FinM 34500/ Stat 39000
STOCHASTIC CALCULUS

Lecture 1

Greg Lawler

January 3-4, 2011

STOCHASTIC CALCULUS (FINM 34500 and STAT 39000

Instructor: Greg Lawler, 415 Eckhart
Information for the course will be posted on my web page:
There will be no text for the course, but those interested in purchasing books might consider
G. Lawler, Introduction to Stochastic Processes
S. Shreve, Stochastic Calculus for Finance II: Continuous-time models

- weekly homework, midterm (on Monday, February 7) , final exam (Grading: approx. 10%, 40%,50%)
- (Tentatively) the lectures of February 21 and February 28 will be given at the Singapore site.
- There will be overlap with material in the first quarter, especially Roger Lee's course. However, there courses are not necessary for this course.

- I assume that people have a familiarity with probability and statistics at an "undergraduate, post calculus" level. While stochastic "calculus" is the primary topic of this course, other topics in probability and stochastic processes will be discussed as we go along.
- Stochastic calculus needs measure theory to develop the subject rigorously. However, this is not needed to understand most of the important ideas. We will not assume measure theory in this course, but I may make occasional references to measure theoretic ideas for those who have this background.
- Although this course will not discuss numerical and simulation techniques in detail, there will be occasional exercises that will require students to write a simple code to do a simulation. These can be done in any language or software that students want. (Any student who feels uncomfortable writing a simple program should learn how to do so quickly before trying to get a job in the finance industry.)

APPLIED MATHEMATICS

- Studying some "real world phenomenon", e.g., stock prices
- Make a mathematical model.
  - Most of the time this requires some simplification in the description of the system
  - The model itself is precise — know what assumptions we are making.
  - Would like model to be "robust" — in other words, if the true situation is slightly different than the model than we hope our predictions will also be only slightly different.
- Use mathematical techniques to study the model and make predictions.
  - Rigorous mathematical proof
  - Nonrigorous (heuristic) mathematical analysis
  - Computer calculation
  - Monte Carlo simulation

- Check predictions to real-world. If the predictions do not agree there are two possible reasons:
  - Error in the mathematical analysis
  - The mathematical model is not good

This course will emphasize the mathematical models (assumptions) and the mathematical analysis.

MARTINGALES — FAIR GAMES

A martingale is the mathematical model of a fair game. Its definition uses the idea of conditional expectation.
Let $X_1, X_2, \ldots$ be random variables which we think of as a time series, that is arriving one at a time. At time n we know the value of $X_1, \ldots, X_n$. Let Y be another random variable. Then the conditional expectation $E[Y \mid X_1, \ldots, X_n]$ is the best guess for Y given the values of $X_1, \ldots, X_n$. For notational ease, we will write this as $E[Y \mid \mathcal{F}_n]$ where $\mathcal{F}_n$ denotes the "information" available at time n. (We assume that information is not lost over time!) We will assume that Y is an integrable random variable, that is, $\mathbb{E}[|Y|] < \infty$.

- If $n = 0$, then $E[Y \mid \mathcal{F}_0]$ is the best guess for Y with no information. We let this be $\mathbb{E}[Y]$.
- $E[Y \mid \mathcal{F}_n]$ depends only on $X_1, \ldots, X_n$,

$$E[Y \mid \mathcal{F}_n] = \phi(X_1, \ldots, X_n).$$

  We say that $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$-measurable.

- Suppose (X, Y) have joint density $f(x, y)$. Given X, the conditional density for Y is

$$f(y \mid X) = \frac{f(X, y)}{\int_{-\infty}^{\infty} f(X, z)\, dz}.$$

Then

$$E[Y \mid X] = \int_{-\infty}^{\infty} y\, f(y \mid X)\, dy.$$

We can write a similar formula for $E[Y \mid X_1, \ldots, X_n]$ if $(X_1, \ldots, X_n, Y)$ have joint density $f(x_1, \ldots, x_n, y)$.

$$E[Y \mid \mathcal{F}_n] = \frac{\int_{-\infty}^{\infty} y\, f(\mathbf{X}, y)\, dy}{\int_{-\infty}^{\infty} f(\mathbf{X}, y)\, dy}, \quad \mathbf{X} = (X_1, \ldots, X_n).$$

Using examples such as this, one arrives at a characterizing property for conditional expectation.

Definition

If Y is an integrable random variable, then $E[Y \mid \mathcal{F}_n]$ is the unique random $\mathcal{F}_n$-measurable random variable such that for each $\mathcal{F}_n$-measurable event,

$$\mathbb{E}\left[1_A\, Y\right] = \mathbb{E}\left[1_A\, E[Y \mid \mathcal{F}_n]\right].$$

- Here 1_A denotes the "indicator function" of the random variable which equals 1 or 0 respectively, depending if A does or does not occur. The event A (and hence also 1_A) is $\mathcal{F}_n$ measurable if one can determine whether it occured from $X_1, \ldots, X_n$.
- One can show existence of this using the "Radon-Nikodym theorem" — however, in most cases one can compute it directly.
- Knowing the properties of conditional expectation is most important.

Properties of Conditional Expectation

- if Y is $\mathcal{F}_n$-measurable, then $\mathbb{E}(Y \mid \mathcal{F}_n) = Y$.
- If A is an $\mathcal{F}_n$-measurable event,

$$\mathbb{E}\left[1_A \, E[Y \mid \mathcal{F}_n]\right] = \mathbb{E}\left[1_A \, Y\right].$$

- $\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\right] = \mathbb{E}[Y]$. (This is a special case of the previous fact with A being the entire space.)
- Suppose $X_1, \ldots, X_n$ are independent of Y. Then, there is no useful information in $\mathcal{F}_n$ and

$$E[Y \mid \mathcal{F}_n] = \mathbb{E}[Y].$$

- Linearity: $E[aY + bZ \mid \mathcal{F}_n] = aE[Y \mid \mathcal{F}_n] + bE[Z \mid \mathcal{F}_n]$
- Suppose Z is a random variable that is measurable with respect to $\mathcal{F}_n$. Then, it acts like a constant.

$$E[YZ \mid \mathcal{F}_n] = Z \, E[Y \mid \mathcal{F}_n].$$

- Projection property: If $n < m$, then

$$E\left[\, E[Y \mid \mathcal{F}_m] \mid \mathcal{F}_n\right] = E[Y \mid \mathcal{F}_n].$$

EXAMPLES

- Suppose $X_1, X_2, \ldots$ are independent, identically distributed (i.i.d.) random variables with $\mathbb{E}[X_j] = \mu$. Let

$$S_n = X_1 + \cdots + X_n.$$

If $m < n$, then $S_n - S_m$ is independent of $\mathcal{F}_m$ and

$$\begin{aligned} E[S_n \mid \mathcal{F}_m] &= E[S_m \mid \mathcal{F}_m] + E[S_n - S_m \mid \mathcal{F}_m] \\ &= S_m + E[S_n - S_m] = S_m + \mu\,(n-m) \end{aligned}$$

- Suppose $\mu = 0$ and $\mathbb{E}[X_j^2] = \sigma^2$.

$$\begin{aligned} &E[S_n^2 \mid \mathcal{F}_m] \\ &\quad = E([S_m + (S_n - S_m)]^2 \mid \mathcal{F}_m) \\ &\quad = E[S_m^2 \mid \mathcal{F}_m] + 2\,E[S_m(S_n - S_m) \mid \mathcal{F}_m] + E[(S_n - S_m)^2 \mid \mathcal{F}_m]. \end{aligned}$$

Since S_m is $\mathcal{F}_m$-measurable and $S_n - S_m$ is independent of $\mathcal{F}_m$,

$$E[S_m^2 \mid \mathcal{F}_m] = S_m^2$$

$$E[(S_n - S_m)^2 \mid \mathcal{F}_m] = \mathbb{E}[(S_n - S_m)^2] = \mathrm{Var}[S_n - S_m] = (n-m)\,\sigma^2$$

$$\begin{aligned} 2\,E[S_m(S_n - S_m) \mid \mathcal{F}_m] &= 2\,S_m\,E(S_n - S_m) \mid \mathcal{F}_m] \\ &= 2\,S_m\,\mathbb{E}[S_n - S_m] = 0. \end{aligned}$$

Combining the terms, we get

$$\mathbb{E}[(S_n - S_m)^2 \mid \mathcal{F}_m] = S_m^2 + (n-m)\,\sigma^2.$$

- What is $E(X_1 \mid S_n)$? By symmetry,

$$E(X_1 \mid S_n) = E(X_2 \mid S_n) = \cdots \mathbb{E}(X_n \mid S_n).$$

$$\begin{aligned} n\,E(X_1 \mid S_n) &= \sum_{j=1}^{n} E[X_j \mid S_n] \\ &= E(X_1 + \cdots + X_n \mid S_n) \\ &= E(S_n \mid S_n) = S_n. \end{aligned}$$

Hence $E(X_1 \mid S_n) = S_n/n$. (This does not depend on the distribution of X_1)

Given a sequence of random variables $X_1, \ldots, X_n$, we call the increasing sequence of information $\mathcal{F}_n$ a (discrete time) filtration.

- The rigorous definition of a filtration uses σ-algebras. However, this intuitive definition will suffice for our purposes.

Definition

A sequence of random variables $M_0, M_1, \ldots$ is called a martingale with respect to the filtration $\{\mathcal{F}_n\}$ if:

- For each n, M_n is an integrable $\mathcal{F}_n$-measurable random variable.
- If $m < n$,

$$E(M_n \mid \mathcal{F}_m) = M_m. \tag{1}$$

Remarks

- If the filtration is not given explicitly, the implicit assumption is that $\mathcal{F}_n$ is the information in $M_1, \ldots, M_n$. (We always need this much information, but we sometimes have more.)
- The condition (1) can be written as

$$E(M_n - M_m \mid \mathcal{F}_m) = 0.$$

- To prove (1), it suffices to show for each n,

$$E(M_{n+1} \mid \mathcal{F}_n) = M_n.$$

To see this, suppose this is true for each n. Then,

$$\begin{aligned} E(M_{n+2} \mid \mathcal{F}_n) &= \mathbb{E}[\, E(M_{n+2} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n] \\ &= E[M_{n+1} \mid \mathcal{F}_n] = M_n, \end{aligned}$$

etc.

- If instead we require for $m < n$,

$$E[M_n \mid \mathcal{F}_m] \geq M_m$$

the process is called a submartingale (a game in our favor). If

$$E(M_n \mid \mathcal{F}_m) \leq M_m$$

it is called a supermartingale (unfair game). Martingales are both submartingales and supermartingales.

EXAMPLES

- Let $X_1, X_2, \ldots$ be independent, mean zero, random variables. Let $S_n = X_1 + \cdots + X_n$. Then S_n is a martingale with respect to $X_1, \ldots, X_n$.
- Under the same assumption assume also that $\mathrm{Var}(X_n) = \mathbb{E}(X_n^2) = \sigma_n^2$. Let

$$S_n = X_1 + \cdots + X_n$$

$$A_n = \sigma_1^2 + \cdots + \sigma_n^2.$$

Then $M_n = S_n^2 - A_n$ is a martingale (with respect to $X_1, \ldots, X_n$). Clearly M_n is $\mathcal{F}_n$-measurable with $\mathbb{E}[|M_n|] < \infty$. Also

$$\begin{aligned} & E[M_{n+1} \mid \mathcal{F}_n] \\ &= E[(S_n + X_{n+1})^2 - A_{n+1} \mid \mathcal{F}_n] \\ &= E[S_n^2 \mid \mathcal{F}_n] + 2E[S_n X_{n+1} \mid \mathcal{F}_n] + E[X_{n+1}^2 \mid \mathcal{F}_n] - A_{n+1} \\ &= S_n^2 + 2S_n E[X_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] - A_{n+1}. \end{aligned}$$

Since X_{n+1} is independent of $\mathcal{F}_n$,

$$E[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0,$$

$$E[X_{n+1}^2 \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}^2] = \sigma_{n+1}^2.$$

Combining this we get

$$E[M_{n+1} \mid \mathcal{F}_n] = S_n^2 + \sigma_{n+1}^2 - A_{n+1} = S_n^2 - A_n = M_n.$$

Discrete stochastic integral

Let M_n be a martingale with respect to $\mathcal{F}_n$. We can think of M_n as the value of some asset. Let B_n be a random variable that is $\mathcal{F}_{n-1}$-measurable. We think of B_n as the amount of money invested right before the value of M_n is released. (B_n can be negative which can represent shorting positions.) Write

$$M_n = M_0 + \sum_{j=1}^{n} \Delta_j, \quad \Delta_j = M_j - M_{j-1}.$$

Let

$$Z_n = \sum_{j=1}^{n} B_j \, \Delta_j = \sum_{j=1}^{n} B_j \, [M_j - M_{j-1}].$$

Then (assuming $\mathbb{E}[|B_j \, \Delta_j|] < \infty$ for each j), Z_n is a martingale with respect to $\mathcal{F}_n$. To see this

$$\begin{aligned} E(Z_{n+1} \mid \mathcal{F}_n) &= E(Z_n + B_{n+1} \, \Delta_{n+1} \mid \mathcal{F}_n) \\ &= Z_n + E(B_{n+1} \, \Delta_{n+1} \mid \mathcal{F}_n). \end{aligned}$$

Since B_{n+1} is $\mathcal{F}_n$-measurable,

$$E(B_{n+1}\,\Delta_{n+1} \mid \mathcal{F}_n) = B_{n+1}\,E(\Delta_{n+1} \mid \mathcal{F}_n),$$

and since M_n is a martingale

$$E(\Delta_{n+1} \mid \mathcal{F}_n) = E(M_{n+1} - M_n \mid \mathcal{F}_n) = 0.$$

Example: Martingale Betting Strategy

The martingale betting strategy is a way to "beat a fair game". The idea is to keep doubling your bet until you win and then you stop. Formally, the martingale is

$$M_n = X_1 + \cdots + X_n$$

where X_j are independent with

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

We start with bet $B_0 = 1$, and at time j we bet 2^j if we have not won before time j and otherwise we bet 0.

$$B_j = \begin{cases} 2^j & \text{if } X_1 = \cdots = X_{j-1} = -1 \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the probability distribution of Z_n is

$$\mathbb{P}\{Z_n = 1 - 2^n\} = 2^{-n}, \quad \mathbb{P}\{Z_n = 1\} = 1 - 2^{-n}.$$

With probability one,

$$Z_\infty = \lim_{n \to \infty} Z_n = 1.$$

Note that for each n, $\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 0$ (this must be true for a martingale!), but

$$1 = \mathbb{E}[Z_\infty] \neq \mathbb{E}[Z_0] = 0.$$

OPTIONAL SAMPLING THEOREM (OST)

The optional sampling or optional stopping theorem (OST) states that "you cannot beat a fair game". We will prove this theorem even though we have already given a counterexample. The way around this is to make some hypotheses that rule out the bad example.

STOPPING TIMES

Given random variables $X_0, X_1, X_2, \ldots$ and the corresponding filtration, we define a stopping time (with respect to $\mathcal{F}_n$) to be a random (nonnegative integer) time T (the time we stop) with the property that we cannot look into the future before stopping. More formally,

- A stopping time T is a random variable taking values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ such that for each n, the event $\{T = n\}$ is $\mathcal{F}_n$-measurable.

Examples

Suppose $\{M_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$.

- Let A be a set of real numbers and let

$$T = \min\{n : M_n \in A\}.$$

- Let k be a fixed integer, and let $T = k$.
- If T and $\hat{T}$ are stopping times, so are

$$T \vee \hat{T} := \max\{T, \hat{T}\}, \quad T \wedge \hat{T} := \min\{T, \hat{T}\}.$$

  In particular, if k is a fixed integer, $T \wedge k$ is a stopping time.
- In the martingale betting strategy, let T be the first j such that $X_j = 1$, or equivalently, the first j such that $Z_j = 1$. Then T is a stopping time.

OST 1 *If M_n is a martingale w.r.t $\{\mathcal{F}_n\}$ and T is a stopping time, then*

$$N_n = M_{n \wedge T}$$

is a martingale. In particular,

$$\mathbb{E}[N_n] = \mathbb{E}[N_0].$$

Note that we observed this in the martingale betting strategy. After n plays either we had lost n times (with probability 2^{-n} for total loss $2^n - 1$) or we were up one dollar. The mean is

$$1\,(1 - 2^{-n}) - (2^n - 1)\,2^{-n} = 0.$$

To see why OST1 is true, we write

$$
\begin{aligned}
M_{n\wedge T} &= \sum_{j=0}^{n} M_j \, 1\{T \wedge n = j\} \\
&= M_n \, 1\{T \geq n\} + \sum_{j=0}^{n-1} M_j \, 1\{T = j\}
\end{aligned}
$$

So,

$$
M_{(n+1)\wedge T} - M_{n\wedge T} = (M_{n+1} - M_n) \, 1\{T \geq n+1\}.
$$

(In other words, if we allow ourselves one extra time unit before we have to stop, the values will differ only if our stopping rule did not want us to stop before time n.) The important observation is that the event $\{T \geq n+1\}$ is the same as the event $\{T > n\}$ which is measurable with respect to $\mathcal{F}_n$! Therefore,

$$
\begin{aligned}
& E\left[(M_{(n+1)\wedge T} - M_{n\wedge T}) \mid \mathcal{F}_n\right] \\
&= E\left[(M_{n+1} - M_n) \, 1\{T > n\} \mid \mathcal{F}_n\right] \\
&= 1\{T > n\} \, E\left[M_{n+1} - M_n \mid \mathcal{F}_n\right].
\end{aligned}
$$

But $E\left[M_{n+1} - M_n \mid \mathcal{F}_n\right] = 0$ since M_n is a martingale.

Now suppose T is not bounded but that we are guaranteed to stop at some point,

$$\mathbb{P}\{T < \infty\} = 1.$$

We know $\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_0]$ for each n. Also,

$$M_T - M_{n\wedge T} = M_T\, 1\{T > n\} - M_n\, 1\{T > n\}.$$

- If $\mathbb{E}\left[|M_T|\right] < \infty$, the random variable $M_T\, 1\{T > n\} \to 0$ and $\mathbb{E}[M_T\, 1\{T > n\}] \to 0$.
  (In the martingale betting strategy example, $M_T = 1$, and

  $$\mathbb{E}[M_T\, 1\{T > n\}] = \mathbb{P}\{T > n\} = 2^{-n}$$

  so this is not the problem.)

We also need

$$\mathbb{E}\left[M_n\, 1\{T > n\}\right] \to 0.$$

This is the condition that does not hold for martingale betting strategy:

$$\mathbb{E}\left[M_n\, 1\{T > n\}\right] = \left(-2^n + 1\right) 2^{-n} \longrightarrow -1.$$

How do we make sure this does not happen? Mathematicians do it the easy way — they just assume that it does not happen.

OST2 *Suppose M_n is a martingale with respect to $\{\mathcal{F}_n\}$ and T is a stopping time such that*

$$\mathbb{P}\{T < \infty\} = 1,$$

$$\mathbb{E}\left[|M_T|\right] < \infty,$$

$$\lim_{n\to\infty} \mathbb{E}\left[|M_n|\, 1\{T > n\}\right] = 0.$$

Then,

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

Example: Gambler's Ruin Estimate

Simple random walk starting at 1,

$$S_n = 1 + X_1 + \cdots + X_n.$$

$X_1, X_2, \ldots$ independent,

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

Let

$$T = \min\{n : S_n = 0 \text{ or } K\}.$$

Then $M_n = S_{n \wedge T}$ is a bounded martingale. Therefore, it satisfies the condition of the OST and

$$1 = \mathbb{E}[M_0] = \mathbb{E}[M_T].$$

But,

$$\mathbb{E}[M_T] = K\,\mathbb{P}\{S_T = K\} + 0\,[1 - \mathbb{P}\{S_T = K\}].$$

Hence.

$$\mathbb{P}\{S_T = K\} = \frac{1}{K}.$$

The condition

$$\lim_{n\to\infty} \mathbb{E}\left[|M_n|\, 1\{T > n\}\right] = 0$$

can be hard to check. It is true if T is bounded (OST1) or if $\mathbb{P}\{T < \infty\} = 1$ and one of the following holds:

- There exists $K < \infty$ such that $|M_n| \le K$ for all n,
- There exists $K < \infty$ such that $\mathbb{E}[M_n^2] < K$ for each n.
  - To see this, for each $x < \infty$,

$$\mathbb{E}\left[|M_n|\, 1\{T > n\}\right] =$$

$$\mathbb{E}\left[|M_n|\, 1\{T > n, |M_n| \le x\}\right] + \mathbb{E}\left[|M_n|\, 1\{T > n, |M_n| > x\}\right].$$

$$\mathbb{E}\left[|M_n|\, 1\{T > n, |M_n| \le x\}\right] \le |x|\, \mathbb{P}\{T > n\},$$

$$\mathbb{E}\left[|M_n|\, 1\{T > n, |M_n| > x\}\right] \le |x|^{-1}\, \mathbb{E}[M_n^2] \le |x|^{-1}\, K.$$

Example

$$S_n = X_1 + \ldots + X_n$$

simple random walk.

$$T = \min\{n : S_n = K \text{ or } -J\}.$$

The gambler's ruin estimate implies that

$$\mathbb{P}\{S_T = K\} = \frac{J}{J+K}.$$

$M_n = S_n^2 - n$ is a martingale (not uniformly bounded). One can show that there exists N such that $\mathbb{E}[M_{n\wedge T}^2] \leq N$ for each n. Therefore,

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[S_T^2 - T].$$

Therefore,

$$\mathbb{E}[T] = \mathbb{E}[S_T^2] = K^2 \frac{J}{J+K} + J^2 \frac{K}{J+K} = JK.$$

In particular, the expected amount of time to reach distance K from the origin is K^2.

Example

$$S_n = X_1 + \cdots + X_n \quad \text{simple random walk.}$$
$$T = \min\{n : S_n = 1\}$$
$$T_J = \min\{n : S_n = 1 \text{ or } -J\}.$$

Then,

$$\mathbb{P}\{T < \infty\} \geq \mathbb{P}\{S_{T_J} = 1\} = \frac{J}{J+1},$$
$$\mathbb{E}[T] \geq \mathbb{E}[T_J] = J.$$

Therefore,

$$\mathbb{P}\{T < \infty\} = 1, \quad \mathbb{E}[T] = \infty.$$

Also, note that

$$1 = \mathbb{E}[S_T] \neq \mathbb{E}[S_0] = 0.$$

Hence this does *not* satisfy the OST. It must be the case that

$$\mathbb{E}[|S_n| \, 1\{T > n\}] \not\longrightarrow 0.$$

In fact, $\mathbb{P}\{T > n\} \sim c\, n^{-1/2}$ and $|S_n| \sim c'\, n^{1/2}$.

MARTINGALE CONVERGENCE THEOREM

Suppose $M_0, M_1, \ldots$ is a martingale with respect to $\{\mathcal{F}_n\}$ such that there exists $C < \infty$ with $\mathbb{E}[|M_n|] \leq C$ for all n. Then with probability one, the limit

$$M_\infty = \lim_{n \to \infty} M_n$$

exists.

- I will not give the proof of this result, but you may wish to look it up in my book. The proof uses a basic idea in trading — buy when the price is low and sell when the price is high. If the martingale (think, stock price) does not converge, then it fluctuates infinitely often and a trader can take advantage of this to make an unlimited amount of money. This contradicts the assumption $\mathbb{E}[|M_n|] \leq C < \infty$.

Example: Polya's urn

Consider an urn with red and green balls. At time $n = 0$, there is one red and one green ball. At each integer time, a ball is chosen at random from the urn and it and a ball of the same color are returned to the urn. At time n, there are $n+2$ ball of which R_n are red and $G_n = (n+2) - R_n$ are green. The fraction of red balls at time n is

$$M_n = \frac{R_n}{R_n + G_n} = \frac{R_n}{n+2}.$$

Let $\mathcal{F}_n$ denote the filtration generated by $M_0, M_1, \ldots, M_n$.

- M_n is a Markov chain. That is to say, the conditional distribution of M_{n+1} given $\mathcal{F}_n$ is the same as the distribution given M_n. (In general, a process is Markov if the only relevant information from the past for predicting the future is the current value.)

- M_n is a martingale with respect to $\mathcal{F}_n$. Verification:

$$\mathbb{P}\left\{ M_{n+1} = \frac{j+1}{n+3} \mid M_n = \frac{j}{n+2} \right\} = \frac{j}{n+2} = M_n,$$

$$\mathbb{P}\left\{ M_{n+1} = \frac{j}{n+3} \mid M_n = \frac{j}{n+2} \right\} = \frac{n+2-j}{n+2} = 1 - M_n$$

$$\begin{aligned} \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \frac{(n+2)\,M_n + 1}{n+3} M_n + \frac{(n+2)\,M_n}{n+3} [1 - M_n] \\ &= \frac{M_n}{n+3} + \frac{(n+2)\,M_n}{n+3} = M_n \end{aligned}$$

- For every n, the distribution of M_n is uniform on the set

$$\left\{ \frac{1}{n+2}, \frac{2}{n+2}, \cdots, \frac{n+1}{n+2} \right\}.$$

- The condition $\mathbb{E}[|M_n|] \leq C$ is trivial to verify because M_n is positive. Indeed, since M_n is a positive martingale, $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[M_0] = 1/2$.

- Therefore, with probability one, the limit

$$M_\infty = \lim_{n \to \infty} M_n$$

  exists. M_∞ is a random variable — the value is not the same every time we perform the experiment.
- Not surprisingly (given the distribution of M_n), the distribution of M_∞ is uniform on $[0, 1]$.
- This toy model can be interpreted in terms of Bayesian statistics as we now discuss.

- Suppose we perform independent, identically distributed trials of an experiment with probability θ of success each time (Bernoulli trials). Suppose that we do not know θ, but we observe

$$X_1, X_2, \ldots$$

with

$$\mathbb{P}\{X_j = 1\} = 1 - \mathbb{P}\{X_j = 0\} = \theta.$$

Goal of statistics: try to determine θ given the observations. (Mathematically, this is not a precise goal.)

- The law of large numbers tells us that

$$\lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n} = \theta.$$

Therefore, if we can observe an infinite number of data points we can determine θ.

- Of course, in practice we can never sample an infinite number of data points so we do the best that we can with a finite number.

- Bayesian approach: since we do not know θ, we will consider it as a random variable. We start with a prior distribution and with every collection of information $X_1, \ldots, X_n$ we determine a posterior distribution.
- Assume the prior distribution is uniform on $[0, 1]$, i.e., has density $f_0(x) = 1, 0 < x < 1$.
- Let $S_n = X_1 + \cdots + X_n$. Given θ,

$$\mathbb{P}\{S_n = k\} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

- When predicting θ from $\mathcal{F}_n$, the only important quantity is S_n, the number of successes. S_n is called a sufficient statistic for θ.

- Let $f_{n,k}(\theta) = f_n(\theta \mid S_n = k)$ be the posterior density at time n given $S_n = k$. A form of Bayes theorem, gives

$$f_{n,k}(\theta) = \frac{\binom{n}{k}\,\theta^k\,(1-\theta)^{n-k}}{\int_0^1 \binom{n}{k}\,s^k\,(1-s)^{n-k}\,ds} = C_{n,k}\,\theta^k\,(1-\theta)^{n-k}.$$

  This is the *beta* density with parameters $k+1$ and $n-k+1$.

- With a little calculus, we get

$$\mathbb{P}\{X_{n+1} = 1 \mid S_n = k\} = \int_0^1 \theta\, f_{n,k}(\theta)\, d\theta = \frac{k+1}{n+2}.$$

  This is exactly the transition probability for Polya's urn.

- The martingale convergence theorem for Polya's urn can be interpreted as the law of large numbers for θ.

Example

- Let $S_n = X_1 + \ldots + X_n$ denote simple (symmetric) random walk which is a martingale.
- It is not the case that the limit

$$\lim_{n \to \infty} S_n$$

  exists.
- In this case, there is no uniform bound for $\mathbb{E}[|S_n|]$.
- For a martingale M_n, if $\mathbb{E}[|M_n|]$ grows to infinity, we cannot (without further information) conclude that it will converge.