Suppose $Y$ is normally distributed with mean 0: $Y \sim \mathcal{N}(0, \sigma_y^2)$, and conditional on $Y = y$, $X$ is also, with mean that is proportional to $y$: $X|Y = y \sim \mathcal{N}(\alpha y, \beta^2 \sigma_x^2)$, where $\alpha$ and $\beta^2 \sigma_x^2$ can take any real and nonnegative real values, respectively. The variance has been written this way for later convenience; right now $\sigma_x^2$ has no interpretation other than as a factor of it.) What is the joint density function for $X$ and $Y$?

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$$

$$= \frac{1}{\sqrt{2\pi \beta^2 \sigma_x^2}} e^{-(x-\alpha y)^2/(2\beta^2 \sigma_x^2)} \cdot \frac{1}{\sqrt{2\pi \sigma_y^2}} e^{-y^2/(2\sigma_y^2)}$$

$$= \frac{1}{2\pi \beta \sigma_x \sigma_y} \exp\left(-\frac{1}{2\beta^2} \left(\frac{x^2}{\sigma_x^2} - \frac{2 \alpha xy}{\sigma_x^2} + \frac{\alpha^2 y^2}{\sigma_y^2} + \frac{\beta^2 y^2}{\sigma_y^2}\right)\right).$$

To make the $xy$ term symmetric, we introduce $\rho$ such that $\alpha = \rho \sigma_x / \sigma_y$, to get:

$$= \frac{1}{2\pi \beta \sigma_x \sigma_y} \exp\left(-\frac{1}{2\beta^2} \left(\frac{x^2}{\sigma_x^2} - \frac{2 \rho xy}{\sigma_x \sigma_y} + \frac{(\rho^2 + \beta^2)y^2}{\sigma_y^2}\right)\right),$$

and then set $\beta^2 = 1 - \rho^2$, to get:

$$= \frac{1}{2\pi \sqrt{1 - \rho^2} \sigma_x \sigma_y} \exp\left(-\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2 \rho xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2}\right)\right). \quad (1)$$

We have changed the names of the arbitrary parameters in the conditional density function for $X|Y = y$ so that it is $\mathcal{N}(\rho \sigma_x / \sigma_y, (1 - \rho^2) \sigma_x^2)$, where $|\rho| \leq 1$, but the mean and variance can still be chosen to be arbitrary real and nonnegative real numbers, respectively.
To understand the meaning of the parameters, notice that the exponent in (1) can be written as:

\[-\frac{1}{2(1-\rho^2)}(x \ y)^T \begin{pmatrix} 1/\sigma_x^2 & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & 1/\sigma_y^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.\]

(2)

If we define another $2 \times 2$ matrix:

\[
\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix},
\]

then

\[
\Sigma^{-1} = \frac{1}{\det \Sigma} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} \sigma_x^2 & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & 1/\sigma_y^2 \end{pmatrix},
\]

so

\[
f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x \ y)^T \Sigma^{-1} (x \ y)}.\]

(3)

By symmetry, $X \sim \mathcal{N}(0, \sigma_x^2)$ and $Y \mid X = x \sim \mathcal{N}(\rho \sigma_x^2 x, (1-\rho^2)\sigma_y^2)$. Since the means of both $X$ and $Y$ are 0,

\[
\text{Cov}[X, Y] = E[XY]
\]

\[
= \int \int xyf_{X,Y}(x, y) \, dx \, dy
\]

\[
= \int \int xyf_{X}(x|y)f_{Y}(y) \, dx \, dy
\]

\[
= \int yf_{Y}(y) \, dy \int xf_{X}(x|y) \, dx
\]

\[
= \int yf_{Y}(y) \, dy \cdot \frac{\sigma_x}{\sigma_y} y
\]

\[
= \frac{\sigma_x}{\sigma_y} \int y^2 f_{Y}(y) \, dy
\]

\[
= \frac{\sigma_x}{\sigma_y} \text{Var}[Y] = \rho \sigma_x \sigma_y.
\]

Thus $\Sigma$ is called the covariance matrix since it has the covariance of $X$ and $Y$ as its off-diagonal elements (and the variances of $X$ and $Y$ as its diagonal elements). Since

\[
\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \text{Var}[Y]}} = \rho,
\]

$\rho$ is the correlation of $X$ and $Y$.

The joint density function (3) generalizes to the situation in which the marginal densities are $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ by translating the variables so that

\[
f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu_x \ y-\mu_y)^T \Sigma^{-1} (x-\mu_x \ y-\mu_y)}.\]

(4)
This is the general \textit{bivariate normal distribution} $\mathcal{N}((\mu_x, \mu_y), \Sigma)$.

Finally, (4) can be generalized to the joint density function of $n$ random variables $X = (X_1, \ldots, X_n)$, each of which has marginal density $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$; it is

$$f_X(x) = \frac{1}{(2\pi \det \Sigma)^{n/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}.$$

This is the \textit{multivariate normal distribution} with $n \times n$ covariance matrix $\Sigma$; \textit{i.e.}, $\Sigma_{ij} = \text{Cov}[X_i, X_j]$. 