TVD

$$\|\mu-\nu\|_{TV} := \sup_{A\in\Omega} |\mu(A)-\nu(A)| \overset{①}{=} \sup_{h:\mathcal{X}\to[0,1]} \left|\int h\,d\mu - \int h\,d\nu\right|$$

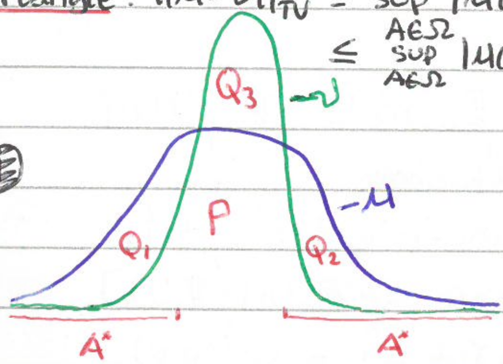
$$\overset{②}{=} \frac{1}{2} \sup_{h:\mathcal{X}\to[-1,1]} \left|\int h\,d\mu - \int h\,d\nu\right| \overset{③}{=} \inf_{\gamma\in\Gamma}\{Pr(X\neq Y) : (X,Y)\sim\gamma\}$$

when Γ = set of joints on (X,Y) s.t. $X\sim\mu$, $Y\sim\nu$ marginally.

METRIC: $D(\mu,\nu)\geq 0$; $D(\mu,\nu)=0 \Leftrightarrow \mu=\nu$; $D(\mu,\nu)=D(\nu,\mu)$; $D(\mu,\nu)\leq D(\mu,\rho)+D(\rho,\nu)$.

Triangle: $\|\mu-\nu\|_{TV} = \sup_{A\in\Omega} |\mu(A)-\rho(A)+\rho(A)-\nu(A)| \leq \sup_{A\in\Omega} |\mu(A)-\rho(A)| + |\rho(A)-\nu(A)|$

$\leq \sup_{A\in\Omega} |\mu(A)-\rho(A)| + \sup_{A\in\Omega} |\rho(A)-\nu(A)| = \|\mu-\rho\|_{TV} + \|\rho-\nu\|_{TV}$.



$P+Q_1+Q_2 = 1 = P+Q_3 \Rightarrow Q_1+Q_2 = Q_3 =: \|\mu-\nu\|_{TV}$

$A^* = \arg\max_{A\in\Omega} |\mu(A)-\nu(A)| = A^{*c}$

① $h(x) = \mathbb{1}(x\in A)$

② $h(x) = \mathbb{1}(x\in A) - \mathbb{1}(x\in A^c)$

③ Need to show $1-\|\mu-\nu\|_{TV} = \sup_{\gamma\in\Gamma}\{Pr(X=Y) : (X,Y)\sim\gamma\}$; but this is P.

Wasserstein

$$W_p(\mu,\nu) = \inf_{\gamma\in\Gamma}\left[\int_{\mathcal{X}^2} D(x,y)^p\,\gamma(dx,dy)\right]^{1/p} = \inf_{\gamma\in\Gamma} \mathbb{E}_\gamma[D(X,Y)^p]^{1/p}$$ any metric, D.

Triangle for W_1: $\inf_{\gamma\in\Gamma} \mathbb{E}_\gamma[D(X,Y)] = \inf_{\gamma\in\tilde\Gamma} \mathbb{E}[D(X,Y)] \leq \inf_{\gamma\in\tilde\Gamma} \mathbb{E}[D(X,Z)+D(Z,Y)]$

$\leq \inf_{\gamma\in\tilde\Gamma} \mathbb{E}[D(X,Z)] + \inf_{\gamma\in\tilde\Gamma} \mathbb{E}[D(Z,Y)] = \inf_{\gamma\in\Gamma_1} \mathbb{E}[D(X,Z)] + \inf_{\gamma\in\Gamma_2} \mathbb{E}[D(Z,Y)]$.

where $\tilde\Gamma$ is set of joints for (X,Y,Z) with marginals μ,ν & ρ.

Γ_1 . . . . (X,Z) with margs μ & ρ & Γ_2 is joints for (Z,Y) with margs ρ,ν.

- For $p_1<p_2$, $W_{p_1}(\mu,\nu)\leq W_{p_2}(\mu,\nu)$.

Pf Jensen $\Rightarrow \mathbb{E}[T]^{p_2/p_1} \leq \mathbb{E}[T^{p_2/p_1}]$; so $\mathbb{E}_\gamma[D(X,Y)^{p_1}]^{\frac{p_2}{p_1}} \leq \mathbb{E}_\gamma[D(X,Y)^{p_2}]$. So convergence in $p_1 \Rightarrow$ cvge in p_2.

$\|\mu-\nu\|_{TV} = W_p(\mu,\nu)$ when $D(X,Y) = D_0(x,y) := \mathbb{1}_{\{x\neq y\}}$.

Pf $\inf_{\gamma\in\Gamma} \mathbb{E}[\mathbb{1}(X\neq Y)] = \inf_{\gamma\in\Gamma} P(X\neq Y) = \|\mu-\nu\|_{TV}$ by ③.

If $D(X,Y)\leq\bar{D}$ then $W_1(\mu,\nu)\leq\bar{D}\,\|\mu-\nu\|_{TV}$.

Pf $D(X,Y)\leq\bar{D}\,\mathbb{1}(X\neq Y)$ $\therefore$ $\mathbb{E}_\gamma[D(X,Y)]\leq\bar{D}\,P_\gamma(X\neq Y)$. Take inf $\gamma\in\Gamma$.