08/11/2016
Tuesday

Total Variation:

$$\|P-Q\|_{TV} = \max_{A\in\mathcal{F}} [P(A)-Q(A)]$$

Properties:

1) Continuity of expectation: (Most useful Property)

If $f: \mathcal{X} \to [B_L, B_u]$ then

$$|\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f(X)]| \leq (B_u - B_L)\,\|P-Q\|_{TV}$$

Example: $f(x) = 1_{\{X\in A\}}$

$$\Rightarrow |P(A) - Q(A)| \leq \|P(A) - Q(A)\|_{TV}$$

2 a. Relation to ROC (detection) (Next most useful property)
↪ receiver operator characteristic.

$\alpha+\beta \geq 1 - \|P-Q\|_{TV}$

$\alpha+\beta \geq 1 - \|P-Q\|_{TV}$

Proof

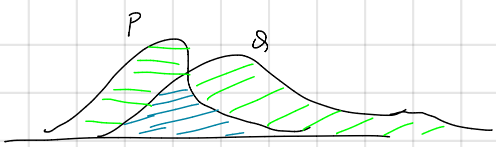
$\alpha = P(B)$ $\quad$ B is decision region

$$\begin{aligned}\beta &= Q(B^c)\\ &= 1 - Q(B)\\ &\geq 1 - P(B) - \|P-Q\|_{TV}\\ &= 1 - \alpha - TV\end{aligned}$$

2 b. $D(P\|Q)$ also related to ROC →

3. Coupling: $X = Y$

$$\min_{\substack{P_{XY}\\ P_X = P\\ P_Y = Q}} \mathbb{P}[X\neq Y] = \|P-Q\|_{TV}$$



4. Marginal is closer than Joint:

$$\|P_X - Q_X\|_{TV} \leq \|P_{XY} - Q_{XY}\|_{TV}$$

Proof 1: LHS limits detection rules to only X whereas the RHS uses X and Y

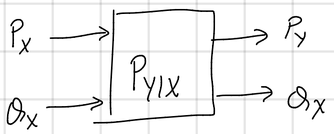
Proof 2: $\|P_{XY} - Q_{XY}\|_{TV} = \frac{1}{2}\sum_{xy} |P_{XY} - Q_{XY}| \geq \frac{1}{2}\sum_{x} \Big| \underbrace{\sum_{x} P_{XY}(x,y) - Q_{XY}(x,y)}_{P_Y - Q_Y} \Big|$

5. Same conditional:

If $P_{Y|X} = Q_{Y|X}$ then $\|P_X - Q_X\|_{TV} = \|P_{XY} - Q_{XY}\|_{TV}$ → see the proof above $P_{Y|X}$ goes out and sums to 1 $\forall x$

6. Data Processing Inequality:

If $P_{Y|X} = Q_{Y|X}$ $\quad \|P_Y - Q_Y\|_{TV} \leq \|P_X - Q_X\|_{TV}$



7. Same Marginal Distribution

If $P_X = Q_X$ then $\|P_{XY} - Q_{XY}\|_{TV} = \mathbb{E}\left[\|P_{Y|X}(\cdot|X) - Q_{Y|X}(\cdot|X)\|_{TV}\right]$

Compare to $D(P\|Q)$

Chain rule $D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + \mathbb{E}_{P_X}\left[D(P_{Y|X=\bar{X}} \| Q_{Y|X=\bar{X}})\right]$

where $\bar{X} \sim P_X$

$$= D(P_Y\|Q_Y) + \mathbb{E}_{P_{\bar{Y}}}\left[D(P_{X|Y=\bar{Y}} \| Q_{X|Y=\bar{Y}})\right]$$

→ 5, 6, 7 hold for $D(P_X\|Q_X)$ as well ↗

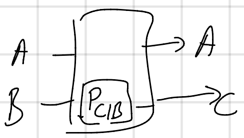
D.P.I.: If $P_{Y|X} = Q_{Y|X}$ then $D(P_Y \| Q_Y) \leq D(P_X \| Q_X)$

D.P.I.: $A - B - C \Rightarrow I(A;C) \leq I(A;B)$

$$D(P_{AC} \| P_A P_C)$$

$P_X = P_{AB}$

$Q_X = P_A P_B$



$P_Y = P_{A,C}$

$Q_Y = P_A P_C$

Pinsker's Inequality:

$$\|P - Q\|_{TV} \leq \sqrt{\tfrac{1}{2} D(P\|Q)} \quad \text{in nats}$$

Other direction: (?) (Not without conditions)

If Q_{X^n} is iid and $|\mathcal{X}| < \infty$, $P_{X^n} \ll Q_{X^n}$

Then $D(P_{X^n} \| Q_{X^n}) \in O\left(\left(n + \log \frac{1}{\|P_{X^n} - Q_{X^n}\|_{TV}}\right) \|P_{X^n} - Q_{X^n}\|_{TV}\right)$

proof of other direction.

$$D(P_{X^n} \| Q_{X^n}) = \mathbb{E}_P\left[\log \frac{P_{X^n}(X^n)}{Q_{X^n}(X^n)}\right] = \mathbb{E}_P\left[\log \frac{1}{Q(X^n)}\right] - H(P_{X^n})$$

$$= \sum_i \mathbb{E}\Big[\underbrace{\log \frac{1}{Q(X_i)}}_{\text{bdd}}\Big] - H(P_{X^n})$$

$$\leq \sum_i \left[\mathbb{E}_Q\left[\log \frac{1}{Q(X_i)}\right] + \log \frac{1}{\min_x Q(x)} \|P - Q\|_{TV}\right] - H(P_{X^n})$$

So

$$D(P_{X^n} \| Q_{X^n}) \leq \left[\underbrace{H(Q_{X^n}) - H(P_{X^n})} + n \log \frac{1}{\min_x Q(x)} \|P - Q\|_{TV}\right]$$

$\|P - Q\|_{TV} \leq \varepsilon \Rightarrow |H(Q) - H(P)| \leq \varepsilon \log |\mathcal{X}| + h(\varepsilon)$ (A generalization of Fano's ineq.)

In Wiretap channel,

Secrecy Objective: $\|P_{MZ^n} - P_M P_{Z^n}\|_{TV} < \varepsilon$ will have powerful consequences

The weakness: $\frac{1}{2^{nR}} \sum_m \|P_{Z^n|M=m} - P_{Z^n}\|_{TV} < \varepsilon$

Again secure on average.