

# STATS 310A - Lecture 1

Michael Howes

October 26, 2021

## 1 A typical problem

The basic problem in probability is as follows. We are given a set  $\Omega$ , a subset  $A \subseteq \Omega$  and a function  $\theta : \Omega \rightarrow [0, 1]$  satisfying  $\sum_{\omega \in \Omega} \theta_{\omega} = 1$ . The goal is to calculate or approximate  $\mathbb{P}(A) = \sum_{\omega \in A} \theta_{\omega}$ .

**Example 1.** [*The birthday paradox*] How many people  $k$  do you need in a group to have a 50/50 chance that two or more have the same birthday? Let's answer this when there are  $N$  days in the year. We can then set  $\Omega = [N]^k$  where  $[N] = \{1, 2, \dots, N\}$ . The set where everyone has a different birthday is  $A = \{(x_1, \dots, x_k) \in \Omega \mid x_i \neq x_j \text{ if } i \neq j\}$ . Suppose  $\theta_{\omega} = \frac{1}{N^k}$  which means every combination of birthdays is equally likely. Then  $\mathbb{P}(A) = \frac{|A|}{N^k}$  and  $|A| = N(N-1)(N-2)\dots(N-k+1)$ . So we can conclude that

$$\begin{aligned}\mathbb{P}(A) &= \frac{N(N-1)(N-2)\dots(N-k+1)}{N^k} \\ &= \prod_{i=0}^{k-1} \frac{N-i}{N} \\ &= \prod_{i=0}^{k-1} \left(1 - \frac{i}{N}\right).\end{aligned}$$

While this is an exact answer for  $\mathbb{P}(A)$  it is not very interpretable and it isn't clear which  $k$  satisfies  $\mathbb{P}(A) = \frac{1}{2}$ . We will thus do an approximation using  $\log(1-x) = -x + O(x^2)$ . Thus

$$\begin{aligned}\mathbb{P}(A) &= e^{\sum_{i=1}^{k-1} \log(1 - \frac{i}{N})} \\ &= e^{\sum_{i=1}^{k-1} (-\frac{i}{N} + O((i/N)^2))} \\ &= e^{-\frac{1}{N} \binom{k}{2} + O(k^3/N^2)} \\ &\sim e^{-\frac{k^2}{2N}} \quad \text{if } k = o(N^{\frac{2}{3}}),\end{aligned}$$

where  $a_n \sim b_n$  means  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$ . Setting  $e^{-\frac{k^2}{2N}} = \frac{1}{2}$  gives  $\frac{k^2}{2N} = \log(2)$  and thus  $k = \sqrt{2 \log(2)N}$ , when  $N = 365$ , we have  $k \approx 23$ .

This model has been applied to analysing claims of voting fraud. There were claims that the election was faked because many people had the same name and date of birth. This model can inform us how people we would expect to have the birthday once we have grouped them by name and year. Not that if the probability of each day is not uniform then  $\mathbb{P}(A)$  increases.

**Exercise 1.** For what value of  $k$  are there even odds that 3 or more people have the same bday.

## 2 Admin and overview

In this course, we will be taught three things.

(a) Measure theoretic probability. Many people ask why we should worry about these details of probability/statistics. One answer is

- We need it to work with infinite sequence spaces.
- We often work with probabilities on manifolds for example the space

$$\left\{ x \in \mathbb{R}^n \mid \prod_{i=1}^n x_i = p, \sum_{i=1}^n x_i = s \right\},$$

can be used to assess goodness of fit for a Gamma distribution.

- We will work with probabilities on large sets for example  $\mathcal{P}(\Omega)$  [the set of all probability distributions on  $\Omega$ ]. This is one framework for Bayesian analysis.
- (b) Basic limit theorems: law of large numbers, CLT, Stein's method, Fourier analysis, Poisson convergence, stable laws.
- (c) Probabilistic intuition.

The book we will use is Pat Billingsley's "Probability and Measure". We will study the first two thirds. Many assignment questions will come from Billingsley.

We will have weekly homework, a midterm and a final. The weighting is 30% homework, 30 % midterm and 40 % final.

### 3 Coin flips model

Set  $\Omega = (0, 1]$  and for  $a < b$  define  $p((a, b]) = b - a$ . If  $A = \cup_{i=1}^n (a_i, b_i]$  and the intervals  $(a_i, b_i]$  are disjoint, then define

$$p(A) = \sum_{i=1}^n b_i - a_i.$$

For each  $\omega \in \Omega$ , we can express  $\omega$  in binary

$$\omega = \sum_{i=1}^{\infty} \frac{d_i(\omega)}{2^i} = 0.d_1(\omega)d_2(\omega)\dots$$

If there are two binary expansions for  $\omega$ , pick the non-terminating one. The function  $d_i$  selects the  $i^{th}$  bit of  $\omega$ .

**Theorem 1.** [Weak Law of Large Numbers] *For every  $\varepsilon > 0$ ,*

$$\lim_{n \rightarrow \infty} p \left( \left\{ \omega \in \Omega \mid \left| \frac{1}{n} \sum_{i=1}^n d_i(\omega) - \frac{1}{2} \right| > \varepsilon \right\} \right) = 0.$$

Note that implied in this theorem is that the set

$$A_n = \left\{ \omega \in \Omega \mid \left| \frac{1}{n} \sum_{i=1}^n d_i(\omega) - \frac{1}{2} \right| > \varepsilon \right\},$$

can be written as a finite union of intervals.

*Proof.* Note that each  $d_i$  is a step function of intervals and thus so is  $\frac{1}{n} \sum_{i=1}^n d_i$  and thus the preimage of  $(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon)$  under  $d_i$  is a finite union of intervals. Thus  $p(A_n)$  is well defined.

Now define  $r_i(\omega) = 2d_i(\omega) - 1 \in \{1, -1\}$ . Thus  $d_i = \frac{r_i+1}{2}$  and so

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n d_i(\omega) &= \frac{\sum_{i=1}^n r_i(\omega) + 1}{2n} \\ &= \frac{1}{2n} \sum_{i=1}^n r_i(\omega) + \frac{1}{2}. \end{aligned}$$

Thus we have

$$A_n = \left\{ \left| \frac{1}{2n} \sum_{i=1}^n r_i \right| > \varepsilon \right\}.$$

Note that  $\int_0^1 r_i(\omega) d\omega = 0$  and

$$\int_0^1 r_i(\omega) r_j(\omega) d\omega = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

These two facts give that  $\int_0^1 r_i^2(\omega) d\omega = 1$ . Thus by Markov's inequality we have

$$p(A_n) = p\left(\left|\sum_{i=1}^n r_i\right|^2 > n^2 \varepsilon^2\right) \leq \frac{n}{n^2 \varepsilon^2} = \frac{1}{n \varepsilon^2}.$$

Thus  $p(A_n) \rightarrow 0$  as  $n \rightarrow \infty$ . □

**Proposition 1.** [Markov's inequality] *Let  $f$  be a non-negative step function on  $(0, 1]$ , then*

$$p(\{\omega \mid f(\omega) > a\}) \leq \frac{1}{a} \int_0^1 f(\omega) d\omega$$

for all  $a > 0$ .

*Proof.* Note that

$$\begin{aligned} ap(\omega \mid f(\omega) > a) &= \int_{\{\omega \mid f(\omega) > a\}} a d\omega \\ &\geq \int_{\{\omega \mid f(\omega) > a\}} f(\omega) d\omega \\ &\geq \int_0^1 f(\omega) d\omega. \end{aligned}$$

Dividing by  $a$  gives the result. □

The weak law says that  $\frac{1}{n} \sum_{i=1}^n d_i(\omega)$  is close to  $\frac{1}{2}$  for each  $n$  large with high probability. The strong law asks for which  $\omega$  do we have  $\frac{1}{n} \sum_{i=1}^n d_i(\omega) \rightarrow \frac{1}{2}$  as  $n \rightarrow \infty$ . This does not hold for all  $\omega$  for  $\omega = 0.01111111 \dots$ , the above limit converges to 1. For  $\omega = 0.011000011111111000000000000000 \dots$ , this limit does not even exist. We wish to say that the limit is equal to  $\frac{1}{2}$  for “most”  $\omega \in \Omega$ .

**Definition 1.** A set  $A \subseteq \Omega$  is *negligible* if for every  $\varepsilon > 0$ , there exists  $\{(a_i, b_i]\}_{i=1}^\infty$  such that

$$A \subseteq \bigcup_{i=1}^\infty (a_i, b_i] \text{ and } \sum_{i=1}^\infty b_i - a_i < \varepsilon.$$

**Theorem 2.** *Let*

$$A = \left\{ \omega \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i(\omega) = \frac{1}{2} \right\},$$

*then  $A^c$  is negligible.*

This theorem will be proved next lecture. For now note that

$$A = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \omega \mid \left| \frac{1}{n} \sum_{i=1}^n d_i(\omega) - \frac{1}{2} \right| < \frac{1}{k} \right\}.$$

A last remark on the weak law. The estimate  $p(A_n) < \frac{1}{n\varepsilon^2}$  is very loose. For instance if  $\varepsilon = \frac{1}{100}$ , then we need to take  $n$  much greater than 10000 to be guaranteed a low error. Better bounds are possible.