Group E: Dhatri Chennavajula, Endreshak Shylha Sue, Hernandez,Beatriz

1. **a.** Let Z be a continuous random variable and $Y = aZ + b$, where a and b are constants with $a \neq 1$. Let y_p and z_p denote the p-quantile of the distribution of Y and Z, respectively, where $0 < p < 1$. Show that

$$y_p = az_p + b$$

Solution :

$$\begin{aligned}
F_{y_p}(y_p) = (P(Y \leq y_p) &= p \\
P(aZ + b \leq y_p) &= p \\
P(Z \leq \frac{y_p - b}{a}) &= p \\
F_z(\frac{y_p - b}{a}) &= p \\
\frac{y_p - b}{a} &= F^{-1}(p) \\
\frac{y_p - b}{a} &= z_p \\
y_p &= az_p + b
\end{aligned}$$

b. Suppose Z has a standard normal distribution. Let $Y = aZ + b$ where $a > 0$, $a \neq 1$, and $b \neq 0$. Let $p_1 < ... < p_m$ with $0 < p_i < 1$. Let y_{p_i} and z_{p_i} denote the p_i-quantile of the distribution of Y and Z, respectively. Would the points (y_{p_i}, z_{p_i}), $i = 1, ..., m$, fall on a 45° line? Justify your answer (mathematically, rather than intuitively).

Solution :
The Cumulative distribution function determines the distribution of random variables. For the points ($y_{p_i} = Q_Y(p_i)$ and $z_{p_i} = Q_Z(p_i)$) to fall on the 45° line, they must be from same distribution and the quantile functions of Y and Z must be identical. From 1a, we know that the quantiles of Y are linearly related to the quantiles of Z. They would be identical only if a=1 and b=0 which is not the case in this problem

$$\begin{aligned}
F_Z(y) &= P(Z \leq y) \\
F_Y(y) &= P(Y \leq y) \\
&= P(aZ + b \leq y) \\
&= P\left(Z \leq \frac{y - b}{a}\right) \\
\Rightarrow F_Z(y) &\neq F_Y(y)
\end{aligned}$$

Since the CDFs are not equal, the points (y_{p_i}, z_{p_i}) will not fall on the 45° line.

c. The departure of the points from 45° indicates that the mystery distribution is not a standard normal distribution. The mystery distribution is a normal distribution with a some mean not equal to zero, and some variance not equal to 1.

2. Let Y be a random variable with cumulative distribution function

$$F(a;\theta) = \begin{cases} 1 - e^{-\theta a} & \text{for } a > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter. Let Z be a random variable with cumulative distribution $F(a; 1)$.

a. Find the quantile functions of Y and Z, denoted Q_Y and Q_Z, respectively.

Solution :

$\frac{dF(a)}{da} = \theta e^{-\theta a} > 0$, F is strictly increasing

$$\begin{aligned} F_Y(Q_Y) &= p \\ 1 - e^{-\theta Q_Y} &= p \\ e^{\theta Q_Y} &= 1 - p, \text{ where } p \neq 1 \\ log(e^{-\theta Q_Y}) &= log(1-p) \\ Q_Y &= -\frac{1}{\theta} log(1-p), \text{ for } 0 \leq p < 1 \end{aligned}$$

Solve Q_Z, by plugging for $\theta =$ in 1

$$Q_Z = -log(1-p), \text{ for } 0 \leq p < 1$$

b. Let $p_1 < ... < p_m$ with $0 < p_i < 1$. Let y_{p_i} and z_{p_i} denote the p_i-quantile of the distribution of Y and Z, respectively. Would the points (y_{p_i}, z_{p_i}), $i = 1, ..., m$, fall on a straight line? If so, give the intercept and slope of the line. If not, explain why.

Solution :
The random variables Y and Z both have exponential distributions but with different parameters (with different mean and variances).

The relationship between Q_Y and Q_Z:

$$Q_Y = \frac{1}{\theta} Q_Z$$

Since there is a linear relationship between the quantiles of Y and the quantiles of Z, the points (y_{p_i}, z_{p_i}) will fall on a straight line, but not on the 45° line. That will only happen if the parameters for Y and Z are identical. From the above equation the slope of the line is $\frac{1}{\theta}$ and the Y-intercept will be zero.

Intercept of the line is the height of the line at Z=0. To find the Y intercepts, we can plot the two medians against each other. Since the points at the median fall on the straight line we can calculate the intercept by plugging the median values in the linear equation $Y = aZ + b$.

$Median = p = 0.5$ and we want to have $q_{0.5}$

$F(q_{0.5}) \geq 0.5$

$\frac{dF(a)}{da} = \theta e^{-\theta a} > 0$, F is strictly increasing

$Calculating\ Median:$

$$\begin{aligned} Y_{0.5} = median(Z) = F(q_{0.5}) &= 0.5 \\ 1 - e^{-\theta a} &= 0.5 \\ a &= \frac{1}{\theta} log(0.5) \\ Y_{0.5} &= \frac{1}{\theta} 0.693 \\ Z_{0.5} = median(Z) = F(q_{0.5}) &= 0.5 \\ 1 - e^{-a} &= 0.5 \\ a &= \log(0.5) \\ Z_{0.5} &= 0.693 \end{aligned}$$

$slope = a = \frac{1}{\theta}, Y_{0.5} = \frac{1}{\theta} 0.693, Z_{0.5} = 0.693$

$$\begin{aligned} Y_{0.5} &= aZ_0.5 + b \\ \frac{1}{\theta} 0.693 &= \frac{1}{\theta} 0.693 + b \\ intercept &= b = 0 \end{aligned}$$

The points (ypi, zpi) will fall on a straight line but NOT on the 45° line. That will only happen if the parameters for Y and Z are identical.

3. Let Y be a random variable with cumulative distribution function

$$F(a; \lambda, \gamma) = \begin{cases} 1 - e^{-\lambda a^{\gamma}} & \text{for } a > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$ and $\gamma > 0$ are parameters. Let Z be a random variable with cumulative distribution function $F(a; 1, 1)$.

a. Find the quantile functions of Y and Z, denoted Q_Y and Q_Z, respectively.

$Solution:$ $\frac{dF(a)}{da} = \lambda a^{\gamma} e^{-\lambda a^{\gamma}} > 0$, F is strictly increasing

$$\begin{aligned} F_Y(Q_Y) &= p \\ 1 - e^{-\lambda Q_Y^{\gamma}} &= p \\ e^{\lambda Q_Y^{\gamma}} &= 1 - p, \text{ where } p \neq 1 \\ log(e^{-\lambda Q_Y^{\gamma}}) &= \log(1-p) \\ Q_Y^{\gamma} &= -\frac{1}{\lambda}\log(1-p) \\ Q_Y &= \Big(-\frac{1}{\lambda} log(1-p) \Big)^{\frac{1}{\gamma}}, \text{ for } 0 \leq p < 1 \end{aligned}$$

To solve Q_Z, plug in for $\lambda = \gamma = 1$:

$$Q_Z = -log(1-p), \text{ for } 0 \leq p < 1$$

b. Let $p_1 < ... < p_m$ with $0 < p_i < 1$. Let y_{p_i} and z_{p_i} denote the p_i-quantile of the distribution of Y and Z, respectively. Would the points (y_{p_i}, z_{p_i}), $i = 1, ..., m$, fall on a straight line? If so, give the intercept and slope of the line. If not, explain why.

$Solution:$

The relationship between Q_Y and Q_Z:

$$Q_Y = \Big(\frac{1}{\lambda} Q_Z \Big)^{\frac{1}{\gamma}}$$

From the above equation it is clear that the Y and Z are not linearly related, the Z quantiles are raised to an exponent. The points (y_{p_i}, z_{p_i}) will not fall on a straight line.

4. **[Extra Credit]** Discuss the implications of the results in problems 1-3 with regard to the use of quantile-quantile plots for comparing two distributions.

Solution :

Problems 1-3 are meant to show the pitfalls of using q-q plots without a deeper understanding of the graph. People often create q-q plots and focus on ensuring that the data fall on the 45-degree line to conclude that the sample data are normally distributed.

Problem 1 shows that data may still be normally distributed even when the quantiles from such data may not fall on the 45-degree line (plotted against the quantiles of the standard normal). When that is the case, we expect to see a line on the q-q plot that will not coincide with the 45-degree line. This is because there is a linear relationship between any non-standard normal distribution and the normal distribution with mean=0, variance=1.

Problem 2 highlights the fact that there may be a linear relationship shown on the q-q plot, but that does not mean that the data are normally distributed. In problem 2, although there is a linear relationship between the quantiles of Y and Z, these random variables have exponential distributions, with Z as the standard version with theta=1. A q-q plot with the quantiles of these 2 variables falling on the 45-degree line would indicate that both Y and Z are exponential (not normal) with theta=1.

Problem 3 highlights the fact that two random variables may have the same type of distribution, but there may not be a linear relationship between the standard and non-standard distributions of the same type. The quantiles of Y and Z may have a non-linear relationship as illustrated in problem 3.