MVNDemo

Empirical Market Microstructure

(2006, Oxford University Press)

Companion *Mathematica* notebook

Joel Hasbrouck

This notebook documents and demonstrates the use of the MVN package to work with multivariate normal distributions.

```
Text @ Style["Notebook evaluated " <> DateString["DateTime"], "Subtitle"]
```

Notebook evaluated Monday 4 June 2007 20:12:55

Documents and demonstrates the use of the MVN package to work with multivariate normal distributions. The MVN package defines a multivariate normal distribution "object" (i.e, a representation for variables that are multivariate normal), and various routines to work with them. It is not a substitute for the *Mathematica* package `Statistics`MultinormalDistribution``. The *Mathematica* package has more extensive facilities for working with density functions, distribution functions and related distributions. The MVN package has more facilities for analyzing linear combinations, joint and conditional densities, and for labeling the results (for purposes of documentation and clarity).

Initializations

```
SetDirectory[
  "c:/Active/Empirical Market Microstructure/Mathematica/Spring 2007"];
```

```
<< MVN.m
```

Properties of the multivariate normal distribution:

The package implements functionality based on the following relations.

Consider a multivariate normal distribution:

$$\underset{n\times 1}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix};\; E[X] = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix};\; Var[X] = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

The partitions here must be consistent, but are otherwise arbitrary. Either X_1 or X_2 or both may be scalars.

If $X \sim MVN(\mu, \Sigma)$, then

$$E[X_1 \mid X_2] = \mu_1 + (X_2 - \mu_2)\, \Sigma_{22}^{-1}\, \Sigma_{21} \tag{2}$$

$$Var[X_1 \mid X_2] = \Sigma_{11} - \Sigma_{12}\, \Sigma_{22}^{-1}\, \Sigma_{21} \tag{3}$$

Consider a linear transform $Y = AX + b$, where A is a rectangular matrix and b is a column vector. Then:

$$Y \sim MVN(A\mu + b, A\Sigma A') \tag{4}$$

The MVN "object"

```
? MVN
```

MVN[μ,Σ,v] is an object that represents a multivariate normal distribution with mean vector μ and covariance matrix Σ. The third parameter, v, is list of Mathematica expressions that correspond to the variables. (Typically they are variable names or formulas for the variables.)

Note: `MVN[.]` is simply a "wrapper" for grouping a mean vector, a covariance matrix and a set of variable labels. Among other things, it provides a descriptive output. The default display form of an `MVN` object is:

```
MVN[μ, Σ, v]
```

$$v \sim \mathcal{N}(\mu, \Sigma)$$

An `MVN` can be built by explicitly supplying the vector and matrix arguments:

```
NewDistribution = MVN[{μ1, μ2}, {{v11, v12}, {v21, v22}}, {x1, x2}]
```

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}\right)$$

The above statement assigns to `NewDistribution` the newly defined `MVN` object. `NewDistribution` may subsequently be used, displayed or passed as an argument to another routine.

```
NewDistribution
```

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}\right)$$

If the components of `MVN` are vectors and/or matrices of length 1, then they are displayed as scalars:

```
MVN[{μ1}, {{var1}}, x]
```

$x \sim \mathcal{N}(\mu_1, \text{var1})$

```
?StandardMVN
```

StandardMVN[n,x] returns an MVN of order n with zero mean and a covariance matrix equal to the identity matrix. The variables are labeled $x_1,\ldots,x_n$. If the second argument is omitted, the variables are labeled $z_1,\ldots,z_n$

```
d1 = StandardMVN[1]
```

$z \sim \mathcal{N}(0, 1)$

```
d2 = StandardMVN[3]
```

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

Get/Set routines (to get the parameters, or to reset them)

```
{GetMean[d1], GetMean[d2] // MatrixForm}
```

$$\left\{0, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right\}$$

```
{GetVariance[d1], GetVariance[d2] // MatrixForm}
```

$$\left\{1, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right\}$$

```
{GetLabel[d1], GetLabel[d2] // MatrixForm}
```

$$\left\{z, \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}\right\}$$

```
SetMean[d1, 123]
```

$z \sim \mathcal{N}(123, 1)$

```
SetMean[d2, {1, 2, 3}]
```

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

```
SetVariance[d1, 56]
```

$$z \sim \mathcal{N}(0, 56)$$

```
SetVariance[d2, {{61, 1, 2}, {1, 100, 0}, {2, 0, 99}}]
```

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 61 & 1 & 2 \\ 1 & 100 & 0 \\ 2 & 0 & 99 \end{pmatrix}\right)$$

```
SetLabel[d1, NewVariable]
```

$$\text{NewVariable} \sim \mathcal{N}(0, 1)$$

```
SetLabel[d2, {Huey, Dewey, Louie}]
```

$$\begin{pmatrix} \text{Huey} \\ \text{Dewey} \\ \text{Louie} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

Routines to construct new variables

```
?LinearForm
```

LinearForm[mvn1,b,c] (where b is a matrix and c is a vector) returns an MVN that describes a linear transformation of mvn1. If mvn1=MVN[μ,Σ,v] then the return MVN is the distribution of b.v+c. LinearForm[mvn1,b] returns an MVN giving the distribution of b.v

```
LinearForm2[d1, {{6}, {7}}]
```

$$\text{LinearForm2}\left[z \sim \mathcal{N}(0, 1), \{\{6\}, \{7\}\}\right]$$

```
b = {{1, 2, 3}, {5, 2, 1}, {4, 1, 0}}; b // MatrixForm
```

$$\begin{pmatrix} 1 & 2 & 3 \\ 5 & 2 & 1 \\ 4 & 1 & 0 \end{pmatrix}$$

```
a = LinearForm[StandardMVN[3], b]
```

$$\begin{pmatrix} z_1 + 2 z_2 + 3 z_3 \\ 5 z_1 + 2 z_2 + z_3 \\ 4 z_1 + z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 14 & 12 & 6 \\ 12 & 30 & 22 \\ 6 & 22 & 17 \end{pmatrix}\right)$$

Often, after we build new variables, we relabel them:

```
b = SetLabel[a, {c, d, e}]
```

$$\begin{pmatrix} c \\ d \\ e \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 14 & 12 & 6 \\ 12 & 30 & 22 \\ 6 & 22 & 17 \end{pmatrix}\right)$$

Sometimes, we'd just like to work with the linear expressions themselves (without having to construct the coefficient matrix directly). `MakeLinearForm` shows how this works.

```
? MakeLinearForm
```

MakeLinearForm[MVN0,b] returns an MVN object representing the distribution of a set of linear combinations of the variables in MVN0=MVN[μ,Σ,v]. b is a list of expressions defining linear combinations of v (and a constant, if desired).

```
MakeLinearForm[MVN[0, 1, x], {2 x + 3, 3 x + 10}]
```

$$\begin{pmatrix} 3 + 2 x \\ 10 + 3 x \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 3 \\ 10 \end{pmatrix}, \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix}\right)$$

(In the above example, since both new variables are linear functions of `x`, the covariance matrix is singular.)

```
MakeLinearForm[StandardMVN[2], {2 z1 + 3, 3 z1 + 2 z2}]
```

$$\begin{pmatrix} 3 + 2 z_1 \\ 3 z_1 + 2 z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 6 \\ 6 & 13 \end{pmatrix}\right)$$

Conditional distributions

A multivariate normal distribution is usually a joint distribution of all the component random variables. One of the most common steps in analysis of an economic model involves constructing one or more conditional distributions. `MVNConditional` does this easily. You just have to specify the conditioning variables, and the target variables.

```
? MVNConditional
```

MVNConditional[MVN0,v1,v2] returns an MVN object representing a conditional multivariate normal distribution. MVN0=MVN0[μ,Σ,v] is the joint distribution. v2 is a list of the conditioning variables; v1 is a list of the target variables. Both v1 and v2 must map to (correspond to) v. MVNConditional[MVN0,v1] returns the conditional distribution of v1 conditioned on all variables that are not in v1.

```
MVNConditional[b, c, {d, e}]
```

$$c \sim \mathcal{N}\left(12\left(\frac{17\,d}{26}-\frac{11\,e}{13}\right)+6\left(-\frac{11\,d}{13}+\frac{15\,e}{13}\right),\ \frac{2}{13}\right)$$

The conditioning set might be a subset of the remaining variables.

```
MVNConditional[b, c, d]
```

$$c \sim \mathcal{N}\left(\frac{2\,d}{5},\ \frac{46}{5}\right)$$

If you omit the second argument, the conditioning variables are all the remaining variables:

```
MVNConditional[b, c]
```

$$c \sim \mathcal{N}\left(12\left(\frac{17\,d}{26}-\frac{11\,e}{13}\right)+6\left(-\frac{11\,d}{13}+\frac{15\,e}{13}\right),\ \frac{2}{13}\right)$$