Ergodicity Economics

Ole Peters and Alexander Adamou
London Mathematical Laboratory
14 Buckingham Street, London WC2N 6DF, UK

2015/11/25 at 12:02:01

Abstract

Contents

1 Tossing a coin

1.1 The game

{section:The_game}

Imagine we offer you the following game: we toss a coin, and if it comes up heads we increase your monetary wealth by 50%; if it comes up tails we reduce your wealth by 40%. We're not only doing this once, we will do it many times, for example once per week for the rest of your life. Would you accept the rules of our game? Would you submit your wealth to the dynamic our game will impose on it?

Your answer to this question is up to you and will be influenced by many factors, such as the importance you attach to wealth that can be measured in monetary terms, whether you like the thrill of gambling, your religion and moral convictions and so on.

In these notes we will mostly ignore these. We will build an extremely simple model of your wealth, which will lead to an extremely simple powerful model of the way you make decisions that affect your wealth. We are interested in analyzing the game mathematically, which requires a translation of the game into mathematics. We choose the following translation: we introduce the key variable, $x(t)$, which we refer to as "wealth". We refer to t as "time". It should be kept in mind that "wealth" and "time" are just names that we've given to mathematical objects. We have chosen these names because we want to compare the behaviour of the mathematical objects to the behaviour of wealth over time, but we emphasize that we're building a model – whether we write $x(t)$ or wealth(time) makes no difference to the mathematics.

The usefulness of our model will be different in different circumstances, ranging from completely meaningless to very helpful. There is no substitute for careful consideration of any given situation, and labeling mathematical objects in one way or another is certainly none.

Having got these words of warning out of the way, we define our model of the dynamics of your wealth under the rules we specified. At regular intervals of duration Δt we randomly generate a factor $r(t)$ with each possible value $r_i \in \{0.6, 1.5\}$ occurring with probability 1/2,

$$r(t) = \begin{cases} 0.6 & \text{with probability } 1/2 \\ 1.5 & \text{with probability } 1/2 \end{cases} \tag{1}$$

{eq:law}

and multiply current wealth by that factor,

$$x(t + \Delta t) = r(t + \Delta t) \times x(t). \tag{2}$$

{eq:gamble}

Without discussing in depth how realistic a representation of your wealth this model is (for instance your non-gambling related income and spending are not represented in the model), and without discussing whether randomness truly exists and what the meaning of a probability is we simply switch on a computer and simulate what might happen. You may have many good ideas of how to analyze our game with pen and paper, but we will just generate possible trajectories of your wealth and pretend we know nothing about mathematics or economic theory. Figure 1 is a trajectory of your wealth, according to our computer model as it might evolve over the course of 52 time steps (corresponding to one year given our original setup).

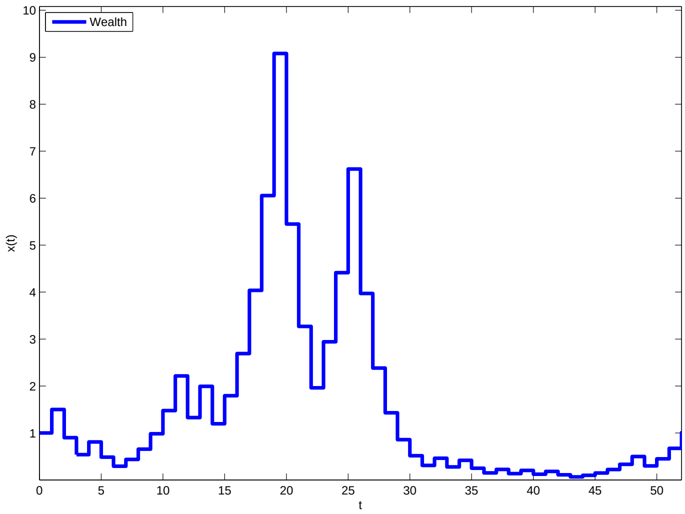


Figure 1: Wealth $x(t)$ resulting from a computer simulation of our game, repeated 52 times. {fig:1_1}

A cursory glance at the trajectory does not reveal much structure. Of course there are regularities, for instance at each time step $x(t)$ changes, but no trend is discernible – does this trajectory have a tendency to go up, does it have a tendency to go down? Neither? What are we to learn from this simulation? Perhaps we conclude that playing the game for a year is quite risky, but is the risk worth taking?

1.1.1 Averaging over many trials

A single trajectory doesn't tell us much about overall tendencies. There is too much noise to discern a clear signal. A common strategy for getting rid of noise is to try again. And then try again and again, and look at what happens on average. For example, this is very successful in imaging – the 2014 Nobel Prize in chemistry was awarded for a technique that takes a noisy image again and again. By averaging over many images the noise is reduced and a resolution beyond the diffraction limit is achieved.

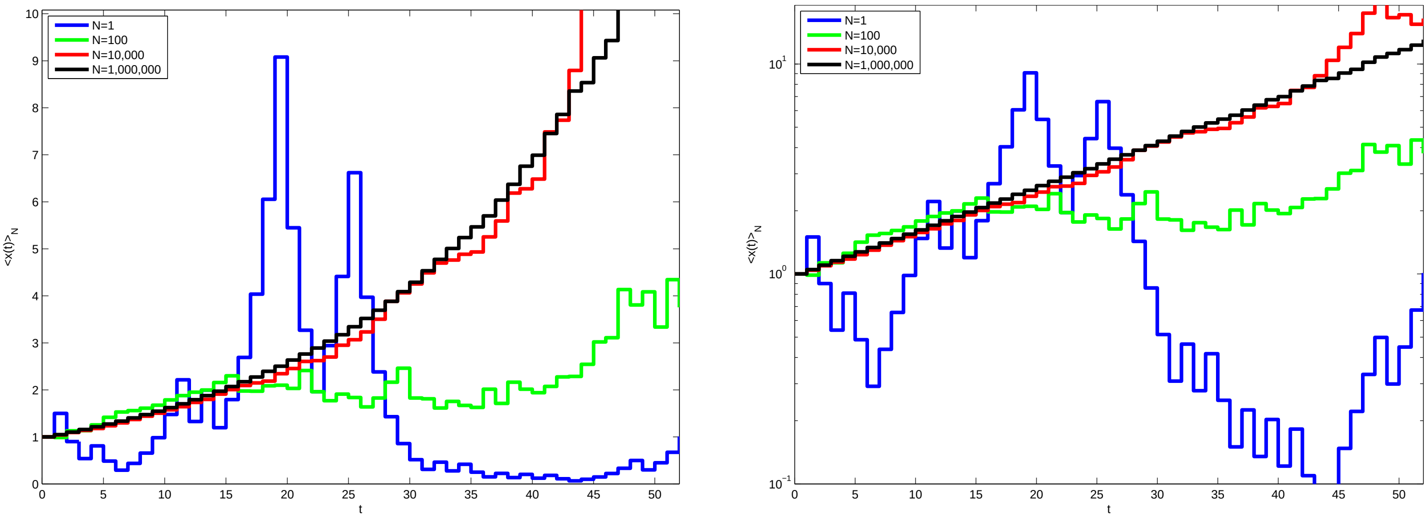
So let's try this in our case and see if we can make sense of the game. In Fig. 2 we average over finite ensembles of N trajectories, that is, we plot the finite-ensemble average.

DEFINITION: **Finite-ensemble average**

The finite-ensemble average of the observable A is

$$\langle A(t) \rangle_N = \frac{1}{N} \sum_i^N A_i(t), \tag{3}$$ {eq:f_ens}

where i indexes a particular realization of A and N is the number of realizations included in the average.



(A) (B)

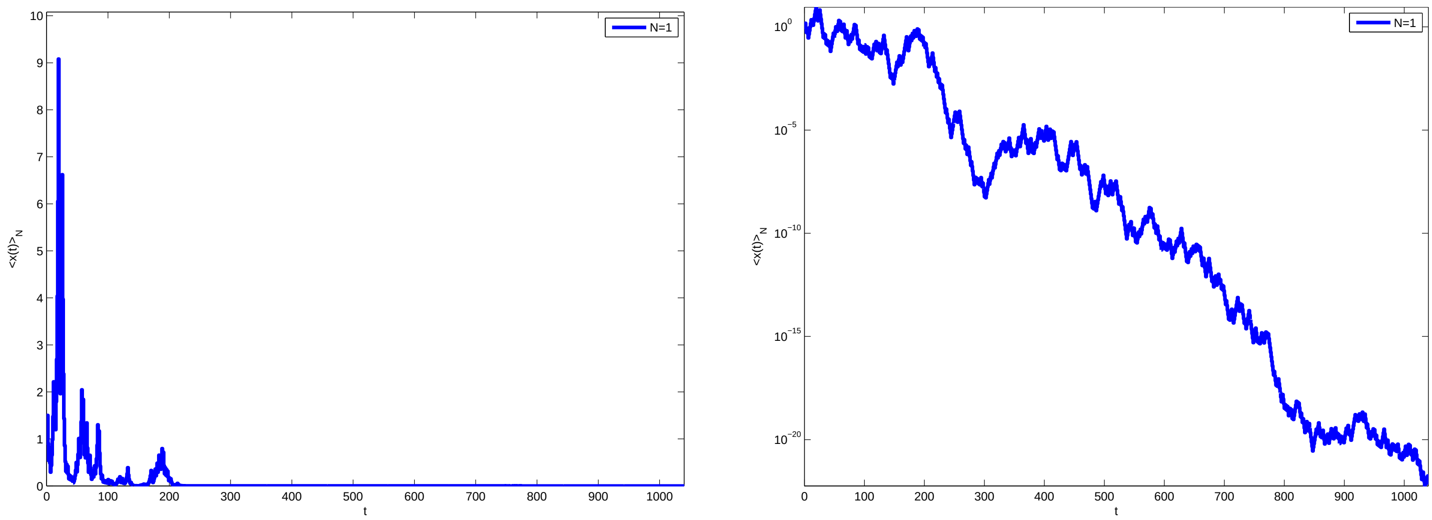
Figure 2: Partial ensemble averages $\langle x(t)\rangle_N$ for ensemble sizes $N = 1, 10^2, 10^4, 10^6$. (A) linear scales, (B) logarithmic scales. {fig:1_2}

As expected, the more trajectories are included in the average, the smaller the fluctuations of that average. For $N = 10^6$ hardly any fluctuations are visible. Since the noise-free trajectory points up it is tempting to conclude that the risk of the game is worth taking. This reasoning has dominated economic theory for about 350 years now. But it is flawed.

1.1.2 Averaging over time

Does our analysis necessitate the conclusion that the gamble is worth taking? Of course it's not, otherwise we wouldn't be belaboring this point. Our critique will focus on the type of averaging we have applied – we didn't play the game many times in a row as would correspond to the real-world situation of repeating the game once a week for the rest of your life. Instead we played the game many times in parallel, which corresponds to a different setup.

We therefore try a different analysis. Figure 3 shows another simulation of your wealth, but this time we don't show an average over many trajectories but a simulation of a single trajectory over a long time. Noise is removed also in this case but in a different way: to capture visually what happens over a long time we have to zoom out – more time has to be represented by the same amount of space on the page. In the process of this zooming-out, small short-time fluctuations will be diminished. Eventually the noise will be removed from the system just by the passage of time.



(A) (B)

Figure 3: Single trajectory over 1,040 time units, corresponding to 20 years in our setup. (A) linear scales, (B) logarithmic scales. {fig:1_3}

Of course the trajectory in Fig. 3 is random, but the apparent trend emerging from the randomness strongly suggests that our initial analysis does not reflect what happens over time in a single system. Jumping ahead a little, we reveal that this is indeed the case. If it seems counter-intuitive then this is because our intuition is built on so-called "ergodic processes", whereas x is non-ergodic. We will say more about this in Sec. 1.1.5. Several important messages can be derived from the observation that an individual trajectory grows more slowly over time than an average of a large ensemble.

1. An individual whose wealth follows (Eq. 2) will make poor decisions if he uses the finite-ensemble average of wealth as an indication of what is likely to happen to his own wealth.

2. The performance of the average (or aggregate) wealth of a large group of individuals differs systematically from the performance of an individual's wealth. In our case large-group wealth grows (think GDP), whereas individual wealth decays.

3. For point 2 to be possible, *i.e.* for the average to outperform the typical individual, wealth must become increasingly concentrated in a few extremely rich individuals. The wealth of the richest individuals must be so large that the average becomes dominated by it, so that the average can grow although almost everyone's wealth decays. Inequality increases in our system.

The two methods we've used to eliminate the noise from our system are well known. The first method is closely related to the mathematical object called the expectation value, and the second is closely related to the object called the time average.

1.1.3 Expectation value

In this section we validate Fig. 2 by computing analytically the average of x over infinitely many trials, a quantity known as the expectation value. The expectation value is usually introduced as the sum of all possible values, weighted by their probabilities. We will define it as a limit instead, and then show that this limit is identical to the familiar expression.

> Definition: **Expectation value i**
>
> The expectation value of a quantity x is the large-ensemble limit of the finite-ensemble average (Eq. 3),
>
> $$\langle x \rangle = \lim_{N\to\infty} \langle x \rangle_N \, . \tag{4}$$

{eq:ens}

This implies that in our first analysis of the problem – by averaging over N trajectories – we were approximately using the expectation value as a gauge of the desirability of the game. We will now prove that letting $N \to \infty$ is indeed the same as working with the more familiar definition of the expectation value.

> Definition: **Expectation value ii**
>
> The expectation value of a quantity x that can take discrete values x_j is the sum of all possible values weighted by their probabilities p_j
>
> $$\langle x \rangle = \sum_j p_j x_j. \tag{5}$$
>
> {eq:exp_sum}
>
> If x is continuous, the expectation value is the integral
>
> $$\langle x \rangle = \int_{-\infty}^{+\infty} s \mathcal{P}_x(s) ds, \tag{6}$$
>
> where $\mathcal{P}_x(s)$ is the probability density function (PDF) of the random variable x at value s.

We now show that the two definitions of the expectation value are equivalent.

Proof. Consider the number of times the value x_j is observed at time t in an ensemble of N trajectories. Call this n_j. The finite-ensemble average can then be re-written as

$$\begin{aligned} \langle x(t) \rangle_N &= \frac{1}{N} \sum_i x_i(t) && (7) \\ &= \sum_j \frac{n_j}{N} x_j(t), && (8) \end{aligned}$$

where the subscript i indexes a particular realization of x, and the subscript j indexes a possible value of x. The fraction $\frac{n_j}{N}$ in the limit $N \to \infty$ is the probability p_j, and we find

$$\lim_{N\to\infty} \langle x(t)\rangle_N = \sum_j p_j x_j(t) \tag{9}$$

The LHS is the expectation value by the first definition as a limit, the RHS is the expectation value by the second definition as a weighted sum. This shows that the two definitions are indeed equivalent. □

We will use the terms "ensemble average" and "expectation value" as synonyms, carefully using the term "finite-ensemble average" for finite N.

We pretended to be mathematically clueless when carrying out the simulations, with the purpose to gain a deeper conceptual understanding of the expectation value. We now compute the expectation value exactly instead of approximating it numerically. Consider the expectation value of (Eq. 2)

$$\langle x(t+\Delta t)\rangle = \langle x(t) r(t)\rangle \,. \tag{10}$$

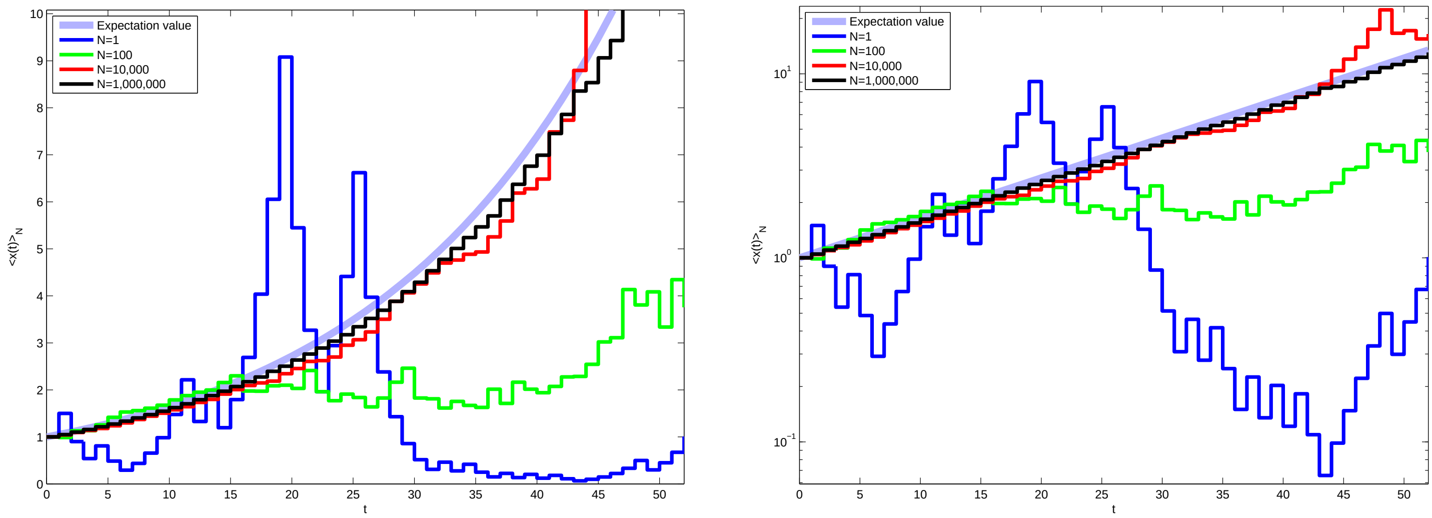
Since $r(t)$ is independent of $x(t)$ (we generate $r(t)$ independently of $x(t)$ in each time step), this can be re-written as

$$\langle x(t+\Delta t)\rangle = \langle x(t)\rangle \langle r\rangle \,, \tag{11}$$

wherefore we can solve recursively for

$$\langle x(t)\rangle = x(0) \langle r\rangle^t \,. \tag{12}$$

The expectation value $\langle r\rangle$ is easily found from (Eq. 1) as $\langle r\rangle = \frac{1}{2} \times 0.6 + \frac{1}{2} \times 1.5 = 1.05$. Since this number is greater than one, $\langle x(t)\rangle$ grows exponentially in time at rate 1.05 per time unit, or expressed as a continuous growth rate, at $\frac{1}{\Delta t} \ln \langle r\rangle \approx 4.9\%$ per time unit. This is what might have led us to conclude that the gamble is worth taking. Figure 4 compares the analytical result for the infinite ensemble to the numerical results of Fig. 2 for finite ensembles.



(A) (B)

Figure 4: Expectation value (thick light blue line) finite-ensemble averages. (A) linear scales, (B) logarithmic scales. {fig:cf_exp}

We stress that the expectation value is just some mathematical object – someone a long time ago gave it a suggestive name, but we certainly shouldn't give any credence to a statement like "we expect to see $\langle x \rangle$ because it's the expectation value." Mathematical objects are quite indifferent to the names we give them.

History: The invention of the expectation value

Expectation values were not invented in order to assess whether a gamble is worth taking. Instead, they were developed to settle a moral question that arises in the following somewhat contrived context: Imagine playing a game of dice with a group of gamblers. The rules of the game are simple: we roll the dice three times, and whoever rolls the most points gets the pot to which we've all contributed equal amounts. We've already rolled the dice twice when suddenly the police burst in because they've heard of our illegal gambling ring. We all avoid arrest, most of us escape through the backdoor, and to everyone's great relief you had the presence of mind to grab the pot before jumping out of a conveniently located ground-floor window. Later that day, under the cover of dusk, we meet behind the old oak tree just outside of town to split the pot in a fair way. But hold on – what does "fair" mean here? Some of us had acquired more points than others in the first two rolls of the dice. Shouldn't they get more? The game was not concluded, so wouldn't it be fair to return to everyone his wager and thank our lucky stars that we weren't arrested? Should we split the pot in proportion to each player's points? All of these solutions were proposed [4]. The question is fundamentally moral, and there is no mathematical answer. But Blaise Pascal, now famous for addressing theological questions using expectation values, put the problem to Fermat, and over the course of a few months' correspondence (the two never met in person) Fermat and Pascal agreed that fairness is achieved as follows: Imagine all (equally likely) possible outcomes of the third round of throws of the dice, call the number of all possibilities N. Now count those possibilities that result in player j winning, call this n_j. If Δx^+ is the amount of money in the pot, then we split the pot fairly by giving each player $\frac{n_j}{N} \times \Delta x^+$. This is $\langle \Delta x^+ \rangle$, according to (Eq. 5) because $\frac{n_j}{N} = p_j$ is the probability that player j wins the amount Δx^+. Later researchers called this amout the "mathematical expectation" or simply "expectation value". But this is really an unfortunate choice – no player "expected" to receive $\langle \Delta x^+ \rangle$. Instead, each player expected to receive either nothing or Δx^+.

1.1.4 Time average

In this section we validate Fig. 3 and compute analytically what happens in the long-time limit. The blue line in Fig. 3 is not completely smooth, there's still some noise. It has some average slope, but that slope will vary from realization to realization. The longer we observe the system, *i.e.* the more time is represented in a figure like Fig. 3, the smoother the line will be. In the long-time limit, $t \to \infty$, the line will be completely smooth, and the average slope will be a deterministic number – in any realization of the process it will come out identical.

The dynamic is set up such that wealth at time t is

$$x(t) = \prod_{t'=1}^{t} r(t'), \tag{13}$$

which we can split up into two products, one for each possible value of $r(t)$, which we call r_1 and r_2. Let's denote the number of r_1s by n_1 and r_2s by n_2, so that

$$x(t) = r_1^{n_1} r_2^{n_2}. \tag{14}$$

We denote by r_{time} the factor by which $x(t)$ changes per time step when the change is computed over a long time. This quantity is found by taking the t^{th} root of $x(t)$ and considering the long-time limit

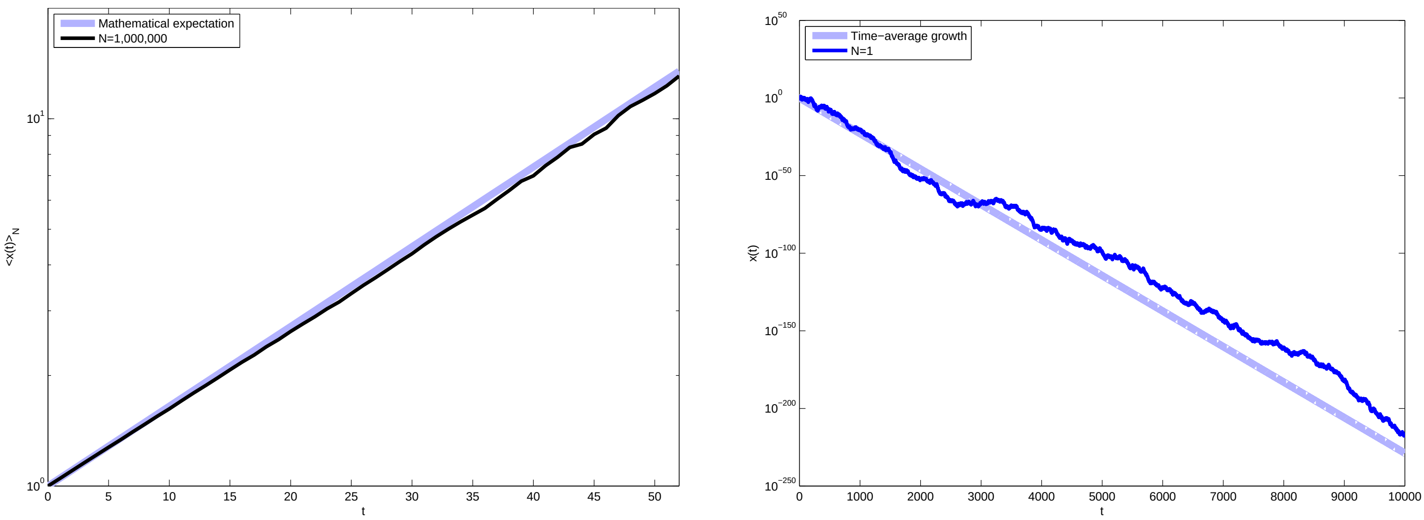
$$\begin{aligned} r_{\text{time}} &= \lim_{t\to\infty} x(t)^{1/t} && (15)\\ &= \lim_{t\to\infty} r_1^{n_1/t} r_2^{n_2/t}. && (16) \end{aligned}$$

Identifying $\lim_{t\to\infty} n_1/t$ as the probability p_1 for r_1 to occur (and similarly $\lim_{t\to\infty} n_2/t = p_2$) this is

$$\lim_{t\to\infty} x(t)^{1/t} = (r_1 r_2)^{1/2}, \tag{17}$$ {eq:long_t}

or $\sqrt{0.9} \approx 0.95$, *i.e.* a number smaller than one, reflecting decay in the long-time limit for the individual trajectory. The trajectory in Fig. 3 was not a fluke: *every* trajectory will decay in the long run at a rate of $(r_1 r_2)^{1/2}$ per time unit.

Figure 5 (B) compares the trajectory generated in Fig. 3 to a trajectory decaying exactly at rate r_{time} and places it next to the average over a million systems.



(A) (B)

Figure 5: (A) Finite-ensemble average for $N = 10^6$ and 52 time steps, the light blue line is the expectation value.[2] (B) A single system simulated for 10,000 time steps, the light blue line decays exponentially with the time-average decay factor $\bar{r}$ in each time step. {fig:1_4}

Excursion: Scalars

$r(t)$ is a random variable, whereas both $\langle r \rangle$ and r_{time} are scalars. Scalars have the so-called "transitive property" that is heavily relied upon in economic theory. Let a_i be a set of scalars. Transitivity means that if $a_1 > a_2$ and $a_2 > a_3$ we have $a_1 > a_3$. Notice that we cannot rank random variables in such a way. The "greater than" relation, $>$, is not defined for a pair of random variables, which is the mathematical way of saying that it is difficult to choose between two gambles, and it is why we went to the trouble of removing the randomness from the stochastic process $x(t)$. Removing randomness by averaging always involves a limiting process, and results are said to hold "with probability one". In the case of $r_{\langle\rangle}$ we considered

[2]In Fig. 5 (A) a slight discrepancy between the expectation value and the $N = 10^6$ finite-ensemble average is visible, especially for later times. This is not coincidence – in the long run, also the finite-ensemble average for 10^6 systems will decay at the time-average growth rate. A simple proof for this for the continuous process is in [7]. Moreover, much is known about the distribution for $\langle x(t) \rangle_N$ for any N and t (because it happens to be mathematically identical to the partition function of the random energy model, introduced and solved by Derrida [3]. We thank J.-P. Bouchaud for pointing this out to us).

the infinite-ensemble limit, $N \to \infty$, and in the case of r_{time} we considered the infinite-time limit, $t \to \infty$. If we use the scalars a_i to represent preferences, we can test for consistency among preferences. For instance, in such a model world where preferences are represented by scalars, the facts that "I prefer kangaroos to Beethoven" and "I prefer mango chutney to kangaroos" imply the fact "I prefer mango chutney to Beethoven". Translating back to reality, economists like to call individuals who make the first two statements but not the third "irrational."

Because transitivity makes for a nicely ordered world, it is useful to find scalars to represent preferences. We are skeptical about the attempt to map all preferences into scalars because the properties of mango chutney are too different, *qualitatively*, from the properties of Beethoven. We will restrict our analysis to money – the amount of money we will receive is random and this introduces a complication, but at least we know how to compare one amount to another in the limit of no randomness – there is no qualitative differences between \$1 and \$3, only a quantitative difference.

Both $r_{\langle\rangle}$ and r_{time} are scalars, and both are therefore potentially powerful representations of preferences. Your decision whether to accept our gamble could now be modelled as a choice between the value of the scalar r_{time} if you do not accept our game, namely $a_1 = 1$, and the value of the scalar r_{time} if you do accept, namely approximately $a_2 = 0.95$. In this model of your decision-making you would prefer not to play because $1 > 0.95$.

We have two averages, $r_{\langle\rangle}$ and r_{time} that we have determined numerically and analytically. Neither average is "wrong" in itself; instead each average corresponds to a different property of the system. Each average is the answer to a different question. Saying that "wealth goes up, on average" is clearly meaningless and should be countered with the question "on what type of average?" An observable that neatly summarizes the two different aspects of multiplicative growth we have illustrated is the exponential growth rate, observed over finite time in a finite ensemble

$$g_{\text{est}}(N,T) = \frac{1}{T} \ln \left(\frac{1}{N} \sum_i^N x_i(T) \right). \tag{18}$$ {eq:gest}

For N, t finite this is a random variable. The relevant scalars arise as two different limits of the same stochastic object. The exponential growth rate of the expectation value (that's also $\frac{1}{\Delta t} \ln r_{\langle\rangle}$) is

$$g_{\langle\rangle} = \lim_{N\to\infty} g_{\text{est}}, \tag{19}$$

and the exponential growth rate followed by every trajectory when observed for a long time (that's also $\frac{1}{\Delta t} \ln r_{\text{time}}$) is

$$g_{\text{time}} = \lim_{t\to\infty} g_{\text{est}}. \tag{20}$$ {eq:gt}

For a single trajectory, $N = 1$, (Eq. 20) becomes

$$\begin{aligned} g_{\text{time}} &= \lim_{t\to\infty} \frac{1}{t} \ln x(t) && (21) \\ &= \lim_{T\to\infty} \frac{1}{T\Delta t} \sum_t^T \Delta \ln x(t). && (22) \end{aligned}$$

This leads us to a technical definition of the time average.

DEFINITION: **Finite-time average**

If the observable $A(t)$ only changes at T discrete times $\Delta t, 2\Delta t$ *etc.*, then the "finite-time average" is

$$\overline{A}_T = \frac{1}{T}\sum_t^T A(t). \tag{23}$$

If it changes continuously the finite-time average is

$$\overline{A}_T = \frac{1}{T\Delta t}\int_0^{T\Delta t} A(t)dt. \tag{24}$$

{eq:t_ave_f}

DEFINITION: **Time average**

The "time average" is the long-time limit of the finite-time average

$$\overline{A} = \lim_{T\to\infty} \overline{A}_T. \tag{25}$$

{eq:t_ave}

According to this definition, g_{time} is the time average of the observable $\Delta \ln x$.

History: William Allen Whitworth

$r_{\langle\rangle}$ and r_{time} are two different properties of the game. $r_{\langle\rangle}$ is computed by taking the large-ensemble limit, r_{time} is the long-time limit. The Victorian mathematician William Allen Whitworth was aware that r_{time} is the property relevant for an individual deciding whether to take part in a repeated gamble, and used this knowledge to write an appendix entitled "The disadvantage of gambling" to the 1870 edition of his book "Choice and Chance" [9]. His argument, in slightly different notation, went as follows. Imagine that you either win or lose, with equal probability, an amount $x(t)\sigma\sqrt{\Delta t}$ in each round of a game. This scaling in Δt is not arbitrarily chosen – it would be observed if the multiplicative increment in each time interval were the product of many independent smaller multiplicative increments, $\ln r = \sum_i \ln r_i \delta t$. In the long run, positive and negative changes will occur equally frequently, and to determine the overall effect we just need to consider the effect of one positive and one negative change in a row. Wealth changes by the factor

$$(1+\sigma\sqrt{\Delta t})(1-\sigma\sqrt{\Delta t}) \tag{26}$$

We take the square root of this factor to determine what happens time-averaged per round. This is

$$[(1+\sigma\sqrt{\Delta t})(1-\sigma\sqrt{\Delta t})]^{1/2} = (1-\sigma^2\Delta t)^{1/2} \tag{27}$$

Letting Δt become infinitesimal we replace it by dt, and the first term in a Taylor expansion becomes exact. We find

$$\left[(1+\sigma\sqrt{\Delta t})(1-\sigma\sqrt{\Delta t})\right]^{1/2} \to 1-\frac{\sigma^2}{2}dt. \tag{28}$$

We choose this notation (also in Sec. 1.4) to anticipate Itô 's work of 1944 that is now the basis of much of financial mathematics.

Whitworth was arguing against a dogma of expectation values of wealth, that had been established almost immediately following Fermat and Pascal's work. He hoped to show mathematically that gambling may not be a good idea even if the odds are favourable, and was a proponent of the notion that commerce should and does consist of mutually beneficial interactions rather than one winner and one loser. In the end his voice was not heard in the economics or mathematics communities. He quit mathematics to become a priest at All Saints Church in London's Margaret Street, only a 22 minute stroll away from the London Mathematical Laboratory, according to Google.

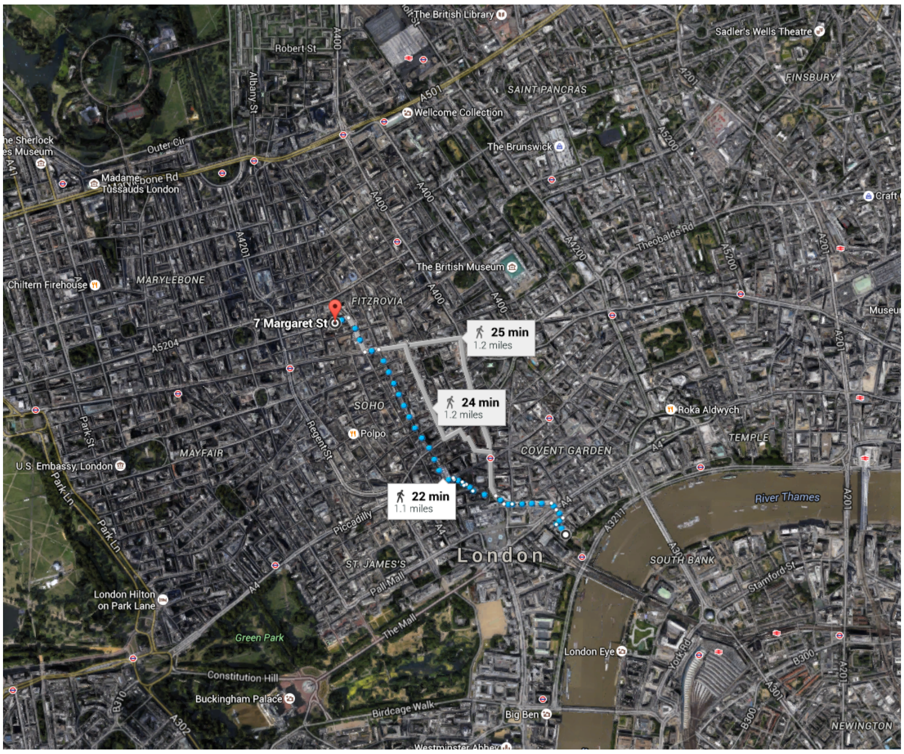


Figure 6: Location of All Saints Church and the London Mathematical Laboratory. {fig:all_saints}

1.1.5 Ergodic observables

{section:Ergodic_observab

We have encountered two types of averaging – the ensemble average and the time average. In our case – assessing whether it will be good for you to play our game, the time average is the interesting quantity because it tells you what happens to your wealth as time passes. The ensemble average is irrelevant because you do not live your life as an ensemble of many yous who can average over their wealths. Whether you like it or not, you will experience yourself owning your own wealth at future times; whether you like it not, you will never experience yourself owning the wealth of a different realization of yourself. The different realizations, and therefore the expectation value, are fiction, fantasy, imagined.

We are fully aware that it can be counter-intuitive that with probability one, a different rate is observed for the expectation value than for any trajectory over time. It sounds strange that the expectation value is completely irrelevant to the problem. A reason for the intuitive discomfort is history: since the 1650s we have been trained to compute expectation values, with the implicit belief that they will reflect what happens over time. It may be helpful to point out that all of this trouble has a name that's well-known to certain people, and that an entire field of mathematics is devoted to dealing with precisely this problem. The field of mathematics is call "ergodic theory." It emerged from the question under what circumstances the expectation value is informative of what happens over time, first raised in the development of statistical mechanics by Maxwell and Boltzmann starting in the 1850s.

History: Randomness and ergodicity in physics

The 1850s were about 200 years after Fermat and Pascal introduced expectation values into the study of random systems. Following the success of Newton's laws of motion, established around the same time as the expectation value, the notion of "proper science" had become synonymous with mechanics. Mechanics had no use for randomness and probability theory, and the success of mechanics was interpreted as a sign that the world was deterministic and that sooner or later we would understand what at the time still seemed random. At that point probability theory would become obsolete.

When Boltzmann hit upon the ingenious idea of introducing randomness into physics, to explain the laws of thermodynamics in terms of the underlying dynamics of large numbers of molecules, he was fighting an uphill battle. Neither molecules nor randomness were much liked in the physics community, especially in continental Europe, right up until the publication of Einstein's 1905 paper on diffusion [5]. Boltzmann had to be more careful than Fermat and Pascal. He had to pre-empt predictable objections from his peers, and the question of ergodicity had to be answered – the usefulness of probability theory relies heavily on expectation values, but they are technically an average over imagined future states of the universe. Boltzmann's critics were aware of this and were not shy to voice their concerns. Under what circumstances are expectation values meaningful? Boltzmann gave two answers. Firstly, expectation values are meaningful when the quantity of interest really is an average (or a sum) over many approximately independent systems. An average over a finite ensemble will be close to the expectation value if the ensemble is large enough. Secondly, expectation values are meaningful, even if only a single system exists, if they reflect what happens over time.

Boltzmann called a system "ergodic[a]" if the possible states of the system could be assigned probabilities in such a way that the expectation value of any observable with respect to those probabilities would be the same as its time average with probability 1.

[a]The word "ergodic" was coined by Boltzmann. He initially proposed the word "monodic", from Greek $\mu o \nu o$ (unique)+$o \delta o \sigma$ (path) suggesting that a single path when followed for a sufficiently long time will explore all there is to explore and reflect what happens in an ensemble. The term "ergodic" refers to the specific system Boltzmann was considering, namely an energy ($\epsilon \rho \gamma o \nu$) shell across which a path is being traced out.

To convey concisely that we cannot use the expectation value and the time average interchangeably in our game, we would say "x is not ergodic."

DEFINITION: **Ergodic property**

In these notes, an observable A is called ergodic if its expectation value is constant in time, $\frac{d\langle A\rangle}{dt} = 0$, and its time average converges to this value with probability one

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T A(t)dt = \lim_{N\to\infty}\frac{1}{N}\sum_i^N A_i. \tag{29}$$ {eq:def_ergodic}

We stress that in a given setup, some observables may have this property even if others do not. Language therefore must be used carefully. Saying our game is non-ergodic really means that some key observables of interest, most notably wealth x, are not ergodic. Wealth $x(t)$, defined by (Eq. 1), is clearly not ergodic – with $A = x$ the LHS of (Eq. 29) is zero, and the RHS is not constant in time but grows. The expectation value $\langle x\rangle(t)$ is simply not informative about the temporal behavior of $x(t)$.

This does not mean that no ergodic observables exist that are related to x. Such observables do exist, and we have already encountered two of them. In fact, we will encounter a particular type of them frequently – in our quest for an observable that tells us what happens over time in a stochastic system we will find them automatically. However, again, the issue is subtle: an ergodic observable may or may not tells us what we're interested in. It may be ergodic but not indicate what happens to x. For example, the multiplicative factor $r(t)$ is an ergodic observable that reflects what happens to the expectation value of x, whereas changes in the logarithm of wealth, $\Delta \ln x = \ln r$, are also ergodic and reflect what happens to x over time.

Proposition: $r(t)$ and $\Delta \ln x$ are ergodic.

Proof. According to (Eq. 39) and (Eq. 3), the expectation value of $r(t)$ is

$$\langle r\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_i^N r_i, \tag{30}$$ {eq:e_r}

and according to (Eq. 25), the time average of $r(t)$ is

$$\bar{r} = \lim_{T\to\infty}\frac{1}{T}\sum_t^T r_t. \tag{31}$$ {eq:t_r}

The only difference between the two expressions is the label we have chosen for the dummy variable (i in (Eq. 30) and t in (Eq. 31)). Clearly, the expressions yield the same value. $r(t)$ is stationary and takes independent identically distributed values.

The same argument holds for $\Delta \ln x$. $\square$

Whether we consider (Eq. 31) an average over time or over an ensemble is only a matter of our choice of words. Mathematically, time t and realization i only enter as dummy variables.

$\langle \Delta \ln x \rangle$ is important, historically. Bernoulli noticed in 1738 [1] that people tend to optimize $\langle \Delta \ln x \rangle$, whereas it had been assumed that they should optimize $\langle \Delta x \rangle$. Unaware of the issue of ergodicity (200 years before the concept was discovered and the word was coined), Bernoulli had no good explanation for this empirical fact and simply stated that people tend to behave as though they valued money non-linearly. We now know what is actually going on: Δx is not ergodic, and $\langle \Delta x \rangle$ is of no interest – it doesn't tell us what happens over time. However, $\Delta \ln x$ is ergodic, $\langle \Delta \ln x \rangle$ specifies what happens to x over time, and it is the right object to optimize.

When the foundations of economic theory were laid, specifically in Daniel Bernoulli's seminal paper of 1738 [1], the distinction between ergodic and non-ergodic observables was unknown. Researchers thought that the expectation value of Δx reflected what happens over time but observed that real people behaved according to what the expectation value of $\Delta \ln x$ would suggest. The origin of this discrepancy remained mysterious and numerous puzzles and paradoxes in economic theory arose as a result. The paradigm we outline here resolves these puzzles.

1.2 Rates

{section:Rates}

The ergodic observable $\Delta \ln x$, identified in the previous section, is basically a rate. If we divide it by the duration of the time step we obtain exactly the exponential growth rate of x, namely $\frac{1}{\Delta t} \Delta \ln x$. Finding good rates of change will be important, wherefore we now discuss the notion of a rate of change and the notion of stationarity. To do this properly let's think about the basic task of science. This may be described as the search for stable structure. Science attempts to build models of the world whose applicability does not vary over time. This does not mean that the world does not change, but the way in which the models describe change does not change. The model identifies something stable. This is implied by the fact that we can write equations (or English sentences) in ink on paper, with the equation (or sentence) remaining useful over time. The ink won't change over time, so if an article written in 1905 is useful today then it must describe something that hasn't changed in the meantime. These "somethings" are often somewhat grandiosely called laws of nature.

Newton's laws are a good illustration of this. They are part of mechanics, meaning that they are an idealized mathematical model of the behavior of positions, time, and masses. For instance, Newton's second law, $F = m \frac{d^2 x}{dt^2}$, states that the mass multiplied by the rate of change of the rate of change of its position equals the force. The law is an unchanging law about positions, time, and masses, but it does not say that positions don't change, it doesn't even say that rates of change of positions don't change. It does say that the rate of change of the rate of change of a position remains unchanged so long as the force and the mass remain unchanged. Newton's deep insight was to transform an unstable thing – the position of a mass – until it became stable: he fixed the force and considered rates of changes of rates of changes, et voilá!, a useful equation could be written down in ink, remaining useful for 350 years so far.

Like Newton's laws (a mathematical model of the world), our game is a prescription of changes. Unlike Newton's laws it's stochastic, but it's a prescription of changes nonetheless. Our game is also a powerful mathematical model of the world, as we will see in subsequent lectures.

We're very much interested in changes of x – we want to know whether we're winning or losing – but changes in x are not stable. Under the rules of the game the rate of change of wealth, $\frac{\Delta x(t)}{\Delta t}$, is a different random variable for each t because it is proportional to $x(t)$. But not to worry, in Newton's case changes in the position are not stable either, even in a constant force field. Nonetheless Newton found a useful stable property. Maybe we can do something similar. We're looking for a function $f(x)$ that satisfies two conditions: it should indicate what happens to x itself, and its changes should be stationary.

The first condition is that $f(x)$ must tells us whether $x(t)$ is growing or shrinking – this just means that $f(x)$ has to be monotonic in x. We know that there is something stationary about x because we were able to write down in ink how x changes. So we only need to find the monotonic function of x that inherits the stationarity of the ink in (Eq. 1). The game is defined by a set of factors of increase in $x(t)$, (Eq. 2). Therefore, the fractional change in x, namely $\frac{x(t+\Delta t)}{x(t)}$, has a stationary distribution. Which function responds additively to a multiplicative change in its argument? The answer is the logarithm, *i.e.* only the logarithm satisfies

$$f[x(t+\Delta t)] - f[x(t)] = f\left[\frac{x(t+\Delta t)}{x(t)}\right] \tag{32}$$

and we conclude that for our game $f(x) = \ln(x)$. For multiplicative dynamics, *i.e.* if $\frac{x(t+\Delta t)}{x(t)}$ is stationary, the expectation value of the rate of change of the logarithm of $x(t)$ determines whether the game is long-term profitable for an individual.

More generally, when evaluating a gamble that is represented as a stochastic process

1. Find a monotonically increasing function $f[x(t)]$ such that $\frac{\Delta f[x(t)]}{\Delta t}$ are independent instances of a stationary random variable.

2. Compute the expectation value of $\frac{\Delta f[x(t)]}{\Delta t}$. If this is positive then $x(t)$ grows in the long run, if it is negative then $x(t)$ decays.

1.3 Brownian motion

In the previous section we established that the discrete increments of the logarithm of x, which we called f, are stationary and independent in our game. A quantity for which this is the case performs a random walk. Indeed, the blue line for a single system in Fig. 2 (B) shows 52 steps of a random walk trajectory.

Random walks come in many forms – in all of them f changes discontinuously by an amount Δf drawn from a stationary distribution, in time steps Δt that are themselves drawn from a stationary distribution.

We are interested only in the simple case where f changes at regular intervals, $\Delta t, 2\Delta t ...$. For the distribution of increments we only insist on the existence of the variance, meaning we insist that $\text{var}(\Delta f) = \left\langle \Delta f^2 \right\rangle - \langle \Delta f \rangle^2$ be finite. The change in f after a long time is then the sum of many stationary independent increments,

$$f(t + T\Delta t) - f(t) = \sum_i^T \Delta f_i \tag{33}$$

The Gaussian central limit theorem tells us that such a sum, properly rescaled, will be Gaussian distributed

$$\lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_i^T \Delta f_i - T \langle \Delta f \rangle \sim \mathcal{N}(0, \text{var}(\Delta f)), \tag{34}$$

where we call $\langle \Delta f \rangle$ the drift term. The logarithmic change in the long-time limit that was of interest to us in the analysis of the coin toss game is thus Gaussian distributed. Imagine simulating a single long trajectory of f and plotting it on paper. The amount of time that has to be represented by a fixed length of paper increases linearly with the simulated time because the paper has a finite width to accommodate the horizontal axis. If $\langle \Delta f \rangle \neq 0$ then the amount of variation in f that has to be represented by a fixed amount of paper also increases linearly with the simulated time. However, the departures of $f(t)$ from its expectation value $t \langle \Delta f \rangle$ only increase as the square-root of t. Thus, the amount of paper-space given to these departures scales as $t^{-1/2}$, and for very long simulated times the trajectory will look like a straight line on paper.

In an intermediate regime, fluctuations will still be visible but they will also be approximately Gaussian distributed. In this regime it is often easier to replace the random walk model with the corresponding continuous process. That process is called Brownian motion. We define Brownian motion as follows:

DEFINITION: **Brownian motion i**

If a stochastic process has continuous paths and is distributed according to $\mathcal{N}(\mu t, \sigma t)$ then it is a Brownian motion.

The process can be defined in different ways. Another illuminating definition is this:

DEFINITION: **Brownian motion ii**

If a stochastic process is continuous, with stationary independent increments, then the process is a Brownian motion.

We quote from [6]: *"This beautiful theorem shows that Brownian motion can actually be defined by stationary independent increments and path continuity alone, with normality following as a consequence of these assumptions. This may do more than any other characterization to explain the significance of Brownian motion for probabilistic modeling."*

Indeed, BM is not just a mathematically rich model but also – due to its emergence through the Gaussian central limit theorem – a model that represents a large universality class, *i.e.* it is a good description of what happens over long times in many other models. BM is a process with two parameters, μ and σ. It can be written as a stochastic differential equation

$$df = \mu dt + \sigma dW \tag{35}$$ {eq:BM_dx}

where dW is a Wiener increment. The Wiener increment can be defined by its distribution and correlator,

$$\begin{aligned} dW &\sim \mathcal{N}(0, \sqrt{dt}) & (36) \\ \langle dW(t) dW(t') \rangle &= dt\ \delta(t, t'), & (37) \end{aligned}$$

where $\delta(t, t')$ is the Kronecker delta – zero if its two arguments differ ($t \neq t'$), and one if they are identical ($t = t'$).[3] In simulations Brownian motion paths can be constructed from a discretized version of (Eq. 35)

$$f(t + \Delta t) = f(t) + \mu \Delta t + \sigma \sqrt{\Delta t} \xi_t, \tag{38}$$

where ξ_i are instances of a standard normal distribution.

BM itself is not stationary, which makes it non-ergodic according to our definition. This is easily seen by comparing time average and expectation value

$$\langle f \rangle_N \sim \mu t + \mathcal{N}(0, t/N) \tag{39}$$ {eq:ens}

$$\bar{f}_t \sim \mu t/2 + \sigma \mathcal{N}(0, t/3) \tag{40}$$ {eq:tim}

The expectation value, *i.e.* the limit $N \to \infty$ of (Eq. 39), converges to μt with probability one, so it's not stationary, it depends on time, and it's unclear what to compare it to. Its limit $t \to \infty$ does not exist.

The time average, the limit $t \to \infty$ of (Eq. 40) diverges unless $\mu = 0$, but even with $\mu = 0$ the limit is a random variable with diverging variance – something whose density is zero everywhere. In no meaningful sense do the two expressions converge to the same scalar in the relevant limits.

Clearly, BM, whose increments are independent and stationary, is not ergodic. That doesn't make it unmanageable or unpredictable – we know the distribution of BM at any moment in time. But the non-ergodicity has surprising consequences of which we mention one now. We already mentioned that if we plot a Brownian trajectory on a piece of paper it will turn into a straight line for long enough simulation times. This suggests that the randomness of a Brownian trajectory becomes irrelevant under a very natural rescaling. Inspired by this insight let's hazard a guess as to what the time-average of zero-drift BM might be. The simplest form of zero-drift BM starts at zero, $f(0) = 0$ and has variance $\text{var}(f(t)) = t$ (this process is also known as the Wiener process). The process is known to be recurrent – it returns to zero, arbitrarily many times,

[3] Physicists often write $dW = \eta dt$, where $\langle \eta \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$, in which case $\delta(t - t')$ is the Dirac delta function, defined by the integral $\int_{-\infty}^{\infty} f(t)\delta(t - t')dt = f(t')$. Because of its singular nature ($\eta(t)$ does not exist ("is infinite"), only its integral exists) it can be difficult to develop an intuition for this object, and we prefer the dW notation.

with probability one in the long-time limit. We would not be mad to guess that the time average of zero-drift BM,

$$\bar{f} = \lim_{t\to\infty} \frac{1}{t} \int_0^t dt' f(t') \tag{41}$$

will converge to zero with probability one. But we would be wrong. Yes, the process has no drift, and yes it returns to zero infinitely many times, but its time average is not a delta function at zero. It is, instead normally distributed with infinite variance according to the following limit

$$\bar{f} \sim \lim_{t\to\infty} \mathcal{N}(0, t/3). \tag{42}$$

Averaging over time, in this case, does not remove the randomness. A sample trajectory of the time average is shown in Fig. 7. In the literature the process $\frac{1}{t}\int_0^t dt' f(t')$ is known as the "random acceleration process" [2].

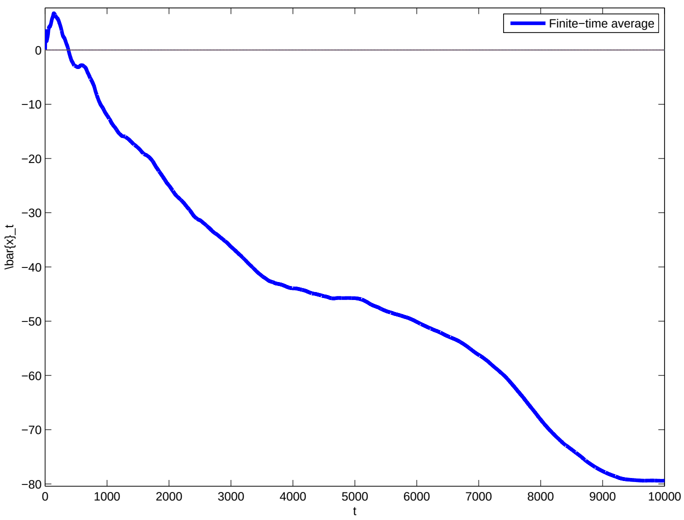


Figure 7: Trajectory of the finite-time average of a zero-drift Brownian motion. The time average does not converge to some value with probability one, but is instead distributed according to $\mathcal{N}(0, t/3)$ for all times. It is the result of integrating a BM; integration is a smoothing operation, and as a consequence the trajectories are smoother than BM (unlike a BM trajectory, they are differentiable). {fig:1_6}

1.4 Geometric Brownian motion {section:Geometric_Browni

DEFINITION: Geometric Brownian motion

If the logarithm of a quantity performs Brownian motion, the quantity itself performs "geometric Brownian motion."

While in the previous chapter $f(x) = \ln(x)$ performed Brownian motion, x itself performed geometric Brownian motion. The change of variable from x to

$f(x) = \ln(x)$ is trivial in a sense but it has interesting consequences. It implies, for instance, that

- $x(t)$ is log-normally distributed
- increments in x are neither stationary nor independent
- $x(t)$ cannot become negative
- the most likely value of x (the mode) does not coincide with the expectation value of x. The median is the same as the mode.

The log-normal distribution is not symmetric, unlike the Gaussian distribution.

Again, it is informative to write geometric Brownian motion as a stochastic differential equation.

$$dx = x(\mu dt + \sigma dW). \quad (43)$$ {eq:GBM_c}

Trajectories for GBM can be simulated using the discretized form

$$\Delta x = x(\mu \Delta t + \sigma \sqrt{\Delta t} \xi_t), \quad (44)$$ {eq:GBM_d}

where $\xi_t \sim \mathcal{N}(0,1)$ are instances of a standard normal variable. In such simulations we must pay attention that the discretization does not lead to negative values of x. This happens if the expression in brackets is smaller than -1 (in which case x changes negatively by more than itself). To avoid negative values we must have $\mu \Delta t + \sigma \sqrt{\Delta t} \xi_t > -1$, or $\xi_t < \frac{1+\mu\Delta t}{\sigma\sqrt{\Delta t}}$. As Δt becomes large it becomes more likely for ξ_t to exceed this value, in which case the simulation fails. But ξ_t is Gaussian distributed, meaning it has thin tails, and choosing a sufficiently small value of Δt makes these failures essentially impossible.

On logarithmic vertical scales, GBM looks like BM, and we've already seen some examples. But it is useful to look at a trajectory of GBM on linear scales to develop an intuition for this important process.

The basic message of the game from Sec. 1.1 is that we may obtain different values for growth rates, depending on how we average – an expectation value is one average, a time average is quite a different thing. The game itself is sometimes called the multiplicative binomial process [8], we thank S. Redner for pointing this out to us. GBM is the continuous version of the multiplicative binomial process, and it shares the basic feature of a difference between the growth rate of the expectation value and time-average growth.

The expectation value is easily computed – the process is not ergodic, but that does not mean we cannot compute its expectation value. We simply take expectations of both sides of (Eq. 43),

$$\begin{aligned} \langle dx \rangle &= \langle x(\mu dt + \sigma dW) \rangle & (45)\\ = d\langle x \rangle &= \langle x \rangle \mu dt. & (46) \end{aligned}$$

This differential equation has the solution

$$\langle x(t) \rangle = x_0 \exp(\mu t), \quad (47)$$

which determines the growth rate of the expectation value as

$$g_{\langle\rangle} = \mu. \quad (48)$$

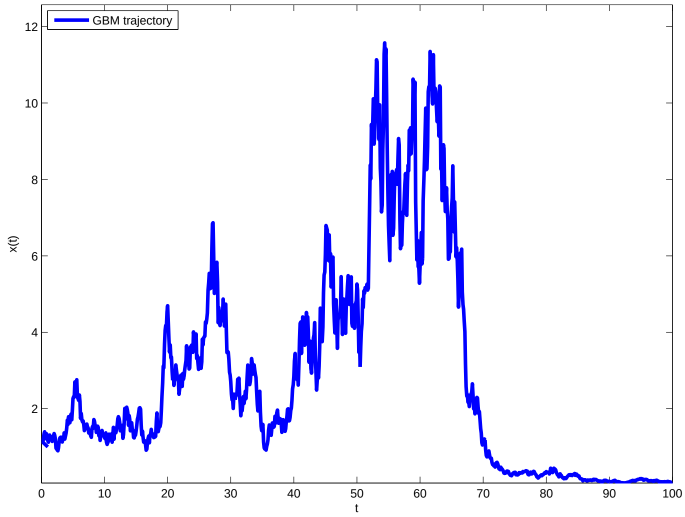


Figure 8: Trajectory of a GBM. The trajectory seems to know about its history – for instance, unlike for BM, it is difficult to recover from a low value of x, and trajectories are likely to get stuck near zero. Occasional excursions are characterised by large fluctuations. Parameters are $\mu = 0.05$ per time unit and $\sigma = \sqrt{2\mu}$, corresponding to zero growth rate in the long run. It would be easy to invent a story to go with this (completely random) trajectory – perhaps something like "things were going well in the beginning but then a massive crash occurred that destroyed morale." {fig:1_7}

As we know this growth rate is different from the growth rate that materializes with probability 1 in the long run. Computing the time-average growth rate is only slightly more complicated. We will follow this plan: consider the discrete process (Eq. 44) and compute the changes in the logarithm of x, then we will let Δt become infinitesimal and arrive at the result for the continuous process. We know $\Delta \ln(x(t))$ to be ergodic, wherefore we will proceed to take its expectation value to compute the time average of the exponential growth rate of the process.

The change in the logarithm of x in a time interval Δt is

$$\begin{aligned} \ln x(t+\Delta t) - \ln x(t) &= \ln[x(1+\mu\Delta t + \sigma\sqrt{\Delta t}\xi_t)] - \ln x(t) && (49)\\ &= \ln x + \ln(1+\mu\Delta t + \sigma\sqrt{\Delta t}\xi_t) - \ln x(t) && (50)\\ &= \ln(1+\mu\Delta t+\sigma\sqrt{\Delta t}\xi_t), && (51) \end{aligned}$$

which we Taylor-expand as $\ln(1 + \text{something small})$ because we will let Δt become small. Expanding to second order,

$$\ln x(t+\Delta t) - \ln x(t) = \mu\Delta t + \sigma\sqrt{\Delta t}\xi_t - \frac{1}{2}\left(\mu\sigma\Delta t^{3/2}\xi_t + \sigma^2\Delta t\xi_t^2\right) + o(\Delta t^2), \quad (52)$$

using "little-o notation" to denote terms that are of order Δt^2 or smaller. Finally, because $\Delta \ln x(t)$ is ergodic, by taking the expectation value of this equation we find the time average of $\Delta \ln x(t)$

$$\langle \ln x(t+\Delta t) - \ln x(t) \rangle = \mu \Delta t - \frac{1}{2}\left(\mu^2 \Delta t^2 + \sigma^2 \Delta t\right) + o(\Delta t^2). \tag{53}$$

Letting Δt become infinitesimal the higher-order terms in Δt vanish, and we find

$$\langle \ln x(t+dt) - \ln x(t) \rangle = \mu dt - \frac{1}{2}\sigma^2 dt \tag{54}$$

so that the time-average growth rate is

$$g_{\text{time}} = \frac{d\,\langle \ln x \rangle}{dt} = \mu - \frac{1}{2}\sigma^2. \tag{55}$$

We have chosen to work with the discrete process here and have arrived at a result that is more commonly shown using Itô's formula. We will not discuss Itô calculus in depth in these notes but we will use some of its results in a later chapter. The above computation should give the reader intuitive confidence that the otherwise surprising results of Itô calculus can be trusted. Why is the result surprising?

Consider the case of no noise $dx = x\mu dt$. Here we can identify $\mu = \frac{1}{x}\frac{dx}{dt}$ as the infinitesimal increment in the logarithm, $\frac{d\ln(x)}{dt}$, using the chain rule of calculus. A naïve application of the chain rule to (Eq. 43) would therefore also yield $\frac{d\langle \ln(x) \rangle}{dx} = \mu$, but the fluctuations in GBM have a non-linear effect, and it turns out that the usual chain rule does not apply. Itô calculus provides a modified chain rule, which leads to the difference $-\frac{\sigma^2}{2}$ between the expectation-value growth rate and the time-average growth rate.

At LML we call this difference the "Weltschmerz" because it is the difference between the many worlds of our dreams and fantasies, and the one cruel reality that the passage of time imposes on us.

References

[1] D. Bernoulli. Specimen Theoriae Novae de Mensura Sortis. Translation "Exposition of a new theory on the measurement of risk" by L. Sommer (1954). *Econometrica*, 22(1):23–36, 1738.

[2] T. W. Burkhardt. The random acceleration process in bounded geometries. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07004, 2007.

[3] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, September 1981.

[4] K. Devlin. *The unfinished game.* Basic Books, 2008.

[5] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, IV(17):549–560, 1905.

[6] J. M. Harrison. *Brownian motion of performance and control.* Cambridge University Press, 2014.

[7] O. Peters and W. Klein. Ergodicity breaking in geometric Brownian motion. *Phys. Rev. Lett.*, 110(10):100603, March 2013.

[8] S. Redner. Random multiplicative processes: An elementary tutorial. *Am. J. Phys.*, 58(3):267–273, March 1990.

[9] W. A. Whitworth. *Choice and chance.* Deighton Bell, 2 edition, 1870.