

1 Introducción

El objetivo del pte. estudio es medir la Cantidad de Información que contiene (lleva) un determinado suceso o mensaje.

Hasta ahora la información tenía un valor cualitativo como ser precisa, estar en tiempo y forma disponible para la toma de decisiones, etc. Ahora bien si la información debe ser transmitida se debe cuantificar a fin de poder dimensionar el medio por el cual va a ser transmitida

Entonces para cuantificar la cantidad de información que aporta un suceso o mensaje debemos pensar qué función la puede representar, en principio vemos que un suceso cualquiera nos aporta más información cuanto menos certidumbre tengamos de dicho suceso, es decir si sabemos que un suceso va a ocurrir no nos trae mucha información en cambio si ocurre algo que no se espera la cantidad de información que aporta dicho suceso es mayor, recordemos que la certidumbre o incertidumbre de ocurrencia de un suceso no es más que la probabilidad de ocurrencia de dicho suceso, entonces la **Cantidad de Información $I_{(A)}$** que aporta un suceso **A** es función inversa de la probabilidad de ocurrencia de ese suceso es decir de la **$p_{(A)}$**

$$I_{(A)} = \text{función } 1/ p_{(A)}$$

Esta función de la probabilidad debe cumplir las siguientes tres condiciones:

-Para un suceso del cual tenemos total certidumbre de ocurrencia (**$p_{(A)}=1$**) la cantidad de información **$I_{(A)}$** que trae es cero, es decir

Dado el suceso **A** con **$p_{(A)} = 1$** \implies la Cant. de Inf. es cero $\rightarrow I_{(A)} = 0$

-Para un suceso del cual tenemos la certidumbre de ocurrencia es muy baja, la cantidad de información que trae es muy alta, es decir

Dado el suceso **A** con **$p_{(A)} \rightarrow 0$** \implies la Cant. de Inf. es muy elevada **$I_{(A)} \rightarrow \infty$**

-Para dos sucesos independientes **A** y **B** cada uno con su respectiva probabilidad de ocurrencia **$p_{(A)}$ y $p_{(B)}$** , donde **A** trae cantidad de información **$I_{(A)}$** y **B** trae cantidad de información **$I_{(B)}$**

Resulta que la probabilidad del suceso conjunto es **$p_{(A.B)} = p_{(A)} * p_{(B)}$**

Así mismo la Información del suceso conjunto es la suma de las Informaciones $I_{(A,B)} = I_{(A)} + I_{(B)}$

Por lo tanto, la única función que en 1 es cero, en cero tiende a infinito y transforma un producto en una suma es la función logaritmo, y en particular como trabajamos con mensajes binarios usaremos el **logaritmo en base 2** cuya unidad de medida será el “bit” reservando por ahora el término “Binit” para el dígito binario.

Entonces dado un suceso **A** con probabilidad de ocurrencia $p_{(A)}$, la Cantidad de Información $I_{(A)}$ que aporta ese suceso es :

$$I_{(A)} = \lg_2 1/p_{(A)} = (\text{Bit}) \quad \text{ó también } I_{(A)} = -\lg_2 p_{(A)}$$

Observación:

Recordando que la función logaritmo (decimal) de un número es el exponente al que hay que elevar la base (10) para obtener el número:

$$\log N = X \Rightarrow 10^X = N$$

De igual manera para el logaritmo en base 2 que nos ocupa:

$$\lg_2 N = X \Rightarrow 2^X = N$$

Aplicando logaritmo en base 10 a esta última expresión

$$\log 2^X = \log N \Rightarrow X \cdot \log 2 = \log N \Rightarrow X = \log N / \log 2$$

o sea $\lg_2 N = \log N / \log 2 \simeq 3.32 \log N$

Mismo análisis se puede hacer partiendo del logaritmo Neperiano (en base e), es decir

$$\lg_2 N = \ln N / \ln 2$$

1.2 Entropía, Fuente y Canal

En la formación de la teoría, encontramos una magnitud que cumple con las especificaciones que se le imponen a la medida de la información y que por tener una expresión formal mente idéntica a la que posee la magnitud termodinámica conocida como “entropía”, se la conoce con el mismo nombre.

1.2.1 Modelos aleatorios univariados

El modelo aleatorio univariado es un modelo probabilístico del comportamiento de una fuente de información o de un receptor de la misma.

Nuestro universo de sucesos elementales estará constituido por símbolos de la fuente, x_i (binario, etc.).



De acuerdo a las características de los símbolos se le asigna a cada uno su probabilidad de ocurrencia.

$$x_i \longrightarrow p(x_i)$$

Tenemos así una caracterización univariada, esto es unidimensional, de la fuente.

Su representación es:

$$\begin{bmatrix} x \\ p(x) \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ p(X_1) & p(X_2) & \dots & p(X_n) \end{bmatrix}$$

Función que provea una magnitud que represente la cantidad de información asociada al suceso que representa la aparición de un símbolo de la fuente. Se propone:

$$I(X_i) = -\lg p(X_i) \quad (1)$$

donde:

$I(x_i)$ es la cantidad de información asociada al suceso x_i

$p(x_i)$ es la probabilidad de ocurrencia del suceso x_i .

OBSERVACIÓN: Centraremos el estudio en fuentes ergódicas (probabilidad de los símbolos independiente del tiempo) y fuentes de memoria nula

INTRODUCCIÓN:

Supongamos que tenemos una fuente **S** que emite tres símbolos **R**, **A**, y **V** y que una **muestra significativa** es la siguiente:



Cada símbolo al aparecer nos dará una cantidad de información dada $I_{(A)}$, $I_{(V)}$ e $I_{(R)}$

Si pretendemos tener la cantidad de Información promedio será:

$$\text{Cant.Inf. Prom} = \frac{N_{(A)} \cdot I_{(A)} + N_{(V)} \cdot I_{(V)} + N_{(R)} \cdot I_{(R)}}{N}$$

N es el número total de la muestra

Ahora bien cada cociente $N_{(i)}/N = fr_i = P_{(i)}$

fr_i es la frec. relativa del símbolo i

$P_{(i)}$ es la probabilidad de ocurrencia del símbolo i y por ser una muestra significativa se puede igualar a la frecuencia relativa

Por lo tanto, la Cantidad de Información promedio que emite una fuente será:

$$H(S) = \sum p_i \cdot \lg 1/p_i = - \sum p_i \cdot \lg p_i$$

ENTROPÍA DE LA FUENTE

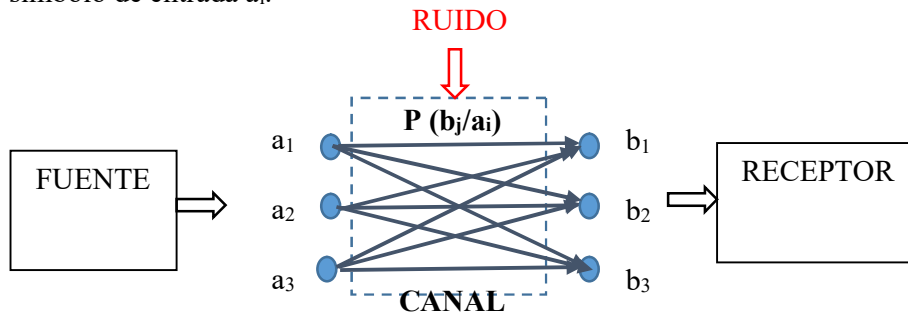
Para una caracterización univariada denominamos “entropía” – y la notamos $H(x)$ – a la esperanza matemática de la información por símbolo de la fuente, esto es:

$$H(x) = - \sum P(X_i) \lg(p(X_i)) \quad \text{bit/smbolo} \quad (2)$$

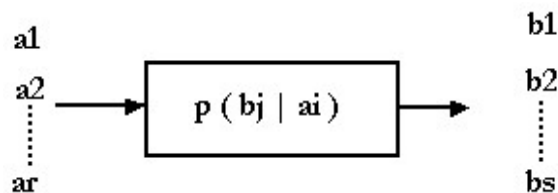
La función entropía tiene su máximo valor, para un determinado número de símbolos fuente, cuando estos son equiprobables.

1.2.2- Modelos aleatorios bivariados

Se puede definir un canal de información como determinado por un alfabeto de salida $B = (b_j)$, $j = 1, 2, 3, \dots, s$; y un conjunto de probabilidades condicionales $p(b_j/a_i)$ siendo esta la probabilidad de recibir a la salida el símbolo b_j cuando se envía el símbolo de entrada a_i .



Puede Representarse esquemáticamente:



Un canal sin memoria queda descrito por medio de la matriz de probabilidades condicionales siguiente:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1s} \\ P_{21} & P_{22} & \dots & P_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r1} & P_{r2} & \dots & P_{rs} \end{bmatrix}$$

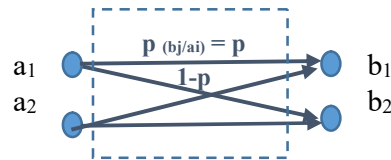
donde $P_{ij} = P(b_j/a_i)$

Cada fila de la matriz corresponde a una entrada del canal y cada columna a una salida.

A un canal se lo denomina transparente cuando no introduce error, esto es si, su matriz de probabilidades condicionales es diagonal.

Un canal de importancia teórica es el binario simétrico (BSC). Este canal posee dos símbolos de entrada ($a_1 = 0$, $a_2 = 1$) y dos de salida ($b_1 = 0$, $b_2 = 1$). Es simétrico por

ser iguales las probabilidades de recibir un cero al enviar un uno y viceversa. Esta probabilidad se denomina probabilidad de error (p).



Dado un canal definido por los símbolos de entrada y salida y una matriz de probabilidades condicionales P . Considerando la probabilidades de los símbolos de entrada $p(a_i)$, $i = 1, 2, \dots, r$ y la matriz de probabilidades condicionales del canal (p), se puede determinar otro conjunto de probabilidades $p(b_j)$, $j = 1, 2, \dots, s$ que representa la probabilidad de aparición del símbolo b_j a la salida del canal.

$$P(b_j) = \sum_{i=1}^r P(a_i) * P_{ij} \quad (3)$$

Cabe destacar que conociendo las probabilidades de salida y la matriz de

$$P(a_i/b_j) = \frac{P_{ij} * P(a_i)}{\sum_{i=1}^r P(a_i) * P_{ij}} \quad (4)$$

probabilidades condicionales, en general no pueden calcularse las probabilidades de entrada.

Se verán las llamadas probabilidades hacia atrás, esto es, $p(a_i/b_j)$ que significa la probabilidad de que habiéndose dado b_j a la salida, provenga de un a_i a la entrada del canal. Según Bayes:

El numerador de la expresión (4), es la probabilidad del suceso (a_i, b_j) , es decir la probabilidad conjunta de la aparición de a_i a la entrada y b_j a la salida.

Entonces:

$$P(a_i, b_j) = P(a_i) * P(b_j/a_i)$$

$$P(a_i) * P(b_j/a_i) = P(b_j) * P(a_i/b_j)$$

Por lo tanto se puede calcular la incertidumbre de un símbolo de entrada dado un símbolo a la salida

$$P(a_i/b_j) = P(a_i) * P(b_j/a_i) / P(b_j)$$

Se puede luego, obtener una matriz de probabilidades conjunta $P(A,B)$.

La caracterización bivariada del sistema, se da:

$$(A, B) = [(a_i, b_j)] \quad P(A, B) = [P(a_i, b_j)]$$

La incertidumbre de aparición en el sistema, la información media al emitir a_i y recepcionar b_j – Se denomina entropía del sistema:

$$H(A, B) = - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) * \lg P(a_i, b_j) \quad (6)$$

Se denomina $P(a_i)$ a la probabilidad a priori de los símbolos de entrada. Y a $P(a_i/b_j)$ se la denomina probabilidad a posteriori, por ser esta la probabilidad del símbolo, luego de sucedido b_j . Es posible calcular la entropía del conjunto de símbolos de entrada según ambas probabilidades.

Así la entropía a priori será:

$$H(A) = - \sum_{i=1}^r P(a_i) * \lg P(a_i) \quad (7)$$

$$H(A) = \sum_i p(a_i) * \lg 1/p(a_i)$$

y la entropía a posteriori:

$$H(A/B) = \sum P(b_j) * P(a_i/b_j) * \lg P(a_i/b_j)$$

$$\begin{aligned} H(A/B) &= \sum_{j=1}^s P(b_j) * H(A/b_j) \\ &= \sum_{j=1}^s P(b_j) * - \sum_{i=1}^r P(a_i/b_j) * \lg P(a_i/b_j) \end{aligned} \quad (8)$$

$$\begin{aligned} &= - \sum_{j=1}^s P(b_j) * \sum_{i=1}^r P(a_i/b_j) * \lg P(a_i/b_j) \\ H(A/B) &= - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) * \lg P(a_i/b_j) \end{aligned} \quad (9)$$

$$H(A/B) = - \sum_{i,j} P(b_j) * P(a_i/b_j) * \lg P(a_i/b_j)$$

Como los símbolos a la salida presentan una probabilidad $P(b_j)$ es posible calcular la entropía media a posteriori como

esto no es más que la entropía de la fuente conociendo la recepción (equivocación de A respecto de B).

Como se conoce la $P(b_j/a_i)$ que caracteriza al canal, en forma similar se obtiene la entropía del canal (equivocación de B respecto de A).

$$H(B/A) = - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) * \lg P(b_j/a_i) \quad (10)$$

$$H(B/A) = - \sum_{i,j} P(a_i) * P(b_j/a_i) * \lg P(b_j/a_i)$$

1.3- Información Mutua.

La información mutua es la información transferida desde la fuente al receptor a través del canal, esto es, la información real que llega al receptor en promedio.

Analizando la ecuación (3) se observa que ella representa la incertidumbre con respecto a la fuente luego de recibir un símbolo b_j . Al promediarla a todo el alfabeto de salida (según -9-) resulta la información promedio (de la fuente) que bloquea el canal. Luego para calcular la información promedio real recibida bastará obtener la diferencia entre la entropía a priori y la equivocación de A respecto a B.

Una importante propiedad de la información mutua es la simetría con respecto a las variables de entrada y salida, entonces es válido decir que:

$$I(A, B) = I(B, A) = H(B) - H(B/A) \quad (12)$$

En el caso de canales transparentes:

$$\begin{aligned} H(A) &= H(B) = H(A, B) = I(A, B) \\ H(A/B) &= H(B/A) = 0 \end{aligned}$$

1.4- Capacidad y eficiencia de Canales.

1.4.1- Capacidad

Supuesto un canal de alfabeto de salida B y probabilidades condicionales $P(b_j/a_i)$ la información mutua resulta:

$$\begin{aligned} I(A, B) &= H(A) - H(A/B) \\ &= \sum_{j,i} P(a_i, b_j) * \lg \frac{P(a_i/b_j)}{P(a_i)} \end{aligned} \quad (13)$$

ya que:

$$\begin{aligned} I(A, B) &= H(A) - H(A/B) \\ &= - \sum_i P(a_i) * \lg P(a_i) + \sum_{j,i} P(a_i, b_j) * \lg P(a_i/b_j) \\ &= \sum_{j,i} P(a_i, b_j) * \lg P(a_i/b_j) - P(a_i) * \lg P(a_i) \\ &\stackrel{(*)}{=} \sum_{j,i} P(a_i, b_j) * \lg P(a_i/b_j) - P(a_i, b_j) * \lg P(a_i) \\ &\stackrel{(**)}{=} \sum_{j,i} P(a_i, b_j) * \lg \frac{P(a_i, b_j)}{P(a_i) * P(b_j)} \end{aligned}$$

Analizando la expresión (13) se nota que $I(A, B)$ depende de la probabilidad de los símbolos de entrada y de las características del canal. Entonces para un canal determinado se estudia la variación de la información mutua en función de las probabilidades de entrada.

Con el solo hecho que la probabilidad de entrada de un símbolo sea 1, la información mutua será nula. Es decir que el valor mínimo $I(A, B)$ es cero.

Quedaría ahora analizar el valor máximo de la misma, este valor es posible lograrlo sintonizando la fuente que provee los símbolos de entrada al canal. Este valor máximo es lo que se denomina capacidad del canal, simbolizándola con C .

(*) Se a introducido dentro de la sumatoria sobre j la expresión de la $H(A)$, esto es posible ya que ella es independiente de ese subíndice. Ahora, si reemplazamos $P(a_i)$ por $P(a_i, b_j)$ la expresión no cambia ya que al sumar sobre todos los símbolos de la salida (cuando se realiza la sumatoria sobre j) resulta $P(a_i)$ nuevamente.

(**) Es válido ya que $P(a_i/b_j) P(b_j) = P(a_i, b_j)$.

La capacidad de un canal será función exclusivamente de las probabilidades condicionales y no dependerá de las probabilidades de los símbolos de entrada.

En los canales transparentes $H(A/B) = 0$ y resulta:

$$I(A, B) = H(A) - H(A/B) = H(A)$$

entonces:

$$C = \max H(A)$$

Si la fuente resultara equiprobable (caso de máxima entropía) y emitiera un símbolo:

$$H(A) = \lg n \quad (\text{bit/simb})$$

$$C = \lg n \quad (\text{bit/simb})$$

Notar que si n aumenta, la capacidad sube, lo cual es lógico ya que el canal no bloquea la información por es transparente.

Para una fuente binaria equiprobable:

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ P(0) & P(1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

resulta $C = \lg 2 = 1 \text{ bit/simb}$

La capacidad puede expresarse también, en unidades b.p.s. (bit/seg.). Esto proviene de que si cada símbolo dura τ seg., y estamos en el caso binario equiprobable la capacidad C resulta:

$$C = 1/\tau \quad \text{b.p.s.}$$

1.4.2- Eficiencia

Se define a eficiencia de transmisión para cada canal y fuente, a:

$$\text{FF} = I(A, B) / C \quad (16)$$

Y redundancia de transmisión a:

$$\text{RED} = 1 - \text{FF}$$

Esta redundancia es la parte no aprovechada de la capacidad portante de información del canal.

1.5- Otros Conceptos Fundamentales

1.5.1- Velocidad de información y señalización

Se denomina velocidad o tasa de información (R), a la cantidad de bits en promedio que se transmiten en 1 segundo, esto es, la cantidad de información promedio producida por una fuente en 1 (un) segundo.

$$R = H'(A)$$

Si se considera un canal con ruido, a la salida se tiene una tasa de información de:

$$R = H'(A) - H'(A/B)$$

Se denomina velocidad o tasa de dígitos o binitos (r) a la cantidad de dígitos por segundo transmitidos.

Como la entropía es la información promedio (o bits por símbolos en promedio $-H$), resulta:

$$R = r H(A)$$

R y H son dos formas de medir información promedio, una en bits por segundos y otra en bits por símbolo.

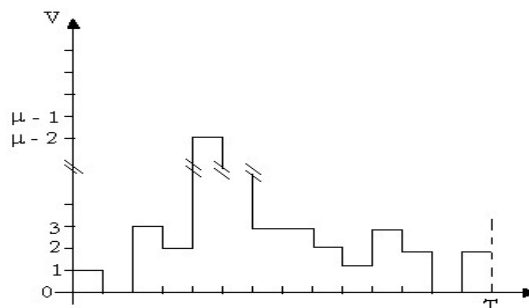
Se denomina velocidad de señalización (s) a la inversa del tiempo mínimo que dura un pulso.

$$s = 1/\tau_{\min} \quad 1/\text{seg.} \text{ ó } \text{baudios}$$

1.5.2- Capacidad de un canal y velocidad de información

La capacidad de un canal es su capacidad de transportar información. Se mide en bits/seg. y una fuente con tasa de información promedio R , si $R \leq C$ existe una codificación tal que la salida de la fuente puede ser transmitida sobre el canal con una frecuencia arbitrariamente pequeña de error, a pesar de la existencia de ruido. Si $R > C$ no es posible transmitir sin error.

Se supone de duración T compuesto por pulsos de duración τ , resulta $r = s$, esto es, la cantidad de pulsos por segundo transmitidos es el máximo para un cierto s . Los pulsos pueden tener μ niveles distintos.



Es posible preguntarse, ¿Cuál será la cantidad de símbolos distintos que podrán representarse?. Esta cantidad resulta igual a los μ niveles. ¿Pero que cantidad de mensajes distintos pueden transmitirse?. Esta cantidad será igual al número de posibilidades de arreglos con repetición en T segundos.

$$M(T) = \mu^n$$

Si todos los pulsos son de inf. y de igual duración $n = T/\tau$

Si son equiprobables se tiene la máxima cantidad de información, la cual resulta:

$$P(M) = 1/M(T)$$

luego:

$$I = \lg M(T)$$

Para obtener R basta con dividir por T , con ello se halla la tasa de información, R , en bits por segundos. No es necesario calcular la entropía, ya que al ser equiprobables coincide con I .

Entonces el cálculo de R es:

$$R = 1/T \lg M(T) = 1/T \lg \mu^n$$

$$R = 1/\tau \lg \mu \quad (\text{solo si todos los pulsos son de inf. y de igual duraci3n})$$

Si $\tau = \tau_{\min}$ luego $1/\tau = s$

Si son sucesos equiprobables y $\tau = \tau_{\min}$, entonces R es máximo y deberá ser $C = R_{\max}$. Esto es, el canal deberá tener una capacidad al menos igual a R. Si el sistema es binario, la cantidad de niveles será μ , lo que implica una capacidad de canal coincidente con la de señalización.

$$C = s$$

La velocidad de señalización está ligada con una característica del canal que se denomina ancho de banda, B. Este valor resulta:

$$B = s/2$$

Ensayando el canal el ancho de banda resulta aproximadamente igual al valor para el cual la relación entre la salida y la entrada disminuye una determinada cantidad.

