STAT611 Homework 03

Name: Lu Sun

UIN: 228002579

Question 1: Ex 7.10 in C&B

(a) Find a two-dimensional sufficient statistic for (α, β)

(b) Find the MLEs of α, β

(c) Given the data, find the MLEs of α, β

(a) Find a two-dimensional sufficient statistic for (α, β)

Answer:

$f(x; \alpha, \beta) = 1_{\{0 \leq x \leq \beta\}} \frac{\partial (x/\beta)^\alpha}{\partial x} = \alpha \beta^{-\alpha} x^{\alpha - 1} 1_{\{0 \leq x \leq \beta\}}$

Assume $X = (x_1, ..., x_n), x_{(1)} = \min\limits_{i=1}^{n}\{x_1, ..., x_n\}, x_{(n)} = \max\limits_{i=1}^{n}\{x_1, ..., x_n\}$:

$f(X; \alpha, \beta) \overset{iid}{=} \prod\limits_{i=1}^{n} f(x_i; \alpha, \beta) = \alpha^n \beta^{-n\alpha} (\prod\limits_{i=1}^{n} x_i)^{\alpha - 1} 1_{\{x_{(n)} \leq \beta\}} 1_{\{x_{(1)} \geq 0\}}$

Suppose $T(X) := (\prod\limits_{i=1}^{n} x_i, x_{(n)}), g(T(X); \alpha, \beta) := \alpha^n \beta^{-n\alpha} (\prod\limits_{i=1}^{n} x_i)^{\alpha-1} 1_{\{x_{(n)} \leq \beta\}}, h(X) := 1_{\{x_{(1)} \geq 0\}}$, then by Factorization Theorem,

$T(X) := (\prod\limits_{i=1}^{n} x_i, x_{(n)})$ is the 2-dimensional sufficient statistaic for (α, β)

(b) Find the MLEs of α, β

Answer:

$L(\alpha, \beta; X) := f(X; \alpha, \beta)$

(1) $x_{(1)} < 0$ or $x_{(n)} > \beta$:

$L(\alpha, \beta; X) = 0.$

(2) $0 \leq x_{(1)}, x_{(n)} \leq \beta$, fix α:

$L(\alpha, \beta | X) > 0$ decreases with β. Thus, $\hat{\beta}_{MLE} = x_{(n)}$.

(3) $0 \leq x_{(1)}, x_{(n)} \leq \beta$, fix $\hat{\beta}_{MLE} = x_{(n)}$:

$$
\begin{aligned}
\frac{\partial \ln L(\alpha,\beta|X)}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left[n\ln\alpha - n\alpha\ln\beta + (\alpha-1)\ln\prod_{i=1}^{n} x_i\right] \\
&= \frac{n}{\alpha} - n\ln\beta + \ln\prod_{i=1}^{n} x_i = 0
\end{aligned}
$$

$$\Longrightarrow \hat{\alpha} = \frac{n}{n\ln\beta - \ln\prod_{i=1}^{n} x_i} \overset{\hat{\beta}_{MLE}=x_{(n)}}{=} \left[\frac{1}{n}\sum_{i=1}^{n}(\ln x_{(n)} - \ln x_i)\right]^{-1}$$

Moreover, $\frac{\partial^2 \ln L}{\partial \alpha^2} = -n/\alpha^2 < 0$, so $\hat{\alpha}_{MLE} = \left[\frac{1}{n}\sum_{i=1}^{n}(\ln x_{(n)} - \ln x_i)\right]^{-1}$

By(1)(2)(3), $\hat{\beta}_{MLE} = x_{(n)}, \hat{\alpha}_{MLE} = \left[\frac{1}{n}\sum_{i=1}^{n}(\ln x_{(n)} - \ln x_i)\right]^{-1}$

(c) Given the data, find the MLEs of α, β

Answer:

$\hat{\beta}_{MLE} = x_{(n)} = 25.0$

$\ln\prod_{i=1}^{14} x_i = \sum_{i=1}^{14}\ln x_i = 43.9526978$

$\hat{\alpha}_{MLE} = \left[\frac{1}{n}\sum_{i=1}^{n}(\ln x_{(n)} - \ln x_i)\right]^{-1} = 1/(\ln 25 - 43.9526978/14) = 12.5948692$

Question 2: Ex 7.11 in C&B

(a) Find the MLE of θ and show that its variance goes to 0 as n goes to ∞

(b) Find the method of moments estimator of θ

(a) Find the MLE of θ and show that its variance goes to 0 as n goes to ∞

Answer:

$L(\theta|X) = f(X|\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \Big(\prod_{i=1}^{n} x_i \Big)^{\theta-1}$

$\frac{\partial \ln L}{\partial \theta} = \frac{\partial}{\partial \theta} \Big[n \ln \theta + (\theta - 1) \ln \prod_{i=1}^{n} x_i \Big] = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i = 0, \hat{\theta} = (-\frac{1}{n} \sum_{i=1}^{n} \ln x_i)^{-1}$

$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} < 0$

$\Longrightarrow \hat{\theta}_{MLE} = (-\frac{1}{n} \sum_{i=1}^{n} \ln x_i)^{-1}$

According to $\hat{\theta}_{MLE} = (-\frac{1}{n} \sum_{i=1}^{n} \ln x_i)^{-1}$, the distribution of $(-\sum_{i=1}^{n} \ln x_i)^{-1}$ is needed. Namely, $-\ln x_i$ distribution is needed. Since $f(x_i|\theta) = \theta x_i^{\theta-1}$, $X_i \sim Beta(\theta, 1)$. $Y_i := -\ln X_i \sim Exp(\theta)$, with $P(Y_i \leq y_i) = P(-\ln X_i \leq y_i) = P(X_i \geq \exp(-y_i)) = \int_{x=\exp(-y_i)}^{1} f(x|\theta)dx = 1 - \exp(-\theta y_i)$.

Suppose $T := \sum_{i=1}^{n} Y_i = -\frac{1}{n} \sum_{i=1}^{n} \ln X_i$. Since $Y_i \sim Exp(\theta)$, then $T \sim Gamma(n, \theta), 1/T \sim Inverse - Gamma(n, \theta)$

$Var(\hat{\theta}_{MLE}) = Var(\frac{n}{-\sum_{i=1}^{n} \ln X_i}) = n^2 Var(1/T) \overset{InverseGamma}{=} n^2 \frac{\theta^2}{(n-1)^2(n-2)}$

$\lim_{n\to\infty} Var(\hat{\theta}_{MLE}) = \lim_{n\to\infty} \frac{n^2}{(n-1)^2(n-1)} \theta^2 = 0$

(b) Find the method of moments estimator of θ

Answer:

$X \sim Beta(\theta, 1), E(X) = \frac{\theta}{\theta+1} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\Longrightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n - \sum_{i=1}^{n} x_i}$

Question 3:

(a) Construct a normal probability plot. Is normality plausible?

(b) Construct a Weibull probability plot. Is the Weibull distribution family plausible?

i. Find the quartile of order p, $F^{\leftarrow}(p; a, b)$

ii. Find the sample median and 3rd quartile of the data set

iii. Equate

$$F^{\leftarrow}(1/2; a, b) = Data\ Media$$

$$F^{\leftarrow}(3/4; a, b) = Data\ 3rdQuartile$$

iv. Solver for a,b

v. Make a Weibull QQ plot using $(\hat{a}, \hat{b})$. Dose it look worse than the normal QQ plot?

(a) Construct a normal probability plot. Is normality plausible?

Answer:

The R code for generate the normal probability plot is shown as follow:

```
library(fitdistrplus)
serving = read.csv("LoadLife.csv",header=F)
serving = serving[,]

fit_n <- fitdist(serving,"norm")
rbind(fit_n$estimate)
png('norm.png')
plot.legend <- c("norm")
opar <- par(no.readonly=TRUE)
par(mfrow=c(2,2))
denscomp(list(fit_n), legendtext = plot.legend)
cdfcomp (list(fit_n), legendtext = plot.legend)
qqcomp  (list(fit_n), legendtext = plot.legend)
ppcomp  (list(fit_n), legendtext = plot.legend)
dev.off()
```

The plot is shown as follow:

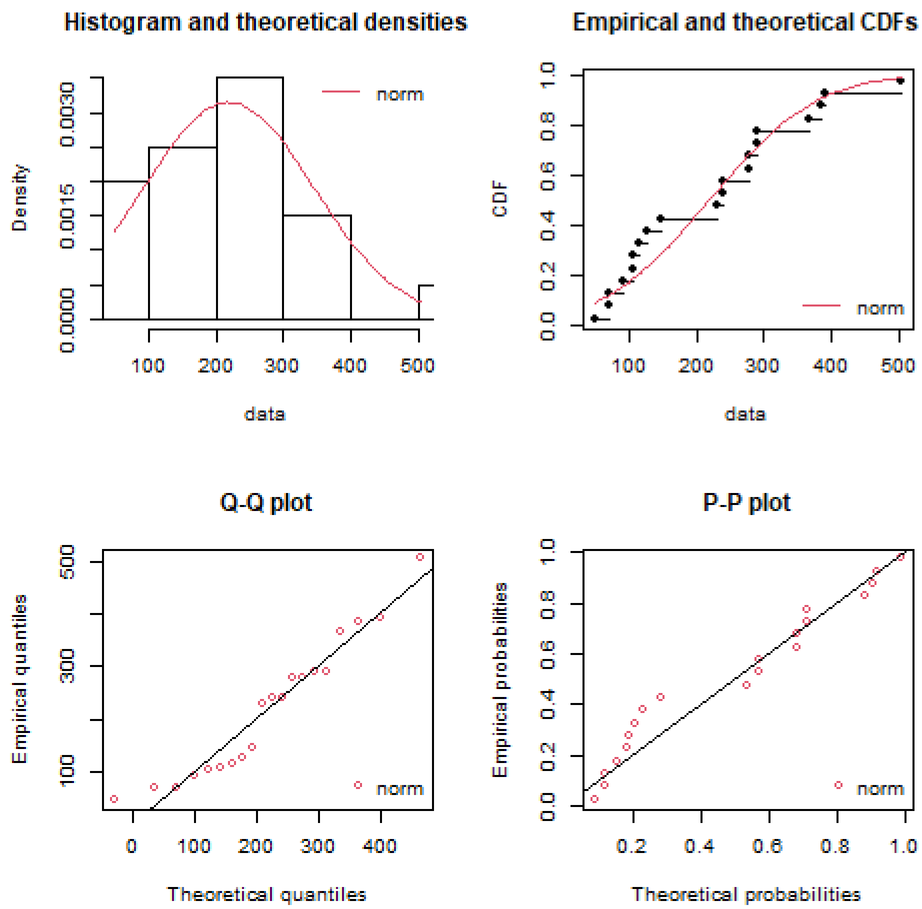

Figure 1: Normal probability plot

From the above plot, normality is plausible.

(b) Construct a Weibull probability plot. Is the Weibull distribution family plausible?

Answer:

The R code for generate the Weibull probability plot is shown as follow:

```
fit_w  <- fitdist(serving, "weibull")
rbind(fit_w$estimate)
png('weibull.png')
plot.legend <- c("weibull")
opar <- par(no.readonly=TRUE)
```

```
par(mfrow=c(2,2))
denscomp(list(fit_w), legendtext = plot.legend)
cdfcomp (list(fit_w), legendtext = plot.legend)
qqcomp  (list(fit_w), legendtext = plot.legend)
ppcomp  (list(fit_w), legendtext = plot.legend)
dev.off()
```

The plot is shown as follow:

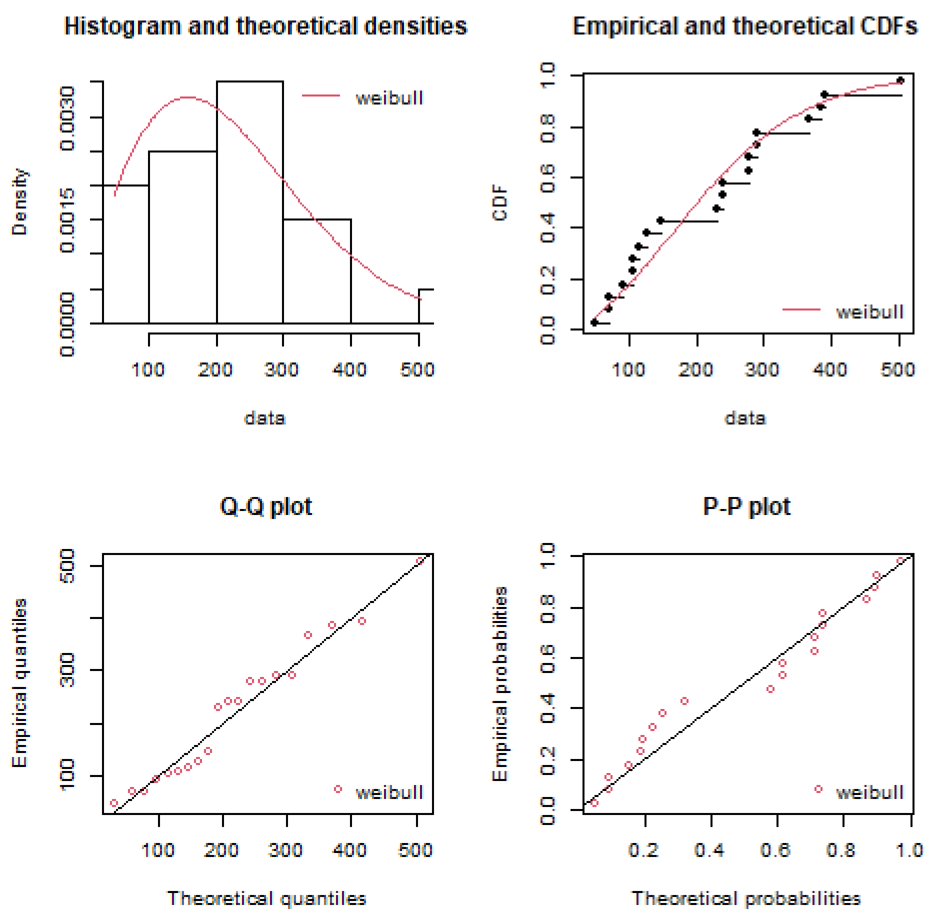

Figure 2: Weibull probability plot

From the above plot, the Weibull distribution family is plausible.

(i) Find the quartile of order p, $F^{\leftarrow}(p; a, b)$

Suppose $x_p = F^{\leftarrow}(p; a, b)$. Then by the definition of quartile, the following equations are obtained.

$$p = P(X \leq x_p) = \int_0^{x_p} f(x; a, b)dx = \int_0^{x_p} \frac{a}{b}(\frac{x}{b})^{(a-1)} \exp[-(x/b)^a]dx$$

$$\implies p = \int_0^{x_p} d(-\exp[-(x/b)^a]) = 1 - \exp[-(x_p/b)^a]$$

$$\implies x_p = b[-\ln(1-p)]^{1/a}$$

Namely,
$F^{\leftarrow}(p; a, b) = b[-\ln(1-p)]^{1/a}$

(ii) Find the sample median and 3rd quartile of the data set

By the following R code, the median and 3rd quartile are shown in the Table 1.

```r
summary(serving)
```

median	3rd quartile
234.5	289.0

Table 1: Median and 3rd quartile

(iii) Equate

$$F^{\leftarrow}(1/2; a, b) = Data\ Media$$

$$F^{\leftarrow}(3/4; a, b) = Data\ 3rdQuartile$$

Suppose $x_{1/2} = F^{\leftarrow}(1/2; a, b), x_{3/4} = F^{\leftarrow}(3/4; a, b), \hat{X} := (x_1, ..., x_n)$ the data set, x_{12} Data Media, x_{34} Data 3rd Quartile. By the definition of $x_{1/2}, x_{3/4}$ in (i), the following formula can be obtained:

$$P(X \leq x_{1/2}) = 1/2 \approx \frac{1}{n} \sum_{x_i \in \hat{X}} I_{\{x_i \leq x_{12}\}} = P(\hat{X} \leq x_{12})$$

$$P(X \leq x_{3/4}) = 3/4 \approx \frac{1}{n} \sum_{x_i \in \hat{X}} I_{\{x_i \leq x_{34}\}} = P(\hat{X} \leq x_{34})$$

$\implies x_{1/2} = F^{\leftarrow}(1/2; a, b) \approx Data\ Media = x_{12}$,
$x_{3/4} = F^{\leftarrow}(3/4; a, b) \approx Data\ 3rd\ Quartile = x_{34}$

(iv) Solver for a,b

By the equation in (i) and (iii),

$$x_{12} = b[-\ln(1-1/2)]^{1/a}, x_{34} = b[-\ln(1-3/4)]^{1/a}$$
$$\Longrightarrow x_{12} = b(\ln 2)^{1/a}, x_{34} = b(2\ln 2)^{1/a}$$
$$\Longrightarrow \hat{a} = \frac{\ln 2}{\ln x_{34} - \ln x_{12}}, \hat{b} = \frac{x_{12}}{(\ln 2)^{1/\hat{a}}} = \frac{x_{12}}{(\ln 2)^{(\ln x_{34} - \ln x_{12})/\ln 2}}$$

(v) Make a Weibull QQ plot using $(\hat{a}, \hat{b})$. Dose it look worse than the normal QQ plot?

First, $(\hat{a}, \hat{b})$ is computed according to formula in (iv) and value in (ii). The value is shown as follow:

a (shape)	b(scale)
3.316952	261.8973

Table 2: value of $\hat{a}, \hat{b}$

Then according to the empirical quantiles in the data set, the theoretical quntiles of Weibull distribution $(\hat{a}, \hat{b})$ and norm distribution (mean and std in Question 3(a)) are generated with the following code.

```
rk = (rank(serving,ties.method = "first")-0.5)/length(serving)
theo_norm = qnorm(rk, mean = fit_n$estimate[1],
                    sd = fit_n$estimate[2],
                    lower.tail = TRUE, log.p = FALSE)
x_12 = summary(serving)[3]
x_34 = summary(serving)[5]
a = log(2)/(log(x_34)-log(x_12))
b = x_12/((log(2))^(1/a))
theo_weilbull = qweibull(rk, shape = a, scale = b,
                        lower.tail = TRUE, log.p = FALSE)
```

The code for plot is show as follow:

```
png('compare.png')
par(pin = c(4,2.75))
plot(theo_norm,serving,xlab = "Theoretical quantiles",
    ylab="Empirical quantiles", col="red",
    main = "Q-Q plot",xlim = c(-20,500),ylim = c(20,500))
lines(0:500, 0:500, col="black")
par(new=TRUE)
plot(theo_weilbull,serving,col="green",new= TRUE,
    axes = FALSE, xlab = "Theoretical quantiles",
     ylab="Empirical quantiles",xlim = c(-20,500),
     ylim = c(20,500),pch=c(3))
legend("bottomright", inset=.05, title="Distribution Type",
```

```
        c("Norm","Weibull"),pch=c(1, 3), col=c("red", "green"))
dev.off()
```

The Weibull QQ plot is shown as follow: From the above plot, it looks worse

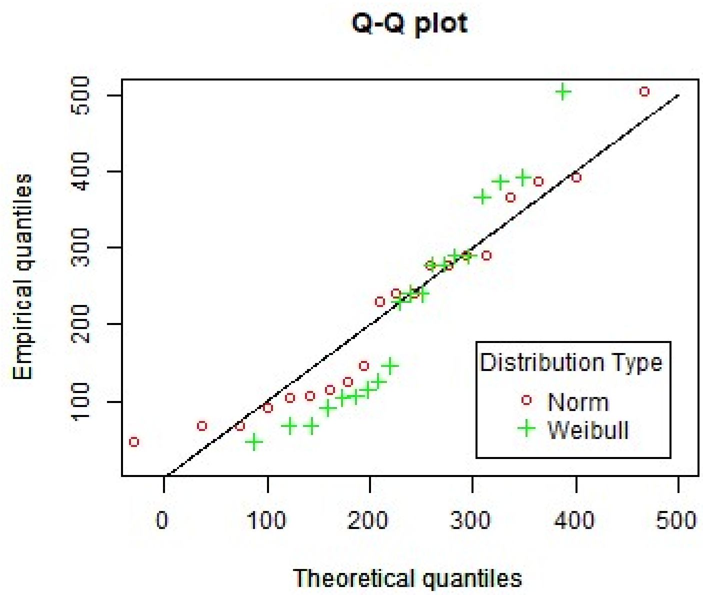

Figure 3: QQ plot

than the normal QQ plot, whose distance with $y = x$ is bigger than that of normal distribution.