Cryptocurrency trends analyzer: platform for trend analysis, portfolio construction and hedging with advanced econometrics

Cromolab

Abstract

Cromolab creates statistical and technological infrastructure necessary for effective investing in cryptocurrencies. Cromolab follows a purely scientific approach combining advanced econometrics methods with up-to-date theory of financial mathematics and computational techniques. The engine behind the platform is based on advanced econometrics methods developed by team of Cromolab. The team consists of world class professors in econometrics, mathematical finance and information technology, supported by PhDs and researchers in econometrics, machine learning, financial mathematics as well as IT. From econometric point of view, the technology that we develop is based on a selection of papers published by memebrs of our team in the last few years. The main components of the methodology stem from (Van Dijk et al. 2013), (Johansen and Gatarek 2017), (Ardia et al. 2016) and (Ardia, Hoogerheide, and Gatarek 2017), but this list is definitely not exhaustive. Due to combinations of up-to-date methods sourced from a selection of disciplines in the contemporary econometrics, the resulting platform constitutes a coherent and complete investment analysis tool for cryptocurrency market. It covers entire chain of investment decision making process: signal processing, portfolio construction, risk analysis. To our understanding the most distinctive feature of this project is its innovative character in terms of combining two spuriously opponent fields of econometrics into one coherent mechanism. By that We mean mixing classical econometrics philosophy on one side, represented by leading world-class econometrician in our team, **Prof. Soren Johansen**, and Bayesian econometric paradigm, represented by another leading worold-class econometrician in our group **Prof. Herman K. van Dijk**. Bringing together classical, large sample based statistical inference, and the computational beauty of Bayesian simulation based approach to econometrics, position this project as one of the most innovative project on the ledger of modern econometric application. It allows for solid platform which covers the plethora of modern analytical techniques to deliver the best possible insight around portfolio analysis in cryptocurrency market. The main components of the platform are: **trend signalling tool**, **portfolio constrution tool**, **market making tool**. The trend signalling tool identifies the current trend in the market (rising or declining trend) for each cryptocurrency, so that the investor can decide on the sign of position, she enters (long or short) in the underlying cryptocurrency. Portfolio construction tool is designed to hedge the abovementioned position in the cryptocurrency with oppsite positions in other cryptocurrencies, statistically related to the undelying to provide the investor with variance minimizing portfolio. Finaly the market making tool is though for big players, which have ambition of setting the price in the illiquid cryptocurrencies, accordingly to their statistical relation to more liquid cryptocurrencies. TO DO: add word technology

Contents

1 Introduction

Bitcoin *puts a question mark on the fractional banking model we know today*- Christine Lagarde, the Managing Director of the International Monetary Fund said in a reamrakbably frank talk at a Bank of England conference. She assumes that the cryptocurrency can displace conventional banking and has unquestioned potential to challenge the monopoly of national monies as payment mean. Apart from direct impact on worldwide economy and settlement systems, it can have a giant influence on financial industry. A huge part of financial institutions, hedge funds, investment banks and high net worth private investors that are currently active in the foreign exchange market might observe a rapid decrease of market activity and thus declining returns.

Nowadays the investment industry makes use of plethora of analytical tools to analyze the foreign exchange markets. However, because of the rapid changes in the market and increasing exposure to the cryptocurrency the standard analytical tools become obsolete. There is a scarcity of tools which provide proper signaling and portfolio construction methods for cryptocurrency market. For the time being they hardly exist. Cromolab sets an objective to develop such technology Cromolab.io. It brings together modern econometric knowhow and melts it with the most modern information technology to create a very flexible cryptocurrency investment platform based solely on analytical methods. The team of Cromolab has ambition to create technology centered around machine intelligence based on most advanced methods of computational statistics.

Cromolab follows a purely scientific approach combining advanced econometrics methods with up-to-date theory of financial mathematics and computational techniques. The team consists of world class professors in the subjects of econometrics, finance and information technology, supported by PhDs and researchers in econometrics, machine learning, financial mathematics as well as IT. From statistical point of view, the technology that we develop is based on a selection of papers published by members of our team in the last few years. The main components of the methodology stem from (Van Dijk et al. 2013), (Johansen and Gatarek 2017), (Ardia et al. 2016) and (Ardia, Hoogerheide, and Gatarek 2017), but this list is definitely not exhaustive. Due to combinations of up-to-date methods sourced from a selection of disciplines in contemporary econometrics, the resulting platform constitutes a coherent and complete investment analysis tool for cryptocurrency market. It covers entire chain of investment decision making process: signal processing, portfolio construction, risk analysis.

This white paper presents the methodological underpinnings of econometric methods behind the technology that Cromolab.io creates. It presents the functionalities of the technology and gives some understanding about services, products, that the technology offers to the investment industry.

2 Technical introduction

It is well know that the recording of price of asset over time forms into a sequence of measurements called time series. The analysis of time series dates back to 1940s and it has remained a well established discipline of econometrics henceforth. Definitely one of the most distinguished papers in this stream is an scientific paper of Prof. Johansen, member of Cromolab.io, (Johansen 1988), which has defined the discipline of economic time series analysis for the next 30 years. This piece of work shows how to test the exact form of dependency across a set of time series associated to a group of economic variable. In a nutshell this philosophy assumes that there exist a common trend, or a group of common trends, which drive the time series. The methodology developed by

Prof. Johansen makes it possible to identify those trends. From statistical point of view those trends form time series themselves. They can be measured upon measurements of underlying economic variables in the analyzed system.

The subdiscipline of time series analysis designed for identification of common trends among many time series is called cointegration analysis. Prof. Johansen is perceived as a father of this stream. His book (Johansen 2006) is a bible of time series econometrics and constitute a must in every master and PhD track in econometrics worldwide.

In the early papers of 1980s and 1990s, cointegration analysis has mostly been applied to the analysis of macroeconomic time series, which describe main indicators of economy in a macroscale. With deeper and deeper understanding of financial markets dynamics, the cointegration analysis has played substantial role in modeling of asset prices and their linkages. That was somehow natural as asset prices are modeled with random walks, which from the theoretical point of view define processes for time series underlying cointegration analysis. Random walk is a simple yet very general process that ensembles the idea of randomness over time. In a nutshell a random walk should be perceived as a process which, starting at some level, can grow or decrease by one unit every period. If the process is symmetric the probability of up and down movement are equal. In case of asymmetric random walk the probability of increase differs from the probability of downward moveement. Both are defined by a proper probability distribution associating a variable $p \in (0, 1)$ with an increase of the process. Then, $1 - p$ automatically stands for the probability of decrease. Typically, the mathematical theory of random walk assumes that those probabilites equal 0.5 for both up- and downmove. In case of application to financial markets those probabilities are not equal. That is an aftermath of alternating trends which occur in the asset price. The probabilities fluctuate, what results in asset price trending immediately. Momentum strategy and trend following is a direct consequence of time varying probabilities of ups/downs which determine a current market cycle over rising and declining trends. The estimation of random walk distribution over time is one of the key components of the statistical engine behind the technology of Cromolab.io. Second building block is the cointegration analysis itself.

Why is the cointegration analysis so important in terms of price series in cryptocurrency market ? First of all it allows for modeling and relating one cryptocurrency price time series to other cryptocurrencies. Secondly, based on methodology recently developed in (Johansen and Gatarek 2017) we can build portfolios of assets with minimal variance. Further, as cointegration assumes a random walk nature of asset price, but does not impose any other conditions, it is definitely applicable to the situation where the probability distribution of random walk evolves over time. To that end it needs to be combined with proper estimation methods. We apply the techniques developed by team of Cromolab.io in the publication (Van Dijk et al. 2013). Combining the detailed analysis of cointegration with accurate time specific probability distribution for random walk up- and down-movements constitutes a complete methodology for optimal portfolio construction and hedging. The estimate of the probability give indication on the market position to enter. Cointegration analysis allows proper portfolio construction. The question arises around the applicability of traditional analysis of financial markets to the new economy of cryptocurrencies. The intuitive answer is rather simple. The nature of price formation might be very different in the cryptocurrency market due to decentralization of the system that stands behind the cryptocurrencies. There is in fact no central bank the drives the supply of money as in case of traditional currencies in foreign exchange. However, based on research performed by Cromolab.io we state that in terms of statistics, nothing differs substantially. The price process can be modeled with random walk and as a consequence cointegration can be applied. (TO DO: reference). Thus, most of the results that have been developed in the financial econometrics so far, can be transplanted into analysis of cryptocurrencies.

The use of cointegration for analyzing financial data is well established over the last 20 years. Regarding the most influential papers in the discipline, the problem of price discovery is discussed by (Hasbrouck 1988), (Lehmann 2002), (Jong and Schotman 2010), and (Grammig and Schlag 2005). Gatev, Goetzmann, and Rouwe (2006) study pairs trading, and continuous time models with a heteroscedastic error process are developed by Duan and Pliska (2004) and Nakajima and Ohashi (2011). Alexander (1999), and more recently Juhl, Kawaller, and Koch (2012), studied optimal hedging using cointegration. (Ardia et al. 2016) have presented how to apply a specific restriction on cointegration model to make it fully applicable to financial applications. Finally (Johansen and Gatarek 2017) have developed a methodology for optimal portfolio construction based on asset prices driven by random walks with portfolio weights depending on the hedging horizon.

To sum up, there are a few key building blocks of the statistical engine behind the platform. First of all we assume that the cryptocurrency rate as any other exchange rate or financial asset price can be modeled with a random walk. Secondly we assume that there is a probability distribution that stands behind this random walk, in terms of up- and downmovement. Then, it is assumed that the evolution of this probability over time can lead to periods of rising trends and downturns. Furthermore, we assume that the historical quotations of the cryptocurrency rate can be explored for estimating on the relation between the price movements and the probability distribution governing the random walk behind. Finally we assume that the entire spectrum of price series which are representing the cryptocurrencies can be modeled by means of cointegration analysis and as it is the case we have the entire spectrum of methods developed by econometrics available for portfolio construction in such a market.

3 Target group

Analytical reports bring in an impressive amount of money all over the world. In 2015 alone, professional traders spent over $50 billion purchasing financial market data, of which $4 billion was spent on professional analytical services and systems based on predictive analytics. By 2020 this figure will increase approximately six times. And these are only professional analytical systems. The B2C financial information market for non-professionals is huge: for example, 54% of US residents have bought shares at least once in their lives, and in China about 30% of residents are engaged in stock trading

This trend is somehow natural as an overwhelming information make it almost impossible to read and analyze everything published on a specific asset. Therefore Cromolab.io bets for simplicity. We offer the technology which combines signaling and portfolio construction in the same platform. Cromolab.io bases inference on statistical analysis solely. We provide platform which in an easy way allows for building the cryptocurrency portfolio based on solid statistical evidence, which is digested by the statistical engine behind the platform.

4 Products

The analytical platfrom is the main product developed by Cromolab.io. However the customers have options to buy insight at different level of detail. In what follows we present the products

Which we are going to offer to the market based on this methodology.

Trend signaling for each cryptocurrency traded Based upon the probability filtering model the customer obtains signal which present current market sentiment in the market: upward or downward trend with information on the exact timing of the start of such a trend and its lenght up to date. Initial analysis show that the trends are definitely alternating in the sense of succession of donwward movements and upward trends ensmble the shape of trigonometric series. Cromolab.io has researched these cycles based on methods developed by Prof. H.K. van Dijk in his publication (Van Dijk, Harvey, and Trimbur 2007) with Prof. Harvey, the key world expert in time series filtering. This methodology has been combined with techniques presented in (Van Dijk and Kleijn 2006) into common methodology. Based on it, the system is able to deliver the expected time of arrival at the equilibrium which is interpreted as an end of the currently observed trend (cycle) and, potentially, beginning of the contrary trend. In that sense the equilibrium is interpreted as a phase when market has no direction or is just after an end of a trend. Such analysis allows for momentum trading based on the identified trend and presents a possibility for trend following.

Hedging portfolio construction Apart from the signaling functionality, which is the key component of the technology, the platform allows for portfolio construction around a selected cryptocurrency. The investor is supposed to select a trending cryptocurrency and express willingness to invest in it. If the selected cryptocurrency follows an upward trend, the investor probably wishes to enter a long position in this cryptocurrency. She can decide to follow an outright position, unhedged by portflio, or, alternatively, she can hedge this position with a proper portfolio around it. To that end the methodology in (Johansen and Gatarek 2017) is applied. The fact that cryptocurrencies can be modeled by means of random walks, which are definitely driven by common trends, and the fact that they are characterized by extensive amount of correlation among each other, opens a wide application potential for methods developed in (Johansen and Gatarek 2017). The methodology assumes that the assets can be modeled with a cointegration model. Given the selected cryptocurrency traded by the investor, the system selects the optimal set of variables to enter the cointegration model. This set is selected among all the other cryptocurrencies analyzed by the platform. the statistical engine estimates thousands of models in the background and selected the statistically most appealing specification. Based on this specification, the methodology developed in (Johansen and Gatarek 2017) is applied to deliver the optimal portfolio weights leading to minimum portfolio variance in an assumed horizon. This horizon is also indicated by the model based on the time series properties of the evolution of the probability which has been discussed above and are estimated based on (Van Dijk, Harvey, and Trimbur 2007) and (Van Dijk and Kleijn 2006). The initial research performed by Cromolab.io indicates that the first derivative of the probability curve indicates highly cyclic behaviour and its dynamics is able to time the trend reversls extremely accurately. Naturally, despite of the automatic horizon length estimate, the investor can define the horizon in an ad hoc fashion and select corresponding portfolio instead of the one inidicated by the cycle model.

Market making TO DO: describe cointegration model for market making

5 Motivation: fluctuations of the random walk

How random is the random walk?

The basic idea underlying the cryptocurrency trend signaling, the main component of the methodology behind the engine, is derived from the theory of random walk fluctuations. This theory has been put forward in the middle of twentieth century and was a very appealing discipline at that time and so it remains today. For introduction to the topic the reader is refferred to (Feller 1957). The results concerning fluctuations in coin tossing show that widely held beliefs about the law of large numbers are fallacious. Coin tossing presents a most basic model of symmetric random walk. Therefore it is often brought as an example, or rather an illustration, of random walk process. The implications of random walk fluctuations are so amazing and so at the variance with common intuition that even sophisticated professors of statistics have doubted that coins actually misbehave to what would be expected from the random walk based upon common intuition.

The theory has been developed addressing the coin tossing but the results are representative of a fairly general situation that can be described in form of a random path in two dimensional space of time and value (of cryptocurrency price in tis case). In this paper we present the part of the theory of random walk fluctuations which is necessary for understanding the main idea of methodolgy we develop. We are motivated principally by the unexpected discovery that this theory can be treated by elemantary method of mathematics, in particular basic combinatorics and fairly basic statistical theory.

We start with defining the rules of typical coin-tossing game and its relation to the simple random walk. Lets imagine there is some initial fortune equal to zero at the beginning of the game. Every period (it can be a second, day, month etc.) we toss a fair coin, each time either collecting 1 unit of value (can be 1 usd for instance) from our opponent or paying it out, depending if the random walk in a given period increases or decreases by one unit. The line representing the evolution of wealth generated by the player in such a game is a perfect example of random walk evolution over the time. In Figure Figure 1 we present a simple example of random walk over time. In this case the random walk evolves over 12 period which might be interpreted as 12 games played between two players as skizzed above.

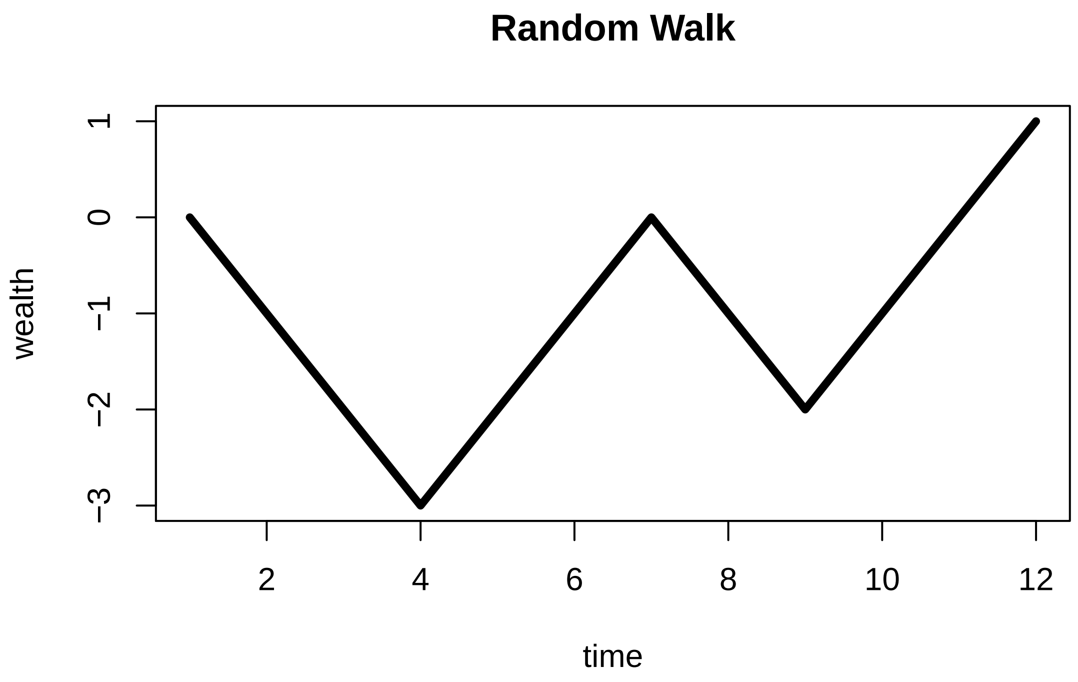


Figure 1: Example of simple random walk.

As the number of games played increases, by the Law of Large Numbers, the fraction of games with positive outcome approaches 1/2 (and negative as well). In the course of the game, however, fortune of a player is likely to fluctuate, changing sign from positive to negative and back. We pose a question: after a large number of n games have been played, what is the fraction of time when the total wealth remained in the winning zone? In what follows we try to answer this question.

By the obvious symmetry, in the time between the consequitive draws, when the total wealth becomes zero, the fortune is equally likely to stay in the positive and negative area. As more games are played, more draws will occur, which seem to imply that the fraction of time spent in positive area goes to 1/2. This would be standard intuition put forward by everyman asked the above mentioned question. In fact this intuition is completely wrong! The fraction of time spent in positive area is more likely to be close to extremities 0 or 1 than to be close to 1/2.

To understand this phenomena we need to take a closer look at the nature of chance fluctuations in random walks. The results are startling. As mentioned above, according to the widespread belief a so-called law of large numbers should ensure that in a long coin-tossing game each player will be on the winning and losing side for about half the time, and that the lead will pass not infrequently from one player to another one. Lets imagine then a huge sample of records of ideal coin-tossing games, each consisting of $2n$ trials. (It is $2n$ rather than n as the draw can only happen in even periods.) We pick one such period at random and identify the period of last draw, where the number of accumulated heads and tails were equal. This number is even, and we denote it by $2k$ (so that $0 < k < n$). Frequent changes of the lead would imply that k is likely to be relaitvely close to n, but in practice this does not need to be so. In fact, the distribution of k is symmetric (any value of k has exactly the same probability as the value of $n - k$). Furthermore, the probabilities near the end points, so near $2n$ are greatest. The most possible values for k are the extremes of 0 and n. The intuition lead to an erroneous picture of probable effects of coin fluctuations. Figure Figure 2 presents profile of such probability for each k, where $n = 4$, thus $2n = 8$.

The understand the relation of point where random walk crosses a zero line, period k and its relation to the total length of the path, we need to take a closer look at the nature of the problem. Let us assume that we work with the simplest possible definition of a random walk, which is defined as a process which starts at period 0 with value 0 and spans over $2n$ periods, and each period it can increase or decrease in value by 1 unit with probability 1/2 both. From a formal point of view we shall be concerned with arrangment of $2n$ plus ones and minus ones. If we refer to $2n$ as the length of the random walk path, then there are 2^{2n} different possible paths which can realize with this length of the process. Why? Because each period, called alternatively as an epoch, $k = 1, \ldots, n$, the process have two options to evolve: either $+1$ or -1. Then in two consequtive periods it has $2x2$ options to evolve, and in $2n$ periods, it can evolve in $\underbrace{2 \times 2 \times 2 \ldots 2}_{\text{2n}}$ different ways. In Figure **??** we present a lattice which represents all the possible paths for the random walk over 6 periods with 3 upward and 3 downward movements.

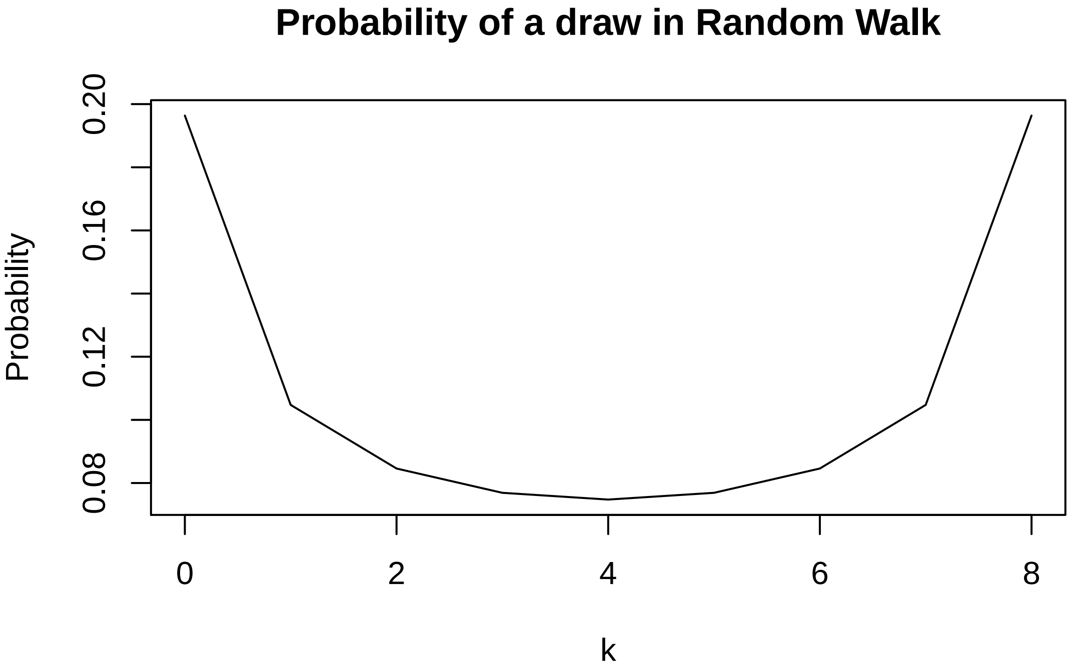
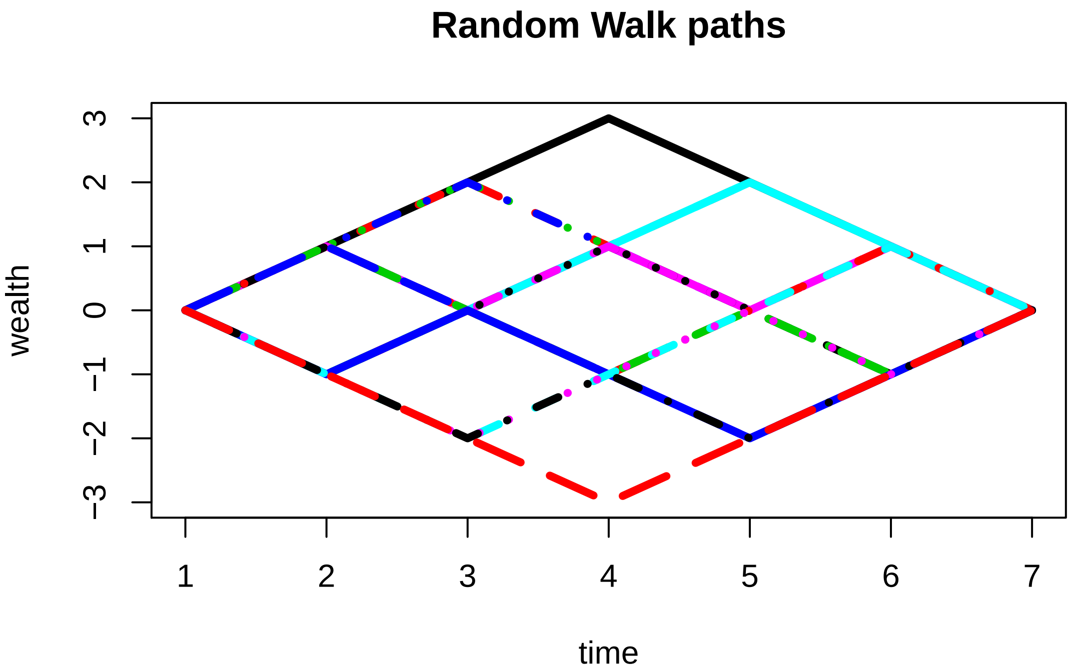


Figure 2: Probability of a draw in a simple random walk with number of periods 2n=8. A draw is represent by a period of time when random walk crosses a zero line.



An immediate observation which arises is the fact that, irrespective of the trajectory the random walk traverses, the initial wealth and the terminal wealth are equal. This is a consequence of the fact that the number of $+1$ and -1 on each path are equal. If the process rises in 3 out of 6 periods and decreases in remaiing 3 periods, then it will always go back to the initial value irrespective of the order of ups and downs.

If we loosen the restriction of equal number of ups and downs and start to geeralize the process, the initial and terminal wealth would not be equal. For instance lets imagine that we assume that among the $2n$ epochs we observe p plus ones and q minus ones. Then, naturally $2n = p + q$ and the value at the period $2n$, so the terminal value of all such paths equals $p \times 1 + q \times (-1) = p - q$.

In Figure Figure 3 the situation with $p = 4$ and $q = 6$ is presented and in Figure Figure 4 a reverse order of p and q is applied. In Figure Figure 3 all paths end in terminal wealth of $p-q = 4-6 = -2$, whereas in Figure Figure 4 they end in $p - q = 6 - 4 = 2$. We observe that the dominance of number of ups over the number of down movements automatically ends in a terminal value in the positive territory, and vice versa.

If we denote $p - q$ with x, then a path from origin to an arbitrary point $(2n, x)$ exists if and only if $2n$ and x remain in the abovementioned relation to p and q i.e.

$$2n = p + q$$
$$x = p - q.$$

Thus each point $(2n, x)$ in the twodimensional space automatically implies the corresponding values of p and q. For instance, let us imagine we are interested in the random walk paths that connect the origin $(0, 0)$ with the point $(6, 2)$. Then, according to the formulas we need to solve the system

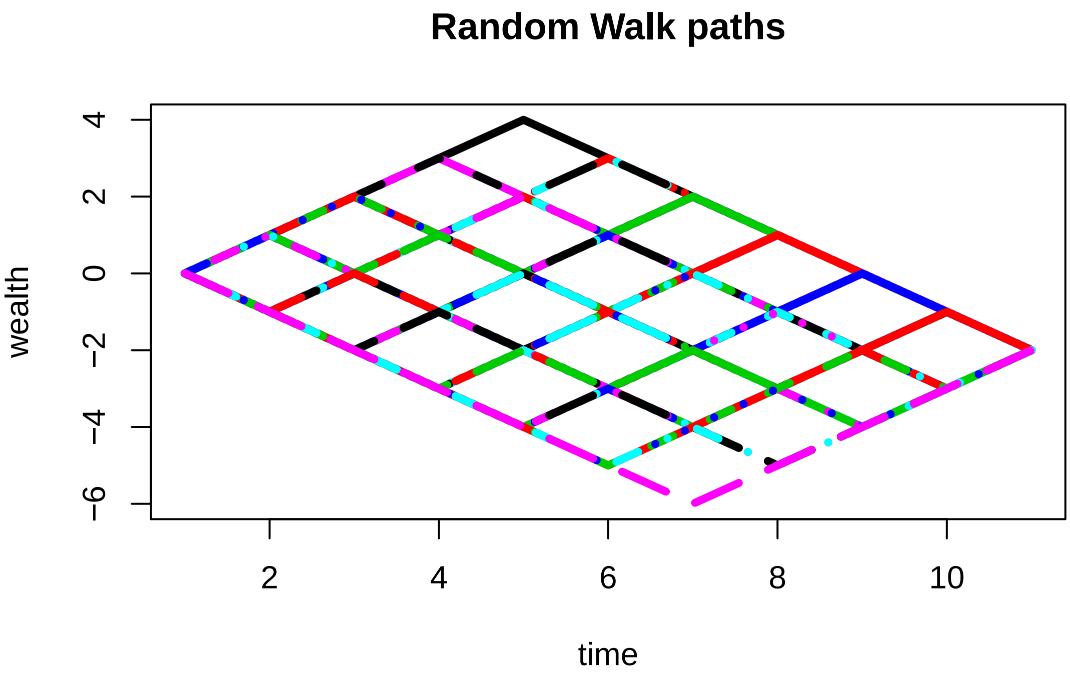


Figure 3: All possible paths for a random walk with 2n = 10 with p=4 and q=6

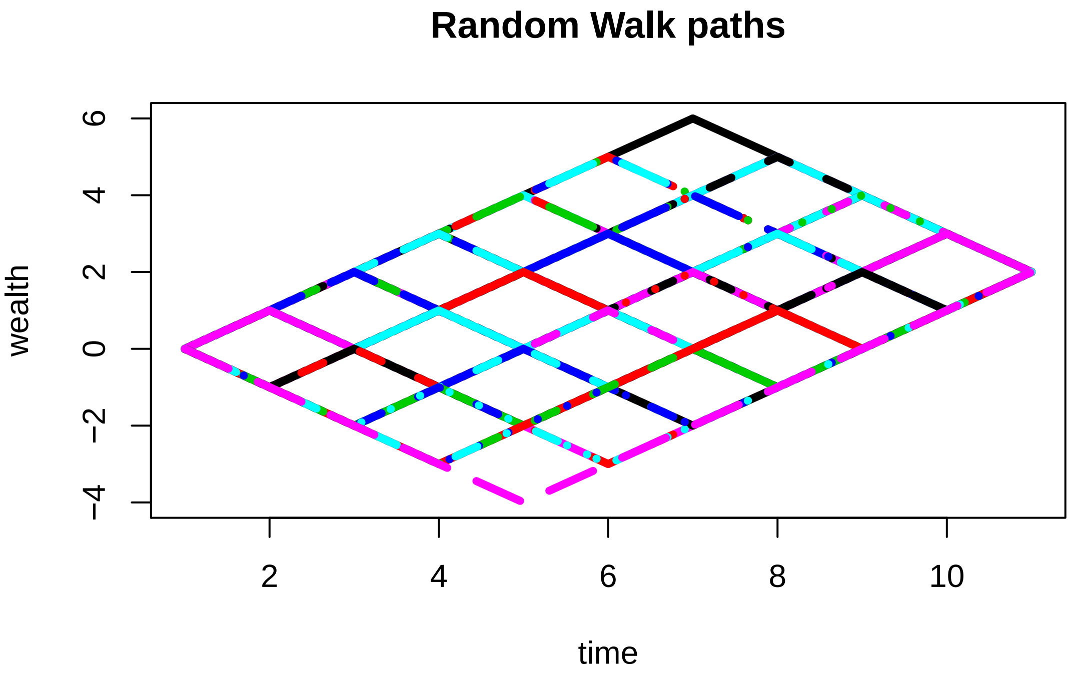


Figure 4: All possible paths for a random walk with 2n = 10 with p=6 and q=4

of simple equations

$$6 = p + q$$
$$2 = p - q$$

which is satisfied by $p = 4$ and $q = 2$. All the paths which lead the process to the value of 2 in 6 steps result from 4 plus ones and 2 minus ones. How many paths of that sort can be constructed? The answer to that question is fairly simple and stems from the basic combinatorial mathematics: that is a number of ways we can distribute 4 balls with a ticker plus ones in 6 bins. Thus the answer is: $\binom{6}{4}$ which can be generalized to

$$\binom{2n}{p.} \tag{1}$$

Importantly

$$\binom{2n}{p} = \binom{2n}{q}.$$

In this chapter we are mostly devoted to the paths which revert to a zero at some point of time, so where a draw occurs at some point on the random walk path. A piece of random walk path between zeroes (draws) is called an excursion from zero. Let us denote by u_{2n} the probability of a return to zero at period $2n$

$$u_{2n} = P(S_{2n} = 0),$$

where S denotes a path of random walk (S stems from the fact that we define the random walk as a sum of a sequence of plus and minus ones). Given the probability we have derived in Equation 1 we can compute the u_{2n} as

$$u_{2n} = \binom{2n}{n} 2^{-2n}. \tag{2}$$

The first term $\binom{2n}{n}$ is motivated by the fact that n out of $2n$ plus ones must occur on the path so that it reverts back to 0 in $2n$ periods. That means that $p = n$ in Equation 1. The term 2^{-2} is nothing else as a total number of all the possible paths between 0 and $2n$, what has been explained above. The important theorem that we base upon is well known from the theory of random walk:

The probability that up to and including period $2n$ the random walk visits the positive area for $2k$ periods ($0 <= 2k <= 2n$) and the negative one for $2n - 2k$ periods is given by

$$p_{2k,2n} = u_{2k} u_{2n-2k}. \tag{3}$$

For a detailed proof we refer for instance to (Feller 1957).

According to the definition of u_{2n} in Equation 2

$$p_{2k,2n} = \binom{2k}{k} 2^{-2k} \binom{2(n-k)}{n-k} 2^{-2(n-k)} = \binom{2k}{k} \binom{2(n-k)}{n-k} 2^{-2n}. \tag{4}$$

It is easy to see that the numbers obtained for different k add to unity. Therefore they constitute a

Table 1: Discrete Arc Sine Distribution of order 8

k=0	k=1	k=2	k=3	k=4
k=8	k=7	k=6	k=5	
0.1964	0.1047	0.0846	0.0769	0.0748

probability distrobution. Such a distribution that attaches a weight $p_{2k,2n}$ to the point $2k$ is called the discrete arc sine distribution of order n. An example of such a distribution has been presented in Table Table 1 for $n = 8$, and has previously been drawn in Figure Figure 2.

As we can see the central term is always the smallest. That reflects the central idea of this theorem, that the probability that both sides of the axis, which are visited by the random walk would be visited with exactly the same frequency is a completely wrong intuition. In fact the opposite is true: such a scenario has the lowest probability. Table Table 1 presents the probability for all possible k's from 0 to $n = 8$. The value associated with a given k represents the probability that the random walk visits the positive area (area above the zero axis) for $\frac{2k}{2n} = \frac{k}{n}$ fraction of periods. For the analyzed case of $n = 8$, the probability that $k = 0$ periods are visited on the positive side is equal to 0.1964. The probability is symmetric in the sense that $p_{2k,2n} = p_{2n-2k,2n}$. That means for instance that the probability that 2 out of 16 periods are spent in the positive area is equal to the probability that 14 out of 16 periods are spent there. Such a probability equals 0.1047 according to the Table Table 1. The fact that the intuition fails completely in case of random walk (coin tossing procedure) is striking. Looking at those probabilities we can observe that in terms of $n = 8$, so 16 periods the probability of not being at all, or being all the time in positve area is $0.1964/0.0784 = 2.6256$ times higher than observing a situation where the same amount of time is spent in both positive and negative areas. Plot Figure 5 presents such ratios for different values of n. Each line in the graph corresponds to one value of n, running from 2 (green line) to 40 (red line). We can see that the extreme values of $k/n = 0$ and $k/n = 1$ are so much more probable to realize than the intuitive scenario of both negative and positive territory being visited with equal frequency of $k/n = 0.5$.

To present the full spectrum of results associated with the arc sine distribution we need to express the binomial coefficient in terms of factorials, according to Stirling's formula. It can be shown that

$$u_{2n} = \frac{1}{\sqrt{pn}}$$

. Applying that formula to Equation 4 we obtain a result

$$p_{2k,2n} = \frac{1}{\pi\sqrt{(k)}\sqrt{(n-k)}}. \tag{5}$$

The latter one is in fact equivalent to

$$p_{2k,2n} = \frac{1}{n\pi\sqrt{(k/n)}\sqrt{((n-k)/n)}}. \tag{6}$$

If we denote $x_k = \frac{k}{n}$ we can evaluate the last equation as

$$p_{2k,2n} = \frac{1}{n}f(x_k), \tag{7}$$

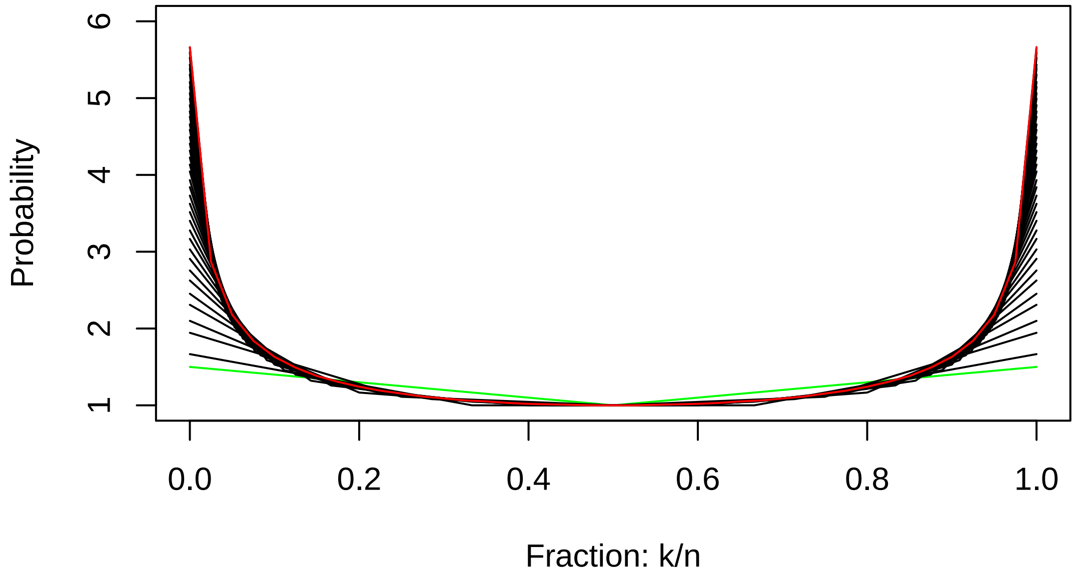


Figure 5: Probability of a draw for different n

with $f(x) = \frac{1}{\pi\sqrt{(x(1-x))}}$. This function gives a shape shown in Figure Figure 6 with its height varying depending of n. This function is referred to as arc sine probability density function. It integrates to 1.

Now, let us imagine that we need to compute the probability that fraction k/n spends in the positive area for a period which is higher than 1/2 and lower than some fraction α, i.e. $1/2 < \alpha < 1$. According to our derivation this probability is given by

$$\sum_{n/2<k<\alpha n} p_{2k,2n} = \frac{1}{\pi n} \sum_{n/2<k<\alpha n} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{1/2}}, \tag{8}$$

which can be approximated by a Riemannian integral

$$\pi^{-1} \int_{1/2}^{\alpha} \frac{dx}{\left(x(1-x)\right)^{1/2}} = 2\pi^{-1} \arcsin \alpha^{1/2} - 1/2. \tag{9}$$

And because of the fact that

$$\pi^{-1} \int_{0}^{1/2} \frac{dx}{\left(x(1-x)\right)^{1/2}} = 1/2 \tag{10}$$

we obtain

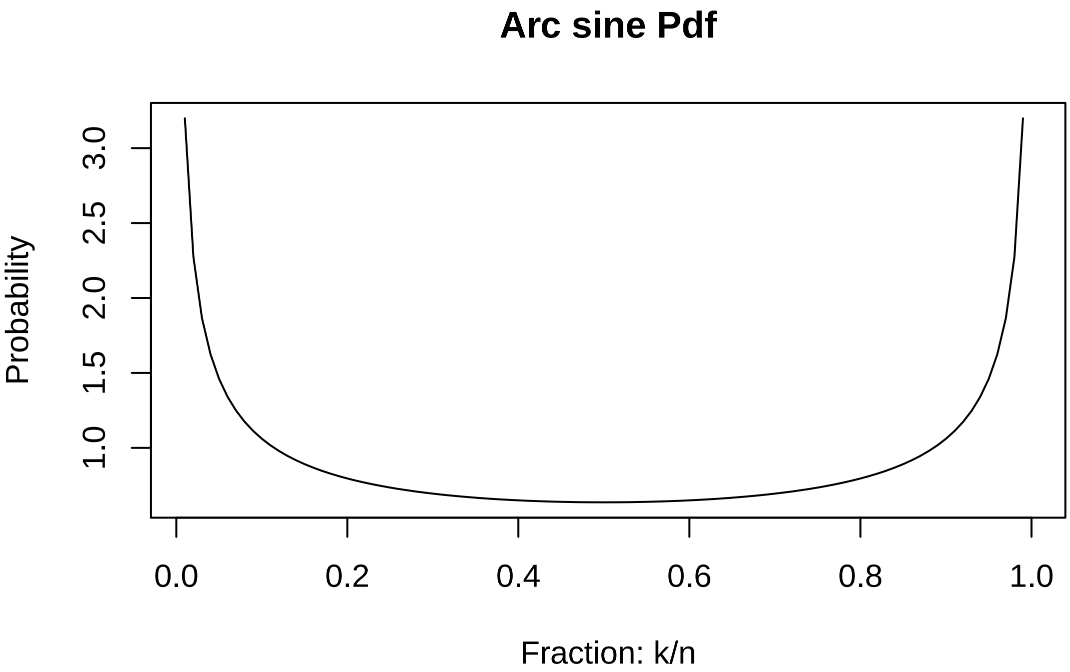


Figure 6: Limiting probability density function of a draw for different fractions k/n

$$\pi^{-1} \int_0^{\alpha} \frac{dx}{\left(x(1-x)\right)^{1/2}} = 2\pi^{-1} \arcsin \alpha^{1/2}, \tag{11}$$

which is a cumulative distribution function of the arc sine distribution. The shape of this function is presented below. It is an immediate result from the shape of the fuction that the fraction of time spent in the positive area is much more likely to be close to zero or one, than to be close to the expected value of 1/2. Figure Figure 7

The cumulative probability distribution function (cdf) is very useful when one needs to read the probability of a less fortunate player to win. Let us come back to the previous example and imagine the random walk identifies with two players who play a game. If a player A is currently winning the random walk visits the positive area. If she is losing, the negative area is visited. In that sense the long period of winning by a particular player is equivalent with a fluctuation of a random walk. Then from the cdf we can compute the probability that corresponds to the situation that a less fortunate player is going to win for a certain proportion of time. That proportion can be denoted by X, where $(X < 1/2)$. Then

$$P_X = 2\pi^{-1} \arcsin X^{1/2}, \tag{12}$$

corresponds to the probability that the fraction of time that the less fortunate player has been in a winning position was smaller than X. As we work with a symmetic problem, there is always some winning player, (and each of two players can be less fortunate), so in fact we need to multiply P_X by a factor of 2 to account for this symmetry. For instance if we take a fraction of 10%, then the

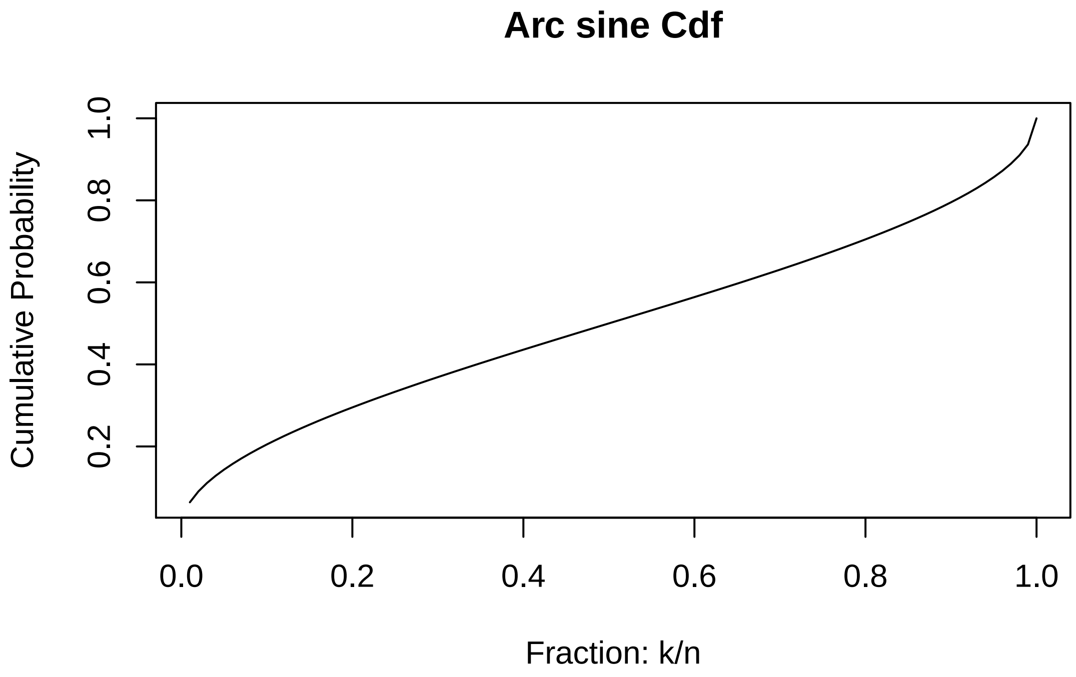


Figure 7: Limiting cumulative probability distribution of a draw for different fractions k/n

corresponding cumulative probability is given by $2 \times 2\pi^{-1} \arcsin 0.1^{1/2} = 0.4$. This means that we can state that there is 40% probability that the less fortunate player would win in maximally 10% of cases. What is more shocking corresponds to extreme levels of for instance $X = 0.01$. Then $2 \times 2\pi^{-1} \arcsin 0.01^{1/2} = 0.13$. That value implies the 13% probability that the less fortunate player would be winning not more than 1% of cases. This results are striking! although the pdf and cdf that we work with are based on asymptotic approximations, they work fine even for so low numbers as $n = 5$.

In what follos we present a short simulation study that confirms the result of this theoretical result. We simulate 1000 paths of random walks of length 100, i.e. for each out of 100 periods we sample either plus one or minus one, both with probability 1/2. We repeat the process for 1000 trajectories of the random walk. Then for each trajectory we count a number of periods that the random walk has been observed in the positive trajectory. Those counts are recorded in a vector. In figure Figure 8 we plot the cumulative distribution function of this vector, the blue line in the plot. The theoretical counterparty based on analytical form of the respective arc sine cdf is plotted with a red line. We observe that the approximation is almost ideal. In the next step we reduce the length of each random walk path to 10 to confirm, if the approximation with the limiting cdf is still valid. In Figure Figure 9 we present the outcome of this experiment. We observe that despite of very limited data input, the law is still holding.

Now imagine we start perturbing the random walk probability distribution, in the sense that instead of the equal probability of occurrence of plus and minus ones set at 0.5, we consider 10 additional scenarios in each of them increasing the probability of plus one by 0.01. Thus the probabilities of plus one in the alternative scenarios vary from 0.51 to 0.6. For each of those scenarios we repeat

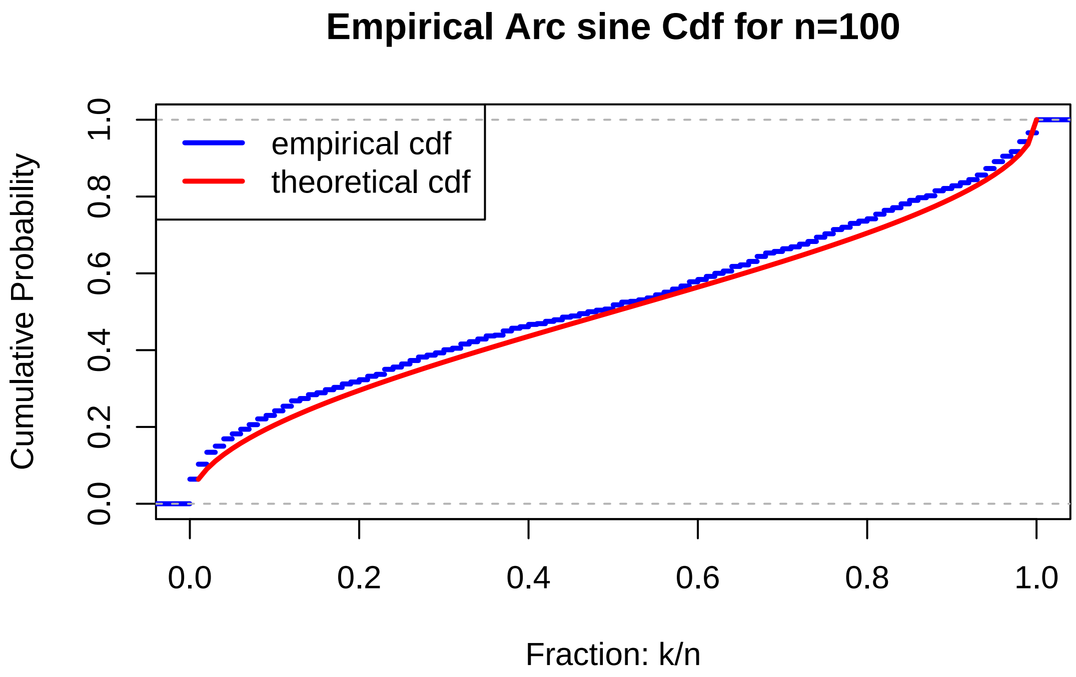


Figure 8: Empirically estimated cumulative probability distribution of a draw for different fractions k/n

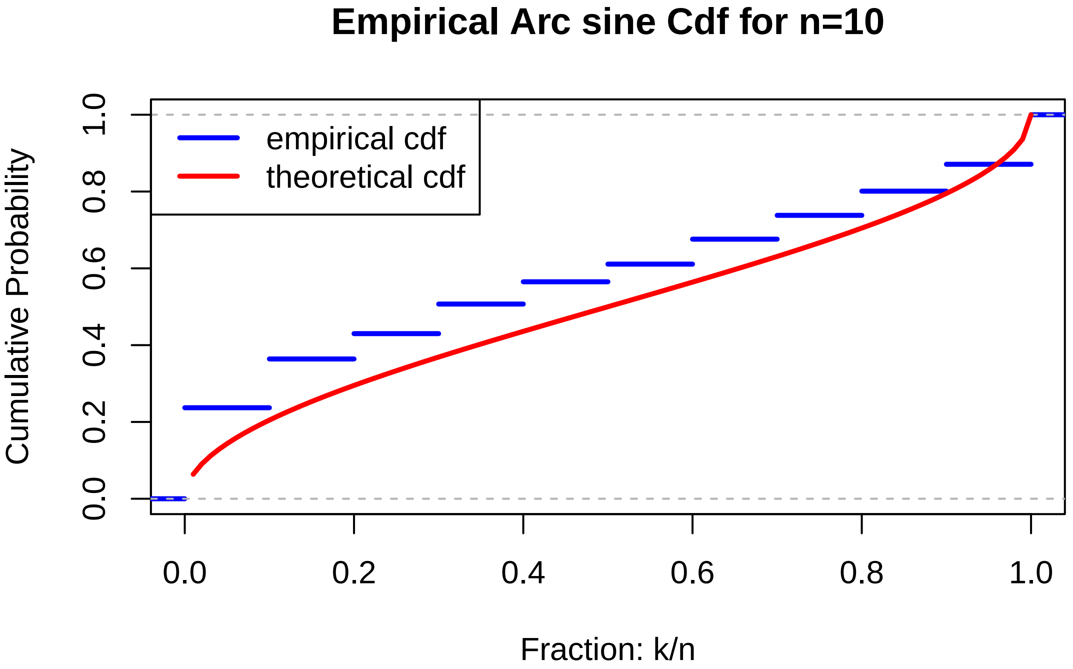


Figure 9: Empirically estimated cumulative probability distribution of a draw for different fractions k/n

exactly the same experiment. We simulate 1000 of random walks of length 100 with a selected level of probability of plus one. Figure Figure 10 presents outcome of this experiment. This is the key plot in this white paper. It shows that perturbing the random walk probability by even a few percent, leads to exteme fluctuations in the random walk. The length of the fluctuations is even higher than in case of random walk with equal probability of up- and downmovement. The most bottom curve in Figure Figure 10 corresponds to the experiment with dedicated to random walks with probability of up-movement equal 0.6. We can observe that in this case there is so much probability cumulated around the fractions $k/n > 0.8$, that the probability of random walk to traverse in the negative side is negligible. If we observe a process that is characterized by probability of increase at the levels which are higher than 0.5, betting for such process to be in positive area is almost a sure deal. If we think of cryptocurrency price modeled by random walk, this is exactly the same situation. If we observe that the probability of increase for the process modeling a given cryptocurrency has dominating probability of incrase, we should go long in this cryptocurrency. If, on the other hand, the probability of decrease is dominating, we should short this cryptocurrency. But… who knows this probability?

```
# length of random walk
T <- 100
# number of simulations of random walk
nSim<-1000

randomWalkRecordings <- matrix(NA, nSim, T)
# number of scenarios for probability of plus and minus ones
J <- 11
# beingInPositive is avariable indicating percentage of time that the random walk goes into th
beingInPositive <- matrix(NA, nSim,J)
for (j in 1:J){
  prob <- 0.50 + 1*(j-1)/100
  for (i in 1:nSim){
    eps <- rbinom(T, 1, prob)
    eps[eps == 0] <- -1
    randomWalkRecordings[i, ] <- cumsum(eps)
    beingInPositive[i,j] <- sum(randomWalkRecordings[i,] > 0) / T
  }
}
plot(ecdf(beingInPositive[,1]), do.points=F, lty=1, col = 'blue', xlim = c(0,1), lwd = 2.5,xlab
for (j in 2:J){
  lines(ecdf(beingInPositive[,j]), do.points=F, lwd = 1.5)
}
p <- (1:100) /100
cdf <- 2/pi * asin(sqrt(p))
lines(p, cdf, col = "red", type = "l", lwd = 2.5)
legend('topleft', c('empirical cdf for p/n=q/n'=1/2,'theoretical cdf', 'cdfs based on perturbed
```

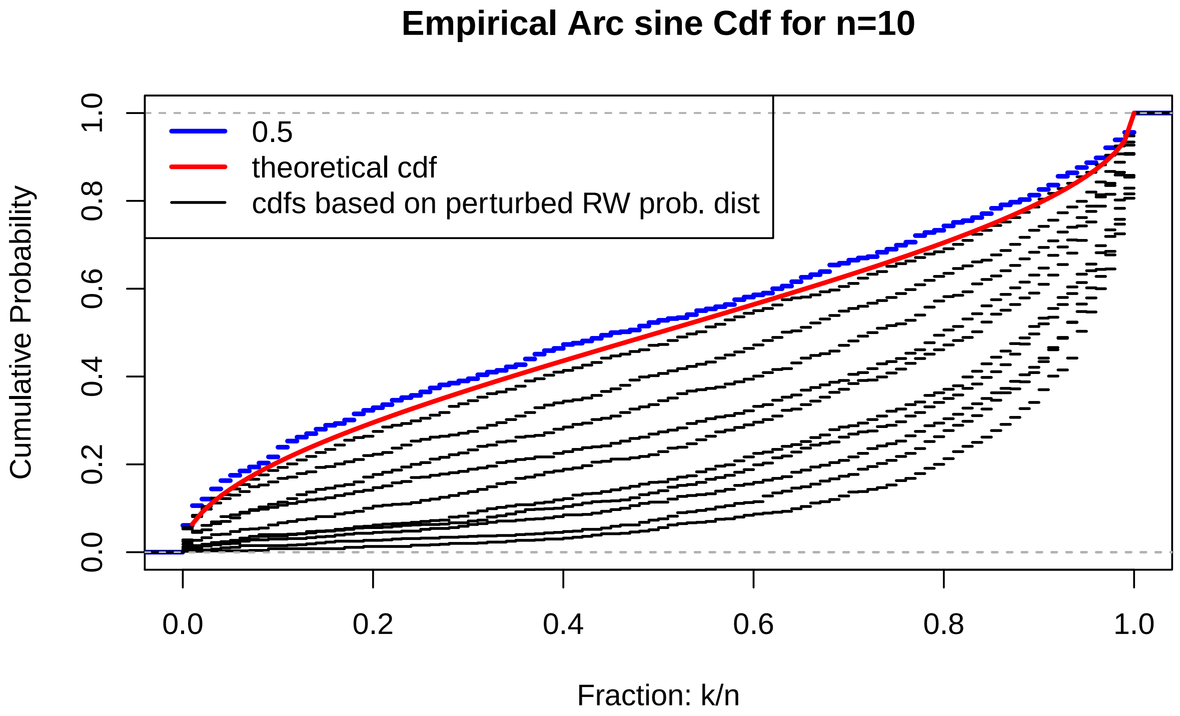


Figure 10: Empirically estimated cumulative probability distribution of a draw for different fractions k/n

6 Cryptocurrency time series model

(Johansen and Gatarek 2017) allows for heteroscedastic time series in the methodology they present. Heteroscedasticity describes the process characterized by sub-populations that have different variabilities from others. By variability we mean the variance of the time series. Although cryptocurrencies can be considered to be relatively new, there has already been some initial analysis into the crypto–currency price data generating process. (Hencic and Gourieroux 2014) applied a non–causal autoregressive model to detect the presence of bubbles in the Bitcoin/USD exchange rate. (Sapuric and Kokkinaki 2014) measures volatility of Bitcoin exchange rate against six major currencies. (Chu, Nadarajah, and Chan 2015) provide a statistical analysis of the log–returns of the exchange rate of Bitcoin versus the USD. They found that the Generalized Hyperbolic distribution seems to be the most appropriate choice to model the unconditional distribution of crypto–currencies time–series. Finally, (Grassi and Catania 2017) find that the . We found that a robust filter for the volatility of crypto–currencies time–series is strongly required by the data. Moreover, we find evidence of long memory in the volatility for some series, while for others a simpler specification is enough. We also find that, differently from foreign exchange currencies, leverage effect has a substantial contribution in the volatility dynamic. On average, volatility increases more after negative shock than after positive shock as in the equity market, hence, crypto–currencies time–series incorporate the so–called leverage effect with some degree of heterogeneity across the series. We find evidence of time–varying skewness for some series and absence of time–varying kurtosis for the whole sample. Stefan Grassi, who hs published on statistical properties of cryptocurrency, is an advisor of Cromolab.io

7 Filtering of time series: estimation of probability evolution

The models applied in case of cryptocurrency modeling need to tackle an important aspect in the sense that inference must be made online, before the data collection ends. For those types of applications one must have, at any time, an up-to-date estimate of the current state of the system. By the state of the system, the current value of parameters is considered. In that case we estimate the probability of an price move from the currency price of the assets and as such we need to have the best possible current estimate of the probability. The density of the current estimate of the parameters is usually called the filtering density. Let us denote it by $p(\theta_t|y_{1:t})$. It denotes the best knowledge about the parameter θ_t that we have based on the information in the data up to the period t.

7.1 Bayesian approach to econometrics and its parallel to probability filtering

Time series filtering has a lot to do with discipline of Bayesian econometrics. In the analysis of real data, nonnecessary economic or financial ones, we rarely have perfect information on the phenomenon of interest. Even when an accurate deterministic model for the system under study is available, there is always some measurements error, or imperfections. We deal with uncertainty. A basic point in Bayesian econometrics is that all the uncertainty that we might have on a phenomenon should be described by means of probability. In this perspective, probability has a subjective interpretation, being a way of formalizing incomplete information that the researcher has about the events of interest. Probability theory presecribes how to assign probabilities coherently. How about the data based inference on probability?

The Bayesian approach postulates learning from experience. The learning process consists of the application of probability rules: one simply has to compute the conditional probability of the event of interest, given the experimental information. Bayes' theorem is the basic rule to be applied to this aim. Given two events A and B, the joint probability of $A \ \ \& \ \ B$ to occur is given by $P(A \ \ \& \ \ B) = P(A|B)P(B) = P(B|A)P(A)$, where $P(A|B)$ is the conditional probability of A given B and $P(B)$ is the refferred to as the marginal probability of B. Bayes' theorem is a direct consequence of the above equalites and says that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{13}$$

The importance of this formula results from its implications for inductive learning process. In the world of Bayesian econometrics, A represents the event of interest for the analyst, B an experimental result which she believes can provide information about A. Given $P(A)$ and having assigned the conditional probabilities $P(B|A)$ of the experimental fact B conditionally on A, the problem of learning about A from the experimental evidence B is solved by computing the conditional probability $P(A|B)$ according to the Bayes' theorem.

The event of interest A in the statistical inference is usually represented by the vector of parameters θ and the experimental result is usually described by the sample of observed data Y. More specifically, based on the knowledge of the problem, the researcher can assign a conditional distribution $p(y|\theta)$ for Y given θ, called the likelihood. $p(\theta)$ expresses the uncertainty on the parameter θ. $p(\theta)$ is usually referred to as a prior distribution. Upon observing $Y = y$, we can apply Bayes' formula to compute the conditional density of θ given y. This density is referred to as posterior distribution

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{14}$$

where $p(y)$ is called marginal distribution of Y,

$$p(y) = \int p(y|\theta)p(\theta)d\theta. \tag{15}$$

The marginal dstribution is only a normalizing factor for $p(y|\theta)p(\theta)$, thus we can use the proprtionality sign instead of equality relation

$$p(\theta|y) \propto \frac{p(y|\theta)p(\theta)}{p(y)}. \tag{16}$$

Equation Equation 16 is the key equation of Bayesian inference and filtering at the same time.

What it basically says is that the posterior distribution is proportional to the product of likelihood and prior. It presents the underpinnings of the Bayesian paradigm which says that the posterior distribution is based upon the prior distribution and the information coming from the observed dataset (likelihood).

This concept is envisaged in Figure Figure 11 with example of beta distribution. The prior distribution which expresses the prior uncertainty of the researcher with respect to the true value of probability parameter p (x-axis) is enriched by the information from the dataset (also in form of beta density) to result in a posterior distribution in the same family of distributions.

```
# parameters of prior distribution
a <- 3
b <- 7
# parameters of the likelihood
s <- 12
f <- 18
curve(dbeta(x, a + s, b + f), from = 0, to = 1, xlab = "p", ylab = "Density",
    lty = 1, lwd = 4, main = "Prior updating and beta conjugacy")
curve(dbeta(x, s + 1, f + 1), from = 0, to = 1, xlab = "p", ylab = "Density",
    lty = 2, lwd = 4, add = TRUE)
curve(dbeta(x, a, b), from = 0, to = 1, xlab = "p", ylab = "Density", lty = 3,
    lwd = 4, add = TRUE)
legend("topright", c("Prior", "Likelihood", "Posterior"), lty = c(3, 2, 1),
    lwd = c(3, 3, 3))
```

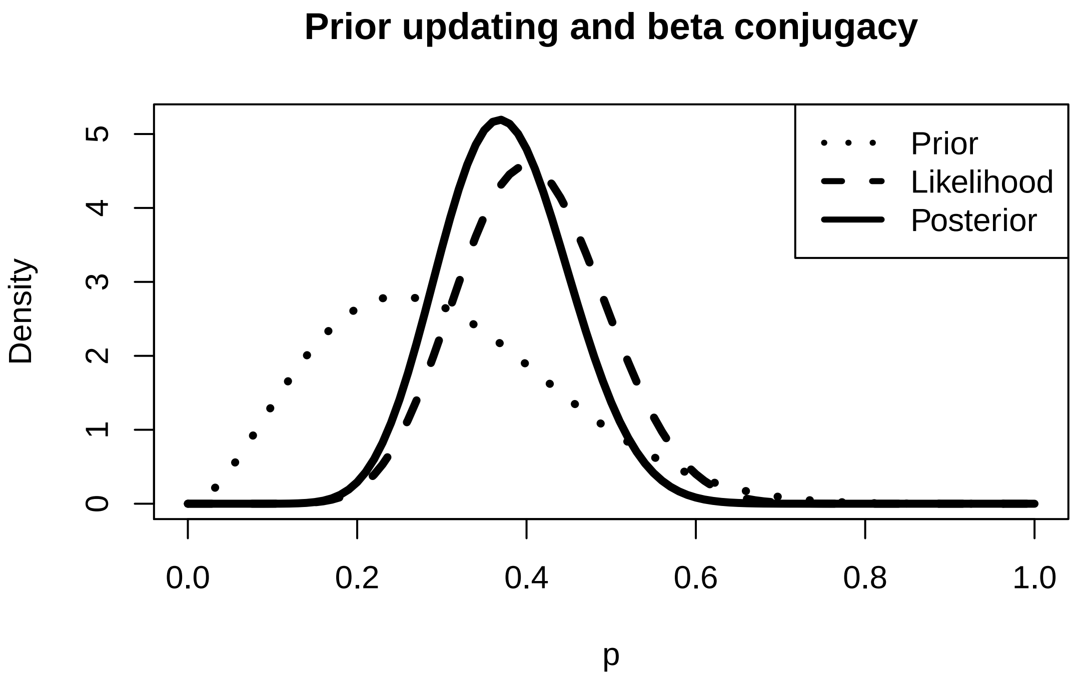


Figure 11: Bayesian updating

Figure 11 presents the concept of updating the prior with data information. We see how the density moves from prior to posterior based on the likelihood data in between. This concerns the fixed data sample. However in case of real time analysis of time series, filtering methods, this picture repeats every period. The estimates of the parameter in the previous period becomes a prior distribution for the estimate of the parameter in the current period. With the data observed in the current period, the estimate is update what shifts the density of the event of interest accordingly to what has been recorded with the new observation. We discuss it shortly based on formulas of filtering density mentioned above.

The density of $p(\theta_t|y_{1:t})$ can derived based on definition of the conditional probability, again, as

$$p(\theta_t|y_t, y_{1:t-1}) = \frac{p(\theta_t, y_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \tag{17}$$

At the same time by reapplicability of the formula for conditional probability we obtain

$$p(\theta_t, y_t|y_{1:t-1}) = p(y_t|\theta_t, y_{1:t-1})p(\theta_t|y_{1:t-1}) = p(y_t|\theta_t)p(\theta_t|y_{1:t-1}), \tag{18}$$

in which we recognize the likelihood $p(y_t|\theta_t)$, associated with currently recorded observation y_t, and the prior density $p(\theta_t|y_{1:t-1})$ which describes the prior knowledge about the state of the parameter θ_t before the observation y_t is recorded.

By joining the equations together we can obtain exact form of the filtering density

$$p(\theta_t|y_t, y_{1:t-1}) = \frac{p(y_t|\theta_t)p(\theta_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \propto p(y_t|\theta_t)p(\theta_t|y_{1:t-1}) \tag{19}$$

The density $p(\theta_t|y_{1:t-1})$ can be described as a one-step-ahead predictive density for the states. Given that we know where the system was in the previous period, so that we know the density $p(\theta_{t-1}|y_{1:t-1})$, and given that we know the form of dependence of state of the parameter today as a function of the state of the parameter in the previous period $p(\theta_t|\theta_{t-1})$, we can derive

$$\begin{aligned} p(\theta_t|y_{1:t-1}) = \int p(\theta_t, \theta_{t-1}|y_{1:t-1})d\theta_{t-1} &= \int p(\theta_t|\theta_{t-1}, y_{1:t-1})p(\theta_{t-1}|y_{1:t-1})d\theta_{t-1} \\ &= \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})d\theta_{t-1} \end{aligned}$$

Thus $p(\theta_t|y_{1:t-1})$, which is interpreted as a prior gives some vision on where potentially the system can go in the current period, based on where it was in the previous epoch. This vision refers to the situation before the new observation enters into the environemnt. The measurement of data bringing this new observation revises this vision and corrects it accordingly to where the data direct the parameters of the system.

The techniques of time series filtering constitute basis for filtering of probability distribution of random walk which models the price of the cryptocurrencies. We apply the method of particle filtering to implement this idea in a fully computationally manner. The project assumes applying this technique at all possible level of data granularity, starting from daily prices and going deeper and deeper to obtain full understanding of the probability distribution behind the ups and downs of the system under study.

7.2 Particle filtering

In many simple applications with Normally distributed variables the usual way of filtering is by means of Kalman filter, which is provided based on closed form analytical expressions for the filtering distribution etc. In case of the technology developed by Cromolab.io this approach is not satisfactory due to highly nonnormal data and, most of all, data which have discrete character and needs to be modeled by means of discrete distributions (Binomial, Poisson and Negative-Binomial). The probability distribution of random walk modeling the data is filtered under assumption of Negative-Binomial distribution for the process of forming up- and downmovement in the cryptocurrency price process.

The Cromolab.io is based on particle approach to filtering. Particle filtering is how sequential Monte Carlo is usually referred to in applications to state space model. It is easiest to understand when viewed as an extension of importance sampling. For this reason this subsection is started with a short discussion of importance sampling. The main difficulty of filtering is to evaluate the integrals entering the formulas in the filter as for instance in Equation 19.

In general we suppose that we are interested in evaluating the expected value of some function $f(X)$

$$E_\pi(f(X)) = \int f(x)\pi(x)dx. \tag{20}$$

Due to difficulty of sampling the target density π, we sample another density, g, known as an importance density having the property that $g(x) = 0$ implies $\pi(x) = 0$, then one can write

$$E_\pi(f(X)) = \int f(x)\frac{\pi(x)}{g(x)}g(x)dx = E_g(f(X)w^*(X)), \tag{21}$$

where $w^*(x) = \pi(x)/g(x)$ is the so-called importance function. This suggestsapproximating the expected value of interest by generating a random sample $x^i,\ iin1, \dots, N$ from g and computing

$$\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w^*(x^{(i)}) = E_\pi(f(X)).$$

The sample $x^i,\ iin1, \dots, N$, with the associated weights $w^i,\ iin1, \dots, N$, canbe viewed as a sample from target density π. This is a great idea which has been applied in Bayesian econometrics over more than 3 decades. The Cromolab.io team member, Prof. H.K. van Dijk, has brought importance sampling to econometrics in his joint paper with T. Kloek, (Van Dijk and Kloek 1978). That is where the modern, computational, econometrics has its root.

In filtering problem, the main challenge concerns the target distribution which changes every time a new observation is made, moving from $p(\theta_{0:t-1}|y_{1:t-1})$ to $p(\theta_{0:t}|y_{1:t})$. How to efficiently update the former to get the proper estimate of the latter? Actually we follow the logic explained in formulas Equation 19 and Equation 7.1. We see that the prior for new period is necessary to obtain the posterior accordingly to in Equation 19, denoted as $\theta_t|y_{1:t-1}$. We follow logic in Equation 7.1 to update this prior properly. To that end for each θ^i_{t-1} from the support of $\hat{\pi}_{t-1}$ we simulate θ^i_t according to $p(\theta_t|\theta_{t-1})$ in Equation 7.1. Then based on the simulated candidate for new set of parameter values, we update the system of weights, w^i_{t-1} to w^i_t. Those weights together with the simulated θ^i_t constitute an proper approximation of $\hat{\pi}_t$.

What remain to be explained is the way to approximate the weights. Dropping the superscripts for notational simplicity, the weights are given by

$$w_t \propto \frac{\pi(y_t|\theta_t)\cdot\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})}\cdot w_{t-1} \tag{22}$$

for discussion we refer for instance to… TODO(add the) These weights need to be normalized of course to lead to a proper density $\hat{\pi}_t$.

Below we present a example of implementation of this algorithm for binomial distribution. It is assumed that the each period a random variable is drawn from the binomial probability. There is another variable behind which evolves over time and represents the parameter of the binomial distribution.

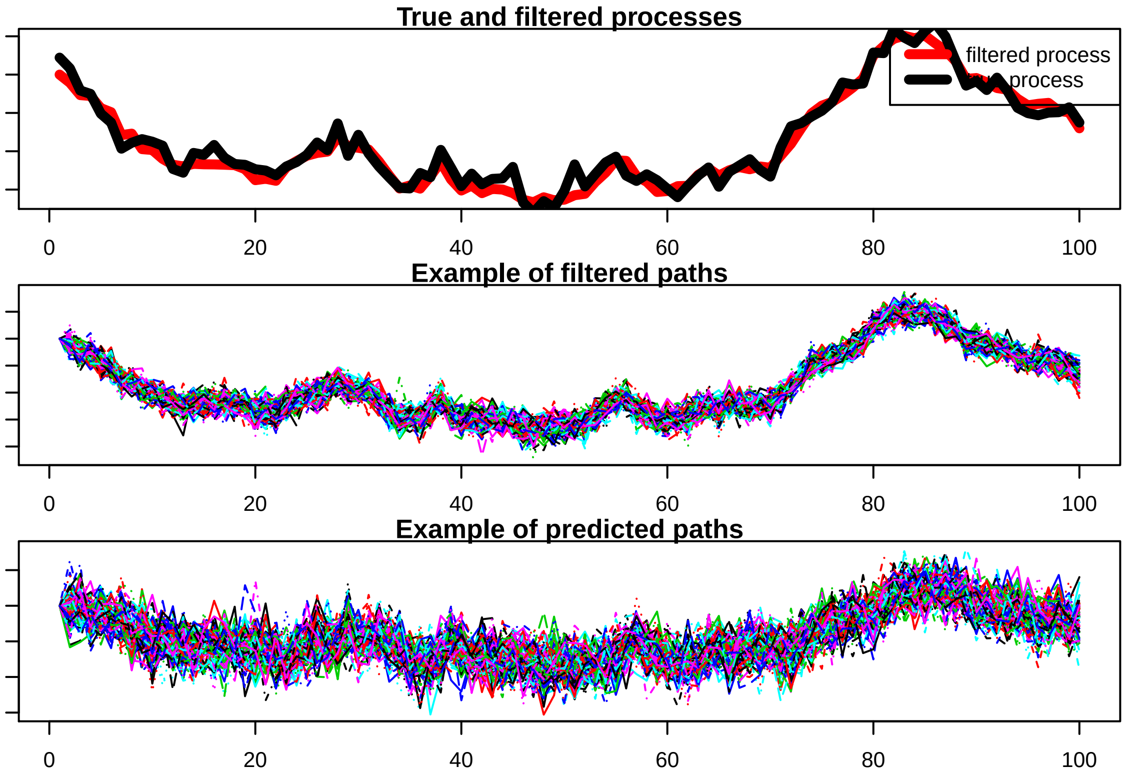


Figure 12: Example of particle filtering

7.3 Particle filtering with unknown parameters

7.4 Sample results

http://mathworld.wolfram.com/RandomWalk1-Dimensional.html

8 Hedging portfolio construction for cryptocurrency market

The main goal of the discussion above was to understand the idea that the evolution of the probability curve for cryptocurrencies analyzed by the system allows for selection of cryptocurrency with the expected behavior: rising or declining trend. As soon as that step has been accomplished, the investor has two options: either to invest in this cryptocurrency in an outright manner (creating a portfolio consisiting only of this cryptocurrency), or investing in a wider portfolio, consisting also of cryptocurrencies, which constitute a proper hedge for the underlying crptocurrency.

Consider the situation that an investor enters an unhedged, outright position. She is worried about the risk due to changing prices which might developed in an unpredictable manner due to the idiosyncratic shocks which are definitely observed in the market and add on substantially to the risk of this investment. To mitigate this risk the investor decides to hedge by short selling other cryptocurrencies which are correlated with the underlying cryptocurrecny. (Johansen and Gatarek 2017) has shown that to properly hedge the assets which developes as random walks evolving accrodingly to some common trends, the investor is better off is she hedges with assets which are not only correlated but also cointegrated with the underlying. The problem is which amounts, the hedge ratios, the hedging cryptocurrencies should be bought/sold to hedge optimally the risk due to the variation of the prices, as measured by conditional variance. Note that instead of holding the first asset, we are buying it and short seeling the hedging assets. In case of a short position in the underlying asset (i.e. when the trend is declining), we are short selling the undelying asset and buying the hedges according to the hedge ratios.

8.1 Cointegration modeling

Before we skizz the methodology of (Johansen and Gatarek 2017) which constitute a basis for the portfolio construction part of the platform, we shortly explain the idea behind the cointegration model, which is a necessary element on the way to compute the hedge ratios.

The cointegrated vector autoregressive model (CVAR) is assumed to describe the variation of the cryptocurrency prices. This model allows for nonstationary prices (random walk, which have nonstationary behaviour) with stationary linear combinations, that is cointegration. Cointegration models time series which follow common trends. Linear combination of individual time series is nothing more as a common trend which is followed by the individual prices with a different strentgh. Some of the first attempts to model common trends with cointegration can be found in (Kasa 1992). Since then, cointegration has been found and tested for in many financial markets.

The cointegration approach to trading in financial markets has been implemented in many important studies studies, for instance, (Lin, McCrae, and Gulati 2006), (Vidyamurthy 2004), (Gillespie and Ulph 2001), (Alexander and Dimitriu 2005a), (Alexander and Dimitriu 2005b) and (Gutierrez and Tse 2011). Recently, researchers, who are part of the project group have published a paper, see (Ardia et al. 2016) which shows important theoretical development necessary for succesfull application of cointegration to financial market analysis. Those techniques are directly applicable to the case of cryptocurrencies and will be part of the engine behind the platform.

To present an example of how cointegration model works in practice, we have estimated the cointegration model on a set of cryptocurrencies presented in figure Figure 13. This figure present the evolution of value of Bitcoin and Ripple over last few months, with prices normalized as 1 on the 1st of April 2017. With the red line the common trend estimated with cointegration model has been displayed. It is interesting to observe that the common trend is, usually, smoother than the time series which combine into it.

The general outcome of cointegration modeling is a set of parameters which define the number of common trends driving the system of time series. Those parameters are necessary for implementing hedging methodology presented in (Johansen and Gatarek 2017). In what follows we present shortly the idea behind this methodology.

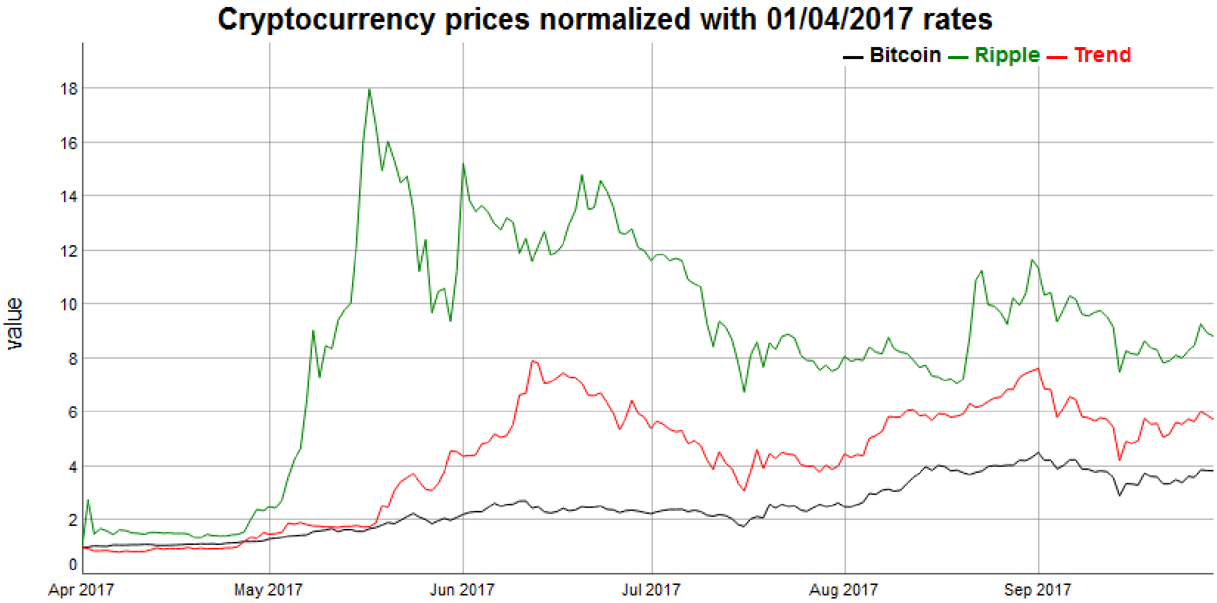


Figure 13: Example of common trend for two selected cryptocurrencies

8.2 Hedging portfolio for cryptocurrency based on cointegration model

In general, the hedging methods can be divided in two classes: static and dynamic methods. The static hedging techniques assume that the hedging portfolio is selected, given information available in period t, and remains unchanged during the entire holding period $t+1,\dots,t+h$. This is opposed to the dynamic hedging methods which allows for rebalancing the portfolio during the holding period, but we are only concerned with static hedging. The holding period h is implied by the cyclic behaviour in the evolution of probability distribution of random walk. The expected holding period can be derived from properties of this time series.

(Johansen and Gatarek 2017) study optimal hedging for an h-period investment in a system that consists of a set of assets, which follow random walks driven by some common trends. It is assumed that there are n assets, in this case cryptocurrencies, with prices modeled with time series $y_t = (y_{1t},\dots,y_{nt})'$, and that the first asset is held for h periods, using the other cryptocurrency to hedge the risk, as measured by conditional variance of returns $\Sigma_{t,h} = Var_t(y_{t+h} - y_t)$ given information at time t, that is y_s, $s = 1,\dots,t$.

There are two main results of (Johansen and Gatarek 2017) which find their application in the cryptocurrency analyzing platform, that we develop. The first set of results to be implmented in the platform concerns the derivation of an expression for the risk, $\Sigma_{t,h}$, which depends on conditional (given the period t) volatility of the error term. Based on this expression, the optimal $h-$period hedging portfolio, which minimizes this risk is derived. By the optimal hedging portfolio we mean the weights corresponding to particular cryptocurrencies entering the portfolio. The limit for $h \to \infty$ of the inverse risk matrix, $\Sigma_{t,h}^{-1}$, is found and used to show that the optimal portfolio approaches a variance minimal cointegrating portfolio, which has a bounded risk.

Thus for longer horizons we should choose the variance minimal cointegrating portfolio, which has a bounded risk, and for shorter horizons we should take conditional volatility into account. The period in between constitute a balance between the long term cointegration based hedge and the short term correlation hedge, closely connected to the conditional volatility.

This result is crucial and stands in opposition to the literature in financial econometrics so far, which has positioned correlation as a main source of insight for hedging. According to the statistically based technology developed in (Johansen and Gatarek 2017), the correlation is only informative as soon as one-day ahead holding period is concerned.

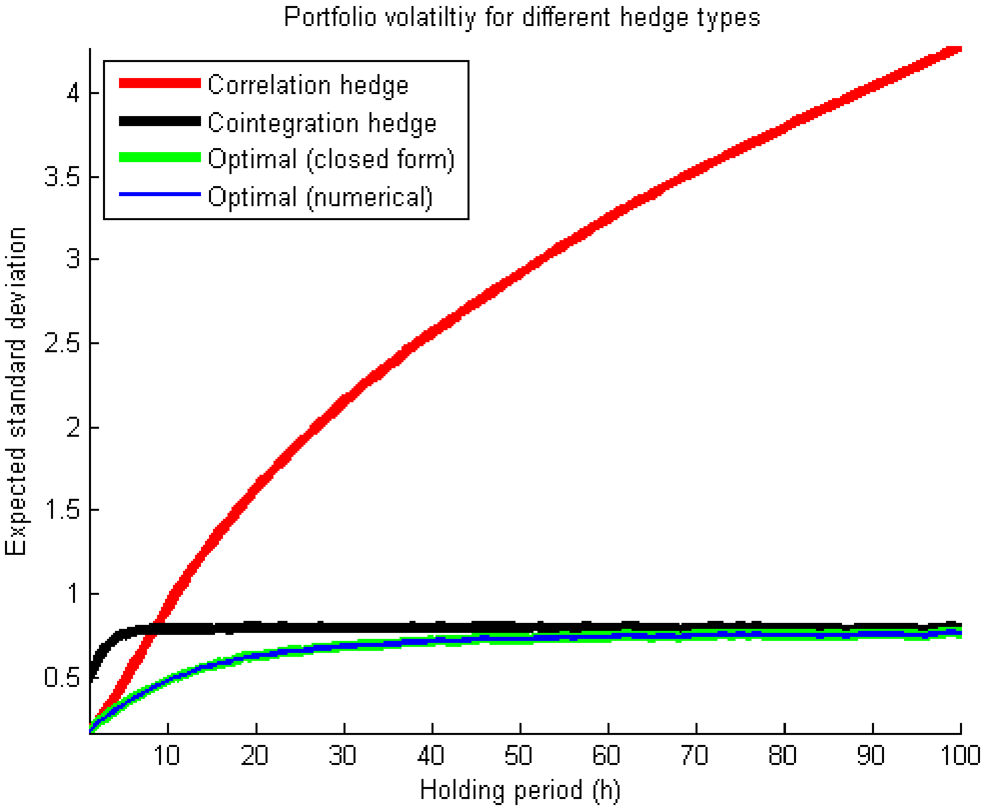


Figure 14: Optimal hedging according to Gatarek and Johansen (2017) versus alternative solutions in terms of minimizing variance (expected standard deviation of portfolio)

The main implications of the methodology is presented in Figure Figure 14. We can observe the variance of portfolio is unbounded if the correlation based hedge is applied. At the same time the cointegration based hedge is inefficient for shorter holding period. In that case the correlation hedge is advantageous. Figure Figure 14 is based on a portfolio which consists of two assets only. In case of multiple assets entering the portfolio the extent of inefficiency corresponding to the unoptimal hedging is substantial. First of all it can lead to losses in a consequence of improper hedging (too little hedging assets in the portfolio) as well as inflated transaction costs implied by overhedging. WE shall keep in mind that overhedging leads to an additional risk in terms of an outright position resulting from too high position in a hedging asset.

The second set of results concerns estimation of risk, and the optimal h-period hedging portfolio based on data $y_t, t = 1, \ldots, T$. Under assumptions on the error term that allows for heteroscedasticity, we show two results. First we show that a regression of returns $y_{t+h} - y_t$ on information at time t gives a consistent estimator for Σ_h, and a similar result holds if the CVAR is estimated by reduced rank regression. Next it is shown that a regression of y_{1t} on the other prices and a constant gives a consistent estimator of the optimal limiting hedging portfolio.

The conclusion of this is, that if the conditional variance is used as risk measure, in the case of conditional volatility, this has to be modelled by a multivariate GARCH model, like the BEKK model, see for instance Engle and Kroner (1995) and Comte and Lieberman (2003), or a multivariate ARCH model like Li, Ling, and Wong (2001). The combined theory of cointegration and a model for heteroscedasticity is challenging. The obvious two-step procedure of first estimating the CVAR assuming i.i.d. Gaussian errors and then use the estimated residuals as input in a BEKK model has not been work out in details.

The well-known formula

$$Var(y_{t+h} - y_t) = E(Var_t(y_{t+h} - y_t)) + Var(E_t(y_{t+h} - y_t)),$$

shows that the choice between the conditional variance, $\Sigma_{h,t}$ and its expectation, Σ_h, does not involve the variation of the information y_t given at the time of investment.

If a consistent estimator of $\Sigma_{t,h} = Var_t(y_{t+h} - y_t)$ is needed, one has to model conditional volatility, but if the first term $\Sigma_h = E(Var_t(y_{t+h} - y_t))$ can be used, it can be estimated by the simple regression methods or from the CVAR.

The role of cointegration for hedging was analysed by Juhl, Kawaller, and Koch (2011). They considered a special case of the CVAR, and we want in this paper to generalize their results to a CVAR with more lags and more cointegrating relations and allow for a some degree of heteroscedasticity in the martingale error term.

Finally we analyze some daily data for futures of electricity prices, and compare the optimal hedging portfolio with the cointegrating portfolio. All proofs are given in the Appendix.

We conclude that cointegration plays an important role in hedging. It allows for the possibility that an h-period hedging portfolio has a risk that is bounded in the horizon h, as opposed to the unhedged risk. As important is the result that for moderate horizons, it is important not to use the cointegrating portfolio, but to use the optimal hedging portfolio which interpolates between the short and long-horizon cointegrating portfolio.

8.3 Portfolio risk analysis

Despite of advanced hedging methods which are responsible for portfolio construction component of the platform the risk can never be fully eliminated. Therefore, the risk analysis component plays important part of the platform. To that end we apply the methodology in (Ardia, Hoogerheide, and Gatarek 2017) that is fully applicable for risk analysis of portfolio with short holding periods. In case of the platfrom in development, that is extemely important aspect, as based on prototyping we know that the cycles in the cryptocurrency trends extend usually over a few days, maximally to a few weeks. Details of the methodology are presented in the referred paper.

9 Market making

10 Trend signals example

11 Information management

12 IT side

13 Products

13.1 Product 1

13.2 Product 2

13.3 Product 3

14 Development plan

15 Token

References

Alexander, C., and A. Dimitriu. 2005a. "Indexing and Statistical Arbitrage: Tracking Error or Cointegration?" *Journal of Portfolio Management* 31: 50–63.

———. 2005b. "Indexing, Cointegration and Equity Market Regimes." *International Journal of Finance and Economics* 10: 213–31.

Ardia, D., Gatarek L.T., L. Hoogerheide, and H.K. Van Dijk. 2016. "Return and Risk of Pairs Trading Using a Simulation-Based Bayesian Procedure for Predicting Stable Ratios of Stock Prices." *Econometrics* 4 (1).

Ardia, D., L. Hoogerheide, and L.T. Gatarek. 2017. "A New Bootstrap Test for Multiple Assets Joint Risk Testing." *Journal of Risk* 19 (4).

Chu, J., S. Nadarajah, and S. Chan. 2015. "Statistical Analysis of the Exchange Rate of Bitcoin." *PloS One*, 1–27.

Feller, W. 1957. *An Introduction to Probability Theory and Its Applications (Volume 1)*. John Wiley & Sons Inc., (2. ed.).

Gillespie, T., and C. Ulph. 2001. "Pair Trades Methodology: A Question of Mean Reversion." Proceedings of International Conference on Statistics, Combinatorics and Related Areas and the 8th International Conference of Forum for Interdisciplinary Mathematics, NSW.

Grammig, Melvin, J., and C. Schlag. 2005. "Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects." *Journal of Financial Econometrics*

12: 139–64.

Grassi, S., and L. Catania. 2017. "Modelling Crypto-Currencies Financial Time-Series." https://ssrn.com/abstract=3028486.

Gutierrez, J. A., and Y. Tse. 2011. "Illuminating the Profitability of Pairs Trading: A Test of the Relative Pricing Efficiency of Markets for Water Utility Stocks." *The Journal of Trading* 6 (2): 50–64.

Hasbrouck, J. 1988. "One security, many markets: determining the contributions to price discovery." *The Journal of Finance* 50: 1175–99.

Hencic, A., and C. Gourieroux. 2014. "Noncausal Autoregressive Model in Application to Bitcoin USD Exchange Rate." Proceedings of the 7th Financial Risks International Forum, 125–25.

Johansen, S. 1988. "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control* 12: 231–54.

———. 2006. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford (2. ed.).

Johansen, S., and L. Gatarek. 2017. "The Role of Cointegration for Optimal Hedging with Heteroscedastic Error Term." *Journal of Econometrics* final revision. http://pure.au.dk/portal/en/publications/id(c0d43 8cba-45fe-b554-f0b716c4d389).html.

Jong, F. de, and P.C. Schotman. 2010. "Price discovery in fragmented markets." *Journal of Financial Econometrics* 8: 1–28.

Kasa, K. 1992. "Common stochastic trends in international stock markets." *Journal of Monetary Economics* 29: 95–124.

Lehmann, B.N. 2002. "Some desiderata for the measurement of price discovery across markets." *Journal of Financial Markets* 5: 259–76.

Lin, Y.-X., M. McCrae, and C. Gulati. 2006. "Loss Protection in Pairs Trading Through Minimum Profit Bounds: A Cointegration Approach." *Journal of Applied Mathematics and Decision Sciences*, 1–14.

Sapuric, S., and A. Kokkinaki. 2014. "Bitcoin Is Volatile. Isnt That Right." Business Information Systems Workshops, 255–65.

Van Dijk, H., and R. Kleijn. 2006. "Bayes Model Averaging of Cyclical Decompositions in Economic Time Series." *Journal of Applied Econometrics* 21 (2). http://dx.doi.org/10.1002/jae.823: 191–212.

Van Dijk, H., and T. Kloek. 1978. "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo." *Econometrica* 46 (1).

Van Dijk, H., A.C. Harvey, and T.M. Trimbur. 2007. "Trends and Cycles in Economic Time Series: A Bayesian Approach." *Journal of Econometrics* 140 (2). http://dx.doi.org/10.1016/j.jeconom.2006.07.006: 618–49.

Van Dijk, H., Billio M., Casarin R., and Ravazzolo F. 2013. "Time-Varying Combinations of Predictive Densities Using Nonlinear Filtering." *Journal of Econometrics* 177 (2).

Vidyamurthy, G. 2004. *Pairs Trading: Quantitative Methods and Analysis.* New York: Wiley.