Convergence Revisited

We've considered several modes of convergence of random variables:

$X_n \to X$ a.s.

$X_n \to X$ in L^p

$X_n \to_p X$

All of these require information about the joint distribution of $\{X, X_n\}$.

We're now going to turn to some convergence notions that only care about the individual distributions.

Total Variation

Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of probability measures on $(S, \mathcal{B})$.
They're just $\mathbb{R}$-valued functions on $\mathcal{B}$, so we can use any function convergence notion we like.

Def: Let μ, ν be probability measures on $(S, \mathcal{B})$. The total variation distance between them is

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

If X, Y are $(S, \mathcal{B})$-valued random variables, we set

$$d_{TV}(X, Y) = d_{TV}(\mu_X, \mu_Y)$$

$$= \sup_{B \in \mathcal{B}} |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)|$$

Lemma: (Scheffé) If α is a finite measure on $(S, \mathcal{B})$ such that $\mu, \nu \ll \alpha$ with $d\mu = u\,d\alpha$, $d\nu = v\,d\alpha$, then

$$d_{TV}(\mu, \nu) = \frac{1}{2} \| u - v \|_{L^1(\alpha)}.$$

Pf. For $B \in \mathcal{B}$,

$$|\mu(B) - \nu(B)|$$

$$|\mu(B^c) - \nu(B^c)|$$

Note: it is always possible to find such an α.

In fact: if $\{\mu_n\}_{n=1}^{\infty}$ is any countable collection of finite measures,

take $\alpha = \sum_{n=1}^{\infty} 2^{-n} \mu_n$.

Cor: d_{TV} is a complete metric on $\mathrm{Prob}(S, \mathcal{B})$.

Pf. (This can be shown directly, but the present approach is slicker.)

Let $\{\mu_n\}_{n=1}^{\infty} \subset \mathrm{Prob}(S, \mathcal{B})$. Fix α as above; then $d\mu_n = u_n \, d\alpha$.

- $0 = d_{TV}(\mu_1, \mu_2)$
- $d_{TV}(\mu_1, \mu_3)$
- If $\{\mu_n\}_{n=1}^{\infty}$ is d_{TV}-Cauchy,

$$d_{TV}(\mu_n, \mu_m) \to 0$$

Cor: If h is a bounded r.v. on $(S, \mathcal{B})$, then $\forall \mu, \nu \in Prob(S, \mathcal{B})$

$$\left| \int_S h\,d\mu - \int_S h\,d\nu \right| \leq 2\, d_{TV}(\mu, \nu) \cdot \sup_S |h| .$$

Moreover, $d_{TV}(\mu, \nu) = \frac{1}{2} \sup \left\{ \left| \int_S h\,d\mu - \int_S h\,d\nu \right| : \sup_S |h| \leq 1 \right\}$

Pf.

Total variation works well when S is countable.

Lemma: If S is countable, and $\mu, \nu \in Prob(S, \mathcal{B})$, then

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{k \in S} |\mu(\{k\}) - \nu(\{k\})|$$

$$\therefore \mu_n \to_{TV} \mu \text{ iff } \mu_n(\{k\}) \to \mu(\{k\}) \quad \forall k \in S.$$

Pf.

Eg. $\mu_p \overset{d}{=} \text{Bernoulli}(p)$.

$$d_{TV}(\mu_p, \mu_q)$$

Eg. $\nu_\lambda \overset{d}{=} \text{Poisson}(\lambda)$.

$$d_{TV}(\nu_\lambda, \nu_\eta)$$

Eg. $d_{TV}(\mu_p, \nu_p)$