Probabilidades y Estadística
Cs. de la Computación

Introducción

Breve reseña histórica:

La teoría de Probabilidades comienza a partir de una disputa entre jugadores en 1654. Los dos matemáticos que participaron de tales discusiones fueron Blaise Pascal y Pierre de Fermat, y su intercambio de correspondencia sentó las bases de la teoría de Probabilidades. Un matemático holandés, Christian Huygens tomó contacto con esa correspondencia y escribió el primer libro sobre Probabilidades en 1657, el cual trataba fundamentalmente sobre problemas relacionados con los juegos de azar.
Durante el siglo XVIII la teoría se desarrolló y se enriqueció con los aportes de Jacob Bernoulli y Abraham de Moivre. En 1812 Pierre de Laplace introdujo una serie de nuevas ideas y técnicas matemáticas en su libro Theorie Analytique des Probabilités y fundamentalmente sacó a la teoría del marco exclusivo de los juegos de azar y aplicó las ideas a muchos problemas científicos y prácticos. Algunas de las importantes aplicaciones desarrolladas en el siglo XIX fueron: teoría de errores, matemática actuarial y mecánica estadística.

Una de las dificultades para el desarrollo de la teoría matemática de las probabilidades fue llegar a una definición de probabilidad matemáticamente rigurosa, pero al mismo tiempo amplia para permitir su aplicación a un amplio rango de fenómenos. En el siglo XX se llegó a una definición axiomática de las Probabilidades (Kolmogorov, 1933).

¿Porqué estudiar Probabilidades y Estadística en Ciencias de la Computación?:

Posibles preguntas que queremos responder:
- ¿Cuál es el máximo número de terminales que pueden estar conectadas en un servidor antes de que el tiempo medio de espera se haga inaceptable?
- En una base de datos, ¿Cómo deberían ser guardados los datos para minimizar el tiempo medio de acceso?

Los sistemas de computación no son determinísticos. Pensemos, por ejemplo, en el delay en el envío de paquetes, comunicaciones en una red, equilibrio de “carga” en servidores, requerimientos de memoria, etc.

¿Para qué sirven las Probabilidades? Si bien estamos frente a procesos aleatorios, no son necesariamente “caóticos”, en el sentido que podemos descubrir un patrón de comportamiento que pueda ser modelado.

Veamos un ejemplo de uso frecuente.

Compresión de archivos: El código ASCII contiene 256 caracteres, cada uno de los cuáles se representa con un número consistente en 8 dígitos binarios, por ejemplo, á se representa por 160 ≡ 10100000.
Para simplificar el problema, supongamos que contamos con sólo 4 caracteres: A, B, C y D. Para representarlos necesitamos 2 bits. Por ejemplo, podríamos representarlos así:

A → 00
B → 01
C → 10
D → 11

Si un texto constara de n caracteres necesitaríamos 2n bits para guardarlo. Esta cantidad de bits es *determinística*.
Supongamos que sabemos que ciertas letras aparecen con más frecuencia que otras, por ejemplo, supongamos que sabemos que las frecuencias con que aparecen las 4 letras en un texto son:

A	0.70 (70%)
B	0.12 (12%)
C	0.10 (10%)
D	0.08 (8%)

El método de codificación de Huffman utiliza la información disponible sobre la frecuencias de aparición de los caracteres y asigna códigos de longitud variable. Por ejemplo, podríamos asignar a los 4 caracteres de nuestro ejemplo los siguientes códigos:

A → 1
B → 00
C → 011
D → 010

¿Cuánto espacio (en bits) ocuparía ahora un texto de n caracteres? No lo sabemos, pero podemos suponer que tendremos en promedio:

0.70 n	veces A's
0.12 n	veces B's
0.10 n	veces C's
0.08 n	veces D's

y el número de bits requerido sería:

$$0.70\,n * (1) + 0.12\,n *(2) + 0.10\,n * (3) + 0.08\,n * (3) = 1.48\,n.$$

Como se observa, el método produce una disminución del espacio promedio requerido para almacenar un texto.

Probabilidad

El término Probabilidad se refiere al estudio del azar y la incertidumbre. En aquellas situaciones en las cuáles se puede producir uno de varios resultados posibles, la Teoría de la Probabilidad provee métodos para cuantificar la chance de ocurrencia de cada uno de ellos.

Ejemplos:

- Se arroja un dado dos veces y se registra la suma de puntos. ¿Cuál es la probabilidad de que se obtenga una suma mayor que 10?
- En un juego de ruleta, ¿cuál es la probabilidad de ganar apostando a primera columna?
- En un juego de ruleta, ¿cuál es la ganancia esperada apostando repetidamente a primera columna?
- ¿Cuál es la probabilidad de que un servidor que atiende a 20 terminales se sature en un determinado momento?
- Dada la información disponible, ¿cuál es la probabilidad de que llueva el próximo fin de semana?

Definiciones:

Experimento: Es cualquier proceso o acción que genera observaciones y que puede ser repetible. Por ejemplo, arrojar un dado, seleccionar un individuo y registrar su peso y su altura, seleccionar una muestra de productos elaborados por una empresa para hacer un control de calidad, seleccionar un día al azar y registrar el número de veces que se satura un servidor.

Espacio muestral asociado a un experimento: Es el conjunto de todos los resultados posibles del experimento. Lo notaremos S.

Ejemplos:

1) Se arroja una moneda una vez.
   S={cara,ceca} ó S={1,0} ó S={éxito,fracaso}

2) Se arroja una moneda dos veces.
   S={(1,1),(1,0),(0,1),(0,0)}

3) Se arroja una moneda hasta que aparece por primera vez una cara.
   $S=\{(1),(0,1),(0,0,1),(0,0,0,1),....\} = \{(x_1,x_2,...x_n) \,/\, n\in N,\ x_i=0 \text{ si } i < n\ ,\ x_n=1\}$

4) Se registra el tiempo transcurrido desde que se intenta la conexión a un servidor hasta que la conexión se efectiviza.

$S=\Re^{+}=(0,\infty)$
Si el sistema tiene un time-out en el tiempo t_o , tendríamos $S=(0, t_o)$.

Como se observa, un espacio muestral puede ser *finito*, como en los ejemplos 1) y 2), *infinito numerable*, como en el ejemplo 3) o *infinito no numerable*, como en el ejemplo 4).

Sucesos o eventos: No sólo estamos interesados en resultados individuales de un experimento sino que pueden interesarnos colecciones o conjuntos de ellos. Se denomina suceso o evento a cualquier subconjunto del espacio muestral. Si S es finito o infinito numerable, cualquier subconjunto es un evento. Si S es infinito "casi todo" subconjunto de S es un evento. Los eventos los designaremos en general con las primeras letras del abecedario en mayúscula: A, B, C,...

Evento elemental o simple: consiste de un único resultado individual.

Evento compuesto: consiste de más de un evento elemental.

Ejemplos: En los ejemplos anteriores, posibles eventos son

1) A = "sale cara" = {cara}={1}.

2) A = "número de caras es menor o igual que 1" ={(1,0),(0,1),(0,0)}.

3) A = "número de tiros requeridos es menor o igual que 5" = $\{(x_1,x_2,...x_n)\in S \ / \ n\leq 5 \}$.

   B = "número de tiros requeridos es par" = $\{(x_1,x_2,...x_n) \in S \ / \ n=2k, \ k \in N\}$.

4) A = "el tiempo es mayor de 10 minutos" = $(10,\infty)$ (en el caso de un sistema sin time-out)

Relación con Teoría de conjuntos: Como un evento o suceso es un conjunto, valen las mismas relaciones que en teoría de conjuntos.

S es un subconjunto de S denominado *suceso cierto o seguro* .

$\varnothing$ es un subconjunto de S denominado *suceso imposible.*

$A \cup B$ es el suceso *unión.* Ocurre cuando A ocurre ó B ocurre.

$A \cap B$ es el suceso *intersección.* Ocurre cuando ocurre A y ocurre B.

A^c ó $\overline{A}$ es el *opuesto* o *complemento* de A. Ocurre cuando no ocurre A.

$A - B = A \cap B^c$ es el suceso *diferencia*. Ocurre cuando ocurre A y no ocurre B.

Se dice que A *está contenido* en B o que A implica B y se denota A $\subseteq$ B si la realización de A conduce a la realización de B, es decir si todo elemento de A pertenece a B.

Dos sucesos A y B se dicen *mutuamente excluyentes* o *disjuntos* si A $\cap$ B = $\emptyset$.

Recordemos algunas propiedades:

Asociatividad: A $\cup$ B $\cup$ C = (A $\cup$ B) $\cup$ C = A $\cup$ (B $\cup$ C)
A $\cap$ B $\cap$ C = (A $\cap$ B) $\cap$ C = A $\cap$ (B $\cap$ C)

Conmutatividad: A $\cup$ B = B $\cup$ A
A $\cap$ B = B $\cap$ A

Distributividad: (A $\cup$ B) $\cap$ C = (A $\cap$ C) $\cup$ (B $\cap$ C)
(A $\cap$ B) $\cup$ C = (A $\cup$ C) $\cap$ (B $\cup$ C)

Leyes de De Morgan: $\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$ y $\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$

Interpretación intuitiva de la Probabilidad: Supongamos que se repite *n* veces un mismo experimento aleatorio en forma independiente y bajo las mismas condiciones. Sea n_A el número de veces que ocurre el suceso A en las *n* repeticiones. Se denomina *frecuencia relativa* de A en la secuencia de *n* repeticiones a

$$fr(A) = \frac{n_A}{n}$$

La evidencia empírica muestra que cuando *n* crece, $fr(A)$ tiende a estabilizarse alrededor de un número que llamaremos P*(A)*.

¿Qué propiedades tiene la frecuencia relativa?

1) $fr(A) = \frac{n_A}{n} \geq 0$

2) $fr(S) = \frac{n_S}{n} = \frac{n}{n} = 1$

3) Si A $\cap$ B = $\emptyset \Rightarrow$ $fr(A \cup B) = \dfrac{n_{A\cup B}}{n} = \dfrac{n_A + n_B}{n} = \dfrac{n_A}{n} + \dfrac{n_B}{n} = fr(A) + fr(B)$

La definición axiomática de Probabilidad, que daremos a continuación, es consistente con la idea intuitiva que se tiene de ella.

Axiomas de Probabilidad: Dado un experimento aleatorio y un espacio muestral asociado S, a cada evento A se le asociará un número que notaremos P(A) y que llamaremos probabilidad del evento A. Esta asignación debe satisfacer los siguientes axiomas:

A1. P(A) $\geq$ 0 para todo evento A.

A2. P(S) = 1

A3a. Si $A_1, A_2, ..., A_n$ es una colección finita de sucesos mutuamente excluyentes, es decir que $A_i \cap A_j = \emptyset \;\; \forall\, i \neq j$, entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

A3b. Si $A_1, A_2, ..., A_n ,...$ es una colección infinita numerable de sucesos mutuamente excluyentes, es decir si $A_i \cap A_j = \emptyset \;\; \forall\, i \neq j$, entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Ejemplo: Consideremos el ejemplo en que se arroja una moneda una vez, para el cual el espacio muestral es S={cara,ceca}. Si denominamos E_1 = {cara} y E_2 ={ceca} a los dos eventos elementales, como P(S) = 1 = P(E_1)+P(E_2), entonces P(E_2) = 1- P(E_1). Por lo tanto, cualquier asignación de probabilidades de la forma: P(E_1) = p y P(E_2)=1-p con 0 $\leq$ p $\leq$ 1, satisface los axiomas.

Propiedades de la Probabilidad:

1) $P(A^c) = 1 - P(A)$ para todo suceso *A*

Dem: $1 \underset{A2}{=} P(S) = P(A \bigcup A^c) \underset{A3a}{=} P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$

En la tercera igualdad usamos el axioma 3 pues $A \cap A^c = \emptyset$.

2) P(∅) = 0

Dem: $P(\emptyset) \underset{P1}{=} 1 - P(\emptyset^c) = 1 - P(S) \underset{A2}{=} 1 - 1 = 0$

3) Si $A \subseteq B \Rightarrow P(A) \leq P(B)$ y $P(B-A) = P(B) - P(A)$

Dem: Si $A \subseteq B \;\Rightarrow\; B = A \cup (B-A)$ y éstos dos eventos son excluyentes. Por el axioma A3a

$$P(B) = P(A) + P(B-A)$$

Dado que, por el axioma A1, *P(B-A)* $\geq$ *0* , resulta *P(B)* $\geq$ *P(A)* y, despejando, se obtiene la segunda afirmación.

4) Dados dos sucesos cualesquiera A y B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dem: $A \cup B = A \cup (B - A) = A \cup (B \cap A^c)$ y estos dos eventos son excluyentes, entonces, por el axioma A3a,

$$P(A \cup B) = P\big(A \cup (B \cap A^c)\big) = P(A) + P(B \cap A^c) \qquad (1)$$

Por otra parte, $B = (B \cap A) \cup (B \cap A^c)$ y estos dos eventos son disjuntos, entonces

$$P(B) = P(B \cap A) + P(B \cap A^c) \;\Rightarrow\; P(B \cap A^c) = P(B) - P(B \cap A) \qquad (2)$$

De (1) y (2) resulta que $P(A \cup B) = P(A) + P(B) - P(B \cap A)$ como queríamos demostrar.

5) Dados dos sucesos cualesquiera A y B, $P(A \cup B) \leq P(A) + P(B)$.

Dem: Esta propiedad se deduce inmediatamente de la propiedad anterior y del axioma A1.

Ejercicios: a) Demostrar, usando la propiedad 4) que, dados tres sucesos cualesquiera, A_1, A_2 y A_3,

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) \\ & - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

b) Probar, usando inducción que, dados $A_1, A_2, ..., A_n$ sucesos cualesquiera,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

<u>Asignación de probabilidades</u>: Supongamos que el espacio muestral S asociado con cierto experimento es finito o infinito numerable. En este caso, una manera simple de trabajar es asignar probabilidades a los sucesos elementales, ya que cualquier suceso A será unión de sucesos elementales y éstos son obviamente mutuamente excluyentes.

Designando E_i a los sucesos elementales de S, S = $\bigcup_{i=1}^{\infty} E_i$ (la unión podría ser finita si el espacio muestral fuese finito). Si conocemos $p_i = P(E_i) \geq 0 \ \forall i$, de manera que $\sum_{i=1}^{\infty} p_i = 1$, entonces dado cualquier suceso A, su probabilidad se puede obtener sumando las probabilidades de los elementales que lo componen, es decir:

$$P(A) = \sum_{E_i \subset A} p_i$$

<u>Ejemplos</u>: 1) Se arroja un dado equilibrado. En este caso, S={1,2,3,4,5,6} y, por suponerlo equilibrado, los sucesos elementales E_i = {i} para i=1,..,6 tienen probabilidad p_i = 1/6. Si deseamos calcular la probabilidad del suceso A = “el resultado es par”, usando que

A= $E_2 \cup E_4 \cup E_6$ se obtiene P(A) = P(E_2)+ P(E_4)+ P(E_6)=1/2

2) Supongamos ahora que se arroja un dado en el cual la probabilidad de las caras pares es el doble que la probabilidad de las caras impares, o sea que, si llamamos p a la probabilidad de cada cara impar,

P(E_1) = P(E_3) = P(E_5) = p y P(E_2) = P(E_4) = P(E_6) = 2 p

Como la suma de las probabilidades debe ser igual a 1,

$$\sum_{i=1}^{6} P(E_i) = 3p + 6p = 9p = 1 \qquad \Rightarrow \qquad p = \frac{1}{9}$$

y, en este caso, P(A) = P(E_2)+ P(E_4)+ P(E_6) = $3\frac{2}{9} = \frac{2}{3}$.

3) Arrojamos una moneda equilibrada hasta obtener cara. ¿Cuál es la probabilidad de que la cara sea obtenida en un número par de lanzamientos?

Si representamos el espacio muestral tal como lo hicimos más arriba, tendríamos

A={(0,1),(0,0,0,1),(0,0,0,0,0,1),.....}

Veremos más adelante que en las condiciones de este experimento es razonable asumir que

$$\text{P(obtener cara en el k - ésimo lanzamiento)} = \left(\frac{1}{2}\right)^k$$

Por lo tanto:

$$P(A) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k} = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^{k} = \frac{1}{1-\frac{1}{4}} - 1 = \frac{4}{3} - 1 = \frac{1}{3}$$

ya que si 0<p<1, entonces

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$$

<u>Espacios de equiprobabilidad</u>: Sea un experimento aleatorio cuyo espacio muestral asociado S es finito y sea n = # S (el símbolo # representa el cardinal del conjunto). Diremos que el espacio es de equiprobabilidad si los n sucesos elementales tienen igual probabilidad, es decir si

$$P(E_i) = p \qquad \forall\, i$$

Como $1 = P(S) = \sum_{i=1}^{n} P(E_i) = \sum_{i=1}^{n} p = np \quad \Rightarrow \quad p = \frac{1}{n} = \frac{1}{\# S}$.

Dado cualquier suceso A, $P(A) = \sum_{E_i \subset A} P(E_i) = \sum_{E_i \subset A} \frac{1}{n} = \frac{\# A}{\# S}$.

<u>Ejemplos</u>: 1) De una urna que contiene 2 bolillas blancas y 3 rojas se extraen 2 bolillas *con reposición*.

a) ¿Cuál es la probabilidad de que se extraiga al menos una bolilla roja?
b) ¿Cuál es la probabilidad de que la primera bolilla extraída sea roja y la segunda blanca?

Supondremos que las bolillas están numeradas, de manera de poder considerar que se trata de un espacio de equiprobabilidad, entonces $S = \{(x_1, x_2) / x_i \in \{R_1, R_2, R_3, B_1, B_2\}\}$ y su cardinal es #S = 5 · 5 = 25

a) $P(A) = 1 - P(A^c)$ siendo $A^c = \{(x_1, x_2) \in \mathrm{S} \,/\, x_i \in \{B_1, B_2\}\}$. Como $\# A^c = 2 \cdot 2 = 4$, resulta $P(A^c) = \dfrac{4}{25} \;\Rightarrow\; P(A) = 1 - \dfrac{4}{25} = \dfrac{21}{25}$.

b) $B = \{(x_1, x_2) \in \mathrm{S} \,/\, x_1 \in \{R_1, R_2, R_3\}, x_2 \in \{B_1, B_2\}\}$. Como $\# B = 3 \cdot 2 = 6 \;\Rightarrow\; P(B) = \dfrac{6}{25}$.

2) Consideremos el ejemplo 1) pero suponiendo ahora que las extracciones se realizan *sin reposición*.

En este caso, $S = \{(x_1, x_2) \,/\, x_i \in \{R_1, R_2, R_3, B_1, B_2\}, x_1 \neq x_2\} \;\Rightarrow\; \# S = 5 \cdot 4 = 20.$

a) $P(A) = 1 - P(A^c)$ siendo $A^c = \{(x_1, x_2) \in \mathrm{S} \,/\, x_i \in \{B_1, B_2\}\}$. Como $\# A^c = 2 \cdot 1 = 2$, resulta $P(A^c) = \dfrac{2}{20} = \dfrac{1}{10} \;\Rightarrow\; P(A) = 1 - \dfrac{1}{10} = \dfrac{9}{10}$.

b) $B = \{(x_1, x_2) \in \mathrm{S} \,/\, x_1 \in \{R_1, R_2, R_3\}, x_2 \in \{B_1, B_2\}\}$. Como $\# B = 3 \cdot 2 = 6 \;\Rightarrow\; P(B) = \dfrac{6}{20}$.

Observación: ¿Qué pasaría si en los ejemplos anteriores eligiésemos como espacio muestral $S = \{(B, B), (B, R), (R, B), (R, R)\}$, denotando *B*: bolilla blanca y *R*: bolilla roja? ¿Sería razonable suponer equiprobabilidad?.