

Rincón de la Bioestadística

Cálculo de probabilidades. Primera parte

Gabriel Cavada Ch.

División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

Introducción

El cálculo de probabilidades se origina en la época post renacentista, fruto del estudio de los juegos de azar, por el deseo de cuantificar las posibilidades de ganar o perder ante una baza de naipes, el lanzamiento de un dado o de una moneda al aire. Este interés lúdico inicial trascendió en la historia del pensamiento, pues el análisis fino de cualquier situación real lleva a considerar la porción de azar (imponderable) presente en la misma. Tanto es así que la única seguridad que tenemos es nuestra muerte “biológica”. Cuando decimos que algo será “seguro” en realidad estamos diciendo que es altamente probable que ello ocurra.

En el estudio de la realidad se distinguen dos tipos de experimentos: los determinísticos y los probabilísticos:

1) Los experimentos determinísticos son aquellos que tienen sólo un resultado posible y, además, éste es predecible. Por ejemplo, si el experimento consiste en soltar una piedra desde una cierta altura y observar el sentido del movimiento, el único resultado posible será que la piedra caiga en la dirección vertical y hacia el suelo.

2) Los experimentos probabilísticos son aquellos que tienen más de un resultado posible y cada resultado es impredecible. Por ejemplo, lanzar una moneda al aire y observar que lado mostrará después de caer. Notemos que si bien sabemos de antemano que la moneda puede mostrar “cara” o “sello”, no podemos predecir que ocurrirá y el resultado lo conoceremos sólo después que la moneda haya finalizado su movimiento.

Dado un experimento cualquiera, que llamaremos E, definimos como ESPACIO MUESTRAL, denotado por Ω , al conjunto de todos los posibles resultados de E. Como ejemplos tenemos:

a) E: Se lanza una moneda al aire

$$\Omega = \{\text{cara, sello}\}$$

b) E: Se lanza un dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Se llama suceso o evento a cualquier subconjunto de Ω , los que se identifican por letras mayúsculas. El hecho que A sea un suceso de Ω , lo denotamos por $A \subset \Omega$. Observamos inmediatamente que el conjunto vacío (\emptyset) es un suceso, pues $\emptyset \subset \Omega$ y le llamamos suceso vacío o suceso imposible.

Si jugamos al “cara” o “sello” y previamente se nos pregunta por la “probabilidad” de sacar cara, diremos que es 50%, pues hay sólo dos posibles resultados, pero además hemos supuesto que las posibilidades de obtener cara son idénticas a las de obtener sello, concepto que se denomina EQUIPROBABILIDAD. En lenguaje técnico diremos que los resultados, cara y son equiprobables, obviamente si es que la moneda que se lanza no está cargada.

Se llama medida de un conjunto a algún número que nos indique el tamaño del conjunto; así, la medida del conjunto A se denota por $m(A)$. Si el conjunto es finito y se pueden contar sus elementos, la medida natural que aparece es $m(A) = \text{“número de elementos del conjunto”}$. Por ejemplo si $\Omega = \{1, 2, 3, 4, 5, 6\}$ entonces $m(\Omega) = 6$.

Si el conjunto es un intervalo de la recta real o una porción del plano cartesiano puede considerarse como $m(A) = \text{“longitud del intervalo”}$ o $m(A) = \text{“área de la porción del plano cartesiano”}$ según sea el caso.

- Definición clásica de probabilidad, probabilidad de la unión, sucesos excluyentes, probabilidad condicional, probabilidad total y fórmula de Bayes.

Introducido el concepto de medida, podemos definir la probabilidad de un suceso A como: “medida de A dividida por medida de Ω ”. En símbolos:

Rincón de la Bioestadística

$$P(A) = \frac{m(A)}{m(\Omega)}$$

De esta definición aparecen dos resultados fundamentales:

$P(\emptyset)=0$, la probabilidad del suceso imposible es nula.

$P(\Omega)=1$, la probabilidad del espacio muestral es 1.

Dos sucesos A y B se dicen excluyentes, si es IMPOSIBLE que ocurran al mismo tiempo; en símbolos $A \cap B = \emptyset$. Por ejemplo si al lanzar un dado, éste muestre “un número par e impar” a la vez.

Hechas las consideraciones anteriores, enunciamos los AXIOMAS del cálculo de probabilidades:

a) $0 \leq P(A) \leq 1$

b) Para una colección de sucesos excluyentes: $P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$

Para enfrentar un problema de cálculo de probabilidades, se deben definir cuidadosamente los sucesos de interés. Ejemplifiquemos con algunas situaciones elementales como el experimento de “lanzar un dado”:

E: Se lanza un dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Definamos los sucesos siguientes y calculemos sus probabilidades de ocurrencia

A: “el dado muestra 1”, así: $A = \{1\}$ y $m(A)=1$, con lo que:

$$P(A) = \frac{1}{6}$$

B: “el dado muestra un número impar”, así $B = \{1, 3, 5\}$ y $m(B)=3$, con lo que:

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

Sin embargo, la realidad presenta sucesos compuestos que se forman por unión, intersección y complementación. Efectuemos estas composiciones y estudiemos su significado. En consecuencia, dados los sucesos A y B se tiene:

$A \cap B$: sucede A y sucede B (ambos a la vez)

$A \cup B$: sucede A ó B, así: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

A^c : no sucede A, así $P(A^c) = 1 - P(A)$

Decimos que los sucesos A y B son INDEPENDIENTES, si la ocurrencia de uno de ellos no altera la ocurrencia o no ocurrencia del otro; la hipótesis de independencia se expresa así: $P(A \cap B) = P(A)P(B)$.

Además, la realidad presenta abundantes SUCESOS CONDICIONALES, es decir, aquellos que condicionan su ocurrencia a la presencia de otros. Por ello, podemos preguntar por la probabilidad de que ocurra un evento DADO EL HECHO que ocurrió tal o cual otro evento. Si consideramos los sucesos A y B, de modo que B condiciona la ocurrencia de A, entonces la probabilidad de que “ocurra A dado el hecho que ocurrió B” es:

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

Nota: En la realidad es difícil acceder a la $P(A \cap B)$, sin embargo, se puede disponer de $P(A|B)$ y $P(B)$, luego hay que tener presente que $P(A \cap B) = P(A|B) P(B)$.

Rincón de la Bioestadística

En múltiples oportunidades la ocurrencia de un suceso principal A se debe a la ocurrencia previa de causas, que también son sucesos, de modo que en el cálculo de la probabilidad de la ocurrencia de A las probabilidades de los sucesos causales deben ser incluidas según la ponderación o influencia que tengan sobre A. Si el suceso principal A se debe a las causas E1, E2,..., En, entonces:

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \dots + P(A|E_n)P(E_n)$$

ó

$$P(A) = \sum_{i=1}^n P(A|E_i)P(E_i)$$

Esta fórmula recibe el nombre de TEOREMA DE LA PROBABILIDAD TOTAL.

Consideremos ahora el problema de calcular la probabilidad de que el suceso principal A se haya producido debido a la causa E_k. En símbolos buscamos P(E_k|A) y este resultado es simple de encontrar, pues proviene de:

$$P(E_k|A) = \frac{P(A|E_k) \cdot P(E_k)}{P(A|E_1) \cdot P(E_1) + P(A|E_2) \cdot P(E_2) + \dots + P(A|E_n) \cdot P(E_n)}$$

ó

$$P(E_k|A) = \frac{P(A|E_k) \cdot P(E_k)}{\sum_{i=1}^n P(A|E_i) \cdot P(E_i)}$$

Esta fórmula es conocida como el TEOREMA DE BAYES, donde el denominador no es más que la probabilidad total.

Ejemplo: En un hospital hay tres servicios: Urgencia, Cirugía y Medicina. El porcentaje de hospitalizados por servicio es: Urgencia 30%, Cirugía 20% y Medicina 50%. Si la mortalidad en cada servicio es 10%, 5% y 3%, respectivamente. ¿Cuál es la probabilidad de que un paciente hospitalizado muera?

Suceso principal, A : el paciente muere

Causas: E1: el paciente está en urgencia

E2: el paciente está en cirugía

E3: el paciente está en medicina

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3)$$

$$P(A) = 0.1 \cdot 0.3 + 0.05 \cdot 0.2 + 0.03 \cdot 0.5 = 0.055$$

Y si se nos comunica que ha ocurrido una muerte, ¿Cuál es la probabilidad que haya ocurrido en Urgencia?

Suceso principal, A: el paciente muere

Causas: E1: el paciente está en urgencia

E2: el paciente está en cirugía

E3: el paciente está en medicina

Es decir se pide:

$$P(E_1|A) = \frac{P(E_1 \cap A)}{P(A)} = \frac{P(A \cap E_1)}{P(A)} = \frac{P(A|E_1) \cdot P(E_1)}{P(A|E_1) \cdot P(E_1) + P(A|E_2) \cdot P(E_2) + P(A|E_3) \cdot P(E_3)}$$

Rincón de la Bioestadística

$$P(E_1 | A) = \frac{0.1 \cdot 0.3}{0.055} = 0.545$$

Una situación clásica es lo que ocurre con los exámenes diagnósticos, cuyos resultados se cruzan con la verdadera situación del paciente. Es decir, se distinguen los sucesos: “El paciente está sano”, “El paciente está enfermo”, “El examen da positivo” y “El examen da negativo”; la situación la observamos en la siguiente tabla:

Verdadero estado del paciente				
Resultado del examen	Enfermo (E)		Sano (E')	
	Positivo (+)	A	B	A+B
	Negativo (-)	C	D	C+D
		A+C	B+D	N=A+B+C+D

Así:

$$P(+ | E) = \frac{A}{A+C} \text{ ,recibe el nombre de sensibilidad del examen}$$

$$P(- | E') = \frac{D}{B+D} \text{ ,recibe el nombre de especificidad del examen}$$

Falsos positivos son aquellos sanos cuyo examen da positivo; falsos negativos son aquellos enfermos cuyo examen da negativo.

Una configuración similar a la anterior es usada comúnmente en epidemiología, pues el sujeto se clasifica doblemente, según una respuesta (enfermo o sano) y una variable explicativa (expuesto o no expuesto), así:

	Enfermo (E)	Sano (E')	
Expuesto	A	B	A+B
No expuesto	C	D	C+D
	A+C	B+D	N=A+B+C+D

Así es posible calcular $P(\text{Enfermo} | \text{Expuesto})$, $P(\text{Sano} | \text{Expuesto})$, $P(\text{Enfermo} | \text{No expuesto})$ y $P(\text{Sano} | \text{No expuesto})$ y como medidas de la asociación se proponen:

$$\text{El Riesgo Relativo: } RR = \frac{P(\text{Enfermo} | \text{Expuesto})}{P(\text{Enfermo} | \text{No Expuesto})} = \frac{A(C+D)}{C(A+B)}$$

$$\text{El "Odds Ratio": } OR = \frac{\frac{P(\text{Enfermo} | \text{Expuesto})}{P(\text{Sano} | \text{Expuesto})}}{\frac{P(\text{Enfermo} | \text{No Expuesto})}{P(\text{Sano} | \text{No Expuesto})}} = \frac{AD}{BC}$$

Rincón de la Bioestadística

Ambas medidas de asociación tienen valor de nulidad igual a 1, es decir si RR u OR son 1 significa que no hay asociación entre enfermedad y exposición.

El riesgo relativo sólo se puede calcular en diseños transversales o prospectivos, ya que siempre los totales de exposición o no exposición no son aleatorios, mientras que el OR se puede calcular en diseños transversales, prospectivos y retrospectivos; además, el OR tiene propiedades matemáticas muy deseables lo que lo hacen una medida de asociación más universal.

Si el número relativo de enfermos es pequeño (prevalencia baja) RR y OR son numéricamente similares.

Tradicionalmente si RR u OR es menor que 1 se habla que la exposición es un factor protector, mientras que si son mayores que 1 se habla de un factor de riesgo. Las inferencias estadísticas asociadas a estos indicadores se verán en un artículo futuro.