SOME APPLICATIONS OF MARTINGALES TO PROBABILITY THEORY

WATSON LADD

Abstract. Martingales are a very simple concept with wide application in probability. We introduce the concept of a martingale, develop the theory, and use the theory to prove some important and interesting theorems from probability theory.

Contents

1. Motivation

In the early eighteenth century the martingale betting strategy was very popular in France[8]. Each time a gambler lost he would bet enough to make up all their previous bets. Before the introduction of table limits this seemed like a sure bet, as the eventual win would reverse all of ones losses. But what would happen when you could not make the next bet?

In roulette, one is gambling on which sector of a wheel a small white ball will come to rest in. The roulette wheel has 38 sectors. 18 are black, 18 are red, and two are green. Each time roulette is played the wheel is spun, bets are placed, and the ball is dropped onto the spinning wheel. Friction gradually slows down the wheel, and the ball eventually comes to rest in one of the slots.

The gambler bets on whether the color of the slot the ball rests in will be red or black, with even odds. If the gambler puts down k dollars, they receive $2k$ if they are correct and otherwise receive 0.

Suppose X_i is the gambler's money after i bets. X_i is X_{i-1} plus or minus the amount the gambler bet. But while the results of the roulette wheel are independent, the bets the gambler makes are not, and so X_i is not the sum of a sequence of independent random variables. However, the gambler expects with probability

Date: August 26, 2011.

18/38 to make his bet, and 20/38 to lose it. Therefore the expected value of X_{i+1}, knowing the past of the wheel and hence of the gambler's purse, is less then X_i was. The gambler is always expecting to lose money on the next bet. At some time, T he will stop, and we want to know what the expected value of his purse is.

Our intuition tells us that this expected value must be less then the money he started with. If he is losing francs every time he takes a spin, then he should not be able to win francs overall. This intuition is correct under very weak conditions.

This problem is not the only motivation for studying martingales. Martingales are well behaved enough to have many theorems about their behavior, but can still encompass a wide range of processes. Using martingale techniques we can determine when sums of random variables converge as well as solve problems about Markov chains and other probabilistic situations.

2. Foundation

Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. Define I_G for a subset G of Ω to be the function such that I_G is 1 on G and 0 on the rest of Ω. This I_G is called the indicator function of G. We write

$$E(X;G) = \int_G X d\mu = E(XI_G),$$

the conditional expectation of X given G.

Often we will write $\mu(S)$ where S is a formula involving a random variable. Then we mean the measure of the set of elements of Ω that make S true.

Random variables $X_1, X_2, \ldots$ are identically distributed when $\mu(X_i^{-1}([a,b])) = \mu(X_j^{-1}([a,b]))$ for all i and j.

Independence is an important concept of elementary probability that can be adapted to our setting. Events A, B in some σ-algebra are independent when $\mu(A \cap B) = \mu(A)\mu(B)$. For k events $A_1, \ldots A_k$ we would say they are independent when $\mu(A_1 \cap \ldots \cap A_k) = \prod_{n=1}^{k} \mu(A_n)$.

A countable set of events is independent when every finite subset of events is independent. A countable set of σ-algebras $\mathscr{F}_1 \ldots \mathscr{F}_n$ is independent when every sequence of events $A_i \in \mathscr{F}_i$ is independent.

Definition 2.1. Let X be a random variable. The σ-algebra generated by X is the smallest σ-algebra containing the sets $X^{-1}((a,b))$ for all real a and b, written $\sigma(X)$.

Definition 2.2. Let A, B be sets of subsets of Ω. Then $\sigma(A, B)$ is the smallest σ-algebra containing A and B.

Definition 2.3. A set of random variables $X_1, \ldots,$ is independent if $\sigma(X_1), \ldots$ is independent.

IID random variables are random variables that are independent and identically distributed.

Let X be a random variable on $\mathscr{F}$. Suppose $E(|X|) < \infty$. Then we say X is in $\mathcal{L}^1$. $\mathcal{L}^1(\Omega)$, the space of all $\mathcal{L}^1$ functions has a norm on it, defined by $\|X\|_1 = E(|X|)$. It is a vector space as well as a complete metric space with the metric induced by the norm, and hence a Banach space. Details of this construction are in [2].

Let $\mathscr{G} \subset \mathscr{F}$ also be a σ-algebra. Then we want to define a $\mathscr{G}$ measurable random variable that is "all the information" we know about X if we are given some event in $\mathscr{G}$.

The existence of such a random variable is a consequence of the Radon-Nikodým theorem. First suppose X positive. Define $\nu(B) = \int_B X \mathrm{d}\mu$. Then ν is an absolutely continuous measure with respect to μ so the Radon-Nikodým theorem tells us that there exists a $\mathscr{G}$ measurable Y such that $\int_B Y \mathrm{d}\mu = \int_B X \mathrm{d}\mu$ for all B in $\mathscr{G}$. To extend to general random variables we split a random variable into its positive and negative part.

From this we have the existence up to a set of measure zero of a $\mathscr{G}$ measurable Y such that $E(XI_G) = E(YI_G)$ for all $G \in \mathscr{G}$ and that such a Y is unique up to a set of measure zero. We write $Y = E(X \mid \mathscr{G})$ to be such a random variable. and call it the conditional expectation with respect to the σ-algebra $\mathscr{G}$.

Conditional expectation satisfies a number of important laws which can be proven from the definition of conditional probability without delving deep into analysis.

Theorem 2.4.
(1) *If Y is a version of $E(X \mid \mathscr{G})$ then $E(Y) = E(X)$*
(2) *If X is $\mathscr{G}$ measurable then $E(X \mid \mathscr{G}) = X$ almost surely*
(3) *$E(a_1X_1 + a_2X_2) = a_1E(X_1 \mid \mathscr{G}) + a_2E(X_2 \mid \mathscr{G})$ almost surely*
(4) *If $X \geq 0$, then $E(X \mid \mathscr{G}) \geq 0$*
(5) *If X_n is an increasing sequence of random variables converging to X pointwise, then $E(X_n \mid \mathscr{G})$ converges monotonically upwards to $E(X \mid \mathscr{G})$ almost surely.*
(6) *If $|X_n(\omega)| < V(\omega)$ for all n, $E(V) < \infty$ and X_n converge to X pointwise, $E(X_n \mid \mathscr{G})$ converges to $E(X \mid \mathscr{G})$*
(7) *If c is a convex function on $\mathbb{R}$ and $E(|c(X)|) < \infty$ then*
$$E(c(X) \mid \mathscr{G}) \geq c(E(X \mid \mathscr{G}))$$
(8) *If $\mathscr{H} \subset \mathscr{G}$, then*
$$E(E(X \mid \mathscr{G}) \mid \mathscr{H}) = E(X \mid \mathscr{H})$$
(9) *If Z is a $\mathscr{G}$ measurable and bounded random variable,*
$$E(ZX \mid \mathscr{G}) = ZE(X \mid \mathscr{G})$$
almost surely
(10) *If $\mathscr{H}$ is independent of $\sigma(\sigma(X), \mathscr{G})$, then $E(X \mid \sigma(\mathscr{G}, \mathscr{H})) = E(X \mid \mathscr{G})$ almost surely*

Proof. See [7]. □

3. Martingales and Stopping Times

Much of this discussion follows [7].

When we are given a sequence of random variables on some probability space there is often an interpretation in terms of events taking place over time, for instance independent observations or values of time varying values. As time goes on we gain more information, and therefore have a larger σ algebra of distinguishable events.

For example if $\mathscr{F}$ is the powerset of the set of sequences of N zeros or ones, we have a sequence of σ-algebras where $\mathscr{F}_i$ is the σ-algebra generated by the sets of coin flips that agree on the first i flips. As time goes on, properties that we previously could not determine, such as "does this sequence have three heads in the first eight flips?" become ones that we are certain about. The σ-algebras represent the world of questions we can answer at point i based on what we know.

Definition 3.1. A *filtered σ-algebra* is a σ-algebra $\mathscr{F}$ and a sequence of σ-algebras $\mathscr{F}_n$ such that $\mathscr{F}_0 \subset \mathscr{F}_1 \ldots \subset \mathscr{F}$. Such a sequence is called a *filtration* on $\mathscr{F}$.

Definition 3.2. A sequence X_n of random variables is adapted with respect to some filtration $\mathscr{F}_0, \mathscr{F}_1, \ldots$ if X_i is $\mathscr{F}_i$-measurable.

In our coin flip example let X_i be the purse of a gambler gambling on each coin flip. His wagers can be set arbitrarily, but the result of his gambling at time i is not dependent on the future coin flips. This condition is a very natural one, and implies that X_i is adapted.

Definition 3.3. A *martingale* is a sequence of random variables $X_1, X_2 \ldots$ and a filtration $\mathscr{F}_1, \mathscr{F}_2, \ldots$ such that

1. X_i is adapted.
2. $E(|X_i|) < \infty$ for all i
3. $E(X_{i+1} \mid \mathscr{F}_i) = X_i$ almost surely

We will sometimes drop the filtration when it is clear from the context or if we are creating a number of martingales with the same filtration.

Definition 3.4. A *supermartingale* is a sequence of random variables $X_1, X_2 \ldots$ such that

1. X_i is adapted.
2. $E(|X_i|) < \infty$ for all i
3. $E(X_{i+1} \mid \mathscr{F}_i) \leq X_i$

Definition 3.5. A *submartingale* is a sequence of random variables M_n such that $-M_n$ is a supermartingale.

What if our gambler varies his bets? He can only vary his bets depending on the history of the system up to time n. Let C_n be his n-th bet, and assume he starts betting at time 1. Then C_n is $\mathscr{F}_{n-1}$ measurable. If the gambler could change his bet knowing the outcome of the next coin toss or spin of the wheel, he could make a lot of money.

Definition 3.6. If C_n is a sequence of random variables such that C_n is measurable with respect to $\mathscr{F}_{n-1}$, C_n is a *previsible process*

Definition 3.7. A stopping time T is a random variable such that $\{\omega \mid T(\omega) = n\}$ is $\mathscr{F}_n$ measurable.

We will have to know when we have to stop at step n at step n so we do not play at $n+1$. This definition again rules out using knowledge about the future. "The time when I have the most money" is not a stopping time while "The time when I have more money then before" is.

Definition 3.8. If C is previsible, and X is a martingale, we define $(C \circ X)_n = \sum_{j=1}^{k} C_j(X_j - X_{j-1})$, the *martingale transform* of X by C.

Proposition 3.9. *The martingale transform of a martingale by a bounded previsible process is a martingale.*

Proof. Let $Y = (C \circ X)$. Then

$$E(Y_n - Y_{n-1} \mid \mathscr{F}_{n-1}) = E(C_n(X_n - X_{n-1}) \mid \mathscr{F}_{n-1}).$$

C_n is $\mathscr{F}_{n-1}$ measurable so we can pull it out and get $C_n E(X_n - X_{n-1} \mid \mathscr{F}_{n-1}) = 0$. So $E(Y_n \mid \mathscr{F}_{n-1}) = Y_{n-1}$, and we have that the Y_n form a martingale. □

Proposition 3.10. *If C_n is a positive previsible process, then the martingale transform of a supermartingale remains a supermartingale.*

Proof. Carry on as above, but change the last line to $C_n E(X_n - X_{n-1} \mid \mathscr{F}_{n-1}) \leq 0$, therefore $E(Y_n \mid \mathscr{F}_{n-1}) = Y_{n-1}$. □

Definition 3.11. Let T be a stopping time. We define the stopped process X^T by

$$X_n^T = \begin{cases} X_n & \text{if } n \leq T, \\ X_T & \text{if } n > T. \end{cases}$$

Stopping is a martingale transform. Our bettor bets until the stopping time tells him to stop, that is he puts down 1 until T, then puts down 0 thereafter. Therefore the stopped martingale is a martingale by our previous proposition about martingale transforms. X_T is the value of X at time T, a random variable. X_T might not always exist, as T might be infinite. However, X^T, a sequence of random variables, always exists: if we do not stop, it is simply X.

The next result is Doob's Optional Stopping Theorem.

Theorem 3.12. *If X is a supermartingale and T a stopping time, $E(X_T) \leq E(X_0)$ if*

(1) *X is bounded,*
(2) *T is bounded,*
(3) *or $E(T) < \infty$ and $|X_n(\omega) - X_{n-1}(\omega)|$ is bounded by K.*

Proof. Consider X_n^T. We have that $E(X_n^T - X_0) \leq 0$ by induction. Now if T is bounded by N then $E(X_T) \leq E(X_0)$ as the sequence X_n^T converges to $E(X^T)$ uniformly. If X is bounded then the X_n^T are converging to X_T and are bounded, so the dominated convergence theorem applies, giving $E(X_T) \leq E(X_0)$. In the last case we have

$$|X_n^T - X_0| = |\sum_{k=1}^{\min\{T,n\}} (X_k - X_{k-1})| \leq KT.$$

$E(KT) < \infty$ so we can use the dominated convergence theorem on $X_n^T - X_0$ to get the desired result. □

All betting strategies are therefore doomed to fail in the real world. Casinos have bet limits, so any betting strategy that you expect to take less then infinite time will have to lose money by the third part of Doob's Optional Stopping Theorem, provided that the game is always not in your favor. Even if the odds of the game are varying, so long as they remain unfavorable there is no opportunity to make money.

4. Martingale Convergence and Sums of Random Variables

We say a sequence of random variables is bounded in $\mathcal{L}^1$ if there is a real K such that $E(|X_n|) < K$ for all n.

Theorem 4.1. *If X_n is a supermartingale bounded in $\mathcal{L}^1$ then*

$$X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$$

exists almost surely.

Proof. First we define an *upcrossing.* Let $[a, b]$ be an interval, and fix ω in some underlying probability space. Let X be a supermartingale. The number of upcrossings by time N is defined to be the largest k such that $0 \le s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k \le N$ and $X_{s_i}(\omega) < a$, $X_{t_i}(\omega) > b$. An upcrossing is a pair of times between which the value of the sequence $X_n(\omega)$ goes from below a to above b. Let $U_N[a, b]$ be the number of upcrossings by time N.

We now use a martingale transform argument to get information about upcrossings. If our gambler bets 1 starting when X_n is below a, and quitting when X_n is above b they will make $b - a$ for each upcrossing. If they continue this betting strategy until time N they will only lose $a - X_N$ after making $b - a$ for each upcrossing. The total amount they have won until time n is a supermartingale by the proposition on martingale transforms. Therefore the Optional Stopping Theorem applies so we get $(b - a)E(U_N[a, b]) \le E(\max\{a - X_N, 0\})$, the Doob Upcrossing Lemma.

We now consider the number of upcrossings over all time when X is an $\mathcal{L}^1$ - bounded martingale. Clearly $U_N[a, b]$ is an increasing function of N. Using our lemma we have $(b - a)EU_N[a, b] \le |a| + E(|X_n|) \le |a| + \sup_n E(|X_n|)$. Using the monotone convergence theorem we have $(b - a)EU_\infty[a, b] \le |a| + \sup_n E(|X_n|)$.

We now want to know the probability of $X_n(\omega)$ converging. In a style of proof very common in analysis we will look at the measure of a "bad" set, which we split up into sets of "increasing badness".

$$\begin{aligned}
\Lambda &= \{\omega \mid X_n(\omega) \text{does not converge to a limit in} [-\infty, \infty]\} \\
&= \{\omega \mid \liminf X_n(\omega) < \limsup X_n(\omega)\} \\
&= \bigcup_{\{a,b\in\mathbb{Q}|a<b\}} \{\omega \mid \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\} \\
&= \bigcup_{\{a,b\in\mathbb{Q}|a<b\}} \Lambda_{a,b}
\end{aligned}$$

Now $\Lambda_{a,b} \subset \{\omega \mid U_\infty[a, b](\omega) = \infty\}$. However, $E(U_\infty[a, b])$ is finite, so $\mu(\Lambda_{a,b}) = 0$. Therefore X_∞ almost surely exists. □

5. Uniform Integrability and Martingales

We begin with a general note on random variables. If $X \in \mathcal{L}^1$ then for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\mu(F) < \delta$ implies $E(XI_F) < \epsilon$. The proof of this statement is not hard but is not enlightening.

Given this result we can show that for any $X \in \mathcal{L}^1$ and $\epsilon > 0$ there exists a $K \in \mathbb{R}$ such that $E(|X|; |X| > K) < \epsilon$. Note that $KP(|X| > K) \le E(|X|)$ so that we can make $\mu(|X| > K) < \delta$ by picking a large enough K. $\mathcal{L}^1$ variables are not overly "spread out": we can get as much of the expectation of the variable within some bounded range.

Definition 5.1. A family of random variables $\mathcal{C}$ is *uniformly integrable* if for each $\epsilon > 0$ we can pick a real K such that for all $X \in \mathcal{C}$, $E(|X|; |X| > K) < \epsilon$.

This immediately implies that a set of uniformly integrable random variables is bounded in $\mathcal{L}^1$, as $E(|X|) = \epsilon + K$ because the expectation on the region where $|X| < K$ must be less then K, while outside that region the expectation is ϵ.

Uniform integrability goes nicely with conditioning on subalgebras.

Lemma 5.2. *Let $X \in \mathcal{L}^1$. Then the class $\{E(X \mid \mathscr{G}) \mid \mathscr{G}\text{sub-}\sigma\text{-algebra of } \mathscr{F}\}$ is UI*

Proof. Let $\epsilon > 0$. Then choose $\delta > 0$ such that for $F \in \mathscr{F}$ $\mu(F) < \delta$ implies $E(|X| \mid F) < \epsilon$. Then choose K such that $E(|X|) < \delta K$. Now let $\mathscr{G}$ be a subσ-algebra of $\mathscr{F}$, and $Y = E(X \mid \mathscr{G})$. By Jensen's inequality $|Y| \leq E(|X| \mid \mathscr{G})$, so $E(|Y|) \leq E(|X|)$, and $K\mu(|Y| > K) \leq E(|Y|) \leq E(|X|)$ so $\mu(|Y| > K) < \delta$. This implies

$$E(|Y| \mid |Y| \geq K) \leq E(|X| \mid |Y| \geq K) < \epsilon.$$

□

Uniform integrability makes almost sure convergence into $\mathcal{L}^1$ convergence.

Theorem 5.3. *If X_n is a sequence of random variables in $\mathcal{L}^1$ and $X_n \to X$ almost surely, and X_n is UI, then $E(|X_n - X|) \to 0$.*

Proof. Pick $\epsilon > 0$. Let

$$\phi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if } -K \leq x \leq K \\ -K & \text{if } x \leq -K \end{cases}$$

Then we can pick a K such that

$$E(|\phi_K(X_n) - X_n|) < \epsilon/3,$$

and so that

$$E(|\phi_K(X) - X|) < \epsilon/3.$$

Now as $|\phi_K(x) - \phi_K(y)| \leq |x - y|$, $\phi_K(X_n) \to \phi_K(X)$ almost surely. But $\phi_K(X_n)$ is bounded. Pick an n_0 such that $\mu(|\phi_K(X_n) - \phi_K(X)|) < \frac{\epsilon}{3K}$ for $n > n_0$. Then $E(|\phi_K(X_n) - \phi_K(X)|) < \epsilon/3$.

Then by the triangle inequality $E(|X_n - X|) < \epsilon$ and we are done. □

This theorem can be immediately combined with Doob's Martingale Convergence Theorem. Recall that uniform integrability implies being bounded in $\mathcal{L}^1$.

Proposition 5.4. *Let M be a UI martingale. Then $M_\infty = \lim M_n$ exists almost surely and $E(|M_\infty - M_n|) \to 0$.*

Proof. Since M is a $\mathcal{L}^1$ bounded martingale it converges to M_∞ almost surely. But since it is UI this limit is a limit in $\mathcal{L}^1$. □

M_∞ is related to M_n in another way as well.

Lemma 5.5. *If M_n is a UI martingale and $M_\infty = \lim_{n\to\infty} M_n$ then $M_n = E(M_\infty|\mathscr{F}_n)$.*

Proof. Note that for $F \in \mathscr{F}_n$ and $r \geq n$ we have $E(M_r; F) = E(M_n; F)$. Now $E(|M_\infty - M_r|; F) \leq E(|M_\infty - M_r|)$. By the triangle inequality and some substitution $|E(M_\infty; F) - E(M_n; F)| < E(|M_\infty - M_r|)$ for all r. But this implies $E(M_\infty; F) = E(M_n; F)$, which is what we needed to prove □

For our next theorem we will need to define the limit of a series of σ- algebras. Let $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \mathscr{F}$ be a filtration, then define $\mathscr{F}_\infty$ to be the union of the $\mathscr{F}_i$.

Theorem 5.6 (Lévy Upwards Theorem)**.** *Let $\xi \in \mathcal{L}^1(\Omega, \mathscr{F}, \mu)$. Let $\mathscr{F}_n$ be a filtration. Define $M_n = E(\xi \mid \mathscr{F}_n)$. Then M_n is a martingale and M_n converges to $\eta = E(\xi \mid \mathscr{F}_\infty)$ almost surely and in $\mathcal{L}^1$.*

Proof. We begin by looking at $Y = E(M_{n+1} \mid \mathscr{F}) = E(E(\xi \mid \mathscr{F}_{n+1}) \mid \mathscr{F}_n)$. This is a random variable that is $\mathscr{F}_n$ measurable and $E(Y; F) = E(\xi; F)$ for $F \in \mathscr{F}_n$. Therefore this random variable is almost surely M_n. Each M_n is bounded in $\mathcal{L}^1$ by 2.4. Therefore M_n is a martingale. It is uniformly integrable by our earlier lemma. Ergo M_∞ exists and is in $\mathcal{L}^1$.

Now we need only show $M_\infty = E(\xi \mid \mathscr{F}_\infty)$. Let $F \in \mathscr{F}_\infty$. Then $F \in \mathscr{F}_n$ for some n. By above $E(M_\infty \mid F) = E(M_n \mid F) = E(E(\xi \mid \mathscr{F}_n) \mid F) = E(\xi \mid F)$. So then $M_\infty = E(\xi \mid \mathscr{F}_\infty)$ almost surely. □

The Lévy Upwards Theorem is a powerful tool for reconstructing a random variable from various observations of it. Each observation expands the σ-algebra we know about, and so the Lévy Upwards Theorem implies that our best estimates of ξ, $E(\xi \mid \mathscr{F}_n)$ converge to $E(\xi \mid \mathscr{F}_\infty)$, which is our estimate knowing all of the information.

Definition 5.7. Let $\mathscr{F}_1, \mathscr{F}_2, \ldots$ be a sequence of σ-algebras such that $\mathscr{F}_i \supset \mathscr{F}_j$ whenever $i < j$. Then the *tail algebra* of the $\mathscr{F}_i$ is $\cap_j \mathscr{F}_j$. The tail algebra of a sequence of random variables $X_1, X_2, \ldots$ is the tail algebra of $\mathscr{F}_i = \sigma(X_i, X_{i+1}, \ldots)$.

Tail algebras are very special algebras, as the Kolmogorov 0-1 law shows.

Theorem 5.8 (Kolmogorov 0-1 Law)**.** *Suppose $X_1, X_2, \ldots$ is a sequence of independent random variables. Let $\mathscr{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. Then if F is in the tail algebra of the X_i, $\mu(F) = 0$ or 1.*

Proof. Let $\eta = I_F$. Define $\mathscr{F}_n = \sigma(X_1, \ldots X_n)$. Then let $M_n = E(\eta \mid \mathscr{F}_n)$. $\eta = E(\eta \mid \mathscr{F}_\infty)$ because $\mathscr{F}_\infty$ contains T. Note $E(\eta \mid \mathscr{F}_\infty) = \lim_{n\to\infty} E(\eta \mid \mathscr{F}_n)$ by the Lévy Upward Theorem. Now, η is $\mathscr{T}_n$ measurable. But if we look at the definitions of $\mathscr{T}_n$ and $\mathscr{F}_n$ we see that η is independent of $\mathscr{F}_n$, as F is independent of all the events in $\mathscr{F}_n$ for any n. So $E(\eta \mid \mathscr{F}_n) = E(\eta) = \mu(F)$. So $\eta = \mu(F)$, but as η is either 0 or 1 we are done. □

If we have a bunch of monkeys we can conclude that the probability they retype the works of Shakespeare infinitely often is either 0 or 1. Many times the Kolmogorov Zero-One Law is useful it is very hard to determine which of 0 or 1 the answer is. Having two choices sometimes helps.

The Lévy Upward Theorem covers increasing sequences of σ-algebras. In the Kolmogorov 0-1 Law we would have liked to have a theorem directly applicable to the case of decreasing sequences of σ-algebras.

Theorem 5.9 (Lévy Downards Theorem)**.** *Let $\mathscr{G}_{-n} \subset \ldots \subset \mathscr{G}_{-1} \subset \mathscr{F}$. Define $\mathscr{G}_{-\infty} = \cap_k \mathscr{G}_{-k}$. Let γ be a $\mathcal{L}^1$ random variable. Then*

$$E(\gamma \mid \mathscr{G}_{-\infty}) = \lim_{n\to\infty} E(\gamma \mid \mathscr{G}_{-n}).$$

Proof. Let $M_{-n} = E(\gamma \mid \mathscr{G}_{-n})$. As a direct consequence of 2.4 M_{-n} is a martingale if we start at some fixed $-N$ and go until M_{-1}. We can apply the Doob Upcrossing Lemma to show that this converges to M_∞ and then carry out a very similar proof as in the Upward Theorem. □

From this we can prove the celebrated Strong Law of Large Numbers.

Theorem 5.10. *Let $X_1, X_2, \ldots$ be independent identically distributed variables, with $E(|X_1|) < \infty$ and $m = E(X_1)$. Then define $S_n = \sum_{i=1}^n X_i$. Then $n^{-1}S_n \to m$ almost surely, and $E(n^{-1}S_n - m) \to 0$.*

Proof. Write $\mathscr{G}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots)$. Then we have $E(X_1|\mathscr{G}_{-n}) = E(X_1|\sigma(S_n))$ by 2.4. However as $E(X_1|\sigma(S_n)) = E(X_i|\sigma(S_n))$ for $i < n$ by independence and identical distribution we have

$$E(X_1|\sigma(S_n)) = n^{-1}E(X_1 + X_2 + \ldots + X_n|\sigma(S_n)) = n^{-1}S_n$$

Now $E(X_1 \mid \mathscr{G}_{-n})$ satisfies the conditions of the Lévy Downward Theorem. Therefore $L = \lim n^{-1}S_n$ exists and is a limit in $\mathcal{L}^1$. $L = \limsup \frac{X_{k+1}+\ldots+X_{k+n}}{n}$ for each k. This lets us apply Kolmogorov's 0-1 Law to get that L is a constant, as $L = c$ does not depend on the first k S_k for any k. But $L = E(L) = m$. □

This theorem is widely applicable to statistics. If we think of our X_n as independent observations of a process where each observation has the same statistics, then we can apply to the strong law to see that averaging all the observations we have will gradually approach the mean of the distribution.

6. Exchangability

We will now follow [4] and prove the Hewitt-Savage theorem about symmetric random variables.

We define Σ as the group of permutations on $\mathbb{N}$ that only move finitely many elements of $\mathbb{N}$. Let $X_1, X_2, \ldots$ be independant identically distributed random variables. Suppose $A = f(X_1, X_2, \ldots)$ such that $A \circ f_\sigma = A$ for all $f_\sigma \in \Sigma$ and A in $\mathcal{L}^1$. We call such an A symmetric. Let $\mathscr{T}$ be the tail algebra of the X_i. All of this is a prequel to the following.

Theorem 6.1 (Hewitt-Savage 0-1 Law)**.** *With A as above, $A = E(A \mid \mathscr{T})$ almost surely.*

Proof. Let $\mathscr{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. Let $B_n = E(A \mid \mathscr{F}_n)$. Then by the Lévy Upwards Theorem B_n converges in $\mathscr{L}^1$ to A. Now define $\mathscr{T}_n = \sigma(X_n, X_{n+1}, \ldots)$. Let $C_n = E(A \mid \mathscr{T}_n)$. Then by the Lévy Downwards Theorem $D = E(A \mid \mathscr{T}) = \lim C_n$ in L^1. By another application of the Lévy Downwards Theorem $E(B_n \mid \mathscr{T}_m)$ converges to $E(B_n \mid \mathscr{T})$ as m increases.

Now fix $\epsilon > 0$. Then we have an N and a M such that for all $n > N$, $m > M$, and $n > m$ we have $E(|B_n - A|) < \epsilon/3$, $E(|C_m - D|) < \epsilon/3$, and $E(|B_n - E(B_n \mid \mathscr{T}_m)|) < \epsilon/3$. Let ρ be a permutation that exchanges $1, \ldots n$ with $m, \ldots n + m - 1$. Then

$$B_n = E(A \mid \mathscr{F}_n) = E((A \circ \rho) \mid \mathscr{F}_n).$$

But now we are conditioning on variables that are in the slots $m, \ldots, m+n-1$ and we can take $E(A \mid \sigma(X_m, \ldots X_{m+n-1}))$ instead by invoking Fubini's theorem. But this σ algebra is a subset of $\mathscr{T}_m$.Ergo

$$E(B_n \mid \mathscr{T}_m)E(E((A \circ \rho) \mid \mathscr{F}_n) \mid \mathscr{T}_m) = E(A \mid \mathscr{T}_m).$$

But then we have $E(|B_n - C_m|) < \epsilon/3$, and by the triangle inequality $E(|A-D|) < \epsilon/3$. □

Our next application requires the concept of exchangability. Let $X_1, X_2 \ldots$ be random variables over some $(\Omega, \mathscr{F}, \mu)$. We say the X_i are exchangeable when for any $[a_n, b_n]$, and a permutation ρ that fixes all but finitely many n, $\mu(\bigcap_k X_k^{-1}([a_k, b_k])) = \mu(\bigcap_k X_{\rho(k)}^{-1}[a_k, b_k])$ where $[a_k, b_k]$ are real intervals.

Notice that exchangability implies identically distributed, but does not necessarily imply independent. For example we can have all X_i either 0 or 1 depending on the same coin being heads or tails. However, IID random variables are exchangable.

We can get a result classifying all exchangable random variables, a result called de Finetti's Theorem. There are many formulations of de Finetti's Theorem. This one is essentially found in [1] but is proved with ergodic theory there.

Theorem 6.2 (de Finetti's Theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of exchangable random variables in $\mathcal{L}^1$. Let $\mathscr{T}$ be the tail algebra. Then $E(X_1 \mid \mathscr{T}), E(X_2 \mid \mathscr{T}), \ldots$ is an independent identically distributed sequence of random variables.*

Proof. Let $\mathscr{T}_m = \sigma(X_m, X_{m+1}, \ldots)$. Then $\mathscr{T}$ is the intersection of all the $\mathscr{T}_m$. Clearly $E(X_i \mid \mathscr{T})$ is an exchangeable sequence since the X_i are exchangeable. Independence of $X_1 \ldots X_n$ is equivalent to $X_1 \leq x_1, X_2 \leq x_2, \ldots X_n \leq x_n$ being independent events. Let $f_n(x)$ be the indicator function of $x \leq x_n$. Then we must show that

$$E(f_1(X_1) \ldots f_n(X_n) \mid \mathscr{T}) = E(f_1(X_1)\mathscr{T}) \ldots E(f_n(X_n)\mathscr{T}).$$

Consider $S_m^k = f_k(X_1) + f_k(X_2) + \ldots + f_k(X_m)$. By exchangability and a symmetry argument we have $E(f_k(X_i) \mid \mathscr{T}_m) = S_m^k/m$. Then applying the Lévy Downward Theorem we have that

$$E(f_k(X_i) \mid \mathscr{T}) = \lim_{m\to\infty} S_m^k/m.$$

Then

$$E(f_1(X_1) \mid \mathscr{T}) \ldots E(f_n(X_n) \mid \mathscr{T}) = \lim_{m\to\infty} \prod_{k=1}^n S_m^k/m.$$

Now

$$m^{-n} \prod_{k=1}^n S_m^k = m^{-n} \sum_{1 \leq m_1, \ldots m_n \leq n} \prod f_{m_i}(X_i).$$

Each term of the sum is either 0 or 1, and there are m^{n-1} terms that have two or more indices the same. As we take the limit the contributions of these terms vanishes since we are dividing by m^n and we are left with

$$\lim_{m\to\infty} \prod_{k=1}^n S_m^k/m = \lim_{m\to\infty} \frac{1}{m(m-1)\ldots(m-n+1)} \sum_{m_i \text{nonequal}} \prod_{i=1}^n f_i(X_{m_i})$$
$$= \lim_{m\to\infty} E(f_1(X_1)f_2(X_2)\ldots f_n(X_n) \mid \mathscr{T})$$

□

de Finetti's theorem is important in Bayesian statistics. Suppose we have a sequence of exchangeable Bernoulli random variables, that is variables that are either 0 or 1. de Finetti's theorem tells us that they are conditionally independent given the tail algebra, which is determined by the mean of the Bernoulli random variables.

Furthermore, the joint probability distribution of the observations and the mean is easy to find because of conditional independence. We therefore can find the probability distribution of the mean of the exchangeable Bernoulli random variables given a finite number of samples through Bayes' Theorem. Absent de Finetti's theorem we would have to assume the conditional independence to get any useful results. Bayesian methods and the use of de Finetti's theorem are developed further in [6].

7. Random Walks, Markov Chains, and Martingales

In what follows we will follow [3] up to a point. Let S be a countable set. We say a sequence of random variables $X_1, X_2, \ldots$ taking values in S is *Markov* when $E(X_i \mid \sigma(X_{i-1}, X_{i-2}, \ldots)) = E(X_i \mid \sigma(X_{i-1}))$. Such a sequence is called a Markov chain or discrete time Markov process. For simplicity we will assume $\mu(X_n = k \mid X_{n-1} = j) = \mu(X_m = k \mid X_{m-1} = j)$ for all m and n, that is to say our chain is time homogeneous. A state in S is absorbing if once entered it is never left, and nonabsorbing otherwise. We have a matrix T with entries $t_{ij} = \mu(X_{k+1} = i; X_k = j)$, the transition matrix for the Markov chain. Let P be the operator that takes a function $f : S \to \mathbb{R}$ to $P(f)(x) = \sum_i f(i)t_{ix}$. We can think of P as averaging f over the possible next states, with the weights given by the Markov chain. We say f is superharmonic when $f \geq Pf$, and subharmonic when $f \leq Pf$. f is harmonic when $f = Pf$.

When we write $\sigma(X_1, \ldots X_n)$ this is the σ-algebra generated by the events $X_i = s_i$. This definition requires S to be discrete.

Theorem 7.1. *If f is superharmonic with respect to a Markov chain then $f(X_n)$ is a supermartingale with respect to $\sigma(X_1, \ldots, X_n)$, and similarly for submartingales and martingales.*

Proof. Note that without loss of generality f is superharmonic as we can take $-f$ instead if necessary. Clearly $f(X_n)$ is adapted with respect to the given filtration, and $E(f(X_{n+1}) \mid \sigma(X_1, \ldots X_n)) = E(f(X_{n+1}) \mid \sigma(X_n))$ by the Markov property. But this is just $P(f)(X_n)$ and the result follows. □

Many gambling games are gambling on Markov chains, such as craps. Given a known function $f : S \to \mathbb{R}$ and a Markov chain $X_1, X_2, \ldots$ with known transition matrix we play a game where we can cash out at any time and earn $f(X_i)$ when we do so or stay in, hopefully increasing our payout. Ideally we would make the choices that maximize the payout when we do cash out. This is only an interesting problem when there are some states that do not lead to others, as otherwise we can just keep playing until we get to the state with the most payout.

First note that any reasonable strategy under these conditions is just a previsible stopping time T. Secondly note that such a stopping time should only depend on the current state, as the past doesn't matter to the Markov process.

Therefore such a stopping time is a partition of the state space S into S_1 and S_2. One stops in S_1 and continues play in S_2. The challenge is to calculate S_1 and S_2. We will write X_n^T to be the sequence of states we are in when playing using the strategy T. Naturally we stop at some point. We define an optimal strategy to be a stopping time (or strategy) that maximizes $E(f(X_T))$. The value of a state is the expected payout when we use an optimal strategy and our Markov chain starts in that state.

Theorem 7.2. *Let T be an optimal strategy. Let $V(s)$ be the value of state s under strategy T. Then $V(X_1^T), V(X_2^T), \ldots$ is a martingale, and for a suboptimal strategy S $V(X_1^S), V(X_2^S) \ldots$ is a supermartingale.*

Proof. For an optimal strategy $E(V(X_{n+1}^T) \mid \mathscr{F}_n) = V(X_n^T)$ as $X_{n+1}^T = X_n^T$ if we have stopped, and otherwise we are playing onwards, in which case the value of our current position is $E(V(X_{n+1}^T) \mid \mathscr{F}_n)$. Any suboptimal strategy will be equivalent to an optimal strategy until the suboptimal strategy says to stop and the optimal one says to go on, or vice versa. If the suboptimal strategy stops, $X_{n+1}^S = X_n^S$ and the martingale property holds. If it does not stop $E(V(X_{n+1}^S) \mid \mathscr{F}_n) \leq V(X_n^S)$ because otherwise an optimal strategy would not tell us to stop at that point. □

Our proof has relied on a simple observation, that an optimal strategy is the best strategy no matter what state we are in. Clearly $V(s)$ does not depend on which optimal strategy we pick. Once we know $V(s)$ we will stop if $f(s) = V(s)$ and otherwise continue playing. Optimal strategies only differ when $f(X_i) = E(V(X_{i+1}) \mid \mathscr{F}_n)$, that is when going on and stopping have the same value.

This theorem shows that $V(S)$ is the lowest superharmonic function with respect to P such that $V \geq f$. As a result we can find $V(S)$ as the limit of $u_i(x) = \max\{Pu_{i-1}, f(x)\}$ where u_0 is initially $\max\{f(x)\}$ for non absorbing states and $f(x)$ for absorbing states. This is essentially reverse induction: since we know what the optimal strategy is for states that are absorbing, we can work backwards to states that are transient, and eventually we will get to all states.

Our penultimate theorem considers random walks on groups. Suppose we have an abelian group G and a probability measure μ on G such that there are IID random variables $X_1, X_2, \ldots$ taking values in G such that $\mu(X_i \in S) = \mu(S)$ for all measurable subsets S in G. Now we can define a Markov chain $Y_i = \sum_{k=1}^n X_i$. This is not a Markov chain in our usual sense as the state space is not finite anymore, but it does satisfy the appropriate generalization of the Markov property. We again have an operator P, but in this case $(Pf)(x) = \int_G f(x+y)\mu(y)$. We define the support of μ to be the smallest set with $\mu(S) = 1$.

Theorem 7.3 (Choquet-Deny)**.** *For a Markov chain as above all bounded harmonic functions are constant on cosets of the group generated by the support of μ.*

Proof. We follow an exercise in [5]. Let $Z_1, \ldots$ be IID random variables on G with distributions μ. Let ϕ be a harmonic function, and fix $x \in G$ Then let $M_n = \phi(x + Z_1 + Z_2 + \ldots + Z_n)$. Let $\mathscr{F}_n = \sigma(Z_1, \ldots Z_n)$. Then M_n is a martingale. Since ϕ is bounded M_n is bounded, hence a $\mathcal{L}^1$ bounded martingale and UI as well. So M_n converges in $\mathcal{L}^1$ to M_∞. But permuting the Z_i does not change M_∞ as it does not change M_n beyond the last modified index. So M_∞ is symmetric in the $Z_1, \ldots Z_n$ and hence constant by the Hewitt-Savage 0-1 Law.

But $M_n = E(M_\infty \mid \mathscr{F}_n)$, so M_n is almost surely constant, and in particular $\phi(x + Z_1) = \phi(x)$ almost surely since both are almost surely constant and $E(\phi(x +$

$Z_1)) = \phi(x)$. Ergo $\phi(y) = \phi(x)$ for all y in the same coset of the group generated by the support of μ as x except for those in some set of measure zero. Note $\phi(y) = E(\phi(y + Z_1))$ for all y and $E(\phi(y + Z_1))$ ignores sets of measure zero. So ϕ is constant everywhere on the coset of the group generated by the support of μ containing x. □

The function $f(x, y) = x$ is harmonic with respect to any measure that is positive only on the x axis. The Markov chain does not "see" where f is not constant because it never goes off the horizontal line it started on. This example shows why we need to constantly worry about cosets in stating and proving the theorem.

As an application of this theorem we will prove a theorem from complex analysis.

Theorem 7.4. *Let $f : \mathbb{C} \to \mathbb{C}$ be an analytic function. Then if f is bounded, f is constant.*

Proof. Since f is analytic $f(x)$ is equal to the mean of f over a ball around x. Let μ be the uniform probability measure on $\bar{B}(0, 1)$. Then f is harmonic with respect to the random walk with law μ. But then f is constant on the cosets of the support of μ by Choquet-Deny. However the support of this measure is all of $\bar{B}(0, 1)$ which generates $\mathbb{C}$. Ergo f is constant on $\mathbb{C}$. □

Acknowledgments

I would like to thank my REU mentors Shawn Drenning and Casey Rodrieguez for their guidance during the course of this REU project.

References

[1] Olav Kallenberg. Foundations of Modern Probability. Springer-Verlag, New York. 1997.
[2] A.N. Kolmogorov and S.V. Fomin. Introductory Real Analysis. Trans: Richard Silverman. Dover, New York. 1975.
[3] G.F. Lawler. Introduction to Stochastic Processes. Chapman and Hall, New York. 1995.
[4] Paul-André Meyer and Claude Dellacherie. Probabilités et Potentiel:Chapitres V à VIII. Hermann, Paris. 1980.
[5] D. Revuz. Markov Chains. Elsevier, Amsterdam. 1984.
[6] Mark J. Schervish. Theorey of Statistics. Springer-Verlag, New York. 1995.
[7] David Williams. Probability with Martingales. Cambridge University Press, Cambridge. 1991.
[8] Wikipedia contributors. "Martingale (betting system)". Cited August 2, 2011