Multivariate normal distributions

Standard bivariate normal with correlation $\rho \in (-1,1)$: Consider the continuous bivariate RV $Z=(X,Y)$ with density $f_{(X,Y)}(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}}$. Completing the square in the exponent (there are two ways to do that!) in the form $x^2-2\rho xy+y^2=(x-\rho y)^2+(1-\rho^2)y^2$ we can rewrite

$$f_{(X,Y)}(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{(x-\rho y)^2+(1-\rho^2)y^2}{2(1-\rho^2)}}.$$

Integrating this formula and using the change of variable $u=\frac{1}{\sqrt{1-\rho^2}}(x-\rho y),\quad v=y$ with $dudv=\frac{1}{\sqrt{1-\rho^2}}dxdy$ shows that

$$\int\int f_{(X,Y)}(x,y)dxdy=\frac{1}{2\pi}\int\int e^{-u^2/2-v^2/2}dudv=1.$$

It also shows that the marginal density f_Y is a normal $\mathcal{N}(0,1)$ density because

$$f_Y(y)=\int f_{(X,Y)}(x,y)dx=\frac{1}{2\pi\sqrt{1-\rho^2}}\int e^{-\frac{(x-\rho y)^2+(1-\rho^2)y^2}{2(1-\rho^2)}}dx=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\times\frac{1}{\sqrt{2\pi(1-\rho^2}}$$

By exchanging the roles of X and Y, we see that the first marginal density f_X is also a normal $\mathcal{N}(0,1)$ density. This shows that $\mathbf{Var}(X)=\mathbf{Cov}(X,X)=\mathbf{Cov}(Y,Y)=\mathbf{Var}(Y)=1.$ We can also observe that X and Y are not independent unless $\rho=0$ because when $\rho\neq 0$, $f_{(X,Y)}(x,y)\neq f_X(x)f_Y(y).$ Next we compute the covariance of X,Y.

$$\begin{aligned}\mathbf{Cov}(X,Y)&=\frac{1}{2\pi\sqrt{1-\rho^2}}\int\int xye^{-\frac{(x-\rho y)^2+(1-\rho^2)y^2}{2(1-\rho^2)}}dxdy\\&=\frac{1}{\sqrt{2\pi}}\int ye^{-y^2/2}\left\{\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int xe^{-\frac{(x-\rho y)^2}{2(1-\rho^2)}}dx\right\}dy\end{aligned}$$

The quantity inside the large parentheses is the expected value of a normal with mean ρy and variance $(1-\rho^2)$ so it must be equal to ρy. This gives

$$\mathrm{Cov}(X,Y) = \frac{1}{\sqrt{2\pi}} \int \rho y^2 e^{-y^2/2} dy = \rho.$$

Here we used the fact that the second moment of a normal $\mathcal{N}(0,1)$ is 1. The correlation $\mathrm{Corr}(X,Y)$ is also equal to ρ.

Standard bivariate normal distributions, $\rho = 0$: The standard uncorrelated bivariate normal $Z = (X,Y)$, (i.e., $\rho = 0$ above) has density

$$f_Z(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \times \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

It follows that X and Y are independent because $f_Z(x,y) = f_X(x) f_Y(y)$. This is a very special property because, **in general**, uncorrelated random variable may very well be dependent.

In dimension n, a standard normal random vector $X \sim \mathcal{N}(0, I_n)$ (I_n stands for the $n \times n$ identity matrix, see below), is a vector with density

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} e^{-\|\mathbf{x}\|^2/2} \text{ where } \|\mathbf{x}\|^2 = x_1^2 + \cdots + x_n^2.$$

Observe this is the product of n univariate normal $\mathcal{N}(0,1)$ densities (one for each coordinate). This means the the coordinates of X, $X = (X_1, \ldots, X_n)$ (in any orthonormal basis!) are independent normal $\mathcal{N}(0,1)$ random variables. Of course, this also means that each of the n univariate marginals $X_1, X_2, \ldots, X_n$ is a standard normal.

General bivariate normal distributions There are several **equivalent** definitions of what a general n dimensional normal random vector is (it is an important subject!). We discuss some of these definitions in dimension 2 (general bivariate random vector).

One rather natural point of view is to call normal random vector any vector that is obtained through very basic simple transformations from a standard normal vector $\mathbf{Z} = (X,Y)$ with independent coordinates, i.e., a random vector with density $f_{\mathbf{Z}}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$. What do we mean by basic simple transformations"? We mean we can transform our random vectors using two types of moves:

- Invertible linear transformation (in the form of matrix multiplication): $\mathbf{W}^T = (U,V)^T = A\mathbf{Z}^T = A(X,Y)^T$ (we need the transpose to go from row-vectors to column-vectors) where A is any fixed invertible 2×2 matrix so that $\mathbf{Z}^T = A^{-1}\mathbf{W}^T$.
- Translation (addition of a fix vector): $\mathbf{W} = (U,V) = \mathbf{Z} + \mathbf{w}_0 = (X + u_0, Y + v_0)$.

Bivariate normal (general form) I: A two dimensional random vector $\mathbf{W} = (U, V)$ is {\em bivariate normal} if there is a vector $\mathbf{w}_0 = (u_0, v_0)$ and an invertible matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that

$$\mathbf{W}^T = \begin{bmatrix} U \\ V \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}$$

where $\mathbf{Z} = (X, Y)$ is $\mathcal{N}(0, I_2)$. Equivalently, the vector (X, Y) given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} = A^{-1} \begin{bmatrix} U - u_0 \\ V - v_0 \end{bmatrix}$$

is a standard normal $\mathcal{N}(0, I_2)$ vector.

We can compute the density of the vector $\mathbf{W}$ by a simple (linear) change of variable

$$\mathbf{w}^T = \begin{bmatrix} u \\ v \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \quad \begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} \text{ with } dxdy = \frac{1}{|\det A|} dudv.$$

This gives

$$f_{\mathbf{W}}(u, v) = \frac{1}{2\pi |\det A|} e^{-\frac{1}{2}[A^{-1}(\mathbf{w}-\mathbf{w}_0)^T \cdot A^{-1}(\mathbf{w}-\mathbf{w}_0)^T]}.$$

Observe that the vector $\mathbf{w}_0 = (u, v)$ has a simple interpretation because $U = aX + bY + u_0$ has expectation

$$E(U) = aE(X) + bE(Y) + u_0 = u_0.$$

The components u_0, v_0 of the vector $\mathbf{w}_0$ are the marginal expectations, $u_0 = E(U), v_0 = E(V)$. We can also compute $\sigma_U^2 = \mathrm{Var}(U), \sigma_V^2 = \mathrm{Var}(V)$ and

$$\sigma_{UV} = \mathrm{Cov}(U, V) = E((U - E(U))(V - E(V)))$$

because $U - E(U) = aX + bY, V - E(V) = cX + dY$ so

$$\mathrm{Var}(U) = a^2 + b^2, \ \ \mathrm{Var}(V) = c^2 + d^2, \ \ \mathrm{Cov}(U, V) = ac + bd.$$

If we call ρ the correlation coefficient of (U, V), by definition, we have $\sigma_{UV} = \rho \sigma_U \sigma_V$.

We can rewrite the formula for $f_{\mathbf{W}}$ using the following notation. Set $\Sigma = AA^T$. This is an invertible symmetric positive definite matrix. Recall that symmetric means that $\Sigma^T = \Sigma$ and positive definite means that, for any row-vector $\mathbf{w}$, $\mathbf{w}\,\Sigma\,\mathbf{w}^T = \mathbf{w}^T \cdot \Sigma\,\mathbf{w}^T \geq 0$ with equality only when $\mathbf{w} = 0$. If we compute Σ in terms of the entries of A, we get

$$\Sigma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix} = \begin{bmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{bmatrix}.$$

We call Σ the **covariance matrix** of the vector (U,V) because its entries are the variances and covariances of the pair (U,V), namely,

$$\Sigma]_{11} = \sigma_U^2,\ [\Sigma]_{12} = [\Sigma]_{21} = \sigma_{UV},\ [\Sigma]_{22} = \sigma_V^2.$$

Covariance matrices are always symmetric. The inverse of Σ, Σ^{-1}, is given by $\Sigma^{-1} = (A^{-1})^T(A^{-1})$ and it is also an invertible symmetric positive definite matrix. Now, we can see that the previous formula for $f_{\mathbf{W}}$ gives

$$\begin{aligned} f_{\mathbf{W}}(u,v) &= \frac{1}{2\pi\sqrt{\det\Sigma}} e^{-\frac{1}{2}(\Sigma^{-1}(\mathbf{w}-\mathbf{w}_0)^T)\cdot(\mathbf{w}-\mathbf{w}_0)^T} \\ &= \frac{1}{2\pi\sqrt{\det\Sigma}} e^{-\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)\Sigma^{-1}(\mathbf{w}-\mathbf{w}_0)^T}. \end{aligned}$$

Bivariate normal (general form) II: Let $\mathbf{w}_0$ be a two dimensional constant vector and Σ be a two by two symmetric positive definite matrix. A two dimensional random vector $\mathbf{W} = (U,V)$ is called {\em bivariate normal} $\mathcal{N}(\mathbf{w}_0, \Sigma)$ if it has density

$$f_{\mathbf{W}}(u,v) == \frac{1}{2\pi\sqrt{\det\Sigma}} e^{-\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)\Sigma^{-1}(\mathbf{w}-\mathbf{w}_0)^T}.$$

This "new" definition is compatible with our previous definition.

Bivariate normal (general form) I \\(\Longleftrightarrow\\) **II:** Let $\mathbf{w}_0$ be a two dimensional constant vector, A an invertible two by two matrix and Σ be a two by two symmetric positive definite matrix with $AA^T = \Sigma$. The two dimensional random vector $\mathbf{W} = (U,V)$ satisfies $W^T = AZ^T + \mathbf{w}_0^T$ with $Z \sim \mathcal{N}(0, I_2)$ if and only if W is normal $\mathcal{N}(\mathbf{w}_0, \Sigma)$ with Σ symmetric positive definite.

We now explains why we can always find such a matrix A if we are given the covariance matrix Σ. Recall that any symmetric matrix has real eigenvalues and is diagonalizable in an orthonormal basis. If this matrix is positive definite, its eigenvalues are (strictly) positive. We apply this to the matrix Σ to see that

$$\Sigma = QDQ^{-1} \text{ with } Q \text{ orthogonal and } D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},\ \lambda_1 > 0,\ \lambda_2 > 0.$$

The fact that Q is orthogonal means that its column vectors are orthonormal or, equivalently $Q^TQ = QQ^T = I$. This gives us a way to find a square root'' for the matrix Σ, namely, the matrix

$$M = Q \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} Q^{-1}.$$

Indeed, this matrix satisfies $M = M^T, \ \ M^2 = \Sigma$ and M is invertible. Now, we check (using a change of variables) that if $\mathbf{W} \sim N(\mathbf{w}_0, \Sigma)$ then $\mathbf{Z}^T = M^{-1}(\mathbf{W}^T - \mathbf{w}_0^T)$ is $N(0, I_2)$. This is equivalent to say that $\mathbf{W}^T = M\mathbf{Z}^T + \mathbf{w}_0^T$ where $\mathbf{Z} \sim \mathcal{N}(0, I_2)$. We know that

$$f_{\mathbf{W}}(u,v) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)\Sigma^{-1}(\mathbf{w}-\mathbf{w}_0)^T}.$$

We use the change of variables provided by the matrix M

$$\mathbf{w}^T = M\mathbf{z}^T + \mathbf{w}_0^T, \ \ \mathbf{w} = (u,v), \ \ \mathbf{z} = (x,y), \ \ dudv = |\det M| dxdy = \sqrt{\det \Sigma}\, dxdy$$

and find that the vector $\mathbf{Z} = (X, Y)$ has density $\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$. This means that $\mathbf{Z}$ is indeed $\mathcal{N}(0, I)$ as desired.

General multivariate normal distributions Everything we said for bivariate normal random variables generalizes to higher dimensions. A general n-dimensional normal random vector $\mathbf{X} = (X_1, \dots, X_n)$ (also called Gaussian) is a random vector with density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T}, \ \ \mathbf{x} = (x_1, \dots, x_n), \text{ with } \boldsymbol{\mu} = (\mu_1, \dots, \mu_n).$$

The vector $\boldsymbol{\mu}$ gives the marginal means (i.e., expectations) $\mu_i = E(X_i)$ and the $n \times n$ matrix Σ gives the variance and covariances $[\Sigma]_{ij} = \mathrm{Cov}(X_i, X_j)$. We call this distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, normal (or Gaussian) with mean $\boldsymbol{\mu}$ and covariance matrix Σ.