Assignment 2 Solution

Ted Ladas -s2124289

Question 1

$F(y;\theta) = 1 - e^{-y^2/(2\theta)}$ and $y \geq 0, \theta \geq 0$ and \

$X_i = \begin{cases} Y_i & Y_i \leq C \\ C & Y_i > C \end{cases}$

$R_i = \begin{cases} 1 & Y_i \leq C \\ 0 & Y_i > C \end{cases}$

a) Calculation of $\hat{\theta}_{MLE}$

We need to calculate the pdf of the data given theta, the survival function of the upper censor point, and the Likelihood of the data given theta. In order to reach the conclusion of the exercise I will define first these.

$X_i = Y_i R_i + C(1-R_i) \iff X_i^2 = Y_i^2 R_i^2 + 2Y_i R_i C(1-R_i) + C^2(1-R_i)^2$

We know that $\ R_i = 0 \ \ \vee \ \ R_i = 1 \ $ therefore $\ 2Y_i R_i C(1-R_i) = 0 =$

If $\ R_i = 0 \ $ then $\ R_i^2 = R_i \ \ and \ \ (1-R_i)^2 = (1-R_i)$

If $\ R_i = 1 \ $ then $\ R_i^2 = R_i \ \ and \ \ (1-R_i)^2 = (1-R_i)$

Therefore $\ R_i^2 = R_i \ \ and \ \ (1-R_i)^2 = (1-R_i)$

So, Finally: $\ X_i^2 = Y_i^2 R_i + C^2(1-R_i)$

We also have to have an expression for the Survival function:

$S(C;\theta) = 1 - F(y;\theta) \iff S(C;\theta) = e^{-y^2/(2\theta)}$

And finally we need to derive the pdf of the data

$f(y;\theta) = \frac{\partial}{\partial y} F(y;\theta) \iff f(y;\theta) = \frac{1}{\theta} y e^{-y^2/(2\theta)}$

So we can start:

$L(\theta) = \prod_i \{(f(y;\theta))^{r_i} (S(C;\theta))^{(1-r_i)}\} = \prod_i \{(\frac{1}{\theta} y e^{-y^2/(2\theta)})^{r_i} (e^{-C^2/(2\theta)})^{(1-r_i)}\} =$

$\theta^{-\sum r_i} \sum y_i^{r_i} e^{-\sum y_i^2 r_i + C^2(1-r_i)} = \theta^{-\sum r_i} \sum y_i^{r_i} e^{-\sum X_i^2}$

$\ln L(\theta) = \ell(\theta) = -\sum r_i \ln\theta - \frac{1}{2\theta} \sum X_i^2 + \ln \sum y_i^{r_i}$

$\frac{\partial}{\partial \theta} \ell(\theta) = 0 \iff -\frac{\sum r_i}{\theta} + \frac{\sum X_i^2}{2\theta^2} + 0 = 0 \iff \frac{-2\theta \sum r_i + \sum X_i^2}{2\theta^2} = 0$

and since $\theta \geq 0$

$-2\theta \sum r_i + \sum X_i^2 = 0 \iff \hat{\theta_{MLE}} = \frac{\sum X_i^2}{2 \sum r_i}$

b) We know that $I(\theta) = -\mathbb{E}\left[\frac{d^2 l(\theta)}{d\theta^2}\right]$

and also that $\mathbb{E}[R_i] = 1P(R_i = 1) + 0P(R_i = 0) = P(Y \leq C) = F(C;\theta) = 1 - e^{-C^2/(2\theta)}$ and therefore $\mathbb{E}[1 - R_i] = 1 - \mathbb{E}[R_i] = e^{-C^2/(2\theta)}$

so if we decompose $\sum X_i^2$ again we have:

$I(\theta) = -\frac{\sum \mathbb{E}[r_i]}{\theta^2} + \frac{\sum \mathbb{E}[Y_i^2 r_i]}{\theta^3} + \frac{\sum \mathbb{E}[C^2(1-r_i)]}{\theta^3} =$

$n/\theta^2(1 - e^{-C^2/(2\theta)})n/\theta^3(-C^2 e^{-C2/2\theta} + 2\theta(1 - e^{-C^2/(2\theta)}) - e^{-C^2/(2\theta)}))$

$I(\theta) = \frac{n}{\theta^2}(1 - e^{-C^2/(2\theta)})$

c) Since we know that $\hat{\theta_{MLE}} \sim N(\theta, I(\theta)^{-1})$ we can Normalize $\hat{\theta_{MLE}}$ and produce the 95% confidence interval as follows

$Z = \frac{\hat{\theta_{MLE}} - \theta}{\sqrt{I(\theta)^{-1}}} \sim N(0,1)$

$P(z_{a/2} \leq Z \leq z_{a/2}) = 1 - a$

$a = 0.05$ hence $z_{-a/2} = -1.959964, z_{a/2} = 1.959964$

$[z_{-a/2}\sqrt{I(\theta)-1}, z_{a/2}\sqrt{I(\theta)-1}]$

Question 2

a)

Like question 1a our likelihood is the product of the censored and the noncencored data. Since we know that the Normal distribution belongs to the exponential family, we could recompute the likelihood and prove the statement. However, this involves a lot of calculations, mostly to calculate the CDF of a normal distribution and to then manipulate it by multiplying it with the PDF in order to derive a distribution of X_i. Fortunately, there is an easier way to show this result.

We know that $Y_i \sim N(\mu, \sigma^2)$ If we can show that $X_i \sim N(\mu, \sigma^2)$ then the result is trivial.

$X_i = YiRi + D(1 - R_i)$

Since our X is an affine transformation of a normal, we can write it as $X = g(Y)$ where g linear. Together with the fact that R_i have a binary nature of $0 \vee 1$ we have our result.

So. $X_i \sim N(\mu, \sigma^2)$

Hence: $\log L(\mu, \sigma^2 | x, r) = \log \prod[\phi(x_i; \mu, \sigma^2)^{r_i} \times \Phi(x_i; \mu, \sigma^2)] = \sum_{i=1}^{n}\{r_i \log \phi(x_i, \mu, \sigma^2) + (1 - r_i)\Phi(x_1; \mu, \sigma^2)\}$

b) –> $\mu = 5.559766$

```
# https://cran.r-project.org/web/packages/maxLik/maxLik.pdf
load(file = "dataex2.Rdata")

ll <- function(m, df) {
    mu = m
    sigma = sqrt(1.5)
    x_i = df[, 1]
    r_i = df[, 2]
    likelihood_f = sum(r_i * dnorm(x_i, mean = mu, sd = sigma, log = TRUE) + (1 -
        r_i) * pnorm(x_i, mean = mu, sd = sigma, log = TRUE))  #log=TRUE: to avoid underflow, instead o
    # Nice to have: modify the optimization newton-raphson function we wrote on
    # statistical programming to optimize this problem as well.
    return(likelihood_f)
}
```

```
start = c(mu = 1)
mle = maxLik(logLik = ll, df = dataex2, start = start, method = "nr")  # Newton-Raphson method is the f
summary(mle)
```

```
## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 4 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -339.7295
## 1  free parameters
## Estimates:
##    Estimate Std. error t value Pr(> t)
## mu  5.55977    0.08761   63.46  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

Question 3

Based on the lecture 'Likelihood based inference with incomplete data' we know that a missing data mechanism is ignorable for likelihood inference iff:

1) The missing data are MAR (or MCAR)
2) the parameter ψ which is the parameter of the missing mechanism and the parameters of the model θ are distinct.

On this examples we know that point true is true because it is stated as such in all the sub-questions. So we only have to consider the first point to answer the question.

What we are doing on each example is the following We express the probability of a datapoint missing as a linear model of various other variables and parameters. The logit() function serves the purpose of transforming the support of our probability $R_f \in [0, 1]$ to $R_{logitf} \in (-\infty, +\infty)$ so that our line makes sense.

(a) Missing data mechanism –> MAR. This is true because the missingness depends on Y_1 which is fully observed. Therefore the mechanism is ignorable for likelihood-based estimation

(b) Missing data mechanism –> MNAR This is true because the missingness depends on Y_2 which is not fully observed. Therefore the mechanism is **not** ignorable for likelihood-based estimation

(c) Missing data mechanism –> MAR. This is true because the missingness depends on Y_1 and it's mean μ_1 which are fully observed. Therefore the mechanism is ignorable for likelihood-based estimation

Question 4

```
load("dataex4.Rdata")

# modifying dataset to be easier to work with
df4 <- dataex4 %>% arrange(Y) %>% mutate(R = (Y == 0 | Y == 1) * 1) %>% replace_na(list(R = 0)) %>%
    replace_na(list(Y = 2))
```

```r
# helper function to evaluate the sigmoid easier
sigmoid <- function(bb, xx) {
    b0 <- bb[1]
    b1 <- bb[2]
    sig = exp(b0 + b1 * xx)/(1 + exp(b0 + b1 * xx))
    return(sig)
}

# E step
q_fun <- function(bb, df) {
    b0 <- bb[1]
    b1 <- bb[2]
    xx <- df[, 1]
    yy <- df[, 2]
    rr <- df[, 3]
    # bb_vector = c(b0 + b1*xx) q <- -sum(yy*rr*log(1 + exp(-b0-b1*xx)) +
    # (1-rr)*log(1+exp(b0+b1*xx)) + sigmoid(bb_minus, xx)*log(1 + exp(-b0-b1*xx)) +
    # sigmoid(bb_plus, xx)*log(1+exp(b0+b1*xx)))
    q <- sum((rr * yy * (b0 + b1 * xx)) - log(1 + exp(b0 + b1 * xx)))  #+ (1 - rr)*(b0 + b1*xx)*sigmoid
    return(q)
}

# M step
epsilon = 1e-08  # tolerance set to machine precesion
beta_old = c(1, 1)

repeat {
    beta_new = coef(maxLik(logLik = q_fun, df = df4, start = beta_old))
    convergence = max(abs(beta_old - beta_new)) < epsilon
    print(convergence)
    if (convergence) {
        break
    }

    beta_old = beta_new
}

print(beta_new)
```

```
## [1] FALSE
## [1] FALSE
## [1] TRUE
## [1]  0.8196668 -3.5546004
```