6.889 Lecture 12

Amartya Shankha Biswas

April 6, 2017

1 Poisson Binomial Distribution

A Poisson binomial distribution is constructed as a generalization of the binomial distribution. Consider n independent Bernoulli random variables $\{X_1, X_1, \cdots, X_n\}$. Each of these has a different bias, i.e. $\mathbb{E}[X_i] = p_i$. The PBD distribution is defined as the sum of all these variables.

$$PBD(p_1, p_1, \cdots, p_n) = \sum_{i=1}^{n} X_i$$

1.1 Poisson Approximation

We can crudely approximate the PBD distribution by a Poisson distribution with $\lambda = \sum p_i$ [1]. Formally, for $X = \sum_{i=1}^{n} X_i$, where $\mathbb{E}[X_i] = p_i$, we have

$$\left\| X - Poi\left(\sum_{i=1}^{n} p_i\right) \right\|_1 \leq 4 \sum_{i=1}^{n} p_i^2$$

Note that this distance is only small when the probabilities are small. In the worst case, if some $p_i = \Theta(1)$, then this bound will also be $\Theta(1)$. However, let's say that there are k elements, each with probability ϵ. Then, the L_1 distance gets bounded by $k \cdot \epsilon^2 = \epsilon$.

1.1.1 Translated Poisson Distribution

We can obtain a better approximation to the PBD distribution by using a *translated Poisson distribution*. A random variable Y is distributed as the translated Poisson distribution $TP(\mu, \sigma^2)$ iff. we can write it as $Y = \lfloor \mu - \sigma^2 \rfloor + Z$. Here Z is distributed as $Poisson(\sigma^2 + \{\mu - \sigma^2\})$, where $\{x\} \equiv x - \lfloor x \rfloor$.

This gives us the following theorem from [2].

Theorem 1. *Given a PBD* $X = \sum_{i=1}^{n} X_i$ *with* $\mathbb{E}[X_i] = p_i$*, define* $\mu = \sum p_i$ *and* $\sigma^2 = \sum p_i(1-p_i)$*. Then*

$$\left\| \sum_{i=1}^{n} X_i - TP(\mu, \sigma^2) \right\|_1 \leq 2 \cdot \frac{\sqrt{\sum_{i=1}^{n} p_i^3(1-p_i)} + 2}{\sum_{i=1}^{n} p_i^(1-p_i)}$$

2 Structure Theorem

The main theorem here concerns the structure of PBDs. Specifically, we will construct S_ϵ, an ϵ cover for the set S_n of all PBDs with support size n. This theorem tells us that every PBD is either close to a PBD whose support is sparse ($\mathcal{O}(1/\epsilon^3)$), or is close to a translated "heavy" Binomial distribution.

Note that we can effectively ignore the contribution of variables that have $p_i = 0$. The translation is caused by variables that have bias that is exactly 1. If there are k such variables, we can ignore all of them and simply subtract k from our random variable.

Theorem 2 (Structure Theorem)**.** *We construct a cover* S_ϵ *for the set of PBDs* S_n*. Let's define* $k = \mathcal{O}(1/\epsilon)$*. The theorem states that for every* $\{X_i\} \in S_n$*, there exists* $\{Y_i\}$*, such that*

1. $\left\| \sum X_i - \sum Y_i \right\|_1 \leq \epsilon$
2. *One of the following holds*
    - *(k-Sparse) –* $\exists l \leq k^3$ *such that* $\forall i \leq l$, $\mathbb{E}[Y_i] \in \left\{ \frac{1}{k^2}, \frac{2}{k^2}, \cdots, \frac{k^2-1}{k^2} \right\}$ *and* $\forall i > l$, $\mathbb{E}[Y_i] \in \{0, 1\}$.
    - *(Heavy Binomial) – There exists some* $l \in [n]$*, and some* $q \in \{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n}{n}\}$*, such that* $\forall i \leq l$, $\mathbb{E}[Y_i] = q$ *and* $\forall i > l$, $\mathbb{E}[Y_i] = 0$*. Additionally, we have* $lq \geq k^2$ *and* $lq(1-q) \geq k^2 - k - 1$.

2.1 Learning

We will now use the Structure Theorem and the Cover theorem for testing hypotheses to make an algorithm that can learn PBDs. Essentially, our algorithm will output one member of the cover set S_ϵ. So, this distribution will be close to our PBD and we can find it using $|S_\epsilon|$ samples.

All we need is to estimate the size of the cover.

- For the sparse case, there are k^3 possible values of l and each of the l important random variables can take k^2possible vales. This gives us $k^3 \cdot (k^2)^{k^3}$ possibilities. Additionally, every other variable can be either 0 or 1. This leads to at most $n+1$ more possibilities (denoting the number of variables that have bias zero). So, the total possible number of possible distributions is $\mathcal{O}((n+1) \cdot k^3 \cdot (k^2)^{k^3})$.
- For the heavy case, there are n possible values for q and at most n possible values for l. This gives us a total of $\mathcal{O}(n^2)$ distributions.

The size of the cover is simply the sum of these two numbers. So, we find that $\log |S_\epsilon| = \mathcal{O}(\log n \cdot k^3 \cdot \log k)$. This means that the number of samples required to learn the distribution is just $\mathcal{O}(\log n \cdot poly(1/\epsilon))$.

2.2 Testing

Testing is also easy to do, given the structure theorem. First, we consider the sparse case. Here, the effective support size is tiny i.e. there are only l possible values. For the heavy Binomial case, we have a binomial distribution on $l \leq n$ elements. Now, we know that almost the entire probability mass of this Binomial is concentrated on the middle $\mathcal{O}(\sqrt{n})$ elements. So, testing against this distribution will only require $\mathcal{O}(n^{1/4})$ samples.

3 Proving the Structure Theorem

We will sketch an outline of the proof of the Structure Theorem. First let us define a trivial bias as any bias that is either zero or one i.e. non-trivial biases actually have some randomness. This will proceed in two steps.

- Step 1 – Eliminate all the non-trivial variables that have expectation in $(0, 1/k)$ or $(1-1/k, 1)$ without changing the L_1 distance too much. Formally, we will construct $\{Z_i\}$ such that $\|\sum X_i - \sum Z_i\|_1 \leq \mathcal{O}(k)$, and for all non-trivial i, $1/k < \mathbb{E}[Z_i] < 1 - 1/k$.
- Step 2 – Construct the final variables $\{Y_i\}$ which satisfy the second property in the structure theorem, such that $\|\sum Y_i - \sum Z_i\|_1 \leq \mathcal{O}(k)$.

3.1 Eliminating Outliers

In Step 1, we will define a new set of variables $\{Z_i\}$ where $\mathbb{E}[Z_i] = p'_i$ and for each non-trivial bias, $1/k < p'_i < 1 - 1/k$. For all the biases where $p_i \in (0, 1/k) \cup (1 - 1/k, 1)$, we simply set $p'_i = p_i$.

Now consider the set L of biases where $0 < p_i < 1/k$. We want to construct p'_i such that $|\sum_L p_i - \sum_L p_i| < 1/k$. This can be dome by setting r values of p'_i to

$1/k$ and the remaining to zero, where $r = \lfloor k \cdot \sum_L p_i \rfloor$. Similarly, we can set the p'_i values for all the non-trivial biases that are larger than $1 - 1/k$ by rounding to either one or $1 - 1/k$.

Now, we need to bound the L_1 distance between $\sum Z_i$ and $\sum X_i$. Fist, we consider the distance between each of these and the corresponding Poissonizations.

$$\left\|\sum X_i - Poisson(\sum p_i)\right\|_1 \leq 2\sum_{i=1}^{n} p_i^2 \leq 2 \cdot \frac{1}{k} \cdot \sum_{i=1}^{n} p_i = \frac{2}{k}$$

$$\left\|\sum Z_i - Poisson(\sum p'_i)\right\|_1 \leq 2\sum_{i=1}^{n} {p'_i}^2 \leq 2 \cdot \frac{1}{k} \cdot \sum_{i=1}^{n} p'_i = \frac{2}{k}$$

Finally, we bound the distance between $Poisson(\lambda_1)$ and $Poisson(\lambda_2)$, where $\lambda_1 = \sum p_i$ and $\lambda_2 = \sum p'_i$.

$$\left\|Poisson(\sum p_i) - Poisson(\sum p'_i)\right\|_1 \leq e^{|\lambda_1 - \lambda_2|} - e^{-|\lambda_1-\lambda_2|} \leq e^{\frac{1}{k}} - e^{-\frac{1}{k}} \leq \frac{3}{k}$$

So, we can now use the triangle inequality to show that

$$\left\|\sum X_i - \sum Z_i\right\|_1 \leq 2\epsilon + 3\epsilon + 2\epsilon = 7\epsilon$$

This concludes the first step of our construction.

3.2 Constructing the Cover

For the k-sparse case, we will simply round each of the biases p'_i. In the original proof, the rounding is performed to the nearest multiple of $\frac{1}{k^2}$. However, to simplify our analysis, we will instead round to the nearest multiple of $\frac{1}{k^4}$ i.e. $q_i = \lfloor k^2 p'_i \rfloor \cdot \frac{1}{k^2}$. So, we have at most k^3 variables with non-trivial bias, and each of the biases is changed by at most $\frac{1}{k^2}$ meaning $|p'_i - q_i| \leq \frac{1}{k^2}$. The total L_1 distance is then bounded by $k^3 \cdot \frac{1}{k^2} = \frac{1}{k}$.

For the non-sparse case, we will approximate the distribution by a Binomial – $B(m', q)$, such that

$$m' = \frac{(\sum p'_i + t)^2}{(\sum {p'_i}^2 + t)}$$

where t is the number of variables whose bias is exactly 1.

To find the bias q, we find l^* such that $\frac{\sum p'_i + t}{m'} \in \left[\frac{l^*-1}{n}, \frac{l^*}{n}\right]$.

Finally, we let $q = l^*/n$.

References

[1] Lucien Le Cam et al. An approximation theorem for the poisson binomial distribution. *Pacific J. Math*, 10(4):1181–1197, 1960.

[2] Adrian Röllin. Translated poisson approximation using exchangeable pair couplings. *The Annals of Applied Probability*, pages 1596–1614, 2007.