3ª Emisión

Módulo 03

INFERENCIA BAYESIANA

Dr. Roberto Bárcenas Curtis

DGTIC

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Dirección General de Cómputo y de Tecnologías de información y Comunicación
Dirección de Docencia en TIC

50 AÑOS
Educación Continua
1971 - 2021

Presentación

La estadística bayesiana es un enfoque que recientemente ha adquirido relevancia en diversas aplicaciones.

Una razón es la manera en la que permite asignar probabilidades a determinados sucesos y actualizar el papel que juega la información previa sobre el valor de un parámetro, a la luz de nueva evidencia (datos).

Objetivo

El participante obtendrá un panorama del enfoque de inferencia bayesiana y lo relacionará con la idea de modelación estadística en general, para ser capaz de realizar una estimación de la distribución posterior de un parámetro.

Contenido

4. INFERENCIA BAYESIANA

4. 1 Especificación del modelo muestral

4.2 Distribución inicial o *a priori*

4.3 Distribución final o *a posteriori*

4.4 Actualización del conocimiento previo

Cuestiones clave en inferencia estadística

1. Estudiar la verosimilitud de una hipótesis:

Validez de una hipótesis *H* en función de los datos obtenidos.

Por ejemplo: *H*: efecto del tratamiento *A* =efecto del tratamiento *B*,

H: costo del tratamiento *A* - Costo del tratamiento *B* > 0

1. Estimar el valor de un parámetro(θ):

Por ejemplo: ¿cuál es la mejor estimación para la tasa de supervivencia de un tratamiento, cuál es su costo medio anual o qué varianza tiene la distribución de la función de costos?

Enfoque de Fisher

Fisher propuso valorar una hipótesis (H_0) a través de una cantidad concreta (d_0), así como la construcción de “p-valor” (p-value):

$$p = Prob(d \geq d_0 \mid H_0)$$

que es la probabilidad de observar algo mayor o igual que lo que objetivamente se observó, suponiendo que sea válida la hipótesis que se valora.

Así, Fisher lo estableció como una medida de la discrepancia de los datos con la hipótesis.

Antecedentes

Probabilidades condicionales

$$\Pr(A, B) = \Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A)$$

Teorema de Bayes:

$$\Pr(A \mid B) \cdot \Pr(B) = \Pr(A, B)$$

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}$$

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}$$

De acuerdo a Bayes

Sean "*H*" = H_0 y "*Datos*"=datos observados

$$\Pr(H \mid Datos) = \frac{\Pr(Datos \mid H) \cdot \Pr(H)}{\Pr(Datos)}$$

donde:

- Pr(H) es la probabilidad *a priori* de que la hipótesis *H* sea cierta.
- Pr(H|Datos) es la probabilidad *a posteriori* de que la hipótesis *H* sea cierta, una vez que se han observado los datos.
- Pr(Datos|H) es la verosimilitud, es decir, la probabilidad de haber observado esos datos si la hipótesis *H* es cierta.
- Pr(Datos) es la probabilidad de haber observado estos datos independientemente de que *H* sea cierta o no.

Utilidad de la inferencia Bayesiana

1. Dado un modelo, la inferencia bayesiana hace uso de los datos obtenidos empíricamente.
2. Inferencias que resulten inaceptables, han de ser consecuencia de afirmaciones inapropiadas, no de un inadecuado sistema de inferencia. Por tanto, todas los factores del modelo, incluidas las distribuciones a *priori*, son susceptibles de ser modificadas.

Regla de Bayes

Describe el proceso de aprendizaje a partir de la experiencia, y muestra cómo el conocimiento del estado del mundo físico, representado por los parámetros, está continuamente en cambio según aparecen nuevos datos.

Probabilidad subjetiva

Es una probabilidad (entre cero y uno) de que un evento ocurra, cuantificando la opinión particular (subjetiva) de una persona, acerca de cuán probable es que este evento ocurra o haya ocurrido.

Visión personal ante el mismo fenómeno

No sólo diferentes personas tendrán diferentes probabilidades subjetivas de un mismo suceso, sino que incluso la misma persona puede cambiar su probabilidad subjetiva al tiempo que se obtiene más información.

Estimaciones puntuales

Desde el punto de vista Bayesiano, el cálculo de un estimador $\hat{\theta}$ de un parámetro θ está basado en la distribución *a posteriori*. Lo más conveniente es estimar parámetros de localización, como la media o la moda.

- Esperanza: $E(\theta|\ \underline{x}) = \int_{-\infty}^{\infty} \theta f(\theta|\ \underline{x}) d\theta$
- Moda: $Mod(\theta|\ \underline{x}) = arg\,\text{máx}_\theta\, f(\theta|\ \underline{x})$
- Mediana: el valor a que satisfece:

$$\int_{-\infty}^{a} f(\theta|\underline{x})d\theta = \frac{1}{2} \text{ y } \int_{a}^{\infty} f(\theta|\underline{x})d\theta = \frac{1}{2}$$

Intervalos de credibilidad

En el modelo bayesiano se utilizan los llamados intervalos de credibilidad. Estos se construyen de la siguiente manera:

Habiendo observado *datos*, se quiere encontrar dos valores $[a_1,a_2]$, de tal manera que $\theta \in [a_1,a_2]$ con una probabilidad alta, o sea,

$$P(a_1 < \theta < a_2) \mid x) \geq 1 - \alpha,$$

siendo $1-\alpha$ el nivel de credibilidad.

Esta definición de intervalos de credibilidad presenta un inconveniente;

Hay muchos valores de a_1 y a_2 que cumplen esta propiedad. ¿Entonces, cómo elegir estos dos valores? Una buena solución es considerar el intervalo $C = \{ \theta : p (\theta \mid x) > \gamma \}$ donde γ es el mayor número real que satisface $p(\theta \in C \mid x) \geq 1 - \alpha$.

Esta regla da origen a un intervalo de amplitud mínima, de modo que cualquier punto excluido del intervalo no tenga mayor credibilidad que cualquier punto incluido en él.

Estos intervalos se conocen como intervalos HPD (highest posterior density).

Función Pérdida

Para determinar qué estimación tomar en cada caso específico, definiremos primero la función de pérdida:

La función de pérdida $L(\hat{\theta}, \theta) \in R$ cuantifica el error obtenido al estimar el parámetro θ con $\hat{\theta}$.

Si el estimador es igual al parámetro, entonces a la función de pérdida se le asocia el valor cero: $L(\hat{\theta}, \theta) = 0$

Funciones de pérdida usadas comúnmente

- Función de pérdida cuadrática: $\mathcal{L}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$
- Función de pérdida lineal: $\mathcal{L}(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$
- Función de pérdida cero-uno: $\mathcal{L}_\varepsilon(\hat{\theta}, \theta) = \begin{cases} 0, si|\hat{\theta} - \theta| \leq \varepsilon \\ 1, si|\hat{\theta} - \theta| > \varepsilon \end{cases}$ donde ε es un parámetro fijado.

Comparación

Estadística clásica	Estadística bayesiana
La probabilidad es la frecuencia relativa que se obtiene tras repetir muchas veces el experimento. La probabilidad es objetiva y es igual para todos los observadores.	La probabilidad es una medida del grado de incertidumbre que tiene un observador sobre el resultado de un experimento. La probabilidad inicial es subjetiva y depende del observador.
La inferencia se basa en calcular las probabilidades de los datos observados o de datos hipotéticos más extremos, dada una hipótesis.	La inferencia se basa en evaluar la probabilidad de que un modelo, o hipótesis, sea cierto dados unos datos observados.
Un intervalo de confianza es el resultado de un proceso que tiene un (1 - α) % de probabilidades de producir un intervalo que contenga al parámetro poblacional.	La probabilidad de que el parámetro poblacional se encuentre en un intervalo de credibilidad del (1 - α) % es esa probabilidad.
El nivel de significancia de un contraste de hipótesis es la probabilidad de que, dada la hipótesis nula, se obtenga un resultado igual o más extremo como el observado.	Se evalúa la probabilidad del modelo (o parámetros de éste) a partir únicamente de los datos observados.

Inferencia Bayesiana

Cuantificar la incertidumbre sobre los parámetros de manera probabílistica al considerarlos como **variables aleatorias**.

Probabilidad condicional

Sean A y B dos eventos y $P(B) > 0$, la *probabilidad condicional* de A dado B es

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

Teorema de Bayes

Sea $\{E_n\}$ una partición finita o numerable de Ω y supóngase $P(A) > 0$, entonces

$$\mathbb{P}(E_n|A) = \frac{P(A|E_n)P(E_n)}{P(A)} = \frac{P(A|E_n)P(E_n)}{\sum_n P(A|E_n)P(E_n)}.$$

Elementos

Sean Θ el espacio de todos los valores que pueden representar a $\boldsymbol{\theta}$ y $\mathcal{F} = \{f(x|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ como el conjunto que define una familia parámetrica a través de la función de densidad $f(x|\boldsymbol{\theta})$.

- Partimos de la especificación de una distribución inicial o *a priori* para $\boldsymbol{\theta}$: $\pi(\boldsymbol{\theta})$.
- Dada una muestra de observaciones $\boldsymbol{x} = (x_1, \ldots, x_n)'$, su distribución conjunta (bajo el supuesto de independencia) es

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}).$$

Esta es llamada *verosimilitud* del modelo y contiene toda la información disponible proporcionada por la muestra observada.

Actualización de la información

El objetivo es determinar la distribución de probabilidad final o *a posteriori* a través del Teorema de Bayes como

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(\boldsymbol{x})}. \tag{1}$$

Se trata de una distribución de la *variable aleatoria* $\boldsymbol{\theta}$ dada la muestra observada $\boldsymbol{x}$, actualizando la distribución inicial más la información contenida en los datos.

Usualmente, el interés se centra en el cálculo de la fórmula (1), que es la base para realizar inferencias sobre $\boldsymbol{\theta}$.

Simplificación

Al observar que $f(\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ es la densidad conjunta de la muestra $\boldsymbol{x} = (x_1, \ldots, x_n)'$ y es constante respecto a $\boldsymbol{\theta}$. Entonces, la distribución final puede expresarse de forma proporcional de la siguiente manera

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{2}$$

incluyendo solamente a la distribución inicial y la información contenida en los datos expresada por la verosimilitud, y donde $f(\boldsymbol{x})$ será la constante que haga que $f(\boldsymbol{\theta}|\boldsymbol{x})$ sea una densidad.

La especificación de la distribución inicial es importante en la inferencia bayesiana ya que su influencia se verá determinada en la distribución final.

Ejemplo

Consideremos un conjunto de n datos que representan el número de éxitos x_i en N_i (con $i = 1, \ldots, n$) realizaciones. **Interesa la cuantificación de la tasa de éxito** θ.

Esto es, $x_i \sim \text{Bin}(\theta, N_i)$. La verosimilitud puede expresarse como

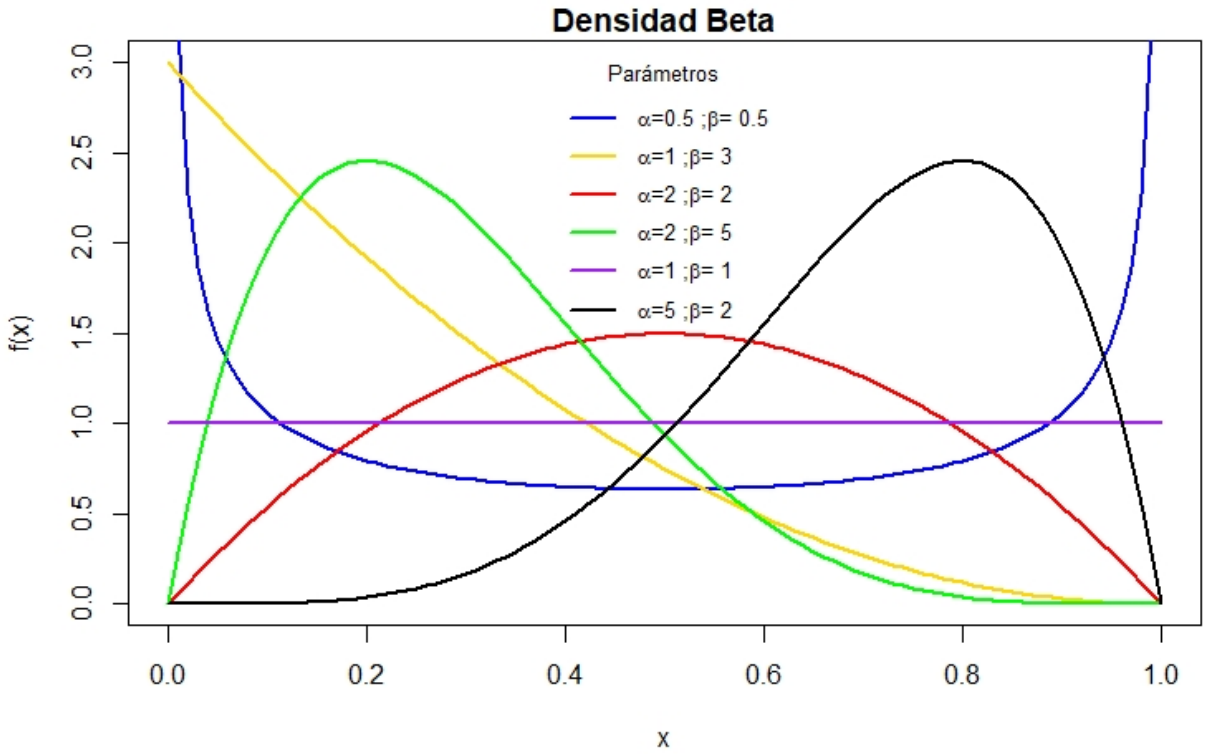
$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \left[\binom{N_i}{x_i} \theta^{x_i} (1-\theta)^{N_i - x_i} \right] \\
&= \prod_{i=1}^{n} \binom{N_i}{x_i} \cdot \theta^{\Sigma x_i} (1-\theta)^{\Sigma N_i - \Sigma x_i} \\
&= \prod_{i=1}^{n} \binom{N_i}{x_i} \cdot \theta^{n\bar{x}} (1-\theta)^{N - n\bar{x}},
\end{aligned}
$$

donde $N = \sum N_i$ y $\bar{x} = \frac{\sum x_i}{n}$.

A priori

Dado que $0 \leq \theta \leq 1$, es natural considerar una distribución cuyo soporte sea el intervalo [0,1] $\rightarrow$ Distribución *a priori* **Beta** para θ:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$



La distribución posterior se obtiene como

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta)\pi(\theta) \\
&\propto \prod_{i=1}^{n} \binom{N_i}{x_i} \theta^{n\bar{x}}(1-\theta)^{N-n\bar{x}} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^{n\bar{x}+\alpha-1}(1-\theta)^{N-n\bar{x}+\beta-1}
\end{aligned}
$$

Es el *kernel* de una distribución Beta, y la constante necesaria para que conforme completamente a una densidad es aquella c tal que $\int c \cdot f(\theta|\boldsymbol{x})d\theta = 1$.

Por lo tanto, puede decirse que $\theta|\boldsymbol{x} \sim \text{Beta}(n\bar{x}+\alpha, N-n\bar{x}+\beta)$.

Ejemplo (cont...)

Con la siguiente información:

- A priori uniforme para θ i.e. Beta de parámetros $\alpha = \beta = 1$,
- $N = 100$ y $\bar{x} = 8$ con $n = 10$.

Vamos a representar explícitamente la inferencia bayesiana sobre el parámetro θ.

- A priori $\pi(\theta)$
- Verosimilitud $f(\boldsymbol{x}|\theta)$
- Distribución posterior $p(\theta|\boldsymbol{x})$

Referencias

1. Bernardo J. and Smith A. (1994) Bayesian theory. Wiley.
2. Erdely A. and Gutiérrez-Peña E. (2007) Monografía de Estadística Bayesiana.
3. Ortiz Padilla, Í. (2018). Inferencia bayesiana. https://idus.us.es/handle/11441/77565
4. Scotto, M. G., & Tobías-Garcés, A. (2003). Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y contrastes de hipótesis. salud pública de méxico, 45(6), 506-511.
5. Puza, B. (2015). Bayesian methods for statistical analysis. ANU Press. https://library.oapen.org/bitstream/handle/20.500.12657/32424/611011.pdf?sequen
6. Robert C.P. (2007) The Bayesian Choice. Springer.

Contacto

Dr. Roberto Bárcenas Curtis

rbarcenas@ciencias.unam.mx