Department of Electrical and Computer Engineering
ECE 776 - Information Theory

Reading

Cover and Thomas, Chapter 12.

Homework 8

1. **Problem 12.1**

   (a) $f_i = \mathcal{N}(0, \sigma_i^2)$, $i = 1, 2$.

$$\begin{aligned} D(f_1 \| f_2) &= \int_{-\infty}^{\infty} f_1(x) \left[\frac{1}{2} \ln \frac{\sigma_2^2}{\sigma_1^2} - \left(\frac{x^2}{\sigma_1^2} - \frac{x^2}{\sigma_2^2} \right) \right] dx \\ &= \frac{1}{2} \left[\ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} - 1 \right] \end{aligned}$$

   (b) $f_i = \lambda_i e^{-\lambda_i x}$,$x \geq 0$, $i = 1, 2$.

$$\begin{aligned} D(f_1 \| f_2) &= \int_0^{\infty} f_1(x) \left[\ln \frac{\lambda_1}{\lambda_2} - (\lambda_1 x - \lambda_2 x) \right] dx \\ &= \ln \frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} - 1 \end{aligned}$$

   (c) $f_1 = \text{unif}[0, 1]$, $f_2 = \text{unif}[a, a+1]$, $0 < a < 1$.

$$\begin{aligned} D(f_1 \| f_2) &= \int_0^1 f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \\ &= \int_0^a f_1(x) \ln \frac{f_1(x)}{0} dx + \int_a^1 f_1(x) \ln 1 dx \\ &= \infty \end{aligned}$$

   In this case, $D(f_1 \| f_2) = \infty$ means that the error exponent is ∞, which means for large number of trial, we can certainly distinguish f_1 and f_2, since the probability of $x \in [0, a]$ or $x \in [1, a+1]$ is non-zero.

   (d) $f_1 = \text{Bern}(1/2)$, $f_2 = \text{Bern}(1)$.

$$\begin{aligned} D(f_1 \| f_2) &= \frac{1}{2} \ln \frac{1/2}{1} + \frac{1}{2} \ln \frac{1/2}{0} \\ &= \infty \end{aligned}$$

   The implication is the same as before (a tail indicates fair coin).

2. **Problem 12.4**
$\mathcal{P}_1$ is define by the constraints $\{h_i\}$ and $\mathcal{P}_2$ is define by the constraints $\{g_i\}$. Hence, $\mathcal{P}_1 \cap \mathcal{P}_2$ is define by the union of constraints $\{h_i\}$ and $\{g_i\}$. Q is given with $q(x)$ distribution.

Assuming that all constraints are active, we can solve the two minimization of $D(p\|q)$ and $D(r\|q)$ as

$$\begin{aligned} p^*(x) &= \arg\min_{p\in\mathcal{P}_1} D(p\|q) \\ &= c_1 q(x) \exp\left(\sum_{i=1}^{r} \lambda_i h_i(x)\right), \\ r^*(x) &= \arg\min_{p\in\mathcal{P}_1\cap\mathcal{P}_2} D(p\|q) \\ &= c_2 q(x) \exp\left(\sum_{i=1}^{r} \lambda_i h_i(x) + \sum_{j=1}^{s} \nu'_j g_i(x)\right), \end{aligned}$$

where the constants are so chosen as to satisfy the constraints. Now we can project p^* onto $\mathcal{P}_1 \cap \mathcal{P}_2$ and get

$$\begin{aligned} p^{**}(x) &= \arg\min_{p\in\mathcal{P}_1\cap\mathcal{P}_2} D(p\|p^*) \\ &= c_3 p^*(x) \exp\left(\sum_{j=1}^{s} \nu_j g_i(x)\right), \\ &= c_3 c_1 q(x) \exp\left(\sum_{i=1}^{r} \lambda_i h_i(x) + \sum_{j=1}^{s} \nu_j g_i(x)\right). \end{aligned}$$

Notice that r^* and p^{**} have the same form. Furthermore, since the constants are chosen to satisfy the same constraints, $r^* = p^{**}$.

3. **Problem 12.5**
We wish to count the number of sequences satisfying a certain property. Instead of directly counting the sequences, we will calculate the probabilities of the set under an uniform distribution. Since the uniform distribution puts a probability of $1/m^n$ on every sequence of length n, we can count the sequences by multiplying the probability of the set by m^n.

The probability of the set can be calculated from Sanov's theorem. Let Q be the uniform distribution, and let E be the set of sequences of length n satisfying $(1/n)\sum g(x_i) \geq \alpha$. Then, by Sanov's theorem, we have

$$Q^n(E) \doteq 2^{-nD(P^*\|Q)},$$

where P^* is the type in E that is closest to Q. Since Q is the uniform distribution, $D(P\|Q) = \log m - H(P)$, and therefore P^* is the type in E that has maximum entropy. Thus, if we let

$$H^* = \max_{P:\sum \Pr(i)g(i)\geq\alpha} H(P),$$

we have

$$Q^n(E) \doteq 2^{-n(\log m - H^*)}.$$

Multiplying this by m^n to find the number of sequences in this set, we get

$$|E| \doteq 2^{-n(\log m - H^*)} m^n = 2^{nH^*}.$$

4. **Problem 12.8**

(a) Given $f_\theta(x) = \mathcal{N}(0,\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}$, therefore

$$f'_\theta = -\frac{1}{2}\frac{1}{\sqrt{2\pi\theta^3}} e^{-\frac{x^2}{2\theta}} + \frac{x^2}{2\theta^2}\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}},$$

and

$$\frac{f'_\theta}{f_\theta} = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}.$$

Therefore, the Fisher information is

$$\begin{aligned} j(\theta) &= E_\theta \left(\frac{\partial f_\theta(X)/\partial\theta}{f_\theta(X)} \right)^2 \\ &= E_\theta \left(\frac{1}{4\theta^2} - 2\frac{1}{2\theta}\frac{x^2}{2\theta^2} + \frac{x^4}{4\theta^4} \right) \\ &= \frac{1}{4\theta^2} - 2\frac{1}{2\theta}\frac{\theta}{2\theta^2} + \frac{3\theta^2}{4\theta^4} \\ &= \frac{1}{2\theta^2} \end{aligned}$$

(b) Given $f_\theta(x) = \theta e^{-\theta x}$, $x \geq 0$, therefore $\ln f_\theta = \ln\theta - \theta x$ and

$$\frac{d\ln f_\theta}{d\theta} = \frac{1}{\theta} - x,$$

and

$$\begin{aligned} J(\theta) &= E_\theta \left(\frac{d\ln f_\theta}{d\theta} \right)^2 \\ &= E_\theta \left(\frac{1}{\theta^2} - 2\frac{1}{\theta}x + x^2 \right) \\ &= \frac{1}{\theta} \end{aligned}$$

(c) For Cramer-Rao bound, $\mathrm{var}(T) \geq 1/J(\theta)$. Thus, the bounds are $2\theta^2$ and θ respectively.