Lecture 32: Strong uniform stopping rules

The variation distance between two probability dists Q_1 and Q_2 is

$$\|Q_1 - Q_2\| = \frac{1}{2} \sum_{\pi \in S_n} |Q_1(\pi) - Q_2(\pi)|$$

Write $Q_i(S) = \sum_{\pi \in S} Q_i(\pi)$.

We have: $\|Q_1 - Q_2\| = \max_{S \subseteq S_n} |Q_1(S) - Q_2(S)|$

proof: Let $B = \{\pi \in S_n : Q_1(\pi) > Q_2(\pi)\}$.

$Q_1(A) - Q_2(A) \leq Q_1(A \cap B) - Q_2(A \cap B)$, $A \subseteq S_n$

because $\pi \in A \cap B^c$ satisfies $Q_1(\pi) - Q_2(\pi) < 0$, so difference in prob cannot decrease when such elements are eliminated. Then

$$Q_1(A) - Q_2(A) \leq Q_1(B) - Q_2(B)$$

since including more elements from B cannot decrease the prob. difference.

But by the same reasoning,

$$Q_2(A) - Q_1(A) \leq Q_2(B^c) - Q_1(B^c)$$

Difference in upper bounds:

$$\big(Q_1(B) - Q_2(B)\big) - \big(Q_2(B^c) - Q_1(B^c)\big)$$

$$= Q_1(B + B^c) - Q_2(B + B^c) = 1 - 1 = 0!$$

Hence, $$Q_1(A) - Q_2(A) \leq C$$

$$Q_2(A) - Q_1(A) \leq C$$

ok

Now take $A = B$: $Q_1(A) - Q_2(A) \leq Q_1(A) - Q_2(A)$

Hence, $A = B$ saturates the upper bound.

(Same for $A = B^c$ for second inequality.)

Hence,

since these are the same

$$\max_{A \subseteq S_n} |Q_1(A) - Q_2(A)| = \frac{1}{2}\Big(\underbrace{Q_1(B) - Q_2(B)}_{>0} + \underbrace{Q_2(B^c) - Q_1(B^c)}_{>0}\Big)$$

$$= \frac{1}{2}\left[\sum_{\pi \in B}\big(Q_1(\pi) - Q_2(\pi)\big) + \sum_{\pi \in B^c} |Q_1(\pi) - Q_2(\pi)|\right]$$

$$= \frac{1}{2} \sum_{\pi \in S_n} |Q_1(\pi) - Q_2(\pi)|.$$

$\boxed{//}$

Recall the initial distribution I and the uniform distribution U.

At the start, we have:

$$\|I - U\| = \frac{1}{2} \sum_{\pi \in S_n} |I(\pi) - U(\pi)|$$

$$= \frac{1}{2} |\underbrace{I(id)}_{1} - \underbrace{U(id)}_{\frac{1}{n!}}| + \frac{1}{2} \sum_{\substack{\pi \in S_n \\ \pi \neq id}} |U(\pi)|$$

$$= \frac{1}{2}\left(1 - \frac{1}{n!}\right) + \frac{1}{2}(n! - 1)\frac{1}{n!}$$

$$= 1 - \frac{1}{n!}$$ ← very close to 1 (max distance)

Now perform one top-in-at-random shuffle. We have:

$$\mathrm{Top}(\tau_i) = 1/n, \text{ zero otherwise}$$

$$\tau_i = (2, 3, \ldots, i, 1, i+1, \ldots, n) \quad 1 \le i \le n,$$

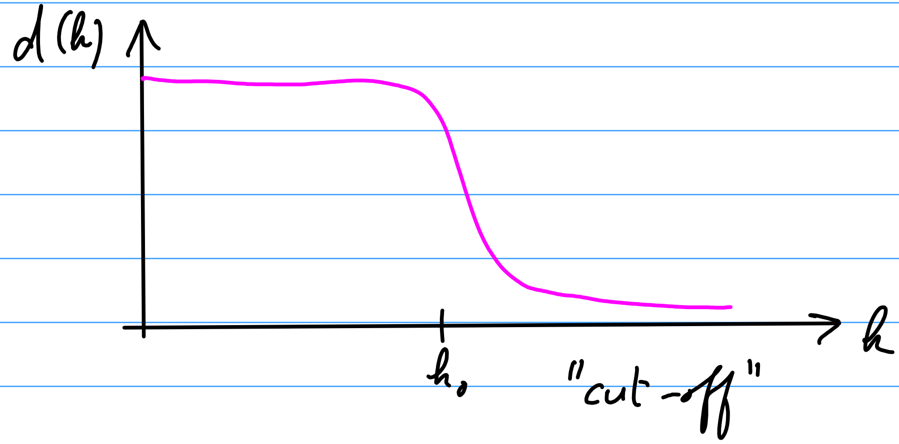
$$\| Top - U \| = \frac{1}{2} \sum_{\tau_i} \left| \frac{1}{n} - \frac{1}{n!} \right| + \frac{1}{2} \sum_{\pi \neq \tau_i} \frac{1}{n!}$$

$$= \frac{1}{2} n \left(\frac{1}{n} - \frac{1}{n!} \right) + \frac{1}{2} (n! - n) \frac{1}{n!}$$

$$= 1 - \frac{1}{(n-1)!}$$ Not a big improvement!

Let $d(k) := \| Top^{*k} - U \|$

k times



Strong uniform stopping rules: [Aldous & Diaconis]

Observe permutations, then yell STOP! using some stopping rule. Strong uniform if $\forall k \geq 0$

If the process is stopped after exactly k steps, then the resulting permutations of the deck have uniform distribution.

Let T be the number of steps performed until the stopping rule applies.

(T is a random var.)

Ordering of deck after k shuffles is $X_k \in S_n$.

Stopping rule is strong uniform if

$$\text{Prob}\left[X_k = \pi \mid T = k\right] = \frac{1}{n!}, \quad \forall \pi \in S_n$$

example: Top-in-at-random.

STOP after the original bottom card (n) is first inserted back into the deck.

is a strong uniform stopping rule.

This works because all the cards below card n are necessarily uniformly distributed.

Now let T_i = # of shuffles until for the first time i cards lie below card n.

We have to determine the distribution of

$$T = T_1 + (T_2 - T_1) + \cdots + (T_{n-1} - T_{n-2}) + (T - T_{n-1})$$

$T_i - T_{i-1}$ = time until top card is inserted at one of the i possible places below card n.

But this is the same as the time a coupon collector takes from coupon $(n-i)$ to $(n-i+1)$!

Think of it as reverse coupon collecting: at first it will be unlikely to insert the top card below the $n^{\underline{th}}$ one, which is like the collector trying to get his/her last coupon.

Thus T has the same distribution as V_n, and

$$\mathrm{Prob}[T > k] \le e^{-c}, \quad k = \lceil n \log n + cn \rceil$$

(from previous lecture)

Next lecture we will relate this to variation dist.