Proof that the Difference of Two Correlated Normal Random Variables is Normal

Sahand Rabbani

Problem Statement

Given two normal random variables X and Y

$$\begin{aligned} X &\sim \mathcal{N}\left(\mu_X, \sigma_X^2\right) \\ Y &\sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right) \end{aligned}$$

that are correlated such that

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

where

$$\begin{aligned} \rho &\triangleq \operatorname{corr}(X, Y) \\ \sigma_{XY} &\triangleq \operatorname{cov}(X, Y) \end{aligned}$$

we endeavor to show that

$$Z \triangleq X - Y \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$$

To solve this problem, we appeal to the bivariate normal probability density function. The proof that follows will make significant use of variables and lemmas to condense notation.

Proof

To prove the above, we will first argue that given two normal random variables X_0 and Y_0

$$\begin{aligned} X_0 &\sim \mathcal{N}\left(0, \sigma_X^2\right) \\ Y_0 &\sim \mathcal{N}\left(0, \sigma_Y^2\right) \end{aligned}$$

such that $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ and

$$Z_0 \triangleq X_0 - Y_0 \sim \mathcal{N}\left(0, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$$

then it necessarily follows that

$$Z \triangleq X - Y \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$$

To show this, we take the former as an assumption and prove this consequence. It is clear that

$$\begin{aligned} X &= X_0 + \mu_X \\ Y &= Y_0 + \mu_Y \end{aligned}$$

It also follows that $\text{cov}(X,Y) = \text{cov}(X_0, Y_0)$ from the below:

$$\text{cov}(X,Y) = E\left\{(X-\mu_X)(Y-\mu_Y)\right\} = E\left\{X_0 Y_0\right\} = \text{cov}(X_0, Y_0)$$

We have that

$$E\left\{Z\right\} = E\left\{X-Y\right\} = E\left\{(X_0+\mu_X)-(Y_0+\mu_Y)\right\} = E\left\{X_0\right\} - E\left\{Y_0\right\} + \mu_X - \mu_Y = \mu_X - \mu_Y$$

Considering that a normal random variable plus a constant is itself a normal random variable, it is clear, then, that if $Z_0 \sim \mathcal{N}\left(0, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$, then necessarily $Z \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$. Now, we endeavor to show that $Z_0 \sim \mathcal{N}\left(0, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$. To do this, consider the bivariate PDF describing the joint probabilities of events X_0 and Y_0:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} - \frac{2\rho xy}{\sigma_X\sigma_Y}\right)\right)$$

It is clear that the PDF for Z_0 will obey

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, x-z)dx$$

We now endeavor to calculate this integral. Before we do so, we define

$$\alpha \triangleq \frac{x^2}{\sigma_X^2} + \frac{(x-z)^2}{\sigma_Y^2} - \frac{2\rho x(x-z)}{\sigma_X\sigma_Y}$$

and

$$A \triangleq \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

to simplify notation. The integral then becomes

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, x-z) = \int_{-\infty}^{\infty} A\exp\left(-\frac{1}{2(1-\rho^2)}\alpha\right)dx$$

From Lemma 1, we have that

$$\alpha = \beta' x^2 - \gamma' x + \delta'$$

where the Greek parameters, defined in the lemma, are functions of z and not functions of the integration variable x. We define

$$\xi = \xi'\left(\frac{1}{2(1-\rho^2)}\right) \qquad \xi \in \{\beta, \gamma, \delta\}$$

This reduces the integral to

$$f_Z(z) = A\int_{-\infty}^{\infty} \exp(-\beta x^2 + \gamma x - \delta)dx$$

We now employ some creative techniques to evaluate the integral:

$$\begin{aligned} f_Z(z) &= A\int_{-\infty}^{\infty} \exp(-\beta x^2 + \gamma x - \delta)dx \\ &= A\exp(-\delta)\int_{-\infty}^{\infty} \exp(-\beta x^2 + \gamma x)dx \\ &= A\exp(-\delta)\int_{-\infty}^{\infty} \exp\left(-\beta x\left(x - \frac{\gamma}{\beta}\right)\right)dx \end{aligned}$$

We note that we can shift the variable of integration by a constant without changing the value of the integral, since it is taken over the entire real line. With this mind, we make the substitution $x \to x + \frac{\gamma}{2\beta}$, which creates a difference of squares in the exponent and allows us to easily evaluate the integral:

$$\begin{aligned}
f_Z(z) &= A\exp(-\delta)\int_{-\infty}^{\infty} \exp\left(-\beta x\left(x - \frac{\gamma}{\beta}\right)\right)dx \\
&= A\exp(-\delta)\int_{-\infty}^{\infty} \exp\left(-\beta\left(x + \frac{\gamma}{2\beta}\right)\left(x - \frac{\gamma}{2\beta}\right)\right)dx \\
&= A\exp(-\delta)\int_{-\infty}^{\infty} \exp\left(-\beta\left(x^2 - \frac{\gamma^2}{4\beta^2}\right)\right)dx \\
&= A\exp\left(-\delta + \frac{\gamma^2}{4\beta}\right)\int_{-\infty}^{\infty} \exp(-\beta x^2)dx
\end{aligned}$$

From Lemma 2, we have that

$$\int_{-\infty}^{\infty} \exp(-\beta x^2)dx = \sqrt{\frac{\pi}{\beta}}$$

giving

$$f_Z(z) = A\sqrt{\frac{\pi}{\beta}}\exp\left(-\delta + \frac{\gamma^2}{4\beta}\right)$$

From Lemma 3, we have that

$$A\sqrt{\frac{\pi}{\beta}} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}}$$

and from Lemma 4, we have that

$$-\delta + \frac{\gamma^2}{4\beta} = -\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}$$

Plugging these two into our expression for $f_Z(z)$ gives

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}}\exp\left(-\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}\right)$$

This is clearly the PDF for a normal random variable with zero mean and variance $\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$. Thus, we see that

$$Z_0 \sim \mathcal{N}\left(0, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$$

and so it follows from the analysis above that

$$Z \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}\right)$$

□

Lemma 1

From the definition of α, we have

$$\begin{aligned}
\alpha &\triangleq \frac{x^2}{\sigma_X^2} + \frac{(x-z)^2}{\sigma_Y^2} - \frac{2\rho x(x-z)}{\sigma_X\sigma_Y} \\
&= \frac{x^2}{\sigma_X^2} + \frac{x^2}{\sigma_Y^2} + \frac{z^2}{\sigma_Y^2} - \frac{2xz}{\sigma_Y^2} - \frac{2\rho x^2}{\sigma_X\sigma_Y} + \frac{2\rho xz}{\sigma_X\sigma_Y} \\
&= x^2\left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}\right) - x\left(2z\left(\frac{1}{\sigma_Y^2} - \frac{\rho}{\sigma_X\sigma_Y}\right)\right) + \frac{z^2}{\sigma_Y^2} \\
&= \beta' x^2 - \gamma' x + \delta'
\end{aligned}$$

where

$$\begin{aligned}
\beta' &\triangleq \frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X \sigma_Y} \\
\gamma' &\triangleq 2z\left(\frac{1}{\sigma_Y^2} - \frac{\rho}{\sigma_X \sigma_Y}\right) \\
\delta' &\triangleq \frac{z^2}{\sigma_Y^2}
\end{aligned}$$

Lemma 2

It is a well known result that

$$\int_{-\infty}^{\infty} \exp(-\beta x^2)dx = \sqrt{\frac{\pi}{\beta}}$$

but we will confirm it using Fourier transforms. We know that the Fourier transform of the integrand is

$$F(f) = \mathcal{F}\left(\exp(-\beta x^2)\right)(f) = \sqrt{\frac{\pi}{\beta}} \exp\left(-\frac{(\pi f)^2}{\beta}\right)$$

We also know that

$$F(0) = \int_{-\infty}^{\infty} \exp(-\beta x^2)dx$$

Evaluating $F(f)$ at $f = 0$ gives

$$F(0) = \int_{-\infty}^{\infty} \exp(-\beta x^2)dx = \sqrt{\frac{\pi}{\beta}} \exp\left(-\frac{(\pi(0))^2}{\beta}\right) = \sqrt{\frac{\pi}{\beta}}$$

Lemma 3

Plugging in our definitions for A and β gives

$$\begin{aligned}
A\sqrt{\frac{\pi}{\beta}} &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \sqrt{\frac{2\pi(1-\rho^2)}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_X^2\sigma_Y^2\left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}\right)}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}}
\end{aligned}$$

Lemma 4

From the definitions of δ, γ, and β, we have

$$\begin{aligned}
-\delta + \frac{\gamma^2}{4\beta} &= \frac{1}{2(1-\rho^2)}\left(-\delta' + \frac{(\gamma')^2}{4\beta'}\right) \\
&= \frac{1}{2(1-\rho^2)}\left(-\frac{z^2}{\sigma_Y^2} + \frac{\left(2z\left(\frac{1}{\sigma_Y^2} - \frac{\rho}{\sigma_X\sigma_Y}\right)\right)^2}{4\left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}\right)}\right) \\
&= -\frac{z^2}{2(1-\rho^2)}\left(\frac{\frac{1}{\sigma_X^2\sigma_Y^2} + \frac{1}{\sigma_Y^4} - \frac{2\rho}{\sigma_X\sigma_Y^3} - \left(\frac{1}{\sigma_Y^2} - \frac{\rho}{\sigma_X\sigma_Y}\right)^2}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}}\right) \\
&= -\frac{z^2}{2(1-\rho^2)}\left(\frac{\sigma_X^2\sigma_Y^2 + \sigma_X^4 - 2\rho\sigma_X^3\sigma_Y - (\sigma_X^2 - \rho\sigma_X\sigma_Y)^2}{\sigma_X^2\sigma_Y^4 + \sigma_Y^2\sigma_X^4 - 2\rho\sigma_X^3\sigma_Y^3}\right) \\
&= -\frac{z^2}{2(1-\rho^2)}\left(\frac{\sigma_X^2\sigma_Y^2 + \sigma_X^4 - 2\rho\sigma_X^3\sigma_Y - \sigma_X^4 + 2\rho\sigma_X^3\sigma_Y - \rho^2\sigma_X^2\sigma_Y^2}{\sigma_X^2\sigma_Y^4 + \sigma_Y^2\sigma_X^4 - 2\rho\sigma_X^3\sigma_Y^3}\right) \\
&= -\frac{z^2}{2(1-\rho^2)}\left(\frac{\sigma_X^2\sigma_Y^2 - \rho^2\sigma_X^2\sigma_Y^2}{\sigma_X^2\sigma_Y^2(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)}\right) \\
&= -\frac{z^2}{2(1-\rho^2)}\frac{\sigma_X^2\sigma_Y^2(1-\rho^2)}{\sigma_X^2\sigma_Y^2(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)} \\
&= -\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}
\end{aligned}$$