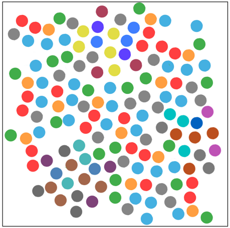
Extrapolating the profile of a finite population

Soham Jana[1], Yury Polyanskiy[2], Yihong Wu[1]

[1]Yale University
[2]Massachusetts Institute of Technology

July 11, 2020

Bernoulli sampling model [Bunge and Fitzpatrick, 1993]



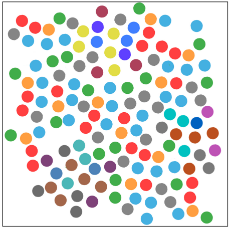
θ_1 Balls of color 1
θ_2 Balls of color 2
...
θ_k Balls of color k

- Setup: Population of k balls, each belonging to one of k types (color).
- Want to estimate: Profile of the urn [Orlitsky et al., 2005]:

$$\pi = \frac{1}{k}\sum_{j=1}^{k} \delta_{\theta_j}$$

  in total variation distance. Note that π_j gives us the proportion of color with exactly j balls.
- Data: X_j balls of color j, distributed as $\mathrm{Binom}(\theta_j, p)$. p might be vanishing, i.e. $p \xrightarrow{k\to\infty} 0$.
- Note that the empirical distribution of color contains more information, but requires more samples to learn. In particular, it cannot be estimated consistently in the sub-linear regime.

Bernoulli sampling model [Bunge and Fitzpatrick, 1993]



θ_1 Balls of color 1
θ_2 Balls of color 2
...
θ_k Balls of color k

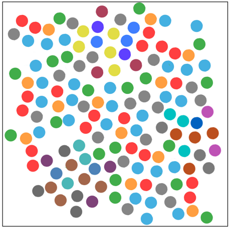
- Setup: Population of k balls, each belonging to one of k types (color).
- Want to estimate: Profile of the urn [Orlitsky et al., 2005]:

$$\pi = \frac{1}{k}\sum_{j=1}^{k} \delta_{\theta_j}$$

in total variation distance. Note that π_j gives us the proportion of color with exactly j balls.

- Data: X_j balls of color j, distributed as $\mathrm{Binom}(\theta_j, p)$. p might be vanishing, i.e. $p \overset{k\to\infty}{\longrightarrow} 0$.
- Note that the empirical distribution of color contains more information, but requires more samples to learn. In particular, it cannot be estimated consistently in the sub-linear regime.

Bernoulli sampling model [Bunge and Fitzpatrick, 1993]



θ_1 Balls of color 1
θ_2 Balls of color 2
...
θ_k Balls of color k

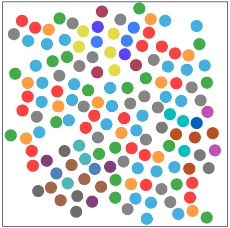
- Setup: Population of k balls, each belonging to one of k types (color).
- Want to estimate: Profile of the urn [Orlitsky et al., 2005]:

$$\pi = \frac{1}{k}\sum_{j=1}^{k}\delta_{\theta_j}$$

in total variation distance. Note that π_j gives us the proportion of color with exactly j balls.

- Data: X_j balls of color j, distributed as $\text{Binom}(\theta_j, p)$. p might be vanishing, i.e. $p \overset{k\to\infty}{\longrightarrow} 0$.
- Note that the empirical distribution of color contains more information, but requires more samples to learn. In particular, it cannot be estimated consistently in the sub-linear regime.

Bernoulli sampling model [Bunge and Fitzpatrick, 1993]



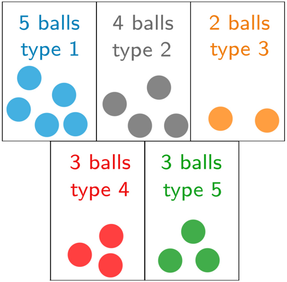
θ_1 Balls of color 1
θ_2 Balls of color 2
...
θ_k Balls of color k

- Setup: Population of k balls, each belonging to one of k types (color).
- Want to estimate: Profile of the urn [Orlitsky et al., 2005]:

$$\pi = \frac{1}{k}\sum_{j=1}^{k}\delta_{\theta_j}$$

  in total variation distance. Note that π_j gives us the proportion of color with exactly j balls.
- Data: X_j balls of color j, distributed as $\mathrm{Binom}(\theta_j, p)$. p might be vanishing, i.e. $p \overset{k\to\infty}{\longrightarrow} 0$.
- Note that the empirical distribution of color contains more information, but requires more samples to learn. In particular, it cannot be estimated consistently in the sub-linear regime.

Example

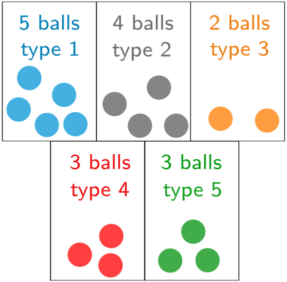


Urn of size 17. The distribution of color is

- 5 blue balls
- 4 gray balls
- 2 orange balls
- 3 red balls
- 3 green balls.

Then the empirical distribution of colors (μ) is given by

$$\mu(\text{blue}) = \frac{5}{17}, \mu(\text{gray}) = \frac{4}{17}, \mu(\text{orange}) = \frac{2}{17},$$
$$\mu(\text{red}) = \frac{3}{17}, \mu(\text{green}) = \frac{3}{17}.$$

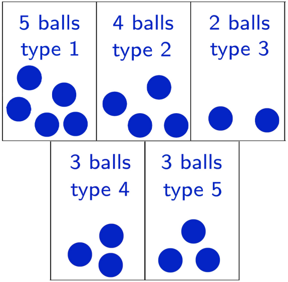


Urn of size 17. The distribution of color is

- 5 blue balls
- 4 gray balls
- 2 orange balls
- 3 red balls
- 3 green balls.

Then the empirical distribution of colors (μ) is given by

$$\mu(\text{blue}) = \frac{5}{17}, \mu(\text{gray}) = \frac{4}{17}, \mu(\text{orange}) = \frac{2}{17},$$
$$\mu(\text{red}) = \frac{3}{17}, \mu(\text{green}) = \frac{3}{17}.$$

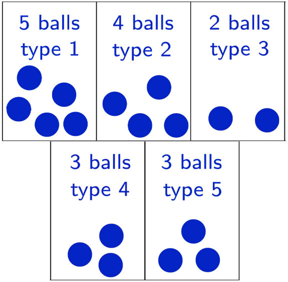


- The profile (π) depends on the color-deleted version of the urn.
- π is supported on $\{0, 1, \ldots, 17\}$ and is given by

$$\pi_m = \begin{cases} 1 - \frac{5}{17} & \text{if } m = 0 \\ \frac{1}{17} & \text{if } m = 2, 4, 5 \\ \frac{2}{17} & \text{if } m = 3 \\ 0 & \text{otherwise.} \end{cases}$$

- π_0 gives us total number of distinct colors (C)

$$\pi_0 = 1 - \frac{C}{k}.$$



- The profile (π) depends on the color-deleted version of the urn.

- π is supported on $\{0, 1, \ldots, 17\}$ and is given by

$$\pi_m = \begin{cases} 1 - \frac{5}{17} & \text{if } m = 0 \\ \frac{1}{17} & \text{if } m = 2, 4, 5 \\ \frac{2}{17} & \text{if } m = 3 \\ 0 & \text{otherwise.} \end{cases}$$

- π_0 gives us total number of distinct colors (C)

$$\pi_0 = 1 - \frac{C}{k}.$$

Usefulness

Usefulness of profile

[Orlitsky et al., 2005] Important label invariant properties (e.g. entropy, number of distinct species) are learnable through π.

Consider small sample regime (sample size vanishing fraction of k)

- Consistent estimation of μ is impossible.
- Consistent estimation of π is possible.
- Useful implication towards estimating label invariant properties from small sample.

The problem of estimating π is part of the program of "empirical Bayes" [Robbins, 1951, Robbins, 1956].

- Want to estimate functional f of $\vec{\theta} = (\theta_1, \ldots, \theta_j)$.
- The goal is to compete with the oracle estimator $\hat{f}(\vec{X}, \pi)$ which one can compute when the true π is known.
- When π is unknown we get estimator $\hat{\pi}$ of π and then substitute to get $\hat{f}(\vec{X}, \hat{\pi})$.

Question : How well the estimation of π can be done?

Usefulness of profile

[Orlitsky et al., 2005] Important label invariant properties (e.g. entropy, number of distinct species) are learnable through π.

Consider small sample regime (sample size vanishing fraction of k)

- Consistent estimation of μ is impossible.
- Consistent estimation of π is possible.
- Useful implication towards estimating label invariant properties from small sample.

The problem of estimating π is part of the program of "empirical Bayes" [Robbins, 1951, Robbins, 1956].

- Want to estimate functional f of $\vec{\theta} = (\theta_1, \ldots, \theta_j)$.
- The goal is to compete with the oracle estimator $\hat{f}(\vec{X}, \pi)$ which one can compute when the true π is known.
- When π is unknown we get estimator $\hat{\pi}$ of π and then substitute to get $\hat{f}(\vec{X}, \hat{\pi})$.

Question : How well the estimation of π can be done?

Usefulness of profile

[Orlitsky et al., 2005] Important label invariant properties (e.g. entropy, number of distinct species) are learnable through π.

Consider small sample regime (sample size vanishing fraction of k)

- Consistent estimation of μ is impossible.
- Consistent estimation of π is possible.
- Useful implication towards estimating label invariant properties from small sample.

The problem of estimating π is part of the program of "empirical Bayes" [Robbins, 1951, Robbins, 1956].

- Want to estimate functional f of $\vec{\theta} = (\theta_1, \ldots, \theta_j)$.
- The goal is to compete with the oracle estimator $\hat{f}(\vec{X}, \pi)$ which one can compute when the true π is known.
- When π is unknown we get estimator $\hat{\pi}$ of π and then substitute to get $\hat{f}(\vec{X}, \hat{\pi})$.

Question : How well the estimation of π can be done?

Existing literature

Distinct elements problem

Find estimator $\hat{\pi}_0$ such that

$$\max_{k \text{ ball urn}} \mathbb{E}|\hat{\pi}_0 - \pi_0| \xrightarrow{k\to\infty} 0.$$

- [Bunge and Fitzpatrick, 1993, Charikar et al., 2000, Raskhodnikova et al., 2009, Valiant and Valiant, 2011, Wu and Yang, 2018, ...]
- [Wu and Yang, 2018] If $\frac{1}{\log k} \lesssim p \lesssim 1$ the optimal rate of estimating π_0 is $k^{-\Theta(p)}$.
- [Valiant, 2012, Wu and Yang, 2018] Consistent estimation is not possible if $p = O\left(\frac{1}{\log k}\right)$.

Estimation of π_m, $m \geq 1$

- Our results refines the above for other atoms of π. We show that the polynomial rate $k^{-\Theta(p)}$ holds for all π_m with $m = o(\log k)$.
- For $m = \Theta(\log k)$ estimation is much harder, with rate $\Omega_p\left(\frac{1}{(\log k)^2}\right)$.

Distinct elements problem

Find estimator $\hat{\pi}_0$ such that

$$\max_{k \text{ ball urn}} \mathbb{E}|\hat{\pi}_0 - \pi_0| \xrightarrow{k\to\infty} 0.$$

- [Bunge and Fitzpatrick, 1993, Charikar et al., 2000, Raskhodnikova et al., 2009, Valiant and Valiant, 2011, Wu and Yang, 2018, ...]
- [Wu and Yang, 2018] If $\frac{1}{\log k} \lesssim p \lesssim 1$ the optimal rate of estimating π_0 is $k^{-\Theta(p)}$.
- [Valiant, 2012, Wu and Yang, 2018] Consistent estimation is not possible if $p = O\left(\frac{1}{\log k}\right)$.

Estimation of π_m, $m \geq 1$

- Our results refines the above for other atoms of π. We show that the polynomial rate $k^{-\Theta(p)}$ holds for all π_m with $m = o(\log k)$.
- For $m = \Theta(\log k)$ estimation is much harder, with rate $\Omega_p\left(\frac{1}{(\log k)^2}\right)$.

Distinct elements problem

Find estimator $\hat{\pi}_0$ such that

$$\max_{k \text{ ball urn}} \mathbb{E}|\hat{\pi}_0 - \pi_0| \xrightarrow{k\to\infty} 0.$$

- [Bunge and Fitzpatrick, 1993, Charikar et al., 2000, Raskhodnikova et al., 2009, Valiant and Valiant, 2011, Wu and Yang, 2018, ...]
- [Wu and Yang, 2018] If $\frac{1}{\log k} \lesssim p \lesssim 1$ the optimal rate of estimating π_0 is $k^{-\Theta(p)}$.
- [Valiant, 2012, Wu and Yang, 2018] Consistent estimation is not possible if $p = O\left(\frac{1}{\log k}\right)$.

Estimation of π_m, $m \geq 1$

- Our results refines the above for other atoms of π. We show that the polynomial rate $k^{-\Theta(p)}$ holds for all π_m with $m = o(\log k)$.
- For $m = \Theta(\log k)$ estimation is much harder, with rate $\Omega_p\left(\frac{1}{(\log k)^2}\right)$.

Distinct elements problem

Find estimator $\hat{\pi}_0$ such that

$$\max_{k \text{ ball urn}} \mathbb{E}|\hat{\pi}_0 - \pi_0| \xrightarrow{k\to\infty} 0.$$

- [Bunge and Fitzpatrick, 1993, Charikar et al., 2000, Raskhodnikova et al., 2009, Valiant and Valiant, 2011, Wu and Yang, 2018, ...]
- [Wu and Yang, 2018] If $\frac{1}{\log k} \lesssim p \lesssim 1$ the optimal rate of estimating π_0 is $k^{-\Theta(p)}$.
- [Valiant, 2012, Wu and Yang, 2018] Consistent estimation is not possible if $p = O\left(\frac{1}{\log k}\right)$.

Estimation of π_m, $m \geq 1$

- Our results refines the above for other atoms of π. We show that the polynomial rate $k^{-\Theta(p)}$ holds for all π_m with $m = o(\log k)$.
- For $m = \Theta(\log k)$ estimation is much harder, with rate $\Omega_p\left(\frac{1}{(\log k)^2}\right)$.

Distinct elements problem

Find estimator $\hat{\pi}_0$ such that

$$\max_{k \text{ ball urn}} \mathbb{E}|\hat{\pi}_0 - \pi_0| \xrightarrow{k\to\infty} 0.$$

- [Bunge and Fitzpatrick, 1993, Charikar et al., 2000, Raskhodnikova et al., 2009, Valiant and Valiant, 2011, Wu and Yang, 2018, ...]
- [Wu and Yang, 2018] If $\frac{1}{\log k} \lesssim p \lesssim 1$ the optimal rate of estimating π_0 is $k^{-\Theta(p)}$.
- [Valiant, 2012, Wu and Yang, 2018] Consistent estimation is not possible if $p = O\left(\frac{1}{\log k}\right)$.

Estimation of π_m, $m \geq 1$

- Our results refines the above for other atoms of π. We show that the polynomial rate $k^{-\Theta(p)}$ holds for all π_m with $m = o(\log k)$.
- For $m = \Theta(\log k)$ estimation is much harder, with rate $\Omega_p\left(\frac{1}{(\log k)^2}\right)$.

- Want to evaluate (for possibly vanishing p)

$$R(k) = \inf \sup \mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right].$$

- Sorted version of μ ($\mu^{\downarrow}$) and π are related

$$\|\pi^1 - \pi^2\|_{\mathrm{TV}} \le \|\mu^{1\downarrow} - \mu^{2\downarrow}\|_{\mathrm{TV}}.$$

- [Orlitsky et al., 2005, Valiant and Valiant, 2013, Han et al., 2018] studied estimation of $\mu^{\downarrow}$ for general population (might not be finite). There exists $\hat{\mu^{\downarrow}}$ based on sample of size $\Theta(k)$ such that

$$\mathbb{E}\left[\|\mu^{\downarrow} - \hat{\mu^{\downarrow}}\|_{\mathrm{TV}}\right] \le O\left(\frac{1}{\sqrt{\log k}}\right).$$

- This implies existence of $\hat{\pi}$ such that $\mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right] \le O\left(\frac{1}{\sqrt{\log k}}\right)$.
- For finite population this upper bound is loose by a square root factor in view of our main result.

- Want to evaluate (for possibly vanishing p)

$$R(k) = \inf \sup \mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right].$$

- Sorted version of μ ($\mu^{\downarrow}$) and π are related

$$\|\pi^1 - \pi^2\|_{\mathrm{TV}} \leq \|\mu^{1\downarrow} - \mu^{2\downarrow}\|_{\mathrm{TV}}.$$

- [Orlitsky et al., 2005, Valiant and Valiant, 2013, Han et al., 2018] studied estimation of $\mu^{\downarrow}$ for general population (might not be finite). There exists $\hat{\mu^{\downarrow}}$ based on sample of size $\Theta(k)$ such that

$$\mathbb{E}\left[\|\mu^{\downarrow} - \hat{\mu^{\downarrow}}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right).$$

- This implies existence of $\hat{\pi}$ such that $\mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right)$.
- For finite population this upper bound is loose by a square root factor in view of our main result.

- Want to evaluate (for possibly vanishing p)

$$R(k) = \inf \sup \mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right].$$

- Sorted version of μ ($\mu^{\downarrow}$) and π are related

$$\|\pi^1 - \pi^2\|_{\mathrm{TV}} \leq \|\mu^{1\downarrow} - \mu^{2\downarrow}\|_{\mathrm{TV}}.$$

- [Orlitsky et al., 2005, Valiant and Valiant, 2013, Han et al., 2018] studied estimation of $\mu^{\downarrow}$ for general population (might not be finite). There exists $\hat{\mu^{\downarrow}}$ based on sample of size $\Theta(k)$ such that

$$\mathbb{E}\left[\|\mu^{\downarrow} - \hat{\mu^{\downarrow}}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right).$$

- This implies existence of $\hat{\pi}$ such that $\mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right)$.
- For finite population this upper bound is loose by a square root factor in view of our main result.

- Want to evaluate (for possibly vanishing p)

$$R(k) = \inf \sup \mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right].$$

- Sorted version of μ ($\mu^{\downarrow}$) and π are related

$$\|\pi^1 - \pi^2\|_{\mathrm{TV}} \le \|\mu^{1\downarrow} - \mu^{2\downarrow}\|_{\mathrm{TV}}.$$

- [Orlitsky et al., 2005, Valiant and Valiant, 2013, Han et al., 2018] studied estimation of $\mu^{\downarrow}$ for general population (might not be finite). There exists $\hat{\mu^{\downarrow}}$ based on sample of size $\Theta(k)$ such that

$$\mathbb{E}\left[\|\mu^{\downarrow} - \hat{\mu^{\downarrow}}\|_{\mathrm{TV}}\right] \le O\left(\frac{1}{\sqrt{\log k}}\right).$$

- This implies existence of $\hat{\pi}$ such that $\mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right] \le O\left(\frac{1}{\sqrt{\log k}}\right)$.
- For finite population this upper bound is loose by a square root factor in view of our main result.

- Want to evaluate (for possibly vanishing p)

$$R(k) = \inf \sup \mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right].$$

- Sorted version of μ ($\mu^{\downarrow}$) and π are related

$$\|\pi^1 - \pi^2\|_{\mathrm{TV}} \leq \|\mu^{1\downarrow} - \mu^{2\downarrow}\|_{\mathrm{TV}}.$$

- [Orlitsky et al., 2005, Valiant and Valiant, 2013, Han et al., 2018] studied estimation of $\mu^{\downarrow}$ for general population (might not be finite). There exists $\hat{\mu^{\downarrow}}$ based on sample of size $\Theta(k)$ such that

$$\mathbb{E}\left[\|\mu^{\downarrow} - \hat{\mu^{\downarrow}}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right).$$

- This implies existence of $\hat{\pi}$ such that $\mathbb{E}\left[\|\pi - \hat{\pi}\|_{\mathrm{TV}}\right] \leq O\left(\frac{1}{\sqrt{\log k}}\right)$.
- For finite population this upper bound is loose by a square root factor in view of our main result.

Results

Main theorem

For all $k \geq \frac{d_0}{\bar{p}}$, where d_0 is some absolute constant, the following holds.

1. There exists absolute constant C such that

$$R(k) \leq \min\left\{\frac{C}{p \log k}, 1\right\}.$$

   The upper bound is achieved by minimum-distance estimator computable in polynomial time.

2. There exists absolute constant c such that

$$R(k) \geq \min\left\{\frac{\bar{p}}{p}, \sqrt{\log k}\right\} \frac{c}{\log k}$$

   where $\bar{p} = 1 - p$.

- This shows that in linear regime (i.e. constant p) the optimal TV rate is $\Theta\left(\frac{1}{\log k}\right)$.
- Consistent estimation is possible in vanishing regime $p = \omega\left(\frac{1}{\log k}\right)$.

Main theorem

For all $k \geq \frac{d_0}{\bar{p}}$, where d_0 is some absolute constant, the following holds.

1. There exists absolute constant C such that

$$R(k) \leq \min\left\{\frac{C}{p\log k}, 1\right\}.$$

   The upper bound is achieved by minimum-distance estimator computable in polynomial time.

2. There exists absolute constant c such that

$$R(k) \geq \min\left\{\frac{\bar{p}}{p}, \sqrt{\log k}\right\}\frac{c}{\log k}$$

   where $\bar{p} = 1 - p$.

- This shows that in linear regime (i.e. constant p) the optimal TV rate is $\Theta\left(\frac{1}{\log k}\right)$.
- Consistent estimation is possible in vanishing regime $p = \omega\left(\frac{1}{\log k}\right)$.

Main theorem

For all $k \geq \frac{d_0}{\bar{p}}$, where d_0 is some absolute constant, the following holds.

1. There exists absolute constant C such that

$$R(k) \leq \min\left\{\frac{C}{p\log k}, 1\right\}.$$

   The upper bound is achieved by minimum-distance estimator computable in polynomial time.

2. There exists absolute constant c such that

$$R(k) \geq \min\left\{\frac{\bar{p}}{p}, \sqrt{\log k}\right\}\frac{c}{\log k}$$

   where $\bar{p} = 1 - p$.

- This shows that in linear regime (i.e. constant p) the optimal TV rate is $\Theta\left(\frac{1}{\log k}\right)$.
- Consistent estimation is possible in vanishing regime $p = \omega\left(\frac{1}{\log k}\right)$.

Main theorem

For all $k \geq \frac{d_0}{\bar{p}}$, where d_0 is some absolute constant, the following holds.

1. There exists absolute constant C such that

$$R(k) \leq \min\left\{\frac{C}{p\log k}, 1\right\}.$$

   The upper bound is achieved by minimum-distance estimator computable in polynomial time.

2. There exists absolute constant c such that

$$R(k) \geq \min\left\{\frac{\bar{p}}{p}, \sqrt{\log k}\right\}\frac{c}{\log k}$$

   where $\bar{p} = 1 - p$.

- This shows that in linear regime (i.e. constant p) the optimal TV rate is $\Theta\left(\frac{1}{\log k}\right)$.
- Consistent estimation is possible in vanishing regime $p = \omega\left(\frac{1}{\log k}\right)$.

Connection to minimum distance estimator

General set up:

- Parameter space Θ.
- Distribution family $\{P_\theta : \theta \in \Theta\}$ with distance measure ρ.
- $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{\theta_j}$.
- Want to analyze

$$R(k) = \inf_{\hat{\pi}} \sup_{\theta_1,\ldots,\theta_k} \mathbb{E}\left[d(\hat{\pi},\pi)\right]$$

under some cost constraint $\frac{1}{k}\sum_{j=1}^{k} c(\theta_j) \leq 1$.

Data: $X_j \sim P_{\theta_j}$ independently for $j = 1, \ldots, k$.

Empirical estimate $\hat{\nu} = \frac{1}{k}\sum_{j=1}^{k} \delta_{X_j}$ satisfies

$$\mathbb{E}\left[\hat{\nu}\right] = \pi P.$$

This motivates estimation of π as

$$\hat{\pi} = \operatorname*{argmin}_{\pi'} \left\{\rho(\hat{\nu}, \pi' P) : \mathbb{E}_{\pi'}\left[c(\theta)\right] \leq 1\right\}.$$

Connection to minimum distance estimator

General set up:

- Parameter space Θ.
- Distribution family $\{P_\theta : \theta \in \Theta\}$ with distance measure ρ.
- $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{\theta_j}$.
- Want to analyze

$$R(k) = \inf_{\hat{\pi}} \sup_{\theta_1,\ldots,\theta_k} \mathbb{E}\left[d(\hat{\pi},\pi)\right]$$

under some cost constraint $\frac{1}{k}\sum_{j=1}^{k} c(\theta_j) \leq 1$.

Data: $X_j \sim P_{\theta_j}$ independently for $j = 1,\ldots,k$.

Empirical estimate $\hat{\nu} = \frac{1}{k}\sum_{j=1}^{k} \delta_{X_j}$ satisfies

$$\mathbb{E}\left[\hat{\nu}\right] = \pi P.$$

This motivates estimation of π as

$$\hat{\pi} = \operatorname*{argmin}_{\pi'} \left\{\rho(\hat{\nu}, \pi' P) : \mathbb{E}_{\pi'}\left[c(\theta)\right] \leq 1\right\}.$$

Connection to minimum distance estimator

General set up:

- Parameter space Θ.
- Distribution family $\{P_\theta : \theta \in \Theta\}$ with distance measure ρ.
- $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{\theta_j}$.
- Want to analyze

$$R(k) = \inf_{\hat{\pi}} \sup_{\theta_1,\ldots,\theta_k} \mathbb{E}\left[d(\hat{\pi},\pi)\right]$$

under some cost constraint $\frac{1}{k}\sum_{j=1}^{k} c(\theta_j) \le 1$.

Data: $X_j \sim P_{\theta_j}$ independently for $j = 1,\ldots,k$.

Empirical estimate $\hat{\nu} = \frac{1}{k}\sum_{j=1}^{k} \delta_{X_j}$ satisfies

$$\mathbb{E}\left[\hat{\nu}\right] = \pi P.$$

This motivates estimation of π as

$$\hat{\pi} = \operatorname*{argmin}_{\pi'} \left\{\rho(\hat{\nu}, \pi' P) : \mathbb{E}_{\pi'}\left[c(\theta)\right] \le 1\right\}.$$

Connection to linear programming

Suppose additionally we have concentration of $\hat{\nu}$ around πP

$$\mathbb{P}\left[\rho(\pi P, \hat{\nu}) > t_k\right] \leq \epsilon_k$$

for some $t_k, \epsilon_k \to 0$. Then we can argue to get results of type

$$R(k) \lesssim \delta(2t_k)$$

where δ is the linear program given by

$$\delta(t) = \sup\left\{d(\pi, \pi') : \rho(\pi P, \pi' P) \leq t, \mathbb{E}_\pi\left[c(\theta)\right] \leq 1, \mathbb{E}_{\pi'}\left[c(\theta)\right] \leq 1\right\}.$$

Choice of total variation: Choosing $\rho(\cdot,\cdot) = \|\cdot - \cdot\|_{\mathrm{TV}}$ gives us

$$\delta(1/k) \lesssim R(k) \lesssim \delta(t_k).$$

When $\delta(1/k) \asymp \delta(t_k)$ we get the rate.

Suppose additionally we have concentration of $\hat{\nu}$ around πP

$$\mathbb{P}\left[\rho(\pi P, \hat{\nu}) > t_k\right] \leq \epsilon_k$$

for some $t_k, \epsilon_k \to 0$. Then we can argue to get results of type

$$R(k) \lesssim \delta(2t_k)$$

where δ is the linear program given by

$$\delta(t) = \sup\left\{d(\pi, \pi') : \rho(\pi P, \pi' P) \leq t, \mathbb{E}_\pi\left[c(\theta)\right] \leq 1, \mathbb{E}_{\pi'}\left[c(\theta)\right] \leq 1\right\}.$$

Choice of total variation: Choosing $\rho(\cdot,\cdot) = \|\cdot - \cdot\|_{\mathrm{TV}}$ gives us

$$\delta(1/k) \lesssim R(k) \lesssim \delta(t_k).$$

When $\delta(1/k) \asymp \delta(t_k)$ we get the rate.

Linear programming: BSM and the total variation case

- For Bernoulli sampling model the family is given by Markov kernel P

$$P_{im} = \binom{i}{m} p^m (1-p)^{i-m}, \quad i, m \geq 0.$$

- Note that profile has mean less than 1. Define linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup \left\{ \|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \leq t; \mathbb{E}_\pi[\theta], \mathbb{E}_{\pi'}[\theta] \leq 1 \right\}.$$

- $\delta_{\mathrm{TV}}(t)$ is a modulus of continuity type linear program that appears in previous work of statistical estimation.

Theorem

There exist absolute constants C_1, C_2, d_0 such that for all $k \geq d_0$

$$\frac{1}{72}\delta_{\mathrm{TV}}\left(\frac{1}{6k}\right) - \frac{C_2}{\sqrt{k}} \leq R(k) \leq 2\delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right), \quad (1)$$

where the upper bound is attained by the minimum distance estimator

$$\hat{\pi} = \underset{\pi'}{\mathrm{argmin}} \left\{ \|\hat{\nu} - \pi' P\|_{\mathrm{TV}} \right\}.$$

Linear programming: BSM and the total variation case

- For Bernoulli sampling model the family is given by Markov kernel P

$$P_{im} = \binom{i}{m} p^m (1-p)^{i-m}, \quad i, m \geq 0.$$

- Note that profile has mean less than 1. Define linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup \left\{ \|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \leq t; \mathbb{E}_{\pi}[\theta], \mathbb{E}_{\pi'}[\theta] \leq 1 \right\}.$$

- $\delta_{\mathrm{TV}}(t)$ is a modulus of continuity type linear program that appears in previous work of statistical estimation.

Theorem

There exist absolute constants C_1, C_2, d_0 such that for all $k \geq d_0$

$$\frac{1}{72}\delta_{\mathrm{TV}}\left(\frac{1}{6k}\right) - \frac{C_2}{\sqrt{k}} \leq R(k) \leq 2\delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right), \tag{1}$$

where the upper bound is attained by the minimum distance estimator

$$\hat{\pi} = \operatorname*{argmin}_{\pi'} \left\{ \|\hat{\nu} - \pi' P\|_{\mathrm{TV}} \right\}.$$

Linear programming: BSM and the total variation case

- For Bernoulli sampling model the family is given by Markov kernel P

$$P_{im} = \binom{i}{m} p^m (1-p)^{i-m}, \quad i, m \geq 0.$$

- Note that profile has mean less than 1. Define linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup \left\{ \|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \leq t; \mathbb{E}_\pi[\theta], \mathbb{E}_{\pi'}[\theta] \leq 1 \right\}.$$

- $\delta_{\mathrm{TV}}(t)$ is a modulus of continuity type linear program that appears in previous work of statistical estimation.

Theorem

There exist absolute constants C_1, C_2, d_0 such that for all $k \geq d_0$

$$\frac{1}{72}\delta_{\mathrm{TV}}\left(\frac{1}{6k}\right) - \frac{C_2}{\sqrt{k}} \leq R(k) \leq 2\delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right), \tag{1}$$

where the upper bound is attained by the minimum distance estimator

$$\hat{\pi} = \operatorname*{argmin}_{\pi'} \left\{ \|\hat{\nu} - \pi' P\|_{\mathrm{TV}} \right\}.$$

Linear programming: BSM and the total variation case

- For Bernoulli sampling model the family is given by Markov kernel P

$$P_{im} = \binom{i}{m} p^m (1-p)^{i-m}, \quad i, m \geq 0.$$

- Note that profile has mean less than 1. Define linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup \left\{ \|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \leq t; \mathbb{E}_{\pi}[\theta], \mathbb{E}_{\pi'}[\theta] \leq 1 \right\}.$$

- $\delta_{\mathrm{TV}}(t)$ is a modulus of continuity type linear program that appears in previous work of statistical estimation.

Theorem

There exist absolute constants C_1, C_2, d_0 such that for all $k \geq d_0$

$$\frac{1}{72} \delta_{\mathrm{TV}}\left(\frac{1}{6k}\right) - \frac{C_2}{\sqrt{k}} \leq R(k) \leq 2\delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right), \tag{1}$$

where the upper bound is attained by the minimum distance estimator

$$\hat{\pi} = \underset{\pi'}{\operatorname{argmin}} \left\{ \|\hat{\nu} - \pi' P\|_{\mathrm{TV}} \right\}.$$

Analyzing $\delta_{\rm TV}$

Lemma

1. *There exists absolute constant* $C_3 > 0$ *such that for all* p, t *we have*

$$\delta_{\rm TV}(t) \leq \min\left\{\frac{C_3}{p\log(1/t)}, 1\right\}. \tag{2}$$

2. *There exist absolute constants* $C_4, t_0 > 0$ *such that for any* $p \in (0,1)$, $t \leq t_0$,

$$\delta_{\rm TV}(t) \geq \min\left\{\frac{\bar{p}}{p}, \sqrt{\log(1/t)}\right\}\frac{C_4}{\log(1/t)}.$$

In view of previous theorem this gives us the rate.

Sketch of proof
($\delta_{\mathrm{TV}}(t)$ bounds)

- To bound $\delta_{\mathrm{TV}}(t)$ we first relate it to another linear program $\delta_*(t)$ in terms of generating functions.
- For any $g(z) = \sum_{n=0}^{\infty} a_n z^n$ define its $\|\cdot\|_A$ norm as

$$\|g\|_A = \sum_{n=0}^{\infty} |a_n|.$$

- Define the new linear program

$$\begin{aligned}\delta_*(t) &\triangleq \sup_{\Delta} \left\{ \sum_{m=0}^{\infty} |\Delta_m| : \|\Delta P\|_1 \le t, \sum_{m=0}^{\infty} m|\Delta_m| \le 1 \right\}. \\ &= \sup_f \left\{ \|f\|_A : \|f_p\|_A \le t, \|f'\|_A \le 1 \right\}\end{aligned}$$

where $f_p(z) = f(\bar{p} + pz)$ and the sup is over all analytic functions f.

Lemma

For all $t \in [0,1]$ we have

$$\frac{1}{2}(\delta_*(t) - t) \le \delta_{\mathrm{TV}}(t) \le \delta_*(t).$$

We bound $\delta_*(t)$ using complex analytic techniques.

- To bound $\delta_{\mathrm{TV}}(t)$ we first relate it to another linear program $\delta_*(t)$ in terms of generating functions.
- For any $g(z) = \sum_{n=0}^{\infty} a_n z^n$ define its $\|\cdot\|_A$ norm as

$$\|g\|_A = \sum_{n=0}^{\infty} |a_n|.$$

- Define the new linear program

$$\delta_*(t) \triangleq \sup_{\Delta} \left\{ \sum_{m=0}^{\infty} |\Delta_m| : \|\Delta P\|_1 \le t, \sum_{m=0}^{\infty} m|\Delta_m| \le 1 \right\}.$$
$$= \sup_{f} \left\{ \|f\|_A : \|f_p\|_A \le t, \|f'\|_A \le 1 \right\}$$

where $f_p(z) = f(\bar{p} + pz)$ and the sup is over all analytic functions f.

Lemma

For all $t \in [0,1]$ we have

$$\frac{1}{2}(\delta_*(t) - t) \le \delta_{\mathrm{TV}}(t) \le \delta_*(t).$$

We bound $\delta_*(t)$ using complex analytic techniques.

- To bound $\delta_{\mathrm{TV}}(t)$ we first relate it to another linear program $\delta_*(t)$ in terms of generating functions.
- For any $g(z) = \sum_{n=0}^{\infty} a_n z^n$ define its $\|\cdot\|_A$ norm as

$$\|g\|_A = \sum_{n=0}^{\infty} |a_n|.$$

- Define the new linear program

$$\begin{aligned}\delta_*(t) &\triangleq \sup_{\Delta}\left\{\sum_{m=0}^{\infty}|\Delta_m| : \|\Delta P\|_1 \le t, \sum_{m=0}^{\infty} m|\Delta_m| \le 1\right\}.\\ &= \sup_{f}\left\{\|f\|_A : \|f_p\|_A \le t, \|f'\|_A \le 1\right\}\end{aligned}$$

where $f_p(z) = f(\bar{p} + pz)$ and the sup is over all analytic functions f.

Lemma

For all $t \in [0,1]$ we have

$$\frac{1}{2}(\delta_*(t) - t) \le \delta_{\mathrm{TV}}(t) \le \delta_*(t).$$

We bound $\delta_*(t)$ using complex analytic techniques.

- To bound $\delta_{\mathrm{TV}}(t)$ we first relate it to another linear program $\delta_*(t)$ in terms of generating functions.
- For any $g(z) = \sum_{n=0}^{\infty} a_n z^n$ define its $\|\cdot\|_A$ norm as

$$\|g\|_A = \sum_{n=0}^{\infty} |a_n|.$$

- Define the new linear program

$$\begin{aligned}\delta_*(t) &\triangleq \sup_{\Delta} \left\{ \sum_{m=0}^{\infty} |\Delta_m| : \|\Delta P\|_1 \le t, \sum_{m=0}^{\infty} m|\Delta_m| \le 1 \right\}. \\ &= \sup_{f} \left\{ \|f\|_A : \|f_p\|_A \le t, \|f'\|_A \le 1 \right\}\end{aligned}$$

where $f_p(z) = f(\bar{p} + pz)$ and the sup is over all analytic functions f.

Lemma

For all $t \in [0,1]$ we have

$$\frac{1}{2}(\delta_*(t) - t) \le \delta_{\mathrm{TV}}(t) \le \delta_*(t).$$

We bound $\delta_*(t)$ using complex analytic techniques.

- To bound $\delta_{\mathrm{TV}}(t)$ we first relate it to another linear program $\delta_*(t)$ in terms of generating functions.
- For any $g(z) = \sum_{n=0}^{\infty} a_n z^n$ define its $\|\cdot\|_A$ norm as

$$\|g\|_A = \sum_{n=0}^{\infty} |a_n|.$$

- Define the new linear program

$$\begin{aligned}\delta_*(t) &\triangleq \sup_{\Delta} \left\{ \sum_{m=0}^{\infty} |\Delta_m| : \|\Delta P\|_1 \le t, \sum_{m=0}^{\infty} m|\Delta_m| \le 1 \right\}. \\ &= \sup_{f} \left\{ \|f\|_A : \|f_p\|_A \le t, \|f'\|_A \le 1 \right\}\end{aligned}$$

where $f_p(z) = f(\bar{p} + pz)$ and the sup is over all analytic functions f.

Lemma

For all $t \in [0, 1]$ we have

$$\frac{1}{2}(\delta_*(t) - t) \le \delta_{\mathrm{TV}}(t) \le \delta_*(t).$$

We bound $\delta_*(t)$ using complex analytic techniques.

Upper bound on $\delta_*(t)$

- Note that the objective function in $\delta_*(t)$ can be written as

$$\|f\|_A = \sum_{m=0}^{\infty} \frac{\left|f^{(m)}(0)\right|}{m!}.$$

- The contribution from $\sum_{m \geq \log(1/t)} \frac{|f^{(m)}(0)|}{m!}$ is at most $\frac{C_p}{\log(1/t)}$ from the derivative constraint.
- For each other terms we use the LP's

$$\delta_m(t) = \sup_f \left\{ \frac{|f^{(m)}(0)|}{m!} : \|f_p\|_A \leq t, \|f'\|_A \leq 1 \right\}$$

  to bound $\delta_*(t)$ as

$$\delta_*(t) \leq \sum_{m=0}^{\log(1/t)} \delta_m(t) + \frac{C_p}{\log(1/t)}.$$

- We then show that each of $\delta_m(t)$ is negligible by using Hadamard three line theorem.

Upper bound on $\delta_*(t)$

- Note that the objective function in $\delta_*(t)$ can be written as

$$\|f\|_A = \sum_{m=0}^{\infty} \frac{\left|f^{(m)}(0)\right|}{m!}.$$

- The contribution from $\sum_{m \geq \log(1/t)} \frac{|f^{(m)}(0)|}{m!}$ is at most $\frac{C_p}{\log(1/t)}$ from the derivative constraint.
- For each other terms we use the LP's

$$\delta_m(t) = \sup_f \left\{ \frac{|f^{(m)}(0)|}{m!} : \|f_p\|_A \leq t, \|f'\|_A \leq 1 \right\}$$

to bound $\delta_*(t)$ as

$$\delta_*(t) \leq \sum_{m=0}^{\log(1/t)} \delta_m(t) + \frac{C_p}{\log(1/t)}.$$

- We then show that each of $\delta_m(t)$ is negligible by using Hadamard three line theorem.

Upper bound on $\delta_*(t)$

- Note that the objective function in $\delta_*(t)$ can be written as

$$\|f\|_A = \sum_{m=0}^{\infty} \frac{\left|f^{(m)}(0)\right|}{m!}.$$

- The contribution from $\sum_{m \geq \log(1/t)} \frac{|f^{(m)}(0)|}{m!}$ is at most $\frac{C_p}{\log(1/t)}$ from the derivative constraint.
- For each other terms we use the LP's

$$\delta_m(t) = \sup_f \left\{ \frac{|f^{(m)}(0)|}{m!} : \|f_p\|_A \leq t, \|f'\|_A \leq 1 \right\}$$

to bound $\delta_*(t)$ as

$$\delta_*(t) \leq \sum_{m=0}^{\log(1/t)} \delta_m(t) + \frac{C_p}{\log(1/t)}.$$

- We then show that each of $\delta_m(t)$ is negligible by using Hadamard three line theorem.

Upper bound on $\delta_*(t)$

- Note that the objective function in $\delta_*(t)$ can be written as

$$\|f\|_A = \sum_{m=0}^{\infty} \frac{\left|f^{(m)}(0)\right|}{m!}.$$

- The contribution from $\sum_{m\geq\log(1/t)} \frac{|f^{(m)}(0)|}{m!}$ is at most $\frac{C_p}{\log(1/t)}$ from the derivative constraint.
- For each other terms we use the LP's

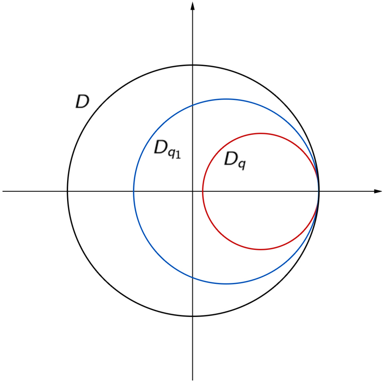
$$\delta_m(t) = \sup_f \left\{ \frac{|f^{(m)}(0)|}{m!} : \|f_p\|_A \leq t, \|f'\|_A \leq 1 \right\}$$

to bound $\delta_*(t)$ as

$$\delta_*(t) \leq \sum_{m=0}^{\log(1/t)} \delta_m(t) + \frac{C_p}{\log(1/t)}.$$

- We then show that each of $\delta_m(t)$ is negligible by using Hadamard three line theorem.

Hadamard's three line theorem



- For any analytic function f define its $\|\cdot\|_{H^\infty(C)}$ norm over set C

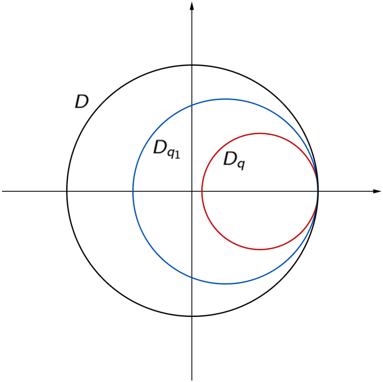
$$\|f\|_{H^\infty(C)} = \sup_{z\in C} |f(z)|.$$

Denote by D the unit disc on $\mathbb{C}$ and let $D_p = \bar{p} + pD$.

- Consider $0 < q < q_1 < 1$
- Then Hadamard's three line theorem says that

$$\|f\|_{H^\infty(D_{q_1})} \leq \|f\|_{H^\infty(D)}^{1-\frac{q\bar{q}_1}{\bar{q}q_1}} \|f\|_{H^\infty(D_q)}^{\frac{q\bar{q}_1}{\bar{q}q_1}}.$$

Hadamard's three line theorem



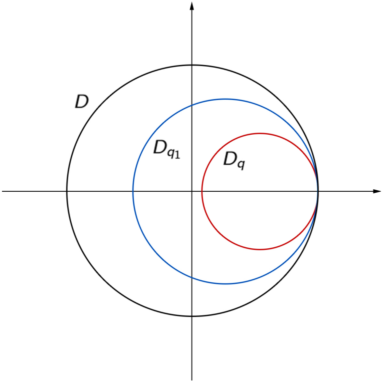
- For any analytic function f define its $\|\cdot\|_{H^\infty(C)}$ norm over set C

$$\|f\|_{H^\infty(C)} = \sup_{z\in C} |f(z)|.$$

  Denote by D the unit disc on $\mathbb{C}$ and let $D_p = \bar{p} + pD$.

- Consider $0 < q < q_1 < 1$
- Then Hadamard's three line theorem says that

$$\|f\|_{H^\infty(D_{q_1})} \leq \|f\|_{H^\infty(D)}^{1-\frac{q\bar{q}_1}{\bar{q}q_1}} \|f\|_{H^\infty(D_q)}^{\frac{q\bar{q}_1}{\bar{q}q_1}}.$$



- For any analytic function f define its $\|\cdot\|_{H^\infty(C)}$ norm over set C

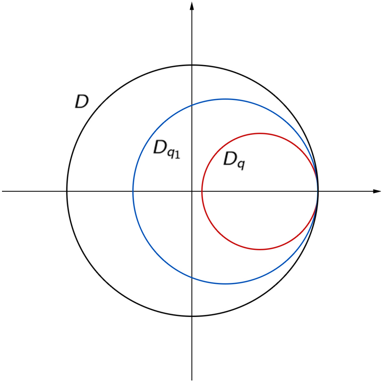
$$\|f\|_{H^\infty(C)} = \sup_{z\in C} |f(z)|.$$

Denote by D the unit disc on $\mathbb{C}$ and let $D_p = \bar{p} + pD$.

- Consider $0 < q < q_1 < 1$
- Then Hadamard's three line theorem says that

$$\|f\|_{H^\infty(D_{q_1})} \leq \|f\|_{H^\infty(D)}^{1-\frac{q\bar{q}_1}{\bar{q}q_1}} \|f\|_{H^\infty(D_q)}^{\frac{q\bar{q}_1}{\bar{q}q_1}}.$$

Hadamard's three line theorem



- For any analytic function f define its $\|\cdot\|_{H^\infty(C)}$ norm over set C

$$\|f\|_{H^\infty(C)} = \sup_{z\in C} |f(z)|.$$

  Denote by D the unit disc on $\mathbb{C}$ and let $D_p = \bar{p} + pD$.
- Consider $0 < q < q_1 < 1$
- Then Hadamard's three line theorem says that

$$\|f\|_{H^\infty(D_{q_1})} \le \|f\|_{H^\infty(D)}^{1-\frac{q\bar{q}_1}{\bar{q}q_1}} \|f\|_{H^\infty(D_q)}^{\frac{q\bar{q}_1}{\bar{q}q_1}}.$$

Upper bounding $\delta_m(t)$ using Hadamard's theorem

- Using ordering between A norm and $H^\infty(D)$ norm, the constraints on $\delta_m(t)$, and Cauchy integral formula we get bounds on $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$.
- Cauchy's integral formula also implies

$$\frac{|f^{(m)}|}{m!} \le 2^m \|f\|_{H^\infty(D_{1/2})}.$$

- For $p < \frac{1}{2} < 1$ we get $D_p \subset D_{1/2} \subset D_1$. So we can bound $\|f\|_{H^\infty(D_{1/2})}$ in terms of $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$ by Hadamard's theorem.
- For $p \ge \frac{1}{2}$ as $D_{1/2} \subset D_p$ we get $\|f\|_{H^\infty(D_{1/2})} \le \|f\|_{H^\infty(D_p)} \le t$.
- Adding up all the bounds we get $O_p\left(\frac{1}{\log(1/t)}\right)$ bound on $\sum_{m=0}^{\log(1/t)} \delta_m(t)$.

Upper bounding $\delta_m(t)$ using Hadamard's theorem

- Using ordering between A norm and $H^\infty(D)$ norm, the constraints on $\delta_m(t)$, and Cauchy integral formula we get bounds on $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$.
- Cauchy's integral formula also implies

$$\frac{|f^{(m)}|}{m!} \leq 2^m \|f\|_{H^\infty(D_{1/2})}.$$

- For $p < \frac{1}{2} < 1$ we get $D_p \subset D_{1/2} \subset D_1$. So we can bound $\|f\|_{H^\infty(D_{1/2})}$ in terms of $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$ by Hadamard's theorem.
- For $p \geq \frac{1}{2}$ as $D_{1/2} \subset D_p$ we get $\|f\|_{H^\infty(D_{1/2})} \leq \|f\|_{H^\infty(D_p)} \leq t$.
- Adding up all the bounds we get $O_p\left(\frac{1}{\log(1/t)}\right)$ bound on $\sum_{m=0}^{\log(1/t)} \delta_m(t)$.

Upper bounding $\delta_m(t)$ using Hadamard's theorem

- Using ordering between A norm and $H^\infty(D)$ norm, the constraints on $\delta_m(t)$, and Cauchy integral formula we get bounds on $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$.
- Cauchy's integral formula also implies

$$\frac{|f^{(m)}|}{m!} \leq 2^m \|f\|_{H^\infty(D_{1/2})}.$$

- For $p < \frac{1}{2} < 1$ we get $D_p \subset D_{1/2} \subset D_1$. So we can bound $\|f\|_{H^\infty(D_{1/2})}$ in terms of $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$ by Hadamard's theorem.
- For $p \geq \frac{1}{2}$ as $D_{1/2} \subset D_p$ we get $\|f\|_{H^\infty(D_{1/2})} \leq \|f\|_{H^\infty(D_p)} \leq t$.
- Adding up all the bounds we get $O_p\left(\frac{1}{\log(1/t)}\right)$ bound on $\sum_{m=0}^{\log(1/t)} \delta_m(t)$.

Upper bounding $\delta_m(t)$ using Hadamard's theorem

- Using ordering between A norm and $H^\infty(D)$ norm, the constraints on $\delta_m(t)$, and Cauchy integral formula we get bounds on $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$.
- Cauchy's integral formula also implies

$$\frac{|f^{(m)}|}{m!} \leq 2^m \|f\|_{H^\infty(D_{1/2})}.$$

- For $p < \frac{1}{2} < 1$ we get $D_p \subset D_{1/2} \subset D_1$. So we can bound $\|f\|_{H^\infty(D_{1/2})}$ in terms of $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$ by Hadamard's theorem.
- For $p \geq \frac{1}{2}$ as $D_{1/2} \subset D_p$ we get $\|f\|_{H^\infty(D_{1/2})} \leq \|f\|_{H^\infty(D_p)} \leq t$.
- Adding up all the bounds we get $O_p\left(\frac{1}{\log(1/t)}\right)$ bound on $\sum_{m=0}^{\log(1/t)} \delta_m(t)$.

Upper bounding $\delta_m(t)$ using Hadamard's theorem

- Using ordering between A norm and $H^\infty(D)$ norm, the constraints on $\delta_m(t)$, and Cauchy integral formula we get bounds on $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$.
- Cauchy's integral formula also implies

$$\frac{|f^{(m)}|}{m!} \le 2^m \|f\|_{H^\infty(D_{1/2})}.$$

- For $p < \frac{1}{2} < 1$ we get $D_p \subset D_{1/2} \subset D_1$. So we can bound $\|f\|_{H^\infty(D_{1/2})}$ in terms of $\|f\|_{H^\infty(D)}$ and $\|f\|_{H^\infty(D_p)}$ by Hadamard's theorem.
- For $p \ge \frac{1}{2}$ as $D_{1/2} \subset D_p$ we get $\|f\|_{H^\infty(D_{1/2})} \le \|f\|_{H^\infty(D_p)} \le t$.
- Adding up all the bounds we get $O_p\left(\frac{1}{\log(1/t)}\right)$ bound on $\sum_{m=0}^{\log(1/t)} \delta_m(t)$.

Lower bound on $\delta_*(t)$

- Using $\|\cdot\|_{H^\infty(D)} \le \|\cdot\|_A$ we relax the constraints and objective functions of $\delta_*(t)$ to get related linear program

$$\delta_{H^\infty}(t) = \sup_f \left\{ \|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \le t, \|f'\|_{H^\infty(D)} \le 1 \right\}$$

- We have $\delta_{H^\infty}(t) = \Theta_p\left(\frac{1}{\log(1/t)}\right)$ and is achieved by the function

$$f(z) = \frac{c_p}{\log(1/t)}(1-z)^2 t^{\frac{p}{p}\frac{1+z}{1-z}}.$$

- We modify $f(z)$ by further dilation $z \to \alpha z$ to get feasible solution to $\delta_*(t)$.
- The coefficients of the modified version can be related to the Laguerre polynomial. The sum of the coefficients gives us the desired logarithmic lower bound. The proof relies on properties of Laguerre polynomials.

Lower bound on $\delta_*(t)$

- Using $\|\cdot\|_{H^\infty(D)} \le \|\cdot\|_A$ we relax the constraints and objective functions of $\delta_*(t)$ to get related linear program

$$\delta_{H^\infty}(t) = \sup_f \left\{ \|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \le t, \|f'\|_{H^\infty(D)} \le 1 \right\}$$

- We have $\delta_{H^\infty}(t) = \Theta_p\left(\frac{1}{\log(1/t)}\right)$ and is achieved by the function

$$f(z) = \frac{c_p}{\log(1/t)}(1-z)^2 t^{\frac{p}{\bar{p}}\frac{1+z}{1-z}}.$$

- We modify $f(z)$ by further dilation $z \to \alpha z$ to get feasible solution to $\delta_*(t)$.
- The coefficients of the modified version can be related to the Laguerre polynomial. The sum of the coefficients gives us the desired logarithmic lower bound. The proof relies on properties of Laguerre polynomials.

Lower bound on $\delta_*(t)$

- Using $\|\cdot\|_{H^\infty(D)} \le \|\cdot\|_A$ we relax the constraints and objective functions of $\delta_*(t)$ to get related linear program

$$\delta_{H^\infty}(t) = \sup_f \left\{ \|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \le t, \|f'\|_{H^\infty(D)} \le 1 \right\}$$

- We have $\delta_{H^\infty}(t) = \Theta_p\left(\frac{1}{\log(1/t)}\right)$ and is achieved by the function

$$f(z) = \frac{c_p}{\log(1/t)}(1-z)^2 t^{\frac{p}{\bar{p}}\frac{1+z}{1-z}}.$$

- We modify $f(z)$ by further dilation $z \to \alpha z$ to get feasible solution to $\delta_*(t)$.
- The coefficients of the modified version can be related to the Laguerre polynomial. The sum of the coefficients gives us the desired logarithmic lower bound. The proof relies on properties of Laguerre polynomials.

Lower bound on $\delta_*(t)$

- Using $\|\cdot\|_{H^\infty(D)} \le \|\cdot\|_A$ we relax the constraints and objective functions of $\delta_*(t)$ to get related linear program

$$\delta_{H^\infty}(t) = \sup_f \left\{ \|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \le t, \|f'\|_{H^\infty(D)} \le 1 \right\}$$

- We have $\delta_{H^\infty}(t) = \Theta_p\left(\frac{1}{\log(1/t)}\right)$ and is achieved by the function

$$f(z) = \frac{c_p}{\log(1/t)}(1-z)^2 t^{\frac{p}{p}\frac{1+z}{1-z}}.$$

- We modify $f(z)$ by further dilation $z \to \alpha z$ to get feasible solution to $\delta_*(t)$.
- The coefficients of the modified version can be related to the Laguerre polynomial. The sum of the coefficients gives us the desired logarithmic lower bound. The proof relies on properties of Laguerre polynomials.

Summary

- The profile of population gives us information about many important label invariant properties.
- In small sample regime of $p = \omega\left(\frac{1}{\log k}\right)$ we can consistently estimate the profile in total variation distance..
- When $p = \Theta(1)$ the optimal rate is $\Theta\left(\frac{1}{\log k}\right)$.
- The estimator which achieves optimal rate is based of minimum distance type and can be computed in polynomial time.
- We device a single infinite dimensional linear program that characterizes the estimator and also proves its minimax optimality. We solve the LP using complex analytic techniques.

Bunge, J. and Fitzpatrick, M. (1993).
Estimating the number of species: a review.
Journal of the American Statistical Association, 88(421):364–373.

Charikar, M., Chaudhuri, S., Motwani, R., and Narasayya, V. (2000).
Towards estimation error guarantees for distinct values.
In *Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, pages 268–279. ACM.

Han, Y., Jiao, J., and Weissman, T. (2018).
Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under wasserstein distance.
In *Proc. 2018 Conference On Learning Theory (COLT)*, pages 3189–3221.

Orlitsky, A., Santhanam, N., Viswanathan, K., and Zhang, J. (2005).
Convergence of profile based estimators.
In *Proc. 2005 IEEE Int. Symp. Inf. Theory (ISIT)*, pages 1843–1847. IEEE.

Raskhodnikova, S., Ron, D., Shpilka, A., and Smith, A. (2009).
Strong lower bounds for approximating distribution support size and the distinct elements problem.
SIAM Journal on Computing, 39(3):813–842.

Robbins, H. (1951).
Asymptotically subminimax solutions of compound statistical decision problems.
In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California.

Robbins, H. (1956).
An empirical bayes approach to statistics.
In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

Valiant, G. (2012).
Algorithmic Approaches to Statistical Questions.
PhD thesis, EECS Department, University of California, Berkeley.

Valiant, G. and Valiant, P. (2011).
Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs.
In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 685–694.

Valiant, G. and Valiant, P. (2013).
Estimating the unseen: improved estimators for entropy and other properties.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 2157–2165.

Wu, Y. and Yang, P. (2018).
Sample complexity of the distinct element problem.
Mathematical Statistics and Learning, 1(1):37–72.

Proof of upper bound: Connection of risk and $\delta_{\mathrm{TV}}(t)$

- $\hat{\nu}$ concentrates around $\mathbb{E}[\hat{\nu}] = \pi P$

$$\mathbb{E}[\|\hat{\nu} - \nu\|_{\mathrm{TV}}] = \left(\sqrt{\frac{C_1 \log k}{k}}\right).$$

- Using McDiarmid's inequality we get

$$\mathbb{P}[|\|\hat{\nu} - \nu\|_{\mathrm{TV}} - \mathbb{E}[\|\hat{\nu} - \nu\|_{\mathrm{TV}}]| \geq \epsilon] \leq exp\left(C_0 k \epsilon^2\right)$$

which implies

$$\mathbb{P}\left[\|\hat{\nu} - \nu\|_{\mathrm{TV}} \geq \sqrt{\frac{C_1 \log k}{k}}\right] \leq \frac{1}{k}.$$

Proof of upper bound: Connection of risk and $\delta_{\mathrm{TV}}(t)$

- $\hat{\nu}$ concentrates around $\mathbb{E}[\hat{\nu}] = \pi P$

$$\mathbb{E}\left[\|\hat{\nu} - \nu\|_{\mathrm{TV}}\right] = \left(\sqrt{\frac{C_1 \log k}{k}}\right).$$

- Using McDiarmid's inequality we get

$$\mathbb{P}\left[\left|\|\hat{\nu} - \nu\|_{\mathrm{TV}} - \mathbb{E}\left[\|\hat{\nu} - \nu\|_{\mathrm{TV}}\right]\right| \geq \epsilon\right] \leq exp\left(C_0 k \epsilon^2\right)$$

which implies

$$\mathbb{P}\left[\|\hat{\nu} - \nu\|_{\mathrm{TV}} \geq \sqrt{\frac{C_1 \log k}{k}}\right] \leq \frac{1}{k}.$$

Proof of upper bound: Connection of risk and $\delta_{\mathrm{TV}}(t)$

- Linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup\left\{\|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \le t; \mathbb{E}_{\pi}[\theta], \mathbb{E}_{\pi'}[\theta] \le 1\right\}.$$

- The minimum distance estimator $\hat{\pi} = \mathrm{argmin}_{\pi}\left\{\|\hat{\nu} - \pi P\|_{\mathrm{TV}} : \mathbb{E}_{\pi}[\theta] \le 1\right\}$ satisfies

$$\|\hat{\pi} P - \pi P\|_{\mathrm{TV}} \le \|\pi P - \hat{\nu}\|_{\mathrm{TV}} + \|\hat{\pi} P - \hat{\nu}\|_{\mathrm{TV}} \le 2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}.$$

- This implies

$$\mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] + \frac{1}{k}$$

and hence

$$R(k) = \inf_{\hat{\pi}} \sup_{\pi} \mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] \le \delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right) + \frac{1}{k}.$$

Proof of upper bound: Connection of risk and $\delta_{\mathrm{TV}}(t)$

- Linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup\left\{\|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \le t; \mathbb{E}_\pi[\theta], \mathbb{E}_{\pi'}[\theta] \le 1\right\}.$$

- The minimum distance estimator $\hat{\pi} = \mathrm{argmin}_\pi \{\|\hat{\nu} - \pi P\|_{\mathrm{TV}} : \mathbb{E}_\pi[\theta] \le 1\}$ satisfies

$$\|\hat{\pi} P - \pi P\|_{\mathrm{TV}} \le \|\pi P - \hat{\nu}\|_{\mathrm{TV}} + \|\hat{\pi} P - \hat{\nu}\|_{\mathrm{TV}} \le 2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}.$$

- This implies

$$\mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] + \frac{1}{k}$$

and hence

$$R(k) = \inf_{\hat{\pi}} \sup_{\pi} \mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] \le \delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right) + \frac{1}{k}.$$

- Linear program

$$\delta_{\mathrm{TV}}(t) \triangleq \sup\left\{\|\pi - \pi'\|_{\mathrm{TV}} : \|\pi P - \pi' P\|_{\mathrm{TV}} \le t; \mathbb{E}_{\pi}[\theta], \mathbb{E}_{\pi'}[\theta] \le 1\right\}.$$

- The minimum distance estimator $\hat{\pi} = \mathrm{argmin}_{\pi} \{\|\hat{\nu} - \pi P\|_{\mathrm{TV}} : \mathbb{E}_{\pi}[\theta] \le 1\}$ satisfies

$$\|\hat{\pi} P - \pi P\|_{\mathrm{TV}} \le \|\pi P - \hat{\nu}\|_{\mathrm{TV}} + \|\hat{\pi} P - \hat{\nu}\|_{\mathrm{TV}} \le 2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}.$$

- This implies

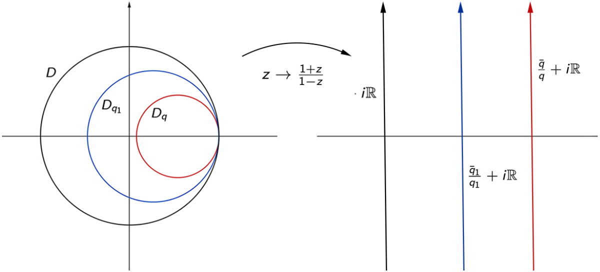
$$\mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] + \frac{1}{k}$$

  and hence

$$R(k) = \inf_{\hat{\pi}} \sup_{\pi} \mathbb{E}\left[\|\hat{\pi} - \pi\|_{\mathrm{TV}}\right] \le \mathbb{E}\left[\delta_{\mathrm{TV}}\left(2\|\pi P - \hat{\nu}\|_{\mathrm{TV}}\right)\right] \le \delta_{\mathrm{TV}}\left(\sqrt{\frac{C_1 \log k}{k}}\right) + \frac{1}{k}.$$

Proof of lower bound: Solution to auxiliary program $\delta_{H^\infty(t)}$

- $\delta_{H^\infty}(t) = \sup_f \left\{ \|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \le t, \|f'\|_{H^\infty(D)} \le 1 \right\}$
- We use the transform $w : z \to \frac{1+z}{1-z}$



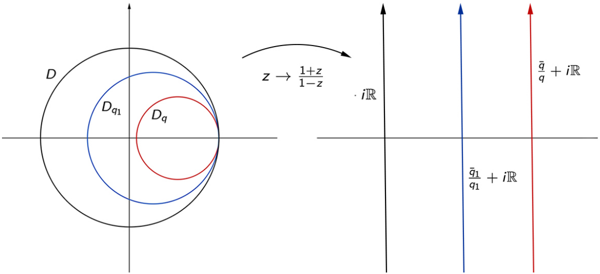
and re-parameterize $f(z) = g(w)$.

- Then we have $g'(w) = \frac{2}{(1+w)^2} f'\left(\frac{w-1}{w+1}\right)$ and using constraints get the bound

$$\|g'\|_{H^\infty(\Re=\epsilon)} \le 2C_p t^{\min\left\{\frac{\epsilon p}{2\bar{p}}, 1\right\}}, \quad \epsilon \in \left(0, \frac{\bar{p}}{p}\right).$$

Proof of lower bound: Solution to auxiliary program $\delta_{H^\infty(t)}$

- $\delta_{H^\infty}(t) = \sup_f \left\{\|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \leq t, \|f'\|_{H^\infty(D)} \leq 1\right\}$
- We use the transform $w : z \to \frac{1+z}{1-z}$



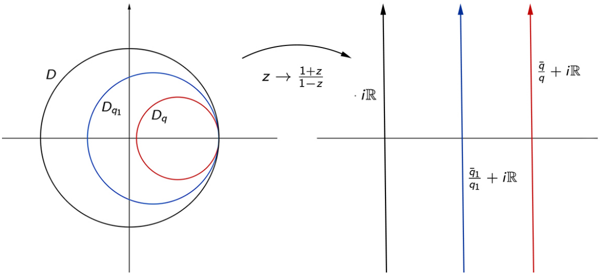
and re-parameterize $f(z) = g(w)$.

- Then we have $g'(w) = \frac{2}{(1+w)^2} f'\left(\frac{w-1}{w+1}\right)$ and using constraints get the bound

$$\|g'\|_{H^\infty(\Re=\epsilon)} \leq 2C_p t^{\min\left\{\frac{\epsilon p}{2\bar{p}},1\right\}}, \quad \epsilon \in \left(0, \frac{\bar{p}}{p}\right).$$

Proof of lower bound: Solution to auxiliary program $\delta_{H^\infty(t)}$

- $\delta_{H^\infty}(t) = \sup_f \left\{\|f\|_{H^\infty}(D) : \|f_p\|_{H^\infty(D)} \leq t, \|f'\|_{H^\infty(D)} \leq 1\right\}$
- We use the transform $w : z \to \frac{1+z}{1-z}$



and re-parameterize $f(z) = g(w)$.

- Then we have $g'(w) = \frac{2}{(1+w)^2} f'\left(\frac{w-1}{w+1}\right)$ and using constraints get the bound

$$\|g'\|_{H^\infty(\Re=\epsilon)} \leq 2C_p t^{\min\left\{\frac{\epsilon p}{2\bar{p}},1\right\}}, \quad \epsilon \in \left(0, \frac{\bar{p}}{p}\right).$$

Proof of lower bound: Solution to auxiliary program $\delta_{H^\infty(t)}$

- Using this we integrate the derivatives to get bound

$$\left|g(iw) - g\left(iw + \frac{\bar{p}}{p}\right)\right| \le C_p \int_0^{\frac{\bar{p}}{p}} t^{\frac{\epsilon p}{2\bar{p}}} \le \frac{C_p}{\log\left(\frac{1}{t}\right)}.$$

- As $\|g\|_{H^\infty(\Re=\bar{p}/p)} \le t$ we get $\|g\|_{H^\infty(\Re=0)} = \|f\|_{H^\infty}(D) \le C_p \frac{1}{\log(1/t)}$.
- As exponential function saturates the Hadamard three line theorem, the guess is to choose exponential function for lower bound. The choice

$$f(z) = \frac{c_p}{\log(1/t)}(1-z)^2 t^{\frac{p}{\bar{p}}\frac{1+z}{1-z}}$$

comes from modifications to satisfy the constraints.

Solving $\delta_*(t)$

- For $\beta > 0, 0 < \alpha < 1$ to be chosen later define

$$h(z) = \exp\left(-\beta \frac{1+\alpha z}{1-\alpha z}\right) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right).$$

- We use the Laguerre polynomial relation

$$h(z) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right) = e^{-\beta}\sum_{n=0}^{\infty} \alpha^n L_n^{-1}(2\beta).$$

- Denote $\Delta_m = e^{-\beta}\alpha^n L_n^{-1}(2\beta)$ and show that for sufficiently large β

$$|\Delta_m| + |\Delta_{m+1}| \geq \alpha^{3\beta/2}\beta^{-1/2}.$$

- We bound $\|h\|_A$ from below by $\sum_{\beta \leq m \leq 3\beta/2} |\Delta_m|$.
- For the choice $\beta = \max\left\{\frac{4p}{\bar{p}}\log(1/t), \sqrt{\frac{\log(1/t)}{p}}\right\}$ and $\alpha = \frac{1}{\beta}$ we get desired lower bound.

Solving $\delta_*(t)$

- For $\beta > 0, 0 < \alpha < 1$ to be chosen later define

$$h(z) = \exp\left(-\beta \frac{1+\alpha z}{1-\alpha z}\right) = \exp(-\beta) \exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right).$$

- We use the Laguerre polynomial relation

$$h(z) = \exp(-\beta) \exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right) = e^{-\beta} \sum_{n=0}^{\infty} \alpha^n L_n^{-1}(2\beta).$$

- Denote $\Delta_m = e^{-\beta} \alpha^n L_n^{-1}(2\beta)$ and show that for sufficiently large β

$$|\Delta_m| + |\Delta_{m+1}| \geq \alpha^{3\beta/2} \beta^{-1/2}.$$

- We bound $\|h\|_A$ from below by $\sum_{\beta \leq m \leq 3\beta/2} |\Delta_m|$.
- For the choice $\beta = \max\left\{\frac{4p}{\bar{p}} \log(1/t), \sqrt{\frac{\log(1/t)}{p}}\right\}$ and $\alpha = \frac{1}{\beta}$ we get desired lower bound.

Solving $\delta_*(t)$

- For $\beta > 0, 0 < \alpha < 1$ to be chosen later define

$$h(z) = \exp\left(-\beta \frac{1+\alpha z}{1-\alpha z}\right) = \exp(-\beta) \exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right).$$

- We use the Laguerre polynomial relation

$$h(z) = \exp(-\beta) \exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right) = e^{-\beta} \sum_{n=0}^{\infty} \alpha^n L_n^{-1}(2\beta).$$

- Denote $\Delta_m = e^{-\beta} \alpha^n L_n^{-1}(2\beta)$ and show that for sufficiently large β

$$|\Delta_m| + |\Delta_{m+1}| \geq \alpha^{3\beta/2} \beta^{-1/2}.$$

- We bound $\|h\|_A$ from below by $\sum_{\beta \leq m \leq 3\beta/2} |\Delta_m|$.
- For the choice $\beta = \max\left\{\frac{4p}{p} \log(1/t), \sqrt{\frac{\log(1/t)}{p}}\right\}$ and $\alpha = \frac{1}{\beta}$ we get desired lower bound.

Solving $\delta_*(t)$

- For $\beta > 0, 0 < \alpha < 1$ to be chosen later define

$$h(z) = \exp\left(-\beta \frac{1+\alpha z}{1-\alpha z}\right) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right).$$

- We use the Laguerre polynomial relation

$$h(z) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right) = e^{-\beta}\sum_{n=0}^{\infty} \alpha^n L_n^{-1}(2\beta).$$

- Denote $\Delta_m = e^{-\beta}\alpha^n L_n^{-1}(2\beta)$ and show that for sufficiently large β

$$|\Delta_m| + |\Delta_{m+1}| \geq \alpha^{3\beta/2}\beta^{-1/2}.$$

- We bound $\|h\|_A$ from below by $\sum_{\beta \leq m \leq 3\beta/2} |\Delta_m|$.
- For the choice $\beta = \max\left\{\frac{4p}{p}\log(1/t), \sqrt{\frac{\log(1/t)}{p}}\right\}$ and $\alpha = \frac{1}{\beta}$ we get desired lower bound.

Solving $\delta_*(t)$

- For $\beta > 0, 0 < \alpha < 1$ to be chosen later define

$$h(z) = \exp\left(-\beta \frac{1+\alpha z}{1-\alpha z}\right) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right).$$

- We use the Laguerre polynomial relation

$$h(z) = \exp(-\beta)\exp\left(-2\beta \frac{\alpha z}{1-\alpha z}\right) = e^{-\beta}\sum_{n=0}^{\infty} \alpha^n L_n^{-1}(2\beta).$$

- Denote $\Delta_m = e^{-\beta}\alpha^n L_n^{-1}(2\beta)$ and show that for sufficiently large β

$$|\Delta_m| + |\Delta_{m+1}| \geq \alpha^{3\beta/2}\beta^{-1/2}.$$

- We bound $\|h\|_A$ from below by $\sum_{\beta \leq m \leq 3\beta/2} |\Delta_m|$.
- For the choice $\beta = \max\left\{\frac{4p}{\bar{p}}\log(1/t), \sqrt{\frac{\log(1/t)}{p}}\right\}$ and $\alpha = \frac{1}{\beta}$ we get desired lower bound.