

Summary of the book<sup>1</sup>

# A First Course in Quantitative Finance

Joshua Gloor

August 20, 2023

---

<sup>1</sup>[Maz18]

<b>I</b>	<b>Technical Basics</b>	<b>3</b>
<b>1</b>	<b>A Primer on Probability</b>	<b>3</b>
1.1	Probability and Measure	3
1.2	Filtrations and the Flow of Information	4
1.3	Conditional Probability and Independence	4
1.4	Random Variables and Stochastic Processes	5
1.5	Moments of Random Variables	6
1.6	Characteristic Function and Fourier-Transform	7
<b>A</b>	<b>Distributions and stochastic processes</b>	<b>9</b>
A.1	Binomial distribution	9
A.2	Normal distribution	9
A.2.1	Standard normal distribution	10
A.3	Wiener process	10
<b>B</b>	<b>Prerequisites</b>	<b>11</b>
B.1	Analysis	11
B.2	Combinatorial Analysis	11
<b>C</b>	<b>Derivations and Proofs</b>	<b>12</b>
C.1	De Morgan's Laws	12
	<b>Bibliography</b>	<b>13</b>

## 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.
$S^c$	Given a set $S$ , $S^c$ denotes the complement of $S$ .
$\#S$	Given a set $S$ , $\#S$ denotes the cardinality (number of elements) of $S$ .
$\mathbb{R}$	$\mathbb{R}$ denotes the real numbers.
$\mathbb{C}$	$\mathbb{C}$ denotes the complex numbers.
$\triangleq$	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  $A^c$  is called the complement of  $A$ .  $A^c$  is the set of elements that are not in  $A$ .

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

#### D. 4: $\sigma$ -algebra

A family  $\mathcal{F}$  of sets (events)  $A, A_1, A_2, \dots$  is called a  $\sigma$ -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  $A^c \in \mathcal{F}$ ,
- (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:

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

From this, the subsequent two laws, known as the De Morgan's laws, can be derived<sup>1</sup>.

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

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

#### D. 5: Measurable space

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

#### D. 6: Power set

The power set<sup>2</sup> of a set  $S$ , denoted as  $2^S$ , is the set of all subsets of  $S$ , including  $\emptyset$  and  $S$  itself.

<sup>1</sup>See appendix C.1 for derivations.

<sup>2</sup>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- $\sigma$ -algebra on  $\mathbb{R}$** 

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

**D. 8: Generated  $\sigma$ -algebra**

The  $\sigma$ -algebra generated<sup>3</sup> by the event  $A$  is  $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ , denoted as  $\sigma(A)$ .

**D. 9: Measure**

A function  $\mu : \mathcal{F} \rightarrow \mathbb{R}_0^+$ , 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  $A_1, A_2, \dots \in \mathcal{F}$  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 measurable space  $(\Omega, \mathcal{F})$  and a measure  $\mu$  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  $\mu$  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  $\sigma$ -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  $\sigma$ -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  $P(B | A)$  is called the conditional probability of  $B$  given  $A$ , and is defined as

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

**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  $(A, \mathcal{F}_A, P(\cdot | A))$  forms a new probability space.

<sup>3</sup>It can be shown that  $\sigma(A)$  is the smallest  $\sigma$ -algebra containing  $A$ .

## T. 2: Bayes' rule

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)} = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | B^c)P(B^c)}$$

## 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 independence is that if events  $A$  and  $B$  are independent, then the conditional probability of  $A$  given  $B$  collapses to the unconditional one:

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

## 1.4 Random Variables and Stochastic Processes

## T. 3

There exists a map from the measurable space  $(\Omega, \mathcal{F})$  onto another measurable space<sup>4</sup>  $(E, \mathcal{B})$ , 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  $X : \Omega \rightarrow E$ .

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<sup>5</sup>.

## D. 16: Stochastic process

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

Note that given a stochastic process  $X_t$ , there is also a family of  $\sigma$ -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  $\sigma$ -algebra.

## D. 18: Null set

Nonempty<sup>7</sup> 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.

<sup>4</sup>Usually  $E$  is a subset of  $\mathbb{R}$ , whereas  $\mathcal{B}$  is the corresponding Borel- $\sigma$ -algebra. For countable  $E$ ,  $\mathcal{B}$  may be chosen as the power set of  $E$ .

<sup>5</sup>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  $P$ .

<sup>6</sup>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  $(X_{t_0}, X_{t_1}, \dots, X_{t_T})$ .

<sup>7</sup>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

$$E[X] = \int x f(x) dx.$$

### 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<sup>8</sup> moment, which means a moment around the expectation value, called the variance,  $M_2 = \text{Var}[X]$ . It is defined as

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

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

Notice that by definition, 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:

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

It is a measure of the asymmetry of a distribution.

### D. 23: Kurtosis

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

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

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

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

Covariance is a linear measure of dependence between two random variables  $X$  and  $Y$ , because the expectation value is a linear functional<sup>9</sup>.

<sup>8</sup>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.

<sup>9</sup>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:

$$\begin{aligned}\text{Cov}[X, Y] = 0 &\not\Rightarrow X \text{ and } Y \text{ are independent} \\ X \text{ and } Y \text{ are independent} &\Rightarrow \text{Cov}[X, Y] = 0\end{aligned}$$

## 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

$$\begin{aligned}\hat{f}(u) &= \sqrt{\frac{|b|}{(2\pi)^{1-a}}} \int_{-\infty}^{\infty} e^{ibu} f(x) \, dx \\ f(x) &= \sqrt{\frac{|b|}{(2\pi)^{1+a}}} \int_{-\infty}^{\infty} e^{-ibu} \hat{f}(u) \, du.\end{aligned}$$

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

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

and its inverse transformation as

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

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 characteristic function, D. 27 together with  $(a, b) = (1, 1)$  is used.

This is consistent 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

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

where  $i$  is the imaginary unit.

Using the definition of the expectation value (D. 20), we see that

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



This is exactly the Fourier-transform<sup>10</sup> of the probability density (or mass) function  $f$ .  
The inverse transform is given by

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

#### T. 6: Convolution theorem for probability distributions

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

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

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

<sup>10</sup>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.

## 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 1: 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

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

### A.2 Normal distribution

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

#### Property 2: 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

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

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

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

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:

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

(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 3: 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. 28: 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 4: 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

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

**T. 7**

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

## B Prerequisites

### B.1 Analysis

#### T. 8: 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. 9: Integration by parts

Given two functions  $f$  and  $g$ , we have

$$\int_a^b f'g \, dx = [fg]_a^b - \int_a^b fg' \, dx.$$

#### T. 10: Integration by substitution

Given two functions  $f$  and  $g$ , we have

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

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. 29: 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.<sup>1</sup>

#### L. 3

A useful combinatorial identity is

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

#### T. 11: The binomial theorem

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

<sup>1</sup>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):

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

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

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

## 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>.