Lecture 17:

Poisson Binomial Distributions

(PBD's)

+

Local Computation Algorithms

(see slides)

Recall: Poisson Distribution

$$\Pr[X=k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\lambda = E[X] = Var[X]$$

if $X_i \sim Pois(\lambda_i)$

then $Y = \sum X_i \sim Pois(\sum \lambda_i)$

(also converse: if $X_1 + X_2 \sim Pois(\lambda_1+\lambda_2)$

then $X_1 \sim Pois(\lambda_1)$
$X_2 \sim Pois(\lambda_2)$)

if n large + p small enough, $\underset{\lambda=np}{\text{Poisson}} \sim \underset{n,p}{\text{Binomial}}$

$$d_{TV}(Poiss(\lambda_1), Pois(\lambda_2)) \leq \frac{1}{2}\left(e^{|\lambda_1-\lambda_2|} - e^{-|\lambda_1-\lambda_2|}\right)$$

Poisson's Binomial Distribution (PBD)

$PBD(p_1 \cdots p_n) : X = \sum_{i=1}^{n} X_i$ $\quad X_i$ independent, $\{0,1\}$ r.v.'s

$E[X_i] = p_i$ $\quad$ not necessarily identically distributed

examples 1) all p_i's $= 1/2$ $\quad X \sim$ Binomial distribution

2) $p_1 = 1/2 \quad p_2 = 1 \quad p_3 = p_4 = \cdots = p_n = 0$

$\Pr[X=0] = 0$

$\Pr[X=1] = 1/2$

$\Pr[X=2] = 1/2$ $\qquad X \sim 1 + \$$

$\Pr[X=3,4,\ldots n] = 0$

PBD vs Poisson$(\sum_{i=1}^{n} p_i)$: $\leq 2 \sum_{i=1}^{n} p_i^2$ [LeCam] (1)

$\leq 2 \frac{\sum p_i^2}{\sum p_i}$ (2)

so Poisson approximation is pretty good, but not arbitrarily good (ie. you can't say you want ϵ accuracy & get that close)

Translated Poisson Distribution:

$TP(\mu, \sigma^2)$: $Y = \lfloor \mu - \sigma^2 \rfloor + Z$

$\uparrow$ $\sim$ Poisson $(\sigma^2 + \underbrace{\{\mu - \sigma^2\}}_{\text{fractional part of } \mu - \sigma^2})$

PBD vs TPD:

Thm $d_{TV}(PBD(p_1 \cdots p_n), TP(\mu, \sigma^2)) \leq \frac{\sqrt{\sum p_i^3 (1-p_i)} + 2}{\sum p_i (1-p_i)}$

still not there

Different Approach:

<u>Thm</u> every PBD is unimodal over $[n]$

$\Rightarrow$ use "Birge" learning to learn with

$O(\frac{1}{\varepsilon^3} \log n)$ samples

Question: do we need dependence on n?

Structure Thm:

<u>Thm</u> PBD "looks like" (to within ε L_1 error) either:

(i) ($\frac{1}{\varepsilon}$-sparse) support of PBD is almost all (as fctn of ε) on $\underbrace{\text{interval of length}}$ $O(1/\varepsilon^3)$

i.e. all but $O(\frac{1}{\varepsilon^3})$ variables have p_i close to 0 or 1 + can be viewed as "fixed"

so we have PBD on $O(\frac{1}{\varepsilon^3})$ variables that can "move"

$\Rightarrow$ tiny effective support size, so can <u>learn</u> each probability of elements in support.

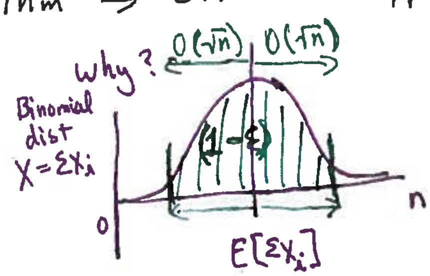
(ii) ($\frac{1}{\varepsilon}$-heavy Binomial) PBD looks like a binomial on $\underbrace{\text{large}}_{> poly(\frac{1}{\varepsilon})}$ number of iid vars.

Use of structure Thm:

learning: Thm $\Rightarrow$ (i) small cover $\left(\frac{1}{\varepsilon}\right)^{O(\log^2(\frac{1}{\varepsilon}))}$ $\Rightarrow$ $\frac{1}{\varepsilon^2}\log^3\frac{1}{\varepsilon}$ learning

(ii) PBD close to Translated binomial

testing: Thm $\Rightarrow$ effective support of distribution is $O(n^{1/2})$ $\Rightarrow$ $O(n^{1/4})$ samples needed



maximized in case 2. But Binomial puts almost all of its weight on $\sqrt{n}$ places in the middle.

More detailed structure: for $X = \sum X_i$, let $k \leftarrow O(1/\varepsilon)$

Thm $\exists\, Y_1, \dots, Y_n$ s.t.

1. $\| \sum X_i - \sum Y_i \|_1 \le O(1/k)$

2. One of following holds:

(1) (k-sparse) $\exists\, \ell \le k^3$ s.t. $\forall\, i \le \ell$ $E[Y_i] \in \{ \frac{1}{k^2}, \frac{2}{k^2}, \dots, \frac{k^2-1}{k^2} \}$ + $\forall\, i > \ell$ $E[Y_i] \in \{0, 1\}$

so $0 \le \sum Y_i \le k^3$

cover size: $(k^3+1)(k^2)^{k^3} \cdot (n+1)$

choices of ℓ; choices of $E[Y_i]$ for $i \le \ell$; # choices for $\sum_{i > \ell} Y_i$

OR

(2) ((n, k)-Binomial form) $\exists\, \ell \in [n]$ + $q \in \{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n} \}$ st. $\forall\, i \le \ell$ $E[Y_i] = q$ + $\forall\, i > \ell$ $E[Y_i] = 0$

$\leftarrow$ actually, $q \ge \frac{1}{k}$, $q \le \frac{k-1}{k}$

$\Rightarrow$ cover size for this part $\le n^2$

also $\underbrace{\ell q}_{E[\sum Y_i]} \ge k^2$ + $\ell q(1-q) \ge k^2 - k - 1$

} note $\ell \ge k^2$; if q small $\sim \frac{1}{k}$ then $\ell \ge k^3$

Cover = union of (1) + (2) covers