

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 $x_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, $y = \sum_{i=1}^n \frac{x_i^2}{\sigma^2} \sim \chi_n^2$.

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

So

$$\begin{aligned} -\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n x_i^2 \geq \alpha^2\right) &= -\frac{1}{n} \log \mathbb{P}\left(Y \geq \frac{n\alpha^2}{\sigma^2}\right) \\ &= -\frac{1}{n} \log \frac{\Gamma(\frac{n}{2}, \frac{n\alpha^2}{2\sigma^2})}{\Gamma(\frac{n}{2})} \end{aligned} \quad (1)$$

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

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n x_i^2 \geq \alpha^2\right) = \inf_{\mathbb{E}_P[X^2] \geq \alpha^2} D(P \parallel \mathcal{N}(0, \sigma^2)) \quad (2)$$

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

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

Define

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

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

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

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

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

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

4.2. (a) To prove the following lemma,

$$\left(\frac{n}{e}\right)^n \leq n! \leq n \left(\frac{n}{e}\right)^n \quad (7)$$

Which is equivalent to,

$$n \ln n - n \leq \ln(n!) \leq (n+1) \ln n - n \quad (8)$$

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

$$(i+1) \ln(i+1) - i \ln i - 1 < \ln(i+1) \quad (9)$$

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

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

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

And for $n \geq 8$, 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 $i = 7, \dots, n-1$

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

So

$$\left(\frac{n}{e}\right)^n \leq n! \leq n \left(\frac{n}{e}\right)^n \quad (12)$$

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

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

Therefore

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

Another explanation using Sanov's theorem, suppose

$X_1, X_2, \dots, X_n \stackrel{i.i.d}{\sim} \text{Bernoulli}(\frac{1}{2})$ consider

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i = p\right) \quad (15)$$

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

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

On the other hand, using Sanov's theorem,

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i = p\right) &= D(\text{Bernoulli}(p) \parallel \text{Bernoulli}(\frac{1}{2})) \\ &= p \log 2p + (1-p) \log(2(1-p)) \\ &= 1 + p \log p + (1-p) \log(1-p) \\ &= 1 - H(p) \end{aligned} \quad (17)$$

Also we can get

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

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\binom{n}{\lfloor np_1 \rfloor \lfloor np_2 \rfloor \cdots \lfloor np_{m-1} \rfloor} \left(n - \sum_{i=1}^{m-1} \lfloor np_i \rfloor \right) \right) = - \sum_{i=1}^m p_i \log p_i \quad (19)$$

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

- 4.3. (a) $\mathbb{E}_p[y] = 0$ means that $p_0 = 1, p_1 = p_2 = 0$. So \mathcal{L}_0 is just a single point $(1, 0, 0)$.

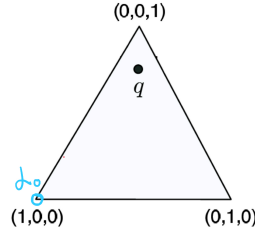


Figure 1: \mathcal{L}_0

- (b) $\mathbb{E}_p[y] = \frac{1}{2}$ means $\mathcal{L}_{\frac{1}{2}} = \{p = (p_0, p_1, p_2) : p_0 + p_1 + p_2 = 1, p_1 + 2p_2 = \frac{1}{2}\}$ which is a line passing $(\frac{1}{2}, \frac{1}{2}, 0)$ and $(\frac{3}{4}, 0, \frac{1}{4})$.

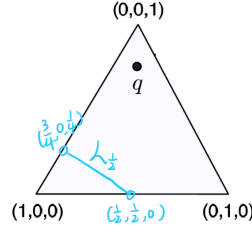


Figure 2: $\mathcal{L}_{\frac{1}{2}}$

- (c) The exponential family is $\mathcal{E} = \{\tilde{q} : \tilde{q} = qe^{sf(y) - \alpha(s)}\}$. Following Pythagoras's Identity, let $f(y) = y$ so \mathcal{E} is orthogonal to $\mathcal{L}_{\frac{1}{2}}$. $\mathcal{E} = \{\tilde{q} : \tilde{q} = qe^{sy - \alpha(s)}\}$. Denote $\lambda = e^s$, so $\tilde{q}_0 = \frac{1}{1+\lambda+4\lambda^2}, \tilde{q}_1 = \frac{\lambda}{1+\lambda+4\lambda^2}, \tilde{q}_2 = \frac{4\lambda^2}{1+\lambda+4\lambda^2}$. Notice that \mathcal{E} passes $(1, 0, 0)$ and $(0, 0, 1)$ and $\tilde{q}_1 \leq \frac{1}{5}$.
- (d) Using the Lagrange-Multiplier method we can induce that the I-projection p^* of q onto $\mathcal{L}_{\frac{1}{2}}$ belongs to \mathcal{E} . So $p^* \in \mathcal{L}_{\frac{1}{2}} \cap \mathcal{E}$. By $\tilde{q}_1 + 2\tilde{q}_2 = \frac{1}{2}$ we can solve $\lambda = \frac{1}{4}, p^* = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$.
- (e) As figure 5

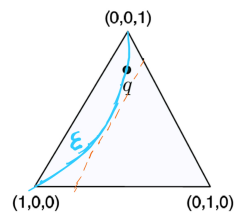


Figure 3: \mathcal{E}

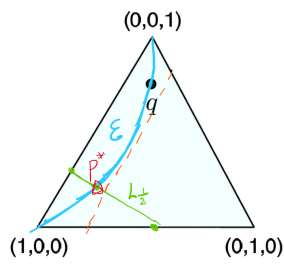


Figure 4: p^*

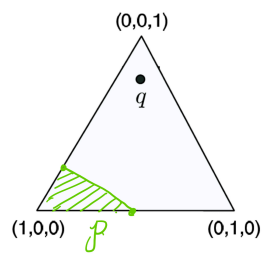


Figure 5: \mathcal{P}

- (f) First, for any $p \in \mathcal{P}$, it belongs to some $\mathcal{L}_\gamma = \{p : \mathbb{E}_p[y] = \gamma\}$ and $\gamma \leq \frac{1}{2}$. So $D(p\|q) \geq D(p_\gamma^*\|q)$ where p_γ^* is the I-projection of q onto \mathcal{L}_γ . And $p_\gamma^* \in \mathcal{E}$. Thus,

$$p^* = \arg \min_{p \in \mathcal{P}} D(p\|q) = \arg \min_{p \in \mathcal{P} \cap \mathcal{E}} D(p\|q) \quad (20)$$

For $\tilde{q}_s \in \mathcal{P} \cap \mathcal{E}$, $\gamma = \mathbb{E}_{\tilde{q}}[y] = \tilde{q}_1 + 2\tilde{q}_2 = \frac{\lambda + 8\lambda^2}{1 + \lambda + 4\lambda^2}$, $\lambda = e^s$.

$$\frac{d\gamma}{d\lambda} = \frac{1 + 16\lambda + 4\lambda^2}{(1 + \lambda + 4\lambda^2)^2} \geq 0 \quad (21)$$

So γ strictly increases with λ , then γ strictly increases with s , vice versa. And when $\gamma = \frac{1}{2}$, $\lambda = \frac{1}{4}$, $s = -\ln 4$

And $D(\tilde{q}_s\|q) = s\mathbb{E}_{\tilde{q}_s}[y] - \alpha(s)$.

$$\frac{\partial D(\tilde{q}_s\|q)}{\partial s} = s \text{Var}_{\tilde{q}_s}[y] \leq 0, \quad \text{for } s \leq -\ln 4 < 0 \quad (22)$$

So $\frac{\partial D(\tilde{q}_s\|q)}{\partial \gamma} \leq 0$ for $\gamma \leq \frac{1}{2}$. To minimize $D(\tilde{q}_s\|q)$, $\gamma^* = \frac{1}{2}$, $p^* = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$

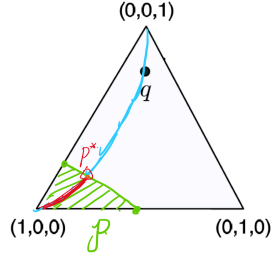


Figure 6: p^*