ACT-11302: Cálculo Actuarial III

Sesion 02 - Modelos de Probabilidad

Juan Carlos Martínez-Ovando

ITAM

Primavera 2019

Agenda

Verosimilitud

Consideremos un conjunto de datos $X_1 = x_1, \ldots, X_n = x_n$ (no átomos) en el caso absolutamente continuo.

Función de verosimilitud

- Enfoque frecuentista: Independencia

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

- Enfoque bayesiano: Independencia condicional

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{n} f(x_i|\theta)\pi(\theta)\mathrm{d}\theta. \tag{2}$$

Ejercicio

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

Verosimilitud

Consideremos un conjunto de datos $X_1 = x_1, \ldots, X_n = x_n$ (no átomos) en el caso absolutamente continuo.

Función de verosimilitud

- Enfoque frecuentista: Independencia

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

- Enfoque bayesiano: Independencia condicional

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{n} f(x_i | \theta) \pi(\theta) \mathrm{d}\theta. \tag{2}$$

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 partición $\{c_j\}_{j=1}^{J}$ tal que,

$$c_1 < c_2 < \ldots < c_J, \tag{3}$$

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

$$\mathcal{C} = \{(c_j, c_{j+1}] : c_j < c_{j+1}, j = 1, \ldots, J\}. \tag{4}$$

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 partición $\{c_j\}_{j=1}^J$ tal que,

$$c_1 < c_2 < \ldots < c_J, \tag{3}$$

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

$$\mathcal{C} = \{(c_j, c_{j+1}] : c_j < c_{j+1}, j = 1, \ldots, J\}. \tag{4}$$

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 anaĺíticas 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,

$$f(x|\theta) = p(x)q(\theta)^{-1}exp\{-\theta x\}, \tag{5}$$

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

Prior conjugada

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

$$\pi(\theta) = c(k_0, m_0)q(\theta)^{-k_0}exp\{-\theta m_0\}, \tag{6}$$

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,

$$f(x|\theta) = p(x)q(\theta)^{-1}exp\{-\theta x\}, \tag{5}$$

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

Prior conjugada

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

$$\pi(\theta) = c(k_0, m_0)q(\theta)^{-k_0}exp\{-\theta m_0\}, \tag{6}$$

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.

$$X^f|x_1\ldots,x_n \sim f(x^f|\widehat{\theta}_n), \tag{7}$$

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

$$p(x^f|x_1\ldots,x_n) = \int_\Theta f(x^f|\theta)\pi(\theta|x_1\ldots,x_n)\mathrm{d}\theta, \tag{8}$$

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.

$$X^f|x_1 \ldots, x_n \sim f(x^f|\widehat{\theta}_n), \tag{7}$$

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

$$p(x^f|x_1 \ldots, x_n) = \int_\Theta f(x^f|\theta)\pi(\theta|x_1 \ldots, x_n)\mathrm{d}\theta, \tag{8}$$

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.

$$X^f | x_1 \dots, x_n \sim f(x^f | \widehat{\theta}_n), \tag{7}$$

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

Enfoque bayesiano

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

$$p(x^f | x_1 \dots, x_n) = \int_{\Theta} f(x^f | \theta) \pi(\theta | x_1 \dots, x_n) \mathrm{d}\theta, \tag{8}$$

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

Ejercicio

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

Intercambiabilidad

Definición

Se dice que un conjunto (numerable) de variables aleatorias $\{X_j\}_{j=1}^{\infty}$ es intercambiabiable con respecto a $\Pr$ si para todo n finito,

$$\Pr(X_1, \ldots, X_n) = \Pr(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{9}$$

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 permutaciones).

Ejercicio

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

Intercambiabilidad

Definición

Se dice que un conjunto (numerable) de variables aleatorias $\{X_j\}_{j=1}^{\infty}$ es intercambiabiable con respecto a $\Pr$ si para todo n finito,

$$\Pr(X_1, \ldots, X_n) = \Pr(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{9}$$

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 permutaciones).

Ejercicio

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

Intercambiabilidad

Definición

Se dice que un conjunto (numerable) de variables aleatorias $\{X_j\}_{j=1}^{\infty}$ es intercambiabiable con respecto a $\Pr$ si para todo n finito,

$$\Pr(X_1, \ldots, X_n) = \Pr(X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{9}$$

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 permutaciones).

Ejercicio

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

Intercambiabilidad

Representación: 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 existe un ente estocástico $\theta \in \Theta$ acompañado de una medida de probabilidad Π, tal que

$$\Pr(X_1, \ldots, X_n) = \int_{\Theta} \left\{ \prod_{j=1}^{n} \Pr(X_j | \theta) \right\} \Pi(\mathrm{d}\theta), \tag{10}$$

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

Comentarios

- El resultado anterior es de *existencia*, i.e. no nombra cómo 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.

Intercambiabilidad

Representación: 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 existe un ente estocástico $\theta \in \Theta$ acompañado de una medida de probabilidad Π, tal que

$$\Pr(X_1, \ldots, X_n) = \int_{\Theta} \left\{ \prod_{j=1}^{n} \Pr(X_j|\theta) \right\} \Pi(\mathrm{d}\theta), \tag{10}$$

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

Comentarios

- El resultado anterior es de *existencia*, i.e. no nombra cómo 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 para casa

Consideremos el casosencillo donde X es discreta con $\mathcal{X} = \{0, 1\}$.

1. Defina el modelo de probabilidad para X.
2. Identifique el parámetro del modelo.
3. Defina la varosimilitud para el parámetro basado en el supuesto de $independencia$ con base en tres datos $x_1 = 1$, $x_2 = 1$ y $x_3 = 1$.
4. Calcule la distribución predictiva para X_4.
5. Examine la forma genérica de la función de verosimilitud para el parámetro del modelo.
6. Identifique la distribucion Π conjugada.
7. Calcule la distribución predictiva para X_4 usando los mismos datos empleados anteriormente, $x_1, x - 2, x_3$.

Gracias por su atención...

juan.martinez.ovando@itam.mx