1

Probability and measure

1.1 Do probabilists need measure theory?

Measure theory provides the theoretical framework essential for the development of modern probability theory. Much of elementary probability theory can be carried through with only passing reference to underlying sample spaces, but the modern theory relies heavily on measure theory, following *Kolmogorov's* axiomatic framework (1932) for probability spaces. The applications of *stochastic processes*, in particular, are now fundamental in physics, electronics, engineering, biology and finance, and within mathematics itself. For example, *Itô's stochastic calculus* for *Brownian Motion* (BM) and its extensions rely wholly on a thorough understanding of basic measure and integration theory. But even in much more elementary settings, effective choices of sample spaces and σ-fields bring advantages – good examples are the study of random walks and branching processes. (See [S], [W] for nice examples.)

1.2 Continuity of additive set functions

What do we mean by saying that we pick the number $x \in [0, 1]$ at random? 'Random' plausibly means that in each trial with uncertain outcomes, each outcome is 'equally likely' to be picked. Thus we seek to impose the *uniform probability distribution* on the set (or *sample space*) Ω of possible outcomes of an experiment. If Ω has n elements, this is trivial: for each outcome ω, the probability that ω occurs is $\frac{1}{n}$. But when $\Omega = [0, 1]$ the 'number' of possible choices of $x \in [0, 1]$ is infinite, even uncountable. (Recall that the set $\mathbb{Q}$ of *rational* numbers is countable, while the set $\mathbb{R}$ of *real* numbers is uncountable.) We cannot define the 'uniform probability' on $[0,1]$ as a function of points x or singletons $\{x\}$; however, we can first define our probability function or (*Lebesgue*) *measure* m just for *intervals*: if $0 \leq a < b \leq 1$, we set $m([a,b]) = b-a$. Thus measure, or 'probability', is a function of *sets*, not of points. The challenge is to extend this idea to more general sets in $[0,1]$.

With such an extension we can determine $m(\{x\})$ for fixed $x \in [0,1]$: for $\varepsilon > 0$, $\{x\} \subset \left[x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2}\right]$, so if we assume m to be *monotone* (i.e. $A \subset B$ implies $m(A) \leq m(B)$), then we must conclude that $m(\{x\}) = 0$. On the other hand, since *some* number between 0 and 1 *is* chosen, it is *not impossible* that it could be our x. Thus a *non-empty* set A can have $m(A) = 0$.

The probability that any one of a *countable* set of reals $A = \{x_1, x_2, \ldots, x_n, \ldots\}$ is selected should also be 0, since for any $\varepsilon > 0$ we can *cover* each x_n by an interval $I_n = \left[x_n - \frac{\varepsilon}{2^{n+2}}, x_n + \frac{\varepsilon}{2^{n+2}}\right]$ so that $A \subset \cup_{n=1}^{\infty} I_n$ with *total length* $\Sigma_{n=1}^{\infty} m(I_n) < \varepsilon$. We just need the 'obvious' property that $m(A) = \sum_{n=1}^{\infty} m(\{x_n\})$ for our conclusion.

We generalise this to demand the *countable additivity property* of *any* probability function $A \mapsto P(A)$, i.e. if $(A_n)_{n\geq 1}$ are disjoint, then $P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$. We shall formalise this in Definition 1.9.

This demand looks very reasonable, and is an essential feature of the calculus of probabilities. It implies *finite* additivity: if $A_1, \ldots, A_n$ $(n \in \mathbb{N})$ are disjoint, then $P\left(\cup_{i=1}^{n} A_i\right) = \Sigma_{i=1}^{n} P(A_i)$. Simply let $A_i = \emptyset$ for $i > n$ (see Proposition 1.3).

Remark 1.1 Our example suggests a useful description of the 'negligible' (or *null*) sets in $[0,1]$ (and by the same token in $\mathbb{R}$) for Lebesgue measure m: the set A is m-*null* if for every $\varepsilon > 0$ there is a sequence $(I_n)_{n\geq 1}$ of intervals of total length $\sum_{n=1}^{\infty} m(I_n) < \varepsilon$, so that $A \subset \cup_{n=1}^{\infty} I_n$. (Note that the I_n need not be disjoint.) This requirement will characterise sets $A \subset \mathbb{R}$ with Lebesgue measure $m(A) = 0$.

Example 1.2 The *Cantor set* provides an uncountable m-null set in $[0,1]$. Start with the interval $[0,1]$, remove the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$, obtaining the set C_1, which consists of the two intervals $\left[0, \frac{1}{3}\right]$ and $\left[\frac{2}{3}, 1\right]$. Next remove the 'middle thirds' $\left(\frac{1}{9}, \frac{2}{9}\right)$, $\left(\frac{7}{9}, \frac{8}{9}\right)$ of these two intervals, leaving

C_2, consisting of four intervals each of length $\frac{1}{9}$, etc. At the nth stage we have a set C_n, consisting of 2^n disjoint closed intervals, each of length $\frac{1}{3^n}$. Thus the total length of C_n is $\left(\frac{2}{3}\right)^n$, which goes to 0 as $n \to \infty$.

We call $C = \bigcap_{n=1}^{\infty} C_n$ the *Cantor set*, which is contained in each C_n, hence is m-null. Using ternary expansions, you may now show that C is uncountable (just as with decimal expansions for $[0,1]$).

We develop some abstract probability theory. Any given set Ω can serve as *sample space*, and we consider ('future') *events* A, B from a given class $\mathcal{A}$ of subsets of Ω. We wish to define the probability $P(A)$ (resp. $P(B)$) as numbers in $[0,1]$. Clearly, we would then also wish to know $P(A \cup B)$, $P(A^c)$, $P(A \cap B)$, etc. Thus the class $\mathcal{A}$ of sets on which P is defined should contain Ω (and $P(\Omega) = 1$) and together with A, B it should also contain $A \cup B$ and A^c. This ensures that $\mathcal{A}$ also contains $A \cap B$: $A^c \cup B^c = (A \cap B)^c$ is in $\mathcal{A}$, hence also $A \cap B$. Such a class $\mathcal{A}$ is a *field.* (This, now standard, use of the term 'field' in probability theory is somewhat unfortunate, and invites confusion with its usual algebraic meaning. Some authors seek to avoid this by using the term 'algebra' instead. We shall not do so.)

We demand that P is *additive,* i.e. for disjoint $A, B \in \mathcal{A}$ we have $P(A \cup B) = P(A) + P(B)$. This suffices for our first result.

Proposition 1.3 *If $A, B \in \mathcal{A}$ and $A \subset B$, then we have $P(B \backslash A) = P(B) - P(A)$. Hence $P(\emptyset) = 0$. Moreover, P is monotone: $A \subset B$ implies $P(A) \leq P(B)$.*

Proof $B \backslash A = B \cap A^c$, so $B \backslash A$ is in $\mathcal{A}$. But $B = A \cup (B \backslash A)$ and these sets are disjoint. Hence $P(B) = P(A) + P(B \backslash A)$. For the second claim, use $B = A$. The final claim follows as P is non-negative.

Exercise 1.4 Show that for any A, B in $\mathcal{A}$ (disjoint or not)

$$P(A \cup B) + P(A \cap B) = P(A) + P(B).$$

Given any probability P (see Definition 1.9 below), we call an *event* A *P-null* if $P(A) = 0$. An event B is called *almost sure* (or *full*) if $P(B) = 1$ (so that B^c is P-null, since $P(B) + P(B^c) = P(\Omega)$). A property (e.g. of some function) holds *almost surely* if it holds on a full set (i.e. except possibly on some null set).

Example 1.5 (i) Let $\Omega = \{1, 2, \ldots, n\}$ (or any finite set) $\mathcal{A} = 2^\Omega$ its *power set* (the class of all subsets) and $P(A) = \frac{\#A}{n}$, where $\#A$ is the number of points in A.

(ii) Let Ω be any set, $\mathcal{A} = 2^\Omega$, $\delta_x(A) = 1$ for $x \in A$, else 0. Then $P = \delta_x$ is the *point mass at* x (also called a *Dirac* δ*-measure*). Of course, the power set is always a field.

(iii) For finite or countable sample spaces, probability distributions can be built from Dirac measures: call a probability P on $\mathbb{R}$ *discrete* if there is a countable full subset C (i.e. $P(C) = 1$). This is obviously equivalent to P having the form $P = \sum_{i=1}^{\infty} p_i \delta_{x_i}$ for some real sequences $(x_i)_{i\geq 1}, (p_i)_{i\geq 1}$ with $p_i > 0$ and $\sum_{i=1}^{\infty} p_i = 1$.

The following distributions should be familiar:

(a) *Bernoulli*: $P = p\delta_1 + (1-p)\delta_0, 0 < p < 1$.

(b) *Binomial Bi(n,p)*: $P = \sum_{i=1}^{n} p_i \delta_i$,
where $p_i = \binom{n}{i} p^i (1-p)^{n-i}, 0 \leq i \leq n, 0 < p < 1$.

(c) *Geometric Geo(p)*: $P = \sum_{i=1}^{\infty} p_i \delta_i$,
where $p_i = p(1-p)^{i-1}, i \geq 1, 0 < p < 1$.

(d) *Negative binomial NegB(n,p)*: $P = \sum_{i=n}^{\infty} p_i \delta_i$,
where $p_i = \binom{i-1}{n-1} p^i (1-p)^{i-n}, i \geq n, 0 < p < 1$.

(e) *Poisson Po(λ)*: $P = \sum_{i=1}^{\infty} p_i \delta_i$,
where $\lambda > 0$ and $p_i = e^{-\lambda} \frac{\lambda^k}{i!}, i \geq 0$.

You may know these better as distributions of well-known classes of *random variables*.

Example 1.6 For a different example, we consider a field that enables us to generate Lebesgue measure on $\mathbb{R}$.

Let $\Omega = \mathbb{R}$. Consider left-open, right-closed intervals, i.e. of the form $(a, b]$ for $a, b \in [-\infty, \infty]$ (the set of *extended reals*, which consists of $\mathbb{R} \cup \{-\infty, \infty\}$) and where by convention we set $(a, \infty] = (a, \infty)$ for $-\infty \leq a \leq \infty$. We then define

$$\mathcal{A}_0 = \{\cup_{i=1}^{n} (a_i, b_i] : a_1 \leq b_1 \leq a_2 \leq \ldots \leq b_n, \ \ n \geq 1\},$$

so that $\mathcal{A}_0$ is the class of all finite disjoint unions of such intervals. We define the measure of such a union as

$$m\left(\cup_{i=1}^{n} (a_i, b_i]\right) = \sum_{i=1}^{n} (b_i - a_i).$$

This fits with our earlier informal definition for closed intervals, since we showed that $m(\{x\}) = 0$ for any x and m is finitely additive.

Exercise 1.7 Verify that $\mathcal{A}_0$ is a field. Would this remain true if we had used open (or closed) intervals instead?

In practice, as can already be seen from examples (d) and (e) above, we are driven to considering unions and intersections of an *infinite sequence of events*. As another example, in an infinite sequence of coin tosses, what is the probability that 'heads' will occur infinitely often? We shall see that this depends crucially on our assumptions about the probability of success at each stage.

The *Borel–Cantelli (BC) Lemmas* are the archetype of this sort of result. To formulate the first lemma, suppose that $A_1, A_2, \ldots, A_n, \ldots$ is a sequence in $\mathcal{A}$ with $\sum_{n=1}^{\infty} P(A_n) < \infty$. How should we find the probability of the event

$$A_{\text{i.o.}} = \{\omega : \omega \in A_n \text{ for infinitely many } n\}?$$

(As with $A_{\text{i.o.}}$, we shall use 'i.o.' as an abbreviation for 'infinitely often' throughout.) We must ensure that $P(A_{\text{i.o.}})$ makes sense. If ω belongs to infinitely many A_n, then for each $m \geq 1$ there is at least one $n \geq m$ with $\omega \in A_n$. So $\omega \in \cup_{n \geq m} A_n$ for all m. Thus we need to define the probability of the union of infinitely many A_n if we are to get further. This leads first to:

Definition 1.8 A class $\mathcal{F}$ of subsets of a given set Ω is a σ-field (use of the term σ-algebra is also common) of subsets of Ω if:
(i) $\Omega \in \mathcal{F}$.
(ii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$.
(iii) $\{A_n : n \in \mathcal{N}\} \subset \mathcal{F}$ implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Thus $\mathcal{F}$ is closed under complements and countable unions.

A field, and indeed any family $\mathcal{A}$ of subsets of Ω, *generates* a minimal σ-field $\sigma(\mathcal{A})$: we define

$$\sigma(\mathcal{A}) = \cap\{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-field}, \mathcal{A} \subset \mathcal{G}\}.$$

It is easily verified that $\sigma(\mathcal{A})$ satisfies (i)–(iii) of Definition 1.8 A key example is given by the *Borel* σ-field $\mathcal{B} = \sigma(\mathcal{A}_0)$ with the field $\mathcal{A}_0$

defined as in Example 1.6 above. In what follows, the pair $(\mathbb{R}, \mathcal{B})$ plays a central role.

For $\omega \in A_{\text{i.o.}}$, we need to have $\omega \in \cup_{n \geq m} A_n$ for *each* $m \geq 1$, so that we write

$$A_{\text{i.o.}} = \cap_{m \geq 1}(\cup_{n \geq m} A_n) := \limsup_{n \to \infty} A_n.$$

To see that $A_{\text{i.o.}}$ belongs to the σ-field $\mathcal{F}$, we show that $\mathcal{F}$ is also *closed under countable intersections.* Since $\mathcal{F}$ is closed under complements, if $(B_n)_{n \geq 1} \subset \mathcal{F}$ then each $B_n^c \in \mathcal{F}$, and by de Morgan's laws we have $(\cap_{n \geq 1} B_n)^c = \cup_{n \geq 1} B_n^c$, so that $\cap_{n \geq 1} B_n$ is the complement of a set in $\mathcal{F}$, hence is itself in $\mathcal{F}$. Thus we see that $A_{\text{i.o.}}$ is well-defined as soon as the sets A_n belong to a σ-field of sets in Ω.

But this still does not tell us how to find its probability.

Definition 1.9 A triple $(\Omega, \mathcal{F}, P)$ is a probability space if Ω is any set, $\mathcal{F}$ is a σ-field of subsets of Ω and the function $P : \mathcal{F} \to [0, 1]$ satisfies:
(i) $P(\Omega) = 1$.
(ii) $\{A_n : n \in N\} \subset \mathcal{F}$ and $A_n \cap A_m = \emptyset$ for $n \neq m$ imply that

$$P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

We say that the *probability* P is a *σ-additive* (also called *countably additive*) set function.

We may equally define $\liminf_{n \to \infty} A_n = \cup_{n \geq 1}(\cap A_{m \geq n})$. This set contains all points that *eventually* belong to sets in the sequence $(A_n)_n$. We also write A_{ev} for this set.

Exercise 1.10 Check: $(\liminf_{n \to \infty} A_n)^c = \limsup_{n \to \infty} A_n^c$. Compare this with a similar result for the upper and lower limits of a sequence of real numbers – recall that for a real sequence (a_n) we define $\limsup_n a_n$ as $\inf_{n \geq 1}(\sup_{m \geq n} a_m)$ and $\liminf_n a_n$ as $\sup_{n \geq 1}(\inf_{m \geq n} a_m)$. You should prove that (a_n) converges if and only if (iff) these quantities coincide! We use this fact in later chapters.

Any countable union $\bigcup_{k=1}^{\infty} A_k$ can be written as a *disjoint* union: let $B_k = A_k \backslash \bigcup_{j=1}^{k-1} A_j$, then clearly $B_k \subset A_k$ for each k and $B_j \cap B_k = \emptyset$ when $j \neq k$. You should verify that $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$.

Exercise 1.11 Prove that a finitely additive set function $P : \mathcal{F} \to [0, 1]$ is σ-additive iff the following statement holds: whenever $(B_n)_{n\geq 1}$ in $\mathcal{F}$ decreases to the empty set ($\cap_{n\geq 1} B_n = \emptyset$), then $\lim_{n\to\infty} P(B_n) = 0$. (This is often called '*continuity*' of P at $\emptyset$.)

To find $P(\limsup_{n\to\infty} A_n)$, we first need some more simple consequences of the definitions:

Proposition 1.12 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space.*
(i) If $(A_i)_{i\geq 1}$ *in* $\mathcal{F}$*, then* $P(\cup_{i\geq 1} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.
(ii) If $A_i \subset A_{i+1}$, $(i \geq 1)$, *then* $P(\cup_{i\geq 1} A_i) = \lim_{n\to\infty} P(A_n)$.
(iii) If $A_{i+1} \subset A_i$, $(i \geq 1)$, *then* $P(\cap_{i\geq 1} A_i) = \lim_{n\to\infty} P(A_n)$.

Exercise 1.13 Prove Proposition 1.12 and show that (ii) and (iii) are equivalent ways of formulating σ-additivity of additive P. (Here (i) says that P is *countably subadditive*.)

We introduce notation for the *convergence of sets*: write $A_n \to A$ if $\limsup_n A_n = \liminf_n A_n = A$ (alternatively, $A_{\text{i.o.}} = A_{\text{ev}} = A$). Note the analogy with convergent sequences! As a special case, we write $A_n \uparrow A$ if $A_n \subset A_{n+1}$ for all n and $A = \cup_{n=1}^{\infty} A_n$. Similarly, $A_n \downarrow A$ if $A_{n+1} \subset A_n$ for all n and $A = \cap_{n=1}^{\infty} A_n$.

Proposition 1.14 *If* $A_n \to A$, *then* $P(A_n) \to P(A)$.

Proof If $A_n \to A$, then $A = A_{\text{i.o.}} = A_{\text{ev}}$, so $P(A) = P(A_{\text{i.o.}}) = P(A_{\text{ev}})$. We need to show that

$$\limsup_n P(A_n) = \liminf_n P(A_n) = P(A).$$

But $\liminf_n P(A_n) \leq \limsup_n P(A_n)$ always holds, so we just need to show that $P(\liminf_n (A_n)) \leq \liminf_n P(A_n)$. Now for each $k \geq 1$

$$\cap_{n\geq k} A_n \subset A_k, \quad \text{hence } P(\cap_{n\geq k} A_n) \leq P(A_k)$$

and the result follows by Proposition 1.12 (ii) on letting $k \to \infty$, since $\cap_{n\geq k} A_n \uparrow \liminf_n A_k = A$.

Exercise 1.15 Show that $\limsup_n P(A_n) \leq P(\limsup_n A_n)$.

For the first BC Lemma, let $B_m = \cup_{n \geq m} A_n$ so that (B_m) decreases, hence by 1.12(ii) and 1.12(i)

$$P(\cap_{m \geq 1} B_m) = \lim_m P(B_m) = \lim_m P(A_m \cup A_{m+1} \cup \ldots)$$
$$\leq \lim_m (P(A_m) + P(A_{m+1}) + \ldots) = 0$$

since the series (of real numbers!) $\sum_{n=1}^{\infty} P(A_n)$ converges.

We have proved:

Lemma 1.16 *(First Borel–Cantelli (BC) Lemma): If $(A_n)_n$ is a sequence of events with $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\limsup_n A_n) = 0$.*

Thus, if we have a sequence of events whose probabilities decrease quickly enough to keep their sum finite (for example, if $P(A_{n+1}) = 0.999P(A_n)$ for each n), then it is *certain* (i.e. the probability is 1) that only finitely many of them will occur. This may not be unduly surprising, but it did need a proof.

1.3 Independent events

The first BC Lemma is immediate from our definitions. Matters are very different, however, when the series $\sum_{n=1}^{\infty} P(A_n)$ diverges. Then we do have a *second BC Lemma*, but this applies only when the sequence of events $(A_n)_n$ is *independent*. Recall some basic definitions:

Definition 1.17 Let $(\Omega, \mathcal{F}, P)$ be a probability space. For $A, B \in \mathcal{F}$ with $P(B) > 0$, define

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

as the *conditional probability* of A, *given* B.

Exercise 1.18 Verify that the function $P_B : A \to P(A|B)$ is again a probability. (*Hint*: if $(A_n)_n$ are pairwise disjoint sets in $\mathcal{F}$, then $(A_n \cap B)_n$ are also pairwise disjoint.)

Exercise 1.19 Suppose that $A, B_n \in \mathcal{F}$ with $(B_n)_n$ pairwise disjoint, $P(B_n) \neq 0$ for all n and $\cup_{n=1}^{\infty} B_n = \Omega$. Prove that

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n).$$

This is often called the *Theorem of Total Probability.*

Definition 1.20 Events A, B in $\mathcal{F}$ are *independent* if

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

When $P(B) > 0$, this is the same as the more natural requirement that B should 'have no influence' on A; i.e. $P(A|B) = P(A)$. However, our definition still makes sense if $P(B) = 0$.

Care must be taken when generalising this definition to three or more sets (e.g. A, B, C): it is not enough simply to require that $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Exercise 1.21 Find examples of sets in $\mathbb{R}$ to justify this claim.

As our general definition of independence of events, we therefore require:

Definition 1.22 Let $(\Omega, \mathcal{F}, P)$ be a probability space. Events $A_1, A_2, \ldots, A_n$ in $\mathcal{F}$ are *independent* if for each choice of indices $1 \leq i_1 < i_2 < \ldots < i_k \leq n$

$$P\left(\cap_{m=1}^{k} A_{i_m}\right) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}) = \prod_{m=1}^{k} P(A_{i_m}).$$

A sequence $(A_n)_{n \geq 1}$ of events (or any family $(A_\alpha)_\alpha$) is *independent* if every finite subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ of events is independent.

With this machinery we formulate:

Lemma 1.23 *(Second Borel–Cantelli (BC) Lemma): If the sequence $(A_n)_{n \geq 1}$ is independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(\limsup_n A_n) = 1$.*

Proof To prove that $P\left(\cap_{k=1}^{\infty}\left(\cup_{n=k}^{\infty} A_n\right)\right) = 1$ it will suffice to show that for each $k \geq 1$, $P\left(\cup_{n=k}^{\infty} A_n\right) = 1$. This follows from Proposition 1.12 (iii): $\cup_{n=k}^{\infty} A_n$ decreases as k increases, so that

$$\lim_{k\to\infty} P\left(\cup_{n=k}^{\infty} A_n\right) = P\left(\cap_{k=1}^{\infty}(\cup_{n=k}^{\infty} A_n)\right).$$

Now consider $\cap_{n=k}^{m} A_n^c$ for a fixed $m > k$. By de Morgan's laws, we have $\left(\cup_{n=k}^{m} A_n\right)^c = \cap_{n=k}^{m} A_n^c$. The (A_n^c) are also independent (check this yourself!), so for $k \geq 1$

$$P\left(\cap_{n=k}^{m} A_n^c\right) = \prod_{n=k}^{m} P\left(A_n^c\right) = \prod_{n=k}^{m}[1 - P(A_n)].$$

For $x \geq 0$, we have $1 - x \leq e^{-x}$ (use the Taylor series, or simply the derivative), so

$$\prod_{n=k}^{m}[1 - P(A_n)] \leq \prod_{n=k}^{m} e^{-P(A_n)} = e^{-\sum_{n=k}^{m} P(A_n)}.$$

Now recall that we have assumed that the series $\sum_n P(A_n)$ diverges. Hence for fixed k the partial sums $\sum_{n=k}^{m} P(A_n)$ grow beyond all bounds as $m \to \infty$. So, as $m \to \infty$ the right-hand side (RHS) of the inequality becomes arbitrarily small.

This proves that

$$1 - P\left(\cup_{n=k}^{m} A_n\right) = P\left(\cap_{n=k}^{m} A_n^c\right) \to 0 \text{ as } m \to \infty.$$

Now write $B_m = \cup_{n=k}^{m} A_n$. The sequence $(B_m)_m$ is increasing in m and its union is $\cup_{n=k}^{\infty} A_n$. Applying Proposition 1.12 (ii), we have

$$P\left(\cup_{n=k}^{\infty} A_n\right) = \lim_{m\to\infty} P(B_m) = 1$$

and the proof is complete.

1.4 Simple random walk

The following famous example illustrates the power of the BC Lemmas. (i) On the line, imagine a drunkard describing a symmetric random walk from 0, i.e. who at each step is equally likely to go left or right. How often does such a walk return to the starting point? The position reached