Tsinghua-Berkeley Shenzhen Institute
INFORMATION THEORY AND STATISTICAL LEARNING
Fall 2020

Homework 4

Hanmo Chen November 17, 2020

- **Acknowledgments: For Problem 1, I refer to `https://en.wikipedia.org/wiki/Incomplete_gamma_function` for Incomplete Gamma function** None

- **Collaborators:** I finish this homework by myself.

- *I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

Hanmo Chen

4.1. (a) Because $\mathrm{x}_i \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$, $\mathrm{y} = \sum\limits_{i=1}^{n} \frac{\mathrm{x}_i^2}{\sigma^2} \sim \chi_n^2$.

$\mathbb{P}(Y \geqslant n\alpha^2/\sigma^2) = \frac{\Gamma(\frac{n}{2}, \frac{n\alpha^2}{2\sigma^2})}{\Gamma(\frac{n}{2})}$ where $\Gamma(s,x)$ denotes the upper Incomplete Gamma function.

So

$$\begin{aligned} -\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\mathrm{x}_i^2 \geqslant \alpha^2) &= -\frac{1}{n}\log\mathbb{P}(Y \geqslant \frac{n\alpha^2}{\sigma^2}) \\ &= -\frac{1}{n}\log\frac{\Gamma(\frac{n}{2}, \frac{n\alpha^2}{2\sigma^2})}{\Gamma(\frac{n}{2})} \end{aligned} \tag{1}$$

To find the asymptotic property, using Sanov's theorem,

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\mathrm{x}_i^2 \geqslant \alpha^2) = \inf_{\mathbb{E}_P[X^2]\geqslant\alpha^2} D(P\|\mathcal{N}(0,\sigma^2)) \tag{2}$$

Suppose the distribution P has pdf $f(x)$, it can be seen as an optimization problem with constraints, that is,

$$\begin{aligned} \min \qquad & D(P\|\mathcal{N}(0,\sigma^2)) = \int f(x)\log\frac{f(x)}{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}} \\ \text{s.t.} \qquad & \int f(x)x^2 \geqslant \alpha^2 \\ & \int f(x) = 1 \end{aligned} \tag{3}$$

Define

$$J(f) = \int f(x) \log \frac{f(x)}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}} + \lambda(\int f(x)x^2 - \alpha^2) + \mu(\int f(x) - 1) \tag{4}$$

And let $\frac{\partial J}{\partial f} = 0$ we have

$$\frac{\partial J}{\partial f} = \log f(x) + \lambda x^2 + \mu = 0 \tag{5}$$

So $f(x) = \exp^{-\mu - \lambda x^2}$, which is normal distribution and satisfies $\mathbb{E}[X^2] \geqslant \alpha^2$. So $P^* = \mathcal{N}(0, \alpha^2)$ and

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\frac{1}{n} \sum_{i=1}^{n} \mathrm{x}_i^2 \geqslant \alpha^2) &= D(P^* \| \mathcal{N}(0, \sigma^2)) \\ &= \int_{\mathbb{R}} f(x) \log \frac{\frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{x^2}{2\alpha^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}} dx \\ &= \ln \frac{\sigma}{\alpha} + \frac{1}{2}(\frac{\alpha^2}{\sigma^2} - 1) \end{aligned} \tag{6}$$

(b) Using the conclusion from (a), $P^* = \mathcal{N}(0, \alpha^2)$.

4.2. (a) To prove the following lemma,

$$\left(\frac{n}{e}\right)^n \leqslant n! \leqslant n\left(\frac{n}{e}\right)^n \tag{7}$$

Which is equivalent to,

$$n \ln n - n \leqslant \ln(n!) \leqslant (n+1) \ln n - n \tag{8}$$

For the left part, notice that $\ln(1 + \frac{1}{i}) < \frac{1}{i}$ for $i \geqslant 1$ which leads to,

$$(i+1)\ln(i+1) - i \ln i - 1 < \ln(i+1) \tag{9}$$

Sum for $i = 1, 2, \cdots, n-1$, we have

$$n \ln n - (n-1) < \sum_{i=1}^{n-1} \ln(i+1) = \ln(n!) \tag{10}$$

For the right part, it holds only when $n \geqslant 7$. It is easy to check $n = 7$. So $\ln(7!) \leqslant 8 \ln 7 - 7$

And for $n \geqslant 8$, because $\ln(1+x) > \frac{x}{x+1}$ for $x > 0$, $\ln(1 + \frac{1}{i}) > \frac{1}{i+1}$, $\ln i < (i+1)\ln(i+1) - i \ln i - 1$.

Sum for $i = 7, \cdots, n-1$

$$\ln(6!)+\sum_{i=7}^{n-1}\ln i+\ln n < 7\ln 7-7+n\ln n+-7\ln 7+\ln n-(n-7) = (n+1)\ln n-n \tag{11}$$

So

$$\left(\frac{n}{e}\right)^n \leqslant n! \leqslant n\left(\frac{n}{e}\right)^n \tag{12}$$

(b) From (a) we have that as $n \to \infty$,

$$\frac{\ln(n!)}{n} \sim \ln\frac{n}{e} \tag{13}$$

Therefore

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} &= \lim_{n\to\infty}\frac{1}{n}\log\frac{n!}{k!(n-k)!}\\ &= \lim_{n\to\infty}\log\frac{n}{e} - p\log\frac{pn}{e} - (1-p)\log\frac{(1-p)n}{e}\\ &= -p\log p - (1-p)\log(1-p) = H(p)\end{aligned} \tag{14}$$

Another explanation using Sanov's theorem, suppose $X_1, X_2, \cdots, X_n \overset{i.i.d}{\sim} Bernoulli(\frac{1}{2})$ consider

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}X_i = p) \tag{15}$$

On the one hand, $\sum_{i=1}^{n} X_i \sim Binomial(n, \frac{1}{2})$, so

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}X_i = p) = \lim_{n\to\infty} -\frac{1}{n}\log\left[\binom{n}{k}\frac{1}{2^n}\right] = 1-\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} \tag{16}$$

On the other hand, using Sanov's theorem,

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}X_i = p) &= D(Bernoulli(p)\|Bernoulli(\frac{1}{2}))\\ &= p\log 2p + (1-p)\log(2(1-p))\\ &= 1 + p\log p + (1-p)\log(1-p)\\ &= 1 - H(p)\end{aligned} \tag{17}$$

Also we can get

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} = H(p) \tag{18}$$

Using the same way but using categorical and multinomial distribution instead of Bernoulli and bionomial distribution.

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\begin{array}{c} n \\ \lfloor np_1 \rfloor \lfloor np_2 \rfloor \cdots \lfloor np_{m-1} \rfloor \left(n - \sum_{i=1}^{m-1} \lfloor np_i \rfloor \right) \end{array} \right) = - \sum_{i=1}^{m} p_i \log p_i \tag{19}$$

where $\sum_{i=1}^{m} p_i = 1$.

4.3.