Probabildiad II	2020-2

Lecture : Predicción y Esperanza condicional

Lecturer: F. Baltazar-Larios *Scribe:*

1. Motivación

Sea Y una variable aleatoria con esperanza finita. Es fácil verificar que la función, llamada error cuadrático medio (ECM),

$$h(c) := \mathbb{E}[(Y-c)^2], \tag{1}$$

alcanza un mínimo en $c = \mathbb{E}[Y]$.

Ejercicio 1.1. *Probar que la función $h(c)$ definida en (1) tiene un mímimo en $c = \mathbb{E}[Y]$.*

Si queremos "predecir" el valor que tomará Y (cuando no se tiene más información), el mejor candidato (en el sentido de ECM) es su valor esperado.

2. Predicción

Consideremos a las variables aleatorias X y Y (dependientes) y supongamos que estamos interesados en "predecir" el valor que tomará Y. Si hemos observado que $X = x$ y no se cuenta con información sobre Y, una forma de intentar "predecir" a Y es como una función $f(x)$ de x. La varibale aleatoria $f(X)$ (f medible) se puede construir con el proposito y la llamaremos (estimador) de Y.

Para obtener el mejor estimador de Y usaremos como criterio al ECM, es decir, el mejor estimador de Y sera la función medible f tal que minimice $\mathbb{E}[(Y-f(X))^2]$.

Proposición 2.1.
$$\mathbb{E}[(Y-f(X))^2] \geq \mathbb{E}[(Y-E[Y|X])^2],$$
para toda función medible f.

Proof. Tenemos que

$$\begin{aligned}
\mathbb{E}[(Y-f(X))^2|X] &= \mathbb{E}[(Y-\mathbb{E}[Y|X]+\mathbb{E}[Y|X]-f(X))^2|X] \\
&= \mathbb{E}[(Y-\mathbb{E}[Y|X])^2|X]+\mathbb{E}[(\mathbb{E}[Y|X]-f(X))^2|X] \\
&+ 2\mathbb{E}[(Y-\mathbb{E}[Y|X])(\mathbb{E}[Y|X]-f(X))|X]. \qquad (2)
\end{aligned}$$

Por otro lado tenemos que

$$\begin{aligned}
\mathbb{E}[(Y-\mathbb{E}[Y|X])(\mathbb{E}[Y|X]-f(X))|X] &= \mathbb{E}[Y-\mathbb{E}[Y|X]|X](\mathbb{E}[Y|X]-f(X)) \\
&= (\mathbb{E}[Y|X]-\mathbb{E}[Y|X])(\mathbb{E}[Y|X]-f(X)) \\
&= 0. \qquad (3)
\end{aligned}$$

De las ecuaciones (2) y (3) obtenemos que

$$\mathbb{E}[(Y-f(X))^2|X] \geq \mathbb{E}[(Y-\mathbb{E}[Y|X])^2|X]. \tag{4}$$

Tomando esperanza en ambos lados de la ecuación (4) tenemos que

$$\mathbb{E}[(Y - f(X))^2] \geq \mathbb{E}[(Y - \mathbb{E}[Y|X])^2]. \tag{5}$$

□

3. Modelo de Regresión

Supongamos que la relación entre las variables aleatorias Y y X está dada por la ecuación

$$Y = f(X) + \varepsilon, \tag{6}$$

donde ε es el error de predicción. Si la función relación tiene una forma lineal obtenemos el llamada **modelo lineal simple** dado por

$$Y = \alpha + \beta X + \varepsilon. \tag{7}$$

La pregunta natural es ¿cómo elegir a α y β de tal forma que el ECM que se cometa sea el menor posible en el momento de predecir.

Tenemos que

$$\mathbb{E}[\varepsilon^2] = \mathbb{E}[(Y - (\alpha + \beta X))^2].$$

Ejercicio 3.1. *Probar que la función $g(\alpha, \beta) = \mathbb{E}[(Y - (\alpha + \beta X))^2]$ tiene un mínimo en*

$$\alpha^* = \mathbb{E}[Y] - \beta^* \mathbb{E}[X]$$

y

$$\beta^* = \frac{Cov[X,Y]}{Var[X]}.$$

A $f^*(X) = \alpha^* + \beta^+ X$ se le llama el predictor lineal óptimo (PLO).

Nota 3.1. *Cuando $\mathbb{E}[Y|X] = a + bX$ el PLO es la esperanza condicional.*

4. Ejemplo

Definición 4.1. *El vector aleatorio (X,Y) tiene distribución Normal bivariada si su función de densidad está dada por*

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{ -\frac{\left(\sigma_2^2(x-\mu_1)^2 + \sigma_1^2(y-\mu_2)^2 - 2\sigma_1^2\sigma_2^2\rho(x-\mu_1)(y-\mu_2)\right)}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \right\}, x, y \in \mathbb{R},$$

donde $\mathbb{E}[X] = \mu_1, \mathbb{E}[Y] = \mu_2, Var[X] = \sigma_1^2, Var[Y] = \sigma_2^2$ y $\rho = \rho(X,Y)$.

Supongamos que las varibales aleatorias X y Y tienen distribución normal. En la Figura 1 se ilustra la relación entre variables de este tipo cuando $X \sim \mathrm{N}(0,1)$ y tenemos observados 100 valores de cada variable aleatoria.

Ejercicio 4.1. *Encontrar el PLO cuando $X \sim N(0,1)$ y $Y \sim N(0,1)$ tienen distribución normal y hacer predicción sobre Y.*

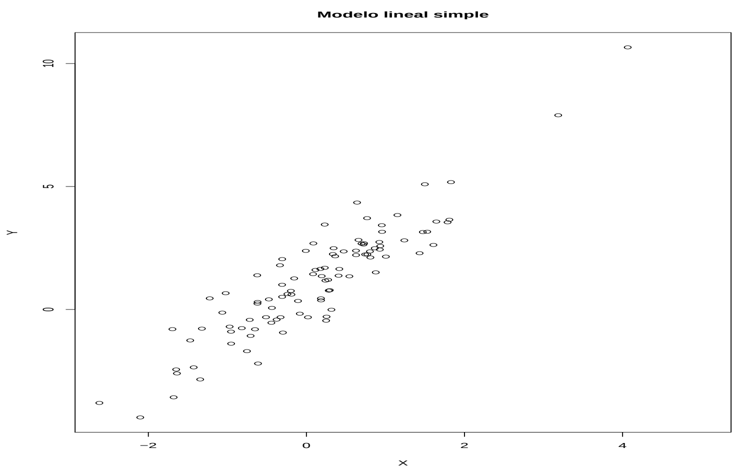

FIGURE 1. Relación entre las variables X e Y.

En la Figura 2 se muestra el PLO con los datos de la Figura 1.

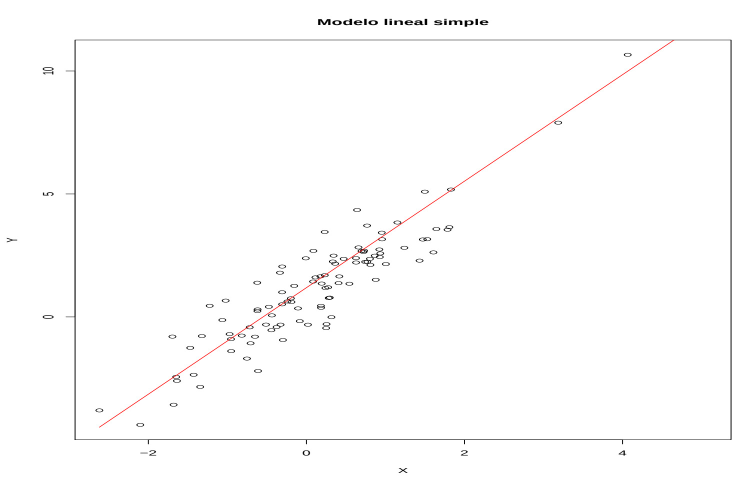

FIGURE 2. Los puntos son los datos y la recta (rojo) es la curva que mejor ajusta y es el PLO.

b

Algorithm 1 Teorema del Límite Central

1: Escribir a la variable aleatoria de interés Z como la suma de v.a.i.i.d.

$$Z = \sum_{i=1}^{n} X_i,$$

donde $\mathbb{E}[X_i] = \mu$ y $Var[X_i] = \sigma^2$

2: Encontrar la $\mathbb{E}[Z] = n\mu$ y $Var[Z] = n\sigma^2$.

3: Concluir que

$$\begin{aligned}
\mathbb{P}[z_1 \leq Z \leq z_2] &= \mathbb{P}\left[\frac{z_1 - n\mu}{\sigma\sqrt{n}} \leq \frac{Z - n\mu}{\sigma\sqrt{n}} \leq \frac{z_2 - n\mu}{\sigma\sqrt{n}}\right] \\
&\approx \Phi\left(\frac{z_2 - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{z_1 - n\mu}{\sigma\sqrt{n}}\right).
\end{aligned}$$

Donde $\Phi(x) = F_X(x)$ para $X \sim N(0,1)$.