NUS National University of Singapore

EE5139/EE6139 Solution 6

Semester I, AY2021/2022
Marco Tomamichel

Exercise 6.1 Data processing inequality for the total variation distance (EE5139)

Let P_X, Q_X be two distributions over some finite set $\mathcal{X}$. Let $\{T(y|x)\}_{y\in\mathcal{Y},x\in\mathcal{X}}$ be some conditional pmf, where $\mathcal{Y}$ is some finite set. Suppose $P_Y(y) = \sum_{x\in\mathcal{X}} T(y|x)\cdot P_X(x)$ and $Q_Y(y) = \sum_{x\in\mathcal{X}} T(y|x)\cdot Q_X(x)$ for each $y\in\mathcal{Y}$. Prove that

$$\delta_{\text{tvd}}(P_Y, Q_Y) \le \delta_{\text{tvd}}(P_X, Q_X).$$

Solution:

$$\begin{aligned}
\delta_{\text{tvd}}(P_Y, Q_Y) &= \sum_{y\in\mathcal{Y}} |P_Y(y) - Q_Y(y)| \\
&= \sum_{y\in\mathcal{Y}} \left| \sum_{x\in\mathcal{X}} \Big[T(y|x)\cdot (P_X(x) - Q_X(x)) \Big] \right| \\
&\le \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} \Big| T(y|x)\cdot (P_X(x) - Q_X(x)) \Big| \\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} T(y|x)\cdot \Big| (P_X(x) - Q_X(x)) \Big| \\
&= \sum_{x\in\mathcal{X}} \Big| (P_X(x) - Q_X(x)) \Big| = \delta_{\text{tvd}}(P_X, Q_X).
\end{aligned}$$

Exercise 6.2 Empirical typical set (all)

Let P_X be a probability distribution over some finite set $\mathcal{X}$. Recall that the empirical typical set (of length n and tolerance ϵ) w.r.t. P_X is defined as

$$\mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_X) := \{\boldsymbol{x}^n \in \mathcal{X}^n : \ \delta_{\text{tvd}}(f_{\boldsymbol{x}^n}, P_X) \le \epsilon\}$$

where $f_{\boldsymbol{x}^n}$ is the empirical distribution on $\mathcal{X}$ induced by the sequence $\boldsymbol{x}^n$, *i.e.*,

$$f_{\boldsymbol{x}^n}(x) := \frac{1}{n} \left| \{i \in \{1,\dots,n\} : \ x_i = x\} \right|.$$

Prove that, for any $\epsilon \in (0,1]$,

$$\lim_{n\to\infty} \mathrm{P}\left[X^n \in \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_X) \right] = 1.$$

Solution: We firstly prove the statement for $|\mathcal{X}| = 2$. Without loss of generality, assume $\mathcal{X} = \{0,1\}$. In this case, we have

$$\delta_{\text{tvd}}(f_{\boldsymbol{x}^n}, P_X) \le \epsilon \iff \left| \frac{\sum_{i=1}^n x_i}{n} - \mathbb{E}[X] \right| \le \epsilon.$$

Hence,

$$\mathrm{P}\left[X^n \in \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_X) \right] = \mathrm{P}\left[\left| \frac{\sum_{i=1}^n X_i}{n} - \mathbb{E}[X] \right| \le \epsilon \right] \ge 1 - \frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)}{\epsilon^2}$$

by Chebyshev's inequality. Since $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\mathrm{Var}\,(X) \to 0$ as $n\to\infty$, above tends to 1 as $n\to\infty$.

For the case when $|X| > 2$, for each $x \in \mathcal{X}$ we consider the "partial" empirical typical sets, and an induced binary random variable Y_x defined as follows

$$\mathcal{A}^{(n)}_{\text{emp},\epsilon,x}(P_X) := \left\{ \boldsymbol{x}^n \in \mathcal{X}^n : \, |f_{\boldsymbol{x}^n}(x) - P_X(x)| \leq \frac{2\epsilon}{|\mathcal{X}|} \right\}$$

$$Y_x = \begin{cases} 0 & \text{if } X \neq x \\ 1 & \text{if } X = x \end{cases}$$

It is not difficult to check that

$$X^n \notin \mathcal{A}^{(n)}_{\text{emp},\epsilon,x}(P_X) \iff Y_x^n \notin \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_{Y_x}).$$

Thus,

$$\begin{aligned} \mathrm{P}\left[X^n \in \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_X)\right] &\geq \mathrm{P}\left[X^n \in \mathcal{A}^{(n)}_{\text{emp},\epsilon,x}(P_X) \quad \forall x \in \mathcal{X}\right] \\ &\geq 1 - \sum_{x\in\mathcal{X}} \mathrm{P}\left[X^n \notin \mathcal{A}^{(n)}_{\text{emp},\epsilon,x}(P_X)\right] \\ &= 1 - \sum_{x\in\mathcal{X}} \mathrm{P}\left[Y_x^n \notin \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_{Y_x})\right] \end{aligned}$$

Since $\mathrm{P}\left[Y_x^n \notin \mathcal{A}^{(n)}_{\text{emp},\epsilon}(P_{Y_x})\right] \to 0$ for each x, as shown in the binary case, we have finished our proof.

Exercise 6.3 Hypothesis testing (all)

Consider the three pmfs P and Q, and S given by the probability vectors $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $s = (0, \frac{1}{2}, \frac{1}{2})$, respectively.

a.) Compute the symmetric error probabilities $\epsilon^*_{\text{sym},1}$ for hypothesis tests for all three pairs.

Solution:

$$\delta_{\text{tvd}}(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| = \frac{1}{6}, \; \delta_{\text{tvd}}(P,S) = \frac{1}{2}, \; \delta_{\text{tvd}}(Q,S) = \frac{1}{3}$$

For pmf pair P and Q:

$$\epsilon^*_{\text{sym},1} = \frac{1}{2}(1 - \delta_{\text{tvd}}(P,Q)) = \frac{5}{12}$$

For pmf pair P and S:

$$\epsilon^*_{\text{sym},1} = \frac{1}{2}(1 - \delta_{\text{tvd}}(P,S)) = \frac{1}{4}$$

For pmf pair Q and S:

$$\epsilon^*_{\text{sym},1} = \frac{1}{2}(1 - \delta_{\text{tvd}}(Q,S)) = \frac{1}{3}$$

b.) Consider the problem of ternary hypothesis testing between the three distributions, when all three have equal priors. Can you find the minimal error probability and optimal test?

Solution: If we divide $\mathcal{X}$ into 3 disjoint subsets, we have the following possible combinations: $\{(1),(2),(3)\}$, $\{(1,2),(3),(\emptyset)\}$,$\{(1,3),(2),(\emptyset)\}$,$\{(2,3),(1),(\emptyset)\}$, $\{(1,2,3),(\emptyset),(\emptyset)\}$. Then we can calculate the following minimal error probability:

$$\epsilon^*_{3,\text{sym},1} = \frac{1}{3} \min_{\mathcal{A}_1,\mathcal{A}_2\subset\mathcal{X},\mathcal{A}_1\cap\mathcal{A}_2=\emptyset} (P(\mathcal{A}_1^{\text{c}}) + Q(\mathcal{A}_2^{\text{c}}) + S(\mathcal{A}_1 \cup \mathcal{A}_2)) = \frac{1}{2},$$

where the optimal test is as follows: if the observed data satisfies $x \in \{1\}$, we say that it is generated from P; if $x \in \{2,3\}$, we say that it is generated from S.

c.) Compute $D(P\|Q)$ and $D(P\|S)$.

Solution:

$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)} = \frac{1}{2}\log\frac{9}{8}.$$

$$D(P\|S) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{S(x)} = \infty.$$

d.) Consider asymmetric hypothesis testing for P and S. For each $\epsilon > 0$, find $N_0(\epsilon) \in \mathbb{N}$ such that $\beta_n(\epsilon) = 0$ for all $n \geq N_0(\epsilon)$. What does that say about the limit $\lim_{n\to\infty} -\frac{1}{n}\log\beta_n(\epsilon)$? Interpret your result in item c) in this light.

Solution: Let $\mathcal{X} = \{1,2,3\}$ and $\mathcal{A} = \{x^n \in \mathcal{X}^n : x^n \text{ contains symbol } 1\}$. Since $P(1) > S(1) = 0$, there exists $N_0(\epsilon) \in \mathbb{N}$ such that $P^n(\mathcal{A}) \geq 1-\epsilon$ for $n \geq N_0(\epsilon)$. However, $S^n(\mathcal{A}) = 0$ and thus, $\beta_n(\epsilon) = \min_{\mathcal{A}\subset\mathcal{X}^n, P^n(\mathcal{A})\geq 1-\epsilon} S^n(\mathcal{A}) = 0$ and $\lim_{n\to\infty} -\frac{1}{n}\log\beta_n(\epsilon) = \infty$, which corresponds to $D(P\|S)$.

Exercise 6.4 A simple parity check code (EE5139)

In the first lecture we encountered a code that stores $k = 4$ bits in $n = 8$ bits by computing the parities $x_1 \oplus x_2$, $x_3 \oplus x_4$, $x_1 \oplus x_3$ and $x_2 \oplus x_4$.

a.) Give the codewords for this code and compute the minimal distance. How many errors can it detect and correct?

Solution:

codeword	min distance	codeword	min distance	codeword	min distance	codeword	min distance
00000000	3	00010101	3	00100110	3	00110011	3
01001001	3	01011100	3	01101111	3	01111010	3
10001010	3	10011111	3	10101100	3	10111001	3
11000011	3	11010110	3	11100101	3	11110000	3

Let $2t+1 = 3$ and then $t = 1$. The code can detect up to 2 bit flip errors, correct up to 1 bit flip errors and correct up to 2 erasures.

b.) Is it a linear code? If so, compute matrices G and H.

Solution: It is a linear code.

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

c.) Use the Hamming bound to determine if this code is perfect or not.

Solution: Hamming bound: $d = 3$, $n = 8$

$$16 = |C| < \frac{2^n}{\sum_{i=0}^{\lfloor (d-1)/2\rfloor}\binom{n}{i}} = \frac{2^8}{9}.$$

So this code is not perfect.

d.) Construct the dual code for this code.

Solution: For the dual code: $G^\perp = H^T$ and $H^\perp = G^T$, that is

$$G^\perp = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H^\perp = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Exercise 6.5 More properties of linear codes (all)

Show the following properties of a (binary) linear $[n, k]$-code C.

a.) The minimal distance $d(C)$ is the minimal Hamming weight of all (non-zero) codewords.

Solution: For a binary linear $[n, k]$-code C, for any $c, c' \in C$, we have $x = c \oplus c' \in C$. Minimal distance:

$$\begin{aligned} d(C) &= \min_{c,c' \in C, c \neq c'} \delta(c, c') = \min_{c,c' \in C, c \neq c'} |\{i : c_i \neq c'_i\}| \\ &= \min_{c,c' \in C, c \neq c'} |\{i : c_i \oplus c'_i = 1\}| = \min_{x \in C, x \text{ non-zero}} |\{i : x_i \neq 0\}|. \end{aligned}$$

Thus, the minimal distance is the minimal Hamming weight of all non-zero codewords.

b.) If H is the parity check matrix of C, then $d(C)$ equals the number of columns of H that are linearly dependent.

Solution: Let $H = [h_1^T, \ldots, h_n^T] \in \{0,1\}^{(n-k)\times n}$, where $h_i^T \in \{0,1\}^{(n-k)\times 1}$. Let $c = [c_1, \ldots, c_n]^T$ be the codeword with minimal Hamming weight. Since H is the parity check matrix of C, then

$$0 = Hc = \sum_{i=1}^{n} h_i^T c_i = \sum_{i: c_i \neq 0}^{n} h_i^T = 0.$$

Thus, $d(C)$ is the number of columns of H that are linearly dependent.

c.) Prove that (after permuting the coordinates if necessary) C has a generator matrix of the form $G = [I_k \; G']^T$ where I_k is the $k \times k$ identity matrix, and where G' is some $k \times (n-k)$ matrix.

Solution: Since k is the minimal number of codewords needed for a basis and k columns of generator matrix $G \in \{0,1\}^{n\times k}$ span C, then G must contain a $k \times k$ identity matrix I_k. Then by permuting the columns of G, we can always write the generating matrix as

$$G = \begin{bmatrix} I_k \\ G' \end{bmatrix}.$$

Exercise 6.6 Asymptotic property of the smooth min-entropy (EE6139)

Given a random variable X on $\mathcal{X}$ distributed according to P_X, we recall that the min-entropy of X is defined as

$$H_{\min}(X) := -\log \max_{x \in \mathcal{X}} P_X(x) = \min_{x \in \mathcal{X}} -\log P_X(x).$$

For each $\epsilon \in [0,1)$, the ϵ-smooth min-entropy of X is defined as

$$H_{\min}^{\epsilon}(X) := \max_{\delta_{\text{tvd}}(P_X,\tilde{P}_X)\leq\epsilon} H_{\min}(X)_{\tilde{P}_X} = \max_{\delta_{\text{tvd}}(P_X,\tilde{P}_X)\leq\epsilon} \left\{ -\log \max_{x\in\mathcal{X}} \tilde{P}_X(x) \right\}.$$

We are interested in showing $\frac{1}{n} H_{\min}^{\epsilon}(X^n) \to H(X)$ as $n \to \infty$, $\epsilon \to 0$ (in that order), where X^n consists of n i.i.d. copies of X, namely $P_{X^n} = \prod_{i=1}^n P_{X_i}$.

a.) For each $\epsilon > 0$, let $\mathcal{A}_\epsilon^{(n)}(X)$ denote the ϵ-typical set of X (see eq. (2.52) from the lecture notes). For each positive integer n, define a distribution on $\mathcal{X}^n$ as

$$\hat{P}_{X^n}(\boldsymbol{x}^n) := \begin{cases} P_{X^n}(\boldsymbol{x}^n) & \boldsymbol{x}^n \in \mathcal{A}_\epsilon^{(n)}(X) \\ \frac{1-P_{X^n}(\mathcal{A}_\epsilon^{(n)}(X))}{|\mathcal{X}|^n - \left|\mathcal{A}_\epsilon^{(n)}(X)\right|} & \boldsymbol{x}^n \notin \mathcal{A}_\epsilon^{(n)}(X) \end{cases}$$

Prove that for n large enough, $\delta_{\text{tvd}}(P_{X^n}, \hat{P}_{X^n}) \leq \epsilon$.

Solution: Note that

$$\begin{aligned} \frac{1}{2}\sum_{\boldsymbol{x}^n} \left| P_{X^n}(\boldsymbol{x}^n) - \hat{P}_{X^n}(\boldsymbol{x}^n) \right| &= \frac{1}{2} \sum_{\boldsymbol{x}^n \in \mathcal{X}^n \setminus \mathcal{A}_\epsilon^{(n)}(X)} \left| P_{X^n}(\boldsymbol{x}^n) - \hat{P}_{X^n}(\boldsymbol{x}^n) \right| \\ &\leq \frac{1}{2} \sum_{\boldsymbol{x}^n \in \mathcal{X}^n \setminus \mathcal{A}_\epsilon^{(n)}(X)} P_{X^n}(\boldsymbol{x}^n) + \hat{P}_{X^n}(\boldsymbol{x}^n) \\ &= 1 - P_{X^n}(\mathcal{A}_\epsilon^{(n)}(X)) \to 0 \end{aligned}$$

as $n \to \infty$ for any fixed $\epsilon > 0$. Thus, these must exist some $N \in \mathbb{N}$ such that $\delta_{\text{tvd}}(P_{X^n}, \hat{P}_{X^n}) \leq \epsilon$ for all $n \geq N$.

b.) Using $\hat{P}_{X^n}$ from the previous step, show that

$$\frac{1}{n} H_{\min}^{\epsilon}(X^n) \geq H(X) - \epsilon$$

for n large enough.

Solution: If $\mathcal{X}^n \setminus \mathcal{A}_\epsilon^{(n)}(X) = \emptyset$,

$$\max_{\boldsymbol{x}^n} \hat{P}_{X^n}(\boldsymbol{x}^n) = \max_{\boldsymbol{x}^n \in \mathcal{A}_\epsilon^{(n)}(X)} P_{X^n}(\boldsymbol{x}^n) \leq 2^{-n(H(X)-\epsilon)}.$$

If $\mathcal{X}^n \setminus \mathcal{A}_\epsilon^{(n)}(X) \neq \emptyset$,

$$\frac{1 - P_{X^n}(\mathcal{A}_\epsilon^{(n)}(X))}{|\mathcal{X}|^n - \left|\mathcal{A}_\epsilon^{(n)}(X)\right|} \leq 1 - P_{X^n}(\mathcal{A}_\epsilon^{(n)}(X)) \leq 2^{-n(H(X)-\epsilon)}$$

as $n \to \infty$. Thus, for n large enough,

$$\begin{aligned} \frac{1}{n} H_{\min}^{\epsilon}(X) &= \frac{1}{n} \max_{\delta_{\text{tvd}}(P_X,\tilde{P}_X)\leq\epsilon} H_{\min}(X)_{\tilde{P}_X} \\ &\geq \frac{1}{n} H_{\min}(X)_{\hat{P}_X} = -\frac{1}{n} \log 2^{-n(H(X)-\epsilon)} = H(X) - \epsilon. \end{aligned}$$

c.) **Continuity of entropy.** For any two distributions P_X, $\tilde{P}_X$ (on the same set) such that $\delta_{\text{tvd}}(P_X, \tilde{P}_X) \leq \epsilon$, prove that for all $\epsilon < 1/2$

$$\left| H(X)_{P_X} - H(X)_{\tilde{P}_X} \right| \leq h(\epsilon) + \epsilon \cdot \log |\mathcal{X}|.$$

where $h : t \mapsto -t \log t - (1-t) \log (1-t)$. **Hint:** Construct a pair of joint random variables (X_1, X_2) with marginals $P_{X_1} = P_X$, $P_{X_2} = \tilde{P}_X$, and $\mathrm{P}[X_1 = X_2]$ large. Use Fano's inequality.

Solution: Define the non-negative functions

$$\begin{aligned} f_{\min} &: x \mapsto \min\{P_X(x), \tilde{P}_X(x)\}, \\ f_+ &: x \mapsto P_X(x) - f_{\min}(x), \\ f_- &: x \mapsto \tilde{P}_X(x) - f_{\min}(x). \end{aligned}$$

Notice that $\sum_x f_+(x) = \sum_x f_-(x) = 1 - \sum_x f_{\min}(x)$. Denote this value by α. Let (X_1, X_2) be a pair of joint random variables (distributing on $\mathcal{X}^2$) with pmf

$$P_{X_1,X_2}(x_1, x_2) = f_{\min}(x_1) \cdot \delta_{x_1,x_2} + \alpha^{-1} \cdot f_+(x_1) \cdot f_-(x_2).$$

This is a valid pmf since: First, it is is obviously non-negative; Second, its marginals match P_X and $\tilde{P}_X$:

$$\begin{aligned} P_{X_1}(x_1) &= f_{\min}(x_1) + f_+(x_1) \cdot \sum_{x_2} \alpha^{-1} \cdot f_-(x_2) = f_{\min}(x_1) + f_+(x_1) = P_X(x_1), \\ P_{X_2}(x_2) &= f_{\min}(x_2) + f_+(x_2) \cdot \sum_{x_1} \alpha^{-1} \cdot f_+(x_1) = f_{\min}(x_2) + f_-(x_2) = \tilde{P}_X(x_2). \end{aligned}$$

Notice that

$$\begin{aligned} \mathrm{P}[X_1 \neq X_2] &= \sum_{x_1 \neq x_2} P_{X_1,X_2}(x_1, x_2) = \sum_{x_1,x_2} \alpha^{-1} \cdot f_+(x_1) \cdot f_-(x_2) = \sum_x f_+(x) \\ &= \sum_{x : P_X(x) > \tilde{P}_X(x)} P_x(x) - \tilde{P}_X(x) = \frac{1}{2} \sum_{x \in \mathcal{X}} \left| P_x(x) - \tilde{P}_X(x) \right| = \delta_{\mathrm{tvd}}(P_X, \tilde{P}_X) \leq \epsilon. \end{aligned}$$

Without loss of generality, we assume $H(X_1) \geq H(X_2)$. In this case, by Fano's inequality,

$$0 \leq H(X_1) - H(X_2) \leq H(X_1 X_2) - H(X_2) = H(X_1|X_2) \leq h(\epsilon) + \epsilon \log(|\mathcal{X}| - 1).$$

Since $\log(|\mathcal{X}| - 1) \leq \log |\mathcal{X}|$, we have

$$|H(X_1) - H(X_2)| \leq h(\epsilon) + \epsilon \cdot \log |\mathcal{X}|.$$

d.) Using the previous step, show that

$$\frac{1}{n} H^{\epsilon}_{\min}(X^n) \leq H(X) + \frac{1}{n} h(\epsilon) + \epsilon \cdot \log |\mathcal{X}|$$

for $0 < \epsilon < 1/2$.

Solution: Suppose $H^{\epsilon}_{\min}(X^n) = H_{\min}(X^n)_{P^{\star}_{X^n}}$, where $\delta_{\mathrm{tvd}}(P_X, P^{\star}_X) \leq \epsilon$. Then,

$$\begin{aligned} \frac{1}{n} H^{\epsilon}_{\min}(X^n) = \frac{1}{n} H_{\min}(X^n)_{P^{\star}_{X^n}} \leq \frac{1}{n} H(X^n)_{P^{\star}_{X^n}} &\leq \frac{1}{n} \left(H(X^n) + h(\epsilon) + \epsilon \cdot \log |\mathcal{X}|^n \right) \\ &= H(X) + \frac{1}{n} \cdot \left(h(\epsilon) + \epsilon \cdot n \log |\mathcal{X}| \right). \end{aligned}$$