Economics

A redevelopment of economic theory without parallel universes

Ole Peters and Alexander Adamou

LONDON MATHEMATICAL LABORATORY

2019/10/27 at 15:57:39

Contents

Part I

Tools

Chapter 1

Tools

In this chapter we motivate and introduce the basic mathematical tools we will use. In Sec. **??** *we play a simple coin-toss game and analyze it numerically, by Monte Carlo simulation, and analytically, with pen and paper. The game motivates the introduction of the expectation value and the time average, which in turn lead to a discussion of ergodic properties. As we have seen, the ergodicity question – whether time averages are identical to expectation values – is the key to our redevelopment of formal economics. This is because ergodicity hadn't been established as a concept when the original formalism was developed. The scientific search for stable structures leads to constants in deterministic settings. When randomness is introduced, the role previously played by constants is taken on by ergodic observables. We also introduce the concepts of a random variable, a stochastic process, scalars as representations of transitive preferences, logarithms and exponentials, and dimensional analysis.*

In Sec. 1.9 we notice that wealth on logarithmic scales follows a random walk in our game, and we relate this to Brownian motion, as the continuous-time limit of the random walk. This allows us to introduce Brownian motion and its scaling properties that are robust enough to yield insights into more complicated models.

Finally, we ask in Sec. 1.10 what wealth in our game is doing in the continuum limit but on linear scales. This takes us to geometric Brownian motion, which will be our starting point for much of the rest of these lectures. We derive ensemble-average and time-average growth rates for geometric Brownian motion, by explicitly taking the continuous-time limit, and then state the key result of Itô calculus, (Eq. 1.63) and (Eq. 1.64), which allows an easier derivation of these growth rates and will be relied on in later chapters.

Some historical perspective is provided to understand the prevalence or absence of key concepts in modern economic theory and other fields. The emphasis is on introducing concepts and useful machinery, with applications in later chapters.

1.1 Random variables

In economics, as elsewhere, we are often interested in 'experiments' whose outcomes we do not yet know. Examples are each coin toss in the game in Chap. **??** and the result of a football match. We might know something about the experiment, such as the possible outcomes and that some are more plausible than others, but we are ignorant of the actual outcome. Luckily, we can use probability theory, a branch of mathematics, to build models of our ignorance.

Often the experimental outcomes have, or can be mapped to, numerical values. For example, each coin toss in the game has possible outcomes heads and tails, which correspond to wealth multipliers 1.5 and 0.6. In such cases, we model the uncertain numerical value as a *random variable*. We assume you have seen random variables before and we will not give a lengthy technical account. Instead, we recommend the two-page discussion in [31, p. 2], whose key points we reproduce here.

A random variable, Z, is defined by:

- the set of its possible values; and
- a probability distribution over this set.

The set of values of Z may be continuous, like the interval $(3, 12)$ or the real numbers, $\mathbb{R}$; discrete, like $\{4, 7.8, 29\}$ or the integers, $\mathbb{Z}$; or a combination of the two. The probability distribution is a function which maps values to probabilities. So

$$\mathrm{P}\left[Z = z\right] = p \tag{1.1}$$

means that the outcome $Z = z$ is associated with the probability p. A specific value, z, is sometimes called an 'instance' or 'realisation' of the random variable, Z. While not obligatory, it is a common convention to denote random variables in upper case and realisations in lower case.

Forget, for the moment, what probabilities might mean in the context of an experiment. In purely mathematical terms, they are just real numbers associated with outcomes. The probability distribution has two constraints:

- the probability of any outcome must be non-negative; and
- the probability of an outcome that is certain to happen, *i.e.* one that includes all possible outcomes, must be one.

The latter is a normalisation condition which fixes the scale of the probabilities.

Discrete random variables

When the set of outcomes is discrete, say $\{z_j\}_{j=1}^{M}$, we assign a probability, p_j, to each outcome, z_j, such that

$$\mathrm{P}\left[Z = z\right] = \begin{cases} p_j & z = z_j \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

and

$$\sum_{j=1}^{M} p_j = 1. \tag{1.3}$$

Continuous random variables
Most of the models we will study contain random variables whose possible outcomes form a continuous set. In this case, Z can take uncountably many values, to which we cannot assign non-zero probabilities while maintaining the normalisation condition. Instead, we must assign probabilities to intervals. We define a probability density function (PDF), $\mathcal{P}_Z(z)$, such that the probability of Z being in the interval (a, b) is given by

$$\mathrm{P}\left[a \leq Z \leq b\right] = \int_a^b \mathcal{P}_Z(z) dz. \tag{1.4}$$

The PDF is a non-negative function, $\mathcal{P}_Z(z) \geq 0$, normalised so that the probability of the certain outcome, *i.e.* the integral over all possible outcomes, is one:

$$\int_{-\infty}^{\infty} \mathcal{P}_Z(z) dz = 1. \tag{1.5}$$

Note the difference between subscript and argument: $\mathcal{P}_Z(z)$ is the probability density of the random variable Z at value z. You might find it helpful to think of $\mathcal{P}_Z(z)\delta z$ as the approximate probability of Z being in a small interval $(z, z+\delta z)$ close to z.

Interpretation
So far we have said nothing of the meaning of probabilities: they are simply numbers assigned to outcomes of random variables. One way to interpret probabilities is to imagine many separate experiments, of whose outcomes we have the same ignorance. For example, we can imagine tossing the same coin at many different times, assuming that each coin toss is equally likely to result in heads. Suppose we perform a sample of N experiments and record the number of times, n, that a particular outcome occurs. Under the so-called 'frequentist' interpretation, the appropriate probability to assign to this outcome is its relative frequency, n/N, in the limit $N \to \infty$. In the coin toss example, the probability assigned to heads would be 0.5 if the coin were unbiased. If biased, it would be some other number between 0 and 1.

Note also that time does not appear in the random variable setup. Of course, there is nothing stopping us from using a probability distribution which depends on time or, indeed, any other variable, like the day of the week or the country we are in. Such parametrisations of the random variable do not change fundamentally its mathematical structure: a set of outcomes and associated probabilities. When we consider probability distributions of random variables that do depend on time, such as wealth, $x(t)$, in the coin tossing game, we will make the time dependence explicit. By default we assume random variables are time-independent

1.2 Expectation values

The expectation value is usually introduced as the sum of all possible values, weighted by their probabilities.

Definition: Expectation value

The expectation value of a random variable z that can take discrete values z_j is the sum of all possible values weighted by their probabilities p_j

$$\langle z \rangle = \sum_j p_j z_j. \tag{1.6}$$

If z is continuous, the expectation value is the integral

$$\langle z \rangle = \int_{-\infty}^{+\infty} s\mathcal{P}_Z(s)ds. \tag{1.7}$$

We will define it as a limit instead, and then show that this limit is identical to the familiar expression. This implies that in our first analysis of the game – by averaging over N trajectories – we were approximately using the (time-dependent) expectation value of a time-dependent random variable 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: Ensemble average

The expectation value of a quantity z is the large-ensemble limit of the finite-ensemble average (Eq. **??**),

$$\langle z \rangle = \lim_{N\to\infty} \langle z \rangle_N . \tag{1.8}$$

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

Proof. Consider the number of times the value z_j is observed in an ensemble of N instances. Call this number n_j. The finite-ensemble average can then be re-written as

$$\langle z \rangle_N = \frac{1}{N}\sum_i z_i \tag{1.9}$$

$$= \sum_j \frac{n_j}{N} z_j, \tag{1.10}$$

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

$$\lim_{N\to\infty} \langle z \rangle_N = \sum_j p_j z_j. \tag{1.11}$$

The left-hand side (LHS) is the expectation value by its first definition as a limit, the right-hand side (RHS) is the expectation value by its second definition as a weighted sum. This shows that the two definitions are indeed equivalent.

Here, or earlier, we must give an interpretation of probability as the limiting frequency of an outcome over many trials. □

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 find $x(t) = \langle x(t) \rangle$ because it's the expectation value." Mathematical objects are quite indifferent to the names we give them.

History: The invention of the expectation value

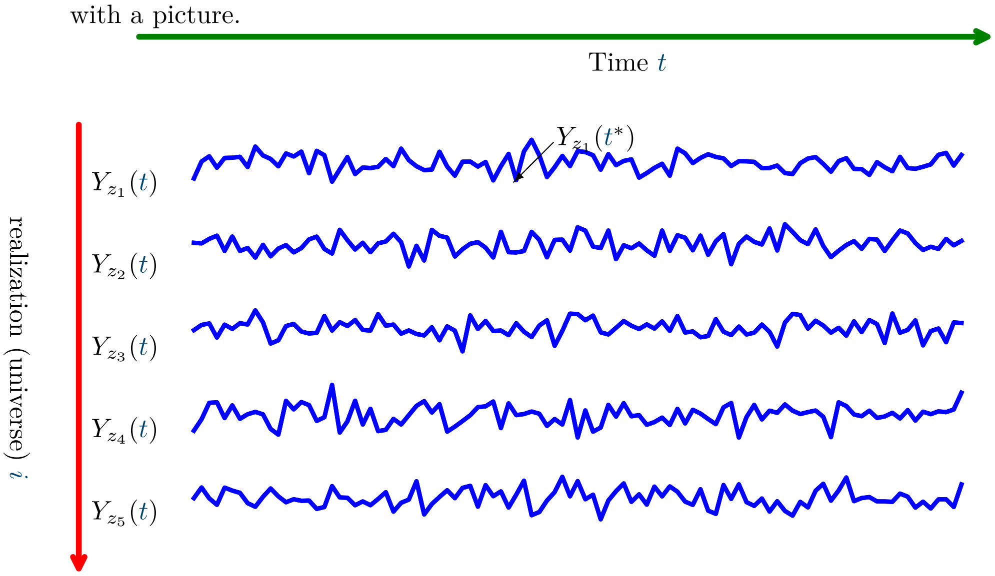
Suppressed.

1.3 Stochastic processes

Once again, we recommend van Kampen [31, p. 52] for a simple definition – this time of stochastic processes. Imagine we've defined a random variable, Z. Any function $Y(Z)$ is then also a random variable. A stochastic process is a special case of such a function, namely one that depends on an additional variable t, a simple scalar parameter, a number, which is interpreted as time, so we write

$$Y_Z(t) = f(Z, t). \tag{1.12}$$

This may not be how you think of a stochastic process, so let's illustrate this with a picture.



When we simulate stochastic processes, we often start with some value and modify it iteratively, for example in each step of a for-loop. In each step we generate a new instance of a random number and thereby construct the trajectory of the stochastic process. In (Eq. 1.12) it's not generated that way. Instead, in this picture, we generate an instance z of the random variable Z only once and insert that into (Eq. 1.12). The value z specifies a simple function of time

$$Y_z(t) = f(z, t) \tag{1.13}$$

meaning that all the randomness is contained in z. Once z is specified, $Y_z(t)$ is specified for all time, and we call it a "realization" or "trajectory" of the stochastic process. Note the use of capital Z for the random variable in (Eq. 1.12) and small z for a realization of it in (Eq. 1.13). As an example you can think of drawing at random a single uniformly distributed real number from the interval $(0, 1)$. With probability 1, this number will be irrational and correspond to an infinite sequence of random decimal digits, which can be interpreted as a stochastic process, where t is given by the decimal place of the digit.

We can also do this: fix a specific time, t^*, and consider the stochastic process at that time, $Y_Z(t^*)$. That's again a random variable, an instance of which may be $Y_{z_1}(t^*)$.

Just as a function of a random variable is another random variable, a function of a stochastic process is another stochastic process. We will often use the noun "observable" to refer to a quantity that is derived from a stochastic process. For example, the growth rate of wealth is an observable of the wealth process.

We will suppress the random variable Z in our notation, and just write $x(t)$ for the stochastic wealth process (instead of writing $x(Z, t)$, *c.f.* (Eq. 1.12)). We will also write $x(t)$ for a specific realization of this process, or $x_i(t)$ when it's important to distinguish different realizations.

1.4 Time averages

An observable that neatly captures the two different aspects of multiplicative growth we have illustrated is the exponential growth rate, $g_{\mathrm{m}}(\langle x(t)\rangle_N, \Delta t)$ observed over finite time Δt, in a finite ensemble of N realisations. Exponential growth rates are ubiquitous and may be familiar, but because they are the origin of the logarithmic function, which will be important for us later on, we will intro them properly in a little excursion that will also clarify what a logarithm is.

Excursion: Compounding growth, exponentials, and the logarithm

Suppressed.

The exponential growth rate of average wealth in an ensemble of N systems, observed over time Δt is

$$g_{\mathrm{m}}(\langle x(t)\rangle_N, \Delta t) = \frac{\Delta \ln \langle x\rangle_N}{\Delta t}, \tag{1.14}$$

where the Δ in the numerator corresponds to the change over the Δt in the denominator. For N and Δt 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 $\frac{1}{\delta t} \ln \langle r\rangle$) is

$$g_{\mathrm{m}}(\langle x\rangle) = \lim_{N\to\infty} g_{\mathrm{m}}, \tag{1.15}$$

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

$$\overline{g} = \lim_{\Delta t\to\infty} g_{\mathrm{m}}. \tag{1.16}$$

We can also write (Eq. 1.14) as a sum of the logarithmic differences in the T individual rounds of the gamble that make up the time interval $\Delta t = T\delta t$

$$g_{\mathrm{m}}(\langle x(t)\rangle_N, \Delta t) = \frac{1}{T\delta t}\sum_{\tau=1}^{T} \Delta \ln \langle x(t+\tau\delta t)\rangle_N . \tag{1.17}$$

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

DEFINITION: **Finite-time average**

The "finite-time average" of the quantity $x(t)$ is

$$\overline{x}_{\Delta t} = \frac{1}{\Delta t}\int_t^{t+\Delta t} x(s)ds. \tag{1.18}$$

If x only changes at $T = \Delta t/\delta t$ discrete times $t+\delta t$, $t+2\delta t$, *etc.*, then this can be written as

$$\overline{x}_{\Delta t} = \frac{1}{T\delta t}\sum_{\tau=1}^{T} x(t+\tau\delta t). \tag{1.19}$$

DEFINITION: **Time average**

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

$$\overline{x} = \lim_{\Delta t\to\infty} \overline{x}_{\Delta t}. \tag{1.20}$$

According to this definition, $\overline{g}$ is the time average of the observable $\frac{\delta \ln x}{\delta t}$. It can be shown that the time-average growth rate of a single trajectory is the same as that of a finite-ensemble average of trajectories, $\lim_{\Delta t\to\infty}\frac{\Delta \ln x}{\Delta t} = \lim_{\Delta t\to\infty}\frac{\Delta \ln\langle x\rangle_N}{\Delta t}$, [26]. In Sec. **??** we will derive this result as well as growth rates in finite ensembles and finite time.

Excursion: Dimensional analysis

Suppressed.

1.5 The game – revisited

We pretended to be mathematically clueless when we ran the simulations, with the purpose to gain a deeper conceptual understanding of the expectation value. We now compute exactly the expectation value of the stochastic process $x(t)$, instead of approximating it numerically. Consider the expectation value of (Eq. **??**)

$$\langle x(t+\delta t)\rangle = \langle x(t) r(t+\delta t)\rangle . \tag{1.21}$$

We've just learned what to call objects like $r(t)$: it's another stochastic process, or an observable. This one is especially simple: in a given realization $x(t)$ it's one instance of the same random variable for each time t. one, namely one that is ergodic. We note here that its ensemble average is time-independent (and in Sec. 1.6 we will see that it's an example of an ergodic observable). Since $r(t+\delta t)$ is independent of $x(t)$, (Eq. 1.21) can be re-written as

$$\langle x(t+\delta t)\rangle = \langle x(t)\rangle \langle r\rangle . \tag{1.22}$$

Therefore, we can solve recursively for the wealth after T rounds, corresponding to a playing time of $\Delta t = T\delta t$:

$$\langle x(t+\Delta t)\rangle = \langle x(t+T\delta t)\rangle = x(t) \langle r\rangle^{T} . \tag{1.23}$$

δt is the duration of a single round of a gamble, while Δt is the amount of time spent gambling.

The expectation value $\langle r\rangle$ is easily found from (Eq. **??**) 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 by a factor 1.05 each 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 1.1 compares the analytical result for the infinite ensemble to the numerical results of Fig. **??** for finite ensembles.

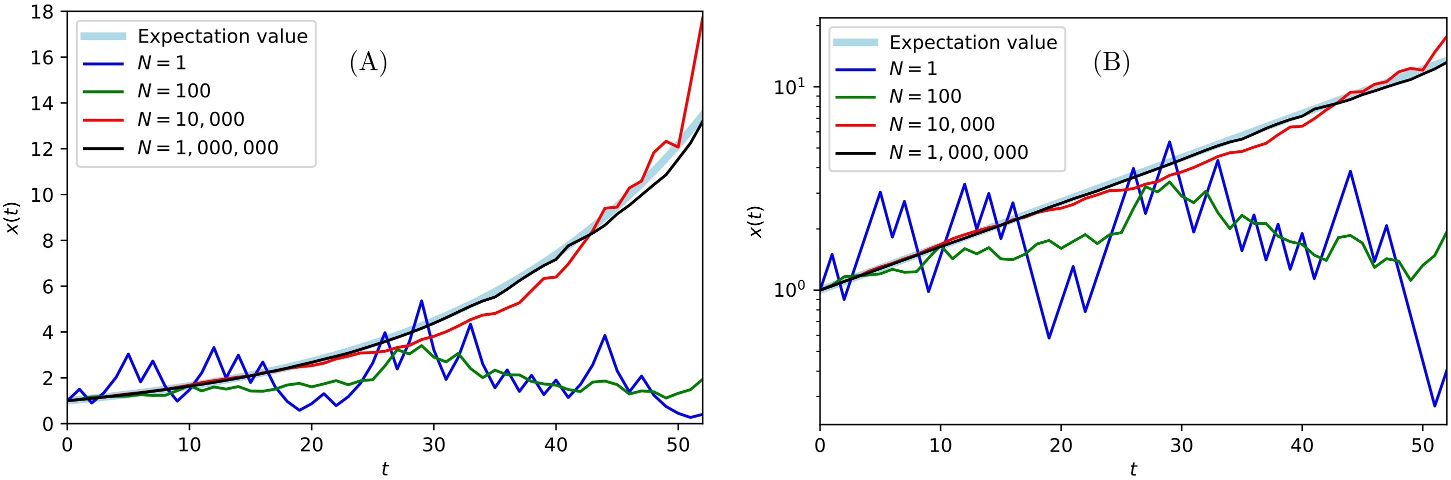


Figure 1.1: Expectation value (thick light blue line) and finite-ensemble averages. (A) linear scales, (B) logarithmic scales.

In this section we validate Fig. **??** and compute analytically what happens in the long-time limit. The blue line in Fig. **??** is not completely smooth, there's still some noise (see panel B). It has some average slope, but that slope will vary from realisation to realisation. The longer we observe the system, *i.e.* the more time is represented in a figure like Fig. **??**, the smoother the line will be. In the long-time limit, $\Delta t \to \infty$, 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 $t + \Delta t$, where $\Delta t = T\delta t$ as before, is

$$x(t+\Delta t) = x(t) \prod_{\tau=1}^{T} r(t+\tau\delta t), \tag{1.24}$$

with the dummy variable τ indicating the round of the game. We can split this into two products, one for each possible value of $r(t)$, which we call r_1 and r_2, *i.e.*

$$r(t) = \begin{cases} r_1 & \text{with probability } p_1 \\ r_2 & \text{with probability } p_2 = 1 - p_1. \end{cases} \tag{1.25}$$

Let's denote the number of occurrences of r_1 by n_1 and of r_2 by n_2, so that

$$x(t+\Delta t) = x(t) r_1^{n_1} r_2^{n_2}. \tag{1.26}$$

We denote by $\overline{r}$ the effective factor by which $x(t)$ is multiplied per round when the change is computed over a long time, *i.e.* $x(t+\Delta t) \sim x(t)(\overline{r})^T$ as $\Delta t \to \infty$. This quantity is found by taking the T^{th} root of $\frac{x(t+\Delta t)}{x(t)}$ and considering its long-time limit:

$$\overline{r} = \lim_{\Delta t \to \infty} \left(\frac{x(t+\Delta t)}{x(t)} \right)^{1/T} \tag{1.27}$$

$$= \lim_{T \to \infty} r_1^{n_1/T} r_2^{n_2/T}. \tag{1.28}$$

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

$$\lim_{T \to \infty} \left(\frac{x(t+T\delta t)}{x(t)} \right)^{1/T} = (r_1 r_2)^{1/2}, \tag{1.29}$$

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. **??** was not a fluke: *every* trajectory will decay in the long run at a rate of $(r_1 r_2)^{1/2}$ per round.

Figure 1.2 (B) compares the trajectory generated in Fig. **??** to a trajectory decaying exactly at rate $\overline{r}$ and places it next to the average over a million systems.

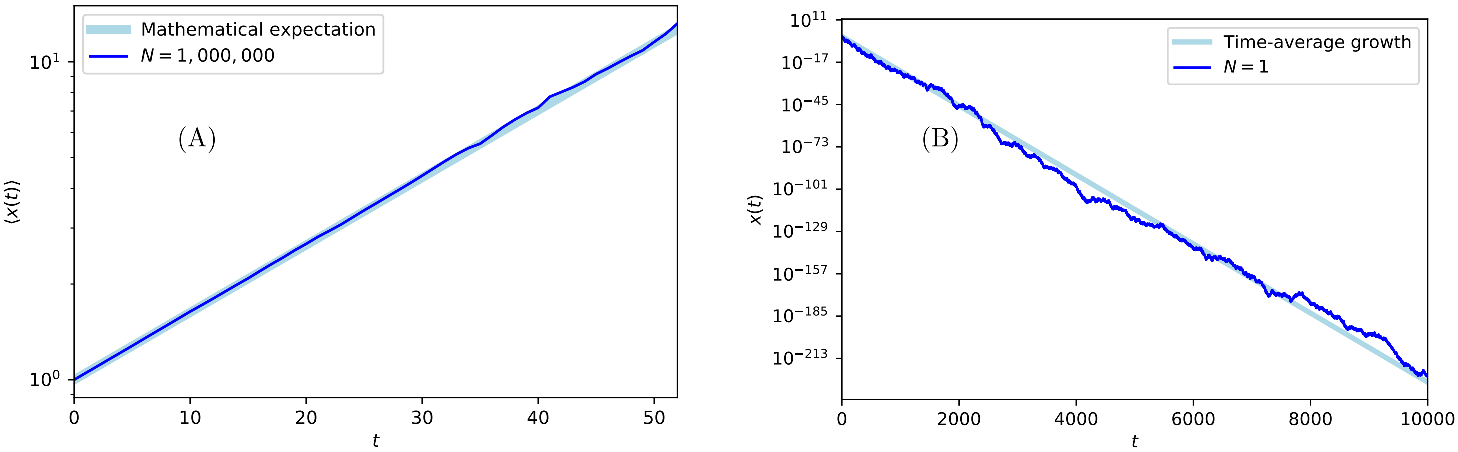


Figure 1.2: (A) Finite-ensemble average for $N = 10^6$ and 52 time steps, the light blue line is the expectation value. (B) A single system simulated for 10,000 time steps, the light blue line decays exponentially with the time-average decay factor $\overline{r}$ in each time step.

Excursion: Scalars

Suppressed.

There are two averages, $r_{\langle\rangle}$ and $\overline{r}$ 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?"

History: William Allen Whitworth

Suppressed.

1.6 Ergodicity

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 called "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. These lecture notes are our attempt to use precisely the insights of these physicists to re-develop economic theory from the foundations up.

History: Randomness and ergodicity in physics

Suppressed.

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

DEFINITION: **Ergodic property**

In these notes, an observable A is called ergodic if its expectation value is constant in time and its time average converges to this value with probability one[a]

$$\lim_{\Delta t\to\infty}\frac{1}{\Delta t}\int_t^{t+\Delta t}A(s)ds=\lim_{N\to\infty}\frac{1}{N}\sum_i^N A_i(t). \tag{1.30}$$

[a]Some researchers would call A "mean ergodic" and require further observables derived from it to be (mean) ergodic in order to call A "wide-sense ergodic." This extra nomenclature is not necessary for our work, but we leave a footnote here to avoid confusion.

The RHS of (Eq. 1.30) is evaluated at time t, and unlike the LHS could be a function of time. For now, we restrict our definition of ergodicity to a setup where that is not the case, *i.e.* where the ergodic property holds at all times. In Sec. **??** we will discuss transient behavior, where the distribution of A is time dependent. We then also consider an observable "ergodic" if its expectation value only converges to the time-average in the $t\to\infty$ limit.

In terms of random variables, Z, and stochastic processes, $Y_Z(t)$, the ergodic property can be visualized as in Fig. 1.3. Averaging a stochastic process over time or over the ensemble are completely different operations, and only under very rare circumstances (namely under ergodicity) can the two operations be interchanged. In our coin-tossing game the operations are clearly not interchangeable. An implicit assumption of interchangeability in the early days is the Original Sin of economic theory.

We stress that in a given setup, some observables may have the ergodic property even if others do not. Language therefore must be used carefully. Saying

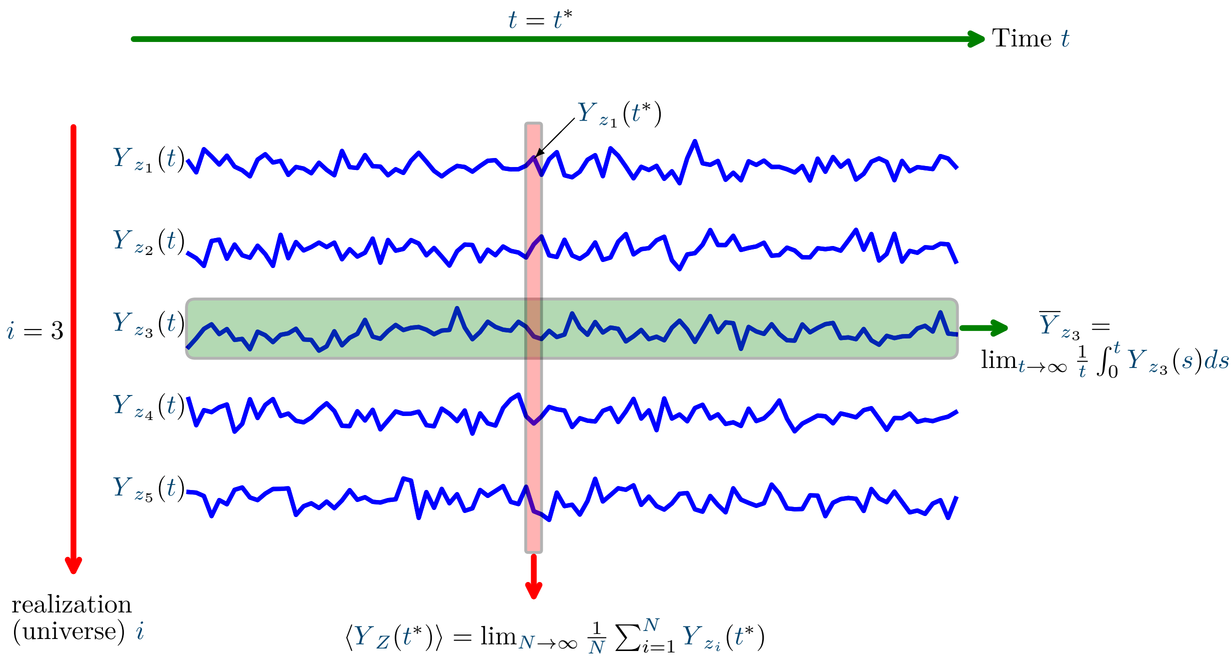


Figure 1.3: Extending the figure on p. 7, averaging over time means averaging along one trajectory from left to right; averaging over the ensemble means averaging at a fixed time across different trajectories from top to bottom.

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. **??**), is clearly not ergodic – with $A = x$ the LHS of (Eq. 1.30) is zero, and the RHS is not constant in time but grows. The expectation value $\langle x\rangle(t)$ simply doesn't give us the relevant information 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 tell 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 per-round 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 for the wealth dynamic defined by (Eq. **??**) and (Eq. **??**).

Proof. According to (Eq. 1.8) and (Eq. **??**), the expectation value of $r(t)$ is

$$\langle r\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{i}^{N} r_i, \tag{1.31}$$

and, according to (Eq. 1.19), the time average of $r(t)$ is

$$\overline{r} = \lim_{T\to\infty} \frac{1}{T} \sum_{\tau}^{T} r_\tau, \tag{1.32}$$

where we have written $r_\tau = r(t+\tau\delta t)$ to make clear the equivalence between the two expressions. The only difference is between the labels we have chosen for the dummy variable (i in (Eq. 1.31) and τ in (Eq. 1.32)). Clearly, the expressions yield the same value.

The same argument holds for $\delta \ln x$. □

Whether we consider (Eq. 1.32) an average over time or over an ensemble is only a matter of our choice of words.

The expectation value $\langle \delta \ln x \rangle$ is important, historically. Daniel Bernoulli noticed in 1738 [2] 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: multiplicative dynamics are a fairly realistic model for real wealth, and under those dynamics 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, and $\langle \delta \ln x \rangle$ does tell us what happens to x over time, wherefore seeing people optimise $\langle \delta \ln x \rangle$ just means seeing them optimise wealth over the one trajectory that describes a financial life, rather than across the ensemble of possibilities.

Ergodicity is not the same concept as stationarity. As an illustration of the difference, consider the following process: $f(t) = z_i$, where z_i is an instance of a random variable Z. Explicitly, this means a realisation of the stochastic process $f(t)$ is generated as follows: we generate the random instance z_i once, and then fix $f(t)$ at that value for all time. The distribution of $f(t)$ is independent of t and in that sense $f(t)$ is stationary. But it is not ergodic: averaging over the ensemble, we obtain $\langle f(t) \rangle = \langle z \rangle$, whereas averaging over time in the i^{th} trajectory gives $\overline{f} = z_i$. Thus the process is stationary but not ergodic.

1.7 Changes and stability

Deleted – the material now appears in the next chapter. However, we must say enough about rates to consider the growth rates of Brownian motion (BM) and geometric Brownian motion (GBM).

1.8 Normal distribution

1.9 Brownian motion

We motivate the model called BM as a limiting process, the continuous-time limit, that arises from random walks. In the previous section we established that the discrete increments of the logarithm of x, which we called v, are instances of a time-independent random variable in our game. A quantity making such random steps over time is said to perform a "random walk." Indeed, the blue line for a single system in Fig. **??** (B) shows 52 steps of a random walk trajectory. Random walks come in many forms – in all of them v changes discontinuously by an amount δv drawn from a time-independent distribution, over time intervals which may be regular or which may be drawn from a time-independent distribution themselves.

We are interested only in the simple case where v changes at regular intervals, $\delta t, 2\delta t, \ldots$. For the distribution of increments we only insist on the existence of the variance, meaning we insist that $\mathrm{var}(\delta v) = \langle \delta v^2 \rangle - \langle \delta v \rangle^2$ be finite. Increments whose distributions are heavier-tailed do not lead to BM (BM has continuous paths, and that continuity is broken by such increments).

The change in v after a long time is the sum of many independent increments,

$$v(t + T\delta t) - v(t) = \sum_i^T \delta v_i. \tag{1.33}$$

The Gaussian central limit theorem tells us that such a sum will become Gaussian-distributed as we add more terms to the sum and re-scale it appropriately, namely so as to keep the width of the distribution finite and remove any systematic drift,

$$\lim_{T\to\infty} \overbrace{\frac{1}{\sqrt{T}}}^{\text{keep width finite}} \sum_i^T (\delta v_i \underbrace{- \langle \delta v \rangle}_{\text{remove systematic drift}}) \sim \mathcal{N}(0, \mathrm{var}(\delta v)), \tag{1.34}$$

where we call $\frac{\langle \delta v \rangle}{\delta t}$ the "drift term." The notation $\sim \mathcal{N}(0, \mathrm{var}(\delta v))$ is short-hand for "is Gaussian distributed, with mean 0 and variance $\mathrm{var}(\delta v)$." 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.

Let's also ask about the re-scaling that was applied in (Eq. 1.34). Scaling properties are very robust, and especially the scaling of random walks for long times will be useful to us.

We work with the simplest setup: at time zero we start at zero, $v(0) = 0$, and in each time step, we either increase or decrease v by 1, with probability 1/2. To avoid notation clutter, we'll set the duration of a time step to $\delta t = 1$, so that T is both the number of steps and the time.

We are interested in the variance of the distribution of v as T increases, which we obtain by computing the first and second moments of the distribution.

The first moment (the expectation value) of v is $\langle v \rangle (T) = 0$, by symmetry for all times.

We obtain the second moment by induction[1]: Whatever the second moment, $\langle v(T)^2 \rangle$, is at time T, we can write down its value at time $T + 1$ as

$$\langle v(T+1)^2 \rangle = \frac{1}{2} \left[\langle (v(T)+1)^2 \rangle + \langle (v(T)-1)^2 \rangle \right] \tag{1.35}$$

$$= \frac{1}{2} \left[\langle v(T)^2 + 1 + 2v(T) \rangle + \langle (v(T)^2 + 1 - 2v(T)) \rangle \right] \tag{1.36}$$

$$= \langle v(T)^2 \rangle + 1. \tag{1.37}$$

[1] The argument is nicely illustrated in [10, Volume 1, Chapter 6-4], where we first came across it.

In addition, we know the initial value of $v(0) = 0$. By induction it follows that the second moment is

$$\left\langle v(T)^2 \right\rangle = T \tag{1.38}$$

and, since the first moment is zero, the variance is

$$\text{var}(v(T)) = T. \tag{1.39}$$

The standard deviation – the width of the distribution – of changes in a quantity following a random walk thus scales as the square-root of the number of steps that have been taken, $\sqrt{T}$.

This square-root behaviour leads to many interesting results. It can make averages stable (because $\sqrt{T}/T$ converges to zero for large T), and sums unstable (because $\sqrt{T}$ diverges for large T). Consequently, we may expect that as the size of some system increases, some properties become stable and others unstable.

Imagine simulating a single long trajectory of v and plotting it on paper[2]. 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 $\langle \delta v \rangle \neq 0$ then the amount of variation in v that has to be represented by a fixed amount of paper also increases linearly with the simulated time. However, the departures of Δv from its expectation value $T \langle \delta v \rangle$ 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 – finally – is BM.

We think of BM as the limit of a random walk where we shorten the duration of a step $\delta t \to 0$, and scale the width of an individual step so as to maintain the random-walk scaling of the variance, meaning $|\delta v| = \sqrt{\delta t}$. In the limit $\delta t \to 0$, this implies that the local slope of a BM trajectory diverges, $\frac{\delta v}{\delta t} \to \infty$. This means that BM trajectories are infinitely jagged, or – in mathematical terms – they are not differentiable. However, the way in which they become non-differentiable, through the $\sqrt{\delta t}$ factor, just leaves the trajectories continuous (this isn't the case for $|\delta v| = \delta t^{\alpha}$, where α is less than 0.5).

Continuity of v means that it is possible to make the difference $|v(t) - v(t + \epsilon)|$ arbitrarily small by choosing ϵ sufficiently small. Trajectories (of non-BM processes) that don't have this property contain what are appropriately called "jumps." Continuity therefore means that there are no jumps. These subtleties make BM a topic of great mathematical interest, and many books have been written about it. We will pick from these books only what is immediately useful to us. To convey the universality of BM we define it formally as follows:

[2] This argument is inspired by a colloquium presented by Wendelin Werner in the mathematics department of Imperial College London in January 2012. Werner started the colloquium with a slide that showed a straight horizontal line and asked: what is this? Then answered that it was the trajectory of a random walk, with the vertical and horizontal axes scaled equally.

Definition: Brownian motion i

If a stochastic process has continuous paths, stationary independent increments, and is distributed according to $\mathcal{N}(\mu t, \sigma^2 t)$ 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 [11]: *"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 that produce random trajectories.

The power of BM lies in its simplicity and analytic tractability, involving only two parameters, μ and σ. We will often work with its representation as a stochastic differential equation (SDE)

$$dv = \mu dt + \sigma dW \tag{1.40}$$

where dW is the so-called "Wiener increment," the beating heart of many SDEs. The Wiener increment can be defined by two properties: its distribution and its auto-correlation,

$$dW \sim \mathcal{N}(0, dt) \tag{1.41}$$
$$\langle dW(t)dW(t')\rangle = dt\ \delta(t, t'), \tag{1.42}$$

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 BM paths can be constructed from a discretized version of (Eq. 1.40)

$$v(t + \delta t) = v(t) + \mu\delta t + \sigma\sqrt{\delta t}\xi_t, \tag{1.43}$$

where ξ_t are instances of a standard normal distribution ($\mathcal{N}(0, 1)$).

BM itself is not a time-independent random variable – it is a non-ergodic stochastic process. This is easily seen by comparing expectation value and time average. We start with the expressions (stated without proof here) for the

[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.

finite-ensemble average and the finite-time average of BM. The finite-ensemble average (easy to derive) is distributed as

$$\langle v \rangle_N \sim \mu t + \mathcal{N}(0, t/N), \tag{1.44}$$

and the finite-time average (a little harder to derive) of a single BM trajectory is distributed as

$$\bar{v}_t \sim \mu t/2 + \sigma \mathcal{N}(0, t/3). \tag{1.45}$$

The expectation value, *i.e.* the limit $N \to \infty$ of (Eq. 1.44), converges to μt with probability one, so it depends on time, and it's unclear how to compare that to a time average (which cannot depend on time). Its limit $t \to \infty$ does not exist.

The time average, the limit $t \to \infty$ of (Eq. 1.45) 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 ergodic, is itself not ergodic. However, 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 with non-zero drift 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, $v(0) = 0$ and has variance $\mathrm{var}(v(t)) = t$ (this process is also known as the "Wiener process"). The process is known to be recurrent – it returns to zero, arbitrarily many times, with probability one in the long-time limit. We would not be mad to guess that the time average of zero-drift BM,

$$\bar{v} = \lim_{t \to \infty} \frac{1}{t} \int_0^t dt' v(t'), \tag{1.46}$$

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

$$\bar{v} \sim \lim_{t \to \infty} \mathcal{N}(0, t/3). \tag{1.47}$$

Averaging over time, in this case, does not remove the randomness. A sample trajectory of the finite-time average (not of BM but of the average over a BM) is shown in Fig. 1.4. In the literature this process, $\frac{1}{t}\int_0^t dt' v(t')$, is known as the "random acceleration process" [3].

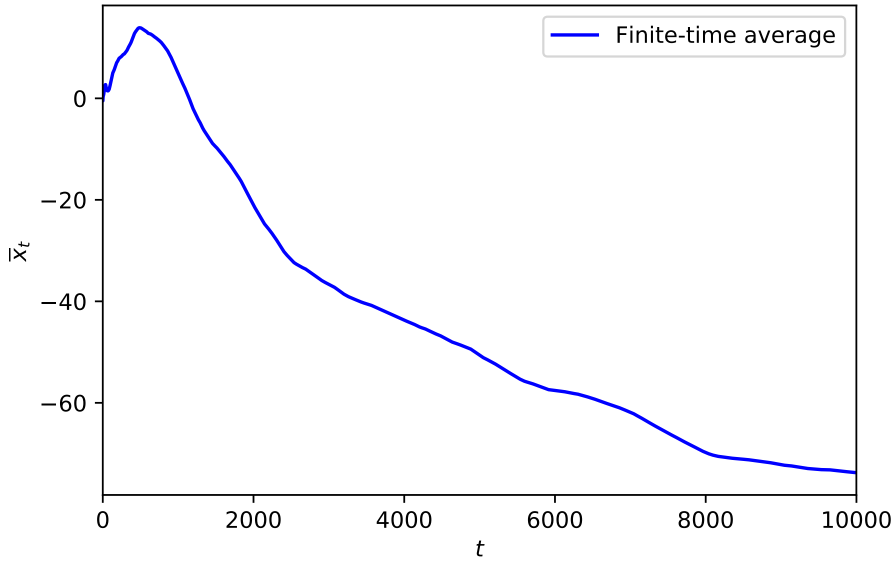


Figure 1.4: Trajectory of the finite-time average of a zero-drift BM. The process is not ergodic: the time average does not converge to a number, but is instead distributed according to $\mathcal{N}(0, t/3)$ for all times, while the expectation value is zero. It is the result of integrating a BM; integration is a smoothing operation, and as a consequence the trajectories are smoother than BM (unlike a BM trajectory, they are differentiable).

1.10 Geometric Brownian motion

Definition: **Geometric Brownian motion**

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

While in Sec. 1.9 $v(x) = \ln(x)$ performed BM, x itself performed GBM. The change of variable from x to $v(x) = \ln(x)$ 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.

These and other properties of the log-normal distribution will be discussed in detail in Sec. **??**.

Again, it is informative to write GBM as a stochastic differential equation.

$$dx = x(\mu dt + \sigma dW). \tag{1.48}$$

Similarly to BM, trajectories for GBM can be simulated using the discretized form (*cf.* (Eq. 1.43))

$$\delta x = x(\mu \delta t + \sigma \sqrt{\delta t} \xi_t), \tag{1.49}$$

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 in (Eq. 1.49) 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.

GBM on logarithmic vertical scales looks like BM on linear vertical scales. Figure **??** is, in fact, and example of a very coarse discretisation of GBM. But it's useful to look at a more finely discretised trajectory of GBM on linear scales to develop an intuition for this important process.

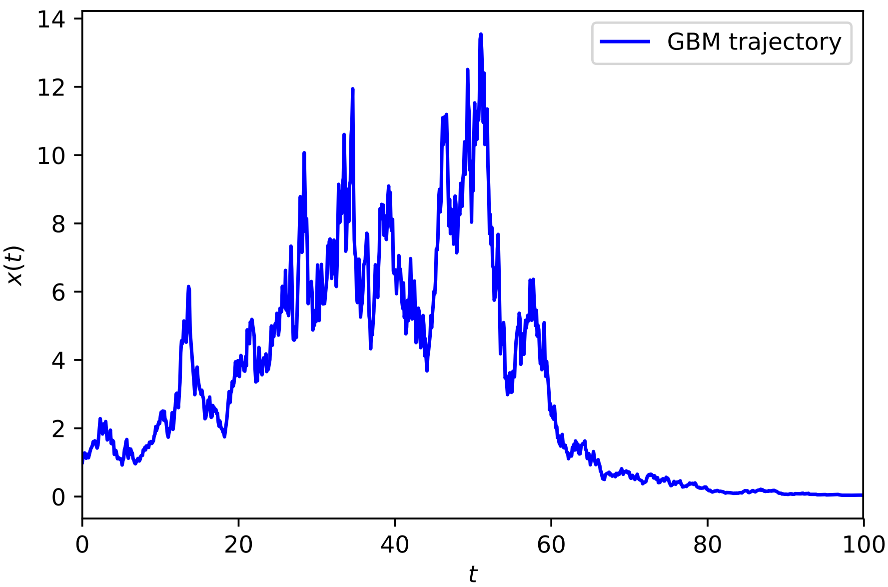


Figure 1.5: Trajectory of a GBM. What happens to the trajectory tomorrow depends strongly on where it is today – for instance, unlike for BM, it is difficult to recover from a low value of x, and trajectories are likely to get stuck near zero. Occasional excursions are characterised by large fluctuations. 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."

The basic message of the game from Sec. **??** 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 another. The game itself is sometimes called the multiplicative binomial process [27], 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 the expectations values of both sides of (Eq. 1.48) to get

$$\begin{aligned} \langle dx \rangle &= \langle x(\mu dt + \sigma dW) \rangle && (1.50) \\ = d\langle x \rangle &= \langle x \rangle \mu dt. && (1.51) \end{aligned}$$

This differential equation has the solution

$$\langle x(t) \rangle = x(t_0)\exp(\mu t), \tag{1.52}$$

which determines the growth rate of the expectation value as

$$g_{\mathrm{m}}(\langle x \rangle) = \mu. \tag{1.53}$$

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, and it will get even simpler once we've introduced Itô calculus in Sec. 1.11. But for now we will follow this plan: consider the discrete process (Eq. 1.49) 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 and reflective of performance over time, 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

$$\begin{aligned} \ln x(t+\delta t) - \ln x(t) &= \ln[x(1+\mu\delta t + \sigma\sqrt{\delta t}\xi_t)] - \ln x(t) && (1.54) \\ &= \ln x + \ln(1+\mu\delta t + \sigma\sqrt{\delta t}\xi_t) - \ln x(t) && (1.55) \\ &= \ln(1+\mu\delta t + \sigma\sqrt{\delta t}\xi_t), && (1.56) \end{aligned}$$

which we Taylor-expand as ln(1+something small) because we will let δt become small. Expanding to second order,

$$\ln x(t+\delta t) - \ln x(t) = \mu\delta t + \sigma\sqrt{\delta t}\xi_t - \frac{1}{2}\left(\mu\sigma\delta t^{3/2}\xi_t + \sigma^2\delta t\xi_t^2\right) + o(\delta t^2), \tag{1.57}$$

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 $\delta \ln x(t)$

$$\langle \ln x(t+\delta t) - \ln x(t) \rangle = \mu\delta t - \frac{1}{2}\left(\mu^2\delta t^2 + \sigma^2\delta t\right) + o(\delta t^2). \tag{1.58}$$

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

$$\langle \ln x(t+dt) - \ln x(t) \rangle = \mu dt - \frac{1}{2}\sigma^2 dt, \tag{1.59}$$

so that the time-average growth rate is

$$\overline{g} = \frac{d\langle \ln x \rangle}{dt} = \mu - \frac{1}{2}\sigma^2. \tag{1.60}$$

The non-ergodicity of GBM leads to a difference between the behaviour of the expectation value (which grows at $g_{\mathrm{m}}(\langle x \rangle)$) and the long-time behaviour of any given trajectory (which grows at $\overline{g}$). Because people experience their wealth over time (which may be described by GBM) and have not access to the ensemble of other possible trajectories, they quite reasonably behave closer to optimising $g_{\mathrm{m}}(\langle x \rangle)$ than to $\overline{g}$.

We could have guessed the result by combining Whitworth's argument on the disadvantage of gambling with the scaling of BM. Let's re-write the factor $1-\epsilon$ in (Eq. **??**) as $1-\sigma\sqrt{\delta t}$. According to the scaling of the variance in a random walk, (Eq. 1.39), this would be a good coarse-graining of some faster process (with shorter time step) underlying Whitworth's game. To find out what happens over one single time step we take the square root of (Eq. **??**),

$$[(1+\sigma\sqrt{\delta t})(1-\sigma\sqrt{\delta t})]^{1/2} = [1-\sigma^2\delta t]^{1/2}. \tag{1.61}$$

Letting δt become infinitesimally small, we replace δt by dt, and the first-order term of a Taylor-expansion becomes exact,

$$[(1+\sigma\sqrt{\delta t})(1-\sigma\sqrt{\delta t})]^{1/2} \to 1-\frac{\sigma^2}{2}dt, \tag{1.62}$$

in agreement with (Eq. 1.60) if the drift term $\mu = 0$, as assumed by Whitworth.

1.11 Itô calculus

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 but we will use some of its results. The key insight of Itô was that the non-differentiability of so-called Itô processes leads to a new form of calculus, where in particular the chain rule of ordinary calculus is replaced. An Itô process is a SDE of the following form

$$dx = a(x,t)dt + b(x,t)dW. \tag{1.63}$$

If we are interested in the behaviour of some other quantity that is a function of x, let's say $v(x)$, then Itô's formula tells us how to derive the relevant SDE as follows:

$$dv = \left(\frac{\partial v}{\partial t} + a(x,t)\frac{\partial v}{\partial x} + \frac{b(x,t)^2}{2}\frac{\partial^2 v}{\partial x^2}\right)dt + b(x,t)\frac{\partial v}{\partial x}dW. \tag{1.64}$$

Derivations of this formula can be found on Wikipedia. Intuitive derivations, such as [13], use the scaling of the variance, (Eq. 1.39), and more formal derivations, along the lines of [11], rely on integrals. We simply accept (Eq. 1.64) as given. It makes it very easy to re-derive (Eq. 1.60), which we leave as an exercise: use (Eq. 1.64) to find the SDE for $\ln(x)$, take its expectation value and differentiate with respect to t. We will use (Eq. 1.64) in Sec. **??**. The above computations are intended to give the reader intuitive confidence that Itô calculus can be trusted[4]. We find that, though phrased in different words, our key

[4]Itô calculus is one way of interpreting the non-differentiability of dW. Another interpretation is due to Stratonovich, which is not strictly equivalent. However, the key property of GBM that we make extensive use of is the difference between the growth rate of the expectation value, $g_{\mathrm{m}}(\langle x \rangle)$, and the time-average growth rate, $\overline{g}$. This difference is the same in the Stratonovich and the Itô interpretation, and all our results hold in both cases.

insight – that *the growth rate of the expectation value is not the time-average growth rate* – has appeared in the literature not only in 1870 but also in 1944. And in 1956 [16], and in 1966 [30], and in 1991 [5], and at many other times. Yet the depth of this insight remained unprobed.

Equation (1.60), which agrees with Itô calculus, may be 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 ordinary calculus. A naïve application of the chain rule to (Eq. 1.48) 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 is a modified chain rule, (Eq. 1.64), which leads to the difference $-\frac{\sigma^2}{2}$ between the expectation-value growth rate and the time-average growth rate.

This difference is sometimes called the "spurious drift", but at the London Mathematical Laboratory (LML) we call it 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.

Summary of Chap. ??

Suppressed.

Part II

Microeconomics

Chapter 2

Decisions in a riskless world

Decision theory is a cornerstone of formal economics. As the name suggests, it models how people make decisions. In this chapter we will generalise and formalise the treatment of the coin tossing game to introduce our approach to decision theory. Our central axiom will be that people attempt to maximize the rate at which wealth grows when averaged over time. This is a surprisingly powerful idea. In many cases it eliminates the need for well established but epistemologically troublesome techniques, such as utility functions.

2.1 Models and science fiction

We will do decision theory by using mathematical models, and since this can be done in many ways we will be explicit about how we choose to do it. We will define a wealth process – a model of how wealth changes with time – and a decision criterion. The wealth process and the decision criterion may or may not remind you of the real world. We will not worry too much about the accuracy of these reminiscences. Instead we will "shut up and calculate" – we will let the mathematical model create its world. Writing down a mathematical model is like laying out the premise for a science-fiction novel. We may decide that people can download their consciousness onto a computer, that medicine has advanced to eliminate ageing and death – these are premises we are at liberty to invent. Once we have written them down we begin to explore the world that results from those premises. A decision criterion is really a model of human behaviour – what makes us who we are if not our decisions? It therefore implies a long list of specific behaviours that will be observed in a given model world. For example, some criteria will lead to cooperation, others will not, some will lead to the existence of insurance contracts, others will not *etc.* We will explore the worlds created by the different models. Once we have done so we invite you to judge which model you find most useful for your understanding of the world. Of course, having spent many years thinking about these issues we have come to our own conclusions, and we will put them forward because we believe them to be helpful.

To keep the discussion to a manageable volume we will only consider a setup that corresponds to making purely financial decisions. We may bet on a horse or take out personal liability insurance. This chapter will not tell you whom you should marry or even whose economics lectures you should attend.

2.2 The decision axiom

A "decision theory" is a model of human behaviour. We will write down such a model phrased as the following simple axiom:

Decision axiom

People optimize the growth rate of their wealth.

Without discussing why people might do this, let's step into the world created by this axiom. To do that, we need to be crystal clear about what a growth rate is, so we'll discuss that first, in Sec. 2.3. Traditionally, decision theory deals with an uncertain future: we have to decide on a course of action now although we don't know with certainty what will happen to us in the future under any of our choices. We will systematically work our way towards this setup, beginning with trivial decisions where neither time nor uncertainty matters Sec. 2.4.1, next introducing time Sec. 2.4.2 (where we will shed light on what's called "discounting"). In the next chapter we will introduce uncertainty, see Sec. **??** (where we will shed light on what's called "expected utility theory").

2.3 Growth rates

You may have wondered why both

$$g_a = \frac{x(t+\Delta t) - x(t)}{\Delta t} \tag{2.1}$$

and

$$g_e = \frac{\ln x(t+\Delta t) - \ln x(t)}{\Delta t} \tag{2.2}$$

are sometimes called a growth rate – they're different objects, why the same name? By the end of this section, the answer to this question should be clear.

When we say that $x(t)$ is a growth process, we mean that it is a monotonic function of t. If you're thinking about randomness – don't, we'll come to that later. For now, we will just work with a deterministic function $x(t)$ – we even use an unusual font to indicate that this is a deterministic function.

For a given process, the appropriate growth rate, g, solves the following problem for us: how do we characterise how fast x grows? A growth rate is a mathematical object of the form

$$g = \frac{\Delta v(x)}{\Delta t}, \tag{2.3}$$

were $v(x)$ is a monotonically increasing function of wealth x. Comparing to (Eq. 2.1) and (Eq. 2.2), we find that $v(x) = x$ for additive dynamics and $v(x) = \ln x$ for multiplicative dynamics. The transformation $x \to v(x)$ ensures temporal stability of g.

How does this work, and what does it mean? Specifically,

1. why the transformation?
2. how do we know which transformation to use?

We will start with the mathematically simple case of the additive growth rate, (Eq. 2.1), and discuss its properties by applying it to additive growth. Next we will ask under what conditions the exponential growth rate, (Eq. 2.2), is appropriate, and that will lead us to the general growth rate, (Eq. 2.3).

2.3.1 Additive growth rate

If I want to know how fast $x(t)$ grows, the most obvious thing to compute is its rate of change – that's the additive growth rate, (Eq. 2.1). This tells me by how much x grows in the interval $[t, t+\Delta t]$. If $x(t)$ is linear in t, so that

$$x(t) = x(0) + \gamma t \tag{2.4}$$

then this is a very informative quantity. We'll now state carefully why it is informative in this case. That may seem pedantic at this point, but it will become useful when we generalise in Sec. **??**.

The additive growth rate (Eq. 2.1) is informative of how fast x grows under additive dynamics (Eq. 2.4) because in this case the t-dependence drops out: we can measure g_a whenever we want, and we'll always get the same value, $g_a = \gamma$. Not to get too philosophical about it, but this kind of time-translation invariance (fancy word) is a key concept in science: the search for laws is the search for universal structure – especially for time-translation invariant structure, for something "timeless."

Let's re-write the linear dynamic (Eq. 2.4) in differential form

$$d\mathsf{x} = \gamma dt \tag{2.5}$$

Because γ depends neither on t nor on x, we can re-write this as

$$d\mathsf{x} = d(\gamma t) \tag{2.6}$$

This second way of writing tells us that the growth rate γ is really a sort of clock speed. There's no difference between rescaling t and rescaling γ (by the same factor) – that means γ is a time scale.

We make a mental note: *the growth rate is a clock speed.* But what kind of clock speed are we talking about? What's a clock speed anyway?

Or: what's a clock? A clock is a process that we believe does something repeatedly at regular intervals. It lets us measure time by counting the repetitions. By convention, after 9,192,631,770 cycles of the radiation produced by the transition between two levels of the caesium 133 atom we say "one second has elapsed." That's just something we've agreed on. But any other thing that does something regularly would work as a clock – like the Earth completing one full rotation around its axis *etc.*

When we say "the growth rate of the process is γ," we mean that x advances by γ units on the process-scale (meaning in x) in one standard time unit (in finance we often choose one year as the unit, Earth going round the Sun). So it's a conversion factor between the time scales of a standard clock and the process clock.

Of course, a clock is no good if it speeds up or slows down. For processes other than additive growth we have to be quite careful before we can use them as clocks, i.e. before we can state their growth rates.

2.3.2 Exponential growth rate

Now what about the exponential growth rate, (Eq. 2.2)? This first thing to notice is that it's not time-translation invariant for additive growth, (Eq. 2.4). Substituting (Eq. 2.4) in (Eq. 2.2) gives

$$\begin{aligned} g_{\mathrm{e}} &= \frac{\ln \mathsf{x}(t+\Delta t) - \ln \mathsf{x}(t)}{\Delta t} &(2.7)\\ &= \frac{\ln\left[\mathsf{x}(t) + \gamma\Delta t\right] - \ln \mathsf{x}(t)}{\Delta t} &(2.8)\\ &= \frac{\ln\left(1 + \frac{\gamma\Delta t}{\mathsf{x}(t)}\right)}{\Delta t}. &(2.9)\end{aligned}$$

That means the exponential growth rate does not extract the clock speed γ from linear growth. There's a mismatch between the process and the form of the rate with which we're measuring its speed. The exponential growth rate of additive growth is not a constant but (see RHS of (Eq. 2.9)) depends on $\mathsf{x}(t)$, *i.e.* on the time when we started measuring. It also depends on how long we measured, Δt. If we used it to characterise the growth in described by (Eq. 2.4), we would find lots of contradictions – some people would say the growth is faster, others slower because they measured at different times or for different periods.

But the exponential growth rate is commonly used, and for good reasons. Let's see what it's good for, by imposing that it's useful and then working backwards to find the process we should use it for (we expect to find exponential growth).

We require that (Eq. 2.2) yield a constant, let's call that γ again, irrespective of when we measure it.

$$g_{\mathrm{e}} = \frac{\Delta \ln \mathsf{x}}{\Delta t} = \gamma, \tag{2.10}$$

or

$$\Delta \ln \mathsf{x} = \gamma \Delta t, \tag{2.11}$$

or indeed, in differential form, and revealing that again the growth rate is a clock speed: γ plays the same role as t,

$$d \ln \mathsf{x} = d(\gamma t). \tag{2.12}$$

This differential equation can be directly integrated and has the solution

$$\ln \mathsf{x}(t) - \ln \mathsf{x}(0) = \gamma t. \tag{2.13}$$

We solve for the dynamic $\mathsf{x}(t)$ by writing the log difference as a fraction

$$\ln \left[\frac{\mathsf{x}(t)}{\mathsf{x}(0)}\right] = \gamma t, \tag{2.14}$$

and exponentiating

$$\mathsf{x}(t) = \mathsf{x}(0) \exp(\gamma t) \tag{2.15}$$

As expected, we find that the *exponential* growth rate, (Eq. 2.2), is the appropriate growth rate (meaning time-independent) for *exponential* growth.

In terms of clocks, what just happened is this: we insisted that (Eq. 2.2) be a good definition of a clock speed. That requires it to be constant, meaning that the process has to advance on the logarithmic scale, specified in (Eq. 2.2), by the same amount in every time interval (measured on the standard clock, of course Earth or caesium).

2.3.3 General growth rate

Finally let's be more ambitious and posit a general process, $\mathsf{x}(t)$, of which we only assume that it grows according to a dynamic that can be written down as a separable differential equation. We could be even more general, but this is bad enough.

How do we define a growth rate now?

Well, as in Sec. **??**, we insist that the thing we're measuring will be a clock speed, *i.e.* a time-independent rescaling of time. We enforce this by writing down the dynamic in differential form, containing the growth rate as a time rescaling factor. Then we'll work backwards and solve for g:

$$d\mathsf{x} = f(\mathsf{x}) d(gt) \tag{2.16}$$

(for linear growth, like in (Eq. 2.4), $f(\mathsf{x})$ would just be $f(\mathsf{x}) = 1$, and for exponential growth, (Eq. **??**), it would be $f(\mathsf{x}) = \mathsf{x}$, but we're leaving it general). We separate variables in (Eq. 2.16) and integrate the differential equation

$$\int_{\mathsf{x}(t)}^{\mathsf{x}(t+\Delta t)} \frac{1}{f(\mathsf{x})} \delta \mathsf{x} = g \Delta t, \tag{2.17}$$

and we've got what we want, namely the functional form of g:

$$g = \frac{\int_{\mathsf{x}(t)}^{\mathsf{x}(t+\Delta t)} \frac{1}{f(\mathsf{x})} d\mathsf{x}}{\Delta t}. \tag{2.18}$$

This doesn't quite look like our stated aim: the general expression for a growth rate, (Eq. 2.3). But we get there, by simplifying (Eq. 2.18) and denoting the definite integral with the letter v, so that

$$\Delta v = \int_{\mathsf{x}(t)}^{\mathsf{x}(t+\Delta t)} \frac{1}{f(\mathsf{x})} d\mathsf{x}. \tag{2.19}$$

Equation (2.3) now follows exactly by substituting (Eq. 2.19) in (Eq. 2.18). This answers the second question of Sec. 2.3: "how do we know which transformation to use?"

But there's a simpler way of finding the transformation $v(\mathsf{x})$ that doesn't involve integrals and differential equations. Let $\mathsf{x}(t)$ be whatever function it wants – we know one transformation of $\mathsf{x}(t)$ that's linear in time, namely the inverse function of $\mathsf{x}(t)$, which we denote

$$\mathsf{x}^{(-1)}(\mathsf{x}) = t \tag{2.20}$$

$\mathsf{x}^{(-1)}(\mathsf{x})$ is the transformation that pulls t out of x: I give you the value of x, you take $\mathsf{x}^{(-1)}(\mathsf{x})$, and you know what t is.

So far, so good – now we know how to get t, which is of course linear in t. But that no longer tells us how fast something grows: we can't use $\mathsf{x}^{(-1)}(\mathsf{x})$ as $v(\mathsf{x})$ because $\frac{\mathsf{x}^{(-1)}[\mathsf{x}(t+\Delta t)] - \mathsf{x}^{(-1)}[\mathsf{x}(t)]}{\Delta t} = 1$, always. So something is missing.

If we use $\mathsf{x}^{(-1)}$ as the transformation in the general growth rate, we're in effect measuring the speed of the process on the scale of the process, which is why the answer is trivial: we will always find a growth rate of 1. Remember that the growth rate is a *conversion factor* between time measured on the standard clock (one that ticks once a second, say), and time measured on the process clock (one that advances γ units on the x-scale in each second).

So $\mathsf{x}^{(-1)}$ has the right form but not the right scale. Instead, let's try the following: take the process $\mathsf{x}(t)$ at unit rate on the standard clock. We'll denote this as $\mathsf{x}_1(t)$. If we now take its inverse as the transformation, $v(\mathsf{x}) = \mathsf{x}_1^{(-1)}(\mathsf{x})$, it will of course produce a rate 1 if $\gamma = 1$. But if γ is something else, it will extract that something else for us!

Here's the algorithm for measuring the growth rate for a general process $\mathsf{x}(t)$.

- Write down the process at rate 1 on the standard clock, $\mathsf{x}_1(t)$.
- Invert it, to find the transformation $v = \mathsf{x}_1^{(-1)}(\mathsf{x})$.

- Finally, evaluate the rate of change of the transformation of the process at the unknown growth rate,

$$g = \frac{\mathsf{x}_1^{(-1)}[\mathsf{x}(t+\Delta t)] - \mathsf{x}_1^{(-1)}[\mathsf{x}(t)]}{\Delta t} \tag{2.21}$$

The key conceptual message from this section is this: any growth process defines an appropriate functional form of a growth rate. If we measure a process with the wrong form of growth rate, we obtain nonsense. Measurements will be inconsistent, depending on arbitrary circumstances like the time of measurement.

The key formal result is this: with (Eq. 2.3) tells us there is a special transformation of wealth that enters into the only meaningful way of stating how fast wealth grows. Equation (2.21) tells us what that transformation is.

The transformation is a linearisation. At the moment we could call it the stationarity transformation because it appropriately removes time dependence. Later – when we generalise to random growth processes – we will call it the ergodicity mapping. In the economics literature, what appears in its place is called the utility function – a term we will mostly avoid because it comes with unhelpful connotations.

2.4 Decisions in a deterministic world

Having clarified what a growth rate is, we can now apply our decision axiom to different situations: act so as to maximize the growth rate of your wealth. In other words, we can explore the world generated by this axiom. We will build from the ground up. First, in Sec. 2.4.1, we will look at comparing two simultaneous payments of different magnitude – which one will the model human choose who obeys our axiom? This is a sanity check: does the model human choose the bigger payment? Next, in Sec. 2.4.2 we add time – what if the model human chooses between two payments of different magnitudes that will occur at different times? This is already a far more complex situation, where the decision criterion requires knowledge of the dynamic. It will shed light on what's called "discounting."

In the next chapter, Sec. **??**, we will add fluctuations, noise, uncertainty: what if the model human doesn't know the magnitude (or time) of the payments with certainty? But for now, everything will be perfectly known.

2.4.1 Different magnitudes

I'm off to the bank to withdraw some money for you. I offer to give you either

(1) $10 when I get back or

(2) $25 when I get back. You tell me what you prefer.

Let's see what our decision axiom says you'll do. Remember there's no uncertainty, I'm not lying to you, no one will rob me on my way from the bank *etc.*

I haven't told you how long it will take me to get to the bank, so we have to keep that general. We'll call that time interval Δt. Because we know that

Δt is the same under options (1) and (2) we don't actually need to know its value to compare the growth rates for the two options. Nor do we have to know the functional form of the growth rate. In this simple case, we can work with a general $v(\mathsf{x})$ in (Eq. 2.3), and any growth rate will give the same answer. Let's see. Under option (1) we have

$$g^{(1)} = \frac{v(\mathsf{x} + \$10) - v(\mathsf{x})}{\Delta t}, \tag{2.22}$$

and under option (2) we have

$$g^{(2)} = \frac{v(\mathsf{x} + \$25) - v(\mathsf{x})}{\Delta t}. \tag{2.23}$$

To find out which growth rate is larger, we subtract $g^{(1)}$ from $g^{(2)}$

$$g^{(2)} - g^{(1)} = \frac{v(\mathsf{x} + \$25) - v(\mathsf{x})(-v(\mathsf{x} + \$10) - v(\mathsf{x}))}{\Delta t} \tag{2.24}$$

$$= \frac{v(\mathsf{x} + \$25) - v(\mathsf{x} + \$10)}{\Delta t}. \tag{2.25}$$

Because $v(\mathsf{x})$ is monotonically increasing, any proper growth rate will be greater under option (2), and our model humans will always go for option (2). That's good – because we would have chosen option (2) if we were you, and our model reproduces this intuitive result.

More generally, our model says: of two certain payments of different sizes at the same time, choose the bigger one.

2.4.2 Different magnitudes and times: discounting

Let's make the decision a little harder: what if we offer you the same amounts as before, but now at different times:

(1) \$10 in a month or

(2) \$25 in two months?

Again, we will compute the two growth rates corresponding to options (1) and (2), and then choose the bigger one – that's how we have been programmed to behave in the world that our axiom is creating. But unlike in the previous case, the functional form of the growth rate will now be important.

Discounting under additive dynamics

Let's start with the additive growth rate, (Eq. 2.1), which is nothing but using the identity function in the general growth rate, $v(\mathsf{x}) = x$ in (Eq. 2.3). Which payment do the model humans choose according to this rate? We've got all the parameters, so this is just a matter of substitution

$${g_{\mathrm{a}}}^{(1)} = \frac{\mathsf{x} + \$10 - \mathsf{x}}{1 \text{ month}} = \$120 \text{ p.a.}, \tag{2.26}$$

and

$${g_{\mathrm{a}}}^{(2)} = \frac{\mathsf{x} + \$25 - \mathsf{x}}{2 \text{ months}} = \$150 \text{ p.a..} \tag{2.27}$$

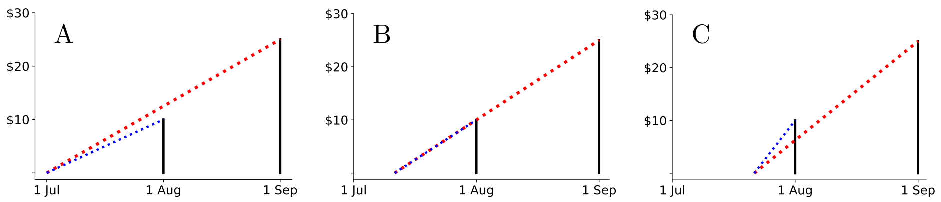


Figure 2.1: Slopes with linear vertical scales are *additive* growth rates. (A) At the beginning option (2) yields the highest additive growth rate; (B) after 1/3 of a month, the options are equally good; (C) after 2/3 of a month, preference reversal has taken place, and option (1) now yields the highest growth rate. As the first payment approaches, the associated growth rate diverges.

The result is clear: the decision axiom, using this growth rate, produces model humans that prefer payment (2). The additive growth rate has a unique feature: initial wealth, x cancels out. I didn't need to know your initial wealth to compute the rate! Only under additive dynamics does initial wealth not enter into the computation of the growth rate, and growth rates can be computed with knowledge of only the payouts and waiting times.

But perhaps the setup is more interesting than it seems at first glance. In the economics literature, decision-making based on additive growth rates is called "hyperbolic discounting" because this case is mathematically equivalent to discounting payments in the future with the hyperbolic function $\frac{1}{\Delta t}$.

An interesting feature of optimizing additive growth rates is what's called "preference reversal:" let's keep our example as it is, except we now let time march forward, holding fixed the moments in time when the payments are to be made. Under these conditions, there comes a time, precisely after a third of a month, when option (2) is no longer preferred, see Fig. 2.2.

You may wonder why someone might model wealth as an additive process. Here is one possibility: if wealth is mostly affected by income and expenses then it will be described by an additive dynamic. Imagine you have a monthly salary of $\$1,000$, and you spend $900 every month on all your expenses. So long as any investment income, like interest payments *etc.* , is negligible, your wealth will follow (Eq. 2.4) with $\gamma = \$100$ per month.

Discounting under multiplicative dynamics

What about the exponential growth rate, with $v(\mathsf{x}) = \ln \mathsf{x}$ in (Eq. 2.3)? We now have growth rates

$$g_{\mathrm{m}}{}^{(1)} = \frac{\ln(\mathsf{x} + \$10) - \ln(\mathsf{x})}{1 \text{ month}}, \tag{2.28}$$

and

$$g_{\mathrm{m}}{}^{(2)} = \frac{\ln(\mathsf{x} + \$25) - \ln(\mathsf{x})}{2 \text{ months}}. \tag{2.29}$$

Curiously, which is greater depends on your initial wealth in our model world. If your wealth is $100, then $g_{\mathrm{m}}{}^{(1)} \approx 114\%$ p.a., and $g_{\mathrm{m}}{}^{(2)} \approx 134\%$ p.a., wherefore you will choose option (2).

But if your initial wealth is $1, then $g_{\mathrm{m}}{}^{(1)} \approx 2,877\%$ p.a. and $g_{\mathrm{m}}{}^{(2)} \approx 1,955\%$ p.a., and you'll choose option (1).

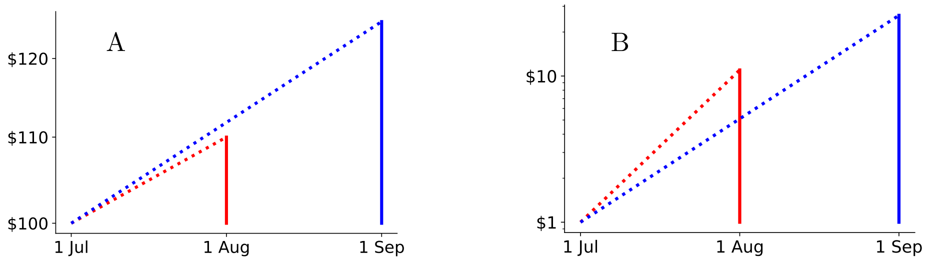


Figure 2.2: Slopes with logarithmic vertical scales. (A) If you have a lot of money (here \$100), exponential growth-rate optimization tells you to be patient and choose the later, larger, payment of \$25. (B) If you have little money (here \$1), the same criterion – exponential growth-rate optimization – tells you to get the cash as fast as possible, and choose the earlier, smaller, payment of \$10.

Notice how the poorer decision maker seems to be more impatient, despite his use of the exact same decision axiom. Using $v(\mathsf{x}) = \ln(\mathsf{x})$ in (Eq. 2.3) is related to what's called "exponential discounting" in the economics literature [**?**]. The word "exponential" is there because we're discounting under multiplicative dynamics (which means exponential growth), and the logarithm is the inverse function of the exponential, which we need to define the appropriate growth rate, see (Eq. 2.21).

Again, let's ask why one might want to model wealth as a multiplicative process. Multiplicative processes have the property that how much you gain tomorrow is proportional to how much you currently have. Increases in wealth are proportional to wealth – this multiplicative property is virtually ubiquitous in Nature. You can imagine many reasons why it applies to wealth. For instance, wealth can be invested in interest-paying bonds. Or it can be invested in oneself: I may pay for a roof over my head, which transforms my life and earning potential from that of someone living on the streets to that of someone with a home address. Similarly, I can invest in my health and education. In Nature, multiplicativity and exponential growth occurs whenever something lives. That's because the most successful definition we have of life is "that which self-reproduces," and self-reproduction implies multiplicative growth.

We learn: in this slightly more complex though still fully deterministic case, which option is preferable does not only depend on the options available but also on the personal circumstances of the decision maker. Both the way the decision maker thinks about wealth as a dynamical process, and his wealth influence his preferences.

Perhaps the most significant message is the richness of this problem. We're applying nothing but our simple axiom, but it forces us to choose how we think about the dynamics of our wealth, and in reality that may depend strongly on many difficult to specify circumstances. In real life payments are not just offered at some point in time, but usually in return for something – an asset or work. Depending on the specific exchange, an additive, multiplicative, or more general model will be appropriate. Such models are explored in even greater detail in [**?**].

Importantly, we need not resort to psychology to generate a host of behaviours, such as impatience of poorer individuals or preference reversal as time ticks on.

Acronyms

BM Brownian motion.

GBM geometric Brownian motion.

LHS left-hand side.

LML London Mathematical Laboratory.

PDF probability density function.

RHS right-hand side.

SDE stochastic differential equation.

List of Symbols

A An observable.

d Differential operator in Leibniz notation, infinitesimal.

δt A time interval corresponding to the duration of one round of a gamble or, mathematically, the period over which a single realisation of the constituent random variable of a discrete-time stochastic process is generated..

δ Most frequently used to express a difference, for instance δx is a difference between two wealths x. It can be the Kronecker delta function, a function of two arguments with properties $\delta(i, j) = 1$ if $i = j$ and $\delta(i, j) = 0$ otherwise. It can also be the Dirac delta function of one argument, $\int f(x)\delta(x - x_0)dx = f(x_0)$.

Δ Difference operator, for instance Δv is a difference of two values of v, for instance observed at two different times.

Δt A general time interval..

η Langevin noise with the properties $\langle\eta\rangle = 0$ and $\langle\eta(t_1)\eta(t_2)\rangle = \delta(t_1 - t_2)$.

f Generic function.

g Growth rate.

$g_{\mathbf{e}}$ Ergodic growth rate for exponential growth.

$g_{\mathbf{a}}$ Ergodic growth rate under additive dynamics, *i.e.* rate of change, $g_{\mathrm{a}}(x; t, \Delta t) = \frac{\Delta x(t)}{\Delta t}$.

$\overline{g_{\mathbf{a}}}$ Time-average ergodic growth rate under additive dynamics, *i.e.* long-term rate of change, $\overline{g_{\mathrm{a}}}(x) = \lim_{\Delta t \to \infty} \frac{\Delta x(t)}{\Delta t}$.

$g_{\mathbf{m}}$ Ergodic growth rate for multiplicative dynamics, *i.e.* exponential growth rate, $g_{\mathrm{m}}(x; t, \Delta t) = \frac{\Delta \ln x}{\Delta t}$.

$\overline{g}$ Time-average ergodic growth rate.

i Label for a particular realization of a random variable.

j Label of a particular outcome.

μ Drift term in BM.

n n_j is the number of times outcome j is observed in an ensemble.

N Ensemble size, number of realizations.

$\mathcal{N}$ Normal distribution, $x \sim \mathcal{N}(\langle p \rangle, \mathrm{var}(p))$ means that the variable p is normally distributed with mean $\langle p \rangle$ and variance $\mathrm{var}(p)$..

o Little-o notation.

p Probability, p_j is the probability of observing event j in a realization of a random variable.

$\mathcal{P}$ Probability density function.

q Possible values of Q. We denote by q_i the value Q takes in the i^{th} realization, and by q_j the j^{th}-smallest possible value.

Q Random variable defining a gamble through additive wealth changes.

r Random factor whereby wealth changes in one round of a gamble.

$r_{\langle\rangle}$ Expectation value of growth factor r.

$\overline{r}$ Average growth factor over a long time.

s Dummy variable in an integration.

σ Magnitude of noise in a Brownian motion.

t Time.

T Number of sequential iterations of a gamble, so that $T\delta t$is the total duration of a repeated gamble..

t_0 Specific value of time t, usually the starting time of a gamble..

τ Dummy variable indicating a specific round in a gamble.

v Stationarity mapping function, so that $v(x)$ has stationary increments.

var Variance.

W Wiener process, $W(t) = \int_0^t dW$ is continuous and $W(t) \sim \mathcal{N}(0, t)$.

x Wealth.

$\overline{x}$ Time-average of x. With subscript Δt, this is a finite-time average, without the subscript it refers to the infinite-time average..

x Deterministic wealth.

ξ A standard normal variable, $\xi \sim \mathcal{N}(0, 1)$.

Y Random variable that is a function of another random variable, Z.

z Generic value of a random variable.

Z Generic random variable.

Bibliography

[1] G. I. Barenblatt. *Scaling.* Cambridge University Press, 2003.

[2] 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.

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

[4] J. Y. Campbell. *Financial decisions and markets, A course in asset pricing.* Princeton University Press, 2017.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, 1991.

[6] J. C. Cox, J. E. Ingersoll, and S. A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica*, 53(2):363–384, 1985.

[7] M. A. B. Deakin. G. I. Taylor and the Trinity test. *Int. J. Math. Edu. Sci. Tech.*, 42(8):1069–1079, 2011.

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

[9] 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.

[10] R. P. Feynman. *The Feynman lectures on physics.* Addison-Wesley, 1963.

[11] J. M. Harrison. *Brownian motion of performance and control.* Cambridge University Press, 2013.

[12] H. Hinrichsen. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys.*, 49(7):815–958, 2000.

[13] J. C. Hull. *Options, Futures, and Other Derivatives.* Prentice Hall, 6 edition, 2006.

[14] K. Itô. Stochastic integral. *Proc. Imperial Acad. Tokyo*, 20:519–524, 1944.

[15] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory.* Springer, 2 edition, 2008.

[16] J. L. Kelly, Jr. A new interpretation of information rate. *Bell Sys. Tech. J.*, 35(4):917–926, July 1956.

[17] P. S. Laplace. *Théorie analytique des probabilités.* Paris, Ve. Courcier, 2 edition, 1814.

[18] J. Marro and R. Dickman. *Nonequilibrium Phase Transitions in Lattice Models.* Cambridge University Press, 1999.

[19] K. Menger. Das Unsicherheitsmoment in der Wertlehre. *J. Econ.*, 5(4):459–485, 1934.

[20] P. R. Montmort. *Essay d'analyse sur les jeux de hazard.* Jacque Quillau, Paris. Reprinted by the American Mathematical Society, 2006, 2 edition, 1713.

[21] H. Morowitz. *Beginnings of cellular life.* Yale University Press, 1992.

[22] O. Peters. Comment on D. Bernoulli (1738).

[23] O. Peters. The time resolution of the St Petersburg paradox. *Phil. Trans. R. Soc. A*, 369(1956):4913–4931, December 2011.

[24] O. Peters and A. Adamou. Insurance makes wealth grow faster. *arXiv:1507.04655*, July 2015.

[25] O. Peters and M. Gell-Mann. Evaluating gambles using dynamics. *Chaos*, 26(2):23103, 2016.

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

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

[28] P. Samuelson. St. Petersburg paradoxes: Defanged, dissected, and historically described. *J. Econ. Lit.*, 15(1):24–55, 1977.

[29] G. Taylor. The formation of a blast wave by a very intense explosion. II. the atomic explosion of 1945. *Proc. R. Soc. A*, 201:175–186, 1950.

[30] E. O. Thorp. *Beat the dealer.* Knopf Doubleday Publishing Group, 1966.

[31] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier (Amsterdam), 1992.

[32] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1944.

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