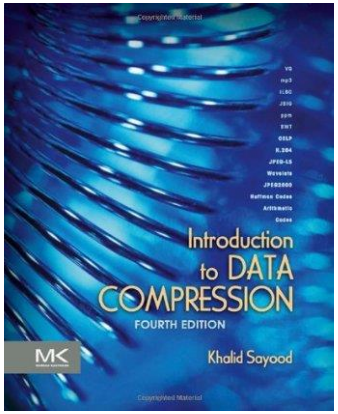
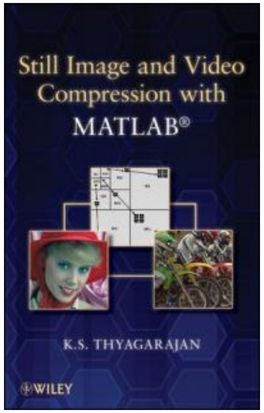
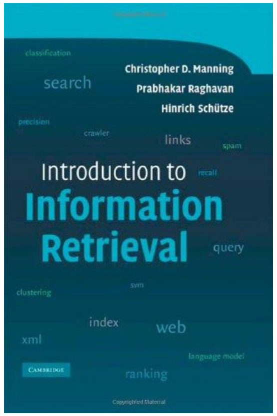






Conceptos para la compresión con pérdida

Rafael Molina

Depto Ciencias de la Computación

e Inteligencia Artificial

Universidad de Granada

Contenidos

I. Objetivos del tema

En la compresión sin pérdida la preocupación estaba en el factor de compresión o equivalentemente la tasa (rate) o número de bits por muestra, ahora en la compresión con pérdida hemos de mirar también a la distorsión que la pérdida de información produce.

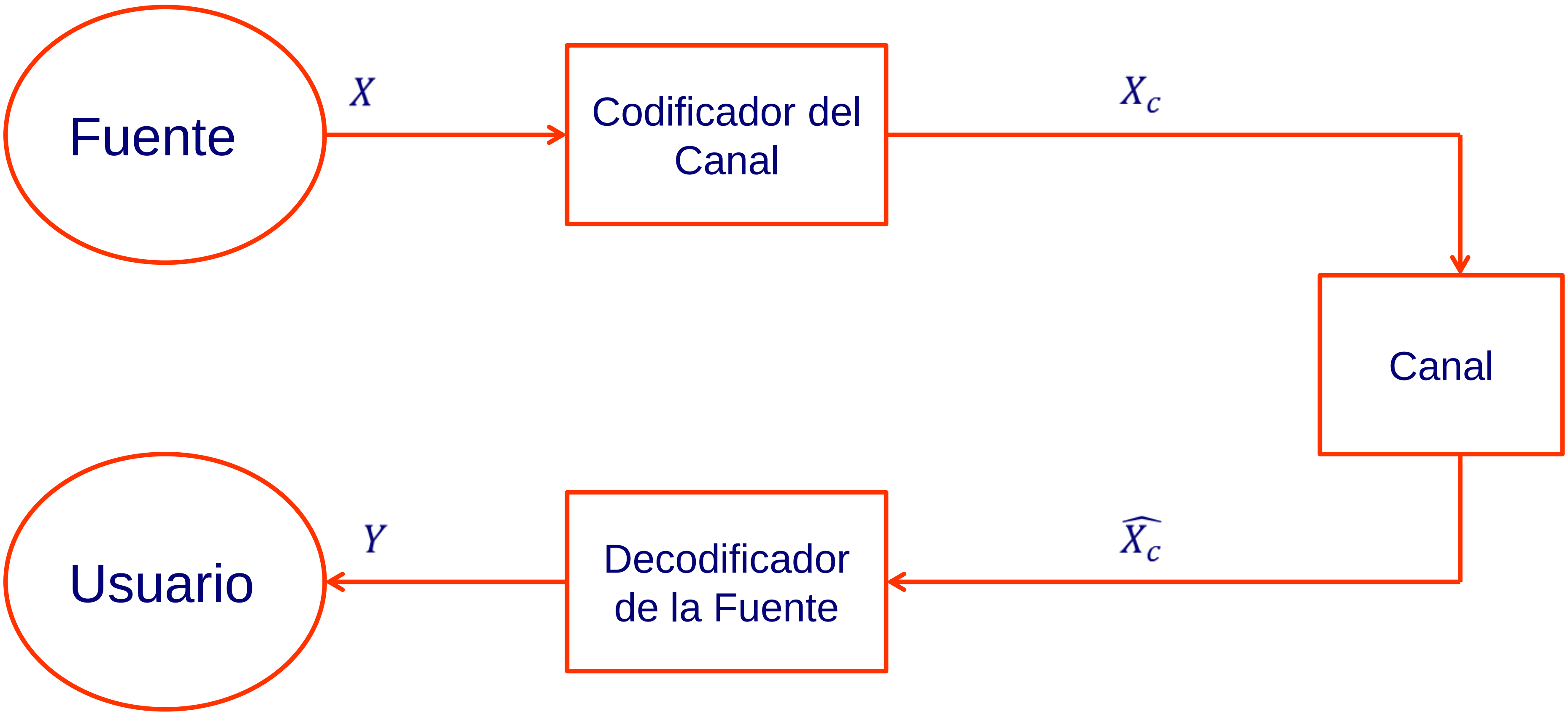
La teoría matemática subyacente al compromiso entre reducir la tasa y el aumento de la distorsión que esta reducción produce es Teoría tasa/distorsión, en inglés, (Rate/Distortion Theory) que estudiamos en este tema.

II. Introducción.

En la codificación sin pérdida nunca tuvimos que preocuparnos por la diferencia entre la secuencia original y la reconstruida una vez comprimida dicha secuencia. Además, la entropía nos marcaba el límite inferior de nuestra tasa.

En la codificación con pérdida queremos disminuir lo más posible la tasa pero esto nos va a introducir una distorsión.

Los dos casos extremos son: enviamos cero información (tasa=0) entonces la distorsión en muy grande o hacemos que la distorsión sea nula en cuyo caso la tasa es grande. El estudio de qué pasa entre estos dos extremos recibe el nombre de Teoría Tasa/Distorsión.



$\widehat{X_c}=X_c$ Canal sin ruido (supondremos que se cumple)

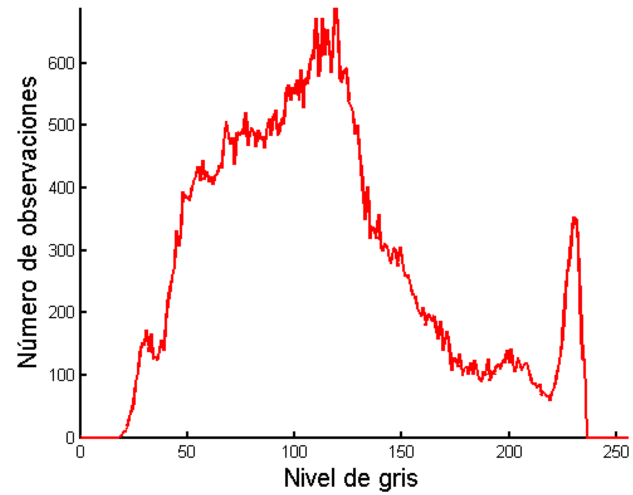
Si X_c reproduce fielmente X compresión sin pérdida, en caso contrario compresión con pérdida.

Ejemplo II.1

Consideremos la imagen de la derecha sin compresión, almacenada en 8 bits por píxel.

Sabemos que si consideramos los valores de los píxeles como realizaciones i.i.d., la entropía nos proporciona la menor tasa que podemos alcanzar sin distorsión.

H(p)= 7.4716 bits



¿Y si tenemos menos bits o estamos dispuestos a soportar algo de distorsión?

```
Img=imread('goldhill.pgm');
[histograma,bin]= hist(Img(:),[0:255]);
prob=histograma/sum(histograma(:));
positivos=prob(histograma>0.0);
H=-sum(positivos.*log2(positivos))
```

```
imshow(Img); figure;
plot(bin,histograma,'-r','LineWidth',2),
axis('tight');
xlabel('Nivel de gris','FontSize',15);
ylabel('Número de
observaciones','FontSize',15);
set(gca,'LineWidth',2,'box','off');
```

III Criterios de distorsión

¿Cómo medimos la fidelidad de la fuente reconstruida a la original?.

Una forma natural de medir la fidelidad es examinar la diferencia entre el original y su reconstrucción. Sin embargo, existe un amplio debate sobre la utilización de otros criterios que después comentaremos.

Los criterios basados en diferencias utilizan normalmente la diferencia en valor absoluto o bien la diferencia al cuadrado.

Sean X la señal original e Y la reconstrucción:

Preliminares

$$x_n = \text{señal original}$$

$$y_n = \text{señal reconstruida}$$

$$1 \leq n \leq N$$

Varianza de la señal

$$\sigma_x^2 = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - Media\,(X\,)\right)^2$$

Error cuadrático medio

$$\sigma_d^2 = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - y_n\right)^2$$

Relación señal ruido

$$SNR_{RMS} = \frac{\sum_{n=1}^{N}\left(x_n - Media\,(X\,)\right)^2}{\sum_{n=1}^{N}\left(x_n - y_n\right)^2} = \frac{\sigma_x^2}{\sigma_d^2}$$

Relación señal ruido en decibelios

$$SNR(dB) = 10\log_{10}\frac{\sigma_x^2}{\sigma_d^2}$$

Pico de la relación señal ruido

$$PSNR(dB) = 10\log_{10}\frac{x_{pico}^2}{\sigma_d^2}$$

$$x_{pico} = \max(X\,)$$

Junto con la utilización de criterios matemáticos, podemos también usar como criterio la pregunta a un grupo de expertos (este sistema estará obviamente bastante sesgado) o intentar medir las discrepancias utilizando conocimiento sobre el sistema de percepción humano (OJO y OÍDO).

Ejemplo III.1

Volvamos a nuestro ejemplo.

¿Cuál es el mejor codificador que puedo utilizar, si lo mejor significa: el menor error cuadrático medio entre la señal original y la reconstruida, y sólo dispongo de una tasa de bits por píxel pequeña?

Vamos a comprimir la imagen usando 2^i valores distintos i=8,..,0. Vamos a calcular la distorsión entre la imagen original y la comprimida y la entropía de la imagen comprimida

```
close all; clear all;
Img=imread('goldhill.pgm');
distorsion=zeros(1,9);
rate=zeros(1,9);
H=zeros(1,9);
[M,N]=size(Img),
nele=M*N;
```

Volveremos sobre él

```
for i=0:8
        k=2^i;
        Imgq=uint8(floor(double(Img)/k));
        Imgreconst=uint8(k*(double(Imgq)+0.5));
        Error=double(Img)-double(Imgreconst);
        Error2=Error.*Error;
        distorsion(i+1)=sum(Error2(:))/nele;
        rate(i+1)=8-i;
        [histograma,bin]= hist(Imgq(:),[0:255]);
        prob=histograma/sum(histograma(:));
        positivos=prob(histograma>0.0);
        H(i+1)=-sum(positivos.*log2(positivos));
        figure;
        imshow (Imgreconst);
        title(sprintf('numero de niveles %d',256/k))
end
figure;
plot(distorsion,rate,'*');  hold on;
plot(distorsion,H,'+'); legend('Uniforme','Entropia')
xlabel('Distorsion','FontSize',15);
ylabel('Rate','FontSize',15);
set(gca,'LineWidth',2,'box','off');
```

numero de niveles 256

numero de niveles 32

numero de niveles 4

numero de niveles 128

numero de niveles 16

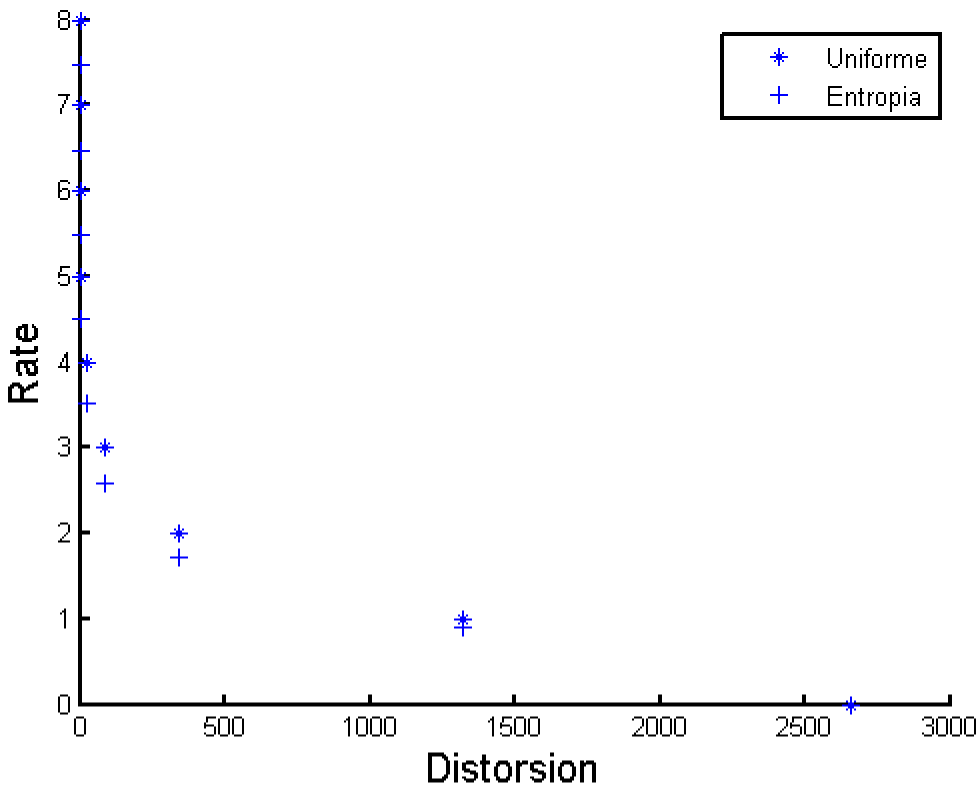
numero de niveles 2

numero de niveles 64

numero de niveles 8

numero de niveles 1

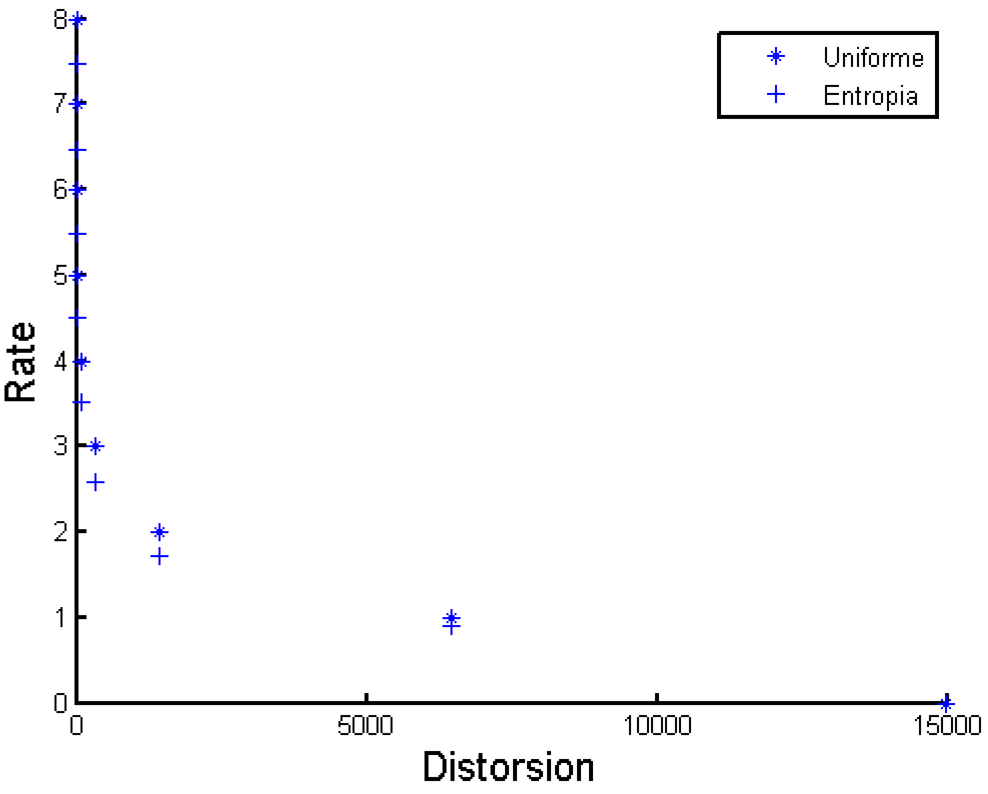
Distorsion	No. Bits (uniforme)	Entropía fuente cuantizada
1.0	8	7.47
.5	7	6.47
1.5	6	5.47
5.5	5	4.49
21.3	4	3.51
86.4	3	2.59
343.7	2	1.71
1319.0	1	0.89
2657.0	0	0.0



Observa, al ejecutar el programa de Matlab que la distorsión con 8 bits es 1 por haber introducido el 0.5 en Imgreconst

Si suprimimos el 0.5 obtenemos la siguiente tabla y gráfica

Distorsion	No. Bits (uniforme)	Entropía fuente cuantizada
0.0	8	7.47
1.0	7	6.47
4.0	6	5.47
18.0	5	4.49
76.0	4	3.51
320.0	3	2.59
1421.0	2	1.71
6471.0	1	0.89
15000.0	0	0.0

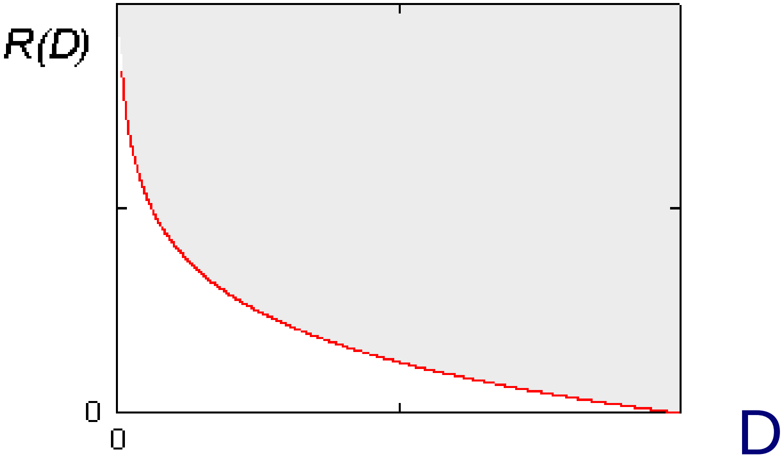


Dados los ejemplos anteriores cabe preguntarse:

- Dada una distorsión, D, ¿cual es el menor número de bits (la menor tasa) que tenemos que utilizar para que la distorsión sea D?

- Dado un presupuesto (tasa) en bits, B, ¿Cuál es la menor distorsión que podemos obtener con un método de compresión?

Gráfico R/D típico



Estamos buscando un equivalente a la cota que nos proporciona la entropía cuando no hay distorsión para el caso en el que permitimos distorsión.

Tenemos que revisitar, por tanto, la teoría de la información.

IV.Teoría de la información revisitada

- Dada una medida de distorsión, para estudiar el compromiso entre tasa y distorsión en compresión con pérdida, nos gustaría definir explícitamente la tasa en función de la distorsión.

- Como veremos esto no es siempre posible (casi nunca) y tendremos que solucionar este problema. Antes de ello necesitamos algunos conceptos más sobre teoría de la información.

Ejemplo 4.1

Un sistema de compresión con pérdida simple consiste en suprimir un cierto número de los bits menos significativos de la entrada.

Por ejemplo si tenemos imágenes monocromo de 8 bits, y sólo podemos ver 64 niveles de gris distintos, podemos suprimir los dos bits menos significativos. Con un ejemplo:

Nuestro alfabeto es {0,1,2,…,15}. El codificador codifica la fuente eliminando el bit menos significativo. La salida es entonces {0,1,2,…,7}. Nuestra estimación de la señal original se obtendrá añadiendo como bit menos significativo un cero. En otras palabras la reconstrucción será {0,2,4,…,14}.

IV.1 Entropía Condicionada

Sea X una variable aleatoria que toma los valores del alfabeto *X*={x_0,x_1,…,x_{n-1}} e Y la variable aleatoria de reconstrucción que toma valores en el alfabeto *Y*={y_0,y_1,…,y_{M-1}}. Sabemos que la entropía de la fuente y de la reconstrucción valen respectivamente

$$H(X) = -\sum_{i=0}^{N-1} P(x_i) \log P(x_i)$$

$$H(Y) = -\sum_{j=0}^{M-1} P(y_j) \log P(y_j)$$

Supongamos que ahora fijamos un valor observado de Y, que lo notamos y_j, este valor produce la distribución condicionada $P(x_0|y_j)$, $P(x_1|y_j)$,….,$P(x_{N-1},y_j)$. Su entropía es

$$H(X \mid Y = y_j) = -\sum_{i=0}^{N-1} P(x_i \mid y_j) \log P(x_i \mid y_j)$$

La media, con la distribución de Y, de estos valores recibe el nombre de entropía condicionada de X dado Y, H(X|Y), es decir

$$H(X \mid Y) = \sum_{j=0}^{M-1} P(y_j) H(X \mid Y = y_j)$$

$$= -\sum_{j=0}^{M-1} P(y_j) \sum_{i=0}^{N-1} P(x_i \mid y_j) \log P(x_i \mid y_j)$$

$$= -\sum_{j=0}^{M-1} \sum_{i=0}^{N-1} P(y_j) P(x_i \mid y_j) \log P(x_i \mid y_j)$$

De igual forma podemos definir la entropía condicionada de Y a X, H(Y|X),

$$H(Y|X) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i) P(y_j | x_i) \log P(y_j | x_i)$$

La entropía condicional H(X|Y) puede interpretarse como la cantidad de incertidumbre que queda en la variable X, dado que conocemos el valor que Y tomó.

Obviamente el conocimiento de Y debería reducir la incertidumbre sobre X. Puede probarse que

$$H(X|Y) \leq H(X)$$

Ejemplo 4.2

Supongamos que tenemos una fuente con alfabeto X={0,1,…,15} que son igualmente probables y que los codificamos haciendo cero el bit menos significativo. Tenemos

$$H(X) = \sum_{i=0}^{15} \frac{1}{16} \log 16 = 4\text{bits}$$

La variable codificada la notamos Y y tenemos Y={0,2,4,6,8,10,12,14}

$$P(Y = j) = P(X = j) + P(X = j+1) = \frac{1}{8}$$

y

$$H(Y) = 3\,\text{bits}$$

Además las probabilidades condicionadas valen

$$P(X = i \mid Y = j) = \begin{cases} 1/2 & \text{Si i = j ó i = j+1, j = 0,2,..,14} \\ 0 & \text{en el resto} \end{cases}$$

De donde

$$H(X \mid Y) = 1 \quad \text{bit}$$

Intuitivamente, de la observación de Y conocemos los tres bits más significativos de X y sólo nos queda por saber el restante.

IV.2 Información Mutua Media

Dado y_j podemos construir la distribución de probabilidad $P(x_k|y_j)$ y podemos medir la discrepancia entre $P(x_k|y_j)$ y $P(x_k)$. Esto nos lleva a considerar la llamada información mutua que viene definida por

$$i(x_k; y_j) = \log\left(\frac{P(x_k \mid y_j)}{P(x_k)}\right)$$

comprueba que

$$\frac{P(x_k \mid y_j)}{P(x_k)} = \frac{P(y_j \mid x_k)}{P(y_j)}$$

¿ Qué ocurre si x e y son independientes?

Usando la información mutua, podemos definir la información mutua media mediante

$$I(X;Y)=\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}P(y_j)P(x_k\mid y_j)\log\left(\frac{P(x_k\mid y_j)}{P(x_k)}\right)$$

Utilizando la entropía y la entropía condicionada, tendremos

$$I(X;Y)=\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}P(y_j)P(x_k\mid y_j)\log\left(\frac{P(x_k\mid y_j)}{P(x_k)}\right)$$

$$=\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}P(y_j)P(x_k\mid y_j)\log\left(P(x_k\mid y_j)\right)$$

$$-\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}P(y_j)P(x_k\mid y_j)\log\left(P(x_k)\right)=H(X)-H(X\mid Y)$$

Además como
$$\frac{P(x_k \mid y_j)}{P(x_k)} = \frac{P(y_j \mid x_k)}{P(y_j)}$$

se cumple

$$I(X;Y) = \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} P(y_j)P(x_k \mid y_j)\log\left(\frac{P(y_j \mid x_k)}{P(y_j)}\right)$$

$$= H(Y) - H(Y \mid X) = I(Y;X)$$

Ejemplo 7.4.3.

Para el ejemplo 7.4.2 tenemos H(X)=4 bits y H(X|Y)=1 bit.
Por tanto I(X;Y)=I(Y;X)=3 bits

Todos los conceptos estudiados en esta sección se extienden a variables continuas pero no los vamos a estudiar aquí.

V. Teoría Razón/Distorsión

La teoría de la razón-distorsión estudia el compromiso entre distorsión y la tasa en esquemas de compresión con pérdida.

La función R(D), que recibe el nombre de función de tasa-distorsión, especifica la tasa más pequeña a la que la salida de una fuente puede ser codificada mientras se mantiene la distorsión menor o igual a D.

Vamos a definir estos conceptos más formalmente. Tenemos que empezar definiendo lo que es la distorsión.

Supongamos que tenemos una fuente X con una distribución de probabilidad P(X).

Si X ha tomado el valor x_i lo podemos codificar, tal vez con uno sólo, pero en general con un conjunto de valores $\{y_j\}$.

Si el valor de X es x_i y el de Y es y_j entonces habrá un error entre ellos que lo mediremos mediante $d(x_i,y_j)$.

¿Cómo se define entonces la distorsión?

$$D = \sum_i P(x_i)\left(\sum_j P(y_j \mid x_i) d(x_i, y_j)\right)$$

Observa

- si tenemos una fuente X con una distribución de probabilidad P(X) y la medida de error $d(x_i,y_j)$
- teniendo en cuenta la definición de distorsión lo que hará que ésta sea mayor o menor es la probabilidad condicionada que utilicemos.

Consideremos el ejemplo 6.4.2. y como medida de distorsión $d(x_i,y_j) =(x_i-y_j)^2$. Entonces la distorsión vale

$$D = \sum_{i=0,2,\dots,14} \frac{1}{16} \times 0 + \sum_{i=1,3,\dots,15} \frac{1}{16} \times 1 = \frac{1}{2}$$

Sea ahora D* una distorsión (un valor de la distorsión) y consideremos el conjunto

$$\Gamma(D^*) = \left\{ P(Y \mid X) \quad \middle| \quad \sum_i P(x_i) \left(\sum_j P(y_j \mid x_i) d(x_i, y_j) \right) \leq D^* \right\}$$

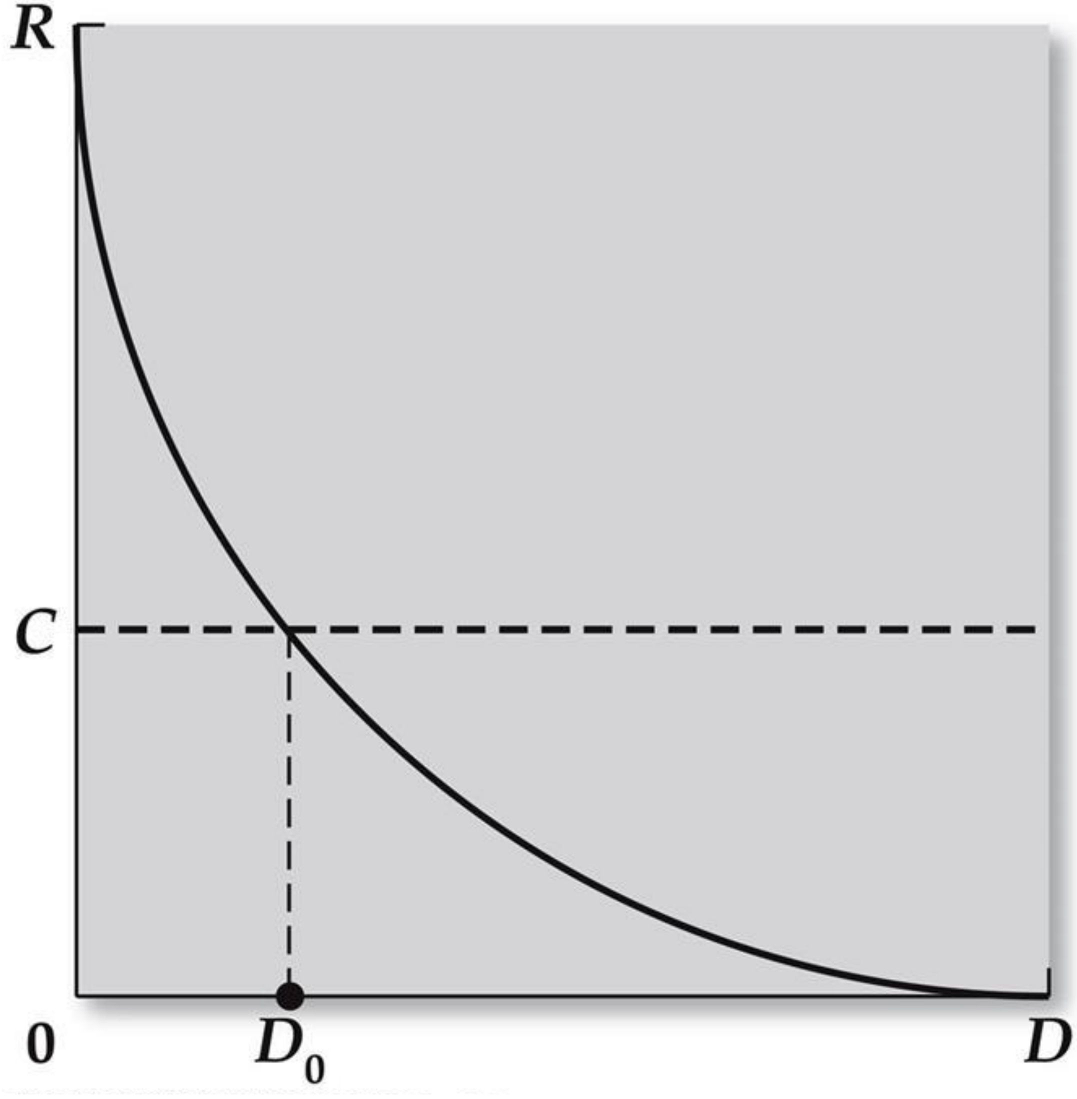
Shannon probó que la mínima tasa (número de bits) necesarios para codificar una fuente con una distorsión D* viene dado por

$$R(D^*) = \min_{\Gamma(D^*)} I(X;Y)$$

y además demostró que si describimos X con una tasa menor que R(D) nuestra distorsión nunca será menor que D.

Veamos un gráfico R/D típico que ya conoces

Gráfico R/D típico



VI. Modelos para las fuentes

Distribución Uniforme: Si X~U[a,b] su función de densidad viene dada por

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & si \quad a \le x \le b \\ 0 & en\ caso\ contrario \end{cases}$$

que cumple $E(X) = (b+a)/2, \quad \mathrm{var}(X) = (b-a)^2/12$

Si tenemos sólo dos valores para representar una realización de esta distribución ¿cómo lo harías?.

Llamemos Y a la representación. Sería razonable medir la distancia entre la realización X y su representación Y ¿cómo?

mediante $(x-y)^2$

¿Y si tenemos 4 valores?. ¿Y si las realizaciones fuesen de las siguientes distribuciones?

Distribución Normal o de Gauss: Si X~N(μ,σ) su función de densidad viene dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2 \right]$$

que cumple

$$E(X) = \mu, \quad \mathrm{var}(X) = \sigma^2$$

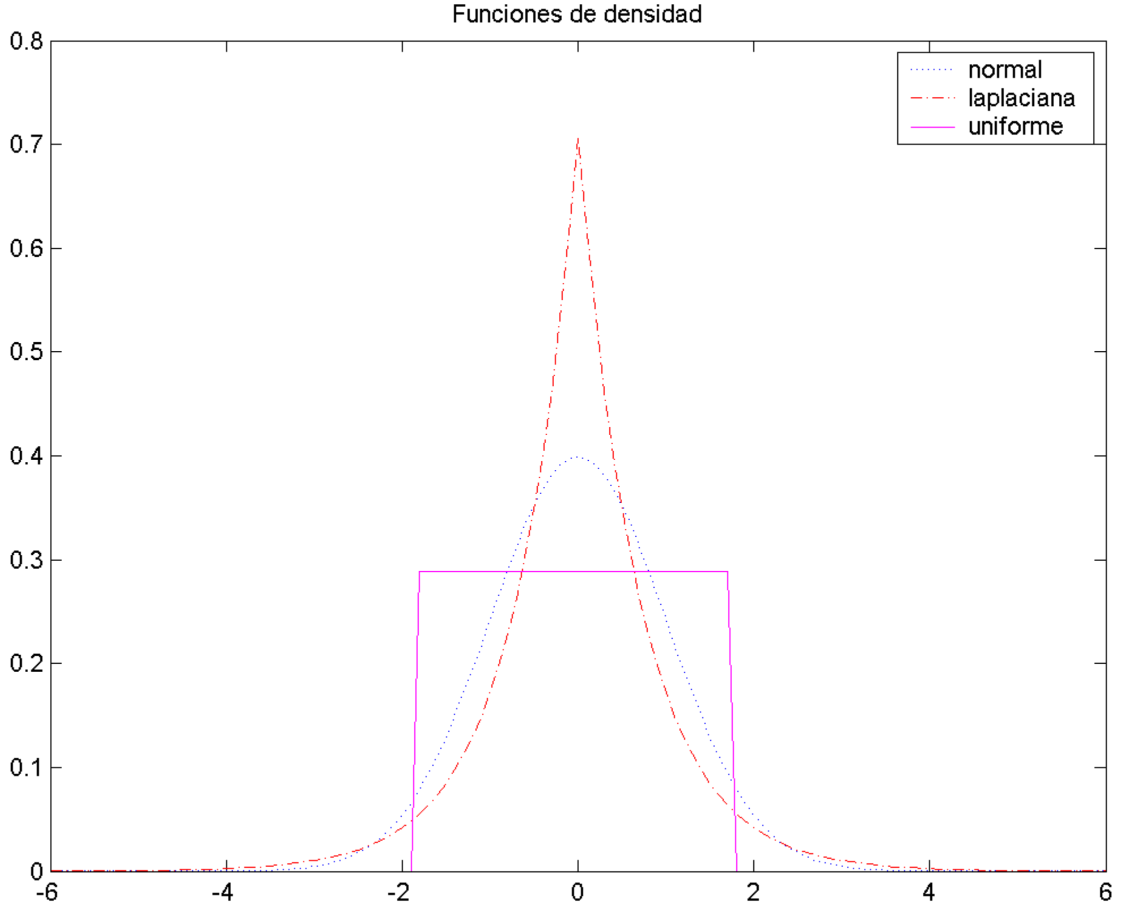
Distribución Laplaciana: Son distribuciones más picudas. Su densidad tiene la forma

$$f_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left[-\frac{\sqrt{2}}{\sigma}|x| \right]$$

que cumple

$$E(X) = 0, \quad \mathrm{var}(X) = \sigma^2$$

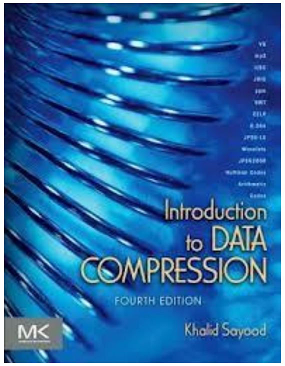
Gráfico de las funciones de densidad correspondientes a variables aleatorias con distribuciones uniforme, normal y de Laplace de media cero y varianza uno.



Además de los modelos descritos que suponen que las fuentes son independientes e idénticamente distribuidas, existen modelos que relacionan variables con las observadas con anterioridad.

Son los modelos autorregresivos y medias móviles y la mezcla de ambos autorregresivos-medias móviles que no estudiaremos aquí

V Bibliografía



K. Sayood, “Introduction to Data Compression”, Morgan and Kaufmann, 2012.