# 16.36: Ingeniería de sistemas de comunicación

## Clase 2: entropía

**Eytan Modiano**

# Información contenida en una variable aleatoria

- **Variable aleatoria X**
  - **Resultado de un experimento aleatorio**
  - **Una V.A. discreta adopta valores de un conjunto infinito de resultados posibles**

    **PMF: $P(X = y) = P_x(y)$**

- **¿Cuánta información contiene el evento X = y?**
  - **¿Saldrá hoy el sol?**

    **Revelar el resultado de este experimento no proporciona información alguna**
  - **¿Ganarán los Celtics el campeonato de la NBA?**

    **Ya que esto es improbable, si es que sí, se proporciona más información que si es que no**

- **Los acontecimientos más improbables proporcionan más información que los que son más probables**

# Medida de la información

- $I(x_i)$ = Cantidad de información revelada por un resultado $X = x_i$
- Propiedades deseables de $I(x)$:

1. Si $P(x) = 1$ o $P(x) = 0$, entonces $I(x) = 0$
2. Si $0 < P(x) < 1$, entonces $I(x) > 0$
3. Si $P(x) < P(y)$, entonces $I(x) > I(y)$
4. Si x e y son acontecimientos independientes, entonces $I(x,y) = I(x)+I(y)$

- Las condiciones anteriores se satisfacen con: $I(x) = \mathrm{Log}_2(1/P(x))$
- La base del *log* no es crítica
  - Base 2 => información medida en *bits*

# Entropía

- Medida del contenido de información de una variable aleatoria
- $X \in \{x_1,\ldots,X_M\}$
- $H(X) = E[I(X)] = \sum P(x_i) \, Log_2(1/P(x_i))$
- Ejemplo: experimento binario
  - $X = x_1$ con probablidad p
  - $X = x_2$ con probabilidad (1-p)
  - $H(X) = pLog_2(1/p) + (1\text{-}p)Log_2(1/(1\text{-}p)) = H_b(p)$
  - H(X) se maximiza con p=1/2, $H_b(1/2) = 1$

    **No es de extrañar que el resultado de un experimento binario se pueda expresar con un solo *bit***

# Límites simples de la entropía

- **Teorema: dada una variable aleatoria con M valores posibles**
  - **$0 <= H(X) <= Log_2(M)$**

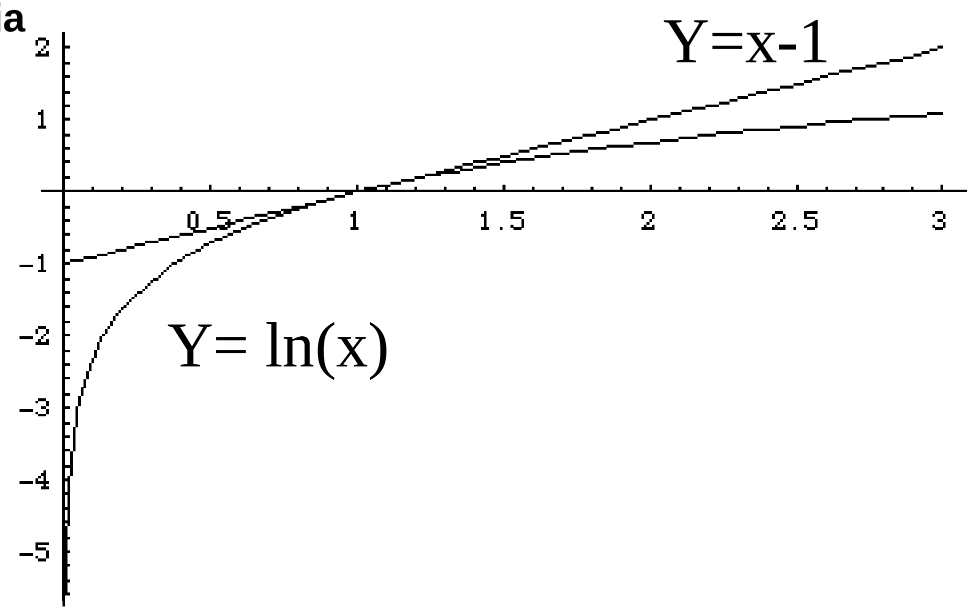
  **A) $H(X) = 0$ si y sólo si $P(x_i) = 1$ para cierto i**

  **B) $H(X) = Log_2(M)$ si y sólo si $P(x_i) = 1/M$ para todo i**

– **La demostración de A es obvia**

– **La de B requiere**
– **la desigualdad log:**

– **si x>0 entonces $ln(x) <= x-1$**
– **Igualdad si x=1**



# Demostración, (cont.)

Considere la suma $\sum_{i=1}^{M} P_i Log(\frac{1}{MP_i})$, mediante desigualdad log:

$$\leq \sum_{i=1}^{M} P_i(\frac{1}{MP_i} - 1) = \sum_{i=1}^{M} (\frac{1}{M} - P_i) = 0, \text{ igualdad si } \quad P_i = \frac{1}{M}$$

Si lo escribimos de otra manera:

$$\sum_{i=1}^{M} P_i Log(\frac{1}{MP_i}) = \sum_{i=1}^{M} P_i Log(\frac{1}{P_i}) + \sum_{i=1}^{M} P_i Log(\frac{1}{M}) \leq 0, \text{igualdad si } \quad P_i = \frac{1}{M}$$

Esto es, $\sum_{i=1}^{M} P_i Log(\frac{1}{P_i}) \leq \sum_{i=1}^{M} P_i Log(M) = Log(M)$

# Entropía conjunta

Entropía conj.: $H(X,Y) = \sum_{x,y} p(x,y)\log(\frac{1}{p(x,y)})$

Entropía condicional H(X | Y) = incertidumbre en X dado Y

$$H(X \mid Y=y) = \sum_{x} p(x \mid Y=y)\log(\frac{1}{p(x \mid Y=y)})$$

$$H(X \mid Y) = E[H(X \mid Y=y)] = \sum_{\mathrm{y}} \mathrm{p(Y=y)}H(X \mid Y=y)$$

$$H(X \mid Y) = \sum_{x,y} p(x,y)\log(\frac{1}{p(x \mid Y=y)})$$

En general: $X_1,...,X_n$ variables aleatorias

$$H(X_n \mid X_1,...,X_{n-1}) = \sum_{x_1,...,x_n} p(x_1,...,x_n)\log(\frac{1}{p(x_n \mid x_1,...,x_{n-1})})$$

# Reglas para la entropía

1. **Regla de la cadena:**

$$H(X_1, .., X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_2,X_1) + \ldots + H(X_n|X_{n-1}\ldots X_1)$$

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

3. **Si $X_1, .., X_n$ son independientes, entonces:**

$$H(X_1, .., X_n) = H(X_1) + H(X_2) + \ldots + H(X_n)$$

**Si también se distribuyen de forma idéntica (I.I.d), entonces:**

$$H(X_1, .., X_n) = nH(X_1)$$

4. **$H(X_1, .., X_n) \le H(X_1) + H(X_2) + \ldots + H(X_n)$ (con igualdad si son independientes)**

**Demostración: utilice la regla de la cadena y observe que $H(X|Y) < H(X)$**
**la entropía no aumenta con información nueva**

# Información mútua

- X, Y variables aleatorias
- Definición: $I(X;Y) = H(Y) - H(Y|X)$
- Observe que $H(Y|X) = H(X,Y) - H(X) \Rightarrow I(X;Y) = H(X)+H(Y) - H(X,Y)$
- $I(X;Y) = I(Y;X) = H(X) - H(X|Y)$
- Nota: $I(X,Y) >= 0$ (igualdad si son independientes)
  - Porque $H(Y) >= H(Y|X)$