Sample problems for the 3rd lecture (Mar. 15)

1. Suppose X_1, X_2, X_3, ... is a random process with

$$E[X_1] = E[X_2] = E[X_3] = \cdots = \mu.$$

Which consequences below are conventionally referred to as the *strong law of large numbers*?

(a) $\frac{X_1+\cdots+X_n}{n}$ converges *in probability* to μ.

(b) $\frac{X_1+\cdots+X_n}{n}$ converges *almost surely* to μ.

(c) $\frac{X_1+\cdots+X_n}{n}$ converges *with probability one* to μ.

(d) $\Pr\left[\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right] = 1$

(e) $\lim_{n\to\infty} \Pr\left[\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| < \epsilon\right] = 1$

Solution. (b), (c) and (d).

2. Suppose $0 < \epsilon < \frac{1}{3}$. Which of the following sets are convex sets?

(a) $\{(p_1, p_2, p_3) \in \mathbb{R}^3 : p_1 \ge 0, p_2 \ge 0, p_3 \ge 0 \text{ and } p_1 + p_2 + p_3 = 1\}$

(b) $\{(p_1, p_2, p_3) \in \mathbb{R}^3 : p_1 > 0, p_2 > 0, p_3 > 0 \text{ and } p_1 + p_2 + p_3 = 1\}$

(c) $\{(p_1, p_2, p_3) \in \mathbb{R}^3 : p_1 > \epsilon, p_2 > \epsilon, p_3 > \epsilon \text{ and } p_1 + p_2 + p_3 = 1\}$

(d) $\{(p_1, p_2) \in \mathbb{R}^2 : p_1 \ge 0, p_2 \ge 0 \text{ and } p_1 + p_2 \le 1\}$

(e) $\{(p_1, p_2) \in \mathbb{R}^2 : p_1 > \epsilon, p_2 > \epsilon \text{ and } p_1 + p_2 < 1 - \epsilon\}$

Solution. These are sets of probabilities that may encounter in the information-theoretical optimization.

(a) is a convex set because for $0 \le \lambda \le 1$, $(\lambda p_1 + (1-\lambda)p_1', \lambda p_2 + (1-\lambda)p_2', \lambda p_3 + (1-\lambda)p_3')$ is in the set as long as (p_1, p_2, p_3) and (p_1', p_2', p_3') are both in the set.

(b), (c), (d) and (e) are all convex sets, which can be similarly verified as (a).

3. Let $f(\boldsymbol{x}) = \sum_{i=1}^n x_i \ln(x_i)$ (i.e., multiplying entropy by -1) and let the set $\mathcal{Q} = \{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1 \text{ and each } x_i \ge 0\}$.

(a) Is $\mathcal{Q}$ a convex set? Justify your answer.

(b) Show that $f(\boldsymbol{x})$ a convex function over $\boldsymbol{x} \in \mathcal{Q}$.
Hint: Subtract $f(\lambda \cdot \boldsymbol{x} + (1-\lambda) \cdot \tilde{\boldsymbol{x}})$ from $\lambda \cdot f(\boldsymbol{x}) + (1-\lambda) \cdot f(\tilde{\boldsymbol{x}})$ and apply the fundamental inequality.

(c) Why isn't it theoretically sound to verify that $f(\boldsymbol{x})$ is a convex function over a *non-convex* set $\tilde{\mathcal{Q}}$?

(d) What is the Lagrange dual function $L(\boldsymbol{\lambda}, \boldsymbol{\nu})$ for the minimization of $f(\boldsymbol{x})$ over $\boldsymbol{x} \in \mathcal{Q}$?

(e) Determine the minimizer $\boldsymbol{x}^\diamond = \boldsymbol{x}^\diamond(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that achieves the Lagrange dual function $L(\boldsymbol{\lambda}, \boldsymbol{\nu})$.

(f) Take $\boldsymbol{x}^\diamond$ into $L(\boldsymbol{\lambda}, \boldsymbol{\nu})$ such that it is expressed as a function of $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ only. Is it a concave function of $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$? Justify your answer.

(g) Determine non-negative $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ that maximizes $L(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Then, find $\boldsymbol{x}^* = \boldsymbol{x}^\diamond(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$.

(h) Do $\boldsymbol{x}^*, \boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ satisfy the KKT condition? Justify your answer.

(i) Does the strong duality hold? Justify your answer.

Solution.

(a) It is a convex set, which can be verified by a similar approach in Problem 2.

(b) For $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ in $\mathcal{Q}$ and $0 \le \lambda \le 1$,

$$
\begin{aligned}
&[\lambda \cdot f(\boldsymbol{x}) + (1-\lambda)\cdot f(\tilde{\boldsymbol{x}})] - f(\lambda \cdot \boldsymbol{x} + (1-\lambda)\cdot \tilde{\boldsymbol{x}}) \\
&= \lambda \sum_{i=1}^{n} x_i \ln(x_i) + (1-\lambda)\sum_{i=1}^{n} \tilde{x}_i \ln(\tilde{x}_i) \\
&\quad - \sum_{i=1}^{n} (\lambda x_i + (1-\lambda)\tilde{x}_i) \ln(\lambda x_i + (1-\lambda)\tilde{x}_i) \\
&= \lambda \sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{\lambda x_i + (1-\lambda)\tilde{x}_i}\right) + (1-\lambda)\sum_{i=1}^{n} \tilde{x}_i \ln\left(\frac{\tilde{x}_i}{\lambda x_i + (1-\lambda)\tilde{x}_i}\right) \\
&\ge \lambda \sum_{i=1}^{n} x_i \left(1 - \frac{\lambda x_i + (1-\lambda)\tilde{x}_i}{x_i}\right) + (1-\lambda)\sum_{i=1}^{n} \tilde{x}_i \left(1 - \frac{\lambda x_i + (1-\lambda)\tilde{x}_i}{\tilde{x}_i}\right) \\
&\quad \text{(By the fundamental inequality)} \\
&= \lambda \sum_{i=1}^{n} (x_i - \lambda x_i - (1-\lambda)\tilde{x}_i) + (1-\lambda)\sum_{i=1}^{n} (\tilde{x}_i - \lambda x_i - (1-\lambda)\tilde{x}_i) \\
&= 0.
\end{aligned}
$$

Hence, $f(\boldsymbol{x})$ a convex function over $\boldsymbol{x} \in \mathcal{Q}$.

(c) Because $\lambda \cdot \boldsymbol{x} + (1-\lambda) \cdot \tilde{\boldsymbol{x}}$ may not lie in the non-convex $\tilde{\mathcal{Q}}$ even if both $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ are.

(d) Here we have n inequality constraints $g_i(\boldsymbol{x}) = -x_i \leq 0$ for $1 \leq i \leq n$ and 1 equality constraint $h(\boldsymbol{x}) = \sum_{i=1}^n x_i - 1 = 0$. Thus,

$$L(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} \left(\sum_{i=1}^n x_i \ln(x_i) + \sum_{i=1}^n \lambda_i(-x_i) + \nu \left(\sum_{i=1}^n x_i - 1 \right) \right).$$

(e) Taking the derivative of $L(\boldsymbol{x}; \boldsymbol{\lambda}, \boldsymbol{\nu}) = \sum_{i=1}^n x_i \ln(x_i) + \sum_{i=1}^n \lambda_i(-x_i) + \nu \left(\sum_{i=1}^n x_i - 1 \right)$ with respect to x_j for $1 \leq j \leq n$ yields

$$[1 + \ln(x_j)] - \lambda_j + \nu = 0 \text{ for } 1 \leq j \leq n.$$

This implies

$$x_j^\diamond = e^{\lambda_j - \nu - 1}.$$

(f)

$$\begin{aligned}
L(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \sum_{i=1}^n e^{\lambda_i - \nu - 1} \ln(e^{\lambda_i - \nu - 1}) + \sum_{i=1}^n \lambda_i(-e^{\lambda_i - \nu - 1}) + \nu \left(\sum_{i=1}^n e^{\lambda_i - \nu - 1} - 1 \right) \\
&= \sum_{i=1}^n (\lambda_i - \nu - 1) e^{\lambda_i - \nu - 1} - \sum_{i=1}^n \lambda_i e^{\lambda_i - \nu - 1} + \nu \sum_{i=1}^n e^{\lambda_i - \nu - 1} - \nu \\
&= -e^{-\nu - 1} \sum_{i=1}^n e^{\lambda_i} - \nu
\end{aligned}$$

The function can be shown concave with respect to $\boldsymbol{\lambda}$ and ν by taking derivatives.

(g)

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\lambda}, \boldsymbol{\nu})}{\lambda_j} &= -e^{-\nu - 1} e^{\lambda_j} < 0 \Rightarrow \lambda_j^* = 0 \text{ for } 1 \leq j \leq n. \\
\frac{\partial L(\boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \nu} &= e^{-\nu} e^{-1} n - 1 = 0 \Rightarrow \nu^* = \ln(n) - 1.
\end{aligned}$$

We conclude that $x_j^* = x^\diamond(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \frac{1}{n}$ for $1 \leq j \leq n$.

(h) The answer is yes because

$$\begin{cases}
g_i(\boldsymbol{x}^*) = -\frac{1}{n} \leq 0, \quad \lambda_i = 0 \geq 0, \quad \lambda_i g_i(\boldsymbol{x}) = 0 & 1 \leq i \leq n \\
h(\boldsymbol{x}^*) = \sum_{i=1}^n \frac{1}{n} - 1 = 0 & \\
\frac{\partial L}{\partial x_k}(\boldsymbol{x}^*; \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \frac{\partial f}{\partial x_k}(\boldsymbol{x}^*) + \sum_{i=1}^n \lambda_i^* \frac{\partial g_i}{\partial x_k}(\boldsymbol{x}^*) + \nu^* \frac{\partial h}{\partial x_k}(\boldsymbol{x}^*) & \\
\qquad = [1 + \ln(\frac{1}{n})] + 0 + [\ln(n) - 1] = 0 & 1 \leq k \leq n
\end{cases}$$

(i) Since i) $f(\boldsymbol{x})$ is convex, ii) each $g_i(\boldsymbol{x})$ is affine (and hence convex), iii) $h(\boldsymbol{x})$ is affine, iv) these functions are all differentiable, and v) the solutions $\boldsymbol{x}^*$, $\boldsymbol{\lambda}^*$ and ν^* satisfy the KKT condition, the strong duality holds.

4. Prove the self-information expression in Theorem 2.1.

5. Prove the log-sum inequality in Lemma 2.7.

   Hint: Subtract one side from the other side and apply the fundamental inequality.