Homework 2

Feng Zhao June 22, 2021

2.1. Using the definition of Fisher information, we have

$$\begin{aligned}I(\theta) &= \int \frac{1}{\theta}\left(\frac{\partial}{\partial\theta}\log[\frac{f(x/\theta)}{\theta}]\right) f(\frac{x}{\theta})dx \\ &= \int \frac{1}{\theta}\left(\frac{f'(x/\theta)(-x)}{f(x/\theta)\theta^2} - \frac{1}{\theta}\right) f(\frac{x}{\theta})dx\end{aligned}$$

After changing of variable, we get

$$I(\theta) = \frac{1}{\theta^2}\int\left(\frac{xf'(x)}{f(x)} + 1\right) f(x)dx$$

2.2. Let $\ell(x,y;\rho) = f_\rho(x,y)$, we first compute its partial derivative about ρ.

$$-\frac{\partial \ell(x,y;\rho)}{\partial\rho} = \frac{1}{1-\rho^2}[-\rho + \frac{\rho(x^2+y^2-\rho xy) - xy}{1-\rho^2}]$$

The maximal value of $\sum_{i=1}^n \ell(x_i, y_i;\rho)$ is achieved when $\sum_{i=1}^n \frac{\partial\ell(x,y;\rho)}{\partial\rho} = 0$, from which we get:

$$g(\rho) := \rho^3 - B\rho^2 + (A-1)\rho - B = 0$$

where $A = \sum_{i=1}^n x_i^2 + y_i^2$ and $B = \sum_{i=1}^n x_i y_i$. Since $g(1) = \sum_{i=1}^n (x_i - y_i)^2 > 0$, $g(-1) = -g(1) < 0$ and $g'(\rho) > 0$ for $\rho \in [-1,1]$, there is a unique root of $g(\rho) = 0$ in the interval $(-1,1)$, which is our MLE estimator $\hat{\rho}$. By the theorem of asymptotic normality, $\sqrt{n}(\hat{\rho}-\rho) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{MLE}})$ where $\sigma^2_{\mathrm{MLE}} = \frac{1}{I(\rho)}$. The second order derivative of $\ell(x,y;\rho)$ is given by

$$-\frac{\partial^2\ell(x,y;\rho)}{\partial\rho^2} = -\frac{1+\rho^2}{(1-\rho^2)^2} + \frac{(1-\rho^2)[-2\rho xy + x^2 + y^2] + 4\rho[-\rho^2 xy + \rho(x^2+y^2) - xy]}{(1-\rho^2)^3}$$

Using $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$ and $\mathbb{E}[XY] = \rho$, we can get

$$I(\rho) = -\mathbb{E}\left[\frac{\partial^2\ell(X,Y;\rho)}{\partial\rho^2}\right] = \frac{1+\rho^2}{(1-\rho^2)^2}$$

Therefore, $\sigma^2_{\mathrm{MLE}} = \frac{(1-\rho^2)^2}{1+\rho^2}$.

2.3. (a) By trapezoidal rule, $\sum_{i=2}^{n-1}\ln i \leq \int_1^n \ln x dx - \frac{\ln n}{2}$. Then $\ln(n!) = \sum_{i=2}^{n-1} + \ln n \leq \int_1^n x dx - \frac{\ln n}{2} + \ln n = \ln n + n\ln n - n \Rightarrow n! \leq n(\frac{n}{e})^n$. On the other hand, $\ln(n!) = \sum_{i=1}^n \ln i \geq \int_1^n \ln x dx = n\ln n - n + 1 \geq n\ln n - n \Rightarrow n! \geq (\frac{n}{e})^n$.

(b) We first prove that $\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} = H(p)$ where $H(p) := -p\log p - (1-p)\log(1-p)$ is the entropy of Bernoulli random variable Bern(p). Indeed, using the inequality in (a), we have

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{(n-k)!k!} \leq \frac{n(n/e)^n}{((n-k)/e)^{n-k}(k/e)^k} \\ \Rightarrow \frac{1}{n}\log\binom{n}{k} &\leq \frac{\log n}{n} - \frac{n-k}{n}\log\frac{n-k}{n} - \frac{k}{n}\log\frac{k}{n}\end{aligned}$$

Similarly $\frac{1}{n}\log\binom{n}{k} \geq \frac{-\log k - \log(n-k)}{n} - \frac{n-k}{n}\log\frac{n-k}{n} - \frac{k}{n}\log\frac{k}{n}$. As $n\to\infty$, $k/n\to p$. Therefore, $\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} = H(p)$. Similar argument can be made to show that

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{\lfloor np_1\rfloor \ldots \lfloor np_{m-1}\rfloor (n-\sum_{i=1}^{m-1}\lfloor np_i\rfloor)} = H(p_1,\ldots,p_m)$$

where $H(p_1,\ldots,p_m) := -\sum_{i=1}^m p_i\log p_i$ is the entropy of categorical distribution with parameter $(p_1,\ldots,p_m)$.

2.4. (a) We will show another general inequality:

$$\sum_{i=0}^{d}\binom{k}{i}q^i(1-q)^{k-i} \leq \exp(-kD(\frac{d}{k}||q)) \text{ for } d < kq \tag{1}$$

for $q\in(0,1)$. If $q=\frac{1}{2}$, then (1) becomes $\sum_{i=0}^{d}\binom{k}{i} \leq \exp(kh(d/k))$, which is the inequality required. We also observe the fact that if we make variable transformation $d' = k-d, q' = 1-q$, then (1) becomes

$$\sum_{i=d}^{k}\binom{k}{i}q^i(1-q)^{k-i} \leq \exp(-kD(\frac{d}{k}||q)) \text{ for } d > kq \tag{2}$$

Equation (2) is equivalent with (1) and we only need to show (2). Consider a Binomial distribution $X\sim \text{Binom}(k,q)$, then the probability $P(X\geq d)$ equals the left hand of (2). Using Chernoff inequality, we have

$$P(X\geq d) \leq \frac{\mathbb{E}[\exp(sX)]}{e^{sd}} = \exp(k[\ln(1-q+qe^s) - s\frac{d}{k}])$$

We choose $s>0$ to minimize the right hand side of the above equation: $s = \ln\frac{(1-q)d}{q(k-d)}$. Then after simplification we get $P(X\geq d) \leq \exp(-kD(\frac{d}{k}||q))$.

(b) We have already shown that $P(X\geq d)$ is upper bounded by $\exp(-kD(\frac{d}{k}||q))$ for $d>kq$, and $P(X\leq d)\leq \exp(-kD(\frac{d}{k}||q))$ for $d<kq$.

To get the lower bound, we just need the inequality

$$\binom{k}{d} \geq \frac{1}{k+1}(d/k)^{-k}(\frac{k-d}{k})^{d-k} \tag{3}$$

which is (11.50) from [1]. Then $P(X \geq d) \geq \binom{k}{d} q^d (1-q)^{k-d} \geq \frac{1}{k} q^d (1-q)^{k-d} (d/k)^{-k} (\frac{k-d}{k})^{d-k} = \frac{1}{k+1} \exp(-kD(\frac{d}{k}||q))$

References

[1] Cover, Thomas M. Elements of information theory. John Wiley & Sons, 1999.