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

Lecture 1: September 26

Lecturer: Persi Diaconis *Scribes: Kenneth Tay*

1.1 Admin Details

This course aims to teach 3 broad things:

- Measure theoretic probability
- Basic theorems
  - Strong law of large numbers
  - Central limit theorem
  - Poisson convergence
  - Stein's method
- Probabilistic thinking

1.2 Basic Problem of Probability

Let X be a countable set $X = \{x_1, x_2, \dots\}$. Define a function p such that $p(x) \geq 0$ for each x, and $\sum_{x \in X} p(x) = 1$. Let A be some subset of X. Then the basic problem of probability can be stated as follows:

$$\text{Compute/Approximate } p(A) = \sum_{x \in A} p(x).$$

1.3 Example (Birthday Problem)

Question: How many people do you need in a group so that the chance of 2 (or more) people having the same birthday is $\frac{1}{2}$?

The birthday problem can be viewed as trying to fit people into 365 categories (i.e. days). A generalization of this would be trying to fit people into N categories for some positive integer N.

Let's try and set up X, p and A for this problem. Let there be k people present. Assuming that people are uniformly distributed across the categories, we have

$$\begin{aligned} X &= [N]^k, && \text{where } [N] := \{1, 2, \dots, N\} \\ p(x) &= \frac{1}{N^k}, && \forall x \in X \\ A &= \{\text{sequences where } x_i\text{'s are distinct}\}. \end{aligned}$$

(Note here that A actually represents the *complement* of the event whose probability we are trying to find.) We have

$$\begin{aligned}
p(A) &= \sum_{x \in A} \frac{1}{N^k} \\
&= \frac{|A|}{N^k} \\
&= \frac{N(N-1)\dots(N-k+1)}{N^k} \\
&= \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\dots\left(1 - \frac{k-1}{N}\right).
\end{aligned}$$

This is a perfectly reasonable answer but it's not particularly insightful. Let's use some approximations to get a better feel for what's going on:

$$\begin{aligned}
p(A) &= \exp\left[\sum_{i=1}^{k-1} \log\left(1 - \frac{i}{n}\right)\right] \\
&= \exp\left[-\sum_{i=1}^{k-1} \left(\frac{i}{n} + O\left(\frac{i^2}{n^2}\right)\right)\right] && \text{using } \log(1-x) = -x + O(x^2) \\
&= \exp\left[-\binom{k}{2}\Big/N + O\left(\frac{k^3}{N^2}\right)\right].
\end{aligned}$$

To pick k such that $p(A) = \frac{1}{2}$:

$$\begin{aligned}
\exp\left[-\binom{k}{2}\Big/N\right] &= \frac{1}{2}, \\
\frac{k(k-1)}{2} &= N \log 2, \\
k &\approx 1.2\sqrt{N}.
\end{aligned}$$

This means that as N grows, the required k for $p(A) = \frac{1}{2}$ grows as $\sqrt{N}$. In particular, for $N = 365$, we have $k = 22.9 \approx 23$.

Extension (Hard): What is the probability of 3 (or more) people having the same birthday?

1.4 Example (Classical Model for Tossing a Fair Coin)

Let $\Omega = (0, 1]$. Define p on intervals $p\left((a, b]\right) := b - a$. For $A = \bigcup_{i=1}^{k} (a_i, b_i]$ disjoint, define $p(A) := \sum_{i=1}^{k} (b_i - a_i)$.

For each $\omega \in \Omega$, write

$$\omega = \sum_{i=1}^{\infty} \frac{d_i(\omega)}{2^i},$$

where d_i represents the i^{th} digit of the binary expansion of ω. (If there is more than one way of writing the binary expansion, using the non-terminating one.)

1.4.1 Weak Law of Large Numbers

We have the following theorem:

Theorem 1.1 (Weak Law of Large Numbers) *For all* $\varepsilon > 0$,

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

Proof: For convenience, define $r_i = 2d_i - 1$. (r_i's take on the values 1 and -1 instead of 1 and 0.) Then, we can rewrite

$$\begin{aligned}\frac{1}{n}\sum d_i &= \frac{1}{n}\sum\left(\frac{r_i+1}{2}\right)\\ &= \frac{1}{2n}\left(\sum r_i\right) + \frac{1}{2}.\end{aligned}$$

Thus, it suffices to prove that

$$\lim_{n\to\infty} p\left\{\left|\frac{1}{n}\sum r_i\right| > \varepsilon\right\} = 0.$$

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

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

As a result, we have

$$\int_0^1 \left(\sum r_i(\omega)\right)^2 d\omega = n.$$

Hence,

$$\begin{aligned} p\left\{\left|\frac{1}{n}\sum r_i\right| > \varepsilon\right\} &= p\left\{\left(\sum r_i\right)^2 > n^2\varepsilon^2\right\} \\ &\leq \frac{n}{n^2\varepsilon^2} && \text{by Markov's Inequality} \\ &= \frac{1}{n\varepsilon^2} \to 0 \end{aligned}$$

as $n \to \infty$. ■

For completeness, we will state and prove (a version of) Markov's inequality, which we used in the proof above:

Theorem 1.2 (Markov) *For a function* $f : (0,1] \to \mathbb{R}_+$ *and* $a > 0$,

$$p\{\omega : f(\omega) \geq a\} \leq \frac{\int_0^1 f(\omega)d\omega}{a}.$$

Proof:

$$\begin{aligned}\int_0^1 f(\omega)d\omega &= \int_{f(\omega)<a} f(\omega)d\omega + \int_{f(\omega)\geq a} f(\omega)d\omega \\ &\geq \int_{f(\omega)<a} f(\omega)d\omega + a\int_{f(\omega)\geq a} f(\omega)d\omega \\ &\geq 0 + ap\{f \geq a\}.\end{aligned}$$

■

Note: In the proof of the weak law of large numbers, we actually proved that the probability of the event in question was bounded above by $\frac{1}{n\varepsilon^2}$. This bound is actually pretty far off: for example, if $\varepsilon = 0.01$, we would need n to be something like $100,000$ before we see good convergence. The upper bound on the probability is more like $\frac{e^{-n\varepsilon}}{\sqrt{n}}$.

1.4.2 Strong Law of Large Numbers

The weak law of large numbers says that for any *fixed* n, $\frac{1}{n}\sum d_i$ is close to $\frac{1}{2}$. In contrast, the strong law looks at a particular realization of ω and asks, is it true that

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} = \frac{1}{2}? \tag{1.1}$$

This relationship may not hold for all ω.

- Let $\omega^* = 0.01111\ldots$. Then $\lim \frac{S_n(\omega^*)}{n} = 1$.
- The limit may not even exist! For example, if $\omega^* = 0.011000011111111\ldots$, where each series of 2^k ones is followed by 2^{k+1} zeros.

With this in mind, the best we can hope for is for Equation 1.1 to hold for ”most” ω. The following definition will help to make the idea of “most” precise:

Definition 1.3 *A set $B \subseteq (0,1]$ is negligible if for every $\varepsilon > 0$, there exist intervals $(a_i, b_i]$ such that*

$$B \subset \bigcup (a_i, b_i] \qquad \text{and} \qquad \sum p(a_i, b_i] < \varepsilon.$$

Theorem 1.4 (Strong Law of Large Numbers) *Let*

$$A = \left\{\omega : \lim_n \frac{S_n}{n} = \frac{1}{2}\right\}.$$

Then A^c is negligible.

In fact, A can be written out explicitly:

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{\omega : \left|\frac{S_n}{n} - \frac{1}{2}\right| < \frac{1}{k}\right\}.$$