Problem set 2

Math 212a

Sept. 28, 2000, Due October 10

Contents

1 The L_2 Martingale Convergence Theorem.

Let $\mathbf{H}$ be a Hilbert space and let $\{x_n\}$ be a sequence of elements of $\mathbf{H}$ which satisfy the following two conditions:

- The $\|x_n\|$ are bounded, i.e. there is a constant M such that

$$\|x_n\| \leq M \quad \text{for all } n$$

  and

- For all pairs m and n with $n > m$

$$(x_n - x_m, x_m) = 0. \tag{1}$$

1. Show that the x_n converge to an element x and that $\|x\| \leq M$. In fact, show that $\|x\| = L$ where L is the greatest lower bound of the M in the first condition.

Once you have figured out this problem, you will agree that the proof is quite transparent. You may wonder about the fancy title. In fact, this result is (part of, or a version of) a key result in probability theory which unifies many so called "limit theorems". It is also a very good entry into getting an intuitive understanding of these theorems. The appropriate language for stating these theorems is measure theory, and therefore any advanced book on probability theory either assumes that the reader has had a course on measure theory, or spends the early chapters on an expostion of measure theory. The martingale convergence theorem usually does not make its appearance until the last third of the book. I plan to get into measure theory in the next few weeks, but in this problem set I want to develop a good bit of the language and some of the results of probability theory using Hilbert space tools.

Here is an example of an "abstract" limit theorem which is in fact a consequence of the martingale convergence theorem, but is easier to prove directly:

2. Let $y_k \in \mathbf{H}$ be such that

$$(y_i, y_j) = 0 \quad i \neq j$$

and such that there is a constant K such that

$$\|y_j\|^2 \leq K$$

for all j. Show that for any real number $s > \frac{1}{2}$ we have

$$\lim_{n\to\infty} \frac{1}{n^s}(y_1 + \cdots + y_n) \to 0.$$

A matter of great importance is to understand what happens (under additional hypotheses) at the critical value $s = \frac{1}{2}$. To understand why Problem **2** is in fact a consequence of Problem **1**, consider the case $s = 1$ and take

$$x_n := \sum_1^n \frac{1}{k} y_k.$$

Then the $\|x_n\|$ are bounded because $\sum 1/k^2$ converges, and condition (1) clearly holds. We have

$$y_k = k(x_k - x_{k-1})$$

so

$$y_1 + \cdots + y_n = \sum_1^n k(x_k - x_{k-1}) = nx_n - \sum_1^n x_{k-1}.$$

Dividing by n gives

$$\frac{1}{n}(y_1 + \cdots + y_n) = x_n - \frac{1}{n}\sum_1^n x_{k-1}.$$

Since we know that $x_n \to x$ we also know that the "Cesaro average" $(1/n)\sum_1^n x_{k-1}$ approaches x so the right hand side tends to $x - x = 0$. The same "summation by parts" argument works for any $s > \frac{1}{2}$ to derive Problem **2** from Problem **1** but is a little harder to write down. The direct proof is easier.

2 Probabilistic language in Hilbert space terms.

In what follows we will take $\mathbf{H}$ to be $L_2([0,1])$, the completion of the space of continuous functions on the unit interval relative to the metric determined by the scalar product

$$(f,g) = \int_0^1 f(t)\overline{g(t)}dt.$$

(There is, of course, no difference between $L_2(\mathbf{T})$ and $L_2([0,1])$, but since the convention is to let probabilities range between 0 and 1, we will take $\mathbf{H} = L_2([0,1])$.)

The space $L_2([0,1])$ has certain additional structures, beyond merely being a Hilbert space. Once we have extracted the necessary additional structures we could assemble them into a collection of axioms for what we might call a "probability Hilbert space".

2.1 Expectation.

The space $\mathbf{H}$ has a preferred element $\mathbf{1}$ which we might think of as the function which is identically one. More precisely, it is the linear function on $\mathcal{C}([0,1])$ which assigns to any continuous function its integral over the unit interval:

$$f \mapsto \int_0^1 f(t)dt = (f, \mathbf{1}).$$

We have

$$(\mathbf{1}, \mathbf{1}) = 1.$$

We make the definition

$$E(X) := (X, \mathbf{1}) \quad \text{for any } X \in \mathbf{H}$$

and call $E(X)$ the **expectation** of X. We will call an arbitrary element of $\mathbf{H}$ a **random variable**. (Strictly speaking we should call these "square integrable random variables", but we won't be considering any other kind in this problem set.)

Suppose that $A \subset [0,1]$ is an interval, or a finite union of intervals. The linear function which assigns to any continuous function its integral over A,

$$f \mapsto \int_A f(t)dt$$

is continuous in the $\|\cdot\|_2$ norm:

$$\left|\int_A f(t)dt\right| \leq \int_A |f(t)|dt \leq \int_{[0,1]} |f(t)dt = (|f|, \mathbf{1}) \leq \|f\|_2\|\mathbf{1}\|_2.$$

so it is given by scalar product with an some element of $\mathbf{H}$ which we shall denote by $\mathbf{1}_A$. The meaning of this notation is that we can "represent" this element by the indicator function of the set A, i.e. by the function which is one on A and zero on the complement of A. Indeed, we can find a sequence of continuous functions which converge to this indicator function in the L_2 norm. There is little danger in thinking of $\mathbf{1}_A$ as being an actual function, or as representing the set A, provided that we understand that modifying the set A by inserting or removing a finite number of points (or, as we shall see later, by a set of measure zero) gives the same element $\mathbf{1}_A \in \mathbf{H}$. Clearly

$$E(\mathbf{1}_A) = (\mathbf{1}_A, \mathbf{1}) = (\mathbf{1}_A, \mathbf{1}_A) = \|\mathbf{1}_A\|^2$$

is just the sum of the lengths of the disjoint intervals which comprise A. (It requires a little combinatorial lemma at this point to prove that this value does not depend on how we break A up into a union of disjoint intervals.) We will denote this value by $P(A)$ and call it the **probability** of the **event** A. So

$$P(A) := E(\mathbf{1}_A) = \|\mathbf{1}_A\|^2. \tag{2}$$

2.2 Conditional probability is a Fourier coefficient.

Suppose that $\mathbf{1}_A \neq 0$, and let π_A denote orthogonal projection onto the one dimensional space spanned by $\mathbf{1}_A$. Let B be some other finite union of intervals, so that $A \cap B$ is again a finite union of intervals.

3. Show that

$$\pi_A(\mathbf{1}_B) = \frac{P(A \cap B)}{P(A)}\mathbf{1}_A.$$

In elementary probability theory the expression $\frac{P(A\cap B)}{P(A)}$ is known as the "conditional probability of B given A" and is denoted by $P(B|A)$,

$$P(B|A) := \frac{P(A \cap B)}{P(A)}.$$

So we can write

$$\pi_A(\mathbf{1}_B) = P(B|A)\mathbf{1}_A. \tag{3}$$

In what follows, we will enlarge our set of "events" and take (2) and (3) as definitions. For this we need to make use of some additional structure of $L_2([0,1])$.

2.3 Reality.

If $\{g_n\}$ is a Cauchy sequence (in the $\|\cdot\|_2$ norm) of elements of $\mathcal{C}([0,1])$ then so is the sequence $\{\overline{g_n}\}$, where, of course,

$$(\overline{f})(t) := \overline{f(t)}.$$

This implies that the complex conjugation operation extends to $\mathbf{H}$ as an anti-linear map of $\mathbf{H}$ onto itself, and $\|\overline{g}\|_2 = \|g\|_2$. Thus we can define $\mathbf{H}_\mathbf{R}$ to consist of those $g \in \mathbf{H}$ which satisfy $\overline{g} = g$. It is a vector space over the real numbers, and it is easy to check that every such element is the limit of a Cauchy sequence of real continuous functions. The scalar product of any two elements of $\mathbf{H}_\mathbf{R}$ is real.

Alternatively, we could have developed the whole theory of Hilbert spaces over the real numbers, and then obtined $\mathbf{H}_\mathbf{R}$ as the completion of the real valued continuous functions on $[0,1]$ under the $\|\cdot\|_2$ norm.

2.4 Covariance and Variance.

Let X and Y be real valued random variables. Define the **covariance** of X and Y by

$$\mathrm{cov}(X,Y) := (X - E(X)\mathbf{1}, Y - E(Y))\mathbf{1}) = (X,Y) - E(X)E(Y).$$

Define the **variance** of X by

$$\mathrm{var}(X) := \mathrm{cov}(X,X) = (X,X) - E(X)^2.$$

We say that X and Y are **uncorrelated** if $\mathrm{cov}(X,Y) = 0$. Notice that if $X = \mathbf{1}_A$ and $Y = \mathbf{1}_B$ are events, then X and Y are uncorrelated if and only if $P(A \cap B) = P(A)P(B)$ or (if $P(A) \neq 0$) if $P(B|A) = P(B)$. We say that the events are **independent**. For more general random variables, we will introduce a notion of independence which is much more restrictive than being uncorrlated.

2.5 The L_2 law of large numbers.

Suppose $\{Z_i\}$ is a family of pairwise uncorrelated random variables whose variances are uniformly bounded:

$$\mathrm{var}(Z_i) \leq K$$

for some constant K for all i. We can then apply Problem **2** to $y_i := Z_i - E(Z_i)\mathbf{1}$ to conclude that

$$\frac{1}{n}(Z_1 + \cdots + Z_n) - \frac{1}{n}\left(E(Z_1) + \cdots + E(Z_n)\,\mathbf{1} \to 0.\right.$$

For example, if all the $E(Z_i)$ are equal, say equal to some number m, then

$$\frac{1}{n}(Z_1 + \cdots + Z_n) \to m\mathbf{1}.$$

In this equation, convergence is in the sense of L_2. It will require more work for us to formulate and prove a corresponding result where convergence is "almost everywhere". We will do this after developing the machinery of measure theory - we will then be able to formulate and prove an "almost everywhere" version of the martingale convergence theorem.

But all versions of the law of large numbers are meant to justify the identification of the intuitive notion of probability with that of long term behavior.

We are going to greatly enlarge the class of "events" and this is going to take a lot of technical work. The idea is this: let S be any subset of $[0,1]$. We can consider the function 1_S where $1_S(x) = 1$ for $x \in S$ and $1_S(x) = 0$ for $x \in S$. Conversely, let f be any function on $[0,1]$ which takes on only the values 1 and 0, then we can let S be the set of x where $f(x) = 1$, and then $f = 1_S$. Now a function f which takes on only the values 0 and 1 can be characterized as follows:

$$0 \leq f(x) \leq 1 \quad \forall x,$$

$$\max\{f(x), 1 - f(x)\} = 1, \quad \forall x,$$

and

$$\min\{f(x), 1 - f(x)\} = 0 \quad \forall x.$$

Since the elements of $\mathbf{H}_\mathbf{R}$ are not functions, we can not use this characterization of the indicator function of a set directly. The trouble is with the quantifier $\forall x$. Changing "values" at a set of measure zero does not have any effect on an element of $\mathbf{H}_\mathbf{R}$. So what we need is to show that there is a notion of partial order $\preceq$ on which is compatable with the Hilbert space structure which is a substitute for the order relation $f(x) \leq g(x)\ \forall x$. Before embarking on this technical stuff, I thought it might be useful to develop some probablistic inuition for the concepts we are introducing.

3 Simulation.

The m.file randomwalk.m in the MATLAB folder simulates a one dimensional random walk with N steps. For convenience, I reproduce it here:

1. % plots a random walk of length N
2. close all
3. clc
4. N=input('N= ')
5. Q=zeros(1,N);
6. R=rand(1,N);
7. Q=Q+(R>.5);
8. P=2*Q-1;

9. W=[0 cumsum(P)];

10. plot(W)

Step 1 is not a command at all, and is just there to describe what the m.file does. Steps 2 and 3 clear the display and command window and Step 4 calls for the input of the number of random walk steps desired. Step 5 is just an initialization. The key steps are 6 and 7. The MATLAB command rand(M,N) produces a M $\times$ N matrix each of whose entries is a number "chosen at random" between 0 and 1, and all the matrix entries chosen completely independently of one another.

There is obviously a deep paradox involved in how a purely mechanical device like a computer can produce numbers "chosen at random". I will leave this issue for courses in the CS department or the philosophy department. Perhaps also the theology department. I will just take this computer capacity for granted.

In step 7, the entry R>.5 produces a matrix the same size as the matrix R, whose entries are 1 or 0 according as the corresponding entry of R is or is not >.5. In other words, it has the effect of applying the function $\mathbf{1}_{[.5,1]}$ to each of the entries of R. The net effect of step 7 is to produce a vector of length N whose entries are 0 or 1, each occurring with probability one-half. Step 8 converts this into a vector whose entries are ± 1 each with probability one-half. Step 9 produces a vector whose first entry is 0 and whose next entries are the cumulative sums of the entries of P and step 10 plots the entries of W against the index of the entry, i.e plots the points (i,W(i)) and joins these points by straight line segments.

Suppose we wanted to simulate an unfair coin where the probability of heads is p, not necessaily .5. We could then apply the function $\mathbf{1}_{[1-p,1]}$ to each of entries of R. This will produce a matrix whose etries are 0 or 1, with 0 occurring with probability $1-p$ and 1 with probability p.

More generally, suppose we want to simulate an experiment whose outcome can be any integer from 0 to n, where the probability of getting i is p_i. Then all we have to do is construct a "step function" whose values are these integers, where the value i is taken on on an interval of length p_i. For example, the function f graphed in Figure 1 will do. Then applying f to each entry of R will produce a simluation of independent trials of the experiment, one trial for each position of R.

In general, if we want to simulate a random variable which takes on non-negative integer values, we don't have to restrict ourselves to the function graphed in Figure 1. In fact, let g be any function defined on $[0,1]$ which takes on non-negative integer values, and suppose (for the sake of simplicity) that for each non-negative integer k, the set

$$g^{-1}(k)$$

is a finite union of intervals. (An interval can be open (without end points), closed (with both end points) or half open (with one end point as in Figure 1).) Suppose that

$$\mu[g^{-1}(k)] = p_k$$

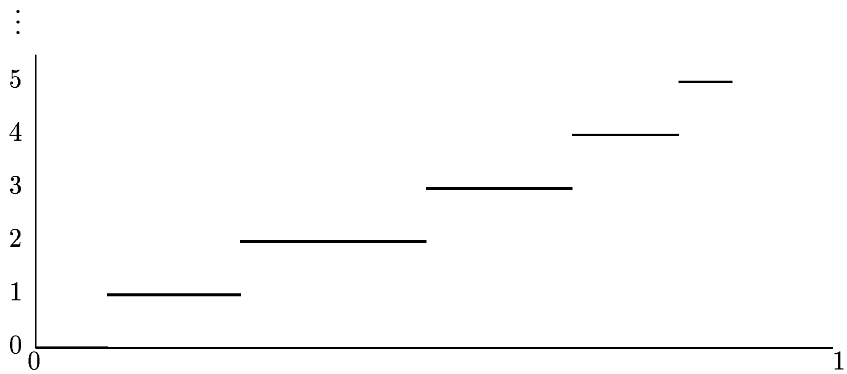

Figure 1: The graph of function f such that $f(hboxrand)$ simulates an experiment which has outcome i with probabiity p_i. The length of the $i-$th interval of constancy is p_i.

Here $\mu(S)$ denotes the "measure " of a set S. If S is a disjoint union of intervals, then $\mu(S)$ is the sum of the lengths of constituent intervals. If g satisfies this condition, then clearly

$$X = g(\text{rand})$$

simultate an experiment with

$$\text{Prob}(X = i) = p_i, \quad i = 0, 1, 2, \ldots, \quad \sum p_i = 1.$$

So the function f in Figure 1 is just one of many. It is characterized by being monotone (non-decreasing). Any function g that we choose belongs to (more precisely "represents") an element of $\mathbf{H_R}$, and this is why we call elements of $\mathbf{H_R}$ random variables.

We can allow for the possibility that any non-negative integer value be achieved. The condition that a g simulating such an experiment belong to $\mathbf{H_R}$ is that

$$\sum i^2 p_i < \infty.$$

Thus our space $\mathbf{H_R}$ will not simulate *all* random experiments, only the square integrable ones.

3.1 The Poisson distribution.

This describes an experiment whose outcome can be any non-negative integer, and where the probablity of having the outcome i is

$$p_i(\lambda) = \frac{\lambda^i}{i!} e^{-\lambda}.$$

Here λ is any non-negative real number. (In the mathematics literature, λ is a standard name for this parameter, but MATLAB prefers the English alphabet, so we will also use letters such as r or p for this parameter.)

4. Let X be a random variable which takes on only non-negative integer values i with probablity p_λ as given above. Show that

$$E(X) = \lambda \quad \text{and} \;\; \text{var}(X) = \lambda.$$

[A neat way of doing this is (for any random variable X that takes on only non-negative integer values n, with probability p_n) is to introduce the **generating function**

$$p_X(z) := \sum_0^\infty p_n z^n$$

and then to evaluate $p'(1)$ and $p''(1)$.]

As mentioned above, many different random variables can simulate the same experiment, but it will be convenient for us to choose the monotone step function as a standard simulator. So we let Y_λ denote the monotone step function with values in the non-negative integers, and where the interval where $Y_\lambda = i$ has length $p_i(\lambda)$.

The m.file poisson.m calls for the input of the parameter r $(= \lambda)$ and an integer N, and produces a row vector Q of size N whose entries are non-negative integers chosen independently according to the Poisson law. Here is the m.file:

```
%produces a vector Q of size N with integer entries
%distributed according to the Poisson distribution
N=input('N= ')
r=input('r= ')
p=exp(-r);
F=p;
i=1;
Q=zeros(1,N);
R=rand(1,N);
while p > eps;
Q=Q+(R > F);
p=(r/i)*p;
F=F+p;
i=i+1;
end
```

It works by first checking whether the entry is $> p_0(r) = e^{-r}$. If so, it adds 1 to that entry. Then it checks whether the entry is $> p_0(r) + p_1(r)$. If so, it adds another 1 etc. It computes $p_i(r)$ as it goes along. We do not want the machine to continue once i is so large that $p_i(r)$ is below the tolerance that MATLAB uses in its computations. This is the meaning of line 10: the command "eps" returns the value of the machine accuracy, and we want the procedure to stop when $p_i(r)$ gets below this value.

This program is not the most efficient when the parameter r is large, since it spends a lot of time checking events of low probability, but it is fine for low values of r and useful for our theoretical discussion.

3.2 Comparing Poisson and Bernoulli.

We will let B_p denote the monotone step function which is equal to 0 on the interval $[0, 1-p)$ and equal to 1 on the interval $[1-p, 1]$. The letter B stands for Bernoulli, who considered random variables which could take on only the values 0 and 1, and takes on the value 1 with probability p. Of course B_p is only one of many such random variables, but all have the same expectation and variance, namely

$$E(B_p) = p \quad \text{and} \quad \text{var}(B_p) = p - p^2.$$

We also note the following useful fact:

$$e^{-p} = 1 - p + \frac{p^2}{2!} - \frac{p^3}{3!} + \cdots$$

is (for $0 \leq p \leq 1$) an alternating series with terms decreasing in absolute value, and so $e^{-p} - 1 + p > 0$ or

$$1 - p < e^{-p}.$$

This says that the interval where $B_p = 0$ is shorter than the interval where $Y_p = 0$.

5. Show that the set $A := \{x \in [0,1] | B_p(x) \neq Y_p(x)\}$ consists of two intervals, and

$$P(A) \leq p^2.$$

We can write this result in the more convenient shorter notation

$$P(B_p \neq Y_p) \leq p^2.$$

We now interupt this discussion of simulation and return to $\mathbf{H}_{\mathbf{R}}$.

4 Order.

It makes sense to say that a real valued function is non-negative:

$$\phi \succeq 0 \quad \Leftrightarrow \quad \phi(t) \geq 0 \text{ for all } t.$$

We have two natural candidates for how to extend this notion to $\mathbf{H}_{\mathbf{R}}$ and we want to check that these two extended definitions are in fact the same. We could define

1. $g \succeq 0$ if $(g, \phi) \geq 0$ for all $\phi \succeq 0 \in \mathcal{C}([0,1])$ or we could define
2. $g \succeq 0$ if there is a sequence of $g_n \in \mathcal{C}([0,1])$ with $g_n \succeq 0$ and $g_n \to g$.

Clearly $2 \Rightarrow 1$, since $(g, \phi) = \lim(g_n, \phi) \geq 0$. We must show the reverse implication.

For any real number a define

$$a^+ = \frac{1}{2}(a + |a|) = \begin{cases} a & \text{if } a \geq 0 \\ 0 & \text{if } a \leq 0 \end{cases}.$$

By examining the four possibilities for the signs of a and b we check that

$$|a^+ - b^+| \leq |a - b|.$$

For a function ϕ we define ϕ^+ by

$$\phi^+(t) = (\phi(t))^+.$$

It follows from the above that if $\{g_n\}$ is a Cuachy sequence of elements of $\mathcal{C}([0,1])$ then so is g_n^+. So if $g_n \to g$ in $\mathbf{H}_{\mathbf{R}}$, then g_n^+ converges to some element, call it g^+. We want to show that $g^+ = g$. Since $g_n^+ - g_n \geq 0$, we have

$$(g, g_n^+ - g_n) \geq 0$$

by condition 2, and passing to the limit we get

$$\|g\|_2^2 = (g, g) \leq (g, g^+).$$

By the Cauchy-Schwartz inequality the right hand side is $\leq \|g\|_2 \|g^+\|_2$ so we get

$$\|g\|_2 \leq \|g^+\|_2.$$

On the other hand, from its very definition we know that

$$\|g_n^+\|_2^2 \leq \|g_n\|_2^2$$

since $|g_n^+(t)| \leq |g_n(t)|$ for all t. Passing to the limit gives $\|g^+|_2 \leq \|g\|_2$ so

$$\|g\|_2 = \|g^+\|_2.$$

But then

$$\|g-g^+\|_2^2 = \|g\|_2^2+\|g^+\|_2^2-2(g,g^+) \le |g^+\|_2^2+\|g\|_2^2-2\|g\|_2^2 = \|g^+\|_2^2-\|g\|_2^2 = 0$$

so $g = g^+$.

In short, it makes sense to ask whether or not $g \succeq 0$ for any $g \in \mathbf{H_R}$. That is, it makes sense to talk of non-negative random variables. If $g_n \in \mathbf{H_R}$ are non-negative random variables, and $g_n \to g$ then it follows from the first definition above that $g \geq 0$. In other words, the set of non-negative random variables is a closed subset of $\mathbf{H_R}$. The sum of two non-negative random variables is again non-negative, and the product of a non-negative random variable by a non-negative real number is again non-negative. Applying the first definition to f and the second definition to g we see that if $f \succeq 0$ and $g \succeq 0$ then $(f,g) \geq 0$.

In fact, the above proof shows that for *any* $g \in \mathbf{H_R}$ we can define its "non-negative part" g^+ as the limit of g_n^+ where $g_n \in \mathcal{C}([0,1]) \to g$ and define its non-positive part g^- as $(-g)^+$ and its absolute value $|g|$ as $|g| := g^+ + g^-$. Equally well we could define $|g|$ to be the limit of $|g_n|$ if g_n is a sequence of elements in $\mathcal{C}[0,1]$ converging to g. We have $\|g^+\|_2 \leq \|g\|_2$ and $\|g^-\|_2 \leq \|g\|_2$ while $|\ |g\ \|_2 = \|g\|_2$. We can also check that the map $g \mapsto g^+$ is continuous, and hence so is the map $g \mapsto |g|$.

We can now write

$$f \succeq g \quad \text{iff } f - g \succeq 0$$

and so

$$f \succeq g \ \Rightarrow f + h \succeq g + h$$

for any $h \in \mathbf{H_R}$.

For any real numbers a and b we have

$$\max\{a,b\} = \frac{1}{2}(a+b+|a-b|) \ \text{ and } \quad \min\{a,b\} = \frac{1}{2}(a+b-|a-b|).$$

So if f and g are functions, and we define

$$f \vee g := \frac{1}{2}(f+g+|f-g|), \quad f \wedge g := \frac{1}{2}(f+g-|f-g|),$$

then $(f \vee g)(t) = \max\{f(t), g(t)\}$ and $(f \wedge g)(t) = \min\{f(t), g(t)\}$. But the above defintions make sense in $\mathbf{H_R}$ and we shall use them.

For example, if f and g are elements of $\mathbf{H_R}$ and $f \succeq 0$ and $g \succeq 0$, then we can find $f_n \succeq 0$, $\ g_n \succeq 0$ in $\mathcal{C}([0,1])$ with $f_n \to f$ and $g_n \to g$. We know that $(f_n \wedge g_n)(t) \geq 0$ for all t, and so $f_n \wedge g_n \succeq 0$. From the continuity of the map $u \mapsto |u|$ we conclude that $f \wedge g \succeq 0$.

Therefore, if $f \succeq 0$ and $g \succeq 0$ then

$$(f,g) \geq (f, f \wedge g) \geq (f \wedge g, f \wedge g) > 0 \ \text{ unless } f \wedge g = 0.$$

Also, $f \vee g$ for any $f, g \in \mathbf{H_R}$ can be characterized as being the "least upper bound" of f and g in the sense that $f \vee g \succeq f$, $f \vee g \succeq g$, and if $h \succeq f$ and $h \succeq g$

then $h \succeq (f \vee g)$. similarly $f \wedge g$ can be characterized as being the "greates lower bound" of f and g.

If $f, g, h \in \mathbf{H_R}$ then

$$\begin{aligned} f - (g \vee h) &= f - \frac{1}{2}(g + h + |g - h|) \\ &= \frac{1}{2}(f - g + f - h - |f - g - (f - h)|) \\ &= (f - g) \wedge (f - h). \end{aligned} \tag{4}$$

Similarly, if follows directly from the definition that the "distributive laws"

$$f \wedge (g \vee h) = (f \wedge g) \vee (f \wedge h) \tag{5}$$

$$f \vee (g \wedge h) = (f \vee g) \wedge (f \vee h) \tag{6}$$

hold.

Finally, from the "triangle inequality" $|u + v| \preceq |u| + |v|$ one can conlude that the oprations $f \vee g$ and $f \wedge g$ are continous¡ e.g. $f_n \to f$ and $g_n \to g$ implies that $f_n \vee g_n \to f \vee g$ and similarly of $\wedge$. We could keep going on, so I will just assign

6. Put in the details for these and similar arguments until you get tired. But be sure to include the following fact that we will use:

$$\text{If } f \wedge g = 0 \text{ then } f \vee g = f + g. \tag{7}$$

4.1 More Events.

TO BE CONTUNED.