# STATS 217: Introduction to Stochastic Processes I

Lecture 17

# Total variation distance

"goal: prove the conv. thm for markov chains irr. & aperiodic"

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

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


$\mu$ and $\nu$ are prob. distributions here.

$|\Omega| = n$. there are $2^n$ possible subsets of $\Omega$

$A_1, A_2, \ldots, A_{2^n}$

$$\max_{i=1,\ldots,2^n} |\mu(A_i) - \nu(A_i)|$$

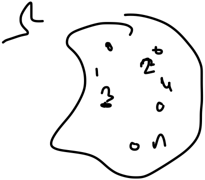
# Total variation distance

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

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

need to take the maximum of $2^{|\Omega|}$ subsets

- On the homework, you will show that




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

$$\frac{1}{2}\left[ |\mu(\{1\}) - \nu(\{1\})| + \cdots + | \quad | \right]$$

# Total variation distance

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

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

- On the homework, you will show that

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

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

distance

prob meas on $\Omega$
for every $x \in \Omega \to \mu(x)$

$S$
metric on $S$
$d: S \times S \longrightarrow \mathbb{R}^{\geq 0}$

# Total variation distance is a metric

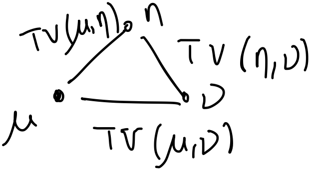
→ $d(x,y) \geq 0$ (✓)
→ $d(x,y) = d(y,x)$ (✓)
→ $d(x,y) = 0 \iff x = y$ (✓)
→ triangle ineq: $d(x,z) \leq d(x,y) + d(y,z)$


- $\text{TV}(\mu, \nu) \geq 0$ and $\text{TV}(\mu, \nu) = \text{TV}(\nu, \mu)$.
- If $\text{TV}(\mu, \nu) = 0$, then $|\mu(x) - \nu(x)| = 0$ for all $x \in \Omega$ so that $\mu \equiv \nu$.

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

# 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 $\mu, \nu, \eta$ on $\Omega$, $\mathsf{TV}(\mu,\nu) \leq \mathsf{TV}(\mu,\eta) + \mathsf{TV}(\eta,\nu)$.



# 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 $\mu, \nu, \eta$ on $\Omega$, $\mathsf{TV}(\mu, \nu) \leq \mathsf{TV}(\mu, \eta) + \mathsf{TV}(\eta, \nu)$.
- Indeed,

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

# 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 $\mu, \nu, \eta$ on $\Omega$, $\mathsf{TV}(\mu,\nu) \leq \mathsf{TV}(\mu,\eta) + \mathsf{TV}(\eta,\nu)$.
- Indeed,

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

# 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 $\mu, \nu, \eta$ on $\Omega$, $\mathsf{TV}(\mu, \nu) \leq \mathsf{TV}(\mu, \eta) + \mathsf{TV}(\eta, \nu)$.
- Indeed,

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

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

# Dual characterization of total variation distance

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

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

$\Omega \xrightarrow{f} \mathbb{R}.$

we demand that
$|f(x)| \leq 1$
$\forall x \in \Omega,$

$\mathbb{E}_{\mu}[f] - \mathbb{E}_{\nu}[f]$

# Dual characterization of total variation distance

- Let $\mu, \nu$ be probability measures on $\Omega$. 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 $f \in \mathcal{F}$,

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

$|f(x)| \leq 1$

- You will prove the reverse inequality on the homework.

# Coupling

joint distributions of Random vars $(X, Y)$

$$\mathbb{P}[X = x] = \sum_{y \in Y} \mathbb{P}[X = x, Y = y]$$

- Let $\mu$ and $\nu$ be two probability measures on $\Omega_1$ and $\Omega_2$ respectively.
- A **coupling** of $\mu$ and $\nu$ is a probability measure $\gamma$ on $\Omega_1 \times \Omega_2$ such that

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

$X \,,\, Y \longrightarrow$ construct joint distribution $(\hat{X}, \hat{Y})$

$$\text{ex}: \left\{ \begin{array}{l} \mathbb{P}[\hat{X} = x, \hat{Y} = y] \\ = \mathbb{P}[X = x]\,\mathbb{P}[Y = y] \end{array} \right\}. \qquad \begin{array}{l} \mathbb{P}[\hat{X} = x] = \mathbb{P}[X = x] \\ \mathbb{P}[\hat{Y} = y] = \mathbb{P}[Y = y] \end{array}$$

# Coupling

- Let $\mu$ and $\nu$ be two probability measures on $\Omega_1$ and $\Omega_2$ respectively.
- A **coupling** of $\mu$ and $\nu$ is a probability measure $\gamma$ 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 $\Omega$ 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

$+1$ w.p. $p$ $\quad$ $+1$ w.p. $q$

$0$ w.p. $1-p$ $\quad$ $0$ w.p. $1-q$

- 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.

$$Y(r) = \begin{cases} 0 & \text{if } r \leq 1-q \\ 1 & \text{if } r > 1-q \end{cases}$$

$X : \Omega \to \mathbb{R}$ $\qquad$ $Y : \Omega \to \mathbb{R}$

$\Omega = [0,1]$

# 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,

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

$\mathbb{P}[\hat{X} = 0]$
$= 1-p$
$= \mathbb{P}[X = 0]$

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

# 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 $\widehat{Y}$.

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

- $\Omega = [0,1]$ with the uniform measure.

$p = 1/2$ H $\quad q = 2/3$ H

- gen. $r \sim unif([0,1])$

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

Can you use just one sample from $U([0,1])$ to generate both coins.

- $r \le 1/2$ : $X = H, Y = H$

$1/2 \le r \le 2/3$ : $X = T, Y = H$

$r > 2/3$ : $X = T, Y = T$

$$\widehat{Y}(r) = \begin{cases} 0 \text{ if } r \le 1-q \\ 1 \text{ if } r > 1-q. \end{cases}$$

- Then,

$$\mathbb{P}[\widehat{X} = 1] = \mathbb{P}[r > 1-p] = p = \mathbb{P}[X = 1]$$

0 $\quad 1/2 \quad$ 1

and similarly for $\widehat{Y}$.

- The name monotone coupling comes from the observation that if $p \le q$, then

$$\widehat{X} \le \widehat{Y} \quad \text{determistically.}$$

feature & not a bug.

# An application of monotone coupling

coin gen. Room (J. T.)

extreme example!

X w.p. 1/2 Y w.p. 1/2

X Room H w.p. 1/2

Y ROOM H w.p. 2/3

- 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 $q$.

$P_t$ starts at 0

$\pm 1$ steps

$+1$ w.p. $p$

$-1$ w.p. $1-p$

$p = 1/2$

$Q_t$ starts at 0

$\pm 1$ steps

$+1$ w.p. $q$

$-1$ w.p. $1-q$

$q = 2/3$

# 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, e.g. $q = 2/3$, $p = 1/2$

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

e.g. $t = 100$


$\mathbb{P}[Q_{100} \leq 0]$ vs. $\mathbb{P}[P_{100} \leq 0]$

all poss. paths, try to look at prob. of those paths a ....

# 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,

this statement is only about "marginals".

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

- Monotone coupling lets us see this directly.

# An application of monotone coupling

$\hat{X} \leq \hat{Y}$ w.p. 1

$\frac{1}{2}$ $\frac{2}{3}$

- 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 $(\widehat{X}, \widehat{Y})$.

$\boxed{\text{Room}}$

$\to (\hat{X}_1, \hat{Y}_1)$

$\to (\hat{X}_2, \hat{Y}_2)$

$\vdots$

$\hat{X}_1, \hat{X}_2, \hat{X}_3 \ldots$ $\qquad$ $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \ldots$

$\underline{\underline{P}}$ $\qquad$ $\underline{\underline{Q}}$

# 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 $t$.

just to go from 0/1 to ± 1

# An application of monotone coupling

- Let $(\widehat{X}, \widehat{Y})$ denote the monotone coupling of $\mathrm{Ber}(p)$ and $\mathrm{Ber}(q)$.
- 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 $t$.

$$\begin{aligned} \mathbb{P}[\underline{P_t} \leq z] &= \mathbb{P}[\hat{P}_t \leq z] \\ &\leq \mathbb{P}[\hat{Q}_t \leq z] \\ &= \mathbb{P}[\underline{Q_t} \leq z]. \end{aligned} \Bigg\} \text{ monotone coupling.}$$

# 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 $\mu$ and $\nu$ be two probability distributions on $\Omega$.
- The **coupling lemma** asserts that

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

# Coupling and total variation

- Let $\mu$ and $\nu$ be two probability distributions on $\Omega$.
- 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 $\mu = \mathsf{Ber}(p)$ and $\nu = \mathsf{Ber}(q)$ with $0 \leq p \leq q \leq 1$.
- Then, by direct computation,

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

# Coupling and total variation

- Let $\mu$ and $\nu$ be two probability distributions on $\Omega$.
- 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

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

# Coupling and total variation

- Let $\mu$ and $\nu$ be two probability distributions on $\Omega$.
- 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?

## 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 $\mu, \nu$. Let $A \subseteq \Omega$.

# 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 $\mu, \nu$. Let $A \subseteq \Omega$. Then,

$$\mu(A) - \nu(A) = \mathbb{P}[X \in A] - \mathbb{P}[Y \in A]$$

# 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 $\mu, \nu$. Let $A \subseteq \Omega$. Then,

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

# 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 $\mu, \nu$. Let $A \subseteq \Omega$. Then,

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

# 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 $\mu, \nu$. Let $A \subseteq \Omega$. Then,

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

# Coupling and total variation

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