Statistical Theory Problem set 3

Bohao Tang

Problem 1: Theoretical Part:

In this problem we use likelihood in part (ii) and assumption in part (iii) of the previous problem set:

(i) $$E[Y_i \mid I_i = 1, \theta] = \frac{\cancel{\theta_2}\, E[Y_i^2]}{E[Y_i]} = \frac{\theta_2 + \theta_1^2}{\theta_1}$$

$$var[Y_i \mid I_i = 1, \theta] = \frac{E[Y_i^3]}{E[Y_i]} - \left[\frac{E[Y_i^2]}{E[Y_i]}\right]^2 = \frac{\theta_2 + \theta_1^2}{\theta_1} \cdot \frac{\theta_2}{\theta_1}$$

$$\Rightarrow \theta_1 = E[Y_i \mid I_i = 1, \theta] - \frac{var[Y_i \mid I_i = 1, \theta]}{E[Y_i \mid I_i = 1, \theta]}$$

$$\theta_2 = \frac{Var[Y_i \mid I_i = 1, \theta]}{E[Y_i \mid I_i = 1, \theta]} \cdot \theta_1$$

$\Rightarrow$ the moment estimator of θ_1, θ_2 are

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{500} Y_i}{500} - \frac{\sum_{i=1}^{500} (Y_i - \bar{Y})^2}{499} \Big/ \bar{Y}$$

$$\theta_2 = \hat{\theta}_1 \cdot \left[\frac{\sum_{i=1}^{500} (Y_i - \bar{Y})^2}{499} \Big/ \bar{Y} \right]$$

(ii) Recall the likelihood will be.

$$\Pr(Y^{obs} \mid I^{obs}, n_1=500, \theta, \alpha) = \frac{(\theta_1/\theta_2)^{500\cdot\frac{\theta_1^2}{\theta_2}}}{\Gamma\left[\frac{\theta_1^2}{\theta_2}\right]^{500}\theta_1^{500}} \cdot \left(\prod Y_i\right)^{\theta_1^2/\theta_2} \cdot e^{-\frac{\theta_1}{\theta_2}\sum Y_i}$$

$$-\log[\text{Likelihood}] = 500\log\Gamma\left[\frac{\theta_1^2}{\theta_2}\right] + 500\log\theta_1 - 500\frac{\theta_1^2}{\theta_2}\log\frac{\theta_1}{\theta_2}$$

$$-\left[\theta_1^2/\theta_2\right]\underbrace{\log\left(\prod Y_i\right)}_{\sum\log(Y_i)} + \frac{\theta_1}{\theta_2}\sum Y_i$$

(iii) Pass., since we minimize the minus log likelihood we want the hessian to ~~to~~ be positive definite.

(iv) denote θ_p be the 95th percentile of the distribution. then θ_p is function of θ_1, θ_2, we have ~~$\theta_p =$ qgamma(0.95,~~

$$\theta_p = \text{qgamma}\left(0.95,\ \text{shape} = \frac{\theta_1^2}{\theta_2}+1,\ \text{scale} = \frac{\theta_2}{\theta_1}\right)$$

so $$\hat{\theta}_{p\ mle} = \text{qgamma}\left(0.95,\ \text{shape} = \frac{\hat{\theta}_{1\ mle}^2}{\hat{\theta}_{2\ mle}}+1,\ \text{scale} = \frac{\hat{\theta}_{2\ mle}}{\hat{\theta}_{1\ mle}}\right)$$

(v) Pass.

Problem 2: Theoretical Part.

(i) The likelihood function is :

$$L(X_1 \cdots X_{50} \mid \mu) = \prod_{i=1}^{50} e^{-\mu} \frac{\mu^{X_i}}{X_i!} \sim e^{-50\mu} \mu^{\sum_{i=1}^{50} X_i}$$

To maximize L, is equivilent to maximize $e^{-50\mu} \mu^{\sum X_i}$

use derivative we have $\hat{\mu}_{MLE} = \dfrac{\sum_{i=1}^{50} X_i}{50} = \bar{X}$

(ii) $g(\mu) = pr(X=0 \mid \mu) = e^{-\mu} \dfrac{\mu^0}{0!} = e^{-\mu}$

So the MLE for $g(\mu)$ is $\hat{g}(\mu) = g(\hat{\mu}_{MLE}) = e^{-\bar{X}}$

the estimator is biased because $g(\mu)$ is convex

so $E\, e^{-\bar{X}} \geq e^{-E\bar{X}} = e^{-\mu}$

and the equality is satifised only when $\bar{X}$ is a.s. a constant which is not always the case.

(iii) Simply $\overline{X_i^*} = \sum_{i=1}^{50} X_i^* / 50$ is an unbiased estimator of $g(\mu)$

because $E\overline{X^*} = E\, 1(X_i=0) = P(X=0) = e^{-\mu} = g(\mu)$

(iv) The likelihood is $pr(X_1 \cdots X_{50} \mid \mu) = \left(\prod_{i=1}^{50} \dfrac{1}{X_i!} \right) e^{-50\mu} \mu^{\sum_{i=1}^{50} X_i}$

$$pr(\bar{X} \mid \mu) = \sum_{X_1+X_2+\cdots+X_{50} = 50\bar{X}} \left(\prod_{i=1}^{50} \frac{1}{X_i!} \right) e^{-50\mu} \mu^{50\bar{X}}$$

$$= \left(\sum_{\substack{X_1+X_2+\cdots+X_{50}=50\bar{X} \\ X_i \geq 0}} \frac{(X_1+\cdots X_{50})!}{X_1!\, X_2! \cdots X_{50}!} \right) \frac{e^{-50\mu} \mu^{50\bar{X}}}{(50\bar{X})!}$$

$$= e^{-50\mu} (50\mu)^{50\bar{X}} / (50\bar{X})!$$

Then $pr(X_1 \cdots X_{50}, \bar{X} | \mu) = \left(\prod_{i=1}^{50} \frac{1}{X_i!} \right) e^{-50\mu} \mu^{50\bar{X}} 1_{\{X_1 + \cdots X_{50} = 50\bar{X}\}}$

$$\Rightarrow pr(X_1 | \bar{X}, \mu) = \sum_{X_2, X_3 \cdots X_{50}} pr(X_1 \cdots X_{50}, \bar{X} | \mu) / pr(\bar{X} | \mu)$$

$$= \sum_{X_2 + X_3 + \cdots X_{50} = 50\bar{X} - X_1} \frac{(50\bar{X})!}{X_1! \cdots X_{50}!} \left(\frac{1}{50}\right)^{50\bar{X}}$$

$$= \sum_{X_2 + \cdots + X_{50} = 50\bar{X} - X_1} \frac{(50\bar{X})!}{X_1!\,(50\bar{X} - X_1)!} \frac{(50\bar{X} - X_1)!}{X_2! X_3! \cdots X_{50}!} \left(\frac{1}{50}\right)^{50\bar{X}}$$

$$= \binom{50\bar{X}}{X_1} \left(\frac{49}{50}\right)^{50\bar{X} - X_1} \left(\frac{1}{50}\right)^{X_1} \sim \text{Binom}\left(50\bar{X}, \frac{1}{50}\right)$$

(V) $\hat{g}(\mu)_{BW} = E\left[\bar{X^*} \mid \bar{X}\right]$

$$= \frac{1}{50} \sum_{i=1}^{50} E\left[1(X_i = 0) \mid \bar{X}\right]$$

$$= \frac{1}{50} \cdot 50 \cdot P\left[X = 0 \mid \bar{X}, \mu\right] = \left(\frac{49}{50}\right)^{50\bar{X}}$$

Since Likelihood $\sim e^{-50\mu} \mu^{\sum_{i=1}^{50} X_i}$, $\bar{X}$ is sufficient statistics.

and for $\sum_{i=1}^{50} X_i$, its distribution is Poisson(50μ)

so $\forall g$ if $E\left[g\left(\sum_{i=1}^{50} X_i\right) \mid 50\mu\right] = 0$ $\forall \mu \geq 0$

Then ~~s~~ $\sum_{k=0}^{+\infty} g(k) \frac{e^{-50\mu} (50\mu)^k}{k!} = 0$ $\forall \mu \geq 0$

$\Rightarrow \sum_{k=0}^{+\infty} \frac{g(k)}{k!} (t)^k = 0$ $\forall t \geq 0$ since the uniqueness of power series

$g(k) = 0$ for all k $\Rightarrow g = 0 \Rightarrow \sum_{i=1}^{50} X_i$ is complete statistics

so as $\bar{X}$

$\Rightarrow \bar{X}$ is sufficient and complete so $\hat{g}(\mu)_{BW}$ is MVUE.

Coding Part

Problem 1

(i).

```r
Y = source("data8.txt")$value

m = mean(Y)
v = var(Y)

theta1 = m - v / m
theta2 = theta1 * v / m
mme = c(theta1,theta2)
print(mme)
```

```
## [1]    80.32726 11879.65785
```

(ii).

```r
f <- function(Y){
    function(para){
        t1 = para[1]
        t2 = para[2]
        500 * log(gamma(t1^2/t2)) + 500 * log(t1) -
        500 * t1^2 / t2 * log(t1/t2) -
        (t1^2/t2 - 1) * sum(log(Y)) + t1 / t2 * sum(Y)
    }
}

mle = optim(mme, f(Y))
```

```
## Warning in log(t1): NaNs produced

## Warning in log(t1/t2): NaNs produced

## Warning in log(t1): NaNs produced

## Warning in log(t1/t2): NaNs produced

## Warning in log(t1): NaNs produced

## Warning in log(t1/t2): NaNs produced

## Warning in log(t1): NaNs produced

## Warning in log(t1/t2): NaNs produced
```

```r
print(mle$par)
```

```
## [1]    90.97896 12481.86544
```

(iii). The eigenvalues of hessian are all positive, therefore hessian is positive definite, so this is indeed the local minimum of minus log likelihood, which is the local maximum of likelihood.

```r
mle = optim(mme, f(Y), hessian = T)
eigenvalues = svd(mle$hessian)$d
print(eigenvalues)
```

```
## [1] 2.693534e-02 9.405684e-07
```

(iv).

```r
t1 = mle$par[1]
t2 = mle$par[2]
p_mle = qgamma(0.95, shape = t1^2 / t2 + 1, scale = t2 / t1)
print(p_mle)
```

```
## [1] 574.3776
```

(V).

```r
estimator <- function(path){
    Y = source(path)$value
    m = mean(Y)
    v = var(Y)
    theta1 = m - v / m
    theta2 = theta1 * v / m
    mme = c(theta1,theta2)
    mle = optim(mme, f(Y))$par
    mle
}

theta12 = c()
for(i in 1:10){
    path = paste("data", i, ".txt",sep = "")
    theta12 = rbind(theta12, estimator(path))
}
print(theta12)
```

```
##            [,1]     [,2]
##  [1,]  69.78007 11005.55
##  [2,]  84.28554 13234.61
##  [3,] 101.84094 13087.02
##  [4,]  96.01610 12650.58
##  [5,]  90.91716 11963.89
##  [6,]  94.40069 12315.02
##  [7,]  89.07779 12758.49
##  [8,]  90.97896 12481.87
##  [9,]  80.31894 10673.38
## [10,] 109.80055 13355.51
```

```r
var1 = var(theta12[,1])
var2 = var(theta12[,2])
cov12 = cov(theta12[,1], theta12[,2])

print(var1)
```

```
## [1] 124.4669
```

```r
print(var2)
```

```
## [1] 820024.6
```

```r
print(cov12)
```

```
## [1] 7559.009
```

Problem 2

(i).

```
X = source("p2data")$value

u_mle = mean(X)
print(u_mle)
```

```
## [1] 3.56
```

(ii).

```
gu_mle = exp(-mean(X))
print(gu_mle)
```

```
## [1] 0.02843882
```

(iii).

```
X0 = (X == 0)
gu_ub = mean(X0)
print(gu_ub)
```

```
## [1] 0.06
```

(v).

```
gu_BW = (49 / 50)^(50 * mean(X))
print(gu_BW)
```

```
## [1] 0.02743099
```