MATH 557 - EXERCISES 1 : SOLUTIONS

1 (a) Suppose, for two points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in the parameter space, $c(\underset{\sim}{x}, \underset{\sim}{y})$ is a function of these two arguments. We have that

$$\begin{aligned}\frac{f_i(\underset{\sim}{x})}{f_i(\underset{\sim}{y})} = c(\underset{\sim}{x}, \underset{\sim}{y}),\ i = 0, \ldots, k \quad &\iff \quad \frac{f_i(\underset{\sim}{x})}{f_i(\underset{\sim}{y})} = \frac{f_0(\underset{\sim}{x})}{f_0(\underset{\sim}{y})},\ i = 1, \ldots, k \\ &\iff \quad \frac{f_i(\underset{\sim}{x})}{f_0(\underset{\sim}{x})} = \frac{f_i(\underset{\sim}{y})}{f_0(\underset{\sim}{y})},\ i = 1, \ldots, k \\ &\iff \quad T_i(\underset{\sim}{x}) = T_i(\underset{\sim}{y})\end{aligned}$$

for the given statistic

$$\underset{\sim}{T}(\underset{\sim}{X}) = (T_1(\underset{\sim}{X}), \ldots, T_k(\underset{\sim}{X}))^{\mathsf{T}} = (T_i(\underset{\sim}{X}))^{\mathsf{T}}_{i=1,\ldots,k}$$

Therefore, by the Theorem from lectures, $\underset{\sim}{T}(\underset{\sim}{X})$ is minimal sufficient. This result implies that for selecting a *model* from a group of models, likelihood ratios provide summary statistics.

(b) If $\underset{\sim}{T}'(\underset{\sim}{X})$ is sufficient for θ, then by the Neyman Factorization Theorem there exist g and h such that

$$f_\theta(\underset{\sim}{x}) = g\left(\underset{\sim}{T}'(\underset{\sim}{x}); \theta\right) h(\underset{\sim}{x}).$$

Thus, choosing θ_i, $i = 1, \ldots, k$, such that $f_i = f_{\theta_i}$,

$$\underset{\sim}{T}(\underset{\sim}{X}) = \left(\frac{f_i(\underset{\sim}{X})}{f_0(\underset{\sim}{X})}\right)^{\mathsf{T}}_{i=1,\ldots,k} = \left(\frac{g_i(\underset{\sim}{T}'(\underset{\sim}{X}); \theta_i) h(\underset{\sim}{X})}{g_0(\underset{\sim}{T}'(\underset{\sim}{X}); \theta_0) h(\underset{\sim}{X})}\right)^{\mathsf{T}}_{i=1,\ldots,k} = \left(\frac{g_i(\underset{\sim}{T}'(\underset{\sim}{X}); \theta_i)}{g_0(\underset{\sim}{T}'(\underset{\sim}{X}); \theta_0)}\right)^{\mathsf{T}}_{i=1,\ldots,k}$$

Hence for any sufficient $\underset{\sim}{T}'$, $\underset{\sim}{T}$ is a function of $\underset{\sim}{T}'$. Therefore $\underset{\sim}{T}$ is minimal sufficient.

(c) Let T be minimal sufficient for θ, so that if W is also sufficient for θ, there exists h such that $h(W) = T$. Let $T' = r(T)$ where r is a 1-1 mapping. Then $T' = r(h(W))$, so T' is a function of every sufficient statistic.

(d) In the notation of the earlier parts, let $\mathcal{F}_1 = \{f_\beta : \beta = 1, 2\}$ and $\mathcal{F} = \{f_\beta : \beta > 0\}$, where f_β is the Exponential density with expectation β, so that

$$f_\beta(x) = \frac{1}{\beta} e^{-x/\beta} I_{(0,\infty)}(x).$$

If $X_{(1)} = \min\{X_1, \ldots, X_n\}$, define

$$T(\underset{\sim}{X}) = \frac{f_2(\underset{\sim}{X})}{f_1(\underset{\sim}{X})} = \frac{\frac{1}{2^n} e^{-\frac{1}{2}\sum_{i=1}^n X_i} I_{(0,\infty)}(X_{(1)})}{e^{-\sum_{i=1}^n X_i} I_{(0,\infty)}(X_{(1)})} = 2^{-n} \exp\left\{\frac{n\overline{X}}{2}\right\}.$$

However, since

$$f_\beta(\underset{\sim}{x}) = \frac{1}{\beta^n} e^{-n\overline{x}/\beta}\, I_{(0,\infty)}(x_{(1)}) = \frac{1}{(2\beta)^n} \frac{1}{T(\underset{\sim}{x})^{2/\beta}}\, I_{(0,\infty)}(x_{(1)})$$

$T(\underset{\sim}{X})$ is sufficient for β by the Neyman factorization theorem. Therefore $T(\underset{\sim}{X})$ is minimal sufficient for β by part (b). Since $T(\cdot)$ is a 1-1 transformation, $\overline{X}$ is minimal sufficient for β by part (c).

2 (a) The joint pmf takes the form of an n-sample multinomial without the combinatorial term (here we observe the X_is individually, not merely the totals in each of the categories):

$$f_{\underset{\sim}{X}|\underset{\sim}{\theta}}(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{j=1}^{3} \theta_j^{\sum_{i=1}^{n} I_{\{j\}}(x_i)} = \prod_{j=1}^{3} \theta_j^{n_j}$$

where

$$n_j = \sum_{i=1}^{n} I_{\{j\}}(x_i)$$

counts the number of times X_i takes the value j, for $j = 1, 2, 3$. But $n_3 = n - n_1 - n_2$, hence

$$f_{\underset{\sim}{X}|\underset{\sim}{\theta}}(\underset{\sim}{x}|\underset{\sim}{\theta}) = \theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n-n_1-n_2}$$

Thus, by the Neyman Factorization Theorem, $\underset{\sim}{T} = (N_1, N_2)^{\mathsf{T}}$ is sufficient for $\underset{\sim}{\theta} = (\theta_1, \theta_2)^{\mathsf{T}}$.

(b) When looking at N_1 and N_2, we have the traditional multinomial pmf, so

$$f_{\underset{\sim}{T}(\underset{\sim}{X})|\underset{\sim}{\theta}}(\underset{\sim}{T}(\underset{\sim}{x})|\underset{\sim}{\theta}) = \frac{n!}{n_1!n_2!(n-n_1-n_2)!}\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)_3^n$$

where $\underset{\sim}{T}(\underset{\sim}{x}) = (n_1, n_2)^{\mathsf{T}}$ and $n_3 = n - n_1 - n_2$.

(c) Using the using calculus approach, we form the log likelihood, partially differentiate in turn with respect to θ_1 and θ_2, equate to zero, and then solve the resulting two equations simultaneously. We have

$$l(\theta_1, \theta_2|\underset{\sim}{x}) = c(\underset{\sim}{n}) + n_1 \log\theta_1 + n_2 \log\theta_2 + n_3 \log(1-\theta_1-\theta_2)$$

so that

$$\begin{aligned} \frac{\partial l}{\partial \theta_1} &= \frac{n_1}{\theta_1} - \frac{n_3}{1-\theta_1-\theta_2} \\ \frac{\partial l}{\partial \theta_2} &= \frac{n_2}{\theta_2} - \frac{n_3}{1-\theta_1-\theta_2} \end{aligned}$$

Equating to zero and subtracting the second from the first equation, we obtain that

$$\frac{n_1}{\theta_1} = \frac{n_2}{\theta_2} \qquad \therefore \qquad \frac{\theta_1}{\theta_2} = \frac{n_1}{n_2}.$$

Substituting back into the first equation, we have

$$\frac{n_1}{n_3} = \frac{\theta_1}{1-\theta_1-\theta_2} = \frac{\theta_1}{1-\theta_1-n_2\theta_1/n_1}$$

Cross multiplying, we get

$$n_1 - n_1\theta_1 - n_2\theta_1 = n_3\theta_1$$

and hence

$$n_1 = (n_1+n_2+n_3)\theta_1 \qquad \therefore \qquad \theta_1 = \frac{n_1}{n_1+n_2+n_3} = \frac{n_1}{n}$$

Thus, by a similar argument for θ_2, we deduce

$$\widehat{\underset{\sim}{\theta}}(\underset{\sim}{x}) = (n_1/n, n_2/n)^{\mathsf{T}}.$$

and hence the estimator is

$$\widehat{\underset{\sim}{\theta}}(\underset{\sim}{X}) = (N_1/n, N_2/n)^{\mathsf{T}}.$$

Note that the case covered here presumes $n_1, n_2, n_3 > 0$. If this does not hold, the maximum likelihood estimator does not lie in the parameter space and is not defined.

3 (a) We have for the likelihood

$$L(\theta|\underset{\sim}{x}) = \prod_{i=1}^{n} \frac{1}{\pi} \frac{1}{1+(x_i-\theta)^2} = \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1+(x_i-\theta)^2}$$

so that

$$l(\theta|\underset{\sim}{x}) = -n \log \pi - \sum_{i=1}^{n} \log(1+(x_i-\theta)^2).$$

yielding the score equation

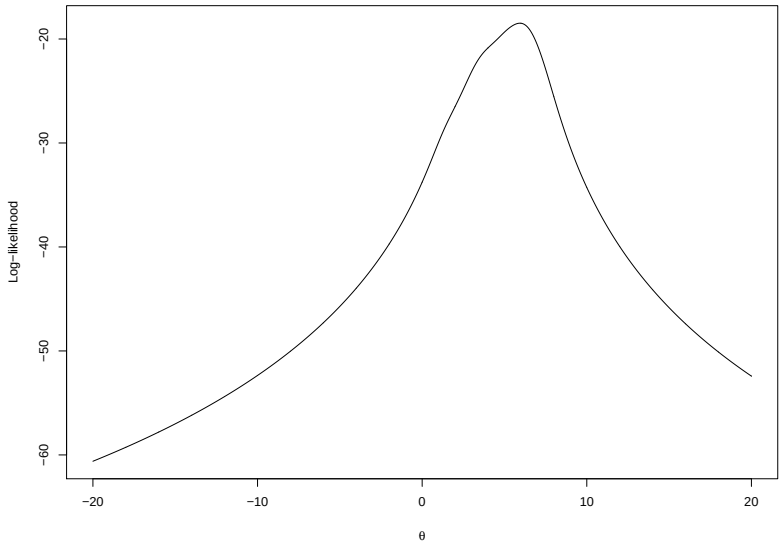
$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^{n} \frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = 0$$

which cannot be solved analytically.

(b) The code below evaluates the log-likelihood in `R` on a fine grid on the range (-20,20):

```
x.dat<-c(7.36,5.14,3.71,3.15,6.00,6.38,1.34,6.73)
thvec<-seq(-20,20,length=20001)
log.pdf.func<-function(x,th){return(-log(pi)-log(1+(x-th)^2))}
log.pdf.mat<-outer(thvec,x.dat,log.pdf.func)
log.like.vec<-apply(log.pdf.mat,1,sum)
plot(thvec,log.like.vec,type=''l'',xlab=expression(theta),ylab=''Log-likelihood'')
th.max<-thvec[which.max(log.like.vec)]
```

In this code, the variable `th.max` contains the numerical value of the estimate of θ; to two decimal places, we find that $\widehat{\theta}(\underset{\sim}{x}) = 5.94$. The code also produces the following plot. The



following code uses the `optimize` function to produce the same result: the commands

```
log.like.fx<-function(x,xd){return(sum(log.pdf.func(xd,x[1])))}
optimize(f=log.like.fx,xd=x.dat,interval=c(-20,20),maximum=TRUE)
```

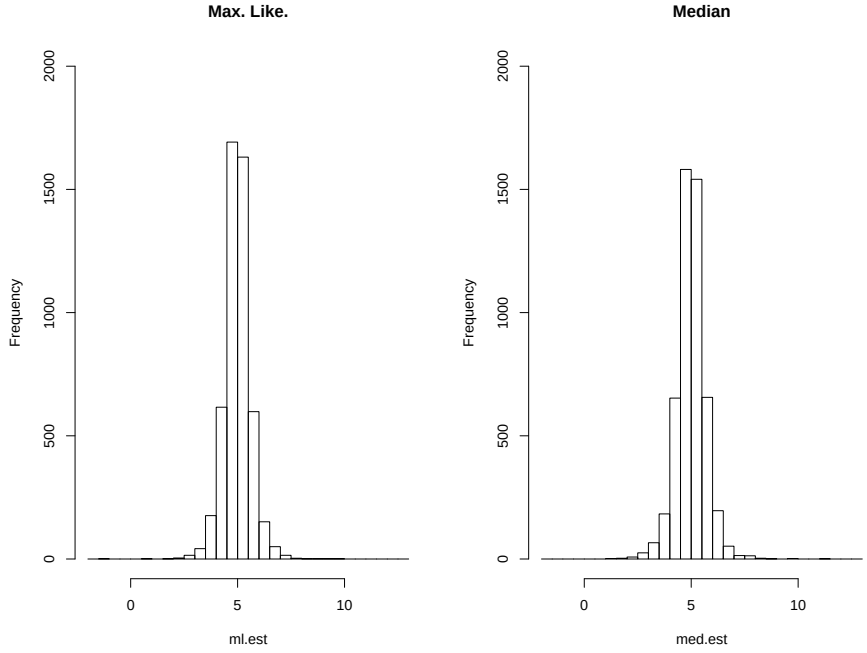
yield

```
$maximum [1] 5.93949
$objective [1] -18.48059
```

4 The following R code computes the $N = 5000$ estimates.

```
N<-5000
n<-8
ml.est<-med.est<-rep(0,N)
theta0<-5
for(i in 1:N){
    x<-rcauchy(n)+theta0
    ml<-optimize(f=log.like.fx,xd=x,interval=c(-20,20),maximum=TRUE)
    ml.est[i]<-ml$maximum
    med.est[i]<-median(x)
}
```

The histograms below depict the 5000 maximum likelihood and median estimates: It can be



discerned from these figures that the variance of the median estimator is slightly larger than for the maximum likelihood estimator; in one simulation, I got that the variance for the sample of median estimates was 0.468, and the variance for the sample of maximum likelihood estimates was 0.387.

4 The likelihood in the original parameterization is

$$L(\theta_1, \theta_2 | x_1, x_2) = \binom{n_1}{x_1} \theta_1^{x_1} (1-\theta_1)^{n_1 - x_1} \binom{n_2}{x_2} \theta_2^{x_2} (1-\theta_2)^{n_2 - x_2}$$

If $\phi = \theta_2/(1-\theta_2)$, then

$$\theta_2 = \frac{\phi}{1+\phi} \qquad \theta_1 = \frac{\phi\psi}{1+\phi\psi}$$

Either by writing out the likelihood in full in the new parameterization, and maximizing in the usual way, or by using invariance properties of maximum likelihood estimates, we conclude that

$$\widehat{\psi}(x_1, x_2) = \frac{\widehat{\theta}_1/(1-\widehat{\theta}_1)}{\widehat{\theta}_2/(1-\widehat{\theta}_2)} = \frac{(x_1/n_1)/(1-x_1/n_1)}{(x_2/n_2)/(1-x_2/n_2)} = \frac{x_1/(n_1-x_1)}{x_2/(n_2-x_2)}$$