MATH 3160 - Probability - Fall 2013

The Infinite Monkey Theorem (Lecture W 9/11)

"It was the best of times, it was the ***blurst*** *of times?! ...You stupid monkey!... Oh, shut up!"*
— Charles Montgomery Burns

The main goal of this handout is to prove the following statement:

Theorem 1 (Infinite monkey theorem). *If a monkey hits the typewriter keys at random for an* ***infinite*** *amount of time, then he will* ***almost surely*** *produce any type of text, such as the entire collected works of William Shakespeare.*

To say that an event E happens *almost surely* means that it happens with probability 1: $P(E) = 1$[1].
We will state a slightly stronger version of Theorem 1 and give its proof, which is actually not very hard. The key ingredients involve:

- basics of convergent/divergent infinite series (which hopefully you already know!);
- basic properties of the probability (measure) P (see Ross §2.4, done in lecture F 9/6);
- the probability of a monotone increasing/decreasing sequence of events, which is the main content of Ross, Proposition 2.6.1;
- the notion of *independent* events (which we will discuss properly next week, see Ross §3.4, but it's quite easy to state the meaning here).

1 Preliminaries

In this section we outline some basic facts which will be needed later. The first two are hopefully familiar (or at least believable) if you have had freshman calculus:

[1] While *almost surely* and *surely* often mean the same in the context of finite discrete probability problems, they differ when the events considered involve some form of infinity. For instance, in the "dart throwing while blindfolded" example we did in class [F 9/6], I claimed that with probability 1 (almost surely) that the dart cannot land on single points (or a countable union thereof). It's not that it can't happen outright: but the probability that it happens, *measured* with respect to the (uncountably infinite) continuum of all events, is smaller than any positive number.

Proposition 2. *Let $(a_n)_{n=1}^{\infty}$ be a sequence of nonnegative real numbers. Then the series $\sum_n a_n$ converges (write: $\sum_n a_n < \infty$) if and only if its tail series $\sum_{n\geq m} a_n$ tends to 0 as $m \to \infty$.*

Proposition 3. *For any $x \in \mathbb{R}$, $1 - x \leq e^{-x}$.*

Proof. If you wish, plot the graphs of $f_1(x) = 1 - x$ and $f_2(x) = e^{-x}$, and you'll see that $f_1(0) = f_2(0) = 1$, while for all $x \neq 0$, $f_1(x) \lneq f_2(x)$. For a calculus proof, what you need to show is that the differentiable function $g(x) = (1 - x) - e^{-x}$ has a global maximum at the point $x = 0$ with value $g(0) = 0$. ☐

The remaining items involve probabilistic notions.

Proposition 4 (Union bound, Boole's inequality)**.** *Let $(E_n)_n$ be any (finite or infinite) sequence of events. Then $P\left(\bigcup_n E_n\right) \leq \sum_n P(E_n)$.*

Proof. Notice that the right-hand side of the inequality is the first sum appearing in the inclusion-exclusion identity. But as we argued many times in class, this sum in general overestimates $P(\bigcup_n E_n)$ (unless all the E_n's are pairwise disjoint). For a formal proof, see Self-Test Problem 2.14. ☐

Definition 5. We say that an infinite sequence of events $(E_n)_{n=1}^{\infty}$ is *monotone* if either $E_1 \subset E_2 \subset \cdots \subset E_n \subset \cdots$ (monotone increasing) or $E_1 \supset E_2 \supset \cdots \supset E_n \supset \cdots$ (monotone decreasing). Define $\lim_{n\to\infty} E_n$ to mean $\bigcup_{n=1}^{\infty} E_n$, if the sequence is monotone increasing; respectively, $\bigcap_{n=1}^{\infty} E_n$, if the sequence is monotone decreasing.

Proposition 6 (Ross, Proposition 2.6.1)**.** *Let $(E_n)_{n=1}^{\infty}$ be a monotone sequence of events. Then*

$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n).$$

Definition 7 (Ross, §3.4)**.** A finite sequence of events $(E_n)_{n=1}^{N}$ is said to be *independent* if $P(E_1 \cap E_2 \cap \cdots \cap E_N) = P(E_1)P(E_2)\cdots P(E_N)$. An infinite sequence of events $(E_n)_{n=1}^{\infty}$ is said to be independent if all its finite subsequences are themselves independent.

For instance, when you roll a die consecutively, the event that you roll any of $\{1, 2, 3, 4, 5, 6\}$ on the 1st attempt is independent from the event that you roll any of $\{1, 2, 3, 4, 5, 6\}$ on the 2nd attempt, etc. So the probability that you roll a 1 first then a 4 next is the product $P(\{1\})P(\{4\}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

2 The Borel-Cantelli lemma

In this section we will put the preliminary statements to good use. Let $(E_n)_{n=1}^{\infty}$ be an arbitrary, infinite sequence of events. We would like to consider the event that infinitely many of the E_n's happen. To state this in another way, if I give you any positive integer m, you can always find a bigger integer $n \geq m$ such that E_n happens. (If not, then there can't be an infinite number of these E_n's happening.) The set-theoretic way of expressing this is

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n,$$

which stands for the event that the E_n's happen *infinitely often*, abbreviated as (E_n i.o.).

The following result, known as the **Borel-Cantelli lemma**[2], says that the probability of (E_n i.o.) depends on whether the series $\sum_n P(E_n)$ converges or not.

[2]This is related to **Kolmogorov's 0-1 law**, introduced by the same Andrei Kolmogorov whose axioms of probability you're learning in this class. The law says that any "tail event," such as the event (E_n i.o.) (under the independence assumption), will either *almost surely happen* or *almost surely not happen*.

Lemma 8 (Borel-Cantelli). *Let $(E_n)_{n=1}^{\infty}$ be any infinite sequence of events.*

(i) If $\sum_{n=1}^{\infty} P(E_n) < \infty$, then $P(E_n \text{ i.o.}) = 0$.

(ii) Suppose $(E_n)_{n=1}^{\infty}$ are independent. If $\sum_{n=1}^{\infty} P(E_n) = \infty$, then $P(E_n \text{ i.o.}) = 1$.

Proof. For Part (i), notice that $(F_m)_{m=1}^{\infty}$, where $F_m := \bigcup_{n\geq m} E_n$, is a monotone decreasing sequence of events. Hence

$$P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n\right) \underset{\text{Def.5}}{=} P\left(\lim_{m\to\infty}\left(\bigcup_{n=m}^{\infty} E_n\right)\right) \underset{\text{Prop.6}}{=} \lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} E_n\right) \underset{\text{Prop.4}}{\leq} \lim_{m\to\infty}\left(\sum_{n=m}^{\infty} P(E_n)\right).$$

Since it is assumed that the series $\sum_n P(E_n)$ converges, Proposition 2 implies that the right-hand side must be 0. This establishes Part (i).

For Part (ii), recall that an application of DeMorgan's laws gives $\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n\right)^c = \bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}(E_n)^c$. So proving that $P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n\right) = 1$ is equivalent to proving that $P\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}(E_n)^c\right) = 0$. In fact, it is enough just to show that for some positive integer m that

$$P\left(\bigcap_{n=m}^{\infty}(E_n)^c\right) = 0.$$

(Why?), and this is what we prove below. Observe that the sequence $(H_k)_{k=m+1}^{\infty}$, where $H_k := \bigcap_{n=m}^{k}(E_n)^c$, is a monotone decreasing sequence of events. Thus under the assumption that the E_n's (and hence the $(E_n)^c$'s) are independent, we have

$$\begin{aligned} P\left(\bigcap_{n=m}^{\infty}(E_n)^c\right) &\underset{\text{Def.5}}{=} P\left(\lim_{k\to\infty}\bigcap_{n=m}^{k}(E_n)^c\right) \underset{\text{Prop.6}}{=} \lim_{k\to\infty} P\left(\bigcap_{n=m}^{k}(E_n)^c\right) \underset{\text{Def.7}}{=} \lim_{k\to\infty}\prod_{n=m}^{k} P((E_n)^c) \\ &= \lim_{k\to\infty}\prod_{n=m}^{k}(1-P(E_n)) \underset{\text{Prop.3}}{\leq} \lim_{k\to\infty}\prod_{n=m}^{k} e^{-P(E_n)} = \lim_{k\to\infty}\exp\left(-\sum_{n=m}^{k} P(E_n)\right). \end{aligned}$$

Finally, since it is assumed that $\sum_n P(E_n) = \infty$, its tail series $\sum_{n=m}^{\infty} P(E_n)$ must also diverge, so the right-hand side must be 0. □

3 Proof of the infinite monkey theorem

Of course the monkey is just a metaphor here: you could make a robot or a human perform the same task as well! The essence of the problem is really the following (now suitably abstracted):

Theorem 9 (Infinite monkey theorem, mathematical version). *Consider an infinite string of letters $a_1a_2\cdots a_n\cdots$ produced from a finite alphabet (of, say, 26 letters) by picking each letter independently at random, and uniformly from the alphabet (so each letter gets picked with probability $\frac{1}{26}$). Fix a string S of length m from the same alphabet (which is the given "text"). Let E_j be the event that the substring $a_ja_{j+1}\cdots a_{j+m-1}$ is S. Then with probability 1, infinitely many of the E_j's occur.*

Proof. Consider the sequence of events $(E_{mj+1})_{j=0}^{\infty}$. Observe that they are independent events: the event that $a_1a_2\cdots a_m$ is S is independent from the event that $a_{m+1}a_{m+2}\cdots a_{2m}$ is S, etc., since they belong to different "blocks" of the infinite string. Moreover, for every j, $P(E_{mj+1}) = \left(\frac{1}{26}\right)^m$. Therefore $\sum_{j=0}^{\infty} P(E_{mj+1}) = \sum_{j=0}^{\infty}(\frac{1}{26})^m = \infty$. So by Part (ii) of the Borel-Cantelli lemma (Lemma 8), the probability that infinitely many of the E_{mj+1}'s occur is 1. □

To paraphrase this in the language of Theorem 1, but in a stronger sense:

Corollary 10 (Infinite monkey theorem, restated)**.** *If a monkey hits the typewriter keys at random for an* ***infinite*** *amount of time, then he will* ***almost surely*** *produce any type of text, such as the entire collected works of William Shakespeare,* ***infinitely often****.*

What if one doesn't have an infinite amount of time/patience to watch a monkey type away randomly? Then none of the above is guaranteed (with full probability). Realistically, one could ask if a monkey could type up some readable text (or any comprehensible gibberish for that matter) in a reasonably long but finite amount of time. See Jesse Anderson's successful attempt at deploying millions of virtual monkeys to type up the entire collected works of Shakespeare[3]. And for amusement, you can always watch the Simpsons episode where they spoofed the infinite monkey theorem.

[3] See `http://www.jesse-anderson.com/2011/10/a-few-million-monkeys-randomly-recreate-every-work-of-shakespeare` and the hyperlinks therein.