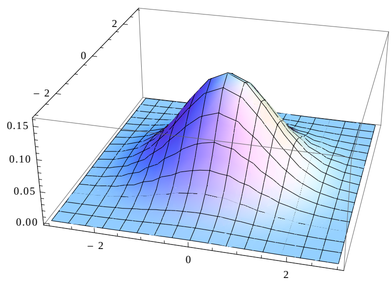
The Multivariate Normal Distribution

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, $\mathbb{R}^2$, with the density clustered around some mean before tapering off in all directions, as 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.

This generalization will give us the *Multivariate Normal (MVN) Distribution.* This turns out to be an extremely useful distribution. Let's highlight a few applications here:

1. The MVN distribution has applications in Bayesian multivariate regressions (i.e. Bayesian regressions with more than one "independent," "predictor," or "X" variable). In such applications, we hope to find the posterior distribution of the β regression vector, and it's common to assume a MVN prior on β that results in an MVN posterior for $p(\beta|y)$.

2. Next, we often use the MVN distribution to describe *recurrence relations*, where we try to forecast an MVN RV one period into the future, given current information about the RV, like it's current value and distribution. One example would include jointly forecasting height *and* weight in the future given current height and weight.

3. Finally, the MVN distribution forms the basis for VARs, which are used to forecast different economic variables one period into the future, expressing each future variable as a function of lags of itself and other economic variables. However, VARs are too broad and interesting a subject to elicit only a cursory summary in this note. I'll write a more extensive summary of them elsewhere.

Contents

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,

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

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:

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

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:

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

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

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

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.

6. Linear Gaussian Recurrences

So far, we've only looked at the properties of a multivariate normal RV at one point in time. This is already useful for cross-section regressions, but it would it would also be relevant to explore the MVN RV in in the context of time series and stochastic processes. For that, we turn to *recurrence relations*, which will conveniently characterize a discrete-time, continuous space stochastic process, $\{\mathbf{X}_t\}$.

6.1. Single Lag, Order 1 Markov Chains

Suppose we know that the stochastic process $\{\mathbf{X}_t\}$ evolves according to the following recurrence relation, where $\mathbf{Z}_t$ is independent of $\mathbf{X}_t$:

$$\mathbf{X}_{t+1} = A\mathbf{X}_t + B\mathbf{Z}_t \qquad \mathbf{Z}_t \sim \mathrm{N}_d(0,\ I_d) \tag{6}$$

Because $\mathbf{Z}_t$ is Gaussian, this implies that $\mathbf{X}_{t+1}$ will *also* be Gaussian for reasons we saw in Section 3 above. So let's define

$$\mathbf{X}_t \sim \mathrm{N}_d(\mu_t,\ \Sigma_t) \tag{7}$$

Forward Mean: Now our task is to compute μ_{t+1} and Σ_{t+1} given the parameters μ_t and Σ_t. Let's start with the mean:

$$\begin{aligned} \mu_{t+1} = E\mathbf{X}_{t+1} &= E\left[A\mathbf{X}_t + B\mathbf{Z}_t\right] \\ &= AE[\mathbf{X}_t] + BE[\mathbf{Z}_t] \\ \Rightarrow \qquad \mu_{t+1} &= A\mu_t \end{aligned}$$

Forward Covariance Matrix: Now, let's compute the covariance matrix, rewriting the recurrence relation somewhat suggestively:

$$\begin{aligned} \mathbf{X}_{t+1} &= A\mathbf{X}_t + B\mathbf{Z}_t \\ \Leftrightarrow \qquad \mathbf{X}_{t+1} - \mu_{t+1} &= A(\mathbf{X}_t - \mu_t) + B\mathbf{Z}_t \end{aligned}$$

Now, we move to the computation of Σ_{t+1}, which simplifies nicely (omitting the explicit simplification and derivation process):

$$\begin{aligned} \Sigma_{t+1} &= E\left\{\left[\mathbf{X}_{t+1} - \mu_{t+1}\right]\left[\mathbf{X}_{t+1} - \mu_{t+1}\right]^T\right\} \\ &= E\left\{\left[A(\mathbf{X}_t - \mu_t) + B\mathbf{Z}_t\right]\left[A(\mathbf{X}_t - \mu_t) + B\mathbf{Z}_t\right]^T\right\} \\ &= A\Sigma_t A^T + BB^T \end{aligned}$$

This is an example of a *forward equation* because the distribution of $\mathbf{X}_{t+1}$ is determined entirely by the distribution of $\mathbf{X}_t$ and the recurren relation.

Finally, note that in some cases, the noise vector won't necessarily be of dimension d. It might have dimension $m < d$, in which case B will have dimension $d \times m$. You can think of m correlated shocks being distributed over d separate variables.

6.2. Multiple Lag, Higher Order Markov Chains

Let's generalize a bit. Suppose we want a particular RV, $\mathbf{X}_{t+1}$, in a stochastic process, $\{\mathbf{X}_t\}$, to depend upon more than one lag, say "k" lags. Namely, suppose that

$$\mathbf{X}_{t+1} = A_0\mathbf{X}_t + A_1\mathbf{X}_{t-1} + \cdots + A_{k-1}\mathbf{X}_{t-k+1} + B\mathbf{Z}_t \tag{8}$$

We can use a *state space expansion* to augment the matrices and write Equation 8's more general recurrence relation as follows:

$$\begin{aligned} & \tilde{\mathbf{X}}_{t+1} = \tilde{A}_t\tilde{\mathbf{X}}_t + \tilde{B}_t\mathbf{Z}_t \\ \Leftrightarrow \quad & \begin{pmatrix} \mathbf{X}_{t+1} \\ \mathbf{X}_t \\ \vdots \\ \mathbf{X}_{t-k+2} \end{pmatrix} = \begin{pmatrix} A_0 & A_1 & \cdots & & A_{k-1} \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X}_t \\ \mathbf{X}_{t-1} \\ \vdots \\ \mathbf{X}_{t-k+1} \end{pmatrix} + \begin{pmatrix} B \\ 0 \\ \vdots \\ 0 \end{pmatrix} \mathbf{Z}_t \end{aligned}$$

Note that $\mathbf{Z}_t$ needs no augmenting because it is memoryless.

Forward Moments: Using the results from the previous subsection, we immediately know that if the stochastic process $\{\mathbf{X}_t\}$ satisifes Equation 8, $E\tilde{\mathbf{X}}_t = \tilde{\mu}_t$, and $\mathrm{Cov}(\tilde{\mathbf{X}}_t) = \tilde{\Sigma}_t$, then

$$\begin{aligned} \tilde{\mu}_{t+1} &= \tilde{A}\tilde{\mu}_t \\ \tilde{\Sigma}_{t+1} &= \tilde{A}\tilde{\Sigma}_t\tilde{A}^T + \tilde{B}\tilde{B}^T \end{aligned}$$

And if we want to find the covariance matrix Σ_{t+1} for $\mathbf{X}_{t+1}$, then we can just find the $kd \times kd$ covariance matrix $\tilde{\Sigma}_{t+1}$ and look at the top left $d \times d$ block.

State Space Expansion: Note what we did here. We had a stochastic process $\{\mathbf{X}_t\}$ that was not a Markov Chain—future values depended upon more than one lag of $\mathbf{X}_t$, violating the Markov Property. However, *state space expansion* allowed us to form a new stochastic process $\{\tilde{\mathbf{X}}_t\}$ that was a Markov Chain, obeying the Markov property. Therefore, anytime a stochastic process does not display the Markov Property, blame that on the size of the state space.

Also note, state space expansion brings us to the sitaution we mentioned in the previous subsection where the noise matrix, $\tilde{B}$, has fewer sources of noise than state variables.

A. Cholesky Decomposition

Recall the motivation: we want to find a matrix A such that $AA^T = \Sigma$. This will be useful when we want to transform a standard MVN RV into a general one via

$$\mathbf{X} = \mu + A\mathbf{Z}, \qquad \mathbf{X} \sim \mathrm{MVN}(\mu, AA^T) \tag{9}$$