

S&DS 351: Stochastic Processes - Homework 6

Bryan SebaRaj

Professor Ilias Zadik

April 4, 2025

Problem 4.6

[Wright-Fisher process] This is a famous urn model from genetics. An urn contains d balls, some black and some white. (These might represent two forms of a gene. We might ask questions like: as genes are sampled randomly to form successive generations, what is the probability that the “white” gene will take over and completely replace the other form?) Here is a way to form the next “generation.” Sample a ball at random from the urn d times, *with replacement*. We get a new, random set of d balls, containing $0, 1, \dots, d$ white balls, with various probabilities. Call the resulting sample generation 1. Then we sample in the same way from generation 1 to form generation 2, and so on. Let X_t denote the number of white balls in generation t . After many generations, the population will have become “fixed”, that is, it will consist of just one color. Suppose that $X_0 = x_0$, some number between 0 and d .

- (a) (2 points) Show that the process $\{X_t\}$ is a martingale.

Suppose that in generation t the urn contains $X_t = x$ white balls (with $0 \leq x \leq d$). To form generation $t + 1$, we sample d times with replacement from the urn. Thus, each ball in generation $t + 1$ is white with probability

$$p = \frac{x}{d}.$$

It follows that, conditioned on $X_t = x$, the number of white balls in generation $t + 1$, denoted by X_{t+1} , follows a binomial distribution:

$$X_{t+1} \mid \{X_t = x\} \sim \text{Binomial}\left(d, \frac{x}{d}\right).$$

Hence, the conditional expectation is

$$\mathbb{E}[X_{t+1} \mid X_t = x] = d \cdot \frac{x}{d} = x.$$

Since this holds for all x , we have

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = X_t,$$

where \mathcal{F}_t is the natural filtration. Thus, $\{X_t\}$ is a martingale.

- (b) (2 points) Prove (and not assume like the Problem 4.6 statement says) that X_t will eventually become fixed, that is, it will consist of one color, with probability 1.

The state space for $\{X_t\}$ is finite, namely $\{0, 1, 2, \dots, d\}$. Notice that the states 0 and d are absorbing because once $X_t = 0$ (all black) or $X_t = d$ (all white), every subsequent generation will remain in that state.

For any state x with $0 < x < d$, the probability that all d balls sampled are white is

$$\left(\frac{x}{d}\right)^d > 0,$$

and the probability that all d balls sampled are black is

$$\left(\frac{d-x}{d}\right)^d > 0.$$

Thus, from any nonabsorbing state there is a strictly positive probability that the next generation will be absorbed into either state 0 or d .

Moreover, because $\{X_t\}$ is a bounded martingale (since $0 \leq X_t \leq d$ for all t), the Martingale Convergence Theorem implies that X_t converges almost surely to a limit, say X_∞ . Given the dynamics of the process, the only possible limit values are 0 and d . Hence,

$$\mathbb{P}(X_\infty \in \{0, d\}) = 1,$$

which means the process eventually becomes fixed (i.e., the population becomes all one color) with probability 1.

- (c) (3 points) Use the martingale to show that the probability that the population eventually becomes all white is x_0/d .

Let

$$p = \mathbb{P}(X_\infty = d)$$

be the probability that the process eventually fixes at state d (all white). Since $\{X_t\}$ is a martingale and converges almost surely to X_∞ , we have

$$\mathbb{E}[X_\infty] = \lim_{t \rightarrow \infty} \mathbb{E}[X_t] = x_0,$$

where x_0 is the initial number of white balls.

Since X_∞ takes only the values 0 and d , we can write

$$\mathbb{E}[X_\infty] = 0 \cdot \mathbb{P}(X_\infty = 0) + d \cdot \mathbb{P}(X_\infty = d) = dp.$$

Thus,

$$dp = x_0 \implies p = \frac{x_0}{d}.$$

Therefore, the probability that the population eventually becomes all white is $\frac{x_0}{d}$.

Problem 4.10 (3 points)

[Doob's inequality for submartingales] Let X_0, X_1, \dots be a nonnegative submartingale, and let b be a positive number. Prove that

$$\mathbb{P}(\max(X_0, \dots, X_n) \geq b) \leq \frac{\mathbb{E}(X_n)}{b},$$

using the following steps as hints:

- (a) Define τ to be the first time t such that $X_t \geq b$, or n , whichever comes first; that is, $\tau = \inf\{t : X_t \geq b\} \wedge n$. Argue that $\{\max(X_0, \dots, X_n) \geq b\} = \{X_\tau \geq b\}$.

Let $\{X_t\}_{t \geq 0}$ be a nonnegative submartingale and let $b > 0$. Define the stopping time

$$\tau = \inf\{t \geq 0 : X_t \geq b\} \wedge n.$$

We first show that

$$\left\{ \max_{0 \leq t \leq n} X_t \geq b \right\} = \{X_\tau \geq b\}.$$

(a) If $\max_{0 \leq t \leq n} X_t \geq b$, then there exists a time $t \leq n$ such that $X_t \geq b$. By the definition of τ , it follows that $\tau \leq t$ and hence $X_\tau \geq b$. Conversely, if $X_\tau \geq b$, then clearly

$$\max_{0 \leq t \leq n} X_t \geq X_\tau \geq b.$$

Thus, the events are identical.

- (b) Apply Markov's inequality, and use an Optional Sampling theorem.

By Markov's inequality,

$$\mathbb{P}(X_\tau \geq b) \leq \frac{\mathbb{E}[X_\tau]}{b}.$$

Since τ is bounded by n , the Optional Sampling Theorem applies to the submartingale $\{X_t\}$ and gives

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_n].$$

Therefore, combining the above estimates we obtain

$$\mathbb{P}\left(\max_{0 \leq t \leq n} X_t \geq b\right) = \mathbb{P}(X_\tau \geq b) \leq \frac{\mathbb{E}[X_\tau]}{b} \leq \frac{\mathbb{E}[X_n]}{b}.$$

Problem 4.13 (5 points)

In Theorem (4.38) there is nothing special about assuming nonnegativity, that is, a lower bound of zero. Show that if $\{X_t\}$ is a supermartingale and there is a random variable X with $\mathbb{E}[|X|] < \infty$ and $X_t \geq X$ for all t , then X_t converges with probability 1 as $t \rightarrow \infty$.

Theorem (4.38): A nonnegative supermartingale converges with probability 1.

Let $\{X_t\}_{t \geq 0}$ be a supermartingale and suppose there exists a random variable X with $\mathbb{E}[|X|] < \infty$ such that

$$X_t \geq X \quad \text{for all } t \geq 0.$$

Define the process

$$Y_t = X_t - X, \quad t \geq 0.$$

Since $X_t \geq X$, we have $Y_t \geq 0$ for all t , so $\{Y_t\}$ is a nonnegative process. Next, we check the supermartingale property for $\{Y_t\}$:

$$\mathbb{E}[Y_{t+1} | \mathcal{F}_t] = \mathbb{E}[X_{t+1} - X | \mathcal{F}_t] = \mathbb{E}[X_{t+1} | \mathcal{F}_t] - X.$$

Since $\{X_t\}$ is a supermartingale, we have

$$\mathbb{E}[X_{t+1} | \mathcal{F}_t] \leq X_t.$$

Thus,

$$\mathbb{E}[Y_{t+1} | \mathcal{F}_t] \leq X_t - X = Y_t.$$

Therefore, $\{Y_t\}$ is a nonnegative supermartingale.

By Theorem (4.38), every nonnegative supermartingale converges almost surely. That is, there exists a random variable Y_∞ such that

$$Y_t \rightarrow Y_\infty \quad \text{almost surely as } t \rightarrow \infty.$$

Since

$$X_t = Y_t + X,$$

it follows that

$$X_t \rightarrow Y_\infty + X \quad \text{almost surely as } t \rightarrow \infty.$$

Thus, $\{X_t\}$ converges with probability 1.

Problem 4.16 (10 points)

Let $\{M_t\}$ be a likelihood ratio martingale as discussed in Example 4.9.

(a) Show that $\mathbb{E}M_t = 1$ for all t .

are given that $M_0 = 1$ and, for each $t \geq 0$,

$$M_{t+1} = M_t \cdot \frac{g(X_{t+1})}{f(X_{t+1})}.$$

Since the random variables X_1, X_2, \dots are independent and M_t is \mathcal{F}_t -measurable (where $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$), we have

$$\mathbb{E}(M_{t+1} | \mathcal{F}_t) = M_t \cdot \int \frac{g(x)}{f(x)} f(x) dx = M_t.$$

Taking expectations on both sides yields

$$\mathbb{E}[M_{t+1}] = \mathbb{E}(\mathbb{E}(M_{t+1} \mid \mathcal{F}_t)) = \mathbb{E}[M_t].$$

Since $\mathbb{E}[M_0] = 1$, by induction we conclude that

$$\mathbb{E}[M_t] = 1 \quad \text{for all } t \geq 0.$$

(b) Show that as $t \rightarrow \infty$, we have $M_t \rightarrow 0$ with probability 1.

Assume that $g \neq f$, so that the likelihood ratio $\frac{g(x)}{f(x)}$ is not identically equal to 1. Note that under the measure corresponding to the density f , by Jensen's inequality we have

$$\mathbb{E}_f \left[\log \frac{g(X_1)}{f(X_1)} \right] \leq \log \mathbb{E}_f \left[\frac{g(X_1)}{f(X_1)} \right] = \log 1 = 0,$$

with strict inequality if $g \neq f$. Hence, defining

$$\mu := \mathbb{E}_f \left[\log \frac{g(X_1)}{f(X_1)} \right],$$

we have $\mu < 0$.

Since X_1, X_2, \dots are i.i.d. with density f , by the Strong Law of Large Numbers,

$$\frac{1}{t} \sum_{i=1}^t \log \frac{g(X_i)}{f(X_i)} \rightarrow \mu \quad \text{almost surely.}$$

That is,

$$\frac{1}{t} \log M_t \rightarrow \mu < 0 \quad \text{almost surely.}$$

Thus, $\log M_t \rightarrow -\infty$ almost surely, which implies

$$M_t \rightarrow 0 \quad \text{almost surely, i.e. with probability 1.}$$

Problem 4.19 (5 points)

Suppose T is a stopping time and $\{X_t\}$ is a submartingale. Define $Y_t = X_{t \wedge T}$. Show that $\{Y_t\}$ is a submartingale.

Let $\{X_t\}_{t \geq 0}$ be a submartingale with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$, and let T be a stopping time. Define the stopped process $\{Y_t\}$ by

$$Y_t = X_{t \wedge T}, \quad t \geq 0.$$

We wish to show that $\{Y_t\}$ is a submartingale, i.e., that for all $0 \leq s < t$,

$$\mathbb{E}[Y_t \mid \mathcal{F}_s] \geq Y_s \quad \text{a.s.}$$

Fix $0 \leq s < t$. Note that the event $\{T \leq s\}$ and its complement $\{T > s\}$ partition the probability space.

Case 1: On the event $\{T \leq s\}$, we have

$$s \wedge T = T \quad \text{and} \quad t \wedge T = T,$$

so that

$$Y_s = X_T \quad \text{and} \quad Y_t = X_T.$$

Thus,

$$\mathbb{E}[Y_t \mid \mathcal{F}_s] = X_T = Y_s \quad \text{on } \{T \leq s\}.$$

Case 2: On the event $\{T > s\}$, we have $s \wedge T = s$, and

$$t \wedge T = \begin{cases} t, & \text{if } T > t, \\ T, & \text{if } s < T \leq t. \end{cases}$$

In either subcase, using the submartingale property of $\{X_t\}$ and the Optional Sampling Theorem (which applies because $t \wedge T$ is bounded by t), we obtain

$$\mathbb{E}[X_{t \wedge T} \mid \mathcal{F}_s] \geq X_s.$$

That is,

$$\mathbb{E}[Y_t \mid \mathcal{F}_s] \geq X_s = Y_s \quad \text{on } \{T > s\}.$$

Combining the two cases, we conclude that

$$\mathbb{E}[Y_t \mid \mathcal{F}_s] \geq Y_s \quad \text{almost surely for all } 0 \leq s < t.$$

Thus, $\{Y_t\}$ is a submartingale.

1. (Snell's Envelope (Optimal Stopping))

This problem will review the optimal employee (in expected value) selection problem. Suppose that we have N candidates that come in for an interview, one at a time. We can model the fit of every candidate to the position of interest as a stochastic process $X_n \in \mathbb{R}_{\geq 0}$ for $0 \leq n \leq N$ and we assume that we know the joint distribution of the X_n .

Assumption: We assume that if we pass on a candidate, we cannot ask them to return.

Note: The selection procedure must follow a stopping time because we cannot “look into the future”. In contrast, if we could ask a candidate to return, then our choice would clearly be to select $\max_{1 \leq n \leq N} X_n$.

Therefore, our goal in this problem is to find which stopping time T maximizes $\mathbb{E}X_T$, i.e., it satisfies that for any other stopping time S , $\mathbb{E}X_T \geq \mathbb{E}X_S$.

We will construct a supermartingale called Snell's envelope Y_n recursively (downwards) as follows:

$$\begin{aligned} Y_N &= X_N \\ Y_n &= \max(X_n, \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n]), \quad 1 \leq n \leq N-1. \end{aligned}$$

Turns out the optimal stopping time is given by $T = \inf\{n \geq 0 \mid X_n = Y_n\}$. Let's prove it!

- (a) (2 points) Prove that $Y_n \geq X_n$ for all n .

We use backward induction.

Base case: At time N ,

$$Y_N = X_N,$$

so $Y_N \geq X_N$ trivially.

Inductive step: Suppose for some $n+1 \leq N$ we already know $Y_{n+1} \geq X_{n+1}$. Then by the definition of Y_n ,

$$Y_n = \max(X_n, \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n]).$$

Clearly

$$Y_n \geq X_n \quad \text{and} \quad Y_n \geq \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n].$$

Hence $Y_n \geq X_n$. By induction, $Y_n \geq X_n$ for all n .

- (b) (3 points) Prove that $Y_n, n \geq 0$ is a supermartingale with respect to $X_n, n \geq 0$.

To show (Y_n) is a supermartingale, we must check

$$\mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n] \leq Y_n \quad \text{for each } n.$$

From the recursive definition,

$$Y_n = \max(X_n, \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n]) \geq \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n].$$

Taking the conditional expectation on both sides of Y_{n+1} and noting that Y_n is $\sigma(X_0, \dots, X_n)$ -measurable, we immediately obtain

$$\mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n] \leq Y_n,$$

which is precisely the supermartingale property.

- (c) (10 points) Use the proof of the optional stopping theorem for supermartingales to conclude $Y_{n \wedge T}$ is a supermartingale with respect to $X_n, n \geq 0$. Here $n \wedge T := \min\{n, T\}$.

Since $(Y_n)_{n=0}^N$ is a supermartingale and T is a stopping time (with respect to the natural filtration generated by X_0, \dots, X_n), the Optional Stopping Theorem (valid under suitable integrability or boundedness conditions, which hold here since N is finite and $X_n \geq 0$) ensures that the stopped process $Y_{n \wedge T}$ is also a supermartingale.

Concretely, for $n \geq 0$,

$$\mathbb{E}[Y_{(n+1) \wedge T} \mid X_0, \dots, X_n] \leq Y_{n \wedge T}.$$

Thus $Y_{n \wedge T}$ is a supermartingale with respect to (X_n) .

- (d) (10 points) Improve your argument to show that $Y_{n \wedge T}$ is in fact a *martingale* with respect to $X_n, n \geq 0$.

We strengthen the supermartingale property to a martingale property. We must show

$$\mathbb{E}[Y_{(n+1) \wedge T} \mid X_0, \dots, X_n] = Y_{n \wedge T} \quad \text{for each } n.$$

Notice by definition

$$T = \inf\{n \geq 0 : X_n = Y_n\}.$$

Hence *before* T occurs (i.e. for all $n < T$), we have

$$X_n < Y_n \implies Y_n = \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n].$$

At such $n < T$, the process is in the “continuation region” where $Y_n = \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]$.

Case 1: If $n < T$, then $n \wedge T = n$. We claim

$$\mathbb{E}[Y_{(n+1) \wedge T} \mid X_0, \dots, X_n] = \mathbb{E}[Y_{n+1} \mid X_0, \dots, X_n] = Y_n = Y_{n \wedge T}.$$

The first equality holds because $(n+1) \wedge T = n+1$ on $\{n < T\}$. The second equality follows from $Y_n = \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]$ in the continuation region.

Case 2: If $n \geq T$, then $n \wedge T = T$ and $(n+1) \wedge T = T$, so both sides are just Y_T , making the conditional expectation identity trivial.

Hence, for *all* $n \geq 0$, we have the martingale property

$$\mathbb{E}[Y_{(n+1) \wedge T} \mid X_0, \dots, X_n] = Y_{n \wedge T}.$$

Thus, $Y_{n \wedge T}$ is a true martingale (again, using boundedness ensures no pathologies with optional stopping).

- (e) (2 points) Use the above to conclude that $\mathbb{E}X_T = \mathbb{E}Y_0$.

From part (d), $(Y_{n \wedge T})$ is a martingale. In particular,

$$\mathbb{E}[Y_{n \wedge T}] = \mathbb{E}[Y_0] \quad \text{for all } n.$$

Since $T \leq N$ (the process length is finite), if we choose $n = N$, we have

$$Y_{N \wedge T} = Y_T.$$

But by definition of T , on the event $\{T < N\}$ we have $Y_T = X_T$. Also at $T = N$ obviously $Y_T = X_T$. So

$$Y_{N \wedge T} = Y_T = X_T \implies \mathbb{E}[Y_T] = \mathbb{E}[X_T].$$

Hence

$$\mathbb{E}[X_T] = \mathbb{E}[Y_T] = \mathbb{E}[Y_{N \wedge T}] = \mathbb{E}[Y_0].$$

- (f) (3 points) Suppose that we are given an alternative stopping rule S . Prove that $\mathbb{E}Y_S \geq \mathbb{E}X_S$.

Since we proved in part (a) that $Y_n \geq X_n$ for every n , it follows for any stopping time S that

$$Y_S \geq X_S \quad \text{pointwise.}$$

Taking expectations preserves the inequality:

$$\mathbb{E}[Y_S] \geq \mathbb{E}[X_S].$$

- (g) (10 points) Now prove that $\mathbb{E}Y_S \leq \mathbb{E}Y_0$. Put the pieces together to prove that $\mathbb{E}X_T \geq \mathbb{E}X_S$.

Since (Y_n) is a supermartingale, for the stopped process $(Y_{n \wedge S})$ the Optional Stopping Theorem implies

$$\mathbb{E}[Y_{n \wedge S}] \leq \mathbb{E}[Y_0] \quad \text{for all } n.$$

By choosing $n = N$ (which exceeds or meets all possible values of S), we get

$$\mathbb{E}[Y_S] \leq \mathbb{E}[Y_0].$$

Putting this together with part (f),

$$\mathbb{E}[X_S] \leq \mathbb{E}[Y_S] \leq \mathbb{E}[Y_0] = \mathbb{E}[X_T].$$

Hence

$$\mathbb{E}[X_T] \geq \mathbb{E}[X_S],$$

showing T is indeed an optimal stopping time.

- (h) (Bonus, 5 points) Can you intuitively explain why T is the optimal stopping time?

Note: there is no “correct” answer to this, but curious to read your thoughts.

(Bonus) Intuitive explanation of why T is optimal.

The time T is defined as the first time the observed value X_n meets (or exceeds) the “continuation value” $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]$. This continuation value is exactly the expected payoff if we choose *not* to stop now. Therefore, if X_n is already larger than the expected value of continuing, it makes sense to stop immediately. Conversely, if the current X_n is smaller than that future expectation, one gains in expectation by waiting. This procedure encodes the best tradeoff between stopping now vs. continuing, hence yields the best expected payoff among all stopping rules.

- (i) Suppose that the data $X_n, n \geq 0$ are drawn uniformly from $[0, 1]$ and are i.i.d.

- (i) (2 points) What is $\mathbb{E}Y_0$ for $N = 2$?

Finally, suppose X_0, X_1, \dots, X_N are i.i.d. uniform(0, 1) random variables. We compute $\mathbb{E}[Y_0]$ for small N .

(i) $N = 2$ case. Here we have three random variables X_0, X_1, X_2 , each uniform(0, 1). By definition:

$$Y_2 = X_2, \quad Y_1 = \max(X_1, \mathbb{E}[Y_2 \mid X_1]) = \max(X_1, \mathbb{E}[X_2]) = \max(X_1, 1/2),$$

since X_2 is independent of X_1 and has mean $1/2$. Next,

$$Y_0 = \max(X_0, \mathbb{E}[Y_1 \mid X_0]) = \max(X_0, \mathbb{E}[Y_1]),$$

because Y_1 is also independent of X_0 . So we need $\mathbb{E}[Y_1]$.

Compute

$$Y_1 = \max(X_1, 1/2).$$

Hence

$$\mathbb{E}[Y_1] = \int_0^1 \max(x, 1/2) dx = \int_0^{1/2} \frac{1}{2} dx + \int_{1/2}^1 x dx = \frac{1}{2} \cdot \frac{1}{2} + \left[\frac{x^2}{2} \right]_{1/2}^1 = \frac{1}{4} + \left(\frac{1}{2} - \frac{1}{8} \right) = \frac{5}{8}.$$

Therefore

$$Y_0 = \max(X_0, 5/8).$$

Taking expectation again,

$$\mathbb{E}[Y_0] = \int_0^1 \max(x, 5/8) dx = \int_0^{5/8} \frac{5}{8} dx + \int_{5/8}^1 x dx = \frac{5}{8} \cdot \frac{5}{8} + \left[\frac{x^2}{2} \right]_{5/8}^1.$$

First part: $5/8 \times 5/8 = 25/64$. Second part:

$$\left[\frac{x^2}{2} \right]_{5/8}^1 = \frac{1}{2} - \frac{(5/8)^2}{2} = \frac{1}{2} - \frac{25}{64} \cdot \frac{1}{2} = \frac{1}{2} - \frac{25}{128} = \frac{64}{128} - \frac{25}{128} = \frac{39}{128}.$$

Hence

$$\mathbb{E}[Y_0] = \frac{25}{64} + \frac{39}{128} = \frac{50}{128} + \frac{39}{128} = \frac{89}{128} \approx 0.6953.$$

Thus for $N = 2$, $\mathbb{E}[Y_0] = 89/128$.

(ii) (3 points) What about $N = 3$?

We have X_0, X_1, X_2, X_3 . Then

$$Y_3 = X_3, \quad Y_2 = \max(X_2, 1/2),$$

and

$$Y_1 = \max(X_1, \mathbb{E}[Y_2]) = \max(X_1, \mathbb{E}[\max(X_2, 1/2)]).$$

We already computed

$$\mathbb{E}[\max(X_2, 1/2)] = \frac{5}{8}.$$

Hence

$$Y_1 = \max(X_1, 5/8).$$

Next,

$$Y_0 = \max(X_0, \mathbb{E}[Y_1]) = \max(X_0, \mathbb{E}[\max(X_1, 5/8)]).$$

We just adapt the same integral for $c = 5/8$ again. As before,

$$\mathbb{E}[\max(X_1, 5/8)] = \int_0^1 \max(x, 5/8) dx = \frac{5}{8} \cdot \frac{5}{8} + \int_{5/8}^1 x dx = \frac{25}{64} + \left(\frac{1}{2} - \frac{25}{128} \right) = \frac{25}{64} + \frac{39}{128} = \frac{89}{128}.$$

Thus

$$Y_0 = \max(X_0, 89/128).$$

Finally,

$$\mathbb{E}[Y_0] = \int_0^1 \max(x, 89/128) dx = \int_0^{89/128} \frac{89}{128} dx + \int_{89/128}^1 x dx = \frac{89}{128} \cdot \frac{89}{128} + \left[\frac{x^2}{2} \right]_{89/128}^1.$$

The first part is $\frac{89}{128} \times \frac{89}{128} = \frac{7921}{16384}$. For the second part,

$$\left[\frac{x^2}{2} \right]_{89/128}^1 = \frac{1}{2} - \frac{(89/128)^2}{2} = \frac{1}{2} - \frac{7921}{16384} \cdot \frac{1}{2} = \frac{1}{2} - \frac{7921}{32768} = \frac{16384}{32768} - \frac{7921}{32768} = \frac{8463}{32768}.$$

Hence

$$\mathbb{E}[Y_0] = \frac{7921}{16384} + \frac{8463}{32768} = \frac{15842}{32768} + \frac{8463}{32768} = \frac{24305}{32768} \approx 0.742.$$

So for $N = 3$, $\mathbb{E}[Y_0] = 24305/32768 \approx 0.742$.

2. (The “ABRACADABRA” problem)

A monkey is sitting next to a typewriter. We assume that the typewriter has exactly 26 keys corresponding to the 26 letters of the English alphabet.

We also assume the following action is taking place: at each time the monkey hits one out of the 26 keys chosen with equal probability and independently from everything else.

(a) (2 points) Prove that the expected number of hits until the letter A is typed is 26.

Let $T := \min\{n \geq 1 : \text{the } n\text{-th typed letter is } A\}$. Then T is a geometric random variable with success probability $p = \frac{1}{26}$. Indeed,

$$\mathbb{P}(T > n) = \left(\frac{25}{26} \right)^n, \quad \mathbb{P}(T = n) = \left(\frac{25}{26} \right)^{n-1} \frac{1}{26}.$$

The expectation of such a geometric($p = \frac{1}{26}$) random variable is

$$\mathbb{E}[T] = \frac{1}{p} = 26.$$

Thus the expected number of keystrokes to see the first A is 26.

- (b) Now let's prove part (a) with martingale theory. Let us assume that a gambling game takes place at a casino based on the keys the monkey hits.

Gambling game: Before each new hit of a key from the monkey, a new gambler appears and bets one dollar that the key the monkey is about to hit is going to be the letter A. If the monkey hits A the casino pays 26 dollars to the gambler, otherwise the gambler leaves the game and the casino wins the gambler's one dollar.

- (3 points) Prove that the “game is fair”, i.e., the total budget of the casino before every new hit is a martingale.

Let M_n be the total budget of the casino *prior* to the n -th keystroke (so $M_0 = 0$ initially). Just before the n -th letter appears:

The new gambler pays \$1, so the casino's budget increases by 1.

Then, when the n -th letter is realized:

$$\begin{cases} \text{If it is A, the casino pays \$26 back, for a net } -25 \text{ (relative to } +1), \\ \text{If it is not A, the gambler loses \$1, and the casino keeps that \$1.} \end{cases}$$

Hence the increment from M_{n-1} to M_n is:

$$M_n - M_{n-1} = \begin{cases} +1 - 26 = -25, & \text{with probability } 1/26, \\ +1, & \text{with probability } 25/26. \end{cases}$$

Therefore

$$\mathbb{E}[M_n - M_{n-1} \mid M_{n-1}, \dots] = \frac{1}{26} \times (-25) + \frac{25}{26} \times (+1) = 0.$$

So

$$\mathbb{E}[M_n \mid M_{n-1}, \dots] = M_{n-1}.$$

This means $(M_n)_{n \geq 0}$ is indeed a (discrete-time) martingale with respect to its natural filtration.

- (5 points) Provide an alternative proof of part (a) using the Optional Stopping Theorem.

Let T be the *first time the letter A appears*, exactly as in part (a). By construction, T is a stopping time with respect to (M_n) . Because M_n is bounded below (in fact $M_n \geq -25n$) and T cannot exceed the time we first see A—or we may simply note T has a finite expectation for a geometric distribution—we can apply the Optional Stopping Theorem (OST) safely under standard conditions for nonnegative or bounded increments. *Step 1: Calculate M_T .* Observe that up to time T (the first A), we have had exactly T gamblers, each paying \$1, so the casino collects \$ T . But at time T , the letter is indeed A, and the gambler at time T wins \$26, so the casino must pay \$26 to that gambler. The net is:

$$M_T = \underbrace{(+1) + \dots + (+1)}_{T \text{ times}} - 26 = T - 26.$$

Step 2: Apply OST. Since M_n is a martingale and T a bounded stopping time (geometric implies finite expectation, so the standard version of OST applies):

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] = 0.$$

Hence

$$\mathbb{E}[T - 26] = 0, \implies \mathbb{E}[T] = 26.$$

Thus the expected number of keystrokes before the letter A appears is 26, matching the direct calculation in part (a).

- (c) (15 points) You may be thinking that part (b) is a very complicated way to prove part (a) which is simple. That is true, but this proof idea is in fact very clever and can give other much harder results, like the following.

Build upon the proof technique of part (b), and prove that the expected number of hits until the monkey types **ABRACADABRA** is $26^{11} + 26^4 + 26$.

We define a process (M_n) , the casino's capital at time n , that places side-bets on whether the next typed letters complete any partial progress toward **ABRACADABRA**. Each time we transition from one "partial match state" to another, an appropriate net amount is paid in or out so that the increments of M_n have zero expectation. The process is carefully designed so that once **ABRACADABRA** first appears, M_τ can be expressed (almost surely) in a way that reveals τ in the formula. The OST then concludes with the desired $\mathbb{E}[\tau]$. Step 1: Label the prefix-matching states.

Define the states $s = 0, 1, 2, \dots, 10$ to represent "how many initial letters of **ABRACADABRA** have been matched so far (most recent)".

ABRACADABRA = A (0) B (1) R (2) A (3) C (4) A (5) D (6) A (7) B (8) R (9) A (10),

where $s = 0$ means "currently matched 0 letters (no partial match)," $s = 1$ means "the most recent letter typed was A, matching 1 letter of the pattern so far," and so on, up to $s = 10$ meaning we have matched all but one final letter A so that if the next letter is A, we complete the entire pattern **ABRACADABRA**.

Step 2: Overlap function.

Crucially, note that **ABRACADABRA** has a length-4 prefix **ABRA** that is also a suffix of that word. In effect:

ABRACADABRA has suffix **ABRA** of length 4 (and that suffix is also the first 4 letters).
length 11

Hence if we are in state $s = 10$ (i.e., we have matched **ABRACADABR**), and the next letter typed is A, then we enter the success state $s = 11$ (meaning we've matched the entire word **ABRACADABRA** for the first time). But if the next letter typed is B—which might partially match a new pattern—then we effectively jump to a state that reflects how many letters of **ABRACADABRA** (from the beginning) are matched by the overlap. For **ABRACADABR** + B, the new situation is $s = 2$ (since the last two letters typed are now **AB**, which is the first 2-letter prefix of **ABRACADABRA**). There is a systematic way to define these transitions for each possible letter from any state s . We won't write out all transitions in detail here, but only highlight that:

Next letter typed \longrightarrow (jump to new state reflecting the new overlap).

Step 3: Define the casino's payoff structure / side-bets.

We construct a capital process M_n (the casino's profit at time n) as follows:

- At every step n , depending on the *current* state $s_n \in \{0, 1, \dots, 10\}$ (the number of matched letters so far) and the letter typed, the casino either *wins or loses a precisely defined amount* so that

$$\mathbb{E}[\Delta M_n \mid \text{previous info}] = 0.$$

- Concretely, from each state s , if the next letter is the correct letter to advance to $s + 1$ (or $s = 10$ to $s = 11$, i.e. success), a certain payoff is made to the gambler. Otherwise, the gambler pays the casino. The size of the bets is chosen so that each step is a fair bet from the viewpoint of the casino.

One explicit (though slightly tedious) construction is:

At each step, a new gambler invests an amount X_s if we are in state s .

If the typed letter transitions $s \rightarrow t$, the gambler is paid X_t and leaves. The net effect on the casino is

$$\Delta M = -X_t + X_s.$$

We want $\mathbb{E}[\Delta M \mid s] = 0$, thus

$$X_s = \sum_{\ell=\text{all letters}} \left[\frac{1}{26} (-X_{\text{nextState}(s,\ell)} + X_s) \right].$$

Solving this system, plus specifying boundary conditions $X_{11} = 0$ at the absorbing success state, yields unique solutions

$$\{X_0, X_1, \dots, X_{10}, X_{11} = 0\}.$$

(Here $X_{11} = 0$ because once the pattern is fully matched, that gambler's bet is over.)

A standard result (or by direct algebra) shows:

$$X_0 = 26^{11} + 26^4 + 26, \quad X_4 = 26^4 + 26, \quad X_{10} = 26, \quad X_{11} = 0, \quad \text{and similarly for the other } X_s.$$

(The powers $26^{11}, 26^4, 26$ directly reflect how partial overlaps restart the pattern matching.)

Step 4: The process M_n is a martingale and at time τ we can solve for τ .

Thus M_n is built so that each jump $\Delta M_n = M_n - M_{n-1}$ has zero conditional expectation given the state s_{n-1} . Hence (M_n) is a martingale. Let τ be the time we first reach state $s = 11$ (i.e., ABRACADABRA is fully matched). By the Optional Stopping Theorem,

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

But $M_0 = X_0$ since we start in state $s = 0$ with that gambler's stake, i.e. $M_0 = X_0 = 26^{11} + 26^4 + 26$. Meanwhile, once τ is reached, the gambler in state 10 receives $X_{11} = 0$ but had invested $X_{10} = 26$. So the net effect for the casino from that final transition is $-X_{11} + X_{10} = +26$ minus any prior partial reimbursements. Collecting all partial contributions, one can check carefully that

$$M_\tau = \tau - (26^{11} + 26^4 + 26)$$

(a direct analogy to the single-letter case). More precisely:

- Summed across all steps from 1 to τ , gamblers have collectively contributed τ dollars into the casino (1 dollar each time).
- At times the pattern partially matches, the casino pays out intermediate amounts, culminating in a final payoff at time τ that ensures the gambler in state 10 (or some partial overlap sequence) is compensated.

A more direct or carefully enumerated argument obtains

$$M_\tau = \tau - (X_0 - X_{11}) = \tau - ((26^{11} + 26^4 + 26) - 0).$$

Hence

$$M_\tau = \tau - (26^{11} + 26^4 + 26).$$

Using $\mathbb{E}[M_\tau] = M_0 = 26^{11} + 26^4 + 26$,

$$\mathbb{E}\left[\tau - (26^{11} + 26^4 + 26)\right] = 26^{11} + 26^4 + 26,$$

which immediately gives

$$\mathbb{E}[\tau] = (26^{11} + 26^4 + 26) + (26^{11} + 26^4 + 26) - 0 = 26^{11} + 26^4 + 26.$$

Therefore the expected time to see the entire ABRACADABRA is

$26^{11} + 26^4 + 26.$