Summary of the book[1]

A First Course in Quantitative Finance

Joshua Gloor

September 9, 2023

[1] [Maz18]

Notation

$\emptyset$	The empty set.
$A \subset S$	Given a set S, $A \subset S$ denotes that A is a subset of S. Following the convention of the book, there's no notational difference between proper and improper subsets.
$S^{\complement}$	Given a set S, $S^{\complement}$ denotes the complement of S.
$\#S$	Given a set S, $\#S$ denotes the cardinality (number of elements) of S.
$\mathbb{R}$	The real numbers.
$\mathbb{C}$	The complex numbers.
$\mathbb{N}_0$	The non-negative integers.
$\triangleq$	Definition symbol.

Part I
Technical Basics

1 A Primer on Probability

1.1 Probability and Measure

D. 1: Sample space

A set $\Omega = \{\omega_1, \omega_2, \dots\}$ with elementary states $\omega_1, \omega_2, \dots$ which may or may not realize is called a sample space. It is the set of all possible outcomes of an experiment.

D. 2: Event

An event is a set of elementary states of the world, for each of which we can tell with certainty whether or not it has realized after the random experiment is over.
Any subset E of the sample space, $E \subset \Omega$, is known as an event. In other words, an event is a set consisting of possible outcomes of the experiment.

D. 3: Complement

Let U be the set of all elements under study (the "universe") and let $A \subset U$. Then $A^{\complement}$ is called the complement of A. $A^{\complement}$ is the set of of elements that are not in A.

$$A^{\complement} = \{x \in U : x \notin A\}.$$

D. 4: σ-algebra

A family $\mathcal{F}$ of sets (events) $A, A_1, A_2, \dots$ is called a σ-algebra, if it satisfies the following conditions

(i) $\mathcal{F}$ is nonempty, i.e., $\mathcal{F} \neq \emptyset$,

(ii) if $A \in \mathcal{F}$ then $A^{\complement} \in \mathcal{F}$,

(iii) if $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

T. 1: De Morgan

The De Morgan's rule:

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^{\complement}\right)^{\complement}. \tag{1}$$

From this, the subsequent two laws, known as the De Morgan's laws, can be derived[1].

$$\bigcup_{n=1}^{\infty} A_n^{\complement} = \left(\bigcap_{n=1}^{\infty} A_n\right)^{\complement} \tag{2}$$

$$\bigcap_{n=1}^{\infty} A_n^{\complement} = \left(\bigcup_{n=1}^{\infty} A_n\right)^{\complement} \tag{3}$$

D. 5: Measurable space

Given a sample space Ω and a σ-algebra $\mathcal{F}$, the pair $(\Omega, \mathcal{F})$ is called a measurable space.

D. 6: Power set

The power set[2] of a set S, denoted as 2^S, is the set of all subsets of S, including $\emptyset$ and S itself.

[1] See appendix C.1 for derivations.

[2] The power set is often (like in this book) denoted as 2^S. The reason for this is that a power set of S has $2^{\#S}$ elements (subsets of S). Intuitively, one can either include an element of S in a subset or not, i.e., for each element of S there are two choices, leading to $2^{\#S}$ possible subsets.

D. 7: Borel-σ-algebra on $\mathbb{R}$

The Borel-σ-algebra on $\mathbb{R}$, denoted as $\mathcal{B}(\mathbb{R})$, is the σ-algebra generated by all open sets (a, b), where $a, b \in \mathbb{R}$ and $a \leq b$.

D. 8: Generated σ-algebra

The σ-algebra generated[3] by the event A is $\mathcal{F} = \{\emptyset, A, A^{\complement}, \Omega\}$, denoted as $\sigma(A)$.

D. 9: Measure

A function $\mu : \mathcal{F} \to \mathbb{R}_0^+$, with the properties

(i) $\mu(\emptyset) = 0$,

(ii) $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$, for $A_1, A_2, \ldots \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$,

is called a measure on the measurable space $(\Omega, \mathcal{F})$.

D. 10: Measure space

Given a measureable space $(\Omega, \mathcal{F})$ and a measure μ on $(\Omega, \mathcal{F})$, the triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

D. 11: Probability space

A measure space $(\Omega, \mathcal{F}, \mu)$ where the measure satisfies $\mu(\Omega) = 1$ is called a probability space. The associated measure μ is then called probability and is abbreviated as $P(A)$ for $A \in \mathcal{F}$. Therefore, the probability space triple is written as $(\Omega, \mathcal{F}, P)$.

1.2 Filtrations and the Flow of Information

D. 12: Filtration

The ascending sequence of σ-algebras $\mathcal{F}_t$, with $\mathcal{F}_0 \subset \mathcal{F}_t \subset \mathcal{F}$, is called a filtration.
If a filtration is generated by successively observing the particular outcomes of a process (like a coin toss), it is called the natural filtration of that process.

1.3 Conditional Probability and Independence

L. 1

Given a probability space $(\Omega, \mathcal{F}, P)$ and an event $A \in \mathcal{F}$ with $P(A) > 0$. Now define

$$\mathcal{F}_A = \{A \cap B \ : \ B \in \mathcal{F}\},$$

the family of all intersections of A with every event in $\mathcal{F}$. Then $\mathcal{F}_A$ is itself a σ-algebra on A and the pair $(A, \mathcal{F}_A)$ is a measurable space.

D. 13: Conditional probability

Given a probability space $(\Omega, \mathcal{F}, P)$ and two events $A, B \in \mathcal{F}$. For $P(A) > 0$, the probability measure $P(B \mid A)$ is called the conditional probability of B given A, and is defined as

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}.$$

L. 2

Given a probability space $(\Omega, \mathcal{F}, P)$ and an event $A \in \mathcal{F}$ such that $P(A) > 0$. Then the triple $(A, \mathcal{F}_A, P(\cdot \mid A))$ forms a new probability space.

[3] It can be shown that $\sigma(A)$ is the smallest σ-algebra containing A.

T. 2: Bayes' rule

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^{\complement})P(B^{\complement})}$$

D. 14: Independence

Two events A and B are said to be independent, if

$$P(A \cap B) = P(A)P(B)$$

A direct consequence of indepence is that if events A and B are independent, then the conditional probability of A given B collapses to the unconditional one:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

1.4 Random Variables and Stochastic Processes

T. 3

There exists a map from the measurable space $(\Omega, \mathcal{F})$ onto another measurable space[4] $(E, \mathcal{B})$, equipped with a distribution function F, induced by the original probability measure P.

D. 15: Random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, \mathcal{B})$ a measurable space. Then the random variable X is a function $X : \Omega \to E$.
If for every set $B \in \mathcal{B}$, there is also a $X^{-1}(B) \in \mathcal{F}$, where the inverse mapping of a random variable X is defined by

$$X^{-1}(B) = \{\omega \in \Omega \; : \; X(\omega) \in B\},$$

we call X a measurable function[5].

D. 16: Stochastic process

Given a continuous or discrete index set $0 \leq t \leq T$, a family[6] of random variables X_t is called a stochastic process.

Note that given a stochastic process X_t, there is also a family of σ-algebras $\mathcal{F}_t$ induced by X_t^{-1} in the original probability space. This is nothing else than the concept of filtrations. To reiterate on the previously stated D. 12, if the filtration $\mathcal{F}_t$ is generated by the process X_t, it is called the natural filtration of this process.

D. 17: Adapted process

If the process X_t is measurable with respect to $\mathcal{F}_t$, it is called adapted to this σ-algebra.

D. 18: Null set

Nonempty[7] sets with probability measure zero are called null sets.

D. 19: Complete probability space

A probability space is called complete, if all subsets of null sets are elements of $\mathcal{F}$.

If a property holds "almost surely", it means that a property is at most violated by events with probability zero.

[4]Ususally E is a subset of $\mathbb{R}$, whereas $\mathcal{B}$ is the corresponding Borel-σ-algebra. For countable E, $\mathcal{B}$ may be chosen as the power set of E.

[5]Note, if for every $B \in \mathcal{B}$, there is a $X^{-1}(B) \in \mathcal{F}$, means that we have a way to measure the preimage, $X^{-1}(B)$. Namely using the measure P.

[6]The notation used in this book might make it non-obvious that the stochastic process X_t is actually a sequence that stretches over the defined index set $0 \leq t \leq T$. It might be helpful to keep in mind that when we talk about the stochastic process X_t, we actually mean the sequence $(X_{t_0}, X_{t_1}, \ldots, X_{t_T})$.

[7]Example: If $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}(\mathbb{R})$, then the whole set $\mathbb{Q}$ of rational numbers has probability zero. Actually, in this case, every countable set has probability zero because it is a union of singletons and singletons have probability zero.

1.5 Moments of Random Variables

D. 20: Expectation value

The first moment of a random variable X is its expectation value $m_1 = E[X]$.
For discrete random variables, it is defined as

$$E[X] = \sum_n x_n f(x_n)$$

and for continuous random variables, provided that the density function f exists, it is defined as

$$E[X] = \int x f(x)\, \mathrm{d}x.$$

T. 4: Linearity of expectation

The expectation value is a linear functional, i.e., for $a, b \in \mathbb{R}$, it holds that

$$E[aX + bY] = aE[X] + bE[Y].$$

D. 21: Variance and standard deviation

The second moment is usually understood as a central[8] moment, which means a moment around the expectation value, called the variance, $M_2 = \mathrm{Var}[X]$. It is defined as

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

The positive root of the variance is called standard deviation, $\mathrm{StD}[X] = \sqrt{\mathrm{Var}[X]}$.

Notice that by definiton, if $E[X] = 0$, then central and raw moments are identical, i.e., $m_k = M_k$.

D. 22: Skewness

The (standardized) third moment is called the skewness of the distribution:

$$\text{skewness of } X = \frac{E\left[(X - m_1)^3\right]}{\left(E\left[(X - m_1)^2\right]\right)^{3/2}}.$$

It is a measure of the asymmetry of a distribution.

D. 23: Kurtosis

The (standardized) fourth moment is called the kurtosis of the distribution:

$$\text{kurtosis of } X = \frac{E\left[(X - m_1)^4\right]}{\left(E\left[(X - m_1)^2\right]\right)^{4/2}}.$$

It is a measure of the proportion of probability mass located in the tails of the distribution. The more massive the tails, the higher the likelihood for extreme events.

D. 24: Covariance

For two random variables X and Y, the covariance is defined as

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

Covariance is a linear measure of dependence between two random variables X and Y, because the expectation value is a linear functional[9].

[8] Note, in general $m_1 \neq M_1$. $m_1 = E[X]$ is the first raw moment (the expectation value), $M_1 = E[X - E[X]] = E[X] - E[X] = 0$ is the first central moment.

[9] See T. 4.

T. 5: Covariance and independence

If two random variables have covariance zero this does not mean they are independent. However, if two random variables are independent then their covariance is zero.
Summarized, we have:

$$\text{Cov}[X,Y] = 0 \nRightarrow X \text{ and } Y \text{ are independent}$$
$$X \text{ and } Y \text{ are independent} \Rightarrow \text{Cov}[X,Y] = 0$$

D. 25: Correlation

Correlation is defined as

$$\varrho_{XY} = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

and falls in the range of $-1 \leq \varrho_{XY} \leq 1$. The term "uncorrelated" can be used to describe zero covariance.

1.6 Characteristic Function and Fourier-Transform

D. 26: Fourier-transform

The Fourier-transform $\hat{f}$ of a function f may be defined using two arbitrary constants a and b as

$$\hat{f}(u) = \sqrt{\frac{|b|}{(2\pi)^{1-a}}} \int_{-\infty}^{\infty} e^{ibux} f(x)\,\mathrm{d}x$$
$$f(x) = \sqrt{\frac{|b|}{(2\pi)^{1+a}}} \int_{-\infty}^{\infty} e^{-ibux} \hat{f}(u)\,\mathrm{d}u.$$

For D. 26, we used the general definition from [Wei], adapted to the notation used in [Maz18]. This is to avoid the subsequently described confusion.

Thomas Mazzoni makes an attempt to provide more insight into the definition of the Fourier-transform in footnote 4 of chapter 2. However, he states a "commonly" used definition of the Fourier-transform as

$$\hat{f}(u) = \frac{1}{(2\pi)^{1-a}} \int_{-\infty}^{\infty} e^{-iux} f(x)\,\mathrm{d}x, \tag{4}$$

and its inverse transformation as

$$f(x) = \frac{1}{(2\pi)^{a}} \int_{-\infty}^{\infty} e^{iux} \hat{f}(u)\,\mathrm{d}u, \tag{5}$$

where he mentions that a is usually chosen to be 1 or $\frac{1}{2}$ and $i = \sqrt{-1}$.
While this may very well be true, we believe that it may cause confusion. The problem is that for eq. (4) and eq. (5), he clearly uses D. 26 together with $b = -1$. But in probability theory and for the computation of the charactersitic function, D. 27 together with $(a, \boldsymbol{b}) = (1, \mathbf{1})$ is used.

This is consistend with the discussion about the different conventions in [Wei]. If not otherwise stated, we will use $(a, b) = (1, 1)$ when we talk about the Fourier-transform in this summary.

D. 27: Characteristic function

The characteristic function of any real-valued random variable X fully describes X's probability distribution. The characteristic function is defined as

$$\varphi(u) = E\left[e^{iuX}\right],$$

where i is the imaginary unit.
Using the definition of the expectation value (D. 20), we see that

$$\varphi(u) = E\left[e^{iuX}\right] = \int_{-\infty}^{\infty} e^{iux} f(x)\,\mathrm{d}x.$$

This is exactly the Fourier-transform[10] of the probability density (or mass) function f.
The inverse transform is given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \varphi(u) \, \mathrm{d}u.$$

T. 6: Convolution theorem for probability distributions

Let $X_1, \ldots, X_N$ be N **independent**, not necessarily identically distributed random variables with characteristic functions $\varphi_n(u)$ for $n = 1, \ldots, N$, then

$$X = \sum_{n=1}^{N} X_n \quad \Longleftrightarrow \quad \varphi(u) = \prod_{n=1}^{N} \varphi_n(u),$$

and the probability density function of the sum X is obtained by inverse transforming its characteristic function.

[10]Notice that because of the way we defined the Fourier-transform (D. 26 with $(a, b) = (1, 1)$), it now makes sense to call this the Fourier-transform.

2 Complex Analysis

2.1 Introduction to Complex Numbers

D. 28: Complex number

Every complex number $z \in \mathbb{C}$ is a pair (x, y) of real numbers. A complex number is represented as

$$z = x + iy.$$

where i is the complex unit satisfying $i^2 = -1$.

D. 29: Components and complex conjugate

Let $z = x + iy$ be a complex number, we define:

$$\begin{aligned} \text{Re}[z] &\triangleq x && \text{The real part of } z. \\ \text{Im}[z] &\triangleq y && \text{The imaginary (complex) part of } z. \\ z^* &\triangleq x - iy && \text{The complex conjugate of } z. \end{aligned}$$

Property 1: Basic operations

The following are the basic operations using two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$.

(i) Addition and subtraction:

$$z_1 \pm z_2 = x_1 \pm x_2 + i(y_1 \pm y_2).$$

(ii) Multiplication:

$$z_1 \cdot z_2 = (x_1 + iy_1) \cdot (x_2 + iy_2) = x_1x_2 - y_1y_2 + i(x_1y_2 + x_2y_1).$$

(iii) Division:

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{x_1 + iy_1}{x_2 + iy_2} \cdot \frac{x_2 - iy_2}{x_2 - iy_2} = \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\frac{x_2y_1 - x_1y_2}{x_2^2 + y_2^2}.$$

D. 30: Absolute value

Let $z = x + iy$ be a complex number, we call

$$|z| \triangleq \sqrt{x^2 + y^2} = \sqrt{zz^*}$$

the absolute value or modulus of z.

Every complex number can also be represented in another way using polar coordinates. For every $z \in \mathbb{C}$, there exists $r \geq 0$ and $\theta \in (-\pi, \pi]$ such that

$$z = r(\cos\theta + i\sin\theta).$$

Subsequently, we show how one can transform from cartesian coordinates to polar coordinates and vice versa.

Cartesian $\rightarrow$ polar

The Cartesian coordinates x and y can be converted to polar coordinates $r \geq 0$ and $\theta \in (-\pi, \pi]$ using:

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \text{atan2}(y, x), \end{aligned}$$

where

$$\text{atan2}(y, x) = \begin{cases} 2\arctan\left(\frac{y}{\sqrt{x^2+y^2}+x}\right) & \text{if } x > 0 \text{ or } y \neq 0, \\ \pi & \text{if } x < 0 \text{ and } y = 0, \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0. \end{cases}$$

If r is calculated before θ, one can use it for finding θ:

$$\theta = \begin{cases} \arccos\left(\frac{x}{r}\right) & \text{if } y \geq 0 \text{ and } r \neq 0, \\ -\arccos\left(\frac{x}{r}\right) & \text{if } y < 0, \\ \text{undefined} & \text{if } r = 0. \end{cases}$$

Polar $\rightarrow$ Cartesian
The polar coordinates r and θ can be converted to Cartesian coordinates x and y using:

$$x = r\cos\theta,$$
$$y = r\sin\theta.$$

D. 31: exp, sin, cos

Let $z \in \mathbb{C}$, the exponential, sine, and cosine of z are:

$$\exp(z) \triangleq \sum_{n=0}^{\infty} \frac{z^n}{n!}$$
$$\sin(z) \triangleq \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$
$$\cos(z) \triangleq \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

The exponential is often written as e^z instead of $\exp(z)$.

T. 7: Euler

For every $t \in \mathbb{R}$, it holds that

$$e^{it} = \cos(t) + i\sin(t).$$

We can use Euler's theorem to write a complex number in the following form:

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Property 2: Mult. and div. in polar form

Multiplication and division for complex numbers z_1, z_2 are much easier using the polar form:

(i) Multiplication:

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}.$$

(ii) Division:

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}.$$

A Distributions and stochastic processes

This section will list distributions and stochastic processes and the properties that were discussed in the book.

A.1 Binomial distribution

$X \sim B(p, N)$, where p is the success probability of a single yes-no trial and N is the total number of trials.

Example: Imagine tossing a coin N times, where each trial is independent of the previous one. Assume that heads is up with probability p and tails with $1-p$. We are now interested in the probability of getting exactly k times heads.

Property 3: Binomial distribution

(i) The probability mass function is given by

$$f(k) = P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

(ii) The distribution function is given by

$$F(k) = P(X \leq k) = \sum_{n=0}^{k} \binom{N}{n} p^n (1-p)^{N-n}.$$

(iii) The expectation value of N trials is

$$E[X] = N \cdot p.$$

(iv) The variance is

$$\mathrm{Var}[X] = N \cdot p(1-p).$$

A.2 Normal distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$, where μ is its expectation value and σ^2 is its variance.

Property 4: Normal distribution

(i) The probability density function is given by

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

(ii) The expectation value is

$$E[X] = \mu.$$

(iii) The variance is

$$\mathrm{Var}[X] = \sigma^2.$$

(iv) The central moments M_k for $k \geq 1$ are

$$M_k = E\left[(X - E[X])^k\right] = \begin{cases} 0 & \text{for odd } k, \\ (k-1)!!\sigma^k & \text{for even } k, \end{cases}$$

where $k!! = k \cdot (k-2)!!$ and $1!! = 1$.

(v) In contrast to T. 5, if $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then the following holds:

$$X \text{ and } Y \text{ are independent} \iff \mathrm{Cov}[X, Y] = 0.$$

(vi) If $Z \sim \mathcal{N}(0, 1)$ and $X = \sigma Z + \mu$, then $X \sim \mathcal{N}(\mu, \sigma^2)$.

(vii) The characteristic function is given by

$$\varphi(u) = \exp\left(iu\mu - \frac{1}{2}u^2\sigma^2\right)$$

A.2.1 Standard normal distribution

$X \sim \mathcal{N}(0,1)$.

Property 5: Standard normal distribution

(i) The probability density function is given by

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

(ii) The following holds:

$$x \cdot \phi(x) = -\phi'(x).$$

(iii) The characteristic function is given by

$$\varphi(u) = e^{-\frac{1}{2}u^2}.$$

A.3 Wiener process

D. 32: Wiener-process

The stochastic process W_t, characterized by the properties

(i) $W_0 = 0$

(ii) W_t has independent increments

(iii) $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $0 \leq s < t$

is called the Wiener-process (or Brownian motion).

Property 6: Wiener-process

(i) For any given time interval $t - s$, W is a continuous random variable with probability density function

$$f(w) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{1}{2}\frac{w^2}{t-s}}.$$

(ii) The distribution function is given by

$$F(w) = \int_{-\infty}^{w} f(x)\,\mathrm{d}x.$$

T. 8

The Wiener-process has almost surely continuous but non-differentiable trajectories (paths).

B Prerequisites

B.1 Analysis

T. 9: Completing the square

For a second degree polynomial $ax^2 + bx + c$, we have

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}.$$

T. 10: Integration by parts

Given two functions f and g, we have

$$\int_a^b f'g \,\mathrm{d}x = [fg]_a^b - \int_a^b fg' \,\mathrm{d}x.$$

T. 11: Integration by substitution

Given two functions f and g, we have

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(y)\,\mathrm{d}y,$$

where we can use $\frac{dy}{dx} = g'(x) \iff dy = g'(x)dx$ to determine the new integral.

B.2 Combinatorial Analysis

If not otherwise stated, the content in this section is taken from [Ros10].

D. 33: Binomial coefficient

We define $\binom{n}{r}$, for $r \leq n$, by

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

and say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time.[1]

L. 3

A useful combinatorial identity is

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \qquad 1 \leq r \leq n.$$

T. 12: The binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

[1] By convention, $0!$ is defined to be 1. Thus, $\binom{n}{0} = \binom{n}{n} = 1$. We also take $\binom{n}{i}$ to be equal to 0 when either $i < 0$ or $i > n$.

C Derivations and Proofs

C.1 De Morgan's Laws

The first law, eq. (2), can be derived by taking the complement on both sides of eq. (1):

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^{\complement} = \left(\left(\bigcup_{n=1}^{\infty} A_n^{\complement}\right)^{\complement}\right)^{\complement} = \bigcup_{n=1}^{\infty} A_n^{\complement}.$$

The second law ,eq. (3), can be found by replacing A_n with $B_n \triangleq A_n^{\complement}$:

$$\bigcap_{n=1}^{\infty} B_n^{\complement} \triangleq \bigcap_{n=1}^{\infty} A_n \overset{\text{eq. (1)}}{=} \left(\bigcup_{n=1}^{\infty} A_n^{\complement}\right)^{\complement} \triangleq \left(\bigcup_{n=1}^{\infty} B_n\right)^{\complement}.$$

D Solutions to Quick Calculations

As a self-learner, I really appreciate that the book has solutions to end of chapter problems.
I do understand that the quick calculations are meant for self-testing ones understanding before continuing to read and that the calculations are meant to be solvable using the material covered. However, I am also a fan of making things as easy to understand as possible and having solutions to all problems.
Sometimes, one might get stuck even thought she understands the material or one might solve it in a wrong way without noticing. That is why I will try to provide some solution for quick calculations.

D.1 A Primer on Probability

Quick calculation 2.1

Using the conditions in D. 4 and T. 1, we derive:

$$\begin{aligned} A_1, A_2 \in \mathcal{F} &\overset{\text{item (ii)}}{\Longrightarrow} A_1^{\complement}, A_2^{\complement} \in \mathcal{F} \\ &\overset{\text{item (iii)}}{\Longrightarrow} A_1^{\complement} \cup A_2^{\complement} \in \mathcal{F} \\ &\overset{\text{eq. (2)}}{\Longrightarrow} (A_1 \cap A_2)^{\complement} \in \mathcal{F} \\ &\overset{\text{item (ii)}}{\Longrightarrow} A_1 \cap A_2 \in \mathcal{F}. \end{aligned}$$

Quick calculation 2.2

We must show that $\mathcal{F} = \{\emptyset, \{2,4,6\}, \{1,3,5\}, \Omega\}$ satisfies the conditions in D. 4:

(i) Holds by definiton, $\mathcal{F}$ has four elements.

(ii) For each element in $\mathcal{F}$, we check whether its complement is also in $\mathcal{F}$:

$$\begin{array}{lll} \emptyset^{\complement} = \Omega & \text{and } \Omega \in \mathcal{F} & \checkmark \\ \{2,4,6\}^{\complement} = \{1,3,5\} & \text{and } \{1,3,5\} \in \mathcal{F} & \checkmark \\ \{1,3,5\}^{\complement} = \{2,4,6\} & \text{and } \{2,4,6\} \in \mathcal{F} & \checkmark \\ \Omega^{\complement} = \emptyset & \text{and } \emptyset \in \mathcal{F} & \checkmark \end{array}$$

(iii) This can be seen by enumerating all possible unions of the elements. One will either get $\{2,4,6\}, \{1,3,5\}$ or Ω from the union of the elements. These are all in $\mathcal{F}$.

Quick calculation 2.3

We show that given $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$ and observing events $A_1 = \{(H,H),(H,T)\}$ and $A_2 = \{(H,T)\}$, $\mathcal{F}_2$ has eight elements by calculating $\mathcal{F}_2$ as described in the book:
At $t = 0$, nothing has happened and the only thing we know is that one of the four possible states will realize, i.e., $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
At $t = 1$, event A_1 has happened. To find the corresponding sigma-algebra we can take the sigma-algebra generated by event A_1, see D. 8.

$$\sigma(A_1) = \{\emptyset, A_1, A_1^{\complement}, \Omega\},$$

therefore,

$$\mathcal{F}_1 = \sigma(A_1) = \{\emptyset, \{(H,H),(H,T)\}, \{(T,H),(T,T)\}, \Omega\}.$$

At $t = 2$, event A_2 has happened. Taking $A_2, A_2^{\complement}$ and unions with elements in $\mathcal{F}_1$ into account, we get:

$$\mathcal{F}_2 = \\ \{\emptyset, \{(H,H),(H,T)\}, \{(T,H),(T,T)\}, \Omega, \underbrace{\{(H,T)\}}_{A_2}, \underbrace{\{(T,T),(T,H),(H,H)\}}_{A_2^{\complement}}, \underbrace{\{(T,H),(T,T),(H,T)\}}_{A_1^{\complement} \cup A_2}, \underbrace{\{(H,H)\}}_{(A_1^{\complement} \cup A_2)^{\complement}}\}$$

We see that $\mathcal{F}_2$ has eight elements as desired.

Quick calculation 2.4

Using D. 13, we can derive the following:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \frac{P(B)}{P(B)} = \frac{P(A \mid B) P(B)}{P(A)}.$$

Left to show is whether $P(A)$ equals $P(A \mid B)P(B) + P(A \mid B^{\complement})P(B^{\complement})$.
We can now use the fact that P is a measure, together with the properties of a measure (D. 9), to derive:

$$\begin{aligned}
P(A) &= P(A \cap \Omega) \\
&= P(A \cap (B \cup B^{\complement})) \\
&\overset{\text{distributive law}}{=} P((A \cap B) \cup (A \cap B^{\complement})) \\
&\overset{\text{item (ii)}}{=} P(A \cap B) + P(A \cap B^{\complement}) \\
&= P(A \cap B)\frac{P(B)}{P(B)} + P(A \cap B^{\complement})\frac{P(B^{\complement})}{P(B^{\complement})} \\
&\overset{\text{D. 13}}{=} P(A \mid B)P(B) + P(A \mid B^{\complement})P(B^{\complement}).
\end{aligned}$$

Quick calculation 2.5

We have to show that $\frac{P(\{2\})}{P(B)} = \frac{1}{3}$ holds.
Using $B = \{1, 2, 3\}$ and the definition of the measure (from Example 2.2 in the book), $P(C) = \frac{\#C}{6}$, we have:

$$\frac{P(\{2\})}{P(B)} = \frac{\#\{2\}}{\#\{1, 2, 3\}} = \frac{1}{3},$$

as desired.

Quick calculation 2.6

The two events are:

- Throwing an even number: $A = \{2, 4, 6\}$.
- Throwing a number less than or equal to four: $B = \{1, 2, 3, 4\}$.

According to D. 14, to show whether A and B are independent, we must show that $P(A \cap B) = P(A)P(B)$ holds.
Using the measure from Example 2.2 in the book again, we have:

$$P(A \cap B) = P(\{2, 4\}) = \frac{1}{3}$$

and

$$P(A)P(B) = P(\{2, 4, 6\})P(\{1, 2, 3, 4\}) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3},$$

which proves the statement.

Quick calculation 2.7

The key things to understand for this derivation are:

(i) The linearity of expectation (T. 4).

(ii) For any constant $c \in \mathbb{R}$, $E[c] = c$.

To give some (informal) intuition for the second property, think of the different values a non-random (i.e., constant) value can take. There is just one value, so the "average" of all the values is just the value itself.
To make it more obvious that $E[X]$ is a constant, we will use $m_1 = E[X]$ in the derivation. We have:

$$\begin{aligned}
E[(X - E[X])^2] &= E\left[X^2 - 2XE[X] + E[X]^2\right] \\
&\overset{\text{(i)}}{=} E[X^2] - 2E\left[XE[X]\right] + E\left[E[X]^2\right] \\
&= E[X^2] - 2E[Xm_1] + E[m_1^2] \\
&\overset{\text{(i)}}{=} E[X^2] - 2m_1E[X] + E[m_1^2] \\
&\overset{\text{(ii)}}{=} E[X^2] - 2m_1E[X] + m_1^2 \\
&= E[X^2] - 2m_1^2 + m_1^2 \\
&= E[X^2] - E[X]^2.
\end{aligned}$$

Quick calculation 2.8

For this derivation, we again use the linearity of expectation (T. 4). Also, remember, $E[X]$ and $E[Y]$ are simply constants.

$$\begin{aligned}E[(X-E[X])(Y-E[Y])] &= E[XY - XE[Y] - YE[X] + E[X]E[Y]]\\ &= E[XY] - E[XE[Y]] - E[YE[X]] + E[E[X]E[Y]]\\ &= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y]\\ &= E[XY] - E[X]E[Y].\end{aligned}$$

Quick calculation 2.9

Given $Z \sim \mathcal{N}(0,1)$ and $X = \sigma Z + \mu$, we have to show that $E[X] = \mu$ and $\text{Var}[X] = \sigma^2$.
First, we have

$$E[X] = E[\sigma Z + \mu] = \sigma \underbrace{E[Z]}_{=0} + \mu = \mu,$$

where we used the linearity of expectation for the second equality.
Also,

$$\begin{aligned}\text{Var}[X] &= \text{Var}[\sigma Z + \mu]\\ &= E\left[(\sigma Z + \mu - \mu)^2\right] \qquad (6)\\ &= E\left[\sigma^2 Z^2\right]\\ &= \sigma^2 E\left[Z^2\right] \qquad (7)\\ &= \sigma^2\Big(\underbrace{\text{Var}[Z]}_{=1} + \underbrace{E[Z]}_{=0}{}^2\Big) \qquad (8)\\ &= \sigma^2,\end{aligned}$$

where we used the definition of the variance (D. 21) for eq. (6), linearity of expectation for eq. (7), and for eq. (8), we again used the definition of the variance. However, this time we solved the equation for $E[Z^2]$, i.e., $\text{Var}[Z] = E\left[Z^2\right] - E[Z]^2 \Longleftrightarrow E\left[Z^2\right] = \text{Var}[Z] + E[Z]^2$.

Bibliography

[Ros10] Sheldon M. Ross. *A First Course in Probability*. 8th ed. Pearson Prentice Hall, 2010.

[Maz18] Thomas Mazzoni. *A First Course in Quantitative Finance*. Cambridge University Press, 2018. DOI: 10.1017/9781108303606.

[Wei] Eric W. Weisstein. *Fourier Transform. From MathWorld–A Wolfram Web Resource*. Last visited on 19/08/2023. URL: http://mathworld.wolfram.com/FourierTransform.html.