Capítulo 5

Límites fundamentales en los sistemas de comunicaciones digitales

El principal objetivo de un sistema de comunicaciones es la transmisión fiable de información. Una fuente produce la información, y el propósito del sistema de comunicaciones es transmitir la salida de la fuente al destino de la información, como se ilustra en la Figura 5.1.

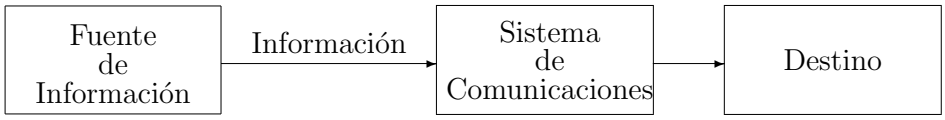

Figura 5.1: Esquema funcional simplificado de un sistema de comunicaciones.

Existe una gran variedad de fuentes de información, y cada una de ellas produce información de naturaleza distinta. Algunos tipos de fuentes de información, con el correspondiente tipo de información que generan, podrían ser, por ejemplo

- Radiodifusión de radio: fuente de voz o de audio.
- Radiodifusión de TV: fuente de vídeo.
- Transmisión de FAX: imagen fija.
- Comunicación entre PC's: fuente binaria o ASCII.
- Almacenamiento de datos: fuente binaria.

A la hora de analizar un sistema de comunicaciones y medir sus prestaciones o sus límites, es posible plantearse cuantificar la cantidad de información transmitida, o la fiabilidad en la transmisión de información. En los capítulos anteriores se han estudiado dos tipos de sistemas de comunicaciones: sistemas de comunicaciones analógicos y sistemas de comunicaciones digitales. En cuanto a sistemas de comunicaciones analógicos, ya se ha estudiado el compromiso entre prestaciones y consumo de recursos de las distintas variantes de modulaciones, especialmente potencia y ancho de banda. En este capítulo se considerarán únicamente los sistemas de comunicaciones digitales, y el objetivo será estudiar los límites fundamentales que se pueden alcanzar en la transmisión fiable

de información con este tipo de sistemas. Conviene recordar que la utilización de un sistema de comunicaciones digital no implica la exclusión de fuentes analógicas. La transmisión digital permite en general una mayor inmunidad frente a ruido, mayor flexibilidad, la aplicación de encriptado y facilita la implementación de los equipos, lo que ha hecho que los sistemas digitales predominen sobre los analógicos. Pero es posible la tranmisión de información de naturaleza analógica a través de un sistema digital, realizando una conversión analógico/digital en el lado del transmisor y la correspondiente conversión digital/analógico en el lado del receptor.

Al estudiar las prestaciones de sistemas de comunicaciones digitales en el capítulo anterior, se ha visto que al aumentar la velocidad binaria de transmisión mediante el aumento del número de bits por símbolo de la constelación, para una cierta energía media por símbolo fija, la probabilidad de error aumenta y por tanto las prestaciones se degradan. Este hecho se consideraba algo ineludible hasta que en los años cuarenta Claude Shannon demostró que es posible transmitir con una probabilidad de error tan baja como se desee a una velocidad binaria arbitraria siempre que esa velocidad esté por debajo de la denominada *capacidad del canal*. Esta demostración se considera el principio de la llamada *teoría de la información*, y establece un límite a los sistemas de comunicaciones en cuanto a la máxima cantidad de información que se puede transmitir de forma fiable. Es preciso hacer notar que esta demostración establece el límite, pero no especifica cómo se puede alcanzar dicho límite. De momento, la teoría de la información no tiene respuesta a esta pregunta.

En este capítulo se pretende analizar cuál es la máxima cantidad de información que se puede transmitir de forma fiable utilizando un sistema digital de comunicaciones, y de qué factores depende el valor de ese límite. Aunque casi todo el mundo tiene una noción intuitiva de lo que es información, para realizar el análisis de un sistema de comunicaciones no basta con esa noción intuitiva, sino que es necesario disponer de medidas cuantitativas de información. Hartley, Nyquist o Shannon fueron pioneros en el desarrollo de definiciones de medidas para la información que resultan de utilidad para el propósito de este capítulo. Estas medidas de información están relacionadas con las distribuciones de probabilidad de los elementos cuya información se trata de cuantificar. Para poder hacer un uso de ellas en el análisis de un sistema digital de comunicaciones es preciso definir en primer lugar modelos probabilísticos que puedan utilizarse para representar el comportamiento de fuentes de información digitales. Dado que la información se transmitirá a través de un canal haciendo uso del sistema de comunicaciones, será preciso también disponer de modelos probabilísticos que permitan representar el comportamiento del canal y el comportamiento del sistema de comunicaciones a distintos niveles. A partir de esos modelos probabilísticos, y utilizando medidas cuantitativas de información, se podrán finalmente obtener los límites que puede alcanzar un cierto sistema de comunicaciones.

Este capítulo se dividirá en cuatro partes. En primer lugar se presentarán modelos probabilísticos adecuados para representar el comportamiento de fuentes de información. A continuación se definirán modelos probabilísticos que permitan representar el comportamiento del canal de comunicaciones y de los elementos del sistema de comunicaciones a distintos niveles. Genéricamente se denominarán todos estos modelos como modelos de canal, ya que los distintos elementos del sistema de comunicaciones pueden en cierto modo considerarse como parte del canal a través del que se transmite la información. La tercera parte presentará distintas medidas cuantitativas de información. Finalmente, en la última parte, haciendo uso de los modelos probabilísticos de fuente y de canal, y de las medidas cuantitativas de información presentadas en la tercera parte, se obtendrán los límites fundamentales alcanzables por un sistema digital de comunicaciones.

5.1. Modelado de las fuentes de información

Una fuente de información produce como salida algo, que se denomirá genéricamente información, que es de interés para el receptor de la misma, que no la conoce de antemano. La misión del sistema de comunicaciones es asegurar que la información se transmite correctamente.

Como la salida de la fuente de información es una función variante en el tiempo e impredecible (si es predecible no tiene demasiado interés transmitirla), esta salida puede ser modelada mediante un proceso aleatorio. Las características de ese proceso aleatorio dependerán de las características y naturaleza de la fuente en cuestión. Por ejemplo, podrá ser un proceso aleatorio en tiempo continuo o en tiempo discreto dependiendo de si la fuente genera una información de naturaleza analógica o digital. Aunque este capítulo se centrará en el estudio de sistemas digitales, y los modelos más relevantes serán por tanto los modelos para fuentes digitales (las fuentes analógicas serán convertidas a formato digital antes de su transmisión), por completitud se estudirán también de forma breve los modelos utilizados para fuentes analógicas.

5.1.1. Fuentes analógicas

El modelo utilizado para caracterizar una fuente analógica será habitualmente un proceso aleatorio en tiempo continuo, $X(t)$, cuyas propiedades estadísticas dependerán de la naturaleza de la fuente. Se puede tomar como ejemplo el modelado de una fuente de voz. Es conocido que la señal de voz tiene la mayor parte de su potencia distribuida esencialmente en la banda de frecuencias entre 300 y 4000 Hz. Por tanto, esta fuente se puede modelar mediante un proceso aleatorio cuya densidad espectral de potencia se ajuste a estas características, tal y como se ilustra en la Figura 5.2.

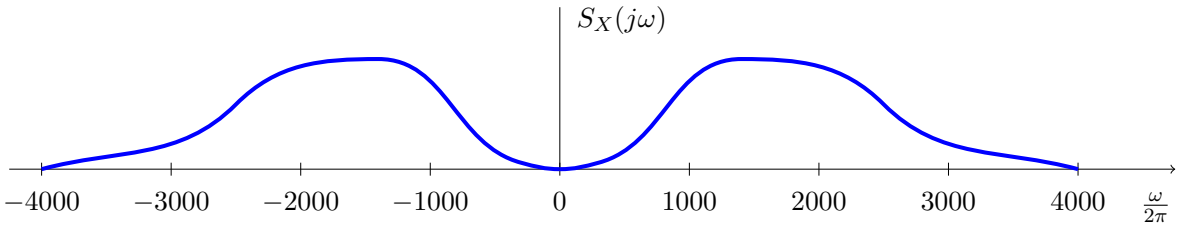

Figura 5.2: Ejemplo de una densidad espectral de potencia que representa un comportamiento típico de una señal de voz.

Habitualmente se considera que el comportamiento medio no cambia con el tiempo, lo que permite asumir que el proceso estacionario. Como además la media de la señal es nula, el proceso aleatorio utilizado para modelar la señal de voz podría ser un proceso aleatorio estacionario, de media nula, y función densidad espectral de potencia como la de la figura. Al ser un proceso estacionario, su función de autocorrelación vendrá dada por la transformada de Fourier inversa de esta densidad espectral de potencia.

El mismo procedimiento es aplicable a distintas fuentes analógicas. Por ejemplo para señales de TV, dependiendo del sistema (PAL, SECAM o NTSC), la banda se encuentra entre 0-6.5 MHz o entre 0-4.5 MHz. Se utilizaría para modelarlas un proceso aleatorio estacionario cuya densidad espectral de potencia representa el comportamiento medio de la respuesta en frecuencia de la señal al cuadrado.

Aunque cada señal analógica tendrá unas características espectrales diferentes, hay algunos aspectos comunes en la mayoría de los modelos utilizados:

1. Se consideran procesos limitados en banda.
2. Esto permite que se puedan muestrear siguiendo el criterio de Nyquist y posteriormente pueden volver a reconstruirse.

5.1.2. Fuentes digitales

En el caso de una fuente digital, su salida se puede modelar mediante un proceso aleatorio discreto en el tiempo. En ese caso, el modelo matemático para una fuente de información es el que muestra la Figura 5.3.

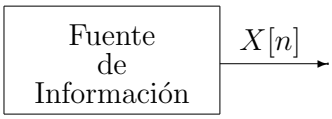

Figura 5.3: Modelo matemático de una fuente de información discreta en el tiempo

La fuente se modela como un proceso aleatorio discreto en el tiempo, $X[n]$. El alfabeto de la fuente puede ser:

- Discreto. Por ejemplo para modelar una fuente de datos digitales o señales analógicas muestreadas y cuantificadas.
- Continuo. Por ejemplo para representar fuentes analógicas muestreadas (como una señal de voz) antes de la cuantificación.

En función del tipo de fuente los parámetros estadísticos del proceso aleatorio serán diferentes. En esta sección se va a estudiar el modelo más sencillo de fuente, que permite realizar todo el desarrollo posterior del capítulo. Este modelo es el *modelo discreto sin memoria*.

Modelo discreto sin memoria

El modelo discreto sin memoria, o DMS (del inglés *discrete memoryless source*), de una fuente de información es un proceso aleatorio discreto en el tiempo y en amplitud, en el que todas las variables aleatorias que forman el proceso aleatorio $X[n]$ son independientes y tienen la misma distribución. Por tanto, una fuente DMS genera una secuencia de variables aleatorias i.i.d. que toman valores de un alfabeto discreto. Para describir de forma completa este tipo de fuente basta con conocer su alfabeto y su distribución, es decir

1. $\mathcal{A}_X = \{x_0, x_1, \cdots, x_{M-1}\}$.
2. $p_X(x_i) = P(X = x_i)$ para $i = 0, 1, \cdots, M-1$.

Ejemplo

Una fuente de información binaria utiliza como modelo un modelo discreto sin memoria que viene descrito por el alfabeto $\mathcal{A}_X = \{0, 1\}$ y las probabilidades de cada símbolo binario, $P(X = 1) = p$, y $P(X = 0) = 1 - p$. En el caso particular en que $p = 1/2$, este tipo de fuente se denomina *fuente binaria simétrica* o BSS (del inglés *binary symmetric source*).

5.2. Modelos probabilísticos de canal

En el Capítulo 4 se ha presentado el modelo general de un sistema de comunicaciones digitales, cuyos elementos funcionales se muestran en la Figura 5.4.

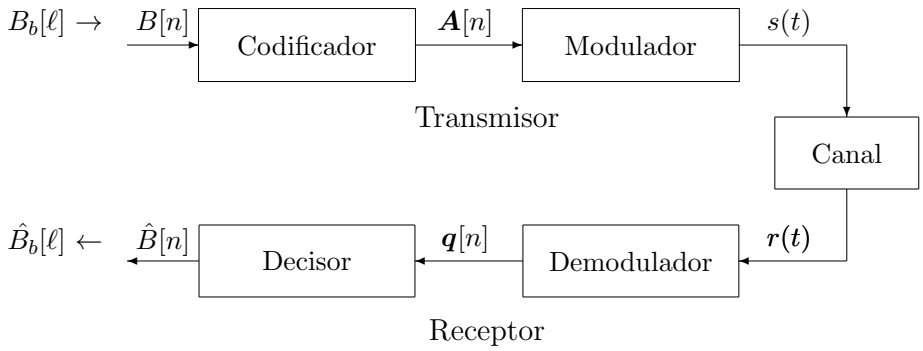

Figura 5.4: Modelo general de un sistema digital de comunicaciones.

En este esquema se denominaba canal a la abstracción del medio físico de transmisión, y se utilizaba como modelo simplificado para representarlo un modelo de canal aditivo gausiano. En esta sección el término canal tendrá un significado más amplio que el considerado en el esquema de la Figura 5.4. Se van a definir varios modelos probabilísticos, que se denominarán genéricamente modelos de canal; estos modelos establecerán la relación probabilística entre la información recibida y la información transmitida entre distintos puntos del sistema de comunicaciones, lo que puede interpretarse como la definición de varios canales sobre el sistema, cada uno de ellos en un diferente nivel de abstracción. La caracterización de estos modelos vendrá dada por la distribución condicional de la salida dada la entrada. En todos los casos se considerará independencia condicional entre entradas y salidas para instantes de tiempo diferentes, por lo que se podrá eliminar la dependencia temporal. Esto significa que el valor de entrada en un cierto instante se modelará mediante una variable aleatoria X, y el valor de la correspondiente salida en el mismo instante se modelará con otra variable aleatoria, Y. En este caso el modelo probabilístico estará caracterizado por la distribución condicional

$$f_{Y|X}(y|x).$$

La diferencia entre los distintos modelos que se van a presentar está en la definición de lo que en cada caso se considera entrada y salida para cada *canal*. Se van a presentar cuatro modelos probabilísticos (o *canales*) diferentes, cada uno considerando lo que es el canal con un distinto nivel de abstracción sobre el modelo general del sistema de comunicaciones. En concreto, los cuatro modelos son los que se muestran en la Figura 5.5, y que se enumeran a continuación:

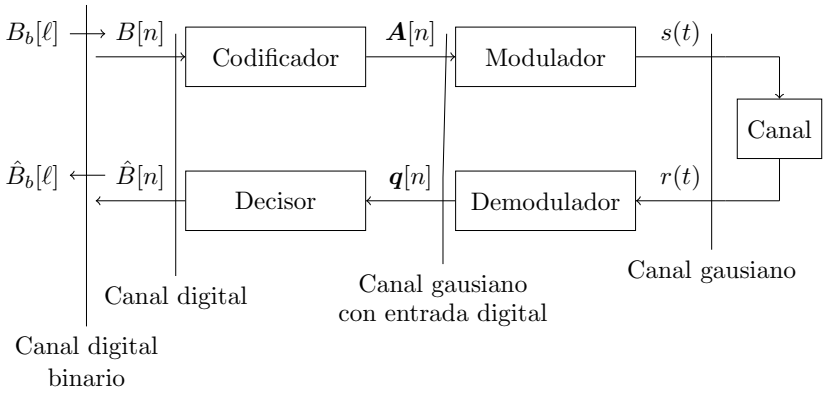

Figura 5.5: Definición de los distintos modelos de canal sobre un sistema de comunicaciones digital.

1. **Canal gausiano**. Es el modelo para el representar el canal físico propiamente dicho, que transforma la señal transmitida mediante la adición de ruido aditivo, modelado mediante un proceso aleatorio estacionario, ergódico, blanco, gausiano, de media nula y densidad espectral de potencia $N_0/2$, de forma que
$$r(t) = s(t) + n(t).$$
En este modelo, como entrada y salida se consideran los valores de la señal transmitida y la señal recibida, respectivamente, en el mismo instante temporal, es decir
$$X \equiv s(t),\ Y \equiv r(t).$$
En ambos casos, las variables aleatorias son variables aleatorias continuas, ya que las señales transmitida y recibida pueden tomar valores en un rango continuo de amplitudes.

2. **Canal gausiano con entrada digital**. Este modelo coincide con lo que en el capítulo anterior se denominó canal discreto equivalente, que presenta como entrada una secuencia de símbolos de un alfabeto discreto (constelación) de M símbolos y como salida tiene una observación con un dominio continuo, la salida del demodulador. En el capítulo anterior se vió que la relación entre ambas era una relación aditiva con el término de ruido a la salida del demodulador
$$\boldsymbol{q}[n] = \boldsymbol{A}[n] + \boldsymbol{z}[n].$$
Las características estadísticas del vector de ruido $\boldsymbol{z}[n]$ se estudiaron en el capítulo anterior. Por tanto, en este modelo la entrada es el símbolo transmitido en un intante discreto n y la salida el valor de observación a la salida del demodulador en ese mismo instante. Se trata por tanto de símbolos y observaciones vectoriales de dimensión N, por lo que para representarlas se utilizarán variables aleatorias multidimensionales (vectoriales) de la misma dimensión
$$\boldsymbol{X} \equiv \boldsymbol{A}[n],\ \boldsymbol{Y} \equiv \boldsymbol{q}[n].$$

3. **Canal digital**. Este modelo considera el conjunto formado por el codificador, modulador, canal, demodulador y el decisor de símbolos. Como entrada tiene un conjunto de símbolos de un alfabeto de M valores y como salida un alfabeto de los mismo M símbolos: $B[n]$ y $\hat{B}[n]$ (o lo que es lo mismo $\boldsymbol{A}[n]$ y $\hat{\boldsymbol{A}}[n]$, ya que hay una equivalencia unívoca entre símbolos como bloques de M bits y vectores N-dimensionales de la constelación que los transportan). Por tanto, en este modelo la entrada y la salida vendrán definidos como
$$X \equiv B[n],\ Y \equiv \hat{B}[n] \text{ ó equivalentemente } X \equiv \boldsymbol{A}[n],\ Y \equiv \hat{\boldsymbol{A}}[n].$$

4. **Canal digital binario**. Este es el último nivel de abstracción, en el que se considera como canal todo el sistema de comunicaciones, cuyas entradas y salidas son los símbolos binarios. Representa el mayor nivel de abstracción que se puede realizar sobre un sistema de comunicaciones: el sistema completo visto como vehículo o canal para la transmisión de bits. En este caso por tanto, la entrada y salida se definirán como los bits transmitido y recibido en el mismo instante de tiempo
$$X \equiv B_b[\ell],\ Y \equiv \hat{B}_b[\ell].$$

Una vez que se han presentado estos cuatro canales, a continuación se obtendrá el modelo probabilístico (distribución condicional de la salida dada la entrada) que se utilizará para representar cada uno de ellos.

5.2.1. Canal gausiano

La relación entre la entrada y la salida del canal gausiano es
$$r(t) = s(t) + n(t),$$
donde el término de ruido $n(t)$ se modela como un proceso aleatorio estacionario, ergódico, blanco, gausiano, de media nula y densidad espectral de potencia $S_n(j\omega) = N_0/2$. La función de autocorrelación del ruido es por tanto
$$R_n(\tau) = \frac{N_0}{2}\,\delta(\tau).$$
Para obtener el modelo probabilístico dado por la probabilidad condicional de la salida dada la entrada, lo primero que hay que tener en cuenta es que estrictamente la potencia del proceso de ruido, al ser blanco, es infinita. Esto implica que en la práctica para minimizar el efecto del ruido en el receptor se utilizará un filtro selectivo en frecuencia. Idealmente, este filtro no introducirá ninguna distorsión en la señal $s(t)$, que se considerará limitada en banda con un ancho de banda B Hz, y al mismo tiempo debe minimizar la potencia del ruido a su salida; el filtro que cumple estas condiciones es un filtro paso bajo ideal, con un ancho de banda igual al de la señal recibida. De este modo el canal gausiano en realidad modelará la relación entre la señal transmitida y la señal recibida después de este filtrado, tal y como se ilustra en la Figura 5.6, donde la respuesta $h_n(t)$ (o su equivalente en frecuencia $H_n(j\omega)$) es la del filtro ideal utilizado para limitar el ruido.

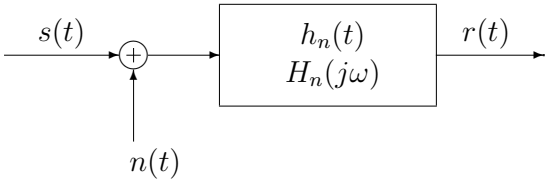

Figura 5.6: Canal gausiano con filtrado en el receptor para la limitación de la potencia de ruido.

Si la señal tiene un ancho de banda B Hz, el ancho de banda del filtro será B Hz, con lo que su respuesta en frecuencia es
$$H(j\omega) = \Pi\left(\frac{\omega}{2W}\right),$$
donde $W = 2\pi B$ denota el ancho de banda en radianes por segundo, y su respuesta al impulso
$$h(t) = 2B\ \mathrm{sinc}(2Bt).$$

La potencia del término de ruido a la salida de este filtro, como se vió en el Capítulo 2, se obtiene integrando la densidad espectral de potencia del proceso de ruido filtrado, que para fitros ideales supone multiplicar N_0 por el ancho de banda en Hz del filtro

$$\sigma^2 = \int_{-\infty}^{\infty} S_n(j\omega)|H_n(j\omega)|^2 \, d\omega = \frac{N_0}{2} \times 2B = N_0 B.$$

Teniendo esto en cuenta, el *canal probabilístico gausiano* se define como aquel que relaciona dos variables aleatorias X e Y con funciones de densidad de probabilidad continuas sobre $\mathbb{R}$ que representan el valor de $s(t)$ y $r(t)$ en un cierto instante de tiempo. Dado que el valor de la salida en un instante será el de la entrada más el valor del ruido filtrado en ese instante, y que ese ruido tiene una distribución gausiana con potencia N_0B, este modelo probabilístico está caracterizado por la función de densidad de probabilidad condicional

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}},$$

donde $\sigma^2 = N_0 B$.

5.2.2. Canal gausiano con entrada digital

La Figura 5.7 ilustra conceptualmente el modelo de canal gausiano con entrada digital, que modela la relación entre los símbolos de la constelación que se transmiten y la observación a la salida del demodulador en un momento dado.

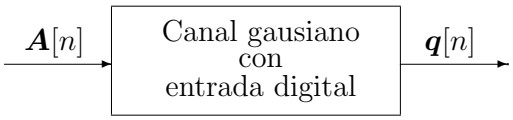

Figura 5.7: Canal gausiano con entrada digital.

Como se vió en el Capítulo 4, la salida $\boldsymbol{q}[n]$ toma la forma

$$\boldsymbol{q}[n] = \boldsymbol{A}[n] + \boldsymbol{z}[n],$$

donde $\boldsymbol{z}[n]$ es la componente de ruido de ruido de la observación, que es un proceso estocástico discreto en el tiempo, multidimensional (de la dimensión del espacio de señales del sistema, N), formado por N variables aleatorias conjuntamente gausianas, e independiente de $\boldsymbol{A}[n]$. Además, sus componentes son a su vez estadísticamente independientes las unas de las otras. Esto hace que se pueda eliminar la dependencia temporal y utilizar la representación

$$\boldsymbol{q} = \boldsymbol{A} + \boldsymbol{z},$$

donde $\boldsymbol{z}$ tiene una función densidad de probabilidad gausiana N-dimensional, de media nula y varianza $N_0/2$ en todas las direcciones del espacio

$$f_{\boldsymbol{z}}(\boldsymbol{z}) = \mathcal{N}^N\left(\boldsymbol{0}, \frac{N_0}{2}\right) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{||\boldsymbol{z}||^2}{N_0}}.$$

A partir de aquí, la obtención del modelo probabilístico es simple. Formalmente se define como aquel que relaciona las variables aleatorias N-dimensionales $\boldsymbol{X}$ e $\boldsymbol{Y}$, la primera con alfabeto

discreto $\{\boldsymbol{x}_i\}$, con $i = 0, \cdots, M-1$, donde cada valor del alfabeto estará identificado con uno de los vectores N-dimensionales que forman la constelación del sistema, y la segunda con una función densidad de probabilidad N-dimensional continua sobre $\mathbb{R}$. Bajo esa premisa, el modelo probabilístico viene dado por la función densidad de probabilidad condicional

$$f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{||\boldsymbol{y}-\boldsymbol{x}_i||^2}{N_0}},$$

es decir, una distribución gausiana N dimensional con media el símbolo transmitido y varianza $\sigma^2 = N_0/2$ en cada dirección del espacio. Como se ve, coincide con el modelo probabilístico que define el canal discreto equivalente, utilizando ahora una notación ligeramente diferente en términos de las variables aleatorias $\boldsymbol{X}$ e $\boldsymbol{Y}$ para denotar, respectivamente, $\boldsymbol{A}$ y $\boldsymbol{q}$.

5.2.3. Canal digital

La Figura 5.8 representa el canal digital. Habitualmente se trabaja con los símbolos de la constelación $\boldsymbol{A}[n]$ en lugar de con los símbolos $B[n]$, aunque dada la relación unívoca entre ambos, se puede aplicar sobre los últimos de igual modo, incluyendo del codificador y decodificador dentro de este modelo.

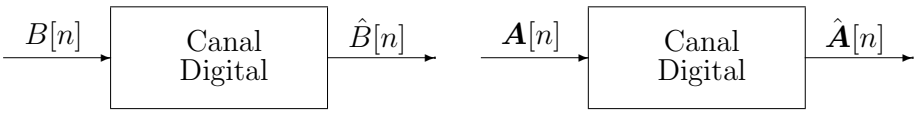

Figura 5.8: Representación conceptual del canal digital.

En este modelo se considera que cada símbolo de la secuencia $B[n]$ es estadísticamente independiente del resto de símbolos de la secuencia. Por otra parte, ya se ha visto en el apartado dedicado a la transmisión indefinida de símbolos, que dado que las señales sólo ocupan el intervalo de duración T dedicado al símbolo y que el ruido es independiente en cada momento, la recepción de cada símbolo es independiente del resto de símbolos. Bajo estas condiciones, la probabilidad de tener un símbolo determinado de la secuencia de salida del canal digital, $\hat{B}[n]$, depende únicamente del símbolo emitido en ese mismo instante $B[n]$. Por tanto, es posible eliminar la dependencia temporal y analizar únicamente el caso de la transmisión de un símbolo aislado, entendiendo que cada vez que se use el canal para transmitir un símbolo, el canal no va a modificar su comportamiento.

Cuando se transmite el símbolo b_i a la entrada del canal, a la salida del mismo vamos a tener un símbolo, b_j, con una probabilidad determinada que es la probabilidad condicional

$$p_{\hat{B}|B}(b_j|b_i).$$

Debido a la asignación unívoca entre un símbolo y la representación vectorial de la señal que lo transmite, esta probabilidad cumple que

$$p_{\hat{B}|B}(b_j|b_i) = p_{\hat{\boldsymbol{A}}|\boldsymbol{A}}(\boldsymbol{a}_j|\boldsymbol{a}_i).$$

En el capítulo anterior se estudio cómo se calculan estas probabilidades integrando la distribución condicional de la observación para el símbolo transmitido en la región de decisión de cada uno de los símbolos de la constelación. Como en este modelo la entrada y la salida eran precisamente $X \equiv B$ e $Y \equiv \hat{B}$, si se conocen estas probabilidades para todas las combinaciones posibles de símbolos transmitido y recibido, el canal estará completamente caracterizado.

Existe un modelo probabilístico ampliamente utilizado que contempla como caso particular el canal digital y se denomina *canal discreto sin memoria* o DMC (del inglés *Discrete Memoryless Channel*). El DMC es un modelo estadístico que relaciona una variable aleatoria X con función densidad de probabilidad discreta que denominamos entrada y otra variable aleatoria Y con función de probabilidad discreta que denominamos salida. En el caso particular del DMC, los alfabetos de entrada y salida pueden ser distintos. En nuestra aplicación comparten el mismo alfabeto, por lo que este es un caso particular del DMC.

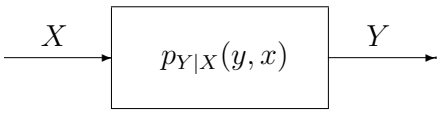

Figura 5.9: Modelo probabilístico del canal discreto sin memoria (DMC).

El canal discreto sin memoria se puede representar mediante el diagrama de la Figura 5.9. Se observa que dentro del bloque que representa el canal no aparecen respuestas al impulso ni respuestas en frecuencia, sino las probabilidades condicionales de la salida respecto a la entrada. El calificativo de *discreto* proviene de la naturaleza de las probabilidades de X e Y, que tienen un alfabeto discreto, y no de *tiempo discreto*, por lo que no hay que confundirlo con el canal discreto equivalente. El calificativo sin memoria proviene del modelo probabilístico de la entrada y la salida, variables aleatorias y no procesos aleatorios (no hay dependencia temporal de los estadísticos). Formalmente, un canal discreto sin memoria se define a través de los siguientes elementos:

1. El *alfabeto de entrada* (de M_X posibles valores)

$$\{x_i\},\ i = 0, \cdots, M_X - 1.$$

2. El *alfabeto de salida* (de M_Y posibles valores)

$$\{y_i\},\ i = 0, \cdots, M_Y - 1.$$

3. El conjunto de probabilidades condicionales

$$p_{Y|X}(y_j|x_i).$$

Estas probabilidades se denominan *probabilidades de transición*, y se suelen agrupar en la denominada *matriz de canal*, que es una matriz de M_X filas y M_Y columnas que ordena las probabilidades de transición de la forma siguiente

$$\boldsymbol{P} = \begin{bmatrix} p_{Y|X}(y_0|x_0) & p_{Y|X}(y_1|x_0) & \cdots & p_{Y|X}(y_{M_Y-1}|x_0) \\ p_{Y|X}(y_0|x_1) & p_{Y|X}(y_1|x_1) & \cdots & p_{Y|X}(y_{M_Y-1}|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p_{Y|X}(y_0|x_{M_X-1}) & p_{Y|X}(y_1|x_{M_X-1}) & \cdots & p_{Y|X}(y_{M_Y-1}|x_{M_X-1}) \end{bmatrix}$$

Tal y como está definida la matriz, una fila está asociada a un cierto valor del alfabeto de entrada, mientras que una columna está asociada a un cierto valor del alfabeto de salida. Por tanto, la suma de los elementos de una fila da como resultado 1 (se tiene la suma de la distribución condicional de probabilidad sobre todo el espacio de observación). Además, si se concatenan dos canales discretos sin memoria, la matriz de canal de la concatenación se obtendrá mediante el producto de las matrices de canal de cada uno de los canales.

En ocasiones, en lugar de la matriz de canal se utiliza una representación gráfica para especificar las probabilidades de transición, utilizando un diagrama de flechas o diagrama de rejilla, como el que muestra la Figura 5.10. En este caso, las probabilidades de transición se incluyen en los pesos asociados a las distintas flechas que forman el diagrama uniendo elementos de la entrada con la salida. Dada la definición de estas probabilidades, las probabilidades asociadas a las flechas que salen de un mismo nodo suman la unidad.

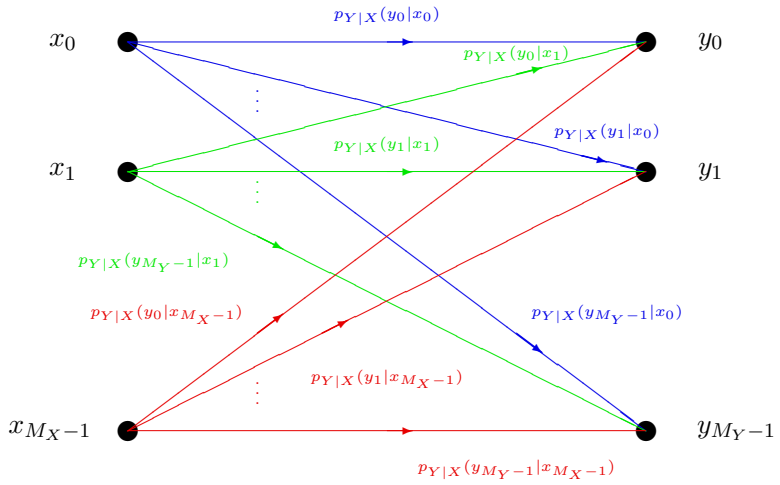

Figura 5.10: Representación de un DMC mediante un diagrama de rejilla.

Hay que resaltar que en la definición de un canal discreto sin memoria aparecen el alfabeto de entrada y el de salida, pero no las distribuciones de probabilidades de la entrada, $p_X(x_i)$ ni de la salida, $p_Y(y_j)$, ya que estas probabilidades no forman parte de la naturaleza del canal.

A continuación se muestra cómo obtener un DMC que represente al canal digital para un cierto sistema de comunicaciones. El modelo se obtiene a partir de las probabilidades de error de símbolo definidas en el capítulo anterior. En primer lugar, para un DMC que modela un sistema de comunicaciones tanto el alfabeto de entrada como el de salida se corresponden con el alfabeto de símbolos del sistema, bien de $\boldsymbol{A}[n]$ o de $B[n]$, dada su equivalencia. Por conveniencia utilizaremos la representación vectorial de los símbolos, es decir

$$x_i \equiv \boldsymbol{a}_i,$$

$$y_j \equiv \boldsymbol{a}_j,$$

donde ahora $M_X = M_Y = M$. De esta forma, la asociación del alfabeto con los símbolos de la constelación está implícita en los subíndices. En cuanto a las probabilidades de transición, es evidente que

$$p_{Y|X}(y_j|x_i) \equiv p_{\hat{\boldsymbol{A}}|\boldsymbol{A}}(\boldsymbol{a}_j|\boldsymbol{a}_i).$$

Es decir, las probabilidad de transición $p_{Y|X}(y_j|x_i)$ indica la probabilidad de recibir el símbolo $\boldsymbol{a}_j$ cuando se ha transmitido el símbolo $\boldsymbol{a}_i$. En este caso, los elementos de la diagonal principal de la matriz de canal, para los que $j = i$, se corresponden con las probabilidades condicionales de acierto para cada símbolo

$$p_{Y|X}(y_i|x_i) = p_{\hat{\boldsymbol{A}}|\boldsymbol{A}}(\boldsymbol{a}_i|\boldsymbol{a}_i) = P_{a|\boldsymbol{a}_i} = 1 - P_{e|\boldsymbol{a}_i}.$$

Los elementos fuera de la diagonal en cambio se corresponden con las probabilidades de error entre distintos símbolos

$$p_{Y|X}(y_j|x_i) = \underset{j\neq i}{p_{\hat{\boldsymbol{A}}|\boldsymbol{A}}(\boldsymbol{a}_j|\boldsymbol{a}_i)} = P_{e|\boldsymbol{a}_i\to\boldsymbol{a}_j}.$$

Consecuentemente, la suma de los elementos de cada fila fuera de la diagonal principal equivale a la probabilidad de error condicional para el símbolo asociado a dicha fila

$$\sum_{\substack{j=0\\ j\neq i}}^{M-1} p_{Y|X}(y_j|x_i) = \sum_{\substack{j=0\\ j\neq i}}^{M-1} P_{e|\boldsymbol{a}_i\to\boldsymbol{a}_j} = P_{e|a_i}.$$

Esto significa que en un sistema ideal, la matriz de canal o el diagrama de rejilla deberían ser como se muestra en la Figura 5.11: una matriz diagonal, o un diagrama de rejilla con una única flecha partiendo de cada símbolo de entrada al mismo símbolo de salida.

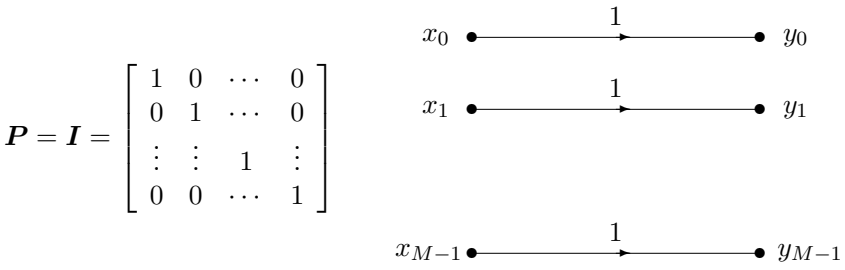

Figura 5.11: Valores ideales de un DMC que modela un canal gausiano.

Para ilustrar el procedimiento de obtención de la matriz de canal para un sistema de comunicaciones se utilizará como ejemplo un sistema con una constelación de cuatro símbolos, $M = 4$, en un espacio unidimensional, $N = 1$, con coordenadas $\boldsymbol{a}_0 = -3$, $\boldsymbol{a}_1 = -1$, $\boldsymbol{a}_2 = +1$, $\boldsymbol{a}_3 = +3$ y equiprobables, con lo que las regiones de decisión vienen dadas por los umbrales $q_{u1} = -2$, $q_{u2} = 0$, $q_{u3} = +2$

$$I_0 = (-\infty, -2],\ I_1 = (-2, 0],\ I_2 = (0, +2],\ I_3 = (+2, +\infty)$$

tal y como muestra la Figura 5.12.

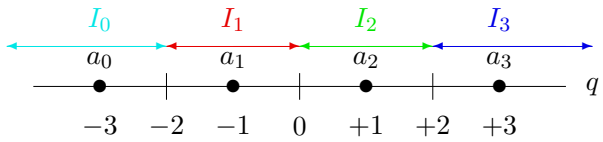

Figura 5.12: Constelación unidimensional de cuatro símbolos equiprobables y sus correspondientes regiones de decisión.

En este caso se tiene la siguiente asociación

$$X \equiv \boldsymbol{A}[n],\ Y \equiv \hat{\boldsymbol{A}}[n],$$

por lo que los alfabetos de entrada y salida coinciden

$$x_i \equiv \boldsymbol{a}_i,\ y_j \equiv \boldsymbol{a}_j \text{ para } i, j \in \{0, 1, \cdots, M-1\}.$$

Las probabilidades de transición $p_{Y|X}(y_j|x_i)$ que definen probabilísticamente el sistema se corresponden en este caso con la probabilidad de recibir el símbolo de índice j cuando se ha transmitido el símbolo de índice i. Estos valores definen la probabilidad de acertar en la transmisión de un símbolo, si $j = i$, o la probabilidad de error entre dos símbolos, si $j \neq i$. Utilizando la notación del capítulo anterior, se tiene

$$p_{Y|X}(y_i|x_i) = P_{a|\boldsymbol{a}_i} = 1 - P_{e|\boldsymbol{a}_i}$$

y

$$p_{Y|X}(y_j|x_i) = P_{e|\boldsymbol{a}_i \to \boldsymbol{a}_j} \text{ para } j \neq i.$$

En el capítulo anterior se explicó cómo se obtenían estos valores, que se volverán a obtener para esta constelación. En primer lugar se calcularán las probabilidades de transición que aparecen en la primera fila de la matriz de canal, $p_{Y|X}(y_j|x_0)$, $\forall j$; es decir, que en este caso se tienen las probabilidades de recibir cada uno de los 4 símbolos cuando se transmite $\boldsymbol{a}_0$ (el símbolo asociado a x_0). La distribución de la observación cuando se transmite el símbolo $\boldsymbol{a}_0$ es gausiana, de media $\boldsymbol{a}_0$ y varianza $N_0/2$. Así que para obtener las probabilidades condicionales, sólo hay que integrar esa distribución gausiana en cada una de las 4 regiones de decisión, tal y como se ilustra en la Figura 5.13.

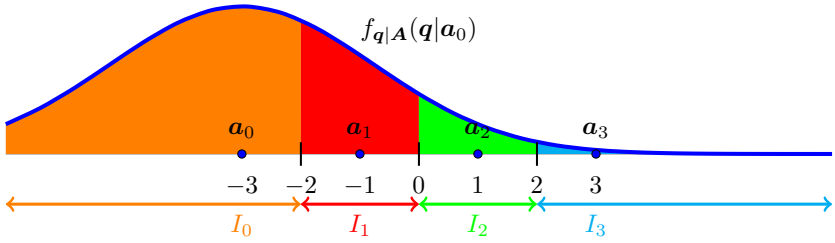

Figura 5.13: Cálculo de las probabilidades de transición asociadas a la primera fila de la matriz de canal.

A continuación se calculan estas probabilidades

- Distribución $f_{\boldsymbol{q}|\boldsymbol{A}}(\boldsymbol{q}|\boldsymbol{a}_0)$: gausiana de media $\boldsymbol{a}_0 = -3$ y varianza $N_0/2$

$$\begin{aligned}
p_{Y|X}(y_0|x_0) =& 1 - P_{e|\boldsymbol{a}_0} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right) \\
p_{Y|X}(y_1|x_0) =& P_{e|\boldsymbol{a}_0 \to \boldsymbol{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) \\
p_{Y|X}(y_2|x_0) =& P_{e|\boldsymbol{a}_0 \to \boldsymbol{a}_2} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) \\
p_{Y|X}(y_3|x_0) =& P_{e|\boldsymbol{a}_0 \to \boldsymbol{a}_3} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)
\end{aligned}$$

Se puede comprobar que estas cuatro probabilidades suman la unidad, como era de esperar.

A continuación se calcularán las probabilidades de transición que aparecen en la segunda fila de la matriz de canal, $p_{Y|X}(y_j|x_1)$, $\forall j$; es decir, que en este caso se tienen las probabilidades de recibir

cada uno de los 4 símbolos cuando se transmite $\boldsymbol{a}_1$ (el símbolo asociado a x_1). La distribución de la observación cuando se transmite el símbolo $\boldsymbol{a}_1$ es gausiana, de media $\boldsymbol{a}_1$ y varianza $N_0/2$. Así que para obtener las probabilidades condicionales, sólo hay que integrar esa distribución gausiana en cada una de las 4 regiones de decisión, tal y como se muestra en la Figura 5.14.

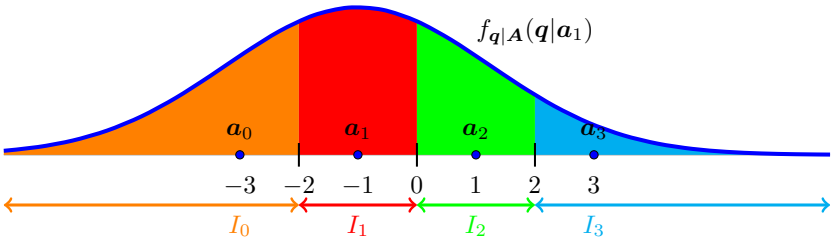

Figura 5.14: Cálculo de las probabilidades de transición asociadas a la primera segunda de la matriz de canal.

A continuación se calculan estas probabilidades

- Distribución $f_{\boldsymbol{q}|\boldsymbol{A}}(\boldsymbol{q}|\boldsymbol{a}_1)$: gausiana de media $\boldsymbol{a}_1$ y varianza $N_0/2$

$$
\begin{aligned}
p_{Y|X}(y_0|x_1) =& P_{e|\boldsymbol{a}_1\to\boldsymbol{a}_0} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_1|x_1) =& 1 - P_{e|\boldsymbol{a}_1} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_2|x_1) =& P_{e|\boldsymbol{a}_1\to\boldsymbol{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_3|x_1) =& P_{e|\boldsymbol{a}_1\to\boldsymbol{a}_3} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)
\end{aligned}
$$

En la tercera fila de la matriz de canal se tienen las probabilidades de transición $p_{Y|X}(y_j|x_2)$, $\forall j$; es decir, que en este caso se tienen las probabilidades de recibir cada uno de los 4 símbolos cuando se transmite $\boldsymbol{a}_2$ (el símbolo asociado a x_2). La distribución de la observación cuando se transmite el símbolo $\boldsymbol{a}_2$ es en este caso gausiana de media $\boldsymbol{a}_2 = +1$ y varianza $N_0/2$. Al igual que en los casos anteriores, para obtener las probabilidades condicionales hay que integrar esa distribución gausiana en cada una de las 4 regiones de decisión, tal y como se ilustra en la Figura 5.15.

A continuación se calculan estas probabilidades

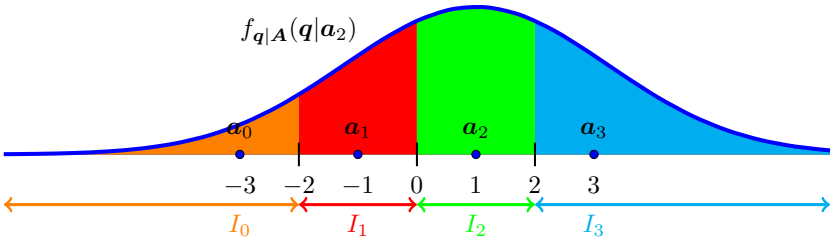

Figura 5.15: Cálculo de las probabilidades de transición asociadas a la tercera fila de la matriz de canal.

- Distribución $f_{\boldsymbol{q}|\boldsymbol{A}}(\boldsymbol{q}|\boldsymbol{a}_2)$: gausiana de media $\boldsymbol{a}_2$ y varianza $N_0/2$

$$
\begin{aligned}
p_{Y|X}(y_0|x_2) =& P_{e|\boldsymbol{a}_2\to\boldsymbol{a}_0} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_1|x_2) =& P_{e|\boldsymbol{a}_2\to\boldsymbol{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_2|x_2) =& 1 - P_{e|\boldsymbol{a}_2} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_3|x_2) =& P_{e|\boldsymbol{a}_2\to\boldsymbol{a}_3} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)
\end{aligned}
$$

Finalmente, en la cuarta y última fila de la matriz de canal se tienen las probabilidades de transición $p_{Y|X}(y_j|x_2)$, $\forall j$; es decir, que en este caso se tienen las probabilidades de recibir cada uno de los 4 símbolos cuando se transmite $\boldsymbol{a}_3$ (el símbolo asociado a x_3). La distribución de la observación cuando se transmite el símbolo $\boldsymbol{a}_3$ es en este caso gausiana de media $\boldsymbol{a}_3 = +3$ y varianza $N_0/2$. Las probabilidades condicionales se obtienen integrando esta distribución condicional en cada una de las 4 regiones de decisión, tal y como se muestra en la Figura 5.16.

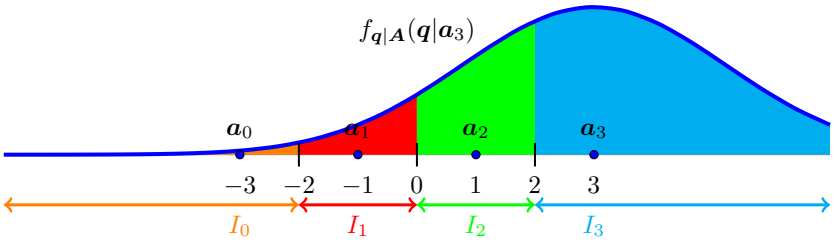

Figura 5.16: Cálculo de las probabilidades de transición asociadas a la cuarta fila de la matriz de canal.

A continuación se calculan estas probabilidades

- Distribución $f_{\boldsymbol{q}|\boldsymbol{A}}(\boldsymbol{q}|\boldsymbol{a}_3)$: gausiana de media $\boldsymbol{a}_3$ y varianza $N_0/2$

$$\begin{aligned}
p_{Y|X}(y_0|x_3) =& P_{e|\boldsymbol{a}_3\to\boldsymbol{a}_0} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_1|x_3) =& P_{e|\boldsymbol{a}_3\to\boldsymbol{a}_1} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_2|x_3) =& P_{e|\boldsymbol{a}_3\to\boldsymbol{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)\\
p_{Y|X}(y_3|x_3) =& 1 - P_{e|\boldsymbol{a}_3} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)
\end{aligned}$$

Para este ejemplo, dada la simetría de la constelación, una vez obtenidos los valores de las proabilidades de transición para las dos primeras filas, las de las dos siguientes se pueden obtener de forma inmediata. En cualquier caso, agrupando todas las probabilidades de transición, el DMC que representa al sistema de comunicaciones que utiliza la constelación de 4 símbolos del ejemplo tiene la siguiente matriz de canal

$$\boldsymbol{P} = \begin{bmatrix}
1 - Q\left(\frac{1}{\sqrt{N_o/2}}\right) & Q\left(\frac{1}{\sqrt{N_o/2}}\right) - Q\left(\frac{3}{\sqrt{N_o/2}}\right) & Q\left(\frac{3}{\sqrt{N_o/2}}\right) - Q\left(\frac{5}{\sqrt{N_o/2}}\right) & Q\left(\frac{5}{\sqrt{N_o/2}}\right)\\
Q\left(\frac{1}{\sqrt{N_o/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_o/2}}\right) & Q\left(\frac{1}{\sqrt{N_o/2}}\right) - Q\left(\frac{3}{\sqrt{N_o/2}}\right) & Q\left(\frac{3}{\sqrt{N_o/2}}\right)\\
Q\left(\frac{3}{\sqrt{N_o/2}}\right) & Q\left(\frac{1}{\sqrt{N_o/2}}\right) - Q\left(\frac{3}{\sqrt{N_o/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_o/2}}\right) & Q\left(\frac{1}{\sqrt{N_o/2}}\right)\\
Q\left(\frac{5}{\sqrt{N_o/2}}\right) & Q\left(\frac{3}{\sqrt{N_o/2}}\right) - Q\left(\frac{5}{\sqrt{N_o/2}}\right) & Q\left(\frac{1}{\sqrt{N_o/2}}\right) - Q\left(\frac{3}{\sqrt{N_o/2}}\right) & 1 - Q\left(\frac{1}{\sqrt{N_o/2}}\right)
\end{bmatrix}.$$

Para distintos valores de N_0, la matriz tendrá distintos valores. A medida que N_0 disminuye, la matriz tiende a la matriz identidad, que es la matriz de canal ideal.

5.2.4. Canal digital binario

El canal digital binario es el modelo probabilístico que supone una mayor abstracción al considerar todo el sistema de comunicaciones como un canal que transmite y recibe bits.

Figura 5.17: Representación conceptual del canal digital binario.

Se trata por tanto de un modelo en el que la descripción probabilística vendrá dada por la probabilidad de recibir cada uno de los posibles valores de la secuencia binaria $\hat{B}_b[\ell]$ dada la secuencia transmitida $B_b[\ell]$. En el caso en que se asume independencia temporal sobre la secuencia de bits transmitida y recibida, se puede obviar la dependencia temporal, y el modelo en este caso

está definido por cuatro probabilidades de transición que, si se utiliza la notación $x_0 \equiv 0$, $x_1 \equiv 1$, $y_0 \equiv 0$ e $y_1 \equiv 1$ y se tiene en cuenta que por definición

$$p_{Y|X}(y_0|x_i) = 1 - p_{Y|X}(y_1|x_i),$$

quedan reducidas únicamente a dos probabilidades de transición relevantes: la probabilidad de error (o de acierto) condicional para cada bit. En ese caso, puede utilizarse un DMC particularizado para el caso $M_X = M_Y = 2$ como modelo del canal digital binario, que tendrá la forma

$$\boldsymbol{P} = \begin{bmatrix} p_{Y|X}(y_0|x_0) & p_{Y|X}(y_1|x_0) \\ p_{Y|X}(y_0|x_1) & p_{Y|X}(y_1|x_1) \end{bmatrix} = \begin{bmatrix} 1-p_{e|0} & p_{e|0} \\ p_{e|1} & 1-p_{e|1} \end{bmatrix}.$$

Las probabilidades $p_{e|0}$ y $p_{e|1}$ denotan la probabilidad de error de bit cuando se transmite un cero y la probabilidad de error de bit cuando de transmite un uno, respectivamente. En la mayor parte de los sistemas de comunicaciones estas dos probabilidades son iguales

$$p_{e|0} = p_{e|1} = \varepsilon,$$

en cuyo caso la matriz de canal es simétrica

$$\boldsymbol{P} = \begin{bmatrix} p_{Y|X}(y_0|x_0) & p_{Y|X}(y_1|x_0) \\ p_{Y|X}(y_0|x_1) & p_{Y|X}(y_1|x_1) \end{bmatrix} = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}.$$

Este caso se conoce como *canal binario simétrico* o BSC (del inglés "*Binary Symmetric Channel*"). La representación en diagrama de rejilla para este modelo se muestra en la Figura 5.18.

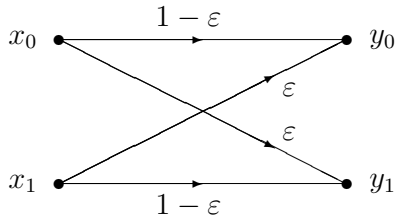

Figura 5.18: Representación en diagrama de rejilla de un canal binario simétrico o BSC.

El BSC es un modelo probabilístico que puede resultar apropiado para representar el canal digital binario. Sin embargo, antes de dar por buena la equivalencia entre canal digital binario y BSC, conviene realizar las siguientes puntualizaciones:

1. Para sistemas binarios, $M = 2$, un símbolo de la secuencia $B[n]$ transporta un único bit, y si los símbolos se transmiten de forma independiente en $B[n]$ eso implica una transmisión independiente a nivel de bit en la secuencia $B_b[\ell]$. En este caso, el modelo BSC representa de forma exacta el canal digital binario.

2. Para sistemas M-ários con $M > 2$, el BSC representa el comportamiento medio a lo largo del tiempo del canal digital binario, porque en el sistema real la transmisión se realiza por bloques de $m = \log_2 M$ bits, símbolo a símbolo (secuencia $B[n]$ y su representación vectorial $\boldsymbol{A}[n]$); se puede por tanto decir que mientras que el canal digital no posee memoria (porque la transmisión se realiza símbolo a símbolo de manera independiente) y encaja perfectamente en el modelo del DMC, el canal digital binario tiene la memoria introducida por el codificador que transforma la secuencia $B_b[\ell]$ en la secuencia $B[n]$. Eso significa que no puede considerarse de forma estricta un canal sin memoria. Desde este punto de vista, el BSC es una aproximación al canal digital binario que representa su comportamiento medio a lo largo del tiempo.

3. El valor de la probabilidad de error ε definida para el BSC es en ambos casos es la probabilidad de error de bit (BER) del sistema.

A pesar de esta discordancia entre las hipótesis del modelo BSC y la naturaleza del canal digital binario, se suele aceptar en la práctica la equivalencia de ambos, asignando a p el valor de la BER del sistema

$$\boldsymbol{P} = \begin{bmatrix} 1-BER & BER \\ BER & 1-BER \end{bmatrix}.$$

5.3. Medidas cuantitativas de información

Una vez establecidos los modelos probabilísticos para fuentes de información, el modelo discreto sin memoria o DMS, y para los canales digitales (canal digital y canal digital binario), en este caso el modelo sin memoria o DMC, en esta sección se analizan varios tipos de medidas cuantitativas de información. Por un lado medidas que se pueden aplicar sobre una variable aleatoria, que puede ser por ejemplo la que modela la entrada o salida de un canal, y otras que se aplican simultáneamente sobre dos variables aleatorias, que en nuestro problema tendrían en cuenta la relación entre la entrada y la salida de un canal. A través de estas medidas se podrá posteriormente calcular la máxima cantidad de información que se puede transmitir de forma fiable con un sistema de comunicaciones digital.

5.3.1. Información y entropía

En primer lugar se va a cuantificar la información que tiene una cierta variable aleatoria discreta. En la aplicación para el estudio de sistemas de comunicaciones, esta variable aleatoria puede representar tanto la salida de una fuente de información como la entrada o salida de un canal digital.

Para obtener una medida cuatitativa de la información que contiene una variable aleatoria, en primer lugar se buscará una medida para la información que contiene un suceso de esa variable aleatoria, es decir, el hecho de que la variable aleatoria tome un cierto valor dentro de su alfabeto, como $X = x_i$. Antes de establecer una media cuantitativa de información para este caso, se van a presentar de forma intuitiva algunas de las propiedades básicas que esta medida debe cumplir, con el objetivo de encontrar luego alguna función que cumpla dichas propiedades y que pueda utilizarse como medida cuantitativa de la información que tiene un suceso de una variable aleatoria.

- Una noción intuitiva de información indica que la cantidad de información de un cierto suceso está relacionada con la probabilidad con la que este se produce. El conocimiento de que se ha producido un suceso poco probable aporta en general más información que el de un suceso más probable. Por poner un ejemplo sencillo, el número medio de días de lluvia al año en algunos puntos del desierto del Sahara es de menos de un día al año, mientras que en la ciudad de Santander es de 151. El hecho de conocer que hoy ha llovido en el desierto del Sahara es más informativo que el hecho de conocer que hoy a llovido en Santander (el primer caso puede aparecer en algún informativo, mientras que es poco probable que el segundo se considere algo noticioso o informativo).
- Por otro lado, cambios pequeños en la probabilidad del suceso deberían dar lugar a pequeños cambios en la información del mismo. Por ejemplo, el suceso de que llueva en Santander

tendrá más o menos la misma cantidad de información que el suceso de que llueva en Gijón, mientras que el suceso de que llueva en el desierto del Sahara tendrá una información similar al suceso de que llueva en el desierto del Gobi o el de Atacama. Esta idea intuitiva se puede interpretar como que la dependencia de la información con respecto de la probabilidad del suceso debe ser continua.

- Finalmente, parece también intuitivo pensar que si se tienen varios sucesos independientes entre sí, la información que tienen en conjunto deberá ser la suma de la información de cada uno de ellos por separado.

Partiendo de estas nociones intuitivas sobre algunas características que debe tener una medida de información sobre un cierto suceso, se llega a la denominada *auto-información*, que es una medida cuantitativa de la información que contiene un suceso de una variable aleatoria.

Auto-información

La auto-información del suceso $X = x_i$ se denota como $I_X(x_i)$. Para obtener la expresión analítica de esta función se han traducido a notación matemática las nociones intuitivas sobre dicha medida que se han comentado anteriormente, lo que da lugar a las cuatro condiciones que ha de cumplir dicha función

1. La medida de información de un suceso deberá depender de la probabilidad de dicho suceso, y no del valor del propio suceso, es decir
$$I_X(x_i) = f(p_X(x_i)).$$
2. Debe además ser una función decreciente de la probabilidad del suceso
$$p_X(x_i) > p_X(x_j) \text{ implicará que } I_X(x_i) < I_X(x_j),$$
lo que significa que la función $f(\cdot)$ deberá ser una función decreciente
$$f(a) < f(b) \text{ para todo } a > b.$$
3. La función $f(\cdot)$ utilizada para la autoinformación deberá también ser una función continua de su argumento, de forma que la variación de información será continua sobre la probabilidad de los sucesos.
4. Finalmente, si dos variables aleatorias son independientes, y se define un suceso conjunto $X = x_i$ e $Y = y_j$, la información del suceso conjunto deberá ser la suma de la información de cada suceso
$$I_{X,Y}(x_i, y_j) = I_X(x_i) + I_Y(y_j).$$
Dado que para variables aleatorias independientes la probabilidad conjunta se puede escribir como el producto de las probabilidades marginales de cada variable
$$p_{X,Y}(x_i, y_j) = p_X(x_i) \times p_Y(y_j),$$
esto significa que la función $f(\cdot)$ elegida para la medida de información deberá ser aditiva sobre el producto de los argumentos, es decir
$$f(a \times b) = f(a) + f(b).$$

Se puede demostrar que la única función que cumple estas propiedades es la función logarítmica. Por tanto, la auto-información se define como

$$I_X(x_i) = -\log(p_X(x_i)).$$

Teniendo en cuenta las propiedades de la función logarítmica, la auto-información se puede escribir alternativamente como

$$I_X(x_i) = \log\left(\frac{1}{p_X(x_i)}\right).$$

La base del logaritmo no es determinante para las características de la medida de información. Lo único que implica son las unidades en que se expresa la información. Las bases utilizadas con mayor frecuencia son 2 y la base natural o número e (logaritmo neperiano). Si la base es 2, las unidades son *bits*, y si se usa el logaritmo natural o neperiano, las unidades son *nats*. En lo sucesivo, cuando no se especifique la base, se asumirán logaritmos de base 2 y por tanto bits como unidades de información. En cualquier caso, el cambio de la base, y por tanto de unidades, no supone más que un escalado, ya que en general los logaritmos en una cierta base se relacionan con los logaritmos en base natural a través de la siguiente relación

$$\log_b x = \frac{\log_e x}{\log_e b} = \frac{\ln x}{\ln b},$$

lo que directamente supone una relación lineal entre los logaritmos en dos bases distintas.

Entropía

La auto-información proporciona una medida cuantitativa de la información de un suceso aislado. Si se quiere cuantificar la información de una variable aleatoria (por ejemplo para modelar una fuente de información) han de tenerse en cuenta todos los posibles sucesos. Una opción razonable consiste en promediar la información de cada suceso considerando su probabilidad. El contenido de información de una variable aleatoria así calculado se denomina *entropía* y se denota mediante $H(X)$. Por tanto, la entropía de la variable aleatoria X se obtiene promediando la auto-información de cada uno de los sucesos que forman parte del alfabeto de la variable aleatoria

$$H(X) = -\sum_{i=0}^{M_X-1} p_X(x_i)\cdot\log p_X(x_i) = \sum_{i=0}^{M_X-1} p_X(x_i)\cdot\log\left(\frac{1}{p_X(x_i)}\right).$$

Las unidades serán bits o nats por símbolo, dependiendo de la base empleada.

A efectos del cómputo hay que tener en cuenta que se considerará que $0\log(0) = 0$. Como se ve, la entropía es una función de la probabilidad de cada suceso, o lo que es lo mismo, de la función densidad de probabilidad discreta, y proporciona un número que representa el contenido de información de esa fuente. No se debe confundir con una función de variable aleatoria que es otra variable aleatoria, tal y como la notación puede hacer parecer.

La entropía puede interpretarse como una cantidad que representa la incertidumbre sobre el valor concreto que toma una variable aleatoria X, que puede modelar, por ejemplo, la salida de una fuente de información. Si X toma siempre el mismo valor x_i, es decir, si $p_X(x_i) = 1$, no se tiene ninguna incertidumbre acerca del valor de la variable aleatoria y la entropía vale 0. Si X deja de tomar siempre el mismo valor, entonces la incertidumbre aumenta y con ella la entropía.

Dos propiedades importantes de la entropía de una variable aleatoria discreta son:

1. La entropía de una variable aleatoria discreta es una función no negativa, es decir

$$H(X) \geq 0.$$

   Esto es evidente ya que el rango de valores de una probabilidad es $0 \leq p_X(x_i) \leq 1$ y $\log(x) \leq 0$ para $0 < x \leq 1$. El valor $H(X) = 0$ sólo se produce si uno de los elementos del alfabeto tiene probabilidad uno y por tanto el resto tiene probabilidad nula.

2. El valor máximo que puede tomar la entropía de una variable aleatoria discreta es el logaritmo del número de elementos de su alfabeto

$$H(X) \leq \log(M_X).$$

   Ese valor máximo se produce únicamente si los símbolos son equiprobables, $p_X(x_i) = 1/M_X$.

Estas dos propiedades establecen los límites para los valores mínimo y máximo que puede tomar la entropía de una variable aleatoria X, e indican bajo que distribuciones se obtienen dichos valores mínimo y máximo, respectivamente. Para ilustrar estas propiedades, a continuación se calcula la entropía en un caso muy sencillo: una variable aleatoria binaria, $M_X = 2$, donde las probabilidades de cada símbolo se parametrizan con la probabilidad de uno de los dos elementos del alfabeto, $p_X(x_0) = p$, $p_X(x_1) = 1 - p$. En este caso la entropía es igual a

$$H(X) = -p \cdot \log(p) - (1-p) \cdot \log(1-p) = p \cdot \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \equiv H_b(p)$$

Esta función, denotada como $H_b(p)$, o en ocasiones como $\Omega(p)$, se denomina *función de entropía binaria*, y se representa en la Figura 5.19 en función de su argumento. Recuerde que dicho argumento representa la probabilidad de uno de los dos elementos del alfabeto de la variable aleatoria binaria.

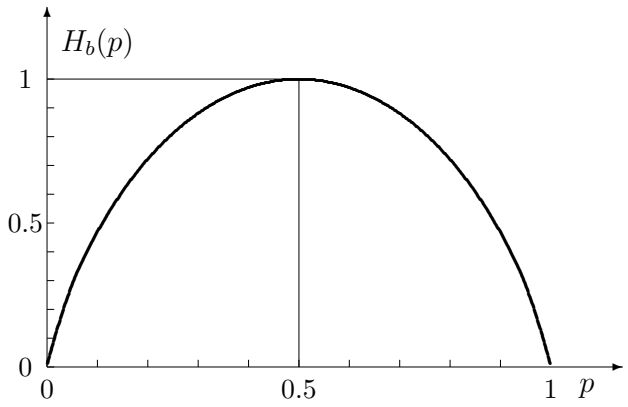

Figura 5.19: Función de entropía binaria, $H_b(p)$, expresada en bits por símbolo.

Se comprueba que en los casos en los que no existe incertidumbre, $p = 0$ o $p = 1$, la entropía es nula. Fuera de estos casos el valor es siempre mayor que cero, tomando el valor máximo, 1 bit por símbolo, cuando los símbolos son equiprobables, lo que representa la situación de máxima incertidumbre posible. Además, dado que la entropía depende únicamente de los valores de la distribución de probabilidad y no de a qué valores está asignado cada uno de esos valores, es una función simétrica respecto a $p = \frac{1}{2}$, lo que es evidente dado que por definición

$$H_b(p) = H_b(1-p).$$

La función entropía binaria puede servir de referencia para definir un bit de información: un bit es la información que se tiene cuando se transmiten dos símbolos con igual probabilidad.

A continuación se muestra un ejemplo de cálculo de la entropía de una variable aleatoria con un alfabeto de más símbolos, en contreto de cinco símbolos.

Ejemplo

Una fuente puede ser modelada con el modelo DMS con un alfabeto

$$\mathcal{A}_X = \{-2, -1, 0, 1, 2\},$$

y unas probabilidades

$$p_X(-2) = \frac{1}{2},\ p_X(-1) = \frac{1}{4},\ p_X(0) = \frac{1}{8},\ p_X(1) = \frac{1}{16},\ p_X(2) = \frac{1}{16}.$$

En este caso la entropía vale

$$H(X) = \frac{1}{2}\log(2) + \frac{1}{4}\log(4) + \frac{1}{8}\log(8) + 2 \times \frac{1}{16}\log(16) = \frac{15}{8} \text{ bits/símbolo.}$$

Se puede ver que el valor de entropía depende únicamente de las probabilidades de los posibles elementos del alfabeto, y no de los valores concretos del alfabeto. Por ejemplo, una fuente con un alfabeto diferente

$$\mathcal{A}_X = \{0, 1, 2, 3, 4\},$$

pero con el mismo conjunto de valores para las probabilidades, aunque con distinta asignación a cada elemento del alfabeto

$$p_X(0) = \frac{1}{2},\ p_X(1) = \frac{1}{4},\ p_X(2) = \frac{1}{8},\ p_X(3) = \frac{1}{16},\ p_X(4) = \frac{1}{16},$$

tiene la misma entropía que la anterior.

5.3.2. Entropía conjunta

La definición de entropía se puede extender a más de una variable aleatoria, lo que tendría aplicación por ejemplo para medir la entropía conjunta de la entrada y la salida de un sistema de comunicaciones. La *entropía conjunta* de dos variables aleatorias X e Y, en general con alfabetos y probabilidades distintos, $\mathcal{A}_X = \{x_i\}_{i=0}^{M_X-1}$, $p_X(x_i)$, y $\mathcal{A}_Y = \{y_i\}_{i=0}^{M_Y-1}$, $p_Y(y_i)$, se define como una extensión trivial de la entropía de una variable aleatoria, considerando en este caso que hay tantos sucesos como casos conjuntos y que la probabilidad de cada uno de ellos está dada por la probabilidad conjunta, lo que lleva a la expresión

$$H(X,Y) = \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log\left(\frac{1}{p_{X,Y}(x_i, y_j)}\right).$$

Al igual que para la entropía de una variable aleatoria, dadas las propiedades de la función logaritmo, se puede cambiar el signo e invertir el argumento del logaritmo,

$$H(X,Y) = -\sum_{i=0}^{M_X-1}\sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log\left(p_{X,Y}(x_i, y_j)\right).$$

Al igual que la entropía, también se mide en bits o nats por símbolo. El concepto se puede extender a N variables aleatorias. En este caso

$$\boldsymbol{X} = (X_1, X_1, \cdots, X_N),$$

y

$$H(\boldsymbol{X}) = -\sum_{x_1, x_2, \cdots, x_N} p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N) \cdot \log(p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)).$$

En esta notación, el sumatorio indica el N sumatorios contemplando todas las posibles combinaciones de los alfabetos de cada variable aleatoria.

La interpretación de la entropía conjunta no difiere de la de entropía para una variable aleatoria. Al fin y al cabo, un par de variables X e Y pueden considerarse como una única variable aleatoria vectorial con un alfabeto de $M_X \cdot M_Y$ símbolos.

Si las variables aleatorias X e Y son independientes se cumple que

$$p_{X,Y}(x_i, y_j) = p_X(x_i) \times p_Y(y_j).$$

En este caso su entropía conjunta es la suma de las entropías individuales. Esto se había apuntado en la definición de las condiciones que debía cumplir la medida de información, y se demuestra de forma muy sencilla, como puede verse en el siguiente desarrollo

$$\begin{aligned}
H(X,Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_X(x_i) \cdot p_Y(y_j) \cdot \log \frac{1}{p_X(x_i) \cdot p_Y(y_j)} \\
&= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_X(x_i) \cdot p_Y(y_j) \cdot \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_X(x_i) \cdot p_Y(y_j) \cdot \log \frac{1}{p_Y(y_j)} \\
&= \sum_{i=0}^{M_X-1} p_X(x_i) \cdot \log \frac{1}{p_X(x_i)} + \sum_{j=0}^{M_Y-1} p_Y(y_j) \cdot \log \frac{1}{p_Y(y_j)} \\
&= H(X) + H(Y).
\end{aligned}$$

Como veremos con más detalle en la siguiente sección, conviene remarcar que esta relación se cumple únicamente bajo la hipótesis de independencia entre las variables aleatorias.

5.3.3. Entropía condicional

El caso en que las variables aleatorias son independientes es en el que la combinación de variables aleatorias produce mayor entropía, ya que si ambas variables no fuesen independientes, el conocimiento del valor de una de ellas eliminaría incertidumbre sobre el valor de la otra. Para medir esa incertidumbre se define la *entropía condicional* de dos variables aleatorias X e Y, $H(X|Y)$, que promedia el valor de la entropía condicional de X dado Y sobre todos los valores del alfabeto de Y, de forma que se define como

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) \cdot H(X|Y = y_j).$$

Desarrollando esta expresión para la entropía $H(X|Y=y_j)$ a partir de la distribución condicional de X dada Y, $p_{X|Y}(x_i|y_j)$, se llega a la expresión equivalente

$$\begin{aligned} H(X|Y) &= \sum_{j=0}^{M_Y-1} p_Y(y_j) \sum_{i=0}^{M_X-1} p_{X|Y}(x_i|y_j) \cdot \log \frac{1}{p_{X|Y}(x_i|y_j)} \\ &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_{X|Y}(x_i|y_j)}. \end{aligned}$$

De acuerdo con la regla de Bayes, esta probabilidad cumple la relación

$$p_{X|Y}(x_i|y_j) \cdot p_Y(y_j) = p_{X,Y}(x_i,y_j).$$

En general, se puede extender de forma natural esta definición cuando el condicionamiento es con respecto a varias variable aleatorias

$$\begin{aligned} H(X_N|X_1,X_2,\cdots,X_{N-1}) = \\ - \sum_{x_1,x_2,\cdots,x_N} p_{\boldsymbol{X}}(x_1,x_2,\cdots,x_N) \log p_{X_N|X_1,X_2,\cdots,X_{N-1}}(x_N|x_1,x_2,\cdots,x_{N-1}). \end{aligned}$$

La entropía condicional puede interpretarse como una medida de la incertidumbre que se tiene sobre el valor de una variable aleatoria, X, cuando se conoce el valor de otra variable aleatoria, Y. O de otra forma, la comparación entre $H(X)$ y $H(X|Y)$ cuantifica la información que el conocimiento de Y aporta sobre el conocimiento de X. Cuando las variables aleatorias X e Y son independientes, el conocimiento del valor de una de ellas no aporta conocimiento sobre la otra y por tanto no elimina incertidumbre sobre su valor. Por tanto, en este caso

$$H(X|Y) = H(X).$$

Por el contrario, si el conocimiento de Y determina por completo el valor de X, conocido el valor de Y no hay ninguna incertidumbre sobre el valor de X, y la entropía condicional sería $H(X|Y)=0$.

La entropía conjunta se relaciona con la entropía y con la entropía condicional a través de la siguiente expresión

$$\begin{aligned} H(X,Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_{X,Y}(x_i,y_j)} \\ &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_X(x_i) \cdot p_{Y|X}(y_j|x_i)} \\ &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_{Y|X}(y_j|x_i)} \\ &= \sum_{i=0}^{M_X-1} p_X(x_i) \cdot \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i,y_j) \cdot \log \frac{1}{p_{Y|X}(y_j|x_i)} \\ &= H(X) + H(Y|X). \end{aligned}$$

A través de un desarrollo equivalente, es fácil demostrar que también se cumple la relación

$$H(X,Y) = H(Y) + H(X|Y).$$

Esto no quiere decir que $H(X|Y)$ sea igual a $H(Y|X)$ en general. Esta relación sólo se cumple cuando $H(X) = H(Y)$.

Es decir, que la entropía conjunta se obtiene como la suma de la entropía de una variable aleatoria más la de la otra condicionada a la primera. Se suma por tanto la incertidumbre de una de las variables aleatorias con la que le queda a la otra cuando se conoce la primera. Esto significa, como ya se ha visto anteriormente, que la entropía conjunta sólo será igual a la suma de la entropía de cada una de las variables aleatorias cuando las variables aleatorias sean independientes.

En general, se puede extender esta relación al caso de un mayor número de variables aleatorias, en cuyo caso aplicando la regla de la cadena se tiene la siguiente relación general

$$H(\boldsymbol{X}) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \cdots + H(X_N|X_1, X_2, \cdots, X_{N-1}).$$

Cuando $(X_1, X_2, \cdots, X_N)$ son variables aleatorias independientes, la entropía conjunta de todas ellas será la suma de la entropía de cada una de las variables aleatorias

$$H(\boldsymbol{X}) = \sum_{i=1}^{N} H(X_i).$$

5.3.4. Información mutua

La entropía representa una medida de incertidumbre sobre el valor de una o varias variables aleatorias. Otro concepto que podríamos definir como "complementario" es la denominada *información mutua* entre dos variables aleatorias X e Y, que se denota $I(X, Y)$. La información mutua representa la información que aporta Y sobre el conocimiento de X. La definición formal a partir de las distribuciones marginales de cada variable aleatoria y de su distribución conjunta es

$$I(X,Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i) \cdot p_Y(y_j)},$$

y se mide en bits.

La información mutua es una medida no negativa, $I(X, Y) \geq 0$, que se puede expresar en términos de la entropía y de la entropía condicional, ya que se cumple la siguiente relación:

$$\begin{aligned}
I(X,Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i) \cdot p_Y(y_j)} \\
&= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log \frac{p_{X|Y}(x_i, y_j)}{p_X(x_i)} \\
&= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log \frac{1}{p_X(x_i)} + \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log(p_{X|Y}(x_i, y_j)) \\
&= \sum_{i=0}^{M_X-1} p_X(x_i) \cdot \log \frac{1}{p_X(x_i)} - \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \cdot \log \frac{1}{p_{X|Y}(x_i, y_j)} \\
&= H(X) - H(X|Y).
\end{aligned}$$

Un desarrollo equivalente demuestra también que

$$I(X,Y) = H(Y) - H(Y|X).$$

Por otro lado, de la propia definición de la información mutua se tiene la propiedad

$$I(X,Y) = I(Y,X),$$

y teniendo en cuenta la relación entre entropías marginal, condicional y conjunta

$$H(X,Y) = H(Y) + H(X|Y),$$

se comprueba de forma trivial que la información mutua se puede también obtener como

$$I(X,Y) = H(X) + H(Y) - H(X,Y).$$

Estas relaciones entre información mutua y las distintas entropías se representan habitualmente de forma gráfica mediante un *diagrama de Venn*, como el que se muestra en la Figura 5.20.

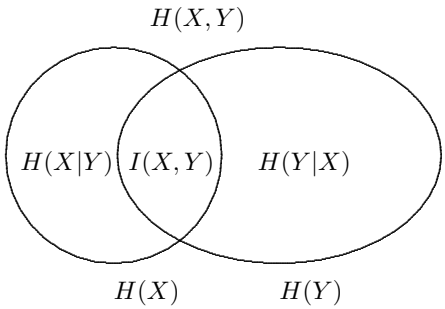

Figura 5.20: Diagrama de Venn que ilustra las relaciones entre entropías e información mutua.

Cada uno de los conjuntos representa la entropía de cada variable, $H(X)$ y $H(Y)$, y la intersección entre ellos es la información mutua $I(X,Y)$, mientras que la unión es la entropía conjunta, $H(X,Y)$. La diferencia entre el conjunto completo y la intersección es la entropía condicional. Si las variables aleatorias son independientes, los conjuntos tendrían intersección nula. Si ambas variables son iguales, la intersección es completa y la entropía condicional es nula. La Figura 5.21 muestra con más claridad como se identifica cada una de las medidas sobre el diagrama de Venn con un secillo código de colores.

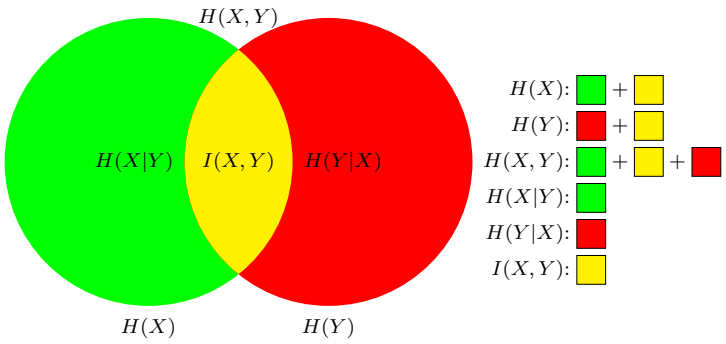

Figura 5.21: Identificación con un sencillo código de colores de las distintas entropías y la información mutua sobre el diagrama de Venn.

A continuación se muestra un ejemplo en el que se calculan las distintas entropías y la información mutua para dos variables aleatorias sencillas.

Ejemplo

Se tienen dos variables aleatorias binarias, X e Y, con identico alfabeto $x_0 = y_0 = 0$, $x_1 = y_1 = 1$, y con la siguiente distribución conjunta

$$p_{X,Y}(0,0) = \frac{1}{3},\ p_{X,Y}(0,1) = \frac{1}{3},\ p_{X,Y}(1,0) = \frac{1}{3},\ p_{X,Y}(1,1) = 0.$$

Para calcular la entropía de cada variable aleatoria hay que conocer las distribuciones marginales, que se obtienen de forma sencilla a partir de la distribución conjunta

$$p_X(x_i) = \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \text{ y } p_Y(y_j) = \sum_{i=0}^{M_X-1} p_{X,Y}(x_i, y_j).$$

En este caso

$$p_X(0) = p_Y(0) = \frac{2}{3},\ p_X(1) = 0_Y(1) = \frac{1}{3}.$$

Por tanto, como en este caso las dos variables tienen la misma distribución, la entropía de ambas variables, parametrizada a través de la función de entropía binaria (al ser las variables aleatorias binarias)

$$H(X) = H(Y) = H_b\left(\frac{2}{3}\right) = H_b\left(\frac{1}{3}\right) = 0{,}919.$$

La entropía conjunta, aplicando su definición, vendrá dada por

$$H(X,Y) = 3 \times \left(\frac{1}{3}\log_2(3)\right) + 0\log_2(0) = \log_2(3) = 1{,}585.$$

Se puede interpretar también este resultado como que (X,Y) es un vector de variables aleatorias con un alfabeto de tres eventos, $(0,0)$, $(0,1)$ y $(1,0)$, todos ellos igualmente probables.

A partir de los resultados anteriores, la entropía condicional se puede obtener a través de la relación

$$H(X|Y) = H(X,Y) - H(Y) = 1{,}585 - 0{,}919 = 0{,}666.$$

Del mismo modo, la información mutua podría obtenerse, por ejemplo, a través de la relación

$$I(X,Y) = H(X) - H(X|Y) = 0{,}919 - 0{,}666 = 0{,}253.$$

La información mutua entre variables aleatorias discretas tiene una serie de propiedades que conviene tener en cuenta. Entre ellas cabe destacar las siguientes:

1. Es siempre no negativa
$$I(X,Y) = I(Y,X) \geq 0.$$
El valor mínimo $I(X,Y) = 0$ se obtiene en el caso en que X e Y son independientes.

2. Su valor máximo está acotado por el valor de la entropía de cada una de las variables aleatorias, con lo que en la práctica está acotado por el valor mínimo de la entropía de las variables aleatorias
$$I(X,Y) \leq \text{mín}(H(X), H(Y)).$$
La información mutua nunca puede ser mayor que la medida de información que tiene cada una de las variables.

3. Se puede definir información mutua condicional como el promedio de la información mutua condicionada a cada uno de los posibles valores de la variable aleatoria con respecto a la que se condiciona
$$I(X,Y|Z) = \sum_{i=0}^{M_z-1} p_Z(z_i) \cdot I(X,Y|Z = z_i).$$

4. La información mutua condicional $I(X,Y|Z)$ también se puede obtener a través de las entropías condicionales como
$$I(X,Y|Z) = H(X|Z) - H(X|Y,Z).$$

5. La regla de la cadena para la información mutua se define a partir de la relación
$$I((X,Y),Z) = I(X,Z) + I(Y,Z|X).$$

6. En general, la regla de la cadena es
$$I((X_1,X_2,\cdots,X_N),Y) = I(X_1,Y) + I(X_2,Y|X_1) + \cdots + I(X_N,Y|X_1,\cdots,X_{N-1}).$$

7. A partir de la definición de información mutua se obtiene la definición de entropía como información mutua de una variable aleatoria consigo misma. Esta relación se demuestra de forma sencilla teniendo en cuenta que la distribución de una variable aleatoria consigo misma toma la forma
$$p_{X,X}(x_i,x_j) = \delta[i-j]\cdot p_X(x_i),$$
por lo que la información mutua de la variable aleatoria X consigo misma es
$$\begin{aligned} I(X,X) &= \sum_{i=0}^{M_X-1}\sum_{j=0}^{M_X-1} \delta[i-j]\cdot p_X(x_i)\cdot \log\frac{\delta[i-j]\cdot p_X(x_i)}{p_X(x_i)\cdot p_X(x_j)} \\ &= \sum_{i=0}^{M_X-1} p_X(x_i)\cdot \log\frac{p_X(x_i)}{p_X(x_i)\cdot p_X(x_i)} \\ &= \sum_{i=0}^{M_X-1} p_X(x_i)\cdot \log\frac{1}{p_X(x_i)} \\ &= H(X). \end{aligned}$$
De aquí es de donde proviene el nombre de *auto-información* (la entropía es el promedio de la auto-información), y a la entropía en ocasiones se la conoce con este nombre.

5.3.5. Entropía diferencial e información mutua

Hasta ahora, las medidas que se han presentado hacen referencia a variables aleatorias discretas, lo que sirve para modelar fuentes de información discretas en el tiempo y con un alfabeto discreto. Para estas variables se han presentado la entropía, $H(X)$ y la información mutua $I(X,Y)$, así como las entropías condicional y conjunta, $H(X|Y)$ y $H(X,Y)$.

Para modelar una fuente discreta en el tiempo pero con alfabeto continuo, por ejemplo una fuente de audio muestreada, es preciso utilizar una variable aleatoria continua. En este caso, el análogo de la entropía para variables aleatorias discretas se llama *entropía diferencial*. Sin embargo, esta medida no tiene el significado intuitivo que tenía la entropía, lo que se debe a varios aspectos, como por ejemplo el hecho de que en una variable aleatoria continua, por definición la probabilidad de un número continuo concreto es nula.

Formalmente, la entropía diferencial de una variable aleatoria continua X, con una función densidad de probabilidad $f_X(x)$, se define como
$$h(X) = \int_{-\infty}^{\infty} f_X(x)\ \log\frac{1}{f_X(x)}\ dx,$$

donde de nuevo se considera $0 \log(1/0) = 0$.

Ejemplo

Se desea determinar la entropía diferencial de una variable aleatoria uniformemente distribuida en un intervalo $[0, a]$.

Utilizando directamente la definición de entropía diferencial, y teniendo en cuenta que la función densidad de probabilidad de la variable aleatoria vale $1/a$ entre 0 y a, la entropía vale

$$h(X) = \int_0^a \frac{1}{a} \log(a)\, dx = \log(a).$$

A partir de este ejemplo se pueden observar algunas propiedades interesantes:

1. Para $a < 1$ se tiene $h(X) < 0$, lo que va en contra de la propiedad de no negatividad de la entropía de una variable aleatoria discreta.
2. Para $a = 1$, se tiene $h(X) = 0$, y en este caso X no es determinista, con lo que tiene un cierto grado de incertidumbre. Esto va también en contra de las propiedades de la entropía para variables aleatorias discretas.

Ejemplo

Se tiene una variable aleatoria X con una función densidad de probabilidad gausiana con media nula y varianza σ^2

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Utilizando logaritmos neperianos, la entropía diferencial para esta variable aleatoria en nats se calcula como

$$\begin{aligned} h(X) =& -\int_{-\infty}^{\infty} f_X(x)\, \ln \frac{1}{\sqrt{2\pi\sigma^2}}\, dx - \int_{-\infty}^{\infty} f_X(x)\, \ln\left(e^{-\frac{x^2}{2\sigma^2}}\right)\, dx \\ =& \ln(\sqrt{2\pi\sigma^2}) + \frac{\sigma^2}{2\sigma^2} = \ln(\sqrt{2\pi\sigma^2}) + \frac{1}{2} \\ =& \frac{1}{2} \ln(2\pi e\sigma^2) \text{ nats.} \end{aligned}$$

Para llegar a esta expresión se han utilizado las siguientes propiedades sobre la distribución gausiana

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \text{ y } \int_{-\infty}^{\infty} x^2\, f_X(x)\, dx = \sigma^2.$$

Cambiando la base del logaritmo a 2, se tiene la entropía diferencial en bits

$$h(X) = \frac{1}{2} \log_2(2\pi e\sigma^2) \text{ bits.}$$

Dependiendo del valor de la varianza, en particular comparándolo con un valor $\sigma^2 = \frac{1}{2\pi e}$, esta entropía puede tomar valores positivos, negativos o nulo.

Al igual que para variables aleatorias discretas, también se definen para variables aleatorias continuas entropías conjuntas y condicionales. Se define la *entropía diferencial conjunta* entre dos variables aleatorias X e Y como

$$h(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{1}{f_{X,Y}(x,y)} dx\, dy.$$

En cuanto a la *entropía diferencial condicional*, su definición es

$$h(X|Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{1}{f_{X|Y}(x|y)} dx\ dy.$$

A menudo se utiliza la definición alternativa pero equivalente

$$h(X|Y) = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{\infty} f_{X|Y}(x|y) \log \frac{1}{f_{X|Y}(x|y)} dx\ dy.$$

Se puede observar que son las extensiones naturales de las definiciones para variables aleatorias discretas. Por tanto se cumplen las mismas relaciones. En concreto,

$$h(X,Y) = h(Y) + h(X|Y).$$

Del mismo modo se define también la *información mutua* para variables aleatorias continuas como

$$I(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dx\ dy,$$

que también se mide en bits, y donde para evitar ambigüedades se define $0 \log \frac{0}{0} = 0$.

Como en el caso de variables aleatorias discretas, la información mutua se puede expresar en función de las entropías

$$I(X,Y) = h(Y) - h(Y|X) = h(X) - h(X|Y) = h(X) + h(Y) - h(X,Y).$$

Al contrario que en el caso de la entropía diferencial para variables aleatorias continuas, donde no se mantiene la interpretación intuitiva de entropía para variables aleatorias discretas como medida de incertidumbre o de información, para la información mutua si se mantiene y tiene el mismo significado; es decir, la información mutua indica el conocimiento que aporta una variable sobre la otra. Además se mantienen la mayoría de propiedades básicas de la información mutua para variables aleatorias discretas. En particular, dada su definición se cumplen las siguientes propiedades:

1. $I(X,Y) \geq 0$, es decir, es una función no negativa.
2. $I(X,Y) = 0$ sólo si las variables X e Y son independientes.
3. $I(Y,X) = I(X,Y)$.

5.4. Capacidad de canal

Una vez definidos los conceptos de entropía e información mutua, en esta sección se va a tratar de determinar la cantidad máxima de información que se puede transmitir por un canal haciendo uso de esas medidas de información. En primer lugar se introducirá el concepto de codificación de canal como mecanismo para lograr una comunicación fiable a través de un cierto canal no fiable, para a continuación definir la capacidad de un canal y estudiar cómo se obtiene dicho valor para dos tipos de canales: el canal digital modelado mediante un DMC y el canal gausiano.

5.4.1. Codificación de canal para una transmisión fiable

El principal objetivo cuando se transmite información sobre cualquier canal de comunicaciones es la *fiabilidad*. Esta fiabilidad en sistemas de comunicaciones digitales se mide mediante la probabilidad de error en el receptor. Como en todo canal de comunicaciones, aparte de las posibles distorsiones, se introduce ruido. A primera vista puede parecer que la probabilidad de error siempre estará acotada por un valor no nulo que dependerá del nivel de ruido, es decir

$$P_e \geq K = f(\sigma^2),$$

donde σ^2 es la potencia del ruido a la entrada del receptor. Sin embargo, como ya se esbozó en la introducción, un resultado fundamental de la teoría de la información dice que la transmisión fiable, entendiéndose por transmisión fiable aquella en la que se tiene una probabilidad de error por debajo de cualquier límite fijado, es posible incluso en canales ruidosos siempre que la velocidad de transmisión o tasa de transmisión esté por debajo de un determinado valor denominado *capacidad de canal*. Este resultado fue presentado por Shannon en 1948 y es lo que se denomina *teorema de codificación de canal*, o en inglés *noisy channel-coding theorem*. A modo de resumen, lo que dice el teorema de codificación de canal es que *la limitación básica que introduce el ruido en un canal de comunicaciones no está en la fiabilidad de la comunicación, sino en la velocidad de la misma*. Así, será posible obtener una comunicación con una probabilidad de error arbitrariamente baja siempre y cuando se transmita información a una tasa por debajo de un valor límite que depedenderá de las características del canal.

En primer lugar se va a ver cómo es posible obtener este límite de una forma intuitiva, siguiendo la misma línea argumental que se sigue en [Proakis y Salehi, 2002], para luego formularlo en términos de la teoría de la información.

La Figura 5.22 muestra un canal discreto sin memoria (DMC) con un alfabeto de entrada formado por cuatro elementos, $\mathcal{A}_X = \{a, b, c, d\}$ y el mismo alfabeto de salida $\mathcal{A}_Y = \mathcal{A}_X$.

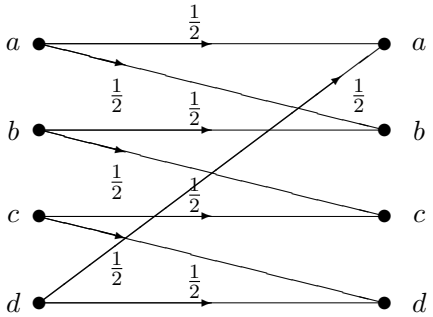

Figura 5.22: Ejemplo de un canal discreto sin memoria.

Dado que no se trata de un canal ideal, si se utiliza este canal para la transmisión en la forma habitual, el receptor no podrá identificar con probabilidad de error nula el símbolo transmitido a la vista del símbolo recibido. Cuando en la salida se tiene el valor a, el receptor no puede discernir si se ha transmitido a o b, ya que la transmisión de ambos puede producir a la salida del canal el símbolo a. Lo mismo ocurre con el resto de posibles valores de salida. Si se observa b, no es posible discernir sin error si se debe a la transmisión de a o de b, y así sucesivamente. Por tanto, hay una probabilidad de error que viene fijada por las características del canal, en este caso por las probabilidades de transición que lo definen.

Sin embargo, dadas las especiales características de este canal, va a ser posible transmitir información sin errores a través del mismo. Es evidente que la imposibilidad de discernir sin error el símbolo transmitido a la vista del valor de salida se debe a que los conjuntos que forman los posibles valores de salida asociados a la transmisión de cada símbolo se "*solapan*". Así, existe un solapamiento entre las salidas que se pueden dar cuando se transmite a (que pueden ser a y b) y cuando se transmite b (que pueden ser b y c). Por eso cuando se observa b en la salida no es posible saber con absoluta certeza qué símbolo se ha transmitido. Pero para este canal es posible elegir un subconjunto de elementos del alfabeto de entrada cuyas salidas no se solapen. Por ejemplo, si sólo se transmiten los símbolos a y c, a la vista de la salida no habrá ambigüedad posible sobre el símbolo transmitido; si en la salida se tienen a o b, se sabe con certeza que el símbolo transmitido es a; lo mismo sucede si en la salida se tienen c o d, en cuyo caso se tiene la certeza de que se ha transmitido c. Por tanto, transmitiendo ese subconjunto de posibles valores de X la probabilidad de error es nula. El precio que se paga por esta fiabilidad es la tasa de transmisión. Transmitiendo los 4 símbolos que forman parte del alfabeto de X en cada uso del canal se transmiten dos bits de información. Sin embargo, si sólo se transmiten 2 posibles valores, en cada uso del canal se transmitirá únicamente un bit de información (recuerde que el número de bits de información por símbolo transmitido es $\log_2 M$, siendo M el número de elementos del alfabeto de símbolos que se transmite).

El mecanismo utilizado en este ejemplo para poder transmitir con probabilidad de error nula ilustra la idea fundamental subyacente en el teorema de codificación de canal para poder transmitir de forma fiable: utilizar en la transmisión sólo aquellos símbolos cuyas salidas correspondientes sean disjuntas. Aquí es preciso matizar que el objetivo de la codificación de canal no es en realidad conseguir una transmisión con probabilidad de error nula, sino con una probabilidad de error por debajo de un cierto valor, que puede ser arbitrariamente bajo. Así, para conseguir una transmisión, si no con probabilidad de error nula, con una probabilidad de error arbitrariamente baja, se podrán utilizar en la transmisión sólo aquellos símbolos que, si no tienen salidas asociadas disjuntas, tengan salidas cuyo solapamiento se produzca con una probabilidad suficientemente baja.

Un problema que surge con esta idea para la transmisión con probabilidad de error arbitrariamente baja, es que en la práctica los canales reales no tienen un comportamiento como el de la Figura 5.22, donde hay símbolos de entrada cuyas salidas no se solapan; y en la gran mayoría de los casos ni siquiera hay un subconjunto de símbolos cuyas salidas se solapen con una probabilidad suficientemente baja. Sin embargo, la teoría de la codificación de canal propone un sencillo mecanismo para generar artificialmente una situación similar a esta. Aunque este mecanismo se puede utilizar para canales digitales con alfabetos M-ários, por simplicidad se ilustrará este mecanismo poniendo como ejemplo un caso binario; en concreto el modelo de canal binario simétrico o BSC, en el que se denotará mediante ε la probabilidad de error de bit (BER), como se muestra en el diagrama de rejilla de la izquierda en la Figura 5.23.

Si se observa el canal BSC y se intenta aplicar el procedimiento aplicado sobre el canal de la Figura 5.22, se ve que directamente no es posible, ya que las salidas de los dos símbolos se solapan por completo con una probabilidad arbitraria (dada por ε) y porque además sólo hay dos símbolos (si sólo se transmite uno de ellos, no habrá en realidad información en la transmisión). En general, no va a ser posible aplicar ese procedimiento de forma directa para casi ningún canal DMC real. Para utilizar esta idea, lo que se hace es aplicarla no directamente sobre el canal sino sobre el denominado *canal extendido* de orden n. El canal extendido de orden n se define como un canal en el que se agrupan bloques de n símbolos para formar símbolos extendidos (de orden n), dando lugar a unos alfabetos de entrada $\mathcal{A}_X^n$ y $\mathcal{A}_Y^n$. La idea es transmitir la información no en cada uso individual del canal, sino de forma conjunta en n usos del canal. El diagrama de la derecha de la Figura 5.23 ilustra la idea de la extensión de orden $n = 3$ para el BSC. La información se

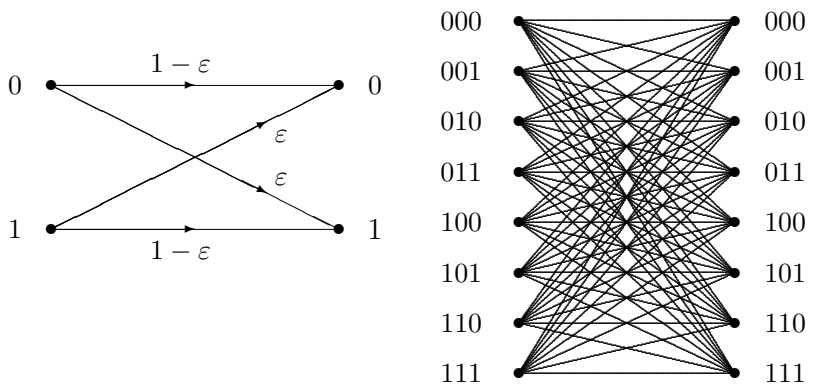

Figura 5.23: Canal binario simétrico (BSC) y su correspondiente canal extendido de orden $n = 3$.

transmitirá por bloques haciendo 3 usos del canal, de forma que el alfabeto de este canal extendido está ahora formado por 8 posibles símbolos extendidos, correspondientes a los 8 posibles valores que pueden tomar los símbolos del canal original en los 3 usos que definen el canal extendido. Se ha pasa por tanto de trabajar con un sistema con un alfabeto de dos símbolos

$$\mathcal{A}_X = \mathcal{A}_Y = \{0, 1\},$$

a un sistema extendido con alfabeto de 2^n símbolos, en este caso $2^3 = 8$ símbolos

$$\mathcal{A}_X^3 = \mathcal{A}_Y^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}.$$

Las probabilidades de transición sobre este canal extendido se obtienen como productos de las n probabilidades de transición a través del canal original asociadas a cada caso sobre el canal extendido. Si se definen los símbolos extendidos como vectores de n elementos, representados estadísticamente por las variables aleatorias vectoriales $\boldsymbol{X}$ y $\boldsymbol{Y}$, con alfabetos

$$\boldsymbol{X} \in \{\boldsymbol{x}_i\}_{i=0}^{M_X^n-1},\ \boldsymbol{Y} \in \{\boldsymbol{y}_j\}_{j=0}^{M_Y^n-1}$$

con

$$\boldsymbol{x}_i = [x_i[0], x_i[1], \cdots, x_i[n-1]]\,,\ \boldsymbol{y}_j = [y_j[0], y_j[1], \cdots, y_j[n-1]]$$

las probabilidades de transición sobre los símbolos extendidos son

$$p_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}_j|\boldsymbol{x}_i) = \prod_{\ell=0}^{n-1} p_{Y|X}(y_j[\ell]|x_i[\ell]).$$

En la figura no se han etiquetado las ramas con las probabilidades de transición por falta de espacio en la misma, pero dichas probabilidades se obtienen de forma muy sencilla, habiendo 4 posibles valores:

- Si el símbolo extendido de entrada y el de salida coinciden, la probabilidad de transición asociada es
$$p_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i) = (1-\varepsilon)^3.$$
- Si entre el símbolo extendido de entrada y el de salida la distancia de Hamming es 1 (cambia sólo un bit), la probabilidad de transición asociada es
$$p_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}_j|\boldsymbol{x}_i) = \varepsilon \cdot (1-\varepsilon)^2.$$

- Si entre el símbolo extendido de entrada y el de salida la distancia de Hamming es 2 (cambian dos bits), la probabilidad de transición asociada es

$$p_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}_j|\boldsymbol{x}_i) = \varepsilon^2 \cdot (1-\varepsilon).$$

- Por último, si entre el símbolo extendido de entrada y el de salida la distancia de Hamming es 3 (cambian los tres bits), la probabilidad de transición asociada es

$$p_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}_j|\boldsymbol{x}_i) = \varepsilon^3.$$

Básicamente, la probabilidad de que un bit coincida entre entrada y salida es $1-\varepsilon$, y la de que sea distinta es ε, de donde de forma directa se obtienen estas cuatro probabilidades.

Evidentemente, para probabilidades de error de bit ε bajas, las dos primeras probablidades son mucho mayores que las dos últimas (es mucho más probable tener cero o un error sobre los tres bits, que tener dos o tres errores). Se podrían por tanto dividir las transiciones en transiciones de alta probabilidad y de baja probabilidad. Esta división se muestra en la Figura 5.24 mediante un código de colores: en negro se representan las transiciones de alta probabilidad y en verde las de baja probabilidad.

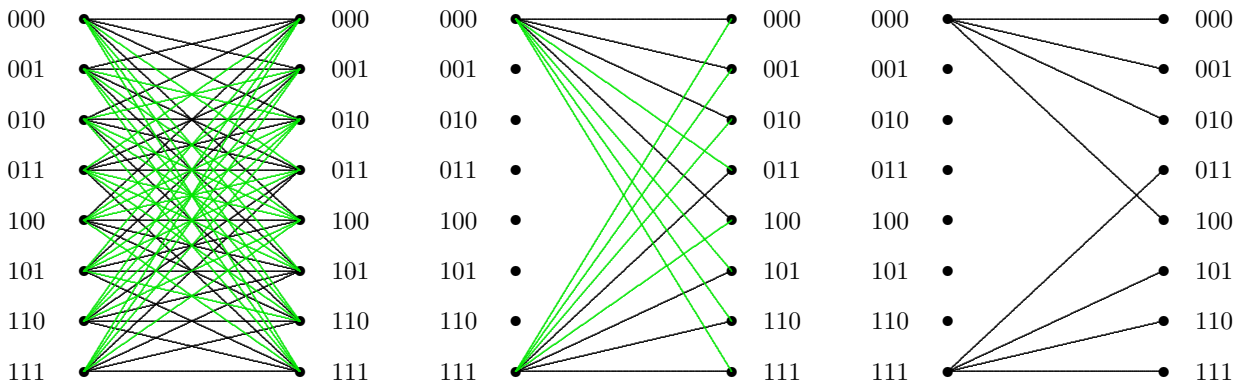

Figura 5.24: Distinción entre transiciones de alta (negro) y de baja (verde) probabilidad y selección de un subconjunto de símbolos extendidos sin solapamiento en los enlaces e alta probabildad.

Hecha esa distinción, ahora es posible buscar un subconjunto de símbolos extendidos a la entrada cuyas salidas sean disjuntas en cuanto a las transiciones de alta probabilidad, lo que sucede por ejemplo con los símbolos extendidos 000 y 111, como se ilustra en la figura. Ahora, cuando se transmite uno de los dos símbolos extendidos, si se producen sólo las transiciones más probables, a partir de la salida se identificará correctamente el símbolo extendido transmitido. Obviamente, también se puede producir alguna transición de baja probabilidad, en cuyo caso se producirá un error al identificar el símbolo extendido transmitido. En ese caso, la probabilidad de error vendrá dada por la probabilidad de que se produzca alguna de las transiciones poco probables, que para este ejemplo será

$$P_e = 1 \times (1-\varepsilon)^3 + 3 \times \varepsilon \cdot (1-\varepsilon)^2.$$

Este valor se obtiene teniendo en cuenta que las transiciones de baja probabilidad son las que correspondientes a dos o tres errores de bit en la transmisión, y que sólo hay un patrón de tres errores y tres de dos errores para cada bloque de tres bits transmitidos.

Como ejemplo numérico para ilustrar el beneficio obtenido, si un sistema de comunicaciones binario tiene una BER $\varepsilon = 0{,}1$, uno de cada 10 bits será erróneo (porcentaje del 10 %). Usando

como base ese sistema y una extensión como la del ejemplo de orden $n = 3$, la probabilidad de error pasará a ser $P_e = 0{,}028$; es decir, se puede conseguir un porcentaje del 2,8 % de bits erróneos usando como base para la transmisión un sistema con un porcentaje de error del 10 %. Para el caso $\varepsilon = 0{,}01$ el porcentaje de bits erróneos del sistema será del 1 %. En ese caso $P_e = 2{,}98 \times 10^{-4}$, lo que significa un porcentaje de error de aproximadamente el 0,03 %.

Esta técnica permite reducir la probabilidad de error de un sistema, pero naturalmente lo hace a costa de algo. Y ese algo es la tasa de transmisión de información. Si se usa el sistema sin la extensión, en cada uso del canal se transmite un bit de información. Si se usa el sistema extendido, como sólo se transmiten 2 de los 8 posibles símbolos extendidos, cada uno de ellos transportará un bit de información real, por ejemplo 000 transportará el "0" y 111 transportará el "1". De esta forma, en cada 3 usos del canal se enviará un sólo bit de información, con lo que la tasa efectiva, que sería 1/3 (1 bit de información por cada 3 usos del canal), ha disminuido respecto a la transmisión directa sin la extensión.

La elección del subconjunto de elementos sin solapamientos en transiciones de baja probabilidad no tiene por qué restringirse a dos símbolos extendidos. Sería por ejemplo posible hacer una extensión de orden $n = 5$, y elegir 4 símbolos extendidos cuyas transiciones de alta probabilidad (definidas como aquellas en las quese producen cero errores o un error de bit sobre los cinco bits transmitidos) no se solapen; podrían ser por ejemplo los símbolos 00000, 10101, 01110 y 11011. En este caso, al haber 4 símbolos extendidos, cada 5 usos del canal se enviarán 2 bits de información real, por lo que la tasa de transmisión efectiva de información será 2/5.

Codificación de canal

En general a esta técnica basada en la definición de símbolos extendidos de orden n que se utiliza para poder realizar una transmisión con una probabilidad de error suficientemente baja sobre un sistema que inherentemente tiene una probabilidad de error mayor se le denomina *codificación de canal*. La técnica no está limitada a su utilización sobre sistemas binarios, como en los ejemplos que se han utilizado para su presentación, sino que se puede emplear también sobre sistemas M-ários. La Figura 5.25 ilustra el funcionamiento de un sistema que utiliza la codificación de canal como mecanismo para controlar la probabilidad de error del sistema.

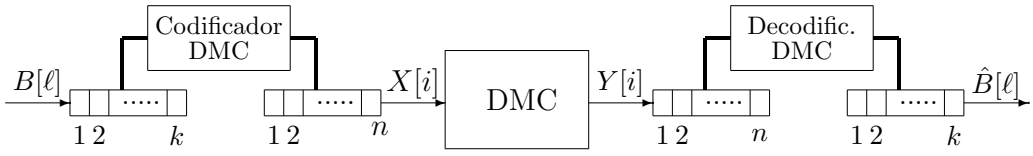

Figura 5.25: Codificador y decodificador de canal para la transmisión sobre un canal digital modelado mediante un canal discreto sin memoria.

Un sistema de codificación de canal está compuesto por un codificador en el lado del transmisor, que lleva a cabo la *codificación de canal* y un decodificador en el lado del receptor, que lleva a cabo la *decodificación de canal*. El codificador tiene como entrada un vector de k símbolos M-ários, que define un alfabeto de entrada, formalmente denominado *conjunto de índices*, de M^k elementos. Este conjunto de índices estaría compuesto en los ejemplos anteriores por los bloques de información efectiva que se va a transmitir en cada n usos del canal. Ante cada entrada, el codificador genera como salida un vector de n símbolos de entrada al DMC, haciendo corresponder a cada una de las combinaciones de k símbolos M-ários que forman el conjunto de índices, una combinación de

n símbolos del DMC ó *palabra código*. En los ejemplos anteriores, estas palabras código serían el subconjunto de símbolos extendidos cuyas salidas se solapan con baja probabilidad. En estas condiciones, se dice que este es un código $\mathcal{C}(k, n)$, donde al valor n se le suele llamar longitud del código. El conjunto de todas las palabra código (M^k), se denomina *diccionario del código*. El decodificador funciona de forma recíproca. Formalmente un código queda definido por el conjunto de índices y las funciones de codificación y decodificación. Los diccionarios del código para los dos ejemplos anteriores, que en ambos casos eran códigos binarios, serían los que se muestran en la Tabla 5.1.

Conjunto de índices	Palabras código
0	000
1	111

Código de ejemplo $\mathcal{C}(1, 3)$

Conjunto de índices	Palabras código
00	00000
01	10101
10	01110
11	11011

Código de ejemplo $\mathcal{C}(2, 5)$

Tabla 5.1: Diccionarios del código para los dos ejemplos binarios que se han utilizado para ilustrar la técnica de codificación de canal.

Un parámetro muy importante en todo código de canal es el que define el cociente entre el número de símbolos a la entrada del codificador, k, y el número de veces que se utiliza el DMC para transmitir la palabra código, n, ya que define la cantidad de información real que se transmite en cada uso de canal. Este parámetro se denomina *tasa de transmisión* o simplemente *tasa del código*. Habitualmente se denota por R y se mide en símbolos por uso del DMC

$$R = \frac{k}{n} \text{ símbolos por uso.}$$

Para sistemas binarios, las unidades de esta tasa del código son bits por uso del canal.

Teorema de codificación de canal

Al presentar el principio de la codificación de canal se ha visto que hay un compromiso entre la tasa de transmisión o tasa del código y las prestaciones obtenidas. De forma natural surgen algunas preguntas respecto a las prestaciones que se pueden obtener con esta técnica. ¿Es posible reducir la probabilidad de error tanto como se desee utilizando esta técnica sobre cualquier tipo de canal? ¿Qué limitaciones imponen las características del canal sobre las prestaciones que puede ofrecer la capacidad de canal? La respuesta a estas preguntas está en el denominado *teorema de codicicación de canal*, presentado por Claude Shannon en 1948.

El teorema de codificación de canal demuestra que existe un límite para la máxima tasa a la que se puede transmitir por un canal discreto sin memoria, que es el que se denomina *capacidad de canal*, y que se obtiene formalmente como el valor máximo, sobre todas las posibles distribuciones para el alfabeto de entrada del canal, de la información mutua entre la entrada y la salida del canal, es decir

$$C = \max_{p_X(x_i)} I(X, Y),$$

donde $I(X, Y)$ es la información mutua entre la entrada X y la salida Y del canal. El teorema además demuestra los siguientes aspectos:

1. Si la tasa de transmisión R es menor que la capacidad del canal C, entonces para cualquier valor arbitrariamante bajo $\delta > 0$ existe un código con una longitud de bloque n suficientemente larga cuya probabilidad de error es menor que δ.

2. Si $R > C$, la probabilidad de error de cualquier código con cualquier longitud de bloque está limitada por un valor no nulo que depende de las características del canal.

3. Existen códigos que permiten alcanzar la capacidad del canal $R = C$.

Es importante hacer aquí una aclaración sobre el tercer punto. El teorema, aunque demuestra que es posible alcanzar dicha capacidad, no responde a la pregunta de cómo se pueden obtener en la práctica dichos códigos. En un problema práctico no se va a alcanzar dicha capacidad, sino que en general se utilizarán códigos que estarán por debajo de la misma.

En la siguiente sección se estudiará cómo calcular la capacidad de canal en primer lugar para un canal digital, y a continuación para el canal gausiano. El diseño y análisis de códigos de canal prácticos no entra dentro del ámbito de esta asignatura, sino que se estudiarán en la asignatura "*Comunicaciones Digitales*".

5.4.2. Capacidad de canal para el canal digital

En primer lugar se va a estudiar el cado el caso del canal digital binario, en el que se agrupan n bits para formar los nuevos símbolos extendidos que se agruparán a la entrada y salida del sistema para implementar la codificación de canal. Para este caso se va a obtener la máxima cantidad de información que se podrá transmitir de forma fiable a través del canal de dos formas: mediante una explicación intuitiva basada en la definición de transiciones de alta probabilidad en la línea de la explicación utilizada anteriormente para explicar el principio de la codificación de canal, tal y como se presenta en [Proakis y Salehi, 2002]; y mediante la definición presentada por Shannon en el teorema de codificación de canal.

Aplicando la ley de los grandes números, para valores de n suficientemente grandes, cuando por una canal binario con probabilidad de error de bit ε se transmite una secuencia de n bits, la salida tendrá con gran probabilidad $n \times \varepsilon$ bits erróneos; es decir, la secuencia recibida tendrá con alta probabilidad $n \times \varepsilon$ bits distintos con respecto a la secuencia transmitida. El número de posibles secuencias de n bits que se diferencian en $n \cdot \varepsilon$ bits viene dado por el número combinatorio

$$\binom{n}{n \cdot \varepsilon}.$$

Teniendo en cuenta que un número combinatorio se puede obtener como

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

y utilizando la aproximación de Stirling para los números factoriales, que viene dada por

$$n! \approx n^n e^{-n} \sqrt{2\pi n}$$

ese número combinatorio puede aproximarse como

$$\binom{n}{n \cdot \varepsilon} \approx 2^{n\, H_b(\varepsilon)},$$

donde $H_b(\varepsilon)$ es la función de entropía binaria con argumento ε. Esto significa que para cada posible secuencia de n bits transmitidos hay aproximadamente $2^{n\,H_b(\varepsilon)}$ secuencias altamente probables en la salida del sistema.

Por otro lado, el número total de secuencias de longitud n bits altamente probables a la salida, depende de la incertidumbre de cada uno de los bits que forman la secuencia, medida a través de su entropía, que a la vez dependerá de la probabilidad de tener un uno o un cero en la salida. Si ambos símbolos son equiprobables, todas las secuencias posibles tendrán la misma probabilidad, y el número de secuencias altamente probables será 2^n. Pero si los bits en la salida no son igualmente probables, la probabilidad de las secuencias de salida será diferente para cada secuencia. Por ejemplo, si el bit "0" es menos probable que el bit "1", las secuencias con muchos ceros serán menos probables. En ese caso, el número de secuencias de bits altamente probables se puede aproximar a partir de la medida de entropía de la salida, $H(Y)$, que será $H(Y) = H_b(p_Y(0))$, utilizando la expresión

$$2^{n\,H(Y)}.$$

Por tanto, el número máximo de secuencias de n bits en la entrada sin solapamiento entre las salidas altamente probables que generan en la salida, y que se denotará como M_{ss}, será el cociente entre el número de secuencias altamente probables en la salida y el número de secuencias que con alta probabilidad se generan en la salida cuando se transmite una cierta secuencia en la entrada, es decir

$$M_{ss} = \frac{2^{nH(Y)}}{2^{nH_b(\varepsilon)}} = 2^{n(H(Y)-H_b(\varepsilon))}.$$

El número de bits de información que se puede asociar a esas M_{ss} secuencias sin solapamiento es

$$\log_2 M_{ss} = n\,(H(Y) - H_b(\varepsilon)) \text{ bits de información.}$$

Por tanto la tasa del código resultante será el cociente entre los bits de información y el número de usos del canal

$$R = \frac{\log_2 M_{ss}}{n} = H(Y) - H_b(\varepsilon).$$

El máximo valor posible de esta tasa R es el que define la denominada *capacidad de canal*, que se denota como C. La entropía de los bits a la salida del canal binario será máxima cuando los bits "0" y "1" son igualmente probables, en cuyo caso $H(Y) = 1$. Por tanto, de forma intuitiva se ha llegado a que la capacidad del canal para un canal digital binario simétrico con probabilidad de error de bit ε es

$$C = 1 - H_b(\varepsilon) \text{ bits por uso del canal.}$$

A este resultado se ha llegado de forma intuitiva para el caso del canal BSC. A continuación se obtendrá el mismo resultado a partir de la teoría de la información. Para ello se calcula la información mutua entre la entrada y la salida del canal y a través de ella se intentará averiguar qué parte de la información se transmite y qué parte se pierde a su paso por el canal.

Para calcular la información entre la entrada, X, y la salida, Y, del canal DMC es necesario conocer las distribuciones de ambas variables. Conociendo la de la entrada, $p_X(x_i)$, como se conocen las probabilidades de transición, se conoce la distribución conjunta de la entrada y la salida

$$p_{X,Y}(x_i, y_j) = p_{Y|X}(y_j|x_i) \cdot p_X(x_i).$$

A partir de esta, la distribución de la salida se obtiene como

$$p_Y(y_j) = \sum_{i=0}^{M_X-1} p_{X,Y}(x_i, y_j) = \sum_{i=0}^{M_X-1} p_{Y|X}(y_j|x_i) \cdot p_X(x_i).$$

De este modo se tiene la caracterización completa entrada/salida. Tenga en cuenta que para un BSC

$$p_{Y|X}(y_j|x_i) = \begin{cases} 1-\varepsilon & \text{si } j = i \\ \varepsilon & \text{si } j \neq i \end{cases}.$$

A partir de estas distribuciones se puede calcular la información mutua entre entrada y salida $I(X,Y)$, por ejemplo a través de las relaciones con las distintas entropías, mediante

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

La información mutua entre la entrada y la salida representa la información que aporta la salida del canal sobre la entrada, o la incertidumbre que se elimina sobre el valor de la entrada cuando se conoce la salida: en definitiva, la información que es transmitida por el canal.

Para aclarar esta idea se utilizará $I(X,Y) = H(X) - H(X|Y)$ y se analizarán dos casos extremos del DMC más sencillo, el BSC. En este caso, el mejor canal posible es aquel libre de errores, es decir, $\varepsilon = 0$, o alternativamente $\varepsilon = 1$. Tenga en cuenta que en un sistema binario equivocarse siempre es una forma de acertar siempre, sólo hay que cambiar la decisión para ello. El peor caso posible es el caso en el que $\varepsilon = 1/2$ (cualquier ε mayor que $1/2$ se puede asimilar, cambiando la decisión, a un caso $1-\varepsilon$).

Para el caso $\varepsilon = 0$, la distribución conjunta entrada salida es

$p_{X,Y}(x_i,y_j)$	x_0	x_1	$p_Y(y_j)$
y_0	$p_X(x_0)$	0	$P_X(x_0)$
y_1	0	$p_X(x_1)$	$P_X(x_1)$

o teniendo en cuenta que no se producen errores, $p_Y(y_i) = p_X(x_i)$ o simplemente se puede establecer la igualdad $Y = X$. Por tanto,

$$I(X,Y) = I(X,X) = H(X),$$

lo que significa que $H(X|Y) = 0$.

Por otro lado, la distribución conjunta cuando $\varepsilon = 1/2$ es

$p_{X,Y}(x_i,y_j)$	x_0	x_1	$p_Y(y_j)$
y_0	$p_X(x_0)/2$	$p_X(x_1)/2$	$1/2$
y_1	$p_X(x_0)/2$	$p_X(x_1)/2$	$1/2$

.

En este caso, Y presenta una distribución equiprobable independientemente de la distribución de X, como era de esperar. Esto quiere decir que X e Y son estadísticamente independientes, y la probabilidad conjunta se puede poner como

$$p_{X,Y}(x_i,y_j) = p_X(x_i) \cdot p_Y(y_j).$$

Como ya se ha deducido con anterioridad, si X e Y son independientes, su información mutua es nula,

$$I(X,Y) = 0,$$

lo que significa que $H(X|Y) = H(X)$.

De estos dos casos se pueden extraer las siguientes conclusiones:

1. La información mutua entre entrada y salida del canal es la cantidad de información que pasa de la entrada a la salida cuando el canal es utilizado. En el caso en que la probabilidad de error es nula, pasa toda la información ($I(X,Y) = H(X)$), y en el caso en que la entrada y la salida son estadísticamente independientes se "pierde" toda la información ($I(X,Y) = 0$).

2. $H(X|Y)$ puede interpretarse como la información que se "pierde" en el canal, y así la información que "atraviesa" el canal, $I(X,Y)$, es igual a la información que hay a la entrada, $H(X)$, menos la que se pierde, $H(X|Y)$. Cuando la probabilidad de error es nula la pérdida es nula, y cuando la entrada y la salida son estadísticamente independientes, la perdida es total, es decir, igual a la información a la entrada del canal.

Estas conclusiones se pueden extender a cualquier DMC con alfabetos de entrada y salida de M_X y M_Y símbolos, respectivamente.

Ahora bien, la información mutua entre la entrada y la salida del canal depende de la distribución de probabilidades a la entrada. Si deseamos saber cuál es la máxima cantidad de información capaz de atravesar un determinado canal, es necesario considerar todas las posibles distribuciones y la que produzca la mayor información mutua será la óptima, y la información mutua para esa distribución será la máxima información capaz de atravesar el canal, es decir, la *capacidad de canal*.

Formalmente, se define la capacidad de canal, C, de un canal discreto sin memoria, como

$$C = \max_{p_X(x_i)} I(X,Y).$$

Sus unidades son bits (o bits por uso de canal).

Algunas propiedades de la capacidad de canal, que derivan de su definición a través de la información mutua, son las que definen las cotas para su valor mínimo y máximo:

1. $C \geq 0$, ya que $I(X,Y) \geq 0$.
2. $C \leq \log M_X$, ya que $C = \max I(X,Y) \leq \max H(X) = \log M_X$.
3. $C \leq \log M_Y$, por la misma razón.

Para algunos casos sencillos, como el canal binario simétrico, es posible calcular de forma directa la capacidad del canal. Para el BSC, la información mutua es

$$\begin{aligned}
I(X,Y) =& H(Y) - H(Y|X) \\
=& H(Y) - \sum_{i=0}^{1} p_X(x_i) H(Y|X = x_i) \\
=& H(Y) - \sum_{i=0}^{1} p_X(x_i) \left(- \sum_{j=0}^{1} p_{Y|X}(y_j|x_i) \log p_{Y|X}(y_j|x_i) \right) \\
=& H(Y) - \sum_{i=0}^{1} p_X(x_i)(-\varepsilon \log(\varepsilon) - (1-\varepsilon)\log(1-\varepsilon)) \\
=& H(Y) - \sum_{i=0}^{1} p_X(x_i) H_b(\varepsilon) \\
=& H(Y) - H_b(\varepsilon).
\end{aligned}$$

Se comprueba que se llega al mismo resultado obtenido con anterioridad. El máximo de esta información mutua se obtiene, dado que el parámetro ε es fijo y no se puede actuar sobre el, cuando la entropía $H(Y)$ es máxima. Esto se produce para una distribución de salida equiprobable, que para el BSC equivale a tener una distribución de entrada equiprobable. En este caso, $H(Y) = 1$ y la capacidad de canal vale por tanto

$$C = 1 - H_b(\varepsilon).$$

Es fácil representar esta dependencia con la probabilidad de error teniendo en cuenta la forma de variación de $H_b(\varepsilon)$. En un canal con probabilidad de error nula podemos transmitir un bit por uso del canal, mientras que en un canal con probabilidad de error 1/2 no se puede enviar ninguna información.

Hay que tener en cuenta que el problema del cálculo de la capacidad de canal en general se puede plantear como un problema de maximización con restricciones. Maximización de la información mutua, que depende de las probabilidades de entrada, con las restricciones impuestas por dichas probabilidades, ya que deben siempre cumplirse las siguientes $2M_X + 1$ restricciones:

- $p_X(x_i) \geq 0$, para $i = 0, 1, \cdots, M_X - 1$.
- $p_X(x_i) \leq 1$, para $i = 0, 1, \cdots, M_X - 1$.
- $\sum_{i=0}^{M_X-1} p_X(x_i) = 1$.

En algunos canales simples será posible calcular la capacidad de canal de forma analítica. Para otro tipo de canales más complicados, a veces no es posible encontrar de forma sencilla la solución de forma analítica. En este caso se recurre a técnicas numéricas, que barren todas las posibles distribuciones de entrada hasta encontrar la que hace máxima la información mutua.

Ejemplo

Se pretende encontrar la capacidad del canal mostrado en la Figura 5.26.

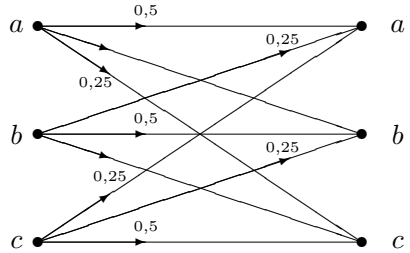

Figura 5.26: Un ejemplo de canal DMC

Para obtener la capacidad, lo primero es obtener una expresión de la información mutua. Una opción es a través de $H(Y)$ y $H(Y|X)$. De la relación entrada salida se observa que para los tres casos, dado $X = a$, $X = b$ o $X = c$, la variable aleatoria de salida Y, tiene tres símbolos con probabilidades 1/2, 1/4 y 1/4 respectivamente. Por tanto,

$$H(Y|X = a) = H(Y|X = b) = H(Y|X = c) = \frac{1}{2}\log_2(2) + 2 \times \frac{1}{4}\log_2(4) = 1{,}5.$$

Como para todos los símbolos se obtiene lo mismo, independientemente de $P_X(x_i)$ se tiene que

$$H(Y|X) = 1{,}5.$$

Por tanto

$$I(Y,X) = H(Y) - 1{,}5$$

Para maximizar $I(X,Y)$, hay que maximizar $H(Y)$, y esto supone encontrar la distribución de probabilidad de la variable aleatoria de entrada que produce salidas igualmente probables. En este caso esto sucede para símbolos de entrada igualmente probables. En este caso

$$H(Y) = \log_2(3) = 1{,}585.$$

Por tanto la capacidad del canal es

$$C = 1{,}585 - 1{,}5 = 0{,}085 \text{ bits/uso del canal.}$$

Se observa que es relativamente baja, pero es normal si se tiene en cuenta que tranmitiendo símbolos a través de este canal se falla un 50 % de las veces.

5.4.3. Capacidad de canal para el canal gausiano

Una vez estudiado el canal digital y cómo se obtiene su capacidad, en esta sección se extiende el estudio para el caso del canal gausiano, donde la salida del mismo en un instante dado es una variable aleatoria continua en lugar de una variable aleatoria discreta. Se tiene en cuanta en este análisis que en un sistema de comunicaciones real existen habitualmente dos importantes limitaciones en cuanto al uso de los recursos:

- Una limitación en la potencia máxima que puede transmitirse, que se aplica sobre la señal de comunicaciones que genera el transmisor.
- Una limitación en el ancho de banda de la señal transmitida.

Por eso en esta sección se estudiará cuál es la máxima cantidad de información que puede transmitirse de forma fiable en un sistema de comunicaciones digitales a través de un canal gausiano cuando la potencia máxima de la señal transmitida está liminada a P_X wattios y el ancho de banda disponible son B Hz. Al igual que en la sección anterior, el análisis se realizará de dos formas. Por un lado se hará una demostración más intuitiva en términos del número de símbolos que pueden transmitirse con baja probabilidad de solapamiento de sus salidas, en la línea del tratamiento seguido en [Proakis y Salehi, 2002], y por otro se obtendrá también la capacidad de canal mediante un desarrollo basado en la teoría de la información, siguiendo la línea del análisis realizado en [Artés-Rodríguez et al., 2007].

Capacidad del canal gausiano: demostración intuitiva

En un canal gausiano la salida del canal se modela mediante una variable aleatoria Y que se relaciona con la entrada, modelada mediante una variable aleatoria X, a través de la relación aditiva

$$Y = X + Z,$$

donde Z es una variable aleatoria que modela el efecto del ruido, con una distribución gausiana, de media nula y con varianza (potencia de ruido) P_Z.

La idea a seguir para obtener la capacidad del canal es similar a la seguida en el canal digital. Se va a muestrear las señales transmitida y recibida en n instantes de tiempo con el objetivo de ver cuál es el máximo número de posibles valores de la señal transmitida que da lugar en la salida a valores con baja probabilidad de solapamiento en el límite cuando n tiende a infinito. La restricción de potencia en la señal transmitida implica que si se tienen n realizaciones de la variable aleatoria X, que se agrupan en un vector $\boldsymbol{x}$

$$\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\},$$

lo que modelaría el valor de la señal transmitida en n instantes, para un valor de n suficientemente grande, se cumple

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P_X.$$

En este caso x_i no denota el alfabeto de una variable aleatoria discreta, sino las n realizaciones de la variable aleatoria continua X que modela la amplitud de la señal transmitida, $s(t)$, en n instantes de tiempo.

Para bloques de longitud n de valores a la entrada, salida y de ruido, agrupados en los vectores $\boldsymbol{x}$, $\boldsymbol{y}$ y $\boldsymbol{z}$, respectivamente, se puede escribir la relación vectorial

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}.$$

Si n es suficientemente grande, por la ley de los grandes números, la limitación en potencia del término de ruido implica la restricción

$$\frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \leq P_Z.$$

Finalmente, dada la independencia entre X y Z, la potencia de Y será la suma de las potencias de X y Z, es decir $P_Y = P_X + P_Z$, de tal modo que sobre las n muestras de la salida se tendrá la restricción

$$\frac{1}{n}\sum_{i=1}^{n} y_i^2 \leq P_Y = P_X + P_N.$$

Teniedo en cuenta que la suma de coordenadas al cuadrado de un vector proporciona su norma al cuadrado, las restricciones que imponen la potencia de la señal transmitida y la potencia de ruido se pueden escribir de la siguiente forma

$$||\boldsymbol{x}||^2 \leq nP_X,\ ||\boldsymbol{y} - \boldsymbol{x}||^2 \leq nP_N,\ ||\boldsymbol{y}||^2 \leq n(P_X + P_N).$$

La interpretación geométrica de las restricciones que imponen la potencia de la señal transmitida y la potencia de la señal de ruido, conduce a las siguientes conclusiones cuando n se incrementa de forma asintótica

- La representación vectorial de las n muestras de la señal transmitida, $\boldsymbol{x}$, se encuentra localizada en una hiper-esfera n-dimensional de radio $\sqrt{nP_X}$ centrada en el origen.
- La representación vectorial de las n muestras de la señal de salida, $\boldsymbol{y}$, se encuentra en una hiper-esfera n-dimensional de radio $\sqrt{nP_Z}$ y centrada en torno a la representación vectorial de las n muestras de la señal transmitida, $\boldsymbol{x}$.

- La representación vectorial de las n muestras de la señal recibida, $\boldsymbol{y}$, se encuentra localizada en una hiper-esfera n-dimensional de radio $\sqrt{n(P_X + P_Z)}$ centrada en el origen.

Haciendo uso de esta interpretación geomética, el cálculo de la capacidad del canal equivale a encontrar cuantas secuencias distintas de n muestras de la señal transmitida $\boldsymbol{x}$ se pueden obtener de tal forma que las salidas a que dan lugar no se solapen sobre el espacio de salida. Obviamente, si se cumple esta condición, entonces las secuencias de salida se pueden decodificar de forma fiable. Por tanto, la pregunta a la que hay que responder para establecer el valor de la capacidad del canal es: *¿Cuantas esfereras de radio $\sqrt{nP_Z}$ se pueden empaquetar en una esfera de radio $\sqrt{n(P_X + P_Z)}$?* La Figura 5.27 ilustra este planteamiento del problema para el cálculo de la capacidad de canal.

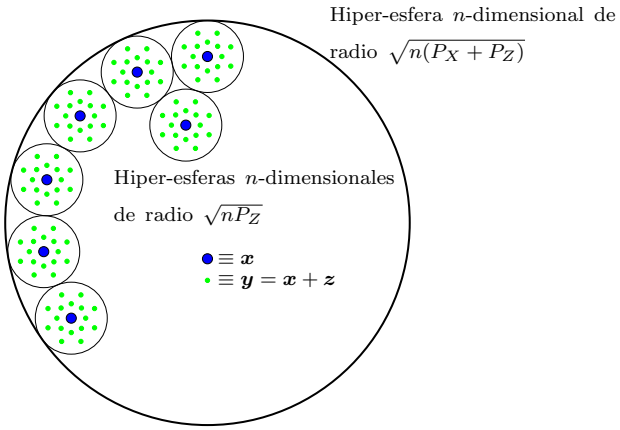

Figura 5.27: Interpretación geométrica del rango de las n muestras en la salida cuando se transmite un cierto valor para las n muestras de la entrada para el cálculo de la capacidad de un canal gausiano.

La respuesta, de forma aproximada, se puede obtener mediante la relación entre los volúmenes de las dos hiper-esferas. El volumen de una hiper-esfera de dimensión n y radio r es proporcional a su radio elevado a la potencia n,

$$V_n = K_n r^n,$$

donde K_n es una constante independiente del radio que depende de la dimensión del espacio. En este caso, el número máximo de símbolos, entendidos como bloques de n muestras de la señal transmitida, que pueden ser transmitidos con fidelidad a través de un canal gausiano se aproxima mediante el cociente entre el volumen de la esfera en que está contenida la representación de las n muestras de la salida del canal, que tiene radio $\sqrt{n(P_X + P_Z)}$, y el volumen de la esfera en que está contenida la representación de la salida cuando se ha transmitido un cierto símbolo $\boldsymbol{x}$, que tiene radio $\sqrt{nP_Z}$. Este cociente es por tanto

$$\begin{aligned} M_{ss} =& \frac{K_n(n(P_X + P_Z))^{n/2}}{K_n(nP_Z)^{n/2}} = \left(\frac{P_X + P_Z}{P_Z}\right)^{n/2} \\ =& \left(1 + \frac{P_X}{P_Z}\right)^{n/2}. \end{aligned}$$

Como con M_{ss} secuencias sin solapamiento se pueden codificar $\log_2 M_{ss}$ bits de información, la capacidad del canal gausiano con restricción de potencia P_X y potencia de ruido P_Z viene dada

por el cociente entre este número de bits y el número de usos del canal, que será n, es decir

$$\begin{aligned} C =& \frac{\log_2 M_{ss}}{n} = \frac{1}{n}\frac{n}{2}\log_2\left(1+\frac{P_X}{P_Z}\right) \\ =& \frac{1}{2}\log_2\left(1+\frac{P_X}{P_Z}\right). \end{aligned}$$

Ahora hay que recordar que al trabajar con una limitación en el ancho de banda, en el receptor se filtra la señal recibida con un filtro ideal de ancho de banda B Hz para minimizar el efecto del ruido, con lo que la potencia del ruido filtrado es

$$P_Z = N_0 B.$$

Por tanto la capacidad del canal gausiano es

$$C = \frac{1}{2}\log_2\left(1+\frac{P_X}{N_0 B}\right) \text{ bits/uso.}$$

Si se multiplica este resultado por el número de usos (transmisiones) por segundo, que cumpliendo el teorema de Nyquist son $2B$, se obtiene la capacidad del canal en bits/s,

$$C = B\log_2\left(1+\frac{P}{N_o B}\right) \text{ bits/s.}$$

Esta es la conocida fórmula de Shannon para la capacidad de un canal con ruido aditivo blanco y gausiano.

Capacidad del canal gausiano a través de la información mutua

A este mismo resultado se puede llegar también a través de una derivación más formal basada en la teoría de la información [Artés-Rodríguez et al., 2007]. En el modelo de canal gausiano la relación entre la entrada y la salida viene dada por

$$Y = X + Z,$$

donde Z es una variable aleatoria gausiana, de media nula y varianza $P_Z = N_0 B$. De este modo la distribución condicional de Y dado X, $f_{Y|X}(y|x)$, es una distribución gausiana de media x y varianza $\sigma^2 = P_Z$. La capacidad del canal se obtendrá a través de la información mutua de la misma forma que para el canal digital, maximizando la información mutua entre la entrada y la salida del canal, pero incluyendo la restricción en la potencia máxima para la señal transmitida (sin esta restricción, la capacidad sería teóricamente infinita). Es decir, se define la capacidad de canal como

$$C = \max_{f_X(x)|E[X^2]\leq P_X} I(X,Y),$$

donde la restricción en potencia viene dada por la limitación $E[X^2] \leq P_X$.

La información mutua entre entrada y salida se puede calcular a través de su relación con las entropías diferenciales

$$I(X,Y) = h(Y) - h(Y|X) = h(X+Z) - h(Z).$$

La entropía diferencial del ruido se calcula de forma muy sencilla, ya que su distribución es gausiana de media nula y varianza $\sigma^2 = P_Z$, caso que ya se ha considerado con anterioridad, por lo que

$$h(Z) = \int_{-\infty}^{\infty} f_Z(z) \log_2 \frac{1}{f_Z(z)}\, dz = \frac{1}{2} \log_2 2\pi e P_Z.$$

Esta entropía diferencial depende únicamente de la varianza del término de ruido, $P_Z = N_0 B$. Por tanto la información mutua se puede escribir como

$$I(X,Y) = h(X+Z) - \frac{1}{2} \log_2 2\pi e P_Z.$$

Para obtener su máximo se hará uso de la siguiente propiedad sobre la entropía diferencial:

- Si una variable aleatoria tiene un valor fijo de varianza, la función densidad de probabilidad que hace máxima su entropía diferencial es la distribución gausiana.

La demostración de esta propiedad se puede ver, por ejemplo, en [Artés-Rodríguez et al., 2007], Capítulo 9, página 565.

Dado que X y Z son estadísticamente independientes y Z tiene media nula, la varianza de Y es la suma de las varianzas de X e Y, por lo que

$$E[Y^2] = E[(X+Z)^2] = E[X^2] + E[Z^2] \leq P_X + P_Z.$$

Como la varianza de Y está acotada, la capacidad se alcanza cuando Y tiene una distribución gausiana, en cuyo caso el valor máximo de la información mutua, y por tanto la capacidad, es

$$C = \frac{1}{2} \log_2 2\pi e (P_X + P_Z) - \frac{1}{2} \log_2 2\pi e \sigma^2 = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{P_Z}\right).$$

Este resultado es el mismo que el obtenido anteriormente.

5.5. Límites en un sistema digital de comunicaciones

En la sección anterior se ha obtenido la capacidad de canal para el canal gausiano

$$C = B \cdot \log_2 \left(1 + \frac{P_X}{N_0 B}\right) \text{ bits/s.}$$

Además de por supuesto del valor de N_0, el valor de la capacidad de canal depende de dos parámetros relevantes de todo sistema de comunicaciones: la potencia de la señal transmitida, P_X, y el ancho de banda disponible, B Hz. En esta sección en primer lugar se analizará la dependencia de la capacidad de canal con estos dos parámetros del sistema de comunicaciones.

En cuanto a la potencia de la señal transmitida, el análisis es sencillo. Si se incrementa la potencia transmitida, se incremente la capacidad de canal: obviamente, cuanto mayor es la potencia, más niveles (o símbolos de longitud n muestras) se pueden poner, y más bits/uso son posibles. Sin embargo hay que tener en cuenta que el incremento sigue una ley logarítmica, de modo que para poder obtener un incremento lineal en la capacidad, es preciso un incremento exponencial en la potencia transmitida. En cualquier caso, teóricamente es posible incrementar la capacidad hasta el infinito incrementando la potencia de la señal transmitida.

El efecto del ancho de banda en la capacidad del canal es diferente. Incrementar B tiene dos efectos contrapuestos. Por un lado, un mayor ancho de banda incrementa la velocidad de transmisión, pero por otro lado incrementa el nivel del ruido y por tanto reduce las prestaciones. Haciendo llegar B al límite infinito, y aplicando la regla de L'Hopital, se obtiene el límite de la capacidad de canal cuando el ancho de banda tiende a infinito

$$\lim_{B\to\infty} C = \frac{P_X}{N_0}\log_2(e) = 1{,}44\frac{P_X}{N_0}.$$

Este resultado significa que, al contrario de lo que sucede con la potencia de la señal transmitida, el incremento del ancho de banda del canal por si sólo no puede aumentar la capacidad hasta cualquier valor deseado, sino que hay un límite máximo alcanzable que depende de la relación señal a ruido (P_X/N_0).

El teorema de codificación de canal de Shannon plantea un límite máximo en la tasa de transmisión de información con un sistema de comunicaciones digitales. Esto quiere decir que en un sistema de comunicaciones práctico, siempre se debe cumplir que la tasa de transmisión efectiva (definida sobre los bits de información) esté por debajo de la capacidad de canal, $R < C$. Para el caso del canal gausiano esto implica que la velocidad de transmisión binaria, R_b bits/s, tiene que estar por debajo de la capacidad del canal. Definiendo la relación señal a ruido como la relación entre la potencia de la señal transmitida y N_0

$$SNR = \frac{P_X}{P_Z} = \frac{P_X}{N_0 B},$$

esto significa que se debe cumplir la relación

$$R_b < B\log\left(1+\frac{P}{N_0 B}\right) \text{ bits/s}.$$

Dividiendo los dos lados de esta desigualdad por B, definiendo la tasa binaria espectral, o eficiencia espectral, como

$$\eta = \frac{R_b}{B} \text{ bits/s/Hz},$$

se obtiene el siguiente límite para la eficiencia espectral en un sistema de comunicaciones práctico

$$\eta < \log_2\left(1+\frac{P_X}{N_0 B}\right).$$

Si ahora se define la energía media por bit como el cociente entre la potencia de la señal transmitida y la tasa de transmisión binaria

$$E_b = \frac{P_X}{R_b},$$

y la correspondiente relación entre E_b y N_0,

$$\frac{E_b}{N_0} = \frac{SNR}{\eta},$$

se puede reescribir la limitación de la tasa binaria espectral o eficiencia espectral en términos de esta relación como

$$\eta < \log_2\left(1+\eta\frac{E_b}{N_0}\right).$$

Esta restricción que relaciona la relación señal a ruido y la eficiencia espectral del sistema se escribe en la literatura con varias expresiones alternativas equivalentes, como por ejemplo

$$\frac{E_b}{N_0} > \frac{2^\eta - 1}{\eta}.$$

Esta expresión indica que hay un valor mínimo de relación E_b/N_0 que es necesario para poder tener una comunicación fiable. En el caso de esta expresión, si se calcula su límite cuando η tiende a infinito se tiene

$$\lim_{\eta\to\infty} \frac{E_b}{N_0} = \ln 2 = 0{,}693 \approx -1{,}6 \text{ dB}.$$

La curva que define el valor límite de la tasa binaria espectral como una función de la relación E_b/N_0, que se representa en la Figura 5.28, divide el plano η vs E_b/N_0 en dos regiones. En una región, debajo de la curva, es posible una comunicación fiable (entendida como una comunicación en la que el uso de técnicas de codificación de canal permite reducir la probabilidad de error hasta niveles arbitrariamente bajos). En la otra, por encima de la curva, no es posible tener una comunicación fiable. Las prestaciones de cualquier sistema se pueden representar mediante un punto en el plano de esta curva. Cuanto más cerca está el punto de la curva, más cerca está el rendimiento de ser óptimo.

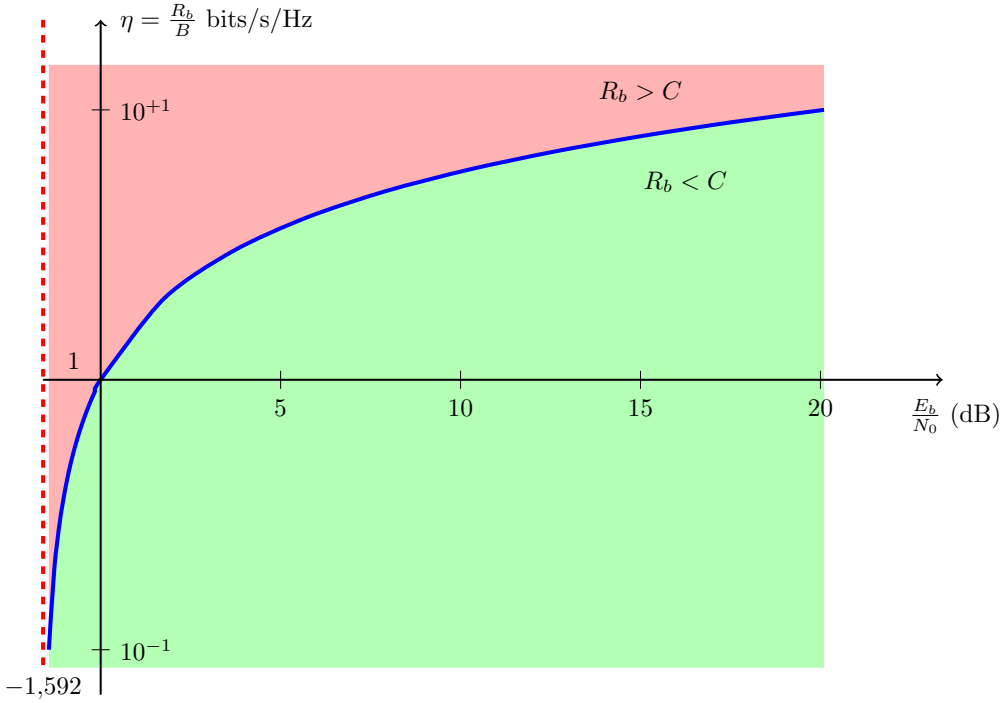

Figura 5.28: Valor límite de la tasa binaria espectral η (bits/s/Hz) frente a relación E_b/N_0 para un canal gausiano. Se distingue la región donde es posible una comunicación fiable (en verde) de la región donde no es posible (en rojo).

Se puede ver que esta curva, cuando η tiende a cero, la relación E_b/N_0 tiende al valor anteriormente calculado

$$\frac{E_b}{N_0} = \ln 2 = 0{,}693 \approx -1{,}6 \text{ dB}.$$

Este es un mínimo absoluto para una comunicación fiable, es decir, que para que se pueda tener una comunicación fiable, la relación E_b/N_0 ha de estar por encima de ese límite.

En esta figura se pueden también apreciar dos cosas en cuanto al valor de η:

1. Cuando $\eta << 1$, el ancho de banda es grande y la única limitación aparece por la potencia. En este caso se habla de *sistemas limitados en potencia*. En este caso hay que utilizar constelaciones sencillas.

2. Cuando $\eta >> 1$ el ancho de banda del canal es pequeño y se habla de *sistemas limitados en ancho de banda*. En este caso se utilizan constelaciones muy densas (p.e. 256-QAM).

La expresión anterior indica un límite mínimo que debe alcanzar la relación entre la energía de la señal y la del ruido para poder transmitir con una cierta eficiencia espectral de forma fiable. Aunque en la literatura ese límite se suele expresar a través de la relación E_b/N_0, en otras ocasiones se expresa a partir de la relación señal a ruido, en cuyo caso la expresión resultante es

$$SNR > 2^{\eta} - 1.$$

Dado este límite mínimo de SNR, en ocasiones se define la denominada "*relación señal a ruido normalizada*", que básicamente compara la relación señal a ruido de trabajo con ese mínimo nivel requerido

$$SNR_{norm} = \frac{SNR}{2^{\eta} - 1}.$$

Por la propia definición de esta relación señal a ruido normalizada, en un sistema práctico debe tomar valores mayores que la unidad, o lo que es lo mismo, mayores que 0 decibelios. Se trata por tanto de una medida relativa que indica lo cerca o lejos que se encuentra el sistema del valor límite de funcionamiento en la zona fiable.

