30 JAN: MCMC READING GROUP
COUPLING

Total Variation Distance

$$\|\underbrace{\mu - \nu}_{\text{Distributions on } \Omega}\|_{TV} = \max_{A \subseteq \Omega} \left(\mu(A) - \nu(A)\right)$$

$\mu(A) = \sum_{x \in A} \mu(x)$

Equivalent Definition:

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

Proof: Suppose A maximizes $\mu(A) - \nu(A)$,
Then A must be $\{i \mid \mu(i) \geq \nu(i)\}$
(other is will make a negative contribution)

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

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

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

$$= \frac{1}{2}\left[\sum_{x\in A} \mu(x)-\nu(x) - \left(1-\sum_{x\in A}\mu(x)\right) + \left(1-\sum_{x\in A}\nu(x)\right)\right]$$

$$= \frac{1}{2}\left[2\mu(A) - 2\nu(A) - 1 + 1\right]$$

$$= \mu(A) - \nu(A)$$

Triangle Inequality: $0.5 \times \text{Norm1}(\mu - \nu)$

satisfies triangle inequality

Coupling

A coupling b/w μ and ν is a pair of random variables (X, Y) defined on a single prob space such that $P(X=x) = \mu(x)$

$P(Y=y) = \nu(y)$

Marginal wrt x

Marginal wrt y

Example Application:

Coin A $\to$ H with prob p, T o/w

Coin B $\to$ H with prob q, T o/w

suppose $H_A(n) = \#$ heads after n tosses of A

$H_B(n) = \#$ heads after n tosses of B

$p < q$.

<u>Claim</u>: $\Pr(H_A(n) > k) \leq \Pr(H_B(n) > k)$

<u>Proof</u>: Define a coupling on $\{H,T\}^2$ as-

1. Toss coin A. If H, set coin B also to H.

2. If T, set coin B to H with some prob p' and tails otherwise.

How to make this a coupling?

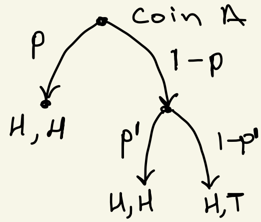
Marginals: Notice Marginal on A is identical to coin A.

Marginal on B -

We want -

$$p + (1-p)\cdot p' = q$$

$$\Rightarrow \boxed{p' = \frac{q-p}{(1-p)}}$$



Let $X_i = \begin{cases} 1, & \text{If the ith toss of the coupled coin has coin A} = H \\ 0, & \text{o/w} \end{cases}$

and similarly Y_i for coin B -

Notice that whenever $X_i = 1$, $Y_i = 1$ by the coupling.

$$\Rightarrow X_i \leq Y_i$$

$$\therefore \sum_{i=1}^{n} X_i \leq \sum_{i=1}^{n} Y_i \Rightarrow H_A(n) \leq H_B(n)$$

$$\Rightarrow \Pr(H_A(n) > k) \leq \Pr(H_B(n) > k)$$

TV Distance and Coupling

<u>Claim</u>: $\|\mu - \nu\|_{TV} = \inf\{\Pr(X \neq Y) \mid (X,Y) \text{ is a coupling of } X \text{ and } Y\}$

<u>Proof</u>: Let $A \subset \Omega$ be the event that maximizes $\mu(A) - \nu(A)$. For any coupling (X,Y) -

$$\Rightarrow \mu(A) - \nu(A) = \Pr(X \in A) - \Pr(Y \in A)$$

$$= \Pr(X \in A, Y \in A) + \Pr(X \in A, Y \notin A)$$

$$- \Pr(Y \in A, X \in A) - \Pr(Y \in A, X \notin A)$$

ignoring

$$\leq \Pr(X \in A, Y \notin A)$$

$$\leq \Pr(X \neq Y)$$

sub-event

Lets explicitly construct a coupling for which
$\Pr(X \neq Y) = \|\mu - \nu\|_{TV}$

Let $p = \sum_{x \in \Omega} \min(\mu(x), \nu(x))$

$$= \sum_{\substack{x \in \Omega \\ \mu(x) \leq \nu(x)}} \mu(x) + \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \nu(x)$$

Adding and subtracting $\sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \mu(x)$ –

$$= \overset{\sum_x \mu(x)}{1} + \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} (\nu(x) - \mu(x))$$

$$p = 1 - \|\mu - \nu\|_{TV}$$

Consider the coupling –

(i) Flip a coin with prob(H) = p.
If H, choose Z acc to the distribution –

$$\frac{\min(\mu(x), \nu(x))}{p}$$

and set $X = Y = Z$

(ii) If T, choose X acc to

$$\begin{cases} \dfrac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{TV}}, & \text{if } \mu(x) > \nu(x) \\ 0, & o/w \end{cases}$$

and choose Y acc to

$$\begin{cases} \dfrac{\nu(x) - \mu(x)}{\|\mu - \nu\|_{TV}}, & \text{if } \mu(x) < \nu(x) \\ 0, & o/w \end{cases}$$

Notice the marginals -

WRT X: if $\mu(x) > \nu(x)$,

$$\Pr(X=x) = p\left(\frac{\min(\mu(x), \nu(x))}{p}\right) + (1-p)\left(\frac{\mu(x)-\nu(x)}{\|\mu-\nu\|_{TV}}\right)$$

$$= \min(\mu(x), \nu(x)) + \|\mu-\nu\|_{TV}\left(\frac{\mu(x)-\nu(x)}{\|\mu-\nu\|_{TV}}\right)$$

$$= \nu(x) + \mu(x) - \nu(x)$$

$$= \mu(x).$$

if $\mu(x) \leq \nu(x)$

$$\Pr(X=x) = \mu(x) + 0 = \mu(x).$$

Similarly wrt Y.

Here, if tails, $X \neq Y$ since they are chosen from disjoint subsets of Ω

$$\Rightarrow \Pr(X \neq Y) = 1 - p = 1 - 1 + \|\mu - \nu\|_{TV}$$
$$= \|\mu - \nu\|_{TV}$$

Mixing Time

Let $$d(t) = \max_{x \in \Omega} \| \underbrace{p^t(x, \cdot)}_{\text{worst starting distribution}} - \underbrace{\pi}_{\text{Stationary Distribution}} \|_{TV}$$

Now $$t_{mix}(\varepsilon) = \min \{ t \mid \underbrace{d(t) \leq \varepsilon}_{\text{First time step where distance less than } \varepsilon.} \}$$

$d(t)$ and $\bar{d}(t)$

$$\bar{d}(t) = \max_{x, y} \| p^t(x, \cdot) - p^t(y, \cdot) \|_{TV}$$

<u>Claim</u>: $d(t) \leq \bar{d}(t)$

$$\| p^t(x,\cdot) - \pi \|_{TV} = \max_{A \subseteq \Omega} \left(p^t(x,A) - \pi(A) \right)$$

Now, since π is stationary,

$$\pi(z) = \sum_{y \in \Omega} \pi(y) \cdot p^t(y,z)$$

$$\Rightarrow \sum_{z \in A} \pi(z) = \sum_{y \in \Omega} \pi(y) \sum_{z \in A} p^t(y,z)$$

$$\therefore \| p^t(x,\cdot) - \pi \|_{TV} = \max_{A \subseteq \Omega} \left(p^t(x,A) - \sum_{y \in \Omega} \pi(y)\, p^t(y,A) \right)$$

$$= \max_{A \subseteq \Omega} \left(\sum_{y \in \Omega} \pi(y) \cdot p^t(x,A) - \sum_{y \in \Omega} \pi(y)\, p^t(y,A) \right)$$

$$\leq \sum_{y \in \Omega} \pi(y) \max_{A \subseteq \Omega} \left(p^t(x,A) - p^t(y,A) \right)$$

$$\leq \max_{y \in \Omega} \max_{A \subseteq \Omega} \left(p^t(x,A) - p^t(y,A) \right)$$

Exc: Show $\bar{d}(t) \leq 2d(t)$ (i.e. $\bar{d}(t)$ is a good approx for $d(t)$)

Markovian Coupling

A Markov Chain (X_t, Y_t) on $\Omega \times \Omega$ s.t.

i. $\Pr(X_{t+1} = x' \mid X_t = x, Y_t = y) = \Pr(x' \mid x)$

ii. $\Pr(Y_{t+1} = y' \mid X_t = x, Y_t = y) = \Pr(y' \mid y)$

Coupling Time

$$\tau_{couple} = \min\{t \mid X_s = Y_s, \text{ for } s \geq t\}$$

Claim: Let (X_t, Y_t) be a Markovian Coupling and $X_0 = x$, $Y_0 = y$.

$$\| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV} \leq P_{xy}(\tau_{couple} > t)$$

Distributions after X starts from x and Y from y.

Proof: $\| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV} \leq \Pr(X_t \neq Y_t)$

(from the previous lemma)

We can always couple so that once $X_t = Y_t$, they remain together. This can be done by making the same transition in both chains.

$$\Rightarrow \Pr(X_t \neq Y_t) = \Pr(\tau_{couple} > t)$$

Since once they meet they will be equal.

Corollary: Let $\forall\, x, y \in \Omega$ there is a coupling (X_t, Y_t) s.t. $X_0 = x$, $Y_0 = y$

$$d(t) = \max_{x} \| P^t(x,\cdot) - \pi \|_{TV}$$

$$\leq \bar{d}(t) = \max_{x,y\in\Omega} \| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV}$$

$$\leq \max_{x,y\in\Omega} \Pr(\tau_{couple} > t)$$

By Markov Inequality –

$$d(t) \leq \max_{x,y\in\Omega} \frac{\mathbb{E}[\tau_{couple}]}{t}$$

For $\tau_{mix}(\varepsilon)$, find min t s.t.

$$\max_{x,y\in\Omega} \frac{\mathbb{E}[\tau_{couple}]}{t} \leq \varepsilon$$

General Theorem

$$\tau_{mix}(\varepsilon) \leq \left\lceil \max_{x,y\in\Omega} \mathbb{E}[\tau_{couple}]\, e \ln\left(\frac{1}{\varepsilon}\right) \right\rceil$$