# The distance approach to approximate combinatorial counting

A. Barvinok, A. Samorodnitsky

presented by Tomáš Gavenčiak

**Problem:** Given a subset $\mathcal{F}$ of $2^X$, $|X| = n$, estimate $|\mathcal{F}|$.

**Variations:** How is $\mathcal{F}$ given? Various kinds of oracles.

**Examples:** Perfect matchings ($X = E_G$), linearly indep. subsets ($X$ matroid elements), heterochromatic spanning trees ($X = E_G$).

## Different estimation methods

### Simple Monte Carlo

Sample a point $x \in 2^X$, see if $x \in \mathcal{F}$. Can estimate only $|\mathcal{F}| \sim \alpha 2^n$, $0 \leq \alpha \leq 1$.

### Markov chain Monte Carlo

Define $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_k = \mathcal{F}$ with $|\mathcal{F}_{i+1}| = O(poly(|\mathcal{F}_i|))$ and known $|\mathcal{F}_0|$. *Product estimators:* estimate $|\mathcal{F}_{i+1}|/|\mathcal{F}_i|$ by Monte Carlo uniform sampling from $\mathcal{F}_{i+1}$. Needs fast-convergent Markov chains for every $\mathcal{F}_i$. Can be very accurate.

### Distance approach

Embed $\mathcal{F}$ as $F \subseteq C_n$, sample a point $x \in C_n$ and compute $\operatorname{dist}(x, F)$, estimate $\Delta(F)$ expected (average) distance to $F$. Can estimate $2^{\alpha_1 n} \leq |\mathcal{F}| \leq 2^{\alpha_2 n}$ for some $0 \leq \alpha_1 < \alpha_2 \leq 1$.

## Problem oracles

**Optimization oracle.** Given weights $\gamma_x, x \in X$, return $\min_{Y \in \mathcal{F}} \sum_{x \in Y} \gamma_x$.

**Hamming distance oracle.** Given $a \in C_n$, return $\min \operatorname{dist}(a, F)$.

For given penalties $d_i : \{0,1\} \times \{0,1\} \to \mathbb{Z}$, define $d(a,b) = \sum_i d_i(a_i, b_i)$.

**Weighted distance oracle.** Given $a \in C_n$ and penalties $d_i$, return $\min d(a, F)$.

**Simple embedding.** Assume $X = \{1 \ldots n\}$. Map $Y \in \mathcal{F}$ to its characteristic vector. To solve dist. oracle for $a$ and penalties $d_i$, set $\gamma_i = d_i(a_i, 1) - d_i(a_i, 0)$.

**Economical embedding.** Assume $X = \{1 \ldots n\}$, $X = X_1 \cup \cdots \cup X_k$ (not disjoint) such that $\forall Y \in \mathcal{F} \forall i : |X_i \cap Y| = 1$. Map $Y \in \mathcal{F}$ to $(y_1 \ldots y_k), y_i = \#_i(X_i \cap Y)$.

## Estimating average distance

**Average distance.** $\Delta(A) = 1/2^n \sum_{x \in C_n} \mathrm{dist}(x, A) = \mathbb{E}_x[\mathrm{dist}(x, A)]$.

**Algorithm estimating $\Delta(A)$.** Given $\epsilon$, sample $\lceil 3n/2\epsilon^2 \rceil$ points $x_i$, return average of $\mathrm{dist}(x_i, A)$.

THEOREM 3.6. Algorithm returns $\alpha$ with $|\Delta(A) - \alpha| \leq \epsilon$ with 0.9 probability.

## Estimating size of $|F|$

**Entropy function.** $H(x) = x \log_2(1/x) + (1-x)\log_2(1/(1-x))$.

THEOREM 3.9. Let $\rho = 1/2 - \Delta(A)/n$, then

$$1 - H\left(\frac{1}{2} - \rho\right) \leq \frac{\log_2 |A|}{n} \leq H(\rho).$$

COROLLARY 3.11. There are $c_1, c_2 > 0$ such that for $\rho = 1/2 - \Delta(A)/n$

$$c_1 \rho^2 \leq \frac{\log_2 |A|}{n} \leq c_2 \rho \log_2 \frac{1}{\rho},$$

and this holds for any $c_1 < 2, c_2 > 1$ for $\rho > 0$ sufficiently small.

## Randomized average distance

**Distance for selected coordinates $l$.** $d_l(a, b) = \sum_i l_i |a_i - b_i|$.

**Randomized average distance.** $\Delta(A, p) = \mathbb{E}_{x \in C_n} \mathbb{E}_{l \in \mathrm{Binom}^n(1,p)} d_l(x, A)$.

**Algorithm estimating $\Delta(A, p)$.** Given $p$ and $\epsilon$, sample $\lceil 3n/\epsilon^2 \rceil$ points $x_i \in C_n$ together with $l_i \in \{0, 1\}^n$. Return average of $d_{l_i}(x_i, A)$.

THEOREM 4.4. Algorithm returns $\alpha$ with $|\Delta(A, p) - \alpha| \leq \epsilon$ with 0.9 probability.

THEOREM 4.5. Let $\rho = p/2 - \Delta(A, p)/n$, then $\rho^2/p \leq \ln(|A|)/n$ and with $\rho \leq 1/2$ and some additional assumptions, $\log_2(|A|)/n \leq H(2\rho)$

COROLLARY 4.6. For any $c_3 < 1/\ln(2)$ and $c_4 > 2$, there is $\delta > 0$ such that for any $A$ with $\ln(|A|)/n \leq \delta$ there is some $p$ such that for $\rho = p/2 - \Delta(A, p)/n$,

$$c_3 \rho^2 \log_2 \frac{1}{\rho} \leq \frac{\log_2(|A|)}{n} \leq c_4 \rho \log_2 \frac{1}{\rho}.$$