S721 HW6

John Koo

From text

7.1

We just need to find the maximum value of $L(\theta \mid x) = f(x \mid \theta)$ for each x. Then we get

x	θ
0	1
1	1
2	2 or 3
3	3
4	3

7.2

Part a

$L(\beta \mid x) = f(x \mid \beta) = \prod_i^n \frac{1}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-x_i/\beta} = \Gamma(\alpha)^{-n} \beta^{-n\alpha} (\prod_i^n x_i)^{\alpha-1} e^{\sum_i^n x_i/\beta}$

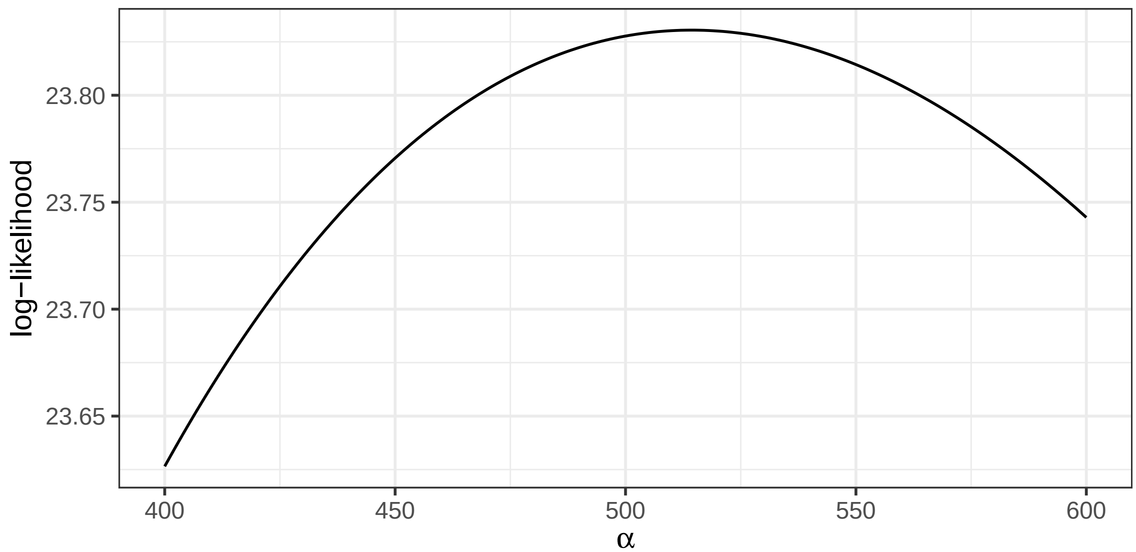
Then $\ell(\beta \mid x) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \log \prod_i^n x_i - \frac{1}{\beta} \sum_i^n x_i$, and so $\frac{\partial \ell}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_i^n x_i$. If we set this to 0, then $0 = -\beta n\alpha + \sum_i^n x_i \implies \hat{\beta} = \frac{1}{n\alpha} \sum_i^n x_i = \frac{\bar{X}}{\alpha}$.

Part b

If we plug in $\hat{\beta} = \frac{\bar{X}}{\alpha}$ into ℓ, we get $\ell = -n \log \Gamma(\alpha) - n\alpha \log \bar{X} + n\alpha \log \alpha + (\alpha - 1) \sum_i \log x_i - n\alpha$, and if we only look at the parts that depend on α, our expression turns into $-n \log \Gamma(\alpha) - n\alpha \log \bar{X} + n\alpha \log \alpha + \alpha \sum_i \log x_i - n\alpha$

```r
library(ggplot2)
theme_set(theme_bw())

x <- c(22, 23.9, 20.9, 23.8, 25, 24, 21.7,
       23.8, 22.8, 23.1, 23.1, 23.5, 23, 23)

n <- length(x)
x.mean <- mean(x)
log.x.sum <- sum(log(x))

alpha <- seq(400, 600)
log.likelihood <- function(alpha) {
  -n * lgamma(alpha) -
    n * alpha * log(x.mean) +
    n * alpha * log(alpha) +
    alpha * log.x.sum -
    n * alpha
}
```

```r
ggplot() +
  geom_line(aes(x = alpha, y = log.likelihood(alpha))) +
  labs(x = expression(alpha), y = 'log-likelihood')
```




```r
alpha.hat <- optimize(log.likelihood, c(400, 600), maximum = TRUE)$maximum
beta.hat <- x.mean / alpha.hat

c(alpha = alpha.hat, beta = beta.hat)
```

```
       alpha         beta
514.33564192   0.04494008
```

7.4

If $X_i \sim \mathcal{N}(\theta, 1)$, then $f(x_i \mid \theta) = \sqrt{\frac{1}{2\pi}} \exp(-\frac{(x_i-\theta)^2}{2})$, and $L(\theta \mid x) = \prod_i^n f(x_i \mid \theta)$. Then $\ell(\theta \mid x) = -\frac{n}{2} \log 2\pi - \sum_i^n \frac{(x_i-\theta)^2}{2}$, so $\partial_\theta \ell = \sum_i^n (x_i - \theta) = -n\theta + \sum_i^n x_i = n(\bar{x} - \theta)$. Then note that if $\bar{x} < \theta$, this is < 0, so the likelihood is decreasing. Therefore, if $\bar{x}$ is outside the domain of θ, then $\hat{\theta} = \min \Theta = 0$.

7.6

Part b

$L(\theta \mid x) = \prod_i^n \theta x_i^{-2} = \theta^n \prod_i^n x_i^{-2}$. Then $\partial_\theta L = n\theta^{n-1} \prod_i^n x_i^{-2}$, and if we set this to 0, $\hat{\theta} = 0$. However, we have that $\theta > 0$, so we cannot use this solution. (We can also show that $\hat{\theta} = 0$ is a minimum, not a maximum.)

Since each x_i is positive, we can see that $\partial_\theta L > 0$, so L is increasing in θ. Therefore, choosing the largest possible value of θ maximizes L. Since $\theta \leq x_i$ for each x_i, we can say $\hat{\theta} = x_{(1)}$.

Part c

$E[X \mid \theta] = \int_\theta^\infty \theta x^{-1} dx = \theta \log x|_\theta^\infty = \infty$. Therefore, $\hat{\theta}_{MOM}$ does not exist.

7.7

$L(0 \mid x) = 1$ and $L(1 \mid x) = 2^{-n} \prod x_i^{-1/2}$.

We can also say $\ell(0 \mid x) = 0$ and $\ell(1 \mid x) = \sum_i^n \log \frac{1}{2\sqrt{x_i}} = -n \log 2 - \frac{1}{2} \sum_i^n \log x_i$. So if this is greater than $\ell(1 \mid x) = 0$, then $\hat{\theta} = 1$, otherwise $\hat{\theta} = 0$. If we manipulate $-n \log 2 - \frac{1}{2} \sum_i^n \log x_i > 0$, then we get $\frac{1}{n} \log x_i < -2 \log 2$. So if the sample mean of the logs is less than $-2 \log 2 \approx -1.386$, then $\hat{\theta} = 1$, and otherwise $\hat{\theta} = 0$.

7.8

Part a

$\sigma^2 = E[X^2] - \mu^2 = E[X^2]$. Since we have a sample size of 1, $\hat{\sigma}^2 = X^2$.

Part b

From before, we have $\ell(\sigma \mid x) = -\frac{1}{2} \log 2\pi - \log \sigma - \frac{x^2}{2} \sigma^{-2}$. Then $\partial_\sigma \ell = -\frac{1}{\sigma} + x^2 \sigma^{-3}$. Setting this to 0, we get $\hat{\sigma} = |X|$.

Part c

From part (a), $\hat{\mu}_2 = \bar{X^2}$ and since $\mu_1 = 0$, we just set $\hat{\sigma}^2 = \hat{\mu}_2 \implies \hat{\sigma} = \sqrt{\bar{X^2}}$, and since we just have a sample size of one, $\hat{\sigma} = |X|$.

7.9

From class, we saw $\hat{\theta}_{MOM} = 2\bar{X}$. Then
$E[\hat{\theta}_{MOM}] = E[2\bar{X}] = \frac{2}{n} \sum_i^n E[X_i] = \frac{2}{n} \frac{n\theta}{2} = \theta$ (since the mean of each X_i is the halfway point between the min and max).

$Var(\hat{\theta}_{MOM}) = Var(2\bar{X}) = 4Var(\bar{X}) = \frac{4}{n^2} \sum_i^n Var(X_i)$.
$Var(X_i) = \int_0^\theta \frac{(x-\theta/2)^2}{\theta} dx = \frac{1}{\theta}(\frac{x^3}{3} - \frac{\theta x^2}{2} + \frac{\theta^2 x}{4})|_0^\theta = \frac{\theta^2}{12}$.
Then $Var(\hat{\theta}_{MOM}) = \frac{4}{n^2} \frac{n\theta^2}{12} = \frac{\theta^2}{3n}$.

From class, we saw $\hat{\theta}_{MLE} = X_{(n)}$, which has pdf $\frac{nx^{n-1}}{\theta^n}$ (from M463 notes).

$E[\hat{\theta}_{MLE}] = E[X_{(n)}] = \int_0^\theta \frac{nx^n}{\theta^n} dx = \frac{1}{\theta} \frac{n}{n+1} x^{n+1}|_0^\theta = \frac{n}{n+1}\theta$.

To get the variance, first we note that $E[X_{(n)}^2] = \int_0^\theta \frac{nx^{n+1}}{\theta^n} dx = \frac{n}{n+2}\theta^2$. Then $Var(X_{(n)}) = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2$ $= \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$

The MOM estimator is unbiased, but the MLE estimator becomes less and less biased as we increase n.

The variance for the MLE estimator decays faster than the variance for the MOM estimator. When $n = 1$, the MOM estimator has a variance of $\theta^2/3$ while the MLE estimator has a variance of $\theta^2/12$, so the MLE estimator will always have less variance.

7.10

Part b

Note that $F(x_i \mid \alpha, \beta) = (\frac{x}{\beta})^\alpha$ for $x \in [0, \beta]$, so $f(x_i \mid \alpha, \beta) = F'(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1}$ when $x \in [0, \beta]$.

$L(\alpha, \beta \mid x) = \prod_i^n \frac{\alpha}{\beta^\alpha} x_i^{\alpha-1} = \alpha^n \beta^{-n\alpha} (\prod_i^n x_i)^{\alpha-1}$.

Then $\ell(\alpha, \beta \mid x) = n \log \alpha - n\alpha \log \beta + (\alpha - 1) \sum_i \log x_i \implies \nabla \ell = \begin{bmatrix} \frac{n}{\alpha} - n \log \beta + \sum_i \log x_i \\ -\frac{n\alpha}{\beta} \end{bmatrix}$

If we look at the part for β, note that $-\frac{n\alpha}{\beta}$ never reaches 0 and it is always negative. Then ℓ is always decreasing in β, so we just set it to the lowest possible value. Since $\beta \geq x_i$ for all x_i, $\hat{\beta} = X_{(n)}$.

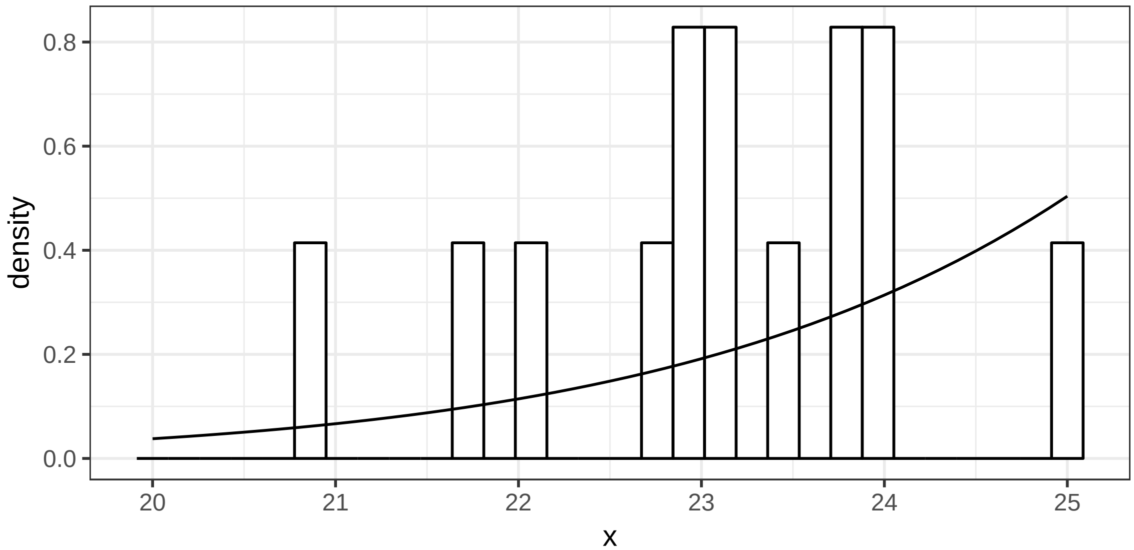
Then if we look at α, we can solve $\frac{n}{\alpha} - n \log \hat{\beta} + \sum_i \log x_i = 0$ to obtain $\hat{\alpha} = \frac{n}{n \log \hat{\beta} - \sum_i \log x_i}$.

Part c

```r
x <- c(22, 23.9, 20.9, 23.8, 25, 24, 21.7,
       23.8, 22.8, 23.1, 23.1, 23.5, 23, 23)

n <- length(x)
beta.hat <- max(x)
alpha.hat <- n / (n * log(beta.hat) - sum(log(x)))

c('alpha' = alpha.hat, 'beta' = beta.hat)
```

```
   alpha     beta
12.59487 25.00000
```

```r
X <- seq(20, 25, by = .1)
p <- alpha.hat / beta.hat ^ alpha.hat * X ^ (alpha.hat - 1)

ggplot() +
  geom_histogram(aes(x = x, y = ..density..),
                 colour = 'black', fill = 'white') +
  geom_line(aes(x = X, y = p))
```



11.38

Part a

$Y_i \sim Poisson(\theta x_i)$, so $E[Y_i] = \theta x_i \implies y_i = \theta x_i + \epsilon_i$.
Then we want to minimize $\sum_i \epsilon_i^2 = \sum_i (y_i - \theta x_i)^2$ w.r.t. θ, which we can do by setting $\partial_\theta \sum_i (y_i - \theta x_i)^2 = 0$.
$\partial_\theta \sum_i (y_i - \theta x_i)^2 = -2 \sum_i x_i(y_i - \theta x_i) = 0 \implies \sum_i x_i y_i - \theta \sum_i x_i^2 = 0 \implies \hat{\theta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$

$Var(\hat{\theta}) = Var(\frac{\sum_i x_i Y_i}{\sum_i x_i^2}) = (\sum_i x_i^2)^{-2} Var(\sum_i x_i Y_i) = (\sum_i x_i^2)^{-2} \sum_i Var(x_i Y_i) = (\sum_i x_i^2)^{-2} \sum_i x_i^2 Var(Y_i)$
$= (\sum_i x_i^2)^{-2} \sum_i x_i^2 \theta x_i = \frac{\sum_i x_i^3}{(\sum_i x_i^2)^2} \theta$

$E[\hat{\theta}] = E[\frac{\sum_i x_i Y_i}{\sum_i x_i^2}] = (\sum_i x_i^2)^{-1} \sum_i x_i \theta x_i = \theta \frac{\sum_i x_i^2}{\sum_i x_i^2} = \theta \implies \hat{\theta}$ is unbiased

Part b

$L(\theta \mid x, y) = \prod_i \frac{e^{-\theta x_i} (\theta x_i)^{y_i}}{y_i!}$, then
$\ell(\theta \mid x, y) = -\theta \sum_i x_i + \sum_i y_i \log \theta x_i - \sum_i \log y_i!$, and
$\partial_\theta \ell = -\sum_i x_i + \frac{1}{\theta} \sum_i \frac{x_i y_i}{x_i} = -\sum_i x_i + \frac{1}{\theta} \sum_i y_i$.
If we set this to 0, we obtain $\hat{\theta} = \frac{\sum_i Y_i}{\sum_i x_i}$.

$Var(\hat{\theta}) = Var(\frac{\sum_i Y_i}{\sum_i x_i}) = (\sum_i x_i)^{-2} \sum_i Var(Y_i) = (\sum_i x_i)^{-2} \theta \sum_i x_i = \frac{\theta}{\sum_i x_i}$

$E[\hat{\theta}] = E[\frac{\sum_i Y_i}{\sum_i x_i}] = (\sum_i x_i)^{-1} \sum_i E[Y_i] = (\sum_i x_i)^{-1} \theta \sum_i x_i = \theta \implies \hat{\theta}$ is unbiased.

Not from text

Problem 1

From before, we saw that $f(x_i \mid \theta) = \theta x_i^{\theta-1}$. Then $L(\theta \mid x) = \theta^n \prod_i x_i^{\theta-1}$ and $\ell(\theta \mid x) = n \log \theta + (\theta - 1) \sum_i \log x_i$. Setting $\partial_\theta \ell = 0$, we get $n/\theta + \sum_i \log x_i = 0 \implies \hat{\theta}_{MLE} = -\frac{1}{n} \sum_i \log x_i$.

$E[X_i] = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1}$. Then we obtain $\hat{\theta}_{MOM}$ by solving $\frac{\theta}{\theta+1} = \bar{X}$ for θ, so $\hat{\theta}_{MOM} = \frac{\bar{X}}{1-\bar{X}}$.

Problem 2

Problem 3