Termo avanzada

Guía 3: entropía e información

Entropía

Definición

Experimento aleatorio con M resultados posibles, distribución de probabilidad p. **Entropía** de esta distribución:

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

Significado

La entropía mide el grado de incertidumbre acerca del resultado del experimento, porque es mínima cuando el resultado se conoce con certeza y máxima cuando se desconoce completamente.

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Certeza** $\Rightarrow$ $S = 0$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Certeza** $\Rightarrow$ $S = 0$

$$\log 1 = 0 \qquad 0 \log 0 = \lim_{\epsilon \to 0} \epsilon \log \epsilon = 0$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Certeza** $\Rightarrow \;\; S = 0$

$$\log 1 = 0 \qquad 0 \log 0 = \lim_{\epsilon \to 0} \epsilon \log \epsilon = 0$$

- **Cualquier** $p \Rightarrow \;\; S \geq 0$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Certeza** $\Rightarrow \;\; S = 0$

$$\log 1 = 0 \qquad 0 \log 0 = \lim_{\epsilon \to 0} \epsilon \log \epsilon = 0$$

- **Cualquier** $p \Rightarrow \;\; S \geq 0$

$$0 \leq p(r) \leq 1$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$.

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$S(p) - \log M$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$S(p) - \log M \quad = \quad S(p) - \sum_{r=1}^{M} p(r) \log M$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$S(p) - \log M \quad = \quad S(p) - \sum_{r=1}^{M} p(r) \log M = -\sum_{r=1}^{M} p(r) \log[Mp(r)]$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$\begin{aligned} S(p) - \log M &= S(p) - \sum_{r=1}^{M} p(r) \log M = -\sum_{r=1}^{M} p(r) \log[Mp(r)] \\ &= \sum_{r=1}^{M} p(r) \log \left[\frac{1}{Mp(r)} \right] \end{aligned}$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$\begin{aligned} S(p) - \log M &= S(p) - \sum_{r=1}^{M} p(r) \log M = -\sum_{r=1}^{M} p(r) \log[Mp(r)] \\ &= \sum_{r=1}^{M} p(r) \log\left[\frac{1}{Mp(r)}\right] \leq \sum_{r=1}^{M} p(r)\left[\frac{1}{Mp(r)} - 1\right] \end{aligned}$$

Significado

$$S(p) = -\sum_{r=1}^{M} p(r) \log p(r)$$

- **Máxima incertidumbre:** $p(r) = 1/M \;\Rightarrow\; S = \log M$
- **Cualquier** $p \Rightarrow\; S \leq \log M$. *Clave:* $\log x \leq x - 1$

$$\begin{aligned} S(p) - \log M &= S(p) - \sum_{r=1}^{M} p(r) \log M = -\sum_{r=1}^{M} p(r) \log[Mp(r)] \\ &= \sum_{r=1}^{M} p(r) \log\left[\frac{1}{Mp(r)}\right] \leq \sum_{r=1}^{M} p(r)\left[\frac{1}{Mp(r)} - 1\right] = 0 \end{aligned}$$

La varianza no mide eso también?

La varianza no mide eso también?

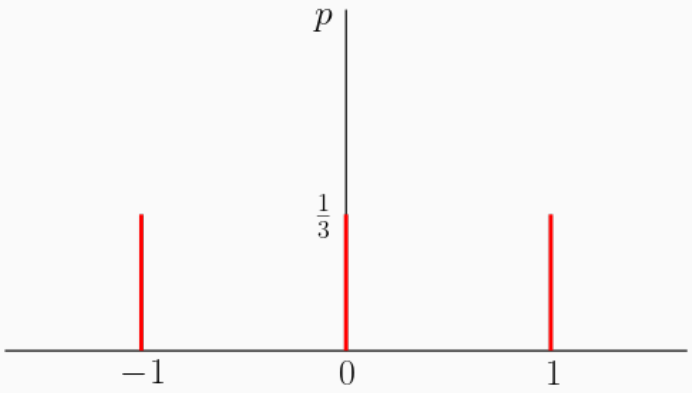
No.

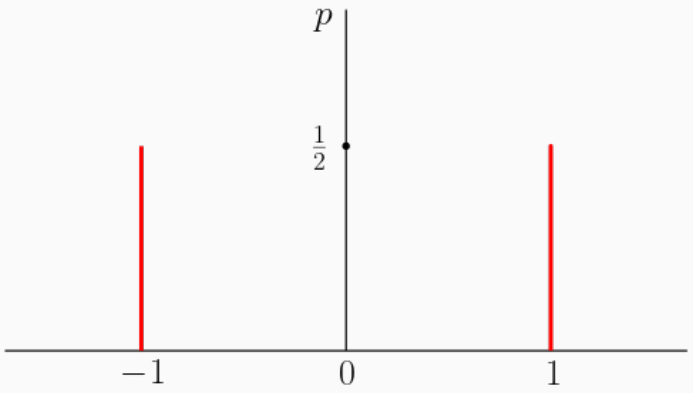
La varianza no mide eso también?

No. Ejemplo: experimento con tres resultados posibles, 0 y ± 1, y dos distribuciones de probabilidad.

La varianza no mide eso también?

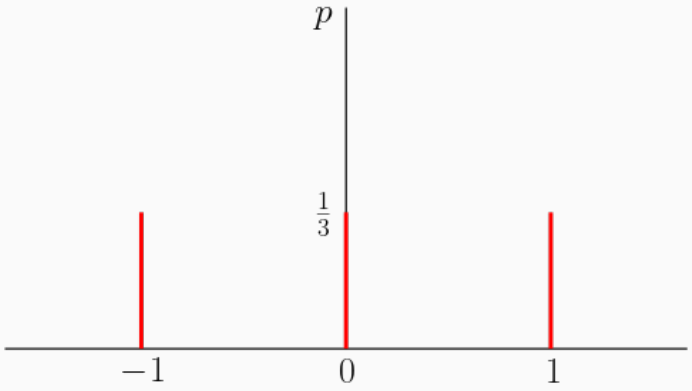
No. Ejemplo: experimento con tres resultados posibles, 0 y ± 1, y dos distribuciones de probabilidad.



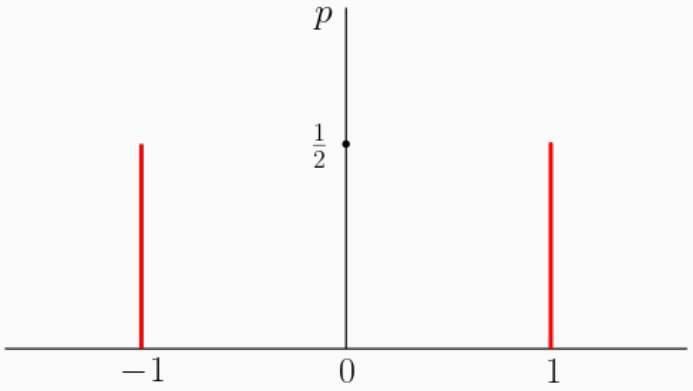


La varianza no mide eso también?

No. Ejemplo: experimento con tres resultados posibles, 0 y ± 1, y dos distribuciones de probabilidad.



$\sigma^2 = 2/3$



$\sigma^2 = 1$

Entropía condicional

Definición

Dos experimentos, A y B, $p(\alpha|\beta)$ probabilidad de que el resultado de A sea α dado que el de B es β.

Definición

Dos experimentos, A y B, $p(\alpha|\beta)$ probabilidad de que el resultado de A sea α dado que el de B es β. **Entropía condicional** de A dado B:

$$S(A|B) \quad = \quad \sum_{\beta} p(\beta) S(p(\ |\beta))$$

Definición

Dos experimentos, A y B, $p(\alpha|\beta)$ probabilidad de que el resultado de A sea α dado que el de B es β. **Entropía condicional** de A dado B:

$$S(A|B) \quad = \quad \sum_{\beta} p(\beta) S(p(\ |\beta)) = -\sum_{\alpha,\beta} p(\beta) p(\alpha|\beta) \log p(\alpha|\beta)$$

Definición

Dos experimentos, A y B, $p(\alpha|\beta)$ probabilidad de que el resultado de A sea α dado que el de B es β. **Entropía condicional** de A dado B:

$$\begin{aligned} S(A|B) &= \sum_{\beta} p(\beta) S(p(\ |\beta)) = -\sum_{\alpha,\beta} p(\beta) p(\alpha|\beta) \log p(\alpha|\beta) \\ &= -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)} \end{aligned}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$S(A|B) - S(A)$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$S(A|B) - S(A) \quad = \quad S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha)$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$\begin{aligned} S(A|B) - S(A) &= S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha) \\ &= S(A|B) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha) \end{aligned}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$\begin{aligned} S(A|B) - S(A) &= S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha) \\ &= S(A|B) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha) \\ &= \sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} \end{aligned}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$\begin{aligned} S(A|B) - S(A) &= S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha) \\ &= S(A|B) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha) \\ &= \sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} \\ &\leq \sum_{\alpha,\beta} p(\alpha,\beta) \left[\frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} - 1 \right] \end{aligned}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$\begin{aligned} S(A|B) - S(A) &= S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha) \\ &= S(A|B) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha) \\ &= \sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} \\ &\leq \sum_{\alpha,\beta} p(\alpha,\beta) \left[\frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} - 1 \right] = 0 \end{aligned}$$

Propiedades

$$S(A|B) = -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

- $S(A|B) \leq S(A)$

$$\begin{aligned} S(A|B) - S(A) &= S(A|B) + \sum_{\alpha} p(\alpha) \log p(\alpha) \\ &= S(A|B) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha) \\ &= \sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} \\ &\leq \sum_{\alpha,\beta} p(\alpha,\beta) \left[\frac{p(\alpha)p(\beta)}{p(\alpha,\beta)} - 1 \right] = 0 \end{aligned}$$

- $S(A|BC) \leq S(A|B)$

Relación con la entropía

$$S(A|B) \quad = \quad -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)}$$

Relación con la entropía

$$
\begin{aligned}
S(A|B) &= -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)} \\
&= -\sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha,\beta) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\beta)
\end{aligned}
$$

Relación con la entropía

$$
\begin{aligned}
S(A|B) &= -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)} \\
&= -\sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha,\beta) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\beta) \\
&= -\sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha,\beta) + \sum_{\beta} p(\beta) \log p(\beta)
\end{aligned}
$$

Relación con la entropía

$$
\begin{aligned}
S(A|B) &= -\sum_{\alpha,\beta} p(\alpha,\beta) \log \frac{p(\alpha,\beta)}{p(\beta)} \\
&= -\sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha,\beta) + \sum_{\alpha,\beta} p(\alpha,\beta) \log p(\beta) \\
&= -\sum_{\alpha,\beta} p(\alpha,\beta) \log p(\alpha,\beta) + \sum_{\beta} p(\beta) \log p(\beta) \\
&= S(AB) - S(B)
\end{aligned}
$$

Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

- $S(A|B) \leq S(A) \;\Leftrightarrow\; S(AB) \leq S(A) + S(B)$

Reescribamos las propiedades

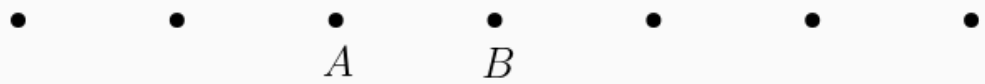
$$S(A|B) = S(AB) - S(B)$$

- $S(A|B) \leq S(A) \;\Leftrightarrow\; S(AB) \leq S(A) + S(B)$ **Subaditividad**

Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

- $S(A|B) \leq S(A) \;\Leftrightarrow\; S(AB) \leq S(A) + S(B)$ **Subaditividad**



Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

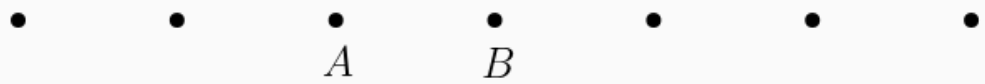
- $S(A|B) \leq S(A) \Leftrightarrow S(AB) \leq S(A) + S(B)$ **Subaditividad**

- $S(A|BC) \leq S(A|B) \Leftrightarrow S(ABC) + S(B) \leq S(AB) + S(BC)$

Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

- $S(A|B) \leq S(A) \Leftrightarrow S(AB) \leq S(A) + S(B)$ **Subaditividad**

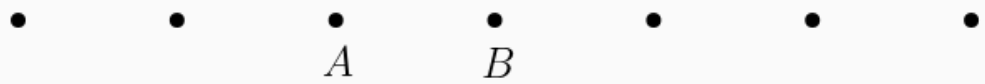


- $S(A|BC) \leq S(A|B) \Leftrightarrow S(ABC) + S(B) \leq S(AB) + S(BC)$
  Subaditividad fuerte

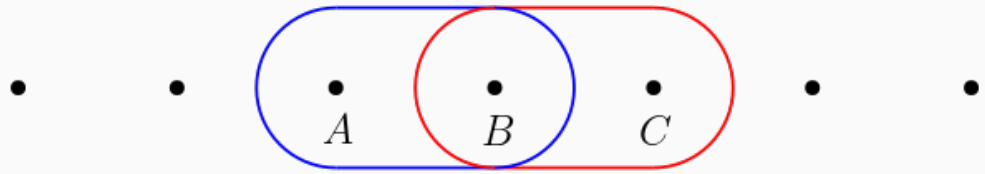
Reescribamos las propiedades

$$S(A|B) = S(AB) - S(B)$$

- $S(A|B) \leq S(A) \Leftrightarrow S(AB) \leq S(A) + S(B)$ **Subaditividad**



- $S(A|BC) \leq S(A|B) \Leftrightarrow S(ABC) + S(B) \leq S(AB) + S(BC)$
  Subaditividad fuerte



Información

Incertidumbre e información

A mayor incertidumbre, mayor es la información que nos va a dar el resultado del experimento $\Rightarrow$ **La entropía mide información**.

Incertidumbre e información

A mayor incertidumbre, mayor es la información que nos va a dar el resultado del experimento $\Rightarrow$ **La entropía mide información**.

Shannon (*A mathematical theory of communication*, 1948) formula esto de manera cuantitativa.

Entropía de una fuente de mensajes

Fuente de mensajes, p_N distribución de probabilidad correspondiente a los mensajes de N caracteres.

Entropía de una fuente de mensajes

Fuente de mensajes, p_N distribución de probabilidad correspondiente a los mensajes de N caracteres. Entropía de esta distribución:

$$S_N = -\sum_{m=1}^{M_N} p_N(m) \log p_N(m)$$

Entropía de una fuente de mensajes

Fuente de mensajes, p_N distribución de probabilidad correspondiente a los mensajes de N caracteres. Entropía de esta distribución:

$$S_N = -\sum_{m=1}^{M_N} p_N(m) \log p_N(m)$$

(si A es el número de caracteres distintos, $M_N = A^N$).

Entropía de una fuente de mensajes

Fuente de mensajes, p_N distribución de probabilidad correspondiente a los mensajes de N caracteres. Entropía de esta distribución:

$$S_N = -\sum_{m=1}^{M_N} p_N(m) \log p_N(m)$$

(si A es el número de caracteres distintos, $M_N = A^N$).

Entropía por caracter de la fuente:

$$s = \lim_{N \to \infty} \frac{S_N}{N}$$

Entropía de una fuente de mensajes

Fuente de mensajes, p_N distribución de probabilidad correspondiente a los mensajes de N caracteres. Entropía de esta distribución:

$$S_N = -\sum_{m=1}^{M_N} p_N(m) \log p_N(m)$$

(si A es el número de caracteres distintos, $M_N = A^N$).

Entropía por caracter de la fuente:

$$s = \lim_{N \to \infty} \frac{S_N}{N}$$

Shannon: s (calculando el log en base 2) es el mínimo número de bits por caracter para codificar los mensajes de la fuente.

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N | x_1, \ldots, x_{N-1}) = p(x_N)$$

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

La entropía por caracter en esta aproximación se denota s_1, y sólo involucra p_1.

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

  La entropía por caracter en esta aproximación se denota s_1, y sólo involucra p_1.

- **Orden 2**: suponer que el proceso de generación de mensajes es de Markov,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N|x_{N-1})$$

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

  La entropía por caracter en esta aproximación se denota s_1, y sólo involucra p_1.

- **Orden 2**: suponer que el proceso de generación de mensajes es de Markov,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N|x_{N-1}) = \frac{p_2(x_{N-1}, x_N)}{p_1(x_{N-1})}$$

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

  La entropía por caracter en esta aproximación se denota s_1, y sólo involucra p_1.

- **Orden 2**: suponer que el proceso de generación de mensajes es de Markov,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N|x_{N-1}) = \frac{p_2(x_{N-1}, x_N)}{p_1(x_{N-1})}$$

  La entropía por caracter en esta aproximación se denota s_2, y sólo involucra p_1 y p_2.

Aproximaciones

Mensaje de N caracteres: $m = (x_1, \ldots, x_N)$

- **Orden 1**: suponer que los caracteres son independientes,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N) = p_1(x_N)$$

  La entropía por caracter en esta aproximación se denota s_1, y sólo involucra p_1.

- **Orden 2**: suponer que el proceso de generación de mensajes es de Markov,

$$p(x_N|x_1, \ldots, x_{N-1}) = p(x_N|x_{N-1}) = \frac{p_2(x_{N-1}, x_N)}{p_1(x_{N-1})}$$

  La entropía por caracter en esta aproximación se denota s_2, y sólo involucra p_1 y p_2.

- Y así sucesivamente.

Aproximaciones

Shannon muestra que $\lim_{n\to\infty} s_n = s$

Aproximaciones

Shannon muestra que $\lim_{n\to\infty} s_n = s$

Además, subaditividad fuerte $\Rightarrow s_{n+1} \leq s_n$

Aproximaciones

Shannon muestra que $\lim_{n\to\infty} s_n = s$

Además, subaditividad fuerte $\Rightarrow s_{n+1} \leq s_n \Rightarrow$ Las s_n dan una cota superior para s, que es mejor cuanto mayor sea n.

Moby Dick

Qué hay que hacer?

Calcular la entropía por caracter de Herman Melville (o mejor dicho sus aproximaciones s_1, s_2 y s_3) a partir de Moby Dick.

Para eso, van a tener que calcular p_1, p_2 y p_3, y eso se hace contando cuántas veces aparece cada caracter, bigrama y trigrama en Moby Dick.

Por ejemplo, en el mensaje 'ah re':

- Caracteres: 'a', 'h', ' ', 'r', 'e'
- Bigramas: 'ah', 'h ', ' r', 're'
- Trigramas: 'ah ', 'h r', ' re'

Algunas referencias

- El paper original de Sannon:

  http://people.math.harvard.edu/ ctm/home/text/others/shannon/entropy/entropy.pdf

- Un paper posterior donde calcula s_1, s_2 y s_3 a partir de las frecuencias tabuladas de los n-gramas en inglés:

  https://www.princeton.edu/ wbialek/rome/refs/shannon_51.pdf

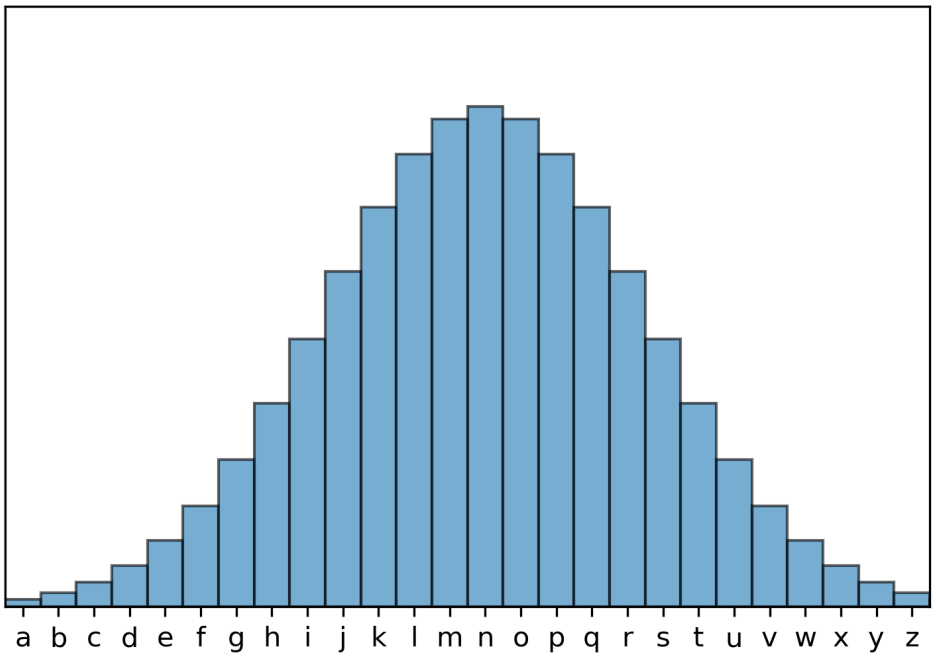
- Una página donde van a encontrar las frecuencias de los n-gramas en varios idiomas, para que puedan chequear sus resultados:

  http://practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/

  En esta página no se contabilizan los espacios, así que para poder comparar deberán eliminar los espacios de sus archivos.

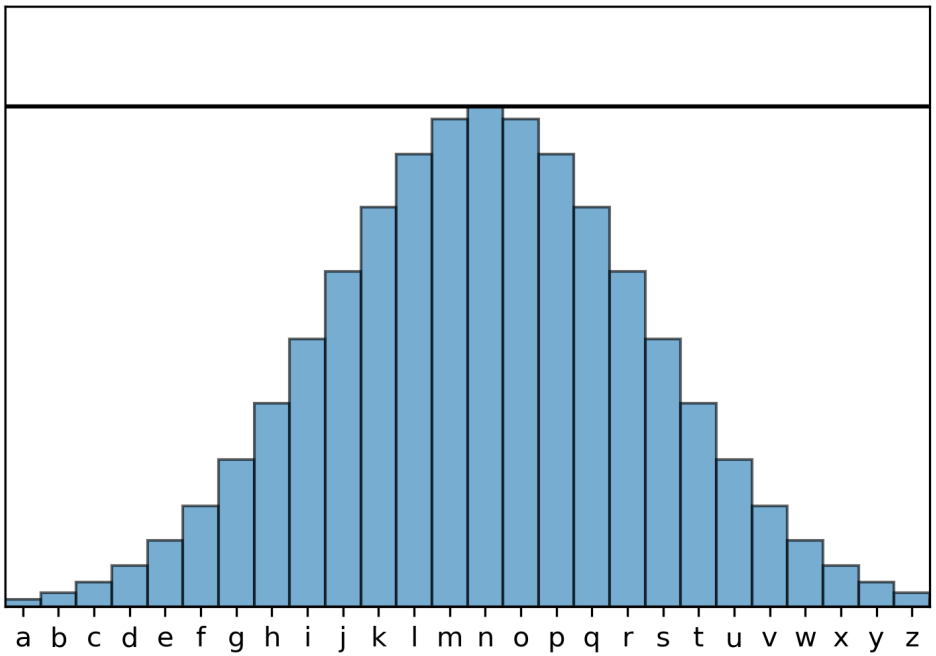
Apéndice

Para los que quieran hacer el ítem *(f)*. Cómo generar letras distribuidas con la distribución que queramos?



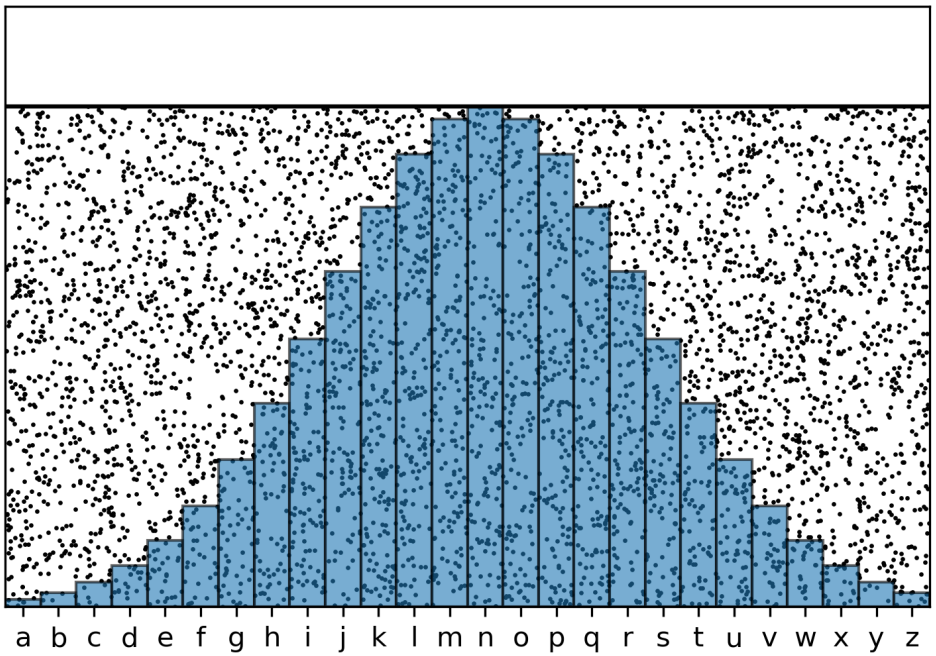
Apéndice

Para los que quieran hacer el ítem *(f)*. Cómo generar letras distribuidas con la distribución que queramos?



Apéndice

Para los que quieran hacer el ítem *(f)*. Cómo generar letras distribuidas con la distribución que queramos?



Apéndice

Para los que quieran hacer el ítem *(f)*. Cómo generar letras distribuidas con la distribución que queramos?