

# SENG 585 Homework #2 - Solution

## From the text problem # 3.5

The supplier of an electromechanical control for a household appliance ran an accelerated life test on sample controls. In the test, 25 controls were put on test and run until 30 thousand cycles had been accumulated. Failures occurred at 5, 21, and 28 thousand cycles. The other 22 controls did not fail by the end of the test.

(a) Compute and plot a nonparametric estimate for  $F(t)$ .

A nonparametric estimate for  $F(t)$  can be easily produced using a ‘life data’ object from the SMRD package. Life data objects are created with the function `frame.to.ld` which requires the data to be formatted into a data frame.

If the data set is pre-loaded as part of the SMRD package (as is the case for the data set in this exercise), it has already been formatted into a data frame. In this case the life data object can be created by simply entering the name of the data set into `frame.to.ld`, for example

```
prob3_5.ld <- frame.to.ld(frame = prob3_5,
                          response.column = 1,
                          censor.column = 2,
                          time.units = "Kilocycles")
```

NOTE: The complete list of pre-loaded data sets in the SMRD package, can be found by running `data(package = 'SMRD')` in the R console.

A plot of the nonparametric estimate for  $F(t)$  is produced by plotting the life.data object, as shown in Figure 1. This exercise does not require us to plot the confidence bands along with the estimate, so we set the `band.type` argument to “none”. The `ylim` argument is added because of the way R dynamically sets axis values for a plot. Try running following code without the `ylim` argument. This problem could easily be fixed such that the `ylim` argument would not need to be set... but the fix would cause bigger issues for other exercises.

```
plot(prob3_5.ld, band.type = "none", ylim = c(0,0.1))
```

To compute the values of  $F(\hat{t}_i)$  and present them in a  $\text{\LaTeX}$ table the following code is used.

**NOTE: to render the  $\text{\LaTeX}$ table you must add the chunk option `results='asis'`.**

```
xtab<-xtable(print(cdfest(prob3_5.ld))$table,
                  digits = c(0,0,0,4,4,4,4),
                  align = rep("c",7),
                  caption = "Value of the nonparametric estimate for Problem 3.5" )
print(xtab, include.rownames = FALSE, comment = FALSE)
```

Kilocycles-lower	Kilocycles-upper	Fhat	SE_Fhat	95% Lower	95% Upper
0	5	0.0000	0.0000	0.0000	0.0000
5	21	0.0400	0.0392	0.0056	0.2355
21	28	0.0800	0.0543	0.0201	0.2694
28	30	0.1200	0.0650	0.0392	0.3130

Table 1: Value of the nonparametric estimate for Problem 3.5

(b) Compute an approximate 95% confidence interval for the probability that an electromechanical device from the same production process, tested in the same way, would fail before 30 thousand cycles. Use the conservative binomial approach.

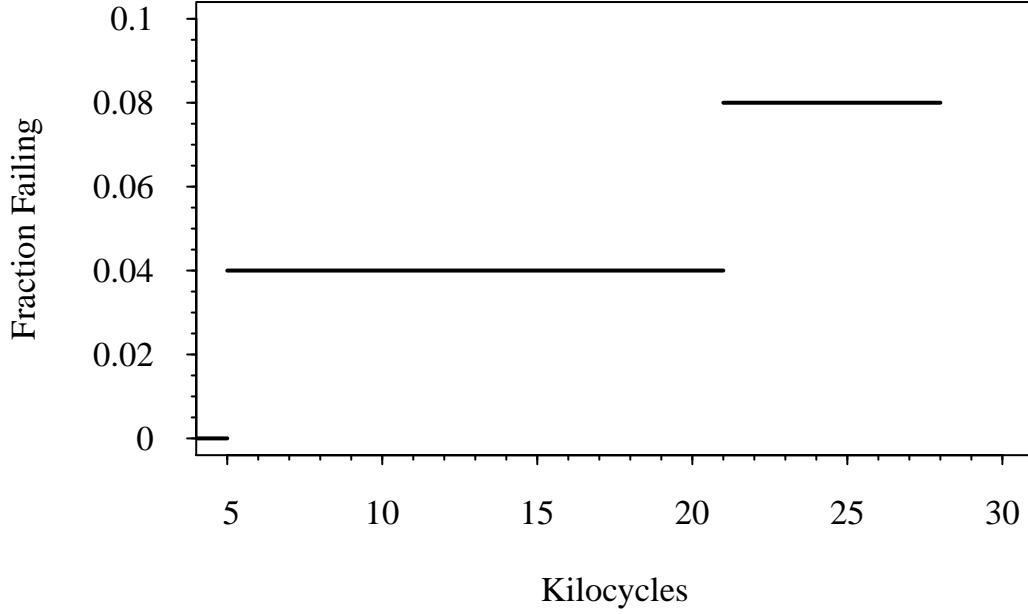


Figure 1: Plot of the nonparametric estimate of the cdf for Problem 3.5

In Example 3.1 we see that the nonparametric estimate for  $\hat{F}(30 \text{ kilocycles})$  with singly censored data is

$$\frac{\# \text{ failed up to 30 kilocycles}}{\# \text{ at risk}} = \frac{3}{25}.$$

The Clopper-Pearson ‘exact’ confidence interval based on the cumulative probabilities of the binomial distribution (i.e., the exact distribution rather than an approximation) gives a lower 95% confidence bound of

$$\left(1 + \frac{(25 - 3 + 1) \times \mathcal{F}(.975, 50 - 6 + 2, 6)}{3}\right)^{-1} = 0.02546$$

and an upper 95% confidence bound of

$$\left(1 + \frac{25 - 3}{(3 + 1) \times \mathcal{F}(.925, 6 + 2, 50 - 6)}\right)^{-1} = 0.31219$$

where  $\mathcal{F}(p, df_1, df_2)$  denotes the  $p^{th}$  quantile of an  $F$  distribution with  $df_1$  and  $df_2$  degrees of freedom.

- 
- (c) *Compute an approximate 95% confidence interval for the probability that an electromechanical device from the same production process, tested in the same way, would fail before 30 thousand cycles. Use the normal-approximation method based on  $Z_{\hat{F}(30)} \sim NOR(0, 1)$*

Using the  $Z_{\hat{F}(30)} \sim NOR(0, 1)$  approximate method results in a lower 95% confidence bound of

$$\frac{3}{25} - 1.96 \times \sqrt{\frac{3}{25} \times \left(1 - \frac{3}{25}\right) \times \frac{1}{25}} = -0.007$$

and an upper 95% confidence bound of

$$\frac{3}{25} + 1.96 \times \sqrt{\frac{3}{25} \times \left(1 - \frac{3}{25}\right) \times \frac{1}{25}} = 0.247$$


---

- (d) *Explain why, in this situation, the approach in part (b) would be preferred to the approach in part (c)*

The CI in part (b) is conservative as it based on the exact distribution of the value of  $F(t)$  (i.e. the binomial) while the method used in part (c) is an approximation of the binomial distribution. Further, the approximation in part (c) has a negative lower bound which is not possible.

---

- (e) *The appliance manufacturer is really interested in the probability of the number of days to failure for its product. Use-rate differs from household to household, but the average rate is 2.3 cycles per day. What can the manufacturer say about the proportion of devices that would fail in 10 years of operation (the expected technological life of the product)?*

We can convert usage in days to usage in cycles by making the following conversion

$$10 \text{ years} \times \left(365 \frac{\text{days}}{\text{year}}\right) \times \left(2.3 \frac{\text{cycles}}{\text{day}}\right) \approx 8.4 \times 10^3 \text{ cycles.}$$

Therefore, we want to compute a 95% CI for  $F(8.4)$  where usage is measured in kilocycles. From the plot in part a) the manufacturer can say that with 95% confidence the proportion of electromechanical control failures after 10 years of use at a use rate of 2.3 cycles per day is in the interval (0,0.24)

Alternatively, if the manufacturer uses the binomial (Clopper-Pearson) approximation, the reported 95% confidence interval would be (0.01, 0.204).

If the manufacturer uses the normal approximation, the reported 95% confidence interval would be (0.01, 0.204).

Finally, if the manufacturer used R to compute the 95% confidence interval by plotting the Kaplan-Meier estimate from a “Surv” object, the reported interval would be (0, 0.114).

If the electromechanical controls experienced only one of two usage rates  $r_1$  or  $r_2$ , where  $r_1 > r_2$  such that the lifetime of any control device follows either  $F_1(t|\theta_1)$  or  $F_2(t|\theta_2)$ . The lifetime of the overall population would then depend on the proportion of units experiencing each usage rate denoted  $\zeta_i, i = 1, 2$ , where  $\zeta_2 = 1 - \zeta_1$ . The probability of failure for the population would then be expressed as

$$F_T(t) = \sum_{i=1}^2 \zeta_i F_i(t|\theta_i)$$


---

- (f) *Refer to part (e). Describe an appropriate model to use when use-rate varies in the population of units. To simplify, start by assuming that there are only two different use-rates. Discuss, using appropriate expressions*

## From the text problem # 3.7

Consider the Plant 1 heat exchanger data in Figure 3.1

- (a) Write the likelihood for these data in terms of  $\pi_1, \pi_2, \pi_3$ , and  $\pi_4$

In terms of  $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$ , the likelihood can be written as

$$L(\boldsymbol{\pi}) = C[\pi_1]^1[\pi_2]^2[\pi_3]^2[\pi_4]^{95}$$

>where  $C$  is a constant that does not depend on  $\pi$ .

- 
- (b) Write the likelihood for these data in terms of  $p_1, p_2, p_3$ , and  $p_4$

In terms of  $\mathbf{p} = (p_1, p_2, p_3, p_4)$ , the likelihood can be written as

$$L(\mathbf{p}) = C[p_1]^1[p_2(1 - p_1)]^2[p_3(1 - p_1)(1 - p_2)]^2[(1 - p_1)(1 - p_2)(1 - p_3)]^{95}$$

because  $p_4 = 1$  in this case.

---

## From the text problem # 3.18

Explain why the nonparametric estimate of  $F(t)$  is a set of points for the heat exchanger data in Example 3.6 but a step-function for the shock absorber data in Example 3.9

The difference in these examples is what is known when a failure occurs. In the heat exchanger problem, failures are only known to occur during the inspections at the end of the year. If a heat exchanger pipe failed prior to the inspection, the exact failure time would be unknown. It would be incorrect to plot the CDF as a step function between the inspections, since zero, one or many failures may have occurred during the year, we just don't know when.

The battery data set lists exact failure times, so if no failures were observed between times  $t_1$  and  $t_2$ , we know that no units failed between these two times and it is reasonable to assume that the value of nonparametric estimate of the cdf is constant between these times.

## From the text problem # 4.7

Consider the Weibull  $h(t)$ . Note that when  $\beta = 1$ ,  $h(t)$  is constant and that when  $\beta = 2$ ,  $h(t)$  increases linearly. Show that:

- (a) If  $0 < \beta < 1$ , then  $h(t)$  is decreasing in  $t$

First, note that for the Weibull distribution

$$\begin{aligned}h(t) &= \frac{\beta}{\eta^\beta} t^{\beta-1}, & t > 0, \beta > 0, \eta > 0 \\h'(t) &= \frac{\beta(\beta-1)}{\eta^\beta} t^{\beta-2}, & t > 0, \beta > 0, \eta > 0 \\h''(t) &= \frac{\beta(\beta-1)(\beta-2)}{\eta^\beta} t^{\beta-3}, & t > 0, \beta > 0, \eta > 0\end{aligned}$$

Thus, for  $0 < \beta < 1$ ,  $h'(t) < 0$ , and  $h(t)$  is decreasing for  $\beta \in (0, 1)$

- (b) If  $1 < \beta < 2$ , then  $h(t)$  is concave increasing.

For  $\beta \in (1, 2)$ ,  $h'(t) > 0$  and  $h''(t) < 0$ , thus  $h(t)$  is concave increasing.

- (c) If  $\beta > 2$ , then  $h(t)$  is convex increasing

For  $\beta \in (2, \infty)$ ,  $h'(t) > 0$  and  $h''(t) > 0$ , thus  $h(t)$  is convex increasing

## From the text problem # 4.15

The coefficient of variation,  $\gamma_2$ , is a useful scale-free measure of relative variability for a random variable.

(a) Derive an expression for the coefficient of variation for the Weibull distribution

Using the results from Exercise 4.5, we can express the Weibull coefficient of variation as

$$\gamma_2 = \sqrt{\frac{\sigma^2}{\mu^2}} = \left[ \Gamma\left(1 + \frac{2}{\beta}\right) \Gamma^{-2}\left(1 + \frac{1}{\beta}\right) - 1 \right]^{0.5}$$

(b) Compute  $\gamma_2$  for all combinations of  $\beta = 0.5, 1, 3, 5$  and  $\eta = 50, 100$ . Also, draw (or use the computer to draw) a graph of the Weibull pdfs for the same combinations of parameters.

As seen in (a), the value of  $\gamma_2$  for the Weibull distribution depends only on the value of the shape parameter  $\beta$ . For a given value of  $\beta$  the value of  $\gamma_2$  can be found using the following R code

```
beta <- c(0.5,1,3,5)

data.frame(CoV = (gamma(1+2/beta)*(gamma(1+1/beta))^-2-1)^0.5,
           row.names = c("eta(0.5) =",
                        "eta(1.0) =",
                        "eta(3.0) =",
                        "eta(5.0) =") )

##                CoV
## eta(0.5) = 2.2360680
## eta(1.0) = 1.0000000
## eta(3.0) = 0.3634465
## eta(5.0) = 0.2290529
```

The plots of the respective pdfs may then be found using

```
beta <- rep(c(0.5,1,3,5),2)
eta <- rep(c(50,100),each = 4)

par(family="serif")

for(i in 1:8) {
  curve(dweibull(x, shape = beta[i], scale = eta[i]),
        add = ifelse(i==1,F,T),
        col = i,
        lty = i,
        lwd = 2.5,
        xlim = c(0,100), ylim = c(0,0.04),
        xlab = 't', ylab = 'f(t)')
}
legend("topright",
      legend = c(expression(beta*','*eta=="0.5,50"),
                  expression(beta*','*eta=="1.0,50"),
                  expression(beta*','*eta=="3.0,50"),
                  expression(beta*','*eta=="5.0,50"),
                  expression(beta*','*eta=="0.5,100"),
                  expression(beta*','*eta=="1.0,100"),
                  expression(beta*','*eta=="3.0,100"),
                  expression(beta*','*eta=="5.0,100")),
      lty = 1:8,
```

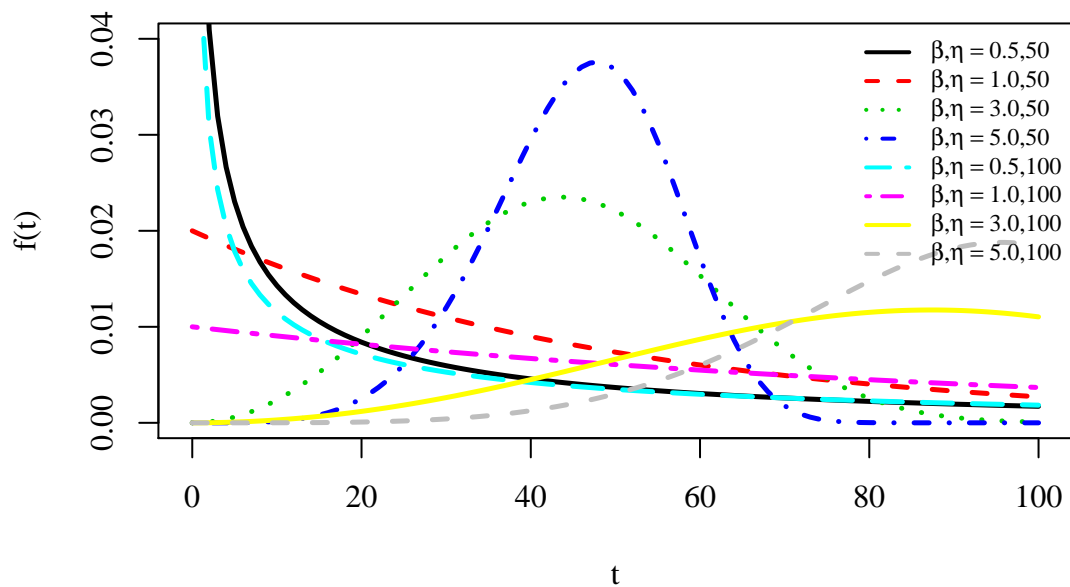


Figure 2: Plots of various Weibull densities

```
lwd = 2,
col = 1:8,
cex = 0.75,
bty = "n",
xjust = 0.5)
```

- (c) Explain the effect that changes in  $\eta$  and  $\beta$  have on the shape of the Weibull density and the effect that they have on  $\gamma_2$ .

In problem 4.7 we discussed the effect changing the shape parameter  $\beta$  has on the shape of the Weibull distribution, thus we need not discuss it again here. Increasing the scale parameter  $\eta$  or “characteristic life” increases the skewness of the distribution by shifting the 0.632 quantile to the right.

The coefficient of variation  $\gamma_2$  is a function of the shape parameter alone. The figure below illustrates the effect of changing  $\beta$

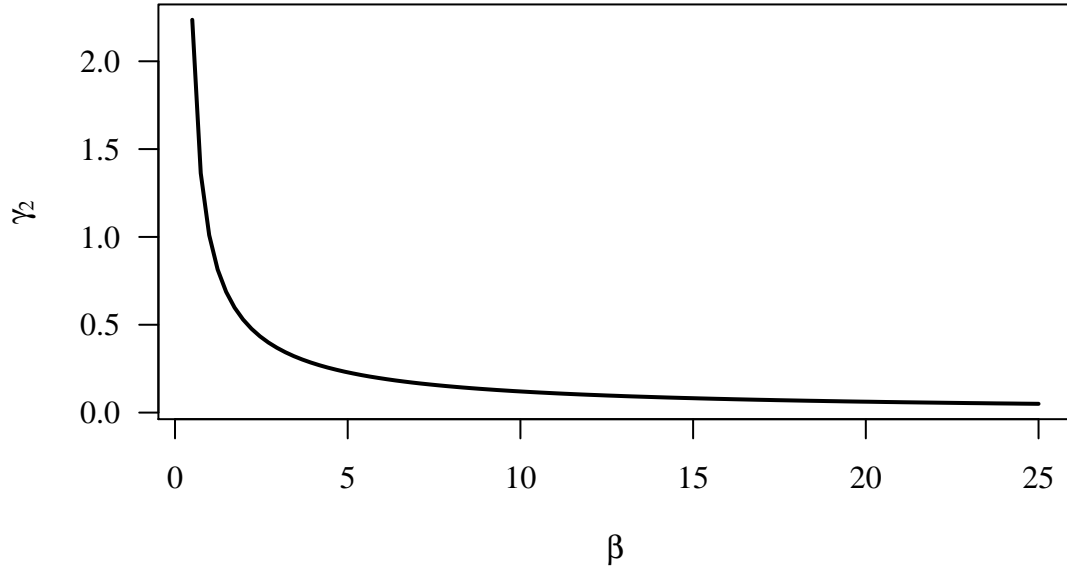


Figure 3: Weibull coefficient of variation vs shape parameter

## From the text problem # 5.2

*In some applications a sample of failure times comes from a mixture of subpopulations.*

- (a) *Write down the expression for the cdf  $F(t)$  for a mixture of two exponential distributions with means  $\theta_1 = 1$  and  $\theta_2 = 10$  (subpopulations 1 and 2, respectively) with  $\zeta$  being the proportion from subpopulation 1.*

In reliability, mixture distributions can arise when multiple failure mechanisms generate failures for which the respective distributions are sufficiently distinct. For the mixture model discussed in this exercise, the pdf is expressed as

$$f_x(x) = \eta \frac{1}{\theta_1} e^{-\left(\frac{x}{\theta_1}\right)} + (1 - \eta) \frac{1}{\theta_2} e^{-\left(\frac{x}{\theta_2}\right)}$$

Integrating wrt  $x$  gives

$$F_x(x) = \int_0^x f_x(t) dt = \eta(1 - e^{-\left(\frac{x}{\theta_1}\right)}) + (1 - \eta)(1 - e^{-\left(\frac{x}{\theta_2}\right)})$$

and finally,

$$F_x(x) = 1 - \eta e^{-x} - (1 - \eta) e^{-\left(\frac{x}{10}\right)}.$$

- (b) *For  $\zeta = 0, 0.1, 0.5, 0.9$ , and 1, compute the mixture  $F(t)$  for a number of values of  $t$  ranging between 0 and 30. Plot these distributions on one graph.*

Figure H2.5 shows the CDF plots for the requested mixture distributions. This plot may be generated with the following R code.



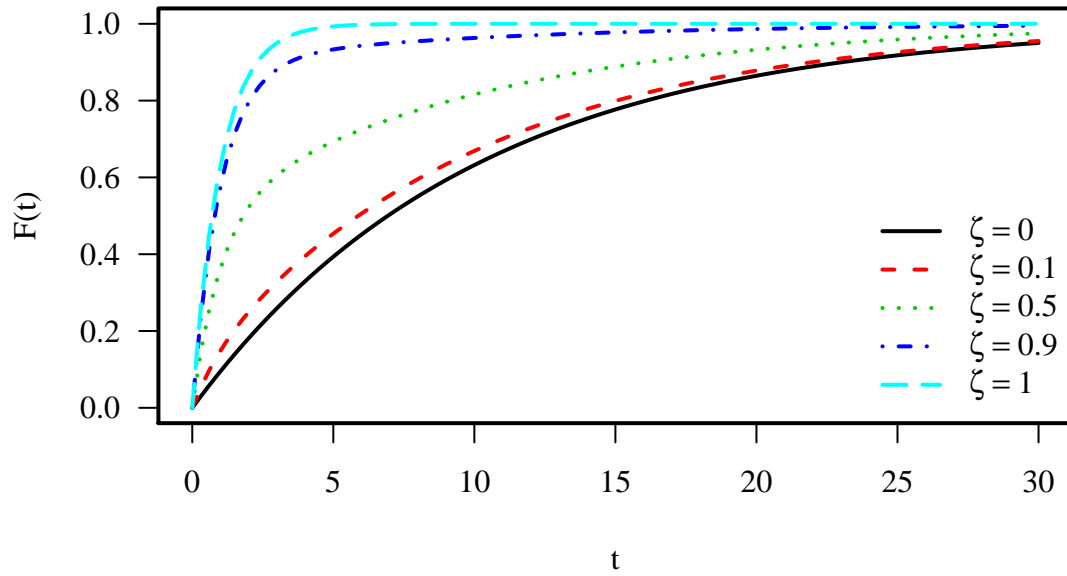


Figure 4: Cdf plots for the mixture distributions

```
par(family = 'serif', lwd = 2, las = 1)

MIX <- function(t,eta) 1-eta*exp(-1*t)-(1-eta)*exp(-1*(t/10))
eta = c(0,0.1,0.5,0.9,1)

for(i in 1:length(eta)) {
  curve(MIX(x, eta = eta[i]),
        add = ifelse(i==1,F,T),
        xlab = 't',
        ylab = 'F(t)',
        col = i,
        lty = i,
        xlim = c(0,30),
        ylim = c(0,1))
}
legend("bottomright",
      c(expression(zeta==0.0),
        expression(zeta==0.1),
        expression(zeta==0.5),
        expression(zeta==0.9),
        expression(zeta==1.0)),
      lty = 1:5,
      col = 1:5,
      bty = "n",
      xjust = 0.5)
```

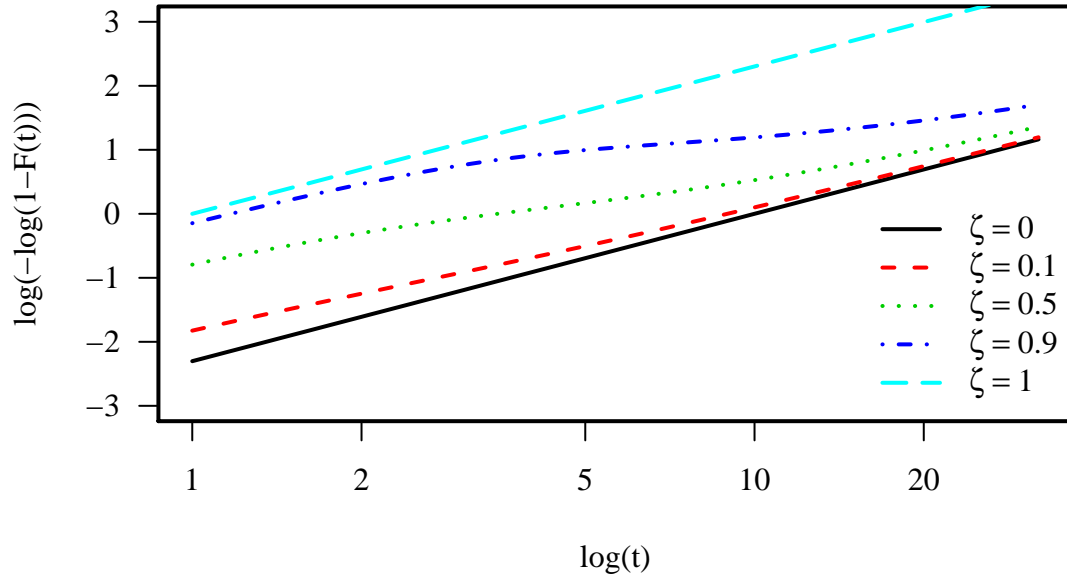


Figure 5: Weibull transformed quantile plots for the mixture distributions

- (c) Plot  $\log(t)$  versus  $\log[-\log[1 - F(t)]]$  for each  $F(t)$  computed in part (b). Comment on the shapes of the mixtures of exponential distributions, relative to a pure exponential distribution or a Weibull distribution.

The same code shown for (b) can be used for (c) applying the transformations.

- (d) Plot the hazard function  $h(t)$  on the mixture distributions in part (b).

As discussed, the hazard function can be expressed as  $h(t) = \frac{f_X(x)}{1-F_X(x)}$ . Though this exercise began with the pdf for the mixture model, the following code can be used to find  $f_X(x)$  symbolically when only the cdf is known.

```
DMIX<-function(t,eta) {
  d.1<-body(MIX)           #Extract body of function MIX
  d.2<-as.expression(d.1)  #Convert body to an expression
  d.3<-D(d.2,"t")          #Take derivative of expression
  d.4<-function(t,eta) eval(d.3) #Convert derivative expression to function
  return(d.4(t,eta))       #Return value of derivative
}
```

Adjusting the plotting code used in parts (b) and (c) results in the following plot of  $h(t)$ .

- (e) Qualitatively, what do the Weibull plots in part (c) suggest about the hazard function of a mixture of two exponential distributions?

The plots shown in (b), (c) and (d) suggest that the mixture of two exponential distributions with  $\eta \in (0, 1)$  is not in the exponential family.

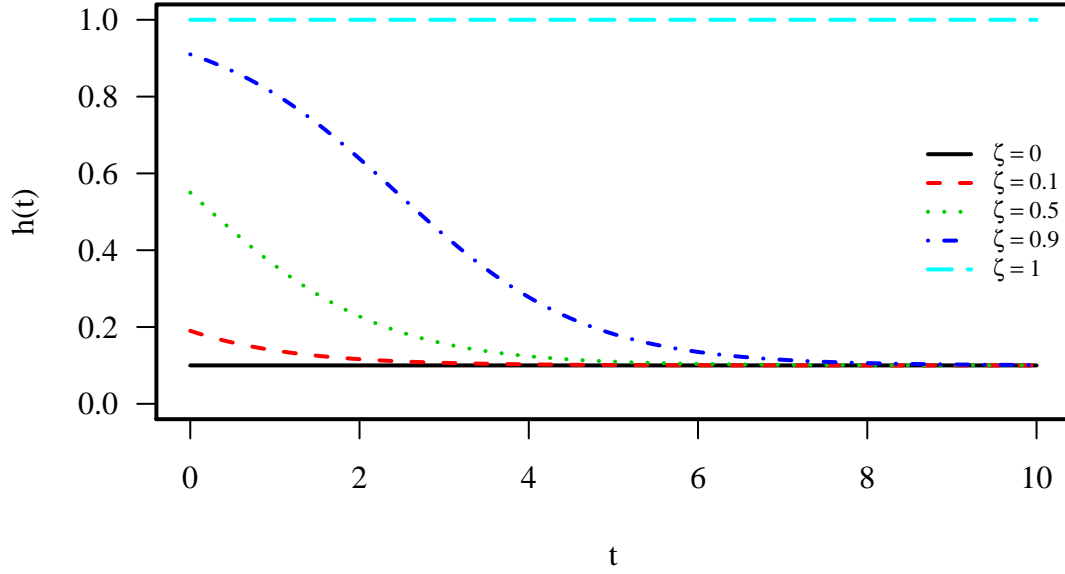


Figure 6: Hazard plots for the mixture distributions

## From the text problem # 5.4

Refer to Exercise 5.1. Show that a mixture of two exponential distributions with different  $\theta$  values with always have a decreasing hazard function

For a mixture of exponentials, where  $\theta_1 \neq \theta_2, \theta_1 > \theta_2$  this problem asks us to show that

$$\frac{dh(x)}{dt} = \left( \frac{f(x)}{S(x)} \right)^2 + \frac{f'(x)}{S(x)} < 0, x < 0.$$

Rearranging this expression gives  $f(x)^2 < -f'(x)S(x)$ . Substituting the density and survival functions for the mixture distribution in this expression results in

$$\left( \frac{\eta}{\theta_1} e^{-\frac{x}{\theta_1}} + \frac{(1-\eta)}{\theta_2} e^{-\frac{x}{\theta_2}} \right) \left( \frac{\eta}{\theta_1} e^{-\frac{x}{\theta_1}} + \frac{(1-\eta)}{\theta_2} e^{-\frac{x}{\theta_2}} \right) < \left( \frac{\eta}{\theta_1} e^{-\frac{x}{\theta_1}} + \frac{(1-\eta)}{\theta_2} e^{-\frac{x}{\theta_2}} \right) \left( \eta e^{-\frac{x}{\theta_1}} + (1-\eta) e^{-\frac{x}{\theta_2}} \right)$$

Expanding this expression gives

$$\frac{\eta^2}{\theta_1^2} e^{-\frac{2x}{\theta_1}} + \frac{2\eta(1-\eta)}{\theta_1\theta_2} e^{-\left(\frac{x}{\theta_1} + \frac{x}{\theta_2}\right)} + \frac{(1-\eta)^2}{\theta_2^2} e^{-\frac{2x}{\theta_2}} < \frac{\eta^2}{\theta_1^2} e^{-\frac{2x}{\theta_1}} + \frac{\eta(1-\eta)}{1} e^{-\left(\frac{x}{\theta_1} + \frac{x}{\theta_2}\right)} \left( \frac{1}{\theta_1^2} + \frac{1}{\theta_2^2} \right) + \frac{(1-\eta)^2}{\theta_2^2} e^{-\frac{2x}{\theta_2}}$$

Simplifying gives

$$\begin{aligned} \frac{2}{\theta_1\theta_2} &< \frac{1}{\theta_1^2} + \frac{1}{\theta_2^2} \\ 0 &< \frac{1}{\theta_1^2} - \frac{2}{\theta_1\theta_2} + \frac{1}{\theta_2^2} \end{aligned}$$

$$\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)^2$$

As we have defined  $\theta_1 > \theta_2$  the quantity  $h'(t) < 0$ .

## From the text problem # 5.12

Let  $T_{(1)}$  denote the minimum of  $m$  independent Weibull random variables with parameters  $\mu_i, i = 1, \dots, m$ , and constant  $\sigma$ . Show that  $T_{(1)}$  has a Weibull distribution.

Let  $T_{(1)} = \min(T_1, \dots, T_m)$  denote the minimum failure time observed from  $m$  independent Weibull random variables having distinct scale parameters  $\mu_i, i = 1, \dots, m$  and an equivalent shape parameter  $\sigma$ . Then

$$\begin{aligned}
 F_{T_{(1)}}(t) &= Pr(T_{(1)} < t) \\
 &= 1 - Pr(T_{(1)} > t) \\
 &= 1 - \prod_{j=1}^m Pr(T_j > t) \\
 &= 1 - \prod_{j=1}^m \exp \left[ - \left( \frac{t}{\exp(\mu_j)} \right)^{\frac{1}{\sigma}} \right] \\
 &= 1 - \exp \left[ - \sum_{j=1}^m \left( \frac{t}{\exp(\mu_j)} \right)^{\frac{1}{\sigma}} \right] \\
 &= 1 - \exp \left[ - t^{\frac{1}{\sigma}} \times \sum_{j=1}^m \left( \frac{1}{\exp(\mu_j)} \right)^{\frac{1}{\sigma}} \right] \\
 &= 1 - \exp \left[ - \left( \frac{t}{\exp[\mu_{(1)}]} \right)^{\frac{1}{\sigma}} \right]
 \end{aligned}$$

which is in the form of a Weibull cdf for which the scale parameter is

$$\exp[\mu_{(1)}] = \left[ \sum_{j=1}^m \left( \frac{1}{\exp(\mu_j)} \right)^{\frac{1}{\sigma}} \right]^{-\sigma}$$