642:621, Mathematical Finance, Notes to Lecture 4, Fall 2011

I. Partial information and σ-algebras

A. Modeling partial information.

It is common in practice to have partial information about the outcome of a random variable or a random process. For example, the complete outcome of a multi-period market model is the future evolution of market prices from time $t = 0$ until expiry T . At any intermediate time t, $0 < t < T$, a market participant has observed the market evolution from time 0 until t. This is partial information about the entire market path over $[0, T]$ and it will change one's assessment of how the market might behave for times later than t.

The simplest form of partial information is knowledge about whether or not some event has occurred. Let $(\Omega, \mathcal{F}, I\!P)$ be a probability space modeling an experiment with random outcome. Suppose the experiment and Let A be an event in $\mathcal{F}$. Suppose that the experiment has been run and the result is ω. However, suppose we do not get to see what ω is, but are only told whether or not it is in the set A. In this case, we say that we have observed event A. Notice that observing A is equivalent to observing its complement A^c, since knowing whether A has occurred or not is the same as knowing whether A^c has occurred or not.

Another simple form of partial information is to observe the value $X(\omega)$ of a random variable defined on $(\Omega, \mathcal{F}, I\!P)$. In general, the value of ω cannot be deduced from the value of $X(\omega)$ and so again this observation amounts to partial information. For a simple example consider the indicator random variable,

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A; \\ 0, & \text{if } \omega \notin A. \end{cases}$$

If $\mathbf{1}_A(\omega) = 1$, we know A has occurred, but if $\mathbf{1}_A(\omega) = 0$, it has not. So observing $\mathbf{1}_A$ is the same as observing A.

Partial information can be much more complicated. For example, in finance it is important to deal with the partial information provided by observing the evolution of the entire market up to a time t. We will want to use this information in making investment decisions at time t about contingent claims that expire at $T > t$. It is necessary to have a uniform and theoretically convenient way to model partial information. Given the framework of a fixed outcome space and a σ-algebra of events, denoted by $(\Omega, \mathcal{F})$, we shall always model partial information by a σ-algebra $\mathcal{G}$, where $\mathcal{G} \subset \mathcal{F}$. (Since $\mathcal{G} \subset \mathcal{F}$, $\mathcal{G}$ is called a sub-σ-algebra of $\mathcal{F}$.) We shall say that we have observed $\mathcal{G}$ if we know whether or not A has occurred for every $A \in \mathcal{G}$, that is if we know the value of $\mathbf{1}_A(\omega)$ for every every $A \in \mathcal{G}$.

The first two examples fit into this framework as follows. Observation of a single event A is the same as observing the σ-algebra $\{A, A^c, \emptyset, \Omega\}$, because we saw above that if A has been observed, so has A^c. Observing a random variable X is the same as observing $\sigma(X)$, the σ-algebra generated by X. Recall that

$\sigma(X)$ is the collection of all events of the form $\{X \in U\}$, where U is a Borel set,

Certainly, if you know the value of $X(\omega)$, you know whether or not $\{X \in U\}$ has occurred, no matter what U is. Thus if A is any event in $\sigma(X)$ you know whether or not it has occurred. Conversely, suppose you have observed A for every event in $\sigma(X)$. The events $\{X \leq a\}$ are in $\sigma(X)$ for all real numbers a, so you know whether or not $X(\omega) \leq a$ for all a. The smallest value of a for which $X(\omega)$ has occurred is exactly $X(\omega)$ and so you know $X(\omega)$. It may seem that saying you have observed $X(\omega)$ by saying you have observed the σ-algebra $\sigma(X)$ is unnecessarily complicated. But it turns out to be very useful.

Why do we insist that partial information be a σ-algebra? Why not allow any collection of subsets $\mathcal{G}$? This answer is that the σ-algebra property makes intuitive sense and is useful for mathematical theory. We have seen already that observation of an event entails observation of its complement. Similarly, if we know for every n whether event A_n has occurred we will know whether or not $\cup_1^\infty A_n$ and $\cap_1^\infty A_n$ have occurred. Thus, the collection of events that can be observed should be closed under the operations of complements, countable unions and intersections, which is just the σ-algebra property.

A remark about σ-algebras. A basic example of a σ-algebra to keep in mind, especially when trying to understand how σ-algebras are used to model partial information, is the σ-algebra induced be a partition of Ω into events. The construction will be repeated here. Suppose $A_1 \ldots, A_K$ is a disjoint partition of Ω by events. Then the σ-algebra induced by this partition is the collection of all events of the form $\bigcup_{i \in I} A_i$, where I is a subset of $\{1, \ldots, K\}$. Thus, it includes, besides $\emptyset$ and Ω: each of the individual events $A_1, \ldots, A_n$; all unions of pairs, $A_i \cup A_j$ for $i \neq j$, all unions of triples $A_i \cup A_j \cup A_k$ for i, j, k distinct, etc.

Discrete random variables give rise to σ-algebras of this type. For example, let X be a random variable taking values only in the set of distinct numbers $\{c_1, \ldots, c_K\}$. Let $B_i = \{X = c_i\}$, $1 \leq i \leq K$. Then $\sigma(X)$ is the σ-algebra induced by the partition $\{B_1, \ldots, B_K\}$.

B. Sub-σ-algebras and partial information in the multi-period, binomial model

All new concepts will be illustrated in detail for the multi-periood, binomial model, which is recalled next, with some new notation.

(i) Time periods: $t_0 = 0, t_1 = h, t_2 = 2h, \ldots, t_n = nh = T$.

(ii) Outcome space: $\Omega = \{\omega = (\omega_1, \ldots, \omega_n);\ \omega_i \in \{-1, 1\}\text{for each } i\}$.

(ii) $\mathcal{F}$ is the collection of all subsets of Ω.

(iii) For $1 \leq i \leq n$, let $\xi(t_i)(\omega) = \omega_i$; thus $\xi(t_i)(\omega)$ is the market movement in period i; $\xi(t_i)(\omega) = \omega_i = 1$ means an 'up' movement, $\xi(t_i)(\omega) = \omega_i = -1$ means a 'down' movement.

(iv) $N_k(\omega) = \frac{1}{2}\sum_1^k (1 + \xi(t_i)(\omega))$. This counts the total number of 'up' movements in the first k periods.

(v) The price process: $S(t_k)(\omega) = u^{N_k(\omega)} d^{k - N_k(\omega)} S(0)$. Here $0 < d < u$.

Here are some frequently encountered sub-σ-algebras of events in this model:

1. σ-algebras associated to partial market histories:

a) $\sigma(\xi(t_i))$.

This is the σ-algebra modeling observation of the market movement in period i. It is instructive to write $\sigma(\xi(t_1))$ out explicitly. Define

$$A_1 = \{(\omega_1, \ldots, \omega_n); \xi(t_1)(\omega) = \omega_1 = 1\}, \qquad A_{-1} = \{(\omega_1, \ldots, \omega_n); \xi(t_1)(\omega) = \omega_1 = -1\}$$

In words, A_1 is the set of all sequences in Ω that start with 1. Since 1 and -1 are the only two possible values of ξ_1, these two events partition Ω (notice that $A_{-1} = A_1^c$), and

$$\sigma(\xi_1) = \Big\{A_1, A_{-1}, \emptyset, \Omega\Big\}.$$

Each $\sigma(\xi(t_i))$ can be represented in a similar, simple way.

b) Let $\sigma(\xi(t_1), \xi(t_2))$ be the smallest σ-algebra containing both $\sigma(\xi(t_1))$ and $\sigma(\xi(t_2))$. This is the partial information market participants have at the end of the first two periods. It is easy to write this out very explicitly also. Define $A_{(1,1)} = \{(\omega_1, \ldots, \omega_n) : (\omega_1, \omega_2) = (1, 1)\}$, $A_{(1,-1)} = \{(\omega_1, \ldots, \omega_n) : (\omega_1, \omega_2) = (1, -1)\}$, $A_{(-1,1)} = \{(\omega_1, \ldots, \omega_n) : (\omega_1, \omega_2) = (-1, 1)\}$, and $A_{(-1,-1)} = \{(\omega_1, \ldots, \omega_n) : (\omega_1, \omega_2) = (-1, -1)\}$. These events represent all possible outcomes of the first two periods and so form a disjoint partition of Ω. $\sigma(\xi(t_1), \xi(t_2))$ is the σ-algebra induced by this partition:

$$\Big\{A_{(1,1)}, A_{(1,1)}, A_{(1,1)}, A_{(1,1)}, A_1 (= A_{(1,1)} \cup A_{(1,1)}), A_{-1}, \emptyset, \Omega\Big\}.$$

c) Generalizing a) and b), let $\sigma(\xi(t_1), \ldots, \xi(t_k))$ be the smallest σ-algebra of events containing each of $\sigma(\xi(t_1)), \ldots, \sigma(\xi(t_k))$. It represents the information market participants have at the end of k periods from observing the first k market movements $\omega_1, \ldots, \omega_k$. It is the σ-algebra induced by the partition of Ω into the events

$$A_{(\eta_1,\ldots,\eta_k)} = \Big\{(\omega_1, \ldots, \omega_n) : (\omega_1, \ldots, \omega_k) = (\eta_1, \ldots, \eta_k)\Big\},$$

where $(\eta_1, \ldots, \eta_k)$ is any sequence of k 1's and -1's.

d) $\sigma(S(t_k))$. From the formula above for the price, it follows that knowing $S(t_k)(\omega)$ is equivalent to knowing $N_k(\omega)$, the number of times in the first k periods that $\omega_i = 1$. Let $B_k = \Big\{\omega : N_k(\omega) = j\Big\}$, $0 \leq j \leq k$. These events form a disjoint partition of Ω and the σ-algebra they induce is $\sigma(S(t_k))$. Notice that if $S(t_1)(\omega), \ldots, S(t_k)(\omega)$ are all known, then so are $\sigma_1, \ldots, \sigma_k$ and vice-versa. Thus $\sigma(S(t_1), \ldots, S(t_k)) = \sigma(\xi(t_1), \ldots, \xi(t_k))$.

C. Filtrations

Let $(\Omega, \mathcal{F}, I\!P)$ be a probability space. A *filtration* is a sequence $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \ldots$ that is increasing in the sense that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$. Filtrations are used to model how partial information accumulates in time. If, as time increases, one takes in new observations but never forgets old information, the partial information $\mathcal{F}_n$ accumulated by time t_n increases as n increases.

The important filtration associated to the multi-period, binomial market model is defined by $\mathcal{F}(t_k) = \sigma(\xi(t_1), \ldots, \xi(t_k))$, $k = 0, 1, \ldots, n$. Here each $\mathcal{F}(t_k)$ represents the information contained in observation of the first k market movements.

D. Measurability and sub-σ-algebras.

Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. Recall that a real valued function Y on Ω is $\mathcal{G}$-measurable if $\{Y \in U\} \in \mathcal{G}$ for every Borel set U. Of course, a random variable Y on a probability space $(\Omega, \mathcal{F}, I\!P)$ is, by definition, $\mathcal{F}$-measurable. But often we shall want to define random variables that are measurable with respect to sub-σ-algebras of $\mathcal{F}$.

A good case to consider as an introduction is the σ-algebra induced by a disjoint partition $A_1, \ldots, A_K$. Call this $\mathcal{G}$. We claim that Y is $\mathcal{G}$-measurable if and only if Y is a discrete random variable that is constant on each event A_i in the partition and, hence, can be written in the form

$$Y = \sum_1^K b_i \mathbf{1}_{A_i}.$$

(Here $b_1, \ldots, b_k$ are not necessarily distinct.) That this is the case was part of problem 7 of Assignment 3. See the solution to understand this point.

A second, very important case is when $\mathcal{G} = \sigma(X_1, \ldots, X_M)$. We have not defined this in general yet, so we do so now. It is the smallest σ-algebra containing all events of the form $\{X_i \in U\}$, where $1 \leq i \leq K$ and U is a Borel set. As a model for partial information, it is equivalent to observing the values of $X_1(\omega), \ldots, X_M(\omega))$. Examples have already been given above (see b) and c)) for the multi-period, binomial model.

Suppose Y is measurable with respect to $\sigma(X_1, \ldots, X_M)$, By definition, for every scalar a, the event $\{Y \leq a\}$ belongs to $\sigma(X_1, \ldots, X_M)$. If we have observed $X_1(\omega), \ldots, X_m(\omega)$ we know whether or not any event in $\sigma(X_1, \ldots, X_M)$, in particular $\{Y \leq a\}$ for any a, has occurred or not. But if we know this, we know the value of $Y(\omega)$. Therefore, since the value of $Y(\omega)$ can be deduced from the values of $X_1(\omega), \ldots, X_M(\omega)$, $Y(\omega)$ must be a function of $X_1(\omega), \ldots, X_M(\omega)$. This is the content of the following result.

Theorem 1 *Suppose, Y is $\sigma(X_1, \ldots, X_M)$-measurable. Then there is a function h such that $Y(\omega) = h(X_1(\omega), \ldots, X_M(\omega))$. h can be chosen to be a Borel function. (The smallest σ-algebra containing all open sets of $\mathbb{R}^M$ is called the Borel σ-algebra of $\mathbb{R}^M$. A Borel function defined on $\mathbb{R}^M$ is one that is measurable with respect to the Borel σ-algebra.)*

The essential thing to understand about this theorem now is that $Y(\omega)$ can be expressed as $h(X_1(\omega), \ldots, X_M(\omega))$. The fact that h can be chosen to be a Borel function is very important theoretically, but is something we will not worry about.

Example; Multi-period, binomial market. Let Y be $\sigma(\xi(t_1), \ldots, \xi(t_k))$-measurable. Then there is a Borel function h such that

$$Y((\omega_1, \ldots, \omega_n)) = Y(\omega) = h(\xi(t_1)(\omega), \ldots, \xi(t_k)(\omega)) = h(\omega_1, \ldots, \omega_k).$$

In other words, the value of $Y(\omega)$ depends only on the value of $(\omega_1, \ldots, \omega_k)$.

We have seen many random variables of this type. The number $N_k(\omega)$ of 1's in $(\omega_1, \ldots, \omega_k)$, the price $S(t_k)(\omega)$ at time t_k, and the value of a contingent claim, $V(t_k)(\omega)$ at time t_k all have the property of actually depending only on $(\omega_1, \ldots, \omega_k)$ in $\omega = (\omega_1, \ldots, \omega_k, \ldots, \omega_n)$. In previous discussion of the multi-period, binomial model, we have expressed this restricted dependence simply by writing $S(t_k)(\omega_1, \ldots, \omega_k)$, $V(t_k)(\omega_1, \ldots, \omega_k)$, etc., even though we were thinking of these as being all defined on Ω. *From now on, we will express this dependence on partial information by saying these random variables are $\sigma(\xi(t_1), \ldots, \xi(t_k))$-measurable.* This may seem a complicated way to deal with a simple situation, but it will turn out to be essential for formulating continuous-time models.

Further important terminology for the multi-period, binomial model: Define, the filtration $\{\mathcal{F}(t_k)\}_{k \geq 0} = \{\mathcal{F}(t_0), \mathcal{F}(t_1), \ldots, \mathcal{F}(t_n)\}$, as in the previous section, by $\mathcal{F}(t_0) =$

$\{\emptyset, \Omega\}$, and $\mathcal{F}(t_k) = \sigma(\xi(t_1), \ldots, \xi(t_k))$ for $k \geq 1$. A random process $\{X(t_k)\}$ is said to be adapted to $\{\mathcal{F}(t_k)\}_{k\geq 0}$ if $X(t_k)$ is $\mathcal{F}(t_k)$-measurable (hence depends only on $(\omega_1, \ldots, \omega_k)$), for every k. All price processes in the multi-period model, whether they represent the price of the underlying or the price of a contingent claim, are $\{\mathcal{F}(t_k)\}_{k\geq 0}$-adapted.

For each k, let $\triangle(t_k)$ be the portfolio held over the time interval from t_k to t_{k+1}. If we suppose it is not possible to look into the future (a reasonable assumption!), this also may depend only on $\omega_1, \ldots, \omega_k$. Thus, we update a definition made in Lecture 1: a portfolio process is *admissible* if it is $\{\mathcal{F}(t_k)\}_{k\geq 0}$-adapted.

II. Conditional expectation

The examples of this section will often involve the expectation of random variables defined on a probability space on a probability space $(\Omega, \mathcal{F}, I\!P)$, for which Ω is a finite set and $\mathcal{F}$ is the collection of all subsets of Ω. The following formula follows directly from the definition of expected value:

$$E[Y] = \sum_{\omega\in\Omega} Y(\omega) I\!P(\{\omega\}). \tag{1}$$

A. Elementary definitions Let X and Y be two discrete random variables. In elementary probability, the conditional expectation of Y given $X = x$ is defined to be

$$E[Y|X=x] = \sum_y y I\!P(Y=y|X=x) = \sum_y y \frac{I\!P(Y=y, X=x)}{\mathbb{P}(X=x)}. \tag{2}$$

Here the sum is over y in the range of Y, and is defined for x such that $I\!P(X=x) > 0$.

If X and Y are two continuous random variables with joint density $f(x, y)$, the definition is

$$E[Y|X=x] = \int y \frac{f(x,y)}{f_X(x)}\, dy, \tag{3}$$

where f_X is the density of X. This formula defines $E[Y|X| = x]$ on the set where $f_X(x) > 0$. Where $f_X(x) = 0$, we can define it any way we like; it is most convenient just to set the conditional expectation to 0 if $f_X(x) = 0$. We don't have to worry about these values because the probability X takes on a value in the set $\{x : f_X(x) = 0\}$ is 0.

In each case, the conditional expectation is a function of the range of values x, that X can take on.

These definitions and their extension are ultimately too restrictive and too clumsy to handle the types of conditioning required in advanced probability theory and in

finance applications. Instead, a more general notion, that allows one to discuss conditioning on the partial information contained in any σ-algebra of events, is required. The general definition will be developed first in the simpler case of a σ-algebra induced by a partition $A_1, \ldots, A_K$, and then extended to the general case.

B. A preliminary definition.

First, it is necessary to define $E[Y|A]$, the expected value of Y given A has occurred, where A is an event for which $I\!P(A) > 0$.

Definition 1 *If $I\!P(A) > 0$ and if $E[Y\mathbf{1}_A]$ is well-defined, define define*

$$E[Y|A] = \frac{E[Y\mathbf{1}_A]}{\mathbb{P}(A)}. \tag{4}$$

Example 1. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Suppose

$$I\!P(\{\omega_1\}) = 3/8, \quad I\!P(\{\omega_2\}) = 1/8, \quad I\!P(\{\omega_1\}) = 1/4, \quad I\!P(\{\omega_1\}) = 1/4,$$

and $Y(\omega_i) = i$. Let $A = \{\omega_1, \omega_2\}$. Then $I\!P(A) = 1/2$, and

$$E[Y\mathbf{1}_A] = \sum_{i=1}^{4} Y(\omega_i)\mathbf{1}_A(\omega_i)I\!P(\{\omega_i\}) = Y(\omega_1)I\!P(\{\omega_1\}) + Y(\omega_2)I\!P(\{\omega_2\}) = \frac{3}{8} + \frac{2}{8}.$$

Thus, $E[Y|A] = \dfrac{5}{4}$. $\diamond$

Example 2. Let X be a discrete random variable and let $A = \{X = z\}$ where $I\!P(X = z) > 0$. Suppose Y is also discrete. Then, since $\mathbf{1}_A(\omega) = \mathbf{1}_{\{z\}}(X(\omega))$, where $\mathbf{1}_{\{z\}}(x) = 1$ if $x = z$ and $\mathbf{1}_{\{z\}}(x) = 0$, if $x \neq z$, Definition 1 says

$$\begin{aligned} E[Y|\{X = z\}] &= \frac{E[Y\mathbf{1}_{\{z\}}(X)]}{I\!P(X = z)} = \frac{1}{I\!P(X = z)} \sum_{y,x} y\mathbf{1}_{\{z\}}(x)I\!P(X = x, Y = y) \\ &= \frac{1}{I\!P(X = z)} \sum_{y} yI\!P(X = z, Y = y). \end{aligned}$$

This is exactly the formula for $E[Z|X = z]$ given in (2) using the definition of conditional expectation in elementary probability. Thus, Definition 1 is consistent with (2). $\diamond$

Note in example 2 that the formula

$$E[Y|X = z] := E[Y|\{X = z\}] = \frac{E[Y\mathbf{1}_{\{X=z\}}]}{\mathbb{P}(\{X = z\})}, \tag{5}$$

coming from Definition 1, makes sense for any Y, whether discrete or not, as long as $E[Y\mathbf{1}_{\{X=z\}}]$ is well-defined.

C. Conditioning on a σ-algebra generated by a partition.

Let $\{A_1, \ldots, A_K\}$ be a partition of Ω into disjoint events. Let $\mathcal{G}$ be the σ-algebra induced by this partition.

Definition 2 *Let Y be a random variable for which $E[Y]$ is defined and finite. Then the conditional expectation of Y given $\mathcal{G}$ is*

$$E[Y|\mathcal{G}](\omega) := \sum_{i=1}^{K} E[Y|A_i]\mathbf{1}_{A_i}(\omega) \tag{6}$$

It is essential to note that this conditional expectation is a *random variable*. For each ω, its value is the conditional expectation of Y given the event A_i that actually occurs. In the simplest case, when $\mathcal{G} = \{A, A^c, \emptyset, \Omega\}$,

$$E[Y|\mathcal{G}](\omega) = E[Y|A]\mathbf{1}_A(\omega) + E[Y|A^c]\mathbf{1}_{A^c}(\omega),$$

returns the value $E[Y|A]$ if A occurs, and the value $E[Y|A^c]$ if it does not.

It is standard practice to omit the dependence on ω and write simply $E[Y|\mathcal{G}]$ when using conditional expectations.

Example 3. Consider the set-up of Example 1 and let $\mathcal{G} = \{A, A^c, \emptyset, \Omega\}$. Here $A^c = \{\omega_3, \omega_4\}$. By a calculation similar to that of Example 2, $E[Y|A^c] = 7/2$. Thus

$$E[Y|\mathcal{G}](\omega) = \frac{5}{4}\mathbf{1}_A(\omega) + \frac{7}{2}\mathbf{1}_{A^c}(\omega) = \frac{5}{4}\mathbf{1}_{\{\omega_1,\omega_2\}}(\omega) + \frac{7}{2}\mathbf{1}_{\{\omega_3,\omega_4\}}(\omega).$$

In keeping with the convention of suppressing the dependence on ω, this would usually be written, $\frac{5}{4}\mathbf{1}_{\{\omega_1,\omega_2\}} + \frac{7}{2}\mathbf{1}_{\{\omega_3,\omega_4\}}$. $\diamond$

D. Conditioning on a discrete random variable. Let X be a discrete random variable taking on values in the finite set $\{c_1, \ldots, c_K\}$. Then $\sigma(X)$ is the σ-algebra induced by the partition $\{X = c_1\}, \ldots, \{X = c_K\}$.

Definition 3 *In this case, $E[Y|X] := E[Y|\sigma(X)]$.*

By applying the original definition (6) and (5)

$$E[Y|X](\omega) = \sum_{i=1}^{K} E[Y|\{X = c_i\}]\mathbf{1}_{\{X=c_i\}}(\omega) = \sum_{i=1}^{K} E[Y|X = c_i]\mathbf{1}_{\{X=c_i\}}(\omega).$$

So, when $X(\omega) = c_i$, $E[Y|X](\omega) = E[Y|X=c_i]$. It is instructive to write this as

$$E[Y|X](\omega) = E[Y|X=x]\Big|_{x=X(\omega)}, \tag{7}$$

This means we can use the formula (2) from elementary probability to compute conditional expectations when conditioning on a random variable.

Example 4. Again consider the set-up of Example 1 and define a second random variable by $X(\omega_1) = X(\omega_2) = 0$ and $X(\omega_3) = X(\omega_4) = 1$. Then $A = \{X=0\}$ and $A^c = \{X=0\}$ and so $\sigma(X) = \mathcal{G} = \{A, A^c, \emptyset, \Omega\}$. Thus $E[Y|X]$ is exactly $E[Y|\mathcal{G}]$ as computed in Example 3. However, it is usual to express it as follows, using (7): since $E[Y|X=0] = E[Y|A] = 5/4$ and $E[Y|X=1] = E[Y|A^c] = 7/2$,

$$E[Y|X] = (5/4)\mathbf{1}_{\{0\}}(X) + (7/2)\mathbf{1}_{\{1\}}(X).$$

This displays the answer clearly as a random variable, since X is a random variable. $\diamond$

Note in Example 4 that $E[Y|X]$ is a function of X. This is a general fact, always true, as will be explained in the next paragraph

E. An alternate characterization of $E[Y|\mathcal{G}]$.

Theorem 2 *Let $\mathcal{G}$ be the σ-algebra induced by a finite disjoint partition of Ω. Suppose $E[|Y|] < \infty$. Then $Z = E[Y|\mathcal{G}]$ is the unique random variable satisfying:*

- *Z is $\mathcal{G}$-measurable;* (8)
- *For every $A \in \mathcal{G}$, $E[\mathbf{1}_A Z] = E[\mathbf{1}_A Y]$.* (9)

In this theorem, 'unique' means 'unique up to sets of probability zero. In other word if $\bar{Z}$ is another random variable satisfying (8) and (9), then $I\!P(Z = E[Y|\mathcal{G}]) = 1$.

Equation (9) with Z replaced by $E[Y|\mathcal{G}]$ is

$$E\Big[\mathbf{1}_A E[Y|\mathcal{G}]\Big] = E[\mathbf{1}_A Y] \text{ for every } A \in \mathcal{G}. \tag{10}$$

Memorize this fact!

Remark on measurability. From this theorem, it follows in general that if X is a discrete random variable, $E[Y|X] = E[X|\sigma(X)]$ is $\sigma(X)$-measurable. By Theorem 1, any $\sigma(X)$-measurable function can be expressed as a function of X. Thus, there is always a function h such that $E[Y|X] = h(X)$. We saw an instance of this already in Example 4.

Here is a proof of the Theorem. Let $\mathcal{G}$ be induced by the disjoint partition $\{A_1, \ldots, A_K\}$. Suppose Z satisfies (8). Then we know that Z can be written in the form $Z = \sum_1^K d_i \mathbf{1}_{A_i}$. Now suppose that Z also satisfies (9). Then for any j, since $A_j \in \mathcal{G}$,

$$E[\mathbf{1}_{A_j} Z] = E[\mathbf{1}_{A_j} Y].$$

But since the sets of the partition are disjoint, $\mathbf{1}_{A_i}\mathbf{1}_{A_j} \equiv 0$ if $i \neq j$ and $\mathbf{1}_{A_j}\mathbf{1}_{A_j} = \mathbf{1}_{A_j}$. Thus

$$\mathbf{1}_{A_j} Z = \mathbf{1}_{A_j}[\sum_{i=1}^{K} d_i \mathbf{1}_{A_i}] = \sum_{i=1}^{L} d_i \mathbf{1}_{A_j}\mathbf{1}_{A_i} = d_j \mathbf{1}_{A_j}.$$

By plugging into the previous equation it follows that

$$E[\mathbf{1}_{A_j} Y] = E[d_j \mathbf{1}_{A_j}] = d_j I\!P(A_j) \quad \text{or} \quad d_j = \frac{E[\mathbf{1}_{A_j} Y]}{I\!P(A_j)}.$$

Thus

$$Z = \sum_{i=1}^{K} \frac{E[\mathbf{1}_{A_i} Y]}{I\!P(A_i)} \mathbf{1}_{A_i} = E[Y|\mathcal{G}].$$

Conversely,

$$E[Y|\mathcal{G}] = \sum_{i=1}^{K} \frac{E[\mathbf{1}_{A_i} Y]}{I\!P(A_i)} \mathbf{1}_{A_i}$$

is certainly $\mathcal{G}$-measurable and if $U = \bigcup j \in J A_j$ is in $\mathcal{G}$,

$$\mathbf{1}_U = \sum_{j \in J} \mathbf{1}_{A_j}$$

,

$$\mathbf{1}_U E[Y|\mathcal{G}] = \sum_{j \in J} \frac{E[\mathbf{1}_{A_j} Y]}{I\!P(A_j)} \mathbf{1}_{A_j},$$

and so

$$\begin{aligned} E[\mathbf{1}_U E[Y|\mathcal{G}]] &= E\Big[\sum_{j \in J} \frac{E[\mathbf{1}_{A_j} Y]}{I\!P(A_j)} \mathbf{1}_{A_j}\Big] \\ &= \sum_{j \in J} \frac{E[\mathbf{1}_{A_j} Y]}{I\!P(A_j)} I\!P(A_j) \\ &= E\Big[(\sum_{j \in J} \mathbf{1}_{A_j}) Y\Big] = E[\mathbf{1}_U Y]. \end{aligned}$$

Thus $E[Y|\mathcal{G}]$ does satisfy conditions (8) and (9).

F. Conditioning on a general σ-algebra.

The Definition 2 will not work for general σ-algebras. To see the problem, suppose that X is a continuous random variable. For every x, $\{X=x\}$ is an event of probability 0. How then can one make sense of conditioning on such an event?

They key idea is to use the characterization of conditional expectation found in Theorem 2. This definition makes no explicit reference to an object such as $E[Y|A]$ requiring to divide by $\mathbb{P}(A)$.

Definition 4 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let Y be a random variable such that $E[|Y|] < \infty$. Let Z be a random variable satisfying*

- *Z is $\mathcal{G}$-measurable;* (11)
- *For every $A \in \mathcal{G}$, $E[\mathbf{1}_A Z] = E[\mathbf{1}_A Y]$.* (12)

Then we say Z is the conditional expectation of Y given $\mathcal{G}$ and denote it by $E[Y|\mathcal{G}]$.

The following result addresses whether such a random variable Z exists.

Theorem 3 *If $E[|Y|] < \infty$, then $E[Y|\mathcal{G}]$ satisfying (11) and 12) exists. It is unique in the sense that if Z and Z' both satisfy (11) and 12), then $\mathbb{P}(Z=Z') = 1$.*

Because of this theorem we can always write down $E[Y|\mathcal{G}]$ in good conscience, once we have checked $E[|Y|] < \infty$. Unfortunately this theorem does not tell us how to calculate the conditional expectation. We have to come up with a formula that we think is correct and then check it satisfies the two conditions (11) and (12).

As before, we define $E[Y|X]$ to be $E[Y|\sigma(X)]$

Example 5. The simplest example occurs when Y is independent of $\mathcal{G}$. Then, $E[Y|\mathcal{G}] = E[Y]$, which makes sense, because knowledge of $\mathcal{G}$ does not affect our assessment of probabilities concerning Y if they are independent. To see this, it is is necessary to check (11) and (12). But $E[Y]$ is just a constant, so it is certainly $\mathcal{G}$-measurable ($\{\omega;\, E[Y] \in U\} = \Omega$ if U contains the constant $E[Y]$ and equals $\emptyset$ if not). Finally, for any $A \in \mathcal{G}$, by independence $E[\mathbf{1}_A Y] = E[\mathbf{1}_A]E[Y] = E[\mathbf{1}_A E[Y]]$, which verifies (12).

Example 6. Let X and Y be jointly continuous random variables with joint density. Let $h(x) = E[Y|X=x]$, as defined in (3):

$$E[Y|X=x] = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_X(x)}\, dy.$$

Claim: In this situation $E[Y|X] = h(X)$.

We will give a partial proof of this claim, only omitting some technical issues. In what follows, if $-\infty < a < b < \infty$, $\mathbf{1}_{(a,b)}(x)$ denotes the indicator function of (a, b); it equals 1 if $a < x < b$ and equals 0 otherwise. This proof is also treated in problem 2.10 in Shreve, volume II.

We need to check that $h(X)$ as defined satisfies (11) and (12). The first condition is easy. Since $h(X)$ is a function of X, it is $\sigma(X)$ measurable.

We need to check (12) for every $A \in \sigma(X)$. This means checking the condition for every event of the form $\{X \in U\}$, where U is a Borel subset of the real line. Actually it suffices to check only U which are intervals, say $U = (a, b)$; we omit the explanation of why this is enough.

First observe that if $A = \{X \in (a, b)\}$, then $\mathbf{1}_A = \mathbf{1}_{(a,b)}(X)$. Thus, to verify (12) for such a set A, it is necessary to show

$$E[\mathbf{1}_{(a,b)}(X)h(X)] = E[\mathbf{1}_{(a,b)}(X)Y]. \tag{13}$$

The right-hand side is

$$E[\mathbf{1}_{(a,b)}(X)Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{(a,b)}(x)yf(x,y)\,dydx.$$

Using the definition of h, the left-hand side is

$$\begin{aligned} E[\mathbf{1}_{(a,b)}(X)h(X)] &= \int_{-\infty}^{\infty} \mathbf{1}_{(a,b)}(x)h(x)f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \mathbf{1}_{(a,b)}(x)\Big[\int_{-\infty}^{\infty} y\frac{f(x,y)}{f_X(x)}\,dy\Big]f_X(x)\,dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{(a,b)}(x)yf(x,y)\,dydx, \end{aligned}$$

since the $f_X(x)$ terms cancel. Hence the two sides of the (13) are the same, which proves the identity. $\diamond$

It is important to read Theorem 2.3.2 in Shreve, Volume II. This lists important properties of the conditional distribution that are all proved using the two defining conditions (11) and (12).

G. Conditioning on multiple random variables. Let $X_1, \ldots, X_M$ be random variables. Let $\sigma(X_1, \ldots, X_M)$ be the smallest σ-algebra containing all events of the form $\{X_i \leq b\}$, where b is any real number and i is any index, $1 \leq i \leq m$. This is called the σ-algebra generated by $X_1, \ldots, X_M$.

Definition 5 $E[Z|X_1,\ldots,X_M] := E[Z|\sigma(X_1,\ldots,x_M)]$.

H. An important special case.

This is the case treated in Lemma 2.3.4 of Shreve, Volume II. It says that if $Y_1,\ldots,Y_L$ are independent of every event in $\mathcal{G}$, if $X_1,\ldots,X_K$ are $\mathcal{G}$-measurable (and hence independent of $Y_1,\ldots,Y_K$), and if $f(x_1,\ldots,x_K,y_1,\ldots,y_L)$ is a function (**it does not represent a density here!**), then

$$E[f(X_1,\ldots,X_K,Y_1,\ldots,Y_L)\Big|\mathcal{G}] = g(X_1,\ldots,X_K), \tag{14}$$

where $g(x_1,\ldots,x_k) = E\Big[f(x_1,\ldots,x_k,Y_1,\ldots,Y_L)\Big]$.

This makes sense, because, by independence, the joint distribution of $Y_1,\ldots,Y_L$ is not affected by knowing the values of $X_1,\ldots,X_K$.

Identity (14) will be applied over and over in Math Finance I and II.

Example 7. Suppose $S(t_1) = X_1S(0)$ and $S(t_2) = X_2S(t_1)$, where X_1 and X_2 are independent and uniformly distributed on $(0,1)$, and $t_2 > t_1$. Here it is supposed that $S(0)$ is not random and is known. Observing $S(t_1)$ is equivalent to observing X_1. What is the conditional expectation of $S(t_2)$ given this information? That is, what is

$$E[S(t_2)|X_1] = E[X_2X_1S(0)|X_1]?$$

Since $g(x_1) = E[X_2x_1S(0)] = x_1S(0)\cdot(1/2)$, it follows from (14) that $E[S(t_2)|X_1] = (1/2)X_1S(0)$. This could also have been derived using the "taking out what is known" principle described in Theorem 2.3.2 of Shreve, Volume II.

More interesting financial examples appear in Assignment 4.