José María Martín Luque
José Luis Ruiz Benito
Ricardo Ruiz Fernández de Alba

Probabilidad
D. G. en Ing. Informática y Matemáticas
Universidad de Granada

23 de diciembre de 2019

Entrega 2. Ejercicio de la relación 4.

Ejercicio 14. *Sea* (X, Y) *un vector aleatorio distribuido uniformemente en el paralelogramo de vértices* $(0, 0)$*;* $(2, 0)$*;* $(3, 1)$*;* $(1, 1)$*. Calcular el error cuadrático medio asociado a la predicción de X a partir de la variable Y y a la predicción de Y a partir de la variable aleatoria X. Determinar la predicción más fiable a la vista de los resultados obtenidos.*

Solución. Comenzamos recordando que si estamos estimando el valor de X a partir de una función φ de Y, podemos calcular el error cuadrático medio de dicha estimación utilizando la expresión

$$\mathrm{ECM}(\varphi(Y)) = \mathrm{E}[X^2] - \mathrm{E}[\mathrm{E}[X/Y]^2].$$

Análogamente, si estimamos el valor de Y a partir de una función ψ de X, el error cuadrático medio es

$$\mathrm{ECM}(\psi(X)) = \mathrm{E}[Y^2] - \mathrm{E}[\mathrm{E}[Y/X]^2].$$

En consecuencia, necesitamos hallar:

1. La función de densidad conjunta del vector aleatorio.
2. Las distribuciones marginales de X e Y.
3. Las distribuciones condicionadas X/Y e Y/X.
4. Las esperanzas $\mathrm{E}[X^2]$ y $\mathrm{E}[Y^2]$.
5. Las esperanzas condicionadas $\mathrm{E}[X/Y]$ y $\mathrm{E}[Y/X]$.
6. Las esperanzas $\mathrm{E}[\mathrm{E}[X/Y]^2]$ y $\mathrm{E}[\mathrm{E}[Y/X]^2]$.
7. Las razones de correlación $\eta^2_{X/Y}$ y $\eta^2_{Y/X}$ para determinar qué ajuste es mejor.

Vayamos paso por paso.

1 Sabemos que el vector aleatorio se distribuye de forma uniforme, luego la función de densidad conjunta será de la forma $f(x, y) = k$, con $k \in \mathbb{R}$. Debemos hallar el valor de k teniendo en cuenta la región en la que está definido, que podemos ver en la Figura 1. Para ello integramos

$$\int_0^1 \int_y^{y+2} k \, \mathrm{d}x \, \mathrm{d}y = \int_0^1 \left[kx\right]_y^{y+2} \mathrm{d}y = \int_0^1 2k \, \mathrm{d}y = 2k.$$

Como se trata de una función de densidad dicha integral debe tener valor 1, luego $k = 1/2$.

Figura 1: Región del plano en la que está distribuido el vector aleatorio (X, Y).

2 La distribución marginal de X viene dada por la función de densidad

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}y = \begin{cases} \displaystyle\int_0^x 1/2 \, \mathrm{d}y = x/2 & \text{para } 0 < x < 1, \\ \displaystyle\int_0^1 1/2 \, \mathrm{d}y = 1/2 & \text{para } 1 < x < 2, \\ \displaystyle\int_{x-2}^1 1/2 \, \mathrm{d}y = (3 - x)/2 & \text{para } 2 < x < 3. \end{cases}$$

La distribución marginal de Y viene dada por la función de densidad

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x = \int_y^{y+2} 1/2 \, \mathrm{d}x = 1 \quad \text{para } 0 < y < 1.$$

3 La distribución condicionada X/Y está determinada por la función de densidad

$$f_{X/Y=y_0}(x) = \frac{f(x, y_0)}{f_Y(y_0)} = \frac{1/2}{1} = 1/2 \quad \text{para } 0 < y_0 < 1, \quad y_0 < x < y_0 + 2.$$

Por otro lado, la distribución condicionada Y/X está determinada por la función de densidad

$$f_{Y/X=x_0}(y) = \frac{f(x_0, y)}{f_X(x_0)} = \begin{cases} \dfrac{1/2}{x_0/2} = 1/x_0 & \text{para } 0 < x_0 < 1, \quad 0 < y < x_0; \\ \dfrac{1/2}{1/2} = 1 & \text{para } 1 < x_0 < 2, \quad 0 < y < 1; \\ \dfrac{1/2}{(3 - x_0)/2} = 1/(3 - x_0) & \text{para } 2 < x_0 < 3, \quad x_0 - 2 < y < 1. \end{cases}$$

④ Las esperanzas $\mathrm{E}[X^2]$ y $\mathrm{E}[Y^2]$ vienen dadas por

$$\begin{aligned}
\mathrm{E}[X^2] &= \int_0^1 x^2 \frac{x}{2}\,\mathrm{d}x + \int_1^2 x^2 \frac{1}{2}\,\mathrm{d}x + \int_2^3 x^2 \frac{3-x}{2}\,\mathrm{d}x \\
&= \left[\frac{x^4}{8}\right]_0^1 + \left[\frac{x^3}{6}\right]_1^2 + \left[\frac{x^3}{2} - \frac{x^4}{8}\right]_2^3 \\
&= \frac{1}{8} + \frac{8}{6} - \frac{1}{6} + \left(\left(\frac{27}{2} - \frac{81}{8}\right) - \left(\frac{8}{2} - \frac{16}{8}\right)\right) \\
&= \frac{8}{3},
\end{aligned}$$

y

$$\mathrm{E}[Y^2] = \int_0^1 y^2\,\mathrm{d}y = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3}.$$

⑤ Las esperanzas condicionadas $\mathrm{E}[X/Y]$ y $\mathrm{E}[Y/X]$ vienen dadas por

$$\mathrm{E}[X/Y] = \int x f_{x/y}\,\mathrm{d}x = \int_y^{y+2} x/2\,\mathrm{d}x = \left[\frac{x^2}{4}\right]_y^{y+2} = 1 + y/2 \quad \text{para } 0 < y < 1,$$

y

$$\mathrm{E}[Y/X] = \int y f_{y/x}\,\mathrm{d}y = \begin{cases} \displaystyle\int_0^x y\frac{1}{x}\,\mathrm{d}y = \frac{1}{x}\left[\frac{y^2}{2}\right]_0^x = \frac{x}{2} & \text{para } 0 < x < 1, \\ \displaystyle\int_0^1 y\,\mathrm{d}y = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2} & \text{para } 1 < x < 2, \\ \displaystyle\int_{x-2}^1 y\frac{1}{3-x}\,\mathrm{d}y = \frac{1}{3-x}\left[\frac{y^2}{2}\right]_{x-2}^1 = \frac{x-1}{2} & \text{para } 2 < x < 3. \end{cases}$$

⑥ Finalmente, calculamos las esperanzas $\mathrm{E}[\mathrm{E}[X/Y]^2]$ y $\mathrm{E}[\mathrm{E}[Y/X]^2]$, que vienen dadas por

$$\begin{aligned}
\mathrm{E}[\mathrm{E}[X/Y]^2] &= \int_0^1 (\mathrm{E}[X/Y])^2 f_Y(y)\,\mathrm{d}y = \int_0^1 (1 + y/2)^2 \cdot 1\,\mathrm{d}y \\
&= \int_0^1 (1 + y + y^2/4)\,\mathrm{d}y = \left[y + y^2/2 + y^3/12\right]_0^1 = 1 + \frac{1}{2} + \frac{1}{12} = \frac{19}{12}
\end{aligned}$$

y

$$\begin{aligned}
\mathrm{E}[\mathrm{E}[Y/X]^2] &= \int_0^3 (\mathrm{E}[Y/X])^2 f_X(x)\,\mathrm{d}x \\
&= \int_0^1 \left(\frac{x}{2}\right)^2 \frac{x}{2}\,\mathrm{d}x + \int_1^2 \left(\frac{1}{2}\right)^2 \frac{1}{2}\,\mathrm{d}x + \int_2^3 \left(\frac{x-1}{2}\right)^2 \frac{3-x}{2}\,\mathrm{d}x \\
&= \frac{1}{32} + \frac{1}{8} + \frac{11}{96} = \frac{13}{48}
\end{aligned}$$

Por tanto, el error cuadrático medio obtenido al estimar X a partir de Y es

$$\mathrm{ECM}(\varphi(Y)) = \mathrm{E}[X^2] - \mathrm{E}[\mathrm{E}[X/Y]^2] = \frac{8}{3} - \frac{19}{12} = \frac{13}{12} \approx 1{,}0833$$

y el obtenido al estimar Y a partir de X,

$$\mathrm{ECM}(\psi(X)) = \mathrm{E}[Y^2] - \mathrm{E}[\mathrm{E}[Y/X]^2] = \frac{1}{3} - \frac{13}{48} = \frac{1}{16} = 0{,}0625$$

7 Calculamos primero las esperanzas de X e Y, que necesitaremos posteriormente para calcular las razones de correlación:

$$\begin{aligned}
&\mathrm{E}[X] \\
&= \int_0^3 x f_X(x)\,\mathrm{d}x \\
&= \int_0^1 x\frac{x}{2}\,\mathrm{d}x + \int_1^2 x\frac{1}{2}\,\mathrm{d}x + \int_2^3 x\frac{3-x}{2}\,\mathrm{d}x \\
&= \frac{1}{6} + \frac{3}{4} + \frac{7}{12} = \frac{3}{2}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}[Y] \\
&= \int_0^1 y f_Y(y)\,\mathrm{d}y \\
&= \int_0^1 y\,\mathrm{d}y \\
&= \frac{1}{2}
\end{aligned}$$

Ahora podemos calcular las razones de correlación:

$$\eta^2_{X/Y} = 1 - \frac{ECM(\varphi(Y))}{\mathrm{Var}[X]} = 1 - \frac{ECM(\varphi(Y))}{\mathrm{E}[X^2] - \mathrm{E}[X]^2} = 1 - \frac{\frac{13}{12}}{\frac{8}{3} - \left(\frac{3}{2}\right)^2} = \frac{-8}{5}???$$

$$\eta^2_{Y/X} = 1 - \frac{ECM(\psi(X))}{\mathrm{Var}[Y]} = 1 - \frac{ECM(\psi(X))}{\mathrm{E}[Y^2] - \mathrm{E}[Y]^2} = 1 - \frac{\frac{1}{16}}{\frac{1}{3} - \left(\frac{1}{2}\right)^2} = \frac{1}{4} = 0{,}25$$

La predicción más fiable es la que tiene una mayor razón de correlación.