Total variation distance:

For distributions μ & π on finite Ω,

$$d_{TV}(\mu, \pi) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \pi(x)|$$

$$= \max_{S \subset \Omega} \mu(S) - \pi(S)$$

choose
$S = \{x : \mu(x) > \pi(x)\}$

$$= \max_{T \subset \Omega} \pi(T) - \mu(T)$$

$T = \{y : \pi(x) > \mu(x)\}$

= "area b/w μ & π"

Properties:

$$0 \leq d_{TV}(\mu, \pi) \leq 1$$

$$d_{TV}(\mu, \pi) \leq d_{TV}(\mu, \nu) + d_{TV}(\nu, \pi)$$

Also written as: $\|\mu - \pi\|_{TV}$ or $d_{TV}(\mu, \pi)$

& if Ω is infinite or continuous space, then replace max over $S \subset \Omega$ by sup.

How to bound variation distance?

<u>Coupling</u> method.

For μ & π on Ω,

distribution ω on product space $\Omega \times \Omega$ is a <u>coupling</u> of μ & π if:

(rows sum to μ) — for all $i \in \Omega$, $\sum_{j \in \Omega} \omega(i,j) = \mu(i)$

(columns sum to ν) — for all $j \in \Omega$, $\sum_{i \in \Omega} \omega(i,j) = \nu(j)$

<u>Example</u>: $\Omega = \{1,2,3,4\}$

$\mu = \left(\frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4}\right)$ $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0\right)$

$d_{TV}(\mu, \pi) = \frac{1}{2}\left(\frac{1}{6} + \frac{1}{12} + \frac{1}{3} + \frac{1}{4}\right) = \frac{5}{12}$

$$\omega = \begin{array}{c} \begin{bmatrix} \frac{1}{3} & \frac{1}{12} & \frac{1}{12} & 0 \\ 0 & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & \frac{1}{8} & 0 \end{bmatrix} \begin{array}{l} = \frac{1}{2} \\ = \frac{1}{4} \\ = 0 \\ = \frac{1}{4} \end{array} \\ \begin{array}{cccc} = \frac{1}{3} & = \frac{1}{3} & = \frac{1}{3} & = 0 \end{array} \end{array} \qquad \omega' = \begin{array}{c} \begin{bmatrix} \frac{1}{3} & \frac{1}{12} & \frac{1}{12} & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \end{bmatrix} \begin{array}{l} = \frac{1}{2} \\ = \frac{1}{4} \\ = 0 \\ = \frac{1}{4} \end{array} \\ \begin{array}{ccc} = \frac{1}{3} & = \frac{1}{3} & = \frac{1}{3} \end{array} \end{array}$$

Both ω & ω' are couplings of μ, π.

For coupling ω of μ, π, take:

Sample $(\sigma, \tau) \sim \omega$

Then $\sigma \sim \mu$ & $\tau \sim \pi$

(but they can be correlated with each other)

if uncorrelated (independent) then product coupling:

$$\omega'' = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 \end{bmatrix} \begin{matrix} = 1/2 \\ = 1/4 \\ = 0 \\ = 1/4 \end{matrix}$$

$$\begin{matrix} = & = & = & = \\ 1/3 & 1/3 & 1/3 & 0 \end{matrix}$$

<u>Coupling Lemma:</u> For <u>coupling</u> ω of μ, π,

let $(\sigma, \tau) \sim \omega$.

Then:

a) $d_{TV}(\mu, \pi) \leq \Pr(\sigma \neq \tau)$

b) $\exists$ coupling ω where

$d_{TV}(\mu, \pi) = \Pr(\sigma = \tau)$.

In words:

a) **Any** coupling upper bounds the variation distance.

b) There is always an *optimal* coupling.

In the previous example, note

for ω: $\Pr_{\omega}(\sigma \neq \tau) = \left(\frac{1}{12} + \frac{1}{12} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = \frac{13}{24} \geq \frac{5}{12}$

for ω': $\Pr_{\omega'}(\sigma \neq \tau) = \left(\frac{1}{12} \times 2 + \frac{1}{4}\right) = \frac{5}{12} = d_{TV}(\mu, \pi)$

So ω' is optimal.

Proof of (a): for $x \in \Omega$, $\omega(x,x) \leq \mu(x)$ & $\omega(x,x) \leq \pi(x)$

thus, $\omega(x,x) \leq \min\{\mu(x), \pi(x)\}$

Thus, $\Pr(\sigma = \tau) = \sum_{x} \omega(x,x) \leq \sum_{x} \min\{\mu(x), \pi(x)\}$

So, $\Pr(\sigma \neq \tau) = 1 - \Pr(\sigma = \tau) \geq 1 - \sum_{x} \min\{\mu(x), \pi(x)\}$

$= \sum_{x} \mu(x) - \min\{\mu(x), \pi(x)\}$

$= \max_{S \subseteq \Omega} \mu(S) - \pi(S) = d_{TV}(\mu, \pi)$. ∎

Proof of (b): we'll construct an optimal ω:

Set $\omega(x,x) = \min\{\mu(x), \pi(x)\}$

for off-diagonal entries set ω to be a product of remaining mass.

Consider a MC defined by P on Ω.

Two chains (X_t) & (Y_t).

Define joint evolution $(X_t, Y_t) \rightarrow (X_{t+1}, Y_{t+1})$

where $\Pr(X_{t+1} = j \mid X_t = i) = P(i,j)$

& $\Pr(Y_{t+1} = l \mid Y_t = k) = P(k,l)$

& if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

By the coupling lemma:

$$d_{TV}\left(P^t(X_0,\cdot), P^t(Y_0,\cdot)\right) \leq \Pr(X_t \neq Y_t)$$

Let

$$T_{couple} = \max_{i,j \in \Omega} \min\left\{t : \Pr\left(X_t \neq Y_t \mid X_0 = i, Y_0 = j\right) \leq \frac{1}{4}\right\}$$

Then, $T_{mix} \leq T_{couple}$.

RW on n-dimensional hypercube $\{0,1\}^n$. Let $\Omega = \{0,1\}^n$

From $X_t \in \Omega$,

1. Pick $i \in \{1, \ldots, n\}$ uar & $b \in \{0,1\}$ uar.
2. Set $X_{t+1}(i) = b$ & $X_{t+1}(j) = X_t(j)$ for $j \neq i$.

Consider 2 copies of this MC: (X_t) & (Y_t).

<u>Product coupling</u>:

if $X_t = Y_t$: choose same i, b.

else choose i, b for X_t

& independently choose i', b' for Y_t.

Let $A_t = \{ j : X_t(j) = Y_t(j) \}$ = "agree bits"

$D_t = \{ j : X_t(j) \neq Y_t(j) \}$ = "disagree"

~~Note~~ $\cancel{D_{t+1}}$

if $i \in D_t$ or $i' \in D_t$ then with Prob. $\frac{1}{2}$ agree afterwards

but if $i \in A_t$ or $i' \in A_t$ Then w.p. $\frac{1}{2}$ disagree after.

So if $|D_t| = 1$ then prob. $\approx \frac{1}{\cancel{2}n}$ that $X_{t+1} = Y_{t+1}$

& prob. $\approx \frac{1}{2}$ $|D_{t+1}| \geq 2$

<u>Roughly</u>: this takes exponential time to reach 0.

<u>Better coupling:</u> (identity coupling)

- choose same i, b for X_t, Y_t

if $i \in D_t$ then $|D_{t+1}| = |D_t| - 1$

& if $i \in A_t$ then $|D_{t+1}| = |D_t|$

note, $D_{t+1} \subseteq D_t$.

$$E\left[\,|D_{t+1}| \;\middle|\; X_t, Y_t\right] = |D_t| - \frac{|D_t|}{n} = |D_t|\left(1 - \frac{1}{n}\right)$$

$$\Pr(X_t \neq Y_t) \leq E\left[|D_t|\right]$$ because if $X_t \neq Y_t$ then $E\left[|D_t|\right] \geq 1$

$$\leq |D_0|\left(1 - \frac{1}{n}\right)^t$$

$$\leq n\left(1 - \frac{1}{n}\right)^t$$

$$\leq n e^{-t/n}$$

$$\leq \frac{1}{4} \text{ for } t \geq n \ln(4n).$$

So $T_{mix} = O(n \log n)$.

Top-to-Random shuffle:

Ω = all $n!$ permutations of $\{1,\dots,n\}$

RW: take top card & insert into random position i.

Ergodic: aperiodic & irreversible
but not symmetric. So why is $\pi = \text{uniform}(\Omega)$?

Note, P is doubly stochastic
(n transitions out of $X \in \Omega$
& n transitions into $X \in \Omega$
& every transition has prob. $1/n$).

HW: π is uniform iff P is doubly stochastic
(see last class notes for proof in case of top-to-random shuffle)

Simpler chain: Random-to-top

- choose random card c & move to top.

Coupling: X_t & Y_t choose same c.

After choose c then c is on top for both,

& c is always in same spot for both.

So just need to choose ~~each~~ every card ≥ 1 time.

Let T = time to get every card at least once.

t_i = time to get i^{th} new card after collecting $(i-1)^{\text{st}}$.

then $T = t_1 + t_2 + \cdots + t_n$

for t_i, need to choose any of $n-(i-1)$ cards

So let $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$

& $t_i = \text{geometric}(p_i)$

$$E[t_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$E[T] = \sum_{j=1}^{n} E[t_i] = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1}$$

$$= n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$$

$$\leq n(1 + \ln n).$$

<u>Markov's inequality</u>: $\Pr(T > 4E[T]) \leq 1/4$

So, $T_{mix} \leq 4n(1 + \ln n) = 4n\ln(en)$

$= O(n \log n).$

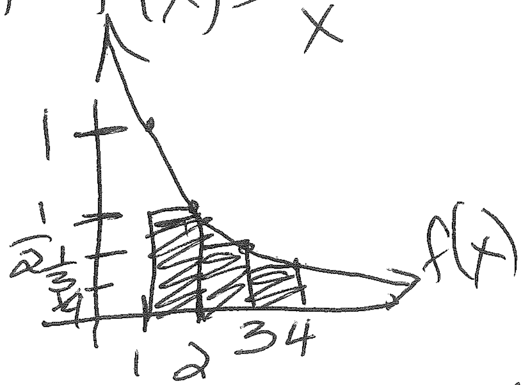
$\blacksquare$

Claim: $1 + \frac{1}{2} + \cdots + \frac{1}{n} \leq 1 + \ln n$

Proof: suffices to show: $\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \leq \ln n.$

let $f(x) = \frac{1}{x}$

then:



$$\frac{1}{2} + \cdots + \frac{1}{n} = \sum_{j=2}^{n} \frac{1}{j} \leq \int_{x=1}^{n} \frac{1}{x}\,dx = \ln x \Big|_{x=1}^{n} = \ln n$$

$\blacksquare$