Discretized multinomial distributions and Nash equilibria in anonymous games

Constantinos Daskalakis and Christos Papadimitriou

The 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008), pp. 25–34.

Speaker: Joseph Chuang-Chieh Lin

Institute of Information Science
Academia Sinica
Taiwan

17 March 2017

Outline

1. Introduction
2. The Main Result
3. Sketch of the proofs
   - The Tickle-down Process
   - The Alternative Sampling of X_i
   - Clustering the Random Vectors
   - Discretization within a Cell
   - Coupling within a Cell

Anonymous games

k-strategy anonymous games

$(n, k, \{u_j^i\}_{i\in[n], j\in[k]})$:

- n: # players;
- k: # pure strategies per player;
- $u_j^i : \prod_{n-1}^k \mapsto [0, 1]$: utility function
  - $\prod_{n-1}^k := \{(x_1, \ldots, x_k) \in ([k] \cup \{0\})^k \mid \sum_{j\in[k]} x_j = n - 1\}$.
  - ⋆ All possible ways to partition $n - 1$ players into the k strategies.

- A mixed strategy profile: $\{\delta_i \in \Delta^k\}_{i\in[n]}$, where Δ^k denotes the set of distributions over $[k]$.

PTAS by Discretization (rough idea)

- Restrict our search to distributions with strategy probabilities being multiples of $1/z$, for some integer $z > 0$.
- Each such quantized mixed strategy can be considered as a *pure* strategy.
    - Utilities of the resulting new game can be computed via dynamic programming.
- Search for approx. NE by solving a corresponding max-flow problem.
- We've seen this for the 2-strategy case.

PTAS by Discretization (rough idea)

- Restrict our search to distributions with strategy probabilities being multiples of $1/z$, for some integer $z > 0$.
- Each such quantized mixed strategy can be considered as a *pure* strategy.
  - Utilities of the resulting new game can be computed via dynamic programming.
- Search for approx. NE by solving a corresponding max-flow problem.
- We've seen this for the 2-strategy case.

Remark on the dynamic programming (e.g., 2-strategy)

- Player n's expected payoff from strategy 1 (given a mixed strategy profile δ):

  $u_1^n((0, n-1)) \cdot \Pr[X = 0] + u_1^n((1, n-2)) \cdot \Pr[X = 1] + \cdots + u_1^n((n-1, 0)) \cdot \Pr[X = n-1]$,

  $X = \#$ players playing strategy 1 under profile δ_{-n}.

- Rewrite $X = \sum_{i=1}^{n-1} X_i$.
  - $X_i \in \{0, 1\}$: whether player i plays strategy 1.

- Consider a $(n-1) \times n$ table $T(i, \ell)$, for $i \in [n-1]$, $\ell \in \{0\} \cup [n-1]$, and $T(i, \ell) = \Pr[\sum_{j \le i} X_j = \ell]$.

$$T(i,\ell) = \begin{cases} \delta_i(1) \cdot T(i-1, \ell-1) + \delta_i(2) \cdot T(i-1, \ell), & \text{if } 0 < \ell < i; \\ \delta_i(1) \cdot T(i-1, i-1), & \text{if } \ell = i; \\ \delta_i(2) \cdot T(i-1, 0), & \text{if } \ell = 0; \\ 0, & \text{if } \ell > i. \end{cases}$$

⋆ Namely, $\Pr[X = \ell] = T(n-1, \ell)$, $\forall \ell \in \{0\} \cup [n-1]$.

Difficulties for the general cases of $k > 2$

- No useful approximations like Poisson approximation for the binomial distribution are known yet.
- Binomial case is easy because it's essentially one-dimensional.
  - In the multinomial case, watching the balls in one bin provides small information about the distribution of the remaining balls in other bins.
- We need something that combines multidimensional Poisson and translated-Poisson approximations *in the same picture*.

Useful lemmas and tools

The total variation distance

$\mathbb{P}, \mathbb{Q}$: two distributions supported by a finite set $\mathcal{A}$.

$$||\mathbb{P} - \mathbb{Q}|| \triangleq ||\mathbb{P} - \mathbb{Q}||_{TV} = \frac{1}{2} \cdot \sum_{\alpha \in \mathcal{A}} |\mathbb{P}(\alpha) - \mathbb{Q}(\alpha)| \, .$$

A simple & useful lemma

Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be two sets of mutually independent random vectors. Then

$$\left\| \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i \right\| \leq \sum_{i=1}^n \|X_i - Y_i\|.$$

Proof of the simple lemma

- By the coupling lemma, for any coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$,

$$\left\|\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i\right\| \le \Pr\left[\sum_{i=1}^n X_i \ne \sum_{i=1}^n Y_i\right] . \le \sum_{i=1}^n \Pr[X_i \ne Y_i] = \sum_{i=1}^n \|X_i - Y_i\|.$$

- By the optimal coupling theorem, there exists a coupling of X_i and Y_i such that $\Pr[X_i \ne Y_i] = \|X_i - Y_i\|$.

- Define a grand coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ such that $\Pr[X_i \ne Y_i] = \sum_{i=1}^n \|X_i - Y_i\|, \forall i$.

Proof of the simple lemma

- By the coupling lemma, for any coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$,

$$\left\| \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i \right\| \leq \Pr\left[\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i \right] . \leq \sum_{i=1}^n \Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|.$$

- By the optimal coupling theorem, there exists a coupling of X_i and Y_i such that $\Pr[X_i \neq Y_i] = \|X_i - Y_i\|$.

- Define a grand coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ such that $\Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|, \forall i$.

Proof of the simple lemma

- By the coupling lemma, for any coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$,

$$\left\|\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i\right\| \le \Pr\left[\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right] . \le \sum_{i=1}^n \Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|.$$

- By the optimal coupling theorem, there exists a coupling of X_i and Y_i such that $\Pr[X_i \neq Y_i] = \|X_i - Y_i\|$.

- Define a grand coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ such that $\Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|, \forall i$.

Proof of the simple lemma

- By the coupling lemma, for any coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$,

$$\left\| \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i \right\| \le \Pr\left[\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right]. \le \sum_{i=1}^n \Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|.$$

- By the optimal coupling theorem, there exists a coupling of X_i and Y_i such that $\Pr[X_i \neq Y_i] = \|X_i - Y_i\|$.
- Define a grand coupling of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ such that $\Pr[X_i \neq Y_i] = \sum_{i=1}^n \|X_i - Y_i\|, \forall i$.

Lemmas for clarification

Lemma A [Daskalakis & Papadimitriou, J. Econ. Theory 2015]

Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be two mixed strategy profiles of an anonymous game of n players and k strategies. Then for all $i \in [n]$ and $\ell \in [k]$:

$$\left| \mathbf{E}\left[u_\ell^i \left(\sum_{j \neq i} X_j \right) \right] - \mathbf{E}\left[u_\ell^i \left(\sum_{j \neq i} Y_j \right) \right] \right| \leq 2 \left\| \sum_{j \neq i} X_j - \sum_{j \neq i} Y_j \right\| .$$

A quick overview of Lemma A's proof

- Observe that $\mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] = \sum_{x\in\prod^k_{n-1}} u^i_\ell(x)\cdot \Pr[\sum_{j\neq i} X_j = x]$.
  - Similarly for $\mathbf{E}[u^i_\ell(\sum_{j\neq i} Y_j)]$.
- $$\begin{aligned} &\left|\mathbf{E}\left[u^i_\ell\left(\sum_{j\neq i} X_j\right)\right] - \mathbf{E}\left[u^i_\ell\left(\sum_{j\neq i} Y_j\right)\right]\right| \\ &= \left|\sum_{x\in\prod^k_{n-1}} u^i_\ell(x)\cdot\left(\Pr\left[\sum_{j\neq i} X_j = x\right] - \Pr\left[\sum_{j\neq i} Y_j = x\right]\right)\right| \\ &\leq \sum_{x\in\prod^k_{n-1}} |u^i_\ell(x)|\cdot\left|\Pr\left[\sum_{j\neq i} X_j = x\right] - \Pr\left[\sum_{j\neq i} Y_j = x\right]\right| \\ &\leq 2\left\|\sum_{j\neq i} X_j - \sum_{j\neq i} Y_j\right\|. \end{aligned}$$

A Lemma for clarification

Lemma B [Daskalakis & Papadimitriou 2015]

Suppose $\{X_i\}_{i\in[n]}$ is a NE of an anonymous game of n players and k strategies, and $\{Y_i\}_{i\in[n]}$ is a mixed strategy profile satisfying:

(a.) the support of Y_i is a subset of the support of X_i, for all i;

(b.) $\exists\epsilon \geq 0$, $\left\|\sum_{j\neq i} X_j - \sum_{j\neq i} Y_j\right\| \leq \epsilon$, for all i.

Then, $\{Y_i\}_{i=1}^{n}$ is a 4ϵ-NE.

For every ℓ in the support of Y_i:

- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)]$ for all $\ell' \in [k]$.
- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} Y_j)] \geq \mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] - 2\epsilon$;
  $\mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} Y_j)] - 2\epsilon$. (by Lemma A)

A Lemma for clarification

Lemma B [Daskalakis & Papadimitriou 2015]

Suppose $\{X_i\}_{i\in[n]}$ is a NE of an anonymous game of n players and k strategies, and $\{Y_i\}_{i\in[n]}$ is a mixed strategy profile satisfying:

(a.) the support of Y_i is a subset of the support of X_i, for all i;

(b.) $\exists\epsilon \geq 0$, $\left\|\sum_{j\neq i} X_j - \sum_{j\neq i} Y_j\right\| \leq \epsilon$, for all i.

Then, $\{Y_i\}_{i=1}^{n}$ is a 4ϵ-NE.

For every ℓ in the support of Y_i:

- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)]$ for all $\ell' \in [k]$.
- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} Y_j)] \geq \mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] - 2\epsilon$;

  $\mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} Y_j)] - 2\epsilon$. (by Lemma A)

A Lemma for clarification

Lemma B [Daskalakis & Papadimitriou 2015]

Suppose $\{X_i\}_{i\in[n]}$ is a NE of an anonymous game of n players and k strategies, and $\{Y_i\}_{i\in[n]}$ is a mixed strategy profile satisfying:

(a.) the support of Y_i is a subset of the support of X_i, for all i;

(b.) $\exists\epsilon \geq 0$, $\left\|\sum_{j\neq i} X_j - \sum_{j\neq i} Y_j\right\| \leq \epsilon$, for all i.

Then, $\{Y_i\}_{i=1}^{n}$ is a 4ϵ-NE.

For every ℓ in the support of Y_i:

- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)]$ for all $\ell' \in [k]$.
- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} Y_j)] \geq \mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] - 2\epsilon$;

  $\mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} Y_j)] - 2\epsilon$. (by Lemma A)

A Lemma for clarification

Lemma B [Daskalakis & Papadimitriou 2015]

Suppose $\{X_i\}_{i\in[n]}$ is a NE of an anonymous game of n players and k strategies, and $\{Y_i\}_{i\in[n]}$ is a mixed strategy profile satisfying:

(a.) the support of Y_i is a subset of the support of X_i, for all i;

(b.) $\exists\epsilon \geq 0$, $\left\|\sum_{j\neq i} X_j - \sum_{j\neq i} Y_j\right\| \leq \epsilon$, for all i.

Then, $\{Y_i\}_{i=1}^{n}$ is a 4ϵ-NE.

For every ℓ in the support of Y_i:

- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)]$ for all $\ell' \in [k]$.
- $\mathbf{E}[u^i_\ell(\sum_{j\neq i} Y_j)] \geq \mathbf{E}[u^i_\ell(\sum_{j\neq i} X_j)] - 2\epsilon$;
  $\mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} X_j)] \geq \mathbf{E}[u^i_{\ell'}(\sum_{j\neq i} Y_j)] - 2\epsilon$. (by Lemma A)

The main theorem

Theorem 2.1

- $\{p_i \in \Delta^k\}_{i=1}^n$;
- $\{X_i \in \mathbb{R}^k\}_{i=1}^n$, $\Pr[X_i = e_\ell] = p_{i,\ell}$, $\forall i \in [n], \ell \in [k]$.
- $z > 0$: a positive integer.

Then $\exists$ $\{\hat{p}_i \in \Delta^k\}_{i=1}^n$, such that:

1. $|\hat{p}_{i,\ell} - p_{i,\ell}| = O(1/z)$, $\forall i \in [n], \ell \in [k]$.
2. $\hat{p}_{i,\ell}$ is an integer multiple of $\frac{1}{2^k}\frac{1}{z}$, $\forall i \in [n], \ell \in [k]$.
3. if $p_{i,\ell} = 0$, then $\hat{p}_{i,\ell} = 0$, $\forall i \in [n], \ell \in [k]$.
4. if $\{\hat{X}_i \in \mathcal{R}^k\}_{i=1}^n$ are independent random unit vectors s.t. $\Pr[\hat{X}_i = e_\ell] = \hat{p}_{i,\ell}$, $\forall i \in [n], \ell \in [k]$, then
$$\left\| \sum_i X_i - \sum_i \hat{X}_i \right\| = O\left(f(k) \frac{\log z}{z^{1/5}} \right) = O(f(k) z^{-1/6}),$$
and $\forall j \in [n]$,
$$\left\| \sum_{i \neq j} X_i - \sum_{i \neq j} \hat{X}_i \right\| = O\left(f(k) \frac{\log z}{z^{1/5}} \right) = O(f(k) z^{-1/6}),$$
where $f(k)$ is an exponential function of k.

The constructive proof for the PTAS for the k-strategy case

Theorem 2.2

There is a PTAS for finding a mixed NE for the k-strategy anonymous game.

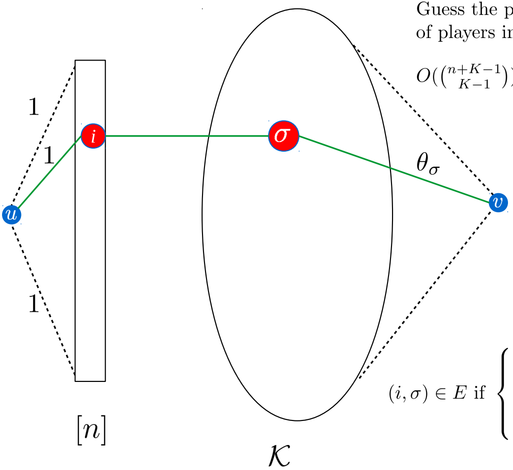
Sketch of the proof:

- Let $(p_1, \ldots, p_n)$ be a mixed NE of the game.
- Take $z = (f(k)/\epsilon)^6$, then $(\hat{p}_1, \ldots, \hat{p}_n)$ is an $O(\epsilon)$-NE.
- How to compute such $\{\hat{p}_i\}_{i=1}^{n}$?

Sketch of the proof of Theorem 2.2

- Remember, $\hat{p}_{i,\ell}$ is an integer multiple of $1/(2^k z)$, for each i, ℓ.
- We proceed with a related K-strategy game, for $K := (2^k z)^k = 2^{k^2}(f(k)/\epsilon)^{6k}$, and seek for its pure NE.
  - Let $\mathcal{K}$ denote the set of such quantized mixed strategies.
- The payoffs resulting from the new game: translating the pure strategy profile into a mixed strategy profile of the original game.

Seeking for the assignment of strategies: the max-flow reduction

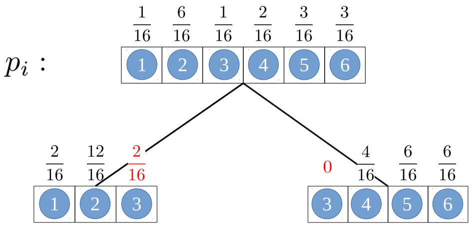


Guess the partition of the number n of players into $\mathcal{K}$

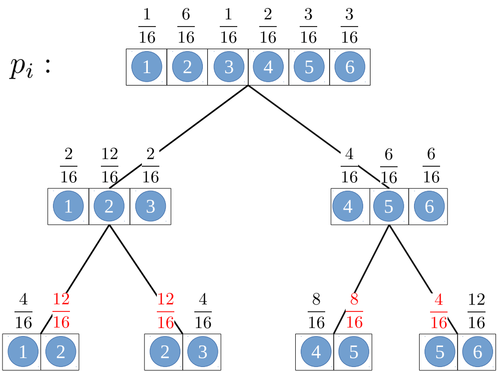
$O(\binom{n+K-1}{K-1}) = O((n+1)^{K-1})$possible guesses

$$(i,\sigma) \in E \text{ if } \begin{cases} \theta_\sigma > 0, \text{ and} \\ \sigma \text{ is an } O(\epsilon)\text{-best response if} \\ \text{the partition of } [n] \setminus \{i\} \text{ into } \mathcal{K} \\ \text{is } \theta \text{ with } \theta_\sigma - 1 \end{cases}$$

The trickle-down process (TDP)



The trickle-down process (TDP)



The properties of TDP

- All leaves have just two strategies.
- At each level, the two sets where the set of strategies is split overlap in at most one strategy.
  - The probability mass of such strategy is divided between its two copies.
- Each node of the tree (T_i, with node set V_i) represents a distribution.
  - $\forall v \in V_i$, v is identified with $(S_v, p_{i,v})$, where $S_v \subseteq [k]$.
  - $p_{i,v}$: is distribution over S_v.

- Based on T_i, we have an alternative way to sample X_i:

The alternative sampling of X_i

SAMPLING X_i

1. (Stage 1) Perform a random walk from the root T_i to the leaves:
   - At every non-leaf node, the left or right child is chosen w.p. $1/2$;
   - $\Phi_i \in \partial T_i$: the leaf chosen by the random walk;
2. (Stage 2) Let (S, p) be the label assigned to the leaf Φ_i.
   - $S = \{\ell_1, \ell_2\}$;
   - set $X_i = e_{\ell_1}$ w.p. $p(\ell_1)$, and $X_i = e_{\ell_2}$ w.p. $p(\ell_2)$.

Lemma 3.1

$\forall i \in [n]$, the process SAMPLING X_i outputs $X_i = e_\ell$ w.p. $p_{i,\ell}$, $\forall \ell \in [k]$.

Clustering the random vectors by Cells

Cell

Two vectors X_i and X_j belong to the same cell if

- $\exists$ isomorphism $f_{i,j} : V_i \mapsto V_j$ between T_i and T_j such that $\forall u \in V_i, v \in V_j$,
  - if $f_{i,j}(u) = v$, then $S_u = S_v$;
  - the elements of S_u and S_v are *ordered the same way* by $p_{i,u}$ and $p_{j,v}$.
- if $u \in \partial T_i$, $v = f_{i,j}(u) \in \partial T_j$, and $\ell^* \in S_u = S_v$ is the strategy with the *smallest* probability mass for both $p_{i,u}$ and $p_{j,v}$, then
  - either $p_{i,u}(\ell^*), p_{i,v}(\ell^*) \leq \frac{\lfloor z^\alpha \rfloor}{z}$ (Type A);
  - or $p_{i,u}(\ell^*), p_{i,v}(\ell^*) > \frac{\lfloor z^\alpha \rfloor}{z}$ (Type B).

Claim 3.3

Any tree resulting from TDP has $\leq k-1$ leaves, and the total number of cells is $\leq g(k) := k^{k^2} 2^{k-1} 2^k k!$.

Clustering the random vectors by Cells

Cell

Two vectors X_i and X_j belong to the same cell if

- $\exists$ isomorphism $f_{i,j} : V_i \mapsto V_j$ between T_i and T_j such that $\forall u \in V_i, v \in V_j$,
  - if $f_{i,j}(u) = v$, then $S_u = S_v$;
  - the elements of S_u and S_v are *ordered the same way* by $p_{i,u}$ and $p_{j,v}$.
- if $u \in \partial T_i$, $v = f_{i,j}(u) \in \partial T_j$, and $\ell^* \in S_u = S_v$ is the strategy with the *smallest* probability mass for both $p_{i,u}$ and $p_{j,v}$, then
  - either $p_{i,u}(\ell^*), p_{i,v}(\ell^*) \leq \frac{\lfloor z^\alpha \rfloor}{z}$ (Type A);
  - or $p_{i,u}(\ell^*), p_{i,v}(\ell^*) > \frac{\lfloor z^\alpha \rfloor}{z}$ (Type B).

Claim 3.3

Any tree resulting from TDP has $\leq k-1$ leaves, and the total number of cells is $\leq g(k) := k^{k^2} 2^{k-1} 2^k k!$.

On the number of cells and leaves

- TDP has $\leq k-1$ leaves by induction.
  - $k=2$: True.
  - For general k,
    the left subtree has j strategies $\Rightarrow\ \leq j-1$ leaves
    the right subtree has $\leq k+1-j$ strategies $\Rightarrow\ \leq k-j$ leaves.
- Let $T(m)$ denote the number of trees for some fixed set of m strategies and their ordering at the tree root.
  - ⋆ $T(m) \approx \sum_{j=2}^{m-1} j \cdot T(j) \cdot T(m+1-j) < k^{k^2}$.

Discretization within a Cell

- Denote all the isomorphic trees of a particular cell w.r.t $\mathcal{I} \subset [n]$ by T.

Rounding

$\forall v \in \partial T$ with $S_v = \{\ell_1, \ell_2\}$, $\ell_1, \ell_2 \in [k]$, do

1. Find a set of probabilities $\{q_{i,\ell_1}\}_{i \in \mathcal{I}}$ with the following properties:
   - $\forall i \in \mathcal{I}$, $|q_{i,\ell_1} - p_{i,v}(\ell_1)| \leq 1/z$;
   - $\forall i \in \mathcal{I}$, q_{i,ℓ_1} is an integer multiple of $1/z$;
   - $|\sum_{i \in \mathcal{I}} q_{i,\ell_1} - \sum_{i \in \mathcal{I}} p_{i,v}(\ell_1)| \leq 1/z$;
2. $\forall i \in \mathcal{I}$, set $\hat{p}_{i,v}(\ell_1) := q_{i,\ell_1}$, $\hat{p}_{i,v}(\ell_2) := 1 - q_{i,\ell_1}$.

- $\hat{p}_i(\ell) := \sum_{\substack{v \in \partial T \\ \ell \in S_v}} 2^{-\mathrm{depth}_T(v)} \hat{p}_{i,v}(\ell)$.

Distribution on the TDP tree leaves

Some notations

- $\Phi_i \in \partial T$: the leaf chosen by Stage 1 of SAMPLING X_i;
  $\hat{\Phi}_i \in \partial T$: the leaf chosen by Stage 1 of SAMPLING $\hat{X}_i$;
- Let $\Phi = (\Phi_i)_{i \in \mathcal{I}}$;
  - let G denote the distribution of Φ;

  Let $\hat{\Phi} = (\hat{\Phi}_i)_{i \in \mathcal{I}}$;
  - let $\hat{G}$ denote the distribution of $\hat{\Phi}$;

Distribution on the TDP tree leaves

Some more notations

$\forall v \in \partial T$, with $S_v = \{\ell_1, \ell_2\}$ and ordering (ℓ_1, ℓ_2):

- $\mathcal{I}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \mathcal{I}_v$ iff $i \in \mathcal{I} \wedge \Phi_i = v$;
  $\hat{\mathcal{I}}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \hat{\mathcal{I}}_v$ iff $i \in \mathcal{I}_v \wedge \hat{\Phi}_i = v$;
- $\mathcal{J}_{v,1} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,1}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_1}$;
  - Let F_v denote the distribution of $|\mathcal{J}_{v,1}|$.

  $\mathcal{J}_{v,2} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,2}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_2}$;
- Let $\mathcal{J} := ((|\mathcal{J}_{v,1}|, |\mathcal{J}_{v,2}|))_{v \in \partial T}$;
  - Let F denote the distribution of $\mathcal{J}$;
- Let $\hat{\mathcal{J}}_{v,1}$, $\hat{\mathcal{J}}_{v,2}$, $\hat{\mathcal{J}}$, $\hat{F}_v$, and $\hat{F}$ be defined similarly.

Distribution on the TDP tree leaves

Some more notations

$\forall v \in \partial T$, with $S_v = \{\ell_1, \ell_2\}$ and ordering (ℓ_1, ℓ_2):

- $\mathcal{I}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \mathcal{I}_v$ iff $i \in \mathcal{I} \wedge \Phi_i = v$;
  $\hat{\mathcal{I}}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \hat{\mathcal{I}}_v$ iff $i \in \mathcal{I}_v \wedge \hat{\Phi}_i = v$;
- $\mathcal{J}_{v,1} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,1}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_1}$;
  - Let F_v denote the distribution of $|\mathcal{J}_{v,1}|$.

  $\mathcal{J}_{v,2} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,2}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_2}$;
- Let $\mathcal{J} := ((|\mathcal{J}_{v,1}|, |\mathcal{J}_{v,2}|))_{v \in \partial T}$;
  - Let F denote the distribution of $\mathcal{J}$;
- Let $\hat{\mathcal{J}}_{v,1}$, $\hat{\mathcal{J}}_{v,2}$, $\hat{\mathcal{J}}$, $\hat{F}_v$, and $\hat{F}$ be defined similarly.

Distribution on the TDP tree leaves

Some more notations

$\forall v \in \partial T$, with $S_v = \{\ell_1, \ell_2\}$ and ordering (ℓ_1, ℓ_2):

- $\mathcal{I}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \mathcal{I}_v$ iff $i \in \mathcal{I} \wedge \Phi_i = v$;
  $\hat{\mathcal{I}}_v \subseteq \mathcal{I}$: the index set s.t. $i \in \hat{\mathcal{I}}_v$ iff $i \in \mathcal{I}_v \wedge \hat{\Phi}_i = v$;
- $\mathcal{J}_{v,1} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,1}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_1}$;
  - Let F_v denote the distribution of $|\mathcal{J}_{v,1}|$.

  $\mathcal{J}_{v,2} \subseteq \mathcal{I}_v$: the index set s.t. $i \in \mathcal{J}_{v,2}$ iff $i \in \mathcal{I}_v \wedge X_i = e_{\ell_2}$;
- Let $\mathcal{J} := ((|\mathcal{J}_{v,1}|, |\mathcal{J}_{v,2}|))_{v \in \partial T}$;
  - Let F denote the distribution of $\mathcal{J}$;
- Let $\hat{\mathcal{J}}_{v,1}$, $\hat{\mathcal{J}}_{v,2}$, $\hat{\mathcal{J}}$, $\hat{F}_v$, and $\hat{F}$ be defined similarly.

Coupling within a cell

Claim 3.4

$\forall \theta \in (\partial T)^{\mathcal{I}}$, $G(\theta) = \hat{G}(\theta)$.

Lemma 3.5

There exists a value of α such that, for all $v \in \partial T$,

$$G\left(\theta : \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - \frac{4}{z^{1/3}},$$

where $F_v(\cdot \mid \Phi)$ (resp., $\hat{F}_v(\cdot \mid \Phi)$) denotes the conditional probability distribution of $|\mathcal{J}_{v,1}|$ (resp., $|\hat{\mathcal{J}}_{v,1}|$) given Φ.

Lemma 3.5

There exists a value of α such that, for all $v \in \partial T$,

$$G\left(\theta : \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - \frac{4}{z^{1/3}},$$

where $F_v(\cdot \mid \Phi)$ (resp., $\hat{F}_v(\cdot \mid \Phi)$) denotes the conditional probability distribution of $|\mathcal{J}_{v,1}|$ (resp., $|\hat{\mathcal{J}}_{v,1}|$) given Φ.

- Roughly speaking, for all $v \in \partial T$, w.p. $\geq 1 - \frac{4}{z^{1/3}}$ over the choices made by Stage 1 of processes $\{\text{SAMPLING } X_i\}_{i \in \mathcal{I}}$ and $\{\text{SAMPLING } \hat{X}_i\}_{i \in \mathcal{I}}$ (assuming these processes are coupled to make the same decisions in Stage 1),

  the total variation distance b/w the conditional distribution of $|\mathcal{J}_{v,1}|$ and $|\hat{\mathcal{J}}_{v,1}|$ is bounded by $O\left(\frac{2^k \log z}{z^{1/5}}\right)$.

Lemma 3.6

Lemma 5 impies $\|F - \hat{F}\| \leq O(k\frac{2^k \log z}{z^{1/5}})$.

- Hence, $\left\|\sum_{i\in\mathcal{I}} X_i - \sum_{i\in\mathcal{I}} \hat{X}_i\right\| = O(k2^k \log z \cdot z^{-1/5})$.

Proof of Lemma 3.6

- Via a union bound, we have

$$G\left(\theta : \forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3}).$$

- Suppose for some $\theta \in (\partial T)^{\mathcal{I}}$, the following is satisfied

$$\forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right).$$

- $\{|\mathcal{J}_{v,1}|\}_{v \in \partial T}$ (resp., $\{|\hat{\mathcal{J}}_{v,1}|\}_{v \in \partial T}$) are conditionally independent given Φ (resp., $\hat{\Phi}$). By the coupling lemma,

$$\|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right).$$

- Therefore,

$$G\left(\theta : \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3})$$

Proof of Lemma 3.6

- Via a union bound, we have

$$G\left(\theta : \forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \ge 1 - O(kz^{-1/3}).$$

- Suppose for some $\theta \in (\partial T)^{\mathcal{I}}$, the following is satisfied

$$\forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(\frac{2^k \log z}{z^{1/5}}\right).$$

- $\{|\mathcal{J}_{v,1}|\}_{v \in \partial T}$ (resp., $\{|\hat{\mathcal{J}}_{v,1}|\}_{v \in \partial T}$) are conditionally independent given Φ (resp., $\hat{\Phi}$). By the coupling lemma,

$$\|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(k \frac{2^k \log z}{z^{1/5}}\right).$$

- Therefore,

$$G\left(\theta : \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(k \frac{2^k \log z}{z^{1/5}}\right)\right) \ge 1 - O(kz^{-1/3})$$

Proof of Lemma 3.6

- Via a union bound, we have

$$G\left(\theta : \forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3}).$$

- Suppose for some $\theta \in (\partial T)^{\mathcal{I}}$, the following is satisfied

$$\forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right).$$

- $\{|\mathcal{J}_{v,1}|\}_{v \in \partial T}$ (resp., $\{|\hat{\mathcal{J}}_{v,1}|\}_{v \in \partial T}$) are conditionally independent given Φ (resp., $\hat{\Phi}$). By the coupling lemma,

$$\|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right).$$

- Therefore,

$$G\left(\theta : \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3})$$

Proof of Lemma 3.6

- Via a union bound, we have

$$G\left(\theta : \forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3}).$$

- Suppose for some $\theta \in (\partial T)^{\mathcal{I}}$, the following is satisfied

$$\forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(\frac{2^k \log z}{z^{1/5}}\right).$$

- $\{|\mathcal{J}_{v,1}|\}_{v \in \partial T}$ (resp., $\{|\hat{\mathcal{J}}_{v,1}|\}_{v \in \partial T}$) are conditionally independent given Φ (resp., $\hat{\Phi}$). By the coupling lemma,

$$\|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right).$$

- Therefore,

$$G\left(\theta : \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \leq O\left(k \frac{2^k \log z}{z^{1/5}}\right)\right) \geq 1 - O(kz^{-1/3}).$$

Proof of Lemma 3.6

- Via a union bound, we have

$$G\left(\theta : \forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(\frac{2^k \log z}{z^{1/5}}\right)\right) \ge 1 - O(kz^{-1/3}).$$

- Suppose for some $\theta \in (\partial T)^{\mathcal{I}}$, the following is satisfied

$$\forall v \in \partial T, \|F_v(\cdot \mid \Phi = \theta) - \hat{F}_v(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(\frac{2^k \log z}{z^{1/5}}\right).$$

- $\{|\mathcal{J}_{v,1}|\}_{v \in \partial T}$ (resp., $\{|\hat{\mathcal{J}}_{v,1}|\}_{v \in \partial T}$) are conditionally independent given Φ (resp., $\hat{\Phi}$). By the coupling lemma,

$$\|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(k\frac{2^k \log z}{z^{1/5}}\right).$$

- Therefore,

$$G\left(\theta : \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(k\frac{2^k \log z}{z^{1/5}}\right)\right) \ge 1 - O(kz^{-1/3}).$$

Proof of Lemma 3.6 (contd.)

- $$\text{Good} := \left\{\theta : \theta \in (\partial T)^{\mathcal{I}}, \|F(\cdot \mid \Phi = \theta) - \hat{F}(\cdot \mid \hat{\Phi} = \theta)\| \le O\left(k\frac{2^k \log z}{z^{1/5}}\right)\right\},$$

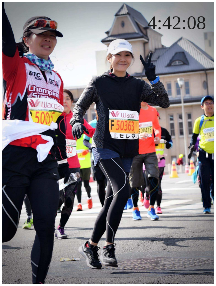
  and $\text{Bad} = (\partial T)^{\mathcal{I}} - \text{Good}$.
  - We knew that $G(\text{Bad}) \le O(kz^{-1/3})$.

- $$\begin{aligned} \|F - \hat{F}\| &= \frac{1}{2}\sum_t |F(t) - \hat{F}(t)| \\ &= \frac{1}{2}\sum_t \left|\sum_\theta F(t \mid \Phi = \theta)G(\Phi = \theta) - \hat{F}(t \mid \hat{\Phi} = \theta)\hat{G}(\Phi = \theta)\right| \\ &= \frac{1}{2}\sum_t \left|\sum_\theta F(t \mid \Phi = \theta) - \hat{F}(t \mid \hat{\Phi} = \theta))G(\theta)\right| \\ &\le \frac{1}{2}\sum_t \sum_\theta \left|F(t \mid \Phi = \theta) - \hat{F}(t \mid \hat{\Phi} = \theta))\right| G(\theta) \end{aligned}$$

Proof of Lemma 3.6 (contd.)

- $$\begin{aligned}\|F - \hat{F}\| \leq \ldots &= \frac{1}{2}\sum_{t}\sum_{\theta\in\text{Good}}\left|F(t \mid \Phi=\theta) - \hat{F}(t \mid \hat{\Phi}=\theta))\right| G(\theta)\\ &\quad + \frac{1}{2}\sum_{t}\sum_{\theta\in\text{Bad}}\left|F(t \mid \Phi=\theta) - \hat{F}(t \mid \hat{\Phi}=\theta))\right| G(\theta)\\ &\leq \sum_{\theta\in\text{Good}} G(\theta)\left(\frac{1}{2}\sum_{t}\left|F(t \mid \Phi=\theta) - \hat{F}(t \mid \hat{\Phi}=\theta))\right|\right)\\ &\quad + \sum_{\theta\in\text{Bad}} G(\theta)\left(\frac{1}{2}\sum_{t}\left|F(t \mid \Phi=\theta) - \hat{F}(t \mid \hat{\Phi}=\theta))\right|\right)\\ &\leq \sum_{\theta\in\text{Good}} G(\theta)\cdot O\left(k\frac{2^k \log z}{z^{1/5}}\right) + \sum_{\theta\in\text{Bad}} G(\theta)\\ &\leq O\left(k\frac{2^k \log z}{z^{1/5}}\right) + O(kz^{-1/3}).\end{aligned}$$

Thank you.


4:42:08


2:56:56


山中伸彌 3:27:45
Nobel Prize winner (2012)
© Center for iPS Cell Research and Application,Kyoto University

京都マラソン 2017