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.

```
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)
m = c(0,0)
library(ks)
set.seed(2000)
  plotdensity = function(r){
    sd = c(sqrt(2), sqrt(2))
    corMat = matrix(c(1,r,r,1), ncol = 2, byrow = T)
```

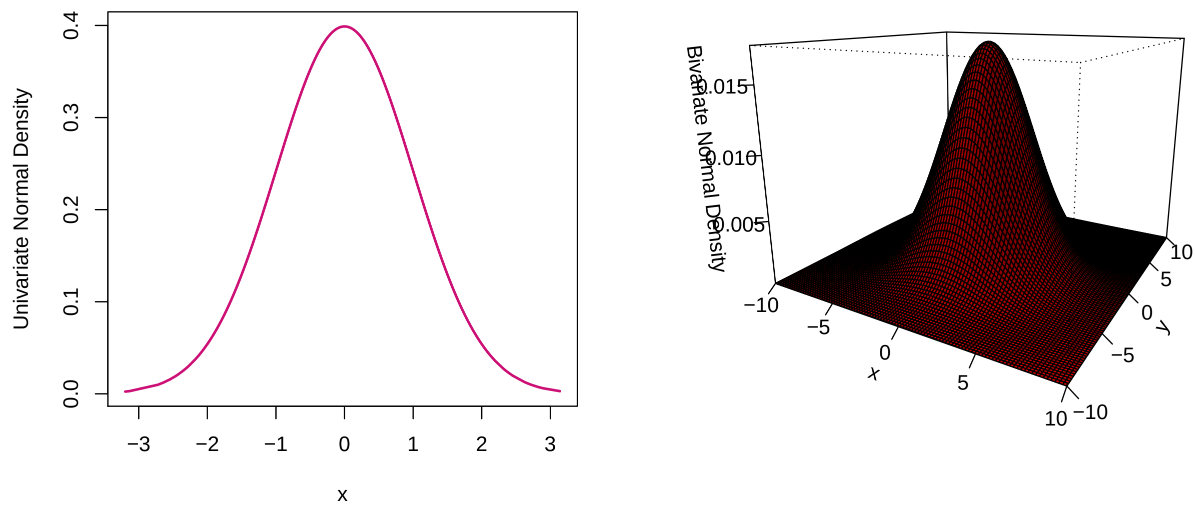

Figure 1: Normal Desity Plots

```
    S = sd%*%t(sd)*corMat
    data =mvrnorm(n = 2000,
           mu = m,
           Sigma = S,
           tol = 1e-6,
           empirical = T)
    cor = kde(data)
    plot(cor, display="persp", 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(-.9,-.75,-.5,-0.25,0,0.25,0.5,.75,.9)){
    plotdensity(i)
  }
```

1.1 Constant probability density contour

From the multivariate normal density function of a $p-$dimensional normal variable, it should be clear that the $x's$ having a constant height for the density are ellipsoids (elliptical in the case of $p = 2$). That is, the multivariate normal density is constant on surfaces where the squared of the distance $(x - \mu)^T \Sigma^{-1}(x - \mu)$ is constant. The collection of those points are said to trace a *contour*:

$$\begin{aligned}\text{Constant probability density contour} &= \{x : (x - \mu)^T \Sigma^{-1}(x - \mu) = c^2\} \\ &= \text{Surface of an ellipsoid centered at } \mu.\end{aligned}$$

See Figure 3 for the contours of the bivariate normal density plots shown in Figure 2.

```
library(MASS)
library(ks)
```

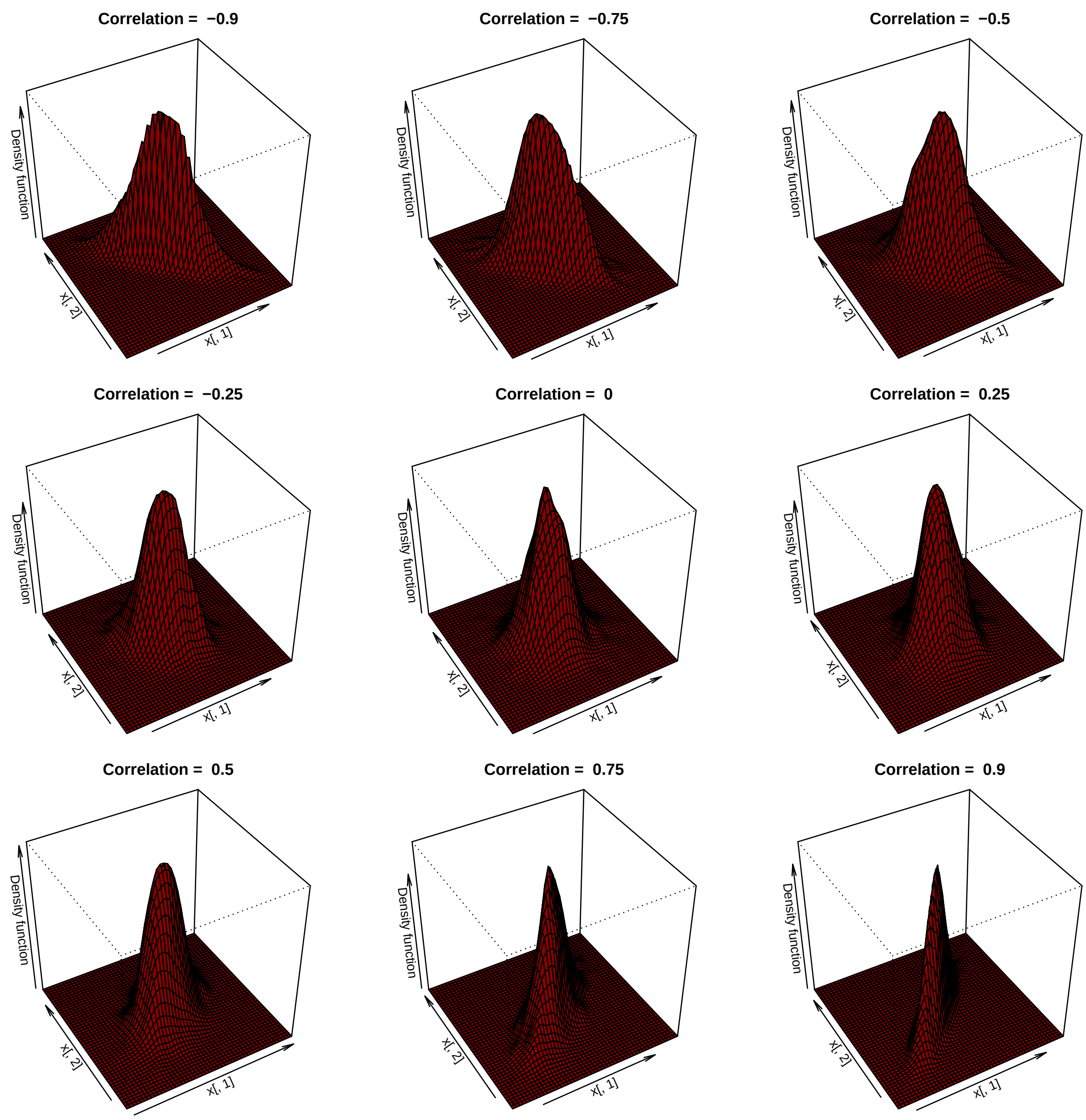

Figure 2: Plots for different correlation values

```
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)
}
```

Remarks

- The axes of the each ellipsoid of constant density are in the direction of the eigenvectors of Σ^{-1}, and their lengths are proportional to the reciprocals of the square toots of the eigenvalues of Σ^{-1}.
- Notethat, we can avoid the calculation of Σ^{-1} when determining the axes, since these ellipsoids are also determined by the eigenvalues and eigenvectors of Σ. We state this formally for later reference.

Result If Σ is positive definite, so that Σ^{-1} exists, then

$$\Sigma x = \lambda x \text{ implies } \Sigma^{-1}x = \frac{1}{\lambda}x$$

where λ is an eigenvalue and x is the corresponding eigenvector.

Question: What is contours of constant density for the $p-$dimensional normal distribution?

Contours of constant density for the $p-$dimensional normal distribution are ellipsoid defined by x such that

$$(x-\mu)^T\Sigma(x-\mu) = c^2$$

where these ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}e_i$, where $\Sigma e_i = \lambda_i e_i$ for $i = 1, 2, \ldots, p$.

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 ellispe 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))
```

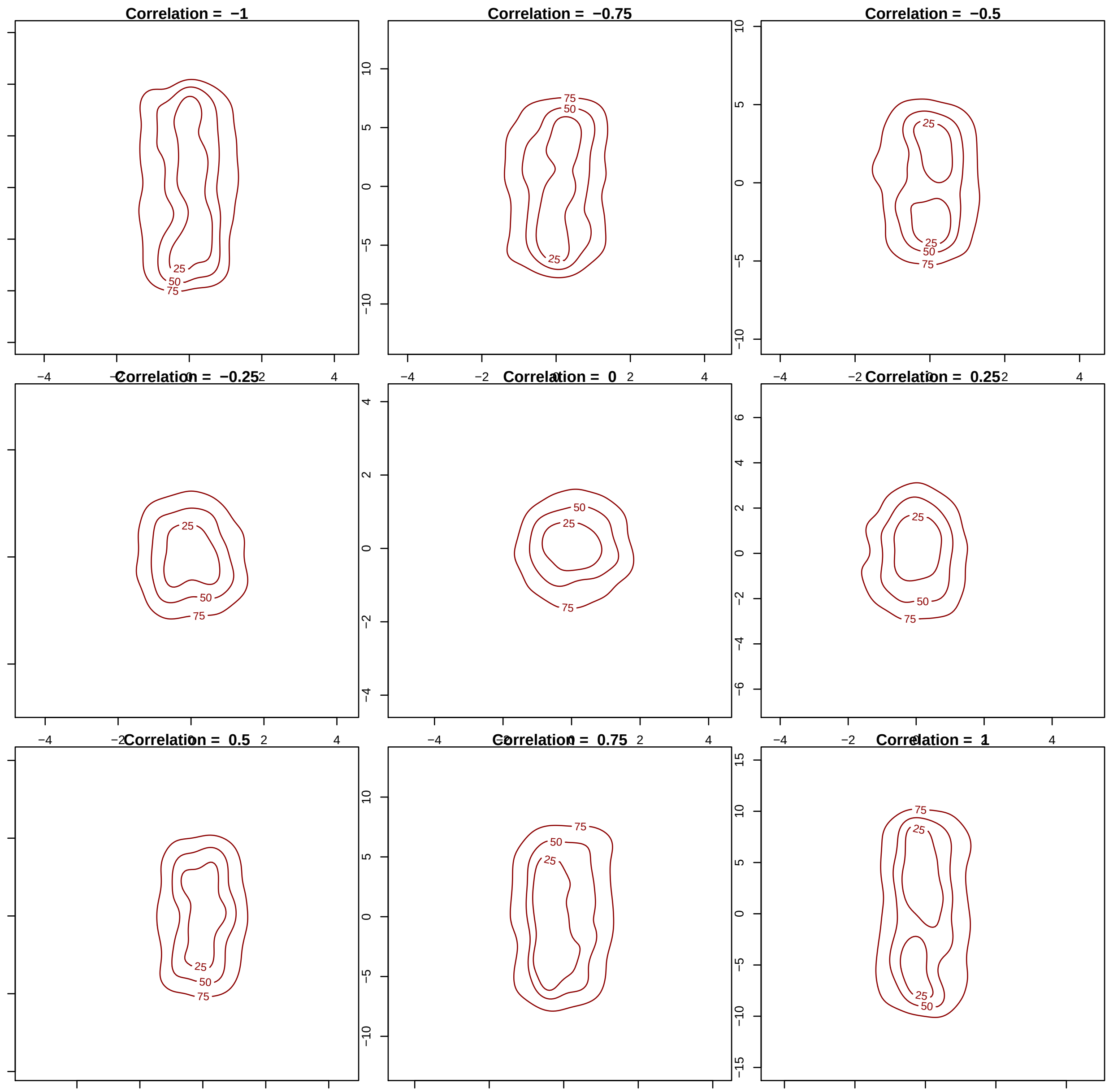

Figure 3: Contour plots for different correlation values

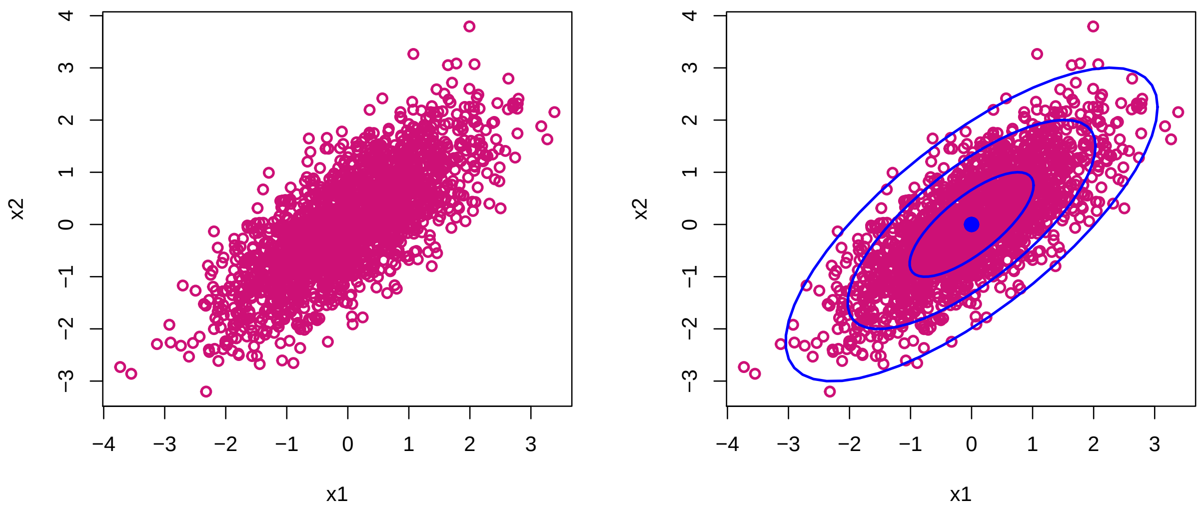

Figure 4: Constant density Contours

```
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 $x = (x_{i1}, x_{i2}, \dots, x_{ip})$ on p variables the *Mahalanobis distance* (or statistical distance) from the mean vector $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ 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.
- If $x \sim N_p(\mu, \Sigma)$ with $|\Sigma| > 0$, then

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

  where χ^2_p denotest eh chi-square distribution with p degrees of freedom.
- $(1-\alpha) \times 100\%$ **Confidence Region:** If $x \sim N_p(\mu, \Sigma)$ with $|\Sigma| > 0$, then

  $$P\left\{x : (x-\mu)^T\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)\right\} = 1-\alpha$$

  where $\chi^2_p(\alpha)$ denotes the upper (100α)th percentile of the χ^2_p distribution. The region

  $$\left\{x : (x-\mu)^T\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)\right\}$$

is called as $(1-\alpha) \times 100\%$ *Confidence Region.*

For example, 95% confidence region for bivariate normally distributed data is given by

$$\left\{x : (x-\mu)^T \Sigma^{-1}(x-\mu) \leq \chi^2_2(0.5)\right\}.$$

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.

Similar to generalized variance, one can define the generalized variance of the standardized variance as determinant of correlation matrix. That is,

$$\text{Generalized population variance of the standardized variables} = \det(P_\rho) = \rho_{11} + \rho_{22} + \cdots + \rho_{pp}$$

and

$$\text{Generalized sample variance of the standardized variables} = \det(R) = r_{11} + r_{22} + \cdots + r_{pp}.$$

Relationship between the $|S|$ and $|R|$:

The quantities $|S|$ and $|R$ are connected by the relationship

$$|S| = s_{11}s_{22}\ldots s_{pp}|R|.$$

Example:

Let $S = \begin{bmatrix} 4 & 3 & 1 \\ 3 & 9 & 2 \\ 1 & 2 & 1 \end{bmatrix}$. Then $R = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{2}{3} \\ \frac{1}{2} & \frac{2}{3} & 1 \end{bmatrix}$ and

$$|S| = 14$$
$$|R| = \frac{7}{18}$$
$$s_{11}s_{22}s_{33}|R| = 4 \times 9 \times 1 \times \frac{7}{18} = 14 = |S|$$

which verifies the relationship between $|S|$ and $|R|$.

3.1 Total Variance

The total population variance is defined as the trace of the population variance-covariance matrix. Similarly the total sample variance is defiend as the trace of the sample variance-covariance matrix. That is,

$$\text{Total population variance} = \text{Trace}(\Sigma) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$$

and

$$\text{Total sample variance} = \text{Trace}(S) = s_{11} + s_{22} + \cdots + s_{pp}.$$

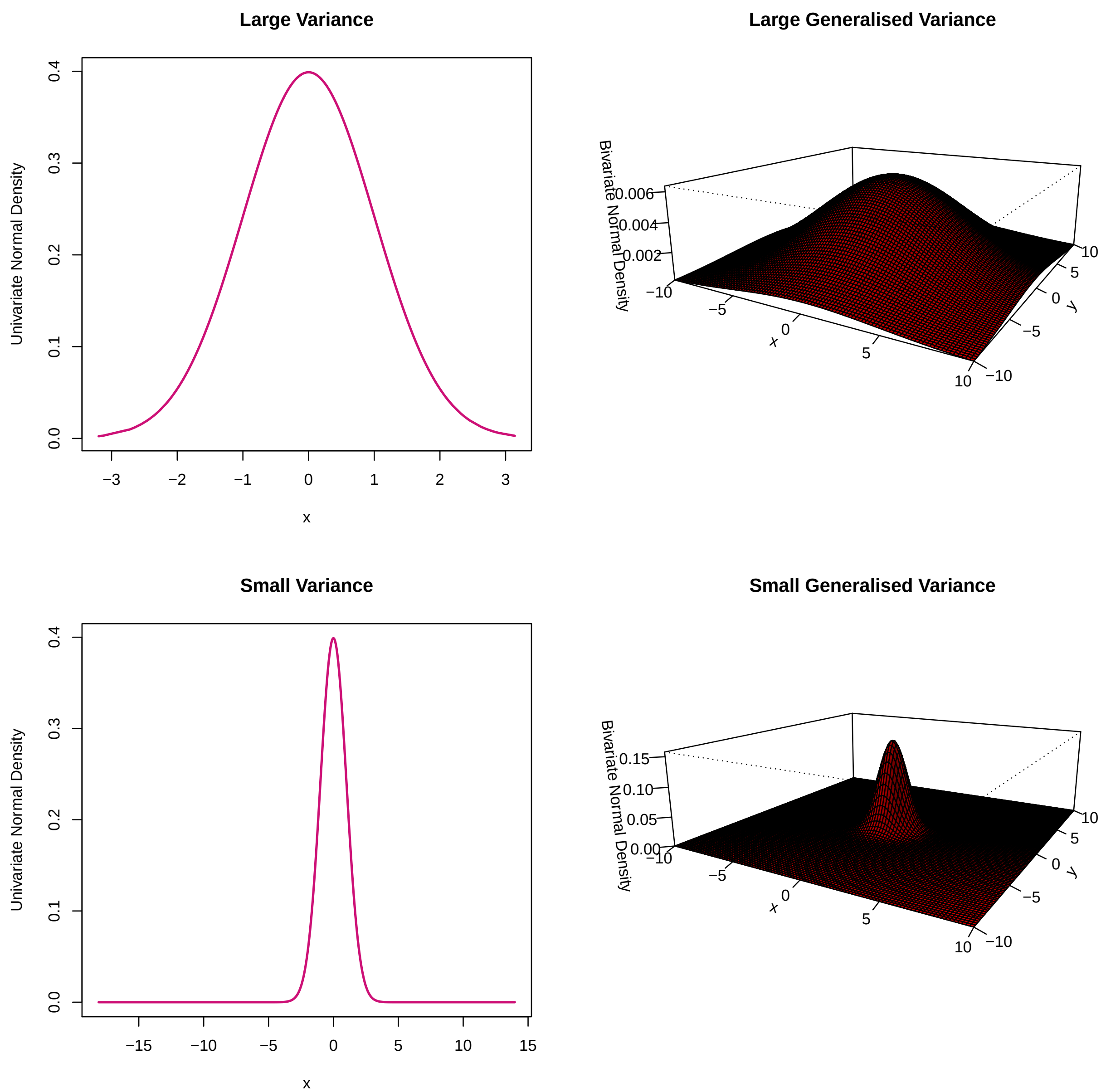

Figure 5: Generalized Sample Variances and Densities