Monkeys typing and martingales

Christopher Lutsko

Abstract

The goal of these notes is to answer the question: how long will it take a monkey to type the word ABRACADABRA? The answer comes from an ingenious application of the theory of Martingales and the optimal stopping theorem. The proof was shown to me in my advisor, Bálint Tóth's wonderful course on Martingales. I highly suggest anyone interested in this subject look at the notes from that course https://people.maths.bris.ac.uk/~mabat/MARTINGALE_THEORY_2016/ which are my primary source for the background. Another great source is the introductory book by Grimmett and Stirzaker [GS20]. I will assume some basic probability and measure theory, but will assume no prior knowledge of martingales.

The goal of this note is to answer the age-old question, "how long would it take a monkey to type the word ABRACADABRA?" (well... commercial typewriters went to market in 1873, so let's say the 250 year old question). Now we may have to make some simplifying assumptions. To help our monkey we will design a typewriter with 26 keys. Moreover let us assume that typing (like most things) is something that monkeys excel at when compared to humans. Thus, we assume they type 250 characters per minute and we also assume monkeys type uniformly at random with no preference for characters. With that, we have the following theorem

Theorem 1. *Under the above assumptions, to type ABRACADABRA, a monkey will take exactly* $26 + 26^4 + 26^{11} = 3,670,344,487,444,778$ *keystrokes. Or approximately* $14,681,377,949,779$ *minutes or* $27,932,606$ *years or* $1,034,541$ *lifespans of the Japanese Macaque.*

Remark. If instead we wanted to type out all of Shakespeare, the answer is not that shocking. One minute on the internet reveals there are $884,421$ words in all of Shakespeare's 43 works. The average (modern) English word has 4.7 characters, and is followed by a space. Thus the number of keystrokes is approximately $4,156,779$. To be safe let's give a range of $3,800,000 - 4,400,000$. Note that we will forgive the monkey for omitting punctuation although we will be very cruel and insist that Shakespeare's work comes in one complete block, in chronological order. Then working approximately it will take $26^{3,800,000} - 26^{4,400,000}$

keystrokes. The lower order terms are absorbed in the approximation and the answer is not that interesting.

The proof relies on martingale theory. The term martingale refers to a betting strategy which is relatively simple: go to a roulette wheel and bet \$1 on red. If you win, go home. If you lose, bet \$2 on red. If you win on your second try, go home with your extra dollar. If you lose, bet \$4 on red. Then keep going. If you win, go home, if you lose double your bet and go again. As long as you have infinite money to sustain your bets, you are guaranteed to win \$1. The probabilistic concept was introduced by Lévy [Lév35] in 1935 and is frequently used to study gambling of all sorts. Martingales are tremendously useful ways to understand 'fair games'. That is, games where the player's expected fortune does not change from move to move. For example, if you go to a casino and they offer you a \$1 bet on a fair coin flip, then your fortune after n flips forms a martingale. Probabilists are tremendously good at using this simple object to study a variety of unexpected problems.

1 Probability spaces and Martingales

Throughout, let

$$\mathbf{X} := (\Omega, \mathcal{F}, \mathbf{P})$$

be a probability space. That is, Ω is the set of all possible outcomes (*sample space*), $\mathcal{F}$ is a σ-algebra on Ω (*event space*), and $\mathbf{P} : \mathcal{F} \to [0, 1]$ denotes the probability measure on the space. A **stochastic process** is a sequence of random variables $X_0, X_1, X_2, \ldots$, jointly defined on $\mathbf{X}$.

A **filtration** of $\mathbf{X}$ is an increasing set sequence of sub σ-algebras

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

The point being that at time n, the σ-algebra contains all information available at time n. Henceforth assume $\mathbf{X} = (\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$ is a filtered probability space.

Example 1.1 Now considering our example, let us set Ω to be the sequence of infinite words taken from our 26-letter alphabet (which we denote $\mathscr{A} := \{a_1, \ldots, a_{26}\}$), $\mathcal{F} = P(\Omega)$ the power set of Ω, and let $\mathbf{P}$ be the probability measure assigning equal probability to each letter in the word. To construct a natural filtration, let $F_{i_1,\ldots,i_n}$ denote the set of all words in Ω whose j^{th} letter is a_{i_j} for $j = 1, \ldots n$. One possible filtration is to set $\mathcal{F}_0$ to be the trivial σ-algebra (i.e $\mathcal{F}_0 = \{\emptyset, \Omega\}$), when no information is known. Then

$$\mathcal{F}_1 = \mathcal{F}_0 \cup \bigcup_{i=1}^{26} F_i$$

Then $\mathcal{F}_1$ contains all information after the monkey has typed one letter. Following this pattern, we set

$$\mathcal{F}_n = \mathcal{F}_{n-1} \cup \bigcup_{\mathbf{i}} F_{\mathbf{i}},$$

where the union is taken over vectors $\mathbf{i} \in \{1, 2, \ldots, 26\}^n$.

Given a stochastic process $\{X_n\}_{n=0}^{\infty}$ on $\mathbf{X}$ we say that a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is **adapted** to $\{X_n\}_{n=0}^{\infty}$ if X_k is $\mathcal{F}_k$ measurable. In our example, let $\varphi_n : \Omega \to \mathbb{R}$ be a measurable function which depends only on the first n entries of ω. Then the filtration we gave is adapted to φ_n.

Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$, a stochastic process $\{X_n\}$ is a **martingale** if

i) The stochastic process $\{X_n\}$ is adapted to the filtration $\mathcal{F}_n$.

ii) The expectation $\mathbf{E}(|X_n|) < \infty$.

iii) For any time $n \geq 0$ we have

$$\mathbf{E}\left(X_{n+1} \mid \mathcal{F}_n\right) = X_n$$

almost surely.

The first two properties are natural technical conditions. The third property is the real defining property of a martingale (known as the *martingale property*), it says that playing another round of the game won't affect the expectation of the stochastic process.

The original application of martingales (and the source of the name) is to fair betting and illustrates the intuition fairly well. Suppose Alice and Bob are playing a game with probability p that Alice wins and probability $1 - p$ Bob wins. Each time they play, Alice places a 1 dollar bet. If Alice wins, she receive a $1/p$ dollar pay-out. This is a fair game since on average neither Alice nor Bob has an advantage. Let X_n denote Alice's fortune at time n. Then X_n is a martingale since every time she plays her expected fortune should be the same as before she has played.

2 Optimal Stopping

Given a filtered probability space $(\Omega, \mathcal{F}_n, \{\mathcal{F}_n\}, \mathbf{P})$, a **stopping time** is a function $T : \Omega \to \overline{\mathbb{N}} = \{0, 1, 2, \ldots\} \cup \{\infty\}$ such that for all $n \in \overline{\mathbb{N}}$ we have

$$\{\omega \ : \ T(\omega) \leq n\} \in \mathcal{F}_n.$$

In other words, at time n we can decide whether or not T has happened.

Returning to Example 1.1, if we write the word $\omega = \omega_1\omega_2\omega_3\dots$ then an example of a stopping time is

$$T(\omega) = \min(n \in \overline{\mathbb{N}} \ : \ \omega_n = a_1).$$

That is, the first time the letter a_1 appears in the word. This is a stopping time for our filtration since, if one knows the first n letters, one can decide whether or not T has happened.

An example of a function which is not a stopping time is

$$T(\omega) = \min(n \in \overline{\mathbb{N}} \ : \ \omega_{n+1} = a_1),$$

the first time the next letter is a_1. This is not a stopping time for our filtration since it requires us to peak into the future to decide whether it has happened at time n.

Given an adapted process $\{X_n\}$ and a stopping time T we can define the so-called **stopped process** $\{X_n^T\}$ to be

$$X_n^T := \min(X_n, X_T).$$

That is, we run X_n until T happens, then we stop. The nice thing about stopped processes is that, if X_n is a martingale, then X_n^T is a martingale (we leave this as an exercise). This leads us to the following theorem, known as Doob's optimal stopping theorem

Theorem 2 (Doob's optimal stopping theorem [GS20]). *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$ be a filtered probability space. Let $\{X_n\}$ be a martingale, and let T be a stopping time with $\mathbf{P}(T < \infty) < \infty$. Then each of the following conditions alone imply the equality*

$$\mathbf{E}\left(X_T\right) = \mathbf{E}\left(X_0\right). \tag{2.1}$$

i) T is almost surely bounded (i.e $\mathbf{P}\left(T \leq N\right) = 1$ for some $N < \infty$).

ii) The stopped martingale $\{X_n^T\}$ is almost surely bounded (i.e $\mathbf{P}\left(\sup_n \left|X_n^T\right| \leq K\right)$ for some $K < \infty$).

iii) The expectation $\mathbf{E}\left(T\right) < \infty$ and there exists a $K < \infty$ such that $\mathbf{E}\left(\left|X_{n+1} - X_n\right| \,\middle|\, \mathcal{F}_n\right) < K$ almost surely.

We are rarely so lucky to be in the first two cases, but the third case is surprisingly useful.

3 Random Typing

Now that we have the definitions of martingales and stopping times, and the optimal stopping theorem, we are ready to try and answer questions about randomly typed characters. The challenge now is to set up an appropriate martingale and to exploit the optimal stopping theorem. For this we start with a toy example:

3.1 Example 1: One heads from a coin toss

First, suppose a coin is tossed with probability p of getting heads, and $1-p$ of getting tails. What is the expected number of tosses needed to get a head?

Formally our sample space Ω, will be set of infinite sequences of heads and tails, the σ-algebra will be the power set and the probability measure is the coin toss probability measure.

Now imagine a game: Bill will pay 1 dollar for each coin toss, if the coin comes up tails, the house wins the dollar. If Bill wins he receives back $1/p$ dollars. To play this game the house will need to start with $1/p - 1$ dollars (otherwise they won't be able to pay Bill his winnings if he wins in the first toss). Let $X_0 = 1/p - 1$ and let X_n denote the house's cash after the n^{th} toss. Let $T : \omega \to \overline{\mathbb{N}}$ the first time a heads is tossed.

We leave it as an exercise to show that X_n is a martingale, T is a stopping time, and together they satisfy the conditions of the optimal stopping theorem. Thus

$$\begin{aligned} \mathbf{E}\left(X_0\right) = 1/p - 1 &= \mathbf{E}\left(X_T\right) \\ &= \mathbf{E}\left(1/p - 1 + T - 1/p\right). \end{aligned}$$

where the last line comes from the fact that at step T Bill has paid T dollars to play and if he wins on that toss he will receive $1/p$ dollars. Thus

$$\mathbf{E}\left(T\right) = 1/p.$$

It is worth noting that if we calculate this expectation explicitly we can prove a cute identity

$$\sum_{n=1}^{\infty} \mathbf{P}\left(T = n\right) = p\sum_{n=0}^{\infty}(1-p)^n(n+1) = p\sum_{n=0}^{\infty} n(1-p)^n + p\sum_{n=0}^{\infty}(1-p)^n = 1/p.$$

Therefore

$$\sum_{n=0}^{\infty} n(1-p)^n = 1/p^2 + 1/p.$$

In general, one can arrive at some interesting series identities using martingales.

Returning to our main aim, this tells us that the expected time for our monkey to type the letter A is $1/26$.

3.2 Example 2: Typing AB

Now returning to the typing example. The probability space and filtration will be the one from Example 1.1. And the game will be similar: Regardless of the previous letter Bill will bet 1 dollar that each letter is an A. If he wins he receives 26 dollars. If the n^{th} letter is A, then Bill will place an additional bet that the $n+1^{th}$ letter is B. The buy-in for this second bet is 26 dollars, and the pay-out is 26^2 dollars. Again, let X_n denote the house's cash at time n. To play the game the house needs at least 26^2-2 dollars. So let $X_0 = 26^2 - 2$ be the house's initial cash.

Now the stopping time, T will be the first time AB appears. Leaving the condition-checking to the reader we again apply the optimal stopping theorem and find

$$\begin{aligned} 26^2 - 2 &= \mathbf{E}\left(X_T\right) \\ &= \mathbf{E}\left(T\right) - 2, \end{aligned}$$

for the second line note that Bill has returned any money he has won up to this point, he has bet T dollars, and has received the pay-out of 26^2. Thus $\mathbf{E}\left(T\right) = 26^2 = 676$.

3.3 Typing ABRACADABRA

Now to type $ABRACADABRA$. For this we follow the same strategy. Regardless of the previous letter, Bill will bet 1 dollar that the next letter is A, with a payout of 26 dollars. If the previous letter was an A he will additionally bet 26 dollars that the next letter is B, with a possible pay-out of 27^2. If the previous two letters were AB he will bet 26^2 dollars that the next letter is R with a pay-off of 26^3. And so forth until he finishes.

Again we let X_n denote the house's net cash at time n. To get started the house will need to account for the worst case scenario (Bill wins in the first instance). In this case Bill wins three bets! One for the word $ABRACADABRA$, one for the word $ABRA$ and one for the word A. Thus his payout will be $26 + 26^4 + 26^{11}$ dollars, but he has put in 11 dollars of his own cash to bet each round. Thus we set $X_0 = 26 + 26^4 + 26^{11} - 11$. Now using optimal stopping we find that

$$\mathbf{E}\left(X_0\right) = 26 + 26^4 + 26^{11} - 11 = \mathbf{E}\left(T\right) - 11.$$

Thus we find that $\mathbf{E}\left(T\right) = 26 + 26^4 + 26^{11} = 3,670,344,487,444,778$.

Now to find the time taken we assume that monkeys type 250 keystrokes per minute giving $14,681,377,949,779$ minutes, or $27,932,606$ years, or $1,034,541$ lifespans of the Japanese Macaque.

References

[GS20] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford University Press, Oxford, 2020. Fourth edition [of 0667520].

[Lév35] P. Lévy. Proprietés asymptotiques des sommes de variables aléatoires énchaînees. *Bull. Sci. Math.*, 59:84–96 & 109–128, 1935.

Department of Mathematics, Rutgers University, Hill Center - Busch Campus, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA. *E-mail:* ***chris.lutsko@rutgers.edu***