STOCHASTIC PROCESSES II

PART I – MARTINGALES

Conditional expectations

- *Measure theory*
  - In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sigma field $\mathcal{A}$ is a collection of events, each of which as a subset of Ω. It satisfies (i) $\varnothing \in \mathcal{F}$ (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (iii) $A_i \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$. Notes:
    - (i) and (ii) $\Rightarrow \Omega \in \mathcal{F}$
    - $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$, so also closed under infinite (and finite) intersection.
  - A random variable maps $X(\omega) : \Omega \to \mathbb{R}$. When we say X is *measurable with respect to* $\mathcal{F}$ and write $X \in \mathcal{F}$, we mean $\left\{\omega : X(\omega) \le x\right\} \in \mathcal{F}\ \forall x$.
- *Conditional expectations*
  - $\mathbb{E}(X \mid Y)$ is a random variable. $\mathbb{E}(X \mid Y)(\omega) = \mathbb{E}\left(X \mid Y = Y(\omega)\right)$. In other words, the fact $Y = Y(\omega)$ "reveals" a "region" of Ω in which we are located. We then find the expected value of X *given we are in that "region"*.
    - In terms of the definition below, we can write $\mathbb{E}(X \mid Y) = \mathbb{E}(X \mid \sigma(Y))$, where $\sigma(Y)$ is the sigma-field *generated by* Y – in other words, $\sigma(Y) = \left\{\left\{\omega : Y(\omega) \le x\right\} : x \in \mathbb{R}\right\}$ – every event can would be revealed by Y.
  - $W = \mathbb{E}(X \mid \mathcal{A})$ is a random variable. $\mathbb{E}(X \mid \mathcal{A})(\omega)$ is a bit harder to understand – effectively, it takes the expectation of X over the smallest $\mathcal{A}$ that contains ω. In other words, let A be the smallest element of $\mathcal{A}$ that contains ω – then we restrict ourselves to some region of Ω and find the expectation over that region; $\mathbb{E}(X \mid \mathcal{A})(\omega) = \mathbb{E}(X\mathbb{I}_A)$. Formal properties:

- $W \in \mathcal{A}$: information as to where we are in Ω only ever "reaches" us via knowledge of which part of $\mathcal{A}$ we're in, so this is obvious.
- $\mathbb{E}\left(W\mathbb{I}_A\right) = \mathbb{E}\left(X\mathbb{I}_A\right)$ for all $A \in \mathcal{A}$: we are now restricting ourselves to a region of Ω that is $\mathcal{A}$-measurable. Provided A is the smallest element for which $\omega \in A$, $W(\omega) = \mathbb{E}(X\mathbb{I}_A)$, and the result follows trivially. (If it is not the smallest element, the result requires additional thought).

- Some properties
  i. $\mathbb{E}\left[X \mid \mathcal{A}\right]$ if $X \in \mathcal{A}$
  ii. $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{A}\right]\right] = \mathbb{E}(X)$
  iii. $\mathbb{E}\left(XZ \mid \mathcal{A}\right) = Z\mathbb{E}\left(X \mid \mathcal{A}\right)$ if $Z \in \mathcal{A}$
  iv. Tower: $\mathbb{E}\left[\mathbb{E}\left(X \mid \mathcal{B}\right) \mid \mathcal{A}\right] = \mathbb{E}\left(X \mid \mathcal{A}\right)$ if $\mathcal{A} \subseteq \mathcal{B}$: in this case, $\mathcal{B}$ is "more descriptive" than $\mathcal{A}$, so the result makes sense.

     Proof: Use $\mathbb{E}\left(\mathbb{E}\left(\mathbb{E}\left(X \mid \mathcal{B}\right) \mid \mathcal{A}\right)\mathbb{I}_A\right) = \mathbb{E}\left(\mathbb{E}\left(X \mid \mathcal{B}\right)\mathbb{I}_A\right)$ for $A \in \mathcal{A}$. Then use the fact that $A \in \mathcal{B}$ to show this is equal to $\mathbb{E}\left(X\mathbb{I}_A\right)$. ■
  v. Linearity
  vi. Jensen's: for convex *f*, $\mathbb{E}\left[f(X) \mid \mathcal{G}\right] \geq f\left(\mathbb{E}\left[X \mid \mathcal{G}\right]\right)$

- Notes
  - $\mathbb{E}\left[X\right] = \mathbb{E}\left[X \mid \{\varnothing, \Omega\}\right]$ (the RHS is a constant, because whatever ω we choose, the only element of $\{\varnothing, \Omega\}$ that contains it will be Ω). Thus, (ii) is a special case of (iv).
  - Integrability of X implies integrability of $\mathbb{E}\left[X \mid \mathcal{A}\right]$:

$$\mathbb{E}\left[\left|\mathbb{E}(X \mid \mathcal{A})\right|\right] \overbrace{\leq}^{\text{(vi)}} \mathbb{E}\left[\mathbb{E}\left(\left|X\right| \mid \mathcal{A}\right)\right] \overbrace{=}^{\text{(ii)}} \mathbb{E}\left(\left|X\right|\right)$$

- ***Example***: Let Ω be countable. Let $\mathbb{B} = \left\{\mathcal{B}_1, \mathcal{B}_2, \cdots\right\}$ be a partition of Ω, and $\mathcal{A}$ be the set of all subsets of $\mathbb{B}$. Then $\mathbb{E}(X \mid \mathcal{A})$ takes value $\frac{\sum_{\omega \in \mathcal{B}_i} X(\omega)\mathbb{P}(\omega)}{\mathbb{P}(\mathcal{B}_i)}$ with probability $\mathbb{P}(\mathcal{B}_i)$.

Proof: Clearly, the RV is $\mathcal{A}$ measurable, because each value it can take is defined by a $\mathcal{B}_i$. Also, $\mathbb{E}\left[\mathbb{E}(X \mid \mathcal{A})\mathbb{I}_A\right]$ is the expected value over those $\mathcal{B}_i \subseteq A$. Clearly, $= \mathbb{E}\left[X\mathbb{I}_A\right]$. ■ □

Martingales

- ***Definition***: $\left\{X_n\right\}$ is a sub-martingale with respect to $\left\{\mathcal{F}_n\right\}$ (where $\mathcal{F}_n \in \mathcal{F}_{n+1}$) if
  i. $X_n \in \mathcal{F}_n$
  ii. $\mathbb{E}(X_n) < \infty$ [it is often convenient to work with the stronger condition $\mathbb{E}\left|X_n\right| < \infty$].
  iii. $\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \geq X_n$ [$\leq$ gives a super-martingale, $=$ gives a martingale]. Implies the weaker property $\mathbb{E}\left[X_{n+1}\right] \geq \mathbb{E}\left[X_n\right]$
- ***Remarks***:
  - A convex function of a martingale is a submartingale.
  - An increasing convex function of a submartingale is a submartingale.

  Proof: (i) and (ii) are simple. $\mathbb{E}\left[f(X_{n+1}) \mid \mathcal{F}_n\right] \geq f\left(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]\right) \geq f(X_n)$. ■
- ***Example***: Let $S_n = \sum_{i=1}^n X_i$, where the X_i are IID with $\mathbb{E}(X_i) = 0, \mathbb{E}\left|X_i\right| < \infty$
  - S_n is a martingale [$\mathbb{E}\left|S_n\right| \leq n\mathbb{E}\left|X_1\right|$] (the *mean martingale*).
  - If $\mathbb{V}\text{ar}(X_i) = \sigma^2 < \infty$, $X_n^2 - \sigma^2 n$ is a martingale (the *variance martingale*). □
- ***Example (the exponential martingale)***: Let $\varphi(\theta) = \mathbb{E}(e^{\theta X_1})$. $M_n = e^{\theta S_n} / \varphi^n(\theta)$ is a martingale. For example, if $S_n = \sum_{i=1}^n X_i$ is an asymmetric random walk with $p = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = -1)$, then $M_n = \left(\frac{1-p}{p}\right)^{S_n}$ is an exponential martingale, with $e^\theta = \frac{1-p}{p}$ and $\varphi(\theta) = 1$. □
- ***Example***: Suppose an urn starts with one black and one white ball. We pull out balls from the urn, and return them to the urn with *another*, *new* ball of the same color. Y_n, the proportion of white balls after n draws, is a martingale (mean ½). □
- ***Example***: Let $\left\{X_n\right\}$ be a Markov Chain with transition matrix $P(x, y)$ and let $h(x)$ be a bounded function with $h(x) = \sum_y p(x,y)h(y)$. $\left\{h(X_n)\right\}$ is then a martingale. □

Modes of convergence, etc…

- ***Modes of convergence***
  - ***Almost sure***: $\mathbb{P}\left(\lim_n X_n(\omega) = X(\omega)\right) = 1$, or $\mathbb{P}\left(X_n \not\to X \text{ i.o.}\right) = 0$
  - ***In probability***: $\lim_{n\to\infty} \mathbb{P}\left(\left|X_n - X\right| \geq \varepsilon\right) = 0 \quad \forall \varepsilon > 0$
    - Can extract a subsequence $\left\{X_{n_k}\right\}$ that tends to X almost surely. Chose $\mathbb{P}\left(\left|X_{n_{k+1}} - X\right| \geq \frac{1}{k}\right) \leq \frac{1}{2^k}$.
  - ***L_1***: $\lim_{n\to\infty} \mathbb{E}\left|X_n - X\right| = 0$ (assuming $X_n \in L_1$)
    - Implies $\lim_{n\to\infty} \mathbb{E}\left|X_n\right| = \mathbb{E}\left|X\right|$, since $\pm\left(\left|X_n\right| - \left|X\right|\right) \leq \left|X_n - X\right|$ (derive by writing $a = a - b + b$ and using triangle inequality).
    - Implies $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$, since $\left|\mathbb{E}(X_n - X)\right| \leq \mathbb{E}\left|X_n - X\right|$, by Jensen.
    - Implies $X_n \to_p X$ by Markov's.
- ***Interchange arguments***
  - Concerned with whether $X_n \to X$ a.s. $\Rightarrow \lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.
  - Holds if
    - ***Bounded convergence***: $\left|X_n\right| \leq c$, c is constant.

      Proof: Define $B_n^c = A_n = \left\{\left|X_n - X\right| < \varepsilon\right\}$. Write $\left|\mathbb{E}(X_n - X)\right|$ as an integral and split over A_n and B_n. Use the fact that $\mathbb{P}(B_n) \to 0$. [Can replace with convergence in probability; $\left|X_n\right| \leq c$ still holds for subsequence]. ■
    - ***Monotone Convergence***: $0 \leq X_n \uparrow X$ almost surely

      Proof: $X_n \leq X \Rightarrow \mathbb{E}(X_n) \leq \mathbb{E}(X) \Rightarrow \limsup_n \mathbb{E}(X_n) \leq \mathbb{E}(X)$. Together with Fatou, gives our result. ■
    - ***Dominated convergence***: $\left|X_n\right| \leq Y$, Y integrable.

      Proof: Condition implies $Y \pm X_n \geq 0$. Apply Fatou to both, subtract $\mathbb{E}(Y) < \infty$ from both sides. ■
  - ***Fatou***: If $X_n \geq 0$, $\mathbb{E}\left(\liminf_n X_n\right) \leq \liminf_n \mathbb{E}(X_n)$

Proof: Let $Y_m = \inf_{n>m} X_n$, and note that $\inf_{n>m} \mathbb{E}(X_n) \geq \mathbb{E}(Y_m)$. Replace Y_m with $\min[Y_m, k]$ – bounded because X_n's positive. Use bounded convergence and let $k \to \infty$. ■

- ***Uniform integrability***
  - ***Definition***: $\{X_n\}$ is uniformly integrable if

$$\lim_{a\to\infty} \sup_n \mathbb{E}\left[\left|X_n\right| \mathbb{I}_{|X_n|>a}\right] = 0$$
$$\Leftrightarrow \forall \varepsilon > 0, \exists a_0 \text{ s.t. } \mathbb{E}\left[\left|X_n\right| \mathbb{I}_{|X_n|>a}\right] \leq \varepsilon \quad \forall n, a \geq a_0$$

  - ***Lemma***: $\{X_n \text{ u.i.}\} \Rightarrow \sup_n \mathbb{E}\left|X_n\right| \leq K$

    Proof: Write $\mathbb{E}\left|X_n\right| = \mathbb{E}\left(\left|X_n\right| \mathbb{I}_{|X_n|\leq a}\right) + \mathbb{E}\left(\left|X_n\right| \mathbb{I}_{|X_n|>a}\right) \leq a + \varepsilon$. ■

  - ***Lemma***: $\mathbb{E}\left[\sup_n \left|X_n\right|\right] < \infty \Rightarrow \{X_n\}$ u.i.

    Proof: Let $Y = \sup_n \left|X_n\right|$. $\mathbb{E}[Y] = \mathbb{E}[Y\mathbb{I}_{Y\leq a}] + \mathbb{E}[Y\mathbb{I}_{Y\geq a}]$. By monotone convergence, first term approaches $\mathbb{E}(Y)$, second gets arbitrary close to 0, and $\left|X_n\right| \leq Y$, so u.i. follows. ■

  - ***Lemma***: $\exists \delta > 0$ s.t. $\mathbb{E}\left|X_n\right|^{1+\delta} \leq K < \infty \Rightarrow \{X_n\}$ u.i.

    Proof: $\mathbb{E}\left(\left|X_n\right| \mathbb{I}_{|X_n|>a}\right) \leq \mathbb{E}\left(\left|X_n\right| \left[\frac{|X_n|}{a}\right]^\delta \mathbb{I}_{|X_n|>a}\right) = \frac{1}{a^\delta} \mathbb{E}\left(\left|X_n\right|^{1+\delta} \mathbb{I}_{|X_n|>a}\right) \leq \frac{K}{a^\delta}$. By splitting into intervals $\left|X_n\right| \leq 1$ and $\left|X_n\right| > 1$, can also show that for any $0 < \delta' \leq \delta$, also integrable. ■

  - ***Theorem***: $\{X_n\}$ u.i. & $X_n \to_p X \Rightarrow X_n \to_{L_1} X$

    Proof: First, show X is integrable by writing $\mathbb{E}\left[\left|X\right|\right] = \mathbb{E}\left[\liminf_n \left|X_n\right|\right]$, using Fatou's, writing lim inf $\leq$ sup and using the definition of u.i [use subsequences for convergence in p]. Then, set $Y_n = \left|X_n - X\right| \leq \left|X_n\right| + \left|X\right|$ – it is u.i. because X integrable. Split $\mathbb{E}(Y_n)$ into $Y_n > a$ and $Y_n \leq a$. First one can be made arbitrarily small by u.i. Bounded convergence applies to the second one, and since $Y_n \to 0$, it can also be made arbitrarily small by letting $n \to \infty$. ■

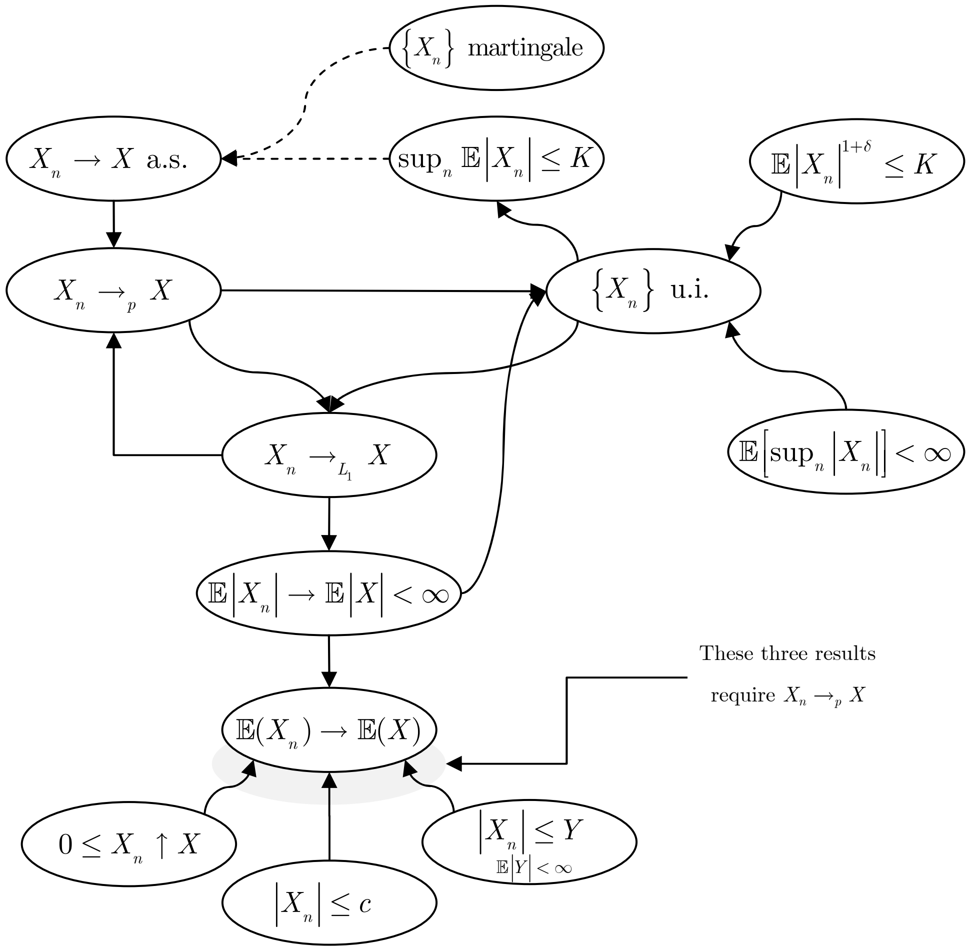
  - ***Theorem***: $X_n \to_p X$ & $\mathbb{E}\left|X_n\right| \to \mathbb{E}\left|X\right| < \infty \Rightarrow \{X_n\}$ u.i.

Proof: Define $f_a(x)$ as a function that links (0, 0), (a − 1, a − 1) and (a, 0). Clearly $x\mathbb{I}_{x\le a-1} \le f_a(x) \le x\mathbb{I}_{x\le a}$. Write $\mathbb{E}\left(\left|X_n\right|\mathbb{I}_{\left|X_n\right|>a}\right) = \mathbb{E}\left|X_n\right| - \mathbb{E}\left[\left|X_n\right|\mathbb{I}_{\left|X_n\right|\le a}\right]$ $\le \mathbb{E}\left|X_n\right| - \mathbb{E}\left[f_a(\left|X_n\right|)\right]$. By L_1 convergence (first term) and monotone convergence (second term), this can be made arbitrarily close to $\mathbb{E}\left|X\right| - \mathbb{E}\left[f_a\left|X\right|\right]$ $\le \mathbb{E}\left|X\right| - \mathbb{E}\left[\left|X\right|\mathbb{I}_{\left|X\right|\le a-1}\right]$. We can make this arbitrarily small since $\mathbb{E}\left|X\right| < \infty$. ■

- It can also be shown that $\left\{X_n\right\}$ u.i. if and only if $\mathbb{P}(A) < \delta \Rightarrow \mathbb{E}\left[\left|X_n\right|\mathbb{I}_A\right] < \varepsilon$ and $\sup_n \mathbb{E}\left|X_n\right| < \infty$. For the only if part, split $\mathbb{E}\left[\left|X_n\right|\mathbb{I}_A\right]$ into $\left|X_n\right| > a$ and $\left|X_n\right| \le a$. For the if part, use Markov's Inequality $\sup_{n\ge 1} \mathbb{P}\left(\left|X_n\right| > a\right) \to 0$, and then use the assumption.

- ***Summary***

(Note: when arrows enter a given bubble at the same point, *both* are required for the bubble to hold. When arrows enter at different points, each arrow independently is sufficient to prove the bubble. The result pertaining to the convergence of martingales (in dotted lines) will be discussed later.)



Optional stopping

- ***Definition (stopping time)***: If T is an integer valued random variable, we say it is a *stopping time* with respect to a filtration $\mathcal{F}_n$ if $\{T = n\} \in \mathcal{F}_n$ for all n (or $\{T \le t\} \in \mathcal{F}_t$ for all t, in continuous time).

  Remark: If T_1 and T_2 are stopping times, so are $T_1 + T_2, T_1 \wedge T_2, T_1 \vee T_2$.
- ***Theorem***: If $\{X_n\}$ is a (sub)martingale and T is a stopping time, $\{X_{T\wedge n}\}$ is also a (sub)martingale. If $\{X_n\}$ is u.i., so it $\{X_{T\wedge n}\}$.

  Proof: Write $X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbb{I}_{\{T=k\}} + X_n \mathbb{I}_{\{T>n-1\}}$. Clearly, this is $\mathcal{F}_n$ measurable, and $|X_{T\wedge n}| \le \sum_{k=0}^{n-1} |X_k| + |X_n|$ so it is integrable. Conditioning follows. $\boxed{u.i.}$: to show u.i., first note that $\{X^+_{T\wedge n}\}$ is also a submartingale, and so $\mathbb{E}(X^+_{T\wedge n}) \le \mathbb{E}|X^+_n| \le \mathbb{E}|X_n|$. Taking limits, $\sup_n \mathbb{E}(X^+_{T\wedge n}) \le \sup_n \mathbb{E}|X_n| < \infty$ [the last inequality follows by u.i. of $\{X_n\}$]. By the theorem in the martingale convergence section, $X^+_{T\wedge n} \to_{n\to\infty} X_T$ with $\mathbb{E}\left|X_T\right| < \infty$. Finally, consider $\mathbb{E}\left[\left|X_{T\wedge n}\right| \mathbb{I}_{X_{T\wedge n} > a}\right]$. Simply split it over $\mathbb{I}_{T\le n}$ and $\mathbb{I}_{T>n}$, drop these indicators and use integrability of X_T and u.i. of $\left\{X_n\right\}$. ■

- ***Example***: Consider a gambler's ruin with wealth S_t at time t with $S_0 = i$ and with probably ½ of going each direction at each time step. Let

$$p = \mathbb{P}\left(\text{Probability we hit } N > i, \text{ at which point we stop}\right)$$
$$1 - p = \mathbb{P}\left(\text{Probability we hit 0, at which point we're ruined}\right)$$

  And let

$$T = \inf\left\{n : S_n = N \text{ or } S_n = 0\right\}$$

  We can now use the OST

  - $\boxed{\text{On } S_t}$ OST says that $\mathbb{E}(S_T) = \mathbb{E}(S_0) = i$. Logic says that $\mathbb{E}(S_T) = pN + 0$. Together, we obtain $p = i / N$.
  - $\boxed{\text{On } S_t^2 - n}$ OST says that $\mathbb{E}(S_T^2 - T) = \mathbb{E}(S_0^2) = i^2$, and so $\mathbb{E}(T) = \mathbb{E}(S_T^2) - i^2$. Logic says that $\mathbb{E}(S_T^2) = pN^2 + 0$. Together, $\mathbb{E}(T) = i\left(N - i\right)$. □

- ***Counterexample***: Consider the example above, but with $T' = \inf\left\{n : S_n = N > i\right\}$. This is well defined, in that $\mathbb{P}(T' < \infty) = 1$ (because the random walk is an irreducible Markov chain, which means every state will eventually be visited) but blindly applying the OST gives $\mathbb{E}(S_{T'}) = i$, which implies that $N = i$. Clearly, something has gone awry.□
- We need to develop conditions under which the OST works. One such condition is…
- …***Theorem***: If $T \le n_0$ a.s. then $\mathbb{E}[X_T] = \mathbb{E}[X_{n_0}]$.

  Proof: $\mathbb{E}[X_T] = \mathbb{E}[X_{T\wedge n_0}] = \mathbb{E}[X_0]$ ■

  Remark: This condition is not satisfied in the counterexample above because even though $\mathbb{P}(T' < \infty) = 1$, it is *not* bounded – following the MC analogy, $\mathbb{E}(T') = \infty$.
- More generally, there are three key "ingredients" to optional stopping

$$\mathbb{E}[X_0] \overbrace{=}^{(1)} \lim_n \mathbb{E}[X_{T\wedge n}] \overbrace{=}^{(2)} \mathbb{E}[\lim_n X_{T\wedge n}] \overbrace{=}^{(3)} \mathbb{E}[X_T]$$

  Notes:

  1. This equality holds for every n (since $X_{T\wedge n}$ is also a martingale) and so it still holds after we take t he limit.
  2. This equality requires an interchange argument – uniform integrability of $\left\{X_{T\wedge n}\right\}$ will achieve this, for example.

3. This equality requires the martingale to converge. We will see later that uniform integrability of $\left\{X_{T\wedge n}\right\}$ also implies convergence of this martingale.

- ***Example***: Let us return to the example above, and justify applying the OST in retrospect:
  - In the first case, $\left|S_{T\wedge n}\right| \le a \vee b \ \forall n$, and so $\sup_n \left|S_{T\wedge n}\right| \le a \vee b$. As such, $\left\{S_{T\wedge n}\right\}$ is u.i.
  - In the second case, $\left|S_{n\wedge T}^2 - (n \wedge t)\right| \le a^2 \vee b^2 - T$, so provided $\mathbb{E}(T) < \infty$, the martingale is u.i. We can see this is true in one of two ways (1) symmetric random walk limited to an interval forms an irreducible Markov chain with finite state space. (2) $\mathbb{E}[S_{T\wedge n_0}^2 - (T \wedge n_0)] = \mathbb{E}[S_0^2 - 0] = 0 \Rightarrow \mathbb{E}[S_{T\wedge n_0}^2] = \mathbb{E}\left(T \wedge n_0\right)$. Taking limits as $n \to \infty$ on both sides ($S_{T\wedge n_0}^2$ is bounded, and $T \wedge n_0$ is monotone increasing) so $\mathbb{E}(S_T^2) = \mathbb{E}(T)$. Since S_T is two-valued and has finite expectation, T has finite expectation. ☐
- Though uniform inetgrability is enough to ensure optional stopping, some weaker conditions are sometimes sufficient…
- ***Theorem***: $\mathbb{E}\left[\left|X_{n+1} - X_n\right| \mid \mathcal{F}_n\right] \le C < \infty$ for $n \le T$ and $\mathbb{E}(T) < \infty$ is enough.
- ***Example (Wald's Identity)***: Let $\left\{X_i\right\}$ be IID with $\mathbb{E}\left|X_i\right| < \infty$ and $\mathbb{E}(X_i) = \mu$. Let T be a stopping time with $\mathbb{E}(T) < \infty$. Then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$.

  Proof: Consider $\left\{S_n - n\mu\right\}$. Applying the OST gives the required result, given that $\mathbb{E}\left(\left|S_n - S_{n-1}\right| \mid \mathcal{F}_n\right) = \mathbb{E}\left|X_n\right| < \infty$. ■ ☐
- ***Example***: Consider an asymmetric random walk with $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = -1)$, and let $T = \inf\left\{n : S_n = -a \text{ or } b\right\}$. We would like to find $p_b = \mathbb{P}(S_T = b)$. (1) Try the exponential martingale ρ^{S_n} (where $\rho = \frac{1-p}{p}$). Note that $0 \le M_{n\wedge T} = \rho^{S_{n\wedge T}} \le \rho^{a+b} \vee 1$; it is therefore bounded, and we can apply the OST to deduce that $1 = \mathbb{E}(\rho^0) = \mathbb{E}(\rho^T)$. Also, $\mathbb{E}(\rho^T) = \rho^b p_b + \rho^a (1 - p_b)$. This allows us to find p_b. (2) Now use the mean martingale $\left\{S_n - n(p-q)\right\}$. OST gives $\mathbb{E}[S_T - T(p-q)] = 0 \Rightarrow \mathbb{E}(S_T) = (p-q)\mathbb{E}(T)$. Using p_b, we can work out $\mathbb{E}(S_T)$ and find $\mathbb{E}(T)$.
- Note that the OST does not necessarily require $\mathbb{P}(T < \infty) = 1$. Indeed, $\mathbb{E}(X_T) = \mathbb{E}\left(X_T \mathbb{I}_{T<\infty} + X_\infty \mathbb{I}_{T=\infty}\right)$, and if the stopped martingale is u.i., X_∞ must exist.

- It is important to remember that when we invoke the martingale convergence theorem so say that $X_{T\wedge n} \to_{n\to\infty} X_T$, we are implicitly implying that $\mathbb{P}(T<\infty)=1$.

Martingale Inequalities

- ***Doob's Inequality/The Maximal Inequality***
  - ***Motivation***: Markov's Inequality states that $a\mathbb{P}\left(\left|X\right|>a\right)\le \mathbb{E}\left|X\right|$.
  - ***Theorem***: If $\left\{X_n\right\}$ is a submartingale and $A=\left\{\max_{0\le k\le n} X_k \ge a\right\}$, then

    $$a\mathbb{P}(A)\le \mathbb{E}[X_n\mathbb{I}_A]\le \mathbb{E}[X_n^+]$$

    Proof: Define $T=\inf\left\{k : X_k \ge a \text{ or } k\ge n\right\}$. Clearly, it's a stopping time, and since $T\le n$, $\mathbb{E}[X_T]=\mathbb{E}[X_n]$. Write both sides of equation by splitting over A and A^c, and note $X_T=X_n$ over A^c. Finally, note $a\mathbb{P}(a)\le \mathbb{E}[X_T\mathbb{I}_A]$. Finally, note that $X_n\le X_n^+ \Rightarrow \mathbb{E}[X_n\mathbb{I}_A]\le \mathbb{E}[X_n^+]$. ■
  - If we let $S_0=0$ and $S_n=\sum_{i=1}^n X_n$ were the X_n are IID with $\mathbb{E}[X_n]=0, \mathbb{V}\text{ar}[X_n]=\sigma^2$, then $\left\{S_n\right\}$ is a martingale, and $\left\{K_n\right\}=\left\{S_n^2\right\}$ is a submartingale. As such, we get *Kolmogorov's Inequality*:

    $$\mathbb{P}\left(\max_{0\le k\le n}\left|S_k\right|\ge x\right)=\mathbb{P}\left(\max_{0\le k\le n} X_k\ge x^2\right)\le \frac{\mathbb{E}[S_n^2]}{x}=\frac{n\sigma^2}{x^2}$$

- ***Azuma's Inequality***
  - ***Theorem***: $\left(x^{-1}-x^{-3}\right)\varphi(x)\le \overline{\Phi}(x)\le x^{-1}\varphi(x)$ for all $x>0$ (where φ is the normal density function).

    Proof: Notice that $\overline{\Phi}(x)\propto \int_x^\infty e^{-y^2/2}\,\mathrm{d}y$. For the upper bound, multiply the integrand by $\left(1+y^{-2}\right)$. For the lower bound, multiply the integral by $\left(1-3y^{-4}\right)$. Integrating gives the required result. ■
  - ***Motivation***: Let $S_0=0$ and $S_n=X_1+\cdots+X_n$, with X_i IID $\mu,\sigma^2<\infty$. Define $Y_n=\frac{S_n}{n}-\mu$. We then have, by the Central Limit Theorem

    $$\mathbb{P}\left(\left|Y_n\right|>\varepsilon\right)\approx \mathbb{P}\left(\left|Z\right|>\tfrac{\sqrt{n}}{\sigma}\varepsilon\right)=2\overline{\Phi}\left(\tfrac{\sqrt{n}}{\sigma}\varepsilon\right)$$

Applying the result above gives an exponential approximation. (It is, by the way, summable, so we automatically recover the SLLN).

- ***Theorem (Azuma's Inequality)***: Let $\{Z_n\}$ be a zero-mean martingale with bounded MG differences (ie: $-\alpha \le Z_i - Z_{i-1} \le \beta$ for $\alpha, \beta \ge 0$). Then
$$\mathbb{P}\left(\bigcup_{n=m}^{\infty} \left|Z_n\right| > n\varepsilon\right) \le 2\exp\left(-\frac{2m\varepsilon^2}{(\alpha+\beta)^2}\right)$$
This bound is not as tight as the CLT's, but it requires less.

- ***Example***: S_n = number of heads in n flips, where $\mathbb{P}(\text{Heads}) = p$. $Z_n = S_n - np$ is a martingale with $-p \le Z_i - Z_{i-1} \le 1 - p$. As such, we can use Azuma's inequality and obtain $\mathbb{P}\left(\bigcup_{n=m}^{\infty} \left|\frac{S_n}{n} - p\right| > \varepsilon\right) \le 2\exp\left(-2m\varepsilon^2\right)$. □

- ***Definition (Doob Martingale)***: Let X be a random variable in L_1 and $\mathcal{F}_n$ be a set of filtrations. Then $X_n = \mathbb{E}\left(X \mid \mathcal{F}_n\right)$ is a martingale.

  Proof: $\mathbb{E}\left(X_{n+1} \mid \mathcal{F}_n\right) = \mathbb{E}\left[\mathbb{E}\left(X \mid \mathcal{F}_{m+1}\right) \mid \mathcal{F}_n\right] = \mathbb{E}\left(X \mid \mathcal{F}_n\right) = X_n$. ■

- Let $\boldsymbol{X} = (X_1, \cdots, X_n)$, where the X_i are independent and with CDF F_i. Define $\mathcal{F}_i = \sigma(X_1, \cdots, X_i)$. Finally, let $h : \mathbb{R}^n \to \mathbb{R}$ such that, if $\boldsymbol{x}$ differs from $\boldsymbol{y}$ in only one component, $\left|h(\boldsymbol{x}) - h(\boldsymbol{y})\right| \le \ell$, for some $\ell \ge 0$. Then $S_i = \mathbb{E}\left[h(\boldsymbol{X}) \mid \mathcal{F}_i\right]$ is a Doob martingale. Provided we can prove $\left|S_i - S_{i-1}\right| \le \ell$, we can apply Azuma's Inequality with $\alpha + \beta = \ell$ to $S_n = h(\boldsymbol{X})$
$$\mathbb{P}\left(\left|h(\boldsymbol{X}) - \mathbb{E}\left[h(\boldsymbol{X})\right]\right| > n\varepsilon\right) \le 2\exp\left(-\frac{n\varepsilon^2}{2\ell^2}\right)$$
  Proof: To prove $\left|S_i - S_{i-1}\right| \le \ell$, note that
$$\begin{aligned} S_i &= \mathbb{E}\left[h(\boldsymbol{X}) \mid \mathcal{F}_i\right] = \int_{x_{i+1}} \cdots \int_{x_n} h\left(X_1, \cdots, X_i, x_{i+1}, \cdots, x_n\right) \mathrm{d}F_n(x_n) \cdots \mathrm{d}F_{i+1}(x_{i+1}) \\ S_{i-1} &= \int_{x_i} \cdots \int_{x_n} h\left(X_1, \cdots, x_i, x_{i+1}, \cdots, x_n\right) \mathrm{d}F_n(x_n) \cdots \mathrm{d}F_{i+1}(x_{i+1}) \mathrm{d}F_i(x_i) \end{aligned}$$
  As such, remembering that densities integrate to 1
$$\begin{aligned} \left|S_i - S_{i-1}\right| &\le \int \cdots \int \left| \quad \right| \, \mathrm{d}\cdots \\ &\le \ell \int \cdots \int 1 \, \mathrm{d}\cdots \\ &= \ell \end{aligned}$$
  As required. ■

- ***Example***: Consider a system of n components (indexed by i) and m experiments (indexed by j). Let $X_i = 1$ if component i works (probability p_i) and 0 otherwise. For experiment j to work, all the components in the set A_j must work, and we assume each component is involved in at most 3 experiments. Let Y be the number of experiments that can be performed. We then have

$$Y = \sum_{j=1}^m \mathbb{I}_{\{\text{Experiment } j \text{ can be performed}\}} = h(\boldsymbol{X}) \qquad \mathbb{E}(Y) = \sum_{j=1}^m \prod_{i \in A_j} p_i$$

Since $\left|h(\boldsymbol{X}) - h(\boldsymbol{Y})\right| \leq 3$, we can apply Azuma's. □

The Upcrossing Inequality

- ***Definition***: A process $\{C_n\}$ is *predictable* with respect to $\{\mathcal{F}_n\}$ if $C_n \leq \mathcal{F}_{n-1}$ for all $n \geq 1$. This is stronger than measurability; it requires C_n to be predictable from the *previous* step's information.
- ***Definition (Martingale Transformation)***: Let $\{X_n\}$ be a submartingale and $\{C_n\}$ be a predictable sequence. We write

$$\left(C \circ X\right)_n = \sum_{k=1}^n C_k \left(X_k - X_{k-1}\right) \qquad \left(C \circ X\right)_0 = 0$$

Intuition: X_n may be the state of a market and C_n the action we take at each step. The theorem states that provided we can't see the future (predictability) and bet infinite amount (boundedness), the game stays fair (martingale).

Proof: Measurability is obvious, and integrability follows from the boundedness of C_n. Then $\mathbb{E}\left(Y_{n+1} \mid \mathcal{F}_n\right) = Y_n + \mathbb{E}\left[C_{n+1}\left(X_{n+1} - X_n\right) \mid \mathcal{F}_n\right]$. Using predictability and the submartingale property leads to the result. [Note: the result also applies to supermartingales and martingales. For the latter, predictability is not required, since the expectation is 0 anyway.] ■

- ***Example***: We can write $\{X_{\tau \wedge n}\}$ as a martingale transformation with $C_n = \mathbb{I}_{\{T \geq n\}} = 1 - \mathbb{I}_{\{T \leq n-1\}} \in \mathcal{F}_{n-1}$. Similarly, we can write $\{X_n - X_{T \wedge n}\}$ as a transformation with $\overline{C}_n = 1 - C_n$. □
- The upcrossing inequality is concerned with transitions of the martingale from a or below to b or above. We define $T_0 = 0$ and

$$T_{2k-1} = \inf\left\{m > T_{2k-2} : X_m \le a\right\}$$
$$T_{2k} = \inf\left\{m > T_{2k-1} : X_m \ge b\right\}$$

Odd stopping times mark the first time the martingale downcrosses a since its last upcrossing of b. Even stopping times are the other way round. These are clearly stopping times.

- Consider a gambling strategy that sets $C_m = 1$ if $T_{2k-1} < m \le T_{2k}$ and $C_m = 0$ otherwise. We effectively "buy" only if we're below a and sell otherwise. Every time there is an upcrossing, a profit of at least $b - a$ is realized.
- ***Theorem***: Let $U_n = \sup\left\{k : T_{2k} \le n\right\}$ – this is the number of upcrossings up to time n – and assume X_n is a submartingale. Then
$$(b-a)\mathbb{E}\left[U_n\right] \le \mathbb{E}\left[(X_n - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right]$$
(This inequality would also apply for a martingale).

  Remark: This is hardly very encouraging. The RHS is the amount that would have been earned by buying at the start and getting out at n.

  Proof: WLOG, shift the upcrossing range to $[0, b-a]$, and define $Y_n = (X_n - a)^+$ – this is also a submartingale and has the same number of upcrossings. Define C_m as above, and since we earn at least b – a at every upcrossing, $(b-a)U_n \le (C \circ Y)_n \Rightarrow (b-a)\mathbb{E}[U_n] \le \mathbb{E}[(C\circ Y)_n]$. Letting $\bar{C}_n = 1 - C_n$, we get $(C\circ Y)_n + (\bar{C}\circ Y)_n = \sum_{k=1}^n (Y_k - Y_{k-1}) = Y_n - Y_0$. Since $(\bar{C}\circ Y)_n$ is also a martingale, $\mathbb{E}\left[(\bar{C}\circ Y)_n\right] \ge \mathbb{E}\left[(\bar{C}\circ Y)_0\right] = 0$. Thus, $\mathbb{E}[(C\circ Y)_n] \le \mathbb{E}[Y_n - Y_0]$. ■

Martingale Convergence

- ***Theorem***: If $\left\{X_n\right\}$ is a submartingale and $\sup_n \mathbb{E}\left|X_n\right| < \infty$ (this is a weaker condition than u.i.), then $X_n \to X$ a.s. and $\mathbb{E}\left|X\right| < \infty$.

  Remark: $\sup_n \mathbb{E}\left|X_n\right| < \infty$ is equivalent to $\sup_n \mathbb{E}(X_n^+) < \infty$ because $x^+ \le \left|x\right| = 2x^+ - x$, and so, for example, $\mathbb{E}\left|X_n\right| = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \le 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0]$.

  Proof: Note that $(X_n - a)^+ \le X_n^+ + \left|a\right|$ and write the upcrossing inequality as $\mathbb{E}\left[U_n\right] \le \frac{\mathbb{E}(X_n^+) + \left|a\right|}{b-a} = K < \infty$ [using the fact $|X|$ dominates X^+ and the condition in the theorem). However, U_n is increasing by definition and therefore tends to some U (possibly ∞), but by monotone convergence, $\mathbb{E}(U_n) \to \mathbb{E}(U) < \infty$. Thus, the number of up-crossings must be finite, and so $\mathbb{P}(\liminf X_n < a < b < \limsup X_n) = 0$. This is true for *any* a and b, and so $\mathbb{P}(\liminf X_n = \limsup X_n) = 1$. Integrability of the limit follows by Fatou. ■

- ***Corollary***: If $\{X_n\}$ is a supermartingale and $X_n \geq 0$, then $X_n \to X$ and $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.
- ***Proof***: Let $Y_n = -X_n$ – this is a submartingale with $\mathbb{E}[Y_n^+] = 0$. The condition of the theorem above (see the remark) is therefore satisfied. ■
- Note, however, that almost sure converge *does not* imply convergence in the means or variances, as the next two examples illustrate..
- ***Example***: Assume $X_0 = i > 0$, and $X_n \mid X_{n-1} \sim \text{Po}(X_{n-1})$. Clearly, this is a martingale, and once we hit 0, we stay there. Let $T = \inf\{n : X_n = 0 \text{ or } X_n \geq b\}$. By optional stopping, $\mathbb{E}(X_T) = \tilde{b}(1-p)$, where $\tilde{b} \geq b$ and $p = \mathbb{P}(X_T = 0)$. But $\mathbb{E}(X_T) = i$. As such, $1 - p \to \frac{i}{b} \to 0$ as $b \to \infty$. Thus, $p \to 1$. How do we know stopping time is finite? Note, however, that $\mathbb{E}[X_n^2] = \mathbb{E}[\mathbb{E}(X_n^2 \mid X_{n-1})] = \mathbb{E}(X_{n-1}^2) + i$. As such, $\mathbb{V}\text{ar}(X_n) = ni$; the variable itself tends to 0, but the variance blows up. □
- ***Example***: Let $X_1 \sim U[0,1]$, and $X_n \mid X_{n-1} \sim U[0, X_{n-1}]$. Let $Y_n = 2^n X_n$. We can write $X_n = U_1 \cdots U_n$ with each U IID U[0,1]. This is a martingale, and, by the SLLN, $\frac{1}{n}\log Y_n = \log 2 + \frac{1}{n}\sum_{i=1}^n \log U_i \to \log 2 + \mathbb{E}(\log U_1) < \infty$ a.s. So $Y_n \to 0$. Note, however, that $\mathbb{V}\text{ar}(X_n) = \left(\frac{4}{3}\right)^n - 1$. Again, the variance blows up. □
- Coupled with uniform integrability, however…
- …***Theorem***: When $\{X_n\}$ is a martingale, the following are equivalent
    i. $\{X_n\}$ is u.i. (and therefore converges almost surely)
    ii. $X_n \to_{L_1} X$
    iii. X_n can be written as a Doob martingale; $X_n = \mathbb{E}\left[X \mid \mathcal{F}_n\right]$, with $\mathbb{E}\left|X\right| < \infty$.

  For submartingales, only (i) and (ii) are equivalent.

  Proof: $\boxed{(i) \Rightarrow (ii)}$ follows from the fact u.i. implies the boundedness condition in the convergence theorem. $\boxed{(ii) \Rightarrow (iii)}$: Let X be the L_1 limit of $\{X_n\}$; clearly, $\mathbb{E}\left[X \mid \mathcal{F}_n\right]$ is a Doob martingale. To show this is equal to X_n, all we need to show is that $\mathbb{E}[X_n \mathbb{I}_A] = \mathbb{E}[X \mathbb{I}_A]$ for any $A \in \mathcal{F}_n$. Two steps. (1) Use the martingale property to write, for any $m > n$, $\mathbb{E}[X_n \mathbb{I}_A] = \mathbb{E}[\mathbb{E}[X_m \mid \mathcal{F}_n]\mathbb{I}_A] = \mathbb{E}[\mathbb{E}[X_m \mathbb{I}_A \mid \mathcal{F}_n]] = \mathbb{E}[X_m \mathbb{I}_A]$. (2) Notice that as $m \to \infty$, $\left|\mathbb{E}[X_m \mathbb{I}_A] - \mathbb{E}[X\mathbb{I}_A]\right| \le \mathbb{E}\left[\left|X_m - X\right|\mathbb{I}_A\right] \to 0$, by L_1 convergence. Letting $m \to \infty$ in step 1 therefore gives the required result. $\boxed{(iii) \Rightarrow (i)}$ Note that by Jensen's, $\left|\mathbb{E}[X \mid \mathcal{F}_n]\right| \le \mathbb{E}\left[\left|X\right| \mid \mathcal{F}_n\right]$, and so we can, WLOG, assume $X \geq 0$. Want to show that for any ε, there exists a_0 such that $\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{F}_n]\mathbb{I}_{\mathbb{E}[X|\mathcal{F}_n]>a}\right] < \varepsilon$. The indicator is $\mathcal{F}_n$ measurable, and so this is equivalent to $\mathbb{E}[X\mathbb{I}_{\mathbb{E}[X|\mathcal{F}_n]>a}] < \varepsilon$. Integrability of X means that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for any A with $\mathbb{P}(A) < \delta$, $\mathbb{E}(X\mathbb{I}_A) \le \varepsilon$. As sch, all we need to show is that $\mathbb{P}(A) = \mathbb{P}(\mathbb{E}[X \mid \mathcal{F}_n] > a) \le \delta$. Applying Markov's, we get $\mathbb{P}(A) \le \mathbb{E}[X] / a$.

- ***Example (Polya's Urn)***: Consider the example of Polya's Urn, discussed above. Let X_n be the proportion of red balls in the urn after draw n. By symmetry, $\mathbb{E}(X_n) = \frac{1}{2}$.

Part II – Brownian Motion

Introduction

- A ***brownian motion*** $\{B(t)\}$, sometimes denoted $\{B_t\}$, is a continuous process with independent and stationary increments which follow a normal distribution; in particular, $B_t \sim N(0,t)$
- A useful way to motivate BM is using a random walk $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, with the X_n IID equally likely to be 1 and –1. Consider, then, a new random walk $S(t)$ which moves every $\Delta t = \frac{t}{n}$ time units (instead of every 1 time unit), and moves by a quantity Δx instead of $\pm$ 1. We then have

$$S(t) = \Delta x\left(X_1 + \cdots + X_n\right)$$

$$\mathbb{E}\left[S(t)\right] = 0 \qquad \mathbb{V}\text{ar}\left[S(t)\right] = (\Delta x)^2 \cdot n \cdot \mathbb{V}\text{ar}(X_1) = \frac{(\Delta x)^2}{\Delta t} t$$

  As $\Delta t \to 0$, this approaches Brownian motion, provided that $(\Delta x)^2 \sim \Delta t$.
- ***Distributional Properties***
  - For notational convenience, we write $\mathbb{P}(B_t = x) = f_t(x)$.
  - Since $B(t) \sim N(0,t)$, its PDF is $f_t(x) = \frac{1}{\sqrt{2\pi t}}\exp(-\frac{x^2}{2t}) = \frac{1}{\sqrt{t}}\varphi(\frac{x}{\sqrt{t}})$.

- ***Covariance***: Consider $s < t$. Then

$$\mathbb{C}\text{ov}\left[B(s), B(t)\right] = \mathbb{C}\text{ov}\left[B(s), B(t) - B(s) + B(s)\right] = 0 + \mathbb{V}\text{ar}\left[B(s)\right] = s = s \wedge t$$

- ***Joint density***: Let $t_1 < t_2$. Then

$$\begin{aligned}\mathbb{P}\left[B(t_1) = x_1, B(t_2) = x_2\right] &= \mathbb{P}\left[B(t_1) = x_1, B(t_2) - B(t_1) = x_2 - x_1\right] \\ &= \mathbb{P}\left[B(t_1) = x_1\right]\mathbb{P}\left[B(t_2 - t_1) = x_2 - x_1\right] \\ &= f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1)\end{aligned}$$

- ***Conditional density***: Let $s < t$. Then

$$\begin{aligned}\mathbb{P}\left[B(s) = x \mid B(t) = b\right] &= \frac{\mathbb{P}\left[B(s) = x, B(t) = b\right]}{\mathbb{P}\left[B(t) = b\right]} \\ &= \frac{f_s(x) f_{t-s}(b-x)}{f_t(b)} \\ &= \frac{\frac{1}{\sqrt{2\pi x}}\exp\left(-\frac{x^2}{2a}\right)\frac{1}{\sqrt{2\pi(t-s)}}\exp\left(-\frac{(b-x)^2}{2(t-s)}\right)}{\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{b^2}{2t}\right)}\end{aligned}$$

By multiplying out and then completing the square, we find that

$$\left(B(s) \mid B(t) = b\right) \sim N\left(b\tfrac{s}{t}, s\left(1 - \tfrac{s}{t}\right)\right)$$

The mean is analogous to what we obtained in a Poisson process. The variance is less intuitive.

- ***Hitting time***: Let $T_a = \inf\left\{t : B(t) = a\right\}$, $a > 0$ (the Brownian path is symmetric, so the result should be identical for $-a$). Now consider

$$\begin{aligned}\mathbb{P}\left(B_t \geq a\right) &= \mathbb{P}\left(B_t \geq a \mid T_a \leq t\right)\mathbb{P}\left(T \leq t\right) + \cancel{\mathbb{P}\left(B_t \geq a \mid T_a > t\right)\mathbb{P}\left(T > t\right)} \\ &= \tfrac{1}{2}\mathbb{P}\left(T \leq t\right)\end{aligned}$$

The last line follows from the so-called *reflecting principle*; once B_t has hit a, it is equally likely to be above and below a at a later time, by symmetry. Now

$$\mathbb{P}\left(T_a \leq t\right) = 2\mathbb{P}\left(B_t \geq a\right) = 2\overline{\Phi}\left(\tfrac{a}{\sqrt{t}}\right)$$

Now, as $t \to \infty$, we would expect this probability to tend to 1, since the Brownian motion should *eventually* hit a. Indeed, $\mathbb{P}\left(T_a \leq t\right) \to 2\overline{\Phi}(0) = 1$. However, Brownian motion is a null-recurrent Markov chain – the expected value of *any* hitting time is infinity:

$$\mathbb{E}(T_a) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{a}{\sqrt{t}} e^{-a^2/2t} \geq \frac{a}{\sqrt{2\pi}} \int_1^\infty \frac{1}{\sqrt{t}} e^{-a^2/2t} \, \mathrm{d}t \geq \frac{ae^{-a^2/2}}{\sqrt{2\pi}} \int_1^\infty \frac{\mathrm{d}t}{\sqrt{t}} = \infty$$

Note also that if $M_t = \max_{s \leq t} B_s$ (a quantity always increasing in t), we have

$$\mathbb{P}\left(M_t \geq a\right) = \mathbb{P}\left(T_a \leq t\right) = 2\overline{\Phi}\left(\tfrac{a}{\sqrt{t}}\right)$$

Similarly, note that $\min_{s \leq t} B_s = -\max_{s \leq t}(-B_s) =_{\mathrm{d}} -\max_{s \leq t} B_s$.

- ***Arcsine law***: Let $X(t)$ be a Brownian Motion starting at a point other than 0. Note that

$$\begin{aligned}\mathbb{P}\left(\min_{0 \leq u \leq t} X(u) \leq 0 \mid X(0) = a\right) &= \mathbb{P}\left(\max_{0 \leq u \leq t} X(u) \geq 0 \mid X(0) = -a\right) \\ &= \mathbb{P}\left(\max_{0 \leq u \leq t} X(u) \geq a \mid X(0) = 0\right) \\ &= \mathbb{P}\left(T_a \leq t\right)\end{aligned}$$

As such

$$\mathbb{P}\left(\text{At least one 0 in } [t_0, t_1] \mid X(t_0) = a\right) = P(a) = \mathbb{P}(T_{|a|} \leq t)$$

Now, assume we know $X(0) = 0$, but do not know $X(t_0)$. The probability α of there being at least one 0 in $[t_0, t_1]$. We can find this by conditioning on the value of $X(t_0)$, and we get

$$\mathbb{P}\left(\text{At least one 0 in } (t_0, t_1) \mid X(0) = 0\right) = 1 - \frac{2}{\pi}\arcsin\sqrt{\frac{t_0}{t_1}} = \frac{2}{\pi}\arccos\sqrt{\frac{t_0}{t_1}}$$

- Consider

$$A_t(x,y) = \mathbb{P}\left(X(t) > y, \min_{0 \leq u \leq t} X(u) > 0 \mid X(0) = x\right)$$

Note that

$$\mathbb{P}\left(X(t) > y \mid X(0) = x\right) = A_t(x,y) + \mathbb{P}\left(X(t) > y, \min_{0 \leq u \leq t} \leq 0 \mid X(0) = x\right)$$

Consider the last term; by reflecting at the first time the process hits 0, we find that every sample path satisfying that term has a corresponding path that falls below 0 and with $X(t) < -y$ (and vice versa). As such

$$\begin{aligned}A_t(x,y) &= \mathbb{P}\left(X(t) > y \mid X(0) = x\right) - \mathbb{P}\left(X(t) < -y \mid X(0) = x\right) \\ &= \mathbb{P}\left(B_t > y - x\right) - \mathbb{P}\left(B_t < -(y + x)\right) \\ &= \mathbb{P}\left(B_t \in [y - x, y + x]\right)\end{aligned}$$

- Consider $M_t = \max_{0 \leq u \leq t} B_t$ and $Y_t = M_t - X_t$. Consider

$$\mathbb{P}\left(M_t \geq m, B_t \leq x\right) = \mathbb{P}\left(B_t \geq 2m - x\right)$$

Because for every path satisfying the first expression, we can reflect at the first point M_t hits m. The first path is more than $m - x$ away from m at t, and therefore so is the second, but in the opposite direction.

- Finally, note that Brownian motion is a special case of a Gaussian Process, since $B(t_1),\cdots,B(t_n)$ can be expressed as a linear combination of IID standard normals:

$$B(t_1) = \sqrt{t_1}\cdot Z_1$$
$$B(t_2) = B(t_1) + B(t_2) - B(t_1) = B(t_1) + B(t_2 - t_1) = \sqrt{t}Z_1 + (\sqrt{t_2 - t_1})Z_2$$
$$\vdots$$

- ***Example***: Consider $X(t) = tB(\frac{1}{t})$. The mean and covariance are still as before. To show the process is continuous, show that $\lim_{t\to 0} X(t) = 0 \Leftrightarrow \lim_{u\to 0}\frac{B(u)}{u} \to 0$. Express $B(u)$ as $\left[B(u) - B(u-1)\right] + \left[B(u-1) - B(u-2)\right] + \cdots$ and use SLLN. Therefore, it is a Brownian motion; Brownian motion is self-symetric; all we need to do is to consider the interval $t \in [0,1]$ and we have everything we need about the process. □

- ***Example***: Let $s < t < u$ and consider $\mathbb{E}\left[B_s B_t B_u\right]$. Another way of stating the Markov property is that *given the present*, past and future are independent. Thus

$$\mathbb{E}\left(\mathbb{E}\left[B_s B_t B_u \mid B_t\right]\right) = \mathbb{E}\left(B_t \mathbb{E}[B_s \mid B_t]\mathbb{E}[B_u \mid B_t]\right) = \tfrac{s}{t}\mathbb{E}\left(B_t^3\right) = 0$$ □

- ***Example (the Brownian Bridge)***: Let $X_t = B_t \mid B_1 = 1$, with $t \in [0,1]$. Now, using the result about the conditional distribution and for $0 < s < t < 1$,

$$\begin{aligned}\mathbb{E}\left[X_s X_t\right] &= \mathbb{E}\left[B_s B_t \mid B_1 = 0\right] \\ &= \mathbb{E}\left(B_t \mathbb{E}\left[B_s \mid B_t\right] \mid B_1 = 0\right) \\ &= \mathbb{E}\left(B_t B_t \tfrac{s}{t} \mid B_1 = 0\right) \\ &= \tfrac{s}{t}\mathbb{E}\left(B_t^2 \mid B_1 = 0\right) \\ &= s(1-t)\end{aligned}$$

(Note that $\mathbb{V}\text{ar}(X_t) = t(1-t)$. Thus, even though this is a Gaussian process, it is *not* Brownian motion). Consider the following argument instead:

$$\begin{aligned}\mathbb{E}\left[X_s X_t\right] &= \mathbb{E}\left(B_s \mathbb{E}[B_t \mid B_s] \mid B_1 = 0\right) \\ &= \mathbb{E}\left(B_s^2 \mid B_1 = 0\right) \\ &= s(1-s)\end{aligned}$$

This second argument is *incorrect*, because in going from the first to the second line, we condition B_t on the past while ignoring that we are at the same time condition B_t on the future (because we are conditioning on $B_1 = 0$). $\square$

- ***Martingales Associated with Brownian Motion***
  - Recall that in continuous time, the martingale property reads, for all $t > s$, $\mathbb{E}\left(X(t) \mid \mathcal{F}_s\right) = X(s)$. Defining a stopping time is more tricky; if T satisfies $\left\{T \le t\right\} \in \mathcal{F}_t$, then since $\mathcal{F}$ is an increasing family, $\left\{T < t\right\} = \bigcup_{n=1}^{\infty} \left\{T \le t - \frac{1}{n}\right\} \in \mathcal{F}_t$. However, if T only satisfies $\left\{T < t\right\} \in \mathcal{F}_t$, then $\left\{T \le t\right\} = \bigcap_n \left\{T < t + \frac{1}{n}\right\} \notin \mathcal{F}_t$ (again, since the family is increasing). To conclude this, we need to assume right-continuity of the filtration.
  - ***Theorem***: The following three processes are martingales:
    - $\left\{B_t\right\}$ (the "mean martingale")
    - $\left\{B_t^2 - t\right\}$ (the "variance martingale")
    - $\left\{\exp\left(\theta B_t - \frac{1}{2}\theta^2 t\right)\right\}$, where θ is a deterministic parameter (the "exponential martingale").

    Proof: The first two parts are trivial. For the last, recall that $\mathbb{E}\left(e^{\theta N(0,1)}\right) = e^{\theta^2/2}$ and $\mathbb{E}\left(e^{\theta B_t - \frac{1}{2}\theta^2 t} \mid \mathcal{F}_s\right) = \left(e^{\theta B_s - \frac{1}{2}\theta^2 t}\right)\mathbb{E}\left(e^{\theta(B_t - B_s)} \mid \mathcal{F}_s\right) = \left(e^{\theta B_s - \frac{1}{2}\theta^2 t}\right)\mathbb{E}\left(e^{\theta[N(0,1)]\sqrt{t-s}}\right)$. $\blacksquare$
  - The exponential martingale can be used to generate many other martingales. Let

$$f(\theta; t, x) = e^{\theta x - \frac{1}{2}\theta^2 t} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} H_n(t, x)$$

    Where H_n is the n^{th} Hermite polynomial, $H_n(t,x) = f^{(n)}(0; t, x)$. Feeding this into the martingale property and exchanging summation and expectation, we can use the fact that this holds for *any* θ including $\theta = 0$, and conclude that for each n, $\left\{H_n(t, B_t)\right\}$ is also a martingale.
  - We can apply the Optional Stopping Theorem to these martingales to get some interesting results.
  - ***Example***: Define $T = \inf\left\{t : B_t = -a \text{ or } b\right\}$. Using the mean martingale, we can find $p_b = a / (a+b)$. Using the variance martingale, we can find $\mathbb{E}(T) = ab$. $\square$

- *Example*: Let $X_t = \mu t - \sigma B_t$, with $\mu, \sigma > 0$; this could be seen as the "net demand up to time t", where σB_t is a production process. We are interested in $T = \inf\left\{t : X_t = b > 0\right\}$ (the first time stock depletes). $\mathbb{E}(T) = \frac{b}{\mu}$ is easily found using the mean martingale on B_T. Use the variance martingale for $\mathbb{V}\text{ar}(T) = \frac{\sigma^2 b}{\mu^3}$ □
- *Example*: $X_t = \mu t + \sigma B_t$, $T = \inf\left\{t : X_t \notin (-a, b)\right\}$. Use $e^{\theta B_T - \frac{1}{2}\theta^2 T} = e^{\theta\frac{X_T - \mu T}{\sigma} - \frac{1}{2}\theta^2 T}$. Choose $\theta = -2\mu / \sigma$, and use the OST (stopped martingale is bounded); $\mathbb{E}(e^{\frac{\theta}{\sigma}X_T}) = 1$. Directly gives p_b and p_a. Use OST on $B_T = (X_T - \mu T)/\sigma$ to find $\mathbb{E}(T)$. Can let $\mu < 0$ and $a \to -\infty$ and find $\mathbb{P}\left(\sup_t X_t \geq b\right) = e^{-2b\mu/\sigma^2}$. □
- *Example*: Following from the above example and letting $T = \inf\left\{t : X_t = b\right\}$, suppose we want $\mathbb{E}(e^{-\gamma T})$. Write $-\gamma T = \theta B_T - \frac{1}{2}\theta^2 T - \beta b$.
  - Use the OST on $\exp(\theta B_{T\wedge t} - \frac{1}{2}\theta^2[T \wedge t]) \leq \exp(\theta B_{T\wedge t})$ [bounded.]
  - Substitute $b = \mu T + \sigma B_T$ and equate coefficients of T and B_T.

Ito's Formula

- For a deterministic x_t, $\mathrm{d}x_t = \dot{x}\ \mathrm{d}t$ and $\mathrm{d}f(x_t) = f'(x_t)\mathrm{d}x_t + \frac{1}{2}f''(x_t)(\mathrm{d}x_t)^2 + \cdots$, but $(\mathrm{d}x_t)^2 = \dot{x}^2(\mathrm{d}t)^2$ which vanishes. In BM, $(\mathrm{d}B_t)^2 \sim \mathrm{d}t$, and so

$$\mathrm{d}f(B_t) = f'(B_t)\ \mathrm{d}B_t + \tfrac{1}{2}f''(B_t)\ \mathrm{d}t$$
$$f(B_t) = f(0) + \int_0^t f'(B_s)\ \mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\ \mathrm{d}s$$

This is Ito's Formula. The first integral above is called Ito's Integral and can be approximated as

$$\int_0^t X_s\ \mathrm{d}B_s \approx \sum_{i=1}^n X_{t_{i-1}}\left[B_{t_i} - B_{t_{i-1}}\right]$$

This is a martingale transformation, and is therefore a martingale under boundedness and predictability ($\Leftarrow$ left-continuity) of X_t. Furthermore, as a martingale, the mean of the integral is 0. To find its variance, consider

$$\mathbb{E}\left(\int_0^t X_s \mathrm{d}B_s\right)^2 \approx \mathbb{E}\left(\sum_{i=1}^n X_{t_{i-1}}\left[B_{t_i} - B_{t_{i-1}}\right]\right)^2 = \sum_{i=1}^n \mathbb{E}(X_{t_{i-1}}^2)(t_i - t_{i-1}) \approx \mathbb{E}\left(\int_0^t X_s^2 \mathrm{d}s\right)$$

Where we have used orthogonality of martingale differences, and conditioned on $\mathcal{F}_{t_{i-1}}$. This is knows as *Ito's Isometry*. More generally, for a bivariate $f(t, x)$

$$\mathrm{d}f(t,B_t) = f_t'\,\mathrm{d}t + f_x'\,\mathrm{d}B_t + \tfrac{1}{2}f_{xx}''\,\mathrm{d}t$$
$$f(T,B_T) = f(0,0) + \int_0^T f_t'\,\mathrm{d}t + \int_0^T f_x'\,\mathrm{d}B_t + \frac{1}{2}\int_0^T f_{xx}''\,\mathrm{d}t$$

Two results

- Since the second integral above is a martingale, we have that if $f(x,t)$ satisfies $f_t' = -\tfrac{1}{2}f_x''$, then $f(t,B_t)$ is a martingale.
- We can develop an analogue of integration by parts

$$\int_0^t B_s\,\mathrm{d}s = tB_t - \int_0^t s\dot{B}_s\,\mathrm{d}t = tB_t - \int_0^t s\,\mathrm{d}B_s$$

  This follows rigorously from Ito's Formula with $f(x,t) = tx$.

- ***Example***: Letting $f(x) = \tfrac{1}{2}x^2$ and $f'(x) = x$, we get

$$\tfrac{1}{2}B_t^2 = \int_0^t B_s\,\mathrm{d}B_s + \tfrac{1}{2}t \Rightarrow \int_0^t B_s\,\mathrm{d}B_s = \tfrac{1}{2}\left(B_t^2 - t\right)$$ □

- ***Example***: Let $f(t,x) = e^{\mu t + \sigma B_t}$ and let $Y_t = f(t,B_t) = e^{\mu t+\sigma B_t}$. Taking derivatives, we have $f_t' = \mu f$, $f_x' = \sigma f$, $f_{xx}'' = \sigma^2 f$ and so

$$\mathrm{d}f(t,B_t) = \mathrm{d}Y_t = \left[(\mu + \tfrac{1}{2}\sigma^2)\mathrm{d}t + \sigma\,\mathrm{d}B_t\right]Y_t$$

  We could have obtained the same result using $f(x) = e^x$ and $Y_t = f(X_t)$ with $X_t = \mu t + \sigma B_t$. □

- ***Example***: We define *integrated Brownian Motion* as $X_t = \int_0^t B_s\,\mathrm{d}s$. This is a Gaussian process, and clearly $\mathbb{E}(X_t) = \int_0^t \mathbb{E}(B_s)\,\mathrm{d}s = 0$. Furthermore, using integration by parts, we can write $X_t = \int_0^t B_s\,\mathrm{d}s = \int_0^t (t-s)\,\mathrm{d}B_s$, and so, for $t < u$ and repeatedly using independent increments

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(X_t, X_u\right) &= \mathbb{C}\mathrm{ov}\left(\int_0^t (t-s)\,\mathrm{d}B_s, \int_0^u (u-s)\,\mathrm{d}B_s\right)\\ &= \mathbb{C}\mathrm{ov}\left(\int_0^t (t-s)\,\mathrm{d}B_s, \int_0^t (u-s)\,\mathrm{d}B_s\right)\\ &= \mathbb{E}\left(\int_0^t (t-s)\,\mathrm{d}B_s \int_0^t (u-s)\,\mathrm{d}B_s\right)\\ &= \int_0^t (t-s)(u-s)\,\mathbb{E}\left[(\mathrm{d}B_s)^2\right]\\ &= \int_0^t (t-s)(u-s)\,\mathrm{d}s\\ &= \tfrac{1}{2}t^2\left(u - \tfrac{t}{3}\right)\end{aligned}$$

  So this is *not* Brownian Motion. □

- *Example*: Consider a more complicated integration by parts

$$\int_0^t B_s^2 \, \mathrm{d}s = tB_t^2 - \int_0^t s \, \mathrm{d}(B_s^2) = tB_t^2 - \tfrac{1}{2}t^2 - 2\int_0^t sB_s \, \mathrm{d}B_s$$

Where the last step used Ito's Formula with $f(x) = x^2$ to deduce $\mathrm{d}(B_s^2) = 2B_s \, \mathrm{d}B_s + \mathrm{d}s$. □

- *Example*: Consider a inventory problem in which the cumulative (net) demand up to time t is $D_t = \lambda t + \sigma B_t$ with initial inventory $S > 0$. When the inventory drops to 0, it is instantly replenished to S. Let $\tau = \inf\left\{t : D_t = S\right\}$. The total inventory-time in $[0,\tau]$ is

$$\begin{aligned}\int_0^\tau (S - D_t) \, \mathrm{d}t &= \int_0^\tau (S - \lambda t - \sigma B_t) \, \mathrm{d}t \\ &= S\tau - \tfrac{1}{2}\lambda\tau^2 - \sigma\tau B_\tau + \sigma\int_0^\tau t \, \mathrm{d}B_t \\ &= \tfrac{1}{2}\lambda\tau^2 + \sigma\int_0^\tau t \, \mathrm{d}B_t\end{aligned}$$

We have used integration by parts and $D_\tau = S$. Taking expectations, and using the OST on the uniformly integrable martingale $\int_0^t s \, \mathrm{d}B_s$, we get

$$\mathbb{E}\left(\int_0^\tau (S - D_t) \, \mathrm{d}t\right) = \tfrac{1}{2}\lambda\mathbb{E}(\tau^2)$$

Assuming the holding-cost per unit time is h and each replenishment costs $cS + K$, the long-term average cost which we seek to minimize is

$$\min_{\lambda,S} \frac{\tfrac{1}{2}\lambda h\mathbb{E}(\tau^2) + cS + K}{\mathbb{E}(\tau)}$$

From the previous section, we have $\mathbb{E}(\tau) = \frac{S}{\lambda}$ and $\mathbb{E}(\tau^2) = \frac{\sigma^2}{\lambda^2}\mathbb{E}(\tau) + \mathbb{E}^2(\tau)$. Feeding into the above and minimizing, we find

$$S = \sqrt{\frac{2K\lambda}{h}} \qquad \lambda = \sqrt{\frac{hS\sigma^2}{2(cs+K)}}$$ □

Part III – Stationary Processes

Introduction

- ***Definition (Strong Stationarity)***: The process $\left\{X(t) : t \geq 0\right\}$ is *stationary* if

$$\left(X(t_1), X(t_2), \cdots, X(t_n)\right) =_d \left(X(t_1+n), X(t_2+n), \cdots, X(t_n+n)\right)$$

For all $t_1 < t_2 < \cdots < t_n$ and $n \geq 0$. In other words, a process is stationary if it is "shift invariant".

- ***Definition (Weak Stationarity)***: The process $\{X(t) : t \geq 0\}$ is said to be *covariance stationary* if $\mathbb{E}[X(t)] = m$ (ie: it has constant mean) and $\mathbb{C}\text{ov}[X(t), X(s)] = \mathbb{E}[(X(t)-m)(X(s)-m)] = R(|t-s|)$, where R is some function of $|t-s|$ only.
- Note that a Gaussian process is entirely defined by its mean and covariance; as such, for a Gaussian process, stationarity implies covariance stationarity.

L_p-convergence

- We define the p-norm of a random variable X as $\|X\|_p = \mathbb{E}^{1/p}\left[|X|^p\right]$.
- ***Holder's Inequality***: For $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then $\mathbb{E}|XY| \leq \|X\|_p \|Y\|_q$. Applications:
  - $\boldsymbol{p = q =}$ 2 gives the Cauchy-Schwartz Inequality: $\mathbb{E}|XY| \leq \|X\|_2 \|Y\|_2$.
  - $\boldsymbol{p = s/r}$ and $Y = 1$ gives Lyapounov's Inequality $\|X\|_r \leq \|X\|_s$ for all $0 \leq r \leq s$.
  - ***Minkowski's Inequality***: $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$.
- Definitions
  - $X \in L_p$ means that $\|X\|_p < \infty$.
  - A sequence is said to be L_p-bounded if $\sup_n \|X\|_p < \infty$.
  - $X_n \to_{L_p} X$ means that $\|X_n - X\|_p \to 0$.
  - $X_n \in L_p$ and $X_n \to X \Rightarrow X \in L_p$, because $\|X\|_p = \|(X - X_n) + X_n\|_p$, and use Minkowski to find $\|X\|_p \leq \|X - X_n\|_p + \|X_n\|_p \leq \varepsilon + \|X_n\|_p < \infty$.
  - Let L_p limit is unique: if $X_n \to_{L_p} X$ and $X_n \to_{L_p} Y$ then use Minkowski's Inequality on $X - Y = X - X_n + (X_n - Y)$.
- ***Theorem***: If $\{X_n\}$ if a martingale that is L_p bounded with $p > 1$, then $X_n \to X$ a.s. in L_p. The same holds if $\{X_n\}$ is a non-negative submartingale.
- ***Cauchy Criterion***: $\|X_m - X_n\|_p \to 0$ for $n, m \to \infty \Leftrightarrow \exists X$ s.t. $\|X_n - X\|_p \to 0$.

Theorems

- ***Theorem (Weak Ergodic Theorem)***: If $\{X_n\}$ is covariance stationary, then $\bar{X}_n \to_{L_2} \bar{X}$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Note that in general, $\bar{X}$ will be a random variable. Under one of two equivalent conditions, it is a deterministic constant equal to $\mathbb{E}(X_1) = m$:
  - o $\mathbb{V}\text{ar}(\bar{X}_n) \to 0$
  - o $\mathbb{C}\text{ov}\left(\bar{X}_n, X_1\right) \to 0$
- ***Theorem (Strong Ergodic Theorem)***: If $\{X_n\}$ is stationary and $X_n \in L_1$, then $\bar{X}_n \to \bar{X}$ a.s. and $\bar{X}_n \to_{L_1} \bar{X}$. Again, $\bar{X}$ is generally a random variable. Under *ergodicity* (defined below), it is constant. An intermediate step in proving this is…

  Lemma (Maximal Ergodic Lemma): Let $\{Y_n\}$ be stationary and let $S_{i,n} = Y_i + \cdots + Y_{i+n-1}$ and $M_{i,n} = \max\left\{0, S_{i,1}, \cdots, S_{i,n}\right\}$. Then $\mathbb{E}[Y_0 Y \mathbb{1}_{\{M_{0,n}>0\}}] \geq 0 \ \forall n \geq 1$.
- ***Definition (Shift Operator)***: φ is a shift operation. If $x = (x_1, x_2, \cdots)$ then $\varphi(x) = (x_2, x_3, \cdots)$. We say the set A is *shift invariant* if $x \in A \Rightarrow \varphi(x) \in A$. For example, the following two sets are shift invariant (we can replace $\leq$ by $\geq$ and =, and lim sup by lim inf)

  $$A_1 = \left\{x : \lim_{n\to\infty} \tfrac{x_1 + \cdots + x_n}{n} \leq a\right\} \qquad A_2 = \left\{x : \limsup_{n\to\infty} \tfrac{x_1 + \cdots + x_n}{n} \leq a\right\}$$
- ***Definition (Ergodicity)***: $X = \left(X_1, \cdots, X_n, \cdots\right)$ is ergodic if $\mathbb{P}\left(X \in A\right) = 0 \text{ or } 1$ for all shift invariant A.

  Note: We can use this definition of ergodicity to show that $\bar{X}$ is constant.

  $$\mathbb{P}\left(X \in A\right) = \mathbb{P}\left(\lim_{n\to\infty} \bar{X} \leq a\right) =_{\text{a.s.}} \mathbb{P}\left(\bar{X} \leq a\right) = 0 \text{ or } 1$$

  Thus, $\bar{X}$ is a constant.
- ***Definition (Mixing)***: $X = \left(X_1, X_2, \cdots, X_n, \cdots\right)$ is *mixing* if

  $$\lim_{n\to\infty} \mathbb{P}\left((X_1, X_2, \cdots) \in A, (X_{n+1}, X_{n+2}, \cdots) \in B\right) = \mathbb{P}\left((X_1, X_2, \cdots) \in A\right) \cdot \mathbb{P}\left((X_1, X_2, \cdots) \in A\right)$$

  Where A and B are sets of infinite sequences (not necessarily shift invariant). We can look at pieces of size k instead $\lim_{n\to\infty} \mathbb{P}\left((X_n, \cdots, X_{n+k}) \in A, (X_{n+m}, \cdots, X_{n+m+k}) \in B\right)$.

  Theorem: If a sequence is mixing, it is ergodic.

  Proof: Let A be any shift invariant set

$$\begin{aligned} p &= \mathbb{P}\Big((X_0, X_1, \cdots) \in A\Big) \\ &=_{A \text{ shift invariant}} \mathbb{P}\Big((X_0, X_1, \cdots) \in A, (X_n, X_{n+1}, \cdots) \in A\Big) \\ &\to_{\text{mixing}} \mathbb{P}\Big((X_0, X_1, \cdots) \in A\Big) \cdot \mathbb{P}\Big((X_0, X_1, \cdots) \in A\Big) \\ &= p^2 \end{aligned}$$

And so p must be equal to 0 or 1. ■

- ***Example***: Let $X_0 = Y$ w.p. p and Z w.p $1-p$, where $Y, Z \in L_1$. Consider two cases:
    - If $X_n \equiv X_0$, $\bar{X} = Y$ w.p p and Z w.p $1-p$
    - If $X_n =_d X_0$, $\bar{X} = \mathbb{E}(Y)$ w.p p and $\mathbb{E}(Z)$ w.p $1-p$

  Neither case is ergodic.

Part IV – Stochastic Orders

Introduction

- The Hazard Rate of a random variable X with CDF F is

$$\lambda_X(t) = \lim_{\mathrm{d}t \to 0} \mathbb{P}\Big[X \in (t, t+\mathrm{d}t) \mid X > t\Big] = \frac{f(t)}{\bar{F}(t)} \qquad t \geq 0$$

  As such

$$\lambda_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}\Big[-\ln \bar{F}(t)\Big] \qquad\qquad \bar{F}(t) = \exp\left(-\int_0^t \lambda_X(s)\ \mathrm{d}s\right)$$

  X follows the exponential distribution $\bar{F}(t) = e^{-t/\mu}$ if and only if $\lambda_X(t) = \frac{1}{\mu}$.

  Example: If $W = \max(X, Y)$, then $\bar{F}_W = \bar{F}_X \bar{F}_Y$, and $\lambda_W = \lambda_X + \lambda_Y$. ■

- Assume X and Y have densities F and G. We define the following variable orderings
    - ***Likelihood ratio ordering***: $X \geq_{\text{LR}} Y$ if $\dfrac{f(y)}{f(x)} \leq \dfrac{g(y)}{g(x)}$ for all $x \geq y$.
    - ***Hazard rate ordering***: $X \geq_{\text{HR}} Y$ if $\lambda_X(x) \leq \lambda_Y(x)$ for all x.
    - ***Stochastic ordering***: $X \geq_{\text{ST}} Y$ if $\bar{F}(x) \geq \bar{G}(y)$ for all x.
    - ***Increasing convex ordering***: $X \geq_{\text{ICX}} Y$ if $\mathbb{E}\big[\phi(X)\big] \geq \mathbb{E}\big[\phi(Y)\big]$ for all increasing and convex functions $\phi(x)$,
- ***Theorem***: $X \geq_{\text{LR}} Y \Rightarrow X \geq_{\text{HR}} Y \Rightarrow X \geq_{\text{ST}} Y \Rightarrow X \geq_{\text{ICX}} Y$

Proof: For $(\text{i}) \to (\text{ii})$, collect x and y terms on each side, integrate both sides on $x \geq y$. For $(\text{ii}) \to (\text{iii})$, use $\bar{F}(x) = \exp\left(-\int_{-\infty}^{x} f(t) / \bar{F}(t)\ \mathrm{d}t\right)$. For $(\text{iii}) \to (\text{iv})$ see next theorem. ■

- ***Theorem***: $X \leq_{\text{icx}} Y$ and $\mathbb{E}(X) = \mathbb{E}(Y) \Leftrightarrow \mathbb{E}\left[f(X)\right] \leq \mathbb{E}\left[f(Y)\right]$ for all convex f. (In other words, when the means are equal, the requirement for the f to be increasing is dropped).
- ***Theorem***: $X \geq_{\text{ST}} Y$ iff $\mathbb{E}\left[\phi(X)\right] \geq \mathbb{E}\left[\phi(Y)\right]$ for any increasing ϕ.

  Proof: $\boxed{\Rightarrow}$ Let $U \sim U[0,1]$. Then $X =_{\text{d}} F^{-1}(U)$ and $Y =_{\text{d}} G^{-1}(U)$. Stochastic ordering implies $F^{-1}(u) \geq G^{-1}(u)$, and so $\mathbb{E}\left[\phi(F^{-1}(u))\right] \geq \mathbb{E}\left[\phi(G^{-1}(u))\right]$. $\boxed{\Leftarrow}$ Pick $\phi(x) = \mathbb{1}_{\{x \geq a\}}$. ■
- Note $\phi(x)$ increasing + convex $\Leftrightarrow -\phi(-x)$ increasing + concave. Thus, we can define a concave ordering ICV for increasing *concave* functions, and $X \geq_{\text{ICX}} Y \Leftrightarrow -X \leq_{\text{ICV}} -Y$.
- ***Theorem***: Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors with independent components.
  - If $X_i \geq_{\text{ST}} Y_i$ for all i, then $h(\boldsymbol{X}) \geq_{\text{ST}} h(\boldsymbol{Y})$ for any increasing $h : \mathbb{R}^n \to \mathbb{R}$.
  - If $X_i \geq_{\text{ICX}} Y_i \geq 0$ for all i, then $h(\boldsymbol{X}) \geq_{\text{UCX}} h(\boldsymbol{Y})$ for any increasing $h : \mathbb{R}_+^n \to \mathbb{R}_+$ that is convex in each variable.

  Proof:

IFR & NBUE Distributions

- We let X_t denote the *residual lifetime* of X given that X has survived up to time t. In other words, $\mathbb{P}\left(X_t > x\right) = \mathbb{P}\left(X > x + t \mid X > t\right) = \frac{\bar{F}(x+t)}{\bar{F}(t)}$. Differentiating, we find that the density of X_t is $f(x+t) / \bar{F}(x)$, and so the failure rate is $\lambda_{X_t}(x) = \lambda_X(t+x)$.
- ***Definition***: IFR denotes the class of distributions with *increasing failure rates* – in other words, $X \in \text{IFR}$ if $\lambda_X(t)$ is an increasing function of t.

  Theorem: $X \in \text{IFR} \Leftrightarrow X_t$ stochastically decreasing.

  Proof: $\bar{F}_{X_t}(x) = \exp\left(-\int_0^t \lambda_{X_t}(s)\ \mathrm{d}s\right) = \exp\left(-\int_t^{t+x} \lambda_X(s)\ \mathrm{d}s\right)$. Clearly, this is increasing in t if and only if $\lambda_X(s)$ is a decreasing function in s. ■
- Now, consider a renewal process with IID inter-arrival time following the distribution of X. Let X_e be the equilibrium distribution of that renewal process – in other words, at any given time t far into the future, the time since the last renewal. Then $\bar{F}_{X_e}(t) = \frac{1}{\mathbb{E}(X)} \int_t^{\infty} \bar{F}(x)\ \mathrm{d}x$, and $\lambda_{X_e}(t) = \bar{F}(t) / \int_t^{\infty} \bar{F}(x)\ \mathrm{d}x$.

- ***Definition***: NBUE denotes the class of distributions that satisfies the property of "new better than used in expectation". In other words, $X \in \text{NBUE}$ if

$$\mathbb{E}\left[X - t \mid X > t\right] = \frac{\mathbb{E}(X-t)^+}{\mathbb{P}(X > t)} = \frac{1}{\bar{F}(t)}\int_t^\infty \bar{F}(x)\ \mathrm{d}x \le \mathbb{E}(X) = \mu \qquad \forall t \ge 0$$

  Theorem: $X \in \text{NBUE} \Leftrightarrow X_e \le_{\text{st}} X \Leftrightarrow \lambda_{X_e}(t) \ge \frac{1}{\mu}$ for all $t \ge 0$. ***Proof***: Elementary.

  Theorem: $X \in \text{IFR} \Rightarrow X \in \text{NBUE}$.

  Proof: $\mathbb{E}(X - t \mid X > t) =_{\text{def of } X_t} \mathbb{E}(X_t) \le_{X_t \text{ stoch. dec.}} \mathbb{E}(X_0) =_{\text{def of } X_t} \mathbb{E}(X)$ ■

- ***Theorem***: $X \in \text{NBUE} \Rightarrow X \le_{\text{icx}} \exp(1/\mu)$, where $\mu = \mathbb{E}(X)$.

  Proof: From the previous theorem, we have that since $X \in \text{NBUE}$, and using $\bar{F}(0) = 1$,

$$\lambda_{X_e}(t) \ge \frac{1}{\mu} \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left[-\ln \bar{F}_e(t)\right] \ge \frac{1}{\mu} \Rightarrow \ln \bar{F}_{X_e}(0) - \ln \bar{F}_{X_e}(x) \ge \frac{x}{\mu} \Rightarrow \bar{F}_{X_e}(x) \le e^{-x/\mu}$$

  Using the definition of $\bar{F}_{X_e}$, the last inequality can be re-written as

$$\int_x^\infty \bar{F}(y)\ \mathrm{d}y \le \mu e^{-x/\mu} = \int_x^\infty e^{-y/\mu}\ \mathrm{d}y$$

  Proposition 9.5.1. in Ross shows that this implies icx ordering. ■

- Note that $X \le_{\text{icx}} Y$ and $\mathbb{E}(X) = \mathbb{E}(Y)$ together imply that $\mathbb{V}\text{ar}(X) \le \mathbb{V}\text{ar}(Y)$, because $\varphi(x) = x^2$ is convex. Now, consider the coefficient of variation, $\sqrt{\mathbb{V}\text{ar}(Y) / \mathbb{E}^2(Y)}$. Using the previous theorem, we can conclude that if $X \in \text{NUBE}$, it's coefficient of variation is ≤ 1.

  - Let us revisiting problem 1: $X_0 = I$ and $X_n \sim Po(X_{n-1})$. Here, we know $X_\infty = 0$, and $\mathbb{E}(X_0) = i \ge \mathbb{E}(X_\infty) = 0$. Furthermore, let $T = \inf\left\{n : X_n \ge b > i\right\}$. But we can be absorbed in 0, so this isn't finite. $\left\{T_{T \wedge n}\right\}$ is non-negative, but NOT uniformly integrable. Thus, we need to use our alternate result…

$$\begin{aligned} i = \mathbb{E}(X_0) &\geq \mathbb{E}\left[B\mathbb{I}_{T<\infty} + 0 \cdot \mathbb{I}_{T=\infty}\right] \\ &\geq B\mathbb{P}(T < \infty) \\ &\geq b \cdot \tfrac{i}{b} \end{aligned}$$

In the last stage, we use *something* like this

$$\begin{aligned} &\tilde{T} = \inf\left\{n : X_n \geq b \text{ or } X_n = 0\right\} \\ &OST \quad \mathbb{E}(X_{\tilde{T}}) = \mathbb{E}(X_0) = i = 0(1-\pi) + B\pi \quad [B \geq b] \end{aligned}$$

So let $\mathbb{P}(X_{\tilde{T}} \geq b) = \pi = \frac{i}{B} \leq \frac{i}{b}$. So, $\mathbb{P}(X_{\tilde{T}} = 0) \geq 1 - \frac{i}{b}$.

- EXAMPLE: Symmetric random walk, with $S_0 = 1$, $T = \inf\{n : S_n = 0\}$. $S_{T \wedge n}$ is a non-negative martingale. So clearly, it must converge. And to 0. If we have $S_0 = -1$ (ie: it's a barrier on the left), then we lose non-negativity, and the whole thing becomes irrelevant.