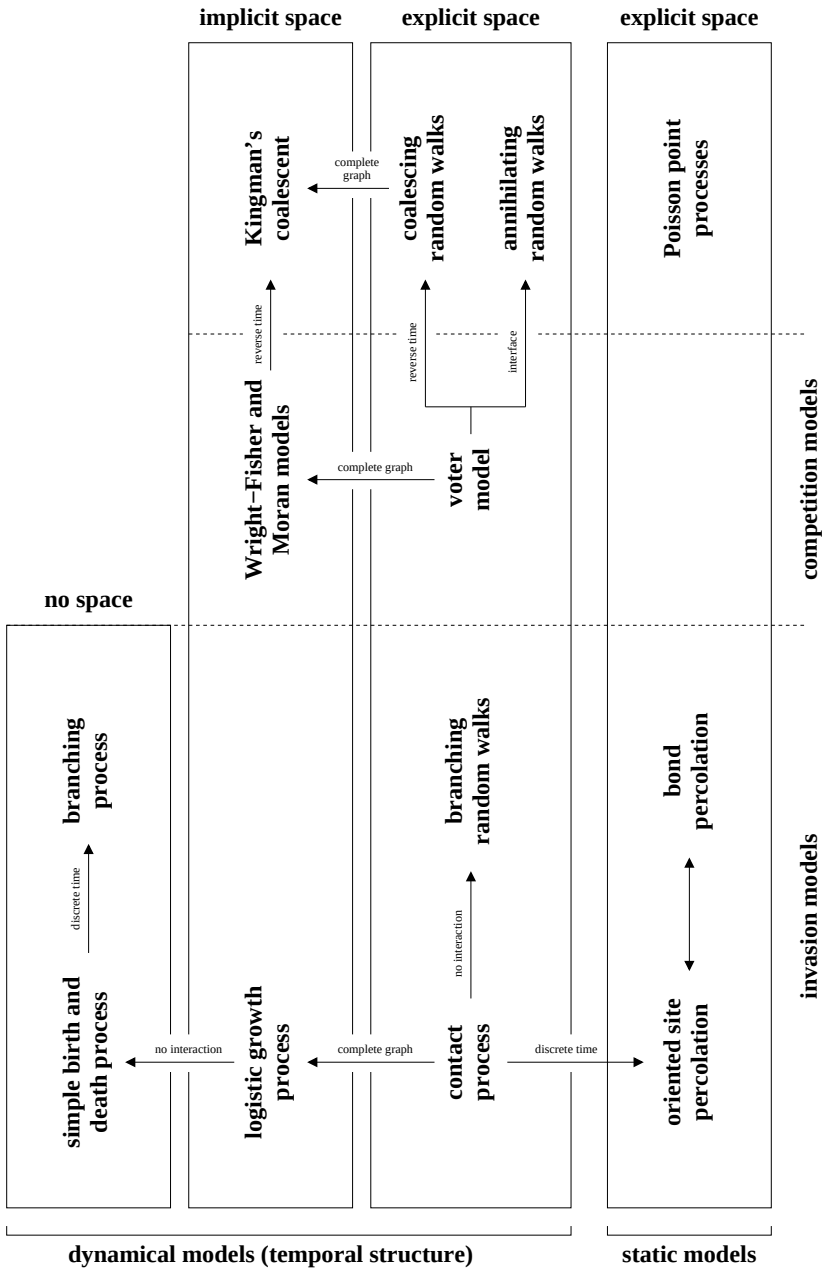
APM 541 – Stochastic modeling

Nicolas Lanchier
School of Mathematical and Statistical Sciences, Arizona State University.

Abstract – These notes give a brief overview of measure theory and stochastic processes before exploring key models that arise from physics, biology and sociology. In contrast with traditional textbooks on this topic, we review a wide variety of modeling perspectives, each of which being illustrated with a simple model, rather than focusing on the analysis of an increasingly complex sequence of processes derived from the same modeling approach. Here, by simple model, we do not mean simple to study but minimal model with the smallest possible number of parameters simple to formulate. This choice intends to promote the learning of important techniques. First, this allows one to review a more diversified set of techniques since the analyses of different processes derived from the same modeling approach are somewhat similar whereas they strongly differ between different modeling perspectives. Moreover, techniques can be better understood looking at simple models since this prevents the true spirit of the proofs from being hidden behind lengthy calculations. In particular, proofs are not provided with the intention of completeness but as illustrations of these techniques. But the main reason for looking at a wide variety of modeling perspectives is to have the opportunity to exhibit important connections among them: even though the models we review emerge from different modeling perspectives, they are general invasion and competition models that mimic similar phenomena and only differ in their level of complexity but are otherwise connected. Understanding how these models are connected in terms of their level of complexity is obviously important from a modeling point of view, but also from a mathematical point of view by making powerful techniques that have been developed for complex modeling perspectives available for less complex ones.

Table of contents



1. Basics of measure and integration theory

To begin with, we give a short description of the most basic concepts of measure and integration theory since modern probability is built from this framework. The use of concepts from measure theory in the context of probability theory is due to Kolmogorov. Measure theory and probability theory mostly differ in their terminology and interpretation so we explain both at once. A good reference in measure theory is Rudin [23] but see also Durrett [12] for a more probabilistic approach and additional results about conditional expectation. To explain why measure theory is useful in probability theory, consider the following simple problem: Find the probability that a number chosen uniformly at random in the unit interval is rational. To answer this question, let

$$X \;:=\; \text{Uniform}\,(0,1) \quad \text{and} \quad f_X \;:=\; \mathbf{1}_{(0,1)}$$

be the density function of X. Then, the probability to be found is

$$P\,(X \in \mathbb{Q}) \;=\; \int_{\mathbb{Q}} f_X(x)\,dx \;=\; \int_0^1 \mathbf{1}_{\mathbb{Q}}(x)\,dx. \tag{1.1}$$

The function $\mathbf{1}_{\mathbb{Q}}$ is known as the Dirichlet function. To find its integral, recall that the Riemann sum of a function is defined from a tagged partition of the domain of the function as illustrated on the left-hand side of Figure 1. The function is said to be Riemann-integrable if and only if the sequence of the Riemann sums converges to a limit that does not depend on the choice of the sequence of the tagged partitions as the partitions get finer and finer. In particular, since each open intervals contains infinitely many rational numbers and infinitely many irrational numbers, the Dirichlet function is not Riemann-integrable so the integral in (1.1) does not exist. Using the framework of measure theory, however, we can properly define and compute this integral, as it is done at the end of the next subsection.

Construction of the abstract integral. This subsection is devoted to the construction of what we shall call the abstract integral, as opposed to the Riemann integral. This new concept of integral has been introduced by Henri Lebesgue in his 1902 Ph.D. dissertation in an attempt to generalize and make more flexible the Riemann integral. Similarly to the Riemann integral, the abstract integral is a linear operator defined on a set of functions. The Lebesgue integral, however, is more powerful for the following reasons. First, it is much more general than the Riemann integral, namely there is one integral associated to each so-called measure. Second, the special case of the Lebesgue measure gives rise to an integral that extends the Riemann integral to a much larger set of functions, including the Dirichlet function. To construct this integral, we first need a few definitions.

Definition 1.1 – Let Ω be a set. From the point of view of probability theory, one thinks of this set as a **sample space**: set of all possible **outcomes** of a random experiment.

1. A collection $\mathcal{F}$ of subsets of Ω is said to be a **σ-algebra** whenever
   - we have $\Omega \in \mathcal{F}$
   - for all $A \in \mathcal{F}$, we have $A^c \in \mathcal{F}$
   - for each sequence $(A_n) \subset \mathcal{F}$, we have $\bigcup_n A_n \in \mathcal{F}$.
2. The pair $(\Omega, \mathcal{F})$ is then called a **measurable space**.
3. Members of the σ-algebra are called **measurable sets** in measure theory and are interpreted as **events** in probability theory.

It is obvious from the definition that the σ-algebra also contains the empty set and is stable under countable intersections. Typically, any set obtained from elementary set operations involving countably many measurable sets is again measurable. In the context of probability theory, the σ-algebra or set of all possible events represents the **information available**: the largest σ-algebra, which consists of all subsets of the sample space, means perfect information whereas the smallest one, which reduces to the sample space and the empty set, means no information. It can be proved that, for any collection $\mathcal{H}$ of subsets of Ω, there exists a smallest σ-algebra that contains $\mathcal{H}$. It is called the **σ-algebra generated by the collection $\mathcal{H}$** and it is written $\sigma(\mathcal{H})$. The usual σ-algebra on the real line is the one generated by the open intervals:

$$\mathcal{B} \;:=\; \sigma\,\{(a,b) : a,b \in \mathbb{R}, a < b\} \tag{1.2}$$

and is called the **Borel σ-algebra**. More generally, the Borel σ-algebra on a topological space is the one generated by the open sets. In view of the large collection of sets that can be produced using elementary set operations involving countably many open intervals, it seems that any subset of the real line is a Borel set. In fact, a result due to Vitali implies that there exists uncountably many subsets of the real line, called **Vitali sets**, which are not Borel sets. His proof, however, relies of the axiom of choice and does not give an explicit construction of such sets. The next step to construct the Lebesgue integral is to define the concept of **measurability** of a function, which can be seen as the analog of continuity in topology.

Definition 1.2 – A function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is called a **measurable function** in measure theory and a **random variable** in probability theory whenever

$$X^{-1}(B) \;:=\; \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{for all} \quad B \in \mathcal{B}. \tag{1.3}$$

The set of all measurable functions is denoted by $\mathcal{M}(\Omega, \mathcal{F})$.

Since the Borel σ-algebra (1.2) is generated by the open intervals, it can be proved that a function is a measurable function/random variable, if and only if the inverse image of any open interval is measurable. In fact, to prove that X is measurable, it is even enough to prove that

$$X^{-1}((-\infty, a]) \;:=\; \{\omega \in \Omega : X(\omega) \le a\} \;\in\; \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{1.4}$$

Using this characterization, one can prove that the positive part, negative part and absolute value of a measurable function are again measurable. Likewise, the supremum, infimum, lim sup and lim inf of a sequence of measurable functions are measurable. It is easy to check that the collection of the inverse images of the Borel sets by a function X, i.e.,

$$\sigma(X) \;:=\; \{A \subset \Omega : A = X^{-1}(B) \text{ for some } B \in \mathcal{B}\}$$

is the smallest σ-algebra that makes X measurable. This σ-algebra is called the **σ-algebra generated by the function X**. In the context of probability theory, the measurability of a random variable is a natural assumption that expresses the respect for the information. Indeed, the information that can be observed from the random variable X is represented by $\sigma(X)$ and we will see later that the probability of $A \in \sigma(X)$ is only well-defined when the set A is an event. In particular, to have a well-defined theory, all sets in $\sigma(X)$ must be measurable, which is exactly the definition of the measurability (1.3) of a function. Returning to the general context of measure theory, key measurable functions to define the integral are called simple functions.

Definition 1.3 – A function $s : \Omega \to \mathbb{R}$ is said to be a **simple function** whenever its range is finite. In particular, letting $a_1, a_2, \ldots, a_N \neq 0$ be the distinct values of s,

$$s \;=\; \textstyle\sum_{n=1,2,\ldots,N} a_n \mathbf{1}_{A_n} \quad \text{where} \quad A_n := \{\omega : s(\omega) = a_n\}. \tag{1.5}$$

Note that, the set Ω being equipped with a σ-algebra $\mathcal{F}$, the simple function s is measurable if and only if, for $n = 1, 2, \ldots, N$, the set A_n is measurable. We write $\mathcal{S}(\Omega, \mathcal{F})$ the set of all measurable simple functions defined on the measurable space $(\Omega, \mathcal{F})$.

Definition 1.4 – Let $(\Omega, \mathcal{F})$ be a measurable space.

1. A **positive measure** on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ which is **σ-additive**, i.e., for each sequence $(A_n) \subset \mathcal{F}$,
$$\mu\left(\textstyle\bigcup_{n\geq 1} A_n\right) \;=\; \textstyle\sum_{n\geq 1} \mu(A_n) \quad \text{whenever} \quad A_i \cap A_j = \varnothing \text{ for } i \neq j.$$
2. If in addition $\mu(\Omega) = 1$ then μ is called a **probability measure**.
3. The triplet $(\Omega, \mathcal{F}, \mu)$ is then called a **measure space** in measure theory and a **probability space** in probability theory.

Probability measures have a number of very useful properties that can be guessed typically using a so-called Venn diagram: assuming that $\mu(\Omega) = 1$,

- the probability of the empty set is zero: $\mu(\varnothing) = 0$,
- $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$ whenever $A_i \cap A_j = \varnothing$ for $i \neq j$,
- $\mu(A) \leq \mu(B)$ whenever $A \subset B$,
- $\lim_{n\to\infty} \mu(A_n) = \mu\left(\lim_{n\to\infty} A_n\right)$ for every monotone sequence $(A_n) \subset \mathcal{F}$.

Another property particularly useful is the so-called inclusion-exclusion identity:

$$\mu\left(\textstyle\bigcup_{n=1,2,\ldots,N} A_n\right) \;=\; \textstyle\sum_{k=1,2,\ldots,N} (-1)^{k+1} \sum_{i_1<\cdots<i_k} \mu\left(A_{i_1} \cap \cdots \cap A_{i_k}\right).$$

In words, the probability of the union is equal to the sum of the probabilities of each set minus the sum of the probabilities of all the double intersections plus the sum of the probabilities of all the triple intersections, and so on. These properties are true more generally for finite positive measures but not in general for infinite measures. Having defined positive measures, we are now ready for the construction of the integral. The construction is divided into three steps starting from positive simple measurable functions, then positive measurable functions, and finally measurable integrable functions. It is good to have this construction in mind since a number of results in measure theory can be proved by following these steps, i.e., by first showing the result for simple functions and then extending the result to more general functions.

Definition 1.5 – Let $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ be a measurable function and $A \in \mathcal{F}$. Then, the integral of X over the set A with respect to the measure μ is defined as follows.

Simple measurable functions – Assume that $X := s$ as in (1.5). Then,

$$\int_A X\, d\mu \;= \int_A \sum_{n=1}^{N} a_n \mathbf{1}_{A_n}\, d\mu \;:=\; \sum_{n=1}^{N} a_n\, \mu(A_n \cap A).$$

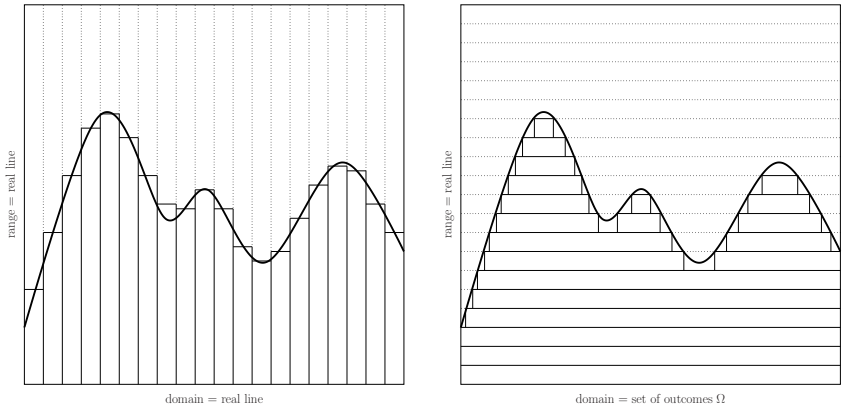


Figure 1. Partitioning the domain (construction of the Riemann integral) versus partitioning the range (construction of the Lebesgue/abstract integral).

Positive measurable functions – Assume that $X \geq 0$. Then,

$$\int_A X\, d\mu \;:=\; \sup_{0\leq s\leq X} \left\{ \int_A s\, d\mu : s \in \mathcal{S}(\Omega, \mathcal{F}) \right\}.$$

Integrable functions – Define the set of **integrable** functions as

$$L^1(\Omega, \mathcal{F}, \mu) \;:=\; \left\{ X \in \mathcal{M}(\Omega, \mathcal{F}) : \int_\Omega |X|\, d\mu < \infty \right\}.$$

Then, for all $X \in L^1(\Omega, \mathcal{F}, \mu)$,

$$\int_A X\, d\mu \;:= \int_A X^+\, d\mu \;- \int_A X^-\, d\mu.$$

Note that this quantity is well-defined because $|X| = X^+ + X^-$ so the integral of the positive and negative parts are both finite. This is the reason why one has to restrict the definition to functions which are either positive measurable or integrable.

By convention, we assume that $0 \cdot \infty = \infty \cdot 0 = 0$. In particular, the integral of the function identically equal to zero over a set with infinite measure is equal to zero. In probability theory, the integral of a random variable X is nothing else than the **expected value** and we write

$$EX \;=\; E_P(X) \;= \int_\Omega X\, dP$$

where P is the probability measure. Returning to the general context of measure theory, a particular case of interest in analysis is the integral with respect to the **Lebesgue measure**, usually written λ, which can be proved to be the unique measure on the Borel sets such that

$$\lambda((a, b]) \;=\; b - a \quad \text{for all} \quad a \leq b.$$

It turns out that the integral with respect to the Lebesgue measure coincides with the Riemann integral on the set of Riemann-integrable functions but the set of functions which are Lebesgue-integrable is much larger. To illustrate this aspect, we now answer the question raised at the beginning of this section and compute the integral (1.1). Since the set of rational numbers is countable, there exists a sequence (A_n) of Borel sets such that

$$A_0 \subset A_1 \subset \cdots \subset A_n \subset \mathbb{Q} \quad \text{and} \quad \text{card}(A_n) = n \quad \text{and} \quad \bigcup_n A_n = \mathbb{Q} \cap (0,1).$$

Then, let $B_n = A_n \setminus A_{n-1}$ for all $n > 0$ so that

$$A_n = \bigcup_{j=1,2,\ldots,n} B_j \quad \text{and} \quad \mathbb{Q} \cap (0,1) = \bigcup_j B_j \quad \text{and} \quad B_i \cap B_j = \varnothing \text{ for } i \neq j.$$

Using σ-additivity and basic properties of measures, we deduce

$$\begin{aligned} P(X \in \mathbb{Q}) &= \lambda(\mathbb{Q} \cap (0,1)) = \textstyle\sum_j \lambda(B_j) \\ &= \lim_{n\to\infty} \textstyle\sum_{j\leq n} \lambda(B_j) = \lim_{n\to\infty} \lambda(A_n) = 0. \end{aligned}$$

More generally, our reasoning shows that the Lebesgue measure of any countable set is equal to zero but we point out that the converse is not true. For instance, the Cantor set is an example of uncountable set with Lebesgue measure zero. The example of the Dirichlet function reveals another important shortcoming of the Riemann integral: the functions $\mathbf{1}_{A_n}$ define an increasing sequence of Riemann-integrable functions with integral equal to zero that converges pointwise to a function which is not Riemann-integrable. In contrast, the integral with respect to the Lebesgue measure or any other positive measure is consistent in such a context: the sequence of the integrals converges to the integral of the limit, which is discussed in detail in the next subsection.

Inversion limits and integrals. As for the Riemann integral, the abstract integral has a number of nice properties. In particular, it is easy to prove that it is a linear and monotone operator on the space of positive measurable or integrable functions. In this subsection, we prove additional properties specific to the abstract integral which are particularly useful in both analysis and probability theory since they give sufficient conditions under which, having a sequence of functions, limit and integral can be interchanged: the monotone convergence theorem and the dominated convergence theorem. We also state without proof another very useful result: Fubini's theorem. The monotone convergence theorem and dominated convergence theorem will be stated for random variables because this is in this context that they will be used later in those notes. Fubini's theorem, however, will be stated in generality for so-called σ-finite measures.

Theorem 1.6 (monotone convergence) – Assume that (X_n) is a nondecreasing sequence of positive random variables with pointwise limit X. Then,

$$X \in \mathcal{M}(\Omega, \mathcal{F}) \quad \text{and} \quad \lim_{n\to\infty} EX_n = EX.$$

Proof. By monotonicity of the sequence (X_n), for all $a \geq 0$, we have

$$\{X \leq a\} = \{\sup_n X_n \leq a\} = \{X_n \leq a \text{ for all } n \geq 1\} = \textstyle\bigcap_{n\geq 1} \{X_n \leq a\}$$

therefore $\{X \leq a\}$ is an event, and so X is a random variable according to (1.4). Using again the monotonicity of the sequence of functions, we also have

$$\lim_{n\to\infty} EX_n = \sup_n EX_n \leq EX. \tag{1.6}$$

To prove the reverse inequality, we let $\epsilon > 0$ be small and $0 \leq s \leq X$ be an arbitrary simple random variable, and we define the sequence of sets

$$A_n = \{\omega \in \Omega : X_n(\omega) \geq (1-\epsilon)\, s(\omega)\} \quad \text{for all } n \in \mathbb{N}.$$

Note that each set A_n is an event and that

$$EX_n \geq E(X_n \mathbf{1}_{A_n}) \geq E((1-\epsilon)\, s\, \mathbf{1}_{A_n}) = (1-\epsilon)\, E(s\, \mathbf{1}_{A_n}).$$

Observing that $\mu(A) = E(s\, \mathbf{1}_A)$ is a positive measure and that $A_n \uparrow \Omega$ because the sequence of random variables is nondecreasing and converges to X, we deduce

$$\begin{aligned} \lim_{n\to\infty} EX_n &\geq (1-\epsilon) \lim_{n\to\infty} E(s\, \mathbf{1}_{A_n}) \\ &= (1-\epsilon) \lim_{n\to\infty} \mu(A_n) = (1-\epsilon)\, \mu(\Omega) = (1-\epsilon)\, E(s). \end{aligned}$$

Since this holds for all $\epsilon > 0$ and all simple measurable functions $0 \leq s \leq X$, by taking the limit as $\epsilon \to 0$ and then the supremum over all simple functions, we obtain

$$\lim_{n\to\infty} EX_n \geq \sup_{0\leq s\leq X} E(s) = EX \tag{1.7}$$

where the last equality follows from the definition of the integral. Combining (1.6) and (1.7) gives the desired equality, and completes the proof. □

Lemma 1.7 (Fatou's lemma) – For any sequence (X_n) of positive random variables,

$$\liminf_{n\to\infty} X_n \in \mathcal{M}(\Omega, \mathcal{F}) \quad \text{and} \quad E(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} EX_n.$$

Proof. The idea is to define the new sequence

$$Z_n(\omega) := \inf_{i\geq n} X_i(\omega) \quad \text{for all} \quad \omega \in \Omega.$$

This is a nondecreasing sequence of positive random variables so it converges pointwise. Moreover, according to the monotone convergence theorem 1.6, its limit

$$\lim_{n\to\infty} Z_n = \lim_{n\to\infty} (\inf_{i\geq n} X_i) = \liminf_{n\to\infty} X_n$$

is a random variable and we have

$$\begin{aligned} E(\liminf_{n\to\infty} X_n) &= E(\lim_{n\to\infty} Z_n) = \lim_{n\to\infty} EZ_n \\ &= \liminf_{n\to\infty} EZ_n \leq \liminf_{n\to\infty} EX_n \end{aligned}$$

which completes the proof. □

Theorem 1.8 (dominated convergence) – Assume that (X_n) is a sequence of random variables with pointwise limit X and that there exists

$$Y \in L^1(\Omega, \mathcal{F}, P) \quad \text{such that} \quad |X_n| \leq Y \quad \text{for all} \quad n \geq 1.$$

Then, X is integrable,

$$\lim_{n\to\infty} E|X_n - X| = 0 \quad \text{and} \quad \lim_{n\to\infty} EX_n = EX.$$

Proof. Since $|X| \le Y$, that Y is integrable and that X is measurable, X is integrable. The rest of the proof follows from Fatou's lemma and the fact that

$$2Y - |X_n - X| \;\ge\; 2Y - |X_n| - |X| \;\ge\; 0 \quad \text{and} \quad \limsup_{n\to\infty} |X_n - X| \;=\; 0.$$

Indeed, by Lemma 1.7, we have

$$\begin{aligned} E(2Y) &= E(2Y - \limsup_{n\to\infty} |X_n - X|) = E(\liminf_{n\to\infty} (2Y - |X_n - X|)) \\ &\le \liminf_{n\to\infty} E(2Y - |X_n - X|) = E(2Y) - \limsup_{n\to\infty} E|X_n - X| \end{aligned}$$

therefore $\limsup_{n\to\infty} E|X_n - X| = 0$. This, together with

$$|EX_n - EX| \;=\; |E(X_n - X)| \;\le\; E|X_n - X|,$$

also implies the last statement of the theorem. □

We now state Fubini's theorem. While Theorems 1.6 and 1.8 give conditions under which limit and integral can be interchanged, Fubini's theorem gives conditions under which two integrals can be interchanged when looking at functions of two variables. To describe the framework, we first need to define the product of measure spaces. Let $(S, \mathcal{S}, \mu_S)$ and $(T, \mathcal{T}, \mu_T)$ be two measure spaces. In order to define the concepts seen so far to functions defined on the product $\Omega = S \times T$, we first introduce the σ-algebra $\mathcal{F}$ generated by the rectangles:

$$\mathcal{F} \;=\; \mathcal{S} \times \mathcal{T} \;:=\; \sigma\{A \times B : A \in \mathcal{S} \text{ and } B \in \mathcal{T}\}.$$

It can be proved that there is a unique measure μ on $\mathcal{F}$ such that

$$\mu(A \times B) \;=\; \mu_S(A)\,\mu_T(B) \quad \text{for all } A \in \mathcal{S} \text{ and } B \in \mathcal{T}.$$

This measure is called **product measure** and is denoted by $\mu := \mu_S \times \mu_T$.

Definition 1.9 – A measure space $(\Omega, \mathcal{F}, \mu)$ is said to be **σ-finite** whenever Ω can be written as a countable union of measurable sets A_n such that $\mu(A_n) < \infty$ for all n.

Theorem 1.10 (Fubini) – Let $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ be a positive measurable function or an integrable function on a σ-finite measure space. Then,

- For all $s \in S$ and all $t \in T$, we have $X(s, \cdot) \in \mathcal{M}(T, \mathcal{T})$ and $X(\cdot, t) \in \mathcal{M}(S, \mathcal{S})$.
- We have $\phi \in \mathcal{M}(S, \mathcal{S})$ and $\psi \in \mathcal{M}(T, \mathcal{T})$ where

$$\phi(s) \;:=\; \int_T X(s,t)\,\mu_T(dt) \;\text{ for } s \in S \quad \text{and} \quad \psi(t) \;:=\; \int_S X(s,t)\,\mu_S(ds) \;\text{ for } t \in T.$$

- The order of integration does not matter in the sense that

$$\int_S \int_T X(s,t)\,\mu_T(dt)\,\mu_S(ds) \;=\; \int_{S\times T} X\,d\mu \;=\; \int_T \int_S X(s,t)\,\mu_S(ds)\,\mu_T(dt).$$

In practice, we often use that the double integral on the left-hand side equals the double integral on the right-hand side. Rather than giving a proof of this theorem, which would require the introduction of more definitions that are not useful for after, we conclude this subsection with some counter-examples. Starting with the monotone and dominated convergence theorems, we note that the

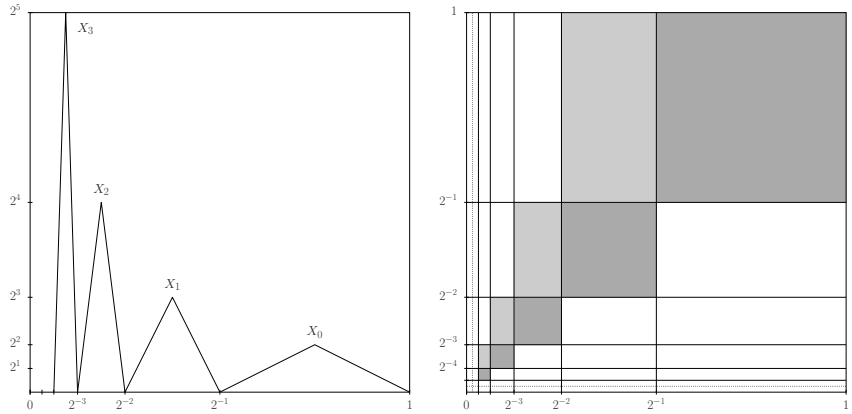


Figure 2. Left: example of sequence for which limit and integral cannot be interchanged. Right: schematic representation of the domain of the function of two variables for which the order of integration matters.

sequence (X_n) illustrated on the left-hand side of Figure 2, which is neither monotone nor dominated by an integrable function, satisfies

$$\int X_n(x)\,dx \;=\; 1 \;\text{ for all } n \ge 1 \quad \text{and} \quad \int \lim_{n\to\infty} X_n(x)\,dx \;=\; 0$$

since the sequence converges pointwise to the function identically equal to zero. In particular, the fact that limit and integral can be interchanged as in Theorems 1.6 and 1.8 does not hold for general sequences of measurable functions. We use the sequence (X_n) to also construct a counter-example for Fubini's theorem. Consider the function X defined on the unit square by

$$X(s,t) \;:=\; \begin{cases} +\,X_n(s)\,X_n(t) & \text{if } \; 2^{-(n+1)} < s, t < 2^{-n} \\ -\,X_{n+1}(s)\,X_n(t) & \text{if } \; 2^{-(n+2)} < s < 2^{-(n+1)} < t < 2^{-n} \end{cases}$$

and $X(x, y) := 0$ otherwise. The right-hand side of Figure 2 shows a partition of the unit square where the dark grey squares represent the region where X is positive, the pale grey rectangles the region where it is negative, and the white rectangles the region where it is equal to zero. Using that the integral of each X_n is equal to one, some basic algebra shows that, one variable being fixed, the integral of X with respect to other variable depends on whether or not the corresponding cross-section intersects a dark/pale grey rectangle and we find

$$\int X(s,t)\,ds \;=\; 0 \quad \text{and} \quad \int X(s,t)\,dt \;=\; \begin{cases} 0 & \text{for } s < 1/2 \\ X_1(s) & \text{for } s > 1/2. \end{cases}$$

It follows that

$$\iint X(s,t)\,ds\,dt \;=\; 0 \quad \text{and} \quad \iint X(s,t)\,dt\,ds \;=\; \int X_1(s)\,ds \;=\; 1.$$

In particular, the conclusion of Fubini's theorem does not hold.

Radon-Nikodým theorem. This theorem is a general result in measure theory that has interesting implications in probability theory. These implications are discussed in the next section. To motivate the theorem, note that, given a positive measurable function ϕ on $(\Omega, \mathcal{F}, \mu)$,

$$\nu(A) \;:=\; \int_A \phi \, d\mu \;=\; \int \phi \, \mathbf{1}_A \, d\mu \quad \text{for all } A \in \mathcal{F}$$

defines a new measure ν on $(\Omega, \mathcal{F})$. Indeed, $\nu(\varnothing) = 0$ while the monotone convergence theorem implies that for all sequences of mutually exclusive measurable sets $(A_n) \subset \mathcal{F}$, we have

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) \;=\; \int \lim_{N\to\infty} \sum_{n=1}^{N} (\phi \, \mathbf{1}_{A_n}) \, d\mu \;=\; \lim_{N\to\infty} \sum_{n=1}^{N} \int \phi \, \mathbf{1}_{A_n} \, d\mu \;=\; \sum_{n=1}^{\infty} \nu(A_n).$$

The Radon-Nikodým theorem is in some sense the converse of the previous statement as it gives the existence and uniqueness of the function ϕ under certain conditions on the measures.

Definition 1.11 (absolute continuity) – The measure ν is said to be absolutely continuous with respect to the measure μ, which we write $\nu \ll \mu$, whenever

$$\text{for all } A \in \mathcal{F}, \quad \mu(A) = 0 \;\text{ implies that }\; \nu(A) = 0. \tag{1.8}$$

Theorem 1.12 (Radon-Nikodým) – Let μ and ν be two σ-finite measures such that $\nu \ll \mu$. Then, there exists a positive measurable function $\phi : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ such that

$$\nu(A) \;=\; \int_A \phi \, d\mu \quad \text{for all } A \in \mathcal{F}.$$

The function ϕ is written $\phi := d\nu/d\mu$ and is called a **Radon-Nikodým derivative** of ν with respect to μ. It is unique in the sense that two such derivatives are equal μ-almost everywhere, i.e., the set where both functions differ has measure zero for the measure μ.

To understand the assumption of the theorem, assume for a moment that the measures ν and μ are simply nonnegative functions defined on the real line. Then, it is obvious that there exists a function ϕ such $\nu = \phi \, \mu$ if and only if $\mu(x) = 0$ implies $\nu(x) = 0$ for all $x \in \mathbb{R}$. This last condition can be viewed as the analog of the absolute continuity for positive measures (1.8). In order to prove the Radon-Nikodým theorem, we now introduce signed measures and Hahn decompositions.

Definition 1.13 – A **signed measure** on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to (-\infty, \infty]$ that satisfies the two axioms of a positive measure.

The value $-\infty$ is excluded from the range of μ to avoid situations in which the measure of the union of a set with measure $+\infty$ and a set with measure $-\infty$ is not defined.

Theorem 1.14 (Hahn decomposition) – Let μ be a signed measure on $(\Omega, \mathcal{F})$. Then, there exist two measurable sets Ω_+ and Ω_- such that

1. $\Omega_+ \cup \Omega_- = \Omega$ and $\Omega_+ \cap \Omega_- = \varnothing$.
2. Ω_+ is a positive set for μ, i.e., $\mu(A) \geq 0$ for all $A \in \mathcal{F}$, $A \subset \Omega_+$.
3. Ω_- is a negative set for μ, i.e., $\mu(A) \leq 0$ for all $A \in \mathcal{F}$, $A \subset \Omega_-$.

Proof. The proof is divided into two steps.

Step 1 – Any $B \in \mathcal{F}$ with $\mu(B) \leq 0$ contains a negative set A with $\mu(A) \leq \mu(B)$.

The idea is to remove from B all possible subsets with a positive measure, which is done inductively by defining a sequence (A_n) whose members are mutually exclusive subsets with a "significantly" large measure. More precisely, having defined the first $n-1$ members of this sequence, we set

$$a_n \;:=\; \sup\{\mu(D) : D \in \mathcal{F} \text{ and } D \subset B \setminus (A_1 \cup A_2 \cup \cdots \cup A_{n-1})\} \;\geq\; 0$$

and let A_n be any measurable set such that

$$A_n \;\subset\; B \setminus (A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \quad \text{and} \quad \mu(A_n) \;\geq\; \min(1, a_n/2).$$

Then, defining $A := B \setminus (A_1 \cup A_2 \cup \cdots)$, by σ-additivity we obtain

$$\begin{aligned} \mu(A) \;&=\; \mu(B) - \textstyle\sum_{n\geq 1} \mu(A_n) \\ &\leq\; \mu(B) - \textstyle\sum_{n\geq 1} \min(1, a_n/2) \;\leq\; \mu(B). \end{aligned}$$

Moreover, if $\mu(D) > 0$ for some $D \subset A$ then we must have $a_n \geq \mu(D) > 0$ for all n and the previous inequality would imply that $\mu(A) = -\infty$ which is not possible by definition of a signed measure, therefore A is indeed a negative set for μ.

Step 2 – Construction of the Hahn decomposition.

This follows from a similar construction: the idea is to express our candidate for Ω_- as a countable union of mutually exclusive negative sets (B_n) with a "significantly" small measure and defined inductively using step 1. Having defined the first $n-1$ members of this sequence, we set

$$b_n \;:=\; \inf\{\mu(D) : D \in \mathcal{F} \text{ and } D \subset \Omega \setminus (B_1 \cup B_2 \cup \cdots \cup B_{n-1})\} \;\leq\; 0.$$

By definition of b_n and step 1, there exist $D_n \in \mathcal{F}$ and $B_n \subset D_n$ negative such that

$$B_n \;\subset\; D_n \;\subset\; D \setminus (B_1 \cup \cdots \cup B_{n-1}) \quad \text{and} \quad \mu(B_n) \;\leq\; \mu(D_n) \;\leq\; \max(-1, b_n/2).$$

Then, we define Ω_- as the union of (B_n). Using the σ-additivity of the signed measure and the fact that each B_n is a negative set, we obtain for all $D \subset \Omega_-$ that

$$\mu(D) \;=\; \mu(\textstyle\bigcup_{n\geq 1} D \cap B_n) \;=\; \textstyle\sum_{n\geq 1} \mu(D \cap B_n) \;\leq\; 0$$

therefore Ω_- is a negative set. Finally, let $\Omega_+ := \Omega \setminus \Omega_-$ and observe that if $\mu(D) < 0$ for some $D \subset \Omega_+$ then we must have $b_n \leq \mu(D) < 0$ for all n and so

$$\mu(\Omega_-) \;=\; \textstyle\sum_{n\geq 1} \mu(B_n) \;\leq\; \sum_{n\geq 1} \max(-1, b_n/2) \;=\; -\infty$$

which is not possible. In conclusion, Ω_+ is a positive set. □

We are now ready for the proof of the Radon-Nikodým theorem.

Proof. We only prove the result for finite positive measures, from which the result for σ-finite measures easily follows. Let $\mathcal{G}$ be the set of all measurable functions g such that

$$\int_A g \, d\mu \;\leq\; \nu(A) \quad \text{for all } A \in \mathcal{F}.$$

Also, define $(g_n) \subset \mathcal{G}$ such that

$$\lim_{n\to\infty} \int g_n \, d\mu \;=\; \sup_{g\in\mathcal{G}} \int g \, d\mu \quad \text{and} \quad \phi_n \;:=\; \max(g_1, g_2, \ldots, g_n)$$

for all $n \geq 1$. Observing that

$$A_1 \;:=\; A \cap \{g_1 > g_2\} \;\in\; \mathcal{F} \quad \text{and} \quad A_2 \;:=\; A \cap \{g_1 \leq g_2\} \;\in\; \mathcal{F}$$

for each measurable set $A \in \mathcal{F}$, we have

$$\begin{aligned} \int_A \phi_2 \, d\mu \;&=\; \int_A \max(g_1, g_2) \, d\mu \;=\; \int_{A_1} \max(g_1, g_2) \, d\mu \;+\; \int_{A_2} \max(g_1, g_2) \, d\mu \\ &=\; \int_{A_1} g_1 \, d\mu \;+\; \int_{A_2} g_2 \, d\mu \;\leq\; \nu(A_1) + \mu(A_2) \;=\; \nu(A_1 \cup A_2) \;=\; \nu(A) \end{aligned}$$

therefore $\phi_2 \in \mathcal{G}$ and a simple induction implies that $\phi_n \in \mathcal{G}$ for all n. Moreover, since $(\phi_n) \subset \mathcal{G}$ is nondecreasing, it converges pointwise to a measurable function $\phi \in \mathcal{G}$ and according to the monotone convergence theorem 1.6, we obtain

$$\sup_{g\in\mathcal{G}} \int g \, d\mu \;\geq\; \int \phi \, d\mu \;=\; \lim_{n\to\infty} \int \phi_n \, d\mu \;\geq\; \lim_{n\to\infty} \int g_n \, d\mu \;=\; \sup_{g\in\mathcal{G}} \int g \, d\mu.$$

The function ϕ is the natural candidate for $d\nu/d\mu$. To check this, define

$$\nu_0(A) \;:=\; \nu(A) \;-\; \int_A \phi \, d\mu \quad \text{for all } A \in \mathcal{F}.$$

It is easy to check that ν_0 is a positive measure, and assuming by contradiction that $\nu_0 \neq 0$, the fact that the measure μ is finite implies that

$$\nu_0(\Omega) - \epsilon\,\mu(\Omega) \;>\; 0 \quad \text{for some } \epsilon > 0 \text{ small.}$$

Fix a Hahn decomposition (Ω_+, Ω_-) for the measure $\nu_0 - \epsilon\,\mu$. Then, for all $A \in \mathcal{F}$

$$\begin{aligned} \nu(A) \;&=\; \int_A \phi \, d\mu \;+\; \nu_0(A) \;\geq \int_A \phi \, d\mu \;+\; \nu_0(A \cap \Omega_+) \\ &\geq\; \int_A \phi \, d\mu \;+\; \epsilon\,\mu(A \cap \Omega_+) \;=\; \int_A \phi \, d\mu \;+\; \epsilon \int_A \mathbf{1}_{\Omega_+} \, d\mu \;=\; \int_A (\phi + \epsilon\,\mathbf{1}_{\Omega_+}) \, d\mu \end{aligned}$$

therefore $\phi + \epsilon\,\mathbf{1}_{\Omega_+} \in \mathcal{G}$. In particular, taking $A = \Omega$, we obtain

$$\int \phi \, d\mu \;+\; \epsilon\,\mu(\Omega_+) \;=\; \int (\phi + \epsilon\,\mathbf{1}_{\Omega_+}) \, d\mu \;\leq\; \sup_{g\in\mathcal{G}} \int g \, d\mu \;=\; \int \phi \, d\mu$$

therefore $\mu(\Omega_+) = 0$, and $\nu_0(\Omega_+) \leq \nu(\Omega_+) = 0$ since $\nu \ll \mu$. But then

$$(\nu_0 - \epsilon\,\mu)(\Omega) \;=\; (\nu_0 - \epsilon\,\mu)(\Omega_-) \;\leq\; 0$$

thus contradicting the definition of ϵ. In particular, $\nu_0 = 0$, which proves the existence of the Radon-Nikodým derivative. To also prove uniqueness in the sense stated above, assume that there exist two measurable functions ϕ and ψ such that

$$\nu(A) \;=\; \int_A \phi \, d\mu \;=\; \int_A \psi \, d\mu \quad \text{for all } A \in \mathcal{F}.$$

Taking $A_1 := \{\phi < \psi\}$ and $A_2 := \{\phi > \psi\}$, we obtain

$$\begin{aligned} \int |\phi - \psi| \, d\mu \;&=\; \int (\phi - \psi)^+ + (\phi - \psi)^- \, d\mu \\ &=\; \int_{A_1} (\phi - \psi) \, d\mu \;+\; \int_{A_2} (\phi - \psi) \, d\mu \;=\; \nu(A_1) - \nu(A_1) + \nu(A_2) - \nu(A_2) \;=\; 0 \end{aligned}$$

therefore $\mu(\phi \neq \psi) = 0$, i.e., $\phi = \psi$ μ-almost everywhere. □

2. Implications in probability theory

The monotone convergence theorem, the dominated convergence theorem and Fubini's theorem are frequently used in probability theory. The Radon-Nikodým theorem is important as well since it allows to define more rigorously and in a more general context key probability concepts typically seen at the undergraduate level: the measure induced by a random variable, whose Radon-Nikodým derivative is the probability mass function in the special case of a discrete random variable and the probability density function in the special case of a continuous random variable, and the conditional expectation. We now describe these two concepts from the point of view of measure theory.

Induced measure and distribution. Having a random variable X on a probability space, one can define a probability measure ν_X on the Borel σ-algebra by letting

$$\nu_X(B) \;:=\; P(X \in B) \;=\; \int_\Omega \mathbf{1}_{X^{-1}(B)} \, dP \quad \text{for all } B \in \mathcal{B}.$$

It is called the **measure induced by** X in measure theory and **distribution of** X in probability theory. To study a random variable in practice, probabilists do not work with the measure P but with its distribution by using the following **change of variables formula**.

Theorem 2.1 – Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a random variable and $f : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then, whenever f is positive or integrable,

$$E\,f(X) \;=\; \int_{\mathbb{R}} f \, d\nu_X. \tag{2.1}$$

Proof. The steps of the proof follow the construct of the integral.

Step 1 – Assume first that $f := \mathbf{1}_B$ for some Borel set B. Then,

$$E\,\mathbf{1}_B(X) \;=\; E\,\mathbf{1}\{X \in B\} \;=\; P(X \in B) \;=\; \nu_X(B) \;=\; \int_{\mathbb{R}} \mathbf{1}_B \, d\nu_X.$$

Step 2 – Assume now that $f := a_1\,\mathbf{1}_{B_1} + \cdots + a_N\,\mathbf{1}_{B_N}$ is a simple measurable function. Using the previous step and the linearity of the integral/expected value, we obtain

$$E\,f(X) \;=\; \sum_{n=1}^{N} a_n\,E\,\mathbf{1}_{B_n}(X) \;=\; \sum_{n=1}^{N} a_n \int_{\mathbb{R}} \mathbf{1}_{B_n} \, d\nu_X \;=\; \int_{\mathbb{R}} f \, d\nu_X.$$

Step 3 – Assume that f is positive and let $s_n(x) := \min(n, 2^{-n}\lfloor 2^n f(x) \rfloor)$. This results in a nondecreasing sequence of simple measurable functions that converges to the function f. In particular, the monotone convergence theorem 1.6 implies that

$$E\,f(X) \;=\; \lim_{n\to\infty} E\,s_n(X) \;=\; \lim_{n\to\infty} \int_{\mathbb{R}} s_n \, d\nu_X \;=\; \int_{\mathbb{R}} f \, d\nu_X.$$

Step 4 – Finally, when f is integrable, we write $f = f^+ - f^-$ and observe that f^+ and f^- are both positive (so the previous step applies) and integrable therefore

$$E\,f(X) \;=\; E\,f^+(X) - E\,f^-(X) \;=\; \int_{\mathbb{R}} f^+\,d\nu_X \;-\; \int_{\mathbb{R}} f^-\,d\nu_X \;=\; \int_{\mathbb{R}} f\,d\nu_X.$$

This completes the proof. □

In practice, it is also convenient to express the distribution ν_X as a measurable function times a standard measure such as the Lebesgue measure. This idea is related to the Radon-Nikodým theorem. In the context of probability theory, if X is a discrete random variable, i.e., its range is finite or countable, say a subset of $\mathbb{Z}$, then

$$\nu_X \;\ll\; \mu \;:=\; \sum_{x\in\mathbb{Z}} \delta_x \quad\text{and}\quad d\nu_X \;=\; \phi_X\,d\mu \;=\; \sum_{x\in\mathbb{Z}} \phi_X(x)\,d\delta_x$$

for some measurable function ϕ_X since μ is obviously σ-finite and thus the Radon-Nikodým theorem applies. Here, δ_x is the pointmass measure, $\delta_x(A) := \mathbf{1}\{x \in A\}$. In the discrete case, the Radon-Nikodým derivative ϕ_X is called the **probability mass function**. Similarly, calling X a continuous random variable whenever its distribution is absolutely continuous with respect to the Lebesgue measure, which is a σ-finite measure, the Radon-Nikodým theorem gives the existence of a measurable function ϕ_X such that $d\nu_X = \phi_X\,d\lambda$. The derivative ϕ_X is then called the **probability density function** of the random variable. In conclusion, random variables can be characterized by the Radon-Nikodým derivative ϕ_X (probability mass or probability density function) of their distribution with respect to some standard measures. Using the Radon-Nikodým derivative in both contexts, the **change of variables formula** (2.1) becomes

$$\begin{aligned} E\,f(X) \;&=\; \int \sum_{x\in\mathbb{Z}} \phi_X(x)\,f\,d\delta_x \;=\; \sum_{x\in\mathbb{Z}} f(x)\,\phi_X(x) && \text{(discrete case)} \\ &=\; \int f\,\phi_X\,d\lambda \;=\; \int_{\mathbb{R}} f(x)\,\phi_X(x)\,dx && \text{(continuous case)} \end{aligned}$$

which are the formula used in practice since random variables are characterized by their distribution through their probability mass function or probability density function.

Conditional expectation. The last step before introducing stochastic processes is to define conditional expectation since it is a key concept through which one can express certain dependency relationships among random variables and define martingales and Markov chains. This concept of conditional expectation is introduced in the following definition. The fact that it exists and is unique follows from the Radon-Nikodým theorem.

Definition 2.2 – Let $X \in L^1(\Omega, \mathcal{F}, P)$ and let $\mathcal{G} \subset F$ be a σ-algebra.

- The **conditional expectation of X given $\mathcal{G}$** is any random variable Z such that

$$Z \in \mathcal{M}(\Omega, \mathcal{G}) \quad\text{and}\quad E\,(X\,\mathbf{1}_A) = E\,(Z\,\mathbf{1}_A) \;\text{ for all } A \in \mathcal{G}.$$

  The variable Z is called a **version** of $E\,(X\,|\,\mathcal{G})$.
- Having a second random variable Y, we define $E\,(X\,|\,Y) := E\,(X\,|\,\sigma(Y))$.
- Also, we define the **conditional probability** by $P\,(X \in B\,|\,\mathcal{G}) := E\,(\mathbf{1}\{X \in B\}\,|\,\mathcal{G})$.

Theorem 2.3 – The conditional expectation exists and is unique in the sense that two different versions of the conditional expectation $E\,(X\,|\,\mathcal{G})$ are equal P-almost surely.

Proof. Assuming first that X is positive, since

$$\nu(A) \;:=\; E\,(X\,\mathbf{1}_A) \;=\; \int_A X\,dP \quad \text{for all } A \in \mathcal{G}$$

defines a finite measure $\nu \ll P$ on the space $(\Omega, \mathcal{G})$, there is a unique Z that satisfies the statement of the theorem: the Radon-Nikodým derivative $d\nu/dP$. In the general case when X is integrable, the first part of the proof applies to its positive part and its negative part. In particular, there exist two random variables Z_+ and Z_- measurable with respect to $\mathcal{G}$ such that

$$\begin{aligned} E\,(X\,\mathbf{1}_A) \;&=\; E\,(X^+\,\mathbf{1}_A) - E\,(X^-\,\mathbf{1}_A) \\ &=\; E\,(Z_+\,\mathbf{1}_A) - E\,(Z_-\,\mathbf{1}_A) \;=\; E\,((Z_+ - Z_-)\,\mathbf{1}_A) \end{aligned}$$

for all $A \in \mathcal{G}$. The uniqueness of the conditional expectation $Z_+ - Z_-$ can be established as in the proof of the Radon-Nikodým theorem. □

Intuitively, one can think of $E\,(X\,|\,\mathcal{G})$ as the best possible approximation of X given that the information available is encoded in the σ-algebra $\mathcal{G}$, i.e., the best possible approximation by a random variable which is $\mathcal{G}$-measurable. As the σ-algebra gets larger, the approximation gets better. Having perfect information $\mathcal{G} = \mathcal{F}$, it is clear intuitively and from the definition that the conditional expectation is equal to X itself whereas having no information $\mathcal{G} = \{\varnothing, \Omega\}$, the only functions which are measurable with respect to $\mathcal{G}$ are the constants so the best guess for the conditional expectation is no better than EX. In short, one can prove that

$$E\,(X\,|\,\mathcal{G}) = X \;\text{ when } \mathcal{G} = \mathcal{F} \quad\text{and}\quad E\,(X\,|\,\mathcal{G}) = EX \;\text{ when } \mathcal{G} = \{\varnothing, \Omega\}. \tag{2.2}$$

The right-hand side of (2.2) extends to the more general case when X and $\mathcal{G}$ are independent, i.e., the σ-algebra does not provide any information about the random variable.

Definition 2.4 – Two σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ are said to be **independent** if each event in $\mathcal{G}$ and each event in $\mathcal{H}$ are pairwise independent, i.e.,

$$P\,(A \cap B) \;=\; P\,(A)\,P\,(B) \quad \text{for all } A \in \mathcal{F} \text{ and } B \in \mathcal{G}.$$

Two random variables X and Y are independent whenever $\sigma(X)$ and $\sigma(Y)$ are.

Lemma 2.5 – Assume that X and $\mathcal{G}$ are independent. Then, $E\,(X\,|\,\mathcal{G}) = EX$.

Proof. First, we observe that the joint distribution of two independent random variables is equal to the product of their distributions. Therefore, if X and Y are independent, by applying Fubini's theorem 1.10 to the function $h(x, y) := xy$, we obtain

$$\begin{aligned} E(XY) \;&=\; \int h\,d\nu_{X,Y} \;=\; \int h\,d(\nu_X \times \nu_Y) \;=\; \int\int xy\;\nu_X(dx)\,\nu_Y(dy) \\ &=\; \int x\,\nu_X(dx) \int y\,\nu_Y(dy) \;=\; EX\,EY. \end{aligned}$$

In particular, since for all $A \in \mathcal{G}$ the random variables X and $\mathbf{1}_A$ are independent,

$$E\,(X\,\mathbf{1}_A) \;=\; E\,(X)\,E\,(\mathbf{1}_A) \;=\; E\,(E\,(X)\,\mathbf{1}_A).$$

Moreover, since $Z = EX$ is constant, it is also measurable with respect to $\mathcal{G}$. $\square$

In practice, $Z := E(f(X)\,|\,Y)$ where f is a measurable function and X and Y two random variables can be computed explicitly from their probability mass/density functions. In the discrete case, the measurability of Z with respect to $\sigma(Y)$ implies that it must be of the form

$$Z \;=\; \textstyle\sum_y a(y)\,\mathbf{1}\{Y=y\}$$

where the sum is over the range of Y. Moreover, since $\{Y=y\}\in\sigma(Y)$,

$$E(f(X)\,\mathbf{1}\{Y=y\}) \;=\; E(Z\,\mathbf{1}\{Y=y\}) \;=\; a(y)\,P(Y=y)$$

from which it follows that

$$Z \;=\sum_y \frac{E(f(X)\,\mathbf{1}\{Y=y\})}{P(Y=y)}\;\;\mathbf{1}\{Y=y\}$$

therefore we obtain the following result.

Lemma 2.6 – Assume that X and Y are discrete random variables and let $\phi_{X,Y}$ be their joint probability mass function. Then, $E(f(X)\,|\,Y)=h(Y)$ where

$$h(y) \;=\; \frac{E(f(X)\,\mathbf{1}\{Y=y\})}{P(Y=y)} \;=\sum_x f(x)\;\phi_{X,Y}(x,y)\Big/\sum_x \phi_{X,Y}(x,y). \tag{2.3}$$

By analogy, in the continuous case, we have

Lemma 2.7 – Assume that X and Y are continuous random variables and let $\phi_{X,Y}$ be their joint density function. Then, $E(f(X)\,|\,Y)=h(Y)$ where h is any function such that

$$h(y)\int\phi_{X,Y}(x,y)\,dx \;=\int f(x)\;\phi_{X,Y}(x,y)\,dx \quad\text{for all } y\in\mathbb{R}. \tag{2.4}$$

Proof. Observe first that, since h is measurable with respect to the Borel σ-algebra, $h(Y)$ is measurable with respect to $\sigma(Y)$. Moreover, for all $A\in\sigma(Y)$ with $P(A)\neq 0$, using the change of variables formula and writing $A:=\{Y\in B\}$, we obtain

$$\begin{aligned}
E(h(Y)\,\mathbf{1}_A) \;&=\; \int h(Y)\,\mathbf{1}_A\,dP \;=\; \int h(Y)\,\mathbf{1}\{Y\in B\}\,dP\\
&=\; \int_B\!\int h(y)\,\phi_{X,Y}(x,y)\,dx\,dy \;=\; \int_B\Big(h(y)\int\phi_{X,Y}(x,y)\,dx\Big)\,dy\\
&=\; \int_B\!\int f(x)\;\phi_{X,Y}(x,y)\,dx\,dy \;=\; E(f(X)\,\mathbf{1}\{Y\in B\}) \;=\; E(f(X)\,\mathbf{1}_A)
\end{aligned}$$

which completes the proof. $\square$

3. Limit theorems

4. Stochastic processes: general definition

Stochastic processes typically describe random phenomena that evolve in time but can also describe random static objects such as random graphs. In general, a stochastic process is a finite or infinite collection of random variables $\{X_i : i\in I\}$.

For all $i\in I$, we have $X_i:(\Omega,\mathcal{F},P)\to(S,\mathcal{S})$ where $(S,\mathcal{S})$ is a measurable space, a topological space equipped with its Borel σ-algebra. In most of our examples, $S\subset\mathbb{R}$.

- The set Ω is called the **set of realizations** (sample space).
- The set S is called the **state space** of the process.
- For all $\omega\in\Omega$, the set $\{X_i(\omega) : i\in I\}$ is called a **sample path**.

When there is one, the **temporal structure** is encoded in the index set I. Usually, to model discrete time we take $I:=\mathbb{N}$ and to model continuous time $I:=\mathbb{R}_+$. Stochastic processes are usually classified according to some dependence relationships among the random variables, and when time is present, we distinguish two important classes of processes: Markov chains and martingales. To define such processes, it is convenient to introduce the sequence of σ-algebras

$$\mathcal{F}_n \;:=\; \sigma(X_0,X_1,\ldots,X_n) \quad\text{for all } n\in\mathbb{N}.$$

In words, $\mathcal{F}_n$ represents the information collected up to time n so conditioning on $\mathcal{F}_n$ means that we know the sample path of the process up to time n.

Definition 4.1 (Markov chain) – The process $(X_n : n\in\mathbb{N})$ is said to be a (temporally homogeneous) discrete-time Markov chain whenever

$$P(X_{n+1}\in B\,|\,\mathcal{F}_n) \;=\; P(X_{n+1}\in B\,|\,X_n) \quad\text{for all } B\in\mathcal{S}.$$

In words, thinking of n as the present time, the previous equation means that the future of the process depends on the past only through the present.

Definition 4.2 (martingale) – The process $(X_n : n\in\mathbb{N})$ is said to be a martingale, respectively a submartingale, a supermartingale, whenever

$$\begin{array}{rcll}
E(X_{n+1}\,|\,\mathcal{F}_n) &=& X_n \text{ for all } n & \text{fair game (martingale)}\\
&\geq& X_n \text{ for all } n & \text{favorable game (submartingale)}\\
&\leq& X_n \text{ for all } n & \text{unfavorable game (supermartingale).}
\end{array}$$

In particular, S must be equipped with an order relationship but we will assume $S=\mathbb{R}$. One can think of X_n as the fortune of a gambler after she has played n times a fair game (martingale), a favorable game (submartingale), or an unfavorable game (supermartingale).

Markov chains and martingales can also be defined in continuous time. It is important to identify the type of stochastic processes we are dealing with in order to analyze them as the techniques available strongly vary between Markov chains, martingales, etc.

Example 4.3 (Ehrenfest chain) – In this model, a total of M particles is distributed among two compartments. At each time step, one particle is chosen uniformly at random from the system and moved from its original compartment to the other compartment. Let X_n denote the number of particles in one of the two compartments, say compartment 1. Then,

$$P(X_{n+1}=X_n-1\,|\,\mathcal{F}_n)=X_n/M \qquad\text{and}\qquad P(X_{n+1}=X_n+1\,|\,\mathcal{F}_n)=1-X_n/M$$

so the process is a Markov chain. However, using (2.3), we get

$$\begin{aligned} E\,(X_{n+1}\,|\,\mathcal{F}_n) &= (X_n-1)(X_n/M) \;+\; (X_n+1)(1-X_n/M) \\ &= 1+(1-2/M)\,X_n \end{aligned}$$

so the process is not a martingale.

Example 4.4 (gambler) – A gambler plays a fair game: at each step she can either win or lose the amount of money she bets both with probability one half. At each game, she decides to bet either 1 dollar or 2 dollars depending on whether she lost or won the previous game. Let X_n denote the gambler's fortune after n games. Then,

$$P\,(X_{n+1}=X_n\pm 1\,|\,\mathcal{F}_n) \;=\; \mathbf{1}\{X_n<X_{n-1}\}$$

so the process is not a Markov chain. However, using again (2.3), we get

$$\begin{aligned} E\,(X_{n+1}\,|\,\mathcal{F}_n) &= ((1/2)(X_n-1)+(1/2)(X_n+1))\,\mathbf{1}\{X_n<X_{n-1}\} \\ &\quad +((1/2)(X_n-2)+(1/2)(X_n+2))\,\mathbf{1}\{X_n>X_{n-1}\} \;=\; X_n \end{aligned}$$

so the process is a martingale.

5. Discrete-time Markov chains

Recall that stochastic processes including a temporal structure are Markov if they have a certain property called loss of memory, i.e., the future of the process depends on the past only through the present. All the processes including a temporal structure discussed in this course and a number of stochastic models that arise from physics, biology and sociology have this property. This section gives the main techniques to study discrete-time Markov chains.

Transition probabilities. Recall that the process $(X_n : n\in\mathbb{N})$ is a discrete-time Markov chain with finite or countable state space S whenever

$$P\,(X_{n+1}\in B\,|\,\mathcal{F}_n) \;=\; P\,(X_{n+1}\in B\,|\,X_n) \quad \text{for all } B\in\mathcal{S}.$$

This implies that the evolution is uniquely defined by the **transition probabilities**

$$p(x,y):=P\,(X_{n+1}=y\,|\,X_n=x) \;\text{ for all } x,y\in S$$

that we assume to be the same for all n (time homogeneous Markov chain). We call **transition matrix** the matrix of transition probabilities

$$P \;:=\; \begin{pmatrix} p(1,1) & p(1,2) & \cdots & p(1,y) & \cdots \\ p(2,1) & p(2,2) & \cdots & p(2,y) & \cdots \\ \vdots & \vdots & & \vdots & \\ p(x,1) & p(x,2) & \cdots & p(x,y) & \cdots \\ \vdots & \vdots & & \vdots & \end{pmatrix} \tag{5.1}$$

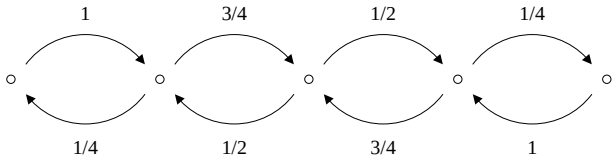


Figure 3. Ehrenfest chain with $M=4$ particles.

In practice, probabilities of interest are the **n-step transition probabilities** which, similarly to the transition probabilities, are defined by

$$p_n(x,y):=P\,(X_n=y\,|\,X_0=x) \;\text{ for all } x,y\in S.$$

The reason why it is relevant to represent the transition probabilities in a matrix form is that the coefficients of the nth power of the transition matrix are simply the n-step transition probabilities as shown in the following theorem.

Theorem 5.1 (Chapman-Kolmogorov equations) – For all states $x,y\in S$,

$$p_n(x,y) \;=\; (P^n)_{x,y} \;:=\; \text{coefficient } (x,y) \text{ of the matrix } P^n. \tag{5.2}$$

Proof. For $n=1$, this follows from the definition of the transition matrix (5.1). Assume that the result holds for some $n>1$. Conditioning on all possible values of X_n we have

$$\begin{aligned} p_{n+1}(x,y) &= \textstyle\sum_{z\in S} P\,(X_{n+1}=y\,|\,X_n=z)\,P\,(X_n=z\,|\,X_0=x) \\ &= \textstyle\sum_{z\in S} p_n(x,z)\,p(z,y) \;=\; \sum_{z\in S}(P^n)_{x,z}\,P_{z,y} \;=\; (P^{n+1})_{x,y} \end{aligned}$$

from which (5.2) follows by induction. □

Another alternative that helps visualizing the evolution of the process is to represent the transition probabilities in the form of a weighted directed graph in which the vertices represent the states and the weight of the arrow $x\to y$ the transition probability $p(x,y)$. For the Ehrenfest chain with $M=4$ particles discussed in Example 4.3, this gives the picture of Figure 5.

Classification of states. A powerful tool to study a Markov chain is to classify the elements of the state space into so-called **communication classes**, which allows to decompose the process into elementary pieces.

Definition 5.2 (communication classes) – Two states $x,y\in S$ are said to communicate, which we write $x\leftrightarrow y$, whenever

$$p_n(x,y)>0 \quad\text{and}\quad p_m(y,x)>0 \quad \text{for some } n,m\geq 0.$$

Lemma 5.3 – The binary relation $\leftrightarrow$ is an equivalence relation, thus inducing a partition of the state space S into communication classes.

Proof. Since $p_0(x,x)=1$ for all $x\in S$, the relation is reflexive. The symmetry follows from the definition. To prove the transitivity, assume that $x\leftrightarrow y$ and $y\leftrightarrow z$ so

$$p_n(x,y)>0 \quad\text{and}\quad p_m(y,z)>0 \quad \text{for some } n,m\geq 0.$$

Conditioning on the possible values after n time steps, we deduce

$$\begin{aligned} p_{n+m}(x,z) &= \sum_{w\in S} P(X_{n+m}=z \,|\, X_n=w)\, P(X_n=w \,|\, X_0=x) \\ &= \sum_{w\in S} p_m(w,z)\, p_n(x,w) \;\geq\; p_n(x,y)\, p_m(y,z) \;>\; 0. \end{aligned}$$

By symmetry, we also have $p_k(z,x)>0$ for some $k\in\mathbb{N}$. □

Definition 5.4 (irreducibility) – If there is only one communication class, the Markov chain is said to be irreducible.

In practice, natural questions are whether each state is recurrent or transient, and periodic. These properties turn out to be class properties, i.e., properties which are invariant over all the elements of the same communication class. Recurrence and transience are defined via the so-called **time of first return**: the time of first return to state x is defined by

$$T_x \;:=\; \inf\{n\geq 1 : X_n = x\} \quad \text{for } x\in S.$$

Definition 5.5 (recurrence and transience) – State x is said to be

$$\begin{aligned} &\text{recurrent} \quad \text{if} \quad P(T_x<\infty \,|\, X_0=x) \;=\; 1 \\ &\text{transient} \quad \text{if} \quad P(T_x<\infty \,|\, X_0=x) \;<\; 1. \end{aligned}$$

Lemma 5.6 – State x is recurrent or transient depending on whether

$$\sum_{n\geq 0} p_n(x,x)=\infty \;\; \text{(recurrent)} \qquad \text{or} \qquad \sum_{n\geq 0} p_n(x,x)<\infty \;\; \text{(transient)}.$$

Proof. By the Markov property, the number of visits in state x starting from state x is geometrically distributed therefore, if x is recurrent, the expected number of visits in state x is infinite, whereas if x is transient the expected number of visits is finite. To conclude, we simply observe that, by the monotone convergence theorem 1.6, the expected number of visits is equal to

$$E\left(\sum_{n\geq 0} \mathbf{1}\{X_n=x\} \,|\, X_0=x\right) \;=\; \sum_{n\geq 0} E\left(\mathbf{1}\{X_n=x\} \,|\, X_0=x\right) \;=\; \sum_{n\geq 0} p_n(x,x)$$

which completes the proof. □

Theorem 5.7 – Recurrence and transience are class properties.

Proof. Assume that $x\leftrightarrow y$. Then,

$$p_n(x,y)>0 \quad \text{and} \quad p_m(y,x)>0 \quad \text{for some } n,m\geq 0.$$

Observing also that

$$p_{k+n+m}(y,y) \;\geq\; p_m(y,x)\; p_k(x,x)\; p_n(x,y)$$

we deduce that, if x is recurrent then

$$\sum_{k\geq 0} p_k(y,y) \;\geq\; \sum_{k\geq 0} p_{k+n+m}(y,y) \;\geq\; p_m(y,x)\, p_n(x,y)\; \sum_{k\geq 0} p_k(x,x) \;=\; \infty$$

therefore y is recurrent according to Lemma 5.6. By symmetry, state x is recurrent if and only if state y is recurrent, and so x is transient if and only if y is transient. □

If a Markov chain is irreducible then, according to Theorem 5.7, all the states are recurrent or all the states are transient, and the process itself is said to be recurrent/transient. Note that irreducible Markov chains with a finite state space are recurrent, but infinite state space does not imply transience, as shown in Section 6 for one- and two-dimensional random walks.

Stationary distribution and limiting behavior. Another natural question is to know whether the process equilibrates at some distribution and how often the recurrent states are visited. This question can be answered using simple linear algebra.

Definition 5.8 (stationary distribution) – The vector $(\pi_x : x\in S)$ is said to be a stationary distribution of the Markov chain (X_n) whenever

$$P(X_0=x)=\pi_x \text{ for all } x\in S \;\;\Longrightarrow\;\; P(X_n=x)=\pi_x \text{ for all } x\in S \text{ and } n\in\mathbb{N}.$$

Theorem 5.9 – Any stationary distribution π satisfies $\pi\cdot P=\pi$.

Proof. Assume that $P(X_0=x)=\pi_x$ for all $x\in S$. Then, for all $y\in S$,

$$\begin{aligned} \pi_y &= P(X_{n+1}=y) \;=\; \sum_{x\in S} P(X_{n+1}=y \,|\, X_n=x)\, P(X_n=x) \\ &= \sum_{x\in S} p(x,y)\,\pi_x \;=\; (\pi\cdot P)_y \end{aligned}$$

which implies that $\pi\cdot P=\pi$. □

Definition 5.10 (period) – The period of x is $\mathrm{per}(x) := \gcd\{n\geq 1 : p_n(x,x)>0\}$.

Lemma 5.11 – The period is a class property.

Proof. Assume that $x\leftrightarrow y$. Then,

$$p_n(x,y)>0 \quad \text{and} \quad p_m(y,x)>0 \quad \text{for some } n,m\geq 0.$$

For all $k>0$ such that $p_k(y,y)>0$, we have

$$\begin{aligned} p_{n+m}(x,x) &\geq p_n(x,y)\; p_m(y,x) \;>\; 0 \\ p_{n+k+m}(x,x) &\geq p_n(x,y)\; p_k(y,y)\; p_m(y,x) \;>\; 0 \end{aligned}$$

therefore $\mathrm{per}(x)$ divides both $n+m$ and $n+k+m$ so it divides k. Since $\mathrm{per}(y)$ is the greatest integer that divides all such k, we deduce that $\mathrm{per}(x)\leq\mathrm{per}(y)$. By symmetry, we also have that $\mathrm{per}(y)\leq\mathrm{per}(x)$ from which it follows that both periods are equal. □

Theorem 5.12 – If the Markov chain is finite, irreducible and aperiodic (all states have period one) then there is a unique stationary distribution π. Moreover,

$$\lim_{n\to\infty} P(X_n=x)=\pi_x \quad \text{for all } x\in S$$

regardless of the initial state/distribution.

Proof. This can be proved by using the Perron-Frobenius theorem in linear algebra but here is a more probabilistic approach that can be found in Durrett [12]. Consider two independent copies of the Markov chain, say (X_n) and (Y_n). Then,

$$Z_n := (X_n, Y_n) \;\text{ is a Markov chain with }\; \bar{p}((x,y),(x',y')) = p(x,x')\, p(y,y').$$

Denote by T the first time this Markov chain hits the diagonal:

$$T \;:=\; \inf\{n\geq 0 : X_n = Y_n\}.$$

Using that P is irreducible and aperiodic, it can be proved that (Z_n) is irreducible, and since the state space is finite, it is also recurrent, therefore

$$P(T > n) \;\to\; 0 \quad \text{as } n \to \infty. \tag{5.3}$$

After the hitting time T, both coordinates have the same distribution therefore

$$\begin{aligned} P(X_n = y) &= P(Y_n = y \text{ and } T \le n) + P(X_n = y \text{ and } T > n) \\ &\le P(Y_n = y) + P(X_n = y \text{ and } T > n). \end{aligned} \tag{5.4}$$

By symmetry, we also have

$$P(Y_n = y) \;\le\; P(X_n = y) + P(Y_n = y \text{ and } T > n). \tag{5.5}$$

Combining (5.4) and (5.5), we get

$$\begin{aligned} \textstyle\sum_{y\in S} |P(X_n = y) - P(Y_n = y)| &\le \textstyle\sum_{y\in S} P(X_n = y \text{ and } T > n) \\ &+ \textstyle\sum_{y\in S} P(Y_n = y \text{ and } T > n) \;\le\; 2\,P(T > n). \end{aligned} \tag{5.6}$$

To conclude, we let $P(Y_0 = y) := \pi_y$ for all $y \in S$ so that the second coordinate is always at stationarity and $X_0 = x$ in which case the previous inequality (5.6) becomes

$$\textstyle\sum_{y\in S} |P(X_n = y \,|\, X_0 = x) - \pi_y| \;\le\; 2\,P(T > n) \;\to\; 0$$

as $n \to \infty$ according to (5.3). This completes the proof. □

Example 5.13 – The following two examples show that irreducibility and aperiodicity are necessary conditions in the previous theorem:

$$P_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad P_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The Markov chain with transition matrix P_1 is not irreducible and has infinitely many stationary distributions whereas the Markov chain with transition matrix P_2 is not aperiodic (periodic with period 2) and the probabilities in Theorem 5.12 do not converge.

6. Symmetric simple random walks

Symmetric simple random walks describe the position of a walker moving on the d-dimensional regular lattice. At each time step, the walker chooses one direction uniformly at random and moves one unit length in this direction. This process is important to study a number of stochastic processes, and will appear again in the analysis of the voter model in Section 20. Symmetric random walks can more generally describe random motions on a graph: the walker located at vertex x moves to one of the $\deg(x)$ vertices connected to x chosen uniformly at random. This defines a Markov chain which is irreducible if and only if the graph is connected. In particular, if the graph is connected, the states/vertices are all transient or all recurrent according to Theorem 5.7, and we call the graph itself transient or recurrent. Note that this notion is not well defined for non-connected graphs since some vertices can be recurrent and other ones transient.

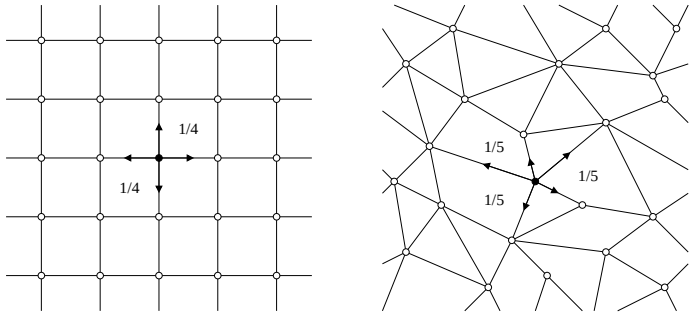


Figure 4. Symmetric simple random walks on graphs.

Recurrence in low dimensions. In the one dimensional case, we will look more generally at the symmetric and asymmetric random walks, i.e., the discrete-time Markov chain (X_n) whose state space is given by $S := \mathbb{Z}$ and transition probabilities given by

$$X_{n+1} := \begin{cases} X_n + 1 & \text{with probability } p \\ X_n - 1 & \text{with probability } q \end{cases} \qquad \text{where } \; p + q = 1.$$

Theorem 6.1 – The random walk (X_n) is recurrent if and only if $p = q$. In particular, the one-dimensional regular lattice is a recurrent graph.

Proof. To begin with, we observe that

$$\sum_{n=0}^{\infty} p_n(0,0) \;=\; \sum_{n=0}^{\infty} P(X_{2n} = 0 \,|\, X_0 = 0) \;=\; \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n.$$

Using Stirling's formula

$$n! \;\sim\; n^n e^{-n} \sqrt{2\pi n} \quad \text{as } \; n \to \infty \tag{6.1}$$

to estimate the previous sum, we get

$$\binom{2n}{n} p^n q^n \;=\; \frac{(2n)!}{n!\,n!}\, p^n q^n \;\sim\; \frac{(2n)^{2n}\, e^{-2n}\, \sqrt{4\pi n}}{n^{2n}\, e^{-2n}\, 2\pi n}\, p^n q^n \;=\; \frac{(4\,pq)^n}{\sqrt{\pi n}}$$

which gives recurrence if and only if $p = q = 1/2$ according to Lemma 5.6. □

Here is another, more elegant and less technical, proof based on a symmetry argument used to prove site recurrence of a system of annihilating random walks by Adelman [1].

Proof. Assume without loss of generality that the random walk first jumps to the right. Then, the event A that it never returns to 0 is the event that it reaches state 2 before state 0 starting from 1, then state 4 before state 0 starting from 2, and so on. The proof follows from the fact that all these events have probability one half. To make the argument precise, let

$$T_x \;:=\; \inf\{n > 0 : X_n = x\} \quad \text{for all } x \in \mathbb{Z}.$$

Then, using a first step analysis and symmetry, we obtain

$$\begin{aligned} P(A) &= P_0(A \,|\, X_1 = -1)\, P_0(X_1 = -1) \;+\; P_0(A \,|\, X_1 = 1)\, P_0(X_1 = 1) \\ &= P_0(A \,|\, X_1 = 1) \;=\; P(T_2 < T_0 \,|\, X_0 = 1)\, P(T_4 < T_0 \,|\, X_0 = 2) \;\cdots \\ &\qquad \cdots\; P(T_{2^n} < T_0 \,|\, X_0 = 2^{n-1}) \;\cdots\; = \; (1/2)^\infty \;=\; 0. \end{aligned}$$

This complete the proof. □

Theorem 6.2 – The two-dimensional lattice is recurrent.

Proof. Thinking of loops with length $2n$ as being made of $2k$ horizontal and $2(n-k)$ vertical segment lines for some $k = 0, 1, \ldots, n$, we obtain

$$p_{2n}(0,0) \;=\; \sum_{k=0}^{n} \binom{2n}{2k}\binom{2k}{k}\binom{2n-2k}{n-k}\left(\frac{1}{4}\right)^{2n} = \; \binom{2n}{n}\left(\frac{1}{4}\right)^{2n} \sum_{k=0}^{n} \binom{n}{k}^2.$$

Now, observe that

$$\sum_{k=0}^{n} \binom{n}{k}^2 \;=\; \sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k} \;=\; \binom{2n}{n}$$

since both sides represent the number of subsets of cardinal n of a set of cardinal $2n$. Using again Stirling's formula (6.1) indicates that for n large we have

$$p_{2n}(0,0) \;=\; \binom{2n}{n}^2\left(\frac{1}{4}\right)^{2n} \sim\; \frac{1}{\pi n} \qquad \text{and} \qquad \sum_{n=0}^{\infty} p_n(0,0) \;=\; \infty$$

therefore the two-dimensional lattice is a recurrent graph according to Lemma 5.6. □

Transience in high dimensions. We focus on the three-dimensional case only and will use later the connection with electrical networks to deduce results about symmetric random walks in higher dimensions and on connected graphs more general than regular lattices.

Theorem 6.3 – The three-dimensional lattice is transient.

Proof. Thinking of loops with length $2n$ as being made of $2i$, $2j$ and $2k$ segment lines in the direction of each of the three axes, respectively, we obtain the equation

$$\begin{aligned} p_{2n}(0,0) \;&=\; \sum_{i+j+k=n} \binom{2n}{2i}\binom{2n-2i}{2j}\binom{2i}{i}\binom{2j}{j}\binom{2k}{k}\left(\frac{1}{6}\right)^{2n} \\ &=\; \binom{2n}{n} \sum_{i+j+k=n} \left[\frac{n!}{i!\cdot j!\cdot k!}\right]^2 \left(\frac{1}{6}\right)^{2n} \;=\; \binom{2n}{n}\left(\frac{1}{12}\right)^{n} \sum_{i+j+k=n} \left[\frac{n!}{i!\cdot j!\cdot k!}\right]^2 \left(\frac{1}{3}\right)^{n}. \end{aligned}$$

Since in addition

$$i!\cdot j!\cdot k! \;\geq\; \big((n/3)!\big)^3 \qquad \text{and} \qquad \sum_{i+j+k=n} \frac{n!}{i!\cdot j!\cdot k!}\left(\frac{1}{3}\right)^{n} \;=\; \left(\frac{1}{3}+\frac{1}{3}+\frac{1}{3}\right)^{n} \;=\; 1$$

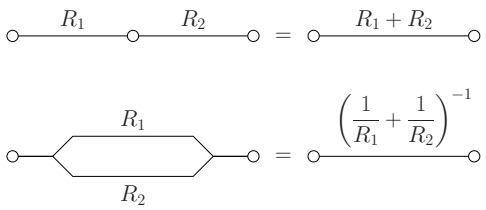


Figure 5. The basic rules to compute electrical resistance.

we deduce that

$$\begin{aligned} p_{2n}(0,0) \;&\leq\; \binom{2n}{n}\left(\frac{1}{12}\right)^{n} \frac{n!}{[(n/3)!]^3} \;=\; \left(\frac{1}{12}\right)^{n} \frac{(2n)!}{n!\cdot[(n/3)!]^3} \\ &\sim\; \left(\frac{1}{12}\right)^{n} \frac{(2n)^{2n}\sqrt{4\pi n}}{n^n\sqrt{2\pi n}\cdot(n/3)^n(2\pi n/3)^{3/2}} \;=\; \frac{\text{constant}}{n^{3/2}} \end{aligned}$$

which implies that the three-dimensional lattice is transient. □

Electrical networks. To study symmetric simple random walks on more general connected graphs, one trick is to think of the graph as an electrical network in which each edge has a certain conductance. For a complete picture of the connections between Markov chains and electrical networks, we refer to Doyle and Snell [9]. Here, we only focus on the relationship between recurrence/transience of symmetric random walks and electrical resistance. The random walk (X_n) associated to an electrical network is defined by the transition probabilities

$$p(x,y) \;:=\; P(X_{n+1} = y \,|\, X_n = x) \;:=\; c(x,y)/\textstyle\sum_{z\sim x} c(x,z)$$

where $c(x,y)$ is the conductance along the edge (x,y). The symmetric simple random walk is obtained by assuming that all the edges have the same conductance since in this case

$$p(x,y) \;:=\; P(X_{n+1} = y \,|\, X_n = x) \;:=\; (\deg(x))^{-1}.$$

Then, we have the following important result.

Theorem 6.4 – The random walk (X_n) is recurrent if and only if the effective resistance of the electrical network between a nominated point and points at infinity is infinite.

The resistance of an edge is defined as the inverse of its conductance, and the resistance between two vertices of the network can be computed by applying successively the two basic rules illustrated in Figure 6. In practice, the resistance between vertices of a complex network can be difficult, if not impossible, to compute, however the following (intuitively obvious) monotonicity law due to Rayleigh is sufficient to obtain very powerful results.

Rayleigh's monotonicity law – If the resistance of an edge is increased/decreased, then the resistance between any two vertices of the network can only increase/decrease.

The last step is to observe that removing an edge from an electrical network is equivalent to setting its conductance equal to zero, i.e., its resistance equal to infinity. From this, the Rayleigh's monotonicity law, and Theorems 6.2–6.3, we deduce that any subgraph of the two-dimensional lattice is recurrent whereas any supergraph of the three-dimensional lattice is transient.

7. Convergence of the martingales

Similarly to Markov chains, martingales include a temporal structure. Martingales can usually be thought of as modeling fair games which translates in the context of competition models into fair/neutral competitions. Recall that the discrete-time process $(X_n : n \in \mathbb{N})$ is said to be a martingale, respectively a submartingale, a supermartingale, whenever

$$\begin{array}{rcll} E(X_{n+1} \,|\, \mathcal{F}_n) & = & X_n \text{ for all } n & \text{fair game (martingale)} \\ & \geq & X_n \text{ for all } n & \text{favorable game (submartingale)} \\ & \leq & X_n \text{ for all } n & \text{unfavorable game (supermartingale).} \end{array}$$

Optimal stopping theorem. Note that a simple induction implies that

$$EX_n \;=\; E(E(X_n \,|\, \mathcal{F}_{n-1})) \;=\; EX_{n-1} \;=\; \cdots \;=\; EX_0. \tag{7.1}$$

The optimal stopping theorem shows that, under certain assumptions, this property is true looking more generally at **stopping times** instead of deterministic times: a random time T is a stopping time such that the event that $T = n$ belongs to $\mathcal{F}_n$ for all n, which can be proved to satisfy

$$T \;:=\; \inf\{n : X_n \in B\} \quad \text{for some Borel set } B \subset \mathbb{R}.$$

More precisely, we have the following theorem.

Theorem 7.1 (optimal stopping theorem) – Assume that (X_n) is a bounded martingale and that T is an almost surely finite stopping time. Then, $EX_T = EX_0$.

Proof. Using that T is a stopping time, it can be proved that (X_n^T), the process stopped at T, is again a martingale. Moreover, we have

$$E|X_n^T| \;\leq\; \sup_n E|X_n| \;<\; \infty \quad \text{and} \quad \lim_{n\to\infty} X_n^T \;=\; X_T \;\text{ a.s.}$$

so it follows from (7.1) and the dominated convergence theorem 1.8 that

$$EX_T \;=\; \lim_{n\to\infty} EX_n^T \;=\; \lim_{n\to\infty} EX_0^T \;=\; EX_0^T \;=\; EX_0.$$

This completes the proof. □

As an application of the optimal stopping theorem, we study the gambler problem. Consider a gambler who plays until she is ruined or has N dollars. At each game, the gambler wins one dollar with probability p or loses one dollar with probability $q := 1 - p$ therefore the gambler's capital after the nth game, X_n, defines a Markov chain.

Theorem 7.2 (gambler ruin problem) – Let $a := q/p$. Then, given that the gambler starts with $X_0 = x$ dollars, the probability that she quits winner is

$$\begin{array}{rcll} p_x & = & (1-a^x)(1-a^N)^{-1} & \text{when} \quad p \neq q \\ & = & x/N & \text{when} \quad p = q. \end{array} \tag{7.2}$$

First, we recall the traditional proof based on a first-step analysis.

Proof. Decomposing according to the possible values of X_1 yields

$$p_x \;=\; \textstyle\sum_{z=x\pm1} P(X_n = N \text{ for some } n \,|\, X_1 = z)\; P_x(X_1 = z) \;=\; p\,p_{x+1} + q\,p_{x-1}$$

therefore $(p+q)\,p_x = p\,p_{x+1} + q\,p_{x-1}$ and

$$\begin{array}{rcl} p_{x+1} - p_x & = & (q/p)\,(p_x - p_{x-1}) \\ & = & a\,(p_x - p_{x-1}) \;=\; \cdots \;=\; a^x\,(p_1 - p_0) \;=\; a^x\,p_1. \end{array}$$

Summing the differences we thus obtain

$$\begin{array}{rcll} p_x \;=\; \sum_{z=0,1,\ldots,x-1} (p_{z+1} - p_z) & = & (1-a^x)(1-a)^{-1}\,p_1 & \text{when} \quad a \neq 1 \\ & = & x\,p_1 & \text{when} \quad a = 1 \end{array}$$

where, since $p_N = 1$, we have

$$\begin{array}{rcll} p_1 & = & (1-a)(1-a^N)^{-1} & \text{when} \quad a \neq 1 \\ & = & 1/N & \text{when} \quad a = 1. \end{array}$$

The result follows. □

The following is a less computational and more elegant proof based on Theorem 7.1.

Proof. Let $T := \min\{n : X_n \in \{0, N\}\}$. When $p = q$, the process (X_n) is a martingale therefore the optimal stopping theorem implies that

$$\begin{array}{rcl} EX_T & = & EX_0 \;=\; x \\ & = & N\,p_x + 0\,(1-p_x) \;=\; N\,p_x \quad \text{so} \quad p_x \;=\; x/N. \end{array}$$

When $p \neq q$, we notice that (a^{X_n}) is a martingale. Indeed,

$$E(a^{X_{n+1}} \,|\, \mathcal{F}_n) \;=\; p\,a^{X_n+1} + q\,a^{X_n-1} \;=\; a^{X_n}(pa + qa^{-1}) \;=\; a^{X_n}.$$

In particular, the optimal stopping theorem implies that

$$\begin{array}{rcl} E\,a^{X_T} & = & E\,a^{X_0} \;=\; a^x \\ & = & p_x\,a^N + (1-p_x)\,a^0 \;=\; p_x\,(a^N - 1) + 1 \quad \text{so} \quad p_x \;=\; (1-a^x)(1-a^N)^{-1} \end{array}$$

which completes the proof. □

Dubins' inequality. Let $(X_n : n \in \bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\})$ be a supermartingale defined at time infinity. For all $a < b$, we define the number of upcrossings

$$u(a,b) \;:=\; \#\text{ upcrossings of the interval } (a,b).$$

More precisely, let $T_0 := 0$ and define

$$T_{2k+1} \;:=\; \inf\{n \geq T_{2k} : X_n \leq a\} \quad \text{and} \quad T_{2k+2} \;:=\; \inf\{n \geq T_{2k+1} : X_n \geq b\}$$

for all $k \geq 0$. Then $u(a,b) := \sup\{k : T_{2k} < \infty\}$.

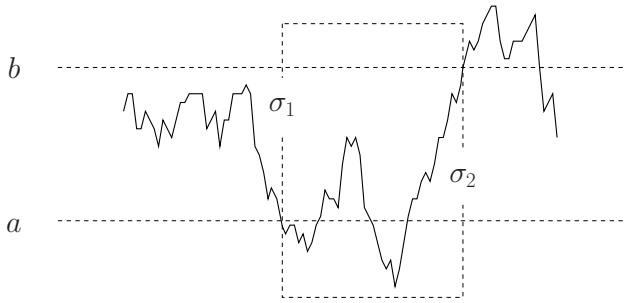


FIGURE 6. Picture of the $(k+1)$th upcrossing from a to b.

Lemma 7.3 (Dubins' inequality) – For all $k \in \mathbb{N}$, we have

$$P\left(u(a,b) > k\right) \;\leq\; (b-a)^{-1}\, E\left[(X_\infty - a)^-\, \mathbf{1}\{u(a,b) = k\}\right]. \tag{7.3}$$

PROOF. To prove this inequality, we first observe that, since the process $(X_n - a)$ also is a supermartingale, we can assume that $a = 0$ without loss of generality. Let the time interval when the $(k+1)$th upcrossing occurs be $(\sigma_1, \sigma_2) := (T_{2k+1}, T_{2k+2})$. Note that

$$\{u(0,b) > k\} \;=\; \{\sigma_2 < \infty\} \;\subset\; \{X_{\sigma_2} \geq b\}$$

from which it follows that

$$\begin{aligned} P\left(u(0,b) > k\right) &= E\left(\mathbf{1}\{u(0,b) > k\}\right) \;\leq\; (1/b)\, E\left[X_{\sigma_2}\, \mathbf{1}\{u(0,b) > k\}\right] \\ &\leq (1/b)\, E\left[X^+_{\sigma_2}\, \mathbf{1}\{u(0,b) > k\}\right] \;\leq\; (1/b)\, E\left[X^+_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty\}\right]. \end{aligned} \tag{7.4}$$

Since (X_n) is a supermartingale and $X_{\sigma_1} \leq 0$ on the event that $\sigma_1 < \infty$, we have

$$E\left[X_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty\}\right] \;\leq\; E\left[X_{\sigma_1}\, \mathbf{1}\{\sigma_1 < \infty\}\right] \;\leq\; 0$$

which implies that

$$E\left[X^+_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty\}\right] \;\leq\; E\left[X^-_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty\}\right]. \tag{7.5}$$

Combining (7.4) and (7.5), we deduce that

$$\begin{aligned} P\left(u(0,b) > k\right) &\leq (1/b)\, E\left[X^-_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty\}\right] \\ &\leq (1/b)\, E\left[X^-_{\sigma_2}\, \mathbf{1}\{\sigma_1 < \infty,\; X_{\sigma_2} \leq 0\}\right] \\ &\leq (1/b)\, E\left[X^-_{\sigma_2}\, \mathbf{1}\{u(0,b) = k\}\right] \;\leq\; (1/b)\, E\left[X^-_\infty\, \mathbf{1}\{u(0,b) = k\}\right] \end{aligned}$$

which completes the proof. □

Martingale convergence theorem. This theorem follows from Dubins' inequality and states that, under certain assumptions, almost all trajectories of a supermartingale converge to a fixed value, i.e., do not oscillate or drift off to infinity. Note however that this limiting value might depend on the trajectory so we have almost sure convergence of the process to a random variable, as opposed to a deterministic limit, as time goes to infinity.

Theorem 7.4 (martingale convergence theorem) – Let (X_n) be either a supermartingale or a submartingale such that $\sup_n E\,|X_n| < \infty$. Then,

$$X_n \;\longrightarrow\; X_\infty \quad \text{with probability 1}$$

where X_∞ is a random variable with $E\,|X_\infty| < \infty$.

PROOF. First, we fix $n > 0$ and let

$$u_n(a,b) \;:=\; \#\text{ upcrossings that occur by time } n.$$

Applying Dubins' inequality (7.3) to the process stopped at time n and using Fubini's theorem 1.10 for positive random variables imply that

$$\begin{aligned} E\,u_n(a,b) &= \textstyle\sum_{k\geq 1} k\, P\left(u_n(a,b) = k\right) \;=\; \sum_{k\geq 1} \sum_{j\leq k-1} P\left(u_n(a,b) = k\right) \\ &= \textstyle\sum_{j\geq 0} \sum_{k\geq j+1} P\left(u_n(a,b) = k\right) \;=\; \sum_{j\geq 0} P\left(u_n(a,b) > j\right) \\ &\leq (b-a)^{-1} \textstyle\sum_{j\geq 0} E\left[(X_n - a)^-\, \mathbf{1}\{u_n(a,b) = j\}\right] \\ &= (b-a)^{-1}\, E\left((X_n - a)^-\right). \end{aligned}$$

By the monotone convergence theorem 1.6, we deduce that

$$\begin{aligned} E\,u(a,b) \;=\; \lim_{n\to\infty} E\,u_n(a,b) &\leq (b-a)^{-1} \lim_{n\to\infty} E\left((X_n - a)^-\right) \\ &\leq (b-a)^{-1} \sup_n E\left(X_n^- + |a|\right) \;<\; \infty \end{aligned}$$

therefore $P\left(u(a,b) < \infty\right) = 1$ for all $a < b$. Since $\mathbb{Q}$ is countable, we also have

$$P\left(u(a,b) < \infty \text{ for all } a, b \in \mathbb{Q}, a < b\right) \;=\; 1$$

which proves, using that the topological closure of $\mathbb{Q}$ is equal to $\mathbb{R}$ (each real number can be written as the limit of a sequence of rational numbers), that

$$X_n \;\longrightarrow\; X_\infty \quad \text{with probability 1}$$

where the random variable X_∞ satisfies $E\,|X_\infty| \leq \sup_n E\,|X_n| < \infty$. □

Finally, note that the trajectories of the one-dimensional symmetric random walk do not converge. The reason is that, though the process is a martingale, it is not bounded. The trajectories do not converge either when looking at the random walk on a finite interval and reflecting boundary conditions. In this case, the reason is that the process is no longer a martingale. However, the theorem holds with absorbing boundaries, i.e., the gambler ruin model.

8. Branching processes

Branching processes are some of the simplest stochastic processes describing single-species dynamics. They ignore the presence of a spatial structure but allow to account for general offspring distributions. In this section, we prove that the long-term behavior of the process depends on the expected value of the offspring distribution.

Description of the model. Branching processes evolve in discrete time with the state at generation n, say X_n, being the number of individuals in the population. Independently of each other, individuals present in the system at generation n produce a random number of offspring, say k offspring with probability p_k. The population at generation $n+1$ consists exclusively of the offspring produced at the previous generation n.

Mathematically, the process (X_n) is a discrete-time Markov chain with state space the set of all nonnegative integers. To describe the dynamics, we let $Y_{n,j}$ be a collection of independent identically distributed random variables with the same probability mass function

$$P(Y_{n,j} = k) \;:=\; p_k \quad \text{for all } n, j \in \mathbb{N}.$$

Thinking of $Y_{n,j}$ as the number of offspring produced by individual j at generation n if there are indeed at least j individuals at that generation, we have

$$X_{n+1} \;=\; \textstyle\sum_{j=1,2,\dots,X_n} Y_{n,j}. \tag{8.1}$$

Connection with martingales. To study branching processes, the first step is to exhibit some connection with martingales, which is a very powerful tool to understand their behavior. To begin with, we note that, according to (8.1), we have

$$E(X_{n+1} \,|\, \mathcal{F}_n) \;=\; E(X_{n+1} \,|\, X_n) \;=\; E(Y_{n,1} + \cdots + Y_{n,X_n} \,|\, X_n) \;=\; \mu\, X_n$$

where $\mu := E(Y_{n,i})$. This implies that

$$E(\mu^{-(n+1)}\, X_{n+1} \,|\, \mathcal{F}_n) \;=\; \mu^{-n}\, X_n$$

therefore the process $(\mu^{-n}\, X_n)$ is a martingale. This suggests that the process exhibits three types of behaviors depending on whether μ is smaller than, equal to, or larger than one, which we refer to as subcritical, critical, and supercritical regimes, respectively.

In the subcritical case $\mu < 1$, starting with one individual, we have

$$P(X_n > 0) \;=\; \textstyle\sum_{k\geq 1} P(X_n = k) \;\leq\; \sum_{k\geq 1} k\, P(X_n = k) \;=\; E(X_n) \;=\; \mu^n$$

therefore the population goes extinct exponentially fast. In the critical case, the martingale convergence theorem 7.4 gives the following result.

Theorem 8.1 – Assume that $p_1 \neq 1$ and $\mu = 1$. Then $\lim_{n\to\infty} P(X_n = 0) = 1$.

Proof. Since the process (X_n) is a martingale with $E\,|X_n| = EX_0 = 1$, the martingale convergence theorem 7.4 implies the existence of X_∞ such that

$$X_n \;\longrightarrow\; X_\infty \quad \text{with probability 1.}$$

Moreover, since $p_1 \neq 1$, for all $k \geq 1$,

$$P(X_n \to k \text{ as } n \to \infty) \;=\; 0.$$

This implies that $X_\infty = 0$ almost surely and proves the theorem. □

The fact that the population dies out at criticality is common in stochastic processes, e.g., bond and site percolations, contact process, but is sometimes difficult to prove.

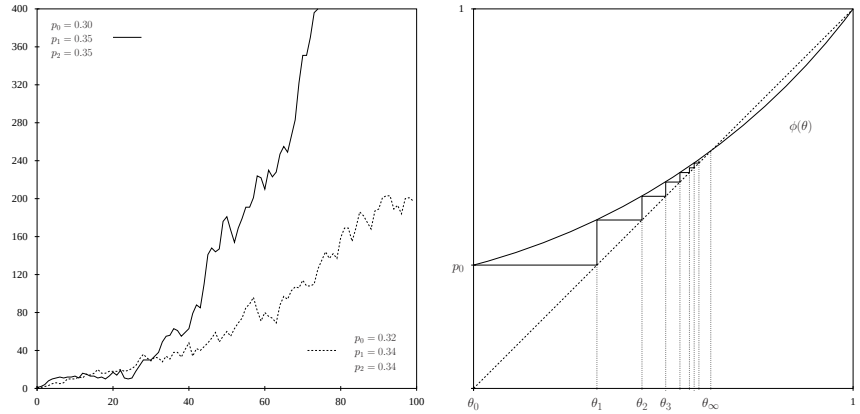


Figure 7. Two realizations of branching processes with different offspring distribution and schematic representation of the graph of the moment generating function of the offspring distribution when $\mu > 1$.

Probability of survival. In both the subcritical and critical cases, the population goes extinct almost surely. We now prove that the survival probability starting with one individual is strictly positive in the supercritical case $\mu > 1$ and compute its value explicitly using the moment generating function of the offspring distribution.

Theorem 8.2 – Assume that $\mu > 1$. Then $P(X_n > 0 \text{ for all } n > 0) > 0$.

Proof. The moment generating function of the offspring distribution is given by

$$\phi(\theta) \;:=\; \textstyle\sum_{k\geq 0} p_k\, \theta^k \quad \text{for all } \;\theta \in [0,1].$$

Letting $\theta_n := P(X_n = 0)$, conditioning on the possible values of the number X_1 of individuals at the first generation, and using independence, we obtain

$$\begin{aligned} \theta_{n+1} \;&=\; \textstyle\sum_{k\geq 0} P(X_{n+1} = 0 \,|\, X_1 = k)\, P(X_1 = k \,|\, X_0 = 1) \\ &=\; \textstyle\sum_{k\geq 0} p_k\, [P(X_{n+1} = 0 \,|\, X_1 = 1)]^k \;=\; \sum_{k\geq 0} p_k\, \theta_n^k \;=\; \phi(\theta_n). \end{aligned}$$

In particular, observing that

$$\theta_\infty \;:=\; P(X_n = 0 \text{ for some } n \geq 0) \;=\; \lim_{n\to\infty} P(X_n = 0) \;=\; \lim_{n\to\infty} \theta_n$$

we deduce that $\phi(\theta_\infty) = \theta_\infty$, i.e., the probability of extinction θ_∞ is a fixed point of the moment generating function. Now, since $\mu > 1$, we must have $p_k > 0$ for some $k \geq 2$, therefore

$$\begin{aligned} \phi'(\theta) \;&=\; \textstyle\sum_{k\geq 1} k\, p_k\, \theta^{k-1} \;>\; 0 \\ \phi''(\theta) \;&=\; \textstyle\sum_{k\geq 2} k\,(k-1)\, p_k\, \theta^{k-2} \;>\; 0 \end{aligned}$$

for all $\theta > 0$. Observing in addition that

$$\phi(0) \;=\; p_0 \;<\; 1 \quad \text{and} \quad \phi'(1) \;=\; \textstyle\sum_{k\geq 1} k\, p_k \;=\; \mu \;>\; 1$$

we obtain the graph of Figure 7, which implies that the moment generating function has a unique fixed point in the open interval $(0,1)$. Since $\theta_0 = 0$, the sequence (θ_n) converges to this fixed point: the probability of survival is equal to one minus the value of this fixed point. □

9. Poisson point and Poisson processes

Poisson point processes model the spatial distribution at a given time of objects that appear randomly and homogeneously in space, e.g., distribution of stars and galaxies in three dimensions, distribution of plants in two dimensions, etc. However, Poisson point processes are most useful in one dimension where the position of an object or entity can be thought of as the time of an event. In this context, they allow one to define a temporal structure and are employed to construct continuous-time Markov chains such as interacting particle systems.

Defining a spatial distribution. For concreteness, think of the d-dimensional Euclidean space as the universe and that countably many entities (points) are scattered in a random manner in this infinite universe. There are multiple ways to define the spatial distribution of these entities but natural assumptions are the following ones:

1. **Independence** – The number of entities in non-overlapping bounded regions of the universe are independent random variables.
2. **Homogeneity** – The number of entities in a given bounded region only depends upon the volume or Lebesgue measure of this region.
3. **Intensity** – The probability that a bounded region with Lebesgue measure v contains one entity is given by $\mu v + o(v)$ as $v \to 0$.
4. **Separability** – The probability that a bounded region with Lebesgue measure v contains at least two entities is given by $o(v)$ as $v \to 0$.

The process that takes into account the four assumptions above is the Poisson point process with intensity μ. Formally, this is a process $X := \{X(A) : A \in \mathcal{B}(\mathbb{R}^d)\}$ where

$$\begin{aligned} \mathcal{B}(\mathbb{R}^d) &:= \text{the Borel } \sigma\text{-algebra on the } d\text{-dimensional Euclidean space} \\ X(A) &:= \text{the number of entities in the Borel set } A. \end{aligned}$$

From now on, we write $X = ppp(\mu)$ for short. The four axioms above can be rewritten in terms of this process, which gives the following more rigorous definition.

Definition 9.1 (Poisson point process) – The process $X = ppp(\mu)$ whenever

(1) $X(A_1), X(A_2), \ldots, X(A_n)$ are independent whenever $A_i \cap A_j = \varnothing$ for $i \neq j$.
(2) The distribution of $X(A)$ only depends on $v(A) :=$ the Lebesgue measure of A.
(3) $P(X(A) = 1) = \mu v(A) + o(v(A))$ as $v(A) \to 0$.
(4) $P(X(A) \geq 2) = o(v(A))$ as $v(A) \to 0$.

The following theorem gives the probability mass function of $X(A)$ and shows that it is related to the Poisson distribution, which explains the name of the process.

Theorem 9.2 – The process $X = ppp(\mu)$ if and only if

(a) $X(A_1), X(A_2), \ldots, X(A_n)$ are independent whenever $A_i \cap A_j = \varnothing$ for $i \neq j$.
(b) and for every bounded Borel set A, we have $X(A) = \text{Poisson}(\mu v(A))$.

Proof. We need to prove two implications.

Sufficient condition – Assume that X satisfies (a) and (b). Then, it is clear that conditions (1) and (2) in Definition 9.1 are satisfied. Moreover,

$$\begin{aligned} P(X(A) = 1) &= \mu v\, e^{-\mu v} \\ &= \mu v\,(1 - \mu v + o(v)) \;=\; \mu v + o(v) \quad \text{as} \quad v := v(A) \to 0. \end{aligned}$$

Similarly, we have

$$\begin{aligned} P(X(A) \geq 2) &= 1 - (1 + \mu v)e^{-\mu v} \\ &= 1 - (1 + \mu v)(1 - \mu v + o(v)) \;=\; o(v) \quad \text{as} \quad v := v(A) \to 0. \end{aligned}$$

Necessary condition – Conversely, assume that $X = ppp(\mu)$. Condition (a) is obvious so we only need to check (b). Fix a bounded Borel set A and introduce a sequence of partitions of A

$$\{A_{1,N}, A_{2,N}, \ldots, A_{N,N}\} \quad \text{for all } N \geq 1 \quad \text{with} \quad v(A_{i,N}) = v/N := v(A)/N.$$

Let Ω_N be the event that $X(A_{i,N}) \geq 2$ for some $i = 1, 2, \ldots, N$. Then,

$$\lim_{N\to\infty} P(\Omega_N) \;\leq\; N \lim_{N\to\infty} P(X(A_{1,N}) \geq 2) \;=\; \lim_{N\to\infty} N\, o(v/N) \;=\; 0.$$

Since in addition the variables $X(A_{i,N})$ are independent,

$$\begin{aligned} P(X(A) = k) &= \lim_{N\to\infty} P(X(A) = k \,|\, \Omega_N^c) \\ &= \lim_{N\to\infty} P(X(A_{1,N}) + \cdots + X(A_{N,N}) = k \,|\, \Omega_N^c) \\ &= \lim_{N\to\infty} \binom{N}{k} \left(\frac{\mu v}{N}\right)^k \left(1 - \frac{\mu v}{N}\right)^{N-k} \;=\; \frac{(\mu v)^k}{k!}\; e^{-\mu v} \end{aligned}$$

which implies that $X(A) = \text{Poisson}(\mu v(A))$. □

Lemma 9.3 (superposition) – Assume that black and white entities are distributed according to two independent Poisson point processes: $X_1 = ppp(\mu_1)$ and $X_2 = ppp(\mu_2)$. Then,

- The resulting set of entities is distributed according to $X = ppp(\mu_1 + \mu_2)$.
- Each entity is independently black with probability $\mu_1(\mu_1 + \mu_2)^{-1}$.

Proof. The first assertion directly follows from the definition but can also be proved from the alternative formulation given by Theorem 9.2 observing that the sum of two independent Poisson random variables is a Poisson random variable with parameter the sum of the parameters:

$$\begin{aligned} P(X(A) = k) &= \sum_{i+j=k} P(X_1(A) = i)\, P(X_2(A) = j) \;=\; \sum_{i+j=k} \frac{(\mu_1 v)^i}{i!}\; e^{-\mu_1 v}\; \frac{(\mu_2 v)^j}{j!}\; e^{-\mu_2 v} \\ &= \sum_{i+j=k} \binom{i+j}{i}\; \mu_1^i\; \mu_2^j\; \frac{v^k}{k!}\; e^{-(\mu_1+\mu_2)v} \;=\; \frac{((\mu_1 + \mu_2)v)^k}{k!}\; e^{-(\mu_1+\mu_2)v} \end{aligned}$$

where $v := v(A)$. To prove the second assertion, we observe that the conditional probability that there is one black entity in A given that there is exactly one entity in A is given by

$$\begin{aligned} P(X_1(A) = 1 \,|\, X(A) = 1) &= P(X_1(A) = 1)\, P(X_2(A) = 0) \\ &\times (P(X_1(A) = 1)\, P(X_2(A) = 0) + P(X_1(A) = 0)\, P(X_2(A) = 1))^{-1} \\ &= \mu_1\, e^{-\mu_1 v}\, e^{-\mu_2 v}\, (\mu_1\, e^{-\mu_1 v}\, e^{-\mu_2 v} + e^{-\mu_1 v}\, \mu_2\, e^{-\mu_2 v})^{-1} \;=\; \mu_1(\mu_1 + \mu_2)^{-1} \end{aligned}$$

which completes the proof. □

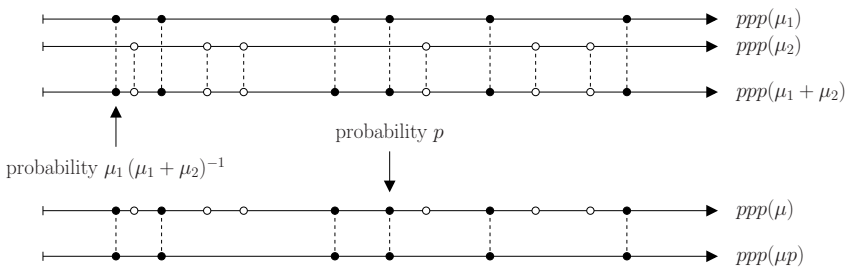


Figure 8. Illustration of the results of Lemmas 9.3 and 9.4.

Lemma 9.4 (extraction) – Assume that $X = ppp(\mu)$ describes the distribution of a set of entities and paint each entity independently in black with probability p. Then,

- The set of black entities is distributed according to $Z = ppp(\mu p)$.

Proof. Given that the number of entities in A is equal to n, the number of black entities in A is binomial with parameters n and p. In particular, using the same notation as above, we have

$$\begin{aligned} P(Z(A)=k) &= \textstyle\sum_{n\geq k} P(Z(A)=k \,|\, X(A)=n)\, P(X_A=n) \\ &= \sum_{n\geq k} \binom{n}{k} p^k (1-p)^{n-k} \frac{(\mu v)^n}{n!} e^{-\mu v} = \frac{p^k}{k!} e^{-\mu v} \sum_{n\geq k} \frac{(\mu v)^n (1-p)^{n-k}}{(n-k)!} \\ &= \frac{(\mu p v)^k}{k!} e^{-\mu v} \sum_{n\geq k} \frac{(\mu(1-p)v)^{n-k}}{(n-k)!} = \frac{(\mu p v)^k}{k!} e^{-\mu v} \sum_{n\geq 0} \frac{(\mu(1-p)v)^n}{n!} \\ &= \frac{(\mu p v)^k}{k!} e^{-\mu v} e^{\mu v(1-p)} = \frac{(\mu p v)^k}{k!} e^{-\mu v p}. \end{aligned}$$

The lemma then follows from Theorem 9.2. □

Defining a temporal structure. In one dimension, one can think of space as time. In this case, Poisson point processes naturally induce continuous-time processes, called Poisson processes, which turn out to be continuous-time Markov chains.

Definition 9.5 (rate) – We say that an event occurs at rate μ if the number $X(A)$ of occurrences of this event in a Borel set $A \subset \mathbb{R}$ satisfies the axioms of Definition 9.1.

Definition 9.6 (Poisson process) – The continuous-time process $(X_t : t \in \mathbb{R}_+)$ is a Poisson process with intensity μ whenever the variable X_t equals the number of entities in the temporal interval $[0,t]$ for a Poisson point process with intensity μ.

By Theorem 9.2, the process (X_t) is a Poisson process with intensity μ if and only if

(a) **Independent increments** – For all $t_1 < t_2 < \cdots < t_{2n}$,

$$X_{t_2} - X_{t_1}, X_{t_4} - X_{t_3}, \ldots, X_{t_{2n}} - X_{t_{2n-1}} \text{ are independent.}$$

(b) **Poisson increments** – For all $s < t$, we have $X_t - X_s = \text{Poisson}(\mu(t-s))$.

Theorem 9.7 – Let (X_t) be a Poisson process with intensity μ. Then, times between consecutive jumps are independent exponential random variables with parameter μ.

Proof. Since according to (a) and (b) the Poisson process has independent stationary increments, the inter-arrival times are independent and identically distributed. Moreover, the distribution function of the time of the first event, say T_1, is given by

$$F_{T_1}(t) = P(T_1 \leq t) = 1 - P(T_1 < t) = 1 - P(X_t = 0) = 1 - e^{-\mu t}$$

which implies that $T_1 = \text{Exponential}(\mu)$. Another approach that also emphasizes the connection between discrete and continuous time is as follows: let $Z_{i,N}$ be the number of events that occur between time $(i-1)(t/N)$ and time $i\,(t/N)$. Then,

$$\begin{aligned} P(T_1 > t) &= \lim_{N\to\infty} P(Z_{i,N} = 0 \text{ for all } i = 1, 2, \ldots, N) \\ &= \lim_{N\to\infty} (1 - \mu t/N + o(t/N))^N = e^{-\mu t} \end{aligned}$$

which as previously gives the result. □

Motivated by the previous theorem, we now state important properties of the exponential random variables that also help to understand Poisson processes.

Theorem 9.8 – We have $T = \text{Exponential}(\mu)$ for some $\mu > 0$ if and only if

$$T \text{ is memoryless}: \quad P(T > t+s \,|\, T > t) = P(T > s) \text{ for all } s, t > 0. \tag{9.1}$$

Proof. Assume that $T = \text{Exponential}(\mu)$. Then,

$$\begin{aligned} P(T > t+s \,|\, T > t) &= P(T > t+s)/P(T > t) \\ &= e^{-(t+s)}/e^{-t} = e^{-s} = P(T > s). \end{aligned}$$

Now, assume (9.1) and let $g(t) := P(T > t)$ for all $t \geq 0$. Then,

$$g(t+s) = g(t)\, g(s) \quad \text{for all } s, t > 0.$$

There exists a unique $\mu > 0$ such that $g(1) = e^{-\mu}$ and we want to prove that $g(t) = e^{-\mu t}$ on the set of integers, then rational numbers, then real numbers.

Proof for the integers – Let $n \in \mathbb{N}^*$. By induction, we get

$$g(n) = g(n-1)\, g(1) = \cdots = (g(1))^n = e^{-\mu n}.$$

Proof for the rational numbers – Let $r = p/q \in \mathbb{Q}_+$. Since $p = qr \in \mathbb{N}^*$, we have

$$g(p) = e^{-\mu p} = g(qr) = (g(r))^q \quad \text{therefore} \quad g(r) = (e^{-\mu p})^{1/q} = e^{-\mu r}.$$

Proof for the real numbers – Let $t \in \mathbb{R}_+$. Then, there exists an increasing sequence of positive rational numbers (t_n) that converges to t. Since g is left-continuous,

$$g(t) = g(\lim_{n\to\infty} t_n) = \lim_{n\to\infty} g(t_n) = \lim_{n\to\infty} e^{-\mu t_n} = e^{-\mu t}.$$

It follows that $F_T(t) = 1 - g(t) = 1 - e^{-\mu t}$ therefore $T = \text{Exponential}(\mu)$. □

Assume that $T_1, T_2, \ldots, T_n$ are independent and $T_i = \text{Exponential}(\mu_i)$.

Lemma 9.9 – Let $T := \min(T_1, T_2, \dots, T_n)$. Then,

$$T \;=\; \text{Exponential}\,(\mu_1 + \cdots + \mu_n) \quad \text{and} \quad P(T_i = T) \;=\; \mu_i\,(\mu_1 + \cdots + \mu_n)^{-1}.$$

Proof. Recalling the connection between exponential random variables and Poisson processes given by Theorem 9.7, this is simply a reformulation of Lemma 9.3 looking at the time between consecutive events instead of the number of events that occur in a time interval. □

10. Continuous-time Markov chains

As in discrete time, Markov chains in continuous time are processes in which the future depends on the past only through the present. More precisely, $(X_t : t \in \mathbb{R}_+)$ is a continuous-time Markov chain with finite or countable state space S whenever

$$P(X_{t+s} \in B \,|\, \mathcal{F}_s) \;=\; P(X_{t+s} \in B \,|\, X_s) \quad \text{for all} \quad B \in \mathcal{S}$$

where $\mathcal{F}_s := \sigma(X_r : 0 \le r \le s)$ is the σ-algebra generated by the past.

Transition rates. Markov chains are more difficult to conceptualize in continuous than discrete time because there is not successor (no next time) on the real line, and therefore the evolution cannot be specified by one-step transition probabilities. Instead, we define the **transition probabilities** more generally for all time intervals:

$$p_t(x,y) \;:=\; P(X_t = y \,|\, X_0 = x) \quad \text{for all } x, y \in S \text{ and all } t \ge 0.$$

In discrete time, the Chapman-Kolmogorov equations (5.2) show that the transition probabilities can be expressed in term of the one-step transition probabilities. In continuous time, the transition probabilities satisfy a differential equation that involves the so-called **transition rates**

$$c(x,y) \;:=\; \lim_{h\to 0} h^{-1}\, p_h(x,y) \quad \text{for all } x \neq y \tag{10.1}$$

provided these quantities are finite, which we assume from now on. The analog of the transition matrix in continuous time is the so-called **infinitesimal matrix** defined by

$$Q \;:=\; (q(x,y))_{x,y\in S} \quad \text{where} \quad q(x,y) \;:=\; \begin{cases} c(x,y) & \text{if} \quad x \neq y \\ -\sum_{z\neq z} c(x,z) & \text{if} \quad x = y. \end{cases} \tag{10.2}$$

Connection with discrete-time Markov chains. To begin with, we prove that continuous-time Markov chains can be seen as the combination of a discrete-time Markov chain that defines a skeleton – the sequence of states visited by the process – and a collection of independent Poisson processes that defines the temporal structure – the times at which the process jumps.

Theorem 10.1 – Let $T_x := \inf\{t \ge 0 : X_t \neq X_0 = x\}$. Then,

$$T_x \;=\; \text{Exponential}\,(-q(x,x)) \quad \text{and} \quad P(X_{T_x} = y) \;=\; -\,q(x,y)/q(x,x). \tag{10.3}$$

Proof. For $h > 0$ and $n \in \mathbb{N}$, let $Z_n := X_{nh}$. Note that (Z_n) is a discrete-time Markov chain, which we think of as what we would know about the continuous-time process if we were only able to see it at times multiple of h. Let τ_h be the time of the first jump

$$\tau_h \;:=\; \inf\{n \ge 0 : Z_n \neq Z_0 = x\}.$$

When h is very small, in which case we can see the continuous-time process very often, the probability of two jumps in an interval of length h before time T_x is very small so with high probability the random variables T_x and $h\tau_h$ differ by at most h. In the limit, we have

$$\lim_{h\to 0} h\,\tau_h \;=\; T_x \quad \text{with probability 1.}$$

Also, by the Markov property, we have

$$P_x(\tau_h = n \text{ and } Z_{\tau_h} = y) \;=\; p_h(x,x)^{n-1}\, p_h(x,y). \tag{10.4}$$

Both assertions follow from this equation by summing over all states or all times, and by observing that, by definition of the transition rates (10.1) and the infinitesimal matrix (10.2),

$$p_h(x,x) \;=\; 1 - \textstyle\sum_{z\neq x} p_h(x,z) \;\sim\; 1 + q(x,x)\,h \quad \text{and} \quad p_h(x,y) \;\sim\; q(x,y)\,h \;\text{ for } x \neq y$$

when time $h > 0$ is small.

Proof of the first assertion – Sum (10.4) over all states $z \neq x$ to obtain

$$P_x(\tau_h = n) \;=\; p_h(x,x)^{n-1}\,(1 - p_h(x,x))$$

then take the limit as $h \to 0$ to get

$$\begin{aligned} P_x(T_x > u) \;&=\; \lim_{h\to 0} P_x(\tau_h > u/h) \;=\; \lim_{h\to 0} \textstyle\sum_{n\ge u/h} p_h(x,x)^{n-1}\,(1 - p_h(x,x)) \\ &=\; \lim_{h\to 0} p_h(x,x)^{u/h-1} \textstyle\sum_{n\ge 0} p_h(x,x)^n\,(1 - p_h(x,x)) \\ &=\; \lim_{h\to 0} p_h(x,x)^{u/h-1} \;=\; \lim_{h\to 0} (1 + q(x,x)h)^{u/h-1} \;=\; \exp(q(x,x)u) \end{aligned}$$

therefore $T_x = \text{Exponential}\,(-q(x,x))$.

Proof of the second assertion – Sum (10.4) over all times $n \ge 1$ to obtain

$$P_x(Z_{\tau_h} = y) \;=\; \textstyle\sum_{n\ge 1} p_h(x,x)^{n-1}\, p_h(x,y) \;=\; p_h(x,y)(1 - p_h(x,x))^{-1}$$

then take the limit as $h \to 0$ to get

$$P_x(X_{T_x} = y) \;=\; \lim_{h\to 0} P_x(Z_{\tau_h} = y) \;=\; -\,q(x,y)/q(x,x)$$

which is the second assertion in (10.3). □

Theorem 10.1 shows the following: given that the process is in state x, the time to the next jump is exponentially distributed with parameter $-q(x,x)$ and the next state visited is state y with probability $-q(x,y)/q(x,x)$ so that the sequence of states visited is described by the discrete-time Markov with transition probabilities

$$p(x,y) := -\,q(x,y)/q(x,x) \;\text{ if } q(x,x) \neq 0 \;\text{ and }\; p(x,y) := 0 \;\text{ otherwise.}$$

This can be understood from the properties of the exponential distribution and Poisson processes as follows. Let $T(x,y)$ be independent exponential random variables with rate $q(x,y)$. Then, given that the process is in state x at time $T(x,y)$, it jumps to state y. In particular, for the process starting from state $X_0 = x$, the time to the first jump is

$$T_x \;=\; \min_{z\neq x} T(x,z) \;=\; \text{Exponential}\,(\textstyle\sum_{z\neq x} q(x,z)) \;=\; \text{Exponential}\,(-q(x,x))$$

according to Lemma 9.9, while the probability that the next state visited is y is given by

$$\begin{aligned} P(X_{T_x} = y) \;&=\; P(T(x,y) = T_x) \\ &=\; P(T(x,y) = \min_{z\neq x} T(x,z)) \;=\; -\,q(x,y)/q(x,x) \end{aligned}$$

according again to Lemma 9.9.

Stationary distribution and limiting behavior. The next two theorems, called Kolmogorov's forward and backward equations, derive differential equations for the transition probabilities that involves the transition rates through the infinitesimal matrix from which the transition probabilities can be computed. These are key results to determine the stationary distributions and understand the limiting behavior of the process.

Theorem 10.2 (Kolmogorov's forward equation)

$$p_t' \;=\; p_t \cdot Q \quad \text{where } p_t \text{ is the matrix with coefficients } p_t(x,y). \tag{10.5}$$

Proof. Decomposing according to the possible values of X_t we have

$$\begin{aligned} p_{t+h}(x,y) &= \textstyle\sum_{z\in S} p_t(x,z)\, p_h(z,y) \\ &= \textstyle\sum_{z\neq y} p_t(x,z)\, p_h(z,y) + p_t(x,y)\, p_h(y,y) \\ &= \textstyle\sum_{z\neq y} p_t(x,z)\, q(z,y)\, h + p_t(x,y)\,(1+q(y,y)\,h) + o(h) \\ &= \textstyle\sum_{z\in S} p_t(x,z)\, q(z,y)\, h + p_t(x,y) + o(h) \\ &= (p_t\cdot Q)(x,y)\, h + p_t(x,y) + o(h) \end{aligned}$$

from which it follows that

$$p_t'(x,y) \;=\; \lim_{h\to 0} h^{-1}\,(p_{t+h}(x,y) - p_t(x,y)) \;=\; (p_t\cdot Q)(x,y)$$

for all $x, y \in S$. This proves (10.5). □

Theorem 10.3 (Kolmogorov's backward equation)

$$p_t' \;=\; Q\cdot p_t \quad \text{where } p_t \text{ is the matrix with coefficients } p_t(x,y). \tag{10.6}$$

Proof. This is the same as the proof of Kolmogorov's forward equation but conditioning on the possible values of X_h instead of X_t. □

Note that solving Kolmogorov's backward equation (10.6) gives

$$p_t \;=\; e^{Qt} \;:=\; \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} \;=\; \sum_{n=0}^{\infty} Q^n \cdot \frac{t^n}{n!}.$$

Definition 10.4 (stationary distribution) – The vector $(\pi_x : x \in S)$ is said to be a stationary distribution of the Markov process (X_t) whenever

$$P(X_0 = x) = \pi_x \text{ for all } x\in S \;\;\Longrightarrow\;\; P(X_t = x) = \pi_x \text{ for all } x\in S \text{ and } t\in\mathbb{R}_+.$$

Note that this is equivalent to $\pi p_t = \pi$ for all $t > 0$ since πp_t is the distribution at time t of the process starting from the distribution π.

To compute the stationary distribution in practice, we use the following theorem.

Theorem 10.5 – π is a stationary distribution if and only if $\pi\cdot Q = 0$.

Proof. Assume that $\pi p_t = \pi$ for all $t > 0$. Then, according to (10.5), we have

$$\pi\cdot Q \;=\; \pi p_t\cdot Q \;=\; \pi p_t' \;=\; (\pi p_t)' \;=\; \pi' \;=\; 0.$$

Conversely, if $\pi\cdot Q = 0$ then (10.6) implies that

$$(\pi p_t)' \;=\; \pi p_t' \;=\; \pi\cdot Q\cdot p_t \;=\; 0 \quad\text{so}\quad \pi p_t \;=\; \pi p_0 \;=\; \pi\, I_S \;=\; \pi$$

for all times $t > 0$. This completes the proof. □

Theorem 10.6 – If the Markov chain (X_t) is finite and irreducible then there is a unique stationary distribution π. Moreover, regardless of the initial distribution,

$$\lim_{t\to\infty} P(X_t = x) \;=\; \pi_x \quad \text{for all } x\in S.$$

Proof. This is essentially the same as the proof of Theorem 5.12. □

11. Simple birth and death processes

Birth and death processes denote a class of continuous-time Markov chains with state space the set of or a subset of all nonnegative integers and in which the transition rate

$$c(x,y) \;\neq\; 0 \;\text{ only if }\; |x-y| = 1.$$

By convention, we write

$$c(x,x+1) \;:=\; \beta_x \;\text{ for } x\geq 0 \quad\text{and}\quad c(x,x-1) \;:=\; \delta_x \;\text{ for } x\geq 1$$

which fully characterizes the process. From a biological point of view, we think of the state as the number of individuals in a population, the parameters (β_x) as birth parameters and (δ_x) as death parameters. The simplest example, still biologically relevant, of birth and death process is the Markov chain in which individuals independently give birth at rate β to one individual and independently die at rate one. In this simple case, the transition rates become

$$\beta_x \;=\; \beta x \;\text{ for } x\geq 0 \quad\text{and}\quad \delta_x \;=\; x \;\text{ for } x\geq 1.$$

The process can be constructed from a countable collection of independent Poisson processes with parameter β and another countable collection of independent Poisson processes with parameter one, but Lemma 9.3 implies that it can also be constructed and simulated numerically as follows: given that the number of individuals is x, the next event occurs after an exponential amount of time with rate $(\beta+1)\,x$, and is a birth with probability $\beta\,(\beta+1)^{-1}$ and a death otherwise. The following is the algorithm written in C with $\beta := 2$. The program displays the time of the updates and the corresponding population size until time 8 unless the population dies out earlier.

```
/* Birth and Death Process */

    #include <stdlib.h>
    #include <stdio.h>
    #include <time.h>
    #include <math.h>
    #define T 8
    #define beta 2
```

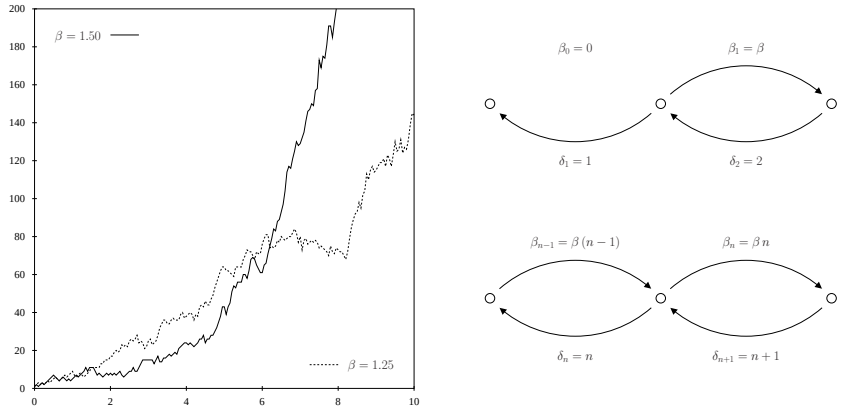


FIGURE 9. Realizations of the simple birth and death process with birth parameter $\beta = 1.50$ and $\beta = 1.25$, respectively, and schematic representation of the evolution rules of the process.

```
    int size ;
    float s, t, u, mu ;

// Take a number uniformly at random in (0, 1)
    float uniform () {
      u = (float) rand() / RAND_MAX ;
      return u ;
    }
// Exponential random variable with parameter mu
    float exponential (float u, float mu) {
      s = (float) - log (u) / mu ;
      return s ;
    }
// Birth and death process
    int main (int argc, char * argv[]) {
      time_t amorce ;
      time (&amorce) ;
      srand (amorce) ;
      t = 0 ;
// Start with one individual
      size = 1 ;
      while (t < T && size != 0) {
// Time of the next update
        u = uniform () ;
        t = t + exponential (u, (beta + 1) * size) ;
        u = uniform () ;
        u = u * (beta + 1) ;
// Death with probability 1 / (beta + 1)
        if (u < 1) {
          size -- ;
        }
// Birth with probability beta / (beta + 1)
        else {
          size ++ ;
        }
// Display time and population size at each update
        printf ("time = %f - population size = %i \n", t, size) ;
      }
    }
```


Probability of survival. The survival probability of such a population can be easily computed by comparing the process with processes studied previously.

Theorem 11.1 – Starting with $X_0 = x$ individuals, we have

$$\begin{aligned} P(X_t \neq 0 \text{ for all } t) &= 1 - \beta^{-x} \quad \text{when } \beta > 1 \\ &= 0 \qquad\quad \text{when } \beta \leq 1. \end{aligned}$$

PROOF. First, when $\beta = 1$, the embedded Markov chain associated with the simple birth and death process is the one-dimensional symmetric random walk absorbed at 0. In particular, almost sure extinction of the population follows from the fact that the unrestricted symmetric random walk is recurrent. When $\beta \neq 1$, the process is related to the gambler's ruin model. Introduce

$$T_N := \inf\{t : X_t = N\} \quad \text{for } N > x$$

and observe that the probability that the population size reaches N individuals is given by the winning probability (7.2) in the gambler's ruin model:

$$p_N := P(X_t = N \text{ for some } t) = (1 - a^x)(1 - a^N)^{-1} = (1 - \beta^{-x})(1 - \beta^{-N})^{-1}$$

according to Theorem 7.2, from which it follows that the survival probability is

$$\begin{aligned} P(X_t \neq 0 \text{ for all } t) = \lim_{N\to\infty} p_N &= 1 - \beta^{-x} \quad \text{when } \beta > 1 \\ &= 0 \qquad\quad \text{when } \beta < 1. \end{aligned}$$

This completes the proof. □

12. Logistic growth process

The logistic growth process can be seen as a spatially implicit version of the simple birth and death process. As in the latter, each individual gives birth at rate β to one individual and independently dies at rate one. However, there is now a set of N sites, each of which can be either empty or occupied by one individual, which reduces the state space to $N + 1$ states. At birth, the offspring is sent to a site chosen uniformly at random. If the site is empty it becomes occupied, otherwise the birth is suppressed, which models competition for space. Letting X_t denote the number of individuals at time t we obtain a birth and death process with transition rates

$$c(x, x+1) := \beta_x = \beta x(1 - x/N) \quad \text{and} \quad c(x, x-1) := \delta_x = x \tag{12.1}$$

for all $x = 0, 1, \ldots, N$. Numerical simulations of the process can be obtained by using these transition rates or alternatively going back to the interpretation of the process using an approach common in the field of interacting particle systems: after an exponential amount of time with parameter $(\beta + 1)N$, we choose a site uniformly at random, say i. If site i is occupied then

- with probability $\beta(\beta + 1)^{-1}$, we choose another site uniformly at random and place an individual at this site if it is empty,
- with probability $(\beta + 1)^{-1}$, we kill the individual at site i.

On the other hand, if site i is empty, nothing happens. As for the simple birth and death process, we give the algorithm written in C. The birth parameter is given by $\beta := 2$ and the number of spatial locations $N := 200$. The program displays the times of potential update and the corresponding population size until time 20 unless the population dies out earlier.

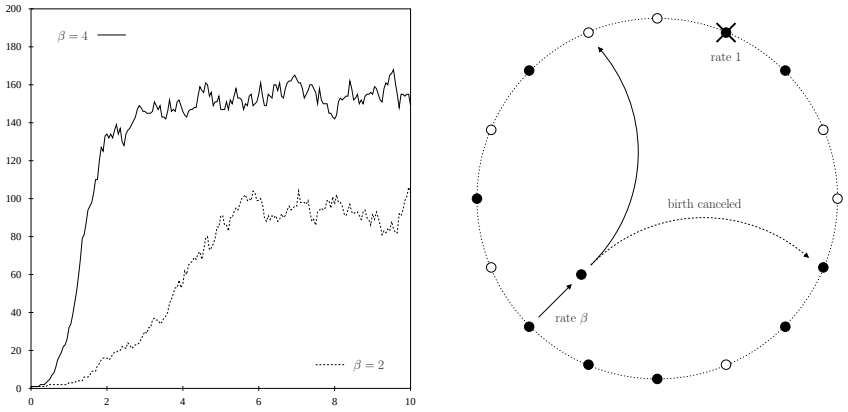


FIGURE 10. Realizations of the logistic growth process with $N = 200$ sites and birth parameter $\beta = 2$ and $\beta = 4$, respectively, and schematic illustration of the evolution rules of the process when $N = 16$.

```
/* Logistic Growth Process */

    #include <stdlib.h>
    #include <stdio.h>
    #include <time.h>
    #include <math.h>
    #define T 20
    #define beta 2
    #define N 200

    int X [N] ;
    int i, j, size ;
    float s, t, u, mu ;

// Take a number uniformly at random in (0, 1)
    float uniform () {
      u = (float) rand() / RAND_MAX ;
      return u ;
    }
// Exponential random variable with parameter mu
    float exponential (float u, float mu) {
      s = (float) - log (u) / mu ;
      return s ;
    }
// Logistic growth process
    int main (int argc, char * argv[]) {
      time_t amorce ;
      time (&amorce) ;
      srand (amorce) ;
      t = 0 ;
// Start with one individual
      size = 1 ;
      X [0] = 1 ;
      for (i = 1 ; i < N ; i ++) {
        X [i] = 0 ;
      }
      while (t < T && size != 0) {
// Time of the next potential update
        u = uniform () ;
        t = t + exponential (u, (beta + 1) * N) ;
// Choose a site i unifromly at random
```



```
        u = uniform () ;
        i = floor (N * u) ;
        i = i % N ;
// If site i is occupied
        if (X [i] == 1) {
          u = uniform () ;
          u = u * (beta + 1) ;
// The particle at i dies with probability 1 / (beta + 1)
          if (u < 1) {
            X [i] = 0 ;
            size -- ;
          }
// The particle at x gives birth with probability beta / (beta + 1)
          else {
            u = uniform () ;
            j = floor (N * u) ;
            j = j % N ;
            if (X [j] == 0) {
              X [j] = 1 ;
              size ++ ;
            }
          }
        }
// Display time and population size
        printf ("time = %f - population size = %i \n", t, size) ;
      }
    }
```

Long-term behavior. Recall that the popular (deterministic) logistic growth model is described by the ordinary differential equation

$$\dot{x} \;=\; rx\,(1 - x/K) \tag{12.2}$$

where r is the intrinsic rate of growth and K the carrying capacity. Starting with a positive density of individuals, the solution to this differential equation converges to zero when the intrinsic rate of growth $r < 0$, and to the carrying capacity K when $r > 0$. The logistic growth process has some similarities with its deterministic counterpart. First, letting

$$r \;:=\; \beta - 1 \qquad \text{and} \qquad K \;:=\; N\,(1 - \beta^{-1}),$$

we obtain that the difference between birth and death rates is given by

$$\begin{aligned} \beta_x - \delta_x &= \beta x\,(1 - x/N) - x \\ &= rx\,(\beta - 1)^{-1}\,(\beta\,(1 - x/N) - 1) \;=\; rx\,(1 - x/K) \end{aligned}$$

which also explains the *a priori* mysterious expression of the differential equation (12.2), and the reason why it indeed takes space into account. In contrast with the deterministic model, since there are only two communication classes in model (12.1), namely

$$C_0 := \{0\} \quad \text{and} \quad C_+ := \{1, 2, \ldots, N\}$$

with state 0 absorbing and all states in C_+ transient, we deduce that the stochastic process is almost surely driven to extinction. However, when $r > 0$, the expected time to extinction increases exponentially with the size of the system and the transient behavior before extinction, which is characterized by a so-called quasi-stationary distribution, is somewhat reminiscent of the limiting behavior of the ordinary differential equation (12.2).

Quasi-stationary distribution. The quasi-stationary distribution of the logistic growth process is the stationary distribution of the process conditioned on non-extinction, which is not equal but close to the stationary distribution of the new process, say (Z_t), obtained by removing state zero from the state space, so all states become recurrent.

Theorem 12.1 – The stationary distribution π of the process (Z_t) satisfies

$$\pi_x \;=\; \frac{\pi_1}{x}\left(\frac{\beta}{N}\right)^{x-1}\frac{(N-1)\,!}{(N-x)\,!} \quad \text{for } x = 1, 2, \ldots, N. \tag{12.3}$$

Proof. The infinitesimal matrix of (Z_t) is

$$Q \;=\; \begin{pmatrix} -\beta_1 & \beta_1 & 0 & \cdots \\ \delta_2 & -(\beta_2+\delta_2) & \beta_2 & \cdots \\ 0 & \delta_3 & -(\beta_3+\delta_3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

By Theorem 10.5 and a simple induction, the stationary distribution thus satisfies

$$\pi_x \;=\; \frac{\beta_1\cdots\beta_{x-1}}{\delta_2\cdots\delta_x}\,\pi_1 \;=\; \frac{\pi_1}{\delta_x}\,\frac{\beta_1\cdots\beta_{x-1}}{\delta_1\cdots\delta_{x-1}} \;=\; \frac{\pi_1}{\delta_x}\prod_{y=1}^{x-1}\frac{\beta_y}{\delta_y}. \tag{12.4}$$

Plugging (12.1) into (12.4), we obtain

$$\begin{aligned}\pi_x \;&=\; \frac{\pi_1}{x}\prod_{y=1}^{x-1}\frac{\beta_y}{\delta_y} \;=\; \frac{\pi_1}{x}\prod_{y=1}^{x-1}\beta\left(1-\frac{y}{N}\right)\\ &=\; \frac{\pi_1}{x}\prod_{y=1}^{x-1}\frac{\beta}{N}\,(N-y) \;=\; \frac{\pi_1}{x}\left(\frac{\beta}{N}\right)^{x-1}\frac{(N-1)\,!}{(N-x)\,!}\end{aligned}$$

This completes the proof. □

We also note that

$$\pi_{x+1}/\pi_x \;=\; \beta_x/\delta_{x+1} \;=\; x\,(x+1)^{-1}\,(\beta_x/\delta_x) \;\approx\; 1 \quad \text{when} \quad \beta_x \approx \delta_x$$

which indicates that the most visited state lies near the carrying capacity K. In conclusion, before extinction, the logistic growth process oscillates randomly around the carrying capacity. We now prove that this transient behavior lasts for a time whose expected value increases exponentially with the size N of the system.

Time to extinction. The time to extinction is defined formally as

$$T \;:=\; \inf\{t > 0 : X_t = 0\}$$

which is a random variable whose distribution depends on the initial state. To compute the expected time to extinction, we also introduce

$$\tau_x \;:=\; E\,(T \,|\, X_0 = x) \quad \text{and} \quad \sigma_x \;:=\; \tau_x - \tau_{x+1}.$$

Theorem 12.2 – For all $x = 0, 1, 2, \ldots, N-1$, we have

$$\sigma_x \;=\; \prod_{z=1}^{x}\frac{\delta_z}{\beta_z}\left[\sigma_0 + \sum_{y=1}^{x}\frac{1}{\beta_y}\prod_{z=1}^{y}\frac{\beta_z}{\delta_z}\right].$$

Proof. A first-step analysis gives

$$\tau_x \;=\; (\beta_x+\delta_x)^{-1} + \;\beta_x\,(\beta_x+\delta_x)^{-1}\,\tau_{x+1} \;+\; \delta_x\,(\beta_x+\delta_x)^{-1}\,\tau_{x-1}$$

for all $x = 1, 2, \ldots, N$. In particular,

$$\begin{aligned}(\beta_x+\delta_x)\,\tau_x \;&=\; \beta_x\,\tau_{x+1} \;+\; \delta_x\,\tau_{x-1} \;+\; 1\\ \beta_x\,(\tau_x - \tau_{x+1}) \;&=\; \delta_x\,(\tau_{x-1} - \tau_x) \;+\; 1\end{aligned}$$

therefore, recalling that $\sigma_x := \tau_x - \tau_{x+1}$, we get

$$\begin{aligned}\sigma_x \;&=\; \frac{1}{\beta_x} + \frac{\delta_x}{\beta_x}\,\sigma_{x-1} \;=\; \frac{1}{\beta_x} + \frac{\delta_x}{\beta_x}\,\frac{1}{\beta_{x-1}} + \frac{\delta_x}{\beta_x}\,\frac{\delta_{x-1}}{\beta_{x-1}}\,\sigma_{x-2} \;=\; \ldots\\ &=\; \sum_{y=1}^{x}\frac{1}{\beta_y}\prod_{z=y+1}^{x}\frac{\delta_z}{\beta_z} \;+\; \sigma_0\prod_{z=1}^{x}\frac{\delta_z}{\beta_z} \;=\; \prod_{z=1}^{x}\frac{\delta_z}{\beta_z}\left[\sigma_0 + \sum_{y=1}^{x}\frac{1}{\beta_y}\prod_{z=1}^{y}\frac{\beta_z}{\delta_z}\right].\end{aligned}$$

This completes the proof. □

Since $\sigma_x \le 0$ and $\sigma_0 = -\tau_1$ the previous theorem implies that

$$\begin{aligned}\tau_1 \;&\ge\; \sum_{y=1}^{x}\frac{1}{\beta_y}\prod_{z=1}^{y}\frac{\beta_z}{\delta_z} \;\ge\; \frac{1}{\beta_{K/2}}\prod_{z=1}^{K/2}\frac{\beta_z}{\delta_z} \;\ge\; \frac{1}{\beta_{K/2}}\left[\beta\left(1-\frac{K}{2N}\right)\right]^{K/2}\\ &\ge\; \frac{1}{\beta_{K/2}}\left[\beta\left(1-\frac{1-\beta^{-1}}{2}\right)\right]^{K/2} \;=\; \frac{1}{\beta_{K/2}}\left(\frac{\beta+1}{2}\right)^{K/2}\end{aligned}$$

so the expected time to extinction starting with one individual grows exponentially fast with the size of the system when $\beta > 1$. In fact, using that the process behaves almost like the simple birth and death process when the fraction of occupied sites is small, we obtain a bimodal behavior of the logistic growth process with N large and $\beta > 1$. More precisely, by Theorem 11.1,

- With probability β^{-1}, which is the probability of extinction of the simple birth and death process, the population goes extinct quickly.
- With probability $1 - \beta^{-1}$, which is the probability of survival of the simple birth and death process, the population first grows and then evolves as described by the quasi-stationary distribution (12.3) for a very long time.

13. Wright-Fisher and Moran models

The Wright-Fisher model is similar to the logistic growth process in the sense that it includes implicit space in the form of global interactions. However, whereas in the logistic growth process spatial locations are either empty or occupied, in the Wright-Fisher model all spatial locations are occupied by one individual of two possible types, which results in a competition model rather than

an invasion model. The Wright-Fisher model was originally introduced to understand the evolution of the gene frequency in a population of haploid individuals with constant size, but can be seen as the basic spatially implicit competition model, in the same way as the logistic growth process can be seen as the basic spatially implicit invasion model. In this section, we focus on the neutral Wright-Fisher model and refer to Neuhauser [21] for an overview of the model in the presence of selection and mutations, as well as a derivation of the so-called ancestral selection graph, which is the analog of Kingman's coalescent in the presence of selection.

The binomial distribution. In the Wright-Fisher model, each site of a set of N sites is occupied by an individual of type 0 or type 1. The parent of each individual at each generation (non overlapping generations) is chosen uniformly at random from the population at the previous generation, and each individual has the same type as its parent. In particular, letting X_n denote the number of type 1 individuals at generation n induces a discrete-time Markov chain. Since each site at generation $n+1$ is independently of type 1 with probability the fraction of type 1 at generation n, we obtain transition probabilities that are binomial:

$$p(x,y) \;:=\; P\left(X_{n+1} = y \,|\, X_n = x\right) \;=\; \binom{N}{y} p^y \,(1-p)^{N-y} \quad \text{where} \;\; p := x/N. \tag{13.1}$$

To exhibit later the connection between the Wright-Fisher model and the voter model, it is useful to also introduce the Moran model that describes similar dynamics but in continuous time: each individual updates its type at rate one by choosing a parent uniformly at random from the population. In particular, the number of type 1 individuals at time t in the Moran model is the birth and death process with birth and death rates

$$c(x,x+1) \;:=\; \beta_x \;=\; (N-x)(x/N) \quad \text{and} \quad c(x,x-1) \;:=\; \delta_x \;=\; x\,(1-x/N)$$

for all $x = 0, 1, \ldots, N$. As for the logistic growth process, the Wright-Fisher model can be simulated numerically by either using the expression of its transition probabilities (13.1) or going back to the interpretation of the process in the following manner: for each site, say i, choose a site uniformly at random, say j, and fix the type of the individual at site i at generation $n+1$ to be the same as the type of the individual at site j at the previous generation. Below is an algorithm following this approach written in C. The program gives the number of type 1 individuals until time 100 for the process with $N := 100$ spatial locations starting with 50% of individuals of type 1.

```
/* Wright-Fisher Model */

    #include <stdlib.h>
    #include <stdio.h>
    #include <time.h>
    #include <math.h>

    #define T 100
    #define N 100

    int X [T][N] ;
    int i, j, n, size ;
    float u ;

// Take a number uniformly at random in (0, 1)
    float uniform () {
      u = (float) rand() / RAND_MAX ;
      return u ;
    }
```



```
// Wright-Fisher model
    int main (int argc, char * argv[]) {

      time_t amorce ;
      time (&amorce) ;
      srand (amorce) ;

// Start with half of the individuals of type 1
      for (i = 0 ; i < N / 2 ; i ++) {
        X [0][i] = 0 ;
      }
      for (i = N / 2 ; i < N ; i ++) {
        X [0][i] = 1 ;
      }

      for (n = 0 ; n < T - 1 ; n ++) {

        size = 0 ;

        for (i = 0 ; i < N ; i ++) {

// Choose a parent of site i uniformly at random
          u = uniform () ;
          j = floor (N * u) ;
          j = j % N ;
          X [n + 1][i] = X [n][j] ;
          size = size + X [n + 1][i] ;
        }

// Display time and number of type 1 individuals
        printf ("time = %2i - number of type 1 = %3i \n", n + 1, size) ;
      }
    }
```

Probability of fixation. Note that the Wright-Fisher model has three communication classes: two of them are the absorbing states 0 (type 0 wins) and N (type 1 wins), while the third one is a transient class. In particular, the system converges almost surely to one of its two absorbing states and the first important question is: who wins?

Theorem 13.1 – Starting with $X_0 = x$ individuals of type 1,

$$P\left(\text{type 1 wins}\right) \;:=\; P\left(X_n = N \text{ for some } n\right) \;=\; x/N.$$

Proof. First, we observe that

$$E\left(X_{n+1} \,|\, X_n = x\right) \;=\; E\left(\text{Binomial}\left(N, x/N\right)\right) \;=\; x$$

from which it follows that (X_n) is a martingale. Let p_0 and p_N be respectively the probability of fixation to all type 0 and all type 1, and introduce the time to fixation

$$T \;:=\; \inf\{t : X_t \in \{0, N\}\}.$$

Since (X_n) is a bounded martingale and time T is an almost surely finite stopping time, the optimal stopping theorem 7.1 implies that

$$E\,X_T \;=\; E\,X_0 \;=\; N \times P\left(X_T = N\right) \;+\; 0 \times P\left(X_T = 0\right) \;=\; N\,p_N$$

therefore, starting with $X_0 = x$ individuals of type 1, we have $p_N = x/N$. $\square$

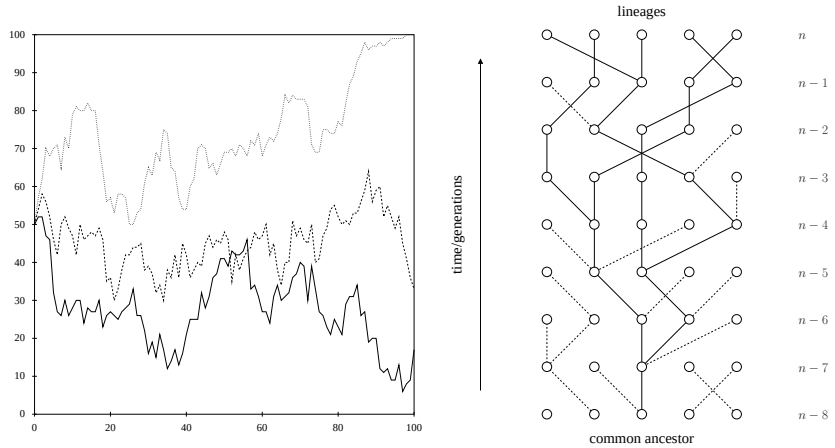


Figure 11. Three independent realizations of the Wright-Fisher model with population size $N = 100$ and schematic illustration of Ewen's argument.

Ewens [13] gave an intuition of this result: after a long enough time, all the individuals must have descended from just one of the individuals present at generation zero and the probability that this common ancestor was of type 1 is equal to the relative frequency x/N of type 1 at generation zero. The following is an alternative proof of Theorem 13.1 that follows this idea.

Proof. The idea is to represent the evolution graphically. Let

$$U(x,n) \;:=\; \text{Uniform}\,\{1,2,\ldots,N\} \;\text{ for all }\; (x,n) \in \{1,2,\ldots,N\}\times\mathbb{N}^*$$

be independent and draw an edge between (x,n) and $(U(x,n),n-1)$ to indicate that the individual at the first space-time point is the offspring of the individual at the second space-time point. Keeping track of the ancestors of the individuals at generation n going backwards in time using the previous collection of uniform random variables induces a system of coalescing random walks on the complete graph with N vertices. Provided n is large enough, there is only one random walk left, the common ancestor, at generation zero. Since the location of this random walk at generation zero is chosen uniformly at random from the entire population, the common ancestor, which determines the common type of the individuals at generation n, is of type 1 with probability the initial frequency of type 1. See Figure 11 for an illustration of this argument. □

Diffusion approximation and time to absorption. The second important question about the model is: how long does it take for the population to fixate? Good approximations of the expected value of the time to absorption for large populations can be found by looking at the so-called diffusion approximation. Diffusion processes are continuous-time Markov processes with continuous sample paths, and so uncountable state space. Recalling that N denotes the population size, the diffusion approximation of the Wright-Fisher model is defined by

$$Y_t \;=\; \lim_{N\to\infty} N^{-1}\, X_{\lfloor Nt\rfloor} \tag{13.2}$$

where $\lfloor Nt\rfloor$ = integer part of Nt. The state space is now the interval $[0,1]$, which represents the frequency, rather than the number, of type 1 individuals. It can be proved that diffusion processes are fully characterized by the following two quantities:

$$\begin{aligned}\mu(x) &= \text{drift parameter} := \lim_{h\to 0} h^{-1}\, E\,(Y_{t+h} - Y_t \,|\, Y_t = x)\\ \sigma^2(x) &= \text{diffusion parameter} := \lim_{h\to 0} h^{-1}\, E\,((Y_{t+h} - Y_t)^2 \,|\, Y_t = x).\end{aligned}$$

Moreover, letting $T := \inf\{t : Y_t \in \{0,1\}\}$ denote the time to absorption, the theory of diffusion processes gives the following differential equation

$$\mu(x)\, u'(x) \;+\; (1/2)\,\sigma^2(x)\, u''(x) \;=\; -1 \quad \text{where} \quad u(x) = E\,(T \,|\, Y_0 = x). \tag{13.3}$$

For the Wright-Fisher model, drift and diffusion parameters are given by

- Drift parameter: $\mu(x) = 0$ because (X_n) is a martingale.
- Diffusion parameter: $\sigma^2(x) = x(1-x)$ because

$$\begin{aligned}\sigma^2(x) &= \lim_{N\to\infty} N\; E\,((Y_{t+1/N} - Y_t)^2 \,|\, Y_t = x)\\ &= \lim_{N\to\infty} N\, E\,((N^{-1}X_{\lfloor Nt\rfloor+1} - N^{-1}X_{\lfloor Nt\rfloor})^2 \,|\, X_{\lfloor Nt\rfloor} = Nx)\\ &= \lim_{N\to\infty} N^{-1}\; E\,((X_{\lfloor Nt\rfloor+1} - Nx)^2 \,|\, X_{\lfloor Nt\rfloor} = Nx)\\ &= \lim_{N\to\infty} N^{-1}\,\text{Var}\,(\text{Binomial}\,(N,x)) \;=\; \lim_{N\to\infty} N^{-1} N\, x\,(1-x).\end{aligned}$$

In particular, according to (13.3), the function $u(x)$ satisfies

$$(1/2)\, x\,(1-x)\, u''(x) \;=\; -1 \quad \text{with} \quad u(0) = u(1) = 1$$

which has solution

$$u(x) \;=\; -2\,(x\ln(x) + (1-x)\ln(1-x)).$$

For instance, $u(0.5) \approx 1.386$ which indicates that, when the population size is large and half of the individuals are of type 1 at generation zero, the expected value of the time to fixation is approximately equal to 1.386 times the population size.

14. Kingman's coalescent

The third important question about the Wright-Fisher model is: given a sample of n individuals at the present time, how far do we need to go back in time to find their common ancestor? More precisely, let the process evolve long enough so that all the individuals have a common ancestor, take a random sample of n individuals, and denote by $T_{\text{MRCA}}(n)$ the amount of time divided by the population size we need to go backwards to find their most recent common ancestor, i.e., the amount of time we need to go backwards until the random walks in the proof of Theorem 13.1 coalesce. This gives rise to a new process called Kingman's coalescent, which allows to answer the question for the diffusion approximation (13.2). This process can be represented graphically as a random tree with n branches in which branches coalesce by pair at some random times.

Theorem 14.1 – In the diffusion approximation, $E\,T_{\text{MRCA}}(n) = 2\,(1 - n^{-1})$.

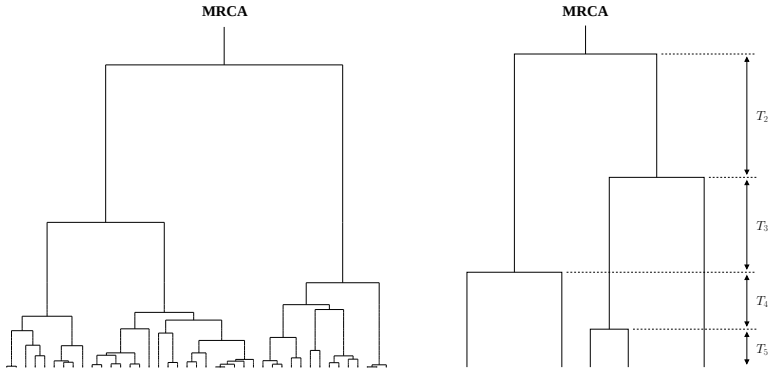


Figure 12. Realization of Kingman's coalescent and picture of the sequence T_j.

Proof. For $j = 2, 3, \ldots, n$, we denote by T_j the time to the first coalescing event going backwards in time when starting with j lineages. Then,

$$T_{\mathrm{MRCA}}(n) \;=\; T_n \;+\; T_{n-1} \;+\; \cdots \;+\; T_2. \tag{14.1}$$

Since at each time step going backwards in time two given lineages intersect, i.e., jumps to the same site, with probability $1/N$, in the diffusion approximation, we have

$$\begin{aligned} P\,(T_2 > t) \;&=\; \lim_{N\to\infty} P\,(\text{no intersection for } \lfloor Nt \rfloor \text{ generations}) \\ &=\; \lim_{N\to\infty} (1 - 1/N)^{\lfloor Nt \rfloor} \;=\; \exp(-t) \end{aligned}$$

therefore $T_2 = \text{Exponential}\,(1)$. By independence, this also implies that each pair of lineages coalesce at rate one, from which it follows that

$$T_j \;=\; \text{Exponential}\,(\#\text{ remaining pairs}) \;=\; \text{Exponential}\,(j\,(j-1)/2). \tag{14.2}$$

Combining (14.1) and (14.2), we conclude that

$$E\,T_{\mathrm{MRCA}}(n) \;=\; \sum_{j=2}^{n} E\,T_j \;=\; \sum_{j=2}^{n} \frac{2}{j\,(j-1)} \;=\; 2\,\sum_{j=2}^{n} \left(\frac{1}{j-1} - \frac{1}{j}\right) \;=\; 2\left(1 - \frac{1}{n}\right).$$

This completes the proof. □

Note that Theorem 14.1 implies that

$$E\,T_{\mathrm{MRCA}}(2) \;=\; 1 \quad \text{and} \quad \lim_{n\to\infty} E\,T_{\mathrm{MRCA}}(n) \;=\; 2,$$

which indicates that the time to the most recent common ancestor for a very large sample is only about twice as large as that for a sample of size two.

15. Bond percolation in two dimensions

The first percolation model was motivated by the simple question: what is the probability that the center of a porous stone immersed in water is wet? Percolation models now arise in a wide

variety of applied sciences such as physics, biology and sociology, and percolation theory is one of the most important topics of probability theory which is also used to understand other models such as interacting particle systems. Problems related to percolation theory are typically easy to formulate but very challenging mathematically. In this section, we focus on bond percolation in two dimensions, a model which can be well understood using certain duality properties of the two-dimensional regular lattice. See Grimmett [14] for more details about percolation models.

Model description. For concreteness, we formulate the model in the context of epidemiology. Imagine that each site of the two-dimensional lattice is occupied by an individual, and that independently of each other any two nearest neighbors are friends with probability p. If an infection can only spread from friend to friend, what is the probability that an infection introduced at the origin will result in an epidemic? Bond percolation on the two-dimensional lattice is simply the process that consists of the countable collection of independent Bernoulli random variables with success probability p that keep track of the friendship relationships. More precisely, let

$$\xi(e) \;:=\; \text{Bernoulli}\,(p) \quad \text{for each} \quad e \in E := \{(x, y) \in \mathbb{Z}^2 \times \mathbb{Z}^2 : \|x - y\| = 1\}.$$

In the terminology of percolation theory, edge e is said to be

$$\textbf{open} \text{ when } \xi(e) = 1 \;\text{ and }\; \textbf{closed} \text{ when } \xi(e) = 0.$$

We say that there is an **open path** between two vertices x and y, which we write $x \leftrightarrow y$, whenever there exists a sequence of vertices $x_0 = x, x_1, \ldots, x_n = y$ such that

$$(x_i, x_{i+1}) \in E \;\text{ is open }\; \text{for all } i = 0, 1, \ldots, n-1. \tag{15.1}$$

Define the **open cluster** at vertex x to be

$$C_x \;:=\; \{y \in \mathbb{Z}^2 : x \leftrightarrow y\}.$$

Bond percolation induces a random subgraph of the lattice: the subgraph of the two-dimensional lattice induced by the set of open edges. The cluster C_0 is the connected component of this subgraph containing the origin. In particular, the event that an epidemic occurs, i.e., infinitely many individuals are ultimately infected, is the event that this cluster is infinite, thus the probability of an epidemic is given by the so-called **percolation probability**

$$\theta(p) \;:=\; P_p\,(|C_0| = \infty). \tag{15.2}$$

Monotonicity of the percolation probability. It should be intuitively clear that the percolation probability (15.2) is a nondecreasing function of the density p of open edges. This can be made rigorous using a so-called **coupling argument**. Coupling is a very powerful technique in probability theory in general and is not limited to percolation models. The basic idea is to find a joint construction of two random variables or stochastic processes, which is called a **coupling**, such that one dominates the other one with probability one. This allows to easily deduce a stochastic domination. In addition to establishing the monotonicity of the percolation probability, the following theorem gives an illustration of how coupling can be used in practice.

Theorem 15.1 – The function $p \mapsto \theta(p)$ is nondecreasing.

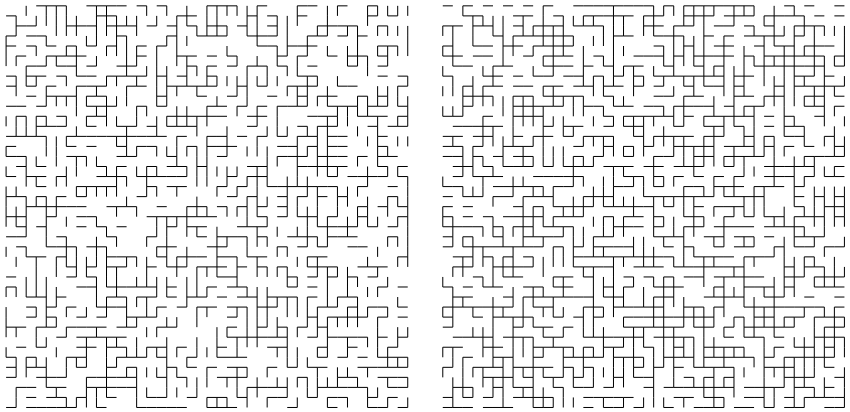


Figure 13. Bond percolation with parameter $p = 0.45$ and $p = 0.55$.

Proof. Having two parameters $p_1 \leq p_2$, the objective is to find a joint construction of bond percolation with parameter p_1 and bond percolation with parameter p_2 in such a way that the latter dominates the former with probability one. To do this, we let

$$U(e) \;:=\; \text{Uniform}\,(0,1) \quad \text{for each edge } e \in E \tag{15.3}$$

be independent and define two collections of Bernoulli random variables by letting

$$\xi_i(e) := 1 \;\text{ if and only if }\; U(e) \leq p_i \;\text{ for each edge } e \in E.$$

These random variables have the following three properties:

- The random variables $\xi_i(e)$, $e \in E$, are independent since they are constructed from (15.3) which are independent random variables.
- In addition, $P\,(\xi_i(e) = 1) = P\,(U(e) \leq p_i) = p_i$ therefore $\{\xi_i(e) : e \in E\}$ is indeed a bond percolation process with parameter p_i as desired.
- Being constructed from the same collection of uniform random variables (15.3), the two percolation processes are not independent and we have

$$\{\xi_1(e) = 1\} \;=\; \{U(e) \leq p_1\} \;\subset\; \{U(e) \leq p_2\} \;=\; \{\xi_2(e) = 1\} \quad \text{for each } e \in E.$$

  In particular, we have $\xi_1 \leq \xi_2$ with probability one.

In conclusion, we have the desired coupling. To complete the proof, we define

$$f(\xi) \;:=\; \mathbf{1}\,\{|C_0(\xi)| = \infty\} \quad \text{where} \quad \xi : E \to \{0,1\} \tag{15.4}$$

and where $C_0(\xi)$ is the cluster at the origin of the graph induced by ξ. Since we have $\xi_1 \leq \xi_2$ and since the function defined in (15.4) is clearly nondecreasing, we deduce that

$$\theta(p_1) \;:=\; P\,(|C_0(\xi_1)| = \infty) \;=\; E\,f(\xi_1) \;\leq\; E\,f(\xi_2) \;=\; \theta(p_2).$$

This completes the proof. □

The critical phenomenon. The fact that the percolation probability is nondecreasing with respect to p implies the existence of a **critical value**: above this critical value, there is a positive probability of an epidemic, while below this critical value the probability of an epidemic is equal to zero. More precisely, we introduce

$$p_c \;:=\; \sup\{p : \theta(p) = 0\}. \tag{15.5}$$

Interesting but also challenging results that describe two-dimensional bond percolation as well as other percolation models in the subcritical and supercritical cases state that below the critical value, there is no infinite cluster and the cluster size decays exponentially, whereas above the critical value there exists exactly one infinite cluster.

Theorem 15.2 – Assume that $p < p_c$. Then,

$$P\,(|C_0| > n) \;\leq\; \exp(-\alpha\, n) \quad \text{for some } \alpha > 0.$$

Theorem 15.3 – Assume that $p > p_c$. Then, the graph induced by the open edges has a unique infinite connected component called the infinite percolation cluster.

Critical values for spatially explicit models such as percolation processes and interacting particle systems are usually impossible to compute but this can be done for bond percolation in two dimensions relying on Theorem 15.3 and **planar duality**.

Definition 15.4 (planar graph) – A planar graph is a graph that can be drawn on the plane in such a way that its edges do not intersect each other. For instance, the complete graph with N vertices is a planar graph if and only if $N \leq 4$.

Definition 15.5 (planar duality) – A dual graph of a planar graph G is a graph that has

- a vertex corresponding to each plane region of G,
- an edge joining two neighboring regions for each edge in G.

Note that the graph with vertex set and edge set

$$\begin{aligned} \mathbb{D}^2 \;&:=\; \{x + (1/2, 1/2) : x \in \mathbb{Z}^2\} \\ E' \;&:=\; \{(x, y) \in \mathbb{D}^2 \times \mathbb{D}^2 : \|x - y\| = 1\} \end{aligned} \tag{15.6}$$

is a dual graph of the two-dimensional lattice. We call this graph the dual lattice. Since it is also isomorphic to the two-dimensional regular lattice, we say that the two-dimensional regular lattice is self-dual, which is one of the keys to computing the critical value of bond percolation. Since both graphs are dual of each other, each edge of the dual lattice intersects exactly one edge of the original lattice, and given a realization of bond percolation on $\mathbb{Z}^2$, we declare an edge of the dual lattice to be open if and only if the unique edge of the original lattice it intersects is open. Note that this simple construction has two important consequences:

- Each realization of bond percolation on the original lattice induces a realization of bond percolation with the same parameter on the dual lattice (15.6).
- Through this coupling of bond percolation processes, open paths of the original lattice and closed paths of the dual lattice (15.6) cannot intersect.

With these preliminary results in hands, we can now compute explicitly the critical value (15.5) of bond percolation in two dimensions, a result due to Kesten [18].

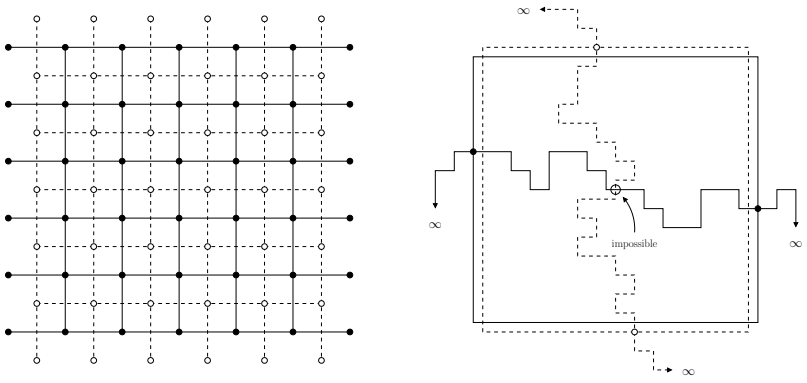


FIGURE 14. Pictures related to the proof of Theorem 15.6.

Theorem 15.6 – For bond percolation in two dimensions, $p_c = 1/2$.

PROOF. First, we prove that $p_c \le 1/2$. Referring to the left-hand side of Figure 14, we consider the following two subgraphs of the lattice and dual lattice:

- G_N is the $(N+1) \times N$ subgraph of the original lattice drawn in continuous lines.
- H_N is the $N \times (N+1)$ subgraph of the dual lattice drawn in dashed lines.

We also introduce the corresponding events

$$\begin{aligned} A_N &:= \text{there is an open path connecting the left and the right sides of the subgraph } G_N \text{ of the original lattice} \\ B_N &:= \text{there is a closed path connecting the top and the bottom sides of the subgraph } H_N \text{ of the dual lattice.} \end{aligned}$$

Through the coupling of the processes, the event A_N occurs if and only if B_N does not occur, which implies that both events form a partition of the sample space. Also, the probability of A_N when edges are open with probability $1-p$ is equal to the probability of B_N when edges are open with probability p. These two properties imply that

$$P_p(A_N) + P_p(B_N) = 1 \quad \text{and} \quad P_{1-p}(A_N) = P_p(B_N)$$

where the index refers to the percolation parameter. In particular,

$$\begin{aligned} p < p_c &\Rightarrow \lim_{N\to\infty} P_p(A_N) = 0 \\ &\Rightarrow \lim_{N\to\infty} P_{1-p}(A_N) = \lim_{N\to\infty} P_p(B_N) = 1 \\ &\Rightarrow 1-p \ge p_c \Rightarrow p \le 1-p_c \end{aligned}$$

which implies that $p_c \le 1 - p_c$ therefore $p_c \le 1/2$. To establish the converse, we prove a stronger result, namely the lack of percolation when the density of open edges is equal to one half. To prove

this result, we proceed by contradiction and consider the new events

$$\begin{aligned} A_N &:= \text{there are two infinite open paths outside the subgraph } G_N \text{ one starting from its left side and one from its right side} \\ B_N &:= \text{there are two infinite closed paths outside the subgraph } H_N \text{ one starting from its bottom side and one from its top side.} \end{aligned}$$

Assuming by contradiction that $\theta(1/2) > 0$, we have

$$\lim_{N\to\infty} P_{1/2}(A_N \cap B_N) = 1.$$

The uniqueness of the infinite percolation cluster stated in Theorem 15.3 then implies that the two paths in the event A_N are linked by an open path and the two paths in the event B_N are linked by a closed path. As shown in the right-hand side of Figure 14, this leads to the existence of an open path and a closed path intersecting each other, thus a contraction. In conclusion, percolation does not occur when the density of open edges equals one half therefore $p_c \ge 1/2$. □

Bond percolation can be defined in higher dimensions as well. In this case, planar duality is no longer available since the regular lattice is not a planar graph. However, it can be easily proved that the critical value is not degenerate, i.e., it belongs to $(0,1)$.

Theorem 15.7 – For bond percolation in d dimensions, $(2d-1)^{-1} \le p_c \le 1/2$.

PROOF. Bond percolation processes in different dimensions but with the same parameter can be coupled to show that the critical value (15.5) is nonincreasing with respect to the dimension. This and Theorem 15.6 give the upper bound. To deal with the lower bound, we observe that

$$\begin{aligned} P_p(|C_0| = \infty) &= \lim_{n\to\infty} P_p(C_0 \not\subset B(0,n)) \\ &\le \lim_{n\to\infty} P_p(\text{there exists an open path of length } n \text{ starting at } 0) \\ &\le \lim_{n\to\infty} 2d(2d-1)^{n-1} p^n = 0 \end{aligned}$$

for all $p < (2d-1)^{-1}$. This completes the proof. □

16. Oriented site percolation in two dimensions

In oriented site percolation, the vertices rather than the edges are open or closed. Moreover, edges are oriented, resulting in a directed graph, so the water, disease, rumor, etc, can only spread in one direction. A good reference on this topic is the review paper of Durrett [10].

Formulation of the model. Let $\mathcal{G}$ be the set of sites

$$\mathcal{G} := \{(x,n) \in \mathbb{Z} \times \mathbb{Z}_+ : x+n \text{ is even}\}.$$

Site percolation is simply the process that consists of a collection of independent Bernoulli random variables with the same success probability p indexed by this set, i.e.,

$$\xi(x,n) := \text{Bernoulli}(p) \quad \text{for each site} \quad (x,n) \in \mathcal{G}.$$

Following the same terminology as before, site (x,n) is said to be

$$\textbf{open} \text{ when } \xi(x,n) = 1 \text{ and } \textbf{closed} \text{ when } \xi(x,n) = 0.$$

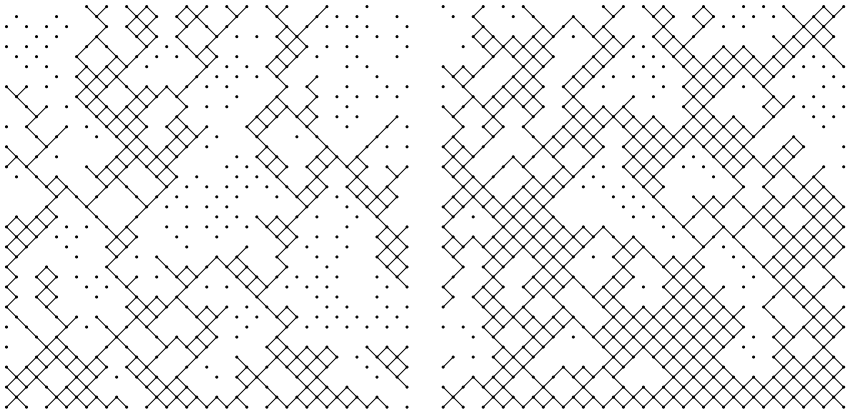


FIGURE 15. Oriented site percolation with parameter $p = 0.70$ and $p = 0.75$.

We say that there is an **open path** from $(0,0)$ to (x,n), which we write $(0,0) \to (x,n)$, whenever there are $x_0 = 0, x_1, \ldots, x_n = x$ such that

$$|x_{m+1} - x_m| = 1 \text{ for } m = 0, 1, \ldots, n-1 \quad \text{and} \quad (x_m, m) \text{ is open for } m = 0, 1, \ldots, n.$$

Finally, we define the **open cluster** starting at the origin by

$$C_0 \ := \ \{(x,n) \in \mathcal{G} : (0,0) \to (x,n)\}.$$

As for bond percolation, the previous construction induces a random subgraph C_0 and we define the **percolation probability** as in (15.2) by setting

$$\theta(p) \ := \ P_p\left(|C_0| = \infty\right), \tag{16.1}$$

which can be proved to be nondecreasing with respect to the percolation parameter p relying again on a coupling argument. Putting the graph upside down, and thinking of the orientation of the graph as gravity, we define the set of **wet sites** at level n as

$$W_n \ := \ \{y \in \mathbb{Z} : (x,0) \to (y,n) \text{ for some } x \in 2\mathbb{Z}\}.$$

Note that wet sites are open but open sites are not necessarily wet as shown in the simulation pictures of Figure 15 in which open sites are represented by black dots whereas wet sites are the one that can in addition be reached by a path of open sites. Using the terminology of epidemiology, one can think of open as susceptible and wet as infected: susceptible individuals become infected only if they can be reached by a path of infection.

Critical value and contour argument. The monotonicity of the percolation probability implies again the existence of a critical value, which is defined as in (15.5) by

$$p_c \ := \ \sup\{p : \theta(p) = 0\}.$$

Unlike the critical value of bond percolation in two dimensions, this critical value cannot be computed explicitly, however it can be proved that it is not degenerate in the sense that it belongs to

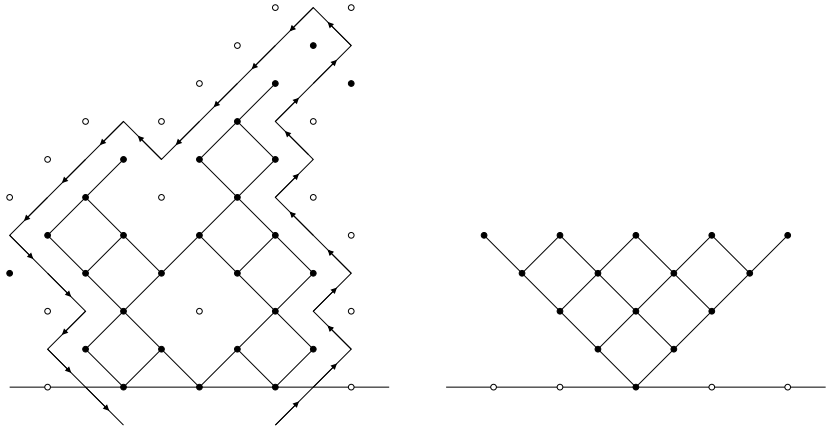


FIGURE 16. Pictures of the contour argument and the event $W_N = \{-N, \ldots, N\}$.

the open interval $(0,1)$. The fact that the critical value is strictly positive simply follows from the same argument as in the proof of Theorem 15.7. The proof that it is strictly smaller than one is more involved and relies on a so-called **contour argument**.

Theorem 16.1 – For oriented site percolation in two-dimensions, $p_c \le 1 - 3^{-4}$.

PROOF. First, we let C_N denote the set of wet sites when starting from

$$W_0 \ = \ \{-2N, -2N+2, \ldots, 0\}$$

and observe that

$$\begin{aligned} \theta(p) \ &:= \ P_p\left(|C_0| = \infty\right) \\ &\ge \ P_p\left(|C_N| = \infty\right) \times P_p\left(W_N = \{-N, \ldots, N\} \,|\, W_0 = \{0\}\right) \\ &\ge \ P_p\left(|C_N| = \infty\right) \times p^{1+2+\cdots+N+(N+1)}. \end{aligned} \tag{16.2}$$

In particular, it suffices to prove that

$$P_p\left(|C_N| < \infty\right) < 1 \quad \text{for } p > 1 - 3^{-4} \text{ and some } N \text{ large enough.}$$

For each realization such that $|C_N| < \infty$ we define the oriented contour Γ as indicated in the left-hand side of Figure 16. The idea of the contour argument is to bound

$$P_p\left(|C_N| < \infty\right) \ = \ P_p\left(\Gamma \text{ exists}\right). \tag{16.3}$$

There are three keys to estimating this probability.

- There are at most 3^m contours of length m.
- Sites to the right of $\swarrow$ and $\nwarrow$ must be closed. Also, sites to the right of two such arrows going up or two such arrows going down cannot be identical. Since in addition,

  $$\text{number of } \swarrow \ + \text{number of } \nwarrow \ \ge \ m/2$$

  for any contour of length m, at least $m/4$ sites must be closed.

- The shortest possible contour has length $2N + 2$.

It follows that for all $p > 1 - 3^4$ there is N large such that

$$P_p\,(\Gamma \text{ exists}) \;=\; \textstyle\sum_{m\geq 2N+2} P_p(\Gamma \text{ has length } m) \;\leq\; \sum_{m\geq 2N+2} 3^m\,(1-p)^{m/4} \;<\; 1.$$

This, together with (16.2) and (16.3), implies that $p_c \leq 1 - 3^{-4}$ as desired. □

17. Interacting particle systems

Most mathematical models introduced in the life and social sciences literature that describe inherently spatial phenomena of interacting populations consist of systems of ordinary differential equations or stochastic counterparts assuming global interactions. These models, however, leave out any spatial structure, while past research has identified the spatial component as an important factor in how communities are shaped and spatial models can result in predictions that differ from nonspatial models. In contrast, the framework of interacting particle systems, introduced independently by Frank Spitzer in the United States and Roland Dobrushin in Soviet Union in the seventies, is ideally suited to investigate the consequences of the inclusion of a spatial structure in the form of stochastic and local interactions. In these models, members of the population (particles) such as cells, plants or agents, are traditionally located on the set of vertices of a regular lattice but the framework can be extended to more general connected graphs. The dynamics is dictated by the geometry of the graph as particles can only interact locally with their neighbors, thus modeling the presence of a spatial structure. The mathematical analysis of interacting particle systems aims to deduce the macroscopic behavior of the system and the emergence of spatial patterns from the microscopic rules described by the local interactions. For more details than what can be found in these notes about particle systems, we refer to Liggett [19, 20] and Durrett [11].

General framework. From a mathematical point of view, an interacting particle system is a continuous-time Markov process with state space

$$\{0, 1, \ldots, \kappa - 1\}^{\mathbb{Z}^d} :=\; \text{all functions } \eta \text{ that map } \mathbb{Z}^d \text{ into } \{0, 1, \ldots, \kappa - 1\}.$$

In the terminology of interacting particle systems,

- Elements $x \in \mathbb{Z}^d$ are called **sites** or, using the terminology of graph theory, **vertices**, and have to be thought of as spatial locations.
- Elements $i \in \{0, 1, \ldots, \kappa - 1\}$ are called **types** or **colors**, e.g., in population dynamics, one can think of the types as empty and occupied, in disease dynamics as healthy, infected and immune, in opinion dynamics as Democrat and Republican.
- The function η which represents the state of the Markov chain is traditionally called a **spatial configuration**: $\eta_t(x)$ is equal to the type of vertex x at time t.

Note that there is a natural one-to-one correspondence

$$\psi \;:\; [0, 1] \;\longrightarrow\; \{0, 1\}^{\mathbb{N}}$$

which implies that the cardinality of the state space is at least equal to $2^{\aleph_0} = \aleph_1$ therefore the state space is not countable. In particular, the dynamics of an interacting particle system cannot be described by an infinitesimal matrix (10.2). Instead, the dynamics is described by local interactions

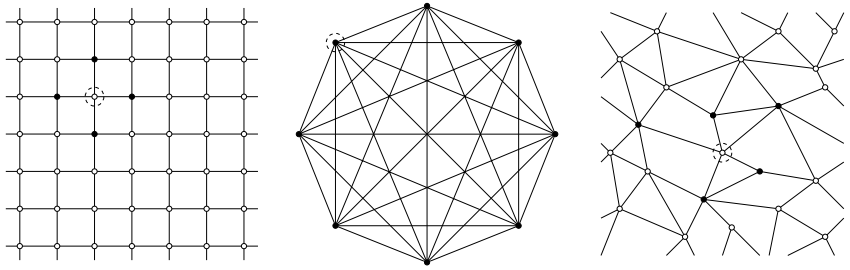


Figure 17. Regular lattice, complete graph and planar graph.

indicating the rate at which a vertex changes its type based on the state of the system in a neighborhood of the vertex. The fact that this rate only depends on a neighborhood models explicit space. To define the dynamics mathematically, we let

$$N_x \;:=\; \{y \in \mathbb{Z}^d : \|x - y\| = 1\} \quad \text{for } x \in \mathbb{Z}^d \tag{17.1}$$

be the **interaction neighborhood** of vertex x. Then, it is assumed that the type at vertex x flips from type i to type $j \neq i$ at a rate

$$\lim_{h\to 0} h^{-1}\,P\,(\eta_{t+h}(x) = j \,|\, \eta_t(x) = i) \;=\; c_{i\to j}(x, \eta_t) \;=\; c_{i\to j}(\eta_t(y) : y \in N_x) \tag{17.2}$$

that only depend on the configuration of the system in the neighborhood (17.1) of vertex x, which is referred to as local interactions. Note that the framework can be easily extended to more general connected graphs $G = (V, E)$ by simply redefining

$$N_x \;:=\; \{y \in V : (x, y) \in E\} \quad \text{for all} \quad x \in V. \tag{17.3}$$

The graph G has to be thought of as the network of interactions: vertices x and y can only interact if they are connected by an edge. The importance of space into dynamical stochastic models can thus be divided into three different levels:

1. **No space** means no interaction: the branching process and the simple birth and death process are examples of such processes since the number of offspring produced by a given individual is independent of the state of the process.
2. **Implicit space** means global interactions: there are spatial locations but all pairs of spatial locations are equally likely to interact so there is no geometrical structure, e.g., the logistic growth process, the Moran model and the Wright-Fisher model.
3. **Explicit space** means local interactions: there are spatial locations along with a geometry encoded by the geometry of a graph and pairs of spatial locations can only interact if they are connected by an edge, a component which is taken into account by the framework of interacting particle systems.

To understand how the general local transition rates (17.2) defined above work in practice, we now focus on the contact process and the voter model, two of the simplest and most popular examples of interacting particle systems.

Invasion: the contact process. The contact process introduced by Harris [16] is the basic invasion model based on the framework of interacting particle systems. We describe the model in the context of population dynamics but it can be turned into an epidemic model by replacing empty by healthy and occupied by infected. Each site of the lattice is either empty or occupied by an individual. Each individual gives birth at rate β and dies at rate one. At birth, the offspring is sent to one of the $2d$ nearest neighbors of the parent's vertex chosen uniformly at random. If this neighbor is empty it becomes occupied, whereas if it is occupied the birth is suppressed. To compute the local transition rates, we observe that an empty site receives an offspring from each of its occupied neighbor at rate $\beta/2d$, hence, the dynamics is described by the rate functions

$$c_{0\to 1}(x,\eta) \;:=\; (\beta/2d)\,\textstyle\sum_{y\in N_x}\eta(y) \qquad \text{and} \qquad c_{1\to 0}(x,\eta) \;:=\; 1. \tag{17.4}$$

As mentioned above, the contact process can as well be defined on more general graphs. An interesting particular example is the **complete graph** with N vertices. Recall that in such a graph all the vertices are connected to each other and for our purpose it is also convenient to assume that each vertex is connected to itself. In particular, the interaction neighborhood (17.3) of each vertex is equal to the entire vertex set therefore the description of the contact process on the complete graph reduces to the following:

> There is a set of N sites or spatial locations, each of which can be either empty or occupied by an individual. Individuals give birth at rate β and die at rate one. At birth, the offspring is sent to a site chosen uniformly at random. If the site is empty it becomes occupied, otherwise the birth is suppressed.

In particular, the number of individuals in the contact process on the complete graph is nothing else than the **logistic growth process** so the contact process on the lattice can be seen as the spatial analog of the logistic growth process.

Competition: the voter model. The voter model, introduced independently by Clifford and Sudbury [6] and Holley and Liggett [17], is the basic competition model based on the framework of interacting particle systems. We describe the model in the context of population genetics but it can be turned into an opinion model by replacing both types of allele by two differing opinions. In the voter model, each site of the lattice is occupied by a haploid individual with an allele of type 0 or an allele of type 1. Each individual gives birth at rate one. The offspring then replaces one of the $2d$ nearest neighbors of the parent's vertex chosen uniformly at random. To find the local transition rates, we observe that a site of type 0 receives an offspring of type 1 from each of its type 1 neighbor at rate $1/2d$, hence, the dynamics is described by the rate functions

$$\begin{aligned} c_{0\to 1}(x,\eta) &:= (1/2d)\,\textstyle\sum_{y\in N_x}\eta(y) \\ c_{1\to 0}(x,\eta) &:= (1/2d)\,\textstyle\sum_{y\in N_x}(1-\eta(y)). \end{aligned} \tag{17.5}$$

As previously, the model can be defined on more general graphs, and we look at the particular example of the complete graph with N vertices. In this case,

> There is a set of N sites, each of which is occupied by a haploid individual with either allele 0 or allele 1. Each individual gives birth to offspring of its own kind at rate one. The offspring then replaces one of the N individuals chosen uniformly at random.

In particular, the number of type 1 individuals in the voter model on the complete graph is nothing else than the **Moran model** therefore the voter model on the lattice can be seen as the spatial analog of the Moran model and its cousin the **Wright-Fisher model**. The advantage of thinking of spatially implicit stochastic models as interacting particle systems on the complete graph is that very powerful techniques developed for interacting particle systems can also be used for spatially implicit models, which often gives very strong results.

18. Graphical representation

The graphical representation introduced by Harris [15] is a powerful tool from both a theoretical and a practical point of view. It was originally introduced to prove the existence and uniqueness of a large class of interacting particle systems. However, the graphical approach also gives an explicit construction of spatial processes which allows for instance to couple different models or to define duality relationships between the system and other stochastic processes. The idea is to construct the process under consideration from a collection of independent Poisson processes. For concreteness, we focus our attention on the examples of the basic contact process and the voter model.

Collections of independent Poisson processes. As for continuous-time Markov chains with finite state space, interacting particle systems can be constructed from a collection of independent Poisson processes. However, since there are *a priori* infinitely many sites, there are also infinitely many collections of Poisson processes but because the dynamics is described by local interactions, there are in fact finitely many Poisson processes at each site. Looking at the d-dimensional lattice and adding one dimension for time, the graphical representation can be seen as a random graph in dimension $d+1$, that we call a percolation structure. For concreteness, we now deal with the contact process. There are two types of events in (17.4): births and deaths, so we let

- $\{T_n(x,y) : n \geq 1\}$ for $\|x-y\| = 1$ be the times of a rate $\beta/2d$ Poisson process,
- $\{U_n(x) : n \geq 1\}$ for $x \in \mathbb{Z}^d$ be the times of a rate one Poisson process.

To turn this into a graph, we draw an arrow $x \to y$ at times $T_n(x,y)$ to indicate that a birth occurs if the tail is occupied and the tip is empty, and a cross at site x at times $U_n(x)$ to indicate that a death occurs if the site is occupied. A realization of the resulting graph is represented on the left-hand side of Figure 18 where it is used to construct the contact process starting with all sites occupied: space-time points which are occupied are drawn in thick lines. The construction of the voter model (17.5) is easier as it only involves one type of events. Let

- $\{T_n(x,y) : n \geq 1\}$ for $\|x-y\| = 1$ be the times of a rate one Poisson process

and draw an arrow $x \to y$ at times $T_n(x,y)$ to indicate that the individual at the tip takes on the type of the individual at the tail. The right-hand side of Figure 18 shows how to construct the voter model given a particular initial configuration and a realization of this graphical representation.

Numerical simulations in finite volume. In practice, the Harris' graphical representation can be used to simulate numerically interacting particle systems. Here, we first give an algorithm to simulate the contact process on the $N\times N$ lattice with periodic boundary conditions. This algorithm relies on basic properties of the exponential random variable and Poisson processes.

(a) WHEN – According to Lemma 9.3, the time of the first potential update, i.e., the first time a symbol appears in the graphical representation, is given by

$$T \;:=\; \min\left(T_1(x,y) : (x,y) \in E\right) \wedge \min\left(U_1(x) : x \in V\right) \;=\; \text{Exponential}\left((\beta+1)\,N^2\right).$$

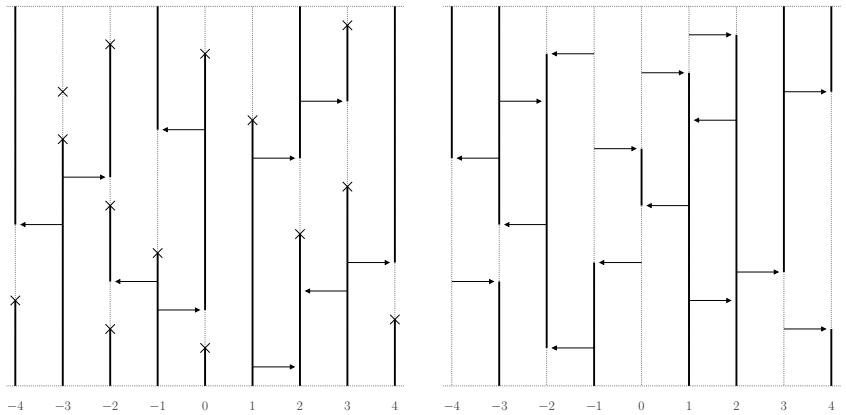


Figure 18. Graphical representation of the contact process and the voter model.

(b) WHERE – According again to Lemma 9.3, each of the vertices is equally likely to be the first one at which a birth or death occurs therefore this vertex is

$$X \;=\; \text{Uniform}\,\{1,2,\ldots,N\}^2.$$

(c) WHAT – If site X is empty then nothing happens so we go to (d). Else, since Lemma 9.3 implies that the first symbol that appears in the graphical representation is a death mark with probability $(\beta+1)^{-1}$ and an arrow with probability $\beta\,(\beta+1)^{-1}$

- With probability $(\beta+1)^{-1}$ we kill the individual at vertex X.
- With probability $\beta\,(\beta+1)^{-1}$ we choose one of the four neighbors of vertex X uniformly at random and put a particle at this site if it is empty.

(d) By memoryless of the exponential random variable proved in Theorem 9.8, the characteristics of the next update are the same as for the first update so we go to (a).

Note that step (a) is irrelevant to understand the long-term behavior of the contact process, in which case the algorithm reduces to the loop (b) → (c) → (b). The following is the algorithm written in C for the process with birth parameter $\beta := 2$ on the 50×50 lattice with periodic boundary conditions. The program displays the configuration at time 100.

```
/* Contact Process */

    #include <stdlib.h>
    #include <stdio.h>
    #include <time.h>
    #include <math.h>
    #define T 100
    #define beta 2
    #define N 50

    int eta [N * N] ;
    int x, y, z ;
    float s, t, u, mu ;
```



```
// Take a number uniformly at random in (0, 1)
    float uniform () {
      u = (float) rand() / RAND_MAX ;
      return u ;
    }
// Exponential random variable with parameter mu
    float exponential (float u, float mu) {
      s = (float) - log (u) / mu ;
      return s ;
    }
// Choose a neighbor of x on the 2D torus
    int neighbor (int x, float u) {
      z = floor (4 * u) ;
      switch (z) {
        case 0 : y = x + 1 ; break ;
        case 1 : y = x - 1 ; break ;
        case 2 : y = x + N ; break ;
        case 3 : y = x - N ; break ;
      }
      y = (y + N * N) % (N * N) ;
      return y ;
    }
// Contact process
    int main (int argc, char * argv[]) {
      time_t amorce ;
      time (&amorce) ;
      srand (amorce) ;
      t = 0 ;
// All sites occupied initially
      for (x = 0 ; x < N * N ; x ++) {
        eta [x] = 1 ;
      }
      while (t < T) {
// Time of the next potential update
        u = uniform () ;
        t = t + exponential (u, (beta + 1) * N * N) ;
// Choose a site x uniformly at random
        u = uniform () ;
        x = floor (N * N * u) ;
        x = x % (N * N) ;
// If site x is occupied
        if (eta [x] == 1) {
          u = uniform () ;
          u = u * (beta + 1) ;
// The particle at x dies with probability 1 / (beta + 1)
          if (u < 1) {
            eta [x] = 0 ;
          }
// The particle at x gives birth with probability beta / (beta + 1)
          else {
            u = uniform () ;
            y = neighbor (x, u) ;
            eta [y] = 1 ;
          }
        }
      }
// Display the final configuration in the terminal: crosses are occupied sites
      for (y = 0 ; y < N ; y ++) {
        for (x = 0 ; x < N ; x ++) {
          if (eta [x + N * y] == 1) {
            printf ("x") ;
          }
          else {
            printf (" ") ;
          }
        }
        printf ("\n") ;
      }
```

```
    }
```

The algorithm to simulate the voter model is simpler. The times between consecutive potential updates are exponentially distributed with parameter the number of sites. At these times, we choose a site, say x, uniformly at random, then a neighbor of this site, say y, and finally set the type at site x equal to the type at site y. The C program is the same as for the contact process except for the initial configuration that should involve both types to avoid trivialities and the while loop describing the local interactions. The corresponding parts of the program become

```
/* Voter Model */
// Start with a product measure with density 1/2
      for (x = 0 ; x < N * N ; x ++) {
        u = uniform () ;
        if (u < 0.5) {
          eta [x] = 0 ;
        }
        else {
          eta [x] = 1 ;
        }
      }
      while (t < T) {
// Time of the next potential update
        u = uniform () ;
        t = t + exponential (u, N * N) ;
// Choose a site x uniformly at random
        u = uniform () ;
        x = floor (N * N * u) ;
        x = x % (N * N) ;
// Choose a neighbor y
        u = uniform () ;
        y = neighbor (x, u) ;
// Site x mimics site y
        eta [x] = eta [y] ;
      }
```

Existence on infinite graphs. On infinite graphs, interacting particle systems cannot be generated by a computer and even their existence is not obvious. Indeed, since there are infinitely many Poisson processes, the time to the first update does not exist. To prove the existence of the process, one must show that the state of any given space-time point can be deduced from only finitely many interactions which can therefore be ordered in time, a result due to Harris [15]. For concreteness, we focus on the contact process but the proof easily extends to the general framework described in Section 17 provided the degree of the graph is uniformly bounded. First, we let $t_0 > 0$ be small and, having a realization of the graphical representation, declare an edge (x, y) to be open if and only if there is an arrow that connects both vertices by time t_0. Then, defining open paths as for bond percolation (15.1), we call island of influence of x the random cluster

$$C_x \;:=\; \{y \in \mathbb{Z}^d : \text{there is an open path connecting } x \text{ and } y\}.$$

It follows from this construction that the graphical representation outside C_x is not relevant in determining the state of vertex x up to time t_0. Moreover,

Lemma 18.1 – There is $t_0 > 0$ small such that $P(|C_x| = \infty \text{ for some } x \in \mathbb{Z}^d) = 0$.

Proof. From the graphical representation, it is clear that edges are independently open with probability $1 - \exp(-\beta t_0/d)$ which is $< (2d-1)^{-1}$ for $t_0 > 0$ small. The result then follows from the

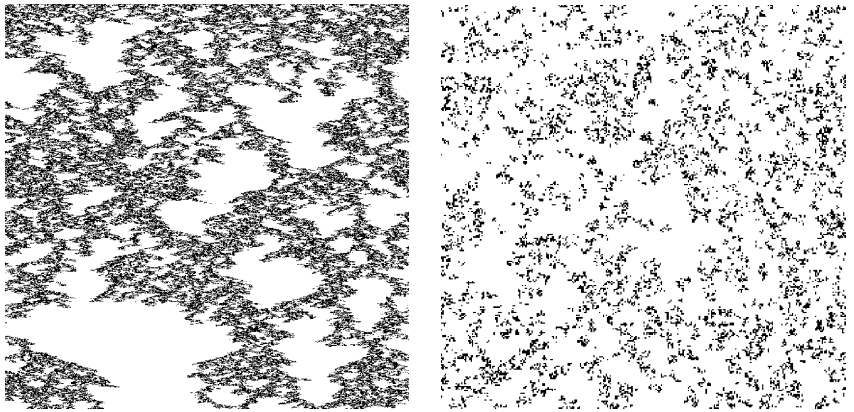


Figure 19. Realization of the contact process with birth parameter $\beta = 3.30$ from time 0 at the top of the picture to time 10,000 at the bottom on the one-dimensional torus with 600 vertices, and snapshot at time 1000 of the contact process with birth parameter $\beta = 1.70$ on a 300×300 lattice with periodic boundary conditions.

bound for the critical value of bond percolation given in Theorem 15.7. □

The previous lemma breaks down the lattice into finite islands that do not interact with each other before time t_0, therefore the process can be constructed independently on each of those islands up to this time. Since the exponential random variable is memoryless, and so the graphical representation translation invariant in time, we can restart the procedure to construct the process up to time $2t_0$ and so on. In conclusion, the process is well defined.

19. Contact process on lattices

In this section, we prove that, similarly to the other invasion models introduced in this course, the contact process exhibits a phase transition: there is a critical birth parameter above which the process starting with one individual survives with positive probability but below which it goes extinct eventually with probability one. We also give a lower bound for the critical value on the lattice to show an important feature of the contact process: due to the inclusion of space in the form of local interactions, the critical value is strictly larger than one, which contrasts with nonspatial and spatially implicit invasion models.

Probability of survival. First, we note that the inclusion of local interactions allows to define multiple notions of survival. We distinguish two "levels" of survival. Assuming that the process starts with a single individual at the origin, we say that

- The process **dies out** when $\eta_t = \varnothing$ for some t.
- The process **survives weakly** when $\eta_t \neq \varnothing$ for all t.
- The process **survives strongly** when $\limsup_{t\to\infty} \eta_t(0) = 1$.

To prove the existence of a phase transition, the first step is to prove that the probability of weak/strong survival is nondecreasing with respect to the birth rate. Let

$$\begin{aligned} \theta_W(\beta) &:= P_\beta\,(\eta_t \neq \varnothing \text{ for all } t) \\ \theta_S(\beta) &:= P_\beta\,(\eta_t(0) = 1 \text{ infinitely often}) \end{aligned} \tag{19.1}$$

be the probability of weak and strong survival when starting with a single individual.

Theorem 19.1 – The functions $\beta \mapsto \theta_W(\beta)$ and $\beta \mapsto \theta_S(\beta)$ are nondecreasing.

Proof. This follows from a coupling argument. The idea is to construct the process (η_t) with parameter β_1 and the process (ξ_t) with parameter $\beta_2 \geq \beta_1$ starting from the same initial configuration and using the same graphical representation in such a way that $\eta_t \leq \xi_t$ at all times.

- For all $x, y \in \mathbb{Z}^d$ with $\|x - y\| = 1$, we draw a continuous arrow $x \to y$ at the arrival times of a Poisson process with parameter $\beta_1/2d$.
- For all $x, y \in \mathbb{Z}^d$ with $\|x - y\| = 1$, we draw a dashed arrow $x \to y$ at the arrival times of a Poisson process with parameter $(\beta_2 - \beta_1)/2d$.
- For all $x \in \mathbb{Z}^d$, we put a death mark $\times$ at vertex x at the arrival times of a Poisson process with parameter one.

Then, the contact process (η_t) can be constructed from this graphical representation by ignoring the dashed arrows while the process (ξ_t) can be constructed by using both types of arrows. Since the set of arrows for the first process is included in the set of arrows for the second process, the space-time region occupied by individuals in the first process is included in its counterpart in the second process, as illustrated in Figure 20, i.e., $\eta_t \leq \xi_t$ at all times. To conclude, let

$$f(\zeta_t : t \in \mathbb{R}_+) := \mathbf{1}\{\zeta_t \neq \varnothing \text{ for all } t\} \quad \text{for} \quad \zeta_t : \mathbb{Z}^d \to \{0, 1\}$$

and observe that the function f is nondecreasing to get

$$\begin{aligned} \theta_W(\beta_1) &:= P\,(\eta_t \neq \varnothing \text{ for all } t) \\ &= E\,f((\eta_t : t \in \mathbb{R}_+)) \leq E\,f((\xi_t : t \in \mathbb{R}_+)) = \theta_W(\beta_2). \end{aligned}$$

The proof for the probability of strong survival is similar. □

The critical value. The monotonicity of the probability of weak/strong survival (19.1) with respect to the birth parameter β motivates the introduction of the two critical values

$$\beta_W := \sup\{\beta : \theta_W(\beta) = 0\} \quad \text{and} \quad \beta_S := \sup\{\beta : \theta_S(\beta) = 0\}.$$

Since strong survival implies weak survival, we have $\beta_W \leq \beta_S$. In fact, it is known that both critical values are equal for the contact process on regular lattices, and we let β_c be their common value. The contact process modified so that births onto occupied sites are not suppressed is called **branching random walk**. The number of individuals in this process evolves according to the simple birth and death process with birth rate β. In particular, a standard coupling argument and Theorem 11.1 imply that $\beta_c \geq 1$. As previously explained, an interesting aspect of the contact process is that, in contrast with branching processes and the simple birth and death process, the critical value is strictly larger than one. More precisely, we have the following result.

Theorem 19.2 – On the d-dimensional regular lattice, $\beta_c \geq 2d\,(2d-1)^{-1} > 1$.

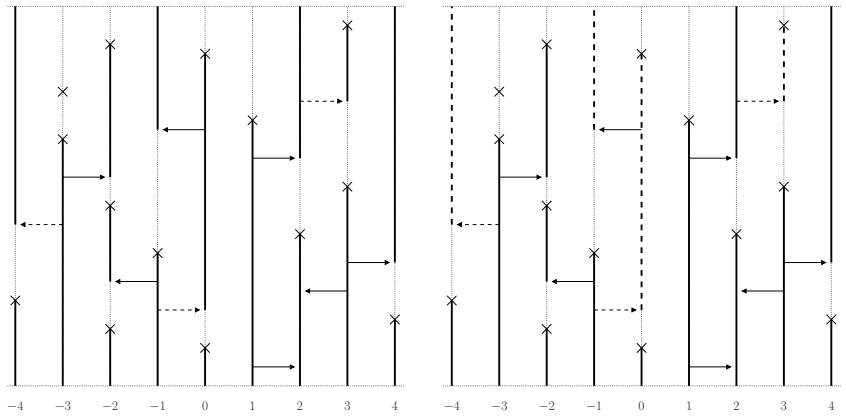


Figure 20. Coupling of contact processes with different birth parameters.

Proof. The first step is to introduce

$$\sigma(A) := P\,(\eta_t \neq \varnothing \text{ for all } t \,|\, \eta_0 = A) \quad \text{for all} \quad A \subset \mathbb{Z}^d \text{ finite.} \tag{19.2}$$

Then, using another coupling argument, i.e., constructing contact processes with the same parameter but starting from different initial configurations using the same graphical representation, it can be proved that the probability (19.2) is submodular, i.e.,

$$\sigma(A \cup B) + \sigma(A \cap B) \leq \sigma(A) + \sigma(B). \tag{19.3}$$

The second step is to use a first-step analysis to decompose $\sigma(A)$. Recall that

- an individual at site $x \in A$ dies at rate one and
- an empty site z with an occupied neighbor $x \in A$ becomes occupied due to a birth that originates from x at rate $\beta/2d$.

In particular, a first-step analysis gives

$$((\beta + 1)\,\mathrm{card}(A))\,\sigma(A) = \textstyle\sum_{x \in A} (\sigma(A - \{x\}) + (\beta/2d) \sum_{\|x-z\|=1} \sigma(A \cup \{z\})). \tag{19.4}$$

Then, we introduce the two probabilities

$$\sigma_1 := \sigma(\{x\}) \text{ and } \sigma_2 := \sigma(\{x, y\}) \quad \text{where} \quad \|x - y\| = 1.$$

Using equation (19.4) with $A := \{x, y\}$ and the submodularity (19.3), we obtain

$$\begin{aligned} 2\,(\beta + 1)\,\sigma_2 &= 2\sigma_1 + (\beta/d) \textstyle\sum_{\|x-z\|=1} \sigma(\{x, y, z\}) \\ &= 2\sigma_1 + (\beta/d)\,\sigma_2 + (\beta/d) \textstyle\sum_{\|x-z\|=1,\, z \neq y} \sigma(\{x, y, z\}) \\ &\leq 2\sigma_1 + (\beta/d)\,\sigma_2 + (\beta/d)(2d-1)(2\sigma_2 - \sigma_1). \end{aligned}$$

Some basic algebra gives

$$(\sigma_1 - \sigma_2)\,(2d - \beta\,(2d-1)) \geq 0. \tag{19.5}$$

Finally, we use that if $\sigma_1 > 0$ then $\sigma_1 < \sigma_2$, together with (19.5), to get

$$\beta < 2d\,(2d-1)^{-1} \implies \sigma_1 \geq \sigma_2 \implies \sigma_1 = 0 \implies \beta \leq \beta_c$$

therefore $\beta_c \geq 2d\,(2d-1)^{-1}$. □

Overview of the contact process. The contact process was only well understood in one dimension until the nineties when Bezuidenhout and Grimmett [2, 3] proved a number of open problems about the process in higher dimensions. The common background of their proofs is the use of block construction techniques from percolation theory. As previously explained, $\beta_W = \beta_S$ for the contact process on regular lattices, indicating that above the common critical value the process survives strongly whereas below the critical value the process dies out. Even though this critical value is not known, Bezuidenhout and Grimmett [2] proved using a continuity argument that the process dies out at criticality, just as in the branching process and the simple birth and death process.

Theorem 19.3 – On the d-dimensional regular lattice, $\theta_W(\beta_c) = \theta_S(\beta_c) = 0$.

The contact process has a certain property called attractiveness, which again can be proved easily using a coupling argument through its graphical representation. This property guarantees that the contact process starting with all sites occupied converges in distribution to a certain invariant measure μ that we call the **upper invariant measure**. Using another property of the process called **self-duality**, one can prove that

$$\mu \;\neq\; \delta_\varnothing \quad \text{if and only if} \quad \theta_W(\beta) = \theta_S(\beta) \;>\; 0$$

where the measure $\delta_\varnothing$ is the measure that concentrates on the configuration in which all sites are empty. The next theorem, called a **complete convergence theorem**, states that if the contact process survives then it converges to its upper invariant measure.

Theorem 19.4 – For all $A \subset \mathbb{Z}^d$, we have $\eta_t \Rightarrow \mu \cdot \sigma(A) + \delta_\varnothing \cdot (1 - \sigma(A))$ as $t \to \infty$.

Conditional on survival, the process converges in distribution to μ and we let K_t be the set in which the contact process starting with a single individual at the origin is distributed according to the upper invariant measure μ at time t. Our definition of this set is not rigorous but can be made rigorous by using a coupling between the process starting with a single individual and the process starting with all sites occupied. The next theorem, called a **shape theorem**, states that this set grows asymptotically linearly in each direction.

Theorem 19.5 – Conditioned on survival, there is a convex set $U \subset \mathbb{R}^d$ such that

$$(1-\epsilon)(U \cap \mathbb{Z}^d) \;\subset\; t^{-1} K_t \;\subset\; (1+\epsilon)(U \cap \mathbb{Z}^d) \quad \text{eventually for all } \epsilon > 0.$$

Bezuidenhout and Grimmett [3] also proved an important result in the subcritical case that basically states that the time to extinction of the subcritical contact process starting with a single individual decays exponentially. This is the analog of Theorem 15.2 regarding bond percolation.

Theorem 19.6 – Assume that $\beta < \beta_c$. Then, $P(\eta_t \neq \varnothing) \leq \exp(-\alpha\, t)$ for some $\alpha > 0$.

Shortly after the work of Bezuidenhout and Grimmett, researchers in the field of interacting particle systems studied the contact process on other graphs. The first remarkable result in this topic is due to Pemantle [22] who proved that $\beta_W < \beta_S$ for the contact process on regular trees with degree larger or equal to three. In other words, on regular trees, there is an intermediate phase in which the process survives weakly but not strongly. In this intermediate phase, any finite region gets empty eventually but the population persists by drifting off to infinity.

20. Voter model on lattices

Even though it does not depend on any parameter, the voter model exhibits very rich dynamics with different types of behaviors depending on the spatial dimension. In this section, we first rely on the graphical representation introduced in Section 18 to describe the so-called duality relationship between the voter model and coalescing random walks looking at the evolution backwards in time, i.e., the ancestry of a finite set of individuals evolves according to a system of coalescing random walks. This duality relationship is then used to prove the main result about the voter model: it clusters in one and two dimensions whereas coexistence occurs in higher dimensions. We conclude with an overview of additional important results.

Duality with coalescing random walks. To exhibit the so-called duality relationship between the voter model and coalescing random walks, we first think of the process as being constructed from the Harris' graphical representation given in section 18. Then, we say that there exists a path from $(y, t-s)$ and (x, t) if there are sequences of times and vertices

$$s_0 \;=\; t-s \;<\; s_1 \;<\; \cdots \;<\; s_{n+1} \;=\; t \quad \text{and} \quad x_0 \;=\; y,\, x_1,\, x_2,\, \ldots,\, x_n \;=\; x$$

such that the following two conditions hold:

- for $i = 1, 2, \ldots, n$, there is an arrow $x_{i-1} \to x_i$ at time s_i and
- for $i = 0, 1, \ldots, n$, there is no arrow pointing at the segment $\{x_i\} \times (s_i, s_{i+1})$.

This also defines a **dual path** from (x, t) to $(y, t-s)$, i.e., dual paths evolve backwards in time. For each subset $A \subset \mathbb{Z}^d$ finite, the **dual process** starting at (A, t) is

$$\hat{\eta}_s(A, t) \;:=\; \{y : \text{there is a dual path from } (x, t) \text{ to } (y, t-s) \text{ for some } x \in A\}. \tag{20.1}$$

Note that the dual process consists of a system of **coalescing random walks**: there is one symmetric simple random walk starting at each point of A at time t evolving at rate one backwards in time, and each collision among these random walks results in a coalescing event, as shown in the right-hand side of Figure 22. Note also that the dual process (20.1) keeps track of the ancestors of the individuals in A at time t. In particular, setting

$$Z_s(x) \;:=\; \hat{\eta}_s(\{x\}, t) \quad \text{for all } x \in \mathbb{Z}^d$$

we have the **duality relationship**

$$\eta_t(x) \;=\; \eta_{t-s}(Z_s(x)) \;=\; \eta_0(Z_t(x)) \quad \text{for all } 0 \leq s \leq t. \tag{20.2}$$

This indicates that the type of site x at time t can be deduced based on the duality relationship from the configuration of the system at earlier times.

Clustering versus coexistence. From now on, we assume that the voter model starts from the product measure with density $\theta \in (0, 1)$ in which sites are of type 1 with probability θ independently of each other. Using the duality between the voter model and coalescing random walks, we first prove first that the system clusters in one and two dimensions.

Theorem 20.1 – Assume that $d \leq 2$. Then,

$$\lim_{t\to\infty} P(\eta_t(x) \neq \eta_t(y)) \;=\; 0 \quad \text{for all} \quad x, y \in \mathbb{Z}^d.$$

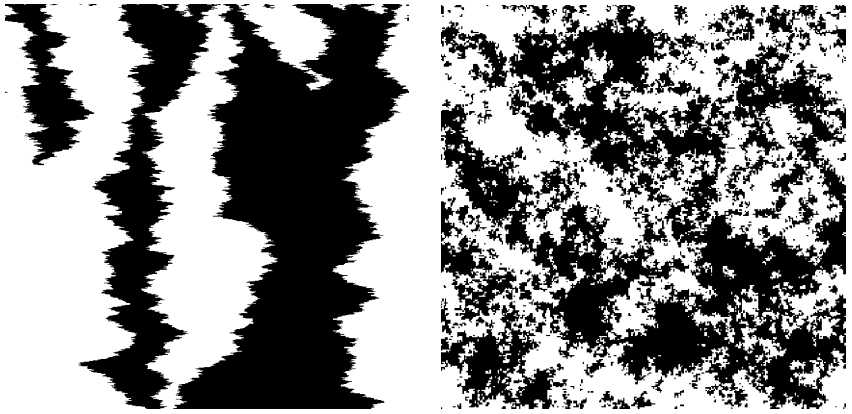


FIGURE 21. Realization of the voter model from time 0 to time 10,000 on the one-dimensional torus with 600 vertices, and snapshot at time 1000 of the voter model on a 300×300 lattice with periodic boundary conditions.

PROOF. Since $Z_s(x)$ and $Z_s(y)$ evolve according to independent random walks run at rate one until they coalesce, the difference between the random walks $Z_s(x) - Z_s(y)$ is a continuous-time random walk run at rate two absorbed at site 0. But since random walks are recurrent in one and two dimensions according to Theorems 6.1 and 6.2, we have

$$\lim_{t\to\infty} P(Z_t(x) \neq Z_t(y)) \;=\; \lim_{t\to\infty} P(Z_s(x) - Z_s(y) \neq 0 \text{ for all } s < t) \;=\; 0.$$

By duality (20.2), we conclude that

$$\lim_{t\to\infty} P(\eta_t(x) \neq \eta_t(y)) \;\leq\; \lim_{t\to\infty} P(Z_t(x) \neq Z_t(y)) \;=\; 0.$$

This completes the proof. □

In contrast, the process converges to a stationary distribution in which both types are present in higher dimensions: both types coexist at equilibrium.

Theorem 20.2 – Assume that $d \geq 3$. Then, η_t converges in distribution to an invariant measure in which there is a positive density of both types.

PROOF. To prove convergence to a stationary distribution, we first observe that there is no type 1 individual in the set A at time t if and only if there is no type 1 individual in the corresponding dual process at time 0, therefore

$$P(\eta_t \cap A = \varnothing) \;=\; E\left((1-\theta)^{|\hat{\eta}_t(A,t)|}\right). \tag{20.3}$$

The dominated convergence theorem 1.8 implies that both terms in (20.3) have a limit when time goes to infinity, which proves the existence of a stationary distribution. Moreover, using again duality (20.2) and the fact that symmetric simple random walks are transient in three or higher dimensions according to Theorem 6.3, we obtain

$$\begin{aligned}\lim_{t\to\infty} P(\eta_t(x) \neq \eta_t(y)) &= \lim_{t\to\infty} P(\eta_0(X_t(x)) \neq \eta_0(X_t(y)))\\ &= \lim_{t\to\infty} 2\theta(1-\theta)\, P(X_t(x) \neq X_t(y)) \;>\; 0.\end{aligned}$$

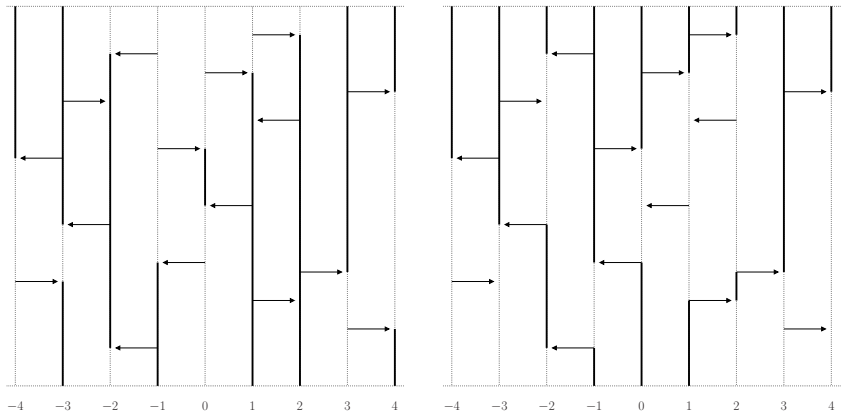


FIGURE 22. Graphical representation and dual process of the voter model.

This shows that the process converges to a stationary distribution in which the density of type 0 and the density of type 1 are both positive. □

Overview of the voter model. The duality between the voter model and coalescing random walks allows to answer additional important questions: How fast do clusters grow in one and two dimensions? How strong are the spatial correlations at equilibrium in higher dimensions? Does the type of any given site stay the same after a finite random time in low dimensions when clustering occurs? In this subsection, we answer these questions without proof. Observing that the density of type 1 is preserved by the voter model dynamics, the combination of Theorems 20.1 and 20.2 gives the following result that summarizes the long-term behavior of the voter model.

Theorem 20.3 – Starting with a density θ of type 1,

$$\begin{aligned}\eta_t &\;\Rightarrow\; (1-\theta)\,\delta_0 + \theta\,\delta_1 && \text{in } d \leq 2 && \text{(consensus)}\\ \eta_t &\;\Rightarrow\; \eta_\infty \text{ where } P(\eta_\infty(x) = 1) = \theta && \text{in } d \geq 3 && \text{(coexistence).}\end{aligned}$$

In one and two dimensions, two fixed sites, say x and y, are eventually totally correlated, and to understand how fast clusters grow, the idea is to look at the correlations between two evolving sites whose distance increases with time, say $t^\alpha\, x$ and $t^\alpha\, y$ which we identify with their nearest neighbor on the lattice. Bramson and Griffeath [5] proved that the size of the clusters scales asymptotically like the square root of time in one dimension.

Theorem 20.4 (cluster size in $d = 1$) – As $t \to \infty$,

$$\begin{aligned}P(\eta_t(t^\alpha\, x) \neq \eta_t(t^\alpha\, y)) &\;\to\; 2\theta(1-\theta) && \text{when } \alpha > 1/2 && \text{(independence)}\\ &\;\to\; 0 && \text{when } \alpha < 1/2 && \text{(total correlations).}\end{aligned} \tag{20.4}$$

The behavior of the process in two dimensions is more interesting: though sites are again independent when $\alpha > 1/2$, Cox and Griffeath [8] proved the following result which indicates that there is no natural scale for the cluster size in two dimensions.

Theorem 20.5 (cluster size in $d = 2$) – As $t \to \infty$,

$$\begin{array}{rcll} P\left(\eta_t(t^\alpha\, x) \neq \eta_t(t^\alpha\, y)\right) & \to & 2\theta(1-\theta) & \text{when } \alpha > 1/2 \quad \text{(independence)} \\ & \to & 4\theta(1-\theta)\,\alpha & \text{when } \alpha < 1/2 \quad \text{(some correlations).} \end{array} \tag{20.5}$$

In higher dimensions, since coexistence occurs, two fixed sites are never totally correlated, even at equilibrium. However, due to the presence of local interactions, sites are not independent either and a natural question is: How strong are the spatial correlations? The answer has been given by Bramson and Griffeath [4] in three dimensions and by Zähle [25] in higher dimensions.

Theorem 20.6 (spatial correlations in $d \geq 3$) – As $n \to \infty$,

$$n^{-(d+2)/2} \sum_{\|x\|<n} (\eta_\infty(x) - \theta) \;\Rightarrow\; \mathcal{N}(0, \sigma^2).$$

In the theorem, $\mathcal{N}(0, \sigma^2)$ refers to the centered normal distribution. Note that, since the number of sites in the sum is of the order of n^d, spatial correlations decaying exponentially fast at equilibrium would give a renormalization constant of $n^{-d/2}$ so the theorem shows that, even in $d \geq 3$, spatial correlations due to local interactions are strong. Finally, we look at the fraction of time a given site, say x, is of type 1 in the long run, a random variable called the **occupation time**. When coexistence occurs, sites change their type infinitely often therefore it is expected that the fraction of time site x is of type 1 converges almost surely to θ. In contrast, when clustering occurs, one can naively think that a cluster will eventually cover site x and that the previous law of large number does not hold. In fact, Cox and Griffeath [7] proved convergence in $d \geq 2$.

Theorem 20.7 (occupation time) – Assume that $d \geq 2$. Then,

$$\frac{1}{t} \int_0^t \eta_s(x)\, ds \;\to\; \theta \;\text{ as }\; t \to \infty. \tag{20.6}$$

This law of large numbers does not hold in one dimension. However, it is known that the system of annihilating random walks that describes the evolution of the boundaries of the voter model in one dimension is site recurrent, which implies that, even in one dimension, the type of any given site keeps changing indefinitely. From the combination of (20.4)–(20.6), we obtain the following description of the voter model in one and two dimensions: clusters form and appear to grow indefinitely so only one type can be present at equilibrium. However, any given site flips infinitely often, which indicates that clusters are not fixed in space but move around, and thus may give the impression of local transience though, strictly speaking, coexistence does not occur.

References

[1] Adelman, O. (1976). Some use of some "symmetries" of some random process. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **12**, 193–197.

[2] Bezuidenhout, C. and Grimmett, G. (1990). The critical contact process dies out. *Ann. Probab.* **18** 1462–1482.

[3] Bezuidenhout, C. and Grimmett, G. (1991). Exponential decay for subcritical contact and percolation processes. *Ann. Probab.* **19** 984–1009.

[4] Bramson, M. and Griffeath, D. (1979). Renormalizing the 3-dimensional voter model. *Ann. Probab.* **7** 418–432.

[5] Bramson, M. and Griffeath, D. (1980). Clustering and dispersion rates for some interacting particle systems on $\mathbb{Z}$. *Ann. Probab.* **8** 183–213.

[6] Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika* **60** 581–588.

[7] Cox, J. T. and Griffeath, D. (1983). Occupation time limit theorems for the voter model. *Ann. Probab.* **11** 876–893.

[8] Cox, J. T. and Griffeath, D. (1986). Diffusive clustering in the two-dimensional voter model. *Ann. Probab.* **14** 347–370.

[9] Doyle, P. G. and Snell, E. J. (1984). *Random walks and electrical networks.* Carus Math. Monographs **22**, Math. Assoc. Amer,. Washington, D. C.

[10] Durrett, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.* **12** 999–1040.

[11] Durrett, R. (1995). Ten lectures on particle systems. In *Lectures on probability theory (Saint-Flour, 1993)*, volume 1608 of *Lecture Notes in Math.*, pages 97–201. Springer, Berlin.

[12] Durrett, R. (1996). *Probability: theory and examples*, second edition, Duxbury Press, Belmont, CA.

[13] Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3** 87–112.

[14] Grimmett, G. R. (1989). *Percolation.* Springer, Berlin.

[15] Harris, T. E. (1972). Nearest neighbor Markov interaction processes on multidimensional lattices. *Adv. Math.* **9** 66–89.

[16] Harris, T. E. (1974). Contact interactions on a lattice. *Ann. Probability*, 2:969–988.

[17] Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* **3** 643–663.

[18] Kesten, H. (1980). The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Comm. Math. Phys.* **74** 41–59.

[19] Liggett, T. M. (1985). *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York.

[20] Liggett, T. M. (1999). *Stochastic interacting systems: contact, voter and exclusion processes.* Berlin Heidelberg New York : Springer.

[21] Neuhauser, C. (2001). Mathematical models in population genetics. In *Handbook of Statistical Genetics* (ed. Cannings C), pages 153–178. John Wiley and Sons Ltd., Chichester, UK.

[22] Pemantle, R. (1992). The contact process on trees. *Ann. Probab.* **20** 2089–2116.

[23] Rudin, W. (1987). *Real and complex analysis*, third edition, McGraw-Hill Book Co., New York.

[24] Spitzer, F. (1976). *Principles of random walks*, second edition, volume 34 of *Graduate Texts in Mathematics.* Springer-Verlag, New York.

[25] Zähle, I. (2001). Renormalization of the voter model in equilibrium. *Ann. Probab.* **29** 1262–1302.