# Solution 1

The $\chi^2$-divergence is defined as $D_{\chi^2}(P||Q) = \sum_x \frac{(P(x)-Q(x))^2}{Q(x)} = \sum_x Q(x)\left(\frac{P(x)}{Q(x)} - 1\right)^2$. This corresponds to the f-divergence with $f(t) = (t-1)^2$. The chain rule for f-divergences states that for any $P_{XY}$ and $Q_{XY}$:

$$D_f(P_{XY}||Q_{XY}) = D_f(P_X||Q_X) + \sum_x P_X(x)D_f(P_{Y|X=x}||Q_{Y|X=x})$$

Let $Q_{XY} = Q_X Q_Y$, meaning $Q_{XY}(x,y) = Q_X(x)Q_Y(y)$. The marginal $Q_X(x) = \sum_y Q_{XY}(x,y) = Q_X(x)\sum_y Q_Y(y) = Q_X(x)$. The conditional $Q_{Y|X=x}(y) = Q_{XY}(x,y)/Q_X(x) = Q_X(x)Q_Y(y)/Q_X(x) = Q_Y(y)$. Applying the chain rule for $D_{\chi^2}$ with $Q_{XY} = Q_X Q_Y$:

$$D_{\chi^2}(P_{XY}||Q_X Q_Y) = D_{\chi^2}(P_X||Q_X) + \sum_x P_X(x)D_{\chi^2}(P_{Y|X=x}||Q_Y)$$

We need to prove $D_{\chi^2}(P_{XY}||Q_X Q_Y) \geq D_{\chi^2}(P_X||Q_X) + D_{\chi^2}(P_Y||Q_Y)$. This is equivalent to showing:

$$\sum_x P_X(x)D_{\chi^2}(P_{Y|X=x}||Q_Y) \geq D_{\chi^2}(P_Y||Q_Y)$$

Let $h(P) = D_{\chi^2}(P||Q_Y)$, where $P$ is a probability distribution on the alphabet of $Y$. The function $f(t) = (t-1)^2$ is convex. An f-divergence $D_f(P||Q)$ is convex in $P$ for a fixed $Q$. Thus, $h(P)$ is convex in $P$. We know that the marginal distribution $P_Y$ is a mixture of the conditional distributions $P_{Y|X=x}$:

$$P_Y(y) = \sum_x P_X(x)P_{Y|X=x}(y)$$

Let $P_x = P_{Y|X=x}$. Then $P_Y = \sum_x P_X(x)P_x$. By Jensen's inequality for the convex function $h(P)$:

$$\sum_x P_X(x)h(P_{Y|X=x}) \geq h\left(\sum_x P_X(x)P_{Y|X=x}\right)$$

Substituting the definition of $h$:

$$\sum_x P_X(x)D_{\chi^2}(P_{Y|X=x}||Q_Y) \geq D_{\chi^2}\left(\sum_x P_X(x)P_{Y|X=x}||Q_Y\right)$$

Since $\sum_x P_X(x)P_{Y|X=x} = P_Y$, we have:

$$\sum_x P_X(x)D_{\chi^2}(P_{Y|X=x}||Q_Y) \geq D_{\chi^2}(P_Y||Q_Y)$$

This completes the proof.

# Solution 2

We are given $P_{XYZ}(x,y,z) = P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x)$. This means $Y$ and $Z$ are conditionally independent given $X$. We need to prove the subadditivity property (2) for the Kullback-Leibler (KL) divergence:

$$D_{KL}(P_{XYZ}||P_X P_Y P_Z) \leq D_{KL}(P_{XY}||P_X P_Y) + D_{KL}(P_{XZ}||P_X P_Z)$$

Here, $P_X P_Y P_Z$ denotes the product of marginal distributions $Q(x,y,z) = P_X(x)P_Y(y)P_Z(z)$. The KL divergence is defined as $D_{KL}(P||Q) = \sum_w P(w)\log\frac{P(w)}{Q(w)}$.

Left Hand Side (LHS):

$$D_{KL}(P_{XYZ}||P_X P_Y P_Z) = \sum_{x,y,z} P_{XYZ}(x,y,z)\log\frac{P_{XYZ}(x,y,z)}{P_X(x)P_Y(y)P_Z(z)}$$

Using $P_{XYZ}(x,y,z) = P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x)$:

$$LHS = \sum_{x,y,z} P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x)\log\frac{P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x)}{P_X(x)P_Y(y)P_Z(z)}$$

$$= \sum_{x,y,z} P_{XYZ}(x,y,z) \left( \log \frac{P_{Y|X}(y|x)}{P_Y(y)} + \log \frac{P_{Z|X}(z|x)}{P_Z(z)} \right)$$

We can split the sum:

$$= \sum_{x,y,z} P_{XYZ}(x,y,z) \log \frac{P_{Y|X}(y|x)}{P_Y(y)} + \sum_{x,y,z} P_{XYZ}(x,y,z) \log \frac{P_{Z|X}(z|x)}{P_Z(z)}$$

Summing over $z$ in the first term gives $P_{XY}(x,y)$, and summing over $y$ in the second term gives $P_{XZ}(x,z)$:

$$LHS = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{Y|X}(y|x)}{P_Y(y)} + \sum_{x,z} P_{XZ}(x,z) \log \frac{P_{Z|X}(z|x)}{P_Z(z)}$$

Right Hand Side (RHS): The marginal distributions are $P_{XY}(x,y) = P_X(x)P_{Y|X}(y|x)$ and $P_{XZ}(x,z) = P_X(x)P_{Z|X}(z|x)$. The first term of RHS is:

$$D_{KL}(P_{XY}||P_X P_Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$$

$$= \sum_{x,y} P_X(x)P_{Y|X}(y|x) \log \frac{P_X(x)P_{Y|X}(y|x)}{P_X(x)P_Y(y)} = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{Y|X}(y|x)}{P_Y(y)}$$

This term is the mutual information $I(X;Y)$. The second term of RHS is:

$$D_{KL}(P_{XZ}||P_X P_Z) = \sum_{x,z} P_{XZ}(x,z) \log \frac{P_{XZ}(x,z)}{P_X(x)P_Z(z)}$$

$$= \sum_{x,z} P_X(x)P_{Z|X}(z|x) \log \frac{P_X(x)P_{Z|X}(z|x)}{P_X(x)P_Z(z)} = \sum_{x,z} P_{XZ}(x,z) \log \frac{P_{Z|X}(z|x)}{P_Z(z)}$$

This term is the mutual information $I(X;Z)$. So, RHS $= I(X;Y) + I(X;Z)$.

Comparing LHS and RHS: We found $LHS = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{Y|X}(y|x)}{P_Y(y)} + \sum_{x,z} P_{XZ}(x,z) \log \frac{P_{Z|X}(z|x)}{P_Z(z)}$. We found $RHS = I(X;Y) + I(X;Z) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{Y|X}(y|x)}{P_Y(y)} + \sum_{x,z} P_{XZ}(x,z) \log \frac{P_{Z|X}(z|x)}{P_Z(z)}$. Thus, $LHS = RHS$. Since equality holds, the inequality $LHS \le RHS$ holds. The KL divergence is subadditive in the sense of (2) (and in fact, satisfies it with equality).

## Solution 3

The symmetric KL divergence is defined by the function $f_{SKL}(x) = (x-1)\log x$. The divergence is $D_{SKL}(P||Q) = D_{f_{SKL}}(P||Q) = \sum_x Q(x)f_{SKL}(P(x)/Q(x))$.

$$D_{SKL}(P||Q) = \sum_x Q(x) \left( \frac{P(x)}{Q(x)} - 1 \right) \log \frac{P(x)}{Q(x)} = \sum_x (P(x) - Q(x)) \log \frac{P(x)}{Q(x)}$$

$$= \sum_x P(x) \log \frac{P(x)}{Q(x)} - \sum_x Q(x) \log \frac{P(x)}{Q(x)} = D_{KL}(P||Q) + D_{KL}(Q||P)$$

We need to show that $D_{SKL}(P_{XYZ}||P_X P_Y P_Z) = D_{SKL}(P_{XY}||P_X P_Y) + D_{SKL}(P_{XZ}||P_X P_Z)$, assuming $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$. Let $P = P_{XYZ}$ and $Q = P_X P_Y P_Z$.

LHS: $D_{SKL}(P||Q) = D_{KL}(P||Q) + D_{KL}(Q||P)$. From problem 2, we know $D_{KL}(P||Q) = I(X;Y) + I(X;Z)$. Let's compute $D_{KL}(Q||P)$:

$$D_{KL}(P_X P_Y P_Z||P_{XYZ}) = \sum_{x,y,z} P_X(x)P_Y(y)P_Z(z) \log \frac{P_X(x)P_Y(y)P_Z(z)}{P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x)}$$

$$= \sum_{x,y,z} P_X(x)P_Y(y)P_Z(z) \left( \log \frac{P_Y(y)}{P_{Y|X}(y|x)} + \log \frac{P_Z(z)}{P_{Z|X}(z|x)} \right)$$

$$= \sum_x P_X(x) \sum_y P_Y(y) \log \frac{P_Y(y)}{P_{Y|X}(y|x)} \sum_z P_Z(z) + \sum_x P_X(x) \sum_z P_Z(z) \log \frac{P_Z(z)}{P_{Z|X}(z|x)} \sum_y P_Y(y)$$

Since $\sum_z P_Z(z) = 1$ and $\sum_y P_Y(y) = 1$:

$$D_{KL}(Q||P) = \sum_{x,y} P_X(x) P_Y(y) \log \frac{P_Y(y)}{P_{Y|X}(y|x)} + \sum_{x,z} P_X(x) P_Z(z) \log \frac{P_Z(z)}{P_{Z|X}(z|x)}$$

$$= \sum_x P_X(x) D_{KL}(P_Y||P_{Y|X=x}) + \sum_x P_X(x) D_{KL}(P_Z||P_{Z|X=x})$$

Where $D_{KL}(P_Y||P_{Y|X=x})$ denotes $\sum_y P_Y(y) \log \frac{P_Y(y)}{P_{Y|X}(y|x)}$. So, LHS $= I(X;Y) + I(X;Z) + \sum_x P_X(x) D_{KL}(P_Y||P_{Y|X=x}) + \sum_x P_X(x) D_{KL}(P_Z||P_{Z|X=x})$.

RHS: $D_{SKL}(P_{XY}||P_X P_Y) + D_{SKL}(P_{XZ}||P_X P_Z)$.

$$D_{SKL}(P_{XY}||P_X P_Y) = D_{KL}(P_{XY}||P_X P_Y) + D_{KL}(P_X P_Y||P_{XY})$$

We know $D_{KL}(P_{XY}||P_X P_Y) = I(X;Y)$.

$$D_{KL}(P_X P_Y||P_{XY}) = \sum_{x,y} P_X(x) P_Y(y) \log \frac{P_X(x) P_Y(y)}{P_{XY}(x,y)}$$

$$= \sum_{x,y} P_X(x) P_Y(y) \log \frac{P_X(x) P_Y(y)}{P_X(x) P_{Y|X}(y|x)} = \sum_{x,y} P_X(x) P_Y(y) \log \frac{P_Y(y)}{P_{Y|X}(y|x)}$$

$$= \sum_x P_X(x) D_{KL}(P_Y||P_{Y|X=x})$$

So, $D_{SKL}(P_{XY}||P_X P_Y) = I(X;Y) + \sum_x P_X(x) D_{KL}(P_Y||P_{Y|X=x})$.

Similarly,

$$D_{SKL}(P_{XZ}||P_X P_Z) = D_{KL}(P_{XZ}||P_X P_Z) + D_{KL}(P_X P_Z||P_{XZ})$$

We know $D_{KL}(P_{XZ}||P_X P_Z) = I(X;Z)$.

$$D_{KL}(P_X P_Z||P_{XZ}) = \sum_{x,z} P_X(x) P_Z(z) \log \frac{P_X(x) P_Z(z)}{P_{XZ}(x,z)}$$

$$= \sum_{x,z} P_X(x) P_Z(z) \log \frac{P_X(x) P_Z(z)}{P_X(x) P_{Z|X}(z|x)} = \sum_{x,z} P_X(x) P_Z(z) \log \frac{P_Z(z)}{P_{Z|X}(z|x)}$$

$$= \sum_x P_X(x) D_{KL}(P_Z||P_{Z|X=x})$$

So, $D_{SKL}(P_{XZ}||P_X P_Z) = I(X;Z) + \sum_x P_X(x) D_{KL}(P_Z||P_{Z|X=x})$.

Therefore, RHS $= [I(X;Y) + \sum_x P_X(x) D_{KL}(P_Y||P_{Y|X=x})] + [I(X;Z) + \sum_x P_X(x) D_{KL}(P_Z||P_{Z|X=x})]$. Comparing LHS and RHS, we see they are equal. Thus, the symmetric KL divergence satisfies (2) with equality.

## Solution 4

The squared Hellinger distance is $H^2(P||Q) = \frac{1}{2} \sum_w (\sqrt{P(w)} - \sqrt{Q(w)})^2 = 1 - \sum_w \sqrt{P(w)Q(w)}$. Let $A(P||Q) = \sum_w \sqrt{P(w)Q(w)}$ be the affinity between $P$ and $Q$. Then $H^2(P||Q) = 1 - A(P||Q)$. We are given $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$ and $Q_{XYZ} = Q_X Q_{Y|X} Q_{Z|X}$. We need to prove $H^2(P_{XYZ}||Q_{XYZ}) \le H^2(P_{XY}||Q_{XY}) + H^2(P_{XZ}||Q_{XZ})$. This is equivalent to proving:

$$1 - A(P_{XYZ}||Q_{XYZ}) \le (1 - A(P_{XY}||Q_{XY})) + (1 - A(P_{XZ}||Q_{XZ}))$$

$$1 - A(P_{XYZ}||Q_{XYZ}) \le 2 - A(P_{XY}||Q_{XY}) - A(P_{XZ}||Q_{XZ})$$

$$A(P_{XY}||Q_{XY}) + A(P_{XZ}||Q_{XZ}) - 1 \le A(P_{XYZ}||Q_{XYZ})$$

Let's compute the affinities.

$$A(P_{XYZ}||Q_{XYZ}) = \sum_{x,y,z} \sqrt{P_{XYZ}(x,y,z)Q_{XYZ}(x,y,z)}$$

$$= \sum_{x,y,z} \sqrt{P_X P_{Y|X} P_{Z|X} Q_X Q_{Y|X} Q_{Z|X}}$$

$$= \sum_x \sqrt{P_X Q_X} \left( \sum_y \sqrt{P_{Y|X} Q_{Y|X}} \right) \left( \sum_z \sqrt{P_{Z|X} Q_{Z|X}} \right)$$

Let $a_x = \sqrt{P_X(x)Q_X(x)}$. Let $b_x = A(P_{Y|X=x}||Q_{Y|X=x}) = \sum_y \sqrt{P_{Y|X}(y|x)Q_{Y|X}(y|x)}$. Let $c_x = A(P_{Z|X=x}||Q_{Z|X=x}) = \sum_z \sqrt{P_{Z|X}(z|x)Q_{Z|X}(z|x)}$. Then $A(P_{XYZ}||Q_{XYZ}) = \sum_x a_x b_x c_x$.

Now compute the marginal affinities. $P_{XY}(x,y) = P_X(x)P_{Y|X}(y|x)$. $Q_{XY}(x,y) = Q_X(x)Q_{Y|X}(y|x)$.

$$A(P_{XY}||Q_{XY}) = \sum_{x,y} \sqrt{P_{XY}(x,y)Q_{XY}(x,y)} = \sum_{x,y} \sqrt{P_X P_{Y|X} Q_X Q_{Y|X}}$$

$$= \sum_x \sqrt{P_X Q_X} \sum_y \sqrt{P_{Y|X} Q_{Y|X}} = \sum_x a_x b_x$$

$P_{XZ}(x,z) = P_X(x)P_{Z|X}(z|x)$. $Q_{XZ}(x,z) = Q_X(x)Q_{Z|X}(z|x)$.

$$A(P_{XZ}||Q_{XZ}) = \sum_{x,z} \sqrt{P_{XZ}(x,z)Q_{XZ}(x,z)} = \sum_{x,z} \sqrt{P_X P_{Z|X} Q_X Q_{Z|X}}$$

$$= \sum_x \sqrt{P_X Q_X} \sum_z \sqrt{P_{Z|X} Q_{Z|X}} = \sum_x a_x c_x$$

We need to prove $\sum_x a_x b_x + \sum_x a_x c_x - 1 \le \sum_x a_x b_x c_x$. Since $P_{Y|X=x}$ and $Q_{Y|X=x}$ are probability distributions for each $x$, their affinity $b_x = A(P_{Y|X=x}||Q_{Y|X=x}) \in [0,1]$. Similarly $c_x \in [0,1]$. Consider the inequality $1 - b_x c_x \le (1-b_x) + (1-c_x)$. This is equivalent to $1 - b_x c_x \le 2 - b_x - c_x$, or $b_x + c_x - b_x c_x \le 1$, or $1 - (1-b_x)(1-c_x) \le 1$. Since $1 - b_x \ge 0$ and $1 - c_x \ge 0$, we have $(1-b_x)(1-c_x) \ge 0$, so the inequality holds. $1 - b_x c_x \le (1-b_x) + (1-c_x)$. Multiply by $a_x = \sqrt{P_X(x)Q_X(x)} \ge 0$ and sum over $x$:

$$\sum_x a_x(1 - b_x c_x) \le \sum_x a_x(1 - b_x) + \sum_x a_x(1 - c_x)$$

$$\sum_x a_x - \sum_x a_x b_x c_x \le \sum_x a_x - \sum_x a_x b_x + \sum_x a_x - \sum_x a_x c_x$$

Let $A_X = A(P_X||Q_X) = \sum_x a_x$.

$$A_X - A(P_{XYZ}||Q_{XYZ}) \le (A_X - A(P_{XY}||Q_{XY})) + (A_X - A(P_{XZ}||Q_{XZ}))$$

$$A_X - A(P_{XYZ}||Q_{XYZ}) \le 2A_X - A(P_{XY}||Q_{XY}) - A(P_{XZ}||Q_{XZ})$$

Rearranging gives:

$$A(P_{XY}||Q_{XY}) + A(P_{XZ}||Q_{XZ}) - A_X \le A(P_{XYZ}||Q_{XYZ})$$

We know that $A_X = A(P_X||Q_X) = \sum_x \sqrt{P_X(x)Q_X(x)}$. By Cauchy-Schwarz, $A_X \le \sqrt{\sum_x P_X(x)}\sqrt{\sum_x Q_X(x)} = 1 \times 1 = 1$. Since $A_X \le 1$, we have $-1 \le -A_X$. Therefore, $A(P_{XY}||Q_{XY}) + A(P_{XZ}||Q_{XZ}) - 1 \le A(P_{XY}||Q_{XY}) + A(P_{XZ}||Q_{XZ}) - A_X$. Combining the inequalities:

$$A(P_{XY}||Q_{XY}) + A(P_{XZ}||Q_{XZ}) - 1 \le A(P_{XYZ}||Q_{XYZ})$$

This is the inequality we needed to prove.

# Solution 5

We need to determine if the Hellinger divergence $H^2$ is subadditive in the sense of (2), which means:

$$H^2(P_{XYZ}||P_X P_Y P_Z) \leq H^2(P_{XY}||P_X P_Y) + H^2(P_{XZ}||P_X P_Z)$$

where $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$. Let $P = P_{XYZ}$. Let $Q = P_X P_Y P_Z$. We want to apply the result from Problem 4. For this, we need $P$ and $Q$ to factor in the specific form: $P = P_X P_{Y|X} P_{Z|X}$ (given). $Q = Q_X Q_{Y|X} Q_{Z|X}$. Let's identify the components of $Q = P_X P_Y P_Z$. The marginal $Q_X(x) = \sum_{y,z} P_X(x) P_Y(y) P_Z(z) = P_X(x)(\sum_y P_Y(y))(\sum_z P_Z(z)) = P_X(x)$. So we identify $Q_X = P_X$. The conditional $Q_{Y|X=x}(y) = Q_{XY}(x,y)/Q_X(x)$. $Q_{XY}(x,y) = \sum_z Q(x,y,z) = \sum_z P_X(x) P_Y(y) P_Z(z) = P_X(x) P_Y(y)$. So $Q_{Y|X=x}(y) = (P_X(x) P_Y(y))/P_X(x) = P_Y(y)$. We identify $Q_{Y|X} = P_Y$. Similarly, $Q_{XZ}(x,z) = P_X(x) P_Z(z)$, so $Q_{Z|X=x}(z) = (P_X(x) P_Z(z))/P_X(x) = P_Z(z)$. We identify $Q_{Z|X} = P_Z$. Now we check if the factorization holds for $Q$: $Q_X(x) Q_{Y|X}(y|x) Q_{Z|X}(z|x) = P_X(x) P_Y(y) P_Z(z)$. This matches the definition of $Q$. So, the distributions $P = P_{XYZ}$ and $Q = P_X P_Y P_Z$ satisfy the conditions of Problem 4. We can apply the result of Problem 4:

$$H^2(P_{XYZ}||Q_{XYZ}) \leq H^2(P_{XY}||Q_{XY}) + H^2(P_{XZ}||Q_{XZ})$$

Substituting $Q_{XYZ} = P_X P_Y P_Z$, $Q_{XY} = Q_X Q_{Y|X} = P_X P_Y$, and $Q_{XZ} = Q_X Q_{Z|X} = P_X P_Z$:

$$H^2(P_{XYZ}||P_X P_Y P_Z) \leq H^2(P_{XY}||P_X P_Y) + H^2(P_{XZ}||P_X P_Z)$$

This is exactly the subadditivity inequality (2) for Hellinger divergence. Yes, we can conclude that the Hellinger divergence is subadditive in the sense of (2).

# Solution 6

We investigate subadditivity for Total Variation (TV) distance and $\chi^2$-divergence. The subadditivity condition (2) is:
$$D_f(P_{XYZ}||P_X P_Y P_Z) \leq D_f(P_{XY}||P_X P_Y) + D_f(P_{XZ}||P_X P_Z)$$

where $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$. Let $Q = P_X P_Y P_Z$.

$\chi^2$**-divergence:** $f(t) = (t-1)^2$. Let's test with a counterexample. Let $X$ be Bernoulli(1/2), $Y = X$, and $Z = X$. Then $P_{XYZ}(x,y,z) = P_X(x)\mathbb{I}(y=x)\mathbb{I}(z=x)$. Here $P_{Y|X}(y|x) = \mathbb{I}(y=x)$ and $P_{Z|X}(z|x) = \mathbb{I}(z=x)$, so the condition $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$ holds. $P_X(x) = 1/2$ for $x \in \{0,1\}$. $P_Y(y) = P_X(y) = 1/2$. $P_Z(z) = P_X(z) = 1/2$. $Q(x,y,z) = P_X(x) P_Y(y) P_Z(z) = (1/2)(1/2)(1/2) = 1/8$ for all $x,y,z \in \{0,1\}$. $P_{XYZ}(x,y,z)$ is $1/2$ if $x = y = z$ (2 cases) and 0 otherwise (6 cases). LHS: $D_{\chi^2}(P_{XYZ}||Q) = \sum_{x,y,z} \frac{(P_{XYZ}(x,y,z) - Q(x,y,z))^2}{Q(x,y,z)}$

$$= \frac{(1/2 - 1/8)^2}{1/8} \times 2 + \frac{(0 - 1/8)^2}{1/8} \times 6$$

$$= \frac{(3/8)^2}{1/8} \times 2 + \frac{(-1/8)^2}{1/8} \times 6 = \frac{9/64}{1/8} \times 2 + \frac{1/64}{1/8} \times 6$$

$$= (9/8) \times 2 + (1/8) \times 6 = 18/8 + 6/8 = 24/8 = 3$$

RHS: $D_{\chi^2}(P_{XY}||P_X P_Y) + D_{\chi^2}(P_{XZ}||P_X P_Z)$. $P_{XY}(x,y) = P_X(x) P_{Y|X}(y|x) = (1/2)\mathbb{I}(y=x)$. $P_X P_Y = (1/2)(1/2) = 1/4$. $P_{XY}$ is $1/2$ if $x = y$ (2 cases) and 0 otherwise (2 cases). $D_{\chi^2}(P_{XY}||P_X P_Y) = \sum_{x,y} \frac{(P_{XY}(x,y) - P_X(x) P_Y(y))^2}{P_X(x) P_Y(y)}$

$$= \frac{(1/2 - 1/4)^2}{1/4} \times 2 + \frac{(0 - 1/4)^2}{1/4} \times 2$$

$$= \frac{(1/4)^2}{1/4} \times 2 + \frac{(-1/4)^2}{1/4} \times 2 = \frac{1/16}{1/4} \times 2 + \frac{1/16}{1/4} \times 2$$

$$= (1/4) \times 2 + (1/4) \times 2 = 1/2 + 1/2 = 1$$

Similarly, $P_{XZ}(x,z) = (1/2)\mathbb{I}(z=x)$ and $P_X P_Z = 1/4$. $D_{\chi^2}(P_{XZ}||P_X P_Z) = 1$. RHS $= 1 + 1 = 2$. We need to check if $LHS \leq RHS$, which is $3 \leq 2$. This is false. Therefore, $\chi^2$-divergence is NOT subadditive in the sense of (2).

**Total Variation (TV) distance:** $D_{TV}(P||Q) = \frac{1}{2}\sum_w |P(w) - Q(w)|$. This corresponds to $f(t) = |t - 1|/2$. We want to prove $D_{TV}(P_{XYZ}||Q) \leq D_{TV}(P_{XY}||Q_{XY}) + D_{TV}(P_{XZ}||Q_{XZ})$. Let $\Delta(x,y,z) = P_{XYZ}(x,y,z) - Q(x,y,z) = P_X P_{Y|X} P_{Z|X} - P_X P_Y P_Z$. Let $\Delta_Y(x,y) = P_{XY}(x,y) - Q_{XY}(x,y) = P_X P_{Y|X} - P_X P_Y$. Let $\Delta_Z(x,z) = P_{XZ}(x,z) - Q_{XZ}(x,z) = P_X P_{Z|X} - P_X P_Z$. We can write $\Delta(x,y,z)$ as:

$$\Delta(x,y,z) = P_X(x)P_{Y|X}(y|x)P_{Z|X}(z|x) - P_X(x)P_Y(y)P_{Z|X}(z|x) + P_X(x)P_Y(y)P_{Z|X}(z|x) - P_X(x)P_Y(y)P_Z(z)$$

$$= P_{Z|X}(z|x)[P_X(x)P_{Y|X}(y|x) - P_X(x)P_Y(y)] + P_Y(y)[P_X(x)P_{Z|X}(z|x) - P_X(x)P_Z(z)]$$

$$= P_{Z|X}(z|x)\Delta_Y(x,y) + P_Y(y)\Delta_Z(x,z)$$

Now we take the absolute value and sum over $y, z$:

$$\sum_{y,z}|\Delta(x,y,z)| = \sum_{y,z}|P_{Z|X}(z|x)\Delta_Y(x,y) + P_Y(y)\Delta_Z(x,z)|$$

By the triangle inequality:

$$\leq \sum_{y,z}|P_{Z|X}(z|x)\Delta_Y(x,y)| + \sum_{y,z}|P_Y(y)\Delta_Z(x,z)|$$

Since $P_{Z|X} \geq 0$ and $P_Y \geq 0$:

$$= \sum_{y,z}P_{Z|X}(z|x)|\Delta_Y(x,y)| + \sum_{y,z}P_Y(y)|\Delta_Z(x,z)|$$

$$= \sum_y |\Delta_Y(x,y)|\left(\sum_z P_{Z|X}(z|x)\right) + \sum_z |\Delta_Z(x,z)|\left(\sum_y P_Y(y)\right)$$

Since $\sum_z P_{Z|X}(z|x) = 1$ and $\sum_y P_Y(y) = 1$:

$$= \sum_y |\Delta_Y(x,y)| + \sum_z |\Delta_Z(x,z)|$$

Now, sum over $x$:

$$\sum_{x,y,z}|\Delta(x,y,z)| \leq \sum_x \left(\sum_y |\Delta_Y(x,y)| + \sum_z |\Delta_Z(x,z)|\right)$$

$$= \sum_{x,y} |\Delta_Y(x,y)| + \sum_{x,z} |\Delta_Z(x,z)|$$

Multiplying by 1/2:

$$\frac{1}{2}\sum_{x,y,z}|P_{XYZ} - Q| \leq \frac{1}{2}\sum_{x,y}|P_{XY} - Q_{XY}| + \frac{1}{2}\sum_{x,z}|P_{XZ} - Q_{XZ}|$$

$$D_{TV}(P_{XYZ}||Q) \leq D_{TV}(P_{XY}||Q_{XY}) + D_{TV}(P_{XZ}||Q_{XZ})$$

Therefore, the Total Variation distance IS subadditive in the sense of (2).

Summary: Total variation is subadditive, $\chi^2$-divergence is not.

# Solution 7

We are looking for convex functions $f$ with $f(1) = 0$ such that the induced f-divergence $D_f$ satisfies the subadditivity property (2):

$$D_f(P_{XYZ}||P_X P_Y P_Z) \leq D_f(P_{XY}||P_X P_Y) + D_f(P_{XZ}||P_X P_Z)$$

where $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$.

From the previous parts, we know:

- KL divergence ($f(t) = t \log t$) satisfies (2) with equality.

- Symmetric KL divergence ($f(t) = (t-1)\log t$) satisfies (2) with equality.

- Hellinger divergence ($f(t) = (\sqrt{t} - 1)^2$, or $f(t) = \frac{1}{2}(\sqrt{t} - 1)^2$) satisfies (2).

- Total Variation distance ($f(t) = \frac{1}{2}|t - 1|$) satisfies (2).

- $\chi^2$-divergence ($f(t) = (t-1)^2$) does NOT satisfy (2).

We also showed that the Reverse KL divergence ($f(t) = -\log t$) satisfies (2) with equality. Note that $f_{SKL}(t) = (t-1)\log t = t\log t - \log t = f_{KL}(t) + f_{RKL}(t)$. Since KL and RKL satisfy (2) with equality, their sum SKL also satisfies it with equality.

This property (subadditivity under the specific tensor product structure) is known to hold for the $\alpha$-divergences $f_\alpha(t) = \frac{t^\alpha - \alpha t + \alpha - 1}{\alpha(\alpha - 1)}$ (for $\alpha \neq 0, 1$) when $0 \leq \alpha \leq 1$.

- The limit $\alpha \to 1$ gives $f_1(t) = t\log t$ (KL divergence).

- The limit $\alpha \to 0$ gives $f_0(t) = -\log t$ (Reverse KL divergence).

- For $\alpha = 1/2$, $f_{1/2}(t) = 2(\sqrt{t} - 1)^2$ which is proportional to the function for Hellinger divergence. Since the inequality is linear in $f$, Hellinger satisfies (2).

Therefore, all $\alpha$-divergences with $\alpha \in [0, 1]$ satisfy the subadditivity inequality (2). Examples beyond KL and RKL (and SKL which is their sum) are the functions $f_\alpha(t)$ for $\alpha \in (0, 1)$.

The Total Variation distance, corresponding to $f(t) = c|t - 1|$ for $c > 0$, also satisfies (2). This function is convex and $f(1) = 0$. It is not an $\alpha$-divergence.

Furthermore, if $f_1$ and $f_2$ are convex functions satisfying $f_i(1) = 0$ and their induced divergences satisfy (2), then any non-negative linear combination $f = af_1 + bf_2$ ($a, b \geq 0$) is also convex, satisfies $f(1) = 0$, and its divergence satisfies (2):

$$D_f(P||Q) = aD_{f_1}(P||Q) + bD_{f_2}(P||Q)$$

$$\leq a(D_{f_1}(P_{XY}||Q_{XY}) + D_{f_1}(P_{XZ}||Q_{XZ})) + b(D_{f_2}(P_{XY}||Q_{XY}) + D_{f_2}(P_{XZ}||Q_{XZ}))$$

$$= D_f(P_{XY}||Q_{XY}) + D_f(P_{XZ}||Q_{XZ})$$

where $Q = P_X P_Y P_Z$, $Q_{XY} = P_X P_Y$, $Q_{XZ} = P_X P_Z$.

So, apart from the examples mentioned in the problem set (KL, SKL, Hellinger), other examples include:

1. Reverse KL divergence: $f(t) = -\log t$.

2. $\alpha$-divergences for $\alpha \in (0, 1)$ excluding $\alpha = 1/2$ if Hellinger is already listed. E.g., $f_{1/3}(t) = \frac{t^{1/3} - t/3 - 2/3}{(1/3)(-2/3)} = -\frac{9}{2}(t^{1/3} - t/3 - 2/3)$.

3. Total Variation distance: $f(t) = |t - 1|$. (Up to a constant factor).

4. Any non-negative linear combination of functions whose divergences satisfy (2), such as combinations of $f_\alpha(t)$ for $\alpha \in [0, 1]$ and $f_{TV}(t) = |t - 1|$. For instance, $f(t) = t\log t + |t - 1|$.