Contents

Remarks

These lecture notes are based on several textbooks. The probability part in the first few chapters follows Blitzstein & Hwang (2015). This is an excellent introduction into probability and covers many more topics than we are able to discuss here. We rely mostly on Chapters 1, 3, 4, 5, 7, and 10 in this book.

One good reference for random number generations and simulations is for example Ross (2006).

For the financial part involving the binomial tree model, we follow the book by Shreve (2004), one of the standard textbooks in this area. We rely mostly on Chapter 1.

We then study more advanced Monte Carlo techniques, in particular variance reduction techniques. Some material in the lecture notes is taken from or was inspired by the books by Ross (2006) and Glasserman (2004).

There are more advanced books available which cover related material far beyond what we are able to cover in this introductory course. See, for example, Glasserman (2004), Seydel (2009), and Asmussen & Glynn (2007).

For several concepts in this lecture notes, also Wikipedia turns out to be a good reference.

For Python, you can find much reading material online for free.

Course overview

This course consists of 12 lectures, each of 3-hour length. There are 12 additional classes to deepen the understanding of the course material and to practice programming in Python.

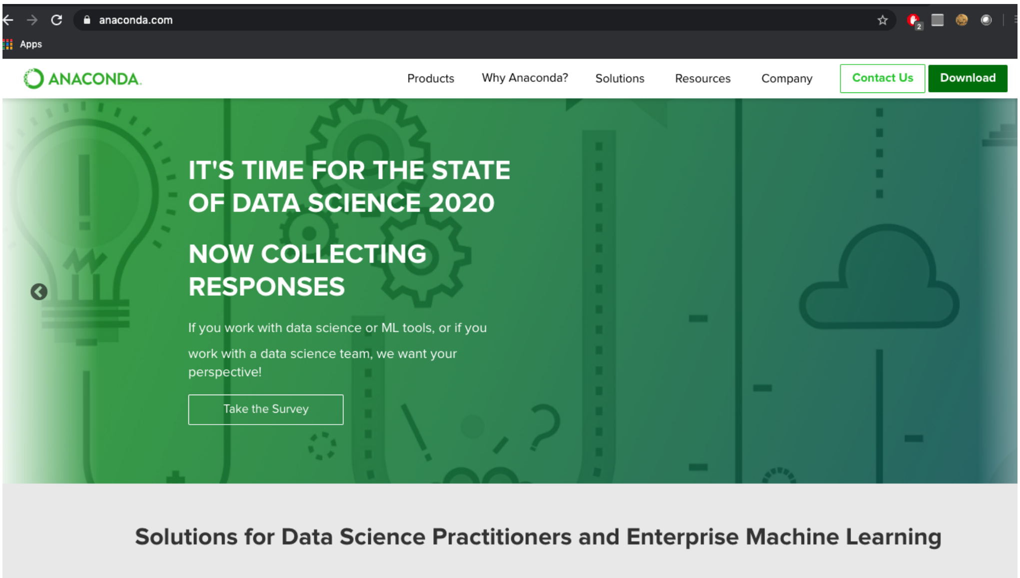
The first few chapters of the lecture notes serve as an introduction into probability. Probability is a very important field with many different applications, one of which is finance - the main application area studied in this course. Unfortunately, its concepts are often difficult to grasp and sometimes counterintuitive. For this reason, we spend a considerable amount of time on introducing these concepts. The programming language Python is introduced by implementing the concepts we have learned – learning by doing.

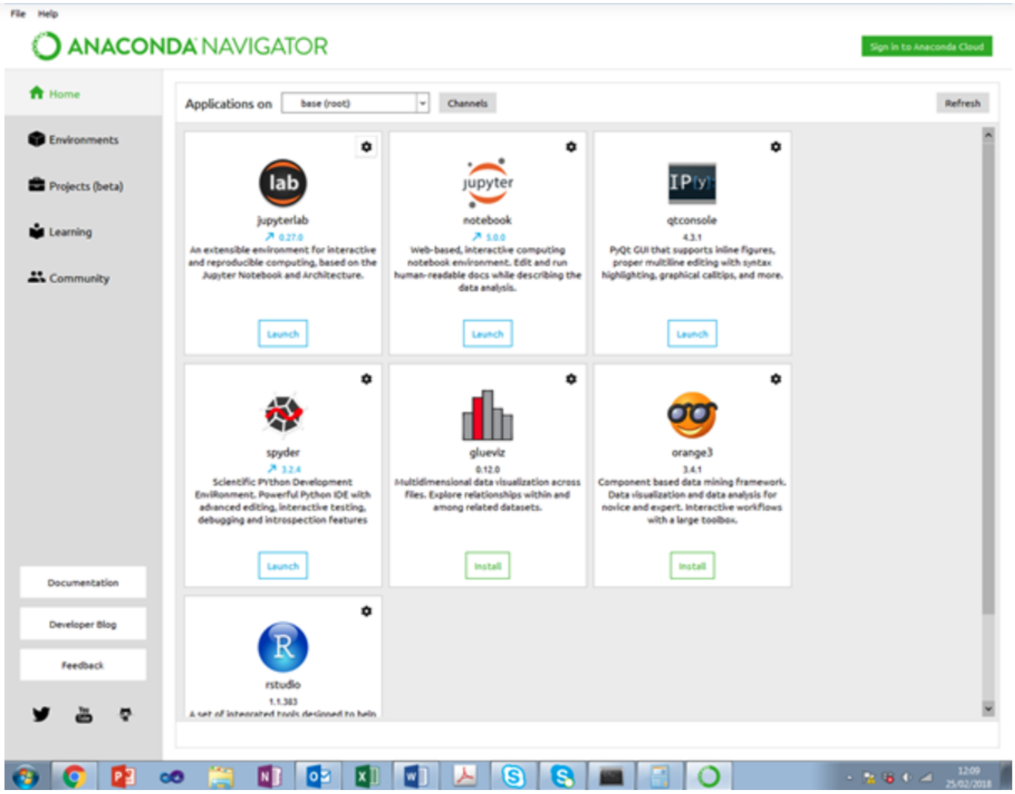
We then discuss how random numbers can be generated and how to do Monte Carlo simulations, which will be important tools to compute option prices in practice.

In the second part of the course, we concentrate on the application of the probabilistic methods introduced in finance. We first study the fundamental ideas used to price financial products by considering the well-studied binomial model for the financial market. We then introduce the Nobel-prize winning Black-Scholes formula for option pricing. Finally we show how the probabilistic methods introduced (Monte Carlo techniques, random number generation etc.) can be used to approximate prices of financial derivatives which cannot be computed analytically.

Anaconda and Jupyter notebooks

https://www.anaconda.com




File Help
ANACONDA NAVIGATOR
Sign in to Anaconda Cloud
Home
Environments
Projects (beta)
Learning
Community
Applications on
base (root)
Channels
Refresh
jupyterlab
0.27.0
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch
notebook
5.0.0
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch
qtconsole
4.3.1
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch
spyder
3.2.4
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch
glueviz
0.12.0
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install
orange3
3.4.1
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.
Install
rstudio
1.1.383
Documentation
Developer Blog
Feedback

Chapter 1

Probability

1.1 What is probability?

- Appears in daily life: *Luck. Coincidence. Randomness. Uncertainty. Risk. Doubt. Fortune. Chance.*
- Often used in a vague, casual way.
- Probability (theory): mathematical theory (a logical framework) to make these concepts precise.
- Probability quantifies uncertainty and randomness in a principled way.
- Can be deeply counterintuitive.
- Probability is extremely useful in many fields, since it provides tools for understanding and explaining variation, separating signal from noise, and modeling complex phenomena.

Applications of probability:

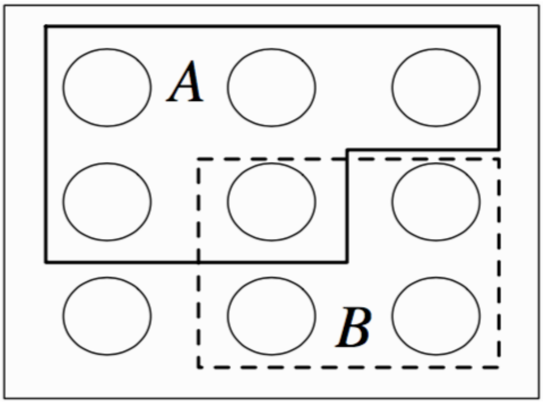
- Statistics
- Physics
- Biology
- Computer science
- Meteorology
- Political science
- Medicine
- Gambling
- **Finance**

1.2 Sample spaces and pebble world

The mathematical framework for probability is built around *sets*. Imagine

- performing an experiment,
- resulting in one out of a set of possible outcomes.
- Before the experiment is performed, it is unknown which outcome will be the result;
- after, the result "crystallizes" into the actual outcome.

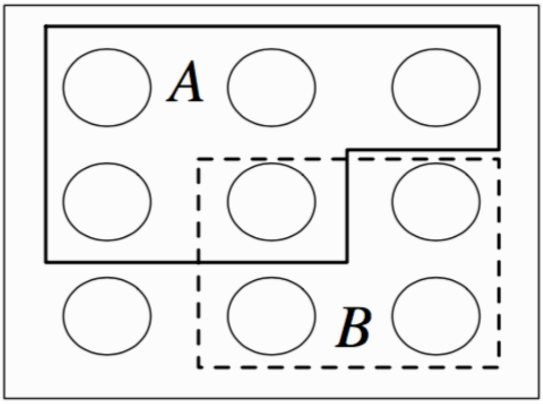
Definition 1. The *sample space* Ω of an experiment is the set of all possible outcomes of the experiment. An *event* A is a subset of the sample space Ω, and we say that A occurred if the actual outcome is in A.



If the sample space is finite, we can illustrate it as a pebble world. Each pebble represents an outcome, and an event is a set of pebbles.

Some set theory ...

- Let Ω be the sample space of an experiment.
- Let $A, B \subset \Omega$ be events.
- Then the *union* $A \cup B$ is the event that occurs if and only if at least one of A and B occurs.
- The *intersection* $A \cap B$ is the event that occurs if and only if both A and B occur.
- The complement A^c is the event that occurs if and only if A does not occur.



A is a set of 5 pebbles, B is a set of 4 pebbles, $A \cup B$ consists of the 8 pebbles in A or B (including the pebble that is in both), $A \cap B$ consists of the pebble that is in both A and B, and A^c consists of the 4 pebbles that are not in A.

De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c \qquad \text{and} \qquad (A \cap B)^c = A^c \cup B^c$$

since

- saying that it is not the case that at least one of A and B occur
- is the same as saying that A does not occur and B does not occur;
- saying that it is not the case that both occur
- is the same as saying that at least one does not occur.

Exercise: Check De Morgan's laws for the example sample space of the last two slides.

Example 2 (Coin tosses)**.**

- A coin is flipped 10 times.
- Writing Heads as H and Tails as T, a possible outcome (pebble) is $HHHTHHTTHT$.
- The sample space is the set of all possible strings of length 10 of H's and T's.
- We can (and will) encode H as 1 and T as 0, so that an outcome is a sequence $(\omega_1, \cdots, \omega_{10})$ with $\omega_j \in \{0, 1\}$, and the sample space is the set of all such sequences. [Here, '$\in$' stands for 'is an element of.']

Example (Coin tosses (continued))**.** Now let's look at some events:

- Let A_1 be the event that the first flip is Heads. As a set,

$$A_1 = \{(1, \omega_2, \cdots, \omega_{10}) : \omega_j \in \{0,1\} \text{ for } 2 \leq j \leq 10\}.$$

  Saying that A_1 occurs is the same thing as saying that the first flip is Heads. Similarly, let A_j be the event that the jth flip is Heads for $j = 2, 3, \cdots, 10$.

- Let B be the event that at least one flip was Heads. As a set,

$$B = \bigcup_{j=1}^{10} A_j$$

Example (Coin tosses (continued))**.**

- Let C be the event that all the flips were Heads. As a set,

$$C = \bigcap_{j=1}^{10} A_j$$

- Let D be the event that there were at least two consecutive Heads. As a set,

$$D = \bigcup_{j=1}^{9} (A_j \cap A_{j+1})$$

Example 3 (Pick a card, any card)**.**

- Pick a card from a standard deck of 52 cards. The sample space Ω is the set of all 52 cards (so there are 52 pebbles, one for each card)
- Consider the following events
  - A: card is an ace.
  - B: card has a black suit.
  - D: card is a diamond.
  - H: card is a heart.
- As a set, H consists of 13 cards.
- We can create various other events in terms of A, B, D, H.
- For example, $A \cap H$ is the event that the card is the Ace of Hearts, $A \cup D \cup H$ is the event that the card is red or an ace.

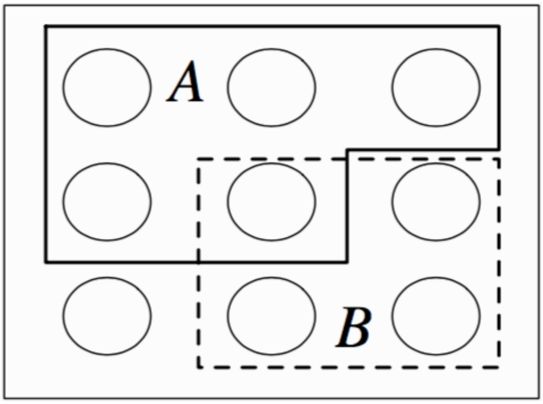
1.3 Naive definition of probability

Historically, the earliest definition of the probability of an event was to count the number of ways the event could happen and divide by the total number of possible outcomes for the experiment.

Definition 4. Let A be an event for an experiment with a finite sample space Ω. The naive probability of A is

$$P_{\text{naive}}[A] = \frac{|A|}{|\Omega|} = \frac{\text{number of outcomes in } A}{\text{total number of outcomes in } \Omega}$$

(We use $|A|$ to denote the size of A.)



$$P_{\text{naive}}[A] = \frac{5}{9};$$
$$P_{\text{naive}}[B] = \frac{4}{9};$$
$$P_{\text{naive}}[A \cup B] = \frac{8}{9};$$
$$P_{\text{naive}}[A \cap B] = \frac{1}{9}.$$

Moreover,

$$P_{\text{naive}}[A^c] = \frac{4}{9};$$
$$P_{\text{naive}}[B^c] = \frac{5}{9}; \dots$$

In general, for an arbitrary event C,

$$P_{\text{naive}}[C^c] = \frac{|C^c|}{|\Omega|} = \frac{|\Omega| - |C|}{|\Omega|} = 1 - \frac{|C|}{|\Omega|} = 1 - P_{\text{naive}}[C].$$

1.4 Counting & combinatorics

Calculating the naive probability of an event A involves counting the number of pebbles in A and in the sample space Ω.

Theorem 5 (Multiplication rule)**.** *Consider a compound experiment consisting of two sub-experiments, Experiment A and Experiment B. Suppose that Experiment A has n possible outcomes, and for each of those outcomes Experiment B has m possible outcomes. Then the compound experiment has nm possible outcomes.*

Corollary 6 (Subsets)**.** *A set with n elements has 2^n subsets, including the empty set $\emptyset$ and the set itself. This follows from the multiplication rule since for each element, we can choose whether to include it or exclude it. For example, the set $\{1, 2, 3\}$ has the 8 subsets: $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$.*

Theorem 7 (Sampling with replacement)**.** *Consider n objects and making k choices from them, one at a time with replacement (i.e., choosing a certain object does not preclude it from being chosen again). Then there are n^k possible outcomes.*

For example, imagine a jar with n balls, labeled from 1 to n. We sample balls one at a time with replacement, meaning that each time a ball is chosen, it is returned to the jar. Each sampled ball is a sub-experiment with n possible outcomes, and there are k sub-experiments. Thus, by the multiplication rule there are n^k ways to obtain a sample of size k.

Theorem 8 (Sampling without replacement)**.** *Consider n objects and making k choices from them, one at a time without replacement (i.e., choosing a certain object precludes it from being chosen again). Then there are $n(n-1)\cdots(n-k+1)$ possible outcomes, for $k \leq n$ (and 0 possibilities for $k > n$).*

Example 9 (Two dice)**.** If we roll two fair dice, which is more likely: a sum of 11 or a sum of 12?

- Label the dice A and B, and consider each die to be a sub-experiment.
- By the multiplication rule, there are 36 possible outcomes for ordered pairs of the form (value of A, value of B), and they are equally likely by symmetry.
- Of these, (5, 6) and (6, 5) are favorable to a sum of 11, while only (6, 6) is favorable to a sum of 12.
- Therefore a sum of 11 is twice as likely as a sum of 12; the probability is $1/18$ for the former, and $1/36$ for the latter.

Example 10 (Committees and teams)**.** Consider a group of four people. How many ways are there to break the people into two teams of two?

- By the multiplication rule, there are 4 ways to choose the first person on the committee and 3 ways to choose the second person on the committee,
- but this counts each possibility twice, since picking 1 and 2 to be on the committee is the same as picking 2 and 1 to be on the committee.
- Since we have overcounted by a factor of 2, the number of possibilities is $(4 \cdot 3)/2 = 6$.

1.5 Binomial coefficients

A *binomial coefficient* counts the number of subsets of a certain size for a set, such as the number of ways to choose a committee of size k from a set of n people. We are counting the number of ways to choose k objects out of n, without replacement and without distinguishing between the different orders in which they could be chosen.

Definition 11. For any nonnegative integers k and n, the binomial coefficient $\binom{n}{k}$, read as "n choose k", is the number of subsets of size k for a set of size n.

For example, $\binom{4}{2} = 6$. [see the previous slide]

Theorem 12 (Binomial coefficient formula)**.** *For $0 \leq k \leq n$, we have*

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

For $k > n \geq 0$, we have $\binom{n}{k} = 0$.

Proof. Let A be a set with $|A| = n$. Any subset of A has size at most n, so $\binom{n}{k} = 0$ for $k > n$. Now let $k \leq n$. By the theorem "sampling without replacement", there are $n(n-1)\cdots(n-k+1)$ ways to make an ordered choice of k elements without replacement. This overcounts each subset of interest by a factor of k! (since we don't care how these elements are ordered), so we can get the correct count by dividing by k!. □

Example 13 (Permutations of a word)**.** How many ways are there to permute the letters in the word LALALAAA?

- We just need to choose where the 5 A's go (or, equivalently, just decide where the 3 L's go).
- So there are

$$\binom{8}{5} = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3!} = 56 \text{ permutations.}$$

1.6 General definition of probability

The naive definition requires equally likely outcomes and can't handle an infinite sample space. We now generalize this notion of probability.

Definition 14 (General definition of probability)**.** A probability space consists of a sample space Ω and a probability function P which takes an event $A \subset \Omega$ as input and returns $P[A]$, a real number between 0 and 1, as output. The function P must satisfy the following axioms:

1. $P[\emptyset] = 0$, $P[\Omega] = 1$.
2. If $A_1, A_2, \cdots$ are disjoint events, then

$$P\left[\bigcup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} P[A_j]$$

   (Saying that these events are disjoint means that they are mutually exclusive: $A_i \cap A_j = \emptyset$ for $i \neq j$.)

Theorem 15 (Properties of probability)**.** *Probability has the following properties, for any events A and B:*

1. $P[A^c] = 1 - P[A]$.
2. *If* $A \subset B$ *then* $P[A] \leq P[B]$.
3. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$.

1.7 Recap

- Probability allows us to quantify uncertainty and randomness in a principled way.
- Pebble World can help us visualize sample spaces and events when the sample space is finite.
- In Pebble World, each outcome is a pebble, and an event is a set of pebbles.
- If all the pebbles have the same mass (i.e., are equally likely), we can apply the naive definition of probability, which lets us calculate probabilities by counting.
- There exist several tools for counting (e.g., multiplication rule).

- Moving beyond the naive definition, we define probability to be a function that takes an event and assigns to it a real number between 0 and 1.
- We require a valid probability function to satisfy two axioms.
- Many useful properties can be derived just from these axioms.
- It is important to distinguish between events and probabilities. The former are sets, while the latter are numbers. Before the experiment is done, we generally don't know whether or not a particular event will occur (happen). So we assign it a probability of happening, using a probability function P.

Chapter 2

Random variables and their distributions

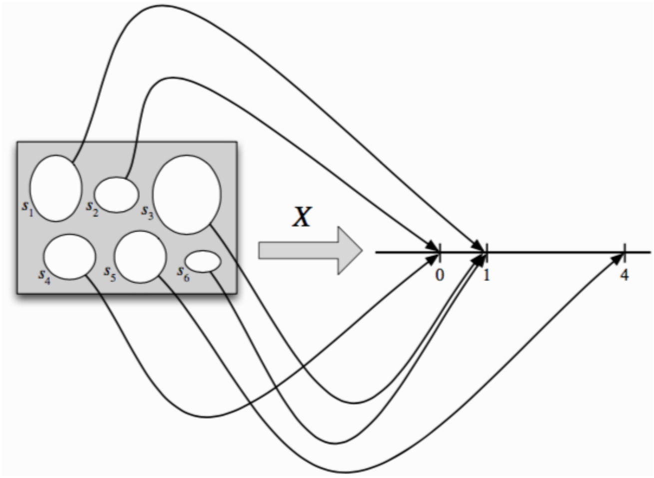
2.1 Introduction

- In this chapter, we introduce random variables,
- an incredibly useful concept that simplifies notation and expands our ability to quantify uncertainty and summarize the results of experiments.
- Random variables are essential throughout the rest of the course.

2.2 Random variables

Definition 16 (Random variable)**.** Given an experiment with sample space Ω, a *random variable* (r.v.) is a function from the sample space Ω to the real numbers $\mathbb{R}$.

- It is common, but not required, to denote random variables by capital letters.
- A random variable X assigns a numerical value $X(\omega)$ to each possible outcome ω of the experiment.
- The randomness comes from the fact that we have a random experiment (with probabilities described by the probability function P).
- The mapping itself is **deterministic**.



A random variable maps the sample space into the real line. The r.v. X depicted here is defined on a sample space with 6 elements, and has possible values 0, 1, and 4. The randomness comes from choosing a random pebble according to the probability function P for the sample space.

- The source of the randomness in a random variable is the experiment itself, in which a sample outcome $\omega \in \Omega$ is chosen according to a probability function P.
- Before we perform the experiment, the outcome ω has not yet been realized, so we don't know the value of X, though we could calculate the probability that X will take on a given value or range of values.
- After we perform the experiment and the outcome ω has been realized, the random variable crystallizes into the numerical value $X(\omega)$.

Example 17 (Coin tosses).

- Consider an experiment where we toss a fair coin twice.
- The sample space consists of four possible outcomes: $\Omega = \{HH, HT, TH, TT\}$.

Here are some random variables on this space. Each r.v. is a numerical summary of some aspect of the experiment.

- Let X be the number of Heads. This is a random variable with possible values 0, 1, and 2. Hence,
$$X(HH) = 2,\ X(HT) = X(TH) = 1,\ X(TT) = 0.$$
- Let Y be the number of Tails. In terms of X, we have $Y = 2 - X$. In other words, $Y(\omega) = 2 - X(\omega)$ for all $\omega \in S$.
- Let I be 1 if the first toss lands Heads and 0 otherwise. Then I assigns the value 1 to the outcomes HH and HT and 0 to the outcomes TH and TT.

Example (Coin tosses (continued)).

- We can also encode the sample space as $\Omega = \{(1,1), (1,0), (0,1), (0,0)\}$, where 1 is the code for Heads and 0 is the code for Tails.
- Then we can give explicit formulas for X, Y, I:

$$\begin{aligned}
X(\omega_1, \omega_2) &= \omega_1 + \omega_2; \\
Y(\omega_1, \omega_2) &= 2 - \omega_1 - \omega_2; \\
I(\omega_1, \omega_2) &= \omega_1.
\end{aligned}$$

2.3 Distributions; the discrete case

Definition 18 (Discrete random variable). A random variable X is said to be *discrete* if there is a finite list of values $a_1, a_2, \cdots, a_n$ or an infinite list of values $a_1, a_2, \cdots$ such that

$$P[X = a_j \text{ for some } j] = 1.$$

If X is a discrete r.v., then the finite or countably infinite set of values x such that $P[X = x] > 0$ is called the *support* of X.

- Most commonly in applications, the support of a discrete r.v. is a set of integers.
- In contrast, a *continuous* r.v. can take on any real value in an interval (possibly even the entire real line). [A continuous r.v. has to satisfy more properties; see Definition 29 below.]

- Given a random variable, we would like to be able to describe its behavior using the language of probability.
- For example, we might want to answer questions about the probability that the r.v. will fall into a given range.
- E.g., if L is the lifetime earnings of a randomly chosen U.S. college graduate, what is the probability that L exceeds a million dollars?
- The *distribution* of a random variable provides the answers to these questions; it specifies the probabilities of all events associated with the r.v., such as the probability of it equaling 3 and the probability of it being at least 110.

Definition 19 (Probability mass function)**.** The probability mass function (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P[X = x]$.

Note that $p_X(x)$ is positive if x is in the support of X, and 0 otherwise.

Example (Coin tosses (continued))**.**

- X is the number of Heads. Since X equals 0 if TT occurs, 1 if HT or TH occurs, and 2 if HH occurs, the PMF of X is the function p_X given by

$$\begin{aligned} p_X(0) &= P[X = 0] = \frac{1}{4}; \\ p_X(1) &= P[X = 1] = \frac{1}{2}; \\ p_X(2) &= P[X = 2] = \frac{1}{4}. \end{aligned}$$

and $p_X(x) = 0$ for all other values of x.

Example (Coin tosses (continued)).

- $Y = 2 - X$ is the number of Tails. Hence

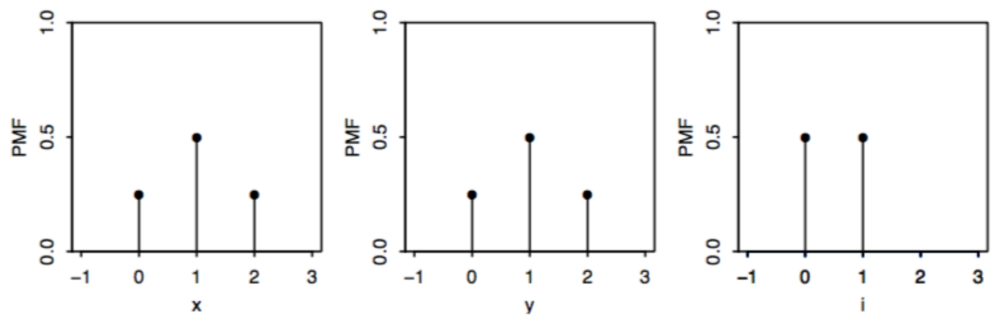
$$p_Y(0) = P[Y = 0] = \frac{1}{4};$$
$$p_Y(1) = P[Y = 1] = \frac{1}{2};$$
$$p_Y(2) = P[Y = 2] = \frac{1}{4}.$$

  and $p_Y(y) = 0$ for all other values of y.

- I is the indicator of the first toss landing Heads.

$$p_I(0) = P[I = 0] = \frac{1}{2};$$
$$p_I(1) = P[I = 1] = \frac{1}{2}.$$

Example (Coin tosses (continued)).



Example 20 (Two fair dice rolls). Let T be the sum of two fair dice rolls.

- Suppose we're interested in the probability that T is in the interval $[1, 4]$.
- There are only three values in the interval $[1, 4]$ that T can take on, namely, 2, 3, and 4.
- Hence

$$\begin{aligned} P[1 \le T \le 4] &= P[T = 2] + P[T = 3] + P[T = 4] \\ &= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}. \end{aligned}$$

2.4 Bernoulli and binomial

Definition 21 (Bernoulli distribution)**.** An r.v. X is said to have the *Bernoulli distribution* with parameter $p \in (0, 1)$ if $P[X = 1] = p$ and $P[X = 0] = 1 - p$.

We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".
Any r.v. whose possible values are 0 and 1 has a $\text{Bern}(p)$ distribution, with p the probability of the r.v. equaling 1.

Definition 22 (Indicator random variable)**.** The *indicator random variable* of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$.

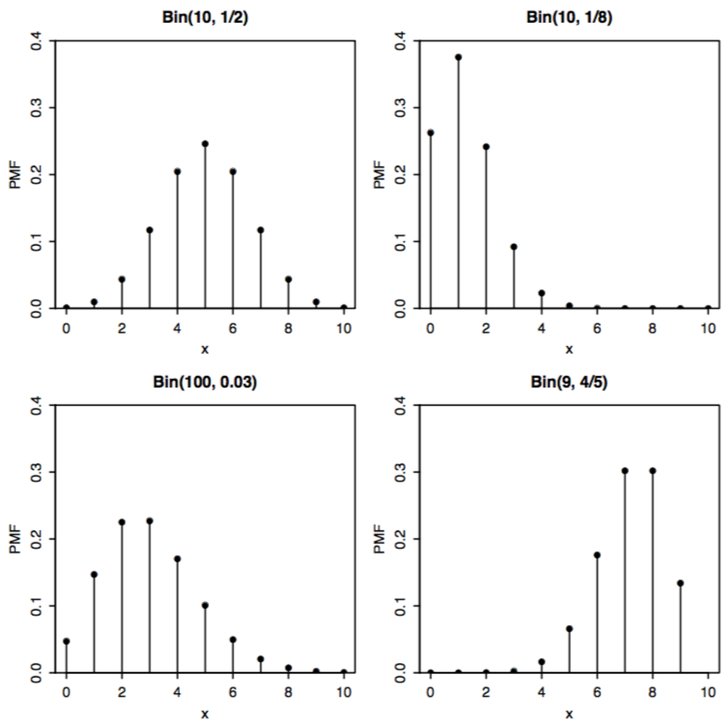
Note that $I_A \sim \text{Bern}(p)$ with $p = P[A]$.

Definition 23 (Binomial distribution)**.** Suppose that n *independent* Bernoulli trials are performed, each with the same success probability $p \in (0, 1)$. Let X be the number of successes. The distribution of X is called the *binomial distribution* with parameters n and p. We write $X \sim \text{Bin}(n, p)$.

Theorem 24. *If $X \sim \text{Bin}(n, p)$, then the PMF of X is*

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, \cdots, n$ (and $P[X = k] = 0$, otherwise).



2.5 Discrete uniform

Definition 25 (Discrete uniform distribution)**.** Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the *discrete uniform distribution* with parameter C.

We write this as $X \sim \text{DUnif}(C)$.

- The PMF of $X \sim \text{DUnif}(C)$ is
$$P[X = x] = \frac{1}{|C|}$$
for $x \in C$ (and 0 otherwise), since a PMF must sum to 1.

- For any $A \subset C$, we have
$$P[X \in A] = \frac{|A|}{|C|}.$$

Example 26 (Random slips of paper)**.** There are 100 slips of paper in a hat, each of which has one of the numbers 1,2,...,100 written on it, with no number appearing more than once. Five of the slips are drawn, one at a time.
Consider random sampling with replacement (with equal probabilities).

1. What is the distribution of how many of the drawn slips have a value of at least 80 written on them?
2. What is the distribution of the value of the jth draw (for $1 \leq j \leq 5$)?
3. What is the probability that the number 100 is drawn at least once?

Example (Random slips of paper (continued))**.** Solutions:

1. The distribution is $\text{Bin}(5, 0.21)$.
2. Let X_j be the value of the jth draw. By symmetry, $X \sim \text{DUnif}(1, 2, \cdots 100)$.
3. Taking complements,

$$\begin{aligned} P[X_j = 100 \text{ for at least one } j] &= 1 - P[X_1 \neq 100, \cdots, X_5 \neq 100] \\ &= 1 - \left(\frac{99}{100}\right)^5 \approx 0.049. \end{aligned}$$

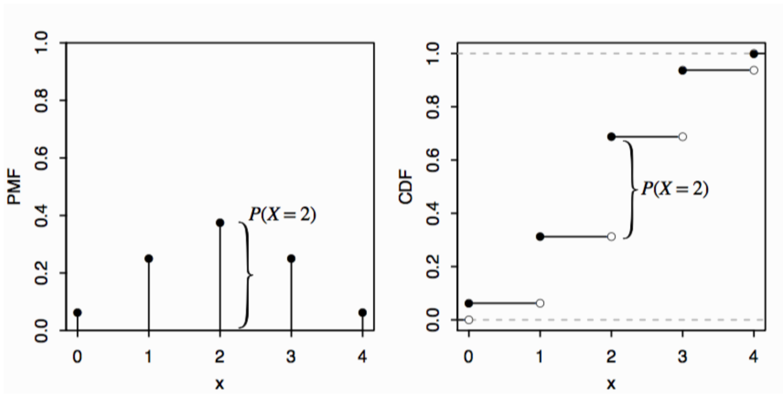
2.6 Cumulative distribution functions

Another function that describes the distribution of an r.v. is the cumulative distribution function (CDF).

Definition 27 (Cumulative distribution function)**.** The *cumulative distribution function (CDF)* of an r.v. X is the function F_X given by $F_X(x) = P[X \leq x]$.

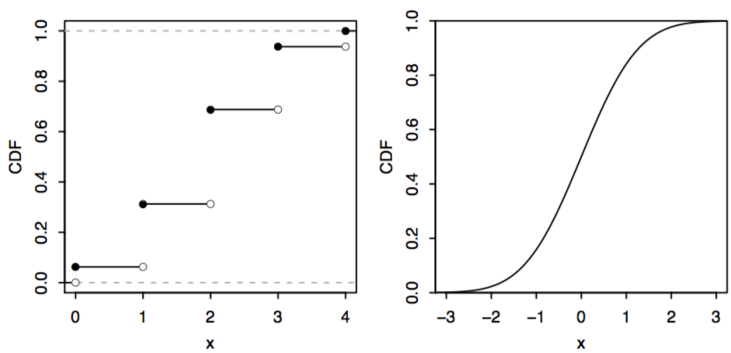
When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

Example 28 (Binomial distribution)**.** $X \sim \mathrm{Bin}(4, 0.5)$



The height of the vertical bar $P[X = 2]$ in the PMF is also the height of the jump in the CDF at 2.

2.7 Distributions; the continuous case



Discrete vs. continuous r.v.s. Left: The CDF of a discrete r.v. has jumps at each point in the support. Right: The CDF of a continuous r.v. increases smoothly.

Definition 29 (Continuous random variable)**.** An r.v. has a *continuous distribution* if its CDF is differentiable. A *continuous random variable* is a random variable with a continuous distribution.

Definition 30 (Probability density function)**.** For a continuous r.v. X with CDF F, the *probability density function (PDF)* of X is the derivative f of the CDF, given by $f(x) = F'(x)$. The support of X, and of its distribution, is the set of all x where $f(x) > 0$.

A continuous r.v. can take on any value in an interval, although the probability that it equals any particular value is exactly 0.

Proposition 31 (PDF to CDF)**.** Let X be a continuous r.v. with PDF f. Then the CDF of X is given by

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

In particular, we have

$$P[a \leq X \leq b] = \int_{a}^{b} f(t)dt.$$

2.8 Continuous uniform

Definition 32 (Continuous uniform distribution)**.** A continuous r.v. U is said to have the *uniform distribution* on the interval (a, b) if its PDF is

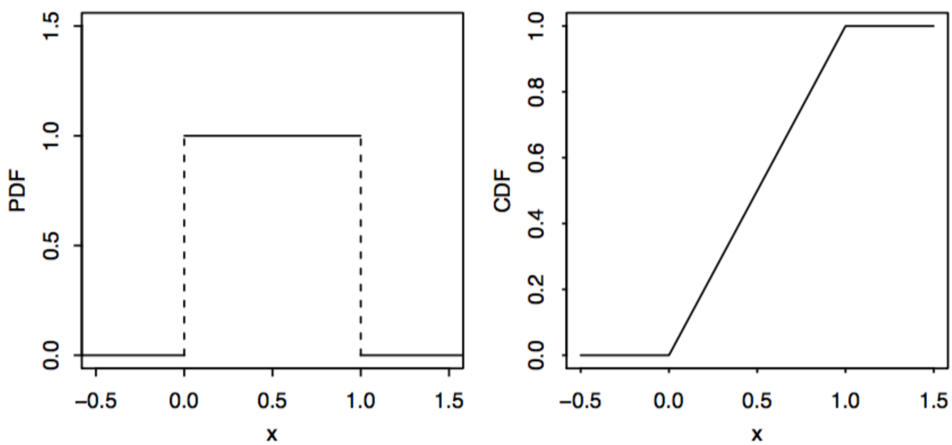
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

We write this as $U \sim \text{Unif}(a, b)$.

Proposition 33. Let $U \sim \text{Unif}(a, b)$, and let (c, d) be a subinterval of (a, b), of length l (so $l = d - c$). Then the probability of U being in (c, d) is proportional to l.

For example, a subinterval that is twice as long has twice the probability of containing U, and a subinterval of the same length has the same probability.

Example 34 (Uniform distribution)**.** The PDF and CDF of $U \sim \text{Unif}(0, 1)$:



2.9 Standard normal distribution

- The normal distribution is a famous continuous distribution with a bell-shaped PDF.
- It is extremely widely used in statistics because of a theorem, the central limit theorem, which says that under very weak assumptions, the sum of a large number of i.i.d. random variables [we shall define below what "i.i.d." means exactly] has an approximately normal distribution, regardless of the distribution of the individual r.v.s.

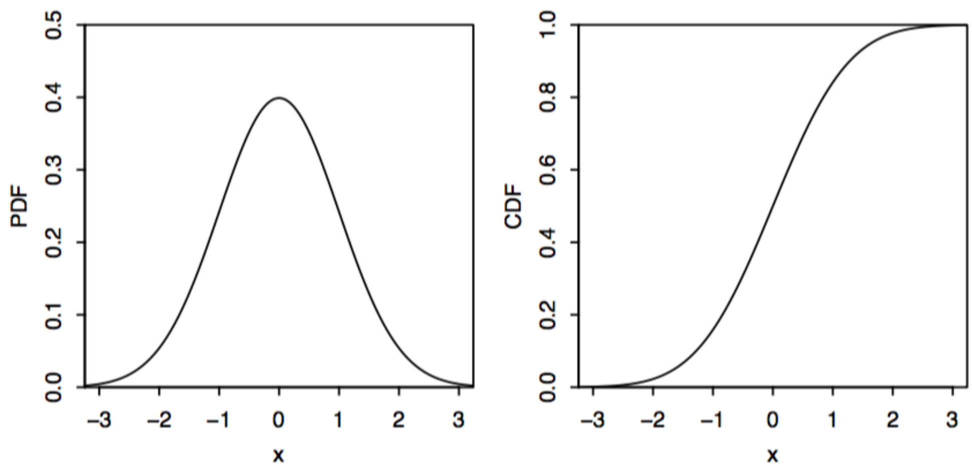
Definition 35 (Standard normal distribution)**.** A continuous r.v. Z is said to have the *standard normal distribution* if its PDF is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}.$$

We write this as $Z \sim \mathcal{N}(0, 1)$.
The standard Normal CDF is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \phi(t) dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} dt.$$

The PDF and CDF of $Z \sim \mathcal{N}(0, 1)$:



Symmetry of tail areas
The area under the PDF curve to the left of -2, which is $P[Z \leq -2] = \Phi(-2)$ by definition, equals the area to the right of 2, which is $P[Z \geq 2] = 1 - \Phi(2)$.
In general, we have

$$\Phi(-z) = 1 - \Phi(z)$$

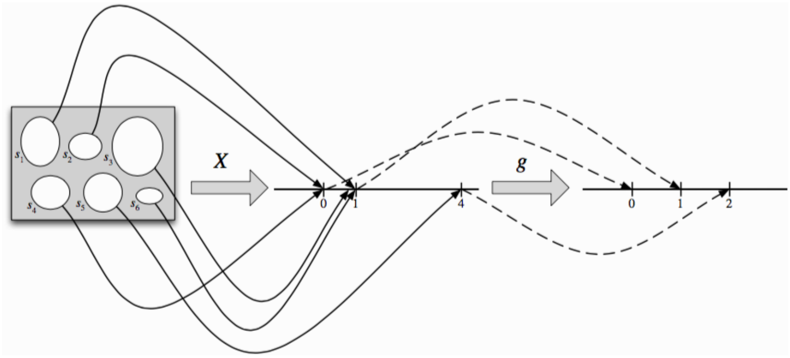
for all z.
This can be seen visually by looking at the PDF curve, and mathematically by substituting $u = -t$ below and using the fact that PDFs integrate to 1:

$$\Phi(-z) = \int_{-\infty}^{-z} \phi(t) dt = \int_{z}^{\infty} \phi(u) du = 1 - \int_{-\infty}^{z} \phi(u) du = 1 - \Phi(z).$$

2.10 Functions of random variables

- If X is a random variable, then X^2, e^X, and $\sin(X)$ are also random variables, as is $g(X)$ for any function $g : \mathbb{R} \mapsto \mathbb{R}$.
- For example, contingent claims (derivatives, options) are functions of the underlying stock price.

Example:



The r.v. X is defined on a sample space with 6 elements, and has possible values 0, 1, and 4. The function g is the square root function. Composing X and g gives the random variable $g(X) = \sqrt{X}$, which has possible values 0, 1, and 2.

Example 36 (Maximum of two dice rolls)**.** We roll two fair 6-sided dice. Let X be the number on the first die and Y the number on the second die. The following table gives the values of X, Y, and $\max(X, Y)$ under 7 of the 36 outcomes in the sample space.

s	X	Y	$\max(X, Y)$
$(1, 2)$	1	2	2
$(1, 6)$	1	6	6
$(2, 5)$	2	5	5
$(3, 1)$	3	1	3
$(4, 3)$	4	3	4
$(5, 4)$	5	4	5
$(6, 6)$	6	6	6

Example (Maximum of two dice rolls (continued))**.** The PMF can be computed as follows:

$$\begin{aligned}
P[\max(X,Y)=5] &= P[X=5, Y\leq 4] + P[X\leq 4, Y=5] + P[X=5, Y=5] \\
&= 2P[X=5, Y\leq 4] + \frac{1}{36} \\
&= 2\left(\frac{4}{36}\right) + \frac{1}{36} \\
&= \frac{9}{36} \\
&= \frac{1}{4}.
\end{aligned}$$

2.11 Independence of random variables

Intuitively, if two r.v.s X and Y are independent, then knowing the value of X gives no information about the value of Y, and vice versa.

Definition 37 (Independence)**.** Random variables X and Y are said to be *independent* if

$$P[X\leq x, Y\leq y] = P[X\leq x]P[Y\leq y],$$

for all $x, y \in \mathbb{R}$.

If X and Y are discrete, then independence of X and Y is the same as

$$P[X=x, Y=y] = P[X=x]P[Y=y],$$

for all $x, y \in \mathbb{R}$.

Example 38 (Two fair dice)**.**

- If X is the number on the first die and Y is the number on the second die, then $X+Y$ is not independent of $X-Y$.
- To see why, note that

$$0 = P[X+Y=12, X-Y=1] \neq P[X+Y=12]P[X-Y=1] = \frac{1}{36}\cdot\frac{5}{36}.$$

- This also makes sense intuitively: knowing the sum of the dice is 12 tells us their difference must be 0, so the r.v.s provide information about each other.

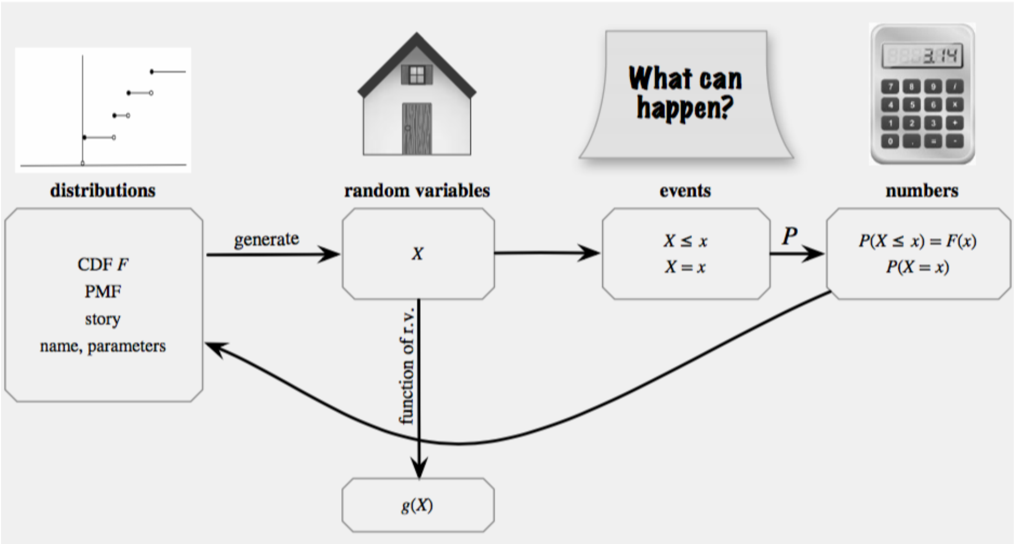
Definition 39 (I.i.d.)**.** We will often work with random variables that are independent and have the same distribution. We call such r.v.s independent and identically distributed, or *i.i.d.* for short.

2.12 Recap

- A random variable (r.v.) is a function assigning a real number to every possible outcome of an experiment.
- The distribution of an r.v. X is a full specification of the probabilities for the events associated with X, such as $\{X = 3\}$ and $\{1 \leq X \leq 5\}$.
- The CDF of a random variable X is the function $P[X \leq x]$.
- It is very important to distinguish between a random variable and its distribution: the distribution is a blueprint for building the r.v., but different r.v.s can have the same distribution.
- A function of a random variable is still a random variable.
- Two random variables are independent if knowing the value of one r.v. gives no information about the value of the other.

- The PMF of a discrete random variable X is the function $P[X = x]$.
- Examples for discrete distributions are the following:
  - A Bern(p) r.v. is the indicator of success in a Bernoulli trial with probability of success p.
  - A Bin(n, p) r.v. is the number of successes in n independent Bernoulli trials, all with the same probability p of success.
  - A DUnif(C) r.v. is obtained by randomly choosing an element of the finite set C, with equal probabilities for each element.

- A continuous r.v. can take on any value in an interval, although the probability that it equals any particular value is exactly 0.
- The CDF of a continuous r.v. is differentiable, and the derivative is called the probability density function (PDF).
- Probability is represented by area under the PDF curve, not by the value of the PDF at a point.
- We must integrate the PDF to get a probability.
- Two important continuous distributions are the uniform and the standard normal distribution.


distributions
random variables
events
numbers
What can happen?
CDF F
PMF
story
name, parameters
generate
X
$X \leq x$
$X = x$
P
$P(X \leq x) = F(x)$
$P(X = x)$
function of r.v.
$g(X)$

Chapter 3

Expectation

3.1 Introduction

- Yesterday, we introduced the distribution of a random variable, which gives us full information about the probability that the r.v. will fall into any particular set.
- For example, we can say how likely it is that the r.v. will exceed 100 or that it will equal 5.
- However, often we want just one number summarizing the "average" value of the r.v.
- There are several senses in which the word "average" is used, but by far the most commonly used is the mean of an r.v., also known as its expected value.

3.2 Definition of expectation

- Given a list of numbers $x_1, x_2, \cdots, x_n$, the familiar way to average them is to add them up and divide by n. This is called the arithmetic mean, and is defined by

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j.$$

- More generally, we can define a weighted mean as

$$\text{weighted-mean}(x) = \sum_{j=1}^{n} x_j p_j,$$

  where the weights $p_1, p_2, \cdots, p_n$ are pre-specified nonnegative numbers that add up to 1.

Definition 40 (Expectation of a discrete r.v.)**.** The *expected value* (also called the *expectation* or *mean*) of a discrete r.v. X is given by

$$E[X] = \sum_x xP[X = x],$$

where the sum is over the support of X (in any case, $xP[X = x]$ is 0 for any x not in the support). The expectation is undefined if

$$\sum_x |x|P[X = x],$$

diverges, since then the series for $E[X]$ diverges or its value depends on the order in which the x_j are listed.

In words, the expected value of X is a weighted average of the possible values that X can take on, weighted by their probabilities.

With this definition, the expected value can be interpreted as the center of mass.

Example 41 (Fair dice)**.**

- X takes on the values $1, 2, 3, 4, 5, 6$, with equal probabilities.
- Intuitively, we should be able to get the average by adding up these values and dividing by 6.
- Using the definition, the expected value is

  $$E[X] = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5$$

  as guessed.
- Note though that X never equals its mean in this example.

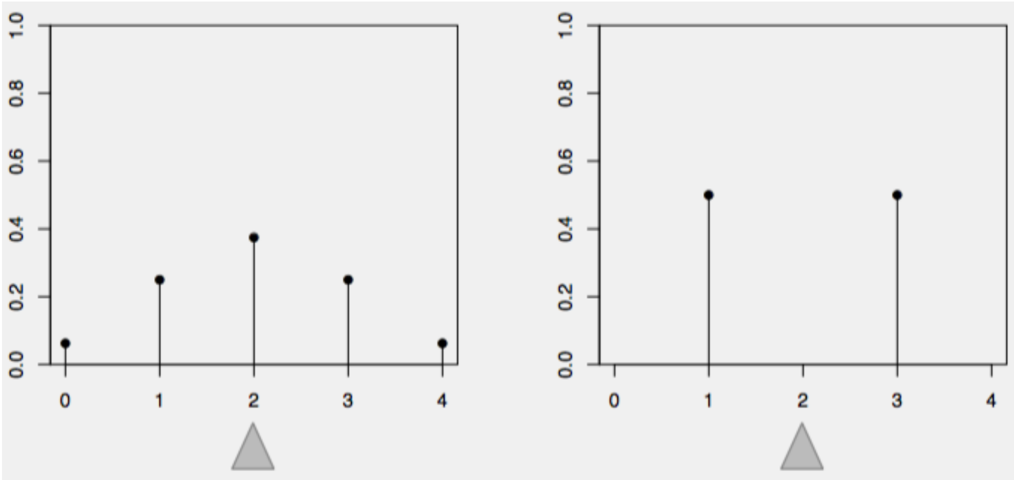
Example 42 (Bernoulli distribution)**.**

- Let $X \sim \text{Bern}(p)$ and $q = 1 - p$.
- Using the definition, the expected value is

  $$E[X] = 1 \cdot p + 0 \cdot q = p.$$

- This makes sense intuitively since it is between the two possible values of X, compromising between 0 and 1 based on how likely each is.

If X and Y are discrete r.v.s with the same distribution, then $E[X] = E[Y]$ (if either side exists). The converse holds not true, as the following plot illustrates:



Definition 43 (Expectation of a continuous r.v.)**.** The *expected value* (also called the *expectation* or *mean*) of a continuous r.v. X with PDF f is given by

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

The expectation is undefined if

$$\int_{-\infty}^{\infty} |x| f(x) dx = \infty.$$

With this definition, the expected value retains its interpretation as a center of mass.

3.3 Linearity of expectations

Theorem 44 (Linearity of expectation)**.** *For any r.v.s X, Y and any constant c,*

$$E[X + Y] = E[X] + E[Y];$$
$$E[cX] = cE[X].$$

Proposition 45 (Monotonicity of expectation)**.** Let X and Y be r.v.s such that $X \leq Y$ with probability 1. Then $E[X] \leq E[Y]$, with equality holding if and only if $X = Y$ with probability 1.

Example 46 (Binomial distribution)**.** Let $X \sim \text{Bin}(n, p)$. Then we can write

$$X = I_1 + \cdots + I_n,$$

where each I_j has expectation $E[I_j] = 1 \cdot p + 0 \cdot q = p$. By linearity,

$$E[X] = E[I_1] + \cdots + E[I_n] = np.$$

3.4 Indicator r.v.s and the fundamental bridge

Recall from above that the indicator r.v. I_A (or $I(A)$) for an event A is defined to be 1 if A occurs and 0 otherwise. So I_A is a Bernoulli random variable, where success is defined as "A occurs" and failure is defined as "A does not occur".

Theorem 47 (Indicator r.v. properties)**.** *Let A and B be events. Then the following properties hold.*

$$\begin{aligned} (I_A)^k &= I_A \text{ for any positive integer } k. \\ I_{A^c} &= 1 - I_A. \\ I_{A \cap B} &= I_A I_B. \\ I_{A \cup B} &= I_A + I_B - I_A I_B. \end{aligned}$$

Theorem 48 (Fundamental bridge between probability and expectation)**.** *There is a one-to-one correspondence between events and indicator r.v.s, and the probability of an event A is the expected value of its indicator r.v. I_A:*

$$P[A] = E[I_A].$$

3.5 Law of the unconscious statistician (LOTUS)

- If g is a function and X an r.v., then $g(X)$ is also an r.v.
- It is often of interest to compute $E[g(X)]$.
- As it turns out, $E[g(X)]$ does not equal $g(E[X])$ in general if g is not linear.
- So how do we correctly calculate $E[g(X)]$?
- Since $g(X)$ is an r.v., one way is to first find the distribution of $g(X)$ and then use the definition of expectation.
- Perhaps surprisingly, it turns out that it is possible to find $E[g(X)]$ directly using the distribution of X, without first having to find the distribution of $g(X)$.
- This is done using the law of the unconscious statistician (LOTUS).

Theorem 49 (Law of the unconscious statistician (LOTUS))**.** *If X is a discrete r.v. and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then*

$$E[g(X)] = \sum_x g(x)P[X = x],$$

where the sum is taken over all possible values of X.
If X is a continuous r.v. with PDF f, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

3.6 Variance

- Like expected value, variance is a single-number summary of the distribution.
- While the expected value tells us the center of mass of a distribution, the variance tells us how spread out the distribution is.

Definition 50 (Variance and standard deviation)**.** The *variance* of an r.v. X is

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

The square root of the variance is called the *standard deviation*:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

- The variance of X measures how far X is from its mean on average,
- but instead of simply taking the average difference between X and its mean $E[X]$, we take the average squared difference.
- To see why, note that the average deviation from the mean, $E[X - E[X]]$, always equals 0 by linearity; positive and negative deviations cancel each other out.
- By squaring the deviations, we ensure that both positive and negative deviations contribute to the overall variability.

Theorem 51. *For any r.v. X,*

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2.$$

Proof. Let $\mu = E[X]$. Expand $(X - \mu)^2$ and use linearity:

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2.\end{aligned}$$

□

Theorem 52. *For any r.v. X and any constant $c \in \mathbb{R}$, its variance has the following properties.*

1. $\mathrm{Var}(X + c) = \mathrm{Var}(X)$.
2. $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$.
3. $\mathrm{Var}(X) \geq 0$.
4. *If Y is an r.v., independent of X then* $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Example 53 (Binomial variance)**.**

- Let's find the variance of $X \sim \mathrm{Bin}(n, p)$ using indicator r.v.s to avoid tedious sums.
- Represent $X = I_1 + I_2 + \cdots + I_n$, where I_j is the indicator of the jth trial being a success.
- Each I_j has variance

$$\mathrm{Var}(I_j) = E[I_j^2] - (E[I_j])^2 = p - p^2 = p(1-p).$$

- Since the I_j are independent, we can add their variances to get the variance of their sum:

$$\mathrm{Var}(X) = \mathrm{Var}(I_1) + \cdots + \mathrm{Var}(I_n) = np(1-p).$$

Example 54 (Uniform expectation and variance)**.**

- Let's derive the mean and variance of $U \sim \text{Unif}(a, b)$.
- The expectation is extremely intuitive: the PDF is constant, so its balancing point should be the midpoint of (a, b).
- This is exactly what we find by using the definition of expectation for continuous r.v.s:

$$E[U] = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2}.$$

- For the variance, we first find $E[U^2]$, using the continuous version of LOTUS:

$$E[U^2] = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right) = \frac{b^3 - a^3}{3(b-a)}.$$

Example (Uniform expectation and variance (continued))**.**

- Then

$$\text{Var}(U) = E[U^2] - (E[U])^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2.$$

- This simplifies, after factoring $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$, to

$$\text{Var}(U) = \frac{(b-a)^2}{12}.$$

3.7 Normal distribution

- Above, we introduced the standard normal distribution. Recall that continuous r.v. Z is said to have the *standard normal distribution* if its PDF is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}.$$

- We write this as $Z \sim \mathcal{N}(0, 1)$.
- Let's compute its mean:

$$E[Z] = \int_{-\infty}^{\infty} z\phi(z) dz = 0,$$

where the last equality follows from the fact that $z\phi(z)$ is an odd function (the area under the function from $-\infty$ to 0 cancels the area under the function from 0 to ∞.)

- For the variance, we use LOTUS as follows:

$$\begin{aligned}\text{Var}(Z) &= E[Z^2] - (E[Z])^2 = E[Z^2] \\ &= \int_{-\infty}^{\infty} z^2\phi(z)dz = 2\int_{0}^{\infty} z(z\phi(z))dz\end{aligned}$$

  Now we use integration by parts to obtain

$$\begin{aligned}\text{Var}(Z) &= \frac{2}{\sqrt{2\pi}}\left(-z\mathrm{e}^{-z^2/2}\Big|_0^\infty + \int_0^\infty \mathrm{e}^{-z^2/2}dz\right) \\ &= \frac{2}{\sqrt{2\pi}}\left(0 + \frac{\sqrt{2\pi}}{2}\right) = 1.\end{aligned}$$

Definition 55 (Normal distribution)**.** If $Z \sim \mathcal{N}(0,1)$ then

$$X = \mu + \sigma Z$$

is said to have the Normal distribution with mean μ and variance σ^2.

We write this as $X \sim \mathcal{N}(\mu, \sigma^2)$.

It's clear by properties of expectation and variance that X does in fact have mean μ and variance σ^2.

Theorem 56 (Normal CDF and PDF)**.** *Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is*

$$F(x) = \Phi\left(\frac{X-\mu}{\sigma}\right),$$

and the PDF of X is

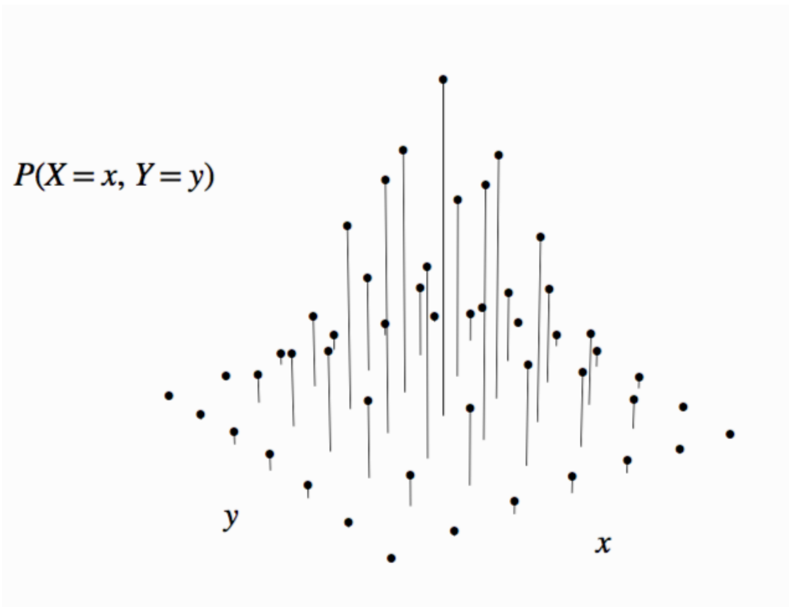
$$f(x) = \phi\left(\frac{X-\mu}{\sigma}\right)\frac{1}{\sigma}.$$

3.8 Covariance and correlation

- Individual distributions of two r.v.s do not tell us anything about whether the r.v.s are independent or dependent.
- For example, two Bern(1/2) r.v.s X and Y could be independent if they indicate Heads on two different coin flips, or dependent if they indicate Heads and Tails respectively on the same coin flip.
- Of course, in real life, we usually care about the relationship between multiple r.v.s in the same experiment.
- Joint distributions capture the previously missing information about how multiple r.v.s interact.

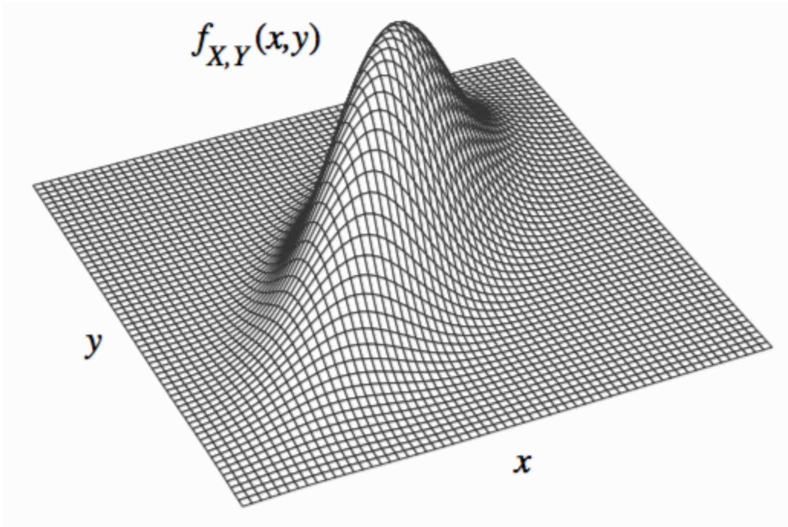
Definition 57 (Joint PMF)**.** The joint PMF of discrete r.v.s X and Y is the function $p_{X,Y}$ given by

$$p_{X,Y}(x, y) = P[X = x, Y = y].$$



Definition 58 (Joint PDF)**.** The joint PDF of continuous r.v.s X and Y is the function $f_{X,Y}$ given by

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} P[X \leq x, Y \leq y].$$



Theorem 59 (2D LOTUS)**.** *Let g be a function from $\mathbb{R}^2$ to $\mathbb{R}$. If X and Y are discrete, then*

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) P[X = x, Y = y].$$

If X and Y are continuous with joint PDF $f_{X,Y}$, then

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy.$$

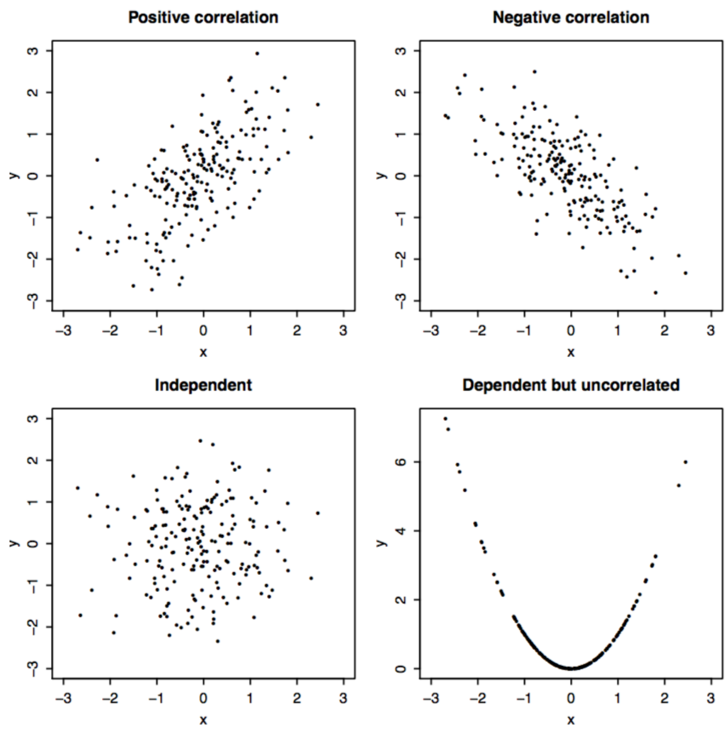
- Just as the mean and variance provided single-number summaries of the distribution of a single r.v., covariance is a single-number summary of the joint distribution of two r.v.s.
- Roughly speaking, covariance measures a tendency of two r.v.s to go up or down together, relative to their expected values:
- positive covariance between X and Y indicates that when X goes up, Y also tends to go up,
- and negative covariance indicates that when X goes up, Y tends to go down.

Definition 60 (Covariance)**.** The *covariance* between r.v.s X and Y is

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

If $\text{Cov}(X,Y) = 0$ then X and Y are said to be *uncorrelated.*

- Let's think about the definition intuitively.
- If X and Y tend to move in the same direction, then $X - E[X]$ and $Y - E[Y]$ will tend to be either both positive or both negative, so $(X - E[X])(Y - E[Y])$ will be positive on average, giving a positive covariance.
- If X and Y tend to move in opposite directions, then $X - E[X]$ and $Y - E[Y]$ will tend to have opposite signs, giving a negative covariance.



Proposition 61. For two r.v.s X and Y, we have

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

Definition 62 (Correlation)**.** The *correlation* between r.v.s X and Y is

$$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Theorem 63 (Correlation bounds)**.** *For any two r.v.s X and Y,*

$$-1 \le \text{Cor}(X,Y) \le 1.$$

3.9 Recap

- The expectation of a discrete r.v. X is

$$E[X] = \sum_x xP[X = x],$$

  and of a continuous r.v. X with PDF f is

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

- Expectation is a single number summarizing the center of mass of a distribution.
- A single-number summary of the spread of a distribution is the variance, defined by

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

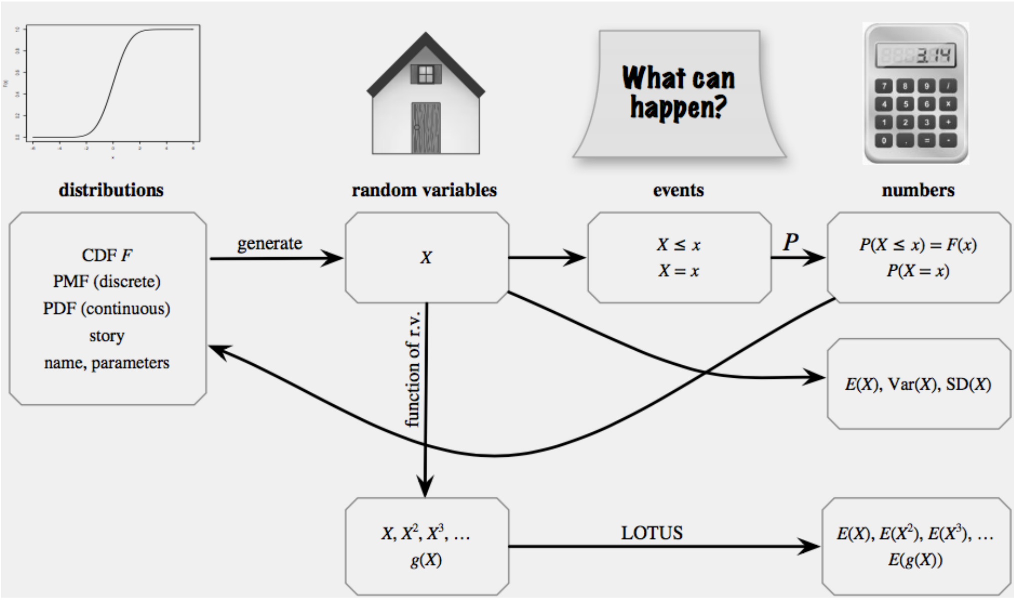
- The square root of the variance is called the standard deviation.
- Expectation is linear, but variance is not.
- An important tool for computing expectations is LOTUS, which says we can calculate the expectation of $g(X)$ using only the PMF/PDF of X.

- We also discussed normally distributed random variables.
- If $Z \sim \mathcal{N}(0, 1)$ then

$$X = \mu + \sigma Z$$

  is said to have the Normal distribution with mean μ and variance σ^2.
- A $\mathcal{N}(\mu, \sigma^2)$ r.v. has a symmetric bell-shaped PDF centered at μ, with σ^2 controlling how spread out the curve is.

- Covariance is a single-number summary of the tendency of two r.v.s to move in the same direction.
- If two r.v.s are independent, then they are uncorrelated (but the converse does not hold).
- Correlation is a unitless, standardized version of covariance that is always between -1 and 1.



What can happen?
3.14
distributions
random variables
events
numbers
CDF F
PMF (discrete)
PDF (continuous)
story
name, parameters
generate
X
$X \leq x$
$X = x$
P
$P(X \leq x) = F(x)$
$P(X = x)$
$E(X)$, Var(X), SD(X)
function of r.v.
$X, X^2, X^3, \ldots$
$g(X)$
LOTUS
$E(X), E(X^2), E(X^3), \ldots$
$E(g(X))$

Chapter 4

Generating random numbers, Part I

4.1 Introduction

"Random numbers should not be generated with a method chosen at random", Knuth (1981).

- In many applications in finance one needs to compute the expectation $E[X]$ of a random variable X, e.g., in option pricing.
- It is not always possible to compute $E[X]$ analytically.
- Suppose we have a sequence of random variables $(X_i)_{i\in\mathbb{N}}$ which are mutually independent and identically distributed with the same distribution as X, then

$$P\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = E[X]\right] = 1$$

  by the so called *Strong Law of Large Numbers.*
- How can one obtain (realisations of) X_i?

Example 64 (Fair die).

- Consider a random variable X which can take values in $\{1,2,3,4,5,6\}$ with $P[X=i]=\frac{1}{6}$ for $i=1,\cdots,6$.
- Then $E[X]$ can be computed analytically and is given by

$$E[X]=\sum_{i=1}^{6} i \underbrace{P[X=i]}_{=1/6}=3.5.$$

- We can interpret the random variable X as the **outcome of the role of a fair die**.
- We can compute an approximation for $E[X]$ by doing a random experiment: We role a fair die n times and therefore generate some numbers $x_1,\cdots,x_n$. Then $\frac{1}{n}\sum_{i=1}^{n} x_i$ can be used as an approximation for $E[X]$.

- Here: Realisations of X can be obtained by rolling a fair die.
- Of course, we already know how to simulate certain distributions with Python (e.g., the commands in numpy.random)
- We want to understand how these commands work; and we need more general methods than this!
- Need methods that can generate a **large number** of random numbers **from any distribution on a computer**.

- We are interested in **generating random numbers** from various distributions.
- As soon as we generate random numbers **on a computer, the generation will have to be based on a completely deterministic mechanism**.
- The generated output is therefore sometimes also referred to as **pseudo-random numbers**.
- The idea is to **develop mechanisms** such that the **computer-generated random numbers mimic the properties of true random numbers** as much as possible.

4.2 The linear congruential generator for generating samples from the uniform distribution

- Initially we want to generate samples from a uniform distribution.
- As we will later see, these samples can also be used to derive samples from other distributions.

Recall the following facts of a uniformly distributed random variable $U \sim \text{Unif}(0,1)$.

- The probability density function (PDF) of U is given by

$$f(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

- The cumulative distribution function (CDF) of U is given by

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \in [0,1] \\ 1, & \text{if } x > 1. \end{cases}$$

The linear congruential generator can be used to generate a sample from the uniform distribution on $[0,1]$.

Definition 65 (Linear congruential generator)**.** Let $m \in \mathbb{N}$ and $a, x_0 \in \{1, 2, \cdots, m-1\}$. A *linear congruential generator* is a recurrence of the following form. For $i = 0, 1, 2, \ldots$ set

$$\begin{aligned} x_{i+1} &= a x_i \bmod m, \\ u_{i+1} &= x_{i+1}/m. \end{aligned}$$

x_0 is called the *seed*, a the *multiplier* and m the *modulus*.

- We do not allow $x_0 = 0$, since for $x_0 = 0$ we would get $x_i = 0 \ \forall i \in \mathbb{N}$.
- The sequence $(u_i)_{i \in \mathbb{N}}$ is a sequence in $[0,1)$.
  It is **deterministic** (so-called **pseudo-random numbers**). However, for appropriate values of a and m, it **does resemble a sequence of samples from the uniform distribution on** $[0,1]$.

- Recall that for $x, m \in \mathbb{N}$ the expression $x \bmod m$ (say x modulo m) returns the remainder of x after division by m, i.e.,

$$x \bmod m = x - m \left\lfloor \frac{x}{m} \right\rfloor \in \{0, 1, \cdots, m-1\},$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

- Examples:

$$5 \mod 7 = 5, \qquad 20 \mod 10 = 0, \qquad 13 \mod 7 = 6.$$

Example 66. For $a = 6$, $m = 11$ and $x_0 = 1$ the linear congruential generator yields

$$1, 6, 3, 7, 9, 10, 5, 8, 4, 2, 1, 6, \cdots$$

for the x_i and the corresponding sequence (u_i) is obtained by setting $u_{i+1} = x_{i+1}/11$.
Observe that this sequence periodically returns back to the seed 1!

- Each of the numbers in a sequence (x_i) resulting from the linear congruential generator takes values in the set $\{0, 1, \cdots, m-1\}$.
- The sequence (x_i) (and hence also (u_i)) will **repeat itself after at most m steps**.
- Hence, **large values of m are necessary for a long cycle**.
- Large values of **m** are not sufficient for a long cycle.

The following example shows that a large m itself does not guarantee a long cycle.

Example 67. Let $m = 11$ as in Example 66, but now we choose the multiplier $a = 3$.

1. With seed $x_0 = 1$ we obtain

$$1, 3, 9, 5, 4, 1, \cdots$$

for the x_i.

2. Changing the seed to $x_0 = 2$ yields

$$2, 6, 7, 10, 8, 2, \cdots.$$

We see that the possible values are split into two cycles.

- One can show that for a **prime number m**, a **full period** is achieved for any $x_0 \neq 0$, if
  - $a^{m-1} - 1$ is a multiple of m,
  - $a^j - 1$ is not a multiple of m for $j = 1, \cdots, m-2$.

  (Full period means that all $m-1$ distinct values $1, 2, \cdots, m-1$ are produced before repeating.)

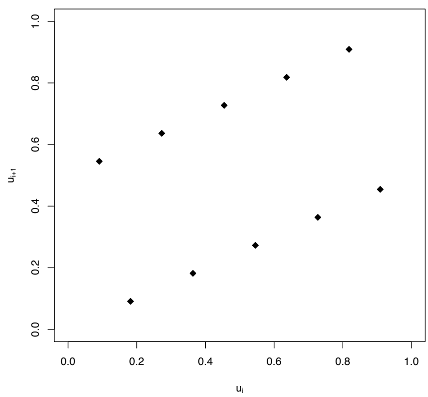
- Such a number a is called a **primitive root of** m.
- The sequence (x_i) is then given by

$$x_0,\ ax_0 \mod m,\ a^2x_0 \mod m, \cdots .$$

- It returns to x_0 at the smallest k for which $a^k x_0 \mod m = x_0$, which is the smallest k for which $a^k \mod m = 1$, i.e., the smallest k for which $(a^k - 1)$ is a multiple of m.

Example 68 (Lattice structure)**.**

- Consider the linear congruential generator in Example 66 with $a = 6$, $m = 11$, $x_0 = 1$.
- Plot consecutive overlapping pairs $(u_1, u_2), (u_2, u_3), \cdots, (u_{10}, u_{11})$.
- You will find that the ten points lie on just two parallel lines through the unit square (with u_i on x-axis and u_{i+1}-on y axis).
- This lattice structure distinguishes those samples from genuine random numbers.



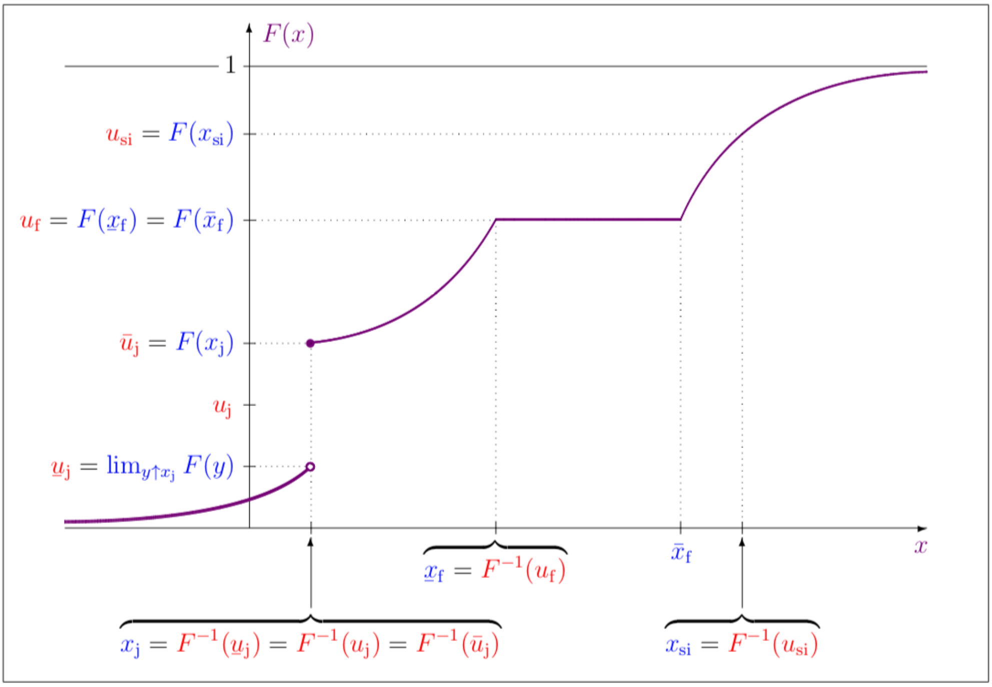
- As we have seen, the linear congruential generator is started with some seed x_0.
- The existence of this **seed is a common feature of random number generators**.
- On a computer, a **time stamp is often used as a seed** for a random number generator.
- If one would like to reproduce results one will need to start the random number generator with the same seed.
- The linear congruential generator is a very **simple** random number generator and has some undesirable properties (e.g. the lattice structure).
- There are much more advanced methods for generating uniform random numbers but these are beyond the scope of this course.

4.3 The inverse transform method for generating samples from a CDF F

We consider a random variable X with cumulative distribution function F, i.e., $F(x) = P[X \leq x]$.

Definition 69 (Inverse of CDF)**.** Given a cumulative distribution function (CDF) F, its (generalized) *inverse* F^{-1} is defined by

$$F^{-1}(u) = \inf \{x \in \mathbb{R} \mid \ F(x) \geq u\} \quad \text{for } u \in (0,1).$$



Theorem 70 (Universality of the Uniform)**.** *If $U \sim \mathrm{Unif}(0,1)$, then the random variable $F^{-1}(U)$ has cumulative distribution function F.*

Proof. We first note that, given $x \in \mathbb{R}$ and $u \in (0,1)$, the right-continuity of F implies that

$$F^{-1}(u) \leq x \iff u \leq F(x).$$

Hence,

$$P\left[F^{-1}(U) \leq x\right] = P[U \leq F(x)] = F(x),$$

where the last equality follows because U is uniform on $[0,1]$. □

- Let's make sure we understand what the theorem is saying.
- If we start with $U \sim \text{Unif}(0,1)$ and a CDF F,
- then we can create an r.v. whose CDF is F
- by plugging U into the inverse CDF F^{-1}.
- Since F^{-1} is a function (known as the quantile function), U is a random variable, and a function of a random variable is a random variable, $F^{-1}(U)$ is a random variable;
- universality of the Uniform says its CDF is F.

Example 71. (Exponential distribution)

- Consider the exponential distribution with parameter μ. Its CDF F is given by

$$F(x) = 1 - e^{-\mu x} \quad \text{for } x > 0.$$

- From

$$u = F(F^{-1}(u)) = 1 - e^{-\mu F^{-1}(u)},$$

we can see that

$$F^{-1}(u) = -\frac{\log(1-u)}{\mu}.$$

- Noting that, if $U \sim \text{Unif}(0,1)$, then $1-U \sim \text{Unif}(0,1)$, we conclude that the random variable $-\frac{1}{\mu}\log(U)$ has the exponential distribution with parameter μ.

- In many cases, F^{-1} cannot be calculated in closed analytical form.
- However, we can still calculate F^{-1} **numerically**.
  Let $u \in (0,1)$. There are many situations (e.g., if F is continuous and strictly increasing) in which one can compute $F^{-1}(u)$ by just solving the following equation for x:

$$F(x) - u = 0.$$

- For instance, if F is C^1, then we can use **Newton's method**. Choose an initial point x_0 and compute recursively

$$x_{n+1} = x_n - \frac{F(x_n) - u}{F'(x_n)}.$$

  For a suitable choice of x_0 and for a suitably well-behaved function F this sequence converges to the required solution.

If a discrete random variable is considered, the evaluation of F^{-1} reduces to a table lookup.

- Suppose X is a random variable which has possible values $c_1 < \cdots < c_n$. (Note: This ordering is important!)
  Let

$$P[X = c_i] = p_i,\ i = 1, \cdots, n \text{ with } p_i \geq 0 \text{ and } \sum_{i=1}^{n} p_i = 1.$$

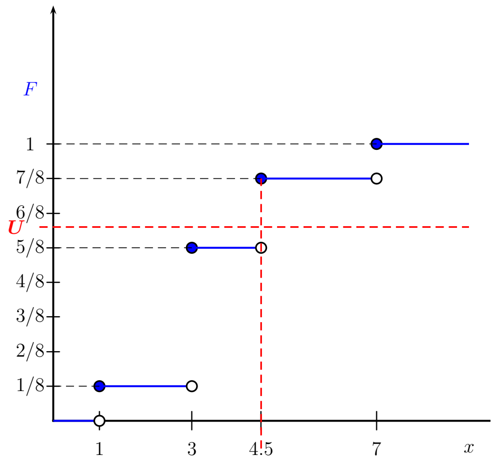
- We define cumulative probabilities

$$q_0 = 0, \qquad\qquad q_i = \sum_{j=1}^{i} p_j, \quad i = 1, \cdots, n.$$

- Then the cumulative distribution function F of X satisfies $F(c_i) = P[X \leq c_i] = q_i$.
- Now we can sample from this distribution as follows:
  1. Generate $U \sim \text{Unif}(0,1)$.
  2. Find $K \in \{1, \cdots, n\}$ such that $q_{K-1} < U \leq q_K$.
  3. Set $\tilde{X} = c_K$.

Example 72 (Discrete distribution). Plot $F(x) = P[X \leq x]$ for a random variable X taking only the values $c_1 = 1, c_2 = 3, c_3 = 4.5, c_4 = 7$.

- Let $P[X = 1] = \frac{1}{8} = p_1$, $P[X = 3] = \frac{1}{2} = p_2$, $P[X = 4.5] = \frac{2}{8} = p_3$, $P[X = 7] = \frac{1}{8} = p_4$.
- Here $q_0 = 0 = P[X < 1]$, $q_1 = \frac{1}{8} = 0.125 = F(1)$, $q_2 = \frac{5}{8} = 0.625 = F(3)$, $q_3 = \frac{7}{8} = 0.875 = F(4.5)$, $q_4 = 1 = F(7)$.
- Suppose $U = 0.7$, then $q_2 = F(c_2) < U \leq F(c_3) = q_3$. Therefore, $X = c_3 = 4.5$.



4.4 Recap

- The linear congruential generator map is a method to construct pseudo-random numbers that resemble samples from a uniform distribution.
- If we have a sample from a uniform distribution, then the inverse transform method yields samples for an arbitrary CDF.
- This method uses the universality of the uniform, namely the fact that $F^{-1}(U)$ for some $U \sim \text{Unif}(0, 1)$ and CDF F yields a sample from F.

Chapter 5

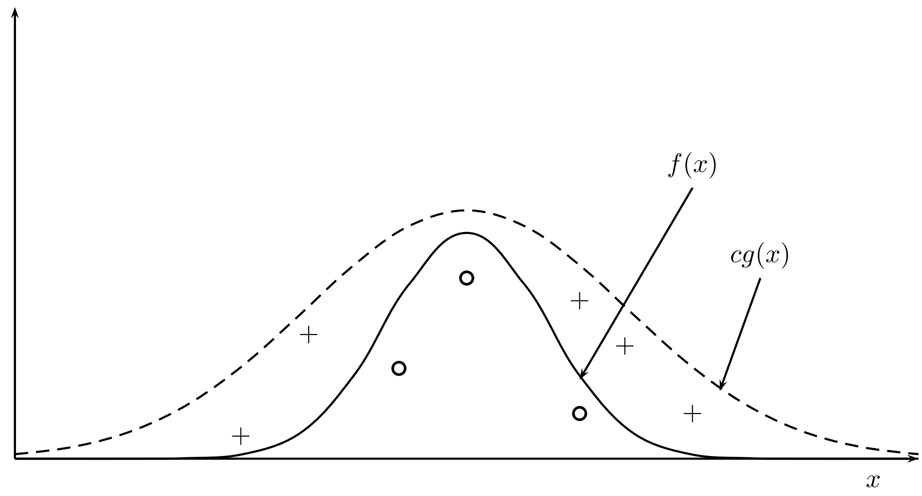
Generating random numbers, Part II

5.1 Von Neumann's acceptance-rejection algorithm for samples from a PDF f

- We would like to sample from a target distribution which has corresponding PDF f.
- Suppose there is a PDF g from which we know how to generate samples from and for which

$$f(x) \leq cg(x) \quad \text{for all } x \in \mathbb{R}, \tag{5.1}$$

  for a constant c.
- The idea is now to generate X from g and accept the sample with probability $\frac{f(X)}{cg(X)}$.



Simulation by acceptance-rejection-sampling. Graphical interpretation:

- Sample X_0 from g.
- Given this X_0 sample $\tilde{U}$ from a uniform distribution on $[0, cg(X_0)]$.
- If $\tilde{U} \leq f(X_0)$ (labeled as '$\circ$') accept X_0, otherwise (labeled as '$+$') reject X_0.

We can use the following modified algorithm:

Definition 73 (Acceptance-rejection algorithm)**.** Suppose condition (5.1) is satisfied. Then, the *Von Neumann's acceptance-rejection algorithm* is given by

1. Generate X from the PDF g.
2. Generate $U \sim \mathrm{Unif}(0, 1)$.
3. If $U \leq \frac{f(X)}{cg(X)}$, then accept X and return it. Otherwise, go back to step 1.

- Hence, we see that we first generate samples from another distribution (with PDF g) which might be easier to sample from and then reject some of those candidates.
- The rejection mechanism is designed such that the accepted candidates are indeed a sample from our target distribution.

Theorem 74. *The algorithm in Definition 73 works, i.e., it indeed returns a sample from the PDF f.*

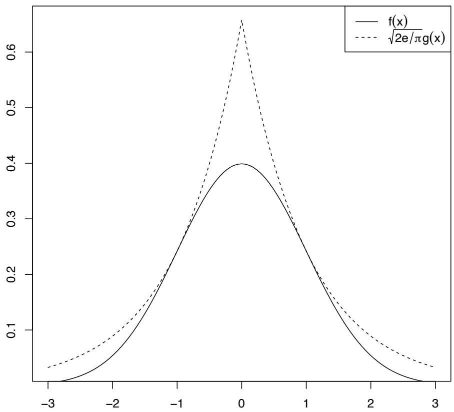
Example 75. We can use Von Neumann's acceptance-rejection algorithm to generate samples from the standard normal distribution using the doubly exponential distribution as follows.

- Consider the probability density function (PDF) of the standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } x \in \mathbb{R},$$

- and the PDF of the doubly exponential distribution

$$g(x) = \frac{1}{2} e^{-|x|} \quad \text{for } x \in \mathbb{R}.$$



Example (continued)**.**

- We next look for a constant $c \geq 1$ such that (5.1) is true.
- We calculate

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+|x|} \\ &= \sqrt{\frac{2}{\pi}} e^{-\frac{x^2-2|x|+1}{2}+\frac{1}{2}} \\ &= \sqrt{\frac{2e}{\pi}} e^{-\frac{(|x|-1)^2}{2}} \\ &\leq \sqrt{\frac{2e}{\pi}} =: c \approx 1.3155.\end{aligned}$$

Example (continued).

- Von Neumann's algorithm:

  1. Generate X from the doubly exponential PDF g and generate U from the uniform distribution on $[0, 1]$.
  2. If
  $$U > \frac{f(X)}{cg(X)} = e^{-\frac{(|X|-1)^2}{2}},$$
  then reject X and go back to step 1.
  3. Return X.

- The proportion of rejected samples is
$$1 - \frac{1}{c} = 1 - \sqrt{\frac{\pi}{2e}} \approx 23.98\%.$$

5.2 The Box-Muller method for generating samples from the standard normal distribution

Definition 76 (Box-Muller algorithm). The *Box-Muller algorithm* is defined as follows:

1. Generate independent random variables $U_1, U_2 \sim \mathrm{Unif}(0, 1)$;
2. Set $R = -2\log(U_1)$;
3. Set $\theta = 2\pi U_2$;
4. Set $X_1 = \sqrt{R}\cos(\theta)$;
5. Set $X_2 = \sqrt{R}\sin(\theta)$;
6. Return X_1 and X_2.

Theorem 77 (Box-Muller algorithm). *The algorithm in Definition 76 returns two independent standard normally distributed random variables.*

5.3 Recap

- Von Neumann's acceptance-rejection algorithm provides a tool how to sample from a specific distribution, if one only can sample from a different distribution.
- This algorithm requires that the target PDF is bounded by a constant times the original PDF.
- The Box-Muller algorithm yields a transformation, that turns two independent standard uniformly distributed random variables into two independent standard normally distributed random variables.

Chapter 6

Monte Carlo integration

6.1 Introduction to Monte Carlo integration

- Monte Carlo simulation is an extremely powerful technique, and there are many problems where it is the only reasonable approach currently available.
- It is a standard technique in option pricing (the pricing of option often reduces to computing an expectation).

- The **numerical evaluation of definite integrals** is one of the main applications that **Monte Carlo simulation** is concerned with.
- To fix ideas, let h be a given function and consider the integral

$$I = \int_0^1 h(s)\, ds.$$

- We can **view this integral as the expectation of the random variable** $X = h(U)$, where $U \sim \text{Unif}(0,1)$. Indeed,

$$I = E[X] = E[h(U)].$$

- This representation gives rise to the **Monte Carlo estimator** of I, which is given by

$$I_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} h(U_i),$$

where $U_1, \cdots, U_n$ are i.i.d. random variables from the Unif$(0,1)$ distribution and $X_i = h(U_i)$.

6.2 Monte Carlo estimators

Definition 78 (Monte Carlo estimator). Suppose that X is a random variable. Let $X_1, \cdots, X_n$ be i.i.d. random variables from the distribution of X. A *Monte Carlo estimator* of $\mu = E[X]$ is given by the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{6.1}$$

- The mean of $\bar{X}_n$ is

$$E\left[\bar{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X] = \mu,$$

  so this estimator is **unbiased**.

- Its variance is given by

$$\operatorname{Var}\left(\bar{X}_n\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(X_i) = \frac{1}{n}\operatorname{Var}(X). \tag{6.2}$$

- The so called *Strong Law of Large Numbers* implies that the Monte Carlo estimator $\bar{X}_n$ of $\mu = E[X]$ given by (6.1) is **consistent**, i.e.,

$$\lim_{n\to\infty} \bar{X}_n = \mu.$$

- The variance in (6.2) can be related to the **number of X_i needed to achieve a certain degree of accuracy**.

- A major challenge is to reduce the variance of the estimator, provided in (6.2). We will consider this problem below.

6.3 Recap

- Monte Carlo gives a way to compute expectations of the form $E[g(Y)]$, where Y is a random variable, which can be sampled.

- To do Monte Carlo, choose a large N, sample N realizations $(y_1, \cdots, y_N)$ of Y, compute $(x_1, \cdots, x_N) = (g(y_1), \cdots, g(y_N))$, and take the sample average.

- The variance of the Monte Carlo estimator is $\operatorname{Var}(g(Y))/n$.

Chapter 7

Introduction to option pricing - the one-period binomial asset pricing model

7.1 What are options?

- A **derivative security** is a security whose value depends on the values of basic underlying variables.
- Such basic underlying variables often are prices of stocks, interest rates, exchange rates etc.
- Derivative securities are also sometimes called **contingent claims**.
- Example: A stock option is a derivative security (here: an option) whose value is contingent on the price of a stock.

Examples of derivative securities:

- A **forward contract** is an **agreement to buy or sell** an asset at a certain future time for a certain price.
- An **option** is a special type of derivative security that gives the **holder the right but not the obligation** to do something, e.g., to buy or sell an asset at a certain future time for a certain price.

- Two basic types of options: **call options and put options**.
- A **call option** gives the holder the **right to buy** the underlying asset by a certain date for a certain price.
- A **put option** gives the holder the **right to sell** the underlying asset by a certain date for a certain price.

- Common names used for the date in the contract: *expiration date*, *exercise date*, **maturity**.
- Common names for price in the contract: *exercise price*, **strike price**.
- **American options** are option that can be **exercised at any time up to the maturity**.
- **European options** are options that can be **exercised only at the maturity date** itself.

- An **option** gives the holder **a right but not an obligation** to do something.
- Hence, the holder can choose not to exercise their right.
- The holder needs to pay for the right to do something.
- **What should the price of an option be?**

There are other derivative securities (e.g. forwards and futures) where the holder is obligated to buy or sell the underlying asset.

- We now focus on **stock options**.
- We need to **model the price of the underlying**, i.e., the stock.
- We do this using the **one-period binomial asset pricing model**.
- We need to develop a **concept for determining the price of an option**.
- We do this using the **principle of no-arbitrage and risk-neutral pricing**.

7.2 The one-period binomial model - model description

- We use the **one-period binomial asset pricing model** to introduce the concept of the **no-arbitrage price of an option**.
- One-period binomial asset pricing model considers two points in time: **Time zero** (=beginning of the period) and **time one** (=end of the period).
- The financial market consists of **two assets**: a **riskless asset** (bank account) and a **risky asset** (stock).
- We model **riskiness** of an asset by modelling it as a **random variable**.

Definition 79 (Riskless asset in one-period binomial model)**.** The time-0 price of the riskless asset is given by $B_0 = 1$ and the time-1 price by $B_1 = (1 + r)$ where $r > -1$.

Remark 80.

- You can think of the riskless asset as a bank account with interest rate r.
- Interest rate is the **same for investing and for borrowing**:
  - Investing £1 at time 0 yields £$(1 + r)$ at time 1.
  - Borrowing £1 at time 0 yields £$(1 + r)$ debt at time 1.

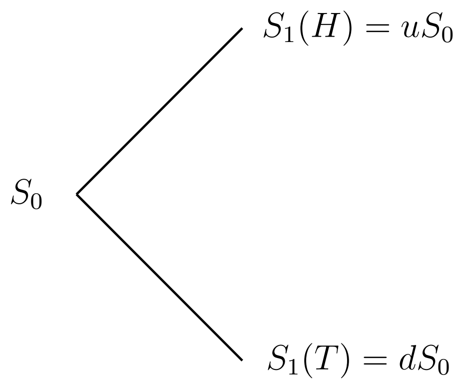
Definition 81 (Risky asset in one-period binomial model)**.** The time-0 price of the risky asset is given by S_0 where $S_0 > 0$. The time-1 price of the risky asset is given by the random variable S_1 defined on the (coin toss) sample space $\Omega = \{H, T\}$ (H=head, T=tail):

$$S_1(\omega) = \begin{cases} uS_0, & \text{if } \omega = H, \\ dS_0, & \text{if } \omega = T, \end{cases}$$

where $u > d > 0$. Furthermore, $P(S_1 = uS_0) = p$, $P(S_1 = dS_0) = 1 - p$, where $p \in (0, 1)$.

Remark 82.

- Price of stock at time 0 is known (S_0). Price of stock at time 1 (S_1) is unknown at time 0 but known at time 1.
- Intuition: Toss a coin and the outcome determines the price of the stock per share at time 1.
- Intuitively think of u as **up factor** and d as **down factor** (even though we do not require that $d < 1 < u$).



7.3 The concept of no-arbitrage

Important: Need to make sure that our mathematical model for the financial market does **not** allow for **arbitrage opportunities**.

Definition 83 (Arbitrage - intuitive definition)**.** A trading strategy is an **arbitrage** if it begins with zero money, has zero probability of losing money and has a positive probability of making money.

Definition 84 (Trading strategy in the one-period binomial model)**.** We refer to a tuple $\varphi = (\beta_0, \Delta_0)^\top \in \mathbb{R}^2$ as a **trading strategy** in the one-period binomial model, where β_0 denotes the **number of riskless assets held at time** 0 and Δ_0 denotes the **number of shares of the stock held at time** 0.

Definition 85 (Wealth corresponding to trading strategy)**.** Consider a trading strategy $\varphi = (\beta_0, \Delta_0)^\top$ in the one-period binomial model. The corresponding **time-0 wealth** is given by

$$X_0^\varphi = X_0 = \beta_0 B_0 + \Delta_0 S_0 = \beta_0 + \Delta_0 S_0,$$

and the corresponding **time-1 wealth** is given by

$$X_1^\varphi = X_1 = \beta_0 B_1 + \Delta_0 S_1 = \beta_0(1 + r) + \Delta_0 S_1.$$

Remark 86. We sometimes write X_t rather than X_t^φ for the wealth at time $t \in \{0, 1\}$ to simplify notation.

Definition 87 (Arbitrage in one-period binomial model)**.** Consider a trading strategy $\varphi = (\beta_0, \Delta_0)^\top$ in the one-period binomial model. We refer to φ as an **arbitrage** if

$$X_0^\varphi = 0, \quad P[X_1^\varphi \geq 0] = 1, \quad P[X_1^\varphi > 0] > 0. \tag{7.1}$$

Theorem 88. *Consider a one-period binomial model. Then, there is no arbitrage if and only if*

$$d < 1 + r < u. \tag{7.2}$$

Proof of Theorem 88.

First, assume that there is no arbitrage in the one-period binomial model. We show that $d < 1 + r < u$ by contradiction.

- Assume that $d \geq 1 + r$.
  - Time 0: Start with zero wealth $X_0 = 0$ by borrowing $\beta_0 = -S_0$ riskless assets and buying $\Delta_0 = 1$ share of stock. Hence, $\varphi = (\beta_0, \Delta_0)^\top = (-S_0, 1)$.
    Indeed $X_0^\varphi = \beta_0 B_0 + \Delta_0 S_0 = -S_0 + \Delta_0 S_0 = 0$.
  - Time 1: Wealth at time 1:
    Outcome Tail:

    $$\begin{aligned} X_1^\varphi(T) &= \beta_0 B_1 + \Delta_0 S_1(T) = -S_0(1+r) + dS_0 \\ &\geq -S_0(1+r) + (1+r)S_0 = 0. \end{aligned}$$

    Outcome Head:

    $$\begin{aligned} X_1^\varphi(H) &= \beta_0 B_1 + \Delta_0 S_1(H) = -S_0(1+r) + uS_0 \\ &\geq -S_0 d + uS_0 = (u-d)S_0 > 0. \end{aligned}$$

    This is a contradiction to the assumption that there is no arbitrage.

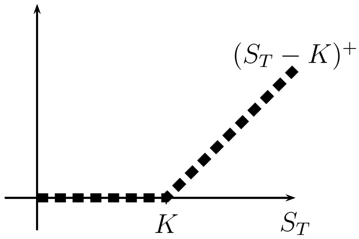
- Assume that $u \leq 1 + r$.
  - Time 0: Start with zero wealth $X_0 = 0$ by short-selling $\Delta_0 = -1$ share of stocks and investing the proceeds S_0 in the riskless asset, i.e., $\beta_0 = S_0$. Hence, $\varphi = (\beta_0, \Delta_0)^\top = (S_0, -1)$.
    Indeed $X_0^\varphi = \beta_0 B_0 + \Delta_0 S_0 = S_0 + \Delta_0 S_0 = 0$.
  - Time 1: Wealth at time 1:
    Outcome Tail:
$$\begin{aligned}X_1^\varphi(T) &= \beta_0 B_1 + \Delta_0 S_1(T) = S_0(1+r) - dS_0 \\ &\geq uS_0 - dS_0 = (u-d)S_0 > 0.\end{aligned}$$
    Outcome Head:
$$\begin{aligned}X_1^\varphi(H) &= \beta_0 B_1 + \Delta_0 S_1(H) = S_0(1+r) - uS_0 \\ &\geq S_0 u - uS_0 = 0.\end{aligned}$$
    This is a contradiction to the assumption that there is no arbitrage.

We prove the second direction (i.e., (7.2) implies that the one-period model is free of arbitrage), in the class.

□

7.4 Pricing a European call option in the one-period binomial model

- A **European call option** gives its owner the right but not the obligation to buy one share of stock at the **maturity date** T for the **strike price** K.
- Assumptions in the one-period binomial model: maturity $T = 1$ and strike K satisfies $S_1(T) < K < S_1(H)$.
- Outcome **Tail**: Option expires worthless. (It would be cheaper to buy the share on the stock market directly for $S_1(T)$ rather than exercising the option and paying K for it.)
- Outcome **Head**: The option will be exercised yielding profit $S_1(H) - K$.
- **Payoff or value of the option at time 1**: $(S_1 - K)^+$,
  (where we omit argument of random variable S_1).
  Notation: $x^+ = \max\{x, 0\}$.



- **How much is the European call option worth at time 0**, i.e., before we know the outcome of the coin toss?
- **Arbitrage pricing theory** approach to option pricing:
  - Replicate option by trading in stock and money market.
  - Price of replicating strategy = price of option.

Example 89 (European call option)**.**

- Consider one-period binomial model with $S_0 = 4$, $u = 2$, $d = \frac{1}{u} = \frac{1}{2}$, $r = \frac{1}{4}$. Hence, $S_1(H) = 8$, $S_1(T) = 2$ and in particular, $d = \frac{1}{2} < 1 + r = \frac{5}{4} < 2 = u$.
- Consider European call option with maturity $T = 1$ and strike price $K = 5$.
- **Option payoff at time 1**:
  - outcome head: $(S_1(H) - 5)^+ = 3$,
  - outcome tail: $(S_1(T) - 5)^+ = 0$.

Example (European call option (continued))**.**

- Consider the **investment strategy** $\varphi = (\beta_0, \Delta_0)^\top = (-0.8, 0.5)$, i.e.:
  - borrow £0.8 in riskless asset,
  - buy $\Delta_0 = 0.5$ shares of stock at time 0.
- Corresponding **time-0 wealth**: $X_0^\varphi = \beta_0 + \Delta_0 S_0 = -0.8 + \Delta_0 S_0 = -0.8 + 0.5 \cdot 4 = 1.2$.
- Corresponding **time-1 wealth**: $X_1^\varphi = -0.8(1 + 0.25) + 0.5 S_1 = -1 + 0.5 S_1$.
  - Outcome head: $X_1^\varphi(H) = -1 + 0.5 S_1(H) = 3$,
  - outcome tail: $X_1^\varphi(T) = -1 + 0.5 S_1(T) = 0$.

Example (European call option (continued))**.**

- The time-1 wealth corresponding to the trading strategy $\varphi = (-0.8, 0.5)^\top$ corresponds to the option payoff at time 1. Hence, we have **replicated** the option by trading in the stock and money markets.
- The initial time-0 wealth $X_0^\varphi = 1.2$ corresponding to the replicating strategy $\varphi = (-0.8, 0.5)^\top$ is the **no-arbitrage price of the option at time 0**.

Example (European call option (continued))**.**

- Why is the time-0 price of the European call indeed £1.2?
- We show that there exist an arbitrage if the price is either greater or smaller than £1.2.

Example (European call option (continued))**.** Consider the case where the price is larger, e.g. £1.21.

- Strategy: At time 0,
  - sell the option for £1.21,
  - use £1.2 to replicate the option (using $\varphi = (-0.8, 0.5)$),
  - invest the additional £0.01 in the riskless asset.
- At time 1, the seller can pay off the option using the replicating portfolio and in addition has gained £0.01 $\cdot$ 1.25 = £0.0125.
- Arbitrage! Seller needs no money initially and without risk of loss gains £0.0125 at time 1.

Example (European call option (continued)). Consider the case where the price is smaller, e.g. £1.19.

- Strategy: At time 0, reverse the replicating strategy, i.e., use $\widetilde{\varphi} = (0.8, -0.5)$:
  - short-sell 0.5 shares of stock which generates an income of £$0.5 \cdot 4$ =£2,
  - buy the option for £1.19,
  - invest £0.8 in the riskless asset,
  - invest the remaining £0.01 in a separate riskless asset account.
- At time 1:
  - outcome head: Option is worth £3, cash in the riskless asset has grown to £$(0.8+0.01) \cdot 1.25$ =£1.0125, i.e., it is £4.0125 in total available to replace the 0.5 shares of stocks for £$0.5 \cdot 8$ = £4.
  - outcome tail: Option is worth 0, investment in riskless asset is worth £1.0125 which is sufficient to replace the 0.5 shares of stock at a price of £$0.5 \cdot 2$ = £1.
- Arbitrage! Buyer needs no money initially and without risk of loss gains £0.0125 at time 1.

Key assumptions to derive the time-0 price:

- Shares of stocks can be subdivided (for sale or purchase),
- same interest rate for investing and borrowing,
- purchase price of stock = selling price of stock (no *bid-ask spread*),
- at any point in time, only two possible values for the stock price in the next period.

7.5 Pricing more general derivatives in the one-period binomial model

Definition 90. In the one-period binomial model we define a **derivative security** to be a security that pays $V_1(H)$ at time 1 if the outcome of the coin toss is head and $V_1(T)$ if the outcome of the coin toss is tail, i.e., V_1 is a random variable on the coin toss sample space.

Example 91 (Derivative securities).

- European call option: $V_1 = (S_1 - K)^+$.
- European put option: $V_1 = (K - S_1)^+$.
- Forward contract: $V_1 = S_1 - K$

Goal: Compute time-0 price of derivative security with payoff V_1 at time-1 by computing the time-0 wealth corresponding to the **replicating strategy**.

Definition 92 (Replicating strategy)**.** Consider the one-period binomial model with a derivative security V_1. We refer to a trading strategy $\varphi = (\beta_0, \Delta_0)^\top \in \mathbb{R}^2$ that satisfies

$$X_1^\varphi(\omega) = \beta_0(1+r) + \Delta_0 S_1(\omega) = V_1(\omega),$$

for all $\omega \in \Omega = \{H, T\}$, as a *replicating strategy* (for V_1).

- Goal: Determine trading strategy $\varphi = (\beta_0, \Delta_0)^\top$ such that

$$X_1^\varphi = \beta_0(1+r) + \Delta_0 S_1 = V_1.$$

- Since there are two possible outcomes for the stock, we obtain a system of two equations:

$$\begin{aligned} \beta_0(1+r) + \Delta_0 S_1(H) &= V_1(H), \\ \beta_0(1+r) + \Delta_0 S_1(T) &= V_1(T). \end{aligned} \tag{7.3}$$

  We need to solve (7.3) for β_0 and Δ_0.

We solve (7.3) for β_0 and Δ_0:

$$\begin{aligned} &\begin{cases} \beta_0(1+r) + \Delta_0 S_1(H) = V_1(H), \\ \beta_0(1+r) + \Delta_0 S_1(T) = V_1(T) \end{cases} \\ \Longleftrightarrow &\begin{cases} \beta_0 = \frac{1}{1+r}\left(V_1(H) - \Delta_0 S_1(H)\right), \\ \beta_0 = \frac{1}{1+r}\left(V_1(T) - \Delta_0 S_1(T)\right) \end{cases} \end{aligned} \tag{7.4}$$

Hence,

$$\begin{aligned} &\frac{1}{1+r}\left(V_1(H) - \Delta_0 S_1(H)\right) = \frac{1}{1+r}\left(V_1(T) - \Delta_0 S_1(T)\right) \\ \Longleftrightarrow &\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}. \end{aligned} \tag{7.5}$$

Then, plugging (7.5) into (7.4) yields

$$\begin{aligned} \beta_0 &= \frac{1}{1+r}\left(V_1(H) - \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} S_1(H)\right) \\ &= \frac{1}{1+r} \frac{V_1(H)S_1(T) - V_1(T)S_1(H)}{S_1(T) - S_1(H)}. \end{aligned}$$

Consider the trading strategy derived from (7.3), i.e., $\varphi = (\beta_0, \Delta_0)^\top$ with

$$\begin{aligned}\Delta_0 &= \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)},\\ \beta_0 &= \frac{1}{1+r} \frac{V_1(H)S_1(T) - V_1(T)S_1(H)}{S_1(T) - S_1(H)}.\end{aligned} \tag{7.6}$$

Then, the corresponding time-0 wealth is given by

$$\begin{aligned}X_0^\varphi = \beta_0 + \Delta_0 S_0 = \beta_0 + \frac{V_1(H) - V_1(T)}{S_0 u - S_0 d} S_0 &= \beta_0 + \frac{V_1(H) - V_1(T)}{u - d}\\ &= \frac{1}{1+r} \left(\frac{V_1(H)d - V_1(T)u}{d-u} + \frac{(1+r)(V_1(H) - V_1(T))}{u-d} \right)\\ &= \frac{1}{1+r} \left(V_1(H) \frac{1+r-d}{u-d} + V_1(T) \frac{u-1-r}{u-d} \right).\end{aligned} \tag{7.7}$$

We collect the results of the calculations in the following theorem.

Theorem 93. *Consider a one-period binomial model satisfying the no-arbitrage condition (7.2) and consider a derivative security paying V_1 at time 1. Then,*

1. *there exists a replicating strategy $\varphi = (\beta_0, \Delta_0)^\top$ such that $X_1^\varphi = V_1$.*
2. *In particular, the replicating strategy $\varphi = (\beta_0, \Delta_0)^\top$ is given by (7.6).*
3. *The time-0 price V_0 of the financial security is given by*

$$V_0 = X_0^\varphi = \frac{1}{1+r} \left(V_1(H) \frac{1+r-d}{u-d} + V_1(T) \frac{u-1-r}{u-d} \right). \tag{7.8}$$

Remark 94.

- Note that the time-0 price V_0 does not depend on the probability p of head occurring!
- The formula (7.6) for Δ_0 is sometimes referred to as *delta-hedging* formula.

Remark 95. The trading strategy given in (7.6) provides a hedge for the seller of the derivative security. The seller has a short position in the derivative security.

One could also derive a hedging strategy for the long position, i.e., for the buyer of the derivative security. The number of shares of a long position hedge is the negative of the Δ_0 given in (7.6). We will look at an example in an exercise in the class.

We consider an alternative representation of the pricing formula (7.8).

Lemma 96. *Suppose the no-arbitrage condition (7.2) holds and $u > d > 0$, define*

$$\tilde{p} = \frac{1+r-d}{u-d}. \tag{7.9}$$

Then $\tilde{p} \in (0,1)$ and $1 - \tilde{p} = \frac{u-1-r}{u-d}$.

Proving this lemma is an exercise that you will do in class.

Theorem 97. *Consider a one-period binomial model satisfying the no-arbitrage condition (7.2) and consider a derivative security paying V_1 at time 1. Then, the time-0 price V_0 of the financial security is given by*

$$V_0 = \frac{1}{1+r}\left(V_1(H)\tilde{p} + V_1(T)(1-\tilde{p})\right) = \tilde{E}\left(\frac{V_1}{B_1}\right), \tag{7.10}$$

where $\tilde{p}$ is given in (7.9) and $\tilde{E}$ is the expectation corresponding to the risk-neutral probability $\tilde{p}$.

Remark 98. The pricing formula (7.10) shows that the price of the derivative security can be expressed as an expectation of the discounted payoff $\frac{V_1}{B_1} = \frac{V_1}{1+r}$ under a probability measure under which head occurs with probability $\tilde{p}$ and tail occurs with probability $1-\tilde{p}$.
The probability $\tilde{p}$ is referred to as **risk-neutral probability** and formula (7.10) as **risk-neutral pricing formula**.
Again, the actual (real world) probability p of head occurring does not matter for the price.

The proof of Lemma 96 is given in the class.

Note that

$$\begin{aligned}
\tilde{E}\left[\frac{S_1}{B_1}\right] &= \frac{1}{1+r}\left(\tilde{p}S_1(H) + (1-\tilde{p})S_1(T)\right) \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d}uS_0 + \frac{u-1-r}{u-d}dS_0\right) \\
&= \frac{S_0}{1+r}\frac{u(1+r)-du+ud-(1+r)d}{u-d} \\
&= S_0.
\end{aligned} \tag{7.11}$$

In particular,

$$(1+r)S_0 = \tilde{p}S_1(H) + (1-\tilde{p})S_1(T).$$

Hence, under the risk-neutral probability the average rate of growth of the stock is equal to the average rate of growth of an investment in the money market account.

7.6 Recap

- The binomial asset pricing models considers a **financial market with two assets**:
  - a riskless asset with price $B_0 = 1$ and $B_1 = (1 + r)$ (where $r > -1$),
  - a risky asset with price $S_0 > 0$ and S_1 is a random variable taking two possible values uS_0 with probability $p \in (0, 1)$ and dS_0 with probability $1 - p$. Assumption: $u > d > 0$.
- We introduced the concept of an **arbitrage**, which is a trading strategy that begins with zero money, trades in the two available assets and at time 1 with positive probability makes money without any possibility of losing money.
- We have seen that the one-period binomial asset pricing model is **free of arbitrage if and only if** $d < 1 + r < u$.

- Idea of **arbitrage pricing theory**: Price of a derivative security must be such that one cannot form an arbitrage by trading in the two underlying assets (the riskless and the risky assets) and the derivative security.
- Idea: **Replicate the payoff of the derivative security at time** 1 **by a portfolio** that trades in the riskless and the risky assets. By the no-arbitrage assumption, the **time-0 price** of the derivative security must be exactly the **price of the replicating strategy at time** 0.
- The one-period binomial model is **complete**, i.e., every derivative security can be replicated by trading in the two underlying assets.
- In the one-period binomial model the **time-0 price** of a derivative security is **uniquely determined** by the time-0 price of the replicating portfolio.
- **Risk-neutral pricing formula**: The time-0 price of a derivative security can be expressed as an expectation of the discounted payoff under the risk-neutral probability measure, see formula (7.10).

- Reading: This chapter followed closely (Shreve, 2004, Section 1.1).

Chapter 8

The multiperiod binomial asset pricing model, Part I

8.1 Model description

We extend the one-period binomial model to the N-period binomial model where $N \in \mathbb{N}$.

Definition 99 (Riskless asset in N-period binomial model)**.** The **price of the riskless asset** at time n is given by $B_n = (1+r)^n$ for all $n \in \{0, 1, \ldots, N\}$ where $r > -1$.

- Note that $B_0 = 1$.
- The initial stock price is $S_0 > 0$.
- Again we assume $0 < d < u$ and these constants model the "up" and "down" movements of the stock price.
- The **price of the stock at time** 1 is $S_1(H) = uS_0$ if the first coin toss is head and $S_1(T) = dS_0$ otherwise.
- The **price of the stock at time** 2 (after two coin tosses) is:

$$S_2(HH) = uS_1(H) = u^2 S_0, \qquad S_2(HT) = dS_1(H) = duS_0,$$
$$S_2(TH) = uS_1(T) = udS_0, \qquad S_2(TT) = dS_1(T) = d^2 S_0.$$

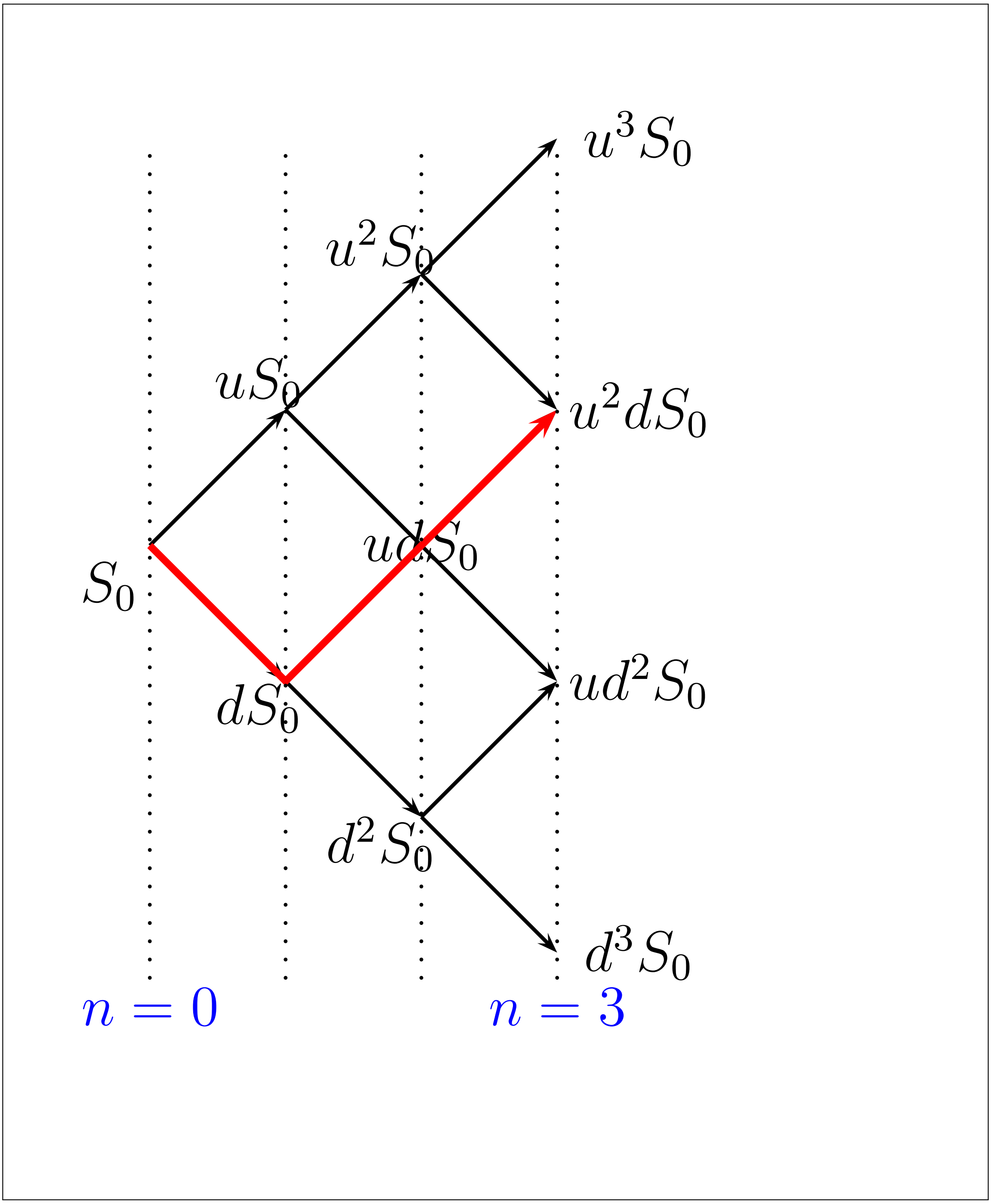
Definition 100 (Risky asset in the N-period binomial model)**.** Let $S_0 > 0$ be the price of the risky asset at time 0. The **price of the risky asset at time** $n \in \{1, \dots, N\}$, denoted by S_n is a **random variable** and depends on the outcome of the first n coin tosses $(\omega_1 \dots \omega_n)$ where for all $i \in \{1, \dots, N\}$ $\omega_i \in \{H, T\}$. In particular,

$$S_n(\omega_1 \dots \omega_n) = S_0 c_{\omega_1} \cdot \ldots \cdot c_{\omega_n}, \tag{8.1}$$

where for all $i \in \{1, \dots, N\}$

$$c_{\omega_i} = \begin{cases} u, & \text{if } \omega_i = H, \\ d, & \text{if } \omega_i = T. \end{cases}$$

Remark 101. Observe that there are $n + 1$ **possible distinct outcomes** for the time-n price of the risky asset.



$u^3 S_0$
$u^2 S_0$
$u S_0$
$u^2 d S_0$
S_0
$ud S_0$
$d S_0$
$u d^2 S_0$
$d^2 S_0$
$d^3 S_0$
$n = 0$
$n = 3$

8.2 Pricing derivatives in the multi-period binomial model

- We now want to compute the time-0 price of a **European call option** in the **two-period binomial model**.
- We use the same idea as in the one-period model: We derive a replicating portfolio $\varphi = (\varphi_t)_{t\in\{0,1\}}$, where $\varphi_t = (\beta_t, \Delta_t)^\top$, $t \in \{0,1\}$.
- As before, β_0 **and** Δ_0 **are constants, but** β_1 **and** Δ_1 **are random variables**, i.e., they will depend on the outcome of the first coin toss.
- In the following we will always assume that the no-arbitrage condition (7.2) is satisfied, i.e., that $d < 1 + r < u$.

- At time 0 the wealth is given by $X_0^\varphi = \beta_0 + \Delta_0 S_0$.
- At time 1 the wealth is given by $X_1^\varphi = \beta_0(1+r) + \Delta_0 S_1$.
- We allow **rebalancing of the portfolio at time** 1, i.e., new numbers of riskless and risky assets β_1, Δ_1 can be chosen, but only such that the current wealth stays the same, i.e., we require

$$X_1^\varphi = \beta_0(1+r) + \Delta_0 S_1 = \beta_1(1+r) + \Delta_1 S_1, \tag{8.2}$$

  i.e., φ is a **self-financing strategy**.
- At time 2 the wealth is given by $X_2^\varphi = \beta_1(1+r)^2 + \Delta_1 S_2$.
- **Goal: Find** β_0**,** Δ_0**,** $\beta_1(H)$**,** $\beta_1(T)$**,** $\Delta_1(H)$**,** $\Delta_1(T)$**.**

Strategy: First solve for $\beta_1(H)$, $\beta_1(T)$, $\Delta_1(H)$, $\Delta_1(T)$ by using the replication argument:

$$X_2^\varphi = \beta_1(1+r)^2 + \Delta_1 S_2 \overset{!}{=} V_2.$$

This in fact leads to 4 linear equations in 4 unknowns:

$$\beta_1(H)(1+r)^2 + \Delta_1(H)S_2(HH) = V_2(HH), \tag{8.3}$$
$$\beta_1(H)(1+r)^2 + \Delta_1(H)S_2(HT) = V_2(HT), \tag{8.4}$$
$$\beta_1(T)(1+r)^2 + \Delta_1(T)S_2(TH) = V_2(TH), \tag{8.5}$$
$$\beta_1(T)(1+r)^2 + \Delta_1(T)S_2(TT) = V_2(TT). \tag{8.6}$$

We can solve (8.3) and (8.4) for $\beta_1(H)$ and $\Delta_1(H)$ and obtain

$$\begin{aligned}
\Delta_1(H) &= \frac{V_2(HH) - V_2(HT)}{S_2(HH) - S_2(HT)}, \\
\beta_1(H) &= \frac{1}{(1+r)^2} \frac{V_2(HT)S_2(HH) - V_2(HH)S_2(HT)}{S_2(HH) - S_2(HT)}.
\end{aligned}$$

We can solve (8.5) and (8.6) for $\beta_1(T)$ and $\Delta_1(T)$ and obtain

$$\begin{aligned}
\Delta_1(T) &= \frac{V_2(TH) - V_2(TT)}{S_2(TH) - S_2(TT)}, \\
\beta_1(T) &= \frac{1}{(1+r)^2} \frac{V_2(TT)S_2(TH) - V_2(TH)S_2(TT)}{S_2(TH) - S_2(TT)}.
\end{aligned}$$

Strategy: Second compute $X_1^\varphi(H), X_1^\varphi(T)$ using the just derived $\beta_1(H)$, $\beta_1(T)$, $\Delta_1(H)$, $\Delta_1(T)$ and then solve for β_0, Δ_0 using the self-financing condition (8.2):

$$\begin{aligned}
X_1^\varphi(H) &= \beta_1(H)(1+r) + \Delta_1(H)S_1(H) = \beta_0(1+r) + \Delta_0 S_1(H), \\
X_1^\varphi(T) &= \beta_1(T)(1+r) + \Delta_1(T)S_1(T) = \beta_0(1+r) + \Delta_0 S_1(T).
\end{aligned}$$

We obtain

$$\begin{aligned}
\Delta_0 &= \frac{X_1^\varphi(H) - X_1^\varphi(T)}{S_1(H) - S_1(T)}, \\
\beta_0 &= \frac{1}{1+r} \frac{X_1^\varphi(T)S_1(H) - X_1^\varphi(H)S_1(T)}{S_1(H) - S_1(T)}.
\end{aligned}$$

We rewrite the expressions of $X_1^\varphi(H)$ and $X_1^\varphi(T)$ to see that they can be interpreted as expectations under the risk-neutral probability:

$$\begin{aligned}
X_1^\varphi(H) &= \beta_1(H)(1+r) + \Delta_1(H)S_1(H) \\
&= \frac{1}{(1+r)^2} \frac{V_2(HT)S_2(HH) - V_2(HH)S_2(HT)}{S_2(HH) - S_2(HT)}(1+r) \\
&\quad + \frac{V_2(HH) - V_2(HT)}{S_2(HH) - S_2(HT)} S_1(H) \\
&= \frac{1}{1+r} \frac{V_2(HT)u^2 - V_2(HH)ud}{u^2 - ud} + \frac{V_2(HH) - V_2(HT)}{u^2 - ud} u \\
&= \frac{1}{1+r} \frac{V_2(HT)u - V_2(HH)d}{u - d} + \frac{V_2(HH) - V_2(HT)}{u - d} \\
&= \frac{1}{1+r} \left(V_2(HH) \frac{1+r-d}{u-d} + V_2(HT) \frac{u-1-r}{u-d} \right) \\
&= \frac{1}{1+r} \left(V_2(HH)\tilde{p} + V_2(HT)(1-\tilde{p}) \right),
\end{aligned}$$

$$
\begin{aligned}
X_1^{\varphi}(T) &= \beta_1(T)(1+r) + \Delta_1(T)S_1(T) \\
&= \frac{1}{1+r}\frac{V_2(TT)du - V_2(TH)d^2}{du-d^2} + \frac{V_2(TH)-V_2(TT)}{du-d^2}d \\
&= \frac{1}{1+r}\frac{V_2(TT)u - V_2(TH)d}{u-d} + \frac{V_2(TH)-V_2(TT)}{u-d} \\
&= \frac{1}{1+r}\left(V_2(TH)\frac{1+r-d}{u-d} + V_2(TT)\frac{u-1-r}{u-d}\right) \\
&= \frac{1}{1+r}\left(V_2(TH)\tilde{p} + V_2(TT)(1-\tilde{p})\right).
\end{aligned}
$$

Note that by the idea of replication the time-1 price of the derivative is given by $V_1 = X_1^{\varphi}$. In particular,

$$
V_1(\omega_1) = X_1^{\varphi}(\omega_1) = \frac{1}{1+r}\left(V_2(\omega_1 H)\tilde{p} + V_2(\omega_1 T)(1-\tilde{p})\right), \tag{8.7}
$$

where $\omega_1 \in \{H, T\}$ is the outcome of the first coin toss.
Note that the time-1 price is a random variable.

Then the time-0 price of V_2 is given by

$$
\begin{aligned}
V_0 = X_0^{\varphi} &= \beta_0 + \Delta_0 S_0 && (8.8) \\
&= \frac{1}{1+r}\frac{X_1^{\varphi}(T)S_1(H) - X_1^{\varphi}(H)S_1(T)}{S_1(H)-S_1(T)} + \frac{X_1^{\varphi}(H)-X_1^{\varphi}(T)}{S_1(H)-S_1(T)}S_0 \\
&= \frac{1}{1+r}\frac{X_1^{\varphi}(T)u - X_1^{\varphi}(H)d}{u-d} + \frac{X_1^{\varphi}(H)-X_1^{\varphi}(T)}{u-d} \\
&= \frac{1}{1+r}\left(X_1^{\varphi}(H)\frac{1+r-d}{u-d} + X_1^{\varphi}(T)\frac{u-1-r}{u-d}\right) \\
&= \frac{1}{1+r}\left(X_1^{\varphi}(H)\tilde{p} + X_1^{\varphi}(T)(1-\tilde{p})\right) && (8.9)
\end{aligned}
$$

Since $X_1^{\varphi} = V_1$ we obtain:

$$
\begin{aligned}
V_0 &= \frac{1}{1+r}\left(V_1(H)\tilde{p} + V_1(T)(1-\tilde{p})\right) && (8.10) \\
&= \frac{1}{1+r}\left(\frac{1}{1+r}\left(V_2(HH)\tilde{p} + V_2(HT)(1-\tilde{p})\right)\tilde{p}\right. \\
&\qquad \left. + \frac{1}{1+r}\left(V_2(TH)\tilde{p} + V_2(TT)(1-\tilde{p})\right)(1-\tilde{p})\right) \\
&= \frac{1}{(1+r)^2}\left(V_2(HH)\tilde{p}^2 + V_2(HT)\tilde{p}(1-\tilde{p})\right. \\
&\qquad \left. + V_2(TH)\tilde{p}(1-\tilde{p}) + V_2(TT)(1-\tilde{p})^2\right).
\end{aligned}
$$

Note that the time-0 price is deterministic.

- The ideas presented in the one- and two-period binomial model generalise to N-periods.
- In particular, from (8.7) and (8.10) we see that the time-n price V_n can be computed from V_{n+1} for all $n \in \{0, 1, \ldots, N-1\}$. Hence, we can compute V_N from the payoff of the option and then V_{N-1}, $V_{N-2}, \ldots V_0$.

We summarise the insights in the following two theorems.

Definition 102 (Arbitrage in N-period binomial model)**.** Consider an N-period binomial model and a trading strategy $\varphi = \big((\beta_n, \Delta_n)^\top\big)_{n\in\{0,\ldots,N-1\}}$, where β_n denotes the numbers of riskless assets and Δ_n the number of risky assets held at time n.

1. The trading strategy φ is referred to as **self-financing** if and only if

$$\beta_{n-1}B_n + \Delta_{n-1}S_n = \beta_n B_n + \Delta_n S_n$$

   for all $n \in \{1, \ldots, N-1\}$.

2. Assume that φ is a self-financing trading strategy. We refer to φ as an **arbitrage** if

$$X_0^\varphi = 0, \quad P(X_N^\varphi \geq 0) = 1, \quad P(X_N^\varphi > 0) > 0. \tag{8.11}$$

Theorem 103. *Consider a multi-period binomial model. Then, there is no arbitrage if and only if*

$$d < 1 + r < u. \tag{8.12}$$

One can generalise the ideas used in the proof of the one-dimensional binomial model to prove the result for the multi-period model. We will omit the proof here.

Theorem 104. *Consider an N-period binomial model satisfying the no-arbitrage condition (8.12) and consider a derivative security paying V_N at time N. Then,*

1. *there exists a replicating self-financing strategy $\varphi = (\varphi_n)_{n\in\{0,\ldots,N-1\}}$ such that $X_n^\varphi = V_n$ for all $n \in \{0, \ldots, N\}$.*
2. *For $n \in \{0, \ldots, N-1\}$ V_n can be derived recursively backwards in time by setting*

$$V_n(\omega_1 \ldots \omega_n) = \frac{1}{1+r}\left(\tilde{p}V_{n+1}(\omega_1 \ldots \omega_n H) + (1-\tilde{p})V_{n+1}(\omega_1 \ldots \omega_n T)\right). \tag{8.13}$$

Theorem (Theorem 104 continued)**.**

3. In particular, the replicating strategy at time $n \in \{0, \ldots, N-1\}$ $\varphi_n = (\beta_n, \Delta_n)^\top$ is given by

$$\Delta_n(\omega_1, \ldots, \omega_n) = \frac{V_{n+1}(\omega_1 \ldots \omega_n H) - V_{n+1}(\omega_1 \ldots \omega_n T)}{S_{n+1}(\omega_1 \ldots \omega_n H) - S_{n+1}(\omega_1 \ldots \omega_n T)}, \tag{8.14}$$

$$\beta_n(\omega_1, \ldots, \omega_n) = \frac{V_{n+1}(\omega_1 \ldots \omega_n T)u - V_{n+1}(\omega_1 \ldots \omega_n H)d}{B_{n+1}(u-d)}. \tag{8.15}$$

4. The time-0 price V_0 can be written as

$$V_0 = \sum_{\omega=(\omega_1 \ldots \omega_N) \in \Omega} \tilde{p}_{\omega_1} \cdot \ldots \tilde{p}_{\omega_N} \frac{V_N(\omega)}{B_N} =: \tilde{E}\left(\frac{V_N}{B_N}\right), \tag{8.16}$$

where for all $n \in \{1, \ldots, N\}$

$$\tilde{p}_{\omega_n} = \begin{cases} \tilde{p}, & \text{if } \omega_n = H, \\ 1-\tilde{p}, & \text{if } \omega_n = T, \end{cases}$$

where $\tilde{p}$ is given in (7.9).

Theorem 104 can be proved by induction and using the same arguments we developed in the one-period and two-period model we will therefore omit it. You can find it in (Shreve, 2004, p. 13–14 (Proof of Theorem 1.2.2)).

Remark 105. The pricing formula (8.16) shows that the price of the derivative security can be expressed as an expectation of the discounted payoff $\frac{V_N}{B_N}$ under a **risk-neutral** probability. Formula (8.16) is referred to as **risk-neutral pricing formula**.
Again, the actual (real world) probability p of head occurring does not matter for the price.

The next result is an immediate consequence of (8.16) in Theorem 104:

Corollary 106. *Consider an N-period binomial model satisfying the no-arbitrage condition (8.12) and consider a derivative security paying $V_N = v(S_N)$ at time N, where v is a deterministic function and S_N is the stock price at time N. Then, the time-0 price of $V_N = v(S_N)$ is given by*

$$V_0 = \tilde{E}\left[\frac{v(S_N)}{B_N}\right] = \frac{1}{B_N}\sum_{k=0}^{N}\binom{N}{k}\tilde{p}^k(1-\tilde{p})^{N-k}v(S_0 u^k d^{N-k}), \tag{8.17}$$

where $\tilde{p} = \frac{1+r-d}{u-d}$.

Definition 107. For $n = 1, \ldots, N$, the **price of the derivative security at time** n if the outcome of the first coin tosses is $\omega_1 \ldots \omega_n$ is defined to be the random variable $V_n(\omega_1 \ldots \omega_n)$ of Theorem 104. The price of the derivative security at time 0 is defined to be V_0 as given in Theorem 104.

Remark 108 (Completeness)**.** The N-period binomial model is **complete** since every derivative security can be replicated by trading in the underlying riskless and risky asset.

In the following we consider an example of a **derivative security** whose payoff does not only depend on the stock price at the maturity date but **on stock prices at previous points in time** as well.
Such derivative securities are called **path-dependent**.

Example 109 (Lookback option)**.** Consider a 3-period binomial model with $S_0 = 4$, $u = 2$, $d = \frac{1}{2}$, $r = \frac{1}{4}$. Consider a **lookback option** that has payoff

$$V_3 = \max_{n \in \{0,1,2,3\}} S_n - S_3.$$

We compute its time-0 price V_0 via backward recursion.
Note that in this example

$$\tilde{p} = \frac{1 + r - d}{u - d} = \frac{\frac{5}{4} - \frac{1}{2}}{\frac{3}{2}} = \frac{1}{2} = 1 - \tilde{p}.$$

Example (Lookback option (continued))**.** V_3 is determined from the payoff directly:

$$\begin{aligned}
V_3(HHH) &= S_3(HHH) - S_3(HHH) = 32 - 32 = 0, \\
V_3(HHT) &= S_2(HH) - S_3(HHT) = 16 - 8 = 8, \\
V_3(HTH) &= S_1(H) - S_3(HTH) = 8 - 8 = 0, \\
V_3(THH) &= S_3(THH) - S_3(THH) = 8 - 8 = 0, \\
V_3(HTT) &= S_1(H) - S_3(HTT) = 8 - 2 = 6, \\
V_3(THT) &= S_2(TH) - S_3(THT) = 4 - 2 = 2, \\
V_3(TTH) &= S_0 - S_3(TTH) = 4 - 2 = 2, \\
V_3(TTT) &= S_0 - S_3(TTT) = 4 - 0.5 = 3.5.
\end{aligned}$$

Example (Lookback option (continued))**.** For any outcome of the first two coin tosses $\omega_1\omega_2$ the time-2 value V_2 is determined from the backward recursion:

$$\begin{aligned}
V_2(\omega_1\omega_2) &= \frac{1}{1+r} \left(\tilde{p} V_3(\omega_1\omega_2 H) + (1 - \tilde{p}) V_3(\omega_1\omega_2 T)\right) \\
&= \frac{4}{5} \left(\frac{1}{2} V_3(\omega_1\omega_2 H) + \frac{1}{2} V_3(\omega_1\omega_2 T) \right) \\
&= \frac{2}{5} \left(V_3(\omega_1\omega_2 H) + V_3(\omega_1\omega_2 T) \right).
\end{aligned}$$

Example (Lookback option (continued))**.** Hence,

$$\begin{aligned}
V_2(HH) &= \frac{2}{5}(V_3(HHH) + V_3(HHT)) = \frac{2}{5}(0+8) = \frac{16}{5} = 3.2,\\
V_2(HT) &= \frac{2}{5}(V_3(HTH) + V_3(HTT)) = \frac{2}{5}(0+6) = \frac{12}{5} = 2.4,\\
V_2(TH) &= \frac{2}{5}(V_3(THH) + V_3(THT)) = \frac{2}{5}(0+2) = \frac{4}{5} = 0.8,\\
V_2(TT) &= \frac{2}{5}(V_3(TTH) + V_3(TTT)) = \frac{2}{5}(2+\frac{7}{2}) = \frac{11}{5} = 2.2.
\end{aligned}$$

Example (Lookback option (continued))**.** For any outcome of the first coin toss ω_1 the time-1 value V_1 is determined from the backward recursion:

$$\begin{aligned}
V_1(\omega_1) &= \frac{1}{1+r}\left(\tilde{p}V_2(\omega_1 H) + (1-\tilde{p})V_2(\omega_1 T)\right)\\
&= \frac{4}{5}\left(\frac{1}{2}V_2(\omega_1 H) + \frac{1}{2}V_2(\omega_1 T)\right)\\
&= \frac{2}{5}\left(V_2(\omega_1 H) + V_2(\omega_1 T)\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
V_1(H) &= \frac{2}{5}(V_2(HH) + V_2(HT)) = \frac{2}{5}(3.2+2.4) = 2.24,\\
V_1(T) &= \frac{2}{5}(V_2(TH) + V_2(TT)) = \frac{2}{5}(0.8+2.2) = 1.2.
\end{aligned}$$

Example (Lookback option (continued))**.** Finally, the time-0 price of the lookback option is given by

$$V_0 = \frac{1}{1+r}\left(\tilde{p}V_1(H) + (1-\tilde{p})V_1(T)\right) = \frac{2}{5}(2.24+1.2) = 1.376.$$

Example (Lookback option (continued))**.** Consider the following hedging strategy for the seller of the lookback option:

- Sell the lookback option at time 0 for 1.376.
- Buy

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} = \frac{2.24 - 1.2}{8 - 2} = \frac{13}{75} \approx 0.1733$$

  shares of stocks at a price of $S_0 = 4$ per share, i.e., the total costs are £$\frac{52}{75} \approx 0.6933$ and invest the remaining $1.376 - \frac{52}{75} \approx 0.6827$ in the riskless asset. Note that indeed

$$\beta_0 = \frac{V_1(T)u - V_1(H)d}{(1+r)(u-d)} = \frac{1.2 \cdot 2 - 2.24 \cdot 0.5}{\frac{15}{8}} = \frac{256}{375} \approx 0.6827.$$

Example (Lookback option (continued))**.**

- At time 1 the investment in the riskless asset is worth $\frac{256}{375} \cdot 1.25 = \frac{64}{75} \approx 0.8533$.
- If the stock price goes up to 8 the position in the stock is worth $\frac{13}{75} \cdot 8 = \frac{104}{75} \approx 1.3867$ and hence the portfolio value is $\frac{64}{75} + \frac{104}{75} = \frac{168}{75} = 2.24 = V_1(H)$.
- If the stock price goes down to 2 the position in the stock is worth $\frac{13}{75} \cdot 2 = \frac{26}{75} \approx 0.3467$ and hence the portfolio value is $\frac{64}{75} + \frac{26}{75} = \frac{90}{75} = 1.2 = V_1(T)$.
- Continuing like this we see that the seller of the lookback option has a portfolio worth V_3 at time 3 no matter what the outcomes of the three coin tosses are.

8.3 Computational aspects in the binomial model, Part I

- **In practice**, one considers binomial models with **100 or more periods**.
- There are $2^{100} \approx 10^{30}$ possible outcomes for a sequence of 100 coin tosses.
- We cannot use an algorithm that starts by tabulating 2^{100} values for V_{100}, but need to find a computationally efficient approach.

We illustrate how the implementation of the binomial model can be set up in a more efficient manner by looking at two examples.

Example 110 (European put option)**.**

- Market model: We consider the **3-period binomial model** with $S_0 = 4$, $u = 2$, $d = \frac{1}{2}$, $r = \frac{1}{4}$.
- Derivative security: We would like to compute the time-0 price of a **European put option** with strike $K = 5$ and maturity $T = 3$, its payoff at time 3 is given by $V_3 = (K - S_3)^+ = \max\{0, K - S_3\}$. In particular,

$$\begin{aligned} V_3(HHH) &= 0, \\ V_3(HHT) = V_3(HTH) = V_3(THH) &= 0, \\ V_3(TTH) = V_3(THT) = V_3(HTT) &= 3, \\ V_3(TTT) &= 4.5. \end{aligned}$$

- There are $2^3 = 8$ possible outcomes for 3 coin tosses but they do not lead to 8 distinct stock values or option payoffs (see also Remark 101 and Corollary 106).

Example (European put option (continued))**.**

- Notation: $v_3(s)$ denotes the payoff of the option at time three for a time-3 stock price s. (V_3 has sequence of three coin tosses as arguments, v_3 has stock price as argument.)
- At time 3: Only four possible stock values: 32, 8, 2, 0.5 and therefore

$$\begin{aligned} v_3(32) &= (5 - 32)^+ = 0, \\ v_3(8) &= 0, \\ v_3(2) &= 3, \\ v_3(0.5) &= 4.5. \end{aligned}$$

- Note that in a 100-period binomial model there are 2^{100} possible outcomes of the coin tosses, but only 101 possible stock prices at time 100.
  Huge reduction in complexity to consider v rather than V!

Example (European put option (continued)).

- From Theorem 104 we know that

$$V_2(\omega_1\omega_2) = \frac{2}{5}\left(V_3(\omega_1\omega_2 H) + V_3(\omega_1\omega_2 T)\right), \tag{8.18}$$

  which represents 4 equations for 4 different choices of $\omega_1\omega_2$ (HH, HT, TH, TT).

- We write $v_2(s)$ for the price of the European put option where s is the stock price at time 2. Then formula (8.18) can be written as

$$v_2(s) = \frac{2}{5}\left(v_3(2s) + v_3(0.5s)\right),$$

  and hence there are only 3 equations corresponding to three stock prices at time 2:

Example (European put option (continued)).

$$\begin{aligned} v_2(16) &= \frac{2}{5}\left(v_3(32) + v_3(8)\right) = 0, \\ v_2(4) &= \frac{2}{5}\left(v_3(8) + v_3(2)\right) = \frac{6}{5} = 1.2, \\ v_2(1) &= \frac{2}{5}\left(v_3(2) + v_3(0.5)\right) = 3. \end{aligned}$$

Example (European put option (continued)).

- Again, from Theorem 104 we know that

$$V_1(\omega_1) = \frac{2}{5}\left(V_2(\omega_1 H) + V_2(\omega_2 T)\right), \tag{8.19}$$

  which represents 2 equations for 2 different choices of ω_1 (H, T).

- We write $v_1(s)$ for the price of the European put option where s is the stock price at time 1. Then formula (8.19) can be written as

$$v_1(s) = \frac{2}{5}\left(v_2(2s) + v_2(0.5s)\right),$$

  and hence there are 2 equations corresponding to two stock prices at time 1;

$$\begin{aligned} v_1(8) &= \frac{2}{5}\left(v_2(16) + v_2(4)\right) = \frac{12}{25} = 0.48, \\ v_1(2) &= \frac{2}{5}\left(v_2(4) + v_2(1)\right) = \frac{42}{25} = 1.68 \end{aligned}$$

Example (European put option (continued)).

- Again, from Theorem 104 we know that

$$V_0 = \frac{2}{5}\left(V_1(H) + V_1(T)\right), \tag{8.20}$$

  which represents only one equation.

- We write $v_0(s)$ for the price of the European put option where s is the stock price at time 0. Then formula (8.20) can be written as

$$v_0(s) = \frac{2}{5}\left(v_1(2s) + v_1(0.5s)\right).$$

  Since the initial stock price is $S_0 = 4$, the price of the European put option at time 0 is

$$v_0(4) = \frac{2}{5}\left(v_1(8) + v_1(2)\right) = \frac{2}{5}(0.48 + 1.68) = 0.864.$$

Example (European put option (continued)).

- Similarly, the replicating strategy can also be characterised as a function of the stock price.

- The number of shares of stock that should be held in the replicating portfolio at time n if the stock price at time n is s is given by

$$\tilde{\Delta}_n(s) = \frac{v_{n+1}(2s) - v_{n+1}(0.5s)}{2s - 0.5s},$$

  which is the analogue of formula (8.14) in Theorem 104.

- The number of riskless assets that should be held in the replicating portfolio at time n if the stock price at time n is s is given by

$$\tilde{\beta}_n(s) = \frac{2v_{n+1}(0.5s) - 0.5v_{n+1}(2s)}{\left(\frac{5}{4}\right)^{n+1}\frac{3}{2}}.$$

  which is the analogue of formula (8.15) in Theorem 104.

- Key idea in Example 110 to reduce the complexity of the computations:
  Since in this example the **option price at time** n **only depends on the stock price at time** n we can write

$$V_n = v(S_n),$$

  i.e., we can **relate the random variable** V_n **to the random variable** S_n **via a deterministic function** v_n.

- Generalisations of this idea are possible even if the **payoff does not just depend on the current stock price but on the whole path of the stock price** as we will see in the next example.

8.4 Recap

- The N-period binomial asset pricing models considers a **financial market with two assets**:
  - a riskless asset with price $B_n = (1+r)^n$ (where $r > -1$) at time $n \in \{0, \dots, N\}$,
  - a risky asset with price $S_0 > 0$ and S_N is a random variable given in (9.5) and hence depends on the outcome of N coin tosses. At any point in time the stock price can only move up by a factor u or down by a factor d and this is determined by the outcome of the coin tosses. Assumption: $u > d > 0$.
- For trading strategies in the multiperiod binomial model we introduced the additional concept of a trading strategy being **self-financing**. A self-financing trading strategy rebalances a portfolio after observing the new asset prices at the next point in time in such a way that only the composition of the portfolio but not the total wealth corresponding to the portfolio changes. In particular, no additional money is used to invest and also no money is taken out of the portfolio.

- When defining arbitrage in multiperiods we only consider self-financing strategies, but otherwise the concept of an arbitrage stays the same as in the one-period model.
- The multiperiod binomial asset pricing model is **free of arbitrage if and only if** $d < 1 + r < u$.

- Idea of **arbitrage pricing theory**: Price of a derivative security must be such that one cannot form an arbitrage by trading in the two underlying assets (the riskless and the risky assets) and the derivative security.
- Idea: **Replicate the payoff of the derivative security at time N by a portfolio** that trades in the riskless and the risky assets. By the no-arbitrage assumption, the **time-0 price** of the derivative security must be exactly the **price of the replicating strategy at time** 0.
- The multiperiod binomial model is **complete**, i.e., every derivative security can be replicated by trading in the two underlying assets.
- In the multiperiod binomial model the **time-0 price** of a derivative security is **uniquely determined** by the time-0 price of the replicating portfolio.
- **Risk-neutral pricing formula**: The time-0 price of a derivative security can be expressed as an expectation of the discounted payoff under the risk-neutral probability measure, see formula (8.16).

- Reading: This chapter followed closely (Shreve, 2004, Sections 1.2, 1.3).

Chapter 9

The multiperiod binomial asset pricing model, Part II

9.1 Computational aspects in the binomial model, Part II

Example 111 (Lookback option - computational aspects)**.** We again consider the example of the 3-period binomial model with $S_0 = 4$, $u = 2$, $d = \frac{1}{2}$, $r = \frac{1}{4}$ and the lookback option with payoff

$$V_3 = \max_{n \in \{0,1,2,3\}} S_n - S_3$$

as introduced in Example 109. As before, $\tilde{p} = \frac{1}{2}$.
Key idea: express the price of the option at time n in terms of the two-dimensional vector of random variables (S_n, M_n), where

$$M_n = \max_{k \in \{0,\ldots,n\}} S_k.$$

Example (Lookback option - computational aspects (continued)).

- At time 3: 4 possible values for the stock price S_3: 32, 8, 2, 0.5.
- At time 3: 6 possible pairs of (S_3, M_3), namely

$$(32,32),(8,16),(8,8),(2,8),(2,4),(0.5,4).$$

- We denote by $v_n(s,m)$ the value of the option at time n if the stock price at time n is $S_n = s$ and $M_n = m$.
- Then the value of the option at time 3 is just its payoff:

$$\begin{aligned} v_3(32,32) &= 0, & v_3(8,16) &= 8, & v_3(8,8) &= 0, \\ v_3(2,8) &= 6, & v_3(2,4) &= 2, & v_3(0.5,4) &= 3.5. \end{aligned}$$

- We again rewrite the recursion provided in Theorem 104 in terms of (s,m) rather than the outcome of the coin tosses:

$$v_n(s,m) = \frac{2}{5}\left(v_{n+1}(2s, \max\{m, 2s\}) + v_{n+1}(0.5s, m)\right). \tag{9.1}$$

Example (Lookback option - computational aspects (continued)).

- At time 2 there are three possible values for the stock price S_2, namely 16, 4, 1, and four possible pairs of (S_2, M_2), namely $(16,16),(4,8),(4,4),(1,4)$.
- Then using (9.1) we obtain the option price at time 2:

$$\begin{aligned} v_2(16,16) &= \frac{2}{5}\left(v_3(32, \max\{16,32\}) + v_3(8,16)\right) \\ &= \frac{2}{5}\left(v_3(32,32) + v_3(8,16)\right) = \frac{16}{5} = 3.2, \\ v_2(4,8) &= \frac{2}{5}\left(v_3(8,8) + v_3(2,8)\right) = \frac{12}{5} = 2.4, \\ v_2(4,4) &= \frac{2}{5}\left(v_3(8,8) + v_3(2,4)\right) = \frac{4}{5} = 0.8, \\ v_2(1,4) &= \frac{2}{5}\left(v_3(2,4) + v_3(0.5,4)\right) = \frac{2}{5} \cdot 5.5 = 2.2. \end{aligned}$$

Example (Lookback option - computational aspects (continued)).

- At time 1 there are two possible values for the stock price S_1, namely 8, 2 and hence 2 possible pairs of (S_1, M_1), namely $(8,8),(2,4)$, therefore using (9.1) we obtain

$$
\begin{aligned}
v_1(8,8) &= \frac{2}{5}\left(v_2(16,16)+v_2(4,8)\right) = 2.24,\\
v_1(2,4) &= \frac{2}{5}\left(v_2(4,4)+v_2(1,4)\right) = 1.2.
\end{aligned}
$$

- Hence, using (9.1) the time-0 price is given by

$$
v_0(4,4) = \frac{2}{5}\left(v_1(8,8)+v_1(2,4)\right) = 1.376.
$$

Example (Lookback option - computational aspects (continued)).

- Similarly, the replicating strategy at time n can also be characterised as a function of (s,m), where s is the stock price at time n and m the maximum stock price to date.
- The number of shares of stock that should be held in the replicating portfolio at time n if the stock price at time n is s is given by

$$
\tilde{\Delta}_n(s,m) = \frac{v_{n+1}(2s,\max\{2s,m\}) - v_{n+1}(0.5s,m)}{2s-0.5s},
$$

  which is the analogue of formula (8.14) in Theorem 104.
- The number of riskless assets that should be held in the replicating portfolio at time n if the stock price at time n is s is given by

$$
\tilde{\beta}_n(s,m) = \frac{2v_{n+1}(0.5s,m) - 0.5v_{n+1}(2s,\max\{2s,m\})}{\left(\frac{5}{4}\right)^{n+1}\frac{3}{2}}.
$$

  which is the analogue of formula (8.15) in Theorem 104.

- Reading: This section followed closely (Shreve, 2004, Section 1.3).

9.2 Option pricing in a special N-period binomial model

- We will see that a special version of the N-period binomial model approximates the famous option pricing model by Black and Scholes (BS).
- Idea: Consider a fixed time interval $[0, T]$ where $T > 0$ is the time horizon. We develop a binomial asset model that describes asset prices at $N + 1$ discrete points in time within the time-interval $[0, T]$.
- Consider the time step size $\frac{T}{N}$, where $N \in \mathbb{N}$. Then we consider a N-period binomial model which determines the prices at time $0, \frac{T}{N}, \frac{2T}{N}, \dots, \frac{NT}{N} = T$.
- We will later see that for $N \to \infty$ we obtain the Black-Scholes model.
- Hence, we study **convergence from a model in discrete-time to a model in continuous time**.

- We first consider the riskless asset.
- Let $r \geq 0$ **be an interest rate in a continuous time model**, i.e., consider a riskless asset worth $B^{(c)}(t) = e^{rt}$ at time $t \in [0, T]$.
- Define an **interest rate in the N-period binomial model r_N by setting $r_N = e^{r\frac{T}{N}} - 1$.**
- Then the time-n price of the riskless asset in the N-period binomial model at the discrete points in time $n \in \{0, \frac{T}{N}, \frac{2T}{N}, \dots, \frac{NT}{N}\}$ is given by

$$B_n = (1 + r_N)^n = e^{rn\frac{T}{N}} = B^{(c)}\left(\frac{nT}{N}\right)$$

  and hence coincides with the price of the riskless asset in continuous time at the discrete points in time.

Definition 112 (Riskless asset in BS-approximating N-period binomial model)**.** The **price of the riskless asset in the BS-approximating N-period binomial model with time horizon T at time n** is given by

$$B_n = (1 + r_N)^n,$$

where

$$r_N = e^{r\frac{T}{N}} - 1 \tag{9.2}$$

for an $r \geq 0$.

- Next we consider a special definition for the risky asset. We consider the following up- and down factors:

$$\begin{aligned} u_N &= \exp\left(\sigma\sqrt{\frac{T}{N}}\right), \\ d_N &= \frac{1}{u_N} = \exp\left(-\sigma\sqrt{\frac{T}{N}}\right) \end{aligned} \tag{9.3}$$

for a $\sigma > 0$ which will be referred to as **volatility**.

- Clearly for $\sigma > 0, T > 0, N > 0$:

$$u_N > d_N > 0.$$

- Note that for large N the **no arbitrage condition**

$$d_N < 1 + r_N < u_N \tag{9.4}$$

is satisfied.

- Consider the right hand side of (9.4):

$$1 + r_N < u_N \Leftrightarrow e^{r\frac{T}{N}} < e^{\sigma\sqrt{\frac{T}{N}}} \Leftrightarrow r\frac{T}{N} < \sigma\sqrt{\frac{T}{N}} \Leftrightarrow \frac{r}{\sigma}\sqrt{T} < \sqrt{N}$$

which is satisfied for large N.

- Consider the left hand side of (9.4):

$$d_N < 1 + r_N \Leftrightarrow e^{-\sigma\sqrt{\frac{T}{N}}} < e^{r\frac{T}{N}} \Leftrightarrow -\sigma\sqrt{\frac{T}{N}} < r\frac{T}{N} \Leftrightarrow \sqrt{N} > -\frac{r}{\sigma}\sqrt{T}$$

which is always satisfied.

Definition 113 (Risky asset in the BS-approximating N-period binomial model)**.** Let $S_0 > 0$ be the price of the risky asset at time 0. The **price of the risky asset in the BS-approximating N-period binomial model at time** $n \in \{1, \dots, N\}$, denoted by S_n is a **random variable** and depends on the outcome of the first n coin tosses $(\omega_1 \dots \omega_n)$ where for all $i \in \{1, \dots, N\}$ $\omega_i \in \{\text{Head}, \text{Tail}\}$. In particular,

$$S_n(\omega_1 \dots \omega_n) = S_0 c_{\omega_1} \cdot \dots \cdot c_{\omega_n}, \tag{9.5}$$

where for all $i \in \{1, \dots, N\}$

$$c_{\omega_i} = \begin{cases} u_N = \exp\left(\sigma\sqrt{\frac{T}{N}}\right), & \text{if } \omega_i = \text{Head}, \\ d_N = \exp\left(-\sigma\sqrt{\frac{T}{N}}\right), & \text{if } \omega_i = \text{Tail}, \end{cases}$$

where $\sigma > 0$.

- We consider a European call option with maturity $T > 0$ and strike $K > 0$ in the BS-approximating N-period binomial model.
- Hence, its payoff at time T which corresponds to step N in the binomial model is given by $V_N = v(S_N) = (S_N - K)^+$.
- According to Corollary 106 its time-0 price is given by

$$V_0^{(N)} = \tilde{E}\left[\frac{v(S_N)}{B_N}\right] = \frac{1}{B_N}\sum_{k=0}^{N}\binom{N}{k}\tilde{p}_N^k(1-\tilde{p}_N)^{N-k}(S_0 u_N^k d_N^{N-k} - K)^+, \tag{9.6}$$

where $\tilde{p}_N = \frac{1+r_N-d_N}{u_N-d_N}$.

9.3 From the N-period binomial model to the Black and Scholes option pricing formula

Theorem 114. *Let $V_0^{(N)}$ be as in (9.6) with $\tilde{p}_N = \frac{1+r_N-d_N}{u_N-d_N}$, u_N, d_N as in (9.3), r_N as in (9.2). Then,*

$$C_0^{BS} = \lim_{N\to\infty} V_0^{(N)} = S_0\Phi(D_1) - Ke^{-rT}\Phi(D_1 - \sigma\sqrt{T}),$$

where

$$D_1 = \frac{\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}, \tag{9.7}$$

and Φ is the CDF of the standard normal distribution.

We provide some details of the proof of Theorem 114 in Subsection 9.3.2.

Theorem 115. *The Black-Scholes formula for the price of the European call option with maturity T and strike K at time t is given by*

$$C_t^{BS}(S_t) = S_t \Phi(D_1(t)) - K e^{-r(T-t)} \Phi(D_1(t) - \sigma\sqrt{T-t}),$$

$$D_1(t) = \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{9.8}$$

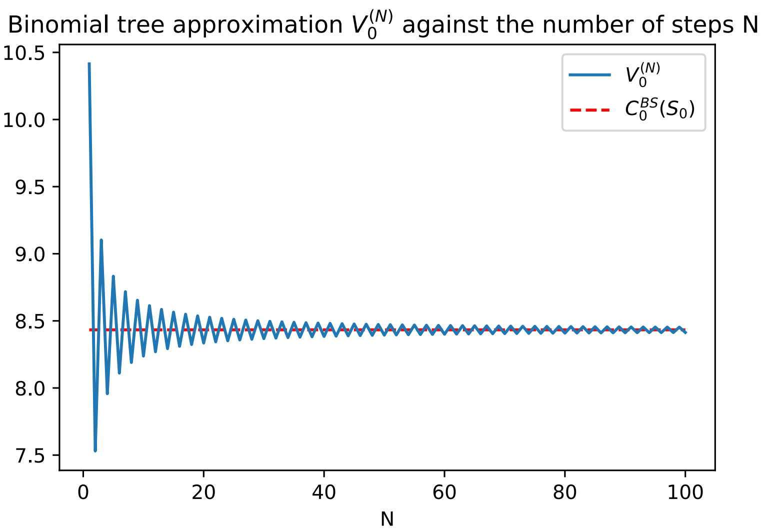
and S_t is the time-t price of the risky asset.

Meaning of model parameters:

- $r \geq 0$: interest rate,
- $\sigma > 0$: volatility of the price of the risky asset,
- S_t: stock price at time t,
- T: maturity date of option,
- K: strike price of option.

Remark 116. From (9.7) and (12.1) we see that $D_1 = D_1(0)$.

Example 117 (Convergence of the multiperiod binomial model)**.** We consider a European call with $K = 100$ maturity $T = 1$ and model parameters for $S_0 = 100$, $r = 0.01$, $\sigma = 0.2$. The following plot shows the convergence of the price given by the N-period binomial model $V_0^{(N)}$ to the Black-Scholes price $C_0^{BS}(S_0)$ as $N \to \infty$.



The proof of Theorem 114 relies on the Central Limit Theorem.

9.3.1 The Central Limit Theorem

- The Central Limit Theorem (CLT) states that the suitably standardised sample mean of independent and identically distributed random variables is approximately normally distributed if the sample size is large, more precisely:

Theorem 118. *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables having finite mean μ and finite variance σ^2. Then, for all $x \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left[\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq x\right] = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy.$$

Remark 119. Note that $E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$ and $\operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{\sigma^2}{n}$.

- Note how general the results of the CLT is.
- The distribution of the individual X_i can be anything in the world, as long as the mean and variance are finite.
- The act of averaging will cause Normality to emerge.
- We will consider an example of a discrete distribution (Poisson) and two continuous distributions (continuous uniform and exponential) in the following.

Remark 120 (Empirical CDF)**.**

- Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with CDF F. For every $z \in \mathbb{R}$, let $R_n(z)$ count how many of $Z_1, \ldots, Z_n$ are less than or equal to z, i.e.,

$$R_n(z) = \sum_{j=1}^{n} I(Z_j \leq z).$$

The indicators $I(Z_j \leq z)$ are i.i.d. with probability of success $F(z)$. Hence, $R_n(z)$ is Binomial with parameters n and $F(z)$.

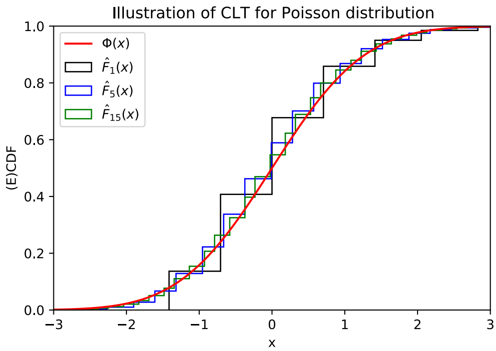
- Then, the empirical CDF of $Z_1, \ldots, Z_n$ is defined as

$$\hat{F}_n(z) = \frac{R_n(z)}{n}$$

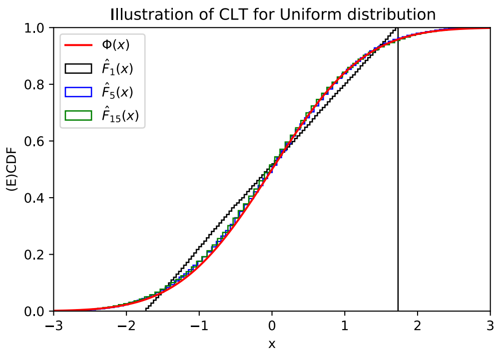
and is a function in z.

- Before $Z_1, \ldots, Z_n$ are observed, $\hat{F}_n(z)$ is a random variable for each z. After $Z_1, \ldots, Z_n$ have been observed, $\hat{F}_n(z)$ just reduces to one value for each z. In particular, $\hat{F}_n$ then is a particular CDF which can be used to estimate the CDF F of Z if F is unknown.
- In the following we will study the empirical CDF of the random variable $\frac{\frac{1}{n}\sum_{i=1}^n X_i - E\left[\frac{1}{n}\sum_{i=1}^n X_i\right]}{\sqrt{\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right]}}$ for different choices of n and see how this approaches the CDF of the standard normal distribution as n goes to infinity.

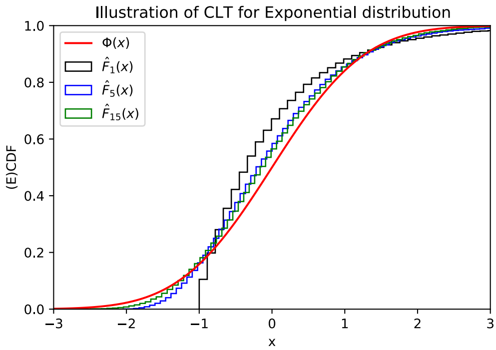
Example 121 (Applying CLT to Poisson distribution with mean 2)**.**



Example 122 (Applying CLT to Uniform distribution)**.**



Example 123 (Applying CLT to Exponential distribution with rate 2).



- We have already introduced the Monte Carlo estimator $\frac{1}{n}\sum_{i=1}^{n} X_i$ of $E[X] = \mu$ where $X_1, \dots, X_n$ are i.i.d. with the same distribution as X.
- We have also already seen using properties of expectation and variance that $E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E[X] = \mu$ and $\text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{\sigma^2}{n}$ where $\sigma^2 = \text{Var}[X]$.
- From the CLT we see that for large n it holds that $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- Hence the CLT gives the additional information that $\frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately *Normal* with the given mean and variance.

9.3.2 Appendix

- In the following we give some intuition on the proof of Theorem 114.
- To simplify the presentation we only consider the case $t = 0$.

- We set $a_N = \min\{k \in \mathbb{N}_0 : S_0 u_N^k d_N^{N-k} - K > 0\}$, then

$$\begin{aligned}
V_0^{(N)} =& \frac{1}{B_N} \sum_{k=0}^{N} \binom{N}{k} \tilde{p}_N^k (1-\tilde{p}_N)^{N-k} (S_0 u_N^k d_N^{N-k} - K)^+ \\
&= \frac{1}{B_N} \sum_{k=a_N}^{N} \binom{N}{k} \tilde{p}_N^k (1-\tilde{p}_N)^{N-k} (S_0 u_N^k d_N^{N-k} - K) \\
&= S_0 \sum_{k=a_N}^{N} \binom{N}{k} \left(\frac{\tilde{p}_N u_N}{1+r_N}\right)^k \left(\frac{(1-\tilde{p}_N) d_N}{1+r_N}\right)^{N-k} \\
&\quad - \frac{K}{(1+r_N)^N} \sum_{k=a_N}^{N} \binom{N}{k} \tilde{p}_N^k (1-\tilde{p}_N)^{N-k} \\
&= S_0 \sum_{k=a_N}^{N} \binom{N}{k} \left(\underbrace{\frac{\tilde{p}_N u_N}{1+r_N}}_{=q_N}\right)^k \left(1 - \underbrace{\frac{\tilde{p}_N u_N}{1+r_N}}_{=q_N}\right)^{N-k} \\
&\quad - \frac{K}{(1+r_N)^N} \sum_{k=a_N}^{N} \binom{N}{k} \tilde{p}_N^k (1-\tilde{p}_N)^{N-k} \\
&= S_0 \bar{B}_{N,q_N}(a_N) - \frac{K}{(1+r_N)^N} \bar{B}_{N,\tilde{p}_N}(a_N).
\end{aligned}$$

Note that $(1+r_N)^N = e^{rT}$.

- To obtain the BS-formula one needs to prove that:

$$\lim_{N\to\infty} \bar{B}_{N,q_N}(a_N) = \lim_{N\to\infty} \sum_{k=a_N}^{N} \binom{N}{k} q_N^k \left(1-q_N\right)^{N-k} = \Phi(D_1),$$

$$\lim_{N\to\infty} \bar{B}_{N,\tilde{p}_N}(a_N) = \lim_{N\to\infty} \sum_{k=a_N}^{N} \binom{N}{k} \tilde{p}_N^k \left(1-\tilde{p}_N\right)^{N-k} = \Phi(D_1 - \sigma\sqrt{T}).$$

- Recall that if $Z_N \sim \text{Bin}(N, q_N)$ then

$$P[Z_N = k] = \binom{N}{k} q_N^k (1-q_N)^{N-k},$$

for $k \in \{0, \dots, N\}$.

- Let $z \in \{0, 1, \dots, N\}$, then

$$P[z \le Z_N \le N] = \sum_{k=z}^{N} \binom{N}{k} q_N^k (1-q_N)^{N-k} = \bar{B}_{N,q_N}(z).$$

- Hence $\bar{B}_{N,q_N}(a_N) = P[a_N \le Z_N \le N]$ where $Z_N \sim \text{Bin}(N, q_N)$.
- Similarly, $\bar{B}_{N,\tilde{p}_N}(a_N) = P\left[a_N \le \hat{Z}_N \le N\right]$ where $\hat{Z}_N \sim \text{Bin}(N, \tilde{p}_N)$

We have already seen that $\tilde{p} \in (0,1)$. The next lemma shows that also indeed $q_N \in (0,1)$.

Lemma 124.

$$q_N = \frac{\tilde{p}_N u_N}{1+r_N} \in (0,1).$$

Proof. It is obvious that $q_N > 0$. To see that $q_N < 1$ consider

$$\begin{aligned}
&\frac{\tilde{p}_N u_N}{1+r_N} < 1 \\
\Leftrightarrow\ &\frac{1+r_N-d_N}{u_N-d_N}\frac{u_N}{1+r_N} < 1 \\
\overset{u_N \geq d_N}{\Leftrightarrow}\ &(1+r_N-d_N)u_N < (u_N-d_N)(1+r_N) \\
\Leftrightarrow\ &u_N + r_N u_N - d_N u_N < u_N + r_N u_N - d_N - d_N r_N \\
&\Leftrightarrow 0 < d_N(u_N - 1 - r_N) = d_N(u_N - (1+r_N))
\end{aligned}$$

which holds due to (9.4). $\square$

- Note that a $Z_N \sim \text{Bin}(N, q_N)$ can be written as a sum of N i.i.d. $\text{Bern}(q_N)$ random variables I_j. Hence,

$$\tilde{Z}_N = \frac{Z_N - E[Z_N]}{\sqrt{\text{Var}(Z_N)}} = \frac{Z_N - Nq_N}{\sqrt{Nq_N(1-q_N)}} = \frac{\sum_{j=1}^N I_j - Nq_N}{\sqrt{Nq_N(1-q_N)}}$$

  where I_j i.i.d. with $\text{Bern}(q_N)$ distribution.

- We also normalise a_N and N and define:

$$\alpha_N = \frac{a_N - Nq_N}{\sqrt{Nq_N(1-q_N)}}, \quad \beta_N = \frac{N(1-q_N)}{\sqrt{Nq_N(1-q_N)}}.$$

- One can check that $\lim_{N\to\infty} \alpha_N = -D_1$ and $\lim_{N\to\infty} \beta_N = +\infty$.

- Then according to the Central Limit Theorem $\tilde{Z}_N$ follows a standard normal distribution (for large N) and hence:

$$\begin{aligned}
\lim_{N\to\infty} B_{N,q_N}(a_N) &= \lim_{N\to\infty} P\left[a_N \le Z_N \le N\right] = \lim_{N\to\infty} P\left[\alpha_N \le \tilde{Z}_N \le \beta_N\right] \\
&= \Phi(\lim_{N\to\infty} \beta_N) - \Phi(\lim_{N\to\infty} \alpha_N) \\
&= \Phi(\infty) - \Phi(-D_1) = 1 - \Phi(-D_1) = \Phi(D_1).
\end{aligned}$$

- One can use similar arguments to show that $\lim_{N\to\infty} \overline{B}_{N,\tilde{p}_N} = \Phi(D_1 - \sigma\sqrt{T})$.

9.4 Recap

- We have investigated what happens in a special N-period binomial model as N tends to ∞.
- We have seen that the corresponding option price for a European call option in the N-period binomial model converges to the Black-Scholes option pricing formula.
- The important argument used to show convergence is the Central Limit Theorem.

Chapter 10

The Black and Scholes option pricing formula as an expectation

10.1 Option prices as expectations in the Black-Scholes market

- We will now see how the Black-Scholes formula for a European call option can be written as an expectation of a suitable random variable.
- Let X be a random variable with a standard normal distribution and again let $r \geq 0$, $\sigma > 0$, $T \geq 0$.
- We define a random variable S_T by setting

$$S_T = S_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X\right). \tag{10.1}$$

- S_T represents the time-T price of a stock per share.

Theorem 125. *Let S_T be given by (10.1) and let $K \geq 0$. Then,*

$$E\left[e^{-rT}(S_T - K)^+\right] = C_0^{BS}(S_0).$$

We will prove the theorem in the class.

Remark 126. Hence, we see that the time-0 price of a European call option can be written as an expectation of the discounted payoff of the option. We now study properties of S_T in more detail.

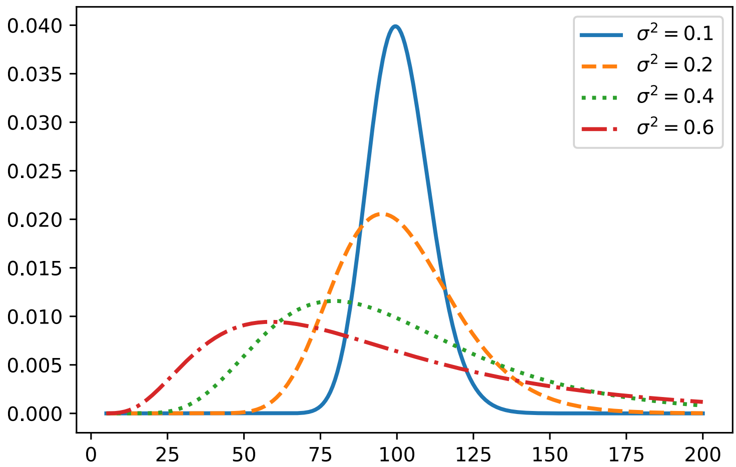
Definition 127 (Log-normal distribution)**.** A positive random variable Z is said to have a *log-normal distribution* if the (natural) logarithm of Z is normally distributed, i.e., $\log(Z) \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}, \sigma^2 > 0$. We write this as $Z \sim \text{Lognorm}(\mu, \sigma^2)$.

- Recall that if $X \sim \mathcal{N}(0,1)$ then $a + bX \sim \mathcal{N}(a, b^2)$.
- Hence, from (10.1) we obtain that

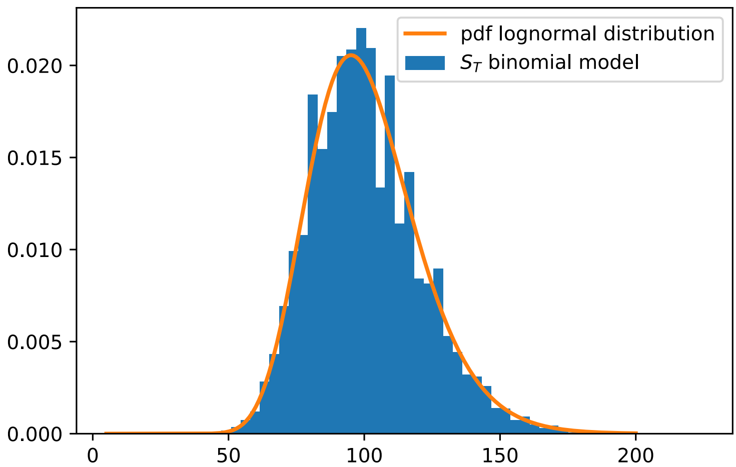
$$\log(S_T) = \log(S_0) + (r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X$$

  has a normal distribution with mean $\log(S_0) + (r - \frac{\sigma^2}{2})T$ and variance $\sigma^2 T$.
- Therefore, $S_T \sim \text{Lognorm}(\log(S_0) + (r - \frac{\sigma^2}{2})T, \sigma^2 T)$.

Plot of the pdf of S_T for different choices of σ and $S_0 = 100$, $r = 0.01$, $T = 1$.



Example 128 (Convergence of the stock price in the multiperiod binomial model)**.** We simulate 10000 stock prices at time $T = 1$ in the $N = 1000$-period binomial model with parameters $S_0 = 100$, $r = 0.01$, $\sigma = 0.2$ and compare their distribution to the lognormal distribution with parameters $\log(S_0) + (r - \frac{\sigma^2}{2})T$ and $\sigma^2 T$.



- Hence, we see that the log-normal distribution can be obtained as a limit of the multi-period binomial model as well.
- Next, we show that the expectation of the discounted log-normally distributed stock price coincides with the initial stock price - a property we have also found in the binomial model.

Proposition 129. Let S_T be given by (10.1). Then,

$$E[e^{-rT} S_T] = S_0 \tag{10.2}$$

and $E[S_T] = S_0 e^{rT}$.

Note that (10.2) corresponds to (7.11) in the one-period binomial model.

Proof of Proposition 129.

$$\begin{aligned}E[S_T] &= E\left[S_0 \exp\left((r-\frac{\sigma^2}{2})T + \sigma\sqrt{T}X\right)\right]\\ &= S_0 \exp\left((r-\frac{\sigma^2}{2})T\right) E\left[\exp\left(\sigma\sqrt{T}X\right)\right].\end{aligned}$$

We compute the expectation:

$$\begin{aligned}E\left[\exp\left(\sigma\sqrt{T}X\right)\right] &= \int_{-\infty}^{\infty} e^{\sigma\sqrt{T}x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx\\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(x^2-2\sigma\sqrt{T}x+\sigma^2T-\sigma^2T\right)}dx\\ &= e^{\frac{\sigma^2T}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(x-\sigma\sqrt{T}\right)^2}dx = e^{\frac{\sigma^2T}{2}}.\end{aligned}$$

Hence, $E[S_T] = S_0e^{rT}$ and in particular $E[e^{-rT}S_T] = S_0$. $\square$

Definition 130 (Black-Scholes Market)**.** The *Black-Scholes market* consists of a risky asset whose time-T price is given by (10.1) and a riskless asset whose time-T price is given by $B_T = e^{rT}$ for $r \geq 0$.

Remark 131. In the above definition the time-T stock price is a log-normally distributed random variable. We have specified its distribution under the *risk-neutral probability measure* and E denotes the expectation under the risk-neutral probability measure.

Theorem 132 (Option pricing in the Black-Scholes market)**.** *We consider a derivative security in the Black-Scholes market with payoff $h(S_T)$ at time T, where h is a suitable function. Then its time-0 price is given by*

$$V_0 = E\left[e^{-rT}h(S_T)\right], \tag{10.3}$$

where S_T is given by (10.1).

10.2 Option pricing by Monte Carlo in the Black-Scholes market

- From Theorem 132 we know how to obtain the time-0 price of a derivative security as an expectation.
- If we cannot compute the expectation analytically, we can approximate it using Monte Carlo methods.

Theorem 133. *The time-0 price V_0 of a derivative security with payoff $h(S_T)$ as specified in Theorem 132 can be approximated by the following Monte Carlo estimator:*

$$V_0^{MC}(n) = \frac{1}{n}\sum_{i=1}^{n} e^{-rT} h(S_i),$$

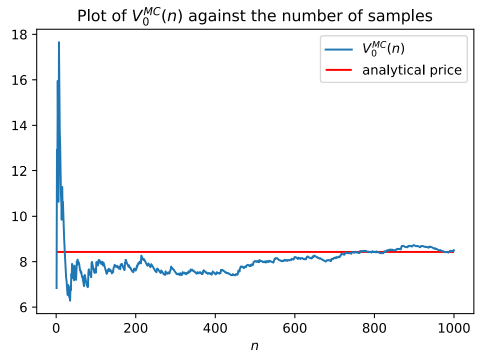
where $S_1, \ldots, S_n$ are i.i.d. from the Lognorm$(\log(S_0) + (r - \frac{\sigma^2}{2})T, \sigma^2 T)$ distribution.

Corollary 134. *The time-0 price V_0 of a derivative security with payoff $h(S_T)$ as specified in Theorem 132 can be approximated by the following Monte Carlo estimator:*

$$V_0^{MC}(n) = \frac{1}{n}\sum_{i=1}^{n} e^{-rT} h\left(S_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X_i\right)\right),$$

where $X_1, \ldots, X_n$ are i.i.d. from the $\mathcal{N}(0,1)$ distribution.

Example 135 (European call: Convergence of the MC estimator $V_0^{\mathrm{MC}}(n)$ to the BS formula C_0^{BS}).



10.3 Measuring the error of Monte Carlo estimation - confidence intervals

We assume that all random variables considered in the following have finite mean and variance.

- In the following we study how we can measure the error of a Monte Carlo estimator.
- Recall that for a random variable X the Monte Carlo estimator of $\mu = E[X]$ is given by $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ where $X_1, \dots, X_n$ are i.i.d. with the same distribution as X.
- How close is the estimator $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ to the true value μ?
- Since the Monte Carlo estimator $\bar{X}_n$ is a random variable, also $|\bar{X}_n - \mu|$ is a random variable.
- Goal: Find bound $b(n, \epsilon)$ such that

$$P\left[\left|\bar{X}_n - \mu\right| < b(n, \epsilon)\right] \approx 1 - \epsilon,$$

where $\epsilon > 0$ is usually small, e.g. 0.05.

- Note, that

$$\begin{aligned} 1 - \epsilon \approx P\left[\left|\bar{X}_n - \mu\right| < b(n, \epsilon)\right] &= P\left[-b(n, \epsilon) < \bar{X}_n - \mu < b(n, \epsilon)\right] \\ &= P\left[-b(n, \epsilon) - \bar{X}_n < -\mu < -\bar{X}_n + b(n, \epsilon)\right] \\ &= P\left[b(n, \epsilon) + \bar{X}_n > \mu > \bar{X}_n - b(n, \epsilon)\right] \\ &= P\left[\mu \in \left(\bar{X}_n - b(n, \epsilon), \bar{X}_n + b(n, \epsilon)\right)\right]. \end{aligned}$$

- Hence, we try to find an interval with random lower bound $\bar{X}_n - b(n, \epsilon)$ and random upper bound $\bar{X}_n + b(n, \epsilon)$ such that the probability that this random interval contains the true parameter μ is high, i.e., $(1 - \epsilon)$.
- Such an interval is called $(1 - \epsilon)$**-confidence interval**.

- How can we compute the probability that the confidence interval contains the true parameter?
- What should $b(n, \epsilon)$ be?
- Solution: Set $b(n, \epsilon) = a_\epsilon \frac{\sigma}{\sqrt{n}}$ for a suitable a_ϵ and apply the Central Limit Theorem.

- Let $\mu = E[X]$ and $0 < \sigma = \sqrt{\mathrm{Var}[X]} < \infty$. From the Central Limit Theorem we know that

$$\lim_{n \to \infty} P\left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right] = \Phi(x) \quad \text{for all } x \in \mathbb{R},$$

where Φ is the CDF of the standard normal distribution.

- It follows that

$$\begin{aligned}
&\lim_{n\to\infty} P\left[\mu \in \left(\bar{X}_n - a_\epsilon \frac{\sigma}{\sqrt{n}}, \bar{X}_n + a_\epsilon \frac{\sigma}{\sqrt{n}}\right)\right] \\
&\qquad = \lim_{n\to\infty} P\left[\bar{X}_n - a_\epsilon \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + a_\epsilon \frac{\sigma}{\sqrt{n}}\right] \\
&\qquad = \lim_{n\to\infty} P\left[-a_\epsilon \frac{\sigma}{\sqrt{n}} < \mu - \bar{X}_n < a_\epsilon \frac{\sigma}{\sqrt{n}}\right] \\
&\qquad = \lim_{n\to\infty} P\left[-a_\epsilon < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < a_\epsilon\right] \\
&\qquad = \lim_{n\to\infty} P\left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < a_\epsilon\right] - \lim_{n\to\infty} P\left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le -a_\epsilon\right] \\
&\qquad = \Phi(a_\epsilon) - \Phi(-a_\epsilon) \qquad (10.4) \\
&\qquad = \Phi(a_\epsilon) - (1 - \Phi(a_\epsilon)) = 2\Phi(a_\epsilon) - 1, \qquad \text{for all } a_\epsilon \ge 0, \qquad (10.5)
\end{aligned}$$

  which provides **asymptotic confidence intervals for the mean** μ.

- Hence, for given $\epsilon \in (0,1)$ we would like to determine the level of confidence by requiring that $2\Phi(a_\epsilon) - 1 = 1 - \epsilon$.

- Hence, we consider the unique point a_ϵ such that

$$\Phi(a_\epsilon) = 1 - \frac{\epsilon}{2}, \qquad (10.6)$$

  and we note that $\Phi(-a_\varepsilon) = \frac{\varepsilon}{2}$, since $\Phi(-x) = 1 - \Phi(x)$ for all x.

- (10.5) implies that

$$\lim_{n\to\infty} P\left[\mu \in \left(\bar{X}_n - a_\varepsilon \frac{\sigma}{\sqrt{n}}, \bar{X}_n + a_\varepsilon \frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \varepsilon. \qquad (10.7)$$

- We conclude that the mean $\mu = E[X]$ belongs to the $(1-\varepsilon)$-confidence interval

$$\left(\bar{X}_n - a_\varepsilon \frac{\sigma}{\sqrt{n}}, \bar{X}_n + a_\varepsilon \frac{\sigma}{\sqrt{n}}\right)$$

  as $n \to \infty$.

Remark 136.

- The Central Limit Theorem provides information about the distribution of the error between the Monte Carlo estimator and μ, in particular $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ is in the limit (i.e., as $n \to \infty$) standard normally distributed.
- Since a standard normal distribution has support $\mathbb{R}$ there is no finite bound on the error between the MC estimator $\bar{X}_n$ and μ.
- Suppose $Z \sim \mathcal{N}_1(0,1)$ and we choose $\varepsilon = 0.05$. Then $a_\varepsilon \approx 1.96$ and

$$P[|Z| < 1.96] \approx 0.95.$$

- Hence, we can say that with a probability close to 0.95, for n large enough, the error between the MC estimator $\bar{X}_n$ and the true μ is bounded by $1.96\frac{\sigma}{\sqrt{n}}$.

- In practice, the **variance σ^2 of X is typically unknown**.
- In such a case, one can for example replace σ by the **sample standard deviation**

$$s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}.$$

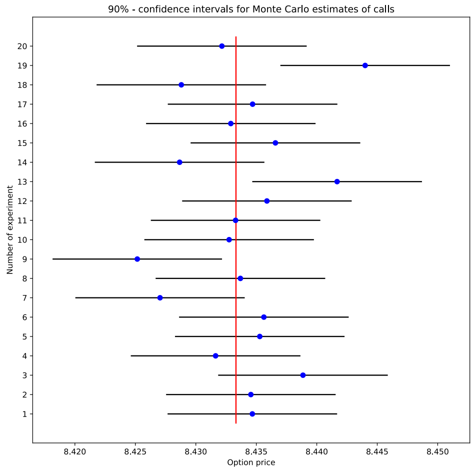
Then,

$$\left(\bar{X}_n - a_\varepsilon \frac{s_n}{\sqrt{n}}, \bar{X}_n + a_\varepsilon \frac{s_n}{\sqrt{n}}\right)$$

is an **asymptotically valid $(1-\varepsilon)$-confidence interval**.

- We see that the length of a confidence interval is given by $\bar{X}_n + a_\varepsilon \frac{\sigma}{\sqrt{n}} - (\bar{X}_n - a_\varepsilon \frac{\sigma}{\sqrt{n}}) = 2a_\epsilon \frac{\sigma}{\sqrt{n}}$.
- Ideally we would like to have short confidence intervals.
- One can see that the length of the confidence interval depends on the variance σ^2 of X.
- One way of **improving the efficiency of Monte Carlo simulation schemes** is to consider sampling that is associated with **reduced variance**, which is the topic of the next chapter.

Example 137 (European call: Different realisations of 0.9 - confidence intervals for the MC estimator of the European call price in the Black Scholes market)**.**



10.4 Recap

- In the Black-Scholes model stock prices follow a log-normal distribution under the risk-neutral probability.
- Option prices in the Black-Scholes market can be expressed as expectations of discounted payoffs using risk-neutral probabilities.
- One can approximate option prices by using a Monte Carlo estimator.

- Monte Carlo estimators are random variables. We have seen that we can quantify the error of a Monte Carlo estimator using **confidence intervals**.
- A confidence interval is a (random) interval that contains the true (unknown) parameter with high probability.
- The main idea for deriving a confidence interval for the Monte Carlo estimator is to use the Central Limit Theorem.
- The length of the standard confidence interval depends (among other things) on the variance of the random variable of which we try to estimate the expectation.

Chapter 11

Advanced Monte Carlo techniques - variance reduction

We assume that all random variables considered in the following have finite mean and variance.

- We consider the problem of estimating the expectation of a random variable X, i.e., we would like to estimate $\mu = E[X]$.
- The classical Monte Carol estimator is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ where $X_1, \ldots, X_n$ are i.i.d. random variables with the same distribution as X.
- The variance of $\bar{X}_n$ is $\mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X)}{n}$.
- We consider two variance reduction methods for estimating μ. They both try to find another random variable Z such that $E[Z] = \mu$.
- Then, they estimate μ using a Monte Carlo estimator $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$, where $Z_1, \ldots, Z_n$ are i.i.d. random variables with the same distribution as Z.
- If $\mathrm{Var}(\bar{Z}_n) \leq \mathrm{Var}(\bar{X}_n)$, then the Monte Carol estimator $\bar{Z}_n$ has a reduced variance compared to the original Monte Carlo estimator $\bar{X}_n$.
- How to find Z?

11.1 Control variates

We can introduce the context of the control variates technique by means of the following example.

Example 138.

- Suppose we want to estimate $E[f(X)]$, where f is a given function and X is a random variable with known $\mu = E[X]$.
- Let (X_i) be an an i.i.d. sequence of random variables with the same distribution as X, and define $Y_i = f(X_i)$, then the sample mean

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{11.1}$$

is an unbiased Monte Carlo estimator of $E[Y] = E[f(X)]$.

- The sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, which arises as a byproduct in the computation of $\bar{Y}_n$, is **an unbiased estimator of the *known* mean** $\mu = E[X]$.
- The **control variates technique uses this extra information** to construct a Monte Carlo estimator of $E[Y] = E[f(X)]$ that performs **better** than the estimator $\bar{Y}_n$ given by (11.1).

- Let X and Y be random variables such that $E[X]$ **is known**.
- The **control variates method** for estimating $E[Y]$ can be described as follows.
  - Given a sequence (X_i, Y_i) of i.i.d. random vectors from the joint distribution of (X, Y), we define

$$Y_i(b) = Y_i - b\left(X_i - E[X]\right), \quad \text{for } i = 1, 2, \ldots,$$

    where $b \in \mathbb{R}$ is a constant.
  - For each choice of $b \in \mathbb{R}$, $(Y_i(b))$ is a sequence of i.i.d. random variables such that

$$E\left[Y_i(b)\right] = E[Y_i] - b\left(E[X_i] - E[X]\right) = E[Y], \tag{11.2}$$

$$\mathrm{Var}\left(Y_i(b)\right) = \mathrm{Var}(Y_i - bX_i) = \mathrm{Var}(Y) - 2b\mathrm{Cov}(X, Y) + b^2\mathrm{Var}(X). \tag{11.3}$$

Definition 139. Suppose the pairs (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d. and that the expectation $E[X]$ is known.
((X, Y) denotes a generic pair of random variables with the same distribution as each (X_i, Y_i).)
The **control variate estimator with parameter** b *of* $E[Y]$ is defined by

$$\bar{Y}_n(b) := \bar{Y}_n - b\left(\bar{X}_n - E[X]\right) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - b\left(X_i - E[X]\right)\right] = \frac{1}{n}\sum_{i=1}^{n} Y_i(b). \tag{11.4}$$

Note that the observed error $\bar{X}_n - E[X]$ is used to control the estimation of $E[Y]$.

Remark 140.

- The *Strong Law of Large Numbers* implies that $\lim_{n\to\infty} \bar{Y}_n(b) = E[Y]$.
- The mean of the control variate estimator is

$$E\left[\bar{Y}_n(b)\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[Y_i(b)\right] \stackrel{(11.2)}{=} E[Y], \tag{11.5}$$

  so it is unbiased.

Remark 141.

- The variance of the control variate estimator is

$$\begin{aligned}\operatorname{Var}(\bar{Y}_n(b)) &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(Y_i(b)) \\ &= \frac{1}{n}\operatorname{Var}(Y_i(b)) \\ &\overset{(11.3)}{=} \frac{1}{n}\left[\operatorname{Var}(Y) - 2b\operatorname{Cov}(X,Y) + b^2\operatorname{Var}(X)\right]. \qquad (11.6)\end{aligned}$$

- This is a function in b and can be minimized with respect to b.
- The value b^* of the parameter b that minimizes the variance of $\bar{Y}_n(b)$ is given by

$$b^* = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}. \qquad (11.7)$$

- Substituting b^* for b in (11.6), we obtain

$$\operatorname{Var}(\bar{Y}_n(b^*)) = \frac{1}{n}\left[\operatorname{Var}(Y) - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}(X)}\right]. \qquad (11.8)$$

- This expression and the fact that

$$\operatorname{Var}(\bar{Y}_n) = \frac{1}{n}\operatorname{Var}(Y)$$

  imply that

$$\frac{\operatorname{Var}(\bar{Y}_n(b^*))}{\operatorname{Var}(\bar{Y}_n)} = 1 - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}(X)\operatorname{Var}(Y)} = 1 - \rho_{XY}^2, \qquad (11.9)$$

  where ρ_{XY} is the correlation between X and Y.

- The control variates method is useful provided that the **correlation ρ_{XY} of X and Y is big** and the extra computational effort associated with generating the samples X_i is relatively small.

- In practice, if $E[Y]$ is not known and we need simulations to estimate it, then it is **unlikely that $\mathbf{Cov(X, Y)}$, which is needed for determining b^* in (11.7), is known**.
- In such a case, we can use an **unbiased estimator $\hat{b}_n^*$ of b^***.
- In particular, we can choose

$$\hat{b}_n^* = \frac{\sum_{j=1}^n \left(X_j - \bar{X}_n\right)\left(Y_j - \bar{Y}_n\right)}{\sum_{j=1}^n \left(X_j - \bar{X}_n\right)^2}.$$

- The corresponding control variates estimator is given by $\frac{1}{n}\sum_{i=1}^n Y_i(\hat{b}_n^*)$, where

$$Y_i(\hat{b}_n^*) = Y_i - \frac{\sum_{j=1}^n \left(X_j - \bar{X}_n\right)\left(Y_j - \bar{Y}_n\right)}{\sum_{j=1}^n \left(X_j - \bar{X}_n\right)^2}\left(X_i - E[X]\right).$$

Example 142. In the following we consider an example which shows the potential benefit of using control variates.

- Suppose we would like to compute the integral

$$\mu = \int_0^1 e^u du = e - 1$$

  using Monte Carlo simulation.
- We can express μ as the expectation of a function of a Unif[0, 1] random variable by setting $\mu = \int_0^1 e^u du = E[e^U]$, where $U \sim \text{Unif}[0,1]$.
- Hence, using the previous notation we are interested in $E[Y]$, where $Y = f(X)$, $f(x) = e^x$ and $X = U \sim \text{Unif}[0,1]$.
- The control variate estimator is

$$\bar{Y}_n(b) = \frac{1}{n}\sum_{i=1}^n (e^{U_i} - b(U_i - \underbrace{E[U]}_{=1/2})),$$

  where $U_1, \ldots, U_n \sim \text{Unif}[0,1]$ i.i.d..

Using (11.7) and (11.8) we obtain

$$\begin{aligned}
\mathrm{Var}(U) &= \frac{1}{12}, \\
\mathrm{Var}(e^U) &= E[e^{2U}] - (E[e^U])^2 = \frac{1}{2}(e^2-1) - (e-1)^2, \\
\mathrm{Cov}(U, e^U) &= E[e^U U] - E[U]E[e^U] = \int_0^1 e^u u du - \frac{1}{2}(e-1) \\
&= e^u u|_{u=0}^{u=1} - \int_0^1 e^u du - \frac{1}{2}(e-1), \\
&= e - (e-1) - \frac{1}{2}(e-1) = -\frac{e}{2} + \frac{3}{2}, \\
\mathrm{Var}(\bar{Y}_n(b^*)) &= \frac{1}{n}\left[\mathrm{Var}(e^U) - \frac{\mathrm{Cov}(U,e^U)^2}{\mathrm{Var}(U)}\right] \approx \frac{1}{n}0.003940.
\end{aligned}$$

- Then $\mathrm{Var}(\bar{Y}_n(b^*)) = \frac{1}{n}\left[\mathrm{Var}(e^U) - \frac{\mathrm{Cov}(U,e^U)^2}{\mathrm{Var}(U)}\right] \approx \frac{1}{n}0.003940.$
- If we compare this to the variance of the standard Monte Carlo estimator given by

$$\mathrm{Var}(\bar{Y}_n) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n e^{U_i}) = \frac{1}{n}\mathrm{Var}(e^U) \approx \frac{1}{n}0.242036$$

we find that

$$1 - \frac{\mathrm{Var}(\bar{Y}_n(b^*))}{\mathrm{Var}(\bar{Y}_n)} \approx 0.9837.$$

Hence, the control variate estimator has reduced the variance by 98.37 % compared to the Monte Carlo estimator.

We consider now an example from option pricing.

Example 143.

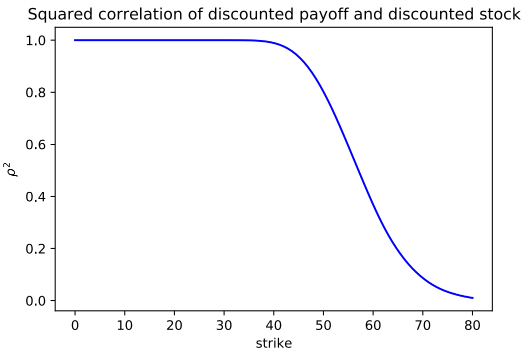
- Suppose we want to price a **European call** written on an underlying stock with:
  - time-T price $S(T)$,
  - maturity T,
  - strike price K.
  - We assume constant interest rate $r \geq 0$.
- Then, $Y = f(S_T) = e^{-rT}(S(T) - K)^+$.
- We have seen that the discounted stock price satisfies $E\left(e^{-rT}S(T)\right) = S(0)$ (where the expectation is taken with respect to the risk-neutral measure).

- Let S_i be i.i.d. random variables with the same distribution as $S(T)$ and $Y_i = e^{-rT}(S_i - K)^+$.
- Then a control variate estimator is given by

$$\overline{Y}_n(b) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - b(e^{-rT}S_i(T) - S(0))).$$

- Here b might be replaced by $\hat{b}_n^*$.
- The effectiveness of the control variate depends heavily on K, see the next Python example. For $K = 0$ we would have perfect correlation.

Example 144 (Control variate estimator of European call option for different strike prices)**.** Black Scholes market with $S_0 = 100$, $r = 0$, $\sigma = 0.3$, $T = 1$, $N = 10000$ and strike prices in $K \in \{0, 1, \dots, 80\}$. We plot $1 - \frac{\mathrm{Var}(\bar{Y}_n(\hat{b}^*))}{\mathrm{Var}(\bar{Y}_n)}$ with respect to the strike price.



11.2 Antithetic variates

- The method of **antithetic variates** attempts to **exploit negative correlation** between random variables to reduce the variance.
- **Uniform distribution:** $U \sim \mathrm{Unif}(0,1)$ if and only if $1 - U \sim \mathrm{Unif}(0,1)$.
- This observation can also be applied to more **general distributions** which can be generated from the inverse transform method. Let $U \sim \mathrm{Unif}(0,1)$:
  $F^{-1}(U)$ and $F^{-1}(1-U)$ both have cdf F and are antithetic since F^{-1} is monotone.
- **Standard normal distribution:** $X \sim \mathcal{N}_d(0, I_d)$ if and only if $-X \sim \mathcal{N}_d(0, I_d)$.
- Pairs of random variables, such as $(U, 1-U)$, $(F^{-1}(U), F^{-1}(1-U))$ and $(X, -X)$, are called **antithetic pairs**.
- The **method of antithetic variates uses antithetic pairs to produce estimators with reduced variance**.

Example 145.

- Suppose that we want to estimate $E[f(U)]$, where f is a given function and $U \sim \text{Unif}(0,1)$.
- Given a sequence (U_i) of i.i.d. random variables from Unif(0, 1),

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i) \quad \text{and} \quad \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} f(1-U_i)$$

  are unbiased Monte Carlo estimators of $E[f(U)]$.
- In particular, $\lim_{n\to\infty} \bar{X}_n = E[f(U)]$ and $\lim_{n\to\infty} \bar{Y}_n = E[f(U)]$
- The **antithetic variates estimator** simply takes the average of these two estimators and considers

$$\frac{\overline{X}_n + \overline{Y}_n}{2}$$

  as an estimator for $Ef(U)$.

Definition 146. Let (X, Y) be a random vector such that X and Y have the **same distribution**. Given a sequence (X_i, Y_i) of i.i.d. random vectors from the joint distribution of (X, Y), the **antithetic variates estimator of** $E[X] \equiv E[Y]$ is given by

$$\bar{Z}_n = \frac{\bar{X}_n + \bar{Y}_n}{2}, \tag{11.10}$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

- The antithetic variates estimator is strongly consistent because

$$\lim_{n\to\infty} \bar{Z}_n = \frac{1}{2}\lim_{n\to\infty} \bar{X}_n + \frac{1}{2}\lim_{n\to\infty} \bar{Y}_n = E[X] \equiv E[Y]. \tag{11.11}$$

  - Note, that the pairs (X_i, Y_i) are assumed to be i.i.d., but obviously for each i the X_i and Y_i have the same distribution but are not necessarily independent.

Since the random variables in the sequence $((X_i, Y_i), i = 1, 2, \ldots)$ are independent, we can see that the variance of the antithetic variates estimator $\bar{Z}_n$ given by (11.10) is

$$\begin{aligned}
\operatorname{Var}(\bar{Z}_n) &= \operatorname{Var}\left(\frac{\bar{X}_n + \bar{Y}_n}{2}\right) = \operatorname{Var}\left(\frac{1}{2n}\sum_{i=1}^{n}(X_i + Y_i)\right) \\
&= \frac{1}{4n^2}\operatorname{Var}\left(\sum_{i=1}^{n}(X_i + Y_i)\right) = \frac{1}{4n}\operatorname{Var}(X_i + Y_i) \\
&= \frac{1}{4n}\left[\operatorname{Var}(X_i) + 2\operatorname{Cov}(X_i, Y_i) + \operatorname{Var}(Y_i)\right] \\
&= \frac{1}{2n}\left[\operatorname{Var}(Y) + \operatorname{Cov}(X, Y)\right].
\end{aligned}$$

Comparing this result with

$$\operatorname{Var}(\bar{Y}_{2n}) = \frac{1}{2n}\operatorname{Var}(Y),$$

we can see that $\operatorname{Var}(\bar{Z}_n) < \operatorname{Var}(\bar{Y}_{2n})$ if $\operatorname{Cov}(X, Y) < 0$.

- Hence, the antithetic variates technique can be useful in practice when the random variables X and Y are negatively correlated and it takes roughly twice (or less than twice) the computational effort to generate a sample (X_i, Y_i) relative to generating a sample Y_i.

- In applications, the following result will be very useful:

  Lemma 147. *If $h : [0,1]^d \to \mathbb{R}$ is a monotone function of each of its arguments, then for $U_1, \ldots, U_d$ i.i.d. with distribution Unif$[0,1]$:*

  $$\mathit{Cov}(h(U_1, \ldots, U_d), h(1 - U_1, \ldots, 1 - U_d)) \leq 0.$$

Remark 148.
• We know already, that we can use the inverse transform method to generate a sample from a distribution with cumulative distribution function F by setting $X_i = F^{-1}(U_i)$ and $U_1, \ldots, U_n \sim$ Unif$[0,1]$ i.i.d..
• Note that for a cumulative distribution function F the generalized inverse F^{-1} is non-decreasing. Hence, if we consider a function $g : \mathbb{R} \to \mathbb{R}$ that is monotone, then $h := g \circ F^{-1}$ is monotone as well. This also generalizes to the case $g : \mathbb{R}^d \to \mathbb{R}$.

Example 149.

- We consider again the integral from Example 142:

$$\mu = \int_0^1 e^u du = E[e^U] = e - 1, \qquad \text{where } U \sim \text{Unif}[0,1].$$

- Let $U_1, \ldots, U_{2n} \sim \text{Unif}[0,1]$ i.i.d.. Then, the standard Monte Carlo estimator using $2n$ random numbers has variance

$$\text{Var}\left(\frac{1}{2n}\sum_{i=1}^{2n} e^{U_i}\right) = \frac{\text{Var}(e^U)}{2n} = \frac{\frac{1}{2}(e^2-1)-(e-1)^2}{2n} \approx \frac{1}{n}0.1210.$$

- Note that

$$\begin{aligned}\text{Cov}(e^U, e^{1-U}) &= E[e^U e^{1-U}] - E[e^U]E[e^{1-U}] = e - (e-1)(-1)(1-e) \\ &= e - (e-1)^2 \approx -0.23.\end{aligned}$$

- Let $U_1, \ldots, U_n \sim \text{Unif}[0,1]$ i.i.d.. Then the variance of the antithetic variates estimator is

$$\begin{aligned}\text{Var}\left(\frac{1}{2n}\left(\sum_{i=1}^{n} e^{U_i} + \sum_{i=1}^{n} e^{1-U_i}\right)\right) &= \frac{1}{2n}\left(\text{Var}(e^U) + \text{Cov}(e^U, e^{1-U})\right) \\ &\approx \frac{1}{n}0.003913.\end{aligned}$$

- Hence,

$$1 - \frac{\text{Var}\left(\frac{1}{2n}\left(\sum_{i=1}^{n} e^{U_i} + \sum_{i=1}^{n} e^{1-U_i}\right)\right)}{\text{Var}\left(\frac{1}{2n}\sum_{i=1}^{2n} e^{U_i}\right)} \approx 0.9677$$

  and therefore the antithetic variates method reduced the variance by 96.77 % .

Example 150. The antithetic variates method eliminates variance due to the antisymmetric part of an integrand.

- Suppose that we want to estimate $E[f(Z)]$, where f is a given function and $Z \sim \mathcal{N}_d(0, I_d)$.
- The standard Monte Carlo estimator of $E[f(Z)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} f(Z_i),$$

- while the antithetic variates estimator of $E[f(Z)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{f(Z_i)+f(-Z_i)}{2}.$$

- The variances of these two estimators are as follows:

$$\text{standard Monte Carlo scheme:} \quad \frac{1}{n}\mathrm{Var}\left(f(Z)\right), \tag{11.12}$$

$$\text{antithetic variates scheme:} \quad \frac{1}{n}\mathrm{Var}\left(\frac{f(Z)+f(-Z)}{2}\right). \tag{11.13}$$

- To investigate how these variances compare, we define

$$f_{\mathrm{s}}(z) = \frac{f(z)+f(-z)}{2} \quad \text{and} \quad f_{\mathrm{a}}(z) = \frac{f(z)-f(-z)}{2},$$

so that $f(z) = f_{\mathrm{s}}(z) + f_{\mathrm{a}}(z)$.

- Now, since both of Z and $-Z$ have the same distribution $\mathcal{N}_d(0, I_d)$, we can calculate

$$\begin{aligned} E\left[f_{\mathrm{s}}(Z)f_{\mathrm{a}}(Z)\right] &= \frac{1}{4}\left(E\left[f^2(Z) - f^2(-Z)\right]\right) \\ &= \frac{1}{4}\left(E\left[f^2(Z)\right] - E\left[f^2(-Z)\right]\right) \\ &= 0 \\ &= E\left[f_{\mathrm{s}}(Z)\right]\underbrace{E\left[f_{\mathrm{a}}(Z)\right]}_{=0}, \end{aligned}$$

- which implies that

$$\mathrm{Cov}\left(f_{\mathrm{s}}(Z), f_{\mathrm{a}}(Z)\right) \equiv E\left[f_{\mathrm{s}}(Z)f_{\mathrm{a}}(Z)\right] - E\left[f_{\mathrm{s}}(Z)\right]E\left[f_{\mathrm{a}}(Z)\right] = 0.$$

- It follows that

$$\begin{aligned}\operatorname{Var}(f(Z)) &= \operatorname{Var}(f_{\mathrm{s}}(Z) + f_{\mathrm{a}}(Z)) \\ &= \operatorname{Var}(f_{\mathrm{s}}(Z)) + \operatorname{Var}(f_{\mathrm{a}}(Z)) + 2\underbrace{\operatorname{Cov}(f_{\mathrm{s}}(Z), f_{\mathrm{a}}(Z))}_{=0} \\ &= \operatorname{Var}\left(\frac{f(Z)+f(-Z)}{2}\right) + \operatorname{Var}(f_{\mathrm{a}}(Z)).\end{aligned}$$

- In the view of this calculation and (11.12)–(11.13), we can see that
  - if f **is symmetric (even)**, i.e., if $f = f_{\mathrm{s}}$, then

$$\operatorname{Var}(f(Z)) = \operatorname{Var}\left(\frac{f(Z)+f(-Z)}{2}\right),$$

    and **antithetic sampling eliminates *no* variance**;
  - if f **is antisymmetric (odd)**, i.e., if $f = f_{\mathrm{a}}$, then

$$\operatorname{Var}\left(\frac{f(Z)+f(-Z)}{2}\right) = 0,$$

    and the method of **antithetic variates eliminates *all* variance**.

Note that if f is symmetric, then $f(z) = f_s(z) + f_a(z) = f_s(z)\ \forall z$, since $f_a(z) = 0\ \forall z$. Hence, in particular $f(z) = f(-z)$, so f is symmetric with respect to the y-axis ($z = 0$).
If f is antisymmetric, then $f(z) = f_s(z) + f_a(z) = f_a(z)\ \forall z$ and in particular $-f(z) = f(-z)\ \forall z$. Hence f is symmetric with respect to the origin.
Note that f_s is symmetric and f_a is antisymmetric.

Example 151.

- Let $X \sim \mathcal{N}(0,1)$ and suppose we would like to estimate $E[X] = 0$ using an antithetic variate estimator.
- We take an i.i.d. sample $X_1, \ldots, X_n$ from the $\mathcal{N}(0,1)$ distribution. Then $(X_i, -X_i)$ are antithetic pairs for all i.
- The antithetic variates estimator is then given by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{X_i + (-X_i)}{2} = 0$$

  and this estimator has zero variance (because it is a deterministic constant). Note that here $f(x) = x$ is antisymmetric ($-f(x) = -x = f(-x)\ \forall x$).

- An antithetic variates estimator for the second moment $E[X^2]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_i^2+(-X_i)^2}{2}=\frac{1}{n}\sum_{i=1}^{n}X_i^2,$$

  which is the same as the Monte Carlo estimator and hence there is no reduction in variance. Note that here $f(x)=x^2$ is symmetric ($f(x)=f(-x)\ \forall x$).

Example 152 (Antithetic variate estimator for European option)**.**

- We now consider an antithetic variates estimator for a European option with payoff $h(S_T)$ where S_T is modelled as in the Black Scholes market. We use ideas from Corollary 134.
- We want to approximate $E\left[e^{-rT}h\left(S_0\exp\left((r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}X\right)\right)\right]$ where $X\sim\mathcal{N}(0,1)$.
- Let $X_1,\dots,X_n$ be an i.i.d. sample from the $\mathcal{N}(0,1)$ distribution.
- Then $(X_i,-X_i)$ are antithetic pairs for all i.

- An antithetic variate estimator for the time-0 price of a European option with payoff $h(S_T)$ is given by:

$$\begin{aligned}
V_0^{\mathrm{AV}}(n) = &\\
&\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}e^{-rT}h\left(S_0\exp\left((r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}X_i\right)\right)\\
&+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}e^{-rT}h\left(S_0\exp\left((r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}(-X_i)\right)\right).
\end{aligned}$$

Example (Antithetic variate estimator for a European call option - example)**.** For $S_0=100$, $K=100$, $r=0$, $\sigma=0.3$, $T=1$, $N=10000$ variance is reduced by antithetic variate estimator by about 66 %.

11.3 Recap

- We have studied two techniques to reduce the variance of a Monte Carlo estimator: Control variates, antithetic variates.
- There is no general rule which tells you which variance reduction method is best.
- The best method to use will always be problem specific.

Chapter 12

Applications and outlook

12.1 Recipe for computing option prices

Suppose you want to compute the time-0 price of a derivative security, i.e., you know the payoff of the derivative security.

Remark 153 (Recipe for pricing options)**.**

- You need to decide which model you use for the underlying asset. In this course we have seen two possible models (a large number of alternative models exists!):
  - the binomial asset pricing model,
  - the Black-Scholes model.
- Compute the time-0 price by computing the risk-neutral expectation of the discounted payoff.
- If this can be done analytically you are done!
- Otherwise: Monte Carlo.
  - Simulate the underlying asset price, plug it into the formula for the payoff and compute the Monte Carlo estimator.
  - Compute a confidence interval for you MC estimator.
  - If it is too large consider using variance reduction methods.

12.2 Case study - cash or nothing call

Consider the following European digital option with payoff at maturity T given by

$$LI_{\{S_T \geq K\}},$$

where $K, L > 0$ and the indicator $I_{\{S_T \geq K\}}$ is 1 if $S_T \geq K$ and 0 otherwise. Hence, this is a European option on S_T that pays nothing if at the maturity date $S_T < K$ and otherwise it pays L. What is its time-0 price?

- We use our recipe and decide on the model for the underlying, for example the Black-Scholes model, i.e., $S_T = S_0 \exp((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X)$ where $X \sim \mathcal{N}(0,1)$ and $B_T = e^{rT}$ where $S_0 > 0$, $r, \sigma \geq 0$.
- We write down the risk-neutral expectation:

$$\begin{aligned} V_0 &= E\left[e^{-rT} LI_{\{S_T \geq K\}}\right] = Le^{-rT} E\left[I_{\{S_T \geq K\}}\right] \\ &= Le^{-rT} \int_{-\infty}^{\infty} I_{\{S_0 \exp((r-\frac{\sigma^2}{2})T + \sigma\sqrt{T}x) \geq K\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \\ &= Le^{-rT} \int_{-\infty}^{\infty} I_{\{S_0 \exp((r-\frac{\sigma^2}{2})T + \sigma\sqrt{T}x) \geq K\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \end{aligned}$$

- Note that

$$\begin{aligned} & S_0 \exp((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}x) \geq K \\ \Longleftrightarrow & (r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}x \geq \log\left(\frac{K}{S_0}\right) \\ \Longleftrightarrow & x \geq \frac{-\log\left(\frac{S_0}{K}\right) - (r - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} = -D_1 + \sigma\sqrt{T}, \end{aligned}$$

  where D_1 is as in (9.7).
- Hence, the time-0 price of the cash or nothing call is given by

$$\begin{aligned} V_0 &= Le^{-rT} \int_{-D_1+\sigma\sqrt{T}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = Le^{-rT}(1 - \Phi(-D_1 + \sigma\sqrt{T})) \\ &= Le^{-rT}\Phi(D_1 - \sigma\sqrt{T}), \end{aligned}$$

  where we used the fact that $\Phi(-x) = 1 - \Phi(x)$ for all $x \in \mathbb{R}$.

- What would be a Monte Carlo estimator for the time-0 price of the cash or nothing call?
- We generate an i.i.d. sample from the stock price at time T by generating an i.i.d. sample $X_1, \ldots, X_n$ from the standard normal distribution and setting $S_i = S_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X_i\right)$.
- Then the Monte Carlo estimator is given by

$$\frac{1}{n}\sum_{i=1}^{n} Le^{-rT} I_{\{S_i \geq K\}}$$

12.3 Black-Scholes formula revisited - implied volatilities

Next, we look into more detail into the Black-Scholes option pricing formula. Recall the following result from Lecture 9.

Theorem 154. *The Black-Scholes formula for the price of the European call option with maturity T and strike K at time t is given by*

$$C_t^{BS}(S_t, \sigma, K, T) = S_t\Phi(D_1) - Ke^{-r(T-t)}\Phi(D_1 - \sigma\sqrt{T-t}),$$

$$D_1 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{12.1}$$

and S_t is the time-t price of the risky asset.

Meaning of model parameters:

- $r \geq 0$: interest rate,
- $\sigma > 0$: volatility of the price of the risky asset,
- S_t: stock price at time t,
- T: maturity date of option,
- K: strike price of option.

- At time t, the stock price S_t is known (observable).
- The strike price K and the maturity date of the option T are known from the option contract.
- The interest rate r is also observable.
- The only *unknown* parameter in the Black-Scholes option pricing formula is the volatility σ.

- How to estimate σ?
- Two approaches:
  - Statistical approach: Consider empirical data of the the stock price and use a statistical method to estimate the unknown parameter. For example, a Maximum Likelihood Estimator can be used, but we will not consider this here.
  - Concept of implied volatility: Consider market data of option prices and find σ such that the Black-Scholes formula matches the observed market price.

The concept of implied volatility:

- Suppose we observe a market price $p_t^{\text{market}}(K,T)$ of a European Call option with strike price K and maturity T at time t.
- We can then try to find σ^{implied} such that

$$C_t^{BS}(S_t, \sigma^{\text{implied}}, K, T) = p_t^{\text{market}}(K,T).$$

|

- According to Rebonato (2005), implied volatility is "the wrong number to put in the wrong formula to get the right price".
- This is one equation for one unknown parameter σ^{implied}.
- We cannot solve this equation analytically but we can solve it numerically.

- There are several numerical methods to compute implied volatilities.
- We will use the Newton method here.
- Recall, that the Newton method is an iterative procedure to compute the root of a function f, i.e., to find x such that $f(x) = 0$:
  Start with an initial point x_0 and compute recursively

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

  For a suitable choice of x_0 and for a suitably well-behaved function f this sequence converges to the required solution as n tends to ∞.

- We write

$$f(\sigma) = C_t^{BS}(S_t, \sigma^{\text{implied}}, K, T) - p_t^{\text{market}}(K, T)$$

and want to solve $f(\sigma^{\text{implied}}) = 0$ for σ^{implied}, i.e., we want to apply Newton's method to f.

- To be able to do this we need to compute the derivative of f.

Theorem 155 (Vega). *The first derivative of the European Call option price with respect to the volatility parameter σ, which is also referred to as* Vega*, is given by*

$$\frac{\partial C_t^{BS}(S_t, \sigma^{implied}, K, T)}{\partial \sigma} = S_t \varphi(D_1)\sqrt{T-t},$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.

To prove Theorem 155, we will use the following Lemma which will be proved in the class.

Lemma 156. *Let*

$$\begin{aligned} D_1 &= \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \\ D_2 &= D_1 - \sigma\sqrt{T-t}. \end{aligned} \tag{12.2}$$

Then, it holds that

$$S_t\varphi(D_1) = Ke^{-r(T-t)}\varphi(D_2).$$

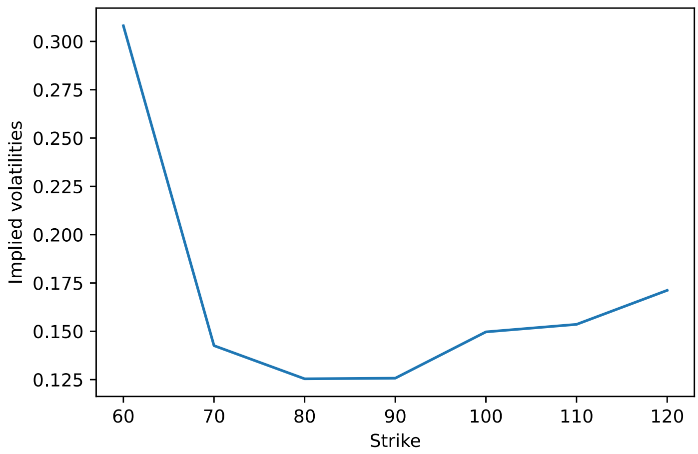
Proof of Theorem 155. First, note that using the chain rule of differentiation we obtain

$$\begin{aligned} \frac{\partial C_t^{BS}(\sigma)}{\partial \sigma} &= S_t\varphi(D_1)\frac{\partial D_1}{\partial \sigma} - Ke^{-r(T-t)}\varphi(D_2)\frac{\partial D_2}{\partial \sigma} \\ &= S_t\varphi(D_1)\frac{\partial D_1}{\partial \sigma} - Ke^{-r(T-t)}\varphi(D_2)\left(\frac{\partial D_1}{\partial \sigma} - \sqrt{T-t}\right) \\ &= \left(S_t\varphi(D_1) - Ke^{-r(T-t)}\varphi(D_2)\right)\frac{\partial D_1}{\partial \sigma} + Ke^{-r(T-t)}\varphi(D_2)\sqrt{T-t}, \end{aligned}$$

where the second equality follows from $D_2 = D_1 - \sigma\sqrt{T-t}$.
Next, we simplify the last line of the equation, using Lemma 156, i.e., we use $S_t\varphi(D_1) = Ke^{-r(T-t)}\varphi(D_2)$.
Then,

$$\frac{\partial C_t^{BS}(\sigma)}{\partial \sigma} = 0 + Ke^{-r(T-t)}\varphi(D_2)\sqrt{T-t} \qquad = S_t\varphi(D_1)\sqrt{T-t},$$

where we used Lemma 156 again in the last step. □



- Example of implied volatilities for 7 different strike prices.
- We see that the implied volatility is not constant!
- The shape of the implied volatilities as a function of strike prices is typical; one refers to this shape as *volatility smile*.
- If the Black-Scholes model was the correct model for the observed prices, the volatility parameter should not change for different strike prices.

- The analysis of implied volatilities is not just restricted to implied volatilities as a function of strike.
- Implied volatilities are also often analysed as functions in strike and time to maturity. This gives a two-dimensional surface - the *volatility surface*.

12.4 Outlook

- So far we have considered the Black-Scholes market only at a fixed point in time and seen that the stock price has a lognormal distribution at this point in time.
- The Black-Scholes model is in fact a model in continuous time, i.e., it can be used to characterise the price of a stock at time t for $t \in [0, T]$.
- To be able to define it for all $t \in [0, T]$ one needs a famous stochastic process - the so-called **Brownian motion** which is also known as **Wiener process**. Defining it formally is beyond the scope of this course.

- In practice, there is a wide range of available models (and you can try to develop new ones!) to describe the dynamics of underlying assets.
- Some for example relax the assumption of a constant volatility in the Black Scholes model, and assume that volatility is time-dependent and possible random as well. Other models, include jumps in stock price etc.
- In the binomial model and the Black-Scholes model many standard option prices of interest can be computed analytically. This is not the case for many more advanced models used in practice.
- The Monte Carlo methods discussed in this course are universal and can be applied in situations in which analytical formulae are not available.

Bibliography

Asmussen, S. & Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer.

Blitzstein, J. K. & Hwang, J. (2015). *Introduction to Probability*. Texts in Statistical Science Series, CRC Press, Boca Raton, FL.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer.

Rebonato, R. (2005). *Volatility and Correlation (2nd edition)*. Wiley.

Ross, S. N. (2006). *Simulation*. 4 ed., Elsevier Academic Press.

Seydel, R. U. (2009). *Tools for Computational Finance*. 4 ed., Springer.

Shreve, S. E. (2004). *Stochastic Calculus for Finance. I*. Springer Finance, Springer-Verlag, New York. The binomial asset pricing model.