1)\ldots$; pues ambas bolas caerían en la misma mano a la vez tras $n-2$ lanzamientos después.

No obstante, depender de un diagrama para el cálculo de bolas y comprobación de que la secuencia puede ser un malabar requiere de un gran esfuerzo en algunos casos, además de nos ser una opción que pueda ser fácilmente programable (un objetivo importante del desarrollo de esta teoría de modelización matemática de los malabares es también el hecho de poder ser estudiados y ejecutados por computadoras). Para evitar estos problemas, vamos a desarrollar conceptos para poder concluir con dos teoremas que resolverán nuestros dilemas plateados: el Teorema de la Media para el número de bolas y el Test de Permutación para comprobar si una secuencia es malabar.

Teorema de la Media

Dada una secuencia, nuestro objetivo es determinar cuántas bolas son necesarias para hacer malabares siguiendo el patrón.

Sea $J : \mathbb{Z} \to \mathbb{Z}_{\geq 0}$ una función de enteros en enteros no negativos tal que a cada momento i le corresponde un lanzamiento de altura $k(i) = J(i) - i$, es decir, $J(i)$ representa el momento exacto en el tiempo (o el lanzamiento) en el que la bola lanzada en i es recibida (y vuelta a lanzar. En el caso de una secuencia malabar finita $\{a_i\}_{i=0}^{p-1}$, entonces $J(i) = i + a_{i \mod p}$. Si J es biyectiva y las alturas son no negativas, $J(i) \geq i$, entonces a cada momento le corresponde un único lanzamiento y tendrá una única llegada. En ese caso se conoce J como *función malabar*. Calculemos el número de bolas:

Teorema 3.1.1. Teorema de la Media

Sea J una función malabar tal que las alturas $k(i) = J(i) - i$ están acotadas entonces el límite

$$\lim_{|I|\to\infty} \frac{\sum_{i\in I} k(i)}{|I|} \tag{3.1}$$

existe, es finito y representa el número de bolas con las que se realiza un juego malabar siguiendo la función J. En el límite, este se calcula sobre todos los intervalos $I = \{a, a+1, a+2\ldots, b\}$, siendo $|I| = b - a + 1$.

Corolario 3.1.1. Teorema de la Media para Secuencias Malabares
El número de bolas necesario para realizar una secuencia malabar es su media aritmética.

Demostración. El corolario es una consecuencia inmediata del teorema. Demostremos este: Sea H la altura máxima posible (la cota superior de k). Consideremos un intervalo I tal que $|I| > H$. Entonces para cada arista del diagrama tiene al menos un nodo en I. Por lo tanto, el número de bolas b es finito, en tanto que que las órbitas también lo son. Si tomamos una bola en concreto, la suma de los valores $k(i)$ en los momentos en los que la bola es recogida/lanzada está acotada inferiormente por $|I| - 2H$ y superiormente por $|I| + 2H$(la bola recorre el intervalo I más los dos lanzamientos con los que entra y sale del intervalo, de altura máxima H). Aplicado a todas las bolas, se cuentan todos los lanzamientos del intervalo, y dividiendo entre $|I|$:

$$\frac{b(|I| - 2H)}{|I|} \leq \frac{\sum_{i\in I} k(i)}{|I|} \leq \frac{b(|I| + 2H)}{|I|}$$

Haciendo tender $|I| \to \infty$, ambas cotas tienden a b. □

Este teorema nos sirve también como una forma de comprobación si una secuencia puede ser malabar. Obviamente, si la media de una secuencia no es un número entero es imposible que esta sea una secuencia malabar, pues no se pueden hacer malabares con un número no entero de bolas. Nótese que esta condición es necesaria pero no suficiente, la secuencia 432, por ejemplo, se realizaría con tres bolas pero no es una secuencia malabar, pues las tres bolas terminarían en la misma mano.

Test de Permutación

Centrémonos ahora en dar un método para comprobar si una secuencia de números enteros no negativos dada es una secuencia malabar. Para ello, existe el siguiente método:

Teorema 3.1.2. Test de Permutación
Sea $s = \{a_i\}_{i=0}^{p-1}$ una secuencia de enteros no negativos. s es una secuencia malabar sí y sólo sí la función

$$\phi_s : [p] \to [p]\,; \quad \phi_s(i) = i + a_i \mod p$$

es una permutación del conjunto $[p] = \{0, 1, 2, \ldots, p-1\}$.

Demostración. Con la notación que hemos usado anteriormente, sea J la función malabar asociada a esta secuencia, de forma que $a_i = k(i)$. Para $i \in [p]$, $J(i) = \phi_s(i)$. Si s es una secuencia malabar, J debe ser biyectiva, y por lo tanto es necesariamente una permutación de $[p]$.
Sea ahora ahora una secuencia $\{a_i\}_{i=0}^{p-1}$ tal que $i + a_i \mod n$ es una permutación de $[p]$. Extendiendo la secuencia periódicamente podemos definir la función $J(i) = i + a_{i \mod p}$. J es inyectiva $\mod p$ al ser una permutación $\mod p$. Entonces sí $J(i) = J(j)$, $i \equiv j \mod p$ y entonces $a_i = a_j$. Al ser $J(i) = i + a_i = j = a_j = J(j)$ se concluye que J es inyectiva. Por otro lado, sea $u \in \mathbb{Z}$. Dado que $i + a_i \mod p$ es una permutación, existe un t tal que $J(t) = t + a_t \equiv u \mod p$. Podemos tomar un apropiado múltiplo n de p de forma que $J(t') = J(t + np) = t + np + a_t = u$; con lo que se concluye que J es sobreyectiva y por lo tanto biyectiva. □

A partir de este teorema podemos relacionar algunas secuencias malabares con otras mediante varias operaciones. El primer cambio es una consecuencia directa del teorema:

Lema 3.1.1. Cambio vertical
Sea $s = \{a_i\}_{i=0}^{p-1}$ *una secuencia de enteros no negativos y* d *un número entero mayor o igual que* $-\min\{a_0, a_1, \ldots, a_{p-1}\}$. *Sea* $s' = \{a_i + d\}_{i=0}^{p-1}$. s *es secuencia malabar sí y sólo sí* s' *lo es. Se conoce esta transformación como* cambio vertical de distancia d.

A partir del teorema de la media, es fácil observar que un cambio vertical de distancia d supone el aumento (o disminución en el caso negativo) de d bolas. Esto es lo que ocurrió en nuestro ejemplo de la Figura 3.2 con las secuencias 414 y 552, aunque previamente debemos conocer otro cambio posible.
Ciertamente, en una secuencia periódica los patrones se repiten indefinidamente, por lo que el propio diagrama malabar se vería inalterado si empezamos en un instante previo o posterior, aunque la propia secuencia cambia los números. Esta caso es el correspondiente al *cambio cíclico*.

Lema 3.1.2. Cambio cíclico
Sea $s = \{a_i\}_{i=0}^{p-1}$ *con* $p \geq 2$ *una secuencia de enteros no negativos y sea* s' *la secuencia* $a_{p-1}a_0a_1 \ldots a_{p-2}$. s *es secuencia malabar sí y sólo sí* s' *lo es. En ese caso el número de bolas (y la media) de ambas secuencias es el mismo.*

Demostración. Aunque su descripción gráfica es bastante concluyente, es muy simple aportar una demostración en vías del Teorema 3.1.2. Con la misma notación, si denominamos vector de prueba al vector $(\phi_s(0), \phi_s(1), \ldots, \phi_s(p-1))$, ϕ_s es una permutación sí y sólo sí el vector contiene todos los elementos de $[p]$. El vector de prueba de s' es

$$(1, 1, \ldots, 1) + (\phi_s(p-1), \phi_s(0), \phi_s(1), \ldots, \phi_s(p-2)) \mod p$$

el cual representa una permutación sí y sólo sí ϕ_s lo es.
El número de bolas es directo, pues la media es la misma. □

En general, se considera que dos secuencias que sólo varían en cambios cíclicos son prácticamente la misma (el malabar es idéntico salvo por un desplazamiento de tiempo).

Procedamos con el último de nuestros cambios: supongamos que al realizar una secuencia dada, en un instante dado i decidimos cambiar la altura a_i de este lanzamiento, de forma que la bola sea recogida en el momento en el que también caería la bola lanzada en un momento posterior j. Si cambiamos también la altura a_j para encajar el espacio que se ha creado, estamos obteniendo una nueva secuencia malabar igual a la anterior en todas las posiciones excepto en las posiciones i y j, de forma que se intercambian sus alturas:

Lema 3.1.3. Intercambio lateral
Sea $s = \{a_i\}_{i=0}^{p-1}$ *con* $p \geq 2$ *una secuencia de enteros no negativos y sean* i, j *tales que* $0 \leq i < j \leq p-1$ *y* $j - i \leq a_i$. *Sea* $s_{i,j}$ *la secuencia igual a* s *salvo por los elementos* i *y* j*-ésimos, cuyos nuevos valores son* $a_j + j - i$ *y* $a_i - j + i$ *respectivamente.* s *es secuencia malabar sí y sólo sí* s' *lo es. En ese caso el número de bolas (y la media) de ambas secuencias es el mismo.*

Demostración. Sea $d = j - i$ la distancia entre los momentos. Si tomamos el vector de prueba que utilizamos en la anterior definición, sólo las entradas i y j de este han cambiado con respecto a la secuencia original. La posición i ha cambiado de $i + a_i \mod p$ ha cambiado a $i + d + a_{i+d}$ mód p, mientras que en la posición $j = i + d$ el cambio es de $i + d + a_{i+d} \mod p$ a $(i + d) + a_i - d$ mód $p = i + a_i \mod p$. Es decir, estas dos posiciones también se han intercambiado en el vector de prueba. Entonces ϕ_s es una permutación sí y sólo sí $\phi_{s_{i,j}}$ lo es también.
El número de bolas es directo, pues la media es la misma. □

En concreto, a partir de las dos últimas operaciones (que conservan el número de bolas) podemos variar entre todas las secuencias malabares del mismo número de bolas. Para ello, podemos establecer un algoritmo que a partir de cualquier secuencia s de enteros no negativos opere hacía la secuencia constante de p veces bolas, utilizando sólo cambios cíclicos e intercambios laterales:

1. Si s es constante, entonces terminamos el algoritmo, en otro caso,

2. Mediante cambios cíclicos, reordenar la secuencia con cambios cíclicos de manera que el primer elemento a_0 sea de altura máxima y el segundo, a_1 sea estrictamente menor que a_0. Si esta diferencia es de una unidad, tenemos una secuencia de la forma $n(n-1)\ldots$, por lo que no es malabar: detener el algoritmo. En otro caso:

3. Ejecutar un intercambio lateral entre las posiciones 0 y 1. Con la nueva secuencia, volver al paso 1.

Cada vez que el algoritmo ejecuta un intercambio lateral, el número de lanzamientos de altura máxima desciende (en tanto que tiende a igualar lanzamientos de altura máxima con otros menores), por lo que el algoritmo termina necesariamente en un número finito de pasos. Como el número de bolas (y por tanto la media) se mantiene constante, el algoritmo ha de terminar en la secuencia pedida. Además, las operaciones descritas transforman secuencias malabares en malabares e idéntico para las no malabares. En tanto que la secuencia constante de p veces b es malabar, es imposible llegar hasta ella desde una secuencia no malabar. Por ello, al ir igualando las alturas de sus lanzamientos estas secuencias terminan en alguna secuencia del tipo $n(n-1)\ldots$
Puesto que todas las dos operaciones se pueden realizar de manera inversa con las misma propiedades, obtenemos el siguiente resultado:

Lema 3.1.4. *La secuencia malabar b-constante de periodo p se puede transformar en cualquier secuencia malabar de de periodo p y b bolas mediante cambios cíclicos e intercambios laterales.*

Por ejemplo, siguiendo con los diagramas expuestos, es fácil obtener la secuencia 414 desde 333:

$$333 \xrightarrow{lateral} 423 \xrightarrow{lateral} 441 \xrightarrow{cíclico} 144 \xrightarrow{cíclico} 414$$

Para terminar con esta sección, el Test de Permutación nos puede servir también para obtener un método con el que hallar todas las posibles secuencias de b bolas y periodo p, a partir de implementar un un razonamiento inverso al de la comprobación de una secuencia. Si comenzamos con una permutación de $[p]$ entonces podemos llegar a las secuencias malabares que al aplicar el test nos llevarían a esta permutación.
Sea P el vector de prueba de una permutación de $[p]$. Calculamos el vector

$$P' = (P - (0, 1, \ldots, p-1)) \quad \text{mód } p$$

La suma de los elementos de P y el vector $(0, 1, \ldots, p-1)$ es la misma, por lo que la suma de los elementos de P' mód p es 0, es decir, su suma es cierta cantidad ap, por lo que la media de P' es un entero a, necesariamente $0 \leq a \leq p-1$. Sea $b' = b - a$. Cualquier vector Q de dimensión p de enteros no negativos cuya suma de elementos sea b' esta relacionado con una secuencia malabar que deriva en la permutación de P de forma que la secuencia forma los elementos de $P' + bQ$.
A partir de todas las permutaciones de periodo p se pueden construir todas las secuencias malabares.

Secuencias malabares de 3 bolas y periodo 3						
P	$(0,1,2)$	$(2,0,1)$	$(1,2,0)$	$(1,0,2)$	$(2,1,0)$	$(0,2,1)$
P'	$(0,0,0)$	$(1,1,1)$	$(2,2,2)$	$(1,2,0)$	$(2,0,1)$	$(0,1,2)$
$b'=b-a$	3	2	1	2	2	2
Q	$(3,0,0)$ $(0,3,0)$ $(0,0,3)$ $(2,1,0)$ $(2,0,1)$ $(1,2,0)$ $(1,0,2)$ $(0,2,1)$ $(0,1,2)$ $(1,1,1)$	$(2,0,0)$ $(0,2,0)$ $(0,0,2)$ $(1,1,0)$ $(1,0,1)$ $(0,1,1)$	$(1,0,0)$ $(0,1,0)$ $(0,0,1)$	$(2,0,0)$ $(0,2,0)$ $(0,0,2)$ $(1,1,0)$ $(1,0,1)$ $(0,1,1)$	$(2,0,0)$ $(0,2,0)$ $(0,0,2)$ $(1,1,0)$ $(1,0,1)$ $(0,1,1)$	$(2,0,0)$ $(0,2,0)$ $(0,0,2)$ $(1,1,0)$ $(1,0,1)$ $(0,1,1)$
$P'+bQ$	**(9,0,0)** $(0,9,0)$ $(0,0,9)$ **(6,3,0)** **(6,0,3)** $(3,6,0)$ $(3,0,6)$ $(0,6,3)$ $(0,3,6)$ **(3,3,3)**	**(7,1,1)** $(1,7,1)$ $(1,1,7)$ **(4,4,1)** $(4,1,4)$ $(1,4,4)$	**(5,2,2)** $(2,5,2)$ $(2,2,5)$	**(7,2,0)** **(1,8,0)** **(1,2,6)** **(4,5,0)** **(4,2,3)** **(1,5,3)**	$(8,0,1)$ $(2,6,1)$ $(2,0,7)$ $(5,3,1)$ $(5,0,4)$ $(2,3,4)$	$(6,2,1)$ $(0,7,2)$ $(0,1,8)$ $(3,4,2)$ $(3,1,5)$ $(0,4,5)$

Cuadro 3.1: Construcción de las 37 secuencias de 3 bolas y periodo 3. Salvo cambios cíclicos, se pueden resumir en 13 (en negrita)

En la Tabla 3.1, podemos observar un ejemplo de la construcción de todas las secuencias de 3 bolas y periodo 3. Intentemos ahora conocer el número de secuencias malabares que se pueden dar con un número de bolas b y periodo p. Para ello, supongamos que en un momento dado asignamos a cada bola un valor entero j tal que $1 \leq j \leq b$, que trataremos de estado de la bola. En el momento en que una bola con estado j_0 es recibida/lanzada, cambiamos el estado de esa bola a estado 1, mientras que aumentamos en una unidad el estado del resto de de bolas con estado $1 \leq j < j_0$. De esta forma, cada bola mantiene un estado diferente al resto en todo momento y el estado de una bola es mayor cuanto más tiempo se haya mantenido en el aire (proviene por lo tanto de un alto lanzamiento). Si anotamos el estado j_0 de cada bola que es recibida, podemos crear un nuevo patrón para definir los malabares, en función de que bola se recoge en cada momento. Por ejemplo, en la Figura 3.3 podemos observar que la secuencia 3 recoge siempre la bola de su estado más alto, mientras que la secuencia 441 recoge las bolas de estados 3,3 y 1 respectivamente. Nótese que podría ocurrir que el último varios de los últimos

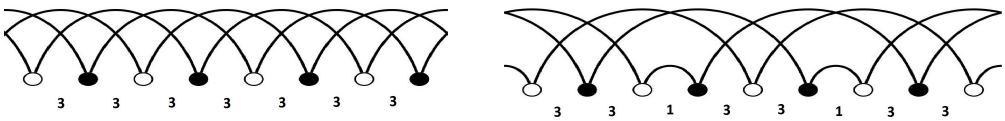

Figura 3.3: Secuencias 3 y 441 respecto al estado de las bolas

estados nunca fueran mencionados. En ese caso se trata de que hay una (o más bolas) que no son usadas, por lo que el juego se está realizando con menos bolas. Por lo tanto, esta notación

nos puede describir cualquier secuencia de periodo p y máximo b bolas. En cada momento, tenemos $b+1$ opciones sobre que bola recoger (desde los b estados o no recoger ninguna, lo que supone realizar un lanzamiento de altura $k = 0$). Esto nos lleva al siguiente resultado:

Teorema 3.1.3. *El número de secuencias malabares de periodo p y número de bolas menor o igual que b es*

$$S^{\leq}(p, b) = (b + 1)^p$$

Así mismo, el número de secuencias malabares de periodo p y número de bolas b es

$$S(p, b) = S^{\leq}(p, b) - S^{\leq}(p, b - 1) = (b + 1)^p - b^p$$

Como nota adicional, al no haberse realizado ninguna restricción, ha de tenerse en cuenta que este número supone la totalidad de las secuencias malabares, a pesar de que en algunos casos el malabar pueda ser el mismo. Es este caso en que se han contado de manera diferente secuencias que difieren únicamente en un cambio cíclico, que suponen el mismo malabar.

Estados Malabares y Grafos de Estados

Hasta ahora hemos descrito las secuencias malabares y varias de sus propiedades. No obstante, estas secuencias no son más que la repetición de un patrón indefinidamente, introduzcamos a continuación que opciones tenemos para pasar de unas secuencias de otras, no limitarnos a una simple repetición, si no "hacer trucos".

Supongamos que nos encontramos en el caso de tres bolas y comenzamos ejecutando tres lanzamientos (alternadamente) de altura 4, por lo que realizaríamos una secuencia de 3 bolas 4440. En el instante en el lanzamos la tercera bola, tenemos que las bolas van a ser recibidas en los instantes posteriores 2, 3 y 4 respectivamente. Si damos un valor 1 al momento en que una bola es recibida; y valor 0 si no, nos encontramos en un estado

$$011100\ldots$$

Puesto que ninguna bola se puede recibir, hemos de realizar un lanzamiento de altura nula, el siguiente estado es entonces

$$11100\ldots$$

En este estado, hemos de lanzar la bola que es recibida. No obstante, no podemos hacer ningún lanzamiento de altura 1 o 2, pues esos momentos posteriores ya están ocupados, entonces debemos realizar un lanzamiento mayor o igual que 3. Supongamos que es de altura 4, el nuevo estado es entonces

$$110100\ldots$$

y para la siguiente bola solo tenemos disponibles alturas 2 o mayor que 3. Esta combinación de ceros y unos se conocen como estados malabares. En general, se puede considerar que la longitud del estado es infinita añadiendo ceros una vez que todas las bolas han sido cubiertas. No obstante, por simplicidad vamos a suponer que la altura máxima está acotada.

Definición 3.1.2. *Un* estado malabar de b bolas y altura máxima k, *con* $0 \leq b \leq k$ *y* $k \geq 1$, *es una secuencia de b unos y $k - b$ ceros de forma que las posiciones $0 \leq i \leq k - 1$ que contienen la unidad representan el instante en que una bola va a ser recibida.*

Es directo observar que en este caso tenemos un total de $\binom{k}{b}$ estados. Desde cualquier estado, tenemos una serie de alturas permitidas para el siguiente lanzamiento, de forma que alturas diferentes conducen a estados diferentes. Si representamos los estados como vértices de un grafo y las alturas en aristas, podemos obtener un grafo orientado que conocemos como el *Grafo de Estados*. En la figura 3.4 podemos observar el grafo de estados para tres bolas y altura 5. Como se puede observar, hay dos casos principales de estados:

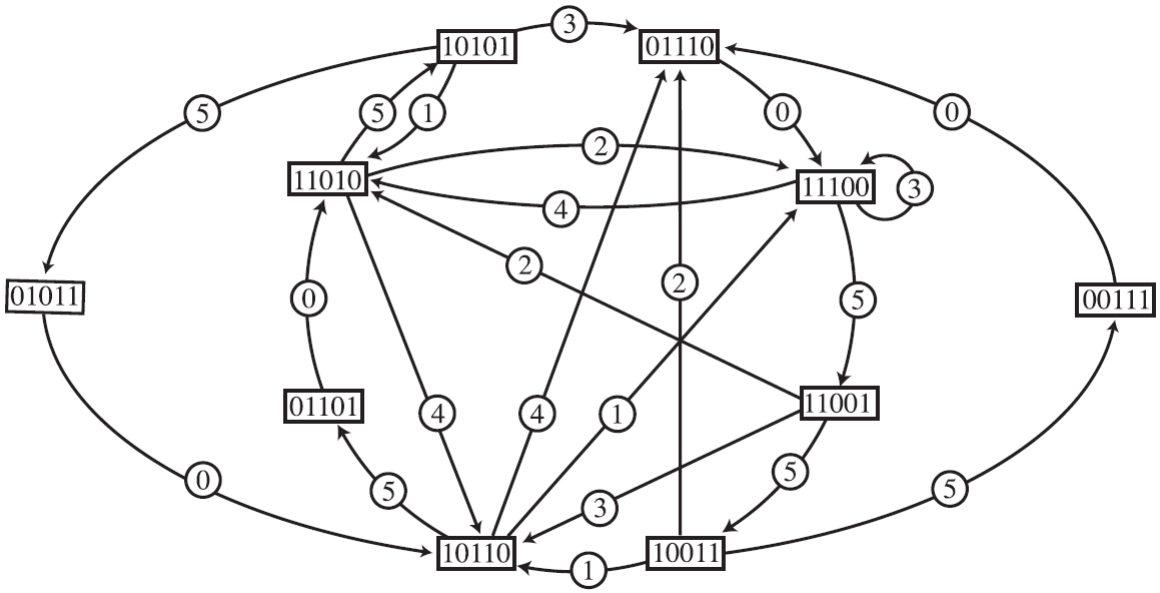

Figura 3.4: Grafo de estados para 3 bolas y altura 5

1. Si la primera entrada de la secuencia es nula , entonces solo se puede realizar un lanzamiento de altura 0. En ese caso, se conduce a un nuevo estado formado eliminando la primera entrada del estado anterior y añadiendo un 0 al final de la secuencia (un desplazamiento a hacia la izquierda). En este caso, sólo se puede ir a otro estado, luego hay una única arista orientada que se origina este vértice. Además , un simple ejercicio de la combinatoria nos permite adivinar que hay $\binom{k-1}{b}$ estados de este tipo.

2. Si la primera entrada de la secuencia es 1, entonces se puede realizar un lanzamiento de altura máxima k o una altura menor $0 < h < k$ si la posición $h+1$ del estado es nula. En el primero, se dirige al estado que se que se produce si se desplazan todos los valores una posición a la izquierda, pasando el primer 1 de la primera a la última posición. En el segundo, el nuevo estado se forma desplazando los valores a la izquierda, eliminando el primero, añadiendo un 0 al final y cambiando el cero de la nueva posición h por un 1. Hay por lo tanto $h - b + 1$ aristas originadas en cada uno de estos vértices. Asimismo, el número de estos estados en $\binom{k-1}{b-1}$.

- *Nota:* De manera análoga podemos deducir que hay un total de $b+1$ aristas que se dirigen a cada estado terminado en 0, mientras que una única arista se dirige a cada estado terminado en 1 (solo se puede concluir en este estado con un lanzamiento de altura máxima k).

Dentro de un grafo de estados, si realizamos un ciclo cerrado sobre este, un malabarista podría realizar los lanzamientos descritos por las aristas del ciclo de manera periódica, en tanto que seguiría el ciclo cerrado infinitas veces. Por lo tanto, todas las secuencias de b bolas y altura máxima k, con $0 \leq b \leq k$ y $k \geq 1$ se corresponden con los ciclos cerrados orientados del grafo de estados, de forma que empiezan y terminan en el mismo vértice y contienen al menos un vértice y una arista.

A partir de un grafo de b bolas y altura máxima k podemos construir su *complementario*, que sería el grafo de $k - b$ bolas y altura k. La construcción de este grafo se realiza cambiando el sentido de todas las aristas y su altura h por $k - h$; mientras que en cada estado cada 0 se cambia por 1 y viceversa. Si tenemos una secuencia $s = \{a_i\}_{i=0}^{p-1}$ del grafo original, se corresponde con la secuencia $c = \{h - a_{p-1-k}\}_{i=0}^{p-1}$ que se encuentra en el grafo complementario.

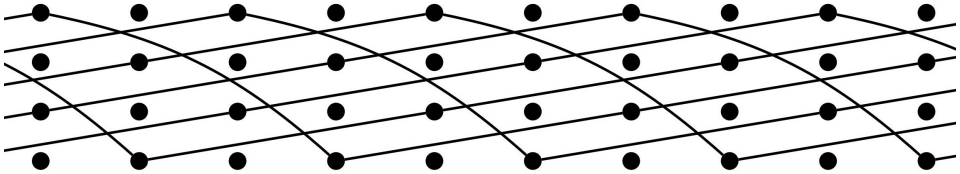

Figura 3.5: Diagrama de malabares con 4 manos

3.2. Múltiples manos

En esta sección vamos a introducir algunas de las principales ideas sobre los malabares en el momento en que dejamos de operar con malabares simples. En este caso, veamos que ocurre cuando el número de manos no se limita a ser simplemente dos, sino que puede que ser cualquier número natural h. Para ello, numeramos las manos desde 0 hasta $h-1$, y representamos el malabar como un diagrama de h filas, donde cada fila representa una mano diferente, mientras que en horizontal se representa el paso del tiempo. En la figura 3.5 vemos un ejemplo con cuatro manos y 6 bolas, donde dos malabaristas intercambiarían una bola siempre con la misma mano. Las propiedades descritas para los diagramas de malabares simples se mantienen en este caso. Para representar estos diagramas, abandonamos el concepto de secuencia malabar para introducir el de matriz malabar:

Definición 3.2.1. *Una* matriz malabar *M de h manos y periodo p es una matriz de dimensión $h \times p$ de números enteros no negativos con subíndices enteros l, $0 \leq l \leq h-1$. Cada entrada la entrada (i,j) de la matriz es representa la altura del lanzamiento en el instante $j-1$ del periodo de la bola que se encuentra en la mano $i-1$ (entendiendo una altura k de manera que la bola se recibe en el instante $k-1$ posterior). Además, el subíndice de cada entrada indica la mano hacía la que es lanzada dicha bola.*

Como ejemplo, la matriz malabar de nuestro diagrama anterior, de 4 manos, 6 bolas y periodo 2 es:

$$\begin{pmatrix} 3_3 & 0 \\ 0 & 3_0 \\ 3_1 & 0 \\ 0 & 3_2 \end{pmatrix}$$

Un importante de las matrices malabares es que se permite realizar más de un lanzamiento en el mismo instante (uno por cada mano sin recaer en malabares múltiples). En el caso de dos manos $h=2$, la representación en matrices permite representar más patrones que los permitidos por la secuencias malabares.

Con ciertas modificaciones, los principales resultados para malabares simples se mantienen. Aunque no daremos una demostración rigurosa, conviene enunciarlos:

Teorema 3.2.1. Teorema de la Media
El número de bolas necesarias para realizar malabares de un patrón correspondiente a una matriz malabar es la suma de todas las entradas de la matriz, ignorando subíndices, dividida entre el periodo.

Teorema 3.2.2. Test de Permutación
Sea M una matriz $h \times p$ cuyas entradas son de la forma de matriz malabar. Sea M' la matriz generada intercambiando las entradas no nulas m de cada columna j por el entero $m+j-1$ mód *p, manteniendo el subíndice. M es una matriz malabar sí y sólo sí por cada entrada (i,j) no nula de M, hay una entrada con el número $(j-1)_{i-1}$ en M'.*

Demostración. Sin detallarnos excesivamente en la demostración, si conviene resaltar que, al igual que la permutación de las secuencias malabares, la matriz M' es una comprobación de que, dentro del periodo, a cada bola lanzada en algún momento y mano es necesariamente recibida desde algún otro momento y/o mano, condición indispensable para que se de el malabar. □

En nuestro ejemplo, es directo que se está trabajando con 6 bolas. Apliquemos el test de permutación para dar una idea más completa a este. La matriz M' para la matriz anterior es

$$\begin{pmatrix} 1_3 & \\ & 0_0 \\ 1_1 & \\ & 0_2 \end{pmatrix}$$

Efectivamente, las entradas $1_3, 0_0, 1_1$ y 0_2 se corresponden a las entradas no nulas $(4,2), (1,1), (2,2)$ y $(3,1)$ de la matriz respectivamente.

Una vez descrito el caso de múltiples manos, tratemos finalmente con las aportaciones de Shannon:

3.3. Malabares uniformes y Teoremas de Shannon

A la hora de trabajar con los malabares, uno de los principales objetivos de Shannon era la creación de un robot que pudiera realizar los malabares por sí sólo. Ciertamente lo consiguió en uno de los casos, un robot que representaba un payaso capaz de hacer malabares de rebote (lanzar las pelotas contra el suelo en vez de hacia arriba representa, a todos los efectos de lo expuesto en este capítulo, los mismos patrones). Para ello, definió un tipo de malabares similar al simple, pero al que incluyó varias definiciones diferentes de los tiempos periódicos: los malabares uniformes.

Definición 3.3.1. *Se dice que un patrón malabar es* uniforme *si no es múltiple (hay un máximo de una bola en cada mano) y los tiempos de permanencia D de una bola en la mano; de vuelo F del objeto; y de vacante V en el que la mano permanece vacía son constantes durante todo el proceso, realizándose además los mismos movimientos constantemente.*

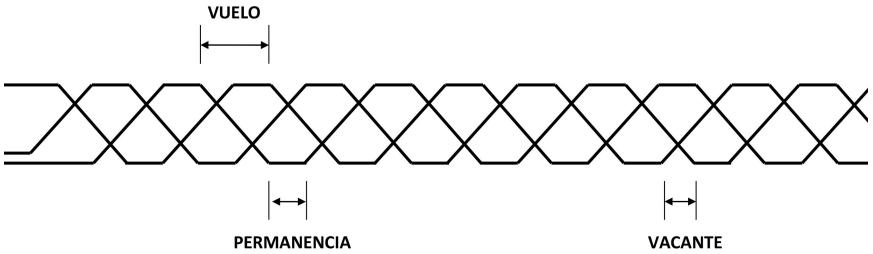

Figura 3.6: Diagrama de malabares con 4 manos

En la Figura 3.6 podemos ver un diagrama de la secuencia 3 representada con estas nuevas notaciones.

En el caso de una matriz malabar, está describe un malabar uniforme si

- Cada entrada de la matriz es o bien nula o igual a un entero positivo fijo k (Todos los lanzamientos han de ser de la misma altura pues el tiempo de vuelo es constante).

- Existe una constante c tal que el número de ceros entre cualquier par de enteros k consecutivos en una misma fila es igual c. Para completar el periodo, el número de ceros antes de la primera entrada no nula de una fila y después de su última entrada no nula debe ser también c.

Por ejemplo, la matriz que representamos en la sección de múltiples es uniforme.

En este tipo de malabares y cualquier resultado derivado para ellos hay una dualidad entre las bolas y las manos, de forma cualquier resultado es cierto si intercambiamos "bolas" por "manos"; y en consecuencia el tiempo de vuelo por el de vacante. La propiedad de un tiempo de permanencia constante es auto-dual, pues se aplica de la misma manera según se trate de manos o bolas.

Antes de comenzar con los teoremas, conviene destacar que todos ellos se aplican si el tiempo es superior al que requiere una bola para recorrer todas las manos, es decir, supondremos un tiempo superior a $h(D+F)$.

Primer Teorema de Shannon

Teorema 3.3.1. *En malabares uniformes con tiempo de permanencia D, tiempo de vacante V y tiempo de vuelo F, se cumple que*

$$\frac{F+D}{V+D} = \frac{b}{h} \tag{3.2}$$

siendo b el número de bolas y h el número de manos.

Demostración. El teorema nos indica que hay una proporcionalidad entre el número de bolas y manos, y el tiempo total del circuito de una bola $F+D$ y de una mano $V+D$, algo que parece bastante lógico. Veamos su demostración: tomemos una bola en concreto y sigamos su recorrido hasta el momento en que ha sido recogida h veces. Por el Principio del Palomar, necesariamente ha pasado al menos dos veces por alguna de las manos. Si nos centramos en esa mano, entre la primera y segunda vez que esa bola pasa por esa mano ha habido otros m lanzamientos de esta bola, por lo que el tiempo es $(m+1)(F+D)$. Por otro lado, la mano seleccionada ha debido hacer un número n de lanzamientos, luego el tiempo es de $(n+1)(V+D)$ y consecuentemente

$$(m+1)(F+D) = (n+1)(V+D)$$

y podemos reducir esta igualdad de eliminando divisores comunes de $(m+1)$ y $(n+1)$ de manera que

$$\frac{F+D}{V+D} = \frac{p}{q} \tag{3.3}$$

con p y q primos entre sí.
Si tomamos ahora un periodo de tiempo $q(F+D)$ tal que ninguna bola es capturada en el instante inicial, sino que las recepciones se producen en tiempos $t_1, t_2, t_3 \dots$; y denominemos s_i el número de recepciones que se producen en el instante t_i (s_i manos recogen simultáneamente s_i bolas). Entonces cada bola ha sido recogida q veces exactamente en ese periodo, mientras que cada mano ha de haber recibido bolas p veces. En total,

$$\sum s_i = qb = ph \tag{3.4}$$

Y $p = q\dfrac{b}{h}$. Combinando esta ecuación con (3.3), llegamos al resultado deseado del teorema. □

Este teorema nos permite también calcular el rango de posibles periodos o frecuencia $V+D$ (tiempo entre lanzamientos de una mano) para un malabar uniforme dado y un tiempo de vuelo F fijo. Con $b > h$ tenemos que

$$V + D = \frac{(F+D)h}{b} = \frac{(F-V)h}{b-h}$$

por lo que el mínimo rango se alcanza si $D = 0$ y el máximo si $V = 0$. Esto nos permite concluir de la siguiente forma:

Corolario 3.3.1. *En un malabar uniforme con tiempo de vuelo fijado, el rango de posibles periodos es*

$$\frac{b}{b-h} \tag{3.5}$$

Segundo Teorema de Shannon

Teorema 3.3.2. *Si b y h son primos entre sí, entonces hay esencialmente una única manera (salvo por diferente etiquetado) de realizar un malabar uniforme de b bolas y h manos. Las bolas pueden ser numeradas desde 0 a $b-1$ y las manos desde 0 a $h-1$ de tal forma que cada bola progresa a través de las manos en una secuencia cíclica y cada mano recibe cada bola de una manera cíclica.*

Tercer Teorema de Shannon

Teorema 3.3.3. *Si b y h no son primos entre sí, sea n el máximo común divisor de b y h. Entonces hay tantas maneras diferentes de malabares uniformes con b bolas y h manos como maneras hay de representar n como suma de enteros positivos.*

Demostración. Hemos omitido ninguna demostración en el Segundo Teorema pues la demostración es exactamente la mismo. En concreto, se puede entender el Segundo Teorema como un caso particular del Tercero en el que $n = 1$, por lo que hay una única manera de realizar los malabares. Veamos la demostración.

Si fijamos un tiempo $t = t_0$, denotamos como n_0 el número de bolas que se lanzan en ese instante. Por los tiempos constantes del malabar uniforme, entonces este grupo de bolas se recibirán y lanzarán siempre en los mismos instantes, por lo que se encontrarán en una mano (necesariamente diferentes) o en el aire al mismo tiempo. Además, ninguna otra bola adicional puede actuar como ellas, es decir, ninguna otra puede ser recogida/lanzada en esos momentos. Por otro lado, las n_0 manos donde se encontraban las bolas en t_0 actúan de manera análoga, recibiendo bolas simultáneamente y de la igualdad (3.3) se deduce que las bolas volverán a las mismas manos por primera vez tras realizar q lanzamientos (con p y q definidos por la ecuación 3.3, respecto de D, V y F).

Precisamente, este conjunto de manos recibirán otras n_0 bolas en el instante $t_0 + D + V$, y este segundo conjunto de bolas recorrerá el mismo circuito (las mismas manos) que el primer conjunto en el mismo orden. Continuando con este razonamiento, hay un número p de grupos de n_0 bolas siguiéndose unas a otras de manera cíclica, pasando por los mismos q grupos de n_0 manos. Entonces hay pn_0 bolas manipuladas por qn_0 manos y este sistema es independiente por la uniformidad, de manera que ninguna otra bola o mano exterior puede interactuar con ninguna de las manos o bolas del sistema.

Si $b > qv_0$, hay más bolas aún en juego, entonces podemos repetir el mismo proceso en un

instante diferente con n_1 bolas. Habiendo pn_1 bolas en este nuevo sistema. Se repite este proceso hasta cubrir todas las bolas (finitas) en un número k de pasos. Entonces

$$b = p\sum_{i=0}^{k-1} n_i \qquad \text{y} \qquad h = p\sum_{i=0}^{k-1} n_i \tag{3.6}$$

Dado que p y q son primos entre sí, entonces $n = \sum_{i=0}^{k-1} n_i$ es el máximo común divisor de b y h.

En el caso del Segundo Teorema, $n = 1$, por lo que $v_0 = 1$, $p = b$, $h = q$ y hay un único sistema independiente , la única de manera hacer el malabar, según la forma descrita (salvo por diferentes etiquetados que se les dé a las bolas y manos).
En el caso del Tercer Teorema, podemos crear tantos sistemas independientes como maneras diferentes hay para representar la suma $\sum_{i=0}^{k-1} n_i$. □

Un ejemplo del Tercer Teorema lo encontramos en nuestro ejemplo de 2 malabaristas (4 manos) y 6 bolas. El máximo común divisor de 4 y 6 es 2, por lo que hay dos formas realizar los malabares, pues 2 se puede descomponer en las sumas 2 o $1 + 1$. Podemos interpretar como el primer caso la situación en la que los dos malabaristas comienzan simultáneamente: entonces los malabaristas tienen la opción de realizarse lanzamientos a sí mismos o al compañero, las cuatro manos forman un único sistema. En el caso $1 + 1$, los dos malabaristas no están sincronizados, por lo que cada uno de ellos es un sistema completamente independiente.

Teoremas de Shannon para secuencias malabares

Si introducimos ciertos conceptos, podemos relacionar los Teoremas de Shannon con algunas secuencias malabares que hemos estudiado previamente, sin necesidad de que estas sean uniformes. Si empezamos con una secuencia malabar simple $s = \{a_i\}_{i=0}^{p-1}$, podemos introducir más manos así como tiempos de vuelo, permanencia y vacante de manera que

- Definimos los instantes como los momentos en los que alguna bola es atrapada.
- Se usan h manos, de forma que h divide al periodo p. Podemos numerar estas manos de 0 a $h - 1$, de forma que cada lanzamiento corresponde a una mano que se van turnando cíclicamente (es decir, hemos representado la matriz como una secuencia).
- Para un lanzamiento a_i, los tiempos de permanencia y vuelo son respectivamente $a_i\omega_i$ y $a_i(1 - \omega_i)$, con $0 \leq \omega_i \leq 1$ para que se cumpla la primera propiedad que hemos introducido. Si no hay ningún lanzamiento de altura 0, una mano recibe una bola cada h lanzamientos. Para evitar que más de una bola caigan a la vez en la misma mano, se exige que $a_i\omega_i < h, \quad i = 0, 1, \ldots, p - 1$.

Durante un periodo, el tiempo total de vuelo de todas las bolas f_T será la suma del tiempo de vuelo de cada lanzamiento:

$$f_T = \sum_{i=0}^{p-1} a_i(1 - \omega_i)$$

pues el tiempo de vuelo no incluido de los últimos lanzamientos (cuyos vuelos se salen del periodo) se complementa con el tiempo de vuelo de las bolas que están el aire al inicio del periodo, siguiendo la estructura cíclica. De manera análoga, el tiempo de permanencia de todas las bolas durante ese periodo es

$$d_T = \sum_{i=0}^{p-1} a_i\omega_i$$

Por último, el tiempo de vacante ha de ser resto del tiempo del periodo en el que una las manos no están ocupadas:

$$v_T = hp - d_T$$

Combinando todas estas expresiones obtenemos que

$$\frac{f_T + d_T}{v_T + d_T} = \frac{\sum_{i=0}^{p-1} a_i(1-\omega_i) + \sum_{i=0}^{p-1} a_i\omega_i}{hp - d_T + d_T} = \frac{\sum_{i=0}^{p-1} a_i}{hp} = \frac{b}{h}$$

Donde la última igualdad se obtiene aplicando el Teorema de la Media. Concluimos entonces

Teorema 3.3.4. Primer Teorema de Shannon para Secuencias Malabares
Sea una secuencia malabar de b bolas y h manos, con tiempos fijados de permanencia, vuelo y vacante, con las especificaciones hechas previamente; de forma que la suma total sobre un periodo de estos tiempos es respectivamente d_T, f_T y v_T. Entonces

$$\frac{f_T + d_T}{v_T + d_T} = \frac{b}{h} \tag{3.7}$$

Bibliografía

[Buh] J. Buhler; D.Eisenbud; R.Graham; C.Wright. *Juggling Drops and Descents*. Amer. Math. Monthly 101, 1994.

[Jus] J. Justesen y T. Høholdt. *A Course in Error-Correcting Codes*. European Mathematic Society, 2004.

[Knu] A.Knutson. *Mathematics of Juggling*. Cornell University. Recuperado de https://www.youtube.com/watch?v=38rf9FLhl-8. Publicado 29/04/2010.

[Pel] R. Pellikaan; X. Wu; S. Bulygin; R. Jurrius. *Error-Correcting Codes and Cryptology*. Versión Preliminar. Cambridge University Press. 2012.

[Pol] B. Polster. *The Mathematics of Juggling*. Springer-Verlag, New York, 2003.

[Sha] C. E. Shannon. *A Mathematical Theory of Communication*. The Bell System Technical Journal, vol. 27, pp. 379-423, 623-656. 1948.

[Sha] C. E. Shannon. *Communication in the Presence of Noise*. Proc. Institute of Radio Engineers, vol. 37, no. 1, pp. 10-21. 1949.

[Sha] C. E. Shannon. *Scientific Aspects of Juggling* en *Claude Elwood Shannon. Collected Papers*. Editado por N. Sloane y A. Wyner. IEEE Press, New York, 1993.

[UCTV] University of California Television. *Claude Shannon - Father of the Information Age*. [1/2002][Science][Show ID: 6090]. Recuperado de https://www.youtube.com/watch?v=z2Whj_nL-x8. Publicado 16/01/2008.