STAT 509 Homework 4 Solution

Chong Ma

March 21, 2017

Problem 1

Note that the time to failure T $\sim$ exponential($\lambda = 1/2000$). Therefore, the pdf and cdf are respectively

$$f_T(t) = \frac{1}{2000}e^{-t/2000}I(t > 0)$$
$$F_T(t) = 1 - e^{-t/2000}I(t > 0)$$

(a) The probability that this piece of equipment will fail before 4000 hours is

$$P(T < 4000) = F_T(4000) = 1 - e^{-4000/2000} = 0.8647$$

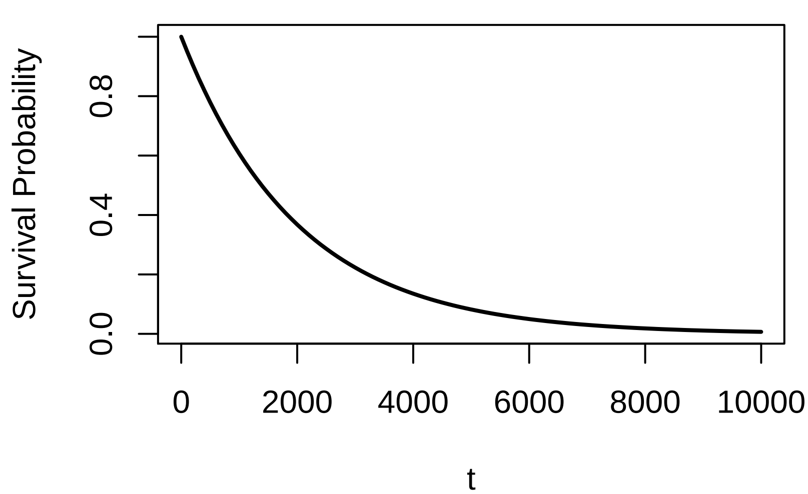
(b) The median of this distribution, $\phi_{0.5}$, satisfies the equation $F_T(\phi_{0.5}) = 1 - e^{-\phi_{0.5}/2000} = 0.5$. Sovling this equation, we can have

$$\phi_{0.5} = 2000\log(2) = 1386.2944$$

(c) Since the hazard function of T is $h(t) = \frac{f_T(t)}{S_T(t)} = \frac{\frac{1}{2000}e^{-t/2000}}{e^{-t/2000}} = \frac{1}{2000}$, therefore the hazard function of T is constant everywhere.

(d) Note that $S_T(t) = 1 - F_T(t) = e^{-t/2000}$, then $S_T(3000) = e^{-3000/2000} = 0.2231$. It indicates that the piece of equipment will survive more than 3000 hours with probability 0.2231.

```
S=function(t) exp(-t/2000)
curve(S,0,10000,xlab="t",ylab="Survival Probability",lwd=2)
```



Problem 2

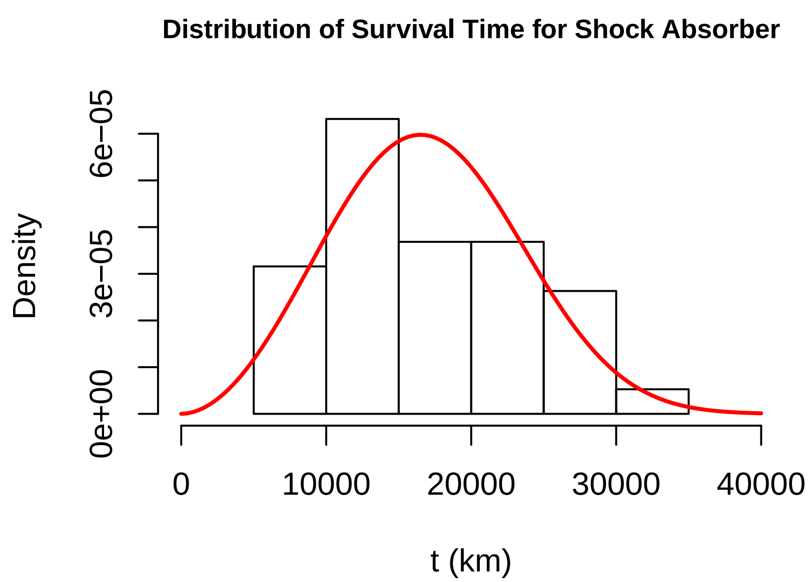
(a) Read the data into R console and plot a histogram superimposed with the fitted Weibull distribution. Under the Weibull distribution assumption, that is, $T \sim \text{Weibull}(\beta, \eta)$, the maximum likelihood estimates for the data are $\hat{\beta} = 2.9$ and $\hat{\eta} = 1.9 \times 10^4$, respectively.

```
shock=c(6700, 6950, 7820, 9120, 9660, 9820, 11310, 11690, 11850,
        11880, 12140, 12200, 12870, 13150, 13330, 13470, 14040,
        14300, 17520, 17540, 17890, 18450, 18960, 18980, 19410,
        20100, 20100, 20150, 20320, 20900, 22700, 23490, 26510,
        27410, 27490, 27890, 28100, 30050)
```

```
fitdistr(shock,densfun = "weibull")
```

```
##       shape          scale
##    2.900248e+00   1.911757e+04
##   (3.671152e-01) (1.136677e+03)
```

```
hist(shock,freq=F,xlim=c(0,40000),xlab="t (km)",cex.main=0.8,
     main="Distribution of Survival Time for Shock Absorber")
curve(dweibull(x,2.9,1.91*10^4),0,40000,col='red',lwd=2,add=TRUE)
```



(b) Note that $T \sim \text{Weibull}(\beta = 2.9, \eta = 1.9 \times 10^4)$, therefore

$$P(T > 15000) = 1 - F_T(15000) = e^{-(15000/19000)^{2.9}} = 0.6042$$

Interpretation: The shock absorber would survive more than 15000 km with probability 0.6042.

(c) The 25th percentile of the distance to failure distribution T satisfies

$$F_T(\phi_{0.25}) = 1 - e^{-(\phi_{0.25}/\eta)^{\beta}} = 0.25$$

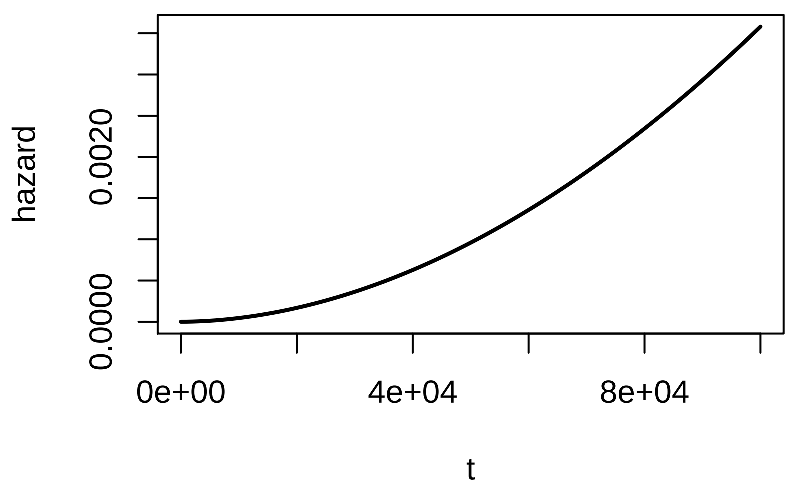
After algebric manipulation, we get $\phi_{0.25} = \eta e^{log(log(\frac{1}{0.75}))/\beta}$. Plug in the according arguments $\hat{\beta} = 2.9$ and $\hat{\eta} = 1.9 \times 10^4$, we have $\phi_{0.25} = 1.2364 \times 10^4$.

(d) Note that the hazard function for Weibull model is

$$h(t) = \frac{f_T(t)}{S_T(t)} = \frac{\beta}{\eta}(\frac{t}{\eta})^{\beta-1}$$

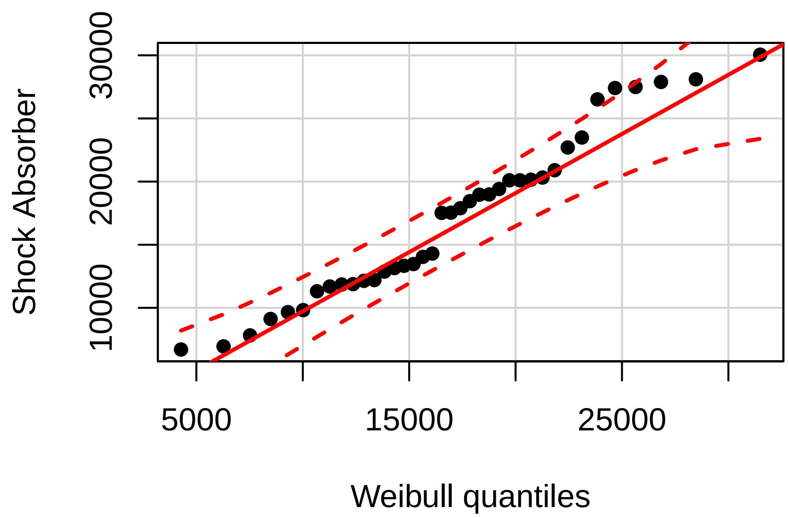
the hazard graph below indicates that the hazard for distance to failure for shock absorber is increasing in an exponential rate.

```
beta.hat=2.9; eta.hat=1.9*10^4
curve(dweibull(x,beta.hat,eta.hat)/pweibull(x,beta.hat,eta.hat,lower.tail=F),0,100000,
      col=1,lwd=2,xlab="t",ylab="hazard")
```



(e) The Weibull QQ plot below shows that there is no serious departure from the assumed weibull model, since almost all data points are along a straight line.

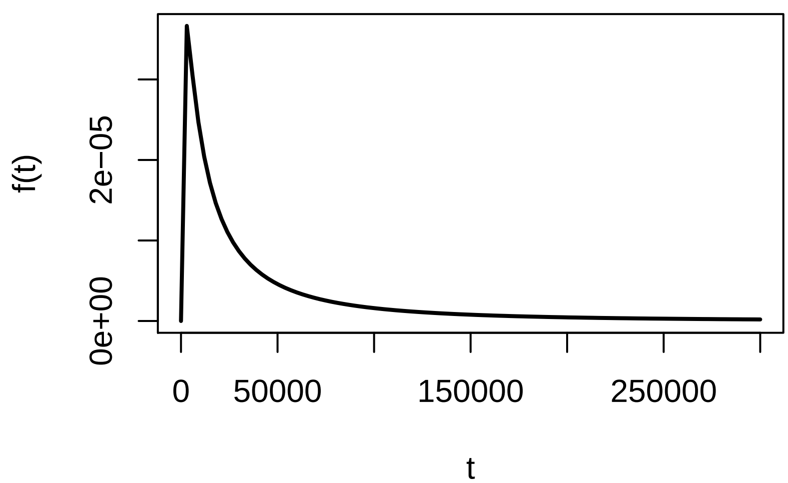
```
qqPlot(shock,distribution="weibull",shape=beta.hat,scale=eta.hat,
       xlab="Weibull quantiles",ylab="Shock Absorber",pch=16)
```



Problem 3

(a) Note that the lifetime T(in hours) of a semiconductor laser follows a lognormal distribution with $\mu = 10$ and $\sigma^2 = 2.25$, that is, $T \sim \text{lognorm}(\mu = 10, \sigma^2 = 2.25)$. The pdf of T is as follows,

```
curve(dlnorm(x,10,sqrt(2.25)),0,300000,lwd=2,xlab="t",ylab="f(t)")
```



(b) The probability that the lifetime exceeds 50,000 hours is

$$P(T > 50000) = \int_{50000}^{\infty} \frac{1}{\sqrt{2\pi}1.5t} exp\left(-\frac{1}{2}\left(\frac{\log(t)-10}{1.5}\right)^2\right) dt = 1 - \text{plnorm}(50000, 10, 1.5) = 0.2924$$

(c) The proportion of lifetimes between 100,000 hours and 200,000 hours is $P(100000 < T < 200000) = \text{plnorm}(200000, 10, 1.5) - \text{plnorm}(100000, 10, 1.5) = 0.0859$

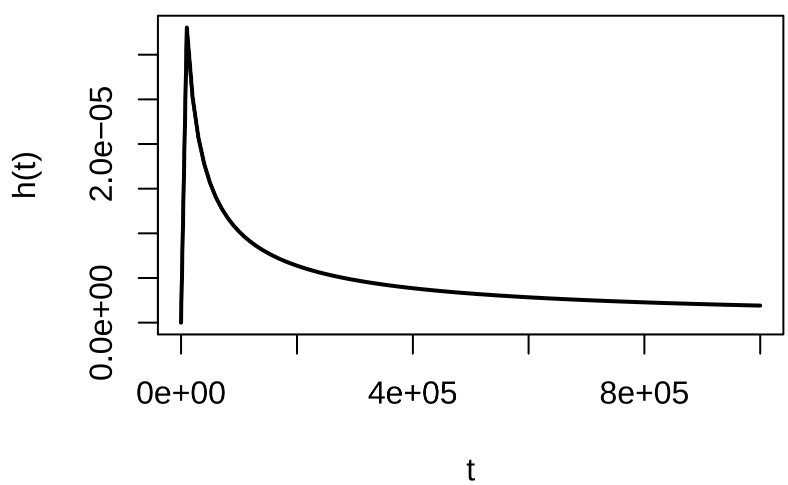
(d) We need find the 90th percentile of the lifetime, which satisfies the equation

$$F_T(\phi_{0.9}) = \int_{-\infty}^{\phi_{0.9}} \frac{1}{\sqrt{2\pi}1.5t} exp\left(-\frac{1}{2}\left(\frac{\log(t)-10}{1.5}\right)^2\right) dt = 0.9$$

There is no closed-form for CDF, but we can readily find the 90th percentile in R. That is, $\phi_{0.9} = \text{qlnorm}(0.9, 10, 1.5) = 1.5059 \times 10^5$.

(e) Obviously the hazard function under the log-normal model is more complex than the one under the weibull model. More importantly, it is less interpretable.

```r
##hazard function given T ~ lognormal(10,2.25)
h=function(t,mu,sigma) dlnorm(t,mu,sigma)/plnorm(t,mu,sigma,lower.tail = F)
curve(h(x,10,1.5),0,1e6,xlab='t',ylab = "h(t)",lwd=2)
```



(f) A possible consequence for using a wrong model might lead to an incorrect warrant claim for the product. For example, under the correct model, the 90th percentile of the lifetime for a product is 50000 hours, whereas the 90th percentile of the lifetime under an incorrect model could be 70000 hours. In this sense, the warranty under the wrong model is longer than it is supposed to be, which could lead to safety security or customer complains.

Problem 4

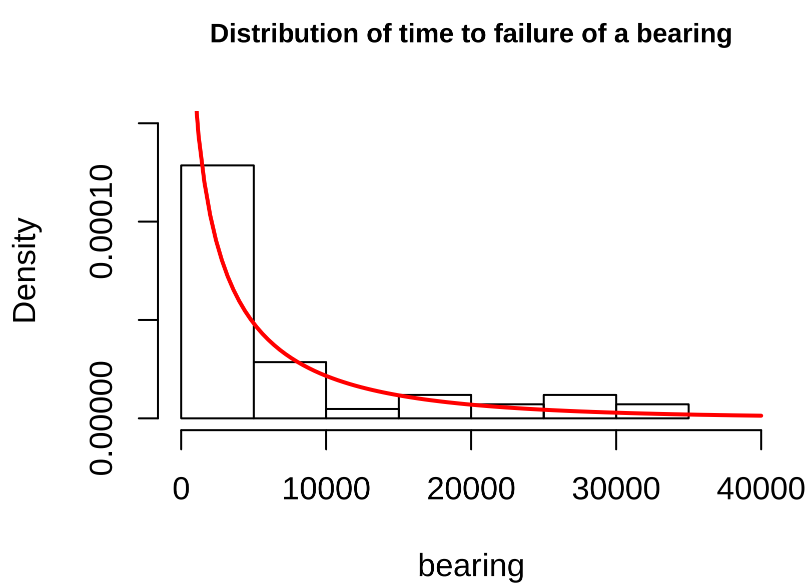
(a) Read the data in R and fit the data under the assumption of Weibull distribution.

```
bearing=c(
27135.9, 3293.5, 16380.9, 3406.6, 12455.7, 1014.4, 4627.3,
22800.0, 425.5, 30859.3, 202.4, 3508.5, 5475.3, 6924.4,
534.9, 288.6, 1363.4, 9200.5, 154.3, 325.2, 4197.5,
5592.2, 682.1, 172.3, 25287.2, 125.0, 1753.7, 1010.5,
5118.9, 2217.6, 175.8, 3142.7, 5094.3, 33154.8, 427.5,
6142.1, 15180.4, 971.4, 103.2, 2691.4, 2406.8, 3961.8,
1814.8, 1074.0, 812.8, 21022.6, 4548.1, 9877.8, 707.9,
1903.9, 2293.7, 2581.4, 31597.1, 25994.2, 3661.3, 2004.0,
238.5, 2164.7, 22304.2, 15.9, 157.8, 17673.4, 143.0,
672.2, 2671.0, 417.8, 5421.3, 290.9, 17286.1, 4263.6,
8492.6, 8885.1, 16947.8, 29890.5, 4102.9, 11009.1, 15.3,
5663.3, 41.4, 2002.8, 1329.6, 29821.9, 139.0, 130.6
)
```

```
##fit the data under the weibull assumption
fitdistr(bearing,densfun = "weibull")
```

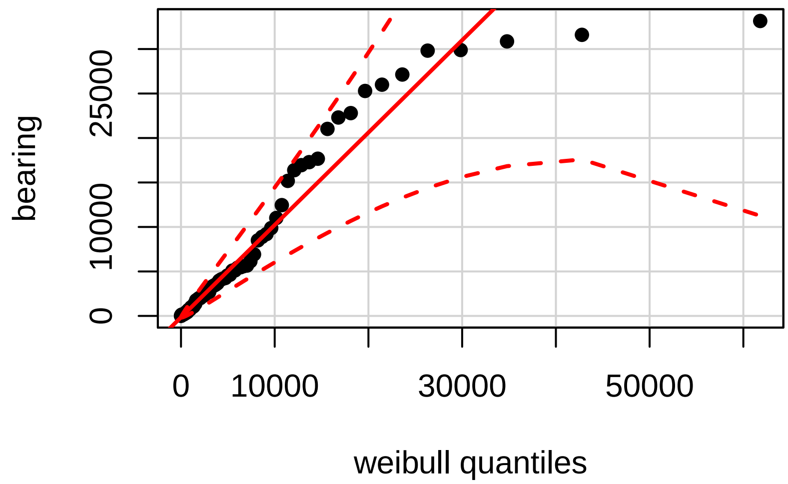
```
##       shape          scale
##   6.556016e-01   5.115829e+03
##  (5.639342e-02) (9.006570e+02)
```

```
## plot the histogram superimposed with the fitted
## weibull distribution
beta.hat=6.56*1e-1; eta.hat= 5.12e3
hist(bearing,freq=F,xlim=c(0,4e4),ylim=c(0,1.5e-4),cex.main=0.8,
     main="Distribution of time to failure of a bearing")
curve(dweibull(x,beta.hat,eta.hat),0,4e4,lwd=2,col="red",add=TRUE)
```



(b) The Weibull qq plot for the bearing data indicates a curvature feature, albeit the data points are within the bounds. It overall implies that the weibull model might not be appropraite for the bearing data.

```
##Weibull qq plot for bearing
qqPlot(bearing, distribution = "weibull",shape=beta.hat,scale=eta.hat,pch=16)
```

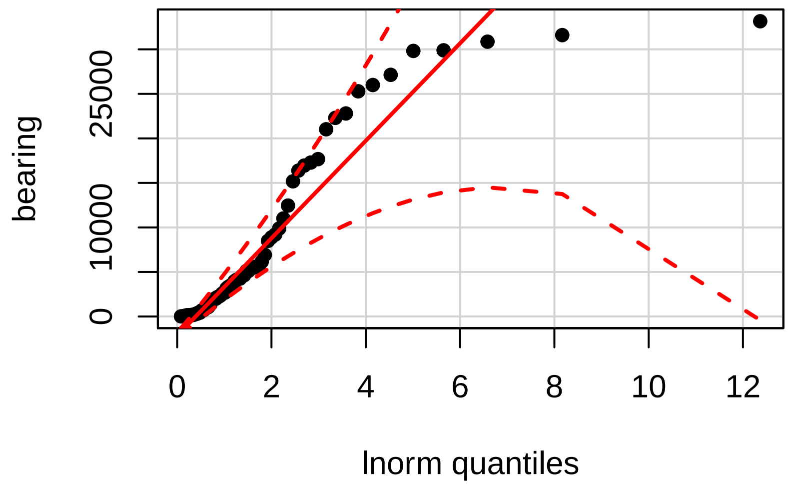


(c-d) By fitting the bearing data using lognormal model, the log normal qqplot also shows a curvature trend, whereas the data points are with the bounds. In contrast to the qq plot under the weibull model, it has more curvature than that for weibull model. In this sense, weibull model is a better one for fitting the bearing data.

```
##Weibull qq plot for bearing
fitdistr(bearing, densfun = "log-normal")
```

```
##     meanlog       sdlog
##   7.6532263   1.8494551
##  (0.2017921) (0.1426886)
```

```
mu=7.65; sigma=1.85
qqPlot(bearing,distribution = "lnorm",pch=16)
```



(e) We fit a very bad model for the bearing data, and the according qq plot seems much more departured from the straightline than either one under the weibull and log-normal models.

```
##Weibull qq plot for bearing
qqPlot(bearing,distribution = "norm",pch=16)
```