EST-46111: Fundamentos de Estadística

–Notas de Clase–

Juan Carlos Martínez-Ovando

ITAM

25 de agosto de 2016

Agenda

Modelo estadístico

Definición

Notación

- X es una variable aleatoria observable (discreta, continua o del tipo mixta).

Modelo estadístico

Sin pérdida de generalidad, refirámonos a X. El modelo estadístico se define como la distribución de probabilidades de X indizada por θ, i.e.

$$X \sim F(x|\theta). \quad (1)$$

El soporte de X, denotado por $\mathcal{X}$ se define como,

$$\mathcal{X} = \{x : F(x|\theta) > 0\}, \quad (2)$$

donde $\mathcal{X}$ toma valores en un espacio Euclidiano de dimensión finita. El parámetro θ, toma valores en el espacio parametral Θ (generalmente de dimensión finita tambén).

Definición

Notación

- X es una variable aleatoria observable (discreta, continua o del tipo mixta).

Modelo estadístico

Sin pérdida de generalidad, refirámonos a X. El modelo estadístico se define como la distribución de probabilidades de X indizada por θ, i.e.

$$X \sim F(x|\theta). \tag{1}$$

El soporte de X, denotado por $\mathcal{X}$ se define como,

$$\mathcal{X} = \{x : F(x|\theta) > 0\}, \tag{2}$$

donde $\mathcal{X}$ toma valores en un espacio Euclidiano de dimensión finita. El parámetro θ, toma valores en el espacio parametral Θ (generalmente de dimensión finita tambén).

Definición

Densidades y masa de probabilidad

1. Cuando X es absolutamente continua, $F(x|\theta)$ admite una densidad, $f(x|\theta)$ y $\mathbb{P}(X=x)=0$ para todo $x \in \mathcal{X}$.
2. Cuando X es discreta, $\mathbb{P}(X=x)>0$ para todo $x \in \mathcal{X}$. Los valores de x que satisfacen lo anterior se llaman átomos.
3. Cuando X es del tipo mixta, admite una parte absolutamente continia al mismo tiempo de admitir átomos, i.e.

$$\mathbb{P}(X \leq x) = F(x|\theta) = F_c(x|\theta_c) + \sum_{x_k \leq x} p(X = x_k|\theta_d), \tag{3}$$

donde

- F_c es el componente continuo de la distribución
- $\{x_k\}_{k\geq 1}$ son los átomos de la distribución
- θ_c y θ_d son los parámetros asociados con la parte continua y discreta, respectivamente.

Verosimilitud

Consideremos un conjunto de datos $X_1 = x_1, \ldots, X_n = x_n$ en el caso continuo.

Función de verosimilitud

- Enfoque frecuentista: Independencia

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta). \tag{4}$$

- Enfoque bayesiano: Independencia condicional

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{n} f(x_i | \theta) \pi(\theta) \mathrm{d}\theta. \tag{5}$$

Ejercicio

Cómo será la expresión de la función de verosimilitud en el caso discreto y tipo mixta?

Verosimilitud

Consideremos un conjunto de datos $X_1 = x_1, \ldots, X_n = x_n$ en el caso continuo.

Función de verosimilitud

- Enfoque frecuentista: Independencia

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta). \tag{4}$$

- Enfoque bayesiano: Independencia condicional

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{n} f(x_i|\theta)\pi(\theta)\mathrm{d}\theta. \tag{5}$$

Ejercicio

Cómo será la expresión de la función de verosimilitud en el caso discreto y tipo mixta?

Verosimilitud

Tipos de datos

- Las expresiones anteriores son correctas cuando los datos son exactos.
- Cuando trabajamos con datos agrupados en $\Re_+$, modificamos el soporte $\mathcal{X}$ por una partici ón $\{c_j\}_{j=1}^{J}$ tal que,

$$c_1 < c_2 < \ldots < c_J, \tag{6}$$

sustituyendo $\mathcal{X}$ por el conjunto,

$$\mathcal{C} = \{(c_j, c_{j+1}] : c_j < c_{j+1}, j = 1, \ldots, J\}. \tag{7}$$

Ejercicio

Cómo será la expresión de la función de verosimilitud para datos agrupados?

Verosimilitud

Tipos de datos

- Las expresiones anteriores son correctas cuando los datos son exactos.
- Cuando trabajamos con datos agrupados en $\Re_+$, modificamos el soporte $\mathcal{X}$ por una partici ón $\{c_j\}_{j=1}^J$ tal que,

$$c_1 < c_2 < \ldots < c_J, \tag{6}$$

sustituyendo $\mathcal{X}$ por el conjunto,

$$\mathcal{C} = \{(c_j, c_{j+1}] : c_j < c_{j+1}, j = 1, \ldots, J\}. \tag{7}$$

Ejercicio

Cómo será la expresión de la función de verosimilitud para datos agrupados?

Distribuciones conjugadas

En el análisis bayesiano de datos, el uso de familias conjugadas entre $f(x|\theta)$ y $\pi(\theta)$ es de utilidad para obtener expresiones analíticas cerradas en el proceso de aprendizaje.

Familia Exponencial

Las familias conjugadas están definidas dentro de la Familia Exponencial de Distribuciones (lineal), para las que la función de densidad o masa de probabilidad admiten la siguiente expresión,

$$f(x|\theta) = p(x)q(\theta)^{-1}exp\{-\theta x\}, \tag{8}$$

considerando que el soporte $\mathcal{X}$ no depende de θ.

Prior conjugada

Las distribución inicial conjugada para la representación atenrior toma la forma,

$$\pi(\theta) = c(k_0, m_0)q(\theta)^{-k_0}exp\{-\theta m_0\}, \tag{9}$$

donde k_0 y m_0 son hiper-parámetros.

Distribuciones conjugadas

En el análisis bayesiano de datos, el uso de familias conjugadas entre $f(x|\theta)$ y $\pi(\theta)$ es de utilidad para obtener expresiones analíticas cerradas en el proceso de aprendizaje.

Familia Exponencial

Las familias conjugadas están definidas dentro de la Familia Exponencial de Distribuciones (lineal), para las que la función de densidad o masa de probabilidad admiten la siguiente expresión,

$$f(x|\theta) = p(x)q(\theta)^{-1}exp\{-\theta x\}, \tag{8}$$

considerando que el soporte $\mathcal{X}$ no depende de θ.

Prior conjugada

Las distribución inicial conjugada para la representación atenrior toma la forma,

$$\pi(\theta) = c(k_0, m_0)q(\theta)^{-k_0}exp\{-\theta m_0\}, \tag{9}$$

donde k_0 y m_0 son hiper-parámetros.

Predicción

Enfoque frecuentista

Bajo el enfoque frecuentista, la predicción de un valor futuro de X, X^f, se obtiene a través de la imputación del EMV de θ en el modelo, i.e.

$$X^f | x_1 \ldots, x_n \sim f(x^f | \widehat{\theta}_n), \tag{10}$$

donde $\widehat{\theta}_n = \widehat{\theta}_n(x_1 \ldots, x_n)$.

Enfoque bayesiano

Bajo el enfoque bayesiano, la predicción se obtiene usando argumentos probabilistas, como

$$p(x^f | x_1 \ldots, x_n) = \int_{\Theta} f(x^f | \theta) \pi(\theta | x_1 \ldots, x_n) \mathrm{d}\theta, \tag{11}$$

donde $\pi(\theta | x_1 \ldots, x_n) \propto f(x_1 \ldots, x_n | \theta) \pi(\theta)$ es la distribución de θ actualizada con la información contenida en $x_1 \ldots, x_n$.

Ejercicio

Muestra que el modelo Bernoulli-beta, visto en las clases previas, es un tipo de distribuciones conjugadas.

Predicción

Enfoque frecuentista

Bajo el enfoque frecuentista, la predicción de un valor futuro de X, X^f, se obtiene a través de la imputación del EMV de θ en el modelo, i.e.

$$X^f|x_1 \ldots, x_n \sim f(x^f|\widehat{\theta}_n), \tag{10}$$

donde $\widehat{\theta}_n = \widehat{\theta}_n(x_1 \ldots, x_n)$.

Enfoque bayesiano

Bajo el enfoque bayesiano, la predicción se obtiene usando argumentos probabilistas, como

$$p(x^f|x_1 \ldots, x_n) = \int_\Theta f(x^f|\theta)\pi(\theta|x_1 \ldots, x_n)\mathrm{d}\theta, \tag{11}$$

donde $\pi(\theta|x_1 \ldots, x_n) \propto f(x_1 \ldots, x_n|\theta)\pi(\theta)$ es la distribución de θ actualizada con la información contenida en $x_1 \ldots, x_n$.

Ejercicio

Muestra que el modelo Bernoulli-beta, visto en las clases previas, es un tipo de distribuciones conjugadas.

Predicción

Enfoque frecuentista

Bajo el enfoque frecuentista, la predicción de un valor futuro de X, X^f, se obtiene a través de la imputación del EMV de θ en el modelo, i.e.

$$X^f | x_1 \ldots, x_n \sim f(x^f | \widehat{\theta}_n), \tag{10}$$

donde $\widehat{\theta}_n = \widehat{\theta}_n(x_1 \ldots, x_n)$.

Enfoque bayesiano

Bajo el enfoque bayesiano, la predicción se obtiene usando argumentos probabilistas, como

$$p(x^f | x_1 \ldots, x_n) = \int_\Theta f(x^f | \theta) \pi(\theta | x_1 \ldots, x_n) \mathrm{d}\theta, \tag{11}$$

donde $\pi(\theta | x_1 \ldots, x_n) \propto f(x_1 \ldots, x_n | \theta) \pi(\theta)$ es la distribución de θ actualizada con la información contenida en $x_1 \ldots, x_n$.

Ejercicio

Muestra que el modelo Bernoulli-beta, visto en las clases previas, es un tipo de distribuciones conjugadas.

Intercambiabilidad

Definición

Se dice que unconjunto (numerable) de variables aleatorias, $\{X_j\}_{j=1}^{\infty}$, es (numerable) intercabiable con respecto a $\mathbb{P}$ si para todo n finito,

$$\mathbb{P}(X_1, \ldots, X_n) = \mathbb{P}(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{12}$$

donde $(\sigma(1), \ldots, \sigma(n))$ es cualquier permutación del vector $(1, \ldots, n)$.

Comentarios

- Cualquier sucesión de variables aleatorias iid es naturalmente intercambiable.
- La noción de intercambiabilidad, como la de independencia, se refiere a que el orden de la información es irrelevante (i.e. los resultados analíticos son invariantes ante permitaciones).

Ejercicio

· Describe un ejemplo donde los datos podrían asociarse con el supuesto de intercambiabilidad, mas no de independencia. · Describe también un ejemplo donde ni intercambiabilidad ni independencia serían supuestos viables.

Intercambiabilidad

Definición

Se dice que unconjunto (numerable) de variables aleatorias, $\{X_j\}_{j=1}^{\infty}$, es (numerable) intercabiable con respecto a $\mathbb{P}$ si para todo n finito,

$$\mathbb{P}(X_1, \ldots, X_n) = \mathbb{P}(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{12}$$

donde $(\sigma(1), \ldots, \sigma(n))$ es cualquier permutación del vector $(1, \ldots, n)$.

Comentarios

- Cualquier sucesión de variables aleatorias iid es naturalmente intercambiable.
- La noción de intercambiabilidad, como la de independencia, se refiere a que el orden de la información es irrelevante (i.e. los resultados analíticos son invariantes ante permitaciones).

Ejercicio

· Describe un ejemplo donde los datos podrían asociarse con el supuesto de intercambiabilidad, mas no de independencia. · Describe también un ejemplo donde ni intercambiabilidad ni independencia serían supuestos viables.

Intercambiabilidad

Definición

Se dice que unconjunto (numerable) de variables aleatorias, $\{X_j\}_{j=1}^{\infty}$, es (numerable) intercabiable con respecto a $\mathbb{P}$ si para todo n finito,

$$\mathbb{P}(X_1, \ldots, X_n) = \mathbb{P}(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{12}$$

donde $(\sigma(1), \ldots, \sigma(n))$ es cualquier permutación del vector $(1, \ldots, n)$.

Comentarios

- Cualquier sucesión de variables aleatorias iid es naturalmente intercambiable.
- La noción de intercambiabilidad, como la de independencia, se refiere a que el orden de la información es irrelevante (i.e. los resultados analíticos son invariantes ante permitaciones).

Ejercicio

· Describe un ejemplo donde los datos podrían asociarse con el supuesto de intercambiabilidad, mas no de independencia. · Describe también un ejemplo donde ni intercambiabilidad ni independencia serían supuestos viables.

Intercambiabilidad

Teorema de de Finetti

Una consecuencia del supuesto de intercambiabilidad (numerable) es el teorema de representación en el que se admite que para toda sucesión de variables aleatorias intercambiables, para toda n finita, se tiene que .existeün ente estocástico $\theta \in \Theta$ acompañado de una medida de probabilidad Π, tal que

$$\mathbb{P}(X_1, \ldots, X_n) = \int_{\Theta} \prod_{j=1}^{n} \mathbb{P}(X_j|\theta)\Pi(\mathrm{d}\theta). \tag{13}$$

Comentarios

- El resultado anterior es de .existencia", i.e. no nombra como se lleva a cabo tal representación.
- Para un conjunto de variables aleatorias intercambiables existen más de una posible representación como la anterior (en términos de diferentes especificaciones de θ y/o de Π).
- Este teorema de representación brinda una interpretación al paradigma bayesiano de inferencia.

Ejercicio

Intercambiabilidad

Teorema de de Finetti

Una consecuencia del supuesto de intercambiabilidad (numerable) es el teorema de representación en el que se admite que para toda sucesión de variables aleatorias intercambiables, para toda n finita, se tiene que .existeün ente estocástico $\theta \in \Theta$ acompañado de una medida de probabilidad Π, tal que

$$\mathbb{P}(X_1, \ldots, X_n) = \int_{\Theta} \prod_{j=1}^{n} \mathbb{P}(X_j | \theta) \Pi(\mathrm{d}\theta). \tag{13}$$

Comentarios

- El resultado anterior es de .existencia", i.e. no nombra como se lleva a cabo tal representación.
- Para un conjunto de variables aleatorias intercambiables existen más de una posible representación como la anterior (en términos de diferentes especificaciones de θ y/o de Π).
- Este teorema de representación brinda una interpretación al paradigma bayesiano de inferencia.

Ejercicio

Intercambiabilidad

Teorema de de Finetti

Una consecuencia del supuesto de intercambiabilidad (numerable) es el teorema de representación en el que se admite que para toda sucesión de variables aleatorias intercambiables, para toda n finita, se tiene que .existeün ente estocástico $\theta \in \Theta$ acompañado de una medida de probabilidad Π, tal que

$$\mathbb{P}(X_1, \ldots, X_n) = \int_\Theta \prod_{j=1}^{n} \mathbb{P}(X_j | \theta) \Pi(\mathrm{d}\theta). \tag{13}$$

Comentarios

- El resultado anterior es de .existencia", i.e. no nombra como se lleva a cabo tal representación.
- Para un conjunto de variables aleatorias intercambiables existen más de una posible representación como la anterior (en términos de diferentes especificaciones de θ y/o de Π).
- Este teorema de representación brinda una interpretación al paradigma bayesiano de inferencia.

Ejercicio

Gracias por su atención...

juan.martinez.ovando@itam.mx