

Recap

$$u \rightarrow \gamma u$$

f-divergences

- Divergence:  $\delta(P, Q) = 0 \iff P = Q$

- Metric: Divergence + Symmetry + Triangle Ineq.

- Convexity

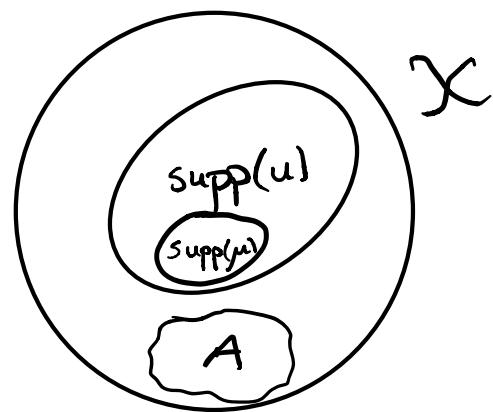
---

## Additional Technical Background

of  $\mu$  measures, NOT function  
set of non-neg. on  $X$

Definition (Absolute Continuity): Let  $\mu, \nu \in \underline{P(X)}$ .  
We say  $\mu$  is absolutely continuous with respect to  $\nu$ , denoted by  $\mu \ll \nu$ , if  $\nu(A) = 0 \Rightarrow \mu(A) = 0$ .

Illustration:



i.e.  $\text{supp}(\mu) \subseteq \text{supp}(\nu)$

Theorem (Radon-Nikodym)

Let  $\mu, \nu \in \mathcal{M}_+(\mathcal{X})$  with  $\mu \ll \nu$ . Then there exists a function  $f \in L^1(\nu)$  such that

$$\underbrace{\mu(A)}_{\int_A d\nu(x)} = \int_A f(x) d\nu(x) \quad \text{for measurable sets } A$$

The function  $f$  is called the Radon-Nikodym derivative of  $\mu$  w.r.t  $\nu$ , often denoted by

$$f = \frac{d\mu}{d\nu}$$

Examples - Let  $\mu \ll \nu$

① Let  $\nu = \#\#$  be the counting measure

where

$$\# = \begin{cases} |A|, & \text{if } |A| < \infty \\ \infty, & \text{o/w} \end{cases}$$

If  $\mu \ll \#$ , then the  $\text{supp}(\mu)$  is countable and

$P = \frac{d\mu}{d\#}$  is the pmf of  $\mu$

( $p(x) = \mu(\{x\})$ , and  $\mu(A) = \sum_{x \in A} p(x)$ )

② Let  $\mu = \lambda$  be the Lebesgue measure.

$$(\lambda(A) = \text{vol}(A) = \int_A dx)$$

If  $\mu \ll \lambda$  then  $p = \frac{d\mu}{d\lambda}$  is the pdf induced by  $\mu$ , and

$$\mu(A) = \int_A p(x) d\lambda(x)$$

shorthand  $dx$

## ► f-divergences

Definition (f-divergence): Let

$$f: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$$

be a convex function such that

①  $f(1) = 0$

②  $f$  is strictly convex around 1:

$$f(\underbrace{\alpha x + (1-\alpha)y}_1) < \alpha f(x) + (1-\alpha)f(y) \quad \forall x, y \in \mathbb{R}_{\geq 0}$$

$\alpha \in [0, 1] \text{ s.t. } \alpha x + (1-\alpha)y = 1$

The f-divergence between two probability measures on the same space  $P, Q \in \mathcal{P}(X)$  that are both dominated by the same measure  $\lambda$  i.e.  $P, Q \ll \lambda$ , is

Such a  $\lambda$  always exists  $\rightarrow$  non-obvious  
<sup>\* doesn't need to be a prob. measure</sup>

$$D_f(P||Q) := \mathbb{E} \left[ f \left( \frac{dP/d\lambda}{dQ/d\lambda} \right) \right]$$

$$= \int_{\text{supp}(Q)} f \left( \frac{dP/d\lambda(x)}{dQ/d\lambda(x)} \right) dQ(x)$$

If  $P \ll Q$ , then

$$D_f(P||Q) = \mathbb{E}_Q \left[ f \left( \frac{dP}{dQ} \right) \right]$$

## Examples

① If  $P, Q \ll \#$  then

$$D_f(P||Q) = \sum_{x \in \text{supp}(Q)} q(x) f \left( \frac{p(x)}{q(x)} \right)$$

where  $p, q$  are the pmfs of  $P, Q$  respectively.

② If  $P, Q \ll \lambda$  then

$$D_f(P||Q) = \int_{\text{Supp}(Q)} q(x) f\left(\frac{p(x)}{q(x)}\right) dx$$

where  $p, q$  are pdfs of  $P, Q$ , respectively.

Conventions:

$$(i) f(0) = f(0^+)$$

$$(ii) 0 = 0 f(0)$$

## Important Special Cases

Shortenels: We'll use

$$\frac{dP}{d\lambda} = dP, \quad \frac{dQ}{d\lambda} = dQ$$

and

$$\int g\left(\frac{dP}{d\lambda}, \frac{dQ}{d\lambda}\right) d\lambda = \int g(dP, dQ)$$

# Kullback - leibler Divergence

$$f(x) = x \log x$$

$$D_{KL}(P \parallel Q) = D_{x \log x}(P \parallel Q)$$

$$= \mathbb{E}_Q \left[ \frac{dP}{dQ} \log \left( \frac{dP}{dQ} \right) \right]$$

$$= \mathbb{E}_P \left[ \log \left( \frac{dP}{dQ} \right) \right]$$

Radon Theorem

## Comments

① If  $P, Q, \ll \#$ :  $D_{KL}(P \parallel Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$

② If  $P, Q, \ll \mathcal{X}$ :  $D_{KL}(P \parallel Q) = \int_X P(x) \log \frac{P(x)}{Q(x)}$

③  $D_{KL}$  is a divergence, but definitely NOT a metric. In fact, it is not a metric and it does not satisfy triangle inequality.

④ If  $P \not\ll Q$ ,  $D_{KL}(P \parallel Q) = \infty$

⑤ Let  $f(x) = -\log x$

$$D_{-\log x}(P \parallel Q) = \mathbb{E}_Q \left[ \log \left( \frac{dQ}{dP} \right) \right] = D_{KL}(Q \parallel P)$$

# Total Variation Distance

$$f(x) = \frac{1}{2} |x - 1|$$

$$\delta_{TV}(P, Q) \triangleq D_{\frac{1}{2}|X-1|}(P||Q)$$

$$= \frac{1}{2} \mathbb{E}_Q \left[ \left| \frac{dP}{dQ} - 1 \right| \right]$$

$$= \frac{1}{2} \int |dP - dQ|$$

Comments:

$$\textcircled{1} \quad P, Q \ll \# \quad \delta_{TV}(P, Q) = \frac{1}{2} \sum_{x \in X} |P(x) - Q(x)| \\ = \frac{1}{2} \|P - Q\|_1$$

$$\textcircled{2} \quad P, Q \ll \lambda: \quad \delta_{TV}(P, Q) = \frac{1}{2} \|P - Q\|_{L^1(\mathbb{R}^d)} \quad \|f\|_{L^1(\mathbb{R}^d)} = \int |f| dx$$

\textcircled{3} \quad \delta\_{TV} \text{ is a metric on } \mathcal{P}(X)

\textcircled{4} \quad If \ \text{Supp}(P) \cap \text{Supp}(Q) = \emptyset, \text{ then } \delta\_{TV}(P, Q) = 1

## $\chi^2$ -Divergence

$$f(x) = (x-1)^2$$

$$\chi^2(P \parallel Q) = D_{(x-1)^2}(P \parallel Q)$$

$$= \mathbb{E}_Q \left[ \left( \frac{dP}{dQ} - 1 \right)^2 \right]$$

$$= \int \left( \frac{dP}{dQ} \right)^2 dQ - 2 \underbrace{\int \frac{dP}{dQ} dQ}_{1} + \underbrace{\int dQ}_{1 \leftarrow a P \text{ measure}}$$

$$= \mathbb{E}_Q \left[ \left( \frac{dP}{dQ} \right)^2 \right] - 1$$

$$= \mathbb{E}_Q \left[ \left( \frac{dP}{dQ} \right)^2 - 1 \right] = D_{(x-1)^2}(P \parallel Q)$$

Comments:

① wrt above, mapping  $f \mapsto D_f$  is NOT one-to-one  
 (injective).

② If  $P \ll Q$ , then  $\chi^2(P \parallel Q) = \infty$

## Proposition (Properties of f-divergences)

① Non-negativity.  $D_f(P \parallel Q_f) \geq 0$  w/ equality iff  $P = Q_f$

② Convexity:  $(P, Q_f) \mapsto D_f(P \parallel Q_f)$  is jointly convex. In particular,  $P \mapsto D_f(P \parallel Q_f)$  is convex for fixed  $Q_f$ , and  $Q_f \mapsto D_f(P \parallel Q_f)$  is convex for fixed  $P$ .

③ Conditioning Increased f-Divergences

Let  $P_{Y|X}$  and  $Q_{Y|X}$  be transition kernels and  $P_x \in \mathcal{P}(X)$ . Define the conditional f-divergence

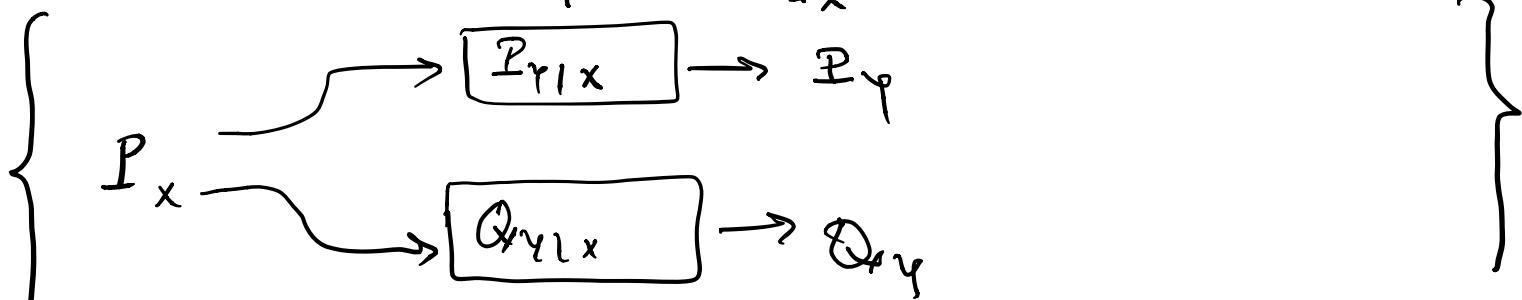
$$D_f(P_{Y|X} \parallel Q_{Y|X} | P_x) := \int_x D_f(P_{Y|X}(\cdot|x) \parallel Q_{Y|X}(\cdot|x)) dP_x(x)$$

$$= \mathbb{E}_{P_x} [D_f(P_{Y|X}(\cdot|x) \parallel Q_{Y|X}(\cdot|x))]$$

and recall

$$P_y(\cdot) = \mathbb{E}_{P_x} [P_{Y|X}(\cdot|x)]$$

$$Q_y(\cdot) = \mathbb{E}_{P_x} [Q_{Y|X}(\cdot|x)]$$



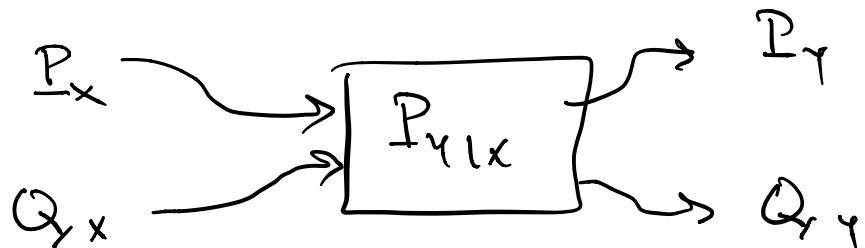
then

$$D_f(P_\gamma \| Q_\gamma) = D_f(\mathbb{E}_{P_x}[P_{\gamma|x}(\cdot|x)] \| \mathbb{E}_{P_x}[Q_{\gamma|x}(\cdot|x)]) \\ \leq D_f(P_{\gamma|x} \| Q_{\gamma|x} | P_x)$$

f-divergence  
Grows!

④ Let  $P_{x\gamma} = P_x P_{\gamma|x}$

$$Q_{x\gamma} = Q_x P_{\gamma|x}$$



$$D_f(P_{x\gamma} \| Q_{x\gamma}) = D_f(P_x \| Q_x)$$