

Ley de propagación de errores, justificación de la
ley $\Sigma_y = J_x \Sigma_x J_x^T$

Joaquín Gómez

24 de mayo de 2025

1. Introducción

A la hora de medir un evento cuya entrada es una variable aleatoria X , y cuya salida o variable dependiente es Y , otra variable aleatoria de la cual no sabemos más que su relación con X , surge el siguiente problema: ¿cómo podemos aproximar Y para cualquier valor, dado que conocemos X y su relación con Y ? Esta pregunta es trivial si se trata de una relación totalmente lineal entre X e Y , lo que implica

$$Y = aX \quad (1)$$

para un cierto número real a . Sin embargo, en el mundo real estas relaciones no se cumplen de forma perfecta.

2. Caso unidimensional

Para este caso, donde tenemos un número de entradas que definiremos como $N = 1$ y salidas $P = 1$, podemos encontrar una función $f : R \rightarrow R$ tal que

$$Y = f(X) \quad (2)$$

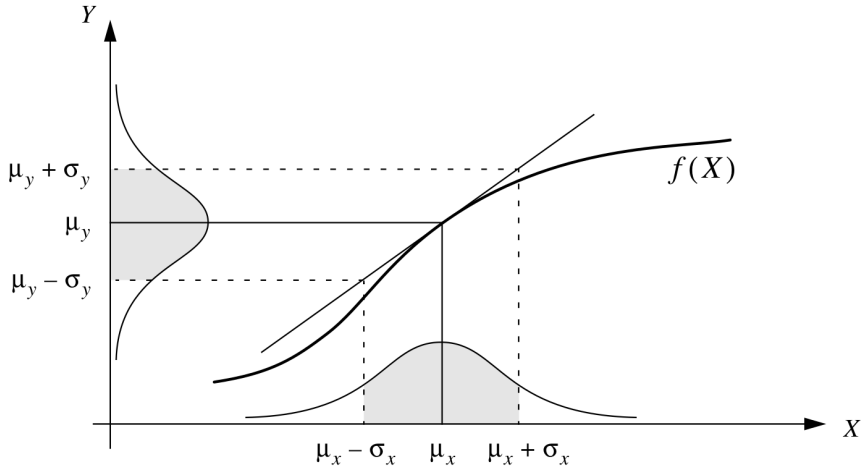


Figura 1: Propagación de error unidimensional

En la figura 1 se muestra un ejemplo, en el cual, a través de una aproximación lineal, podemos determinar la distribución Gaussiana de Y . Para hacerlo, debemos utilizar una aproximación de Taylor de primer orden. En este caso es

$$Y = f(\mu_y) + \left. \frac{\partial f}{\partial X} \right|_{X=\mu_x} (X - \mu_x) \quad (3)$$

El significado del signo de “aproximado” es que Y no va a representar el valor *real* de la función, pero puede ser una aproximación suficientemente buena, siempre y cuando la correspondencia entre X e Y sea “suficientemente lineal” en un intervalo dado. Se considera el intervalo $[\mu_x - \sigma_x, \mu_x + \sigma_x]$, pero también

es válido, en algunos casos, considerar $[\mu_x - 2\sigma_x, \mu_x + 2\sigma_x]$. Esto último siempre y cuando la desviación estándar no sea muy grande.

Analicemos lo siguiente, dados los valores reales de Y , podríamos conocer su varianza y su media. Supongamos que sabemos la varianza real de Y , dada como σ_0^2 y su media μ_0 . Ahora, si suponemos

$$\begin{aligned}\sigma_o^2 &\approx \sigma_y^2 \\ \mu_o &\approx \mu_y\end{aligned}\tag{4}$$

entonces podemos utilizar la aproximación para encontrar

$$\begin{aligned}\mu_y &= f(\mu_x) \\ \sigma_y &= \left. \frac{\partial f}{\partial X} \right|_{X=\mu_x} \sigma_x\end{aligned}\tag{5}$$

En una situación real, debemos inferir los valores de μ_x y σ_x , o intentar adivinarlos. En esta situación, se tiene que dar que nuestro valor de muestra $X = x^*$ sea cercano a $E[X]$. Basándonos en este principio, podemos a su vez intentar hallar un valor aproximado para s_x^* , que deberá ser cercano a $E[(X - \mu_x)^2]^{1/2}$. Si cumplimos ambas condiciones, obtendremos un valor de y^* no tan alejado de $E[Y]$, y una desviación estándar s_y^* también suficientemente cerca de $E[(Y - \mu_y)^2]^{1/2}$.

Matemáticamente, dada una muestra $X = x^* \approx E[X]$ y un valor inferido $s_x^* \approx \sigma_x$, tenemos que, la aproximación $Y = y^*$ y s_y^* cumplen

$$\begin{aligned}y^* &\approx E[Y] \\ s_y^* &\approx \sigma_y\end{aligned}\tag{6}$$

En este caso, lo que se intenta decir es que las medidas reales que nosotros tomemos, pueden ser similares a las que resulten de nuestra aproximación, siempre y cuando se cumpla que los valores de entrada sean representativos. En otras palabras, si tomamos una muestra y corresponde a un valor cercano al promedio, entonces debería cumplirse que la salida de nuestro sistema también sea cercana a su promedio.

3. Sistema con varias entradas y una salida (campo escalar)

Para este caso, tendremos una serie de $N = n$ entradas X_1, X_2, \dots, X_n y $P = 1$ salida. La función $f : R^n \rightarrow R$ da la relación

$$f(X_1, X_2, \dots, X_n) = Y\tag{7}$$

Nuevamente, utilizaremos la aproximación de Taylor de primer orden

$$Y = f(\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)}) + \sum_{i=1}^n \frac{\partial f}{\partial X_i} (\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)}) (X_i - \mu_x^{(i)})\tag{8}$$

(nótese que la derivada parcial está aplicada en el vector μ_x). Así, podemos encontrar la media y desviación estándar de Y , definiremos dos variables para simplificar el cálculo, que son $a_0 = f(\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)})$ y $a_i = \frac{\partial f}{\partial X_i}(\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)})$, de forma que obtenemos

$$\begin{aligned}
\mu_y &= E[Y] \\
&= E \left[a_0 + \sum_i a_i (X_i - \mu_x^{(i)}) \right] \\
&= E[a_0] + E \left[\sum_i a_i (X_i - \mu_x^{(i)}) \right] \\
&= a_0 + \sum_i a_i E[X_i] - \sum_i a_i E[\mu_x^{(i)}] \\
&= a_0 + \sum_i a_i \mu_x^{(i)} - \sum_i a_i \mu_x^{(i)} \\
&= a_0
\end{aligned} \tag{9}$$

Para hallar la desviación estándar, el camino es más largo, para ello utilizamos la varianza

$$\begin{aligned}
\sigma_y^2 &= E[(Y - \mu_y)^2] \\
&= E \left[\left(\sum_i a_i (X_i - \mu_x^{(i)}) \right)^2 \right] \\
&= E \left[\sum_i a_i (X_i - \mu_x^{(i)}) \sum_j a_j (X_j - \mu_x^{(j)}) \right] \\
&= E \left[\sum_i a_i^2 (X_i - \mu_x^{(i)})^2 + \sum_{i \neq j} a_i a_j (X_i - \mu_x^{(i)}) (X_j - \mu_x^{(j)}) \right] \\
&= E \left[\sum_i a_i^2 (X_i - \mu_x^{(i)})^2 \right] + E \left[\sum_{i \neq j} a_i a_j (X_i - \mu_x^{(i)}) (X_j - \mu_x^{(j)}) \right] \\
&= \sum_i a_i^2 E[(X_i - \mu_x^{(i)})^2] + \sum_{i \neq j} a_i a_j E[(X_i - \mu_x^{(i)}) (X_j - \mu_x^{(j)})] \\
&= \sum_i a_i^2 \text{Var}(X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j) \\
&= \sum_i \left(\frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \right)^2 \text{Var}(X) + \sum_{i \neq j} \frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \frac{\partial f}{\partial X_j} \Big|_{X=\mu_x} \text{Cov}(X_i, X_j) \\
&= \sum_i \left(\frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \right)^2 \sigma_x^2 + \sum_{i \neq j} \frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \frac{\partial f}{\partial X_j} \Big|_{X=\mu_x} \Sigma_{ij}^2
\end{aligned} \tag{10}$$

Entonces, la desviación estándar es

$$\sigma_y = \sqrt{\sum_i \left(\frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \right)^2 \sigma_x^2 + \sum_{i \neq j} \frac{\partial f}{\partial X_i} \Big|_{X=\mu_x} \frac{\partial f}{\partial X_j} \Big|_{X=\mu_x} \Sigma_{ij}^2} \quad (11)$$

Si las variables independientes no están relacionadas entre sí, es decir, $\text{Cov}(X_i, X_j) = 0$ para $j \neq i$, entonces la fórmula se resume en

$$\sigma_y^2 = \sum_i \left(\frac{\partial f}{\partial X_i} \right)^2 \sigma_x^2 \quad (12)$$

Donde suponemos que aplicamos la derivada parcial en un punto fijo, que será en este caso, la media μ_x . Omitimos esto para abreviar la fórmula y evitar así demasiada sintaxis.

4. Sistema con múltiples entradas y salidas (campo vectorial)

Este caso es una generalización del caso anterior, tomamos las ideas que ya conocemos, para encontrar la fórmula más recurrente para la propagación de errores, que es

$$\Sigma_y = F_x \Sigma_x F_x^T \quad (13)$$

¿Cómo encaramos el problema? Antes que nada, definiremos la función que relaciona una serie de entradas X_1, X_2, \dots, X_n a una serie de salidas f_1, f_2, \dots, f_p . Tendremos que el sistema tiene $N = n$ y $P = p$ entradas y salidas respectivamente. De forma que la función que lo representa es $f : R^n \rightarrow R^p$, así

$$\begin{cases} Y_1 = f_1(X_1, X_2, \dots, X_n) \\ Y_2 = f_2(X_1, X_2, \dots, X_n) \\ \vdots \\ Y_p = f_p(X_1, X_2, \dots, X_n) \end{cases} \quad (14)$$

Para esta sección, definiremos $Y = (Y_1, Y_2, \dots, Y_p)$. Buscaremos una fórmula para la aproximación de Taylor de primer orden, como vimos en la sección anterior

$$Y_i = f_i(X_1, X_2, \dots, X_n) + \sum_{k=1}^n \frac{\partial f_i}{\partial X_k} (X_1, X_2, \dots, X_n) (X_k - \mu_x^k) \quad (15)$$

Si observamos bien, podemos representar al segundo miembro del lado derecho de la ecuación como el producto escalar $\nabla f_i \cdot (X - \mu_x)$

$$Y_i = f_i(X_1, X_2, \dots, X_n) + \nabla f_i \cdot (X - \mu_x) \quad (16)$$

Asímismo, podemos escribir Y como una suma de dos vectores, donde $X = (X_1 \ X_2 \ \dots \ X_n)$ y $\mu_x = (\mu_x^{(1)} \ \mu_x^{(2)} \ \dots \ \mu_x^{(n)})$

$$Y = \begin{bmatrix} f_1(\mu_x) \\ f_2(\mu_x) \\ \vdots \\ f_p(\mu_x) \end{bmatrix} + \begin{bmatrix} \nabla f_1(\mu_x) \cdot (X - \mu_x) \\ \nabla f_2(\mu_x) \cdot (X - \mu_x) \\ \vdots \\ \nabla f_p(\mu_x) \cdot (X - \mu_x) \end{bmatrix} \quad (17)$$

Si separamos el vector con los gradientes ∇f_i , y los expandimos, obtendríamos la matriz Jacobiana

$$J_f(X) = \begin{bmatrix} \nabla f_1(\mu_x) \\ \nabla f_2(\mu_x) \\ \vdots \\ \nabla f_p(\mu_x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(X)}{\partial X_1} & \cdots & \frac{\partial f_1(X)}{\partial X_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(X)}{\partial X_1} & \cdots & \frac{\partial f_p(X)}{\partial X_n} \end{bmatrix} \quad (18)$$

de forma tal que, podemos reescribir (17) como

$$Y = f(\mu_x) + J_f(\mu_x) \cdot (X - \mu_x) \quad (19)$$

Debemos encontrar la covarianza de Y , para ello, sin perder generalidad, buscamos el elemento en la posición ij en la matriz de covarianza Σ_y y el elemento i en el vector μ_y . Definimos $a_{i0} = f_i(\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)})$ y $a_{ij} = \frac{\partial f_i}{\partial X_j}(\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(n)})$

$$\begin{aligned} \mu_y^{(i)} &= E[Y_i] \\ &= E \left[a_{i0} + \sum_j a_{ij}(X_j - \mu_x^{(j)}) \right] \\ &= E[a_{i0}] + E \left[\sum_j a_{ij} X_j \right] - E \left[\sum_j a_{ij} \mu_x^{(j)} \right] \\ &= a_{i0} + \sum_j a_{ij} E[X_j] - \sum_j a_{ij} \mu_x^{(j)} \\ &= a_{i0} + \sum_j a_{ij} \mu_x^{(j)} - \sum_j a_{ij} \mu_x^{(j)} \\ &= a_{i0} \end{aligned} \quad (20)$$

$$\begin{aligned}
\Sigma_y^{(ij)} &= \text{Cov}(Y_i, Y_j) \\
&= E \left[(Y_i - \mu_y^{(i)})(Y_j - \mu_y^{(j)}) \right] \\
&= E \left[\sum_k a_{ik}(X_k - \mu_x^{(k)}) \sum_l a_{jl}(X_l - \mu_x^{(l)}) \right] \\
&= E \left[\sum_k a_{ik}^2 (X_k - \mu_x^{(k)})^2 + \sum_{l \neq k} \sum a_{jl} a_{ik} (X_l - \mu_x^{(l)})(X_k - \mu_x^{(k)}) \right] \\
&= E \left[\sum_k a_{ik}^2 (X_k - \mu_x^{(k)})^2 \right] + E \left[\sum_{l \neq k} \sum a_{jl} a_{ik} (X_l - \mu_x^{(l)})(X_k - \mu_x^{(k)}) \right] \\
&= \sum_k a_{ik}^2 E \left[(X_k - \mu_x^{(k)})^2 \right] + \sum_{l \neq k} \sum a_{jl} a_{ik} E \left[(X_l - \mu_x^{(l)})(X_k - \mu_x^{(k)}) \right] \\
&= \sum_k a_{ik}^2 \Sigma_x^{kk} + \sum_{l \neq k} \sum a_{jl} a_{ik} \Sigma_x^{(lk)} \\
&= \sum_{k,l} \sum a_{jl} a_{ik} \Sigma_x^{(lk)} \\
&= \sum_{k,l} \sum \frac{\partial f_j}{\partial X_l} \frac{\partial f_i}{\partial X_k} \Sigma_x^{(lk)}
\end{aligned} \tag{21}$$

Recordemos que la matriz de covarianza de X está dada por

$$\Sigma_x = \begin{bmatrix} \Sigma_x^{(11)} & \Sigma_x^{(12)} & \cdots & \Sigma_x^{(1n)} \\ \Sigma_x^{(21)} & \Sigma_x^{(22)} & \cdots & \Sigma_x^{(2n)} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_x^{(n1)} & \Sigma_x^{(n2)} & \cdots & \Sigma_x^{(nn)} \end{bmatrix} \tag{22}$$

La expresión (21) indica que estamos multiplicando un vector columna de la matriz J_x , por un vector fila de J_x , por todos los elementos de la matrix Σ_x .

$$\sum_{k,l} \frac{\partial f_j}{\partial X_l} \frac{\partial f_i}{\partial X_k} \Sigma_x^{(lk)} = (J_x)_{:,j} \cdot \Sigma_x \cdot (J_x)_{i,:} \tag{23}$$

Aquí, utilizamos la notación “ $:,j$ ” para indicar que extraemos la columna j de la matriz, y “ $i,:$ ” la fila i . Finalmente, si cada entrada de $\Sigma_y^{(ij)}$ corresponde a (23), tenemos que la matriz de covarianza de Y es

$$\Sigma_y = J_x \Sigma_x J_x^T \tag{24}$$

La transpuesta de J_x indica que estamos multiplicando sus filas.

La ecuación (24) expresa un cambio de base (lineal) aplicado a la matriz de covarianza de X . Básicamente, estamos transformando la covarianza de X a través de su relación con Y , esta relación está dada por la matriz Jacobiana, ya que expresa la tasa de cambio de cada salida Y_i con respecto a sus variables X_j . El significado intuitivo es el mismo que en el caso unidimensional.