Introducción a la Teoría de la Información

Milan S. Derpich
Departamento de Electrónica
Universidad Técnica Federico Santa María
SETIA-2010

Información

Información

¿Es posible hallar una noción intuitiva y a la vez cuantificable de la idea de información?

Información

¿Es posible hallar una noción intuitiva y a la vez cuantificable de la idea de información?

Información

¿Es posible hallar una noción intuitiva y a la vez cuantificable de la idea de información?

Información

Información

Qué tan informativo es un mensaje depende del contexto:

Información

Qué tan informativo es un mensaje depende del contexto:
Más improbable $\Rightarrow$ Más Información

Información

Qué tan informativo es un mensaje depende del contexto:
Más improbable $\Rightarrow$ Más Información

Luego, la información $\mathcal{I}(E)$ asociada a un evento E con probabilidad $p(E)$ es:

$$\mathcal{I}(E) = f(p(E))$$

para alguna función $f(\cdot)$ decreciente.

INFORMACIÓN

$$\mathcal{I}(E) = f(p(E))$$

Información

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Información

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Sean E_1 y E_2 dos eventos distintos.

Información

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Sean E_1 y E_2 dos eventos distintos.
Si E_1 y E_2 son independientes, resulta natural considerar que

$$\mathcal{I}(E_1, E_2) = \mathcal{I}(E_1) + \mathcal{I}(E_2) = f(p(E_1)) + f(p(E_2))$$

Información

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Sean E_1 y E_2 dos eventos distintos.
Si E_1 y E_2 son independientes, resulta natural considerar que

$$\mathcal{I}(E_1, E_2) = \mathcal{I}(E_1) + \mathcal{I}(E_2) = f(p(E_1)) + f(p(E_2))$$

Por otro lado,

$$\mathcal{I}(E_1, E_2) = f(p(E_1, E_2)) = f(p(E_1)p(E_2))$$

Información

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Sean E_1 y E_2 dos eventos distintos.
Si E_1 y E_2 son independientes, resulta natural considerar que

$$\mathcal{I}(E_1, E_2) = \mathcal{I}(E_1) + \mathcal{I}(E_2) = f(p(E_1)) + f(p(E_2))$$

Por otro lado,

$$\mathcal{I}(E_1, E_2) = f(p(E_1, E_2)) = f(p(E_1)p(E_2))$$

Por lo tanto, $f(\cdot)$ es tal que

$$f(xy) = f(x) + f(y)$$

INFORMACIÓN

$$\mathcal{I}(E) = f(p(E))$$

¿Qué función $f(\cdot)$ es la más adecuada?

Sean E_1 y E_2 dos eventos distintos.
Si E_1 y E_2 son independientes, resulta natural considerar que

$$\mathcal{I}(E_1, E_2) = \mathcal{I}(E_1) + \mathcal{I}(E_2) = f(p(E_1)) + f(p(E_2))$$

Por otro lado,

$$\mathcal{I}(E_1, E_2) = f(p(E_1, E_2)) = f(p(E_1)p(E_2))$$

Por lo tanto, $f(\cdot)$ es tal que

$$f(xy) = f(x) + f(y)$$

Entonces

$$\mathcal{I}(E) = f(p(E)) = -\log(p(E))$$

Información

Claude E. Shannon llamó a

$$\mathcal{I}(E) = -\log(p(E))$$

la "auto-información" (*self-information*) asociada a un eveno E.

INFORMACIÓN

Claude E. Shannon llamó a

$$\mathcal{I}(E) = -\log(p(E))$$

la "auto-información" (*self-information*) asociada a un eveno E.

Con esta noción como como punto de partida, Shannon estableció en 1948 las bases de lo que se conoce hoy como Teoría de la Información (*Information Theory*).

Sumario

Fuentes de Información - Entropía

Fuentes de Información - Entropía

Todo sistema que origine variables aleatorias es una fuente de información.

Fuentes de Información - Entropía

Todo sistema que origine variables aleatorias es una fuente de información.
Sea x una variable aleatoria discreta.
¿Cuánta información (incertidumbre) "emana" de x?

FUENTES DE INFORMACIÓN - ENTROPÍA

Todo sistema que origine variables aleatorias es una fuente de información.
Sea x una variable aleatoria discreta.
¿Cuánta información (incertidumbre) “emana” de x?

Entropía de una variable aleatoria:

$$H(\mathrm{x}) \triangleq \mathrm{E}[\mathcal{I}(\mathrm{x})] = -\sum_i p(\mathrm{x} = x_i) \log(p(\mathrm{x} = x_i))$$

Fuentes de Información - Entropía

Todo sistema que origine variables aleatorias es una fuente de información.
Sea x una variable aleatoria discreta.
¿Cuánta información (incertidumbre) "emana" de x?

Entropía de una variable aleatoria:

$$H(\mathrm{x}) \triangleq \mathrm{E}[\mathcal{I}(\mathrm{x})] = -\sum_i p(\mathrm{x} = x_i) \log(p(\mathrm{x} = x_i))$$

Si usamos logaritmo base 2, las unidades de $H(\mathrm{x})$ son bits

Entropía - Ejemplos

Ejemplo

x distribuye uniformemente sobre N valores discretos:

ENTROPÍA - EJEMPLOS

EJEMPLO

x distribuye uniformemente sobre N valores discretos:

$$H(\mathrm{x}) = -\sum_{i=1}^{N} \frac{1}{N} \log_2(1/N) = \log_2(N) \text{ [bit]}$$

Entropía - Ejemplos

Ejemplo

x distribuye uniformemente sobre N valores discretos:

$$H(\mathrm{x}) = -\sum_{i=1}^{N} \frac{1}{N} \log_2(1/N) = \log_2(N) \text{ [bit]}$$

Ejemplo

Sea $\mathrm{x} \sim$ Bernoulli(α): $p(\mathrm{x}=0) = \alpha$, $p(\mathrm{x}=1) = 1-\alpha$.

$$H(\mathrm{x}) = -[\alpha \log_2 \alpha + (1-\alpha)\log_2(1-\alpha)] \quad \text{[bit]}$$

Entropía - Ejemplos

Ejemplo

x distribuye uniformemente sobre N valores discretos:

$$H(\mathrm{x}) = -\sum_{i=1}^{N} \frac{1}{N} \log_2(1/N) = \log_2(N) \text{ [bit]}$$

Ejemplo

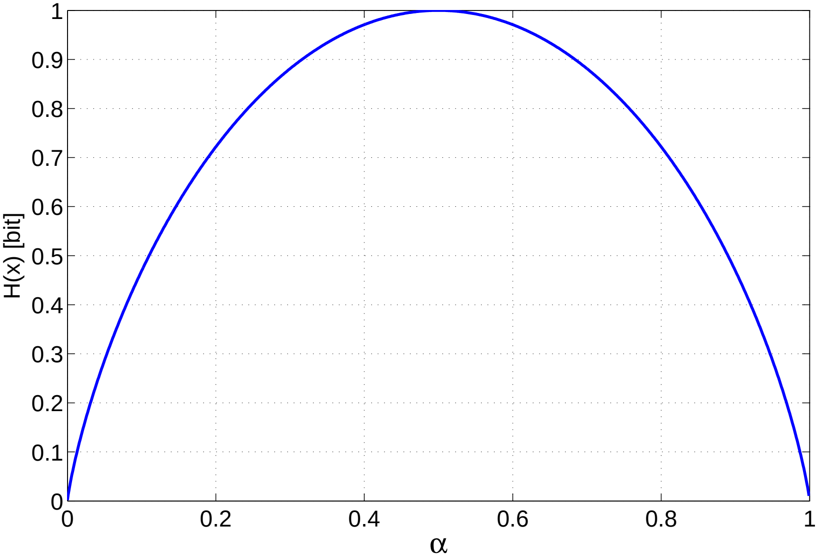
Sea $\mathrm{x} \sim$ Bernoulli(α): $p(\mathrm{x}=0)=\alpha$, $p(\mathrm{x}=1)=1-\alpha$.

$$H(\mathrm{x}) = -[\alpha \log_2 \alpha + (1-\alpha)\log_2(1-\alpha)] \quad \text{[bit]}$$

$$\alpha = 1/4 \Longrightarrow H(\mathrm{x}) = 0{,}8113 \text{ [bit]}$$

$$\alpha = 1/2 \Longrightarrow H(\mathrm{x}) = 1 \text{ [bit]}$$

Entropía de x $\sim Bernoulli(\alpha)$



Entropía

Entropía

Propiedades:

- $H(x) \geq 0$.

Entropía

Propiedades:

- $H(\mathrm{x}) \geq 0$.
- $H(\mathrm{x}) = 0$ sí y sólo si x no es aleatorio.

Entropía

Propiedades:

- $H(\mathrm{x}) \geq 0$.
- $H(\mathrm{x}) = 0$ sí y sólo si x no es aleatorio.
- Para todo mapa $\mathcal{M}$ determinístico, $H(\mathcal{M}\,\mathrm{x}) \leq H(\mathrm{x})$. Hay igualdad sí y sólo si $\mathcal{M}$ es invertible.

Entropía Condicional

Entropía Condicional

Sean x, y dos v.a. discretas. La entropía de x dado y se define como

$$H(\mathrm{x} \,|\, \mathrm{y}) = \sum_j p(\mathrm{y} = y_j) H(\mathrm{x} \,|\, \mathrm{y} = y_j)$$

Entropía Condicional

Sean x, y dos v.a. discretas. La entropía de x dado y se define como

$$H(\mathrm{x}\,|\,\mathrm{y}) = \sum_j p(\mathrm{y} = y_j) H(\mathrm{x}\,|\,\mathrm{y} = y_j)$$

Claramente,

$$H(\mathrm{x}\,|\,\mathrm{y}) \leq H(\mathrm{x})$$

ya que la incertidumbre acerca de x se reduce (o se mantiene) al tener conocimiento de las realizaciones de y.

Entropía Condicional

Sean x, y dos v.a. discretas. La entropía de x dado y se define como

$$H(\mathrm{x}\,|\,\mathrm{y}) = \sum_j p(\mathrm{y} = y_j) H(\mathrm{x}\,|\,\mathrm{y} = y_j)$$

Claramente,

$$H(\mathrm{x}\,|\,\mathrm{y}) \leq H(\mathrm{x})$$

ya que la incertidumbre acerca de x se reduce (o se mantiene) al tener conocimiento de las realizaciones de y.

Además:

$$H(\mathrm{x}, \mathrm{y}) = H(\mathrm{x}\,|\,\mathrm{y}) + H(\mathrm{y})$$

Información Mutua

Información Mutua

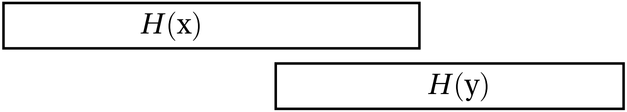
La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$

Información Mutua

La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

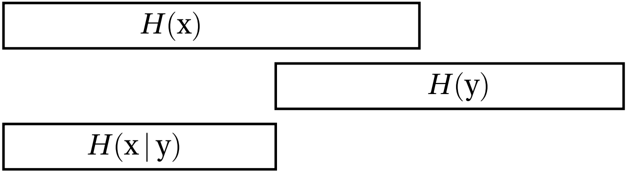
$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$



Información Mutua

La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

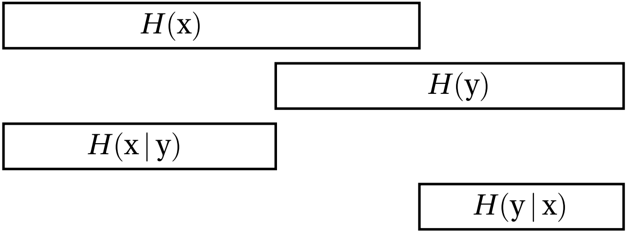
$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$



Información Mutua

La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

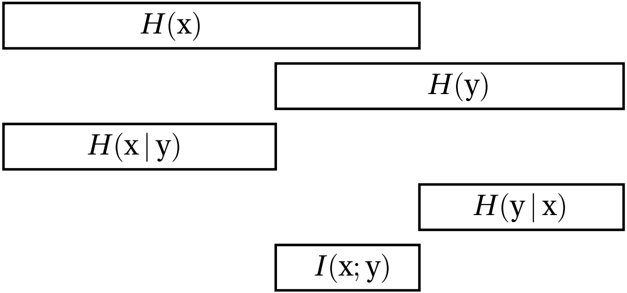
$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$



Información Mutua

La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

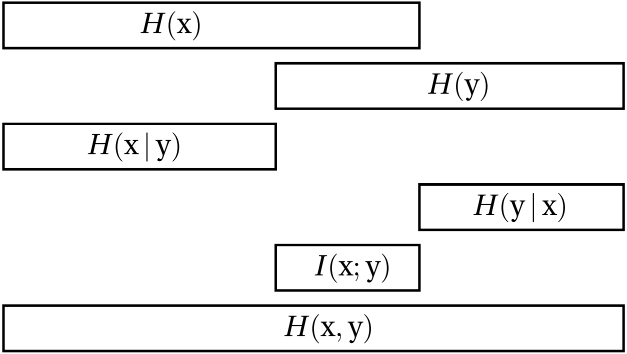
$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$



Información Mutua

La información mutua entre dos v.a. x, y, denotada por $I(\mathrm{x};\mathrm{y})$, mide cuánto nos "dice" y acerca de x:

$$I(\mathrm{x};\mathrm{y}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{y}) = H(\mathrm{y}) - H(\mathrm{y}\,|\,\mathrm{x}) = H(\mathrm{x}) + H(\mathrm{y}) - H(\mathrm{x},\mathrm{y})$$



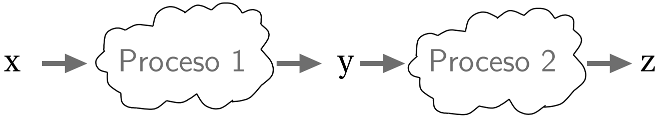
"Data Processing Inequality"

"Data Processing Inequality"

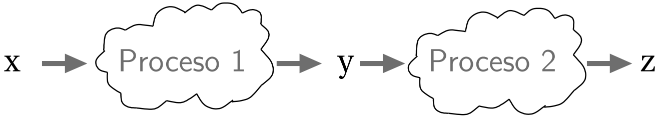
$\mathbf{X}$

"Data Processing Inequality"

"DATA PROCESSING INEQUALITY"



"Data Processing Inequality"

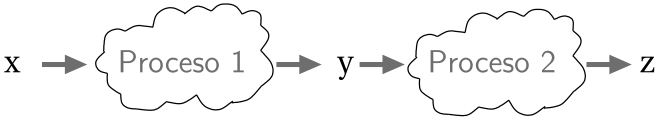


Si Proceso 2 es independiente de x así como del Proceso 1, entonces

$$p_{\mathrm{x},\mathrm{z}\,|\,\mathrm{y}}(x,z|y) = p_{\mathrm{x}\,|\,\mathrm{y}}(x|y)\,p_{\mathrm{z}\,|\,\mathrm{y}}(z|y), \quad \forall x,y,z,$$

es decir, x se vuelve independiente de z cuando y es conocido.

"DATA PROCESSING INEQUALITY"



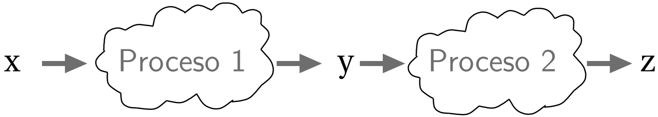
Si Proceso 2 es independiente de x así como del Proceso 1, entonces

$$p_{\mathrm{x,z}\,|\,\mathrm{y}}(x,z|y) = p_{\mathrm{x}\,|\,\mathrm{y}}(x|y)p_{\mathrm{z}\,|\,\mathrm{y}}(z|y), \quad \forall x,y,z,$$

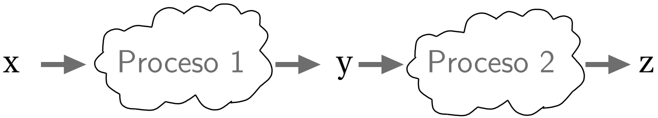
es decir, x se vuelve independiente de z cuando y es conocido. Se dice entonces que x, y, z satisfacen la Cadena de Markov

$$\mathrm{x} \leftrightarrow \mathrm{y} \leftrightarrow \mathrm{z}$$

"Data Processing Inequality"



“Data Processing Inequality”

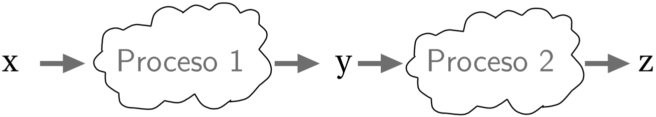


Teorema (Data Processing Inequality)

Si la cadena de Markov $\mathrm{x} \leftrightarrow \mathrm{y} \leftrightarrow \mathrm{z}$ *se cumple, entonces*

$$I(\mathrm{x};\mathrm{y}) \geq I(\mathrm{x};\mathrm{z})$$
$$I(\mathrm{y};\mathrm{z}) \geq I(\mathrm{x};\mathrm{z})$$

"Data Processing Inequality"



Teorema (Data Processing Inequality)

Si la cadena de Markov $\mathrm{x} \leftrightarrow \mathrm{y} \leftrightarrow \mathrm{z}$ *se cumple, entonces*

$$I(\mathrm{x};\mathrm{y}) \geq I(\mathrm{x};\mathrm{z})$$
$$I(\mathrm{y};\mathrm{z}) \geq I(\mathrm{x};\mathrm{z})$$

Demostración.

$$I(\mathrm{x};\mathrm{z}) = H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{z}) \leq H(\mathrm{x}) - H(\mathrm{x}\,|\,\mathrm{z},\mathrm{y}) = I(\mathrm{x};\mathrm{y})$$ □

La Divergencia de Kullback-Leibler

La Divergencia de Kullback-Leibler

- Medida de discrepancia entre dos distribuciones estadísticas p, q:

$$D_{KL}(p;q) \triangleq \sum_i p(i) \log \frac{p(i)}{q(i)}$$

La Divergencia de Kullback-Leibler

- Medida de discrepancia entre dos distribuciones estadísticas p, q:

$$D_{KL}(p;q) \triangleq \sum_i p(i) \log \frac{p(i)}{q(i)}$$

- $D_{KL}(p;q) \geq 0$, con $D_{KL}(p;q) = 0 \iff p \equiv q$.

La Divergencia de Kullback-Leibler

- Medida de discrepancia entre dos distribuciones estadísticas p, q:

$$D_{KL}(p;q) \triangleq \sum_i p(i) \log \frac{p(i)}{q(i)}$$

- $D_{KL}(p;q) \geq 0$, con $D_{KL}(p;q) = 0 \iff p \equiv q$.
- ¡No es simétrica!: $D_{KL}(p;q) \neq D_{KL}(q;p)$

La Divergencia de Kullback-Leibler

- Medida de discrepancia entre dos distribuciones estadísticas p, q:

$$D_{KL}(p;q) \triangleq \sum_i p(i) \log \frac{p(i)}{q(i)}$$

- $D_{KL}(p;q) \geq 0$, con $D_{KL}(p;q) = 0 \iff p \equiv q$.
- ¡No es simétrica!: $D_{KL}(p;q) \neq D_{KL}(q;p)$
- La información mutua entre dos variables x, y, puede escribirse como

$$I(\mathrm{x};\mathrm{y}) = D_{KL}(p(\mathrm{x},\mathrm{y})\,;\,p(\mathrm{x})p(\mathrm{y}))$$

Propiedad de la Equipartición Asintótica (AEP)

Sea x^n una secuencia de n v.a. discretas i.i.d., donde $\mathrm{x}[k] \sim p$, $\forall k$. Luego

$$\frac{1}{n}\log p(x^n) = \frac{1}{n}\sum_{k=1}^{n}\log p(\mathrm{x}[k])$$

es una variable aleatoria.

Propiedad de la Equipartición Asintótica (AEP)

Sea x^n una secuencia de n v.a. discretas i.i.d., donde $\mathrm{x}[k] \sim p$, $\forall k$. Luego

$$\frac{1}{n} \log p(x^n) = \frac{1}{n} \sum_{k=1}^{n} \log p(\mathrm{x}[k])$$

es una variable aleatoria. Más aún, por la Ley de los Grandes Números,

$$-\frac{1}{n} \sum_{k=1}^{n} \log p(\mathrm{x}[k]) \xrightarrow{n \to \infty} \mathrm{E}[-\log(p(\mathrm{x}))] = H(\mathrm{x})$$

en probabilidad.

Propiedad de la Equipartición Asintótica (AEP)

Sea x^n una secuencia de n v.a. discretas i.i.d., donde $\mathrm{x}[k] \sim p$, $\forall k$. Luego

$$\frac{1}{n}\log p(x^n) = \frac{1}{n}\sum_{k=1}^{n}\log p(\mathrm{x}[k])$$

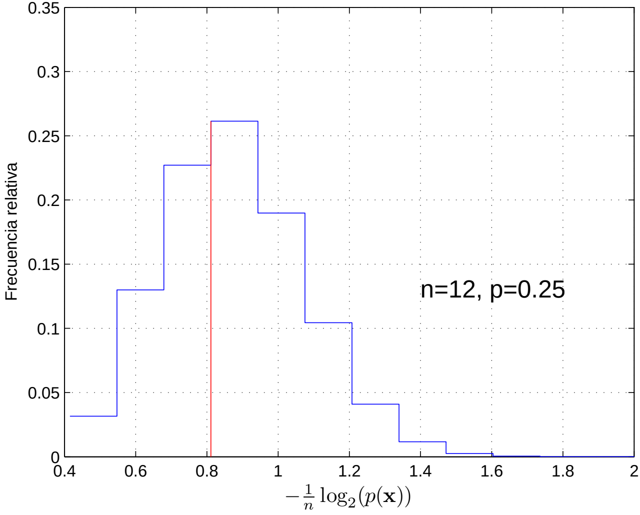
es una variable aleatoria. Más aún, por la Ley de los Grandes Números,

$$-\frac{1}{n}\sum_{k=1}^{n}\log p(\mathrm{x}[k]) \xrightarrow{n\to\infty} \mathrm{E}[-\log(p(\mathrm{x}))] = H(\mathrm{x})$$

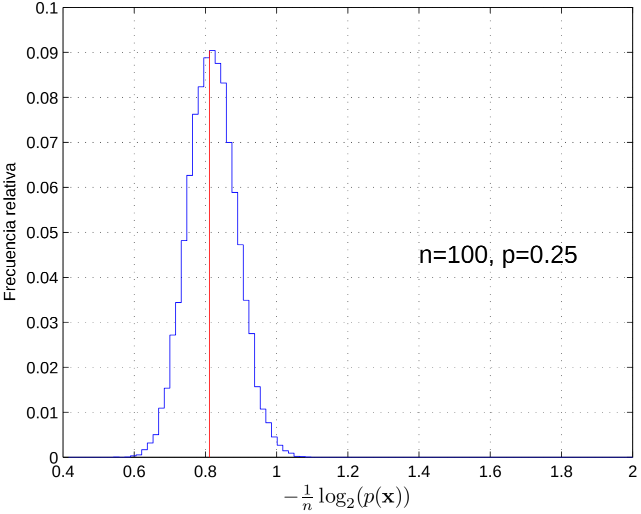
en probabilidad. Ésto es, para n grande, cualquier realización x^n de x^n es tal que, con alta probabilidad,

$$-\frac{1}{n}\sum_{k=1}^{n}\log p(x[k]) \simeq H(\mathrm{x})$$

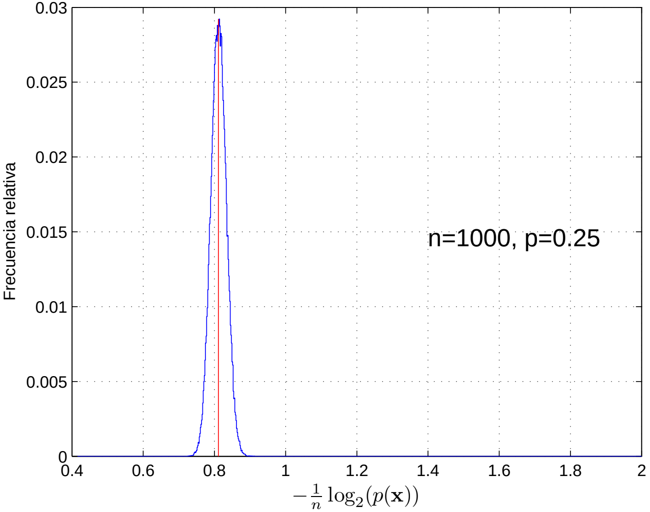
Propiedad de la Equipartición Asintótica (AEP)



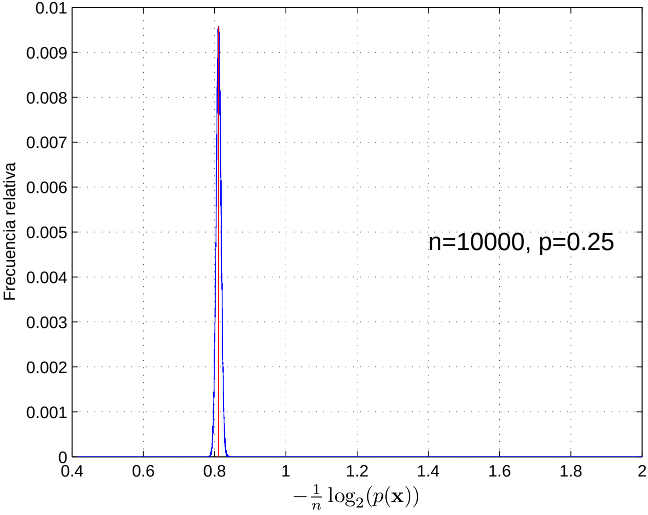
Propiedad de la Equipartición Asintótica (AEP)



Propiedad de la Equipartición Asintótica (AEP)



Propiedad de la Equipartición Asintótica (AEP)



Propiedad de la Equipartición Asintótica (AEP)

Definición (Conjunto Típico)

El conjunto típico $\mathcal{A}_\epsilon^{(n)}$ asociado a la secuencia de v.a. x^n es aquel formado por las secuencias $\mathcal{A}_\epsilon^n$ tales que

$$2^{-n[H(\mathrm{x})+\epsilon]} \leq \Pr\{x^n\} \leq 2^{-n[H(\mathrm{x})-\epsilon]}$$

Propiedad de la Equipartición Asintótica (AEP)

Definición (Conjunto Típico)

El conjunto típico $\mathcal{A}_\epsilon^{(n)}$ asociado a la secuencia de v.a. x^n es aquel formado por las secuencias $\mathcal{A}_\epsilon^n$ tales que

$$2^{-n[H(\mathrm{x})+\epsilon]} \leq \Pr\{x^n\} \leq 2^{-n[H(\mathrm{x})-\epsilon]}$$

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$, *entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1-\epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$*, entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$*, entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n}\log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1-\epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- Para toda secuencia x^n en $\mathcal{A}_\epsilon^{(n)}$, $-\frac{1}{n}\log p(x^n)$ es muy cercano a $H(\mathrm{x})$.

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$*, entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- Para toda secuencia x^n en $\mathcal{A}_\epsilon^{(n)}$, $-\frac{1}{n}\log p(x^n)$ es muy cercano a $H(\mathrm{x})$.
  $\Rightarrow$ Las secuencias en x^n son “cuasi” equi-probables.

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$, *entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- Con probabilidad tendiente a 1, las realizaciones de x^n pertenecen $\mathcal{A}_\epsilon^{(n)}$.

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$*, entonces* $\left| -\frac{1}{n} \sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x}) \right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- Con probabilidad tendiente a 1, las realizaciones de x^n pertenecen $\mathcal{A}_\epsilon^{(n)}$.
  $\Rightarrow$ Si n es grande, podemos olvidarnos de las secuencias $x^n \notin \mathcal{A}_\epsilon^{(n)}$ (que son la mayoría).

PROPIEDAD DE LA EQUIPARTICIÓN ASINTÓTICA (AEP)

TEOREMA (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$, *entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n}\log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1-\epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- $\mathcal{A}_\epsilon^{(n)}$ contiene "aproximadamente" $2^{nH(\mathrm{x})}$ secuencias.

Propiedad de la Equipartición Asintótica (AEP)

Teorema (AEP)

- *Si* $x^n \in \mathcal{A}_\epsilon^{(n)}$*, entonces* $\left|-\frac{1}{n}\sum_{k=1}^{n} \log p(x[k]) - H(\mathrm{x})\right| \leq \epsilon$.
- $\Pr\{\mathcal{A}_\epsilon^{(n)}\} > 1 - \epsilon$ *para* n *suficientemente grande.*
- $|\mathcal{A}_\epsilon^{(n)}| \leq 2^{n[H(\mathrm{x})+\epsilon]}$.
- $|\mathcal{A}_\epsilon^{(n)}| \geq 2^{n[H(\mathrm{x})-\epsilon]}$ *para* n *suficientemente grande.*

En palabras...

- $\mathcal{A}_\epsilon^{(n)}$ contiene "aproximadamente" $2^{nH(\mathrm{x})}$ secuencias.
  $\Rightarrow$ Basta utilizar (aproximadamente) $nH(\mathrm{x})$ bits para indexar cada secuencia de largo n con probabilidad no-negligible de ocurrir.

Propiedad de la Equipartición Asintótica (AEP)

Ejemplo

Sea x^n una secuencia de variables aleatorias independientes, $\mathrm{x}[k] \sim$ Bernoulli$(1/4)$. En este caso, $H(\mathrm{x}) = 0{,}8113$ [bit]. Para $n = 12$:

- Nro. total de secuencias binarias posibles: $2^{12} = 4096$.
- Nro. de secuencias en $\mathcal{A}_\epsilon^{(n)} \simeq 2^{nH(\mathrm{x})} = 852$.
- Porcentaje de secuencias en $\mathcal{A}_\epsilon^{(n)}$ respecto del total:

$$\frac{2^{nH(\mathrm{x})}}{2^n} = 2^{-n[1-H(\mathrm{x})]} = 2^{-0{,}1887n} \simeq 20\,\%$$

- Para $n = 20$, dicho porcentaje cae a $7\,\%$.
- Para $n = 40$, se reduce a $0{,}5\,\%$.

Sumario

Entropía Diferencial

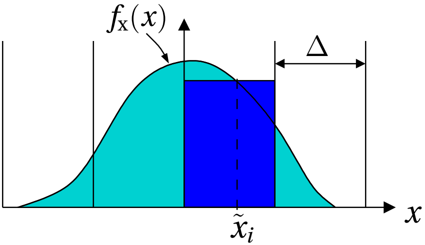
Entropía Diferencial

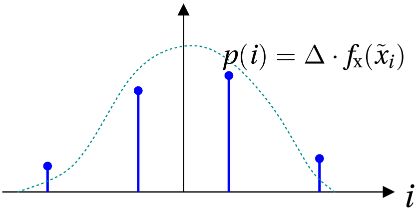
¿Cómo extender las nociones anteriores a variables aleatorias continuas?

ENTROPÍA DIFERENCIAL

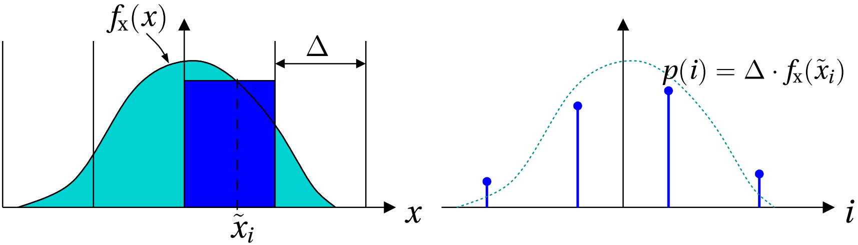
¿Cómo extender las nociones anteriores a variables aleatorias continuas? Sea x una v.a. continua con densidad de probabilidad f_{x}, y sea x_{Δ} la v.a. discreta correspondiente a la versión cuantizada de x con paso Δ. Por simplicidad, asumamos que f_{x} es continua.

Entropía Diferencial



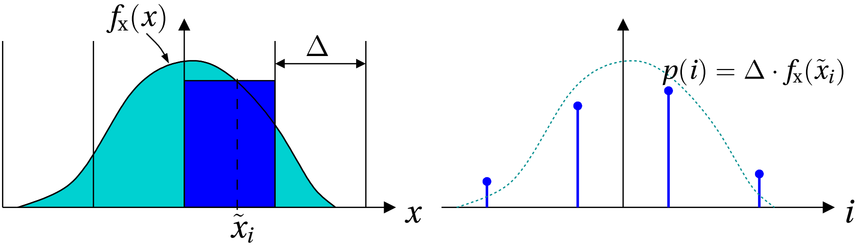


Entropía Diferencial



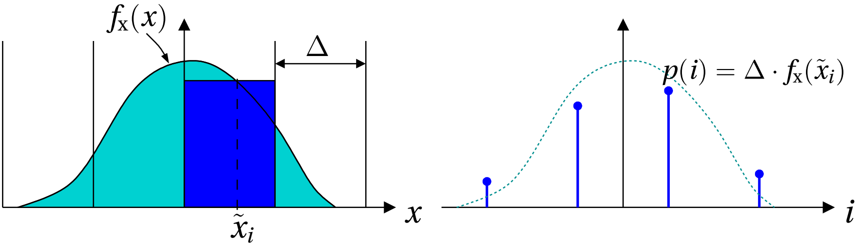
$$H(x_\Delta) = -\sum_i p(i) \log(p(i))$$

ENTROPÍA DIFERENCIAL



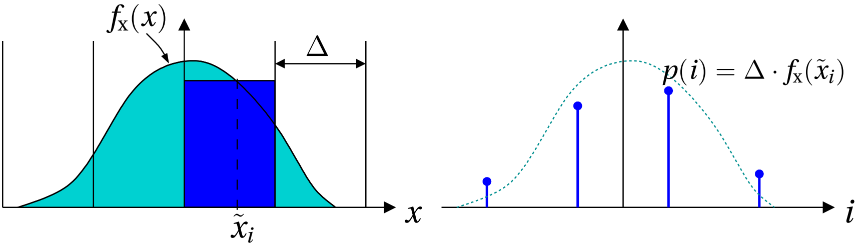
$$H(\mathrm{x}_\Delta) = -\sum_i p(i)\log(p(i)) = -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i))$$

Entropía Diferencial



$$
\begin{aligned}
H(\mathrm{x}_\Delta) &= -\sum\nolimits_i p(i)\log(p(i)) = -\sum\nolimits_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)) \\
&= -\sum\nolimits_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(f_{\mathrm{x}}(\tilde{x}_i)) - \log(\Delta)\sum\nolimits_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)
\end{aligned}
$$

Entropía Diferencial

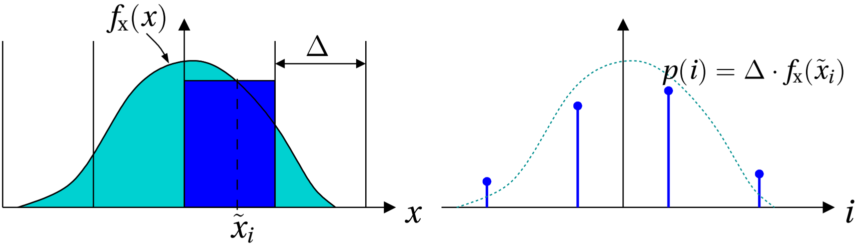


$$H(\mathrm{x}_\Delta) = -\sum_i p(i)\log(p(i)) = -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i))$$
$$= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(f_{\mathrm{x}}(\tilde{x}_i)) - \log(\Delta)\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)$$

Tomando límite cuando $\Delta \to 0$,

$$H(\mathrm{x}) = \lim_{\Delta \to 0} H(\mathrm{x}_\Delta) = \int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx - \lim_{\Delta \to 0} \log(\Delta),$$

ENTROPÍA DIFERENCIAL



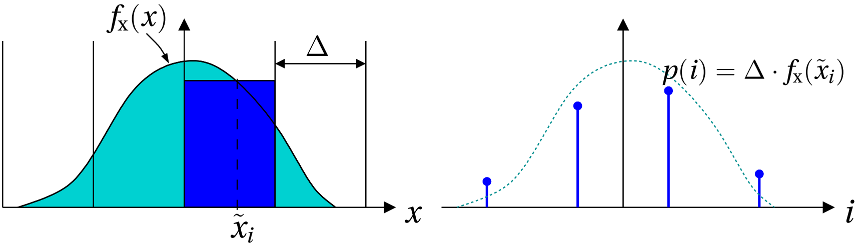
$$H(\mathrm{x}_\Delta) = -\sum_i p(i)\log(p(i)) = -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i))$$
$$= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(f_{\mathrm{x}}(\tilde{x}_i)) - \log(\Delta)\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)$$

Tomando límite cuando $\Delta \to 0$,

$$H(\mathrm{x}) = \lim_{\Delta \to 0} H(\mathrm{x}_\Delta) = \int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx - \lim_{\Delta \to 0} \log(\Delta),$$

¡La entropía de x es infinita!

Entropía Diferencial



$$H(\mathrm{x}_\Delta) = -\sum_i p(i)\log(p(i)) = -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i))$$
$$= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(f_{\mathrm{x}}(\tilde{x}_i)) - \log(\Delta)\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)$$

Tomando límite cuando $\Delta \to 0$,

$$H(\mathrm{x}) = \lim_{\Delta\to 0} H(\mathrm{x}_\Delta) = \int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx - \lim_{\Delta\to 0}\log(\Delta),$$

¡La entropía de x es infinita! ¿Cómo evaluar la entropía de x entonces?

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{“algo”}.$$

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{“algo”}.$$

Sea $y \sim \mathrm{U}(-1/2, 1/2)$.

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{“algo”}.$$

Sea $\mathrm{y} \sim \mathrm{U}(-1/2, 1/2)$.
Entonces

$$H(\mathrm{x}_\Delta) - H(\mathrm{y}_\Delta) = -\sum_i p(i)\log(p(i)) + \sum_j \Delta\log(\Delta)$$

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{``algo''}.$$

Sea $\mathrm{y} \sim \mathrm{U}(-1/2, 1/2)$.
Entonces

$$\begin{aligned} H(\mathrm{x}_\Delta) - H(\mathrm{y}_\Delta) &= -\sum_i p(i)\log(p(i)) + \sum_j \Delta \log(\Delta) \\ &= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)) + \log(\Delta) \end{aligned}$$

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{“algo”}.$$

Sea $\mathrm{y} \sim \mathrm{U}(-1/2, 1/2)$.
Entonces

$$\begin{aligned} H(\mathrm{x}_\Delta) - H(\mathrm{y}_\Delta) &= -\sum_i p(i)\log(p(i)) + \sum_j \Delta \log(\Delta) \\ &= -\sum_i \Delta \cdot f_\mathrm{x}(\tilde{x}_i)\log(\Delta \cdot f_\mathrm{x}(\tilde{x}_i)) + \log(\Delta) \\ &= -\sum_i \Delta \cdot f_\mathrm{x}(\tilde{x}_i)\log(f_\mathrm{x}(\tilde{x}_i)) + \log(\Delta)\left[1 - \sum_i \Delta \cdot f_\mathrm{x}(\tilde{x}_i)\right] \end{aligned}$$

Entropía Diferencial

Idea: Comparar la entropía de x con la de otra variable aleatoria continua (ambas son infinitas).

$$\infty - \infty = \text{“algo”}.$$

Sea $\mathrm{y} \sim \mathrm{U}(-1/2, 1/2)$.
Entonces

$$\begin{aligned} H(\mathrm{x}_\Delta) - H(\mathrm{y}_\Delta) &= -\sum_i p(i)\log(p(i)) + \sum_j \Delta \log(\Delta) \\ &= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(\Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)) + \log(\Delta) \\ &= -\sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\log(f_{\mathrm{x}}(\tilde{x}_i)) + \log(\Delta)\left[1 - \sum_i \Delta \cdot f_{\mathrm{x}}(\tilde{x}_i)\right] \end{aligned}$$

Luego

$$\lim_{\Delta \to 0} \left[H(\mathrm{x}_\Delta) - H(\mathrm{y}_\Delta)\right] = -\int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx.$$

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x) \log(f_{\mathrm{x}}(x)) dx$$

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x) \log(f_{\mathrm{x}}(x)) dx$$

- $h(\mathrm{x})$ también se mide en bits.

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x) \log(f_{\mathrm{x}}(x)) dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).
- $h(\mathrm{x})$ puede ser negativa.

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq - \int f_{\mathrm{x}}(x) \log(f_{\mathrm{x}}(x)) dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).
- $h(\mathrm{x})$ puede ser negativa.
- Si el soporte de f_{x} tiene medida cero (x discreta), entonces $h(\mathrm{x}) = -\infty$.

ENTROPÍA DIFERENCIAL

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).
- $h(\mathrm{x})$ puede ser negativa.
- Si el soporte de f_{x} tiene medida cero (x discreta), entonces $h(\mathrm{x}) = -\infty$.
- Si $x \sim$U(1), entonces $h(x) = 0$.

Entropía Diferencial

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x)\log(f_{\mathrm{x}}(x))dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).
- $h(\mathrm{x})$ puede ser negativa.
- Si el soporte de f_{x} tiene medida cero (x discreta), entonces $h(\mathrm{x}) = -\infty$.
- Si $x \sim$U(1), entonces $h(x) = 0$.
- Si $x \sim \mathcal{N}(0, \sigma^2)$, entonces $h(x) = \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma^2)$.

ENTROPÍA DIFERENCIAL

Así, se define la enropía diferencial de una v.a. continua x como

$$h(\mathrm{x}) \triangleq -\int f_{\mathrm{x}}(x) \log(f_{\mathrm{x}}(x)) dx$$

- $h(\mathrm{x})$ también se mide en bits.
- La diferencia entre dos entroopías diferenciales es entropía "absoluta" (no diferencial).
- $h(\mathrm{x})$ puede ser negativa.
- Si el soporte de f_{x} tiene medida cero (x discreta), entonces $h(\mathrm{x}) = -\infty$.
- Si $x \sim$U(1), entonces $h(x) = 0$.
- Si $x \sim \mathcal{N}(0, \sigma^2)$, entonces $h(x) = \frac{1}{2}\log(2\pi\, \mathrm{e}\, \sigma^2)$.
- Para constantes a, b, se tiene que $h(a\,\mathrm{x} + b) = \log(a) + h(\mathrm{x})$.

Máxima Entropía Diferencial

- Entre las distribuciones con un mismo soporte, la entropía diferencial es maximizada por la distribución uniforme.

Máxima Entropía Diferencial

- Entre las distribuciones con un mismo soporte, la entropía diferencial es maximizada por la distribución uniforme.
- Entre las distribuciones con una misma varianza, la entropía diferencial es maximizada por la distribución normal.

Non-Linear Information Processing

- Sistema desconocido, con entrada $\mathbf{x}$, salida $\mathbf{z}$ (vectores aleatorios).

Non-Linear Information Processing

- Sistema desconocido, con entrada $\mathbf{x}$, salida $\mathbf{z}$ (vectores aleatorios).
- Diseñar sistema $f_{\boldsymbol{w}}(\cdot)$ tal que $f_{\boldsymbol{w}}(\mathbf{x}) \simeq \mathbf{z}$, donde $\boldsymbol{w}$ es un vector de parámetros.
- Ejemplos:
    - Filitros no lineales adaptivos

Non-Linear Information Processing

- Sistema desconocido, con entrada $\mathbf{x}$, salida $\mathbf{z}$ (vectores aleatorios).
- Diseñar sistema $f_{\boldsymbol{w}}(\cdot)$ tal que $f_{\boldsymbol{w}}(\mathbf{x}) \simeq \mathbf{z}$, donde $\boldsymbol{w}$ es un vector de parámetros.
- Ejemplos:
  - Filitros no lineales adaptivos
  - *Feature Extraction*
  - *Clustering*

Non-Linear Information Processing

Idea General:

Non-Linear Information Processing

Idea General:

- Usar entropía en lugar de la varianza (por ejemplo, en filtrado adaptivo, como medida del error):

Non-Linear Information Processing

Idea General:

- Usar entropía en lugar de la varianza (por ejemplo, en filtrado adaptivo, como medida del error):

$$\mathrm{E}\left[\|\mathbf{z} - f_w(\mathbf{x})\|^2\right] \longrightarrow h(\mathbf{z} - f_w(\mathbf{x})).$$

NON-LINEAR INFORMATION PROCESSING

Idea General:

Non-Linear Information Processing

Idea General:

- Usar información mutua en lugar de la correlación.

Non-Linear Information Processing

Idea General:

- Usar información mutua en lugar de la correlación. Por ejemplo:
  - En Independent Component Analysis, la idea es encontrar parámetros $\boldsymbol{w}$ con los que las salidas $\mathbf{y} = f_{\boldsymbol{w}}(\mathbf{x})$ sean tan independientes como sea posible (minimizar redundancia):

$$\min_{\boldsymbol{w}} I(\mathbf{y}) = \min_{\boldsymbol{w}} \left[\left(\sum_i h(\mathrm{y}_i) \right) - h(\mathbf{y}) \right]$$

Compresión de Datos

Compresión de Datos

Compresión Sin Pérdida (lossless): Reversible

Compresión de Datos

Compresión Sin Pérdida (lossless): Reversible

- Genéricos (ZIP, LZ77, RAR, etc.)

Compresión de Datos

Compresión Sin Pérdida (lossless): Reversible

- Genéricos (ZIP, LZ77, RAR, etc.)
- Dependientes de la fuente (Lossless JPEG, FLAC, APE, WavPack)

Compresión de Datos

Compresión Sin Pérdida (lossless): Reversible

- Genéricos (ZIP, LZ77, RAR, etc.)
- Dependientes de la fuente (Lossless JPEG, FLAC, APE, WavPack)

Aplicable a fuentes con entropía finita (ya digitalizadas).

Compresión de Datos

Compresión Sin Pérdida (lossless): Reversible

- Genéricos (ZIP, LZ77, RAR, etc.)
- Dependientes de la fuente (Lossless JPEG, FLAC, APE, WavPack)

Aplicable a fuentes con entropía finita (ya digitalizadas).
Resultado fundamental:

Teorema (Lossless Compression)

Sea x *una fuente aleatoria discreta con entropía* $H(\mathrm{x})$*. Entonces la mínima cantidad de bits necesarios para representar* x *satisface*

$$R_{min} \geq H(\mathrm{x})$$

Compresión de Datos

Compresión de Datos

Compresión Con Pérdida (lossy): Irreversible

Compresión de Datos

Compresión Con Pérdida (lossy): Irreversible

- Audio (OGG, MP3, AAC, etc.)

Compresión de Datos

Compresión Con Pérdida (lossy): Irreversible

- Audio (OGG, MP3, AAC, etc.)
- Imagen (JPEG, JPEG2000, DjVu)

Compresión de Datos

Compresión Con Pérdida (lossy): Irreversible

- Audio (OGG, MP3, AAC, etc.)
- Imagen (JPEG, JPEG2000, DjVu)
- Aplicable a fuentes con entropía finita o infinita.

Compresión de Datos

Compresión Con Pérdida (lossy): Irreversible

- Audio (OGG, MP3, AAC, etc.)
- Imagen (JPEG, JPEG2000, DjVu)
- Aplicable a fuentes con entropía finita o infinita.
- Compromiso Tasa/Distorsión (Rate-Distortion Theory).

Resultado fundamental:

Teorema (Lossy Compression)

Sea x *una fuente aleatoria. Entonces la mínima cantidad de bits necesarios para representar* x *con distorsión menor que* D *satisface*

$$R^D_{min} \geq R(D) \triangleq \min_{\mathrm{y}:d(\mathrm{x},\mathrm{y})\leq D} I(\mathrm{x};\mathrm{y})$$

Compresión de Datos

Ejemplo:

- Fuente: $\mathrm{x} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{x}}^2)$.

Compresión de Datos

Ejemplo:

- Fuente: $\mathrm{x} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{x}}^2)$.
- Distorsión: $d(\mathrm{x}, \mathrm{y}) = \mathrm{E}\left[(\mathrm{x} - \mathrm{y})^2\right]$

Compresión de Datos

Ejemplo:

- Fuente: $\mathrm{x} \sim \mathcal{N}(0, \sigma_{\mathrm{x}}^2)$.
- Distorsión: $d(\mathrm{x}, \mathrm{y}) = \mathrm{E}\left[(\mathrm{x}-\mathrm{y})^2\right]$
- Entonces, si $\mathrm{n} \triangleq \mathrm{y}-\mathrm{x}$, cumpliéndose que $\mathrm{E}\left[\mathrm{n}^2\right] \leq D$:

$$
\begin{aligned}
I(\mathrm{x};\mathrm{y}) &= h(\mathrm{x}) - h(\mathrm{x}\,|\,\mathrm{y}) \\
&= \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{y}-\mathrm{n}\,|\,\mathrm{y}) \\
&= \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{n}\,|\,\mathrm{y}) \\
&\geq \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{n}) \\
&\geq \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - \frac{1}{2}\log(2\pi\,\mathrm{e}\,D) = \frac{1}{2}\log\frac{\sigma_{\mathrm{x}}^2}{D}
\end{aligned}
$$

Compresión de Datos

Ejemplo:

- Fuente: $\mathrm{x} \sim \mathcal{N}(0, \sigma_{\mathrm{x}}^2)$.
- Distorsión: $d(\mathrm{x}, \mathrm{y}) = \mathrm{E}\left[(\mathrm{x}-\mathrm{y})^2\right]$
- Entonces, si $\mathrm{n} \triangleq \mathrm{y}-\mathrm{x}$, cumpliéndose que $\mathrm{E}\left[\mathrm{n}^2\right] \leq D$:

$$
\begin{aligned}
I(\mathrm{x};\mathrm{y}) &= h(\mathrm{x}) - h(\mathrm{x}\,|\,\mathrm{y}) \\
&= \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{y}-\mathrm{n}\,|\,\mathrm{y}) \\
&= \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{n}\,|\,\mathrm{y}) \\
&\geq \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - h(\mathrm{n}) \\
&\geq \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{x}}^2) - \frac{1}{2}\log(2\pi\,\mathrm{e}\,D) = \frac{1}{2}\log\frac{\sigma_{\mathrm{x}}^2}{D} = R(D)
\end{aligned}
$$

Comunicación de Datos

Comunicación de Datos

Canal discreto con errores aleatorios

Comunicación de Datos

Canal discreto con errores aleatorios

- Fuente de datos discreta.

Comunicación de Datos

Canal discreto con errores aleatorios

- Fuente de datos discreta.
- Transmisión de símbolos por canales (Internet, etc.)

Comunicación de Datos

Canal discreto con errores aleatorios

- Fuente de datos discreta.
- Transmisión de símbolos por canales (Internet, etc.)
- Almacenamiento de datos en medios físicos (CD, DVD, Discos Duros).

Comunicación de Datos

Canal discreto con errores aleatorios

- Fuente de datos discreta.
- Transmisión de símbolos por canales (Internet, etc.)
- Almacenamiento de datos en medios físicos (CD, DVD, Discos Duros).
- ¿Cómo transmitir/almacenar a la mayor tasa con baja probabilidad de error?.

COMUNICACIÓN DE DATOS

Canal discreto con errores aleatorios

- Fuente de datos discreta.
- Transmisión de símbolos por canales (Internet, etc.)
- Almacenamiento de datos en medios físicos (CD, DVD, Discos Duros).
- ¿Cómo transmitir/almacenar a la mayor tasa con baja probabilidad de error?.

Resultado fundamental:

TEOREMA (DISCRETE NOISY CHANNEL CAPACITY)

Sea $\mathcal{C}$ un canal aleatorio discreto. Entonces la máxima tasa de datos transmisible por $\mathcal{C}$ con probabilidad de error negligible (Capacidad de Canal) es

$$C = \max_{p(\mathrm{x})} I(\mathrm{x}; \mathcal{C}\,\mathrm{x}).$$

Comunicación de Datos

Comunicación de Datos

Canal analógico con ruido

Comunicación de Datos

Canal analógico con ruido

- Fuente de datos discreta s codificada en señal analógica x.

Comunicación de Datos

Canal analógico con ruido

- Fuente de datos discreta s codificada en señal analógica x.
- Restricción de potencia: $\sigma_{\mathrm{x}}^2 \leq P$.

Comunicación de Datos

Canal analógico con ruido

- Fuente de datos discreta s codificada en señal analógica x.
- Restricción de potencia: $\sigma_{\mathrm{x}}^2 \leq P$.
- Canal analógico: $\mathcal{C}$ (por ejemplo, $\mathcal{C}\,\mathrm{x} = \mathrm{x} + n$).

Comunicación de Datos

Canal analógico con ruido

- Fuente de datos discreta s codificada en señal analógica x.
- Restricción de potencia: $\sigma_{\mathrm{x}}^2 \leq P$.
- Canal analógico: $\mathcal{C}$ (por ejemplo, $\mathcal{C}\,\mathrm{x} = \mathrm{x} + n$).
- Decodificador recupera símbolos a partir de $\mathcal{C}\,\mathrm{x}$.

Comunicación de Datos

Canal analógico con ruido

- Fuente de datos discreta s codificada en señal analógica x.
- Restricción de potencia: $\sigma_{\mathrm{x}}^2 \leq P$.
- Canal analógico: $\mathcal{C}$ (por ejemplo, $\mathcal{C}\,\mathrm{x} = \mathrm{x} + n$).
- Decodificador recupera símbolos a partir de $\mathcal{C}\,\mathrm{x}$.

Resultado fundamental:

Teorema (Analog Noisy Channel Capacity)

Sea $\mathcal{C}$ un canal aleatorio continuo. Entonces la máxima tasa de datos transmisible por $\mathcal{C}$ con probabilidad de error negligible (Capacidad de Canal) es

$$C(P) = \max_{p(\mathrm{x}):E[\mathrm{x}^2]\leq P} I(\mathrm{x};\mathcal{C}\,\mathrm{x}).$$

Comunicación de Datos

Ejemplo: Canal Gaussiano

Comunicación de Datos

Ejemplo: Canal Gaussiano

- Canal: $\mathrm{y} \triangleq \mathcal{C}\,\mathrm{x} = G\,\mathrm{x} + \mathrm{n}$, con n indep. x, $\mathrm{n} \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$.

COMUNICACIÓN DE DATOS

Ejemplo: Canal Gaussiano

- Canal: $\mathrm{y} \triangleq \mathcal{C}\,\mathrm{x} = G\,\mathrm{x} + \mathrm{n}$, con n indep. x, $\mathrm{n} \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$.
- Si $\sigma_{\mathrm{x}}^2 \leq P$, se cumple que

$$\begin{aligned} I(\mathrm{x};\mathrm{y}) &= h(\mathrm{y}) - h(\mathrm{y}\,|\,\mathrm{x}) = h(\mathrm{y}) - h(\mathrm{x}+\mathrm{n}\,|\,\mathrm{x}) \\ &= h(\mathrm{y}) - h(\mathrm{n}\,|\,\mathrm{x}) = h(G\,\mathrm{x}+\mathrm{n}) - h(\mathrm{n}) \end{aligned}$$

Comunicación de Datos

Ejemplo: Canal Gaussiano

- Canal: $\mathrm{y} \triangleq \mathcal{C}\,\mathrm{x} = G\,\mathrm{x} + \mathrm{n}$, con n indep. x, $\mathrm{n} \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$.
- Si $\sigma_{\mathrm{x}}^2 \leq P$, se cumple que

$$\begin{aligned} I(\mathrm{x};\mathrm{y}) &= h(\mathrm{y}) - h(\mathrm{y}\,|\,\mathrm{x}) = h(\mathrm{y}) - h(\mathrm{x}+\mathrm{n}\,|\,\mathrm{x}) \\ &= h(\mathrm{y}) - h(\mathrm{n}\,|\,\mathrm{x}) = h(G\,\mathrm{x}+\mathrm{n}) - h(\mathrm{n}) \end{aligned}$$

Dado que $\sigma_{\mathrm{y}}^2 = G^2\sigma_{\mathrm{x}}^2 + \sigma_{\mathrm{n}}^2 \leq G^2 P + \sigma_{\mathrm{n}}^2$, se deduce que

$$\begin{aligned} I(\mathrm{x};\mathrm{y}) &\leq \frac{1}{2}\log(2\pi\,\mathrm{e}[G^2 P + \sigma_{\mathrm{n}}^2]) - \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{n}}^2) \\ &= \frac{1}{2}\log\left(1 + \frac{G^2 P}{\sigma_{\mathrm{n}}^2}\right) = \frac{1}{2}\log\left(1 + SNR\right) \end{aligned}$$

Comunicación de Datos

Ejemplo: Canal Gaussiano

- Canal: $\mathrm{y} \triangleq \mathcal{C}\,\mathrm{x} = G\,\mathrm{x} + \mathrm{n}$, con n indep. x, $\mathrm{n} \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$.
- Si $\sigma_{\mathrm{x}}^2 \leq P$, se cumple que

$$\begin{aligned} I(\mathrm{x};\mathrm{y}) &= h(\mathrm{y}) - h(\mathrm{y}\,|\,\mathrm{x}) = h(\mathrm{y}) - h(\mathrm{x}+\mathrm{n}\,|\,\mathrm{x}) \\ &= h(\mathrm{y}) - h(\mathrm{n}\,|\,\mathrm{x}) = h(G\,\mathrm{x}+\mathrm{n}) - h(\mathrm{n}) \end{aligned}$$

Dado que $\sigma_{\mathrm{y}}^2 = G^2\sigma_{\mathrm{x}}^2 + \sigma_{\mathrm{n}}^2 \leq G^2 P + \sigma_{\mathrm{n}}^2$, se deduce que

$$\begin{aligned} I(\mathrm{x};\mathrm{y}) &\leq \frac{1}{2}\log(2\pi\,\mathrm{e}[G^2P + \sigma_{\mathrm{n}}^2]) - \frac{1}{2}\log(2\pi\,\mathrm{e}\,\sigma_{\mathrm{n}}^2) \\ &= \frac{1}{2}\log\left(1 + \frac{G^2P}{\sigma_{\mathrm{n}}^2}\right) = \frac{1}{2}\log\left(1 + SNR\right) = C(P) \end{aligned}$$

Control Sobre Redes

Control Sobre Redes

- Planta $G(z)$, con polos inestables $\{p_i\}$.

Control Sobre Redes

- Planta $G(z)$, con polos inestables $\{p_i\}$.
- Canal discreto sin pérdidas ni retardo entre salida de G y entrada del controlador.

CONTROL SOBRE REDES

- Planta $G(z)$, con polos inestables $\{p_i\}$.
- Canal discreto sin pérdidas ni retardo entre salida de G y entrada del controlador.

Resultado fundamental:

TEOREMA (ENTROPÍA TOPOLÓGICA)

La mínima tasa promedio sobre el canal discreto de feedback con la cual se logra estabilidad de lazo cerrado está dada por

$$R_{min}^{stab} = \sum_i \log |p_i| \, .$$

GRACIAS POR SU ATENCIÓN

milan.derpich@usm.cl
http://profesores.elo.utfsm.cl/~mderpich/