Lecture 1: Revision of probability concepts, and the simple random walk

Susana Gomes

October 3^{rd}, 2023

Plan for today

1. Short intro to the topics in this module
2. Revision of some basic probability definitions
3. Basics of random variables
4. The simple random walk, LLN and CLT

Introduction

Introduction

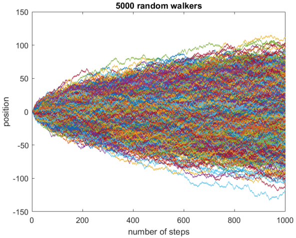
This module will be an overview of (some of) the most common concepts in modelling real world systems.

Introduction

This module will be an overview of (some of) the most common concepts in modelling real world systems.

The first half of the module will broadly cover “stochastic modelling", and it will focus on **Markov chains**, more general stochastic processes (including **diffusion processes**, Brownian motion, etc) and interacting particle systems.

Right: A lot of random walkers

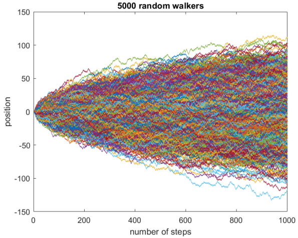


Introduction

This module will be an overview of (some of) the most common concepts in modelling real world systems.

The first half of the module will broadly cover “stochastic modelling", and it will focus on **Markov chains**, more general stochastic processes (including **diffusion processes**, Brownian motion, etc) and interacting particle systems.

Right: A lot of random walkers



After this, we will move on to **graphs** and **networks**. We will explore some basic graph definitions and properties, random graphs, percolation, and application of all this to networks.

Right: A network, taken from wikipedia

Basic probability definitions

Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

- The **sample space** Ω is the space where events we want to study live.

  e.g., if you roll a die twice, Ω is the set of all pairs of numbers between 1 and 6 ($\{(1,1),(1,2),\ldots,(6,6)\}$).

Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

- The **sample space** Ω is the space where events we want to study live.

  e.g., if you roll a die twice, Ω *is the set of all pairs of numbers between 1 and 6 (*$\{(1,1),(1,2),\ldots,(6,6)\}$*).*

- An **event** ω is a realisation of one of the members of the sample space.

  e.g. $(1,6)$.

Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

- The **sample space** Ω is the space where events we want to study live.

  e.g., if you roll a die twice, Ω *is the set of all pairs of numbers between 1 and 6 (*$\{(1,1),(1,2),\ldots,(6,6)\}$*).*

- An **event** ω is a realisation of one of the members of the sample space.

  e.g. $(1,6)$.

- In practice, we work with **sets of events** A and we consider $\omega \in A$. A is always a (measurable) subset of Ω.

  e.g., "the two rolls add up to 8*",* $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$.

Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

- The **sample space** Ω is the space where events we want to study live.

  e.g., if you roll a die twice, Ω *is the set of all pairs of numbers between 1 and 6 (*$\{(1,1),(1,2),\ldots,(6,6)\}$*).*

- An **event** ω is a realisation of one of the members of the sample space.

  e.g. $(1,6)$.

- In practice, we work with **sets of events** A and we consider $\omega \in A$. A is always a (measurable) subset of Ω.

  e.g., "the two rolls add up to 8*",* $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$*.*

- The **set of all events** $\mathcal{F} \subset \mathcal{P}(\Omega)$ is a subset of the powerset $\mathcal{P}(\Omega)$.

  In order to be able to define probability, we need $\mathcal{F}$ *to be a closed system, i.e., it needs to be a* σ***-algebra.***

σ-algebras

Let Ω be a set, and let $\mathcal{P}(\Omega)$ be its power set. We say that $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a **σ-algebra** if it has the following properties:

1. $\mathcal{F}$ is closed under the complement operation:

$$A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}.$$

2. $\Omega \in \mathcal{F}$.

Note that together with 1., this means that the set of no events $\varnothing$ is in $\mathcal{F}$ since $\varnothing = \Omega^C$.

3. $\mathcal{F}$ is closed under *countable* union:

$$A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Probability

We can now define a probability distribution.

Probability

We can now define a probability distribution.

Probability distributions

A **probability distribution** $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function which satisfies the following properties:

(i) it is positive, i.e. $\mathbb{P}(A) \geq 0$.

(ii) it is normalised, i.e. $\quad \mathbb{P}(\emptyset) = 0 \quad$ and $\quad \mathbb{P}(\Omega) = 1$.

(iii) it is additive, i.e. $\quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$,

where $A_1, A_2, \ldots$ is a collection of disjoint events, i.e. $A_i \cap A_j = \emptyset$, $\forall i, j$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Probability

We can now define a probability distribution.

Probability distributions

A **probability distribution** $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function which satisfies the following properties:

(i) it is positive, i.e. $\mathbb{P}(A) \geq 0$.

(ii) it is normalised, i.e. $\quad \mathbb{P}(\emptyset) = 0 \quad$ and $\quad \mathbb{P}(\Omega) = 1$.

(iii) it is additive, i.e. $\quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$,

where $A_1, A_2, \ldots$ is a collection of disjoint events, i.e. $A_i \cap A_j = \emptyset$, $\forall i, j$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Note that (i) and (ii) mean that we have that $\mathbb{P} : \mathcal{F} \to [0, 1]$.

Some relevant properties to note

- From the probability axioms + properties of σ-algebras, we have

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A), \qquad \text{and} \qquad \mathbb{P}(\varnothing) = 0.$$

Some relevant properties to note

- From the probability axioms + properties of σ-algebras, we have

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A), \qquad \text{and} \qquad \mathbb{P}(\emptyset) = 0.$$

- If Ω is a **discrete set**, we always have

$$\mathcal{F} = \mathcal{P}(\Omega) \qquad \textbf{and} \qquad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

e.g. $\mathbb{P}(\textit{the two rolls add up to } 8) =$
$\mathbb{P}((2,6)) + \mathbb{P}((3,5)) + \mathbb{P}((4,4)) + \mathbb{P}((5,3)) + \mathbb{P}((6,2)) = \frac{5}{36}$.

Some relevant properties to note

- From the probability axioms + properties of σ-algebras, we have

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A), \qquad \text{and} \qquad \mathbb{P}(\varnothing) = 0.$$

- If Ω is a **discrete set**, we always have

$$\mathcal{F} = \mathcal{P}(\Omega) \qquad \textbf{and} \qquad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

  e.g. $\mathbb{P}(\textit{the two rolls add up to } 8) =$
  $\mathbb{P}((2,6)) + \mathbb{P}((3,5)) + \mathbb{P}((4,4)) + \mathbb{P}((5,3)) + \mathbb{P}((6,2)) = \frac{5}{36}$.

- If Ω is **continuous** (e.g. $[0,1]$), then we have

$$\mathcal{F} \subsetneq \mathcal{P}(\Omega).$$

Independence and conditional probability

- We say that two events $A, B \subset \Omega$ are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

  e.g. rolling a die repeatedly.

Independence and conditional probability

- We say that two events $A, B \subset \Omega$ are **independent** if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$
e.g. rolling a die repeatedly.
- If $\mathbb{P}(B) > 0$ then we define the **conditional probability** of A given B as
$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Independence and conditional probability

- We say that two events $A, B \subset \Omega$ are **independent** if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$
e.g. rolling a die repeatedly.
- If $\mathbb{P}(B) > 0$ then we define the **conditional probability** of A given B as
$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$
If A and B are independent, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Independence and conditional probability

- We say that two events $A, B \subset \Omega$ are **independent** if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$
e.g. rolling a die repeatedly.
- If $\mathbb{P}(B) > 0$ then we define the **conditional probability** of A given B as
$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$
If A and B are independent, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

A useful result related to conditional probability is the Law of total probability:

Lemma: Law of total probability

Let $B_1, \ldots, B_n$ be a **partition** of Ω such that $\mathbb{P}(B_i) > 0$ for all i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

Independence and conditional probability

- We say that two events $A, B \subset \Omega$ are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

  e.g. rolling a die repeatedly.

- If $\mathbb{P}(B) > 0$ then we define the **conditional probability** of A given B as

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

  If A and B are independent, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

A useful result related to conditional probability is the Law of total probability:

Lemma: Law of total probability

Let $B_1, \ldots, B_n$ be a **partition** of Ω such that $\mathbb{P}(B_i) > 0$ for all i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

Note also that

$$\mathbb{P}(A|C) = \sum_{i=1}^{n} \mathbb{P}(A|C \cap B_i)\,\mathbb{P}(B_i|C) \quad \text{provided} \quad \mathbb{P}[C] > 0.$$

Random variables

Random variables (definitions)

A **random variable** (rv) X is a (measurable) function $X : \Omega \to \mathbb{R}$.

Random variables (definitions)

A **random variable** (rv) X is a (measurable) function $X : \Omega \to \mathbb{R}$.

The **distribution function** of the random variable X is

$$F(x) = \mathbb{P}[X \leq x] = \mathbb{P}\big[\{\omega : X(\omega) \leq x\}\big].$$

Random variables (definitions)

A **random variable** (rv) X is a (measurable) function $X : \Omega \to \mathbb{R}$.

The **distribution function** of the random variable X is

$$F(x) = \mathbb{P}[X \leq x] = \mathbb{P}\big[\{\omega : X(\omega) \leq x\}\big].$$

- The rv X is called **discrete** if it only takes values in a countable subset

  $$\Delta = \{x_1, x_2, \ldots\} \subset \mathbb{R},$$

  and its distribution is characterised by the **probability mass function**

  $$\pi(x) := \mathbb{P}[X = x] , \quad x \in \Delta.$$

Random variables (definitions)

A **random variable** (rv) X is a (measurable) function $X : \Omega \to \mathbb{R}$.

The **distribution function** of the random variable X is

$$F(x) = \mathbb{P}[X \leq x] = \mathbb{P}\big[\{\omega : X(\omega) \leq x\}\big].$$

- The rv X is called **discrete** if it only takes values in a countable subset

  $$\Delta = \{x_1, x_2, \ldots\} \subset \mathbb{R},$$

  and its distribution is characterised by the **probability mass function**

  $$\pi(x) := \mathbb{P}[X = x]\ , \quad x \in \Delta.$$

- The rv X is called **continuous**, if its distribution function is

  $$F(x) = \int_{-\infty}^{x} f(y)\, dy \quad \text{for all } x \in \mathbb{R}$$

  where $f : \mathbb{R} \to [0, \infty)$ is the **probability density function (PDF)** of X.

So... What does this mean?

In general, $f = F'$ is given by the derivative of the distribution function (which always exists for continuous rv's).

For discrete rv's, F is a step function with 'PDF'

$$f(x) = F'(x) = \sum_{y \in \Delta} \pi(y) \delta(x - y).$$

Important concepts (1/2)

There are many ways to describe random variables or events, and we will focus on the usual ones: mean, variance and covariance.

Important concepts (1/2)

There are many ways to describe random variables or events, and we will focus on the usual ones: mean, variance and covariance.

- The **expected value** of X is given by

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \Delta} x\pi(x), & \text{if X is discrete} \\ \int_{\mathbb{R}} x\, f(x)\, dx, & \text{if X is continuous} \end{cases}$$

Important concepts (1/2)

There are many ways to describe random variables or events, and we will focus on the usual ones: mean, variance and covariance.

- The **expected value** of X is given by

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \Delta} x\pi(x), & \text{if X is discrete} \\ \int_{\mathbb{R}} x\, f(x)\, dx, & \text{if X is continuous} \end{cases}$$

- The **variance** is given by $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Important concepts (1/2)

There are many ways to describe random variables or events, and we will focus on the usual ones: mean, variance and covariance.

- The **expected value** of X is given by

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \Delta} x\pi(x), & \text{if X is discrete} \\ \int_{\mathbb{R}} x\, f(x)\, dx, & \text{if X is continuous} \end{cases}$$

- The **variance** is given by $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- And the **covariance** of two rv's by $\mathrm{Cov}[X, Y] := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Important concepts (2/2)

Two random variables X, Y are **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$.

Important concepts (2/2)

Two random variables X, Y are **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$.

This justifies the definition of **joint distributions**:

$$f(x, y) = f^X(x)\, f^Y(y) \quad \text{or} \quad \pi(x, y) = \pi^X(x)\, \pi^Y(y)$$

and their **marginals** $\quad f^X(x) = \int_{\mathbb{R}} f(x, y)\, dy \quad$ and $\quad \pi^X(x) = \sum_{y \in \Delta_y} \pi(x, y)$.

Important concepts (2/2)

Two random variables X, Y are **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$.

This justifies the definition of **joint distributions**:

$$f(x,y) = f^X(x)\, f^Y(y) \quad \text{or} \quad \pi(x,y) = \pi^X(x)\, \pi^Y(y)$$

and their **marginals** $\quad f^X(x) = \int_{\mathbb{R}} f(x,y)\, dy \quad$ and $\quad \pi^X(x) = \sum_{y \in \Delta_y} \pi(x,y)$.

Independence implies $\mathrm{Cov}[X, Y] = 0$, i.e. X and Y are **uncorrelated**.

Important concepts (2/2)

Two random variables X, Y are **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$.

This justifies the definition of **joint distributions**:

$$f(x,y) = f^X(x)\, f^Y(y) \quad \text{or} \quad \pi(x,y) = \pi^X(x)\, \pi^Y(y)$$

and their **marginals** $\quad f^X(x) = \int_{\mathbb{R}} f(x,y)\, dy \quad$ and $\quad \pi^X(x) = \sum_{y \in \Delta_y} \pi(x,y)$.

Independence implies $\mathrm{Cov}[X, Y] = 0$, i.e. X and Y are **uncorrelated**.

The inverse is in general false, but holds if X and Y are Gaussian.

The simple random walk

Simple random walk

Let $X_1, X_2, \dots \in \{-1, 1\}$ be a sequence of independent, identically distributed random variables **(iid rv's)** with

$$p = \mathbb{P}(X_i = 1) \quad \text{and} \quad q = \mathbb{P}(X_i = -1) = 1 - p.$$

The sequence $Y_0, Y_1, \dots$ defined as $Y_0 = 0$ and $Y_n = \sum_{k=1}^{n} X_k$ is called the **simple random walk (SRW)** on $\mathbb{Z}$.

So... What happens when $n \to \infty$?

Turns out we can predict what a simple random walk will do (and we will work on expansions on this) using the Law of Large Numbers:

So... What happens when $n \to \infty$?

Turns out we can predict what a simple random walk will do (and we will work on expansions on this) using the Law of Large Numbers:

Weak law of large numbers (LLN)

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iid rv's with $\mu := \mathbb{E}(X_k) < \infty$ and $\mathbb{E}(|X_k|) < \infty$. Then

$$\frac{1}{n} Y_n = \frac{1}{n} \sum_{k=1}^{n} X_k \to \mu \quad \text{as } n \to \infty$$

in distribution (i.e. the **distribution function** of Y_n converges to $\mathbb{1}_{[\mu,\infty)}(x)$ for $x \neq \mu$).

And...

And in some (most!) situations we can even say more...

And...

And in some (most!) situations we can even say more...

Central limit theorem (CLT)

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iid rv's with $\mu := \mathbb{E}(X_k) < \infty$ and $\sigma^2 := \mathrm{Var}(X_k) < \infty$. Then

$$\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \to \xi \quad \text{as } n \to \infty$$

in distribution, where $\xi \sim N(0,1)$ is a **standard Gaussian** with PDF

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

And...

And in some (most!) situations we can even say more...

Central limit theorem (CLT)

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iid rv's with $\mu := \mathbb{E}(X_k) < \infty$ and $\sigma^2 := \mathrm{Var}(X_k) < \infty$. Then

$$\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \to \xi \quad \text{as } n \to \infty$$

in distribution, where $\xi \sim N(0,1)$ is a **standard Gaussian** with PDF

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

In fact, we can say that, as $n \to \infty$,

$$Y_n = \sum_{k=1}^{n} X_k = n\mu + \sqrt{n}\sigma\xi + o(\sqrt{n}) , \quad \xi \sim N(0,1).$$