Capítulo 1

Introducción a la Teoría de Probabilidad

Para la mayoría de la gente, "probabilidad" es un término vago utilizado en el lenguaje cotidiano para indicar la posibilidad de ocurrencia de un evento futuro. Así, hablamos de la probabilidad de que llueva mañana, la probabilidad de que suba el dolar, etc. En ciencia el concepto de probabilidad se utiliza para hacer *inferencias*, esto es, predicciones acerca de ciertos procesos cuyo resultado es variable cada vez que se repiten. Así, por ejemplo, no se puede predecir con exactitud cual será la presión arterial de una persona en un momento dado, ni si la moneda arrojada caerá de un lado o del otro. Este tipo de procesos, dan un resultado diferente dentro de un cierto conjunto de valores cada vez que se repiten, pero la frecuencia relativa con que ocurre cada valor en una gran serie de observaciones es a menudo estable. Si arrojamos un número grande de veces la misma moneda (suponiendo que no este cargada) observaremos que aproximadamente la mitad de las veces sale cara y la otra mitad seca. Procesos que presentan estas propiedades se denominan **aleatórios** o **estocásticos**. Dicha frecuencia relativa estable nos dá una medida intuitiva (pero significativa) de la posibilidad de ocurrencia de un evento aleatorio en una observación futura y constituye la definición **estadística** de probabilidad. Supongamos que el resultado de un experimento esta descripto por una variable X, que puede tomar valores de un conjunto finito $x_1, x_2, \ldots, x_n$. Supongamos que repetimos el experimento N veces y sea N_i el número de veces que la variable tomo el valor x_i, con $i = 1, \ldots, n$. Podemos entonces definir la probabilidad P_i de observar el valor x_i como:

$$P_i = \lim_{N \to \infty} \frac{N_i}{N} \tag{1.1}$$

Dado que $\sum_{i=1}^{n} N_i = N$, tenemos que $0 \leq P_i \leq 1$ para $i = 1, \ldots, N$ y $\sum_{i=1}^{n} P_i = 1$.

Notemos que la Ec.(1.1) es una definición *a posteriori*, esto es, para conocer la probabilidad de ocurrencia de un resultado en un experimento, primero debemos realizar el experimento un número grande de veces. Por otra parte, no tenemos manera de probar que este límite realmente existe y es independiente del conjunto de experimentos, lo cual debe ser asumido como un axioma. Vamos entonces a buscar una formulación que nos permita estimar probabilidades *a priori*, a partir de propiedades intrínsecas del sistema. Para ello vamos a plantear un conjunto diferente de axiomas. Sin embargo, el concepto anterior nos resultará útil como guía para una definición mas formal, ya que vamos a exigir que ambas definiciones coincidan en la circunstancias apropiadas. Supongamos entonces que un posible resultado de un experimento es A y que por algún método le asignamos una probabilidad de ocurrencia $P(A)$. Entonces si llevamos a cabo N experimentos idénticos, esperamos que el evento A ocurra aproximadamente un número $N\,P(A)$ de veces. De hecho, vamos a exigir que en el límite de N muy grande la fracción relativa de veces que ocurre A converja a $P(A)$.

1.1. Definiciones básicas y axiomas

Vamos a comenzar definiendo el **espacio muestral** de un experimento como un conjunto S, tal que cualquier resultado del experimento corresponda a uno o mas elementos de S. Un **evento** es cualquier subconjunto de S. Un **evento simple** es un elemento del conjunto. Dos eventos A y B se dicen **mutuamente excluyentes** si $A \cap B = \emptyset$; eventos simples son siempre mutuamente excluyentes.

La **probabilidad** de un evento $A \supset S$ se define como una función $P : S \to [0, 1]$ con las siguientes propiedades:

1. **Axioma 1**: $P(S) = 1$

2. **Axioma 2**: Para cualquier secuencia (finita o infinita) de eventos E_i mutuamente excluyentes entre sí, esto es, $E_i \cap E_j = \emptyset$ para todo par $i \neq j$, entonces

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

Las siguientes propiedades se deducen de manera inmediata.

- Si tomamos como E_i todos los eventos simples (esto es, todos los elementos de S) sabemos que $S = \bigcup_i E_i$ y del axioma 2 se sigue que $P(S) = \sum_i P(E_i)$. Del axioma 1 tenemos entonces que

$$\sum_i P(E_i) = 1$$

  Esta propiedad es válida en general para cualquier secuencia de eventos (subconjuntos) mutuamente excluyentes y exhaustiva, esto es, una secuencia tal que $S = \cup_i E_i$. Un caso particular de lo anterior es el siguiente. Sea E un evento cualquiera y sea $\overline{E}$ su complemento, esto es, $S = E \cup \overline{E}$, con $E \cap \overline{E} = \emptyset$. Entonces se sigue que $P(\overline{E}) = 1 - P(E)$.

- Dado que el conjunto vacío es el complemento de S, del resultado anterior entonces $P(S) + P(\emptyset) = 1$. Por el axioma 1 se sigue que $P(\emptyset) = 0$.

Las propiedades anteriores tienen una interpretación bastante directa. $P = 1$ implica una certeza absoluta en el resultado del experimento. Un resultado del experimento con probabilidad uno es determinista. Así, la probabilidad de que *algo* ocurra $P(S) = 1$. Por otra parte, un resultado con probabilidad cero implica que el mismo no ocurrira nunca. Así, la probabilidad de que no ocurra *nada* es $P(\emptyset) = 0$, ya que el experimento siempre arroja algun resultado.

El axioma 2 nos da una regla de aditividad para las probabilidades. La probabilidad de eventos que no pueden ocurrir simultaneamente es la suma de las probabilidades de los eventos correspondientes.

Dados dos eventos A y B, $P(A \cup B)$ nos da la probabilidad de que ocurra A, B o ambos a la vez. Por otra parte, $P(A \cap B)$ es la probailidad de ocurran ambos eventos simultaneamente y solo eso. Esta última se denomina **probabilidad conjunta** de A y B. Estas probabilidades estan relacionadas a través de

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.2}$$

Esta propiedad puede demostrarse facilmente mediante el diagrama de Venn de la Fig.1.1. El conjunto $P(A \cup B)$ puede dividirse en tres conjuntos disjuntos: I, II y III (eventos mutuamente

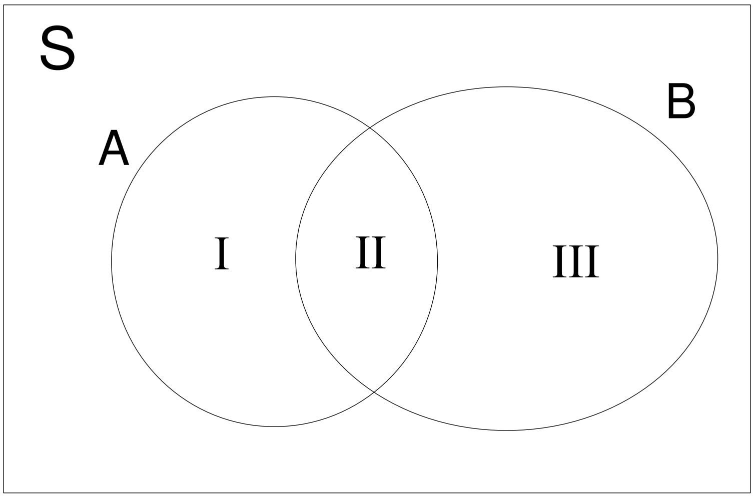

Figura 1.1: Diagrama de Venn para la unión de dos eventos.

excluyentes). En otras palabras, I es el conjunto de todos los elementos en A que no estan en B, III lo inverso y II el conjunto de todos los elementos que pertenecen simultáneamente a A y B. Así $II = A \cap B$, $A = I \cup II$ y $B = III \cup II$. Además

$$A \cup B = I \cup II \cup III$$

Dado que I, II y III son mutuamente excluyentes

$$P(A \cup B) = P(I) + P(II) + P(III)$$

$$P(A) = P(I) + P(II)$$

$$P(B) = P(II) + P(III)$$

de donde se obtiene la Eq.(1.2).

En muchos experimentos resulta natural asumir que todos los resultados posibles (eventos simples) tienen la misma probabilidad de ocurrencia. Así, por ejemplo, si nuestro experimento es arrojar una moneda, el espacio muestral será $S = \{cara, seca\}$. Si la moneda no esta cargada, es de esperar que ambos eventos simples tengan la misma probabilidad. Otro ejemplo puede ser al resultado de arrojar un dado de 6 caras. El espacio muestral consiste de 6 eventos simples $S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$, donde E_i corresponde al resultado i. Para un dado no cargado todos estos resultados deberían tener la misma probabilidad. Consideremos entonces un caso genérico de N posibles resultados E_i con $i = 1, \ldots, N$. Dado que

$$\sum_{i=1}^{N} P(E_i) = 1$$

se sigue inmediatamente que

$$P(E_i) = \frac{1}{N}$$

Así, en el ejemplo de la moneda $P(E_i) = 1/2$ y para el ejmplo del dado $P(E_i) = 1/6$. De esto se sigue que, en un espacio muestral finito, para cualquier evento E

$$P(E) = \frac{\text{número de elementos en E}}{\text{número de elementos en S}} \tag{1.3}$$

En otras palabras, si asumimos que todos los resultados de un experimento son igualmente probables, la probabilidad de cualquier evento E es igual a la fracción de elementos en el espacio muestral contenidos en E. Por ejemplo, para el caso del dado, la probabilidad de que el resultado sea menor que 3 será $P(E) = 2/6 = 1/3$.

Un concepto de gran utilidad es la **probabilidad condicional** $P(A|B)$, la cual nos da la probabilidad de que ocurra el evento A *dado que tenemos la certeza de que ocurre* B. $P(A|B)$ se define por la ecuación

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.4}$$

Podemos entender esta ecuación si consideramos el ejemplo simple de un espacio muestral con N eventos equiprobables. Sea N_A, N_B y N_{AB} el número de elementos contenidos en A, B y $A \cap B$ respectivamente. Entonces, $P(A) = N_A/N$, $P(B) = N_B/N$ y $P(A \cap B) = N_{AB}/N$. Dado que tenemos la certeza de que ocurre B, para esta situación particular B es nuestro espacio muestral. Dado que todos los eventos en B a su vez son equiprobables, tenemos que $P(A|B) = N_{AB}/N_B$. Dividiendo numerador y denominador de la ecuación anterior por N se obtiene la Ec.(1.4). Dado que $P(A \cap B) = P(B \cap A)$, tenemos también que

$$P(A|B)\, P(B) = P(B|A)\, P(A) \tag{1.5}$$

El concepto de probabilidad condicional nos permite introducir el concepto de **eventos independientes**. Dos eventos A y B son independientes *si y solo si*:

$$P(A|B) = P(A) \tag{1.6}$$

De la Ec.(1.5) esto implica inmediatamente que $P(B|A) = P(B)$. Por otra parte, de la Ec.(1.4) tenemos que

$$P(A \cap B) = P(A)\; P(B) \tag{1.7}$$

esto es, si dos eventos son independientes, su probabilidad conjunta es el producto de las probabilidades individuales.

1.2. Cálculo de probabilidades en espacios finitos: elementos de análisis combinatorio

Consideremos el siguiente ejemplo. Se lanza una moneda perfecta tres veces. Cual es la probabilidad de obtener cara en dos de los tres lanzamientos? Un evento simple en este experimento consiste en un conjunto ordenado de tres lementos binarios C ó S (cara o seca), tal como CSS (cara en la primera tirada y seca en las dos siguientes). Dado que la moneda es perfecta, podemos asumir que los eventos simples son equiprobables. Así, el primer paso para resolver el problema es determinar el número total de puntos muestrales (eventos simples) y el segundo paso determinar el número de puntos muestrales que contiene dos C. La manera mas simple es enumerar todos los puntos muestrales:

$$CCC \quad CCS \quad CSC \quad SCC$$

$$CSS \quad SCS \quad SSC \quad SSS$$

Tenemos 8 puntos muestrales y por lo tanto la probabilidad de cada uno de ellos será $1/8$; de todos los puntos muestrales en tres hay dos caras y por lo tanto la probabilidad buscada es $3/8$.

Es evidente que esta técnica tan simple solo es util cuando el número posible de puntos muestrales es pequeño. En la mayoría de los casos de interés, los espacios muestrales son enormes y enumerar todos los puntos muestrales resulta imposible. Sin embargo, para calcular probabilidades de eventos solo necesitamos *contar* cuantos puntos muestrales tenemos y para ello no es necesario enumerarlos a todos. Esto se puede ver en el siguiente ejemplo.

Supongamos una reunión de 20 personas, en principio sin ninguna relación directa entre ellos. Cual es la probabilidad de que al menos 2 de ellas cumplan años el mismo día?

Para empezar a resolver este problema, podemos en primer lugar invertir la pregunta. Si consideramos el evento "todas las 20 personas cumplen años en días diferentes"es evidente que es el complemento del evento "al menos 2 de ellas cumplen años el mismo día"; si llamamos P_1 y P_2 a las probabilidaes de ambos eventos tenemos que $P_1 + P_2 = 1$, o $P_2 = 1 - P_1$. Calculemos entonces P_1 (es mas facil que calcular directamente P_2).

Supongamos que numeramos los días del año de 1 a 365 (ignoramos por simplicidad los años bisiestos). Un punto muestral de este experimento consiste en un conjunto de 20 números (cada uno entre 1 y 365) donde el primero representa el cumpleaños de la primera persona, el segundo número el de la segunda persona, etc.. El número de puntos muestrales N_s será entonces el número de 20-adas diferentes que podemos formar con 365 números. Para cada persona tenemos 365 posibilidades diferentes. Dado que los números (fechas) pueden repetirse para diferentes personas $N_s = 365^{20}$. Calculemos ahora el número de puntos muestrales N_1 en los cuales no se repite ningún número entre los 20. Para la primera persona tenemos 365 posibilidades. Por cada una de ellas, tenemos solo 364 para la segunda persona (todas los números posibles, menos el que se selecciono antes). Para la tercera persona tenemos 363, etc. Así

$$N_1 = 365 \times 364 \times 363 \times \cdots \times 346$$

Así

$$P_1 = \frac{N_1}{N_s} = \frac{365 \times 364 \times 363 \times \cdots \times 346}{365^{20}} \approx 0{,}59$$

y por lo tanto $P_2 \approx 0{,}41$. Existe un 41 % de probabilidad de que en cualquier reunión de 20 personas al menos 2 cumplan años el mismo día. En un grupo de 50 personas esta probabilidad asciende a un 97 %.

Como ilustra el ejemplo anterior, al tratar con grandes cantidades de objetos a menudo tenemos que calcular el numero de diferentes maneras de combinar los mismos. Este tipo de cálculo entra dentro del análisis combinatorio, del cual veremos algunas nociones básicas.

Una **permutación** es cualquier arreglo de un conjunto de N objetos diferentes entre sí en un orden definido. Si queremos contar el número total de posibles permutaciones de de N objetos, podemos pensar que tenemos N casilleros ordenados y distribuímos los objetos uno en cada casillero. Para el primer casillero tenemos N posibilidades. Por cada elección para el primer casillero procedemos a llenar los siguientes con los que quedan. Así, para el segundo tenemos $N - 1$ posibilidades por cada una de las N del primero. A su vez, por cada una de las $N - 1$ posibilidades del segundo tenemos $N - 2$ para el tercero y así sucesivamente. El número total de posibles permutaciones de N objetos es por lo tanto:

$$N \times (N - 1) \times (N - 2) \times \cdots \times 2 \times 1 = N!$$

Supongamos ahora que queremos calcular el número de permutaciones diferentes de R objetos tomados entre un conjunto de $N > R$ (como en el ejemplo de los cumpleaños). Este número se denomina P^N_R: **permutaciones de N objetos tomados de a R**. El procedimiento para obtener

P_R^N es semejante al anterior. En este caso, podemos pensar que tenemos R casilleros ordenados y N objetos diferentes para llenarlos. Al igual que antes, tenemos N posibilidades para llenar el primero, $N-1$ para el segundo y así hasta completar los R. Por lo tanto

$$P_R^N = N \times (N-1) \times (N-2) \times \cdots \times (N-R+1) = \frac{N!}{(N-R)!}$$

Así, en el ejemplo de los cumpleaños $N_1 = P_{20}^{365} = 365!/(365-20)!$.

Una **combinación** es una selección de R objetos tomados entre un conjunto de N, pero sin importar *el orden* de los objetos. El número de posibles combinaciones de N objetos tomados de a R se denota C_R^N. Podemos obtener C_R^N a partir de P_R^N. En P_R^N cada selección de un conjunto particular de R objetos aparece repetida a través de todas las permutaciones posibles de los mismos, esto es, $R!$. Tenemos entonces que $P_R^N = R!\, C_R^N$ y por lo tanto

$$C_R^N \equiv \binom{N}{R} = \frac{P_R^N}{R!} = \frac{N!}{R!\,(N-R)!}$$

C_R^N también se suele llamar número combinatorio o coeficiente binomial, ya que es el coeficiente que aparece en el desarrollo del binomio de Newton:

$$(x+y)^N = \sum_{k=0}^{N} x^k\, y^{N-k} \binom{N}{k} \tag{1.8}$$

Algunas propiedades inmediatas de los coeficientes binomiales son

$$\binom{N}{0} = \binom{N}{N} = 1$$

$$\binom{N}{1} = N$$

En el ejemplo de las tres monedas, el número de eventos con dos caras es $C_2^3 = 3$.

El número de permutaciones de N objetos, en los cuales hay n_1 idénticos (por lo tanto no importa el orden) de un tipo, n_2 de otro tipo, …, n_p elementos idénticos del tipo p, viene dado por el **coeficiente multinomial**:

$$\binom{N}{n_1,\ \ldots,\ n_p} = \frac{N!}{n_1!\ \ldots\ n_p!}$$

dado que los mismo aparecen en la fórmula multinomial

$$(x_1 + \cdots + x_p)^N = \sum x_1^{n_1}\, x_2^{n_2} \cdots\, x_p^{n_p} \binom{N}{n_1,\ \ldots,\ n_p}$$

donde la suma comprende todos los valores de $n_1, \ldots, n_p$ entre cero y N, tal que $\sum_{i=1}^{p} n_i = N$.

1.2.1. Ejercicios

1. Una baraja francesa consta de 52 cartas, dividida en cuatro grupos (palos) con 13 figuras (números) cada grupo. En el juego de poker se reparten 5 cartas elegidas al azar.

    a) Calcule la probabilidad de obtener un poker: cuatro cartas del mismo número.

b) Calcule la probabilidad de obtener un full: tres cartas del mismo número y las dos restantes iguales a otro número.

c) Calcule la probabilidad de obtener un par simple: dos cartas del mismo número y las tres restantes con números diferentes entre sí y diferentes a los anteriores.

d) Calcule la probabilidad de obtener un par doble: dos pares de cartas con números iguales en cada par pero diferentes entre pares y una quinta diferente de los dos grupos anteriores.

2. Se arroja una moneda 3 veces. Sea E_i el evento "sale cara en la tirada i-ésima" $(i = 1, 2, 3)$. Muestre que los tres eventos son independientes entre sí.