Lecture 1: Probability Theory

In this lecture I review some basic concepts in probability theory. Part of my purpose is to demystify the use of these terms, which appear too often in economics papers.

1. Probability spaces

A formal definition of probability requires specifying the events from which a realization is drawn, the sets of events over which probabilities may be defined and the probabilities themselves. Formally, our interest is in \((\Omega, F, \mu)\), called a probability space, where

- \(\Omega\) is the set that constitutes the sample space
- \(F\) is a \(\sigma\) – algebra of sets, each of which is a subset of \(\Omega\)
- \(\mu\) is a probability measure

What is a \(\sigma\) – algebra of sets? An algebra of sets (defined with respect to \(\Omega\) is a collection of subsets of \(\Omega\) such that 1) \(\Omega\) is in the collection and 2) the collection is closed under finite complements and unions. A \(\sigma\) – algebra is an algebra of sets that is also closed under countable unions.

What role is played by the \(\sigma\) – algebra in describing a probability space? In order for probabilities to be defined coherently in a sense made precise with the definition of a probability measure, the assignment of probabilities may not be possible for all subsets of \(\Omega\). The restriction of probabilities to a \(\sigma\) – algebra avoid such technical problems.
A standard example of a $\sigma$–algebra are the Borel sets of $\mathbb{R}$. The Borel sets are the smallest $\sigma$–algebra of subsets of $\mathbb{R}$ that contains all half-open intervals of the form $(a, b]$. It is easy to see that this is a rich collection!

A probability measure $\mu(\cdot)$ is a mapping from $F$ to $\mathbb{R}$ such that 1) $\mu(f) \geq 0 \ \forall f \in F$, 2) $\mu(\Omega) = 1$, 3) for any countable collection of disjoint sets

$$\sum_i \mu(f_i) = \mu\left(\bigcup_i f_i\right)$$

Notice that these requirements are very intuitive. The first simply says that probabilities cannot be negative, the second that the probability something happens in 1 and the third generalizes the idea that if events $A$ and $B$ are mutually exclusive, the probability that either happens is the sum of the probabilities of the individual events.

2. Random variables

Definition

For a probability space $(\Omega, F, \mu)$, suppose there exist a mapping $x$ from $\Omega$ to such that for every member $B$ of the Borel sets, the set defined by $\{\omega : x(\omega) \in B\}$ is an element of $F$. In this case $x(\omega)$ is a random variable. The probability space $(\Omega, F, \mu)$ determines the probabilities associated with $x$.

3. Stochastic processes

One can also use a probability space to construct a stochastic process. In order to do so, it is necessary to define a measure preserving transformation.

Definition
A measure preserving transformation $T$ is a mapping from $F$ to $F$ such that 
\[ \mu(T^{-1}f) = \mu(f) \forall f. \]

One can use a measure preserving transformation to define a sequence of random variables $x_t = x(T^t \omega)$. Such a sequence is strictly stationary. To prove this, recall that strict stationarity requires that $\forall a, b$

\[ \mu(a \leq x_t \leq b) = \mu(\omega : a \leq x(T^t \omega) \leq b) \quad (1.2) \]

Let $C$ denote this set of $\omega$'s. Since $T$ is a measure preserving transformation, it must be the case that for $D = T^{-1}C$

\[ \mu(C) = \mu(D) \quad (1.3) \]

By (1.2),

\[ \mu(D) = \mu(\omega : a \leq x(T^{t+1} \omega) \leq b) = \mu(\omega : a \leq x_{t+1} \leq b) \quad (1.4) \]

which verifies stationarity.

4. Ergodicity

The measure preserving transformation is important as its properties will determine much about what may learned about the probability structure of a time series from a given sample path realization. The key property in this regard is ergodicity. As a preliminary, we need the definition of an invariant set.
Definition. Invariant set

A set $S$ is invariant under a given measure preserving transformation $T$ is invariant if $S = T^{-1}S$. A set $S$ is almost invariant if $\mu(S \Delta T^{-1}S) = 0$

Similarly,

Definition. Almost invariant set

A set $S$ is almost invariant if $\mu(S \Delta T^{-1}S) = 0$

This leads to the definition of ergodicity.

Definition. Ergodicity

A measure preserving transformation is ergodic if for every invariant set $S$, either $\mu(S) = 0$ or $\mu(\Omega - S) = 0$.

This definition will also apply if invariant is replaced with almost invariant,

Ergodicity is deeply related to the law of large numbers.

Ergodic Theorems

1. If $x(\omega)$ is a random variable with finite expectation, then,

$$ \lim_{n \to \infty} n^{-1} \sum_{j=0}^{n-1} x(T^j \omega) = \hat{x}(\omega) $$

(1.5)
where \( \hat{x}(\omega) \) exists and is finite with probability 1.

2. If \( T \) is ergodic, then

\[
\mu(\hat{x}(\omega) = E(x)) = 1
\]

(1.6)

**Mixing**

A measure preserving transformation is said to be mixing if for every pair of sets \( A \) and \( B \),

\[
\lim_{n \to \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)
\]

(1.7)

**Proposition. Relationship between mixing and ergodicity**

Mixing implies ergodicity.

Pf. Suppose \( B \) is invariant. Then \( \mu(A \cap T^{-n}B) = \mu(A \cap B) \). By mixing \( \mu(A \cap B) = \mu(A)\mu(B) \). If this is true for all invariant \( B \), it is true when \( A = B \), therefore \( \mu(B) = \mu(B)^2 \), so \( \mu(B) \) must equal 0 or 1.

There are many refinements of this general mixing definition; one often encounters forms of mixing conditions in the context of central limit theorems for dependence processes. This is to be expected, since central limit theorems for dependent processes typically require stronger restrictions on the dependence between observations than is needed for laws of large numbers.