PROPERTIES OF THE BIVARIATE NORMAL DISTRIBUTION

Suppose that two variables X and Y have a ***bivariate Normal distribution***:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \text{Normal}_2 \left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \right) \equiv \text{Normal}\left(\mu, \Sigma\right).$$

In this form, the parameter σ_{XY} is the ***covariance*** between X and Y,

$$\sigma_{XY} = \mathbb{E}_{X,Y}[(X-\mu_X)(Y-\mu_Y)]$$

and Σ is the ***variance-covariance matrix***. If we write

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

the joint density is

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_X \sigma_Y \sqrt{1-\rho_{XY}^2}} \exp\left\{ -\frac{1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{2\rho_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

To understand this joint density better, we examine the term in the exponent

$$\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{2\rho_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2.$$

Writing

$$z_1 = \frac{x-\mu_X}{\sigma_X} \qquad\qquad z_2 = \frac{y-\mu_Y}{\sigma_Y}$$

the expression becomes simply

$$z_1^2 - 2\rho_{XY} z_1 z_2 + z_2^2.$$

We may re-write this by completing the square in z_2 as

$$(z_2 - \rho_{XY} z_1)^2 + (1-\rho_{XY}^2) z_1^2$$

Hence, for the transformed random variables (Z_1, Z_2), we must have the joint density

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho_{XY}^2}} \exp\left\{ -\frac{1}{2(1-\rho_{XY}^2)} \left[(z_2 - \rho_{XY} z_1)^2 + (1-\rho_{XY}^2) z_1^2 \right] \right\}.$$

From this we can deduce the factorization of this joint density as

$$f_{Z_1,Z_2}(z_1,z_2) = f_{Z_2|Z_1}(z_2|z_1) f_{Z_1}(z_1)$$

where

$$f_{Z_1}(z_1) \propto \exp\left\{ -\frac{1}{2} z_1^2 \right\}$$

$$f_{Z_2|Z_1}(z_2|z_1) \propto \exp\left\{ -\frac{1}{2(1-\rho_{XY}^2)} \left[(z_2 - \rho_{XY} z_1)^2 \right] \right\}$$

that is,

$$Z_1 \sim \text{Normal}(0,1) \qquad\qquad Z_2|Z_1 = z_1 \sim \text{Normal}(\rho_{XY} z_1, (1-\rho_{XY}^2)).$$

Finally, back-transforming to (X, Y), we deduce by elementary properties of the Normal distribution

$$X \sim \text{Normal}(\mu_X, \sigma_X^2) \qquad\qquad Y|X = x \sim \text{Normal}\left(\mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2 (1-\rho_{XY}^2) \right).$$

Note that

$$\rho_{XY} \frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}.$$

Therefore, we can conclude the following important properties of the bivariate Normal distribution:

- **Marginal distributions:**

$$X \sim \text{Normal}(\mu_X, \sigma_X^2) \qquad \text{and} \qquad Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$$

- **Conditional distributions:**

$$Y|X = x \sim \text{Normal}\left(\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho_{XY}^2)\right)$$
$$X|Y = y \sim \text{Normal}\left(\mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho_{XY}^2)\right)$$

- **Linear transformation:** If $\mathbf{A}$ is a non-singular (2×2) matrix, and $(U, V)^\top = \mathbf{A}(X, Y)^\top$, then

$$\begin{bmatrix} U \\ V \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} \mu_U \\ \mu_V \end{bmatrix}, \begin{bmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{bmatrix}\right) \equiv \text{Normal}\left(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^\top\right).$$

The `mvtnorm` package has functions that allow us to study the multivariate normal density. After installing the package using `install.packages('mvtnorm')`, run

```
library(mvtnorm)
```

We examine the case where

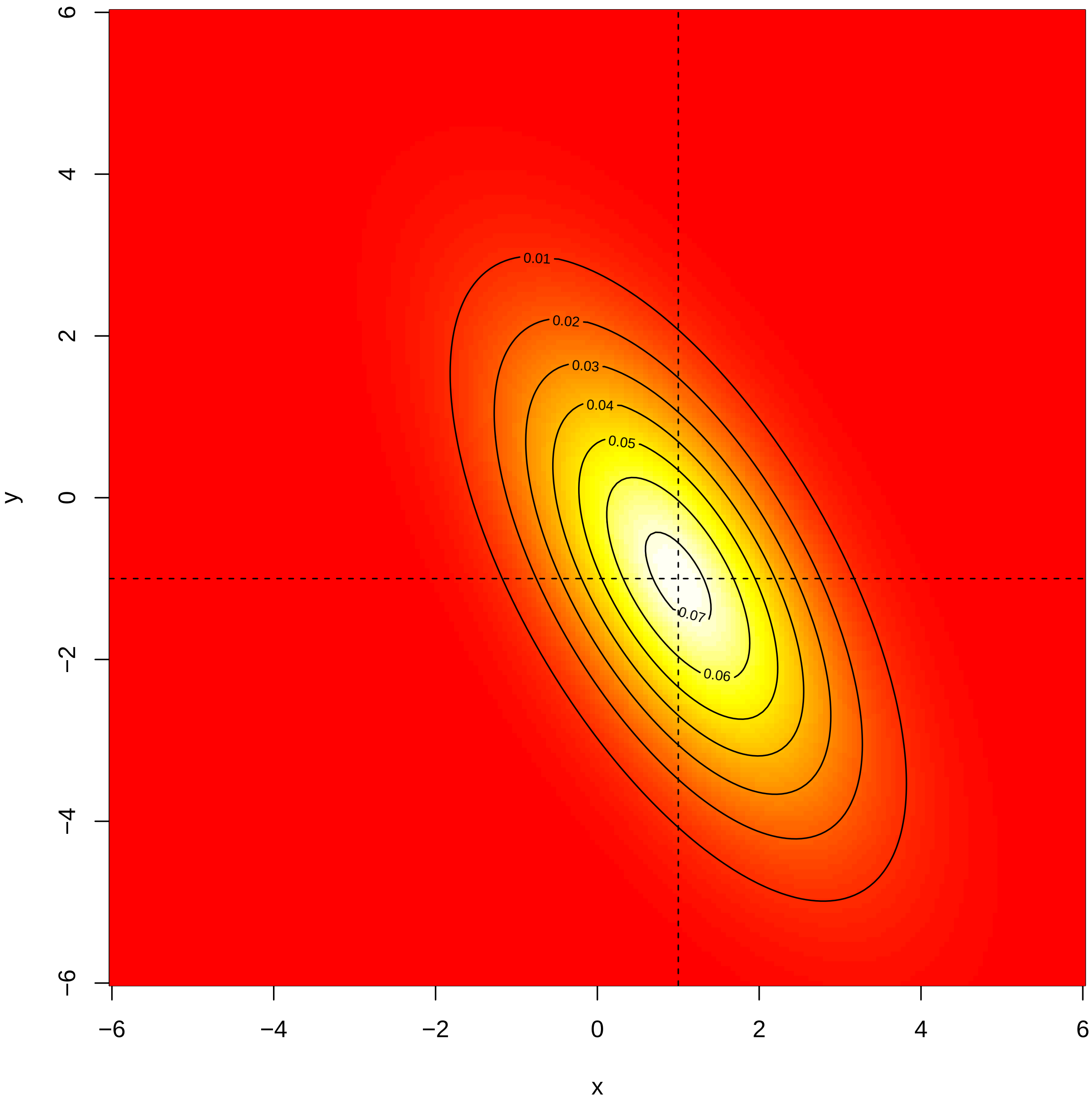
$$\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & -1.8 \\ -1.8 & 4 \end{bmatrix}$$

For this example

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-1.8}{\sqrt{2 \times 4}} = -0.6363.$$

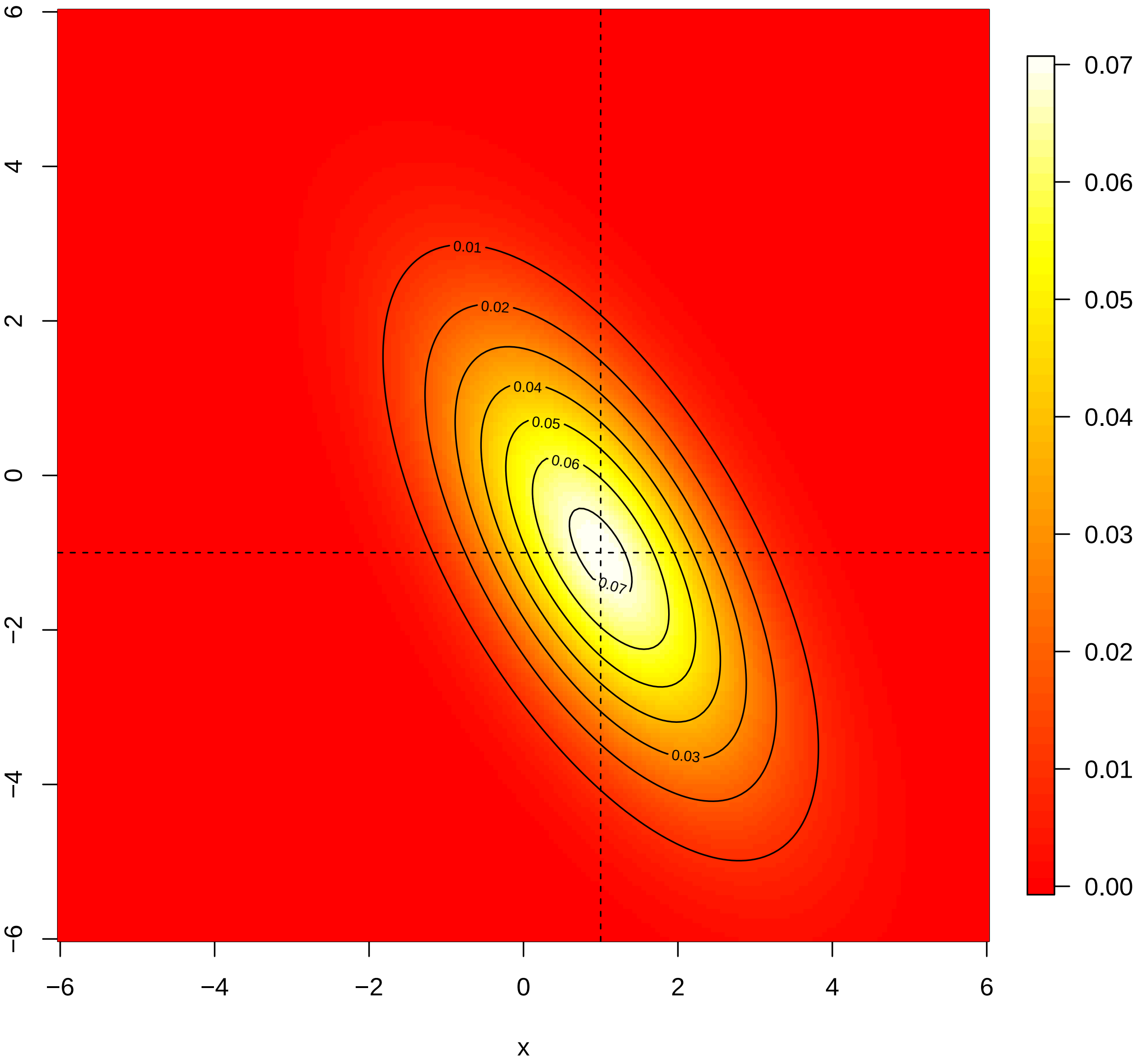
Here is a heatmap and contour plot for the joint density:

```
mu<-c(1,-1)
Sig<-matrix(2*c(1,-0.9,-0.9,2),nrow=2,ncol=2)
xvec<-yvec<-seq(-6,6,length=201)
dens <- matrix(dmvnorm(expand.grid(xvec, yvec),mean=mu,
                   sigma = Sig),ncol = length(xvec))
par(pty='s',mar=c(4,4,0,2))
image(xvec,yvec,dens,xlab='x',ylab='y',col=heat.colors(50))
contour(xvec,yvec,dens,add=T)
abline(v=mu[1],h=mu[2],lty=2)
```



A plot with a colour scale can be obtained using the **image.plot** function from the **fields** package: after loading this package using `library(fields)` we have

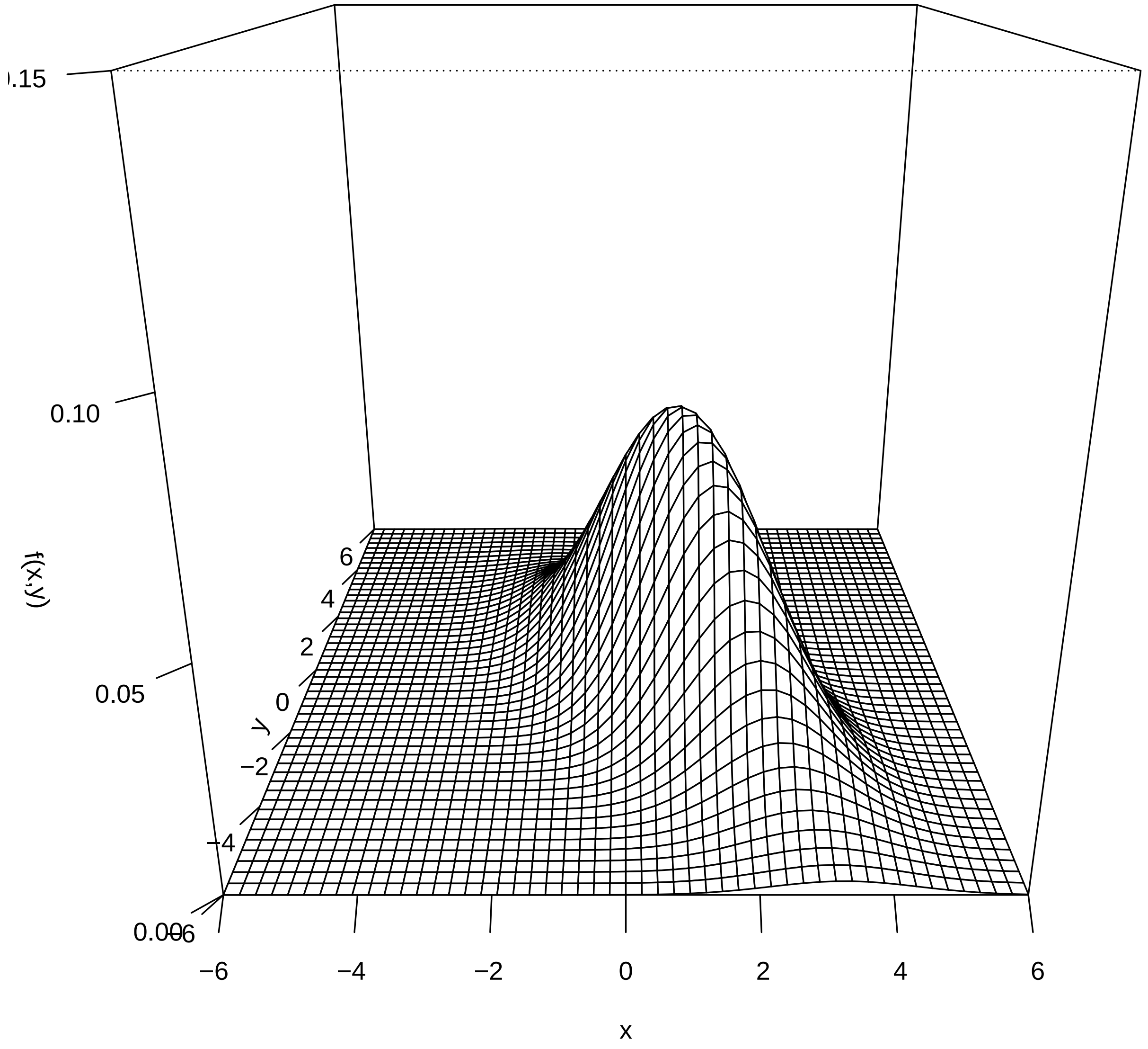
```
library(fields,quietly=TRUE,verbose=FALSE)

+ Spam version 1.4-0 (2016-08-29) is loaded.
+ Type 'help( Spam)' or 'demo( spam)' for a short introduction
+ and overview of this package.
+ Help for individual functions is also obtained by adding the
+ suffix '.spam' to the function name, e.g.  'help( chol.spam)'.
+
+ Attaching package:  'spam'
+ The following objects are masked from 'package:base':
+
+    backsolve, forwardsolve

par(pty='s',mar=c(4,3,0,6))
image(xvec,yvec,dens,xlab='x',ylab='y',col=heat.colors(50))
image.plot(zlim=c(0,0.07),legend.only=TRUE,col=heat.colors(50))
contour(xvec,yvec,dens,add=T)
abline(v=mu[1],h=mu[2],lty=2)
```



6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6
x
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00

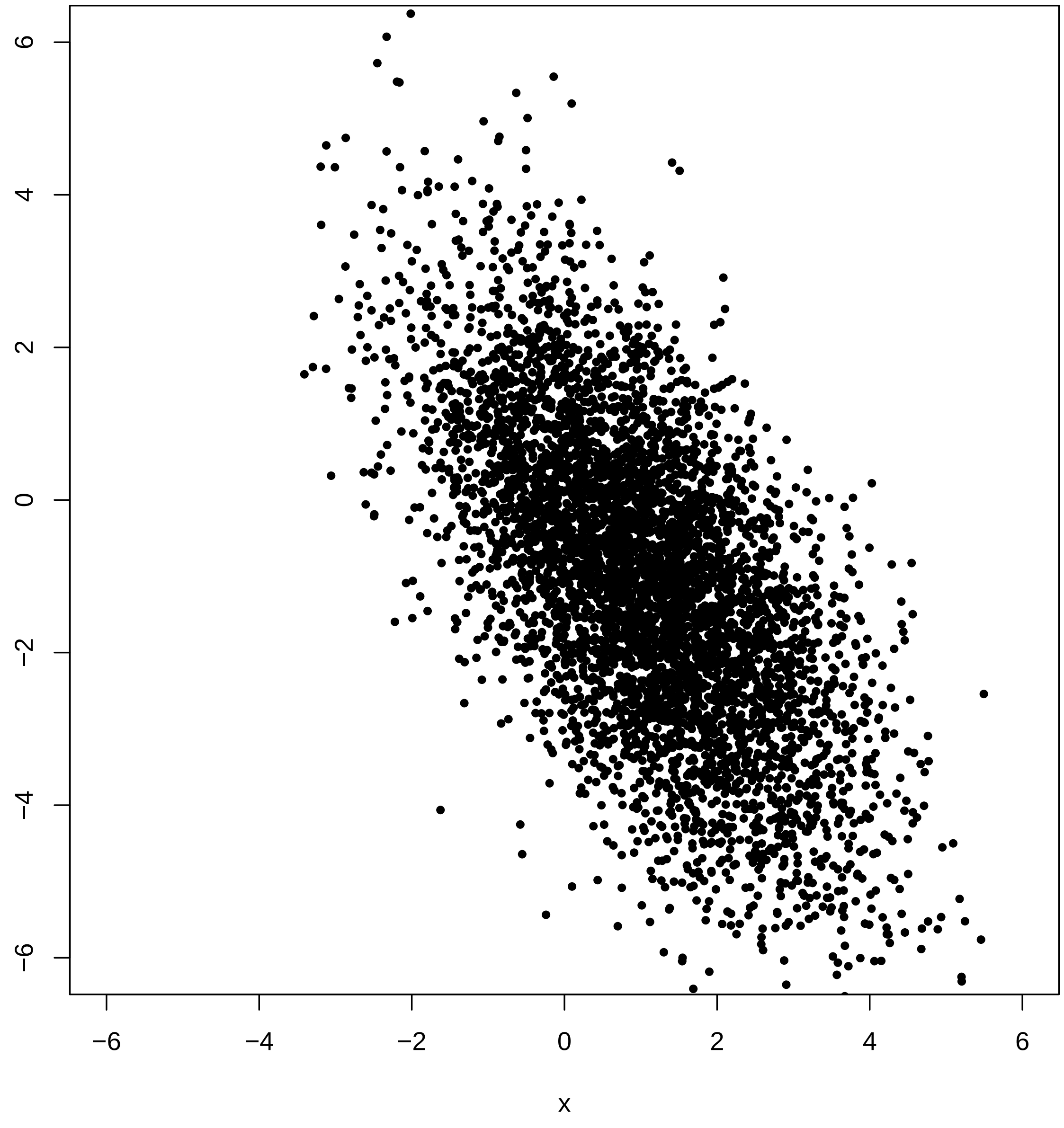
A perspective plot of the joint density can be obtained using `persp`:

```
xvec<-yvec<-seq(-6,6,length=51)
dens <- matrix(dmvnorm(expand.grid(xvec, yvec),mean=mu,
                       sigma = Sig),ncol = length(xvec))
par(pty='s',mar=c(4,2,0,2))
persp(xvec,yvec,dens,xlab='x',ylab='y',zlab='f(x,y)',
           phi=20,zlim=range(0,0.15),ticktype='detailed')
```



A random sample can be obtained from the joint multivariate normal using `rmvnorm`

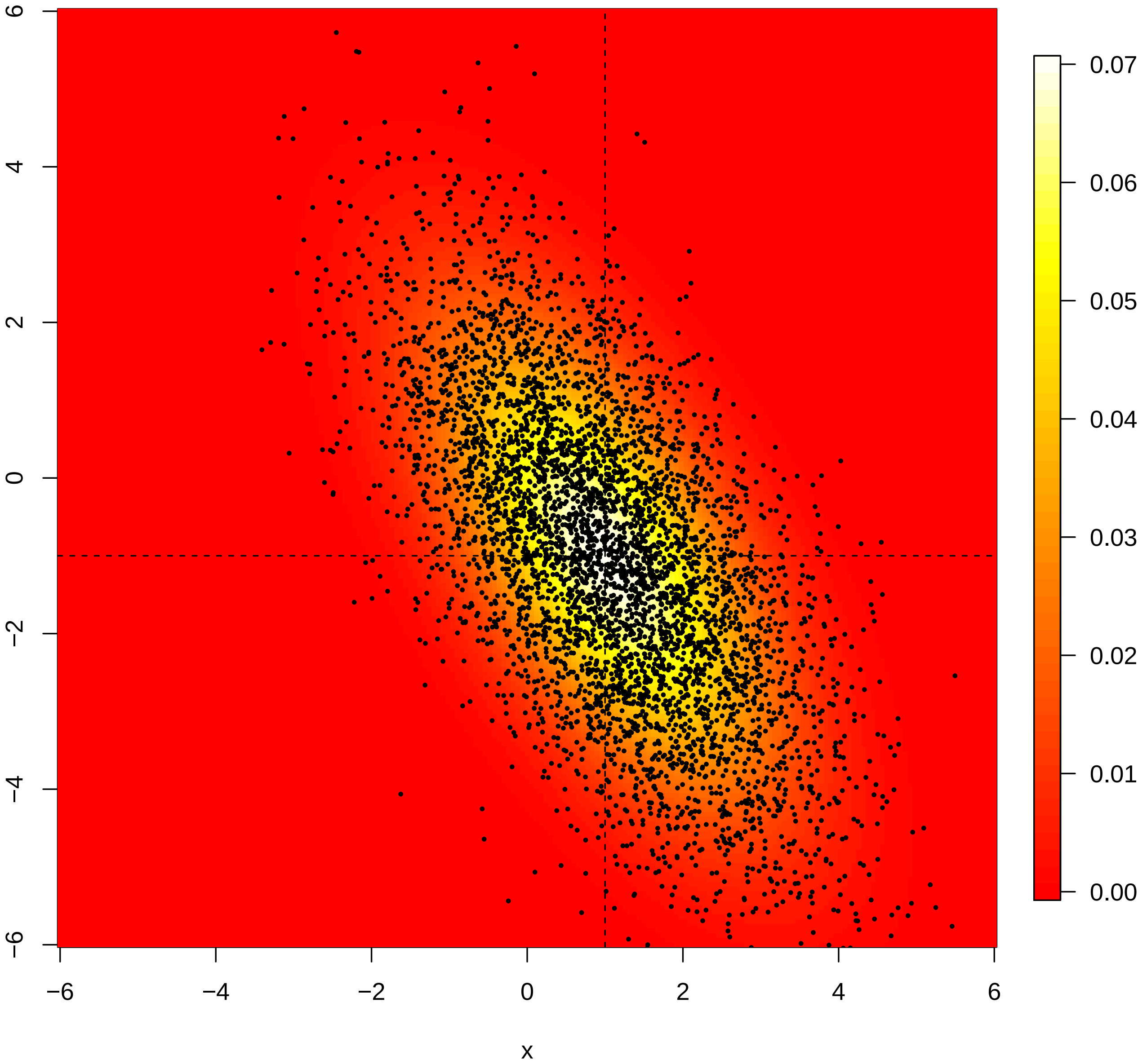
```
N<-5000
set.seed(1234)
XY.samp<-rmvnorm(N,mean=mu,sigma = Sig)
par(pty='s',mar=c(4,2,0,6))
plot(XY.samp,pch=19,cex=0.6,xlab='x',ylab='y',xlim=range(-6,6),ylim=range(-6,6))
```



```
X<-XY.samp[,1];Y<-XY.samp[,2]
```

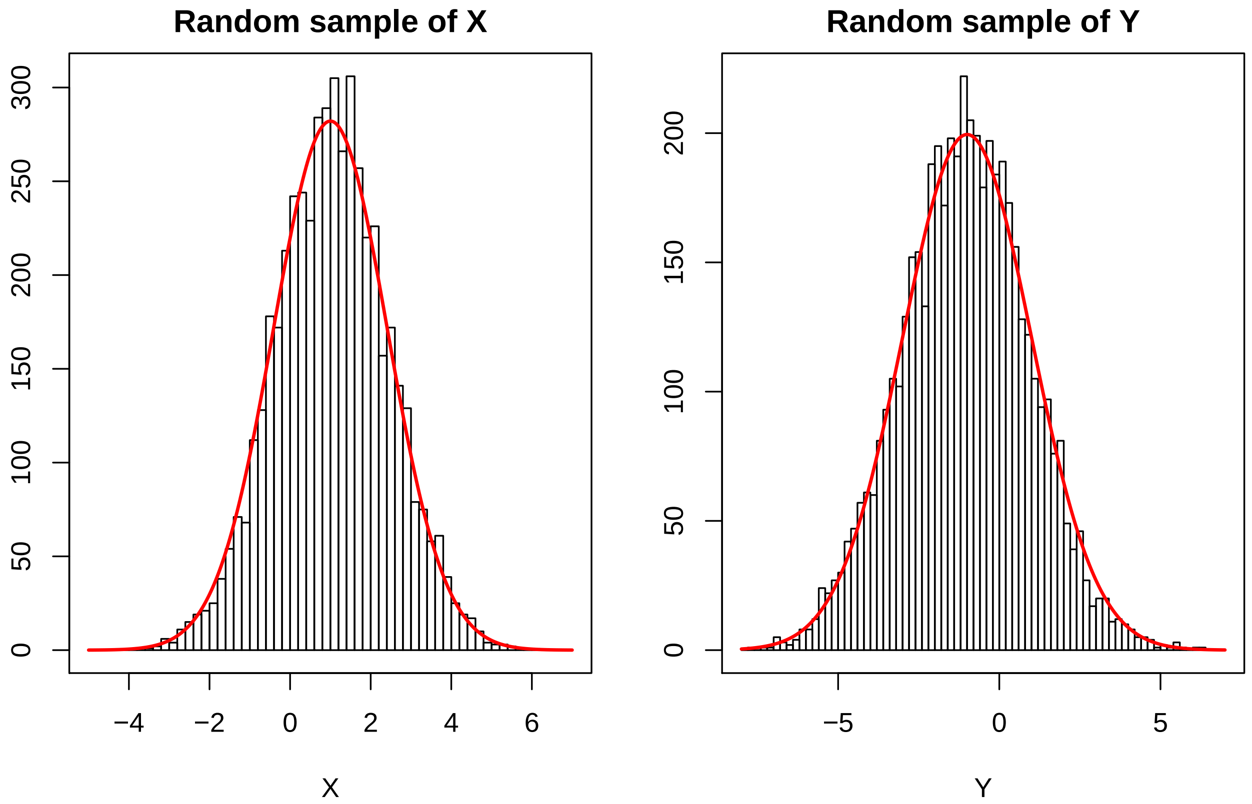
Superimposing the points on the heatmap plot, we see that the random sample is correctly drawn from the high density region;

```
xvec<-yvec<-seq(-6,6,length=201)
dens <- matrix(dmvnorm(expand.grid(xvec, yvec),mean=mu,
                   sigma = Sig),ncol = length(xvec))
par(pty='s',mar=c(4,2,0,6))
image(xvec,yvec,dens,xlab='x',ylab='y',col=heat.colors(50))
image.plot(zlim=c(0,0.07),legend.only=TRUE,col=heat.colors(50))
points(X,Y,cex=0.3,pch=19)
abline(v=mu[1],h=mu[2],lty=2)
```



We can verify that the marginal distributions are correct:

```
par(mfrow=c(1,2),mar=c(4,2,2,2))
muX<-1;sigmaX<-sqrt(2)
xv<-seq(-5,7,length=1001)
yv<-dnorm(xv,muX,sigmaX)
hist(X[X>-5 & X<7],breaks=seq(-5,7,by=0.2),xlab='X',main='Random sample of X');box()
lines(xv,yv*N*0.2,lwd=2,col='red')
muY<--1;sigmaY<-sqrt(4)
xv<-seq(-8,7,length=1001)
yv<-dnorm(xv,muY,sigmaY)
hist(Y[Y>-8 & Y<7],breaks=seq(-8,7,by=0.2),xlab='Y',main='Random sample of Y');box()
lines(xv,yv*N*0.2,lwd=2,col='red')
```

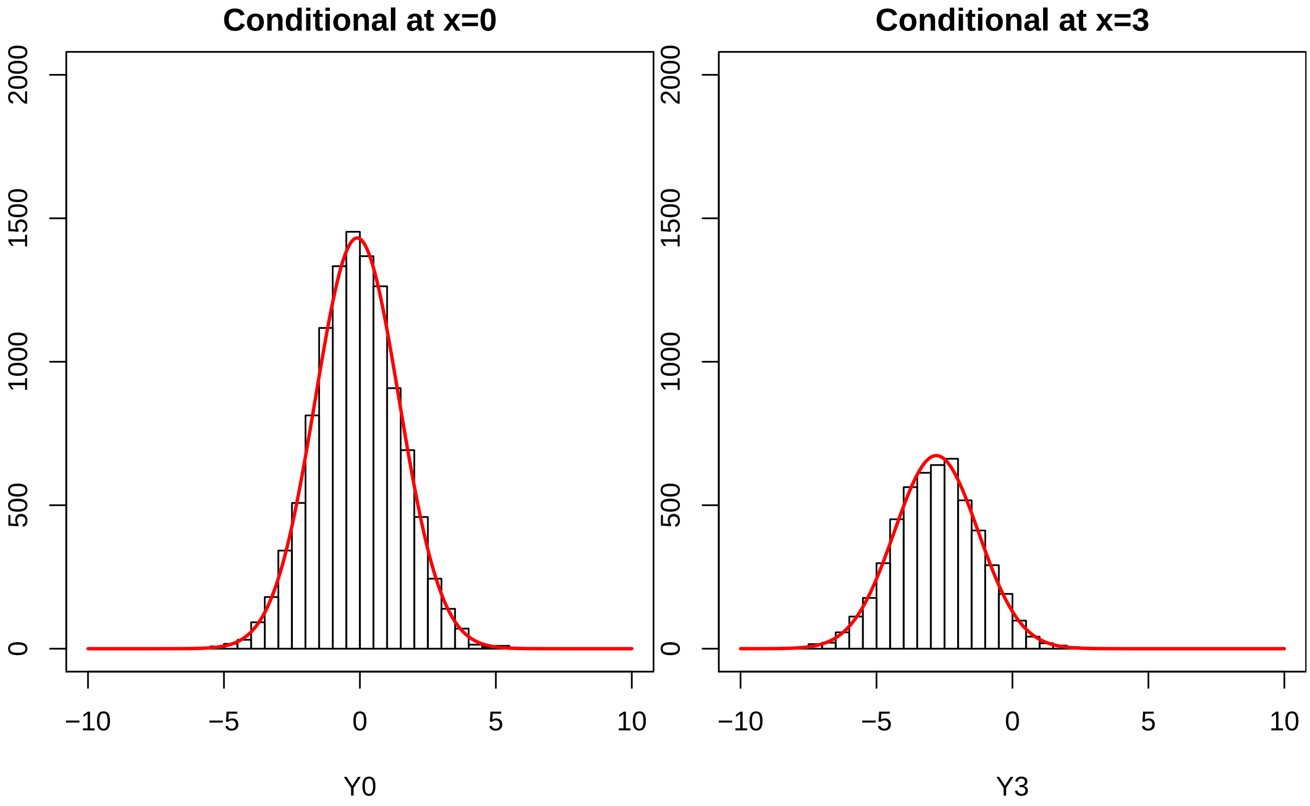


```
mean(X);var(X)

+ [1] 1.044067
+ [1] 1.930187

mean(Y);var(Y)

+ [1] -1.044438
+ [1] 3.909567
```

For the conditional distribution, we take a particular narrow range of x, and look at the samples of y for which the sampled x lie in that range. Here we will take a narrow range around $x = 0$ and then around $x = 3$. We do this with a much larger sample size for effective illustration.

```
sigmaXY<-Sig[1,2]
rhoXY<-sigmaXY/(sigmaX*sigmaY)
XY.big<-rmvnorm(100*N,mean=mu,sigma = Sig)
Xbig<-XY.big[,1];Ybig<-XY.big[,2]
```

```
par(mfrow=c(1,2),mar=c(4,2,2,0))
Y0<-Ybig[Xbig > -0.05 & Xbig < 0.05];N0<-length(Y0)
hist(Y0,breaks=seq(-10,10,by=0.5),ylim=range(0,2000),main='Conditional at x=0');box()
cond.mean0<-muY+(sigmaXY/sigmaX^2)*(0-muX)
cond.var0<-sigmaY^2*(1-rhoXY^2)
yv<-seq(-10,10,by=0.01)
dv<-dnorm(yv,cond.mean0,sqrt(cond.var0))
lines(yv,dv*0.5*N0,col='red',lwd=2)
Y3<-Ybig[Xbig > 2.95 & Xbig < 3.05];N3<-length(Y3)
hist(Y3,breaks=seq(-10,10,by=0.5),ylim=range(0,2000),main='Conditional at x=3');box()
cond.mean3<-muY+(sigmaXY/sigmaX^2)*(3-muX)
cond.var3<-sigmaY^2*(1-rhoXY^2)
dv<-dnorm(yv,cond.mean3,sqrt(cond.var3))
lines(yv,dv*0.5*N3,col='red',lwd=2)
```



At $x = 0$, the conditional mean is

$$\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(0 - \mu_X) = -0.1$$

whereas at $x = 3$, the conditional mean is

$$\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(3 - \mu_X) = -2.8.$$

However, in both cases, the conditional variance is

$$\sigma_Y^2(1 - \rho_{XY}^2) = 2.38$$

RELATIONSHIP WITH SIMPLE LINEAR REGRESSION

To see the relation to regression, we may factorize the joint density

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$$

where $X \sim \text{Normal}(\mu_X, \sigma_X^2)$ and

$$Y|X = x \sim \text{Normal}\left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_X), \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\right) \equiv \text{Normal}\left(\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho_{XY}^2)\right)$$

Equating the conditional expectation of Y given $X = x$

$$\mathbb{E}_{Y|X}[Y|x] = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

with a simple linear regression

$$\mathbb{E}_{Y|X}[Y|x] = \beta_0 + \beta_1 x$$

we identify

$$\beta_0 = \mu_Y - \mu_X\rho_{XY}\frac{\sigma_Y}{\sigma_X} \qquad\qquad \beta_1 = \rho_{XY}\frac{\sigma_Y}{\sigma_X}.$$

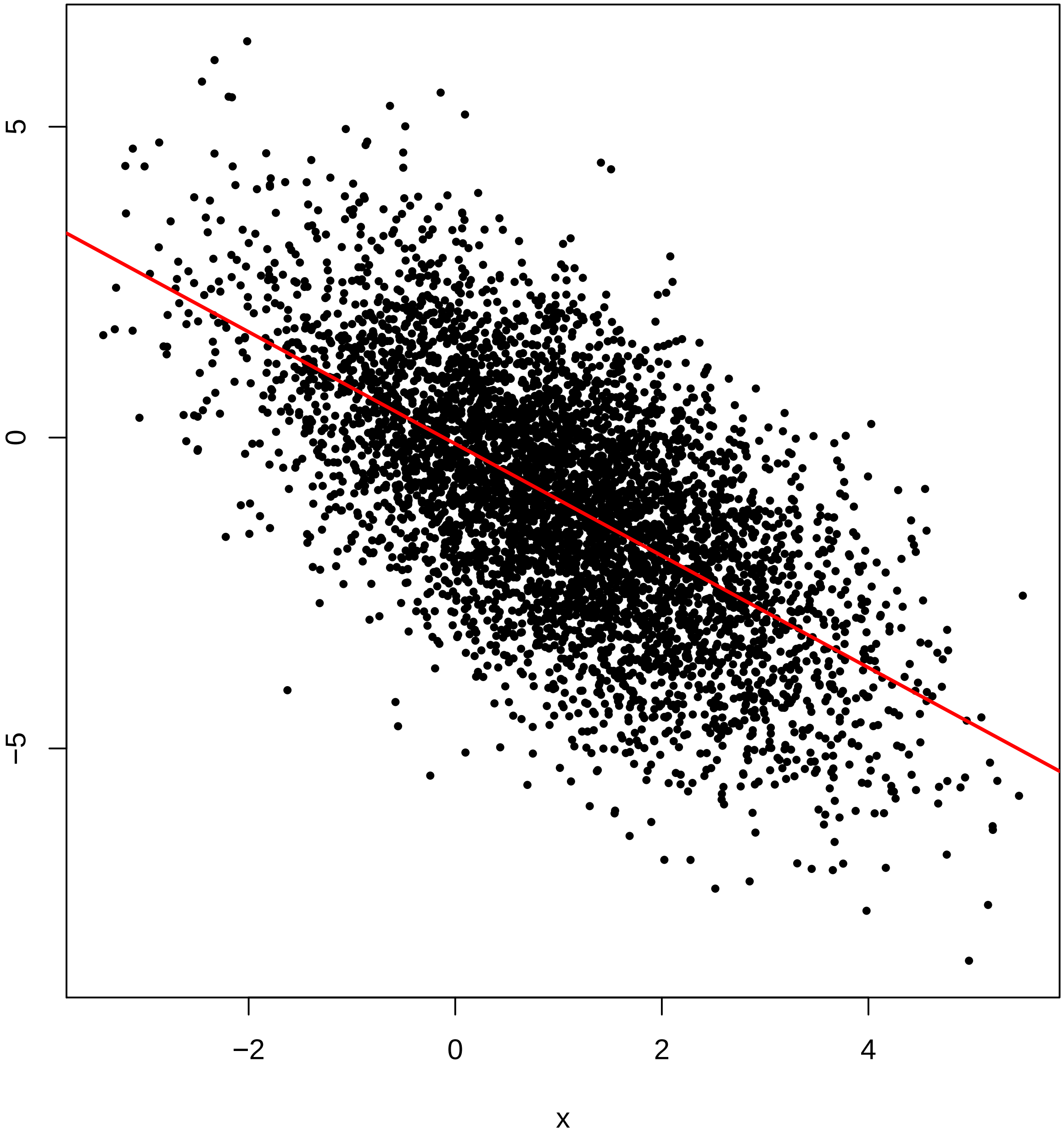
In the example above, we have for the conditional mean of Y at $X = x$

$$\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_X) = -1 + \frac{-1.8}{2}(x-1) = -0.1 - 0.9x$$

and the conditional variance is

$$\sigma_Y^2(1-\rho_{XY}^2) = 4(1-(1.8^2/(2\times 4))) = 2.38.$$

```
be1<-Sig[1,2]/Sig[1,1]
be0<-mu[2]-be1*mu[1]
c(be0,be1)

+ [1] -0.1 -0.9

rho<-Sig[1,2]/sqrt(Sig[1,1]*Sig[2,2])
sigY.X<-Sig[2,2]*(1-rho^2)
rho

+ [1] -0.6363961

sigY.X

+ [1] 2.38

par(pty='s',mar=c(4,2,2,2))
plot(X,Y,pch=19,cex=0.5,xlab='x',ylab='y')
abline(be0,be1,col='red',lwd=2)
```

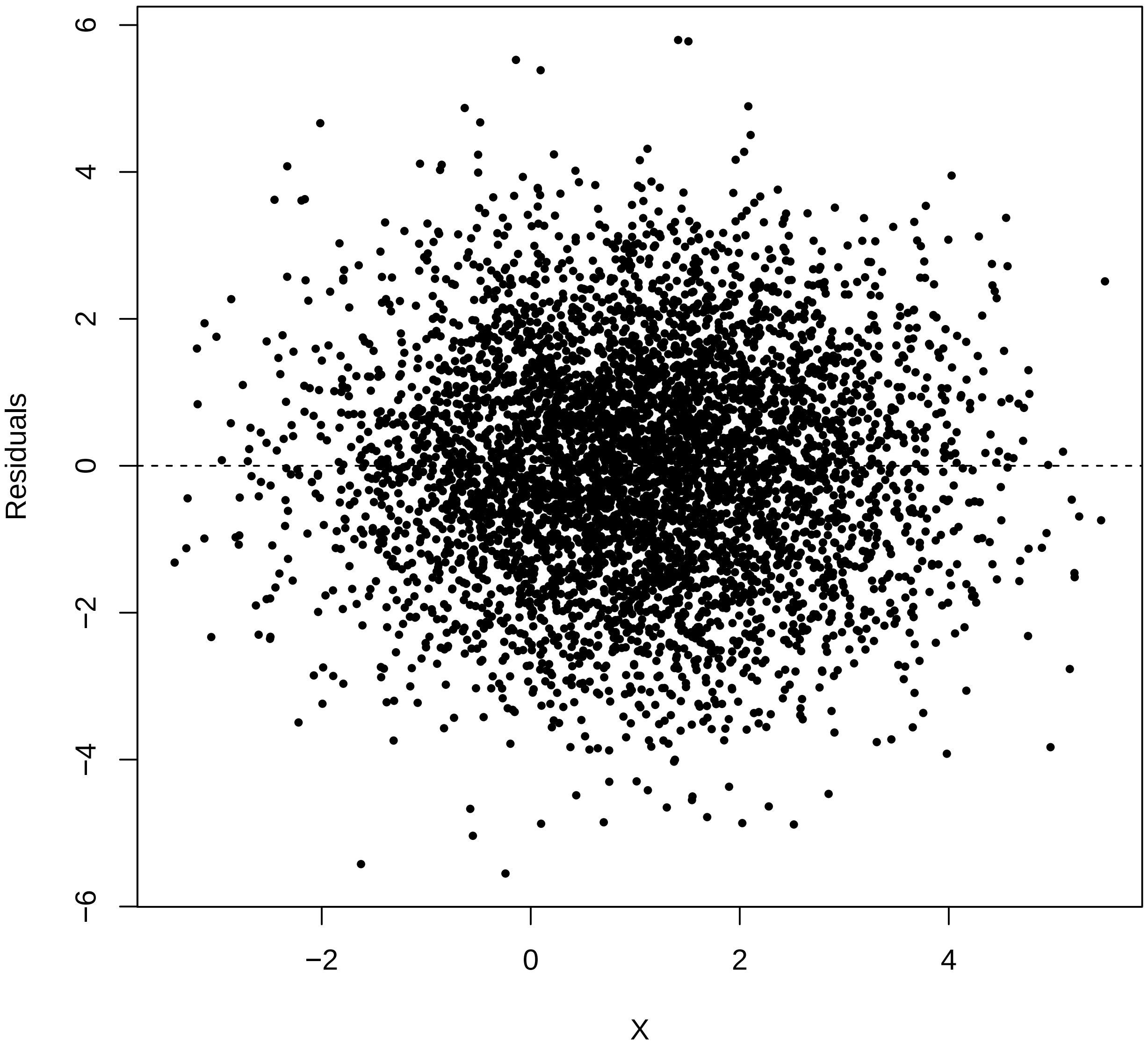


Fitting a simple linear regression model recovers the correct parameters:

```
fitY.X<-lm(Y~X)
summary(fitY.X)$coef

+                Estimate Std. Error   t value     Pr(>|t|)
+ (Intercept) -0.1039638 0.02708447  -3.83850 0.0001253172
+ X           -0.9007801 0.01558568 -57.79536 0.0000000000

summary(fitY.X)$sigma^2  #Estimate of sigma^2

+ [1] 2.343873
```

The residual plot still seems adequate: the values of the x-variable are normally distributed around the value 1, so we no longer see a 'band' around zero, however the residuals do appear to be centered at zero for all x, with constant variability.

```
plot(X,residuals(fitY.X),ylab='Residuals',pch=19,cex=0.5)
abline(h=0,lty=2)
```



To confirm correct recovery of the slope and intercept, we may test using a t-test. The test statistics to test $\mathrm{H}_0 : \beta_0 = -0.1$ and $\mathrm{H}_0 : \beta_1 = -0.9$ are computed as follows:

```
t0<-(coef(fitY.X)[1]+0.1)/summary(fitY.X)$coef[1,2]
t1<-(coef(fitY.X)[2]+0.9)/summary(fitY.X)$coef[2,2]
```

Here we have

$$t_0 = \frac{\widehat{\beta}_0 - (-0.1)}{\text{e.s.e.}(\widehat{\beta}_0)} = \frac{-0.1039638 + 0.1}{0.0270845} = -0.1463479 \simeq -0.1463$$

and

$$t_1 = \frac{\widehat{\beta}_1 - (-0.9)}{\text{e.s.e.}(\widehat{\beta}_1)} = \frac{-0.9007801 + 0.9}{0.0155857} = -0.0500538 \simeq -0.0501$$

neither of which is significant at $\alpha = 0.05$ when compared against the appropriate quantile of the null distribution; here the null distribution is the Student$(n-2) \equiv$ Student(4998) distribution (with 0.975 quantile 1.9604387). The sample correlation is

$$r = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 \sum\limits_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}\mathrm{SS_T}}}$$

which is computed in `R` as follows:

```
cor(X,Y)

+ [1] -0.6329277
```