Families of σ-fields.

A set, Ω, may have many different σ-fields defined on it. Certainly there is a smallest and a largest. These being $\{\emptyset, \Omega\}$ and the set of all subsets of Ω respectively. One way of thinking of σ-fields is to regard them as <u>information</u>. This information may be rather "coarse" or as "refined" as is possible, the former may be modeled by the σ-field $\{\emptyset, \Omega\}$ where the information may be summarised as "something in Ω occurs". This is not altogether enlightening since Ω is chosen to represent all possible outcomes in any case. On the other hand the information may be as detailed as is possible and this may be modeled by the power set of Ω — if this is appropriate. In this case all possible out-comes and <u>any</u> combination of them constitutes an event — a piece of information.

In using Probability Theory to model actual situations one usually encounters something (intermediate) between coarse and refined information. Indeed the information usually changes as time progresses. We can model a simple case of this by considering an increasing sequence of σ-fields $\mathcal{F}_n$, $n = 0, 1, 2, \ldots$ with $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$.

Here we regard $\mathcal{F}_n$ as "the information available at time $t = n$". There is no loss of information as time progresses and so the σ-field at time $n+1$ is at least as big as the one at time n.

<u>Example</u>

Let $\Omega = (0,1]$, $\mathcal{F}_0 = \{\phi, \Omega\}$, $\mathcal{F}_1 = \{\phi, \Omega, (0,\frac{1}{2}], (\frac{1}{2},1]\}$

and $\mathcal{F}_n = \{E : \; E = \bigcup_{i=1}^{k} J_i$ where $0 \le k \le 2^n$ and $J_i \in \{(0,\frac{1}{2^n}], (\frac{1}{2^n},\frac{2}{2^n}], \dots, (1-\frac{1}{2^n},1]\}\}$

It is clear, that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots$.

A natural question arises when we have a system like this: Suppose that we have an increasing family of σ-fields $(\mathcal{F}_n)$ defined on Ω with $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$, where $\mathcal{F}$ is some (large) σ-field on Ω and suppose also $\mathbb{P}$ is a probability measure on $\mathcal{F}$. Then $\mathbb{P}$ is a probability measure on $\mathcal{F}_n$ too. We think of $(\Omega, \mathcal{F}, \mathbb{P})$ as being the probability space that governs the situation we are investigating while $(\Omega, \mathcal{F}_n, \mathbb{P})$ is the probability space "of events known at time n". One can think of $(\Omega, \mathcal{F}_n, \mathbb{P})$ as an "estimate" of $(\Omega, \mathcal{F}, \mathbb{P})$. We are interested in the analysis of $(\Omega, \mathcal{F}, \mathbb{P})$ and the ~~the~~ random variables associated with it. But all we know, at time n, is $(\Omega, \mathcal{F}_n, \mathbb{P})$, so we would like to <u>predict</u> things about objects associated with $(\Omega, \mathcal{F}, \mathbb{P})$ by means of objects associated with $(\Omega, \mathcal{F}_n, \mathbb{P})$

In order to give a mathematically precise idea of prediction we must digress.

Probabalistically Equivalent Random Variables

If X and Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}\{\omega : X(\omega) = Y(\omega)\} = 1$ then we say X and Y are equivalent or $\mathbb{P}$-equivalent. For the purposes of integration, and much probability theory, X and Y are the same random variable. So from now on, for this section of the course, we will identify random variables which are $\mathbb{P}$-equivalent

NB. Precisely we define $X \sim Y \iff X = Y$ $\mathbb{P}$-a.s. Then $\sim$ is an equivalence relation. We regard a random variable and the equivalence class it belongs to as the same thing. Addition of classes, etc, are defined as follows, let $\langle f \rangle$ denote the class of f.

$$\langle f \rangle + \langle g \rangle = \langle f + g \rangle,$$

$$\langle f \rangle \langle g \rangle = \langle fg \rangle,$$

$$\lambda \langle f \rangle = \langle \lambda f \rangle, \quad \lambda \in \mathbb{R}.$$

END

Let $L^2(\Omega, \mathcal{F}, \mathbb{P})$ be the set of (equivalence classes of) random variables, f, for which $\int_\Omega |f(\omega)|^2 d\mathbb{P} < \infty$

Then, with the operations of addition and scalar multiplication defined above $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a linear space:

(a)

$$\int_\Omega |\lambda f(\omega)|^2 \, d\mathbb{P} = |\lambda|^2 \int_\Omega |f(\omega)|^2 d\mathbb{P} < \infty \quad \text{when}$$

$\int_\Omega |f|^2 d\mathbb{P} < \infty$ and $\lambda \in \mathbb{R}$.

(b)

$$\int_\Omega |f(\omega) + g(\omega)|^2 d\mathbb{P} \leqslant \int_\Omega |f(\omega) + g(\omega)|^2 d\mathbb{P} + \int_\Omega |f(\omega) - g(\omega)|^2 d\mathbb{P}$$

$$= \int_\Omega \left((f(\omega)+g(\omega))^2 + (f(\omega)-g(\omega))^2 \right) d\mathbb{P}$$

$$= 2\int_\Omega |f(\omega)|^2 d\mathbb{P} + 2\int_\Omega |g(\omega)|^2 d\mathbb{P} < \infty$$

whenever $\int_\Omega |f(\omega)|^2 d\mathbb{P} < \infty > \int_\Omega |g(\omega)|^2 d\mathbb{P}$.

In fact $L^2(\Omega, \mathcal{F}, \mathbb{P})$ has even more structure

Define,

$$\langle f, g \rangle = \int_\Omega f(\omega) g(\omega) \, d\mathbb{P} \qquad f, g \in L^2$$

Then $\langle \circ, \circ \rangle$ is an inner product on L^2, so L^2 is an <u>inner product space.</u>

Further, $\|f\| = \langle f, f \rangle^{1/2} = \left(\int_\Omega |f(\omega)|^2 d\mathbb{P} \right)^{1/2}$ is a norm on L^2, making it a complete normed

42

Insert 1

There is one fact about Hilbert Spaces I want you to learn here and now. It may seem rather curious to you at first but its use will make it sensible.

Let $f, g \in \mathcal{H}$ a Hilbert Space. Then

$$f = g \iff \forall h \in \mathcal{H} \ \langle f, h \rangle = \langle g, h \rangle.$$

In other words, if the inner products of f and g against any h in $\mathcal{H}$ always agree they must be equal.

space. If you have not met norms yet, don't worry, you have met the idea of a <u>metric space</u>. Now $d(f,g) = \|f-g\| = \left(\int_\Omega |f-g|^2 d\mathbb{P}\right)^{1/2}$ is a metric on $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and L^2 is <u>complete</u>, as a metric space, in the metric $d(\cdot,\cdot)^{(t)}$.
For those that have met them, $L^2(\Omega,\mathcal{F},\mathbb{P})$ is a Hilbert Space — a complete inner product space.

(t) Recall that complete means every Cauchy sequence in the metric space converges to an element of the metric space.

$\boxed{\text{INSERT 1}}$ HERE

Consider now the <u>two</u> L^2-spaces $L^2(\Omega,\mathcal{F},\mathbb{P})$ and, for some fixed $n \in \mathbb{N}$, $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$.
Since $\mathcal{F}_n \subseteq \mathcal{F}$ any $\mathcal{F}_n$-simple random variable is also a $\mathcal{F}$-simple random variable. Hence $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$ is a subspace of $L^2(\Omega,\mathcal{F},\mathbb{P})$ and, as $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$ is complete, it is a <u>closed subspace</u> of $L^2(\Omega,\mathcal{F},\mathbb{P})$. This means simply that $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$ is a subspace of $L^2(\Omega,\mathcal{F},\mathbb{P})$ (as a vector space) and any Cauchy sequence in $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$ converges to something in $L^2(\Omega,\mathcal{F}_n,\mathbb{P})$.

At this point we need to import a result from the theory of Hilbert spaces. It is quite simple to state and easy to visualise and we will accept what it says without proof."

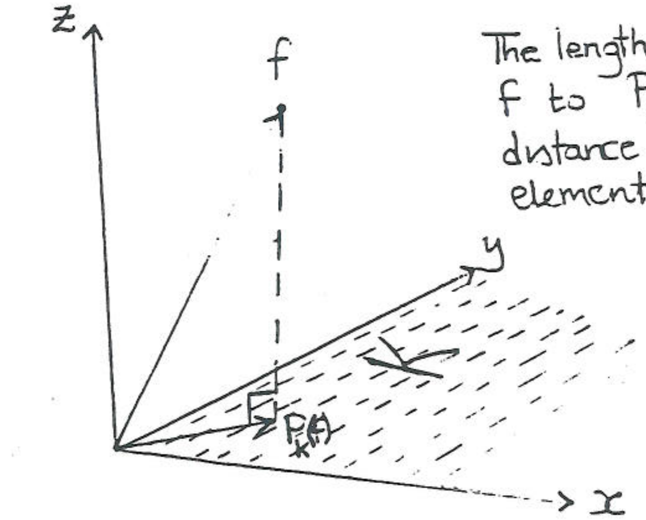
<u>Theorem</u> (H 1)

Let H be a Hilbert Space and K a closed linear subspace of H. If $f \in H$ there is a unique element of K, $P_K(f)$ say, such that

$$\| f - P_K(f) \| \leqslant \| f - k \| \quad \forall k \in K.$$

In other words, there is, in K, a unique point which is <u>closest to f</u>.

<u>Remark</u> In $\mathbb{R}^3$ — a Hilbert Space — this amounts to nothing other than orthogonal projection :



The length of the line from f to $P_K(f)$ is the shortest distance from f to an element of K. (the xy plane.

The remarkable thing is that this picture holds good in <u>any</u> Hilbert Space and hence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Returning, briefly, to our ideas of using objects associated with $(\Omega, \mathcal{F}_n, \mathbb{P})$ to predict things about objects in $(\Omega, \mathcal{F}, \mathbb{P})$ we see that if we restrict ourselves to $L^2(\Omega, \mathcal{F}, \mathbb{P})$ then there is a natural choice of elements of $L^2(\Omega, \mathcal{F}_n, \mathbb{P})$. We simply take the element of $L^2(\Omega, \mathcal{F}_n, \mathbb{P})$ which is closest to f, ie. $P_{L^2(\Omega, \mathcal{F}_n, \mathbb{P})}(f)$.

As $f \in \mathcal{H}$ varies the values of $P_K(f)$ will vary. There is a fundemental relationship from which all of the properties of the <u>map</u> $f \longmapsto P_K(f)$ follow it is,

$$\langle P_K(f), k \rangle = \langle f, k \rangle \quad - **$$

for $f \in \mathcal{H}$ and $k \in K$. This now becomes an <u>assumption</u> of our course.

In the particular case where $\mathcal{H}$ is $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and K is $L^2(\Omega, \mathcal{G}, \mathbb{P})$, for some σ-field, $\mathcal{G} \subseteq \mathcal{F}$, the map P_K is called the conditional expectation of $\mathcal{F}$ onto $\mathcal{G}$. Note that when $\mathcal{G} \subseteq \mathcal{F}$, $\mathcal{G}$ and $\mathcal{F}$ σ-fields on Ω then $L^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$, indeed $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, it is a closed subspace for $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is <u>complete</u> as a metric space.

<u>Properties of the Conditional Expectation</u>

These fall into two parts. There are those that follow from the fact that

8

It is the orthogonal projection onto a closed subspace of a Hilbert space, and those that arise from the fact that the Hilbert space is a space of random variables.

Theorem (M1)

Let M_G denote the orthogonal projection of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then,

(i) M_G is a linear mapping.

(ii) $\| M_G(f) \| \leqslant \| f \|$

(iii) If $f(\omega) \geqslant 0$ for $\omega \in \Omega$ then $M_G(f) \geqslant 0$ too.

(iv) If $f(\omega) \leqslant C$ $\forall\, \omega \in \Omega$ then $M_G(f) \leqslant C$ too.

(v) $M_G(I_\Omega) = I_\Omega$

(vi) $M_G \circ M_G = M_G$

(vii) $M_G(fg) = M_G(f) \cdot g$ for $f \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and bounded $g \in L^2(\Omega, G, \mathbb{P})$.

(viii) $\mathbb{E}(M_G(f)) = \mathbb{E}(f)$.

Proof

(i) For this situation $\langle f, k \rangle = \int_\Omega f.k \, d\mathbb{P}$.

Note that,

$$\langle \alpha f_1 + \beta f_2, k \rangle = \int_\Omega (\alpha f_1 + \beta f_2) k \, d\mathbb{P} = \alpha \int_\Omega f_1 k \, d\mathbb{P} + \beta \int_\Omega f_2 k \, d\mathbb{P}$$

$$= \alpha \langle f_1, k \rangle + \beta \langle f_2, k \rangle$$

Now, from the fact that $\langle M_G(f) . k \rangle = \langle f, k \rangle$

$$= \langle (\alpha M_G(f_1) + \beta M_G(f_2)), k \rangle .$$ Now this is true for every k in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. This means that (†)

$$M_G(\alpha f_1 + \beta f_2) = \alpha M_G(f_1) + \beta M_G(f_2) .$$

(†) In a Hilbert Space, if $\langle x, k\rangle = \langle y, k\rangle$ for <u>all</u> k then in particular $\langle x, x-y\rangle = \langle y, x-y\rangle$ i.e. $\langle x-y, x-y\rangle = 0$ or $\|x-y\|^2 = 0$ (i.e. $d(x,y)=0$) so $x = y$.

(vi) Recall that $M_G(f)$ is the element of $L^2(\Omega, G, \mathbb{P})$ which is closest to f. If f is actually <u>in</u> $L^2(\Omega, G, \mathbb{P})$ this must be f itself, so $M_G(M_G(f)) = M_G(f)$.

(v) $\Omega \in G$ so $I_\Omega \in L^2(\Omega, G, \mathbb{P})$ so

$$M_G(I_\Omega) = I_\Omega \text{ by (vi).}$$

(ii) Observe $\langle f, k\rangle = \int_\Omega fk \, d\mathbb{P} = \int_\Omega kf \, d\mathbb{P} = \langle k, f\rangle$.

From the relation, **,

$$\langle M_G(f), k\rangle = \langle f, k\rangle$$

we reason that

$$\langle M_G(f), M_G(f)\rangle = \langle f, M_G(f)\rangle = \langle M_G(f), f\rangle .$$

Now $f = M_G(f) + f - M_G(f)$ so,

$$\langle f, f\rangle = \langle M_G(f) + (f - M_G(f)), M_G(f) + (f - M_G(f))\rangle$$

$$= \langle M_G(f), M_G(f)\rangle + \langle M_G(f), f - M_G(f)\rangle +$$

$$\langle f - M_G(f), M_G(f)\rangle + \langle f - M_G(f), f - M_G(f)\rangle$$

Now

$$\langle M_G(f), f - M_G(f) \rangle = \langle f - M_G(f), M_G(f) \rangle$$

using the relation ** $= \langle M_G(f - M_G(f)), M_G(f) \rangle$

and as M_G is linear $= \langle M_G(f) - M_G(M_G(f)), M_G(f) \rangle$

which by (vi) is $= \langle M_G(f) - M_G(f), M_G(f) \rangle$

$$= \langle 0, M_G(f) \rangle$$

$$= 0.$$

So $\|f\|^2 = \langle f, f \rangle = \langle M_G(f), M_G(f) \rangle + \langle f - M_G(f), f - M_G(f) \rangle$

$$= \| M_G(f) \|^2 + \| f - M_G(f) \|^2$$

$$\geqslant \| M_G(f) \|^2$$ proving (ii).

(iii) Let $g = I_{\{\omega : M_G(f) < 0\}}$ then as $f \geqslant 0$,

$0 \leqslant fg$ and so $0 \leqslant \int_\Omega fg \, d\mathbb{P}$ but note

that $\{\omega : M_G(f) < 0\} \in G$ so $g \in L^2(\Omega, G, \mathbb{P})$,

hence,

$$0 \leqslant \int_\Omega fg \, d\mathbb{P} = \langle f, g \rangle = \langle M_G(f), g \rangle \text{ by **}$$

$$= \int_\Omega M_G(f) I_{\{M_G(f) < 0\}} \, d\mathbb{P}$$

$$\leqslant 0$$

So $\int_\Omega M_G(f) I_{\{M_G(f) < 0\}} \, d\mathbb{P} = 0$