1 Prerequisites

Probability space, $\sigma$-field, measurability, Expectation of random variables

2 Summary

Definition and some discussion of Conditional Expectation.

3 Conditional Expectation

This definition of conditional expectation is due to Kolmogorov (1933).

**Definition 1** Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, some sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$, and a random variable $X \in L^1(\mathcal{F})$ (meaning that $X$ is $\mathcal{F}$-measurable and $\mathbb{E}|X| < \infty$), the conditional expectation of $X$ given $\mathcal{G}$ is the (almost surely unique) random variable $\hat{X}$ such that

i. $\hat{X} \in L^1(\mathcal{G})$ that is, $\hat{X}$ is $\mathcal{G}$-measurable; and

ii. $\mathbb{E}(\hat{X}1_G) = \mathbb{E}(X1_G)$ for all $G \in \mathcal{G}$: that is, $\hat{X}$ integrates like $X$ over all $\mathcal{G}$-sets.

Recall that $\mathbb{E}(\hat{X}1_G) = \int_G X d\mathbb{P}$. The random variable $\hat{X}$ is denoted by $\mathbb{E}(X|\mathcal{G})$. This definition is motivated by elementary considerations. Recall the ‘undergraduate’ definition of conditional probability given by Bayes’ Rule

$$\mathbb{P}(A|B) \equiv \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)}$$
for $\mathbb{P}(B) > 0$. Now if $G_1, G_2, \ldots$ is a partition of $\Omega$ into measurable sets, then
\[
\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap G_i) = \sum_i \mathbb{P}(A|G_i)\mathbb{P}(G_i) \tag{1}
\]
This is a form of the law of total probability. $\mathbb{P}(\cdot|B)$ is a new probability measure on $\Omega$ which concentrates on $B$. This generalizes naturally to conditional expectations because $\mathbb{P}(d\omega|B)$ can be used to integrate as any $\mathbb{P}(d\omega)$. Using this fact, Bayes’ rule, and the law of total probability,
\[
\mathbb{E}(X|B) = \int_{\Omega} X(\omega)\mathbb{P}(d\omega|B) = \frac{\int X(\omega)1(\omega \in B)\mathbb{P}(d\omega)}{\mathbb{P}(B)} = \frac{\mathbb{E}(X1_B)}{\mathbb{P}(B)}
\]
Note that equation (1) is obtained by multiplying the identity $1 = \sum_i 1_{G_i}$ on both sides by $1_A$ and taking expectations. This can easily be generalized to the following by multiplying the identity on both sides by $X$ and taking expectations.
\[
\mathbb{E}(X) = \sum_i \mathbb{E}(X1_{G_i}) = \sum_i \mathbb{E}(X|G_i)\mathbb{P}(G_i) \tag{2}
\]
This will be true provided $\mathbb{E}|X| < \infty$. A variation of equation (2) can be obtained as follows. Let $G$ be any union of the $G_i$, i.e., $G \in \mathcal{G}$ where $\mathcal{G} = \sigma(G_i, i = 1, 2, \ldots)$. Again, multiplying the identity $1_G = \sum_{i: G_i \subset G} 1_{G_i}$ on both sides by $X$ and taking expectations, the following is true:
\[
\mathbb{E}(X1_G) = \sum_{i: G_i \subset G} \mathbb{E}(X1_{G_i}) = \sum_{i: G_i \subset G} \mathbb{E}(X|G_i)\mathbb{P}(G_i) \tag{3}
\]
The R.H.S. of equation (3) can interpreted as the expectation of a random variable $\hat{X} = \sum_{i: G_i \subset G} \mathbb{E}(X|G_i)1_{G_i}$, that is, $\hat{X}$ takes the value $\mathbb{E}(X|G_i)$ if $G_i$ occurs. We have just shown that $\mathbb{E}(X1_G) = \mathbb{E}(\hat{X}1_G)$ for every $G$ which is a union of the $G_i$s’ (refer to condition (ii) of definition 1). Since $\hat{X}$ is measurable w.r.t. $\mathcal{G}$, condition (i) of definition 1 is satisfied. $\hat{X} = \mathbb{E}(X|\mathcal{G})$ has been constructed explicitly in the case when $\mathcal{G} = \sigma(G_1, G_2, \ldots)$.

4 References