

Teoría de la Comunicación

Capítulo 4

Límites fundamentales en comunicaciones digitales (Teoría de la Información)

Marcelino Lázaro

Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid



1/88

Índice de contenidos

- **Introducción**
- **Modelos (probabilísticos) de fuentes de información**
- **Modelos (probabilísticos) de canal**
- **Medidas cuantitativas de información**
 - ▶ **Entropías**
 - ▶ **Información mutua**
- **Límites en un sistema digital de comunicaciones**
 - ▶ **Introducción al concepto de codificación**
 - ★ Mecanismo para la detección y corrección de errores
 - ▶ **Capacidad de canal**
 - ★ Capacidad de un canal digital
 - ★ Capacidad de un canal gaussiano

Introducción

- Finalidad de un sistema de comunicaciones:

- ▶ Transmisión de información



- Teoría de la información

- ▶ Medidas cuantitativas de información
- ▶ Análisis de un sistema de comunicaciones
 - ★ Cantidad de información generada
 - ★ Cantidad de información transmitida de forma efectiva (recibida)
 - ★ Límites fundamentales en la transmisión de información

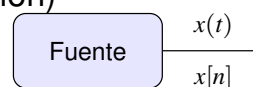
- Organización del capítulo:

- ▶ Modelos (probabilísticos) para las fuentes de información
- ▶ Modelos (probabilísticos) del sistema (canales)
- ▶ Medidas cuantitativas de información
- ▶ Límites fundamentales en un sistema digital de comunicaciones

Modelado de fuentes de información

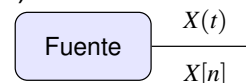
- Salida de la fuente: señal (flujo de información)

- ▶ En tiempo continuo: $x(t) \equiv s(t)$
- ▶ En tiempo discreto: $x[n] \equiv B_b[n]$ (o $A[n]$)



- Modelo de la salida de la fuente (información)

- ▶ Proceso aleatorio, $X(t)$, ó $X[n]$



- Modelo para fuentes en tiempo continuo (analógicas)

- ▶ Proceso aleatorio en tiempo continuo $X(t)$

- ★ Caracterización: $m_X(t), R_X(t + \tau, t), S_X(j\omega)$
 - Normalmente son procesos limitados en banda (B Hz)
 - $S_X(j\omega)$ refleja el comportamiento espectral medio de la fuente

- Modelo para fuente en tiempo discreto

- ▶ Proceso aleatorio en tiempo discreto $X[n]$
- ▶ Tipos de alfabeto de la fuente (posibles valores de $X[n]$)
 - ★ Continuo (e.g., señales muestreadas)
 - ★ Discreto (fuentes digitales)
 - Modelo más simple: **fuentes discretas sin memoria**

Fuente discreta sin memoria

DMS : *Discrete Memoryless Source*

- Fuente: Proceso aleatorio en tiempo discreto $X[n]$

- ▶ Discreta

- ★ Alfabeto de $X[n]$: discreto, de M_X valores $\mathcal{A}_X = \{x_0, x_1, \dots, x_{M_X-1}\}$

- ▶ Sin memoria

- ★ $p_{X[n]}(x_i)$ no depende de n : $p_{X[n]}(x_i) \equiv p_X(x_i)$
 - $\{X[n]\}$: V.A.'s independientes e idénticamente distribuidas (i.i.d.)

- Descripción del proceso (caracterización estadística)

- ▶ Variable aleatoria X

- ★ Al ser $X[n]$ i.i.d., la estadística es la misma para todo n
 - ★ Alfabeto $\mathcal{A}_X = \{x_0, x_1, \dots, x_{M_X-1}\}$
 - ★ Probabilidades $\{p_X(x_i)\}_{i=0}^{M_X-1} = \{p_X(x_0), p_X(x_1), \dots, p_X(x_{M_X-1})\}$

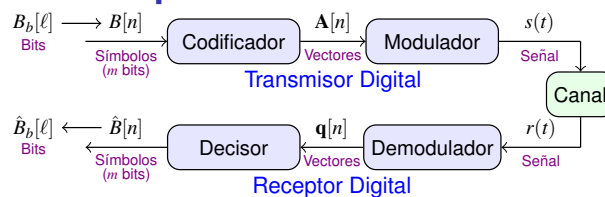
- Ejemplo: fuente binaria simétrica

BSS: *Binary Symmetric Source*

- ▶ Alfabeto $\mathcal{A}_X = \{x_0, x_1\}$, típicamente $x_0 \equiv 0, x_1 \equiv 1$

- ▶ Probabilidades $p_X(x_0) = p_X(x_1) = \frac{1}{2}$

Sistemas Digitales - Modelos probabilísticos de canal



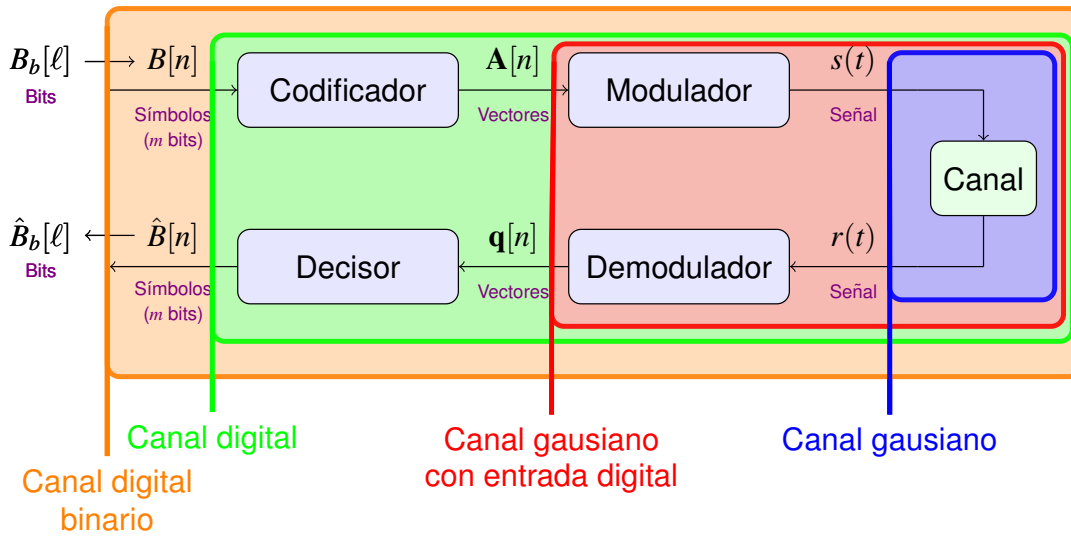
- Modelos probabilísticos de canal

- ▶ Establecen la relación probabilística entre la información recibida y la transmitida en distintos puntos de este modelo general del sistema de comunicaciones
 - ★ Distintos niveles de abstracción en la definición de *canal*
 - ★ Modelo probabilístico: Entrada (X) , salida (Y) , distribución $f_{Y|X}(y|x)$

- Modelos de canal

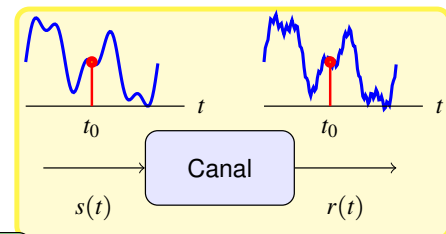
- ▶ Canal gaussiano: $Y \equiv r(t) \mid X \equiv s(t)$
 - ★ Representa el canal físico
 - ▶ Canal gaussiano con entrada digital: $Y \equiv \mathbf{q}[n] \mid X \equiv \mathbf{A}[n]$
 - ★ Representa el canal discreto equivalente
 - ▶ Canal digital: $Y \equiv \hat{B}[n] \mid X \equiv B[n]$
 - ★ Representa la transmisión de símbolos
 - ▶ Canal digital binario: $Y \equiv \hat{B}_b[\ell] \mid X \equiv B_b[\ell]$
 - ★ Representa la transmisión de bits

Modelos probabilísticos de canal - Representación



Canal gaussiano

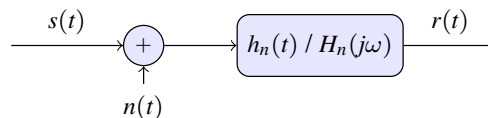
- Relación entrada / salida
 - ▶ Entrada: $X \equiv s(t)$, para un instante dado t
 - ▶ Salida: $Y \equiv r(t)$, para el mismo instante t
- Modelo canal gaussiano



$$r(t) = s(t) + n(t)$$

- ▶ $n(t)$: p.a. estacionario, blanco y gaussiano $m_n = 0, S_n(j\omega) = \frac{N_0}{2}, R_n(\tau) = \frac{N_0}{2} \delta(\tau)$

- Limitación de la potencia de ruido - Filtrado a la entrada del receptor



- ▶ Señal de ancho de banda B Hz ($W = 2\pi B$ rad/s)
 - ★ Filtro $h_n(t)$ ideal, ancho de banda B Hz: potencia de ruido $N_0 B$ Wats
- Distribución de $Y|X$ cuando $Y \equiv r(t_0)$ y $X \equiv s(t_0)$
 - ▶ Distribución de Y cuando $X = s(t_0) \equiv x$
 - ▶ Gaussiana, de media $x = s(t_0)$ y varianza $\sigma^2 = N_0 B$

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

Canal gaussiano con entrada digital

- Relación entrada / salida
 - ▶ Entrada: $\mathbf{X} \equiv \mathbf{A}[n]$, para un instante dado n
 - ★ Vector de N variables aleatorias (alfabeto discreto, $M = 2^m$ valores)
 - ★ Si $\mathbf{A}[n] = \mathbf{a}_i \rightarrow \mathbf{X} = \mathbf{x}_i \equiv \mathbf{a}_i$
 - ▶ Salida: $\mathbf{Y} \equiv \mathbf{q}[n]$, para el mismo instante n
 - ★ Vector de N variables aleatorias (alfabeto continuo)
- Es igual al canal discreto equivalente (Capítulo 3)
 - ▶ Se le da este nombre en el ámbito de Teoría de la Información (TI)
- Modelo del canal gaussiano con entrada digital

$$\mathbf{q}[n] = \mathbf{A}[n] + \mathbf{z}[n] \quad \equiv \quad \mathbf{q} = \mathbf{A} + \mathbf{z}$$

- ▶ Independiente del instante discreto n
- ▶ Distribución de los N elementos del vector de ruido $\mathbf{z} = [z_0, z_1, \dots, z_{N-1}]^T$
 - ★ Son N V.A.'s Independientes, gaussianas, media nula, varianza $N_0/2$
- Distribución condicional de la salida dada la entrada
 - ▶ Distribución gaussiana (N -dimensional)
 - ★ Media: el símbolo transmitido ($\mathbf{x}_i \equiv \mathbf{a}_i$)
 - ★ Varianza: $N_0/2$ en cada dirección del espacio N -dimensional

$$f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{q}-\mathbf{a}_i\|^2}{N_0}} \rightarrow f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_i) = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{y}-\mathbf{x}_i\|^2}{N_0}}$$

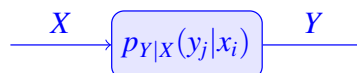
Canal digital

- Relación entrada / salida
 - ▶ Entrada: símbolo para un instante dado n
 - ★ $X \equiv B[n]$ (o alternatively $X \equiv \mathbf{A}[n]$)
 - ▶ Salida: símbolo para el mismo instante n
 - ★ $Y \equiv \hat{B}[n]$ (o alternatively $Y \equiv \hat{\mathbf{A}}[n]$)
 - ▶ Alfabeto de X e Y : símbolos (bloques de m bits / vectores)
 - ★ $M = 2^m$ posibles símbolos (Alfabeto de $B[n]$, o alternatively de $\mathbf{A}[n]$)

$$X, Y \in \{x_i = y_i \equiv b_i\}_{i=0}^{M-1} \quad \text{ó} \quad X, Y \in \{x_i = y_i \equiv \mathbf{a}_i\}_{i=0}^{M-1}$$

- Modelo probabilístico: canal discreto sin memoria

DMC: Discrete Memoryless Channel



- Caracterización del canal discreto sin memoria
 - 1 Alfabeto de entrada: M_X valores $\mathcal{A}_X = \{x_0, x_1, \dots, x_{M_X-1}\}$
 - 2 Alfabeto de salida: M_Y valores $\mathcal{A}_Y = \{y_0, y_1, \dots, y_{M_Y-1}\}$
 - 3 Conjunto de $M_X \times M_Y$ probabilidades condicionales (de transición) $p_{Y|X}(y_j|x_i)$,
 $\forall i \in \{0, 1, \dots, M_X - 1\}, \forall j \in \{0, 1, \dots, M_Y - 1\}$

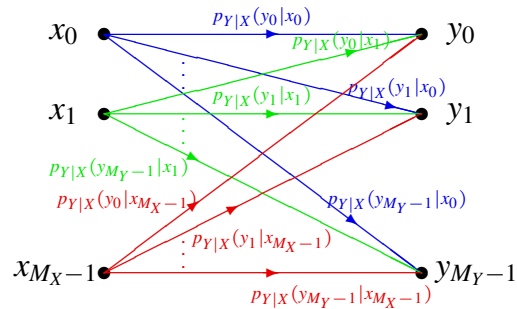
DMC: Representación de las probabilidades de transición

- Matriz de canal

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

► Elementos de una fila suman 1

- Diagrama de flechas



► Flechas que salen de un mismo nodo suman 1

Aplicación del DMC al canal digital

- Alfabetos de X e Y

► Alfabeto de $B[n]$ (m bits): $M = 2^m$ símbolos:

$$\mathcal{A}_X = \mathcal{A}_Y = \{b_0, b_1, \dots, b_{M-1}\}$$

$$x_i \equiv b_i, y_j \equiv b_j, M_X = M_Y = M, i, j \in \{0, 1, \dots, M-1\}$$

- Probabilidades de transición $p_{Y|X}(y_j|x_i) \equiv p_{\hat{B}|B}(b_j|b_i) = p_{\hat{A}|A}(\mathbf{a}_j|\mathbf{a}_i)$

► Probabilidades de acierto: $p_{Y|X}(y_i|x_i) = p_{\hat{B}|B}(b_i|b_i)$

► Probabilidades de error: $p_{Y|X}(y_j|x_i) = p_{\hat{B}|B}(b_j|b_i)$ para $j \neq i$

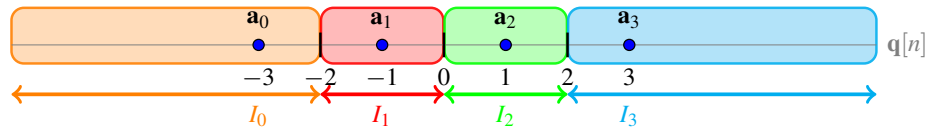
► Valores ideales: matriz de transición / diagrama de flechas

$$\mathbf{P} = \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Cálculo de las probabilidades de transición - Ejemplo

- $M = 4$, símbolos equiprobables $p_A(\mathbf{a}_i) = \frac{1}{4}$
 - ▶ Constelación: $\mathbf{a}_0 = -3$, $\mathbf{a}_1 = -1$, $\mathbf{a}_2 = +1$, $\mathbf{a}_3 = +3$
 - ▶ Regiones de decisión: umbrales $q_{u1} = -2$, $q_{u2} = 0$, $q_{u3} = +2$

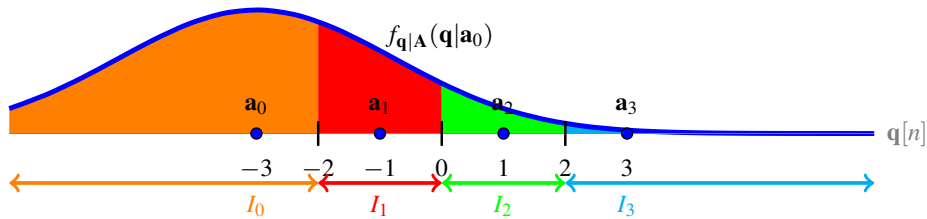
$$I_0 = (-\infty, -2], I_1 = (-2, 0], I_2 = (0, +2], I_3 = (+2, +\infty)$$



- Probabilidades de transición (matriz de canal)

$$\mathbf{P} = \begin{bmatrix} 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{5}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{1}{\sqrt{N_0/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{3}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) \\ Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) & Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) & 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right) \end{bmatrix}$$

Elementos de la primera fila: $x_0 \equiv \mathbf{a}_0 \rightarrow p_{Y|X}(y_j|x_0), \forall j$



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

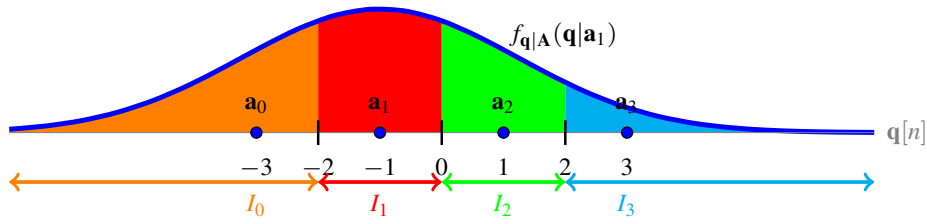
$$p_{Y|X}(y_0|x_0) = P_{a|\mathbf{a}_0} = 1 - P_{e|\mathbf{a}_0} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_0) = P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_0) = P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_2} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_0) = P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_3} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

Elementos de la segunda fila: $x_1 \equiv \mathbf{a}_1 \rightarrow p_{Y|X}(y_j|x_1), \forall j$



- Distribución $f_{q|A}(\mathbf{q}|\mathbf{a}_1)$: gaussiana de media $\mathbf{a}_1 = -1$ y varianza $N_0/2$

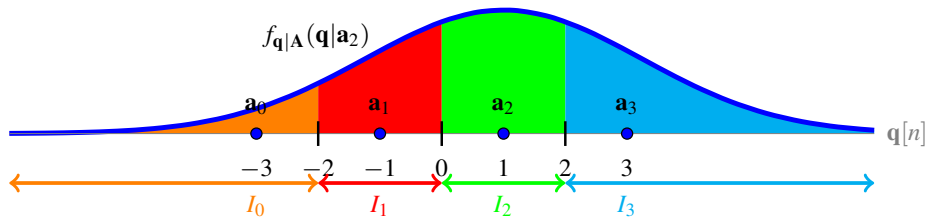
$$p_{Y|X}(y_0|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_0} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_1) = P_{a|\mathbf{a}_1} = 1 - P_{e|\mathbf{a}_1} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_1) = P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_3} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

Elementos de la tercera fila: $x_2 \equiv \mathbf{a}_2 \rightarrow p_{Y|X}(y_j|x_2), \forall j$



- Distribución $f_{q|A}(\mathbf{q}|\mathbf{a}_2)$: gaussiana de media $\mathbf{a}_2 = +1$ y varianza $N_0/2$

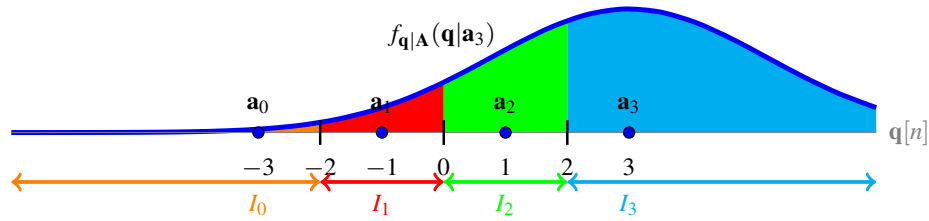
$$p_{Y|X}(y_0|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_0} = Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_1|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_1} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_2) = P_{a|\mathbf{a}_2} = 1 - P_{e|\mathbf{a}_2} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_2) = P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_3} = Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

Elementos de la cuarta fila: $x_3 \equiv \mathbf{a}_3 \rightarrow p_{Y|X}(y_j|x_3), \forall j$



- Distribución $f_{q|A}(q|\mathbf{a}_3)$: gaussiana de media $\mathbf{a}_3 = +3$ y varianza $N_0/2$

$$p_{Y|X}(y_0|x_3) = P_{e|\mathbf{a}_3 \rightarrow \mathbf{a}_0} = Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

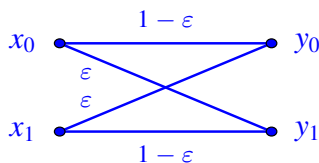
$$p_{Y|X}(y_1|x_3) = P_{e|\mathbf{a}_3 \rightarrow \mathbf{a}_1} = Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_2|x_3) = P_{e|\mathbf{a}_3 \rightarrow \mathbf{a}_2} = Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right)$$

$$p_{Y|X}(y_3|x_3) = P_{a|\mathbf{a}_3} = 1 - P_{e|\mathbf{a}_3} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

Canal digital binario

- Relación entrada / salida
 - ▶ Entrada: $X \equiv B_b[\ell]$, bit para un instante dado ℓ
 - ▶ Salida: $Y \equiv \hat{B}_b[\ell]$, bit para el mismo instante ℓ
 - ★ Alfabeto de X e Y : Bits $x_0 = y_0 \equiv 0$, $x_1 = y_1 \equiv 1$
- Modelo probabilístico
 - ▶ Particularización del DMC para $M_X = M_Y = 2$
 - ▶ 4 probabilidades condicionales $p_{Y|X}(y_j|x_i)$, para $i, j \in \{0, 1\}$
 - ★ 2 probabilidades de acierto de bit ($j = i$)
 - ★ 2 probabilidades de error de bit ($j \neq i$)
- Ejemplo: **Canal binario simétrico** (BSC: *Binary Symmetric Channel*)
 - ▶ Misma probabilidad de error para los dos símbolos: $p_{Y|X}(y_1|x_0) = p_{Y|X}(y_0|x_1) = \varepsilon$
 - ▶ Probabilidad (tasa) de error de bit (BER): $BER = \varepsilon$
 - ★ $BER = p_X(x_0) p_{Y|X}(y_1|x_0) + p_X(x_1) p_{Y|X}(y_0|x_1) = \varepsilon$



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

MEDIDAS CUANTITATIVAS DE INFORMACIÓN

Autoinformación (de un suceso de una V.A. discreta)

- $I_X(x_i)$: mide la información de un suceso de la variable aleatoria ($X = x_i$)
- Requerimientos para una medida de información de un suceso
 - 1 Debe depender de la probabilidad del suceso
 - ★ $I_X(x_i) = f(p_X(x_i))$
 - 2 Debe ser una función decreciente de la probabilidad
 - ★ $p_X(x_i) > p_X(x_j) \rightarrow I_X(x_i) < I_X(x_j)$
 - 3 Debe ser una función continua de la probabilidad
 - ★ $p_X(x_i) \approx p_X(x_j) \rightarrow I_X(x_i) \approx I_X(x_j)$
 - 4 Para el suceso conjunto ($X = x_i, Y = y_j$) independientes ($p_{X,Y}(x_i, y_j) = p_X(x_i) \times p_Y(y_j)$)
 - ★ $I_{X,Y}(x_i, y_j) = I_X(x_i) + I_Y(y_j)$
- Función que cumple estas propiedades - [Autoinformación](#)

$$I_X(x_i) = -\log_b(p_X(x_i))$$

- La base del logaritmo define las unidades de la medida
 - ★ Base 2 : bits
 - ★ Base e (logaritmo natural o neperiano $\ln$): nats
- NOTA: Relación $\log_b(x) = \ln(x) / \ln(b)$

Entropía (de una V.A. discreta)

- Medida de incertidumbre sobre el resultado de una variable aleatoria (discreta)
 - Alfabeto: $\mathcal{A}_X = \{x_0, x_1, \dots, x_{M_X-1}\}$ (M_X símbolos)
 - Probabilidades: $\{p_X(x_0), p_X(x_1), \dots, p_X(x_{M_X-1})\}$
- Promedio de la autoinformación de cada suceso

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

NOTA: Por convención: $0 \times \log 0 = 0$

- Unidades: bits/símbolo (base 2) ó nats/símbolo (base e)
- Valores límite de la entropía de variables aleatorias discretas

1 $H(X) \geq 0$

★ $0 \leq p_X(x_i) \leq 1$ y, consecuentemente, $-\log(p_X(x_i)) \geq 0$

★ $H(X) = 0$ cuando $p_X(x_i) = 1, p_X(x_j) = 0 \forall j \neq i$

- No hay ninguna incertidumbre sobre X

2 $H(X) \leq \log(M_X)$

★ $H(X) = \log(M_X)$ si los símbolos son equiprobables, $p_X(x_i) = 1/M_X$

- Situación de máxima incertidumbre sobre X

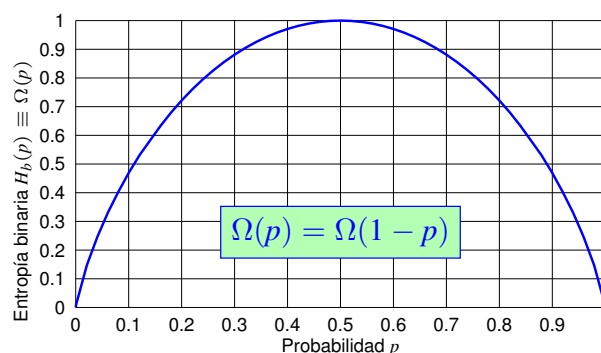
Ejemplo - Entropía binaria: $H_b(p) \equiv \Omega(p)$

- Variable aleatoria binaria

Alfabeto: $\{x_0, x_1\}$

Probabilidades: $\{p_X(x_0) = p, p_X(x_1) = 1 - p\}$

$$\begin{aligned} H(X) \equiv H_b(p) \equiv \Omega(p) &= -p \log_2(p) - (1-p) \log_2(1-p) \\ &= p \log_2\left(\frac{1}{p}\right) + (1-p) \log_2\left(\frac{1}{1-p}\right) \text{ bits/símbolo} \end{aligned}$$



- Valor máximo: $\max \Omega(p) = 1$ bit/símbolo
 - Se alcanza para $p = 0,5$ (valor de referencia)

Entropía conjunta (de dos V.A.'s discretas)

- Medida de la información conjunta de dos (o más) variables aleatorias

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

- Variables aleatorias independientes

- Probabilidad conjunta: $p_{X,Y}(x_i, y_j) = p_X(x_i) p_Y(y_j)$

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

Entropía condicional (de dos V.A.'s discretas)

- Incertidumbre en una v.a. cuando se conoce el valor de otra

- Entropía de X dado que $Y = y_j$: $p_X(x_i) \rightarrow p_{X|Y}(x_i|y_j)$

$$H(X|Y = y_j) = \sum_{i=0}^{M_X-1} p_{X|Y}(x_i|y_j) \log \frac{1}{p_{X|Y}(x_i|y_j)}$$

- Entropía condicional: Promedio de $H(X|Y = y_j)$

- Se promedia sobre todos los valores del alfabeto de Y

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

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

- Para variables aleatorias independientes

- Independencia: $p_{X|Y}(x_i|y_j) = p_X(x_i) \rightarrow H(X|Y = y_j) = H(X), \forall y_j \rightarrow H(X|Y) = H(X)$

Relación entre entropía conjunta y condicional

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

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

Información mutua (entre dos V.A.'s discretas)

- Mide la información que aporta una variable aleatoria X sobre el conocimiento de otra variable aleatoria Y

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

- Propiedades

1 $I(X, Y) = I(Y, X) \geq 0$

* La igualdad se cumple en el caso de que X e Y sean independientes

2 $I(X, Y) \leq \min(H(X), H(Y))$

- 3 Se puede definir una información mutua condicional

$$I(X, Y|Z) = \sum_{i=0}^{M_Z-1} p_Z(z_i) I(X, Y|Z = z_i)$$

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

- 4 La regla de la cadena para la información mutua es $I((X, Y), Z) = I(X, Z) + I(Y, Z|X)$

$$I((X_1, X_2, \dots, X_N), Y) = I(X_1, Y) + I(X_2, Y|X_1) + \dots + I(X_N, Y|X_1, \dots, X_{N-1})$$

- 5 A partir de la definición de información mutua se obtiene la definición de entropía

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

Relaciones información mutua y entropía

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

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

$$H(X, Y) = H(Y) + H(X|Y) \rightarrow H(X|Y) = H(X, Y) - H(Y)$$

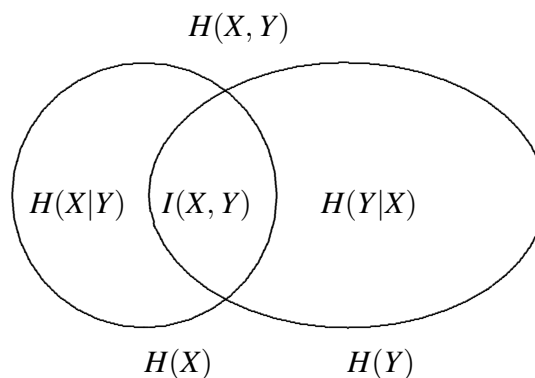
$$H(X, Y) = H(X) + H(Y|X) \rightarrow H(Y|X) = H(X, Y) - H(X)$$

Relaciones información mutua y entropía (II)

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

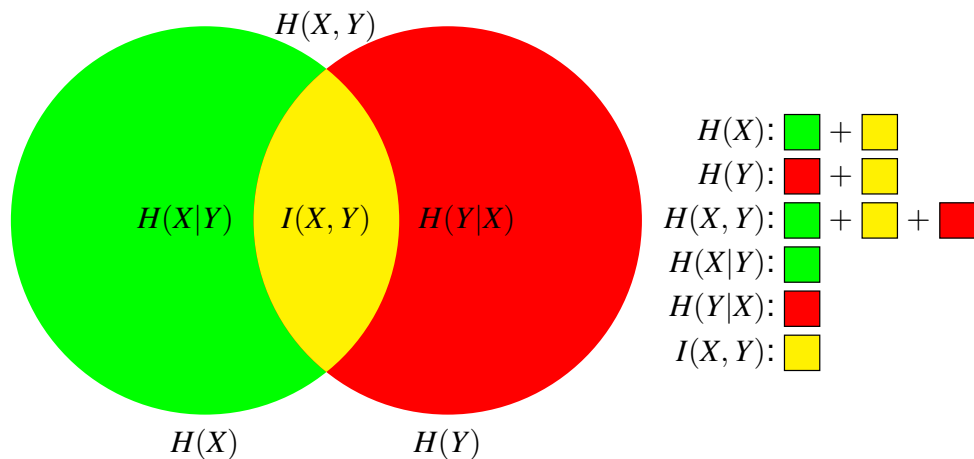
● Representación en un diagrama de Venn

- ▶ Entropías e información mutua representadas por áreas



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

Diagrama de Venn - Entropías e información mutua



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

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

Entropía diferencial

- Extensión de las definiciones de entropía a variables aleatorias continuas

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \log \frac{1}{f_X(x)} dx$$

Definición de *entropía diferencial conjunta*

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

Lo mismo se hace para la *entropía diferencial condicional*

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

A menudo se utiliza la definición alternativa pero equivalente

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

Entropía diferencial e información mutua - Relaciones

- Definición de la información mutua

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

- Se mantienen las mismas relaciones que para variables discretas

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

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

Entropías diferenciales e información mutua : Propiedades

- Las entropías diferenciales no mantienen las propiedades de las entropías para variables aleatorias discretas en cuanto a interpretación como cantidad de información

- ▶ Ejemplo: $X \sim \mathcal{U}(0, a)$

$$h(X) = \log(a)$$

- ▶ Ejemplo: $Y \sim \mathcal{N}(0, \sigma^2)$

$$h(Y) = \frac{1}{2} \log(2\pi e \sigma^2)$$

Se puede ver que dependiendo de los valores de a o σ^2 ambas entropías pueden tomar valores positivos, nulos o negativos (en función de si $a \gtrless 1$, o $\sigma^2 \gtrless \frac{1}{2\pi e}$)

- La información mutua sí mantiene esa interpretación intuitiva y las correspondientes propiedades. En particular

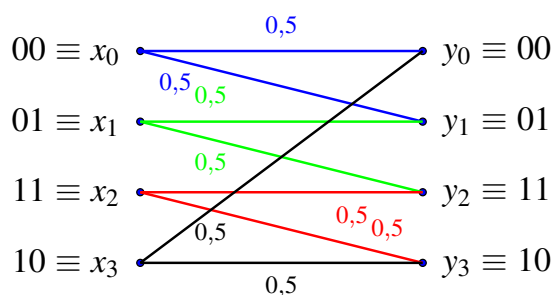
- ▶ $I(X, Y) \geq 0$ (función no negativa)
- ▶ $I(X, Y) = 0$ sólo si X e Y son independientes
- ▶ $I(X, Y) = I(Y, X)$

LÍMITES FUNDAMENTALES

EN

SISTEMAS DE COMUNICACIONES

Transmisión fiable sobre canales no fiables - Ejemplo

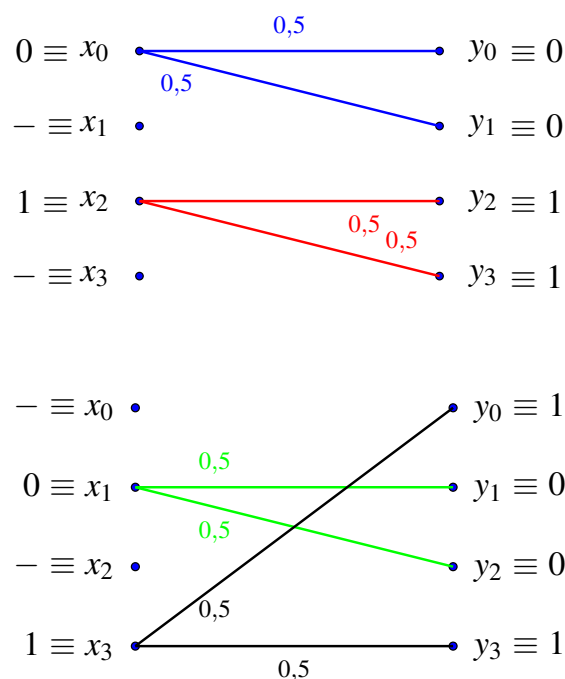


- 4 símbolos $\equiv$ 2 bits de información por uso del canal
- El canal no es fiable - Se comenten errores
 - ▶ Probabilidad de error de símbolo es $P_e = 1/2$
 - ▶ Con asignación binaria de Gray $BER = 1/4$
- Causa de los errores
 - ▶ Dado $Y = y_j$ no es posible identificar de forma unívoca $X = x_i$
 - ▶ Ejemplo: se observa $Y = y_0$
 - ★ El símbolo transmitido puede ser x_0 (sin error)
 - ★ El símbolo transmitido puede ser x_3 (con error)

Transmisión fiable sobre canales no fiables - Ejemplo (II)

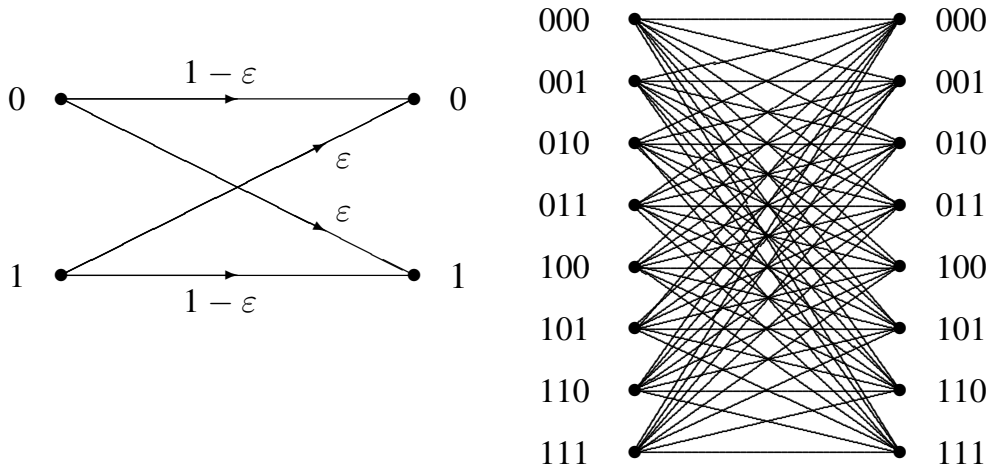
- Opción para transmitir información de forma fiable
 - ▶ Transmitir un subconjunto de los símbolos
 - ★ Símbolos que generen salidas “sin solapamiento”
 - ▶ Ejemplo: transmitir sólo x_0 y x_2
 - ★ x_0 puede dar como salidas y_0 o y_1
 - ★ x_2 puede dar como salidas y_2 o y_3
- Dada una salida no hay incertidumbre en el símbolo transmitido !!!
- Es posible transmitir información sobre este canal sin errores
 - ▶ Coste de la transmisión fiable
 - ★ Menos información transmitida por cada uso del canal
 - En este caso: 2 símbolos $\equiv$ 1 bit por uso del canal
- Los canales habituales no permiten esto de forma directa
 - ▶ Alternativa: forzar un comportamiento similar - Codificación de canal
 - ★ No se busca probabilidad de error nula (sin solapamientos)
 - ★ Se busca poder reducir la probabilidad de error de forma arbitraria
 - Solapamientos con probabilidad arbitrariamente baja

Transmisión fiable sobre canales no fiables - Ejemplo (III)



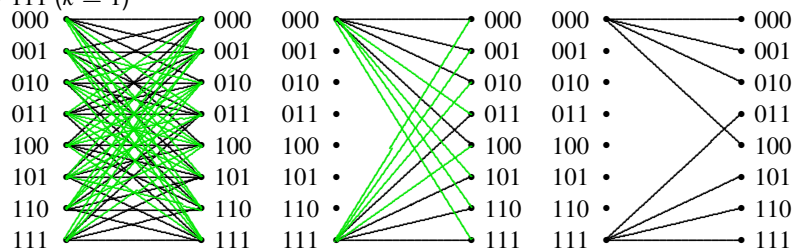
Codificación de canal

- Se utiliza el canal n veces de forma conjunta
 - Definición de símbolos extendidos: grupo de n símbolos
- Se busca un subconjunto de símbolos (2^k) que produzcan “*bajo solapamiento*” en la salida
 - Se transportan k bits de información por cada n usos del canal
- Ejemplo: canal binario simétrico ($BER = \varepsilon$) con $n = 3$



Codificación de canal (II)

- Situaciones más probables (para ε razonablemente bajo)
 - 0 errores o 1 error sobre 3 bits - 4 ramas/símbolo (en negro)
- Situaciones menos probables
 - 2 errores o 3 errores sobre 3 bits - 4 ramas/símbolo (en verde)
- Subconjunto de 2^k ($k < n$) elementos con “*bajo solapamiento*”
 - Ejemplo: 000 y 111 ($k = 1$)



- Despreciando los enlaces de “*baja probabilidad*” no hay solapamiento
 - Probabilidad de error: $P_e = 3 \times [\varepsilon^2 (1 - \varepsilon)] + \varepsilon^3$
 - Ejemplos: $\varepsilon = 0,1 \rightarrow P_e = 0,028$ | $\varepsilon = 10^{-3} \rightarrow P_e = 2,998 \times 10^{-6}$
 - Información transmitida: 1 bit (k) de información por cada 3 (n) usos del canal
 - Tasa de codificación: $R = k/n = 1/3$

- Intuición: aumentando n y k (con k/n constante) se puede reducir más

Existe un límite: **Capacidad de canal**

Codificación de canal (III)

- Ejemplo $n = 3$: 8 posibles tripletes de bits (símbolos extendidos)

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Subconjunto de $2^k = 2$ símbolos extendidos con “bajo solapamiento”

- No se solapan si: hay 0 o 1 bit erróneos en el triplete transmitido
- Información transmitida con cada triplete: 1 bit ($k = 1$)
- Tasa de codificación: $R = \frac{k}{n} = \frac{1}{3}$

$$\begin{aligned} 0 &\equiv \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ 1 &\equiv \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

$$P_a = \underbrace{(1 - \varepsilon)^3}_{P(0 \text{ err.})} + 3 \times \underbrace{[\varepsilon^1(1 - \varepsilon)^2]}_{P(1 \text{ err.})}, \quad P_e = 3 \times \underbrace{[\varepsilon^2(1 - \varepsilon)^1]}_{P(2 \text{ err.})} + \underbrace{\varepsilon^3}_{P(3 \text{ err.})}$$

Codificación de canal (IV)

- Ejemplo $n = 6$: 64 posibles tuplas de bits (símbolos extendidos): subconjunto de $2^k = 4$ con “bajo solapamiento”

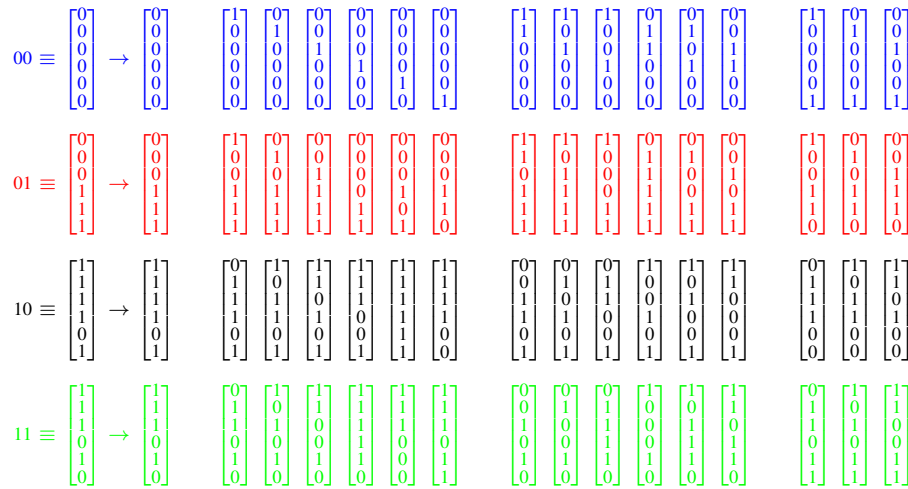
$$\begin{aligned} &\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \\ &\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

- No se solapan si:
 - Hay 0 o 1 bit erróneos en la 6-tupla transmitida
 - Hay dos errores en los 4 primeros bits de la 6-tupla transmitida
 - Hay dos errores: uno en el último bit, y otro en el 1º, 2º ó 3º

- Información transmitida con cada 6-tupla: 2 bit ($k = 2$): Tasa de codificación $R = \frac{k}{n} = \frac{2}{6} = \frac{1}{3}$

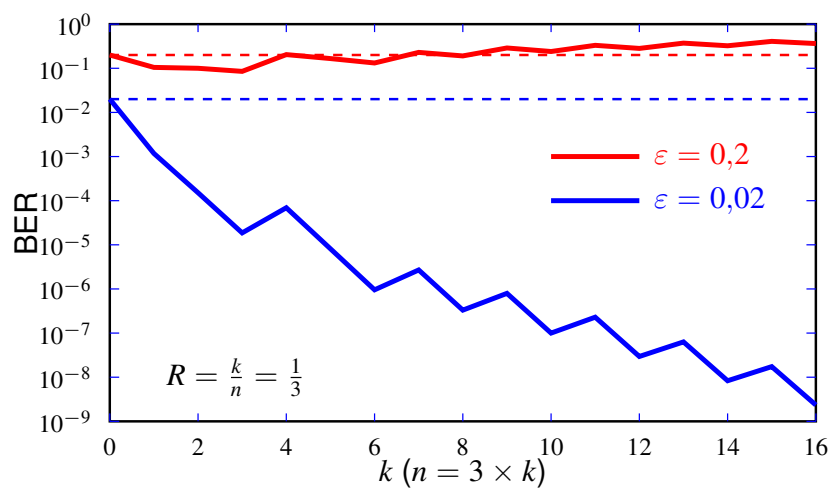
- Probabilidad de acierto: $P_a = \underbrace{(1 - \varepsilon)^6}_{P(0 \text{ err.})} + 6 \times \underbrace{[\varepsilon^1(1 - \varepsilon)^5]}_{P(1 \text{ err.})} + 9 \times \underbrace{[\varepsilon^2(1 - \varepsilon)^4]}_{P(2 \text{ err.})}$

Codificación de canal (V)



$$P_e = (15 - 9) \times \underbrace{\left[\varepsilon^2 (1 - \varepsilon)^4 \right]}_{P(2 \text{ err.})} + 20 \times \underbrace{\left[\varepsilon^3 (1 - \varepsilon)^3 \right]}_{P(3 \text{ err.})} + 15 \times \underbrace{\left[\varepsilon^4 (1 - \varepsilon)^2 \right]}_{P(4 \text{ err.})} + 6 \times \underbrace{\left[\varepsilon^5 (1 - \varepsilon) \right]}_{P(5 \text{ err.})} + \underbrace{\varepsilon^6}_{P(6 \text{ err.})}$$

Codificación de canal (VI)



● Capacidad del canal

$$\varepsilon = 0,2 \rightarrow C = 0,278 \quad (R > C) \quad \varepsilon = 0,02 \rightarrow C = 0,858 \quad (R < C)$$

Codificación de canal para detección y corrección de errores

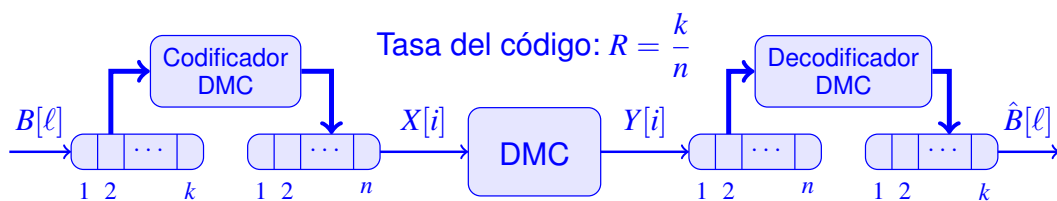
- Codificación de canal: introducción de redundancia estructurada
 - ▶ k símbolos de información se transportan transmitiendo n símbolos (símbolos extendidos, con $n > k$)
 - ★ Tasa del código: $R = \frac{k}{n}$
 - ★ Diccionario del código $\mathcal{C}(k, n)$

Ejemplo de diccionario del código para dos códigos binarios

Conj. de índices (k)	Palabras código (n)	Conj. de índices (k)	Palabras código (n)
0	000	00	000000
1	111	01	000111
		10	111101
		11	111010

Código de ejemplo $\mathcal{C}(1, 3)$

Código de ejemplo $\mathcal{C}(2, 6)$



Teorema de codificación de canal

Capacidad de Canal: C

Teorema de codificación de canal (Shannon 1948):

- 1 Si la tasa de transmisión R es menor que C , entonces para cualquier $\delta > 0$ existe un código con una longitud de bloque n suficiente cuya probabilidad de error es menor que δ
 - ▶ Codificación de canal: permite reducir la probabilidad de error hasta cualquier nivel arbitrariamente bajo
- 2 Si $R > C$, la probabilidad de error de cualquier código con cualquier longitud de bloque n está limitada por un valor no nulo
 - ▶ Codificación de canal: **NO** permite reducir la probabilidad de error hasta cualquier nivel arbitrariamente bajo
- 3 Existen códigos que permiten alcanzar la capacidad del canal $R = C$

Capacidad de canal

- Máxima cantidad de información que se puede transmitir de forma fiable a través de un canal de comunicaciones en un sistema digital de comunicaciones
 - ▶ En la transmisión se produce una distorsión
 - ★ Potencial pérdida de información
 - ▶ Transmisión fiable - Definición
 - ★ Ideal: transmisión sin pérdida potencial de información
 - ★ En la práctica: transmisión con capacidad para reducir la probabilidad de error tanto como sea necesario
 - ▶ Concepto de codificación de canal
 - ★ Mecanismo que permite una transmisión fiable
 - ▶ Capacidad de canal
 - ★ Límite en el número de símbolos extendidos con solapamiento arbitrariamente bajo cuando el número de usos del canal tiende a infinito
- Estudio de la capacidad de canal:
 - ▶ Canal digital (DMC)
 - ★ El canal digital binario se considera un caso particular
 - ▶ Canal gaussiano

DMC: Capacidad de canal a través de la información mutua

- Información mutua entre la entrada y la salida de un DMC

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

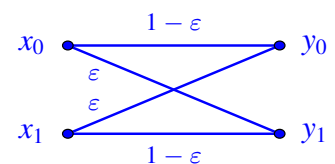
- Análisis para un BSC con $BER = \varepsilon$ en dos casos extremos

- ▶ Caso óptimo (canal ideal): $\varepsilon = 0$

$$H(X|Y) = 0 \rightarrow I(X, Y) = H(X)$$

- ▶ Peor caso: $\varepsilon = 1/2$

$$H(X|Y) = H(X) \rightarrow I(X, Y) = 0 \quad X \text{ e } Y \text{ independientes}$$



- Se pueden extraer las siguientes conclusiones

- 1 La información mutua entre entrada y salida del canal se puede ver como la cantidad de información que pasa de la entrada a la salida al usar el canal

- ★ En un canal ideal ($\varepsilon = 0$) pasa toda la información: $I(X, Y) = H(X)$

- ★ Si entrada y salida son independientes no pasa información: $I(X, Y) = 0$

- 2 $H(X|Y)$ puede interpretarse como la información que se “pierde” en el canal

- ★ En un canal ideal ($\varepsilon = 0$) la pérdida es nula: $H(X|Y) = 0$

- ★ Si entrada y salida son independientes, la pérdida es total: $H(X|Y) = H(X)$

- ★ $I(X, Y) = H(X) - H(X|Y)$: información a la entrada, menos la que se pierde

Capacidad de canal para un canal digital

- Definición formal para un DMC

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

- Sus unidades son bits (o bits por uso de canal)
- Maximización sobre $p_X(x_0), p_X(x_1), \dots, p_X(x_{M_X-1})$

- Valores límite

$$0 \leq C \leq \min\{\log M_X, \log M_Y\}$$

- $C \geq 0$
 - ★ Ya que $I(X, Y) \geq 0$
- $C \leq \log M_X$
 - ★ Ya que $I(X, Y) \leq H(X)$, y $H(X) \leq \log M_X$
- $C \leq \log M_Y$
 - ★ Ya que $I(X, Y) \leq H(Y)$, y $H(Y) \leq \log M_Y$

Capacidad: problema de maximización con restricciones

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

- Planteamiento: maximización de una función con restricciones

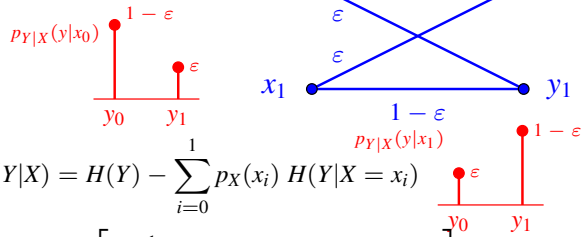
- Función a maximizar
 - ★ $I(X, Y)$
- Variables sobre las que se maximiza (incógnitas)
 - ★ $p_X(x_0), p_X(x_1), \dots, p_X(x_{M_X-1})$
- Restricciones (sobre las incógnitas)
 - ★ $0 \leq p_X(x_i) \leq 1$, para $i \in \{0, 1, \dots, M_X - 1\}$
 - ★ $\sum_{i=0}^{M_X-1} p_X(x_i) = 1$

- En general encontrar una solución analítica puede ser difícil

- Soluciones analíticas sólo para canales “*sencillos*”
- Cálculo mediante métodos numéricos utilizando ordenadores

Canal binario simétrico con $BER = \varepsilon$

- Capacidad: $C = \max_{p_X(x_i)} I(X, Y)$
- Cálculo de $I(X, Y)$



$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^1 p_X(x_i) H(Y|X = x_i)$$

$$= H(Y) - \sum_{i=0}^1 p_X(x_i) \left[- \sum_{j=0}^1 p_{Y|X}(y_j|x_i) \log p_{Y|X}(y_j|x_i) \right] = H(Y) - \Omega(\varepsilon)$$

$-\varepsilon \log(\varepsilon) - (1-\varepsilon) \log(1-\varepsilon) = \Omega(\varepsilon)$

- Cálculo de la capacidad de canal
 - ▶ Se busca el máximo de la información mutua
 - ★ Para este canal $H(Y|X) = \Omega(\varepsilon)$ no depende de $p_X(x_i)$
 - $I(X, Y)$ es máxima cuando $H(Y)$ es máxima
 - ★ $H(Y)$ es máxima si $p_Y(y_j) = \frac{1}{M_Y}$: $\max H(Y) = \log M_Y = 1$ bit/símb.

$$\left. \begin{aligned} p_Y(y_0) &= \frac{1}{2} = p_X(x_0)(1 - \varepsilon) + p_X(x_1) \varepsilon \\ p_Y(y_1) &= \frac{1}{2} = p_X(x_0) \varepsilon + p_X(x_1)(1 - \varepsilon) \end{aligned} \right\} \rightarrow p_X(x_0) = \frac{1}{2}, p_X(x_1) = \frac{1}{2}$$

$$C = 1 - \Omega(\varepsilon) \text{ bits/uso} \quad p_X(x_0) = p_X(x_1) = \frac{1}{2}$$

Límites para la transmisión en un canal digital

- Un canal digital tiene una capacidad C bits/uso
 - ▶ Si se utilizan códigos de canal, los códigos prácticos (aquellos que permiten reducir la probabilidad de error de forma arbitraria) deberán tener una tasa de codificación menor que dicha capacidad

$$R < C$$

- Limitación práctica en términos de velocidad de transmisión efectiva cuando se utiliza codificación para protección frente a errores

- ▶ Sistema diseñado para transmitir a R_b bits/s (tasa bruta)

- ★ Tasa efectiva:

$$R_b^{\text{efectiva}} = R \times R_b \text{ bits de información/s}$$

- ★ Límite en la tasa de transmisión efectiva

$$R_b^{\text{efectiva}} < C \times R_b \text{ bits de información/s}$$

Capacidad de canal para el canal gaussiano

- Modelo de relación entrada salida en un canal gaussiano

$$Y = X + Z$$

Z es una variable aleatoria gaussiana, de media nula y varianza P_Z

- Capacidad de canal en las siguientes condiciones:

- Potencia transmitida: P_X watt.
- Ancho de banda: B Hz
 - Potencia de ruido: $P_Z = N_0 B$ watt.

- Cálculo a través de la información mutua $C = \max_{f_X(x) \mid E[X^2] \leq P_X} I(X, Y)$ Restricción

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

- Resultado

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

Se obtiene para $f_X(x)$ gaussiana

Capacidad de canal para canal gaussiano (II)

Condiciones:

- Potencia transmitida: P_X watt.
- Ancho de banda: B Hz
 - Potencia de ruido: $P_Z = N_0 B$ watt.

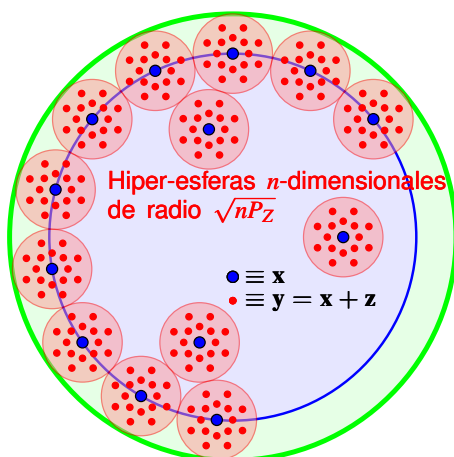
Capacidad: número de vectores para n usos sin solapamiento dado el ruido

$$M_{ss} = \left(1 + \frac{P_X}{P_Z} \right)^{n/2} \quad C = \frac{\log_2 M_{ss}}{n} = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{P_Z} \right)$$

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

Número de transmisiones por segundo: $2B$

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



Hiper-esfera n -dimensional: radio $\sqrt{nP_X}$

Hiper-esfera n -dim.: radio $\sqrt{n(P_X + P_Z)}$

Capacidad del canal gaussiano - Efecto de P_X y B

- La capacidad depende de dos parámetros de diseño

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

- Potencia de la señal transmitida, P_X
- Ancho de banda disponible en Hz, B

- Capacidad de canal en función de la potencia transmitida P_X

$$\lim_{P_X \rightarrow \infty} C = \infty$$

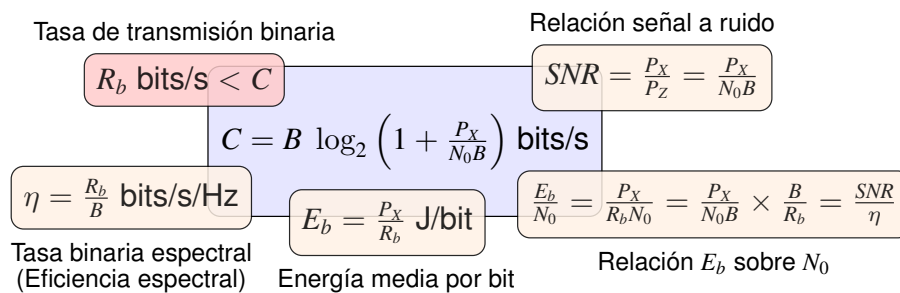
- Se puede aumentar C de forma arbitraria aumentando P_X
- Aumento lineal de C requiere aumento exponencial de P_X

- Capacidad de canal en función del ancho de banda (B Hz)

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

- El incremento de B no permite un incremento arbitrario de C

Límites para la transmisión en un canal gaussiano



- Sistema de comunicaciones práctico $R_b < C \rightarrow R_b < B \log_2(1 + SNR) \text{ bits/s}$

- Dividiendo por B en ambos lados y reordenando

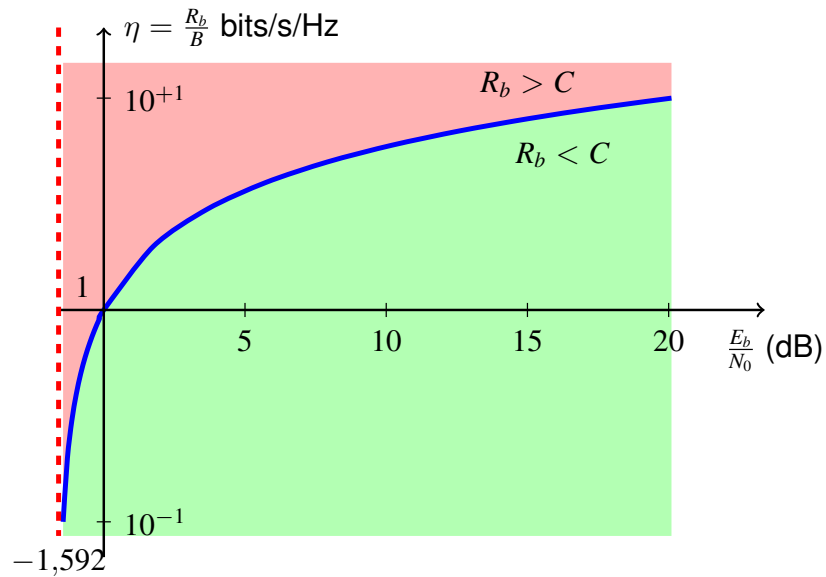
$$\eta < \log_2(1 + SNR), \quad \eta < \log_2 \left(1 + \eta \frac{E_b}{N_0} \right)$$

$$SNR > 2^\eta - 1, \quad \frac{E_b}{N_0} > \frac{2^\eta - 1}{\eta}$$

$$\text{Si } \eta \rightarrow 0 \text{ entonces } \frac{E_b}{N_0} = \ln 2 = 0,693 \approx -1,6 \text{ dB}$$

Tasa binaria espectral frente a E_b/N_0

- Se representa sobre el plano η vs $\frac{E_b}{N_0}$ la curva $\frac{E_b}{N_0} = \frac{2^\eta - 1}{\eta}$
 - Divide el plano en dos regiones: sistemas con $R_b < C$ (prácticos) y con $R_b > C$



Relación señal a ruido normalizada

- Cota inferior para SNR

$$SNR > 2^\eta - 1$$

- Definición de SNR normalizada

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

- Cota inferior sobre SNR_{norm}

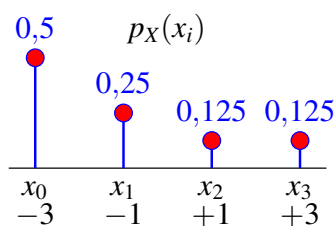
$$SNR_{norm} > 1 \text{ (0 dB)}$$

EJEMPLOS

DE

MEDIDAS DE INFORMACIÓN

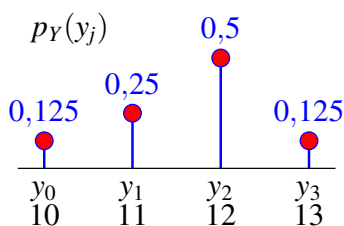
Ejemplo - Entropía



x_i	$x_0 = -3$	$x_1 = -1$	$x_2 = +1$	$x_3 = +3$
$p_X(x_i)$	$1/2$	$1/4$	$1/8$	$1/8$

$$H(X) = - \sum_{i=0}^{M_X-1} p_X(x_i) \log_2 p_X(x_i) = \sum_{i=0}^{M_X-1} p_X(x_i) \log_2 \frac{1}{p_X(x_i)}$$

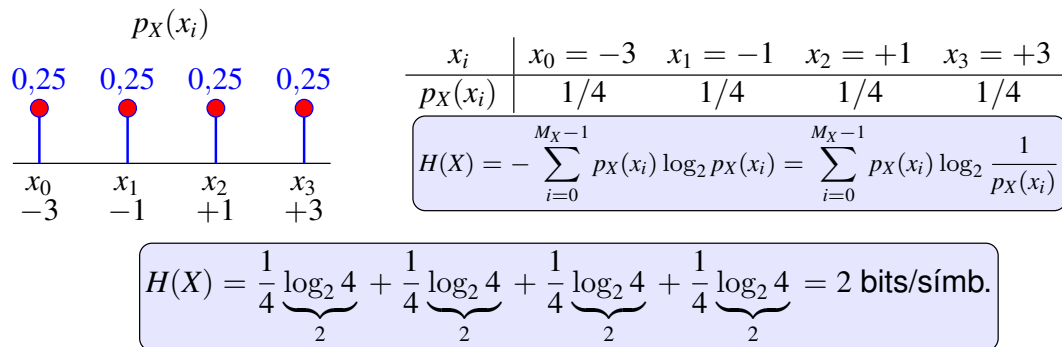
$$H(X) = \frac{1}{2} \underbrace{\log_2 2}_1 + \frac{1}{4} \underbrace{\log_2 4}_2 + \frac{1}{8} \underbrace{\log_2 8}_3 + \frac{1}{8} \underbrace{\log_2 8}_3 = \frac{7}{4} \text{ bits/símbolo}$$



y_j	$y_0 = 10$	$y_1 = 11$	$y_2 = 12$	$y_3 = 13$
$p_Y(y_j)$	$1/8$	$1/4$	$1/2$	$1/8$

$$H(Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) \log_2 \frac{1}{p_Y(y_j)} = ?$$

Ejemplo - Entropía : Símbolos equiprobables



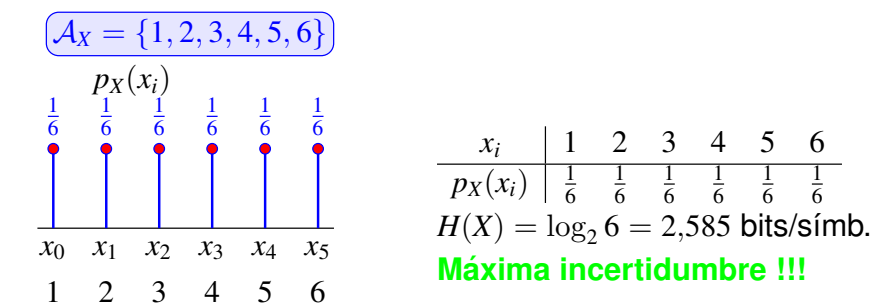
● Máximo valor de la entropía

$$H(X) \leq \log_2 M_X$$

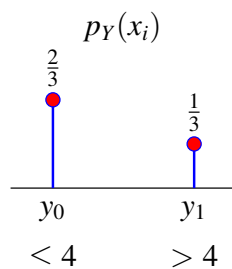
► Se obtiene para símbolos equiprobables

★ Situación de **máxima incertidumbre**

Ejemplo - Dado : Entropía



$$\mathcal{A}_Y = \{\leq 4, > 4\}$$



y_j	≤ 4	> 4
$p_Y(y_j)$	$\frac{2}{3}$	$\frac{1}{3}$

$$H(Y) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3}$$

$$= \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3$$

$$= \Omega\left(\frac{2}{3}\right) = \Omega\left(\frac{1}{3}\right) = 0,918 \text{ bits/símb.}$$

Menor incertidumbre !!!

Ejemplo - Dado : Entropía conjunta

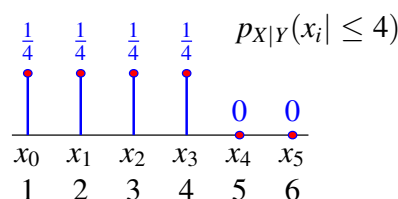
$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	

$$H(X, Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X,Y}(x_i, y_j)}$$

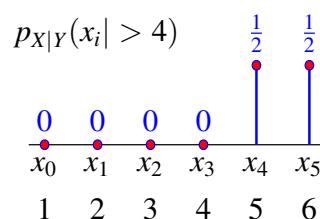
$$\begin{aligned} H(X, Y) &= \frac{1}{6} \log_2 6 + \frac{1}{6} \log_2 6 + \dots \\ &= 6 \times \frac{1}{6} \log_2 6 + \underbrace{6 \times 0 \log_2 \frac{1}{0}}_0 = \log_2 6 = 2,585 \text{ bits/símb.} \end{aligned}$$

$$H(X, Y) = 2,585 \neq H(X) + H(Y) = 3,503 \text{ bits/símb.}$$

Ejemplo - Dado : Entropía condicional $H(X|Y)$



$$\begin{aligned} H(X|Y = \leq 4) &= 4 \times \frac{1}{4} \log_2 4 \\ &= \log_2 4 = 2 \text{ bits/símb.} \end{aligned}$$



$$\begin{aligned} H(X|Y = > 4) &= 2 \times \frac{1}{2} \log_2 2 \\ &= \log_2 2 = 1 \text{ bits/símb.} \end{aligned}$$

Mitad de incertidumbre !!!

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) H(X|Y = y_j) = \frac{2}{3} \times 2 + \frac{1}{3} \times 1 = \frac{5}{3} \text{ bits/símb.}$$

Menor que $H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$

Ejemplo - Dado : Entropía condicional $H(X|Y)$

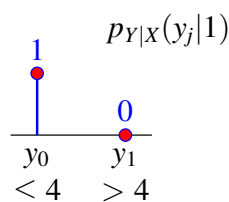
$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

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

$p_{X Y}(x_i y_j)$	1	2	3	4	5	6	$\sum_i p_{X Y}(x_i y_j)$
≤ 4	1/4	1/4	1/4	1/4	0	0	1
> 4	0	0	0	0	1/2	1/2	1

$$\begin{aligned}
 H(X|Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X|Y}(x_i|y_j)} = \frac{1}{6} \log_2 4 + \frac{1}{6} \log_2 4 + \dots \\
 &= 4 \times \frac{1}{6} \log_2 4 + 2 \times \frac{1}{6} \log_2 2 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_0 \\
 &= \frac{2}{3} \log_2 4 + \frac{1}{2} \log_2 2 = \frac{5}{3} \text{ bits/símb.}
 \end{aligned}$$

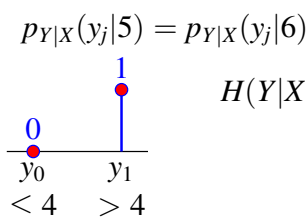
Ejemplo - Dado : Entropía condicional $H(Y|X)$



$$\begin{aligned}
 H(Y|X=1) &= -1 \log_2 1 - 0 \log_2 0 \\
 &= 0 \text{ bits/símb.}
 \end{aligned}$$

Sin incertidumbre !!!

$$\begin{aligned}
 p_{Y|X}(y_j|1) &= p_{Y|X}(y_j|2) = p_{Y|X}(y_j|3) = p_{Y|X}(y_j|4) \\
 H(Y|X=1) &= H(Y|X=2) = H(Y|X=3) = H(Y|X=4) = 0 \text{ bits/símb.}
 \end{aligned}$$



$$\begin{aligned}
 H(Y|X=5) &= H(Y|X=6) = -1 \log_2 1 - 0 \log_2 0 \\
 &= 0 \text{ bits/símb.}
 \end{aligned}$$

Sin incertidumbre !!!

$$H(Y|X) = 0 \text{ bits/símb.}$$

Ejemplo - Dado : Entropía condicional $H(Y|X)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

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

$p_{Y X}(y_j x_i)$	1	2	3	4	5	6
≤ 4	1	1	1	1	0	0
> 4	0	0	0	0	1	1
$\sum_j p_{Y X}(y_j x_i)$	1	1	1	1	1	1

$$\begin{aligned}
 H(Y|X) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{Y|X}(y_j|x_i)} \\
 &= 6 \times \frac{1}{6} \log_2 1 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_0 = 0 \text{ bits/símb.}
 \end{aligned}$$

Ejemplo - Dado : Información mutua $I(X, Y)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/6	1/6	1/6	1/6	0	0	2/3
> 4	0	0	0	0	1/6	1/6	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

$$I(X, Y) = \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i) p_Y(y_j)}$$

$$\begin{aligned}
 I(X, Y) &= \frac{1}{6} \log_2 \frac{\frac{1}{6}}{\frac{1}{6} \times \frac{2}{3}} + \dots + \frac{1}{6} \log_2 \frac{\frac{1}{6}}{\frac{1}{6} \times \frac{1}{3}} \\
 &= 4 \times \frac{1}{6} \log_2 \frac{3}{2} + 2 \times \frac{1}{6} \log_2 3 + 6 \times \underbrace{0 \log_2 \frac{1}{0}}_0
 \end{aligned}$$

$$I(X, Y) = \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3$$

$$I(X, Y) = \Omega\left(\frac{2}{3}\right) = \Omega\left(\frac{1}{3}\right) = 0,9183 \text{ bits/símb.}$$

Relaciones entre medidas cuantitativas de información

$H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$	$H(Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$
$H(X Y) = \frac{5}{3} \text{ bits/símb.}$	$H(Y X) = 0 \text{ bits/símb.}$
$H(X, Y) = \log_2 6 = 2,585 \text{ bits/símb.}$	$I(X, Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$

● Relaciones para entropía conjunta

$$H(X, Y) = H(X) + H(Y|X) = \log_2 6 + 0 = \log_2 6 = 2,585 \text{ bits/símb.}$$

$$H(X, Y) = H(Y) + H(X|Y) = \Omega(1/3) + \frac{5}{3} = 2,585 \text{ bits/símb.}$$

$$H(X, Y) \neq H(X) + H(Y) = 3,503 \text{ bits/símb. (NO Independientes !!!)}$$

Relaciones entre medidas cuantitativas de información (II)

$H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$	$H(Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$
$H(X Y) = \frac{5}{3} \text{ bits/símb.}$	$H(Y X) = 0 \text{ bits/símb.}$
$H(X, Y) = \log_2 6 = 2,585 \text{ bits/símb.}$	$I(X, Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$

● Relaciones para información mutua

$$I(X, Y) = H(Y) - H(Y|X) = \Omega(1/3) - 0 = 0,918 \text{ bits/símb.}$$

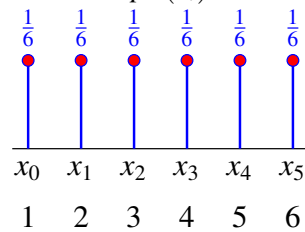
$$I(X, Y) = H(X) - H(X|Y) = \log_2 6 - \frac{5}{3} = 0,918 \text{ bits/símb.}$$

$$I(X, Y) = H(X) + H(Y) - H(X, Y) = \log_2 6 + \Omega(1/3) - \log_2 6 = 0,918 \text{ bits/símb.}$$

$$I(X, Y) \neq 0 \text{ bits/símb. (NO Independientes!!!)}$$

Ejemplo Dos Dados - Entropía

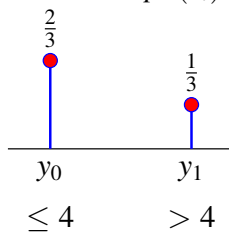
Dado 1 : $p_X(x_i)$



x_i	1	2	3	4	5	6
$p_X(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$$

Dado 2 : $p_Y(y_i)$



y_j	≤ 4	> 4
$p_Y(y_j)$	$\frac{2}{3}$	$\frac{1}{3}$

$$\begin{aligned} H(Y) &= -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \\ &= \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3 \\ &= \Omega\left(\frac{2}{3}\right) = \Omega\left(\frac{1}{3}\right) = 0,918 \text{ bits/símb.} \end{aligned}$$

Ejemplo Dos Dados - Entropía conjunta

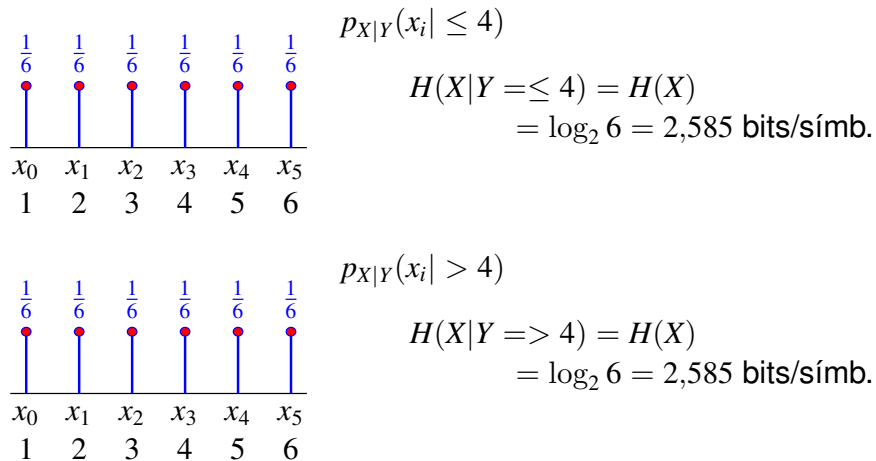
$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{3}$
> 4	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
$p_X(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$

$$\begin{aligned} H(X, Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X,Y}(x_i, y_j)} \\ &= 6 \times \frac{1}{9} \log_2 9 + 6 \times \frac{1}{18} \log_2 18 = 3,503 \text{ bits/símb.} \end{aligned}$$

X e Y son INDEPENDIENTES

$$H(X, Y) = H(X) + H(Y) = 2,585 + 0,918 = 3,503 \text{ bits/símb.}$$

Ejemplo Dos Dados - Entropía condicional $H(X|Y)$



X e Y son INDEPENDIENTES

$$H(X|Y) = \sum_{j=0}^{M_Y-1} p_Y(y_j) H(X|Y = y_j) = H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$$

Ejemplo Dos Dados - Entropía condicional $H(X|Y)$

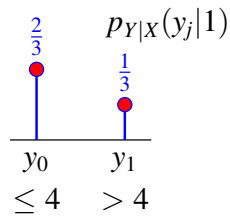
$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/9	1/9	1/9	1/9	1/9	1/9	2/3
> 4	1/18	1/18	1/18	1/18	1/18	1/18	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

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

$p_{X Y}(x_i y_j)$	1	2	3	4	5	6	$\sum_i p_{X Y}(x_i y_j)$
≤ 4	1/6	1/6	1/6	1/6	1/6	1/6	1
> 4	1/6	1/6	1/6	1/6	1/6	1/6	1

$$\begin{aligned}
 H(X|Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{X|Y}(x_i | y_j)} \\
 &= 6 \times \frac{1}{9} \log_2 6 + 6 \times \frac{1}{18} \log_2 6 = \log_2 6 \text{ bits/símb.}
 \end{aligned}$$

Ejemplo Dos Dados - Entropía condicional $H(Y|X)$



$$\begin{aligned} H(Y|X=1) &= H(Y) \\ &= \Omega(1/3) = 0,918 \text{ bits/símb.} \end{aligned}$$

$$p_{Y|X}(y_j|1) = p_{Y|X}(y_j|2) = p_{Y|X}(y_j|3) = p_{Y|X}(y_j|4) = p_{Y|X}(y_j|5) = p_{Y|X}(y_j|6)$$

$$H(Y|X=1) = H(Y|X=2) = H(Y|X=3) = H(Y) = 0,918 \text{ bits/símb.}$$

$$H(Y|X=4) = H(Y|X=5) = H(Y|X=6) = H(Y) = 0,918 \text{ bits/símb.}$$

X e Y son INDEPENDIENTES

$$H(Y|X) = \sum_{i=0}^{M_X-1} p_X(x_i) H(Y|X=x_i) = H(Y) = 0,918 \text{ bits/símb.}$$

Ejemplo Dos Dados - Entropía condicional $H(Y|X)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/9	1/9	1/9	1/9	1/9	1/9	2/3
> 4	1/18	1/18	1/18	1/18	1/18	1/18	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

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

$p_{Y X}(y_j x_i)$	1	2	3	4	5	6
≤ 4	2/3	2/3	2/3	2/3	2/3	2/3
> 4	1/3	1/3	1/3	1/3	1/3	1/3
$\sum_j p_{Y X}(y_j x_i)$	1	1	1	1	1	1

$$\begin{aligned} H(Y|X) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{1}{p_{Y|X}(y_j|x_i)} \\ &= 6 \times \frac{1}{9} \log_2 \frac{3}{2} + 6 \times \frac{1}{18} \log_2 3 \\ &= \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3 = \Omega(1/3) = 0,918 \text{ bits/símb.} \end{aligned}$$

Ejemplo Dos Dados - Información mutua $I(X, Y)$

$p_{X,Y}(x_i, y_j)$	1	2	3	4	5	6	$p_Y(y_j)$
≤ 4	1/9	1/9	1/9	1/9	1/9	1/9	2/3
> 4	1/18	1/18	1/18	1/18	1/18	1/18	1/3
$p_X(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	2/3

$$\begin{aligned}
 I(X, Y) &= \sum_{i=0}^{M_X-1} \sum_{j=0}^{M_Y-1} p_{X,Y}(x_i, y_j) \log_2 \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i) p_Y(y_j)} \\
 &= \frac{1}{9} \log_2 \frac{\frac{1}{9}}{\frac{1}{6} \times \frac{2}{3}} + \dots + \frac{1}{18} \log_2 \frac{\frac{1}{18}}{\frac{1}{6} \times \frac{1}{3}} \\
 &= 6 \times \frac{1}{18} \log_2 1 = 0 \text{ bits/símb.}
 \end{aligned}$$

X e Y son INDEPENDIENTES

$$I(X, Y) = 0 \text{ bits/símb.}$$

Relaciones entre medidas cuantitativas de información

● Ejemplo Dos Dados

$H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$	$H(Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$
$H(X Y) = \log_2 6 = 2,585 \text{ bits/símb.}$	$H(Y X) = \Omega(1/3) = 0,918 \text{ bits/símb.}$
$H(X, Y) = 3,503 \text{ bits/símb.}$	$I(X, Y) = 0 \text{ bits/símb.}$

● Relaciones para la entropía conjunta

$$H(X, Y) = H(X) + H(Y|X) = \log_2 6 + \Omega(1/3) = 3,503 \text{ bits/símb.}$$

$$H(X, Y) = H(Y) + H(X|Y) = \Omega(1/3) + \frac{5}{3} = 3,503 \text{ bits/símb.}$$

$$H(X, Y) = H(X) + H(Y) = 3,503 \text{ bits/símb. (INDEPENDIENTES!!!)}$$

Relaciones entre medidas cuantitativas de información (II)

- Ejemplo Dos Dados

$$H(X) = \log_2 6 = 2,585 \text{ bits/símb.}$$

$$H(Y) = \Omega(1/3) = 0,918 \text{ bits/símb.}$$

$$H(X|Y) = \log_2 6 = 2,585 \text{ bits/símb.}$$

$$H(Y|X) = \Omega(1/3) = 0,918 \text{ bits/símb.}$$

$$H(X, Y) = 3,503 \text{ bits/símb.}$$

$$I(X, Y) = 0 \text{ bits/símb.}$$

- Relaciones para la información mutua

$$I(X, Y) = H(Y) - H(Y|X) = \Omega(1/3) - \Omega(1/3) = 0 \text{ bits/símb.}$$

$$I(X, Y) = H(X) - H(X|Y) = \log_2 6 - \log_2 6 = 0 \text{ bits/símb.}$$

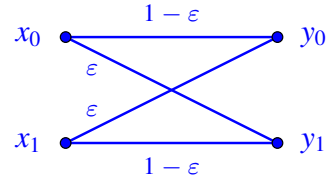
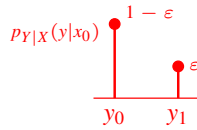
$$I(X, Y) = H(X) + H(Y) - H(X, Y) = \log_2 6 + \Omega(1/3) - 3,503 = 0 \text{ bits/símb.}$$

$$I(X, Y) = 0 \text{ bits/símb. (INDEPENDIENTES!!!)}$$

EJEMPLOS DE CAPACIDAD DE CANAL

Canal binario simétrico con $BER = \varepsilon$

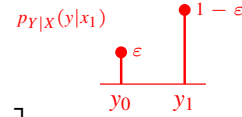
- Capacidad: $C = \max_{p_X(x_i)} I(X, Y)$
- Incógnitas: $p_X(x_0), p_X(x_1)$
- Evaluación de $I(X, Y)$



$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^1 p_X(x_i) H(Y|X = x_i)$$

$$= H(Y) - \sum_{i=0}^1 p_X(x_i) \left[- \sum_{j=0}^1 p_{Y|X}(y_j|x_i) \log p_{Y|X}(y_j|x_i) \right] = H(Y) - \Omega(\varepsilon)$$

$-\varepsilon \log(\varepsilon) - (1-\varepsilon) \log(1-\varepsilon) = \Omega(\varepsilon)$



● Capacidad del canal

- ▶ Máximo de la información mutua

- ★ $H(Y|X) = \Omega(\varepsilon)$ no depende de $p_X(x_i)$ para este canal
- $I(X, Y)$ es máxima cuando $H(Y)$ es máxima
- ★ $H(Y)$ es máxima si $p_Y(y_j) = \frac{1}{M_Y}$: $\max H(Y) = \log M_Y = 1$ bit/símb.

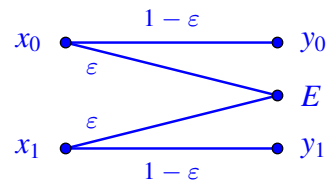
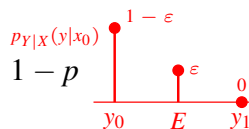
$$\left. \begin{aligned} p_Y(y_0) &= \frac{1}{2} = p_X(x_0)(1 - \varepsilon) + p_X(x_1) \varepsilon \\ p_Y(y_1) &= \frac{1}{2} = p_X(x_0) \varepsilon + p_X(x_1)(1 - \varepsilon) \end{aligned} \right\} \rightarrow p_X(x_0) = \frac{1}{2}, p_X(x_1) = \frac{1}{2}$$

$$C = 1 - \Omega(\varepsilon) \text{ bits/uso}$$

$$p_X(x_0) = p_X(x_1) = \frac{1}{2}$$

Binary Erasure Channel (BEC)

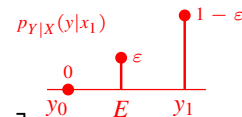
- Capacidad $C = \max_{p_X(x_i)} I(X, Y)$
- Incógnitas: $p_X(x_0) = p, p_X(x_1) = 1 - p$
- Evaluación de $I(X, Y)$



$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^1 p_X(x_i) H(Y|X = x_i)$$

$$= H(Y) - \sum_{i=0}^1 p_X(x_i) \left[- \sum_{j=0}^1 p_{Y|X}(y_j|x_i) \log p_{Y|X}(y_j|x_i) \right] = H(Y) - \Omega(\varepsilon)$$

$-\varepsilon \log(\varepsilon) - (1-\varepsilon) \log(1-\varepsilon) = \Omega(\varepsilon)$



● Capacidad de canal

- ▶ Máximo de la información mutua

- ★ $H(Y|X) = \Omega(\varepsilon)$ no depende de $p_X(x_i)$ para este canal
- $I(X, Y)$ es máxima cuando $H(Y)$ es máxima

- ★ Probabilidades $p_Y(y_j)$ para calcular $H(Y)$

$$\begin{cases} p_Y(y_0) = p(1 - \varepsilon) \\ p_Y(E) = p\varepsilon + (1 - p)\varepsilon = \varepsilon \\ p_Y(y_1) = (1 - p)(1 - \varepsilon) \end{cases}$$

$$H(Y) = -p(1 - \varepsilon) \log[p(1 - \varepsilon)] - (1 - p)(1 - \varepsilon) \log[(1 - p)(1 - \varepsilon)] - \varepsilon \log \varepsilon$$

Binary Erasure Channel (BEC)

$$\begin{aligned}
 H(Y) &= -p(1-\varepsilon) \log[p(1-\varepsilon)] \\
 &\quad - (1-p)(1-\varepsilon) \log[(1-p)(1-\varepsilon)] \\
 &\quad - \varepsilon \log \varepsilon \\
 &= -p(1-\varepsilon) [\log p + \log(1-\varepsilon)] \\
 &\quad - (1-p)(1-\varepsilon) [\log(1-p) + \log(1-\varepsilon)] \\
 &\quad - \varepsilon \log \varepsilon \\
 &= -p(1-\varepsilon) \log p - (1-p)(1-\varepsilon) \log(1-p) \\
 &\quad - \underbrace{[p(1-\varepsilon) + (1-p)(1-\varepsilon)]}_{(1-\varepsilon)} \log(1-\varepsilon) \\
 &\quad - \varepsilon \log \varepsilon \\
 &= (1-\varepsilon) \Omega(p) + \Omega(\varepsilon)
 \end{aligned}$$

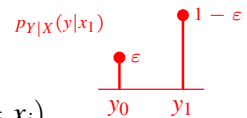
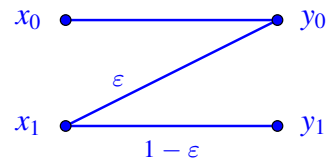
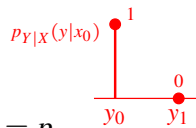
$$I(X, Y) = H(Y) - \Omega(\varepsilon) = (1-\varepsilon) \Omega(p)$$

$$C = 1 - \varepsilon \text{ bits/usd}$$

$$p_X(x_0) = p = \frac{1}{2}, p_X(x_1) = 1 - p = \frac{1}{2}$$

Canal Z con prob. de error ε

- Capacidad: $C = \max_{p_X(x_i)} I(X, Y)$
- Incógnitas: $p_X(x_0) = 1 - p, p_X(x_1) = p$
- Evaluación de $I(X, Y)$



$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=0}^1 p_X(x_i) H(Y|X = x_i)$$

$$\begin{aligned}
 H(Y|X = x_0) &= 0, & H(Y|X = x_1) &= \Omega(\varepsilon) \text{ bits/simb.} \\
 &= H(Y) - p\Omega(\varepsilon)
 \end{aligned}$$

$$\left\{ \begin{aligned} p_Y(y_0) &= 1 - p + p\varepsilon = 1 - p(1 - \varepsilon) \\ p_Y(y_1) &= p(1 - \varepsilon) \end{aligned} \right\} \rightarrow H(Y) = \Omega(p(1 - \varepsilon))$$

$$\begin{aligned}
 I(X, Y) &= H(Y) - H(Y|X) = \Omega(p(1 - \varepsilon)) - p\Omega(\varepsilon) \\
 &= -p(1 - \varepsilon) \log[p(1 - \varepsilon)] - (1 - p(1 - \varepsilon)) \log[1 - p(1 - \varepsilon)] - p\Omega(\varepsilon)
 \end{aligned}$$

Canal Z con prob. de error ε

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) = \Omega(p(1-\varepsilon)) - p\Omega(\varepsilon) \\ &= -p(1-\varepsilon) \log[p(1-\varepsilon)] - (1-p(1-\varepsilon)) \log[1-p(1-\varepsilon)] - p\Omega(\varepsilon) \end{aligned}$$

$$\Omega(x) = -x \log x - (1-x) \log(1-x) \rightarrow \frac{\partial \Omega(x)}{\partial x} = \log \frac{1-x}{x}$$

$$\frac{\partial I(X, Y)}{\partial p} = (1-\varepsilon) \log \frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} - \Omega(\varepsilon) = 0$$

$$\log \frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} = \frac{\Omega(\varepsilon)}{1-\varepsilon} \rightarrow \frac{1-p(1-\varepsilon)}{p(1-\varepsilon)} = 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}} \rightarrow p = \frac{1}{(1-\varepsilon) \left(1 + 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}$$

$$C = \Omega\left(\frac{1}{1 + 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}}\right) - \frac{1}{(1-\varepsilon) \left(1 + 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)} \Omega(\varepsilon) \text{ bits/uso}$$

$$p_X(x_0) = 1-p = 1 - \frac{1}{(1-\varepsilon) \left(1 + 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}, \quad p_X(x_1) = p = \frac{1}{(1-\varepsilon) \left(1 + 2^{\frac{\Omega(\varepsilon)}{1-\varepsilon}}\right)}$$

Otros canales (simétricos)

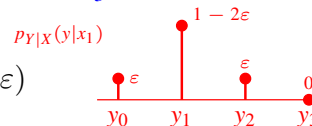
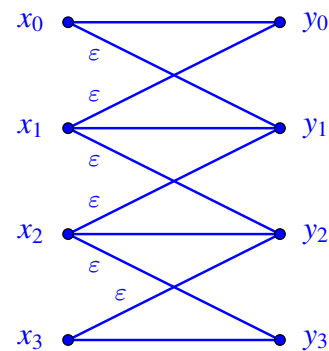
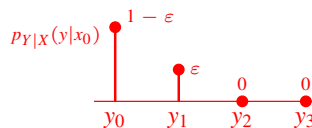
- Capacidad: $C = \max_{p_X(x_i)} I(X, Y)$
- Incógnitas: $p_i = p_X(x_i), i \in \{0, 1, 2, 3\}$
- Evaluación de $I(X, Y) = H(Y) - H(Y|X)$

$$H(Y|X = x_0) = -\varepsilon \log \varepsilon - (1-\varepsilon) \log(1-\varepsilon) = \Omega(\varepsilon)$$

$$\begin{aligned} H(Y|X = x_1) &= -2 \times \varepsilon \log \varepsilon - (1-2\varepsilon) \log(1-2\varepsilon) \\ &= \Omega(2\varepsilon) + 2\varepsilon \end{aligned}$$

$$H(Y|X = x_2) = H(Y|X = x_1), \quad H(Y|X = x_3) = H(Y|X = x_0)$$

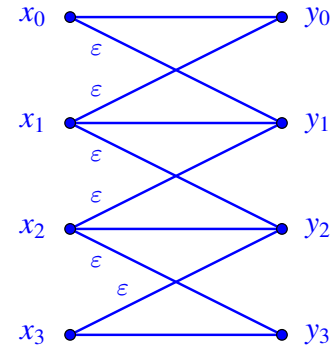
$$H(Y|X) = (p_0 + p_3)\Omega(\varepsilon) + (p_1 + p_2)[\Omega(2\varepsilon) + 2\varepsilon]$$



Otros canales (simétricos)

$$p_X(x_0) = p_0, p_X(x_1) = p_1, p_X(x_2) = p_2, p_X(x_3) = p_3$$

$$\left\{ \begin{array}{l} p_Y(y_0) = p_0(1 - \varepsilon) + p_1\varepsilon \\ p_Y(y_1) = p_1(1 - 2\varepsilon) + (p_0 + p_2)\varepsilon \\ p_Y(y_2) = p_2(1 - 2\varepsilon) + (p_1 + p_3)\varepsilon \\ p_Y(y_3) = p_3(1 - \varepsilon) + p_2\varepsilon \end{array} \right\}$$



$$\begin{aligned} H(Y) = & -(p_0(1 - \varepsilon) + p_1\varepsilon) \log(p_0(1 - \varepsilon) + p_1\varepsilon) \\ & -(p_1(1 - 2\varepsilon) + (p_0 + p_2)\varepsilon) \log(p_1(1 - 2\varepsilon) + (p_0 + p_2)\varepsilon) \\ & -(p_2(1 - 2\varepsilon) + (p_1 + p_3)\varepsilon) \log(p_2(1 - 2\varepsilon) + (p_1 + p_3)\varepsilon) \\ & -(p_3(1 - \varepsilon) + p_2\varepsilon) \log(p_3(1 - \varepsilon) + p_2\varepsilon) \end{aligned}$$

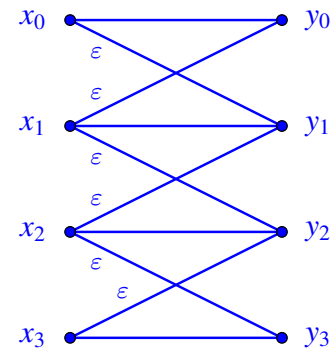
$$H(Y|X) = (p_0 + p_3)\Omega(\varepsilon) + (p_1 + p_2) [\Omega(2\varepsilon) + 2\varepsilon]$$

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

Otros canales (simétricos) - Simetría

$$p_X(x_0) = p_3, p_X(x_1) = p_2, p_X(x_2) = p_1, p_X(x_3) = p_0$$

$$\left\{ \begin{array}{l} p_Y(y_0) = p_3(1 - \varepsilon) + p_2\varepsilon \\ p_Y(y_1) = p_2(1 - 2\varepsilon) + (p_3 + p_1)\varepsilon \\ p_Y(y_2) = p_1(1 - 2\varepsilon) + (p_2 + p_0)\varepsilon \\ p_Y(y_3) = p_0(1 - \varepsilon) + p_1\varepsilon \end{array} \right\}$$



$$\begin{aligned} H(Y) = & -(p_0(1 - \varepsilon) + p_1\varepsilon) \log(p_0(1 - \varepsilon) + p_1\varepsilon) \\ & -(p_1(1 - 2\varepsilon) + (p_0 + p_2)\varepsilon) \log(p_1(1 - 2\varepsilon) + (p_0 + p_2)\varepsilon) \\ & -(p_2(1 - 2\varepsilon) + (p_1 + p_3)\varepsilon) \log(p_2(1 - 2\varepsilon) + (p_1 + p_3)\varepsilon) \\ & -(p_3(1 - \varepsilon) + p_2\varepsilon) \log(p_3(1 - \varepsilon) + p_2\varepsilon) \end{aligned}$$

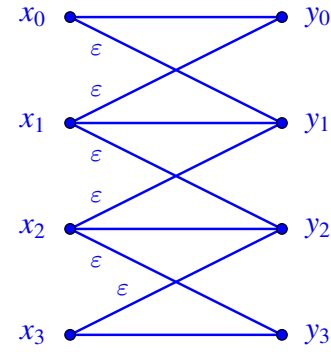
$$H(Y|X) = (p_0 + p_3)\Omega(\varepsilon) + (p_1 + p_2) [\Omega(2\varepsilon) + 2\varepsilon]$$

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

Otros canales (simétricos) - Parametrización

$$p_X(x_0) = p_X(x_3) = \frac{p}{2}, \quad p_X(x_1) = p_X(x_2) = \frac{1-p}{2}$$

$$\left\{ \begin{array}{l} p_Y(y_0) = \frac{p}{2}(1-\varepsilon) + \frac{1-p}{2}\varepsilon = \frac{p+\varepsilon-2p\varepsilon}{2} \\ p_Y(y_1) = \frac{1-p}{2}(1-2\varepsilon) + \frac{1}{2}\varepsilon = \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ p_Y(y_2) = \frac{1-p}{2}(1-2\varepsilon) + \frac{1}{2}\varepsilon = \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ p_Y(y_3) = \frac{p}{2}(1-\varepsilon) + \frac{1-p}{2}\varepsilon = \frac{p+\varepsilon-2p\varepsilon}{2} \end{array} \right\}$$



$$\begin{aligned} H(Y) &= -2 \times \frac{p+\varepsilon-2p\varepsilon}{2} \log \frac{p+\varepsilon-2p\varepsilon}{2} \\ &\quad -2 \times \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \log \frac{1-(p+\varepsilon-2p\varepsilon)}{2} \\ &= -(p+\varepsilon-2p\varepsilon)[\log(p+\varepsilon-2p\varepsilon) - \log 2] \\ &\quad -(1-(p+\varepsilon-2p\varepsilon))[\log(1-(p+\varepsilon-2p\varepsilon)) - \log 2] \\ &= 1 + \Omega(p+\varepsilon-2p\varepsilon) \end{aligned}$$

$$H(Y|X) = p\Omega(\varepsilon) + (1-p)[\Omega(2\varepsilon) + 2\varepsilon]$$

$$I(X, Y) = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)[\Omega(2\varepsilon) + 2\varepsilon]$$

Otros canales (simétricos) - Capacidad

$$I(X, Y) = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)[\Omega(2\varepsilon) + 2\varepsilon]$$

$$\frac{\partial I(X, Y)}{\partial p} = (1-2\varepsilon) \log \frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} - \Omega(\varepsilon) + \Omega(2\varepsilon) + 2\varepsilon = 0$$

$$\log \frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} = \frac{\Omega(\varepsilon) - \Omega(2\varepsilon) - 2\varepsilon}{1-2\varepsilon} \rightarrow \frac{1-(p+\varepsilon-2p\varepsilon)}{p+\varepsilon-2p\varepsilon} = 2^{\frac{\Omega(\varepsilon) - \Omega(2\varepsilon) - 2\varepsilon}{1-2\varepsilon}}$$

$$p = \frac{1-\varepsilon \left(1 + 2^{\frac{\Omega(\varepsilon) - \Omega(2\varepsilon) - 2\varepsilon}{1-2\varepsilon}}\right)}{(1-2\varepsilon) \left(1 + 2^{\frac{\Omega(\varepsilon) - \Omega(2\varepsilon) - 2\varepsilon}{1-2\varepsilon}}\right)}$$

$$C = 1 + \Omega(p+\varepsilon-2p\varepsilon) - p\Omega(\varepsilon) - (1-p)[\Omega(2\varepsilon) + 2\varepsilon] \text{ bits/uso}$$

$$p_X(x_0) = p_X(x_3) = \frac{p}{2}, \quad p_X(x_1) = p_X(x_2) = \frac{1-p}{2}$$