ACT-11302: Cálculo Actuarial III

Sesión 01 - Modelos de Probabilidad

Juan Carlos Martínez-Ovando

ITAM

Otoño 2019

Agenda

Modelo de probabilidad

Definición

Notación

- X se refiere a una variable aleatoria observable (discreta o continua)
- x se refiere a un valor específico de esta variable

Modelo de probabilidad

Sin pérdida de generalidad, refirámonos a X. El modelo de probabilidad 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 $\mathcal{X}$ forma un subconjunto de un espacion Euclidiano de dimensión finita. El parámetro θ, toma valores en el espacio parametral Θ (generalmente de dimensión finita).

Definición

Notación

- X se refiere a una variable aleatoria observable (discreta o continua)
- x se refiere a un valor específico de esta variable

Modelo de probabilidad

Sin pérdida de generalidad, refirámonos a X. El modelo de probabilidad 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 $\mathcal{X}$ forma un subconjunto de un espacion Euclidiano de dimensión finita. El parámetro θ, toma valores en el espacio parametral Θ (generalmente de dimensión finita).

Definición

Densidades y masa de probabilidad

1. Cuando X es absolutamente continua, $F(x|\theta)$ admite una densidad, $f(x|\theta)$, tal que
$$F(x|\theta) = \int_{-\infty}^{x} f(s|\theta) ds,$$
implicando que el soporte no tenga átomos, i.e.
$$\Pr(X = x) = 0$$
para todo $x \in \mathcal{X}$.
2. Cuando X es del tipo discreto, el soporte $\mathcal{X}$ está formado solamente por átomos, i.e. valores especificos de X, digamos $\mathcal{X} = \{x_1^*, \ldots, x_n^*\}$ tales que
$$\Pr(X = x_i^*) = p_i > 0,$$
para todo $x_i^* \in \mathcal{X}$, con
$$\sum_{i=1}^{n} p_i = 1.$$

Definición

Densidades y masa de probabilidad

3 Cuando X es del tipo mixta, el modelo de probabilidad admite una parte absolutamente continia al mismo tiempo de admitir una parte discreta, i.e.

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

donde

- $F_c(\cdot)$ 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.

Densidades y masa de probabilidad

En este tipo de distribuciones, el soporte $\mathcal{X}$ está formado de una parte absolutamente continua (sin átomos), $\mathcal{X}_c$, y una parte discreta (formada solo de átomos), $\mathcal{X}_d$, i.e.

$$\mathcal{X} = \mathcal{X}_c \cup \mathcal{X}_d.$$

Definición

Densidades y masa de probabilidad

3 Cuando X es del tipo mixta, el modelo de probabilidad admite una parte absolutamente continia al mismo tiempo de admitir una parte discreta, i.e.

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

donde

- $F_c(\cdot)$ 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.

Densidades y masa de probabilidad

En este tipo de distribuciones, el soporte $\mathcal{X}$ está formado de una parte absolutamente continua (sin átomos), $\mathcal{X}_c$, y una parte discreta (formada solo de átomos), $\mathcal{X}_d$, i.e.

$$\mathcal{X} = \mathcal{X}_c \cup \mathcal{X}_d.$$

Ejemplo

Ejemplo

Pensemos en el modelo de probabilidad con un átomo en $\{0\}$ que admite la posibilidad de tomar valores en la recta real positiva, i.e.

$$\mathcal{X} = \{0\} \cup (0, \infty).$$

Modelo de probabilidad

El modelo de probabilidad estará definido por una masa de probabilididad en $\{0\}$, i.e.

$$\Pr(X = 0) = \Pr(X \in \{0\}) = \theta_0,$$

y una densidad para la parte continua,

$$f(x|\theta_c) = \theta_c \,\mathsf{exp}\{-x\theta_c\}\mathbb{I}_{(0,\infty)}(x),$$

con $0 < \theta_0 < 1$ y $\theta_c > 0$.

Ejercicio

¿Qué forma toma $F(x|\theta)$ y quién es θ?

Ejemplo

Ejemplo

Pensemos en el modelo de probabilidad con un átomo en $\{0\}$ que admite la posibilidad de tomar valores en la recta real positiva, i.e.

$$\mathcal{X} = \{0\} \cup (0, \infty).$$

Modelo de probabilidad

El modelo de probabilidad estará definido por una masa de probabilididad en $\{0\}$, i.e.

$$\Pr(X = 0) = \Pr(X \in \{0\}) = \theta_0,$$

y una densidad para la parte continua,

$$f(x|\theta_c) = \theta_c \,\mathsf{exp}\{-x\theta_c\}\mathbb{I}_{(0,\infty)}(x),$$

con $0 < \theta_0 < 1$ y $\theta_c > 0$.

Ejercicio

¿Qué forma toma $F(x|\theta)$ y quién es θ?

Ejemplo

Ejemplo

Pensemos en el modelo de probabilidad con un átomo en $\{0\}$ que admite la posibilidad de tomar valores en la recta real positiva, i.e.

$$\mathcal{X} = \{0\} \cup (0, \infty).$$

Modelo de probabilidad

El modelo de probabilidad estará definido por una masa de probabilididad en $\{0\}$, i.e.

$$\Pr(X = 0) = \Pr(X \in \{0\}) = \theta_0,$$

y una densidad para la parte continua,

$$f(x|\theta_c) = \theta_c \,\mathsf{exp}\{-x\theta_c\}\mathbb{I}_{(0,\infty)}(x),$$

con $0 < \theta_0 < 1$ y $\theta_c > 0$.

Ejercicio

¿Qué forma toma $F(x|\theta)$ y quién es θ?

Ejemplo

Ahora, incorporación de datos (información)...

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

Verosimilitud

Función de verosimilitud

- Enfoque frecuentista: Independencia

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

- Enfoque bayesiano: Independencia condicional

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

Ejercicio

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

Verosimilitud

Función de verosimilitud

- Enfoque frecuentista: Independencia

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

- Enfoque bayesiano: Independencia condicional

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

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

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

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

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

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 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,

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

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 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,

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

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 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

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

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

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

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 caso sencillo donde X es discreta con soporte en $\mathcal{X} = \{0, 1\}$.

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