Assignment 5

Eric Mittman

February 16, 2016

7.4 (a)
$$\hat{\theta} = \frac{TTT}{r} = \frac{9430589}{4} = 2357647$$

(b)
```
TTT <- (345+556+712+976 + 9428*1000)
mle <- TTT/4
ll <- function(theta) -4 * log(theta) - 1/theta * TTT
rll <- function(theta) ll(theta) - ll(mle)
to_solve <- function(theta) qchisq(.95,1) + 2 * rll(theta)
ci <- NULL
ci[1] <- uniroot(to_solve, c(mle - 2e6,mle))[[1]]
ci[2] <- uniroot(to_solve, c(mle,mle + 1e7))[[1]]
ci

## [1] 1014883 7594289
```

(c) In large sample sizes, a confidence interval so constructed will cover the true value of θ 95% of the time.

(d) This can be interpreted as the failure rate. If there are N active components, then $N \cdot \lambda$ would be the number of expected failures in the next unit of time.

(e) An approximate 95% confidence interval is $(132, 985)$.

(f) We cannot say with 95% confidence that the new component is better. The estimated improvement is 426 FITs, with a 95% confidence interval allowing values between an 84% reduction in hazard to a 16% increase in hazard.

(g) We can observe that there is some discrepancy between the fitted exponential model and what would be the fit for a Weibull model. This discrepancy, which can be observed by the error bar failing to capture the first failure time, suggests there may be a problem with the current model. Since the exponential model is a special case of the Weibull, this could be addressed with a formal likelihood ratio test.

7.9 (a) $TTT = 20 \times 50 = 1000$. Since $\widehat{\theta}_{mle} = \frac{TTT}{r} \to \infty$ as $r \to 0$, the MLE is ∞.
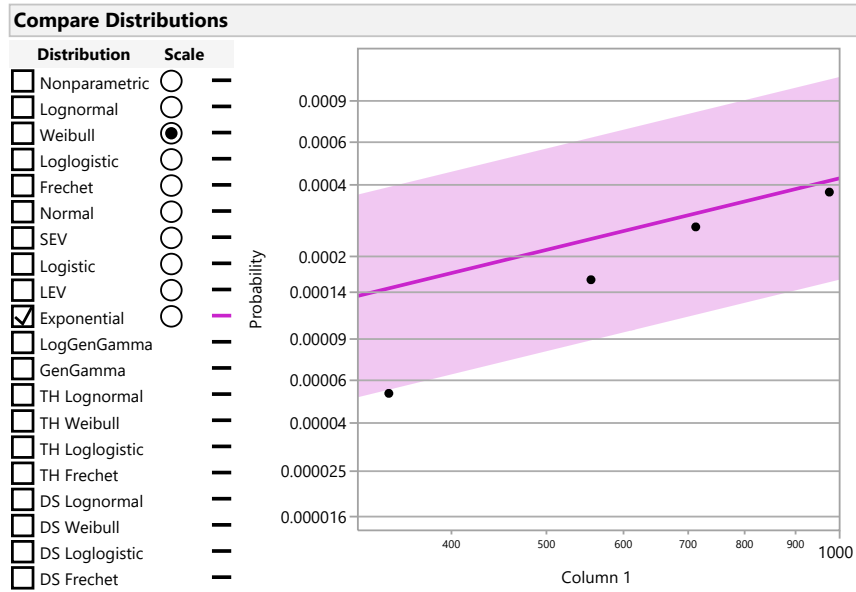
Figure 1: Fitted exponential distribution on Weibull paper for exercise 7.4 (Component A)

(b)
$$\frac{1000}{-\log(.05)} \approx 333.81$$

(c) Yes. This result is strongly dependent on our assumption that the failure times follow an exponential distribution. Units were only observed for 20 hours, and our assumption of constant hazard leads to this large lower bound on θ.

(d)
- lower confidence bound for $t_{.1} = 35.17$
- upper confidence bound for $h(50) = .0030$
- upper confidence bound for $F(50) = .139$.

8.6 Given $SE_{\hat{g}}$, apply the delta method:

$$SE^2_{\log(\hat{g})} \approx \left(\frac{d}{dg}\log(g)\right)^2 \bigg|_{g=\hat{g}} SE_{\hat{g}}$$

$$\implies SE_{\log(\hat{g})} \approx \frac{1}{\hat{g}} SE_{\hat{g}}$$

Using this approximation and transforming a standard normal confidence interval,

$$\left(\log(\hat{g}) - z_{1-\alpha/2} \cdot \frac{1}{\hat{g}} SE_{\hat{g}}, \quad \log(\hat{g}) + z_{1-\alpha/2} \cdot \frac{1}{\hat{g}} SE_{\hat{g}}\right)$$

which yields

$$(\hat{g}/w, \hat{g} \times w), \text{ where } w = \exp(z_{1-\alpha/2} \cdot \frac{1}{\hat{g}} SE_{\hat{g}})$$

7.1 (a) Setting the first derivative of the log likelihood to zero and solving,

$$l(\theta) = -n\log(\theta) - \frac{-1}{\theta}\sum_{i=1}^{r} y_i$$

$$\implies l'(\theta) = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{r} y_i$$

$$l'(\widehat{\theta}) = 0 \iff \widehat{\theta} = \bar{y}$$

(b)
$$R(\theta) = \frac{L(\theta)}{L(\widehat{\theta})}$$

$$= \frac{\theta^{-n}\exp\{-n\bar{y}/\theta\}}{\bar{y}^{-n}\exp\{-n\bar{y}/\bar{y}\}}$$

$$= \exp(n)\left(\frac{\bar{y}}{\theta}\right)^n \exp\left(-\frac{n\bar{y}}{\theta}\right)$$

(c) We can solve for θ where $R(\theta|y) = \chi^2_{1,1-\alpha}$, which should produce two solutions, yielding lower and upper endpoints to a $(1-\alpha)$ approximate confidence interval. (As was done in 7.4(b).)
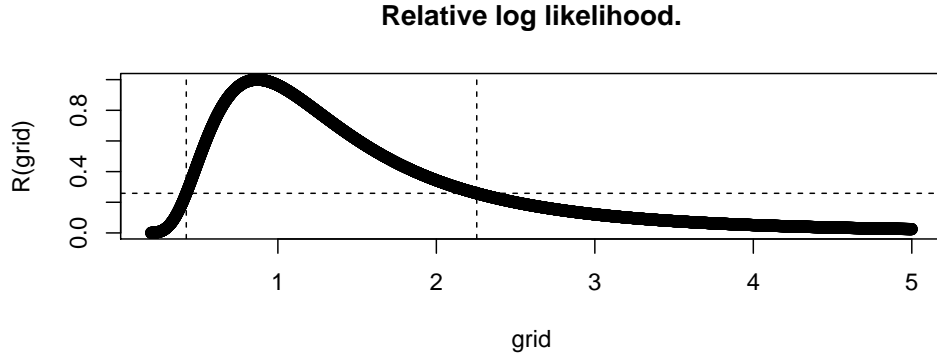
3

(d)
```
n <- 4
ybar <- .87
R <- function(theta) {
   exp(n) * (ybar/theta)^n * exp(-n * ybar / theta)
}
to_solve2 <- Vectorize(function(theta) {2 * log(R(theta)) + qchisq(.9,1)})

low <- uniroot(to_solve2, c(0.2,ybar))[[1]]
high <- uniroot(to_solve2, c(ybar,10))[[1]]

cat(c(low,high))

## 0.4220005 2.254898

grid <- 200:5000 * .001
plot(grid, R(grid), main="Relative log likelihood.")
abline(v=c(low,high),h=exp(-qchisq(.9,1)/2),lty=2)
```

Relative log likelihood.

7.2 (a)

$$L(\theta) = \theta^{-r} \exp\left(-\frac{\sum_{i=1} rt_i)}{\theta}\right) \times \exp\left(-\frac{(n-r)t_c}{\theta}\right)$$

(b)

$$\frac{d}{d\theta} \log(L(\theta)) = \frac{-r}{\theta} + \frac{\sum_{i=1} rt_i + (n-r)t_c}{\theta^2}$$

Setting $\frac{d}{d\theta} \log(L(\hat{\theta})) = 0$,

$$-r\hat{\theta} + \sum_{i=1} rt_i + (n-r)t_c = 0$$

$$\implies \hat{\theta} = \frac{\sum_{i=1} rt_i + (n-r)t_c}{r} = \frac{TTT}{r}$$

(c) Since $t_c > 0$, as $r \to 0$, $\hat{\theta} \to \infty$.

4

(d)
$$R(\theta) = \left(\frac{TTT/r}{\theta}\right)^r \exp(r) \exp\left(\frac{-TTT}{\theta}\right)$$

7.5 This figure is a consistent, unbiased estimate of the mean time to failure *if we assume* that the distribution of time to failure for the population of hard disks follows an exponential distribution. This assumption is false, as these drives wear out over time. This estimate will then be biased high because the hazard function is increasing and is not constant with time.