Introducción a la Probabilidad

Ing. Nicolás G Tripp

Algunas definiciones

Espacio muestral (E): es el conjunto de todos los resultados posibles de un experimento aleatorio.

Experimento aleatorio: cualquier proceso cuyo resultado no se puede predecir de forma exacta.

Suceso o punto muestral: un resultado obtenido al realizar un experimento aleatorio.

Evento (A): subconjunto del espacio muestral.

Teoría de conjuntos

Complemento de un evento A ($\bar{A}$): es el evento que contiene todo lo que no es parte de A.

Intersección de dos eventos (A∩B): es el evento que incluye los elementos compartidos por **"A y B"**

Eventos mútuamente exclusivos o disjuntos: la intersección no tiene elementos.

Unión de dos eventos (A∪B): es el evento que incluye los elementos de **"A o B"**

Conteo de espacios muestrales

Para determinar la cantidad de elementos de un conjunto se puede:

- listar los elementos
- utilizar un diagráma de árbol
- utilizar fórmulas de conteo:
  - Regla de la multiplicacion
  - Permutaciones
  - Combinaciones.

Obligatorio
para conjuntos
grandes!!!

Regla de la multiplicación

Si una operación 1 se puede realizar de n_1 formas distintas y para cada opción se puede realizar una segunda operación de n_2 formas, y así sucesivamente, entonces el conjunto de operaciones se puede realizar de $n_1 * n_2 * \ldots n_m$ formas diferentes.

Ejemplo:

Se sospecha que una mujer porta el gen de la hemofilia y es posible que algunos de sus tres hijos contraigan la enfermedad. Determine cuántos sucesos diferentes contiene el espacio muestral.

E: hijo enfermo, $\bar{E}$: hijo sano.   2*2*2=8 casos posibles

Permutaciones y combinaciones

Para un conjunto de n objetos diferentes elegidos de a r por vez

cada ordenamiento posible se llama "permutación"

$$nPr = \frac{n!}{(n-r)!}$$

cada agrupamiento posible sin importar el orden

se llama "combinación"

$$nCr = \frac{n!}{r!(n-r)!}$$

Conteo de espacios muestrales

Ejemplo:

Los tres hijos de la reina (Juán, María y Ana) aspiran a ser comandante del ejército o del hospital del reino. Determine el espacio muestral.

$$nPr=\frac{n!}{(n-r)!}=\frac{3!}{1!}=6\,opciones$$

Ejército	Juán	María	Juán	Ana	María	Ana
Hospital	María	Juán	Ana	Juán	Ana	María

$$nCr=\frac{n!}{r!(n-r)!}=\frac{3!}{2!1!}=3\,opciones$$

Ejército	Juán o María	Juán o Ana	María o Ana
Hospital	María o Juán	Ana o Juán	Ana o María

Probabilidad

Definimos la “probabilidad de un suceso” como una medida de la certidumbre asociada a la ocurrencia de tal suceso

Se expresa mediante un número real entre 0 (imposible) y 1 (certeza).

$$0 \leq P(A) \leq 1 \qquad P(\emptyset)=0 \qquad P(E)=1$$

Fuentes de probabilidad

- Intuición o saber popular
- Frecuencia observada en experimento real $\lim_{N \to \infty} \frac{n}{N} = P(A)$
- Laplace (experimento lógico) $P(A) = \frac{\text{casos favorables}}{E}$

Laplace asume que cada suceso tiene la misma probabilidad de ocurrencia

Probabilidad de eventos

Probabilidad de un evento: es la suma de las probabilidades de sus puntos muestrales. $P(A)=\sum_i P(a_i)$

Suma de dos eventos: $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

Suma de 3 eventos:

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

Eventos complementarios: $P(A)+P(\overline{A})=1$

Probabilidad de eventos

Ejemplo 1

Un alergista afirma que el 50% de los pacientes que examina son alérgicos a algún tipo de planta. ¿Cuál es la probabilidad de que 3 de sus próximos 4 pacientes sean alérgicos a algún tipo de planta?

A: el paciente es alérgico a un tipo de planta

$\bar{A}$: el paciente no es alérgico a un tipo de planta

$$P(A) = \frac{\text{casos favorables a } A}{E}$$

$$P(A) = \frac{\{(AAA\bar{A}), (AA\bar{A}A), (A\bar{A}AA), (\bar{A}AAA)\}}{2 \text{ x } 2 \text{ x } 2 \text{ x } 2} = \frac{1}{4}$$

Probabilidad de eventos

Ejemplo 2

La probabilidad de que un paciente se recupere de una cirugía del corazón es del 80%. ¿Cuál es la probabilidad de que 2 de los 3 próximos pacientes se recuperen?

R: el paciente se recupera de la cirugía
$\overline{R}$: el paciente no se recupera de la cirugía
A: 2 de los 3 pacientes se recuperan

$$P(A)=P(R\cap R\cap \bar{R})+P(R\cap \bar{R}\cap R)+(\bar{R}\cap R\cap R)=3\ x\ 0{,}2\ x\ 0{,}8^2=0{,}384$$

Probabilidad condicionada

La probabilidad de que suceda un evento B luego de la ocurrencia de un evento A se llama **probabilidad condicionada**.

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

A partir de la probabilidad condicionada se tiene que

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

La regla general se conoce como regla de multiplicación

$$P(A_1\cap A_2)=P(A_1|A_2)P(A_2)$$

Probabilidad condicionada

<u>Ejemplo</u>

El 60% de los individuos de una población están vacunados contra cierta enfermedad. Durante una epidemia se sabe que el 20% ha contraído la enfermedad y 2 de cada 100 personas están vacunadas y están enfermas.

Calcular el porcentaje de vacunados que enferma y el de vacunados entre los que están enfermos.

V: el individuo está vacunado
E: el individuo está enfermo

$$P(E|V)=\frac{P(E\cap V)}{V}=\frac{0,02}{0,6}=0,033 \qquad P(V|E)=\frac{P(V\cap E)}{E}=\frac{0,02}{0,2}=0,1$$

Sucesos independientes

Si la probabilidad de un suceso no cambia con la ocurrencia de otro se dice que ambos sucesos son independientes

$$\text{Sucesos independientes} \Leftrightarrow P(B|A) = P(B) \text{ o } P(A|B) = P(A)$$

Utilizando la regla de multiplicación también es cierto que

$$\text{Sucesos independientes} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

Sucesos independientes

Ejemplo

Dos tratamientos A y B curan una determinada enfermedad el 20% y 30% de los casos respectivamente. Suponiendo que ambos actúan de modo independiente, ¿cuál de las siguientes estrategias utilizaría para curar a un individuo con tal enfermedad?

a) Aplicar ambos tratamientos a la vez

b) Aplicar primero el B, y si no surte efecto aplicar el A

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,2 + 0,3 = 0,5$$

$$P(A|B) = P(A) = 0,2$$

Probabilidad Total

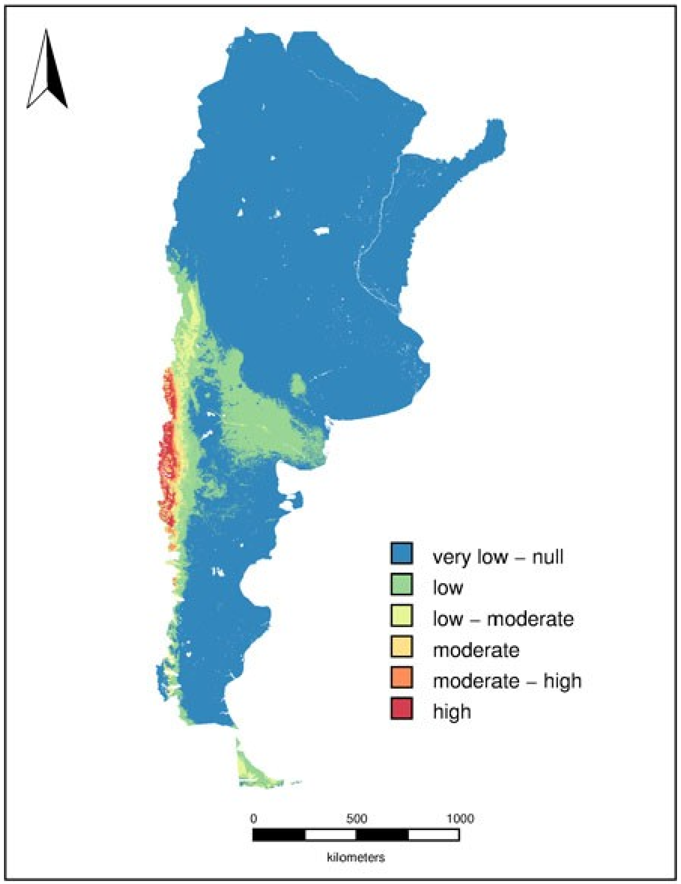
Llamamos "**partición**" del espacio muestral al conjunto de eventos que forman un sistema exhaustivo (contiene todos los sucesos) y excluyente (sin interseccion) de sucesos.

Si un conjunto de eventos $B_1,\ldots,B_n$ forman una partición del espacio muestral E tal que $P(B_i)\neq 0$ para i=1,..,n entonces para un evento A incluido en E, **la probabilidad total** es

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

Probabilidad Total

Niveles de Riesgo para Hantavirus en el sur de Argentina



Andreo et al (2014). Estimating Hantavirus Risk in Southern Argentina: A GIS–based approach combining human cases and host distribution. Viruses 6, 201–222.

Las provincias forman una partición de la población argentina.

E: habitante contagiado de hantavirus

B_i: habitante de la i-ésima provincia argentina

$$P(E)=\sum_{i=1}^{23} P(E|B_i)P(B_i)$$

Se obtiene de la distribución demográfica

Se obtiene de los niveles de riesgo

Probabilidad Total

Ejemplo

En una cierta ciudad se sabe por experiencia que el 5% de los adultos mayores de 40 años padece algún tipo de cáncer. Si la probabilidad de que un doctor diagnostique la enfermedad en un paciente enfermo es 78% y en un paciente sano es 6%, ¿cuál es la probabilidad de que un adulto mayor de 40 años en dicha ciudad sea diagnosticado con cáncer?

C: paciente con cáncer
D^+: paciente diagnosticado como enfermo de cáncer

$$P(C)=0{,}05 \quad P(D^+|C)=0{,}78 \quad P(D^+|\bar{C})=0{,}06$$

$$P(D^+)=P(D^+|C)\cdot P(C)+P(D^+|\bar{C})\cdot P(\bar{C})=0{,}096$$

Teorema de Bayes

Sea un conjunto de eventos $A_1,\ldots,A_n$ una partición del espacio muestral E. Sea B un suceso de E del que conocemos todas las probabilidades condicionadas $P(B \mid A_i)$, entonces

$$P(A_j|B)=\frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

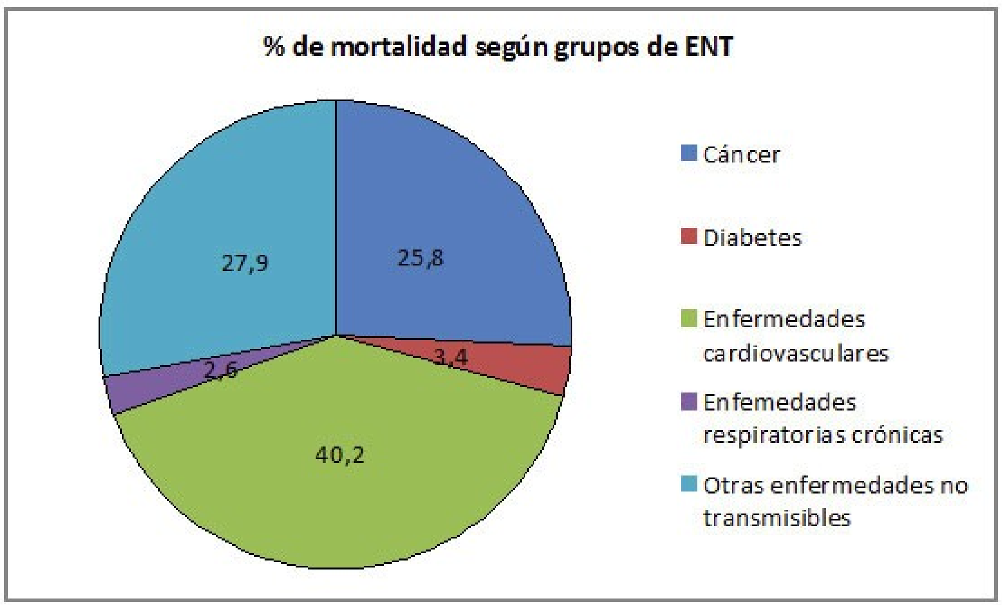
Buscamos la causa más probable del evento

Teorema de Bayes

$$P(M)=(0{,}258)(0{,}00212)+(0{,}034)(0{,}1)+(0{,}402)(0{,}33)+(0{,}026)(0{,}145)$$

$$P(\text{Cáncer}|M)=\frac{P(M|\text{Cáncer})P(\text{Cáncer})}{P(M)}=\frac{0{,}000547}{0{,}1404}=0{,}004$$

$$P(\text{Cardiovascular}|M)=0{,}94$$



P(Cáncer): 212 casos por 100.000 habitantes. Instituto Nacional del cáncer (2018)

P(Diabetes): 1 de cada 10 habitantes. Diario Infobae (2018)

P(Cardiovascular): 1 de cada 3. Diario TN (2018)

P(EPOC): 14,5%. Estudio EPOC.AR (2016)

Ministerio de salud en base a datos de la Dirección de Estadísticas e Información en Salud (2013).

Teorema de Bayes

Ejemplo 2

El 10% de la población de Málaga es alcohólica. En los decesos que reporta el ministerio de salud no figura el alcoholismo pero sí diagnósticos de hepatopatías. Se realizó un estudio que mostró que el 85% de los individuos alcohólicos y el 7% de los no alcohólicos sufrían tales patologías. Se busca saber cuál es la probabilidad de que un individuo con las hepatopatías sea alcohólico.

A: el individuo es alcohólico

H: el individuo padece una hepatopatía

$$P(H|A)=0{,}85 \qquad P(H|\bar{A})=0{,}07 \qquad P(A)=0{,}1$$

$$P(A|H)=\frac{P(H|A)\,P(A)}{P(H|A)\,P(A)+P(H|\bar{A})\,(1-P(A))}=\frac{0{,}085}{0{,}148}=0{,}57$$

Test diagnósticos

La aplicación del teorema de Bayes a la Medicina se denomina "tests diagnósticos"

Los resultados se denominan positivos T^{+} si la evidencia a favor de la enfermedad es alta y negativos T^{-} en caso contrario.

Los test poseen las siguientes características

- Sensibilidad (tasa de verdaderos positivos) $P\left(T^{+}|E\right)$
- Especificidad (tasa de verdaderos negativos) $P\left(T^{-}|\bar{E}\right)$
- Índice predictivo de verdaderos positivos $P\left(E|T^{+}\right)$
- Índice predictivo de verdaderos negativos $P\left(\bar{E}|T^{-}\right)$

Test diagnósticos

Ejemplo 1

Con el objeto de diagnosticar la colelietasis se usan los ultrasonidos. La técnica tiene una sensibilidad del 91% y una especificidad del 98%. En cierta población el 20% de los habitantes padece la enfermedad. Si a un habitante se le aplica el test y da positivo, ¿cuál es la probabilidad de que sufra la colelietasis?

E: el paciente padece colelietasis

$$P(T^+|E)=0{,}91 \qquad P(T^-|\bar{E})=0{,}98 \qquad P(E)=0{,}2$$

$$P(E|T^+)=\frac{P(T^+|E)P(E)}{P(T^+|E)P(E)+P(T^+|\bar{E})P(\bar{E})}=0{,}9192$$

éste es 1−falso verdadero

éste es 1−complemento

Test diagnósticos

Ejemplo 2

Se conoce que 7 de cada 100 mujeres mayores de 60 años desarrolla un cierto tipo de cáncer. Existe un test de sangre que produce falsos negativos el 10% de las veces y falsos positivos el 5%. Si una mujer recibe un resultado negativo, ¿cuál es la probabilidad de que esté realmente enferma?

C: la mujer padece cáncer

$$P\left(T^{+}|\bar{C}\right)=0,05 \qquad P\left(T^{-}|C\right)=0,1 \qquad P\left(C\right)=0,07$$

$$P\left(C|T^{-}\right)=\frac{P\left(T^{-}|C\right)P\left(C\right)}{P\left(T^{-}|C\right)P\left(C\right)+P\left(T^{-}|\bar{C}\right)P\left(\bar{C}\right)}$$

$$=\frac{P\left(T^{-}|C\right)P\left(C\right)}{P\left(T^{-}|C\right)P\left(C\right)+\left(1-P\left(T^{+}|\bar{C}\right)\right)\left(1-P\left(C\right)\right)}=\frac{0,007}{0,8905}=0,008$$