Teoría Estadística Elemental I
Teoría (resumida) del 1^{er} Tema

Raúl Jiménez
Universidad Carlos III de Madrid

Septiembre 2011

Muchos de los eventos que estamos acostumbrados a observar no pueden ser predeterminados. Por ejemplo, ¿cuánto variará el euro respecto al dólar de hoy a una semana? ¿Cuánto lloverá durante el próximo mes? El escenario dispuesto para observar lo que está por ocurrir se denomina **experimento aleatorio**. Los juegos de azar nos brindan ejemplos clásicos de experimentos aleatorios. Aunque los objetos que estudiemos con la teoría de probabilidades estén siempre asociados a un determinado experimento aleatorio, los presentamos en un contexto matemático muy general y útil para la modelación de cualquier escenario.

Espacios de probabilidad

El conjunto de todos los posibles resultados de un experimento aleatorio es llamado **espacio muestral** y comúnmente denotado por la letra Ω. Otros conjuntos de interés de posibles resultados son llamados **eventos** y denotados por letras mayúsculas, generalmente las primeras del abecedario.

A lo largo de estas notas se hace uso intensivo de operaciones con conjuntos, es por ello que conviene recordar algunos conceptos básicos, tales como:

- Conjunto vacío.
- Conjunto numerable, infinito numerable y no numerable.
- Unión, intersección y diferencia de conjuntos.
- Complemento y partición de un conjunto.
- Diagramas de Venn.
- Leyes distributivas y leyes de Morgan.

Dado un experimento aleatorio, la clase $\mathcal{F}$ de todos los eventos o conjuntos de interés debe tener ciertas propiedades (razonables):

(i) El espacio muestral es un conjunto de interés,

$$\Omega \in \mathcal{F}.$$

(ii) Si un conjunto es de interés su complemento también lo es,

$$\text{si} \quad A \in \mathcal{F} \quad \text{entonces} \quad A^c \in \mathcal{F}.$$

(III) La unión de una colección contable de eventos es un evento de interés,

$$\text{si} \quad A_1, A_2, \ldots \quad \text{son eventos de } \mathcal{F} \quad \text{entonces} \cup_{n\geq 1} A_n \in \mathcal{F}.$$

Una clase de eventos que satisface las tres propiedades anteriores se denomina **σ-álgebra**. Es fácil comprobar que si $\mathcal{F}$ es una σ-álgebra entonces cumple propiedades tales como:

(I') $\emptyset \in \mathcal{F}$.

(II') Si $A, B \in \mathcal{F}$ entonces $A - B \in \mathcal{F}$.

(III') Si $A_1, A_2, \ldots$ son eventos de $\mathcal{F}$ entonces $\cap_{n\geq 1} A_n \in \mathcal{F}$.

Aún más general, se puede demostrar que $\mathcal{F}$ **es cerrada bajo operaciones numerables de conjuntos.**

Uno de nuestros objetivos es medir el chance de que eventos asociados a un experimento aleatorio ocurran: ¿cuál es el chance de que llueva más este otoño que el pasado? ¿Cuál es el chance de que el euro retroceda ante el dólar? ¿Cuál es el chance de ganar un juego de póker?

Una **medida de probabilidad** es una función que asigna a cada evento el chance o probabilidad que tiene de ocurrir al observar un experimento aleatorio. Si asignamos a los eventos que no tienen chance de ocurrir probabilidad 0 y a los eventos que tienen chance seguro de ocurrir probabilidad 1, entonces una medida de probabilidad es una función $P : \mathcal{F} \to [0, 1]$ que debe satisfacer las siguientes propiedades:

$$P(\Omega) = 1, \tag{1}$$

Si $A_1, A_2, \ldots$ son eventos disjuntos de $\mathcal{F}$, es decir si $A_i \cap A_j = \emptyset$ para todo $i \neq j$, entonces

$$P(\cup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n) \tag{2}$$

Esta última propiedad es conocida como **σ-aditividad** y es natural exigírsela a casi cualquier medida: área, volúmen, etc. La idea subyacente es que toda medida debe permitir medir por partes.

A partir de (1) y (2) las siguientes propiedades de las medidas de probabilidad pueden (y deben) ser demostradas todas de manera directa:

P1. $P(\emptyset) = 0$

P2. Aditividad: si $A_1, A_2, \ldots, A_n$ son eventos disjuntos, entonces

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

P3. $P(A^c) = 1 - P(A)$

P4. $P(B - A) = P(B) - P(B \cap A)$

P5. Si $A \subset B$ entonces $P(B - A) = P(B) - P(A)$

P6. Monotonía: si $A \subset B$ entonces $P(A) \leq P(B)$

P7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

P8. Subaditividad: $P(\cup_{n\geq 1} A_n) \leq \sum_{i=1}^{n} P(A_i)$

Otras propiedades que se demuestran con un poco más de trabajo (el profesor puede escoger un par de ellas, recomendamos P10 y P11) son:

P9. Fórmula de inclusión exclusión:

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) \\ &+ \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) \\ &- \cdots + (-1)^{n+1} P(\cap_{i=1}^{n} A_i) \end{aligned}$$

Note que el caso $n = 2$ corresponde a P7. El caso $n = 3$ se requiere para resolver varios ejercicios.

P10. σ-subaditividad: para cualquier sucesión de eventos, no necesariamente disjuntos,

$$P(\cup_{n\geq 1} A_n) \leq \sum_{n\geq 1} P(A_n)$$

P11. Continuidad por la izquierda: si $A_1, A_2, \ldots$ es una sucesión creciente de eventos, es decir, para cualquier n se verifica que $A_n \subset A_{n+1}$, entonces

$$P(\cup_{n\geq 1} A_n) = \lim_{n} P(A_n)$$

P12. Continuidad por la derecha: si $A_1, A_2, \ldots$ es una sucesión decreciente de eventos, es decir, para cualquier n se verifica que $A_{n+1} \subset A_n$, entonces

$$P(\cap_{n\geq 1} A_n) = \lim_{n} P(A_n)$$

Dado un espacio muestral Ω, una σ-álgebra $\mathcal{F}$ de subconjuntos de Ω y una medida de probabilidad $P : \mathcal{F} \to [0, 1]$, la terna $(\Omega, \mathcal{F}, P)$ es llamada **espacio de probabilidad**.

Probabilidad condicional

Información adicional, no inicialmente contemplada, de un experimento puede modificar el escenario de tal forma que la probabilidad que le hayamos asignado a un evento cambie. Por ejemplo, la probabilidad que le hayamos dado a que el euro se revalorizará frente al dólar durante la próxima semana cambiará si sabemos que acaba de ocurrir una caída importante en Wall Street. En general, consideremos que A y B son eventos que ocurren con probabilidad $P(A)$ y $P(B)$. Si sabemos que B ha ocurrido, la probabilidad de que A ocurra no tiene por que seguir siendo $P(A)$, ya que A ocurrirá sí y sólo sí $A \cap B$ ocurre. Lo anterior sugiere que, dado que B ocurre, la probabilidad de A es proporcional a $P(A \cap B)$. Ya que, dado que B ocurre, B en un evento seguro, la constante de proporcionalidad a la que hacemos referencia debe ser $1/P(B)$. La siguiente definición pone orden al trabalenguas anterior.

Definición (Probabilidad Condicional). Sean A y B eventos con $P(B) > 0$, entonces la probabilidad condicional de A dado B se denota por $P(A|B)$ y se define por

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

Para cada evento A, $P(A|B)$ es un número positivo, es decir, la probabilidad condicional establece un correspondencia entre los eventos y los números reales positivos. Más específicamente, la probabilidad condicional es una medida de probabilidad.

Proposición 1. *Sea B un evento con $P(B) > 0$, entonces*

(i) *Para todo evento A, $0 \leq P(A|B) \leq 1$*

(ii) $P(\Omega|B) = 1$

(iii) *Si $A_1, A_2, \ldots$ son eventos disjuntos entonces*

$$P(\cup_{n\geq 1} A_n|B) = \sum_{n\geq 1} P(A_n|B)$$

Por la proposición anterior, todas las propiedades que satisfacen las medidas probabilidad también las satisface la probabilidad condicional. Por ejemplo, la probabilidad condicional es monótona, subaditiva, continua por la derecha y por la izquierda.

La probabilidad condicional brinda una importante fórmula para el cálculo de probabilidades, cuando se tiene una partición apropiada del espacio muestral. Una **partición de un conjunto** A es una sucesión de eventos disjuntos $B_1, B_2, \ldots$ cuya unión sea A.

Fórmula de probabilidad total. Sea $B_1, B_2, \ldots$ una partición del espacio muestral, Supongamos que $P(B_i) > 0$ para $i \geq 1$. Entonces, para cualquier evento A,

$$P(A) = \sum_{i\geq 1} P(A|B_i)P(B_i). \tag{3}$$

La aplicación de esta fórmula se basa en la apropiada escogencia de la partición, de manera que $P(A|B_i)$ sea sencillo de calcular. Comúnmente esta fórmula simplifica engorrosos cálculos.

Ejemplo. Se tienen dos cajas. La primera tiene b_1 bolas blancas y r_1 rojas. La segunda caja tiene b_2 bolas blancas y r_2 rojas. Si se pasa una bola al azar de la primera caja a la segunda y luego se extrae un bola al azar de la segunda caja. Use la fórmula de probabilidad total para calcular la probabilidad de extraer una bola blanca de la segunda caja.

Son comunes las situaciones en las que se tiene conocimiento preciso, o al menos información estadística, acerca de $P(A|B)$ cuando en realidad se requiere conocer $P(B|A)$. La siguiente es una sencilla y poderosa fórmula, que relaciona ambas probabilidades.

Fórmula de Bayes. Sean A y B eventos con probabilidad no nula, entonces

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{4}$$

Ejemplo. Continuando con el ejemplo anterior, use la fórmula de Bayes para calcular la probabilidad de haber pasado una bola roja de la primera caja a la segunda caja cuando la que se extrajo de la segunda caja fue blanca.

Otra fórmula de mucha utilidad para cálculo de probabilidades, cuando se consideran **experimentos secuenciales**, comúnmente modelizados con **árboles de decisión**, es la llamada fórmula de multiplicación:

Fórmula de multiplicación. Sean $A_1, A_2, \ldots, A_n$ eventos con probabilidad no nula. Entonces, para $n \geq 2$,

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)\cdots P(A_n|\cap_{i=1}^{n-1} A_i) \tag{5}$$

Independencia

La noción de independencia en teoría de probabilidades está tomada de su significado cotidiano. En general, decimos que un par es independiente cuando el resultado de las acciones de uno no afecta en el resultado de las acciones del otro. En términos probabilísticos, diremos que dos eventos son independientes si la ocurrencia de uno de ellos no afecta la probabilidad de ocurrecia del otro. Es decir, A es independiente de B si

$$P(A|B) = P(A)$$

Para que la ecuación anterior esté bien definida, es necesario que $P(B) > 0$, en cuyo caso, podemos reescribir la ecuación como

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

De esta última ecuación podemos observar que:

- La independencia es recíproca, esto es, si A es independiente de B entonces B es independiente de A.
- La condición $P(B)$ o $P(A) > 0$ no es requerida.

Ahora estamos en capacidad de definir formalmente la independencia e interpretarla.

Independencia de dos eventos. Decimos que el par de eventos A, B son independientes respecto a P si

$$P(A \cap B) = P(A)P(B) \tag{6}$$

¿Cómo generalizar la noción de independencia de una par de eventos a una familia? Pues igual que en el sentido cotidiano: para que una familia sea independiente cualquier subgrupo debe serlo, no basta que sean independientes por pares o que lo sea un subgrupo en particular.

Independencia de una Familia de Eventos. Decimos que la familia de eventos $\{A_i, i \in I\}$ es independiente si para cualquier $J \subset I$

$$P(\cap_{i \in J} A_i) = \Pi_{i \in J} P(A_i) \tag{7}$$

Ejemplo. Considere $\Omega = \{1, 2, 3, 4\}$ y $P(\{\omega\}) = 1/4$ para todo $\omega \in \Omega$. Sean $A = \{1, 2\}$, $B = \{1, 3\}$ y $C = \{1, 4\}$. Note que la probabilidad de cada uno de estos eventos es $1/2$ y por tanto cada par de eventos son independientes. Por ejemplo, A y B son independientes ya que

$$P(A \cap B) = 1/4 = P(A)P(B).$$

Sin embargo, $P(A \cap B \cap C) = 1/4 \neq P(A)P(B)P(C)$, y por tanto A, B y C no son independientes.

Para determinar la no independencia (dependencia) de una familia de eventos basta verificar que la ecuación (7) no se cumple para un subgrupo particular (para algún J). Sin embargo, la independencia de una colección de eventos puede ser una propiedad dura de comprobar. Por ejemplo, para verificar por definición la independencia de apenas 10 eventos habría que verificar más de 1000 ecuaciones! Afortunadamente, consideraremos muchos casos en que la independencia de una familia de eventos es una consecuencia directa de la manera en que son observados. El caso que queremos destacar trata de eventos asociados a **repeticiones independientes de experimentos aleatorios**, tales como lanzamientos sucesivos de un dado o una moneda. Si se tienen n experimentos independientes, en el sentido de que los resultados de unos no afectan los resultados de los otros, y $A_1, A_2, \ldots$ son eventos asociados al primer experimento, al segundo, etc., entonces $A_1, A_2, \ldots$ son independientes.

Los siguientes dos resultados conciernen con sucesiones de eventos asociados a experimentos independientes.

Proposición 2. *Sea A es un evento con probabilidad no nula, asociado a un determinado experimento aleatorio. Si repetimos el experimento infinitas veces, entonces A ocurre alguna vez con probabilidad 1.*

Para demostrar este resultado aplicamos varias propiedades que hemos aprendido. Llamemos A_n al evento A ocurre en el n-ésimo experimento y $p = P(A_n)$. Usando P2, las leyes de Morgan, P12 y la independencia de $A_1, A_2 \ldots, A_m$, obtenemos

$$\begin{aligned}
P(A \text{ ocurre alguna vez}) &= P(\cup_{n\geq 1} A_n) \\
&= 1 - P\left([\cup_{n\geq 1} A_n]^c\right) \\
&= 1 - P(\cap_{n\geq 1} A_n^c) \\
&= 1 - \lim_{m\to\infty} P(\cap_{n\geq 1}^m A_n^c) \\
&= 1 - \lim_{m\to\infty} (1-p)^m = 1
\end{aligned}$$

Proposición 3. *Sean A y B son eventos mutuamente excluyentes, asociados a un determinado experimento, con probabilidad no nula. Entonces, si repetimos el experimento infinitas veces, A ocurre antes que B con probabilidad*

$$P(A \textit{ ocurra antes que } B) = \frac{P(A)}{P(A) + P(B)}.$$

Para probar esta proposición observemos que

$$\begin{aligned}
P(A \text{ ocurra antes que } B) &= \sum_{k\geq 1} P(A \text{ ocurre antes que } B \text{ en el experimento } k) \\
&= \sum_{k\geq 0} [P(\text{ni } A \text{ ni B ocurren})]^k P(A) \\
&= \frac{P(A)}{1 - P(\text{ni } A \text{ ni B ocurren})} \\
&= P(A)\frac{1}{P(A\cup B)} = \frac{P(A)}{P(A)+P(B)}.
\end{aligned}$$

Una elegante aplicación de la conjunción de este resultado con la fórmula de probabilidad total, que sugerimos que o bien el profesor o bien el estudiante demuestre, determina que la probabilidad de ganar en el juego de dados es

$$\frac{8}{36} + 2\left(\frac{3}{36}\frac{3}{3+6} + \frac{4}{36}\frac{4}{4+6} + \frac{5}{36}\frac{5}{5+6}\right) = \frac{244}{495} = 0{,}493.$$

En el juego tiras los dados en una primera ronda. Si sale 7 o 11 ganas. Si sale 2, 3 o 12 pierdes. Si tiras 4, 5, 6, 8, 9 o 10 hay que seguir lanzando hasta que o bien repitas el número que lanzaste en la primera ronda o bien salga un 7. En el primer caso ganas, en el segundo pierdes.

Espacios equiprobables

En muchos experimentos aleatorios; por ejemplo, en la mayoría de los juegos de azar; el cálculo de probabilidades puede reducirse a contar el número de elementos de un conjunto.

Denotemos por $|A|$ el número de elementos o **cardinal** del conjunto A. Si Ω es finito y todos los resultados del experimento tienen igual probabilidad de ocurrencia decimos que el espacio es **equiprobable**. En ese caso, la probabilidad de un resultado cualquiera del experimento debe ser $1/|\Omega|$, ya que $P(\Omega) = 1$. Así, la probabilidad de un evento A de un espacio equiprobable es

$$P(A) = |A|/|\Omega|.$$

A continuación, vamos a presentar dos esquemas elementales de conteo.

Variaciones y Permutaciones. Sean E y F dos conjuntos finitos. Supongamos sin pérdida de generalidad que $E = \{1, 2, \dots, p\}$ y $F = \{1, 2, \dots, n\}$. Denotemos por I_n^p el número de funciones inyectivas que van de E a F. Claramente, si $p > n$ entonces $I_n^p = 0$. Si $p \leq n$, podemos construir una función inyectiva $f : E \to F$ usando el siguiente esquema recursivo:

Empezamos seleccionando $f(1)$ entre los n elementos pertenecientes a F. Una vez escogido $f(1)$, existe $n - 1$ posibles escogencias para $f(2)$, ya que $f(2)$ debe diferir de $f(1)$ para que f sea inyectiva. Siguiendo este procedimiento, $f(i)$ puede ser escogido entre los $n - (i - 1)$ elementos $F - \{f(1), \dots, f(i-1)\}$. En total, tenemos $n(n-1)\dots(n-p+1)$ posibilidades para construir f.

En resumen, si $p \leq n$, el número de inyecciones de E a F es

$$I_n^p = n(n-1)\dots(n-p+1) = \frac{n!}{(n-p)!},$$

siendo $n!$ el factorial de n, definido por

$$n! = 1 \cdot 2 \cdot 3 \cdots n \tag{8}$$

para $n \geq 1$ y $0! = 1$.

Varios problemas de conteo se reducen a calcular el número de funciones inyectivas entre dos conjuntos. Por ejemplo, ¿de cuántas maneras podemos colocar p bolas enumeradas en n cajas? Otro problema típico es: ¿cuántos arreglos, o conjuntos ordenados, pueden construirse extrayendo sin reposición p elementos de un conjuntos con n elementos. La respuesta a ambas preguntas es I_n^p.

El caso especial $I_n^n = P_n = n!$ es comunmente interpretado como el total de permutaciones de n elementos, lo cual no es más que el número de funciones biyectivas sobre un conjunto de n elementos.

Números Combinatorios. Sea F un conjunto con n elementos, a continuación vamos a responder la pregunta de cuántos subconjuntos de F con p elementos hay.

Ya que un arreglo de p elementos de F $(x_1, x_2, ..., x_p)$ puede identificarse como una función inyectiva $f : \{1, \dots, p\} \to F$ definida por $f(i) = x_i$, el número de arreglos o subconjuntos ordenados de F con p elementos es I_n^p. Ahora, las $p!$ permutaciones del arreglo $(x_1, \ldots, x_p)$ representan el mismo subconjunto de F. En consecuencia, el número de subconjuntos diferentes de F con p elementos es I_n^p dividido por el número $p!$ de permutaciones de un conjunto con p elementos. Así, si $p \leq n$, el número de subconjuntos de F con p elementos es

$$\binom{n}{p} = \frac{n!}{(n-p)!p!} \tag{9}$$

De la fórmula del binomio de Newton y de los cálculos anteriores podemos deducir que el número de subconjuntos de un conjunto de n elementos es 2^n, ya que

$$\sum_{p=0}^{n} (\text{número de subconjuntos con } n \text{ elementos}) = \sum_{p=0}^{n} \binom{n}{p} = 2^n. \tag{10}$$

Una propiedad útil de los números combinatorios es

$$\binom{n}{p} = \binom{n}{n-p}. \tag{11}$$

Otra, conocida como fórmula de Pascal, es

$$\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}. \tag{12}$$

Varios problemas clásicos del cálculo de probabilidades, que se reducen a contar el número de elementos de un conjunto son versiones del siguiente problema de muestreo sin reposición:

De una caja que contiene N_1 bolas negras y N_2 bolas rojas y escogemos aleatoriamente n bolas $(n \leq N_1 + N_2)$ sin reposición. ¿Cuál es la probabilidad de escoger exactamente k bolas negras? Si k es mayor que N_1 o n, la probabilidad de escoger k bolas negras es cero, así que supondremos que $0 \leq k \leq \text{mín}(N_1, n)$. El conjunto Ω de todos los posibles resultados del experimento aleatorio es la familia de todos los subconjuntos ω de n bolas de las $N_1 + N_2$ bolas de la caja. De manera que

$$|\Omega| = \binom{N_1 + N_2}{n}$$

Debemos contar los subconjuntos ω con k bolas negras y $n-k$ bolas rojas. Para formar tal conjunto debemos formar un conjunto de k bolas negras entre las N_1 bolas negras. Sabemos que hay $\binom{N_1}{k}$ posibilidades de hacer lo anterior. Para cada subconjunto de k bolas negras, debemos asociar un subconjunto de $n-k$ bolas rojas. Este conjunto lo formamos de entre las N_2 bolas rojas y hay $\binom{N_2}{n-k}$ maneras de hacerlo. Así que, si A es el evento que consiste en escoger k bolas negras y $n-k$ bolas rojas, de las $N_1 + N_2$ bolas que hay en la caja, entonces

$$|A| = \binom{N_1}{k}\binom{N_2}{n-k}$$

Por lo tanto, la probabilidad de A es

$$P(A) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N_1+N_2}{n}} \tag{13}$$