

5 The multivariate normal distribution

We met the bivariate normal distribution in MAS1604. The bivariate normal distribution is a generalization of the normal distribution to pairs of random variables, or equivalently, to a distribution on vectors in $\mathbb{R}^2$. The multivariate normal distribution is the analogous distribution on vectors in $\mathbb{R}^n$.

5.1 The bivariate normal distribution

Suppose we have constants $\mu_x, \mu_y, \sigma_x, \sigma_y$ and ρ such that $\sigma_x \geq 0, \sigma_y \geq 0$ and $-1 \leq \rho \leq 1$. Define the 2×2 matrix Σ by

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

Then define a joint probability density function by

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}Q(x, y)\right) \quad (1)$$

where

$$Q(x, y) = (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) \quad (2)$$

and the column vectors $\underline{x}, \underline{\mu}$ are defined by $\underline{x} = (x, y)^T, \underline{\mu} = (\mu_x, \mu_y)^T$. If random variables (X, Y) have joint probability density given by $f_{X,Y}$ above, then we say that (X, Y) have a **bivariate normal distribution** and write

$$(X, Y)^T \sim N_2(\underline{\mu}, \Sigma).$$

Some of the terms in the definition can be expanded:

$$\sqrt{\det \Sigma} = \sigma_x \sigma_y \sqrt{1 - \rho^2}$$

and

$$\begin{aligned} Q(x, y) &= \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)} \begin{pmatrix} x - \mu_x & y - \mu_y \end{pmatrix} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \\ &= \frac{1}{1 - \rho^2} \left[\left(\frac{x - \mu_x}{\sigma_x} \right)^2 + \left(\frac{y - \mu_y}{\sigma_y} \right)^2 - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} \right]. \end{aligned}$$

However, equations (1) and (2) are **much** more important than the expansions just given.

It can be proved that the function $f_{X,Y}(x, y)$ in equation (1) integrates to 1 and therefore defines a valid joint PDF. The proof is omitted.

Remarks:

1. The vector $\underline{\mu} = (\mu_x, \mu_y)^T$ is called the **mean vector** and the matrix Σ is called the **covariance matrix** (or sometimes **variance-covariance matrix**).
2. Functions of the form $F(\underline{x}) = \underline{x}^T \Sigma^{-1} \underline{x}$ are called **quadratic forms**. Quadratic forms are functions $\mathbb{R}^n \rightarrow \mathbb{R}$ which satisfy certain properties. They crop up in several areas of mathematics and statistics.

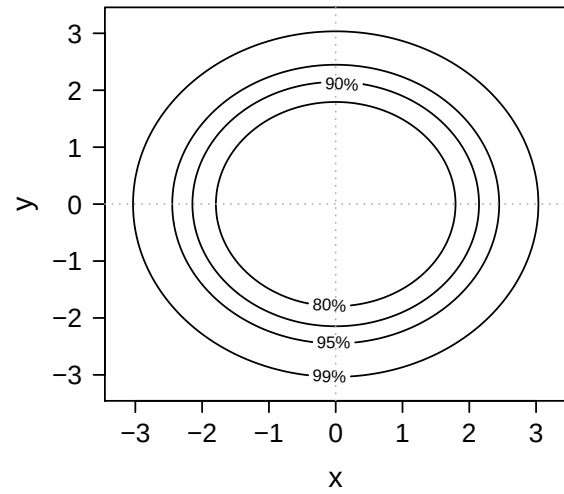
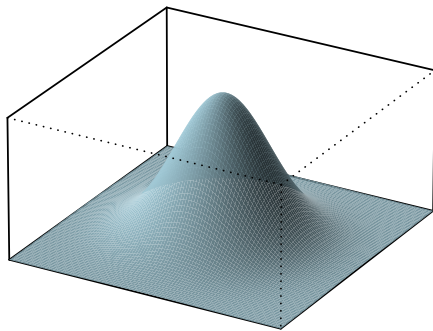
3. The matrix Σ and its inverse Σ^{-1} are **positive definite**. A matrix A is positive definite if

$$\underline{x}^T A \underline{x} \geq 0$$

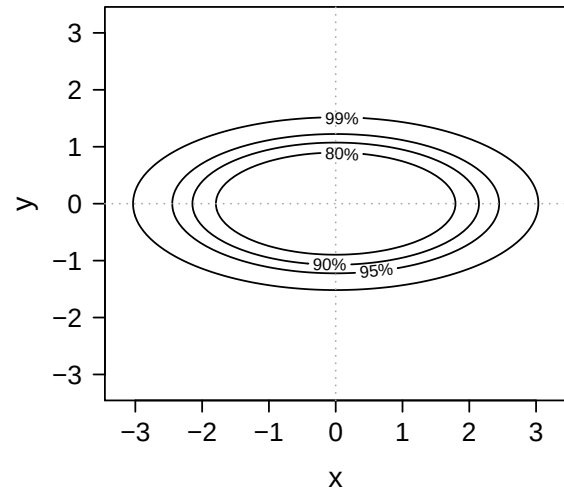
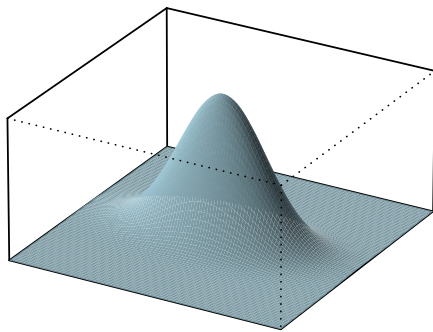
for all non-zero vectors $\underline{x}$.

4. It follows that when $\mu_x = \mu_y = 0$, $Q(x, y)$ is of positive definite quadratic form.

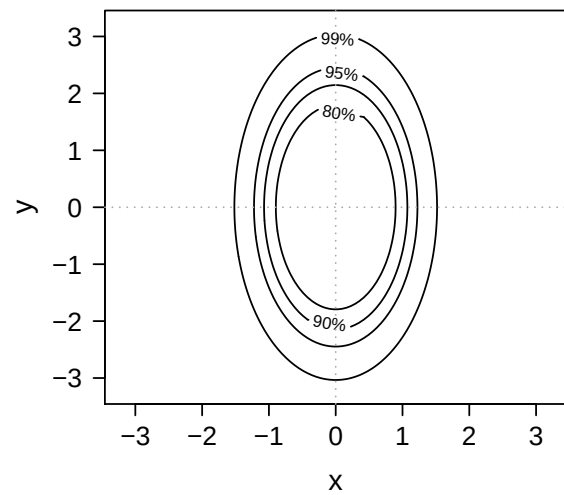
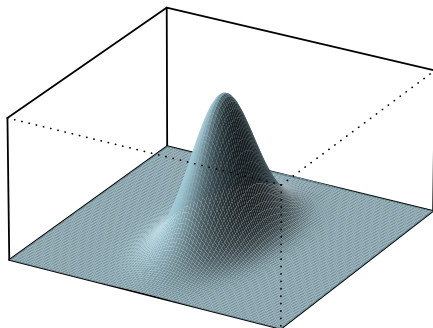
$$\sigma_x = \sigma_y, \rho = 0$$



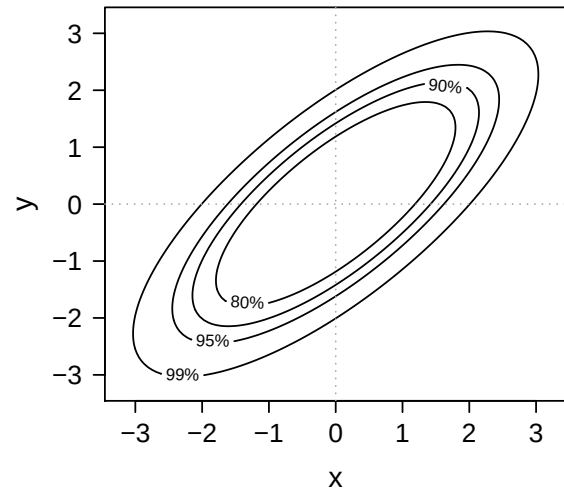
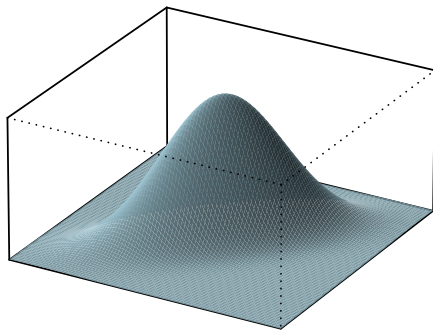
$$\sigma_x = 2\sigma_y, \rho = 0$$



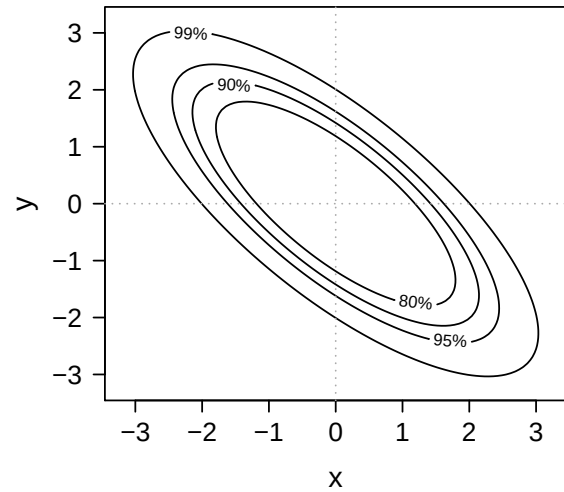
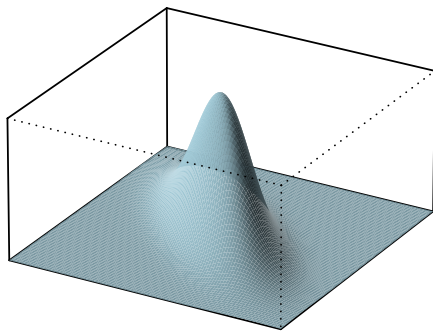
$$2\sigma_x = \sigma_y, \rho = 0$$



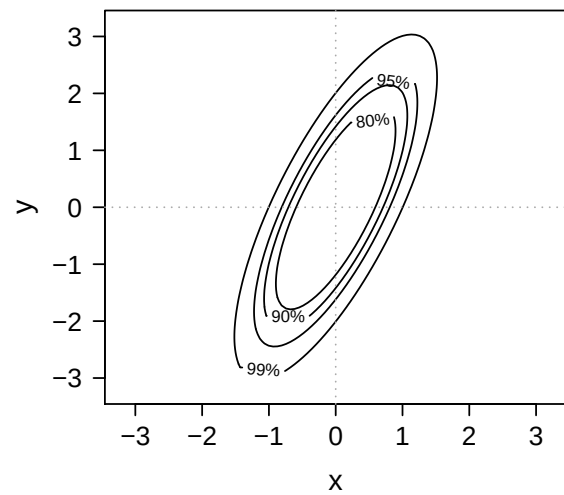
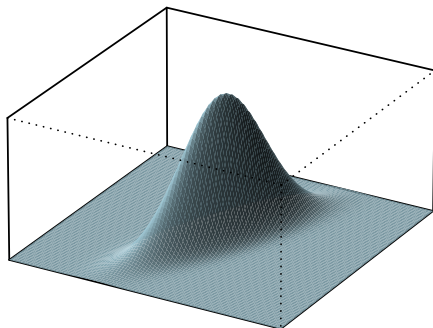
$$\sigma_x = \sigma_y, \rho = 0.75$$



$$\sigma_x = \sigma_y, \rho = -0.75$$



$$2\sigma_x = \sigma_y, \rho = 0.75$$



The previous two pages show contour and surface plots for the density function on the xy -plane for different parameter values. Comments:

1. $Q(x, y) \geq 0$ with equality only when $\underline{x} = \underline{\mu}$. It follows that the density function has its mode at $\underline{x} = \underline{\mu}$. The above plots all have $\mu_x = \mu_y = 0$.
2. Changing the values of μ_x, μ_y does not change the shape of the plots, but corresponds to a translation of the xy -plane i.e. changing μ_x, μ_y just shifts the contours / surface to a new mode position.
3. The contours of equal density are **circular** when $\sigma_x = \sigma_y$ and $\rho = 0$ and **elliptical** when $\sigma_x \neq \sigma_y$ or $\rho \neq 0$.
4. σ_x and σ_y control the extent to which the distribution is **dispersed**.
5. The parameter ρ is the **correlation** of X, Y i.e. $\text{Cor}(X, Y) = \rho$. (The proof is omitted.) Thus for non-zero ρ , the contours are at an angle to the axes.

When (X, Y) has a bivariate normal distribution then given a general set $A \subseteq \mathbb{R}^2$, calculation of the probability $\Pr((X, Y) \in A)$ requires a computer. However, the marginal and conditional distributions of the bivariate normal distribution are particularly simple – one of the reasons the distribution is so useful in statistics.

Suppose $(X, Y)^T \sim N_2(\underline{\mu}, \Sigma)$. The following results hold.

1. The marginal distributions are normal:

$$X \sim N(\mu_x, \sigma_x^2) \quad \text{and} \quad Y \sim N(\mu_y, \sigma_y^2).$$

2. The conditional distributions are normal:

$$X|Y = y \sim N\left(\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\right) \quad \text{and}$$

$$Y|X = x \sim N\left(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right).$$

3. When $\rho = 0$, X and Y are independent.
4. Linear combinations of X and Y are also normally distributed:

$$aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$$

where a, b are constants. Thus a **linear combination** of normal random variables is also normally distributed itself.

Proofs are omitted.

5.2 Simulating bivariate normal random variables

R has built-in functions to simulate bivariate and multivariate normal random variables. However, we can simulate using the strategy introduced in section 4.1: first simulate from the marginal for X (or marginal for Y if preferred), and then simulate Y from the conditional $Y|X = x$.

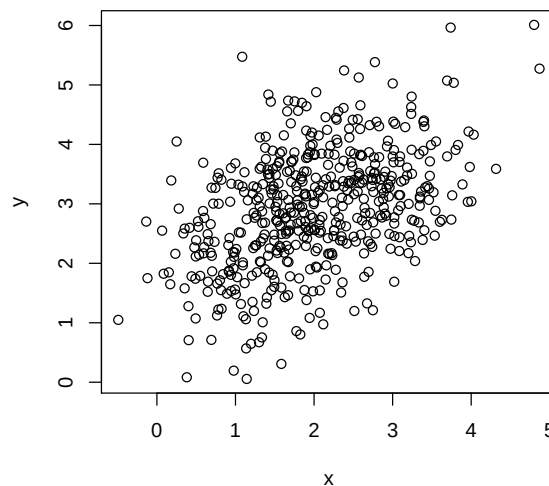
Example 5.1: Suppose $(X, Y)^T \sim N_2(\underline{\mu}, \Sigma)$ where $\mu_x = 2$, $\mu_y = 3$, $\sigma_x = 1$, $\sigma_y = 1$ and $\rho = 0.5$. Simulate a sample of size 500 from this distribution and draw a scatter plot. Use simulation to find $\Pr(X^2 + Y^2 < 9)$.

The marginal distribution of X is $X \sim N(2, 1^2)$. Using the formula for the conditional

$$\begin{aligned} Y|X = x &\sim N\left(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right) \\ &\sim N(3 + 0.5(x - 2), 0.75). \end{aligned}$$

To sample:

```
1 npts = 500
2 x = rnorm(npts, mean = 2, sd = 1)
3 y = rnorm(npts, mean = 3+0.5*(x-2), sd = sqrt(0.75))
```



To find $\Pr(X^2 + Y^2 < 9)$ approximately, count the number of points in the region:

```
1 npts = 10000
2 x = rnorm(npts, mean = 2, sd = 1)
3 y = rnorm(npts, mean = 3+0.5*(x-2), sd = sqrt(0.75))
4 f = x^2+y^2
5 sum(f<9)/npts
6 [1] 0.2877
```

5.3 The multivariate normal distribution

The multivariate normal distribution is a distribution on vectors in $\mathbb{R}^n$. Suppose that $\underline{X}$ is a random vector with n entries, i.e. $\underline{X} = (X_1, \dots, X_n)^T$. Then

$$\underline{X} \sim N_n(\underline{\mu}, \Sigma)$$

if $X_1, \dots, X_n$ have joint PDF given by

$$f_{\underline{X}}(\underline{x}) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}Q(\underline{x})\right)$$

where

$$Q(\underline{x}) = (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}).$$

This definition makes sense for any column vector $\underline{\mu} \in \mathbb{R}^n$ and any **positive definite** $n \times n$ matrix Σ . Some remarks:

1. The vector $\underline{\mu}$ is the **mean** of the distribution and Σ is called the **covariance** matrix.
2. All the marginal distributions of $\underline{X}$ are normal. (We do not specify their parameters here, however).
3. Similarly, all the conditional distributions of $\underline{X}$ are normal. (Again, we do not specify the parameters of these distributions here).

Example 5.2: Suppose $\underline{X} \sim N_3(0, \Sigma)$ where

$$\Sigma = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

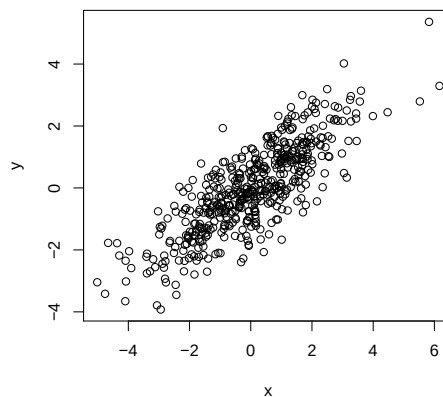
Use R to sample from this distribution. Create a scatterplot of (X, Y) samples and compute the correlation.

In R, you use the `mvnrm` function to sample:

```
1 library(MASS)
2 npts = 500
3 mu = c(0, 0, 0)
4 sigma = matrix(data = c(3,2,1,2,2,1,1,1,3), nrow = 3, byrow = TRUE)
5 x = mvnrm(npts, mu, sigma)
```

Now `x` has dimension `npts`×3, so column 1 is X and column 2 is Y :

```
1 plot(x[,1], x[,2], xlab = 'x', ylab = 'y')
```



The sample correlation is:

```
1 cor(x[,1], x[,2])
2 [1] 0.812629
```