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, with $\beta^2 \leq 1$.) What is the joint density function for $X$ and $Y$?

\[
f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y) = \frac{1}{\sqrt{2\pi \beta^2 \sigma_x^2}} e^{-\frac{(x-\alpha y)^2}{2\beta^2 \sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi \sigma_y^2}} e^{-\frac{y^2}{2\sigma_y^2}}
= \frac{1}{2\pi \beta \sigma_x \sigma_y} \exp\left(-\frac{1}{2\beta^2} \left(\frac{x^2}{\sigma_x^2} - 2 \frac{\alpha xy}{\sigma_x^2} + \frac{\alpha^2 y^2}{\sigma_x^2} + \frac{\beta^2 y^2}{\sigma_y^2}\right)\right).
\]

To make the $xy$ term symmetric, we introduce $-1 \leq \rho \leq 1$ such that $\alpha = \rho \sigma_x / \sigma_y$, to get:

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

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

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

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.

© David A. Meyer 2017
To understand the meaning of the parameters, notice that the exponent in (1) can be written as:
\[-\frac{1}{2(1-\rho^2)}(x \ y) \left( \begin{array}{cc} \frac{1}{\sigma_x^2} & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & \frac{1}{\sigma_y^2} \end{array} \right) \left( \begin{array}{c} x \\ y \end{array} \right). \tag{2}\]

If we define another $2 \times 2$ matrix:
$$
\Sigma = \left( \begin{array}{cc} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{array} \right),
$$
then
$$
\Sigma^{-1} = \frac{1}{\det \Sigma} \left( \begin{array}{cc} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{array} \right) = \frac{1}{1-\rho^2} \left( \begin{array}{cc} \frac{1}{\sigma_x^2} & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & \frac{1}{\sigma_y^2} \end{array} \right),
$$
so
$$
f_{X,Y}(x,y) = \frac{1}{2\pi \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x y)\Sigma^{-1}(x y)}. \tag{3}\]

By symmetry, $X \sim \mathcal{N}(0, \sigma_x^2)$ and $Y|X = x \sim \mathcal{N}(\frac{\rho \sigma_y x}{\sigma_x}, (1-\rho^2)\sigma_y^2)$. Since the means of both $X$ and $Y$ are 0,
$$
\text{Cov}[X,Y] = \mathbb{E}[XY] = \iint x y f_{X,Y}(x,y) \, dx \, dy = \iint x y f_{X|Y}(x|y)f_Y(y) \, dx \, dy = \int y f_Y(y) \, dy \int x f_{X|Y}(x|y) \, dx = \int y f_Y(y) \, dy \int x f_{X|Y}(x|y) \, dx = \rho \frac{\sigma_x}{\sigma_y} \int y^2 f_Y(y) \, dy = \rho \frac{\sigma_x}{\sigma_y} \text{Var}[Y] = \rho \sigma_x \sigma_y.
$$

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)\Sigma^{-1}(x-\mu_x, y-\mu_y)}. \tag{4}\]

© David A. Meyer 2017
This is the general bivariate normal distribution \( N((\mu_x, \mu_y), \Sigma) \).

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 \( X_i \sim N(\mu_i, \sigma_i^2) \); 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 multivariate normal distribution with \( n \times n \) covariance matrix \( \Sigma \); i.e., \( \Sigma_{ij} = \text{Cov}[X_i, X_j] \).