This article intends to prove the well known facts about bivariate normal distribution. I mainly refered to the book "DeGroot, Probability and Statistics, 4ed" and `https://statproofbook.github.io`.

1 main definition and theorems

Theorem 3.9.5 : bivariate transformation of PDF

Let U and V be random variables with joint PDF $f_{UV}(u,v)$. Suppose there is a set $S \subset \mathbb{R}^2$ such that $P((u,v) \in S) = 1$ and a differentiable injective function r of S into $\mathbb{R}^2$ Denoting $T = r(S)$, r is a one to one correspondence between S and T. Let s be the inverse $s = r^{-1}$ of r and denote

$$\begin{cases} x = r_1(u,v) \\ y = r_2(u,v) \end{cases}, \qquad \begin{cases} u = s_1(x,y) \\ v = s_2(x,y) \end{cases}$$

Then, the joint PDF $f_{XY}(x,y)$ of two random variables $X = r_1(U,V)$ and $Y = r_2(U,V)$ is given by

$$f_{XY}(x,y) = \begin{cases} f_{UV}(s_1(x,y), s_2(x,y)) \times |J| & (x,y) \in T \\ 0 & \text{(otherwise)} \end{cases}$$

where

$$J = \det \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}$$

There is no proof in the book and we regard it as true.

Example 3.9.9

Let U and V be random variables such that

$$f_{UV}(u,v) = \begin{cases} 4uv & 0 < u,v < 1 \\ 0. & \text{(otherwise)} \end{cases}$$

For random variables X and Y defined by $X = \frac{U}{V}$ and $Y = UV$, find the PDF $f_{XY}(x,y)$.

Let $S = (0,1)^2 = \{(u,v) \in \mathbb{R}^2 : 0 < u,v < 1\}$. Then,

$$\begin{aligned} P((u,v) \in S) &= \iint_S f_{UV}(u,v)\, du\, dv \\ &= \int_0^1 \int_0^1 4uv\, du\, dv \\ &= 1. \end{aligned}$$

Let r_1 and r_2 be functions on S defined by

$$\begin{aligned} x &= r_1(u,v) = \frac{u}{v} \\ y &= r_2(u,v) = uv \end{aligned} \tag{$*$}$$

The function r on S defined by $r(u,v) = (r_1(u,v), r_2(u,v))$ is differentiable on S. It is also injective ; if $\left(\frac{u_1}{v_1}, u_1v_1\right) = \left(\frac{u_2}{v_2}, u_2v_2\right)$, then

$$\begin{aligned} {u_1}^2 &= \left(\frac{u_1}{v_1}\right) \times (u_1v_1) = \left(\frac{u_2}{v_2}\right) \times (u_2v_2) = {u_2}^2 \\ {v_1}^2 &= \left(\frac{u_1}{v_1}\right) \div (u_1v_1) = \left(\frac{u_2}{v_2}\right) \div (u_2v_2) = {v_2}^2 \end{aligned}$$

that is, $(u_1,v_1) = (u_2,v_2)$ if $r(u_1,v_1) = r(u_2,v_2)$.

Let $T = r(S)$. Then $r : S \to T$ is bijective. To find its inverse s, we can make use of $(*)$ to get

$$\begin{aligned} u &= s_1(x,y) = \sqrt{xy} \\ v &= s_2(x,y) = \sqrt{\frac{y}{x}}. \end{aligned}$$

Now we find T.

$$
\begin{aligned}
(x,y)\in T &\iff (s_1(x,y),x_2(x,y))\in S\\
&\iff (u,v)\in S\\
&\iff 0<u<1 \quad \& \quad 0<v<1\\
&\iff 0<\sqrt{xy}<1 \quad \& \quad 0<\sqrt{\frac{y}{x}}<1\\
&\iff x>0 \quad \& \quad y>0 \quad \& \quad y<\frac{1}{x} \quad \& \quad y<x\\
&\iff 0<y<\min\left(x,\frac{1}{x}\right),\\
T&=\left\{(x,y)\in\mathbb{R}^2 : 0<y<\min\left(x,\frac{1}{x}\right)\right\}
\end{aligned}
$$

Finally, evaluate $|J|$

$$
\begin{aligned}
|J| &= \det\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{bmatrix}\\
&= \begin{vmatrix} \frac{1}{2}\sqrt{\frac{y}{x}} & \frac{1}{2}\sqrt{\frac{x}{y}} \\ -\frac{1}{2}\sqrt{\frac{y}{x^3}} & \frac{1}{2}\sqrt{\frac{1}{xy}}\end{vmatrix}\\
&= \frac{1}{2x}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f_{XY}(x,y) &= \begin{cases} f_{UV}\left(\sqrt{xy},\sqrt{\frac{y}{x}}\right) & (x,y)\in T\\ 0 & \text{(otherwise)}\end{cases}\\
&= \begin{cases} 4y & 0<y<\min\left(x,\frac{1}{x}\right)\\ 0 & \text{(otherwise)}.\end{cases}
\end{aligned}
$$

Theorem 5.10.1

Suppose that $U,V\sim N(0,1)$ are independent. Let $\mu_X,\mu_Y\in\mathbb{R}$, $\sigma_X,\sigma_Y>0$ and $-1<\rho<1$. Let X and Y be random variables defined by

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

Then, the joint PDF of X and Y is given by

$$
f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}
$$

Since U and V are standard normal and independent, the joint PDF of U and V is

$$
\begin{aligned}
f_{UV}(u,v) &= f_U(u)f_V(v)\\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)\times\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}v^2\right)\\
&= \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(u^2+v^2)\right\}
\end{aligned}
$$

And, $P\left((u,v)\in\mathbb{R}^2\right)=1$ since

$$
\begin{aligned}
P\left((u,v)\in\mathbb{R}^2\right) &= \iint_{\mathbb{R}^2} f_{UV}(u,v)\,du\,dv \\
&= \int_{\mathbb{R}} f_U(u)\,du \times \int_{\mathbb{R}} f_V(v)\,dv \\
&= 1
\end{aligned}
$$

A differentiable function $r:\mathbb{R}^2\to\mathbb{R}^2$ such that $r(u,v)=(r_1(u,v),r_2(u,v))$ is given by

$$
\begin{aligned}
x &= r_1(u,v) = \sigma_X u + \mu_X \\
y &= r_2(u,v) = \sigma_Y\left[\rho u + \sqrt{1-\rho^2}v\right] + \mu_Y
\end{aligned}
$$

It has s as its inverse where $s(x,y)=(s_1(x,y),s_2(x,y))$ and r is invertible ;

$$
\begin{aligned}
u &= s_1(x,y) = \frac{x-\mu_X}{\sigma_X} \\
v &= s_2(x,y) = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)
\end{aligned}
$$

The jacobian $J=\frac{\partial(u,v)}{\partial(x,y)}$ is given by

$$
\begin{aligned}
J &= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} \\
&= \begin{vmatrix} \frac{1}{\sigma_X} & 0 \\ \frac{\rho}{\sigma_X\sqrt{1-\rho^2}} & \frac{1}{\sigma_Y\sqrt{1-\rho^2}} \end{vmatrix} \\
&= \frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}
\end{aligned}
$$

By theorem 3.9.5, the PDF of X and Y is given by

$$
\begin{aligned}
f_{XY}(x,y) =& f_{UV}\left(s_1(x,y),s_2(x,y)\right)\times|J| \\
=& \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \frac{1}{1-\rho^2}\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)^2\right]\right\}\times\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\
=& \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}
\end{aligned}
$$

Theorem 5.10.2(Main Theorem)

Suppose that X, Y are random variables where the PDF is given by

$$
f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}.
$$

Then,

(a) the random variables U and V given by

$$
\begin{aligned}
U &= \frac{X-\mu_X}{\sigma_X} \\
V &= \frac{1}{\sqrt{1-\rho^2}}\left(\frac{Y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)
\end{aligned}
$$

are independent and standard normal.

(b) $X\sim N(\mu_X,{\sigma_X}^2)$, $Y\sim N(\mu_Y,{\sigma_Y}^2)$, $\rho_{XY}=\rho$.

We prove (a) only. The proof of (b) depends on the notion of moment generating function and is discussed in section 2. To prove (a), we first evaluate the function $f_{UV}(u,v)$, where we apply 3.9.5 in the reverse order. Since

$$\begin{aligned} x &= \sigma_X u + \mu_X \\ y &= \sigma_Y \left(\rho u + \sqrt{1-\rho^2}v\right) + \mu_Y, \end{aligned}$$

the jacobian $J = \frac{\partial(x,y)}{\partial(u,v)}$ is given by

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \\ &= \begin{vmatrix} \sigma_X & 0 \\ \rho\sigma_Y & \sigma_Y\sqrt{1-\rho^2} \end{vmatrix} \\ &= \sigma_X\sigma_Y\sqrt{1-\rho^2}. \end{aligned}$$

Now, by theorem 3.9.5,

$$\begin{aligned} f_{UV}(u,v) =& f_{XY}\left(\sigma_X u + \mu_X, \sigma_Y\left(\rho u + \sqrt{1-\rho^2}v\right) + \mu_Y\right) \times |J| \\ =& \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right.\right. \\ &\left.\left. + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \times \sigma_X\sigma_Y\sqrt{1-\rho^2} \\ =& \frac{1}{2\pi}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2 - 2\rho u(\rho u + \sqrt{1-\rho^2}v) + (\rho u + \sqrt{1-\rho^2}v)^2\right]\right\} \\ =& \frac{1}{2\pi}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2 - \rho u^2 + (1-\rho^2)v^2\right]\right\} \\ =& \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(u^2+v^2)\right\} \end{aligned}$$

Integrating with respect to v yields the marginal distribution $f_U(u)$.

$$\begin{aligned} f_V(v) &= \int_{\mathbb{R}} f_{UV}(u,v)\,dv \\ &= \frac{e^{-\frac{1}{2}u^2}}{2\pi}\int_{\mathbb{R}} e^{-\frac{1}{2}v^2}\,dv \\ &\overset{\star}{=} \frac{e^{-\frac{1}{2}u^2}}{2\pi} \times \sqrt{2\pi} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} \end{aligned}$$

where $\star$ can be justified by

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-\frac{1}{2}v^2}\,dv\right)^2 &= \iint_{\mathbb{R}^2} e^{-\frac{1}{2}({v_1}^2+{v_2}^2)}\,dv_1\,dv_2 \\ &= \int_0^{2\pi}\int_0^\infty e^{-\frac{1}{2}r^2}\,dr\,d\theta \\ &= 2\pi\int_0^\infty e^{-R}\,dR \\ &= 2\pi. \end{aligned}$$

Similarly,

$$f_V(v) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}v^2}.$$

Thus, U and V are standard normal. Furthermore, since $f_{UV}(u,v) = f_U(u)f_V(v)$, U and V are independent.

Definition 5.10.1 : bivariate normal distributions

If random variables X and Y have

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right] \right\}.$$

as PDF, X and Y is said to have the bivariate normal distributions with means μ_X and μ_Y, variances ${\sigma_X}^2$ and ${\sigma_Y}^2$, and correlation coefficient ρ.

2 remaining proof using moment generating functions

4.4.2 Definition and 4.4.2 Theorem : moment genrating functions

Let X be a random variable and let

$$M_X(t) = E[e^{tX}]$$

for each $t \in \mathbb{R}$, where the value is well defined. The function M_X is called the *moment generating function* of X. Then,

$$M_X{}^{(n)}(0) = E[X^n]$$

where the superscript (n) is n-times derivative and n is a positive integer.

The proof introduced here depends on infinite sum of random variable, where the convergence is just assumed to be guaranteed. By the Maclaurin series of the exponential function and the linearity of E,

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] \\ &= E\left[1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\right] \\ &= 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \cdots \end{aligned}$$

Then,

$$\begin{aligned} M_X'(t) &= E[X] + tE[X^2] + \frac{t^2}{2!}E[X^3] + \frac{t^3}{3!}E[X^4] + \cdots \\ M_X''(t) &= E[X^2] + tE[X^3] + \frac{t^2}{2!}E[X^4] + \frac{t^3}{3!}E[X^5] + \cdots \\ &\vdots \\ M_X^{(n)}(t) &= E[X^n] + tE[X^{n+1}] + \frac{t^2}{2!}E[X^{n+2}] + \frac{t^3}{3!}E[X^{n+3}] + \cdots . \end{aligned}$$

It follows that

$$M_X^{(n)}(0) = E[X^n].$$

Example 4.4.3 If we use the above theorem to the exponential distribution,

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x}, \qquad x > 0 \\ M_X(t) &= E\left[e^{tX}\right] \\ &= \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{(t-\lambda)x}\, dx \\ &= \lambda \left[\frac{1}{\lambda - t} e^{(t-\lambda)x}\right]_0^\infty \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

provided that $t < \lambda$. Thus,

$$\begin{aligned} E[X] = M_X'(0) &= \left.\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = \frac{1}{\lambda} \\ E[X^2] = M_X''(0) &= \left.\frac{2\lambda}{(\lambda - t)^3}\right|_{t=0} = \frac{2}{\lambda^2} \\ V[X] &= \frac{1}{\lambda^2} \end{aligned}$$

4.2.6 Theorem and 4.4.4 Theorem : moment genrating functions

Let X and Y be independent random variables. Then,

(a) $E[XY] = E[X]E[Y]$

(b) $M_{X+Y}(t) = M_X(t) \times M_Y(t)$

The proof of (b) below depends on the fact that functions of independent variables are also independent ;

$$\begin{aligned} (a)\ E[XY] &= \iint_{\mathbb{R}^2} xy f_{XY}(x,y)\,dx\,dy \\ &= \iint_{\mathbb{R}^2} xy f_X(x) f_Y(y)\,dx\,dy \\ &= \int_{\mathbb{R}} x f_X(x) dx \times \int_{\mathbb{R}} y f_Y(y) dy \\ &= E[X]E[Y] \\ (b)\ M_{X+Y}(t) &= E\left[e^{t(X+Y)}\right] \\ &= E\left[e^{tX} \times e^{tY}\right] \\ &= E\left[e^{tX}\right] \times \left[e^{tY}\right] \\ &= M_X(t) \times M_Y(t) \end{aligned}$$

5.6.2 Theorem / 5.6.7 Theorem / 5.6.1. Corollary

(a) $X \sim N(\mu, \sigma^2)$ if and only if $M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$

(b) If $X \sim N(\mu_X, {\sigma_X}^2)$ and $Y \sim N(\mu_Y, {\sigma_Y}^2)$ are independent, then

$$aX + bY + c \sim N\left(a\mu_X + b\mu_Y + c, a^2{\sigma_X}^2 + b^2{\sigma_Y}^2\right).$$

(a) Suppose the former. Then

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \int_{\mathbb{R}} e^{tx} \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left(tx - \left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2\right) dx. \end{aligned}$$

Let $u = \frac{x-\mu}{\sigma}$ so that $dx = \sqrt{2}\sigma\,du$. Thus

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left(t(\mu + \sqrt{2}\sigma u) - u^2\right) \times \sqrt{2}\sigma\,du. \\ &= \frac{e^{\mu t}}{\sqrt{\pi}} \int_{\mathbb{R}} \exp\left(\sqrt{2}\sigma t u - u^2\right) du. \\ &= \frac{e^{\mu t}}{\sqrt{\pi}} \int_{\mathbb{R}} \exp\left(-\left(u - \frac{\sigma t}{\sqrt{2}}\right)^2 + \frac{\sigma^2 t^2}{2}\right) du. \end{aligned}$$

Substitute again, $u - \frac{\sigma t}{\sqrt{2}}$ by v, it becomes

$$\begin{aligned} M_X(t) &= \frac{e^{\mu t + \frac{1}{2}\sigma^2 t^2}}{\sqrt{\pi}} \int_{\mathbb{R}} \exp\left(-v^2\right) du \\ &\overset{*}{=} \frac{e^{\mu t + \frac{1}{2}\sigma^2 t^2}}{\sqrt{\pi}} \times \sqrt{\pi} \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

as desired. $(*)$ is because of the usual double integral and coordinate change.

The converse part of (a) can be justified by 4.4.5 Theorem, which state that the moment generating function of a distribution is unique. That is, if Y satisfies $M_Y(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$, then $M_X = M_Y$ and $X = Y$. It follows that $Y \sim N(\mu, \sigma^2)$. That uniqueness theorem is said to be proved using graduate level tools such as probability measures.

(b) By the properties of moment generating function and expectation of independent variables and the only if part of (a),

$$\begin{aligned} M_{aX+bY+c}(t) &= E\left[e^{t(aX+bY+c)}\right] \\ &= E\left[e^{atX} \times e^{btY} \times e^{ct}\right] \\ &= E\left[e^{atX}\right] \times E\left[e^{btY}\right] \times E\left[e^{ct}\right] \\ &= M_X(at) \times M_Y(bt) \times e^{ct} \\ &= \exp\left(\mu_X(at) + \frac{1}{2}{\sigma_X}^2(at)^2\right) \times \exp\left(\mu_Y(bt) + \frac{1}{2}{\sigma_Y}^2(bt)^2\right) \times \exp(ct) \\ &= \exp\left((a\mu_X + b\mu_Y + c)t + \frac{1}{2}\left(a^2{\sigma_X}^2 + b^2{\sigma_Y}^2\right)t^2\right) \end{aligned}$$

By the if part of (a), $aX + bY + c$ has normal distribution of mean $a\mu_X + b\mu_Y + c$ and variance $a^2{\mu_X}^2 + b^2{\mu_Y}^2$.

Proof of the main theorem (5.10.2) (b)

Proof. Since $X = \sigma_X U + \mu_X$ and $U \sim N(0,1)$, X follows a normal distribution with mean

$$E[X] = \sigma_X \times 0 + \mu_X = \mu_X$$

and variance

$$V[X] = {\sigma_X}^2 \times 1^2 = {\sigma_X}^2.$$

And, since $Y = \sigma_Y\left(\rho U + \sqrt{1-\rho^2}V\right) + \mu_Y$, Y follows a normal distribution with mean

$$E[Y] = \sigma_Y \rho \times 0 + \sigma_Y\sqrt{1-\rho^2} \times 0 + \mu_Y = \mu_Y$$

and variance

$$V[Y] = {\sigma_Y}^2\rho^2 \times 1^2 + {\sigma_Y}^2(1-\rho^2) \times 1^2 = {\sigma_Y}^2.$$

To prove $\rho_{XY} = \rho$, let's prove the bilinearity of Cov. Since

$$\begin{aligned} \mathrm{Cov}(aX + bY, Z) &= E\left[((aX + bY) - (a\mu_X + b\mu_Y))(z - \mu_Z)\right] \\ &= E\left[(a(X - \mu_X) + b(Y - \mu_Y))(z - \mu_Z)\right] \\ &= E\left[a(X - \mu_X)(z - \mu_Z) + b(Y - \mu_Y)(z - \mu_Z)\right] \\ &= aE\left[(X - \mu_X)(z - \mu_Z)\right] + bE\left[(Y - \mu_Y)(z - \mu_Z)\right] \\ &= a\mathrm{Cov}(X, Z) + b\mathrm{Cov}(Y, Z), \end{aligned}$$

Cov is linear with respect to the first coordinate. Similarly, it is linear with respect to the second coordinate, and Cov is binlinear. Moreover, $\mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y)$ if c is a constant, since

$$\begin{aligned} \mathrm{Cov}(X + c, Y) &= E\left[((X + c) - (\mu_X + c))(Y - \mu_Y)\right] \\ &= E\left[(X - \mu_X)(Y - \mu_Y)\right] \\ &= \mathrm{Cov}(X, Y). \end{aligned}$$

Similarly, $\mathrm{Cov}(X, Y + c) = \mathrm{Cov}(X, Y)$.

Therefore,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\left(\sigma_X U + \mu_X, \sigma_Y \rho U + \sigma_Y \sqrt{1-\rho^2} V + \mu_Y\right) \\
&= \mathrm{Cov}\left(\sigma_X U, \sigma_Y \rho U + \sigma_Y \sqrt{1-\rho^2} V\right) \\
&= \mathrm{Cov}\left(\sigma_X U, \sigma_Y \rho U\right) + \mathrm{Cov}\left(\sigma_X U, \sigma_Y \sqrt{1-\rho^2} V\right) \\
&= \sigma_X \sigma_Y \rho \mathrm{Cov}\left(U, U\right) + \sigma_X \sigma_Y \sqrt{1-\rho^2} \mathrm{Cov}\left(U, V\right) \\
&= \sigma_X \sigma_Y \rho \times 1 + \sigma_X \sigma_Y \sqrt{1-\rho^2} \times 0 \\
&= \sigma_X \sigma_Y \rho
\end{aligned}$$

and $\rho_{XY} = \rho$. $\square$