“The market can stay irrational much longer than you can stay solvent” (John Maynard Keynes)

“I’d be a bum on the street with a tin cup if the markets were always efficient” (Warren Buffett)

“Ничего не хоти, умрешь веселым” (Елизавета Быликина)

ISBN 978-3-00-046520-8
http://www.yetanotherquant.com

Contents

Preface

There are thousands of books on trading[1]. For instance, the query "trading" yielded (on 04.02.2014) 62986 results on amazon.com. However, the content is extremely repetitive from book to book. First of all most of them sell hope, not knowledge. It is actually obvious: a professional writer (professional in the sense that he writes for a living) wants his books to be sold. It is much easier to sell hope than knowledge (which is often bitter) to a mass reader. Such books usually tell you that there are several markets (equities, commodities, bonds, FX) and either one of these markets is claimed to suit a private trader particularly well or it is recommended to concentrate on profit opportunities, not on asset classes. Then comes a common warning that "most time the markets are unpredictable", or alternatively an encouragement that the stock prices grow in the long term. Then some very basic information on brokers and trading orders, probably something on fundamental analysis and market psychology. Most of these books are about technical analysis such as chart patterns, since the quantitative indicators like moving averages require some mathematics, which the most of mass readers do not possess.

As to concrete trading strategy recommendations, they are distinctly clustered (and sorted in order of the cluster size) as follows: nothing concrete, trend following, swing trading, buy and hold, pair/spread trading and "exotics" like seasonal trading and chasing penny stocks. With very few exceptions no book contains a detailed strategy description. But even if the latter is there, there is still no word about the backtesting. Money management is either completely disregarded or its importance is noted but without detailed arguments. Concrete recommendations are rare, the rules of thumb are to not risk more than x% of the capital per trade (x% is often 1% or 2%, sometimes up to 5%). Another common advice is to take the volatility into account but all they recommend is usually just to take smaller positions in volatile stocks.

On the other hand there are a lot of books on "modern" quantitative finance. While the most of them are mathematically flawless, their content is both unreadable for a mere mortal and has little to do with

[1] In this book the words "investment" and "trading" are used interchangeably.

reality. And of course they are very costly; $100+ is not an uncommon price.

In this sense this book is brand new. First of all, it does not sell hope and is not intended for "dummies". Rather it is for retail investors who already have some experience. It is also very useful for students who study [quantitative] finance. In order to read this book, you need a *working* knowledge of college mathematics[2]. On the other hand, this book is practically oriented and completely void of mathematical arrogance. Moreover, the content of the chapters 2 and 3 is very elementary but at the same time thoroughly explained. Even if you cope only with these chapters, the book is worth its price since you get a clear perception of what you can (and what you cannot) achieve in the market in the long term.

Modern investment is unthinkable without modern IT, so some programming skills are very helpful (though not necessary if you are ready to learn in parallel).

This book does not present any out of box super winning strategy (though I do disclose my personal investment approach). Such strategies - even if they do work - expire quickly. However, this book does provide you with a toolset to create (and what is more important, to test and evaluate) trading strategies.

I also take into account that (like me) the most of my readers have a fulltime job and a family. This implies that a *practical* investment activity should not be too time-consuming. For those, who are too busy and have very little time but still wish to invest this book provides some hints how to avoid common errors and pitfalls.

Since I live in Germany, I mostly watch the German market. Thus most of the examples are about German stocks. Though the German market does have its peculiarities (strong dependence on export, conservatism of German investors), these examples are straightforwardly applicable for any developed market.

[2]I mean US college level, which is approximately equal to the German Abitur or Russian high-school with emphasis on math.

Finally, it is worth explaining what motivated me to write this book. Well, first of all “when a man has anything to tell in this world, the difficulty is not to make him tell it, but to prevent him from telling it"[3]. Like every person I also want my efforts to be rewarded. However, I expect the main profit not from the book sales but rather from the increase in my goodwill: being an author of a *good* book usually helps in career. Keeping the book price under €15 I try to make it affordable for everyone and in particular for students. I hope this book will be interesting and useful to the readers and help to navigate in turbulent financial markets.

Writing a book is always a big undertaking, which requires a lot of time and concentration. This book would never be written without my wife Olesja who inspired and supported me all the time. Karla Penter did her best to translate my “Russian English” to the British English (all remained typos and grammar mistakes are completely my fault, not hers). I am very indebted to Dr. Thomas Rupp for his valuable remarks.

In order to improve the book I actively used ~~crowd~~folksourcing. I am grateful to everybody who has contributed, in particular to Matthias Siemering, Jörg Hiermayr, Dimitri Semenchenko, Dmitriy Bogdanis, Dr. Frank Wittemann, Dr. Peter Schwendner, Dr. Roland Stamm, Dr. Aleksey Min, Natalia Shenkman, Stanislav Narivonchik, Alex Ocnariu, Alexander Mora Araya, Daniel Lamparter, Daniel Schroeter, Elena Tichij, Florin Leist, Franziska Zerweck, Gloria Straub, Laila Unkauf, Oksana Mook, Roman Wenger, Sekou Cissé, Christiane Kandeler, Celine de Sousa and Vincenzo de Matteo.

[3] Caesar and Cleopatra by George Bernard Shaw (http://www.gutenberg.org/files/3329/3329-h/3329-h.htm#link2H_4_0003).

Chapter 1: A brief review of the probability theory

As I warned you, you need a working knowledge of college mathematics in order to read this book. For the sake of completeness, we briefly review the main ideas of the probability theory. We start with a quiz, which although elementary is not trivial. Moreover, it addresses both the typical problems you will encounter as an investor and some fine ideas of probability theory that often remain beyond the scope of the first course on probability. If the quiz is no problem for you then neither will be the rest of the book. But if you cannot comprehend it even after reading the solutions then... well, you might still succeed as investor. I know a couple of people that have trouble with this quiz but are still successful on the market (or at least they say that they are). However, I believe the investment should be a healthy blend of science and life experience rather than a curios mixture of subconsciousness and voodoo craft.

Q1: Initial prices of stocks A and B were, respectively, $10 and $100. Later the prices grew, respectively, to $20 and $150. Which stock performed better?

Q2: A portfolio consists of three stocks; their weights in portfolio are equal. After a year the first stock yielded 15%, the second -20% and the third one 10%. What is portfolio's total return?

Q3: A trader invested his capital in an ETF[4] on RTS[5]. The return for the first year was 10%, for the second year 15% but for the third year -20%. After the third year a trader sold his portfolio. How much has he earned?

Q4: A similar situation as by Q3 but now the returns were -20%, 15% and 10%. Does the order of returns really matter?

[4] Exchange Traded Fund

[5] Russian stock index

Q5: A stock dropped by *x%* then grew by *y%* and returned to its initial price. Write a formula to express *y* via *x*.

Q6: A trader bought a stock, held it for three months and then sold, yielding 5%. What is his annualized growth rate?

Q7: If we want to double our wealth in five years, which annualized growth rate do we need?

Q8: One thousand people took a new medicine. One of them had an allergic reaction. What is the probability of the allergic response?

Q9: A deck consists of 36 cards[6]. What are the probabilities a) to get a queen, b) to get spades, c) to get the queen of spades?

Q10: One has tossed a coin ten times and got ten heads. What is the probability to get a head by the eleventh toss?

Q11: The first bank account pays 6% annually, the second pays 3% semi-annually, the third pays 1.5% quarterly and the fourth pays 0.5% monthly. Which bank account is better (assuming they all can be considered risk-free)?

[6] Such "abridged" decks are most common in Russia. There are 9 values from six to ace and 4 suits.

A1: The return on the first stock is (20-10)/10 = 1 = 100%, whereas the return on the second stock is (150-100)/100 = 0.5 = 50%. Thus the first stock has performed better. The absolute price changes (in this case, respectively, $10 and $50) do not matter: if an investor has, e.g. $100, he can buy either ten stocks A or one stock B. In the first case his gain will be 10*$10 = $100 and in the second case just 1*$50=$50.

A2: Assume a trader initially had *x* dollars. The stock weights in portfolio are equal, i.e. he has invested *x/3* dollars in each stock. His terminal wealth is then (1+0.15)*x/3 +* (1-0.2)*x/3* + (1+0.1)*x/3* = (3 + 0.05)*x/3* = *x* + 0.0167*x*. Since his initial capital was *x*, the total return is equal to (*x* + 0.0167*x - x) / x = 0.0167.* Of course we could just have calculated the [simple] average of returns: (0.15 - 0.2 + 0.1) / 3 = 0.0167 = 1.67%. In general, the weights of stocks in a portfolio are not the same. In this case the portfolio return is equal to $\sum\limits_{i=1}^{n} w_i r_i / \sum\limits_{i=1}^{n} w_i$, where w_i and r_i are the weight and the return of the *i-th* stock.

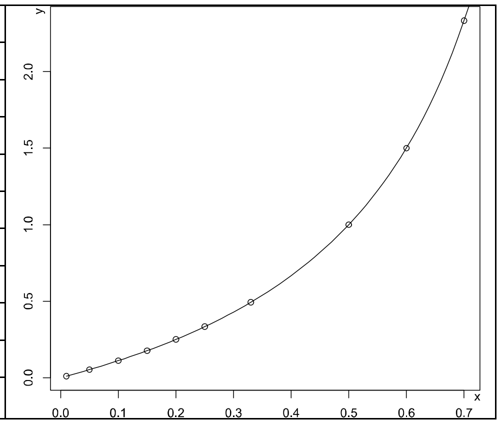
A3: Assume a trader initially had *x* dollars. After the first year the return was 10%, so the total wealth after the first year is *x + 0.1x =x(1+0.1)= 1.1x .* The return for the second year was 15% and a trader had *1.1x* dollars at the beginning of the second year. That is, his wealth at the end of the second year is *1.1x(1+0.15) = 1.265x* . Analogously, the wealth at the end of the third year is *1.265x(1-0.2) = 1.012x* . So the total return is 1.2%. And the *annualized* return (s. A6) is just 0.4%. The lesson from Q2 and Q3: never confuse an arithmetic mean and a compound annual growth rate (CAGR).

A4: As you likely noticed, the formula to calculate the total return for the case of Q3 is (1+0.1)(1+0.15)(1-0.2) - 1 . Since the product is commutative, the order of annual returns does not matter. However, it likely matters from a psychological point of view: especially for a newbie the scenario -20%, 15%, 10% is much less comfortable than 10%, 15%, -20%. As a matter of fact, most of trading software displays only the acquisition price and the current price. So in the first case a trader will always see a profit. Though the profit decreases from 26.5% to 1.2%, comparing the acquisition price with the current price we are still in the black. But in the second case we first need to endure the loss of 20%. Note that the experienced traders usually consider the *maximum drawdown*, which does not depend on the order of returns and is equal to -20% in both cases. Further we will use the maximum drawdown as the main measure of risk.

A5: It holds: 1 = (1-*x*)(1+*y*) thus $y = \dfrac{1}{1-x} - 1$

Let us calculate some concrete values, which you should learn by heart:

-x	**y**
-1.00%	1.01%
-5.00%	5.26%
-10.00%	11.11%
-15.00%	17.65%
-20.00%	25.00%
-25.00%	33.33%
-33.33%	50.00%
-50.00%	100.00%
-60.00%	150.00%
-70.00%	233.33%



A6: The question is not as elementary as it may seem to be. First of all, the answer depends on the so-called day count convention. The simplest one is 30/360, i.e. for simplicity's sake we assume 30 days in every month and 360 days in every year. Thus [any] three months make exactly a quarter of year and it holds: $(1+r)^{0.25} = 1.05$ Taking logs we obtain $0.25\ln(1+r) = \ln(1.05) = 0.04879$ Exponentiate and get $1 + r = \exp(0.19516) = 1.21550$ Finally $r = 0.21550$ We check the solution: $(1 + 0.21550)^{0.25} = 1.04999865 \approx 1.05$

The general formula 1-A6-1 is not the only feasible one (s. A11). However, it is most common in [practical] finance.

$$(1+r)^t = 1+R$$
$$r = \exp\left(\frac{\ln(1+R)}{t}\right) - 1$$
Formula 1-A6-1 CAGR or the annualization of the interest rate *R* accrued during the time *t*.

Note that the 5% that was yielded after the three months are called [quarterly] *simple return.* The average (or the arithmetic mean) of the simple returns over several periods is called *arithmetic return*. The arithmetic return is an important number for a series of bets (or trades) *without* reinvestment of winnings. It is also used to find the optimal portfolios (s. the chapter on Kelly criterion). But in order to characterize a series of trades with reinvestments we need CAGR or, alternatively, a *geometric mean* return over several periods (CAGR is more convenient since it also allows to compare two strategies with different investment periods).

A7: It holds $(1+r)^5 = 2$ thus according to the formula 1-A6-1 $r = 0.14870$ which is pretty ambitious. For comparison: the DAX (German counterpart of Dow Jones Industrial Average) has a long-term CAGR about 8%. On the other hand Warren Buffet achieves 20% CAGR.

A8: This question is not trivial. A correct but an incomplete answer is as follows: one in a thousand is 0.1% thus the probability of the allergic response is also 0.1%. According to Thilo Sarrazin only 46%

of Germans and 25% of Americans can provide this answer[7]. The problem is, however, that 0.1% is just an empirical estimation of the [genuine] probability but not the genuine probability itself. The more the number of trials (in this case the number of persons that took the medicine) the better is the convergence of the estimate to the genuine probability. The underlying theory is far from being easy. Fortunately there is an accessible alternative: the Monte Carlo simulation, which will be one of the most important tools for us.

A9: Among 36 cards there are 4 queens. For a well-shuffled deck each card can be drawn with equal probability, which is 1/36. In four cases there can be a queen, so the total probability is 4/36 = 1/9 = 11.11%. Analogously there are 9 spades thus the probability to get a spade is 1/4 = 25%. Finally, there is only one queen of spades, so the probability to get it is 1/36 = 2.78%. Note that the suit of a card is obviously independent from its value. On the other hand the probability to draw the queen of spades is numerically equal to the product of probabilities to draw a queen and to draw spades. In general, if the events A and B are independent then the probability that both A and B will occur is equal to the product of their probabilities. Moreover, the converse is also true, i.e. if the formula 1-A9-1 holds then the events A and B are independent.

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

Formula 1-A9-1 Probability law for independent events

A10: This case is a perfect demonstration of the "contradictions" between the empirical and the genuine probability or, as some people say, between mathematics and physics. A "physicist" will conclude that the coin is heavily biased, so the probability to get a

[7] T. Sarrazin, "Deutschland schafft sich ab". 6th ed(2010), p. 196
Thilo Sarrazin (born in 1945) was a very successful senator for finance in Berlin and a member of the Executive Board of the German Central Bank (Bundesbank).

head by the next toss shall be close to 1. However, a "mathematician" will assume a fair coin and thus state that the probability to get a head by the 11th toss is still 0.5, since the toss outcomes are independent from each other[8]. Note that *if* the coin is fair, the probability to get 10 heads by 10 tosses is, according to the formula 1-A9-1 equal to $(0.5)^{10} = 0.0009765625 \approx 0.1\%$. It is experience and not mathematics that lets us decide whether to consider an event virtually impossible or pretty rare but still possible. In case of a *fair* coin I would rather say, it is virtually impossible to get 10 heads by 10 tosses. But an allergic reaction of one patient among one thousand does not imply (at least to me) that the allergy is virtually impossible, though the probability of this event is 0.1% as well... Continuing on with this idea, there is one more example, which I learnt from the wonderful textbook on probability theory by Elena Wenzel[9]: if you are an artillerist, you likely can accept that one of a thousand shells will not explode by hitting a target. But if you are a paratrooper, the probability of 0.1% to have your parachute undeployed will likely be too high for you.

A11: The first bank account pays 6% return to the end of the year. The second pays off $(1+0.03)^2 = 1.0609$ (first payment takes place after six months and the second to the end of the year). Analogously for the third account we have $(1+0.015)^4 = 1.06136$ and $(1+0.005)^{12} = 1.06168$ for the fourth. So there is a difference, although not really a big one. What if we continue this process to its limit (and does this limit exist at all)?! Yes, it does! This is known as *continuously compounded interest* and it is very popular in the theory of mathematical finance.

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{\frac{t}{n}} = \exp(rt)$$

Formula 1-A11-1 Continuously compounded interest

8 In other words they are serially independent.

9 Elena Sergeevna Wentzel (1907-2002) was a prominent Soviet mathematician and writer (she wrote under pseudonym I. Grekova that literally means "Mrs. Y").

We can also obtain the formula 1-A11-1 the other way around. Assume that the wealth in a bank account grows proportionally to its current value W_t with coefficient r, i.e. the following differential equation holds: $dW_t = rW_t dt$. The solution is $W_t = W_0 \exp(rt)$ where W_0 is the initial wealth.

Very close is the idea of the *logarithmic returns*, i.e. if the initial wealth was W_0 , the terminal wealth is equal to W_1 and the investment period was t years then the logarithmic return is

$$r = \frac{1}{t}\ln\left(\frac{W_1}{W_0}\right) = \frac{\ln(W_1) - \ln(W_0)}{t}$$

Formula 1-A11-2 Logarithmic returns

Indeed, if the interest is compounded continuously then $W_1 = W_0 \exp(rt)$ thus 1-A11-A2 holds. Logarithmic returns are convenient for continuous time models[10]. Since the stock trading is essentially discrete, we will mostly use compound returns according to 1-A6-1.

Now we turn to the probability theory. As you may know, it originates from gambling. The easiest case is a symmetric coin: the probabilities of head (H) and tail (T) are both equal to 0.5. However, the question "what is the probability to get *at least* one head if we toss a symmetric coin twice" is not completely trivial. Still we can enumerate all possible outcomes. They are HH, HT, TH and TT. The first three outcomes suit our criterion. Since the coin is symmetric, they all have the same probability, which is equal to 0.25; recall that the probabilities of all possible (and mutually exclusive) outcomes

[10] In particular due to the fact that a compound return on a long and non-leveraged position can take values in [-1, ∞], whereas a logarithmic return can lie in [$-\infty$, ∞] and thus can be straightforwardly modeled by continuous probability distribution like the normal distribution.

must sum to one. Thus the probability of the event to get at least one head by two tosses is 0.75 or 75%.

So far so good but what if we want to calculate, say, the probability to get at least 60 heads by 100 tosses?! The enumeration of the outcomes does not seem to be feasible anymore. We still could have solved this problem analytically but we actually need not. Instead we can engage R, an opensource and royalty-free statistical software.

```
#exact value
print("Exact Solution")
1 - pbinom(59, 100, 0.5)

#approximate value by Monte Carlo simulation
tosses = rbinom(100000, 100, 0.5)
nSuccessfulOutcomes = 0
for(i in 1:100000)
{
 if(tosses[i] >=60 )
   nSuccessfulOutcomes = nSuccessfulOutcomes + 1
}
print("Approximate Solution")
nSuccessfulOutcomes / 100000
```

R-code 1.1 Probability of at least 60 heads by 100 tosses

If you never heard about R before, don't worry. So far just go to http://www.r-project.org, download and install R (the installation on Windows is straightforward) and enter the R-code 1.1 to the command line. Further we will learn the basics of R iteratively step by step. Let us discuss the R-code 1.1 in detail. The command `pbinom(59, 100, 0.5)` in the 3rd line gives the exact value of the probability that there will be no more than 59 heads by 100 tosses of a symmetric coin (0.5 stands for the probability of head). Respectively, the probability that there will be 60 or more heads is `1 - pbinom(59, 100, 0.5)`.

Here we made use of our knowledge of the probability distribution. But this will not always be the case. For example, we may draw the stock returns from a (sophisticated) distribution and have no idea of the distribution of terminal wealth (though the distribution of returns is known). But it is absolutely no problem as long as we have R at hand[11]. All we need to do is just to run a Monte Carlo simulation, which is, in our case, nothing else but a computer simulation of the coin tosses. The command `tosses = rbinom(100000, 100, 0.5)` tells R to simulate 100 tosses of a symmetric coin, calculate the sum of heads, repeat the process 100000 times and put the values to the array "tosses". Command `nSuccessfulOutcomes = 0` initiates the counter of outcomes with 60 heads or more. Then we run through all outcomes with a *for*-loop and check whether we got the desired number of heads. If yes, we increment *nSuccessfulOutcomes* by one. In the last code line `nSuccessfulOutcomes / 100000` we divide the number of successful outcomes by the total number of trials. The more it is, the better the final result converges to the genuine probability of the event to get 60 heads or more.

In my case the genuine probability was 0.02844397 (and will obviously be the same in your case). As to the approximate probability, I yielded 0.02896 which is pretty close to the genuine value. However, in your case it will be slightly different, though most likely also close to the true value.

If you are completely new to programming, do the following: 1. Read more about arrays (in principle, an array is just a sequence of values). 2. Read about loops (loops are used to repeat the same or the similar actions many times). Besides *for*-loop, learn also *while*-loop. 3. Learn the conditional operator *if* and its extension *if ... else* as well as logical operators *and*, *or*, *not* (in R they are, respectively, written as *&&*, *||*, *!*) . 4. Note the difference of the assignment operator = and its synonym <- vs. "equal to" operator == .
Exercise 1.1 (programming)

[11] The only problem we may encounter is the computational intensity.

By means of this simple Monte Carlo simulation we addressed two important topics: the (discrete) stochastic processes and (the convergence to) the expectation of a random variable. Let us so far postpone the former and discuss the latter. If we assign 1 to head and 0 to tail, the expected value of a single coin toss is 0.5. Recall the definition of the expectation:

$$\mathbb{E}[X] := \sum_{i=1}^{n} p_i X_i$$

Definition 1.1 Expectation of a discrete random variable

where X_i is the value of a random variable in case of outcome i and p_i is the probability of this outcome. In our example let[12] $X_1 := \{X = T\} = 0$, $X_2 := \{X = H\} = 1$
Since the coin is symmetric, $p_1 = p_2 = 0.5$ and $\mathbb{E}[X] = p_1 X_1 + p_2 X_2 = 0.5 \cdot 0 + 0.5 \cdot 1 = 0.5$
Note that sometimes we can calculate the expectation just by intuition[13]. Indeed, if we toss a symmetric coin 100 times, we should get approximately the same number of heads and tails. Since we assigned 1 to head and 0 to tail, the expectation is 50.

Now let us note a very important idea, which you (or your instructor) may have missed out in your first course on probability.

In case of a <u>single</u> toss (or just <u>a few</u> tosses) the expectation is not really meaningful!

☝

[12] ":=" means "equal by definition". The expression $X_1 := \{X = T\}$ is read as follows: by definition let the first outcome take place when we got a tail by a coin toss.

[13] But be careful, in mathematics the intuition can easily let you down! That's why always verify it by a Monte Carlo simulation!

Indeed, the expectation is just the average value, but does averaging really make sense in case of just one coin toss? The situation radically changes if we toss a coin many times (the more, the better). By hundred tosses the probability to get *exactly* the expected value, i.e. 50, is about 8% which does not seem to be very large. However, the probability that the outcome deviates from the expected value by no more than ±10 is more than 95%! You can check it with R command `pbinom(60, 100, 0.5) - pbinom(39,100,0.5)`. It means that in a sense *the random outcome becomes less and less random as we increase the number of trials*. In other words the mean value of the experiment converges to its expectation.

Here we smoothly step to the next important concept: the variance and the standard deviation (that are direct relatives of the assets volatility). Let us slightly modify our example: instead of a coin toss consider a stock A, which can either fall 5% down or grow 10% up. Further let stock B go either 30% up or 25% down. Note that in practice we can often estimate these numbers relatively precisely and even *define* them, e.g. if we set our take profit and stop loss orders accordingly[14]. Let the probabilities[15] of both events be 0.5. According to the definition 1.1 both stocks have the same expected return. Indeed: $0.5 \cdot 0.1 + 0.5 \cdot (-0.05) = 0.025 = 2.5\%$ and $0.5 \cdot 0.3 + 0.5 \cdot (-0.25) = 0.025 = 2.5\%$.
However, you (should) intuitively feel that the stock B is much more risky. Indeed, for both stocks we expect the same return but we know that by the stock A we will not lose more than 5% of our investment even in the worst case. Or in other words, the worst possible deviation from the expected return is equal to -7.5%. But for the stock B it is -27.5%!

[14] Of course there is always a chance that the traded asset will reach neither stop loss nor take profit. However, experienced traders rarely encounter this case, later we will discuss why.
[15] As to the probabilities, they are really hard to estimate. However, it is not an insuperable hindrance. Later we will consider how to deal with it.

At this point we are getting somewhat messy since we mix the variance and the (maximum) drawdown risk. They are not the same. However, as long as the returns never deviate too far from their expectations[16], the drawdown risk and the variance are very closely related! You will see this in the next chapter.

Stop reading for a while and have a cup of tea. Imagine that you have lost 5% of your savings. How do you feel? Now imagine that the loss is 25%. Have you planned to buy a house or a new car? Or maybe an expensive vacation? Which loss would be acceptable for you so that you still could afford all your plans?
Exercise 1.2 (psychology)

Now let us formally define the variance and the standard deviation.

$$\mathbb{VAR}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Definition 1.2 Variance of a random variable

$$\mathrm{StdDev}[X] := \sqrt{\mathbb{VAR}[X]}$$

Definition 1.3 Standard deviation of a random variable

Essentially, *the standard deviation measures the dispersion of the outcomes from their expected value*. You may wonder why we at first consider $\mathbb{E}[(X - \mathbb{E}[X])^2]$ and then take the square root instead of simply considering $\mathbb{E}[(X - \mathbb{E}[X])]$. As a matter of fact, $\mathbb{E}[(X - \mathbb{E}[X])] = 0$ (as a good exercise, prove this statement! Just apply the definition 1.1. As a useful corollary you will get that

[16] Or, in mathematical terms, as long as the return distribution has no fat tails.

$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$ and, in general, the expectation of a constant is this constant itself). Additionally, the advantages of the definitions 1.2 and 1.3 become evident as soon as one encounters the *linear regression*.

Let us calculate the variance of returns on stocks A and B. For convenience, we summarize the possible outcomes and their probabilities. In other words, we specify the probability distributions for the random returns of the both stocks.

Probability	Return on stock A	Return on stock B
0.5	-0.05	-0.25
0.5	0.1	0.3

We have already calculated that $\mathbb{E}[A] = \mathbb{E}[B] = 0.025$ According to the definitions 1.2 and 1.3, we calculate
$\mathbb{VAR}[A] = 0.5 \cdot (0.1 - 0.025)^2 + 0.5 \cdot (-0.05 - 0.025)^2 = 0.005625$
$\mathbb{VAR}[B] = 0.5 \cdot (0.3 - 0.025)^2 + 0.5 \cdot (-0.25 - 0.025)^2 = 0.075625$
Hence $\mathrm{StdDev}[A] = 0.075$ and $\mathrm{StdDev}[B] = 0.275$

Chapter 2: Money Management according to Kelly criterion: the first encounter

Every highly qualified telecommunication engineer knows the (basics of) information theory. This theory does not tell him how to design the devices with the maximum bandpass but it does tell him which bandpass he can theoretically achieve. Since our engineer is highly qualified, he knows that practically achievable bandpass will be somewhat below the theoretical one.
We - the financiers - have our own analogue of the information theory that tells us, which maximum expected growth rate, i.e. which CAGR, we can achieve in the long term. This "fortune's formula" is known as Kelly criterion. Surprisingly, it was Claude Shannon (the father of the information theory), who significantly contributed to its discovery. Even more surprising is that very few financial professionals are (really) aware about Kelly criterion. At the same time every gambler knows it.

A short historical review of Kelly criterion is quite appropriate here[17]. The original paper "A New Interpretation of Information Rate"[18] was written by J. L. Kelly, JR. in 1956. In essence, it considers a favorable game[19] and the optimal fraction of gambling capital, which one should bet by each stake. But from the traders' and bettors' point of view the published version is pretty vague. Though one can grasp the main idea, the paper is rather about communication channels than about optimal stakes. Allegedly, there was an earlier version, written both by Kelly and Shannon. In this version the author(s) freely talked about bookies and insiders. But both authors worked for AT&T, whose management "was never

---

[17] For the most part I follow James Case's paper "The Kelly Criterion: Fallacy or Kuhnian Paradigm Shift Waiting to Happen?" (SIAM News, Volume 39, Number 3, April 2006, available at http://www.siam.org/pdf/news/930.pdf)

[18] Available at https://www.princeton.edu/~wbialek/rome/refs/kelly_56.pdf

[19] Favorable game is a game with a positive expectation. The simplest case is tossing of a non-symmetric coin. Oppositely, in case of a symmetric coin the expectation is zero, so one speaks about *fair* game.

keen to advertise the fact that bookies long represented an embarrassingly large fraction of the firm's customer base". As a result, the authors prepared a "more politically correct" version, signed by Kelly alone.

A couple of years later Edward Thorp[20], a professor of mathematics at MIT found out how to get an edge in blackjack. Shannon was a good friend of Thorp and drew his attention to Kelly's paper. Kelly's approach allowed Thorp to maximize *in long term* the advantage of his edge in blackjack. Thorp applied his strategy in Las Vegas casinos and made good money. Successful gamblers are quickly recognized and expelled from casinos, so Thorp had to disguise himself in a combination of wraparound glasses and a beard in order to avoid expulsion.

Yet a casino affair was just a warm-up for Thorp. His real deal was the statistical arbitrage, i.e. the exploration of the market imperfections that though not necessarily guarantees a riskless profit but provides a sure edge "on average". In an efficient market there are no such imperfections but Thorp never believed the markets were efficient. ‹‹In the late 1970s affordable, powerful computers and high quality databases were becoming more affordable, making a revolution in Finance possible... The idea was to rank stocks by their percentage change in price, corrected for splits and dividends, over a recent past period such as the last two weeks. We found that the stocks that were most up tended to fall relative to the market over the next few weeks and the stocks which were the most down tended to rise relative to the market. Using this forecast our computer simulations showed approximately a 20 percent annualized return from buying the "best" decile of stocks, and selling short the "worst" decile[21]››. Prof. Thorp is a seminal writer; however, he never disclosed *in detail* how he applied Kelly criterion to his portfolio of *many* assets. We will

[20] http://en.wikipedia.org/wiki/Edward_O._Thorp

[21] Ed Thorp, A Mathematician on Wall Street. Statistical Arbitrage - Part II. http://www.wilmott.com/pdfs/080630_thorp.pdf

discuss this issue later but so far consider the Kelly criterion in its simplest univariate form.

Assume a gambler tosses a biased coin so that the probability *p* to get a tail is known and larger than 0.5. (In our terms, it is a favorable game and a gambler has an edge). After each bet a gambler loses or doubles the money at stake. Obviously, he wants to exploit his edge completely and at first glance the idea to maximize his *expected terminal wealth* (recall definition 1.1) does not look implausible. Let W_0 be his initial capital and let *u* be the fraction that the gambler bets at each stake. Respectively, he lays aside *(1-u)*. Then the expected capital after the first bet is

$$W_0[2pu + (1-p)\cdot 0 + (1-u)] = W_0[2pu + 1 - u]$$

Respectively, the expected capital after the second bet is

$$\begin{aligned} W_0[2pu + 1 - u]2pu + W_0[2pu + 1 - u](1-u) \\ = W_0[2pu + 1 - u]^2 \end{aligned}$$

For *n* bets one has $\mathbb{E}[W_n] = W_0(2pu + 1 - u)^n$. Obviously, this expression is growing with *u* (recall, *p* > 0.5 thus *2p* > 1). Thus in order to maximize the expected terminal wealth a gambler should put at stake all his capital by each bet. However, this is too risky because each bet looms the danger to lose everything and as the number *n* of bets gets larger and larger a gambler *will* eventually go bankrupt (recall Q10 from the quiz). Thus – if the winnings are *re-invested* - the idea to maximize the expected terminal wealth is actually bad and our gambler needs another approach. On the other hand if he bets nothing (i.e. lets *u* = 0) he will make no use of his edge. So the optimal fraction is somewhere between zero and one, but where?!

As an alternative to the maximization of the expected terminal wealth, Kelly suggested to maximize the *expected growth rate* (recall Q5 and Q7). Indeed, one can make an outstanding series of returns, say 20%, 30% and 80%, which corresponds to a total return of (1+0.2)(1+0.3)(1+0.8) - 1 = 184%!
But then one may have a bad luck and get a return of -70%. Then the total return will be -34.48% (recall Q5)! In this sense the

maximization of the expected growth rate means that we limit the aftermath of severe negative returns, because just one such negative return can drastically reduce our wealth. At the same time we sufficiently participate in positive returns. Thus in the long run Kelly's approach beats *any other* approach!

As you may have noted, in case of re-investment of winnings the wealth grows exponentially in time, i.e. the time (or the number of investment periods) stays in the exponent (recall formula 1-A6-1). Thus the maximization of the expected growth rate actually means the *maximization of expected logarithmic terminal wealth* $\ln(W_n)$. If we hold an asset for *n* periods for which the returns are random than we have

$$W_n = W_0 \cdot (1+r_1) \cdot (1+r_2) \cdot \ldots \cdot (1+r_n)$$

Taking logarithm we turn this product into a sum and maximize

$$\mathbb{E}[\ln(1+r_1) + \ln(1+r_2) + \ldots + \ln(1+r_n)]$$

A reader, who is familiar with utility functions, will readily ask whether we just maximize the logarithmic utility. In a sense, yes, and Kelly did recognize this aspect of his approach. But he always pointed at the fact that his argumentation has nothing to do with utility functions and this connection to the logarithmic utility is just a coincidence. However, it is a very happy coincidence! First of all because the logarithmic utility is *myopic*[22]. That is, though we have many bets (or investment periods), we need to act by each bet as *if it were the last one*! Normally, this is not the case and one optimizes running backwards[23]. Indeed, as Elena Wenzel explains, why should we optimize the first step when the second step takes

[22] Logarithmic utility has many other nice properties. In general, some utility functions (first of all the logarithmic and negative power utility) are much more useful than the others. Unfortunately, the textbooks on economics usually consider the utility functions very superficially, so that the students get a wrong impression that utility functions is a nice theoretical concept, which is, however, useless in practice.

[23] This approach was introduced by Richard Bellman(1920 - 1984), who called it dynamic programming. By that time the words "computers" and "programming" were the buzz words and Bellman made use of it in order to obtain a grant for his research.

us – directly or figuratively – in a swamp?! That's why we should rather trace our route backward from destination to the start and optimize iteratively. For this we need to know all possible routes and the number of steps. However, in case of an [active] wealth management we usually do not know how many trades we will commit. But in case of acting according to the Kelly criterion we simply do not need to know it! We just optimize every trade as if it were the last one!

Moreover, under fairly general conditions the Kelly criterion maximizes the *median*[24] of the terminal wealth. As a measure of central tendency, the median is often preferred to the expectation. Indeed, assume we need to sum up with a single number the information about the wealth of a group of five people (such as an owner of a company and his employees). If a business owner earns \$1000000 per year but pays his employees relatively modest annual salaries of \$30000, \$40000, \$50000 and \$60000 then the mean (or "expected" income) is \$236000 but the median is \$50000, which is obviously much more representative.

Now let us consider the following example[25]: assume there is a stock[26] that yields for each dollar invested, \$2.70 or only \$0.30, both with probability 1/2. According to the definition 1.1 this is a perfect investment opportunity: the expected wealth is $0.5 \cdot (2.7 + 0.3) = 1.5$, i.e. the expected return is 50%! So far, so good and it does not seem implausible to invest all our capital in

---

[24] Ethier, S. N. (2004). The Kelly system maximizes median fortune. Journal of Applied Probability, 41(5):1230 – 1236

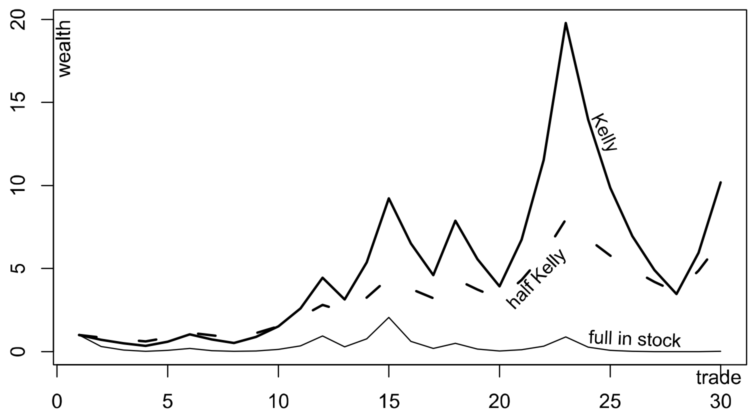
[25] It is Samuelson's example. Paul Samuelson (1915 - 2009) was a Nobel Prize winner in economics and the main critic of Kelly criterion, see "The "fallacy" of maximizing the geometric mean in long sequences of investing or gambling". Proceedings of the National Academy of Sciences, 68(10), 2493-2496). In this sense, it is especially interesting to beat such a prominent critic with his own example. Samuelson actually imposed a constraint either to invest in stock *completely* or not at all. But who can impose such constraint to a trader in a free society and a liberal market?!

[26] And there *are* suchlike stocks, e.g. Nokia (FI0009000681)

this stock. But look what happens in the long run! Since the probabilities to go up and down are equal, we can expect that the number of ups and downs will be approximately equal. Recall our example with 100 coin tosses, it is completely analogous: if we hold this stock for 100 investment periods, the probability that the number of downs remains between 40 and 60 is more than 95%. By 100 trades and a typical 50/50 scenario we yield $[(0.3)(2.7)]^{50} = [0.81]^{50} \approx 0$, i.e. we are broke! In order to make a profit we need at least 55 ups: $(0.3)^{45} \cdot (2.7)^{55} = 1.56842$. But the probability to get 55 or more ups is just 18.41% (check it with R analogously to R-Code 1.1)! Respectively, the probability to make loss is 81.59%, which is too much for a typical risk-averse investor.
But look what happens if we always hold exactly 42% of our wealth in stock and the rest in cash. It means that after each trade we sell a part of stocks if they went up and buy if they went down so that by each trade the wealth fraction invested in stock remains 42%. In this case we yield by a typical scenario

$$([\underbrace{(1-0.42)}_{\text{in cash}} + 2.7 \cdot \underbrace{(0.42)}_{\text{in stock}}][\underbrace{(1-0.42)}_{\text{in cash}} + 0.3 \cdot \underbrace{(0.42)}_{\text{in stock}}])^{50} = [(1+1.7\cdot 0.42)(1-0.7\cdot 0.42)]^{50} = [(1.714)\cdot(0.706)]^{50} = 13828.53$$

What an improvement just by means of a clever money management! Moreover, even if we encounter the worst of the typical scenarios, i.e. just 40 ups and 60 downs we still make profit since $(1.714)^{40} \cdot (0.706)^{60} = 1.94$. That is, with probability of more than 95% we do make profit!




```
install.packages("tseries")
library(tseries)
N_TRADES = 30
outcomes = rbinom(N_TRADES, 1, 0.5)
wealth = array(1.0, dim=N_TRADES)
wealthKelly = array(1.0, dim=N_TRADES)
wealthHalfKelly = array(1.0, dim=N_TRADES)
for( i in 2:(length(outcomes)) )
{
  if(outcomes[i] == 0) {
    wealth[i] = wealth[i-1] * (1 - 0.7)
    wealthKelly[i] = wealthKelly[i-1] * 0.42 *
     (1 - 0.7) + wealthKelly[i-1] * (1 - 0.42)

    wealthHalfKelly[i] = wealthHalfKelly[i-1] * 0.21 *
     (1 - 0.7) + wealthHalfKelly[i-1] * (1 - 0.21)
  }
  else {
    wealth[i] = wealth[i-1] * (1 + 1.7)
    wealthKelly[i] = wealthKelly[i-1] * 0.42 *
     (1 + 1.7) + wealthKelly[i-1] * (1 - 0.42)

    wealthHalfKelly[i] = wealthHalfKelly[i-1] * 0.21 *
     (1 + 1.7) + wealthHalfKelly[i-1] * (1 - 0.21)
    }
}
```

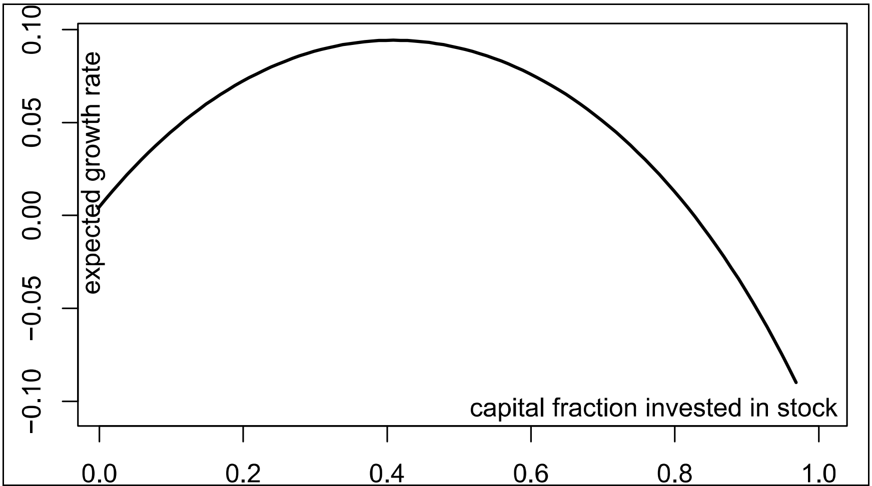
```
chartYmin = min(c(wealth, wealthKelly, wealthHalfKelly))
chartYmax = max(c(wealth, wealthKelly, wealthHalfKelly))
ts.plot(wealth, lwd=1, ylim=c(chartYmin, chartYmax))
lines(wealthKelly, lwd=2,)
lines(wealthHalfKelly, lty=2, lwd=2)
```

Figure 2.1 Kelly, half-Kelly and full investment after 30 trades (a sample scenario).

The constant fraction of 42% is nothing else but the optimal solution according to the Kelly criterion! Since the logarithmic utility is myopic, it suffices to consider a single trade. So we need to maximize

$$0.5\ln(1+1.7u)+0.5\ln(1-0.7u)$$

It is $u=0.42$ that maximizes this expression. You can check it analytically if you have a working knowledge of calculus. Or, alternatively, we can find it numerically with R.




```
u = seq(1,100)/100 #fractions
expectedGrowthRates = 0.5*(log(1 + 1.7*u) + log(1 - 0.7*u))
plot(u, expectedGrowthRates, type="l", lwd=2)
abline(v=0.42, col="grey")
abline(h=expectedGrowthRates[42], col="grey")
abline(h=expectedGrowthRates[100], col="grey")
which.max(expectedGrowthRates)
```

Figure 2.2 Expected growth rate vs. capital fraction in stock.

But looking at figure 2.1 you will probably say that even Kelly strategy is too risky: though the terminal wealth is good the intermediate drop from 20 to 4 (i.e. the drawdown of -80%) is not acceptable. In this case you should reduce the optimal fraction; a common practice is to halve it. This is so called half-Kelly strategy. Though its terminal performance is not as good as Kelly's (asymptotically Kelly beats *any* other strategy) but it is much less risky. Look at figure 2.2! You see that if you bet less or more than the optimal Kelly fraction, your expected growth rate decreases. So if you underbet, you decrease both expected growth rate *and* risk. But if you overbet, the growth rate decreases but the risk grows! Since in practice we can only approximately estimate the optimal Kelly fraction (recall the case "genuine probability vs. its empirical estimate"), we would better underbet. This is also the reason why the half-Kelly strategy is so popular among practitioners.

We see that in Samuelson's example the *maximum* possible expected growth rate is about 9.5%. It tells us what we can, *in principle,* "squeeze" from this investment opportunity in the long term[27]. Actually, such growth rate is not bad (it is better than by DAX) but it would not even double your wealth in five years (recall Q7).

Thus never believe books or people that promise to make you millionaire. I do not affirm that it is principally impossible to turn a couple [of thousands] of bucks into millions. Sometimes it happens, as for instance James "RevShark" DePorre describes in his book "Invest Like a Shark: How a Deaf Guy with No Job and Limited Capital Made a Fortune Investing in the Stock Market". The trick is that Mr. DePorre was lucky to invest in vastly trending markets. Actually, he bought bubbles and sold them at the right time. Currently, he offers wealth management according to his "Sharkfolio" strategy. In this sense I say: I believe he did indeed

[27] Completely analogous to the information theory, which tells an engineer what he can, in principle, squeeze from his communication device!

make several millions from $30000 but I do not believe he will reproduce this success with his “Sharkfolio” in the long term.

Once again: the beauty and usefulness of the Kelly criterion is that it lets you estimate what you can *realistically* expect in the *long* term. Can one beat Kelly in short run?! Oh, yes[28]! But in the short run (or more precisely with a small number of trades or investment periods) it is rather a lottery than a systematic investment. Recall, it makes little sense to speak about the expectation by a single coin toss but if you toss a coin 100 times, the number of heads and tails will be quite close to the expected value. The same holds for trading and investment: if you do have an edge, you will profit from it in the long term. But in the short term the randomness dominates and what you get is more a matter of chance.

In this sense, I highly recommend you to read the book “What I have learnt losing a million dollars” by Jim Paul (former governor of the Chicago Mercantile Exchange). It is a rare book, which also sells a bitter truth rather than an [unrealistic] hope. The leitmotiv of this book is that the author got used to breaking the rules and succeeded for a while. But once he failed … and lost everything. Does it not resemble the cases that we have considered: to lose most of your wealth it suffices to get a severe negative return *just once*, if you overbet. But if you follow the rules, i.e. bet optimally, even a severe negative return will not make you bankrupt.

☝

You are probably wondering, how we can apply Kelly criterion to the stock market. Finally, the stock trading is not coin tossing and there are not just two possible outcomes like heads and tails but much, much more. This is of course true but surprisingly a continuum of outcomes by stock trading can be very well approximated by a simple model á la coin toss.

[28] For the technical details have a look at Browne, S.(2000) Can you do better than Kelly in the short run? (Available at: http://www.sbrisk.com/Browne/beat_kelly.pdf)

In order to release the superiority of the Kelly criterion we need a lot or trades (or bets) and the reinvestment of winnings. The latter is very important and probably makes the main difference between gambling and investment. Indeed, according to the Kelly criterion we invest a constant[29] *fraction* of our capital, e.g. 42% as in the case above. But if we play poker with friends or roulette in casino, we usually put a constant *amount* of money at stake, say, $10. In this case, if we have an edge, the capital grows not exponentially but linearly thus it is plausible to optimize the expected terminal wealth.

☝

I, for one, do not play roulette since it is unfavorable game due to the [double] zero. Nor do I play poker. But sometimes I buy out-of-the-money options[30]. Such options, when carefully selected, may be considered as roulette with a positive expectation. The chance to lose everything is still big but there are also some chances to yield a lot of money. My monthly pocket money budget is about €100 but I usually have no time to spend it ☺. On the other hand I sometimes buy a pretty expensive toy, say, a new accessory for my computer. Thus if I invest my pocket money in options and make on average €40 per trade, I yield 12*€40 = €480 to the end of year, which is enough for e.g. an additional monitor (I have four at home). Within the year there are months when I lose my pocket money. If I made profit in previous months, I can compensate my losses with it. But even if I started the year with losses, it is completely no problem for me to refrain from going to a restaurant or bar during the first couple of months. Important is just to get profit to the end of year. But I never confuse gambling with my pocket money and the investment!

---

[29] If the odds change from trade to trade, the optimal Kelly fraction also changes respectively.

[30] In Germany the options, sold to retail investors, are usually called Optionscheine. But they are essentially options, i.e. the rights but not obligations to buy (call option) or sell (put option) an underlying at a given price (strike). The most of Optionscheine are American, i.e. one can exercise them any time before expiry date. Most of traders do not execute but simply sell them (s. Chapter 6).