1 Probability theory

1.1 Basics

Consider a *finite sample space* Ω

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}, \qquad M < \infty$$

Define a *probability measure* P on Ω such that

$$P(\{\omega_i\}) = P(\omega_i) = p_i > 0, \qquad i = 1, \ldots, M$$
$$\sum_{i=1}^{M} p_i = 1.$$

For every subset A of Ω, $A \subseteq \Omega$, we have that

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

A *random variable* X on Ω is a mapping

$$X : \Omega \longrightarrow \mathbb{R}.$$

The *expectation* of X is defined as

$$E[X] = \sum_{i=1}^{M} X(\omega_i) P(\omega_i).$$

1.2 Sigma-algebras and information

It is important to know which information is available to investors. This is formalized using σ-algebras and filtrations.

Definition 1 *A collection $\mathcal{F}$ of subsets of Ω is called a* σ-algebra *(or* σ-field*) if the following hold.*

1. $\emptyset \in \mathcal{F}$.

2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

3. If $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ then $\bigcup_{n=1}^{\infty} \in \mathcal{F}$

Remark 1 When working on a finite sample space Ω condition 3 will reduce to

3.' If A and B $\mathcal{F}$ then $A \cup B \in \mathcal{F}$.

Example 1 The following are examples of σ-algebras.

1. $\mathcal{F} = 2^\Omega = \{A | A \subseteq \Omega\}$, the power set of Ω.
2. $\mathcal{F} = \{\emptyset, \Omega\}$, the trivial σ-algebra.
3. $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$.

□

Definition 2 *A set $\mathcal{P} = \{A_1, \dots, A_n\}$ of nonempty subsets of the sample space Ω is called a (finite)* partition *of Ω if*

1. $\bigcup_{i=1}^{n} A_i = \Omega$
2. $A_i \cap A_j = \emptyset$ *for* $i \neq j$

The σ-algebra consisting of all possible unions of the A_i:s (including the empty set) is called the σ-algebra generated by $\mathcal{P}$ *and is denoted by $\sigma(\mathcal{P})$.*

Remark 2 On a finite sample space every σ-algebra is generated by a partition.

When making decisions investors may only use the information available to them. This is formalized by measurability requirements.

Definition 3 *A function $X : \Omega \longrightarrow \{x_1, \dots, x_K\}$ is* $\mathcal{F}$-measurable *if*

$$f^{-1}(x_i) = \{\omega \in \Omega | X(w) = x_i\} \in \mathcal{F} \quad \textit{for all } x_i$$

If X is $\mathcal{F}$-measurable we write $f \in \mathcal{F}$.

Remark 3 Let $\mathcal{F} = \sigma(\mathcal{P})$. Then a function $f : \Omega \longrightarrow \mathbb{R}$ is $\mathcal{F}$-measurable if and only if f is constant on each set A_i, $i = 1, \dots, n$.

This captures the idea that based on the available information we should be able to determine the value of X.
Measurability is preserved under a lot of operations which is the content of the next proposition.

Proposition 1 *Assume that X and Y are $\mathcal{F}$-measurable. Then the following hold:*

1. *For all real numbers α and β the functions*

   $$\alpha X + \beta Y, \qquad X \cdot Y$$

   are measurable.

2. *If $Y(\omega) \neq 0$ for all ω, then*

   $$\frac{X}{Y}$$

   is measurable.

3. *If $\{X_n\}_{n=1}^{\infty}$ is a (countable) sequence of measurable functions, then the functions*

   $$\sup_n X_n, \quad \inf_n X_n, \quad \limsup_n X_n, \quad \liminf_n X_n,$$

   are measurable.

Definition 4 *Let X be a function $X : \Omega \longrightarrow \mathbb{R}$. Then $\mathcal{F} = \sigma(X)$ is the smallest σ-algebra such that X is $\mathcal{F}$-measurable.*
If $X_1, \dots, X_n$ are functions such that $X_i : \Omega \longrightarrow \mathbb{R}$, then $\mathcal{F} = \sigma(X_1, \dots, X_n)$ is the smallest σ-algebra such that $X_1, \dots, X_n$ are $\mathcal{G}$-measurable.

The next proposition formalizes the idea that if Z is measurable with respect to a certain σ-algebra, then "the value of Z is completely determined by the information in the σ-algebra".

Proposition 2 *Let $X_1, \ldots, X_n$ be mappings such that $X_i : \Omega \longrightarrow \mathbb{R}$. Assume that $Z : \Omega \longrightarrow \mathbb{R}$ is $\sigma(X_1, \ldots, X_n)$-measurable. Then there exists a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that*

$$Z(\omega) = f(X_1(\omega), \ldots, X_n(\omega)).$$

We also need to know what is meant by independence. Recall that two events A and B on a probability space $(\Omega, \mathcal{F}, P)$ are independent if

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

For σ-algebras and random variables on $(\Omega, \mathcal{F}, P)$ we have the following definition.

Definition 5 *The σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are* independent *if*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) \quad \textit{whenever } A_i \in \mathcal{F}_i,\ i = 1 \ldots, n.$$

Random variables $X_1, \ldots, X_n$ are independent if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.

1.3 Stochastic processes and filtrations

Let $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.

Definition 6 *A* stochastic process $\{S_n\}_{n=0}^{\infty}$ *on the probability space $(\omega, \mathcal{F}, P)$ is a mapping*

$$S : \mathbb{N} \times \Omega \longrightarrow \mathbb{R}$$

such that for each $n \in \mathbb{N}$

$$S_n(\cdot) : \Omega \longrightarrow \mathbb{R}$$

is $\mathcal{F}$-measurable

Note that $S_n(\omega) = S(n, \omega)$. We have that for a fixed n

$$\omega \longrightarrow S(n, \omega)$$

is a random variable. For a fixed ω

$$n \longrightarrow S(n, \omega)$$

is a deterministic function of time, called the *realization* or *sample path* of S for the outcome ω.

Remark 4 In this course we will mostly be looking at a fixed time horizon so the process will only live up until time T, that is we will be looking at processes $\{S_n\}_{n=0}^{T}$.

A stochastic process generates information and as before this is formalized in terms of σ-algebras, only now there will be a time dimension as well.

Definition 7 *Let $\{S_n\}_{n=0}^{\infty}$ be random process on $(\Omega, \mathcal{F}, P)$. The σ-algebra generated by S over $[0,t]$ is defined by*

$$\mathcal{F}_t^S = \sigma\{S_n; n \leq t\}.$$

We interpret $\mathcal{F}_t^S$ as the information generated by observing S over the time interval $[0,t]$. More generally information developing over time is formalized by filtrations. The are families of increasing σ-algebras.

Definition 8 *A* filtration $\underline{\mathcal{F}} = \{\mathcal{F}_n\}_{n\geq 0}$ *on $(\Omega, \mathcal{F}, P)$ is an indexed family of σ-algebras on Ω such that*

1. *$\mathcal{F}_n \subseteq \mathcal{F}, \quad n \geq 0$,*
2. *if $m \leq n$ then $\mathcal{F}_m \subseteq \mathcal{F}_n$.*

Remark 5 As stated before, we will mostly be looking at a fixed time horizon in this course so the filtration will only live up until time T, that is we will be looking at filtrations $\underline{\mathcal{F}} = \{\mathcal{F}_n\}_{n=0}^{T}$.

For stochastic process the following measurability conditions are relevant.

Definition 9 *Given a filtration $\underline{\mathcal{F}}$ and a random process S on $(\Omega, \mathcal{F}, P)$ we say that S is* adapted *to $\underline{\mathcal{F}}$ if*

$$S_n \in \mathcal{F}_n \qquad \textit{for all } n \geq 0,$$

and S is predictable *with respect to $\underline{\mathcal{F}}$ if*

$$S_n \in \mathcal{F}_{n-1} \qquad \textit{for all } n \geq 1.$$

1.4 Conditional expectation

Let X be a random variable on $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ a σ-algebra such that $\mathcal{G} \subseteq \mathcal{F}$. In this section we aim to define the expectation of X given the information in $\mathcal{G}$, or conditional on $\mathcal{G}$, $E[X|\mathcal{G}]$. We will do this in three steps.

1. First we will define the expectation of X given a set $B \in \mathcal{F}$, such that $P(B) \neq 0$, i.e. $E[X|B]$. Recall that

   $$E[X] = \sum_{i=1}^{M} X(\omega_i)P(\omega_i) = \sum_{\omega\in\Omega} X(\omega)P(\omega)$$

   Now it would seem natural (?) to use the normalized probabilities

   $$\frac{P(\omega_i)}{P(B)} \qquad \text{on } B.$$

   We thus define

   $$E[X|B] = \sum_{\omega_i\in B} X(\omega_i)\frac{P(\omega_i)}{P(B)} = \frac{1}{P(B)}\sum_{\omega\in B} X(\omega)P(\omega)$$

Example 2 Consider the finite sample space $\Omega = \{\omega_1, \omega_2, \omega_3\}$ endowed with the power σ-algebra $\mathcal{F} = 2^\Omega$, and a probability measure P such that $P(\omega_i) = 1/3,\ i = 1, 2, 3$. Furthermore let $B_1 = \{\omega_1, \omega_3\}$, $B_2 = \{\omega_2\}$, $\mathcal{P} = \{B_1, B_2\}$, and $\mathcal{G} = \sigma(\mathcal{P})$. Finally, let

$$X(\omega) = I_{\{\omega_1\}}(\omega) = \begin{cases} 1, & \text{if } \omega = \omega_1 \\ 0, & \text{otherwise.} \end{cases}$$

Then we have that

$$E[X] = \sum_{\omega \in \Omega} X(\omega)P(\omega) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{3}$$

and that

$$E[X|B_1] = \frac{1}{P(B_1)} \sum_{\omega \in B_1} X(\omega)P(\omega) = \frac{1}{1/3 + 1/3}\left(1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}\right) = \frac{1}{2},$$

whereas

$$E[X|B_2] = \frac{1}{P(B_2)} \sum_{\omega \in B_2} X(\omega)P(\omega) = \frac{1}{1/3} \cdot 0 \cdot \frac{1}{3} = 0.$$

□

2. Next we will define the expectation of X conditional on a partition $\mathcal{P}$ of Ω. Suppose that $\mathcal{P} = \{B_1, \ldots, B_K\}$ and that $P(B_i) \neq 0,\ i = 1, \ldots, K$. Note that for any random variable Y measurable with respect to $\sigma(\mathcal{P})$ we have that if $\omega_i \in B_j$

$$Y(\omega_i) = E[Y|B_j]$$

since Y is constant on each B_i. This means that

$$Y(\omega) = \sum_{i=1}^{K} E[Y|B_i] I_{B_i}(\omega),$$

where I_{B_i} denotes the indicator function of B_i, i.e.

$$I_{B_i}(\omega) = \begin{cases} 1, & \text{if } \omega \in B_i \\ 0, & \text{otherwise.} \end{cases}$$

We now define

$$E[X|\mathcal{P}](\omega) = \sum_{i=1}^{K} E[X|B_i] I_{B_i}(\omega).$$

Note that this means that $E[X|\mathcal{P}]$ is a random variable Z such that

(a) $Z \in \sigma(\mathcal{P})$ and that

(b) for all $B \in \sigma(\mathcal{P})$ we have that

$$\sum_{\omega \in B} Z(\omega)P(\omega) = \sum_{\omega \in B} X(\omega)P(\omega).$$

Example 3 Continuing on Example 2 we can compute

$$E[X|\mathcal{P}] = \sum_{i=1}^{2} E[X|B_i] I_{B_i}(\omega) = \frac{1}{2} \cdot I_{B_1}(\omega) + 0 \cdot I_{B_2}(\omega).$$

□

3. Now we are ready to give the general definition of $E[X|\mathcal{G}]$.

   Definition 10 *Consider a random variable X on $(\Omega, \mathcal{F}, P)$ and a σ-algebra $\mathcal{G}$ such that $\mathcal{G} \subseteq \mathcal{F}$. The* conditional expectation of X given $\mathcal{G}$ *denoted $E[X|\mathcal{G}]$ is any random variable Z such that*

   (a) $Z \in \mathcal{G}$, and

   (b) for all $A \in \mathcal{G}$ we have that

   $$\sum_{\omega \in A} Z(\omega) P(\omega) = \sum_{\omega \in A} X(\omega) P(\omega).$$

The proposition below states some properties of the conditional expectation.

Proposition 3 *The conditional expectation has the following properties. Suppose that $\alpha, \beta \in \mathbb{R}$ and that X and Y are random variables on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a σ-algebra such that $\mathcal{G} \subseteq \mathcal{F}$. Then the following hold.*

1. *Linearity.*
   $$E[\alpha X + \beta Y|\mathcal{G}] = \alpha E[X|\mathcal{G}] + \beta E[Y|\mathcal{G}].$$
2. *Monotonicity. If $X \leq Y$ then*
   $$E[X|\mathcal{G}] \leq E[Y|\mathcal{G}].$$
3. $$E[E[X|\mathcal{G}]] = E[X].$$
4. *If $\mathcal{H}$ is σ-algebra such that $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ then*

   (a)
   $$E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{H}],$$

   (b)
   $$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}].$$

   Thus "the smallest σ-algebra always wins".
5. *Jensen's inequality. If φ is a convex function, then*
   $$\varphi(E[X|\mathcal{G}]) \leq E[\varphi(X)|\mathcal{G}].$$
6. *If X is independent of $\mathcal{G}$ then*
   $$E[X|\mathcal{G}] = E[X].$$
7. *Taking out what is known. If $X \in \mathcal{G}$ then*
   $$E[XY|\mathcal{G}] = X \cdot E[Y|\mathcal{G}].$$