Poisson's Binomial Distribution (PBD)

\[ \text{PBD} \left( p_1, \ldots, p_n \right) \overset{d}{=} X = \sum_{i=1}^{n} X_i \quad \text{where} \quad X_i \text{ independent, } \mathbb{E}(X_i) = p_i, \text{ not necessarily identically distributed} \]

**Examples**

1) all \( p_i = \frac{1}{2} \) \( \implies X \sim \text{Binomial distribution} \)

2) \( p_1 = \frac{1}{2}, \quad p_2 = 1 \) \( \implies P_3 = P_4 = \ldots = p_n = 0 \)

\[
\begin{align*}
\Pr[\sum X = 0] &= 0 \\
\Pr[X = 1] &= \frac{1}{2} \\
\Pr[\sum X = 2] &= \frac{1}{2} \\
\Pr[X = 3, 4, \ldots, n] &= 0
\end{align*}
\]

**PBD vs Poisson**

\[
\text{PBD vs Poisson} \left( \sum_{i=1}^{n} p_i \right) \leq 2 \sum_{i=1}^{n} p_i^2 \tag{1}
\]

\[
\leq 2 \sum_{i=1}^{n} \frac{p_i^2}{p_i} \tag{2}
\]

**Translated Poisson Distribution:**

\[ TP \left( \mu, \sigma^2 \right) : \quad Y = \left( \mu - \sigma^2 \right) + Z \]

\[ \sim \text{Poisson} \left( \mu^2 + \frac{\left( \mu - \sigma^2 \right)^2}{2} \right) \]

**PBD vs TPD:**

\[ \text{Thm} \quad d_{TV} \left( \text{PBD} \left( p_1, \ldots, p_n \right), TP \left( \mu, \sigma^2 \right) \right) \leq \frac{\sum p_i^3 (1-p_i) + 2}{\sum p_i (1-p_i)} \]

\[ \text{still not there} \]
Structure Thm:

Thm: PBD "looks like" (to within ε L, error) either:

(i) \( (\frac{1}{\varepsilon}\)-sparse) support of PBD is almost all (as \( \varepsilon \)

on interval of length \( O(\frac{1}{\varepsilon^3}) \)

i.e. all but \( O(\varepsilon^3) \) variables have \( p_i \) close to 0 or 1

it 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 learn each probability of

elements in support.

(ii) \( (\frac{1}{\varepsilon}\)-heavy Binomial) PBD looks like a binomial

on large number of iid vars.

\( \Rightarrow \) \( poly(\frac{1}{\varepsilon}) \)

Use of structure Thm:

learning: Thm \( \Rightarrow \) small cover

testing: Thm \( \Rightarrow \) effective support of distribution is \( O(n^{\varepsilon/4}) \)

\( \Rightarrow \) \( O(n^{\varepsilon/4}) \) samples needed

maximized in case 2.

But Binomial puts almost all

of its weight on \( \frac{n}{2} \) places in the middle.
More detailed structure: for $X = \sum_{X}^{h}$, let $k = O(1/k)$

**Thm.** \exists \ Y_{1}, ..., Y_{n} \ st.

1. $\|X - \sum_{Y}^{h}\|_{1} = O(1/k)$

2. One of following holds:

   (i) (k-sparse) \exists \ l \in \mathbb{K}^{3} \ st. \ \forall \ i \leq l

   \[
   E[Y_{i}] \in \{ \frac{1}{K^2}, \frac{2}{K^2}, ..., \frac{K^2-1}{K^2} \}
   \]

   + \ \forall \ i > l \ E[Y_{i}] \in [0, 1]

   \[
   0 \leq \sum_{Y}^{h} \leq K^{3}
   \]

   or

   (ii) ((n, k)-Bernoulli form) \exists \ l \in [n] \ \forall \ q \in \left[ \frac{1}{K^2}, \frac{2}{K^2}, ..., \frac{K^2-1}{K^2} \right]

   s.t. \ \forall \ i \leq l \ E[Y_{i}] = q

   + \ \forall \ i > l \ E[Y_{i}] = 0

   also \ \forall \ g \geq k^{2} + \ln(1-q) \geq K^{2} - k - 1

   \[
   E[Y_{i}]
   \]

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

---

Coversize:

\[
\binom{K^{3}+1}{K}
\]

\[
\binom{K^{3}+1}{K} \cdot \binom{n+1}{K}
\]

\[
\binom{K^{3}+1}{K} \cdot \binom{n+1}{K}
\]

Choice of $Y_{i}$ for $i \leq l$

Choice of $Y_{i}$ for $i > l$

Choice of $q$

\[
\sum_{Y}^{h} = 0
\]

Cover size

For this part $\leq n^{2}$

\[
\text{Cover = union of (1) + (2) covers}
\]

\[
\text{Cover = union of (1) + (2) covers}
\]

\[
\text{Cover = union of (1) + (2) covers}
\]
Proof Outline \[ \text{let } k = 0 \left( \frac{1}{k} \right) \]

Step 1: eliminate rare with expectation in \((0, \frac{1}{k})\) or \((\frac{k-1}{k}, 1)\) w/o much change

for all \(i \) st. \[ p_i \leq 0, \frac{1}{k} \]

take the sum to figure out how many \((p_i)\)'s with prob \( \frac{1}{k} \) would have similar sum

ie. \[ \sum \frac{1}{k} \approx \sum p_i \]

set 1st \( r \) such \( i \) to \( \frac{1}{k} \) rest to 0

use \[ d_{TV} \left( \sum_{i \leq \lambda_1} \chi_i, \text{Poiss} \left( \sum_{i \leq \lambda_1} p_i \right) \right) \leq \frac{1}{k} \sum_{i \leq \lambda_1} p_i \leq \frac{1}{k} \sum_{i \leq \lambda_1} p_i = \frac{1}{k} \]

\[ d_{TV} \left( \text{Poiss} \left( \sum_{i \leq \lambda_1} p_i \right), \text{Poiss} \left( \sum_{i \leq \lambda_1} \chi_i \right) \right) \leq \frac{1}{2} \left( e^{\lambda_1 - \lambda_2} - e^{-\lambda_1 - \lambda_2} \right) \]

\[ \chi_1 \]

\[ \chi_2 = \frac{1}{2} \left( e^{\frac{1}{k}} - e^{-\frac{1}{k}} \right) \]

\[ d_{TV} \left( \text{Poiss} \left( \sum_{i \leq \lambda_1} p_i \right), \sum_{i \leq \lambda_1} \chi_i \right) \leq \frac{1}{k} \leq \frac{1.5}{k} \]

\[ \Delta \Rightarrow \text{dist} \leq \frac{3.5}{k} \]

get total \( \frac{7}{k} \) dist when do heavy els.
Step 2:

k-sparse case:

- weaker proof:
  
  \[ \text{Use } d_{TV}(\mathcal{E}X, \mathcal{E}Y) \leq \mathbb{E}_x d_{TV}(X, Y) \text{ when } X, Y \text{ indep} \]

  if round each \( p_i \) to nearest multiple of \( \frac{1}{K^4} \)

  \[ \text{get } d_{TV}(\mathcal{E}X, \mathcal{E}X') \leq K^3 \cdot \frac{1}{K^4} = \frac{1}{K} \]

  they do something smarter!

  idea: something like step 4 (+ relate to Binomial)

  use a different bound on similarity to Binomial

  use different grouping - K groups

  each contributes \( O(\frac{1}{K^2}) \) error

  total \( O(\frac{1}{K}) \) error

if not k-sparse:

approx by Binomial distribution

\[ B(m', q) \quad \text{versus fixed to 1} \]

\[ m' = \frac{(\mathbb{E}p_i + t)^2}{(\mathbb{E}p_i^2 + t)} \quad q = \frac{p}{n} \]

Can show via bound on similarity to translated Poisson dist.

that approx is good.

\[ \square \]

Further improvements:

Can weed more out of cover by using Roos's Thm:

\[ \text{if } \sum p_i^t = \sum q_i^t \quad \forall \quad t = 1, 2 \ldots (\log n) \]

\[ \Rightarrow \| p - q \|_1 \leq \varepsilon \]

\[ \varepsilon \text{ not quite as stated here unless all } p_{ij} \leq \frac{1}{2} \]

(Otherwise need to separate 2 st. \( p_{ij} > \frac{1}{2} \)

from \( n \) st. \( p_{ij} > \frac{1}{2} \))