TOPIC. Representing and quantile functions. This section deals with the quantiles of a random variable. We start by discussing how quantiles are related to the inverse df of the preceding section, and to the closely related notion of a representing function. We then look at a useful parametric family of distributions, the so-called Tukey family, that are defined in terms of their quantiles. Finally, we present a graphical technique for comparing two distributions, the so-called Q/Q-plot.

Representing functions. Let X be a random variable with df F. A mapping R from $(0,1)$ to $\mathbb{R}$ such that

(R1) R is nondecreasing and

(R2) $R(U) \sim X$ whenever $U \sim \text{Uniform}(0,1)$

is called a ***representing function*** for X. According to the IPT Theorem (Theorem 1.5), the left-continuous inverse F^* to F is a representing function. The motivating idea behind a representing function R is the use we put F^* to in proving Theorem 1.6, namely, if you want to prove something about X, it may be helpful to try representing X as $R(U)$ for a standard uniform variable U.

Theorem 1. *Suppose R is a representing function for X and $Y := T(X)$ is a transformation of X.*

(1) *If T is nondecreasing, then the mapping $u \rightsquigarrow T\big(R(u)\big)$ is a representing function for Y.*

(2) *If T is nonincreasing, then the mapping $u \rightsquigarrow T\big(R(1-u)\big)$ is a representing function for Y.*

Proof • (1) The mapping $u \rightsquigarrow T\big(R(u)\big)$ is nondecreasing in u and for $U \sim \text{Uniform}(0,1)$, $R(U) \sim X \Longrightarrow T\big(R(U)\big) \sim T(X) = Y$.

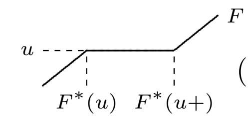
(2) In this case the mapping $u \rightsquigarrow T\big(R(1-u)\big)$ is nondecreasing in u for $U \sim \text{Uniform}(0,1)$, $1-U \sim U \Longrightarrow R(1-U) \sim R(U) \sim X \Longrightarrow T\big(R(1-U)\big) \sim T(X) = Y$. ∎

(R1) R is nondecreasing. (R2) $U \sim \text{Uniform}(0,1) \Longrightarrow R(U) \sim X$.

The theorem covers the important special case where $T(x) = a+bx$; this transformation is nondecreasing if $b \geq 0$, and nonincreasing if $b \leq 0$. There is always at least one representing function for X, namely F^*. Are there any others? That question is answered by:

Theorem 2. *Suppose X has df F and left-continuous inverse df F^*. A mapping R from $(0,1)$ to $\mathbb{R}$ is a representing function for X if and only if*

$$F^*(u) \leq R(u) \leq F^*(u+) \ \text{ for all } u \in (0,1). \tag{1}$$



Consequently F^* is the only representing function iff F is strictly increasing over $\{\, x \in \mathbb{R} : 0 < F(x) < 1 \,\}$ (see Exercise 1).

Proof • ($\Longleftarrow$). Suppose (1) holds.

⋆ (R1) *holds*: Indeed, suppose $0 < u < w < 1$. Then

$$\begin{aligned} R(u) &\leq F^*(u+) && \text{(by (1) for } u) \\ &\leq F^*(w) && \text{(since } F^* \text{ is nondecreasing)} \\ &\leq R(w) && \text{(by (1) for } w). \end{aligned}$$

⋆ (R2) *holds*: Indeed, suppose $U \sim \text{Uniform}(0,1)$. Then $F^*(U) \sim X$. I am going to show that

$$P[\, R(U) = F^*(U)\,] = 1; \tag{2}$$

by Exercise 2 this implies $R(U) \sim F^*(U)$ and hence $R(U) \sim X$, as required. To get (2), let $J = \{u \in (0,1) : F^*(u) \neq F^*(u+)\,\}$ be the set of jumps of F^*. Exercise 1.9 asserts that J is countable. Thus

$$\begin{aligned} P[\, R(U) \neq F^*(U)\,] &\leq P[\, U \in J\,] && \text{(by (1))} \\ &= \sum\nolimits_{u \in J} P[\, U = u\,] = 0 && \text{(since } P[\, U = u\,] = 0 \text{ for all } u). \end{aligned}$$

• ($\Longrightarrow$). Let R be a representing function for X, and let $U \sim \text{Uniform}(0,1)$.

$\star$ *$F^*(u) \le R(u)$ for all $u \in (0,1)$*: Indeed, let $u \in (0,1)$ be given. By the switching formula (1.5), $F^*(u) \le R(u) \iff u \le F\big(R(u)\big)$. Now use

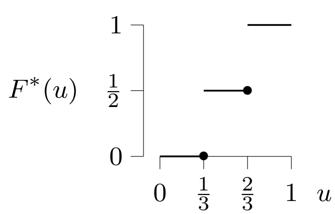
$$\begin{aligned} F\big(R(u)\big) &= P[\,X \le R(u)\,] && \text{(since } X \text{ has df } F) \\ &= P[\,R(U) \le R(u)\,] && \text{(by (R2))} \\ &\ge P[\,U \le u\,] && \text{(by (R1))} \\ &= u && \text{(since } U \text{ is uniform).} \end{aligned}$$

$\star$ *$R(u) \le F^*(u+)$ for all $u \in (0,1)$*: Indeed, let $u \in (0,1)$ be given. By Exercise 1.8, $F^*(u+) \ge R(u) \iff u \ge F\big(((R(u))-\big)$. Now use

$$\begin{aligned} F\big((R(u))-\big) &= P[\,X < R(u)\,] \\ &= P[\,R(U) < R(u)\,] \\ &\le P[U < u] = u\,. \end{aligned}$$
∎

As the proof pointed out, there can be at most countably many values of u such that $F^*(u) < F^*(u+)$.

Example 1. Suppose X takes the values 0, 1/2, and 1 with probabilities 1/3 each. We have seen that F^* has the following graph:



(R1) R is nondecreasing. (R2) $U \sim \text{Uniform}(0,1) \Longrightarrow R(U) \sim X$.
(1): $F^*(u) \le R(u) \le F^*(u+)$.

Consequently $R\colon (0,1) \to \mathbb{R}$ is a representing function for X if and only if

$$R(u) = \begin{cases} 0, & \text{if } 0 < u < 1/3, \\ r_1, & \text{if } u = 1/3, \\ 1/2, & \text{if } 1/3 < u < 2/3, \\ r_2, & \text{if } u = 2/3, \\ 1, & \text{if } 2/3 < u < 1 \end{cases}$$

for some numbers $r_1 \in [0, 1/2]$ and $r_2 \in [1/2, 1]$. •

Suppose R is a representing function for a random variable X with df F. Let n be a positive integer and let U_1, U_2, …, U_n be independent random variables, each uniformly distributed over $(0,1)$. Then

$$X_1 := R(U_1),\ X_2 := R(U_2),\ \ldots,\ X_n := R(U_n)$$

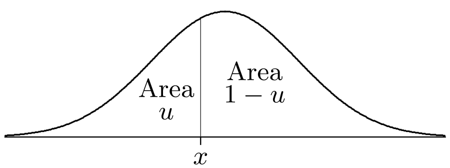
are independent and each distributed like X; in brief, $X_1, \ldots, X_n$ is a random sample of size n from F. Now let $X_{(1)}$, $X_{(2)}$, …, $X_{(n)}$ be the random variables resulting from arranging the X_i's in increasing order; these are called the ***order statistics*** for the X_i's . Similarly, let $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$ be the order statistics for the U_i's. Since R is nondecreasing we have

$$X_{(1)} := R(U_{(1)}),\ X_{(2)} := R(U_{(2)}),\ \ldots,\ X_{(n)} := R(U_{(n)})$$

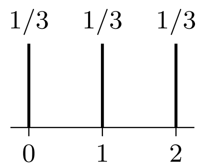
Consequently to study properties of order statistics, it often suffices to first work out the uniform case (i.e., for $X \sim U$) and then transform the result through R. This is a standard technique.

Quantiles. We turn now to the quantiles of a random variable X. Here are a couple of examples to motivate the general definition given below.

Case A: $X \sim N(0,1)$



Case B: $X \sim$ Uniform on 1, 2, 3



In case A, X has a density. For $u \in (0,1)$, the u^{th}-quantile of X is defined to be the number x such that the area under the density to the left of x equals u, or, equivalently, the area to the right of x equals $1-u$. This is the same as requiring

$$P[\, X \le x\,] = u \quad \text{and} \quad P[\, X \ge x\,] = 1-u. \tag{3}$$

In case B, X has a probability mass function. For $u = 0.5$, the u^{th}-quantile, or median, of X should clearly be taken to be $x = 1$. However, (3) doesn't hold for this u and x, rather we have

$$P[\, X \le x\,] \ge u \quad \text{and} \quad P[\, X \ge x\,] \ge 1-u. \tag{4}$$

We make this the general definition. For any random variable X, be it continuous, discrete, or whatever, a ***u***

be it continuous, discrete, or whatever, a ***$\boldsymbol{u}^{\text{th}}$-quantile*** of X is defined to be a number x such that (4) holds. According to the following theorem, there is always at least one such x; in some situations (as in Case B above, with $u = 1/3$ or $u = 2/3$) there may be more than one. Obviously (3) always implies (4). However, (4) implies (3) only when

$$P[\, X = x\,] = 0; \tag{5}$$

see Exercise 3. A mapping Q from $(0,1)$ is called a ***quantile function*** for X if $Q(u)$ is a u^{th}-quantile for X, for all $u \in (0,1)$.

Theorem 2: $R\colon (0,1) \to \mathbb{R}$ is a representing function for a random variable X with df F iff $F^*(u) \le R(u) \le F^*(u+)$ for all $u \in (0,1)$.

x is a u^{th}-quantile for X iff $P[\, X \le x\,] \ge u$ and $P[\, X \ge x\,] \ge 1-u$.

$Q\colon (0,1) \to \mathbb{R}$ is a quantile function for X iff $Q(u)$ is a u^{th}-quantile for each $u \in (0,1)$.

Theorem 3. *Let X be a random variable. A mapping Q from $(0,1)$ to $\mathbb{R}$ is a quantile function for X if and only if Q is a representing function for X.*

Proof Let F be the df of X and F^* the left-continuous inverse to F. A number x is a u^{th} quantile for X if and only

$$u \le P[\, X \le x] = F(x) \iff F^*(u) \le x \tag{6}$$

and

$$\begin{aligned} &1-u \le P[\, X \ge x] = 1 - P[\, X < x\,] = 1 - F(x-) \\ &\quad \iff u \ge F(x-) \iff F^*(u+) \ge x; \end{aligned} \tag{7}$$

here (6) used the switching formula (1.5) and (7) used the alternate switching formula given in Exercise 1.8. Consequently,

$$\begin{aligned} &Q \text{ is a quantile function for } X \\ &\quad \iff F^*(u) \le Q(u) \le F^*(u+) \ \text{ for all } u \in (0,1) \\ &\quad \iff Q \text{ is a representing function for } X, \end{aligned}$$

the last step following from Theorem 2. ∎

Theorems 1 and 3 imply

Theorem 4. *Suppose Q is a quantile function for X and $Y = T(X)$.*

(1) *If T is nondecreasing, then the mapping $u \rightsquigarrow T\big(Q(u)\big)$ is a quantile function for Y.*

(2) *If T is nonincreasing, then the mapping $u \rightsquigarrow T\big(Q(1-u)\big)$ is a quantile function for Y.*

The Tukey family. For $\lambda \in (-\infty, \infty)$ let $R_\lambda \colon (0,1) \to \mathbb{R}$ be defined by

$$R_\lambda(u) := \begin{cases} \dfrac{u^\lambda - (1-u)^\lambda}{\lambda}, & \text{if } \lambda \neq 0 \\ \log\Big(\dfrac{u}{1-u}\Big), & \text{if } \lambda = 0 \end{cases} \tag{8}$$

for $0 < u < 1$. Note that

$$\frac{d}{du} R_\lambda(u) = u^{\lambda-1} + (1-u)^{\lambda-1} > 0$$

for each u, so R_λ is continuous and strictly increasing. Consequently if $U \sim \text{Uniform}(0,1)$, then $X_\lambda := R_\lambda(U)$ is a random variable having R_λ as its (unique) representing (= quantile) function. The distribution of X_λ is called the ***Tukey distribution*** $\boldsymbol{\mathfrak{T}(\lambda)}$ ***with shape parameter*** $\boldsymbol{\lambda}$ (and scale parameter 1). The Tukey family is $\{\, \mathfrak{T}(\lambda) : -\infty < \lambda < \infty \,\}$. This family has a number of interesting/useful properties.

1°. It is very easy to simulate a draw from $\mathfrak{T}(\lambda)$: you just take a uniform variate u and compute $R_\lambda(u)$. The Tukey family is therefore well-suited for empirical robustness studies, and that is why Tukey introduced it.

2°. The distributions vary smoothly with λ. This is clear for $\lambda \neq 0$. Moreover

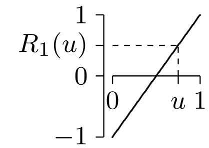
$$\begin{aligned} \lim_{\lambda \to 0} R_\lambda(u) &= \lim_{\lambda \to 0} \frac{u^\lambda - (1-u)^\lambda}{\lambda} \\ &= \lim_{\lambda \to 0} \frac{\frac{d}{d\lambda}\big(u^\lambda - (1-u)^\lambda\big)}{\frac{d}{d\lambda}\lambda} \qquad \text{(by l'Hospital's rule)} \\ &= \lim_{\lambda \to 0} \frac{u^\lambda \log u - (1-u)^\lambda \log(1-u)}{1} \\ &= \log\Big(\frac{u}{1-u}\Big) = R_0(u). \end{aligned}$$

In statistics, R_0 is known as ***logit transformation***.

$R_\lambda(u) = \big(u^\lambda - (1-u)^\lambda\big)/\lambda$ for $\lambda \neq 0$. $\qquad R_0(u) = \log\big(u/(1-u)\big)$.

3°. The Tukey family contains (or almost contains) some common distributions.

- For $\lambda = 1$, we have



$$R_1(u) = \frac{u - (1-u)}{1} = 2u - 1$$

for each u. Thus $X_1 = R_1(U) \sim \text{Uniform}(-1,1)$.

- $\mathfrak{T}(0.14)$ is very close to $N(0,2)$; see Figure 4 below.

- For any λ, the df F_λ of X_λ is just the ordinary mathematical inverse R_λ^{-1} to the continuous strictly increasing function R_λ, so $F_\lambda(x)$ is the u such that $x = R_\lambda(u)$. For $\lambda = 0$ we can get a closed form expression for $F_0(x)$, as follows:

$$x = R_0(u) = \log\Big(\frac{u}{1-u}\Big) \Longrightarrow \frac{u}{1-u} = e^x \Longrightarrow u = \frac{e^x}{1+e^x}\,;$$

thus

$$F_0(x) = \frac{1}{1+e^{-x}} \tag{9}$$

for $-\infty < x < \infty$. Since $F_0(x)$ is continuously differentiable in x, the distribution has a density which we can find by differentiating the df:

$$f_0(x) := \frac{d}{dx} F_0(x) = \frac{e^{-x}}{(1+e^{-x})^2} = F_0(x)\big(1 - F_0(x)\big), \tag{10}$$

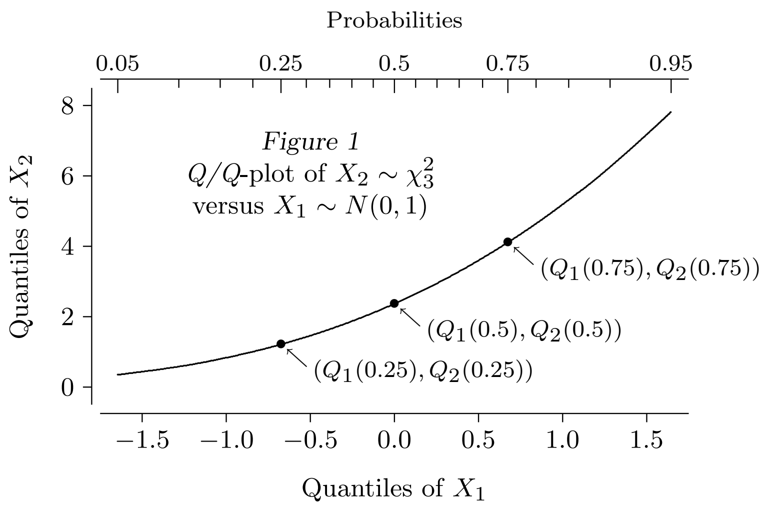
again for $-\infty < x < \infty$. This distribution is called the ***(standard) logistic distribution***.

- $\mathfrak{T}(-1)$ is very close to the distribution of πT_1, where T_1 has a t-distribution with 1 degree of freedom; see Figure 5 below.

4°. As λ goes from 1 to -1 (and the distribution goes (modulo scaling) from uniform, to almost normal, to logistic, to almost t_1), the tails of the $\mathfrak{T}(\lambda)$ distribution go from very light to very heavy. This is another reason why the family is good for robustness studies.

***Q/Q*-plots.** The distributions of two random variables are often compared by making a simultaneous plot of their dfs, or densities, or probability mass functions. This section talks about another way to make the comparison, the so-called Q/Q-plot, in which the quantiles of one of the variables are ploted against the corresponding quantiles of the other variable. To simplify the discussion, suppose that all quantile functions involved are unique, so we can talk about "the" u^{th}-quantile, instead of just "a" u^{th}-quantile.

Let then X_1 be a random variable with df F_1 and quantile function $Q_1 = F_1^*$, and, similarly, let X_2 be a random variable with df F_2 and quantile function $Q_2 = F_2^*$. A plot of the points $\big(Q_1(u), Q_2(u)\big)$ for u in (some subinterval) of $(0,1)$ is called a ***Q/Q-plot*** of X_2 against X_1, or, equivalently, of F_2 against F_1, or of Q_2 against Q_1. As in the illustration below, it is good practice to include a "probability axis" showing the range and spacing of the u's covered by the plot.



Two special cases are worth noting. (A) if X_2 is standard uniform, then the plot displays $\big(Q_1(u), u\big)$ for $u \in (0,1)$; this is essentially just a plot of the df of X_1. (B) if X_1 is standard uniform, then the plot displays $\big(u, Q_2(u)\big)$ for $u \in (0,1)$; this is a graph of the inverse df F_2^*.

Interpreting *Q/Q*-plots. Suppose X_1 and X_2 are two random variables such that

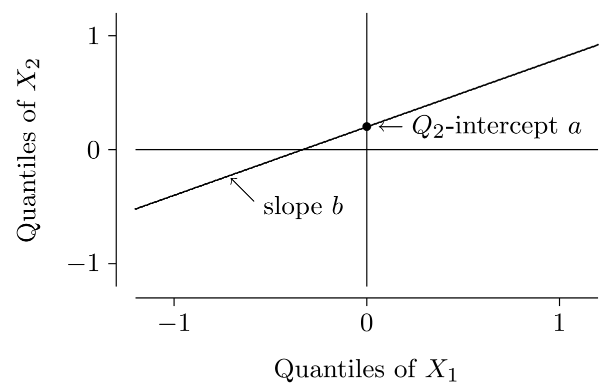
$$X_2 \sim a + bX_1 \tag{11}$$

for some constants a and b with $b > 0$; this says that the distribution of X_2 is obtained from that of X_1 by a change of scale (multiplication by b) and a change in location (addition of a). Let Q_1 be the quantile function for X_1. Since the transformation $x \rightsquigarrow a + bx$ is increasing, Theorem 4 implies that the quantile function for X_2 is given by

$$Q_2(u) = a + bQ_1(u). \tag{12}$$

Consequently a Q/Q-plot of X_2 against X_1 is simply a straight line with Q_2-intercept a and slope b:

Figure 2: Q/Q-plot of $X_2 \sim a + bX_1$ against X_1

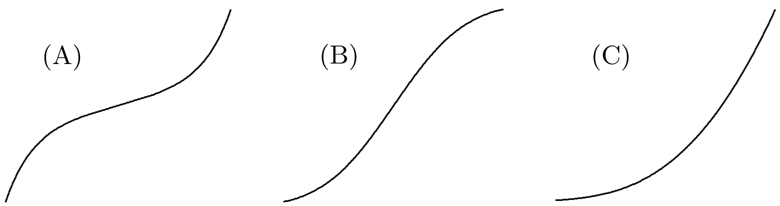


This is one of the primary reasons why Q/Q-plots are of interest.

The Q/Q-plot in Figure 1 is not a straight line. However, at least over the range of probabilities covered in that figure, the plot is not too far from a straight line with intercept a and slope b (for what values of a and b?); that means that the so-called χ^2_3 distribution is approximately the same as that of $a + bZ$ for a standard normal random variable Z.

We have seen that linearity in a Q/Q-plot means that the distributions of the two variables involved are related by a change in location and scale, in the sense that (12) and, equivalently, (11) hold. What does curvature mean? Consider the following three Q/Q-plots; in each plot the quantiles of some random variable X_2 are plotted against the quantiles of some X_1.

Figure 3: Three Q/Q-plots



In case (A), the quantiles of X_2 grow ever more rapidly than do those of X_1 as u tends to 0 and to 1; this means that X_2 has heavier tails than does X_1. In case (B), X_2 has lighter tails than X_1. In case (C), the right tail of X_2 is heavier than that of X_1, but the left tail is lighter. Looking back at Figure 1, you can see that the right tail of the χ^2_3 is somewhat heavier than "normal", and the left tail is somewhat lighter.

The Tukey family, revisited. The next three pages exhibit Q/Q-plots for various members of the Tukey family. In each case you should think about what the plots say about the Tukey distributions. In Figure 4 the solid, nearly-diagonal line is the Q/Q-plot of $\mathfrak{T}(0.14)$ versus standard normal. The dotted line is the least squares linear fit to the solid line; the slope of the least squares line is almost $\sqrt{2}$. In Figure 5 the solid line is the Q/Q-plot of $\mathfrak{T}(-1)$ versus the t distribution with 1 degree of freedom. The dotted line is the straight line with Q_2-intercept 0 and slope π. Figure 6 is a simultaneous Q/Q-plot of the Tukey distributions for $\lambda = -1$ to 1 by 0.2, each against standard normal.

Figure 4

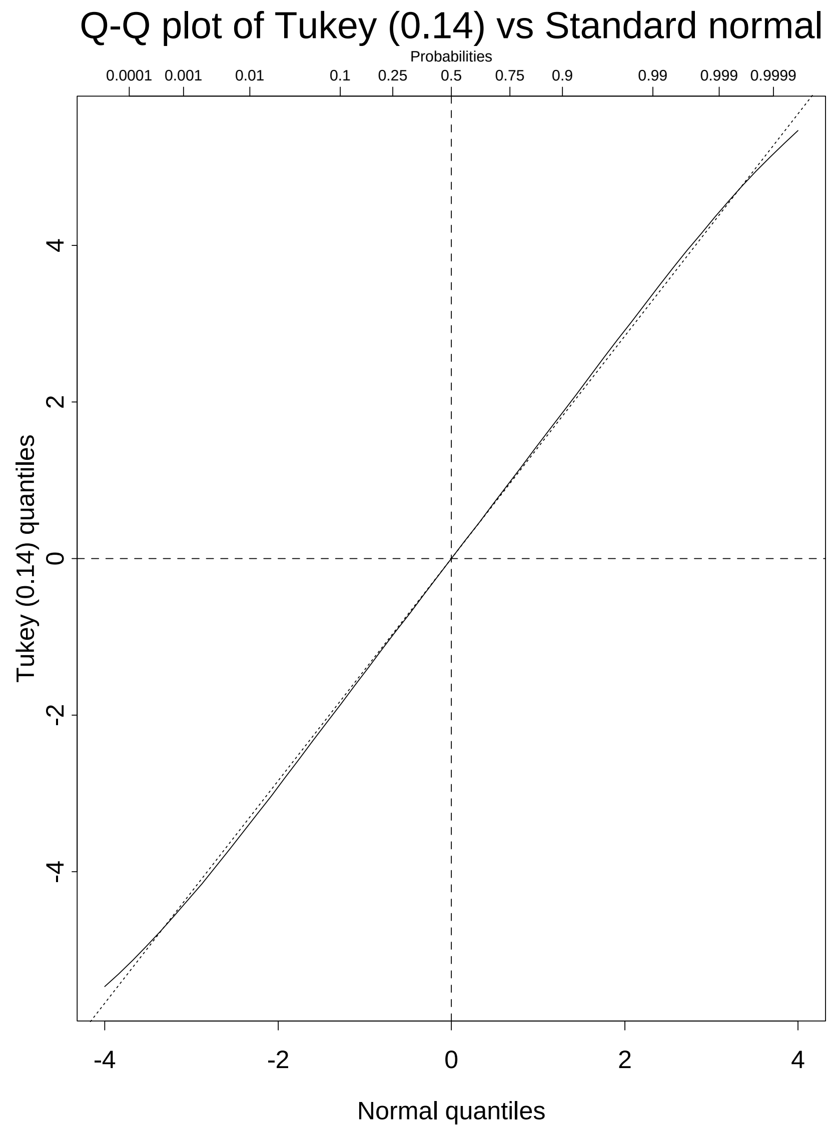


Figure 5

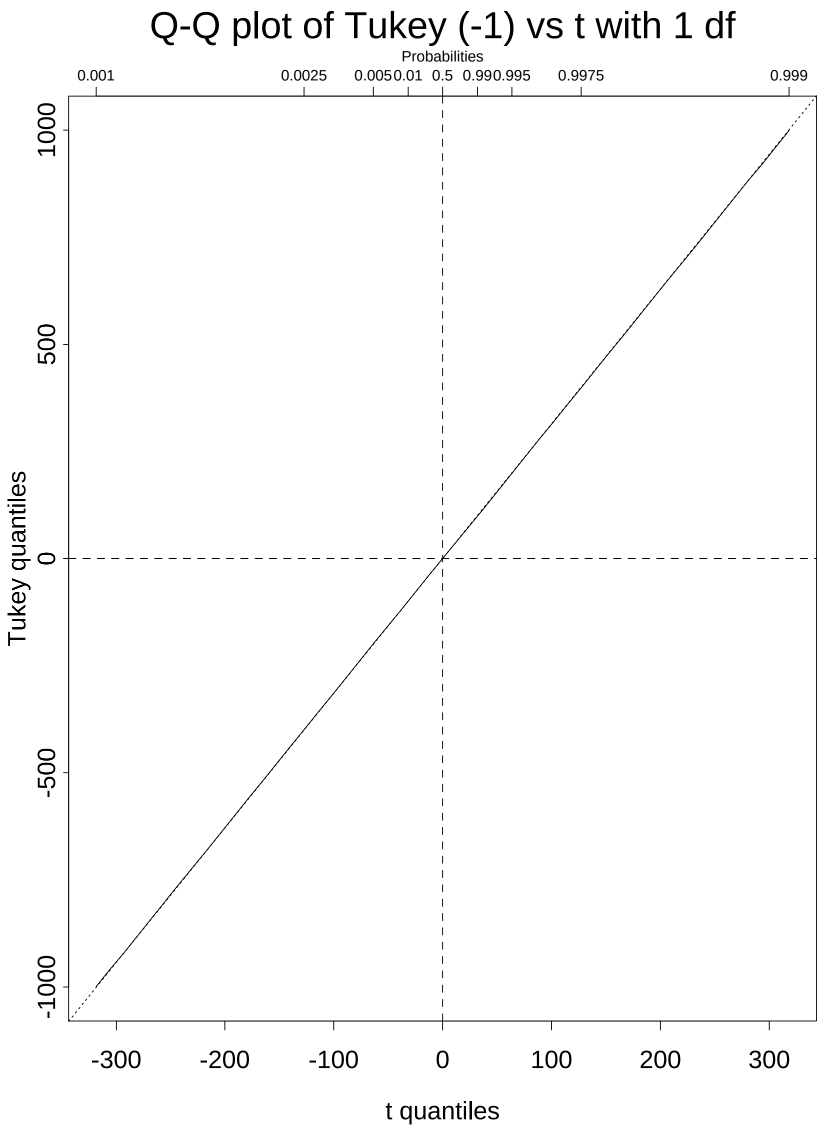
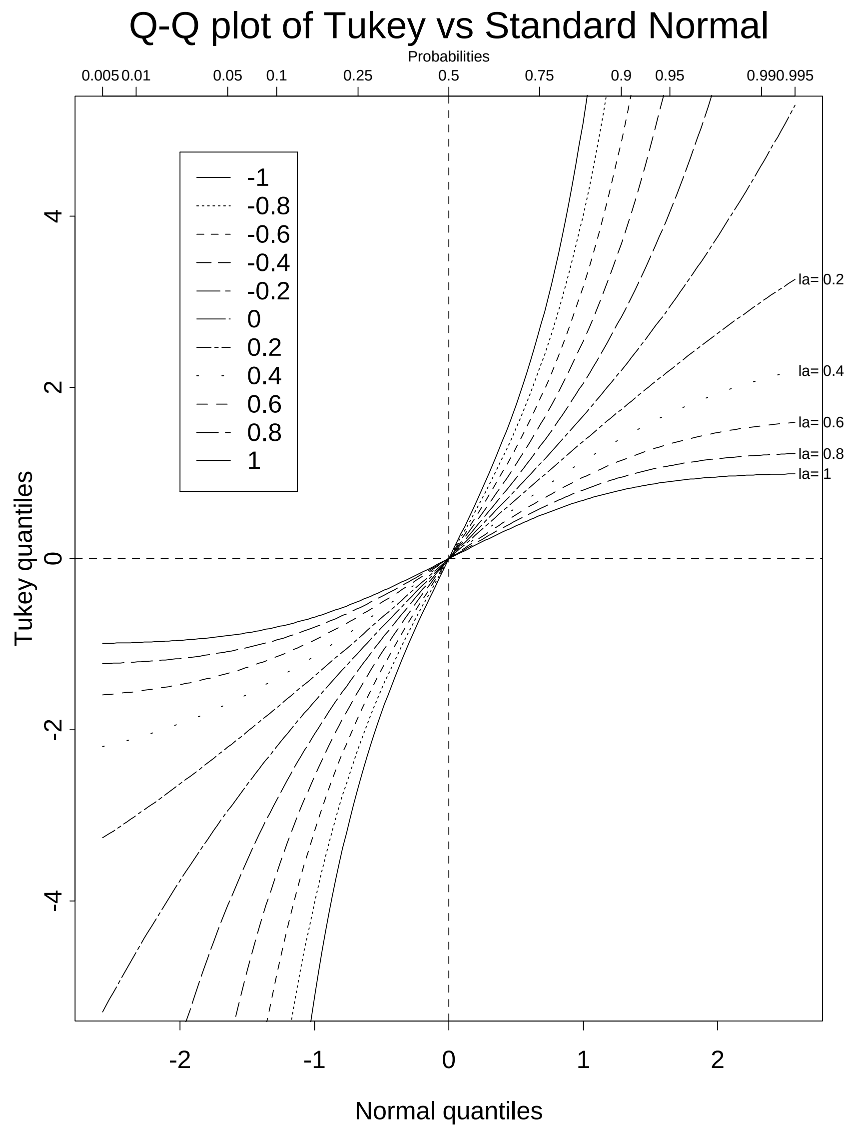


Figure 6



Exercise 1. Prove the claim made after the statement of Theorem 2, namely, that $F^*(u) = F^*(u+)$ for all $u \in (0,1) \iff F$ is strictly increasing over $\{\, x \in \mathbb{R} : 0 < F(x) < 1 \,\}$. [Hint. For $\Longleftarrow$, suppose $x = F^*(u)$ satisfies $F(x) < 1$. Argue that $F^*(u+) \le y$ for each $y > x$.] $\diamond$

For two random variables Y and Z, the notation $Y \sim Z$ means

$$P[\, Y \in B \,] = P[\, Z \in B \,] \text{ for each subset } B \text{ of } \mathbb{R} \tag{13}$$

(technically, B needs to be a so-called Borel measurable set, but we'll ignore that in this course). It is sufficient that (13) hold for sets B of the form $(-\infty, x]$ for $x \in \mathbb{R}$, i.e., that Y and Z have the same df.

Exercise 2. (a) Suppose Y and Z are two random variables such that $Y \sim Z$. Show that $T(Y) \sim T(Z)$ for any transformation T. (b) Suppose Y and Z are defined on the same probability space. Show that if $P[\, Y = Z \,] = 1$, then $Y \sim Z$. $\diamond$

Exercise 3. Show that (3) holds if and only if both (4) and (5) hold. $\diamond$

Exercise 4. The standard exponential distribution has density

$$f(x) = \begin{cases} e^{-x}, & \text{for } x \ge 0, \\ 0, & \text{for } x < 0. \end{cases}$$

What are its quantiles? $\diamond$

Exercise 5. Suppose Z is a standard normal random variable, with density $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ for $-\infty < z < \infty$. (a) Show that $P[Z \ge z] = (1 + o(1))\phi(z)/z$ as $z \to \infty$; here $o(1)$ denotes a quantity that tends to 0 as $z \to \infty$. [Hint: integrate by parts.] (b) Let q_α be the $(1-\alpha)^{\text{th}}$-quantile of Z. Show that

$$q_\alpha = \sqrt{2\log(1/\alpha) - \log(\log(1/\alpha)) - \log(4\pi) + o(1)} \tag{14}$$

as $\alpha \downarrow 0$; here $o(1)$ denotes a quantity that tends to 0 as $\alpha \downarrow 0$. $\diamond$

Exercise 6. The text pointed out several reasons why the Tukey family is good for robustness studies. Give at least one reason why it is not good. $\diamond$

Exercise 7. Show that $\mathfrak{T}(1)$ and $\mathfrak{T}(2)$ are the same distribution, up to a change in scale. Is there a generalization of this fact? $\diamond$

Exercise 8. Below Figure 1, the text states that a plot of $(Q_1(u), u)$ for $u \in (0,1)$ is essentially just a plot of the df of X_1. The term "essentially just" was used because the two plots can in fact be slightly different. Explain how. $\diamond$

Exercise 9. Figure 3 shows the Q/Q-plots for three random variables, X_1, X_2, and X_3, each plotted against $Z \sim N(0,1)$. The plots are drawn to different scales. In scrambled order, the distributions of the X_i's are: (i) standard uniform, (ii) standard exponential, and (iii) t with 1 degree of freedom. Match the plots with the distributions. $\diamond$

Exercise 10. Let X_1 be a standard exponential random variable (see Exercise 4) and let X_2 be a random variable distributed like $-X$. (i) Draw a Q/Q-plot of X_2 against X_1, by plotting $Q_2(u)$ against $Q_1(u)$. (ii) Now plot $Q_2(1-u)$ versus $Q_1(u)$. What is the general lesson to be learned from (i) and (ii)? $\diamond$

Exercise 11. Let Q_1 be the quantile function of the t-distribution with 1 degree of freedom, with density $1/\big(\pi(1+x^2)\big)$ for $-\infty < x < \infty$, and let Q_2 be the quantile function of the Tukey distribution $\mathfrak{T}(-1)$ with parameter $\lambda = -1$. (a) Show that $r(u) := Q_2(u)/Q_1(u)$ is a nondecreasing function of u for $1/2 < u < 1$ (this is hard to do analytically, but easy to demonstrate by a graph) with $r(1/2+) = 8/\pi$ and $r(1-) = \pi$ (this is not difficult). (b) Use the result of part (a) to explain why the Q/Q-plot in Figure 5 looks like a straight line with Q_2-intercept 0 and slope π. $\diamond$