1 Normal bivariada general

Sea $Z_1, Z_2 \sim N(0,1)$ que usaremos para construir una distribución bivariada normal.

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(z_1^2 + z_2^2)\right],$$

Queremos transformar estas distribuciones normales para tener los siguientes parámetros arbitrarios: $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$.

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y \end{aligned}$$

Primero examinemos las distribuciones marginales de X e Y.

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ &= \sigma_X N(0,1) + \mu_X \\ &= N(\mu_X, \sigma_X^2) \\ Y &= \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y \\ &= \sigma_Y[\rho N(0,1) + \sqrt{1-\rho^2} N(0,1)] + \mu_Y \\ &= \sigma_Y[N(0,\rho^2) + N(0, 1-\rho^2)] + \mu_Y \\ &= \sigma_Y N(0,1) + \mu_Y \\ &= N(\mu_Y, \sigma_Y^2). \end{aligned}$$

Segundo podemos encontrar la covarianza y el coeficiente de correlación.

$$\begin{aligned} \text{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E})(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}[(\sigma_X Z_1 + \mu_X - \mu_X)(\sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y - \mu_Y)] \\ &= \mathbb{E}[(\sigma_X Z_1)(\sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2])] \\ &= \sigma_X \sigma_Y \mathbb{E}[\rho Z_1^2 + \sqrt{1-\rho^2} Z_1 Z_2] \\ &= \sigma_X \sigma_Y \rho \mathbb{E}(Z_1^2) \\ &= \sigma_X \sigma_Y \rho \end{aligned}$$

Por tanto :

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \rho.$$

En consecuencia, si queremos generar una variable aleatoria Normal Bivariada con $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ donde la correlación de X e Y es ρ podemos generar dos variables aleatorias normales independientes Z_1 y Z_2 y usar la transformación:

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y \end{aligned}$$

También podemos usar este resultado para encontrar la densidad conjunta de la Bivariada Normal utilizando un cambio de variables en dos dimensiones.

2 Cambio de variable multivariado

Sea $X_1, \ldots, X_n$ tiene una distribución conjunta continua con PDF f definido en S. Podemos definir n nuevas variables aleatorias $Y_1, \ldots, Y_n$ como sigue:

$$Y_1 = r_1(X_1, \ldots, X_n) \quad Y_n = r_n(X_1, \ldots, X_n)$$

Si asumimos que las n funciones $r_1, \ldots r_n$ definen una transformación diferenciable uno a uno de S a T, entonces sea inversa de esta transformación:

$$x_1 = s_1(y_1, \ldots, y_n) \quad x_n = s_n(y_1, \ldots, y_n)$$

Entonces e PDF conjunto de $Y_1, \ldots Y_n$ es,

$$g(y_1, \ldots, y_n) = \begin{cases} f(s_1, \ldots, s_n)|J| & \text{para } (y_1, \ldots, y_n) \in T \\ 0 & \text{en otros casos} \end{cases}$$

donde

$$J = \det \begin{bmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{bmatrix}$$

Lo primero que tenemos que encontrar son las inversas de la transformación. Si $x = r_1(z1, z_2)$ y $y = r_2(z_1, z_2)$ necesitamos encontrar las funciones h_1 y h_2 de manera que $Z_1 = s_1(X, Y)$ y $Z_2 = s_2(X, Y)$.

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ Z_1 &= \frac{X - \mu_X}{\sigma_X} \\ Y &= \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y \\ \frac{Y - \mu_Y}{\sigma_Y} &= \rho \frac{X - \mu_X}{\sigma_X} + \sqrt{1-\rho^2} Z_2] \\ Z_2 &= \frac{1}{\sqrt{1-\rho^2}} \left[\frac{Y - \mu_Y}{\sigma_Y} - \rho \frac{X - \mu_X}{\sigma_X} \right]. \end{aligned}$$

Por tanto,

$$s_1(x, y) = \frac{x - \mu_X}{\sigma_X} \quad s_2(x) = \frac{1}{\sqrt{1-\rho^2}} \left[\frac{y - \mu_Y}{\sigma_Y} - \rho \frac{x - \mu_X}{\sigma_X} \right]$$

Ahora calculamos el Jacobiano,

$$J = \det \begin{bmatrix} \dfrac{\partial s_1}{\partial x} & \dfrac{\partial s_1}{\partial y} \\ \dfrac{\partial s_2}{\partial x} & \dfrac{\partial s_2}{\partial y} \end{bmatrix} = \det \begin{bmatrix} \dfrac{1}{\sigma_X} & 0 \\ \dfrac{-\rho}{\sigma_X \sqrt{1-\rho^2}} & \dfrac{1}{\sigma_Y \sigma_X \sqrt{1-\rho^2}} \end{bmatrix} = \frac{1}{\sigma_X \sigma_Y \sqrt{1-\rho^2}}.$$

La densidad conjunta de X e Y es dado por,

$$\begin{aligned} f(x,y) &= f(z_1,z_2)|J| \\ \frac{1}{2\pi}\exp\left[-\frac{1}{2}(z_1^2+z_2^2)\right]|J| &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}(z_1^2+z_2^2)\right] \\ &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\frac{1}{1-\rho^2}\left(\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right)^2\right]\right] \\ &= \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-2\rho\frac{(x-\mu_X)}{\sigma_X}\frac{(y-\mu_Y)}{\sigma_Y}\right)\right]. \end{aligned}$$

3 Densidad de una distribución normal bivariada- Notación Matricial

Obviamente, la densidad para la Normal Bivariada es complicada y solo empeora cuando consideramos densidades conjuntas de distribuciones normales de dimensiones más altas. Podemos escribir la densidad en una forma más compacta usando la notación matricial,

$$\mathbf{x}=\begin{pmatrix}x\\y\end{pmatrix}\quad \boldsymbol{\mu}=\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}\quad \boldsymbol{\Sigma}=\begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}$$

$$f(\boldsymbol{x})=\frac{1}{2\pi}(\det\boldsymbol{\Sigma})^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

Podemos confirmar nuestros resultados verificando el valor de $(\det\boldsymbol{\Sigma})^{-1/2}$ y $(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ para el caso bivariado.

$$(\det\boldsymbol{\Sigma})^{-1/2}=(\sigma_X^2\sigma_Y^2-\rho^2\sigma_X^2\sigma_Y^2)^{-1/2}=\frac{1}{\sigma_X\sigma_Y(1-\rho^2)^{1/2}}$$

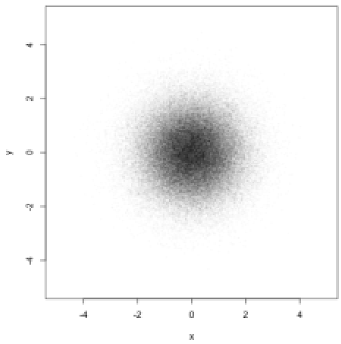
También,

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

$$\begin{aligned} &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{pmatrix}x-\mu_X\\y-\mu_Y\end{pmatrix}^T\begin{pmatrix}\sigma_Y^2 & -\rho\sigma_X\sigma_Y\\ -\rho\sigma_X\sigma_Y & \sigma_X^2\end{pmatrix}\begin{pmatrix}x-\mu_X\\y-\mu_Y\end{pmatrix} \\ &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{pmatrix}\sigma_Y^2(x-\mu_X)-\rho\sigma_X\sigma_Y(y-\mu_Y)\\ -\rho\sigma_X\sigma_Y(x-\mu_X)+\sigma_X^2(y-\mu_Y)\end{pmatrix}^T\begin{pmatrix}x-\mu_x\\y-\mu_y\end{pmatrix} \\ &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}(\sigma_Y^2(x-\mu_X)^2-2\rho\sigma_X\sigma_Y(x-\mu_X)(y-\mu_Y)+\sigma_X^2(y-\mu_Y)^2) \\ &= \frac{1}{1-\rho^2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right). \end{aligned}$$

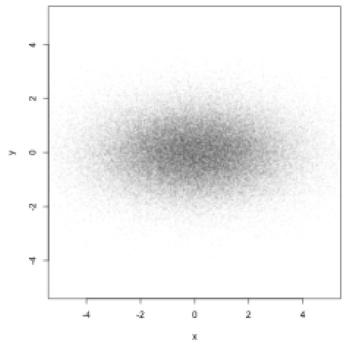
La distribución normal estándar bivariada tiene un máximo en el origen. Debes tener en cuenta que el único parámetro en la distribución normal estándar bivariada es la correlación ρ entre x e y. Si x e y son independientes ($\rho=0$), las superficies de la constante $f(x,y)$ son círculos concéntricos alrededor del origen. A medida que aumenta ρ, la distribución se estira diagonalmente, formando isolíneas elípticas con ejes mayores con pendiente positiva. Para ρ negativo, los ejes mayores tienen una pendiente negativa.

Estos gráficos fueron tomados desde las notas de Colin Rundel de la universidad de Duke.



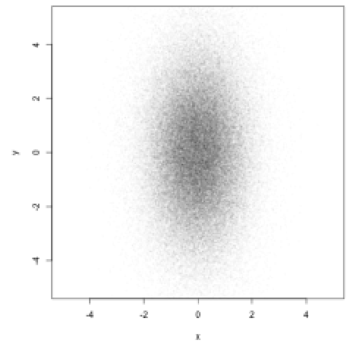
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0$



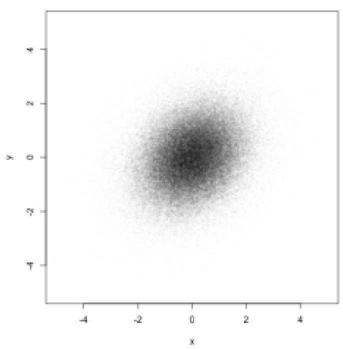
$X \sim \mathcal{N}(0,2),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0$



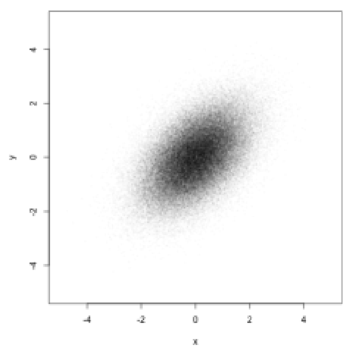
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,2)$

$\rho = 0$



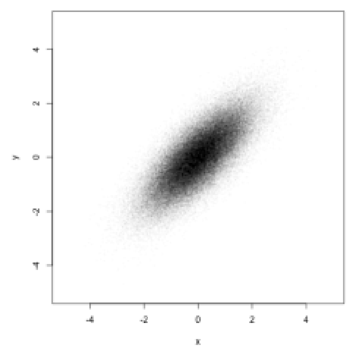
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.25$



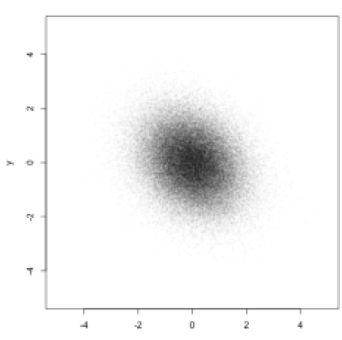
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.5$



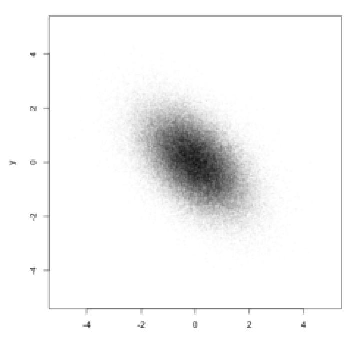
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.75$



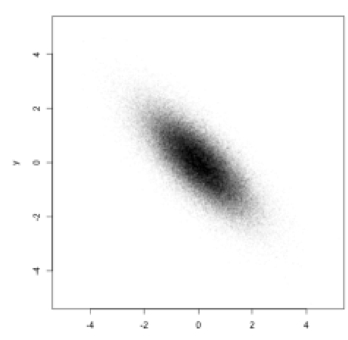
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.25$



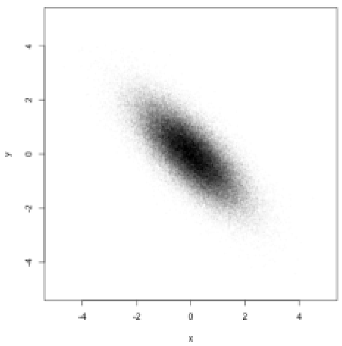
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.5$



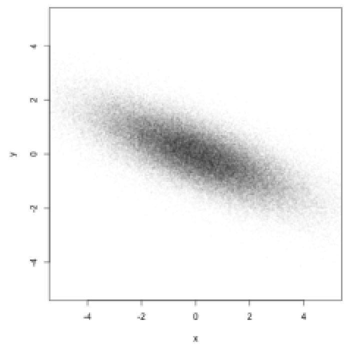
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



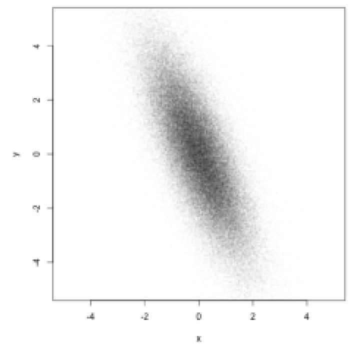
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



$X \sim \mathcal{N}(0,2),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,2)$

$\rho = -0.75$

3.1 Distribución de probabilidad condicional normal bivariada

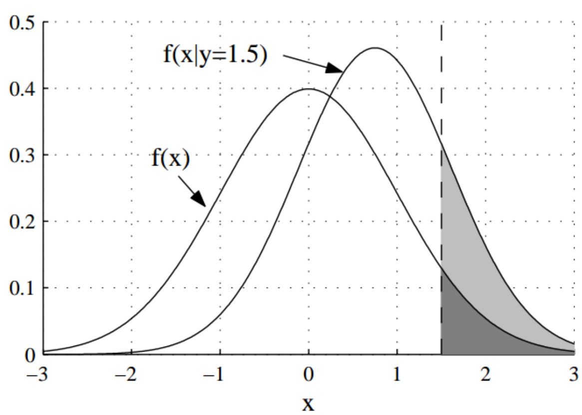
Una característica poderosa de la distribución normal bivariada es que la función de distribución de probabilidad condicional para una de las variables, dado un valor conocido para la otra variable, es distribuida normalmente, es decir, $f(x|y = y_0) \sim N(\mu_{X|y=y_0}, \sigma^2_{X|y=y_0})$. La media condicional y la varianza de x, dado $y = y_0$ es:

$$\mu_{X|y=y_0} = \mu_X + \rho\sigma_X \frac{(y - y_0)}{\sigma_y}$$
$$\sigma^2_{X|y=y_0} = \sigma_X \sqrt{1 - \rho^2}.$$

A partir de estos parámetros, podemos determinar la probabilidad de que x caiga en un rango dado, $x_0 \leq x \leq x_1$ de la distribución normal.

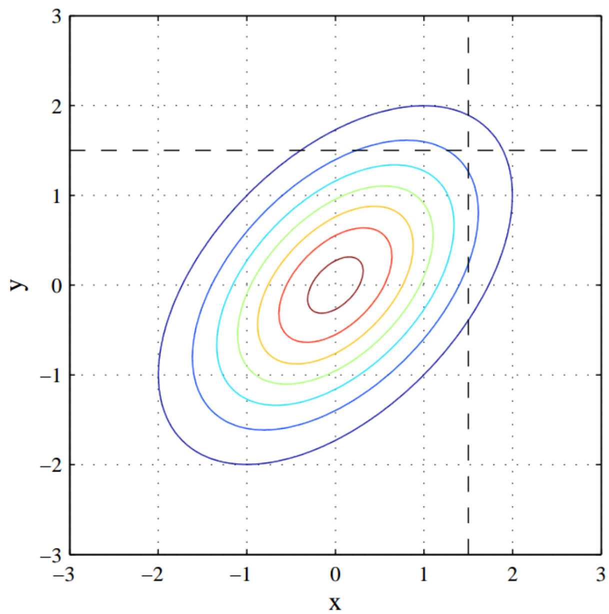
Este punto se ilustra en la siguiente figura, donde la distribución condicional para la variable x, dada la variable $y = 1.5$. El sombreado indica la probabilidad de que x exceda 1.5 la desviación estńdar si no se sabe nada sobre y (sombreado oscuro para la curva inferior), y si se sabe que $y = 1.54$ (sombreado claro para la curva superior).

Ten en cuenta la gran diferencia en las probabilidades, aunque la correlación no es excepcionalmente alta.



Hay un par de características adicionales a tener en cuenta. Primero, el eje mayor de la elipse definido por la distribución normal bivariada (ver la siguiente figura donde $\rho = 0.4$) puede obtenerse mediante regresión lineal de mínimos cuadrados. En este caso, se toma la regresión de mínimos cuadrados para minimizar la distancia perpendicular entre los datos y la línea de regresión. Esto se puede lograr con el análisis de la función ortogonal empírica (empirical orthogonal function (EOF)), en ese análisis de regresión simplemente gira el sistema de coordenadas.

El vector propio principal del análisis EOF apunta en la dirección del eje mayor de la elipse y la distancia desde el máximo de la distribución a una isolínea de la distribución.



4 Distribución normal multivariada

La notación matricial nos permite expresar fácilmente la densidad de la distribución normal multivariada para un número arbitrario de dimensiones. Expresamos la distribución normal multivariable k dimensional como sigue,

$$\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

donde $\boldsymbol{\mu}$ es el vector columna $k \times 1$ de medias y $\boldsymbol{\Sigma}$ es la matriz de covarianza $k \times k$ donde $\{\boldsymbol{\Sigma}\}_{i,j} = \text{Cov}(X_i, X_j)$.

La densidad de la distribución es

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}}(\det \boldsymbol{\Sigma})^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

En el caso bivariado, tuvimos una transformación de modo que pudimos generar dos valores aleatorias normales independientes y transformarlos en una muestra de una distribución normal bivariada arbitraria.

Existe un método similar para la distribución normal multivariable que aprovecha la descomposición de Cholesky de la matriz de covarianza. La descomposición de Cholesky se define para una matriz X definida positiva y simétrica como,

$$L = \text{Cholesky}(\boldsymbol{X})$$

donde $\boldsymbol{L}$ es una matriz triangular inferior tal que $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{X}$.