Independencia y condicionamiento

Come podemos cambiar la probabilidad de un evento si uno tiene una información adicional? El concepto que nos permite responder esta pregunta se llama probabilidad condicional. El contexto ideal de su uso es cuando uno enfrenta un problema donde dos características diferentes se relacionan si tener una causalidad completa entre las dos.

Vacunas y enfermedades Nos interesamos al problema de conocer la probabilidad que una persona tomada al azar en una población tenga un enfermedad en particular. Supongamos que por esta enfermedad existe una vacuna que permite reducir (sin proteger completamente) la proporción de la población que sera sensible de ser infectado. Ademas, supongamos que solo una proporción p de la población está vacunada contra esa enfermedad. Podemos denotar E por el evento "la persona está enferma" y V el evento "la persona está vacunada". Se siente que la realización de uno de los dos eventos debe de influir sobre la probabilidad de realización del otro. Por ejemplo, frecuentemente las autoridades quieren saber la probabilidad de E condicionalmente a V que se nota

$$\mathbb{P}(E|V).$$

1 Probabilidad condicional a un evento, Independencia

Definición 1. *Sea B un evento de probabilidad positiva $\mathbb{P}(B) > 0$. Para todo evento $A \in \mathcal{A}$ definimos la **probabilidad condicional** de A condicionalmente a B por*

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

Propiedades de la probabilidad condicional La probabilidad condicional tiene la propiedad fundamental siguiente.

- El mapeo $A \mapsto \mathbb{P}(A|B)$ es una probabilidad sobre Ω.

De eso deducimos directamente las propiedades básicas heredadas de las propiedades de una probabilidad.

1. $\mathbb{P}(\emptyset|B) = 0$ y $\mathbb{P}(\Omega|B) = 1$.
2. $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.
3. Si $A \subset C$, entonces $\mathbb{P}(A|B) \leq \mathbb{P}(C|B)$.
4. Si $B \subset A$, $\mathbb{P}(A|B) = 1$. $(\star)$
5. Para toda sucesión A_n de eventos disjuntos
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \middle| B\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B)$$
6. Si $A_n \uparrow A$ (so si $A_n \downarrow A$) entonces $\mathbb{P}(A_n|B) \to \mathbb{P}(A|B)$.

Formula para intersecciones y independencia Hasta ahora, hemos visto una manera de descomponer una probabilidad en eventos mas sencillos con uniones disjuntas. Uno podría tener ganas de encontrar una formula general para intersecciones de eventos. Una tal formula no existe en general y de eso nace la definición de independencia. Usando la definición de la probabilidad condicional vemos que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$. La independencia simplifica esa escritura.

Definición 2. *Dos eventos son **independientes** cuando*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

De manera equivalente, lo son cuando $\mathbb{P}(A|B) = \mathbb{P}(A)$ (una información sobre B no cambia la probabilidad de A) o $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Un ejemplo de condicionamiento Una urna contiene r bolas rojas y v bolas verdes. Sacamos dos bolas una después de la otra sin reemplazo en la urna. Denotamos

$$A = \{\text{sacamos una bolas roja al primer sorteo}\}, \quad B = \{\text{sacamos una bolas roja al segundo sorteo}\}.$$

Un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ que modela el problema debe de cumplir

$$\mathbb{P}(A) = \frac{r}{r+v}$$
$$\mathbb{P}(B|A) = \frac{r-1}{r+v-1}.$$

Deducimos que la probabilidad de observar dos bolas rojas es

$$\mathbb{P}(\text{dos rojas}) = \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(A) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}$$

Del otro lado, podemos calcular la probabilidad de B usando una enumeración de los arreglos de dos bolas en un conjuntos de $r+v$. Entonces $|\Omega| = (r+v)(r+v-1)$ y $|B| = (r+v-1)r$ lo que sucede $\mathbb{P}(B) = \frac{r}{r+v}$. Vemos que $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\,\mathbb{P}(B)$ y los dos eventos no son independientes.

2 Condicionamiento en cascada, formula de Bayes

Condicionamiento en cascada El condicionamiento al respeto de varios eventos es posible. La única condición para poder definirlo es de siempre condicionar por eventos de probabilidad positiva. Sean eventos $A_1, \ldots, A_n$ tal que $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. En este caso,

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Prueba. Primero debemos de verificar que los condicionamientos no se hacen al respeto de eventos de probabilidad nula. Pero, $\bigcap_{i=1}^{n-1} A_i \subset \bigcap_{i=1}^{k} A_i$ y entonces

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right) \geq \mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0.$$

Finalmente,

$$\begin{aligned}
&\mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \\
&= \mathbb{P}(A_1)\,\frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)}\,\frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \times \cdots \times \frac{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1} \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})} \\
&= \mathbb{P}(A_1 \cap \cdots \cap A_{n-1} \cap A_n)
\end{aligned}$$

□

Formula de condicionamiento total Cuando se puede razonar con una disyunción de casos y que los diferentes casos se representan por eventos de probabilidad positiva, la formula adecuada sera la formula de condicionamiento total.

Teorema 1. *Sea $B_1, \ldots, B_n$ una partición de Ω en eventos de probabilidad positiva ($\forall i,\ \mathbb{P}(B_i) > 0$). Entonces $\forall A \in \mathcal{A}$,*

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

Un caso particular es cuando $n = 2$, y $\mathbb{P}(A) = \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)$.

Prueba. Observando que $B_1, \ldots, B_n$ es una partición de Ω, tenemos

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} (A \cap B_i),$$

y como $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$, el resultado sucede. □

Formula de Bayes Obviamente, la formula de definición se puede manipular para decir que $\mathbb{P}(A\cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$. Aunque, por el momento, no es mas que un juego de escritura, podemos deducir la formula de Bayes

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

que permite intervertir el condicionamiento de A al respeto de B. Ademas, si $B = B_1$ es un evento que pertenece a una partición $B_1, B_2, \ldots, B_n$ de Ω, sucede que

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(A|B_1)\,\mathbb{P}(B_1)}{\sum_{i=1}^{n}\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}$$

Esa idea es al origen de la teoría fundamental de la estadística bayesiana.

3 Pares de variables aleatorias y independencia de variables

Una manera agradable de manipular conjuntos de variables aleatorias (independientes o no) es de construir el vector aleatorio cuyas entradas son las dichas variables. Por ejemplo, si supongamos dadas dos variables aleatorias X y Y, construimos el **vector aleatorio** (X,Y) como la variable aleatoria

$$\omega \mapsto (X(\omega), Y(\omega)).$$

En este caso, construimos un objecto que tiene valores en $\mathbb{R}^2$. Un par de variables discretas forma un vector de soporte discreto en $\mathbb{R}^2$. $(\star)$
En este contexto, la variable X se llama **primera marginale** y Y la **segunda marginale**. Como en el caso real, podemos definir la ley del vector aleatorio.

Definición 3. *Sea (X,Y) un vector aleatorio discreto. Llamamos **ley del vector aleatorio** el mapeo $P_{X,Y}$ dado por*

$$\forall B \subset \mathbb{R}^2, \quad P_{X,Y}(B) = \mathbb{P}\left(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}\right).$$

*Las leyes P_X y P_Y se llaman **ley marginales** del vector (X,Y).*

Como en el caso real el soporte de (X,Y) se reduce a una parte numerable de $\mathbb{R}^2$ y la ley del vector está completamente definido por los valores $p_{x,y} = \mathbb{P}(X = x, Y = y)$ donde $x \in X(\Omega)$ y $y \in Y(\Omega)$. El conocimiento de $P_{X,Y}$ contiene la información de las marginales P_X y P_Y por las identidades que ya encontramos,

$$\forall x \in X(\Omega), \quad \mathbb{P}(X = x) = \sum_{y\in Y(\Omega)} \mathbb{P}(X = x, Y = y),$$
$$\forall y \in Y(\Omega), \quad \mathbb{P}(Y = y) = \sum_{x\in X(\Omega)} \mathbb{P}(X = x, Y = y)$$

Al inverso, no es verdad que P_X y P_Y son suficientes para describir la ley del vector!!

Un contra-ejemplo Lanzamos dos dados, un rojo y un azul. Notamos X el resultado del dado rojo y Y el resultado del dado azul. Ademas, definimos $Z = 7 - X$. Los vectores (X,Y) y (X,Z) tienen las mismas leyes marginales (uniformes sobre $\{1,2,3,4,5,6\}$) pero no tienen la misma ley conjunta.

Un caso particular es cuando las dos variables no se influyen una sobre la otra. El ejemplo muestra que es el caso de X y Y pero no de X y Z. Unas variables con un tal comportamiento son particulares en probabilidad y da nacimiento a la definición siguiente.

Definición 4. *Dos variables X y Y se dicen **independientes** si para todo A y todo B subconjuntos de $\mathbb{R}$, los eventos $\{X \in A\}$ y $\{Y \in B\}$ son independientes (como eventos), entonces*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).$$

Ejemplos y comentarios

- Dos dados diferentes producen variables aleatorias independientes.
- Dos sorteos de bolas en urnas distintas son independientes.
- Dos sorteos sucesivos en una misma urna con reemplazo son independientes.
- La definición pide que se verifique la independencia por todos los subconjuntos A y B. No es suficiente de verificarlo solo sobre un par A, B particular.

Teorema 2. *Sean X, Y dos variables aleatorias independientes, entonces la ley del vector (X, Y) es completamente caracterizada por las marginales P_X y P_Y.*

Para ver eso, podemos verificar que la independencia de X y Y es equivalente a tener que, para cada $x \in X(\Omega)$ y $y \in Y(\Omega)$,

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y).$$

En efecto,

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \sum_{x \in X(\Omega)\cap A} \sum_{y \in Y(\Omega)\cap B} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in X(\Omega)\cap A} \sum_{y \in Y(\Omega)\cap B} \mathbb{P}(X = x)\,\mathbb{P}(Y = y) \\
&= \sum_{x \in X(\Omega)\cap A} \mathbb{P}(X = x) \times \sum_{y \in Y(\Omega)\cap B} \mathbb{P}(Y = y) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).
\end{aligned}$$

Finalmente, sucede que por cada x, y tenemos $p_{x,y} = p_x p_y$, lo que caracteriza completamente la ley de (X, Y).

4 Suma de dos variables independientes

Un estudio muy común cuando uno tiene una colección de variables aleatorias independientes es de calcular la suma. El supuesto de independencia es muy importante, porque si no la ley de la suma de dos variables no sería caracterizada por la ley de las marginales. En efecto, la suma de dos variables Bernoulli $\mathcal{B}(0.5)$ independientes es una variable binomiale $\mathcal{B}(2, 0.5)$ aunque la suma de $X \sim \mathcal{B}(0.5)$ y de $1 - X$ (que es también B(0.5)) vale siempre $X + (1 - X) = 1$.

Teorema 3. *Sean dos variables X y Y independientes de soportes en $\mathbb{N} \cup \{0\}$. Entonces la ley de la variable aleatoria $X + Y$ está dada por*

$$\forall n \in \mathbb{N} \cup \{0\}, \quad \mathbb{P}(X + Y = n) = \sum_{i=0}^{n} \mathbb{P}(X = i)\,\mathbb{P}(Y = n - i).$$

Prueba. Como las variables X y Y tienen un soporte en los naturales, la variable $X + Y$ también. Descomponemos el evento $\{X + Y = n\}$ en la unión disjunta de los eventos $\{X = i, Y = j\}$ con $i + j = n$. Entonces,

$$\mathbb{P}(X + Y = n) = \sum_{i+j=n} \mathbb{P}(X = i, Y = j) = \sum_{i+j=n} \mathbb{P}(X = i)\,\mathbb{P}(Y = j) = \sum_{i=0}^{n} \mathbb{P}(X = i)\,\mathbb{P}(Y = n - i).$$

Cuidado, sin la independencia, la prueba sería falsa... □

Suma de dos variables de Poisson Si X y Y son dos variables aleatorias de Poisson con parámetros respectivos α y β. Calculamos la ley de la suma $S = X + Y$.

$$\begin{aligned}
\mathbb{P}(S = n) &= \sum_{i=0}^{n} \mathbb{P}(X = i)\,\mathbb{P}(Y = n - i) = \sum_{i=0}^{n} e^{-\alpha}\frac{\alpha^i}{i!} e^{-\beta}\frac{\beta^{n-i}}{(n-i)!} \\
&= \frac{e^{-(\alpha+\beta)}}{n!} \sum_{i=0}^{n} \binom{n}{i} \alpha^i \beta^{n-i} = \frac{e^{-(\alpha+\beta)}(\alpha + \beta)^n}{n!}.
\end{aligned}$$

Entonce S tiene la ley de probabilidad de una Poisson de parámetro $\alpha + \beta$.

5 Problemas

1. Muestre las afirmaciones siguientes:
   - Si $P(A) = 0$ o 1, el evento A es independiente de todos los eventos.
   - Dos eventos incompatibles jamas son independientes.
   - La independencia no es absoluta. (i.e. depende del espacio $(\Omega, \mathcal{A}, \mathbb{P})$).
   - Si A y B son independientes entonces igualmente, A y B^c, A^c y B, A^c y B^c lo son.

2. Cada miembro de un grupo de n jugadores lanza un dado.
   a) Para cualquier pareja de jugadores que lanzan el mismo número, el grupo obtiene un punto. Encuentre la media y varianza del total de puntos obtenidos por el grupo.
   b) Encuentre la media y la varianza del total de puntos obtenidos si cualquier pareja de jugadores que lanzan el mismo número obtienen dicho número de puntos.

3. Si X es una variable aleatoria geométrica. Mostrar que
$$\mathbb{P}(X = n + k | X > n) = \mathbb{P}(X = k), \qquad k, n \geq 1.$$
   ¿ Existe alguna otra distribución en los enteros positivos con esta propiedad?

4. Sean X y Y dos variables independientes de ley geométrica de parámetro p. Calcular la ley de $X + Y$.

5. ($\star$)Supongamos que una persona cae y tira n parejas de cartas y sobres por las escaleras. Cada pareja de cartas y sobres tienen el mismo destinatario. Al recoger las cartas y los sobres, la persona ordena cada pareja de manera aleatoria. Muestre que el X, elnúmero de parejas de cartas y sobres que coinciden, tiene media y varianza unitaria, para $n \geq 2$. Muestre que la función masas de probabilidades de X converge a una función masas de probabilidades de una variable aleatoria Poisson conforme n se va a infinito.

6. ($\star$)Se considera una sucesión de experimentos idénticos y independientes de probabilidad de éxito p. Denotamos S_i el evento de éxito al i-ésimo experimento. Denotamos X_1 el tiempo de primer éxito y $X_1 + X_2$ el tiempo de segundo éxito.
   6.1. Para $j, k \in \mathbb{N}$, expresar los eventos $\{X_1 = j, X_2 = k\}$ en función de los S_i. Deducir la ley de (X_1, X_2) y luego la de X_2. Verificar que X_1 y X_2 son independientes.
   6.2. Calcular $\mathbb{P}(X_1 = k | X_1 + X_2 = n)$ por $k = 1, \ldots, n - 1$. Comentar.