Probability and Applications to Finance

David Karapetyan

Contents

Part 1

Basic Probability

CHAPTER 1

Outcomes, Events, and Likelihood

Throughout this course, we seek to rigorously describe the likelihood of outcomes of an experiment, or, more generally, combinations of outcomes of an experiment. For example, we would like to know the likelihood of the Dodgers winning the pennant, or the likelihood that stock of a company will go up in a week. We have used words to describe these events—since we are doing mathematics, we would like to translate everything to numbers, without losing any of the meaning.

Example. A fair coin is tossed. There are two possible outcomes: heads (denoted by 0) and tails (denoted by 1). We call $\{0, 1\}$ the *outcome space* or *sample space*. The *event space* consists of

(i) Outcome is heads or tails
(ii) Outcome is heads and tails
(iii) Outcome is heads
(iv) Outcome is tails

which we translate, via numbers, to the set $\{\{0\} \cup \{1\}, \{0\} \cap \{1\}, \{0\}, \{1\}\}$. Observe that the event "Heads or tails" is the union of the events "Outcome is Heads" and "Outcome is tails", and "Heads and tails" is their intersection.

Example. A die with six distinct faces is thrown. The outcome space is $\{1, 2, 3, 4, 5, 6\}$. Some possible events are

(i) The outcome is even
(ii) The outcome is odd
(iii) The outcome is greater than 2
(iv) The outcome is less than 2.

These events, respectively, are expressed via the sets $\{2, 4, 6\}, \{1, 3, 5\}, \{3, 4, 5, 6\}$, and $\{1\}$. Observe that there are many other possible outcomes.

Observe that the larger the outcome space, the more complex our event space may become. Also observe that the event "The outcome is even" is the *complement* (with respect to the outcome space) of the event "The outcome is odd". That is, $\{2, 4, 6\}^c = \{1, 3, 5\}$.

Motivated by these two examples, we seek to generalize our notion of event space such that it completely models all possible events that can arise from a given outcome space. We have the following.

Definition. Let Ω be an outcome space. Then a *σ-algebra* or *filter* on Ω, denoted by $\mathcal{F}(\Omega)$ is defined to be a set of subsets of Ω that is closed under countable unions and complements. More precisely

(i) $A_i \in \sigma \implies \cup_i A_i \in \sigma$
(ii) $A_i \in \sigma \implies A_i^c \in \sigma$

EXAMPLE. Suppose we are tossing a fair coin repeatedly until the first tail shoes up, and wish to know the number of tosses we must use. Then our outcome space is $\{\{0\},\{0,1\},\{0,0,1\},\ldots\}$. This is an infinite outcome space.

EXAMPLE. The following are all filters of Ω:

(i) $\{\emptyset, \Omega\}$
(ii) $\{\emptyset, A, A^c, \Omega\}$
(iii) The power set of Ω

Observe that a natural outcome space to use is the real line, or subsets t and to build appropriate filters on top of it. We will return to this idea later.

CHAPTER 2

Discrete Probability

1. Unconditional Probability

Suppose we roll a 6-sided die N times. Let $N(A)$ denote the number of $4's$ rolled in N tries. One can observe that

$$\lim_{N\to\infty} N(A)/N = 1/6.$$

A similar phenomenon holds for many other real-world events.

Definition. Let $N(A)$ denote the number of "successes" S in N trials. If the limit $N(A)/N \doteq P$ exists, we say S occurs with probability P.

Observe that $0 \leq P \leq 1$, by definition. Next, let A and B be two disjoint events. Then it is easy to check that

$$N(A \cup B) = N(A) + N(B)$$

and that, in general (i.e. for possibly disjoint A, B),

$$N(A \cup B) = N(A) + N(B) - N(A \cap B).$$

Furthermore, since $\Omega = A \cup A^c$, it follows immediately that $N(\Omega) = N$. A similar argument shows $N(\emptyset) = 0$. This discussion motivates the following.

Definition. A *probability measure* on $(\Omega, \mathcal{F})$ is a continuous function $\mathbb{P} : \mathcal{F} \to [0, 1]$ which satisfies

(i) $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$
(ii) For disjoint A_i, $\mathbb{P}(\cup_i^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$.

Observe that our construction immediately rules out silly probability measures such as $P \equiv 0$ or $P \equiv 1$. Furthermore, we can show the following, whose proof we leave to the reader.

Lemma 1. *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$*

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
(ii) $\mathbb{P}(A \cup B) = P(A) + P(B) - P(A \cap B)$.
(iii) If $B \subset A$, then $P(A) \geq P(B)$.
(iv) (Set Continuity). *If $\cup_i^\infty A_i = A$ and $\cap B_i^\infty = B$, with $A_i \subset A_j$ and $B_j \subset B_i$ for $j > i$, then*

$$\lim \mathbb{P}(A_i) = \mathbb{P}(\lim \cup_i^n A_i) = \mathbb{P}(A)$$
$$\lim \mathbb{P}(B_i) = \mathbb{P}(\lim \cap_i^n B_i) = \mathbb{P}(B)$$

A last technical distinction: the empty-set denotes the event "nothing occurs" and has probability 0, by definition. However, there are many events in a given sigma-algebra that have probability 0 but that are not the empty set.

Example. What is the probability that, using a fair coin, one flips only heads in infinitely many tries?

Letting H_N denote the event N heads in the first N tries, set continuity yields continuity,

$$P(N_\infty) = \lim_{N\to\infty} \mathbb{P}(N_H) = \lim_{N\to\infty} 2^{-N} = 0.$$

2. Conditional Probability

We now wish to tackle statement that often occur in practice, such as "What is the probability that it will snow today, *given* that the sky is grey today", or "What is the probability that a mystery word is 'zebra', *given* that the first three letters are 'z', 'e', 'b'." Observe that when we are given information, it allows us to adjust the probability of the event we are interested in. We wish to express this mathematically. Motivated by the "zebra" example, we have the following.

Definition. Let $\Omega, \mathcal{F}$ be a filter, and $A, B \in \mathcal{F}$ be events, where $\mathbb{P}(B) > 0$. Then $\mathbb{P}(A|B)$ is defined to be the probability of A, given that $A \in \mathcal{F}_B$, where $\mathcal{F}_B \doteq \{U \in \mathcal{F} : B \cap U \neq \emptyset\}$.

Imagine that we roll a die, and wish to know the probability that we have rolled a 4. To compute this, we let A denote the event that we roll a 4. Then in the last section we saw that

$$\mathbb{P}(A) = \lim_{N\to\infty} N(A)/N.$$

Suppose now that we are given the event B, which is the event that we haven't rolled a 6. What's the probability now that we have rolled a 4? Stated a bit differently, how can we modify our ratio above to reflect the new value? We simply discard all trials whose rolls gave a 6, as this is clearly impossible with the new die. Of course, we must also discard the number of times we rolled a 6. Putting this all together, we obtain

$$\begin{aligned}
\mathbb{P}(A|B) &= \lim_{N\to\infty} N(A \cap B)/(N - \text{number of occurrences of 6}) \\
&= \lim_{N\to\infty} \frac{N(A\cap B)}{N} \times \frac{1}{1 - (\text{number of occurrences of 6})/N} \\
&= \frac{\mathbb{P}(A\cap B)}{(1 - \mathbb{P}(B^c))} \\
&= \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.
\end{aligned}$$

Motivated by this, we have the following definition.

Definition. Let A, B be events, and suppose B occurs. Then

$$\mathbb{P}(A|B) \doteq \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

Observe that it is implicit in the definition that $\mathbb{P}(B) > 0$, otherwise B does not occur.

Example. A family has two children. What is the probability that both are boys, given that at least one of them is a boy?

Example. A family has two children. What is the probability that both are boys, given that the youngest is a boy?

Observe that if no information is given, the probability of two boys is 1/4. Hence, the conditions given in the example above improve the probability. Sometimes, given information may reduce the probability of an event. If the given information has no effect on the probability of an event, we say that the event and the information are *independent*. We will discuss this more in the upcoming lectures.

Often, when information is given, it makes computing probabilities easier. Consequently, the following result is extremely useful.

Lemma 2. *Let $\{B_i\}_{i=1}^n$ be a partition of Ω. Then for any $A \in \mathcal{F}$*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

the above lemma has the limitation that sometimes $\mathbb{P}(A|B)$ is difficult to compute as well. In such situations, it is often easier to compute $\mathbb{P}(B|A)$. Fortunately, we have the following.

Lemma 3 (Bayes' Theorem)**.** *For events A, B, we have*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

Example. We are given two urns, each containing an assortment of colored balls. Urn I contains two white and three blue balls, and urn II contains three white and four blue balls. A ball is drawn at random from urn I and put into urn II, and then a ball is picked at random from urn II. What is the probability that it is blue?

Example. You are on "Let's Make a Deal". There are three doors, with a new convertible behind one, and a goat behind each of the others. You pick door number I. To tease you, Monty opens door number II, revealing a goat, then offers to let you switch your choice to door number III. Should you?

Example (Symmetric Random Walk). Las Vegas has decided to offer a new game, with 50/50 odds of winning or losing in each iteration. The player begins with \$$k$ and the house begins with \$$N$, where $N \gg k$. If the player wins in a round, he wins a dollar; otherwise, he loses a dollar. The player adopts the strategy to continue playing until either he or the casino is bankrupted. What the probability that the player goes bankrupt?

Solution. Let A denote the event that the player is eventually bankrupted, and let B denote victory for the player in the first trial. Then

$$\begin{aligned}\mathbb{P}_k(A) &= \mathbb{P}_k(A|B)\,\mathbb{P}(B) + \mathbb{P}_k(A|B^c)\,\mathbb{P}(B^c)\\ &\approx \frac{p_{k+1}}{2} + \frac{p_{k-1}}{2}\end{aligned}$$

We have the boundary conditions $p_0 = 1$ and $p_N = 0$, which we combine with the above *difference equation* to obtain

$$\mathbb{P}_k(A) = 1 - k/N.$$

□

Observe that if we start off with $k \approx N$, the player stands a chance to bankrupt the casino! This is one of the reasons all games in Vegas have odds favoring the house. Also, the above example motivates casinos to have a lot of cash on hand, so that the amount you have at any given time is dwarfed by comparison. If you are a high roller with a bankroll rivaling the casino's, the casino will try to get you to play a game that give you terrible odds over the long run (for example, craps). If you decide to play blackjack (the game offering the best odds), the casino will try to distract you with drinks, entertainment, women/men, etc. It usually works.

Example (False Positives). A rare disease affects one person in 10^5. A test for the disease is wrong with probability $1/100$; that is, it is positive with probability $1/100$ for someone who is in fact healthy, and negative with probability $1/100$ for someone who is in fact ill. What is the probability that you have the disease given that you took the test and it is positive?

Solution. Let A be the event that we have the disease, and B be the event that the test if positive. Then we apply Bayes' Theorem to obtain

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B|A)\,\mathbb{P}(A) + \mathbb{P}(B|A^c)\,\mathbb{P}(A^c)} \\ &= \frac{99/100 \times 1/10^5}{99/100 \times 1/10^5 + 1/100 \times (10^5-1)/10^5} \\ &\approx 1/1000. \end{aligned}$$

□

Moral: don't freak out if your doctor says you *might* have cancer. Take the test again.

3. Independence

Definition. We say two events A, B, $\mathbb{P}(B) > 0$ are *independent* if

$$\mathbb{P}(A|B) = \mathbb{P}(A),$$

or, equivalently, that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

More generally, a family $\{A_i\}_{i=1}^n$ is *independent* if

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).$$

Example. What is the probability that Connie's first child will be a masculine child, given that a neighboring mobster recently had a boy?

Example. What is the probability of flipping heads with a fair coin on the $10th$ trial, given that heads is flipped on all previous trials?

CHAPTER 3

Translating Outcomes to Numbers

As we have seen, the sample space Ω, equipped with an associated filter $\mathcal{F}$ and probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ are the linchpins of our theory of probability. However, in practice, abstract outcomes $\omega \in \Omega$ and sets $E \in \mathcal{F}$ are cumbersome to work with. We would like to "translate" our analysis on $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, dx)$, where $\mathcal{B}_{\mathbb{R}}$ denotes the Borel filter on $\mathbb{R}$ (the smallest filter containing all the open sets in $\mathbb{R}$), and dx is Lebesgue measure. This translation will be given by continuous functions from Ω, equipped with its filter structure, to $\mathbb{R}$, equipped with the Borel filter. We will call such functions *random variables.*

1. Random Variables

Definition. A *random variable* is a function $X : \Omega \to \mathbb{R}$ with the property that $X^{-1}(U) \in \mathcal{F}$ for every open $U \in \mathbb{R}$. If the range of X is a countable subset of $\mathbb{R}$, we say X is a *discrete random variable.* If Ω is countable, we say that $\{\Omega, \mathcal{F}, \mathbb{P}\}$ is a *discrete probability space.*

As shorthand, we shall often denote $\{\omega : X(\omega) \in B\}$ by $\{X \in B\}$.

Remark. Observe that, for a discrete random variable $X : \Omega \to [\alpha_1, \dots, \alpha_n]$, $n \leq \infty$,

$$\mathbb{P}(\Omega) = \sum_{i=1}^{n} \mathbb{P}(X = x_i).$$

Lastly, we say two random variables X, Y are *independent* if $\{X \in B_1\}$ and $\{Y \in B_2\}$ are independent for all Borel sets B_1, B_2.

Example. Suppose we are interested in studying a single flip of a fair coin. Let $\Omega = \{H, T\}$ be the outcomes of a coin flip, and $\mathcal{F} = \{\emptyset, \Omega, H, T\}$. Then $X : \Omega \to \mathbb{R}$, given by $X(H) = 0$, $X(T) = 1$ is a discrete random variable.

Example. A traveler is lost in the woods, and starts walking aimlessly, but never west. We can assign numbers to the directions he takes. Then Ω is the set of all possible directions (North, South, East and the *continuum* of values in between these). Then $X : \Omega \to \mathbb{R}$ given by $X(North) = 1$, $X(South) = -1$, $X(East) = 0$ and $F_X \doteq \{X^{-1}(U) : U \in \mathbb{R}\}$ is a continuous random variable.

Definition. For a space $\{\Omega, \mathcal{F}\}$, we say that a measure u is *absolutely continuous* respect to a measure v, denoted $u \ll v$, if $v(E) = 0$ implies $u(E) = 0$ for every $E \in \mathcal{F}$.

Theorem 4 (Radon-Nikodym). *Let $u, v : \Omega \to \mathbb{R}^n$ be measures on $\mathcal{F}_\Omega$, with $u \ll v$. Then there exists a positive $f : \Omega \to \mathbb{R}^n$ such that*

$$u(E) = \int_E f(\omega) dv(\omega)$$

for every $E \in \mathcal{F}$.

The proof of this theorem lies outside the scope of this course, but can be found in almost every graduate textbook on analysis.

We now apply the Radon-Nikodym theorem to define continuous random variables.

Definition. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *distribution measure* of X is the probability measure $\mu_X : \mathcal{B}_X \to [0,1]$ given by $\mu_X(B) = \mathbb{P}(\{\omega\} : X(\omega) \in B)$.

Definition. A random variable X is discrete if and only if it takes values in a countable subset of $\mathbb{R}$, and continuous if and only if $\mu_X \ll dm$, where dm denotes Lebesgue measure on $\mathbb{R}^n$. By Radon-Nikodym, its *distribution function* $F(x) \doteq \mathbb{P}(X \leq x)$ is given by

$$F(x) = \int_{-\infty}^{x} f(z)\, dz.$$

for some unique, positive $f(z)$, which we call *probability density* function of X. If X is discrete, we call $f(x) \doteq \mathbb{P}(X = x)$ its *mass function*.

Lastly, we remark that there exist random variables which are neither discrete nor continuous, but rather a mixture of the two.

Example (A Random Variable that is Neither Discrete Nor Continuous). A coin is tossed. We assign the event "lands heads" the number -1, and if it lands tails, we toss a rod, and assign "lands tails" how far the rod has landed from us. In this case, our random variable X is neither continuous nor discrete: it has a point mass at $X = H$, but is continuous otherwise. Observe that $X|$lands tails and $X|$lands heads are continuous and discrete, respectively.

2. Common Discrete Distributions

2.1. Bernoulli Distribution. We begin with the most basic distribution, from which we are able to derive a multitude of others.

Definition. Let X be the random variable denoting the number of successes in n independent trials, where p is the probability of success in an individual trial. Then

$$b(k, n, p) \doteq \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

is called the *Bernoulli distribution* of k successes in n independent trials, where p is the probability of success on an individual trial. The special case $p = 1/2$ is called a *binomial distribution*.

Example. Let X denote the number of heads one obtains from n flips of a two-sided, non-weighted coin. Then X is binomially distributed. If the coin is weighted, then X follows a Bernoulli distribution.

2.2. Poisson Distribution. We wish now to study events that are distributed sparsely in time. Suppose we have a large number n of independent trials, with success in each individual trial unlikely. In order to have a sparse distribution of events that is nontrivial (i.e. we have no successes), we need to assume that there is a positive probability that we achieve at least one success in n trials. Hence, we assume the expected number of successes $\lambda \doteq np$ to be nonzero.

Observe that with the above assumptions, we obtain

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \approx \frac{n^k}{k!}$$

and so

$$\begin{aligned} P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &\approx \frac{n^k p^k q^{n-k}}{k!} \\ &= \frac{\lambda^k q^{n-k}}{k!} \\ &= \frac{\lambda^k (1-p)^{n-k}}{k!} \\ &= \frac{\lambda^k (1-\lambda/n)^{n-k}}{k!} \\ &\approx \frac{\lambda^k (1-\lambda/n)^{n}}{k!}, \quad n >> 1 \\ &\approx \frac{\lambda^k e^{-\lambda}}{k!} \end{aligned}$$

Definition. We call

$$p(k,\lambda) \doteq \frac{\lambda^k e^{-\lambda}}{k!}$$

the *Poisson distribution* of k successes in a large number of trials, where, on average, we expect λ successes, where λ is much less than the number of trials. That is, the trials have the feature that successes are sparsely distributed.

We will now extend this result. Assume that in the interval $[0,1]$, we have n equally spaced points, representing trials. Split this interval into two pieces, $[0,t]$ and $[t,1]$. Then we have nt trials in interval $[0,t]$ and $n-nt$ trials in interval $[t,1]$. Then the average number of successes in $[0,t]$ is given by $\lambda_t \doteq ntp = \lambda t$. Repeating our preceding computation, but with λ replaced by λ_t and n by nt, we obtain the following.

Theorem 5. *Let λ be the average number of successes in the interval $[0,1]$. The probability of finding exactly k successes in the subinterval $[0,t]$, $t \leq 1$, is given by*

$$p(k,\lambda t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

with average λt.

Exercise. Generalize this result to intervals of arbitrary length.

2.3. Negative Binomial Distribution. For an experiment, it is often important to know just how many trials one needs to achieve a certain number of successes. To tackle this problem, we ask a simpler question: out of a total of n independent trials with either success or failure as the outcome, what is the probability that the $r'th$ success occurs at the $v'th$ trial, where $r \leq v \leq n$? We reason this out as follows: there is a total of $\binom{v-1}{r-1}$ ways to arrange $r-1$ successes out of $v-1$ trials, each with associated probability $p^{r-1}(1-p)^{(v-1)-(r-1)} = p^{r-1}q^{v-r}$.

Right after the $v-1$ trial, we would like to have the $r'th$ success. This success has associated probability p. Putting everything together, we obtain the following.

THEOREM 6. *Let X denote the number of successes achieved after v independent trials of an experiment, where we assume a success occurs at trial v, and where the probability of success if p. Then the probability that $X = r$, where $r \leq v$, is given by*

$$p(r, v, p) \doteq \mathbb{P}(X = r) = \binom{v-1}{r-1} p^r q^{v-r}$$

DEFINITION. We call X a random variable with a *negative binomial distribution.*

3. Common Continuous Distributions

3.1. Uniform Distribution.

DEFINITION. Let X be a continuous random variable on a probability space $\Omega, \mathcal{F}, \mathbb{P}$ such that the minimum and maximum values of X are $a, b \in \mathbb{R}$, respectively, and $\mathbb{P}(c \leq X \leq d) = \mathbb{P}(c+t \leq X \leq d+t)$ for all c, d in $[a, b]$ and $t \in \mathbb{R}$ such that $c+t, d+t \in [a, b]$. Then we say X is *uniformly distributed* in $[a, b]$. It is easy to check that X has density

$$f(x) = \begin{cases} x/(b-a), & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

3.2. Normal Distribution.

DEFINITION. Let X be a continuous random variable on a probability space $\Omega, \mathcal{F}, \mathbb{P}$. We say X is *normally distributed* with standard deviation σ and mean μ if X has density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

If $\mu = 0$ and $\sigma = 1$, we say X is *standard normally distributed.*

CHAPTER 4

Joint Distributions

Up to now, we have considered random variables taking values in $\mathbb{R}$. It is easy to extend this theory to random variables taking values in $\mathbb{R}^n$: our distribution of X is then given by $\mu_X(B) = \mathbb{P}(X^{-1}(B))$ for every Borel set $B \in \mathbb{R}^2$. Using our intuition, we ask if our distribution function should be of

$$F(x) = \int_{-\infty}^{x} f(s)\,ds$$

where now ds denotes Lebesgue measure on $\mathbb{R}^2$ and $x = (x_1, x_2) \in \mathbb{R}^2$. However, what do we mean when we say $x \le y$ for $x, y \in \mathbb{R}^2$?

First, observe that if $X : \Omega \to \mathbb{R}^n$, then $X = (X_1, X_2, \dots, X_n)$, where $X_i : \Omega \to \mathbb{R}$. Hence, our study of X will be simplified if we can study it component-wise. Motivated by this, we define $x \le y$ if and only if $x_i \le y_i$.

Definition. Let $X = (X_1, X_2, \dots, X_n)$. Then the *joint distribution* $F_X : \mathbb{R}^n \to \mathbb{R}$ is defined as $F_X(x) = \mathbb{P}(X \le x)$.

Lemma 7. *Let F_X be a joint distribution. Then*

(i) $\lim_{x\to\infty} F_X(x) = 1$ *and* $\lim_{x\to-\infty} F_X(x) = 0$
(ii) If $x \le y$, then $F_X(x) \le F_X(y)$.
(iii) $F(x)$ is continuous from above.

If all the X_i are discrete, we say X has a *joint discrete distribution.* If all the X_i are continuous, we say X has a *joint continuous distribution.* If the distribution is jointly discrete, then for $x = (x_1, x_2, \dots, x_n)$, $f(x) = \mathbb{P}(X_1 = x_1, X_2, = x_2, \dots, X_n = x_n)$. If the distribution is jointly continuous, then by Radon-Nikodym

$$F(x) = \int_{-\infty}^{x} f(s)\,ds = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots x_n)\,ds_1\,ds_2 \dots\,ds_n.$$

One of the primary applications of joint distributions is in finding the distributions and associated density functions of sums and products of random variables.

Example. Let X and Y be continuous random variables, with associated densities $f_X(x)$ and $f_Y(y)$, respectively. What are the density functions of $Z = X + Y$ and $Z = XY$? What if X and Y are independent?

Recall that a continuous random variable $X : \Omega \to \mathbb{R}^n$ is defined to be a random variable whose distribution

$$u_X(B) \doteq \mathbb{P}(X \in B)$$

is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^n$. Hence, by the Radon-Nikodym theorem

$$\begin{aligned} F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}((X,Y) \in B_z) &= \int_{B_z} f\, dm \\ &= \int_{B_y} \int_{B_x} f(x,y)\, dx dy \end{aligned}$$

Applying the observation that $B_z = \{(x,y) : -\infty < x < \infty, y \le z - x\}$, we obtain the distribution function

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{(X,Y)}(x,y) dy dx.$$

Differentiating with respect to z, we arrive at the density

$$f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x) dx.$$

If X and Y are independent, then $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$, giving

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = f_X * f_Y(z).$$

Similarly, for $Z = XY$, the distribution function is given by

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z/x} f_{(X,Y)}(x,y) dy dx$$

with density

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{x} f_{(X,Y)}(x, z/x) dx.$$

If X and Y are independent, this density becomes

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{x} f_X(x) f_Y(z/x) dx.$$

CHAPTER 5

Expectation

Let $X : \Omega \to [\alpha_1, \dots, \alpha_k]$, $k < \infty$ be a random variable denoting the outcome of some experiment, and let $X_1, \dots, X_n$ be independent trials of the same experiment, where $X_i : \Omega \to [\alpha_1, \dots, \alpha_k]$. Then we call

$$\mathbb{E}(X) \doteq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \doteq \mu$$

the *expectation*, or *mean*, or X. If the limit does not exist, or $\mu = \pm\infty$, we say that the expectation does not exist.

Example. Let X be a discrete random variable with mass function[1]

$$f(k) = \begin{cases} Ak^{-2}, & k \in \mathbb{Z}^+ \\ 0, & \text{otherwise} \end{cases}$$

Then it is easy to check that

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kAk^{-2} = A\sum_{k=1}^{\infty} k^{-1} = \infty.$$

Observe that we can write

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\alpha_i(\#\alpha_i) + \cdots + \alpha_n(\#\alpha_n)}{n} \to \alpha_1\,\mathbb{P}(\alpha_1) + \cdots + \alpha_n\,\mathbb{P}(\alpha_n).$$

This motivates the following.

Definition. The *expectation* of a random variable X with mass function f is

$$\mathbb{E}(X) = \sum_{x\in\mathbb{R}:f(x)>0} xf(x), \quad \text{provided} \quad \sum_{x\in\mathbb{R}:f(x)>0} |xf(x)| < \infty$$

If X is continuous with density $f(x)$, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)\,dx, \quad \text{provided} \quad \int_{-\infty}^{\infty} |xf(x)|\,dx < \infty.$$

Lemma 8. *If $g : \mathbb{R} \to \mathbb{R}$ is continuous, then*

$$\mathbb{E}(g(X)) = \sum_{x\in\mathbb{R}:f(x)>0} g(x)f(x) \longleftrightarrow \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

Theorem 9 (Properties of Expectation). *Let X be a random variable. Then*

(i) $\mathbb{E}(aX + bY) = aX(X) + b\,\mathbb{E}(Y)$.

(ii) $E\ (1) = 1$

[1] Observe that for f to be a mass function, A must be chosen such that $\sum_{k=1}^{\infty} k^{-2} = 1$.

(iii) If X, Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

PROOF. We shall only do the discrete case; the proof for the continuous case is analogous.

(i) This follows immediately from the linearity of sums and integrals.
(ii) We have $\mathbb{E}(1) = \sum xf(x) = \sum_{x=1} xf(x) = 1$.
(iii) Let $Z = XY$. Then

$$\mathbb{E}(Z) = \sum_z zf_Z(z) = \sum_{x,y} xyf_{(X,Y)}(x,y) = \sum_x xf_X(x) \sum_y yf_Y(y) = \mathbb{E}(X)\,\mathbb{E}(Y).$$

□

Having defined expectation, we can use it to define the *variance* of a random variable X

$$\mathbb{E}([X - \mathbb{E}(X)]^2),$$

the *standard deviation*

$$\sigma_X \doteq \sqrt{\mathrm{Var}(X)}$$

and *covariance* of random variables X, Y

$$\mathrm{Covar}(X, Y) \doteq \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)])$$

The standard deviation of a random variable X is the average spread of the trials of an experiment away from the mean, and is always positive.

In practice, we consider the normalized covariance of two random variables, commonly called the *correlation coefficient*

$$\rho(X, Y) = \frac{\mathrm{Covar}(X, Y)}{\sigma_X \sigma_Y}$$

The correlation coefficient has the following important property.

THEOREM 10. *For random variables X, Y, $|\rho(x, y)| \leq 1$, where $|\rho(x, y)| = 1$ if and only if $aX + bY = 1$ almost surely, for some $a, b \in \mathbb{R}$.*

PROOF. It is a standard application of Cauchy-Schwarz. □

From this theorem, we see that the correlation index of two random variables is a measure of the linear dependence of one random variable on another. If the index is negative, a rise in one random variables implies a fall in the other, and vice verse. If the coefficient is positive, then the variables rise or fall together.

DEFINITION. We say two random variables X and Y are *correlated* if

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y).$$

Observe that independent random variables are uncorrelated. However, lack of correlation does not, in general, imply independence. Heuristically, this is because correlation is only a measure of the degree of linear dependence between variables, and does not capture higher order (for example, quadratic) types of dependence.

EXAMPLE. Let X be a continuous random variable with the standard normal distribution, and let $Y = X^2$. Clearly X and Y are dependent. However,

$$\mathbb{E}(X^3) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 e^{-x^2/2}\,dx = 0$$

by anti-symmetry, and so

$$\mathbb{E}(XY) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X)\,\mathbb{E}(Y).$$

which implies X, Y are uncorrelated.

CHAPTER 6

Conditional Expectation

DEFINITION. The *conditional distribution function* of Y given $X = x$, written $F_{Y|X}(\cdot|x)$ is defined by

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x).$$

In the discrete case, the *conditional mass function* is given by

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x).$$

Observe that, in the discrete case

$$\begin{aligned} f_{(X,Y)}(x, y) &= \mathbb{P}_{\mathcal{F}\times\mathcal{F}}(X \leq x, Y \leq y) \\ &= \mathbb{P}_{\mathcal{F}}(X \leq x \cap Y \leq y) \\ &= \mathbb{P}_{\mathcal{F}}(X \leq x|Y \leq y)\,\mathbb{P}_{\mathcal{F}}(Y \leq y) \\ &= f_{X|Y}(x, y) f_y(y) \end{aligned}$$

and so

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

and similarly

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}.$$

We will heretofore suppress the distinction between $\mathbb{P}_{\mathcal{F}\times\mathcal{F}}$ and $\mathbb{P}_{\mathcal{F}}$, for the sake of clarity.

Observe that in the continuous case, we have some difficulties, since we are not allowed to condition on null events. However, we may consider

$$\mathbb{P}(Y \leq y, x \leq X \leq x + h) = \mathbb{P}(Y \leq y|x \leq X \leq x + h)\,\mathbb{P}(x \leq X \leq x + h)$$

which we can rewrite, using Taylor series, as

$$\begin{aligned} \mathbb{P}(Y \leq y|x \leq X \leq x + h) &= \frac{F_{(X,Y)}(x + h, y) - F_{(X,Y)}(x, y)}{F_X(x + h) - F_X(x)} \\ &= \frac{hf_{(X,Y)}(x, y) + O(h^2)}{hf_X(x) + O(h^2)} \\ &= \frac{f_{(X,Y)}(x, y) + O(h)}{f_X(x) + O(h)} \end{aligned}$$

Letting $h \to 0$, we obtain

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}.$$

Similarly,

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}.$$

Thus, the formulas for conditional distributions functions in both the discrete and continuous cases are identical.

Since $Y|X = x$ is a random variable, we can write $\psi(x) = \mathbb{E}(Y|X = x)$. Then $\psi(X)$ is a random variable, which we call the expectation of Y given X, also written as $\mathbb{E}(Y|X)$.

THEOREM 11. *For any two random variable $X, Y : \Omega \to \mathbb{R}^n$,*

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$$

PROOF. For the discrete case, we have

$$\begin{aligned} \mathbb{E}(\psi(X)) &= \sum_x \psi(x) f_x(x) \\ &= \sum_{x,y} y f_{Y|X}(y|x) f_X(x) \\ &= \sum_{x,y} y f_{(x,y)} \\ &= \sum_y y f_Y(y) \\ &= \mathbb{E}(Y). \end{aligned}$$

The proof for the continuous case is analogous. □

THEOREM 12. *Let X, Y be random variables as before, and $g(x)$ be a function such that $\mathbb{E}(g(X)) < \infty$. Then*

$$\mathbb{E}(\mathbb{E}(Y|X)g(X)) = \mathbb{E}(Yg(X)).$$

PROOF. It is almost identical to that of the preceding theorem. □

The two theorems above are immensely useful in solving a wide variety of discrete and non-discrete problems. Heuristically, they allow us to divide a difficult probability problem into several simpler problems.

EXAMPLE. Suppose we flip a fair coin, and assign -1 points to tails, and $+1$ points to heads. We play a game where we flip a coin repeatedly until we obtain -1 points, or 2 points. What is the expected number of flips before the game terminates?

Let $X_i, -1 \le i \le 2$ denote the number of flips it will take for the game to terminate if we begin with i points. Then we have

$$\begin{aligned}
\mathbb{E}(X_0) &= \mathbb{E}(X_0|\text{first flip heads})\,\mathbb{P}(\text{first flip heads}) + \mathbb{E}(X_0|\text{first flip tails})\,\mathbb{P}(\text{first flip tails}) \\
&= (\mathbb{E}(X_1)+1)\frac{1}{2} + (\mathbb{E}(X_{-1})+1)\frac{1}{2} \\
&= \frac{1}{2}(\mathbb{E}(X_1)+\mathbb{E}(X_{-1})) + 1 \\
&= \frac{1}{2}\,\mathbb{E}(X_1) + 1 \\
&= \frac{1}{2}[\mathbb{E}(X_1|\text{flip heads})\frac{1}{2} + \mathbb{E}(X_1|\text{flip tails})\frac{1}{2}] + 1 \\
&= \frac{1}{4}[\mathbb{E}(X_1|\text{flip heads}) + \mathbb{E}(X_1|\text{flip tails})] + 1 \\
&= \frac{1}{4}[\mathbb{E}(X_2) + 1 + \mathbb{E}(X_0) + 1] + 1 \\
&= \frac{1}{4}[\mathbb{E}(X_0) + 2] + 1
\end{aligned}$$

which gives $\mathbb{E}(X_0) = 2$.

Example. A hen lays N eggs, where N has a Poisson distribution with parameter λ. An egg hatches with probability p, independently of the other eggs. Let K be the number of chicks. Compute $\mathbb{E}(K|N)$, $\mathbb{E}(K)$, and $\mathbb{E}(N|K)$.

To compute $\mathbb{E}(K|N)$, we must first recall the Poisson distribution

$$f_N(n) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Given N eggs, the probability that K out of N eggs hatch obeys a Bernoulli distribution, and so

$$f_{K|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$$

Now, $\mathbb{E}(K|N=n) = pn$, and so $\mathbb{E}(K|N) = pN$. Furthermore, $\mathbb{E}(K) = \mathbb{E}(\mathbb{E}(K|N)) = \mathbb{E}(pN) = p\lambda$. To compute $\mathbb{E}(N|K)$, we apply Bayes Theorem to obtain

$$\begin{aligned}
f_{N|K}(n|k) &= \frac{f_{K|N}(k|n) f_N(n)}{f_K(k)} \\
&= \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} \lambda^n e^{-\lambda}/n! / \sum_{m \ge k} \binom{m}{k} p^k (1-p)^{m-k} (\lambda^m e^{-\lambda})/m!, & n \ge k \\ 0, & n < k \end{cases}
\end{aligned}$$

where we obtained the denominator via the observation

$$\begin{aligned}\mathbb{P}(K=k) &= \mathbb{P}(K=k \cap N \geq k)\\ &= \sum_{m\geq k} \mathbb{P}(K=k \cap N=m)\\ &= \sum_{m\geq k} \mathbb{P}(K=k|N=m)\,\mathbb{P}(N=m)\\ &= \sum_{m\geq k} f_{K|N}(k,m) f_N(m)\\ &= \sum_{m\geq k} \binom{m}{k} p^k (1-p)^{m-k} (\lambda^m e^{-\lambda})/m!.\end{aligned}$$

Simplifying, we obtain

$$f_{N|K}(n|k) = \begin{cases} (q\lambda)^{n-k} e^{-q\lambda}/(n-k)!, & n \geq k\\ 0, & n < k\end{cases}$$

and so

$$\begin{aligned}\mathbb{E}(N|K=k) &= \sum_{n\geq k} n f_{N|K}(n|k)\\ &= \sum_{n\geq 0} (n+k)\frac{(q\lambda)^n e^{-q\lambda}}{n!}\\ &= k\sum_{n\geq 0} \frac{(q\lambda)^n e^{-q\lambda}}{n!} + \sum_{n\geq 0} n\frac{(q\lambda)^n e^{-q\lambda}}{n!}\\ &= k + q\lambda.\end{aligned}$$

Hence, $\mathbb{E}(N|K) = K + q\lambda$.

CHAPTER 7

Characteristic Functions

If $f = f(x)$ is integrable (that is, $\int_{\mathbb{R}} |f|\ dx < \infty$), we define its *Fourier transform* by

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-ix\xi} f(x)\ dx.$$

If $\hat{f}(\xi)$ is integrable as well, then we have the following.

THEOREM 13. *If $f, \hat{f}$ are integrable, then*

$$f(x) = \int_{-\infty}^{\infty} e^{ix\xi} \hat{f}(\xi)\ d\xi.$$

PROOF. See Stein and Shakarchi [?zbMATH02171466]. ☐

Proposition 14. *The Fourier transform has the following properties.*

(i) $\widehat{f(x+h)}(\xi) = \widehat{f(\xi)e^{ih\xi}}$
(ii) $\widehat{f(x)e^{-ixh}}(\xi) = \widehat{f}(\xi + h)$
(iii) $\widehat{\frac{d}{dx}f(x)}(\xi) = i\xi$
(iv) $\widehat{-ixf(x)}(\xi) = \frac{d}{d\xi}\widehat{f}(\xi)$
(v) $\hat{\hat{f}}(x) = f(-x)$
(vi) $\int_{\mathbb{R}} \widehat{f}(x)g(x)\ dx = \int_{\mathbb{R}} f(x)\widehat{g}(x)\ dx$
(vii) $\widehat{f \star g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$.

PROOF. The proofs are a straightforward application of u-substitution, and interchanging order of integration. See Stein and Shakarchi [?zbMATH02171466]. ☐

DEFINITION. The characteristic function of a random variable X is given by

$$\phi_X = \mathbb{E}(e^{-itX}).$$

Observe that if X has a density f, then

$$\phi(t) = \int_{\mathbb{R}} e^{-itx} f(x)\ dx$$

This is just the Fourier transform of the density.[1]

Lemma 15. *If X, Y are independent, then*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

[1] Some texts set $\phi_X(t) = \mathbb{E}(e^{-itX})$, which is equivalent to our definition of the characteristic function, evaluated at $-t$.

PROOF. If X and Y are independent, then e^{-itX} and e^{-itY} are independent. Therefore

$$\phi_{X+Y}(t) = \mathbb{E}(e^{-itX-itY}) = \mathbb{E}(e^{-itX}e^{-itY}) = \mathbb{E}(e^{-itX})\,\mathbb{E}(e^{-itY}) = \phi_X(t)\phi_Y(t)$$

which completes the proof.[2] □

DEFINITION. The joint characteristic function of X and Y is given by

$$\phi_{X,Y}(s,t) = \mathbb{E}(e^{-isX}e^{-itY}).$$

Observe that if (X,Y) has the joint density function $f(x,y)$, then letting $F(x,y) = e^{-isx}e^{-ity}$, we obtain

$$\begin{aligned}\phi_{X,Y}(s,t) &= \mathbb{E}(e^{-isX}e^{-itY})\\ &= \mathbb{E}(F(X,Y))\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)f(x,y)\;dxdy\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-isx}e^{-ity}f(x,y)\;dxdy\end{aligned}$$

which is just the Fourier transform of the joint density in the x and y variables.

Lemma 16. *Two random variables X and Y are independent if and only if*

$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t), \quad \forall s,t$$

PROOF. We prove this result for the case where X and Y have densities. To establish necessity, assume independence. Then e^{-isX} and e^{-itY} are independent for all s,t, and so

$$\phi_{X,Y}(s,t) = \mathbb{E}(e^{-isX}e^{-itY}) = \mathbb{E}(e^{-isX})\,\mathbb{E}(e^{-itY}) = \phi_X(s)\phi_Y(t).$$

To establish the sufficiency, we prove the contrapositive. Assume there exist s,t such that

$$\mathbb{E}(e^{-isX}e^{-itY}) \neq \mathbb{E}(e^{-isX})\,\mathbb{E}(e^{-itY}).$$

Then this is equivalent to saying

$$\int_{-\infty}^{\infty} e^{-isx}g(x)\;dx \int_{-\infty}^{\infty} e^{-ity}h(y)\;dy \neq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-isx}e^{-ity}f(x,y)\;dxdy.$$

If X and Y are independent, then $f(x,y) = g(x)h(y)$, which implies

$$\int_{-\infty}^{\infty} e^{-isx}g(x)\;dx \int_{-\infty}^{\infty} e^{-ity}h(y)\;dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-isx}e^{-ity}f(x,y)\;dxdy$$

which is a contradiction. □

[2] Observe that if X and Y both have densities and are independent, then the density of $X+Y$ is $f\star g$, and so $\phi_{X+Y}(t) = \widehat{f\star g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi) = \phi_X(t)\phi_Y(t)$. Hence, the power of characteristic functions manifests itself when we are summing random variables: convolutions of functions are difficult to deal with, while products of functions, in general, are not.

1. Examples of Characteristic Functions

(1) Bernoulli distribution for one trial with parameter p. That is, the Bernoulli random variable outputs 1 with probability p, and 0 with probability $1-p$.

$$\phi(t) = \mathbb{E}(e^{-itX}) = e^{-it0}(1-p) + e^{-it}p = 1 - p + pe^{-it}$$

(2) Bernoulli distribution with n independent trials and parameter p. Letting $X = X_1 + \cdots + X_n$, where the X_i denote the trials, we have

$$\phi_X(t) = \phi_{X_1+\ldots+X_n}(t) = \prod_{1\le i\le n} \phi_{X_i}(t) = \left[1 - p + pe^{-it}\right]^n.$$

(3) Normal Distribution. If X is $N(0,1)$, then

$$\begin{aligned}
\phi(t) = \mathbb{E}(e^{-itX}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-itx} e^{-x^2/2}\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x^2-2itx)/2}\,dx \\
&= \int_{-\infty}^{\infty} e^{-(x-it)^2/2 - t^2/2}\,dx \\
&= e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-it)^2/2}\,dx \\
&= e^{-t^2/2}
\end{aligned}$$

where the last step follows from complex integration. We now compute the characteristic function of $Y \overset{d}{=} N(\mu, \sigma^2)$ by observing that $Y \overset{d}{=} \sigma X + \mu$, which gives

$$\begin{aligned}
\phi_Y(t) = \phi_{\sigma X+\mu} &= \phi_{\sigma X}(t)\phi_\mu(t) \\
&= e^{it\mu} \int_{-\infty}^{\infty} e^{it\sigma x} e^{-x^2/2}\,dx \\
&= e^{it\mu}\phi_X(\sigma t) \\
&= e^{it\mu} e^{-(\sigma t)^2/2} \\
&= e^{it\mu - (\sigma t)^2/2}
\end{aligned}$$

2. Convergence in Distribution Using Characteristic Functions

THEOREM 17. *Suppose $F_1, \cdots, F_n$ are a sequence of distributions with corresponding characteristic functions $\phi_1, \cdots, \phi_n$, respectively.*

(i) If $F_n \to F$ pointwise in $\mathbb{R}$, where F has characteristic function ϕ, then $\phi_n \to \phi$ pointwise in $\mathbb{R}$.
(ii) Conversely, if $\phi_n \to \phi$ pointwise in $\mathbb{R}$ and ϕ is continuous at $t = 0$, then ϕ is the characteristic function of some distribution F, and $F_n \to F$ pointwise in $\mathbb{R}$

Corollary 18. *If $F_1, \cdots, F_n$ and F are distribution functions with densities and corresponding characteristic functions $\phi_1, \cdots, \phi_n$ and ϕ respectively, then $F_n \to F$ pointwise in $\mathbb{R}$ if and only if $\phi_n \to \phi$ pointwise in $\mathbb{R}$.*

3. The Weak Law of Large Numbers, and the Central Limit Theorem

DEFINITION. We say a sequence of random variables $\{X_i\}$ with corresponding distribution functions $\{F_i\}$ *converges in distribution* to a random variable X with corresponding distribution function F, if $F_n \to F$ pointwise in $\mathbb{R}$. We denote this by $X_n \xrightarrow{\mathrm{d}} X$.

THEOREM 19 (Weak Law of Large Numbers). *Let $\{X_i\}$ be a sequence of independent, identically distributed random variables with finite means μ. Let $S_n \doteq X_1 + \cdots X_n$. Then*

$$\frac{S_n}{n} \xrightarrow{d} \mu.$$

PROOF. By theorem 17, it suffices to show that the characteristic function of S_n/n converges pointwise to $e^{it\mu}$, the characteristic function of μ (and clearly continuous at $t = 0$). Observe that

$$\begin{aligned}\phi_n(t) = \phi_{X_1/n+\ldots+X_n/n}(t) = \prod_{1\le j\le n} \phi_{X_j/n}(t) &= \left[\phi_{X_1/n}(t)\right]^n \\ &= \left[\phi_{X_1}(t/n)\right]^n\end{aligned}$$

where the two steps follow from the identical distribution of the X_i, and the identity

$$\phi_{X_1/n}(t) = \mathbb{E}(e^{iX_1t/n}) = \phi_{X_1}(t/n),$$

respectively. A Maclaurin series expansion gives

$$\begin{aligned}\phi_{X_1}(t/n) &= \phi_{X_1}(0) + \phi'_{X_1}(0)(t/n) + o(t/n) \\ &= 1 + (i\mu t)/n + o(t/n)\end{aligned}$$

and so

$$\lim_{n\to\infty} \left[\phi_{X_1}(t/n)\right]^n = \lim_{n\to\infty} \left[1 + (i\mu t)/n\right]^n = e^{i\mu t}$$

completing the proof. □

THEOREM 20 (Central Limit Theorem). *Let $\{X_i\}$ be independent, identically distributed random variables with means μ and nonzero variance σ^2. Let $S_n \doteq X_1 + \cdots + X_n$. Then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1).$$

PROOF. By theorem 17, it suffices to show that the characteristic function of $S_n/\sqrt{n\sigma^2}$ converges pointwise to $e^{-t^2/2}$, the characteristic function of a standard normal random variable (and clearly continuous at $t = 0$). Let $Y_i = (X_i - \mu)/n$. Then the Y_i have mean 0 and variance 1. Furthermore,

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{X_1 - \mu + \cdots + X_n - \mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\left[Y_1 + \cdots + Y_n\right].$$

Then

$$\phi_{\frac{1}{\sqrt{n}}\sum Y_i}(t) = \phi_{\sum Y_i} = \left[\phi_{Y_1}\left(t/\sqrt{n}\right)\right]^n$$

where the last step follows from the fact that the Y_i are identically distributed. Next, observe that

$$\begin{aligned}\phi'_{Y_1}(0) &= \mathbb{E}(-iY_1e^{-itY_1})|_{t=0} = -i\,\mathbb{E}(Y_1) = 0 \\ \phi''_{Y_1}(0) &= \mathbb{E}(Y_1^2e^{-itY_1})|_{t=0} = \mathbb{E}(Y_1^2) = \mathrm{Var}(Y_1) = 1.\end{aligned}$$

Hence, a Maclaurin expansion gives

$$\phi_{Y_1}(t) = 1 - t^2/2n + o(t^2/n)$$

and so

$$\lim_{n\to\infty} \phi_{\frac{1}{\sqrt{n}}\sum Y_i} = \lim_{n\to\infty} \left[1 - t^2/2n\right]^n = e^{-t^2/2}$$

which completes the proof. □

Part 2

Finance

CHAPTER 8

Risk

Definition. A *riskless asset* is an asset whose future value can be precisely determined. If an asset is not riskless, we say it is a *risky asset*.

The idea of a riskless asset is a theoretical construction. No asset in the world is without some risk. However, the primary example of a (nearly) riskless asset is a bond of a stable government. Examples of risky assets include stock in a limited liability company (especially those on the NASDAQ or outside the Fortune 500), and real estate.

Definition (Axioms of Market Efficiency). In an efficient market, we have the following:

(i) All available information about an asset is contained in its price. In other words, the price of an asset reflects its risk.
(ii) The past price of a stock has no influence on its present price. More precisely, stock movement is a martingale.

We will assume these axioms throughout the course, and add to them as we develop more complex models. Of course, the two axioms above are not true in general, otherwise there would be no bargains in the market, and investors like Warren Buffett would not have had nearly as much success as they have had. However, they do hold for the average investor in general. More precisely, by the time the average investor hears about a deal, the deal vanishes (that is, quicker, savvier investors like Buffett swoop in before them), which in turn makes a deal less of a bargain (demand for it raises the price).

Rephrased differently, our competition with each other allows stock prices to stabilize. Observe further that the riskiness of a stock is related to how high its price is. The more expensive a stock, the more you have to lose if you a purchase a share and the company goes under.

Since the two axioms hold for the overwhelming majority of investors, of what use is finance to them?

Remark. The purpose of finance is not so much to beat the market, but to relate the prices of assets to each other, in order to reduce our exposure to risk. This is known as *hedging*, and will be discussed in detail in the upcoming lectures.

Example. Generally, when the price of pork goes up, the price of beef goes down, since they are goods that can be substituted for one another. If I have a portfolio consisting of pork, I am in danger of losing a considerable amount of money if the price of pork goes down. To hedge against this, I also purchase beef. Now, if the price of pork goes down, my gains from my beef investment will offset them. However, if the price of pork goes up, I will lose money on my beef investment. What is of note is that the potential "swing", or variance, around my expected gains from my portfolio has been diminished by *hedging* our pork purchase with a beef purchase.

1. Introduction to Hedging and Risk Diversification

Suppose there is a contract on the market, which guarantees 100 is we flip heads with a fair coin, and 0 otherwise. We would like to find what the fair price is for the contract.

Our expected winnings is 50 dollars. However, people are naturally *risk averse*, especially when large sums of money are involved. Hence, we are willing to pay $\leq \$50$. Now suppose another contract exists that pays out 100 if we flip tails, and 0 otherwise. Since both contracts above exist on the market, we can buy both and be guaranteed 100. Hence, each contract must be worth 50 dollars *when the other is available.* The ability to hedge has eliminated the effects of our natural risk aversion!.

A closely related idea is *diversification.*

Example. Suppose now that there are N independent coin flips X_i, each paying $100/N$ for heads, and 0 for tails. Let $X \doteq X_1 + X_2 + \cdots + X_n$. Then $\mathbb{E}(X) = 50$, as in the single coin example above. However

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}([X-50]^2) \\
&= \mathbb{E}(X^2) - 2500 \\
&= \sum_{i,j:i\neq j} \mathbb{E}(X_i)\,\mathbb{E}(X_j) + \sum_{i=1}^{N} \mathbb{E}(X_i^2) - 2500 \\
&= \sum_{i,j:i\neq j} 2500/N^2 + \sum_{i=1}^{N} 5000/N^2 - 2500 \\
&= (N^2 - N)(2500/N^2) + 5000/N - 2500 \\
&= 2500/N
\end{aligned}$$

and so

$$\sigma(X) = 50/\sqrt{N} \to 0.$$

Observe, that for the single coin case, $N = 1$ and $\sigma(X) = 50$. Hence, spreading one's investments around, or *diversifying* allows them to control variance and, hence, control risk. However, a consequence of our analysis is that the more diversified a portfolio, the less chance that one will make a great deal of money (i.e. deviate far from the mean).

2. Assets

We now give a brief overview of some of the most common assets.

Stocks. Most shares traded on the market are from *limited liability* companies. What this means is that if the company goes bankrupt, the investor loses their investment, but can't be sued if the company is sued. That is, limited liability stocks can never have negative value.

The investor makes money from the stock via both dividends paid out, and by the price of the share on the market.

Bonds. The most typical is a 30 year bond, with coupons (i.e. interest payments) issued every year. The principal is returned at the end of the 30 years. They are not entirely riskless—interest rates may change, diminishing the value of the bond on the open market. This is reflected by the *yield* of the bond, where we define

$$\text{yield} = \frac{\text{interest payment}}{\text{price of bond}}.$$

Since bonds from stable governments are close to riskless, their yield is typically low (around 3%). Bonds from unstable governments are very risky (good luck getting General Sibanda to return your principal once he takes over Zimbabwe), and so have high yields.

Corporate Bonds. These are loans to a company, and their riskiness varies, depending on the stability of the company.

Derivatives. These are any instrument whose value is determined by the value of another asset, commonly called the *underlying*. At their core, derivatives are instruments that allow us to offset the risk of an asset by adopting the "opposite" risk.

Example. Suppose a firm will be paid in yen one year into the future,and will want to exchange the yen into dollars immediately upon payment. If the exchange rate dips in one year, the firm will lose money. To lessen the risk of this occurring, the firm can purchase a *forward contract*. That is, it can agree today to a fixed exchange rate for a fixed amount of yen one year into the future. Observe that if the exchange rate is greater one year from now than that fixed in the contract, the firm loses potential profits. However, if the exchange rate is less one year from now than that in the contract, the firm has guarded itself from great losses.

If the amount of yen paid at the end of the year will be highly variable, the firm can enter into an *option*, which is a contract to exchange *any* quantity of yen at a fixed interest rate, in the future. However, unlike forwards, options are *not-binding*. That is, the firm would have the option to not exercise the contract if it is not variable to a year from now.

Definition. A *call option* is a contract whose purchaser has the option to buy an asset at an agreed upon rate in the future. A *put option* is the a contract whose purchaser has the option to sell an asset at an agreed upon rate in the future. A *vanilla option* is either an *American option* or *European option*—an American option can be exercised any day before an agreed upon day in the future, while a European option can only be exercised on one agreed upon day in the future.

The pricing of options is a difficult problem, which we shall tackle in future lectures.

CHAPTER 9

Arbitrage

We motivate our discussion of *arbitrage* via an example.

Example. Suppose 1 British pound is worth 1.5 dollars, 1 dollar is worth 100 yen, and 1 pound is worth 149 yen. We now construct a strategy to make a profit *without risk of losses* as follows: We borrow 149 yen, immediately exchange it for 1 pound, which we exchange for 1.5 dollars. Then we exchange our 1.5 dollars for 150 yen, which we use to pay back our debt of 149 yen. All these trades are executed instantaneously, allowing us to avoid interest costs. Therefore, our our net profit is $150 - 149 = 1$. This type of investment opportunity is known as an *arbitrage opportunity.*

We remark that arbitrage opportunities do exist in the market, but they don't exist for long, due to their extreme desirability. That is, once enough people find out about the opportunity in the example, it raises the number of people demanding to exchange yen for pounds, lowering the value of yen relative to pounds until a $150 : 1$ yen to pound ratio is reached, upon which the arbitrage opportunity disappears.

Since arbitrage opportunities are rare and vanish quickly, we will assume absence of arbitrage in our models. We shall also adopt a number of others axioms, extending our previous framework.

Definition (Axioms of Market Efficiency, Updated). In an efficient market, we have the following:

(i) All available information about an asset is contained in its price. In other words, the price of an asset reflects its risk.
(ii) The past price of a stock has no influence on its present price. More precisely, stock movement is a martingale.
(iii) No market moving. That is, one can buy and sell any amount of an asset without affecting its price. Many prop shops attempt to violate this by artificially moving prices, thereby constructing an arbitrage opportunity. This has adverse effects for the economy and market efficiency.
(iv) Liquidity. That is, we can buy or sell any amount of an asset. Observe that Wall St. provides this. Indeed, any stable stock exchange does.
(v) We are allowed to short assets.
(vi) Fractional quantities of shares can be purchased. This is motivated by the idea that if there are 10^6 shares of a particular stock available, then $1/10^6 \sim 0.5/10^6$.
(vii) No transaction costs, commonly known as the *bid-offer spread.* This is of course false in the real world, but we assume it to simplify our models. Observe that our assumption introduces a negligible error if the transaction cost is small relative to the value of the asset being exchanged.

1. Mathematical Definition of Arbitrage, and its Corollaries

DEFINITION. A portfolio is said to be an *arbitrage portfolio* if today it is of x value, and in the future it has zero probability of being $< x$ in value, and nonzero probability of being $> x$ in value.

THEOREM 21 (Monotonicity Theorem). *Consider two portfolios A, B with values $A(t)$ and $B(t)$, respectively. If $A(T) \geq B(T)$ in every state of the world at time T, then $A(t) \geq B(t)$ for every $t < T$. In addition, if we know that $A(T) > B(T)$ in some states of the world at time T, then $A(t) > B(t)$ for all $t < T$.*

PROOF. Consider portfolio C constructed by going long A, and short B. Then $C(T) = A(T) - B(T) \geq 0$ in all states of the world at time T. By the no-arbitrage principle, we must have $C(t) \geq 0$. □

Corollary 22. *A vanilla call option or put option is always of positive value before expiry.*

PROOF. Let A be a portfolio that is long a call option, and B be an empty portfolio. Then B has zero value for all time, while $A(T) = (S_T - K)_+ \geq 0$ in all states of the market at time T. Hence, $A(T) \geq B(T)$ in all states, and hence $A(t) \geq B(t)$ by monotonicity. □

Corollary 23 (Weighted Symmetry of Randomness)**.** *If two portfolios P and Q are equal at some $t < T$ and $P(T) > Q(T)$ in some world states, then $Q(T) > P(T)$ in others.*

PROOF. Assume $P(T) > Q(T)$ in some world states evolving from time t, where $P(t) = Q(t)$. We proceed by contradiction. Suppose $P(T) < Q(T)$ does not hold in any world state. Then $P(T) \geq Q(T)$ in all world states. By monotonicity, $P(t) > Q(t)$, which is a contradiction. □

2. Introduction to Arbitrage-Free Pricing

2.1. Forward Exchange Rates. Suppose a firm wants to enter a contract to exchange up to 1 dollar for yen a year from now at an exchange rate K'. We wish to compute a non-arbitrage inducing exchange rate K' for the contract. Assume we start off with net worth 0 dollars. Recall that we assume that there are no transaction costs, so we are allowed to enter the forward without a fee. Next, observe that we can hedge a long position in the forward by shorting $1/(1+r)$ in bonds, which we then exchange at today's exchange rate K for yen.[1]

We then invest this in the Japanese bond market, which accrues interest at rate d. We then must exchange this back for dollars at year end at the forward exchange rate $1/K'$, and we use the proceeds to pay back our debt from shorting, which is now 1. Hence, in all possible states of the world in one year, our net worth in dollars will be

$$\frac{K}{K'}\left(\frac{1+d}{1+r}\right) - 1.$$

Since we started with 0 dollars, we must end up with 0 dollars, otherwise an arbitrage opportunity is present. Therefore,

$$\frac{K}{K'}\left(\frac{1+d}{1+r}\right) - 1 = 0$$

[1]The way to find the appropriate short is to ask yourself: If the market exchange rate in a year's time is bigger than K', which currency has benefited? Why? We lose out on having agreed to a forward exchange rate K, but can benefit by having invested in the currency that weakened over the course of the year.

which gives[2]

$$K' = K\left(\frac{1+d}{1+r}\right). \tag{2.1}$$

Exercise. Show that our exchange of exactly 1, and our strategy to short $1/(1+r)$ is somewhat artificial: we can short $a/(1+r)$, where $0 < a < 1$, and obtain the same value for K'.

Hence, we have shown that going long a forward contract is equivalent to the bond shorting strategy we have outlined above. The strategy of pricing an asset by decomposing it into instruments whose price can be computed is known as *replication*.

2.2. Options. Recall that a European call option grants the owner the right, but not the obligation, to purchase an asset at price K at some point in the future, but not before. Let S be the value of the underlying at expiry. Then at expiry, the *pay-off* of the option is precisely

$$(S-K)_+ = \max(S-K, 0)$$

Similarly, for a put option we have pay-off

$$(K-S)_+ = \max(K-S, 0)$$

Theorem 24 (Put-call parity). *If a call option, with price C, and a put option, with price P, and a forward contract, of price F, have the same strike and expiry, then*

$$C - P = F$$

Proof. Observe that if we adopt a portfolio long a call option, and short a put, both with strike price K, then the pay off of the portfolio at expiry, *in all states of the world*, is $S - K - C + P$, where C, and P are the prices of the call and put options at starting time $t = 0$, respectively. If we go long a forward contract with strike price K and cost F, this will have pay-off $S - K - F$ at expiry in all states of the world. Next, observe that, at expiry, $(S-K-F)-(S-K-C+P) = C-P-F$. If $C-P-F \geq 0$, then by the monotonicity theorem, a portfolio long a call option, short a put, and short a forward has value ≥ 0 in all states. If $C - P - F > 0$, we have an arbitrage opportunity; that is, we can short a call-put portfolio, invest the proceeds in a forward, and always be guaranteed a profit. Similarly, if $C - P - F < 0$, then we can short a forward, go long a call-put portfolio, and always be guaranteed a profit. Hence, by the no arbitrage principle, we conclude that we must have $C - P - F = 0$, or $C - P = F$. □

Theorem 25 (Bounds on Option Prices). *Let C_t be the price of a call option on a non-dividend paying stock, S_t, with expiry T and strike K. Let Z_t be the price of a zero-coupon bond with maturity T, at which time it is worth \$1. Then we have*

$$S_t > C_t > S_t - KZ_t.$$

In particular, if the interest rate is non-negative, then $C_t > S_t - K$ for $t < T$.

[2]Observe that we did not need to even consider stocks and derivatives in our hedging argument. Perhaps their existence would create an even bigger arbitrage than the one we obtain above if (2.1) does not hold. However, this is irrelevant—the existence of bonds and forwards in a market is enough to force (2.1) to hold.

Proof. At expiry the stock is worth S_T and the option is worth $\max(S_T - K, 0)$. Hence, the stock is always worth at least as much as the option in all world states at time T. By the monotonicity theorem, it follows that $S_t > C_t$.

To establish the lower bound, consider the portfolio long one option and K zero-coupon bonds which mature at time T. Then at expiry, our portfolio is worth

$$(S_T - K)_+ + K = \max(S_T, K) \geq S_T.$$

Observe that our analysis implies that in some world states at time T, our portfolio is worth as much as the stock, and in others it is worth more. By the monotonicity theorem, this is true for all $t < T$, and so

$$C_t + KZ_t > S_t$$

or

$$C_t > S_t - KZ_t$$

completing the proof. □

Remark. In particular, if the interest rate is non-negative, then $Z_t < 1$ for any $t < T$, and so $C_t > S_t - KZ_t > S_t - K$. This implies that, before expiry, an option is always worth more than the value that would be obtained by exercising it today.

This remark, in conjunction with the above theorem, implies the following.

Corollary 26. *If interest rates are non-negative, then a European call option and an American call option with the same non-dividend paying underlying stock with the same strike and expiry, are of equal value.*

Proof. Recall that an American option and a European option are nearly identical, with the caveat that an American option can be exercised any time before expiry. Hence, we are paying an additional premium to have this right. However, by the previous theorem, the American option is worth the most at expiry. Hence, it is only exercised at expiry. Therefore, the right to exercise the American option at any time before T has no value, and so the premium must be zero. Hence, the American option and European option have the same value at all times. □

Definition. Let S be a stock that is fixed in price. We say an option $C = C(K, T, S, t)$ is *time-homogeneous* if $C(K, T, S, t) = C(K, T - t, S, 0)$.

Intuitively, time-homogeneity tells us that, as long as the market remains stable, the cost of an option is influenced by its duration, and not the point in time which we evaluate it. As an example, in terms of price it makes no difference if we evaluate an option starting in January with expiry in July, or an identical option starting in July with expiry in December. The prices will be the same.

Theorem 27. *Let $C(K, T, S(t), t)$ denote a call option on an underlying S, with strike K and expiry T. Then*

(i) $C(K) = C(K, T, S(t), t)$ is a strictly decreasing, strictly convex, Lipschitz continuous function of K, with Lipschitz constant $Z(t)$, where Z is a non-coupon bearing bond with

$Z(T) = 1$. *Furthermore,*

$$-Z(t) < \frac{\partial C}{\partial K}(K) < 0 \text{ if } \frac{\partial C}{\partial K}(K) \text{ exists,}$$
$$\frac{\partial^2 C}{\partial k^2}(K) > 0 \text{ if } \frac{\partial^2 C}{\partial k^2}(K) \text{ exists.}$$

(ii) $C(T) = C(K, T, S(t), t)$ *is a strictly increasing function of* T *when interest rates are non-negative.*

(iii) $C(t) = C(K, T, S(t), t)$ *is a strictly decreasing function of* t*. Furthermore,*

$$\frac{\partial C}{\partial t}(t) < 0 \text{ if } \frac{\partial C}{\partial t}(t) \text{ exists.}$$

(iv) $C(S) = C(K, T, S(t), t)$ *is a strictly increasing, strictly concave, Lipschitz continuous function of* S*. Furthermore,*

$$\frac{\partial C}{\partial S}(S) < 0 \text{ if } \frac{\partial C}{\partial S}(S) \text{ exists,}$$
$$\frac{\partial^2 C}{\partial S^2}(S) < 0 \text{ if } \frac{\partial^2 C}{\partial S^2}(S) \text{ exists.}$$

Proof of item i. Assume $K_2 > K_1$. Then $C(K_2, T, S(t), t) < C(K_1, T, S(t), t)$ in all states. Hence, by monotonicity, $C(K_2, t) < C(K_1, t)$, and so C is strictly decreasing with respect to K. Next, observe that

$$\begin{aligned} C(K_2) + (K_2 - K_1)Z(T) &= (K_2 - K_1) + [S(T) - K_2]_+ \\ &\geq [S(T) - K_1]_+. \end{aligned}$$

Hence, by monotonicity,

$$C(K_2, T, S(t), t) + (K_2 - K_1)Z(t) \geq C(K_1, T, S(t), t)$$

or

$$C(K_1, T, S(t), t) - C(K_2, T, S(T), t) \leq (K_2 - K_1)Z(t).$$

This establishes the Lipschitz continuity. Lastly, dividing both sides by $(K_2 - K_1)$ and invoking the mean-value theorem gives the desired bounds on $\partial C/\partial K$, if it exists. To establish the convexity, observe that applying the identity $(A + B)_+ \leq A_+ + B_+$ gives

$$\begin{aligned} &\theta(S_T - K_2)_+ + (1-\theta)(S_T - K_1) - (S_T - [\theta K_2 + (1-\theta)K_1])_+ \\ &= \theta(S_T - K_2)_+ + (1-\theta)(S_T - K_1)_+ - (S_T - [\theta K_2 + (1-\theta)K_1])_+ \\ &\geq [\theta(S_T - K_2) + (1-\theta)(S_T - K_1)]_+ - (S_T - [\theta K_2 + (1-\theta)K_1])_+ \\ &= [S_T - [\theta K_2 + (1-\theta)K_1]]_+ - (S_T - [\theta K_2 + (1-\theta)K_1])_+ \\ &= 0. \end{aligned}$$

Observe that the lower bound established above is *strict* in some states of the world. Hence, by monotonicity

$$\theta C(K_2, t) + (1-\theta)C(K_1) > C(\theta K_2 + (1-\theta)K_1).$$

Lastly, if $\partial^2 C/\partial K^2$ exists, then it is positive by the second derivative test. □

Proof of item ii. Let $T_1 < T_2$, fix $t \leq T_1$, and adopt the notation $C_1 = C(K, T_1, S(t), t)$, $C_2 = C(K, T_2, S(t), t)$. By corollary 26, C_1 and C_2 are each equal in value to an American option with the same underlying, strike, and expiry. Hence, without loss of generality we may treat C_2 as an American option. Next, observe that $C_2(T_1) > C_1(T_1) = \max\{S(T_1) - K, 0\}$ by theorem 25 and the remark following it. By monotonicity, $C_2(t) > C_1(t)$, completing the proof. □

Proof of item iii. By time homogeneity and item ii, respectively

$$C(K, T, S(0), t) = C(K, T - t, S(t), 0) < C(K, T, S(t), 0)$$

completing the proof. □

Proof of item iv. The argument is analogous to that of item i. We shall leave the proof of concavity to the reader, and tackle the remainder.

Proceeding, let $S_1 < S_2$ be two fixed stock prices. Then $C(K, T, S, T) < C(K, T, S_2, T)$, so by monotonicity $C(K, T, S, t) < C(K, T, S_2, t)$ for all $t < T$. Hence, $C(S)$ is a strictly decreasing function of S. Next, observe that

$$\begin{aligned} C(S_1) + (S_2 - S_1)Z(T) &= [S_1 - K]_+ + S_2 - S_1 \\ &\geq [S_2 - K]_+ \\ &= C(S_2) \end{aligned}$$

so by monotonicity

$$C(K, T, S_1, t) + (S_2 - S_1)Z(t) \geq C(K, T, S_2, t)$$

or

$$C(K, T, S_2, t) - C(k, T, S_1, t) \leq (S_2 - S_1)Z(t).$$

This establishes the Lipschitz continuity. Lastly, dividing both sides by $(S_2 - S_1)$ and invoking the mean-value theorem gives the desired bounds on $\partial C/\partial S$, if it exists. □

CHAPTER 10

Pricing the Multi-step, two State, zero interest Option

In order to accurately price an option, our model must take into account the infinitely many states the future can take. To simplify our discussion, we shall first restrict our model to futures with finitely many states, and zero-interest.

Proceeding, suppose we wish to price an option today with an underlying worth 100, and expiry one day into the future. Furthermore, suppose that we know that the underlying will tomorrow be worth 110 with probability p, and 90 tomorrow with probability $1 - p$, where $p \neq 0, 1$.

REMARK. Observe that if $p = 1$, then we have an arbitrage opportunity, since we can go short 100 worth in bonds, purchase the stock at 100, which then will increase to 110, after which we can pay back our bond debt and be left with a profit of 10, risk free. A similar arbitrage exists for the case when $p = 0$.

1. The Arbitrage-free Approach

Consider a portfolio P that is long δ shares of the underlying, and short the option. Then

$$P(0) = \delta S(0) + C_0 - C(0)$$

$$P(T) = \begin{cases} 110\delta + C_0 - 10, & \text{in some states} \\ 90\delta + C_0, & \text{in others.} \end{cases}$$

Observe that if $P(T)$ were the same in all states, monotonicity would force $P(t) = P(T)$ for all $t \leq T$, and hence, eliminate risk. For this to occur, we must have

$$110\delta + C_0 - 10 = 90\delta + C_0$$

which gives $\delta = 1/2$. Hence, for $\delta = 1/2$, we have $P(0) = P(T)$, which by substitution gives us

$$\frac{100}{2} = \frac{90}{2} + C_0$$

or $C_0 = 5$. Arbitrage considerations have once again allowed us to price an option. Observe furthermore that we were able to price the option independently of the probability $0 < p < 1$. Hence, it is reasonable to ask whether we can choose a probability $0 < p' < 1$ that somehow simplifies our computations.

2. The Risk Neutral Approach

From our arbitrage considerations, we see that, for suitably chosen δ, we had $P(0) = P(T)$. Then of course, for this δ, we must have $\mathbb{E}(P(0)) = \mathbb{E}(P(T))$. Let us weaken our assumptions a little by allowing arbitrary δ, while still assuming $\mathbb{E}(P(0)) = \mathbb{E}(P(T))$. Now, $\mathbb{E}(P(0)) = P(0) = 100\delta$, so

$$(110\delta + C_0 - 10)p + (90\delta + C_0)(1-p) = 100\delta$$

which reduces to

$$20\delta p - 10p + C_0 = 10\delta$$

Observe that for $p = 1/2$, the dependence on δ vanishes, and we obtain $C_0 = 5$. We call $p = 1/2$ the *risk-neutral* probability.

Does such a probability exist in the general case?

THEOREM 28. *Consider a two state, no interest world for an asset S. Suppose $S(0) = s$ and*

$$S(T) = \begin{cases} s+u, & \text{with probability } p \\ s-d, & \text{with probability } 1-p \end{cases}$$

where $u = d \geq 0$. Suppose the option is worth α in the "up" state and β in the "down" state. Then the price of a call option $C(0)$ with underlying S and expiry T is equal to the expected price of the option at time T with $p = 1/2$. That is,

$$C(0) = 1/2(\alpha + \beta).$$

PROOF. Construct a portfolio by buying δ shares of S and shorting a call option with underlying S. Then, to obtain a portfolio without risk, we must find δ such that our portfolio has the same value in either the "up" or "down" states. Hence, we set

$$s + u\delta - \alpha + C_0 = (s-d)\delta - \beta + C_0$$

which gives

$$\delta = \frac{\alpha - \beta}{u + d}$$

By monotonicity, we must have $P(0) = P(T)$, and so

$$\frac{s(\alpha-\beta)}{u+d} = (s+u)\delta - \alpha + C_0$$
$$\frac{s(\alpha-\beta)}{u+d} = (s-d)\delta - \beta + C_0.$$

Summing these two equations, substituting in δ, and simplifying gives

$$C_0 = \frac{1}{2}\left[\frac{\alpha-\beta}{u+d}(d-u)\right] + \frac{1}{2}(\alpha+\beta)$$
$$= \frac{1}{2}(\alpha+\beta) \tag{2.1}$$

which gives $C_0 = 1/2(\alpha + \beta)$ for $u = d$, completing the proof. [1] □

[1] Observe that if $u \neq d$, we can obtain still obtain a risk-neutral p' by setting the right-hand-side of (2.1) equal to $p'\alpha + (1-p')\beta$ and solving for p'.

We now generalize our arguments. As before, we assume zero-interest. Let an underlying S be worth 100 today ($t = 0$), with up and down states of ± 5 tomorrow ($t = 1$), respectively. We let each of these states, in turn, have up and down states of ± 5 the following day. Our option will have a strike price of 100 and expiry $T = 2$ two days from now.

We can see this two step model as a one step model: in the up and down states tomorrow, the option will have some value. Applying theorem 28, we can then take the risk-neutral expectation over these two values to obtain the cost of the option today. That is

$$C_{105}(0) = \mathbb{E}(C_{105}(1)) = \frac{1}{2}(110 - 100)_+ + \frac{1}{2}(100 - 100)_+ = 5$$
$$C_{95}(0) = \mathbb{E}(C_{95}(1)) = \frac{1}{2}(100 - 100)_+ + \frac{1}{2}(90 - 100)_+ = 0$$

and so

$$\begin{aligned} C_{100}(0) = \mathbb{E}(C_{100}(1)) &= \frac{1}{2}C_{105}(0) + \frac{1}{2}C_{95}(0) \\ &= 2.5. \end{aligned}$$

Note that this is nothing but the expected value of $C(T)$ over all possible two-step paths; that is

$$2.5 = C_{100}(0) = \frac{(110 - 100)_+ + 2(100 - 100)_+ + (90 - 100)_+}{4}$$

In general, let C be an option with underlying S, whose price follows a tree with N steps, with a choice of two directions for each step ("up" and "down"). Since the price of the option depends solely on the price of the underlying, we may adopt the notation $C(t) = f(S(t))$. Then an induction shows that

$$\begin{aligned} C_0 = \mathbb{E}(C(T)) &= \sum_{i=0}^{j} \mathbb{P}(S = S_0 + (N-j)u - jd) f(S_0 + (N-j)u - jd) \\ &= \sum_{j=0}^{N} \mathbb{P}(S = S_0 + (N-2j)u) f(S_0 + (N-2j)u), \quad u = d \end{aligned}$$

Next, observe that there are $\binom{N}{j}$ ways to arrange j up movements and $N-j$ down movements. Each arrangement corresponds to a different path; however, each path with j up movements and $N - j$ down movements leads to the same final node. Hence, our preceding computation simplifies to

$$C_0 = \frac{1}{2N} \sum_{j=0}^{N} \binom{N}{j} f(S_0 + (N-2j)u). \tag{2.2}$$

We will see in the next section that this can be written more concisely as or, written more concisely,

$$C_0 = E(f(S_0 + u\sum_{j=1}^{N})X_j) \tag{2.3}$$

where the $\{X_i\}$ are symmetric, independent random variables taking the values ± 1.

CHAPTER 11

The Continuous Time Black-Scholes Formula

1. Introduction

We now seek to price options in continuous time by taking $N \to \infty$ in the right-hand-side of (2.2).

We now generalize and clean up our construction in the previous section. We begin with the following assumptions:

(i) The underlying asset is today worth $S(0) = 0$.
(ii) Expiry is T.
(iii) $\mathbb{E}(S(T)) = S_0$.
(iv) $\text{Var}(S(T)) = \sigma^2 T$.
(v) The asset follows a *symmetric random walk*. That is, for some $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$, we have

$$S(t_{j+1}) = \begin{cases} S(t_j) + \alpha, & p = 1/2 \\ S(t_j) - \alpha, & 1 - p \end{cases}$$

Suppose it takes k steps until we reach the expiry T. Then each step can be modeled by mutually independent random variables Z_i, each taking the values $\pm\beta$ with equal probability. Furthermore,

$$\begin{aligned} \sigma^2 T = \text{Var}(S(T)) &= \text{Var}(Z_1 + Z_2 + \cdots + Z_k) \\ &= \text{Var}(Z_1) + \text{Var}(Z_2) + \cdots + \text{Var}(Z_k) \\ &= k\text{Var}(Z_i), \quad 1 \leq i \leq k \end{aligned}$$

and so

$$\text{Var}(Z_i) = \sigma\sqrt{T/k} \doteq \sigma_k, \quad 1 \leq i \leq k.$$

Now, since each Z_i has mean zero, we have

$$\begin{aligned} \text{Var}(Z_i) &= \frac{1}{2}\beta^2 + \frac{1}{2}(-\beta)^2 \\ &= \beta \end{aligned}$$

and hence

$$\beta = \sigma\sqrt{T/k}.$$

Letting

$$X_i = \begin{cases} +1, & p = 1/2 \\ -1, & p = 1/2 \end{cases}$$

we see that

$$S(T) = S_0 + \sum_{i=0}^{k} \sigma_k X_i.$$

Then for an option that pays $f(S(T))$ at time T, we have

$$f(S(0)) = \mathbb{E}(f(S(T))) = \mathbb{E}[f(S_0 + \sigma_k \sum_{i=1}^{k} X_i)].$$

By theorem 20,

$$S_0 + \sigma_k \sum_{i=1}^{k} X_i = S_0 + \sigma\sqrt{T} \times \frac{1}{\sqrt{k}} \sum_{i=1}^{k} X_i \xrightarrow{\text{dist}} S_0 + \sigma\sqrt{T} N(0,1).$$

and so our option price is given by

$$\begin{aligned} f(S(0)) &= \mathbb{E}(f(S_0 + \sigma_k \sum_{i=1}^{k} X_i)) \\ &\to \mathbb{E}(S_0 + \sigma\sqrt{T} N(0,1)) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(S_0 + \sigma\sqrt{T}x) e^{-x^2/2}\, dx. \end{aligned}$$

Unfortunately, our model is too simplistic, as it allows for stock prices to be negative, and has assumed interest to be zero.

2. Distribution of Stock Price With Zero Interest

Our one step, two state model before involved a stock price S_0 today increasing or decreasing to $S_0 \pm u$, respectively. In the multi-step extension, this has the drawback of allowing for negative stock prices. To remedy this, we alter our model by now having S_0 increase or decrease by a *percentage*. That is, either S_0 goes to $S_0 r_+$ or $S_0 r_-$, where $r_+ > 1$ and $0 < r_- < 1$. Extending this model to multiple steps, where at each step we increase or decrease by a factor of $r_\pm$, we see that the stock price is always bounded below by zero, as desired. However, to recover the benefits of the theory we developed for our previous model (where up states and down states were given by an addition of an increment, rather than a product), we consider the evolution of $\log(S(t))$. Then, by the product rule for logarithms, it follows that $\log(S(t)) \to \log(S(t)) + \log(r_\pm)$.

In general, for an n step symmetric model, adopting the notation $\Delta t = T/n$ for expiry T and initial time $t = 0$, we have

$$\log S_{j\Delta t} = \log S_{(j-1)\Delta t} + X_j$$

where

$$X_j = \begin{cases} \sigma\sqrt{\Delta t}, & p = 1/2 \\ -\sigma\sqrt{\Delta t}, & p = 1/2 \end{cases}$$

and the X_j, $1 \leq j \leq n$ are independent. Observe that, in this construction, $\Delta t \sigma^2$ is the variance, per unit time Δt (or, equivalently, per step). Since the X_j are all independent, the variance from $t = 0$ to $t = T$ of our random stock evolution will be $\sigma^2 T$. Next, letting

$$Y_j = \begin{cases} 1, & p = 1/2 \\ -1, & 1 - p = 1/2 \end{cases}$$

we see that

$$\log S_{j\Delta t} = \log S_{(j-1)\Delta t} + \sigma\sqrt{\Delta t}Y_j$$

Next, observe that

$$\sigma\sqrt{\Delta t}\sum_{j=1}^{n} Y_j = \sum_{j=1}^{n}(\log S_{j\Delta t} - \log S_{(j-1)\Delta t}) = \log S_T - \log S_0$$

which we rewrite as

$$\log S_t = \log S_0 + \sigma\sqrt{T} \times \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Y_j \xrightarrow{d} \log S_0 + \sigma\sqrt{T}N(0,1)$$

where the last step follows from the central-limit theorem. Hence,

$$S_T \stackrel{d}{=} S_0 e^{\sigma\sqrt{T}N(0,1)}.$$

We say that S_T is *log-normally distributed*, since its logarithm is normally distributed. Furthermore, observe that $S_T(\omega) \not< 0$ for all $\omega \in \Omega$.

3. Distribution of Stock Price with Positive Interest

We first restrict our attention to the one step, two state model, with initial stock price S_0 and future states $S_\pm$. Observe that the assumption of no-arbitrage implies $S_- \leq S_0 e^{r\Delta t} < S_+$. We shall now attempt to price a call option C with underlying S at time $t = 0$.

3.1. The No-Arbitrage Pricing Approach. Consider a portfolio P that is long δ shares of the underlying, and short the option. Then

$$P(0) = \delta S(0) \pm C_0$$

$$P(T) = \begin{cases} S_+\delta + C_0 - (S_+ - S_0), & \text{in some states} \\ S_-\delta + C_0, & \text{in others.} \end{cases}$$

Observe that if $P(T)$ were the same in all states, monotonicity would force $P(t) = P(T)$ for all $t \leq T$, and hence, eliminate risk. For this to occur, we set the up state and down state of our portfolio equal to one another, and obtain

$$\delta = \frac{S_+ - S_0}{S_+ - S_-}. \tag{3.1}$$

By monotonicity, our portfolio is worth δS_0 at all times. However, for no arbitrage to occur, our portfolio should be worth $\delta S_0 e^{r\Delta t}$. Hence, our analysis has been misguided.

Where have we gone wrong? Observe that our portfolio has no bond component, and hence fails to take account of the positive interest rate. To attempt to remedy this, we note that once C_0 is obtained from shorting a call option, it can be immediately invested in a bond, which will yield C_0e^{rt} in the future. That is

$$P(0) = \delta S(0) \pm C_0,$$

$$P(T) = \begin{cases} S_+\delta + C_0e^{rt} - (S_+ - S_0), & \text{in some states} \\ S_-\delta + C_0e^{rt}, & \text{in others.} \end{cases}$$

Setting the up state of our portfolio equal to the down, we obtain (3.1), which is the same δ we obtained in the "wrong" model. Applying monotonicity, we obtain that $P(0) = P(T)$, and so our portfolio value of δS_0 initially does not change. This again violates the no-arbitrage principle; that is

$$\delta S_0 < \delta S_0 e^{rT}$$

and so we could construct an arbitrage by shorting P and investing the proceeds in a bond.

We need to revise our model some more. The critical observation that our failed attempts have led us to is the following: if $P(0)$ is a risk neutral portfolio at time $t = 0$, then its value in both the up and down states must be equal to $e^{rT}P(0)$, in order to avoid arbitrage.

We formulate this rigorously by defining

$$P(0) = [\delta S_0 \pm C_0] - \delta S_0$$

$$P(T) = \begin{cases} S_+\delta + C_0e^{rt} - (S_+ - S_0) - e^{rT}\delta S_0, & \text{in some states} \\ S_-\delta + C_0e^{rt} - e^{rT}\delta S_0, & \text{in others.} \end{cases}$$

That is, P is the portfolio we obtain by borrowing δS_0 at r interest, and then investing it into the portfolio that is long δS_0 shares of stock, and short an option.

Setting the up and down states equal to one another, we obtain (3.1). Applying monotonicity, we see that $0 = P(0) = P(T)$, which gives

$$S_+\delta + C_0e^{rt} - (S_+ - S_0) = \delta S_- + C_0e^{rT} = e^{rT}\delta S_0 \tag{3.2}$$

and so

$$C_0 = e^{-rT}\delta(e^{rT}S_0 - S_-) \tag{3.3}$$

Note that since (3.2) holds, no arbitrage using bonds is possible. Contrast this with our previous, failed models.

3.2. The Risk Neutral Approach. Observe that for the no-arbitrage approach, we were able to find a risk-free δ by setting $P(T) = P(0)$ and solving.

To find the risk-neutral probability p, we adopt an analogous approach, and assume $\mathbb{E}(P(T)) = \mathbb{E}(P(0))$. Then

$$p(\delta S_+ + C_0e^{rT} - S_+ + S_0 - e^{rT}S_0) + (1-p)(\delta S_- + C_0e^{rT} - e^{rT}\delta S_0) = 0$$

which we rewrite as

$$\delta\,\mathbb{E}_p(S(T)) + C_0e^{rT} + p(S_0 - S_+) = e^{rT}\delta S_0. \tag{3.4}$$

Suppose we find p such that

$$\mathbb{E}_p(S(T)) = e^{rT} S_0. \tag{3.5}$$

Then (3.4) reduces to

$$C_0 e^{rT} + p(S_0 - S_+) = 0$$

Observe that we have eliminated the dependence on the value of δ entirely. Solving for C_0, we obtain

$$\begin{aligned} C_0 &= e^{-rT} p(S_+ - S_0) \\ &= e^{-rT} \left[p(S_+ - S_0)_+ + (1-p)(S_- - S_0)_+ \right] \\ &= e^{-rT}\, \mathbb{E}_p(f(S_T)) \end{aligned}$$

Lastly, we compute the *risk-neutral* probability p from expression (3.5), which gives

$$p(S_+) + (1-p)S_- = e^{rT} S_0$$

or

$$p = \frac{e^{rT} S_0 - S_-}{S_+ - S_-}. \tag{3.6}$$

We now generalize the risk-neutral approach to multiple timesteps. Suppose, starting from $t = 0$, we have N symmetric timesteps of length Δt, where $N\Delta t = T$. Using the risk-neutral option pricing results in the previous section, and running an induction, it follows that

$$C_0 = e^{-rT}\, \mathbb{E}_{\vec{p}}(f(S_T)) \tag{3.7}$$

where $\vec{p} \doteq \{p_1, p_2, \ldots, p_{2^N-1}\}$ is the vector of risk neutral probabilities at each step. Recall that, for zero interest, $p_i = 1/2$ for all i. Indeed, this is implied by formula (3.6). However, this formula also implies that for $r \neq 0$, the p_i change as $\Delta t \to 0$. Hence, computing $\mathbb{E}_{\vec{p}}(f(S_T))$ via a counting argument as in the zero-interest case is considerably more difficult, due to the ever-changing p_i as the timestep shrinks in size. Furthermore, each timestep size Δt results in a new $\vec{p}$. Since we are interested in taking $\Delta t \to 0$, we see that a counting argument, while possible, is extremely cumbersome. Instead, we will seek to find the distribution of S_T under the risk-neutral vector $\vec{p}$, as this will allow us to compute $E_{\vec{p}}(f(S_T))$ with relative ease.

Proceeding, we express a symmetric, multi-step model via the equation

$$\log S_j = \log S_{j-1} \pm \sigma\sqrt{\Delta t}$$

where we have adopted the notation $S_j \doteq S_{j\Delta t}$. This is equivalent to

$$S_j = S_{j-1} e^{\pm\sigma\sqrt{\Delta t}}.$$

From (3.6), we see that the risk-neutral probability at S_{j-1} is

$$\frac{S_{j-1}e^{r\Delta t} - S_{j-1}e^{-\sigma\sqrt{\Delta t}}}{S_{j-1}e^{\sigma\sqrt{\Delta t}} - S_{j-1}e^{-\sigma\sqrt{\Delta t}}}.$$

Cancelling S_{j-1}, we see that

$$p = \frac{e^{r\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}.$$

Hence, the risk-neutral probabilities at each timestep do not depend on the initial stock price at the timestep or on the time elapsed, but only on the timestep size. Hence, $\vec{p}$ has been simplified considerably, since we have shown that $p_i = p_j$.

Employing a Maclaurin series expansion of e^x, we obtain

$$p = \frac{1}{2}\left[1 + \left(\frac{r - 1/2\sigma^2}{\sigma}\right)\sqrt{\Delta t}\right] + O(\Delta t).$$

Hence, we rewrite our symmetric, multi-step model as the equivalent *risk-neutral*, multi-step model

$$\log S_j = \log S_{j-1} + \sigma\sqrt{T/N} Z_j$$

where the $Z_j, 1 \leq j \leq N$, are independent and

$$Z_j = \begin{cases} 1, & p = p(\Delta t) \\ -1, & 1 - p. \end{cases}$$

We would like to take the limit of this expression as $N \to \infty$, and apply the central limit theorem. We compute

$$\mathbb{E}(Z_j) = \left(\frac{r - \frac{1}{2}\sigma^2}{\sigma}\right)\sqrt{\Delta t} + O(\Delta t)$$

and

$$\begin{aligned} \operatorname{Var}(Z_j) &= \mathbb{E}(Z_j^2) - [\mathbb{E}(Z_j)]^2 \\ &= 1 - [\mathbb{E}(Z_j)]^2 \\ &= 1 - \left(\frac{r - \frac{1}{2}\sigma^2}{\sigma}\right)^2 \Delta t + O[(\Delta t)^{3/2}] \end{aligned}$$

and so

$$\begin{aligned} \operatorname{Var}(\sum_{j=1}^{N} Z_j) &= \sum_{j=1}^{N} [\operatorname{Var}(Z_j)] \\ &= N - N\left(\frac{r - \frac{1}{2}\sigma^2}{\sigma}\right)^2 T/N + NO[(T/N)^{3/2}] \\ &= N - \left(\frac{r - \frac{1}{2}\sigma^2}{\sigma}\right)^2 T + O[N^{-1/2}]. \end{aligned}$$

Hence,

$$\mathbb{E}(\sigma\sqrt{T/N}\sum_{j=1}^{N} Z_j) = \left(r - \frac{1}{2}\sigma^2\right) T + O(N^{-1/2})$$

$$\mathrm{Var}(\sigma\sqrt{T/N}\sum_{j=1}^{N} Z_j) = \sigma^2 T + O(N^{-1}).$$

Applyng the central limit theorem, we obtain

$$\begin{aligned}\sigma\sqrt{T/N}\sum_{j=1}^{N} Z_j &\xrightarrow{\mathrm{d}} N\left[\left(r - \frac{1}{2}\sigma^2\right) T, \sigma^2 T\right] \\ &= \left(r - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T} N(0,1).\end{aligned}$$

Observing

$$\sum_{j=1}^{N} (\log S_j - \log S_{j-1}) = S_T - S_0$$

we conclude that

$$\log S_T \stackrel{\mathrm{d}}{=} \log S_0 + \left(r - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{t} N(0,1)$$

from which we obtain[1]

$$S_T \stackrel{\mathrm{d}}{=} S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{t}N(0,1)}.$$

From (3.7), it follows that, for a call option C with strike price K

$$C_0 = e^{-rT}\,\mathbb{E}\left[(S_T - K)_+\right].$$

4. The Explicit Formula

It turns out the expectation in the previous section can be computed explicitly for our symmetric, risk-neutral model. First, observe that

$$\begin{aligned}\mathbb{P}(S_T \le s) &= \mathbb{P}\left(e^{\sigma\sqrt{T}N(0,1)} \le \frac{s}{S_0 e^{(r-\frac{1}{2}\sigma^2)T}}\right) \\ &= \mathbb{P}\left[N(0,1) \le \frac{1}{\sigma\sqrt{T}} \log\left(\frac{s}{S_0 e^{(r-\frac{1}{2}\sigma^2)T}}\right)\right] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\ell} e^{-x^2/2}\, dx\end{aligned}$$

where

$$\ell = \frac{1}{\sigma\sqrt{T}}\left[\log s - \log S_0 - (r - \frac{1}{2}\sigma^2)T\right].$$

[1]A simple computation shows that $\mathbb{E}_p(S_T) = S_0 e^{rT}$. That is, under the risk neutral probability p, our stock grows, on average, at the bond market rate. This is consonant with our model's built-in no-arbitrage assumptions.

Hence, the density function of S_T is given by[2]

$$g_{S_T}(s) = \frac{d}{ds}\,\mathbb{P}(S_T \le s) = \frac{1}{\sigma s\sqrt{2\pi T}} e^{-\left[\log s - \log S_0 - (r-\frac{1}{2}\sigma^2)T\right]^2/2\sigma^2 T}.$$

Therefore,

$$\mathbb{E}_{p^\star}(f(S_T)) = \frac{1}{\sigma\sqrt{2\pi T}} \int_{-\infty}^{\infty} (s-K)_+ e^{-\left[\log s - \log S_0 - (r-\frac{1}{2}\sigma^2)T\right]^2/2\sigma^2 T} \times \frac{1}{s}\, ds.$$

Since $(s-K)_+ = 0$ for $s < K$, the integral simplifies to

$$\frac{1}{\sigma\sqrt{2\pi T}} \int_{K}^{\infty} (s-K) e^{-\left[\log s - \log S_0 - (r-\frac{1}{2}\sigma^2)T\right]^2/2\sigma^2 T} \times \frac{1}{s}\, ds.$$

Using the substitution $\mu = \left[\log s - \log S_0 - (r - \frac{1}{2}\sigma^2)T\right]/\sigma\sqrt{T}$, this becomes

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}} \int_{\ell_K}^{\infty} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T + s\sigma\sqrt{T}} - K\right] e^{-s^2/2}\, ds \\
&= \frac{1}{\sqrt{2\pi}} \int_{\ell_K}^{\infty} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T + s\sigma\sqrt{T}}\right] e^{-s^2/2}\, ds - \frac{K}{\sqrt{2\pi}} \int_{\ell_K}^{\infty} e^{-s^2/2}\, ds \\
&\doteq \mathrm{I} - \mathrm{II}
\end{aligned}$$

where $\ell_K = \ell(K)$. Since the Gaussian is symmetric, we have

$$\mathrm{II} = KN(-\ell_K).$$

To evaluate integral I, we rewrite it and apply the change of variable $\mu = s - \sigma\sqrt{T}$. Thus

$$\begin{aligned}
\mathrm{I} &= \frac{1}{\sqrt{2\pi}} \int_{\ell_K}^{\infty} S_0 e^{rT} e^{-(s-\sigma\sqrt{T})^2/2}\, ds \\
&= \frac{S_0 e^{rT}}{\sqrt{2\pi}} \int_{\ell_K - \sigma\sqrt{T}}^{\infty} e^{-s^2/2}\, ds \\
&= S_0 e^{rT} N(-\ell_K + \sigma\sqrt{T})
\end{aligned}$$

where the last step follows from the symmetry of the Gaussian. Putting everything together, we obtain the *Black-Scholes Formula*

$$\begin{aligned}
C_0 &= e^{-rT}\,\mathbb{E}\left[(S_T - K)_+\right] \\
&= S_0 N(-\ell_K + \sigma\sqrt{T}) - Ke^{-rT} N(-\ell_K) \\
&= S_0 N(d_1) - Ke^{-rT} N(d_2)
\end{aligned}$$

where

$$\begin{aligned}
d_1 &= \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \\
d_2 &= \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.
\end{aligned}$$

[2]Observe that $g_{S_T}(s)$ is integrable in a neighborhood of 0, due to exponential decay near 0 dominating the polynomial growth near 0 of the $1/s$ term. To see this explicitly, use L'Hôpital's rule.

Bibliography