Take home 2

Pieter Luyten

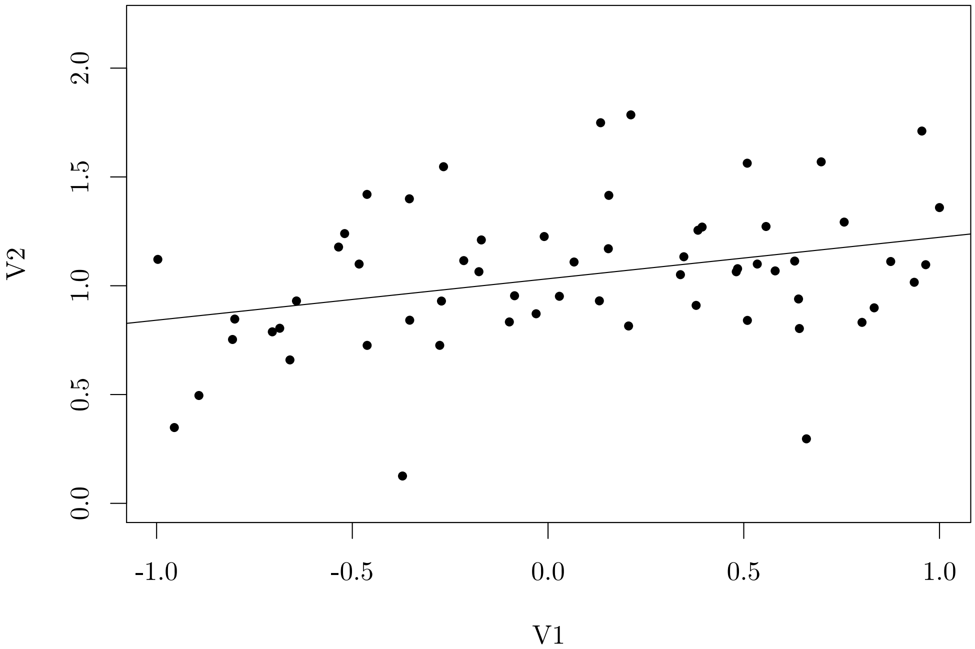
14 december 2019

Excercise 1

(a)

Fit a linear model, assuming that the strong Gaussian assumption is relevant.

The value for the intercept of the fit is 1.0321764 and for the rico of the fit is 0.1904323



Figuur 1: line fit throught the data in Ex1.txt

(b)

Test whether $\beta_1 = 0$ at the level $\alpha = 0.01$.

Using the result from section 7.4.1 in <REFERENCE cursus> we know that the random variable

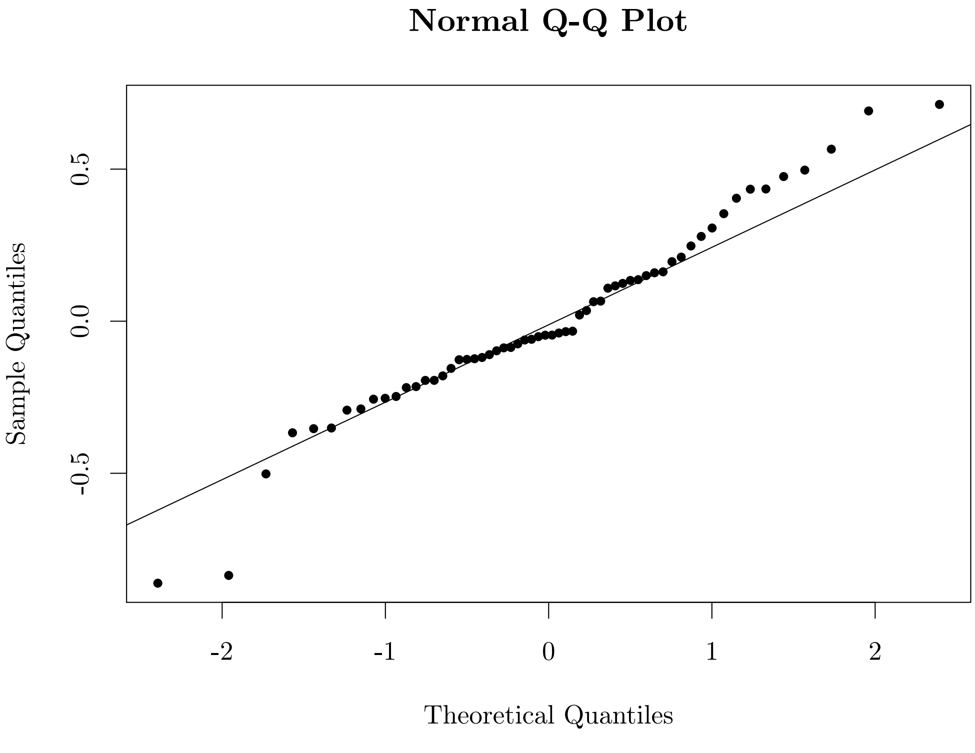
$$T = \frac{\beta_1}{\sqrt{\frac{S^2}{\sum_{i=1}^{n}(x_i - \bar{x}^2)}}} \tag{1}$$

has a Student distribution with $n-2$ degrees of freedom. The test value is 2.603. Using a student-t distribution with $60-2=58$ degrees of freedom we find a p-value of 0.0117. At the confidence level $\alpha = 0.01$, the null hypothesis that $\beta_1 = 0$ holds. The 99% confidence region for the test value is $[-2.663, 2.663]$.

(c)

Explain what a Q-Q plt is and apply this to the residuals.

A q-q plot is a plot where the quantiles of the assumed distribution are plotted against the quantiles from the sample. So in a sample of n points where we label the observation $x_i, i \in \{1, 2, \ldots, n\}$ from lowest to highest, the ith point will be plotted at the $(Q(i/n), x_i)$. Here $Q(x)$ is the quantile function of the assumed distribution. This is a function such that $P(X < Q(p)) = p$. If the assumed distribution is a good model for the observed sample, the points will well fitted by a linear function.

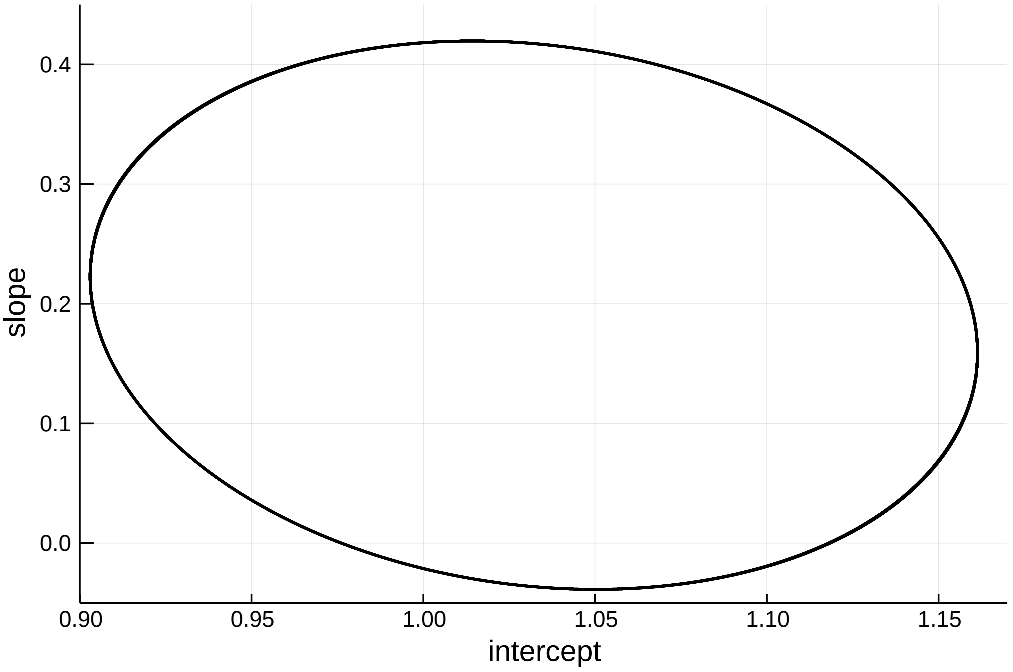


Figuur 2: qq plot for the errors on the fit

(d)

Perform a test that does not require normality of the errors.

To do a test that does not require normality of the errors we first derive an assymptotic result



Figuur 3: A plot of the 99% confidence region for β.

for the distribution of $\hat{\beta}$. We use that $\hat{\beta}$ is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2}$$

And that we can rewrite Y as $Y = X\beta + \epsilon$ where ϵ is the random vector of errors. Filling this in in equation 2 yields

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T (X\beta + \epsilon)\\ &= (X^T X)^{-1}(X^T X)\beta + (X^T X)^{-1} X^T \epsilon\\ &= \beta + (X^T X)^{-1} X^T \epsilon\end{aligned}$$

We see that the distribution of $\hat{\beta}$ is a linear combination of the distributions of the individual errors.

$$\hat{\beta}_i = \beta_i + \sum_{j=1}^{n} [(X^T X)^{-1} X^T]_{ij} \epsilon_j \tag{3}$$

Let $A = (X^T X)^{-1} X$. If we assume that the ϵ_i are i.i.d. with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$ we see that the distribution of $\hat{\beta}_i$ has expectation value $E(\hat{\beta}_i) = \beta_i$ and $Var(\hat{\beta}_i) = (A_i \cdot A_i^T)\sigma^2$.

(e)

Determine a 99% confidence region for $\hat{\beta}$

We use the formula from section 7.5.4 in the book to find a confidence region for β. THis region is of the form:

$$\left\{ \beta \middle| \frac{(\hat{\beta} - \beta)^T (X^T X)(\hat{\beta} - \beta)}{pS^2} \leq F_{p,n-p}(1 - \alpha) \right\} \tag{4}$$

Because the matrix $X^T X$ is positive definite, this is an ellips with center β and axes along the eigenvectors of $X^T X$. The upper bound for this expression is in this case: $F_{2,58(0.99)=4.9910}$ Filling in the values for $X, p = 2, n = 60$ and S^2 we get the following result (obtained using the cAS system sympy):

$$\{\beta | 5.2596\alpha^2 + 0.8222\alpha\beta - 0.4950\alpha + 1.6711\beta^2 - 0.6629\beta + 0.0711 \leq 4.9910\}$$

The resulting ellips is plotted in figure 3.

Question 2

(a)

Compute the ordinary least squares $\hat{\beta}_{OLS}$.

The value for $\hat{\beta}_{OLS}$ obtained using ordinary least squares with the correlated errors is:

$$\hat{\beta}_{OLS} = (3.923161, 1.214355)^T \tag{5}$$

(b)

Suppose we further know that the errors are correlated and satisfy the following equation:

$$\epsilon_i = \rho\epsilon_{i-1} + \eta_i, \quad \text{for} i = 1, \ldots, n \tag{6}$$

Wwhere η_n are i.i.d. standar Gaussian and $\rho = 0.8$. Use 6 to compute the variance of ϵ.

we know that the variance is the same for all ϵ_i and that they obey the recursive relation:

$$\epsilon_i = \rho\epsilon_{i-1} + \eta_i$$

with η_i standard normaly distributed. Using $Var(\epsilon_i) = Var(\epsilon_j)$ for all $i, j \in \{1, 2, \ldots 80\}$ we find:

$$\begin{aligned}
Var(\epsilon_i) &= \rho^2 Var(\epsilon_{i-1}) + Var(\eta_i) \\
\Leftrightarrow Var(\epsilon_i) &= \rho^2 Var(\epsilon_i) + 1 \\
\Leftrightarrow Var(\epsilon_i) &= \frac{1}{1-\rho^2} \\
&= \frac{25}{9} \\
&= 2.777\ldots
\end{aligned}$$

We conclude that the variance of ϵ is $\frac{25}{9}$.

(c)

Transform the model in such a way that the errors are no longer correlated. cCompute the ordinary least squares for this new model and compare it to the one obtained in (a)

To transform the errors to a basis where they are no longer correlated we will use the variance-covariance matrix V of the errors. The off-diagonal elements, so the covariances, are (suppose $i < j$:

$$\begin{aligned}
Cov(\epsilon_i, \epsilon_j) &= E[(\epsilon_i - E(\epsilon_i))(\epsilon_j - E(epsilon_j))] \\
&= E[\epsilon_i \epsilon_j] \\
&= E[\epsilon_i(\rho\epsilon_{j-1}\eta_j)] \\
&\vdots \\
&= E[\epsilon_i(\rho^{j-i}\epsilon_i + \rho^{j-i-1}\eta_i + \rho^{j-i-2}\eta_{i-2} + \ldots + \eta_j)] \\
&= E[\epsilon_i \rho^{i-j}\epsilon_i] \\
&= \rho^{i-j} Var(\epsilon) \\
&= \rho^{i-j}\frac{25}{9}
\end{aligned}$$

So the matrix V has $\frac{25}{9}$ on the diagonal, $\frac{25}{9}\rho$ on the two sub-diagonals next to the diagonal, $\frac{25}{9}\rho^k$ on the kth sub-diagonal. the matrix V can be diagonalized with an orthogonal (or unitary) transformation U: $\Lambda = MVM^T$. When we use the definition of the variance-covariance matrix (where ϵ denotes the column vector with the errors) we find:

$$\begin{aligned} V &= E((\epsilon - E(\epsilon))(\epsilon - E(\epsilon))^T) \\ \Leftrightarrow \Lambda &= UE(\epsilon\epsilon^T)U^T \\ \Leftrightarrow \Lambda &= E((U\epsilon)(U\epsilon)^T) \end{aligned}$$

Where we used that $E(\epsilon = \mathbf{0})$. The vector $U\epsilon$ has variance-covariance matrix Λ so we conclude that it is a vector of uncorrelated stochastic variables.

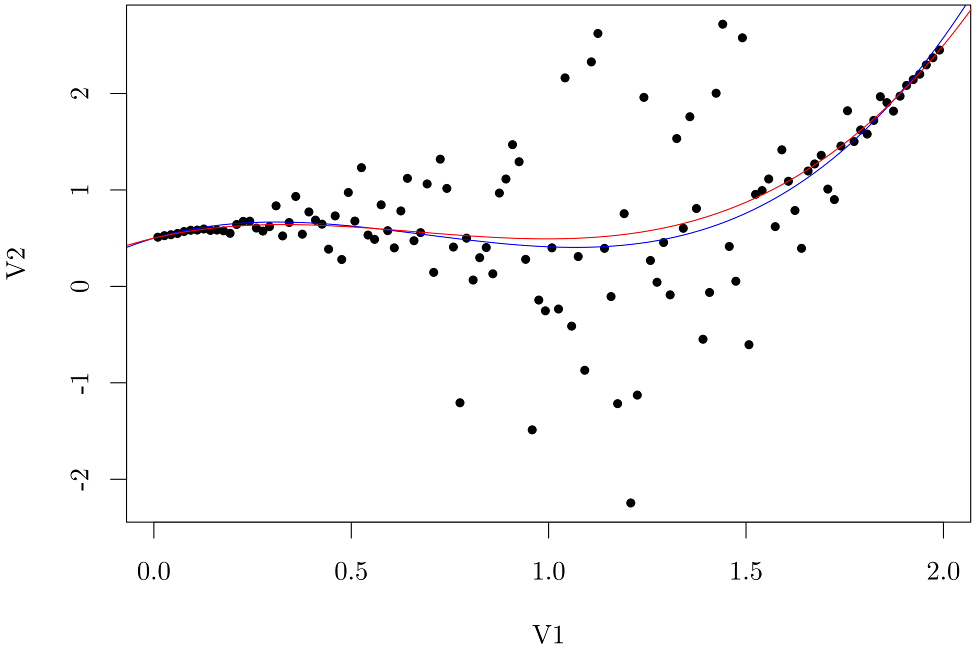
The value for $\hat{\beta}_{OLS}$ that we find using these uncorrelated errors is exactly the same as with the uncorrelated errors:

$$\hat{\beta}_{OLS} = (3.923161, 1.214355)^T$$

This is expected because we used a unitary transformation to get uncorrelated errors and by filling this in the expression for the Sum of squares of the remainders we find:

$$\begin{aligned} RSS(\hat{\beta}_{Uncorrelated}) &= (UY - UX\beta)^T(UY - UX\beta) \\ &= (Y - X\beta)^T U^T U(Y - X\beta) \\ &= (Y - X\beta)^T I(Y - X\beta) \\ &= RSS(\hat{\beta}_{Correlated}) \end{aligned}$$

which means the RSS for the correlated and the uncorrelated errors will have the same optimal $\hat{\beta}$.



Figuur 4: fits of a cubic function using ordinary least square (blue) and weighted least squares (red)

Question 3

(a)

Comput the weighted least square estimator $\hat{\beta}_{WLS}$

The values for the fitted coefficients with ordinary least squares are: $[0.4994, 1.1985, -2.4959, 1.2082]$. In figure 4 the fit is plotted along with the data.

(b)

Suppose we further know that the errors are Gaussian but heteroscedastic as follows:

$$\sigma(x) = \begin{cases} x^2 & \text{if} x \in [0, 4/3] \\ 4(x-2)^2 & \text{if} x \in [4/3, 2] \end{cases}$$

Compute the weighted least square estimator $\hat{\beta}_{WLS}$.

The coeffitients fitted with the weighted least square algorithm are $[0.5003, 0.9669, -1.9655, 0.9910]$.

(c)

Compare $\hat{\beta}_{WLS}$ to $\hat{\beta}_{OLS}$ and with the true parameter i.e. $\beta = (0.5, 1, -2, 1)^T$

The coefficients that are calculated using the least squares method are closer to the real values than the ones calculated using the ordinary least squares. To really say something about the difference and how well they match the given real values of the function we need more information about the distribution of the parameters. In figure 4 the two fits are plotted together with the data to fit through.

(d)

Determine the distribution of $\hat{\beta}_{WLS}$

To derive the distribution of $\hat{\beta}_{WLS}$ we start from the distribution for the errors, which is known:

$$\epsilon \sim N_n(0, W)$$

Where W is a diagonal matrix with at position W_{ii} the variance of ϵ_i. This is the same matrix W used in the formula to calculate the estimate for β with the weighted least squares method. This formula is:

$$\hat{\beta}_{WLS} = (X^T W^{-1} X)^{-1} X^T W^{-1} Y \tag{7}$$

here Y is the vector with samples used to calculate the least squares estimator. X is the matrix as defined on page 193 in the course notes, this matrix is deterministic. Y can be rewritten as follows:

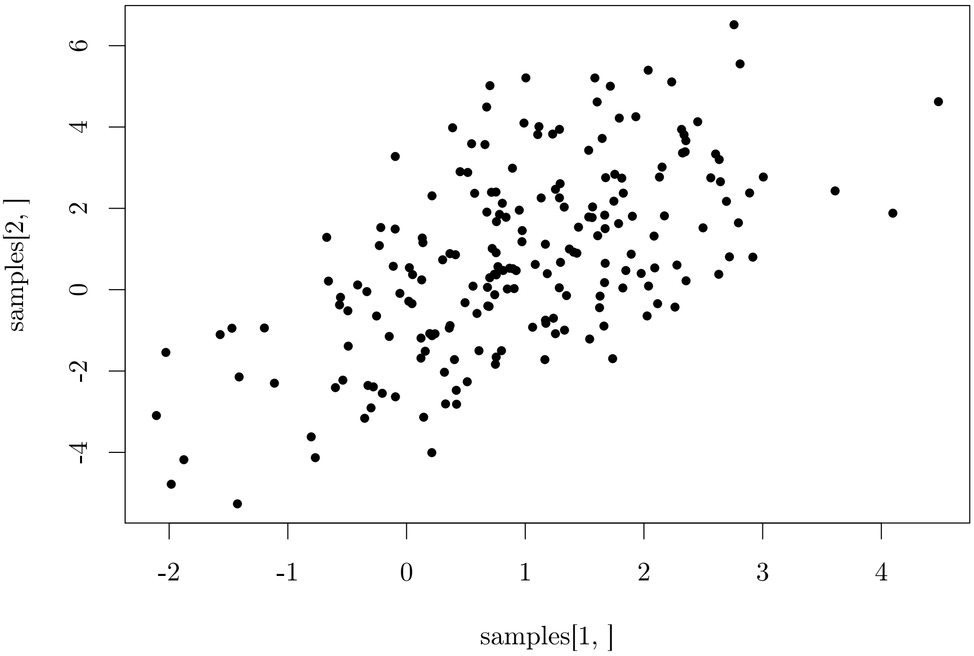
$$Y = X\beta + \epsilon$$

substituting this into equation 7 yields:

$$\begin{aligned}
\hat{\beta}_{WLS} &= (X^T W^{-1} X)^{-1} X^T W^{-1} (X\beta + \epsilon) \\
&= (X^T W^{-1} X)^{-1} (X^T W^{-1} X)\beta + (X^T W^{-1} X)^{-1} X^T W^{-1} \epsilon \\
&= \beta + X^{-1} W X^{-T} X^T W^{-1} \epsilon \\
&= \beta + X^{-1} \epsilon
\end{aligned}$$

We see this is a linear transformation of a multivariate normal distribution with mean **0** and as variance-covariance matrix $\Sigma_\epsilon = W$. Again using the results from section 6.2, page 172 in the course notes we find that the distribution of $\hat{\beta}_{WLS}$ is a multivariate normal distribution with mean β and variance-covariance matrix $X^{-1} W X^{-T} = (X^T W^{-1} X)^{-1}$. We conclude that:

$$\hat{\beta}_{WLS} \sim N_p(\beta, (X^T W^{-1} X)^{-1})$$



Figuur 5: Projection on the xy-plane of the sample from the multivariate normal distribution

Question 4

Let $\mathbf{X}$ be a 3-dimensional Gaussian vector with parameters

$$\mu = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1.25 & 1.50 & 0.5 \\ 1.5 & 5.25 & 3.5 \\ 0.5 & 3.5 & 3.0 \end{pmatrix}$$

(a)

Produce $n = 200$ simulations of $\mathbf{X}$

To generate a sample from a multi-normal distribution we work analogues to the derivation in section 6.2 in the course notes, but backwards. We are given the variance-covariance matrix Σ_X of a multivariate normal model. This matrix must be positive definite, so the square root of this matrix exists. Let A be a matrix such that $AA^T = \Sigma_X$. This matrix can be found by diagonalizing Σ_X: $\Sigma_X = U\Lambda U^T$ and than taking $A = U\sqrt(\Lambda)$. We can do this because all eigenvalues of Σ_X are positive. The result from section 6.2 now tells us that $X = AY + \mu$ where Y is a vector of standard normal distributions. This observation can be used to draw a sample from X by drawing n independent samples from a standard normal distribution, filling them in in a vector y and the calculating the sample from X as $x = Ay + \mu$. This was implemented in R

(b)

Compute $P(X_1 > 1|X_2 = 1, X_3 = -2)$ and $P(X_1 > 1|X_2 + X_3 = -1)$.

To compute these conditional chances we use proposition 6.2 from the course notes with the following vectors and matrices:

$$\Sigma_{11} = (1.25) \qquad \Sigma_{12} = (1.50, 0.5)$$

$$\Sigma_{21} = \begin{pmatrix} 1.50 \\ 0.50 \end{pmatrix} \qquad \Sigma_{22} = \begin{pmatrix} 5.25 & 3.5 \\ 3.5 & 3.0 \end{pmatrix}$$

$$\mathbf{X}_1 = (\mathbf{X}_1) \qquad \mathbf{x}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

When we fill in these values in the the formula for the distribution of the conditional distribution $f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)$ we find the following distribution:

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) = N_1(1, 0.446) \tag{8}$$

The probability $P(X_1 > 1|X_2 = 1, X_3 = -2)$ is equal to 0.5 because of the symetry of the conditional probability distribution around 1.

To calculate the probability $P(X_1 > 1|X_2 + X_3 = 1)$ we first transform the random vector $\mathbf{X}$ to the vector $\mathbf{Y} = (X_1, X_2 + X_3)^T = \mathbf{AX}$ where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Using part (b) of proposition 6.2 we find that the distribution of $\mathbf{Y}$ is given by:

$$\mathbf{Y} \sim N(A\mu, \mathbf{A}\Sigma\mathbf{A}^T) = N(...)$$

Observe that $P(X_1 > 1|X_2 + X_3 = -1) = P(Y_1 > 1|Y_2 = -1)$. We can again use part (c) of proposition 6.2 to calculate this probability. This yields the following conditional probability distribution:

$$f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1|\mathbf{y}_2 = -1) = N(1, 0.9877)$$

Therefore $P(X_1 > 1|X_2 + X_3 = -1) = 0.5$.

(c)

Let $\mathbf{Y} = (X_1, X_2)^T$. Represent graphically the density contours that comprise 95% of the probability mass of $\mathbf{Y}$

To determine the density function of $\mathbf{Y} = (X_1, X_2)^T$ We make again use of proposition 6.2 (c) with the following matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

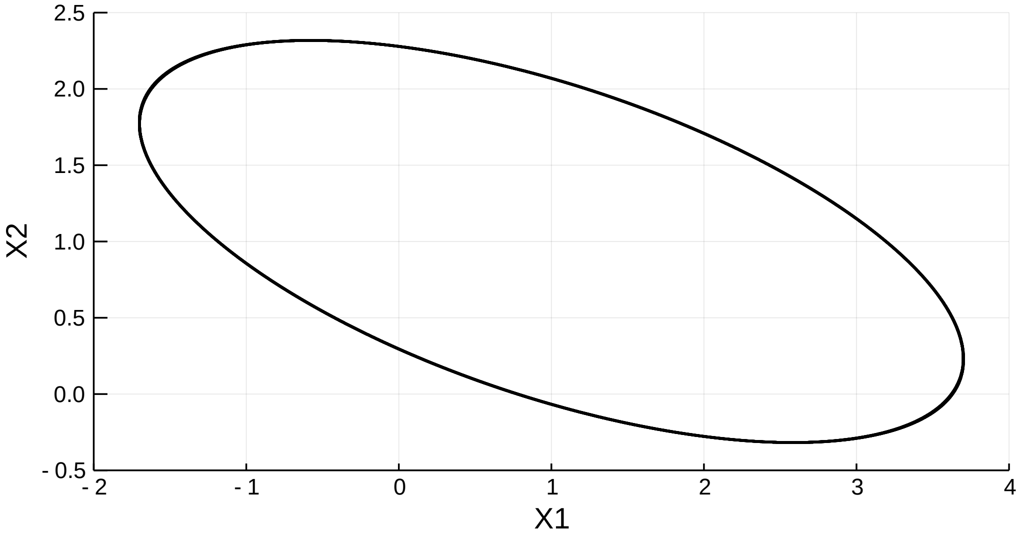
From which we find the probability distribution of $\mathbf{Y}$

$$\mathbf{Y} \sim N_2\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.25 & 1.50 \\ 1.50 & 5.25 \end{pmatrix}\right)$$

We can now use proposition 6.1 to find the contours of of this density function. The contour that comprises 95% of the probability mass is given by the following equation:

$$Q_{\chi^2_2}(0.95) = (\mathbf{Y} - \mu)^T\Sigma^{-1}(\mathbf{Y} - \mu)$$

This region is plotted in figure



Figuur 6: The contour that comprises 95% of the density of the probability density function of the random vector $\mathbf{Y}$.

Acknowledgements

Rune Buckinx and Michaël Maex for helping to fix stupid mistakes in stupid R code Seppe for stupid discussions about how well the questions are asked Thibeau for organising a group crying session Robbe for genuinly good input about trying to solve stupid questions