Juan Pablo Naranjo
201730006
Gabriel Mateo Mejía
201718064
Maria Fernanda Martinez
201716576

CASO DE ESTUDIO 6
Procesos Estocásticos - IELE 4010

20-10-2021

Punto 1

Las mediciones de temperatura en un proceso físico se modelan a través de una variable aleatoria continua Y que sigue una distribución Gaussiana con media 30 y varianza 1.5. Se observa la variable Y con un error aleatorio W aditivo. En particular, se observa la variable aleatoria $X = Y + W$, donde el ruido W se modela como una variable Gaussiana con media 0 y varianza 1. Se sabe que la variable W es independiente de la variable Y.

a)

A continuación se presentan las observaciones de la variable aleatoria Y (en este caso es la variable independiente) graficadas contra las observaciones de la variable aleatoria $X = Y + W$. Se generaron 1000 observaciones de cada variable en Python usando la función *np.random.normal*.

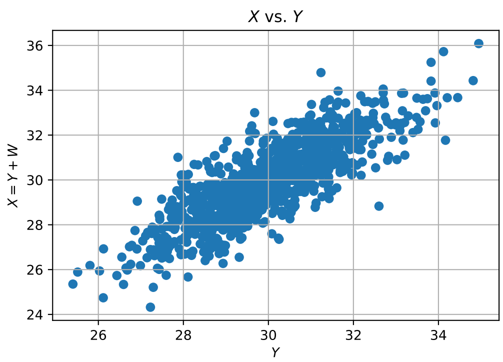

Figura 1: Observaciones de X contra Y

Con la información suministrada, se pueden determinar los siguientes parámetros sobre las variables aleatorias en cuestión:

- $\mu_Y = E[Y] = 30$
- $\mu_W = E[W] = 0$
- $\mu_X = E[X] = E[Y] + E[W] = 30 + 0 = 30$
- $\sigma_Y^2 = 1.5$
- $\sigma_W^2 = 1$

- Como Y y W son independientes, se tiene que $\sigma_X^2 = \text{var}(Y+W) = \sigma_Y^2 + \sigma_W^2 = 1.5 + 1 = 2.5$

b)

Ahora se pide encontrar la distribución Gaussiana multivariable del vector $Z = [Y, X]^T$. Como se sabe que tanto Y como X son variables aleatorias Gaussianas, se sabe que la PDF del vector aleatorio Z corresponde a una PDF Gaussiana conjunta. Esta PDF está dada por la expresión

$$f_Z(z) = \frac{1}{2\pi\sqrt{\det(C_Z)}} e^{\frac{-1}{2}(z-\mu_Z)^T C_Z^{-1}(z-\mu_Z)} \tag{1}$$

en donde se sabe que

- $z = [y, x]^T$
- $\mu_Z = [\mu_Y, \mu_X]^T = [30, 30]^T$
- $C_Z = \begin{bmatrix} \sigma_Y^2 & \text{cov}(Y,X) \\ \text{cov}(Y,X) & \sigma_X^2 \end{bmatrix}$

Aquí, se tiene que como Y y X son independientes y por lo tanto no correlacionadas,

$$\text{cov}(Y,X) = \text{cov}(Y, Y+W) = \text{cov}(Y,Y) + \text{cov}(Y,W) = \sigma_Y^2 + 0 = \sigma_Y^2 = 1.5$$

Entonces, la PDF del vector aleatorio Z está dada por la función en (1), con los parámetros:

$$\begin{aligned} \mu_Z &= \begin{bmatrix} 30 \\ 30 \end{bmatrix} \\ C_Z &= \begin{bmatrix} 1.5 & 1.5 \\ 1.5 & 2.5 \end{bmatrix} \end{aligned} \tag{2}$$

En la figura 2 se pueden observar las curvas de nivel de esta PDF traslapada con los datos generados en el literal anterior.

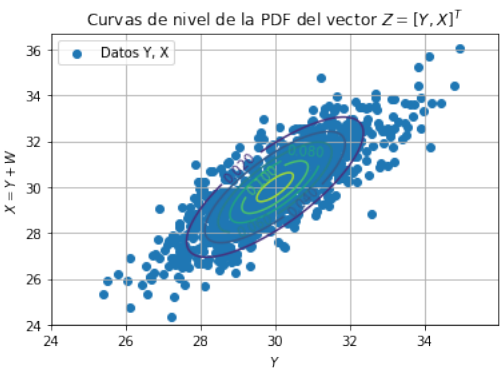

Figura 2: Curvas de nivel de $f_Z(z)$ con las observaciones de Y y X traslapadas

c)

Para encontrar la función del estimador lineal MMSE (LMMSE) de Y basado en X, se partió del hecho de que el estimador lineal tiene la forma

$$G_X = \{a_0 + a_1 X\} \tag{3}$$

para todo a_0, $a_1 \in \mathbb{R}$.

El estimador LMMSE de Y basado en X (dado por la expresión $\hat{Y}_{\text{LMMSE}}(X)$) se encuentra solucionando la siguiente ecuación para $\hat{Y}_{\text{LMMSE}}(X)$:

$$E[(Y - \hat{Y}_{\text{LMMSE}}(X))(a_0 + a_1 X)] = 0 \tag{4}$$

Para solucionar esta ecuación, esta se divide en dos partes del valor esperado. Primero se soluciona la primera mitad

$$\begin{aligned} E[(Y - \hat{Y}_{\text{LMMSE}}(X))a_0] &= 0 \\ E[(Y - \hat{Y}_{\text{LMMSE}}(X))] &= 0 \\ E[Y] &= E[\hat{Y}_{\text{LMMSE}}(X)] \\ E[Y] &= E[a_0 + a_1 X] \\ E[Y] &= a_0 + a_1 E[X] \end{aligned} \tag{5}$$

De las anteriores expresiones se encuentra que

$$a_0 = E[Y] - a_1 E[X] \tag{6}$$

Pasando a la segunda mitad, se tiene que.

$$\begin{aligned} E[(Y - \hat{Y}_{\text{LMMSE}}(X))a_1 X] &= 0 \\ E[YX - \hat{Y}_{\text{LMMSE}}(X)X] &= 0 \\ E[YX - (a_0 + a_1 X)X] &= 0 \\ E[YX] &= a_0 E[X] + a_1 E[X^2] \end{aligned} \tag{7}$$

Ahora, se reemplaza el valor de a_0 en la anterior ecuación:

$$\begin{aligned} E[YX] &= a_0 E[X] + a_1 E[X^2] \\ E[YX] &= (E[Y] - a_1 E[X])\, E[X] + a_1 E[X^2] \\ E[YX] &= E[Y]E[X] - a_1 (E[X])^2 + a_1 E[X^2] \\ E[YX] - E[Y]E[X] &= a_1 (E[X^2] - (E[X])^2) \end{aligned} \tag{8}$$

Conociendo las definiciones de covarianza y de varianza, se sabe que esta última expresión es lo mismo que

$$\text{cov}(Y, X) = a_1 \text{var}(X) \tag{9}$$

y luego

$$a_1 = \frac{\text{cov}(Y, X)}{\text{var}(X)} \tag{10}$$

Finalmente, se encuentra el estimador lineal MMSE de Y basado en X como

$$\begin{aligned} \hat{Y}_{\text{LMMSE}}(X) &= a_0 + a_1 X = E[Y] - \frac{\text{cov}(Y, X)}{\text{var}(X)} E[X] + \frac{\text{cov}(Y, X)}{\text{var}(X)} X \\ \hat{Y}_{\text{LMMSE}}(X) &= E[Y] + \frac{\text{cov}(Y, X)}{\text{var}(X)} [X - E[X]] \end{aligned} \tag{11}$$

d)

A continuación se muestra el estimador $\hat{Y}_{\text{LMMSE}}(X)$ graficado para varios valores de X:

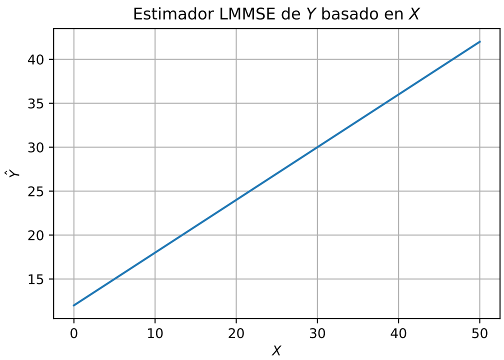

Figura 3: Estimador $\hat{Y}_{\text{LMMSE}}(X)$

Se obtuvo una recta como era de esperarse y se puede decir que, teniendo los valores de X, se pueden estimar los valores de la variable Y ($\hat{Y}_{\text{LMMSE}}(X)$) de acuerdo a la ecuación de esta recta, es decir, a través de

$$\begin{aligned}
\hat{Y}_{\text{LMMSE}}(X) &= E[Y] + \frac{\text{cov}(Y, X)}{\sigma_X^2}[X - E[X]] \\
\hat{Y}_{\text{LMMSE}}(X) &= 30 + \frac{1.5}{2.5}[X - 30] \\
\hat{Y}_{\text{LMMSE}}(X) &= \frac{3}{5}X + 12
\end{aligned} \tag{12}$$

Esta recta es una estimación razonable de la variable aleatoria Y, dado que se sabe que es una variable ruidosa (X está contaminada con un ruido aleatorio). Sin embargo, es precisamente debido a esta contaminación que la estimación no va a ser perfecta y puede llegar a arrojar valores de Y que están un poco alejados de los valores observados en la realidad.

Punto 2

a)

Para empezar, se graficaron las observaciones de los datos como se muestran en la Figura 4.

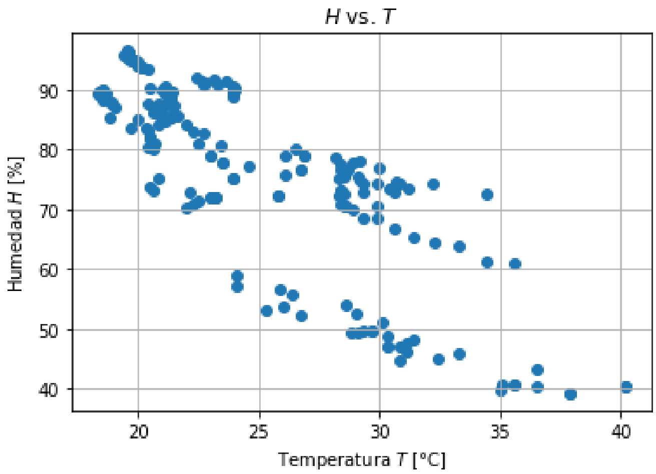

Figura 4: Observaciones en el plano H vs T

Analizando el comportamiento en esta gráfica se puede observar que las variables H y T se encuentran correlacionadas de forma negativa, donde si la temperatura aumenta la humedad disminuye.

Adicionalmente, se observa que los valores más comunes de temperatura se encuentran en el rango de 20 °C a 25 °C y los más comunes de humedad se encuentran entre 80 % y 100 % de humedad relativa.

b)

Se utilizaron las funciones *mean* y *cov* de MATLAB para hallar el vector de medias y la matriz de covarianza de los datos. Los resultados obtenidos fueron los siguientes:

$$\mu_Z = \begin{bmatrix} 24.79 \\ 76.74 \end{bmatrix} \; C_Z = \begin{bmatrix} 26.71 & -68.84 \\ -68.84 & 253.19 \end{bmatrix}$$

Por lo tanto, la función de densidad de probabilidad corresponde a la mostrada a continuación:

$$f_z(z) = \frac{1}{2\pi\sqrt{\det(C_Z)}} e^{\frac{-1}{2}(z-\mu_Z)^T C_Z^{-1}(z-\mu_Z)} \tag{13}$$

Seguido a esto, los datos del archivo *datosCaso6.txt* fueron traslapados con las curvas de nivel de la PDF mostrada en la Ecuación 13 y los resultados se muestran en la Figura 5.

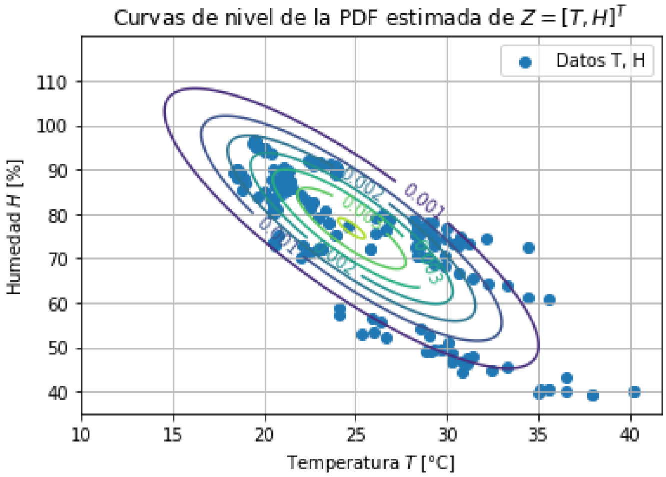

Figura 5: PDF con los datos de Z

c)

Con la varianza y la media del vector Z (Humedad y temperatura) encontrada en el literal b) se requiere calcular los siguientes estimadores: MMSE lineal de T dada la observación de H y MMSE lineal de H dada la observación de T.

Como se definió en la clase magistral, el estimador lineal "Minimum Mean Squared Error" (MMSE) se calcula de la siguiente forma:

$$\hat{H}_{\text{MMSE}}(T) = E[H] + \frac{\text{cov}(H,T)}{\sigma_T^2}(T - E[T]) \tag{14}$$

$$\hat{H}_{\text{MMSE}}(T) = 76.74 + \frac{-68.84}{26.71}(T - 24.79) \tag{15}$$

El valor de los parámetros de cada estimador no coincide dado que la variable a estimar es diferente. Por una parte se estimará la temperatura con base en los datos de humedad y por otra parte se estimará la humedad con base en los datos de temperatura.

$$\hat{T}_{\text{MMSE}}(H) = E[T] + \frac{\text{cov}(T,H)}{\sigma_H^2}(H - E[H]) \tag{16}$$

$$\hat{T}_{\text{MMSE}}(H) = 24.79 + \frac{-68.84}{253,19}(H - 76.74) \tag{17}$$

Con los estimadores ya definidos, se calculan los valores de humedad y temperatura que eran desconocidos del archivo *datosNuevosCaso6.txt* como se ve en la Figura 6.

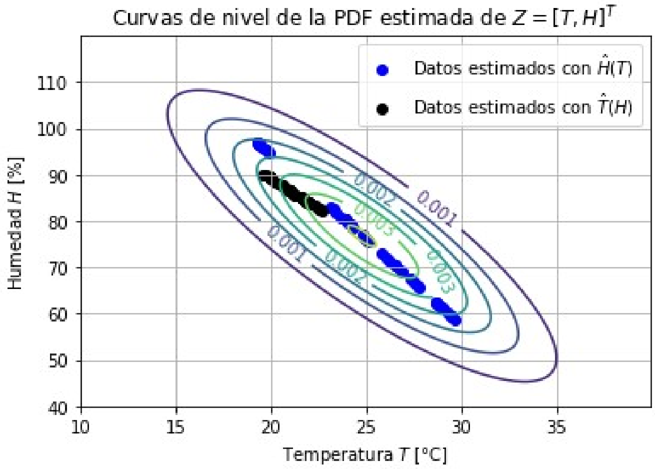

Figura 6: PDF con los datos de Z

En la Figura 6 se puede notar que los datos estimados de cada variable son consistentes con la PDF estimada del vector aleatorio Z, donde se evidencia que se distribuyen de forma similar a los datos mostrados en el literal b.

Finalmente, dado a que el estimador que se encontró para cada variable aleatoria (H y T) es un estimador lineal, los datos que faltaban (y que fueron estimados) dibujan una recta de acuerdo a las ecuaciones 15 y 17, respectivamente. Adicionalmente, se puede notar que estos datos están alrededor de la media μ_Z (temperatura y humedad).