

## Announcements

- ① Scribes for lecture notes
  - ② HW2 released Friday
  - ③ Next Thursday → Examples all lecture from Kia

## Recap

## f-divergences

$$\text{Def: } D_f(P||Q) = \mathbb{E}_Q f\left(\frac{dP}{dQ}\right) = \int f\left(\frac{dP}{dQ}(x)\right) dQ(x)$$

where  $f$  is convex w/  $f(0) = 0$

$f$  strictly convex around 1

Examples: KL, TV,  $\mathcal{K}^2$ -divergence

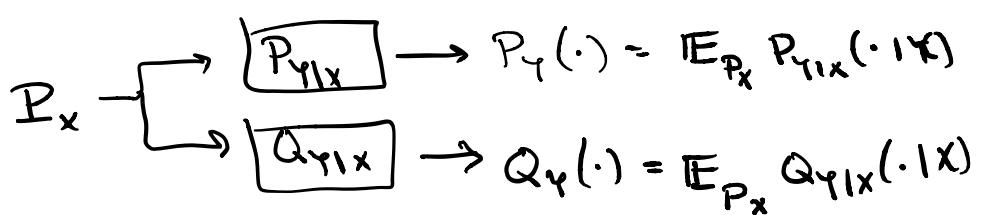
$$x \log x \quad \frac{1}{2}|x-1| \quad x^2-1 \quad (x-1)^2$$

# Properties

(i) Non-negativity :  $D_f(P||Q) \geq 0$  w/ equality iff  $P = Q$

(ii) Convexity  $(P, Q) \mapsto D_f(P||Q)$  is jointly convex

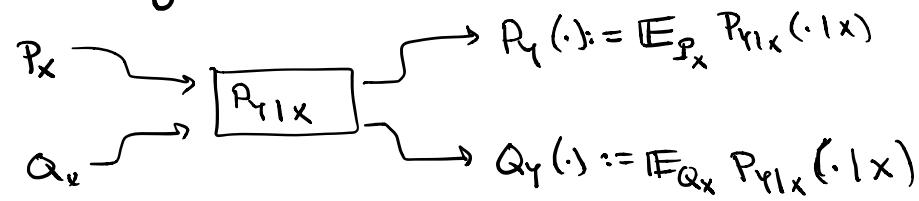
### (iii) Conditioning



$$D_f(P_Y || Q_Y) \leq D_f(P_{Y|X} || Q_{Y|X}|P_X)$$

$\stackrel{:=}{=} \int D_f(P_{Y|X}(\cdot|x) || Q_{Y|X}(\cdot|x)) dP_X(x)$

(iv) Joint vs. Marginal



$$D_f(P_X \| Q_X) = D_f(P_{X,Y} \| Q_{X,Y})$$

$P_X P_{Y|X}$        $Q_X Q_{Y|X}$

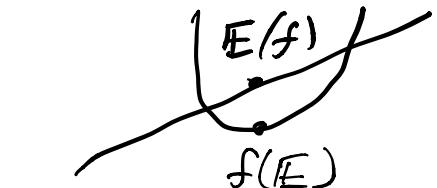
Sketch of Proof

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

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

$$\geq f(1)$$

$$= 0$$



Jensen's  
Inequality

equality iff follows by strict convexity

(ii) For any convex  $f$ , we may define "the perspective of  $f$ " which is given by

$$(x, y) \mapsto y f\left(\frac{x}{y}\right)$$

If  $f$  is convex, then perspective of  $f$  is jointly convex.

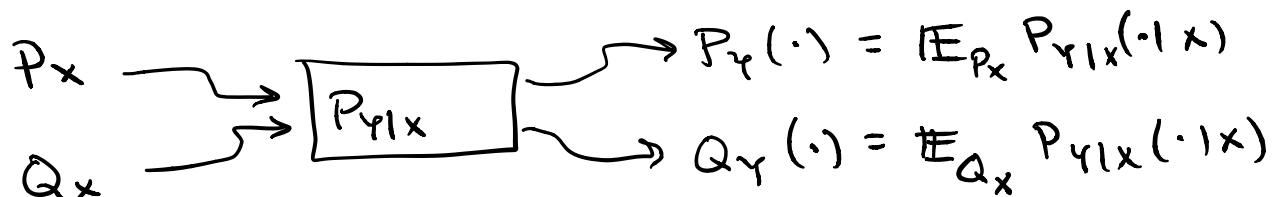
Recalling that  $D_f(P \| Q) = \int f\left(\frac{dP}{dQ}\right) dQ$  and combining w/ the above leads to the conclusion.

(iii)

$$\begin{aligned} D_f(P_Y \parallel Q_Y) &= D_f(\mathbb{E}_{P_X}[P_{Y|X}(\cdot|x)] \parallel \mathbb{E}_{P_X}[Q_{Y|X}(\cdot|x)]) \\ &\leq \mathbb{E}_P D_f(P_{Y|X}(\cdot|x) \parallel Q_{Y|X}(\cdot|x)) \\ &= D_f(P_{Y|X} \parallel Q_{Y|X} | P_X) \end{aligned}$$

$$\begin{aligned} (\text{iv}) \quad D_f(P_{X,Y} \parallel Q_{X,Y}) &= \mathbb{E}_{Q_{X,Y}} f\left(\frac{dP_{X,Y}}{dQ_{X,Y}}\right) \\ &= \mathbb{E}_{Q_{X,Y}} f\left(\frac{dP_X dP_{Y|X}}{dQ_X dP_{Y|X}}\right) \\ &= \mathbb{E}_{Q_X} f\left(\frac{dP_X}{dQ_X}\right) - D_f(P_X \parallel Q_X) \end{aligned}$$

### Data Processing Inequality



Theorem: Let  $P_X, Q_X \in \mathcal{P}(\mathcal{X})$  and  $P_{Y|X}$  be a transition kernel.  
Then:

$$\underline{D_f(P_X \parallel Q_X) \geq D_f(P_Y \parallel Q_Y)}$$

conditioning does NOT increase

## Examples

### ① Gaussian Convolutions

$$X \sim P_x \quad X' \sim Q_x \quad \text{and} \quad \begin{array}{c} \downarrow \text{iid} \\ Z \sim N(0, \sigma^2 \text{Id}) \perp\!\!\!\perp X, X' \\ Z' \sim N_\sigma \end{array}$$

Let

$$Y = X + Z \quad \text{and} \quad Y' = X' + Z'$$

(where the transition kernel is  
 $P_{Y|X}(\cdot|x) = N(x, \sigma^2 \text{Id})$ )

Recall:

$$\text{if } A \sim \mu \perp\!\!\!\perp B \sim \nu$$

$$\text{then } A+B \sim \mu * \nu \text{ where}$$

elaborated below

$$(\mu * \nu)(A) = \boxed{\iint_A \mathbb{1}_{\{x+y\}} d\mu(x) d\nu(y)}$$

$$\Rightarrow Y \sim P_x * N_\sigma \quad \text{and} \quad Y' \sim Q_x * N_\sigma$$

$$\Rightarrow \underline{\text{DPI}}: D_f(P_x \| Q_x) \geq D_f(\underbrace{P_x * N_\sigma}_{Y} \| \underbrace{Q_x * N_\sigma}_{Y'})$$

$$\left. \begin{array}{l} \mu \text{ has pdf } f \\ \nu \text{ has pdf } g \end{array} \right\} f * g$$

$\mu * \nu$  is probability measure when pdf is  $f * g$

## ② Deterministic Functions

$X \sim P_x$ ,  $X' \sim Q_x$  and set  $Y = g(X)$  for a deterministic measurable  $g$ .

$$\left\{ \begin{array}{l} \text{here } P_{Y|X}( \cdot | x) = \delta_{g(x)}, \text{ where } \delta_a \text{ is the} \\ \text{Dirac measure at } a: \delta_a(A) = \begin{cases} 1, & a \in A \\ 0, & \text{o/w} \end{cases} \\ \text{Also notice } \delta_a(A) = \mathbb{1}_A(a) \end{array} \right.$$

(i) Take any event  $E$  and set

$$Y = \mathbb{1}_{\{X \in E\}}$$

$$\bullet Y \text{ is a binary rv w/ } P_Y(1) = \mathbb{E}_{P_X} [\mathbb{1}_{\{X \in E\}}]$$

$$= P_X(E)$$

$$\Rightarrow Y = \mathbb{1}_{\{X \in E\}} \sim \text{Ber}(P_X(E))$$

$$Y' = \mathbb{1}_{\{X' \in E\}} \sim \text{Ber}(Q_X(E))$$

$\Rightarrow \text{DPI: For any measurable } E$

$$D_f(P_X || Q_X) \geq D_f(\text{Ber}(P_X(E)) || \text{Ber}(Q_X(E)))$$

easy  to compute  $\rightarrow$  allows for optimizations

(ii) When  $X = (X_1, X_2) \sim P_{X_1, X_2}$

$$X' = (X'_1, X'_2) \sim Q_{X_1, X_2}$$

and  $Y = g(X_1, X_2) = X_1$  and  $Y' = g(X'_1, X'_2) = X'_1$

$$Y \sim P_{X_1} \quad Y' \sim Q_{X_1}$$

$$\implies \stackrel{\text{DPI}}{D_f(P_{X_1, X_2} || Q_{X_1, X_2})} \geq D_f(P_{X_1} || Q_{X_1})$$

Proof DPI prop (i)

$$D_f(P_X || Q_X) = D_f(P_{XY} || Q_{XY}) = \mathbb{E}_{Q_{XY}} \left[ f\left(\frac{dP_{XY}}{dQ_{XY}}\right) \right]$$

Law of Total Probability

$$= \mathbb{E}_{Q_Y} \left[ \mathbb{E}_{Q_{X|Y}} \left[ f\left(\frac{dP_{XY}}{dQ_{XY}}\right) | Y \right] \right]$$

Jensen's

$$\geq \mathbb{E}_{Q_Y} \left[ f\left( \underbrace{\mathbb{E}_{Q_{X|Y}} \left[ \frac{dP_{XY}}{dQ_{XY}} | Y \right]}_{\textcircled{*}} \right) \right]$$

look at  $\textcircled{*}$

$$\begin{aligned} &= \mathbb{E}_{Q_{X|Y=y}} \left[ \frac{dP_{XY}}{dQ_{XY}} \mid Y=y \right] = \mathbb{E}_{Q_{X|Y=y}} \left[ \frac{\frac{dP_Y(y)P_{X|Y}(x|y)}{dQ_Y(y)Q_{X|Y}(x|y)}}{\frac{dQ_Y(y)Q_{X|Y}(x|y)}{dQ_Y(y)Q_{X|Y}(x|y)}} \mid Y=y \right] \\ &= \mathbb{E}_{P_{X|Y=y}} \left[ \frac{\frac{dP_Y(y)}{dQ_Y(y)}}{\frac{dQ_Y(y)}{dQ_Y(y)}} \mid Y=y \right] \quad \left\{ \begin{array}{l} \sum_{x \in X} \cancel{g_{X|Y}(x|y)} \frac{P_Y(y) P_{X|Y}(x|y)}{\cancel{g_Y(y)} \cancel{g_{X|Y}(x|y)}} \\ = \sum_{x \in X} P_{X|Y}(x|y) \frac{P_Y(y)}{g_Y(y)} = \frac{P_Y(y)}{g_Y(y)} \end{array} \right. \\ &= \frac{dP_Y(y)}{dQ_Y(y)} \end{aligned}$$

$$\Rightarrow \text{from } \textcircled{*} \quad h(Y) = \frac{dP(Y)}{dQ(Y)}$$

$$h(Y) = \frac{dP(Y)}{dQ(Y)} \quad ?$$

$$D_f(P_X || Q_X) \geq \mathbb{E}_{Q_Y} \left[ f \left( \frac{dP_Y(Y)}{dQ_Y(Y)} \right) \right]$$

$$= D_f(P_Y || Q_Y)$$

Q.E.D