Solution: f-Divergence and Hypothesis Testing

Problem 4: f-Divergence and Hypothesis Testing

We address each part in detail, including all intermediate steps.

1. Convexity of $f_\alpha(x)$

Definition 1 (Convex Function). *A twice-differentiable function $g : (0, \infty) \to \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x > 0$.*

Let $\alpha > 0$, $\alpha \neq 1$, and define

$$f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}, \quad x > 0.$$

Compute the first and second derivatives explicitly:

$$f'_\alpha(x) = \frac{d}{dx}\left(\tfrac{x^\alpha - 1}{\alpha - 1}\right) = \frac{\alpha x^{\alpha-1}}{\alpha - 1},$$

$$f''_\alpha(x) = \frac{d}{dx}\left(\frac{\alpha x^{\alpha-1}}{\alpha - 1}\right) = \frac{\alpha(\alpha - 1)x^{\alpha-2}}{\alpha - 1} = \alpha x^{\alpha-2}.$$

Since $\alpha > 0$ and $x^{\alpha-2} > 0$ for $x > 0$, we have $f''_\alpha(x) > 0$. By definition, f_α is strictly convex.

2. Definition of the f-divergence

Let $P = (p_1, \ldots, p_n)$ and $Q = (q_1, \ldots, q_n)$ be two probability distributions on a finite set $\Omega = \{1, 2, \ldots, n\}$, with $q_i > 0$ for all i.

Definition 2 (f-Divergence). *Let $f : (0, \infty) \to \mathbb{R}$ be convex with $f(1) = 0$. The f-divergence of P relative to Q is*

$$D_f(P\|Q) = \sum_{i=1}^{n} q_i\, f\left(\frac{p_i}{q_i}\right).$$

In particular, for $f = f_\alpha$, we recover the α-divergence:

$$D_{f_\alpha}(P\|Q) = \sum_{i=1}^{n} q_i \frac{(p_i/q_i)^\alpha - 1}{\alpha - 1}.$$

3. Data-Processing Inequality for f-Divergence

Suppose we apply any randomized transformation T to the sample space, inducing a Markov kernel $T(j \mid i)$ with output alphabet Γ. Then the output distributions are

$$P_T(j) = \sum_{i=1}^{n} p_i \, T(j \mid i), \quad Q_T(j) = \sum_{i=1}^{n} q_i \, T(j \mid i).$$

Theorem 1 (Data-Processing Inequality)**.** *For any convex f and any Markov kernel T,*

$$D_f(P_T \| Q_T) \le D_f(P \| Q).$$

Proof. Write

$$D_f(P_T \| Q_T) = \sum_{j \in \Gamma} Q_T(j) f\Big(\tfrac{P_T(j)}{Q_T(j)}\Big).$$

For each j, define weights $w_i = T(j \mid i) q_i / Q_T(j)$ (so $\sum_i w_i = 1$). Then

$$\frac{P_T(j)}{Q_T(j)} = \frac{\sum_i p_i T(j \mid i)}{Q_T(j)} = \sum_i w_i \frac{p_i}{q_i}.$$

By convexity of f,

$$f\Big(\sum_i w_i \frac{p_i}{q_i}\Big) \le \sum_i w_i f\Big(\frac{p_i}{q_i}\Big).$$

Multiplying both sides by $Q_T(j)$ and summing over j gives

$$D_f(P_T \| Q_T) \le \sum_j \sum_i T(j \mid i) q_i f\Big(\frac{p_i}{q_i}\Big) = \sum_i q_i f\Big(\frac{p_i}{q_i}\Big) = D_f(P \| Q).$$

$\square$

4. Data-Processing for Rényi α-Divergence

Define the Rényi α-divergence for $\alpha > 0$, $\alpha \neq 1$, by

$$R_\alpha(P \| Q) = \frac{1}{1-\alpha} \ln\Big(1 + (\alpha - 1) D_{f_\alpha}(P \| Q)\Big).$$

Note that the function $g(u) = \frac{1}{1-\alpha} \ln(1 + (\alpha - 1)u)$ is strictly increasing in u since its derivative is positive. Hence, applying the monotone mapping g to both sides of the data-processing inequality for f_α yields

$$R_\alpha(P_T \| Q_T) = g\big(D_{f_\alpha}(P_T \| Q_T)\big) \le g\big(D_{f_\alpha}(P \| Q)\big) = R_\alpha(P \| Q).$$

Thus Rényi divergence also satisfies data-processing.

5. Limits as $\alpha \to 1$

We show

$$\lim_{\alpha\to 1} D_{f_\alpha}(P\|Q) = \lim_{\alpha\to 1} R_\alpha(P\|Q) = D_{\mathrm{KL}}(P\|Q).$$

Proof. Using the Taylor expansion for x^α around $\alpha = 1$:

$$x^\alpha = e^{\alpha \ln x} = 1 + (\alpha - 1)\ln x + o(\alpha - 1),$$

it follows that

$$f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1} = \ln x + o(1).$$

Therefore,

$$D_{f_\alpha}(P\|Q) = \sum_i q_i\big(\ln(p_i/q_i) + o(1)\big) = D_{\mathrm{KL}}(P\|Q) + o(1).$$

Next, since $D_{f_\alpha} \to D_{\mathrm{KL}}$, we have

$$R_\alpha(P\|Q) = \frac{\ln\big(1 + (\alpha-1)D_{f_\alpha}\big)}{1-\alpha} = \frac{(\alpha-1)D_{f_\alpha} + o(\alpha-1)}{1-\alpha} = D_{f_\alpha} + o(1) \to D_{\mathrm{KL}}(P\|Q).$$

□

6. Hypothesis Testing with Rényi Divergence

Let $X^n = (X_1, \ldots, X_n)$ be i.i.d. with distribution either Q_X under H_0 or P_X under H_1. A stochastic test outputs a binary decision $Z \in \{0, 1\}$ with

$$P_{H_1}(Z = 0) = \alpha_n < \epsilon, \quad P_{H_0}(Z = 1) = \beta_n.$$

Then P–error (α_n) and Q–error (β_n) satisfy by data-processing:

$$R_\alpha(P_X^{\otimes n}\|Q_X^{\otimes n}) \geq R_\alpha(\mathrm{Bern}(\alpha_n)\|\mathrm{Bern}(\beta_n)).$$

Since Rényi divergence is additive over product distributions,

$$R_\alpha(P_X^{\otimes n}\|Q_X^{\otimes n}) = n\, R_\alpha(P_X\|Q_X).$$

Hence

$$nR_\alpha(P_X\|Q_X) \geq R_\alpha(\mathrm{Bern}(\alpha_n)\|\mathrm{Bern}(\beta_n)).$$

7. Bounding Type-II Exponent

Now let $\alpha_n < \epsilon = 1/3$, and set $\alpha = 1 + 1/\sqrt{n}$. We use the bound

$$R_\alpha(\mathrm{Bern}(\epsilon)\|\mathrm{Bern}(1-\beta_n)) \approx (1-\alpha)\ln(1-\beta_n) - h(\epsilon)$$

for large n, where $h(\epsilon)$ is the binary entropy. Rearranging yields

$$\ln(1-\beta_n) \leq -\tfrac{1}{\alpha-1}\big(nR_\alpha(P_X\|Q_X) + h(\epsilon)\big) \approx -\sqrt{n}\, R_\alpha(P_X\|Q_X),$$

implying

$$\beta_n \leq e^{-\sqrt{n}\, R_\alpha(P_X\|Q_X)}.$$

Thus for $\beta_n = e^{-nE_n}$, we have

$$E_n \leq R_{1+1/\sqrt{n}}(P_X\|Q_X).$$

8. Conclusion: KL Bound on Error Exponent

Taking $n \to \infty$, since $\alpha = 1 + 1/\sqrt{n} \to 1$, and using

$$\lim_{n\to\infty} R_{1+1/\sqrt{n}}(P_X \| Q_X) = D_{\mathrm{KL}}(P_X \| Q_X),$$

we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \ln \frac{1}{\beta_n} \leq D_{\mathrm{KL}}(P_X \| Q_X).$$

This completes the detailed proof.