

# STATS310A - Lecture 13

Persi Diaconis  
Scribed by Michael Howes

11/04/21

## Contents

<b>1</b>	<b>References and notation</b>	<b>1</b>
<b>2</b>	<b>Law of small numbers</b>	<b>2</b>
2.1	“Approximately”	2
2.2	“Not too dependent”	2
2.3	“Small”	3
<b>3</b>	<b>Examples</b>	<b>3</b>
3.1	Poisson’s original example	3
3.2	Magic factor	3
3.3	Birthday problem	4
3.4	Homework	5
<b>4</b>	<b>Stein’s method</b>	<b>5</b>
<b>5</b>	<b>3 basic problems in probability</b>	<b>5</b>
5.1	Coupon collector	6
5.2	Matching problem	6

## 1 References and notation

Today we will study Poisson approximation and Stein’s method. This is not in the textbook but two references are

- “Poisson Approximation and the Chen-Stein Method” by Arratia, Goldstein and Gord.
- “Exchangeable pairs and Poisson approximation” by Chatterjee, Diaconis and Meckes.

Through out this lecture we will use  $\mathcal{P}_\lambda$  to denote the Poisson distribution with parameter  $\lambda \geq 0$ . That is  $\mathcal{P}_\lambda$  is a probability measure on  $\mathbb{N} = \{0, 1, 2, \dots\}$  and

$$\mathcal{P}_\lambda(\{j\}) = \frac{e^{-\lambda} \lambda^j}{j!}.$$

If  $Z$  is a random variable with distribution  $\mathcal{P}_\lambda$ , we will say that  $Z$  is Poisson( $\lambda$ ). If  $Z$  is Poisson( $\lambda$ ), then

$$\begin{aligned}\mathbb{E}[Z] &= \lambda, \\ \text{Var}(Z) &= \lambda \text{ and,} \\ \mathbb{E}[Z(Z-1)\dots(Z-k+1)] &= \lambda^k.\end{aligned}$$

Also if  $Z_1 \sim \mathcal{P}_{\lambda_1}$  and  $Z_2 \sim \mathcal{P}_{\lambda_2}$  are independent, then  $Z_1 + Z_2 \sim \mathcal{P}_{\lambda_1 + \lambda_2}$ .

## 2 Law of small numbers

Let  $I$  be a finite set and let  $\{X_i\}_{i \in I}$  be a collection of 0 - 1 random variables with expected value  $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = p_i$ . Define

$$\lambda = \sum_{i \in I} p_i,$$

and

$$W = \sum_{i \in I} X_i.$$

If  $p_i$  are “small” and  $\{X_i\}_{i \in I}$  are “not too dependent”, then  $W$  is “approximately” Poisson( $\lambda$ ). We will start this lecture by defining the three terms in quotes.

### 2.1 “Approximately”

To define “approximately” we need to put a topology on the space of probability measures.

**Definition 1.** If  $(\Omega, \mathcal{F})$  is a measurable space and  $P, Q$  are probabilities on  $\Omega$ , then we define the *total variation distance* between  $P$  and  $Q$  to be

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

We will occasionally drop the subscript and simply write  $\|P - Q\|$  for  $\|P - Q\|_{TV}$ .

**Exercise 1.** This is will be on the upcoming homework.

(1) For a function  $f : \Omega \rightarrow \mathbb{R}$ , let  $\|f\|_\infty = \sup_{\omega \in \Omega} |f(\omega)|$ . Then show

$$\|P - Q\| = \sup \left\{ \frac{1}{2} \left| \int f dP - \int f dQ \right| : \|f\|_\infty \leq 1 \right\}.$$

(2) Let  $\mu$  be a  $\sigma$ -finite measure on  $(\Omega, \mathcal{F})$  such that  $P$  and  $Q$  have densities with respect to  $\mu$ . That is there exist measurable functions  $f, g : \Omega \rightarrow [0, \infty]$  such that

$$P(A) = \int_A f d\mu \quad \text{and} \quad Q(A) = \int_A g d\mu,$$

then

$$\|P - Q\|_{TV} = \frac{1}{2} \int |f(\omega) - g(\omega)| d\mu.$$

Note that  $P$  and  $Q$  always have densities with respect to  $\mu = \frac{1}{2}P + \frac{1}{2}Q$ .

### 2.2 “Not too dependent”

A simple undirected graph is a collection of vertices and a collection of edges between the vertices such that there are no loops (edges that start and end at the same vertex) and no multiple edges between vertices.

**Definition 2.** Given a collection of random variables  $\{X_i\}_{i \in I}$ , a *dependency graph* for  $\{X_i\}_{i \in I}$  is a simple undirected graph  $\Gamma$  with vertices  $I$  and edges  $E$  such that for all  $I_1, I_2 \subseteq I$ , if  $I_1 \cap I_2 = \emptyset$  and there are no edges in  $\Gamma$  between  $I_1$  and  $I_2$ , then

$$\{X_i\}_{i \in I_1} \quad \text{and} \quad \{X_j\}_{j \in I_2},$$

are independent of each other.

**Example 1.** If  $\{X_i\}_{i \in I}$  are independent, then the empty graph on  $I$  is a dependency graph for  $\{X_i\}_{i \in I}$ .

**Definition 3.** If  $\Gamma$  is a dependency graph for  $\{X_i\}_{i \in I}$ , then for  $i \in I$ , we define the *neighbourhood* of  $i$  to be the set

$$N_i = \{i\} \cup \{j \in I : (i, j) \in E\}.$$

## 2.3 “Small”

The meaning behind  $p_i$  “small” can be seen in the statement of the Poisson approximation.

**Theorem 1.** Let  $I$  be a finite set and let  $\{X_i\}_{i \in I}$  be 0-1 random variables and let  $\Gamma$  be a dependency graph for  $\{X_i\}_{i \in I}$ . As before let  $p_i = \mathbb{E}[X_i]$ ,  $\lambda = \sum_{i \in I} p_i$  and  $W = \sum_{i \in I} X_i$ . Also define  $p_{ij}$  to be  $\mathbb{P}(X_i = X_j = 1)$ . If we define  $\mathbb{P}_W(A) = \mathbb{P}(W \in A)$  for  $A \subset \mathbb{N}$ , Then

$$\|\mathbb{P}_W - \mathcal{P}_\lambda\| \leq \min\{3, 1/\lambda\} \left( \sum_{i \in I} \sum_{j \in N_i \setminus \{i\}} p_{ij} + \sum_{i \in I} \sum_{j \in N_i} p_i p_j \right).$$

## 3 Examples

### 3.1 Poisson’s original example

Suppose  $I = \{1, \dots, n\}$  and  $X_i$  are i.i.d. with  $\mathbb{P}(X_i = 1) = p$ . Then  $W = \sum_{i=1}^n X_i$ ,  $\lambda = np$  and we can take  $\Gamma$  to be the empty graph. Then the bound becomes

$$\|\mathbb{P}_W - \mathcal{P}_\lambda\| \leq \min\{3, 1/np\} \times np^2.$$

If we take  $p = \frac{1}{n}$ , then  $\frac{1}{\lambda} = 1 < 3$  and so

$$\|\mathbb{P}_W - \mathcal{P}_\lambda\| \leq n \frac{1}{n^2} = \frac{1}{n}.$$

### 3.2 Magic factor

We call the  $1/\lambda$  term “the magic factor.” Suppose again that  $\{X_i\}_{i=1}^n$  are independent and  $\Gamma$  is the empty graph but now  $\mathbb{P}(X_i = 1) = \frac{1}{i}$ . In this example

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

Since the dependency graph is again empty, the first double sum is zero and the second is

$$\sum_{i \in I} \sum_{j \in N_i} p_i p_j = \sum_{i=1}^n \frac{1}{i^2} = \log(n) - \gamma + O(1/n).$$

For large  $n$ ,  $\frac{1}{\log(n)} \leq 3$  and we have the bound

$$\|\mathbb{P}_W - \mathcal{P}_\lambda\| \leq \frac{1}{\log n} \frac{\pi^2}{6}.$$

Note that this example is equivalent to the ESP problem of guessing cards from a deck of  $n$  cards with complete feedback. We can take  $X_i = 1$  if and only if the  $(n - i + 1)^{th}$  guess is correct.

### 3.3 Birthday problem

Suppose we have  $n \in \mathbb{N}$  people,  $C \in \mathbb{N}$  colors and we are interested in matches of size  $k \in \mathbb{N}$ . Define

$$I = \{\alpha : \alpha \subseteq \{1, \dots, n\}, |\alpha| = k\}.$$

Thus  $I$  consists of all group of  $k$  people and  $|I| = \binom{n}{k}$ . Color  $i = 1, \dots, n$  with a colour in  $\{1, 2, \dots, C\}$  uniformly and independently. For  $\alpha \in I$ , define

$$X_\alpha = \begin{cases} 1 & \text{if all elements of } \alpha \text{ have the same color,} \\ 0 & \text{else.} \end{cases}$$

Define  $W = \sum_{\alpha \in I} X_\alpha$ . Then  $W > 0$  if and only if there exists a  $k$ -set that has the same colour. Also

$$p_\alpha = \mathbb{E}[X_\alpha] = C^{1-k}.$$

Thus  $\lambda = \binom{n}{k} C^{1-k}$ . The Poisson approximation suggests that for fixed  $C$  and  $k$  we have

$$\mathbb{P}(W > 0) \approx 1 - \mathbb{P}_\lambda(\{0\}) = 1 - e^{-\lambda} = 1 - e^{-\binom{n}{k} C^{1-k}}.$$

Thus we can set the above equal to  $1/2$  and answer the question of how many people do we need for there to be even odds that at least one group of size  $k$  has all the same color.

- If  $k = 2$  and  $C = 365$ , then we have the usual birthday problem and  $\binom{23}{2} \frac{1}{365} = \log(2)$  to 4 d.p. so  $e^{-\binom{23}{2} \frac{1}{365}} \approx 1/2$  and we recover our previous answer of 23.
- If  $k = 3$  and  $C = 365$ , then  $n = 84$  given  $\lambda \approx 0.7152$  and  $e^{-\lambda} = 0.489$ . Thus in a group of 84 there are roughly equal odds that a group of 3 have the same birthday.
- Taking  $C = 60$  you can ask questions about the odds that a group of  $k$  people have the seconds-hands on their watch in the same position.
- We can also ask other questions. As well as telling us the probability of at least one match, the Poisson approximation also tells us how many matches of size  $k$  we should expect.

Let's calculate the error bound in this example. If  $I$  and  $X_\alpha$  are as above, then  $X_\alpha$  and  $X_\beta$  are independent if and only if  $\alpha \cap \beta = \emptyset$ . Thus we can define a dependency graph  $\Gamma$  where we join  $\alpha$  to  $\beta$  if and only if  $\alpha \cap \beta \neq \emptyset$ .

**Proposition 1.** *With notation and definitions as above we have*

$$\|\mathbb{P}_W - \mathcal{P}_\lambda\| \leq \min\{3, 1/\lambda\} \left( \binom{n}{k} \sum_{a=1}^{k-1} \binom{k}{a} \binom{n-k}{k-a} C^{1-(2k-a)} + \binom{n}{k} \sum_{a=1}^k \binom{k}{a} \binom{n-k}{k-a} C^{2-2k} \right).$$

*Proof.* In the first double sum of the Poisson approximation we have  $\beta \in N_\alpha \setminus \{\alpha\}$  if and only if  $|\alpha \cap \beta| = a$  for some  $a = 1, 2, \dots, k-1$ . The number of such  $\beta$  is precisely  $\binom{k}{a} \binom{n-k}{k-a}$  since we must choose  $a$  elements of  $\alpha$  to be in  $\beta$  and  $k-a$  elements of  $\{1, \dots, n\} \setminus \alpha$  to be in  $\beta$ . Furthermore

$$\begin{aligned} p_{\alpha, \beta} &= \mathbb{P}(\text{everyone in } \alpha \cup \beta \text{ has the same color}) \\ &= C^{1-|\alpha \cup \beta|} \\ &= C^{1-|\alpha| - |\beta| + |\alpha \cap \beta|} \\ &= C^{1-k-k+a} \\ &= C^{1-(2k-a)}. \end{aligned}$$

The calculation for the second double sum is similar. □

### 3.4 Homework

Consider the boys and girls birthday problem. Say we have  $n$  boys and  $n$  girls. We can then ask what is the chance that there is at least one boy-girl pair with the same birthday? We can then ask how large does  $n$  need to be for this probability to equal  $\frac{1}{2}$ . There are many generalizations:

- We can consider any graph that describes how the  $n$  people are connected. We can then ask questions about the probability that there is a collection of  $k$  people that are all connected in the graph and all have the same colour.
- The colours  $i$  can be chosen with probability  $p_i$  instead of uniformly.

## 4 Stein's method

When we prove the Poisson approximation, we will use Stein's method and Stein's equation. We will use the following fact:

**Proposition 2.** *A random variable  $Z$  has  $\mathcal{P}_\lambda$  distribution if and only if for every bounded function  $f : \mathbb{N} \rightarrow \mathbb{R}$ ,*

$$\mathbb{E}[Zf(Z)] = \lambda \mathbb{E}[f(Z+1)]. \quad (1)$$

The equation (1) is called *Stein's equation*. As an exercise one can check that if  $Z$  is  $\text{Poisson}(\lambda)$ , then Stein's equation does indeed hold. The essence of Stein's method is to show that  $W$  approximately satisfies

$$\mathbb{E}[Wf(W)] = \lambda \mathbb{E}[f(W+1)],$$

and show that this implies  $W$  is approximately  $\text{Poisson}(\lambda)$ . To prove Proposition 2 we will prove the following analytic lemma:

**Lemma 1 (\*\*).** *Given  $\lambda > 0$  and  $A \subseteq \mathbb{N}$ , there exists a unique function  $f : \mathbb{N} \rightarrow \mathbb{R}$  such that  $f(0) = 0$  and for all  $j \in \mathbb{N}$ ,  $|f(j)| \leq 1.25$ ,  $|f(j+1) - f(j)| \leq \min\{3, 1/\lambda\}$  and*

$$\lambda f(j+1) - jf(j) = \delta_A(j) - \mathcal{P}_\lambda(A).$$

To see that Lemma (\*\*) implies Proposition 2, note that if  $Z$  is a random variable that satisfies

$$\mathbb{E}[\lambda f(Z+1) - Zf(Z)] = 0,$$

for all bounded  $f$ . Then given a subset  $A \subseteq \mathbb{N}$ , we can find  $f$  as in Lemma (\*\*). This gives

$$\begin{aligned} \mathbb{P}(Z \in A) - \mathcal{P}_\lambda(A) &= \mathbb{E}[\delta_A(Z) - \mathcal{P}_\lambda(A)] \\ &= \mathbb{E}[\lambda f(Z+1) - Zf(Z)] \\ &= 0, \end{aligned}$$

and so  $\mathbb{P}(Z \in A) = \mathcal{P}_\lambda(A)$ . Since this holds for all  $A \subseteq \mathbb{N}$  we can conclude that  $Z$  is  $\text{Poisson}(\lambda)$ .

## 5 3 basic problems in probability

The 3 fundamental problems in combinatorial probability are:

- (1) The birthday problem with bells and whistles.
- (2) Coupon collector problem.
- (3) The matching problem.

All three of these can be solved and studied with Poisson approximation and the law of small numbers. We have already seen how this works for the birthday problem. We will briefly discuss the other two.

### 5.1 Coupon collector

- Drop  $n$  balls into  $B$  boxes where the chance a ball drops into box  $i$  is  $\theta_i$  for  $i = 1, \dots, B$ .
- What is the chance that all boxes are covered?

We can define

$$X_i = \begin{cases} 1 & \text{if box } i \text{ is empty,} \\ 0 & \text{else.} \end{cases}$$

and  $W = \sum_{i=1}^B X_i$ , then  $W = 0$  if and only if all boxes are covered and  $\mathbb{P}(X_i) = (1 - \theta_i)^n$ . We can consider different variations. For example multiple balls could be dropped into distinct boxes at once.

### 5.2 Matching problem

- Two decks containing cards  $1, 2, \dots, N$  are shuffled separately. The cards in the deck are turned over simultaneously. We get a dollar everytime the two cards match.
- What is the probability we make at least one dollar?

Define

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ turn over match,} \\ 0 & \text{else.} \end{cases}$$

and  $W = \sum_{i=1}^N X_i$ . We have  $W > 0$  if and only if there is at least one match. Note that  $\mathbb{P}(X_i = 1) = \frac{1}{N}$ . Again we can consider variations such as multiple cards in the deck with the same label or more than two decks.