Chapter 2

Conditional Expectation and Martingales

Information in probability and its applications is related to the notion of sigma algebras. For example if I wish to predict whether tomorrow is wet or dry $(X_2 = 1 \text{ or } 0)$ based only on similar results for today (X_1) and yesterday (X_0), then the I am restricted to random variables that are functions $g(X_0, X_1)$ of the state on these two days. In other words, the random variable must be measurable with respect to the sigma algebra generated by X_0, X_1. Our objective is, in some sense, to get as close as possible to the unobserved value of X_2 using only random variables that are measurable with respect to this sigma algebra. This is essentially one way of defining conditional expectation. It provides the closest approximation to a random variable X if we restrict to random variables Y measurable with respect so some courser sigma algebra.

Conditional Expectation.

Theorem A23

Let $\mathcal{G} \subset \mathcal{F}$ be sigma-algebras and X a random variable on $(\Omega, \mathcal{F}, P)$. Assume $E(X^2) < \infty$. Then there exists an almost surely unique $\mathcal{G}$-measurable Y such that

$$E[(X-Y)^2] = \inf_Z E(X-Z)^2 \tag{2.1}$$

where the infimum (infimum=greatest lower bound) is over all $\mathcal{G}$-measurable random variables. *Note.* We denote the minimizing Y by $E(X|\mathcal{G})$.

For two such minimizing Y_1, Y_2, i.e. random variables Y which satisfy (2.1), we have $P[Y_1 = Y_2] = 1$. This implies that conditional expectation is almost surely unique.

Suppose $\mathcal{G} = \{\varphi, \Omega\}$. What is $E(X|\mathcal{G})$? What random variables are measurable with respect to $\mathcal{G}$? Any non-trivial random variable which takes two or more possible values generates a non-trivial sigma-algebra which includes sets that are strict subsets of the probability space Ω. Only a constant random variable is measurable with respect to the trivial sigma-algebra $\mathcal{G}$. So the question becomes what constant is as close as possible to all of the values of the random variable X in the sense of mean squared error? The obvious answer is the correct one in this case, the expected value of X because this leads to the same minimization discussed before, $\min_c E[(X-c)^2] = \min_c\{var(X) + (EX-c)^2\}$ which results in $c = E(X)$.

Example

Suppose $\mathcal{G} = \{\varphi, A, A^c, \omega\}$ for some event A. What is $E(X|\mathcal{G})$?

Consider a candidate random variable Z taking the value a on A and b on the set A^c. Then

$$\begin{aligned} E[(X-Z)^2] &= E[(X-a)^2 I_A] + E[(X-b)^2 I_{A^c}] \\ &= E(X^2 I_A) - 2aE(XI_A) + a^2 P(A) \\ &+ E(X^2 I_{A^c}) - 2bE(XI_{A^c}) + b^2 P(A^c). \end{aligned}$$

Minimizing this with respect to both a and b results in

$$\begin{aligned} a &= E(XI_A)/P(A) \\ b &= E(XI_{A^c})/P(A^c). \end{aligned}$$

These values a and b are usually referred to in elementary probability as $E(X|A)$ and $E(X|A^c)$ respectively. Thus, the conditional expected value can be written

$$E(X|\mathcal{G})(\omega) = \begin{cases} E(X|A) & \text{if } \omega \in A \\ E(X|A^c) & \text{if } \omega \in A^c \end{cases}$$

As a special case consider X to be an indicator random variable $X = I_B$. Then we usually denote $E(I_B|\mathcal{G})$ by $P(B|\mathcal{G})$ and

$$P(B|\mathcal{G})(\omega) = \begin{cases} P(B|A) & \text{if } \omega \in A \\ P(B|A^c) & \text{if } \omega \in A^c \end{cases}$$

Note: *Expected value is a constant, but the conditional expected value $E(X|G)$ is a random variable measurable with respect to G. Its value on the atoms (the distinct elementary subsets) of $\mathcal{G}$ is the average of the random variable X over these atoms.*

Example

Suppose $\mathcal{G}$ is generated by a finite partition $\{A_1, A_2, ..., A_n\}$ of the probability space Ω. What is $E(X|\mathcal{G})$?

In this case, any $\mathcal{G}$-measurable random variable is constant on the sets in the partition $A_j, j = 1, 2, ..., n$ and an argument similar to the one above shows that the conditional expectation is the simple random variable:

$$E(X|\mathcal{G})(\omega) = \sum_{i=1}^{n} c_i I_{A_i}(\omega)$$

$$\text{where } c_i = E(X|A_i) = \frac{E(XI_{A_i})}{P(A_i)}$$

Example

Consider the probability space $\Omega = (0, 1]$ together with $P =$ Lebesgue measure and the Borel Sigma Algebra. Suppose the function $X(\omega)$ is Borel measurable. Assume that $\mathcal{G}$ is generated by the intervals $(\frac{j-1}{n}, \frac{j}{n}]$ for $j = 1, 2, ..., n$. What is $E(X|\mathcal{G})$?

In this case

$$E(X|\mathcal{G})(\omega) = n\int_{(j-1)/n}^{j/n} X(s)ds \text{ when } \omega \in (\frac{j-1}{n}, \frac{j}{n}]$$
$$= \text{average of } X \text{ values over the relevant interval.}$$

Theorem A24 Properties of Conditional Expectation

(a) If a random variable X is $\mathcal{G}$-measurable, $E(X|\mathcal{G}) = X$.

(b) If a random variable X independent of a sigma-algebra $\mathcal{G}$, then $E(X|\mathcal{G}) = E(X)$.

(c) For any square integrable $\mathcal{G}$-measurable Z, $E(ZX) = E[ZE(X|\mathcal{G})]$.

(d) (special case of (c)): $\int_A XdP = \int_A E(X|\mathcal{G}]dP$ for all $A \in \mathcal{G}$.

(e) $E(X) = E[E(X|\mathcal{G})]$.

(f) If a $\mathcal{G}$-measurable random variable Z satisfies $E[(X-Z)Y] = 0$ for all other $\mathcal{G}$-measurable random variables Y, then $Z = E(X|\mathcal{G})$.

(g) If Y_1, Y_2 are distinct $\mathcal{G}-$measurable random variables both minimizing $E(X-Y)^2$, then $P(Y_1 = Y_2) = 1$.

(h) *Additive* $E(X + Y|\mathcal{G}) = E(X|\mathcal{G}) + E(Y|\mathcal{G})$.
Linearity $E(cX + d|\mathcal{G}) = cE(X|\mathcal{G}) + d$.

(i) If Z is $\mathcal{G}-$measurable, $E(ZX|\mathcal{G}) = ZE(X|\mathcal{G})$ a.s.

(j) If $\mathcal{H} \subset \mathcal{G}$ are sigma-algebras, $E[E(X|\mathcal{G})|\mathcal{H}] = E(X|\mathcal{H})$.

(k) If $X \leq Y$, $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ a.s.

(l) *Conditional Lebesgue Dominated Convergence*. If $X_n \to X$ a.s. and $|X_n| \leq Y$ for some integrable random variable Y, then $E(X_n|\mathcal{G}) \to E(X|\mathcal{G})$ in distribution

Notes. In general, we define $E(X|Z) = E(X|\sigma(Z))$, the conditional expected value given the sigma algebra generated by X, $\sigma(X)$. We can define the conditional variance $var(X|\mathcal{G}) = E\{(X - E(X|\mathcal{G}))^2|\mathcal{G}\}$.

Proof.

(a) Notice that for any random variable Z that is $\mathcal{G}$-measurable, $E(X-Z)^2 \geq E(X-X)^2 = 0$ and so the minimizing Z is X (by definition this is $E(X|\mathcal{G})$).

(b) Consider a random variable Y measurable with respect $\mathcal{G}$ and therefore independent of X. Then

$$\begin{aligned} E(X-Y)^2 &= E[(X-EX+EX-Y)^2] \\ &= E[(X-EX)^2] + 2E[(X-EX)(EX-Y)] + E[(EX-Y)^2] \\ &= E[(X-EX)^2] + E[(EX-Y)^2] \text{ by independence} \\ &\geq E[(X-EX)^2]. \end{aligned}$$

It follows that $E(X-Y)^2$ is minimized when we choose $Y = EX$ and so $E(X|\mathcal{G}) = E(X)$.

(c) for any $\mathcal{G}-$measurable square integrable random variable Z, we may define a quadratic function of λ by

$$g(\lambda) = E[(X - E(X|\mathcal{G}) - \lambda Z)^2]$$

By the definition of $E(X|\mathcal{G})$, this function is minimized over all real values of λ at the point $\lambda = 0$ and therefore $g'(0) = 0$. Setting its derivative $g'(0) = 0$ results in the equation

$$E(Z(X - E(X|\mathcal{G}))) = 0$$

or $E(ZX) = E[ZE(X|\mathcal{G})]$.

(d) If in (c) we put $Z = I_A$ where $A \in \mathcal{G}$, we obtain $\int_A X dP = \int_A E(X|\mathcal{G}] dP$.

(e) Again this is a special case of property (c) corresponding to $Z = 1$.

(f) Suppose a $\mathcal{G}$-measurable random variable Z satisfies $E[(X-Z)Y] = 0$ for all other $\mathcal{G}$-measurable random variables Y. Consider in particular $Y = E(X|\mathcal{G}) - Z$ and define

$$\begin{aligned} g(\lambda) &= E[(X - Z - \lambda Y)^2] \\ &= E((X-Z)^2 - 2\lambda E[(X-Z)Y] + \lambda^2 E(Y^2) \\ &= E(X-Z)^2 + \lambda^2 E(Y^2) \\ &\geq E(X-Z)^2 = g(0). \end{aligned}$$

In particular $g(1) = E[(X - E(X|\mathcal{G}))^2] \geq g(0) = E(X-Z)^2$ and by the uniqueness of conditional expectation in Theorem A23, $Z = E(X|\mathcal{G})$ almost surely.

(g) This is just deja vu (Theorem A23) all over again.

(h) Consider, for an arbitrary $\mathcal{G}-$measurable random variable Z,

$$\begin{aligned} E[Z(X+Y-E(X|\mathcal{G})-E(Y|\mathcal{G}))] &= E[Z(X-E(X|\mathcal{G}))] + E[Z(Y-E(Y|\mathcal{G}))] \\ &= 0 \text{ by property (c).} \end{aligned}$$

It therefore follows from property (f) that $E(X+Y|\mathcal{G}) = E(X|\mathcal{G}) + E(Y|\mathcal{G})$.

By a similar argument we may prove $E(cX + d|\mathcal{G}) = cE(X|\mathcal{G}) + d$.

(i)-(l) We leave the proof of these properties as exercises ■

2.1 Conditional Expectation for integrable random variables

We have defined conditional expectation as a projection (i.e. a $\mathcal{G}-$measurable random variable that is the closest to X) only for random variables with finite variance. It is fairly easy to extend this definition to random variables X on a probability space $(\Omega, \mathcal{F}, P)$ which are integrable, i.e. for which $E(|X|) < \infty$. We wish to define $E(X|\mathcal{G})$ where the sigma algebra $\mathcal{G} \subset \mathcal{F}$. First, for non-negative integrable X, we may choose as sequence of simple random variables $X_n \uparrow X$. Since simple random variables have only finitely many values, they have finite variance, and we can use the definition above for their conditional expectation. Then $E(X_n|\mathcal{G})$ is an increasing sequence of random variables and so it converges. Define $E(X|\mathcal{G})$ to be the limit. In general, for random variables taking positive and negative values, we define $E(X|\mathcal{G}) = E(X^+|\mathcal{G}) - E(X^-|\mathcal{G})$. There are a number of details that need to be ironed out. First we need to show that this new definition is consistent with the old one when the random variable happens to be square integrable. We can also show that the properties (a)-(i) above all hold under this new definition of conditional expectation. We close with the more common definition of conditional expectation found in most probability and measure theory texts, essentially property (d) above. It is, of course, equivalent to the definition as a projection that we used above when the random variable is square integrable and when it is only integrable, reduces to the aforementioned limit of the conditional expectations of simple functions.

Theorem A25

Consider a random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ for which $E(|X|) < \infty$. Suppose the sigma algebra $\mathcal{G} \subset \mathcal{F}$. Then there is a unique (almost surely P) $\mathcal{G}-$measurable random variable Z satisfying

$$\int_A X dP = \int_A Z dP \text{ for all } A \in \mathcal{G}$$

Any such Z we call the conditional expectation and denote by $E(X|\mathcal{G})$.

2.2 Martingales in Discrete Time

In this section all random variables are defined on the same probability space $(\Omega, \mathcal{F}, P)$. Partial information about these random variables may be obtained from the observations so far, and in general, the "history" of a process up to time t is expressed through a sigma-algebra $H_t \subset \mathcal{F}$. We are interested in stochastic processes or sequences of random variables called martingales, intuitively, the total fortune of an individual participating in a "fair game". In order for the game to be "fair", the expected value of your future fortune given the history of the process up to and including the present should be equal to your present wealth. In a sense you are neither tending to increase or decrease your wealth over time- any fluctuations are purely random. Suppose your fortune at time s is denoted X_s. The values of the process of interest and any other related processes up to time s generate a sigma-algebra H_s. Then the assertion that the game is fair implies that the expected value of our future fortune given this history of the process up to the present is exactly our present wealth $E(X_t|H_s) = X_s$ for $t > s$. In what follows, we will sometimes state our definitions to cover both the discrete time case in which t ranges through the integers $\{0, 1, 2, 3, ...\}$ or a subinterval of the real numbers like $T = [0, \infty)$. In either case, T represents the set of possible indices t.

Definition

$\{(X_t, H_t); t \in T\}$ is a *martingale* if

(a) H_t is increasing (in t) family of sigma-algebras

(b) Each X_t is H_t- measurable and $E|X_t| < \infty$.

(c) For each $s < t$, where $s, t \in T$, we have $E(X_t|H_s) = X_s$ a.s.

Example

Suppose Z_t are independent random variables with expectation 0. Define $H_t = \sigma(Z_1, Z_2, ..., Z_t)$ for $t = 1, 2, ...$and $S_t = \sum_{i=1}^t Z_i$. Then Notice that for integer $s < t$,

$$\begin{aligned} E[S_t|H_s] &= E[\sum_{i=1}^t Z_i|H_s] \\ &= \sum_{i=1}^t E[Z_i|H_s] \\ &= \sum_{i=1}^s Z_i \end{aligned}$$

because $E[Z_i|H_s] = Z_i$ if $i \leq s$ and otherwise if $i > s$, $E[Z_i|H_s] = 0$. Therefore $\{(S_t, H_t),\ t = 1, 2, ...\}$ is a (discrete time) martingale. As an exercise you might show that if $E(Z_t^2) = \sigma^2 < \infty$, then $\{(S_t^2 - t\sigma^2, H_t), t = 1, 2, ...\}$ is also a discrete time martingale.

Example

Let X be any integrable random variable, and H_t an increasing family of sigma-algebras for t in some index set T. Put $X_t = E(X|H_t)$. Then notice that for $s < t$,

$$E[X_t|H_s] = E[E[X|H_t]|H_s] = E[X|H_s] = X_s$$

so (X_t, H_t) is a martingale.

Technically a sequence or set of random variables is not a martingale unless each random variable X_t is integrable. Of course, unless X_t is integrable, the concept of conditional expectation $E[X_t|H_s]$ is not even defined. You might think of reasons in each of the above two examples why the random variables X_t above and S_t in the previous example are indeed integrable.

Definition

Let $\{(M_t, H_t); t = 1, 2, ...\}$ be a martingale and A_t be a sequence of random variables measurable with respect to $H_{t-1.}$ Then the sequence A_t is called **non-anticipating** *(*an alternate term is ***predictable*** but this will have a slightly different meaning in continuous time*)*.

In gambling, we must determine our stakes and our strategy on the $t'th$ play of a game based on the information available to use at time $t-1$. Similarly, in investment, we must determine the weights on various components in our portfolio at the end of day (or hour or minute) $t-1$ *before* the random marketplace determines our profit or loss for that period of time. In this sense both gambling and investment strategies must be determined by non-anticipating sequences of random variables (although both gamblers and investors often dream otherwise).

Definition(Martingale Transform).

Let $\{\{(M_t, H_t), t = 0, 1, 2, ...\}$ be a martingale and let A_t be a bounded non-anticipating sequence with respect to H_t. Then the sequence

$$\tilde{M}_t = A_1(M_1 - M_0) + ... + A_t(M_t - M_{t-1}) \tag{2.2}$$

is called a *Martingale transform* of M_t.

The martingale transform is sometimes denoted $A \circ M$ and it is one simple transformation which preserves the martingale property.

Theorem A26

The martingale transform $\{(\tilde{M}_t, H_t), t = 1, 2, ...\}$ is a martingale.

Proof.

$$\begin{aligned} E[\widetilde{M}_j - \widetilde{M}_{j-1}|H_{j-1}] &= E[A_j(M_j - M_{j-1}|H_{j-1}] \\ &= A_j E[(M_j - M_{j-1}|H_{j-1}] \text{ since } A_j \text{ is } H_{j-1} \text{ measurable} \\ &= 0 \text{ a.s.} \end{aligned}$$

Therefore

$$E[\widetilde{M}_j|H_{j-1}] = \widetilde{M}_{j-1} \text{ a.s.}$$

■

Consider a random variable τ that determines when we stop betting or investing. Its value can depend arbitrarily on the outcomes in the past, as long as the decision to stop at time $\tau = t$ depends only on the results at time $t, t-1,$ ...etc. Such a random variable is called an optional stopping time.

Definition

A random variable τ taking values in $\{0, 1, 2, ...\} \cup \{\infty\}$ is a (optional) *stopping time* for a martingale $\{(X_t, H_t), t = 0, 1, 2, ...\}$ if for each n, $[\tau \leq t] \in H_t$.

If we stop a martingale at some random stopping time, the result continues to be a martingale as the following theorem shows.

Theorem A27

Suppose that $\{(M_t, H_t), t = 1, 2, ...\}$ is a martingale and τ is an optional stopping time. Define a new sequence of random variables $Y_t = M_{t\wedge\tau} = M_{\min(t,\tau)}$ for $t = 0, 1, 2, ...$Then $\{(Y_t, H_t), t = 1, 2, ..\}$ is a martingale.

Proof. Notice that

$$M_{t\wedge\tau} = M_0 + \sum_{j=1}^{t}(M_j - M_{j-1})I(\tau \geq j).$$

Letting $A_j = I(\tau \geq j)$ this is a bounded $H_{j-1}-$measurable sequence and therefore $\sum_{j=1}^{n}(M_j - M_{j-1})I(\tau \geq j)$ is a martingale transform. By Theorem A26 it is a martingale. ■

Example (Ruin probabilities)

A random walk is a sequence of partial sums of the form $S_n = S_0 + \sum_{i=1}^{n} X_i$ where the random variables X_i are independent identically distributed. Suppose that $P(X_i = 1) = p$, $P(X_i = -1) = q$, $P(X_i = 0) = 1 - p - q$ for $0 < p + q \leq 1$, and $p \neq q$. This is a model for our total fortune after we play n games, each game independent, and resulting either with a win of \$1, a loss of \$1 or break-even (no money changes hand). However we assume that the game is not fair, so that the probability of a win and the probability of a loss are different. We can show that

$$M_t = (q/p)^{S_t}, t = 0, 1, 2, ...$$

is a martingale with respect to the usual history process $H_t = \sigma(X_1, Z_2, ..., X_t)$. Suppose that our initial fortune lies in some interval $A < S_0 < B$ and define the optional stopping time τ as the first time we hit either of two barriers at A or B. Then $M_{t\wedge\tau}$ is a martingale. Suppose we wish to determine the probability of hitting the two barriers A and B in the long run. Since $E(M_\tau) = \lim_{t\to\infty} E(M_{t\wedge\tau}) = (q/p)^{S_0}$ by dominated convergence, we have

$$(q/p)^A p_A + (q/p)^B p_B = (q/p)^{S_0} \tag{2.3}$$

where p_A and $p_B = 1 - p_A$ are the probabilities of hitting absorbing barriers at A or B respectively. Solving, it follows that

$$((q/p)^A - (q/p)^B)p_A = (q/p)^{S_0} - (q/p)^B \tag{2.4}$$

or that

$$p_A = \frac{(q/p)^{S_0} - (q/p)^B}{(q/p)^A - (q/p)^B}. \tag{2.5}$$

In the case $p = q$, a similar argument provides

$$p_A = \frac{B - S_0}{B - A}. \tag{2.6}$$

These are often referred to as ruin probabilities, and are of critical importance in the study of the survival of financial institutions such as Insurance firms.

Definition

For an optional stopping time τ define the sigma algebra corresponding to the history up to the stopping time H_τ to be the set of all events $A \in H$ for which

$$A \cap [\tau \leq t] \in H_t, \text{ for all } t \in T. \tag{2.7}$$

Theorem A28

H_τ is a sigma-algebra.

Proof. Clearly since the empty set $\varphi \in H_t$ for all t, so is $\varphi \cap [\tau \leq t]$ and so $\varphi \in H_\tau$. We also need to show that if $A \in H_\tau$ then so is the complement A^c. Notice that for each n,

$$\begin{aligned} &[\tau \leq t] \cap \{A \cap [\tau \leq t]\}^c \\ &\quad = [\tau \leq t] \cap \{A^c \cup [\tau > t]\} \\ &\quad = A^c \cap [\tau \leq t] \end{aligned}$$

and since each of the sets $[\tau \leq t]$ and $A \cap [\tau \leq t]$ are H_t-measurable, so must be the set $A^c \cap [\tau \leq t]$. Since this holds for all t it follows that whenever $A \in H_\tau$ then so A^c. Finally, consider a sequence of sets $A_m \in H_\tau$ for all $m = 1, 2, ...$ We need to show that the countable union $\cup_{m=1}^\infty A_m \in H_\tau$. But

$$\{\cup_{m=1}^\infty A_m\} \cap [\tau \leq t] = \cup_{m=1}^\infty \{A_m \cap [\tau \leq t]\}$$

and by assumption the sets $\{A_m \cap [\tau \leq t]\} \in H_t$ for each t. Therefore

$$\cup_{m=1}^\infty \{A_m \cap [\tau \leq t]\} \in H_t$$

and since this holds for all t, $\cup_{m=1}^\infty A_m \in H_\tau$. ■

There are several generalizations of the notion of a martingale that are quite common. In general they modify the strict rule that the conditional expectation of the future given the present $E[X_t|H_s]$ is exactly equal to the present value X_s for $s < t$. The first, a submartingale, models a process in which the conditional expectation satisfies an inequality compatible with a game that is either fair or is in your favour so your fortune is expected either to remain the same or to increase.

Definition

$\{(X_t, H_t); t \in T\}$ is a *submartingale* if

(a) H_t is increasing (in t) family of sigma-algebras.

(b) Each X_t is H_t measurable and $E|X_t| < \infty$.

(c) For each $s < t$,, $E(X_t|H_s) \geq X_s$ a.s.

Note that every martingale is a submartingale.

There is a very useful inequality, Jensen's inequality, referred to in most elementary probability texts. Consider a real-valued function $\phi(x)$ which has the property that for any $0 < p < 1$, and for any two points x_1, x_2 in the domain of the function, the inequality

$$\phi(px_1 + (1-p)x_2) \leq p\phi(x_1) + (1-p)\phi(x_2)$$

holds. Roughly this says that the function evaluated at the average is less than the average of the function at the two end points or that the line segment joining the two points $(x_1, \phi(x_1))$ and $(x_2, \phi(x_2))$ lies above or on the graph of the function everywhere. Such a function is called a *convex function*. Functions like $\phi(x) = e^x$ and $\phi(x) = x^p, p \geq 1$ are convex functions but $\phi(x) = \ln(x)$ and $\phi(x) = \sqrt{x}$ are not convex (in fact they are concave). Notice that if a random variable X took two possible values x_1, x_2 with probabilities $p, 1-p$ respectively, then this inequality asserts that the function at the point $E(X)$ is less than or equal $E\phi(X)$, i.e.

$$\phi(EX) \leq E\phi(X)$$

There is also a version of Jensen's inequality for conditional expectation which generalizes this result, and we will prove this more general version.

Theorem A29 (Jensen's Inequality)

Let ϕ be a convex function. Then for any random variable X and sigma-field H,

$$\phi(E(X|H)) \leq E(\phi(X)|H). \tag{2.8}$$

Proof. Consider the set $\mathcal{L}$ of linear function $L(x) = a + bx$ that lie entirely below the graph of the function $\phi(x)$. It is easy to see that for a convex function

$$\phi(x) = \sup\{L(x); L \in \mathcal{L}\}.$$

For any such line,

$$\begin{aligned} E(\phi(X)|H) &\geq E(L(X)|H) \\ &\geq L(E(X)|H)). \end{aligned}$$

If we take the supremum over all $L \in \mathcal{L}$, we obtain

$$E(\phi(X)|H) \geq \phi(E(X)|H)).$$

■

The standard version of Jensen's inequality follows on taking H above to be the trivial sigma-field. Now from Jensen's inequality we can obtain a relationship among various commonly used norms for random variables. Define the norm $||X||_p = \{E(|X|^p)\}^{1/p}$ for all $p \geq 1$. The norm allows us to measure distances between two random variables, for example a distance between X and Y can be expressed as

$$||X - Y||_p$$

It is well known that

$$||X||_p \leq ||X||_q \text{ whenever } 1 \leq p < q \tag{2.9}$$

This is easy to show since the function $\phi(x) = |x|^{q/p}$ is convex provided that $q \geq p$ and by the Jensen's inequality,

$$E(|X|^q) = E(\phi(|X|^p) \geq \phi(E(|X|^p)) = |E(|X|^p)|^{q/p}.$$

A similar result holds when we replace expectations with conditional expectations. Let X be any random variable and H be a sigma-field. Then for $1 \leq p \leq q < \infty$

$$\{E(|X|^p|H)\}^{1/p} \leq \{E(|X|^q|H)\}^{1/q}. \tag{2.10}$$

Proof. Consider the function $\phi(x) = |x|^{q/p}$. This function is convex provided that $q \geq p$ and by the conditional form of Jensen's inequality,

$$E(|X|^q|H) = E(\phi(|X|^p)|H) \geq \phi(E(|X|^p|H)) = |E(|X|^p|H)|^{q/p} \text{ a.s.}$$

■

In the special case that H is the trivial sigma-field, this is the inequality

$$||X||_p \leq ||X||_q. \tag{2.11}$$

Theorem A30 (Constructing Submartingales).

Let $\{(S_t, H_t), t = 1, 2, ...\}$ be a martingale. Then $(|S_t|^p, H_t)$ is a submartingale for any $p \geq 1$ provided that $E|S_t|^p < \infty$ for all t.Similarly $((S_t - a)^+, H_t)$ is a submartingale for any constant a.

Proof. Since the function $\phi(x) = |x|^p$ is convex for $p \geq 1$, it follows from the conditional form of Jensen's inequality that

$$E(|S_{t+1}|^p|H_t) = E(\phi(S_{t+1})|H_t) \geq \phi(E(S_{t+1}|H_t)) = \phi(S_t) = |S_t|^p \text{ a.s.}$$

■

Various other operations on submartingales will produce another submartingale. For example if X_n is a submartingale and ϕ is a convex nondecreasing function with $E\phi(X_n) < \infty$, Then $\phi(X_n)$ is a submartingale.

Theorem A31 (Doob's Maximal Inequality)

Suppose (M_n, H_n) is a non-negative submartingale. Then for $\lambda > 0$ and $p \geq 1$,

$$P(\sup_{0 \leq m \leq n} M_m \geq \lambda) \leq \lambda^{-p} E(M_n^p)$$

Proof. We prove this in the case $p = 1$. The general case we leave as a problem. Define a stopping time

$$\tau = \min\{m; M_m \geq \lambda\}$$

so that $\tau \leq n$ if and only if the maximum has reached the value λ by time n or

$$P[\sup_{0 \leq m \leq n} M_m \geq \lambda] = P[\tau \leq n].$$

Now on the set $[\tau \leq n]$, the maximum $M_\tau \geq \lambda$ so

$$\lambda I(\tau \leq n) \leq M_\tau I(\tau \leq n) = \sum_{i=1}^{n} M_i I(\tau = i). \tag{2.12}$$

By the submartingale property, for any $i \leq n$ and $A \in H_i,$

$$E(M_i I_A) \leq E(M_n I_A).$$

Therefore, taking expectations on both sides of (2.12), and noting that for all $i \leq n$,

$$E(M_i I(\tau = i)) \leq E(M_n I(\tau = i))$$

we obtain

$$\lambda P(\tau \leq n) \leq E(M_n I(\tau \leq n)) \leq E(M_n).$$

■

Once again define the norm $||X||_p = \{E(|X|^p)\}^{1/p}$. Then the following inequality permits a measure of the norm of the maximum of a submartingale.

Theorem A32 (Doob's L^p Inequality)

Suppose (M_n, H_n) is a non-negative submartingale and put $M_n^* = \sup_{0 \leq m \leq n} M_n$. Then for $p > 1$, and all n

$$||M_n^*||_p \leq \frac{p}{p-1} ||M_n||_p$$

One of the main theoretical properties of martingales is that they converge under fairly general conditions. Conditions are clearly necessary. For example consider a simple random walk $S_n = \sum_{i=1}^{n} Z_i$ where $Z_i$are independent identically distributed with $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}$. Starting with an arbitrary value of S_0, say $S_0 = 0$ this is a martingale, but as $n \to \infty$ it does not converge almost surely or in probability.

On the other hand, consider a Markov chain with the property that $P(X_{n+1} = j|X_n = i) = \frac{1}{2i+1}$ for $j = 0, 1, ..., 2i$. Notice that this is a martingale and beginning with a positive value, say $X_0 = 10$, it is a non-negative martingale. Does it converge almost surely? If so the only possible limit is $X = 0$ because the nature of the process is such that $P[|X_{n+1} - X_n| \geq 1|X_n = i] \geq \frac{2}{3}$ unless $i = 0$. Convergence to $i \neq 0$ is impossible since in that case there is a high probability of jumps of magnitude at least 1. However, X_n does converge almost surely, a consequence of the martingale convergence theorem. Does it converge in L_1 i.e. in the sense that $E[|X_n - X|] \to 0$ as $n \to \infty$? If it did, then $E(X_n) \to E(X) = 0$ and this contradicts the martingale property of the sequence which implies $E(X_n) = E(X_0) = 10$. This is an example of a martingale that converges almost surely but not in L_1.

Lemma

If $(X_t, H_t), t = 1, 2, ..., n$ is a (sub)martingale and if α, β are optional stopping times with values in $\{1, 2, ..., n\}$ such that $\alpha \leq \beta$ then

$$E(X_\beta | H_\alpha) \geq X_\alpha$$

with equality if X_t is a martingale.

Proof. It is sufficient to show that

$$\int_A (X_\beta - X_\alpha) dP \geq 0$$

for all $A \in H_\alpha$. Note that if we define $Z_i = X_i - X_{i-1}$ to be the submartingale differences, the submartingale condition implies

$$E(Z_j | H_i) \geq 0 \text{ a.s. whenever } i < j.$$

Therefore for each $j = 1, 2, ...n$ and $A \in H_\alpha$,

$$\begin{aligned}
\int_{A \cap [\alpha = j]} (X_\beta - X_\alpha) dP &= \int_{A \cap [\alpha = j]} \sum_{i=1}^{n} Z_i I(\alpha < i \leq \beta) dP \\
&= \int_{A \cap [\alpha = j]} \sum_{i=j+1}^{n} Z_i I(\alpha < i \leq \beta) dP \\
&= \int_{A \cap [\alpha = j]} \sum_{i=j+1}^{n} E(Z_i | H_{i-1}) I(\alpha < i) I(i \leq \beta) dP \\
&\geq 0 \text{ a.s.}
\end{aligned}$$

since $I(\alpha < i)$, $I(i \leq \beta)$ and $A \cap [\alpha = j]$ are all measurable with respect to H_{i-1} and $E(Z_i | H_{i-1}) \geq 0$ a.s. If we add over all $j = 1, 2, ..., n$ we obtain the desired result. ■

The following inequality is needed to prove a version of the submartingale convergence theorem.

Theorem A33 (Doob's upcrossing inequality)

Let M_n be a submartingale and for $a < b$, define $N_n(a, b)$ to be the number of complete upcrossings of the interval (a, b) in the sequence $M_j, j = 0, 1, 2, ..., n$. This is the largest k such that there are integers $i_1 < j_1 < i_2 < j_2 ... < j_k \leq n$ for which

$$M_{i_l} \leq a \text{ and } M_{j_l} \geq b \text{ for all } l = 1, ..., k.$$

Then

$$(b - a) E N_n(a, b) \leq E\{(M_n - a)^+ - (M_0 - a)^+\}$$

Proof. By Theorem A29, we may replace M_n by $X_n = (M_n - a)^+$ and this is still a submartingale. Then we wish to count the number of upcrossings of the interval $[0, b']$ where $b' = b - a$. Define stopping times for this process by $\alpha_0 = 0$,

$$\begin{aligned}
\alpha_1 &= \min\{j; 0 \le j \le n, X_j = 0\} \\
\alpha_2 &= \min\{j; \alpha_1 \le j \le n, X_j \ge b'\} \\
&\ldots \\
\alpha_{2k-1} &= \min\{j; \alpha_{2k-2} \le j \le n, X_j = 0\} \\
\alpha_{2k} &= \min\{j; \alpha_{2k-1} \le j \le n, X_j \ge b'\}.
\end{aligned}$$

In any case, if α_k is undefined because we do not again cross the given boundary, we define $\alpha_k = n$. Now each of these random variables is an optional stopping time. If there is an upcrossing between X_{α_j} and $X_{\alpha_{j+1}}$ (where j is odd) then the distance travelled

$$X_{\alpha_{j+1}} - X_{\alpha_j} \ge b'.$$

If X_{α_j} is well-defined (i.e. it is equal to 0) and there is no further upcrossing, then $X_{\alpha_{j+1}} = X_n$ and

$$X_{\alpha_{j+1}} - X_{\alpha_j} = X_n - 0 \ge 0.$$

Similarly if j is even, since by the above lemma, $(X_{\alpha_j}, H_{\alpha_j})$ is a submartingale,

$$E(X_{\alpha_{j+1}} - X_{\alpha_j}) \ge 0.$$

Adding over all values of j, and using the fact that $\alpha_0 = 0$ and $\alpha_n = n$,

$$E\sum_{j=0}^{n}(X_{\alpha_{j+1}} - X_{\alpha_j}) \ge b' E N_n(a, b)$$

$$E(X_n - X_0) \ge b' E N_n(a, b).$$

In terms of the original submartingale, this gives

$$(b - a) E N_n(a, b) \le E(M_n - a)^+ - E(M_0 - a)^+.$$

■

Doob's martingale convergence theorem that follows is one of the nicest results in probability and one of the reasons why martingales are so frequently used in finance, econometrics, clinical trials and lifetesting.

Theorem A34 (Sub)martingale Convergence Theorem.

Let (M_n, H_n); $n = 1, 2, \ldots$ be a submartingale such that $sup_{n\to\infty} EM_n^+ < \infty$. Then there is an integrable random variable M such that $M_n \to M$ a.s. If $\sup_n E(|M_n|^p) < \infty$ for some $p > 1$ then $||M_n - M||_p \to 0$.

Proof. The proof is an application of the upcrossing inequality. Consider any interval $a < b$ with rational endpoints. By the upcrossing inequality,

$$E(N_a(a, b)) \le \frac{1}{b-a} E(M_n - a)^+ \le \frac{1}{b-a}[|a| + E(M_n^+)]. \tag{2.13}$$

Let $N(a, b)$ be the total number of times that the martingale completes an upcrossing of the interval $[a, b]$ over the infinite time interval $[1, \infty)$ and note that $N_n(a, b) \uparrow N(a, b)$ as $n \to \infty$. Therefore by monotone convergence $E(N_a(a, b)) \to EN(a, b)$ and by (2.13)

$$E(N(a, b)) \leq \frac{1}{b-a} \limsup[a + E(M_n^+)] < \infty.$$

This implies

$$P[N(a, b) < \infty] = 1.$$

Therefore,

$$P(\liminf M_n \leq a < b \leq \limsup M_n) = 0$$

for every rational $a < b$ and this implies that M_n converges almost surely to a (possibly infinite) random variable. Call this limit M. We need to show that this random variable is almost surely finite. Because $E(M_n)$ is non-decreasing,

$$E(M_n^+) - E(M_n^-) \geq E(M_0)$$

and so

$$E(M_n^-) \leq E(M_n^+) - E(M_0).$$

But by Fatou's lemma

$$E(M^+) = E(\liminf M_n^+) \leq \liminf EM_n^+ < \infty$$

Therefore $E(M^-) < \infty$ and consequently the random variable M is finite almost surely. The convergence in L^p norm follows from the results on uniform integrability of the sequence. ■

Theorem A35 (L^p martingale Convergence Theorem)

Let $(M_n, H_n);\ n = 1, 2, \ldots$ be a martingale such that $\sup_{n \to \infty} E|M_n|^p < \infty, p > 1$. Then there is an random variable M such that $M_n \to M$ a.s. and in L^p.

Example (The Galton-Watson process)

. Consider a population of Z_n individuals in generation n each of which produces a random number ξ of offspring in the next generation so that the distribution of Z_{n+1} is that of $\xi_1 + ... + \xi_{Z_n}$ for independent identically distributed ξ. This process $Z_n, n = 1, 2, ...$ is called the Galton-Watson process. Let $E(\xi) = \mu$. Assume we start with a single individual in the population $Z_0 = 1$ (otherwise if there are j individuals in the population to start then the population at time n is the sum of j independent terms, the offspring of each). Then the following properties hold:

1. The sequence Z_n/μ^n is a martingale.
2. If $\mu < 1$, $Z_n \to 0$ and $Z_n = 0$ for all sufficiently large n.
3. If $\mu = 1$ and $P(\xi \neq 1) > 0$, then $Z_n = 0$ for all sufficiently large n.
4. If $\mu > 1$, then $P(Z_n = 0$ for some $n) = \rho$ where ρ is the unique value < 1 satisfying $E(\rho^\xi) = \rho$.

Definition (supermartingale)

$\{(X_t, H_t); t \in T\}$ is a *supermartingale* if

(a) H_t is an increasing (in t) family of sigma-algebras.

(b) Each X_t is H_t measurable and $E|X_t| < \infty$.

(c) For each $s < t$, $s, t \in T$, $E(X_t|H_s) \leq X_s$ a.s.

The properties of supermartingales are very similar to those of submartingales, except that the expected value is a non-increasing sequence. For example if $A_n \geq 0$ is a predictable (non-anticipating) bounded sequence and (M_n, H_n) is a supermartingale, then the supermartingale transform $A \circ M$ is a supermartingale. Similarly if in addition the supermartingale is non-negative $M_n \geq 0$ then there is a random variable M such that $M_n \to M$ a.s. with $E(M) \leq E(M_0)$. The following example shows that a non-negative supermartingale may converge almost surely and yet not converge in expected value.

Example

Let S_n be a simple symmetric random walk with $S_0 = 1$ and define the optional stopping time $N = \inf\{n; S_n = 0\}$. Then

$$X_n = S_{n \wedge N}$$

is a non-negative (super)martingale and therefore X_n converges almost surely. The limit (call it X) must be 0 because if $X_n > 0$ infinitely often, then $|X_{n+1} - X_n| = 1$ for infinitely many n and this contradicts the convergence. However, in this case, $E(X_n) = 1$ whereas $E(X) = 0$ so the convergence is not in the L_1 norm (in other words, $||X - X_n||_1 \nrightarrow 0$) or in expected value.

A martingale under a reversal of the direction of time is a reverse martingale. The sequence $\{(X_t, H_t); t \in T\}$ is a *reverse martingale* if

(a) H_t is decreasing (in t) family of sigma-algebras.

(b) Each X_t is H_t- measurable and $E|X_t| < \infty$.

(c) For each $s < t$, $E(X_s|H_t) = X_t$ a.s.

It is easy to show that if X is any integrable random variable, and if H_t is any decreasing family of sigma-algebras, then $X_t = E(X|H_t)$ is a reverse martingale. Reverse martingales require even fewer conditions than martingales do for almost sure convergence.

Theorem A36 (Reverse martingale convergence Theorem).

If $(X_n, H_n);\ n = 1, 2, \ldots$ is a reverse martingale, then as $n \to \infty$, X_n converges almost surely to the random variable $E(X_1 | \cap_{n=1}^{\infty} H_n)$.

The reverse martingale convergence theorem can be used to give a particularly simple proof of the strong law of large numbers because if $Y_i, i = 1, 2, \ldots$ are independent identically distributed random variables and we define H_n to be the sigma algebra $\sigma(\bar{Y}_n, Y_{n+1}, Y_{n+2}, \ldots)$, where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$, then H_n is a decreasing family of sigma fields and $\bar{Y}_n = E(Y_1|H_n)$ is a reverse martingale.

2.3 Uniform Integrability

Definition

A set of random variables $\{X_i, i = 1, 2, \ldots.\}$ is uniformly integrable if

$$\sup_i E(|X_i| I(|X_i| > c) \to 0 \text{ as } c \to \infty$$

Some Properties of uniform integrability:

1. Any finite set of integrable random variables is uniformly integrable.
2. Any infinite sequence of random variables which converges in L^1 is uniformly integrable.
3. Conversely if a sequence of random variables converges almost surely and is uniformly integrable, then it also converges in L^1.
4. If X is integrable on a probability space (Ω, H) and H_t any family of sub-sigma fields, then $\{E(X|H_t)\}$ is uniformly integrable.
5. If $\{X_n, n = 1, 2, \ldots\}$ is uniformly integrable, then $\sup_n E(X_n) < \infty$.

Uniform integrability is the bridge between convergence almost surely or in probability and convergence in expectation as the following results shows.

Theorem A37

Suppose a sequence of random variables satisfies $X_n \to X$ in probability. Then the following are all equivalent:

1. $\{X_n, n = 1, 2, \ldots\}$ is uniformly integrable
2. $X_n \to X$ in L^1.
3. $E(|X_n|) \to E(|X|)$

As a result a uniformly integrable submartingale $\{X_n, n = 1, 2, ...\}$ not only converges almost surely to a limit X as $n \to \infty$ but it converges in expectation and in L^1 as well, in other words $E(X_n) \to E(X)$ and $E(|X_n - X|) \to 0$ as $n \to \infty$. There is one condition useful for demonstrating uniform integrability of a set of random variables:

Lemma

Suppose there exists a function $\phi(x)$ such that $\lim_{x\to\infty} \phi(x)/x = \infty$ and $E\phi(|X_t|) \leq B < \infty$ for all $t \geq 0$. Then the set of random variables $\{X_t; t \geq 0\}$ is uniformly integrable.

One of the most common methods for showing uniform integrability, used in results like the Lebesgue dominated convergence theorem, is to require that a sequence of random variables be dominated by a single integrable random variable X. This is, in fact a special use of the above lemma because if X is an integrable random variable, then there exists a convex function $\phi(x)$ such that $\lim_{x\to\infty} \phi(x)/x = \infty$ and $E(\phi(|X|) < \infty$.