G5P11

Verifique los resultados analíticos obtenidos en el ítem 10d escribiendo un programa que realice la siguiente simulación numérica:

I para cada x_i genere al azar un y_i de la distribución gaussiana $N(\hat{a}_1 + \hat{a}_2 x_i, \sigma)$.

II ajuste una recta a los (x_i, y_i) generados, y prediga el valor y_a para $x_a = 0.5$.

Repita 1000 veces los pasos I.–II., construyendo un histograma con los valores de y_a, y dibuje sobre éste la gaussiana con el valor esperado y el error de y_a calculado teoricamente en 10d.

Considero N mediciones (x_i, y_i), donde los y_i son independientes y todos con el mismo σ, y los x_i son números sin error. Los parámetros de la recta $y = a_1 + a_2 x$ que mejor ajusta los datos surgen de la fórmula de cuadrados mínimos:

$$\begin{cases} \hat{a}_1 = (\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i)/\Delta \\ \hat{a}_2 = (N \sum x_i y_i - \sum x_i \sum y_i)/\Delta \end{cases}$$

$$\text{con } \Delta = N \sum x_i^2 - (\sum x_i)^2$$

Como los x_i no son variables aleatorias, $a_1 = a_1(y_1, ..., y_n)$ y $a_2 = a_2(y_1, ..., y_n)$.
La matriz de covarianza para estos parámetros es

$$Cov(\hat{a}_1, \hat{a}_2) = \frac{\sigma^2}{\Delta} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{bmatrix}$$

Aplicando este método a los datos del problema 10, obtuve los parámetros $\hat{a}_1 = (1{,}452 \pm 0{,}721)$ y $\hat{a}_2 = (0{,}799 \pm 0{,}286)$.

Una vez conocidos los parámetros a_1 y a_2, puedo definir una recta $y_a = a_1 + a_2 x_a$, donde las variables aleatorias son a_1 y a_2 (y por lo tanto y_a), pero no x_a. Para un x_a dado, puedo calcular su y_a correspondiente, cuya varianza viene dada por

$$Var(y_a) = \left(\frac{\partial y_a}{\partial a_1}\right)^2 Var(a_1) + \left(\frac{\partial y_a}{\partial a_2}\right)^2 Var(a_2) + 2\frac{\partial y_a}{\partial a_1}\frac{\partial y_a}{\partial a_2} Cov(a_1, a_2) =$$
$$Var(a_1) + x_a^2 Var(a_2) + 2x_a Cov(a_1, a_2) = \frac{\sigma^2}{\Delta}(\sum x_i^2 + x_a^2 N - 2x_a \sum x_i) \equiv \sigma_{y_a}$$

Por ejemplo, para $x_a = 0{,}5$, $y_a = (1{,}852 \pm 0{,}579)$.

Para cada x_i del punto 10c, genero un y_i de la distribución $N(\hat{a}_1 + \hat{a}_2 x_i, \sigma)$. Tengo entonces 11 puntos (x_i, y_i). Aplicando el método de cuadrados mínimos, al igual que en el problema 10, obtengo los parámetros $\hat{a}_1$ y $\hat{a}_2$. Con estos parámetros, predigo el valor de y_a para $x_a = 0{,}5$, y guardo el valor obtenido. Repito este proceso 1000 veces, obteniendo el histograma de la Fig. 1.

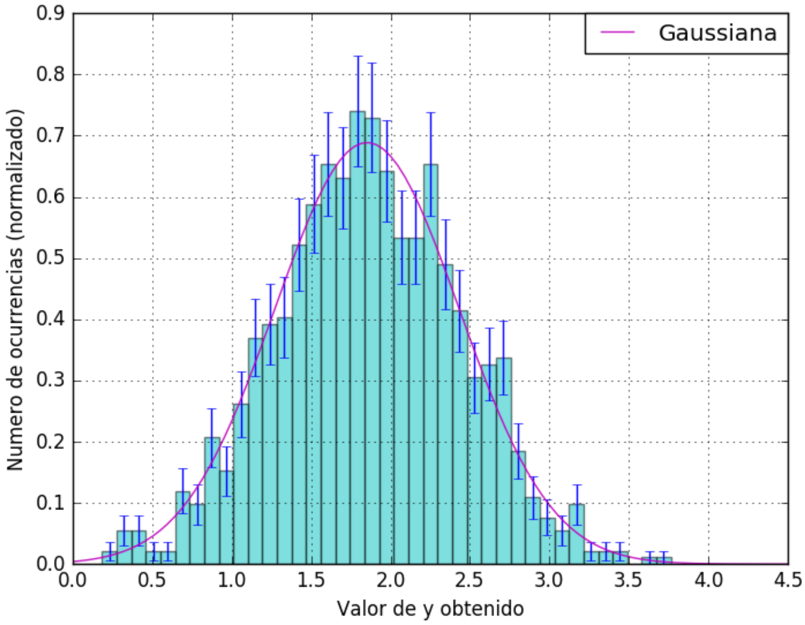

Figura 1: Histograma de los valores obtenidos para y_a.

La curva graficada sobre el histograma corresponde a la gaussiana $N(y_a, \sigma_{y_a})$, en este caso, N(1.852, 0.579).

En el ejercicio 10 había tomado 11 puntos y los había ajusatdo por una recta utilizando cuadrados mínimos. Ahora, con los mismos x_i, genero y_i provenientes de la distribución $N(\hat{a}_1 \ + \ \hat{a}_2 \ x_i, \ \sigma)$: la esperanza de estos y_i es el valor de y correspondiente a x_i en la recta del punto 10. El 68 % de las veces, el valor de y_i va a caer en el intervalo $(\hat{a}_1 + \hat{a}_2 x_i \pm \sigma)$. Realizo un ajuste lineal sobre estos nuevos puntos, obteniendo nuevos valores para $\hat{a}_1$ y $\hat{a}_2$. Repitiendo el proceso 1000 veces, obtengo un conjunto de valores para los parámetros que, por el teorema central del límite, van a tener distribución gaussiana, como se observa en la Fig.2

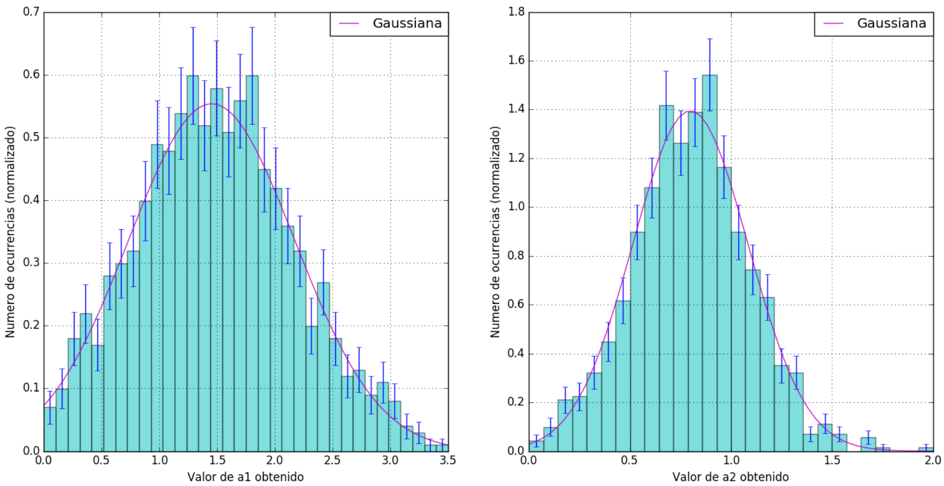

Figura 2: Histograma de los valores obtenidos para a_1 y a_2 respectivamente.

Las gaussianas superpuestas a los histogramas son, respectivamente, $N(1{,}452, 0{,}721)$ y $N(0{,}799, 0{,}286)$, es decir, con los valores de μ y σ obtenidos en el problema 10. Si los parámetros fluctúan de manera gaussiana alrededor del valor obtenido en el punto anterior, e y_a es una variable aleatoria que depende solamente de estos dos (de hecho, la dependencia es lineal), también fluctuará de manera gaussiana alrededor del valor obtenido en el punto 10.

Obs.: Los errores asignados a los histogramas son poissonianos porque el número de entradas en un bin de un histograma es una variable aleatoria con distribución binomial, ya que representa cuántas mediciones "cayeron" en un determinado bin. Si la cantidad de bines tiende a infinito (entonces la probabilidad de caer en un determinado bin tiende a cero), estamos en el límite de Poisson, que es lo que ocurre en este caso.