Chapter 13: The ABRACADABRA Problem

Elvis Cui

April 10, 2020

Contents

1 The ABRACADABRA Problem

Let $\{X_t, t \in \mathbb{N}\}$ denote random letters drawn independently and uniformly from the English alphabet (i.e., a discrete time stochastic process). More precisely, $X_t(\omega) \in E = \{A, B, C, \cdots, Z\}$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ s.t. $\mathbb{P}(\{X_i = l_i, X_j = l_j\}) = \mathbb{P}(X_i = l_i)\mathbb{P}(X_j = l_j) = \frac{1}{26^2}, \forall i \neq j$.

Next, we define T as the first time in which the word "ABRACADABRA" (a magic word) appears (later I will show this is a **stopping time**, let's take it for granted for now):

$$T(\omega) = \inf_{t \in \mathcal{N}} \{t : X_{t-10}(\omega) = A, X_{t-9}(\omega) = B, \cdots, X_t(\omega) = A\}$$

and we let $\inf_t \emptyset = \infty$. The famous ABRACADABRA problem asks: what is the expectation of T ? In other words, how long should I wait in average until some random sequence appears ?

The answer is:

$$\mathbb{E}(T) = 26^{11} + 26^4 + 26$$

2 Formulation via Gambler's Problem

Suppose before each time t, a gambler comes in with 1 dollar in hand, and he bets that at time t, $X_t(\omega) = A$. If he loses, he quitsand if he wins, he gets 26 dollars back. Then at time $t + 1$, he bets all his money on the event $X_{t+1}(\omega) = B$. Similarly, he gets 26^2 back if he wins and quits if $X_{t+1}(\omega) \neq B$. He will repeat this procudure until $(t + 10)^{th}$ round.

Therefore, at time $(t + 10)$, the t^{th} gambler either has 26^{11} dollars or 0 dollars. To be more rigorous, let's introduce the random variable

$$L_t^j(\omega) = \text{the } (t-j+1)^{th} \text{ letter in ABRACADABRA}$$
$$t \in \mathbb{N},\ j = t, \cdots, t+10$$

Therefore, the probability measure induced by each $L_t^j, t \in \mathbb{N}, j = t, \cdots, t+10$ is a delta measure. Next, let's define $M_t^j(\omega)$ to be the money the j^{th} gambler has at time t. Formally, for each $\omega \in \Omega$,

$$M_t^j := \begin{cases} 1 & \text{if } j > t \text{ or } t = 0 \\ 26 M_{t-1}^j \mathbb{1}\{L_t^j = X_t\} & \text{if } j \leq t \leq j+10 \\ M_{j+10}^j & \text{if } t > j+10 \end{cases}$$

Then, let's define the total reward of all gamblers at time t (i.e., minus the initial 1 dollar):

$$R_t = \sum_{j=1}^{\infty}(M_t^j - 1) = \sum_{j=1}^{t} M_t^j - t$$

$$R_0 = \sum_{j=1}^{\infty}(M_0^j - 1) = 0$$

If we regard t as a random variable (i.e., the stopping time we are looking for), then

$$R_T = \sum_{j=1}^{T} M_T^j - T = \sum_{t=1}^{\infty} \mathbb{1}\{T = t\} \left(\sum_{j=1}^{t} M_t^j - t \right)$$

Here the case $T(\omega) = \infty$ is excluded because we have $\mathbb{E}(T) < \infty$. The most important thing about R_T is that for any $\omega \in \{\omega : T(\omega) = t\}$, we have

$$R_t(\omega) = \sum_{j=1}^{t} M_t^j(\omega) - t = 26^{11} + 26^4 + 26 - t$$

The reasons are the following: since T is the first time the sequence "ABRACADABRA" appears, so every gambler before time $(T-10)$ will lose all his money. Again, since T is the stopping time, the $(T-10)^{th}$ gambler will have 26^{11} dollars at the end. Similarly, the $(T-3)^{th}$ and T^{th} gambler will get 26^4 and 26 dollars respectively.

Finally, if we can show that

$$\mathbb{E}(R_T) = \mathbb{E}(R_0) = 0$$

We will have

$$\mathbb{E}(T) = 26^{11} + 26^4 + 26$$

The details are shown in below.

3 Proof

3.1 T is a stopping time

Define the filtration $\{\mathcal{F}_t, t \in \mathbb{N}\}$ as:

$$\mathcal{F}_t = \sigma\left(\sigma(X_1, \cdots, X_t), \sigma(L_m^j, m \in \mathbb{N}, j = m, \cdots, m+10)\right)$$

Then we have

$$\{T = t\} \in \sigma(X_1, \cdots, X_t) \subset \mathcal{F}_t$$

Moreover, $T \wedge t$ is a bounded stopping time.

3.2 $\mathbb{E}(T)$ is finite

To show the expectation of the stopping time T is finite, I use the following lemma:

Lemma 1 *Suppose that T is a stopping time s.t. for some N in $\mathbb{N}$ and some $\epsilon > 0$, we have for every $t \in \mathbb{N}$:*

$$\mathbb{P}(T \leq t + N | \mathcal{F}_t)(\omega) > \epsilon \ a.s.$$

Then we have $\mathbb{E}(T) < \infty$.

Since for any t, the probability of $\{\omega : X_t(\omega) = A, \cdots, X_{t+10}(\omega) = A\}$ is $\frac{1}{26^{11}}$, the condition in the lemma holds.

3.3 R_t has bounded increments

For each $\omega \in \Omega$, we have

$$\begin{aligned}
|R_t - R_{t-1}| &= |\sum_{j=1}^{t-1}(M_t^j - M_{t-1}^j) + M_t^t - 1| \\
&\leq |M_t^t - 1| + \sum_{j=1}^{t-1} |M_t^j - M_{t-1}^j| \\
&\leq 26 + 10 \times 26^{11}
\end{aligned}$$

3.4 $\mathbb{E}(R_T) = \lim_t \mathbb{E}(R_{T \wedge t})$

Since $\mathbb{E}(T) < \infty$, we have that for each $\omega \in \Omega$ except for a set of $\mathbb{P}$-measure 0,

$$R_{T \wedge t} \rightarrow_{a.s.} R_T$$

Therefore, the claim holds if we can show $R_{T \wedge t}$ is bounded by a random variable integrable w.r.t $\mathbb{P}$ (a consequence of Lebesgue's Dominated Convergence Theorem). But we have

$$\begin{aligned}
|R_{T \wedge t} - R_0| &= |\sum_{k=1}^{T \wedge t}(R_k - R_{k-1})| \\
&\leq \sum_{k=1}^{T \wedge t} |R_k - R_{k-1}| \\
&\leq 11 \times 26^{11}(T \wedge t) \\
&\leq 11 \times 26^{11} T
\end{aligned}$$

Note T has finite expectation and $R_0 = 0$.

3.5 $\{R_t, t \in \mathbb{N}\}$ forms a martingale

Now we want to show $\mathbb{E}(R_{T \wedge t}) = \mathbb{E}(R_0) = 0$. If we can show that R_t forms a mean 0 martingale sequence, then by **Doob's Optional Sampling Theorem** (see next section), $R_{T \wedge t}$ will also be a mean 0 martingale.

First, we have to show for each t, R_t is $\mathcal{F}_t$ measurable. Since for each t, for any j, M_t^j depends only on $\{X_1, \cdots, X_t, L_1^j, \cdots, L_t^j, j \in \mathbb{N}\}$, R_t belongs to the sigma-algebra $\mathcal{F}_t$.

Second, we need to show R_t is integrable for any t. But indeed we have

$$|R_t - R_0| \leq 11 \times 26^{11} t$$

Lastly, by the definition of conditional expectation,

$$\mathbb{E}(R_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{t} \mathbb{E}(M_t^j - 1|\mathcal{F}_{t-1}) \tag{1}$$

If we can show that $\{M_t^j - 1\}$ forms a mean 0 martingale, then R_t will also be a mean 0 martingale since we will have

$$\sum_{j=1}^{t} \mathbb{E}(M_t^j - 1|\mathcal{F}_{t-1}) = \sum_{j=1}^{t}(M_{t-1}^j - 1) = \sum_{j=1}^{t-1}(M_{t-1}^j - 1) = R_{t-1}$$

But for any t and j, using the definition of M_t^j, we get

$$\mathbb{E}(M_t^j - 1|\mathcal{F}_{t-1}) = \begin{cases} 1 - 1 = 0 = M_{t-1}^j - 1 & \text{if } t = 0 \text{ or } j > t \\ 26M_{t-1}^j \mathbb{P}(L_t^j = X_t) - 1 = M_{t-1}^j - 1 & \text{if } j \le t \le j + 10 \\ M_{j+10}^j - 1 = M_{t-1}^j - 1 & \text{if } t > j + 10 \end{cases}$$

Thus, $\{M_t^j - 1\}$ forms a mean 0 martingale sequence for any j.

3.6 Doob's Optional Sampling Theorem

If a sequence $\{R_t, t \in \mathbb{N}\}$ forms a martingale w.r.t. $\{\mathcal{F}_t, t \in \mathbb{N}\}$. Then for any 2 stopping times T and S s.t. $\mathbb{P}(T \le N) = \mathbb{P}(S \le N) = 1$ for some N, we have

$$\mathbb{E}(R_T|\mathcal{F}_S) = R_{T\wedge S} = \mathbb{E}(R_S|\mathcal{F}_T)$$

Now if we take $S(\omega) = t - 1$ and T to be $T \wedge t$, then

$$\mathbb{E}(R_{T\wedge t}|\mathcal{F}_{t-1}) = R_{T\wedge(t-1)}$$

Therefore, $\{R_{T\wedge t}, t \in \mathbb{N}\}$ is indeed a martingale sequence.

4 More examples

4.1 Flipping a coin

One is flipping a coin, then the expected waiting time of the occurrence of the sequence "HHTTHH" is

$$\mathbb{E}(T) = 2^6 + 2^2 + 2 = 70$$

4.2 Patterns of genes

The expected length of seeing a pattern "ATGCATG" is

$$\mathbb{E}(T) = 4^7 + 4^3 = 16448$$

4.3 Family names of Chinese

According to "the Book of Family Names" (written by someone in the North Song dynasty), there are 504 family names among Chinese. So if we randomly pick people until the sequence of family name "Zhao, Qian, Sun, Li" appears, the expected number will be

$$\mathbb{E}(T) = 504^4 = 64524128256$$

But of course this is not realistic: the distribution of names if not uniform and the total population is finite.

4.4 Rolling a die

Suppose we roll a die many times, the expected time to observe a sequence "6 6 6 6 6 6" is

$$\mathbb{E}(T) = 6^6 + 6^5 + 6^4 + 6^3 + 36 + 6$$