Repaso Esperanza Condicional

Caso Discreto

Sea (X, Y) un vector aleatorio discreto tal que $\mathrm{E}(\|X\|) < \infty$. Si consideramos los eventos de la forma $\{Y = y\}$ para y en el rango de Y, entoces la probabilidad condicional $\mathrm{P}(X = x|Y = y)$ existe y nos define la distribucion condicional $X|Y = y$.

Moviendo y, definimos la funcion $\psi : Rg\ Y \to \mathrm{R}$

$$\psi(y) = \mathrm{E}(X|Y = y) = \sum_x x\mathrm{P}(X = x|Y = y)$$

Definimos la esperanza condicional de X dado Y como la variable aleatoria que se obtiene al evaluar ψ en Y: $\psi(Y)$.

Chequeen ustedes que

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(X)$$

Caso Continuo

Sea (X, Y) un vector aleatorio continuo con densidad conjunta f_{XY}. Definimos la funcion de densidad de X condicionada a $Y = y$ como

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{XY}(x,y)}{f_Y(y)} & f_Y(y) \neq 0 \\ 0 & f_Y(y) = 0 \end{cases}$$

Es facil chequear que es una funcion de densidad ya que integrar la conjunta en x resulta la marginal de Y que se cancela con la de abajo.

Siguiendo la idea del caso discreto, para cada y defino la funcion

$$\psi(y) = \int_{-\infty}^{+\infty} x f_{X|Y=y}(x)\mathrm{d}x$$

Y definimos la variable aleatoria

$$\mathrm{E}(X|Y) = \psi(Y)$$

. Tambien pueden chequear que

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(X)$$

.

Interpretacion como Proyección Ortogonal

Consideramos el espacio de Hilbert $L^2(\Omega) = \{X : \Omega \to \mathbb{R} : \mathrm{E}(X^2) < \infty\}$ con el producto interno definido por

$$\langle X, Y \rangle = \mathrm{E}(XY).$$

Sean $X, Y \in L^2(\Omega)$ variables aleatorias y sea V el subespacio generado por las funciones medibles de Y.

$$V = \{h(Y) \in L^2(\Omega) : h : \mathbb{R} \to \mathbb{R} \text{ medible}\}$$

Queremos proyectar X sobre el subespacio V. Podemos hacerlo pues V es cerrado con la norma inducida por el producto interno.
Entonces existe Z la proyeccion de X sobre V:

- Minimiza la distancia de X a V.

  $$Z = \mathrm{argmin}_{h(Y) \in V}\ \mathrm{d}(h(Y), X) = \mathrm{E}((h(Y) - X)^2)$$

- $X - Z$ es ortogonal a V.

  $$\langle X - Z, h(Y) \rangle = 0$$

  Esta condicion se puede reescribir como

  $$\mathrm{E}(Xh(Y)) = \mathrm{E}(Zh(Y))$$

  para toda h medible.

Obtenemos de aca la definicion más general de la esperanza condicional.

Definición 1. *Esperanza condicional Dadas X e Y variables aleatorias, definimos la esperanza condicional de X dada Y como la única función Z que cumple*

1. $Z = g(Y)$ *para alguna función g medible.*
2. $\mathrm{E}(Xh(y)) = \mathrm{E}(Zh(Y))$ *para toda h medible.*

Estimación

La idea por detrás de estos primeros ejemplos es tratar de encontrar el parámetro de una cierta distribucin contando solamente con una muestra de valores obtenidos con esa distribución.

Momentos

Supongamos que tenemos variables aleatorias $X_1, ..., X_n$ i.i.d. y $X_1 \sim \text{Exp}(\theta)$. Queremos estimar θ a partir de la muestra. Lo primero que se nos puede ocurrir, amparado por la lei de los grandes números es que como las variables son i.i.d, su promedio se debe parecer a su esperanza.

$$\overline{X} \sim \text{E}(X_1) = \frac{1}{\theta}$$

Entonces puedo despejar

$$\hat{\theta}_m = \frac{1}{\overline{X}}$$

Y obtenemos un estimador de momentos para θ usando el primer momento. Podemos usar momentos de mayor orden, por ejemplo el segundo.

$$\overline{X^2} \sim \text{E}(X_1)^2 + \text{Var}(X_1) = \frac{1}{\theta^2} + \frac{1}{\theta^2} = \frac{2}{\theta^2}$$

Entonces despejamos

$$\hat{\theta}_{m2} = \sqrt{\frac{2}{\overline{X^2}}}$$

Definición 2. *Dadas $X_1, ..., X_n$ i.i.d. cuya distribucion depende del parámetro θ, el estimador de momentos $\hat{\theta}_{mk}$ es el que cumple*

$$\text{E}_{\hat{\theta}_k}(X_1) = \overline{X}$$

Máxima Verosimilitud

.

Sean $X_1, ..., X_n$ son variables aleatorias i.i.d cuya distribucion depende de un parámetro θ. Fijado θ, podemos obtener la densidad conjunta (o la puntual conjunta en el caso discreto) como el producto de las densidades.

$$\text{P}(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} f_X(x_i)$$

Definimos entonces la función de likelihood

$$\mathcal{L}(\theta, x_1, ..., x_n) = \prod_{i=1}^{n} f_{X\theta}(x_i)$$

El estimador de máxima verosimilitud va a ser el θ que maximize esta probabilidad. Es decir, que haga la probabilidad de ver la muestra que vemos sea máxima.

$$\hat{\theta}_{MV} = \text{argmax}_\theta \mathcal{L}(\theta, x)$$

Problema . Sea f una funcion de densidad de una variable aleatoria X.

$$f_X(x) = p\theta^p x^{-p-1} 1_{[\theta,\infty)}(x)$$

Sean $X_1, ..., X_n$ i.i.d con $X_1 \sim X$. Suponiendo θ conocido, calcular $\hat{p}_{MV}$

La función de likelihood es

$$\mathcal{L}(p, x) = \prod p\theta^p x^{-p-1} 1_{[\theta,\infty)}(x_i)$$

Pero en vez de buscar los puntos críticos de esta función, lo haremos para el logaritmo.

$$\ell(p, x) = \prod p\theta^p x^{-p-1} 1_{[\theta,\infty)}(x_i) = \sum \log p + p \log \theta - (p + 1) \log x_i$$

Distribuyendo la suma,

$$\ell(p, x) = n \log p + np \log \theta - (p + 1) \sum \log x_i$$

Queremos maximizar en p. Para eso derivamos y buscamos puntos críticos.

$$\frac{\mathrm{d}}{\mathrm{d}p}\ell(p, x) = \frac{n}{p} + n \log \theta - \sum \log x_i$$

Igualando a cero y despejando p, obtenemos el único punto crítico

$$\hat{p}_{MV} = \frac{n}{\sum \log \frac{x_i}{\theta}}$$

Sabemos que el EMV ya que ℓ es una funcion continua con derivada continua que tiende a cero en los extremos. por lo tanto debe tener un máximo y lo alcanza en su único punto crítico.

Teorema 1. *Invarianza del EMV. $X_1, ..., X_n$ i.i.d. con distribución que depende de θ. Sea g una función biyectiva. Entonces*

$$\widehat{g(\theta)}_{\mathrm{MV}} = g(\hat{\theta}_{\mathrm{MV}})$$