

# BACKGROUND TO THE BINOMIAL MODEL

MICHAEL J. SHARPE,  
MATHEMATICS DEPARTMENT, UCSD

## 1. INTRODUCTION

These notes provide an expansion of the background ideas used in Williams' notes on the Binomial Model for stock option pricing. The ideas discussed here are special cases of more general, somewhat abstract, concepts that are normally treated in more advanced courses.

## 2. A BINARY MODEL

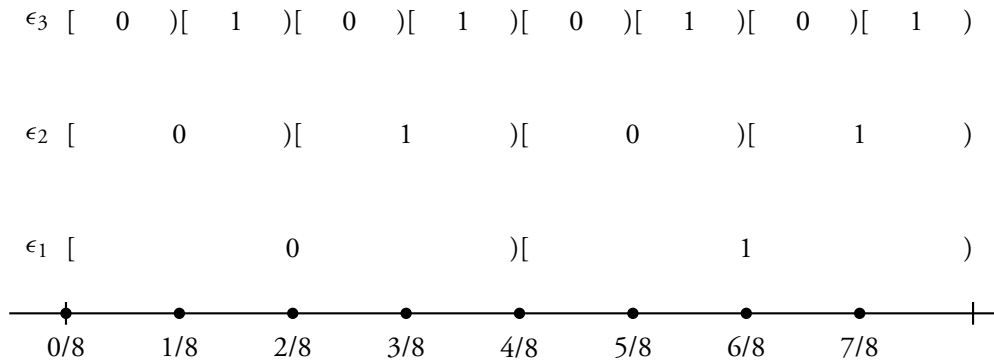
We are going to be discussing sequences of  $T$  binomial trials. One way to think of each such sequence is as a  $T$ -tuple  $(\epsilon_1, \dots, \epsilon_T)$ , where each  $\epsilon_j$  is either 0 or 1. One might take  $\Omega$  to be the set of all such  $T$ -tuples, but it will be conceptually easier (especially for drawing pictures) to define  $\Omega$  in an equivalent way, as the collection of all real numbers  $x$  having a binary expansion of the form

$$(2.1) \quad x = \frac{\epsilon_1}{2} + \frac{\epsilon_2}{2^2} + \dots + \frac{\epsilon_T}{2^T}; \quad \epsilon_j = 0 \text{ or } 1 \text{ for } j = 1, \dots, T.$$

Given  $x \in \Omega$ , it is clear that  $x \geq 0$  and  $x \leq \sum_1^T 2^{-j} = 1 - 2^{-T}$  (summing a geometric series) and, since all summands are integer multiples of  $2^{-T}$ , so is  $x$ . Conversely, if we start with  $x = k2^{-T}$ , where  $k$  is an integer between 0 and  $2^T - 1$ , we may give a binary expansion of  $k$  using the following procedure:

- if  $k < 2^{T-1}$ , let  $\epsilon_1 = 0$ , else let  $\epsilon_1 = 1$ , and let  $k_1 := k - \epsilon_1 2^{T-1}$ , so that  $0 \leq k_1 < 2^{T-1}$ ;
- repeat the first procedure but using  $k_1$  and  $T - 1$  in place of  $k$  and  $T$ , letting  $\epsilon_2 = 0$  if  $k_1 < 2^{T-2}$ , else  $\epsilon_2 = 1$ , setting  $k_2 = k_1 - \epsilon_2 2^{T-2}$ ;
- continue this process  $T$  times, at which point we have  $k = \epsilon_1 2^{T-1} + \epsilon_2 2^{T-2} + \dots + \epsilon_T 2^0$ .

Now we have  $x = k2^{-T}$ , which amounts to the expansion (2.1). As there are exactly  $2^T$   $T$ -tuples of the form  $(\epsilon_1, \dots, \epsilon_T)$ , this shows also that the expansion (2.1) is unique. (That is,  $x$  uniquely determines the coefficients  $\epsilon_1, \dots, \epsilon_T$ .) From now on, we picture  $\Omega$  as a subset of the unit interval  $[0,1)$ , consisting of evenly spaced points, all multiples of  $2^{-T}$ . Here is a picture of  $\Omega$  in the special case  $T = 3$ , indicating the associated values of  $\epsilon_1, \epsilon_2$  and  $\epsilon_3$ .



As the picture illustrates, the points in  $\Omega$  with  $\epsilon_1 = 1$  are precisely those within  $[2^{-1}, 1)$ , while those with  $\epsilon_1 = 0$  are those within  $[0, 2^{-1})$ . Similarly, those with  $\epsilon_2 = 0$  comprise the left half of the intervals  $[2^{-1}, 1)$  and  $[0, 2^{-1})$ , and so on.

**2.1. Partitions and  $\sigma$ -algebras:** The following discussion applies to any space  $\Omega$ , not necessarily finite. By a (finite) partition  $\mathcal{A}$  of  $\Omega$ , we mean a decomposition of  $\Omega$  into a finite number of disjoint subsets, so:  $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$  where  $A_1 \cup \dots \cup A_n = \Omega$  and  $A_i \cap A_j = \emptyset$  if  $i \neq j$ . The sets  $A_j$  are called the **atoms** of the partition  $\mathcal{A}$ . Along with a finite partition  $\mathcal{A}$  of  $\Omega$ , we consider the  $\sigma$ -algebra  $\mathcal{F}$  generated by  $\mathcal{A}$ , defined by

$$(2.2) \quad \mathcal{F} := \{G \subset \Omega : G \text{ is a union of atoms of } \mathcal{A}\}.$$

If  $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$ , then  $\mathcal{F}$  consists of the following sets:

- $\emptyset$ : the union of an empty class of atoms;
- $A_1, \dots, A_n$ : the unions of one atom;
- $A_1 \cup A_2, \dots, A_{n-1} \cup A_n$ : the unions of two atoms;
- ...
- $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$ : the union of all  $n$  atoms.

In more advanced work,  $\sigma$ -algebras are harder to define and work with than in this discrete model. Here are the basic facts we need.

**Proposition 2.3.** *Let  $\mathcal{F}$  be the  $\sigma$ -algebra generated by the finite partition  $\mathcal{A} = \{A_1, \dots, A_n\}$ . (i) Let  $G \in \mathcal{F}$ . Then  $G^c \in \mathcal{F}$ . (ii) Let  $G_1, \dots, G_n \in \mathcal{F}$ . Then  $\cup_1^n G_i \in \mathcal{F}$  and  $\cap_1^n G_i \in \mathcal{F}$ .*

*Proof.* (i) Suppose  $G = A_{j_1} \cup \dots \cup A_{j_k}$ , where each  $A_{j_i} \in \mathcal{A}$ . Then  $G^c \in \mathcal{F}$  also, because  $G^c$  is the union of all the remaining atoms in  $\mathcal{A}$ . (ii) That the union of sets in  $\mathcal{F}$  is also in  $\mathcal{F}$  is clear since, if each of  $G_1$  and  $G_2$  is a union of atoms, then so is  $G_1 \cup G_2$ . In the case of intersection  $G_1 \cap G_2$ , we just notice that because the atoms are disjoint,  $G_1 \cap G_2$  is the union of the atoms that belong to both  $G_1$  and  $G_2$ , and is therefore in  $\mathcal{F}$ .  $\square$

**Definition 2.4.** *Let  $\mathcal{F}$  denote the  $\sigma$ -algebra generated by a finite partition  $\mathcal{A}$  of  $\Omega$ . A function  $X$  on  $\Omega$  is  $\mathcal{F}$ -measurable provided  $X$  is constant on each atom of  $\mathcal{A}$ .*

**Proposition 2.5.** *With  $\mathcal{F}$  as in the definition, let  $X$  be a function of  $\Omega$  taking only a finite number of distinct values  $x_1, \dots, x_k$ . Then  $X$  is  $\mathcal{F}$ -measurable provided  $\{\omega : X(\omega) = x_j\} \in \mathcal{F}$  for  $j = 1, \dots, k$ .*

*Proof.* Let  $X$  be  $\mathcal{F}$ -measurable, and let  $y_1, \dots, y_n$  denote the constant values it takes on the atoms  $A_1, \dots, A_n$ . Fix  $x$  to be one of the  $x_j$ , so that  $x$  is among the  $y_i$ , say  $y_{i_1} = \dots = y_{i_m} = x$ , with all the other  $y_i$  different from  $x$ . Then  $\{\omega : X(\omega) = x\} = A_{i_1} \cup \dots \cup A_{i_m} \in \mathcal{F}$ . Conversely, if  $\{\omega : X(\omega) = x_j\} \in \mathcal{F}$  for  $j = 1, \dots, k$ , then  $X$  must be constant on each atom, otherwise there would exist two distinct values, say  $x_i \neq x_j$ , and an atom  $A$  on which  $X$  takes both these values, and this would entail that  $\{\omega : X(\omega) = x_j\}$  contain part but not all of  $A$ .  $\square$

The simplest  $\mathcal{F}$ -measurable functions are the indicator functions, defined as follows.

**Definition 2.6.** *Let  $G \subset \Omega$ . By the indicator of  $G$ , we mean the function  $1_G$  determined by*

$$1_G(\omega) = \begin{cases} 1 & \text{if } \omega \in G; \\ 0 & \text{if } \omega \in G^c. \end{cases}$$

The function  $1_G$  indicates  $G$  in the sense that  $\{\omega : 1_G(\omega) = 1\} = G$ . The following simple but important point is used repeatedly.

**Proposition 2.7.** *Let  $G \subset \Omega$ . Then  $G \in \mathcal{F}$  if and only if  $1_G$  is  $\mathcal{F}$ -measurable.*

*Proof.* Suppose  $G \in \mathcal{F}$ . In order to prove that  $1_G$  is  $\mathcal{F}$ -measurable, we need to prove that  $\{\omega : 1_G(\omega) = x\} \in \mathcal{F}$  for  $x = 0, 1$ . These respective cases amount to  $G^c$  and  $G$ , which are both in  $\mathcal{F}$ . Suppose, conversely, that  $1_G$  is  $\mathcal{F}$ -measurable. Then  $G = \{\omega : 1_G(\omega) = 1\} \in \mathcal{F}$ .  $\square$

**Proposition 2.8.** *With  $\mathcal{F}$  and  $\mathcal{A}$  as above, let  $X_1, \dots, X_m$  be  $\mathcal{F}$ -measurable, and let  $f$  denote an arbitrary function of  $m$  variables defined on the range of  $(X_1, \dots, X_m)$ . Then  $Y := f(X_1, \dots, X_m)$  is also  $\mathcal{F}$ -measurable.*

*Proof.* It is enough to verify that  $Y$  is constant on every atom of  $\mathcal{A}$ , a fact which is an immediate consequence of the fact that each  $X_j$  is constant on each atom of  $\mathcal{A}$ .  $\square$

**Corollary 2.9.** *Let  $c$  be a scalar, and let  $X_1, \dots, X_m$  be  $\mathcal{F}$ -measurable. Then so are, for example,  $cX_1, X_1 + \dots + X_m, X_1X_2, X_1/X_2$  (provided it is defined). Because a union of sets in  $\mathcal{F}_t$  is also in  $\mathcal{F}_t$ , we see that if  $X$  is  $\mathcal{F}_t$ -measurable, then so is  $\{\omega : X(\omega) \leq x\}$ , for the latter may be written as  $\cup_{y \leq x} \{\omega : X(\omega) = y\}$ , where the union is taken over the finitely many  $y$  in the range of  $X$ .*

**Corollary 2.10.** *Let  $X$  be  $\mathcal{F}$ -measurable. Then  $\{\omega : X(\omega) \leq y\} \in \mathcal{F}$  for every  $y$ .*

*Proof.* Write  $f(x) := 1_{(-\infty, y]}(x)$ . Then  $f(X)$  is  $\mathcal{F}$ -measurable, and it is equal to 1 precisely where  $X(\omega) \leq y$ .  $\square$

**Proposition 2.11.** *Let  $X$  be  $\mathcal{F}$ -measurable. Then we may express  $X$  as a linear combination of indicators of atoms of  $\mathcal{A}$ .*

*Proof.* Let  $X$  take the constant value  $x_k$  on the atom  $A_k$ . Then  $X = \sum_1^n x_k 1_{A_k}$ .  $\square$

**2.2. The filtration of  $\Omega$ :** We return now to the special  $\Omega$  described in the first section, amounting to multiples of  $2^{-T}$  in  $[0, 1)$ . Given any integer  $t$  with  $0 \leq t \leq T$ , define a partition  $\mathcal{A}_t$  of  $\Omega$  by

$$(2.12) \quad \mathcal{A}_t := \{\Omega \cap [0, \frac{1}{2^t}), \Omega \cap [\frac{1}{2^t}, \frac{2}{2^t}), \dots, \Omega \cap [\frac{2^t - 1}{2^t}, 1)\}.$$

For simplicity, we shall omit the  $\Omega$  when we describe the elements of  $\mathcal{A}_t$ . The partition  $\mathcal{A}_t$  generates a  $\sigma$ -algebra  $\mathcal{F}_t$ , in the manner described in the paragraphs above.

There are two extreme cases to be considered. Since  $\mathcal{A}_0$  has just one atom, namely  $\Omega$  itself,  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . On the other hand,  $\mathcal{A}_T$  consists of sets each containing a single point of  $\Omega$ , so that  $\mathcal{F}_T$  consists of all subsets of  $\Omega$ . Because the partitions  $\mathcal{A}_t$  refine as  $t$  increases (that is, each atom  $A \in \mathcal{A}_t$  is a union of atoms of  $\mathcal{A}_{t+1}$ ),

$$(2.13) \quad \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_T.$$

(A family of  $\sigma$ -algebras satisfying (2.13) is called a **filtration** of  $\Omega$ . One should think of sets in  $\mathcal{F}_t$  as representing information which may potentially be determined by time  $t$ .)

For  $\omega \in \Omega$  with  $\omega = \sum_{j=1}^T \epsilon_j 2^{-j}$ , define  $D_j(\omega) = \epsilon_j$ , so that each  $D_j$  is a random variable on  $\Omega$  with possible values 0 and 1. It is clear that  $D_t$  is  $\mathcal{F}_t$ -measurable for  $t = 1, \dots, T$ , as  $D_t$  is constant on each atom of  $\mathcal{A}_t$ . Much more is in fact true.

**Proposition 2.14.** *Let  $Y$  be  $\mathcal{F}_t$ -measurable,  $1 \leq t \leq T$ . Then there is a function  $f$  of  $t$  variables such that  $Y = f(D_1, \dots, D_t)$ .*

*Proof.* By hypothesis,  $Y$  is constant on each atom of  $\mathcal{A}_t$ , namely the intervals of the form  $[k2^{-t}, (k+1)2^{-t})$ . It follows that  $Y$  is completely determined by its values at the points  $k2^{-t}$ ,  $0 \leq k < 2^t$ . The latter points are expressible in terms of the first  $t$  binary digits  $\epsilon_1, \dots, \epsilon_t$ , which is to say that  $Y$  is completely determined by the values of the functions  $D_1, \dots, D_t$ . In other words,  $Y$  is a function of those variables.  $\square$

Note the following special case (from which an alternate proof could have been constructed, making use of Proposition 2.11. If  $Y$  is the indicator of  $[k2^{-t}, (k+1)2^{-t})$ , write  $k2^{-t} = \frac{\epsilon_1}{2} + \dots + \frac{\epsilon_t}{2^t}$  and set  $f(z_1, \dots, z_t) := 1_{\{(\epsilon_1, \dots, \epsilon_t)\}}(z_1, \dots, z_t)$ . Then  $Y = f(D_1, \dots, D_t)$ .

## 3. CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION

The definitions in this section may be made in considerably greater generality than here, but at considerable mathematical cost. To keep the mathematical problems to a minimum, we shall assume that  $\Omega$  is finite, so that all  $\sigma$ -algebras are generated by some partition, and the results of the previous section may be applied. Unless otherwise specified,  $\mathcal{F}$  denotes a  $\sigma$ -algebra generated by a partition  $\mathcal{A} = \{A_1, \dots, A_n\}$  of  $\Omega$ .

Let  $A$  be an event. For any event  $G$ , the elementary conditional probability of  $G$ , given  $A$ , is defined by:

$$P(G | A) := \begin{cases} \frac{P(G \cap A)}{P(A)} & \text{if } P(A) > 0 \\ 0 & \text{if } P(A) = 0. \end{cases}$$

(The value 0 is completely arbitrary. It doesn't matter at all how one defines it when  $P(A) = 0$ .) The probabilistic meaning is the relative frequency of occurrence of  $G$  on those trials where  $A$  occurs. In similar manner, the elementary conditional expectation of a random variable  $X$ , given  $A$ , is defined by

$$E(X | A) := \begin{cases} \frac{E(X1_A)}{P(A)} & \text{if } P(A) > 0 \\ 0 & \text{if } P(A) = 0. \end{cases}$$

We are now going to define a more general conditional probability and conditional expectation, related to the elementary cases, but having cleaner mathematical properties.

$$(3.1) \quad P\{G | \mathcal{F}\} = \sum_{k=1}^n P\{G | A_k\}1_{A_k}.$$

In other words,  $P\{G | \mathcal{F}\}$  is an  $\mathcal{F}$ -measurable random variable. Similarly,

$$(3.2) \quad E\{X | \mathcal{F}\} = \sum_{k=1}^n E(X | A_k)1_{A_k}.$$

The two are related by

$$P\{G | \mathcal{F}\} = E\{1_G | \mathcal{F}\},$$

so that properties of conditional expectation follow easily from properties of conditional expectation.

Here are a couple of special cases of conditional expectation. First, let's assume that  $\mathcal{A} = \{\Omega\}$ . Then  $E\{X | \mathcal{F}\} = E(X | \Omega)1_\Omega$ , which is a constant random variable with value  $E(X | \Omega) = E(X1_\Omega)/P(\Omega) = E(X)$ . In other words, in this special case, conditional expectation reduces to ordinary expectation. At the opposite end of the spectrum, suppose  $\mathcal{A} = \{\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_N\}\}$ , so that the atoms of the partition are singletons. Then  $\mathcal{F}$  consists of all subsets of  $\Omega$ , and from the definition we find

$$\begin{aligned} E\{X | \mathcal{F}\} &= \sum_{i=1}^N E(X1_{\{\omega_i\}})/P(\{\omega_i\})1_{\{\omega_i\}} \\ &= \sum_{i=1}^N X(\omega_i)P(\{\omega_i\})/P(\{\omega_i\})1_{\{\omega_i\}} \\ &= X \quad \text{except possibly on a set of probability 0.} \end{aligned}$$

In other words, in this particular case, the conditional expectation just spits back  $X$  itself.

The following is a useful restatement of the definition of conditional expectation.

**Proposition 3.3.**  $Y := E\{X | \mathcal{F}\}$  is the unique  $\mathcal{F}$ -measurable random variable having the property

$$(3.4) \quad E\{Y1_A\} = E\{X1_A\} \quad \text{for every atom } A \text{ in the partition generating } \mathcal{F}.$$



## 4. BINOMIAL MODEL

Fix now  $p$  with  $0 \leq p \leq 1$ . The binomial probability measure  $P$  (with parameter  $p$  on  $\Omega$  is defined by attaching weight  $p^{\epsilon_1+\dots+\epsilon_T} (1-p)^{T-(\epsilon_1+\dots+\epsilon_T)}$  to the point  $x = \frac{\epsilon_1}{2} + \frac{\epsilon_2}{2^2} + \dots + \frac{\epsilon_T}{2^T} \in \Omega$ . Another way to put this is

$$(4.1) \quad P\{D_1 = \epsilon_1, \dots, D_T = \epsilon_T\} = p^{\epsilon_1+\dots+\epsilon_T} (1-p)^{T-(\epsilon_1+\dots+\epsilon_T)}.$$

As a consequence, we see that

$$\begin{aligned} & P\{D_1 = \epsilon_1, \dots, D_{T-1} = \epsilon_{T-1}\} \\ &= P\{D_1 = \epsilon_1, \dots, D_{T-1} = \epsilon_{T-1}, D_T = 1\} + P\{D_1 = \epsilon_1, \dots, D_{T-1} = \epsilon_{T-1}, D_T = 0\} \\ &= p^{\epsilon_1+\dots+\epsilon_{T-1}+1} (1-p)^{T-(\epsilon_1+\dots+\epsilon_{T-1}+1)} + p^{\epsilon_1+\dots+\epsilon_{T-1}+0} (1-p)^{T-(\epsilon_1+\dots+\epsilon_{T-1}+0)} \\ &= p^{\epsilon_1+\dots+\epsilon_{T-1}} (1-p)^{T-1-(\epsilon_1+\dots+\epsilon_{T-1})} (p + (1-p)) \\ &= p^{\epsilon_1+\dots+\epsilon_{T-1}} (1-p)^{T-1-(\epsilon_1+\dots+\epsilon_{T-1})}. \end{aligned}$$

As this formula is identical to (4.1) except that  $T$  is replaced by  $T-1$ , it follows that we may iterate this  $T-1$  times to get  $P\{D_1 = \epsilon_1\} = p^{\epsilon_1} (1-p)^{1-\epsilon_1}$ , which is to say  $P\{D_1 = 1\} = p$  and  $P\{D_1 = 0\} = 1-p$ . Likewise, if perform the same expansion as above but using variables other than the last, we find in fact  $P\{D_k = 1\} = p$  and  $P\{D_k = 0\} = 1-p$  for all  $k$ . In other words,  $D_1, \dots, D_T$  are iid Bernoulli variates with parameter  $p$ .