Probability Theory

VI Total variation $\sim$ Skorohod's Theorem

Seongho, Joo

August 7, 2025

Seoul National University

Total variation

We've considered several modes of convergence of random variables: almost sure convergence, L^p convergence, and convergence in probability.

All of these require information about the *joint* distribution of $\{X, X_n\}$. We are going to turn to some convergence notions that only care about the individual distributions.

Total variation

We've considered several modes of convergence of random variables: almost sure convergence, L^p convergence, and convergence in probability.

All of these require information about the *joint* distribution of $\{X, X_n\}$. We are going to turn to some convergence notions that only care about the individual distributions. Let $\{\mu_n\}_{n=1}^{\infty}$ be s sequence of probability measures on $(S, \mathcal{B})$.

Definition 1

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

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

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

$$d_{\mathsf{TV}}(X, Y) = d_{\mathsf{TV}}(\mu_X, \mu_Y) = \sup_{B \in \mathcal{B}} |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)|$$

Total variation

Lemma 1 (Scheffé)

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

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

Note: it is always possible to find such α.

Proof.

Total variation

Corollary 1

d_{tv} is a complete metric on $\text{Prob}(S, \mathcal{B})$

Proof. Let $\{\mu_n\}_{n=1}^{\infty} \subset \text{Prov}(S, \mathcal{B})$. Fix α such that $\mathrm{d}\mu_n = u_n \, \mathrm{d}\alpha$.

- $0 = d_{\text{TV}}(\mu_1, \mu_2) \implies u_1 = u_2 \ \alpha - a.s. \implies \mu_1 = \mu_2$
- $d_{\text{TV}}(\mu_1, \mu_3) = \frac{1}{2}\|u_1 - u_3\| \leq \frac{1}{2}(\|u_1 - u_2\| + \|u_2 - u_3\|) =$
  $d_{\text{TV}}(\mu_1, \mu_2) + d_{\text{TV}}(\mu_2, \mu_3)$
- If $\{\mu_n\}_{n=1}^{\infty}$ is a d_{TV}-Cauchy,
  $$\frac{1}{2}\|u_n - u_m\|_{L^1(\alpha)} = d_{\text{TV}}(\mu_n, \mu_m) \to 0 \implies \{u_n\}_{n=1}^{\infty} \text{ is a Cauchy in } L^1(\alpha)$$
  Define $\mathrm{d}\mu = u \, \mathrm{d}\alpha$, $d_{\text{TV}}(\mu_m, \mu) = \frac{1}{2}\|u_n - u\|_{L^1(\alpha)} \to 0$

Total variation

Corollary 2

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

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

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

Proof.

Total variation

Total variation works well when S is countable.

Lemma 2

If S is countable, and $\mu, \nu \in \mathsf{Prov}(S, \mathcal{B})$, then

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

$$\implies \mu_n \to \mu \text{ in TV if and only if } \mu_n(\{k\}) \to \mu(\{k\}) \quad \forall k \in S$$

Example $\nu_\lambda = \mathsf{Poisson}(\lambda)$

$$d_{\mathsf{TV}}(\nu_\lambda, \nu_\eta) = \frac{1}{2} \sum_{k=0}^{\infty} \left| e^{\lambda} \frac{\lambda^k}{k!} - e^{-\eta} \frac{\eta^k}{k!} \right| \leq |\lambda - \eta|$$

Example $d_{\mathsf{TV}}(\mu_p, \nu_p) = p(1 - e^{-p})$

Law of Rare Events

Theorem 1 (The Law of Rare Events)

Let $\{X_j\}_{j=1}^{\infty}$ be independent, $X_j = \text{Bernoulli}(p_j)$. Set $S_n = X_1 + \cdots + X_n$. Let $N = \text{Poisson}(p_1 + \cdots + p_n)$ Then,

$$d_{\mathsf{TV}}\left(S_n, N\right) \leq \sum_{j=1}^{n} p_j^2$$

Lemma 3 (Sub-additivity of TV distance)

Let $\{\mu_j, \nu_j\}_{j=1}^{n}$ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ Then,

$$d_{\mathsf{TV}}\left(\mu_1 * \mu_2 \ldots \mu_n, \nu_1 * \nu_2 \cdots * \nu_n\right) \leq \sum_{j=1}^{n} d_{\mathsf{TV}}\left(\mu_j, \nu_j\right)$$

Coupling

Given probability measure μ, ν on $(S, \mathcal{B})$, a *coupling* is a pair (X, Y) of random variables on a common probability space, taking values in $(S, \mathcal{B})$, such that $\mu_X = \mu, \mu_Y = \nu$

Example: $\mu \otimes \nu$ is a independent coupling of μ, ν.

Lemma 4 (Coupling Estimate)

If (X, Y) is any coupling of the Borel probability measures μ, ν, then $d_{\mathsf{TV}}(\mu, \nu) \leq \mathbb{P}(X \neq Y)$

Proof.

TV distance is somewhat too Strong!

In non-discrete settings, total variation convergence is usually too much to ask.

- Let $a_n \in \mathbb{R}, a_n \to a$.

$$d_{\mathsf{TV}}\left(\delta_{a_n}, \delta_a\right) = \sup_B \left|\delta_{a_n}(B) - \delta(B)\right| \geq 1 \text{ i.o.}$$

- A discrete approximation of Unif($[0,1]$), $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{k/n}$

$$d_{\mathsf{TV}}\left(\mu_n, \mathsf{Unif}[0,1]\right) = 1$$

TV distance is somewhat too Strong!

In non-discrete settings, total variation convergence is usually too much to ask.

- Let $a_n \in \mathbb{R}, a_n \to a$.

$$d_{\mathsf{TV}}\left(\delta_{a_n}, \delta_a\right) = \sup_B |\delta_{a_n}(B) - \delta(B)| \geq 1 \text{ i.o.}$$

- A discrete approximation of Unif($[0,1]$), $\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{k/n}$

$$d_{\mathsf{TV}}\left(\mu_n, \mathsf{Unif}[0,1]\right) = 1$$

Definition 2

Let S be a metric space, $\mu_n, \mu \in \mathsf{Prob}(S, \mathcal{B}(S))$.
Say μ_n converges *weakly* to μ, $\mu_n \underset{w}{\to} \mu$, if $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu \; \forall f \in C_b(S)$

Example: If $a_n \to a$, then $\int f \, \mathrm{d}\delta_{a_n} = f(a_n) \to f(a) = \int f \, \mathrm{d}\delta_a \; \forall f \in C_b$

Weak Convergence

Proposition 1

If $d_{\mathsf{TV}}(\mu_n, \mu) \to 0$, then $\mu_n \underset{w}{\to} \mu$

Proof. Directly followed by Corollary 2.

Proposition 2

If $X_n, X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{B}(S))$ and $X_n \underset{\mathbb{P}}{\longrightarrow} X$, then $X_n \xrightarrow{w} X$.

Proof. First, we prove the following lemma.

Lemma 5

If $X_n \underset{\mathbb{P}}{\longrightarrow} X$ and $g \in C(S)$, then $g(X_n) \underset{\mathbb{P}}{\longrightarrow} g(X)$.

Proof. For $\varepsilon, \delta > 0$, let $B_{\varepsilon,\delta}(g) = \{x \in S : \exists y \in S \; d(x,y) < \delta, |g(x) - g(y)| \geq \varepsilon\}$ Continuity of g means that, for fixed $\varepsilon > 0$, $B_{\varepsilon,\delta}(g) \downarrow \emptyset$ as $\delta \downarrow 0$. Note that $\{|g(x_n) - g(X)| \geq \varepsilon\} \subseteq \{d(X_n, X) \geq \delta\} \cup \left\{X \in B_{\varepsilon,\delta}(G)\right\}$

$$\implies \mathbb{P}(|g(X_n) - g(X)| \geq \varepsilon) \leq \mathbb{P}(d(X_n, X) \geq \delta) + \underbrace{\mathbb{P}(X \in B_{\varepsilon,\delta}(g))}_{\mu_X(B_{\varepsilon,\delta}(g)) \to 0 \text{ as } \delta \downarrow 0}$$

Weak Convergence

Let $f \in C_b(S)$. Then

$$\int f \, \mathrm{d}\mu_{X_n} = \mathbb{E}\, f(X_n)$$

By the lemma, $X_n \underset{\mathbb{P}}{\longrightarrow} X \implies f(X_n) \underset{\mathbb{P}}{\longrightarrow} f(X)$ and note that $|f(X_n)| \leq M$

By dominated convergence theorem, $f(X_n) \to f(X)$ in L^1.

$\mathbb{E}\, f(X_n) \to \mathbb{E}\, f(X) = \int f \, \mathrm{d}\mu_X$

Corollary 3

If $X_n \to X$ a.s., or if $X_n \to X$ in L^p, then $X_n \xrightarrow{w} X$.

Definition 3

Let μ be a Borel measure on metric space S. An event $B \in \mathcal{B}(S)$ is a continuity set for μ if

$$\mu(\partial A) = 0, \quad \partial A = \bar{A} \setminus \text{int}(A)$$

Example. $(-\infty, a]$ is not a continuity set for δ_a.

Example. If $\mu \in \text{Prov}(\mathbb{R}, \mathcal{B}(\mathbb{R})$ and F_μ is continuous, then all intervals are continuity sets for μ

Conditions for Weak convergence

Theorem 2 (Portmanteau)

Let S be a complete, separable metric space. Let $\mu_n, \mu \in \mathsf{Prob}(S, \mathcal{B}(S))$. TFAE:

1. $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu \quad \forall f \in C_b(S)$
2. $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu \quad f \in \mathsf{Lip}_b(S)$
3. $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F) \quad \forall$ closed $F \subseteq S$
4. $\liminf_{n\to\infty} \mu_N(G) \geq \mu(G) \quad \forall$ open $G \subseteq S$
5. $\mu_n(A) \to \mu(A) \quad \forall \mu -$ continuity sets $A \in \mathcal{B}(S)$

Proof.

Weak convergence for $\mathbb{R}^d$

Theorem 3

Let $\mu_n, \mu \in \mathrm{Prob}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then $\mu_n \xrightarrow{w} \mu$ if and only if $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \quad \forall f \in C_c(\mathbb{R}^d) \ \ldots (\dagger)$

Lemma 6

If $\dagger$ holds true, then $\lim_{R\uparrow\infty} \inf_n \mu_n(\bar{B}_R) = 1$

Proof for Theorem.

Weak convergence for $\mathbb{R}^d$

Corollary 4

Let $\mu_n, \mu \in \text{Prob}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then $\mu_n \xrightarrow{w} \mu$ if and only if

$$\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \quad \forall f \in C_c^\infty(\mathbb{R}^d)$$

For any function $F : S \to T$ between topological spaces,

$$\text{Cont}(F) := \{x \in S : F \text{ is continuous at } x\}, \text{ Disc}(F) := S \setminus \text{Cont}(F)$$

Recall that every probability measure on $\mathbb{R}$ is a Stieltjes measure.

Weak convergence for $\mathbb{R}^d$

Corollary 4

Let $\mu_n, \mu \in \mathrm{Prob}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then $\mu_n \xrightarrow{w} \mu$ if and only if

$$\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \quad \forall f \in C_c^\infty(\mathbb{R}^d)$$

For any function $F : S \to T$ between topological spaces,

$$\mathrm{Cont}(F) := \{x \in S : F \text{ is continuous at } x\},\ \mathrm{Disc}(F) := S \setminus \mathrm{Cont}(F)$$

Recall that every probability measure on $\mathbb{R}$ is a Stieltjes measure.

Theorem 4

Let $\mu, \mu \in \mathrm{Prob}(\mathbb{R})$. Set $F_n(T) = \mu_n(-\infty, t], F(t) = \mu(-\infty, t]$.
Then $\mu_n \xrightarrow{w} \mu$ if and only if $F_n(t) \to F(t) \quad \forall t \in \mathrm{Cont}(F)$

Vague Convergence

Consider $\mu_n = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_n \in \mathsf{Prob}(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right))$. Does μ_n have a limit in some sense?
If $f \in C_c(\mathbb{R}), f(x) = 0\, \forall |x| \geq M$, $\int f \,\mathrm{d}\mu_n = \frac{1}{2} f(0), n \geq m$
Note: $\mu_n \overset{w}{\not\longrightarrow} \frac{1}{2}\delta_0$. In fact, $\{\mu_n\}_{n=1}^{\infty}$ possesses no weakly convergent subsequence.
What's going on?

Vague Convergence

Consider $\mu_n = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_n \in \mathsf{Prob}(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right))$. Does μ_n have a limit in some sense?
If $f \in C_c(\mathbb{R}), f(x) = 0 \, \forall |x| \geq M$, $\int f \, \mathrm{d}\mu_n = \frac{1}{2} f(0), n \geq m$
Note: $\mu_n \overset{w}{\not\longrightarrow} \frac{1}{2}\delta_0$. In fact, $\{\mu_n\}_{n=1}^{\infty}$ possesses no weakly convergent subsequence.
What's going on?

Definition 4 (Vague convergence)

Let μ_n, μ be Borel measures on $\mathbb{R}^d$. Say μ_n converges *vaguely* to μ, $\mu_n \underset{v}{\to} \mu$, if $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \, \forall f \in C_c(\mathbb{R}^d)$

It is possible to lose mass, but not gain it under vague convergence.

Vague Convergence

Consider $\mu_n = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_n \in \text{Prob}(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right))$. Does μ_n have a limit in some sense?
If $f \in C_c(\mathbb{R}), f(x) = 0 \; \forall |x| \geq M$, $\int f \, \mathrm{d}\mu_n = \frac{1}{2}f(0), n \geq m$
Note: $\mu_n \overset{w}{\not\longrightarrow} \frac{1}{2}\delta_0$. In fact, $\{\mu_n\}_{n=1}^{\infty}$ possesses no weakly convergent subsequence.
What's going on?

Definition 4 (Vague convergence)

Let μ_n, μ be Borel measures on $\mathbb{R}^d$. Say μ_n converges *vaguely* to μ, $\mu_n \underset{v}{\to} \mu$, if $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n \to \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \; \forall f \in C_c(\mathbb{R}^d)$

It is possible to lose mass, but not gain it under vague convergence.

Definition 5 (Tightness)

A family $\Lambda \subseteq \text{Prob}(S, \mathcal{B}\left(()\, S\right))$ is called *tight* if $\varepsilon > 0, \exists K_\varepsilon \subseteq S$ compact s.t. $\mu(K_\varepsilon) \geq 1 - \varepsilon \; \forall \mu \in \Lambda$

Note. Weakly convergent sequence of probability measures on $\mathbb{R}^d$ are tight.
Example. $\mu_n = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_n$, $\{\mu_n\}_{n=1}^{\infty}$ is not tight.

Vague Convergence

Theorem 5

If $\mu_n \in \text{Prov}(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right))$ and $\mu_n \underset{v}{\rightarrow} \mu$ for some Borel measure, then $\mu(\mathbb{R}^d) = 1$ if and only if $\{\mu_n\}_{n=1}^{\infty}$ is tight.

Proof.

Vague convergence on $\mathbb{R}$

Proposition 3

Let $\mu_n \in \text{Prov}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and let μ be a finite Borel measure on $\mathbb{R}$. Let $F_n(t) = \mu_n(-\infty t], F(t) = \mu(-\infty, t]$. Then $\mu_n \underset{v}{\to} \mu$ if and only if $F_n(b) - F_n(a) \to F(b) - F(a) \quad \forall a, b, \in \text{Cont}(F)$

Prokhorov's Compactness Theorem

Some sequences of probability measures have no weakly convergent subsequences. The one and only obstruction is tightness.

Theorem 6 (Prokhorov)

Let S be a separable metric space. If $\{\mu_n\}_{n=1}^{\infty} \subset$ Prov$(S, \mathcal{B}(S))$, $\exists$ vaguely convergent subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$

Corollary 5

If $\{\mu_n\}_{n=1}^{\infty}$ is also tight, then $\exists$ weakly convergent subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ whose limit μ is a probability measure.

Proof for Theorem

Connection between weak convergence and a.s. convergence

Theorem 7 (Skorohod)

Let S be a separable metric space, and $\mu_n, \mu \in \mathrm{Prob}(S, \mathcal{B}(S))$. If $\mu_n \xrightarrow{w} \mu$, then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $Y_n, Y : (\Omega, \mathcal{F}) \to (S, \mathcal{B}(S))$ with $Y_n^* \mathbb{P} = \mu_{Y_n} = \mu_n$, $Y^* \mathbb{P} = \mu_Y = \mu$, and $Y_n \to Y$ a.s.

Connection between weak convergence and a.s. convergence

Theorem 7 (Skorohod)

Let S be a separable metric space, and $\mu_n, \mu \in \text{Prob}(S, \mathcal{B}(S))$. If $\mu_n \xrightarrow{w} \mu$, then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $Y_n, Y : (\Omega, \mathcal{F}) \to (S, \mathcal{B}(S))$ with $Y_n^* \mathbb{P} = \mu_{Y_n} = \mu_n$, $Y^* \mathbb{P} = \mu_Y = \mu$, and $Y_n \to Y$ a.s.

Corollary 6 (Continuous mapping theorem)

Let $F : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Let $X_n \xrightarrow{w} X$, and suppose $\mathbb{P}(X \in \text{Disc}(f)) = 0$. Then $f(X_n) \xrightarrow{w} f(X)$. If in addition f is bounded, then $\mathbb{E}\, f(X_n) \to \mathbb{E}\, f(X)$