Preliminaries of Probability:

- Random variables are denoted by upper-case letters, e.g., X, & its support set (or alphabet) denoted by calligraphic letters: $\mathcal{X}$.
- we write $X \sim P_X$ to denote X is distributed according the dist P_X
- In this course, we are only concerned with either discrete (X is a discrete set) continuous r.v (X is a continuum).
- Discrete rv are usually characterized by their Pmf: $X \sim P_X \Leftrightarrow$ $P_r(X=x) = P_X(x)$
- continuous rv " " " " " pdf: $X \sim P_X$ & P_X has density p_X $\Leftrightarrow$ $P_r(X \in A) = \int_A p_X(x)\,dx$ (lower case)
- cdf: $F_X(x) := P_r(X \le x)$
- Expected value: $X \sim P_X$, $E[X] := \sum_x x\, P_X(x)$ or $E[X] := \int x\, p_X(x)\,dx$ (pdf)

- k^{th} Moment: $E[X^k]$.
- For example: 2^{nd} moment: $E[X^2] = \sum x^2 p_X(x)$ or $E[X^2] = \int x^2 f_X(x)\,dx$
- Variance: $Var(X) := E[(X - E[X])^2] = E[X^2] - (E[X])^2$

* Examples of discrete r.v:

1. Bernoulli: $X \sim Bernoulli(p)$ ($0 \le p \le 1$)

   pmf: $p_X(x) := Pr(X = x)$ $\quad x \in \{0, 1\}$

   $= p^x (1-p)^{1-x}$

   $= \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$

   $X = \{0, 1\}$

   $E[X] = p$

   $Var(X) = p \cdot (1-p)$

2. Uniform: $X \sim U([k])$ for some integer k: $\{1, 2, \dots, k\}$

   $p_X(x) = \begin{cases} 1/k & x \in [k] \\ 0 & o.w \end{cases}$

   $= \frac{1}{k} \cdot \mathbb{1}_{\{x \in [k]\}}$

   $E[X] = \frac{k+1}{2}$

   $Var(X) = \frac{k^2 - 1}{12}$

3. Binomial : $X \sim Bin(n, p)$ integer n & $p \in [0,1]$

Let $B_1, \dots, B_2 \overset{iid}{\sim} Ber(p)$. Then $X = B_1 + B_2 + \dots + B_n$

It can be easily shown that

$$Pr(X=k) = \begin{cases} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} & 0 \le k \le n \\ 0 & \text{OW} \end{cases}$$

$E[X] = np$

$Var(X) = n \cdot p(1-p)$

Example of continuous dist:

1- Exponential : $X \sim Exp(\lambda)$ with $\lambda > 0$:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{ow} \end{cases} \qquad X = \mathbb{R}^+$$

verify that $\lambda \int_0^\infty e^{-\lambda x} dx = 1$ $\Rightarrow$ the above is a valid pdf

Also, verify that:

$$E[X] = \frac{1}{\lambda} \quad \& \quad Var(X) = \frac{1}{\lambda^2}$$

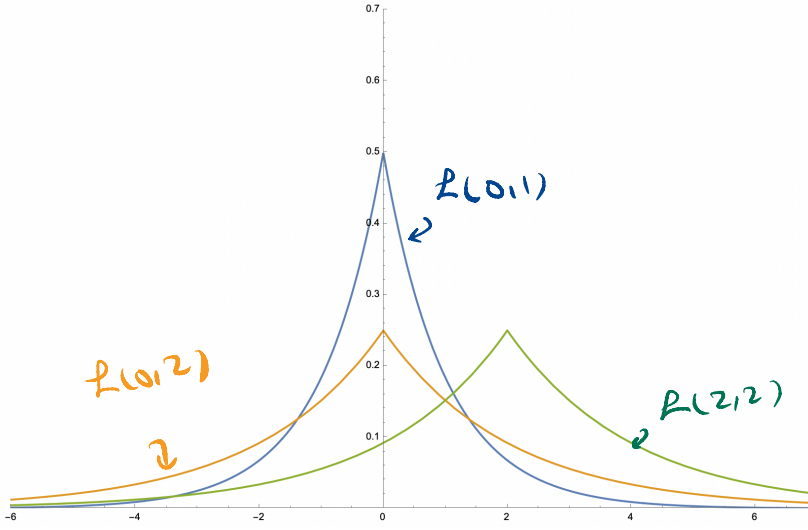
2- Laplace $X \sim \mathcal{L}(\mu, b)$ (b: scale parameter) $\mu \in \mathbb{R}$ & $b > 0$

Pdf: $f(x) = \frac{1}{2b} e^{\frac{-|x-\mu|}{b}}$ $\qquad X = \mathbb{R}$

Verify that $\frac{1}{2b}\int_{-\infty}^{\infty} e^{-\frac{|x-\mu|}{b}}\,dx = 1$.

$E[X] = \mu$

$Var(X) = 2b^2$



Exercise: Let $X \sim L(\mu, b)$

Show $Pr(|X-\mu| \geq t) = e^{-t/b}$

Approach I: Directly

Approach II: First verify that:

$|X-\mu| \sim Exp(b^{-1})$

3- Gaussian $X \sim \mathcal{N}(\mu, \sigma^2)$ $\mu \in \mathbb{R}$ $\sigma > 0$ $x \in \mathbb{R}$

pdf: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Verify that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1$?

Note that Gaussian has a faster decay than Lap

$\mathbb{E}[X] = \mu$ & $\mathrm{Var}(X) = \sigma^2$.

$\mathbb{E}[(X-\mu)^k]$ central moment

$= \begin{cases} 0 & k \text{ odd} \\ \sigma^k (k-1)!! & k \text{ even} \end{cases}$

double factorial

$n!! = n \cdot (n-2)(n-4) \cdots$

Exercise: $\Pr(|X-\mu| \geq t) = 2Q(t/\sigma)$

where

$Q(a) := \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-x^2/2} dx$

Next, we need to define a family of distance measure between distributions. But before that, we need to discuss Jensen's inequality

Detour: Jensen's inequality:

Definition: * Let f be a real-valued function. We say that f is convex if $f(\alpha x + \bar{\alpha} y) \leq \alpha f(x) + \bar{\alpha} f(y) \quad \forall x, y \in \mathbb{R}, \forall \alpha \in [0,1]$ (*)

* Strictly convex if $\leq$ strict inequality.

* Concave if $\geq$

* Strictly concave $>$

We can generalize (*):

(**)

$$f\left(\sum p_i x_i\right) \leq \sum p_i f(x_i) \qquad \text{for all } p_i \text{ such that } \sum p_i = 1.$$

For example:

$$f(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3) = f\Big(\alpha_1 x_1 + (1-\alpha_1)\underbrace{\Big(\tfrac{\alpha_2}{1-\alpha_1} x_2 + \tfrac{\alpha_3}{1-\alpha_1} x_3\Big)}_{y}\Big)$$

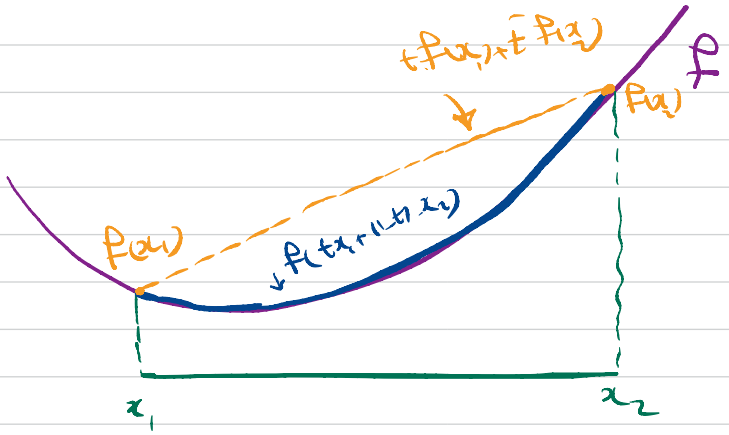
$$\leq \alpha_1 f(x_1) + (1-\alpha_1)\cdot f\Big(\tfrac{\alpha_2}{1-\alpha_1} x_2 + \tfrac{\alpha_3}{1-\alpha_1} x_3\Big)$$

$$\leq \alpha_1 f(x_1) + (1-\alpha_1)\cdot\left[\tfrac{\alpha_2}{1-\alpha_1}\cdot f(x_2) + \tfrac{\alpha_3}{1-\alpha_1} f(x_3)\right]$$

$$= \alpha_1 f(x_1) + \alpha_2 f(x_2) + \alpha_3 f(x_3)$$

More general statement (**) can be proved by induction.

Geometric Interpretation:



Any line segment connecting any two points lies above the function.

Theorem. If f is twice differentiable, then:

f is convex $\iff$ $f''(x) \geq 0$

f is st. convex $\iff$ $f''(x) > 0$

f is concave $\iff$ $f''(x) \leq 0$

f is st. concave $\iff$ $f''(x) < 0$

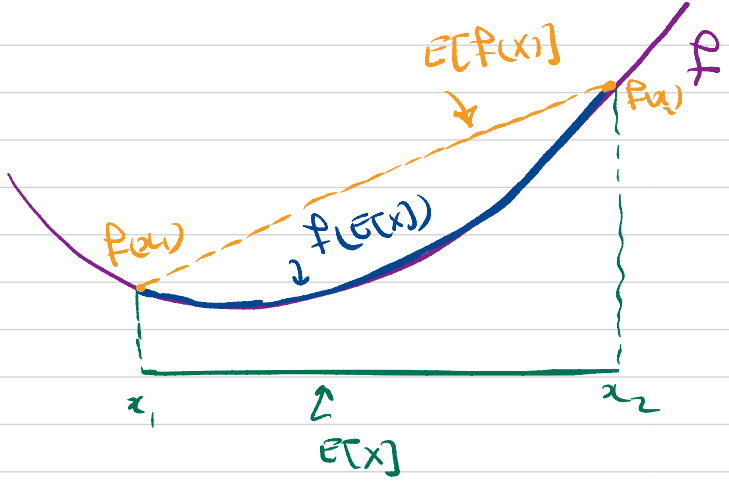
Jensen's Inequality:

Let f be a convex function & X be a random variable.

Then:

$$E[f(x)] \geq f(E[x])$$

"=" iff $X = E[X]$



Distance measure among distributions

Question: Given two distributions P & Q, how to quantify the "distance" between them?

In math, any appropriate measure of "distance" between any pair of objects, must satisfy the following:

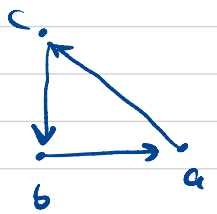
(can be distributions)

1- $d(a, b) = 0 \iff a = b$

2- $d(a, b) = d(b, a)$ Symmetry

3- Triangle inequality: $d(a, c) \leq d(a, b) + d(b, c)$

Example: Euclidean distance



We often ignore some of these properties to define "divergence" instead of distance.

I. Total Variation distance (TV)

$$TV(P,Q) := \frac{1}{2} \sum |P(x) - Q(x)| \qquad \text{discrete}$$

(P(x), Q(x): pmf)

$$:= \frac{1}{2} \int |p(x) - q(x)| \, dx \qquad \text{continuous}$$

(p(x), q(x): pdf)

Fact: TV is in fact a distance metric, meaning it satisfies:

1. $TV(P,Q) = 0 \iff P = Q$
2. $TV(P,Q) = TV(Q,P)$
3. $TV(P,R) \le TV(P,Q) + TV(Q,R)$

Why? $TV(P,R) = \frac{1}{2} \sum |P(x) - R(x)| \le \cdots$ Triangle inequality of 1.1

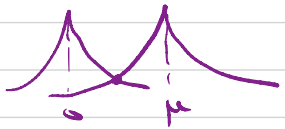
$|P(x) - R(x)| \le |P(x) - Q(x)| + |Q(x) - R(x)|$

Example: $TV(\text{Bernoulli}(p), \text{Bernoulli}(q)) = \frac{1}{2}|p(0)-q(0)| + \frac{1}{2}|p(1)-q(1)|$

$= \frac{1}{2}|1-p-(1-q)| + \frac{1}{2}|p-q|$

$= |p-q|$

$TV(\mathcal{L}(0,b), \mathcal{L}(\mu,b)) = 1 - e^{-|\mu|/2b}$



assume $\mu > 0$

$\uparrow \quad \frac{1}{2}\int \left| \frac{1}{2b} e^{-\frac{|x|}{b}} - \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \right| dx = \frac{1}{4b}\left[\int_{-\infty}^{\mu/2} \cdot + \int_{\mu/2}^{\infty} \cdot\right]$

Exercise: $TV(N(0,\sigma^2), N(\mu,\sigma^2)) = Q(-\frac{\mu}{\sigma}) - Q(\frac{\mu}{\sigma})$

$= 1 - Q(\frac{\mu}{\sigma})$

Theorem. For any dist. P & Q on support set $\mathcal{A}$, we have:

$$TV(P,Q) = \sup_{A \subseteq \mathcal{A}} [P(A) - Q(A)]$$

Equivalent expression for TV

Proof. we need to prove:

① $P(A) - Q(A) \leq TV(P,Q) \quad \forall A \in \mathcal{A}$

② $\exists A^* \in \mathcal{A}$ s.t. $P(A^*) - Q(A^*) = TV(P,Q)$.

Note $P(A) - Q(A) = \frac{1}{2}(P(A) - Q(A)) + \frac{1}{2}(Q(A^c) - P(A^c))$

① Take arbitrary A: $P(A) - Q(A) = \frac{1}{2}(P(A) - Q(A)) + \frac{1}{2}(Q(A^c) - P(A^c)) = \frac{1}{2}\sum_{x \in A}(P(x) - Q(x)) + \frac{1}{2}\sum_{A^c}(Q - P)$

$$\leq \frac{1}{2}\sum_{A} |P(x) - Q(x)| + \frac{1}{2}\sum_{A^c} |Q(x) - P(x)|$$

$$= \frac{1}{2}\sum |P(x) - Q(x)| = TV(P,Q)$$

What is the maximizing A? $A^* = \{x : p(x) \geq q(x)\}$

$$TV(p,q) = \frac{1}{2}\sum_x |p(x)-q(x)| = \frac{1}{2}\left[\sum_{x\in A^*} |p(x)-q(x)| + \sum_{x\in A^{*c}} |p(x)-q(x)|\right]$$

$$= \frac{1}{2}\left[\sum_{x\in A^*} (p(x)-q(x)) + \sum_{x\in A^{*c}} (q(x)-p(x))\right]$$

$$= p(A^*) - q(A^*)$$ ∎

Another equivalent expression for TV:

$$TV(p,q) = \frac{1}{2} \sup_{f:\ |f(x)|\leq 1\ \forall x} E_p[f(x)] - E_q[f(x)]$$

Proof is similar!

There are two other equivalent expressions for TV:

supported on finite set

* **Theorem:** For any P & Q, we have:

$$TV(P,Q) = \inf_{\substack{P_{XY} \\ \{P_X = P \\ P_Y = Q}} \Pr(X \neq Y)$$

← minimization is over all "coupling" of P & Q.

* **Theorem:** For any P, Q, we have:

$$TV(P,Q) = 1 - \int \min\{p(x), q(x)\}\, dx$$

$$= 1 - \sum_x \min\{p(x), q(x)\}$$

II. Kullback-Leibler (KL) Divergence:

$$D(P\|Q) := E_P\left[\log \frac{P(x)}{Q(x)}\right] \overset{\text{discrete}}{=} \sum_x P(x)\cdot \log \frac{P(x)}{Q(x)}$$

↑ natural log

$$\overset{\text{cont.}}{=} \int_x p(x)\cdot \log \frac{p(x)}{q(x)}$$

Example:

- $D(\text{Bernoulli}(p)\,\|\,\text{Bernoulli}(q)) = P(1)\cdot \log \frac{P(1)}{Q(1)} + P(0)\cdot \log \frac{P(0)}{Q(0)}$

$$= p \log \frac{p}{q} + (1-p)\cdot \log \frac{1-p}{1-q}.$$

- $D(L(0,b)\,\|\,L(\mu,b)) = \frac{b\, e^{-|\mu|/b} - b - |\mu|}{b}$

↑ assume $\mu > 0$

$$\frac{1}{2b}\int e^{-\frac{|x|}{b}} \cdot \log \frac{e^{-|x|/b}}{e^{-|x-\mu|/b}}\, dx = \int_{-\infty}^{0} + \int_{0}^{\mu} + \int_{\mu}^{\infty}$$

$$D(N(a,\sigma^2)\,\|\,N(b,\sigma^2)) = \underset{N(a,\sigma^2)}{E}\left[\log \frac{e^{-\frac{(x-a)^2}{2\sigma^2}}}{e^{-\frac{(x-b)^2}{2\sigma^2}}}\right]$$

$$= \frac{1}{2\sigma^2}\cdot E\left[(a-b)(2x-a-b)\right]\log e$$

$$= \frac{1}{2\sigma^2}(a-b)\cdot \log e \underbrace{E\left[(2x-a-b)\right]}_{a-b}$$

$$= \frac{(a-b)^2}{2\sigma^2}$$

Exercise:

$$D(N(a,\sigma_1^2)\,\|\,N(b,\sigma_2^2)) = \frac{1}{2}\log\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2+(a-b)^2}{2\sigma_2^2} - \frac{1}{2}$$

Properties of KL divergence:

1) $D(P\|Q) \geq 0 \quad \forall P, Q$, "$=$" if & only if $P = Q$

2) $D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + \sum P_X(x) D(P_{Y|X=x}\|Q_{Y|X=x})$

($P_X P_{Y|X}$, $Q_X Q_{Y|X}$; P_X: marginal)

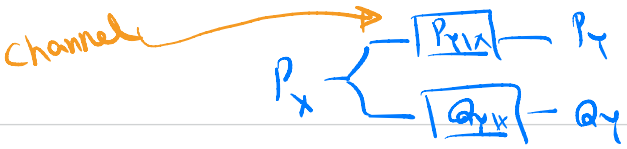
"chain rule"

3) $D(P_{XY}\|Q_{XY}) \geq D(P_X\|Q_X)$ (or $D(P_Y\|Q_Y)$)

monotonicity

4) Let P_Y & Q_Y be output dist. of $P_{Y|X}$ & $Q_{Y|X}$ with common input P_X. Then

$$D(P_Y\|Q_Y) \leq \sum_x P_X(x) \cdot D(P_{Y|X=x}\|Q_{Y|X=x})$$



5) Data Processing Inequality:

$$D(P_Y \| Q_Y) \leq D(P_X \| Q_X)$$

* outputs of a channel are closer than the correspondig inputs.

Proofs. 1. $D(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$

$$= \sum_x q(x) \frac{p(x)}{q(x)} \log \frac{p(x)}{q(x)} = \sum q(x) \cdot f\left(\frac{p(x)}{q(x)}\right)$$

$t \mapsto t \log t$

Jensen's ineq $\geq f\left(\sum_x q(x) \frac{p(x)}{q(x)}\right) = f(1) = 0$

2. $D(P_{XY}\|Q_{XY}) = \sum_{x,y} P(x,y) \cdot \log \frac{P(x,y)}{Q(x,y)} = \sum_{x,y} P(x) \cdot P(y|x) \cdot \log \frac{P(x) \cdot P(y|x)}{Q(x) \cdot Q(y|x)}$

$= \sum_{x,y} P(x) \cdot P(y|x) \cdot \log \frac{P(x)}{Q(x)} + \sum_{x,y} P(x) \cdot P(y|x) \cdot \log \frac{P(y|x)}{Q(y|x)}$

$= \sum_{x} P(x) \cdot \log \frac{P(x)}{Q(x)} + \sum_{x} P(x) \cdot \sum_{y} P(y|x) \cdot \log \frac{P(y|x)}{Q(y|x)}$

3. Note $D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + \underbrace{\sum P(x) \cdot D(P_{Y|X=x}\|Q_{Y|X=x})}_{\geq 0}$

$\Rightarrow D(P_{XY}\|Q_{XY}) \geq D(P_X\|Q_X)$

4.

consider joint dist. $P_{XY} = P_X \cdot P_{Y|X}$ & $Q_{XY} = P_X Q_{Y|X}$

$D(P_{XY}\|Q_{XY}) = D(P_X\|P_X) + \sum P(x) \cdot D(P_{Y|X=x}\|Q_{Y|X=x})$

↑ chain rule

$D(P_X\|P_X) = 0$

By monotonicity: $D(P_Y \| Q_Y) \le D(P_{XY} \| Q_{XY}) = \sum p(x) \cdot D(P_{Y|X=x} \| Q_{Y|X=x})$

5. $P_X P_{Y|X}$ $Q_X P_{Y|X}$

$$D(P_{XY} \| Q_{XY}) = D(P_X \| Q_X) + \sum p(x) \cdot D(P_{Y|X=x} \| P_{Y|X=x})$$

By monotonicity: $D(P_Y \| Q_Y) \le D(P_{XY} \| Q_{XY}) = D(P_X \| Q_X)$

* Note that

1) $D(p\|q) \neq D(q\|p)$ * Not Symmetric *

2) $D(p\|r) \le D(p\|q) + D(q\|p)$ No triangle inequality

* So KL-divergence is not a distance metric

Theorem. (Pinsker's Inequality.)

$$TV(P,Q) \leq \sqrt{\frac{D(P\|Q)}{2}}$$

* You may see different constant in RHS, but 2 is optimal. Note that this doesn't mean that this is the best inequality connecting TV & D.

* Note that $0 \leq TV(P,Q) \leq 1$
  - 0: "=" $P=Q$
  - 1: if P & Q are mutually singular; i.e., they have non-overlapping support.

However;

$0 \leq D(P\|Q) \leq \infty$
- 0: "=" $P=Q$
- ∞: if & only if $\exists\, x_0 \in \text{Supp}(P)$ but $x_0 \notin \text{Supp}(Q)$.

Equivalently $\text{Supp}(Q) \subseteq \text{Sup}(P)$

Proof: We first show that it suffices to prove this inequality for Bernoulli dist. Then, we prove it assuming P & Q are Bernoulli.

I. consider the following channel:

Note that when $X \sim P \Rightarrow Y \sim \text{Ber}(P(E))$ → Because $P_r(Y=1) = \sum_x P_r(X=x) \cdot P_r(Y=1|X=x)$

$= \sum_x P(x) \cdot \underbrace{P_{Y|X}(1|x)}_{1_{\{x \in E\}}} = P(E)$

where $E = \{x : P(x) \geq Q(x)\}$

$X \sim Q \Rightarrow Y \sim \text{Ber}(Q(E))$

$$D(P\|Q) \overset{\text{DPI}}{\geq} D(\text{Ber}(P(E)) \| \text{Ber}(Q(E)))$$

$$\overset{\text{Pinsker's ineq for Bernoulli}}{\geq} 2\,TV^2(\text{Ber}(P(E)), \text{Ber}(Q(E))) = 2 \cdot (P(E) - Q(E))^2$$

$$= 2\,TV^2(P, Q)$$

So we only need to prove it for binary.

II. Suppose $P = \text{Ber}(p)$ & $Q = \text{Ber}(q)$. WLOG, we assume $p \geq q$.

[Note that $D(\text{Ber}(p)\|\text{Ber}(q)) = D(\text{Ber}(\bar{p})\|\text{Ber}(\bar{q}))$ so we can always assume $p \geq q$]

$$D(p\|q) := p\log\frac{p}{q} + \bar{p}\log\frac{\bar{p}}{\bar{q}} \quad \& \quad TV(p,q) = |p-q|$$

Define:

$$g(p,q) := p\log\frac{p}{q} + \bar{p}\log\frac{\bar{p}}{\bar{q}} - 2(p-q)^2.$$ We need to show that $g(p,q) \geq 0 \ \forall p,q$.

$$\frac{\partial g}{\partial q} = -\frac{p}{q} + \frac{\bar{p}}{\bar{q}} - 4(p-q) = \frac{q-p}{q(1-q)} - 4(p-q) \leq 0$$

So $q \mapsto g(p,q)$ is decreasing for $q \leq p \Rightarrow g(p,q) \geq g(p,p) = 0$

* Important Observation: a relationship between TV & KL was obtained just by looking at the binary case!

* We discussed two measures of "distance": TV & KL.

We can construct a family of other divergences:

Def. Let $f: (0,\infty) \to \mathbb{R}$ be a convex function satisfying $f(1)=0$. Given two distributions P & Q, their f-divergence is:

$$D_f(P\|Q) := E_Q\left[f\left(\frac{P(x)}{Q(x)} \right) \right]$$

dis $$= \sum_x Q(x) \cdot f\left(\frac{P(x)}{Q(x)} \right)$$ ← pmfs of P & Q

$$= \int q(x) \cdot f\left(\frac{p(x)}{q(x)} \right) dx$$ ← pdf of P & Q

Examples:

1. $f(t) = \frac{1}{2}|t-1|$: $D_f(p\|q) = E_q\left[\frac{1}{2}\left|\frac{p(x)}{q(x)} - 1\right|\right]$

discrete ↓

$$= \frac{1}{2} \cdot \sum q(x) \cdot \left|\frac{p(x)}{q(x)} - 1\right|$$

$$= \frac{1}{2} \sum |p(x) - q(x)|$$

$$= TV(p, q).$$

2. $f(t) = t \log t$: $D_f(p\|q) = D(p\|q)$

3- $f(t) = (\sqrt{t}-1)^2$: $D_f(p\|q) = \sum q(x)\cdot\left(\sqrt{\frac{p(x)}{q(x)}}-1\right)^2$

$= \sum \left(p(x) + q(x) - 2\sqrt{p(x)\cdot q(x)}\right)$

$= 2 - 2\sum_x \sqrt{p(x)\cdot q(x)}$

$= H(p,q)$

Squared Hellinger distance

Can be obtained by $f(t) = 2(1-\sqrt{t})$

4- $f(t) = (t-1)^2$ χ^2- divergence

Same thing for $f(t) = t^2 - 1$.

or $f(t) = \frac{1}{t} - t$

5- $f(t) = \frac{t^\alpha - 1}{\alpha - 1}$ for $\alpha \in (0,1)\cup(1,\infty)$

χ^α- divergence

$$\chi^\alpha(p\|q) = \frac{1}{\alpha-1}\left[\sum_x p(x)^\alpha \cdot q(x)^{1-\alpha} - 1\right]$$

$\alpha = 1/2 \rightarrow$ Squared Hellinger distance:

we are often interested in a function of χ^α:

$$D_\alpha(p\|q) := \frac{1}{\alpha-1}\left[\log\left(1 + (\alpha-1)\cdot\chi^\alpha(p\|q)\right)\right]$$

* Rényi divergence *

6 $f(t) = t\log t - (1+t)\cdot\log\left(\frac{1+t}{2}\right)$

$$D_f(p\|q) = D\left(p \,\|\, \frac{p+q}{2}\right) + D\left(q \,\|\, \frac{p+q}{2}\right)$$

↑ Symmetric

7. Hockey-Stick divergence: or E_γ-divergence

$$E_\gamma(P\|Q) := D_{f_\gamma}(P\|Q)$$

Fix $\gamma \geq 0$: $f_\gamma(t) = (t-\gamma)_+ - (1-\gamma)_+$

Hence, $E_\gamma(P\|Q) \overset{\gamma \geq 1}{=} \sum_x Q(x) \left(\frac{P(x)}{Q(x)} - \gamma\right)_+ = \sum_x \left(P(x) - \gamma Q(x)\right)_+$

&

$$E_\gamma(P\|Q) \overset{\gamma < 1}{=} \sum_x Q(x) \left(\frac{P(x)}{Q(x)} - \gamma\right)_+ - 1 + \gamma = \sum_x \left(P(x) - \gamma Q(x)\right)_+ - \underbrace{(1-\gamma)}_{= \sum_x (P(x) - \gamma Q(x))}$$

$$= \sum_x Q(x) \left(\gamma - \frac{P(x)}{Q(x)}\right)_+ = \sum_x \left(\gamma Q(x) - P(x)\right)_+$$

we could equivalently define: $f_\gamma(t) = \begin{cases} (t-\gamma)_+ & \gamma \geq 1 \\ (\gamma - t)_+ & \gamma < 1 \end{cases}$