Matemáticas Discretas

L. Enrique Sucar

INAOE

Teoría de Probabilidad

"Considero que la probabilidad representa el estado de la mente con respecto a una afirmación, evento u otra cosa para las que no existe conocimiento absoluto"

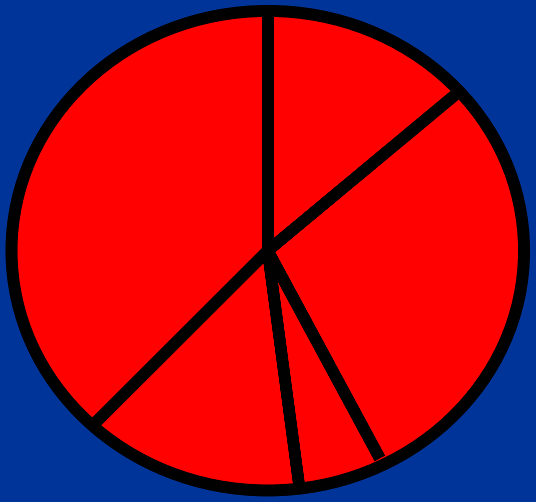
[August De Morgan, 1838]

Conceptos de Probabilidad

- Interpretaciones
- Definición y axiomas
- Probabilidad condicional
- Teorema de Bayes
- Independencia e independencia condicional
- Variables aleatorias y distribuciones básicas
- Teoría de información

¿Qué es probabilidad?

- Interpretaciones
- Definición matemática



Interpretaciones

- Clásica – eventos equiprobables
- Lógica – medida de grado de creencia racional (inferencia respecto a evidencia)
- Subjetiva – medida del grado de creencia personal (factor de apuesta)
- Frecuencia – medida del número de ocurrencias con *muchas* repeticiones
- Propensión – medida del número de ocurrencias bajo condiciones repetibles

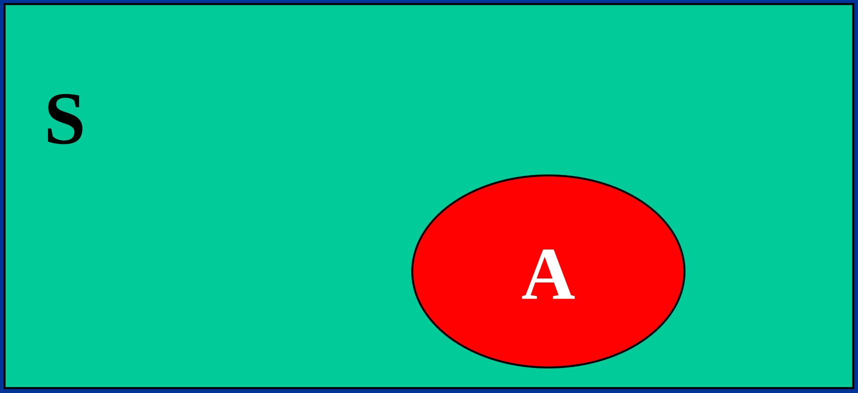
Interpretaciones

Dos principales enfoques:

- Objetiva (clásica, frecuencia, propensión) – las probabilidades existen y se pueden medir en el mundo real
- Epistemológica (lógica, subjetiva) – las probabilidades tienen que ver con el conocimiento humano, medida de creencia

Definición

- Dado un experimento *E* y el espacio de muestreo *S*, a cada evento *A* le asociamos un número real *P(A)*, el cual es la probabilidad de *A* y satisface los siguientes axiomas



Axiomas

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- $P(A \cup B \cup C \ldots) = P(A) + P(B) + P(C) + \ldots$

  A, B, C … mutuamente exclusivos

Justificaciones de Probabilidad

- Argumento del “libro holandés”
- Deducción de Cox
- Demostración lógica

Teoremas

- $P(\emptyset) = 0$
- $P(\neg A) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidad Condicional

$$P(A \mid B) = P(A \cap B) / P(B)$$

- Probabilidad de que ocurra un evento dado que ocurrió otro:
  - Dado que el dado cayó par, cuál es probabilidad de que sea un número primo?
  - Dado que tiene catarro, cuál es la probabilidad de que tenga gripe?

Regla de Bayes

- De la definición de probabilidad condicional se puede deducir:

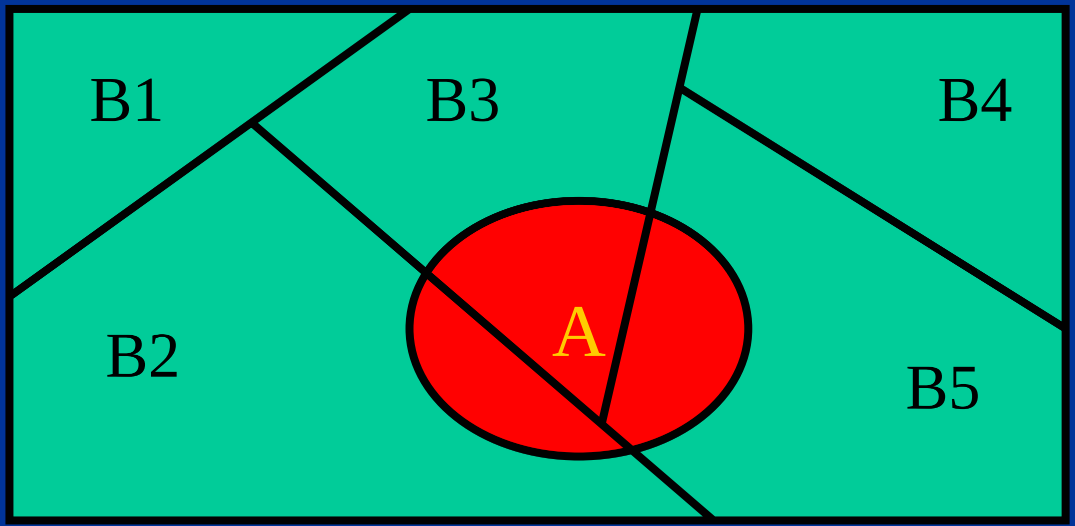
$$P(B \mid A) = P(B)\, P(A \mid B) \,/\, P(A), \text{ dado } P(A) > 0$$

- Esto permite "invertir" las probabilidades, por ejemplo obtener la *P* de una enfermedad dado un síntoma, con conocimiento de la *P* de los síntomas dado que alguien tiene cierta enfermedad

Probabilidad Total

- Dada una partición, *B*, de *S*, la probabilidad de un evento *A* se puede obtener como:

$$P(A) = \Sigma_i \; P(A \mid B_i) \; P(B_i)$$



Teorema de Bayes

- Con la definición de probabilidad total, el teorema de Bayes se puede escribir como:

$$P(B \mid A) = P(B)\, P(A \mid B) \,/\, \Sigma_i\, P(A \mid B_i)\, P(B_i)$$

Eventos independientes

- Dos eventos son independientes si la ocurrencia de uno no altera la probabilidad de ocurrencia del otro:

$$P(A \mid B) = P(A) \text{ ó}$$

$$P(B \mid A) = P(B)$$

- Lo que es equivalente a:

$$P(A \cap B) = P(A)\, P(B)$$

- Independientes $\neq$ mutuamente exclusivos

Independencia condicional

- *A* es condicionalmente independiente de *B* dado *C*, si el conocer *C* hace que *A* y *B* sean independientes:

$$P(A \mid B,C) = P(A \mid C)$$

- Ejemplo:
  - A – regar el jardín
  - B – predicción del clima
  - C – lluvia

Regla de la Cadena

- De la definición de probabilidad condicional, se puede evaluar la probabilidad de $A_1 \cap A_2 \cap A_3 \ldots \cap A_N$ (probabilidad conjunta) como:

$$P(A_1, A_2, \ldots, A_N) =$$
$$P(A_1 \mid A_2, \ldots, A_N)\, P(A_2 \mid A_3, \ldots, A_N) \ldots P(A_N)$$

Variables Aleatorias

- A cada evento *A* se le asigna un valor numérico $X(A) = k$, de forma que a cada valor le corresponde una probabilidad $P(X = k)$
- *X* es una variable aleatoria
- Ejemplos:
  – X = Número de águilas en *N* lanzamientos
  – Y = Número del dado al lanzarlo
  – Z = Número de fallas antes de darle a un blanco

Ejemplos – variables aleatorias

- Suma de las caras superiores al lanzar 2 dados – cuál es la probabilidad de cada valor?
- En un sistema de comunicaciones, se transmiten mensajes de longitud variable, de 1 a N bytes. Número de bytes en el mensaje es una v.a.

Tipos de Variables Aleatorias

- Discretas: el número de valores de X (rango) es finito o contablemente finito
- Continua: puede asumir todos los posibles valores en cierto intervalo $a - b$, ejemplos:
  - X = temperatura ambiente
  - Y = tiempo en el que falle cierto dispositivo
  - Z = distancia del robot a la pared

Distribución de probabilidad

- Variables discretas: p(X):

  $p(X) \geq 0$

  $\Sigma\, p(X) = 1$

- Variables continuas: f(x):

  $f(x) \geq 0$

  $\int f(x) = 1$

Función acumulativa

- Probabilidad de que la variable X tome un valor menor a x
- Discretas: $P(X) = \Sigma_x \, p(X)$
- Continuas: $F(X) = \int_x f(X)$
- Propiedades:
  - $0 \leq F(X) \leq 1$
  - $F(X1) \leq F(X2)$, si $X1 \leq X2$
  - $F(-\infty) = 0$
  - $F(+\infty) = 1$

Estadísticas

- Moda: valor de mayor probabilidad
- Mediana: valor medio (divide el área en 2)
- Promedio: valor "esperado":

$$E(X) = \Sigma_x X \, p(X)$$

- Varianza: dispersión

$$\sigma^2(X) = \Sigma_x (X - E(X))^2 \, p(X)$$

- Desviación estandar

$$\sigma(X) = \sqrt{\sigma^2}$$

Ejemplo

- Para la v.a de la suma de 2 dados:
  - Moda
  - Mediana
  - Promedio
  - Varianza

Variables aleatorias en 2-D

- X y Y son dos funciones que asignan números reales a los eventos en S, entonces (*X*, *Y*) es una variable aleatoria en dos dimensiones
- Propiedades

  $p(X,Y) \geq 0$

  $\Sigma\, \Sigma\, p(X,Y) = 1$
- Ejemplos:
  - Número de artículos terminados en dos líneas de producción
  - Número de pacientes con cáncer y número que fuma

Ejemplo – v.a. en 2D

X \ Y	Y1	Y2	Y3
X1	0.15	0.15	0.2
X2	0.15	0.15	0.2

Probabilidad conjunta, marginal, y condicional

- Probabilidad conjunta:

  $p(X,Y)$

Probabilidad marginal:

$$p(X) = \Sigma_Y \, p(X,Y)$$

- Probabilidad condicional:

$$p(X \mid Y) = p(X,Y) / p(Y)$$

Ejemplo

- $P(x=x1, y=y2) = ?$
- $P(x=x1) = ?$
- $P(y=y3) =?$
- $P(y3 \mid x1) =?$

Independencia y Correlación

- Dos variables aleatorias son independientes si su probabilidad conjunta es el producto de las marginales:

$$p(X,Y) = p(X)\ p(Y)$$

- Correlación: grado de relación lineal entre dos variables aleatorias (diferente independencia):

$$\rho\ (X,Y) = E\{[(X - E(X)][Y - E(Y)]\} / \sigma_X\ \sigma_{Y,,}$$

$$[-1, 1]$$

Ejemplo

- Son X y Y independientes?

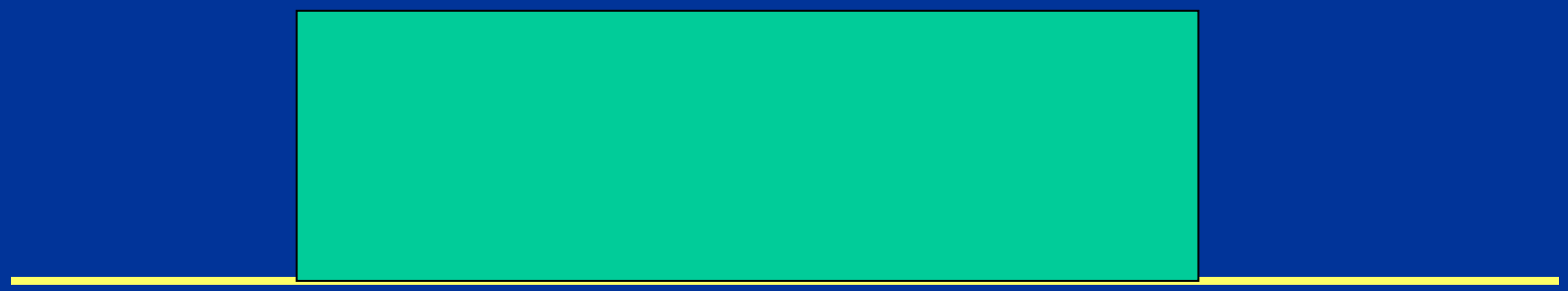
Distribuciones básicas

- Uniforme
- Binomial
- Gaussiana o normal

- Histograma de una variable aleatoria

Uniforme

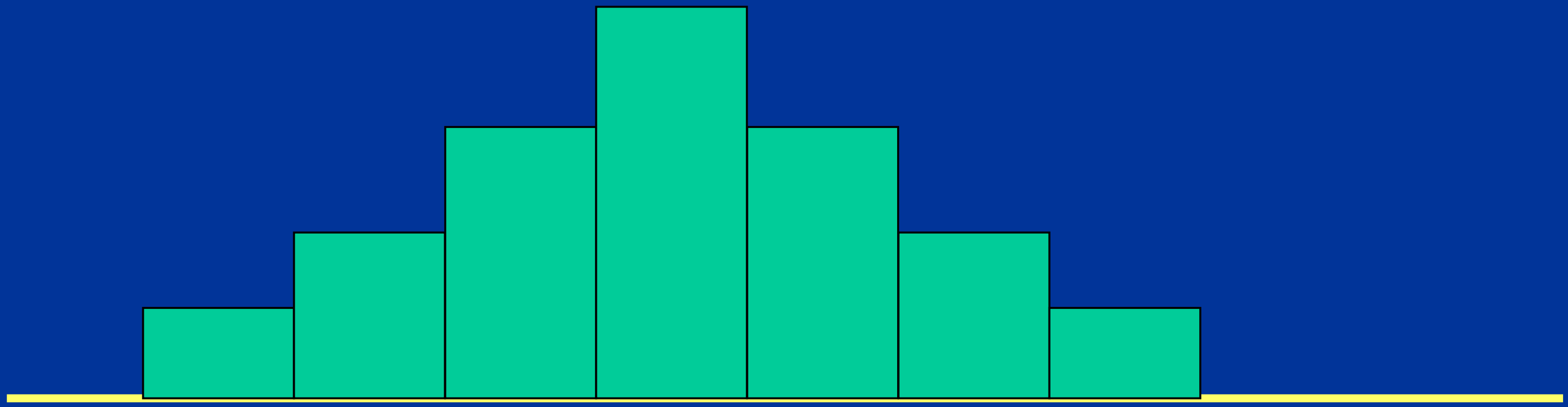
- Todos los valores en el rango son equiprobables



Binomial

- *X* es el número de valores verdaderos en *N* repeticiones de un proceso de Bernoulli con probabilidad *P* de verdadero (éxito)

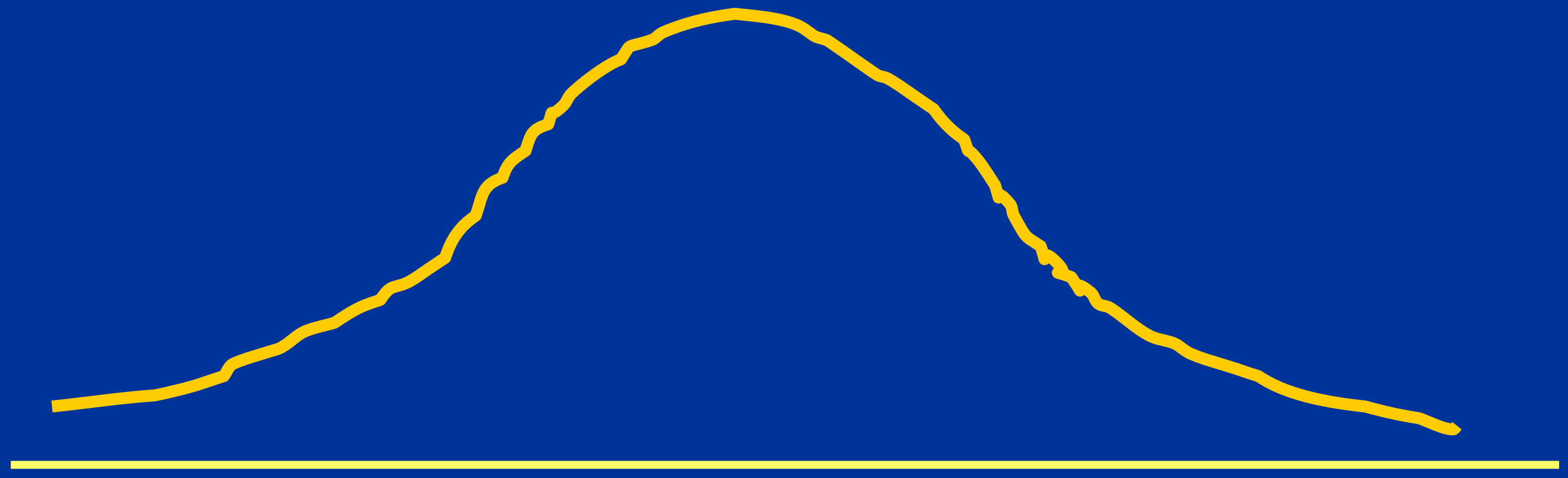
$$P(X=k) = (n\ k)\ p^{k}\ (1-p)^{n-k}$$



Gaussiana

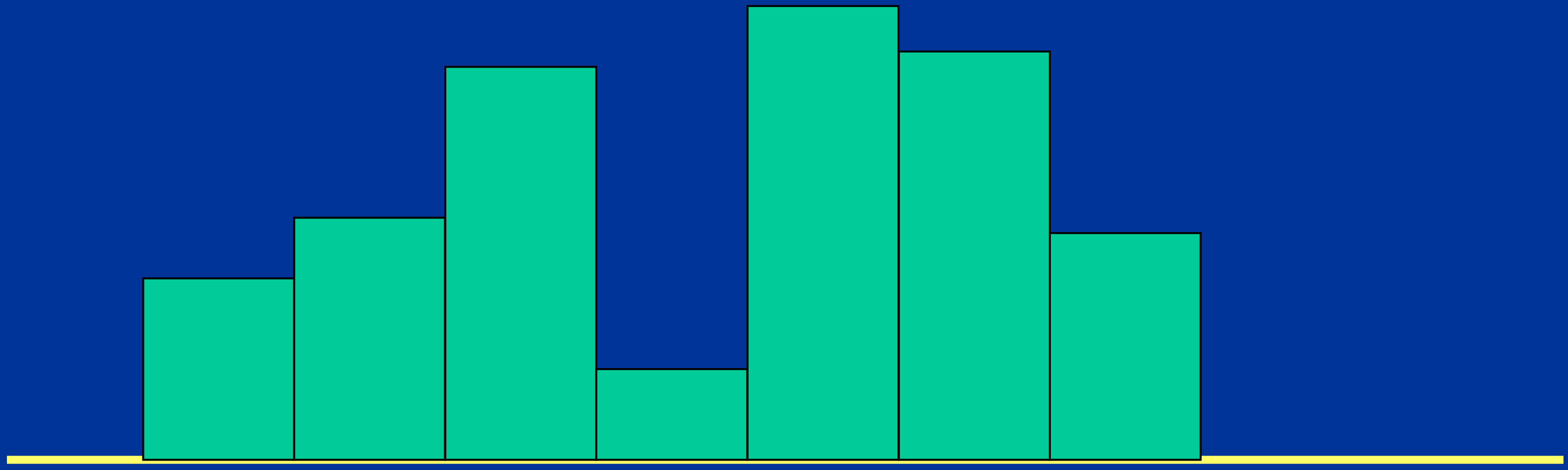
- Aproximación a la binomial con p=0.5 y N muy grande (corresponde a la suma de muchas variables aleatorias independientes)

$$f(x) = 1/\sigma(2\pi)^{1/2} \exp[-1/2 ((x-\mu)/\sigma)^2]$$



Histograma

- Muestra el número de datos por intervalo en forma absoluta o relativa



Conceptos de Teoría de Información

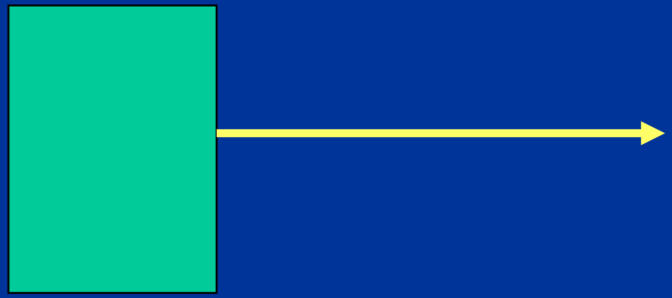
- Definición
- Medida de Información
- Entropía

Información

- La cantidad de información recibida respecto a la ocurrencia de un evento es *inversamente proporcional* a su probabilidad
- Ejemplos
  - Está nublado en Puebla
  - Hizo erupción el Popocatépetl
  - Renunció Calderón

Medidia de Información

- Dada una fuente de información discreta
  - *q* posibles mensajes: $m_1, m_2, ...m_q$
  - con probabilidades: $p_1, p_2, ..., p_q$



$m_1, m_2, ...m_q$

Propiedades de la medida (I)

- $I(m1) > I(m2)$, si $p1 < p2$
- $I(m) \rightarrow \infty$, si $p \rightarrow 0$
- $I(m) \geq 0$, $0 \leq P(A) \leq 1$
- $I(m1 + m2) = I(m1) + I(m2)$, si m1 y m2 son independientes

Medida de Información

- Función logarítmica:

$$I(m_k) = \log (1/p_k)$$

- En base 2 (en bits):

$$I(m_k) = \log_2 (1/p_k)$$

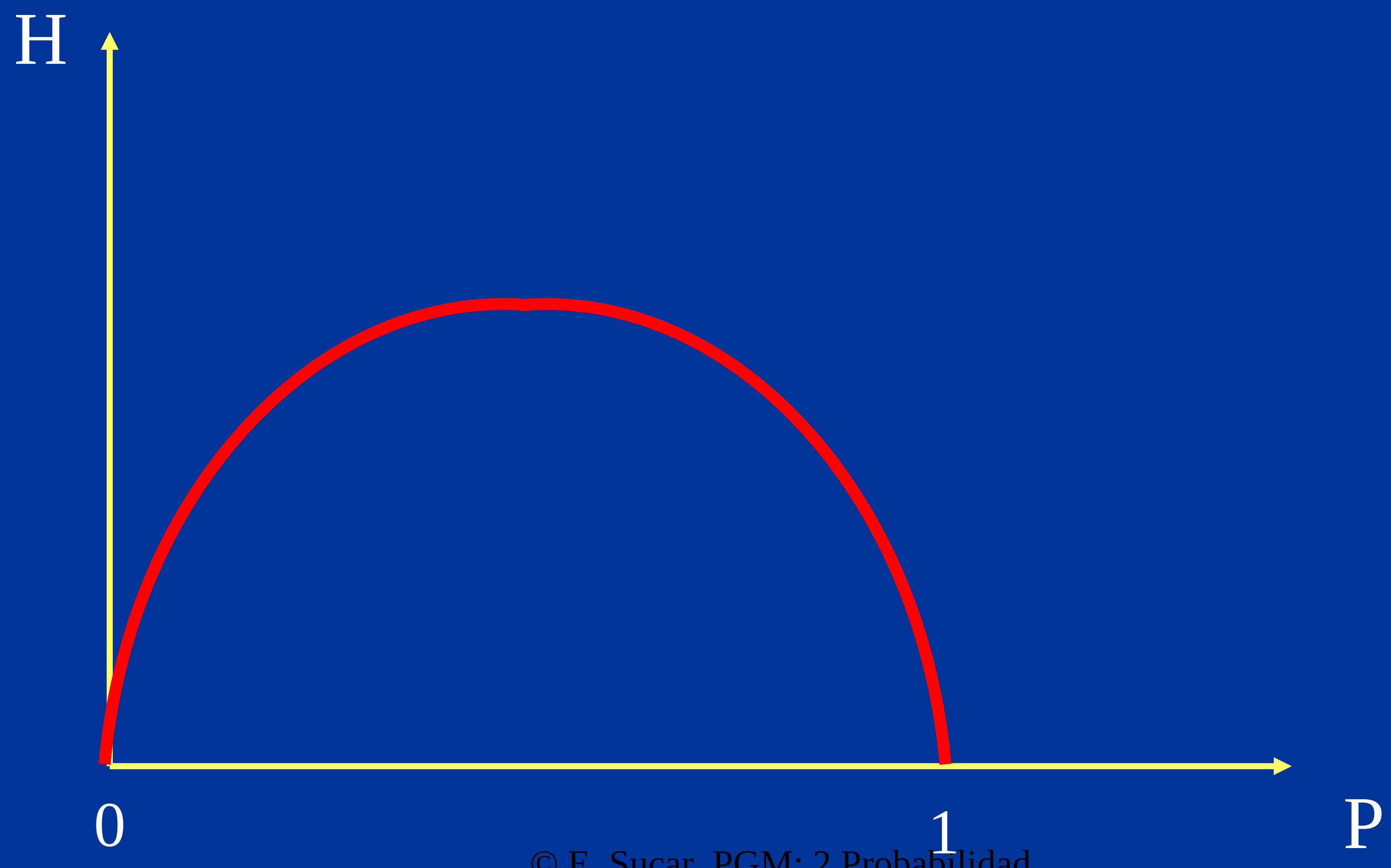
Entropía

- Información promedio de un mensaje que puede tomar *n* valores:

$$H = E(I) = \Sigma_i \, p_i \log_2 (1/p_i)$$

- *H* es la entropía
- En promedio, se esperan recibir *H* bits de información
- Cuándo es H máxima y cuando es mínima?

Ejemplo: H para una fuente binaria



Referencias

- Liu, Capítulo 3
- Libros básicos de probabilidad, por ej.:
  - Meyer, Introductory Probability and Statistical Applications
  - Waserman, All of statistics

Ejercicios

1. Al tirar 2 dados (1 al 6 c/u), que tan probable es que sumen 7?
2. Si tiro una moneda 10 veces, que tan probable es que salgan 4 águilas?
3. Dada la siguiente tabla de probabilidades conjuntas, encuentra las siguientes probabilidades: P(x1), P(y2), P(x1 | y2). Son X y Y independientes?
4. En el problema anterior, son x1 y x2 independientes?

Tabla Conjunta (ej. 3 y 4)

X \ Y	Y1	Y2	Y3
X1	0.1	0.2	0.1
X2	0.3	0.1	0.2

Ejercicios

5. Alguien tiene una, y solo una, de dos posibles enfermedades: tifoidea (T) o hepatitis (H). Hay dos posibles síntomas: dolor de cabeza (D) y fiebre (F), cada uno de los cuales puede ser verdadero (D, F) o falso (~D, ~F). Dados:

   $P(T) = 0.6$
   $P(D|T) = 0.7$
   $p(D|\sim T) = 0.4$
   $P(F|T) = 0.9$
   $P(F|\sim T) = 0.5$

   Describe el espacio de muestreo y completa las tablas de probabilidad.

6. Si generamos en el grupo un comité al azar de 3 estudiantes, que probabilidad hay de que haya exactamente una mujer? Al menos una mujer? (asumiendo 60 alumnos, 40 hombres y 20 mujeres)

Ejercicios

7. Si asumimos que la estatura de los estudiantes sigue una distribución gaussiana, con media 1.70 m y desviación estándar 0.10 m, que tan probable es que haya un estudiante de más de 1.90 m?
8. Demuestra por inducción la regla de la cadena
9. Que probabilidad hay de obtener hacer un “full” (de 5 cartas, 3 con el mismo número y 2 con el mismo número, diferente a las primeras. El poker se puede ver como 4 clases de elementos (figuras), cada clase numerada del 1 al 13.

Ejercicios

10. En cierto lugar el clima se comporta estadisticamente de la siguiente manera: de 365 dias, 200 soleados, 60 nublados, 40 lluvia, 20 nieva, 20 tormenta, 10 graniza, 10 viento y 5 llovizna:
    - Si cada día se envia un mensaje con el clima, que información da para cada tipo de clima?
    - Cuál es el promedio de bits de información que da el mensaje?