Repaso de transformación de variable

Una función de una variable aleatoria es también una variable aleatoria.

Parece natural plantearse: si se conoce la función densidad de probabilidad de la primera, cómo es la expresión de la densidad de probabilidad de la nueva variable.

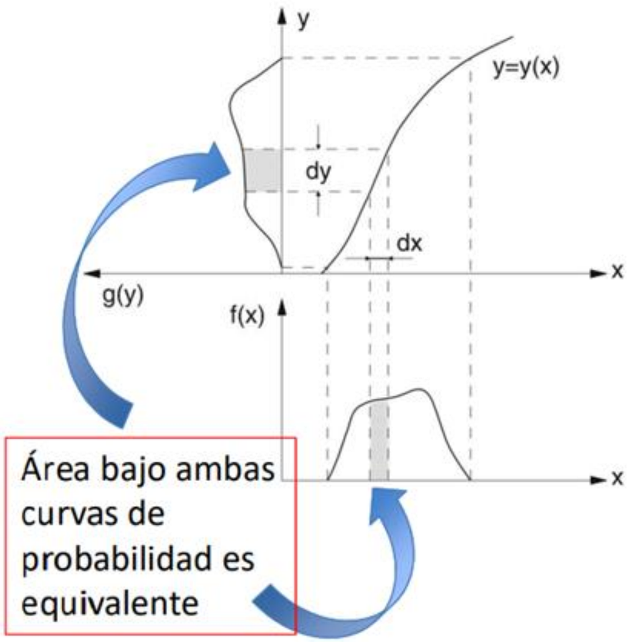
Variable aleatoria con densidad de probabilidad conocida: $X, f(x)$

Función que transforma de X a Y: $Y = Y(X)$

Cuál es la expresión de la función de densidad de probabilidad de Y? $g(y)$

Caso para una función Y(X) **biyectiva** y **estrictamente creciente**

$X, f(x) \qquad Y = Y(X) \qquad g(y)$



$$P(x \leq X < x + dx) = P(y \leq Y < y + dy)$$

Por correspondencia de áreas:

$$f(x)\, dx = g(y)\, dy$$

$$g(y) = f(x(y))\, \frac{dx}{dy}$$

Expresión general para la función densidad de probabilidad de Y:

$$g(y) = f(x(y)) \left|\frac{dx}{dy}\right| \qquad dx = \left|\frac{dx}{dy}\right| dy$$

Transformación de variables para el caso función de densidad de probabilidad conjunta de dos variables

Es la generalización del caso de transformación para una variable.

El punto de partida son dos variables aleatorias con función densidad de probabilidad conjunta:

$$(X, Y) \sim f(x, y)$$

Sean $U = U(X,Y)$ y $V = V(X,Y)$ funciones deterministas que transforman el plano (X,Y) en el plano (U,V)

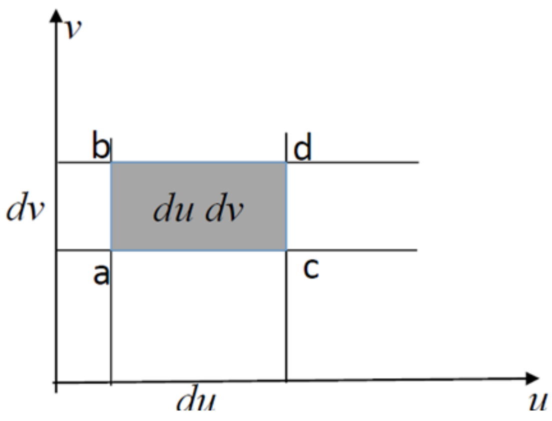
$$U = U(X,Y) \text{ y } V = V(X,Y)$$

Como (X,Y) son variables aleatorias, (U,V) también lo son.

Es natural preguntarse cuál será la expresión de la función de densidad de probabilidad conjunta para (U,V)?

$$(U, V) \sim g(u, v)$$

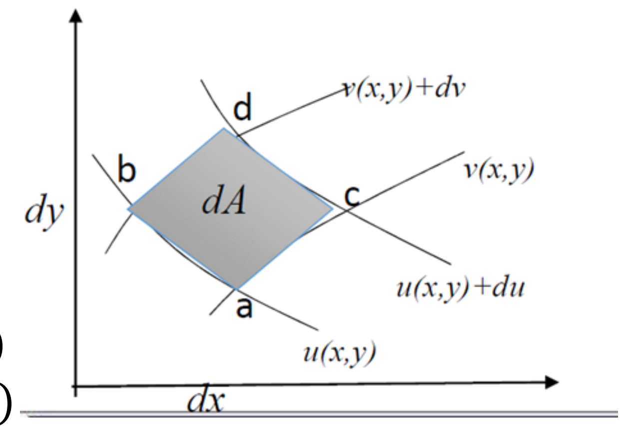
Dominio de $g(u,v)$: plano (u,v)



$$U = U(X,Y)$$
$$V = V(X,Y)$$

$$X = X(U,V)$$
$$Y = Y(U,V)$$

Dominio de $f(x,y)$: plano (x,y)



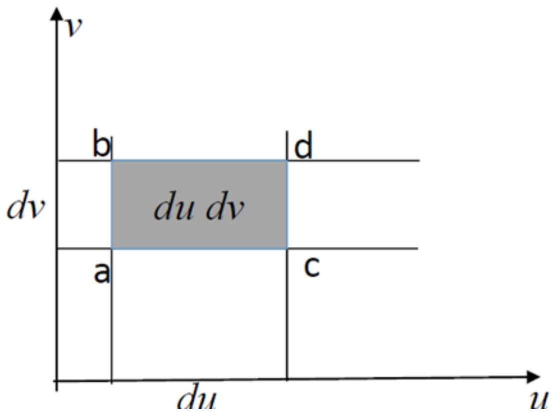
Condición de igualdad de probabilidad entre los dos planos:

$$P\big((X,Y)\ \in dA\big) = P(u < U \leq u + du, v < V \leq v + dv)$$

$$f(x,y)\, dA = g(u,v)\, du\, dv$$

El punto clave es analizar la relación de áreas que hay entre un elemento diferencial en el plano (u,v), definido por los puntos (u,v) y $(u+du, v+dv)$ y el correspondiente elemento en el plano (x,y). Es decir, necesitamos la relación entre dA y $du\, dv$, que es el área diferencial en el plano (u,v).

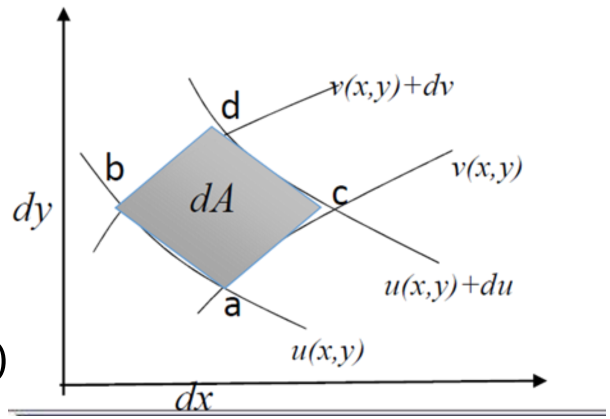
Plano (u, v)



$$U = U(X, Y)$$
$$V = V(X, Y)$$

$$X = X(U, V)$$
$$Y = Y(U, V)$$

Plano (x, y)



$$x_a = x(u, v) \qquad x_b = x(u, v + dv)$$
$$y_a = y(u, v) \qquad y_b = y(u, v + dv)$$

$$x_c = x(u + du, v)$$
$$y_c = y(u + du, v)$$

Se puede demostrar que:

$$dA = \begin{vmatrix} 1 & x_a & y_a \\ 1 & x_b & y_b \\ 1 & x_c & y_c \end{vmatrix}$$

Valor absoluto del determinante.

Luego, por desarrollo de Taylor:

$$x_b = x(u, v + dv) \quad \Rightarrow \quad x_b = x(u,v) + \frac{\partial x}{\partial v} dv$$

$$y_b = y(u, v + dv) \quad \Rightarrow \quad y_b = y(u,v) + \frac{\partial y}{\partial v} dv$$

$$x_c = x(u + du, v) \quad \Rightarrow \quad x_c = x(u,v) + \frac{\partial x}{\partial u} du$$

$$y_c = y(u + du, v) \quad \Rightarrow \quad y_c = y(u,v) + \frac{\partial y}{\partial u} du$$

Reemplazando y haciendo algebra:

$$dA = \begin{vmatrix} 1 & x_a & y_a \\ 1 & x_b & y_b \\ 1 & x_c & y_c \end{vmatrix} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} du\, dv = J\left(\frac{x,y}{u,v}\right) du\, dv$$

A $J\left(\frac{x,y}{u,v}\right)$ se lo denomina ‘Jacobiano de la transformación’ (definido por el valor absoluto del determinante).

Finalmente, reemplazando en:

$$f(x,y)\,dA = g(u,v)\,du\,dv$$

$$f(x,y)\,J\left(\frac{x,y}{u,v}\right)\,du\,dv = g(u,v)\,du\,dv$$

Encontramos la expresión final de la transformación:

$$g(u,v) = f\big(x(u,v),y(u,v)\big)\,J\left(\frac{x,y}{u,v}\right)$$ donde: $$J\left(\frac{x,y}{u,v}\right) = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

Generalización de elementos y conceptos para más de dos variables

Se define un vector de dimensión n, donde cada componente es una v.a.:

$$\vec{X} = (X_1, X_2, \dots X_n)^T$$

La **función de distribución conjunta** se define como:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

$$F(\vec{x}) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$ N.V/M. *

La **función de densidad de probabilidad conjunta** se define como:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1\, \partial x_2\, \dots \partial x_n}$$

$$f(\vec{x}) = \frac{\partial^n F(\vec{x})}{\partial x_1\, \partial x_2\, \dots \partial x_n}$$ N.V/M.

* N.V/M. = Notación vectorial / matricial

La **función de densidad de probabilidad marginal de $\boldsymbol{x_r}$** se define como:

$$f(x_r) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n)\ dx_1\ \dots dx_{r-1} dx_{r+1}\ \dots dx_n$$

$$f(x_r) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(\vec{x})\ dx_1\ \dots dx_{r-1} dx_{r+1}\ \dots dx_n$$

El Operador esperanza matemática se define como:

$$E[H(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} H(x_1, x_2, \dots, x_n)\, f(x_1, x_2, \dots, x_n)\ dx_1 \dots dx_n$$

$$E\left[H\left(\vec{X}\right)\right] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} H(\vec{x})\, f(\vec{x})\ dx_1 \dots dx_n$$ N.V/M.

En particular:

$$E[X_r] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_r\, f(x_1, x_2, \dots, x_n)\, dx_1\ \dots dx_n$$

$$E[x_r] = \int_{-\infty}^{+\infty} x_r\, g(x_r)\, dx_r$$

Parámetros que caracterizan el comportamiento estadístico del vector $\vec{X}$ de v.a.

Vector de medias:

$$\vec{\mu}_{\vec{X}} = \left(\mu_{X_1}, \mu_{X_2}, \dots \mu_n\right)^T$$

$$\vec{\mu}_{\vec{X}} = \left(E[X_1], E[X_2], \dots, E[X_n]\right)^T$$

$$\vec{\mu}_{\vec{X}} = E\left[\vec{X}\right]$$ N.V/M.

Matriz de varianza-covarianza:

$$C_{\vec{X}} = E\left[\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)^T\right]$$

$$C_{\vec{X}} = \begin{bmatrix} E\left[\left(X_1 - \mu_{X_1}\right)^2\right] & \dots & E\left[\left(X_1 - \mu_{X_1}\right)(X_n - \mu_n)\right] \\ \dots & \dots & \dots \\ E\left[\left(X_n - \mu_{X_n}\right)(X_1 - \mu_1)\right] & \dots & E\left[\left(X_n - \mu_{X_n}\right)^2\right] \end{bmatrix}$$

Elementos de la matriz: en la diagonal están las varianzas y fuera de ella están las covarianzas:

$$C_{\vec{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \dots & \sigma_{X_1 X_n} \\ \dots & \dots & \dots \\ \sigma_{X_1 X_n} & \dots & \sigma_{X_n}^2 \end{bmatrix}$$

Caso importante de vector de variables aleatorias es cuando estas son independientes. Luego:

$$f(\vec{x}) = f(x_1, x_2, \dots, x_n) = g_1(x_1) g_2(x_2) \dots g_n(x_n)$$

Transformación de variables, generalización para más de dos variables

El punto de partida es un vector de v.a. con función densidad de probabilidad conjunta:

$$\vec{X} = (X_1, X_2, \dots, X_n)^T$$

$$\vec{X} \sim f(\vec{x})$$

Se genera un vector nuevo de v.a. a partir de relaciones determinitas del vector anterior:

$$\vec{Y} = \vec{Y}(\vec{X})$$

$$\vec{Y} = \left(Y_1(\vec{X}), Y_2(\vec{X}), \dots, Y_n(\vec{X})\right)^T$$

La función densidad de probabilidad para el nuevo vector será:

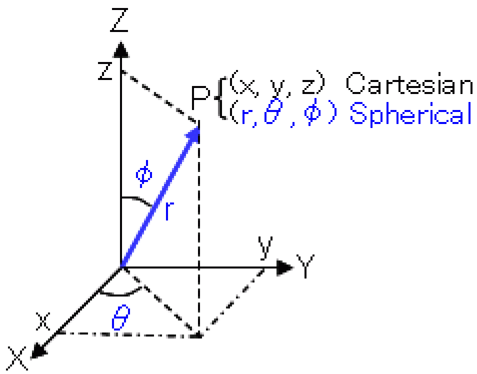
$$g(\vec{y}) = J\left(\frac{\vec{x}}{\vec{y}}\right) f(\vec{x}(\vec{y}))$$

$$J\left(\frac{\vec{x}}{\vec{y}}\right) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Jacobiano de la transformación (valor absoluto del determinante).

Ejemplo de transformación para vector de 3 v.a.

$$\vec{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad \begin{matrix} -\infty < X < +\infty \\ -\infty < Y < +\infty \\ -\infty < Z < +\infty \end{matrix} \qquad \vec{Y} = \begin{bmatrix} R \\ \Theta \\ \Phi \end{bmatrix} \qquad \begin{matrix} 0 < R < +\infty \\ 0 < \Theta < 2\pi \\ 0 < \Phi < \pi \end{matrix}$$



$f(\vec{X})$ Función de densidad de probabilidad conjunta.

$$X = R\,cos(\Theta)\,sin(\Phi)$$

$$\text{Y} = R\,sin(\Theta)\,sin(\Phi)$$

$$\text{Z} = R\,cos(\Phi)$$

Transformaciones de variables.

$$g(\vec{y}) = J\left(\frac{\vec{x}}{\vec{y}}\right) f(\vec{x}(\vec{y}))$$

$$J\left(\frac{\vec{x}}{\vec{y}}\right) = absdet \begin{bmatrix} cos(\theta)\,sin(\varphi) & sin(\theta)\,sin(\varphi) & sin(\varphi) \\ -r\,sin(\theta)\,sin(\varphi) & r\,cos(\theta)\,sin(\varphi) & 0 \\ r\,cos(\theta)\,cos(\varphi) & r\,sin(\theta)\,cos(\varphi) & r\,cos(\varphi) \end{bmatrix} = r^2\,sin(\varphi)$$

$$g(r,\theta,\varphi) = r^2\,sin(\varphi)\,f(x(r,\theta,\varphi), y(r,\theta,\varphi), z(r,\theta,\varphi))$$

Transformaciones lineales de variables – Propagación de errores

En la práctica, es muy común encontrar situaciones con transformaciones lineales de variables aleatorias. El planteo de interés es entender cómo se ‘transforman’/’propagan’ los parámetros estadísticos (media y matriz de var-cov) del primer conjunto de variables al segundo conjunto.

Como se verá más adelante, incluso el problema general (no-lineal) puede simplificarse desarrollando por series de Taylor y aplicando la solución para una transformación lineal.

El punto de partida es un vector de v.a. con su matriz de var-cov.:

$$\vec{X} = (X_1, X_2, \dots, X_n)^T \qquad C_{\vec{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \dots & \sigma_{X_1 X_n} \\ \dots & \dots & \dots \\ \sigma_{X_1 X_n} & \dots & \sigma_{X_n}^2 \end{bmatrix}$$

Por otro lado, se tiene r funciones lineales de $\vec{X}$**:**

$$Y_1 = a_1 + t_{1\,1}X_1 + t_{1\,2}X_2 \ldots + t_{1\,n}X_n$$

$$Y_2 = a_2 + t_{2\,1}X_1 + t_{2\,2}X_2 \ldots + t_{2\,n}X_n$$

$$\ldots$$

$$Y_r = a_r + t_{r\,1}X_1 + t_{r\,2}X_2 \ldots + t_{r\,n}X_n$$

En notación vectorial/matricial:

$$\vec{Y} = T\vec{X} + \vec{a}$$

El objetivo es encontrar una expresión para la matriz var-cov de $\vec{Y}$

Primero, buscamos la expresión para la media de $\vec{Y}$:

$$\vec{\mu}_{\vec{Y}} = E\left[\vec{Y}\right]$$

$$E\left[\vec{Y}\right] = E\left[T\vec{X} + \vec{a}\right] = T\,E\left[\vec{X}\right] + \vec{a} = T\vec{\mu}_{\vec{X}} + \vec{a}$$

Luego, por definición de matriz de var-cov:

$$C_{\vec{Y}} = E\left[\left(\vec{Y} - \vec{\mu}_{\vec{Y}}\right)\left(\vec{Y} - \vec{\mu}_{\vec{Y}}\right)^{T}\right]$$

Reemplazando las expresiones de $\vec{Y}$ y $\vec{\mu}_{\vec{Y}}$:

$$C_{\vec{Y}} = E\left[\left(T\vec{X} + \vec{a} - \left(T\vec{\mu}_{\vec{X}} + \vec{a}\right)\right)\left(T\vec{X} + \vec{a} - \left(T\vec{\mu}_{\vec{X}} + \vec{a}\right)\right)^{T}\right]$$

$$C_{\vec{Y}} = E\left[\left(T\vec{X} - T\vec{\mu}_{\vec{X}}\right)\left(T\vec{X} - T\vec{\mu}_{\vec{X}}\right)^{T}\right] = E\left[T\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)^{T} T^{T}\right]$$

$$C_{\vec{Y}} = T\, E\left[\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)\left(\vec{X} - \vec{\mu}_{\vec{X}}\right)^{T}\right] T^{T}$$

$\therefore \quad C_{\vec{Y}} = T\, C_{\vec{X}}\, T^{T}$ Expresión para la ley de propagación de errores.

Transformaciones no-lineales de variables – Propagación de errores

El punto de partida es un vector de v.a. con su matriz de var-cov.:

$$\vec{X} = (X_1, X_2, \ldots, X_n)^T \qquad C_{\vec{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \ldots & \sigma_{X_1 X_n} \\ \ldots & \ldots & \ldots \\ \sigma_{X_1 X_n} & \ldots & \sigma_{X_n}^2 \end{bmatrix}$$

Se tiene también, r funciones de $\vec{X}$:

$$Y_1 = Y_1(\vec{X})$$
$$Y_2 = Y_2(\vec{X})$$
$$\ldots$$
$$Y_r = Y_r(\vec{X})$$

$$\vec{Y} = \vec{Y}(\vec{X})$$ En notación vectorial / matricial.

Desarrollando por Taylor en torno a la media y quedándonos con los términos lineales:

$$y_1 \approx y_1(\vec{\mu}_{\vec{X}}) + \frac{\partial y_1}{\partial x_{1_{\vec{\mu}_{\vec{X}}}}}\left(x_1 - \mu_{x_1}\right) + \ \ldots + \frac{\partial y_1}{\partial x_{n_{\vec{\mu}_{\vec{X}}}}}\left(x_n - \mu_{x_n}\right)$$

$$\ldots$$

$$y_r \approx y_r(\vec{\mu}_{\vec{X}}) + \frac{\partial y_r}{\partial x_{1_{\vec{\mu}_{\vec{X}}}}}\left(x_1 - \mu_{x_1}\right) + \ \ldots + \frac{\partial y_r}{\partial x_{n_{\vec{\mu}_{\vec{X}}}}}\left(x_n - \mu_{x_n}\right)$$

$$\therefore \quad \vec{Y} \approx T\vec{X} + \vec{a} \qquad\qquad T = \begin{bmatrix} \frac{\partial y_1}{\partial x_{1_{\vec{\mu}_{\vec{X}}}}} & \ldots & \frac{\partial y_1}{\partial x_{n_{\vec{\mu}_{\vec{X}}}}} \\ \ldots & \ldots & \ldots \\ \frac{\partial y_r}{\partial x_{1_{\vec{\mu}_{\vec{X}}}}} & \ldots & \frac{\partial y_r}{\partial x_{n_{\vec{\mu}_{\vec{X}}}}} \end{bmatrix}$$

$$a_i = y_1(\vec{\mu}_{\vec{X}}) - \frac{\partial y_i}{\partial x_{1_{\vec{\mu}_{\vec{X}}}}}\mu_{X_n} - \cdots - \frac{\partial y_i}{\partial x_{n_{\vec{\mu}_{\vec{X}}}}}\mu_{X_n}$$

Luego:

$$C_{\vec{Y}} \approx T\, C_{\vec{X}}\, T^T$$

Expresión para la ley de propagación de errores para el caso no-lineal. Atención a la forma de T, que es diferente del caso lineal.

En algunos libros/publicaciones se puede encontrar la siguiente expresión para las varianzas:

$$\sigma_{Y_i}^2 \approx \sum_{j=1}^{n} \left(\frac{\partial y_i}{\partial x_j}_{\vec{\mu}_{\vec{X}}} \right)^2 \sigma_{X_j}^2$$

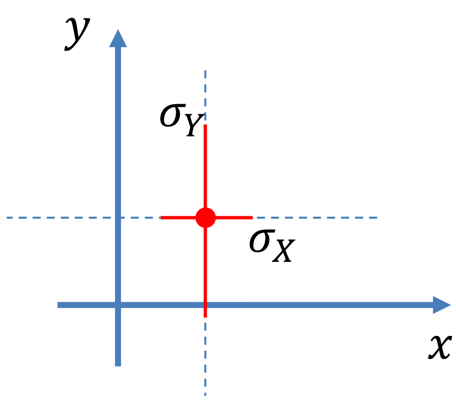
Si llamamos 'error estadístico', y le asignamos la letra Δ, a la desviación estándar de la variable, entonces:

$$\Delta_{Y_i} = \sqrt{\sigma_{Y_i}^2} \approx \sqrt{\sum_{j=1}^{n} \left(\frac{\partial y_i}{\partial x_j}_{\vec{\mu}_{\vec{X}}} \right)^2 {\Delta_{X_j}}^2}$$

Atención: la expresión anterior es válida únicamente para el caso en que las componentes de $\vec{X}$ sean independientes, es decir que $C_{\vec{X}}$ es diagonal para cualquier valor de $\vec{X}$**.**
Si no se cumple esa condición, hay que utilizar la expresión completa que tienen en cuenta las covarianzas.

Ejemplo de transformaciones no-lineales de variables – Propagación de errores

Se miden las coordenadas cartesianas de un punto en el plano (x, y) con un instrumento tal que la desviación estándar en x es de 1m y en y es 3 y estas son independientes entre sí. Nos interesa entender cómo se propagarán los errores cuando transformamos las coordenadas al sistema polar.



$$\vec{X} = \begin{bmatrix} X \\ Y \end{bmatrix} \qquad C_{\vec{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix} \qquad \begin{matrix} \sigma_X = 1 \\ \sigma_Y = 3 \end{matrix}$$

Transformaciones:

$$\vec{Y} = \begin{bmatrix} R \\ \Theta \end{bmatrix} \qquad \begin{matrix} r = \sqrt{x^2 + y^2} \\ \theta = arctan\left(\frac{y}{x}\right) \end{matrix} \qquad T = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix}$$

$$C_{\vec{Y}} = T\, C_{\vec{X}}\, T^T = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} C_{\vec{X}} \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix}^T$$

$$C_{\vec{Y}} = T\, C_{\vec{X}}\, T^{T} = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} C_{\vec{X}} \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix}^{T}$$

Para avanzar en la compresión, analicemos la situación como si estuviéramos midiendo un punto en torno al punto (1,1):

$$T_{(1,1)} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$C_{\vec{Y}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}^{T}$$

Matriz de var-cov en el sistema polar: $C_{\vec{Y}} = \begin{bmatrix} 5 & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & \frac{5}{2} \end{bmatrix}$

Ahora desandemos el camino y volvamos al sistema cartesiano:

$$C_{\vec{Y}} = \begin{bmatrix} 5 & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & \frac{5}{2} \end{bmatrix}$$

$$x = r\cos(\theta)$$
$$y = r\sin(\theta)$$

$$T = \begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix}$$

$$T_{(1,1)} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix}$$

Luego:

$$C_{\vec{X}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix} \begin{bmatrix} 5 & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & \frac{5}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}$$

Consistente, obtuvimos la matriz original.

Sin embargo, si hubiéramos usado

$$\sigma_{Y_i}^2 \approx \sum_{j=1}^{n} \left(\frac{\partial y_i}{\partial x_j}_{\vec{\mu}_{\vec{X}}} \right)^2 \sigma_{X_j}^2 \qquad C_{\vec{X}} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

Inconsistente, no obtuvimos la matriz original.

Absurdo: transformando y anti-transformando modificamos la matriz de var-cov de las mediciones. El error está en no considerar las covarianzas.