Estadística Bayesiana

Irving Gómez Méndez

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

Ejemplo 1

Tenemos una prueba de laboratorio que detecta una cierta enfermedad A. Denotamos el que salga una prueba positiva para la enfermedad como $T = 1$ y negativa como $T = 0$, y que el paciente tenga la enfermedad A como $E = 1$ y $E = 0$ en otro caso. Las características de la prueba son:

$$\mathbb{P}(T = 1|E = 1) = 0.92, \quad \mathbb{P}(T = 0|E = 0) = 0.99,$$

y la prevalencia de la enfermedad (la proporción de personas enfermas en la población en cuestión) es de 0.12. O sea $\mathbb{P}(E = 1) = 0.12$.

Yo voy y me hago la prueba referida (perteneciendo yo a la población en cuestión) y esta sale positiva. ¿Cuál es la probabilidad posterior de que yo tenga la enfermedad A?

$\mathbb{P}(T=1|E=1)$ ó $\mathbb{P}(T=1|E=0)$ **no** es lo que necesitamos. Más bien queremos saber qué sucede dado o una vez que $T=1$.

Teorema de Bayes:

$$\mathbb{P}(E=1|T=1) = \frac{\mathbb{P}(T=1|E=1)\mathbb{P}(E=1)}{\mathbb{P}(T=1|E=1)\mathbb{P}(E=1) + \mathbb{P}(T=1|E=0)\mathbb{P}(E=0)}$$

$$= 0.9262$$

Pero, en realidad, o tengo la enfermedad o no la tengo; entonces...

¡¿Qué quiere decir: La probabilidad de que tenga la enfermedad es 0.9262?!

¡No hay frecuencias o eventos repetidos: o estoy o no estoy enfermo!

El estadístico Bruno de Finetti comenzaba su libro con la frase *LA PROBABILIDAD NO EXISTE*. Las mayúsculas aparecen en el texto original. ¿A qué se refiere?

La probabilidad es condicional y subjetiva

Uno de los puntos de partida básicos en la Estadística Bayesiana es el concepto de probabilidad (y su definición). Para empezar, pongamos unos ejemplos en que usamos la probabilidad y tratemos de encontrar una definición para esta.

1. ¿Cuál es la probabilidad de que al lanzar una moneda caiga "águila"?
2. ¿Cuál es la probabilidad de que haya un eclipse mañana?
3. ¿Cuál es la probabilidad de que llueva mañana?
4. ¿Cuál es la probabilidad de que esté lloviendo en México?
5. ¿Cuál es la probabilidad de que haya más de 10^9 estrellas en nuestra galaxia?

Se podría pensar que para la primer pregunta la respuesta es 1/2, pero ¿de qué moneda estamos hablando? ¿tiene esta moneda un águila? (¡podría ser extranjera!). La moneda que se está usando, ¿se la dieron de cambio en la tiendita al profesor? o ¿es una moneda que se acaba de sacar del bolsillo un ilusionista?

Para la pregunta 4, yo podría informarme si está lloviendo en México (¿la ciudad o el estado completo o alguna parte del país?) y entonces esta probabilidad sería 0 ó 1; pero, en este instante, ¿qué tanto sabemos del evento "está lloviendo en México"?

Sea A el evento "está lloviendo en México". Una persona que vive a algunos kilómetros de Qaanaaq, sin acceso a internet, no tendría ninguna razón para asignar mayor probabilidad a A que a A^c. Si denotamos por H_1 la información de esta persona, entonces $\mathbb{P}(A|H_1) = \mathbb{P}(A^c|H_1)$.

Una persona que vive en el Bajío posee una información distinta H_2 y podría ser que

$$\mathbb{P}(A|H_2) = \begin{cases} 3/4 & \text{si llueve en el Bajío,} \\ 1/4 & \text{si no llueve en el Bajío.} \end{cases}$$

Por el contrario, si una persona vive en México, poseerá una información H_3, tal que

$$\mathbb{P}(A|H_3) = \begin{cases} 1 & \text{si llueve en México,} \\ 0 & \text{si no llueve en México.} \end{cases}$$

Por lo tanto, la probabilidad de un evento A es una medida del grado de incertidumbre que un individuo (o agente) tiene sobre ese evento. Esto quiere decir que la probabilidad es siempre contextual, dada una serie de supuestos y consideraciones, aun para las probabilidades más simples.

Ejemplo 2

Suponga que una bolsa tiene 4 canicas que pueden ser blancas o negras, pero no sabemos cuántas hay de cada color. Sabemos que hay cinco posibilidades: $\{B, B, B, B\}$, $\{N, B, B, B\}$, $\{N, N, B, B\}$, $\{N, N, N, B\}$, $\{N, N, N, N\}$. Llamemos a estas posibilidades *conjeturas*. Deseamos saber cuál de estas conjeturas es más plausible dada cierta evidencia sobre el contenido de la bolsa.
Como no contamos con más información sobre la plausibilidad de cada conjetura, asignamos una probabilidad de $1/5$ a cada una. Sacamos tres canicas al azar de la bolsa, una a la vez, haciendo la selección con reemplazo y observamos: (N, B, N).

Podemos calcular la probabilidad del evento (N, B, N) bajo cada una de las conjeturas.

Conjetura	Probabilidad Previa	Probabilidad Posterior
$\{B, B, B, B\}$	$1/5$	$\propto 1/5 \times 0$
$\{N, B, B, B\}$	$1/5$	$\propto 1/5 \times 3$
$\{N, N, B, B\}$	$1/5$	$\propto 1/5 \times 8$
$\{N, N, N, B\}$	$1/5$	$\propto 1/5 \times 9$
$\{N, N, N, N\}$	$1/5$	$\propto 1/5 \times 0$

Por lo tanto, la conjetura $\{N, N, N, B\}$ es la más plausible.

La posterior de hoy es la previa del mañana

El futuro no es como era antes

Suponga que sacamos otra canica, y que resulta ser una canica negra. Con esta nueva información podemos actualizar la probabilidad de cada conjetura.

Conjetura	Probabilidad Previa	Probabilidad Posterior
$\{B, B, B, B\}$	$\propto 0$	$\propto 0 \times 0 = 0$
$\{N, B, B, B\}$	$\propto 3$	$\propto 3 \times 1 = 3$
$\{N, N, B, B\}$	$\propto 8$	$\propto 8 \times 2 = 16$
$\{N, N, N, B\}$	$\propto 9$	$\propto 9 \times 3 = 27$
$\{N, N, N, N\}$	$\propto 0$	$\propto 0 \times 4 = 0$

Hay que notar que l**a que antes era la probabilidad posterior ahora se ha vuelto nuestra probabilidad previa**.

Hasta ahora la nueva información ha sido de la misma naturaleza que la anterior. Pero en general, la información previa y la nueva pueden ser de distinto tipo. Suponga, por ejemplo, que alguien de la fábrica de canicas le dice que las canicas negras son raras. Pues por cada bolsa del tipo $\{N, N, N, B\}$ hay dos del tipo $\{N, N, B, B\}$ y tres del tipo $\{N, B, B, B\}$. Además, todas las bolsas contienen al menos una canica negra y una blanca.

Con esta información, actualizamos la probabilidad de cada conjetura

Conjetura	Probabilidad Previa	Probabilidad Posterior
$\{B, B, B, B\}$	$\propto 0$	$\propto 0 \times 0 = 0$
$\{N, B, B, B\}$	$\propto 3$	$\propto 3 \times 3 = 9$
$\{N, N, B, B\}$	$\propto 16$	$\propto 16 \times 2 = 32$
$\{N, N, N, B\}$	$\propto 27$	$\propto 27 \times 1 = 27$
$\{N, N, N, N\}$	$\propto 0$	$\propto 0 \times 0 = 0$

Con la información que ahora contamos, resulta que la conjetura $\{N, N, B, B\}$ es ahora más plausible.

Filosofía bayesiana

De una u otra manera, la subjetividad siempre ha estado presente en la actividad científica, comenzando por los supuestos sobre los cuales se decide abordar un determinado problema, típicamente se les denomina *supuestos razonables*, pero son "razonables" de acuerdo a la experiencia e información (subjetiva) particular de quien o quienes estudian un fenómeno en un momento dado.

De acuerdo a Wolpert (1992):

> *[...] la idea de una "objetividad científica" tiene tan solo un valor limitado, ya que el proceso mediante el cual se generan las ideas [o hipótesis] científicas puede ser bastante subjetivo [...] Es una ilusión creer que los científicos no tienen un vínculo emocional con sus convicciones científicas [...] las teorías científicas implican una continua interacción con otros científicos y el conocimiento previamente adquirido [...] así como una explicación que tenga posibilidades de ser aceptada [o al menos considerada seriamente] por el resto de la comunidad científica.*

De acuerdo a Press (2003) la subjetividad es una parte inherente y requerida para la inferencia estadística, y para el *método científico*. Sin embargo, ha sido la manera informal y desorganizada en la que se ha permitido la presencia de la subjetividad, la responsable de diversos errores y malas interpretaciones en la historia de la ciencia. La estadística bayesiana incorpora de manera formal y fundamentada la información subjetiva.

Al hablar de *información subjetiva* nos referimos a toda aquella información previa que se tiene sobre el fenómeno aleatorio de interés, antes de recolectar o realizar nuevas mediciones sobre el mismo, incluyendo: datos históricos, teorías, opiniones y conjeturas de expertos, conclusiones basadas en estudios previos, etc.

Mantras

En resumen:

- La probabilidad de un evento A es una medida del grado de creencia que tiene un individuo en la ocurrencia de A con base en la información H que posee. Por lo tanto, toda probabilidad es condicional (a las circunstancias, el *agente*, etc.)
- Sea $\mathcal{P} = \{p(Y|\theta), \theta \in \Theta\}$ una familia paramétrica de distribuciones. Como toda incertidumbre debe ser medida a través de una probabilidad y contamos con incertidumbre en los parámetros, entonces los parámetros deben ser modelados con una medida de probabilidad (una distribución).

Modelos Uniparamétricos

Consideremos Y una variable aleatoria, cuya verosimilitud, cuando toma el valor y y los parámetros θ toman el valor t, está dada por

$$p_{Y|\theta}(y|t).$$

Por ejemplo, si Y es una variable aleatoria normal con media θ y con varianza unitaria, entonces

$$p_{Y|\theta}(y|t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-t)^2}{2}\right\} \mathbb{1}_{y\in\mathbb{R}}.$$

En el argot bayesiano se suele hacer uso indiscriminado de abusos, a veces hasta peligrosos, de notación, por ejemplo

$$p_{Y|\theta}(y|t)$$

los escribiríamos como

$$p(Y|\theta).$$

Notación

- Variables aleatorias: Y, θ.
- Espacio muestral: $\mathcal{Y}$, $Y \in \mathcal{Y}$.
- Espacio paramétrico: Θ, $\theta \in \Theta$.
- Muestra aleatoria: $\mathbf{Y} = (Y_1, \ldots, Y_n)$.
- Muestra observada: $\mathbf{y} = (y_1, \ldots, y_n)$.
- Distribución previa o *a priori*: $p(\theta)$.
- Verosimilitud: $p(Y|\theta)$.
- Verosimilitud de la muestra: $p(\mathbf{Y}|\theta)$.
- Evidencia: $p(\mathbf{Y})$.
- Distribución posterior o *a posteriori*: $p(\theta|\mathbf{Y})$.

Distribución posterior

La distribución de probabilidad previa se trata de una distribución basada en experiencia previa (experiencia de especialistas, datos históricos, etc.), antes de obtener datos muestrales nuevos.

Luego procedemos a observar los nuevos datos (obtención de evidencia) y combinamos esta información con la distribución previa mediante la Regla de Bayes y obtenemos una distribución de probabilidad posterior:

$$p_{\theta|\mathbf{Y}}(t|\mathbf{y}) = \frac{p_{\mathbf{Y}|\theta}(\mathbf{y}|t)p_\theta(t)}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{p_{\mathbf{Y}|\theta}(\mathbf{y}|t)p_\theta(t)}{\int_\Theta p_{\mathbf{Y}|\theta}(\mathbf{y}|\tilde{t})p_\theta(\tilde{t})d\tilde{t}}$$

Usando los abusos usuales de la Estadística Bayesiana, el resultado anterior lo escribiríamos como:

$$p(\theta|\mathbf{Y}) = \frac{p(\mathbf{Y}|\theta)p(\theta)}{p(\mathbf{Y})} = \frac{p(\mathbf{Y}|\theta)p(\theta)}{\int_{\Theta} p(\mathbf{Y}|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}$$

o también como:

$$p(\theta|\mathbf{Y}) \propto p(\mathbf{Y}|\theta)p(\theta).$$

Beta-Bernoulli

Supongamos que $Y_i|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$ y que la incertidumbre acerca de $\theta \in (0,1)$ la cuantificamos mediante $\theta \sim \mathsf{Beta}(\alpha, \beta)$. Obtenemos $\mathbf{y} = (y_1, \ldots, y_n)$, entonces

$$p(\mathbf{Y}|\theta) = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(y_i)$$

y

$$p(\theta) = B(\alpha,\beta)^{-1}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{1}_{(0,1)}(\theta).$$

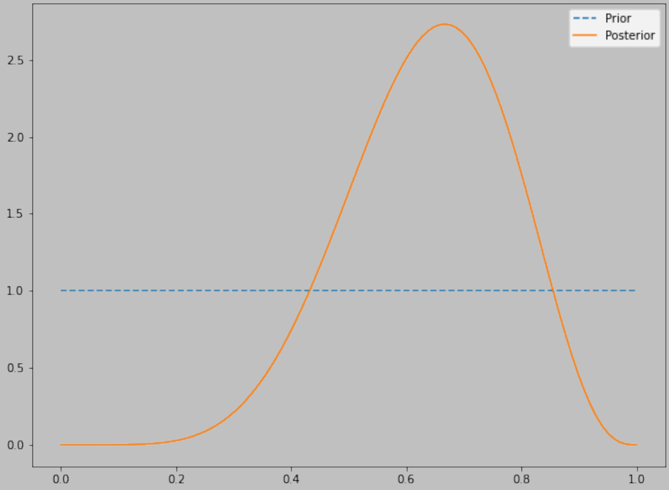
Por lo tanto,

$$p(\theta|\mathbf{Y}) \propto \theta^{\alpha+\sum_{i=1}^n y_i-1}(1-\theta)^{\beta+n-\sum_{i=1}^n y_i-1}\mathbb{1}_{(0,1)}(\theta),$$

es decir $\theta|\mathbf{Y} \sim \mathsf{Beta}\left(\alpha + \sum_{i=1}^n y_i, \beta + n - \sum_{i=1}^n y_i\right)$.

Ejemplo 3

Suponga que $\mathbf{Y}|\theta \sim \mathsf{Bernoulli}(\theta)$ y que no contamos con información sobre θ como para preferir un valor sobre otro, así que modelamos $\theta \sim \mathsf{Beta}(1,1)$. Se obtiene una muestra aleatoria simple (i.e. se obtienen variables i.i.d.) $\mathbf{y} = (1,0,1,1,1,0,1,0,1)$. A continuación se muestra la función de densidad previa y posterior para θ.


Prior
Posterior
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Beta-Binomial

Recuerde que si $Y_1 \ldots, Y_n|\theta \overset{iid}{\sim} \mathsf{Bernoulli}(\theta)$, entonces

$$Z = \sum_{i=1}^{n} Y_i|\theta \sim \mathsf{Binomial}(n, \theta).$$

Luego, si

$$\theta \sim \mathsf{Beta}(\alpha, \beta),$$

concluimos que

$$\theta|Z \sim \mathsf{Beta}(\alpha + Z, \beta + n - Z).$$

Principio de indiferencia

Al analizar el modelo Binomial, Laplace consideró la distribución Uniforme como previa, argumentando lo que llamó el *principio de indiferencia* también llamado *principio de razón insuficiente*, que establece que el supuesto de uniformidad es apropiado cuando no se sabe nada sobre θ, i.e.

$$\theta \sim \mathsf{Beta}(1,1),$$
$$Y|\theta \sim \mathsf{Binomial}(n,\theta)$$
$$\Rightarrow \theta|Y \sim \mathsf{Beta}(Y+1, n-Y+1).$$

Una de las primeras aplicaciones de Laplace fue el estimar la proporción θ de nacimientos de mujeres en una población. Laplace obtuvo que entre 1745 y 1770 nacieron 241945 niñas y 251527 niños en París, si Y denota el número de nacimiento de niñas

$$\Rightarrow \theta | Y \sim \mathsf{Beta}(241946, 251528)$$