Total variation distance

Definition (3.1)

Consider μ, ν probabilities on E, we define the total variation distance between μ and ν as

$$||\mu - \nu||_{TV} = \sup_{A \subset E} |\mu(A) - \nu(A)|.$$

It is straightforward that $||\cdot||_{TV}$ defines a distance on $\mathcal{P}(E)$, the set of probabilities on E.

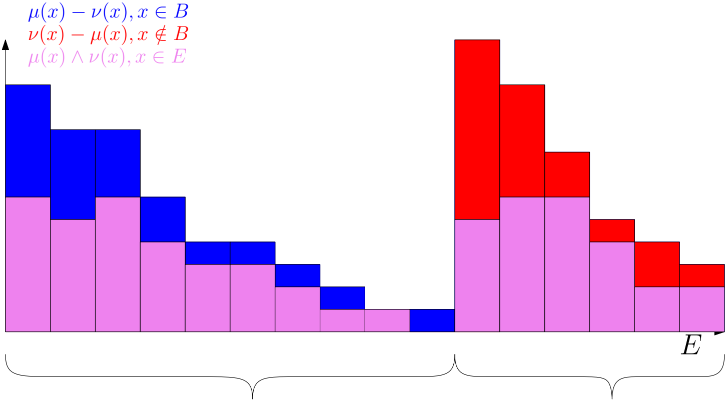
Theorem (3.1.1)

Assume E is finite or countable. Then

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$$

$$= \sum_{\{x : \mu(x) \geq \nu(x)\}} (\mu(x) - \nu(x)) = 1 - \sum_{x \in E} (\mu(x) \wedge \nu(x))$$

Total variation distance : illustration



$B := \{x : \mu(x) \geq \nu(x)\}$ B^c

$$||\mu - \nu||_{TV} = \frac{1}{2} \Sigma_{x \in E} |\mu(x) - \nu(x)| = \Sigma_{x \in B} (\mu(x) - \nu(x)) = 1 - \Sigma_{x \in E} (\mu(x) \wedge \nu(x)).$$

Proof of theorem (3.1.1)

Recall $B = \{x : \mu(x) \geq \nu(x)\}$. For any $A \subset E$, we easily see that

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B),$$

so B realizes the supremum of $\mu(A) - \nu(A), A \subset E$. Similarly $\mu(A) - \nu(A) \geq \mu(B^c) - \nu(B^c)$, so B^c realizes the supremum of $\nu(A) - \mu(A), A \subset E$. Since $\mu(B) + \mu(B^c) = \nu(B) + \nu(B^c) = 1$ both B and B^c realize the supremum of $|\mu(A) - \nu(A)|, A \subset E$. Moreover,

$$|\mu(B) - \nu(B)| = |\mu(B^c) - \nu(B^c)| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|,$$

and the first two equalities of the theorem. Finally $\sum_{x \in E}(\mu(x) \wedge \nu(x)) = \nu(B) + \mu(B^c)$ so that

$$1 - \sum_{x \in E}(\mu(x) \wedge \nu(x)) = 1 - \mu(B^c) - \nu(B) = \mu(B) - \nu(B),$$

finishing the proof.

Total variation distance and coupling

Recall (X, Y) is a coupling of μ and ν as soon as $X \sim \mu, Y \sim \nu$.

Theorem (3.1.2)

$$||\mu - \nu||_{TV} = \inf \{\mathbb{P}(X \neq Y) \;:\; (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\}$$

Proof of Theorem (3.1.2) : First observe that if (X, Y) is a coupling of μ and ν we have

$$\begin{aligned}\mathbb{P}(X \neq Y) &\geq \mathbb{P}(X \in B, Y \in B^c) \\ &= \mathbb{P}(X \in B) - \mathbb{P}(X \in B, Y \in B) \\ &\geq \mathbb{P}(X \in B) - \mathbb{P}(Y \in B) = \mu(B) - \nu(B)\end{aligned}$$

thus for any coupling (X, Y), $\mathbb{P}(X \neq Y) \geq ||\mu - \nu||_{TV}$.
Note that the inequality above becomes an equality provided $\{Y \in B\} \subset \{X \in B\}$.

To finish the proof of the theorem it is enough to find an optimal coupling ensuring $\{Y \in B\} \subset \{X \in B\}$. Let $p = 1 - ||\mu - \nu||_{TV} = \sum_{x \in E} (\mu(x) \wedge \nu(x))$. If $p = 1$, $\mu = \nu$ and we can always take $X = Y$. If $p = 0$, μ and ν have disjoint support, so any coupling works.
Otherwise let $\gamma_0(x) := \frac{\mu(x) \wedge \nu(x)}{p}, x \in E$, a probability on E.
$\gamma_1(x) := \frac{\mu(x) - \nu(x)}{1-p}, x \in B$ a probability on B,
$\gamma_2(x) := \frac{\nu(x) - \mu(x)}{1-p}, x \in B^c$, a probability on B^c,
Define $Z_0 \sim \gamma_0, Z_1 \sim \gamma_1, Z_2 \sim \gamma_2$, $\xi \sim \mathrm{Ber}(p)$ (independently) and

$$X = \xi Z_0 + (1 - \xi) Z_1, \quad Y = \xi Z_0 + (1 - \xi) Z_2.$$

Now (X, Y) is a coupling of μ and ν and $\mathbb{P}(X \neq Y) = \mathbb{P}(\xi = 0) = 1 - p$, as required.

Further considerations

- When E is metric, at most countable, and has no accumulation point (e.g. $E = \mathbb{N}, \mathbb{Z}, \mathbb{Z}^d$), it can be shown that the topology of total variation distance is equivalent to that of convergence in distribution. See exercise III.3 for $E = \mathbb{Z}$.
- Working with variation distance allows for an elegant proof of a relatively general version of the *rare events* theorem (see exercises III.3 to III.7). More precisely if for any $n \in \mathbb{N}$, $(\xi_{n,m})_{0 \leq m \leq n}$ are independent Bernoulli with respective parameters $\{p_{n,m}, 0 \leq m \leq n\}$ satisfying

$$\lambda_n = \sum_{m=0}^{n} p_{n,m} \underset{n\to\infty}{\longrightarrow} \lambda, \qquad \max_{0 \leq m \leq n} p_{n,m} \underset{n\to\infty}{\longrightarrow} 0,$$

then

$$\sum_{m=0}^{n} \xi_{n,m} \underset{n\to\infty}{\overset{\text{(law)}}{\longrightarrow}} Z \sim \mathrm{Poisson}(\lambda).$$