Tema 2

Probabilidad

2.1 ¿Por qué estudiar Probabilidad en Computación?

El ser humano ha convivido desde siempre con la incertidumbre. Estamos tan acostumbrados a aceptar hechos que conocemos de manera fragmentaria, a razonar a partir de premisas incompletas, a tomar decisiones basadas en creencias subjetivas, que la presencia de la incertidumbre nos resulta natural. Si salimos a la calle, lo más probable es que, antes de decidir qué ropa ponernos, consideremos las posibilidades de lluvia, quizás sólo observando el trozo de cielo que nos deja ver la ventana, quizás recordando la estación del año y el tiempo que hizo en los últimos días. En todo caso, lo único que hemos hecho es decidir en base a un razonamiento aproximado y cargado de incertidumbre. La causa de esa presencia ubicua de la incertidumbre es la extraordinaria complejidad de la realidad, la multitud de causas que se esconden detrás de hechos simples y que nos resulta difícil de comprender. Sin embargo, sobrevivimos en medio de esa sopa de incertidumbre que nos rodea: tomamos decisiones, creamos modelos para explicar la realidad, nos esforzamos por comprender esa aleatoriedad, por tratarla y sacar provecho de ella, razonamos en su presencia e incluso acumulamos conocimiento a su pesar.

Hay muchos fenómenos físicos gobernados por la incertidumbre, como por ejemplo los fenómenos microscópicos. Pensemos en el comportamiento de los gases, formados por muchísimas partículas cuyo comportamiento se describe teniendo en cuenta las interacciones aleatorias entre ellas. A pesar de esto, hay una teoría de los gases que predice con bastante exactitud el comportamiento macroscópico, lo cual no deja de sorprendernos. Sin duda, donde reina la incertidumbre por sus fueros es en la Mecánica Cuántica, entronizada por el principio de Heisenberg enunciado en el año 1927. Este principio afirma que cuanto más precisa es la medida de la posición de un electrón, más imprecisa es la medida de su velocidad, de modo que no es posible conocer ambos con precisión absoluta. ¿Hay una afirmación más rotunda de la incertidumbre? Las consecuencias de este principio son profundísimas y alcanzan a la ciencia y la técnica de nuestros días, pues termina con una manera determinista de concebir el conocimiento.

No sólo en los fenómenos cuánticos aparece la incertidumbre; quizás en este campo es más patente a causa del principio de incertidumbre, pero a medida que el progreso científico exigió un conocimiento en profundidad de los fenómenos, con más capacidad de predicción, la incertidumbre empieza a aparecer de modo natural. Antes del desarrollo de la probabilidad y

la estadística los análisis de los problemas eran deterministas, y sus conclusiones, limitadas. La incertidumbre, entre otros muchos campos, aparece en:

- Economía y Ciencias Sociales: comportamiento de mercados, índices bursátiles, tendencias sociales, resultados de elecciones, etc.
- Ingeniería: procesos de fabricación, control de calidad, planificación de tareas, mediciones de características, etc.
- Informática y Computación: tráfico en redes de comunicaciones, tiempo de ejecución de programas, accesos a páginas web, comportamiento de estructuras de datos, gestión de recursos, etc.

Esa incertidumbre es consecuencia de que los fenómenos que estudiamos vienen dados por un alto número de causas, muchas de ellas de pequeño efecto, interdependientes de un modo desconocido, y de comportamiento difícil de explicar o modelizar. De esto se sigue la necesidad de incorporar la incertidumbre al razonamiento, a la deducción, en suma, al método científico. Si pretendemos tener modelos que expliquen la realidad, entonces no podemos ignorar ese aspecto. La Teoría de la Probabilidad es la rama de las Matemáticas que materializa tal incorporación. Podríamos decir que la probabilidad es la lógica de la incertidumbre.

Feller (1906 - 1970), uno de los grandes probabilistas del siglo XX, resaltaba de la probabilidad tres características, que a su juicio, le proporcionan su utilidad y belleza[1]:

- Intuición. La probabilidad es intuitiva porque la usamos en el razonamiento cotidiano. Nos sirve para cuantificar el conocimiento subjetivo que tenemos de un hecho y tomar decisiones.
- Formalismo lógico. La probabilidad es de suma importancia para el método científico. A partir de Kolmogorov, que introduce la definición axiomática de probabilidad, esta se une con la lógica, esto es, con las leyes del pensamiento. Esto permitió que la probabilidad, ahora con el soporte de la lógica, se desarrollase como una rama del conocimiento plenamente independiente. Esta unión de la lógica y la intuición parece que es lo que desconcierta al estudiante en un primer momento.
- Aplicaciones. Son muchas y en los ámbitos más diversos. Nombrar todas sus aplicaciones sería largo, pero, dado que este material está dirigido a alumnos de Informática, merece la pena nombrar algunas de las más relevantes. Sin embargo, dejamos al alumno que las busque él por su cuenta; véase el problema más abajo 2.1.1.

Problema 2.1.1 Buscad en internet aplicaciones de la probabilidad a la Computación. Al menos deberíais encontrar cinco grandes aplicaciones, de relevancia. No traigáis a clase nada que no entendáis.

[1]Cita tomada de su famoso libro *An Introduction to Probability Theory and Its Applications*, cuya primera edición es de 1963.

2.2 Razonamiento probabilístico

El razonamiento probabilístico debe estar presente en la formación de un científico o un técnico por la sencilla razón de que es parte del método científico. ¿Justifica esto que haya que incluir el estudio de la probabilidad? Para contestar a esto querríamos examinar el informe *The Joint Task Force for Computing Curricula 2005*, publicado en 2005 por la ACM americana, la asociación más importante de Informática[2]. En dicho informe daba la siguiente definición de *computación* (nuestra traducción):

> *De modo general, podemos dar el significado de* ***computación*** *a toda actividad que específicamente requiera ordenadores, se beneficie de ellos o los cree. Así pues, la computación incluye: el diseño de sistemas hardware y software para un amplio rango de objetivos; procesamiento, estructuración y gestión de varios tipos de información; la realización de estudios científicos; hacer que los ordenadores se comporten inteligentemente; crear y usar comunicaciones y entretenimiento multimedia; buscar y recopilar información relevante para cualquier objetivo particular, entre otros. La lista es virtualmente interminable y las posibilidades son infinitas. Computación tiene otros significados que son más específicos, basados en el contexto en que se usa el término. Por ejemplo, un especialista en sistemas de información verá el término computación de modo diferente al de un ingeniero de software. Con independencia del contexto, hacer computación de calidad puede ser complicada difícil y complicado. Porque la sociedad necesita gente que haga computación de calidad, concebimos la computación no solamente como una profesión sino como una disciplina científica.*

Dentro de esta definición se encuentran varios conceptos importantes. Resaltaríamos dos: la concepción de la computación como una disciplina científica, la cual, como toda disciplina, requiere de capacidad de abstracción y rigor; y la definición de computación como procesamiento de la información en contextos muy generales y dispares. La definición de disciplina científica incluye el uso del **método científico**, y dentro de este se encuentra el razonamiento probabilístico y estadístico. Respecto al procesamiento, este a su vez descansa en otros dos conceptos relevantes: el de **algoritmo** o procedimiento finito de resolución de un problema y las **estructuras de datos** u organización eficiente de la información implicada en dicho problema (en los próximos temas se ampliarán estos conceptos).

Vamos a desarrollar un ejemplo (informático) que mostrará la utilidad y, a la vez, el carácter del razonamiento probabilístico. Consideremos un problema muy frecuente en informática: la ordenación. Supongamos que tenemos una matriz de n números distintos $M = \{a_1, \cdots, a_n\}$. Un algoritmo muy conocido para ordenar es el llamado *quicksort*. Es este un algoritmo recursivo que funciona del siguiente modo: toma a_1, que llamaremos el pivote, y subdivide la matriz M en dos submatrices $I_1 = \{a_i, \cdots, a_k\}$ y $D_1 = \{a_{k+1}, \cdots, a_j\}$ de tal manera que a_1 es menor que cualquier elemento de I_1 y mayor que cualquiera de los de D_1. El algoritmo, después de esta primera operación, se llama a sí mismo sobre las submatrices I_1 y D_1. Si alguna submatriz tiene tamaño uno o cero, la recursión se para. Es claro que tras la partición de M en I_1 y D_1, el elemento a_1 está ordenado correctamente; véase la figura 2.1 para ver ilustrada esta explicación.

[2]ACM Computing Curricula. The Joint Task Force for Computing Curricula 2005. http://www.acm.org/education/curricvols/CC2005-March06Final.pdf,2005.

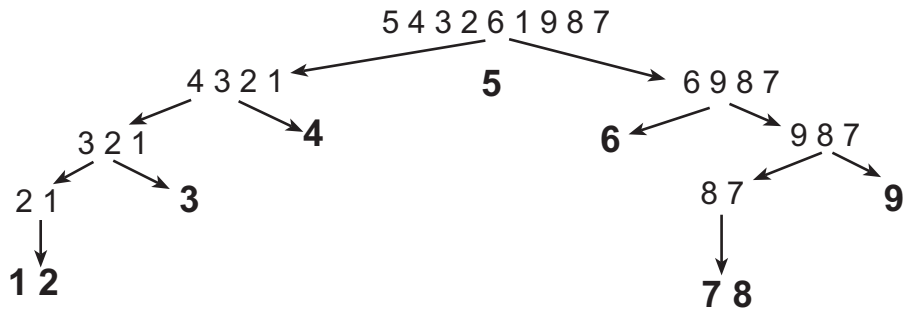

Figura 2.1: Ejecución del algoritmo quicksort

Cabe preguntarse por la velocidad de este algoritmo. La velocidad o tiempo de ejecución de un algoritmo es una de las principales variables a la hora de compararlo con otro algoritmo que resuelva el mismo problema. Otras variables que influyen son la cantidad de memoria, la sencillez conceptual del algoritmo, el uso de estructuras de datos simples, etc. Y aquí aparece la incertidumbre: la velocidad depende del número de datos y del orden en que aparezcan. Ciertamente, la expresión exacta de la velocidad del algoritmo será compleja. Tendrá que llevar la cuenta del número de comparaciones, de intercambios, del paso de parámetros en las llamadas recursivas, etc., y todas estas cantidades varían de una entrada de datos a otra. Por simplificar el análisis, nos detendremos sólo en el número de comparaciones. Llamemos $T(n)$ a ese número, donde n es el número de datos de entrada.

¿Qué ocurre si los números de M están ordenados? En ese caso, la submatriz I_1 está vacía de elementos, ya que no existen elementos menores que el pivote, y D_1 tiene $n-1$ elementos. Queda entonces la siguiente ecuación recursiva:

$$T(n) = O(n-1) + T(n-1),$$

cuya solución es $T(n) \leq cn^2$ ($c > 0$ es una constante que no tiene importancia en este momento.) En este caso, tenemos un algoritmo cuadrático. Démonos cuenta de que este análisis ha considerado sólo la peor situación posible, que los datos vengan ordenados, sin hacer otras consideraciones. Por ejemplo, no revela la frecuencia con que aparece ese caso más desfavorable e ignora la incertidumbre del problema. Este tipo de análisis se llama del **peor caso** y es de tipo determinista.

Cambiemos el enfoque e introduzcamos la incertidumbre y, detrás de ella, el razonamiento probabilístico. El comportamiento del algoritmo dependerá de cómo se hagan las particiones sucesivas de las matrices. Si el pivote deja dos submatrices de aproximadamente el mismo tamaño, el algoritmo irá rápido; si deja una submatriz con muchos elementos y la otra con pocos, el algoritmo será lento. Ahora tendremos en cuenta la elección del pivote y el tiempo de ejecución que resulta de ese pivote. Calculamos como medida de la velocidad del algoritmo el promedio de todos esos tiempos. Este análisis se llama **análisis en media** y es más difícil que el análisis en el peor de los casos visto antes. La ecuación de recurrencia que sale en el análisis en media es (su obtención no es relevante en este momento):

$$T(n) = \frac{1}{n}\left(T(1) + T(n-1) + \sum_{i=1}^{n-1}(T(i) + T(n-i))\right) + O(n),$$

cuya solución es $T(n) = n \log n, c > 0$. Es sorprendente que la velocidad en media sea mucho menor que la velocidad en el peor de los casos.

¿Qué conclusiones podemos sacar de los dos análisis anteriores? En primer lugar, es preferible usar un algoritmo $O(n \log n)$ para ordenar que uno $O(n^2)$ porque es mucho más rápido. En segundo, lugar, podríamos pensar en modificar el algoritmo para que las llamadas recursivas se produjesen sobre matrices del mismo tamaño aproximadamente. Esto, claro, haría más complicado el algoritmo; esta idea ya se ha puesto en práctica a través de los árboles AVL o árboles equilibrados. Pero siendo más finos en el análisis, nos damos cuenta de que el algoritmo en media ya corre en tiempo $O(n \log n)$. En la práctica este algoritmo es de los más usados y se implementa tal cual está descrito arriba. Por ejemplo, compiladores de C o C++ lo traen implementado entre sus funciones básicas.

Este ejemplo muestra cómo funciona el razonamiento probabilístico: un análisis determinista, el del peor de los casos, arrojó unos resultados más bien conservadores, pesimistas diríamos; el análisis probabilístico dio resultados más precisos, si bien fueron más difíciles de obtener.

2.3 Problemas preliminares

Problema 2.3.1 El nombre de la paradoja viene por el nombre del presentador del concurso *Let's make a deal*. En el concurso se presenta al concursante con tres puertas donde detrás de una ellas hay un coche y detrás de las otras dos una cabra. El concursante elige una puerta y en ese momento Monty Hall, el presentador, abre otra que siempre corresponde a la de una cabra. En este momento el presentador ofrece al concursante la posibilidad de cambiar su elección. ¿Debe el concursante mantener su elección original o escoger la otra puerta? ¿Supone alguna diferencia?

Ejercicio 2.3.2 De una urna que contiene 10 bolas distintas, se extraen 3 al azar con reemplazamiento. Determinar la probabilidad de que aparezcan:

(a) Las tres iguales;

(b) Dos iguales;

(c) Las tres distintas.

Ejercicio 2.3.3 De una urna que contiene 10 bolas, 6 son rojas y 4 negras. Se extraen 3 bolas sin reemplazamiento. Determinar la probabilidad de que aparezcan: (a) las tres del mismo color; (b) al menos una de cada color.

Problema 2.3.4 Dos jugadores tiran una moneda no cargada. El que saque cara la primera vez será el ganador. ¿Cuál es la probabilidad de que el primer jugador gana?

2.4 Espacios de probabilidad

Definición 2.4.1 Experimento aleatorio. Un *experimento aleatorio* es cualquier operación que cumple las condiciones:

(1) Antes de realizar el experimento no se sabe cuál va a ser el resultado del mismo.

(2) El conjunto de los resultados posibles sí es conocido a priori.

(3) El experimento deberá ser repetible en idénticas condiciones.

Ejemplo 2.4.2 Por ejemplo, tirar un dado no cargado es un experimento aleatorio. En efecto, conocemos sus posibles resultados, los elementos del conjunto $\{1, \ldots, 6\}$, pero no conocemos el resultado particular de una tirada antes de hacerla; y además podemos tirarlo cuantas veces queramos.

Definición 2.4.3 Espacio muestral. El conjunto de los posibles resultados de un experimento aleatorio se llama *espacio muestral*. Lo designaremos por E. Los elementos de E reciben el nombre de *sucesos elementales*.

Ejemplo 2.4.4 En el caso del dado de más arriba, el espacio muestral es $E = \{1, \ldots, 6\}$. Cada elemento de E es un suceso elemental.

Definición 2.4.5 Espacio de sucesos. El espacio de sucesos es cualquier subconjunto del espacio muestral. Si no se dice nada en contra, el espacio de sucesos por defecto será $\mathcal{P}(E)$

Definición 2.4.6 Sucesos compuestos. Llamaremos *suceso compuesto* a los sucesos no elementales, esto es, a los subconjuntos de dos o más elementos del espacio muestral.

Definición 2.4.7 Suceso seguro y suceso imposible El suceso E se llama *suceso seguro* y el suceso $\emptyset$ se llama *suceso imposible*.

Ejercicio 2.4.8 Siguiendo con el experimento aleatorio de tirar un dado, dad ejemplos de sucesos elementales, sucesos compuestos. Describidlos por enumeración y comprensión.

Problema 2.4.9 Dado que los sucesos son subconjuntos de un conjunto (del espacio muestral E), aplicad la teoría de conjuntos para obtener propiedades de los sucesos.

Definición 2.4.10 Sucesos incompatibles. Dos sucesos $A, B \subseteq E$ se dicen que son *sucesos incompatibles* si $A \cap B = \emptyset$, o dicho de otro modo, no pueden ocurrir a la vez.

Definición 2.4.11 Espacio de probabilidad. Sea una aplicación $P : \mathcal{P}(E) :\longrightarrow \mathbb{R}$ tal que cumple los siguientes tres axiomas:

Axioma 1: Para todo suceso $A \subseteq E$, se tiene que $P(A) \geq 0$.

Axioma 2: $P(E) = 1$.

Axioma 3: Sea $\{A_i \mid i \in I\}$ un conjunto de sucesos disjuntos, esto es, tales que $A_i \cap A_j$ si $i \neq j$. Entonces se tiene que:

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$$

A la terna $(E, \mathcal{P}, P)$ se le llama *espacio de probabilidad*.

Ejercicio 2.4.12 En los problemas de la sección 2.3 identifica el espacio muestral, el espacio de sucesos y la función de probabilidad.

Teorema 2.4.13 *Sea $(E, \mathcal{P}, P)$ un espacio de probabilidad.*

(1) Si A, B son dos sucesos cualesquiera y $A \subseteq B$, entonces $P(A) \leq P(B)$.

(2) La probabilidad es un número entre 0 y 1.

(3) Para todo $A \in \mathcal{P}$, se tiene que $P(A) = 1 - P(\overline{A})$.

(4) Si A, B son dos sucesos cualesquiera, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Nota 2.4.14 La propiedad (4) del teorema anterior se puede generalizar a más sucesos y es una consecuencia inmediata del principio de inclusión-exclusión.

Ejemplo 2.4.15 He aquí varios ejemplos de espacios de probabilidad donde se detalla la terna de un espacio de probabilidad.

(1) Tirar una moneda una vez. Llamaremos $E = \{C, X\}$ al conjunto de resultados posibles de tirar una moneda, donde C es cruz y X cara. El espacio de sucesos es, como fijamos más arriba, $\mathcal{P}(E)$. Si la moneda no está cargada, entonces $P(C) = P(X) = \frac{1}{2}$.

(2) Tirar una moneda dos veces. Ahora el espacio muestral será $E_1 = E \times E$, donde $E = \{C, X\}$. Al usar el producto cartesiano distinguimos el orden en que se tiran las monedas. Un posible espacio de probabilidad podría ser el que asigna estas probabilidades a los sucesos elementales:

$$P((C,C)) = P((C,X)) = P((X,C)) = P((X,X)) = \frac{1}{4}$$

Otro espacio distinto del anterior podría asignar las probabilidades como sigue:

$$P((C,C)) = \frac{1}{5}, P((C,X)) = P((X,C)) = \frac{3}{10}, P((X,X)) = \frac{1}{5}$$

(3) Tirada de dos dados. Se tiran dos dados y se observa la suma de los valores. Es un experimento aleatorio, suponiendo que el dado no está cargado. El espacio muestral es $E = \{2, \ldots, 12\}$. En este ejemplo, en realidad, tenemos la concatenación de dos experimentos aleatorios, donde el segundo experimento, observar la suma de los puntos, depende de los resultados del primer experimento, que es tirar los dados. Para el primer experimento aleatorio su espacio muestral es $E_1 = \{(1,1), (1,2), \ldots, (6,6)\}$. A cada uno de estos sucesos elementales les asignamos $\frac{1}{36}$ (hay otras maneras posibles de hacerlo; esta es la que mejor refleja la idea de que los dados no están cargados). Las probabilidades para el segundo espacio son estas:

$$\begin{array}{lll} P(2)=\frac{1}{36} & P(6)=\frac{5}{36} & P(10)=\frac{3}{36} \\ P(3)=\frac{2}{36} & P(7)=\frac{6}{36} & P(11)=\frac{2}{36} \\ P(4)=\frac{3}{36} & P(8)=\frac{5}{36} & P(12)=\frac{1}{36} \\ P(5)=\frac{4}{36} & P(9)=\frac{4}{36} & \end{array}$$

Nota 2.4.16 Para comprobar que son espacios de probabilidad, hay que verificar que cumplen los axiomas 1-3 dados más arriba. En estos casos basta con observar que las probabilidades son no negativas y que la suma de las probabilidades de los sucesos elementales es uno.

2.4.1 Regla de Laplace

En las aplicaciones de la probabilidad a la computación, hay un caso que aparece con muchísima frecuencia y es de la probabilidad discreta. Aquí por discreta queremos decir que el espacio muestral es o bien finito o bien infinito numerable (el infinito de $\mathbb{N}$). Vamos a estudiar ese caso en detalle.

Supongamos que tenemos un espacio muestral finito $E = \{a_1, \ldots, a_n\}$. Si A es un suceso, entonces A es la unión finita de k sucesos elementales, pongamos $A = \{a_{i_1}\} \cup \ldots \cup \{a_{i_k}\}$. Los sucesos elementales son disjuntos y entonces en virtud del axioma 3 de la probabilidad, tenemos que

$$P(A) = P(\{a_{i_1}\} \cup \ldots \cup \{a_{i_k}\}) = \sum_{j=1}^{k} P(\{a_{i_j}\})$$

En el caso de que todos los sucesos sean *equiprobables*, es decir, que tengan todos la misma probabilidad, la fórmula anterior se transforma en esta:

$$P(A) = \sum_{j=1}^{k} P(\{a_{i_j}\}) = \sum_{j=1}^{k} \frac{1}{n} = \frac{k}{n} = \frac{|A|}{|E|}$$

Esta fórmula se llama **regla de Laplace** y se enuncia con frecuencia, en las condiciones que hemos expuesto aquí, diciendo que la probabilidad de A es el número de casos favorables partido por el número de casos posibles. Los casos favorables se reducen al recuento de los elementos de A, en suma, al cálculo de su cardinal. Aquí es donde entra la combinatoria en juego, como herramienta para conseguir ese cálculo.

2.5 Problemas de probabilidad - I

Ejercicio 2.5.1 Se escoge un entero entre 0 y 100. ¿Cuál es la probabilidad de que no sea ni divisible por 2 ni por 5?