Problem Set Solutions

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

We want to show that the function $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}$ is convex on $\mathbb{R}^+$ for $\alpha > 0$. We compute the second derivative $f''_\alpha(x)$.

Case 1: $\alpha > 0, \alpha \neq 1$. The first derivative is:

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

The second derivative is:

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

Since $\alpha > 0$ and $x \in \mathbb{R}^+$ (i.e., $x > 0$), we have $x^{\alpha-2} > 0$. Therefore, $f''_\alpha(x) = \alpha x^{\alpha-2} > 0$ for all $x > 0$.

Case 2: $\alpha = 1$. We find the limit of $f_\alpha(x)$ as $\alpha \to 1$ using L'Hopital's rule (differentiating numerator and denominator with respect to α):

$$f_1(x) = \lim_{\alpha\to 1} \frac{x^\alpha - 1}{\alpha - 1} = \lim_{\alpha\to 1} \frac{\frac{d}{d\alpha}(x^\alpha - 1)}{\frac{d}{d\alpha}(\alpha - 1)} = \lim_{\alpha\to 1} \frac{x^\alpha \ln x}{1} = x \ln x$$

Now, we check the convexity of $f_1(x) = x\ln x$. The first derivative is:

$$f'_1(x) = \frac{d}{dx}(x\ln x) = 1 \cdot \ln x + x \cdot \frac{1}{x} = \ln x + 1$$

The second derivative is:

$$f''_1(x) = \frac{d}{dx}(\ln x + 1) = \frac{1}{x}$$

For $x > 0$, $f''_1(x) = \frac{1}{x} > 0$.

In both cases, $f''_\alpha(x) > 0$ for $x > 0$. Thus, $f_\alpha(x)$ is strictly convex on $\mathbb{R}^+$ for all $\alpha > 0$.

2. Definition of f-Divergence

Let P and Q be probability distributions on a finite sample space $\Omega = \{1, 2, ..., n\}$, with probability mass functions $p_i = P(i)$ and $q_i = Q(i)$. Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a convex function. The f-divergence $D_f(P||Q)$ is defined as:

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

We adopt the conventions: $0f(0/0) = 0$, and $qf(p/q) = p\lim_{u\to\infty}\frac{f(u)}{u}$ if $q = 0, p > 0$. Assuming $P \ll Q$ (i.e., $p_i = 0$ whenever $q_i = 0$), the sum is effectively over i such that $q_i > 0$:

$$D_f(P||Q) = \sum_{i:q_i>0} q_i f\left(\frac{p_i}{q_i}\right)$$

3. Data Processing Inequality for f-Divergence

Let P, Q on Ω and P', Q' on Ω' be distributions related by a channel $K : \Omega \to \Omega'$, i.e., $p'(y) = \sum_{x\in\Omega} p(x)K(y|x)$ and $q'(y) = \sum_{x\in\Omega} q(x)K(y|x)$. We show $D_f(P'||Q') \leq D_f(P||Q)$ for convex f. Assume $P \ll Q$. If $q'(y) = 0$, then for all x with $q(x) > 0$, $K(y|x) = 0$. Since $P \ll Q$, $p(x) > 0$ only if $q(x) > 0$, so $K(y|x) = 0$ for $p(x) > 0$. Thus $p'(y) = 0$. The term $q'(y)f(p'(y)/q'(y))$ is 0. We sum over y where $q'(y) > 0$.

$$D_f(P'||Q') = \sum_{y:q'(y)>0} q'(y) f\left(\frac{p'(y)}{q'(y)}\right) = \sum_{y:q'(y)>0} q'(y) f\left(\frac{\sum_x p(x)K(y|x)}{\sum_x q(x)K(y|x)}\right)$$

Let $u_x = p(x)/q(x)$. $p(x) = q(x)u_x$.

$$D_f(P'||Q') = \sum_{y:q'(y)>0} q'(y) f\left(\frac{\sum_x q(x)K(y|x)u_x}{q'(y)}\right)$$

Define weights $w_{y,x} = \frac{q(x)K(y|x)}{q'(y)} \geq 0$. $\sum_x w_{y,x} = \frac{\sum_x q(x)K(y|x)}{q'(y)} = \frac{q'(y)}{q'(y)} = 1$. The expression inside f is $\sum_x w_{y,x}u_x$. By Jensen's inequality for convex f:

$$f\left(\sum_x w_{y,x}u_x\right) \leq \sum_x w_{y,x} f(u_x) = \sum_x \frac{q(x)K(y|x)}{q'(y)} f\left(\frac{p(x)}{q(x)}\right)$$

Substitute back:

$$D_f(P'||Q') \leq \sum_{y:q'(y)>0} q'(y) \left[\sum_x \frac{q(x)K(y|x)}{q'(y)} f\left(\frac{p(x)}{q(x)}\right)\right]$$

$$= \sum_{y:q'(y)>0} \sum_x q(x)K(y|x) f\left(\frac{p(x)}{q(x)}\right) = \sum_y \sum_x q(x)K(y|x) f\left(\frac{p(x)}{q(x)}\right)$$

$$= \sum_x q(x) f\left(\frac{p(x)}{q(x)}\right) \left(\sum_y K(y|x)\right)$$

Since $\sum_y K(y|x) = 1$,

$$D_f(P'||Q') \leq \sum_x q(x) f\left(\frac{p(x)}{q(x)}\right) = D_f(P||Q)$$

4. Rényi Divergence and Data Processing Inequality

The Rényi α-divergence (standard definition, using ln): $R_\alpha(P||Q) = \frac{1}{\alpha-1} \ln\left(\sum_i p_i^\alpha q_i^{1-\alpha}\right)$ for $\alpha > 0, \alpha \neq 1$. Let $f_\alpha(x) = \frac{x^\alpha - 1}{\alpha - 1}$.

$$D_{f_\alpha}(P||Q) = \sum_i q_i f_\alpha\left(\frac{p_i}{q_i}\right) = \sum_i q_i \frac{(p_i/q_i)^\alpha - 1}{\alpha - 1} = \frac{1}{\alpha - 1}\left(\sum_i p_i^\alpha q_i^{1-\alpha} - 1\right)$$

Let $S_\alpha(P||Q) = \sum_i p_i^\alpha q_i^{1-\alpha}$. Then $D_{f_\alpha}(P||Q) = \frac{S_\alpha(P||Q)-1}{\alpha-1}$. So $S_\alpha(P||Q) = 1 + (\alpha-1)D_{f_\alpha}(P||Q)$.

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

(This differs from the problem statement's $R_\alpha = \frac{1}{1-\alpha}\ln(1+(\alpha-1)D_{f_\alpha})$ which seems unconventional for $\alpha > 1$).

We show DPI for R_α. Let P', Q' be post-channel distributions. From Q1, f_α is convex. From Q3, D_{f_α} satisfies DPI: $D_{f_\alpha}(P'||Q') \le D_{f_\alpha}(P||Q)$. Let $D' = D_{f_\alpha}(P'||Q')$ and $D = D_{f_\alpha}(P||Q)$. Let $S'_\alpha = 1+(\alpha-1)D'$ and $S_\alpha = 1+(\alpha-1)D$. We have $D' \le D$.

Case 1: $\alpha > 1$. Then $\alpha - 1 > 0$. $D' \le D \implies S'_\alpha \le S_\alpha$. By Jensen's inequality, $S_\alpha \ge 1$. $g(S) = \frac{1}{\alpha-1}\ln S$ is increasing for $S \ge 1$. So $S'_\alpha \le S_\alpha \implies R_\alpha(P'||Q') \le R_\alpha(P||Q)$.

Case 2: $0 < \alpha < 1$. Then $\alpha - 1 < 0$. $D' \le D \implies S'_\alpha \ge S_\alpha$. By Jensen's inequality, $S_\alpha \le 1$. $g(S) = \frac{1}{\alpha-1}\ln S$ is decreasing for $S > 0$. So $S'_\alpha \ge S_\alpha \implies R_\alpha(P'||Q') \le R_\alpha(P||Q)$.

In both cases, DPI holds for R_α.

5. Limit of D_{f_α} and R_α as $\alpha \to 1$

Limit of $D_{f_\alpha}(P||Q)$:

$$\lim_{\alpha\to1} D_{f_\alpha}(P||Q) = \sum_i q_i \lim_{\alpha\to1} f_\alpha(p_i/q_i) = \sum_i q_i\left(\frac{p_i}{q_i}\ln\frac{p_i}{q_i}\right) = \sum_i p_i \ln\frac{p_i}{q_i} = D_{KL}(P||Q)$$

(using $\lim_{\alpha\to1} f_\alpha(x) = x\ln x$ from Q1).

Limit of $R_\alpha(P||Q)$:

$$R_\alpha(P||Q) = \frac{1}{\alpha-1}\ln\left(\sum_i p_i^\alpha q_i^{1-\alpha}\right) = \frac{\ln S(\alpha)}{\alpha-1}$$

As $\alpha \to 1$, $S(\alpha) \to \sum_i p_i = 1$, so we have $\frac{0}{0}$. Use L'Hopital's rule (differentiate w.r.t. α):

$$\lim_{\alpha\to1} R_\alpha(P||Q) = \lim_{\alpha\to1}\frac{S'(\alpha)/S(\alpha)}{1}$$

$$S'(\alpha) = \frac{d}{d\alpha}\sum_i p_i^\alpha q_i^{1-\alpha} = \sum_i p_i^\alpha q_i^{1-\alpha}\ln(p_i/q_i)$$

$$S'(1) = \sum_i p_i \ln(p_i/q_i) = D_{KL}(P||Q)$$

$$\lim_{\alpha\to1} R_\alpha(P||Q) = \frac{S'(1)/S(1)}{1} = \frac{D_{KL}(P||Q)}{1} = D_{KL}(P||Q)$$

Both limits are $D_{KL}(P||Q)$.

6. Rényi Divergence Bound for Hypothesis Testing

$H_0 : X^n \sim Q_X^n$, $H_1 : X^n \sim P_X^n$. Test $Z \in \{0, 1\}$. Type I error: $\epsilon_n = Q_X^n(Z = 1) \le \epsilon$. Type II error: $\delta_n = P_X^n(Z = 0) = e^{-nE_n}$. Let $P^n = P_X^n, Q^n = Q_X^n$. Let P_Z, Q_Z be distributions of Z. $Q_Z = \text{Ber}(\epsilon_n)$, $P_Z = \text{Ber}(1 - \delta_n)$. DPI for R_α: $R_\alpha(P^n||Q^n) \ge R_\alpha(P_Z||Q_Z)$. We showed $R_\alpha(P^n||Q^n) = nR_\alpha(P_X||Q_X)$.

$$nR_\alpha(P_X||Q_X) \ge R_\alpha(\text{Ber}(1 - \delta_n)||\text{Ber}(\epsilon_n))$$

The problem statement asks to show $nR_\alpha(P_X||Q_X) \ge R_\alpha(\text{Ber}(\epsilon)||\text{Ber}(1 - e^{-nE_n}))$. This seems non-standard; it swaps the roles of P_Z, Q_Z arguments and potentially uses the bound ϵ instead of ϵ_n. Let's assume the problem meant the inequality derived from standard DPI: $nR_\alpha(P_X||Q_X) \ge R_\alpha(\text{Ber}(1 - e^{-nE_n})||\text{Ber}(\epsilon_n))$.

7. Bound on Error Exponent E_n

Given $\epsilon = 1/2$, so $\epsilon_n \le 1/2$. Let $\alpha = 1 + 1/\sqrt{n}$. Let $h = 1/\sqrt{n}$. Use the inequality from Q6 derived from standard DPI:

$$nR_{1+h}(P_X||Q_X) \ge R_{1+h}(\text{Ber}(1 - \delta_n)||\text{Ber}(\epsilon_n))$$

$$R_{1+h}(\text{Ber}(1 - \delta_n)||\text{Ber}(\epsilon_n)) = \frac{1}{h} \ln \left((1 - \delta_n)^{1+h}\epsilon_n^{-h} + \delta_n^{1+h}(1 - \epsilon_n)^{-h}\right)$$

Substitute $\delta_n = e^{-nE_n}$.

$$RHS = \frac{1}{h} \ln \left((1 - e^{-nE_n})^{1+h}\epsilon_n^{-h} + (e^{-nE_n})^{1+h}(1 - \epsilon_n)^{-h}\right)$$

Assume $E_n \to E > 0$. Then $e^{-nE_n} \to 0$. $(1 - e^{-nE_n})^{1+h} \to 1$. $e^{-nE_n(1+h)} = e^{-(n+\sqrt{n})E_n}$ is very small.

$$RHS = \frac{1}{h} \ln \left((1 - o(1))\epsilon_n^{-h} + e^{-(n+\sqrt{n})E_n}(1 - \epsilon_n)^{-h}\right)$$

$$\approx \frac{1}{h} \ln(\epsilon_n^{-h}) = -\ln \epsilon_n$$

This approximation ignores E_n. Let's try the inequality stated in the problem:

$$nR_{1+h}(P_X||Q_X) \ge R_{1+h}(\text{Ber}(\epsilon_n)||\text{Ber}(1 - \delta_n))$$

$$R_{1+h}(\text{Ber}(\epsilon_n)||\text{Ber}(1-\delta_n)) = \frac{1}{h} \ln(\epsilon_n^{1+h}(1-(1-\delta_n))^{-h} + (1-\epsilon_n)^{1+h}(1-\delta_n)^{-h})$$

$$= \frac{1}{h} \ln(\epsilon_n^{1+h}\delta_n^{-h} + (1 - \epsilon_n)^{1+h}(1 - \delta_n)^{-h})$$

$$= \frac{1}{h} \ln(\epsilon_n^{1+h}e^{nhE_n} + (1 - \epsilon_n)^{1+h}(1 - e^{-nE_n})^{-h})$$

For large n, $(1-e^{-nE_n})^{-h} \approx 1$. $nhE_n = \sqrt{n}E_n$.

$$RHS \approx \frac{1}{h}\ln(\epsilon_n^{1+h}e^{\sqrt{n}E_n} + (1-\epsilon_n)^{1+h})$$

$$= \frac{1}{h}\ln[(1-\epsilon_n)^{1+h}(1+(\frac{\epsilon_n}{1-\epsilon_n})^{1+h}e^{\sqrt{n}E_n})]$$

Since $\epsilon_n \leq 1/2$, $\epsilon_n/(1-\epsilon_n) \leq 1$. The term $x = (\frac{\epsilon_n}{1-\epsilon_n})^{1+h}e^{\sqrt{n}E_n}$ grows exponentially if $E > 0$. Use $\ln(1+x) \approx \ln x$.

$$RHS \approx \frac{1}{h}\ln[(1-\epsilon_n)^{1+h}(\frac{\epsilon_n}{1-\epsilon_n})^{1+h}e^{\sqrt{n}E_n}]$$

$$= \frac{1}{h}\ln[\epsilon_n^{1+h}e^{\sqrt{n}E_n}] = \frac{1+h}{h}\ln\epsilon_n + \frac{\sqrt{n}E_n}{h} = (\sqrt{n}+1)\ln\epsilon_n + nE_n$$

So, $nR_{1+1/\sqrt{n}}(P_X||Q_X) \geq nE_n + (\sqrt{n}+1)\ln\epsilon_n$.

$$E_n \leq R_{1+1/\sqrt{n}}(P_X||Q_X) - \frac{\sqrt{n}+1}{n}\ln\epsilon_n$$

Since $\epsilon_n \leq 1/2$, $\ln\epsilon_n < 0$. Let $C_n = -\ln\epsilon_n \geq \ln 2$.

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

The term $\frac{\sqrt{n}+1}{n}C_n = O(1/\sqrt{n})$. Thus $E_n \leq R_{1+1/\sqrt{n}}(P_X||Q_X) + o(1)$. For large enough n, this means E_n is upper bounded by $R_{1+1/\sqrt{n}}(P_X||Q_X)$, potentially with a small positive term. The question likely asks for the leading term behavior.

8. Conclusion on the Error Exponent Limit

We want to show $\lim_{n\to\infty}\frac{1}{n}\log\frac{1}{\beta_{1-\epsilon}(P,Q)} \leq D_{KL}(P||Q)$. Let $\delta_n^* = \beta_{1-\epsilon}(P^n, Q^n)$ be the minimum type II error for $\epsilon_n \leq \epsilon$. Let E_n^* be the corresponding exponent, $\delta_n^* = e^{-nE_n^*}$. From Q7, using the inequality assumed there, for any test with $\epsilon_n \leq \epsilon = 1/2$,

$$E_n \leq R_{1+1/\sqrt{n}}(P_X||Q_X) + \frac{\sqrt{n}+1}{n}(-\ln\epsilon_n)$$

This holds for the optimal test sequence with error $(\epsilon_n^*, \delta_n^*)$ and exponent E_n^*:

$$E_n^* \leq R_{1+1/\sqrt{n}}(P_X||Q_X) + \frac{\sqrt{n}+1}{n}(-\ln\epsilon_n^*)$$

Since $\epsilon_n^* \leq \epsilon = 1/2$, $(-\ln\epsilon_n^*)$ is positive and bounded below by $\ln 2$. The term $\frac{\sqrt{n}+1}{n}(-\ln\epsilon_n^*)$ is $O(1/\sqrt{n})$.

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

Taking the limit superior as $n \to \infty$:

$$\limsup_{n\to\infty} E_n^* \le \limsup_{n\to\infty} \left(R_{1+1/\sqrt{n}}(P_X||Q_X) + O(1/\sqrt{n})\right)$$

As $n \to \infty$, $\alpha = 1+1/\sqrt{n} \to 1$. From Q5, $\lim_{\alpha\to 1} R_\alpha(P_X||Q_X) = D_{KL}(P_X||Q_X)$. Assuming continuity,

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

The $O(1/\sqrt{n})$ term vanishes.

$$\limsup_{n\to\infty} E_n^* \le D_{KL}(P_X||Q_X)$$

The error exponent is $E = \lim_{n\to\infty} -\frac{1}{n}\log \beta_{1-\epsilon}(P^n, Q^n) = \lim_{n\to\infty} E_n^*$ (if the limit exists). If the limit exists, $E \le D_{KL}(P_X||Q_X)$. If not, the $\limsup$ result holds. The problem asks to conclude $\lim_{n\to\infty} \frac{1}{n}\log\frac{1}{\beta_{1-\epsilon}(P,Q)} \le D_{KL}(P||Q)$, which is $E \le D_{KL}(P_X||Q_X)$ (using P, Q for P_X, Q_X).