Week 2 Solutions

Semester 1, 2025

Computer Laboratory Session Activity

Exercise

1. (a) Consider the model $y = x\beta + u$ where the observed variables y and x, and the unobserved disturbance u are all n-vectors. Show that the OLS estimator for β is

$$\hat{\beta} = \frac{x'y}{x'x}. \tag{1}$$

Solution:
The OLS estimator is defined according to

$$\hat{\beta} = \underset{\beta}{\text{argmin}}(y - x\beta)'(y - x\beta) = \underset{\beta}{\text{argmin}}[y'y - 2x'y\beta + x'x\beta^2] = \underset{\beta}{\text{argmin}}\, S(\beta),$$

say. Now, $S(\beta)$ is everywhere continuous and differentiable and so calculus is our friend here, meaning that our solution will satisfy the first order condition

$$0 = \left.\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta}\right|_{\beta=\hat{\beta}} = -2x'y + 2x'x\hat{\beta} \implies x'x\hat{\beta} = x'y \implies \hat{\beta} = \frac{x'y}{x'x},$$

as required.

(b) Suppose that, in (1), $y \sim \mathrm{N}(0, I_n)$ and that y and x are statistically independent. Then, show that

$$\alpha = \frac{x'y}{(x'x)^{1/2}} \sim \mathrm{N}(0, 1)$$

and that α is independent of x.

Solution:
If y and x are statistically independent then the conditional distribution of y given x is the same as the unconditional distribution of y. That is,

$$y \sim \mathrm{N}(0, I_n) \implies y \mid x \sim \mathrm{N}(0, I_n).$$

Consequently,

$$\alpha \mid x = \left.\frac{x'y}{(x'x)^{1/2}}\right| x \sim \mathrm{N}\left(\frac{x'0}{(x'x)^{1/2}}, \frac{x'}{(x'x)^{1/2}} I_n \left(\frac{x'}{(x'x)^{1/2}}\right)'\right) = \mathrm{N}(0, 1).$$

Noting that the conditional distribution of $\alpha \mid x$ does not depend on x, it follows that it is also the unconditional distribution of α and that α is independent of x.

(c) Further suppose that $x \sim \mathrm{N}(0, I_n)$. Since $\hat{\beta} = \alpha/(x'x)^{1/2}$, where $\alpha \sim \mathrm{N}(0,1)$ and is independent of x, and since $x'x \sim \chi^2_n$, deduce that $n^{1/2}\hat{\beta} \sim t_n$.

Solution:

We have seen that α is independent of x, because its conditional density is not a function of x. Consequently, it will be independent of functions of x and, specifically, of $(x'x)^{1/2}$. Given that $x \sim \mathrm{N}(0, I_n)$ we know that $x'x \sim \chi^2_n$. Therefore,

$$n^{1/2}\hat{\beta} = \frac{n^{1/2}\alpha}{(x'x)^{1/2}} = \frac{\alpha}{(x'x/n)^{1/2}} \sim \frac{N(0,1)}{\sqrt{\chi^2_n/n}} = t_n,$$

as required, where the final equality follows as a consequence of the numerator and denominator.

2. Suppose that $X = [X_1, X_2, X_3]' \sim \mathrm{N}(\mu, \Sigma)$. If

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & 0 \\ \rho^2 & 0 & 1 \end{bmatrix}$$

show that the conditional distribution of $[X_1, X_2]'$ given X_3 has mean vector

$$[\mu_1 + \rho^2(x_3 - \mu_3), \mu_2]'$$

and covariance matrix

$$\begin{bmatrix} 1-\rho^4 & \rho \\ \rho & 1 \end{bmatrix}.$$

Solution:

Let $Y_1 = [X_1, X_2]'$ and $Y_2 = X_3$, so that $Y = [Y_1', Y_2]' = X$. Equally, let

$$\Omega = \begin{bmatrix} \Omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix},$$

where $\omega_{22} = 1$, $\omega_{12} = [\rho^2, 0]' = \omega_{21}'$ and

$$\Omega_{11} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

so that $\Omega = \Sigma$. If we partition $\mathrm{E}[X] = \mu = [\mu_1, \mu_2, \mu_3]'$, say, and $\mathrm{E}[Y] = \mu = [\mu_{Y_1}', \mu_{Y_2}]'$, with $\mu_{Y_1} = [\mu_1, \mu_2]'$ and $\mu_{Y_2} = \mu_{Y_2}$, we see that this is less a transformation than a re-grouping of the elements of X. (Alternatively, it is a transformation for which the Jacobian of transformation is unity.) From our results on the conditional distribution of jointly normally distributed random variables, we see that $Y_1 \mid Y_2 \sim \mathrm{N}\left(\mu_{Y_1|Y_2}, \Omega_{11\cdot 2}\right)$, where

$$\begin{aligned} \mu_{Y_1|Y_2} = \mu_{Y_1} + \omega_{12}\omega_{22}^{-1}(Y_2 - \mu_{Y_2}) &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \rho^2 \\ 0 \end{bmatrix} (1)^{-1}(X_3 - \mu_{X_3}) \\ &= \begin{bmatrix} \mu_1 + \rho^2(X_3 - \mu_{X_3}) \\ \mu_2 \end{bmatrix} \end{aligned}$$

and

$$\Omega_{11\cdot 2} = \Omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{21} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} \rho^2 \\ 0 \end{bmatrix} (1)^{-1}[\rho^2, 0] = \begin{bmatrix} 1-\rho^4 & \rho \\ \rho & 1 \end{bmatrix},$$

as required.

3. If X_1, X_2, and X_3 are iid. $\mathrm{N}(\mu, \Sigma)$ p-vectors, and if $Y_1 = X_1 + X_2$, $Y_2 = X_2 + X_3$, and $Y_3 = X_1 + X_3$, then obtain

   (a) the conditional distribution of Y_1 given Y_2; and

   (b) the conditional distribution of Y_1 given Y_2 and Y_3.

   Hint: In lectures you were given a formula for K^{-1}, the inverse of a partitioned matrix K, where

$$K = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

   This formula that was given is very general and works even when B and C are not square. All it required was that A was non-singular. However, when B and C are also non-singular there is considerable simplification available. In our case there is even greater simplification arising from the facts that (i) $A = D$ and $B = C$, and (ii) that $A = \alpha B$ for some scalar constant α. Hence, we need the inverse of a matrix of the form

$$K = \begin{bmatrix} \alpha B & B \\ B & \alpha B \end{bmatrix},$$

   with B non-singular. Now, it must be the case that $KK^{-1} = I$. If we partition K^{-1} conformably with K then we can write

$$K^{-1} = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

   say, and so $KK^{-1} = I$ implies a set of 4 equations in 4 unknowns (E, F, G, H):

$$\alpha BE + BG = I, \qquad \alpha BF + BH = 0,$$
$$BE + \alpha BG = 0, \qquad BF + \alpha BH = I,$$

   From the zero equations we see that

$$H = -\alpha F, \qquad E = -\alpha G,$$

   which we can substitute back into the identity equations to obtain

$$-\alpha^2 BG + BG = I \implies (1-\alpha^2)BG = I \implies G = (1-\alpha^2)^{-1}B^{-1},$$
$$BF - \alpha^2 BF = I \implies (1-\alpha^2)BF = I \implies F = (1-\alpha^2)^{-1}B^{-1}.$$

   We can substitute these expressions for G and H into those for E and F above to obtain

$$K^{-1} = \begin{bmatrix} -\frac{\alpha}{1-\alpha^2}B^{-1} & \frac{1}{1-\alpha^2}B^{-1} \\ \frac{1}{1-\alpha^2}B^{-1} & -\frac{\alpha}{1-\alpha^2}B^{-1} \end{bmatrix}$$

   Note that K^{-1} retains the symmetry of K and, like K, all four partitions of K^{-1} are also non-singular.

   Solution:
   Because X_1, X_2, and X_3 are independent we know that $X = [X_1', X_2', X_3']' \sim \mathrm{N}(\delta, \Omega)$, where $\delta = [\mu', \mu', \mu']'$ and $\Omega = \mathrm{diag}(\Sigma, \Sigma, \Sigma)$. Now, $Y = [Y_1', Y_2', Y_3']' = AX$, where

$$A = \begin{bmatrix} I_p & I_p & 0 \\ 0 & I_p & I_p \\ I_p & 0 & I_p \end{bmatrix}.$$

From the properties of Normal random variables we see that

$$\mathrm{E}[Y] = A\,\mathrm{E}[X] = A\delta = 2\delta$$

and

$$\mathrm{Var}[Y] = A\,\mathrm{Var}[X]\,A' = \begin{bmatrix} I_p & I_p & 0 \\ 0 & I_p & I_p \\ I_p & 0 & I_p \end{bmatrix} \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma & 0 \\ 0 & 0 & \Sigma \end{bmatrix} \begin{bmatrix} I_p & 0 & I_p \\ I_p & I_p & 0 \\ 0 & I_p & I_p \end{bmatrix}$$

$$= \begin{bmatrix} 2\Sigma & \Sigma & \Sigma \\ \Sigma & 2\Sigma & \Sigma \\ \Sigma & \Sigma & 2\Sigma \end{bmatrix} = 2I_3 \otimes \Sigma = \Omega \text{ (say)}.$$

(a) We can obtain the conditional distribution of Y_1 given Y_2 from the joint distribution of Y_1 and Y_2, which is just the marginal distribution of Y_1 and Y_2 from the joint distribution of Y. That is,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathrm{N}\left(2\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 2\Sigma & \Sigma \\ \Sigma & 2\Sigma \end{bmatrix}\right).$$

Hence, using the formulae provided in the Solution to Question 2, with appropriate partitioning of the parameters here,

$$Y_1 \mid Y_2 \sim \mathrm{N}\left(2\mu + \Sigma(2\Sigma)^{-1}(y_2 - 2\mu), 2\Sigma - \Sigma(2\Sigma)^{-1}\Sigma\right) = \mathrm{N}\left(\tfrac{1}{2}(y_2 + 2\mu), \tfrac{3}{2}\Sigma\right).$$

Note that this conditional variance is smaller than the unconditional variance of Y_1 in the sense that $\mathrm{Var}[Y_1] - \mathrm{Var}[Y_1 \mid Y_2] = 2\Sigma - \frac{3}{2}\Sigma = \frac{1}{2}\Sigma > 0$.

(b) We can read the condition distribution of Y_1 given Y_2 and Y_3 directly from the joint distribution of Y as follows: $Y_1 \mid Y_2, Y_3 \sim \mathrm{N}\left(\mu_{1\cdot 23}, \Omega_{11\cdot(23)}\right)$, say, where

$$\mu_{1\cdot 23} = 2\mu + [\Sigma, \Sigma] \begin{bmatrix} 2\Sigma & \Sigma \\ \Sigma & 2\Sigma \end{bmatrix}^{-1} \left(\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} - 2\begin{bmatrix} \mu \\ \mu \end{bmatrix}\right) = \tfrac{1}{3}(y_2 + y_3 + 2\mu)$$

and

$$\Omega_{11\cdot(23)} = 2\Sigma - [\Sigma, \Sigma] \begin{bmatrix} 2\Sigma & \Sigma \\ \Sigma & 2\Sigma \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \\ \Sigma \end{bmatrix} = \tfrac{4}{3}\Sigma.$$

Note that this conditional variance is smaller than the unconditional variance of Y_1 in the sense that $\mathrm{Var}[Y_1] - \mathrm{Var}[Y_1 \mid Y_2, Y_3] = 2\Sigma - \frac{4}{3}\Sigma = \frac{2}{3}\Sigma > 0$. Further note that the more you condition on the greater the variance reduction, that is

$$\mathrm{Var}[Y_1 \mid Y_2] - \mathrm{Var}[Y_1 \mid Y_2, Y_3] = \tfrac{3}{2}\Sigma - \tfrac{4}{3}\Sigma = \tfrac{1}{6}\Sigma > 0.$$

4. R offers two commands `qqplot` for the construction of quantile-quantile plots: `qqnorm` and `qqplot`. Quantile-quantile plots, or Q-Q plots, as they are known, are a graphical device often used for comparing the quantiles of an empirical distribution with those of a theoretical distribution. Construct Q-Q plots comparing the following distributions against a t distribution with 5 degrees of freedom: N(0,1), t_{20}, t_{10}, t_5, t_2, and t_1. Describe the patterns that you observe and the lessons that you can take from the exercise.

Solution:

First, a few of comments.

(a) I said in Lecture 1 that I would sometimes use the tutorials as an opportunity for you to teach yourself something useful that does not fit well into the lecture scenario. Q-Q plots is one of those topics: really useful and a good opportunity to think about exactly what a distribution function is telling you, what quantiles are, and about probability in general. My observation is that, as a group, you are not uniformly strong in your understanding of the probability that was taught in first year. So this is a good chance to brush up on that.

(b) Daniel has provided a file, `Week2.R`, which covers the computing pretty well. Below I provide some discussion of what you might observe. I need to issue one disclaimer. I use Matlab much more than I use R and so the figures provided below were actually constructed in Matlab. They were then converted into tikz, which is a graphical format that LaTeX can deal with, using the Matlab add-in matlab2tikz.

(c) Finally, I won't give you the `.tex` files used to produce all the Exercises and Solutions as a matter of course but, if you ever want to see how something is done then just ask.

And now for the actual solution. Q-Q plots plot the quantiles of one distribution against those of some other distribution. Usually one of the sets of quantiles are those from some theoretical distribution, with the other set being from an empirical distribution. This need not be the case, as demonstrated by this exercise, where we are plotting one set of theoretical quantiles against another. The Q-Q plots originally requested can be found in Figures 1–6. As a bonus, Figures 7 and 8 plot the theoretical quantiles of the distributions of centred chi-squared random variables against those of the t_5 distribution. In particular, these figures compare the quantiles of the distributions of random variables X against those of a t_5 distribution where $X = \chi^2_3 - 3$ and $X = \chi^2_{30} - 30$, respectively. These are presented because all the other distributions considered are symmetric, whereas these distributions are skewed (to the right). As a further exercise you should explore what happens if one of the distributions is skewed to the left. Figures 9 and 10 look at what happens when one plots the quantiles of a distribution against those of a transformed version of the same underlying random variable. In particular, in Figure 9 we add a non-zero mean to the t_5 random variable and in Figure 10 we scale the variable by a constant. As a final note, the most common Q-Q plots have standard Normal distributions as the basis of comparison. This exercise illustrates the point that this need not be the case.

Moving forward, we note that all of the distributions considered have unbounded support, so it is just silly to attempt to cover quantiles over the entire support of the distributions. In these figures I have plotted points at 1% increments starting at the 1% quantile through to the 99 quantile, for each distribution. Moreover, each figure contains a solid red line which passes through the points defined by the first and third quartiles of each distribution. For ease of reference we shall, hereafter, simply refer to this as the red line.[1] The red line serves no real purpose beyond acting as a visual aid to look at the 'straightness' of the Q-Q plot. The

[1] Matlab actually provides a solid line joining these two points and then extrapolates that line to the extremes of the plot (which is done using a dash-dot pattern). In my plots I have deleted the shorter line segment (as it isn't actually visible under the Q-Q plot) and converted the dash-dot pattern to a solid line.

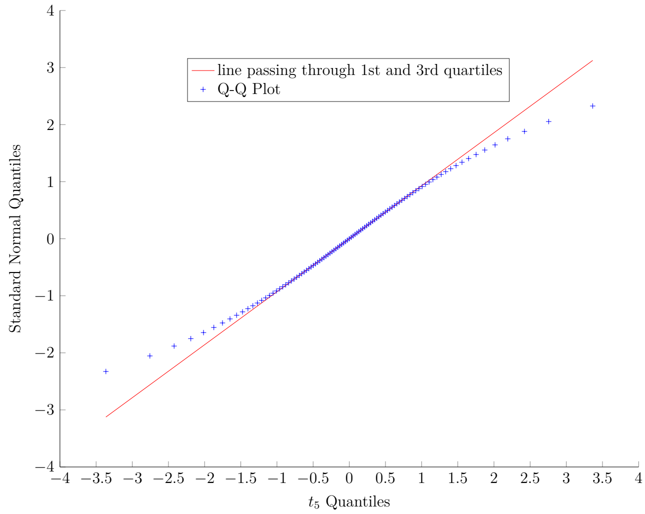

Figure 1: Q-Q Plot: N(0, 1) (or t_∞) versus t_5

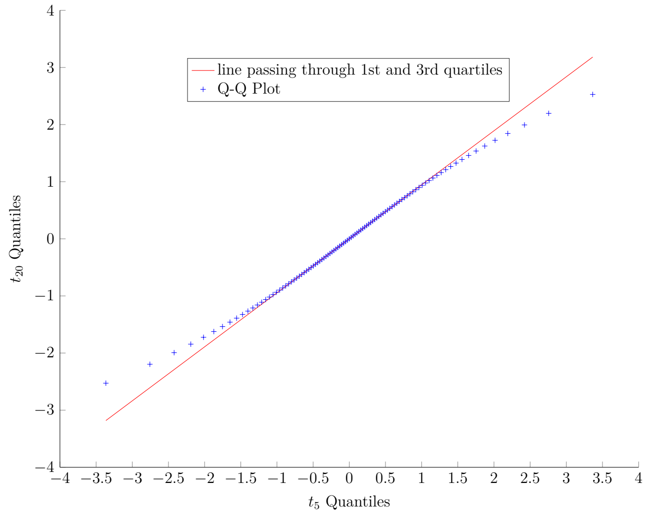

Figure 2: Q-Q Plot: t_{20} versus t_5

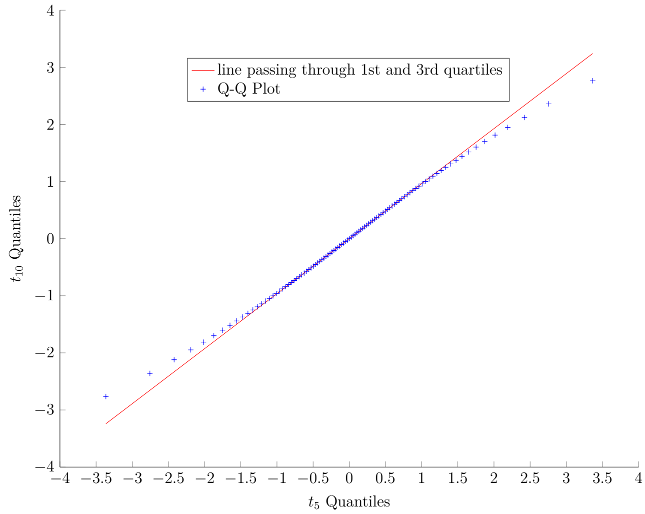

Figure 3: Q-Q Plot: t_{10} versus t_5

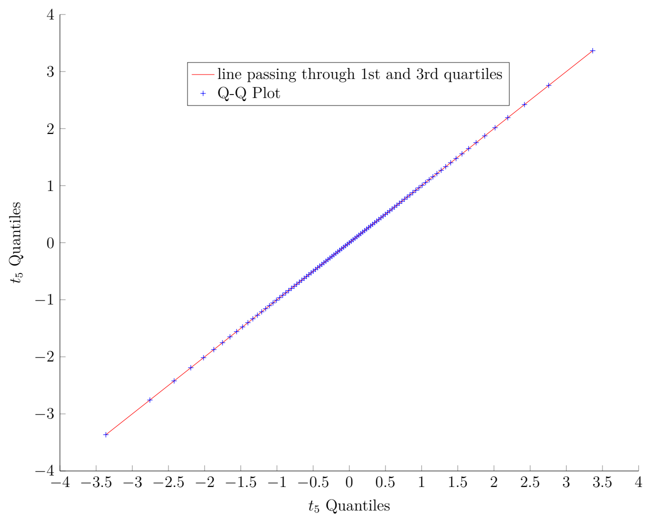

Figure 4: Q-Q Plot: t_5 versus t_5

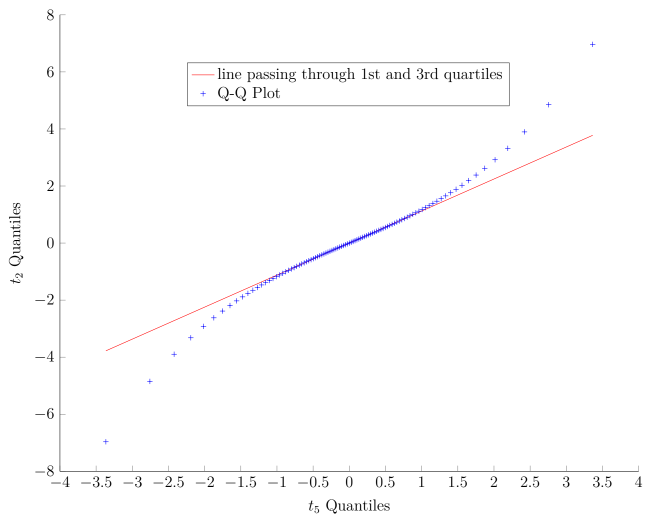

Figure 5: Q-Q Plot: t_2 versus t_5

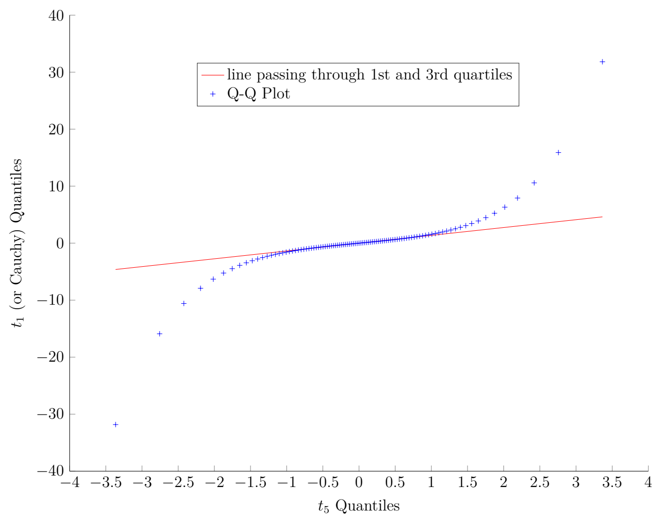

Figure 6: Q-Q Plot: t_1 (or Cauchy) versus t_5

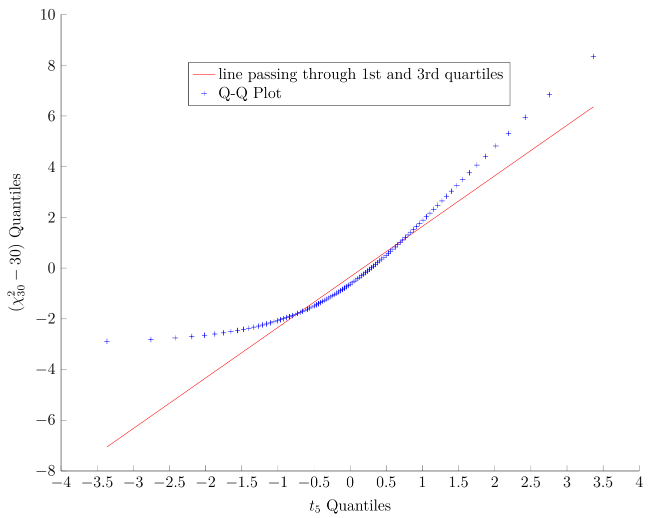

Figure 7: Q-Q Plot: $(\chi^2_3 - 3)$ versus t_5

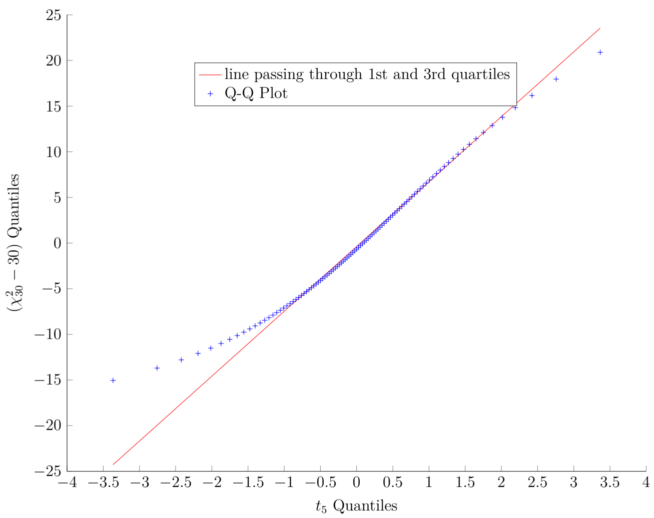

Figure 8: Q-Q Plot: $(\chi^2_{30} - 30)$ versus t_5

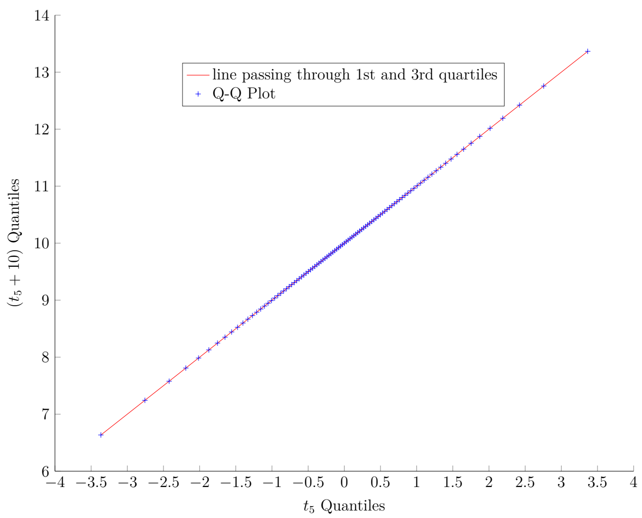

Figure 9: Q-Q Plot: $(t_5 + 10)$ versus t_5

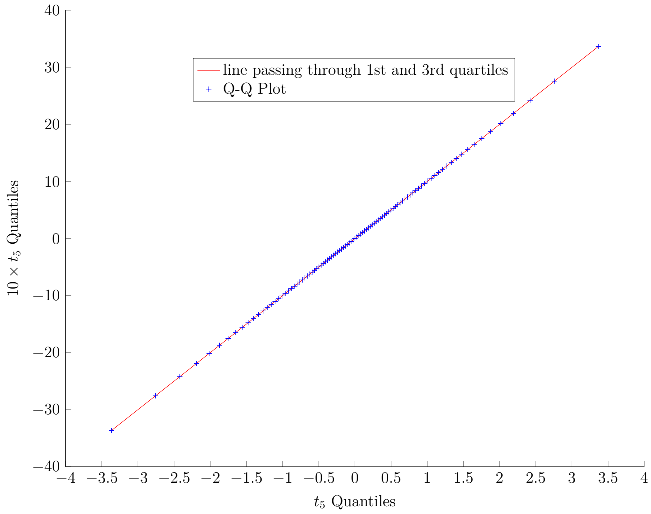

Figure 10: Q-Q Plot: $10 * t_5$ versus t_5

plots of quantile pairs are then marked by blue '+' symbols. All of the relevant commands are contained in the file **Exercise7.m**. Note that I edited the various .tex files created by the `matlab2tikz` command to change axis labels and to modify the legend. There were some other changes made to get things working. I would encourage you to run **Exercise7.m** and compare the output that you obtain with that used in generating this document, to see the sorts of changes that I have made. All of the LaTeX source files are available as a .zip file on the LMS.

Looking now at the patterns in the plots it makes sense to start at Figure 4, which plots the quantiles of a t_5 distribution against themselves. As you would expect, the '+' symbols all lie on the red line which, in this case, is a 45 degree line. In practice if you were to generate a sample of random draws from some distribution then you almost certainly would not observe such a perfect correspondence. In particular, the extremes of the distribution are often under-sampled and so you may get departures there.

Excluding the bonus plots for the moment, the remaining distributions can be broken up into two groups. The first group consists of t distributions with larger degrees of freedom that the reference distribution, namely the standard Normal (t_∞), the t_{20} and the t_{10}. These distributions all have relatively thin tails and so their quantiles are somewhat more tightly clustered about zero than those of the t_5 distribution. Looking at Figures 1–3, we see essentially the same pattern in each. To the left of zero, for a given percentage, the quantile for a t_5 distribution is a smaller (more negative) number than that for the thinner-tailed distributions, e.g. the 1% quantiles for the t_5 and the $\mathrm{N}(0,1)$ are approximately -3.3649 and -2.3263, respectively. Similarly, to the right of zero, the quantiles of the t_5 distribution are larger numbers than those of the thinner-tailed distributions, e.g. the 99% quantiles for the t_5 and the $\mathrm{N}(0,1)$ are approximately 3.3649 and 2.3263, respectively. Because all of these distributions are symmetric about zero we see that they each have their 50% quantile at zero.

The other group of symmetric distributions is comprised of t distributions with fewer degrees of freedom than 5, namely the t_2 and the t_1 or Cauchy distributions, which have fatter tails than does the t_5 distribution. Here we might expect to see the patterns reversed in the tails of the distributions but, again, similarity around the point of symmetry. Not surprisingly this is exactly what we see. Interestingly, the evidence suggests that a Cauchy distribution has much fatter tails relative to a t_5 distribution than does a t_5 relative to a standard Normal distribution.

Figures 7 and 8 plot the quantiles of skewed distributions against the quantiles of the t_5 distribution, which is symmetric about zero. Observe that the skewed distributions have been centred so that these distributions also have zero mean. This is done by subtracting their respective degrees of freedom which, as we know, is the mean of a chi-squared random variable. For both of the skewed distributions there is a much greater probability mass down towards the lower bound of their support than for the t_5 distribution and this is far more pronounced at smaller degrees of freedom. Indeed, we know that as $\nu \to \infty$, $(\chi^2_\nu - \nu)/\sqrt{2\nu} \to \mathrm{N}(0,1)$ and so, at larger degrees of freedom the support of the skewed distribution is increasing and the quantity $\chi^2_\nu - \nu$ is becoming increasingly symmetric about zero. This is reflected in the Q-Q plot in Figure 8. As we move into the right-hand tails of these distributions, all the probability mass at the lower end of the support means that the right-hand tails of the distribution must become thin. Just how thin relative to the t_5 distribution differs with the degrees of freedom. In particular, we see that for the larger the degrees of freedom of the chi-squared variate, the smaller the probability mass at the lower bound of the support and hence the thicker the right-hand tail of the distribution. (You really need to get this off the scale of the y-axis as the location of the red line vis-a-vis the Q-Q plot is not helpful in this regard.) For example, the 99% quantile for $\chi^2_{30} - 30$ distribution is approximately 20.8922, that of the $\chi^2_3 - 3$ distribution is approximately 8.3449, whereas that for the t_5 distribution is only 3.3649 (approximately). We conclude that the probability mass around zero for the t_5 distribution is larger than that at the lower bound of the support of the skewed distributions.

Our final experiments involve transformations of the t_5 variates by either the addition of a positive constant (10 in Figure 9) or by scaling the variate by a positive constant (10 in Figure 10). The remarkable thing is that, as was seen in Figure 4, the Q-Q plots continue to live on a straight-line, albeit no longer a 45 degree line. Nevertheless, affine transformations of random variables, transformations of the type $Y = \alpha + \beta X$ will have the same impact on the quantiles of the distribution. More complicated transformations will have different impacts. For example, the square of a t_5 random variable won't be chi-squared, but it will be something similar and so Figures 7 and 8 provide some insight to what might happen.