STAT 305 Assignment 5

Name: Kevin Nguyen
Student Number: 97703045

Question 1

1. We have the log-likelihood

$$
\begin{aligned}
LL(\theta) &= \log\left[n! \prod_{r,c} \frac{\theta_{r,c}^{n_{r,c}}}{n_{r,c}!}\right] \\
&= \log(n!) + \log\left[\prod_{r,c} \frac{\theta_{r,c}^{n_{r,c}}}{n_{r,c}!}\right] \\
&= \log(n!) + \sum_{r,c} \log\left[\frac{\theta_{r,c}^{n_{r,c}}}{n_{r,c}!}\right] \\
&= \log(n!) + \left[\sum_{r,c} n_{r,c}\log(\theta_{r,c}) - \log(n_{r,c}!)\right]
\end{aligned}
$$

2. Under the null, $\theta_{r,c} = \phi_r \rho_c$

   i) finding the MLE, $\hat{\phi}_r$

$$
\begin{aligned}
\mathcal{L} &= \log(n!) + \left[\sum_{r,c} n_{r,c}\log(\phi_r\rho_c) - \log(n_{r,c}!)\right] \\
\text{constraint} : \sum_r \phi_r &= 1 \\
\frac{\partial}{\partial \phi_r} &= \frac{\partial}{\partial \phi_r}\left[\mathcal{L} - \lambda\left(\sum_r \phi_r - 1\right)\right] \\
&= \left[\sum_{r,c} \frac{\partial}{\partial \phi_r}(n_{r,c}\log(\phi_r\rho_c))\right] - \lambda \\
&= \sum_{r,c} \frac{n_{r,c}}{\phi_r} - \lambda = \frac{1}{\phi_r}\sum_{r,c} n_{r,c} - \lambda \\
&= \frac{n_{r.}}{\phi_r} - \lambda = 0 \\
\phi_r &= \frac{n_{r.}}{\lambda} \propto n_{r.} \\
\hat{\phi}_r &= \frac{n_{r.}}{\sum_r n_{r.}} = \frac{n_{r.}}{n} \quad \square \qquad \text{to make sure } \phi_r \text{ sums to 1}
\end{aligned}
$$

ii) finding the MLE, $\hat{\rho}_c$

$$\mathcal{L} = \log(n!) + \left[\sum_{r,c} n_{r,c} \log(\phi_r \rho_c) - \log(n_{r,c}!)\right]$$

$$\text{constraint} : \sum_{c} \rho_c = 1$$

$$\begin{aligned}
\frac{\partial}{\partial \rho_c} &= \frac{\partial}{\partial \rho_c}\left[\mathcal{L} - \lambda\left(\sum_r \phi_r - 1\right)\right] \\
&= \left[\sum_{r,c} \frac{\partial}{\partial \rho_c}(n_{r,c}\log(\phi_r \rho_c))\right] - \lambda \\
&= \sum_{r,c} \frac{n_{r,c}}{\rho_c} - \lambda = \frac{1}{\rho_c}\sum_{r,c} n_{r,c} - \lambda \\
&= \frac{n_{.c}}{\rho_c} - \lambda = 0 \\
\rho_c &= \frac{n_{.c}}{\lambda} \propto n_{.c} \\
\hat{\rho}_c &= \frac{n_{.c}}{\sum_c n_{.c}} = \frac{n_{.c}}{n} \quad \square \qquad \text{to make sure } \rho_c \text{ sums to 1}
\end{aligned}$$

3. Under the null, $\theta_{r,c}$ depends on ϕ_r and ρ_c, which have the constraint of $\sum_r \phi_r + \sum_c \rho_c = 1$. Thus, since there are 3 categories in r, there are 2 free variables for ϕ_r (ex. if you know ϕ_{red} and ϕ_{green}, ϕ_{blue} is 1 minus those 2 numbers). Similarly, since there are 2 categories in c, there are 1 free variables for ρ_c. Thus we have 3 free variable.

4. Under the alternative, the parameters are $\theta_{r,c}$ for $r = 1, 2, 3$ and $c = 1, 2$, it's obvious to see there's 6 combination of (r, c). We have the constraint that $\sum \theta_{r,c} = 1$ so similar to the question above, we only need the first 5 quantity to figure out the last value. Thus there are 5 free parameters.

5. We can obtain the Wilks statistic via the following R code based on the following formula.

$$W = 2\sum_{r,c} Y_{r,c} \ln\left(\frac{Y_{r,c}}{E_{r,c}}\right)$$

First, we need to compute all the expected value

$$\begin{aligned}
e_{11} &= \phi_1 \cdot \rho_1 \cdot n = \frac{n_{r.}}{n} \cdot \frac{n_{.c}}{n} \cdot n = \frac{50}{150} \cdot \frac{25}{150} \cdot 150 = \frac{25}{3} \\
e_{12} &= \phi_1 \cdot \rho_2 \cdot n = \frac{50}{150} \cdot \frac{125}{150} \cdot 150 = \frac{125}{3} \\
&\dots
\end{aligned}$$

Since $\phi_1 = \phi_2 = \phi_3 = 1/3$, expected value depends on either ρ_c and thus every cell in the Monday column has the same expected value and every cell in the Friday column.

In R,

```
observed <- c(3, 47, 5, 45, 17, 33)
# order is {(1,1), (1,2), (2,1), (2,2), (3,1), (3,2)}

a <- 1/18 * 150
# expected value for every cell in the Monday column

b <- 5/18 *150
# expected value for every cell in the Friday column

expected <- c(a, b, a, b, a, b)
wilks <- 2 * sum(observed * log(observed / expected))
print(wilks) # yields 15.85977
```

So we get a test statistic of 15.85977. From the results of the previous questions, the difference in free parameter between the null and alternative hypothesis is $5 - 3 = 2$. Thus $W \sim \chi^2_2$. The critical value of a 5% confidence level can be obtained via **qchisq**(0.95, 2). This gives a critical value of 5.991465. As you can see, our test statistics exceed the critical value, so we reject the null hypothesis. Alternatively, if we look at the p-value, we get 0.0007196556 (via R by 2∗**pchisq**(15,85977, 2, **lower**.tail = FALSE)). This means that we would reject for any confidence level above 0.007%.