Box–Muller Method and Change of Variables

Reconstructed from handwritten notes

Table of contents

Functions of Several Variables

Suppose (x, y) are jointly distributed and mapped onto (u, v), jointly distributed.

$u = g_1(x, y)$

$v = g_2(x, y)$

and that the transformation can be inverted:

$x = h_1(u, v)$

$y = h_2(u, v)$

Assume h_1 and h_2 have continuous partial derivatives and that the Jacobian

$$J = \begin{vmatrix} \frac{\partial}{\partial u} h_1 & \frac{\partial}{\partial v} h_1 \\ \frac{\partial}{\partial u} h_2 & \frac{\partial}{\partial v} h_2 \end{vmatrix} \neq 0$$

Then

$$f_{UV}(u, v) = f_{XY}(h_1(u, v),\, h_2(u, v))\, |J|.$$

Example: Polar Coordinates from Independent Normals

Let X and Y be independent $\mathcal{N}(0, 1)$ random variables. Find the joint density of the polar coordinates R and Θ. Recall polar coordinates:

$r = \sqrt{x^2 + y^2}$

$\theta = \tan^{-1}\left(\frac{y}{x}\right)$

and the inverse transformation

$x = r\cos\theta$

$y = r\sin\theta$

So

$$f_{R\Theta}(r, \theta) = f_{XY}(x, y)\, |J|.$$

Find J:

$$J = \det\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \det\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

$= \; r\cos^2\theta + r\sin^2\theta$

$= \; r\,(\,\cos^2\theta + \sin^2\theta)$

$= \; r.$

Therefore

$f_{R\Theta}(r,\theta) = r\, f_{XY}(r\cos\theta,\, r\sin\theta).$

Because X and Y are independent standard normals,

$f_X(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad f_Y(y) = \dfrac{1}{\sqrt{2\pi}}\, e^{-y^2/2}.$

Hence

$f_{R\Theta}(r,\theta) = r\, f_X(r\cos\theta)\, f_Y(r\sin\theta)$

$= r\;\dfrac{1}{\sqrt{2\pi}} e^{-r^2\cos^2\theta/2}\;\dfrac{1}{\sqrt{2\pi}} e^{-r^2\sin^2\theta/2}$

$= r\;\dfrac{1}{2\pi}\; e^{-\frac{1}{2}r^2\,(\cos^2\theta+\sin^2\theta)}$

$= r\;\dfrac{1}{2\pi}\; e^{-r^2/2}.$

Thus R and Θ are independent with

$f_R(r) = r\, e^{-r^2/2},\; r \geq 0 \quad \text{(Rayleigh)}$

$f_\Theta(\theta) = \dfrac{1}{2\pi},\; \theta \in (0, 2\pi) \quad \text{(Uniform)}.$

Visualizing the Polar Transformation in R

We can illustrate the mapping of a random Euclidean point (X, Y) to polar coordinates (R, Θ), with Θ measured **counter-clockwise** from the positive X-axis.

```
set.seed(123)

# Generate a single random point from N(0,1) x N(0,1)
x <- rnorm(1)
y <- rnorm(1)

# Polar conversion
r <- sqrt(x^2 + y^2)
theta <- atan2(y, x)                # in (-pi, pi]
theta_ccw <- ifelse(theta < 0, theta + 2*pi, theta)  # map to [0, 2*pi)
```

```r
# Plot base
plot(0, 0, type = "n", xlim = c(-3,3), ylim = c(-3,3),
     xlab = "X", ylab = "Y", main = "Euclidean to Polar Coordinates (CCW angle)")
abline(h = 0, v = 0, lty = 3)  # dashed axes for reference

# Point and radius vector
points(x, y, col = "red", pch = 19, cex = 1.5)
arrows(0, 0, x, y, col = "blue", lwd = 2)

# Dashed circle at radius R
symbols(0, 0, circles = r, inches = FALSE, add = TRUE, lty = 2)

# Dashed line through (0,0) and (x,y) using angle (avoids slope issues when x ~ 0)
L <- 3
lines(c(-L, L) * cos(theta), c(-L, L) * sin(theta), lty = 3)

# Annotate R
text(x, y, labels = "(X,Y)", pos = 4)
text(x/2, y/2, labels = expression(R), pos = 2)

# Counter-clockwise arc from 0 to theta_ccw
arc_r <- max(0.8, min(1.2, 0.8 * r))  # keep the arc visible
arc_t <- seq(0, theta_ccw, length.out = 200)
lines(arc_r * cos(arc_t), arc_r * sin(arc_t), col = "darkgreen", lwd = 2)

# Angle label at midpoint of the arc
mid_t <- theta_ccw / 2
text(arc_r * cos(mid_t), arc_r * sin(mid_t), labels = expression(theta),
     col = "darkgreen", pos = 3)
```

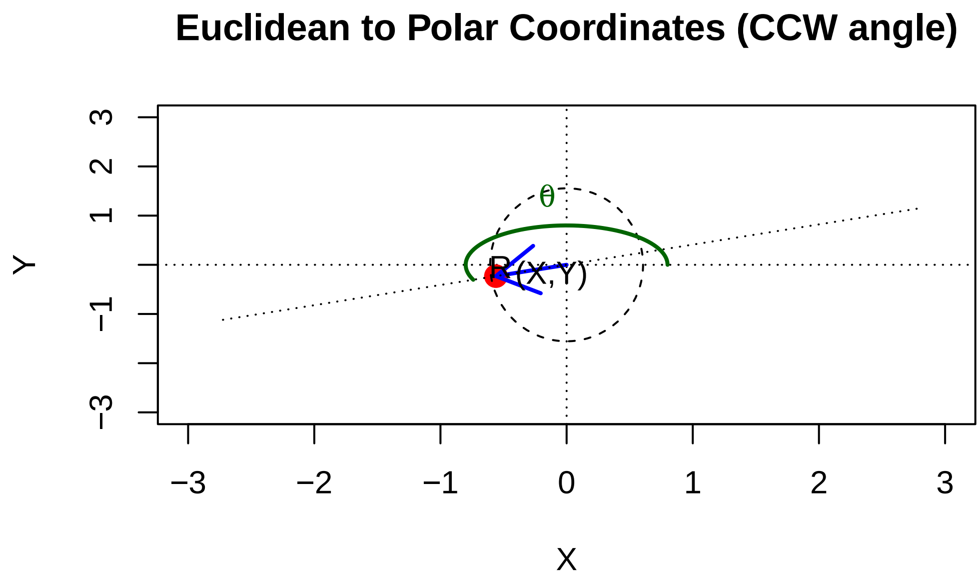

Figure 1: Euclidean point and its polar representation

This plot shows:

- (X, Y) as the red dot in Euclidean space.
- R as the radius of the dashed circle.
- Θ measured **counter-clockwise** from the positive X-axis, indicated by the green arc.
- Dashed lines through the origin along the axes and through (X, Y).

This plot shows:

- (X, Y) as the red dot in Euclidean space.
- R as the radius of the dashed circle.
- Θ measured counter-clockwise from the positive X-axis, indicated by the green arc.
- Dashed lines through the origin along the X-axis and through (X, Y).

Box–Muller Method

Generate standard normal random values (X, Y) from independent uniforms (U_1, U_2).

1. Generate $U_1, U_2 \overset{\text{iid}}{\sim} \text{Unif}(0, 1)$.
2. $-2\,\log U_1 \sim \text{Exp}(\frac{1}{2})$ and $2\pi U_2 \sim \text{Unif}(0, 2\pi)$.
3. Set $X = \sqrt{-2\,\log U_1}\,\cos(2\pi U_2)$ and $Y = \sqrt{-2\,\log U_1}\,\sin(2\pi U_2)$, which yields $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$.

R Implementation

```r
set.seed(42)

# Box-Muller generator: returns n pairs (X, Y)
box_muller <- function(n) {
  u1 <- runif(n)
  u2 <- runif(n)
  r <- sqrt(-2 * log(u1))
  theta <- 2 * pi * u2
  x <- r * cos(theta)
  y <- r * sin(theta)
  data.frame(x = x, y = y)
}

# Example: draw 10,000 samples and check moments
samples <- box_muller(10000)
c(mean_x = mean(samples$x), var_x = var(samples$x),
  mean_y = mean(samples$y), var_y = var(samples$y))
```

```
      mean_x        var_x       mean_y        var_y 
0.0005101192 1.0244204655 0.0006820139 0.9999651650 
```

```r
# Quick visual checks
par(mfrow = c(1, 3))
hist(samples$x, breaks = 40, main = "X ~ N(0,1)")
hist(samples$y, breaks = 40, main = "Y ~ N(0,1)")
plot(samples$x, samples$y, pch = 16, cex = 0.4,
     main = "Scatter of (X,Y)")
```

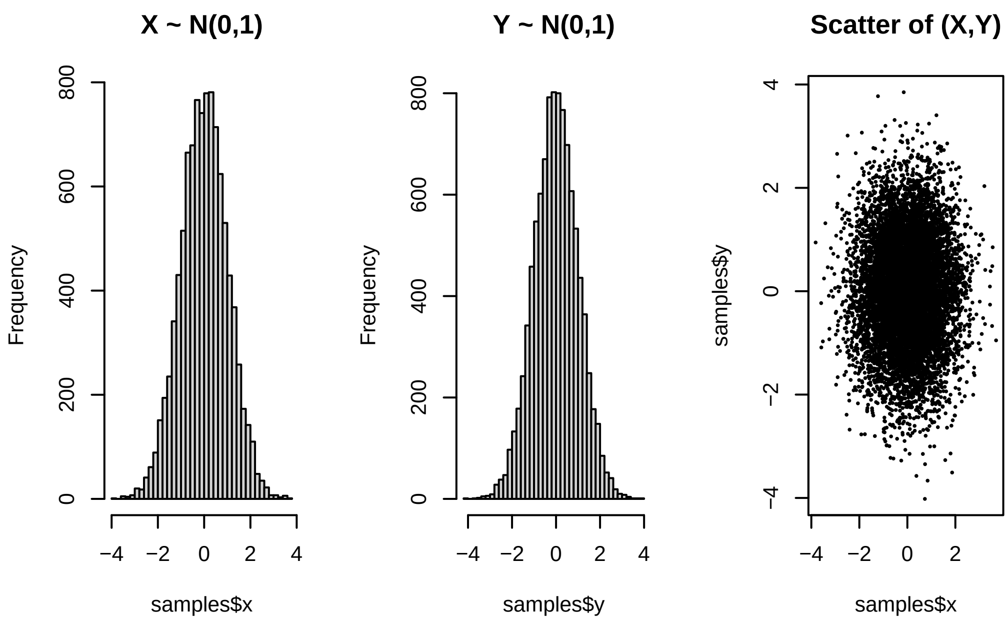

Figure 2: Diagnostics for Box–Muller samples

```
par(mfrow = c(1, 1))
```

> Notes: The derivations above follow the change-of-variables formula with the polar transformation, showing that R is Rayleigh and Θ is Uniform, which leads directly to the Box–Muller algorithm.

Derivation of the Rayleigh Distribution

1. From Uniform to Exponential via the Inverse CDF Method

Let $U \sim \text{Unif}(0,1)$. To generate an exponential random variable T with rate $\lambda = \frac{1}{2}$ (mean 2), we use the Inverse CDF Method.

The CDF of $\text{Exp}(\frac{1}{2})$ is

$F_T(t) = 1 - e^{-t/2}, \quad t \geq 0.$

Set $F_T(t) = U$ and solve for t:

$U = 1 - e^{-t/2}$

$e^{-t/2} = 1 - U$

$-\frac{t}{2} = \ln(1 - U)$

$t = -2\, \ln(1 - U).$

Since $1 - U \sim \text{Unif}(0, 1)$ as well, we can equivalently write

$t = -2\, \ln U.$

Thus

$T = -2 \ln U \sim \text{Exp}(1/2).$

---

2. From Exponential to Rayleigh

We want R such that $R^2 \sim \text{Exp}(\frac{1}{2})$.
That is, if $T \sim \text{Exp}(1/2)$ then $R = \sqrt{T}$ is Rayleigh distributed.

The CDF of R is

$F_R(r) = P(R \leq r) = P(\sqrt{T} \leq r) = P(T \leq r^2).$

Because $T \sim \text{Exp}(1/2)$,

$F_T(t) = 1 - e^{-t/2}.$

So

$F_R(r) = F_T(r^2) = 1 - e^{-r^2/2}, \quad r \geq 0.$

Differentiating gives the Rayleigh PDF:

$f_R(r) = \dfrac{d}{dr} F_R(r) = r e^{-r^2/2}, \quad r \geq 0.$

---

3. Inverse CDF Check for Rayleigh

To generate R directly from $U \sim \text{Unif}(0, 1)$:

Set $F_R(r) = U$:

$U = 1 - e^{-r^2/2}$

$e^{-r^2/2} = 1 - U$

$r^2 = -2\ \ln(1 - U)$

$r = \sqrt{-2\ \ln(1 - U)}.$

Since $1 - U \sim \text{Unif}(0, 1)$, we can write

$r = \sqrt{-2\ \ln U},$

which matches the Box–Muller derivation.

R Implementation

```r
# Generate Rayleigh samples using inverse transform
rayleigh_inverse <- function(n) {
  u <- runif(n)
  r <- sqrt(-2 * log(u))
  r
}

# Check empirical vs theoretical mean
set.seed(123)
samples_r <- rayleigh_inverse(10000)
c(emp_mean = mean(samples_r), emp_var = var(samples_r),
  th_mean = sqrt(pi/2), th_var = (4 - pi)/2)
```

```
 emp_mean   emp_var   th_mean    th_var 
1.2583035 0.4248272 1.2533141 0.4292037 
```

```r
hist(samples_r, breaks = 40, probability = TRUE,
     main = "Rayleigh( =1)", xlab = "r")
curve(x * exp(-x^2/2), from = 0, to = 5, add = TRUE, col = "red", lwd = 2)
```

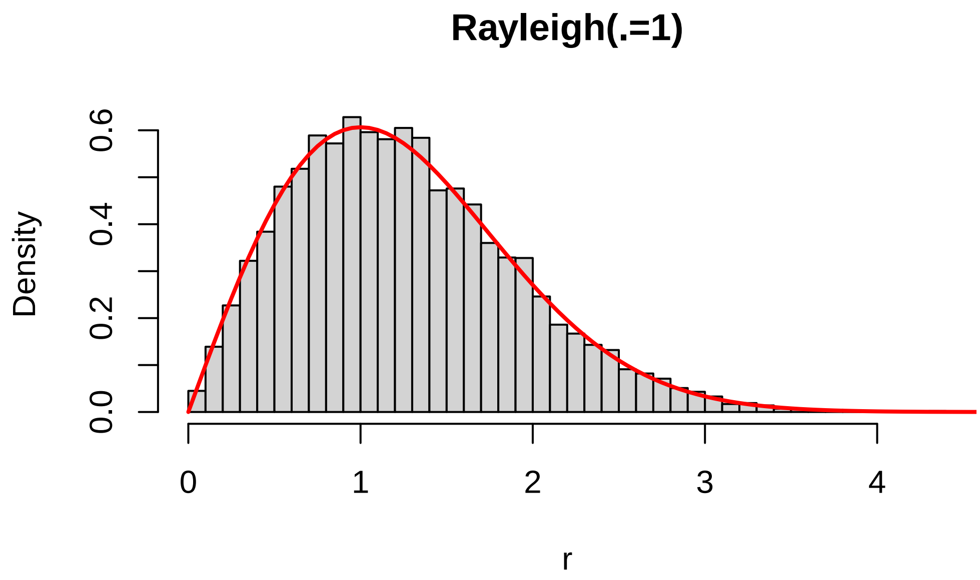

Figure 3: Plot generated by R code

Thus, the Rayleigh distribution naturally arises as the distribution of $R = \sqrt{-2\ln U}$, completing the connection between uniforms, exponentials, and normals via the Box–Muller construction.

Deriving the Rayleigh via the Inverse CDF Method

1) From Uniform to $\mathrm{Exp}(\frac{1}{2})$

Let $U \sim \mathrm{Unif}(0,1)$. The target CDF for $T \sim \mathrm{Exp}(\frac{1}{2})$ is $F_T(t) = 1 - e^{-t/2},\ t \geq 0$.

Set $U = F_T(T)$ and solve for T: $U = 1 - e^{-T/2}$ $e^{-T/2} = 1 - U$ $-\dfrac{T}{2} = \log(1-U)$ $T = -2\log(1-U)$.

Since $1 - U \overset{d}{=} U$ for $U \sim \mathrm{Unif}(0,1)$, an equivalent generator is $T = -2\log U$, which yields $T \sim \mathrm{Exp}(\frac{1}{2})$.

2) From $\mathrm{Exp}(\frac{1}{2})$ to Rayleigh

Define the transformation $R = \sqrt{T}$.

For $r \geq 0$, $\mathbb{P}(R \leq r) = \mathbb{P}(\sqrt{T} \leq r) = \mathbb{P}(T \leq r^2) = F_T(r^2) = 1 - e^{-r^2/2}$.

Thus the CDF of R is $F_R(r) = 1 - e^{-r^2/2},\ r \geq 0$, and the PDF is $f_R(r) = \dfrac{d}{dr}F_R(r) = r\,e^{-r^2/2},\ r \geq 0$, which is the Rayleigh distribution with parameter $\sigma = 1$.

3) Verifying via the Inverse CDF for Rayleigh

Start from $U \sim \text{Unif}(0,1)$ and set $U = F_R(R)$ with $F_R(r) = 1 - e^{-r^2/2},\ r \geq 0$.

Solve for R: $U = 1 - e^{-R^2/2}\ e^{-R^2/2} = 1 - U\ -\dfrac{R^2}{2} = \log(1-U)\ R = \sqrt{-2\,\log(1-U)}$.

Again using $1 - U \overset{d}{=} U$, an equivalent generator is $R = \sqrt{-2\,\log U}$.

Combining parts (1) and (2), if $T = -2\,\log U \sim \text{Exp}(\frac{1}{2})$ and $R = \sqrt{T}$, then $F_R(r) = 1 - e^{-r^2/2}$ and $f_R(r) = r\,e^{-r^2/2}$, confirming that the transformation produces a Rayleigh random variable.

The Bivariate Normal Distribution

We introduce the general **bivariate normal** distribution with different means, variances, and correlation.
Notation: $BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$.

1) Adding Means and Variances

Let the random vector be

$$\mathbf{Z} = \begin{bmatrix} X \\ Y \end{bmatrix}.$$

We specify its mean vector as

$$\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix},$$

and its covariance matrix without correlation as

$$\Sigma_0 = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}.$$

Thus if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ independently, then

$$\mathbf{Z} \sim BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, 0).$$

2) Adding Correlation ρ

To introduce correlation, the covariance matrix becomes

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Thus the full distribution is

$$\mathbf{Z} \sim BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho).$$

Its density is

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

In compact matrix notation, with

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_X^2 & -\rho/(\sigma_X\sigma_Y) \\ -\rho/(\sigma_X\sigma_Y) & 1/\sigma_Y^2 \end{bmatrix} \cdot \frac{1}{1-\rho^2},$$

the density can be written as

$$f_{X,Y}(\mathbf{z}) = \frac{1}{2\pi\,|\Sigma|^{1/2}} \exp\bigl(-\tfrac{1}{2}(\mathbf{z}-\mu)^\top\Sigma^{-1}(\mathbf{z}-\mu)\bigr).$$

This general form compactly encodes the means, variances, and correlation structure of the bivariate normal distribution.

Bivariate Normal $\mathrm{BVN}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$

We derive the joint density in two steps: first with arbitrary means and variances (no correlation), then add correlation ρ using matrix notation.

1) Add means and variances (no correlation)

Let the random vector be $\mathbf{Z} = \begin{bmatrix} X \\ Y \end{bmatrix}$ with mean $\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$ and covariance matrix $\Sigma_0 = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$.

Equivalently, $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ independently. The joint PDF factors: $f_{X,Y}(x,y) = \dfrac{1}{2\pi\,\sigma_X\,\sigma_Y} \exp\left\{-\dfrac{1}{2}\left[\dfrac{(x-\mu_X)^2}{\sigma_X^2} + \dfrac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\}$.

In matrix notation, with $\mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix}$, $f_{X,Y}(\mathbf{z}) = \dfrac{1}{(2\pi)^1} \dfrac{1}{\sqrt{|\Sigma_0|}} \exp\left\{-\dfrac{1}{2}(\mathbf{z}-\mu)^\top \Sigma_0^{-1}(\mathbf{z}-\mu)\right\}$, where $|\Sigma_0| = \sigma_X^2\,\sigma_Y^2$ and $\Sigma_0^{-1} = \begin{bmatrix} 1/\sigma_X^2 & 0 \\ 0 & 1/\sigma_Y^2 \end{bmatrix}$.

Thus $\mathbf{Z} \sim \mathrm{BVN}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, 0)$.

2) Add correlation ρ

Introduce covariance $\mathrm{Cov}(X,Y) = \rho\,\sigma_X\sigma_Y$ with $\rho \in (-1,1)$. The covariance matrix becomes $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$.

Its determinant and inverse are $|\Sigma| = \sigma_X^2\,\sigma_Y^2\,(1-\rho^2)$

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1/\sigma_X^2 & -\rho/(\sigma_X\sigma_Y) \\ -\rho/(\sigma_X\sigma_Y) & 1/\sigma_Y^2 \end{bmatrix}.$$

The bivariate normal density with correlation ρ is

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

In compact matrix form, $f_{X,Y}(\mathbf{z}) = \dfrac{1}{(2\pi)^1} \dfrac{1}{\sqrt{|\Sigma|}} \exp\left\{-\dfrac{1}{2}(\mathbf{z}-\mu)^\top \Sigma^{-1}(\mathbf{z}-\mu)\right\}$.

> Notes: 1) When $\rho = 0$, the off-diagonal terms vanish and the density reduces to the independent case above.
> 2) Existence requires $\sigma_X > 0$, $\sigma_Y > 0$, and $|\rho| < 1$ so that Σ is positive definite.
> 3) One can obtain $\mathbf{Z}$ by an affine transform of a standard normal $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ using a matrix square root (e.g., Cholesky): $\mathbf{Z} = \mu + \mathbf{L}\,\mathbf{N}$ with $\mathbf{L}\,\mathbf{L}^\top = \Sigma$.

Deriving the Bivariate Normal $BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ via PDF Transformations

We use matrix notation. Let $\mathbf{z} = (z_1, z_2)^\top$ with $z_1, z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and joint PDF $f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\,\mathbf{z}^\top \mathbf{z}\right\}$.

Define $\mu = (\mu_X, \mu_Y)^\top$ and $\mathbf{D} = \operatorname{diag}(\sigma_X, \sigma_Y)$.

1) Add Means and Variances (Independent Case)

Consider the affine map $\mathbf{x}_0 = \mu + \mathbf{D}\,\mathbf{z}$. Then $\mathbf{z} = \mathbf{D}^{-1}(\mathbf{x}_0 - \mu)$ and the Jacobian is $\left|\det \frac{\partial \mathbf{z}}{\partial \mathbf{x}_0}\right| = \left|\det \mathbf{D}^{-1}\right| = \frac{1}{\sigma_X \sigma_Y}$.

By the PDF method, $f_{\mathbf{X}_0}(\mathbf{x}_0) = f_{\mathbf{Z}}(\mathbf{D}^{-1}(\mathbf{x}_0 - \mu))\,\left|\det \mathbf{D}^{-1}\right|$.

Therefore $f_{\mathbf{X}_0}(x_0, y_0) = \dfrac{1}{2\pi\,\sigma_X \sigma_Y} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_0 - \mu_X}{\sigma_X}\right)^2 + \left(\frac{y_0 - \mu_Y}{\sigma_Y}\right)^2\right]\right\}$,

which is the independent bivariate normal with means μ_X, μ_Y and variances σ_X^2, σ_Y^2.

2) Add Correlation ρ (Keep Means and Variances)

Work with centered variables $\tilde{\mathbf{x}}_0 = \mathbf{x}_0 - \mu$ and construct $\tilde{\mathbf{x}} = \mathbf{C}\,\tilde{\mathbf{x}}_0$, where $\mathbf{C} = \begin{bmatrix} 1 & 0 \\ \rho\,\frac{\sigma_Y}{\sigma_X} & \sqrt{1-\rho^2} \end{bmatrix}$.

Set the final variables as $\mathbf{x} = \mu + \tilde{\mathbf{x}} = \mu + \mathbf{C}(\mathbf{x}_0 - \mu)$.

The Jacobian of the map $\mathbf{x}_0 \mapsto \mathbf{x}$ equals $\left|\det \frac{\partial \mathbf{x}_0}{\partial \mathbf{x}}\right| = \left|\det \mathbf{C}^{-1}\right| = \frac{1}{\sqrt{1-\rho^2}}$.

Using the PDF method again with translation invariance of the Jacobian, we obtain $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_0}(\mathbf{C}^{-1}(\mathbf{x} - \mu) + \mu)\,\left|\det \mathbf{C}^{-1}\right|$.

After algebra (or by composing the two linear maps on $\mathbf{z}$), the covariance of $\mathbf{x}$ is $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$, and the joint PDF simplifies to the standard BVN form $f_{X,Y}(x,y) = \dfrac{1}{2\pi\,\sigma_X\sigma_Y\,\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$.

Matrix Notation Summary

Let $\mu = (\mu_X, \mu_Y)^\top$ and $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$. Then the density can be written compactly as $f_{\mathbf{X}}(\mathbf{x}) = \dfrac{1}{2\pi\sqrt{\det\Sigma}}\,\exp\Big\{-\dfrac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\Big\}$, where $\det\Sigma = \sigma_X^2\sigma_Y^2(1-\rho^2)$ and $\Sigma^{-1} = \dfrac{1}{(1-\rho^2)\,\sigma_X^2\sigma_Y^2}\begin{bmatrix} \sigma_Y^2 & -\rho\,\sigma_X\sigma_Y \\ -\rho\,\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}$.

The two-step construction explicitly shows how translation and diagonal scaling introduce means and variances (with Jacobian $|\det\mathbf{D}^{-1}|$), and how a subsequent shear/rotation $\mathbf{C}$ introduces correlation (with Jacobian $|\det\mathbf{C}^{-1}|$), yielding the full $BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ density.