

# The Multivariate Normal Distribution

Paul Johnson

June 10, 2013

## 1 Introduction

A one dimensional Normal variable should be very familiar to students who have completed one course in statistics.

The multivariate Normal is a generalization of the Normal which describes the probability of observing  $n$  – *tuples*.

The multivariate Normal distribution looks just like you might expect from your experience with the univariate Normal: symmetric, single-peaked, basically pleasant in every way!

The vectors that are described by the multivariate Normal distribution can have many elements as we choose to assign. Let's work with an example in which each randomly selected case has two attributes: a two dimensional model. A point drawn from a two dimensional Normal would be a pair of real-valued numbers (you can call it a vector or a point if you like).

Given any vector, such as

$$(0.2432, 0.1211)$$

or

$$(-2.2432, 0.1211)$$

the multivariate Normal density function gives back an indicator of how likely we are to draw that particular pair of values.

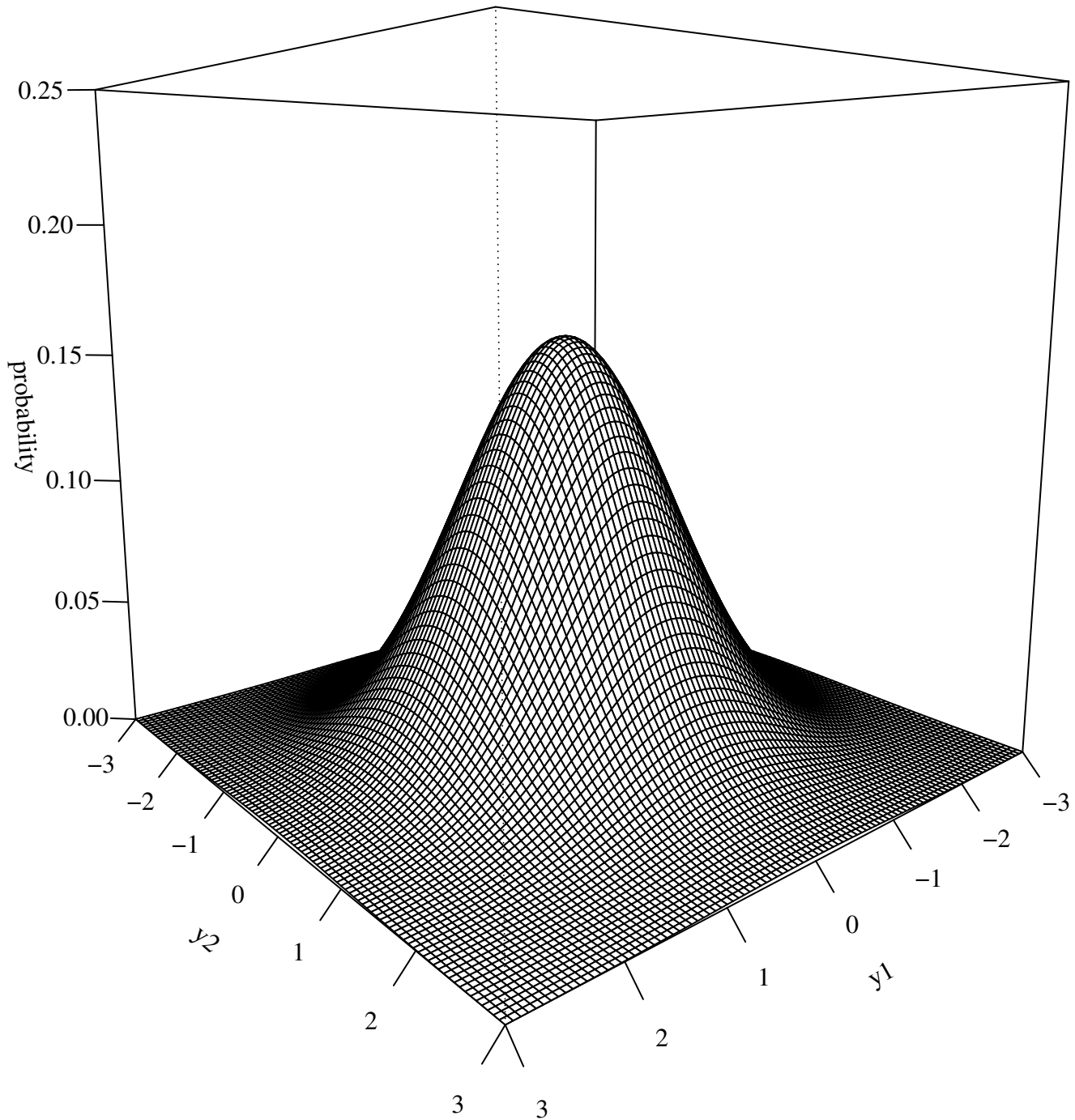
The Multi Variate Normal (MVN) distribution has two parameters, but each of these two parameters contains more than one piece of information. One parameter,  $\mu$ , describes the “location” or “center point” of the distribution, while the other parameter determines the dispersion, the width of the spread from side to side in the observed points.

Figure 1 illustrates the density function of a Normal distribution on two dimensions.

## 2 Mathematical Description

In the multivariate Normal density, there are two parameters,  $\mu$  and  $\Sigma$ . The first is an  $n \times 1$  column vector of “location parameters” and an  $n \times n$  “dispersion matrix”  $\Sigma$ .

Figure 1: Multi Variate Normal Distribution



$$\begin{aligned}
f(y_i|\mu, \Sigma) &= \\
&= f((y_{i1}, y_{i2}, \dots, y_{in})' | \mu = (\mu_1, \dots, \mu_n)', \Sigma = \begin{bmatrix} \sigma_1^2 & & \sigma_{1n} \\ & \ddots & \\ \sigma_{n1} & & \sigma_n^2 \end{bmatrix}) \\
&= \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(y_i - \mu)' \Sigma^{-1} (y_i - \mu)\right) \tag{1}
\end{aligned}$$

The symbol  $|\Sigma|$  refers to the determinant of the matrix  $\Sigma$ . The determinant is a single real number. The symbol  $\Sigma^{-1}$  is the inverse of  $\Sigma$ , a matrix for which  $\Sigma\Sigma^{-1} = I$ . This equation assumes that  $\Sigma$  can be inverted, and one sufficient condition for the existence of an inverse is that the determinant is not 0.

The matrix  $\Sigma$  must be positive semi-definite in order to assure that the most likely point is  $\mu = (\mu_1, \mu_2, \dots, \mu_n)$  and that, as  $y_i$  moves away from  $\mu$  in any direction, then the probability of observing  $y_i$  declines.

The denominator in the formula for the MVN is a normalizing constant, one which assures us that the distribution integrates to 1.0.

In the last section of this writeup, I have included a little bit to help with the matrix algebra and the importance of a “weighted cross product” like  $(y_i - \mu)\Sigma^{-1}(y_i - \mu)$ .

If there is only one dimension, this gives the same value as the ordinary Normal distribution, which one will recall is

$$f(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right]$$

In the one dimensional case,  $\Sigma = \sigma^2$ , meaning  $|\Sigma| = \sigma^2$  and  $\Sigma^{-1} = 1/\sigma^2$ .

In the example with two dimensions, the location parameter has two components

$$\mu = (\mu_1, \mu_2)$$

and (again for the two dimensional case) the scale parameter is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

For the  $2 \times 2$  case, the determinant and inverse are known:

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 \quad \text{and} \quad \Sigma^{-1} = \begin{bmatrix} \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} & \frac{-\sigma_{12}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \\ \frac{-\sigma_{12}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} & \frac{\sigma_1^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \end{bmatrix}$$

so you can actually calculate an explicit value “by hand” if you have enough patience.

$$\begin{aligned}
f(y_i) &= (y_1, y_2)' | \mu, \Sigma) \\
&= \frac{1}{\sqrt{(2\pi)^n \sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} \exp\left(-\frac{1}{2} [y_{i1} - \mu_1, y_{i2} - \mu_2] \begin{bmatrix} \frac{\sigma_2^2}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} & \frac{-\sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \\ \frac{-\sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} & \frac{\sigma_1^2}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \end{bmatrix} \begin{bmatrix} y_{i1} - \mu_1 \\ y_{i2} - \mu_2 \end{bmatrix}\right)
\end{aligned}$$

At this time, I'm unwilling to write out any more of that. But I encourage you to do it.

### 3 The Expected Value and Variance

Recall that expected value is defined as a “probability weighted sum of possible outcomes”.

$$E(y) = \mu$$

$$Var(y) = \Sigma$$

Because we are considering a multivariate distribution here, it is probably lost in the shuffle that the expected value is calculated by integrating over all dimensions.

If we had a discrete distribution, say  $p(x_1, x_2)$ , then the expected value would be calculated by looping over all values of  $x_1$  and  $x_2$  and calculating a probability weighted sum

$$E(x_1) = \sum_{i=1}^n \sum_{j=1}^m p(x_{1i}, x_{2j}) x_{1i}$$

With a probability model defined on continuous variables, the calculation of the expected value requires the use of integrals, but except for the change from notation ( $\Sigma$  to  $f$ ), there is not major conceptual change.

$$E(y_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y_1, y_2) y_1$$

Because the covariance between two variables is the same, whether we calculate  $\sigma_{12} = Covariance(y_1, y_2)$  or  $\sigma_{21} = Covariance(y_2, y_1)$ , the matrix  $\Sigma$  has to be same on either side of the diagonal.

As you can see, the matrix is symmetrical, so that the bottom-left element is the same as the top-right. The matrix is the same as its transpose.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \Sigma'$$

## 4 Illustrations

### 4.1 Contour Plots

The parameters that were used to create Figure 1 were  $\mu = (0,0)$  and  $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . That is the bivariate equivalent of the “standardized normal distribution.” A contour plot offers one way to view this distribution. The equally-probable points are connected by lines. The idea will be familiar to economists as “indifference curves” and to geographers as “topographical maps”. Consider the contours in Figure 2.

### 4.2 Change the Variance

If we change one of the diagonal elements in  $\Sigma$ , the effect is to “spread out” the observations on that dimension. See Figure 3.

### 4.3 The effect of the Covariance parameter

Until this point, we have been considering distributions in which the two variables  $y_1$  and  $y_2$  were statistically independent, in the sense that the expected value of  $y_1$  was not influenced by  $y_2$ .

## 5 “Cross-sections”

A conditional density gives the probability of the outcome  $y = (y_1, y_2)$  variable when one of these is taken as given. For example,

$$f((y_1, y_2) | y_2 = 0.75)$$

represents the probability of observing  $y_1$  when it is known that  $y_2 = 0.75$ .

The Multivariate Normal distribution has the interesting property that the conditional distributions are basically Normal (Normal up to a constant of proportionality). Figure 6 illustrates the MVN when  $y_2 = 0.75$ . The “total area” under that curve will not equal 1.0, but if we normalize it by finding the area (call that  $A$ ), then

$$\frac{f(y_1 | y_2 = 0.75)}{A}$$

is a valid probability density function.

## 6 Appendix: Matrix Algebra.

If you have never studied matrix algebra, I would urge you to do a little study right now. Find any book or one of my handouts.

Figure 2: The View from Above Corresponding to Figure 1

```
#image(x1, x2, probX, col = terrain.colors(100), axes = FALSE)  
contour(x1, x2, probX)
```

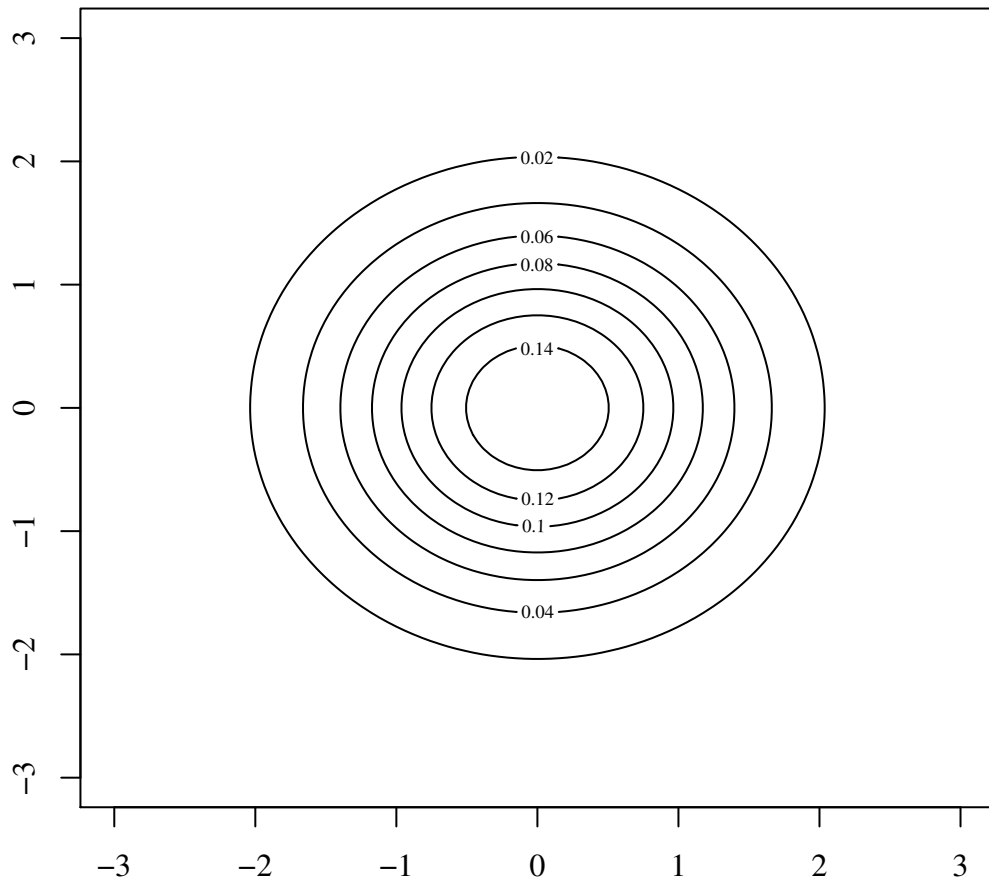


Figure 3: MV Normal( $\mu = (0,0), \sigma^2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ )

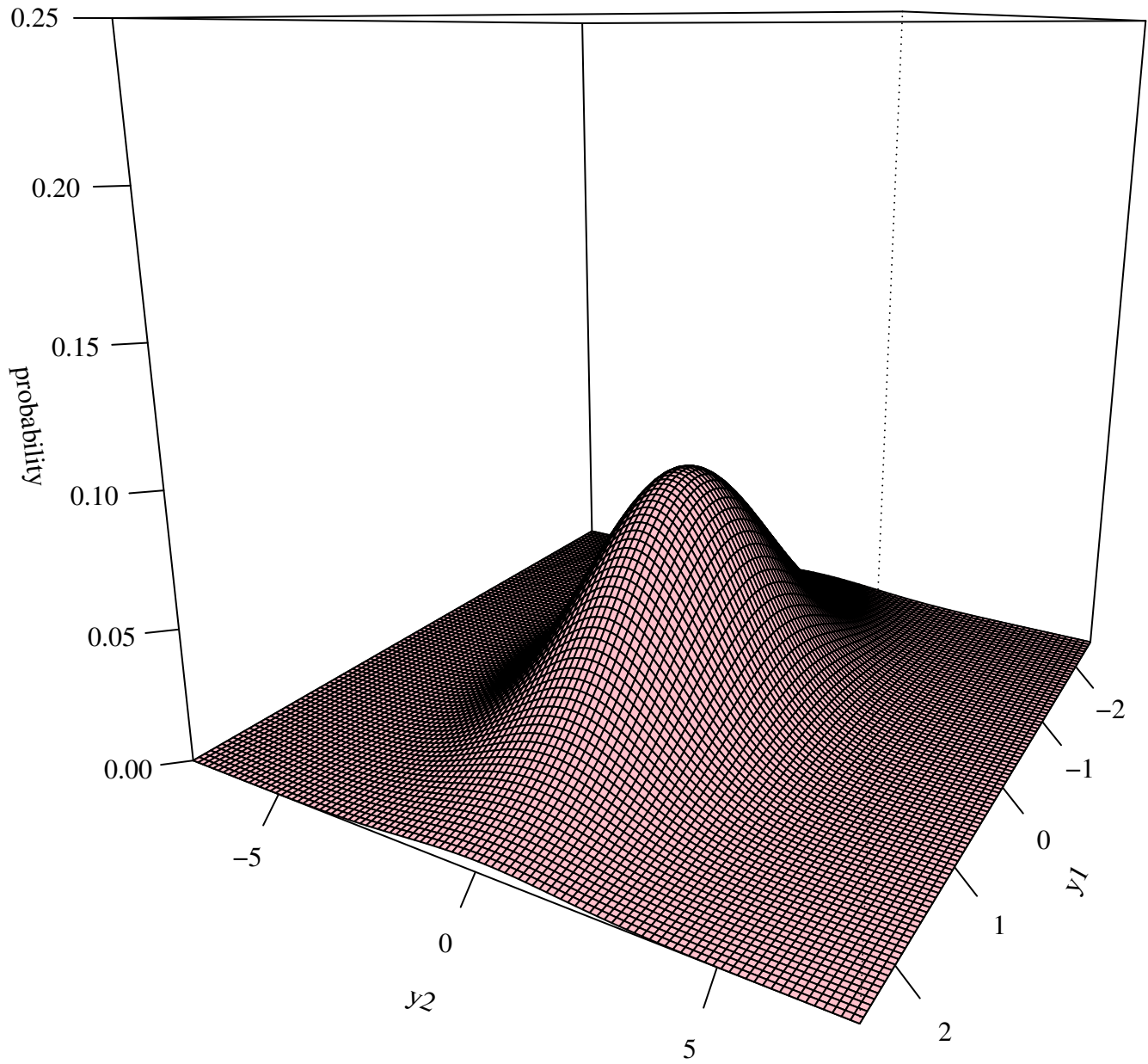


Figure 4: MV Normal( $\mu = (0, 0), \sigma^2 = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$ )

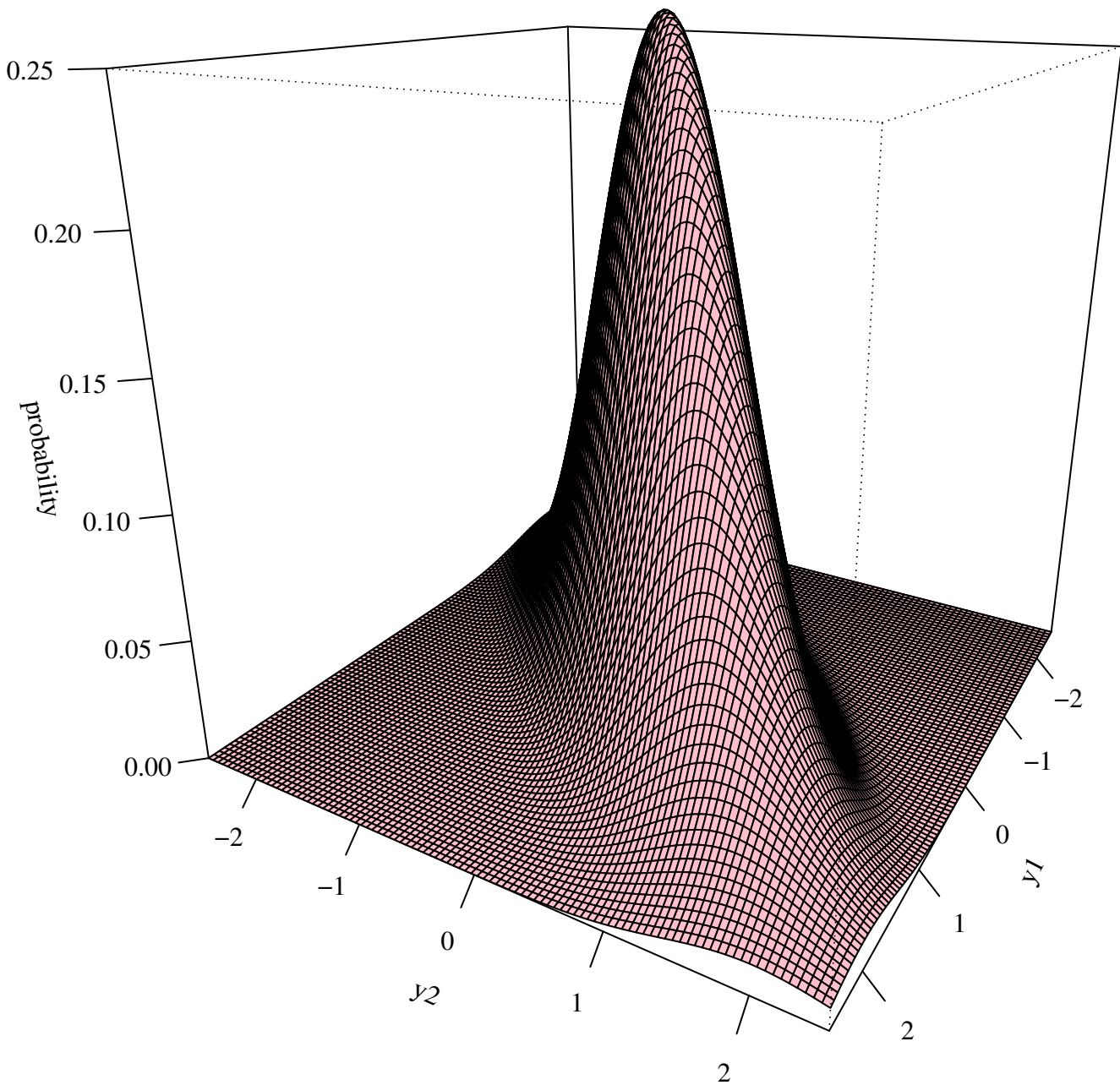




Figure 5: The View from Above

```
#image(x1, x2, probX, col = terrain.colors(100), axes = FALSE)  
contour(x1, x2, probX)
```

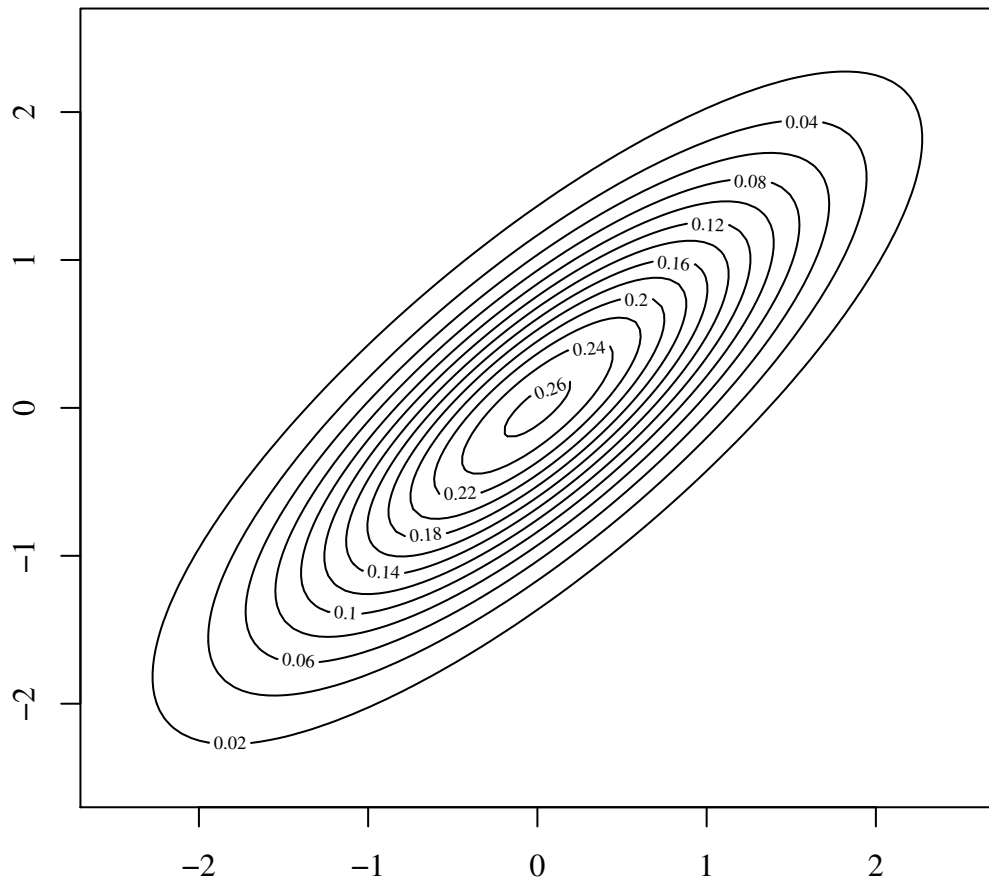
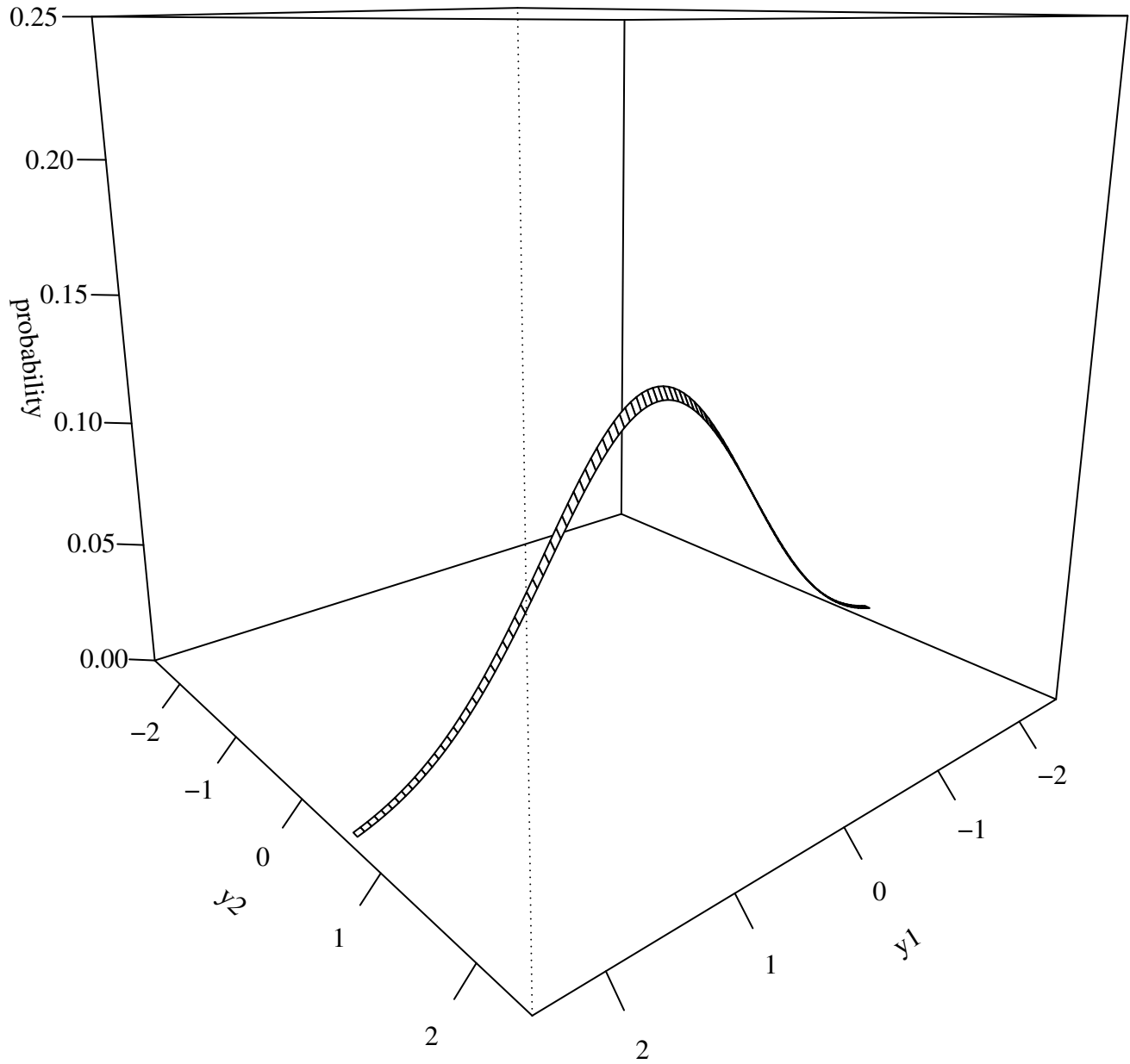


Figure 6: Cross section of Standard MVN for  $y_2 = 0.75$



Your objective is to understand enough about multiplication of matrices and vectors so that you can see the meaning of an expression like the following.

$$x'Mx$$

$x$  is an  $n \times 1$  column vector and  $M$  is an  $n \times n$  matrix. The result is a single number, and if you write it out term-for-term, you find this a “weighted sum of  $x$ ’s squares and cross-products.”

$$\begin{aligned} & (x_1, x_2, \dots, x_n) \begin{bmatrix} m_{11} & & m_{1n} \\ & \ddots & \\ m_{n1} & & m_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= (x_1, x_2, \dots, x_n) \begin{bmatrix} m_{11}x_1 + m_{12}x_2 + \dots + m_{1n}x_n \\ m_{21}x_1 + m_{22}x_2 + \dots + m_{2n}x_n \\ \vdots \\ m_{n1}x_1 + m_{n2}x_2 + \dots + m_{nn}x_n \end{bmatrix} \\ &= m_{11}x_1x_1 + m_{12}x_2x_1 + \dots + m_{1n}x_nx_1 + \\ & \quad m_{21}x_1x_2 + m_{22}x_2x_2 + \dots + m_{2n}x_nx_2 + \\ & \quad \dots + \\ & \quad m_{n1}x_1x_n + m_{n2}x_2x_n + \dots + m_{nn}x_nx_n \\ &= \sum_{i=1}^n \left( \sum_{j=1}^n m_{ij}x_ix_j \right) \end{aligned}$$

If  $M$  is diagonal, meaning

$$M = \begin{bmatrix} m_{11} & 0 & 0 & 0 \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & m_{nn} \end{bmatrix} \tag{2}$$

then there are no “cross products” and the result of  $x'Mx$  is simply a sum of weighted  $x$ -squared terms.

$$= \sum_{i=1}^n m_{ii}x_i^2 \tag{3}$$

A matrix is positive semi-definite if, for all  $x$ ,

$$x'Mx \geq 0 \tag{4}$$

If  $M$  is positive semi-definite, so is  $M^{-1}$ .

In the Normal distribution's density formula, we have:

$$\exp\left(-\frac{1}{2}(y_i - \mu)' \Sigma^{-1}(y_i - \mu)\right) \quad (5)$$

In that formula, the role of  $x$  is played by  $(y_i - \mu)$  and the role of  $M$  is played by  $\Sigma^{-1}$ .

The Normal distribution requires that  $\Sigma$  is positive semi-definite.

$$-\frac{1}{2}(y_i - \mu)' \Sigma^{-1}(y_i - \mu) > 0 \quad (6)$$

Tidbit, The Eigenvectors  $(e_1, e_2, \dots, e_n)$  of  $\Sigma$  are known as its “principal components”. The principal components give the axes of symmetry of the MVN. If the  $i$ 'th eigenvalue (the one associated with the  $i$ 'th eigenvector) is  $\lambda_i$ , the length of the ellipsoid on the  $i$ th axis is proportional to  $\sqrt{\lambda_i}$  (B. Walsh and M. Lynch, Appendix 2, The Multivariate Normal, 2002, [http://nitro.biosci.arizona.edu/zbook/volume\\_2/chapters/vol2\\_A2.html](http://nitro.biosci.arizona.edu/zbook/volume_2/chapters/vol2_A2.html))