Probability Metrics

We've seen the total variation metric on $\mathrm{Prob}(S,\mathcal{B})$

$$d_{TV}(\mu,\nu) = \sup_{B\in\mathcal{B}} |\mu(B)-\nu(B)|$$

$$= \frac{1}{2} \sup_{h\in\mathbb{B}_1(S,\mathcal{B})} \left|\int h\,d\mu - \int h\,d\nu\right|$$

bounded $\mathcal{B}$-measurable w/ $\sup|h|\leq 1$

This is an example of a (dual) probability metric:

$$d(\mu,\nu) = \sup_{h\in\mathcal{H}} \left|\int h\,d\mu - h\,d\nu\right|$$

some class of functions on S.

Always a pseudo-metric; genuine metric if $\mathcal{H}$ is sufficiently rich.

too strong for weak convergence

Eg. Kolmogorov metric: $d_{Kol}(\mu,\nu) = \sup_{t\in\mathbb{R}} |F_\mu(t) - F_\nu(t)|$

(on $\mathrm{Prob}(\mathbb{R},\mathcal{B}(\mathbb{R}))$)

$$= \sup_{t\in\mathbb{R}} \left|\int_{\mathbb{R}} \mathbb{1}_{(-\infty,t]}\,d\mu - \int_{\mathbb{R}} \mathbb{1}_{(-\infty,t]}\,d\nu\right|$$

Wasserstein L^1 Distance

On $\mathrm{Prob}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, define

$$d_{W_1}(\mu,\nu) = \sup_{\|f\|_{Lip}\le 1} \left| \int f\,d\mu - \int f\,d\nu \right|$$

i.e. $\mathcal{H} = Lip_1(\mathbb{R})$

$= \{ f : \mathbb{R} \to \mathbb{R} : |f(x) - f(y)| \le |x-y| \ \forall x,y \in \mathbb{R} \}$

It's a metric (by the Portmanteau theorem)

$\|f\|_{Lip} = \sup |f'|$

That doesn't look like what we called the Wasserstein metric in [Lecture 21.2]. But it actually is.

Prop:

$$d_{W_1}(\mu,\nu) = \sup_{\|f\|_{Lip}\le 1} \left| \int f\,d\mu - \int f\,d\nu \right|$$

$$= \int_{-\infty}^{\infty} |F_\mu(t) - F_\nu(t)|\,dt$$

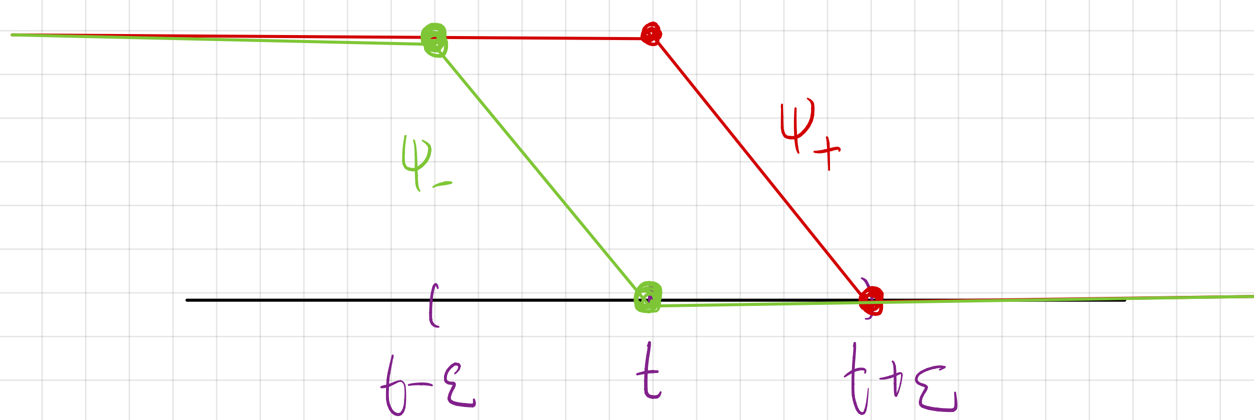
$$= \inf\{ \mathbb{E}[|X-Y|] : (X,Y) \text{ is a coupling of } \mu,\nu \}$$

The Wasserstein distance controls the Kolmogorov distance – at least when one of the measures has a bounded density.

Prop: If $d\nu = \rho\, dx$ and $\rho \leq C < \infty$, then for any $\mu \in \mathrm{Prob}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$d_{Kol}(\mu,\nu) \leq 2\sqrt{C\, d_{W_1}(\mu,\nu)}.$$

Pf. Fix $t \in \mathbb{R}$, $\varepsilon > 0$, and define two continuous approximations to $\mathbb{1}_{(-\infty,t]}$:



$\leftarrow$ slope $\frac{1}{\varepsilon}$

$\|\psi_\pm\|_{Lip} \leq \frac{1}{\varepsilon}.$

$$\|\varepsilon f\|_{Lip} = \sup_{x \neq y} \frac{|\varepsilon f(x) - \varepsilon f(y)|}{|x-y|} = \varepsilon \|f\|_{Lip}$$

$\|\varepsilon\psi_\pm\| \leq 1.$

$$\psi_- \leq \mathbb{1}_{(-\infty,t]} \leq \psi_+$$

$$\therefore \int \underbrace{\mathbb{1}_{(-\infty,t]}}_{\leq \psi_+} d\mu - \int \mathbb{1}_{(-\infty,t]}\, d\nu$$

$$\leq \int \psi_+\, d\mu - \int \psi_+\, d\nu + \int \psi_+\, d\nu - \int \mathbb{1}_{(-\infty,t]}\, d\nu$$

$$\mu(-\infty,t] - \nu(-\infty,t]$$

$$\leq \int \varphi_+ \, d\mu - \int \varphi_+ \, d\nu + \underbrace{\int \underbrace{(\varphi_+ - \mathbb{1}_{(-\infty,t]})}_{\leq\, \mathbb{1}_{(t,t+\varepsilon]}} \underbrace{d\nu}_{\leq\, \underbrace{\rho\, dx}_{\leq\, C}}}_{\leq\, C\varepsilon}$$

$$= \frac{1}{\varepsilon}\left(\int \varepsilon\varphi_+ \, d\mu - \int \varepsilon\varphi_+ \, d\nu \right)$$

$$\leq \frac{1}{\varepsilon} d_{W_1}(\mu,\nu)$$

$\therefore$ $\mu(-\infty,t] - \nu(-\infty,t] \leq \frac{1}{\varepsilon} d_{W_1}(\mu,\nu) + C\varepsilon$, for all $\varepsilon > 0$.

optimize over $\varepsilon > 0$. $\varepsilon = \sqrt{\frac{d_{W_1}(\mu,\nu)}{C}}$

$\therefore$ $\leq 2\sqrt{C\, d_{W_1}(\mu,\nu)}$

Now use φ_- to prove the reverse ineq. for $\nu(-\infty,t] - \mu(-\infty,t]$.

$$d_{Kol}(\mu,\nu) = \therefore \sup_{t\in\mathbb{R}} |\mu(-\infty,t] - \nu(-\infty,t]| \leq 2\sqrt{C\, d_{W_1}(\mu,\nu)}$$

///

As usual, we apply a metric on measures to random variables by $d(X,Y) := d(\mu_X, \mu_Y)$.

Cor: If $Z \overset{d}{=} \mathcal{N}(0,1)$, then for any random variable X,

$$d_{Kol}(X,Z) \leq 2\sqrt{d_{W_1}(X,Z)}$$

b/c $\rho_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$\frac{1}{\sqrt{2\pi}} \leq 1$

One of our goals is to prove (a version of) the:

Berry-Esseen Theorem:

Let $\{X_n\}_{n=1}^{\infty}$ be iid. L^3 random variables, with $\mathbb{E}[X_j]=0$, $\mathbb{E}[X_j^2]=1$, $\mathbb{E}[|X_j|^3]=\sigma^3$.

Let $S_n = X_1 + \cdots + X_n$. If $Z \overset{d}{=} \mathcal{N}(0,1)$, then

$$d_{Kol}\left(\frac{S_n}{\sqrt{n}}, Z\right) \leq C \frac{\sigma^3}{\sqrt{n}} \leftarrow \text{rate = optimal}$$

$\sup_{t\in\mathbb{R}} |F_{S_n/\sqrt{n}}(t) - F_Z(t)|$

$\Phi(t) = \mathrm{Erf}(t)$

best known $C < 0.4748$

0.41