Fatima Khoogar, Ying Gong, Ryan Quigley
MATH 267A
Section 2
Group C

Problem Set 8

1.

a) Since Z and Y are continuous random variables, y_p and z_p are the p-quantiles that satisfy $F_Y(y_p) = F_Z(z_p) = p$. Now we use the definition of the CDF to show the desired result.

$$\begin{aligned}
F_Y(y_p) &= p \\
P(Y \leq y_p) &= p \\
P(aZ + b \leq y_p) &= p \\
P(Z \leq \frac{y_p - b}{a}) &= p && \text{assuming } a > 0 \\
F_Z\left(\frac{y_p - b}{a}\right) &= p
\end{aligned}$$

Thus, $z_p = \frac{y_p - b}{a} \Leftrightarrow y_p = az_p + b$. Note, in the equations above we must assume $a > 0$. If $a = 0$, the right hand side of the inequality will be undefined. If $a < 0$, the inequality will flip and we will end up with $y_p = az_{1-p} + b$.

b) From part (a), $Y = aZ + b$ where $a > 0$ implies $y_p = az_p + b$. This holds for any $p_i \in (0, 1)$, so we have $y_{p_i} = az_{p_i} + b$. For any two points (y_{p_i}, z_{p_i}) and (y_{p_j}, z_{p_j}) where $i \neq j$, the slope between them is,

$$\frac{z_{p_j} - z_{p_i}}{y_{p_j} - y_{p_i}} = \frac{z_{p_j} - z_{p_i}}{(az_{p_j} + b) - (az_{p_i} + b)} = \frac{z_{p_j} - z_{p_i}}{a(z_{p_j} - z_{p_i})} = \frac{1}{a}$$

Since it is given that $a \neq 1$, the points $(y_{p_i}, z_{p_i}), i = 1, ..., m$, do not fall on a 45° line.

c) From part (b), $Z \sim \mathrm{N}(0, 1) \Rightarrow Y \sim \mathrm{N}(b, a^2)$. Therefore, we have previously shown that plotting the corresponding p_i-quantiles of a standard normal distribution against a normal distribution with mean b and variance a^2 gives a straight line with slope that is a function of a. Thus, the departure from the 45° line in the graph does not indicate that the mystery distribution is not a normal distribution since the points still fall on a straight line; however, it does indicate that the mystery distribution is not standard normal.

2.

a) F is a continuous function for $a > 0$. We take the derivative with respect to a to ensure that F is strictly increasing.

$$\frac{\partial}{\partial a} F(a; \theta) = \theta \mathrm{e}^{-\theta a}$$

The exponential function is always greater than 0 and it is given that θ is greater than 0, so the derivative is positive for $a > 0$. Thus, F is strictly increasing. To find $Q_Y(p)$, let $p = F(q_p)$.

$$
\begin{aligned}
p &= 1 - \mathrm{e}^{-\theta q_p} \\
q_p &= -\frac{1}{\theta}\log(1-p) && \text{for } 0 \le p < 1 \\
Q_Y(p) &= -\frac{1}{\theta}\log(1-p) && \text{for } 0 \le p < 1
\end{aligned}
$$

Since Z has the same form of cumulative distribution function with $\theta = 1$,

$$
\begin{aligned}
Q_Z(p) &= -\log(1-p) && \text{for } 0 \le p < 1 \\
\Rightarrow Q_Z(p) &= \theta Q_Y(p)
\end{aligned}
$$

b) From part (a), we have $z_p = \theta y_p$ for $0 \le p < 1$. There is a linear relationship between z_p and y_p, so all points $(y_{p_i}, z_{p_i}), i = 1, ..., m$, where $p_i \in [0, 1)$ will fall on the straight line with slope $= \theta$ and intercept $= 0$.

More formally, the intercept is the point $(y_{p_i}, 0)$ so we want to solve $0 = \theta y_{p_i}$ for y_{p_i}. It is given that $\theta > 0$, so it must be the case that $y_{p_i} = 0$. For the slope, consider any two points (y_{p_i}, z_{p_i}) and (y_{p_j}, z_{p_j}) where $i \neq j$. The slope between them is,

$$
\frac{z_{p_j} - z_{p_i}}{y_{p_j} - y_{p_i}} = \frac{z_{p_j} - z_{p_i}}{\frac{1}{\theta}z_{p_j} - \frac{1}{\theta}z_{p_i}} = \frac{z_{p_j} - z_{p_i}}{\frac{1}{\theta}(z_{p_j} - z_{p_i})} = \theta
$$

See Figure 1 below for QQ-plots of the quantiles of $F(a; 2)$ and $F(a; 0.5)$ against the quantiles of $F(a; 1)$

3.

a) Again we check the derivative of F with respect to a to make sure it is strictly increasing.

$$
\frac{\partial}{\partial a}F(a; \lambda, \gamma) = \lambda\gamma a^{\gamma-1}\mathrm{e}^{-\lambda a^\gamma}
$$

The exponential function is always greater than 0 and it is given that a, λ, and γ are all greater than 0, so the derivative is positive for $a > 0$. Thus, F is strictly increasing. To find $Q_Y(p)$, let $p = F(q_p)$.

$$
\begin{aligned}
p &= 1 - \mathrm{e}^{-\lambda a^\gamma} \\
q_p &= \left[-\frac{1}{\lambda}\log(1-p)\right]^{\frac{1}{\gamma}} && \text{for } 0 \le p < 1 \\
Q_Y(p) &= \left[-\frac{1}{\lambda}\log(1-p)\right]^{\frac{1}{\gamma}} && \text{for } 0 \le p < 1
\end{aligned}
$$

Since Z has the same form of cumulative distribution function with $\lambda = \gamma = 1$,

$$
\begin{aligned}
Q_Z(p) &= -\log(1-p) && \text{for } 0 \le p < 1 \\
\Rightarrow Q_Z(p) &= \lambda[Q_Y(p)]^\gamma
\end{aligned}
$$

b) From part (a), we have $z_p = \lambda y_p^\gamma$ for $0 \leq p < 1$. Unless $\gamma = 1$, there is a non-linear relationship between the z_p and y_p. Thus, the points $(y_{p_i}, z_{p_i}), i = 1, ..., m$ will not fall on a straight line. See Figure 2 below for QQ-plots of the quantiles of $F(a; 2, 1)$ and $F(a; 0.25, 2)$ against the quantiles of $F(a; 1, 1)$

4. The simple test of "does it fall on the 45° line?" can only be used to determine whether or not two particular distributions are identical. This simple test should not be used to rule out entire families of distributions. In problems (1) and (2), we compared the generalized normal and exponential distributions to their respective standard parameterizations. In both cases, the relationship between the corresponding quantiles was linear, so deviations from the 45° line were still a straight line. Thus when comparing the quantiles of an unknown distribution to the quantiles of the standard normal or the standard exponential, a straight line in the QQ-plot indicates that the unknown distribution belongs to the family of distributions containing that particular standard parameterization. Additionally, the intercept and slope provide an estimate of μ and σ for the normal distribution, and the slope provides an estimate of θ for the exponential distribution. This straight line property is not the case for all distributions, which is demonstrated in problem 3. There we are comparing a generalized Weibull distribution to the standard Weibull distribution, and the conclusion is that the corresponding quantiles are not linearly related. Thus when comparing the quantiles of an unknown distribution and the standard Weibull distribution, we would look for a curved line to indicate that an unknown distribution belongs to the family of Weibull distributions. In general when we observe deviations from the 45° line, the conclusions that can be drawn about the unknown distribution depend on the distribution it is being compared to.

Figure 1: Problem 2

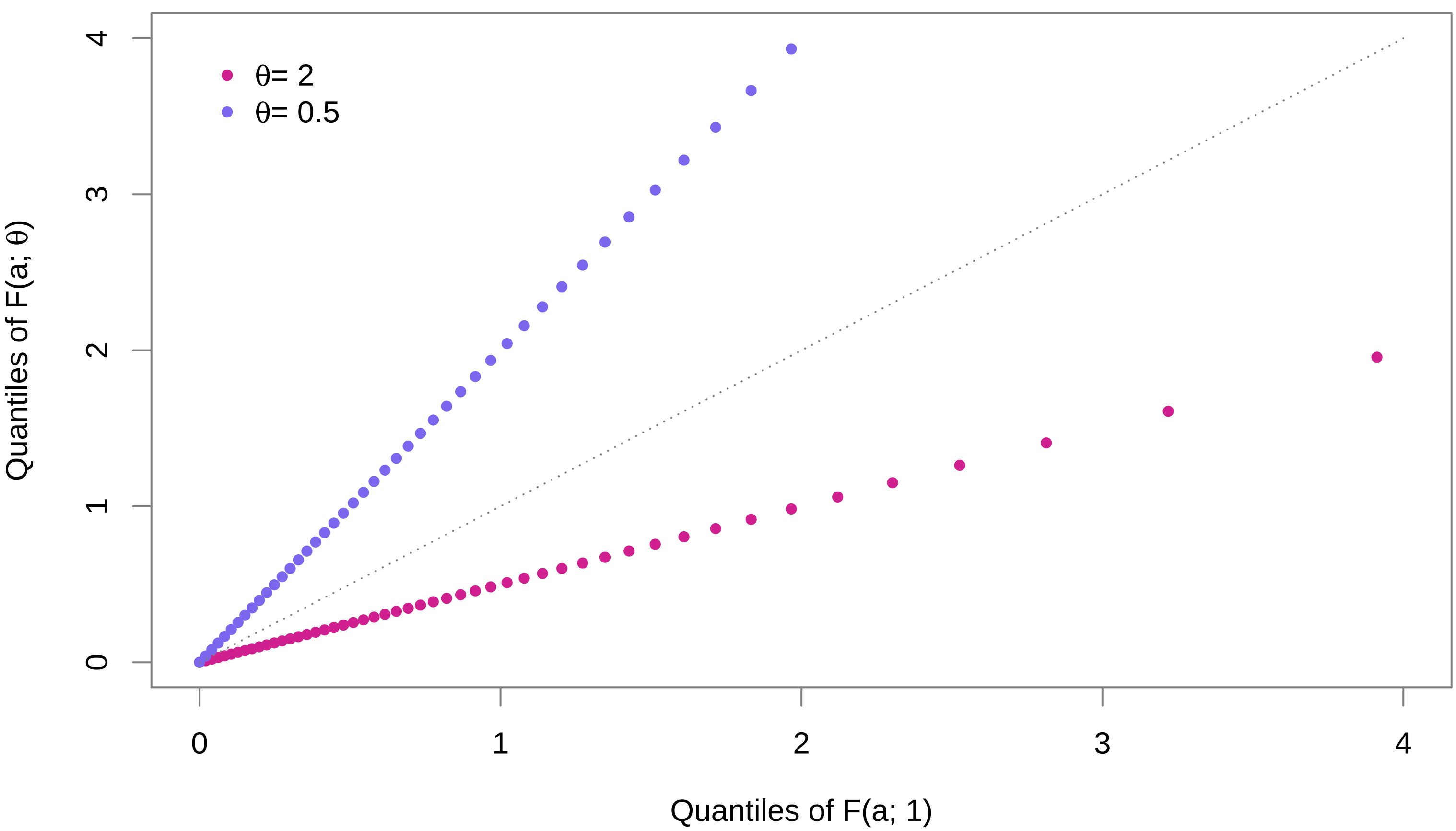

Figure 2: Problem 3