STATS 217: Introduction to Stochastic Processes I

Lecture 18

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/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.

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 $S_n = X_1 + \cdots + X_n$.

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, $\lambda_i = \Lambda/n + O(\Lambda^2/n^2)$, $\lambda = \Lambda + O(\Lambda^2/n)$.
- 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

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

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

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

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

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

Poisson approximation

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

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

Poisson approximation

- 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.
- Then, $S_n \sim \widehat{X}_1 + \cdots + \widehat{X}_n$ and $\text{Pois}(\lambda) \sim \widehat{Y}_1 + \cdots + \widehat{Y}_n$.
- Moreover, by the coupling lemma and our previous calculation

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

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.
- Indeed, we can take $\Lambda = [0, 1]$, Z is uniformly distributed on $[0, 1]$ and

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