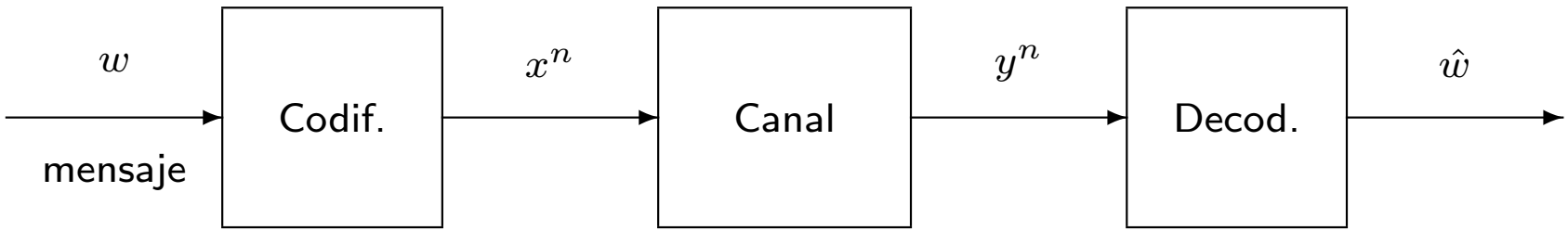
Introducción a la Teoría de la Información

Capacidad del canal

Facultad de Ingeniería, UdelaR

22 de mayo de 2020

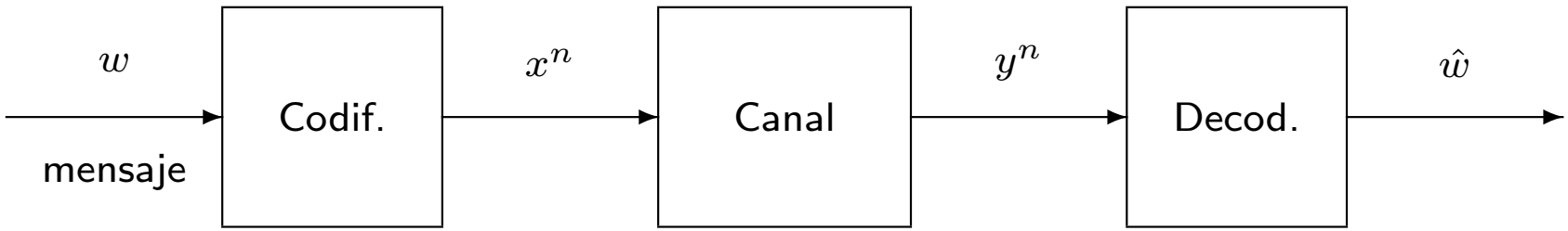
Modelo matemático de sistema de comunicación



Transmisión de un mensaje:

1. **fuente de datos** genera *mensaje* $w \in \mathcal{W}$,
2. **codificador** mapea w a *palabra código* $x^n \in \mathcal{X}^n$,
3. x^n se distorsiona al pasar por **canal**, resultando en $y^n \in \mathcal{Y}^n$,
4. **decodificador** infiere mensaje enviado $\hat{w}$ a partir de y^n.
5. Transmisión exitosa si $\hat{w} = w$.

Sinopsis



Capacidad del canal

- Cuánta información puede transmitirse por uso del canal
- Es una propiedad exclusiva del canal

Segundo teorema de Shannon

- Es posible transmitir por un canal ruidoso con probabilidad de error arbitrariamente pequeña

Importancia

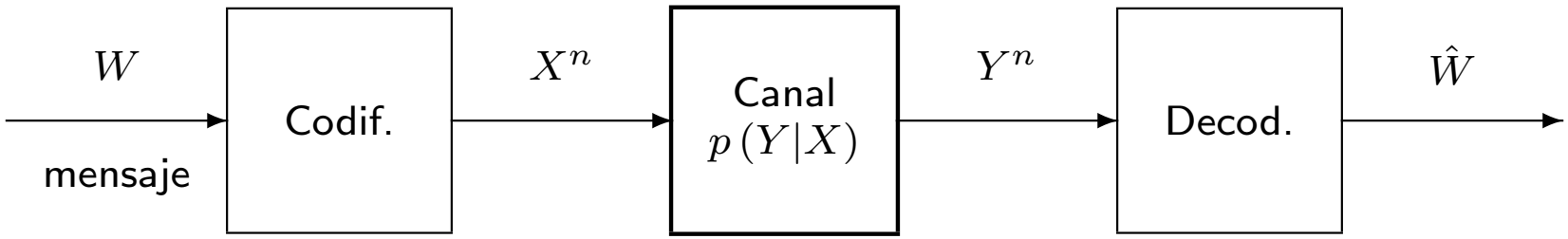
- Uno de los teoremas con más impacto práctico: la era digital
- Internet, celulares, telecomunicaciones, videoconferencias!, etc.
- Transmisión sin error a tasas no triviales se creía imposible

Temas

- Capacidad de canal: informacional vs. operativa
- El segundo teorema de Shannon
- Códigos prácticos de corrección de errores
- Canales con realimentación
- Separabilidad y codificación conjunta fuente-canal

Capacidad de canal

Canal discreto sin memoria (DMC)



Definición (Canal discreto sin memoria C (DMC))

- Canal: terna $(\mathcal{X}, p(Y|X), \mathcal{Y})$
- Entrada al canal: $X \in \mathcal{X}$, (alfabeto de entrada: $\mathcal{X}$)
- Salida del canal: $Y \in \mathcal{Y}$, (alfabeto de salida: $\mathcal{Y}$)
- Alfabetos finitos: $|\mathcal{X}| < \infty$ y $|\mathcal{Y}| < \infty$
- Sin memoria: Y_k *sólo depende de* X_k,
  $p(Y_k = y_k | X_k = x_k, X_{k-1} = x_{k-1}, Y_{k-1} = y_{k-1} \ldots) = p(Y_k = y_k | X_k = x_k)$
- Matriz de transición de prob.: $\Pi \triangleq \{\pi_{ij} = p(Y = j | X = i)\}$

Código de canal (M,n)

Definición (Código de canal (M, n))

- Conjunto de índices $J = \{1, 2, \ldots, M\}$ (mensajes W)
- Función de codificación: $X^n : J \to \mathcal{X}^n$.
- Función de decodificación: $g : \mathcal{Y}^n \to J$.

Definición (Codebook)

Se le llama codebook al conjunto imagen de X^n, es decir: $\{X^n(1) \ldots X^n(M)\}$.

Probabilidad de error de un código para un canal

- Error asociado a mensaje i ($\mathbf{1}(\cdot)$ es la función indicatriz):

$$\begin{aligned} \lambda_i &= \Pr\{g(Y^n) \neq i | X^n = X^n(i)\} \\ &= \sum_{y^n} p\left(y^n | x^n(i)\right) \mathbf{1}(g(y^n) \neq i) \end{aligned}$$

- **Error maximal**

$$\lambda^{(n)} = \max_{i \in J} \lambda_i$$

- **Error promedio**:

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$$

Capacidad Operativa

Definición (Tasa R de un código (M, n))

$$R = \frac{\log M}{n}$$

- Medida en bits por uso del canal (log. en base 2)
- Una tasa R se dice alcanzable para un canal si existe una secuencia de códigos $\left(\lceil 2^{nR} \rceil, n\right)$ tal que $\lambda^{(n)} \to 0$ cuando $n \to \infty$.

Definición (Capacidad operativa del canal)

$$C' = \sup\{R : R \text{ es alcanzable}\}$$

Capacidad Informacional de un canal

Definimos la capacidad informacional de un canal como

$$C = \max_{p(x) \in \mathcal{P}} I(X;Y),$$

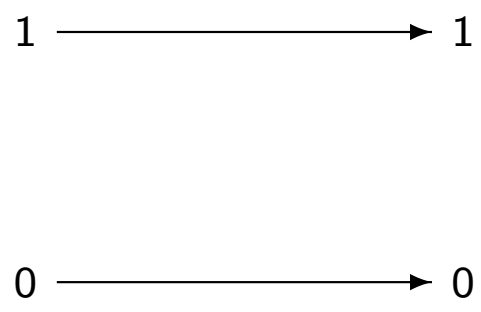
donde $\mathcal{P}$ es el espacio de todas las distribuciones de probabilidad sobre $\mathcal{X}$ para X.

- $0 \leq I(X;Y) \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$.
- El máximo en la definición de C existe, porque $I(X;Y)$ está acotada y es una función continua de $p(x)$ en el dominio compacto $\mathcal{P}$.
- En general, C no tiene forma cerrada; se calcula numéricamente.
- $\mathcal{P}$ convexo, $I(X;Y)$ cóncava en $\mathcal{P}$
  $\Rightarrow C$ **se puede calcular eficientemente.**

Capacidad informacional de un canal sin ruido

$$\Pi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

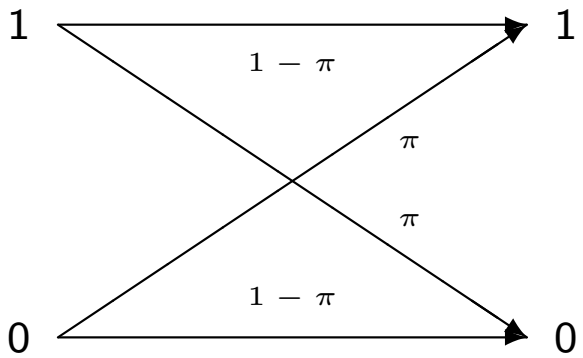
$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} \left(H(X) - H(X|Y)\right) \\ &= \max_{p(x)} \left(H(X) - 0\right) = 1 \end{aligned}$$



Capacidad informacional del canal binario simétrico (BSC)

$$
\begin{aligned}
I(X;Y) = & \; H(Y) - H(Y|X) \\
= & \; H(Y) - \sum_x p(x) \, H(Y|X=x) \\
\leq & \; 1 - \sum_x p(x) \, H(Y|X=x) \\
\overset{sim.}{=} & \; 1 - \sum_x p(x) H(\pi) \\
= & \; 1 - H(\pi)
\end{aligned}
$$

$$
\Pi = \begin{bmatrix} 1-\pi & \pi \\ \pi & 1-\pi \end{bmatrix}
$$

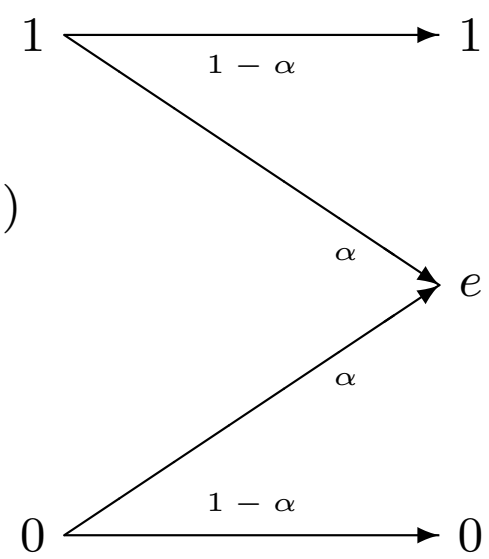


El máximo se alcanza en $p(x)$ uniforme.

Capacidad informacional del canal con borraduras

$$\Pi = \begin{bmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1-\alpha & \alpha \end{bmatrix}$$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - \sum_y p(y)\, H(X|Y=y) \\ &= H(X) - p(e)\, H(X|Y=e) - \\ &\quad p(0)\, H(X|Y=0) - p(1)\, H(X|Y=1) \\ &= (1-\alpha)\, H(X) \end{aligned}$$



Máximo $C = (1-\alpha)$, alcanzado para $p(x)$ uniforme.

Teorema de codificación del canal

Sinopsis

Dos definiciones de capacidad:

- Informacional:

$$C = \max_{p(x)} I\left(X;Y\right)$$

- Operacional:

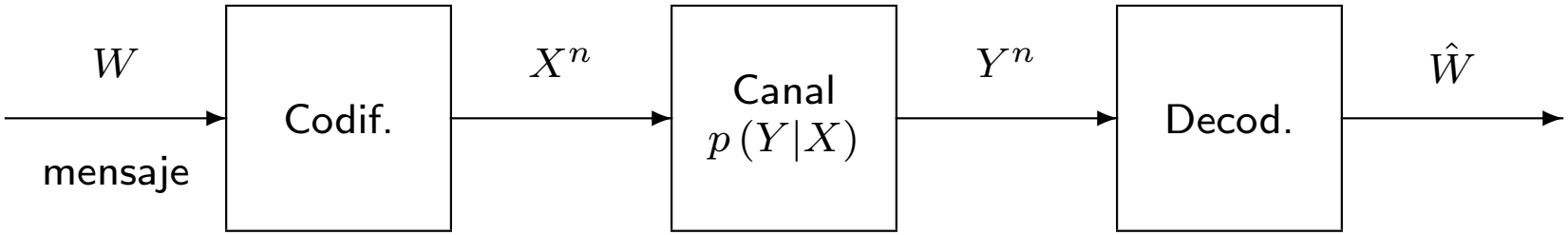
$$C' = \sup\{R : R \text{ es alcanzable}\}$$

Esencia del segundo teorema de Shannon:

- Ambas coinciden:

$$C = C'$$

Planteo formal



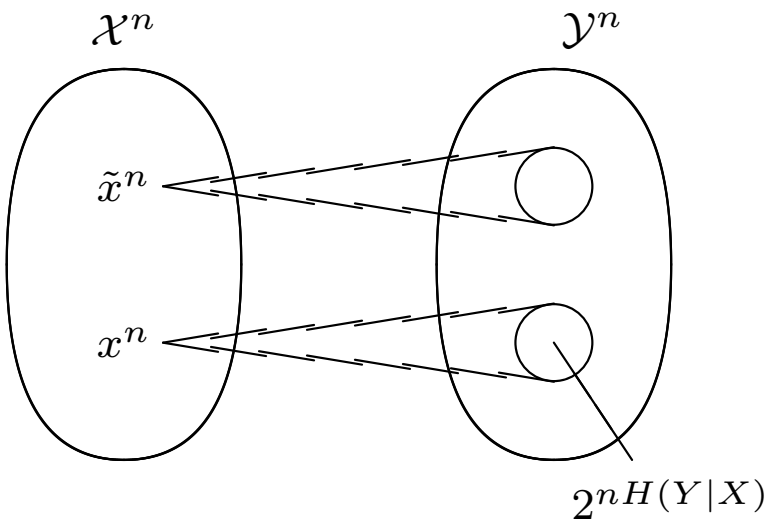
- Mensaje $W \in J,\ J = \{1, 2, \ldots, M\}$
- Palabra de código enviada: $X^n(W)$, donde $X^n : J \to \mathcal{X}^n$ es la *función de codificación* del código.
- **Canal DMC** $(\mathcal{X}^n, p(Y^n|X^n), \mathcal{Y}^n)$:

$$p\left(y_k|x^k, y^{k-1}\right) = p(y_k|x_k), k = 1, 2, \ldots, n\,.$$

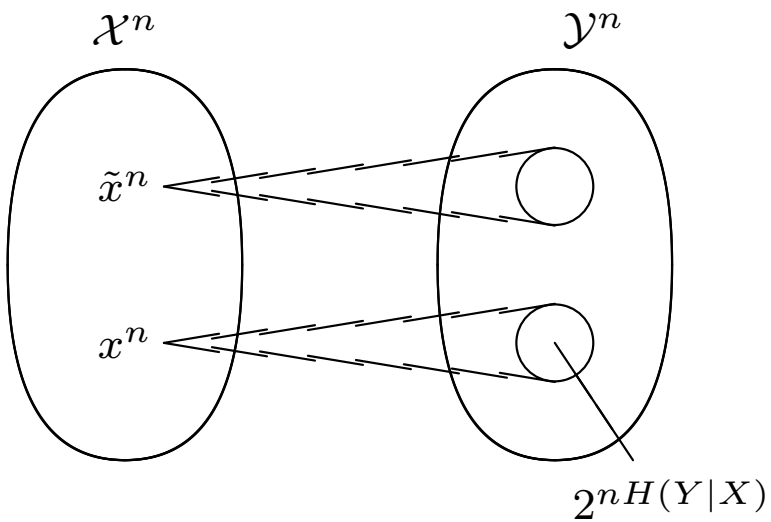
- Cadena recibida: $Y^n \sim p(y^n|x^n)$.
- Mensaje decodificado: $\hat{W} = g(Y^n)$, donde $g : \mathcal{Y}^n \to J$ es la *función de decodificación* del código.
- Error de transmisión: $W \neq \hat{W}$.

Teorema de codificación de canal: intuición

- Basado en AEP para pares (X^n, Y^n).
- Para n grande, cada x^n induce $2^{nH(Y|X)}$ salidas típicas.
- Para que no haya error, los conjuntos de salidas típicas para x^n y $\tilde{x}^n$ deberían ser disjuntos.
- Hay un total de $2^{nH(Y)}$ salidas típicas.
- A groso modo, esto limita la cantidad de palabras de código a un total de $2^{nH(Y)}/2^{nH(Y|X)} = 2^{nI(X;Y)}$.



Paréntesis: AEP conjunta



Repaso: AEP y conjuntos típicos

Sean $X_1 \dots X_n$ variables aleatorias i.i.d., $X_i \sim p(x)$, cada una de ellas con entropía $H = H(X_i)$ finita.

Equipartición asintótica (AEP)

- "Casi toda la probabilidad se concentra en eventos que son casi equiprobables"

$$P\Big\{(x_1 \dots x_n) : p(x_1 \dots x_n) = 2^{-n(H \pm \epsilon)}\Big\} \approx 1\,.$$

- Por la ley de los grandes números,

$$-\frac{1}{n} \log p(X_1 \dots X_n) \to H \quad \text{en probabilidad}$$

Repaso: conjuntos típicos

Definición (Conjunto típico)

El *conjunto típico* $A_\epsilon^{(n)}$ con respecto a $p\,(x)$ es el conjunto de secuencias que cumplen:

$$2^{-n(H+\epsilon)} \leq p(x_1 \dots x_n) \leq 2^{-n(H-\epsilon)}$$

Propiedades

- $(x_1 \dots x_n) \in A_\epsilon^{(n)} \Leftrightarrow H - \epsilon \leq -\frac{1}{n} \log p(x_1 \dots x_n) \leq H + \epsilon$
- $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para n suficientemente grande.
- $|A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$
- $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H-\epsilon)}$ para n suficientemente grande.

Tipicalidad conjunta

Definición

El conjunto $A_\epsilon^{(n)}$ de secuencias conjuntamente típicas son los pares (x^n, y^n) tales que

- x^n es típico según $p(x^n)$: $\left|-\frac{1}{n}\log p(x^n) - H(X)\right| \leq \epsilon$
- y^n es típico según $p(y^n)$: $\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| \leq \epsilon$
- (x^n, y^n) es típico según $p(x^n, y^n)$:

$$\left|-\frac{1}{n}\log p(x^n, y^n) - H(X,Y)\right| \leq \epsilon$$

donde $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$.

AEP conjunta

Teorema

1. $\Pr\left\{(X^n, Y^n) \in A_\epsilon^{(n)}\right\} \to 1$ cuando $n \to \infty$
2. $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+e)}$
3. Si $(\hat{X}^n, \hat{Y}^n) \sim p(x^n)\, p(y^n)$ (o sea que $\hat{X}^n$ e $\hat{Y}^n$ son indep. con marginales idénticas a $p(x^n, y^n)$) entonces

$$\Pr\left\{(\hat{X}^n, \hat{Y}^n) \in A_\epsilon^{(n)}\right\} \leq 2^{-n(I(X;Y)-3\epsilon)}$$

y para n suficientemente grande

$$\Pr\left\{(\hat{X}^n, \hat{Y}^n) \in A_\epsilon^{(n)}\right\} \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}$$

Nota: El punto 3 dice que la prob. de que dos secuencias típicas de $\mathcal{X}^n$ y $\mathcal{Y}^n$, pero no generadas por la transmisión, sean también conjuntamente típicas, es chica, y tiende a 0 con n grande.

AEP conjunta: prueba de 3

- Si $\hat{X}^n$ e $\hat{Y}^n$ son independientes con marginales idénticas a $p\left(x^n, y^n\right)$ entonces, desarrollando por cotas superiores,

$$\begin{aligned}
\Pr\left\{(\hat{X}^n, \hat{Y}^n) \in A_\epsilon^{(n)}\right\} &= \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p\left(x^n\right) p\left(y^n\right) \\
&\leq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \\
&\leq 2^{n(H(X,Y)+\epsilon)} 2^{-n(H(X)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \\
&= 2^{n(H(X,Y)+\epsilon)-n(H(X)-\epsilon)-n(H(Y)-\epsilon)} \\
&= 2^{-n(I(X;Y)-3\epsilon)}
\end{aligned}$$

AEP conjunta: prueba de 3

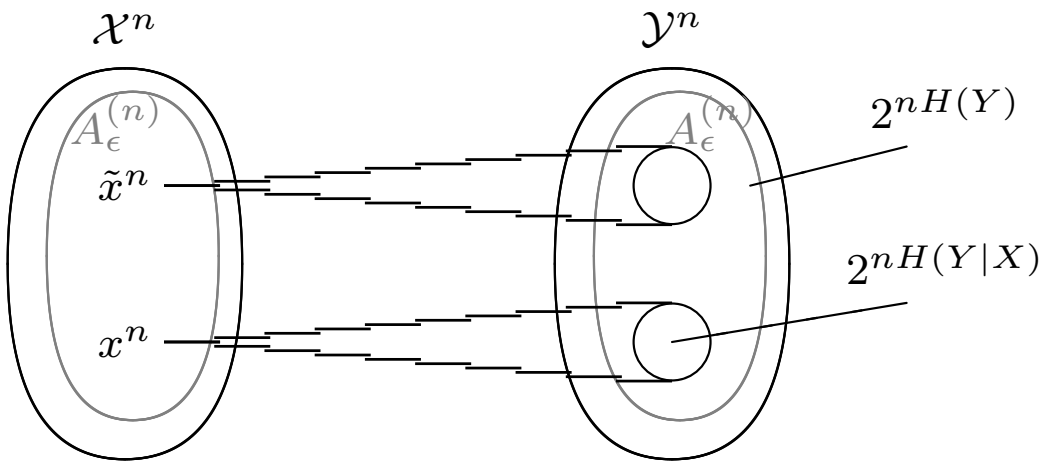
- Haciendo lo mismo pero con las cotas inferiores

$$
\begin{aligned}
\Pr\left\{(\hat{X}^n, \hat{Y}^n) \in A_\epsilon^{(n)}\right\} &= \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p\left(x^n\right) p\left(y^n\right) \\
&\geq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X)+\epsilon)} 2^{-n(H(Y)+\epsilon)} \\
&\geq (1-\epsilon) 2^{n(H(X,Y)-\epsilon)} 2^{-n(H(X)+\epsilon)} 2^{-n(H(Y)+\epsilon)} \\
&= (1-\epsilon) 2^{n(H(X,Y)-\epsilon)-n(H(X)+\epsilon)-n(H(Y)+\epsilon)} \\
&= (1-\epsilon) 2^{-n(I(X;Y)+3\epsilon)}
\end{aligned}
$$

□

Prueba del teorema

Intuición



- Cada palabra de código $X^n(i)$ genera $\approx 2^{nH(Y|X)}$ palabras conj. tip. en $\mathcal{Y}^n$
- Hay $\approx 2^{nH(Y)}$ secuencias típicas en $\mathcal{Y}^n$
- Los subconjuntos inducidos en $\mathcal{Y}^n$ no se deben solapar
- El máximo posibls de conj. disjuntos (o sea, mensajes distintos) es

$$\approx 2^{nH(Y)}/2^{nH(Y|X)} = 2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)}$$

Enunciado

Teorema (Teorema de codificación de canal)

- Todas las tasas R bajo la capacidad del canal C son alcanzables.
- Específicamente, para cada $R < C$ existe una secuencia de códigos $(2^{nR}, n)$ con probabilidad de error máxima $\lambda^{(n)} \to 0$.
- Recíprocamente, la existencia de una secuencia de códigos $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$ implica $R \leq C$.

Demostración del directo

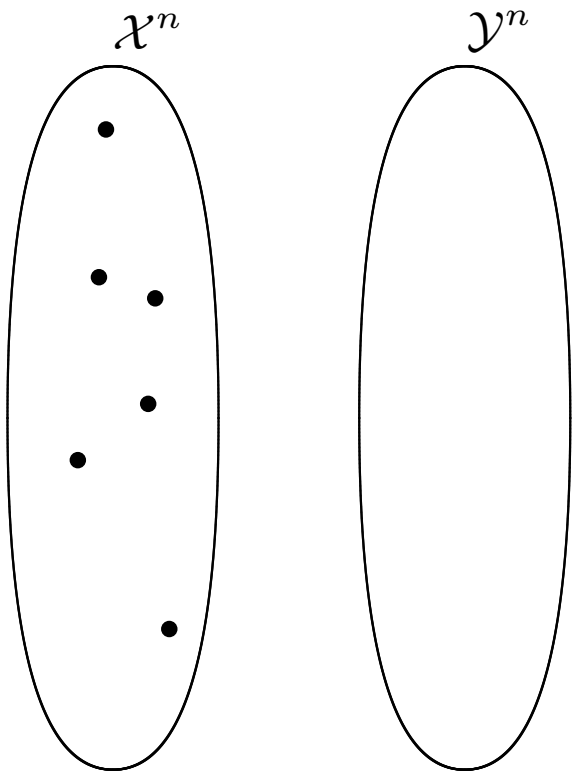
Hoja de ruta

- Vamos a analizar la probabilidad de error en un *experimento imaginario*.
- La aleatoriedad en este experimento involucra:
  - La elección de un código al azar. Las reglas de codificación y decodificación están predeterminadas, pero las palabras de código las vamos a sortear con una distribución a definir más adelante.
  - La elección al azar de un mensaje a transmitir. La distribución para esta elección aleatoria la vamos a definir a conveniencia para nuestro análisis, no está vinculada realmente a una fuente de información.
  - La aleatoriedad en la salida que genera el canal a partir de la palabra de código en la entrada.
- Vamos a llegar a la conclusión de que la probabilidad de error, teniendo en cuenta *todos* estos elementos aleatorios, tiende a cero.
- El código aleatorio es simplemente un artilugio para la demostración. No establece ningún código específico con el cual codificar, pero nos permite concluir que existe algún código "bueno".

Demostración: esquema de codificación aleatoria

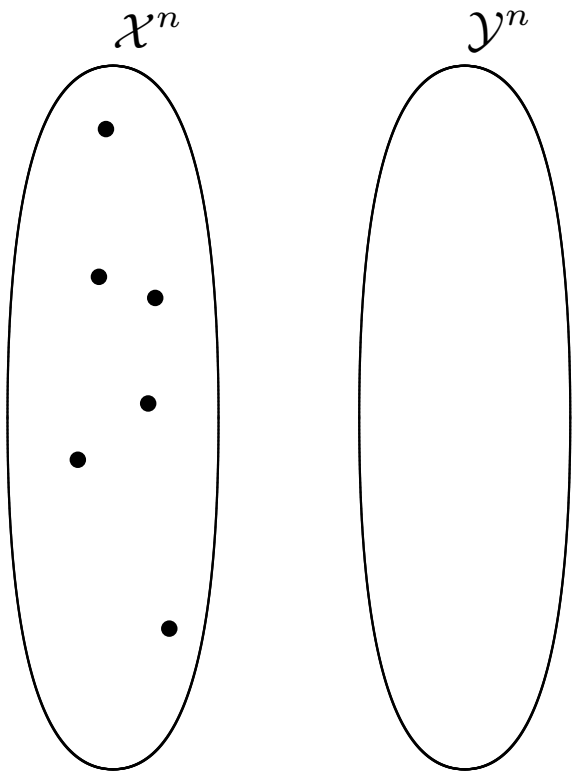
1. Codebook $\mathcal{C} = \{x^n(w)\}_{w=1,\ldots,2^{nR}}$ **aleatorio**.
   Cada palabra se sortea según $p(x)$.

$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$



Demostración: esquema de codificación aleatoria

1. Codebook $\mathcal{C} = \{x^n(w)\}_{w=1,\ldots,2^{nR}}$ **aleatorio**.
   Cada palabra se sortea según $p(x)$.
$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$
2. $\mathcal{C}$ es dada a transmisor y receptor.

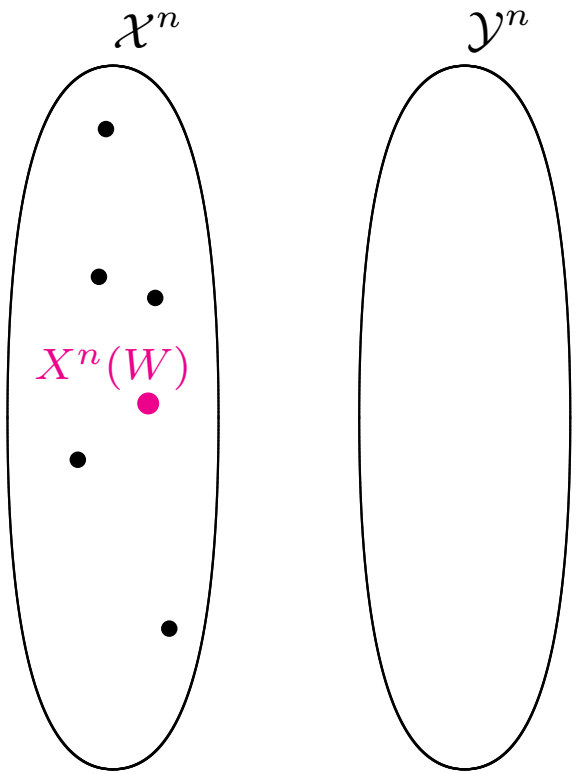


Demostración: esquema de codificación aleatoria

1. Codebook $\mathcal{C} = \{x^n(w)\}_{w=1,\ldots,2^{nR}}$ **aleatorio**. Cada palabra se sortea según $p(x)$.

$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$

2. $\mathcal{C}$ es dada a transmisor y receptor.
3. Se sortea un mensaje W de acuerdo a la dist. uniforme en J: $\Pr\{W = w\} = 2^{-nR}$.

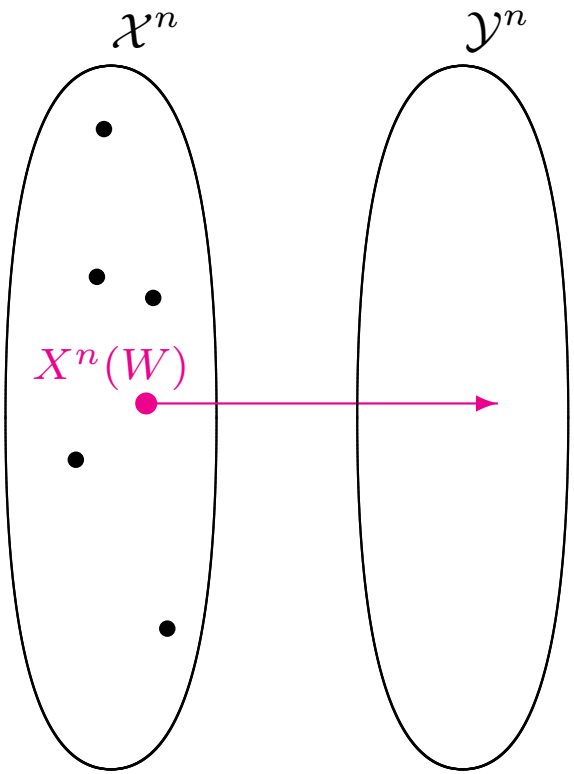


Demostración: esquema de codificación aleatoria

1. Codebook $\mathcal{C} = \{x^n(w)\}_{w=1,\ldots,2^{nR}}$ **aleatorio**. Cada palabra se sortea según $p(x)$.

$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$

2. $\mathcal{C}$ es dada a transmisor y receptor.
3. Se sortea un mensaje W de acuerdo a la dist. uniforme en J: $\Pr\{W = w\} = 2^{-nR}$.
4. La palabra W-ésima del código, $X^n(W)$, es enviada por el transmisor (v.a. porque fun. de W).



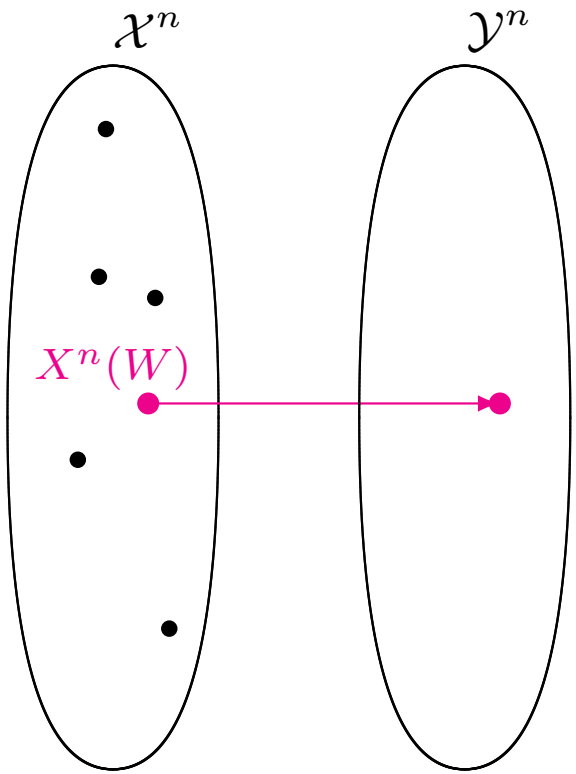
Demostración: esquema de codificación aleatoria

1. Codebook $\mathcal{C} = \{x^n(w)\}_{w=1,\ldots,2^{nR}}$ **aleatorio**. Cada palabra se sortea según $p(x)$.

$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$

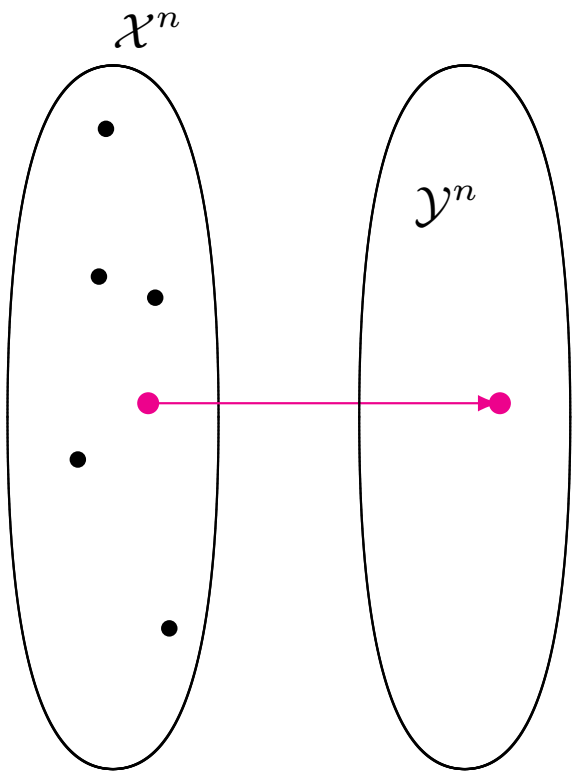
2. $\mathcal{C}$ es dada a transmisor y receptor.
3. Se sortea un mensaje W de acuerdo a la dist. uniforme en J: $\Pr\{W = w\} = 2^{-nR}$.
4. La palabra W-ésima del código, $X^n(W)$, es enviada por el transmisor (v.a. porque fun. de W).
5. El receptor recibe Y^n de acuerdo a:

$$\Pr\{y^n | x^n(W)\} = \prod_{i=1}^{n} p(y_i | x_i(W))$$



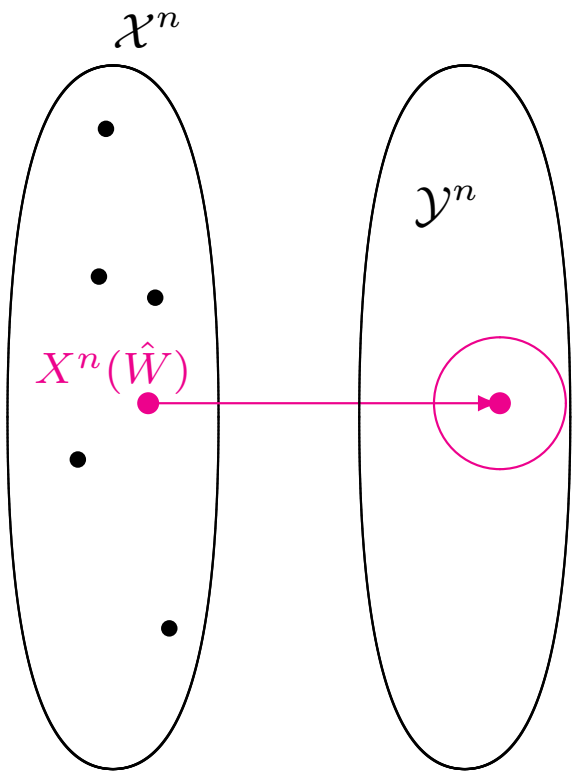
Demostración: decodificación por tipicalidad conjunta

6. El receptor decide que el indice enviado fue $\hat{W}$ si se cumple que



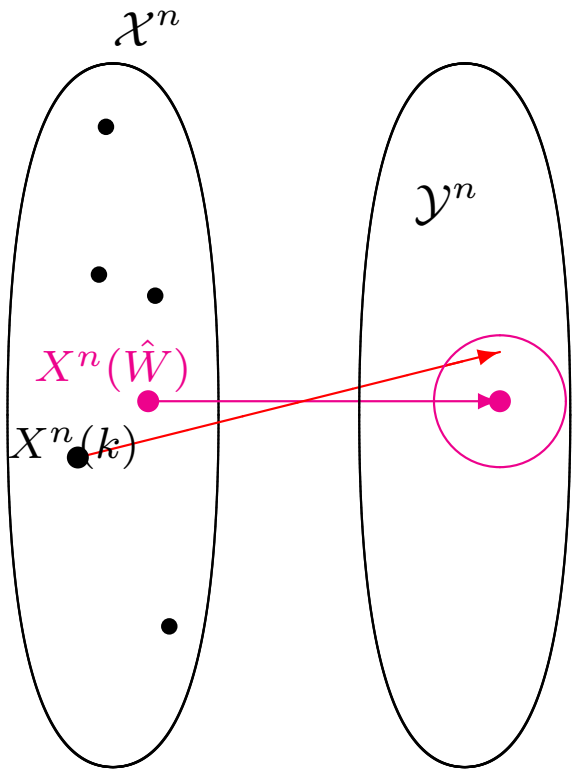
Demostración: decodificación por tipicalidad conjunta

6. El receptor decide que el indice enviado fue $\hat{W}$ si se cumple que
   - $(X^n(\hat{W}), Y^n)$ es conjuntamente típico, es decir, si pertenecen a $A_\epsilon^{(n)}$, con ϵ a definir.



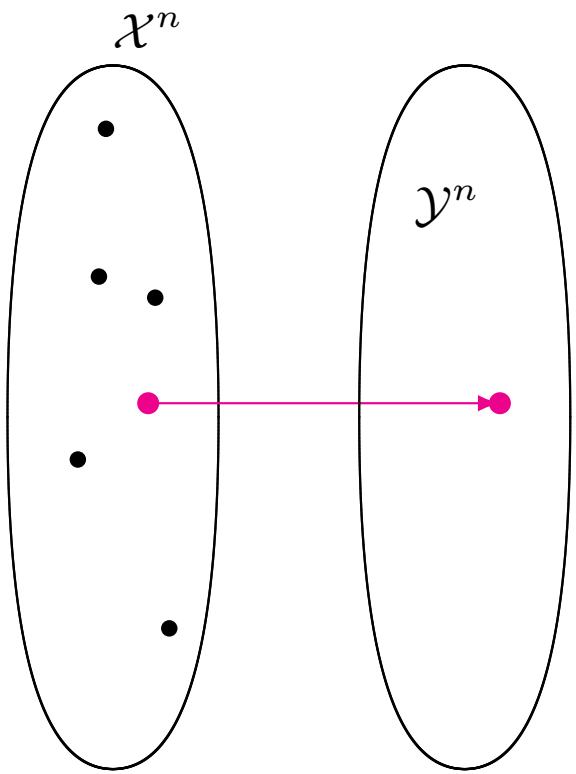
Demostración: decodificación por tipicalidad conjunta

6. El receptor decide que el indice enviado fue $\hat{W}$ si se cumple que
   - $(X^n(\hat{W}), Y^n)$ es conjuntamente típico, es decir, si pertenecen a $A_\epsilon^{(n)}$, con ϵ a definir.
   - No hay otro índice k tal que $(X^n(k), Y^n)$ es conjuntamente típico.



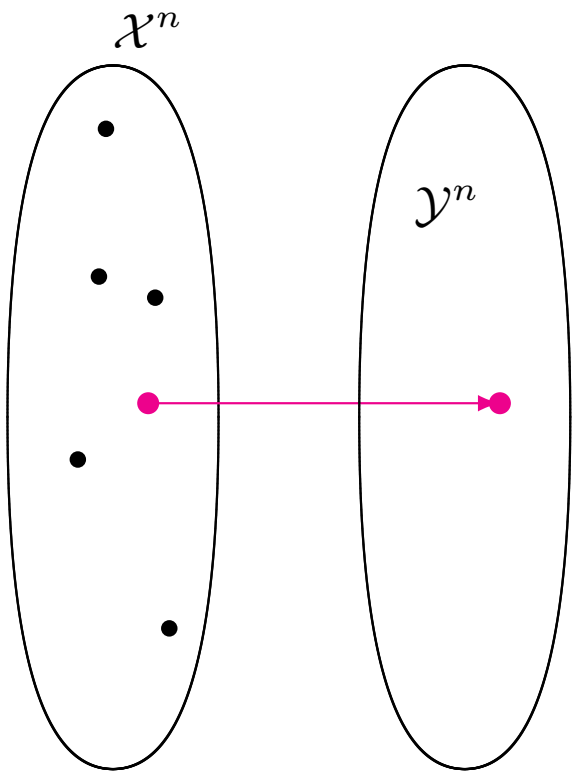
Demostración: decodificación por tipicalidad conjunta

6. El receptor decide que el indice enviado fue $\hat{W}$ si se cumple que
   - $(X^n(\hat{W}), Y^n)$ es conjuntamente típico, es decir, si pertenecen a $A_\epsilon^{(n)}$, con ϵ a definir.
   - No hay otro índice k tal que $(X^n(k), Y^n)$ es conjuntamente típico.
7. Si no existe un $\hat{W} \in J$ que cumpla esto, se declara $\hat{W} = 0$.



Demostración: decodificación por tipicalidad conjunta

6. El receptor decide que el indice enviado fue $\hat{W}$ si se cumple que
   - $(X^n(\hat{W}), Y^n)$ es conjuntamente típico, es decir, si pertenecen a $A_\epsilon^{(n)}$, con ϵ a definir.
   - No hay otro índice k tal que $(X^n(k), Y^n)$ es conjuntamente típico.
7. Si no existe un $\hat{W} \in J$ que cumpla esto, se declara $\hat{W} = 0$.
8. En todo caso, se da un error de decodificación cuando $W \neq \hat{W}$.



Demostración: análisis de la prob. de error

Consideramos un mensaje arbitrario, w, y los siguientes eventos:

$$\begin{aligned} E_u &= \left\{(X^n(u), Y^n) \in A_\epsilon^{(n)}\right\}, 1 \leq u \leq 2^{nR}, \\ \mathcal{E}_w &= \{\hat{W} \neq w\}. \end{aligned}$$

Observamos que

$$\mathcal{E}_w \subseteq E_w^c \cup \bigcup_{u \neq w} E_u,$$

y por lo tanto, para $\mathcal{E} = \{\hat{W} \neq W\}$, se cumple que

$$P(\mathcal{E}|W = w) \leq P(E_w^c|W = w) + \sum_{u \neq w} P(E_u|W = w).$$

Notar que, dado $W = w$, los eventos E_u dependen de la selección aleatoria del código, que determina $X^n(u)$ para $1 \leq u \leq 2^{nR}$, y de la acción aleatoria del canal, que determina Y^n cuando se transmite w.

Demostración: análisis de la prob. de error

$$\begin{aligned} P(\mathcal{E}|W=w) &\leq P(E_w^c|W{=}w) + \sum_{u\neq w} P(E_u|W{=}w) \\ &\leq \epsilon + \sum_{u\neq w} 2^{-n(I(X;Y)-3\epsilon)} \\ &\leq \epsilon + 2^{nR}2^{-n(I(X;Y)-3\epsilon)} \\ &= \epsilon + 2^{-n(I(X;Y)-3\epsilon-R)} \end{aligned}$$

Para n grande y $R < I(X;Y) - 3\epsilon$. obtenemos

$$P(\mathcal{E}|W=w) \leq 2\epsilon$$

Demostración: análisis de la prob. de error

- Para cada w, tenemos $P(\mathcal{E}|W=w) \leq 2\epsilon$, lo cual implica que

$$P(\mathcal{E}) = \sum_{w} \frac{1}{2^{nR}} P(\mathcal{E}|W=w) \leq 2\epsilon\,,$$

  donde recordamos que $\frac{1}{2^{nR}}$ es la probabilidad de elegir un mensaje w.

- Por otra parte tenemos

$$P(\mathcal{E}) = \sum_{\mathcal{C}} P(\mathcal{C}) P_e^{(n)}(\mathcal{C})\,, \tag{1}$$

  donde $P_e^{(n)}(\mathcal{C})$ es la probabilidad de error promediada sobre todas las palabras de código de $\mathcal{C}$,

$$P_e^{(n)}(\mathcal{C}) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C})\,,$$

  que, como W se sortea con distribución uniforme, es igual a la probabilidad de $\mathcal{E}$ dado el código $\mathcal{C}$.

- Como $P(\mathcal{E}) \leq 2\epsilon$, (1) implica que existe un $\mathcal{C}^*$ que cumple $P_e^{(n)}(\mathcal{C}^*) \leq 2\epsilon$.

Demostración: refinamiento del código

- Descártese la peor mitad de las palabras de $\mathcal{C}^*$ (las de mayor probabilidad de error).
- Dado que

$$2\epsilon \geq \frac{1}{2^{nR}} \sum \lambda_i(\mathcal{C}^*)$$

  la mitad restantes de las palabras de código tienen prob. de error $\lambda_i \leq 4\epsilon$ (sino su suma daría más que 2ϵ).
- Quedan 2^{nR-1} palabras, por lo que la tasa baja a $R - \frac{1}{n}$, que tiende a R con $n \to \infty$.

Final

Sea una tasa arbitraria $\bar{R}$, $0 < \bar{R} < C$, y sea $\bar{\epsilon} > 0$ arbitrario. En nuestra construcción

- Tomamos $p(x)$ t.q. $I(X;Y) = C$ para $X \sim p$.
- Definimos $R = (\bar{R} + C)/2$.
- Elegimos $\epsilon < \text{mín}\{(C - R)/3, \bar{\epsilon}/4\}$.

A partir de nuestro razonamiento anterior podemos concluir que

- Como se satisface que $R < I(X;Y) - 3\epsilon$, se cumple que para $\mathcal{C}^*$ la probabilidad de error maximal satisface $\lambda^{(n)} \leq 4\epsilon \leq \bar{\epsilon}$, para n suficientemente grande.
- La tasa de $\mathcal{C}^*$ es $R - \frac{1}{n}$ que, para n suficientemente grande, es mayor que $\bar{R}$.

□

Demostración del recíproco

Esquema de la prueba

- Empezaremos con el caso más simple $P_e^{(n)} = 0$.
- A partir de ese caso, y ayudados por la desigualdad de Fano, se demuestra el recíproco.
- Reveremos la desigualdad de Fano en el contexto del teorema.

Caso simple: probabilidad de error 0

Probabilidad de error 0

- Probaremos que $P_e^{(n)} = 0$ implica $R \leq C$.
- Se asume que $g(Y^n) = W$, o sea que $H(W|Y^n) = 0$.
- Definimos W uniforme sobre J, de modo que se cumple $H(W) = nR$.

Probabilidad de error 0: lema

Lema (Cota de info. mutua)

$$
\begin{aligned}
I(X^n;Y^n) &\overset{\text{def.}}{=} H(Y^n) - H(Y^n|X^n) \\
&\overset{\text{cadena}}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, X^n) \\
&\overset{\text{DMC}}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I(Y_i;X_i)
\end{aligned}
$$

Utilizar el canal varias veces no aumenta su capacidad.

Probabilidad de error 0

Demostración

Desarrollando

$$\begin{aligned} nR = H(W) &= H(W|Y^n) + I(W;Y^n) \\ &= I(W;Y^n) \\ &\overset{(a)}{\leq} I(X^n;Y^n) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} I(X_i;Y_i) \\ &\overset{(c)}{\leq} nC \end{aligned}$$

□

(a) desigualdad de procesamiento de datos. (b) Lema demostrado en diapositiva anterior. (c) definición de capacidad.

Desigualdad de Fano para codificación de canal

Lema (Desigualdad de Fano)

Para un canal discreto sin memoria y un mensaje W distribuido uniformemente sobre un alfabeto de tamaño 2^{nR} se cumple

$$H\left(W|\hat{W}\right) \leq 1 + P_e^{(n)} nR\,.$$

Demostración Sea $J = \left\{1, 2, \ldots, 2^{nR}\right\}$ el alfabeto de los mensajes W. La forma general de la desigualdad de Fano establece que

$$H(\hat{W}|W) \leq 1 + Pr\{\hat{W} \neq W\} \log |J|$$

Sustituyendo: $\log |J| = \log 2^{nR} = nR$, $Pr\{\hat{W} \neq W\} = P_e^{(n)}$ (por def.) se llega al lema.

Recíproco: demostración

Teorema (Recíproco del teorema de codificación de canal.)

Toda secuencia de códigos $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$ debe cumplir $R \leq C$.

Demostración

$$
\begin{aligned}
nR = H(W) &= H(W|Y^n) + I(W;Y^n) \\
&\overset{W \to X^n \to Y^n}{\leq} H(W|Y^n) + I(X^n(W);Y^n) \\
&\overset{\text{Fano}}{\leq} 1 + P_e^{(n)} nR + I(X^n(W);Y^n) \\
&\leq 1 + P_e^{(n)} nR + nC \\
R &\leq 1/n + P_e^{(n)} R + C
\end{aligned}
$$

Cuando $n \to \infty$ los dos primeros términos a la derecha tienden a 0, probando que $R \leq C$.

□

Recíproco: conclusiones

- Error para tasas arriba de C:
  Reescribiendo la penúltima ecuación de la demostración,

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$$

  se muestra que $P_e^{(n)}$ se aleja de 0 para n suficientemente grande.

Códigos prácticos de canal

Códigos de canal

- Shannon promete códigos buenos, pero no nos dice cómo hallarlos.
- Desde la aparición del trabajo de Shannon se están buscando códigos buenos.
- Además de baja probabilidad de error, los códigos deben admitir implementaciones eficientes.
- En esta sección veremos algunos esquemas de codificación de canal muy sencillos

Curso específico ofrecido por el NTI

El NTI dicta un curso específico para este tema "Códigos para Corrección de Errores".

Códigos lineales

Notación

- Denotamos F_q al cuerpo finito de q elementos ($q = p^m$ con p primo).
- Ejemplo: F_2 es el conjunto $\{0, 1\}$ con la suma y el producto módulo 2 (XOR y AND).

Definición

- Decimos que un código (M, n), $\mathcal{C}$, es lineal, si es un subespacio de F^n sobre F, con $|\mathcal{C}| = M$.
- Si k es la dimensión de $\mathcal{C}$ sobre F, entonces es $M = q^k$.

Códigos lineales

Definición

- Decimos que una matriz G es generadora de $\mathcal{C}$ sobre F cuando sus filas forman una base de $\mathcal{C}$ (la matriz generadora no es única).
- Para codificar una palabra u entonces, simplemente se multiplica por G:

$$u \longmapsto uG\,.$$

- La matriz G tiene dimensiones $k \times n$ (si las filas son linealmente independientes).

Ejemplo (Paridad)

El código de paridad $(4, 3)$ es el generado por la siguiente matriz

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Definición

- Decimos que una matriz H es de chequeo de paridad de un código $\mathcal{C}$ sobre F cuando:

$$c \in \mathcal{C} \iff Hc^T = 0\,.$$

- El código $\mathcal{C}$ es entonces el núcleo de H.

Ejemplo (Paridad)

La matriz de chequeo de paridad para el código de paridad $(4, 3)$ es:

$$H = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$$

Ejemplo (Código de Hamming)

El código de Hamming $(2^4, 7)$ está definido por la siguiente matriz de chequeo de paridad:

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

La dimensión de $\mathcal{C}$ es $k = 4$; hay $M = 2^4$ palabras:

0000000	0100101	1000011	1100110
0001111	0101010	1001100	1101001
0010110	0110011	1010101	1110000
0011001	0111100	1011010	1111111

Ejemplo (Código de Hamming)

- La dimensión de $\mathcal{C}$ es $k = 4$
- La cantidad de mensajes es $M = 2^4 = 16$
- La tasa del código es $R = \frac{\log M}{n} = \frac{k}{n} = \frac{4}{7}$
- La distancia mínima del código es $d = 3$

La siguiente matriz genera $\mathcal{C}$:

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Ejemplo (Como decodificamos?)

Supongamos que se envió una palabra c, pero se recibió una palabra r que difiere en un bit con c.
Esto es: $r = c + e_i$, donde e_i tiene un 1 en la posición i y ceros en el resto.

$$Hr = H(c + e_i) = Hc + He_i = He_i$$

que es la $i-$ésima columna de H. Por lo tanto conocemos i, y podemos recuperar la palabra enviada c.

¿Qué pasa si ocurren 2 errores?

Más códigos

Otros códigos lineales

Con ideas más sofisticadas se logran códigos mucho mejores:

- Reed-Solomon
- BCH (Bose, Ray-Chaudhuri, Hocquenghem)
- LDPC (Low Density Parity-Check Codes)
- Turbo-Codes

Curso: Códigos para Corrección de Errores

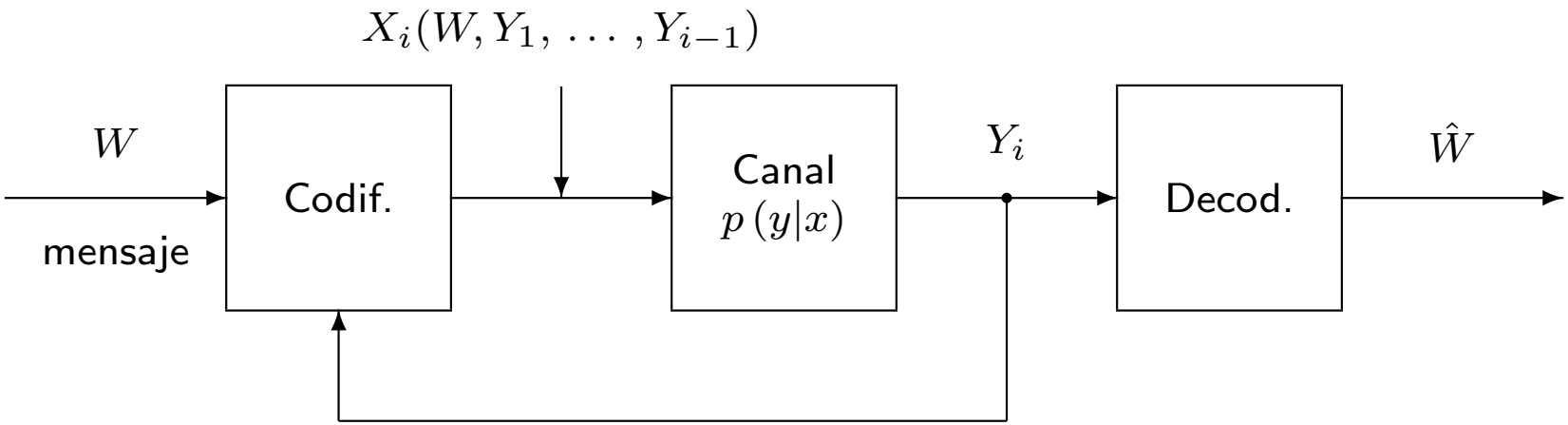
Estos temas se ven en un curso específico dictado por el NTI.

Aplicaciones

Algunas aplicaciones

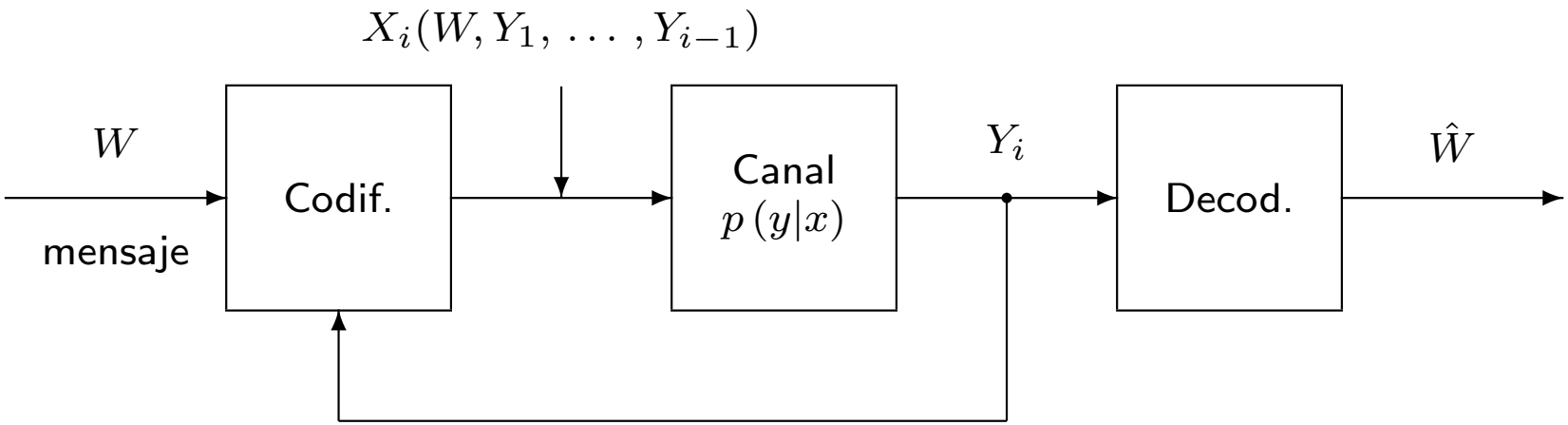
- LDPC
  - IEEE 802.16
  - DVB-S2 (Digital Video Broadcasting - Satellite - Second Generation)
- Turbo-Codes
  - 3G
  - IEEE 802.16
  - NASA (Mars Reconnaissance Orbiter)
- Reed-Solomon
  - CD
  - DVD
  - DSL
  - NASA (Mars Pathfinder)

Canales con realimentación



- Cada símbolo recibido es devuelto inmediatamente y sin ruido al transmisor.
- Puede este esquema mejorar la capacidad?

Canales con realimentación



- Cada símbolo recibido es devuelto inmediatamente y sin ruido al transmisor.
- Puede este esquema mejorar la capacidad?
- La respuesta es no!

Códigos de canal con realimentación

Definición (Código con realimentación)

Un código con realimentación (M, n) está definido por

- un conjunto de índice $J = \{1, \ldots, M\}$,
- una función de codificación determinada por n mapeos, $x_i(W, Y^{i-1})$, $1 \leq i \leq n$, $W \in J$,
- una función de decodificación $g : \mathcal{Y}^n \to J$.

La tasa del código es $R = \frac{\log M}{n}$.

Definición (Tasa alcanzable)

Una tasa R se dice alcanzable para un canal con realimentación si existe una secuencia de códigos con realimentación $\left(\lceil 2^{nR} \rceil, n\right)$ tal que $\lambda^{(n)} \to 0$ cuando $n \to \infty$.

Definición (Capacidad de un canal con realimentación)

La capacidad de un canal con realimentación, C_{FB}, es el supremo de las tasas R alcanzables a por códigos con realimentación.

Capacidad del canal con realimentación

Teorema (Capacidad del canal con realimentación)

$$C_{FB} = C = \max_{p(X)} I(X;Y)$$

Demostración (1):

- Como un código sin realimentación es un caso particular de código con realimentación, toda tasa alcanzable sin realimentación es también alcanzable con realimentación. Por lo tanto tenemos

$$C_{FB} \geq C\,.$$

- Probar que $C_{FB} \leq C$ es parecido a la demostración del recíproco del segundo teorema de Shannon, pero no vale $W \to X^n \to Y^n$. Ejemplo:
  - $J = \{0, 1\}$, $n = 2$,
  - para $W = 0$, el codificador define $x_1 = 0, x_2 = Y_1$,
  - para $W = 1$, el codificador define $x_1 = 0, x_2 = \bar{Y}_1$.

  Si además de X_1, X_2 conocemos W, Y_1 queda determinado, pero no necesariamente si solo conocemos X_1, X_2.

Capacidad del canal con realimentación (2)

$$\begin{aligned} nR = H(W) &= H(W|Y^n) + I(W;Y^n) \\ &\overset{Fano}{\leq} 1 + P_e^{(n)} nR + I(W;Y^n) \end{aligned}$$

Ahora hay que acotar $I(W;Y^n)$:

$$\begin{aligned} I(W;Y^n) &= H(Y^n) - H(Y^n|W) \\ &\overset{cadena}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},W) \end{aligned}$$

Capacidad del canal con realimentación (3)

$$
\begin{aligned}
&\overset{cadena}{=} && H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \ldots, Y_{i-1}, W) \\
&\overset{(a)}{=} && H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \ldots, Y_{i-1}, W, X_i) \\
&\overset{(b)}{=} && H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i)
\end{aligned}
$$

(a) X_i es función de W, Y^{i-1}.
(b) Y_i es condicionalmente independiente de Y^{i-1}, W dado X_i.

Capacidad del canal con realimentación (4)

$$\begin{aligned} I(W;Y^n) = & \quad H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ \leq & \quad \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ = & \quad \sum_{i=1}^{n} I(Y_i|X_i) \\ \leq & \quad nC \end{aligned}$$

Combinando esta cota con la desigualdad de Fano obtenemos:

$$nR \leq 1 + P_e^{(n)} nR + nC$$

Dividiendo por n y con $n \to \infty$ se obtiene $R \leq C$, y luego $C_{FB} \leq C$. Combinando esto y $C_{FB} \geq C$,

$$C = C_{FB}$$

□

Codificación conjunta fuente-canal: uniendo teoremas

- Vimos que $R > H$ (primer teorema de Shannon)
- Tambien vimos que $R < C$ (segundo teorema de Shannon)
- ¿Será cierto que $H < C$ es condición necesaria y suficiente para transmitir los datos de una fuente por un canal con capacidad C?

¿Separar o no separar?

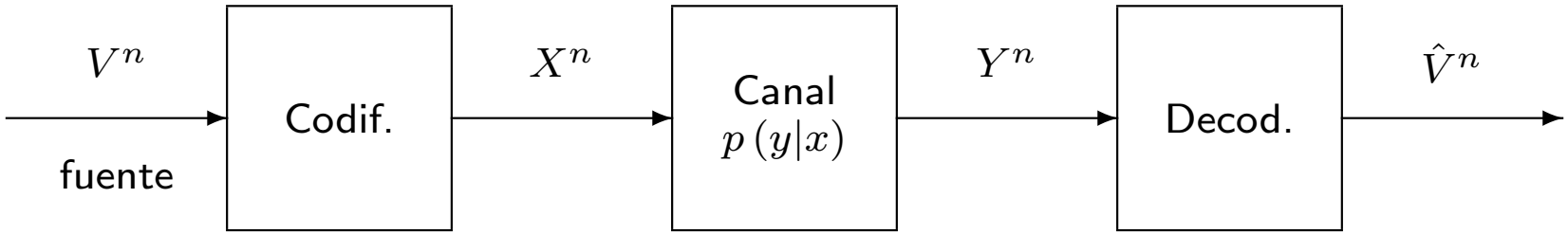
- Supongamos que queremos transmitir audio por un canal.
- Podríamos diseñar un código que mapee directamente las muestras de audio a la entrada del canal.
- También podríamos comprimir el audio al máximo, y luego crear un código adecuado al canal.
- No es obvio que ambos esquemas sean equivalentes

¿Separar o no separar? Lo mismo da!

En esta sección probaremos que los esquemas son equivalentes.

- Esto simplifica muchísimo el diseño de sistemas de transmisión!
- El diseño en capas de redes como Internet es un ejemplo.
- Puede no cumplirse en algunos casos de sistemas multitransmisor/multireceptor.
- En algunos casos la redundancia de la fuente está naturalmente adaptada al canal
  - el texto escrito puede recuperarse incluso habiéndose perdido hasta la mitad de las letras.
  - el oído tiene una capacidad inusitada para recuperar un mensaje hablado frente a un SNR muy bajo.

Codificación conjunta: formalización del problema



- Fuente $\mathcal{V}$ que cumple AEP (por ejemplo cualquier fuente estacionaria y ergódica)
- Secuencia a enviar $V^n = V_1, \ldots, V_n$, $V_i \in \mathcal{V}$
- Codifica como $X^n(V^n)$
- Decodificada como $\hat{V}^n = g(Y^n)$
- Error si $\hat{V}^n \neq V^n$. Prob de error:

$$\Pr\left\{\hat{V}^n \neq V^n\right\} = \sum_{y^n} \sum_{v^n} p\left(v^n\right) p\left(y^n | x^n(v^n)\right) \mathbf{1}\left(g(y^n) \neq v^n\right)$$

Teorema (Codificación conjunta fuente-canal)

$V_1, V_2, \ldots, V_n$ generado por fuente que cumple la AEP con tasa de entropía $H(\mathcal{V})$.

- Si $H(\mathcal{V}) < C$, existe un código que mapea fuente a canal con $\Pr\left\{\hat{V}^n \neq V^n\right\} \to 0$.
- Si $H(\mathcal{V}) > C$ y la fuente es estacionaria, con cualquier codificación que se elija ocurre que $\Pr\left\{\hat{V}^n \neq V^n\right\}$ se aparta de 0, por lo cual no es posible enviar los datos por el canal con probabilidad de error arbitrariamente pequeña.

Codificación conjunta: directo

- Sea $A_\epsilon^{(n)}$ el conjunto típico para la fuente, para un natural n y $0 < \epsilon < C - H(\mathcal{V})$ arbitrarios.
- Transmitimos solo secuencias típicas, identificándolas a través de un índice en una enumeración de $A_\epsilon^{(n)}$.
- Las no típicas producen error. Esto contribuye a lo sumo ϵ a la probabilidad de error.
- La cantidad de mensajes distintos a transmitir es $|A_\epsilon^{(n)}| \leq 2^{n(H(\mathcal{V})+\epsilon)}$.
- Como $\frac{\log |A_\epsilon^{(n)}|}{n} < C$, para n suficientemente grande existe un código que nos permite trasmitir el índice de una secuencia típica con probabilidad de error menor que ϵ haciendo n usos del canal.
- Tenemos entonces

$$\begin{aligned} \Pr\left\{\hat{V}^n \neq V^n\right\} &\leq \Pr\left\{V^n \in (A_\epsilon^{(n)})^c\right\} + \Pr\left\{g(Y^n) \neq V^n \middle| V^n \in A_\epsilon^{(n)}\right\} \\ &\leq \epsilon + \epsilon = 2\epsilon \end{aligned}$$

□

Codificación conjunta: recíproco

Consideremos ahora una secuencia de códigos conjuntos tal que $\Pr\left\{\hat{V}^n \neq V^n\right\} \to 0$:

$$X^n(V^n) : \mathcal{V}^n \to \mathcal{X}^n$$
$$g(Y^n) : \mathcal{Y}^n \to \mathcal{V}^n$$

Tenemos

$$\begin{aligned} H(\mathcal{V}) &\overset{(a)}{\leq} \frac{H(V_1, V_2, \ldots, V_n)}{n} \\ &= \frac{H(V^n)}{n} \\ &= \frac{1}{n} H(V^n|\hat{V}^n) + \frac{1}{n} I(V^n; \hat{V}^n) \end{aligned}$$

donde (a) surge de la monotonía de $\frac{H(V^n)}{n}$ para procesos estacionarios (recordar ejercicio de práctico).

Codificación conjunta: recíproco (2)

Por la desigualdad de Fano tenemos

$$H(V^n|\hat{V}^n) \leq 1 + \Pr\left\{\hat{V}^n \neq V^n\right\} \log|\mathcal{V}^n| = 1 + \Pr\left\{\hat{V}^n \neq V^n\right\} n \log|\mathcal{V}|\,,$$

de donde, sustituyendo en

$$H(\mathcal{V}) \leq \frac{1}{n} H(V^n|\hat{V}^n) + \frac{1}{n} I(V^n;\hat{V}^n)\,,$$

obtenemos

$$\begin{aligned} H(\mathcal{V}) &\leq \frac{1}{n}\left(1 + \Pr\left\{\hat{V}^n \neq V^n\right\} n \log|\mathcal{V}|\right) + \frac{1}{n} I(V^n;\hat{V}^n) \\ &\overset{V^n \to X^n \to Y^n \to \hat{V}^n}{\leq} \frac{1}{n}\left(1 + \Pr\left\{\hat{V}^n \neq V^n\right\} n \log|\mathcal{V}|\right) + \frac{1}{n} I(X^n;Y^n) \\ &\overset{DMC}{\leq} \frac{1}{n} + \Pr\left\{\hat{V}^n \neq V^n\right\} \log|\mathcal{V}| + C \end{aligned}$$

Codificación conjunta: recíproco (3)

Tenemos

$$H(\mathcal{V}) \leq \frac{1}{n} + \Pr\left\{\hat{V}^n \neq V^n\right\} \log|\mathcal{V}| + C\,.$$

Cuando $n \to \infty$, el hecho de que $\Pr\left\{\hat{V}^n \neq V^n\right\} \to 0$ implica que

$$H(\mathcal{V}) \leq C\,.$$

Más aún, si se cumpliera $H(\mathcal{V}) > C$, despejando $\Pr\left\{\hat{V}^n \neq V^n\right\}$ vemos que la probabilidad de error se mantiene alejada de 0 a medida que $n \to \infty$.

□

Codificación conjunta: conclusiones

- Con este resultado se unen los dos teoremas de Shannon.
- Esto muestra que no se pierde nada al separar el proceso de compresión y la codificación de canal
- En ambos casos estamos sujetos a la condición de que la tasa de entropía esté por debajo de la capacidad del canal.