

ST 793

Homework-1 : Solution

Salil Koner

1. Problem: 2.2

```
require(ggplot2)
require(gridExtra)

# CDF
pextval<-function(x,location,scale){
  exp(-exp(-(x-location)/scale))
}

#Quantile
qextval<-function(t,location,scale){
  -scale*log(-log(t))+location
}

# Density
dextval<-function(x, location=0, scale=1, log=T){
  t <- (x - location)/scale
  ld<- -t - exp(-t)
  if (log==T){
    ft <- ld - log(scale)
  }
  else {
    ft <- exp(ld)/scale
  }
  return(ft)
}

#####
#          Finding log-likelihood of           #
#          extreme value distribution        #
#####

llk.extval<- function(mu,sigma, data=flow.max){
  sum(dextval(x=data, location=mu, scale=sigma, log=T))
}

#####
#          Optimizing negative log-likelihood      #
#####

nllk.extval <- function(theta, data=flow.max){
  -llk.extval(mu=theta[1], sigma=theta[2], data=data)
}
```

```

# Data

flow.max<-c(5550, 4380, 2370, 3220, 8050, 4560, 2100,
           6840, 5640, 3500, 1940, 7060, 7500, 5370,
           13100, 4920, 6500, 4790, 6050, 4560, 3210,
           6450, 5870, 2900, 5490, 3490, 9030, 3100,
           4600, 3410, 3690, 6420, 10300, 7240, 9130)

out <- nlm(nllk.extval, c(1,2), data=flow.max)

*****#
#      Maximum likelihood estimates      #
*****#

muhat<-out$estimate[1]
sigmahat<-out$estimate[2]

muhat

## [1] 4395.145

sigmahat

## [1] 1882.495

*****#
#      Finding Q-Q plot and CDF plot      #
#      of extreme value distribution      #
*****#

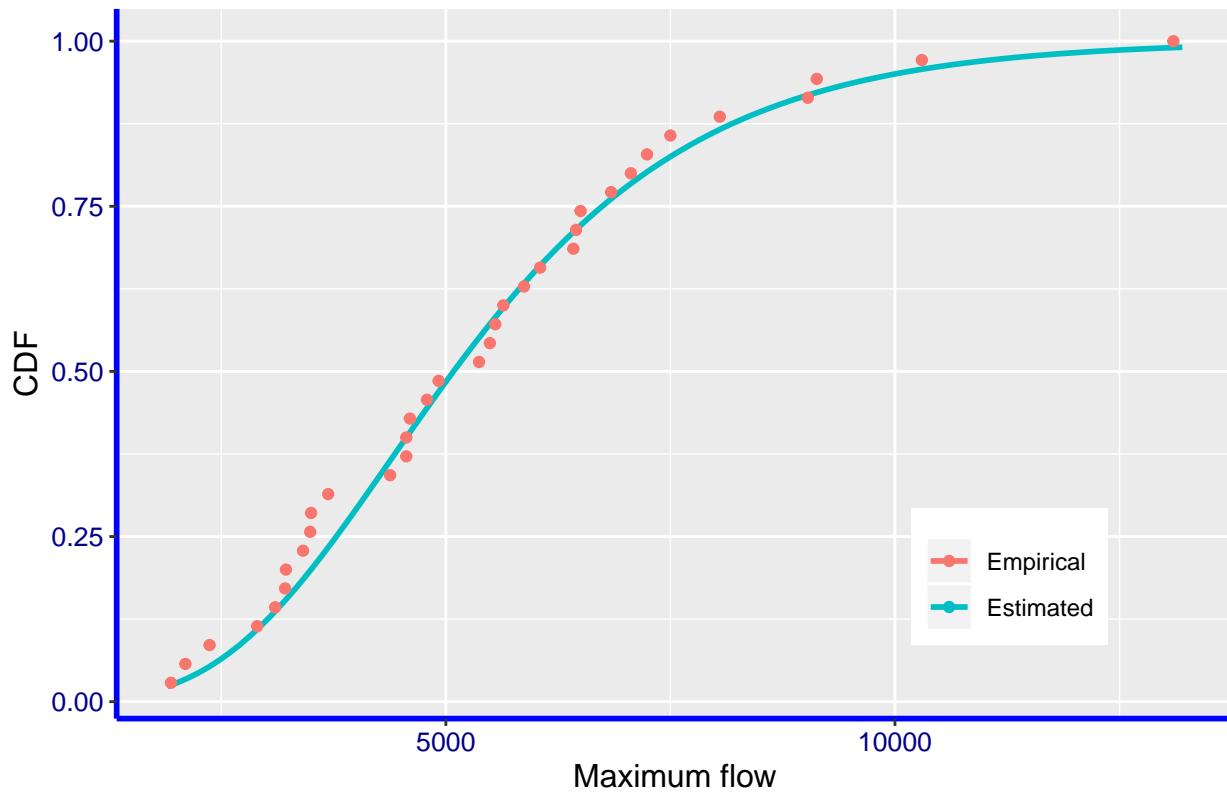
x<-seq(1900,13200,length.out=100) # a grid of values
y<-pextval(x,muhat,sigmahat) # est. cdf for grid
ht <- 1:35/35 # heights for empirical cdf

d1<- data.frame(flow=x, CDF=y, type='Estimated')
d2<- data.frame(flow=sort(flow.max), CDF=ht, type='Empirical')

cdfplot<-ggplot() +
  geom_line(data=d1, aes(flow, CDF, colour='Estimated') , size=1) +
  geom_point(data=d2, aes(flow, CDF, colour='Empirical')) +
  labs(x='Maximum flow', title='Empirical and Estimated CDF (using MLE') +
  theme(plot.title = element_text(color="blue", size=14, hjust = 0.5),
        axis.title = element_text(size=12),
        axis.text = element_text(size=10, colour='darkblue'),
        axis.line = element_line(colour = "blue",
                                 size = 1, linetype = "solid"),
        legend.title = element_blank(),
        legend.position = c(0.8, 0.2))

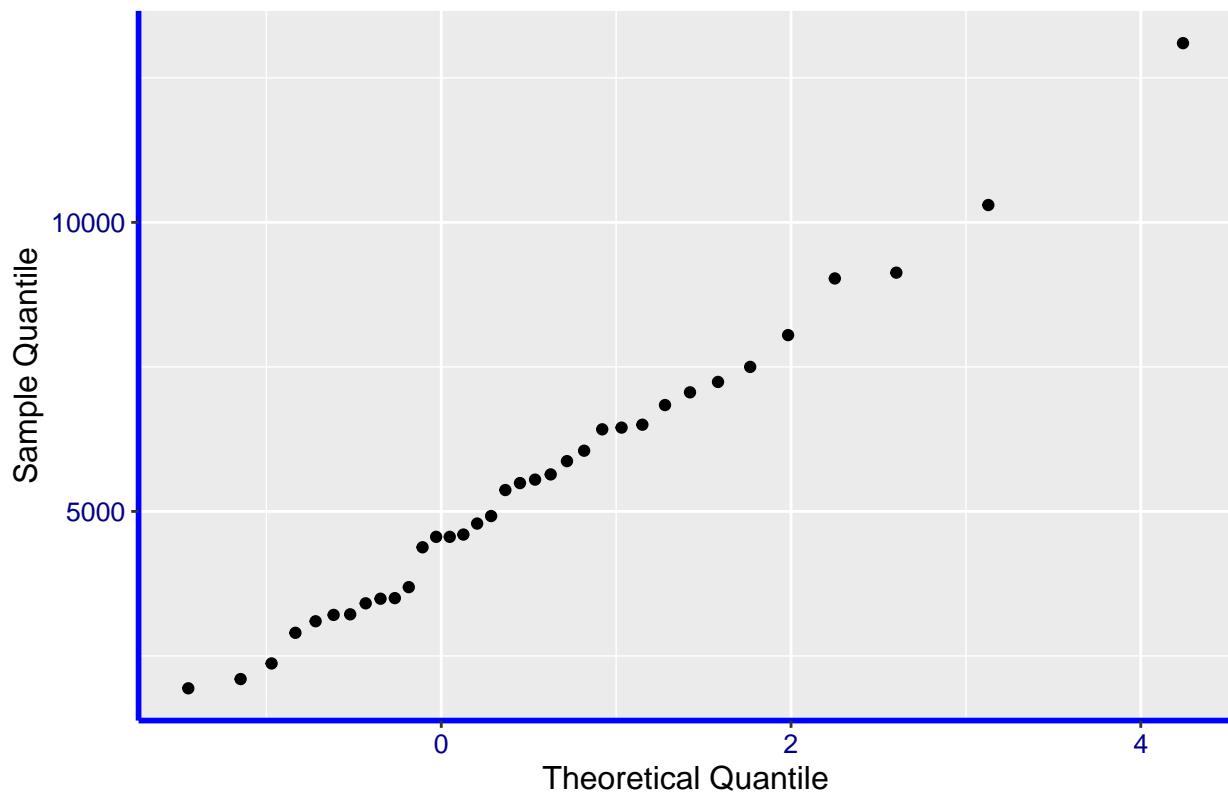
```

Empirical and Estimated CDF (using MLE)



```
qqplot<-qplot(qextval(ppoints(flow.max),0,1),sort(flow.max)) +
  labs(x='Theoretical Quantile', y='Sample Quantile', title='Q-Q plot') +
  theme(plot.title = element_text(color="blue", size=14, hjust = 0.5),
        axis.title = element_text(size=12),
        axis.text = element_text(size=10, colour='darkblue'),
        axis.line = element_line(colour = "blue",
                                size = 1, linetype = "solid"))
```

Q–Q plot



Problem 2

$$2.3.a) \quad \pi = p + (1-p)e^{-\lambda}$$

$$\Rightarrow \pi = p(1-e^{-\lambda}) + e^{-\lambda}$$

$$\Rightarrow \pi - e^{-\lambda} = p(1-e^{-\lambda})$$

$$\Rightarrow p = \frac{\pi - e^{-\lambda}}{1 - e^{-\lambda}} = \frac{\pi e^{\lambda} - 1}{e^{\lambda} - 1}.$$

$$\text{and } 1-p = 1 - \frac{\pi e^{\lambda} - 1}{e^{\lambda} - 1} = \frac{1-\pi}{1-e^{-\lambda}}.$$

Then, the reparametrized model can be rewritten as,

$$P(Y=0) = \pi \quad \text{and}$$

$$P(Y=y) = \frac{(1-\pi)}{(1-e^{-\lambda})} e^{-\lambda} \cdot \frac{\lambda^y}{y!}, \quad y=1, 2, 3, \dots$$

$$\underline{b)} \quad \text{Define } \delta_i = \begin{cases} 1 & \text{if } Y_i = 0 \\ 0 & \text{if } Y_i > 0. \end{cases}$$

Then, the distribution of Y_i under the reparametrized model is

$$f_{Y_i}(y_i) = \pi^{\delta_i} \left[\frac{(1-\pi)}{(e^{-\lambda}-1)} \frac{\lambda^{y_i}}{y_i!} \right]^{1-\delta_i}$$

for all $i=1, 2, \dots, n$.

Note that, $\sum_{i=1}^n \delta_i = n_0 \equiv \text{the number of zeroes in the sample.}$

under the independence assumption, the likelihood function is given by

$$\begin{aligned}
 L(\lambda, \pi) &= \prod_{i=1}^n f_{Y_i}(\gamma_i | \lambda, \pi) \\
 &= \prod_{i=1}^n \left\{ \pi^{\delta_i} \left[\frac{(1-\pi)}{(e^\lambda - 1)} \frac{\lambda^{\gamma_i}}{\gamma_i!} \right]^{1-\delta_i} \right\} \\
 &= \pi^{\sum_{i=1}^n \delta_i} (1-\pi)^{n - \sum_{i=1}^n \delta_i} \frac{\lambda^{\sum_{i=1}^n \gamma_i (1-\delta_i)}}{(e^\lambda - 1)^{n - \sum_{i=1}^n \delta_i} \left(\prod_{i=1}^n \pi^{\gamma_i} \gamma_i! \right)^{1-\delta_i}} \\
 &= \left[\pi^{n_0} (1-\pi)^{n-n_0} \right] \left[\frac{\lambda^{(n-n_0)\bar{Y}_+}}{(e^\lambda - 1)^{n-n_0} \prod_{i=1}^n (\gamma_i!)^{1-\delta_i}} \right] \\
 &= g(\pi) h(\lambda).
 \end{aligned}$$

$$\text{where, } g(\pi) = \pi^{n_0} (1-\pi)^{n-n_0}$$

$$h(\lambda) = \frac{\lambda^{(n-n_0)\bar{Y}_+}}{(e^\lambda - 1)^{n-n_0} \prod_{i=1}^n \pi^{\gamma_i} (\gamma_i!)^{1-\delta_i}}$$

$$\sum_{i=1}^n \gamma_i (1-\delta_i) = \text{sum of all non-zero values} = (n-n_0) \bar{Y}_+$$

Thus, we see that the likelihood function is split into two parts, $g(\pi)$; the part involves only π and $h(\lambda)$ involves only λ . As the likelihood function is separable with respect to π and λ , it is equivalent to optimize π and λ separately.

Taking log of $g(\pi)$ and $h(\lambda)$ separately, we get,

$$\log g(\pi) = n_0 \log \pi + (n-n_0) \log(1-\pi)$$

$$\begin{aligned} \log h(\lambda) &= (n-n_0) \bar{Y}_+ + \log \lambda - (n-n_0) \log(e^\lambda - 1) + \sum_{i=1}^n (\delta_i - 1) \log Y_i \\ &= (n-n_0) \left[\bar{Y}_+ + \log \lambda - \log(e^\lambda - 1) \right] + c(\bar{Y}_+). \end{aligned}$$

Differentiating $\log g(\pi)$ and $\log h(\lambda)$ w.r.t π and λ respectively, we get, after equating to 0,

$$\frac{\partial}{\partial \pi} \log g(\pi) = \frac{n_0}{\pi} - \frac{n-n_0}{1-\pi} = 0$$

$$\Rightarrow \boxed{\hat{\pi} = \frac{n_0}{n} = \frac{n_0}{n}} = 0$$

$$\frac{\partial}{\partial \lambda} \log h(\lambda) = (n-n_0) \left[\bar{Y}_+ - \frac{e^\lambda}{e^\lambda - 1} \right] = 0$$

$$\Rightarrow \frac{\lambda e^\lambda}{e^\lambda - 1} = \bar{Y}_+$$

$$\Rightarrow \boxed{\hat{\lambda} = \bar{Y}_+ (1 - e^{-\bar{Y}_+})} \Rightarrow \text{the non-linear equation involving } \bar{Y}_+.$$

To make sure that we are really maximizing log-likelihood
we take second derivative of $g(\pi)$ and $h(x)$.

$$\begin{aligned} \left. \frac{\partial^2}{\partial \pi^2} \log g(\pi) \right|_{\pi = \hat{\pi}} &= - \frac{n_0}{\hat{\pi}^2} - \frac{n-n_0}{(1-\hat{\pi})^2} \Bigg|_{\pi = \hat{\pi}} \\ &= - \left[\frac{n_0}{\hat{\pi}^2} + \frac{n-n_0}{(1-\hat{\pi})^2} \right] \\ &= - \left[\frac{n^2}{n_0} + \frac{n^2}{n-n_0} \right] = - \frac{n^3}{n_0(n-n_0)} < 0 \end{aligned}$$

As, we don't have an explicit expression for MLE of λ , it is tricky to show that the second derivative is indeed negative at $\lambda = \hat{\lambda}$.

As, we proved that $\hat{\lambda} \geq \lambda$ satisfies,

$$\hat{\lambda} = \bar{y}_+ (1 - \bar{e}^{\hat{\lambda}})$$

$$\Rightarrow \boxed{\bar{y}_+ \bar{e}^{\hat{\lambda}} = \bar{y}_+ - \hat{\lambda}} \quad \dots \dots \textcircled{1}$$

Now, $\bar{y}_+ > 0$ and $\bar{e}^{\hat{\lambda}} > 0$ implies,

$$\bar{y}_+ - \hat{\lambda} \geq 0 \Rightarrow \boxed{\bar{y}_+ > \hat{\lambda}} \quad \dots \textcircled{2}$$

\rightarrow we will need $\textcircled{1}$ and $\textcircled{2}$ while showing

$$\left. \frac{\partial^2}{\partial \lambda^2} \log h(\lambda) \right|_{\lambda = \hat{\lambda}} < 0.$$

$$\left. \frac{\partial^2}{\partial \lambda^2} \log h(\lambda) \right|_{\lambda=\hat{\lambda}}$$

$$= (n - n_0) \left[- \frac{\bar{y}_+}{\hat{\lambda}^2} + \frac{e^{\hat{\lambda}}}{(e^{\hat{\lambda}} - 1)^2} \right]$$

$$= (n - n_0) \left[- \frac{\hat{\lambda} e^{\hat{\lambda}}}{(e^{\hat{\lambda}} - 1) \hat{\lambda}^2} + \frac{e^{\hat{\lambda}}}{(e^{\hat{\lambda}} - 1)^2} \right]$$

Putting
 $\bar{y}_+ = \frac{\lambda e}{e^{\lambda} - 1}$

$$= - \frac{(n - n_0) e^{\hat{\lambda}}}{\cancel{\lambda} (e^{\hat{\lambda}} - 1)} \left[\frac{1}{\hat{\lambda}} - \frac{1}{e^{\hat{\lambda}} - 1} \right]$$

$$= - \frac{n - n_0}{(1 - e^{-\hat{\lambda}})} \left[\frac{e^{\hat{\lambda}} - 1 - \hat{\lambda}}{\hat{\lambda} (e^{\hat{\lambda}} - 1)} \right]$$

$$= - \frac{(n - n_0)}{\hat{\lambda} (e^{\hat{\lambda}} - 1) (1 - e^{-\hat{\lambda}})} \left(\sum_{k=2}^{\infty} \frac{\hat{\lambda}^k}{k!} \right)$$

[Because, $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$]

As, $e^{\hat{\lambda}} > 1$ as $\hat{\lambda} > 0$.

$$\left. \frac{\partial^2}{\partial \lambda^2} \log h(\lambda) \right|_{\lambda=\hat{\lambda}} < 0.$$

So, $(\hat{\pi}, \hat{\lambda})$ is indeed maximizes the likelihood function and hence they are the MLE for (π, λ) .

c) conditional distribution of Y given $Y > 0$ is

$$P(Y=y | Y > 0) = \begin{cases} P(Y=y | Y > 0) & \text{if } y > 0 \\ 0 & \text{if } y=0 \end{cases}$$

$$= \begin{cases} P(Y=y) / P(Y > 0) & \text{if } y > 0 \\ 0 & \text{if } y=0 \end{cases}$$

$$= \begin{cases} \frac{(1-\pi) e^{-\lambda} \times y}{(1-e^{-\lambda}) y!} / (1 - P(Y=0)) & \text{if } y > 0 \\ 0 & \text{if } y=0 \end{cases}$$

$$= \begin{cases} \frac{e^{-\lambda} \times y}{(1-e^{-\lambda}) y!} & \text{if } y > 0 \\ 0 & \text{if } y=0 \end{cases}$$

$$\boxed{\text{Because, } P(Y=0) = \pi}$$

Note that, this is free of π , and involves only one parameter, λ .

conditional likelihood of non-zero values based on the above conditional distribution is given by,

$$\begin{aligned}
 L_c(\lambda | \underline{y}) &= \prod_{\substack{i=1 \\ y_i > 0}}^n \left(\frac{1}{e^\lambda - 1} \right) \frac{\lambda^{y_i}}{(y_i!)} \\
 &= \frac{\left(e^\lambda - 1 \right)^{-(n-n_0)} \lambda^{\sum_{i=1}^n (\delta_i - 1)}}{\prod_{\substack{i=1 \\ y_i > 0}}^n (y_i!)}. \\
 &= \left(e^\lambda - 1 \right)^{-(n-n_0)} \lambda^{\sum_{i=1}^n (\delta_i - 1)} / \prod_{\substack{i=1 \\ y_i > 0}}^n (y_i!).
 \end{aligned}$$

Taking logarithm & and differentiating and equating to zero given,

$$\frac{\partial}{\partial \lambda} \log L_c(\lambda) = \frac{\partial}{\partial \lambda} \left[-(n-n_0) \log(e^\lambda - 1) + (n-n_0) \bar{Y}_+ \log \lambda + c(\underline{y}) \right]$$

$$\Rightarrow 0 = -(n-n_0) \frac{e^\lambda}{e^\lambda - 1} + (n-n_0) \bar{Y}_+ / \lambda$$

$$\Rightarrow \bar{Y}_+ = \frac{\hat{\lambda} e^{\hat{\lambda}}}{e^{\hat{\lambda}} - 1} \quad \text{or,} \quad \hat{\lambda} = \bar{Y}_+ (1 - e^{-\hat{\lambda}})$$

which is the same equation we got from part (b).

Problem 3

2.16. $y_i^{(x)} = \alpha_i^T \beta + e_i, \quad i=1(1)n.$

given, $e_i \stackrel{iid}{\sim} N(0, \sigma^2)$

By the property of normal distribution,

$$y_i^{(x)} \stackrel{\text{ind}}{\sim} N(\alpha_i^T \beta, \sigma^2)$$

Note that, $y_i^{(x)}$ are not identically distributed.

To find the likelihood fn. based y_i 's, we need to find their distribution.

$$P(y_i \leq t | x_i) = P(y_i^{(x)} \leq t^{(x)})$$

$$= \cancel{P(\alpha_i^T \beta + e_i \leq t)}$$

$$= P\left(\frac{y_i^{(x)} - \alpha_i^T \beta}{\sigma} \leq \frac{t^{(x)} - \alpha_i^T \beta}{\sigma}\right)$$

$$= \Phi\left(\frac{t^{(x)} - \alpha_i^T \beta}{\sigma}\right)$$

Taking derivative w.r.t t , we get the distribution of y_i

$$f_{y_i}(y_i) = \frac{1}{\sigma} \phi\left(\frac{y_i^{(x)} - \alpha_i^T \beta}{\sigma}\right) \left| \frac{d}{dt} t^{(x)} \right|_{t=y_i}$$

[ϕ is PDF of $N(0, 1)$ and Φ is the CDF of $N(0, 1)$]

As, y_i 's are independently distributed, likelihood function is given by

$$\begin{aligned}
 L(\beta, \sigma, \gamma | \{(y_i, x_i)\}_{i=1}^n) &= \prod_{i=1}^n f_{Y_i}(y_i | \beta, \sigma, \gamma, x_i) \\
 &= \frac{1}{\sigma^n} \prod_{i=1}^n \phi\left(\frac{y_i^{(\gamma)} - x_i^T \beta}{\sigma}\right) \left| \frac{\partial t^{(\gamma)}}{\partial t} \Big|_{t=y_i} \right| \\
 &= \sigma^{-n} (\sqrt{2\pi})^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \cancel{y_i^{(\gamma)}} \left(y_i^{(\gamma)} - x_i^T \beta\right)^2\right) \times \\
 &\quad \prod_{i=1}^n \left| \frac{\partial t^{(\gamma)}}{\partial t} \Big|_{t=y_i} \right|
 \end{aligned}$$

Hence we proof. \square

Problem 4

2.22. a)

$$Y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma^2), \quad i=1(1)K, \quad j=1(1)n_i;$$

Likelihood function of $(Y_{11}, \dots, Y_{1n_1}, \dots, Y_{K1}, \dots, Y_{Kn_K})$ is

$$L(\mu_1, \dots, \mu_K, \sigma^2) = \prod_{i=1}^K \prod_{j=1}^{n_i} f_{Y_{ij}}(y_{ij} \mid \mu_i, \sigma^2).$$

$$= \prod_{i=1}^K \prod_{j=1}^{n_i} \frac{1}{(\sqrt{2\pi}\sigma)} \exp\left(-\frac{(y_{ij} - \mu_i)^2}{2\sigma^2}\right).$$

$$= (\sqrt{2\pi}\sigma)^{-\sum n_i} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2\right).$$

Log-likelihood fn. is given by

$$\ell(\mu_1, \dots, \mu_K, \sigma^2) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 - \frac{N}{2} \log 2\pi$$

differentiating w.r.t μ_1, \dots, μ_K and σ^2 and equating to 0,

we get

$$\frac{\partial \ell}{\partial \mu_i} = -\cancel{\frac{N}{2} \log} - \frac{1}{2\sigma^2} \sum_{j=1}^{n_i} 2(y_{ij} - \mu_i)(-1) = 0$$

$$\Rightarrow \boxed{\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} = \bar{y}_i, \quad i=1(1)K.}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2$$

$$= \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \frac{\text{SSE}}{N}$$

Thus, MLE of μ_i is $\hat{\mu}_i = \bar{y}_i$ and that of $\hat{\sigma}^2$ is

$$\boxed{\hat{\sigma}^2 = \frac{\text{SSE}}{N}}$$

Taking 2nd derivative of log-likelihood function,

$$\frac{\partial^2 \ell}{\partial \mu_i^2} \Bigg|_{(\mu_1, \dots, \mu_K, \sigma^2) = (\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{\sigma}^2)} = -\frac{n_i}{\hat{\sigma}^2} < 0$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \mu_i \partial \sigma^2} \Bigg|_{(\mu_1, \sigma^2) = (\hat{\mu}_1, \hat{\sigma}^2)} &= -\frac{1}{\hat{\sigma}^4} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i) \\ &= -\frac{1}{\hat{\sigma}^4} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = 0 \quad [\text{as, } \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}] \end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \sigma^4} \Bigg|_{(\mu_1, \hat{\sigma}^2) = (\hat{\mu}_1, \hat{\sigma}^2)} = \frac{N}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2$$

$$= \frac{N}{2\hat{\sigma}^4} - \frac{\text{SSE}}{\hat{\sigma}^6} = \frac{N}{2\hat{\sigma}^4} - \frac{N}{\hat{\sigma}^4}$$

$$= -\frac{N}{2\hat{\sigma}^4} < 0$$

The Hessian matrix is given by,

$$H = \text{diag}\left(-\frac{n_1}{\hat{\sigma}^2}, -\frac{n_2}{\hat{\sigma}^2}, \dots, -\frac{n_K}{\hat{\sigma}^2}, -\frac{N}{2\hat{\sigma}^4}\right)$$

which is clearly negative definite.

So, $\hat{\mu}_i = \bar{y}_i$ and $\hat{\sigma}^2 = \frac{SSE}{N}$ is indeed MLE of μ_i, σ^2 .

by Expand $\bar{y}_i = \sum_{j=1}^n y_{ij}/n_i = \left(\frac{1}{n_i}, \dots, \frac{1}{n_i}\right) \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}$

then, clearly we can write

$$\begin{aligned} \tilde{v}_i &= \begin{pmatrix} 1 - \frac{1}{n_i} & -\frac{1}{n_i} & \dots & -\frac{1}{n_i} & -\frac{1}{n_i} \\ -y_{n_i} & 1 - \frac{1}{n_i} & \dots & -y_{n_i} & -\frac{1}{n_i} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{1}{n_i} & -\frac{1}{n_i} & \dots & 1 - \frac{1}{n_i} & -\frac{1}{n_i} \end{pmatrix} \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} \\ &= A \tilde{y}_i \quad \text{where } A \in \mathbb{R}^{(n_i-1) \times n_i} \end{aligned}$$

$$\text{Note that, } A = \left[I_{n_i-1} \quad -\frac{1}{n_i} J_{n_i-1} \quad \vdots \quad -\frac{1}{n_i} \quad \underbrace{1}_{n_i-1} \right]$$

where, I_K is K th order identity matrix,

J_K is $K \times K$ matrix of all the entries equal to 1.
i.e $J_K = \frac{1}{n_K} \mathbf{1}_K^T$ and $\mathbf{1}_K$ is vector of length K with
all the coordinates as 1..

$$\text{then, } AAT = \left[I_{n_i-1} - J_{n_i-1} \cdot \frac{1}{n_i} \quad \vdots \quad -\frac{1}{n_i} \mathbb{1}_{n_i-1} \right] \begin{bmatrix} I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \\ -\frac{1}{n_i} \mathbb{1}_{n_i-1}^T \end{bmatrix}$$

$$= \left(I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \right) \left(I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \right) + \cancel{\frac{1}{n_i^2} I_{n_i-1} \mathbb{1}_{n_i-1}^T} \\ + \frac{1}{n_i^2} \mathbb{1}_{n_i-1} \mathbb{1}_{n_i-1}^T$$

$$= \left(I_{n_i-1} - \frac{2}{n_i} J_{n_i-1} + \frac{1}{n_i^2} J_{n_i-1} J_{n_i-1}^T \right) + \frac{1}{n_i^2} J_{n_i-1}$$

$$= \left(I_{n_i-1} - \frac{2}{n_i} J_{n_i-1} + \frac{(n_i-1)}{n_i^2} J_{n_i-1} \right) + \frac{1}{n_i^2} J_{n_i-1}$$

$$= I_{n_i-1} + J_{n_i-1} \left(\frac{1}{n_i^2} + \frac{(n_i-1)}{n_i^2} - \frac{2}{n_i} \right)$$

$$= I_{n_i-1} - \frac{1}{n_i} J_{n_i-1}.$$

Also, $(AAT)^{-1} = I_{n_i-1} + J_{n_i-1}$. [This comes from following theorem]

Theorem: for any matrix B of the form $B = aI_K + bJ_K$,

- $B^{-1} = cI_K + dJ_K$ where
- $c = \frac{1}{a}$ and $d = -\frac{b}{a(a+bK)}$

\Rightarrow Prove the theorem and check $(AAT)^{-1} = I_{n_i-1} + J_{n_i-1}$

$$\begin{bmatrix} J_K & J_K \\ 1_K 1_K^T & 1_K 1_K^T \\ K 1_K 1_K^T \\ K J_K \end{bmatrix}$$

So, we showed that,

$$\tilde{v}_i = A \tilde{y}_i \text{ where } \tilde{y}_i \sim N_{n_i}(\mu_i \mathbf{1}_{n_i}, \sigma^2 I_{n_i})$$

$$\text{and } AAT = I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \text{ and } (AT)^T = I_{n_i-1} + J_{n_i-1}.$$

Then, By the property of Multivariate Normal distribution

$$\tilde{v}_i \sim \text{Normal}_{(n_i-1)}\left(\mu_i \mathbf{1}_{n_i}^\top, \sigma^2 (AAT)\right)$$

$$\text{Now, } A \mathbf{1}_{n_i} = \left[I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} : -\frac{1}{n_i} \mathbf{1}_{n_i-1}^\top \right] \begin{bmatrix} \mathbf{1}_{n_i-1} \\ 1 \end{bmatrix}$$

$$= \mathbf{1}_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \cdot \mathbf{1}_{n_i-1}^\top - \frac{1}{n_i} \mathbf{1}_{n_i-1}^\top$$

$$= \mathbf{1}_{n_i-1} - \frac{1}{n_i} (n_i-1) \mathbf{1}_{n_i-1}^\top - \frac{1}{n_i} \mathbf{1}_{n_i-1}^\top \quad \begin{bmatrix} \mathbf{J}_K \mathbf{1}_K \\ = \mathbf{1}_K \mathbf{1}_K^\top \mathbf{1}_K \\ = K \left(\frac{1}{K}\right) \end{bmatrix}$$

$$= 0$$

$$\text{So, } \tilde{v}_i \sim N_{n_i-1}\left(0, \sigma^2 (I_{n_i-1} + J_{n_i-1})^{-1}\right)$$

So, the density of \tilde{v}_i is given by,

$$f_{\tilde{v}_i}(\tilde{v}_i | \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^{n_i-1}} |I_{n_i-1} + J_{n_i-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \tilde{v}_i^\top [I_{n_i-1} + J_{n_i-1}] \tilde{v}_i\right)$$

[where $|A|$ is the determinant of matrix A]

$$\text{Also, } |I_{n_i-1} + J_{n_i-1}| = n_i$$

This is followed by the theorem below:

Theorem: $|aI_K + bJ_K| = a^{K-1}(a+b)$

Proof: Prove that $aI_K + bJ_K$ has eigen values

a with geometric multiplicity $(K-1)$ and $a+b$.

So,

$$f_{V_i}(v_i | \sigma^2) = (2\pi)^{\frac{n_i-1}{2}} \sigma^{-(n_i-1)} \frac{1}{n_i} v_i^T \exp\left(-\frac{1}{2\sigma^2} v_i^T (I_{n_i-1} + J_{n_i-1}) v_i\right)$$

As, v_i are independent across i , so, as the v_i 's.

So, the likelihood function based v_1, \dots, v_K is given by,

$$L(\sigma^2 | v_1, \dots, v_K) = (\sigma^2)^{-\sum_{i=1}^K (n_i-1)/2} \left(\frac{\pi}{n_i} \right)^{\sum_{i=1}^K (n_i-1)/2} (2\pi)^{\sum_{i=1}^K (n_i-1)/2}$$

$$\exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^K v_i^T (I_{n_i-1} + J_{n_i-1}) v_i\right)$$

$$v_i^T \mathbf{1}_{n_i-1} = \mathbf{1}_n^T \begin{pmatrix} I_{n_i-1} - \frac{1}{n_i} J_{n_i-1} \\ -\frac{1}{n_i} \mathbf{1}_{n_i-1}^T \end{pmatrix} \mathbf{1}_{n_i-1}$$

$$= \mathbf{1}_n^T \begin{pmatrix} +1/n_i \mathbf{1}_{n_i-1} \\ +\frac{1}{n_i} - \cancel{1} = 1 \end{pmatrix} = \left(\sum_{j=1}^{n_i} Y_{ij} \right) / n_i - \bar{Y}_{in_i}$$

$$= -(\bar{Y}_{in_i} - \bar{Y}_i)$$

$$\text{So, } \tilde{v}_i^T J_{n_i-1} \tilde{v}_i = \tilde{v}_i^T \mathbf{1}_{n_i-1} \mathbf{1}_{n_i-1}^T \tilde{v}_i \\ = (y_{in_i} - \bar{y}_i)^2$$

$$\text{and } \tilde{v}_i^T \tilde{v}_i = \sum_{j=1}^{n_i-1} (y_{ij} - \bar{y}_i)^2$$

$$\text{So, } \tilde{v}_i^T (I_{n_i-1} + J_{n_i-1}) \tilde{v}_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

So, the log-likelihood fn. based on v_1, \dots, v_K is

$$l(\sigma^2 | v_1, \dots, v_K) = \left(-\frac{1}{2} \sum_{i=1}^K (n_i - 1) \right) \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + c.$$

Differentiating w.r.t σ^2 and setting to 0 gives,

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2\sigma^2} (N-K) + \frac{1}{2\sigma^4} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{N-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\boxed{\hat{\sigma}^2 = \frac{\text{SSE}}{N-K}}$$

$$\text{and } \frac{\partial^2 l}{\partial \sigma^4} \Bigg|_{\hat{\sigma}^2 = \hat{\sigma}^2} = +\frac{N-K}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \text{ SSE} \\ = -\frac{(N-K)}{2\hat{\sigma}^4} < 0.$$

So, the MLE of σ^2 is now,

$$\hat{\sigma}^2 = \frac{SSE}{N-k}$$

c) Differentiating the profile log-likelihood w.r.t σ^2
we get, after setting to zero,

$$\frac{\partial}{\partial \sigma^2} \ell_p(\sigma^2) = -\frac{N}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} GSSSE(\theta) = 0$$

$$\Rightarrow \boxed{\hat{\sigma}_G^2 = \frac{1}{N} GSSSE(\theta)}.$$

$$\frac{\partial^2}{\partial \sigma^2} \ell_p(\sigma^2) \Big|_{\hat{\sigma}^2 = \hat{\sigma}_G^2} = -\frac{N}{2\hat{\sigma}_G^2} < 0.$$

So, the MLE of σ^2 is $\hat{\sigma}_G^2 = \frac{1}{N} GSSSE(\theta)$

d) Similarly as above, differentiating REML log-likelihood
and equating to zero we get,

$$\frac{\partial}{\partial \sigma^2} \ell_R(\sigma^2) = -\frac{(N-p)}{2\sigma^2} + \frac{1}{2\sigma^4} GSSSE(\theta) = 0$$

$$\Rightarrow \boxed{\hat{\sigma}_R^2 = \frac{1}{N-p} GSSSE(\theta)}$$

$$\text{with } \frac{\partial^2}{\partial \sigma^2} \ell_R(\sigma^2) \Big|_{\hat{\sigma}^2 = \hat{\sigma}_R^2} = -\frac{(N-p)}{2\hat{\sigma}_R^2} < 0.$$

Problem 5:

2.24.

$$Y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma^2), \quad j=1,2$$

$$\text{So, } \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N_2 \left(\mu_i \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \sigma^2 I_2 \right).$$

$$T_i = Y_{i1} + Y_{i2}$$

$$\text{So, } \begin{pmatrix} Y_{i1} \\ T_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = BY_i \sim$$

By Property of Normal distribution,

$$\begin{pmatrix} Y_{i1} \\ T_i \end{pmatrix} \sim N_2 \left(\mu_i \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \sigma^2 BB^T \right)$$

$$BB^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad BB^T = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\text{So, } \begin{pmatrix} Y_{i1} \\ T_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_i \\ 2\mu_i \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

By given result,

$$Y_{i1} | T_i = t_i \sim N(\mu_i^*, \sigma^{*2})$$

$$\text{where } \mu_i^* = \mu_i + \frac{\sigma^2}{2\sigma^2} (t_i - 2\mu_i) = t_i$$

$$\sigma^{*2} = \sigma^2 - \frac{\sigma^2}{2\sigma^2} \sigma^2 = \sigma^2/2.$$

$$\text{So, } Y_{i1} | T_i = t_i \sim N\left(\frac{t_i}{2}, \frac{\sigma^2}{2}\right).$$

QED.

