STATS 217: Introduction to Stochastic Processes I

Lecture 18

Coupling and total variation

$\mu \quad \nu \qquad (\tilde{\mu}, \tilde{\nu})$

$TV(\mu, \nu) = \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$

- Let μ and ν be two probability distributions on Ω.
- The **coupling lemma** asserts that (very fundamental/useful).

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

extreme case: $\mu = \nu$

$TV(\mu, \mu) = 0$

a coupling of μ & μ is just (X, X) where $X \sim \mu$.

last time:

$\mu = Ber(p)$

$\nu = Ber(q)$

↳ ind. coupling $\mathbb{P}[X \neq Y]$

↳ monotone coupling $\mathbb{P}[X \neq Y]$

↳ ...

* if you give me any coupling → upper bound on TV.
(this is what we will use for e.g. Poisson approx.)

* if you can compute TV, 'test' quality of a coupling.

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

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

$$\frac{1}{2} |\mu(0) - \nu(0)| + \frac{1}{2} |\mu(1) - \nu(1)|$$

Coupling and total variation

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

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

(intuitively, same as min)

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

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

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

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

$\hat{X} = \hat{Y} = 0$ dis. $\hat{X} = \hat{Y} = 1$

0 1-q 1-p 0

$\Rightarrow$ monotone coupling is optimal.

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

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

it suffices to show that for any coupling (X, Y) of μ & ν, we have

$$\mathrm{TV}(\mu, \nu) \leq \mathbb{P}[X \neq Y]$$

& now take inf on both sides over all couplings

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/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$.

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

enough to show that given $A \subseteq \Omega$

$$|\mu(A) - \nu(A)| \leq \Pr(X \neq Y)$$

and now, take max over A on both sides

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/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

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

$X \sim \mu$
$Y \sim \nu$

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/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

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

$$= \cancel{\mathbb{P}[X \in A, Y \in A]} + \mathbb{P}[X \in A, Y \notin A].$$

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/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

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

$$\{X \in A, Y \notin A\} \subseteq \{X \neq Y\}$$

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/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

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

Coupling and total variation

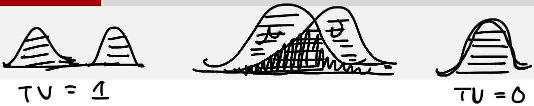
- $\mathsf{TV}(\mu,\nu) = \inf\{\mathbb{P}[X \neq Y] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}$.
- Easy/useful direction: $\leq$. Why?
- Let (X, Y) be any coupling of μ, ν. Let $A \subseteq \Omega$. Then,

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

- The reverse inequality is a starred problem on HW7.

Rough idea: $x \in \Omega$, can only get to agree on x w.p. $\min\{\mu(x), \nu(x)\}$

Poisson approximation



- In our discussion of Poisson random variables, we frequently used the (informal) approximation

$$\text{Pois}(\lambda) \approx X_1 + \cdots + X_n,$$

$\lambda > 0$ fixed
n large
$(n \to \infty)$

where $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with mean λ/n.

- Now, we have the machinery to make this precise.

Poisson approximation

- In our discussion of Poisson random variables, we frequently used the (informal) approximation

$$\mathrm{Pois}(\lambda) \approx X_1 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with mean λ/n.
- Now, we have the machinery to make this precise.
- Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with means $p_1, \ldots, p_n$.
- In other words, for each X_i, $\mathbb{P}[X_i = 1] = p_i$ and $\mathbb{P}[X_i = 0] = (1 - p_i)$.

Poisson approximation

- In our discussion of Poisson random variables, we frequently used the (informal) approximation

$$\text{Pois}(\lambda) \approx X_1 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with mean λ/n.
- Now, we have the machinery to make this precise.
- Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with means $p_1, \ldots, p_n$.
- In other words, for each X_i, $\mathbb{P}[X_i = 1] = p_i$ and $\mathbb{P}[X_i = 0] = (1 - p_i)$.
- Let $S_n = X_1 + \cdots + X_n$.

we are trying to approx S_n by Pois(λ)

Poisson approximation

$\left[\mathbb{P}[\text{Pois}(\lambda_i) = 0] = \mathbb{P}[\text{Ber}(p_i) = 0] \right]$

- Let $\lambda_i = -\log(1 - p_i)$. Equivalently, $e^{-\lambda_i} = (1 - p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.

$S_n = X_1 + \cdots + X_n$

$X_i \approx \text{Pois}(\lambda_i)$

$S_n \approx \text{Pois}(\lambda_1) + \cdots + \text{Pois}(\lambda_n)$

$= \text{Pois}(\lambda)$

Poisson approximation

- Let $\lambda_i = -\log(1 - p_i)$. Equivalently, $e^{-\lambda_i} = (1 - p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.
- We will show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \frac{1}{2} \sum_{i=1}^{n} \lambda_i^2.$$

Poisson approximation

- Let $\lambda_i = -\log(1 - p_i)$. Equivalently, $e^{-\lambda_i} = (1 - p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.
- We will show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \le \frac{1}{2} \sum_{i=1}^{n} \lambda_i^2.$$

- **Example**: $\Lambda > 0$ is fixed, n is large, $p_i = \Lambda/n$ for all $i = 1, \ldots, n$.
- Then, $\lambda_i = \Lambda/n + O(\Lambda^2/n^2)$, $\lambda = \Lambda + O(\Lambda^2/n)$.

$-\log\left(1 - \frac{\Lambda}{n}\right)$

Poisson approximation

- Let $\lambda_i = -\log(1 - p_i)$. Equivalently, $e^{-\lambda_i} = (1 - p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.
- We will show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \le \frac{1}{2} \sum_{i=1}^{n} \lambda_i^2.$$

- **Example**: $\Lambda > 0$ is fixed, n is large, $p_i = \Lambda/n$ for all $i = 1, \ldots, n$.
- Then, $\lambda_i = \Lambda/n + O(\Lambda^2/n^2)$, $\lambda = \Lambda + O(\Lambda^2/n)$.
- On the homework, you will show that $\mathrm{TV}(\mathrm{Pois}(\mu), \mathrm{Pois}(\nu)) \le |\nu - \mu|$.

$$\sum_{i=1}^{n} \lambda_i^2 = O\left(\cancel{n} \cdot \frac{\Lambda^2}{n^{\cancel{2}}}\right) = O\left(\frac{\Lambda^2}{n}\right).$$

- Let $\lambda_i = -\log(1-p_i)$. Equivalently, $e^{-\lambda_i} = (1-p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.
- We will show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \le \boxed{\frac{1}{2}\sum_{i=1}^{n} \lambda_i^2.}$$

- **Example**: $\Lambda > 0$ is fixed, n is large, $p_i = \Lambda/n$ for all $i = 1, \ldots, n$.
- Then, $\lambda_i = \Lambda/n + O(\Lambda^2/n^2)$, $\underline{\underline{\lambda}} = \underline{\underline{\Lambda}} + \underline{\underline{O(\Lambda^2/}}n)$.
- On the homework, you will show that $\mathrm{TV}(\mathrm{Pois}(\mu), \mathrm{Pois}(\nu)) \le |\nu - \mu|$.
- Then, by the triangle inequality,

$$\mathrm{TV}(S_n, \mathrm{Pois}(\Lambda)) \le \underbrace{\mathrm{TV}(S_n, \mathrm{Pois}(\lambda))}_{= O\left(\frac{\Lambda^2}{n}\right)} + \underbrace{\mathrm{TV}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\Lambda))}_{= O\left(\frac{\Lambda^2}{n}\right).}$$

S_n · Poi(λ) · Poi(Λ)

Poisson approximation

- Let $\lambda_i = -\log(1 - p_i)$. Equivalently, $e^{-\lambda_i} = (1 - p_i)$.
- Let $\lambda = \lambda_1 + \cdots + \lambda_n$.
- We will show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \frac{1}{2} \sum_{i=1}^{n} \lambda_i^2.$$

- **Example**: $\Lambda > 0$ is fixed, n is large, $p_i = \Lambda/n$ for all $i = 1, \ldots, n$.
- Then, $\lambda_i = \Lambda/n + O(\Lambda^2/n^2)$, $\lambda = \Lambda + O(\Lambda^2/n)$. $\lambda_i = -\log(1-p_i)$
- On the homework, you will show that $\mathrm{TV}(\mathrm{Pois}(\mu), \mathrm{Pois}(\nu)) \leq |\nu - \mu|$.
- Then, by the triangle inequality,

$$\begin{aligned}
\mathrm{TV}(S_n, \mathrm{Pois}(\Lambda)) &\leq \mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) + \mathrm{TV}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\Lambda)) \\
&\leq O(\Lambda^2/n) + O(\Lambda^2/n) \\
&\leq O(\Lambda^2/n),
\end{aligned}$$

which justifies our approximation from before.

Poisson approximation

- We now show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \frac{1}{2}\sum_{i=1}^{n}\lambda_i^2.$$

let me try to prove this for $n = 1$.

$$\left[\ \mathrm{TV}\left(\mathrm{Ber}(p), \mathrm{Pois}(-\log(1-p))\right) \leq \frac{1}{2}\left(\log(1-p)\right)^2. \ \right]$$

Poisson approximation

- We now show that
$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2.$$
- Let us first prove this for $n = 1$. Let $\lambda = -\log(1-p)$. We want to show:
$$\mathrm{TV}(\mathrm{Ber}(p), \mathrm{Pois}(\lambda)) \leq \frac{1}{2}\lambda^2.$$

Poisson approximation

- We now show that

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2.$$

- Let us first prove this for $n = 1$. Let $\lambda = -\log(1-p)$. We want to show:

$$\mathrm{TV}(\mathrm{Ber}(p), \mathrm{Pois}(\lambda)) \leq \frac{1}{2}\lambda^2.$$

- By the coupling lemma, it suffices to exhibit a coupling $(\widehat{X}, \widehat{Y})$ of $\mathrm{Ber}(p)$ and $\mathrm{Pois}(\lambda)$ such that

$$\mathbb{P}[\widehat{X} \neq \widehat{Y}] \leq \frac{1}{2}\lambda^2.$$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.

$$\widehat{Y} \xleftarrow{Z} \boxed{Z} \xrightarrow{\min\{Z,1\}} \widehat{X}$$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.
- Clearly $\widehat{Y}$ has the correct marginal distribution. As for $\widehat{X}$, note that

$$\mathbb{P}[\widehat{X} = 0] = \mathbb{P}[Z = 0] = e^{-\lambda} = (1 - p) = \mathbb{P}[\text{Ber}(p) = 0].$$

$$\Rightarrow \mathbb{P}[\hat{X} = 1] = 1 - (1-p) = p$$

$\hat{X} = \min\{Z, 1\}$ $\quad\quad$ $\hat{X} = 0 \Leftrightarrow Z = 0$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.
- Clearly $\widehat{Y}$ has the correct marginal distribution. As for $\widehat{X}$, note that

$$\mathbb{P}[\widehat{X} = 0] = \mathbb{P}[Z = 0] = e^{-\lambda} = (1 - p) = \mathbb{P}[\text{Ber}(p) = 0].$$

$\hat{Y} = Z$
$\hat{X} = \min\{Z, 1\}$

- Moreover,

$$\mathbb{P}[\widehat{X} \neq \widehat{Y}] = \mathbb{P}[Z \geq 2]$$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.
- Clearly $\widehat{Y}$ has the correct marginal distribution. As for $\widehat{X}$, note that

$$\mathbb{P}[\widehat{X} = 0] = \mathbb{P}[Z = 0] = e^{-\lambda} = (1 - p) = \mathbb{P}[\text{Ber}(p) = 0].$$

- Moreover,

$$\begin{aligned} \mathbb{P}[\widehat{X} \neq \widehat{Y}] &= \mathbb{P}[Z \geq 2] \\ &= e^{-\lambda} \sum_{j \geq 2} \frac{\lambda^j}{j!} \end{aligned}$$

$$= e^{-\lambda} \left(\frac{\lambda^2}{2} \right) \left(\sum_{j \geq 2} \frac{\lambda^{j-2}}{j!} \cdot 2 \right)$$

change of vars.

$j - 2 = k$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.
- Clearly $\widehat{Y}$ has the correct marginal distribution. As for $\widehat{X}$, note that

$$\mathbb{P}[\widehat{X} = 0] = \mathbb{P}[Z = 0] = e^{-\lambda} = (1 - p) = \mathbb{P}[\text{Ber}(p) = 0].$$

- Moreover,

$$\begin{aligned} \mathbb{P}[\widehat{X} \neq \widehat{Y}] &= \mathbb{P}[Z \geq 2] \\ &= e^{-\lambda} \sum_{j \geq 2} \frac{\lambda^j}{j!} \\ &\leq \frac{\lambda^2}{2} \underbrace{\sum_{j \geq 0} e^{-\lambda} \frac{\lambda^j}{j!}}_{=1} \end{aligned}$$

Poisson approximation

- Here is such a coupling: Generate $Z \sim \text{Pois}(\lambda)$. Then, set $\widehat{Y} = Z$ and $\widehat{X} = \min\{Z, 1\}$.
- Clearly $\widehat{Y}$ has the correct marginal distribution. As for $\widehat{X}$, note that

$$\mathbb{P}[\widehat{X} = 0] = \mathbb{P}[Z = 0] = e^{-\lambda} = (1 - p) = \mathbb{P}[\text{Ber}(p) = 0].$$

- Moreover,

$$\begin{aligned}\mathbb{P}[\widehat{X} \neq \widehat{Y}] &= \mathbb{P}[Z \geq 2] \\ &= e^{-\lambda} \sum_{j \geq 2} \frac{\lambda^j}{j!} \\ &\leq \frac{\lambda^2}{2} \sum_{j \geq 0} e^{-\lambda} \frac{\lambda^j}{j!} \qquad \lambda = -\log(1-p). \\ &= \frac{\lambda^2}{2}. \qquad \Rightarrow TV(\text{Ber}(p), \text{Pois}(\lambda)) \leq \lambda^2/2\end{aligned}$$

Poisson approximation

- At this point, we are almost done.

Poisson approximation

$S_n = X_1 + \cdots + X_n$ ind.

$\mathrm{Poi}(\lambda) = \mathrm{Poi}(\lambda_1) + \cdots + \mathrm{Poi}(\lambda_n)$

independent

- At this point, we are almost done.
- Let $(\widehat{X}_i, \widehat{Y}_i)$ be a coupling of $\text{Ber}(p_i)$ and $\text{Pois}(\lambda_i)$ as above.
- Let $(\widehat{X}_1, \widehat{Y}_1), \ldots, (\widehat{X}_n, \widehat{Y}_n)$ be independent copies of this coupling.

$Z_1 \sim \mathrm{Poi}(\lambda_1)$ $\qquad$ $Z_n \sim \mathrm{Poi}(\lambda_n)$ $\qquad$ $Z_1 \ldots Z_n$ are ind.

$\hat{Y}_1 = Z_1$

$\hat{X}_1 = \min\{Z_1, 1\}$ $\quad \ldots\ldots$

Poisson approximation

- At this point, we are almost done.
- Let $(\widehat{X}_i, \widehat{Y}_i)$ be a coupling of $\mathrm{Ber}(p_i)$ and $\mathrm{Pois}(\lambda_i)$ as above.
- Let $(\widehat{X}_1, \widehat{Y}_1), \ldots, (\widehat{X}_n, \widehat{Y}_n)$ be independent copies of this coupling.
- Then, $S_n \sim \widehat{X}_1 + \cdots + \widehat{X}_n$ and $\mathrm{Pois}(\lambda) \sim \widehat{Y}_1 + \cdots + \widehat{Y}_n$.

"

$X_1 + \cdots + X_n$

"

$\mathrm{Pois}(\lambda_1) + \cdots + \mathrm{Pois}(\lambda_n)$

-

Poisson approximation

- At this point, we are almost done.
- Let $(\widehat{X}_i, \widehat{Y}_i)$ be a coupling of $\mathrm{Ber}(p_i)$ and $\mathrm{Pois}(\lambda_i)$ as above.
- Let $(\widehat{X}_1, \widehat{Y}_1), \ldots, (\widehat{X}_n, \widehat{Y}_n)$ be independent copies of this coupling.
- Then, $S_n \sim \widehat{X}_1 + \cdots + \widehat{X}_n$ and $\mathrm{Pois}(\lambda) \sim \widehat{Y}_1 + \cdots + \widehat{Y}_n$.
- Moreover, by the coupling lemma and our previous calculation

$$\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) \leq \mathbb{P}[\widehat{X}_1 + \ldots \widehat{X}_n \neq \widehat{Y}_1 + \cdots + \widehat{Y}_n]$$

Poisson approximation

- At this point, we are almost done.
- Let $(\widehat{X}_i, \widehat{Y}_i)$ be a coupling of $\mathrm{Ber}(p_i)$ and $\mathrm{Pois}(\lambda_i)$ as above.
- Let $(\widehat{X}_1, \widehat{Y}_1), \dots, (\widehat{X}_n, \widehat{Y}_n)$ be independent copies of this coupling.
- Then, $S_n \sim \widehat{X}_1 + \cdots + \widehat{X}_n$ and $\mathrm{Pois}(\lambda) \sim \widehat{Y}_1 + \cdots + \widehat{Y}_n$.
- Moreover, by the coupling lemma and our previous calculation

$$\begin{aligned}
\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) &\leq \mathbb{P}[\widehat{X}_1 + \dots \widehat{X}_n \neq \widehat{Y}_1 + \cdots + \widehat{Y}_n] \\
&\leq \mathbb{P}[\widehat{X}_1 \neq \widehat{Y}_1] + \cdots + \mathbb{P}[\widehat{X}_n \neq \widehat{Y}_n] \\
&\leq \frac{1}{2} \lambda_1^2 + \cdots\cdots + \frac{1}{2} \lambda_n^2
\end{aligned}$$

Poisson approximation

- At this point, we are almost done.
- Let $(\widehat{X}_i, \widehat{Y}_i)$ be a coupling of $\mathrm{Ber}(p_i)$ and $\mathrm{Pois}(\lambda_i)$ as above.
- Let $(\widehat{X}_1, \widehat{Y}_1), \ldots, (\widehat{X}_n, \widehat{Y}_n)$ be independent copies of this coupling.
- Then, $S_n \sim \widehat{X}_1 + \cdots + \widehat{X}_n$ and $\mathrm{Pois}(\lambda) \sim \widehat{Y}_1 + \cdots + \widehat{Y}_n$.
- Moreover, by the coupling lemma and our previous calculation

$$
\begin{aligned}
\mathrm{TV}(S_n, \mathrm{Pois}(\lambda)) &\le \mathbb{P}[\widehat{X}_1 + \ldots \widehat{X}_n \neq \widehat{Y}_1 + \cdots + \widehat{Y}_n] \\
&\le \mathbb{P}[\widehat{X}_1 \neq \widehat{Y}_1] + \cdots + \mathbb{P}[\widehat{X}_n \neq \widehat{Y}_n] \\
&\le \frac{\lambda_1^2}{2} + \cdots + \frac{\lambda_n^2}{2}.
\end{aligned}
$$

$\forall \ A$

$$\Pr(S_n \in A) = \Pr(\mathrm{Poi}(\lambda) \in A) \pm \left(\frac{\lambda_1^2 + \cdots + \lambda_n^2}{2}\right)$$

Random mapping representation of Markov chains

- We have often specified transitions of Markov chains in words. For instance, for the symmetric simple random walk, instead of writing down the transition matrix, we have used a simple description like: at each step, toss an independent fair coin. If the coin lands heads, move one step right. Else, move one step left.

Random mapping representation of Markov chains

- We have often specified transitions of Markov chains in words. For instance, for the symmetric simple random walk, instead of writing down the transition matrix, we have used a simple description like: at each step, toss an independent fair coin. If the coin lands heads, move one step right. Else, move one step left.
- We can formalize this by using the **random mapping representation** of a transition matrix P on the state space S. This is simply a function $f : S \times \Lambda \to S$ along with a Λ-valued random variable Z which satisfies

$$\mathbb{P}[f(x, Z) = y] = P_{x,y}.$$

- For instance, in the case of the symmetric simple random walk, we can take $\Lambda = \{H, T\}$, Z is a random variable which is H with probability $1/2$ and T otherwise, and $f(x, H) = x + 1$, $f(x, T) = x - 1$.

Random mapping representations of Markov chains

- In fact, every transition matrix on a finite state space $\{1, \ldots, n\}$ has a random mapping representation.

Random mapping representations of Markov chains

- In fact, every transition matrix on a finite state space $\{1, \ldots, n\}$ has a random mapping representation.
- Indeed, we can take $\Lambda = [0, 1]$, Z is uniformly distributed on $[0, 1]$ and

$$f(i, z) = j \iff \sum_{\ell=1}^{j-1} P_{i,\ell} \leq z \leq \sum_{\ell=1}^{j} P_{i,\ell}.$$

P_{11} P_{12} P_{13} P_{1n}

$\underline{P_{11}}$ P_{12} . . . . . P_{1n}

0 $P_{11} + P_{12}$ 1

P_{11}