Teoría Bayesiana

Julián Jiménez-Cárdenas[1]

[1]Universidad Nacional de Colombia, Bogotá.

juojimenezca@unal.edu.co

σ-álgebra I

Definición (Experimento Aleatorio)

Un experimento se dice aleatorio si su resultado no se puede determinar de antemano.

Definición (Espacio de Muestra)

El conjunto Ω de todos los posibles resultados de un experimento aleatorio se llama espacio de muestra. Un elemento $\omega \in \Omega$ se llama resultado o muestra. Ω se dice discreto si es finito o contable.

σ-álgebra II

Definición (σ-álgebra)

Tome $\Omega \neq \emptyset$. Una colección $\Im$ de subconjuntos de Ω se llama $\sigma-$álgebra sobre Ω si:

1. $\Omega \in \Im$,
2. *Si $A \in \Im$, entonces $A^c \in \Im$ y,*
3. *Si $A_1, A_2, \cdots \in \Im$, entonces $\bigcup_{i=1}^{\infty} A_i \in \Im$.*

Los elementos de $\Im$ se llaman eventos.

Teorema

Si $\Omega \neq \emptyset$ y $\Im_1, \Im_2, \ldots$ son $\sigma-$álgebras sobre Ω, entonces $\bigcap_{i=1}^{\infty} \Im_i$ es una σ-álgebra sobre Ω.

σ-álgebra III

Demostración.

Como $\Omega \in \Im_j$, para $j = 1, 2, \ldots$, $\Omega \in \bigcap_{j=1}^{\infty} \Im_j$. Si $A \in \bigcap_{j=1}^{\infty} \Im_j$, $A \in \Im_j$, para $j = 1, 2, \ldots$, de modo que $A^c \in \Im_j$, y $A^c \in \bigcap_{j=1}^{\infty} \Im_j$. Por último, si

$$A_1, A_2, \cdots \in \bigcap_{j=1}^{\infty} \Im_j,$$

para todo $j = 1, 2, \ldots$, $A_1, A_2, \cdots \in \Im_j$, de modo que

$$\bigcup_{i=1}^{\infty} A_i \in \Im_j \text{ y } \bigcup_{i=1}^{\infty} A_i \in \bigcap_{j=1}^{\infty} \Im_j.$$

$\square$

σ-álgebra IV

Definición (σ-álgebra generada)

Tome $\Omega \neq \emptyset$ y $\mathcal{A}$ como una colección de subconjuntos de Ω. Si $\mathcal{M} := \{\Im : \Im$ es una $\sigma -$ álgebra sobre Ω que contiene a $\mathcal{A}\}$,

$$\sigma(\mathcal{A}) := \bigcap_{\Im \in \mathcal{M}} \Im$$

es la $\sigma-$álgebra más pequeña sobre Ω que contiene a $\mathcal{A}$. Esta $\sigma-$álgebra se conoce como σ-álgebra generada por $\mathcal{A}$.

Definición (Espacio de medida)

Tome $\Omega \neq \emptyset$ y sea $\Im$ una σ-álgebra sobre Ω. La pareja $(\Omega, \Im)$ se llama espacio de medida.

σ-álgebra V

$\emptyset$ es el evento imposible. Ω es el evento seguro y $\{\omega\}$, con $\omega \in \Omega$ es un evento simple. Decimos que el evento A ocurre después de llevar a cabo el experimento aleatorio si se obtiene un resultado en A, esto es, A ocurre si el resultado es algún $\omega \in A$.

1. El evento $A \cup B$ ocurre si y sólo si A ocurre, B pasa, o ambos ocurren.
2. El evento $A \cap B$ ocurre si y sólo si A y B ocurren a la vez.
3. El evento A^c ocurre si y sólo si A no ocurre.
4. El evento $A - B$ ocurre si y sólo si A ocurre pero B no ocurre.

Definición (Eventos mutuamente excluyentes)

Dos eventos A y B se dicen mutuamente excluyentes si $A \cap B = \emptyset$.

Espacio de probabilidad I

Definición (Frecuencia relativa)

Para cada evento A, el número $f_r(A) := \frac{n(A)}{n}$ se llama la frecuencia relativa de A, donde $n(A)$ indica el número de veces que ocurre A en n repeticiones del experimento aleatorio.

Cuando $n \to \infty$, se puede hablar de la probabilidad de que ocurra el evento A, normalizada de 0 a 1. La formalización de este concepto se encuentra en la idea del espacio de probabilidad.

Espacio de probabilidad II

Definición (Espacio de probabilidad)

Tome $(\Omega, \Im)$ *como un espacio de medida. Una función real* P *sobre* $\Im$ *que satisface las siguientes condiciones:*

1. $P(A) \geq 0$ *para todo* $A \in \Im$ *(no negativa),*
2. $P(\Omega) = 1$ *(normalizada) y,*
3. *si* $A_1, A_2, \ldots$ *son eventos mutuamente excluyentes en* $\Im$*, esto es, si*
$$A_i \cap A_j = \emptyset \text{ para todo } i \neq j, \text{ entonces}$$
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

se llama medida de probabilidad sobre $(\Omega, \Im)$*. La tripleta* $(\Omega, \Im, P)$ *se llama espacio de probabilidad.*

Notas I

Aplicando el teorema anterior de forma inductiva, para algunos eventos $A_1, A_2, \ldots, A_n \in \Im$:

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_n) = &\sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \cdots \\
&+ (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) \\
&+ \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n).
\end{aligned}
$$

Tome $(\Omega, \Im, P)$ como un espacio de probabilidad con Ω finito o contable y $\Im = \mathbb{P}(\Omega)$. Tome $\emptyset \neq A \in \Im$. Es claro que

Notas II

$$A = \bigcup_{\omega \in A} \{\omega\}, \text{ de modo que}$$

$$P(A) = \sum_{\omega \in A} P(\omega), \text{ donde } P(\omega) := P(\{\omega\}).$$

Así, P queda completamente definido por $p_j := P(\omega_j)$, donde $\omega_j \in \Omega$. El vector $|\Omega|-$dimensional $p := (p_1, p_2, \dots)$ satisface las siguientes condiciones:

- $p_j \geq 0$ y
- $\sum_{j=1}^{\infty} p_j = 1$.

Un vector que satisface las anteriores condiciones se llama **vector de probabilidad**.

Introducción

Tome B como un evento cuya opción de ocurrir debe ser medida bajo la suposición de que otro evento A fue observado. Si el experimento se repite n veces bajo las mismas circunstancias, entonces la frecuencia relativa de B bajo la condición A se define como

$$f_r(B|A) := \frac{n(A \cap B)}{n(A)} = \frac{\frac{n(A\cap B)}{n}}{\frac{n(A)}{n}} = \frac{f_r(A \cap B)}{f_r(A)}, \text{ si } n(A) > 0.$$

Esto motiva la siguiente definición

Probabilidad Condicional I

Definición (Probabilidad condicional)

Tome $(\Omega, \Im, P)$ *como un espacio de probabilidad. Si* $A, B \in \Im$, *con* $P(A) > 0$, *entonces la probabilidad del evento* B *bajo la condición* A *se define como sigue*

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

El siguiente teorema provee algunas propiedades de la probabilidad condicional.

Probabilidad Condicional II

Teorema (Medida de probabilidad condicional)

Tome $(\Omega, \Im, P)$ *como un espacio de probabilidad y* $A \in \Im$*, con* $P(A) > 0$*. Entonces:*

1. $P(\cdot|A)$ *es una medida de probabilidad sobre* Ω *centrada en* A*, esto es,* $P(A|A) = 1$.
2. *Si* $A \cap B = \emptyset$*, entonces* $P(B|A) = 0$.
3. $P(B \cap C|A) = P(B|A \cap C)P(C|A)$ *si* $P(A \cap C) > 0$.
4. *Si* $A_1, A_2, \ldots, A_n \in \Im$*, con* $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$*, entonces*

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

Probabilidad Condicional III

Demostración.

1. Las tres propiedades de una medida de probabilidad deben ser verificadas.
   1. Claramente, $P(B|A) \geq 0$ para todo $B \in \Im$.
   2. $P(\Omega|A) = \frac{P(\Omega \cap A}{P(A)} = \frac{P(A)}{P(A)} = 1$. También se tiene que $P(A|A) = 1$.
   3. Tome $A_1, A_2, \cdots \in \Im$ una sucesión de conjuntos disyuntos. Entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i | A\right) = \frac{P\left(A \cap \bigcup_{i=1}^{\infty} A_i\right)}{P(A)} = \frac{P\left(\bigcup_{i=1}^{\infty} A \cap A_i\right)}{P(A)}$$

$$= \sum_{i=1}^{\infty} \frac{P(A \cap A_i)}{P(A)} = \sum_{i=1}^{\infty} P(A_i|A).$$

$\square$

Teorema probabilidad total I

Los siguientes resultados son vitales para aplicaciones posteriores.

Teorema (Teorema de probabilidad total)

Tome $A_1, A_2, \ldots$ como una partición finita o contable de Ω, esto es, $A_i \cap A_j = \emptyset, \forall i \neq j$ y $\bigcup_{i=1}^{\infty} A_i = \Omega$, tal que $P(A_i) > 0$, para todo $A_i \in \Im$. Entonces, para todo $B \in \Im$:

$$P(B) = \sum_i P(B|A_i)P(A_i).$$

Teorema probabilidad total II

Demostración.

Observe que

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) = \bigcup_{i=1}^{\infty} B \cap A_i,$$

de modo que

$$P(B) = P\left(\bigcup_{i=1}^{\infty} B \cap A_i \right) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

$\square$

Regla de Bayes I

Como corolario del teorema anterior, se obtiene un resultado conocido como **regla de Bayes**, que constituye la base para la **teoría Bayesiana**.

Corolario (Regla de Bayes)

Tome $A_1, A_2, \ldots$ como una partición finita o contable de Ω con $P(A_i) > 0$, para todo i; entonces, para todo $B \in \Im$ con $P(B) > 0$:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(B|A_j)P(A_j)}, \forall i.$$

Regla de Bayes II

Demostración.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_j P(B|A_j)P(A_j)}.$$

Con la partición $A_1 = A, A_2 = A^c$ se obtiene la forma usual de la regla de Bayes. □

Distribuciones a priori y a posteriori

Definición (Distribuciones a priori y a posteriori)

Tome $A_1, A_2, \ldots$ como una partición finita o contable de Ω, con $P(A_i) > 0$, para todo i. Si $P(B) > 0$, con $B \in \Im$, entonces $\{P(A_n)\}_n$ se llama distribución a priori (antes de que B ocurra), y $\{P(A_n|B)\}_n$ se llama distribución a posteriori (después de que B ocurra).

Algunas veces, la ocurrencia de un evento B no afecta la probabilidad de un evento A, es decir,

$$P(A|B) = P(A).$$

En este caso, se dice que el evento A es independiente del evento B. Esto motiva la siguiente definición.

Eventos Independientes I

Definición (Eventos independientes)

Dos eventos A y B se dicen independientes si y sólo si

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

Si esta condición no se tiene, se dice que los eventos son dependientes.

Variables aleatorias I

Definición (Variable aleatoria)

Tome $(\Omega, \Im, P)$ *como un espacio de probabilidad. Una variable aleatoria es un mapa* $X : \Omega \to \mathbb{R}$ *tal que, para todo* $A \in \mathbb{B}$, $X^{-1}(A) \in \Im$, *donde* $\mathbb{B}$ *es la* σ*-álgebra de Borel sobre* $\mathbb{R}$ *(*σ*-álgebra más pequeña que contiene todos los intervalos de la forma* $(-\infty, a]$*).*
El conjunto de posibles valores de X *es* $\mathbb{S} := \{x \in \mathbb{R} : \exists\omega \in \Omega$ *tal que* $X(\omega) = x\}$, *conocido como* **soporte de la variable aleatoria** X.

Si X es una variable aleatoria definida sobre un espacio de probabilidad $(\Omega, \Im, P)$, se introduce la notación

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\}, \text{ con } B \in \mathbb{B}.$$

Variables aleatorias II

Definición (Variable aleatoria discreta)

Una variable aleatoria X se dice discreta cuando el soporte $\mathbb{S}$ de X es un subconjunto finito o contable de $\mathbb{R}$. Para $x \in \mathbb{S}$, la función $f(x) = P(X = x)$ se llama función de densidad de probabilidad (pdf para abreviar).

Definición (Variable aleatoria continua)

Una variable aleatoria X se dice continua si el soporte $\mathbb{S}$ de X es la unión de uno o más intervalos y si existe una función no negativa y real $f(x)$ tal que $P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$. La función $f(x)$ se llama función de densidad de probabilidad (pdf).

Variables aleatorias III

Algunas propiedades de la *pdf* discreta son las siguientes:

1. $f(x) \geq 0, \forall x \in \mathbb{S}$ y $f(x) = 0, \forall x \notin \mathbb{S}$.
2. $\sum_{x \in \mathbb{S}} f(x) = 1$.
3. $P(X \in B) = \sum_{x \in B} f(x)$.

Análogamente, para la *pdf* continua:

1. $f(x) \geq 0, \forall x \in \mathbb{S}$ y $f(x) = 0, \forall x \notin \mathbb{S}$.
2. $\int_{\mathbb{S}} f(x) \mathrm{d}x = 1$.
3. $P(X \in B) = \int_B f(x) \mathrm{d}x$.

Variables aleatorias IV

Definición (Función de distribución acumulativa)

La función de distribución acumulativa (CDF, para abreviar) de una variable aleatoria se define como la función $F(x) = P(X \leq x)$.

El siguiente teorema resume algunas propiedades importantes de una *CDF*.

Variables aleatorias V

Teorema

Si X es una variable aleatoria, con CDF $F(x)$, entonces:

1. $\lim_{x\to-\infty} F(x) = 0$ *y* $\lim_{x\to\infty} F(x) = 1$.
2. $F(x)$ *es no decreciente; esto es,* $F(x) \leq F(y)$*, siempre que* $x \leq y$.
3. $F(x)$ *es continua por derecha.*
4. $P(a < X \leq b) = F(b) - F(a)$.

Variables aleatorias VI

Teorema

Si X es una variable aleatoria discreta con CDF $F(x)$ y soporte $\mathbb{S} = \{x_0, x_1, \ldots\}$, con $x_0 < x_1 < \cdots$, entonces, para $x_k \in \mathbb{S}$,

$$f(x_k) = F(x_k) - F(x_{k-1}).$$

Teorema

Para una variable aleatoria continua, $f(x) = \frac{\mathrm{d}F}{\mathrm{d}x}, \forall x \in \mathbb{R}$.

Vectores Aleatorios I

Definición (Vector aleatorio)

Un vector aleatorio $\vec{X} = (X_1, X_2, \ldots, X_k)$ *es un vector* $k-$*dimensional, donde* $X_1, \ldots, X_k$ *son variables aleatorias. Un vector aleatorio se dice discreto cuando cada una de las variables aleatorias que lo conforman son discretas, y continuo cuando son continuas.*

Definición (Variable aleatoria bivariada)

Un vector aleatorio bidimensional $\vec{X} = (X_1, X_2)$ *se llama variable aleatoria bivariada.*

Vectores Aleatorios II

De modo similar al caso de las variables aleatorias, los vectores aleatorios tienen *pdf*, un soporte y una *CDF*. El soporte de un vector aleatorio $k-$dimensional es el conjunto de valores que puede tomar, denotado por $\mathbb{S}_{\vec{X}} \subseteq \mathbb{R}^k$.

Definición (Función de densidad de probabilidad adjunta discreta)

Tome $\vec{X}$ como un vector aleatorio discreto $k-$dimensional. La pdf adjunta de $\vec{X}$ se define como

$$f(\vec{x}) := f(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$$

para $\vec{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{S}_{\vec{X}}$.

Vectores Aleatorios III

La *pdf* adjunta discreta tiene las siguientes propiedades:

1. $0 \leq f(x_1, x_2, \ldots, x_k) \leq 1, \forall \vec{x} \in \mathbb{S}_{\vec{X}}$.
2. $\sum_{\vec{x} \in \mathbb{S}_{\vec{X}}} f(\vec{x}) = 1$.
3. Para cualquier subconjunto $B \subseteq \mathbb{S}_{\vec{X}},\ P(\vec{X} \in B) = \sum_{\{\vec{x} \in \mathbb{S}_{\vec{X}} : \vec{x} \in B\}} f(\vec{x})$.

Vectores Aleatorios IV

Definición (Función de densidad de probabilidad adjunta continua)

Tome $\vec{X}$ como un vector aleatorio k-dimensional continuo. La pdf continua de $\vec{X}$ se define como cualquier función no negativa $f(\vec{x})$ que satisfaga las siguientes propiedades:

1. $f(x_1, \ldots, x_k) > 0, \forall \vec{x} \in \mathbb{S}_{\vec{X}}$.
2. $\int_{\mathbb{S}_{\vec{X}}} f(x_1, \ldots, x_k) \mathrm{d}x_1 \cdots \mathrm{d}x_k = 1$.
3. *Para cualquier subconjunto* $B \subset \mathbb{S}_{\vec{X}},\ P(\vec{X} \in B) = \int_B f(x_1, \ldots, x_k) \mathrm{d}x_1 \cdots \mathrm{d}x_k$.

Vectores Aleatorios V

Definición (Función de distribución acumulativa adjunta)

Tome $\vec{X} = (X_1, X_2, \ldots, X_k)$ *como un vector aleatorio* $k-$*dimensional. La CDF adjunta de* $\vec{X}$ *se define como*

$$F(x_1, x_2, \ldots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k)$$

$$\forall (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

Vectores Aleatorios VI

Definición (Función de densidad de probabilidad marginal)

Tome $\vec{X} = (X_1, \ldots, X_k)$ como un vector aleatorio $k-$dimensional. La función de densidad de probabilidad marginal de la variable aleatoria X_i es, para los casos discreto y continuo:

$$f_i(x_i) = \underbrace{\sum_{x_1 \in \mathbb{S}_X} \cdots \sum_{x_k \in \mathbb{S}_{X_k}}}_{\text{quitando la suma sobre } x_i} f(x_1, \ldots, x_k)$$

$$f_i(x_i) = \underbrace{\int_{x_1 \in \mathbb{S}_{X_1}} \cdots \int_{x_k \in \mathbb{S}_{X_k}}}_{\text{quitando la integral sobre } x_i} f(x_1, \ldots, x_k) \prod_{n \neq i} \mathrm{d}x_n$$

Vectores Aleatorios VII

Definición (Función de densidad de probabilidad condicional)

Tome $\vec{X}(X_1, \ldots, X_k)$ *como un vector aleatorio* $k-$*dimensional. Para un valor fijo de* x_i, *donde* $f_i(x_i) > 0$, *la función de densidad de probabilidad condicional para* $\vec{Y}|X_i$; *donde* $\vec{Y}$ *es un vector aleatorio* $(k-1)-$*dimensional con todas las variables aleatorias de* $\vec{X}$, *a excepción de* X_i; *es*

$$f(\vec{y}|x_i) = \frac{f(x_1, \ldots, x_k)}{f_i(x_i)},$$

donde $\vec{y} \in \mathbb{S}_{\vec{Y}}$.

Vectores Aleatorios VIII

Definición (Colección independiente de variables aleatorias)

Una colección de variables aleatorias $\{X_1, X_2, \ldots, X_k\}$ *se dice independiente cuando*

$$F(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} F_i(x_i), \forall \vec{x} \in \mathbb{R}^k,$$

donde $F_i(x_i)$ *es la CDF marginal de la variable aleatoria* X_i *(determinada a partir de la pdf marginal:* $F_i(x_i) := P(X_i \leq x_i)$*).*

Vectores Aleatorios IX

Definición (Colección independiente de variables aleatorias)

También se puede definir la independencia entre variables aleatorias usando las pdf, en el sentido de que la misma colección de variables aleatorias se dice independiente cuando

$$f(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} f_i(x_i), \forall \vec{x} \in \mathbb{S}_{\vec{X}},$$

donde $f_i(x_i)$ es la pdf marginal asociada a la variable aleatoria X_i.

Vectores Aleatorios X

Definición (Colección de variables aleatorias independientes idénticamente distribuidas)

Una colección de variables aleatorias $\{X_1, X_2, \ldots, X_k\}$ *se dice independiente e idénticamente distribuidas (iid, para abreviar) si y sólo si* $X_1, X_2, \ldots, X_k$ *son variables aleatorias independientes y la pdf de cada variable aleatoria es idéntica.*

Estimación paramétrica

1. Un modelo probabilístico $f(x, \theta)$, especificado con los valores de los parámetros desconocidos.
2. Un conjunto de posibles valores de θ bajo consideración, llamado el espacio de parámetros, denotado por Θ.
3. Una muestra aleatoria de n observaciones del modelo probabilístico.
4. Un conjunto de estimadores puntuales para los valores de los parámetros desconocidos, basados en la información contenida en la muestra aleatoria.
5. Las propiedades específicas de los estimadores que permiten evaluar la precisión y eficiencia del estimador.

Muestra y Estadística I

Definición (Muestra)

Una colección de variables aleatorias $X_1, \ldots, X_k$ *se llama muestra de tamaño n. Una muestra de n variables aleatorias independientes* $X_1, \ldots, X_n$ *se llama muestra aleatoria.*

Definición (Estadística y estimador)

Dada una muestra $X_1, X_2, \ldots, X_n$*, una estadística* $T = T(X_1, \ldots, X_n)$ *es una función de la muestra que no depende de ningún otro parámetro desconocido. Un estimador es una estadística que se usa para determinar una cantidad desconocida, y el estimado es el valor observado del estimador (evaluando la función en la muestra).*

Muestra y Estadística II

Definición (Distribución muestral)

Para una muestra $X_1, \ldots, X_n$ y una estadística $T = T(X_1, \ldots, X_n)$, la distribución muestral de la estadística T es la distribución de probabilidad asociada a la variable aleatoria T. La pdf de la distribución muestral se denota como $f_T(t; \theta)$.

Muestra y Estadística III

Definición (Valor esperado)

Tome X como una variable aleatoria con pdf $f(x)$ en $\mathbb{S}_X$. El valor esperado de la variable aleatoria X, denotado por $E(X)$, se define como

$$E(X) = \sum_{x \in \mathbb{S}_X} x f(x)$$

cuando X es una variable aleatoria discreta, y como

$$E(X) = \int_{x \in \mathbb{S}_X} x f(x) \mathrm{d}x$$

cuando X es una variable aleatoria continua.

Muestra y Estadística IV

Definición (Estimador imparcial)

Una estadística T *se dice estimador imparcial de un parámetro* θ *cuando* $E(T) = \theta$, $\forall\theta \in \Theta$. *Una estadística se conoce como estimador parcial de* θ *cuando* $E(T) \neq \theta$, *y la parcialidad de una estadística* T *para estimar un parámetro* θ *se define como* $Bias(T;\theta) = E(T) - \theta$.

Definición (Estimador asintóticamente imparcial)

Una estadística $T_n = T(X_1, \ldots, X_n)$ *se conoce como estimador asintóticamente imparcial de un parámetro* θ *cuando*

$$\lim_{n\to\infty} Bias(T_n;\theta) = 0.$$

Muestra y Estadística V

Algunas variables útiles para determinar la precisión y exactitud del estimador son las siguientes:

$$SE(T) := \sqrt{E((T - E(T))^2)} := \sqrt{Var(T)}.$$

$$MSE(T;\theta) = E((T-\theta)^2).$$

Una estadística que contiene toda la información relevante acerca de θ en una muestra se conoce como estadística suficiente.

Muestra y Estadística VI

Definición (Estadística suficiente)

Tome $X_1, \ldots, X_n$ como una muestra de variables aleatorias iid con pdf común $f(x;\theta)$, para $\theta \in \Theta \subseteq \mathbb{R}^d$. Un vector de estadísticas $\vec{S}(\vec{X}) := (S_1(\vec{X}), \ldots, S_k(\vec{X}))$ se dice que es una estadística suficiente $k-$dimensional para un parámetro θ si y sólo si la distribución condicional de $\vec{X}$ dado $S = s$ no depende de θ, para ningún valor de s.

Muestra y Estadística VII

Definición (Función de verosimilitud)

Para una muestra $X_1, \ldots, X_n$, la función de verosimilitud $L(\theta|\vec{X})$ es la pdf adjunta de $\vec{X} = (X_1, \ldots, X_n)$, es decir,

$$L(\theta|\vec{X}) = f(x_1, \ldots, x_n; \theta)$$

La función logarítmica de verosimilitud $\ell(\theta)$ se define como el logaritmo de la función de verosimilitud.

Cuando $X_1, \ldots, X_n$ es una muestra de variables aleatorias *iid*, se puede escribir la función de verosimilitud como

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Muestra y Estadística VIII

Teorema (Teorema de factorización de Neyman-Fisher)

Tome $X_1, \ldots, X_n$ como una muestra de variables aleatorias iid con pdf $f(x;\theta)$, y espacio de parámetros Θ. Una estadística $S(\vec{X})$ es suficiente para θ si y sólo si $L(\theta)$ se puede factorizar como

$$L(\theta) = g(S(\vec{x});\theta)h(\vec{x}),$$

donde $g(S(\vec{x});\theta)$ no depende de $\vec{x} = (x_1, \ldots, x_n)$, excepto a través de $S(\vec{x})$, y $h(\vec{x})$ no depende de θ.

Ley de verosimilitud

Tome $X_1, \ldots, X_n$ como una muestra de variables aleatorias *iid* con *pdf* común $f(x; \vec{\theta})$ y espacio de parámetros Θ. Para $\vec{\theta} \in \Theta$, mientras mayor sea el valor de $L(\vec{\theta})$, el modelo probabilístico con parámetro $\vec{\theta}$ se ajusta más a los datos observados. Entonces, el grado con el cual la información de la muestra da soporte a un parámetro $\vec{\theta}_0 \in \Theta$, en comparación con otro parámetro $\vec{\theta}_1 \in \Theta$ es igual a la razón entre sus verosimilitudes

$$\Lambda(\vec{\theta}_0, \vec{\theta}_1) = \frac{L(\vec{\theta}_0)}{L(\vec{\theta}_1)}.$$

Definición (Función de Score)

Tome $X_1, \ldots, X_n$ como una muestra de variables aleatorias con función de verosimilitud $L(\vec{\theta})$, para $\vec{\theta} \in \Theta$. Si la función de verosimilitud logarítmica $\ell(\vec{\theta})$ es diferenciable, la función de Score se define como

$$Sc(\vec{\theta}) = \nabla_{\vec{\theta}}\, \ell(\theta),$$

de tal modo que una condición necesaria para que $\vec{\theta} \in \Theta$ sea un máximo es que $Sc(\theta) = \vec{0}$.

Estimación bayesiana I

En la estimación paramétrica puntual bayesiana, el parámetro θ se trata como una variable aleatoria, con su propia *pdf* $\pi(\theta;\lambda)$.
Las inferencias de θ en la aproximación bayesiana están basadas en la distribución de θ dados los valores observados de una muestra aleatoria $\vec{x} = (x_1, \ldots, x_n)$, llamada distribución posterior, y denotada por $f(\theta|\vec{x})$.
Usando el teorema de Bayes y el teorema de la probabilidad total, en el caso de que θ es una variable aleatoria continua,

$$f(\theta|\vec{x}) = \frac{f(\vec{x},\theta;\lambda)}{f_{\vec{X}}(\vec{x})} = \frac{f(\vec{x}|\theta)\pi(\theta;\lambda)}{\int_{\mathbb{S}_\theta} f(\vec{x}|\theta)\pi(\theta;\lambda)\mathrm{d}\theta}.$$

De modo similar, cuando θ es una variable aleatoria discreta,

Estimación bayesiana II

$$f(\theta|\vec{x}) = \frac{f(\vec{x}|\theta)\pi(\theta;\lambda)}{\sum_{\theta\in\mathbb{S}_\theta} f(\vec{x}|\theta)\pi(\theta;\lambda)}.$$

La distribución posterior combina la información disponible de θ en la distribución previa y la función de verosimilitud para producir una distribución actualizada que contiene toda la información disponible de θ.

El siguiente teorema indica que la distribución posterior depende de la muestra $\vec{x}$ sólo bajo una estadística suficiente para θ.

Estimación bayesiana III

Teorema

Si $X_1, \ldots, X_n$ es una muestra de variables independientes iid con pdf común $f(x|\theta)$, S es una estadística suficiente para θ, y $\pi(\theta; \lambda)$ una distribución previa para θ, entonces la distribución posterior de θ dado $\vec{X}$ depende de la muestra sólo a través de una estadística suficiente S.

Referencias

L. Blanco, V. Arunachalam, and S. Dharmaraja.
Introduction to Probability and Stochastic Processes with Applications.
Wiley, 2012.

R. Rossi.
Mathematical statistics: an introduction to likelihood based inference.
John Wiley Sons, Inc., 2018.