Multivariate Normal Distribution

Module I (Lecture 4)

David Raj Micheal

August 2018

Contents

A random variable X is normally distributed with mean μ and variance σ^2 if it has the probability density function of X as:

$$f(x) = \frac{1}{\sqrt{2\pi} \times \sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}.$$

As shorhand notation we write

$$X \sim N(\mu, \sigma^2)$$

indicating that X is distributed according to (denoted by the wavey symbol 'tilde') a normal distribution (denoted by N), with mean μ and variance σ^2.

If we have a random vector x with p random variables that is distributed according to a multivariate normal distribution with population mean vector μ and population variance-covariance matrix Σ, then this random vector, x, will have the joint density function as shown in the expression below:

$$f(x_1, \ldots, x_p) = \frac{1}{\left(\sqrt{2\pi}\right)^p \times |\Sigma|^{-1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu) \right\},$$

where $|\Sigma|$ denotes the determinant of the variance-covariance matrix Σ and Σ^{-1} is the inverse of the variance-covariance matrix Σ. We denote x follows multivariate normal distribution with mean μ and variance-covariance matrix Σ by,

$$x \sim N(\mu, \Sigma).$$

1 Bivariate Normal Distribution

Before discussing the multivariate normal distribution we discuss the bivariate normal distribution to understand the multivariate normal distribution further. Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ follows bivariate normal distribution with mean $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and variance-covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. Note that, $\sigma_{12} = \sigma_1 \sigma_2 \rho_{12}$ by the definition of correlation and hence the variance covariance matrix can be written as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix}.$$

Hence the determinant of Σ is

$$|\Sigma| = \sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2).$$

and the inverse of Σ is

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho_{12}^2)} \begin{pmatrix} \sigma_2^2 & -\rho_{12}\sigma_1\sigma_2 \\ -\rho_{12}\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Substituting in the expressions for the determinant and the inverse of the variance-covariance matrix we obtain, after some simplification, the joint probability density function of $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ for the bivariate normal distribution as shown below:

$$f(x_1, x_2) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}.$$

standard multivariate normal distribution: A standard multivariate normal distribution is defined to have mean equal to the zero vector, and variance-covariance matrix equal to the identity matrix.

The shape of a standard multivariate normal is like a nicely rounded hill(3 dimensioal bell shape). Generally, when there is correlation between measurements the hill becomes elongated in the direction of the correlation. Bivariate Normal desity plots are given in Figure 2 for different correlation values.

The multivariate normal density function has contours of equal density that are elliptical. See Figure 3 for the contours of the bivariate normal density plots shown in Figure 2.

```
bivar_norm = function(x,y,mu=c(0,0),sigma=c(3,3),rho=0,Pi = 3.142){
  exp((-1/(2*(1-rho^2)))*
        (((x-mu[1])/sigma[1])^2-2*rho*(x-mu[1])*(y-mu[2])/(sigma[1]*sigma[2])+
           ((y-mu[2])/sigma[2])^2))/(2*Pi*sigma[1]*sigma[2]*sqrt(1-rho^2));
}
set.seed(10100)
dat = sort(rnorm(1000, 0, 1))
  x =  y = seq(-10,10,length=100)
  z = outer(x,y,bivar_norm)

  par(mfrow=c(1,2))
  plot(dat, dnorm(dat),
       type = 'l',
       lwd = 2,
       col ='deeppink3',
       xlab = "x",
       ylab = "Univariate Normal Density" )
    # plot(x,dnorm(x),ylab = "Density", lwd = 2, type = 's' )
   persp(x,y,z,
        theta = 30,
        phi = 15,
        ticktype = 'detailed',
        expand = .75,
        shade = .2,
        col = 'red2',zlab = "Bivariate Normal Density" )
```

```
library(MASS)
library(ks)
  n = 2000; m1 = 0; m2 = 0; s1 =1; s2 = 1;
  plotdensity = function(r){
    x1 = rnorm(n,m1,s1)
    x2 = s2*r*(x-m1)/s1+m2 + s2*rnorm(n,0,sqrt(1-r^2))
    cor = kde(x=cbind('x1' =x1,'x2' = x2))
    plot(cor, display="persp", thin=3, border=1, col="red4",
```

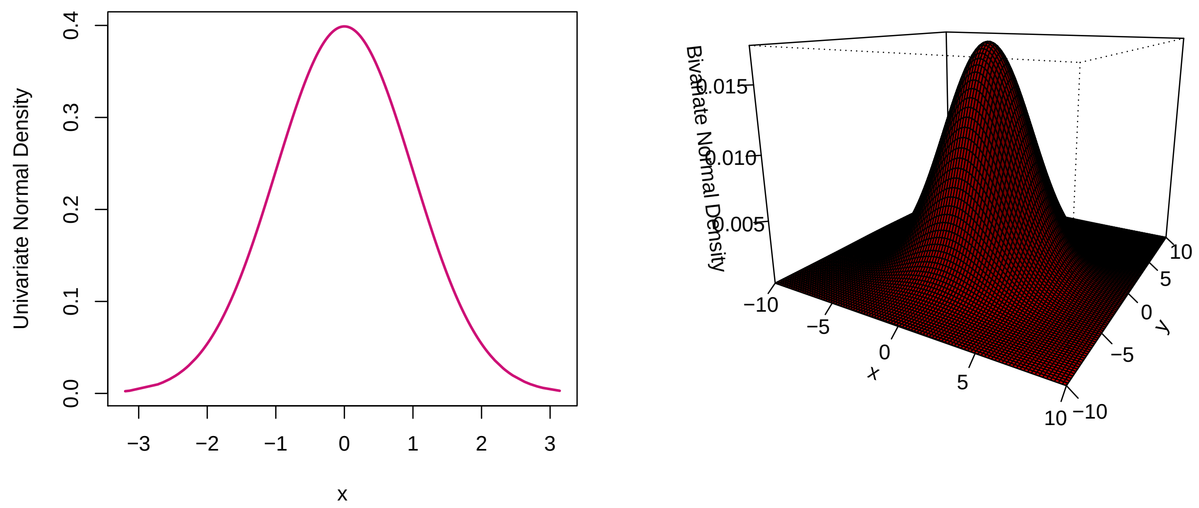

Figure 1: Normal Desity Plots

```
        main = paste("Correlation = ", r))
    }
  par(mfrow=c(3,3),mar=c(1,1,1,1),oma=c(0,0,2,0),new=TRUE)
  for (i in c(-1,-.75,-.5,-0.25,0,0.25,0.5,.75,1)){
    plotdensity(i)
  }
```

```
library(MASS)
library(ks)
  n = 2000; m1 = 0; m2 = 0; s1 =1; s2 = 1;
  plotdensity = function(r){
    x1 = rnorm(n,m1,s1)
    x2 = s2*r*(x-m1)/s1+m2 + s2*rnorm(n,0,sqrt(1-r^2))
    cor = kde(x=cbind('x1' =x1,'x2' = x2))
    plot(cor,
         display="slice",
         thin=3,
         border=1,
         col="red4",
         main = paste("Correlation = ", r))
    }
  par(mfrow=c(3,3),mar=c(1,1,1,1),oma=c(0,0,2,0),new=TRUE)
  for (i in c(-1,-.75,-.5,-0.25,0,0.25,0.5,.75,1)){
    plotdensity(i)
  }
```

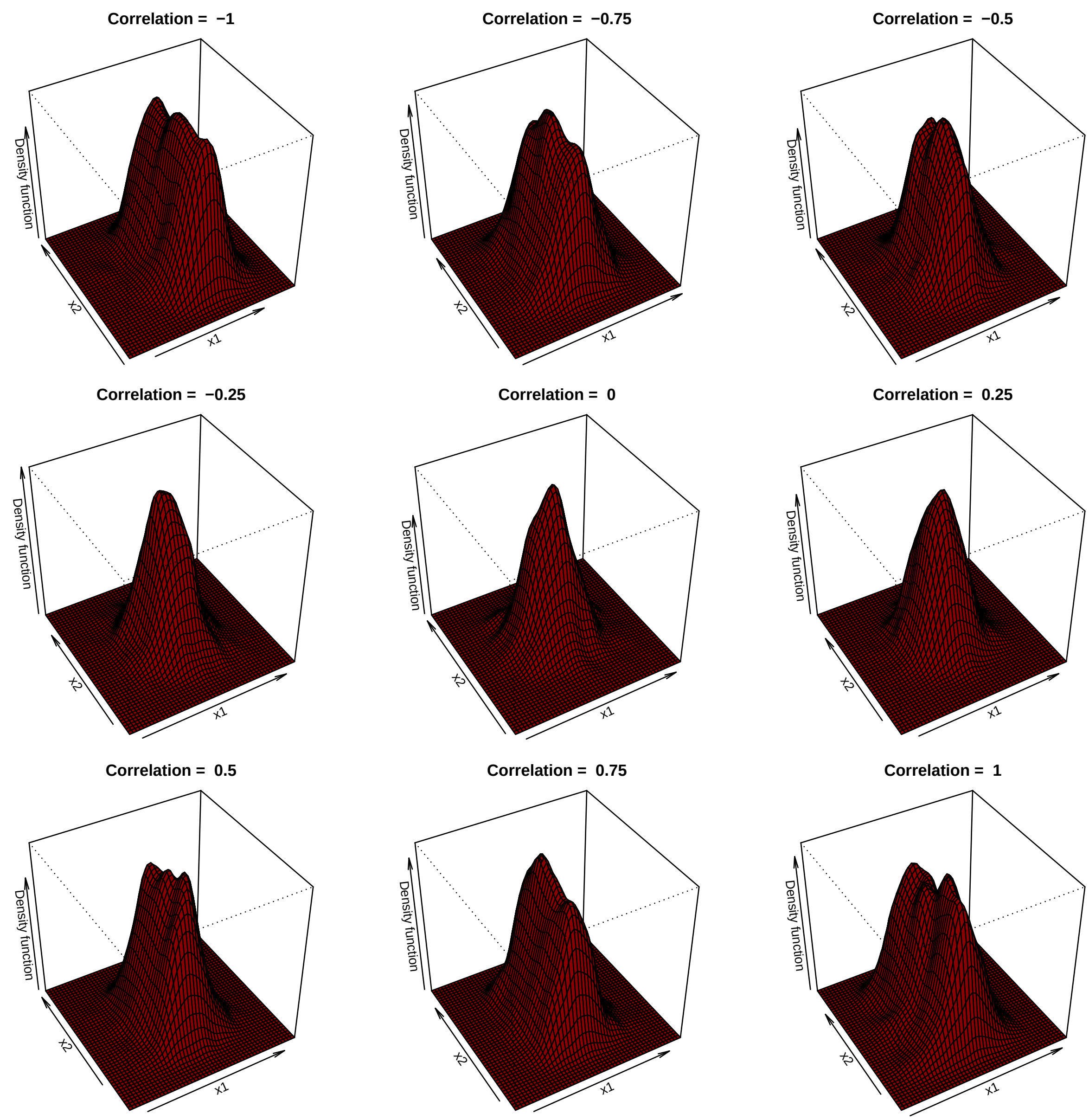

Figure 2: Plots for different correlation values

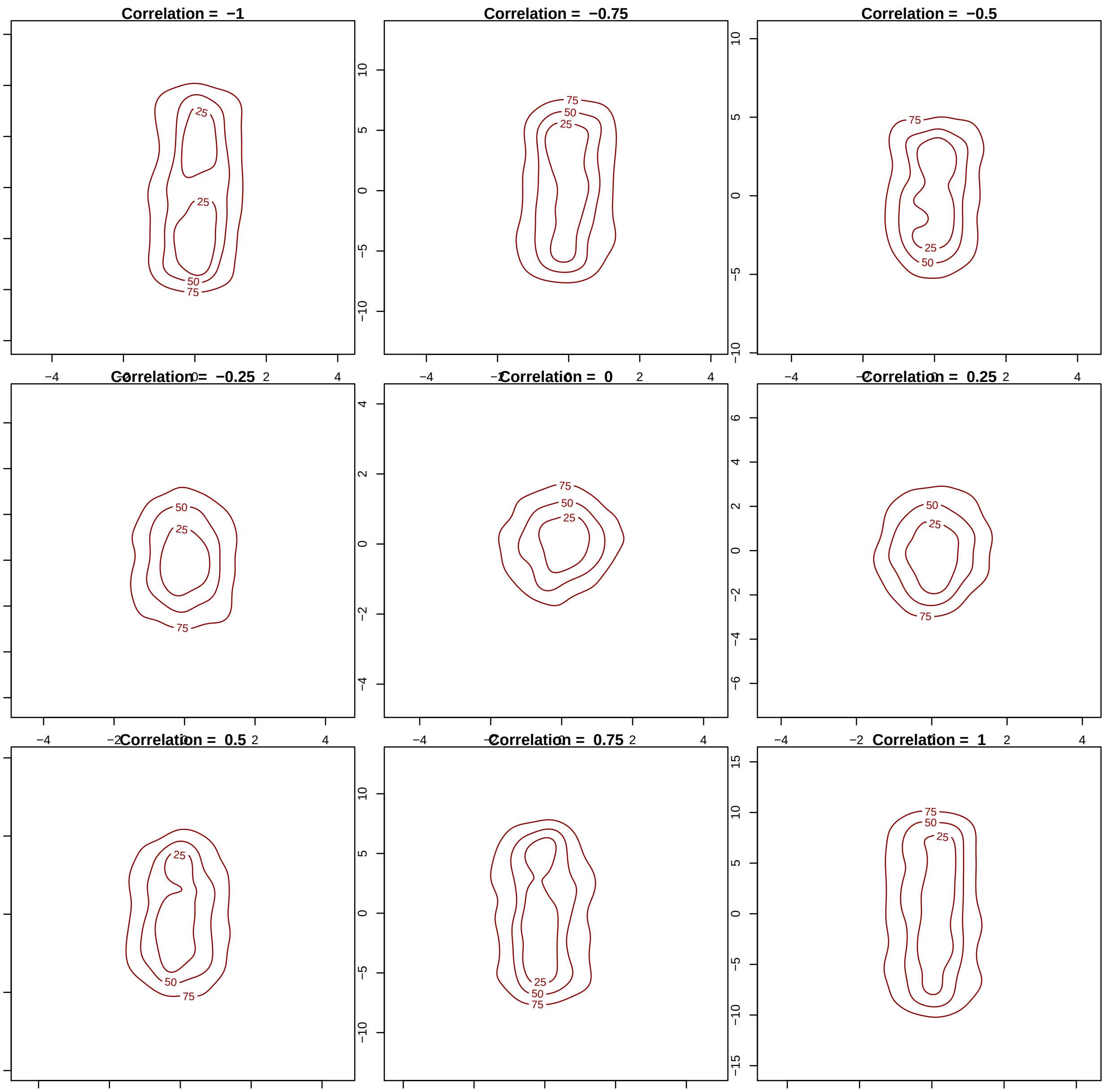

Figure 3: Contour plots for different correlation values

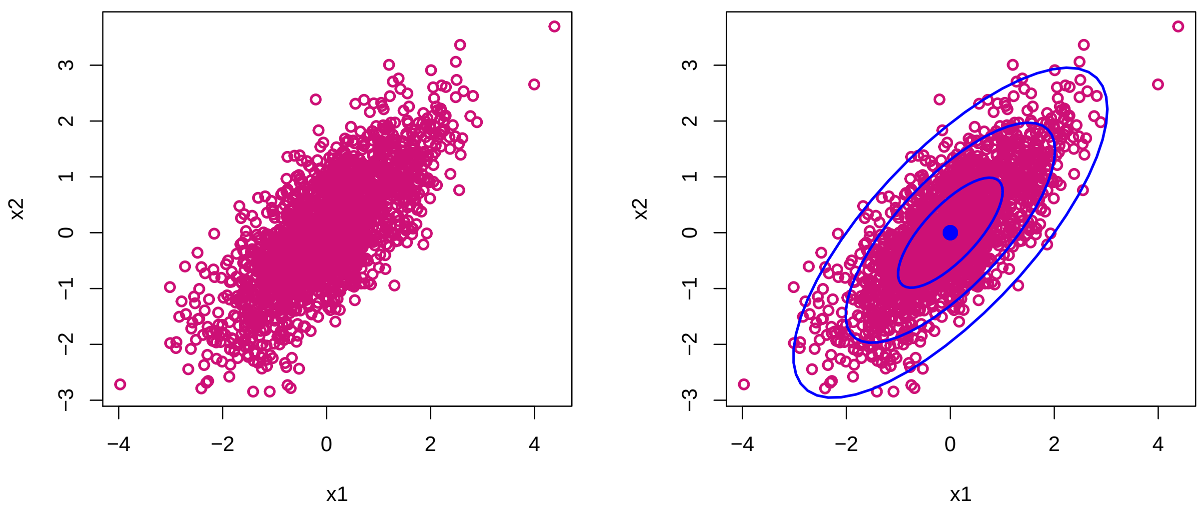

Figure 4: Confidence Ellipse

2 Mahalanobis Distance

Note that the multivariate density functions

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

only depends through the exponent term $(x-\mu)^T\Sigma^{-1}(x-\mu)$, which is the equation for a hyper-ellipse centered at μ. For a bivariate normal, where $p = 2$ variables, we have an ellipse shown in the Figure 4.

```
library(car)
```

```
## Loading required package: carData
```

```
n = 2000; m1 = 0; m2 = 0; s1 =1; s2 = 1; r =0.75;
x1 = rnorm(n,m1,s1)
x2 = s2*r*(x1-m1)/s1+m2 + s2*rnorm(n,0,sqrt(1-r^2))
par(mfrow=c(1,2))
  plot(x1,x2,  lwd = 2, col ='deeppink3', xlab = "x1", ylab = "x2" )
  plot(x1,x2,  lwd = 2, col ='deeppink3', xlab = "x1", ylab = "x2" )
  ellipse(center = c(m1,m2), shape = cov(cbind(x1,x2)),radius = 3, lwd = 2)
  ellipse(center = c(m1,m2), shape = cov(cbind(x1,x2)),radius = 2, lwd = 2)
  ellipse(center = c(m1,m2), shape = cov(cbind(x1,x2)),radius = 1, lwd = 2)
```

The term $\left(\frac{x-\mu}{\sigma}\right)^2$ in the exponent of the univariate normal density measures the squared distance from x and μ in standard deviation units. Similarly, the term $(x-\mu)^T\Sigma^{-1}(x-\mu)$ in the exponent of the multivariate normal density is the squared generalized distance from x to μ. This distance is called as *Mahalanobis distance*. Formally, for the ith observation $x = (x_{i1}, x_{i2}, \ldots, x_{ip})$ on the p variables the *Mahalanobis distance* (or statistical distance) from the mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ is defined as

$$\Delta^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$$

where Σ is the variance-covariance matrix.

- Note that Δ, the square root of Δ^2 in the above equation, is not in standard deviation units as is $\frac{x-\mu}{\sigma}$.
- The most common use for the Mahalanobis distance is to find multivariate outliers, which indicates unusual combinations of two or more variables.

3 Generalized Variance

The generalized population variance of a random vector of p random rvariables x is defined as the determinant of its variance-covariance matrix Σ. Analogously, the generalized sample variance a random vector of p random rvariables x is defined as the determinant of its sample variance-covariance matrix S.

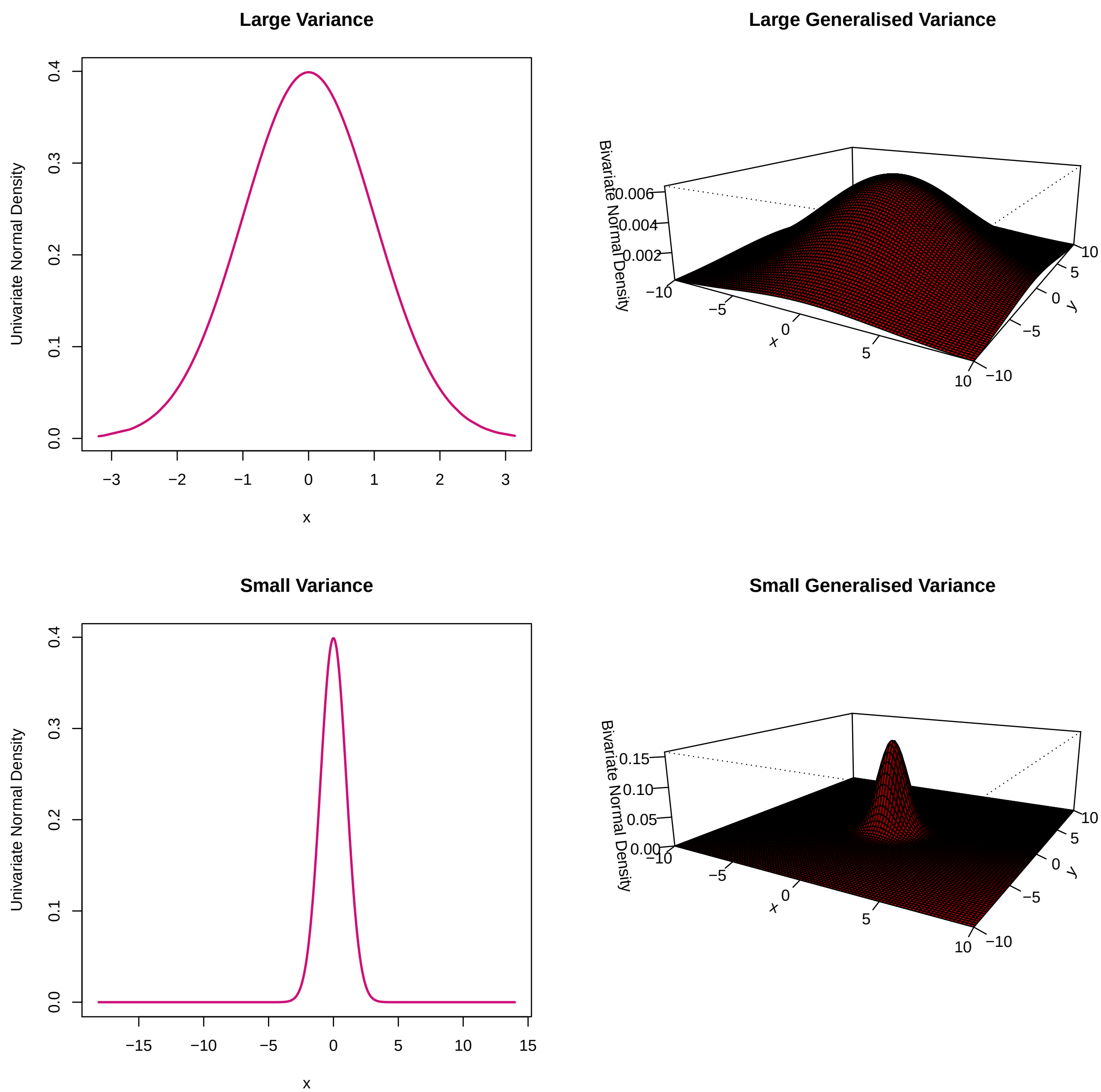

Figure 5: Generalized Sample Variances and Densities