

Estadística Bayesiana

Irving Gómez Méndez



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:

$$\begin{aligned}\mathbb{P}(E = 1|T = 1) &= \frac{\mathbb{P}(T = 1|E = 1)\mathbb{P}(E = 1)}{\mathbb{P}(T = 1|E = 1)\mathbb{P}(E = 1) + \mathbb{P}(T = 1|E = 0)\mathbb{P}(E = 0)} \\ &= 0.9262\end{aligned}$$

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

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

Por lo tanto, la probabilidad de un evento A es una medida del grado de incertidumbre que un individuo 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 la **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 $\mathcal{P} = \{f_{X|\theta}(x|p), p \in \Theta\}$ 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).

Notación

- ▶ Variables aleatorias: X, θ .
- ▶ Valores observados: x, p .
- ▶ Espacio muestral: $\mathcal{X}, X \in \mathcal{X}$.
- ▶ Espacio paramétrico: $\Theta, \theta \in \Theta$.
- ▶ Muestra aleatoria: $\mathbf{X} = (X_1, \dots, X_n)$.
- ▶ Muestra observada: $\mathbf{x} = (x_1, \dots, x_n)$.
- ▶ Distribución previa o *a priori*: $f_{\theta}(p)$.
- ▶ Verosimilitud: $f_{X|\theta}(x|p)$.
- ▶ Verosimilitud observada: $f_{\mathbf{X}|\theta}(\mathbf{x}|p)$.
- ▶ Evidencia: $f_{\mathbf{X}}(\mathbf{x})$.
- ▶ Distribución posterior o *a posteriori*: $f_{\theta|\mathbf{X}}(p|\mathbf{x})$.

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:

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

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

$$f_{\theta|\mathbf{X}}(p|\mathbf{x}) = \frac{f_{\mathbf{X}|\theta}(\mathbf{x}|p)f_{\theta}(p)}{f_{\mathbf{X}}(\mathbf{x})}$$

los escribiríamos como

$$f(\theta|\mathbf{X}) \propto f(\mathbf{X}|\theta)f(\theta).$$

Beta-Bernoulli

Supongamos que $X_i|\theta = p \stackrel{iid}{\sim} \text{Ber}(p)$ y que la incertidumbre acerca de $\theta \in [0, 1]$ la cuantificamos con $\theta \sim \text{Beta}(\alpha, \beta)$.

Obtenemos $\mathbf{x} = (x_1, \dots, x_n)$, entonces

$$f(\mathbf{x}|p) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(x_i)$$

y

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

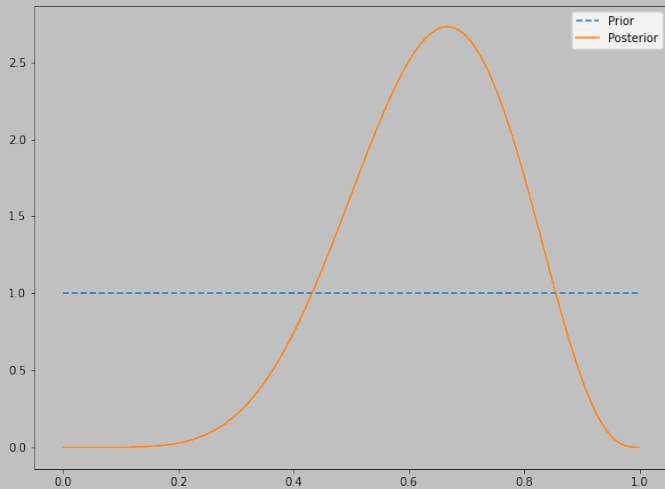
Por lo tanto

$$f(p|\mathbf{x}) \propto p^{\alpha+\sum_{i=1}^n x_i-1} (1-p)^{\beta+n-\sum_{i=1}^n x_i-1} \mathbb{1}_{[0,1]}(p),$$

es decir $\theta|\mathbf{X} = \mathbf{x} \sim \text{Beta}(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$.

Ejemplo 3

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



Ejercicio 1

Suponga que una particular población de células puede estar en uno de los siguientes tres estados de producción de una cierta proteína. Los estados son A, B y C, de producción baja, media y alta. Se toma una muestra al azar de 20 células, dentro de cierta población y se verifica si cada una de estas está en producción de la proteína (el resultado del aparato es si o no: 1 o 0, por cada célula analizada). De esta muestra resultan 12 células en producción (1) y las demás en negativo (0).

Por otro lado sabemos que si la población está en el estado A, solo el 20% de la células producen la proteína, si está en el estado B el 50% de las células la producen y si está en el estado C el 70% la produce.

¿Cuál es la probabilidad de que la población esté en cada uno de estos estados?

Distribuciones predictivas

Muchos son los casos en los que, más que estar interesados en el vector de parámetros θ , lo que queremos es describir el comportamiento de observaciones futuras del fenómeno aleatorio en cuestión, esto es, hacer predicción.

Por lo regular la estadística frecuentista aborda este problema estimando puntualmente a θ con base en la muestra observada y dicho estimador \hat{p} es sustituido en $f_{X|\theta}(x|p)$, es decir, se utiliza $f_{X|\theta}(x|\hat{p})$.

En la estadística bayesiana este problema se aborda marginalizando la distribución conjunta de θ y X .

- Distribución predictiva previa o *a priori*:

$$f_X(x) = \int_{\Theta} f_{X|\theta}(x|p) f_{\theta}(p) dp$$

Una vez obtenida una muestra \mathbf{x} se induce una distribución conjunta para X y θ condicional en la muestra,

$$\begin{aligned} f_{X,\theta|\mathbf{X}}(x, p|\mathbf{x}) &= \frac{f_{X,\theta,\mathbf{X}}(x, p, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \\ &= f_{X|\theta,\mathbf{X}}(x|p, \mathbf{x}) \frac{f_{\theta,\mathbf{X}}(p, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \\ &= f_{X|\theta}(x|p) f_{\theta|\mathbf{X}}(p|\mathbf{x}) \end{aligned}$$

- Distribución predictiva (posterior o *a posteriori*):

$$f_{X|\mathbf{X}}(x|\mathbf{x}) = \int_{\Theta} f_{X|\theta}(x|p) f_{\theta|\mathbf{X}}(p|\mathbf{x}) dp$$

Ejemplo 4

Consideremos una urna que contiene dos monedas: una cargada y la otra equilibrada. Supongamos que la moneda cargada está construida para tener una probabilidad de $3/4$ de que salga águila. Una persona tomará una de las dos monedas de la urna (no necesariamente al azar) y echará volados.

En este caso podemos definir el espacio medible como $\Omega = \{\text{“Sale águila”}, \text{“Sale sol”}\}$ y tomar como sigma-álgebra el conjunto potencia de Ω , $\mathcal{P}(\Omega)$. En este espacio definimos la variable aleatoria

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = \text{“Sale águila”}, \\ 0 & \text{si } \omega = \text{“Sale sol”}. \end{cases}$$

Implícitamente estamos eliminando como posibles resultados eventos como que la moneda caiga vertical, que no sepamos el resultado del volado, etc.

$$\Theta = \left\{ \frac{3}{4}, \frac{1}{2} \right\}, \mathcal{X} = \{0, 1\}, X|\theta = p \sim \text{Ber}(p).$$

Información previa:

$$\mathbb{P}\left(\theta = \frac{3}{4}\right) = \alpha, \mathbb{P}\left(\theta = \frac{1}{2}\right) = 1 - \alpha, \alpha \in (0, 1).$$

$$f_{\theta}(p) = \alpha \mathbb{1}_{\{3/4\}}(p) + (1 - \alpha) \mathbb{1}_{\{1/2\}}(p).$$

Verosimilitud:

$$f_{X|\theta}(x|p) = p^x(1 - p)^{1-x} \mathbb{1}_{\{0,1\}}(x).$$

Predictiva previa:

$$\begin{aligned}f_X(x) &= \sum_{p \in \Theta} f_{X|\theta}(x|p) f_{\theta}(p) \\&= \alpha \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{1-x} \mathbb{1}_{\{0,1\}}(x) + (1 - \alpha) \left(\frac{1}{2}\right) \mathbb{1}_{\{0,1\}}(x) \\&= \alpha \left[\left(\frac{3}{4} - \frac{1}{2}\right) \mathbb{1}_{\{1\}}(x) + \left(\frac{1}{4} - \frac{1}{2}\right) \mathbb{1}_{\{0\}}(x) \right] + \frac{1}{2} \mathbb{1}_{\{0,1\}}(x) \\&= \frac{\alpha}{4} \left[\mathbb{1}_{\{1\}}(x) - \mathbb{1}_{\{0\}}(x) \right] + \frac{1}{2} \mathbb{1}_{\{0,1\}}(x).\end{aligned}$$

Es decir,

$$f_X(1) = \frac{1}{2} + \frac{\alpha}{4}, \quad f_X(0) = \frac{1}{2} - \frac{\alpha}{4}.$$

Verosimilitud observada: Sean $X_1, \dots, X_n | \theta = p \stackrel{iid}{\sim} \text{Ber}(p)$, entonces

$$\begin{aligned} f_{\mathbf{X}|\theta}(\mathbf{x}|p) &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(x_i) \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} g(\mathbf{x}). \end{aligned}$$

Evidencia:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \sum_{p \in \Theta} f_{\mathbf{X}|\theta}(\mathbf{x}|p) f_{\theta}(p) \\ &= \alpha \left(\frac{3}{4}\right)^{\sum_{i=1}^n x_i} \left(\frac{1}{4}\right)^{n-\sum_{i=1}^n x_i} g(\mathbf{x}) + (1-\alpha) \left(\frac{1}{2}\right)^n g(\mathbf{x}) \\ &= \left[\alpha \frac{3^{\sum_{i=1}^n x_i}}{4^n} + (1-\alpha) \frac{1}{2^n} \right] g(\mathbf{x}). \end{aligned}$$

Posterior:

$$\begin{aligned} f_{\theta|\mathbf{X}}(p|\mathbf{x}) &= \frac{f_{\mathbf{x}|\theta}(\mathbf{x}|p)f_{\theta}(p)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \left[\alpha \mathbb{1}_{\{3/4\}}(p) + (1-\alpha) \mathbb{1}_{\{1/2\}}(p) \right]}{\alpha 3^{\frac{\sum_{i=1}^n x_i}{4^n}} + (1-\alpha) \frac{1}{2^n}} \\ &= \frac{[2(1-p)]^n \left(\frac{p}{1-p} \right)^{\sum_{i=1}^n x_i} \left[\alpha \mathbb{1}_{\{3/4\}}(p) + (1-\alpha) \mathbb{1}_{\{1/2\}}(p) \right]}{1-\alpha + \alpha \left(\frac{3^{\sum_{i=1}^n x_i}}{2^n} \right)}. \end{aligned}$$

Sea $v = \frac{\alpha}{1-\alpha} \left(\frac{3 \sum_{i=1}^n x_i}{2^n} \right)$, entonces

$$f_{\theta|\mathbf{X}}(p|\mathbf{x}) = \frac{[2(1-p)]^n \left(\frac{p}{1-p} \right)^{\sum_{i=1}^n x_i} \left[\alpha \mathbb{1}_{\{3/4\}}(p) + (1-\alpha) \mathbb{1}_{\{1/2\}}(p) \right]}{(1-\alpha)(1+v)},$$

es decir

$$f_{\theta|\mathbf{X}}(1/2|\mathbf{x}) = \frac{1}{1+v}$$

y

$$f_{\theta|\mathbf{X}}(3/4|\mathbf{x}) = 1 - \frac{1}{1+v} = \frac{v}{1+v} = \frac{1}{1+v^{-1}}.$$

Predictiva (posterior)

$$\begin{aligned}f_{X|\mathbf{X}}(x|\mathbf{x}) &= \sum_{p \in \Theta} f_{X|\theta}(x|p) f_{\theta|\mathbf{X}}(p|\mathbf{x}) \\&= \left[\frac{1}{2} \left(\frac{1}{v+1} \right) + \frac{3^x}{4} \left(\frac{1}{v^{-1}+1} \right) \right] \mathbb{1}_{\{0,1\}}(x) \\&= \left[\frac{2}{4(v+1)} + \frac{3^x v}{4(v+1)} \right] \mathbb{1}_{\{0,1\}}(x) \\&= \frac{3^x v + 2}{4(v+1)} \mathbb{1}_{\{0,1\}}(x).\end{aligned}$$

Resumen de ejemplo 4

Previa:

$$f_{\theta}(p) = \alpha \mathbb{1}_{\{3/4\}}(p) + (1 - \alpha) \mathbb{1}_{\{1/2\}}(p), \quad \alpha \in (0, 1)$$

Verosimilitud:

$$f_{X|\theta}(x|p) = p^x(1-p)^{(1-x)} \mathbb{1}_{\{0,1\}}(x)$$

Predictiva previa:

$$f_X(1) = \frac{1}{2} + \frac{\alpha}{4}, \quad f_X(0) = \frac{1}{2} - \frac{\alpha}{4}$$

Posterior:

$$f_{\theta|\mathbf{X}}(1/2|\mathbf{x}) = \frac{1}{1+v}, \quad f_{\theta|\mathbf{X}}(3/4|\mathbf{x}) = \frac{1}{1+v^{-1}},$$

$$v = \frac{\alpha}{1-\alpha} \left(\frac{3 \sum_{i=1}^n x_i}{2^n} \right)$$

Predictiva (posterior):

$$f_{X|\mathbf{X}}(1|\mathbf{x}) = \frac{3v+2}{4(v+1)}, \quad f_{X|\mathbf{X}}(0|\mathbf{x}) = \frac{v+2}{4(v+1)}$$

Tarea 1

- ▶ Dejar tarea 1.
- ▶ Mostrar ejemplo Beta-Bernoulli.
- ▶ Mostrar ejemplo moneda cargada.

No hay torta gratis

Cuando se hace inferencia estadística frecuentista, los procedimientos suelen justificarse por el comportamiento asintótico del método. Como consecuencia, su desempeño con muestras pequeñas es cuestionable. Al contrario, la estadística bayesiana es válida para cualquier tamaño de muestra. Esto no quiere decir que tener más datos no sea útil, todo lo contrario.

El precio que hay que pagar por este poder es la dependencia en la información previa. Si se tiene una mala previa, los resultados no serán confiables.

Discusión sobre la previa

Históricamente, algunos detractores de la estadística bayesiana argumentan sobre la arbitrariedad de la distribución previa. Es cierto que las distribuciones previas son lo suficientemente flexibles para codificar distinto tipo de información. Entonces, si la previa puede ser cualquier cosa, ¿no es posible obtener cualquier respuesta que quieras? De hecho sí.

Sin embargo, si tu objetivo es mentir usando la estadística es una tontería hacerlo a través de la previa, pues dicha mentira estaría rápidamente cubierta. Es más fácil modificar la inferencia cambiando la verosimilitud.

Ejercicio 2

Sean $X_1, \dots, X_n | \mu, \sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ y sea $\lambda = \sigma^{-2}$. A λ se le conoce como la “precisión”. Haciendo abuso de notación diremos que $X_1, \dots, X_n | \mu, \lambda \stackrel{iid}{\sim} \mathcal{N}(\mu, \lambda)$.

Suponga que λ es conocida y considere la distribución previa para $\mu \sim \mathcal{N}(\mu_0, \lambda_0)$. Obtenga la distribución posterior de μ .

Regresión (Normal-Inversa Gama)

Detalles sangrientos



Cómo evitar a Procrustes

Un modelo bayesiano es una máquina que toma de entrada la distribución previa de los parámetros y la verosimilitud, y usando el teorema de Bayes como motor produce la distribución posterior. Sin embargo, saber la regla matemática del funcionamiento del motor suele ser de muy poca ayuda. Restringirse únicamente a aquellos modelos que permiten la manipulación matemática es una solución *procrustea*.

Ante este problema es necesario recurrir a alguna técnica numérica que permita aproximar la manipulación matemática.

Aproximación usando una rendija

Una solución sencilla cuando se tienen pocos parámetros (típicamente uno o dos) continuos consiste en generar una rendija de valores para los parámetros. Sea p_j alguno de estos valores, entonces se puede calcular la distribución posterior en p_j (salvo por una constante de proporcionalidad) usando la fórmula:

$$f_{\theta|\mathbf{X}}(p_j|\mathbf{x}) \propto f_{\mathbf{X}|\theta}(\mathbf{x}|p_j)f_{\theta}(p_j).$$

Una importante consecuencia de este hecho es que podemos generar una muestra de la distribución posterior a partir de la rendija de valores propuesta, simplemente basta con seleccionar el valor p_j de manera proporcional a $f_{\mathbf{X}|\theta}(\mathbf{x}|p_j)f_{\theta}(p_j)$.

Simular de la predictiva posterior

También nos puede interesar generar una muestra de la distribución predictiva. Una vez que se cuenta con una muestra de la distribución posterior de los parámetros, p_1, \dots, p_m , se puede generar una muestra x_1, \dots, x_m de la distribución predictiva posterior. Simplemente hay que simular $\mathbf{X}_j \sim f_{X|\theta}(x|p_j)$.

Pruebas de hipótesis

Algunas veces la inferencia estadística puede ser formulada como:

1. Se cuenta con una hipótesis, la cual puede ser cierta o falsa ($H : \theta \in \Theta_1$).
2. Se obtiene evidencia estadística sobre la falsedad de la hipótesis.
3. Usamos (o deberíamos usar) el teorema de Bayes para deducir de manera lógica el impacto de la evidencia en la hipótesis

$$\mathbb{P}(H|\mathbf{x}) = \mathbb{P}(\theta \in \Theta_1|\mathbf{x}) = \int_{\Theta_1} f(p|\mathbf{x})dp.$$

Estimación por intervalo

Si se cuenta con una muestra de la distribución posterior p_1, \dots, p_m , se puede estimar $\mathbb{P}(\theta \in \Theta_1 | \mathbf{x}) = \mathbb{E} [\mathbb{1}_{\theta \in \Theta_1} | \mathbf{x}]$ mediante

$$\frac{1}{m} \sum_{j=1}^m \mathbb{1}_{p_j \in \Theta_1}.$$

También se vuelve sencillo estimar intervalos (θ_1, θ_2) tales que $\mathbb{P}(\theta \in (\theta_1, \theta_2)) = 1 - \alpha$. A estos intervalos se les llama intervalos de credibilidad. Un intervalo de particular interés es el de menor longitud cuya probabilidad es $1 - \alpha$ (highest posterior density interval, HPDI).

Estimación puntual (MAP)

Recuerde que el estimador bayesiano consiste en toda la distribución posterior. Sin embargo, a veces nos es requerido reportar un único valor. En este caso es común reportar el valor más probable a posteriori (maximum a posteriori, MAP). Lamentablemente, dicho estimador puntual puede dar lugar a resultados absurdos.

Ejemplo 5

Considere el ejercicio, planteado en la tarea, del globo terráqueo. Suponga que en 3 lanzamientos se obtiene AAA, en este caso el MAP vale 1. Lo que es un resultado absurdo.

Estimación puntual

En vez de reportar el MAP se podría optar por la media o la mediana de la distribución posterior, pero entonces surge la pregunta de qué estimador puntual es el que deberíamos de reportar. Una manera de tomar esta decisión es a través del uso de alguna función de pérdida $L(p, p_0)$, donde p_0 es el verdadero parámetro. Pero, ¡el verdadero parámetro es desconocido!

Para solucionar este inconveniente se minimiza la perdida esperada, tomando el valor esperado con respecto a la distribución posterior. Es decir se calcula

$$\begin{aligned} L(p) &= \mathbb{E}_{\tilde{p} \sim f_{\theta|\mathbf{x}}} [L(p, \tilde{p})] \\ &= \int_{\Theta} L(p, \tilde{p}) f_{\theta|\mathbf{x}}(\tilde{p}|\mathbf{x}) d\tilde{p} \end{aligned}$$

y se selecciona $p \in \arg \min_{p \in \Theta} L(p)$.

Si se cuenta con muestra p_1, \dots, p_m de la distribución posterior, entonces $L(p)$ puede ser estimado mediante

$$\frac{1}{m} \sum_{j=1}^m L(p, p_j)$$

► Mostrar ejemplo Beta-binomial.

Cómo determinar qué función de pérdida usar

Considere el caso en que se requiere decidir si ordenar una evacuación o no con base en la velocidad del viento provocado por un huracán.

El riesgo a que haya fallecidos y/o personas afectadas aumenta rápidamente conforme aumenta la rapidez del viento. Sin embargo, también se induce un costo al ordenar una evacuación innecesaria, aunque este es mucho menor.

Por lo tanto, se debería usar una función de pérdida muy asimétrica, que crece rápidamente cuando la velocidad del viento excede nuestra inferencia, pero crece lentamente cuando la velocidad del viento es menor que nuestra inferencia.

Breve comentario sobre las pruebas de hipótesis

Retomando nuestra discusión sobre las pruebas de hipótesis. De manera más general, lo que se desea es calcular

$$\mathbb{P}(H|\text{evidencia}) = \frac{\mathbb{P}(\text{evidencia}|H)\mathbb{P}(H)}{\mathbb{P}(\text{evidencia})}.$$

Lo más importante es aumentar $\mathbb{P}(H)$, lo cual requiere un esfuerzo cognitivo y argumentativo y no se limita a una simple prueba estadística.

Simular de la predictiva previa

Note que también se puede simular una muestra de la predictiva previa, para ello basta con simular una muestra de los parámetros p_1, \dots, p_m , a partir de la distribución previa y luego simular $\mathbf{X}_j \sim f_{X|\theta}(x|p_j)$.

Poder simular de la predictiva previa puede ayudar a discriminar entre distintas distribuciones previas. Pues al simular de la predictiva previa podemos observar las consecuencias de las distribución previa sobre la variable de interés X . Muchas de las técnicas convencionales para decidir la distribución previa pueden dar lugar a distribuciones más bien absurdas.

- Mostrar ejemplo de la estatura de la población !Kung.

Regresión de la estatura en el peso

Para estos ejercicios vamos a usar los datos de la comunidad !Kung. Consideraremos sólo a los adultos, esto porque la estatura está fuertemente relacionada con la edad hasta antes de la adultez.

Sea h_i la estatura del i -ésimo individuo, w_i su peso y \bar{w} el peso promedio de los adultos. Vamos a considerar el modelo:

$$h_i \sim \mathcal{N}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta(w_i - \bar{w})$$

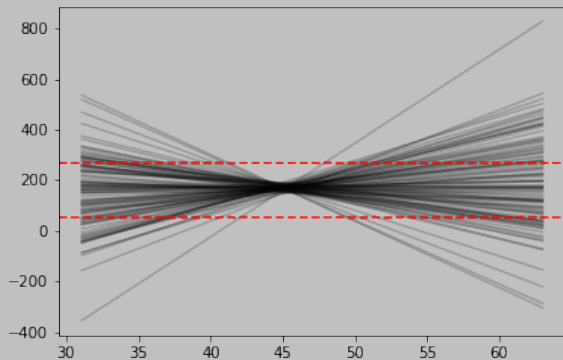
$$\alpha \sim \mathcal{N}(170, 10)$$

$$\beta \sim \mathcal{N}(0, 10)$$

$$\sigma \sim \mathcal{U}(0, 33)$$

Note cuando $w_i = \bar{w}$, $\mu_i = \alpha$. Es decir cuando el peso es igual al promedio de la población la estatura promedio será α , así que hace mucho sentido modelar $\alpha \sim \mathcal{N}(170, 10)$. Para σ simplemente usamos una distribución con poca información. Pero **¿por qué hemos puesto esa previa para β ?**

Una manera de verificar qué tan sensatas son estas previas es simulando valores de α y β , $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m$ y luego graficar $\alpha_j + \beta_j(w_j - \bar{w})$, $j = 1, \dots, m$.



¡Evidentemente esta previa para β es absurda!

Cualquiera sabe que el peso y la estatura guardan (hasta cierto punto) una correlación positiva, por lo que es sensato considerar una distribución previa estrictamente positiva para β . Por lo que consideramos el siguiente modelo:

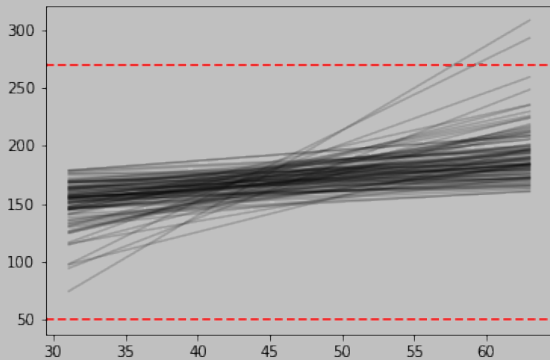
$$h_i \sim \mathcal{N}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta(w_i - \bar{w})$$

$$\alpha \sim \mathcal{N}(170, 10)$$

$$\beta \sim \text{lognormal}(0, 1)$$

$$\sigma \sim \mathcal{U}(0, 33)$$



Estimador MAP

Una vez establecido el modelo podemos crear una rendija de valores para α , β y σ . Y como antes, simular una muestra de la posterior. Con esta muestra podemos estimar puntualmente los parámetros y graficar la recta con mayor probabilidad a posteriori $\hat{\alpha} + \hat{\beta}(w - \bar{w})$.

Intervalos de confianza

Como μ depende de los parámetros y estos tienen una distribución posterior, entonces μ también cuenta con una distribución posterior. En particular fijando un peso w , y contando con una muestra a posteriori $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m$ podemos obtener una muestra a posteriori μ_1, \dots, μ_m usando la relación:

$$\mu_j = \alpha_j + \beta_j(w - \bar{w}).$$

A partir de esta muestra podemos obtener un intervalo de confianza para la estatura media dado el peso w . Variando w desde el valor más bajo hasta el valor más alto podemos obtener un intervalo de confianza para la función de regresión.

Intervalos de predicción

Finalmente, para agregar intervalos de predicción hay que recordar que $h \sim \mathcal{N}(\mu, \sigma)$. Así para un peso fijo w podemos obtener una muestra de la posterior de μ como se explicó antes, μ_1, \dots, μ_m . Además podemos simular una muestra de la posterior de σ , $\sigma_1, \dots, \sigma_m$. Por lo tanto podemos obtener una muestra de la posterior de h , h_1, \dots, h_m , donde $h_j \sim \mathcal{N}(\mu_j, \sigma_j)$. A partir de la muestra h_1, \dots, h_m podemos obtener un intervalo de predicción para el peso w . Variando w desde el valor más bajo hasta el valor más alto podemos obtener un intervalo de predicción para la función de regresión.

- Mostrar ejemplo de regresión de la estatura de la población !Kung.