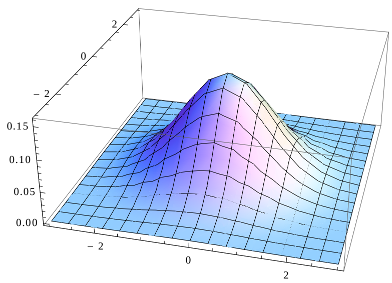
The Multivariate Normal Distribution

The univariate normal distribution is an extremely familiar concept where some random variable X can take values along the real with probabilities that match the famouse bell-curve. Recall the probability density function of

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

However, that's limited to only one dimension, and we would like to generalize to higher dimensions. In the next-simplest 2-dimensional case, we'd like a distribution that actually looks like a bell—where potentional values can range over the real plane, $\mathbb{R}$, where the density is clustered around some mean before tapering off in all directions, as seen below.



This figure has mean zero for both X_1 and and X_2, and which are independent, implying $\sigma = I_2$, the identity matrix. It's easy to see that any vertical cuts parallel to xz or yz planes will yield a traditional normal random variable. This of course generalizes to higher dimensions, although we can't display it so nicely.

1 Notation

In this note, the multivariate distribution will apply to a d-dimensional random vector

$$\mathbf{X} = \begin{pmatrix} X_1 & X_2 & \dots & X_d \end{pmatrix}^T, \qquad \mathbf{X} \sim N_d(\mu, \Sigma)$$

where μ is the n-dimensional *mean vector*,

$$\mu = \begin{pmatrix} EX_1 & EX_2 & \dots & EX_d \end{pmatrix}^T,$$

and where Σ is the $d \times d$ *covariance matrix*, which is defined and has in its i, j entry

$$\sigma^2 = E\left[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T\right] \in \mathbb{R}^{N \times N}$$

$$\Sigma_{ij} = Cov(X_i, X_j), \qquad i, j = 1, \dots, d$$

All MVN random variables (and random vectors, more generally) will be in boldface, such as $\mathbf{X}$ for easy subscripting. $\mathbf{X}_t$ will indicate the random vector $\mathbf{X}$ at time t, while X_i indicates the ith element of $\mathbf{X}$. Constants, vectors of constants, and matrices of constants will not be in boldface.

2 Definition

A random vector $\mathbf{X}$ has a *multivariate normal* distribution if every linear combination of its components,

$$Y = a_1 X_1 + \ldots + a_d X_d$$
$$\Leftrightarrow Y = a\mathbf{X}, \qquad a \in \mathbb{R}^d$$

is *univariate normally distributed*, with a corresponding mean and variance. This gives a joint density function of

$$f_X(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\, e^{-\frac{1}{2}(x-\mu)^T\, \Sigma^{-1}(x-\mu)}, \qquad |\Sigma| = \det \Sigma \qquad x \in \mathbb{R}^d \tag{1}$$

Note, we impose the requirement that Σ is symmetric and positive definite. Symmetric because the correlation of X and Y equals the correlation of Y and X.

Terminology: I'll often us MVN to refer to the case where $\mathbf{X}$ is a vector with $d \geq 2$; however, it should be clear that the univariate normal distribution is just a special case where $d = 1$. Therefore, when speaking about an RV that could be either MVN or univariate normal *or* properties that apply equally well to either type of RV, I'll often use the term *Gaussian* both for convenience and out of reverence to long-dead German-speaking mathematicians. Stimmt?

3 Linear Transformations of MVN Random Variables

It's fairly common to consider linear transformations and functions of a multivariate normally distributed random variable. For instance, we might have an economic or statistical model with a recurrence relation to describe the dynamics of some process:

$$\mathbf{X}_{t+1} = A\mathbf{X}_t + \mathbf{V}_{t+1}$$

where $\mathbf{V}_t$ is some innovation or random noise vector. Therefore, it would be useful to be able to derive the distributions of *functions* or *linear transformation* of multivariate normal random variables.

3.1 Transformation Theory Recap

So let $\mathbf{X} = A\mathbf{Y}$. Suppose we have the distribution of $\mathbf{X}$, denoted f_X, and we want the distribution of $\mathbf{Y}$, f_Y. Then

$$\begin{aligned} f_Y(y) &= |\det(A)|\; f_X(Ay) \\ f_X(x) &= \frac{1}{|\det(A)|}\; f_Y(A^{-1}x) \end{aligned} \tag{2}$$

3.2 Derivation of Probability Distribution

So let's find the probability distribution of a linear transforamtion of an MVN RV. Begin by assuming

$$\begin{aligned} \mathbf{X} = A\mathbf{Y}, &\qquad \mathbf{Y} \sim \text{MVN}(\mu, \Sigma) \\ \Rightarrow \quad f(y) = k\ \exp &\left\{ -\frac{1}{2}(y-\mu)^T\ \Sigma^{-1}(y-\mu) \right\} \end{aligned}$$

where k is a constant.[1] Assuming that A is invertible, we substitute in, using Equation 2, to get the distribution of $\mathbf{X}$:

$$\Rightarrow \quad f_X(A^{-1}x) = k' \exp\left\{ -\frac{1}{2}(A^{-1}x-\mu)^T\Sigma^{-1}(A^{-1}x-\mu) \right\} \tag{3}$$

Next, since the expectation is a linear operator, we can use the fact that

$$\begin{aligned} E\mathbf{X} &= E[A\mathbf{Y}] = AE\mathbf{Y} = A\mu \\ \Rightarrow \quad \mu_* &= A\mu \\ \Leftrightarrow \quad \mu &= A^{-1}\mu_* \end{aligned}$$

where μ' is the mean vector of $\mathbf{X}$. With that, we can substitute back into Equation 3 and simplify even further, using convenient matrix manipulations like the distributivity property, associativity, etc.:

$$\begin{aligned} f_X(A^{-1}x) &= k' \exp\left\{ -\frac{1}{2}(A^{-1}x-\mu)^T\ \Sigma^{-1}(A^{-1}x-\mu) \right\} \\ &= k' \exp\left\{ -\frac{1}{2}(A^{-1}x-A^{-1}\mu_*)^T\ \Sigma^{-1}(A^{-1}x-A^{-1}\mu_*) \right\} \\ &= k' \exp\left\{ -\frac{1}{2}[A^{-1}(x-\mu_*)]^T\ \Sigma^{-1}[A^{-1}(x-\mu_*)] \right\} \\ &= k' \exp\left\{ -\frac{1}{2}(x-\mu_*)^T\ [(A^{-1})^T\ \Sigma^{-1}A^{-1}](x-\mu_*) \right\} \\ &= k' \exp\left\{ -\frac{1}{2}(x-\mu_*)^T\ \Sigma_*^{-1}\ (x-\mu_*) \right\} \\ \Rightarrow \mathbf{X} \sim \text{MVN}(\mu_*, \Sigma_*), &\qquad \text{where } \mu_* = A\mu \text{ and } \Sigma_* = A\Sigma A^T \end{aligned}$$

This is huge! It means *linear transformations of MVN RV's are themselves MVN.*

[1] The constant will come from the definition of the distribution given above in Equation 1, but it's boring algebra not that interesting, so I'll suppress the details.

4 From Standard to General MVN Random Variables

If you're familiar with the standard univariate normal RV, then you can probably guess what the standard MVN RV is:

$$\mathbf{Z} \sim \text{MVN}(0,\ I_d) \tag{4}$$

where I_d is the $d \times d$ identity matrix. This form for the covariance matrix also suggests that the different components of $\mathbf{Z}$ (denoted $Z_1, Z_2, \ldots, Z_d$) are independent and Gaussian.

Moreover, just like we can build from a standard univariate to a general univariate. To do so, we'll use the results from the last section, since building from a standard MVN to general MVN simply involves linear transformations of the components. Specifically, we can express the MVN RV $\mathbf{X}$ as follows

$$\mathbf{X} = \mu + A\mathbf{Z}, \qquad \mathbf{X} \sim \text{MVN}(\mu, AA^T) \tag{5}$$

Computation: Suppose we know that we want $\mathbf{X}$ to be MVN(μ, Σ), and we can only generate $\mathbf{Z}$. How do we choose A such that $AA^T = \Sigma$. Typically, we'll have to use something like a *Cholesky Factorization* algorithm to find the correct A in the form of a lower triangular matrix. And if Σ is symmetric, positive definite, then A is guaranteed to exist and this approach will work. The algorithm itself can be found in the appendix.

5 Facts About Multivariate Normal Random Variables

So to summarize, MVN (or Gaussian) RV's are particularly nice to work with because of some convenient properties:

1. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are *univariate* normally distributed and independent. This then implies that they are *jointly normally distributed.* In other words, $\begin{pmatrix}\mathbf{X} & \mathbf{Y}\end{pmatrix}$ must have a multivariate normal distribution.

2. Linear functions of Gaussians are Gaussian. So if A is a constant matrix and $\mathbf{X}$ is MVN, then $A\mathbf{X}$ is also MVN.

3. Uncorrelated Gaussian random variables are independent.

4. Conditions Gaussian's are normal. So if X_1 and X_2 are two components of a MVN RV, $\mathbf{X}$, then $X_1|X_2$ is normal, and vice versa.