STATS 310A: Theory of Probability I **Autumn 2016/17**

Lecture 14: November 9

Lecturer: Persi Diaconis *Scribes: Kenneth Tay*

14.1 Stein's Method for Poisson Approximation

Recall: for $\lambda > 0$, the Poisson distribution with parameter λ is given by

$$P_\lambda(j) = \frac{e^{-\lambda}\lambda^j}{j!}, \qquad 0 \leq j < \infty.$$

- If $X \sim \text{Poisson}(\lambda)$, then $\mathbb{E}X = \lambda$, $\text{Var } X = \lambda$, and

$$\mathbb{E}[X(X-1)\dots(X-r+1)] = \lambda^r \text{ for all } 0 < r < \infty.$$

- If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ with X and Y independent, then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Consider the following set-up:

- $\{X_i\}_{i \in I}$ a collection of 0/1-valued random variables (I some index set),
- $W := \displaystyle\sum_{i \in I} X_i$,
- $p_i := \mathbb{E}X_i = P\{X_i = 1\}$,
- $\lambda := \sum_{i \in I} p_i = \mathbb{E}W$.

The **Poisson heuristic** can be stated (imprecisely) as follows:

> If $\{X_i\}_{i \in I}$ are "not too dependent", $|I|$ is "large", $\{p_i\}$ are "small", λ is "a number", then W is approximately distributed as a Poisson distribution with parameter λ, i.e. $W \dot\sim \text{Poisson}(\lambda)$.

The rest of this lecture will try to make this statement more precise.

14.1.1 Total Variation Distance

Definition 14.1 *Let P and Q be probabilities on some measurable space $(\Omega, \mathcal{F})$. We define the **total variation distance** between P and Q to be*

$$\|P - Q\|_{TV} := \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.$$

Here are some properties of the total variation distance:

1. $\|P-Q\|_{TV} = \dfrac{1}{2} \sup\limits_{\|f\|_\infty \le 1} |\mathbb{E}_P(f) - \mathbb{E}_Q(f)|$, where $\|f\|_\infty = \sup|f(\omega)|$,.
2. If λ is a σ-finite measure on $(\Omega, \mathcal{F})$, and h, g are probability densities (i.e. $P(A) = \int_A h(\omega)\lambda(d\omega)$, $Q(A) = \int_A g(\omega)\lambda(d\omega)$, then $\|P-Q\|_{TV} = \dfrac{1}{2}\displaystyle\int_\Omega |g(\omega) - h(\omega)|\lambda(d\omega)$.
3. The bounded measurable functions on $(\Omega, \mathcal{F})$ form a Banach space with $\|f-g\| = \sup|f(\omega) - g(\omega)|$. The dual space (i.e. all linear functionals on the space) is all finite measures, and $\|\mu - \nu\|_{TV} = \sup_{\|f\|\le 1}|\mu(f) - \nu(f)|$ is the dual norm.

Definition 14.2 *Let I be a finite set, $\{X_i\}_{i\in I}$ be a collection of 0/1-valued random variables.*

*A **graph** is an ordered pair (I, E), where I is the set of vertices and $E \subseteq I \times I$ is the set of edges. E must be symmetric (i.e. $(i,j) \in E \Leftrightarrow (j,i) \in E$) and has no loops (i.e. $(i,i) notin E$ for all i).*

*A graph (I, E) is a **dependency graph** for $\{X_i\}_{i\in I}$ if for any two disjoint subsets $I_1, I_2 \subseteq I$ with no edges between them, $\{X_i\}_{i\in I_1}$, $\{X_j\}_{j\in I_2}$ are independent. (A dependency graph need not be unique.)*

If $(i,j) \in E$, we write $i \sim j$.

Definition 14.3 *For a vertex $i \in I$, the **neighborhood** of i is $N_i := \{i\} \cup \{j : (i,j) \in E\}$.*

As a simple example, if $\{X_i\}_{i\in I}$ are all independent, $(I, \emptyset)$ is a dependency graph for the X_i's.

14.1.2 Poisson Heuristic

The following theorem makes the Poisson heuristic precise:

Theorem 14.4 *Let $\{X_i\}_{i\in I}$ be a collection of 0/1-valued random variables, and let (I, E) be a dependency graph for $\{X_i\}_{i\in I}$. Let $p_{ij} = P(X_i = X_j = 1)$, and let P_W denote the probability distribution of $W = \sum X_i$. Then*

$$\|P_W - Poisson(\lambda)\|_{TV} \le \min(3, \lambda^{-1}) \left[\sum_{i\in I} \sum_{j\in N_i\setminus\{i\}} p_{ij} + \sum_{i\in I} \sum_{j\in N_i} p_i p_j \right].$$

We will now go through a series of examples to see this theorem in action.

14.1.2.1 Example 1: Everything is independent

Assume $\{X_i\}_{i\in I}$ are all independent. We can take the dependency graph to be $(I, E) = (I, \emptyset)$, in which case

$$\sum_{i\in I} \sum_{j\in N_i - \{i\}} p_{ij} = 0.$$

Thus, Theorem 14.4 gives

$$\|P_W - \text{Poisson}(\lambda)\|_{TV} \le \min(3, \lambda^{-1}) \sum_{i\in I} p_i^2.$$

(Original case by Poisson) Suppose further that $I = [n] = \{1, \dots, n\}$, $p_i = \dfrac{1}{n}$ for all i. Then $W \sim \text{Binom}(n, \frac{1}{n})$, and $\lambda = \mathbb{E}W = 1$. The inequality above further simplifies to

$$\|P_W - \text{Poisson}(1)\|_{TV} \leq 1 \cdot n \left(\frac{1}{n^2}\right) = \frac{1}{n}.$$

14.1.2.2 Example 2: Card guessing

Assume that we have a deck of n cards labeled $1, 2, \dots, n$. We go through the deck, trying to guess each card. After each guess, we are told what card it actually was (i.e. complete feedback).

Let

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ guess is correct,} \\ 0 & \text{otherwise.} \end{cases}$$

In this case, W is equal to the number of correct guesses.

Under the optimal strategy (guess at random a card that you know is still in the deck), the X_i are independent, and

$$\lambda = \sum_{i=1}^{n} \frac{1}{i} = \log n - \gamma + O(1/n) \approx \log n.$$

Using the empty dependence graph, Theorem 14.4 gives

$$\|P_W - \text{Poisson}(\lambda)\|_{TV} \leq \min(3, \lambda^{-1}) \sum_{i=1}^{n} \frac{1}{i^2} \leq \frac{\pi^2/6}{\log n}.$$

This goes to 0 as $n \to \infty$, but slowly.

14.1.2.3 Example 3: Generalized birthday problem

Consider the following set-up:

- Fix $n \in \mathbb{N}$, $2 \leq k \leq n$.
- Let I be the set of all k-element subsets of $[n]$ (there are $\binom{n}{k}$ of them).
- Fix $c \in \mathbb{N}$.
- Let $\{Y_i\}_{i=1}^{n}$ be independent random variables with $P(Y_i = j) = \dfrac{1}{c}$ for $1 \leq j \leq c$.

(For the ordinary birthday problem: $n = 23$, $c = 365$, $k = 2$.)

For $i \in I$, let

$$X_i = \begin{cases} 1 & \text{if all} Y_j \text{ have the same color for } j \in i, \\ 0 & \text{otherwise.} \end{cases}$$

Then $W = \sum_{i\in I} X_i$ represents the number of k-element subsets whose elements all have the same color. $\lambda = \mathbb{E}W = \binom{n}{k}\frac{1}{c^{k-1}}$, and $p_{ij} = P(X_i = X_j = 1) = c^{-|i\cup j|+1}$.

Consider the graph where $i \sim j \Leftrightarrow i \cap j \neq \emptyset$ (i.e. there is an edge between i and j iff they share a common element). This is a dependency graph for $\{X_i\}$.

With this set-up, Theorem 14.4 gives

$$\|P_W - \text{Poisson}(\lambda)\|_{TV} \leq \min(3, \lambda^{-1})\binom{n}{k}\left[\sum_{a=1}^{k-1}\binom{k}{a}\binom{n-k}{k-a}c^{1-(2k-a)} + c^{2-2k}\sum_{a=1}^{k}\binom{k}{a}\binom{n-k}{k-a}\right].$$

(In the above, we can think of a as the size of the intersection between i and j.)

Case of $k = 2$.

In this case, $\lambda = \binom{n}{2}/c$. Usually c is given to you (e.g. $c = 365$), you want to choose n in a way that λ "doesn't depend on c".

- If $n \ll \sqrt{c}$, a birthday match is essentially impossible: the chance of a match is equal to $P(W > 0)$, which is approximately equal to $(\text{Poisson}(\lambda) > 0) = 1 - e^{-\lambda}$.
- If $n = \theta\sqrt{c}$, the bound given by Theorem 14.4 is on the order of $1/\sqrt{c}$.

Case of $k = 3$.

In this case, $\lambda = \binom{n}{3}/c^2 \sim \frac{n^3}{6c^2}$. We need n to be roughly $c^{2/3}$ for λ to "be a number".

- When c is large and n is of order $c^{2/3}$, the bound is of order $c^{-1/3}$.
- When $n = 84$, $c = 365$, $\lambda \approx 0.7152$, probability of failure $= e^{-\lambda} = 0.4891$.

14.1.2.4 3 basic problems of finite probability

Feller (Introduction to Probability & Applications, Vol. 1) states the following 3 basic problems of finite probability:

1. Birthday problem
2. Coupon collector's problem
3. Matching problem

Each of these problems can be framed in the Poisson heuristic set-up above, and hence can be solved using the machinery we've developed.