Problem Set 7

Math 212a

Tuesday November 9, 2001, Due Nov. 18

In problem set **2** we developed a lot of probability theory using purely Hilbert space methods, i.e. no measure theory. Now that we have studied measure theory, we can go back and combine Hilbert space and measure theoretic methods to get some powerful results.

One key idea is the notion of conditional expectation. In Kolmogorov's formulation of the general form of this concept (see below), the existence of a conditional expectation is an immediate consequence of the Radon-Nikodym theorem which I hope to do in class next Tuesday. (If I am not mistaken, this Thursday is a University holiday.) We showed in problem set **2** that for L_2 random variables, conditional expectation is just orthogonal projection. So one of the things we will do here is redo the L_2 version so as to give us an alternative proof of the existence of conditional expectation. Of course I have to define the terms.

Once we have the notion of conditional expectation, it is easy to define a "martingale", in such a way that it generalizes the notion of an L_2 martingale as defined in the first problem in problem set **2**. The proof of the L_2 martingale convergence theorem as given in that problem, was quite transparent. The proof of a more general martingale convergence theorem is a bit harder.

Associated with the concept of a martingale is the notion of a stopping time. Roughly speaking, a stopping time is an "exit strategy" to use current political terminology - it is a decision in advance as to when to quit playing the game represented by the martingale. Doob's stopping time theorem says that under certain hypotheses, a stopping time can not help. You can not do better by playing with an exit strategy that not playing at all. The issues involved in the proof of this theorem are precisely some of the issues we have been discussing in class - when is the limit of an integral equal to the integral of a limit? That *some* condition is necessary was recognized by Bernoulli in his St. Petersburg paradox.

I will start with a pretty application of Doob's stopping time theorem. So I am not writing in logical order. You might want to skim through the whole problem set before beginning the problems.

Contents

1 Expected time until a pattern.

This is a problem of some theological importance. It is often said that if a "billion" monkeys sit in front of a typewriter each typing a letter at random once a second, "eventually" one will type out the text of Shakespeare's Julius Caeser. The question is - how long should we expect to wait until this happens?

1.1 The geometric distribution.

As a warm up question, one might ask the following: Suppose an experiment is repeated indefinitely and independently, and there are two possible outcomes: "A" with probability q and "B" with probability $p = 1 - q$.

1. What is the probability that B will occur for the first time on the k-th trial, where $k = 1, 2, \dots$? What is $E(k)$, the expected time until the first B appears?

1.2 The pattern matters.

When we ask how long should we expect to wait until a given pattern appears, the answer is more tricky. For instance, suppose that $p = q = \frac{1}{2}$, in the previous example.

What is the expected waiting time until the pattern AAAAAA appears? Answer: 126.

What is the expected waiting time until the pattern AABBAA appears? Answer: 70.

So we have to wait a shorter time for AABBAA on average than for AAAAAA even though the probability of the actual occurrence of these sequences in six consecutive trials is the same. For a purely combinatorial (but complicated) explanation of this computation, see Feller *Introduction to Probability Theory and its Applications, I* page 304. For a more conceptual version of this computation using "stopping times", look ahead. Indeed, I want to show (following Ross *Stochastic Processes, 2nd ed.* page 301) how this kind of problem can be solved in a very elementary and conceptual way using Doob's "Martingale stopping theorem". I will try to give an intuitive statement of the conclusion of this theorem. I will defer the statement of the hypotheses.

First some notation and some necessary results:

2 Conditional expectation and martingales.

2.1 Probability measures, probability triples.

Let Ω be a set and $\mathcal{F}$ be a σ-field of subsets of Ω and μ a measure on $\mathcal{F}$. We say that μ is a **probability measure** if $\mu(\Omega) = 1$. We will usually write P instead of μ for a probability measure. For example, we could take $\Omega = [0, 1]$, $\mathcal{F}$ to consist of the Lebesgue measurable sets, and P to be Lebesgue measure.

In what follows, it will frequently be the case that Ω and P are fixed, but we will be varying $\mathcal{F}$. So we will call $(\Omega, \mathcal{F}, P)$ a **probability triple**, or simpy a triple for short.

2.2 Conditional expectation.

Let $(\Omega, \mathcal{F}, P)$ be a probability triple. A real valued function on Ω which is measurable (relative to the Borel sets on $\mathbb{R}$) is called a (real valued) random variable. If $X \in L_1(\Omega, \mathcal{F}, P)$ then its expectation, denoted by $E(X)$ is just another way of writing its integral. i.e.

$$E(X) := \int_\Omega X dP.$$

Let $X \in L_1(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$.

Theorem 1 [**Kolmogorov.**] *There exists a random variable Y such that*

- *Y is $\mathcal{G}$ measurable,*
- *$Y \in L_1(\Omega, \mathcal{G}, P)$ and*
- $$\int_G X dP = \int_G Y dp \qquad \forall G \in \mathcal{G}.$$

Furthermore, if Z is another random variable with these properties, then $Z = Y$ almost everywhere.

Proof of uniqueness. We are assuming that $\int_G (Y-Z)dP = 0$ for all $G \in \mathcal{G}$. If it is not true that $Y = Z$ a.e., then either the set where $Y - Z > 0$ or the set where $Z - Y > 0$ has positive measure. Without loss of generality, we may assume that the set A where $Y - Z > 0$ has positive measure. The set where $Y - Z > 1/n$ belongs to $\mathcal{G}$, call it A_n. We have $A_n \nearrow A$ so we must have $P(A_n) > 0$ for some n. But then

$$\int_{A_n} (Y - Z) > \frac{1}{n} P(A_n) > 0$$

contradicting our assumption.

Proof of the existence using Radon-Nikodym. Write $X = X^+ - X^-$. If we can find Y^+ and Y^- that work we can choose $Y = Y^+ - Y^-$. So it is enough to establish the existence of Y when X is non-negative. Define the function on elements of $\mathcal{G}$ by

$$G \mapsto \int_G X dP.$$

It is immediate that this is a measure which is absolutely continuous relative to the measure P restricted to $\mathcal{G}$. Radon-Nikodym then guarantees the existence of a non-negative $Y \in L_1(\Omega, \mathcal{G}, P)$ such that

$$\int_G X dP = \int_G Y dP \quad \forall\, G \in \mathcal{G}. \quad \square$$

Proof of the existence using projections in Hilbert space and the monotone convergence theorem. Since I will not get to the Radon-Nikodym theorem until Tuesday Nov 16, here is an alternative proof using what we already know:

- **Proof when $X \in L_2$.** We know that $L_2(\Omega, \mathcal{G}, P)$ is complete, hence it is a closed subspace of $L_2(\Omega, \mathcal{F}, P)$. Let π denote orthogonal projection of $L_2(\Omega, \mathcal{F}, P)$ onto $L_2(\Omega, \mathcal{G}, P)$. Set

  $$Y = \pi(X).$$

  By the definition of $L_2(\Omega, \mathcal{G}, P)$, we know that Y is $\mathcal{G}$ measurable, and since $P(\Omega) = 1 < \infty$, we know that $L_2(\Omega, \mathcal{G}, P) \subset L_1(\Omega, \mathcal{G}, P)$, so $Y \in L_1(\Omega, \mathcal{G}, P)$. We must verify that $\int_G X dP = \int_G Y dp \quad \forall G \in \mathcal{G}$. Since $\mathbf{1}_G \in L_2(\Omega, \mathcal{F}, P)$ and $X - Y$ is orthogonal to $L_2(\Omega, \mathcal{F}, P)$ we know that

  $$(X - Y, \mathbf{1}_G) = 0$$

  or

  $$\int_G X dP = (X, \mathbf{1}_G) = (Y, \mathbf{1}_G) = \int_G Y dP$$

  as required. So we have again verified that that for $X \in L_2$ conditional expectation is exactly orthogonal projection. We now want to extend this result from L_2 to L_1. For this we first prove:

- *If U is a non negative bounded random variable (and so an element of L_2) then $W = E(U|\mathcal{G})$ is non-negative almost everywhere.* If not, there is some integer $n > 0$ such that $G \in \mathcal{G}$ has $P(G) > 0$ where

  $$G := \{x|W(x) \leq -\frac{1}{n}\}.$$

  But then

  $$0 \leq \int_G UdP = \int_G WdP \leq -\frac{1}{n}P(G) < 0,$$

  a contradiction.

- **Proof of the existence of conditional expectation for $X \in L_1$.** By splitting X into it positive and negative parts, we may assume that X is non-negative. Choose bounded non-negative random variables X_n with $X_n \nearrow X$. Let $Y_n = E(X_n|\mathcal{G})$ which we know exist and are non-negative and are increasing almost everywhere. Define Y by

  $$Y(\omega) = \lim Y_n(\omega).$$

  Then $Y_n \nearrow Y$ is $\mathcal{G}$) measurable, belongs to $L_1(\Omega, \mathcal{G}, P)$ and $\int_G XdP = \int_G Ydp \quad \forall G \in \mathcal{G}$ all by the monotone convergence theorem applied to $L_1(\Omega, \mathcal{G}, P)$ and to $L_1(\Omega, \mathcal{F}, P)$. $\square$

2. Take $\Omega = [0,1], \mathcal{F} =$ the Lebesgue measurable sets, and P to be Lebesgue measure. For a fixed n, let

$$G_0 := [0, \frac{1}{2^n}], \text{ and } G_k = (\frac{k}{2^n}, \frac{k+1}{2^n}] \; hbox for 1 \leq k < 2^n.$$

Let $\mathcal{G}_n$ be the σ-field generated by these sets, so $\mathcal{G}_n$ contains 2^{2^n} elements. Let X be the random variable $X(x) = x \; \forall x \in [0,1]$. What is $E(X|\mathcal{G}_n)$?

Of course, in this problem, there is nothing special about the particular random variable that we chose.

2.3 Some useful properties of conditional expectation.

1. The map $X \mapsto E(X, \mathcal{G}$ is linear. This is immediate from the defining properties.

2. If $\mathcal{H} \subset \mathcal{G}$ then

   $$E(E(X|\mathcal{G})|\mathcal{H})) = E(X|\mathcal{H}). \tag{1}$$

   This follows from the projection definition for $X \in L_2$ and then from the limiting definition for general $X \in L_1$.

3. If we take $\mathcal{H}$ to consist of just $\emptyset$ and Ω, then $E(X|\mathcal{H})$ is just the constant $E(X)$. So a special but very useful case of the preceding item is

$$E(E(X|\mathcal{G})) = E(X). \tag{2}$$

4. If Z is bounded and $\mathcal{G}$ measurable, then

$$E(ZX|\mathcal{G}) = ZE(X|\mathcal{G})\text{a.e..} \tag{3}$$

   Indeed, we need only prove this for non-negative X as usual. Both sides are $\mathcal{G}$ measurable. If $Z = \mathbf{1}_A$ for $A \in \mathcal{G}$ then for any $B \in \mathcal{G}$ we have

$$\int_B \mathbf{1}_A X dP = \int_{A\cap B} X dP = \int_{A\cap B} E(X|\mathcal{G}) dP = \int_B \mathbf{1}_A E(X|\mathcal{G}) dP.$$

   So the equation is true by the uniqueness. Then it is true for non-negative simple functions by linearity, and then true for all non-negative X by monotone convergence.

5. If X is independent of $\mathcal{G}$ then

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

   Recall that to say that X is independent of $\mathcal{G}$ means that for any $A \in \sigma(X)$, the σ-field determined by X, and any $B \in \mathcal{G}$ we have

$$P(A \cap B) = P(A) \cdot P(B).$$

   For any such A we have

$$\int_B \mathbf{1}_A dP = P(A \cap B) = P(A) \cdot P(B) = \int_B P(A) dP$$

   and the constant function $P(A)$ is measurable for any σ-field. So $E(Z|\mathcal{G}) = E(Z)$ for simple functions which are measurable relative to $\sigma(X)$ and hence for all integrable functions which are measurable relative to $\sigma(X)$ in particular for X.

Let us go back to Problem **2**. If X is now any element of $L_1([0,1])$ we can think of $E(X|\mathcal{G}_n)$ as giving more and more detailed information about X as n increases. This is not a purely theoretical example. Indeed, when you download an image (on a relatively slow computer) you see this happening. First the image is blurred out and then more and more details emerge. The method of image compression and transmission is very close to what is described in Problem **2**. In fact the sequence of σ-fields in Problem **2** are closely related to what are known as Haar wavelets. If we get to study wavelets in the Spring, I will describe how this method is improved upon in current technology.

2.4 Filtered probability spaces and martingales.

So we define a **filtered space** $(\Omega, \mathcal{F}, \mathcal{F}_n, P)$ to consist of a probability triple $(\Omega, , \mathcal{F}, P)$ together with a collection of σ-fields

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \cdots \subset \mathcal{F}.$$

An example is provided by Problem **2**. Another way of obtaining examples is to start with a collection $W_0, W_1, W_2, \ldots$ of random variables, and let

$$\mathcal{F}_n = \sigma(W_0, \ldots, W_n),$$

that is the smallest σ-field relative to which $W_0, \ldots, W_n$ are measurable. In this case, we will frequently write

$$E(X|W_0, \ldots, W_n)$$

instead of $E(X|\mathcal{F}_n)$.

A collection of random variables X_n is called a (discrete time) **random process**. A random process is said to be **adapted** to the filtration $\{\mathcal{F}_n\}$ if X_n is $\mathcal{F}_n$ measurable.

A random process $\{X_n\}$ is said to form a **martingale** relative to the filtration $\{\mathcal{F}_n\}$ if it is adapted, if

$$E(|X_n|) < \infty$$

for all n and

$$E(X_n|\mathcal{F}_{n-1}) = X_{n-1}.$$

If you think of X_n as a gambler's fortune after the nth gamble, then this condition states that his conditional expected fortune after the $(n+1)$st gamble is equal to his fortune after the nth gamble no matter what may have previously occurred. So a martingale is a generalized version of a fair game.

If we have a filtration $\{\mathcal{G}_n\}$ with

$$\mathcal{G}_n \subset \mathcal{F}_n$$

for all n, and if each X_n in a martingale is actually $\mathcal{G}_n$ measurable, then it follows from (1) that it is a martingale relative to the filtration $\mathcal{G}_n$.

So the standard definition is to take $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ and so the condition to be a martingale is

$$E(X_{n+1}|X_0, \ldots, X_n) = X_n. \tag{4}$$

The introduction of the more general filtration is convenient for checking (4). Consider the following example: Let $Y_0, \ldots, Y_n, \ldots$ be random variables and define

$$Z_i = Y_i - E(Y_i|Y_0, \ldots, Y_{i-1})$$

and then

$$X_n := \sum_0^n Z_i.$$

Suppose that $E(|X_n|) < \infty$ for all n. Let $\mathcal{F}_n$ be $\sigma(Y_0, \dots, Y_n)$, the σ field generated by $Y_0, \dots, Y_n$. Then X_n is $\mathcal{F}_n$ measurable and

$$X_{n+1} = X_n + Y_{n+1} - E(Y_{n+1}|\mathcal{F}_n)$$

so

$$E(X_{n+1}|\mathcal{F}_n) = E(X_n|\mathcal{F}_n) = X_n$$

since $E(E(Y|\mathcal{F}_n)|\mathcal{F}_n) = E(Y_{n+1}|\mathcal{F}_n)$ and X_n is $\mathcal{F}_n$ measurable. So the X_n form a martingale.

An extreme special case of this example is where the Y_n are independent and have expectation zero so that $E(Y_i|Y_0, \dots, Y_{i-1}) = E(Y_i) = 0$.

3 Stopping times.

Suppose we have a filtered space $(\Omega, \mathcal{F}, \mathcal{F}_n, P)$. A map

$$\tau : \Omega \to \{0, 1, 2, \dots, \infty\}$$

is called a **stopping time** if it satisfies the following two conditions

- The set $\{\omega|\tau(\omega) = n\}$ belongs to $\mathcal{F}_n$ and
- The probability that $\tau < \infty$ is one.

The first condition has the following intuitive meaning: I will decide to stop playing the game at time n based on my knowledge at time n. I may not use information arriving in the future to decide to stop at time n. It is easy to check that the first item is equivalent to the condition that the set $\{\omega|\tau(\omega) \leq n\}$ belongs to $\mathcal{F}_n$.

The second condition is technical, and some authors drop it.

If we are given a stopping time τ, and a process $\{X_n\}$, we can consider the random variable X_τ. Explicitly, X_τ is the function on Ω given by

$$X_\tau(\omega) := X_{\tau(\omega)}(\omega).$$

Since X_τ is itself a random variable, i.e. a measurable function on Ω, we can consider its expectation

$$E(X_\tau) = \int_\Omega X_\tau dP.$$

This represents the expected fortune of the gambler if he plays according to his strategy of stopping at time τ.

3.1 The conclusion of Doob's Stopping Time Theorem.

We can now state the *conclusion* of the Martingale Stopping Theorem which says that no stopping time strategy (under suitable technical hypotheses called "regularity") can change the expected outcome of a martingale. In symbols:

$$E(X_\tau) = E(X_0). \tag{5}$$

We can see that *some* condition on the stopping time is needed by looking at the example of random walk on the integers: Let $B_n,\ n = 1, 2, \cdots$ be a collection of independent random variables taking on the values 1 and -1 each with probability $\frac{1}{2}$, and let

$$X_0 \equiv 0,\ X_1 := B_1,\ X_2 := X_1 + B_2, \ldots, X_{n+1} := X_n + B_{n+1}.$$

Then

$$E(X_{n+1}|X_1, \ldots, X_n) = E(X_n|X_1, \ldots, X_n) + E(B_{n+1}|X_1, \ldots, X_n) = X_n$$

since B_{n+1} is independent of $X_1, \ldots X_n$ its conditional expectation is the same as its ordinary expectation which is zero, and $E(X_n|X_1, \ldots, X_n) = X_n$. So the X_n, which represent the positions at time n of a particle undergoing a random walk, form a martingale. Equally well, we can think of this as the fortune (positive or negative) at time n of a gambler with unlimited credit who starts at zero and bets according these Bernoulli trials.

Suppose the gambler decides to stop as soon as he is one dollar ahead. So

$$\tau(\omega) = \inf\{n :\ X_n(\omega) = 1\}.$$

It is not hard to prove that $P(\tau < \infty) = 1$. It is clear that the decision whether or not to stop at n depends only on the outcomes of the first n trials. So τ is a stopping time. By its very definition,

$$X_\tau \equiv 1$$

so

$$E(X_\tau) = 1.$$

But $E(X_0) = 0$. This fact, that $E(X_\tau) \neq E(X_0)$, is one way of formulating Bernoulli's famous "St. Petersburg paradox". The trouble with this particular τ is that $E(\tau) = \infty$.

3.2 Application to our problem.

I will illustrate how to apply this to our problem of waiting for a pattern. Suppose that a letter L can appear with probability p. A fair reward for betting one dollar correctly on the appearance of L is $\frac{q}{p}$ since the expected gain on such a bet is

$$p \cdot \frac{q}{p} - q \cdot 1 = 0.$$

Suppose that there are only the letters A,B, and C with probabilities $\frac{1}{2}, \frac{1}{3}$ and $\frac{1}{6}$ respectively. The payoffs for guessing these letters correctly are then $1, 2$ and 5. I want to compute the expected waiting time for the first appearance of ACA. Here is how Ross (page 301) does this problem. Consider the following game from the point of view of the casino: A player arrives on day one and bets one dollar that A will appear. If he wins, he bets his entire fortune, consisting of two dollars, on C on day two. If he loses, he quits. In either event, a second player comes in on day two and bets his entire fortune of one dollar on A. If the first player won on day two, he will have $2+5\cdot 2 = 12$ dollars, which he bets on day three. In the meanwhile, either the second player loses and quits on day two or continues to day three, and a third player comes in to bet on A on day three. If the first player loses on the third day he quits. Also, if he wins on day three he will have $12+12 = 24$ dollars, and he takes his winnings and goes home. The casino agrees to play this game until the first appearance of the pattern ACA, i.e the first winner. This is its chosen stopping time. Let X_n represent the winnings of the casino on day n. It is a martingale with $E(X_n) = 0$. Therefore $E(X_\tau) = 0$ by the Martingale Stopping Theorem. Now the value of X_τ can be computed as follows: All gamblers betting on days $1, \dots \tau - 3$ will have lost one dollar. The gambler betting on day $\tau - 2$ will have won 23 dollars. The gambler betting on A on day $\tau - 1$ will have lost one dollar, and the gambler betting on A on day τ will have won one dollar. So

$$X_\tau = \tau - 3 - 23 + 1 - 1 = \tau - 26.$$

Since $E(X_\tau) = 0$ we conclude that

$$E(\tau) = 26.$$

On average it will take 26 days for the pattern ACA to occur for the first time.

Let us compute how long it takes for AABBAA to appear when the only outcomes are A with probability p and B with probability $q = 1 - p$. For a correct bet on A of a dollar, the gambler wins $\frac{q}{p}$ dollars and has a fortune of $1 + \frac{q}{p} = \frac{1}{p}$ dollars. So the winner who started at day $\tau - 5$ has $p^{-4}q^{-2}$ dollars. The casino will have gained $\tau - 6$ dollars from the players betting before time $\tau - 5$, and will have lost $p^{-4}q^{-2} - 1$ dollars to the winner who started on day $\tau - 5$, will have gained one dollar each from the players starting on days $\tau - 4, \tau - 3$ and $\tau - 2$, will have lost $p^{-2} - 1$ dollars to the player starting on day $\tau - 1$ and $p^{-1} - 1$ dollars to the player who entered on day τ. So

$$X_\tau = \tau - 6 + 6 - p^{-4}q^{-2} - p^{-2} - p^{-1}.$$

Therefore

$$E(\tau) = p^{-4}q^{-2} + p^{-2} + p^{-1}.$$

If $p = q = \frac{1}{2}$ this is $64 + 4 + 2 = 70$ as claimed.

If we are waiting for the pattern AAAAAA then the same computation yields

$$p^{-6} + p^{-5} + p^{-4} + p^{-3} + p^{-2} + p^{-1} = p^{-1} \cdot \frac{1 - p^{-6}}{1 - p^{-1}}.$$

If $p = \frac{1}{2}$ this is $2 \cdot 63 = 126$.

3. Monkeys type one of the 26 capital letters every minute, each with the same probability 1/26. On average, how long with it take a monkey to type until "ABRACADABRA" appears?

We will now work our way toward a formulation and proof of Doob's theorem.

3.3 The stopped process.

Let us call τ a **random time** if we drop the condition $P(\tau < \infty) = 1$ in the definition of stopping time, but retain the condition that the event

$$\{\tau = n\}$$

is determined by the random variables $X_1, \ldots, X_n$. Suppose that the X_n form a martingale, and define the **stopped process** as

$$X_{\tau\wedge n}.$$

In more detail: the value of $Y_n = X_{\tau\wedge n}$ at a point ω is

$$Y_n(\omega) = \begin{cases} X_n(\omega) & \text{if } \quad n \leq \tau(\omega) \\ X_{\tau(\omega)}(\omega) & \text{if } \quad n > \tau(\omega) \end{cases}.$$

Proposition 1 *The stopped process is also a martingale and* $E(X_{\tau\wedge n}) = E(X_0)$.

Proof. [Ross page 298.] Let the random variables I_n be defined by

$$I_n(\omega) = \begin{cases} 1 & \text{if } \quad \tau(\omega) \geq n \\ 0 & \text{if } \quad \tau(\omega) < n \end{cases}.$$

So $I_n = 1$ if we haven't yet stopped after observing $X_1, \ldots, X_{n-1}$ and is zero otherwise. We claim that if we set $Y_n = X_{\tau\wedge n}$ then

$$Y_n = Y_{n-1} + I_n(X_n - X_{n-1}).$$

Indeed consider separately the two possibilities $\tau \geq n$ and $\tau < n$: If at ω, $\tau \geq n$ then $Y_n = X_n,\ Y_{n-1} = X_{n-1}$, and $I_n = 1$ so the equation is true. If $\tau(\omega) < n$ then at ω we have $Y_n = Y_{n-1} = X_\tau$ and $I_n = 0$. The equation is true in both cases.

Taking conditional expectations gives

$$E(Y_n|X_1, \ldots, X_{n-1}) = E(Y_{n-1}|X_1, \ldots, X_{n-1}) + E(I_n(X_n - X_{n-1})|X_1, \ldots, X_{n-1}).$$

Since Y_{n-1} depends only on $X_1, \ldots, X_{n-1}$ we have $E(Y_{n-1}|X_1, \ldots, X_{n-1}) = Y_{n-1}$. As to the second term in the above displayed equation, since I_n depends only on $X_1, \cdots, X_{n-1}$ we can pull it out of the conditional expectation sign by (3). fact. So

$$E(I_n(X_n - X_{n-1})|X_1, \ldots, X_{n-1}) = I_n E(X_n - X_{n-1})||X_1, \ldots, X_{n-1}) = 0$$

since the X_n form a martingale. Thus

$$E(Y_n|X_1, \ldots, X_{n-1}) = Y_{n-1}.$$

But the σ-field determined by the $X_1, \ldots, X_{n-1}$ contains the σ-field generated by the $Y_1, \ldots, Y_{n-1}$ so if $E(Y_n|X_1, \ldots, X_{n-1}) = Y_{n-1}$ then

$$E(Y_n|Y_1, \ldots, Y_{n-1}) = Y_{n-1}$$

proving that the Y_n form a martingale. Since $Y_0 = X_0$ we have

$$E(X_{\tau \wedge n}) = E(X_0) \tag{6}$$

for all n, which was the last assertion of the proposition. QED

3.4 Doob's Stopping Time Theorem.

Suppose we let $n \to \infty$ in $X_{n \wedge \tau}$. If ω is a point such that $\tau(\omega) < \infty$ then

$$\lim_{n \to \infty} X_{n \wedge \tau}(\omega) = X_\tau(\omega). \tag{7}$$

In fact if $n > \tau(\omega)$ then $X_{\tau \wedge n}(\omega) = X_\tau(\omega)$. So the condition that τ be a stopping time asserts that the above limit holds almost everywhere. So the question of whether (6) implies (5) is reduced to a familiar type of problem in measure theory: can we pass to the limit under the integral sign?

Doob's theorem asserts that any one of the following three conditions is enough to conclude that (5) holds:

1. τ is bounded (i.e. there is an N such that $\tau(\omega) \leq N$ for all ω,
2. There is a constant K such that $|Y_n(\omega)| < K$ for all ω,
3. $E(\tau) < \infty$ and there is a constant M such that

$$E(|X_{n+1} - X_n|X_0, \ldots X_n) < M.$$

If condition 1) holds, then we have eventual equality in (7) so there is nothing to prove.

4. Cite the relevant theorem(s) in integration theory to verify that conditions 2) and 3) are enough to guarantee passing to the limit to obtain (5).

5. Show that condition 3) is satisfied in our waiting for a pattern example. [Hint: Use problem 1 to get a crude estimate on $E(\tau)$.]

3.5 Wald's equation.

Let Z_i be independent identically distributed random variables with $E[|Z_i|] < \infty$ and (common) expectation $E[Z_i] = \mu$. So

$$X_n = \sum_1^n (Z_i - \mu)$$

form a martingale (starting at 1 for convenience). Let τ be a stopping time with $E(\tau) < \infty$. Wald's equation says that

$$E\left[\sum_{i=1}^{\tau} Z_i\right] = \mu E(\tau). \tag{8}$$

6. Prove Wald's equation.