Stochastische Finanzmärkte

Quantitative Finance

Lecture Notes

Summer 2024

Alois Pichler

TECHNISCHE UNIVERSITÄT
CHEMNITZ
Faculty of Mathematics

Version as of April 29, 2024

Preface and Acknowledgment

The purpose of these lecture notes is to facilitate the content of the lecture and the course. From experience it is helpful and recommended to attend and follow the lectures in addition. These lecture notes do not cover the lectures completely.

These notes have been started at the University of Vienna, the work was continued at NTNU and now at Technische Universität Chemnitz. I am indebted to these institutions for providing a solid and sound basis for work.

I am particularly indebted to Prof. Georg Ch. Pflug for numerous discussions in the area and significant support over years.

Some parts of these lecture notes follow *Stochastic models in Finance* (lecture notes by Pflug, particularly Chapter 1), some books as Williams [27], stochastic integration particularly Karatzas and Shreve [10, 11] or Øksendal [16]. Albrecher et al. [1] is a valuable and useful introduction in German language. The lecture notes Schmidt [20] have been developed at/ for TU Chemnitz; valuable lecture notes include Žitković [25, 26].

Please report mistakes, errors, violations of copyright, improvements, or necessary completions. Updated version of these lecture notes:
https://www.tu-chemnitz.de/mathematik/fima/public/stochastische_Finanzm__rkte.pdf

Syllabus and content:
https://www.tu-chemnitz.de/mathematik/studium/module/2013/M18.pdf

Contents

Part I

Deterministic Finance and Elements of Stochastics

Terms and Definitions

1

1.1 GLOSSARY OF COMMON TERMS FREQUENTLY USED IN FINANCE

Bid (Geld): the *highest* price a buyer is willing to pay for a given security at a given time; also called bid price.

Ask (Brief): the *lowest* price that any investor or dealer has declared that she/ he will sell a given security or commodity for. For over-the-counter stocks, the ask is the best quoted price at which a Market Maker is willing to sell a stock. For mutual funds, the ask is the net asset value plus any sales charges.

Spread (Spanne): the difference between the current bid and the current ask (in over-the-counter trading) or offered (in exchange trading) of a given security; also called bid/ ask spread.

Basis Point A basis point is 1/100th of a percent. Often denoted as ‱ or bp, and often—colloquially—referred to as "bip" ("bips", pl.).

Liquidity describes the degree to which an asset or security can be quickly bought or sold in the market without affecting the asset's price. Liquidity refers to the extent to which a market, such as a country's stock market or a city's real estate market, allows assets to be bought and sold at stable prices. Cash is the most liquid asset, while real estate, fine art and collectibles are all relatively illiquid.

Orders different orders include

- **Market order**: a buy or sell order in which the broker is to execute the order at the best price currently available, also called *at the market*. These are often the lowest-commission trades because they involve very little work by the broker.
- **Limit order**: in commodities and securities trading, a client's instructions to a broker to buy (or sell) an item at a specific maximum (or minimum) price. If the entire order cannot be filled (executed) at the same time, the balance may be kept for later execution according to the instructions. Also called resting order. See also away from the market order, market order, not held order, and stop order.
- **Stop order**: client's order to a broker to buy or sell a commodity or security when a specified price is reached, either above (on a buy order) or below (on a sell order) the price current at the time the order is given. A stop order becomes a market order when the item is offered at or below the specified price.
- **Stop-Loss**: a stop order for which the specified price is below the current market price and the order is to sell.
- **Stop-Buy**: variation of a stop order in which a broker is instructed to buy a commodity or security when its price reaches a certain level.

OTC (over the counter, also called unlisted): a security which is not traded on an exchange, usually due to an inability to meet listing requirements. For such securities, broker/ dealers negotiate directly with one another over computer networks and by phone, and their activities are monitored by the national association of securities dealers (NASD). OTC stocks are usually very risky since they are the stocks that are not considered large or stable enough to trade on a major exchange. They also tend to trade infrequently, making the bid-ask spread larger. Also, research about these stocks is more difficult to obtain.

Complete market (vollständiger Markt) A complete market (aka Arrow-Debreu market) is a market with two conditions:

(i) Negligible transaction costs and therefore also perfect information,
(ii) there is a price for every asset in every possible state of the world.

Arbitrage it the possibility of a risk-free profit after transaction costs. For instance, an arbitrage is present when there is the opportunity to *instantaneously* buy low and sell high.

Stock split ratio (Bezugsverhältnis)

Warrant (Zertifikat) is a corporate bond, which gives the owner the right/ option to buy at a predetermined date a predetermined number of stocks at a predetermined price.

Derivative (Derivat): contracts based on another asset, contract, etc. Examples include options (based on a stock, e.g.), futures, swaps, credit derivatives (based on a portfolio of loans)

Forward (Forward) is an OTC contract to buy a fixed quantity at a fixed time (in the future) at an agreed, predetermined price.

Future (Future) is a contract to buy (or sell) a standard quantity (a stock, commodity, etc.) at a future date at a fixed price. In contrast to the forward contract the future is traded on a stock exchange.

Option (Option) is a contract which gives the holder the right to buy (sell) a stock at a predetermined price (the strike price) at a fixed date (European option) or during a prespecified period (American Option).

Greeks the Greeks are the quantities representing the sensitivity of the price of derivatives to a change in underlying parameters on which the value of an instrument or portfolio of financial instruments is dependent.

- **Delta** (also called hedge ratio): the change in price for every move in the price of the underlying security: $\Delta = \frac{\partial V}{\partial S}$.[1]
- **Gamma**: Rate at which the delta of an option changes in response to a change in the price of the underlying asset. Positive gamma indicates positive convexity of the trading position: an up or down move in the price of the underlying asset will give the position a value higher than that predicted by delta: $\Gamma = \frac{\partial^2 V}{\partial S^2} = \frac{\partial \Delta}{\partial S}$.
- **Speed** is the derivative of Γ, $\frac{\partial^3 V}{\partial S^3}$
- **Lambda** λ (or **Omega** Ω, also called elasticity or gearing; effektiver Hebel): a measure for leverage.
  The *relative* change in an option's value as a price change in the underlying security. Allows the investor to see the relationship between an option's price and the underlying's price. For example, a stock option with an omega of 2 indicates that the price of the option will increase 2% for every 1% increase in the price of the stock:

$$\lambda = \frac{\partial V}{\partial S} \cdot \frac{S}{V} = \Delta \cdot \frac{S}{V}.$$

[1] The derivative may be approximated by $\frac{\partial V}{\partial S} \approx \frac{V(S+\Delta S)-V(S-\Delta S)}{2\Delta S}$ or $\frac{\partial V}{\partial S} \approx \frac{V(S+\Delta S)-V(S)}{\Delta S}$ for some ΔS which is small and appropriate for the desired accuracy. This holds for other derivatives (Greeks) as well.

- Omega-to-Delta, $\frac{\Omega}{\Delta} = \frac{\lambda}{\Delta} = \frac{S}{V}$;
- **Theta** (also called time decay): the ratio of the change in an option's price to the decrease in its time to expiration: $\Theta = -\frac{\partial V}{\partial t}$.
- **Rho:** the Euro-change in a given option's price that results from changing in interest rates: $\rho = \frac{\partial V}{\partial r}$. $-\frac{\rho}{V}$ is also called the duration.
- **Vega:** the change in the price of an option that results from changing the volatility: $\nu = \frac{\partial V}{\partial \sigma}$.

Leverage (also gearing; Hebel): a general technique of multiplying gains and losses (cf. derivatives).

Volatility a measure for variation of price of a financial instrument over time.

Implied volatility a theoretical value designed to represent the volatility of the security underlying an option as determined by the price of the option. The factors that affect implied volatility are the exercise price, the riskless rate of return, maturity date and the price of the option. Implied volatility appears in several option pricing models, including the Black–Scholes Option Pricing Model.

Moneyness is a measure of the degree to which a derivative is likely to have positive monetary value at its expiration, in the risk-neutral measure. It can be measured in percentage probability, or in standard deviations.

Moneyness is often defined as the probability of a positive monetary value at expiration.

- **in the money** (im Geld): Situation in which an option's strike price is below the current market price of the underlying (for a call option) or above the current market price of the underlying (for a put option). Such an option has intrinsic value.
  The option has positive value if exercised.
- **at the money** (am Geld): A condition in which the strike price of an option is equal to (or nearly equal to) the market price of the underlying security.
  The option has value 0 if exercised.
- **out of the money** (aus dem Geld): A call option whose strike price is higher than the market price of the underlying security, or a put option whose strike price is lower than the market price of the underlying security.
  Option has negative value if exercised.

Par value Stated value, or face value.

- **over par** (über pari): over the par value.
- **at par** (pari): at the par value.
- **under par** (unter pari): under the par value.

Option values

- Intrinsic value: is the value of exercising the option now
- Option value: is the market (purchase) price of the option
- Time value: is the gap between option value − intrinsic value.

Cash settlement or **physical settlement** (Barausgleich): a transaction settled with a cash payment in the amount of profit or loss rather than the physical delivery of a commodity or other underlying. Examples include futures and options contracts for indices, which cannot be delivered.

ETF Exchange Traded Fund. A fund that tracks an index, but can be traded like a stock. ETFs always bundle together the securities that are in an index; they never track actively managed mutual fund portfolios (because most actively managed funds only disclose their holdings a few times a year, so the ETF would not know when to adjust its holdings most of the time). Investors can do just about anything with an ETF that they can do with a normal stock, such as short selling. Because ETFs are traded on stock exchanges, they can be bought and sold at any time during the day (unlike most mutual funds). Their price will fluctuate from moment to moment, just like any other stock's price, and an investor will need a broker in order to purchase them, which means that she/ he will have to pay a commission. On the plus side, ETFs are more tax-efficient than normal mutual funds, and since they track indexes they have very low operating and transaction costs associated with them. There are no sales loads or investment minimums required to purchase an ETF. The first ETF created was the Standard and Poor's Deposit Receipt (SPDR, pronounced "Spider") in 1993. SPDRs gave investors an easy way to track the S&P 500 without buying an index fund, and they soon become quite popular.

1.2 PARTICULAR INSURANCE TERMS

Premium insurance premiums need to cover both the expected cost of losses, plus the cost of issuing and administering the policy, adjusting losses, and supplying the capital needed to reasonably assure that the insurer will be able to pay claims.

Actuarial Reserve a liability equal to the present value of the future expected cash flows of a contingent event

Life table (**mortality table** or **actuarial table**) is a table which shows, for each age, what the probability is that a person of that age will die before his or her next birthday.

Policyholder The person holding an insurance contract.

Insured Person The person whose life is covered by the insurance contract.

Premium payer The person paying the insurance premium

Beneficiary The recipient of the benefits defined in the insurance contract

bid, orders to buy		ask, orders to sell	
price/ €	volume	price/ €	volume
105	400	109	200
103	300	111	100
101	300	113	300

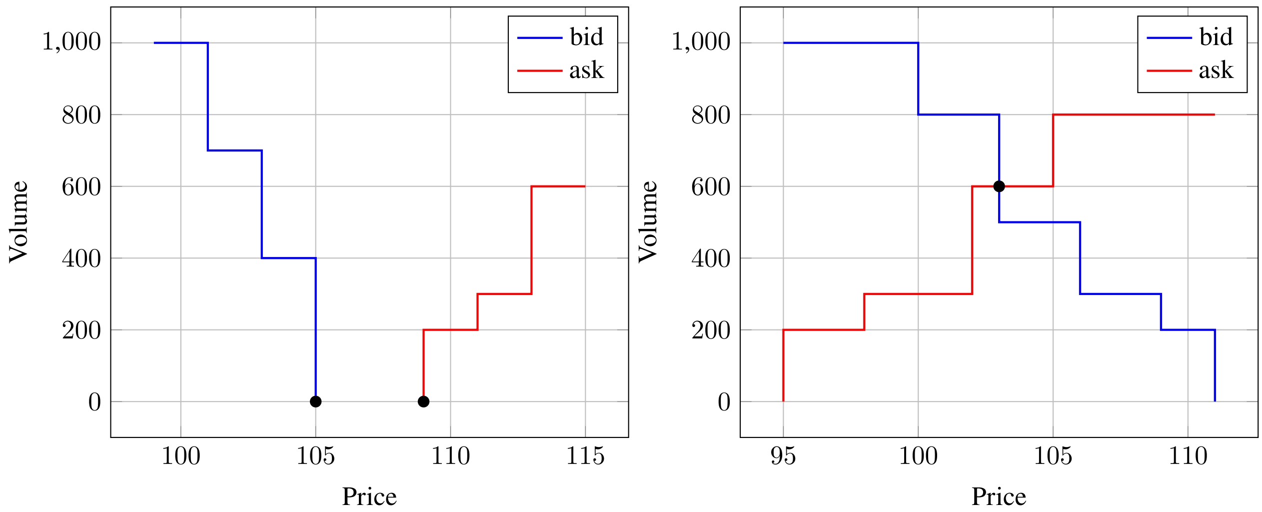
(a) No transaction takes place

bid, orders to buy		ask, orders to sell	
price/ €	volume	price/ €	volume
111	200	95	200
109	100	98	100
106	200	102	300
103	300	105	200
100	200		

(b) Transaction

Table 1.1: Order books

1.3 STOCK EXCHANGE



(a) The bid-ask spread is 4, cf. Table 1.1a

(b) 600 units are traded at a price of 103, cf. Table 1.1b

Figure 1.1: Clearing

To see how a stock exchange determines prices consider the order books depicted in the Tables 1.1. No transaction can take place in the first case, as the prices offered to buy are all below the prices offered to sell a stock. The highest price a market participant is willing to buy is 105, while the lowest price of some market participant is 109: no x-actn (transaction) can take place.

This situation is different in case of the order book in Table 1.1b. Indeed, by fixing the price at 103, 600 units can be traded (cf. Figure 1.1b).

The stock exchange fixes the price in such a way that the transacted volume is maximized, since the fees earned by the stock exchanges are dependent on traded volumes.

In recent years, automated trading by computer algorithms has become the main mechanism for placing and changing orders. High-frequency traders place and cancel thousands of computer calculated orders within seconds. The time ticks for electronic stock exchanges are now measured in microseconds (10^{-6}

seconds).

Interest Rates

2

We employ international actuarial notation and follow Gerber [7].

2.1 DEFINITIONS

Interest is a fee paid by a borrower of assets to the owner as a form of compensation for the use of the assets. It is most commonly the price paid for the use of borrowed money, or money earned by deposited funds. The interest rate[1] is the percentage relative to the initial amount, the principal. Interest rates are usually given on the basis of an entire year, and this is sometimes indicated by adding "p.a." (per annum, lat.) to the interest rate.

Interest rates are denoted by i here. Quantities related to the interest rate i are collected in Table 2.1. Of particular importance in finance (for example in life insurance to compute expected values) is the discount factor $v = \frac{1}{1+i}$. The force of interest, $\delta = \ln(1+i)$,[2] is important to evaluate fixed interest rates during a year, as $1 + i = \exp\delta$.

2.2 COMPOUND INTEREST

Compound interest is the concept of adding *accumulated* interest back to the principal, so that interest is earned on interest from that moment on. Declaring interest to be principal is called compounding (i.e., interest is compounded).

2.3 BASIC RELATIONS

In case the interest stays constant and does not vary over time, then the initial capital will grow in a time period of t years to the amount of $(1+i)^t$. The amount on a savings book thus is

$$B_t = (1+i)^t \cdot B_0$$

[1] sometimes also: yield

[2] Natural Logarithm with basis $e = 2.71828$

	symbol	relation to i	relation to δ	examples	
interest rate	i	–	$e^{\delta} - 1$	3 %	3.045 %
force of interest	δ	$\ln(1+i)$	–	2.96 %	3 %
discount rate	d	$\frac{i}{1+i}$	$1 - e^{-\delta}$	2.91 %	2.955 %
percentage rate	r	$1+i$	e^{δ}	1.03	1.031
discount factor	v	$\frac{1}{1+i}$	$e^{-\delta}$	0.971	0.97

Table 2.1: Quantities related to the interest rate

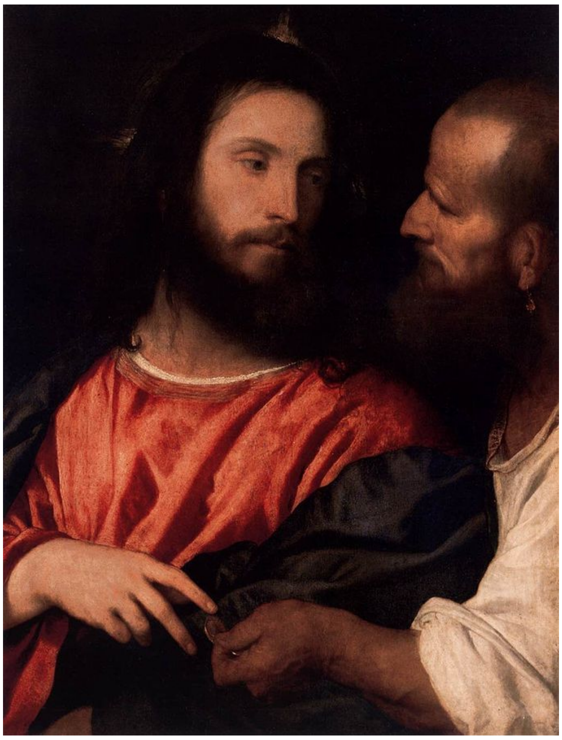


Figure 2.1: Tizian: der Zinsgroschen, 1516. Gemäldegalerie Alte Meister, Dresden

after t years and given a constant, annual interest rate of i. In different words, today's value of a payment B_t, which is due in t years and given a constant interest yield i, is

$$B_0 = B_t \cdot v^t.$$

2.4 THE PRESENT VALUE OF CERTAIN FUTURE PAYMENTS. DEFINITION AND RELATIONS

The present value (PV, aka. time value) is the value of a *future* cash flow C_t (we shall sometimes write P_t to indicate cash flows representing payments) or series of future payments on a given date, discounted to reflect the time value, or today's value of money and other factors such as investment risk. Present value calculations are widely used in business and economics to provide a means to compare cash flows at different times on a meaningful *like to like* basis.

The present value of a certain, single payment being due in t years in the future is

$$PV = \frac{C_t}{(1+i)^t} = C_t \cdot v^t = C_t \cdot e^{-\delta t}$$

assuming constant interest. In presence of a varying interest rates the respective quantity—notice the reverse situation in comparison to (2.11)—is

$$PV = C_t \cdot e^{-\int_{t_0}^{t} \delta(s)ds}.$$

The present value may be considered as a function in time itself, as

$$PV_{t_0} = PV_t \cdot e^{-\int_{t_0}^{t} \delta(s)ds}.$$

In case of a sequence of cash flows in the future the present value is given by

$$PV = \sum_k \frac{C_{t_k}}{(1+i)^{t_k}} = \sum_k C_{t_k} v^{t_k} = \sum_k C_k e^{-\delta \cdot t_k}, \tag{2.1}$$

or again more generally for a varying interest rate

$$PV_{t_0} = \sum_k C_{t_k} \cdot e^{-\int_{t_0}^{t_k} \delta(s)ds}. \tag{2.2}$$

2.5 INTEREST ON MONTH-BY-MONTH BASIS

To facilitate compounding on, say, month-by-month basis, the quantities $i^{(m)}$ and $d^{(m)}$ have proven useful. They reflect interest rates (discount rates, resp.), which are regularly added to the principal. The quantities are defined implicitly by

$$\left(1 + \frac{i^{(m)}}{m}\right)^m = 1 + i \quad \text{and} \quad \left(1 - \frac{d^{(m)}}{m}\right)^m = 1 - d. \tag{2.3}$$

Useful quantities derived from the interest and discount rate include

$$\alpha^{(m)} := \frac{i\, d}{i^{(m)}\, d^{(m)}} \quad \text{and} \quad \beta^{(m)} := \frac{i - i^{(m)}}{i^{(m)}\, d^{(m)}}. \tag{2.4}$$

$i=3\%$	$i^{(m)}$	$d^{(m)}$	$\alpha^{(m)}$	$\beta^{(m)}$
$m=1$	i	d	1	0
$m=1$	3 %	2.913 %	1	0
$m=12$	2.960 %	2.952 %	1.000 07	0.4632
$m=360$	2.956 %	2.956 %	1.000 07	0.5035
$m=\infty$	2.956 %	2.956 %	1.000 07	0.5050
$m=\infty$	$\log(1+i)$	$\log(1+i)$	$\frac{e^\delta-2+e^{-\delta}}{\delta^2}$	$\frac{e^\delta-1-\delta}{\delta^2}$

Table 2.2: Example for an interest rate of $i=3\,\%$

They are useful to simplify the computation of present values of payments which are due regularly m-times during a year, with regular intervals (cf. (2.7) in Lecture 2.6.2 below).

The quantities $i^{(m)}$, $d^{(m)}$, $\alpha^{(m)}$ and $\beta^{(m)}$ satisfy the following relations:

(i) $i^{(m)}=m\left(v^{-1/m}-1\right)$ and $d^{(m)}=m\left(1-v^{1/m}\right)$;

(ii) $\frac{1}{d^{(m)}}-\frac{1}{i^{(m)}}=\frac{1}{m}$;

(iii) $1+\beta^{(m)}d^{(m)}=\frac{i}{i^{(m)}}$;

(iv) $d^{(m)}=\frac{d}{\alpha^{(m)}-\beta^{(m)}d}$ and $d=\frac{\alpha^{(m)}d^{(m)}}{1+\beta^{(m)}d^{(m)}}$;

(v) $d\le d^{(m)}\le\delta\le i^{(m)}\le i$ whenever the interest rate is nonnegative, $i\ge 0$ and $m\ge 1$ (employing Bernoulli's inequality is an option here);

(vi) $d^{(m)}\to\delta$ and $i^{(m)}\to\delta$ whenever $m\to\infty$ (note that $m\to\infty$ corresponds to regular, but smaller payments, which become more and more frequent during the year; one may employ de l'Hôpital's rule).

Taylor series expansions around $\delta=0$ (this choice corresponds to *no interest*, i.e., $i=0\,\%$) are

(vii) $i^{(m)}=i\left(1-\frac{\beta^{(m)}}{\alpha^{(m)}}d\right)=m\left(e^{\frac{\delta}{m}}-1\right)=\delta+\frac{\delta^2}{2m}+\frac{\delta^3}{6m^2}+\frac{\delta^4}{24m^3}+\frac{\delta^5}{120m^4}+\dots$

(viii) $d^{(m)}=\frac{d}{\alpha^{(m)}-d\cdot\beta^{(m)}}=m\left(1-e^{-\frac{\delta}{m}}\right)=\delta-\frac{\delta^2}{2m}+\frac{\delta^3}{6m^2}-\frac{\delta^4}{24m^3}+\dots$

(ix) $\alpha^{(m)}=1+\frac{m^2-1}{12m^2}\delta^2+O\left(\delta^4\right)$

(x) $\beta^{(m)}=\frac{m-1}{2m}+\frac{m^2-1}{6m^2}\delta+\frac{m^2-1}{24m^2}\delta^2+\dots$

Remark 2.1. Banks sometimes use simply $\left(1+\frac{i}{m}\right)^m\approx 1+i$ to account for regular (monthly, say) interest rates. However this is not entirely correct and as $i^{(m)}\le i$ rather an overestimation of the underlying interest rate.

2.6 THE PRESENT VALUE OF IMPORTANT INVESTMENTS

We provide the present value of some important investments. This gives us the opportunity to introduce the notation conventions, which are common in actuarial science and finance.

2.6.1 Lump-sum payment

The present value of a single payment of 1 monetary unit,[3] which is due t years from now, is

$$PV = v^t.$$

2.6.2 Annuity

An annuity consists of a sequence of n payments of 1, each due at the beginning of the corresponding year at the times $t = 0, 1, \dots n-1$. Its present value is denoted $\ddot{a}_{\overline{n}|}$. A closed form expression for $\ddot{a}_{\overline{n}|}$ is available due to the well-known sum of the geometric series,

$$\ddot{a}_{\overline{n}|} = \sum_{t=0}^{n-1} v^t = \frac{1-v^n}{1-v} = \frac{1-v^n}{d}. \tag{2.5}$$

A useful approximation is $\ddot{a}_{\overline{n}|} \simeq \frac{n}{2}(1+v^n)$.

An annuity, which does not pay 1 at the beginning of the respective year, but the fraction $\frac{1}{m}$€ at the beginning of each of all m fractions of the year (m =12 for months) has the present value

$$\ddot{a}^{(m)}_{\overline{n}|} := \sum_{t=0}^{n\cdot m-1} \frac{1}{m} \cdot v^{\frac{t}{m}} = \frac{1-v^n}{m\left(1-v^{\frac{1}{m}}\right)} = \frac{1-v^n}{d^{(m)}} \tag{2.6}$$

$$= (1-v^n)\left(\frac{1}{d}\frac{i\,d}{d^{(m)}i^{(m)}} - \frac{i-i^{(m)}}{d^{(m)}i^{(m)}}\right)$$

$$= \alpha^{(m)} \cdot \ddot{a}_{\overline{n}|} - \beta^{(m)} \cdot (1-v^n), \tag{2.7}$$

the common abbreviation is $\ddot{a}^{(m)}_{\overline{n}|}$.

Note the useful approximation

$$\ddot{a}^{(m)}_{\overline{n}|} \approx \ddot{a}_{\overline{n}|} - \frac{m-1}{2m}(1-v^n), \tag{2.8}$$

which is based on the Taylor expansion in (ix) and (x).

2.6.3 Perpetuity

A perpetuity is an annuity without definite end, or a stream of cash payments that continues forever. As above (let $n \to \infty$ in (2.7)) the present value is

$$\ddot{a} := \ddot{a}_{\overline{\infty}|} = \sum_{t=0}^{\infty} v^t = \frac{1}{1-v} = \frac{1}{d},$$

or

$$\ddot{a}^{(m)} := \ddot{a}^{(m)}_{\overline{\infty}|} = \frac{1}{d^{(m)}}$$

for the perpetuity with payments $\frac{1}{m}$€.

[3] In what follows, we will use € synonymously for *monetary unit*.

2.6.4 Loan

When repaying a loan on annual basis, a series of payments (installments) is due at the times 1, 2, ... n (note this essential difference in comparison with an annuity in Lecture 2.6.2). Based on these installments the present value of the loan is

$$a_{\overline{n}|} := \sum_{t=1}^{n} v^t = v \cdot \sum_{t=0}^{n-1} v^t = v \frac{1-v^n}{1-v} = \frac{1-v^n}{i}.$$

It should be stressed that this is the amount which can be borrowed to a client who will repay 1 every year. In particular, this is the present value just an instant *after* borrowing the specified amount to the customer.

A usual loan, being repaid on fractional (monthly, e.g.) basis, has the present value

$$a_{\overline{n}|}^{(m)} := \frac{1}{m} \sum_{t=1}^{n \cdot m} v^{\frac{t}{m}} = \frac{v^{\frac{1}{m}}}{m} \cdot \sum_{t=0}^{n \cdot m-1} v^{\frac{t}{m}} = \frac{v^{\frac{1}{m}}}{m} \cdot \frac{1-v^n}{1-v^{\frac{1}{m}}} = \frac{1-v^n}{i^{(m)}}. \tag{2.9}$$

This present value represents the amount of cash that can be borrowed to a client, provided he will repay $\frac{1}{m}$ in every month during the following n successive years.

Notice here as well that this is just the present value an instant *after* borrowing. The present value *before* borrowing is 0.

An approximation in common use is

$$a_{\overline{n}|}^{(m)} \approx \ddot{a}_{\overline{n}|} - \frac{m+1}{2m} \left(1 - v^n\right),$$

which is even an exact formula for $m = 1$ (compare with (2.8)).

2.6.5 Perpetual interest

An investment provides a return of 1 at the end of each period. To obtain the present of the investment value let $n \to \infty$ in (2.9), such that

$$a^{(m)} := a_{\overline{\infty}|}^{(m)} = \frac{1}{i^{(m)}}.$$

The formula $1/i$ is in frequent use to obtain the present value for a house or flat, which is rented out.

2.7 VARYING INTEREST RATES AND NON-REGULAR INTEREST PAYMENTS

In a typical situation the interest rate changes frequently, say, on a day-by-day basis. This has to be reflected in the formulae, and the according interest rate has to be taken into account for the respective time period. Assuming that interest is still given on an annual basis (p.a.) the formula thus rewrites

$$B_{t_n} = B_{t_0} \cdot \left(1 + i_{t_0}\right)^{t_1 - t_0} \cdot \left(1 + i_{t_1}\right)^{t_2 - t_1} \cdot \ldots \cdot \left(1 + i_{t_{n-1}}\right)^{t_n - t_{n-1}}, \tag{2.10}$$

where the interest is constantly i_k during the time interval from t_k to t_{k+1} ($k = 0, 1, 2 \ldots$).

The product (2.10), however, is not an easy to handle expression. By employing the force of interest Eq. (2.10) rewrites as

$$\begin{aligned} B_{t_n} &= B_{t_0} \cdot e^{\delta_{t_0}(t_1 - t_0)} \cdot e^{\delta_{t_1}(t_2 - t_1)} \cdot \ldots \cdot e^{\delta_{t_{n-1}}(t_n - t_{n-1})} \\ &= B_{t_0} \cdot e^{\delta_{t_0}(t_1 - t_0) + \delta_{t_1}(t_2 - t_1) + \ldots \delta_{t_{n-1}}(t_n - t_{n-1})}. \end{aligned}$$

This quantity easily is seen to be

$$B_{t_n} = B_{t_0} \cdot e^{\int_{t_0}^{t_n} \delta(t)dt}, \tag{2.11}$$

where $\delta(\cdot)$ is the simple function

$$\delta(t) := \delta_{t_i} \text{ for } t_i \le t < t_{i+1}$$

accounting for the varying force of interest. The expression (2.11) may and will be used now in a general environment with varying interest rates, i.e., for general function δ representing the varying force of interest.

2.8 BONDS, AND THE COUPON CORRECTION

A bond is a formal contract to repay borrowed money with interest at fixed intervals (cf. O'Sullivan and Sheffrin [17]). A bond thus is a debt security, in which the authorized issuer owes the holders a debt. Depending on the terms of the bond the issuer is obliged to pay interest (the *coupon*) and/ or to repay the *principal* at a later date, termed *maturity*.

The bond's price with coupon c at the market is p. The internal rate of return thus satisfies[4]

$$p = c \cdot \tilde{v} \cdot \tilde{\ddot{a}}_{\overline{n}|} + \tilde{v}^n, \tag{2.12}$$

where $\tilde{v} = \frac{1}{1+\tilde{i}}$ and $\tilde{\ddot{a}}_{\overline{n}|} = \frac{1-\tilde{v}^n}{1-\tilde{v}}$ as usual, but computed with interest rate $\tilde{i}$ (internal rate of return, cf. (2.17)).

Coupon Correction

It is well accepted to compare different investment vehicles by comparing their internal rate of return. For this purpose, however, Eq. (2.12) has to be solved. This is typically complicated and there is no general, closed form solution.

It is an alternative to consider the price of the corresponding zero-coupon bond $\tilde{v}^n$ instead of p. Solving Eq. (2.12) numerically can be avoided by accepting the useful (and very good) proxy

$$\tilde{v}^n \cong v^n + (p - 100\%) \frac{n \cdot v^n}{\ddot{a}_{\overline{n}|}} + O\,(p - 100\%)^2, \tag{2.13}$$

where the coupon takes the role of the interest, $v = \frac{1}{1+c}$ and $\ddot{a}_{\overline{n}|} = \frac{1-v^n}{1-v}$.

The formula (2.13) is even *exact* for a bond

- *at par* (i.e., $p = 1$),
- for a zero-coupon bond ($c = 0$, hence $v = 1$ and $\tilde{v}^n = p = 1 + (p-1)\,\frac{n \cdot 1}{n}$), or
- at maturity ($n = 0$).

Formula (2.13) allows to quickly compute an approximation of the internal interest rate, as[5]

$$\tilde{i} = (\tilde{v}^n)^{-\frac{1}{n}} - 1 \cong (1+i)\left(1 + (p - 100\%)\,\frac{n}{\ddot{a}_{\overline{n}|}}\right)^{-\frac{1}{n}} - 1 \cong i - \frac{p-1}{\ddot{a}_n}.$$

[4] compare this to $1 = v^n + i \cdot v \cdot \ddot{a}_{\overline{n}|}$, that is to say the present value of the bond is $p = 1$ if the coupon $c = i$ just reflects the market interest.

[5] The very short approximation $\tilde{i} \cong i - \frac{p-1}{\ddot{a}_{\overline{n}|}}$ is enlightening, though a bit worse.

market price of the bond	60%	80%	100%	120%	140%
internal rate of return	12.11%	7.98%	5%	2.69%	0.82%
zero-coupon bond price $\tilde{v}^n$	31.9%	46.4%	61.39%	76.7%	92.2%
interest rate, approximated	12.39%	8.02%	5%	2.71%	0.87%
by $v^n + (p-1)\,\frac{n\cdot v^n}{\ddot{a}_{\overline{n}|}}$	31.1%	46.3%	61.39%	76.6%	91.7%

Table 2.3: Internal rate of return and the corresponding zero-coupon bond (cf. Example 2.2)

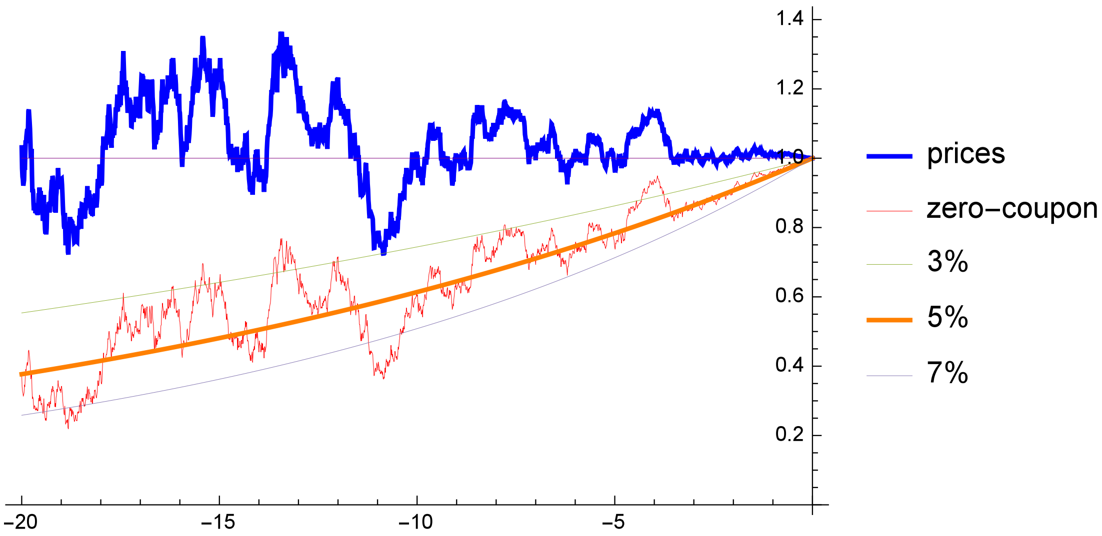


Figure 2.2: Treasury bond with a coupon of 5%: when the bond was issued 5% was a reasonable interest rate.

Remark. The next order to approximate $\tilde{v}^n$ (cf. (2.13)) is

$$\tilde{v}^n \cong v^n + (p-1)\,\frac{n\cdot v^n}{\ddot{a}_{\overline{n}|}} + (p-1)^2\left(\tfrac{n+1}{2}\ddot{a}_{\overline{n}|} - \ddot{a}^{inc}_{\overline{n}|}\right)\frac{n\cdot v^n}{\ddot{a}^3_{\overline{n}|}} + O\,(p-1)^3\,,$$

where

$$\ddot{a}^{inc}_{\overline{n}|} := \sum_{t=0}^{n-1}(t+1)\,v^t = \frac{1-(n+1)\,v^n + n\,v^{n+1}}{(1-v)^2} \tag{2.14}$$

is the linearly increasing annuity.

Example 2.2 (Quality of the coupon correction formula (2.13), exposed in selected situations)**.** Consider a bond with term $n = 10$ years and a coupon $c = 5\%$. Table 2.3 exposes the interest rate, as well as the convincing quality of the approximation according (2.13).

Example 2.3. Figure 2.2 displays observed prices of an exchange traded bond, as well as the price of the corresponding zero-coupon bond. When the bond was issued, the nominal value (100%) (almost exactly) reflects an interest rate of 5%.

2.9 RECURSIONS

The present value, as defined above in (2.2), satisfies the recursion

$$PV_{t_k} = C_{t_k} + \frac{1}{1+i_{t_k}} PV_{t_{k+1}}.$$

This is called the *prospective* recursion, as the recursion takes only values into account which correspond to payments and present values in the future.

The *retrospective* formula for the present value is

$$PV_{t_{k-1}} \left(1 + i_{t_{k-1}}\right) - \left(1 + i_{t_{k-1}}\right) C_{t_{k-1}} = PV_{t_k}.$$

In practice, retrospective recursions are not used, but prospective recursions occur very frequently. In what follows we list prospective recursions for selected present values outlined in Lecture 2.6. Recursion have notably a very natural interpretation, they describe the time evolution of the entire process of accumulated, discounted cashflows.

(i) Lump-sum payment: The recursion is trivial, as there is no cash flow:

$$v^t = v \cdot v^{t-1};$$

(ii) Annuity:

$$\ddot{a}_{\overline{n}|} = 1 + v \cdot \ddot{a}_{\overline{n-1}|};$$

(iii) Monthly annuity:

$$\ddot{a}^{(m)}_{\overline{n}|} = \ddot{a}^{(m)}_{\overline{1}|} + v \cdot \ddot{a}^{(m)}_{\overline{n-1}|};$$

(iv) Perpetuity:

$$\ddot{a} = 1 + v \cdot \ddot{a};$$

(v) Loan: the present value can be considered *before* or *after* the respective payment at given time. For the sake of consistency we shall restrict ourselves to the situation of payments just an instance *before* the respective payments. The present value thus are $PV_0 = 0$ and $PV_t = \ddot{a}^{(m)}_{\overline{n+1-t}|}$, and the respective cash flows are $C_0 = -a^{(m)}_{\overline{n}|}$ and $C_t = \frac{1}{m}$.

The recursion therefore is just as in the situation of an annuity,

$$\ddot{a}^{(m)}_{\overline{n}|} = \ddot{a}^{(m)}_{\overline{1}|} + v \cdot \ddot{a}^{(m)}_{\overline{n-1}|}, \tag{2.15}$$

in particular

$$\ddot{a}_{\overline{n}|} = 1 + v\, \ddot{a}_{\overline{n-1}|}$$

for an annually paid loan.

(For the sake of completeness, the recursion *after* the payments is $a_{\overline{n}|} = v + v \cdot a_{\overline{n-1}|}$.)

2.10 THE ANNUAL PERCENTAGE RATE

In the previous lecture (Lecture 2.6) we have computed the present value of some important investments with constant interest rate following the baseline formula (2.1). In this lecture we consider the inverse question, that is, we are interested in the constant yield corresponding to a sequence of future payments and a given present value.

To make the dependence on the yield i and the cashflows $C = (C_t)$ explicit in (2.1) we write

$$PV^C(i) \coloneqq \sum_t \frac{C_t}{(1+i)^t}. \tag{2.16}$$

Consider a cash flow C_t and the corresponding present value PV^C, computed according (2.16) with a particular interest rate i. Define the cashflow

$$C'_t \coloneqq \begin{cases} C_0 - PV^C & \text{if } t = 0 \\ C_t & \text{else.} \end{cases}$$

By construction, this modified cash flow C' satisfies the equation

$$PV^{C'}(i) = 0$$

for the initial interest i we started with.

For a general sequence of cash flows the particular interest rate $\tilde{i}$ solving the equation

$$PV^C(\tilde{i}) = 0 \tag{2.17}$$

is called annual percentage rate (APR) or internal rate of return. Notice, that the cash flows must change signs in time in order to get a meaningful APR.

Remark 2.4. The solution of (2.17) is – possibly – not unique. In this situation the biggest solution is the APR.

Remark 2.5 (Z-spread). The *Z-spread* of a bond is the number of basis points one needs to apply to a series of zero rates such that the present value of the bond, accounting for accrued interest, equals the sum of all future cash flows discounted using the adjusted zero rate. The spread is calculated iteratively and provides a more accurate reflection of value than other measures as it uses the entire yield curve to value the cash flows.

Remark. For a fixed interest rate the *Z-spread* satisfies the equation

$$PV(i + Z) = 0$$

(i.e., $\tilde{i} = i + Z$).

2.11 PROBLEMS

Exercise 2.1. *Motivate and explain the approximation (2.8).*

Exercise 2.2. *Compute $i^{(m)}$ and $d^{(m)}$ for $m = 1, 12, 365$ and for $m = \infty$ and for an interest rate of your choice.*

Exercise 2.3. *Compute $\alpha^{(m)}$ and $\beta^{(m)}$ for $m = 1, 12, 365$ and for $m = \infty$ (and again choose the interest rate).*

Exercise 2.4. *Show that*

$$1 = d^{(m)} \cdot \ddot{a}_{\overline{n}|}^{(m)} + v^n; \tag{2.18}$$

what does the formula mean economically?

Exercise 2.5. *Verify the last line in Table 2.2.*

Exercise 2.6. *Verify the relations (i)–(x) on page 22.*

Exercise 2.7. *A savings book offers the following annual returns:*

months	*0–11*	*12–23*	*24–35*	*36–47*	*48–60*
yield	*1.1%*	*1.3%*	*2.1%*	*2.6%*	*3.0%*

After the entire period of 5 years the principal amount of 1000 will grow to 1000 € ∗ 1.011 ∗ 1.03 ∗ 1.021 ∗ 1.026 ∗ 1.03 = 1500 €*. What is the average, constant yield (the APR) of the investment?*

Exercise 2.8. *The savings*

time/ year	*0*	*1*	*2*	*3*	*4*
amount	*1000*	*2000*	*0*	*4000*	*7000*

accumulate to 15000 in year 5. What is the APR of these investments?

Exercise 2.9. *Compute δ, d, r and v for $i = 1\,\%$, $2\,\%$ and an up-to-date interest rate of a 10 years governmental bond.*

Exercise 2.10. *Express the interest rate etc. by completing the following table:*

		i	δ	d	r	v
interest rate	i	i	$e^{\delta} - 1$			
force of interest	δ	$\ln(1+i)$	δ			
discount rate	d	$\frac{i}{1+i}$		d		
percentage rate	r	$1+i$			r	
discount factor	v	$\frac{1}{1+i}$				v

3 Hedging Interest: Duration and Convexity

The problem with socialism is that you eventually run out of other people's money.

Margaret Thatcher, 1925 – 2013

3.1 DEFINITION AND RELATIONS

The present value, as defined in its general context in (2.16), depends on the future interest rate. The financial instrument usually is not based on some *fixed*, future interest rate, and an initial amount B_0 will be compounded (by the bank, say) to the quantity $B_t = B_0 \cdot e^{\delta \cdot t}$. Or reverse, a future, fixed payment of B_t, discounted, is worth

$$B_t \cdot e^{-\delta \cdot t}$$

today ($t = 0$), reflected by the present value.

The concept of duration was introduced by Macaulay.[1]

Definition 3.1. The *duration* is defined as

$$D := -\frac{\frac{\partial}{\partial \delta} PV_t}{PV_t} = -\frac{\partial}{\partial \delta} \log PV_t .$$

Note, that the dimension of D_t is time ($[D_t] = [\text{year}]$), which justifies the term duration, at least to some extend.

The duration of the present value given (2.16), in explicit terms, is

$$D = -\frac{\frac{\partial}{\partial \delta} PV_t}{PV_t} = -\frac{1}{PV_t} \frac{\partial}{\partial \delta} \sum_t C_t \cdot e^{-\delta \cdot t} = \frac{1}{PV_t} \sum_t C_t \cdot t \cdot e^{-\delta \cdot t},$$

which is a weighted average.

Remark 3.2. For the present value $PV_t = \sum_{\ell=0}^n e^{-\delta t_\ell} C_{t_\ell}$, the duration is

$$D = \frac{\sum_{\ell=0}^n e^{-\delta t_\ell} C_{t_\ell} \cdot t_\ell}{\sum_{\ell=0}^n e^{-\delta t_\ell} C_{t_\ell}} = \sum_{\ell=0}^n \underbrace{\frac{e^{-\delta t_\ell} C_{t_\ell}}{\sum_{j=0}^n e^{-\delta t_j} C_{t_j}}}_{w_\ell} \cdot t_\ell = \sum_{\ell=0}^n w_\ell \cdot t_\ell ,$$

where w_ℓ are weights with $\sum_{\ell=1}^n w_\ell = 1$.

Recall, that $i = e^\delta - 1$ is a function of δ, and thus $\frac{di}{d\delta} = e^\delta = 1 + i$: this is the reason for

$$D = -\frac{\frac{d}{di} PV_t}{PV_t} \cdot \frac{di}{d\delta} = -\frac{1+i}{PV_t} \cdot \frac{\partial}{\partial i} PV_t ,$$

[1] Frederick Macaulay, 1882–1970, Canadian economist

which is frequently used as an alternative definition for the duration. However, the factor $1+i$ is usually small (1.03, e.g.) and thus may be neglected in real-world situations. Recall moreover that

$$\frac{d\delta}{di} = \frac{d}{di}\log(1+i) = \frac{1}{1+i} = v,$$

which relates the derivative with to the force of interest δ and i the interest directly.

3.2 EXAMPLES

3.2.1 Lump-Sum Payment

Recall, the present value for the lump sum payment of 1 in t years is $PV = v^t = e^{-\delta\cdot t}$. The duration thus is the term of the contract,

$$D = -\frac{1}{e^{-\delta\cdot t}}\frac{d}{d\delta}e^{-\delta\cdot t} = -\frac{1}{e^{-\delta\cdot t}}\left(-t\cdot e^{-\delta\cdot t}\right) = t.$$

3.2.2 Annuity

The closed form

$$\begin{aligned} D - \frac{\frac{d}{d\delta}\ddot{a}^{(m)}_{\overline{n}|}}{\ddot{a}^{(m)}_{\overline{n}|}} &= -\frac{\frac{d}{d\delta}\frac{1}{m}\sum_{j=0}^{nm-1}e^{-\delta j/m}}{\frac{1}{m}\sum_{j=0}^{nm-1}e^{-\delta j/m}} = -\frac{\frac{d}{d\delta}\frac{1-e^{-\delta n}}{1-e^{-\delta/m}}}{\frac{1-e^{-\delta n}}{1-e^{-\delta/m}}} = -\frac{\frac{ne^{-\delta n}}{1-e^{-\delta/m}} - \frac{1}{m}\frac{(1-e^{-\delta n})e^{-\delta/m}}{(1-e^{-\delta/m})^2}}{\frac{1-e^{-\delta n}}{1-e^{-\delta/m}}} \\ &= \frac{1}{i^{(m)}} - \frac{n\cdot v^n}{1-v^n} = \frac{1}{i^{(m)}} - \frac{1}{i^{(1/n)}} \end{aligned} \tag{3.1}$$

can be found by taking the derivative after some simplifications.[2] Asymptotically, it holds that

$$D = \frac{mn-1}{2m} - \frac{n^2m^2-1}{12m^2}\delta + O\left(\delta^2\right).$$

3.2.3 Perpetuity

The closed form

$$D = \frac{1}{i^{(m)}}$$

can be found by letting $n \to \infty$ in (3.1).

This quantity is ∞, if the interest rate $i = 0$. However – and this may somehow contradict the intuition, as the annuity itself never will stop – is finite for positive interest rates.

3.2.4 Loan

The closed form is

$$D = \frac{1}{i^{(m)}} - \frac{n\cdot v^n}{1-v^n}.$$

Notice, that the duration is $D = \frac{n}{2} + \frac{1}{2m} - \mathcal{O}(i)$.

[2]Differentiate (2.5) with respect to v to get $\sum_{k=0}^{n-1} k\cdot v^{k-1} = \frac{1-n\cdot v^{n-1}+(n-1)v^n}{(1-v)^2}$

3.3 REMARK

(i) Notice, usually it does not make sense to compute the duration for cash flows which have different signs. In this situation it may be advantageous to compute the duration for the positive and negative cash flows separately.

(ii) A usual bank has loans on the asset side, and savings books on the liability side of the balance sheet. Both sides have durations, and they may differ: for example, if there are loans with a fixed interest rate in the portfolio, or savings books with fixed interest rates on the other side.

(iii) The duration is related to a quantity denoted ρ, which will be considered and described as well in the sequel; $\rho = \frac{\partial}{\partial i} PV$ is the derivative with respect to the interest rate. ρ is, in contrast, *not* adjusted by the additional quotient $-PV$ of the financial instrument.

3.4 HEDGING INTEREST — DURATION MATCHING

A financial position is immunized against a change of the interest rate, if the weighted durations on long and short positions coincide – this is often referred to as *duration matching*. Such a position may be considered as zero coupon bond.

In this case the profit is not vulnerable with respect to changing interest rates.

3.5 CONVEXITY

Convexity measure of the sensitivity of the duration of a financial instrument to changes in interest rates.

Definition 3.3. Convexity is

$$\begin{aligned} C := & \frac{1}{PV_t} \cdot \frac{\partial^2}{\partial \delta^2} PV_t \\ = & \frac{1}{PV_t} \left((1+i)^2 \frac{\partial^2}{\partial i^2} PV_t + (1+i) \frac{\partial}{\partial i} PV_t \right) \end{aligned} \tag{3.2}$$

The related formulae are

$$C = \frac{1}{PV_t} \cdot \sum_t t^2 \cdot C_t \cdot e^{-\delta \cdot t} = \frac{1}{PV_t} \cdot \sum_t t^2 \cdot \frac{C_t}{(1+i)^t}.$$

3.6 PROBLEMS

Exercise 3.1. *Verify the durations provided in the examples above (Lecture 3.2.1–3.2.4).*

Exercise 3.2. *Verify (3.2) to compute the convexity.*

Duality in Optimization

4

4.1 DUALITY

Proposition 4.1 (*Max–min inequality*)**.** Any *real-valued function* L *on* $D \times \Lambda$ *satisfies the* max-min-inequality

$$\underbrace{\sup_{\lambda\in\Lambda} \underbrace{\inf_{x\in D} L(x;\lambda)}_{=:d(\lambda)}}_{d^*} \quad \le \quad \underbrace{\inf_{x\in D} \sup_{\lambda\in\Lambda} L(x;\lambda)}_{p^*}\,. \tag{4.1}$$

- the inequality $d^* \le p^*$ is called *weak duality*, and
- $p^* - d^* \ge 0$ is the *duality gap*;
- in case of $d^* = p^*$, L is said to have the *strong max-min* property, *strong duality* or *saddle-point* property;
- the function[1]

  $$d(\lambda) \coloneqq \inf_{x\in D} L(x;\lambda) \tag{4.2}$$

  is called *dual function*. Obviously, $d(\lambda) \le d^* \le p^*$.

Definition 4.2 (Saddle point)**.** A point (x^*, λ^*) is a *saddle point* if

$$L\left(x^*;\lambda\right) \le L\left(x;\lambda^*\right) \qquad \text{for all } x \text{ and all } \lambda$$

(in this case, $L(x^*;\lambda) \le L(x^*;\lambda^*) \le L(x;\lambda^*)$).

Remark 4.3. Any saddle point satisfies $\sup_{\lambda\in\Lambda} L\left(x^*;\lambda\right) \le \inf_{x\in D} L\left(x;\lambda^*\right)$.

Existence of a saddle point implies the strong max-min property and $d^* = d\left(\lambda^*\right) = L\left(x^*;\lambda^*\right) = p^*$, because

$$\begin{aligned} p^* &= \inf_{x\in D} \sup_{\lambda\in\Lambda} L\left(x;\lambda\right) \\ &\le \sup_{\lambda\in\Lambda} L\left(x^*;\lambda\right) \\ &\le \inf_{x\in D} L\left(x;\lambda^*\right) = d\left(\lambda^*\right) \\ &\le \sup_{\lambda\in\Lambda} \inf_{x\in D} L\left(x;\lambda\right) = d^*. \end{aligned} \tag{4.3}$$

Definition 4.4. A function $f\colon X \to \mathbb{R}$ is lower semi-continuous (lsc) if $\{x\colon f(x) > \alpha\}$ is open for every $\alpha \in \mathbb{R}$ (i.e., the level sets $\{x\colon f(x) \le \alpha\}$ are closed).

[1] By convention, $\inf\{\} = +\infty$, $\sup\{\} = -\infty$, resp. (cf. Footnote 11 on page 49).

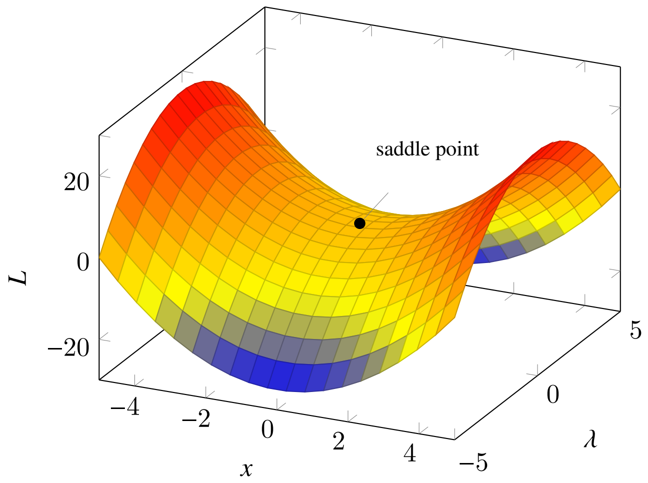


Figure 4.1: Saddle point

Lemma 4.5. *For a metric space X, $f\colon X \to \mathbb{R}$ is lsc. if* $\liminf_{x\to x_0} f(x) \geq f(x_0)$ *for every* $x_0 \in X$.

Sion's minimax theorem provides a sufficient condition for strong duality to hold.

Theorem 4.6 (Sion's minimax theorem, cf. Sion [23] or Komiya [12])**.**

(i) Let D and Λ be convex and (at least) one of these sets compact,

(ii) $x \mapsto L(x,\lambda)$ (quasi-)convex and lsc. for any $\lambda \in \Lambda$ and

(iii) $\lambda \mapsto L(x,\lambda)$ (quasi-)concave and usc. for any $x \in D$,

then L has the strong max-min property.

4.2 LAGRANGIAN

To investigate the *primal problem*

$$\begin{aligned} &\text{minimize (in x)} && f(x) \\ &\text{subject to} && g_j(x) \leq 0,\; j = 1,\dots,n, \\ &&& h_i(x) = 0,\; i = 1,\dots,m, \\ &&& x \in D \end{aligned} \tag{P}$$

define the *Lagrange-function*[2] on $D \times \Lambda$ with $\Lambda := \mathbb{R}^m \times \mathbb{R}^n_{\geq 0}$ as

$$L(x;\lambda,\mu) := f(x) + \sum_{i=1}^{m} \lambda_i\, h_i(x) + \sum_{j=1}^{n} \mu_j\, g_j(x), \tag{4.4}$$

($\lambda_i \in \mathbb{R}$ for all $i = 1,\dots,n$ and $\mu_j \geq 0$, $j = 1,\dots,m$). The *Lagrange dual function*, as defined in (4.2), is the *concave* function

$$d(\lambda,\mu) := \inf_{x\in D} L(x;\lambda,\mu); \tag{4.5}$$

[2]Joseph–Louis Lagrange, 1736–1813

note that $d(\lambda,\mu)$ is concave, as it is the infimum of linear functions.

The (unconstrained) *Lagrange dual problem* is the *concave* problem

$$\begin{aligned} \text{maximize} \quad & d(\lambda,\mu) \\ \text{subject to} \quad & \lambda \in \mathbb{R}^m, \\ & \mu \in \mathbb{R}^n \text{ with } \mu_j \geq 0 \text{ for all } j. \end{aligned} \tag{D}$$

Theorem 4.7. *(x^*,λ^*,μ^*) is a saddle point for the Lagrangian L (Eq. (4.4)) iff*[3]

(i) x^* is primal optimal,

(ii) (λ^*,μ^*) is dual optimal and

(iii) strong duality is obtained.

In addition, $d^* = d(\lambda^*,\mu^*) = L(x^*;\lambda^*,\mu^*) = f(x^*) = p^*$ and $\mu_j^* g_j(x^*) = 0$ (complementary slackness).

Corollary. *Let x^* be primal optimal and (λ^*,μ^*) dual optimal, but with strictly positive duality gap. Then there does not exist any saddle-point, but the following inequalities hold:*

$$d^* = d(\lambda^*,\mu^*) \leq L(x^*;\lambda^*,\mu^*) = L(x^*;0,\mu^*) \leq f(x^*) = p^*,$$

and consequently $0 \leq -\mu^{*\top} g(x^*) \leq p^* - d^*$. The saddle point inequality rewrites

$$L(x^*;\lambda,\mu) - f(x^*) \leq 0 \leq L(x;\lambda^*,\mu^*) - d(\lambda^*,\mu^*)$$

for all x, λ and $\mu \geq 0$.

Proof of Theorem 4.7. Let (x^*,λ^*,μ^*) be a saddle point, then

$$d(\lambda,\mu) \leq L(x^*;\lambda,\mu) \leq \inf_{x\in D} L(x;\lambda^*,\mu^*) = d(\lambda^*,\mu^*),$$

which shows that (λ^*,μ^*) is optimal for the dual, thus (ii).

Strong duality (i.e., (iii)) follows via (4.3) since we assume a saddle-point.

In addition

$$\begin{aligned} f(x^*) + \lambda^\top h(x^*) + \mu^\top g(x^*) &= L(x^*;\lambda,\mu) \\ &\leq L(x^*;\lambda^*,\mu^*) \\ &= f(x^*) + \lambda^{*\top} h(x^*) + \mu^{*\top} g(x^*) \end{aligned}$$

for all λ and $\mu \geq 0$, hence $h(x^*) = 0$ and $g(x^*) \leq 0$, which shows that x^* is feasible for the primal problem. Consequently $\mu^\top g(x^*) \leq \mu^{*\top} g(x^*) \leq 0$ for all $\mu \geq 0$, so we deduce $\mu_j^* g_j(x^*) = 0$ (complementary slackness).

Again from the saddle-point-property

$$\begin{aligned} f(x^*) &= L(x^*;\lambda^*,\mu^*) \\ &\leq L(x;\lambda^*,\mu^*) \\ &= f(x) + \lambda^{*\top} \underbrace{h(x)}_{=0} + \underbrace{\mu^{*\top} g(x)}_{\leq 0} \end{aligned}$$

and so it follows that x^* is indeed optimal for the primal problem, thus (i).

[3] John von Neumann, 1903–1957

Conversely, observe that

$$\sup_{\mu\geq 0,\,\lambda} \underbrace{f(x)+\lambda^\top h(x)+\mu^\top g(x)}_{L(x;\lambda,\mu)} = \begin{cases} f(x) & \text{if } h_i(x)=0 \text{ and } g_j(x)\leq 0,\\ \infty & \text{else.}\end{cases}$$

Thus, assuming that x^* is primal optimal and (λ^*, μ^*) dual optimal,

$$d^* = d(\lambda^*, \mu^*) \leq L(x^*; \lambda^*, \mu^*) \leq f(x^*) = p^*,$$

and consequently $0 \leq {\mu^*}^\top g(x^*) \leq p^* - d^*$.

Moreover,

$$\begin{aligned} L(x^*; \lambda, \mu) &\leq \sup_{\lambda,\mu\geq 0} L(x^*; \lambda, \mu)\\ &= \inf_{x\in D} \sup_{\lambda,\mu\geq 0} L(x; \lambda, \mu)\\ &\underset{(4.1)}{=} \sup_{\lambda,\mu\geq 0} \inf_{x\in D} L(x; \lambda, \mu) + p^* - d^*\\ &= \sup_{\lambda,\mu\geq 0} d(\lambda, \mu) + p^* - d^*\\ &= d(\lambda^*, \mu^*) + p^* - d^*\\ &\leq L(x, \lambda^*, \mu^*) + p^* - d^* \end{aligned}$$

for all x, λ and $\mu \geq 0$, establishing the saddle-point inequality. □

4.3 LINEAR PROGRAMS

The linear programs in Table 4.1 are dual – in the sense described – to each other.[4]

Example 4.8. Consider the fifth primal problem in Table 4.1. With $D = \{x \geq 0\}$ and $g(x) = b - Ax$ the Lagrange-dual function (4.5) is

$$d(\mu) = \inf_{x\geq 0} c^\top x + \mu^\top (b - Ax) = \inf_{x\geq 0} \mu^\top b + \left(c^\top - \mu^\top A\right) x = \begin{cases} \mu^\top b & \text{if } c^\top - \mu^\top A \geq 0,\\ -\infty & \text{else.}\end{cases}$$

The dual problem (D) thus is the first dual in Table 4.1.

Remark 4.9. Suppose that the dual program in Table 4.1 has a unique solution μ and/ or λ. Then the derivative of the primal program with respect to b is μ for inequality constraints, and λ for the equality constraint.

4.4 QUADRATIC PROGRAMS

Consider the problem

$$\begin{aligned} \text{minimize } & \frac{1}{2} x^\top H x + c^\top x\\ \text{subject to } & A_1 x = b_1,\\ & A_2 x \geq b_2. \end{aligned}$$

[4] George Dantzig, 1914–2005; Leonid Kantorovich, 1912–1983

Linear Program (primal)		Dual Program	
minimize (in x) subject to	$c^\top x$ $A_1 x = b_1$ $A_2 x \ge b_2$	maximize (in λ,μ) subject to	$\lambda^\top b_1 + \mu^\top b_2$ $\lambda^\top A_1 + \mu^\top A_2 = c^\top$ $\mu \ge 0$
minimize (in x) subject to	$c^\top x$ $A_1 x = b_1$ $A_2 x \ge b_2$ $x \ge 0$	maximize (in λ,μ) subject to	$\lambda^\top b_1 + \mu^\top b_2$ $\lambda^\top A_1 + \mu^\top A_2 \le c^\top$ $\mu \ge 0$
minimize (in x) subject to	$c^\top x$ $Ax = b$	maximize (in λ) subject to	$\lambda^\top b$ $\lambda^\top A = c^\top$
minimize (in x) subject to	$c^\top x$ $Ax = b$ $x \ge 0$	maximize (in λ) subject to	$\lambda^\top b$ $\lambda^\top A \le c^\top$
minimize (in x) subject to	$c^\top x$ $Ax \ge b$	maximize (in μ) subject to	$\mu^\top b$ $\mu^\top A = c^\top$ $\mu \ge 0$
minimize (in x) subject to	$c^\top x$ $Ax \ge b$ $x \ge 0$	maximize (in μ) subject to	$\mu^\top b$ $\mu^\top A \le c^\top$ $\mu \ge 0$

Table 4.1: Duality of important linear programs

The Lagrangian is

$$\begin{aligned} L(x;\lambda,\mu) &\coloneqq \frac{1}{2}x^\top Hx + c^\top x + \lambda^\top (b_1 - A_1 x) + \mu^\top (b_2 - A_2 x) \\ &= \frac{1}{2}x^\top Hx + \left(c^\top - \lambda^\top A_1 - \mu^\top A_2\right) x + \lambda^\top b_1 + \mu^\top b_2. \end{aligned}$$

Note, that $0 = \frac{\partial L}{\partial x} = x^\top H + (c^\top - \lambda^\top A_1 - \mu^\top A_2)$ and we abbreviate the optimal solution with $u \coloneqq -H^{-1}\left(c - A_1^\top \lambda - A_2^\top \mu\right)$. The dual function is

$$\begin{aligned} d(\lambda,\mu) &= \inf_x L(x;\lambda,\mu) \\ &= L(u;\lambda,\mu) \\ &= \frac{1}{2}u^\top Hu - u^\top Hu + \lambda^\top b_1 + \mu^\top b_2 \\ &= -\frac{1}{2}u^\top Hu + \lambda^\top b_1 + \mu^\top b_2. \end{aligned}$$

Following (D), the dual problem thus is

$$\begin{aligned} &\text{maximize } -\frac{1}{2}u^\top Hu + \lambda^\top b_1 + \mu^\top b_2 \\ &\text{subject to } \lambda^\top A_1 + \mu^\top A_2 - u^\top H = c^\top \text{ and} \\ &\qquad\qquad \mu \ge 0, \end{aligned}$$

as $\lambda^\top A_1 + \mu^\top A_2 - u^\top H = c$. Table 4.2 collects these results and generalizes the linear problems in Table 4.1.

Quadratic Program (primal)		Dual Program	
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $A_1 x = b_1$ $A_2 x \geq b_2$	maximize (in λ,μ) subject to	$-\frac{1}{2}u^\top Hu + \lambda^\top b_1 + \mu^\top b_2$ $\lambda^\top A_1 + \mu^\top A_2 - u^\top H = c^\top$ $\mu \geq 0$
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $A_1 x = b_1$ $A_2 x \geq b_2$ $x \geq 0$	maximize (in λ,μ) subject to	$-\frac{1}{2}u^\top Hu + \lambda^\top b_1 + \mu^\top b_2$ $\lambda^\top A_1 + \mu^\top A_2 - u^\top H \leq c^\top$ $\mu \geq 0$
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $Ax = b$	maximize (in λ) subject to	$-\frac{1}{2}u^\top Hu + \lambda^\top b$ $\lambda^\top A - u^\top H = c^\top$
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $Ax = b$ $x \geq 0$	maximize (in λ) subject to	$-\frac{1}{2}u^\top Hu + \lambda^\top b$ $\lambda^\top A - u^\top H \leq c^\top$
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $Ax \geq b$	maximize (in μ) subject to	$-\frac{1}{2}u^\top Hu + \mu^\top b$ $\mu^\top A - u^\top H = c^\top$ $\mu \geq 0$
minimize (in x) subject to	$\frac{1}{2}x^\top Hx + c^\top x$ $Ax \geq b$ $x \geq 0$	maximize (in μ) subject to	$-\frac{1}{2}u^\top Hu + \mu^\top b$ $\mu^\top A - u^\top H \leq c^\top$ $\mu \geq 0$

Table 4.2: Duality of important quadratic programs

4.5 FENCHEL–TRANSFORM

It is useful here to naturally extend the (concave) Lagrange-dual function by

$$d(\lambda, \mu) = \begin{cases} d(\lambda, \mu) & \text{if } \mu_j \geq 0, \\ -\infty & \text{else.} \end{cases}$$

We may state the dual problem (D) equivalently as

$$\begin{array}{ll} \text{minimize (in } \lambda, \mu) & -d\,(\lambda, \mu) \\ \text{subject to} & -\mu_j \leq 0, \end{array}$$

(the same form as (P) without h, but $g(\mu) \coloneqq -\mu$) and start from this problem as initial problem: the Lagrangian is $\tilde{L}(\lambda, \mu; y) = -d(\lambda, \mu) - y^\top \mu$, the corresponding concave dual function is

$$\begin{aligned} \tilde{d}(y) &= \inf_{\lambda, \mu \geq 0} \tilde{L}\,(\lambda, \mu; y) \\ &= \inf_{\lambda, \mu} -y^\top \mu - d(\lambda, \mu) \\ &= -\sup_{\lambda, \mu}\,(0, y)^\top \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + d(\lambda, \mu) \\ &= -\,(-d)^*\,(0, y)\,, \end{aligned}$$

where

$$f^*(y) \coloneqq \sup_x y^\top x - f(x)$$

is f's *convex conjugate* function (f^* is always convex and lsc.; other names are Fenchel transform, *Legendre-Fenchel transform*; note the *Fenchel-Young* inequality $x^\top y \le f(x) + f^*(y)$; we shall call $-(-f)^*$ *concave conjugate*).

The *dual-dual* problem, in view of (D), thus is

$$\begin{aligned} &\text{maximize (in y)} && \tilde{d}(y) \\ &\text{subject to} && y_j \ge 0. \end{aligned} \tag{DD}$$

We may start here *again* with the Lagrangian $\tilde{\tilde{L}}(y;\tilde{\mu}) = -\tilde{d}(y) - \tilde{\mu}^\top y$, the corresponding dual function thus is $\tilde{\tilde{d}}(\tilde{\mu}) = \inf_{y\ge 0} -\tilde{d}(y) - \tilde{\mu}^\top y = -\sup_y \tilde{\mu}^\top y + \tilde{d}(y) = -\left(-\tilde{d}\right)^*(\tilde{\mu})$, and the *dual-dual-dual* thus is

$$\begin{aligned} &\text{maximize (in } \tilde{\mu}) && -\left(-\tilde{d}\right)^*(\tilde{\mu}) \\ &\text{subject to} && \tilde{\mu}_j \ge 0. \end{aligned} \tag{DDD}$$

This is the same as (DD), but $\tilde{d}$ replaced by its concave conjugate $-\left(-\tilde{d}\right)^*$. The difference to the dual (D) is that we finally got rid of λ.

Repeating the procedure will lead us back to the (DD), as for convex (lsc.) functions $\left(-\tilde{d}\right)^{**} = -\tilde{d}$. Note the optimal values $d^* = d(\lambda^*, \mu^*) = \tilde{d}(y^*) = -\left(-\tilde{d}\right)^*(\tilde{\mu}^*)$ etc..

4.6 KARUSH–KUHN–TUCKER (KKT)

Let L be differentiable in the saddle point (x^*, λ^*, μ^*), then $\nabla L(x^*, \lambda^*, \mu^*) = 0$ (notice the *simultaneous* differentiation with respect to all 3 variables).

From Theorem 4.7 we deduce: For *any* optimization problem with *differentiable objective and constraint functions* for which strong duality obtains, any pair of primal and dual optimal points must satisfy the conditions (KKT)

(i) Stationarity: $0 \in \partial f(x^*) + \sum_i \lambda_i^* \cdot \partial h_i(x^*) + \sum_j \mu_j^* \cdot \partial g_j(x^*)$ $(0 = \nabla_x L)$,

(ii) Primal feasibility: $h_i(x^*) = 0$, $g_j(x^*) \le 0$ $(\nabla_\lambda L = 0,\ \nabla_\mu L = 0)$,

(iii) Dual feasibility: $\mu_j^* \ge 0$ and

(iv) Complementary slackness: $\mu_j^* \cdot g_j(x^*) = 0$.

An element u^* out of the (locally convex) linear space's dual is called *subgradient*, iff the *subgradient inequality* $u^* \in \partial f(x) :\Longleftrightarrow f(z) \ge f(x) + u^*(z - x)$ holds for all z.

$\partial f(x)$, the (convex and closed) set of all subgradients in x, is called *subdifferential*.

Theorem 4.10. *Let x^* be primal optimal for the primal* (P) *(plus some regularity conditions), then there exist λ^* and μ^* such that (KKT).*

Remark. If the primal problem is convex, then (KKT) are also *sufficient* conditions for optimality of x^*, (λ^*, μ^*).

For differentiable f, g and h the problem

$$\begin{aligned} &\text{maximize (in } \lambda, \mu) && L(x;\lambda,\mu) \\ &\text{subject to} && \mu_j \ge 0, \\ & && \nabla_x L(x;\lambda,\mu) = 0 \end{aligned} \tag{WD}$$

is called *Wolfe dual problem*.

4.7 FARKAS' LEMMA

Theorem 4.11 (A Theorem on the Alternative)**.** *Exactly* one *of these following two statements holds true:*

- *There exists y such that $Wy = z$ and $y \geq 0$;*
- *There exists σ such that $\sigma^\top W \leq 0$ and $\sigma^\top z > 0$.*

4.8 DERIVATIVE

Theorem 4.12. *Consider the function*

$$f(x) \coloneqq \min \{ f(x, y) \colon g(x, y) \leq 0 \text{ and } h(x, y) = 0 \}$$

with minimizing argument $y(x)$ satisfying (KKT) for any x. Let f, g, h and $y \in C^1$, then

$$f'(x) = f_x\big(x, y(x)\big) + \lambda(x)^\top h_x\big(x, y(x)\big) + \mu(x)^\top g_x\big(x, y(x)\big)$$

with respective Lagrange multipliers (dependent on x).

Proof. As for the proof notice first that $h(x, y(x)) = 0$, thus $h_x + h_y y'(x) = 0$. Then we find either $g_i(x, y(x)) = 0$ or $g_i(x, y(x)) < 0 \wedge \mu_i = 0$ by complementary slackness, that is again $\mu_i \left(g_{i,x} + g_{i,y} y'(x)\right) = 0$ and $\mu^\top \left(g_x + g_y y'(x)\right) = 0$.

Now recall that $f(x) = f(x, y(x))$ and $f_y + \lambda^\top h_y + \mu^\top g_y = 0$ (KKT). Thus

$$\begin{aligned} f'(x) &= f_x + f_y y'(x) \\ &= f_x - \lambda^\top h_y y'(x) - \mu^\top g_y y'(x) \\ &= f_x + \lambda^\top h_x + \mu^\top g_x. \end{aligned}$$

□

4.9 PROBLEMS

Exercise 4.1. *Verify the duality relations in Table 4.1 as outlined in Example 4.8.*

Part II

Stochastic Finance

5 Elements of Probability

Recap of ingredients from probability theory and statistics to fix the notation.

5.1 THE PROBABILITY SPACE

Definition 5.1 (Probability space)**.** The probability space is a triple

$$(\Omega, \mathcal{F}, P),$$

where Ω is the sample space, $\mathcal{F}$ is the sigma algebra on Ω and $P\colon \mathcal{F} \to [0,1]$ the probability measure.

5.1.1 The Sample Space

The sample space is an arbitrary set and usually denoted Ω. The elements $\omega \in \Omega$ are occasionally called *outcomes* and Ω collects all outcomes.

5.1.2 Sigma Algebra

Definition 5.2 (Algebra)**.** Let Ω be a set. A collection $\mathcal{F}$ of subsets of Ω is called an *algebra* on Ω (or algebra of subsets of Ω) if

(i) $\Omega \in \mathcal{F}$,

(ii) for each $E \in \mathcal{F}$ we have that $E^{\mathrm{c}} \in \mathcal{F}$, where $E^{\mathrm{c}} \coloneqq \{\omega \in \Omega\colon \omega \notin E\} = \Omega \setminus E$ is the complement of E, and

(iii) if $E, F \in \mathcal{F}$, then $E \cup F \in \mathcal{F}$.

Note that $E \cap F = (E^{\mathrm{c}} \cup F^{\mathrm{c}})^{\mathrm{c}}$. An algebra on Ω thus is a family of subsets of Ω which is stable under finitely many set operations.

Definition 5.3 (Sigma algebra)**.** An algebra $\mathcal{F}$ is a *sigma algebra*, or *field of subsets* if

(i) $\mathcal{F}$ is an algebra, and

(ii) for a *countable* sequence $E_1, E_2, \dots$ of events it holds that $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$.

An element $E \in \mathcal{F}$ is often called an *event*, or a *possible event*. Sigma algebras are often denoted by $\mathcal{F}$, sometimes also by Σ.

Remark 5.4. Note that by (ii) and Exercise 5.7 below we can say that a sigma algebra on Ω is a family of subsets of Ω which is *stable under any countable collection of set operations*.

Definition 5.5 (Borel sigma algebra)**.** On a topological space (Ω, τ), the Borel[1] sigma algebra $\mathscr{B}$ is the smallest sigma algebra containing τ and denoted (with slight abuse of notation)

$$\mathscr{B}(\Omega) \coloneqq \sigma(\tau).$$

[1] Émile Borel, 1871–1956

Remark 5.6. It is in fact difficult (but not impossible without the Axiom of Choice) to find a subset $E \subset \mathbb{R}$ which is not contained in $\mathscr{B}(\mathbb{R})$.

Remark 5.7. It was a famous mistake made by Lebesgue to assume that the projection of a Borel set in $\mathbb{R}^2$ is a Borel set in $\mathbb{R}$.

Definition 5.8. The pair $(\Omega, \mathcal{F})$ is called *measurable space*.

Remark 5.9. The sigma algebra is designed to model *information*. More information is available for $\mathcal{F}'$, if $\mathcal{F}' \supseteq \mathcal{F}$.

5.1.3 Probability Measure

Definition 5.10 (Probability measure)**.** A function

$$P \colon \mathcal{F} \to [0, 1]$$

is said to be a *probability measure* if the following are satisfied

(i) $P(\{\}) = 0$ and

(ii) P is *countably additive*, i.e., for a *countable* sequence of events $E_1, E_2, \dots \in \mathcal{F}$ which are pairwise disjoint (i.e., $E_m \cap E_n = \emptyset$ whenever $m \neq n$) it holds that

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n).$$

Remark 5.11. The inequality $P\left(\bigcup_{n=1}^{\infty} E_n\right) \le \sum_{n=1}^{\infty} P(E_n)$ for measurable sets is called *Boole's inequality* of *union bound*.

Remark 5.12. Note that the role of a sigma algebra is to provide sets, such that it is possible to assign a probability. In practice, it is difficult to find a non-measurable set (cf. Remark 5.6).

Theorem 5.13 (Carathéodory's[2] Extension Theorem)**.** *Let Ω be a set, $\mathcal{F}_0$ and define $\mathcal{F} := \sigma(\mathcal{F}_0)$. For a countably additive map $\mu_0 \colon \mathcal{F}_0 \to [0, 1]$ there exists a measure $\mu \colon \mathcal{F} \to [0, 1]$ which extends μ_0, i.e., $\mu = \mu_0$ on $\mathcal{F}_0$.*

5.2 BOREL–CANTELLI LEMMAS

Lemma 5.14 (First Borel–Cantelli Lemma)**.** *Let E_n be a sequence of events such that $\sum_{n=1} P(E_n) < \infty$. Then*

$$P\left(\limsup_{n\to\infty} E_n\right) = P\left(E_n, \text{ infinitely often}\right) = 0,$$

where

$$\limsup E_n := \bigcap_{m\ge 1} \bigcup_{n\ge m} E_n = \{\omega \colon \omega \in E_n \text{ for infinitely many } n\}.$$

Proof. Define $G_m := \bigcup_{n\ge m} E_n$. Then $G_m \supset G := \limsup E_n$ and consequently

$$P(G) \le P(G_m) \le \sum_{n\ge m} P(E_n) \xrightarrow[m\to\infty]{} 0.$$

This is the assertion. □

[2] Constantin Carathéodory, 1873–1950

Lemma 5.15 (Second Borel–Cantelli Lemma). *If E_n is a sequence of* independent *events and $\sum_{n=1} P(E_n) = \infty$, then*

$$P\left(\limsup_{n\to\infty} E_n\right) = 1.$$

Proof. Note first that $G_m \supset G_{m+1}$, and thus $\bigcap_{n\geq m} E_n^c = G_m^c$ is *in*creasing, as m increases. It follows for $m' > m$ that

$$P\left(\bigcap_{n\geq m} E_n^c\right) \leq P\left(\bigcap_{n\geq m'} E_n^c\right) = \prod_{n\geq m'} (1 - P(E_n)) \leq \exp\left(-\sum_{n\geq m'} P(E_n)\right) \to 0.$$

by independence and as $1 - x \leq e^{-x}$ whenever $x \geq 0$, for every m. Hence

$$P\left((\limsup E_n)^c\right) = P\left(\bigcup_m \bigcap_{n\geq m} E_n^c\right) \leq \sum_{m=0} P\left(\bigcap_{n\geq m} E_n^c\right) = 0,$$

from which the assertion follows. □

5.3 RANDOM VARIABLES

Definition 5.16 (Random variable). Let $(\Omega, \mathcal{F})$ and (S, Σ) be measurable spaces. A function

$$X\colon \Omega \to S$$

is said to be *measurable* if

$$X^{-1}(S) \in \mathcal{F} \text{ whenever } S \in \Sigma.$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space and (S, Σ) a measurable space. A *random variable* X is a measurable function defined on a probability space.

Definition 5.17. For a random variable $X\colon (\Omega, \mathcal{F}) \to (S, \Sigma)$,

$$P^X(A) \coloneqq P\left(X^{-1}(A)\right) \qquad (A \in \Sigma)$$

defines a probability measure on Σ, called the *image measure* (a.k.a. pushfoward measure and denoted X_*P or $X\#P$).

$$X^{-1}(A) \coloneqq \{\omega \in \Omega\colon X(\omega) \in A\}$$

is the *preimage* (or *inverse image*).

5.4 EXPECTATION AND INTEGRATION

For a random variable $X\colon \Omega \to S$, the *expectation* is denoted

$$\mathbb{E}\, X = \int X dP = \int_\Omega X(\omega) P(d\omega) = \int_S x\, P(X \in dx).$$

For a simple random variable $X = \sum_{i=1}^n X_i\, \mathbb{1}_{E_i}$ and a function $g\colon S \to \mathbb{R}$ this is $\mathbb{E}\, g(X) = \sum_{i=1}^n g(X_i) P(E_i)$. More generally,

$$\mathbb{E}\, g(X) = \int g(X) dP = \int_\Omega g\big(X(\omega)\big) P(d\omega) = \int_S g(x)\, P(X \in dx).$$

$\mu_1 = \kappa_1$	$\kappa_1 = \mu_1$	mean $\mu = \mathbb{E}\, X$
$\mu_2 = \kappa_2 + \kappa_1^2$	$\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$	variance $\sigma^2 = \mathbb{E}\,(X-\mu)^2$
$\mu_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3$	$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$	skewness $\gamma_1 = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^3$
$\mu_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4$	$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$	kurtosis $\kappa = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^4$
		excess $:=$ kurtosis $-$ 3

Table 5.1: Selected relations of moments and cumulants

Definition 5.18. The moment generating function is

$$m_X(t) := \mathbb{E}\, e^{tX}, \tag{5.1}$$

the cumulant-generating function is

$$K_X(t) := \log \mathbb{E}\, e^{tX}. \tag{5.2}$$

By linearity of the expectation and the Taylor series expansion $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots$ it follows that

$$m_X(t) = 1 + t\,\mathbb{E}\, X + \frac{t^2}{2}\,\mathbb{E}\, X^2 + \frac{t^3}{3!}\,\mathbb{E}\, X^3 + \dots,$$

where $\mu_k(X) := \mathbb{E}\, X^k$ is the kth moment (see Table 5.1 for the derived quantities skewness[3] and kurtosis[4],). The cumulants (aka semi-invariants) κ_n, $n = 1, 2, \dots$ are defined by the Taylor series expansion

$$K(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}.$$

The *characteristic function* (i.e., *Fourier transform*) is

$$\varphi_X(t) := \mathbb{E}\, e^{itX} \tag{5.3}$$

and the Laplace transform $L_X(t) := \mathbb{E}\, e^{-tX}$: note the relations $\varphi_X(t) = m_X(it) = L_X(-it)$.

5.5 REAL-VALUED RANDOM VARIABLES

Let $X : \Omega \to \mathbb{R}$ be a real-valued random variable (often denoted as $X \in \mathbb{R}$).

Definition 5.19. *The variance* is

$$\operatorname{var} X := \mathbb{E}\,(X - \mathbb{E}\, X)^2 = \mathbb{E}\left(X^2\right) - (\mathbb{E}\, X)^2,$$

the *standard deviation* is

$$\sigma(X) := \sqrt{\operatorname{var}(X)}.$$

Proposition 5.20. *It holds that* $\operatorname{var}(\alpha X + \beta) = \alpha^2 \operatorname{var} X$, *where* $\alpha \in \mathbb{R}$.

Proposition 5.21. *If X has the density f, then* $\operatorname{var} X = \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (u-v)^2 f(u) f(v)\, du\, dv$.

Proposition 5.22. *The following hold true:*

[3] Schiefe, dt.
[4] Wölbung, dt.

- ▹ *If* $\mathbb{E}\,X < \infty$, *then* $\mathbb{E}\,X = m'_X(0)$ *and*
- ▹ *if* $\operatorname{var} X < \infty$, *then* $\operatorname{var} X = m''_X(0) + m'_X(0)\big(1 - m'_X(0)\big)$.

Theorem 5.23 (Chebyshev's and Markov's[5] inequality cf. Exercise 5.40)**.** *For a random variable* $X\colon \Omega \to \mathbb{R}$ *it holds that*

(i) $P(|X| \ge \lambda) \le \frac{1}{\lambda^p}\,\mathbb{E}\,|X|^p$ *for* $p > 0$,

(ii) $P(|X - \mathbb{E}\,X| \ge x) \le \frac{M_k}{x^k}$, *where* $M_k = \mathbb{E}\,|X - \mathbb{E}\,X|^k$ *is the absolute* k*th moment;*

(iii) in particular, $P(|X - \mathbb{E}\,X| \ge k\sigma) \le \frac{1}{k^2}$, *where* $\sigma = \sqrt{\operatorname{var} X}$ *is the standard deviation of* X.

Corollary 5.24 (Chernoff[6] bounds)**.** *Let* $X \in \mathbb{R}$ *be a random variable, then*

$$P\,(X \ge a) \le \inf_{t>0} e^{-ta}\,\mathbb{E}\,e^{tX},$$
$$P\,(X \le a) \le \inf_{t<0} e^{-ta}\,\mathbb{E}\,e^{tX}.$$

Theorem 5.25 (Jensen's inequality,[7] cf. Exercise 5.41)**.** *Let* $\varphi\colon \mathbb{R} \to \mathbb{R}$ *be convex, then*

$$\varphi\,(\mathbb{E}\,X) \le \mathbb{E}\,\varphi(X).$$

Theorem 5.26 (Hölder's[8] inequality)**.** *The norm in* L^p*-spaces is* $\|X\|_p \coloneqq (\mathbb{E}\,|X|^p)^{1/p}$. *Then*

$$\mathbb{E}\,|X \cdot Y| \le \|X\|_p \cdot \|Y\|_q$$

whenever the exponents are Hölder conjugate, i.e., $\frac{1}{p} + \frac{1}{q} = 1$.

Definition 5.27 (Cumulative distribution function, cdf)**.** The *cumulative distribution function* (cdf, cf. Figure 5.1a) or just distribution function of a $\mathbb{R}$-valued random variable X is[9]

$$F_X\,(x) \coloneqq P(X \le x). \tag{5.4}$$

The quantile function[10] is the *generalized inverse* distribution function (cdf, cf. Figure 5.1b),[11]

$$\mathsf{V@R}_\alpha(X) \coloneqq q_\alpha\,(X) \coloneqq F_X^{-1}\,(\alpha) \coloneqq \inf\,\{x\colon F_X(x) \ge \alpha\}\,.$$

Further names of the quantile function are *Value-at-Risk*, or *inverse distribution function*, $\mathsf{V@R}_\alpha(X) \coloneqq F_X^{-1}\,(\alpha)$.

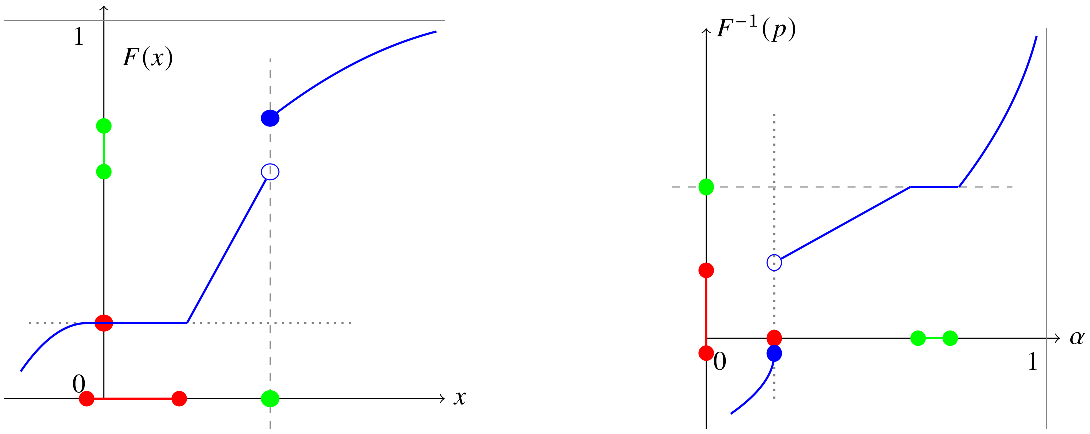
Remark 5.28. The cdf $F_X(\cdot)$ is non-decreasing and right-continuous and hence *càdlàg* (French: "continue à droite, limite à gauche").

Remark 5.29. The cdf and the inverse distribution function can be used to evaluate a expectations. Indeed (cf. also Exercise 5.21),

$$\mathbb{E}\,g(X) = \int_{-\infty}^{+\infty} g(x)dF_X(x) = \int_0^1 g(F_X^{-1}(u))du, \tag{5.5}$$

where the first integral on the real line is a Riemann–Stieltjes integral.

[5] Andrey Markov, 1856–1922
[6] Herman Chernoff, 1923, student of Abraham Wald
[7] Johan Jensen, 1859–1903
[8] Otto Hölder, 1859–1937
[9] Note that $F_X(\cdot)$ is càdlàg, i.e., continue à droite, limite à gauche: right continuous with left limits
[10] This is sometimes called *right quantile*, and $\sup\,\{q\colon P\,(X \le q) \le \alpha\}$ the *left quantile*.
[11] Note, that $\inf \emptyset = +\infty$.



(a) cdf

(b) Quantile function, the generalized inverse of Figure 5.1a

Figure 5.1: Cumulative distribution, and the corresponding quantile function

5.6 DISCRETE PROBABILITY MEASURES

Definition 5.30. A random variable is said to be *discrete* if its range (or image) $X(\Omega)$ is finite or countably infinite.

5.6.1 Elements of discrete probability measures

Definition 5.31. The *Dirac*[12] *measure* is

$$\delta_\omega(E) \coloneqq \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{else.} \end{cases}$$

Note, that the measure is well-defined on Ω's entire power set $\mathcal{P}(\Omega)$, δ_ω further is a *probability* measure taking only the values 0 or 1. As well it holds that $\mathbb{1}_E(\omega) = \delta_\omega(E)$ where $\mathbb{1}_E$ is the indicator function of the event E.

Definition 5.32. A discrete probability measure on Ω can be written in the form (cf. Exercise 5.1)

$$P(\cdot) = \sum_{\omega \in \Omega} p_\omega \cdot \delta_\omega(\cdot). \tag{5.6}$$

For a discrete measure, the function $\omega \mapsto P(\{\omega\})$ is called the *probability mass function*.

[12] *Paul Dirac, 1902–1984, English theoretical physicist*

5.6.2 Selected examples of discrete probability measures

Binomial distribution

A random variable X follows a Binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, denoted $X \sim B(n,p)$, if the probability mass function is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0,\ 1, \dots, n,$$

its moment generating function is $m_X(t) = (1 + p(e^t - 1))^n$.

As a rule of thumb, the Binomial distribution $B(n,p)$ is sufficiently well approximated by a normal distribution $\mathcal{N}\big(np,\ np(1-p)\big)$ if $\sigma^2 = np(1-p) > 9$; cf. Figure 5.2a.

Bernoulli distribution

The distribution

$$B(1,p)$$

is a special case of the binomial distribution ($n = 1$) and called Bernoulli.

Poisson distribution

The probability mass function of the Poisson distribution $P(\lambda)$ is (cf. Exercise 5.44)

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0,\ 1, \dots, .$$

Observe in Figure 5.2b that $X \sim \mathcal{N}(\lambda, \lambda)$.

5.7 CONTINUOUS PROBABILITY MEASURES

Definition 5.33. A random variable is *continuous* if its range $X(\Omega)$ is *un*countably infinite. (Cf. Definition 5.30).

5.7.1 Elements of continuous probability measures

Definition 5.34. The measure Q is said to be absolutely continuous with respect to the measure P if

$$P(A) = 0 \implies Q(A) = 0.$$

Theorem 5.35 (Radon–Nikodým[13])**.** *The probability measure Q is* absolutely continuous *with respect to the measure P, if and only if there is a density $f\colon \Omega \to [0,\infty)$ such that*

$$P(A) = \int_A f\,dQ = \int_A f(\omega) Q(d\omega).$$

[13] Johann Radon (1887–1956) proved the special case $\mathbb{R}^n$ in 1913 and Otton Marcin Nikodým (1887–1974) the general case in 1930.

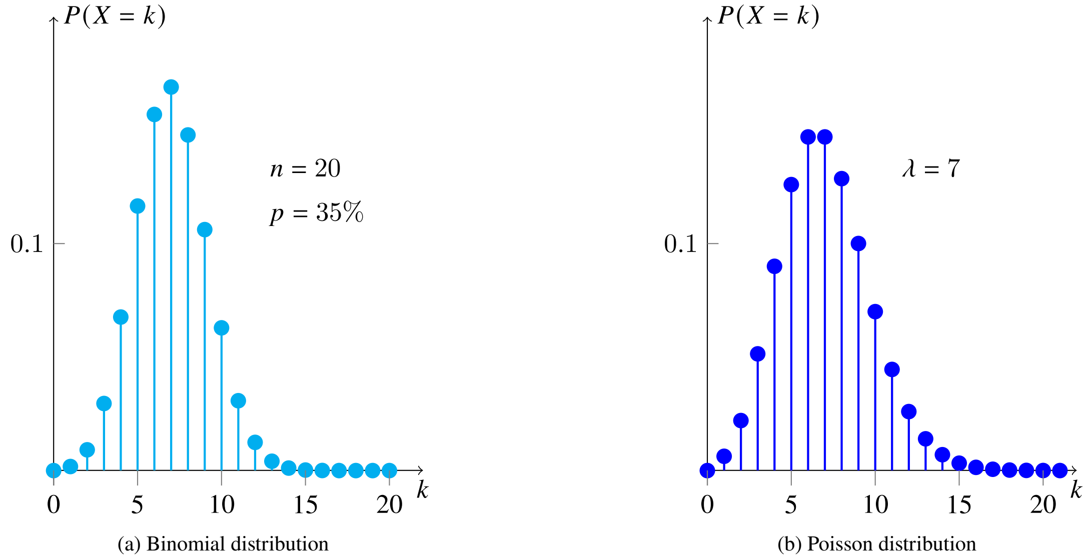


Figure 5.2: Probability mass function

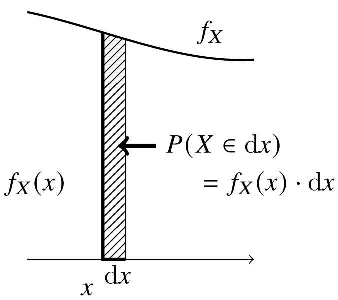


Figure 5.3: Density, cf. Remark 5.38

Definition 5.36. The probability density function (pdf) satisfies

$$P(X \in A) = \int_A f_X(x)dx, \tag{5.7}$$

where dx denotes the usual Lebesgue measure. Note, that $f_X(x) = F_X'(x)$.

Remark 5.37. A random variable X is *absolutely continuous* (with respect to the Lebesgue measure) if it has a density. Note the difference to discrete (Definition 5.30) and continuous (Definition 5.33) random variables.

Remark 5.38. The identity (5.7) justifies the shorthand notation $P(X \in [x, x + dx]) = f_X(x) \cdot dx$ for the density; or even shorter, $P(X \in dx) = f_X(x) \cdot dx$ (cf. Figure 5.3), where dx is an interval with small/ infinitesimal width containing x.

5.8 TRANSFORMATION OF RANDOM VARIABLES

For an invertible, and nondecreasing function g the transform $Z := g(X)$ has distribution function

$$F_{g(X)}(x) = P\big(g(X) \le x\big) = P\big(X \le g^{-1}(x)\big) = F_X\big(g^{-1}(x)\big). \tag{5.8}$$

The density function of $g(X)$ is found by differentiating,

$$f_{g(X)}(x) = f_X\big(g^{-1}(x)\big) \cdot \big(g^{-1}\big)'(x) = \frac{f_X\big(g^{-1}(x)\big)}{g'\big(g^{-1}(x)\big)}.$$

The result for vector valued X follows from the change of variables formula.

Proposition 5.39 (Transformation of densities)**.** *Let g be invertible, then*

$$f_{g(X)}(x) = \frac{f_X\big(g^{-1}(x)\big)}{\big|\det g'\big(g^{-1}(x)\big)\big|} = f_X\big(g^{-1}(x)\big) \cdot \underbrace{\big|\det\big(g^{-1}\big)'(x)\big|}_{Jacobian}. \tag{5.9}$$

Proof. It holds that

$$\begin{aligned}\int_A f_{g(X)}(z)dz &= P\big(g(X) \in A\big) = P\big(X \in g^{-1}(A)\big) \\ &= \int_{g^{-1}(A)} f_X(x)dx = \int_A f_X\big(g^{-1}(x)\big)\big(g^{-1}\big)'(x)dx = \int_A \frac{f_X\big(g^{-1}(x)\big)}{\big|\det g'\big(g^{-1}(x)\big)\big|}dx.\end{aligned}$$

The results follows by comparing the integrands. □

Proof. For another proof, let $h(\cdot)$ be any test function. Observe that

$$\mathbb{E}\, h\big(g(X)\big) = \int h\big(g(x)\big) f_X(x)dx = \int h(y)\frac{f_X\big(g^{-1}(y)\big)}{\big|\det g'\big(g^{-1}(y)\big)\big|}dy$$

and

$$\mathbb{E}\, h\big(g(X)\big) = \int h(y) f_{g(X)}(y)dy.$$

The result follows now by comparing the integrands. Of course, one may choose $h = \mathbb{1}_A$ to get the preceding proof. □

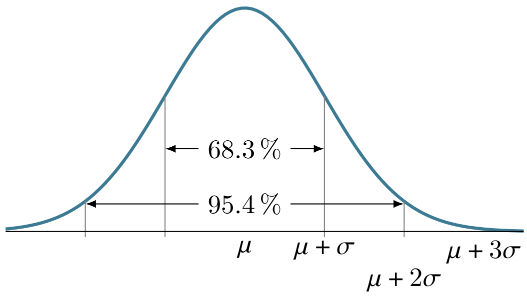


Figure 5.4: Normal distribution: the 68 − 95 − 99.7 rule

5.9 EXAMPLES OF CONTINUOUS PROBABILITY DISTRIBUTIONS

Many probability distributions on $\mathbb{R}^n$ are given in terms of densities of the Lebesgue measure.

Uniform distribution

U follows a continuous uniform distribution, if $P(U \in [c,d]) = \frac{d-c}{b-a}$. The density is $f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$ and the mean is $\mathbb{E}\,U = \frac{1}{2}(a+b)$ and variance $\operatorname{var} U = \frac{1}{12}(b-a)^2$. More generally, the uniform distribution on a compact set $K \subset \mathbb{R}^d$ is the measure $P(U \in A) = \frac{\lambda(A)}{\lambda(K)}$, where λ is the Lebesgue measure.

Normal distribution

The density of the normal distribution in the univariate case, the pdf is often denoted as

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \tag{5.10}$$

the cdf as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du. \tag{5.11}$$

For the univariate normal distribution the moment generation function is (replace $x \leftarrow \mu + \sigma x$)

$$\begin{aligned} m_X(t) \coloneqq \mathbb{E}\,e^{tX} &= \int_{-\infty}^{\infty} e^{tx}\varphi(x)dx = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \int_{-\infty}^{\infty} e^{t(\mu+\sigma x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= e^{t\mu + \frac{t^2}{2}\sigma^2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t\sigma)^2} dx = e^{t\mu+\frac{t^2}{2}\sigma^2}. \end{aligned} \tag{5.12}$$

The moments of the univariate normal distribution $X \sim \mathcal{N}(0,\sigma^2)$ are

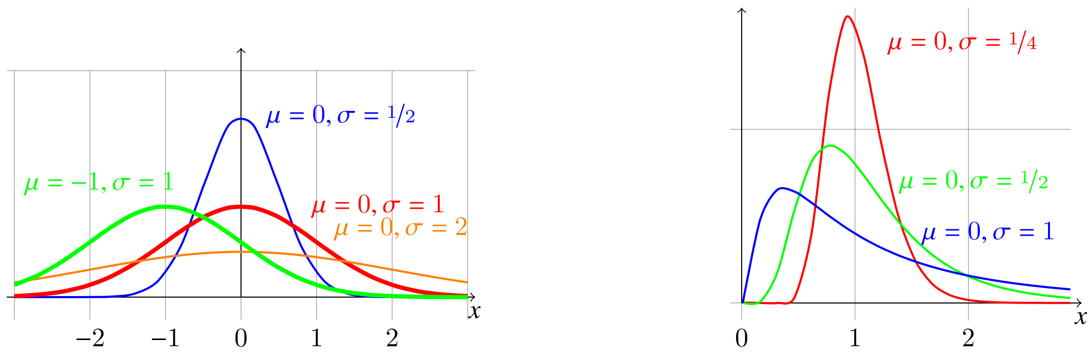
$$m_k(X) = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \frac{k}{2}!\,2^{k/2}\sigma^k & \text{if } k \text{ is even.} \end{cases}$$

Note in particular that

$$m_2(X) = \sigma^2,\ m_4(X) = 3\sigma^4,\ m_6(X) = 15\sigma^6 \text{ and } m_8(X) = 105\sigma^8; \tag{5.13}$$

note as well that $\operatorname{var}(X^2) = 2\sigma^4$.

The cumulants of the normal distribution are $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$, all other cumulants are $\kappa_k = 0$, $k = 3, 4, \dots$.



(a) Normal distribution $\mathcal{N}(\mu, \sigma^2)$

(b) Log-normal distribution

Figure 5.5: Densities of the normal and log-normal distribution

Lemma 5.40 (Stein's lemma[14])**.** *The random variable X is standard normally distributed, iff*

$$\mathbb{E}\left(f'(X) - X \cdot f(X)\right) = 0$$

for every bounded $f \in C^1$. For a general normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ it holds that

$$\sigma^2 \cdot \mathbb{E}\, f'(X) = \mathbb{E}(X - \mu) f(X).$$

Exponential, Erlang and Gamma distribution

The exponential distribution with density $f(x) = \lambda e^{-\lambda x}$ has mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$. The sum of independent Exponential variables follows an Erlang $E(n, \lambda) \sim X_1 + \dots + X_n =: S_n$ distribution with pdf and cdf are (cf. Figure 5.6a)

$$f_n(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, \quad F_n(x) = \frac{\gamma(n, \lambda x)}{\Gamma(n)} \qquad (x \geq 0). \tag{5.14}$$

Its mean is $\mathbb{E}\, S_n = \frac{n}{\lambda}$, the variance is $\operatorname{var} S_n = \frac{n}{\lambda^2}$ and $-\frac{1}{\lambda}\ln(U_1 \cdot \dots \cdot U_n)$ is $E(n, \lambda)$ distributed, whenever U_i are independent and $[0, 1]$ uniform.

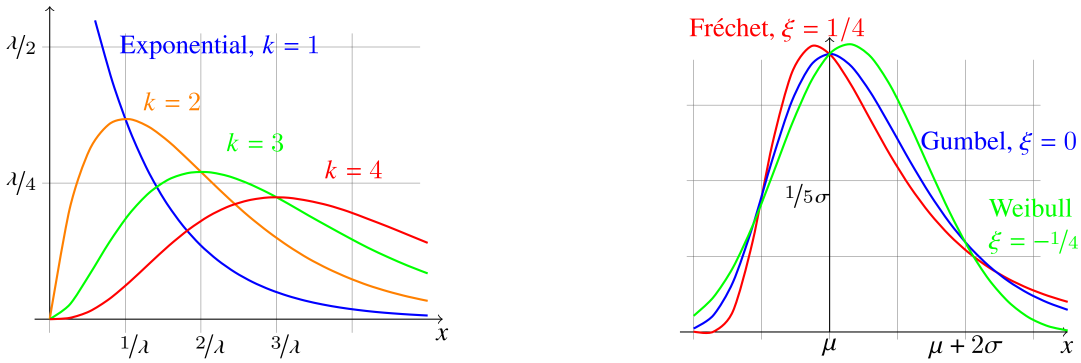
Lemma 5.41. *Let $X \sim E(m, \lambda)$ and $Y \sim E(n, \lambda)$ be independent. Then $X + Y \sim E(m + n, \lambda)$.*

Proof. The convolution is

$$\begin{aligned} f_{X+Y}(x) &= \int_0^x f_m(y) f_n(x-y) dy = \int_0^x \frac{\lambda^m y^{m-1}}{(m-1)!} e^{-\lambda y} \frac{\lambda^n (x-y)^{n-1}}{(n-1)!} e^{-\lambda(x-y)} dy \\ &\underset{y \leftarrow xu}{=} \frac{\lambda^{m+n} x^{m+n-1}}{(m+n-1)!} e^{-\lambda x} \cdot \frac{(m+n-1)!}{(m-1)!(n-1)!} \int_0^1 u^{m-1}(1-u)^{n-1} du. \end{aligned}$$

□

[14] Charles Stein, 1920–2016



(a) Erlang distribution $E_k(\lambda)$ ($\Gamma\left(k, \frac{1}{\lambda}\right)$ distribution)

(b) Generalized extreme value distributions

Figure 5.6: Densities of extreme value distributions

More generally, a distribution with density (5.14) is said to follow a gamma distribution with shape $k > 0$ and rate $\lambda > 0$. The moment generating function of an $E(n, \lambda)$ variable is (cf. Exercise 5.33)

$$m(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \tag{5.15}$$

and notably only finite for $t < \lambda$.

Lemma 5.42. *The sum of independent Gamma variables is Gamma again,*

$$\Gamma\left(\frac{\gamma_1 t}{\lambda}, \frac{\gamma_1 t}{\lambda^2}\right) + \Gamma\left(\frac{\gamma_2 t}{\lambda}, \frac{\gamma_2 t}{\lambda^2}\right) \sim \Gamma\left(\frac{(\gamma_1 + \gamma_2)t}{\lambda}, \frac{(\gamma_1 + \gamma_2)t}{\lambda^2}\right).$$

χ^2-*distribution*

Suppose that $X_i \sim \mathcal{N}(0, 1)$ are independent standard Gaussians, then

$$\sum_{i=1}^{k} X_i^2 \sim \chi_k^2 \tag{5.16}$$

has χ^2 distribution with k degrees of freedom. Its density is

$$f_{\chi_k^2}(x) = \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

(Exercise 5.33 relates χ^2 and Erlang distributions).

Remark 5.43. Occasionally, the χ distribution is $\sqrt{\chi^2}$. Its density, by (5.9), is

$$f_{\chi_k}(x) = \frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{k-1} e^{-\frac{x^2}{2}}.$$

Logistic distribution

The logistic distribution follows the density $f(x;\mu,s)=\frac{1}{s}\frac{1}{e^{\frac{x-\mu}{s}}+2+e^{-\frac{x-\mu}{s}}}$.

5.9.1 Heavy-tailed distributions

Log-normal distribution

The log-normal random variable $X=\exp(Y)$ with $Y\sim\mathcal{N}\left(\mu,\sigma^2\right)$ has pdf

$$\frac{1}{x\cdot\sigma}\varphi\left(\frac{\ln x-\mu}{\sigma}\right)=\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2} \tag{5.17}$$

and cdf $\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$ (cf. the transformation of random variables in Lecture 5.8). All moments of the log-normal distribution exist,

$$\mathbb{E}\,X^n=e^{n\mu+\frac{1}{2}n^2\sigma^2} \tag{5.18}$$

(cf. (5.12) and Exercise 5.47), but they grow very fast and the moment generating function does not exist. It follows that $\mathbb{E}\,X=e^{\mu+\frac{1}{2}\sigma^2}$ and the variance is $\operatorname{var}X=\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$ (cf. Exercise 5.50). An important example is the geometric Brownian motion, i.e., the stochastic process $S_t=S_0\exp\bigl((\mu-\sigma^2/2)t+\sigma B_t\bigr)$, where B_t is a Brownian motion.

Cauchy distribution

The Cauchy distribution (aka Lorentz distribution in physics) follows the density

$$f(x)=\frac{\gamma}{\pi}\frac{1}{\gamma^2+(x-\delta)^2}. \tag{5.19}$$

It does not have finite moments and $\mathbb{E}\,|X|=\infty$. The characteristic function is $\varphi(t)=\exp(i\mu-\gamma|t|)$. Note, that the law of large numbers (LLN, Theorem 7.7 below) does *not* hold for the Cauchy distribution.

Proposition 5.44. *If $X_1\sim C(x_1,\gamma_1)$ and $X_2\sim C(x_2,\gamma_2)$, then $X_1+X_2\sim C(x_1+x_2,\gamma_1+\gamma_2)$. Moreover, for $X_i\sim C(x_i,\gamma_i)$ it holds that $\frac{1}{n}\sum_{i=1}^n X_i\sim C\left(\frac{1}{n}\sum_{i=1}^n x_i,\frac{1}{n}\sum_{i=1}^n\gamma_i\right)$. In particular, for $X_i\sim C(x,\gamma)$, then $\frac{1}{n}\sum_{i=1}^n X_i\sim C(x,\gamma)$.*

Lévy distribution

The Lévy distribution follows the density

$$f(x)=\frac{1}{\sqrt{2\pi x^3}}e^{-\frac{1}{2x}}$$

(if $X\sim\mathcal{N}(0,1)$, then $\frac{1}{X^2}$ is Lévy distributed).

The standard Lévy distribution satisfies the condition

$$X_1+X_2+\dots+X_n\sim n^2X=n^{1/\alpha}X$$

(i.e., α-stable with $\alpha=1/2$), where X_i and X are standard Lévy distributed.

Inverse Gaussian distribution

The cdf of the inverse Gaussian distribution $IG(\lambda,\mu)$ is

$$f(x)=\sqrt{\frac{\lambda}{2\pi x^3}}\,e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}.$$

If $X_i\sim IG(\mu_i,2\mu_i^2)$, then $\sum_{i=1}^n X_i=IG\left(\sum_{i=1}^n\mu,2\left(\sum_{i=1}^n\mu_i\right)^2\right)$.

Student's t-distribution

Let $X_1\,X_2,\dots$ be iid. with $X_i\sim\mathcal{N}(\mu,\sigma^2)$. Let $\overline{X}_n:=\frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean and $S^2:=\frac{1}{n-1}\sum_{i=1}^n\left(X_i-\overline{X}_n\right)^2$ be the (Bessel corrected) sample variance. Then

$$\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}\sim\mathcal{N}(0,1)\quad\text{and}\quad\sqrt{n}\frac{\overline{X}_n-\mu}{S}\sim t_{n-1}. \tag{5.20}$$

The density of Student's[15] t-distribution with ν degrees of freedom is

$$f_{t_\nu}(x)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\cdot\frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

(cf. Exercise 5.32 for a relation between Student's and Cauchy's distribution; $\Gamma(s):=\int_0^\infty x^{s-1}e^{-x}dx$ is Euler's Gamma function).

5.9.2 Extreme value distributions

The cdf of the generalized extreme value (or Fisher–Tippett) distribution is the parametric family $F(x,\xi)=\exp\left(-(1+\xi x)^{-1/\xi}\right)$, $\xi\in\mathbb{R}$, on the domain $\{x:1+\xi x>0\}$ (cf. Figure 5.6b). The following distributions are variates of the generalized extreme value distribution.

5.9.2.1 *Fréchet distribution (type II extreme value distribution,* $\alpha={}^1/_\xi>0$*)*

has the cdf $\Phi_\alpha(x)=\exp(-x^{-\alpha})$, $x\ge 0$.

5.9.2.2 *Gumbel distribution (type I extreme value distribution,* $\xi=0$*)*

has the cdf $\Lambda(x)=\exp(-e^{-x})$, $x\in\mathbb{R}$. The distribution function follows by letting $\xi\to 0$ in the Fréchet or Weibull distribution. The expectation is $\mathbb{E}\,X=\gamma$, where $\gamma=0.577\,215\,66\dots$ is the Euler-Mascheroni constant.

5.9.2.3 *Weibull distribution (type III extreme value distribution,* $\xi={}^{-1}/_\alpha<0$*)*

has the cdf $\Psi_\alpha(x)=\exp(-(-x)^\alpha)$, $x\le 0$.

Proposition 5.45 (cf. Embrechts et al. [6, p. 123])**.** *The extreme value distributions are related as follows:*

$$X\textit{ has cdf }\Phi_\alpha\iff\ln X^\alpha\textit{ has cdf }\Lambda\iff -X^{-1}\textit{ has cdf }\Psi_\alpha.$$

[15] Introduced by William Sealy Gosset (aka "Student"), an employee of the Brewery *Guinness* in Dublin.

Proposition 5.46 (The extreme value distribution is closed under maximization)**.** *Let $(\varepsilon_i)_{i=1}^n$ be independent random variables which are Gumbel distributed with mean μ_i and common scale parameter $b > 0$. Then the maximum $\varepsilon \coloneqq \max\{\varepsilon_i + c_i : i = 1, \dots n\}$ of the shifted variables is again Gumbel distributed with mean*

$$\mathbb{E}(\varepsilon) = \mu \coloneqq b \cdot \log\left(\sum_{i=1}^n \exp\left(\frac{\mu_i + c_i}{b}\right)\right) \tag{5.21}$$

and the same scale parameter b, where $c_i \in \mathbb{R}$ are arbitrary constants.

Proof. From the cumulative distribution function of the Gumbel distributions with respective means it follows that

$$\begin{aligned}
P\left(\max_{i\in\{1,\dots n\}} \varepsilon_i + c_i \le z\right) &= P\Big(\varepsilon_1 + c_1 \le z,\ \varepsilon_2 + c_2 \le z, \dots, \varepsilon_n + c_n \le z\Big)\\
&= \prod_{i=1}^n P(\varepsilon_i \le z - c_i) = \prod_{i=1}^n \exp\left(-e^{-\frac{z-c_i-\mu_i}{b}-\gamma}\right)\\
&= \exp\left(-\sum_{i=1}^n e^{-\frac{z-c_i-\mu_i}{b}-\gamma}\right) = \exp\left(-e^{-\frac{z}{b}-\gamma}\cdot \sum_{i=1}^n e^{\frac{\mu_i+c_i}{b}}\right)\\
&= \exp\left(-e^{-\frac{z}{b}-\gamma}\cdot e^{\frac{\mu}{b}}\right) = \exp\left(-e^{-\frac{z-\mu}{b}-\gamma}\right),
\end{aligned}$$

because $\sum_{i=1}^n e^{\frac{\mu_i+c_i}{b}} = e^{\frac{\mu}{b}}$. This reveals the assertion. □

The following proposition addresses the probability of choice. Again, an explicit formula is available for shifted Gumbel variables.

Proposition 5.47 (Choice probabilities for shifted Gumbel variables)**.** *Let $(\varepsilon_i)_{i=1}^n$ be independent Gumbel distributed random variables with individual mean μ_i and common scale parameter $b > 0$. Then the probability of choice for the variables shifted by c_i is*

$$P\left(\varepsilon_1 + c_1 = \max_{i\in\{1,2,\dots n\}} \varepsilon_i + c_i\right) = \frac{\exp\left(\frac{c_1+\mu_1}{b}\right)}{\exp\left(\frac{c_1+\mu_1}{b}\right) + \dots + \exp\left(\frac{c_n+\mu_n}{b}\right)}.$$

Proof. Without loss of generality one may consider a pair $(\varepsilon_1, \varepsilon_2)$ of independent Gumbel variables with location parameter 0, because the maximum in (5.21) itself is Gumbel distributed by Proposition 5.46.

Thus

$$\begin{aligned}
P(\varepsilon_1 + c_1 \ge \varepsilon_2 + c_2) &= P(\varepsilon_2 \le \varepsilon_1 + c_1 - c_2)\\
&= \int_{-\infty}^{\infty} f(x_1) \int_{-\infty}^{x_1+c_1-c_2} f(x_2)\, dx_2 dx_1\\
&= \int_{-\infty}^{\infty} f(x_1) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) dx_1,
\end{aligned} \tag{5.22}$$

where the cdf of the Gumbel distribution has been substituted. By substituting the probability density function (pdf) f, Eq. (5.22) continues as

$$\begin{aligned}
P\left(\varepsilon_1 + c_1 \geq \varepsilon_2 + c_2\right) &= \int_{-\infty}^{\infty} \frac{1}{b} \exp\left(-\frac{x_1}{b} - e^{-\frac{x_1}{b}}\right) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{b} e^{-\frac{x_1}{b}} \exp\left(-e^{-\frac{x_1}{b}} \left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right) dx_1 \\
&= \left[\frac{\exp\left(-e^{-\frac{x_1}{b}} \left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right)}{1 + e^{-\frac{c_1-c_2}{b}}}\right]_{x_1=-\infty}^{\infty} \\
&= \frac{1}{1 + e^{-\frac{c_1-c_2}{b}}} = \frac{e^{\frac{c_1}{b}}}{e^{\frac{c_1}{b}} + e^{\frac{c_2}{b}}}. \qquad (5.23)
\end{aligned}$$

This completes the proof. □

Finally we give a proof that the difference of Gumbel variables enjoys a logistic distribution (cf. Nadarajah [14]).

Corollary 5.48. *If ε_1 and ε_2 are Gumbel distributed with mean μ_1 and μ_2 and common scale parameter $b > 0$. Then the difference $\varepsilon \coloneqq \varepsilon_2 - \varepsilon_1$ follows a logistic distribution with mean $\mu = \mu_2 - \mu_1$ and cumulative distribution function*

$$F_\varepsilon(z) = \frac{1}{1 + \exp\left(-\frac{z-\mu}{b}\right)},$$

which is the distribution function of a logistic variable.

Proof. If follows from (5.23) in the proof of the preceding theorem that

$$F_\varepsilon(z) = P(\varepsilon_2 - \varepsilon_1 \leq z) = P(\varepsilon_1 + z \geq \varepsilon_2) = \frac{1}{1 + e^{-\frac{z-(\mu_2-\mu_1)}{b}}},$$

which completes the proof. □

5.10 PROBLEMS

Exercise 5.1. *Show, that there are at most countably many $\omega \in \Omega$ for which $p_\omega > 0$ in (5.6).*

Exercise 5.2. *Let $\mathcal{F}$ and $\mathcal{G}$ be sigma algebras. Prove that $\mathcal{F} \cap \mathcal{G}$ is a sigma algebra. Is $\mathcal{F} \cup \mathcal{G}$ necessarily a sigma algebra?*

Exercise 5.3. *Show that $\mathcal{F} \coloneqq \bigcap_{i\in I} \mathcal{F}_i$ is a sigma algebra, where $\mathcal{F}_i$ is a sigma algebra for every $i \in I$.*

Exercise 5.4. *Let $E_1, E_2, \dots \in \mathcal{F}$ be events. Show that there is a sequence of sets $E_1', E_2', \dots \in \mathcal{F}$ which are pairwise disjoint (i.e., $E_m' \cap E_n' = \emptyset$ whenever $m \neq n$) so that $\bigcup_{n=1}^\infty E_n = \bigcup_{n=1}^\infty E_n'$.*

Exercise 5.5. *What is the smallest sigma algebra containing $E_1, E_2, \dots E_n \subset \Omega$?*

Exercise 5.6. *On algebra $\mathcal{F}$ define and discuss the operations $E + F \coloneqq E \Delta F \coloneqq (E \cup F) \backslash (E \cap F)$ and $E \cdot F \coloneqq E \cap F$.*

Exercise 5.7. *Show that $\bigcap_{n=1}^\infty E_n \in \mathcal{F}$ whenever $E_n \in \mathcal{F}$ for all $n \in \mathbb{N}$.*

Exercise 5.8. *Show that $P(\Omega) = 1$.*

Exercise 5.9. *What of the following is a sigma algebra?*

(i) $\mathcal{F} = \{E \subseteq \mathbb{R} : 0 \in E\}$,

(ii) $\mathcal{F} = \{E \subseteq \mathbb{R} : E \text{ is finite}\}$,

(iii) $\mathcal{F} = \{E \subseteq \mathbb{R} : E \text{ or } E^{\mathsf{c}} \text{ is finite}\}$,

(iv) $\mathcal{F} = \{E \subseteq \mathbb{R} : E \text{ is open}\}$,

(v) $\mathcal{F} = \{E \subseteq \mathbb{R} : E \text{ is open or } E \text{ is closed}\}$,

Exercise 5.10. *The preimage X^{-1} preserves all set operations, as $X^{-1}\left(\bigcup_{\alpha\in A} E_\alpha\right) = \bigcup_{\alpha\in A} X^{-1}(E_\alpha)$, $X^{-1}(E^{\mathsf{c}}) = \left(X^{-1}(E)\right)^{\mathsf{c}}$, etc.*

Exercise 5.11. *Describe and plot the cdf and generalized inverse of a Bernoulli distribution with $P(X = 0) = 1 - p$ and $P(X = 1) = p$.*

Exercise 5.12 (A degenerate situation)**.** *Describe and plot the cdf and generalized inverse of a random variable X with $P(X = 7) = 1$. Discuss this situation in detail.*

Exercise 5.13. *Show that for a binomially distributed random variable $X \sim B(n, p)$ it holds that $\mathbb{E}\, X = np$ and $\operatorname{var} X = np(1 - p)$.*

Exercise 5.14. *Give the Value-at-Risk for $\alpha = 10\,\%$ for the following distribution X:*

x_i	−10.1	−7.2	−3.3	−1.8	2.8	3.1	3.2	3.7	4.1	5.1	8.2
$\mathbb{P}\,(X = x_i)$	0.02	0.03	0.10	0.12	0.20	.11	.07	.03	.08	.14	.10

Also, compute X's expected value and variance.

Exercise 5.15. *Show that* $\mathsf{V@R}_\alpha(X) = \inf\{x : F_X(x) \ge \alpha\} = \sup\{x : F_X(x) < \alpha\}$ *and* $\inf\{x : F_X(x) > \alpha\} = \sup\{x : F(x) \le \alpha\}$.

Exercise 5.16. *Show that* $\mathsf{V@R}_\alpha(c + \lambda \cdot X) = c + \lambda \cdot \mathsf{V@R}_\alpha(X)$ *whenever* $\lambda \ge 0$.

Exercise 5.17 (Cf. van der Vaart [24, Lemma 21.1])**.** *For every $0 < \alpha < 1$ and $x \in \mathbb{R}$,*

(i) *$F_X(\cdot)$ is continuous from the right, the quantile function F_X^{-1} is nondecreasing and continuous from the left.*

(ii) *$F^{-1}(\alpha) \le x$ if and only if $\alpha \le F(x)$;*

(iii) *$F\big(F^{-1}(\alpha)\big) \ge \alpha$ for $0 < \alpha < 1$ with equality, iff $\alpha \in \{F(x) : x \in \mathbb{R}\}$, i.e., α is in the range of F;*

(iv) *$F^{-1}\big(F(x)\big) \le x$ for all $x \in \mathbb{R}$; equality fails iff x is in the interior or at the right of a "flat" of F;*

(v) *$F\Big(F^{-1}\big(F(x)\big)\Big) = F(x)$ and $F^{-1}\Big(F\big(F^{-1}(\alpha)\big)\Big) = F^{-1}(\alpha)$;*

(vi) *$P\left(X = F_X^{-1}\left(F_X\left(X\right)\right)\right) = 1$; in words: the* generalized inverse *F_X^{-1} is indeed the inverse* with probability 1.

(vii) *$F \circ F^{-1} \circ F = F$, $\left(F^{-1}\right)^{-1} = F$ and $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$.*

Exercise 5.18. *With reference to the Figures 5.1a and 5.1b:*

- *Discuss, why the red realizations will not happen (i.e., will happen with probability 0);*

▹ *Discuss, why the green realization will happen with strictly positive probability.*

Exercise 5.19. *Let F be continuous, and g monotone and continuous from the left. Show that*

▹ $F^{-1}_{g(X)}(p) = g\big(F^{-1}(p)\big)$ *if g is increasing, and*

▹ $F^{-1}_{g(X)}(p) = g\big(F^{-1}(1-p)\big)$ *if g is decreasing.*

Give the explicit formula to compute $\mathbb{E}\,X$ for the Dirac measure and the discrete probability measure P.

Exercise 5.20. *Give the explicit formula to compute $\mathbb{E}\,X$ for the Dirac measure and the discrete probability measure $P(\cdot) = \sum_{i=1}^n p_i \delta_{x_i}(\cdot)$.*

Exercise 5.21. *Prove formula (5.5) to compute the expectation.*

Exercise 5.22. *Suppose that $X \in \mathbb{R}$ has a density f_X and let $c \in \mathbb{R}$ be a constant. Verify that*

(i) $f_{X+c}(x) = f_X(x-c)$,

(ii) $f_{c\cdot X}(x) = \frac{1}{c} f_X\left(x/c\right)$ *and*

(iii) $f_{c/X}(x) = \frac{c}{x^2} f_X\left(c/x\right)$.

Exercise 5.23. *Equip $[0,1]$ with the usual Lebesgue measure and consider the functions $U_1(u) \coloneqq u$, $U_2(u) \coloneqq 1-u$ and $U_3 \coloneqq \begin{cases} 2u & \text{if } u \le 1/2, \\ 2-2u & \text{if } u \ge 1/2 \end{cases}$. Show that U_1, U_2 and U_3 are all uniformly distributed random variables.*

Exercise 5.24. *Explain $\mathbb{E}\,g(X)$ for some examples, e.g., for*

$$\begin{aligned} g\colon \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto 1 \end{aligned} \qquad \begin{aligned} g\colon \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x \end{aligned} \qquad \begin{aligned} g\colon \mathbb{R} &\to \mathbb{R}. \\ x &\mapsto 1 \end{aligned}$$

How can we denote $\mathbb{E}\,g(X)$ for discrete (continuous, respectively) random variables explicitly?

Exercise 5.25 (Summation by parts, aka Abel's lemma)**.** *For discrete random variables with $P(X \ge x_0) = 1$ and possible outcomes $\{x_k\colon k = 0, 1, \dots n\}$ set $p_k \coloneqq P(X = x_k)$. Discuss, verify and explain*

$$\begin{aligned} \mathbb{E}\,g(X) = \sum_{k=0}^n g(x_k) p_k &= g(x_0) + \sum_{k=0}^{n-1} \big(g(x_{k+1}) - g(x_k)\big) \cdot \sum_{j=k+1}^n p_j \qquad (5.24) \\ &= g(x_0) + \sum_{k=0}^{n-1} \big(g(x_{k+1}) - g(x_k)\big) \cdot P(X > x_k). \end{aligned}$$

Exercise 5.26. *Verify formula (5.24) for the random variable X given by*

k	0	1	2	3	4	5
x_k	−3.1	−1.2	−0.3	−1.8	2.8	3.1
$\mathbb{P}(X = x_k)$	5 %	10 %	15 %	10 %	20 %	40 %
$g(x_k)$	4	7	6	3	5	1

Exercise 5.27 (Integration by parts, Fubini's theorem)**.** *Show that*

$$\begin{aligned} \mathbb{E}\,g(X) &= g(a) + \int_a^\infty g'(x)\big(1 - F(x)\big)dx, \quad \text{if } P(g(X) \ge g(a)) = 1, \qquad (5.25) \\ &= g(b) - \int_{-\infty}^b g'(x) F(x) dx, \qquad \text{if } P(g(X) \le g(b)) = 1. \end{aligned}$$

Exercise 5.28. *Relate and compare (5.24) and (5.25).*

Exercise 5.29. *Verify the 4th moment for the normal distribution in (5.13) and show that* $\operatorname{var} X^2 = 2\sigma^4$.

Exercise 5.30. *Show that Student's t-distribution (5.20) with* ν *degrees of freedom has a finite nth moment provided that* $n < \nu$. *For* $n > \nu$, *the nth moment is* ∞.

Exercise 5.31. *Verify the summation formula for the Gamma distribution, Lemma 5.42.*

Exercise 5.32. *Relate Student's t and Cauchy's distribution.*

Exercise 5.33. *Relate the* χ^2 *and Erlang distributions.*

Exercise 5.34. *Let* X_i *be independent copies of the random variable from the previous Exercise 5.11. Define* $Z \coloneqq \sum_{i=1}^n X_i$ *and compute* $\mathbb{E}\, Z$ *and* $\operatorname{var} Z$ *for this resulting random variable.*

Exercise 5.35. *Show that Z's distribution (the previous example) is*

$$P(Z \le k) = \sum_{i \le k} \binom{n}{i} p^i (1-p)^{n-i}$$

for $0 < p < 1$ *and* $n \in \mathbb{N}$; *explain, using an appropriate plot.*

Exercise 5.36. *Given the random variable Z from Exercise 5.34, state*

$$\mathbb{E}\, g\,(X)$$

in explicit terms.

Exercise 5.37. *Show that the density of* $1/x$ *is* $f_{1/X}(x) = x \cdot f_X(1/x)$.

Exercise 5.38. *Prove Proposition 5.20.*

Exercise 5.39. *Verify Proposition 5.21.*

Exercise 5.40. *Verify Chebyshev's inequality, Theorem 5.23.*

Exercise 5.41. *Verify Jensen's inequality, Theorem 5.25.*

Exercise 5.42. *Use Jensen's inequality (Theorem 5.25) to show that* $\|X\|_p \le \|X\|_{p'}$ *whenever* $p \le p'$.

Exercise 5.43 (Large deviation). *Use Corollary 6.15 to show that* $P\big(\overline{X}_n \ge \alpha\big) \le e^{-t\alpha}\, \mathbb{E}\, e^{t\overline{X}_n} = e^{-t\alpha}\, m_X\left(\frac{t}{n}\right)^n$ *and further*

$$\frac{1}{n} \log P\left(\overline{X}_n \ge \alpha\right) \le -\Lambda^*(\alpha),$$

where $\Lambda^*(z) \coloneqq \sup_{t\in\mathbb{R}} tz - \log m_X(t)$ *is the convex conjugate of* $\Lambda(t) \coloneqq \log m_X(t)$.

Exercise 5.44. *Show that the moment generating function of the Poisson distribution with parameter* λ *is*

$$m_X(t) = \mathbb{E}\, e^{tX} = \exp\left(\lambda(e^t - 1)\right).$$

Exercise 5.45. *Show the moment generating function for the Erlang distribution, Eq. (5.15).*

Exercise 5.46. *Verify and discuss the inverse transform method* $X \sim F_X^{-1}(U)$.

Exercise 5.47. *Compute the moments of the log-normal distribution (cf. (5.18)) and show that the moment generating function does not exist outside* $\{0\}$*, although it is monotone, finite and bounded in* $\{x \in \mathbb{R} \colon x \le 0\}$.

Exercise 5.48. *Verify Stein's lemma, Lemma 5.40.*

Exercise 5.49. *Verify the density (5.17) of the log-normal distribution.*

Exercise 5.50. *Give the variance for the log-normal distribution. What is the formula for small* σ*?*

Exercise 5.51. *Verify the moment generating function (5.12) of the normal distribution (at least in one dimension).*

Exercise 5.52 (Mills ratio,[16] cf. Williams [27, Section 14.8])**.** *Show that*

$$\frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \le \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du \le \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad x>0,$$

i.e.,

$$\frac{x}{x^2+1}\varphi(x) \le 1-\Phi(x) \le \frac{1}{x}\varphi(x), \qquad x>0$$

and thus

$$\Phi(-x) = 1-\Phi(x) \sim \frac{1}{x}\varphi(x), \qquad x \gg 1. \tag{5.26}$$

Hint: for φ *defined in (5.10) we have* $\varphi'(u) = -u\varphi(u)$ *and thus* $\varphi(x) = -\int_x^\infty \varphi'(u)du = \dots$. *For the second inequality note that* $\left(\frac{1}{u}\varphi(u)\right)' = -\left(1+{}^1\!/\!_{u^2}\right)\varphi(u)$. *A similar computation as before for* $\frac{\varphi(x)}{x} = \int_x^\infty \left(1+{}^1\!/\!_{u^2}\right)\varphi(u)du$ *leads to the result.*

Indeed, and more generally,

$$\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du \sim \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\left(\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}-\frac{15}{x^7}+\frac{105}{x^9}-\frac{905}{x^{11}}+\dots\mp\frac{(2k)!}{k!2^k}\pm\dots\right).$$

Exercise 5.53 (The probit function)**.** *Show that (cf. https://dlmf.nist.gov/7.17)*

$$\log\Phi(x) \sim -\frac{x^2}{2} - \log|x| - \frac{1}{2}\log(2\pi) - \frac{1}{x^2} + \frac{5}{2x^4} - \frac{37}{3x^6} + \dots, \qquad x \ll -1 \tag{5.27}$$

and verify that

$$\Phi^{-1}(p) \sim -\sqrt{\log\frac{1}{p^2} - \log\log\frac{1}{p^2} - \log(2\pi) + \frac{\log\log\frac{1}{p^2}}{\log\frac{1}{p^2}}}, \qquad p \sim 0.$$

Hint: substitute $p \leftarrow \Phi(x)$ *in the latter display, note that* $\log\frac{1}{p^2} = -2\log p$ *and expand* $x^2 = \Phi^{-1}(\Phi(x))^2$ *by employing (5.27).*

Exercise 5.54 (Inverse Mills ratio)**.** *Show that*

$$\mathbb{E}\left(X \mid X \ge \mu + \sigma\Phi^{-1}(\alpha)\right) = \mu + \sigma\frac{\varphi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \qquad \text{for } X \sim \mathcal{N}(\mu,\sigma^2)$$

(note that the approximation $\mathbb{E}(X \mid X \ge x) \sim \mu + \sigma(x-\mu)$ *derived from (5.26) is usually not very good and not in use).*

[16] John P. Mills

Exercise 5.55. *For $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ we have that*

$$\mathsf{AV@R}_\alpha(X) = \mu + \frac{\sigma}{1-\alpha} \cdot \alpha\sqrt{-2\log\alpha}\,(1+o(1))$$

for $\alpha \to 0$.

Exercise 5.56. *Show for $X \sim \mathcal{N}(\mu, \sigma^2)$ that $\mathbb{E}\,X_+ = \mu\Phi\left(\frac{\mu}{\sigma}\right) + \sigma\varphi\left(\frac{\mu}{\sigma}\right)$ and $\mathbb{E}\,X_- = \mu\left(1 - \Phi\left(\frac{\mu}{\sigma}\right)\right) - \sigma\varphi\left(\frac{\mu}{\sigma}\right)$.*

Exercise 5.57 (Cf. Øksendal [16, Exercise 2.1])**.** *Suppose that $X\colon \Omega \to \mathbb{R}$ assumes only countably many values $a_1, a_2, \dots \in \mathbb{R}$.*

(i) Show that X is a random variable if and only if $X^{-1}(a_k) \in \mathcal{F}$ for all $k = 1, 2, \dots$

(ii) Show that $\mathbb{E}\,f(X) = \sum_{k=1}^\infty f(a_k) P(X = a_k)$, where $f\colon \mathbb{R} \to \mathbb{R}$ is bounded and measurable.

Use a standard software as Matlab or R for the following exercises.

Exercise 5.58. *Simulate 100 realizations of an*

(i) Exponential distribution using the Inverse Transform Method ($\lambda = 1$ and $\lambda = 5$);

(ii) Compute the sample average of your sample;

(iii) Plot a historgram of your samples obtained;

(iv) Same with 10 000 samples.

Exercise 5.59. *Same as Exercise 5.58, but for a Gumbel distribution.*

Exercise 5.60. *Same as Exercise 5.58, but for a Cauchy distribution. What happens with the mean if you repeat the simulations a couple of times, particularly for large sample sizes? What goes wrong here?*

Exercise 5.61. *Same as Exercise 5.58 for a normal distribution, using the acceptance rejection method (rejection sampling).*

Exercise 5.62. *What are the difficulties to generate variates of the normal distributions using the inverse transform method?*

Exercise 5.63. *Same as Exercise 5.58 for a Binomial distribution.*

Exercise 5.64. *Same as Exercise 5.58 for a Poisson distribution.*

Exercise 5.65. *For some of the distributions mentioned above with mean μ and variance σ, plot the histogram of (100 realizations of) $Z \coloneqq \frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sigma}$.*

Exercise 5.66. *Verify the χ^2 density (5.16) by sampling and plotting the histogram.*

Exercise 5.67. *Verify Student's density (5.20) by sampling and plotting the histogram for some k of your choice.*

Exercise 5.68. *Same as the previous Exercise 5.65 for the Cauchy distribution with $\mu = 0$ and $\sigma = \gamma = 1$ in (5.19). What do you observe? What is wrong here?*

Lemma 5.49. *The* minimum *of independent exponential distributions $X_k \sim E(\lambda_k)$ is exponential again, $\min\{X_1, \dots X_n\} \sim E(\lambda_1 + \dots + \lambda_n)$, as*

$$P\left(\min(X_1, X_2) \geq x\right) = P\left(X_1 \geq x\right) \cdot P\left(X_2 \geq x\right) = e^{-(\lambda_1+\lambda_2)x}.$$

Note, that the maximum *of exponentials is* not *exponentially distributed.*

Random Vectors

6

Random vectors have already be considered in the previous lecture. Here we collect some further details on random vectors to build the idea.

Suppose that f is a measurable function $\mathbb{R}^n$ (with respect to the Lebesgue measure). Then

$$P(A) = \int_A f(t)dt = \int \cdots \int_A f_{X_1,\dots X_n}(t_1,\dots t_n)dt_1,\dots dt_n$$

is a measure on $\mathbb{R}^n$. $f_{X_1,\dots X_n}$ is the density (pdf, cf. Definition 5.36) of the random vector $(X_1,\dots X_n)$ if

$$P(X_1 \in A_1,\dots X_n \in A_n) = \int_{A\times\cdots\times A_n} f_{X_1,\dots X_n}(t)dt = \int_{A_1} \cdots \int_{A_n} f_{X_1,\dots X_n}(t_1,\dots t_n)dt_1,\dots dt_n.$$

The corresponding cdf is

$$F_{X_1,\dots X_n}(x_1,\dots x_n) \coloneqq \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1,\dots t_n)dt_1,\dots dt_n,$$

where integration is effective on rectangle

$$(-\infty, x_1] \times \cdots \times (-\infty, x_n] .$$

Note the relation

$$f_{X_1,\dots X_n}(x_1,\dots x_n) = \frac{\partial}{\partial x_1} \dots \frac{\partial}{\partial x_n} F_{X_1,\dots X_n}(x_1,\dots x_n).$$

Definition 6.1. Let $(X_1,\dots X_n)$ be a random vector. The distribution of the marginal variables (the *marginal distribution*) is obtained by marginalizing over the distribution of the variables being discarded, and the discarded variables are said to have been marginalized out.

The *marginal distribution* of X_k has cdf

$$F_{X_k}(x_k) = \lim_{x_1\to\infty, x_{k-1}\to\infty, x_{k+1}\to\infty, x_n\to\infty} F_{X_1,\dots X_n}(x_1,\dots x_{k-1}, x_k, x_{k+1},\dots x_n). \tag{6.1}$$

More generally,

$$F_{X_{k_1},\dots X_{k_j}}(x_{k_1},\dots x_{k_j}) = \lim_{x_k\to\infty \text{ for all } k\notin\{k_1,\dots k_j\}\to\infty} F(x_1,\dots x_n) \tag{6.2}$$

is the cdf of the (sub-)vector $\left(X_{k_1},\dots X_{k_j}\right)$.

Remark 6.2. The density of the marginal distribution (6.1) is

$$f_{X_k}(x_k) \coloneqq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots X_n}(t_1,\dots t_{k-1}, x_k, t_{k+1},\dots t_n)dt_1,\dots dt_{k-1}dt_{k+1}dt_n,$$

the generalization for (6.2) is apparent.

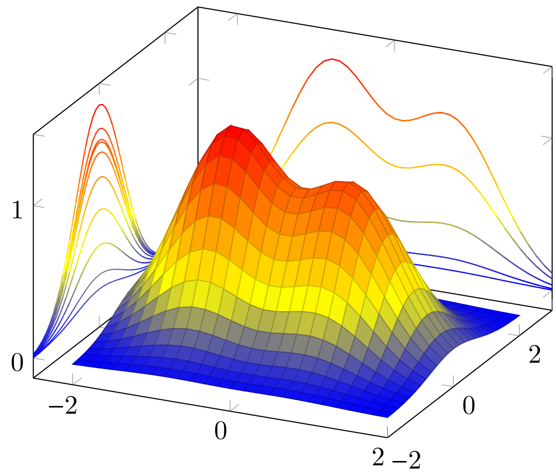


Figure 6.1: Bivariate distribution and its marginals

Example 6.3 (Linear combination of random variables)**.** Consider the transform $g\begin{pmatrix}x\\y\end{pmatrix} := \begin{pmatrix}a\,x+b\,y+c\\ \frac{y}{a}\end{pmatrix}$ with $g^{-1}\begin{pmatrix}z_1\\z_2\end{pmatrix} = \begin{pmatrix}\frac{z_1-c}{a}-b\,z_2\\ a\,z_2\end{pmatrix}$ and $g'\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}\frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y}\\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial x}\end{pmatrix} = \begin{pmatrix}a & b\\ 0 & \frac{1}{a}\end{pmatrix}$ in the setting of Lecture 5.8. Then $f_{a\,X+b\,Y+c,Y/a}(z_1,z_2) = f_{X,Y}\left(\frac{z_1-c}{a}-b\,z_2, az_2\right)$, and thus

$$f_{a\,X+b\,Y+c}(z) = \int_{-\infty}^{\infty} f_{X,Y}\left(\frac{z-c}{a}-bz_2,\; az_2\right)dz_2$$

and particularly

$$f_{X\pm Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}\left(z\mp y, y\right)dy.$$

For independent random variables we get the well-known formula for the convolution,

$$f_{X\pm Y}(z) = \int_{-\infty}^{\infty} f_X\left(z\mp y\right)\cdot f_Y(y)dy. \tag{6.3}$$

Example 6.4 (Ratio distribution)**.** As an example consider $g\begin{pmatrix}x\\y\end{pmatrix} := \begin{pmatrix}x/y\\y\end{pmatrix}$ with $g^{-1}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x\cdot y\\y\end{pmatrix}$ and $(g^{-1})'\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}y & x\\0 & 1\end{pmatrix}$. Then $f_{X/Y,Y}(z,y) = f_{X,Y}(x\cdot y, y)\cdot|y|$, and thus

$$f_{X/Y}(z) = \int_{-\infty}^{\infty} |y|\cdot f_{X,Y}(z\cdot y, y)dy \tag{6.4}$$

(cf. the t-distribution (5.20) as an example).

The product distribution is derived in the same way (note that $\det g'\begin{pmatrix}x\\y\end{pmatrix} = \det\begin{pmatrix}1/y & -x/y^2\\0 & 1\end{pmatrix} = \frac{1}{y}$), it has the density

$$f_{X\cdot Y}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|}\cdot f_{X,Y}\left(x,\frac{z}{x}\right)dx. \tag{6.5}$$

(As an example cf. the χ^2-distribution (5.16)).

6.1 COVARIANCE

Definition 6.5. The covariance of two $\mathbb{R}$-valued random variables X and Y is the number

$$\operatorname{cov}(X,Y) \coloneqq \mathbb{E}\big((X-\mathbb{E}\,X)\cdot(Y-\mathbb{E}\,Y)\big).$$

The variance (Definition 5.19) is a special case of the covariance, $\operatorname{var}(X)=\operatorname{cov}(X,X)$.

Definition 6.6 (Pearsons's ρ, Pearson[1] product-moment correlation coefficient)**.** The correlation of two random variables X and Y is

$$\rho_{X,Y} \coloneqq \operatorname{corr}(X,Y) \coloneqq \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var} X}\cdot\sqrt{\operatorname{var} Y}}.$$

Proposition 6.7 (Cauchy–Schwarz[2] inequality, cf. Exercise 6.16)**.** *It holds that* $-1 \le \operatorname{corr}(X,Y) \le 1$ *(cf. Theorem 5.26).*

Proposition 6.8 (Hoeffding's[3] covariance identity, cf. Lehmann [13] and Exercise 6.14.)**.** *It holds that*

$$\operatorname{cov}(X,Y) = \iint\limits_{\mathbb{R}\times\mathbb{R}} F_{X,Y}(x,y) - F_X(x)F_Y(y)\,dx dy,$$

where $F_{X,Y}(\cdot,\cdot)$ *is the joint distribution function of* (X,Y) *and* $F_X(\cdot)$ *and* $F_Y(\cdot)$ *its marginals.*

Proof. (cf. Lehmann [13]). Let $(\tilde X,\tilde Y)$ be an independent copy of (X,Y), i.e., $P(\tilde X \le x, \tilde Y \le y) = P(X \le x, Y \le y) = F(x,y)$. Then

$$\begin{aligned} 2\,\mathbb{E}(XY) - 2\,\mathbb{E}(X)\,\mathbb{E}(Y) &= \mathbb{E}(X-\tilde X)(Y-\tilde Y) \\ &= \mathbb{E}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(I(u,\tilde X)-I(u,X)\Big)\Big(I(v,\tilde Y)-I(v,Y)\Big)\,du dv, \end{aligned}$$

where $I(u,x) = \begin{cases} 1 & \text{if } x \le u, \\ 0 & \text{else.} \end{cases} = \mathbb{1}_{(-\infty,u]}(x)$. We can exchange the integral and the expectation, as all are assumed to be finite. Hence

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E} I(u,\tilde X)I(v,\tilde Y) - I(u,\tilde X)I(v,Y) - I(u,X)I(v,\tilde Y) + I(u,X)I)(v,Y)\,du dv \\ &\qquad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2F(u,v) - 2F(u)F(v)\,du dv, \end{aligned}$$

from which the assertion follows. □

Definition 6.9. The cross covariance (also dispersion matrix or simply covariance matrix) of a random vector $X \in \mathbb{R}^m$ and a random vector $Y \in \mathbb{R}^n$ is

$$\operatorname{cov}(X,Y) \coloneqq \mathbb{E}\left[(X-\mathbb{E}\,X)\cdot(Y-\mathbb{E}\,Y)^\top\right] \in \mathbb{R}^{m\times n}.$$

Note that the dispersion matrix is an $m \times n$ matrix and particularly not (necessarily) symmetric.

[1] Karl Pearson, 1857–1936
[2] Karl Hermann Amandus Schwarz, 1843–1921
[3] Wassily Hoeffding, 1914–1991

Proposition 6.10. *It holds that*

$$\operatorname{cov}(X,Y) = \mathbb{E}\left(X \cdot Y^\top\right) - (\mathbb{E}\, X) \cdot (\mathbb{E}\, Y)^\top. \tag{6.6}$$

and $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)^\top$.

Proposition 6.11. *For linear transformations (also vector-valued transformations) it holds that*

$$\operatorname{cov}(AX+\alpha, BY+\beta) = A\operatorname{cov}(X,Y)B^\top$$

and

$$\operatorname{var}(AX+BY) = A\operatorname{var}(X)A^\top + B\operatorname{var}(Y)B^\top + A\operatorname{cov}(X,Y)B^\top + B\operatorname{cov}(Y,X)A^\top;$$

for $X, Y \in \mathbb{R}$ *particularly,*

$$\operatorname{var}(\alpha X+\beta Y) = \alpha^2\operatorname{var}(X) + \beta^2\operatorname{var}(Y) + 2\alpha\beta\operatorname{cov}(X,Y). \tag{6.7}$$

Proposition 6.12. *It holds that*

$$\operatorname{cov}\left(\sum_{i=1}^n \alpha_i X_i,\ \sum_{j=1}^m \beta_j Y_j\right) = \sum_{i=1}^n\sum_{j=1}^m \alpha_i\beta_j\operatorname{cov}(X_i,Y_j).$$

6.2 INDEPENDENCE

Definition 6.13. Sub-sigma algebras $\mathcal{G}_1, \mathcal{G}_2, \dots$ of $\mathcal{F}$ are called *independent* if, whenever $G_i \in \mathcal{G}_i$ and $i_1, \dots i_n$ are distinct, then

$$P(G_{i_1}\cap\dots\cap G_{i_n}) = P(G_{i_1})\cdot\ldots\cdot P(G_{i_n}). \tag{6.8}$$

Random variables $X_1, X_2, \dots$ are called independent if the sigma algebras

$$\sigma(X_1),\ \sigma(X_2), \dots$$

are independent. Independence of random variables X and X' is occasionally denoted by

$$X \perp X'.$$

Theorem 6.14 (cf. Exercise 6.17 and Øksendal [16, Exercise 2.5])**.** *Suppose that X and Y are independent. Then* $\mathbb{E}[X\cdot Y] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$ *and particularly* $P(X\le a,\ Y\le b) = P(X\le a)\cdot P(Y\le b)$.

Corollary 6.15. *Suppose that the random variables* X_i*,* $i = 1,2,\dots$ *are independent and identically distributed. Then the moment generating (cf. (5.1)) function of the sample mean* $\overline{X}_n \coloneqq \frac{1}{n}\sum_{i=1}^n X_i$ *is*

$$m_{\overline{X}_n}(t) = m_X\left(\frac{t}{n}\right)^n.$$

6.3 CONVOLUTION

Proposition 6.16 (Convolution)**.** *Let* X_1 *and* X_2 *be independent with density* f_{X_1} *and* f_{X_2}*, then* X_1+X_2 *has density*

$$f_{X_1+X_2}(x) = f_{X_1} * f_{X_2}(x) = \int f_{X_1}(x-y)\, f_{X_2}(y)dy,$$

the cdf is

$$F_{X_1+X_2}(x) = \int F_{X_1}(x-y)\, dF_{X_2}(y).$$

6.4 PROBLEMS

Exercise 6.1. *Show that (6.3), (6.4) and (6.5) are densities.*

Exercise 6.2. *Verify the identity (6.6).*

Exercise 6.3. *Show that* $\operatorname{var}(\alpha X_1 + \beta X_2) = \alpha^2 \operatorname{var} X_1 + \beta^2 \operatorname{var} X_2$*, whenever* X_1 *and* X_2 *are independent.*

Exercise 6.4 (Convolution of Dirac measures)**.** *Suppose that* $X \sim \delta_a$ *and* $Y \sim \delta_a$*. Show that* $X + Y \sim \delta_{a+b}$*.*

Exercise 6.5. *Verify that distributions of X and Y in Tables 6.1a and 6.1b have the same marginals. Moreover, compute the expectation, variance and in particular the covariance of both distributions.*

$\mathbb{P}\left[\begin{matrix} X = x_i, \\ Y = y_i \end{matrix}\right]$	$x_1 = 3$	$x_2 = 4$	$x_3 = 6$
$y_1 = 2$	10 %	10 %	0 %
$y_2 = 5$	0 %	30 %	0 %
$y_3 = 7$	0 %	10 %	40 %

(a) Measure

$\mathbb{P}\left[\begin{matrix} X = x_i, \\ Y = y_i \end{matrix}\right]$	$x_1 = 3$	$x_2 = 4$	$x_3 = 6$
$y_1 = 2$	0 %	0 %	20 %
$y_2 = 5$	0 %	10 %	20 %
$y_3 = 7$	10 %	40 %	0 %

(b) Measure

Table 6.1: Probabilities with coinciding marginals

Exercise 6.6. *Give the marginal probabilities of this following bivariate probability distribution*

1%	2%	5%	2%	1%	4%	1%
3%	5%	1%	3%	7%	2%	1%
0%	0%	1%	2%	3%	20%	2%
1%	2%	9%	2%	1%	15%	4%

Give some (at least two) other bivariate probability distribution, which have the same marginal distribution.

Is there a "natural" one?

Exercise 6.7. *Give another bivariate distribution such that the covariance in Exercise 6.5 vanishes (is 0).*

Exercise 6.8. *Explain and motivate the formula for the convolution for the random variable introduced in (6.3) above.*

Exercise 6.9. *Let* U_1 *and* $U_2 \sim U[0,1]$ *be uniformly distributed and independent. Use (6.3) to compute the density of* $U_1 + U_2$*.*

Exercise 6.10. *Let* U_1 *and* $U_2 \sim U[0,1]$ *be uniformly distributed and independent. Show that*

$$P\left(U_1^2 + U_2^2 \in du\right) = \begin{cases} \frac{\pi}{4} du & \text{if } u \in [0,1], \\ \left(\frac{\pi}{4} - \arctan\sqrt{u-1}\right) du & \text{if } u \in [1,2]. \end{cases}$$

Note particularly that $P\left(U_1^2 + U_2^2 \le u \mid U_1^2 + U_2^2 \le 1\right) = u$*, i.e.,* $R^2 := U_1^2 + U_2^2$ *is uniformly distributed provided that* $R \le 1$*.*

Exercise 6.11. *Give the distribution of*

- $X + Y$,
- $X - Y$,

- $X * Y$,
- $1/X$ and
- X/Y

for X and Y as in Exercise 6.5.

Exercise 6.12 (What is fair about a fair game? Cf. Williams [27, Exercise 4.7])**.** *Let* $X_1, X_2, \dots$ *be the random variables*

$$X_n := \begin{cases} n^2 - 1 & \text{with probability } 1/n^2 \\ -1 & \text{with probability } 1 - 1/n^2. \end{cases}$$

Show that $\mathbb{E}\, X_n = 0$ *for all* $n \in \mathbb{N}$*, but if* $S_n := X_1 + X_2 + \dots X_n$ *it holds that*

$$\frac{S_n}{n} \to -1, \text{ a.s.}$$

(Hint: apply the first Borel–Cantelli lemma, Lemma 5.14).

Exercise 6.13. *Verify Proposition 6.11.*

Exercise 6.14. *Verify Hoeffdings identity, Proposition 6.8 (an optional simplification is by assuming that* X *and* Y *have a common density and* $-C \le X \le 0$ *and* $-C \le Y \le 0$ *and then employing Proposition 6.11).*

Exercise 6.15. *Convolution: verify Proposition 6.16.*

Exercise 6.16. *Show Cauchy–Schwarz's inequality (Proposition 6.7) by employing Hölder's inequality (Theorem 5.26).*

Exercise 6.17. *Verify Theorem 6.14 by assuming that X and Y are bounded.*

Exercise 6.18. *Show that* $X_1 + X_2$ *has a density, provided that the random vector* (X_1, X_2) *has a multivariate density.*

Exercise 6.19. *Is* $X_1 + X_2$ *continuous, provided that both,* X_1 *and* X_2 *are continuous? (Hint: Exercise 5.23.) Relate to the previous example.*

Exercise 6.20. *If both* X_1 *and* X_2 *have a density, does* $X_1 + X_2$ *necessarily have a density?*

Exercise 6.21. *Let the random variables* X_i*,* $i = 1, 2, \dots, n$ *have identical variance* ($\operatorname{var} X_i = \sigma^2$) *and satisfy* $\operatorname{corr}\left(X_i, X_j\right) \le \rho$*. Show that* $\rho \ge -\frac{1}{n-1}$*. (Hint:* $0 \le \operatorname{var}(X_1 + \dots + X_n)$ *and (6.7).)*

7 Convergence of Random Variables

Some parts of this section follow van der Vaart [24, Section 2].

7.1 BASIC THEORY

Definition 7.1. A sequence of random variables (X_n) converges *almost surely* to X iff

$$P\big(X = \lim_{n\to\infty} X_n\big) = 1,$$

i.e., $P\big(\{\omega\colon X(\omega) = \lim_{n\to\infty} X_n(\omega)\}\big) = 1$; we write $X_n \xrightarrow{a.s.} X$.

Definition 7.2. A sequence of random variables (X_n) converges *in probability* to the random variable X iff

$$P\big(|X_n - X| > \varepsilon\big) \xrightarrow[n\to\infty]{} 0$$

i.e., $P\big(\{|X_n(\omega) - X(\omega)| > \varepsilon\}\big) \xrightarrow[n\to\infty]{} 0$ for every $\varepsilon > 0$; we write $X_n \xrightarrow{P} X$.

Theorem 7.3. *Convergence almost surely implies convergence in probability, i.e.,* $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$.

Proof. Define $A_n \coloneqq \bigcup_{m\ge n}\{|X_m - X| > \varepsilon\}$ and observe that $A_n \supseteq A_{n+1} \supseteq \dots \supseteq A_\infty \coloneqq \bigcap_{n>0} A_n$. Choose $\omega \in B \coloneqq \{X = \lim_{n\to\infty} X_n\}$, i.e., $X_n(\omega) \to X(\omega)$ and $|X_n(\omega) - X(\omega)| < \varepsilon$ for $n > N$. Hence $\omega \notin A_n$ and for all $n > N$ and thus $\omega \notin A_\infty$, i.e., $B \subseteq A_\infty^{\mathrm{c}}$ or $B^{\mathrm{c}} \supseteq A_\infty$ and consequently $P(A_\infty) \le P(B^{\mathrm{c}}) = 0$ as we have that $X_n \xrightarrow{a.s.} X$. Note finally that $P\big(|X_n - X| > \varepsilon\big) \le P(A_n) = P(A_\infty) + \sum_{k=n}^\infty P(A_k \setminus A_{k+1})$, as these sets are pairwise disjoint. It follows that $P\big(|X_n - X| > \varepsilon\big) \le P(A_n) \to P(A_\infty) = 0$, the assertion. □

Definition 7.4. A sequence of random variables X_n converges to X in L^p (or p-th *mean*), if $\mathbb{E}|X - X_n|^p \to 0$ (i.e. $\int_\Omega |X - X_n|^p \mathrm{dP} \to 0$), as $n \to \infty$.

Theorem 7.5. *Convergence in L^p implies convergence in probability.*

Proof. It follows from Markov's inequality (Theorem 5.23) that $P(|X - X_n| > \varepsilon) \le \frac{1}{\varepsilon^p}\mathbb{E}|X - X_n|^p \to 0$ as $n \to \infty$. □

Theorem 7.6 (L^2 weak law of large numbers (LLN), cf. Durrett [4, Theorem 2.2.3.], cf. Proposition 7.16 below)**.** *Let X_i be* uncorrelated *random variables (i.e., not necessarily independent) with $\mathbb{E}X_i = \mu$ and* $\operatorname{var} X_i \le C < \infty$*. Then $\overline{X}_n \to \mu$ in L^2 (i.e., $\mathbb{E}\big(\overline{X}_n - \mu\big)^2 \to 0$) and in probability (i.e., $P\big(\big|\overline{X}_n - \mu\big| > \varepsilon\big) \to 0$.*

Proof. Note first that $\mathbb{E}\,\overline{X}_n = \mu$. Then

$$0 \le \mathbb{E}\big(\overline{X}_n - \mu\big)^2 = \operatorname{var}\overline{X}_n = \frac{1}{n^2}\left(\operatorname{var} X_1 + \dots + \operatorname{var} X_n\right) \le \frac{Cn}{n^2} \xrightarrow[n\to\infty]{} 0.$$

Further, by Chebyshev's inequality (Theorem 5.23 or Theorem 7.5 directly) it follows that

$$P\left(\big|\overline{X}_n - \mu\big| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left(\overline{X}_n - \mu\right)^2 \xrightarrow[n\to\infty]{} 0.$$

□

Remark 7.7. More specifically, if the random variables are identically distributed with $\sigma^2 = \operatorname{var} X_i$, then $P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) > 1 - \frac{\sigma^2}{n\cdot\varepsilon}$.

Definition 7.8. A sequence of $\mathbb{R}^k$-valued random variables (X_n) converges *in distribution* to the random variable X iff

$$P(X_n \le x) \xrightarrow[n\to\infty]{} P(X \le x) \tag{7.1}$$

at every point x where the cdf. $F_X \colon x \mapsto P(X \le x)$ is continuous. We (occasionally) write $X_n \xrightarrow{\mathcal{D}} X$ (some authors prefer $X_n \rightsquigarrow X$).

Alternatively, this is called *weak convergence*, *weak* convergence* or *convergence in law*.

Remark 7.9. Note, that a.s. convergence and convergence in probability only make sense if X and X_n are defined on the same probability space. This is, however, *not* the case for convergence in distribution.

Theorem 7.10. *Convergence in probability implies convergence in distribution.*

Proof. Note first that (cf. also Exercise 7.1 and $A = (A \cap B) \cup A \cap B^c$)

$$\begin{aligned}\{X \le x\} &= \{X \le x,\ X_n - X \le \varepsilon\} \cup \{X \le x,\ X_n - X > \varepsilon\}\\ &\subset \{X_n \le x + \varepsilon\} \cup \{|X - X_n| > \varepsilon\}\end{aligned}$$

and thus

$$P(X \le x) \le P(X_n \le x + \varepsilon) + P(|X - X_n| > \varepsilon). \tag{7.2}$$

It follows that

$$P(X \le x - \varepsilon) - P(|X - X_n| > \varepsilon) \le P(X_n \le x) \le P(X \le x + \varepsilon) + P(|X - X_n| > \varepsilon),$$

where the roles of X and X_n are interchanged in the second application of inequality (7.2).

By sending $n \to \infty$,

$$P(X \le x - \varepsilon) \le \liminf_{n\to\infty} P(X_n \le x) \le \limsup_{n\to\infty} P(X_n \le x) \le P(X \le x + \varepsilon).$$

As F_X is continuous at x by assumption, it follows that the limit exists and $P(X_n \le x) \to P(X \le x)$, the assertion. □

Lemma 7.11 (Portmanteau theorem)**.** *The following are equivalent:*

(i) $X_n \xrightarrow{\mathcal{D}} X$ *(see Definition 7.8);*

(ii) $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ *for all bounded and continuous functions* f*;*

(iii) $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ *for all bounded and Lipschitz continuous functions* f*;*

(iv) $\liminf_{n\to\infty} P(X_n \in G) \ge P(X \in G)$ *for every open set* G*;*

(v) $\limsup_{n\to\infty} P(X_n \in F) \le P(X \in F)$ *for every closed set* F*;*

(vi) $P(X_n \in B) \to P(X \in B)$ *for every Borel set* B *with* $P(X \in \partial B) = 0$.

Proof. (i) $\Longrightarrow$ (ii). Assume that the distribution of X is continuous and without loss of generality that $f \in [-1,1]$. Then (i) implies that $P(X_n \in I) \to P(X \in I)$ for every rectangle I. Choose a sufficiently large rectangle, compact I so that $P(X \in I) > 1-\varepsilon$. f is uniformly continuous on I and thus $I = \bigcup_{j=1}^k I_j$ such $|f(x) - f(y)| < \varepsilon$ for every $x, y \in I_j$. Define $f_\varepsilon = \sum_{j=1}^k f(x_j)\, \mathbb{1}_{I_j}$ for $x_j \in I_j$. Let n be large enough so that $P(X_n \notin I) < \varepsilon$. It follows that

$$\begin{aligned}
|\mathbb{E} f(X_n) - \mathbb{E} f_\varepsilon(X_n)| &< \varepsilon + P(X_n \notin I) < 2\varepsilon,\\
|\mathbb{E} f(X) - \mathbb{E} f_\varepsilon(X)| &< \varepsilon + P(X \notin I) < 2\varepsilon \text{ and}\\
|\mathbb{E} f_\varepsilon(X_n) - \mathbb{E} f_\varepsilon(X)| &< \sum_{j=1}^k \left|P(X_n \in I_j) - P(X \in I_j)\right| \left|f(x_j)\right| < \varepsilon
\end{aligned}$$

for n large enough. It follows that $|\mathbb{E} f(X_n) - \mathbb{E} f(X)| < 5\varepsilon$.

For a general random variable the preceding arguments still hold true provided that $P(X \in \partial I) = 0$ — a set A with $P(X \in \partial A) = 0$ is called a *continuity set*. But only countably many intervals can satisfy $P(X \in \partial I) > 0$, so that it is possible to find intervals in the general situation.

(ii) $\Longrightarrow$ (iii) is clear.

(iii) $\Longrightarrow$ (iv) for G open define $f_m(x) \coloneqq \min\left\{1, m \inf_{y \notin G} \|x-y\|\right\}$ so that $0 \le f_m \nearrow \mathbb{1}_G$. Then

$$\liminf_{n\to\infty} P(X_n \in G) \ge \liminf_{n\to\infty} \mathbb{E} f_m(X_n) = \mathbb{E} f_m(X)$$

and $\mathbb{E} f_m(X) \to P(X \in G)$ by the monotone convergence theorem.

(iv)$\Longleftrightarrow$(v) by taking complements.

(iv)+(v) $\Longrightarrow$ (vi).

$$P(X \in B^\circ) \underset{(iv)}{\le} \liminf P(X_n \in B^\circ) \le \limsup P(X_n \in \overline{B}) \underset{(v)}{\le} P(X \in \overline{B}). \tag{7.3}$$

If $P(X \in \partial B) = 0$, then the left and right side are equal and all inequalities in (7.3) are equalities, (vi) follows.

(vi) $\Longrightarrow$ (i). Every point of continuity of the function $x \mapsto P(X \le x)$ satisfies $P(X \in \partial(-\infty, x]) = 0$ (a continuity set), so that (7.1) follows, i.e., (i). □

Theorem 7.12 (Continuous mapping theorem)**.** *Let $g\colon \mathbb{R}^k \to \mathbb{R}^m$ be continuous at every point of C with $P(X \in C) = 1$. Then*

(i) If $X_n \xrightarrow{a.s.} X$, then $g(X_n) \xrightarrow{a.s.} g(X)$;

(ii) If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$;

(iii) If $X_n \xrightarrow{\mathcal{D}} X$, then $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.

Proof. As for (iii) apply (ii) of the portmanteau theorem. □

Lemma 7.13 (Slutsky[1])**.** *Let $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} c$ where $c \in \mathbb{R}$ is a constant. then*

(i) $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$,

(ii) $X_n \cdot Y_n \xrightarrow{\mathcal{D}} cX$ and

(iii) $X_n / Y_n \xrightarrow{\mathcal{D}} X/c$ provided that $c \ne 0$.

Remark 7.14. It is *essential* in Slutsky's lemma that $Y_n \xrightarrow{\mathcal{D}} c$, a constant: there are distributions with $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} Y$, but $(X_n, Y_n) \not\to (X, Y)$ in distribution (Exercise 7.8).

[1] Jewgeni Jewgenjewitsch Sluzki, 1880–1948

7.2 CONVERGENCE IN DISTRIBUTION AND THE CHARACTERISTIC FUNCTION

Cf. (5.3) for the characteristic function.

Theorem 7.15 (Lévy's[2] continuity theorem)**.** *Let X_n and X be random vectors in $\mathbb{R}^k$. Then $X_n \xrightarrow{\mathcal{D}} X$ iff $\varphi_{X_n}(t) \to \varphi_X(t)$ for every $t \in \mathbb{R}^k$. Moreover, if $\varphi_{X_n}(t) \to \varphi(t)$ at every $t \in \mathbb{R}^k$ and $\varphi(\cdot)$ is continuous at $t = 0$, then φ is the characteristic function of a random vector X and $X_n \xrightarrow{\mathcal{D}} X$.*

Proposition 7.16 (Weak law of large numbers, LLN, cf. Theorem 7.6)**.** *Let X_n be i.i.d. random variables with characteristic function φ. Then $\overline{X}_n \xrightarrow{P} \mu$ iff φ is differentiable at 0 and in this case $\mu = -i\,\varphi'(0)$.*

Proof. It holds that

$$\mathbb{E}\, e^{it\overline{X}_n} = \mathbb{E} \prod_{j=1}^n e^{i\frac{t}{n}X_j} = \prod_{j=1}^n \mathbb{E}\, e^{i\frac{t}{n}X_j} = \varphi\left(\frac{t}{n}\right)^n = \left(1 + \frac{t}{n} i\mu + o\left(1/n\right)\right)^n \xrightarrow[n\to\infty]{} e^{it\mu} \tag{7.4}$$

where $\varphi'(0) = i\mu$. But $t \mapsto e^{it\mu}$ is the characteristic function of the constant random variable, $X = \mu$. Hence, by Lévy's continuity theorem, $\overline{X}_n \xrightarrow{\mathcal{D}} \mu$. ☐

Proposition 7.17 (Central limit theorem, CLT)**.** *Let X_i be i.i.d. random variables with $\mathbb{E}\, X_i = \mu$ and $\operatorname{var} X_i = \sigma^2$. Then*

$$\sqrt{n} \cdot \left(\overline{X}_n - \mu\right) \xrightarrow{\mathcal{D}} X,$$

where $X \sim \mathcal{N}(0, \sigma^2)$.

Proof. Assume that $\mu = 0$. Similarly to (7.4) above we have that

$$\begin{aligned}\mathbb{E}\, e^{it\sqrt{n}\cdot\overline{X}_n} &= \mathbb{E} \prod_{j=1}^n e^{i\frac{t}{\sqrt{n}}\cdot X_j} = \prod_{j=1}^n \mathbb{E}\, e^{i\frac{t}{\sqrt{n}}\cdot X_j} = \varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + i\frac{t}{\sqrt{n}}\,\mathbb{E}\, Y_j - \frac{t^2}{2n}\,\mathbb{E}\, Y_j^2 + o\left(1/n\right)\right)^n \\ &= \left(1 + i\frac{t}{\sqrt{n}} \underbrace{\mathbb{E}\, Y_j}_{0} - \frac{t^2}{2n}\sigma^2 + o\left(1/n\right)\right)^n = \left(1 - \frac{\frac{t^2\sigma^2}{2}}{n} + o\left(1/n\right)\right)^n \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}t^2\sigma^2}.\end{aligned}$$

The right hand side $t \mapsto e^{-t^2\sigma^2/2}$ is the characteristic function of the normal distribution with 0 mean and variance σ, cf. (5.12). By Lévy's continuity theorem again, $\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{\mathcal{D}} X$, where $X \sim \mathcal{N}\left(0, \sigma^2\right)$ is normally distributed. ☐

Recall that $g(x + h) = g(x) + g'_x(h) + o(h)$ as $h \to 0$ for differentiable g and the linear form

$$g'_x(h) \coloneqq \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1} & \dots & \frac{\partial g_m}{\partial x_k} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_k \end{pmatrix}.$$

Theorem 7.18 (The Delta method, Δ-theorem, cf. van der Vaart [24], Shapiro et al. [21])**.** *Let $g\colon \mathbb{R}^k \to \mathbb{R}^m$ be differentiable at $\theta \in \mathbb{R}^k$. If $r_n\,(X_n - \theta) \xrightarrow{\mathcal{D}} X$ for some numbers $r_n \to \infty$, then*

$$r_n \cdot \left(g(X_n) - g(\theta)\right) \xrightarrow{\mathcal{D}} g'_\theta(X).$$

Further, $r_n \cdot \left(g(X_n) - g(\theta)\right) - g'_\theta\left(r_n \cdot (X_n - \theta)\right) \to 0$ in probability.

[2] Paul Lévy, 1886–1971

7.3 PROBLEMS

Exercise 7.1. *Verify (7.2) formally.*

Exercise 7.2. *Prove the statement in Remark 7.7.*

Exercise 7.3. *Show that if X_n has a t-distribution with n degrees of freedom, then $X_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.*

Exercise 7.4. *Approximate the χ_k^2 distribution with k degrees of freedom by a normal distribution.*

Exercise 7.5. *Find an example for which $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} Y$, but (X_n, Y_n) does not converge in distribution.*

Exercise 7.6. *Let $P(X_n = i/n) = 1/n$ for $i = 1, \dots, n$. Then $X_n \xrightarrow{\mathcal{D}} X \sim U[0,1]$ (the uniform distribution). Find a Borel set so that $P(X_n \in B) = 1$, but $P(X \in B) = 0$.*

Exercise 7.7. *Find a sequence of random variables with $X_n \xrightarrow{\mathcal{D}} 0$, but $\mathbb{E}\, X_n \to \infty$ (hint: use (ii) of the portmanteau theorem)*

Exercise 7.8. *Verify Remark 7.14 on Slutsky's lemma.*

Exercise 7.9. *Use characteristic functions for $X_n \sim B\,(n, \lambda/n)$(binomial) to show that $X_n \xrightarrow{\mathcal{D}} X$, where $X \sim P(\lambda)$ (Poisson).*

8 Conditional Expectation

8.1 INTRODUCTION

Historically, expectation was already considered by Blaise Pascal,[1] who was interested in a problem by Chevalier de Méré. Conditional expectation was considered much later, about 1780 by Pierre-Simon Laplace[2] for conditional densities. A. Kolmogorov[3] formalized the statement of conditional expectation about 1930 by employing the Theorem by Radon–Nikodym (Theorem 5.35). P. Halmos[4] (1950) and J. Doob[5] (1953) formulated the statement for conditional expectation on sigma sub-algebras on abstract spaces.

8.2 CONDITIONING ON A SIGMA ALGEBRA

Example 8.1. Consider the random variable $X\colon \Omega \coloneqq \{\omega_1, \dots, \omega_5\} \to S \coloneqq \mathbb{R}$ with

$$X(\omega) \coloneqq \begin{cases} 40 & \text{if } \omega = \omega_1, \\ 60 & \text{if } \omega = \omega_2, \\ \dots & \\ 20 & \text{if } \omega = \omega_5, \end{cases}$$

with $\mathbb{E}\, X = 36$, cf. Figure 8.1. This random variable is measurable with respect to the sigma algebra $\mathcal{F} = \mathcal{P}(\Omega)$, but *not* with respect to the sub sigma algebra

$$\mathcal{G} \coloneqq \sigma\big(\{\omega_1, \omega_2\}, \{\omega_3, \omega_4, \omega_5\}\big).$$

Intuitively, it is evident that the conditional expectation should be

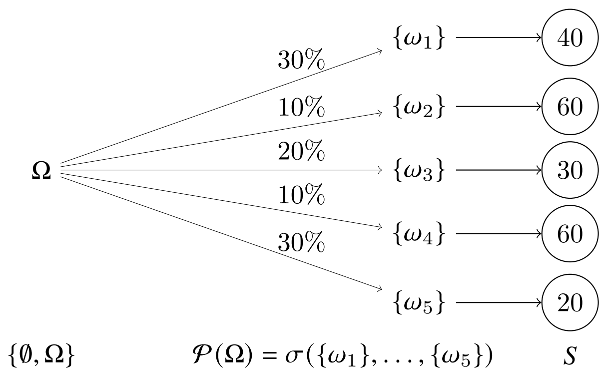
$$Y(\omega) \coloneqq \begin{cases} 45 & (= 40 * 75\% + 60 * 25\%) & \text{if } \omega \in \{\omega_1, \omega_2\}, \\ 30 & (= 30 * 2/6 + 60 * 1/6 + 20 * 3/6) & \text{if } \omega \in \{\omega_3, \omega_4, \omega_5\}. \end{cases}$$

Note, that Y is measurable with respect to the sigma algebra $\mathcal{G}$ (and $\mathcal{F}$). Further, the random variable Y enjoys the property (cf. Exercise **??**)

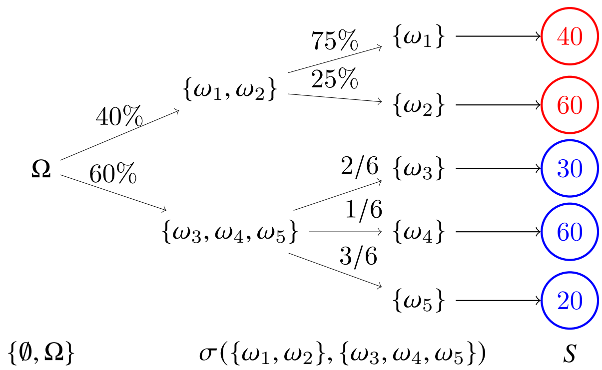
$$\int_G X\, dP = \int_G Y\, dP \quad \text{for every set } G \in \mathcal{G}. \tag{8.1}$$

This motivates the following definition, implicitly contained in the theorem.

[1] Blaise Pascal, 1623–1662
[2] Pierre–Simon Marquis de Laplace, 1749–1827
[3] Andrei N. Kolmogorov, 1903–1987
[4] Paul Halmos, 1916–2006
[5] Joseph L. Doob, 1910–2004

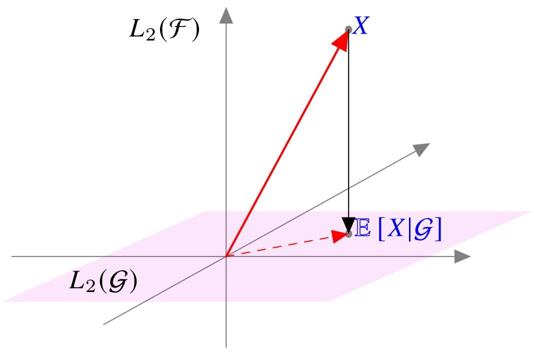


(a) Expectation

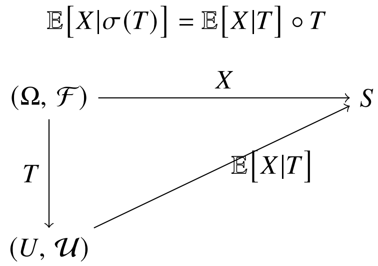


(b) Conditional expectation

Figure 8.1: Random variable $X : \Omega \to S$; indicated are the probabilities and the sigma algebras



(a) Projection



(b) The diagram commutes *on average*

Figure 8.2: Conditional expectation

Theorem 8.2 (Conditional expectation)**.** *Let X be a random variable with $\mathbb{E}|X| < \infty$. Let $\mathcal{G}$ be a sub-sigma algebra of $\mathcal{F}$. Then there exists a random variable Y such that*

(i) Y is $\mathcal{G}$-measurable,

(ii) $\mathbb{E}|Y| < \infty$,

(iii) it holds that
$$\int_G X\,dP = \int_G Y\,dP \text{ for every set } G \in \mathcal{G}. \tag{8.2}$$

Further, if $\tilde{Y}$ is another random variable with these properties (i)–(iii), then $Y = \tilde{Y}$ a.s. A random variable with the properties (i)–(iii) is called a version of the conditional expectation $\mathbb{E}(X \mid \mathcal{G})$ of X given $\mathcal{G}$ and we write $Y = \mathbb{E}(X \mid \mathcal{G})$.

For the proof of Theorem 8.2 we shall verify the following statement on existence of the projection in the Hilbert space first.

Proposition 8.3. *Let $\mathcal{K} \subset H$ be a closed subspace of the Hilbert space H. Let $X \in H$ be given. Then there exists $Y \in \mathcal{K}$ such that (cf. Figure 8.2a)*

- *$\|X - Y\| = \inf_{W \in \mathcal{K}} \|X - W\|$ and*
- *$X - Y \perp Z$, i.e.,*
$$\langle X, Z\rangle = \langle Y, Z\rangle \tag{8.3}$$
for all $Z \in \mathcal{K}$.

Proof. Choose $Y_n \in \mathcal{K}$ such that $\|X - Y_n\| \to \Delta \coloneqq \inf_{W\in\mathcal{K}} \|X - W\|$. By the parallelogram law,[6]
$$\underbrace{\|X - Y_n\|^2}_{\to\Delta^2} + \underbrace{\|X - Y_m\|^2}_{\to\Delta^2} = \underbrace{2\left\|X - \tfrac{1}{2}(Y_m + Y_n)\right\|^2}_{\geq 2\Delta^2} + 2\left\|\tfrac{1}{2}(Y_m - Y_n)\right\|^2. \tag{8.4}$$

[6]The parallelogram law follows by adding the lines
$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \quad \text{and}$$
$$\|x-y\|^2 = \langle x-y, x-y\rangle = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2;$$
here, chose $x \coloneqq X - \frac{1}{2}(Y_m + Y_n)$ and $y \coloneqq \frac{1}{2}(Y_m - Y_n)$.

It follows that $\|Y_m - Y_n\| \to 0$, and Y_n thus is a Cauchy sequence with a limit $Y \in \mathcal{K}$. It follows from (8.4) further that $\|X - Y\| = \Delta$; hence, the first assertion.

Now choose any $Z \in \mathcal{K}$. Clearly, $\|X - Y - t\,Z\|^2 \geq \|X - Y\|^2$, hence $-2t\,\langle Z, X - Y\rangle + t^2\,\|Z\|^2 \geq 0$. This can only hold true for all $t \in \mathbb{R}$ if $\langle Z, X - Y\rangle = 0$. This is (8.3), completing the proof. □

Proof of Theorem 8.2. We divide the proof into three steps.

(i) Uniqueness: suppose that $Y \neq \tilde{Y}$ a.s., then there is (without loss of generality) $\varepsilon > 0$ such that $P\left(Y - \tilde{Y} > \varepsilon\right) > 0$. But $G \coloneqq \left\{Y - \tilde{Y} > \varepsilon\right\} \in \mathcal{G}$, and hence $\int_G Y - \tilde{Y} dP > \varepsilon \cdot P(G) > 0$ which contradicts (iii).

(ii) Existence of $\mathbb{E}\,(X \mid \mathcal{G})$ for $X \in L^2(\mathcal{F})$. We employ the inner product $\langle X, Y\rangle \coloneqq \mathbb{E}\,XY$ on the Hilbert space $H = L^2(\mathcal{F})$. Suppose that $X \in L^2(\mathcal{F})$ and recall that $L^2(\mathcal{G})$ is a closed subspace of the Hilbert space $L^2(\mathcal{F})$. Thus, by Proposition 8.3, there is $Y \in L^2(\mathcal{G})$ such that $X - Y \perp Z$ (i.e., $\mathbb{E}(X - Y)Z = 0$) for all $Z \in L^2(\mathcal{G})$. For $G \in \mathcal{G}$ set $Z \coloneqq \mathbb{1}_G$ and observe with (8.3) that

$$\int_G X dP = \int_\Omega XZ dP = \mathbb{E}\,XZ = \mathbb{E}\,YZ = \int_\Omega YZ dP = \int_G Y dP,$$

so Y has the desired property (8.2).

(iii) Existence of $\mathbb{E}\,(X \mid \mathcal{G})$ for $X \in L^1(\mathcal{F})$. First, write $X = X^+ - X^-$ with $X^+, X^- \geq 0$. Hence $Y_n \coloneqq \mathbb{E}\,(X \wedge n \mid \mathcal{G})$ exists and it holds true that $0 \leq Y_n \leq Y_{n+1}$, so the limit $Y(\omega) \coloneqq \lim Y_n(\omega)$ exists pointwise. For any $Z = \mathbb{1}_G$ with $G \in \mathcal{G}$ it holds that

$$\begin{aligned}\int_G Y dP &= \mathbb{E}\,YZ = \lim_{n\to\infty} \mathbb{E}\,Y_n Z = \lim_{n\to\infty} \int_G Y_n dP \\ &= \lim_{n\to\infty} \int_G X \wedge n\, dP = \lim_{n\to\infty} \mathbb{E}(X \wedge n)Z = \mathbb{E}\,XZ = \int_G X dP\end{aligned}$$

by Lebesgue's monotone convergence theorem.

Hence the result. □

Proposition 8.4. *Suppose that $\mathcal{K} \subset H$ is spanned by independent vectors u_i, $i = 1, \dots, n$, and $X \in \mathcal{H}$. Then closest to X on $\mathcal{K}$ is $Y = \sum_{j=1}^n w_j\, u_j$, where $K\,w = f$, $K \coloneqq \left\langle u_i, u_j\right\rangle_{i,j=1}^n$ is the Gram matrix and $f \coloneqq \left(\langle X, u_i\rangle\right)_{i=1}^n$; more explicitly,*

$$Y = \sum_{i=1}^n \sum_{j=1}^n \langle X, u_i\rangle\, K_{ij}^{-1}\, u_j.$$

Proof. Let $Z = \sum_{i=1}^n a_i\, u_i \in \mathcal{K}$. We have that

$$\langle Y, Z\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i\, w_j \left\langle u_i, u_j\right\rangle = \sum_{i=1}^n a_i \sum_{j=1}^n K_{ij}\, w_j = \sum_{i=1}^n a_i\, f_i = \sum_{i=1}^n a_i\, \langle X, u_i\rangle = \langle X, Z\rangle\,,$$

i.e., $\langle Y, Z\rangle = \langle X, Z\rangle$ for every $Z \in \mathcal{K}$. This is the characterizing equation (8.3). □

Conditional expectation cheat sheet. The following properties of the conditional expectation hold true:

(i) Trivial sigma algebra: $\mathbb{E}\left[X \mid \{\emptyset, \Omega\}\right](\omega) = \mathbb{E}\, X$ for all $\omega \in \Omega$

(ii) The law of total expectation: $\mathbb{E}\,\mathbb{E}\left[X \mid \mathcal{G}\right] = \mathbb{E}\, X$ (choose $G = \Omega$ in (8.1))

(iii) If X is $\mathcal{G}$-measurable, then $\mathbb{E}\left[X \mid \mathcal{G}\right] = X$

Remark 8.5. It follows that $P\colon X \mapsto \mathbb{E}\left[X \mid \mathcal{G}\right]$ is a projection, i.e., $P = P^2$.

(iv) linear: $\mathbb{E}\left[\lambda_1 X_1 + \lambda_2 X_2 \mid \mathcal{G}\right] = \lambda_1\,\mathbb{E}\left[X_1 \mid \mathcal{G}\right] + \lambda_2\,\mathbb{E}\left[X_2 \mid \mathcal{G}\right]$

(v) positive: $\mathbb{E}\left[X \mid \mathcal{G}\right] \ge 0$, if $X \ge 0$

(vi) monotone: if $0 \le X_n \uparrow X$, then $\mathbb{E}\left[X_n \mid \mathcal{G}\right] \uparrow \mathbb{E}\left[X \mid \mathcal{G}\right]$

(vii) Fatou: if $X_n \ge 0$, then $\mathbb{E}\left[\liminf_{n\to\infty} X_n \mid \mathcal{G}\right] \le \liminf_{n\to\infty} \mathbb{E}\left[X_n \mid \mathcal{G}\right]$ a.s.

(viii) dominated convergence: if $X_n \to X$ a.s. and $X_n \le V$, $\mathbb{E}\, V < \infty$, then $\mathbb{E}\left[X_n \mid \mathcal{G}\right] \to \mathbb{E}\left[X \mid \mathcal{G}\right]$

(ix) Jensen: $\varphi\left(\mathbb{E}\left[X \mid \mathcal{G}\right]\right) \le \mathbb{E}\left[\varphi(X) \mid \mathcal{G}\right]$ for $\varphi(\cdot)$ convex; particularly $\left\|\mathbb{E}\left[X \mid \mathcal{G}\right]\right\|_p \le \|X\|_p$, $p \ge 1$ (cf. Theorem 5.25).

(x) Tower property:[7] $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right] = \mathbb{E}\left[X \mid \mathcal{H}\right]$, for $\mathcal{H} \subset \mathcal{G}$ a sub sigma algebra;

(xi) Taking out what is known: $\mathbb{E}\left[Z \cdot X \mid \mathcal{G}\right] = Z \cdot \mathbb{E}\left[X \mid \mathcal{G}\right]$, if Z is $\mathcal{G}$-measurable;

(xii) Role of independence: if $\mathcal{H}$ is independent from $\sigma\left(\mathcal{G}, \sigma(X)\right)$, then $\mathbb{E}\left[X \mid \sigma(\mathcal{G}, \mathcal{H})\right] = \mathbb{E}\left[X \mid \mathcal{G}\right]$

Remark 8.6. For an atom $G \in \mathcal{G}$, we may also set $\mathbb{E}(X \mid G) \coloneqq \frac{1}{P(G)} \int_G X\, dP$. Note, that $\mathbb{E}(X \mid G) \in S$, while $\mathbb{E}(X \mid \mathcal{G})\colon \Omega \to S$ is a random variable. It holds that $\mathbb{E}(X \mid G) = \mathbb{E}(X \mid G)(\omega)$ for a.e. $\omega \in G$.

8.3 CONDITIONING ON A RANDOM VARIABLE

The random variable
$$T\colon (\Omega, \mathcal{F}) \to (U, \mathcal{U})$$
generates the sigma algebra
$$\sigma(T) \coloneqq \sigma\left(\left\{T^{-1}(B)\colon B \in \mathcal{U}\right\}\right) \subset \mathcal{F}. \tag{8.5}$$

Theorem 8.7 (Doob–Dynkin lemma, aka. factorization lemma)**.** *Let $T\colon \Omega \to \mathbb{R}^m$, $X\colon \Omega \to \mathbb{R}^n$ be random variables. Then X is $\sigma(T)$-measurable if and only if there is a Borel measurable function $\varphi\colon \mathbb{R}^m \to \mathbb{R}^n$ such that $X = \varphi \circ T$, i.e., $X(\omega) = \varphi\big(T(\omega)\big)$.*

Proof. Let $X = \sum_{i=1}^n f_i\, \mathbb{1}_{A_i}$ be a simple function with $A_i \cap A_j = \emptyset$ for $i \ne j$ which is $\sigma(T)$ measurable. Then there are $B_i \in \mathbb{R}^m$ such that $A_i = \{\omega\colon T(\omega) \in B_i\} = T^{-1}(B_i)$. Set $\varphi \coloneqq \sum_{i=1}^n f_i\, \mathbb{1}_{B_i}$. For $\omega \in A_i$ we have that $X(\omega) = f_i$ and $\varphi\big(T(\omega)\big) = f_i$ and hence $X = \varphi \circ T$.

For $X \ge 0$ measurable there exists a sequence of simple functions X_n so that $X_n(\omega) \nearrow X(\omega)$ a.s. and $X_n(\omega) = \varphi_n\big(T(\omega)\big)$. Let B' denote the Borel set
$$B' \coloneqq \left\{x \in \mathbb{R}^m\colon \lim_{n\to\infty} \varphi_n(x) \text{ exists}\right\}.$$
On B' define $\varphi(x) \coloneqq \lim_{n\to\infty} \varphi_n(x)$, which is $\sigma(T)$ measurable. Apparently, it holds that $X = \varphi(T)$ a.e. □

[7] Glättungsregel

For the following Doob–Dynkin lemma see as well Kallenberg [9, Lemma 1.13] or Shiryaev [22, Theorem II.4.3].

Definition 8.8. The random variable $\mathbb{E}\big(X \mid \sigma(T)\big)$ is measurable with respect to $\sigma(T)$. By the Doob–Dynkin lemma, there is a measurable function φ so that $\varphi \circ T = \mathbb{E}\big(X \mid \sigma(T)\big)$. For this function we write

$$\mathbb{E}(X \mid T) \coloneqq \varphi.$$

Remark 8.9. The preceding definition also justifies writing $\mathbb{E}(X \mid T = t) = \varphi(t)$, the function is

$$\begin{aligned}\mathbb{E}(X \mid T)\colon U &\to S\\ t &\mapsto \mathbb{E}(X \mid T = t),\end{aligned}$$

cf. Figure 8.2b. By definition,

$$\mathbb{E}(X \mid T) \circ T = \mathbb{E}\big(X \mid \sigma(T)\big). \tag{8.6}$$

There is a factorization *on average* in the following sense.

Theorem 8.10 (Radon–Nikodým)**.** *The function $\mathbb{E}(X \mid T)$ is the Radon–Nikodým derivative of the measure $P_X^T = P_X \circ T^{-1}$ with respect to the image measure $P^T = P \circ T^{-1}$,*

$$\mathbb{E}(X \mid T) = \frac{dP_X^T}{dP^T},$$

where $P_X(C) \coloneqq \int_C X\, dP$. I.e.,

$$\int_{T^{-1}(B)} X\, dP = \int_{T^{-1}(B)} \mathbb{E}(X \mid T) \circ T\, dP \qquad \text{for all } B \in \mathcal{U}. \tag{8.7}$$

Proof. Note that $P^T(B) = P\big(T^{-1}(B)\big) = 0$ implies that $P_X^T(B) = P_X\big(T^{-1}(B)\big) = 0$. Hence, $P_X^T \ll P^T$, i.e., P_X^T is absolutely continuous with respect to P^T. By the change of variables formula it holds that

$$\begin{aligned}\int_B \mathbb{E}(X \mid T)\, dP^T &= \int_{T^{-1}(B)} \mathbb{E}(X \mid T) \circ T\, dP\\ &\underset{(8.6)}{=} \int_{T^{-1}(B)} \mathbb{E}\big(X \mid \sigma(T)\big)\, dP\\ &\underset{(8.2)}{=} \int_{T^{-1}(B)} X\, dP\\ &= P_X^T(B),\end{aligned}$$

i.e., $\mathbb{E}\,(X \mid T)$ is the Radon–Nikodým derivative of the measure P_X^T with respect to the measure P^T, the assertion. □

8.4 LAW OF TOTAL VARIANCE

Definition 8.11. The conditional variance is

$$\operatorname{var}(Y|\, Z) \coloneqq \mathbb{E}\left(\big(Y - \mathbb{E}(Y \mid Z)\big)^2 \Big|\, Z\right) = \mathbb{E}\left(Y^2 - \mathbb{E}(Y \mid Z)^2 \big|\, Z\right) = \mathbb{E}\left(Y^2 \big|\, Z\right) - \big(\mathbb{E}(Y \mid Z)\big)^2. \tag{8.8}$$

Theorem 8.12 (The rule of double variance, aka. law of total variance)**.** *The variance for $Y \in L^2$ can be stated as*

$$\operatorname{var} Y = \mathbb{E} \operatorname{var}(Y|\, Z) + \operatorname{var} \mathbb{E}(Y \mid Z).$$

Proof. We have that

$$\begin{aligned}\operatorname{var} Y &= \mathbb{E}\, Y^2 - (\mathbb{E}\, Y)^2 = \mathbb{E}\,\mathbb{E}\left(Y^2 \middle| Z\right) - \left(\mathbb{E}\,\mathbb{E}\left[Y|\, Z\right]\right)^2\\ &= \mathbb{E}\left[\operatorname{var}\left(Y|\, Z\right) + \left(\mathbb{E}\left(Y \mid Z\right)\right)^2\right] - \left(\mathbb{E}\,\mathbb{E}\left[Y|\, Z\right]\right)^2\\ &= \mathbb{E}\operatorname{var}\left(Y|\, Z\right) + \mathbb{E}\left(\mathbb{E}\left(Y \mid Z\right)\right)^2 - \left(\mathbb{E}\,\mathbb{E}\left(Y|\, Z\right)\right)^2\\ &= \mathbb{E}\operatorname{var}\left(Y|\, Z\right) + \operatorname{var}\mathbb{E}\left(Y \mid Z\right),\end{aligned}$$

which is the assertion. ◻

Definition 8.13. The conditional covariance is

$$\begin{aligned}\operatorname{cov}(X, Y \mid Z) &:= \mathbb{E}\left(\left(X - \mathbb{E}(Y \mid Z)\right) \cdot \left(Y - \mathbb{E}(Y \mid Z)\right) \middle| Z\right) &\quad (8.9)\\ &= \mathbb{E}\left(XY - \mathbb{E}(X \mid Z) \cdot \mathbb{E}(Y \mid Z) | Z\right). &\quad (8.10)\end{aligned}$$

Theorem 8.14 (The law of total covariance)**.** *The covariance for $X \in L^2$ and $Y \in L^2$ can be stated as*

$$\operatorname{cov}(X, Y) = \mathbb{E}\operatorname{cov}(X, Y \mid Z) + \operatorname{cov}\left(\mathbb{E}(X \mid Z),\, \mathbb{E}(Y \mid Z)\right).$$

8.5 CONDITIONAL PROBABILITIES

Notably, we have introduced the conditional expectation without defining conditional probabilities. This was different in the motivating Example 8.1.

Definition 8.15. The conditional probability is

$$P(A \mid \mathcal{G}) := \mathbb{E}\left(\mathbb{1}_A \mid \mathcal{G}\right).$$

Lemma 8.16 (Bayes' theorem[8])**.** *Suppose that $\mathcal{G} = \sigma\left(B_i \colon i = 1, 2, \dots\right)$ with $B_i \cap B_j = \emptyset$. Then*

$$P(A \mid \mathcal{G})(\omega) = \frac{P(A \cap B_i)}{P(B_i)} =: P(A \mid B_i) \quad \text{for almost every } \omega \in B_i,$$

provided that $P(B_i) > 0$, $i = 1, 2, \dots$.

Proof. From measurability and from $B_i \cap B_j = \emptyset$ it follows that $P(A \mid \mathcal{G})$ is constant on each atom B_i, $c = P(A \mid \mathcal{G})$ a.e., say. Then, as $B_i \in \mathcal{G}$,

$$P(A \cap B_i) = \int_{B_i} \mathbb{1}_A \, dP = \int_{B_i} \mathbb{E}(\mathbb{1}_A \mid \mathcal{G}) dP = \int_{B_i} P(A \mid \mathcal{G}) dP = c \cdot P(B_i)$$

and thus the assertion. ◻

Corollary 8.17. *For densities we have that*

$$f_X(x \mid y) \cdot f_Y(y) = f_{X,Y}(x, y), \quad (8.11)$$

where

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx \quad (8.12)$$

is the density of the marginal distribution.

[8] Thomas Bayes, 1701–1761

Proof. Define $X \coloneqq \mathbb{1}_{A\times\mathbb{R}}$ and $T(x,y) \coloneqq y$, such that $T^{-1}(B) = \mathbb{R}\times B$. Then it holds that

$$\begin{aligned}\iint_{A\times B} f(x,y)dxdy &= \iint_{\mathbb{R}\times B} \mathbb{1}_{A\times\mathbb{R}}\, dP = \iint_{\mathbb{R}\times B} P\left(A\times\mathbb{R} \mid y\right) P(dx,dy)\\ &= \int_B \int_{\mathbb{R}} f(x,y)dx\, P\left(X\in A \mid y\right) dy\\ &= \int_B P\left(X\in A\mid y\right) f(y)dy = \int_B\int_A f(x\mid y)dx\, f(y)dy,\end{aligned}$$

from which the assertion follows. □

Remark 8.18. Note, that $P(T=y)=0$.

Remark 8.19. It holds that $\mathbb{E}\left[g(X)\mid Y=y\right] = \int g(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}dx$, cf. (8.12).

Remark 8.20. A different and perhaps more instructive way of writing (8.11) is

$$f_X(x\mid Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \tag{8.13}$$

which relates to Bayes' rule, Remark 8.16.

8.6 MARTINGALES

Definition 8.21 (Martingale)**.** A process X is a (sub-, super-) *martingale* (relative to $\left(\mathcal{F}_t : t\in\mathbf{T},\, P\right)$ if

(i) X is adapted, i.e., X_t is $\mathcal{F}_t$-measurable for every $t\in\mathbf{T}$,

(ii) $\mathbb{E}\,|X_t| < \infty$ for all $t\in\mathbf{T}$, and

(iii) $\mathbb{E}\left(X_{t'}|\,\mathcal{F}_t\right) = X_t$ a.s. for all $t\le t'$ $(t,\, t'\in\mathbf{T})$ (super: $X_t \ge \mathbb{E}\left(X_{t'}|\,\mathcal{F}_t\right)$, sub: $X_t \le \mathbb{E}\left(X_{t'}|\,\mathcal{F}_t\right)$).

8.7 DOOB'S MARTINGALE INEQUALITIES

Definition 8.22. The process $M_t \coloneqq \sup_{s\le t} X_s$ is called the *running maximum process* of the stochastic process X_t.

Doob's submartingale inequality provides an upper bound of the running maximum process in terms of the genuine process:

Theorem 8.23 (Doob's submartingale inequality, aka Doob's maximal inequalities)**.** *Let X be a non-negative and continous submartingale (i.e., $t\mapsto X_t(\omega)$ is continuous a.s.). Then, for $\lambda>0$ and $p\ge 1$,*

$$P\left(\sup_{s\le t} X_s \ge \lambda\right) \le \frac{1}{\lambda^p}\,\mathbb{E}\left(X_t\cdot \mathbb{1}_{\sup_{s\le t} X_s\ge\lambda}\right) \le \frac{1}{\lambda^p}\,\mathbb{E}\,X_t^p. \tag{8.14}$$

For $p>1$,

$$\mathbb{E}\left(\sup_{s\le t} X_s^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\,X_t^p \text{ or } \|M_t\|_p \le \frac{p}{p-1}\,\|X_t\|_p\,.$$

Proof. We prove the first part of the theorem in case of a submartingale in *discrete* time $X = (X_{t_k})_{k=1}^n$ (see also Williams [27, Section 14.6]). Note that $E \coloneqq \left\{\sup_{s\le t} X_s \ge \lambda\right\}$ is the *disjoint* union $E = E_{t_0} \cup E_{t_1} \cup \dots \cup E_{t_n}$, where

$$\begin{aligned} E_0 &:= \left\{X_{t_0} \ge \lambda\right\}, \\ E_k &:= \left\{X_{t_0} < \lambda\right\} \cap \dots \cap \left\{X_{t_{k-1}} < \lambda\right\} \cap \left\{X_{t_k} \ge \lambda\right\}. \end{aligned}$$

Note, that $E_k \in \mathcal{F}_{t_k}$ and $X_{t_k} \ge \lambda$ on E_k. Hence, as X is a submaritingale (with $0 = t_0 < t_1 < \dots < t_n = t$),

$$\mathbb{E}\left(X_t^p \cdot \mathbb{1}_{E_k}\right) \ge \mathbb{E}\left(X_{t_k}^p \cdot \mathbb{1}_{E_k}\right) \ge \lambda^p \cdot P(E_k)$$

by Makov's ineqaulity (Theorem 5.23). Summing over k now yields (8.14), the first result.

We have from (8.14) that $P\left(\sup_{s\le t} X_s \ge \lambda\right) \le \frac{1}{\lambda}\mathbb{E}\left(X_t \cdot \mathbb{1}_{\sup_{s\le t} X_s \ge \lambda}\right)$ and consequently

$$\int_0^\infty \lambda^{p-1} \cdot P\left(\sup_{s\le t} X_s \ge \lambda\right) d\lambda \le \int_0^\infty \lambda^{p-2} \cdot \mathbb{E}\left(X_t \cdot \mathbb{1}_{\sup_{s\le t} X_s \ge \lambda}\right) d\lambda.$$

By Fubini's theorem thus (choose $g(x) = \frac{x^p}{p}$ in (5.25))

$$\int_0^\infty \lambda^{p-1} \cdot P\left(\sup_{s\le t} X_s \ge \lambda\right) d\lambda = \frac{1}{p}\,\mathbb{E} \sup_{s\le t} X_s^p$$

and similarly

$$\begin{aligned} \int_0^\infty \lambda^{p-2} \cdot \mathbb{E}\left(X_t\, \mathbb{1}_{\sup_{s\le t} X_s \ge \lambda}\right) d\lambda &= \int_0^\infty \lambda^{p-2}\, \mathbb{E}\,\mathbb{E}\left(X_t \cdot \mathbb{1}_{\sup_{s\le t} X_s \ge \lambda} \middle| X_t\right) d\lambda \\ &= \int_0^\infty \lambda^{p-2} \int_{\mathbb{R}} \mathbb{E}\left(x \cdot \mathbb{1}_{\sup_{s\le t} X_s \ge \lambda} \middle| X_t = x\right) P(X_t \in dx) d\lambda \\ &= \int_{\mathbb{R}} x \cdot \int_0^\infty \lambda^{p-2} P\left(\sup_{s\le t} X_s \ge \lambda \middle| X_t = x\right) d\lambda\, P(X_t \in dx) \\ &\underset{(5.25)}{=} \int_{\mathbb{R}} x \cdot \mathbb{E}\left[\frac{1}{p-1} \sup_{s\le t} X_s^{p-1} \middle| X_t = x\right] P(X_t \in dx) \\ &= \frac{1}{p-1}\,\mathbb{E}\left(\sup_{s\le t} X_s^{p-1} \cdot X_t\right); \end{aligned}$$

hence, and after applying Hölder's inequality (Theorem 5.26, $\frac{1}{p} + \frac{1}{\frac{p}{p-1}} = 1$),

$$\mathbb{E} \sup_{s\le t} X_s^p \le \frac{p}{p-1}\,\mathbb{E}\left(\sup_{s\le t} X_s^{p-1} \cdot X_t\right) \le \frac{p}{p-1}\left(\mathbb{E} \sup_{s\le t} X_s^p\right)^{\frac{p-1}{p}} \cdot \left(\mathbb{E}\, X_t^p\right)^{\frac{1}{p}}.$$

Consequently,

$$\mathbb{E} \sup_{s\le t} X_s^p \le \left(\frac{p}{p-1}\right)^p \cdot \mathbb{E}\, X_t^p,$$

the assertion. □

8.8 DISINTEGRATION

Theorem 8.24. *Let $P\colon \mathcal{A} \otimes \mathcal{B} \to [0,1]$ be a probability measure on $\mathcal{X} \times \mathcal{Y}$. Then there exists a kernel $P\colon \mathcal{B} \times \mathcal{X} \to [0,1]$ such that*

(i) $x \mapsto P(B \mid x)$ *is measurable for every* $B \in \mathcal{B}$,

(ii) $B \mapsto P(B \mid x)$ *is a probability measure for every* $x \in X$,

(iii) $P(A \times B) = \int_A P(B \mid x)\, P'(\mathrm{d}x)$, *where* $P'(A) \coloneqq P(A \times \mathcal{Y})$ *is the marginal measure.*

Proof. For every $B \in \mathcal{B}$, the mapping $P(\cdot \times B) \colon \mathcal{A} \to [0,1]$, with $A \mapsto P(A \times B)$, is a measure and note that $P(A \times B) \le P'(A)$ for every $A \in \mathcal{A}$. It follows in particular, that $P(\cdot \times B) \ll P'(\cdot)$. By Radon–Nikodým, there exists a measurable function f_B with $0 \le f_B \le 1$ (a density) such that $P(A \times B) = \int_A f_B(x)\, P'(\mathrm{d}x)$. We shall write $P(B \mid x) \coloneqq f_B(x)$ so that $P(A \times B) = \int_A P(B \mid x)\, P'(\mathrm{d}x)$, hence (i) and (iii).

Let $B_n \in \mathcal{B}$ be pairwise disjoint. As $f_{B_n} \ge 0$, it follows with the monotone convergence theorem that

$$\sum_{n \in \mathbb{N}} \int_A f_{B_n}(x)\, P'(\mathrm{d}x) = \int_A \sum_{n \in \mathbb{N}} f_{B_n}(x)\, P'(\mathrm{d}x).$$

Further, as B_n are pairwise disjoint, we have that

$$\sum_{n \in \mathbb{N}} \int_A f_{B_n}(x)\, P'(\mathrm{d}x) = \sum_{n \in \mathbb{N}} P(A \times B_n) = P\Big(A \times \bigcup_{n \in \mathbb{N}} B_n\Big) = \int_A f_{\bigcup_{n \in \mathbb{N}} B_n}(x)\, P'(\mathrm{d}x).$$

As the density is unique, it follows that $P\left(\bigcup_{n \in \mathbb{N}} B_n \mid x\right) = \sum_{n \in \mathbb{N}} P(B_n \mid x)$. That is, $P(\cdot \mid x)$ is a σ-additive measure P'- almost everywhere. Further, it holds that $P'(A) = P(A \times \mathcal{Y}) = \int_A P(\mathcal{Y} \mid x)\, P'(\mathrm{d}x)$ so that $P(\mathcal{Y} \mid x) = 1$ and $P(\cdot \mid x)$ is indeed a probability measure. □

Theorem 8.25. *Let f be a random variable and P be a probability measure on $\mathcal{X} \times \mathcal{Y}$. Then* $\mathbb{E}(f \mid x) = \int_{\mathcal{Y}} f(x,y)\, P(\mathrm{d}y \mid x)$.

Proof. Indeed, for the measurable sets $G \times \mathcal{Y}$ generating the σ-algebra it holds that

$$\begin{aligned}\iint_{G \times \mathcal{Y}} \mathbb{E}(f \mid x)\, \mathrm{d}P &= \int_G \int_{\mathcal{Y}} \mathbb{E}(f \mid x)\, P(\mathrm{d}y \mid x)\, P'(\mathrm{d}x) \\ &= \int_G \mathbb{E}(f \mid x)\, P'(\mathrm{d}x) \\ &= \int_G \int_{\mathcal{Y}} f(x,y)\, P(\mathrm{d}y \mid x)\, P'(\mathrm{d}x) \\ &= \iint_{G \times \mathcal{Y}} f(x,y)\, P(\mathrm{d}x, \mathrm{d}y),\end{aligned}$$

the defining equation of the conditional expectation. □

8.9 PROBLEMS

Exercise 8.1. *Show that (8.5) is a sigma algebra.*

Exercise 8.2. *Show that* $\left\{B \subset U \colon T^{-1}(B) \in \mathcal{F}\right\}$ *(in the setting of Section 8.3) is a sigma algebra.*

Exercise 8.3. *Verify the cheat list (i)–(xii).*

Exercise 8.4. *Show that* $f(x \mid y)$ *given in (8.11) is indeed a density.*

Exercise 8.5. *Discuss the Borel–Kolmogorov paradox.*

Exercise 8.6. *Give the conditional probabilities* $P(Y = y_i \mid X = 3)$ *and* $P(X = x_i \mid Y = 5)$ *for the distribution in Exercise 6.5.*

Exercise 8.7. *Verify the definition of the double variance formula, Eq. (8.8).*

Exercise 8.8 (Wald's formulas[9])**.** *Let* X_i *be iid,* $N \in \mathbb{N}$ *random with* $N \perp X_i$ *for* $i = 1, 2, \dots$. *Then the following hold for* $Z \coloneqq X_1 + X_2 + \cdots + X_N$ *(Hint: the double variance formula, or ¿¿):*

(i) $\mathbb{E}\, Z = \mathbb{E}\, X \cdot \mathbb{E}\, N$,

(ii) $\operatorname{var} Z = \operatorname{var} X \cdot \mathbb{E}\, N + (\mathbb{E}\, X)^2 \cdot \operatorname{var} N$.

[9] Abraham Wald, 1902–1950

9 Stochastic Processes and Martingales

9.1 STOCHASTIC PROCESSES

Definition 9.1. A *filtered probability space* (also known as *stochastic basis*) on the totally ordered set $\mathbf{T}$ is a quadruple

$$\big(\Omega, (\mathcal{F}_t)_{t\in\mathbf{T}}, \mathcal{F}, P\big),$$

where

(i) $\big(\Omega, \mathcal{F}, P\big)$ is a probability space, and

(ii) $(\mathcal{F}_t)_{t\in\mathbf{T}}$ is a filtration, that is, an increasing sequence of sub-sigma algebras of $\mathcal{F}$: $\mathcal{F}_t \subset \mathcal{F}_{t'} \subset \ldots \mathcal{F}$ whenever $t \le t'$.

We define $\mathcal{F}_\infty \coloneqq \sigma\left(\bigcup_{t\in\mathbf{T}} \mathcal{F}_t\right) \subset \mathcal{F}$.

Typical index sets include $\mathbf{T} = \{0, 1, 2 \ldots\}$, $\mathbf{T} = [0, T]$ and $\mathbf{T} = [0, \infty)$.

Definition 9.2. The collection $X = (X_t)_{t\in\mathbf{T}}$ is a stochastic process, provided that $X_t : \Omega \to (S, \Sigma)$ for every $t \in \mathbf{T}$.[1]

Interpretation. One may usually associate $t \in \mathbf{T}$ with *time* and ω with *particle* or *experiment*. With this picture $X_t(\omega)$ describes the position (or result) of the particle (experiment) ω at time t.

Remark 9.3. The stochastic process X can also be seen as a function,

$$\begin{aligned} X\colon \mathbf{T}\times\Omega &\to (S,\Sigma) \\ (t,\omega) &\mapsto X_t(\omega) \end{aligned}$$

assuming that X is *jointly* measurable.

Definition 9.4. For $\omega \in \Omega$, the mapping

$$\begin{aligned} X(\omega)\colon \mathbf{T} &\to S \\ t &\mapsto X_t(\omega) \end{aligned}$$

is called a *path*. Note, that $X(\omega) \in S^T$.

Remark 9.5. One may also define

$$\begin{aligned} \Phi_X : \Omega &\to S^{\mathbf{T}} \\ \omega &\mapsto \Phi_X(\omega)\colon \mathbf{T} \to S \\ &\qquad\qquad t \mapsto X_t(\omega), \end{aligned}$$

where $S^{\mathbf{T}}$ is the collection of all functions from $\mathbf{T}$ to S. The *law* of a stochastic process is the pushforward measure (image measure) $\mathcal{L}_X(\cdot) \coloneqq P \circ \Phi_X^{-1}(\cdot)$.

[1] In almost all situations we consider $(S, \Sigma) = (\mathbb{R}^n, \mathcal{B})$.

Definition 9.6 (Adapted process, natural filtration)**.** A process is said to be *adapted* to the filtration $\mathcal{F}_t$, if X_t is measurable with respect to $\mathcal{F}_t$ for every $t \in \mathbf{T}$.

The *natural filtration* induced by the stochastic process X is

$$\mathcal{F}_t \coloneqq \sigma\left(X_s^{-1}(A) \colon A \in \Sigma,\ s \le t,\ s \in \mathbf{T}\right).$$

Remark 9.7. Notice, that we don't have a sigma algebra on $S^{\mathbf{T}}$. However, we would like to assign probabilities to sets as

$$F \coloneqq \left\{\omega \colon X_{t_1}(\omega) \in A_{t_1},\ X_{t_2}(\omega) \in A_{t_2}, \dots, X_{t_n}(\omega) \in A_{t_n}\right\} \in \mathcal{F}$$

with $A_{t_i} \in \Sigma$ for all $i = 1, \dots n$ as it is natural to ask for the probability

$$P\left(\left\{X_{t_1} \in A_{t_1},\ X_{t_2} \in A_{t_2}, \dots, X_{t_n} \in A_{t_n}\right\}\right).$$

9.2 EXAMPLES OF STOCHASTIC PROCESSES, DISCRETE TIME STOCHASTIC PROCESSES

Often we consider processes, which are driven by a mean zero driving process Z_t. First order processes are of the form

$$X_{t+1} = F(X_t, Z_t),$$

i.e., the value at time $t + 1$ does only depend on the value X_t at time t and the driving process Z_t at time t.

9.2.1 The additive model (random walk model)

The simplest model is the Bernoulli random walk model. Here, the driving process is $Z_t = 2Y_t - 1$, where $Y_t \sim B(1, 1/2)$ (cf. Section 5.6.2). The process is

$$X_{t+1} = X_t + u \cdot (2Y_t - 1)$$

with some starting value $X_0 = x_0$. It is easy to see that $\mathbb{E}(X_t) = x_0$ and $\operatorname{var}(X_t) = u^2 \cdot t$. The process is not stationary (i.e., the distribution of X_t differs from the distribution of $X_{t'}$ whenever $t \neq t'$).

9.2.2 The multiplicative model (Black–Derman–Toy or lattice model)

The additive model has the disadvantage that the process may fall negative. The multiplicative model

$$X_{t+1} = X_t \cdot v^{2Y_t - 1}$$

avoids this. Notice that $\log X_t$ follows the recursion $\log X_{t+1} = \log X_t + (\log\, v) \cdot (2Y_t - 1)$.

9.2.3 The autoregressive process with mean reversion

The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term), as

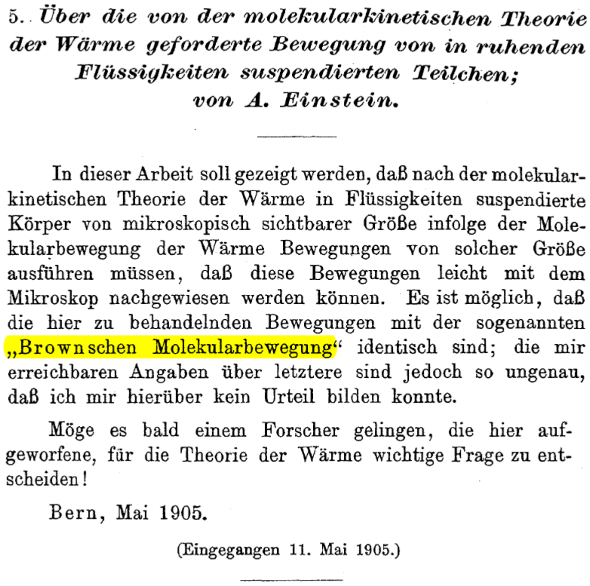
$$X_{t+1} = X_t + a(\mu - X_t) + Z_t,$$

where $(Z_t)_{t=1}$ is a zero mean i.i.d. process.

9.3 PROBLEMS

Example 9.8. Discuss, why B_{2t} is not adapted, while $B_{t/2}$ is.

10 Brownian Motion



5. ***Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen; von A. Einstein.***

In dieser Arbeit soll gezeigt werden, daß nach der molekularkinetischen Theorie der Wärme in Flüssigkeiten suspendierte Körper von mikroskopisch sichtbarer Größe infolge der Molekularbewegung der Wärme Bewegungen von solcher Größe ausführen müssen, daß diese Bewegungen leicht mit dem Mikroskop nachgewiesen werden können. Es ist möglich, daß die hier zu behandelnden Bewegungen mit der sogenannten „Brownschen Molekularbewegung" identisch sind; die mir erreichbaren Angaben über letztere sind jedoch so ungenau, daß ich mir hierüber kein Urteil bilden konnte.

Möge es bald einem Forscher gelingen, die hier aufgeworfene, für die Theorie der Wärme wichtige Frage zu entscheiden!

Bern, Mai 1905.

(Eingegangen 11. Mai 1905.)

Figure 10.1: Einstein [5] in 1905, his miraculous year/ annus mirabilis

10.1 INFORMAL INTRODUCTION OF THE WIENER PROCESS

Let S_t be a symmetric random walk, i.e. (cf. Section 9.2.1),

$$\begin{aligned} S_0 &= 0, \\ S_j &:= S_{j-1} + \xi_j, \qquad j = 1, 2, \dots, \end{aligned} \tag{10.1}$$

where ξ_j is a sequence of independent random variables with $P(\xi_j = \pm 1) = \frac{1}{2}$ (so that $\frac{1}{2}(1+\xi_j)$ is Bernoulli $B(1, 1/2)$). Let

$$W_n(t) := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j = \frac{1}{\sqrt{n}} \sum_{j \le n \cdot t} \xi_j,$$

where $\lfloor x \rfloor$ is the floor function.[1]

The process W_n has the following properties:

(i) $\mathbb{E}\, W_n(t) = 0$, $\operatorname{var} W_n(t) = \operatorname{var} \sum_{j=0}^{\lfloor nt \rfloor} \frac{1}{\sqrt{n}} \xi_j = \sum_{j=0}^{\lfloor nt \rfloor} \frac{1}{n} \operatorname{var} \xi_j = \frac{\lfloor nt \rfloor}{n}$,

[1] $\lfloor \cdot \rfloor$ is the floor function, i.e., $\lfloor x \rfloor = m \in \mathbb{Z}$ and $m \le x < m+1$.

(ii) W_n has independent increments, (i.e., $W_n(t_1) - W_n(s_1)$ and $W_n(t_2) - W_n(s_2)$ are independent whenever $s_1 < t_1 < s_2 < t_2$);

(iii) W_n has stationary increments (i.e., they do not vary in time);

(iv) W_n is a martingale, i.e., $\mathbb{E}\left[W_n(t) \mid W_n(s)\right] = W_n(s)$ for $s < t$ (i.e., $\mathbb{E}\left[W_n(t) \mid W_n(s) = x\right] = x$); further, $\mathbb{E}\left[W_n(t) \mid W_n(v),\ 0 \le v \le s\right] = W_n(s)$ for $s < t$;

(v) W_n is càdlàg (French: "continue à droite, limite à gauche"; the collection of càdlàg functions on a given domain is known as Skorokhod space).

As $n \to \infty$, the process W_n converges (in distribution) to a limiting process, called the *Wiener*[2] *process*

$$W_t := \lim_{n\to\infty} W_n(t).$$

The Wiener process W_t has following properties:

(i) $W_t \sim \mathcal{N}(0, t)$, in particular $\mathbb{E}\, W_t = 0$ and $\operatorname{var} W_t = t$,

(ii) The covariance is

$$\operatorname{cov}\left(W_s, W_t\right) = s \wedge t, \tag{10.2}$$

where $s \wedge t := \min\{s, t\}$.

(iii) W is a martingale: $\mathbb{E}\left(W_t \,|\, W_v,\ 0 \le v \le s\right) = W_s$ for $s < t$,

(iv) W has independent increments (i.e., $W_{t_1} - W_{s_1} \perp W_{t_2} - W_{s_2}$) and

(v) W has stationary increments. More specifically, $W_t - W_s \sim \mathcal{N}(0, t - s)$ for $s < t$.

Proof. Assuming that $s < t$ it holds that

$$\begin{aligned}\operatorname{cov}(W_t - W_s, W_s) &= \mathbb{E}(W_t - W_s)W_s = \mathbb{E}\,\mathbb{E}\left[(W_t - W_s)\cdot W_s \,|\, W_s\right]\\ &= \mathbb{E}\left[W_s \cdot \mathbb{E}\left[(W_t - W_s) \,|\, W_s\right]\right] = 0.\end{aligned}$$

Hence

$$\operatorname{cov}(W_t, W_s) = \operatorname{cov}(W_t - W_s, W_s) + \operatorname{var}(W_s) = 0 + s = s$$

and

$$\begin{aligned}\operatorname{var}(W_t - W_s) &= \operatorname{var} W_t + \operatorname{var} W_s - 2\operatorname{cov}(W_t, W_s)\\ &= t + s - 2s = t - s.\end{aligned}$$

Assuming $s < t < u < v$ it holds further that

$$\begin{aligned}\operatorname{cov}(W_v - W_u, W_t - W_s) &= \operatorname{cov}(W_v, W_t) - \operatorname{cov}(W_v, W_s) - \operatorname{cov}(W_u, W_t) + \operatorname{cov}(W_u, W_s) =\\ &= t - s - t + s = 0,\end{aligned}$$

i.e., the Wiener process has uncorrelated, and by normality hence independent increments.

For the reamining assertions note that

$$W_n(t) - W_n(s) = \frac{1}{\sqrt{n}} \sum_{j=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \xi_j \sim \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt\rfloor-\lfloor ns\rfloor} \xi_j \approx \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n(t-s)\rfloor} \xi_j.$$

□

[2]Norbert Wiener, 1894–1964

10.2 EXISTENCE OF BROWNIAN MOTION

Remark 10.1. More generally one may define the Wiener process as the weak limit of

$$\tilde{W}_n(t) \coloneqq \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j,$$

where ξ_j are i.i.d. variables with mean 0 and variance 1 (so in contrast to (10.1) not necessarily Binomial), or by linear interpolation, as

$$\tilde{\tilde{W}}_n(t) \coloneqq \tilde{W}_n(t) + \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}.$$

Definition 10.2. The Wiener process is characterized by the following four properties:

(i) $W_0 = 0$

(ii) W_t is almost surely continuous

(iii) W_t has independent increments, i.e., $W_{t_1} - W_{s_1}$ and $W_{t_2} - W_{s_2}$ are independent for $0 \le s_1 \le t_1 \le s_2 \le t_2$

(iv) $W_t - W_s \sim \mathcal{N}(0, t - s)$ for all $s < t$.

Lemma 10.3 (Self similarity)**.** *Let W_t be a Wiener process.*

(i) Symmetry: $-W_t$ is a Wiener process;

(ii) Time-reversal: $W_T - W_{T-t}$ is a Wiener process for $t \in [0, T]$;

(iii) Scaling: For $c > 0$, the process $\frac{1}{\sqrt{c}} W_{ct}$ is a Wiener process again;

(iv) Time-inversion: $t \cdot W_{1/t}$ is a Wiener process.

Proof. As for (iv), note that $\mathbb{E}\, \tilde{W}_s \tilde{W}_t = st \left(\frac{1}{s} \wedge \frac{1}{t}\right)$ for $\tilde{W}_t \coloneqq t \cdot W_{1/t}$; see Exercise 10.3. □

10.2.1 Using Kolmogorov's extension theorem

We demonstrate first existence of Brownian motion.[3] We shall denote the transition probability by

$$p(t; x, y) \coloneqq \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}(x-y)^2} \tag{10.3}$$

with the interpretation that $p(t, x, y)dy$ is the probability of a particle located in x to move to dy within time t (i.e., $P(X_{t+\Delta t} \in dy) = p(\Delta t; x, y)dy$, cf. Remark 5.38). As an example consider the times $0 = t_0 < t_1 < \cdots < t_n$ and verify that a Brownian motion, starting at $t_0 = 0$ in $x = 0$, has the cdf. (cf. Exercise 10.2)

$$\begin{aligned} F_{(W_{t_1}, \dots W_{t_n})}(x_1, \dots x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} & p(t_1; 0, y_1) \cdot p(t_2 - t_1; y_1, y_2) \cdot \ldots \\ & \cdot p(t_n - t_{n-1}; y_{n-1}, y_n)\, dy_n \ldots dy_2 dy_1. \end{aligned} \tag{10.4}$$

Existence follows from Kolmogorov's extension theorem:

[3] Robert Brown, 1773–1858, Scottish botanist

Theorem 10.4 (Kolmogorov's extension theorem)**.** *Let $\nu_{t_1,\dots,t_n}$ be probability measures. Suppose these measures satisfy the following two conditions:*

- $\nu_{t_1,\dots t_n}(F_1\times\dots\times F_n)=\nu_{t_{\pi(1)},\dots,t_{\pi(n)}}(F_{\pi(1)}\times\dots\times F_{\pi(n)})$ *for all permutations π of* $\{1,2,\dots n\}$,
- $\nu_{t_1,\dots,t_k}(F_1\times\dots\times F_k)=\nu_{t_1,\dots t_n}(F_1\times\dots\times F_k\times\underbrace{\mathbb{R}^n\times\dots\times\mathbb{R}^n}_{n-k\ times})$.

Then there exists a probability space $(\Omega,\mathcal{F},P)$ and a stochastic process $X\colon \mathbf{T}\times\Omega\to\mathbb{R}^n$ such that

$$P(X_{t_1}\in F_1,\dots,X_{t_n}\in F_n)=\nu_{t_1,\dots t_k}(F_1\times\dots\times F_n)$$

for all $t_1,\dots t_n\in\mathbf{T}$.

10.2.2 Constructive

With the transition kernel (10.3) we find that

$$P\left(W_s\in dx, W_{\frac{1}{2}(t+s)}\in dy, W_t\in dz\right)=p(s;0,x)p\left(\frac{t-s}{2};x,y\right)p\left(\frac{t-s}{2};y,z\right)dxdydz,$$

and by dividing by

$$P\left(W_s\in dx, W_t\in dz\right)=p(s;0,x)p\left(t-s;x,z\right)dxdz$$

we conclude further (and informally) that

$$P\left(W_{\frac{1}{2}(t+s)}\in dy\,\middle|\, W_s=x,\ W_t=z\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy,\tag{10.5}$$

where $\mu:=\frac{1}{2}(x+z)$ and $\sigma^2=\frac{1}{4}(t-s)$.

Formula (10.5) suggests that we can construct a Brownian motion by interpolation.

Haar functions. For $n\ge 1$ and $k\in I(n):=\{i\in\mathbb{N}\colon i \text{ odd and } i\le 2^n\}$ ($I(0)=\{1\}$, $I(1)=\{1\}$, $I(2)=\{1,3\}$ etc.) define the Haar functions on $[0,1]$, i.e.,

$$H_1^{(0)}(t):=1\quad\text{and}\quad H_k^{(n)}(t):=\begin{cases}2^{(n-1)/2} & \text{if } \frac{k-1}{2^n}\le t<\frac{k}{2^n}\\ -2^{(n-1)/2} & \text{if } \frac{k}{2^n}\le t<\frac{k+1}{2^n}\\ 0 & \text{else.}\end{cases}\tag{10.6}$$

Schauder functions

$$S_k^{(n)}(t):=\int_0^t H_k^{(n)}(u)du.$$

Further, consider the stochastic process

$$B_t^{(n)}(\omega):=\sum_{m=0}^n\sum_{k\in I(m)}\xi_k^{(m)}S_k^{(m)}(t),$$

where $\xi_k^{(n)}\sim\mathcal{N}(0,1)$ are all independent standard normals.

Lemma 10.5. *The sequence of functions $t\to W_t^{(n)}(\omega)$ converges uniformly, as $n\to\infty$, to a continuous function for almost every $\omega\in\Omega$.*

Theorem 10.6. *The process $W_t:=\lim_{n\to\infty}W_t^{(n)}$ is a Brownian motion on $[0,1]$ (cf. Figure 10.2 for illustration).*

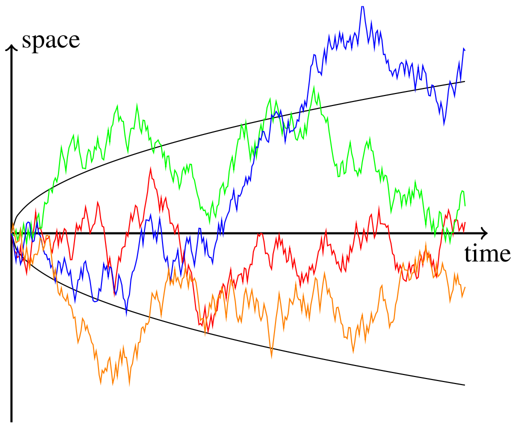


Figure 10.2: 4 Brownian paths

10.3 THE CANONICAL SPACE FOR BROWNIAN MOTION AND BASIC PROPERTIES

Theorem 10.7 (Kolmogorov's continuity theorem, often also Kolmogorov–Čentsov theorem)**.** *Suppose the process X satisfies the condition $\mathbb{E}\,|X_t - X_s|^\alpha \le D\,|t-s|^{\beta+1}$ for some $D>0$, $\alpha>0$ and $\beta>0$ and all $s, t \le T$. Then there is a continuous version, i.e., a stochastic process $\tilde{X}$ so that $P\Big(\big\{\omega\colon X_t(\omega) = \tilde{X}(\omega)\big\}\Big) = 1$ for all t and $P\Big(t \mapsto \tilde{X}_t \text{ is continuous}\Big) = 1$.*

Proof. Here is a nice proof: link. □

For the Brownian motion in $\mathbb{R}^n$ it holds that $\mathbb{E}\,|W_t - W_s|^4 = 3n^2\,|t-s|^2$, so the theorem applies.

The canonical space for the Brownian motion is, the one most convenient for many future developments, is

$$\Omega := C\big([0,\infty)\big), \tag{10.7}$$

the space of all continuous, $\mathbb{R}$-valued functions on $[0,\infty)$ with metric

$$d(\omega_1,\omega_2) := \sum_{n=1}^\infty \frac{1}{2^n} \max_{0\le t\le n} |\omega_1(t) - \omega_2(t)| \wedge 1, \tag{10.8}$$

(cf. Karatzas and Shreve [10, Section 2.4]).

The sigma algebra considered on Ω is the sigma algebra $\mathcal{F}$ generated by the finite dimensional cylinder sets

$$\big\{\omega \in C\big([0,\infty)\big)\colon\ (\omega(t_1),\dots,\omega(t_n)) \in A\big\}, \qquad n \in \{1,2,\dots\},\ A \in \mathcal{B}(\mathbb{R}^n)$$

(the Borel sets). The triple $(\Omega,\mathcal{F},P)$ with $\Omega = C\big([0,\infty)\big)$ is called the *canonical space*. Note, however, that the set (10.7) is not measurable with respect to the Borel sigma algebra on in $(\mathbb{R}^n)^{[0,\infty)}$.

10.4 PROBLEMS

Exercise 10.1. *Simulate a few Brownian paths.*

Exercise 10.2. *Verify (10.4).*

Exercise 10.3. *Prove Lemma 10.3.*

Exercise 10.4. *Show that the Haar functions $H_k^{(n)}(\cdot)$ (cf. (10.6)) are a complete orthonormal system for $L^2([0,1])$ with $\left\|H_k^{(n)}\right\|_2 = 1$, i.e., $\int_0^1 H_k^{(n)}(x) H_{k'}^{(n')}(x)dx = \begin{cases} 1 & \text{if } k = k' \text{ and } n = n' \\ 0 & \text{else} \end{cases}$.*

Exercise 10.5. *Show that $d(\cdot,\cdot)$ is a metric on $C\big([0,\infty)\big)$ and further, $\big(C\big([0,\infty)\big),\, d\big)$ is Polish — a complete, separable, metric space.*

Exercise 10.6. *Verify the covariance relation (10.2) by using the transition probabilities (10.3) explicitly.*

Exercise 10.7. *Show that* $\operatorname{cov}\left(W_s^2, W_t^2\right) = 2\ \operatorname{cov}\left(W_s, W_t\right)^2 = 2\,(s \wedge t)^2$. *In particular,* $\operatorname{var} W_t^2 = 2t^2$.

11 Ito Integral

This section introduces Itō's calculus[1] and again follows Øksendal [16].

11.1 INFORMAL DISCUSSION

Lemma 11.1 (Quadratic variation)**.** *The integrated, squared increments recover the time expired:*

$$\sum_{i=1}^{n}\left(W_{t_i}-W_{t_{i-1}}\right)^2 \xrightarrow[\max_j \Delta t_j\to 0]{} t, \qquad \textit{a deterministic number!!!},$$

where $\Delta t := \max\{t_i - t_{i-1} : i = 1,\dots n\}$ *for the partition* $0 = t_0 < t_2 < \dots < t_n = t$.

Remark 11.2. The statement is immediate for $\Delta W_t = \frac{1}{\sqrt{n}}\xi_i$, where $P(\xi_i = \pm 1) = \frac{1}{2}$.

Proof. Let $0 = t_0 < t_2 < \dots < t_n = t$ be a partition of $[0, t]$.

$$\mathbb{E}\sum_{i=1}^{n}\left(W_{t_i}-W_{t_{i-1}}\right)^2 = \sum_{i=1}^{n}\operatorname{var}\left(W_{t_i}-W_{t_{i-1}}\right) = \sum_{i=1}^{n} t_i - t_{i-1} = t$$

and

$$\operatorname{var}\left(\sum_{i=1}^{n}\left(W_{t_i}-W_{t_{i-1}}\right)^2\right) = \sum_{i=1}^{n}\operatorname{var}\left(\left(W_{t_i}-W_{t_{i-1}}\right)^2\right) = 2\sum_{i=1}^{n}(t_i-t_{i-1})^2 \xrightarrow[\max\Delta t_j\to 0]{} 0,$$

as the partition gets finer; we have used that $\operatorname{var}(X^2) = 2\sigma^4$ for $X \sim \mathcal{N}(0,\sigma^2)$, cf. (5.13). □

Recall that we have from Taylor expansions

(i) $df(x) = f'(x)\,dx$, for example $dt^2 = 2t\,dt$;

(ii) further, informally, $(dt)^2 = 0$.

For the Wiener Process, however,

(iii) $(dW_t)^2 = dt$,

as the preceding lemma demonstrates. Indeed,

$$\sum_{i=1}^{n}\left(W_{t_i}-W_{t_{i-1}}\right)^2 \to \int_0^T (dW_t)^2 = T \text{ (a real number!) and}$$

$$T = \int_0^T dt$$

for all T, whenever the discretization $0 = t_0 < t_2 < \dots < t_n = T$ gets tighter. Thus, informally, $(dW_t)^2 = dt$, the assertion in (iii).

[1] Kiyosi Itō, 1915–2008

11.2 ITO'S INTEGRAL

Example 11.3. Consider the functions

$$\phi_1^{(n)}(t,\omega) := \sum_{j\geq 0} W_{j\cdot 2^{-n}}(\omega)\, \mathbb{1}_{[j\cdot 2^{-n},(j+1)2^{-n})}(t) \text{ and}$$
$$\phi_2^{(n)}(t,\omega) := \sum_{j\geq 0} W_{(j+1)\cdot 2^{-n}}(\omega)\, \mathbb{1}_{[j\cdot 2^{-n},(j+1)2^{-n})}(t).$$

Then it holds (put $t_j := j\cdot 2^{-n}$) that

$$\mathbb{E}\int_0^T \phi_1^{(n)}(t,\omega)\, dW_t(\omega) = \sum_{j\geq 0} \mathbb{E}\, W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) = 0$$

as W_t is independent of later increments. But

$$\mathbb{E}\int_0^T \phi_2^{(n)}(t,\omega)\, dW_t(\omega) = \sum_{j\geq 0} \mathbb{E}\, W_{t_{j+1}}\left(W_{t_{j+1}} - W_{t_j}\right) = \sum_{j\geq 0} t_{j+1} - t_j = T.$$

Hence,

$$\int_0^T \phi_1^{(n)}(t,\omega)\, dW_t(\omega) \not\to \int_0^T \phi_2^{(n)}(t,\omega) dW_t(\omega)$$

as $n \to \infty$.

The choice of the left or right endpoint in the integral makes a difference here. This is in contrast to the usual Riemann–Stieltjes integral, but for these the integrator has bounded variation (which $t \to W_t(\omega)$ lacks almost surely).

Definition 11.4. We consider the following class of functions (processes) $f \in \mathcal{V}$, where

$$f\colon [0,\infty)\times\Omega \to \mathbb{R}$$

(or $f \in \mathcal{V}(S,T)$, for functions $f\colon [S,T)\times\Omega\to\mathbb{R}$) such that

(i) $(t,\omega)\mapsto f(t,\omega)$ is $\mathcal{B}\otimes\mathcal{F}$-measurable ($\mathcal{B}$ is the Borel sigma algebra on $[0,\infty)$),

(ii) $\omega \mapsto f(t,\omega)$ is $\mathcal{F}_t$ adapted for every t fixed and

(iii) $\mathbb{E}\int_S^T f(t,\omega)^2 dt < \infty$.

Note that the functions in Example 11.3 satisfy $\phi_1^{(n)} \in \mathcal{V}$, but $\phi_2^{(n)} \notin \mathcal{V}$.

Definition 11.5. A function $\phi \in \mathcal{V}$ is *elementary*, if $\phi(t,\omega) = \sum_{j=0}^{n-1} e_j(\omega)\cdot \mathbb{1}_{[t_j,t_{j+1})}(t)$[2] (with $S = t_0 < t_1 \cdots < t_n = T$). For an elementary function ϕ we define the Itō-integral as

$$\int_S^T \phi(t,\omega)\, dW_t(\omega) = \sum_{j=0}^{n-1} e_j(\omega)\left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right) \qquad \text{for every } \omega\in\Omega. \tag{11.1}$$

Note, that e_j is necessarily $\mathcal{F}_t$-measurable provided that $\phi \in \mathcal{V}$.

[2]Note that $\phi(\cdot,\omega)$ is càdlàg, cf. Footnote 9 on page 49.

Lemma 11.6 (Itō isometry for elementary functions)**.** *If $\phi(t,\omega)$ is bounded and elementary then*

$$\mathbb{E}\left(\int_S^T \phi(t,\omega)\,dW_t(\omega)\right)^2 = \mathbb{E}\left(\int_S^T \phi(t,\omega)^2\,dt\right).$$

Proof. Note first that

$$\mathbb{E}\,e_i e_j\left(W_{t_{i+1}} - W_{t_i}\right)\left(W_{t_{j+1}} - W_{t_j}\right) = \begin{cases} 0 & \text{if } i \neq j \\ \mathbb{E}\,e_i^2(t_{i+1} - t_i) & \text{if } i = j \end{cases}$$

and hence

$$\begin{aligned}\mathbb{E}\left(\int_S^T \phi\,dW\right)^2 &= \sum_{i,j} \mathbb{E}\,e_i e_j\left(W_{t_{i+1}} - W_{t_i}\right)\left(W_{t_{j+1}} - W_{t_j}\right) = \sum_i \mathbb{E}\,e_i^2(t_{i+1} - t_i)\\ &= \mathbb{E}\sum_i e_i^2(t_{i+1} - t_i) = \mathbb{E}\left(\int_S^T \phi(t,\omega)^2\,dt\right),\end{aligned}$$

the assertion. □

The functions in $f \in \mathcal{V}$ can be approximated by elementary functions $\phi \in \mathcal{V}$ in the norm $\mathbb{E}\int_S^T \left(f(t,\omega) - \phi(t,\omega)\right)^2\,dt$.

Definition 11.7 (Itō's integral for functions in $\mathcal{V}$)**.** For $f \in \mathcal{V}$ we define

$$\int_S^T f(t,\omega)\,dW_t(\omega) \coloneqq \lim_{n\to\infty}\int_S^T \phi_n(t,\omega)\,dW_t(\omega) \qquad \text{in } L^2(P),$$

where $\phi_n(t,\omega)$ is a sequence of elementary functions such that $\mathbb{E}\int_S^T \left(f(t,\omega) - \phi_n(t,\omega)\right)^2\,dt \to 0$.

Itō's integral $\int_S^T \phi\,dW$ defined for general elementary functions in (11.1) can be extended to bounded and continuous functions first, then to bounded functions in $\mathcal{V}$ and finally to all functions in $\mathcal{V}$.

Corollary 11.8 (to Lemma 11.6, Itō isometry)**.** *It holds that*

$$\mathbb{E}\left(\int_S^T f(t,\omega)\,dW_t(\omega)\right)^2 = \mathbb{E}\left(\int_S^T f(t,\omega)^2\,dt\right) \tag{11.2}$$

for $f \in \mathcal{V}$.

Corollary 11.9. *For $f \in \mathcal{V}$ and a sequence $f_n \in \mathcal{V}$ with $\mathbb{E}\left(\int_S^T f_n(t,\omega)^2\,dt\right) \to \mathbb{E}\left(\int_S^T f(t,\omega)^2\,dt\right)$ it holds that*

$$\int_S^T f_n(t,\omega)\,dW_t(\omega) \to \int_S^T f(t,\omega)\,dW_t(\omega) \qquad \textit{in } L^2.$$

Example 11.10 (Cf. Exercise 11.3)**.** It holds that

$$\int_0^t W_s\,dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t. \tag{11.3}$$

Proof. Put $\phi_n(t,\omega) \coloneqq \sum_j W_{t_j}(\omega)\cdot \mathbb{1}_{[t_j,t_{j+1})}(t)$. Then

$$\mathbb{E}\int_0^t (\phi_n - W_s)^2\, ds = \mathbb{E}\sum_j \int_{t_j}^{t_{j+1}} \left(W_{t_j} - W_s\right)^2 ds = \sum_j \int_{t_j}^{t_{j+1}} (s-t_j) ds = \sum_j \frac{1}{2}(t_{j+1}-t_j)^2 \to 0,$$

and by Itō's isometry (Lemma 11.6), thus

$$\int_0^t \phi_n dW_s \xrightarrow[\max \Delta t_j \to 0]{} \int_0^t W_s dW_s.$$

To verify (11.3) in L^2 we need to verify that

$$\mathbb{E}\left(\frac{1}{2}W_t^2 - \frac{t}{2} - \sum_j W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right)\right)^2 \xrightarrow[\max_j \Delta t_j \to 0]{} 0.$$

To this end,

$$\mathbb{E}\left(\frac{1}{2}W_t^2 - \frac{t}{2} - \sum_j W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right)\right)^2 = \mathbb{E}\begin{pmatrix} \frac{1}{4}W_t^4 + \frac{t^2}{4} + \left(\sum_j W_{t_j}\Delta W_{t_j}\right)^2 \\ -\frac{t}{2}W_t^2 - W_t^2 \sum_j W_{t_j}\Delta W_{t_j} + t\cdot \sum_j W_{t_j}\Delta W_{t_j} \end{pmatrix} \tag{11.4}$$

$$= \frac{3}{4}t^2 + \frac{1}{4}t^2 +$$

$$+\,\mathbb{E}\left(\sum_j W_{t_j}\Delta W_{t_j}\right)^2 \tag{11.5}$$

$$-\,\frac{t^2}{2}$$

$$-\,\mathbb{E}\sum_j W_{t_j}\Delta W_{t_j} W_t^2 \tag{11.6}$$

$$+\,0.$$

Then

$$(11.5) = \mathbb{E}\left(\sum_j W_{t_j}\Delta W_{t_j}\right)^2 = \sum_{i,j} \mathbb{E}\, W_{t_i} W_{t_j} \Delta W_{t_i} \Delta W_{t_j} = \sum_i \mathbb{E}\, W_{t_i}^2 \left(\Delta W_{t_i}\right)^2 = \sum_i t_i \Delta t_i$$

and (recall that $t_j < t_{j+1} < t$)

$$\begin{aligned} -(11.6) &= \sum_j \mathbb{E}\, W_{t_j}\Delta W_{t_j} W_t^2 = \sum_j \mathbb{E}\, W_{t_j}\Delta W_{t_j}\left(W_{t_j} + \Delta W_{t_j} + (W_t - W_{t_{j+1}})\right)^2 \\ &= \sum_j \mathbb{E}\, W_{t_j}^3 \underbrace{\Delta W_{t_j}}_{0} + \sum_j \mathbb{E}\, W_{t_j} \underbrace{\left(\Delta W_{t_j}\right)^3}_{0} + \sum_j \mathbb{E}\, W_{t_j}\Delta W_{t_j} \underbrace{\left(W_t - W_{t_{j+1}}\right)^2}_{0} \\ &\quad + 2\sum_j \mathbb{E}\, W_{t_j}^2 \left(\Delta W_{t_j}\right)^2 + 2\sum_j \mathbb{E}\, W_{t_j}^2 \Delta W_{t_j} \underbrace{\left(W_t - W_{t_{j+1}}\right)}_{0} + 2\sum_j \mathbb{E}\, W_{t_j}\left(\Delta W_{t_j}\right)^2 \underbrace{(W_t - W_{t_{j+1}})}_{0} \\ &= 2\sum_j \mathbb{E}\, W_{t_j}^2 \left(\Delta W_{t_j}\right)^2 = 2\sum_j t_j \left(t_{j+1} - t_j\right). \end{aligned}$$

Collecting terms,

$$(11.4) = \frac{1}{2}t^2 - \sum_j t_j(t_{j+1} - t_j) \xrightarrow[\max \Delta t_j \to 0]{} \frac{1}{2}t^2 - \int_0^t s ds = \frac{1}{2}t^2 - \frac{1}{2}t^2 = 0$$

and hence the result. □

11.3 PROPERTIES OF ITO'S INTEGRAL

Theorem 11.11. *For $c \in \mathbb{R}$ and $f, g \in \mathcal{V}$ it holds that (as for the usual integral) that*

(i) $\int_S^T f dW_t = \int_S^U f dW_t + \int_U^T f dW_t$,

(ii) $\int_S^T cf + g dW_t = c \int_S^T f dW_t + \int_S^T g dW_t$.

11.4 THE MARTINGALE PROPERTY

Theorem 11.12. *Itō's integral $M_t(\omega) \coloneqq \int_0^t f(s,\omega) dW_s(\omega)$ is a martingale.*

Proof. Let ϕ_n be a simple function and $t < s$. Then

$$\begin{aligned}\mathbb{E}\left(\int_0^s \phi_n dW \middle| \mathcal{F}_t\right) &= \mathbb{E}\left(\int_0^t \phi_n dW + \int_t^s \phi_n dW \middle| \mathcal{F}_t\right)\\ &= \mathbb{E}\left(\int_0^t \phi_n dW \middle| \mathcal{F}_t\right) + \mathbb{E}\left(\int_t^s \phi_n dW \middle| \mathcal{F}_t\right) = \int_0^t \phi_n dW,\end{aligned}$$

as $\int_0^t \phi_n dW$ is $\mathcal{F}_t$ measurable and $\Delta W_j = W_{t_{j+1}} - W_{t_j}$ is independent from $\mathcal{F}_t$ whenever $t_j \geq t$. □

Theorem 11.13. *The martingale $M_t(\omega) \coloneqq \int_0^t f(s,\omega) dW_s(\omega)$ satisfies Doob's martingale inequality and it holds that*

$$P\left(\sup_{s \leq t} |M_s| \geq \lambda\right) \leq \frac{1}{\lambda^2} \mathbb{E} \int_0^t f(s,\omega)^2 ds.$$

Proof. The result combines Doob's martingale inequality (Theorem 8.23) and Itō's isometry, Corollary 11.8. □

11.5 PROBLEMS

Exercise 11.1. *Use simulations to verify that $\sum_{i=1}^n \left(W_{t_i} - W_{t_{i-1}}\right)^2 \sim \mathcal{N}\left(T, 2\sum_{i=1}^n (t_i - t_{i-1})^2\right) \xrightarrow{\mathcal{D}} T$.*

Exercise 11.2. *Show that $\mathbb{E}\left(W_t^3 \middle| \mathcal{F}_s\right) = W_s^3 + (s-t)W_s$, so that W_t^3 is not a martingale.*

Exercise 11.3 (Integration by parts)**.** *Show that $\int_0^t s dB_s = tB_t - \int_0^t B_s ds$. Discuss why integration by parts holds true here, but not in (11.3).*

12 Stochastic Calculus

12.1 ITO'S LEMMA

Itô's Lemma is the Taylor series extension for stochastic processes. To this end consider the *stochastic differential equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{12.1}$$

However, we first have to give a meaning to (12.1).

Definition 12.1 (Strong solution)**.** We say that (12.1) holds iff

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s,\ 0 \le t < \infty.$$

X_t is called an Itō-process.

Lemma 12.2 (Itō's lemma, also Itō–Döblin[1] lemma)**.** *Suppose that X_t satisfies (12.1) and g is regular (smooth) enough. Then the process*

$$Y_t \coloneqq g(t, X_t)$$

satisfies

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)\,(dX_t)^2\,,$$

where $dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0$, $dW_t \cdot dW_t = dt$ (cf. Table 12.1). Or expanded using (12.1),

$$dY_t = \left(\frac{\partial g}{\partial t}(t, X_t) + b(t, X_t)\frac{\partial g}{\partial x}(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2\frac{\partial^2 g}{\partial x^2}(t, X_t)\right)dt + \sigma(t, X_t)\frac{\partial g}{\partial x}(t, X_t)dW_t. \tag{12.2}$$

Remark 12.3. Note that by Definition 12.1 the equation (12.2) reads

$$g(t, X_t) = g(0, X_0) + \int_0^t \left(\frac{\partial g}{\partial t} + b\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2}\right)ds + \underbrace{\int_0^t \sigma\frac{\partial g}{\partial x}dW_s}_{\text{Martingale}}. \tag{12.3}$$

Informal proof. By a usual Taylor-series expansion for the smooth function $f(\cdot, :)$ we have that

$$\begin{aligned}
g(t + dt, x + dx) =& g(t, x)\\
&+ \frac{\partial}{\partial t}g(t, x)dt + \frac{\partial}{\partial x}g(t, x)dx\\
&+ \frac{1}{2}\left(\frac{\partial^2}{\partial t^2}g(t, x)(dt)^2 + 2\frac{\partial^2}{\partial t\partial x}g(t, x)dtdx + \frac{\partial^2}{\partial x^2}g(t, x)(dx)^2\right)\\
&+ o\left(dt^2 + dx^2\right).
\end{aligned}$$

[1] Wolfgang Döblin, 1915–1940

It follows that

$$\begin{aligned} dY_t &= g(t+dt, X_t+dX_t) - g(t, X_t) \\ &= \frac{\partial}{\partial t} g(t, X_t) dt + \frac{\partial}{\partial x} g(t, X_t) dX_t \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2} g(t, X_t)(dX_t)^2 + o\left((dt)^2 + dt dX_t + (dX_t)^2\right) \end{aligned}$$

By plugging in the formula (12.1) we obtain

$$\begin{aligned} dY_t =& \frac{\partial}{\partial t} g(t, X_t) dt \\ &+ \frac{\partial}{\partial x} g(t, X_t)\left(b(t, X_t)dt + \sigma(t, X_t) dW_t\right) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial x^2} g(t, X_t)\left(b(t, X_t)dt + \sigma(t, X_t) dW_t\right)^2 + o\left((dt)^2 + dt dX_t + (dX_t)^2\right) \end{aligned}$$

and after collecting terms thus

$$dY_t = \left(\frac{\partial}{\partial t} g(t, X_t) + b(t, X_t)\frac{\partial}{\partial x} g(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2 \frac{\partial^2}{\partial x^2} g(t, X_t)\right) dt + \underbrace{\sigma(t, X_t)\frac{\partial}{\partial x} g(t, X_t) dW_t}_{\text{martingale}},$$

the desired result. □

Sketch of the proof. By using Taylor's expansion we have that

$$g(t, X_t) = g(0, X_0) + \sum_j \underbrace{g(t_{j+1}, X_{t_{j+1}}) - g(t_j, X_j)}_{\Delta g(t_j, X_j)}$$

$$= g(0, X_0)$$

$$+ \sum_j \frac{\partial}{\partial t} g(t_j, X_{t_j}) \cdot \underbrace{(t_{j+1} - t_j)}_{=:\Delta t_j} \tag{12.4}$$

$$+ \frac{\partial}{\partial x} g(t_j, X_{t_j}) \cdot \underbrace{\left(X_{t_{j+1}} - X_{t_j}\right)}_{=:\Delta X_{t_j}} \tag{12.5}$$

$$+ \frac{1}{2}\sum_j \frac{\partial^2}{\partial t^2} g(t_j, X_{t_j}) \cdot (\Delta t_j)^2 \tag{12.6}$$

$$+ \frac{1}{2} 2 \sum_j \frac{\partial^2}{\partial t \partial x} g(t_j, X_{t_j}) \cdot \Delta t_j \Delta X_j \tag{12.7}$$

$$+ \frac{1}{2}\sum_j \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(\Delta X_{t_j}\right)^2 \tag{12.8}$$

$$+ o\left((\Delta t_j)^2 + \left(\Delta X_{t_j}\right)^2 + \Delta t_j \Delta X_{t_j}\right).$$

For the first order terms we have

$$(12.4) = \sum_j \frac{\partial g}{\partial t}\left(t_j, X_{t_j}\right) \cdot \left(t_{j+1} - t_j\right) \xrightarrow[\Delta t_j \to 0]{} \int_0^t \frac{\partial g}{\partial t}(s, X_s)\, ds,$$

and

$$(12.5) = \sum_j \frac{\partial g}{\partial x}\left(t_j, X_{t_j}\right) \cdot \left(X_{t_{j+1}} - X_{t_j}\right) \xrightarrow[\Delta t_j \to 0]{} \int_0^t \frac{\partial g}{\partial x}(s, X_s)\, dX_s.$$

The second order terms:

(i) $dt^2 = 0$: it is evident that

$$(12.6) = \frac{1}{2}\sum_j \frac{\partial^2}{\partial t^2} g(t_j, X_{t_j}) \cdot (\Delta t_j)^2 \xrightarrow[\Delta t_j \to 0]{} 0$$

(ii) $dt \cdot dW = 0$: and

$$\begin{aligned}(12.7) = \mathbb{E}\left(\sum_j \frac{\partial^2}{\partial t \partial x} g(t_j, X_{t_j}) \Delta t_j \Delta X_j\right)^2 &= \sum_{i,j} \mathbb{E}\, \frac{\partial^2}{\partial t \partial x} g(t_i, X_{t_i}) \frac{\partial^2}{\partial t \partial x} g(t_j, X_{t_j}) \Delta t_i \Delta t_j \Delta X_j \Delta X_i \\ &= \sum_i \mathbb{E}\left(\frac{\partial^2}{\partial t \partial x} g(t_j, X_{t_j}) \Delta X_i\right)^2 \cdot (\Delta t_i)^2 \xrightarrow[\Delta t_j \to 0]{} 0,\end{aligned}$$

which is convergence in L^2.

(iii) $(dW)^2 = 0$: finally, by using (12.1),

$$\begin{aligned}2 \cdot (12.8) &= \sum_j \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(\Delta X_{t_j}\right)^2 = \sum_j \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(b_j \Delta t_j + \sigma_j \Delta W_{t_j}\right)^2 \\ &= \sum_j \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(b_j^2 \left(\Delta t_j\right)^2 + 2 b_j \sigma_j \Delta t_j \Delta W_{t_j} + \sigma_j^2 \left(\Delta W_{t_j}\right)^2\right) \\ &= \sum_j b^2(t_j, X_{t_j}) \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(\Delta t_j\right)^2 \qquad (12.9) \\ &\quad + 2\sum_j b(t_j, X_{t_j}) \sigma(t_j, X_{t_j}) \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \Delta t_j \Delta W_{t_j} \qquad (12.10) \\ &\quad + \sum_j \sigma^2(t_j, X_{t_j}) \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(\Delta W_{t_j}\right)^2. \qquad (12.11)\end{aligned}$$

Now (12.9) $\xrightarrow[\Delta t_j \to 0]{}$ 0 in the same way as (12.6) above and (12.10) $\xrightarrow[\Delta t_j \to 0]{}$ 0 as (12.7).

It remains to verify that

$$(12.11) = \sum_j \sigma^2(t_j, X_{t_j}) \frac{\partial^2}{\partial x^2} g(t_j, X_{t_j}) \cdot \left(\Delta W_{t_j}\right)^2 \xrightarrow[\Delta t_j \to 0]{} \int_0^t \sigma^2(s, X_s) \frac{\partial^2}{\partial x^2} g(s, X_s) ds \qquad \text{in } L^2.$$

To this end define $\alpha(t,x) \coloneqq \sigma^2(t,x)\frac{\partial^2}{\partial x^2}g(t,x)$ with $\alpha_t \coloneqq \alpha(t,X_t)$. Then the equation in the latter display rewrites as

$$\mathbb{E}\left(\sum_j \alpha_{t_j}\cdot\left(\Delta W_{t_j}\right)^2 - \sum_i \alpha_{t_i}\Delta t_i\right)^2 = \mathbb{E}\left(\sum_j \alpha_{t_j}\cdot\left(\left(\Delta W_{t_j}\right)^2 - \Delta t_j\right)\right)^2$$
$$= \sum_{i,j}\mathbb{E}\,\alpha_{t_i}\alpha_{t_j}\left(\left(\Delta W_{t_i}\right)^2 - \Delta t_i\right)\cdot\left(\left(\Delta W_{t_j}\right)^2 - \Delta t_j\right). \quad (12.12)$$

The summands vanish for $i \neq j$ by independence and as $\mathbb{E}\,(\Delta W_t)^2 = \Delta t$, and thus

$$(12.12) = \sum_i \mathbb{E}\,\alpha_{t_i}^2\left(\left(\Delta W_{t_i}\right)^2 - \Delta t_i\right)^2 = \sum_i \mathbb{E}\,\alpha_{t_i}^2\left(\left(\Delta W_{t_i}\right)^4 - 2\Delta t_i \Delta W_{t_i} + \left(\Delta t_j\right)^2\right)$$
$$= \sum_i \mathbb{E}\,\alpha_{t_i}^2\left(3\,(\Delta t_i)^2 - 2\,(\Delta t_i)^2 + \left(\Delta t_j\right)^2\right) = 2\sum_i \mathbb{E}\,\alpha_{t_i}^2\,(\Delta t_i)^2 \xrightarrow[\Delta t_j\to 0]{} 0,$$

by (5.13), as desired.

□

12.2 APPLICATIONS OF ITO'S LEMMA

In what follows we shall typically consider the process $X_t = W_t$, i.e.,

$$dW_t = \underbrace{0}_{b}\,dt + \underbrace{1}_{\sigma}\,dW_t$$

and apply (12.2). Following (12.2), the stochastic process

$$Y_t \coloneqq g(t,W_t)$$

satisfies the differential equation

$$dY_t = \frac{\partial g}{\partial t}(t,W_t)dt + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,W_t)dt + \frac{\partial g}{\partial x}(t,W_t)dW_t. \quad (12.13)$$

Example 12.4. Let $g(t,x) = \frac{1}{2}x^2$ and $b = 0$, $\sigma = 1$, then $X_t = W_t$ and

$$\frac{1}{2}W_t^2 = g(t,X_t) = W_0 + \int_0^t 0 + 0\cdot\frac{\partial g}{\partial x} + \frac{1}{2}ds + \int_0^t W_s dW_s = W_0 + \frac{t}{2} + \int_0^t W_s dW_s.$$

This is the result obtained in (11.3).

Theorem 12.5 (Integration by parts)**.** *The usual formula (it is essential that f does* not *depend on ω)*

$$W_t\cdot f(t) = \int_0^t W_s\,df(s) + \int_0^t f(s)\,dW_s.$$

Proof. Let $g(t,x) = x\cdot f(t)$. Then $\frac{\partial g}{\partial t} = xf'(t)$, $\frac{\partial g}{\partial x} = f$ and $\frac{\partial^2 g}{\partial x^2} = 0$. Then, by (12.3) with $b = 0$ and $\sigma = 1$,

$$W_t\cdot f(t) = 0 + \int_0^t W_s f'(s)\,ds + \int_0^t f(s)d\,W_s,$$

from which the assertion follows. □

Example 12.6 (Geometric Brownian motion)**.** Consider the process $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right)$. Then it holds that

$$dS_t = \mu S_t dt + \sigma S_t \, dW_t.$$

Indeed, choose $g(t,x) \coloneqq \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma x\right)$ and apply (12.2) to the process

$$dW_t = \underbrace{0}_{b} dt + \underbrace{1}_{\sigma} dW_t.$$

12.3 FURTHER MARTINGALES

Lemma 12.7. *For g smooth enough define* $a(t,x) \coloneqq \left(\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right) g(t,x)$*, then*

$$M_t \coloneqq g(t, W_t) - \int_0^t a(s, W_s) ds$$

is a martingale.

Proof. Put $b = 0$ and $\sigma = 1$ in (12.3), then

$$g(t, W_t) = g(0, W_0) + \int_0^t \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right) ds + \int_0^t \frac{\partial g}{\partial x} dW_s,$$

thus

$$g(t, W_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right) ds = g(0, W_0) + \int_0^t \frac{\partial g}{\partial x} dW_s$$

and the right hand side is a martingale by Theorem 11.12. Thus the result. □

Corollary 12.8. *If* $\left(\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right) g(t,x) = 0$*, then*

$$M_t \coloneqq g(t, W_t)$$

is a martingale.

Example 12.9 (Wald's martingale)**.** The process

$$M_t \coloneqq \exp\left(\sigma \, W_t - \frac{1}{2}\sigma^2 \, t\right) \tag{12.14}$$

is a martingale.

Theorem 12.10 (Novikov-Condition)**.** *Suppose that* $\mathbb{E} \exp\left(\frac{1}{2}\int_0^t \sigma(u)^2 du\right) < \infty$*, then the stochastic process*

$$M_t \coloneqq \exp\left(\int_0^t \sigma(u) dW_u - \frac{1}{2}\int_0^t \sigma(u)^2 du\right) \tag{12.15}$$

is a martingale. The process M_t *is called* stochastic exponential *or* Doléans-Dade exponential *of the process* $\int_0^t \sigma(u) dW_u$.

Proof. For $s < t$ it holds that

$$\begin{aligned}\mathbb{E}\left(\exp\left(\int_0^t \sigma(u)dW_u\right)\middle|\mathcal{F}_s\right) &= \mathbb{E}\left(\exp\left(\int_0^s \sigma(u)dW_u\right)\cdot\exp\left(\int_s^t \sigma(u)dW_u\right)\middle|\mathcal{F}_s\right)\\ &= \exp\left(\int_0^s \sigma(u)dW_u\right)\cdot\mathbb{E}\left(\exp\left(\int_s^t \sigma(u)dW_u\right)\middle|\mathcal{F}_s\right)\\ &= \exp\left(\int_0^s \sigma(u)dW_u\right)\cdot\mathbb{E}\exp\left(\int_s^t \sigma(u)dW_u\right),\end{aligned}$$

as the increments dW_u do not depend on $\mathcal{F}_s$ for $u \geq s$. Recall now from Proposition **??** that $\int_s^t \sigma(u)dW_u \sim \mathcal{N}\left(0, \int_s^t \sigma(u)^2 du\right)$ is normally distributed with mean 0. The expected value of the log-normal random variable follows from (5.18) (with $\mu = 0$ and $n = 1$) and

$$\mathbb{E}\left(\exp\left(\int_0^t \sigma(u)dW_u\right)\middle|\mathcal{F}_s\right) = \exp\left(\int_0^s \sigma(u)dW_u\right)\cdot\exp\left(\frac{1}{2}\int_s^t \sigma(u)^2 du\right).$$

Multiply with $\exp\left(-\frac{1}{2}\int_0^t \sigma(u)^2 du\right)$ and it follows that

$$\begin{aligned}&\mathbb{E}\left(\underbrace{\exp\left(\int_0^t \sigma(u)dW_u - \frac{1}{2}\int_0^t \sigma(u)^2 du\right)}_{M_t}\middle|\mathcal{F}_s\right)\\ &= \underbrace{\exp\left(\int_0^s \sigma(u)dW_u - \frac{1}{2}\int_0^s \sigma(u)^2 du\right)}_{M_s}\cdot\underbrace{\exp\left(\frac{1}{2}\int_s^t \sigma(u)^2 du - \frac{1}{2}\int_s^t \sigma(u)^2 du\right)}_{=1},\end{aligned}$$

which is the assertion. □

12.4 ITO'S LEMMA IN HIGHER DIMENSIONS

Theorem 12.11 (Multi-dimensional Itō formula)**.** *Let $Y(t,\omega) := g(t, X_t)$, where $dX_t = bdt + \sigma\cdot dW_t$, i.e.,*

$$\begin{pmatrix} dX_t^{(1)}\\ \vdots\\ dX_t^{(n)}\end{pmatrix} = \begin{pmatrix} b_1(t,X_t)\\ \vdots\\ b_n(t,X_t)\end{pmatrix}dt + \begin{pmatrix}\sigma_{1,1}(t,X_t) & \dots & \sigma_{1,m}(t,X_t)\\ \vdots & & \vdots\\ \sigma_{n,1}(t,X_t) & \dots & \sigma_{n,m}(t,X_t)\end{pmatrix}\cdot\begin{pmatrix} dW_t^{(1)}\\ \vdots\\ dW_t^{(m)}\end{pmatrix},$$

then

$$dY_k = \frac{\partial g_k}{\partial t}(t,X_t)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X_t)dX_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X_t)dX_i dX_j,$$

where $dW_t^{(i)}dW_t^{(t)} = \delta_{i,j}dt$, etc. Table 12.1 collects the rules for Itō's calculus in higher dimensions.

Proof. The proof of Itō's lemma (Lemma 12.2) applies, but it remains to be shown that $dW_t^{(1)}dW_t^{(2)} = 0$,

	dt	$dW_t^{(1)}$	$dW_t^{(2)}$
dt	0	0	0
$dW_t^{(1)}$	0	dt	0
$dW_t^{(2)}$	0	0	dt

Table 12.1: Itō calculus for Brownian motion including higher dimensions

i.e., we claim that $\int_0^t f dW_s^{(1)} dW_s^{(2)} = 0$. Let ϕ_n be simple. Then

$$\begin{aligned}\mathbb{E}\left(\int_0^t \phi_n dW_t^{(1)} dW_t^{(2)}\right)^2 &= \mathbb{E}\left(\sum_i e_{t_i}\left(W_{t_{i+1}}^{(1)} - W_{t_i}^{(1)}\right)\left(W_{t_{i+1}}^{(2)} - W_{t_i}^{(2)}\right)\right)^2 \\ &= \mathbb{E}\sum_{i,j} e_{t_i} e_{t_j}\left(W_{t_{i+1}}^{(1)} - W_{t_i}^{(1)}\right)\left(W_{t_{i+1}}^{(2)} - W_{t_i}^{(2)}\right)\left(W_{t_{j+1}}^{(1)} - W_{t_j}^{(1)}\right)\left(W_{t_{j+1}}^{(2)} - W_{t_j}^{(2)}\right) \\ &= \mathbb{E}\sum_i e_{t_i}^2\left(W_{t_{i+1}}^{(1)} - W_{t_i}^{(1)}\right)^2\left(W_{t_{i+1}}^{(2)} - W_{t_i}^{(2)}\right)^2 = \sum_i \mathbb{E}\, e_{t_i}^2\,(t_{i+1} - t_i)^2 \xrightarrow[\Delta t_j \to 0]{} 0,\end{aligned}$$

thus the assertion. □

12.5 PROBLEMS

Exercise 12.1. *Use $g(t,x) \coloneqq x^2 - t$ to show that $W_t^2 - t$ is a martingale (and $W_t^2 = t + 2\int_0^t W_s dW_s$).*

Example 12.12. Use $g(t,x) \coloneqq \left(x^2 - t\right)^2$ to show that $\left(W_t^2 - t\right)^2 - 4\int_0^1 W_s^2 ds$ is a martingale.

Exercise 12.2. *Verify that the exponential martingale (12.14) is a martingale for every $\sigma \in \mathbb{R}$ fixed.*

Example 12.13. Define the stochastic processes

$$Y_t \coloneqq \int_0^t \sigma(u) dW_u \text{ and } Z_t \coloneqq f(t, Y_t)$$

with $f(t,x) \coloneqq \exp\left(x - \frac{1}{2}\int_0^t \sigma(u)^2 du\right)$. Use Itō's formula to show that $dZ_t = Z_t \sigma_t dW_t$, i.e., $Z_t = \int_0^t Z_s \sigma_s dW_s$ is a martingale. Use this result to deduce (12.15).

Exercise 12.3 (Depreciation)**.** *Assume that the profit of an asset evolves according a geometric Brownian motion. The discounted value of the asset, which is in use up to time T, is $\mathbb{E}\int_0^T e^{-rt} e^{-(\mu - \sigma^2/2)t + \sigma B_t} dt = \int_0^T e^{-(r+\mu)t} = \frac{1}{r+\mu}\left(1 - e^{-(r+\mu)T}\right)$. Given that the time of a failure is random as well, with exponential distribution, then the time value of the asset is*

$$\mathbb{E}\int_0^\infty \lambda e^{-\lambda t}\int_0^t e^{-rt'} e^{-(\mu - \sigma^2/2)t' + \sigma B_{t'}} dt = \frac{1}{r+\mu} - \frac{\lambda}{(r+\mu)(r+\mu+\lambda)} = \frac{1}{r+\mu+\lambda}.$$

13 Stochastic Differential Equations

13.1 LINEAR EQUATION EXAMPLES

Definition 13.1. A (arithmetic) Brownian motion (ABM) with drift μ is the solution of the SDE $dS_t = \mu\, dt + \sigma\, dW_t$.

Lemma 13.2. *The explicit solution of the Brownian motion with drift is* $S_t = S_0 + \mu\, t + \sigma\, W_t$.

Proof. Note that $S_t = g(t, W_t)$ for the function $g(t,x) \coloneqq S_0 + \mu\, t + \sigma\, x$. The Employing (12.13) (Ito's lemma) we see that $dS_t = \mu\, dt + \sigma\, dW_t$, as required. □

Note, that the pdf of the marginal distribution S_t is $S_t \sim \mathcal{N}\left(S_0 + \mu\, t,\ t\sigma^2\right)$ so that

$$P\left(S_t \in dy\right) = \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{1}{2t\sigma^2}(y - S_0 - \mu t)^2} dy.$$

Definition 13.3. A geometric Brownian motion (GBM) with drift μ is the solution of the SDE $dS_t = \mu S_t dt + \sigma S_t dW_t$.

Lemma 13.4. *The explicit solution of the GBM is*

$$S_t = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t} = S_0\, e^{\mu\, t} \cdot \underbrace{\exp\left(\sigma\, W_t - \frac{1}{2}\sigma^2\, t\right)}_{\textit{exponential martingale, (12.14)}}.$$

The corresponding marginal pdf is

$$P\left(S_t \in dy\right) = \frac{1}{\sigma\, y\, \sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2 t}\left(\log \frac{y}{S_0} - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2} dy. \tag{13.1}$$

Proof. The Employing (12.13) to the function $g(t,x) \coloneqq S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\, x}$.

As for the marginal density consider the cdf

$$P(S_t \le y) = P\left(W_t \le \frac{1}{\sigma}\left(\ln \frac{y}{S_0} - \left(\mu - \frac{1}{2}\sigma^2\right)t\right)\right) = \Phi\left(\frac{1}{\sigma\sqrt{t}}\left(\ln \frac{y}{S_0} - \left(\mu - \frac{1}{2}\sigma^2\right)t\right)\right),$$

where Φ is the cdf of the standard normal distribution. The density (13.1) follows by differentiating. □

13.2 THE GENERAL, ONE-DIMENSIONAL LINEAR EQUATION

The Black–Scholes differential equation (17.1) (also (12.1)) is linear. The general linear stochastic differential equation driven by a Wiener process W_t is

$$dS_t = \big(r(t)S_t + a(t)\big)dt + \big(\sigma(t)S_t + b(t)\big)dW_t. \tag{13.2}$$

Its solution (cf. Karatzas and Shreve [10, Section 5.6]) can be given explicitly. To this end define the auxiliary quantities

$$\zeta_t := \int_0^t \sigma(u) dW_u - \frac{1}{2}\int_0^t \sigma(u)^2 du \qquad \text{and}$$
$$Z_t := \exp\left(\zeta_t + \int_0^t r(u) du\right).$$

Then

$$S_t = Z_t \left(S_0 + \int_0^t \frac{1}{Z_u}\big(a(u) - \sigma(u) b(u)\big) du + \int_0^t \frac{b(u)}{Z_u} dW_u\right). \tag{13.3}$$

In particular, the solution of

$$dS_t = r(t) S_t dt + \sigma(t) S_t dW_t$$

is

$$S_t = S_0 \cdot \exp\left(\int_0^t r(u) - \frac{1}{2}\sigma(u)^2 du + \int_0^t \sigma(u) dW_u\right).$$

Pham [19]

13.3 PROBLEMS

Exercise 13.1. *Discuss (13.3) in the non-stochastic case.*

Exercise 13.2. *Verify (13.1) explicitly.*

14 Kolmogorov Differential Equations

14.1 BACKWARD EQUATION

Consider the stochastic process, or Kolmogorov–Feller diffusion process

$$\begin{aligned} dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ X_0 &= x_0. \end{aligned} \tag{14.1}$$

Define the linear operator

$$(\mathscr{A} g)(t,x) \coloneqq \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{E}\big(g(t+\Delta t, X_{t+\Delta t}) - g(t, X_t) \mid X_t = x\big). \tag{14.2}$$

We shall often write $\mathbb{E}^{t,x} f(X_T) \coloneqq \mathbb{E}\left(f(X_T) | X_t = x\right)$ and $\mathbb{E}^{t,X_t} f(X_T) \coloneqq \mathbb{E}\left(f(X_T) | \mathcal{F}_t\right)$ for $t \leq T$.

Lemma 14.1. *We have that*

$$(\mathscr{A} g)(t,x) = \frac{\partial}{\partial t} g(t,x) + b(t,x)\frac{\partial}{\partial x} g(t,x) + \frac{1}{2}\sigma(t,x)^2 \frac{\partial^2}{\partial x^2} g(t,x).$$

Proof. We have from Itō's representation (12.2) that

$$\begin{aligned} g(t+\Delta t, X_{t+\Delta t}) &= g(t,x) \\ &+ \left(\frac{\partial g}{\partial t}(t, X_t) + b(t, X_t)\frac{\partial g}{\partial x}(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2 \frac{\partial^2 g}{\partial x^2}(t, X_t)\right)\Delta t \\ &+ \sigma(t, X_t)\frac{\partial g}{\partial x}(t, X_t)\Delta W_t. \end{aligned} \tag{14.3}$$

The result follows by taking expectations and by sending $\Delta t \to 0$. □

Remark 14.2. The operator $\mathscr{A}$ is a linear differential operator.

Proposition 14.3 (Dynkin's formula)**.** *We have*

$$\mathbb{E}^x g(\tau, X_\tau) = g(t,x) + \int_0^\tau \mathbb{E}^x (\mathscr{A} g)(s, X_s) ds.$$

Proof. The result follows by integrating the expectation of (14.3). □

Definition 14.4. It is often useful to decompose the operator $\mathscr{A} = \frac{\partial}{\partial t} + G_x$.

Theorem 14.5 (Kolmogorov backward equation)**.** *The function*

$$u(t,x) \coloneqq \mathbb{E}\big(f(X_T) \mid X_t = x\big) \tag{14.4}$$

satisfies the Kolmogorov backward equation

$$\begin{aligned} -u_t = G_x u \quad &\textit{or} \qquad \mathscr{A} u = 0 \textit{ and} \\ u(T,x) = f(x) \quad &(\textit{terminal condition at } t = T). \end{aligned} \tag{14.5}$$

Proof. We have from Ito's Lemma that

$$\begin{aligned}u(t+\Delta t, X_{t+\Delta t}) &= u(t,x) \\ &+ \left(\frac{\partial u}{\partial t}(t,X_t) + b(t,X_t)\frac{\partial u}{\partial x}(t,X_t) + \frac{1}{2}\sigma(t,X_t)^2\frac{\partial^2 u}{\partial x^2}(t,X_t)\right)\Delta t \\ &+ \sigma(t,X_t)\frac{\partial u}{\partial x}(t,X_t)\Delta W_t.\end{aligned} \tag{14.6}$$

Taking expectations with respect to $\mathbb{E}\,(\cdot \mid X_t = x)$ gives

$$\mathbb{E}\,[u(t+\Delta t, X_{t+\Delta t}) \mid X_t = x] = u(t,x) + \Delta t \cdot \mathscr{A} u(t,x). \tag{14.7}$$

From the Markov property we deduce for the particular function (14.4) that

$$\begin{aligned}\mathbb{E}\,[u(t+\Delta t, X_{t+\Delta t}) \mid X_t = x] &= \mathbb{E}\left[\mathbb{E}\big(f(X_T) \mid X_{t+\Delta t}\big) \mid X_t = x\right] \\ &= \mathbb{E}\,(f(X_T) \mid X_t = x) \\ &= u(t,x)\end{aligned}$$

and hence $\mathscr{A} u = 0$ follows from (14.7), the result. □

Remark 14.6. Denote by $p(x,r;y,t)$ the probability to be in y at time t, given that the process is in x at time r. For a *time homogeneous process* this density does not depend on t, i.e.,

$$P_x(X_{t+\Delta t} \in dy \mid X_t = x) = p(\Delta t; x, y)\, dy. \tag{14.8}$$

Then, for y fixed, the density $(t,x) \mapsto p(t;x,y)$ satisfies the backward equation (14.5).

Corollary 14.7. *For every y fixed it holds that*

$$\frac{\partial}{\partial t} p(t;x,y) = G_x p(t;x,y).$$

Proof. Consider the function $f_B(x) \coloneqq \frac{1}{\lambda(B)}\,\mathbb{1}_B(x)$ with $y \in B$. Then

$$u(t,x) = \mathbb{E}\,(f_B(X_T) \mid X_t = x) = \frac{P\,(X_T \in B \mid X_t = x)}{\lambda(B)} \xrightarrow[\lambda(B)\to 0]{} p(T-t;x,y),$$

as the process is time homogeneous. The result follows from (14.5), as the time t reverses. □

14.2 FORWARD EQUATION

Proposition 14.8. *The adjoint (or conjugate) of G_x is the operator*

$$G_x^* f = -\frac{\partial}{\partial x}\big(b(t,x)f(t,x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\sigma(t,x)^2 f(t,x)\right).$$

Theorem 14.9 (Kolmogorov forward equation, Fokker–Planck equation; cf. Corollary 14.7)**.** *For every x fixed it holds that*

$$\begin{aligned}\frac{\partial}{\partial t} p(t;x,y) &= G_y^* p(t;x,y), \\ p(0,x,y) &= \delta_x(y) \qquad \textit{initial condition at } t = 0.\end{aligned}$$

Proof. Consider again $u(t,x) \coloneqq \mathbb{E}\left(f(X_T) \mid X_t = x\right)$. As x is fixed we abbreviate $p(t,y) \coloneqq p(t,x,y)$ so that

$$\mathbb{E}\, f(X_T) = \mathbb{E}\,\mathbb{E}\left[f(X_T) \mid X_t\right] = \mathbb{E}\, u(t,X_t) = \int_{\mathbb{R}} u(t,y) P_x(X_t \in dy) = \int_{\mathbb{R}} u(t,y) p(t,y) dy,$$

which does not depend on time.

Differentiate the latter to get with the backward equation (14.5)

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} u(t,y)\right) p(t,y) + u(t,y) \frac{\partial}{\partial t} p(t,y)\, dy \\ &= -\int_{-\infty}^{\infty} \left(G_y u(t,y)\right) p(t,y)\, dy + \int_{-\infty}^{\infty} u(t,y) \frac{\partial}{\partial t} p(t,y)\, dy \\ &= -\int_{-\infty}^{\infty} u(t,y)\, G_y^* p(t,y)\, dy + \int_{-\infty}^{\infty} u(t,y) \frac{\partial}{\partial t} p(t,y) dy \\ &= \int_{-\infty}^{\infty} u(t,y) \left(\frac{\partial}{\partial t} p(t,y) - G_y^* p(t,y)\right) dy. \end{aligned}$$

As $f(\cdot)$ was arbitrary, the assertion follows. □

14.2.1 Scaled Brownian motion

The process $dX_t = \sigma\, dW_t$ has the probabilities $p(t,y) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2 t} y^2}$ which satisfies $p_t = \frac{1}{2}\sigma^2 p_{yy}$.

14.2.2 Brownian motion with drift

The process $dX_t = \mu dt + \sigma dW_t$ has the probabilities $p(t,y) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2 t}(y - \mu t - x_0)^2}$ which satisfy

$$p_t = -\mu\, p_y + \frac{1}{2}\sigma^2\, p_{yy}.$$

14.2.3 Geometric Brownian motion with drift

The process $dX_t = \mu X_t dt + \sigma X_t dW_t$ has the probabilities $p(t,y) = \frac{1}{y\sigma\sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2 t}\left(\ln\frac{y}{x_0} - (\mu - \frac{1}{2}\sigma^2)t\right)^2}$ which satisfy

$$p_t = -\left(\mu - \sigma^2\right) p - \left(\mu - 2\sigma^2\right) y p_y + \frac{\sigma^2}{2} y^2 p_{yy}.$$

15 Cox-Ross-Rubinstein-Model, or Binomial Model

We start with the Bernoulli model, which we extend to the binomial model, cf. Cox et al. [3]. The binomial model extends asymptotically to the log-normal distribution, describing the evolution of the stock under the risk free measure.

Remark 15.1. It is common in financial mathematics to distinguish the interest rate and the *risk free* interest rate. We follow this practice and denote the *force of interest* by r ($\frac{1}{1+i} = e^{-r}$, instead of δ).

15.1 THE BERNOULLI MODEL

We motivate the idea of hedging with the following example.

Example 15.2. Consider a (simplified) stock which evolves over time (cf. Figure 15.1). Its initial price (time $t = 0$) is S_0, and the price S_t at the later time (t) randomly increases to S_{up} or decreases to S_{down}. Assume we are interested in a contract with the following characteristics:

- if $S_t = S_{up}$, then we expect a payment of V_{up}
- in case the price decreases to $S_t = S_{down}$, then we expect V_{down}.

How much should we pay for a contract, which guarantees these payments? What is a fair price?

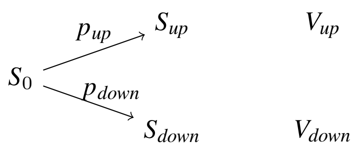


Figure 15.1: Bernoulli setting

We intend to hedge the contract. That is, we buy y_0 stocks and invest an amount of x_0 in money (bond) which is available at an interest rate of i. The idea of hedging means to adjust the numbers x_0 and y_0 in such way, so that the claim $V(S) = \begin{cases} V_{up} & \text{if } S = S_{up} \\ V_{down} & \text{if } S = S_{down} \end{cases}$ is available at the end *without* accepting/ facing any risk. As a buyer, we intend to accept a price which represents a minimum of possible hedging strategies. To this end consider the optimization problem

$$\begin{aligned} \text{minimize}_{\,(\text{in } x_0, y_0)} \quad & x_0 + y_0\, S_0 \\ \text{subject to} \quad & x_0(1+i) + y_0\, S_{up} \geq V_{up}, \\ & x_0(1+i) + y_0\, S_{down} \geq V_{down}. \end{aligned} \tag{15.1}$$

Remark 15.3. Note, that (15.1) does not make any assumption an any transition probability from S_0 to S_{up} or S_{down}.

To solve the linear optimization problem (15.1) consider the Lagrangian (cf. (4.4))

$$L(x_0, y_0; p_{up}, p_{down}) = \left\{ \begin{array}{l} x_0 + y_0\, S_0 \\ \quad -p_{up}\left(x_0 + \frac{y_0\, S_{up} - V_{up}}{1+i}\right) \\ \quad -p_{down}\left(x_0 + \frac{y_0\, S_{down} - V_{down}}{1+i}\right), \end{array} \right.$$

where p_{up} and p_{down} are the Lagrangian dual parameters corresponding to the two constraints in (15.1). To identify the optimal parameters x_0, y_0, p_{up} and p_{down} we take the derivatives of the Lagrangian and obtain the system of equations

$$\begin{pmatrix}\frac{\partial}{\partial x_0}\\ \frac{\partial}{\partial y_0}\\ \frac{\partial}{\partial p_{up}}\\ \frac{\partial}{\partial p_{down}}\end{pmatrix} L = 0, \text{ i.e., } \begin{pmatrix}0 & 0 & 1 & 1\\ 0 & 0 & \frac{S_{up}}{1+i} & \frac{S_{down}}{1+i}\\ 1 & \frac{S_{up}}{1+i} & 0 & 0\\ 1 & \frac{S_{down}}{1+i} & 0 & 0\end{pmatrix}\begin{pmatrix}x_0\\ y_0\\ p_{up}\\ p_{down}\end{pmatrix} = \begin{pmatrix}1\\ S_0\\ \frac{V_{up}}{1+i}\\ \frac{V_{down}}{1+i}\end{pmatrix}.$$

The solution of this linear system of equations is

$$\begin{pmatrix}x_0\\ y_0\\ p_{up}\\ p_{down}\end{pmatrix} = \begin{pmatrix}\frac{S_{up}V_{down}-S_{down}V_{up}}{(1+i)(S_{up}-S_{down})}\\ \frac{V_{up}-V_{down}}{S_{up}-S_{down}}\\ \frac{(1+i)S_0-S_{down}}{S_{up}-S_{down}}\\ \frac{S_{up}-(1+i)S_0}{S_{up}-S_{down}}\end{pmatrix}. \tag{15.2}$$

Now observe the following:

▹ $p_{up} + p_{down} = 1$, i.e., the dual parameters p_{up} and p_{down} can be interpreted as a probability (provided that $p_{up} \geq 0$ and $p_{down} \geq 0$, which is the case under the mild and natural assumption $S_{down} \leq (1+i)S_0 \leq S_{up}$;

▹ It holds that

$$\frac{1}{1+i}\big(p_{up}\,S_{up} + p_{down}\,S_{down}\big) = \frac{1}{1+i}\left(\frac{(1+i)S_0 - S_{down}}{S_{up}-S_{down}}S_{up} + \frac{S_{up}-(1+i)S_0}{S_{up}-S_{down}}S_{down}\right) = S_0,$$

i.e., given the probabilities we may write $S_0 = \frac{1}{1+i}\,\mathbb{E}\,S_T$;

▹ further,

$$\begin{aligned}\frac{1}{1+i}\big(p_{up}\,V_{up} + p_{down}\,V_{down}\big) &= \frac{1}{1+i}\left(\frac{(1+i)S_0 - S_{down}}{S_{up}-S_{down}}V_{up} + \frac{S_{up}-(1+i)S_0}{S_{up}-S_{down}}V_{down}\right)\\ &= \frac{S_{up}\,V_{down} - S_{down}\,V_{up}}{(1+i)(S_{up}-S_{down})} + \frac{V_{up}-V_{down}}{S_{up}-S_{down}}S_0\\ &= x_0 + y_0\,S_0,\end{aligned}$$

i.e., the objective of our problem (15.1) is given in terms of the probabilities identified as

$$v_0 \coloneqq x_0 + y_0\,S_0 = \frac{1}{1+i}\,\mathbb{E}\,V(S_T). \tag{15.3}$$

What did we achieve?

(i) We have found a probability measure $P(\cdot) \coloneqq p_{up}\,\delta_{S_{up}}(\cdot) + p_{down}\,\delta_{S_{down}}(\cdot)$ which depends on S but not on V. The measure is called the *risk free measure*.

Neglecting interest (i.e., $i = 0$) it holds that

(ii) $S_0 = \frac{1}{1+i}\,\mathbb{E}\,S_t$, i.e., $\tilde{S} = \big(S_0, \frac{1}{1+i}S\big)$ is a *martingale* and

(iii) the price for the claim is a simple expectation, $v_0 = \mathbb{E}\,\frac{1}{1+i}V(S)$ and further, the process $v = \big(v_0, \frac{1}{1+i}V(S)\big)$ is a martingale.

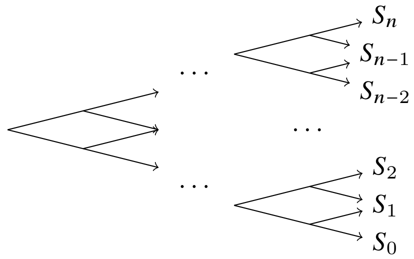


Figure 15.2: Binomial lattice

15.2 BINOMIAL MODEL

The binomial model extends and generalizes the Bernoulli model. Consider a whole lattice (Fig. 15.2) instead of the Bernoulli distribution (Fig. 15.1).

Again we consider a payoff $V(S_T)$ at the end of the period $[0,T]$. In the binomial model we choose

$$\Delta t := \frac{T}{n}, \quad S_{up} = S_0 \cdot e^{\sigma\sqrt{\Delta t}} \quad \text{and} \quad S_{down} = S_0 \cdot e^{-\sigma\sqrt{\Delta t}} \tag{15.4}$$

to describe the evolution of the stock, the percentage rate (interest i) corresponding to the time period $\frac{T}{n}$ is $1+i=e^{r\frac{T}{n}}$. In this exponential setting, the *risk-free* probabilities

$$\begin{aligned}
p := p_{up} &= \frac{(1+i)S_0 - S_{down}}{S_{up}-S_{down}} = \frac{e^{r\frac{T}{n}} - e^{-\sigma\sqrt{\frac{T}{n}}}}{e^{\sigma\sqrt{\frac{T}{n}}} - e^{-\sigma\sqrt{\frac{T}{n}}}} \\
&= \frac{1 + r\frac{T}{n} - \left(1 - \sigma\sqrt{\frac{T}{n}} + \frac{1}{2}\sigma^2\frac{T}{n}\right)}{\left(1 + \sigma\sqrt{\frac{T}{n}} + \frac{1}{2}\sigma^2\frac{T}{n}\right) - \left(1 - \sigma\sqrt{\frac{T}{n}} + \frac{1}{2}\sigma^2\frac{T}{n}\right)} + \mathcal{O}\left(n^{-3/2}\right) \\
&= \frac{r\frac{T}{n} + \sigma\sqrt{\frac{T}{n}} - \frac{1}{2}\sigma^2\frac{T}{n}}{2\sigma\sqrt{\frac{T}{n}}} + \mathcal{O}\left(n^{-3/2}\right) \qquad (15.5) \\
&= \frac{1}{2} + \frac{1}{2\sigma}\sqrt{\frac{T}{n}}\left(r - \frac{\sigma^2}{2}\right) + \mathcal{O}\left(n^{-3/2}\right) \qquad (15.6)
\end{aligned}$$

and p_{down} remain constant (i.e., they do not depend on S_{up}, S_{down} nor S_0). In this setting the stock follows the binomial distribution

$$P\left(S_T = S_0 e^{\sigma(2k-n)\sqrt{\Delta t}}\right) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{15.7}$$

or

$$P\left(\frac{\frac{1}{\sigma}\ln\frac{S_T}{S_0} + n\sqrt{\Delta t}}{2\sqrt{\Delta t}} = k\right) = \binom{n}{k} p^k (1-p)^{n-k},$$

that is, asymptotically,

$$\frac{\frac{1}{\sigma}\ln\frac{S_T}{S_0}+n\sqrt{\frac{T}{n}}}{2\sqrt{\frac{T}{n}}}\sim \operatorname{bin}(n,p)\xrightarrow{\mathcal{D}}\mathcal{N}\big(np,np(1-p)\big),$$

or

$$\frac{\frac{\frac{1}{\sigma}\ln\frac{S_T}{S_0}+n\sqrt{\frac{T}{n}}}{2\sqrt{\frac{T}{n}}}-np}{\sqrt{np(1-p)}}\cdot\sqrt{T}\xrightarrow{\mathcal{D}}\mathcal{N}(0,T)$$

by adjusting the mean and re-scaling. Now note that $p(1-p)=\frac{1}{4}+O(1/n)$ by (15.6) and

$$\begin{aligned}\frac{\frac{\frac{1}{\sigma}\ln\frac{S_T}{S_0}+n\sqrt{\frac{T}{n}}}{2\sqrt{\frac{T}{n}}}-np}{\sqrt{np(1-p)}}\sqrt{T}&\sim\frac{\frac{\frac{1}{\sigma}\ln\frac{S_T}{S_0}}{2\sqrt{\frac{T}{n}}}+\frac{n}{2}-np}{\frac{\sqrt{n}}{2}}\sqrt{T}\\&=\frac{1}{\sigma}\ln\frac{S_T}{S_0}+2\sqrt{Tn}\left(\frac{1}{2}-p\right)\sim\frac{1}{\sigma}\ln\frac{S_T}{S_0}-\frac{T}{\sigma}\left(r-\frac{\sigma^2}{2}\right)\end{aligned}\tag{15.8}$$

so that $\frac{1}{\sigma}\ln\frac{S_T}{S_0}-\frac{T}{\sigma}\left(r-\frac{\sigma^2}{2}\right)\sim\mathcal{N}(0,T)$. It follows that $S_T=S_0\,e^{T(r-\frac{1}{2}\sigma^2)+\sigma\,W_T}$, where $W_T\sim\mathcal{N}(0,T)$.

15.3 LOG-NORMAL MODEL

More generally we see that S_t is log-normally distributed for the risk free measure, that is,

$$S_t=S_0\cdot e^{t(r-\frac{1}{2}\sigma^2)+\sigma W_t}$$

with $W_t\sim\mathcal{N}(0,t)$.

Lemma 15.4. *It holds that $S_0=e^{-rt}\,\mathbb{E}\,S_t$.*

Proof. Indeed,

$$\mathbb{E}\,S_t=\mathbb{E}\,S_0e^{t(r-\frac{1}{2}\sigma^2)+\sigma W_t}=S_0e^{tr}\cdot\mathbb{E}\,e^{-\frac{1}{2}\sigma^2+\sigma W_t},$$

by Wald's identity (12.14). □

Consider now (15.3) to see that the price of a payoff $v(\cdot)$ in the binomial setting thus is

$$v_0=e^{-rT}\cdot\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\cdot v\left(S_0e^{\sigma(2k-n)\sqrt{\Delta t}}\right),$$

or asymptotically

$$v_0=e^{-rT}\cdot\mathbb{E}\,v\Big(S_0e^{T(r-\frac{1}{2}\sigma^2)+\sigma W_T}\Big),\tag{15.9}$$

where the expectation $\mathbb{E}$ is with respect to the risk free measure.

16 Options

16.1 PAY-OFF FUNCTION

The pay-off function of a put option at maturity is $V(S) := \max\{K - S, 0\}$, of a call option it is $V(S) := \max\{S - K, 0\}$ where K is some fixed price, the strike price.

Remark. A private investor may hold *long positions* of

- cash,
- stocks,
- call and put options,

but no short positions of these instruments. However, he may be short in cash by taking a loan from his bank.

The pay-off function thus is always *convex*; Conversely and moreover, any convex pay-off function may be replicated, possibly by different means.

Exercising options: American options may be exercised any time, whereas European options can be exercised only at expiry.

16.2 EXPLICIT FORMULA FOR EUROPEAN PAYOFF FUNCTIONS

European Options are exercised at a specified time (date) T. The explicit formula (15.9) thus applies.

16.2.1 European call option

Following (15.9), the value (or price) v of the European call with payoff

$$V_{call}(S) := \max\{0, S - K\} = (S - K)_+$$

is

$$\begin{aligned} v_{call}(t, S) &= e^{-r(T-t)}\, \mathbb{E}\, V_{call}\Big(S \cdot e^{(T-t)(r-\frac{1}{2}\sigma^2)+\sigma W_{T-t}}\Big) \qquad (16.1) \\ &= e^{-r(T-t)} \int_{-\infty}^{\infty} \Big(Se^{(T-t)(r-\frac{1}{2}\sigma^2)+\sigma x} - K\Big)_+ \cdot \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{1}{2(T-t)}x^2}\, dx \\ &\underset{x \leftarrow -x\sqrt{T-t}}{=} \int_{-\infty}^{\infty} \Big(Se^{-(T-t)\frac{1}{2}\sigma^2 - \sigma\sqrt{T-t}x} - Ke^{-r(T-t)}\Big)_+ \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx. \qquad (16.2) \end{aligned}$$

To evaluate the option further define the auxiliary quantities

$$d_\pm := \frac{1}{\sigma\sqrt{T-t}} \left[\left(r \pm \frac{\sigma^2}{2}\right)(T-t) + \ln\frac{S}{K}\right]$$

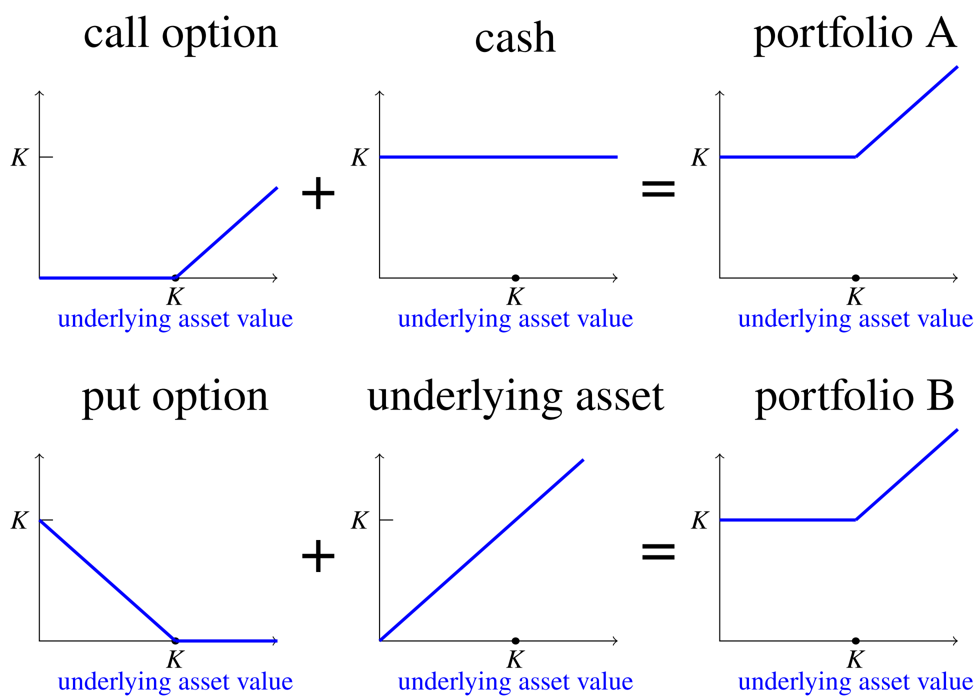


Figure 16.1: Identical portfolios at maturity by parity

and observe that $x \le d_-$ in (16.2). Thus,

$$\begin{aligned} v_{call}(t,S) &= \int_{-\infty}^{d_-} S e^{-(T-t)\frac{1}{2}\sigma^2 - \sigma\sqrt{T-t}x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx - K e^{-r(T-t)} \int_{-\infty}^{d_-} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= S \int_{-\infty}^{d_-} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x+\sigma\sqrt{T-t}\right)^2} dx - K e^{-r(T-t)} \Phi(d_-) \\ &= S \int_{-\infty}^{d_- + \sigma\sqrt{T-t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx - K e^{-r(T-t)} \Phi(d_-) \\ &= S \cdot \Phi(d_+) - K \cdot e^{-r(T-t)} \Phi(d_-). \end{aligned} \tag{16.3}$$

after some algebra, where $\Phi(\cdot)$ is the cdf of the normal distribution given in (5.11) and $\tau \coloneqq T - t$ is the remaining time to maturity.

Remark 16.1. Note that $d_\pm$ are functions of t and S and they can be expressed as

$$d_\pm(t,S) = \frac{\ln \frac{S}{K e^{-r(T-t)}} \pm \frac{1}{2}\sigma^2 (T-t)}{\sigma\sqrt{T-t}};$$

further, $d_+ - d_- = \sigma\sqrt{T-t}$.

16.2.2 European put option

The explicit price of the European put option with pay-off

$$V_{put}(S) \coloneqq \max\{K - S, 0\} \tag{16.4}$$

is

$$\begin{aligned} v_{put}(t,S) &= e^{-r(T-t)}\,\mathbb{E}\,V_{put}\Big(S\cdot e^{(T-t)(r-\frac{1}{2}\sigma^2)+\sigma W_{T-t}}\Big) \\ &= Ke^{-r(T-t)}\Phi(-d_-) - S\cdot\Phi(-d_+) \\ &= Ke^{-r(T-t)}\left(1-\Phi(d_-)\right) - S\left(1-\Phi(d_+)\right) \end{aligned} \tag{16.5}$$

16.2.3 Parity

It holds generally that $\max(x,y)+\min(x,y)=x+y$, hence

$$\begin{aligned} \max(S-K,0)+\min(S-K,0) &= S-K, \text{ or} \\ S+\max(K-S,0) &= K+\max(S-K,0). \end{aligned}$$

Taking expectations gives the following put-call parity.

Theorem 16.2 (Parity for European options)**.** *The identity*

$$S+v_{put}(t,S)=Ke^{-r(T-t)}+v_{call}(t,S), \tag{16.6}$$

is called the put-call parity *for European options.*[1]

Proof. This is immediate from (16.8) and (16.9) (from (16.3) and (16.5), respectively). □

Notice also the relation of the parity (16.6) and Figure 16.1.

Remark 16.3 (Comparison of option values and the intrinsic value, cf. Figure 16.2)*.* Recall from the European call option (from (16.6)) that

$$\underbrace{v_{call}(t,S)}_{\text{option value}} \ge \max\left\{S-Ke^{-r(T-t)},0\right\} \ge \underbrace{\max\left\{S-K,0\right\}}_{\text{intrinsic value}} \tag{16.7}$$

(as $v_{put}(t,S)\ge 0$ and by assuming that $r\ge 0$). It follows that the option value $v_{\text{call}}(t,S)$ is greater than its intrinsic value $S-K$.

The respective inequality for European put options reads

$$v_{put}(t,S)\ge\max\left\{e^{-r(T-t)}K-S,0\right\},$$

but the option value of the put is *not necessarily* greater than its intrinsic value.

16.2.4 Dividend paying stocks

Including continuous *dividend payments* (which are modeled as $qS_t\,dt$ with q representing the dividend payment rate) the prices for the call and put are

$$v_{call}(t,S) := S\cdot e^{-q(T-t)}\cdot\Phi(d_+) - K\cdot e^{-r(T-t)}\cdot\Phi(d_-) \tag{16.8}$$
$$v_{put}(t,S) := K\cdot e^{-r(T-t)}\cdot\Phi(-d_-) - S\cdot e^{-q(T-t)}\cdot\Phi(-d_+) \tag{16.9}$$

for $d_\pm=\frac{\ln\frac{Se^{-q(T-t)}}{Ke^{-r(T-t)}}\pm\frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$ (note again that $d_+-d_-=\sigma\sqrt{T-t}$).

[1] For theta cf. duration.

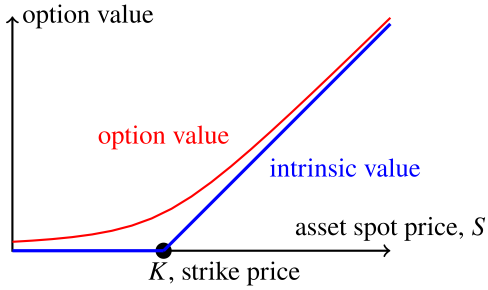


Figure 16.2: Comparison of option value and intrinsic value for a call option

Remark 16.4. Note that these formulae (16.8) and (16.9) are more general than (16.3) and (16.5) (choose $q = 0$ there).

Theorem 16.5 (Parity for European options)**.** *The identity*

$$S \cdot e^{-q(T-t)} + v_{put}(t, S) = Ke^{-r(T-t)} + v_{call}(t, S),$$

or

$$S + v_{put}(t, S) = Ke^{-r(T-t)} + v_{call}(t, S),$$

is called the put-call parity *for European options.*[2]

16.2.5 FX options: the Garman-Kohlhagen model

The pricing formulae may be used for options on foreign currencies (*FX-options*), one simply replaces the interest rate r and the continuous dividend q in the preceding Section 16.2.4 by the domestic and foreign risk free interest rate,

$$r \leftrightarrow r_{domestic}, \qquad q \leftrightarrow r_{foreign}.$$

16.2.6 Moneyness

Moneyness is the probability for a cash transaction to take place at maturity. For a call this evaluates to

$$\begin{aligned}
P(S_T \geq K) &= P\left(S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma W_{T-t}} \geq K\right) \\
&= P\left(\ln \frac{S_0}{K} + \left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma W_{T-t} \geq 0\right) \\
&= P\left(W_{T-t} \geq -\frac{1}{\sigma}\left(\ln \frac{S_0}{K} + \left(r - \frac{1}{2}\sigma^2\right)(T-t)\right)\right) \\
&= P\left(\frac{W_{T-t}}{\sqrt{T-t}} \leq \frac{1}{\sigma\sqrt{T-t}}\left(\ln \frac{S_0}{K} + \left(r - \frac{1}{2}\sigma^2\right)(T-t)\right)\right) \\
&= \Phi\left(d_-\right).
\end{aligned}$$

Moneyness of a put option apparently is

$$P(S_T \leq K) = 1 - \Phi\left(d_-\right) = \Phi\left(-d_-\right).$$

[2]For theta cf. duration.

16.2.7 European Greeks

$\tau \coloneqq T - t$		call	put
Delta	$\Delta = \frac{\partial}{\partial S}$	$e^{-q\tau}\Phi(d_+)$	$-e^{-q\tau}\Phi(-d_+) = -e^{-q\tau}(1-\Phi(d_+))$
Moneyness		$\Phi(d_-)$	$1-\Phi(d_-)$
Gamma	$\Gamma = \frac{\partial^2}{\partial S^2}$	$e^{-q\tau}\frac{\Phi'(d_+)}{S\sigma\sqrt{\tau}}$	
Vega	$\nu = \frac{\partial}{\partial\sigma}$	$Se^{-q\tau}\Phi'(d_+)\sqrt{\tau} = Ke^{-q\tau}\Phi'(d_-)\sqrt{\tau}$	
Theta	$\Theta = \frac{\partial}{\partial t}$	call: $-Se^{-q\tau}\frac{\Phi'(d_+)\sigma}{2\sqrt{\tau}} - rKe^{-r\tau}\Phi(d_-) + qSe^{-q\tau}\Phi(d_+)$	
	$\Theta = \frac{\partial}{\partial t}$	put: $-Se^{-q\tau}\frac{\Phi'(d_+)\sigma}{2\sqrt{\tau}} + rKe^{-r\tau}\Phi(-d_-) - qSe^{-q\tau}\Phi(-d_+)$	
rho	$\rho = \frac{\partial}{\partial r}$	$K\tau e^{-r\tau}\Phi(d_-)$	$-K\tau e^{-r\tau}\Phi(-d_-)$

16.3 AMERICAN OPTIONS

The American options give more rights to the owner (they can be exercised at arbitrary times) and thus they are more expensive.

16.3.1 American call options

Definition 16.6. A random variable τ is a stopping time of the filtration $\mathcal{F}_t$, $t \in [0,T]$, if $\{\tau \le t\} \in \mathcal{F}_t$.

The price of an American call option is

$$v_{call}^A(t,S) = \sup_{\tau\in[t,T]} \mathbb{E}_S\, e^{-r(\tau-t)} \cdot \max(S_\tau - K, 0), \tag{16.10}$$

where $\tau(\cdot)$ is a stopping time (a random variable itself, adapted to the filtration) with values in $[t,T]$ determining the time *when* the option should be exercised.

The following theorem justifies that American and European call options are traded at the same price, $V_{call}^A = V_{call}^E$.

Theorem 16.7 (Merton's no early exercise theorem)**.** *An American call option should* not *be exercised prematurely (assuming that the interest is $r \ge 0$).*

Proof. If exercised at time t, the total value of the option is its intrinsic value $S_t - K$ (this corresponds to the stopping time $\tau = t$ in (16.10)). However, by (16.7) the European call is more valuable than the intrinsic value. Further, the American option provides more rights to its holder and is thus even more expensive than the European option, i.e.,

$$v_{call}^A(t,S) \ge v_{call}^E(t,S).$$

By Jensen's inequality, $\varphi(\mathbb{E}\,S_t) \le \mathbb{E}\,\varphi(S_t)$ for convex functions φ. The function $\varphi(S) \coloneqq (S-K)_+$ is convex, thus

$$\begin{aligned} v_{call}^A(t,S) &\ge v_{call}^E(t,S) = e^{-rt}\,\mathbb{E}\,\varphi(S_t) \ge e^{-rt}\varphi(\mathbb{E}\,S_t) = e^{-rt}\varphi(e^{rt}S) \\ &= e^{-rt}\max(e^{rt}S - K, 0) = \max(S - e^{-rt}K, 0) \ge S - K. \end{aligned}$$

The value $\max\{0, S-K\}$ corresponds to the stopping time $\tau = t$ in (16.10). $\tau = t$ thus is not optimal and it is better to wait. □

In absence of dividends it is optimal to exercise an American call option at its expiry, that is not to exercise an American call. If there are dividends, then it is optimal to exercise only at a time immediately before the stock goes ex dividend.

16.3.2 American put option

The price for the American put option is

$$v_{put}^{A}(t,S) = \sup_{\tau\in[t,T]} \mathbb{E}\, e^{-r(\tau-t)} \cdot \max(K - S_\tau, 0)\,. \tag{16.11}$$

An argument as in Merton's theorem does not hold true here (why?).

The optimal time to exercise an American option is called *fugit*.[3] It follows for $\tau = t$ in (16.11) that

$$P \coloneqq v_{put}^{A}(t,S) \geq \max\{K - S, 0\} \geq K - S. \tag{16.12}$$

Whence the option should be exercised, whenever the criterion (16.12) is violated, i.e.,

$$P + S \leq K \qquad \text{(fugit)}. \tag{16.13}$$

Note, that P and S can be observed on the market: the American put option is exercised once the fugit criterion (16.13) *occasionally* holds true.

The Black–Scholes price of an American put can be evaluated using an algorithm (Algorithm 1).

Result: americanPut($\tau, S, K, r, \sigma, q, n$) — implementation Binomial Model
$\Delta t \coloneqq \tau/n$;
$up \coloneqq e^{+\sigma\sqrt{\Delta t}}$; $down \coloneqq e^{-\sigma\sqrt{\Delta t}}$; — binomial tree setting, cf. (15.4)
$p_{up} \coloneqq \frac{e^{(r-q)\Delta t} - down}{up - down}$; $p_{down} \coloneqq \frac{up - e^{(r-q)\Delta t}}{up - down}$; — the risk free measure, cf. (15.2)
for $i = 0$ **to** n **do**
  $v_i \coloneqq \max\{0,\ K - S * up^{2i-n}\}$; — payoff at maturity, cf. (16.4) and (15.7)
end
for $t \coloneqq n-1$ **downto** 0 **step** -1 **do**
  for $i \coloneqq 0$ **to** t **do**
    $v_i \coloneqq e^{-r\Delta t} \cdot \left(p_{up} * v_{i+1} + p_{down} * v_i\right)$; — the martingale property, cf. (15.3)
    $exerciseNow \coloneqq K - S * up^{2i-t}$;
    if $v_i < exerciseNow$ **then**
      $v_i \coloneqq exerciseNow$; — fugit, cf. (16.13)
    end
  end
end
return *americanPut*:= v_0; — return v_0, i.e., the price of the American put

Algorithm 1: Binomial model: the price of an American put option

Theorem 16.8 (Parity for American options)**.** *For American options it holds that*

$$S - K \leq v_{call}(t,S) - v_{put}^{A}(t,S) \leq S - Ke^{-r(T-t)}. \tag{16.14}$$

Proof. The second inequality $v_{call}(t,S) - v_{put}^{A}(t,S) \leq S - Ke^{-r(T-t)}$ follows from $v_{put}^{A} \geq v_{put}^{E}$ and the parity equation (16.6). As for the remaining inequality consider two portfolios:

A: one American call and K € in cash

B: one American put and one share.

[3] fugit (lat.): flees, flight

Exercising B gives $\max(K - S_t, 0) + S_t = \max(S_t, K) =: B$.

Exercising A at the same time as B gives

$$A \coloneqq \max(S_t - K, 0) + Ke^{rt} = \max(S_t, K) + K\left(e^{rt} - 1\right) \geq B.$$

By taking expectations it follows $v_{call} + K \geq v_{put} + S$, the remaining equation. □

16.3.3 Other options

Payoff functions of different options:

Option	Payoff-function
Call option on F	$\max\big(F(S_T^1, S_T^2) - K, 0\big)$
Put option on F	$\max\big(K - F(S_T^1, S_T^2), 0\big)$
Barrier option on F	$\mathbb{1}_{\{F(S_T^1, S_T^2) > K\}}$
Lookback option on F	$\max\big(K - \min_{0\leq t\leq T} F(S_t^1, S_t^2), 0\big)$

Functions for F include

Option type	Payoff-function
Spread	$F(S_T^1, S_T^2) = a_1 S_T^1 - a_2 S_T^2$
Basket	$F(S_T^1, S_T^2) = a_1 S_T^1 + a_2 S_T^2$
Best of	$F(S_T^1, S_T^2) = \max(S_T^1, S_T^2)$
Worst of	$F(S_T^1, S_T^2) = \min(S_T^1, S_T^2)$
Average	$\int_t^T S_u^1\,\mathrm{d}u + \int_t^T S_u^2\,\mathrm{d}u$

16.4 PROBLEMS

Exercise 16.1. *Set $S_0 = 10$, $S_{up} = 12$, $S_{down} = 7$ and verify (for $i = 0$) that $p_{up} = 60\,\%$ and $p_{down} = 40\,\%$. For a payoff with $V_{up} = 1$ and $V_{down} = 0$, the investments are $x_0 = -1.4$ (cash) and $y = 0.2$ shares.*

Exercise 16.2. *Compute the Black–Scholes price for a put and call option with $r = 1\%$, $\sigma = 15\%$, $T - t = 1$, $S = 10$ and $K = 11$.*

Exercise 16.3. *Verify Remark 16.4.*

Exercise 16.4. *Verify the derivations in Example 15.2.*

Exercise 16.5. *Formulate and solve the dual of the linear optimization problem (15.1) explicitly.*

Exercise 16.6. *Consider the* Bernoulli *setting in Section 15.2 and verify that $\mathbb{E}\, S_T = e^{r\frac{T}{n}} S_0$ and*

$$\operatorname{var} S_T = S_0^2 \frac{T}{n} \sigma^2 + \mathcal{O}\left(\frac{1}{n^2}\right).$$

Exercise 16.7. *Verify the asymptotic expressions (15.6) and (15.8).*

Exercise 16.8. *Verify the price of some options (cf. (16.3) and (16.5)) explicitly by employing (15.9).*

Exercise 16.9. *Implement Algorithm 1.*

Exercise 16.10. *Compute the price of some European/American call/put options and compare your result with actually traded options.*

Black–Scholes Differential Equation

17

Black[1] and Scholes[2] have been awarded the Nobel Memorial Prize in Economic Sciences in 1997 for the method to determine the value of derivatives.

A former model was given by Louis Bachelier.[3]

17.1 DERIVATION OF THE BLACK–SCHOLES DIFFERENTIAL EQUATION

Recall that

- the final solution obtained in (15.9) for the European option with pay-off $V(\cdot)$ is $v(t,S) \coloneqq \mathbb{E}\, V\Big(S \cdot e^{t(r-\frac{1}{2}\sigma^2)+\sigma W_t}\Big)$;
- the geometric Brownian motion $S_t = S_0 \cdot e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W_t}$ satisfies the stochastic differential equation

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t. \tag{17.1}$$

The value of the derivative varies according Itō's rule (12.2) as

$$\begin{aligned} dv(t,S_t) &= v_t\, dt + v_S dS_t + \frac{1}{2} v_{SS} (dS_t)^2 \\ &= v_t\, dt + v_S \cdot \big(\mu S_t\, dt + \sigma S_t\, dW_t\big) + \frac{1}{2} v_{SS} \cdot (dS_t)^2 \\ &= \left(v_t + \mu S_t v_S + \frac{1}{2}\sigma^2 S_t^2 v_{SS}\right) dt + \sigma S_t v_S\, dW_t. \end{aligned}$$

To get rid of the random part we choose a portfolio consisting of (note that this is what is called a Δ-hedge)

(i) $\Delta \coloneqq \frac{\partial}{\partial S} v = v_S$ stocks and

(ii) -1 derivatives.

The value of the new portfolio is $\Pi \coloneqq -v + v_S \cdot S$. We find that

$$\begin{aligned} d\Pi_t = -dv + v_S \cdot dS_t &= -\left(v_t + \mu S_t v_S + \frac{1}{2}\sigma^2 S^2 v_{SS}\right) dt - \sigma v_S S_t\, dW_t + v_S\big(\mu S_t\, dt + \sigma S_t\, dW_t\big) \\ &= \left(-v_t - \frac{1}{2}\sigma^2 S^2 v_{SS}\right) dt. \end{aligned} \tag{17.2}$$

Note, that the price of the portfolio Π_t is *not random* any longer and further, μ is gone.

However, there is only one risk free asset on the market. It has interest r, and thus

$$d\Pi_t = r\Pi_t\, dt = r\,(-v + S \cdot v_S)\, dt. \tag{17.3}$$

[1] Fischer S. Black, 1938–1995, American Economist

[2] Myron S. Scholes, 1941–, Canadian-American financial economist

[3] Louis Jean-Baptiste Alphonse Bachelier, 1870–1946, French mathematician, a student of Henri Poincaré

By comparing (17.2) and (17.3) we find that $\left(-v_t - \frac{1}{2}\sigma^2 S^2 v_{SS}\right) = r\left(-v + S \cdot v_S\right)$, or

$$v_t + rSv_S + \frac{1}{2}\sigma^2 S^2 v_{SS} = r\,v, \tag{17.4}$$

which is the Black–Scholes differential equation. Note in particular that this equation is free of the drift μ but involves the risk free interest rate r instead.

Remark 17.1. The above derivation holds as well for time-dependent μ, σ and r. The Black–Scholes differential equation then is (written in full detail)

$$\frac{\partial}{\partial t}v(t,S) + \frac{1}{2}\sigma(t)^2 S^2 \frac{\partial^2}{\partial S^2}v(t,S) + r(t)S\frac{\partial}{\partial S}v(t,S) - r(t)v(t,S) = 0. \tag{17.5}$$

The function $V(S,t)$ in (16.3) is a solution of the Black–Scholes equation (17.5) (cf. Exercise (17.2)).

Remark 17.2. Both prices (16.8) and (16.9) follow the Black-Scholes partial differential equation (PDE) (17.4),

$$\underbrace{\frac{\partial v}{\partial t}}_{\Theta} + \frac{1}{2}\sigma^2 S^2 \underbrace{\frac{\partial^2 v}{\partial S^2}}_{\Gamma} + rS \underbrace{\frac{\partial v}{\partial S}}_{\Delta} = rv, \tag{17.6}$$

but with different termination conditions $v\,(T,S) = \max\{S-K,\,0\}$ and $v\,(T,S) = \max\{K-S,\,0\}$.

Remark 17.3. Equation (17.6) displays the Black–Scholes differential equation by involving important Greeks.

17.2 GENERAL SOLUTION OF THE BLACK–SCHOLES DIFFERENTIAL EQUATION

We consider a European option with payoff function $V(S)$. The Black–Scholes differential equation with boundary condition is

$$\frac{\partial}{\partial t}v + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2}v + rS\frac{\partial}{\partial S}v - rv = 0 \text{ with} \tag{17.7}$$
$$v(T,S) = V(S),\ S \geq 0 \text{ and}$$
$$v(t,0) = 0,\ 0 \leq t \leq T$$

To solve the equation we introduce the new variables

(i) $\tau = T - t$ (remaining time to maturity) and

(ii) $x = \left(r - \frac{1}{2}\sigma^2\right)\tau + \log\frac{S}{K}$ (i.e., $S = K\exp\left(x - (r - \frac{1}{2}\sigma^2)\tau\right)$

and the new function

$$u(\tau,x) = e^{r\tau}\cdot v(t,S),$$

i.e.,

$$u(\tau,x) = e^{r\tau}\cdot v\big(t(\tau,x), S(\tau,x)\big).$$

Differentiating (i) with respect to the new variables τ and x gives $\frac{dt}{d\tau} = -1$ and $\frac{dt}{dx} = 0$, and differentiating (ii) with respect to τ and x reveals that $\frac{dS}{d\tau} = -S\left(r - \frac{1}{2}\sigma^2\right)$ and $\frac{dS}{dx} = S$.

Hence

$$\begin{aligned}
\frac{\partial}{\partial\tau}u(\tau,x) &= re^{r\tau}v(t,S)+e^{r\tau}\left(v_t(t,S)\frac{\partial t}{\partial\tau}+v_S(t,S)\frac{\partial S}{\partial\tau}\right)\\
&= re^{r\tau}v(t,S)-e^{r\tau}v_t(t,S)-e^{r\tau}\left(r-\frac{1}{2}\sigma^2\right)S\,v_S(t,S),\\
\frac{\partial}{\partial x}u(x,\tau) &= e^{r\tau}\left(v_t(t,S)\frac{\partial t}{\partial x}+v_S(t,S)\frac{\partial S}{\partial x}\right)=e^{r\tau}S\,v_S,\text{ and}\\
\frac{\partial^2}{\partial x^2}u(x,\tau) &= e^{r\tau}\left(\frac{\partial S}{\partial x}\cdot v_S+S\left(v_{S,t}(t,S)\frac{\partial t}{\partial x}+v_{SS}(t,S)\frac{\partial S}{\partial x}\right)\right)=e^{r\tau}S\,v_S+e^{r\tau}S^2\,v_{SS},
\end{aligned}$$

so that

$$\frac{1}{2}\sigma^2u_{xx}-u_\tau=e^{r\tau}\left(\frac{1}{2}\sigma^2\,S\,v_S+\frac{1}{2}\sigma^2\,S^2\,v_{SS}-r\,v+v_t+r\,S\,v_S-\frac{1}{2}\sigma^2\,S\,v_S\right)=0$$

by (17.7). It follows (together with the boundary conditions) that

$$\begin{aligned}
u_\tau &= \frac{1}{2}\sigma^2u_{xx} \qquad (17.8)\\
u(0,x) &= V\left(K\cdot e^x\right),\qquad \tau=0,
\end{aligned}$$

which is the *heat equation* (a linear, parabolic partial differential equation) with basic solution given in Exercise 17.1. Its general solution with initial condition u_0 is

$$u(\tau,x)=\frac{1}{\sqrt{2\pi\tau\sigma^2}}\int_{-\infty}^{\infty}u_0(\xi)e^{-\frac{(x-\xi)^2}{2\tau\sigma^2}}d\xi$$

and thus

$$\begin{aligned}
v(t,S) &= e^{-r(T-t)}u\left(T-t,\log\frac{S}{K}+\left(r-\frac{1}{2}\sigma^2\right)(T-t)\right)\\
&= \frac{e^{-r(T-t)}}{\sqrt{2\pi(T-t)\sigma^2}}\int_{-\infty}^{\infty}V\left(K\cdot e^{\xi}\right)e^{-\frac{\left(\log\frac{S}{K}+\left(r-\frac{1}{2}\sigma^2\right)(T-t)-\xi\right)^2}{2(T-t)\sigma^2}}d\xi\\
&= \frac{e^{-r(T-t)}}{\sqrt{2\pi(T-t)\sigma^2}}\int_{-\infty}^{\infty}V\left(K\cdot e^{\log\frac{S}{K}+\left(r-\frac{1}{2}\sigma^2\right)(T-t)+\xi}\right)e^{-\frac{\xi^2}{2(T-t)\sigma^2}}d\xi\\
&= e^{-r(T-t)}\int_{-\infty}^{\infty}V\left(S\cdot e^{\left(r-\frac{1}{2}\sigma^2\right)(T-t)+\sigma\xi}\right)\frac{1}{\sqrt{2\pi(T-t)}}e^{-\frac{\xi^2}{2(T-t)}}d\xi.
\end{aligned}$$

This is exactly Eq. (16.1), which was evaluated explicitly for the European call option in Section 16.2.1.

17.3 TIME TO MATURITY

It is often convenient to formulate the Black–Schones differential equation in terms of *time to maturity* τ, i.e., to introduce $\tau:=T-t$ and to consider the function $\tilde{v}(\tau,S):=v(T-\tau,S)$ instead of $v(t,S)$. The equation then is (cf. (17.7))

$$\begin{aligned}
-\frac{\partial}{\partial\tau}\tilde{v}+\frac{1}{2}\sigma^2S^2\frac{\partial^2}{\partial S^2}\tilde{v}+rS\frac{\partial}{\partial S}\tilde{v}-r\tilde{v}=0\text{ with}&\\
\tilde{v}(0,S)=V(S),\ S\geq 0\text{ and}&\\
\tilde{v}(\tau,0)=0,\ \tau\geq 0.&
\end{aligned}$$

As an example, it holds that (cf. (16.8) and (16.9))

$$\begin{aligned}\tilde{v}_{call}(\tau, S) &\coloneqq S \cdot \Phi(d_+) - K \cdot e^{-r\tau} \cdot \Phi(d_-),\\ \tilde{v}_{put}(\tau, S) &\coloneqq K \cdot e^{-r\tau} \cdot \Phi(-d_-) - S \cdot \Phi(-d_+),\end{aligned}$$

where $d_\pm(\tau, S) = \frac{1}{\sigma\sqrt{\tau}}\left[\left(r \pm \frac{\sigma^2}{2}\right)\tau + \ln\frac{S}{K}\right]$.

17.4 PROBLEMS

Exercise 17.1. *Verify that the kernel* $k_\xi(t,x) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-\frac{1}{2t\sigma^2}(x-\xi)^2}$ *satisfies the heat equation (17.8) (parabolic equation) for every fixed* $\xi \in \mathbb{R}$.

Exercise 17.2. *Show that* $V(t,S)$ *given in (16.3) solves the Black–Scholes equation (17.5).*

Exercise 17.3. *Define* $d_\pm(t,S) \coloneqq \frac{1}{\sqrt{\int_t^T \sigma(u)^2 du}}\left[\int_t^T r(u) \pm \frac{\sigma(u)^2}{2} du + \ln\frac{S}{K}\right]$ *and show that*

$$v(t,S) \coloneqq S \cdot \Phi\big(d_+(t,S)\big) - K \cdot e^{-\int_t^T r(u)du} \cdot \Phi\big(d_-(t,S)\big) \tag{17.9}$$

solves the general time dependent equation (17.5). Note as well that (17.9) is the price for the European call option.

Parameters in Black–Scholes 18

Very often it is assumed that the price process of an underlying follows a standard model for a geometric Brownian motion (GBM), that is the stochastic differential equation

$$dS_t = r\,S_t\,dt + \sigma\,S_t\,dW_t. \tag{18.1}$$

Here,

- r is the interest rate – the symbol r indicates the risk free interest rate (cf. Figure 18.1[1]);
- σ is the parameter to model stochastic volatility and
- dW_t is normally distributed with mean 0 and variance dt, $\mathcal{N}(0, dt)$.

For $\sigma = 0$ this equation reduces to an ordinary differential equation

$$dS_t = r\,S_t\,dt$$

with well-known solution

$$S_t = S_0\,e^{rt}.$$

The solution of the general stochastic differential equation (18.1) is (cf. Section 13.2)

$$S_t = S_0 \cdot e^{\left(r-\frac{1}{2}\sigma^2\right)t+\sigma W_t}$$

and it is obvious that the special case $\sigma = 0$ is naturally contained in this more general formula.

18.1 Estimation of μ or r

Usually the log-return covering the entire period observed is chosen to estimate a stock's or fund's performance, that is,

$$\hat{\mu} \coloneqq \frac{1}{t}\ln\frac{S_t}{S_0}.$$

In the given context of a geometric Brownian motion, however, $\hat{\mu} \coloneqq \frac{1}{t}\ln\frac{S_t}{S_0} = \left(r - \frac{1}{2}\sigma^2\right) + \sigma\frac{W_t}{t}$,[2] and thus

$$\mathbb{E}\,\hat{\mu} = \mathbb{E}\left(\frac{1}{t}\ln\frac{S_t}{S_0}\right) = \mathbb{E}\left(r - \frac{1}{2}\sigma^2 + \sigma\frac{W_t}{t}\right) = r - \frac{1}{2}\sigma^2,$$

and the variance of the estimator $\hat{\mu}$ is

$$\operatorname{var}\hat{\mu} = \operatorname{var}\left(\frac{1}{t}\ln\frac{S_t}{S_0}\right) = \operatorname{var}\left(\left(r - \frac{1}{2}\sigma^2\right) + \sigma\frac{W_t}{t}\right) = \frac{\sigma^2}{t}. \tag{18.2}$$

[1] https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/Historic-Yield-Data-Visualization.aspx

[2] again the *natural logarithm* with basis $e = 2.718\ldots$.

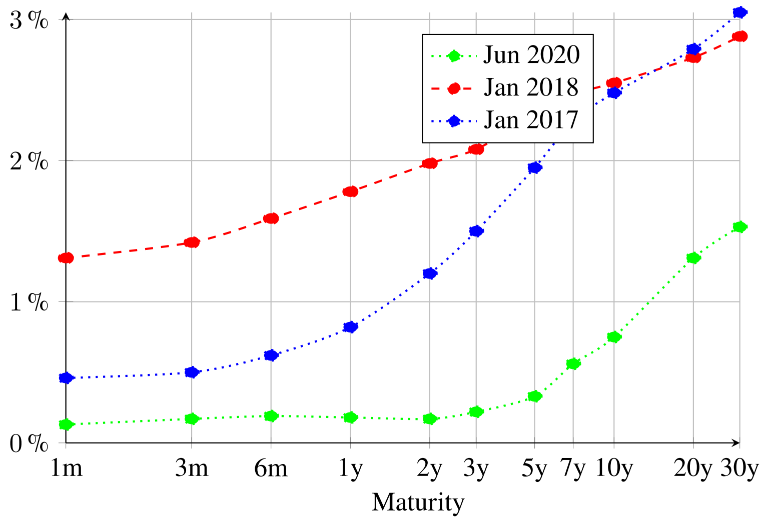


Figure 18.1: The U.S. treasury yield curve: indicator for the risk free rate

Remark 18.1. The variance of the estimator $\hat{\mu}$ thus does *not* tend to 0 unless $t \to \infty$, i.e., we will have to wait for ever (until infinity) to obtain an useful estimator.

In order to get an estimator for μ one might modify the estimator and split the period in n, say, different periods (t_i, t_{i+1}) and try $\hat{\mu} := \frac{1}{t_{i+1}-t_i} \ln \frac{S_{t_{i+1}}}{S_{t_i}}$[3] instead, or even better, the weighted average

$$\hat{\hat{\mu}}_n := \sum_{i=0}^{n-1} \frac{t_{i+1}-t_i}{t_n-t_0} \frac{1}{t_{i+1}-t_i} \ln \frac{S_{t_{i+1}}}{S_{t_i}},$$

where each period (t_i, t_{i+1}) is weighted by its relative length $\frac{t_{i+1}-t_i}{t_n-t_0}$ (note that $\sum_{i=0}^{n-1} \frac{t_{i+1}-t_i}{t_n-t_0} = 1$).

Now notice that $\hat{\hat{\mu}}_n = \frac{1}{t_n-t_0} \sum_{i=0}^{n-1} \ln \frac{S_{t_{i+1}}}{S_{t_i}} = \frac{1}{t} \ln \frac{S_{t_n}}{S_{t_0}} = \hat{\mu}$. This, however, is no improvement over $\hat{\mu}$ at all, and in particular the estimator $\hat{\hat{\mu}}_n$ satisfies

$$\operatorname{var} \hat{\hat{\mu}}_n = \operatorname{var} \frac{1}{t} \ln \frac{S_t}{S_0} = \frac{\sigma^2}{t} \tag{18.3}$$

as well. This quantity is independent of n and does *not* tend to 0, whenever $n \to \infty$. Both, $\hat{\mu}$ and $\hat{\hat{\mu}}_n$ thus are *useless* as estimators for the drift μ.

Remark. Summarizing, we don't have an estimator for μ and this is another reason why we have to get μ from somewhere else: we have to involve the risk free interest rate in place of the drift μ, that is, $\mu \leftarrow r$.

[3]It should be noticed that $\frac{1}{t_{i+1}-t_i} \ln \frac{S_{t_{i+1}}}{S_{t_i}}$ can be interpreted as return per time interval, in which t_i are being measured (usually in years). Then

$$\frac{1}{t_n-t_0} \ln \frac{S_{t_n}}{S_{t_0}} = \sum_{i=0}^{n-1} \frac{t_{i+1}-t_i}{t_n-t_0} \cdot \frac{1}{t_{i+1}-t_i} \ln \frac{S_{t_{i+1}}}{S_{t_i}},$$

that is to say the weighted average of all these returns reflects the average return (because $\sum_{i=0}^{n-1} \frac{t_{i+1}-t_i}{t_n-t_0} = 1$) during the entire time interval.

18.2 ESTIMATION OF THE VOLATILITY σ

Given observations of the stock S_{t_i} at subsequent times t_i *in years* ($[t_i]$ = *year*) the stochastic volatility may be estimated using the unbiased estimator (Bessel correction $n-1$)

$$\hat{\sigma}_n^2 = \left(\frac{1}{n-1} \sum_{i=0}^{n-1} \frac{\left(\ln \frac{S_{t_{i+1}}}{S_{t_i}}\right)^2}{t_{i+1}-t_i} \right) - \frac{1}{n-1} \frac{\left(\ln \frac{S_{t_n}}{S_{t_0}}\right)^2}{t_n - t_0}$$

($\frac{n-1}{n}\hat{\sigma}_n^2$ is actually the maximum likelihood estimator). For the typical and particular situation $t_{i+1} - t_i = \frac{1}{m}$ this reads

$$\hat{\sigma}_n^2 = \frac{m}{n-1} \sum_{i=0}^{n-1} \left(\ln \frac{S_{t_{i+1}}}{S_{t_i}}\right)^2 - \frac{m}{n-1} \frac{\left(\ln \frac{S_{t_n}}{S_{t_0}}\right)^2}{n}.$$

Note now that

$$\begin{aligned}
\mathbb{E}\left(\ln \frac{S_t}{S_0}\right)^2 &= \mathbb{E}\left(\left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)^2 \\
&= \left(r - \frac{1}{2}\sigma^2\right)^2 t^2 + 2\left(r - \frac{1}{2}\sigma^2\right) t\sigma\, \mathbb{E}\, W_t + \sigma^2\, \mathbb{E}\, W_t^2 \\
&= \left(r - \frac{1}{2}\sigma^2\right)^2 t^2 + \sigma^2 t.
\end{aligned}$$

Summing up the terms in above's estimator thus

$$\begin{aligned}
\mathbb{E}\, \hat{\sigma}_n^2 =& \frac{1}{n-1} \sum_{i=1}^{n} \frac{\left(r - \frac{1}{2}\sigma^2\right)^2 (t_{i+1}-t_i)^2 + \sigma^2 (t_{i+1}-t_i)}{t_{i+1}-t_i} \\
&- \frac{1}{n-1} \frac{\left(r - \frac{1}{2}\sigma^2\right)^2 (t_n - t_0)^2 + \sigma^2 (t_n - t_0)}{t_n - t_0} \\
=& \frac{1}{n-1}\left(\left(r - \frac{1}{2}\sigma^2\right)^2 (t_n - t_0) + n \cdot \sigma^2\right) \\
&- \frac{1}{n-1}\left(\left(r - \frac{1}{2}\sigma^2\right)^2 (t_n - t_0) + \sigma^2\right) = \frac{n\sigma^2 - \sigma^2}{n-1} = \sigma^2,
\end{aligned}$$

that is to say this estimator $\hat{\sigma}_n^2$ is unbiased, indeed. Moreover, it can be shown that

$$\operatorname{var} \hat{\sigma}_n^2 \xrightarrow[n\to\infty]{} 0,$$

a very useful property in contrast to $\hat{\mu}_n$, cf. (18.2) and (18.3).

Remark. In contrast to μ we do have a proper estimator for the volatility σ.

18.3 IMPLIED VOLATILITY

The implied volatility of an option contract is that value of the volatility of the underlying instrument which, when input in an option pricing model (such as Black–Scholes) will return a theoretical value equal to the current market price of the option.

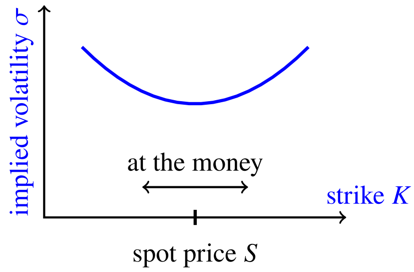


Figure 18.2: Volatility smile

Remark 18.2. The implied volatility is thus

> *a wrong number which, plugged into the wrong formula, gives the right answer*.

Volatility smiles (cf. Figure 18.2) are implied volatility patterns that arise in pricing financial options. In particular for a given expiration, options whose strike price differs substantially from the underlying asset's price command higher prices (and thus implied volatilities) than what is suggested by standard option pricing models.

18.4 PROBLEMS

Exercise 18.1. *Deutsche Bank had the following closing prices at consecutive months in 2016:*

t_i	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Jan*
S_{t_i} / €	*25.2*	*22.8*	*24.2*	*24.9*	*25.0*	*24.8*	*25.1*	*25.2*	*23.3*	*23.0*	*22.8*	*24.0*	*24.4*

Give its volatility.

Exercise 18.2. *Compute the volatility of a stock of your choice based on* daily *observations during the last month.*

19 The Girsanov Theorem: Change of Measure

This lecture follows Pflug [18].

We consider a stochastic process X with drift μ_t under P, i.e.,

$$dX_t = \mu_t dt + \sigma_t d\overline{W}_t.$$

We are interested in a new probability measure Q on the same sample space such that the process X has the same diffusion, but no drift, i.e.,

$$dX_t = \sigma_t dW_t$$

with W being a Wiener process under Q. Further, $Q \ll P$ and $P \ll Q$ (absolutely continuous).

The density of Q with respect to P (under P) is given by the Girsanov formula

$$\left.\frac{dQ}{dP}\right|_P = \exp\left(-\int_0^T \frac{\mu_t}{\sigma_t} d\overline{W}_t - \frac{1}{2}\int_0^T \left(\frac{\mu_t}{\sigma^2}\right)^2 dt\right),$$

where $\overline{W}$ is a Wiener process under P.

Proof. We start with a normal density with mean μ and covariance matrix Σ:

$$f(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} e^{-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)}.$$

If P has density $f(x;\mu,\Sigma)$ and Q has density $f(x;0,\Sigma)$, then

$$\begin{aligned}\frac{dP}{dQ} = \frac{f(x;\mu,\Sigma)}{f(x;0,\Sigma)} &= \exp\left(-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu) + \frac{1}{2}x^t\Sigma^{-1}x\right)\\ &= \exp\left(x^\top\Sigma^{-1}\mu - \frac{1}{2}\mu^\top\Sigma^{-1}\mu\right).\end{aligned}$$

Notice that

$$\mathbb{E}_P\left[h(X)\right] = \mathbb{E}_{\mathbb{Q}}\left[h(X)\frac{dP}{dQ}(X)\right],$$

so that $x \sim \mathcal{N}(0,\Sigma)$ under Q.

We illustrate the proof of the Girsanov formula only for deterministic μ_t and σ. For the given process,

$$\begin{aligned}dX_t &= \mu_t dt + \sigma_t d\overline{W}_t \qquad \text{under } P \text{ and}\\ dX_t &= \sigma_t dW_t \qquad \text{under } Q\end{aligned}$$

we choose a partition $(t_1,\dots,t_n)$ and set

$$D_{t_i} := X_{t_{i+1}} - X_{t_i}.$$

Note that

$$\begin{aligned}\mathbb{E}_P D_{t_i} &= \mu_t(t_{i+1}-t_i) \text{ and}\\ \operatorname{var} D_{t_i} &= \sigma_{t_i}^2(t_{i+1}-t_i).\end{aligned}$$

Then $\frac{dP}{dQ}$ under Q is

$$\frac{f_\mu(D_1,\dots D_n)}{f_0(D_1,\dots D_n)} = \exp\left(\sum \frac{\mu_{t_i}(t_{i+1}-t_i)}{\sigma_{t_i}(t_{i+1}-t_i)}(X_{t_{i+1}}-X_{t_i}) - \frac{1}{2}\sum \frac{\mu_{t_i}^2(t_{i+1}-t_i)^2}{\sigma_{t_i}^2(t_{i+1}-t_i)}\right)$$

which converges, as $n \to \infty$, to

$$\frac{f_\mu(D_1,\dots D_n)}{f_0(D_1,\dots D_n)} = \exp\left(\int \frac{\mu_t}{\sigma_t}dX_t - \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right).$$

We need, however, $\frac{dQ}{dP}$ under Q, which is the inverse,

$$\frac{dQ}{dP} = \exp\left(-\int \frac{\mu_t}{\sigma_t}dX_t + \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right).$$

Now we compute $\frac{dQ}{dP}$ under P. Since

$$dX_t = \sigma_t dW_t \text{ under } Q \text{ and}$$
$$dX_t = \mu_t dt + \sigma_t d\overline{W}_t \text{ under } P$$

we have that

$$\sigma_t dW_t = \mu_t dt + \sigma_t d\overline{W}_t,$$

where $\overline{W}$ is a Wiener process under P. Therefore, under P,

$$\begin{aligned}\frac{dQ}{dP} &= \exp\left(-\int \frac{\mu_t}{\sigma_t}dX_t + \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right)\\ &= \exp\left(-\int \frac{\mu_t}{\sigma_t^2}\sigma_t dX_t + \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right)\\ &= \exp\left(-\int \frac{\mu_t}{\sigma_t}d\overline{W}_t - \int \frac{\mu_t}{\sigma_t^2}\mu_t dt + \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right)\\ &= \exp\left(-\int \frac{\mu_t}{\sigma_t}d\overline{W}_t - \frac{1}{2}\int \frac{\mu_t^2}{\sigma_{t_i}^2}dt\right),\end{aligned}$$

which is the claim. □

20 Important Stochastic Processes

Stochastic processes are employed to model the evolution of stock. However, they can be used to model the evolution of interest rates (short-rate) equally well. This chapter presents general process first and then addresses particular short-rated models used in mathematical finance.

20.1 ORNSTEIN–UHLENBECK

The stochastic differential equation for the mean-reverting Ornstein[1]–Uhlenbeck[2] process is

$$dx_t = \theta(\mu - x_t)dt + \sigma\, dW_t. \tag{20.1}$$

To solve this linear, non-homogeneous equation compare (20.1) and (13.2), i.e., set $r(t) \coloneqq -\theta$, $a(t) = \theta \cdot \mu$, $\sigma(t) = 0$ and $b(t) = \sigma$. Then $\zeta_t = 0$ and $Z_t = e^{-\theta t}$. The closed–form, explicit solution is

$$\begin{aligned} x_t &= e^{-\theta t}\left(x_0 + \int_0^t e^{\theta u}\theta\mu\, du + \int_0^t \sigma e^{\theta u}\, dW_u\right) \\ &= x_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma \int_0^t e^{-\theta(t-u)}\, dW_u. \end{aligned} \tag{20.2}$$

Properties of the Ornstein–Uhlenbeck process:

- $\mathbb{E}\, x_t = x_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right)$, with long term mean $\mathbb{E}\, x_t \xrightarrow[t\to\infty]{} \mu$.
- The covariance is

$$\begin{aligned} \operatorname{cov}(x_s, x_t) &= \mathbb{E}\left(x_s - \mathbb{E}\, x_s\right)\left(x_t - \mathbb{E}\, x_t\right) = \mathbb{E}\,\sigma \int_0^s e^{-\theta(s-u)} dW_u \cdot \sigma \int_0^t e^{-\theta(t-u)} dW_u \\ &= \sigma^2 e^{-\theta(s+t)}\, \mathbb{E} \int_0^s e^{\theta u} dW_u \cdot \int_0^t e^{\theta u} dW_u \\ &\underset{(11.2)}{=} \sigma^2 e^{-\theta(s+t)}\, \mathbb{E} \int_0^{s\wedge t} e^{2\theta u} du = \sigma^2 e^{-\theta(s+t)} \frac{1}{2\theta}\left(e^{2\theta(s\wedge t)} - 1\right), \\ &= \frac{\sigma^2}{2\theta}\left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right), \end{aligned} \tag{20.3}$$

  where we have used Itō's isometry (Corollary 11.8) and the identity $s \wedge t = \frac{1}{2}(s + t - |s - t|)$.

- The marginal distribution at time t is $x_t \sim \mathcal{N}\left(x_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right),\ \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t}\right)\right)$, with limiting distribution $\mathcal{N}\left(\mu,\ \frac{\sigma^2}{2\theta}\right)$. The process thus is said to possess an invariant measure, which is $\mathcal{N}\left(\mu,\ \frac{\sigma^2}{2\theta}\right)$.

[1] Leonard Ornstein, 1880–1941
[2] George Eugene Uhlenbeck, 1900–1988

- The process x_t is *mean reverting*.
- Time–change: the process

$$x_t' = x_0 e^{-\theta t} + \mu \left(1 - e^{-\theta t}\right) + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta t} W_{e^{2\theta t}-1} \tag{20.4}$$

  solves the Ornstein–Uhlenbeck stochastic differential equation (20.1) as well. Indeed, using (10.2), the covariance of the process x_t' is

$$\operatorname{cov}(x_s', x_t') = \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left((e^{2\theta s} - 1) \wedge (e^{2\theta t} - 1)\right) = \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta(s\wedge t)} - 1\right) = (20.3),$$

  from which follows that the processes x_t and x_t' are equal in distribution, as both are normally distributed.
- Note, that the time-changed formula (20.4)—in contrast to (20.2)—does not involve any integration with respect to the Brownian motion.

20.2 BLACK MODEL

Consider the modle for the interest rate which follows an Ornstein–Uhlenbeck process $dr_t = \theta(\mu - r_t)dt + \sigma dW_t$, cf. (20.1); its marginal distribution is

$$r_t \sim \mathcal{N}\left(\underbrace{r_0 e^{-\theta t} + \mu\big(1 - e^{-\theta t}\big)}_{=:\mu_t},\ \underbrace{\frac{\sigma^2}{2\theta}\big(1 - e^{-2\theta t}\big)}_{=:\sigma_t^2}\right).$$

For a call option with payoff $r \mapsto (r - K)_+$, the prices is

$$\begin{aligned}
\mathbb{E}\,(r_t - K)_+ &= \int_K^\infty (x - K)_+ \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{1}{2\sigma_t^2}(x-\mu_t)^2} dx = \int_{\frac{K-\mu_t}{\sigma_t}}^\infty (\mu_t + \sigma_t x - K) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
&= (\mu_t - K) \int_{\frac{K-\mu_t}{\sigma_t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx + \sigma_t \int_{\frac{K-\mu_t}{\sigma_t}}^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \\
&= (\mu_t - K)\, \Phi\left(-\frac{K - \mu_t}{\sigma_t}\right) + \frac{\sigma_t}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{K-\mu_t}{\sigma_t}\right)^2}.
\end{aligned}$$

20.3 COX–INGERSOLL–ROSS

The Cox-Ingersoll-Ross model

$$dr_t = a\,(b - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t \tag{20.5}$$

has an explicit solution as well, although more complicated. The marginal distribution r_t is a non-central χ^2-distribution (and thus not normal).

The continuous version is discretized as $r_{t+\Delta t} = r_t + a\,(b - r_t)\,\Delta t + \sigma\sqrt{r_t \Delta t}\,\varepsilon_t$, or

$$\frac{r_{t+\Delta t} - r_t}{\sqrt{r_t}} = \frac{ab\Delta t}{\sqrt{r_t}} - a\sqrt{r_t}\Delta t + \sigma\sqrt{\Delta t}\,\varepsilon_t,$$

which is eligible for linear regression to uncover the parameters a, b and σ.

20.4 HESTON MODEL

Here, $dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t$, where ν_t is a stochastic volatility following the Cox–Ingersoll–Ross model (20.5).

20.5 ONE-FACTOR SHORT RATE MODELS

20.5.1 Merton's model (1973)

$$r_t = r_0 + at + \sigma dW_t.$$

20.5.2 Vašiček model (1977)

A particular variant of the Ornstein–Uhlenbeck process is the Vasicek[3] model,

$$dr_t = a\left(b - r_t\right) dt + \sigma dW_t.$$

20.5.3 Rendleman–Bartter (1980)

$$dr_t = \theta r_t dt + \sigma r_t dW_t.$$

20.5.4 Cox–Ingersoll–Ross (CIR, 1985)

$$dr_t = a\left(b - r_t\right) dt + \sigma\sqrt{r_t} dW_t.$$

20.5.5 Ho-Lee model (1986)

$$dr_t = \theta_t dt + \sigma dW_t.$$

20.5.6 Hull–White model (1990)

$$dr_t = \left(\theta_t - \alpha r_t\right) dt + \sigma_t dW_t.$$

20.5.7 Black–Derman–Toy (1990)

$$d\ln r_t = \left(\theta_t + \frac{\sigma_t'}{\sigma_t}\ln r_t\right) dt + \sigma_t dW_t,$$

where σ_t' is a parameter (more precisely, a function taking taking the role of a parameter).

20.5.8 Black–Karasinski (1991)

$$d\ln r_t = \left(\theta_t - \phi_t \ln r_t\right) dt + \sigma_t dW_t$$

20.5.9 Kalotay–Williams–Fabozzi (1993)

$$d\ln r_t = \theta_t dt + \sigma_t dW_t$$

the log-normal version of Hull–White and a special case ($\sigma_t' \equiv 0$) of Black–Derman–Toy.

[3] Oldřich Alfons Vašíček, 1942, Czech mathematician

20.6 MULTI-FACTOR SHORT-RATE MODELS

20.6.1 Longstaff–Schwartz model (1992)

$$dr_t = (\mu X_t + \theta_t Y_t)\, dt + \sigma_t \sqrt{Y} W_t^{(3)},$$

where

$$dX_t = (a_t - bX_t)dt + \sqrt{X_t}c_t dW_t^{(1)} \text{ and}$$
$$dY_t = (d_t - eY_t)dt + \sqrt{Y_t}f_t dW_t^{(2)}.$$

(the Brownian motions $W_t^{(1)}$, $W_t^{(2)}$ and $W_t^{(3)}$ are independent.

Derivatives

21

Derivatives are financial instruments. Financial instruments, in general, can be categorized by form depending on whether they are cash instruments or derivative instruments.

- ▹ Cash instruments are financial instruments whose value is determined directly by markets. Examples are securities, loans and deposits.
- ▹ Derivative instruments are financial instruments which derive their value from the value and characteristics of one or more underlying assets.

21.1 CLASSIFICATION

Derivative instruments may be classified in general in three dimensions, which are:

(i) The market, on which they are traded

- ▹ Over-the-counter (OTC) derivatives, and
- ▹ (Stock) Exchange traded derivatives (ETD).

(ii) The relationship between the underlying and the derivative, eg.

- ▹ Future/ Forwards
- ▹ Option
- ▹ Swap

(iii) The type of underlying, eg.

- ▹ Equity,
- ▹ Foreign exchange (FX),
- ▹ Interest rate,
- ▹ Credit (loans),
- ▹ Commodity, ...

Some common examples of derivatives are exemplary and schematically listed in this following table:

DERIVATIVES	CONTRACT TYPE				
	Future/ Forward		Swap	Option	
UNDERLYING	ETD	OTC	OTC	ETD	OTC
Equity	DJIA-index future, Single stock future	Repo, Back-toback	Equity swap	Option on DJIA-index future	Stock Option, Warrant
Interest Rate	€/ $ future, Euribor future	Forward rate agreement	Interest rate swap	Option on Eurodollar future, Option on Euribor future	Interest Rate Cap and Floor, Swaption, Basis Swap, Bond option
Credit	Bond future	Repo	Credit default swap, Total return swap	Option on bond future	Credit Default Option
Foreign Exchange	Currency future	Currency forward	Currency swap	Option on currency future	Currency Option
Commodity			Commodity swap		
Insurance				Weather Derivatives	

21.2 FUTURE/ FORWARD

The main characteristics of futures[1] are

- a specified commodity of (standardized) quality (*commodity future*) or financial contract (*financial future*);
- a specified quantity;
- a fixed date of delivery T in the future (the *delivery date*);
- the type of *settlement*, which can for example be cash settlement or physical settlement,
- the futures price is determined and fixed at the beginning of the contract (*delivery price* K), however, the specified amount is due and will be paid in future at expiry of the futures contract.
- *Both* parties have the obligation to fulfill the contract (the cash and delivery) in future.

Given these characteristics of a future it is pretty obvious that the *delivery price* K of the future contract is simply found by compounding,

$$K := S_t \cdot (1+i)^{T-t};$$

Here, is the current price of the (non-dividend paying) underlying

- S_t is the underlying

[1] For trading futures and forwards see for example Chicago Merchandise Exchange, www.cmegroup.com.

- ▹ $T - t$ is the term of the contract and
- ▹ i the risk free return.

In a situation where there is no market value available but can somehow be estimated, the fair delivery price is

$$K \coloneqq \mathbb{E}\, S_T.$$

Given this future delivery price K the value of the future/ forward will vary in time, as the underlying (and the interest etc.) vary. Today's present value of the contract is

$$p \coloneqq \mathbb{E}\, S_t - \frac{K}{(1+i)^{T-t}},$$

and the payoff is $S_T - K$ (long position.)

In some situations a deposit of, say, 5 % of the contracts amount is required at the beginning of the contract as a deposit (*initial margin*).

21.3 SWAPS

A swap is an agreement between two parties to exchange flows of payments from a given security (the asset) for a different set of cash-flows . The agreement defines how those cash flows will be computed and when the are due.

21.3.1 Asset Swap

21.3.2 Interest Rate Swap

An interest rate swap is a derivative, where two parties agree to exchange a nominal value in future. Usually interest payments are fixed in such way, that one party pays a fixed interest rate, the other party will pay a variable interest rate ("plain vanilla swap"). The variable rate is derived from reference interest rates in inter-banking business.

Interest rate swaps are used to secure positions in the balance sheet against changes of interest rates, but for speculation as well.

The price is $PV_{swap} = PV_{fix} - PV_{variable}$.

21.3.3 Credit Default Swaps (CDS)

A credit default swap (CDS) transfers a potential credit default to another partner, who will receive a premium in exchange.

The premium is determined by the probability of the default and the potential loss, respective (that is to say the credit rating) and is usually paid in installments.

In case of a default the CDS stops, and the premium payment as well. In case no defined credit event occurs during the period, the CDS expires.

The regular premium for constant q (credit default's probability) is

$$\frac{{}_nA}{\ddot{a}_n} = \frac{q}{1+i}.$$

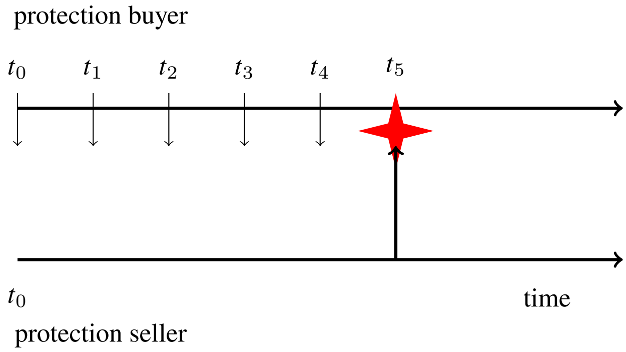


Figure 21.1: Payments of the protection buyer and seller

21.4 SWAPTIONS

A swaption is an option, giving the right to the buyer for a premium to enter a interest-swap at a given date (European swaption).

21.5 INTEREST RATE CAP AND FLOORS

Interest rate cap and floor are interest rate derivatives. They offer an upper and a lower bound for the interest rate.

- An interest rate cap is a derivative in which the buyer receives payments at the end of each period in which the interest rate exceeds the agreed strike price.
- An interest rate floor is a series of European put options or floorlets on a specified reference rate, usually LIBOR. The buyer of the floor receives money if on the maturity of any of the floorlets, the reference rate fixed is below the agreed strike price of the floor.

21.6 EXOTIC OPTIONS

21.7 CREDIT DERIVATIVES

A credit derivative is a derivative whose value is derived from the credit risk on an underlying bond, loan or other financial asset. In this way, the credit risk is on an entity other than the counter parties to the transaction itself. This entity is known as the reference entity and may be a corporate, a sovereign or any other form of legal entity which has incurred debt. Credit derivatives are bilateral contracts between a buyer and seller under which the seller sells protection against the credit risk of the reference entity.

21.7.1 Credit Linked Notes (CLN)

Credit Linked Notes sind Wertpapiere, deren Rückzahlungsprofil zum Beispiel abhängig vom Eintritt sogenannter Kreditereignisse bei einem oder mehreren Referenzschuldnern ist. Kann also der Referenzschuldner seinen Zahlungsverpflichtungen nicht nachkommen, wird dieser Ausfall auf den Anleihegläubiger übertragen. Tritt kein Kreditereignis ein, kommt der Anleger in den Genuss einer attraktiven Rendite.

21.7.2 Collateralized Debt Obligation (CDO)

Collateralized debt obligations (CDOs) are a type of structured asset-backed security (ABS) with multiple "tranches" that are issued by special purpose entities and collateralized by debt obligations including bonds and loans. Each tranche offers a varying degree of risk and return so as to meet investor demand. CDOs' value and payments are derived from a portfolio of fixed-income underlying assets. CDO securities are split into different risk classes, or tranches, whereby "senior" tranches are considered the safest securities. Interest and principal payments are made in order of seniority, so that junior tranches offer higher coupon payments (and interest rates) or lower prices to compensate for additional default risk. In simple terms, think of a CDO as a promise to pay cash flows to investors in a prescribed sequence, based on how much cash flow the CDO collects from the pool of bonds or other assets it owns. If cash collected by the CDO is insufficient to pay all of its investors, those in the lower layers (tranches) suffer losses first.

21.8 REPO

A repurchase agreement (also known as a repo, RP, or sale and repurchase agreement) is the sale of securities together with an agreement for the seller to buy back the securities at a later date. The repurchase price should be greater than the original sale price, the difference effectively representing interest, sometimes called the repo rate. The party that originally buys the securities effectively acts as a lender. The original seller is effectively acting as a borrower, using their security as collateral for a secured cash loan at a fixed rate of interest.

A repo is equivalent to a spot sale combined with a forward contract. The spot sale results in transfer of money to the borrower in exchange for legal transfer of the security to the lender, while the forward contract ensures repayment of the loan to the lender and return of the collateral of the borrower. The difference between the forward price and the spot price is effectively the interest on the loan while the settlement date of the forward contract is the maturity date of the loan.

22 The Value-at-Risk and Risk Measures

22.1 THE BASEL II FRAMEWORK

For market risk, the preferred approach is value-at-risk (V@R). Banks will have flexibility in devising the precise nature of their models, but the following minimum standards will apply for the purpose of calculating their capital charge.

(i) "Value-at-risk" must be computed on a daily basis.

(ii) A 99th percentile, one-tailed confidence interval is to be used.

(iii) An instantaneous price shock equivalent to a 10 day movement in prices is to be used.

(iv) The historical observation period is a minimum length of one year.

(v) Banks should update their data sets no less frequently than once every three months.

V@R cushion: The Basle Committee has decided to establish a cushion of this type by requiring a multiplication factor of 3 to be applied to the Value-at-Risk calculation.

Definition 22.1. We say that

- ▹ $X_1 \leq X_2$ provided that $X_1(\omega) \leq X_2(\omega)$ a.s.
  - $X_1 \leq_{FST} X_2$ provided that $\mathbb{P}[X_1 \leq r] \geq \mathbb{P}[X_2 \leq r]$ for all r, or equivalently $q_\alpha(X_1) \leq q_\alpha(X_1)$;
  - $X_2 \leq_{SST} X_2$ provided that $\int_{-\infty}^{r} \mathbb{P}[X_1 \leq r] \leq \int_{-\infty}^{r} \mathbb{P}[X_2 \leq r]$ for all r;
  - $X_2 \leq_{n-ST} X_2$ provided that $\int_{-\infty}^{r} \cdots \int_{-\infty}^{r_n} \mathbb{P}[X_1 \leq r_n] \leq \int_{-\infty}^{r} \cdots \int_{-\infty}^{r_n} \mathbb{P}[X_2 \leq r_n]$ for all r;

  Financial quantities

  - Sharpe index: $\frac{\mathbb{E}X - r_f}{\sigma}$;

Probabilities for Credit and Insurance Risk

23

Conceptually, insurance risk and credit risk are the same:

- ▹ Payment of an installment corresponds to premium payment, and
- ▹ Credit default corresponds to death (or any insured event).

So we do not distinguish between credit risk and insurance risk and treat them analogously here.
The chapter follows Gerber [7].

23.1 THE EVENT OF DEATH/ CREDIT DEFAULT

Consider a person (a loan) of age x. Let T be a random variable giving the *future lifetime* of the life (default). The pdf is

$$G(t) = P(T \le t)$$

and the density, if it exists, is

$$g(t)\,dt = P(t \le T \le t + dt).$$

A common and accepted notation in insurance is

$$ {}_tq_x \coloneqq G(t) \text{ and } {}_tp_x \coloneqq 1 - G(t),$$

further the symbol

$$\begin{aligned} {}_{s|t}q_x &\coloneqq P(s < T \le s+t) \\ &= G(s+t) - G(s) = {}_{s+t}q_x - {}_sq_x. \end{aligned}$$

One further defines

$${}_tp_{x+s} \coloneqq P(T > s+t \mid T > s) = \frac{1 - G(s+t)}{1 - G(s)}$$

and

$${}_tq_{x+s} \coloneqq P(T \le s+t \mid T > s) = \frac{G(s+t) - G(s)}{1 - G(s)},$$

the conditional probabilities of dying within t years, provided that the person (loan) survived s years. Important relations include

$${}_{s+t}p_x = 1 - G(s+t) = (1 - G(s))\,\frac{1 - G(s+t)}{1 - G(s)} = {}_sp_x \cdot {}_tp_{x+s}$$

and

$${}_{s|t}q_x = G(s+t) - G(s) = (1 - G(s))\,\frac{G(s+t) - G(s)}{1 - G(s)} = {}_sp_x \cdot {}_tq_{x+s}.$$

Example 23.1. The expected further lifetime (remaining lifetime) is

$$\mathbb{E}T = \int_0^\infty t\,g(t)\,dt = \int_0^\infty 1 - G(t)dt = \int_0^\infty {}_tp_x dt.$$

Remark. For $t = 1$ the index is omitted, that is actuaries write

$${}_1p_x = p_x \text{ and } {}_1q_x = q_x.$$

23.1.1 Force of mortality

... is defined as

$$\mu_T(t) \coloneqq \frac{g(t)}{1-G(t)} = -\frac{d}{dt}\log\big(1-G(t)\big)$$

such that, e.g., $\mathbb{E}\,T = \int_0^\infty t\; {}_tp_x \mu_T(t)dt$.

It holds that $\mu_T(t) = -\frac{d}{dt}\log\; {}_tp_x$, whence

$${}_tp_x = e^{-\int_0^t \mu_T(s)ds}.$$

de Moivre (1724) postulated $\mu(t) = \frac{1}{\omega-(x+t)}$

Gompertz (1824) postulated $\mu(t) = Bc^{x+t}$

Makerham (1860) modified to $\mu(t) = A + Bc^{x+t}$

Weibull (1939) modified to $\mu(t) = k\,(x+t)^n$

23.1.2 Mortality within one year

Consider (cf. Footnote 1 on page 93)

$$K \coloneqq \lfloor T \rfloor,$$

so that $P(K=k) = P(k \le T < k+1) = {}_kp_x \cdot q_{x+k}$. Moreover define $S \coloneqq T-K$ and $S^{(m)} \coloneqq \frac{1}{m}\lfloor mS+1\rfloor$.

A usual assumption now is ${}_uq_x = u\cdot q_x$ such that

$$P\left[S^{(m)} = \frac{j}{m}\right] = \frac{1}{m}, \quad j = 1,2\dots m,$$

such that $S^{(m)}$ and K are independent. Note further, that

$$\mathbb{E}\,v^{S^{(m)}-1} = \sum_{j=1}^m \frac{1}{m}v^{j/m-1} = \frac{1}{m}v^{1/m-1}\frac{1-v}{1-v^{1/m}} = \frac{i}{i^{(m)}} \tag{23.1}$$

(cf. (2.3) and Exercise 23.1) and

$$\mathbb{E}\left(1-S^{(m)}\right)v^{S^{(m)}-1} = \frac{i}{i^{(m)}}\left(\frac{1}{d}-\frac{1}{d^{(m)}}\right) \tag{23.2}$$

and

$$\mathbb{E}\left(1-S^{(m)}\right)v^{1-S^{(m)}} = \frac{d-v\,i^{(m)}}{i^{(m)}d^{(m)}} \tag{23.3}$$

23.1.3 Remaining Life Expectancy

The remaining life expectancy is

$$e_x \coloneqq \mathbb{E}\left(K+\frac{1}{2}\right) = \sum_{k=0} {}_kp_x - \frac{1}{2} = \ddot{a}_x - \frac{1}{2},$$

where $\ddot{a}_x$ is computed with interest rate $i = 0$. The variance of the remaining life expectation can be treated as above to give

$$\begin{aligned}
\operatorname{var} K &= \operatorname{var}\left(K + \tfrac{1}{2}\right) \\
&= \sum_{k=0} \left(k + \tfrac{1}{2}\right)^2 \ {}_k p_x \, q_{x+k} - e_x^2 \\
&= \sum_{k=0} \left(k + \tfrac{1}{2}\right)^2 \ {}_k p_x - \sum_{k=0} \left(k + \tfrac{1}{2}\right)^2 \ {}_{k+1} p_x - e_x^2 \\
&= \frac{1}{4} + \sum_{k=0} \left(k + \tfrac{1}{2}\right)^2 \ {}_k p_x - \sum_{k=0} \left(k - \tfrac{1}{2}\right)^2 \ {}_k p_x - e_x^2 \\
&= \frac{1}{4} + 2 \sum_{k=0} k \ {}_k p_x - \left(\ddot{a}_x - \tfrac{1}{2}\right)^2 = 2\ddot{a}_x^{\text{inc}} - \ddot{a}_x - \ddot{a}_x^2.
\end{aligned}$$

23.2 PROBLEMS

Exercise 23.1. *Verify (23.1) and (23.2).*

24 Loans and Lump Sum Insurance Premiums

Life Insurance

Recall that $q_{x+k} = P[K = k \mid K \geq x]$ is the probability of default in year k, provided non-default in the earlier years.

Let the payoff be $Z = v^{K+1}$, where the default happens in year K. Then $\mathbb{P}\left[Z = v^{k+1}\right] = {}_k p_x \cdot q_{x+k}$. Note that

- $A_x = \mathbb{E}_x Z = \mathbb{E}_x \left[v^{K+1}\right] = \sum_{k=0} v^{k+1}\, {}_k p_x \cdot q_{x+k}$, and
- $\operatorname{var}[Z] = \mathbb{E} Z^2 - A_x^2 = \sum_{k=0} {}_k p_x \cdot q_{x+k} \cdot v^{2k+2} - A_x^2$.
- Special case for $q_k = q$:
  - $A_x = \sum_{k=0} v^{k+1} q (1-q)^k = \frac{qv}{1-v(1-q)}$, and
  - $\operatorname{var} Z = \mathbb{E}\left[Z^2\right] - A_x^2 = \sum_{k=0} q(1-q)^k \cdot v^{2k+2} - A_x^2 = \frac{qv^2}{1-(1-q)v^2} - \left(\frac{qv}{1-v(1-q)}\right)^2$
- $\operatorname{var} Z = \mathbb{E}\left(Z^2 \cdot \mathbb{1}_{\{K \leq n\}}\right) - A_{x:n}^2 = \sum_{k=0}^{n-1} {}_k p_x \cdot q_{x+k} \cdot v^{2k+2} - A_{x:n}^2$.
- ${}_n A_x = \mathbb{E}\left(Z \cdot \mathbb{1}_{\{K<n\}}\right) = \mathbb{E}\left(v^{K+1} \cdot \mathbb{1}_{\{K<n\}}\right) = \sum_{k=0}^{n-1} v^{k+1}\, {}_k p_x \cdot q_{x+k}$
- $$ {}_n A_x^{(m)} = \mathbb{E}\, v^{K+S^{(m)}} = \mathbb{E}\, v^{K+1}\, \mathbb{E}\, v^{S^{(m)}-1} = \frac{i}{i^{(m)}}\, {}_n A_x \tag{24.1}$$

  (as K and $S^{(m)}$ are independent, cf. (23.1))

Pure Endowment

Consider the pay-off function $Z \coloneqq v^n\, \mathbb{1}_{\{n,n+1,\dots\}}(K) = \begin{cases} 0 & K < n \\ v^n & k \geq n \end{cases}$. Then $\mathbb{E} Z = {}_n p_x \cdot v^n =: {}_n E_x$.

Here, $\operatorname{var} Z = {}_n p_x\, v^{2n} - \left({}_n p_x\, v^n\right)^2 = {}_n p_x\, {}_n q_x\, v^{2n}$ and $\frac{d}{d\delta}\, {}_n E_x = -n \cdot_n E_x$. The duration thus is $D = -\frac{\frac{d}{d\delta} {}_n E_x}{{}_n E_x} = n$.

Endowment

The payoff function $Z \coloneqq v^{K+1 \wedge n} = \begin{cases} v^{K+1} & K < n, \\ v^n & k \geq n, \end{cases}$ with present value

$$A_{x:\overline{n}|} \coloneqq \mathbb{E}\ Z = {}_n A_x + {}_n E_x = \sum_{k=0}^{n-1} v^{k+1}\, {}_k p_x \cdot q_{x+k} + {}_n p_x v^n,$$

and with monthly payment (cf. (24.1))

$$A_{x:\overline{n}|}^{(m)} = {}_n A_x^{(m)} + {}_n E_x = \frac{i}{i^{(m)}}\, {}_n A_x + {}_n E_x.$$

24.1 CONSTANTLY REPAID LOAN AND ANNUITY

Constant annuity

Recall the annuity $1 + v + v^2 + \cdots v^{K-1} = \frac{1-v^K}{1-v} = \ddot{a}_{\overline{K}|}$. Then the present value is (use ${}_k p_x \coloneqq \mathbb{P}\left[K \ge k\right]$ and $q_{x+k} \coloneqq \mathbb{P}\left[K = x+k\right]$)

$$\begin{aligned}
\ddot{a}_{x:\overline{n}|} &\coloneqq \mathbb{E}\left[\ddot{a}_{\overline{K+1 \wedge n}|}\right] = \mathbb{E}_x \sum_{k=0}^{n-1} v^k\, \mathbb{1}_{\{K \ge k\}} \\
&= \sum_{k=0}^{\infty} {}_k p_x \cdot q_{x+k} \cdot \frac{1 - v^{k+1\wedge n}}{1-v} = \sum_{k=0}^{\infty} {}_k p_x \cdot (1 - p_{x+k}) \cdot \frac{1 - v^{k+1\wedge n}}{1-v} \\
&= \sum_{k=0} {}_k p_x \cdot \frac{1 - v^{k+1\wedge n}}{1-v} - \sum_{k=0}^{n-1} {}_{k+1} p_x \cdot \frac{1 - v^{k+1\wedge n}}{1-v} \\
&= 1 + \sum_{k=1} {}_k p_x \cdot \left(\frac{1 - v^{k+1\wedge n}}{1-v} - \frac{1 - v^{k\wedge n}}{1-v}\right) \\
&= 1 + \sum_{k=1}^{n} {}_k p_x \cdot \frac{v^{k\wedge n} - v^{k+1\wedge n}}{1-v} = \sum_{k=0}^{n-1} {}_k p_x \cdot v^k
\end{aligned} \tag{24.2}$$

For the particular situation $q_k \equiv q$ thus,

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^k \cdot (1-q)^k = \frac{1 - \big((1-q)v\big)^n}{1 - (1-q)v}. \tag{24.3}$$

Note that $(1-q)v = \frac{1}{1+(i+Z)}$, where $Z = q\frac{1+i}{1-q} \approx q$ is an additional interest (spread) to account for a potential default.

Lemma 24.1 (Cf. (2.18) and Exercise 24.2)**.** *It holds that*

$$1 = A_x + d \cdot \ddot{a}_x, \quad \textit{and} \quad 1 = A_{x:\overline{n}|} + d \cdot \ddot{a}_{x:\overline{n}|}. \tag{24.4}$$

Monthly paid annuities

Lemma 24.2. *The present value is*

$$\ddot{a}^{(m)}_{x:\overline{n}|} = \alpha^{(m)} \cdot \ddot{a}_{x:\overline{n}|} - \beta^{(m)}\left(1 - {}_n E_x\right).$$

Proof. For the present value of the monthly paid annuity can be found by considering

$$1 = v^{K+S^{(m)}} + d^{(m)} \cdot \ddot{a}^{(m)}_{\overline{K+S^{(m)}}|}.$$

Take expectations and it follows that

$$1 = A^{(m)}_x + d^{(m)} \cdot \ddot{a}^{(m)}_x = \frac{i}{i^{(m)}} A_x + d^{(m)} \cdot \ddot{a}^{(m)}_x = \frac{i}{i^{(m)}}(1 - d \cdot \ddot{a}_x) + d^{(m)} \cdot \ddot{a}^{(m)}_x$$

by (24.1) and Lemma 24.1. Hence, by (2.4),

$$\ddot{a}^{(m)}_x = \frac{di}{i^{(m)} d^{(m)}} \ddot{a}_x + \frac{1}{d^{(m)}}\left(1 - \frac{i}{i^{(m)}}\right) = \alpha^{(m)} \cdot \ddot{a}_x - \beta^{(m)}.$$

The assertion follows. ◻

Proof. For another, more explicit proof expand the annuity as

$$\begin{aligned}
\ddot{a}_x^{(m)} &= \sum_{k=0}^{\infty} {}_k p_x\, v^k \left(\left(1 - q_{x+k}\frac{0}{m}\right) \frac{v^{\frac{0}{m}}}{m} + \dots + \left(1 - q_{x+k}\frac{m-1}{m}\right) \frac{v^{\frac{m-1}{m}}}{m} \right) \\
&= \sum_{k=0} {}_k p_x\, v^k \left(\ddot{a}_{\overline{1}|}^{(m)} - q_{x+k} \cdot \mathbb{E}\left(1 - S^{(m)}\right) v^{1-S^{(m)}} \right) \\
&= \sum_{k=0} {}_k p_x\, v^k \left(\frac{1-v}{d^{(m)}} - q_{x+k} \frac{d - v\, i^{(m)}}{i^{(m)} d^{(m)}} \right);
\end{aligned}$$

we have use (23.3) and (2.6) (recall that $S^{(m)} \in \left\{\frac{1}{m}, \dots 1\right\}$. It follows that

$$\begin{aligned}
\ddot{a}_x^{(m)} &= \sum_{k=0} {}_k p_x\, v^k \left(\frac{1-v}{d^{(m)}} - \frac{d - v\, i^{(m)}}{i^{(m)} d^{(m)}} + (1 - q_{x+k}) \frac{d - v\, i^{(m)}}{i^{(m)} d^{(m)}} \right) \\
&= \sum_{k=0} {}_k p_x\, v^k \left(\frac{i^{(m)} - d}{i^{(m)} d^{(m)}} + p_{x+k} \frac{d - v\, i^{(m)}}{i^{(m)} d^{(m)}} \right) \\
&= \sum_{k=0} {}_k p_x\, v^k \frac{i^{(m)} - d}{i^{(m)} d^{(m)}} + \sum_{k=0} {}_{k+1} p_x\, v^{k+1} \frac{i - i^{(m)}}{i^{(m)} d^{(m)}} \\
&= \sum_{k=0} {}_k p_x\, v^k \frac{i^{(m)} - d}{i^{(m)} d^{(m)}} + \sum_{k=0} {}_k p_x\, v^k \frac{i - i^{(m)}}{i^{(m)} d^{(m)}} - \frac{i - i^{(m)}}{i^{(m)} d^{(m)}} \\
&= \sum_{k=0} {}_k p_x\, v^k \frac{i - d}{i^{(m)} d^{(m)}} - \frac{i - i^{(m)}}{i^{(m)} d^{(m)}} = \frac{i\, d}{i^{(m)} d^{(m)}} \ddot{a}_x - \frac{i - i^{(m)}}{i^{(m)} d^{(m)}} = \alpha^{(m)} \cdot \ddot{a}_x - \beta^{(m)}
\end{aligned}$$

by employing the definitions (2.4) and $i - d = id$; consequently,

$$\ddot{a}_{x:\overline{n}|}^{(m)} = \alpha^{(m)} \cdot \ddot{a}_{x:\overline{n}|} - \beta^{(m)} \left(1 - {}_n E_x\right).$$

□

Lemma 24.3 (Cf. Lemma 24.1)**.** *By substituting (24.4) and (24.1) it follows for the monthly paid annuity that*

$$1 = A_{x:\overline{n}|}^{(m)} + d^{(m)} \cdot \ddot{a}_{x:\overline{n}|}^{(m)}.$$

Increasing annuity

The increasing annuity is defined as $\ddot{a}^{\text{inc}}_{\overline{K}|} := 1 + 2v + 3v^2 + \cdots K\, v^{K-1} = \frac{1-(K+1)v^K + K v^{K+1}}{(1-v)^2}$ (cf. (2.14)). Then

$$
\begin{aligned}
\ddot{a}^{\text{inc}}_{x:\overline{n}|} &:= \mathbb{E}\, \ddot{a}^{\text{inc}}_{\overline{K+1\wedge n}|} = \mathbb{E} \sum_{k=0}^{n-1} (k+1) v^k\, \mathbb{1}_{\{K \ge k\}} \\
&= \sum_{k=0}^{\infty} {}_k p_x q_{x+k} \ddot{a}^{\text{inc}}_{\overline{k+1\wedge n}|} = \sum_{k=0}^{\infty} {}_k p_x \left(1 - p_{x+k}\right) \ddot{a}^{\text{inc}}_{\overline{k+1\wedge n}|} \\
&= \sum_{k=0} \left({}_k p_x \ddot{a}^{\text{inc}}_{\overline{k+1\wedge n}|} - {}_{k+1} p_x \ddot{a}^{\text{inc}}_{\overline{k+1\wedge n}|} \right) \\
&= 1 + \sum_{k=1} {}_k p_x \left(\ddot{a}^{\text{inc}}_{\overline{k+1\wedge n}|} - \ddot{a}^{\text{inc}}_{\overline{k\wedge n}|} \right) \\
&= \sum_{k=0}^{n-1} {}_k p_x (k+1) v^k.
\end{aligned}
\tag{24.5}
$$

24.2 PROBLEMS

Exercise 24.1. *Verify some formulae for lump-sum premiums in Lecture 24.*

Exercise 24.2. *Verify Lemma 24.1 and Lemma 24.3 (consider perhaps* $1 - v^n = 1 - d\, \ddot{a}_{\overline{n}|}$ *and Exercise 2.4 first).*

Exercise 24.3. *Use (5.24) to verify (24.2) and (24.5).*

Regular Payments, and Constantly Repaid Loan

25

Equivalence Principle

The equivalence principle states that a contract is fair if $\mathbb{E}\, L = 0$ for all, where L incorporates all future premiums.

Pure Endowment

${}_nE_x = P \cdot \ddot{a}_{x:\overline{n}|}$, that is $\frac{{}_nE_x}{\ddot{a}_{x:n}} = P$

Life Insurance

${}_nA_x = P \cdot \ddot{a}_{x:\overline{n}|}$, that is $\frac{{}_nA_x}{\ddot{a}_{x:\overline{n}|}} = P$

Endowment

${}_nA_x + {}_nE_x = P \cdot \ddot{a}_{x:\overline{n}|}$, that is $\frac{{}_nA_x + {}_nE_x}{\ddot{a}_{x:\overline{n}|}} = P$

Endowment with premium refund

${}_nE_x + P \cdot {}_nA_x^{inc} = P \cdot \ddot{a}_{x:\overline{n}|}$, that is $\frac{{}_nE_x}{\ddot{a}_{x:\overline{n}|} - {}_nA_x^{inc}} = P$

Deferred Annuities

${}_nE_x \cdot \ddot{a}_{x+n} = P \cdot \ddot{a}_{x:\overline{n}|}$, that is $\frac{{}_nE_x \cdot \ddot{a}_{x+n}}{\ddot{a}_{x:\overline{n}|}} = P$

25.1 GROSS PREMIUMS

The premium described in Lecture 25 is the net premium. In contrast to the net premium the gross premium is loaded with costs. Of course, the costs are charged to the policyholder. Different types of costs are often associated with different sources of the costs. α-costs are related to closing the insurance contract and, this is the amount often dedicated to the insurance broker. α-costs occur once, at the beginning of the contract. The basis of remuneration for the insurance broker is the value of the insurance contract for the insurance company. This is the lump-sum premium, or the total of all future premiums, in some cases the sum insured.

β-costs are historically related to collecting the insurance premium, they are naturally based on the gross premium. β-costs incur regularly and are not considered for lump-sum contracts.

Associated with managing and governing the insurance contract are γ-costs. These costs are often associated are often based on the sum insured or the current book-value of the contract. These costs vary significantly among insurance companies, which are in competition. However, typical cost rates are $\alpha = 40\,‰$, $\beta = 3\,\%$ and $\gamma = 40\,‰$.

25.1.1 Pure Endowment

The present value of an endowment insurance contract consists of the net premium, α- and γ-costs. β-costs are not present for the lump-sum contract. As costs have to be funded by the premium payer the equations for lump-sum premium and regularly paid premiums are

$$ {}_nE_x + \alpha P + \gamma \ddot{a}_{x:\overline{n}|} = P \text{ and } {}_nE_x + \alpha n P + \gamma \ddot{a}_{x:\overline{n}|} + \beta P \ddot{a}_{x:\overline{n}|} = P \ddot{a}_{x:\overline{n}|}, \tag{25.1} $$

the premiums are

$$ P = \frac{{}_nE_x + \gamma \ddot{a}_{x:\overline{n}|}}{1-\alpha} \text{ and } P = \frac{{}_nE_x + \gamma \ddot{a}_{x:\overline{n}|}}{(1-\beta)\ddot{a}_{x:\overline{n}|} - \alpha n}. $$

This pattern is repeated for other types of insurance contracts.

25.1.2 Life Insurance

Here, the broker's commission and admin costs are based on the sum insured,

$$ P = {}_nA_x + \alpha + \gamma \ddot{a}_{x:\overline{n}|} \text{ and } P = \frac{{}_nA_x + \alpha + \gamma \ddot{a}_{x:\overline{n}|}}{(1-\beta)\ddot{a}_{x:\overline{n}|}}. $$

25.1.3 Endowment

The premiums are

$$ P = \frac{{}_nE_x + {}_nA_x + \gamma \ddot{a}_{x:\overline{n}|}}{1-\alpha} \text{ and } P = \frac{{}_nE_x + {}_nA_x + \gamma \ddot{a}_{x:\overline{n}|}}{(1-\beta)\ddot{a}_{x:\overline{n}|} - \alpha n}. $$

25.1.4 Endowment with premium refund

$$ P = \frac{{}_nE_x + \gamma \ddot{a}_{x:\overline{n}|}}{1 - {}_nA_x - \alpha} \text{ and } P = \frac{{}_nE_x + \gamma \ddot{a}_{x:\overline{n}|}}{(1-\beta)\ddot{a}_{x:\overline{n}|} - {}_nA_x^{inc} - \alpha n}. $$

25.1.5 Deferred Annuities with premium refund

The premiums

$$ P = \frac{{}_nE_x \cdot \ddot{a}_{x+n}(1 + \gamma \ddot{a}_{x:\overline{n}|} + \delta)}{1 - {}_nA_x - \alpha} \text{ and } P = \frac{{}_nE_x \cdot \ddot{a}_{x+n} + \gamma \ddot{a}_{x:\overline{n}|} \cdot {}_nE_x \ddot{a}_{x+n} + \delta_n E_x \ddot{a}_{x+n}}{(1-\beta)\ddot{a}_{x:\overline{n}|} - {}_nA_x^{inc} - \alpha n} $$

are based on the annuity, they derive from $P =_n E_x \cdot \ddot{a}_{x+n} + P_n A_x^{inc} + \beta P \ddot{a}_{x:\overline{n}|} + \alpha n P + \gamma \ddot{a}_{x:\overline{n}|} \cdot {}_nE_x \ddot{a}_{x+n} + \delta_n E_x \ddot{a}_{x+n}$. Here γ are the admin costs as long as no annuity is paid, the costs δ incurring afterward are based on the annuity.

25.1.6 Reminder of Debt of a Constantly Repaid Loan

Recall that a total loan of $L = \ddot{a}_{x:\overline{n}|}$ may be granted to a client at an installment of 1 (say). After some time k, the outstanding amount is

$$ V_k = \ddot{a}_{x+k:\overline{n-k}|}, $$

reflecting the present value of all future payments (prospective calculation). This should be in line with the retrospective computation, that is to say

$$ V_k = \frac{L \cdot r^k - \sum_{i=0}^{k-1} {}_kp_x r^{k-i}}{{}_kp_x} = \ddot{a}_{x+k:\overline{n-k}|}. $$

Indeed, this holds true (multiply by ${}_k p_x v^k$) as

$$L - \sum_{i=0}^{k-1} {}_k p_x v^i = v^k \, {}_k p_x \, \ddot{a}_{x+k:\overline{n-k}|}$$

and $L = \ddot{a}_{x:\overline{n}|}$, as above.

Notice that one may decompose

$$\begin{aligned}
\ddot{a}_{x+k:\overline{n-k}|} &= 1 + v \cdot p_{x+k} \, \ddot{a}_{x+1+k:\overline{n-(k+1)}|} \\
&= 1 + v \left(1 - q_{x+k}\right) \ddot{a}_{x+1+k:\overline{n-(k+1)}|} \\
&= 1 + v \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|} - q_{x+k} \cdot v \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|},
\end{aligned}$$

and the installment thus may be decomposed as

$$\begin{aligned}
-1 &= \underbrace{v \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|} - \ddot{a}_{x+k:\overline{n-k}|}}_{P^{saving}} + \underbrace{-q_{x+k} \cdot v \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|}}_{P^{risk}} \\
&= \underbrace{v \cdot V_{k+1} - V_k}_{P^{saving}} + \underbrace{-q_{x+k} \cdot v \cdot V_{k+1}}_{P^{risk}} . \qquad (25.2)
\end{aligned}$$

Notice the particular interpretation $P^{risk} = q_{x+k} \cdot v \cdot V_{k+1}$ which states that the outstanding amount V_{k+1} has to be covered in case of default, which happens with probability q_{x+k} in year k. Note, that no recovery applies here.

Incorporating Recovery Now

Suppose there is a recovery C_{x+k} in case of default in year k. The present value for the loan then obviously is

$$L = \mathbb{E}\left[\ddot{a}_{\overline{(K+1)\wedge n}|} + v^{K+1} C_K\right],$$

and this is the amount which can be granted. We have investigated the payment stream $\ddot{a}_{x:\overline{n}|} = \mathbb{E}\left[\ddot{a}_{\overline{(K+1)\wedge n}|}\right]$ above, so let's focus on $A_x = \mathbb{E}\left[v^{K+1} C_K\right]$. The recursion obviously is (assuming the $C_K = 0$ for $K \geq n$, that is after termination of the loan),

$$\begin{aligned}
A_x &= \mathbb{E}\, v^{K+1} C_{K+1} \\
&= \sum_{k=0}^{n-1} {}_k p_x \cdot q_{x+k} \cdot v^{k+1} \cdot C_{k+1} \\
&= v \cdot q_x C_1 + \sum_{k=0}^{n-2} {}_{k+1} p_x \cdot q_{x+1+k} \cdot v^{k+2} \cdot C_{k+2} \\
&= v \cdot q_x C_1 + {}_1 p_x \cdot v \cdot \sum_{k=0}^{n-2} {}_k p_{x+1} \cdot q_{x+1+k} \cdot v^{k+1} \cdot C_{k+2} \\
&= v \cdot q_x C_1 + {}_1 p_x \cdot v \cdot A_{x+1}, \qquad (25.3)
\end{aligned}$$

or again and more generally

$$0 = v \cdot V_{k+1} - V_k + v \cdot q \left(C_{k+1} - V_{k+1}\right)$$

for $V_k = A_{x+k}$.

Summing up with (25.2) thus

$$\begin{aligned}
-1 =& v \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|} - \ddot{a}_{x+k:\overline{n-k}|} - v \cdot q_{x+k} \cdot \ddot{a}_{x+1+k:\overline{n-(k+1)}|} \\
&+ v \cdot A_{x+k+1} - A_{x+k} + v \cdot q\left(C_{k+1} - A_{x+k}\right) \\
=& v \cdot \left(\ddot{a}_{x+1+k:\overline{n-(k+1)}|} + A_{x+k+1}\right) - \left(\ddot{a}_{x+k:\overline{n-k}|} + A_{x+k}\right) \\
&+ v \cdot q_{x+k}\left(C_{k+1} - \left(A_{x+k} + \ddot{a}_{x+1+k:\overline{n-(k+1)}|}\right)\right) \\
=& \underbrace{v \cdot V_{k+1} - V_k}_{-P^{saving}} + \underbrace{v \cdot q_{x+k} \cdot \left(C_{k+1} - V_{k+1}\right)}_{P^{risk}}.
\end{aligned} \tag{25.4}$$

and (recall that $d = 1 - v$, Exercise 2.10)

$$\begin{aligned}
-1 &= v \cdot V_{k+1} - V_k + q_{x+k} \cdot v \cdot \left(C_{k+1} - V_{k+1}\right) \\
&= \underbrace{V_{k+1} - V_k}_{savings} - \underbrace{d \cdot V_{k+1}}_{interest} + \underbrace{q_{x+k} \cdot v \cdot \left(C_{k+1} - V_{k+1}\right)}_{risk}
\end{aligned}$$

or alternatively

$$V_{k+1} = V_k - \underbrace{1}_{repayment} + \underbrace{d \cdot V_{k+1}}_{interest} - \underbrace{q_{x+k} \cdot v \cdot \left(C_{k+1} - V_{k+1}\right)}_{risk}$$

25.1.7 The general pattern

To generalize the findings and patterns of the previous displays we define

$$P^{savings} := vV_{k+1} - V_k \text{ and } P^{risk} := vq_{x+k}\left(C_{k+1} - V_{k+1}\right),$$

then every premium P_k (recall that a negative repayment is a premium) rewrites by (25.4) as

$$P_k = P_k^{savings} + P_k^{risk}.$$

We obtain that

$$P_k = \underbrace{v\, V_{k+1} - V_k}_{P^{savings}} + \underbrace{v\, q_{x+k}\left(C_{k+1} - V_{k+1}\right)}_{P^{risk}},$$

or

$$V_{k+1} = V_k + \underbrace{P_k}_{payment} + \underbrace{d \cdot V_{k+1}}_{interest} - \underbrace{q_{x+k} \cdot v \cdot \left(C_{k+1} - V_{k+1}\right)}_{risk},$$

which assembles the next wealth V_{k+1} as previous wealth V_k, plus payment P_k, plus earned interest dV_{k+1} (note that this is not iV_k) minus the payment for risk, $q_{x+k}v(C_{k+1} - V_{k+1})$. Note further that

$$P_k + V_k = v\left\{\left(1 - q_{x+k}\right)V_{k+1} + q_{x+k}\, C_{k+1}\right\},$$

the Markov property (cf. 25.3).

Constant Risk—a special case

For constant risk, again $q_{x+k} \equiv q$ and fixed recovery $C = 1$, and recalling (24.3)

$$\begin{aligned} A_x &= \mathbb{E}\, v^{K+1} C_K \\ &= \sum_{k=0}^{n-1} {}_k p_x \cdot q_{x+k} \cdot v^{k+1} \\ &= \sum_{k=0}^{n-1} (1-q)^k \cdot q \cdot v^{k+1} = q \cdot v \cdot \sum_{k=0}^{n-1} \big((1-q)v\big)^k \\ &= q \cdot v \cdot \ddot{a}_{x:\overline{n}|}, \end{aligned}$$

that is $\frac{A_x}{\ddot{a}_{x:\overline{n}|}} = q \cdot v$, i.e., the premium does not change.

25.2 THE MODEL

The face amount of a loan with installment $\frac{1}{m}$ we have found (cf. (2.9)) to be $a_{\overline{n}|}^{(m)}$. We will extend the notion of present values and incorporate uncertainty, which in turn is described by random variables. The generalized present value is

$$V := \mathbb{E} \sum_t \frac{C_t}{(1+i)^t},$$

which obeys the same features and properties, as recursivity and duration.

To incorporate for example credit risk we may proceed as follows: define the simple binomial random variables

L_t	R_{t+1} "default"	0 "non-default"
	q_t	$1 - q_t =: p_t$

describing the recovery in case of a credit event (credit default): $\mathbb{P}[L_t = 0] = 1 - q_t$ and $\mathbb{P}[L_t = R_{t+1}] = q_t$: the probability for a default during year t is q_t, and in this case the recovery is R_{t+1}.

In case of a defaulted loan, the actual outstanding amount is lost for the bank despite a potential recovery. That is to say, it may not be re-paid at time t and the bank has to write-off the entire outstanding amount.

The present value thus is

$$\begin{aligned} V &= \mathbb{E} \sum_t \frac{C_t}{(1+i)^t} \\ &= \sum_{k=0}^{T} \frac{C_{t_k}}{(1+i)^k}\, {}_k p_{t_0} \cdot q_{t_0+k}, \end{aligned}$$

where we have put ${}_k p_t := (1-q_t)(1-q_{t+1}) \dots (1-q_{t+k-1})$ to account for the *non-default* during all the years from t to $t+k$

We may augment the model by additional parameters and incorporate for example parameters fees and a potential recovery in case of a credit event. Incorporated all this in one formula, the respective prospective recursion rewrites

$$V_t + P_t = F_t + (1-q_t) \cdot v_t V_{t+1} + q_t \cdot v_t R_{t+1}; \tag{25.5}$$

here,

- P_t is the payment, installment or premium,
- F_t is the fee due in year t,
- R_{t+1} is the recovery in case of default (notice the index $t+1$: recovery is considered towards the end of the year; changing the sign of R_t would describe additional risk, which is loaded in case of a credit event), and
- q_t, as above, is the probability for a credit default (credit event) within year t ($q_t := \mathbb{P}$ [default within year t]).

Together with the additional requirement

$$V_T = 0$$

it is possible, using (25.5), to compute all present values of the loan given the fixed payments P_t, and the additional condition

$$V_0 = 0$$

may be used to derive the face amount, which can be borrowed to the client.

The recursion, in addition, has these following interpretations:

- The left hand side, $V_t + P_t$, is the cash available. This amount will be used to finance the right hand side.
- The right hand side $F_t + v_t\,(1 - q_t)\,V_{t+1} + v_t q_t R_t$, consisting of
  - fees,
  - the future present value in case of non-default, and
  - the recovery

  has to be financed.
- both sides, left an right, have to coincide.

Equation (25.5) may be written as

$$\begin{aligned} P_t &= F_t + \underbrace{V_{t+1} - V_t}_{\Delta V} - \underbrace{d_t V_{t+1}}_{\text{interest}} + \underbrace{v_t q_t\,(R_{t+1} - V_{t+1})}_{P_i^R}. \\ &= \underbrace{F_t}_{P_t^C} + \underbrace{v_t V_{t+1} - V_t}_{P_t^S} + \underbrace{v_t q_t\,(R_{t+1} - V_{t+1})}_{P_t^R} \end{aligned}$$

The premium consists of 4 ingredients:

- the component describing the fees, $P_t^F := F_t$,
- the savings premium $P_t^S := v_t V_{t+1} - PV_t$, which consists of two parts,
  - the change of reserves, $\Delta V = V_{t+1} - V_t$ and
  - the respective interest voucher, $d_t V_{t+1}$,
- the premium responsible to quantify the contribution for risk, $P_t^R = q_t \cdot v_t\,(V_{t+1} - R_{t+1})$.

25.3 MARGINS

Consider a usual (personal) loan with some (variable or constant) interest rate i, there will be no recovery $R_{t+1} = 0$ (as typical for a personal loan), a constant default rate $q_t = q$ and neglect the fees, as they have minor impact anyhow. The recursion thus simplifies to

$$V_t = 1 + v\,(1-q)\,V_{t+1}.$$

Introduce the derived interest rate $i_q \coloneqq \frac{i+q}{1-q}$ and observe that $\frac{1}{1+i_q} = \frac{1-q}{1+i}$. So the recursion, again, is

$$V_t = 1 + \frac{1}{1+i_q} V_{t+1},$$

which strikingly reminds to (2.15): i_q, however, is the interest rate with some margin added, and $i_q \approx i + q$ – this observation justifies the fact, that financial institutions base all their expectations on risk into a single number – the margin – which then is added to the interest rate.

Given constant interest rates the solution of this recursion is simply $\ddot{a}_n$, but computed with the augmented interest rate i_q instead of i.

25.4 HATTENDORFF'S THEOREM

The present value is the expected value, which we have computed on a prospective basis. Having a look at the evolution of a specific loan in time, we will observe another pattern: All future present values are then PV_t, provided the loan did not default yet. When defaulted, PV falls back to 0 and stays there. So any individual loan is a process, and any individual loan has incorporated risk.

The random variable L_t introduced above to describe the loss has expected value $q_t c_t$, the variance is $q_t\,(1-q_t)\,c_t^2$.

The easiest way to measure this risk incorporated is the variance. Hattendorff's theorem[1] gives the precise formula, which is

$$\operatorname{var} L = \sum_{k=0} v^{2k+2}\left(R_{k+1} - V_{k+1}\right)^2 \;{}_{k+1}p_t\, q_{t+k}$$

[1] K. Hattendorff, 1968

26 Karhunen–Loève and Donsker's Theorem

26.1 Karhunen–Loève Theorem

Consider a stochastic process $(X_t)_{t\in[0,1]}$ during the times $[0,1]$ with

(i) Every trajectory $t \mapsto X_t(\cdot) \in L^2$,

(ii) $\mathbb{E}\, X_t = 0$.

Define the covariance function $K(s,t) \coloneqq \operatorname{cov}(X_s, X_t)$ (note that K is symmetric, $K(s,t) = K(t,s)$) and consider the operator

$$\begin{aligned} \mathbf{K}\colon L^2[0,1] &\to L^2[0,1] \\ f &\mapsto \mathbf{K}f, \text{ where } (\mathbf{K}f)(s) \coloneqq \int_0^1 K(s,t) f(t) dt. \end{aligned} \tag{26.1}$$

We denote its eigenvalues by λ_k, the eigenvectors (i.e., eigenfunctions) of $\mathbf{K}$ by $e_k(\cdot)$, and assume (without loss of generality) that $e_k(\cdot)$ have norm 1, i.e., $\int_0^1 e_k(t)e_j(t)dt = \delta_{k,j}$. We thus have that

$$\int_0^1 K(s,t)e_j(t)dt = (\mathbf{K}e_j)(s) = \lambda_j e_j(s), \qquad s \in [0,1]. \tag{26.2}$$

Note in particular that

$$\int_0^1 \left(\sum_{k=1} \lambda_k e_k(s) e_k(t) \right) e_j(t) dt = \sum_{k=1} \lambda_k e_k(s) \cdot \underbrace{\int_0^1 e_k(t) e_j(t) dt}_{\delta_{k,j}} = \lambda_j e_j(s), \tag{26.3}$$

so that

$$K(s,t) = \sum_{k=1} \lambda_k e_k(s) e_k(t) \qquad \left(\text{i.e., } K = \sum_{k=1} \lambda_k e_k \otimes e_k^* = \sum_{k=1} \lambda_k \left|e_k\right\rangle \left\langle e_k\right|\right) \tag{26.4}$$

by comparing (26.2) and (26.3). By assuming that $K(s,t) \ge 0$ it follows further from Mercer's theorem that $\lambda_k \ge 0$. Note further that we may assume that the orthonormal system is complete, as otherwise we may augment the system with additional functions $e_k(\cdot)$ until they form a complete orthonormal system of L^2.

Theorem 26.1 (Karhunen[1]–Loève[2]). *The stochastic process has the expansion*

$$X_t(\omega) = \sum_{k=1}^{\infty} Z_k(\omega) e_k(t) \tag{26.5}$$

[1] Kari Karhunen, 1915–1992

[2] Michel Loève, 1907–1979

with convergence in L^2 *for every t, where*

$$Z_k = \int_0^1 X_t e_k(t)dt \qquad (i.e., Z_k(\omega) = \int_0^1 X_t(\omega)e_k(t)dt). \tag{26.6}$$

Further,

$$\mathbb{E}\, Z_k = 0 \text{ and } \mathbb{E}\, Z_j Z_k = \delta_{jk} \cdot \lambda_j. \tag{26.7}$$

Proof. As $t \to X_t(\omega) \in L^2[0,1]$ the expansion (26.5) is clear, as $e_k(\cdot)$ constitute a complete orthonormal system.

We then have

$$\mathbb{E}\, Z_k = \mathbb{E} \int_0^1 X_t e_k(t)dt = \int_0^1 (\mathbb{E}\, X_t)e_k(t)dt = 0$$

and

$$\begin{aligned}\mathbb{E}\, Z_k \cdot Z_j &= \mathbb{E} \int_0^1 X_s e_k(s)ds \cdot \int_0^1 X_t e_j(t)dt = \int_0^1 \int_0^1 (\mathbb{E}\, X_s \cdot X_t)e_k(s)ds \cdot e_j(t)dt \\ &= \int_0^1 \underbrace{\int_0^1 K(s,t)e_k(s)ds}_{\lambda_k e_k(t)} \cdot e_j(t)dt = \int_0^1 \lambda_k e_k(t)e_j(t)dt = \lambda_k \delta_{jk}.\end{aligned}$$

We finally show convergence in L^2. To this end define $S_N(t) \coloneqq \sum_{k=1}^N Z_k e_k(t)$. Then

$$\begin{aligned}\mathbb{E}\left(X_t - S_N(t)\right)^2 &= \mathbb{E}\, X_t^2 + \mathbb{E}\, S_N(t)^2 - 2\,\mathbb{E}\, X_t S_N(t) \\ &= K(t,t) + \mathbb{E} \sum_{k=1}^N \sum_{j=1}^N Z_k Z_j e_i(t)e_k(t) - 2\,\mathbb{E}\, X_t \sum_{k=1}^N Z_k e_k(t) \\ &= K(t,t) + \sum_{k=1}^N \lambda_k e_k(t)^2 - 2\,\mathbb{E}\, X_t \sum_{k=1}^N e_k(t) \int_0^1 X_s e_k(s)ds \\ &= K(t,t) + \sum_{k=1}^N \lambda_k e_k(t)^2 - 2 \sum_{k=1}^N \int_0^1 K(s,t)e_k(t)e_k(s)ds \\ &= K(t,t) + \sum_{k=1}^N \lambda_k e_k(t)^2 - 2 \sum_{k=1}^N \underbrace{\int_0^1 K(s,t)e_k(s)ds\; e_k(t)}_{\lambda_k e_k(t)} \\ &= K(t,t) - \sum_{k=1}^N \lambda_k e_k(t)^2 \xrightarrow[N\to\infty]{} 0\end{aligned}$$

for every $t \in [0,1]$ by (26.4) (or Mercer's theorem) again. □

Corollary 26.2. *The process* $\tilde{X}_t \coloneqq \sum_{k=1}^\infty Z_k e_k(t)$ *has covariance* $\operatorname{cov}\left(\tilde{X}_s, \tilde{X}_t\right) = \operatorname{cov}(X_s, X_t)$ *and thus is a copy of the stochastic process* $(X_t)_{t\in[0,1]}$.

Proof. It follows from (26.7) that $\mathbb{E}\, \tilde{X}_t = 0$. Hence

$$\begin{aligned}\operatorname{cov}\left(\tilde{X}_s, \tilde{X}_t\right) &= \mathbb{E}\, \tilde{X}_s \cdot \tilde{X}_t = \mathbb{E} \sum_{k=1}^\infty Z_k e_k(s) \cdot \sum_{j=1}^\infty Z_j e_j(t) = \sum_{j,k=1}^\infty e_k(s)e_j(t)\,\mathbb{E}\, Z_k Z_j \\ &= \sum_{k=1}^\infty e_k(s)e_k(t)\lambda_k = K(s,t)\end{aligned}$$

by (26.4) and thus the result. □

26.2 KARHUNEN–LOÈVE REPRESENTATION OF THE WIENER PROCESS

The covariance function of the Brownian motion is $K(s,t) = \operatorname{cov}(W_s, W_t) = \min(s,t)$, cf. (10.2). To deduce the eigenvalues of the associated operator $\mathbf{K}$ observe that every eigenfunction (eigenvector) e satisfies $\mathbf{K}e = \lambda \cdot e$, i.e.,

$$\int_0^t s\, e(s)ds + \int_t^1 t\, e(s)ds = \int_0^1 \min(s,t)e(s)ds = \int_0^1 K(s,t)e(s)ds = \lambda \cdot e(t). \tag{26.8}$$

By differentiating with respect to t thus

$$t\, e(t) - t\, e(t) + \int_t^1 e(s)ds = \lambda \cdot e'(t), \tag{26.9}$$

and again

$$-e(t) = \lambda \cdot e''(t). \tag{26.10}$$

The general solution is $e(t) = A \sin \frac{t}{\sqrt{\lambda}} + B \cos \frac{t}{\sqrt{\lambda}}$.

Choose $t = 0$ in (26.8) to see that $e(0) = 0$, thus $B = 0$. Further, choose $t = 1$ in (26.9) to see that $e'(1) = 0$, i.e., $0 = e'(1) = \frac{A}{\sqrt{\lambda}} \cos \frac{1}{\sqrt{\lambda}}$ and thus

$$\lambda_k = \frac{1}{\left(\left(k - \frac{1}{2}\right)\pi\right)^2}.$$

The orthonormal eigenvecturs thus are $e_k(t) = \sqrt{2} \sin\left(k - \frac{1}{2}\right)\pi t$. Note in particular that

$$K(s,t) = \min(s,t) = \sum_{k=1}^{\infty} \frac{2 \sin\left(k - \frac{1}{2}\right)\pi s \cdot \sin\left(k - \frac{1}{2}\right)\pi t}{\left(\left(k - \frac{1}{2}\right)\pi\right)^2},$$

by (26.4).

Remark 26.3. Suppose that X_t is Gaussian (normal) for each $t \in [0,1]$. The sum of Gaussians is Gaussian as well. Thus, by its definition in (26.6), Z_k is Gaussian. With (26.7) it follows that $Z_k \sim \mathcal{N}(0, \lambda_k)$.

Finally let ξ_k be independent standard normals, then $Z_k \coloneqq \frac{1}{\left(k-\frac{1}{2}\right)\pi}\xi_k$ satisfy the conditions (26.7). We infer from (26.5) the representation

$$W_t = \sum_{k=1} \xi_k \frac{\sqrt{2}\sin\left(k - \frac{1}{2}\right)\pi t}{\left(k - \frac{1}{2}\right)\pi}, \qquad t \in [0,1],$$

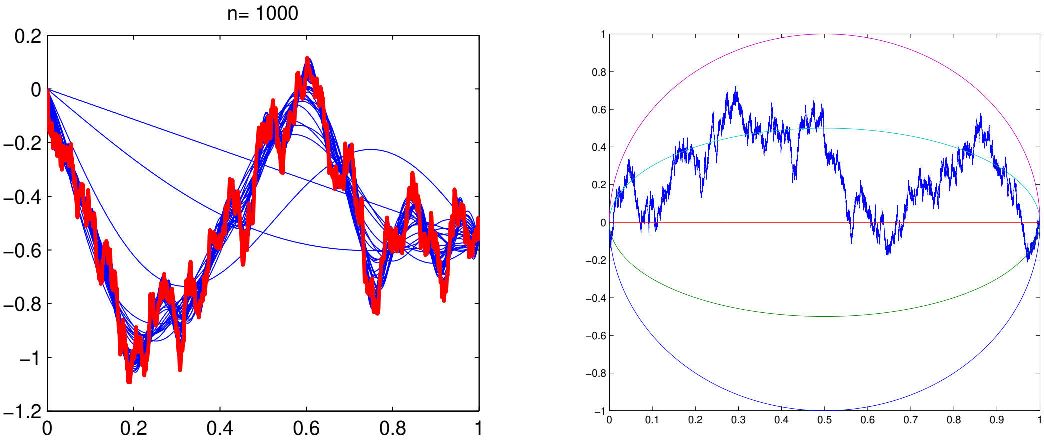
for the Wiener process.

Remark 26.4. By the self similarity (Lemma 10.3) we get a Brownian motion on $[0,c]$ by the transformation $\sqrt{c}W_{t/c}$ for arbitrary $c > 0$.

26.3 KARHUNEN–LOÈVE REPRESENTATION OF THE BROWNIAN BRIDGE

The covariance function of the Brownian bridge

$$B_t \coloneqq W_t - t \cdot W_1, \qquad t \in [0,1] \tag{26.11}$$



(a) The first 1000 partial sums in the Karhunen–Loève representation (26.14)

(b) The scaled empirical process (27.2) converging towards a Brownian bridge for uniform random variables

Figure 26.1: Further constructions of Brownian motion

i.e., a Wiener process conditioned on $W_1 = 0$, is

$$\begin{aligned}\operatorname{cov}(B_s, B_t) &= \mathbb{E}\,(W_s - sW_1)\,(W_t - tW_1) = \mathbb{E}\,W_sW_t - sW_1W_t - tW_sW_1 + stW_1^2\\ &= \min(s,t) - s\min(1,t) - t\min(s,1) + st\\ &= \min(s,t) - st - st + st = \min(s,t) - st. \end{aligned} \tag{26.12}$$

As for the Wiener process the eigenfunctions satisfy the equation (26.10), but for the eigenvalue $\lambda_k = \frac{1}{k^2\pi^2}$ and the eigenfuntion $e_k(t) = \sqrt{2}\sin k\pi t$. We have that

$$\min(s,t) - st = \sum_{k=1} \frac{2\sin k\pi s \cdot \sin k\pi t}{k^2\pi^2}.$$

The Brownian bridge thus has the representation

$$B_t = \sum_{k=1} \xi_k \sqrt{2}\frac{\sin k\pi t}{k\pi}, \qquad t \in [0,1], \tag{26.13}$$

where ξ_k are independent normals.

Remark 26.5. By combining (26.13) and (26.11) we also find that

$$W_t = B_t + tW_1 \sim t\xi_0 + \sum_{k=1} \xi_k \cdot \sqrt{2}\frac{\sin k\pi t}{k\pi}, \qquad t \in [0,1]. \tag{26.14}$$

Figure 26.1a displays a trajectory drawn from (26.14).

Donsker's theorem

27

Consider iid random variables X_i and the *empirical distribution function*

$$F_n(x) \coloneqq \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}(X_i) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i\le x\}}.$$

Apparently,

$$\mathbb{E}\, F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\,\mathbb{1}_{(-\infty,x]}(X_i) = \frac{1}{n}\sum_{i=1}^n P(X_i \le x) = F(x)$$

and

$$\begin{aligned}
\mathbb{E}\, F_n(x)F_n(y) &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\,\mathbb{1}_{(-\infty,x]}(X_i)\,\mathbb{1}_{(-\infty,y]}(X_j)\\
&= \frac{1}{n^2}\sum_{i\ne,j} \mathbb{E}\,\mathbb{1}_{(-\infty,x]}(X_i)\,\mathbb{1}_{(-\infty,y]}(X_j) + \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\,\mathbb{1}_{(-\infty,x]}(X_i)\,\mathbb{1}_{(-\infty,y]}(X_i)\\
&= \frac{1}{n^2}\sum_{i\ne,j} P(X_i\le x)P(X_j\le y) + \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\,\mathbb{1}_{(-\infty,x\wedge y]}(X_i)\\
&= \frac{n(n-1)}{n^2}F(x)F(y) + \frac{1}{n}F(x\wedge y) = \frac{1}{n}\big(F(x)\wedge F(y) - F(x)F(y)\big) + F(x)F(y)
\end{aligned}$$

by independence, so that

$$\operatorname{cov}\big(F_n(x), F_n(y)\big) = \frac{1}{n}\big(F(x\wedge y) - F(x)F(y)\big). \tag{27.1}$$

It follows that the scaled process converges (choose $x = y$ in (27.1)),

$$\sqrt{n}\big(F_n(x) - F(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, F(x)\big(1-F(x)\big)\Big),$$

and comparing (27.1) with (26.12) reveals that the limit

$$\sqrt{n}\big(F_n(x) - F(x)\big) \xrightarrow{\mathcal{D}} B_{F(x)} \tag{27.2}$$

converges towards a transformed (time changed) Brownian bridge, if F is continuous.

Figure 26.1b provides a sample of the process for the uniform distribution (i.e., $F(u) = u$).

Fractional Brownian motion 28

28.1 PROPERTIES

Definition 28.1. A process $(X_t)_{t\in\mathbf{T}}$ is Gaussian, if $\left(X_{t_1},\dots X_{t_n}\right)$ is a multivariate Gaussian random variable for every $n\in\mathbb{N}$ and every selection $(t_1,\dots t_n)\in\mathbf{T}^n$ of indices.

Definition 28.2. A Gaussian process $\left(B_t^H\right)$ is a *fractional Brownian motion* (fBm) if $\mathbb{E}\,B_t^H=0$ and

$$\mathbb{E}\,B_s^H B_t^H=\frac{1}{2}\left(t^{2H}+s^{2H}-|t-s|^{2H}\right).$$

$H\in(0,1)$ is called the *Hurst*[1] *index* (sometimes also Hurst exponent, or Hurst parameter associated with the fractional Brownian motion).

Remark 28.3. The process $B_t^{1/2}$ is a Brownian motion (Wiener process), cf. the covariance relation (10.2).

Proposition 28.4 (Properties of the fractional Brownian motion)**.** *It holds that*

(i) $B_t^H\sim\mathcal{N}(0,t)$,

(ii) $\mathbb{E}\left(B_s^H-B_t^H\right)^2=|t-s|^{2H}$ *(and in particular the increments are stationary),*

(iii) $\operatorname{cov}\left(B_{s_2}^H-B_{s_1}^H,B_{t_2}^H-B_{t_1}^H\right)<0$ *for* $H<1/2$ *and* $s_1<s_2<t_1<t_2$ *(i.e., the increments are negatively correlated for* $H<1/2$*), and*

(iv) $\operatorname{cov}\left(B_{s_2}^H-B_{s_1}^H,B_{t_2}^H-B_{t_1}^H\right)>0$ *for* $H>1/2$ *(positively correlated increments).*

(v) $B_t^H-B_s^H\sim B_H(t-s)$,

(vi) $B_{ct}^H\sim c^H B_t^H$ *for* $c>0$ *(self-similarity of the fBm).*

Differences between the Brownian motion and fBm:

(i) In contrast to the Brownian motion, the process B_t^H is not a semi-martingale ($H\neq 1/2$).

(ii) While the increments in Brownian Motion are independent, the opposite is true for fractional Brownian motion. This dependence means that if there is an increasing pattern in the previous steps, then it is likely that the current step will be increasing as well (if $H\geq 1/2$).

(iii) The fBm are not Markovian, and this becomes a strong difficulty to study and to put these models into practice (the usual techniques assume the Markov property).

[1] Harold Edwin Hurst, 1880–1978, British hydrologist

28.2 SIMULATION

A fBm-path can be simulated at discrete points $(t_1, \dots t_n) \in \mathbf{T}^n$ by finding a matrix L (for example by employing a Cholesky-decomposition) so that

$$L \cdot L^\top = \left(\frac{1}{2}\left(t_i^{2H} + t_j^{2H} - \left|t_i - t_j\right|^{2H}\right)\right)_{i,j=1}^{n}.$$

Then $u \coloneqq L \cdot (v_1, \dots v_n)^\top$ for $v_1, \dots v_n$ independent standard Gaussians is a sample path of an fBm.

28.3 ESTIMATION OF THE HURST PARAMETER

Note that from Proposition 28.4 (ii) we have that

$$\frac{1}{N^{1-2H}} \sum_{i=1}^{N} \left(B^H_{\frac{i+1}{N}} - B^H_{\frac{i}{N}}\right)^2 \to 1 \qquad a.s.$$

Denote $V_N \coloneqq \sum_{i=1}^N \left(B^H_{\frac{i+1}{N}} - B^H_{\frac{i}{N}}\right)^2$, then

$$\hat{H}_n \coloneqq \frac{1}{2}\left(1 - \frac{\ln V_N}{\ln N}\right)$$

is an estimator for H and it holds that

$$H_N \to H \qquad a.s.$$

28.4 PROBLEMS

Exercise 28.1. *Verify the claims in Remark 28.3 and Proposition 28.4.*

Stable Distributions 29

As for references see

- ▹ Nolan [15], http://fs2.american.edu/jpnolan/www/stable/chap1.pdf and
- ▹ Janson [8], http://www2.math.uu.se/~svante/papers/sjN12.pdf
- ▹ http://www.math.nus.edu.sg/~matsr/ProbI/Lecture11.pdf.

29.1 PROPERTIES

Definition 29.1. X follows a *stable distribution* (or *Lévy-alpha stable distribution*) if for iid copies X_i, $i = 1, \dots n$ of X there are numbers $C_n > 0$ and D_n so that $X_1 + X_2 + \dots + X_n \sim C_n \cdot X + D_n$.

Stable distributions do not have a closed form density. They are usually given by their characteristic function (recall that $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ for independent random variables).

Definition 29.2. A random variable X is stable, $X \sim S(\alpha, \beta, \gamma, \delta; 1)$

$$\mathbb{E}\, e^{itX} = \begin{cases} \exp\left(-|\gamma t|^\alpha \left(1 - i\beta\,(\operatorname{sign} t)\tan\frac{\pi\alpha}{2}\right) + i\delta t\right) & \text{if } \alpha \neq 1, \\ \exp\left(-\gamma\,|t|\left(1 + i\beta\frac{2}{\pi}\,(\operatorname{sign} t)\ln t\right) + i\delta t\right) & \text{if } \alpha = 1. \end{cases} \tag{29.1}$$

The parameters given here for $X \sim S(\alpha, \beta, \gamma, \delta; 1)$ are the standard (*type 1*) parametrization,[1] its characteristic function is of simple form and has nice algebraic properties.

Type 0 parametrization (which differs in its location parameter δ) is favored for numerical work and statistical inference, as the characteristic function is continuous for all parameters (note that (29.1) is not continuous at $\alpha = 1$).

The parameters are

(i) $\alpha \in (0, 2)$ the index of the stability or the shape parameter,

(ii) $\beta \in [-1, 1]$ skewness parameter,

while

(i) $\gamma \in (0, \infty)$ is a scale parameter and

(ii) $\delta \in (-\infty, \infty)$ the location parameter.

Example 29.3. The following are stable distributions:

- ▹ the Gaussian distribution ($\alpha = 2$)
- ▹ the Cauchy distribution ($\alpha = 1$)

[1] Although useful, the different parametrizations have cause "a comedy of errors". The parametrization , in use are type 1 and type 0.

▹ Lévy distribution ($\alpha = 1/2$, $\beta = \pm 1$, page 57)

Remark 29.4. The α-stable distribution has the following properties:

(i) $P(|X| > x) \sim x^{-\alpha}$ for $x \to \infty$,

(ii) $\mathbb{E}\,|X|^p < \infty$ if $0 < p < \alpha$, but $\mathbb{E}\,|X|^p = \infty$ for $p \geq \alpha$,

(iii) $\mathbb{E}\,X = \mu$ for $\alpha > 1$, but $\mathbb{E}\,X = \infty$ for $\alpha \leq 1$,

(iv) $Y \coloneqq \sum_{i=1}^n X_i$ is α-stable with parameters $\mu = \sum_{i=1}^n \mu_i$, $\sigma^\alpha = \sum_{i=1}^n \alpha_i^\alpha$, $\beta = \frac{\sum_{i=1}^n \beta_i \sigma_i^\alpha}{\sum_{i=1}^n \sigma_i^\alpha}$.

Cont and Tankov [2]

30 Merton fraction

We consider a riskless bond with $dB_t = rB_t dt$ and a risky asset $dS_t = \mu S_t dt + \sigma S_t dW_t$, the discount rate ρ. Further, for the investment with fraction π_t, the wealth is $w_t \coloneqq (1-\pi_t)B_t + \pi_t S_t$ so that

$$dw_t = \left(\left(\pi_t\mu + (1-\pi_t)r\right) w_t - c_t\right) dt + \pi_t \sigma w_t \, dW_t.$$

Employing the utility $u(\cdot)$ and the consumption c_t, the investor maximizes

$$\mathbb{E}\left[\int_t^T e^{-\rho(s-t)} u(c_s) ds + e^{-\rho(T-t)} p(T) u(w_T) \middle| w_t\right],$$

where $p(\cdot)$ is the terminal, bequest function. For the analysis define the optimal discounted value function

$$v(t,w) \coloneqq \max_{\pi_t, c_t} \mathbb{E}\left[\int_t^T e^{-\rho s} u(c_s) ds + e^{-\rho T} p(T) u(w_T) \middle| w_t = w\right].$$

Theorem 30.1 (Merton's fraction,[1] cf. Karatzas and Shreve [10, 5.8 C])**.** *For the particular utility function* $u(c) \coloneqq \frac{c^{1-\gamma}}{1-\gamma}$ *the optimal policy is* $c_t^* = \frac{\nu\, w_t}{1+(\nu\epsilon-1)e^{-\nu(T-t)}}$ *and*

$$\pi_t^* = \frac{\mu - r}{\sigma^2 \cdot \gamma}$$

(the constant (sic!) Merton fraction), where $\nu \coloneqq \frac{\rho - (1-\gamma)\left(\frac{(\mu-r)^2}{2\sigma^2\gamma} + r\right)}{\gamma}$. *The optimal discounted value function has the explicit form* $v(t,w) = e^{-\rho t} \frac{\left(1+(\nu\epsilon-1)e^{-\nu(T-t)}\right)^\gamma}{\nu^\gamma} \frac{w^{1-\gamma}}{1-\gamma}$.

Remark 30.2. The risk aversion coefficient of the utility function $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ is $-\frac{c u''(c)}{u'(c)} = \gamma$.

Proof. For $t_1 \in (t,T)$ it is apparent that

$$v(t,w) \coloneqq \max_{\pi_t, c_t} \mathbb{E}\left[v(t_1, w_{t_1}) + \int_t^{t_1} e^{-\rho s} u(c_s) ds \middle| w_t = w\right].$$

In stochastic differential form ($t_1 \to t$) this is

$$0 = \max_{\pi_t, c_t} \mathbb{E}\left[dv(t, w_t) + e^{-\rho t} u(c_t)\right].$$

By Ito's lemma thus

$$0 = \max_{\pi_t, c_t} \mathbb{E}\left[\frac{\partial v}{\partial t} + \frac{\partial v}{\partial w} dw_t + \frac{1}{2}\frac{\partial^2 v}{\partial w^2}(dw_t)^2 + e^{-\rho t} u(c_t)\right],$$

or the Hamilton–Jacobi–Bellman equation

$$0 = \max_{\pi, c}\left[\frac{\partial v}{\partial t} + \left(\left(\pi\mu + (1-\pi)r\right) w - c\right)\frac{\partial v}{\partial w} + \frac{\pi^2\sigma^2 w^2}{2}\frac{\partial^2 v}{\partial w^2} + e^{-\rho t} u(c)\right], \tag{30.1}$$

where we have used the martingale property of W_t. □

[1] Robert C. Merton, 1944, US economist; Nobel Memorial Prize in Economic Sciences 1977

The first order conditions in (30.1) are found by computing the derivative with respect to π and c, they are

$$(\mu - r)\frac{\partial v}{\partial w} + \pi_t \sigma^2 w \frac{\partial^2 v}{\partial w^2} = 0$$

and

$$-\frac{\partial v}{\partial w} + e^{-\rho t} u'(c) = 0$$

so that

$$\pi^* = -\frac{(\mu - r)}{\sigma^2} \cdot \frac{\frac{\partial v}{\partial w}}{w \frac{\partial^2 v}{\partial w^2}} \tag{30.2}$$

and

$$c_t^* = u'^{-1}\left(e^{\rho t}\frac{\partial v}{\partial w}\right). \tag{30.3}$$

Substituting the optimal π^* and c^* in (30.1) gives the equation

$$0 = \frac{\partial v}{\partial t} - \frac{(\mu - r)^2 \left(\frac{\partial v}{\partial w}\right)^2}{2\sigma^2 \frac{\partial^2 v}{\partial w^2}} + rw\frac{\partial v}{\partial w} - u'^{-1}\left(e^{\rho t}\frac{\partial v}{\partial w}\right)\frac{\partial v}{\partial w} + e^{-\rho t} u\left(u'^{-1}\left(e^{\rho t}\frac{\partial v}{\partial w}\right)\right). \tag{30.4}$$

To continue we specify the utility function $u(c) \coloneqq \frac{c^{1-\gamma}}{1-\gamma}$. Note, that $u'(c) = c^{-\gamma}$ and $u'^{-1}(x) = x^{-1/\gamma}$, so that

$$c_t^* = \left(e^{\rho t}\frac{\partial v}{\partial w}\right)^{-\frac{1}{\gamma}}$$

from (30.3) and thus

$$-c^*\frac{\partial v}{\partial w} + e^{-\rho t} u(c^*) = -\left(e^{\rho t}\frac{\partial v}{\partial w}\right)^{-\frac{1}{\gamma}}\frac{\partial v}{\partial w} + \frac{e^{-\rho t}}{1-\gamma}\left(e^{\rho t}\frac{\partial v}{\partial w}\right)^{-\frac{1-\gamma}{\gamma}} = \frac{\gamma}{1-\gamma}e^{-\frac{\rho t}{\gamma}}\left(\frac{\partial v}{\partial w}\right)^{\frac{\gamma-1}{\gamma}}$$

and thus

$$0 = \frac{\partial v}{\partial t} - \frac{(\mu - r)^2}{2\sigma^2}\frac{\left(\frac{\partial v}{\partial w}\right)^2}{\frac{\partial^2 v}{\partial w^2}} + rw\frac{\partial v}{\partial w} + \frac{\gamma}{1-\gamma}e^{-\frac{\rho t}{\gamma}}\left(\frac{\partial v}{\partial w}\right)^{\frac{\gamma-1}{\gamma}}. \tag{30.5}$$

To linearize and solve this equation we try the ansatz $v(t, w) \coloneqq e^{-\rho t} f(t)^\gamma \frac{w^{1-\gamma}}{1-\gamma}$ so that

$$\begin{aligned}
\frac{\partial v}{\partial t} &= e^{-\rho t}\left(-\rho f(t)^\gamma + \gamma f(t)^{\gamma-1} f'(t)\right)\frac{w^{1-\gamma}}{1-\gamma}, \\
\frac{\partial v}{\partial w} &= e^{-\rho t} f(t)^\gamma w^{-\gamma}, \\
\frac{\partial^2 v}{\partial w^2} &= -\gamma e^{-\rho t} f(t)^\gamma w^{-\gamma-1}.
\end{aligned}$$

Hence (30.5) reads

$$0 = e^{-\rho t} w^{1-\gamma} f(t)^{\gamma-1}\left(-\frac{\rho}{1-\gamma}f(t) + \frac{\gamma}{1-\gamma}f'(t) - \frac{(\mu - r)^2}{2\sigma^2\gamma}f(t) + rf(t) + \frac{\gamma}{1-\gamma}\right)$$

or

$$f'(t) = \nu f(t) - 1$$

with $\nu = \frac{\rho-(1-\gamma)\left(\frac{(\mu-r)^2}{2\sigma^2\gamma}+r\right)}{\gamma}$. Employing the terminal condition $v(T,w) = e^{-\rho T}\epsilon^\gamma \frac{w^{1-\gamma}}{1-\gamma}$ we get the boundary condition $f(T) = \epsilon$. The general solution thus is

$$f(t) = \frac{1 + (\nu\epsilon - 1)e^{-\nu(T-t)}}{\nu}.$$

Substituting back in (30.2) gives the further results, in particular the (constant) Merton fraction.

Bibliography

[1] H. Albrecher, A. Binder, and P. Mayer. *Einführung in die Finanzmathematik*. Springer Science and Business Media, 2009. doi:10.1007/978-3-7643-8784-6. 3

[2] R. Cont and P. Tankov. *Financial Modeling With Jump Processes*. Chapman & Hall/ CRC, 2004. ISBN 1584884134. doi:10.1201/9780203485217. URL `https://books.google.com/books?id=bVlieV8GBrIC`. 178

[3] J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979. doi:10.1016/0304-405x(79)90015-1. 119

[4] R. A. Durrett. *Probability. Theory and Examples*. Duxbury Press, Belmont, CA, second edition, 2004. ISBN 0-534-42441-4. 73

[5] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, 1905. doi:10.1002/andp.19053220806. 93

[6] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*, volume 33 of *Stochastic Modelling and Applied Probability*. Springer, 1997. URL `http://books.google.com/books?id=BXOI2pICfJUC`. 58

[7] H. U. Gerber. *Life Insurance Mathematics*. Springer-Verlag Berlin and Heidelberg, 3rd edition, 1997. doi:10.1007/978-3-662-03460-6. 19, 153

[8] S. Janson. Stable distributions. Lecture notes, 2011. URL `http://www2.math.uu.se/~svante/papers/sjN12.pdf`. 177

[9] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 2002. doi:10.1007/b98838. 84

[10] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer-Verlag New York, 1991. doi:10.1007/978-1-4612-0949-2. URL `http://books.google.com/books?id=ATNy_Zg3PSsC`. 3, 97, 114, 179

[11] I. Karatzas and S. E. Shreve. *Methods of Mathematical Finance*. Stochastic Modelling and Applied Probability. Springer, 1998. doi:10.1007/b98840. 3

[12] H. Komiya. Elementary proof for Sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988. doi:10.2996/kmj/1138038812. 36

[13] E. L. Lehmann. Some concepts of dependence. In *Selected Works of E. L. Lehmann*, chapter 64, pages 811–827. Springer, 2012. doi:10.1007/978-1-4614-1412-4. 69

[14] S. Nadarajah. Linear combination of Gumbel random variables. *Stochastic Environmental Research and Risk Assessment*, 21(3):283–286, 2007. ISSN 1436-3240. doi:10.1007/s00477-006-0063-4. 60

[15] J. P. Nolan. *Stable distributions*. Springer New York, 2016. URL `http://fs2.american.edu/jpnolan/www/stable/chap1.pdf`. 177

[16] B. Øksendal. *Stochastic Differential Equations*. Springer, 6th edition, 2003. doi:10.1007/978-3-642-14394-6. 3, 65, 70, 99

[17] A. O'Sullivan and S. M. Sheffrin. *Economic Principles in Action*. Upper Saddle River, New Jersey 07458: Pearson Prentice Hall, 2003. 25

[18] G. Ch. Pflug. Stochastic models in finance. Lecture notes, 2016. 139

[19] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009. doi:10.1007/978-3-540-89500-8. 114

[20] T. Schmidt. Stochastische Finanzmärkte. Lecture notes, 2015. URL `http://docplayer.org/905344-Stochastische-finanzmaerkte.html`. 3

[21] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. MOS-SIAM Series on Optimization. SIAM, third edition, 2021. doi:10.1137/1.9781611976595. 76

[22] A. N. Shiryaev. *Probability*. Springer, New York, 1996. doi:10.1007/978-1-4757-2539-1. 84

[23] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. URL `https://projecteuclid.org/euclid.pjm/1103040253`. 36

[24] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi:10.1017/CBO9780511802256. URL `http://books.google.com/books?id=UEuQEM5RjWgC`. 61, 73, 76

[25] G. Žitković. Continuous time finance. Lecture notes, 2005. URL `https://www.ma.utexas.edu/users/gordanz/lecture_notes_page.html`. 3

[26] G. Žitković. Introduction to stochastic processes. Lecture notes, 2014. URL `https://www.ma.utexas.edu/users/gordanz/lecture_notes_page.html`. 3

[27] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991. doi:10.1017/CBO9780511813658. URL `http://books.google.com/books?id=RnOJeRpk0SEC`. 3, 64, 72, 87

Index