STATS 217: Introduction to Stochastic Processes I

Lecture 17

Total variation distance

- Let μ and ν be two probability distributions on Ω.
- The **total variation distance** between them, denoted by $\mathrm{TV}(\mu, \nu)$, is defined by

$$\mathrm{TV}(\mu, \nu) := \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

- On the homework, you will show that

$$\mathrm{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

- Note that $\mathrm{TV}(\mu, \nu)$ is a metric on the set of probability measures on Ω.

Total variation distance is a metric

- $\mathsf{TV}(\mu, \nu) \geq 0$ and $\mathsf{TV}(\mu, \nu) = \mathsf{TV}(\nu, \mu)$.
- If $\mathsf{TV}(\mu, \nu) = 0$, then $|\mu(x) - \nu(x)| = 0$ for all $x \in \Omega$ so that $\mu \equiv \nu$.
- Finally, TV satisfies the triangle inequality: for probability measures μ, ν, η on Ω, $\mathsf{TV}(\mu, \nu) \leq \mathsf{TV}(\mu, \eta) + \mathsf{TV}(\eta, \nu)$.
- Indeed,

$$\begin{aligned} 2\,\mathsf{TV}(\mu, \nu) &= \sum_{x \in \Omega} |\mu(x) - \nu(x)| \\ &= \sum_{x \in \Omega} |\mu(x) - \eta(x) + \eta(x) - \nu(x)| \\ &\leq \sum_{x \in \Omega} |\mu(x) - \eta(x)| + \sum_{x \in \Omega} |\eta(x) - \nu(x)| \\ &= 2\,\mathsf{TV}(\mu, \eta) + 2\,\mathsf{TV}(\eta, \nu). \end{aligned}$$

Dual characterization of total variation distance

- Let μ, ν be probability measures on Ω. Let $\mathcal{F}$ denote the collection of all functions $f : \Omega \to \mathbb{R}$ satisfying $\max_{x \in \Omega} |f(x)| \leq 1$.
- Then,

$$\mathrm{TV}(\mu, \nu) = \frac{1}{2} \sup_{f \in \mathcal{F}} \left\{ \sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x) \right\}.$$

- Why? For any $f \in \mathcal{F}$,

$$\begin{aligned} |\sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x)| &\leq \sum_{x \in \Omega} |f(x)||\mu(x) - \nu(x)| \\ &\leq \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \end{aligned}$$

- You will prove the reverse inequality on the homework.

Coupling

- Let μ and ν be two probability measures on Ω_1 and Ω_2 respectively.
- A **coupling** of μ and ν is a probability measure γ on $\Omega_1 \times \Omega_2$ such that

$$\gamma(A \times \Omega_2) = \mu(A) \quad \forall A \subseteq \Omega_1 \text{ and}$$
$$\gamma(\Omega_1 \times B) = \nu(B) \quad \forall B \subseteq \Omega_2.$$

- Similarly, a coupling of random variables $X : \Omega_1' \to \Omega_1$ and $Y : \Omega_2' \to \Omega_2$ is a pair of random variables $\widehat{X} : \Omega \to \Omega_1$ and $\widehat{Y} : \Omega \to \Omega_2$ defined on a common probability space Ω such that

$$\mathbb{P}[\widehat{X} = x] = \mathbb{P}[X = x] \quad \forall x \in \Omega_1 \text{ and}$$
$$\mathbb{P}[\widehat{Y} = y] = \mathbb{P}[Y = y] \quad \forall y \in \Omega_2.$$

Example: independent coupling

- Let $X \sim \text{Ber}(p)$ and $Y \sim \text{Ber}(q)$ where $0 \leq p \leq q \leq 1$.
- Formally, we can think of $X : [0,1] \to \{0,1\}$ where $[0,1]$ is equipped with the uniform measure and

$$X(r) = \begin{cases} 0 \text{ if } r \leq 1-p \\ 1 \text{ if } r > 1-p. \end{cases}$$

- Y admits a similar interpretation.
- An obvious coupling of X and Y is the **independent coupling** i.e., $\Omega = [0,1] \times [0,1]$ equipped with the uniform measure,

$$\widehat{X}(r_1, r_2) = \begin{cases} 0 \text{ if } r_1 \leq 1-p \\ 1 \text{ if } r_1 \geq 1-p, \end{cases}$$

and similarly for $\widehat{Y}$ (with r_1 replaced by r_2 and p replaced by q).

Example: monotone coupling

A particularly useful coupling in this case is the **monotone coupling**.

- $\Omega = [0, 1]$ with the uniform measure.
- $$\widehat{X}(r) = \begin{cases} 0 \text{ if } r \leq 1 - p \\ 1 \text{ if } r > 1 - p. \end{cases}$$
- $$\widehat{Y}(r) = \begin{cases} 0 \text{ if } r \leq 1 - q \\ 1 \text{ if } r > 1 - q. \end{cases}$$
- Then,
$$\mathbb{P}[\widehat{X} = 1] = \mathbb{P}[r > 1 - p] = p = \mathbb{P}[X = 1]$$
and similarly for $\widehat{Y}$.
- The name monotone coupling comes from the observation that if $p \leq q$, then
$$\widehat{X} \leq \widehat{Y} \quad \text{determistically.}$$

An application of monotone coupling

- Let $(P_t)_{t\geq 0}$ denote a simple random walk starting from 0 where the probability of taking a step to the right is p.
- Let $(Q_t)_{t\geq 0}$ denote a simple random walk starting from 0 where the probability of taking a step to the right is q.
- Then, if $p \leq q$, intuitively,

$$\mathbb{P}[Q_t \leq z] \leq \mathbb{P}[P_t \leq z] \quad \forall t \geq 0, z \in \mathbb{Z}.$$

- Monotone coupling lets us see this directly.

An application of monotone coupling

- Let $(\widehat{X}, \widehat{Y})$ denote the monotone coupling of Ber(p) and Ber(q).
- Let $(\widehat{X}_t, \widehat{Y}_t)_{t \geq 1}$ denote iid copies of $(\widehat{X}, \widehat{Y})$.
- Let $\widehat{P}_t = \sum_{i=1}^{t}(2\widehat{X}_i - 1)$ and $\widehat{Q}_t = \sum_{i=1}^{t}(2\widehat{Y}_i - 1)$.
- Then, by construction, $\widehat{P}_t \leq \widehat{Q}_t$ for all t.
- Moreover, $P_t \sim \widehat{P}_t$ and $Q_t \sim \widehat{Q}_t$ for all t.
- So, for any $z \in \mathbb{Z}$ and any $t \geq 0$,

$$\mathbb{P}[Q_t \leq z] = \mathbb{P}[\widehat{Q}_t \leq z] \leq \mathbb{P}[\widehat{P}_t \leq z] = \mathbb{P}[P_t = z].$$

Coupling and total variation

- Let μ and ν be two probability distributions on Ω.
- The **coupling lemma** asserts that

$$\mathsf{TV}(\mu, \nu) = \inf\{\mathbb{P}[X \neq Y] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

- Example: let $\mu = \mathrm{Ber}(p)$ and $\nu = \mathrm{Ber}(q)$ with $0 \leq p \leq q \leq 1$.
- Then, by direct computation,

$$\mathsf{TV}(\mathrm{Ber}(p), \mathrm{Ber}(q)) = \frac{1}{2}(|q - p| + |1 - q - 1 + p|) = q - p.$$

- For the monotone coupling $(\widehat{X}, \widehat{Y})$, we have

$$\mathbb{P}[\widehat{X} \neq \widehat{Y}] = \mathbb{P}[1 - q \leq r \leq 1 - p] = q - p.$$

- The above characterization shows that the monotone coupling is an **optimal coupling**.

Coupling and total variation

- $\mathsf{TV}(\mu, \nu) = \inf\{\mathbb{P}[X \neq Y] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}$.
- Easy direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \\ &= \mathbb{P}[X \in A] - \mathbb{P}[X \in A, Y \in A] - \mathbb{P}[X \notin A, Y \in A] \\ &= \mathbb{P}[X \in A, Y \notin A] - \mathbb{P}[X \notin A, Y \in A] \\ &\leq \mathbb{P}[X \in A, Y \notin A] \\ &\leq \mathbb{P}[X \neq Y]. \end{aligned}$$

- The reverse inequality is a starred problem on HW7.