Tsinghua-Berkeley Shenzhen Institute
Large Deviation and High Dimensional Statistics
Spring 2021

Problem Set 3

Issued: Wednesday 21st April, 2021 **Due:** Wednesday 5th May, 2021

3.1. For the scale family with density $\theta^{-1}f(x/\theta)$, $\theta > 0$, where $f(x) > 0$ and $f'(x)$ exists for all x, show that the Fisher information is

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

3.2. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be i.i.d. according to a bivariate normal distribution with means $\mathbb{E}[X_i] = \mathbb{E}[Y_i] = 0$, variances $\mathrm{Var}[X_i] = \mathrm{Var}[Y_i] = 1$, and unknown correlation coefficient ρ. We have

$$f_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right), \quad -1 < \rho < 1$$

Please find the MLE of ρ and evaluate its asymptotic normality.

3.3. We hope to derive an asymptotic value of $\binom{n}{k}$.

(a) Firstly, let's prove the lemma about Stirling's approximation of factorials, which we have used before.

$$\left(\frac{n}{e}\right)^n \le n! \le n\left(\frac{n}{e}\right)^n$$

Please justify the following steps:

$$\ln(n!) = \sum_{i=2}^{n-1} \ln i + \ln n \le \cdots$$

$$\ln(n!) = \sum_{i=1}^{n} \ln i \ge \cdots$$

(b) If $0 < p < 1$, and $k = \lfloor np \rfloor$, i.e., k is the largest integer less than or equal to np, then please find

$$\lim_{n\to\infty} \frac{1}{n} \log \binom{n}{k}$$

Now let p_i's be a probability distribution on m symbols. Guess what is

$$\lim_{n\to\infty} \frac{1}{n} \log \binom{n}{\lfloor np_1\rfloor \ \lfloor np_2\rfloor \ \cdots \ \lfloor np_{m-1}\rfloor \ (n - \sum_{i=1}^{m-1}\lfloor np_i\rfloor)}$$

3.4. Show that

(a) If $d < \frac{k}{2}$,

$$\sum_{i=1}^{d} \binom{k}{i} \leq \exp\left(kh\left(\frac{d}{k}\right)\right),$$

where $h(\cdot)$ is the binary entropy function that $h(p) \triangleq -p \log p - (1-p) \log(1-p)$.

(b) The probability that in k independent trials an event of probability q occurs d times or less/more, according to whether d is less or greater than kq, is bounded above by $\exp(-kD(d/k\|q))$, where

$$D(p\|q) \triangleq p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \tag{1}$$

Moreover, show in both cases that the upper bound divided by $k+1$ is a lower bound.