STA355H1 - Assignment 3

1. (a) From the slides of lecture 8, M follows a hypergeometric distribution with parameters n_0, n_1, N and $\mathbb{P}(M = m; N) = \dfrac{\binom{n_0}{m}\binom{N-n_0}{n_1-m}}{\binom{N}{n_1}}$ for $0 \le m \le n_1$, using the convention that $\binom{n}{k} = 0$ if $n < k$. To determine $\mathbb{E}(M)$, first note that

$$m\binom{n_0}{m} = \frac{n_0!}{(m-1)!(n_0-m)!} = \frac{n_0(n_0-1)!}{(m-1)!((n_0-1)-(m-1))!} = n_0\binom{n_0-1}{m-1}.$$

Then,

$$\begin{aligned}\mathbb{E}_N(M) &= \sum_{m=0}^{n_1} m\mathbb{P}(M=m) = \sum_{m=1}^{n_1} m\frac{\binom{n_0}{m}\binom{N-n_0}{n_1-m}}{\binom{N}{n_1}} = \frac{n_0 n_1}{N}\sum_{m=1}^{n_1}\frac{\binom{n_0-1}{m-1}\binom{(N-1)-(n_0-1)}{(n_1-1)-(m-1)}}{\binom{N-1}{n_1-1}} \\ &= \frac{n_0 n_1}{N}\sum_{m'=0}^{n_1-1}\frac{\binom{n_0-1}{m'}\binom{(N-1)-(n_0-1)}{(n_1-1)-m'}}{\binom{N-1}{n_1-1}} = \frac{n_0 n_1}{N}\end{aligned}$$

where the last summation equals 1 since it is the sum of all probabilities of a hypergeometric distribution with parameters $n_0-1, n_1-1, N-1$. By the method of moments, $M = n_0 n_1/\hat{N} \implies \hat{N} = n_0 n_1/M$. Notice that this estimator is unstable for small positive values of M and undefined when $M = 0$.

(b) First note that $t_1 = n_0$. Applying the definition of conditional probability recursively yields

$$\begin{aligned}\mathbb{P}(M_1 = m_1, \ldots, M_k = m_k) &= \mathbb{P}(M_1 = m_1)\prod_{i=2}^{k}\mathbb{P}(M_i = m_i | M_1 = m_1, \ldots, M_{i-1} = m_{i-1}) \\ &= \frac{\binom{n_0}{m_1}\binom{N-n_0}{n_1-m_1}}{\binom{N}{n_1}}\prod_{i=2}^{k}\frac{\binom{t_i}{m_i}\binom{N-t_i}{n_i-m_i}}{\binom{N}{n_i}} = \prod_{i=1}^{k}\frac{\binom{t_i}{m_i}\binom{N-t_i}{n_i-m_i}}{\binom{N}{n_i}} \\ &= \mathcal{L}(N).\end{aligned}$$

where $N - t_k - n_k + m_k \ge 0$ or equivalently $N \ge \Sigma_{j=0}^{k}(n_j - m_j)$ due to $\binom{N-t_i}{n_i-m_i}$. To derive the maximum likelihood estimate, we determine the log-likelihood function:

$$\begin{aligned}\log\mathcal{L}(N) &= \sum_{i=1}^{k}\log\frac{t_i!(N-t_i)!n_i!(N-n_i)!}{m_i!(t_i-m_i)!(n_i-m_i)!(N-t_i-n_i+m_i)!N!} \\ &\propto \sum_{i=1}^{k}\log\frac{(N-t_i)!(N-n_i)!}{(N-t_i-n_i+m_i)!N!}\end{aligned}$$

which we maximize by taking its derivative with respect to N and setting it to 0. The solution for N is then the estimate $\hat{N}$, which must satisfy the constraints $\hat{N} \in \mathbb{N}$ and $\hat{N} \ge \Sigma_{j=0}^{k}(n_j - m_j)$ as before. Finally, we check if $(\log\mathcal{L}(\hat{N}))'' > 0$ to confirm if $\hat{N}$ is a true maximum.

(c) Adapting the likelihood function from the previous part,

$$\mathcal{L}(\omega) = \prod_{i=1}^{k}\frac{\binom{t_i}{m_i}\binom{1/\omega-t_i}{n_i-m_i}}{\binom{1/\omega}{n_i}} \approx \prod_{i=1}^{k}\frac{\exp(-n_i t_i\omega)(n_i t_i\omega)^{m_i}}{m_i!}$$

$$\ln\mathcal{L}(\omega) \approx \sum_{i=1}^{k}\ln\frac{\exp(-n_i t_i\omega)(n_i t_i\omega)^{m_i}}{m_i!} = \sum_{i=1}^{k}(m_i\ln(n_i t_i\omega) - n_i t_i\omega - \ln(m_i!))$$

$$(\ln\mathcal{L}(\omega))' = \sum_{i=1}^{k}(\frac{m_i}{n_i t_i\omega}n_i t_i - n_i t_i) = \sum_{i=1}^{k}(\frac{m_i}{\omega} - n_i t_i) = \sum_{i=1}^{k}\frac{m_i}{\omega} - \sum_{i=1}^{k}n_i t_i \overset{\text{set}}{=} 0$$

which yields $\hat{\omega} = (\sum_{i=1}^k m_i)/(\sum_{i=1}^k n_i t_i)$. Note that this is indeed a maximum since $(\ln \mathcal{L}(\hat{\omega}))'' = -(\Sigma_{i=1}^k m_i)/\hat{\omega}^2 < 0$. An estimator for N is then $\hat{N} = 1/\hat{\omega} = (\sum_{i=1}^k n_i t_i)/(\sum_{i=1}^k m_i)$.

(d) By the definition of the standard error estimator from the slides of lecture 10,

$$\widehat{\text{se}}(\hat{\omega}) = (-\frac{d^2}{d\omega^2} \ln \mathcal{L}(\hat{\omega}))^{-1/2} = (\frac{1}{\hat{\omega}^2} \sum_{i=1}^k m_i)^{-1/2} = \frac{(\sum_{i=1}^k m_i)^{1/2}}{\sum_{i=1}^k n_i t_i}.$$

(e) Since the Poisson distribution is part of the exponential family, we can assume that $\sqrt{nI(\omega)}(\hat{\omega} - \omega)$ approximately $\sim \mathcal{N}(0,1)$, or equivalently $(\hat{\omega} - \omega)/\widehat{\text{se}}(\hat{\omega})$ approximately $\sim \mathcal{N}(0,1)$, so we have

$$\begin{aligned} 0.95 &\approx \mathbb{P}(-1.96\widehat{\text{se}}(\hat{\omega}) \leq \hat{\omega} - \omega \leq 1.96\widehat{\text{se}}(\hat{\omega})) \\ &= \mathbb{P}(\hat{\omega} - 1.96\widehat{\text{se}}(\hat{\omega}) \leq \omega \leq \hat{\omega} + 1.96\widehat{\text{se}}(\hat{\omega})) \\ &= \mathbb{P}\Big(\frac{1}{\hat{\omega} + 1.96\widehat{\text{se}}(\hat{\omega})} \leq N \leq \frac{1}{\hat{\omega} - 1.96\widehat{\text{se}}(\hat{\omega})}\Big) \end{aligned}$$

which indicates that the 95% CIs for ω and N are $[\hat{\omega} - 1.96\widehat{\text{se}}(\hat{\omega}), \hat{\omega} + 1.96\widehat{\text{se}}(\hat{\omega})]$ and $[(\hat{\omega} + 1.96\widehat{\text{se}}(\hat{\omega}))^{-1}, (\hat{\omega} - 1.96\widehat{\text{se}}(\hat{\omega}))^{-1}]$ respectively. See the attached code, numerical results, and plot. Since $N \in \mathbb{N}$, we can round the results to get $\hat{N} = 451$ with a 95% CI of $[322, 752]$.

2. (a) Holding κ fixed,

$$\begin{aligned} \mathcal{L}(\mu) &= \prod_{i=1}^n \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(X_i - \mu)) = \frac{1}{(2\pi I_0(\kappa))^n} \prod_{i=1}^n \exp(\kappa \cos(X_i - \mu)) \\ \log \mathcal{L}(\mu) &= \kappa \sum_{i=1}^n \cos(X_i - \mu) - n \log(2\pi I_0(\kappa)) \\ (\log \mathcal{L}(\mu))' &= \kappa \sum_{i=1}^n \sin(X_i - \mu) = \kappa(\cos(\mu) \sum_{i=1}^n \sin(X_i) - \sin(\mu) \sum_{i=1}^n \cos(X_i)) \stackrel{\text{set}}{=} 0 \end{aligned}$$

which implies that the estimate $\hat{\mu}$ satisfies $\cos(\hat{\mu})\Sigma_{i=1}^n \sin(X_i) - \sin(\hat{\mu})\Sigma_{i=1}^n \cos(X_i) = 0$. Next,

$$\frac{\sum_{i=1}^n \sin(X_i)}{\sum_{i=1}^n \cos(X_i)} = \frac{\sin(\hat{\mu})}{\cos(\hat{\mu})} = \tan(\hat{\mu}) \implies \hat{\mu} = \arctan(\frac{\sum_{i=1}^n \sin(X_i)}{\sum_{i=1}^n \cos(X_i)})$$

which has more than one solution; the solution where $(\log \mathcal{L}(\hat{\mu}))'' = -\kappa\Sigma_{i=1}^n \cos(X_i - \hat{\mu}) < 0$ is therefore the MLE.

(b) i. See the attached code and plots. The mean direction is chosen to be 0.

ii. See the attached code and numerical result. Note that the log-likelihood plot from the previous part shows that the function is concave down at the calculated estimate, confirming that it is the MLE.

iii. The jackknife estimate is $\widehat{\text{se}}(\hat{\mu}) = (\frac{n-1}{n}\Sigma_{i=1}^n(\hat{\mu}_{-i} - \hat{\mu}_\bullet))^{1/2}$ and the information estimate is $\widehat{\text{se}}(\hat{\mu}) = (-\frac{d^2}{d\mu^2} \ln \mathcal{L}(\hat{\mu}))^{-1/2} = (\kappa\Sigma_{i=1}^n \cos(X_i - \hat{\mu}))^{-1/2}$. See the attached code and numerical results. The estimates are similar, with the jackknife estimate being larger. In addition, the estimates are quite large with respect to $\hat{\mu}$, though this is likely due to the initial choice of $\mu = 0$.

(c) Define $\ell(\theta) = \ln \mathcal{L}(\theta)$ for some parameter θ. Assuming regularity conditions, we have $\hat{\mu} \xrightarrow{p} \mu$, meaning we can approximate μ with a second order Taylor expansion around $\hat{\mu}$:

$$\ell(\mu) \approx \ell(\hat{\mu}) + \ell'(\hat{\mu})(\mu - \hat{\mu}) + \frac{1}{2}\ell''(\hat{\mu})(\mu - \hat{\mu}) = \ell(\hat{\mu}) + \frac{1}{2}\ell''(\hat{\mu})(\mu - \hat{\mu}) \text{ since } \ell'(\hat{\mu}) = 0$$
$$\implies 2(\ell(\hat{\mu}) - \ell(\mu)) \approx -\ell''(\hat{\mu})(\mu - \hat{\mu}).$$

Next, the von Mises distribution is part of the exponential family since its density can be written as

$$\begin{aligned} f(x;\kappa,\mu) &= \exp(c(\kappa,\mu)T(x) - d(\kappa,\mu) + h(x)) \\ &= \exp(\kappa\cos(x)\cos(\mu) + \kappa\sin(x)\sin(\mu) - \log(I_0(\kappa)) - \log(2\pi)) \\ &= \exp(\begin{bmatrix}\kappa\cos(\mu)\\ \kappa\sin(\mu)\end{bmatrix} \cdot \begin{bmatrix}\cos(x)\\ \sin(x)\end{bmatrix} - \log(I_0(\kappa)) - \log(2\pi)). \end{aligned}$$

Then, $-\ell''(\hat{\mu}) \approx nI(\mu)$ by the slides of lecture 11, so we have $2(\ln \mathcal{L}(\hat{\mu}) - \ln \mathcal{L}(\mu)) \approx nI(\mu)(\hat{\mu} - \mu)^2$ as needed. Note that $2(\ln \mathcal{L}(\hat{\mu}) - \ln \mathcal{L}(\mu))$ is an approximate pivot for μ since $\sqrt{nI(\mu)}(\hat{\mu} - \mu)$ approximately $\sim \mathcal{N}(0,1)$ implies $nI(\mu)(\hat{\mu} - \mu)^2$ approximately $\sim \chi^2(1)$.

Addendum: We can also show that $-\ell''(\mu) \approx nI(\mu^*)$, which represents the theoretical Fisher information here. WLOG, assume that $\mu < \hat{\mu}$, so the mean value theorem states that $\exists c \in (\mu, \hat{\mu})$ such that $\ell''(\hat{\mu}) - \ell''(\mu) = (\hat{\mu} - \mu)\ell'''(c)$. Since $\hat{\mu} \xrightarrow{p} \mu$ as before and $|\ell'''(c)| = \kappa\Sigma_{i=1}^n \sin(X_i - c) \leq n\kappa$, we have $(\hat{\mu} - \mu)\ell'''(c) \xrightarrow{p} 0$ by continuous mapping, implying $\ell''(\hat{\mu}) \xrightarrow{p} \ell''(\mu)$. Furthermore, $\frac{1}{n}\ell''(\mu) = \frac{1}{n}\Sigma_{i=1}^n \frac{\partial^2}{\partial\mu^2} \ln f(X_i;\kappa,\mu) \xrightarrow{p} \mathbb{E}[\frac{\partial^2}{\partial\mu^2} \ln f(X_i;\kappa,\mu)] = -I(\mu^*)$ by the WLLN.

(d) Assume $I(\mu) \neq 0$. Using the approximate pivot,

$$\begin{aligned} 0.95 &\approx \mathbb{P}(nI(\mu)(\hat{\mu} - \mu)^2 \leq \chi^2_{0.95}) = \mathbb{P}((\hat{\mu} - \mu)^2 \leq \frac{\chi^2_{0.95}}{nI(\mu)}) \\ &= \mathbb{P}\Big(\hat{\mu} - \sqrt{\frac{\chi^2_{0.95}}{nI(\mu)}} \leq \mu \leq \hat{\mu} + \sqrt{\frac{\chi^2_{0.95}}{nI(\mu)}}\Big) \end{aligned}$$

where $\chi^2_{0.95}$ is the chi-squared critical value corresponding to a 95% CI for 1 degree of freedom. Using the observed Fisher information,

$$0.95 \approx \mathbb{P}(-1.96 \leq \sqrt{nI(\mu)}(\hat{\mu} - \mu) \leq 1.96) = \mathbb{P}\Big(\hat{\mu} - \frac{1.96}{\sqrt{nI(\mu)}} \leq \mu \leq \hat{\mu} + \frac{1.96}{\sqrt{nI(\mu)}}\Big)$$

since $\sqrt{nI(\mu)}(\hat{\mu} - \mu)$ approximately $\sim \mathcal{N}(0,1)$. Thus, the 95% CIs are $[\hat{\mu} - \sqrt{\frac{\chi^2_{0.95}}{nI(\mu)}}, \hat{\mu} + \sqrt{\frac{\chi^2_{0.95}}{nI(\mu)}}]$ and $[\hat{\mu} - \frac{1.96}{\sqrt{nI(\mu)}}, \hat{\mu} + \frac{1.96}{\sqrt{nI(\mu)}}]$ respectively.

(e) See the attached code and numerical results. Note that $(nI(\mu))^{-1/2} = \widehat{se}(\hat{\mu})$, which was calculated in part (b) (iii). The two intervals are the same, which makes sense since $\sqrt{\chi^2_{0.95}} = 1.96$. Finally, both intervals contain the true $\mu = 0$.