Problem Set 6

Math 212a

November 29, 2001, Due Dec. 11

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. I will try to develop some of these techniques in this problem set. There are nine problems scattered through the text.

Contents

1 Expected time until a pattern. **1**

1.1 The geometric distribution. 2
1.2 The pattern matters. 2
1.3 Stopping time. 2
1.4 Martingales. 3
1.5 Application to our problem. 4
1.6 The stopped process. 6
1.7 Doob's Stopping Time Theorem. 7

2 Conditional Expectation. **8**

2.1 Convergence in L^p. 8
2.2 Conditional expectation on L^2. 10
2.3 Conditional expectation on L^1. 11
2.4 Conditional expectation on $L^p, 1 \leq p < \infty$. 13

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 $k = 1, 2, \ldots$? 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.

1.3 Stopping time.

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.

Suppose that $X_n,\ n \geq 1$ is a sequence of random variables with

$$E(|X_n|) < \infty \tag{1}$$

for all n. Recall that $E(X)$ is just another way of writing $\int_\Omega X dP$, so the above condition says that $X_n \in L^1$.

For example, X_n might represent the fortune of a gambler after the n-th gamble in a casino. The gambler might have a strategy as to when to quit playing. (This strategy might be more sophisticated than to simply quit when he runs out of money.) His strategy must not, however, involve a foreknowledge of the future. So let τ denote this stopping time. Since τ might depend on the outcome of some of the gambles, τ itself is a random variable, taking on the possible values $1, 2, 3, \ldots, \infty$. We will make two conditions on τ. The first is that the probability that $\tau < \infty$ is one. The second condition - that no foreknowledge of the future is allowed - says that the event

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

is determined by the random variables $X_1, \ldots, X_n$. Let me phrase this condition in terms of measure theory. A (real valued) random variable is a real valued function on our underlying probability space $(\Omega, \mathcal{A}, P)$. Here Ω is a set, $\mathcal{A}$ is a σ-field, and P is a measure on $\mathcal{A}$ with $P(\Omega) = 1$. For example, Ω might be the unit interval with $\mathcal{A}$ the Lebesgue measurable sets and P Lebesgue measure. A function X on Ω is called measurable, if $X^{-1}(B) \in \mathcal{A}$ for any Borel subset of $\mathbf{R}$, and any finite or countable collection of measurable functions determines a sub-σ-field of $\mathcal{A}$ generated by all these inverse images. Thus the collection $X_1, \ldots, X_n$ of functions determines a sub-σ-field $\mathcal{B}_n$ with

$$\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{B}_n \subset \cdots .$$

Our second condition on τ says that

$$\tau^{-1}(\{n\}) \in \mathcal{B}_n.$$

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, and 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.

1.4 Martingales.

The casino is "fair" if the expected value of the gambler's fortune after the $(n+1)$st gamble is equal to his fortune after the nth gamble. We can try to phrase this mathematically by saying that

$$E(X_{n+1}|X_1, \ldots, X_n) = X_n. \tag{2}$$

The symbol on the left is the conditional expectation of X_{n+1} given (the outcome of) $X_1, \ldots, X_n$. We gave a definition of this concept in the Hilbert space setting in Problem Set **2**, and part of our task will be to generalize this definition. But let us assume that we can use the Hilbert space definition as applied to $L_2[0,1]$) for example. In other words, replace condition (1) by the stronger condition:

$$X_n \in L^2 \quad \forall \;\; n$$

and use the results of Problem set 2, or look ahead to the next section where the concept of conditional expectation is developed from scratch, first for L^2 and then for all L^p, $1 \leq p < \infty$.

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 the game. In symbols

$$E(X_\tau) = E(X_0). \tag{3}$$

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 $X_1, \ldots X_n$ so 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 $E(\tau) = \infty$.

1.5 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 $1, \ldots \tau - 3$ 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 $2 \cdot 63 = 126$.

2. 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?

1.6 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

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

Proposition 1 *The stopped process is also a martingale and* $E(X_{\tau \wedge n}) = E(X_1)$.

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

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

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 $\tau \ge n$ and $\tau < n$: If at ω, $\tau geqn$ 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 $I_n = 0$. So

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

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 - see Problem **7** below for a review of this 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. (If you have forgotten this property of conditional expectation, again look at the next section.) Since $Y_1 = X_1$ we have

$$E(X_{\tau \wedge n}) = E(X_1) \tag{4}$$

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

1.7 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

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

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 (4) implies (3) is reduced to a familiar type of problem in measure theory: can we pass to the limit under the integral sign? For example, suppose that τ is bounded, i.e. that there exists a constant C such that

$$\tau(\omega) < C \quad \forall\, \omega. \tag{5}$$

Then the limit process is eventual equality, so there is no problem in passing to the limit.

Suppose that the $X_{\tau \wedge n}$ are uniformly bounded in absolute value, i.e. there is a constant K such that

$$|X_n(\omega)| \le K \tag{6}$$

for all n and ω. Since $P(\Omega) = 1$, in particular is finite, we can apply the dominated convergence theorem to conclude (3).

Suppose that that

$$E(\tau) < \infty \;\text{ and } \exists\, K > 0 \text{ such that } |X_n(\omega) - X_{n-1}(\omega)| < K \tag{7}$$

for all n and ω. Then

$$|X_{\tau \wedge n} - X_1| \le \sum_{k=2}^{\tau \wedge n} |X_k - X_{k-1}| \le K\tau$$

and $E(K\tau) = KE(\tau) < \infty$. So by the dominated convergence theorem we can pass to the limit and obtain (3).

The content of Doob's stopping time theorem is what we have just proved: that any one of the above three conditions is enough to justify (3).

In the case of our problem of the waiting time until a pattern appears, the number of players in play on a given day is at most ℓ, the length of the pattern, so the second of the two conditions in (7) - that the total possible winnings or losses of the casino on a given day be bounded is satisfied. As to the first condition, let w represent the probability of occurrence of the pattern in ℓ successive trials. Then $E(\tau)$ is finite: in fact, consider the geometric distribution with parameter w. By problem **1**, $E(\tau) \leq \ell \cdot (1/w) < \infty$ verifying the first condition.

2 Conditional Expectation.

In this section I will deal with a fixed probability space $(\Omega, \mathcal{A}, P)$ but with varying sub-σ fields $\mathcal{B}$. Unless otherwise specified, I will assume that every such $\mathcal{B}$ contains all the sets of P-measure zero in $\mathcal{A}$. I will let

$$\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{A}, P), \quad 1 \leq p \leq \infty$$

denote the space of spaces of measurable functions f such that $|f|^p$ is integrable (for $p < \infty$), or which are essentially bounded (for $p = \infty$) and only at the very beginning will I distinguish between these spaces and L^p where we identify two functions which are equal except on a set of measure zero.

2.1 Convergence in L^p.

Here is a theorem which I may have forgotten to do in class: Its purpose is to describe what it means for a sequence in $\mathcal{L}^p$, $1 \leq p < \infty$ to be convergent.

Theorem 1 *Suppose $f_n \to f$ in $\mathcal{L}^p$, $1 \leq p < \infty$. Then there is a subsequence which converges to f almost everywhere, and such that for any $\epsilon > 0$ there is a set Z with $P(Z) < \epsilon$ such that the convergence is uniform outside of Z.*

Proof. Replacing f_n by $f_n - f$ we may assume that $f_n \to 0$. Passing to a subsequence we may assume that

$$||f_n||_p < \frac{1}{2^{2n}}.$$

Changing the f_n on a set of measure zero we may assume that the f_n are measurable. Let

$$Y_n := \{x| \ \ |f_n(x)| \geq \frac{1}{2^n}\}$$

so

$$\frac{1}{2^n} P(Y_n)^{\frac{1}{p}} \leq \left(\int_{Y_n} |f_n(x)|^p dP \right)^{\frac{1}{p}} \leq ||f_n||_p \leq \frac{1}{2^{2n}}$$

so

$$P(Y_n) \leq \frac{1}{2^{np}} \leq \frac{1}{2^n}.$$

Set

$$Z_n := Y_n \cup Y_{n+1} \cup \cdots$$

so

$$P(Z_n) \leq \frac{1}{2^{n-1}}$$

and if $x \notin Z_n$ then

$$k \geq n \quad \Rightarrow \quad |f_k(x)| \leq \frac{1}{2^k}$$

so f_k converges to zero uniformly outside Z_n and pointwise on the complement of

$$Z := \bigcap Z_n$$

and Z has measure zero. QED

Corollary 1 *$f \in \mathcal{L}^p$ has $\|f\|_p = 0$ if and only if $f = 0$ almost everywhere.*

Proof. If $\|f\|_p = 0$, then the sequence $\{0, 0, \dots\}$ converges in L_1 to f and hence almost everywhere by the theorem. Hence $f = 0$ a.e. Conversely, if $f = 0$ a.e., then $\{0, 0, \dots\}$ is a sequence which converges in L_p norm to f. QED

In particular, if $\|f - g\| = 0$, then $f = g$ a.e. and conversely. In other words, the projection from $\mathcal{L}_p$ to L_p consists in identifying two functions which agree almost everywhere.

Corollary 2 *If $f_n \to f$ a.e. and also is Cauchy in the L^p norm, then $f \in \mathcal{L}^p$ and is the L^p limit of the f_n.*

Proof. By the completeness of $\mathcal{L}^p$ we know that f_n converges to some $g \in \mathcal{L}^p$ and hence some subsequence converges almost everywhere to g. We also know that this subsequence (in fact we are assuming that the original sequence) converges a.e. to f. Hence $f = g$ a.e. and hence $f \in \mathcal{L}_p$. QED

In the above arguments we did not make use of the fact that P is a probability measure. All statements are true for an arbitrary measure space. But for the rest of this section, it will be simpler to assume that the constant **1** is integrable. So we will only deal with a probability measure. But we will drop the distinction between $(\mathcal{L})^p$ and L^p.

We let $L(\mathcal{B})$ denote the space of (equivalence classes of) finite $\mathcal{B}$ measurable sets, and

$$L^p(\mathcal{B}) = L(\mathcal{B}) \cap L^p(\Omega, \mathcal{A}, P).$$

Clearly

$$L(\mathcal{B}_1) \subset L(\mathcal{B}_2) \Leftrightarrow \mathcal{B}_1 \subset \mathcal{B}_2$$

with similar equivalences with L replaced by L^p and $\subset$ replaced by $=$.

We will (temporarily) be considering real valued functions, and want to characterize which subspaces of $L = L(\mathcal{A})$ or of L^p are of the form $L(\mathcal{B})$ or of the form $L^p(\mathcal{B})$. Any such subspace is closed under under monotone limits. In fact the space $L^p(\mathcal{B})$ is closed as a subspace of L^p by the preceding theorem: If $f_n \in L^p(\mathcal{B})$ converges in L^p we can pass to a subsequence which converges almost everywhere, and hence its limit f is in $\mathcal{B}$. Any monotone sequence in L^p whose pointwise limit also belongs to L^p actually converges in L^p to its pointwise limit since $\|f_n - f\|_p \to 0$ by the dominated convergence theorem.

Also any subspace of L^p of the form $L^p(\mathcal{B})$ contains the constant function $\mathbf{1}$ and is also closed under $\wedge$ and $\vee$.

3. Prove the converse: If M is a subspace of L or of L^p which is closed under monotone limits, contains $\mathbf{1}$ and is closed under $\wedge$ and $\vee$ then M is of the form $L(\mathcal{B})$ or $L^p(\mathcal{B})$. To say that M is closed under monotone limits means that if $f_n \in M$ is a monotone sequence and its pointwise limit f belongs to L or to L^p then $f \in M$. [Hint: define $\mathcal{B} := \{B : \ B \in \mathcal{A}, \ \mathbf{1}_B \in M\}$ and show that it is a σ-field with the desired properties.]

For $1 \leq p < \infty$ being closed as a subspace implies being closed under monotone limits, as we saw above by the monotone convergence theorem. So any subspace of L^p which is closed as a subspace, contains the constant $\mathbf{1}$ and is closed under $\wedge$ and $\vee$ must be of the form $L^p(\mathcal{B})$ for some $\mathcal{B}$. But for $p = \infty$ we have our old friend the essentially continuous functions on $[0,1]$ which is closed as a subspace of L^∞ but is not of the form $L^\infty(\mathcal{B})$.

An important corollary of the preceding remark is:

4. Let $1 \leq p < \infty$. Let

$$U : L^p \to L^p$$

be a linear map which satisfies:

- $\|U\| \leq 1$
- $U(\mathbf{1}) = \mathbf{1}$
- $f \geq 0 \Rightarrow Uf \geq 0$.

Let

$$M := \{f \in L^p : Uf = f\}$$

be the subspace of invariant functions under U. Show that M is of the form $M = L^p(\mathcal{B})$. [Hint: Show that $U(f^+) \geq (Uf)^+$ for all $f \in L^p$ and conclude from this that $f \in M \Rightarrow f^+ \in M$. From this conclude that M is closed under $\wedge$ and $\vee$.]

2.2 Conditional expectation on L^2.

We now revisit the notion of conditional expectation in L^2. The orthogonal projection of L^2 onto a subspace of the form $L^2(\mathcal{B})$ is called **conditional expectation with respect to** $\mathcal{B}$ and is denoted by $E^{\mathcal{B}}$. It is characterized by

1. $E^{\mathcal{B}} f \in L^2(\mathcal{B})$ and
2. $\int E^{\mathcal{B}}(f) g dP = \int f g dP \ \ \forall \ g \in L^2(\mathcal{B})$

for all $f \in L^2$. (Recall that we are dealing with real spaces for the moment.)

In the second condition it is enough to verify the equation for a set of g's whose linear combinations are dense in $L^2(\mathcal{B})$.

5. Show that $E^{\mathcal{B}}(f) \geq 0$ if $f \geq 0$. [Hint: Consider $g = \mathbf{1}_{E^{\mathcal{B}}(f)<0}$ in condition 2.]

6. Show that if U is an orthogonal projection on L^2 then a necessary and sufficient condition that it be of the form $U = E^{\mathcal{B}}$ is that

- $f \geq 0 \Rightarrow Uf \geq 0$ and
- $U(\mathbf{1}) = \mathbf{1}$.

7. Show that if $f \in L^2$ and $h \in L^\infty(\mathcal{B})$ then

$$E^{\mathcal{B}}(hf) = hE^{\mathcal{B}}(f).$$

[Hint: Show that $hE^{\mathcal{B}}(f) - hf$ is orthogonal to $L^2(\mathcal{B})$ and use this fact to deduce the result.]

2.3 Conditional expectation on L^1.

We now want to extend the notion of conditional expectation to L^1 and then to all L^p, $1 \leq p \leq \infty$.

Let $\overline{L}_+$ denote the space of equivalence classes of measurable maps of $(\Omega, \mathcal{A}, P)$ to $[0, \infty]$. So we are requiring a function in $\overline{L}_+$ to be non-negative, but it might take on the value $+\infty$. Our first step will be to generalize the notion of conditional expectation to this class of non-negative functions, with the possibility that this conditional expectation might be infinite at places. So let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. We let $\overline{L}_+(\mathcal{B})$ denote the subspace of $\mathcal{B}$ measurable equivalence classes in $\overline{L}_+$.

Proposition 2 *For any $f \in \overline{L}_+$ there exists a unique element $E^{\mathcal{B}}(f) \in \overline{L}_+(\mathcal{B})$ such that*

$$\int_\omega E^{\mathcal{B}}(f) g dP = \int_\Omega f g dP \tag{8}$$

for all $g \in \overline{L}_+(\mathcal{B})$ (with the understanding that both sides might be infinite). In fact, $E^{\mathcal{B}}(f)$ is already determined by the condition that the above equation hold for $g = \mathbf{1}_B$ for all $B \in \mathcal{B}$. Also, for all $h \in \overline{L}_+(\mathcal{B})$

$$E^{\mathcal{B}}(hf) = hE^{\mathcal{B}}(f). \tag{9}$$

Proof. If $f \in L^2_+$, i.e. if f is a non-negative element of L^2, then the $E^{\mathcal{B}}(f)$ constructed by orthogonal projection as above satisfies condition (8) in the proposition for all $g \in L^2(\mathcal{B})$, in particular for all $g \in L^2_+(\mathcal{B})$. We can approximate any $g \in \overline{L}_+(\mathcal{B})$ from below by the monotone increasing sequence $g \wedge n\mathbf{1}$ each of whose elements belongs to $L^2(\mathcal{B})_+$ so (8) holds. If $f \in \overline{L}_+$ but does not belong to L^2 we can approximate it from below by the monotone increasing sequence $f \wedge n\mathbf{1}$ whose elements do belong to L^2. By Problem **5** the sequence $E^{\mathcal{B}}(f \wedge n\mathbf{1})$ is monotone so we can define its pointwise limit to be $E^{\mathcal{B}}(f)$, and then (8) holds by the monotone convergence theorem. This establishes the existence part of the proposition.

For the uniqueness, suppose that h_1 and h_2 are two elements of $L^2(\mathcal{B})_+$ such that

$$\int_B h_1 dP = \int_B f dP = \int_B h_2 dP \quad \forall\ B \in \mathcal{B}.$$

In particular, for any pair of real numbers a, b with $0 \le a < b$, let

$$B := \{\omega : h_1(\omega) \le a < b \le h_2(\omega)\}.$$

Then $B \in \mathcal{B}$. Let $p = P(B)$. Then

$$\int_B h_1 dP \le ap$$

while

$$\int_B h_2 \ge bp.$$

These two can only be equal if $p = 0$. Letting a and b range over the rationals, this shows that $h_1 \le h_2$ a.e. and interchanging the roles of h_1 and h_2 we conclude that $h_1 = h_2$ a.e. This proves uniqueness.

Finally, if $h \in \overline{L}_+(\mathcal{B})$ then $hE^{\mathcal{B}}(f)$ satisfies

$$\int hE^{\mathcal{B}}(f) g dP = \int f h g dP$$

for all $g \in \overline{L}_+(\mathcal{B})$ and then the uniqueness implies (9). QED

Recall that we defined L^1 to consist of all measurable f such that f^+ and f^- have finite integrals.

Theorem 2 *For all $f \in L^1$ there exists a unique $E^{\mathcal{B}}(f) \in L^1(\mathcal{B})$ such that*

$$\int_\Omega f g dP = \int_\Omega E^{\mathcal{B}}(f) g dP \quad \forall\ g \in L^\infty(\mathcal{B}). \tag{10}$$

Furthermore

- $f \mapsto E^{\mathcal{B}}(f)$ *is linear,*
- $f \mapsto E^{\mathcal{B}}(f)$ *is an idempotent,*
- $f \ge 0 \Rightarrow E^{\mathcal{B}}(f) \ge 0,$
- $\|E^{\mathcal{B}}(f)\|_1 \le \|f\|_1,$

- $E^{\mathcal{B}}(\mathbf{1}) = \mathbf{1}$

and if h is a $\mathcal{B}$ measurable function then

$$E^{\mathcal{B}}(hf) = hE^{\mathcal{B}}(f)$$

if hf is integrable.

Proof. If $f \in L^1$ is non-negative, we can apply the proposition to get $f \in L_+(\mathcal{B})$ satisfying (8) for all $g \in \overline{L}^+(\mathcal{B})$. In particular if we take $g = \mathbf{1}$ we see that $E^{\mathcal{B}}(f)$ is integrable, and then we can extend (8) to all $g \in L^\infty(\mathcal{B})$. For a general $f \in L^1$ we define

$$E^{\mathcal{B}}(f) := E^{\mathcal{B}}(f^+) - E^{\mathcal{B}}(f^-)$$

and this still satisfies (10). The same uniqueness argument we used for the proposition shows that $E^{\mathcal{B}}(f)$ is already determined by the condition

$$\int_B E^{\mathcal{B}}(f)dP = \int_B f dP \quad \forall \;\; B \in \mathcal{B}$$

and this uniqueness implies that the map $f \mapsto E^{\mathcal{B}}(f)$ is linear. By construction $f \geq 0$ implies that $E^{\mathcal{B}}(f) \geq 0$ and $E^{\mathcal{B}}(\mathbf{1}) = \mathbf{1}$. The definitions imply that

$$|E^{\mathcal{B}}(f)| \leq E^{\mathcal{B}}(f^+) + E^{\mathcal{B}}(f^-) = E^{\mathcal{B}}(|f|)$$

and so

$$\|E^{\mathcal{B}}(f)\|_1 = \int_\Omega |E^{\mathcal{B}}(f)dP \leq \int_\Omega E^{\mathcal{B}}(|f|)dP = \int_\Omega |f|dP = \|f\|_1.$$

Since $E^{\mathcal{B}}(f) = f$ if $f \in L^1(\mathcal{B})$ it follows that the operator $E^{\mathcal{B}}$ idempotent. Finally, we know that the last equation in the theorem is true when f and h are non-negative, as this is the content of the proposition. Applying this to the decompositions $f = f^+ - f^-$ and $h = h^+ - h^-$ and using the uniqueness establishes the last assertion of the theorem. QED

2.4 Conditional expectation on $L^p, 1 \leq p < \infty$.

Since the total measure $P(\Omega)$ is finite, $L^p \subset L^1$ so we know that $E^{\mathcal{B}}(f) \in L^1$ is defined and unique by the preceding theorem. We wish to show that it belongs to L^p.

8. Show that if $f \in L^p$ and $g \in L^q$ then

$$|E^{\mathcal{B}}(fg)| \leq E(|f|^p)^{\frac{1}{p}} \cdot E^{\mathcal{B}}(|g|^q)^{\frac{1}{q}}.$$

[Hint: Follow the proof of Hölder's inequality that we gave in class. Here you have to be a bit more careful since, for example the the function $E^{\mathcal{B}}(|f|^p)$ might vanish on some non-trivial set B.]

9. Show that $E^{\mathcal{B}}$ maps L^p to L^p, and that

- $f \mapsto E^{\mathcal{B}}(f)$ is linear,
- $f \mapsto E^{\mathcal{B}}(f)$ is an idempotent,
- $f \geq 0 \Rightarrow E^{\mathcal{B}}(f) \geq 0$,
- $\|E^{\mathcal{B}}(f)\|_p \leq \|f\|_p$,
- $E^{\mathcal{B}}(\mathbf{1}) = \mathbf{1}$

and if $h \in L^q(\mathcal{B})$ then

$$E^{\mathcal{B}}(hf) = hE^{\mathcal{B}}(f).$$