The Modelling of Random Phenomena

Aureli Alabert

This survey is based on a course given by the author in the Università degli Studi dell'Aquila, as a part of the Intensive Programme **Mathematical Models in Life and Social Sciences**, in July 2008.
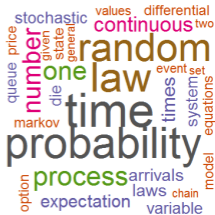
It is a fast-paced introduction, aimed at graduate students, to the mathematical modelling of phenomena and experiments in which randomness is present. After the introductory first section, several examples from daily life, industry, biology, and finance are presented, and the necessary tools are introduced alongside, but without much mathematical detail.

1 Introduction to randomness

1.1 Random phenomena

A **random phenomenon** is a physical phenomenon in which "randomness" takes a place.

So, what is **randomness**? It is something that we do not control, in the sense that it may lead to different outcomes or measurements of the phenomenon in what we believe are "identical" conditions.

There are many keywords associated to the discussion and mathematical foundation of random phenomena: probability, chance, likelihood, statistical regularity, plausibility, . . . There are whole books discussing and trying to explain what is the nature of chance and randomness. It is not worth going into such philosophical depth for the practitioner. One may get lost into the variety of "defini-

tions" or "trends" related to the word **probability** (classical, frequentist, axiomatic, subjective, objective, logical, ...) or **statistics** (frequentist, classical, Bayesian, decision-theoretic, ...).

1.2   The modelling point of view

Instead, take the modelling point of view: Each problem must be treated in its own merits, choosing the appropriate tools provided by mathematics.

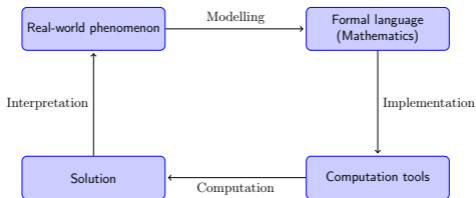In general, the modelling of a real world phenomenon follows the scheme of Figure 1.

Figure 1: General mathematical modelling

When randomness is present, the scheme is the same. The distinguishing feature is the use of the mathematical concept of "probability" (which has an unambiguous and worldwide accepted definition), and the solution to the problem comes usually in the form of a "probability distribution" or some particular property of a probability distribution. See Figure 2.
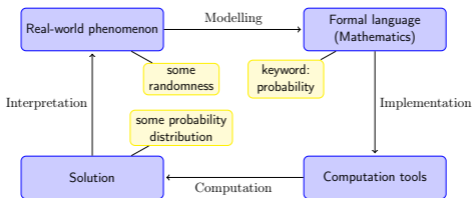

Figure 2: Mathematical modelling in the presence of randomness

1.3 Quantifying randomness: Probability

Take a playing die, for example (Figure 3). Throwing a die is a familiar random phenomenon. We need the outcome to be unpredictable (thus potentially different) each time; otherwise the die is not useful for playing. On the other hand, the experiment is performed each time in identical conditions: We throw the die on the table so that it rebounds several times before stopping. Of course, the conditions are no "truly" identical; in this case, our ignorance about the exact physical conditions provides the desired unpredictability, therefore the randomness.

Suppose we examine the die, and we see that it looks new, homogeneous, balanced and with no visible manufacturing defect. Is there any outcome that looks more likely to appear than some other? If not, then it is logical that any attempt to quan-
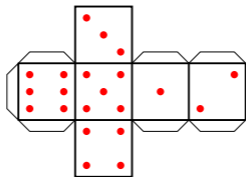
Figure 3: A playing die developed to show all its faces.

tify the likelihood of the outcomes lead to assign the same quantity to all outcomes.

We may think that every outcome takes an equal part of a cake they have to share. Let us say, arbitrarily, that the cake measures 1. Therefore, every outcome has to take $1/6$ of the cake. We say that every possible **result** ω of the random phenomenon "throwing a balanced die" has a **probability** of $1/6$. See Figure 4.

From the probability of all possible **results** $\omega \in \Omega$, we can deduce (define, in fact, but in the only
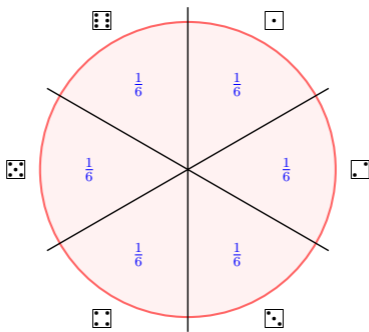
Figure 4: A (presumed) balanced die eating the probability cake.

sensible way) the probability of all possible **events**, that is, subsets $A \subset \Omega$: The event A takes the part of cake that its results $\omega \in A$ take in total.

1.4 The law of Large Numbers

The relative frequency of an event in a series of identical experiments is the quotient

$$\frac{\text{Number of occurrences of the event}}{\text{Number of experiments performed}}.$$

If $1/6$ is the probability of obtaining a 3 when tossing the die, *it can be proved* that the relative frequency of the event $\{3\}$ converges to $1/6$ when the number of experiments tends to infinity.

In general, the relative frequency of an event converges to its probability. This is the **Law of Large Numbers**. It is a Theorem (an important one). It is not a definition of "probability", as it is frequently said.

1.5 Statistical inference

We may think that a die is balanced when in fact it is not. In this case, the relative frequencies will not

converge to the probabilities that we expect. Or, plainly, we suspect that the die is not balanced, and we do not know what to expect.

In any case, the Law of Large Numbers leads to the following idea:

1. Toss the die as many times as you can.

2. Write down the relative frequency of each result.

3. Construct the model of the die by assigning

   Probability of ω := Relative frequency of ω.

This is **Statistical Inference**: We construct a model of a random phenomenon using the data provided by a **sample** of the **population**.

The population here is a (hypothetical) infinite sequence of die throws. In the usual applications, the population is a big, but finite, set of objects

(people, animals, machines or anything), and the sample is a subset of this set.

In another common (and definitely overused) setting of statistical inference, one simply declares the die as balanced unless the relative frequencies deviate too much of the expected values. If they do, then the die is declared "non-balanced".

1.6 Probability. The mathematical concept

We want a mapping that assigns to every event a number called "the probability of the event" satisfying:

1. It is nonnegative.

2. The probability of the whole set Ω of possible results is 1.

3. The probability of the union of two disjoint events is the sum of the probabilities of the two events.

Formally: A **probability** is a mapping

$$P\colon\ \mathcal{P}(\Omega) \longrightarrow [0,1]$$
$$A \longmapsto P(A)$$

such that $P(\Omega) = 1$ and for any countable family $\{A_n\}_n \subset \Omega$, with $A_i \cap A_j = \emptyset$ if $i \neq j$,

$$P\Big(\bigcup_{n=1}^{\infty} A_n \Big) = \sum_{n=1}^{\infty} P(A_n)\,.$$

This definition captures perfectly the idea of the pieces of cake taken by the different events that we saw in Figure 4. The extension to a countably infinite union instead of just finite does not harm and allows to construct a mathematical theory much more in line with the phenomena that we intend

to model. Demanding the same for uncountable unions, on the contrary, would collapse the theory and make it useless. If Ω is a finite set, then of course this discussion is void.

Sometimes it is not possible to define the mapping on the whole collection $\mathcal{P}(\Omega)$ of subsets of Ω preserving at the same time the properties of the definition. In this case, we define it on a subcollection $\mathcal{F} \subset \mathcal{P}(\Omega)$ satisfying some desirable stability properties:

1. $\Omega \in \mathcal{F}$,

2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

3. $\{A_n\}_n \subset \mathcal{F} \Rightarrow \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$,

where $A^c := \Omega - A$ is the complement set of A.

These subcollections are called σ-**fields** or σ-**algebras**. They enjoy the right stability properties

so that the additivity property in the definition of P still makes sense.

Probability Theory is a specialised part of **Measure and Integration Theory**. In general, a **measure** is a function defined on the sets of a σ-field with values in a set which is not necessarily the interval $[0, 1]$.

1.7   Drawing probabilities

Probabilities behave like areas of planar regions. Consider Figure 5.

To compute the area of the region $A \cup B$, we may add the areas of A and B, and then subtract the area of $A \cap B$, which have been counted twice. This leads immediately to the fundamental formula:
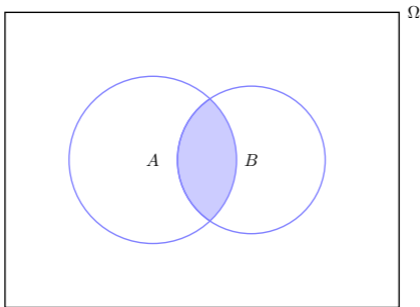
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Figure 5: Probabilities and areas

All usual lists of "properties of the probabilities" are trivial derivations of this formula, and can also be deduced from Figure 5. It is useless to learn them by heart.

1.8 Conditional probabilities

Consider the following example (see Figure 6): We have a box with five white balls, numbered 1 to 5,
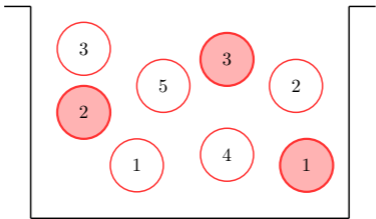
Figure 6: Balls in a box

and three red balls, numbered 1 to 3. We pick a ball "completely at random". What is the probability of drawing an even number?

First of all, what is the probability of picking a particular ball? The expression "completely at random", though imprecise, is used to mean that all outcomes are equally likely, as in the case of the balanced die.

We are interested in the event $A = \{W_2, W_4, R_2\}$, where W means white ball and R red ball. Since each of the balls in A takes $\frac{1}{8}$ of the proba-

bility cake, we have that

$$P(A) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}\,.$$

Now suppose a ball has been picked by someone, who tell us that the ball is white. What is the probability that the ball carries an even number?

In this case the possible results are $W = \{W_1, W_2, W_3, W_4, W_5\}$, all with probability $\frac{1}{5}$, thus the probability of $\{W_2, W_4\}$ is $\frac{2}{5}$. The additional information has led as to change the model, and consequently the value of the probabilities.

Notice that :

$$\frac{2}{5} = \frac{2/8}{5/8} = \frac{P(A \cap W)}{P(W)}\,,$$

where the probabilities in the quotient are those of the original model.

The **conditional probability** of A to B is de-

fined as

$$P\big(A \,\big/\, B\big) := \frac{P(A \cap B)}{P(B)}\,.$$

In relation to Figure 5, the conditional probability of A to B is the proportion of area of A inside B.

We say that A and B are **independent** if the information that B has happened does not change the probability of A:

$$P\big(A \,\big/\, B\big) = P(A)\,.$$

Equivalently,

$$P(A \cap B) = P(A) \cdot P(B)\,.$$

1.9 Random variables

We can now step into a second level of difficulty: the concept of random variable. Let us consider the

following example: We toss two balanced dice, and we are interested in the sum of the points shown. We may consider directly the set $\Omega = \{2, \ldots, 12\}$ and assign probabilities to each element of Ω, but this is difficult; or we may keep the model closer to the real experiment by defining $\Omega = \{(i,j) : 1 \le i \le 6, \ 1 \le j \le 6\}$, and think of the mapping

$$\Omega \xrightarrow{X} \{2, \ldots, 12\}$$
$$(i,j) \longmapsto i + j$$

If the dice really look balanced, and if it is clear that the outcome of one die does not influence the outcome of the other, then it is natural to distribute the same amount of the probability cake to every pair (i,j), that means $P\{(i,j)\} = \frac{1}{36}$.

This setting induces a probability P_X on $\{2,$

..., 12}, which is what we are looking for:

$$P_X\{2\} = P\{(1,1)\} = \frac{1}{36},$$
$$P_X\{3\} = P\{(1,2),(2,1)\} = \frac{2}{36},$$
$$P_X\{4\} = P\{(1,3),(2,2),(3,1)\} = \frac{3}{36} \quad \ldots, \text{ etc.}$$

In general, a **random variable** is a mapping $X \colon \Omega \longrightarrow \mathbb{R}$. ($\mathbb{R}$ can be replaced by other convenient sets; technically, the random variable must take values in another **measure space**, that is, a set endowed with a σ-field.) The **law** of a random variable is the probability P_X on $\mathbb{R}$ induced by P and X as in the example.

From the modelling point of view *the law is the important thing*, not Ω or the mapping X themselves. Typically one says: "I am observing a random phenomenon following the law ...".

From the law of a random variable one may define certain numeric values that carry some infor-

mation, and that sometimes are all that is needed in a particular application. The most important one is the *expectation*, which is the "mean value" that a variable with that law will take. It can be thought as the limit of the arithmetic mean of the observed values of the variable when the number of observations tends to infinity. But this is again a version of the Law of Large Numbers, and not a definition.

The **expectation** $E[X]$ of a random variable X with law P_X and taking a countable number of values k is defined as

$$E[X] := \sum k \cdot P_X\{k\},$$

with the sum extended over all values taken by X. The **variance** of X is a degree of dispersion of its values around the expectation, and defined as

$$\mathrm{Var}[X] := E\left[(X - E[X])^2\right].$$

1.10   The binomial law

Leaving aside the elementary "equiprobable" or "uniform" model of the balanced die, the most basic useful example of probability law is the one appearing in the following situation:

Fix an event A of any random experiment. Call p its probability: $P(A) = p$. Repeat n times the same experiment, and let X be the number of occurrences of A in the n trials. The law of X is then determined by

$$P\{X = k\} = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0, \ldots, n. \tag{1}$$

We write $X \sim \text{Binom}(n, p)$ and say that X follows a **binomial law** with parameters (n, p).

The sentence "repeating n times the *same* experiment" means in particular that one experiment does not influence the result of another, and there-

fore events concerning the outcome or one experiment are independent of events concerning the outcome of the other experiments, in the sense of section 1.8. This fact is key in the deduction of formula (1).

2 Examples from daily life: Arrivals and waiting lines

2.1 The geometric law

Assume the experiments of Section 1.10 are performed continuously and at regular unit time intervals. We want to know the time elapsed between an occurrence of A and the next occurrence of A. Or, in other words, how many experiments are needed before observing again the event A.

This is a situation that may be of interest in manufacturing, where the event A is the occurrence

of a defective item in the production line.

Let N be the number of A^c occurrences before the next occurrence of A. Then it is easy to deduce

$$P\{N = k\} = (1 - p)^k \cdot p, \quad k = 0, 1, 2, \ldots$$

We write $N \sim \text{Geom}(p)$ and say that N follows a **geometric law** with parameter p.

2.2 Tails and the memoryless property

Once we know the **density function** (or **probability function**) $k \mapsto P\{N = k\}$, we can compute, as in the case of the die, the probability of any event $P\{N \in B\}$, where B is any subset of $\mathbb{N}$. In particular, we can compute the right and left **tails** of the law:

$$P\{N > k\} = (1-p)^{k+1}, \quad P\{N \le k\} = 1 - (1-p)^{k+1}.$$

Because of the (hypothesized) independence between the experiments, the law of N is the same if

we define N as the number of A^c occurrences before the first occurrence of A. From this fact one can prove the **memoryless property**:

$$P\{N > m + k \ / \ N > m\} = P\{N > k\}.$$

In words, knowing that the event has not appeared in the first k experiments, it is not more or less likely to appear than if we just start now the sequence.

2.3 Arrivals at random times: The Poisson law

Assume now that the arrivals occur at random times instead of regularly. For example, the arrival of customers to a waiting line may correspond to this situation. To be precise, assume:

1. People arrive alone (never in groups).

2. The probability p that an arrival occurs during a time interval of length h (small) is proportional to h:

$$p = \lambda \cdot h + o(h) \,.$$

3. The number of arrivals on disjoint time intervals are independent random variables.

We would like to know, for instance, the law of the number of arrivals N_t in the interval $[0, t]$, or the number of arrivals per unit time. The hypotheses above are quite suitable for a situation where the arrivals can be considered "completely at random".

Now, divide $[0, t]$ in intervals of length $h = t/n$. For n big enough, inside each interval we will see at most one arrival, and this will happen with probability λh. Therefore, the number of arrivals in

$[0, t]$ follows approximately a law $\text{Binom}(n, \lambda\, t/n)$. Hence, by (1):

$$P\{k \text{ arrivals in } [0,t]\} \approx \binom{n}{k} \cdot \left(\frac{\lambda\, t}{n}\right)^k \cdot \left(1 - \frac{\lambda\, t}{n}\right)^{n-k}.$$

Taking $n \to \infty$,

$$P\{k \text{ arrivals in } [0,t]\} = \frac{(\lambda\, t)^k}{k!} \exp\{-\lambda\, t\}. \quad (2)$$

Let N be the number of arrivals per unit time. We write $N \sim \text{Pois}(\lambda)$ and say that N follows a **Poisson law** with parameter λ:

$$P\{N = k\} = \frac{\lambda^k}{k!} \exp\{-\lambda\}.$$

The parameter λ is called the **traffic intensity**.

2.4 Interarrival times: The exponential law

Let T be the time between two arrivals. As in the case of the geometric law, this random variable

is equal in law to the time when the first arrival takes place. The event $\{T > t\}$ means to have 0 arrivals in $[0, t]$, whose probability according to (2) is $\exp\{-\lambda t\}$.

We observe that this probability is nonzero for all $t \geq 0$, and that it cannot be expressed as the sum of the probability of elementary events. We say that the interarrival times follow a **continuous law**, in contrast with all laws seen so far, called **discrete laws**.

In the case of continuous laws, the **density** is a function $f \colon \mathbb{R} \longrightarrow \mathbb{R}^+$ such that $P\{T \in [a, b]\}$ is the area under its graph between a and b.

$$P\{T \in [a, b]\} = \int_a^b f \,.$$

To compute the density of the interarrival

times, we observe that

$$\int_0^t f = P\{T \in [0, t]\} = 1 - \exp\{-\lambda t\},$$

so that

$$f(t) = \lambda \cdot \exp\{-\lambda t\}.$$

$T \sim \text{Exp}(\lambda)$ is called the **exponential law** with parameter λ.

2.5 Continuous laws

Continuous laws have some features that contrast with those of discrete laws:

- The law is not determined by the probability of the individual outcomes.

- It is the density that determines the law. (This can be said to be true also for discrete laws, but the concept of "density function" is different.)

- It is not possible to assign a probability to all subsets of the real line (this is not obvious). But we do not need to! It is possible to assign a probability to all intervals, and therefore to the members of the minimal σ-field containing the intervals, which is far more than what we need from a practical point of view.

- Continuous laws show why we cannot ask a probability to be additive for collections of arbitrary cardinality. For example: $1 = P\{T \geq 0\} \neq \sum_{t \geq 0} P\{T = t\} = 0$.

- The **expectation** of a variable with a continuous law cannot be defined with sums. It is the integral

$$\mathrm{E}[X] := \int_{-\infty}^{\infty} x\, f(x)\, dx\,,$$

where f is the density. Notice however the analogy with the definition for discrete laws.

In the context of measure theory, the expectation can be expressed in a unified way for all cases.

The correct name of these laws is **absolutely continuous**, for mathematical consistency, but the adverb is frequently dispensed with. "Continuous", strictly speaking, simply means that the so-called **distribution function** $F(x) := P\{X \leq x\}$, which is always non-decreasing and right-continuous, is furthermore continuous; whereas "absolutely continuous" refers to the stronger property that the distribution function is a primitive of another function, the density: $F(x) = \int_{-\infty}^{x} f$.

2.6 Poisson arrivals / Exponential times

Still some remarks about the relation between the Poisson and the exponential laws:

1. If the interarrival times are $\text{Exp}(\lambda)$, then the arrivals per unit time are $\text{Pois}(\lambda)$.

2. This situation is called "completely random arrivals", in the sense that the arrival times $0 < t_1 < t_2 < \cdots < t_k < t$ have the law of k independent uniformly distributed values in $[0, t]$, after sorting them.

3. The exponential laws enjoy the same *memoryless property* as the geometric law,

$$P\{T > t + s \mid T > s\} = P\{T > t\},$$

and is the only continuous law with this property. It is a good model for lifetimes of "ageless devices"; for instance, the lifetime of an electronic device, or living beings in their middle ages, when the death comes from internal or external accidents (electric shocks, heart strokes, ...).
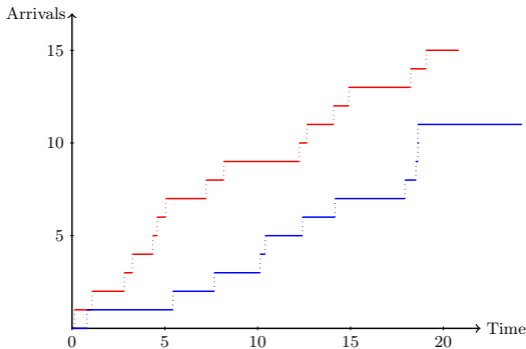
Figure 7: A Poisson sample path with $\lambda = 1$ (red) and with $\lambda = 0.5$ (blue). Lower λ means less frequent arrivals in average.

2.7 The Poisson process

The collection of random variables $\{N_t, \ t \geq 0\}$, counting how many arrivals have occurred in the time interval $[0, t]$, form the **Poisson process**.

When we observe a particular arrival phenomenon, we see, as time passes, a **sample path**

of the Poisson process (see Figure 7). We may also think of the Poisson process as the collection of all its sample paths.

2.8 Stochastic processes

In general, a random evolution in time is modelled by a **stochastic process**. There are two possible points of view of a stochastic process:

1. As a collection of random variables:

$$X := \{X_t, \ t \geq 0\}, \quad \text{with } X_t \colon \Omega \longrightarrow \mathbb{R}.$$

2. As a "random function"

$$X \colon \ \Omega \longrightarrow \mathbb{R}^{\mathbb{R}^+}$$
$$\omega \longmapsto X(\omega)$$

Here $\mathbb{R}^{\mathbb{R}^+}$ denotes the set of all functions $\mathbb{R}^+ \to \mathbb{R}$, which can be identified with the Cartesian product

of "$\mathbb{R}^+$ copies" of $\mathbb{R}$ as a set, as a topological space and as a measure space.

2.9   Queues (waiting lines)

A **queue** is a situation in which **users** arrive to a **service**, wait to be served if the service is not immediately available, and leave after having been served (Figure 8).
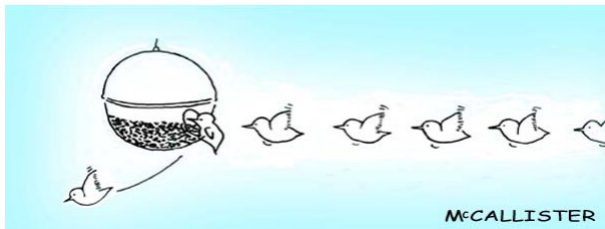

Figure 8: A typical simple queue: Customers arrive, wait in a line, are served, and leave the system. (Illustration appeared in The New Yorker, 1977)

Examples are customers in a supermarket cash,

cars in the highway at the toll booths, and parts in a manufacturing chain.

Its behaviour depends, among other things, on:

1. **Arrival pattern:** Interarrival times, number of users per arrival, patience of the customers, . . .

2. **Service pattern:** Service time, number of users served simultaneously, . . .

3. **Queue discipline:** FIFO (First-In, First-Out), LIFO (Last-In, First-Out), SIRO (Service in Random Order), . . . , with variants specifying priorities, pre-emption, etc.

4. **Capacity:** Number of users allowed to wait.

Moreover, everything may be dependent or not on the state of the system (number of users, etc.) and the elapsed time since the start.

Typical questions posed in these situations are:

- How many users are in the line? (at a given time, in the mean, . . .)

- How long a user must wait? (a given user, in the mean, . . .)

- How much time a service facility is idle?

- How long are the busy/idle periods of the service facility?

The answers are random variables if at least one of the features is random. We would like to know the law of these variables, or at least its expectation, or some other value of interest.

The purpose of knowing these laws or law parameters is, frequently, to take a decision about some controllable inputs of the queue, and with some cost associated to each of the values of these

inputs. For instance, the number of cashiers in a supermarket clearly influences the waiting time of the customers; benefits may increase thanks to that, but the running costs are also higher. Here we enter the realm of **optimisation** and **operations research**.

2.10 The M/M/1 queue. Transition probabilities

Assume that we have Poisson arrivals to a queue, the service time is also random and follows an exponential law (just one among some common situations), and there is a single service channel (only one user at a time is served).

More precisely, we now put in rigorous mathematics the hypothesis of Section 2.3. In the sequel we use the usual notation $o(h)$ to mean any function such that $\lim_{h \to 0} o(h)/h = 0$. Assume that the arrivals satisfy:

1. $P\{\text{more than one arrival in } [t, t+h]\} = o(h)$.

2. $P\{\text{an arrival occurs in } [t, t+h]\} = \lambda h + o(h)$.

3. The number of arrivals in non-overlapping time intervals are independent random variables.

And moreover the service times satisfy

1. $P\{\text{more than one service completed in } [t, t+h]\} = o(h)$.

2. $P\{\text{a service is completed in } [t, t+h]\} = \mu h + o(h)$ (assuming the service is not idle).

3. The number of completed services in non-overlapping time intervals are independent random variables.

All these properties together imply that we have a queue where the interarrival times follow the law $\mathrm{Exp}(\lambda)$ and the service times follow the law $\mathrm{Exp}(\mu)$.

Assume, moreover, than jointly considered, arrivals and services are independent.

Let us call now N_t the number of users in the system at time t. We can compute the probability that the state of the system changes from n users to any other number in some time interval $[t, t + h]$. These are called the **transition probabilities**, and can be considered for any stochastic process. It is easy to find, using the hypotheses above that for all $n \geq 1$

a) $P\{N_{t+h} = n + 1 \ / \ N_t = n\} = \lambda h + o(h)$, for $n \geq 0$.

b) $P\{N_{t+h} = n - 1 \ / \ N_t = n\} = \mu h + o(h)$, for $n \geq 1$.

c) $P\{N_{t+h} = n \ / \ N_t = n\} = 1 - (\lambda + \mu) h + o(h)$, for $n \geq 1$, and
$P\{N_{t+h} = 0 \ / \ N_t = 0\} = 1 - \lambda h + o(h)$.

d) All other transition probabilities are $o(h)$.

2.11 The M/M/1 queue. Differential equations

Fix two times $s \leq t$. Denote $p_{n,m}(s,t)$ the conditional probability of being in state m at time t, conditional to be in state n at time s. Then, for $m > 0$,

$$p_{n,m}(s, t+h) = \sum_{k \in \mathbb{N}} p_{n,k}(s,t) \cdot p_{k,m}(t, t+h)$$

$$= p_{n,m}(s,t) \cdot p_{m,m}(t, t+h)$$
$$+ p_{n,m-1}(s,t) \cdot p_{m-1,m}(t, t+h)$$
$$+ p_{n,m+1}(s,t) \cdot p_{m+1,m}(t, t+h) + o(h)$$
$$= p_{n,m}(s,t) \cdot \big(1 - (\lambda + \mu)\, h + o(h)\big)$$
$$+ p_{n,m-1}(s,t) \cdot \big(\lambda\, h + o(h)\big)$$
$$+ p_{n,m+1}(s,t) \cdot \big(\mu\, h + o(h)\big) + o(h)\,,$$

where the so-called Markov property have been used (see Section 3.3 below). Diving by h and taking $h \to 0$, we obtain

$$\frac{d}{dt} p_{n,m}(s,t) = -(\lambda + \mu)\, p_{n,m}(s,t) + \lambda\, p_{n,m-1}(s,t)$$
$$+ \mu\, p_{n,m+1}(s,t)\,.$$

Analogously, for $m = 0$, one finds

$$\frac{d}{dt}p_{n,0}(s,t) = -\lambda\, p_{n,0}(s,t) + \mu\, p_{n,1}(s,t)\,.$$

This is a countably infinite system of ordinary differential equations for the conditional probabilities $p_{n,m}(s,t) := P\big\{N_t = m \,\big/\, N_s = n\big\}$, for $s \le t$, and $n, m \in \mathbb{N}$.

One can also obtain differential equations for the law of N_t itself: Denote $p_n(t) = P\{N_t = n\}$. For $n > 0$,

$$\begin{aligned}
\frac{d}{dt}\, p_n(t) &= \frac{d}{dt}\Big(\sum_{k \in \mathbb{N}} p_k(0)\, p_{k,n}(0,t) \Big) \\
&= \sum_{k \in \mathbb{N}} p_k(0)\Big[-(\lambda + \mu)\, p_{k,n}(0,t) + \lambda\, p_{k,n-1}(0,t) \\
&\qquad\qquad\qquad\qquad + \mu\, p_{k,n+1}(0,t)\Big] \\
&= -(\lambda + \mu)\, p_n(t) + \lambda\, p_{n-1}(t) + \mu\, p_{n+1}(t)\,.
\end{aligned}$$

And, for $n = 0$,

$$\frac{d}{dt}p_0(t) = -\lambda\, p_0(t) + \mu\, p_1(t)\,.$$

We get again a countably infinite system of ordinary differential equations. The system can be solved exactly but it is difficult and there is a lot of higher mathematics involved.

2.12   The M/M/1 queue. Steady-state law

In the long run, as t grows, does the law of N_t stabilises? If this is true, then the derivatives in the system of Section 2.11 must vanish when $t \to \infty$:

$$0 = -\lambda p_0 + \mu p_1 \,,$$
$$0 = -(\lambda + \mu) p_n + \lambda p_{n-1} + \mu p_{n+1} \,.$$

By induction,

$$p_n = \left(\frac{\lambda}{\mu}\right)^n \cdot p_0 \,.$$

Using as boundary condition $\sum p_n = 1$, we obtain

$$p_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n} \,,$$

hence a necessary condition for the existence of a stabilisation is $\lambda < \mu$. Denote $\rho := \lambda/\mu$. This number is called the **traffic intensity** of the queue.

If $\rho \geq 1$, no steady-state exists; in fact, the queue tends to grow forever, as more and more users accumulate in it.

If, on the contrary, $\rho < 1$, then $p_0 = 1 - \rho$, and we get

$$p_n = \rho^n(1 - \rho)\,,$$

which is the probability of having n users in the system, in the long run.

Knowing the law of the number of users in the system in the long run, it is easy to compute:

- The expectation of the number of users N in the system:

$$\mathrm{E}[N] = \frac{\rho}{1 - \rho}\,.$$

- The expectation of the number of customers N_q in the queue:

$$\frac{\rho^2}{1 - \rho}.$$

- The law of the waiting time T_q in queue:

$$P\{T_q = 0\} = 1 - \rho.$$
$$P\{T_q \leq t\} = 1 - \rho \exp\{-\mu\,(1 - \rho)\,t\} \quad \text{(for } t > 0).$$

- The expectation of T_q:

$$\mathrm{E}[T_q] = \frac{\lambda}{\mu(\mu - \lambda)}.$$

2.13 Complex queueing systems. Simulation

The results above are specific of the M/M/1 queue. There are specific results for other types of queues, and there are also some general results. For in-

stance, the relations

$$\mathrm{E}[N] = \lambda \, \mathrm{E}[T]$$
$$\mathrm{E}[N_q] = \lambda \, \mathrm{E}[T_q] \, ,$$

which one can easily deduce in the M/M/1 queue, are true, no matter the law of arrivals and service times.

However, except for relatively easy queue systems, there is no hope to find analytical results, as computations become intractable very soon. That means that in the real world, one can hardly find closed formulae.

What to do then? One may propose:

- *Idea 1*: Observe the system long enough, take data and do some sort of statistical inference.

- *Idea 2*: Simulate the system in a computer, and do statistical inference as well.

For idea 1 to work, we need the system really running, some mechanism of observation, and a lot of time. In practice, we seldom can afford such luxuries. For idea 2, on the other hand, we only need, essentially, a mechanism to generate random numbers.

There are very good random number generators embodied in software. Their outcome is not really random, but they can fool any detector of "nonrandomness". Anyway, if the quality of a stream of such pseudo-random numbers is a concern, it is very easy to use a true random number generator based in hardware: Nowadays, several internet sites offer potentially infinite streams of true random numbers produced by a quantum device. And such devices are quite cheap, in fact.

2.14 Birth and death processes

A **birth and death process** N_t takes values in $\mathbb{N}$ and the change across an infinitesimal time interval can only be -1, 0, +1:

$$P\{N_{t+h} = n + 1 \,\big/\, N_t = n\} = \lambda_n \cdot h + o(h)\,,$$
$$P\{N_{t+h} = n - 1 \,\big/\, N_t = n\} = \mu_n \cdot h + o(h)\,.$$

This is a generalisation of the M/M/1 queue model to transition probabilities that may depend on the system state.

The corresponding system of differential equations for the state of the system becomes

$$\frac{d}{dt}p_n(t) = -(\lambda_n + \mu_n)\,p_n(t) + \lambda_{n-1}\,p_{n-1}(t)$$
$$+ \mu_{n+1}\,p_{n+1}(t)$$
$$\frac{d}{dt}p_0(t) = -\lambda_0\,p_0(t) + \mu_1\,p_1(t)$$

Birth and death processes have been used, for example, to model the varying size of a biological population under given environmental conditions, or to describe the evolution of an epidemic.

3   Example from industry: Inventories

3.1   Inventory modelling

A company distributes some product, maybe after processing some raw material that arrives to the warehouse. Let us assume that we are dealing only with one product and no processing time. Assume also that the product has an approximately constant level of demand, but the arrival of orders from the clients is not so predictable. The time required to obtain units of prod-

uct from the manufacturer is also subject to some variability.

Two fundamental questions in this situation are:

1. When should more items be ordered?

2. How many items should be ordered when an order is placed?

A couple of things to take into account:

- If a customer wants to purchase but we do not have items, the sale is lost. Therefore, it is important to have enough items in the warehouse.

- The product may become obsolete, and there is also a cost of maintaining the inventory. Therefore, it is not good to keep in storage too many items.

Simple hypothesis for an inventory problem that allow analytical computations similar to the M/M/1 queue are:

- Orders arrive for single items with a random interarrival times following the same law, independent from each other.

- The time to receive items from the manufacturer (**lead times**) follows some law, and are independent, and independent of order arrival.

A commonly used simple strategy is the (r, s)-**policy**: when the inventory drops to r units, order $s - r$ units. One may measure the performance to this policy, given r and s by the average inventory level, or by the average no-inventory time, or by the number of orders that arrive when the inventory is broken, or, most probably, by an combination of

these and other measures that ultimately reduces to a measure of economic benefit that the company wants to maximise.
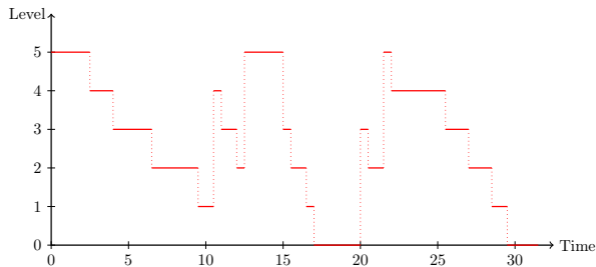

Figure 9: A path of a inventory process. For some time before day 20 and around day 30 the inventory was "empty".

The inventory process, whose paths have the aspect of Figure 9, is not in general a birth and death process: Items may arrive in batches to the warehouse, and the clients' orders may also involve more that one unit. It is therefore a generalisation of the situations seen in the last sections. But it

can still be simulated easily if we know the input distributions.

3.2   Markov chains

We further generalise by abstracting a remarkable property of the inventory process: If we know the state of the system at a particular time t, we do not need to know anything about previous states to predict the future. This is a random analogous of the uniqueness property of deterministic dynamical systems when proper initial conditions are given.

Formally: If $t_1 < \cdots < t_k < t$,

$$P\{N_t = n \ / \ N_{t_1} = n_1, \ \ldots, \ N_{t_k} = n_k\}$$
$$= P\{N_t = n \ / \ N_{t_k} = n_k\}.$$

Stochastic processes satisfying this property are called **Markov chains**, and enjoy an extensive an quite rich theory.

3.3 Chapman–Kolmogorov equation

Consider times $0 \le u < t < s$. Recall the notation of Section 2.11 for the transition probabilities.

The **Chapman–Kolmogorov equation** for Markov chains establishes that the probability of going from state n to state m when time runs from u to s can be computed by decomposing all possible paths at the intermediate time t:

$$p_{n,m}(u,s) = \sum_k p_{n,k}(u,t)\, p_{k,m}(t,s)\,.$$

We have already used this in Section 2.11.

In particular, the law of the random variable N_{t+h} can be obtained from the law of N_t and the transition probabilities from t to $t+h$:

$$p_{n,m}(0, t+h) = \sum_k p_{n,k}(0,t)\, p_{k,m}(t, t+h)\,,$$

$$\sum_n p_n(0)\, p_{nm}(0, t+h) = \sum_k \sum_n p_n(0)\, p_{n,k}(0,t)\, p_{k,m}(t, t+h)$$

$$p_m(t+h) = \sum_k p_k(t)\, p_{k,m}(t, t+h)\,.$$

3.4 Kolmogorov forward and backward equations

Assume

$$1 - p_{n,n}(t, t+h) = q_n(t)\,h + o(h)\,,$$

$$p_{n,m}(t, t+h) = q_{n,m}(t)\,h + o(h) \quad (\text{for } n \neq m),$$

for some continuous functions q_n and q_{nm}. Then, the following two relations hold:

$$\frac{\partial}{\partial t} p_{n,m}(u,t) = q_m(t)\,p_{n,m}(u,t) + \sum_{k \neq m} p_{n,k}(u,t)\,q_{k,j}(t)\,,$$

$$\frac{\partial}{\partial u} p_{n,m}(u,t) = q_n(u)\,p_{n\,m}(u,t) - \sum_{k \neq n} q_{n,k}(t)\,p_{k,m}(u,t)\,.$$

These differential equations for the transition probabilities are known as **Kolmogorov equations**, forward and backward, respectively.

3.5 Differential equations for the laws

Assume that the functions q_n and $q_{n,m}$ above are constant: $q_n(t) \equiv q_n$ and $q_{n,m}(t) \equiv q_{n,m}$. The Markov chain is then called **time-homogeneous**.

From Kolmogorov forward equations, letting $u = 0$, multiplying by $p_n(0)$ and summing over n, one obtains a (infinite) system of differential equations for the laws of N_t:

$$\frac{d}{dt} p_m(t) = -q_m \, p_m(t) + \sum_{k \neq m} p_k(t) \, q_{k,j} \, .$$

3.6 Long-run behaviour of Markov chains

In many applications it is of interest to study the behaviour of the chain *in the long run*. For instance:

- Limiting distributions: Assume that the limits $\lim_{t \to \infty} p_{nm}(u, t)$ exist and are equal, for

all n. That means, the limit is independent of the initial state, when time is large. The limit is a probability law called the **limiting** or **steady-state distribution** of the Markov chain.

- Stationary distributions: If the limit of the laws $\{\lim_{t \to \infty} p_n(t)\}_n$ exists, it is called the **stationary distribution** of the chain. If there is a limiting distribution, then it coincides with the stationary distribution. But the latter may exist independently.

- Ergodicity: Loosely speaking, **ergodictiy** means that some kind of information that can be extracted from a process as a whole, can also be obtained by observing one single path. For instance, ergodicity with respect to the expectation means that the limit

$\lim_{t \to \infty} \mathrm{E}[X(t)]$ coincides with

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t X(s)\, ds$$

for all sample paths $X(s)$. For example, the M/M/1 queue, with traffic intensity $\rho < 1$, satisfies this property.

In particular, ergodicity implies that simulating one only sample path for long enough time is sufficient to estimate the expectation of the process in the long run.

- Classification of states: The elements of the state space of Markov chains are classified according to different interwoven criteria. Among the most important concepts: A state is **transient** if the probability to never returning to it is positive; otherwise it is called **recurrent**, and the process will certainly visit that state an infinite number

of times; a state is **absorbing** if the chain never leaves it once it is reached.

3.7 Stochastic processes in discrete time

A **discrete time stochastic process** is a process where the family of random variables is indexed by a discrete set, usually $\mathbb{Z}$ or $\mathbb{N}$.

A **discrete time Markov chain** has the same definition of a Markov (continuous time) chain, except that the index t runs over a discrete set, usually the non-negative integers.

Another important class of stochastic process in discrete time is the **time series**, that models a different sort of dependency between variables. Figure 10 shows the monthly evolution of the number of passengers of international airlines between January 1949 and December 1960. One observes a *trend* (increasing), a *seasonality* (peaks at the cen-
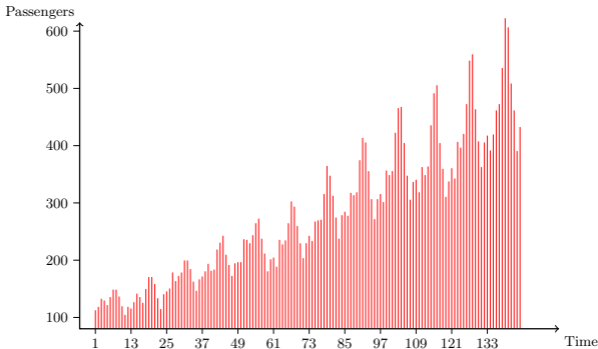
Figure 10: A time series: A discrete time stochastic process with 144 values corresponding to the number of airlines passengers (in thousands) between 1949 and 1960.

tral months of the year) and a residual *noise* (the purely random component of the process). Usually, one tries to fit a suitable model of dependence between the variables, so that the original process is expressed as the sum of these individual components.

4 Example from biology: genes

4.1 Genotype and gene frequencies

Alleles are several forms that a gene may have in a particular place (**locus**) of a chromosome.

For example, sheep haemoglobin presents two forms, produced by two alleles, A and B, of a certain locus. Each individual possesses chromosomes in pairs, one coming from each parent. This implies that there are three possible **genotypes**: AA, AB, BB.

Typically, one allele is **dominant**, while the other is **recessive**. The recessive allele shows up externally in the **phenotype** only if the dominant is not present.

Assume we extract blood from a population of N sheep, and the geno-

Some sheep

types appear in proportions P_{AA}, P_{AB} and P_{BB} (called **genotypic frequencies**). The **gene frequencies** are the proportions of the two alleles:

$$P_A := P_{AA} + \frac{1}{2}P_{AB}$$
$$P_B := P_{BB} + \frac{1}{2}P_{AB}. \tag{3}$$

4.2 Hardy-Weinberg principle

Assume that:

- The proportions are the same for males and females.

- The genotype does not influence mating preferences.

- Each allele of a parent is chosen with equal probability $1/2$.

Then, the probabilities of each mating are, approximately (assuming a large population):

$$P(AA \text{ with } AA) = P_{AA}^2 \,,$$
$$P(AB \text{ with } AB) = P_{AB}^2 \,,$$
$$P(BB \text{ with } BB) = P_{BB}^2 \,,$$
$$P(AA \text{ with } AB) = 2\, P_{AA}\, P_{AB} \,,$$
$$P(AA \text{ with } BB) = 2\, P_{AA}\, P_{BB} \,,$$
$$P(AB \text{ with } BB) = 2\, P_{AB}\, P_{BB} \,.$$

We can deduce easily the law of the genotypes for the next generation:

$$Q_{AA} = P_{AA}^2 + \frac{1}{2}\, 2\, P_{AA}\, P_{AB} + \frac{1}{4}\, P_{AB}^2 = P_A^2 \,,$$
$$Q_{BB} = P_{BB}^2 + \frac{1}{2}\, 2\, P_{BB}\, P_{AB} + \frac{1}{4}\, P_{AB}^2 = P_B^2 \,,$$
$$Q_{AB} = 2\, P_A\, P_B \,.$$

Computing the gene frequencies Q_A and Q_B with (3) we find again P_A and P_B, so the genotype

frequencies must be constant from the first generation onwards. This is the **Hardy-Weinberg principle** (1908).

As an application of this principle, suppose B is recessive and we observe a 4% proportion of individuals showing the corresponding phenotype. Then we can deduce the genotype proportions of the whole population:

$$4\% = P_{BB} = P_B^2, \ \Rightarrow P_B = 20\%, \ P_A = 80\%,$$
$$P_{AA} = 64\%, P_{AB} = 32\%.$$

If the population were small, then randomness in the mating may lead to **genetic drift**, and eventually one of the alleles will disappear from the population. The other gets *fixed*, and time to fixation is one of the typical things of interest. This purely random fact explains the lost of genetic diversity in closed small populations.

4.3 Wright-Fisher model (1931)

If the mating is completely random, it does not matter how the alleles are distributed among the N individuals. We can simply consider the population of $2N$ alleles.

Assume that at generation 0 there are X_0 alleles of type A, with $0 < X_0 < 2N$. We pick alleles from this population independently from each other $2N$ times to form the N individuals of generation 1.

The law of the number of alleles of type A must be Binom$(2N, p)$, with $p = X_0/2N$. Thus

$$P\{X_1 = k\} = \binom{2N}{k}(X_0/2N)^k (1 - X_0/2N)^{2N-k},$$

$$k = 0, \ldots, 2N.$$

In general, the number of alleles A in generation $n+1$ knowing that there are j in generation n is

$$P\{X_{n+1} = k \,/\, X_n = j\} = \binom{2N}{k}(j/2N)^k(1 - j/2N)^{2N-k}.$$

This defines a Markov chain in discrete time. Its expectation is constant, $E[X_n] = E[X_0]$, and the expectation of the random variable X_n, knowing that a past variable X_m $(m < n)$ has taken value k, is equal to k:

$$E\left[X_n \,\middle/\, X_m = k\right] = k, \quad k = 0, \ldots, 2N. \quad (4)$$

However, as we saw in Section 4.2, the process will eventually reach states 0 or $2N$, and it will remain there forever. They are absorbing states (see Section 3.6).

4.4 Conditional expectation

The expression on the right-hand side of (4) is called the **conditional expectation** of X_n given that $X_m = k$. It is exactly the expectation of X_n computed from the conditional probability to the

event $\{X_m = k\}$. One may write (4) as

$$\mathrm{E}\left[X_n \, \big/ \, X_m\right] = X_m$$

and the left-hand side is now a random variable instead of a single number, called the **conditional expectation** of X_n given X_m. For each $\omega \in \Omega$, the random variable $\mathrm{E}\left[X \, \big/ \, Y\right](\omega)$ is equal to the number $\mathrm{E}\left[X \, \big/ \, Y = y\right]$, if $Y(\omega) = y$.

In case the conditioning random variable Y has a continuous law, the definition above does not work, since $\{Y(\omega) = y\}$ is an event of probability zero. The intuitive meaning is however the same. Mathematically, the trick is not to consider the y individually, but collectively: The conditional expectation $\mathrm{E}\left[X \, \big/ \, Y\right]$ is given a sense as the (unique) random variable that can be factorized as a composition $(\varphi \circ Y)(\omega)$, with $\varphi \colon \mathbb{R} \to \mathbb{R}$, and whose expectation, restricted to the events of Y, coincides

with that of X:

$$\mathrm{E}[(\varphi \circ Y) \cdot \mathbf{1}_{\{Y \in B\}}] = \mathrm{E}[X \cdot \mathbf{1}_{\{Y \in B\}}],$$

where $\mathbf{1}_{\{Y \in B\}}$ is equal to 1 if $Y(\omega) \in B$ and 0 otherwise.

4.5 Continuous approximation of discrete laws

Discrete laws involve only elementary discrete mathematics, but they are sometimes cumbersome with computations. For instance, computing exactly the probability density of a $\mathrm{Binom}(n, p)$ distribution when n is large involve making the computer work with higher precision than usual. Although nowadays this is not a big deal (unless n is really very large), it is still useful, and conceptually important, to use continuous laws as a proxy to the real distribution.

Specifically, for the binomial law: If $X \sim$ Binom(n, p), then

$$\frac{X - n\,p}{\sqrt{n\,p\,(1-p)}} \sim N(0, 1) \ , \ \text{(approximately, for } n \text{ large)}$$

where $N(0, 1)$ denotes the so called **Normal** (or **Gaussian**) **law** with expectation 0 and variance 1. Its density function is the *Gaussian bell curve*

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

Figure 11 shows graphically the approximation.

The importance of the Gaussian law comes from the **Central Limit Theorem**, which explains its ubiquity: If X_n is a sequence of independent identically distributed random variables, with finite variance, and $S_n := \sum_{i=1}^{n} X_i$, then

$$\frac{S_n - \mathrm{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \quad \textit{converges in law to} \quad N(0, 1) \ .$$
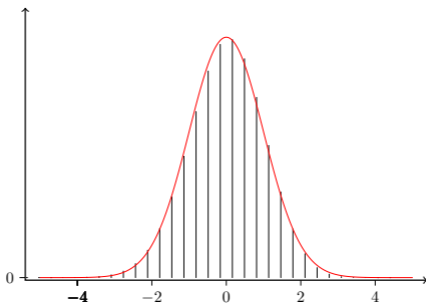
Figure 11: Approximation by the Gaussian law: The probability density of Binom($n = 45, p = 0.7$), after subtracting its expectation (*centring*) and dividing by the square root of its variance (*reducing*), depicted with black vertical lines. In red, the density function of $N(0, 1)$.

We immediately see that the binomial case above is a particular application of this theorem, taking $X_i \sim \text{Binom}(1, p)$, which implies $S_n \sim \text{Binom}(n, p)$. Convergence in law is a non-elementary concept that has to do with duality in

functional spaces: Suppose that $\{Y_n\}$ is a sequence of random variables with respective distributions P_n, and Y is a random variable with distribution P. Then, we say that $\{Y_n\}$ **converge in law to** P if for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n \to \infty} \mathrm{E}[f(Y_n)] = \mathrm{E}[f(Y)].$$

The seemingly natural "setwise" convergence $\lim_{n \to \infty} P_n(A) = P(A)$ for all sets A is too strong, and will not work for the purpose of approximating by continuous distributions.

One practical consequence of the Central Limit Theorem for modelling is that any phenomenon whose result is the sum of many quantitatively small causes (like for instance the height or the weight of a person) will be well described by a Gaussian random variable. The fact that the Gaussian laws may take values in any interval of the real line is not an obstacle due to the rapid decrease of

the bell curve: Outside a short interval, the probability is extremely small.

4.6 Random walk and the Wiener process

Let $\{X_n\}_n$ be a Markov chain taking values in $\mathbb{Z}$ with

$$X_0 = 0 \,,$$
$$P\{X_{n+1} = i + 1 \,\big/\, X_n = i\} = 1/2 \,,$$
$$P\{X_{n+1} = i - 1 \,\big/\, X_n = i\} = 1/2 \,.$$

This process is called **random walk**. It is simply a "walk" on the integer lattice, where at each time step we go to the left or to right according to the toss of a fair coin. In other words, the increments $\varepsilon_n := X_n - X_{n-1}$ are independent and take values 1 and -1 with probability $1/2$.

Define a sequence of continuous-time process

W_t^N by renormalisation of a random walk:

$$W_t^N = \frac{1}{\sqrt{N}} X_{\lfloor N t \rfloor} = \frac{1}{\sqrt{N}} \sum_{k=1}^{\lfloor N t \rfloor} \varepsilon_k \,.$$

By the Central Limit Theorem,

$$= \frac{1}{\sqrt{N t}} \sum_{k=1}^{\lfloor N t \rfloor} \varepsilon_k \quad \text{converges in law to} \quad N(0,1) \,,$$

hence the sequence $\{W_t^N\}_N$ converges in law to a random variable $W_t \sim N(0,t)$, the Gaussian law with variance t, for all $t > 0$, whose density is

$$f(x) = \frac{1}{\sqrt{2 \pi t}} \, e^{-x^2/2t} \,.$$

Analogously, $W_t^N - W_s^N$ converges in law to $W_t - W_s \sim N(0, t-s)$. The limiting process W_t satisfies:

1. The increments in non-overlapping intervals are independent.

2. The expectation is constant equal to zero.

3. The sample paths are continuous functions.

4. The sample paths are non-differentiable at any point.

W is called the **Wiener process**. In fact, a Wiener process is defined by its laws, but usually it is additionally asked to have continuous paths. This particular construction as the limit of random walks leads indeed to continuous paths.

The Wiener process is also called **Brownian Motion** in the mathematical literature. However, the Brownian motion is a physical phenomenon, and the Wiener process is just a mathematical model (and not the best one) to that phenomenon.

4.7 Diffusion approximation of Wright-Fisher model

The Markov chain of the Wright-Fisher model is too complicated to work upon. Instead, define

$Y_t^N = \frac{1}{2N} X_{\lfloor 2Nt \rfloor}^N$. Then,

$$\mathrm{E}\left[(Y_{t+h}^N - Y_t^N)^2 \,\middle/\, Y_t^N = x\right]$$
$$= \left(\frac{1}{2N}\right)^2 \mathrm{E}\left[(X_{\lfloor 2N(t+h)\rfloor}^N - X_t^N)^2 \,\middle/\, X_t^N = x\right]$$
$$= \frac{1}{2N} x\,(1-x)\,, \quad (\text{if } h \sim \frac{1}{2N}).$$

The limiting process Y_t exists, satisfies

$$\mathrm{E}\left[(Y_{t+h} - Y_t)^2 \,\middle/\, Y_t = x\right] = h\,x\,(1-x) + o(h)$$

and it is called the **diffusion approximation** of the original Markov chain.

4.8 Diffusions

A **diffusion** Y is a continuous-time Markov process, with continuous paths, and such that

1. $\mathrm{E}[Y_{t+h} - Y_t \,/\, Y_t = x] = b(t,x)\,h + o(h)\,,$

2. $\mathrm{E}[(Y_{t+h} - Y_t)^2 \,/\, Y_t = x] = a(t,x)h + o(h)$

for some functions a and b. See Section 4.4 for the interpretation of the conditional expectations when the conditioning variable is continuous.

Under mild conditions, Y_t has a continuous law with density $f(t,x)$ satisfying the **Kolmogorov forward and backward equations**:

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[a(t,x) f(t,x) \right] - \frac{\partial}{\partial x} \left[b(t,x) f(t,x) \right],$$
(5)

$$-\frac{\partial}{\partial s} f(t,x) = \frac{1}{2} a(t,x) \frac{\partial^2}{\partial x^2} f(t,x) - b(t,x) \frac{\partial}{\partial x} f(t,x).$$
(6)

The Wright–Fisher model can be expanded to take into account other effects in population dynamics, such as selection or mutation. This complications make even more useful the corresponding diffusion approximations.

5 Example from economy: stock markets

5.1 A binomial economy

Assume an economy with only two states:

- Up (with probability p),

- Down (with probability $1 - p$).

Assume that there are two assets:

- A **risk-free bond** with interest rate R, and

- A **share** with price $S(0)$ at time 0 and $S(1)$ at time 1, given by

$$S(1) = \begin{cases} S(0)u, & \text{if the economy is "up"} \\ S(0)d, & \text{if the economy is "down"} \end{cases}$$

A **trading strategy** for a portfolio is defined by

- $B_0 \in$ allocated to the bond, and

- Δ_0 quantity of shares of the stock at time zero.

The values of the portfolio at times 0 and 1 are

$$V(0) = B_0 + \Delta_0 S(0)$$
$$V(1) = B_0(1 + R) + \Delta_0 S(1)$$

Kuwait stock market

5.2   Free lunch?

As we will see, one can make money for free, unless $d < 1 + R < u$. An **arbitrage opportunity** is the situation in which, without investing any money at time zero, the probability to have a positive portfolio at time one is positive, and the probability of a loss is zero.

For a couple (B_0, Δ_0) such that $V(0) = 0$,

$$V(1) = B_0(1 + R) + \Delta_0 S(1) = \begin{cases} B_0(1 + R) + \Delta_0 u S(0) \\ B_0(1 + R) + \Delta_0 d S(0) \end{cases}$$

$$= \begin{cases} \Delta_0 S(0) \cdot [u - (1 + R)] \\ \Delta_0 S(0) \cdot [d - (1 + R)] \end{cases}$$

with respective probabilities p and $1 - p$.

If $(1 + R) < d$, both quantities are positive and we could borrow money to buy assets to have a sure win. If $(1 + R) > u$, both quantities are negative and we could make money by selling assets and buying bonds. If $V(0) \neq 0$, the argument is equally

valid.

The arbitrage situation is not realistic if all the actors have complete information. Thus, usually there is no free lunch!

5.3 European options

An **European call option** is a financial derivative: It gives the holder the right (not the obligation) to buy a share for an pre-specified amount (**exercise price** K) on a specific later date (**expiry date** T). Similarly, an **European put option** is the right to sell the share.

If $S(T)$ is the value of the share at time T, the payoff of a call is $\left(S(T) - K\right)^+$. If $S(T) < K$, the holder does not exercise the option, since it can buy the share in the market for a cheaper price, so the payoff is never negative.

Correspondingly, the payoff of a put is $\left(K - \right.$

$S(T))^+$, see Figure 12.
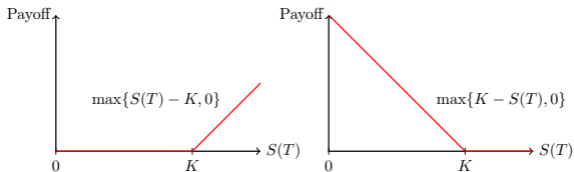
Figure 12: Graphs of an European call and an European put

5.4 Fair price of an European call option. Example

Assume the following data:

- Current price of the share: $S(0) = 100$,

- Interest of the risk-free bond: 10%,

- Possible prices for the share at time 1: 120 or 90,

- Exercise price: $K = 100$.

We have $u = 1.2$, $d = 0.9$, $1 + R = 1.1$.
The payoff will be $C_u := 20$ or $C_d := 0$.

To find the fair price, let us construct a portfolio with a value $V(1)$ equal to the payoff of the option. The fair price will be $V(0)$.

$$V(1) = \begin{cases} B_0 \cdot 1.1 + \Delta_0 S(0) \cdot 1.2 = 20 \\ B_0 \cdot 1.1 + \Delta_0 S(0) \cdot 0.9 = 0 \end{cases}$$
$$\Rightarrow \Delta_0 S(0) = 66.67 , \quad B_0 = -54.55$$

The fair price is thus 12.12 .

5.5 Fair price of an European call option. In general

In general, we have

$$V(1) = \begin{cases} B_0\,(1+R) + \Delta_0\,S(0)\,u = C_u \\ B_0\,(1+R) + \Delta_0\,S(0)\,d = C_d \end{cases}$$

$$\Rightarrow B_0 = \frac{u\,C_d - d\,C_u}{(1+R)\,(u-d)},$$

$$\Delta_0\,S(0) = \frac{C_u - C_d}{u-d},$$

$$\Rightarrow B_0 + \Delta_0\,S(0) = (1+R)^{-1}\big(C_u\,q + C_d\,(1-q)\big),$$

$$\text{where } q = \frac{1+R-d}{u-d}.$$

It follows that the fair price of the option is the expected (and discounted!) payoff of the option under the probability $Q = (q, 1-q)$ for the states of the economy:

$$\mathrm{E}_Q\left[(1+R)^{-1}\big(S(1) - K\big)^+\right]. \qquad (7)$$

Some remarks on the probability Q:

- Under Q, the share and the bond have the same expected return:

$$\begin{aligned}
\mathrm{E}_Q[S(1)] &= S(0)\, u\, q + S(0)\, d\, (1-q) \\
&= S(0)\Big(u\, \frac{1+R-d}{u-d} + d\, \frac{u-1-R}{u-d} \Big) \\
&= S(0)\, (1+R)\,.
\end{aligned}$$

- The probability Q does not depend on the underlying probability $P = (p, 1-p)$ nor on the payoff of the option.

- Q is called the **risk-neutral probability** (or **martingale probability**).

5.6 Fair price of an European call option. Example (cont.)

With the same data as before, we compute now the fair price directly using formula (7), where in this case $q =: 2/3$.

$$E_Q\left[(1+R)^{-1}\left(S(1)-K\right)^+\right]$$
$$= (1+R)^{-1}\left(C_u \cdot q + C_d \cdot (1-q)\right)$$
$$= \frac{1}{1.1}\left[20 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3}\right] = 12.12\,.$$

Assume now that the exercise price is fixed to $K = 95$ instead of $K = 100$, while all other data remain the same. Logically, the option should be more expensive in this case. Applying again formula (7),

$$E_Q\left[(1+R)^{-1}\left(S(1)-K\right)^+\right]$$
$$= (1+R)^{-1}\left(C_u \cdot q + C_d \cdot (1-q)\right)$$
$$= \frac{1}{1.1}\left[25 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3}\right] = 15.15\,.$$

5.7 European call option. Multiperiod

The previous sections dealt with a single time period. Assume now that the expiry time of the option is T and that we can change the composition of the portfolio at any of the intermediate integer times.

A **trading strategy** is then $\{(B_t, \Delta_t),\ 0 \le t \le T-1\}$. It is called a **self-financing strategy** if we do not put new money or take money out of the portfolio.

At time t, we can change the portfolio composition, but the value remains the same:

$$B_t + \Delta_t\, S(t) = B_{t+1} + \Delta_{t+1}\, S(t)\,.$$

The new value at time $t+1$ will be:

$$B_{t+1}\,(1+R) + \Delta_{t+1}\, S(t+1)\,.$$

Therefore, the value increments for a self-financing strategy is

$$V(t+1) - V(t) = B_{t+1} R + \Delta_{t+1}\big(S(t+1) - S(t)\big).$$

We can compute the fair price $F(0)$ at time 0 of an option with exercise value K at expiry date T recursively:

$$F(T-1) = \mathrm{E}_Q\left[(1+R)^{-1}(S(T) - K)^+ \big/ S(T-1)\right],$$
$$F(T-2) = \mathrm{E}_Q\left[(1+R)^{-1}F(T-1) \big/ S(T-2)\right]$$
$$= \mathrm{E}_Q\left[(1+R)^{-2}(S(T) - K)^+ \big/ S(T-2)\right],$$
$$\vdots$$
$$F(0) = \mathrm{E}_Q\left[(1+R)^{-T}(S(T) - K)^+\right].$$

This computation uses essential properties of the conditional expectation that we are not going to detail here. But the conclusion must be quite intuitive.

5.8   Martingales

Under probability Q, the stochastic process $(1 + R)^{-t} S(t)$ enjoys the martingale property. A stochastic process $\{X_t, t \geq 0\}$ is a **martingale** if

$$\mathrm{E}\left[X_t \,/\, X_s\right] = X_s \quad \text{whenever } s < t, \quad (8)$$

meaning that the knowledge of the state of the system at time s makes this the expected value at any later time. The discrete time process defined in Section 4.3 is a discrete time martingale (see Equation (4)).

Martingales are good models for *fair games*: The expected wealth of a player in the future is the current wealth, no matter what happened before, or how long has been playing.

From (8) it can be deduced in particular that the expectation of the process is constant in time.

In our case of the European call option, this means

$$\mathrm{E}_Q\left[(1+R)^{-t}\,S(t)\right] = S(0)\,,$$

implying that

$$\mathrm{E}_Q\left[S(T)\right] = S(0)\cdot(1+R)^T$$

which is precisely the return of the risk-free bond (and this is why Q is called a "risk neutral" probability measure).

5.9 European call option. Continuous time

In continuous time, it can be shown that there is also a probability Q under which $e^{-Rt}S(t)$ is a martingale, and the fair price at time 0 of a call option is given by

$$F(0) = \mathrm{E}_Q\left[e^{-RT}\left(S(T) - K\right)^+\right],$$

although Q is more difficult to describe here.

The evolution of the value of the bond asset $I(t)$ is driven by the well-known differential equation

$$dI(t) = R \cdot I(t)\, dt\,.$$

The evolution of the price of the share can be described as

$$dS(t) = S(t)\big(\mu\, dt + \sigma\, dW(t)\big) \qquad (9)$$

where W is a Wiener process, approximating (in the continuum limit) the Markov chain given by the binomial model. The trend, if $p \neq 1/2$, goes to the **drift** μ. The **volatility** σ is the intensity of the noise. This is a simple example of a **stochastic differential equation**. It is a pathwise description of a diffusion process with $b(t,x) \equiv \mu$ and $a(t,x) \equiv \sigma^2$ (see Section 4.8).

Although the paths of W are non-differentiable

everywhere, Equation (9) has the obvious meaning

$$S(t) = S(0) + \mu \int_0^t S(r)\,dr + \sigma\,W(t)\,.$$

This equation can be solved explicitly (this is not common, of course). The solution is the stochastic process given by

$$S(t) = S(0)\,\exp\left\{\mu\,t - \frac{1}{2}\,\sigma^2\,t + \sigma\,W(t)\right\},$$

and we can compute its law from here.

The evolution of the whole portfolio value will be

$$dV(t) = B_t\,dI(t) + \Delta_t\,dS(t)\,.$$

5.10   Stochastic differential equations

In general, a diffusion process X with characteristic functions $a(t,x)$ and $b(t,x)$ (called respectively **diffusion** and **drift coefficients**) can be repre-

sented pathwise by means of the stochastic differential equation

$$dX(t) = b(t, X(t)) \, dt + a(t, X(t))^{1/2} \, dW(t) \, ,$$

with a suitable definition of the last term, which in general, when the function a depends effectively of its second argument, does not possess an obvious meaning.

Diffusions can therefore be studied at the same time with the tools of partial differential equations that describe the evolution of the laws in time, and with the tools of stochastic processes and stochastic differential equations, that provide the evolution of the paths themselves.

The word "diffusion" is taken from the physical phenomenon with that name: The movement of particles in a fluid from regions of high concentration to regions of low concentration. The heat "diffuses" in the same way, following the negative

gradient of the temperature field $f(t, x)$. In one space dimension, it obeys the partial differential equation

$$\frac{\partial}{\partial t} f(t, x) = D \, \frac{\partial^2}{\partial x^2} f(t, x) \,,$$

where D is called the *thermal diffusivity*. Comparing with Kolmogorov equations (5-6), we see that, with suitable initial conditions, $f(t, x)$ is the density at time t and point x of a diffusion process following the stochastic differential equation

$$dX(t) = \sqrt{2 \, D} \; dW(t) \,,$$

that means, essentially, the Wiener process.

6   Recommended books

- Nelson, Stochastic Modeling, Dover 1995 (Arrivals, queues, Markov chains, simulation)

- Gross-Harris, Fundamentals of queueing theory, Wiley 1998
  (Queues)

- Asmussen-Glynn, Stochastic simulation, Springer 2007
  (Simulation)

- Maruyama, Stochastic problems in population genetics, Springer 1977
  (Diffusions, application to genetics)

- Lamberton-Lapeyre, Introduction au calcul stochastique appliqué à la finance, Ellipses 1997
  (Diffusions, stochastic differential equations, application to finance)

Contents

Dedication: Io non crollo

On April 6th 2009, a strong earthquake destroyed the building where the course was given the year before, and caused more than 300 deaths in the region.

This work is dedicated to the people that died, lost a beloved one, or lost their homes that day. I adhere to the motto that helped the university people to carry on after the disaster:

Io non crollo.

Photo: Renato di Bartolomeo

Department of Mathematics

Universitat Autònoma de Barcelona

Bellaterra, Catalonia

Aureli.Alabert@uab.cat