MAST30001 Stochastic Modelling

Lecturer: Mark Holmes

Stochastic modelling

We study a random experiment in the context of a Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$. Here,

- Ω is the sample space - the set of all possible outcomes of our random experiment,
- $\mathcal{F}$ is a sigma-field - a collection (with certain properties) of subsets of Ω. We view these as events we can *see* or *measure*,
- $\mathbb{P}$ is a probability measure - a function (with certain properties) defined on the elements of $\mathcal{F}$ with certain properties.

The sample space Ω

The (nonempty) set of possible outcomes for the random experiment. Examples:

- coin tossing: $\{H, T\}$, $\{(H,H),(H,T),(T,H),(T,T)\}$, the set $\{H, T\}^{\mathbb{N}}$ of all infinite sequences of *H*s and *T*s.
- battery lifetime: $[0, \infty)$.
- animal population: $\mathbb{Z}_+$.
- queue length over time: the set of piecewise-constant functions from $[0, \infty)$ to $\mathbb{Z}_+$.
- infection status: $\{\textit{susceptible}, \textit{infected}, \textit{immune}\}^n$ (if n is the number of individuals).
- facebook friend network: set of simple networks with number of vertices equal to the number of users: edges connect friends.

Elements of Ω are often written as ω.

Review of basic notions of set theory

- $A \subset B$.
  - A is a subset of B, or if A occurs then B occurs.
- $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\} = B \cup A$.
  - Union of sets (events): at least one occurs.
  - $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$.
- $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\} = B \cap A$.
  - Intersection of sets (events): all occur.
  - $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$.
- $A^c = \{\omega \in \Omega : \omega \notin A\}$.
  - Complement of a set/event: event doesn't occur.
- $\varnothing$: the empty set or impossible event.
- A and B are disjoint (or mutually exclusive) if $A \cap B = \varnothing$.

The *sigma-field* $\mathcal{F}$

A sigma-field $\mathcal{F}$ on Ω is a collection of subsets of Ω such that

- $\Omega \in \mathcal{F}$, and
- if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and
- if $A_1, A_2, \dots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$
- For countable sample spaces, $\mathcal{F}$ is typically the set of all subsets.

  Example: Toss a coin once, $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$
- For uncountable sample spaces, the situation is more complicated - see later.

The probability measure $\mathbb{P}$

A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a set function from $\mathcal{F}$ to $[0, 1]$ satisfying

P1. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$ [probabilities measure long run %'s or certainty]

P2. $\mathbb{P}(\Omega) = 1$ [There is a 100% chance something happens]

P3. Countable additivity: if A_1, A_2 $\cdots$ are disjoint events in $\mathcal{F}$, then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ [Think about it in terms of frequencies]

How do we specify $\mathbb{P}$?

The modelling process consists of

- defining the values of $\mathbb{P}(A)$ for some 'basic events' in $A \in \mathcal{F}$.
- deriving $\mathbb{P}(B)$ for the other 'unknown' more complicated events in $B \in \mathcal{F}$ from the axioms above.

Example: Toss a fair coin 1000 times. Any particular sequence (of length 1000) of H's and T's has chance 2^{-1000}.

- What is the chance there are more than 600 H's in the sequence?
- What is the chance the first time the proportion of heads exceeds the proportion of tails occurs after toss 20?

Properties of $\mathbb{P}$

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- ...
- "Continuity":

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \mathbb{P}(\cup_{i=1}^{n} A_i)$$

$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \mathbb{P}(\cap_{i=1}^{n} A_i)$$

Conditional probability

Let $A, B \in \mathcal{F}$ be events with $\mathbb{P}(B) > 0$. Supposing we know that B occurred, how likely is A given that information? That is, what is the conditional probability $\mathbb{P}(A|B)$?

We define

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

For a frequency interpretation, consider the situation where we have 10 tickets, numbered 1, 2, ..., 10, and we choose one ticket uniformly at random. Then $\Omega = \{1, 2, \ldots, 10\}$.

Let $A = \{\omega \in \Omega : \omega \text{ is even}\} = \{2, 4, 6, 8, 10\}$,
and $B = \{\omega \in \Omega : \omega \text{ is a multiple of } 3\} = \{3, 6, 9\}$.

Then $\mathbb{P}(A) = 5/10$, $\mathbb{P}(B) = 3/10$, $\mathbb{P}(A \cap B) = \mathbb{P}(\{6\}) = 1/10$.
According to the definition, $\mathbb{P}(A|B) = \frac{1/10}{3/10} = 1/3$. This is nothing but the proportion of outcomes in B that are also in A.

Example:

Tickets are drawn consecutively and *without replacement* from a box of tickets numbered $1, \ldots, 10$. What is the chance the second ticket is even numbered given the first is

- even?
- labelled 3?

Law of total probability

Let $B_1, B_2, \cdots, B_n$ be disjoint events with $\mathbb{P}(B_i) > 0$ and $A \subset \bigcup_{j=1}^{n} B_j$, then

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A|B_j)\mathbb{P}(B_j).$$

If also $\mathbb{P}(A) > 0$ then

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1}^{n} \mathbb{P}(A|B_k)\mathbb{P}(B_k)}.$$

Example:

A disease affects 1/1000 newborns and shortly after birth a baby is screened for this disease using a cheap test that has a 2% false positive rate (the test has no false negatives). If the baby tests positive, what is the chance it has the disease?

Independent events

Events A and B are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

If $\mathbb{P}(B) \neq 0$ then this is the same as $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Events $A_1, \cdots, A_n$ are independent if, for each subset $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, n\}$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_k}).$$

This is a stronger requirement than each pair of events being independent.

A disconcerting example

- Let S be the circle of radius 1.
- We say two points on S are in the same family if you can get from one to the other by taking steps of arclength 1 around the circle.

- Each family chooses a single member to be head.
- If X is a point chosen uniformly at random from the circle, what is the chance X is the head of its family?

A disconcerting example

- $A = \{X$ is head of its family$\}$.
- $A_i = \{X$ is i steps clockwise from its family head$\}$.

- $B_i = \{X$ is i steps counterclockwise from its family head$\}$.
- By uniformity, should have $\mathbb{P}(A) = \mathbb{P}(A_i) = \mathbb{P}(B_i)$, BUT
- law of total probability:

$$1 = \mathbb{P}(A) + \sum_{i=1}^{\infty} (\mathbb{P}(A_i) + \mathbb{P}(B_i)) \qquad !$$

The issue is that the set A is not one we can *measure* so should not be included in $\mathcal{F}$.

These kinds of issues are technical to resolve and are dealt with in later probability or analysis subjects which use *measure theory*.

Random variables

A random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that

(Often we want to talk about the probabilities that random variables take values in $(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}$. When we write $\mathbb{P}(X \leq b)$, we mean the probability of the set $\{\omega : X(\omega) \leq b\}$. In order for this to make sense, we need this set to be in $\mathcal{F}$!)

$\{X \leq b\} := \{\omega \in \Omega : X(\omega) \leq b\} \in \mathcal{F}$ for every $b \in \mathbb{R}$

(in measure theory, this is called a *Borel-measurable function*, after Emile Borel (1871-1956)).

Because $\mathcal{F}$ is a sigma-field, this tells us that things like $\{X > b\}$, $\{a < X < b\}$ etc. are also in $\mathcal{F}$.

In this course we will also allow our random variables to take the value $+\infty$ (or $-\infty$). So in general they will be functions from Ω to $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Indicator random variables

The most important examples of random variables are the indicator random variables:

Let $A \in \mathcal{F}$ be an event. Then the function $\mathbb{1}_A : \Omega \to \mathbb{R}$ (or $I_A : \Omega \to \mathbb{R}$) given by

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise,} \end{cases}$$

is called the indicator of A.

Exercise: show that $\mathbb{1}_A$ is a random variable.

Distribution Functions

The function $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega : X(\omega) \leq t\})$ that maps $\mathbb{R}$ to $[0, 1]$ is called the distribution function of the random variable X.

Any distribution function F

F1. is non-decreasing,

F2. is such that $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$,

F3. is 'right-continuous', that is $\lim_{h \to 0^+} F(t+h) = F(t)$ for all t.

Exercise: If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A \in \mathcal{F}$, find the distribution function of $\mathbb{1}_A$.

Distribution Functions

We say that a distribution function F is

- discrete if F only grows in jumps, i.e. there exists a countable set $A \subseteq \mathbb{R}$ and a function f mapping A to $[0, 1]$ such that $F(x) = \sum_{y \in A: y \leq x} f(y)$.
- absolutely continuous if there exists a function f that maps $\mathbb{R}$ to $\mathbb{R}_+$ such that $F(t) = \int_{-\infty}^{t} f(u) du$.

A random variable X has a discrete distribution if F_X is discrete, and an (absolutely) continuous distribution if F_X is absolutely continuous.

The function f is called the probability mass function in the discrete case, and the probability density function in the absolutely continuous case. Note that there are distribution functions that are not even mixtures of the above (see e.g. Cantor Function). In other words there are random variables whose distributions are not mixtures of discrete and absolutely continuous distributions!

Examples of distributions

- Examples of discrete random variables: binomial, Poisson, geometric, negative binomial, discrete uniform
  http://en.wikipedia.org/wiki/Category:Discrete_distributions

- Examples of continuous random variables: normal, exponential, gamma, beta, uniform on an interval (a, b)
  http://en.wikipedia.org/wiki/Category:Continuous_distributions

Random Vectors

A random vector $\mathrm{X} = (X_1, ..., X_d)$ is a vector of random variables on the same probability space.
The distribution function F_{X} of X is

$$F_{\mathrm{X}}(\mathrm{t}) = \mathbb{P}(X_1 \leq t_1, \cdots, X_d \leq t_d), \ \mathrm{t} = (t_1, \cdots, t_d) \in \mathbb{R}^d.$$

One can also write this as

$$F_{\mathrm{X}}(\mathrm{t}) = \mathbb{P}(\cap_{i=1}^{d}\{X_i \leq t_i\}).$$

It follows that

$$\begin{aligned} & \mathbb{P}(s_1 < X_1 \leq t_1, s_2 < X_2 \leq t_2) \\ = \ & F(t_1, t_2) - F(s_1, t_2) - F(t_1, s_2) + F(s_1, s_2). \end{aligned}$$

Independent Random Variables

The random variables $X_1, \cdots, X_d$ are called independent if $F_{\mathbf{X}}(\mathbf{t}) = F_{X_1}(t_1) \times \cdots \times F_{X_d}(t_d)$ for all $\mathbf{t} = (t_1, \cdots, t_d)$.

This turns out to be equivalent to the statement that the events $\{X_1 \in I_1\}, \cdots, \{X_d \in I_d\}$ are independent for all intervals $I_1, \ldots I_d$.

If the random variables are all discrete, or the random vector is absolutely continuous (the latter is stronger than each of its coordinates being absolutely continuous) then this is equivalent to a relevant mass/density function $f_{\mathbf{X}}$ factorising as $f_{\mathbf{X}}(\mathbf{t}) = f_{X_1}(t_1) \cdots f_{X_d}(t_d)$.

Expectation

The expectation (or expected value) of a random variable X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF_X(x).$$

The integral on the right hand side is a Lebesgue-Stieltjes integral. It can be evaluated as

$$= \begin{cases} \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx, & \text{if } X \text{ is absolutely continuous.} \end{cases}$$

In second year, we required that the integral be absolutely convergent. In this course we will allow the expectation to be infinite, provided that we never get in a situation where we have '$\infty - \infty$'.

Expectation of $g(X)$

If X is a random variable then for a measurable (nice) function g, $Y = g(X)$ is a random variable, so by definition

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y dF_Y(y).$$

The law of the unconscious statistician says that also

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

Properties of Expectation

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- If $\mathbb{P}(X \leq Y) = 1$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- If $\mathbb{P}(X = c) = 1$, then $\mathbb{E}[X] = c$.
- ...
- If $0 \leq X_n \uparrow X$ pointwise in ω then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$
- If $X_i \geq 0$ then $\mathbb{E}[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]$
- If $X \geq 0$ then $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > t)dt = \int_0^{\infty} \mathbb{P}(X \geq t)dt$

Moments

- The kth moment of X is $\mathbb{E}[X^k]$.
- The kth central moment of X is $\mathbb{E}[(X - \mathbb{E}[X])^k]$.
- The variance $\mathrm{Var}(X)$ of X is the second central moment $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
- The covariance of $\mathrm{Cov}(X, Y)$ of X and Y is $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.
- If X and Y have finite means and are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$.

Conditioning on random variables

Let X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $A \in \mathcal{F}$.
The conditional probability of event A given a random variable X is written as $\mathbb{P}(A|X)$. The conditional expectation of Y given a random variable X is written as $\mathbb{E}[Y|X]$.
The official definitions are technical, but you should think of these as follows:

- $\mathbb{P}(A|X)$ is a function of X that takes the value $\mathbb{P}(A|X = x)$ on the event that $X = x$.
- $\mathbb{E}[Y|X]$ is a function of X that takes the value $\mathbb{E}[Y|X = x]$ on the event that $X = x$.

As X is a random variable, these two functions of X are also random variables.

Conditioning on random variables - technicality

A random variable has the property that $\{X \leq x\} \in \mathcal{F}$ for every $x \in \mathbb{R}$. It could be however that $\mathcal{F}$ contains a lot more stuff than events regarding X. Let $\mathcal{G}_X$ denote the smallest σ-field on Ω such that X is a random variable on $(\Omega, \mathcal{G}_X, \mathbb{P})$. Suppose that $\mathbb{E}[|Y|] < \infty$. Then $\mathbb{E}[Y|X]$ is defined to be any random variable Z on $(\Omega, \mathcal{G}_X, \mathbb{P})$ such that $\mathbb{E}[Z \mathbb{1}_B] = \mathbb{E}[Y \mathbb{1}_B]$ for each $B \in \mathcal{G}_X$.

- There is typically more than one random variable Z that satisfies this definition. Note that if Z and Z' both satisfy this definition then $\mathbb{P}(Z = Z') = 1$, so we generally don't care which version of $\mathbb{E}[Y|X]$ we are working with.
- $\mathbb{P}(A|X)$ is defined to be $\mathbb{E}[\mathbb{1}_A|X]$.

Example: let $D \in \mathcal{F}$, and let $X = \mathbb{1}_D$. Suppose that Y is a random variable with $\mathbb{E}[|Y|] < \infty$ then the random variable

$$\mathbb{E}[Y|D]\mathbb{1}_D + \mathbb{E}[Y|D^c]\mathbb{1}_{D^c},$$

satisfies the definition of $\mathbb{E}[Y|X]$.

Conditioning on discrete random variables - the punchline

If X has a discrete distribution (taking values $x_1, x_2, \ldots$), and $\mathbb{E}[|Y|] < \infty$ then

$$\mathbb{E}[Y|X] = \sum_i \mathbb{E}[Y|X = x_i] \mathbb{1}_{\{X = x_i\}},$$

i.e. the right hand side satisfies the definition of $\mathbb{E}[Y|X]$.

This is nothing but the statement that if $\eta(x) = \mathbb{E}[Y|X = x]$, then $\eta(X)$ satisfies the definition of $\mathbb{E}[Y|X]$.

Conditional Distribution

The conditional distribution function $F_{Y|X}(y|X)$ of Y evaluated at the real number y is given by $\mathbb{P}(Y \leq y|X)$.

Properties of Conditional Expectation

The following hold with probability 1:

- Linearity: $\mathbb{E}[aY_1 + bY_2|X] = a\mathbb{E}[Y_1|X] + b\mathbb{E}[Y_2|X]$,
- Monotonicity: If $\mathbb{P}(Y_1 \leq Y_2) = 1$, then $\mathbb{E}[Y_1|X] \leq \mathbb{E}[Y_2|X]$,
- $\mathbb{E}[c|X] = c$,
- $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$,

- For any nice (i.e. Borel measurable) function g, $\mathbb{E}[g(X)Y|X] = g(X)\mathbb{E}[Y|X]$
- $\mathbb{E}[Y|X]$ is the function of X that is closest to Y in the mean square sense. This means that $\mathbb{E}[(g(X) - Y)^2]$ is minimised when $g(X) = \mathbb{E}[Y|X]$ (see Borovkov, page 57).

Exercise

Let $\Omega = \{a, b, c, d\}$, and let $\mathcal{F}$ contain all subsets of Ω.
Let $\mathbb{P}$ be the probability measure satisfying
$\mathbb{P}(\{a\}) = \frac{1}{2}$, $\mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \frac{1}{8}$ and $\mathbb{P}(\{d\}) = \frac{1}{4}$.

Define random variables,

$$Y(\omega) = \begin{cases} 1, & \omega = a \text{ or } b, \\ 0, & \omega = c \text{ or } d, \end{cases}$$

$$X(\omega) = \begin{cases} 2, & \omega = a \text{ or } c, \\ 5, & \omega = b \text{ or } d. \end{cases}$$

Compute $\mathbb{E}[X]$, $\mathbb{E}[X|Y]$ and $\mathbb{E}[\mathbb{E}[X|Y]]$.

Example:

Suppose also that the number of individuals M entering a bank in a given day has a Poisson(λ) distribution.

Suppose that individuals entering the bank each hold an Australian passport with probability p, independently of each other and M.

Let N denote the number of individuals holding an Australian passport who enter the bank during that day.

- What is the distribution of N, given that $M = m$?
- Find $\mathbb{E}[N|M = m]$.
- Give an expression for $\mathbb{E}[N|M]$, simplifying where possible.
- Compute $\mathbb{E}[N]$.

Law of Large Numbers (Borovkov §2.9)

The Law of Large Numbers (LLN) states that if $X_1, X_2, \cdots$ are independent and identically-distributed with mean μ, then with probability 1

$$\bar{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu$$

as $n \to \infty$.

Central Limit Theorem (Borovkov §2.9)

The Central Limit Theorem (CLT) states that if $X_1, X_2, \cdots$ are independent and identically-distributed with mean μ and variance $\sigma^2 \in (0, \infty)$, then for any x,

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) \to \Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

as $n \to \infty$.
That is, a suitably-scaled variation from the mean approaches a standard normal distribution as $n \to \infty$.
(Note that writing $Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$, this becomes

$$F_{Z_n}(x) \to F_Z(x),$$

for each x, where $Z \sim \mathcal{N}(0, 1)$.)

Limit Theorems (Borovkov §2.9)

The Poisson Limit Theorem states that that if $X_1, X_2, \cdots$ are independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p_i$, then $X_1 + X_2 + \cdots + X_n$ is well approximated by a Poisson random variable with parameter $\lambda_n = p_1 + \cdots + p_n$.

Specifically, with $W_n = X_1 + X_2 + \cdots + X_n$, then, for each $x \in \mathbb{R}$

$$F_{W_n}(x) \approx F_{Y_n}(x)$$

where $Y_n \sim \mathrm{Poisson}(\lambda_n)$.

(There is, in fact, a bound on the accuracy of this approximation

$$|\mathbb{P}(W_n \in B) - \mathbb{P}(Y_n \in B)| \leq \frac{\sum_{i=1}^{n} p_i^2}{\max(1, \lambda_n)},$$

)

Example

Suppose there are three ethnic groups, A (20%), B (30%) and C (50%), living in a city with a large population. Suppose 0.5%, 1% and 1.5% of people in A, B and C respectively are over 200cm tall.

Suppose that we select at random 50, 50, and 200 people from A, B, and C respectively. What is the probability that at least four of the 300 people will be over 200cm tall?

Stochastic Processes (Borovkov §2.10)

A collection of random variables $\{X_t,\ t \in I\}$ (or $\{X(t), t \in I\}$, or $(X_t)_{t \in I} \ldots$) on a common prob space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a stochastic process. The index variable t is often called 'time'.

- If $I = \{0, 1, 2, \cdots\}$ or $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$, the process is a discrete time process.
- If $I = \mathbb{R}$ or $[0, \infty)$, the process is a continuous time process.

Examples of Stochastic Processes

Standard Brownian Motion is a very special *Gaussian process* $(X_t)_{t\in[0,\infty)}$ where $X_0 = 0$,

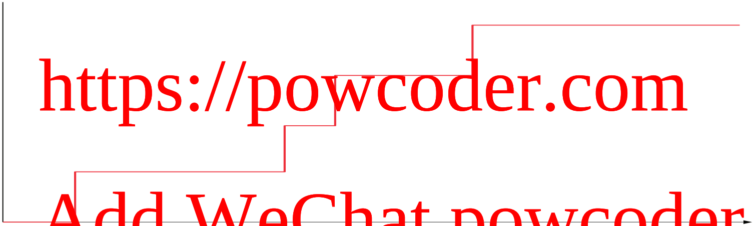
- For each $t > s \geq 0$, $X_t - X_s \sim \mathcal{N}(0, t - s)$

- For any $0 \leq s_1 < t_1 \leq s_2 < \cdots \leq s_k < t_k$, $X_{t_1} - X_{s_1}, \ldots, X_{t_k} - X_{s_k}$ are independent.

- $(X_t)_{t\geq 0}$ is continuous (with probability 1) as a function of t.

Examples of Stochastic Processes

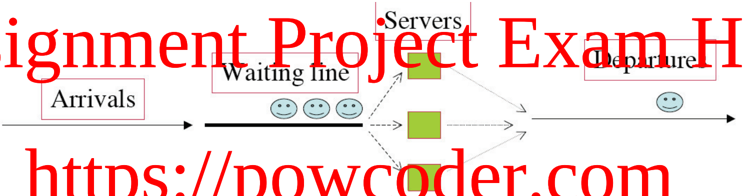
If e.g. X_t is the number of sales of an item up to time t,

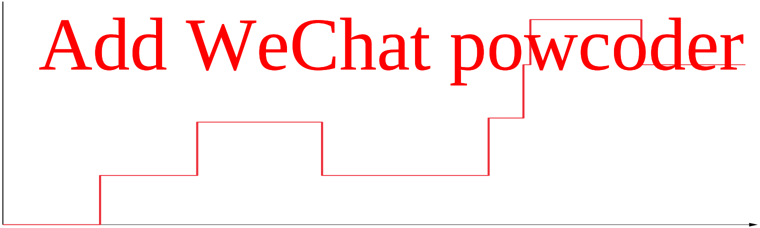


then $(X_t)_{t \geq 0}$ is called a counting process.

Examples of Stochastic Processes

X_t is the number of people in a queue at time t.





Interpretations

- If both ω and t are fixed, then $X_t(\omega)$ is a real number.
- For a fixed t, the function $X_t : \Omega \to \mathbb{R}$ is a random variable.
- For a fixed ω, we can think of t as a variable, and $X_\cdot(\omega) : I \to \mathbb{R}$ as a function (realization, trajectory, sample path).
- If we allow ω to vary, we get a collection of trajectories.

Interpretations

If X_t is a counting process:

- For fixed ω and t, $X_t(\omega)$ is a non-negative integer.
- For fixed ω, $X.(\omega)$ is a non-decreasing step function of t.
- For fixed t, X_t is a non-negative integer-valued random variable.
- For $s < t$, $X_t - X_s$ is the number of events that have occurred in the interval $(s, t]$.

If X_t is the number of people in a queue at time t, then $\{X_t : t \geq 0\}$ is a stochastic process where, for each t, X_t is a non-negative integer-valued random variable but it is NOT a counting process because, for fixed ω, $X_t(\omega)$ can decrease.

Finite-Dimensional Distributions

Knowing just the one-dimensional distributions (i.e. the distribution of X_t for each t) is not enough to describe a stochastic process.

To specify the complete distribution of a stochastic process $(X_t)_{t \in I}$, we need to know the finite-dimensional distributions , i.e. the family of joint distribution functions $F_{t_1, t_2, \cdots, t_k}(x_1, \cdots, x_k)$ of $X_{t_1}, \cdots, X_{t_k}$ for all $k \geq 1$ and $t_1, \cdots, t_k \in I$.

Discrete-Time Markov Chains

We are frequently interested in applications where we have a sequence $X_1, X_2, \cdots$ of outputs (which we model as random variables) in discrete time. For example,

- DNA: A (adenine), C (cytosine), G (guanine), T (thymine).
- Texts: X_j takes values in some alphabet, for example $\{A, B, \cdots, Z, a, \cdots\}$.
  - Developing and testing compression software.
  - Cryptology: codes, encoding and decoding.
  - Attributing manuscripts.

Independence?

Is it reasonable to assume that neighbouring letters are independent?

No, e.g. in English texts

- *a* is very common, but *aa* is very rare.
- *q* is virtually always followed by *u* (and then another vowel).

The Markov Property

The Markov property embodies a natural first generalisation to the independence assumption. It assumes a kind of one-step dependence or memory. Specifically, for $I = \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ (and discrete-valued) processes the Markov property takes the form

$$\begin{aligned}&\mathbb{P}(X_{n+1} = y | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)\\ =&\mathbb{P}(X_{n+1} = y | X_n = x_n)\end{aligned}$$

for each $n \in \mathbb{Z}_+$.

Discrete stopping times

Let $(X_t)_{t\in\mathbb{Z}_+}$ be a sequence of random variables. A random variable T taking values in $\mathbb{Z}_+ \cup \{+\infty\}$ is called a stopping time with respect to $(X_t)_{t\in\mathbb{Z}_+}$ if for each $t \in \mathbb{Z}_+$, there exists a non-random function h_t such that

$$\mathbb{1}_{\{T\le t\}} = h_t(X_0, \ldots, X_t).$$

This says that we can determine whether or not $T \le t$, knowing only the values of $X_0, \ldots, X_t$ (i.e. knowing about the past and the present but without knowing the future).

Example: First hitting time. For a sequence $(X_t)_{t\in\mathbb{Z}_+}$, let $T(x) = \inf\{t \in \mathbb{Z}_+ : X_t = x\}$ denote the first time that the sequence is equal to x. Then

$$\mathbb{1}_{\{T(x)\le t\}} = \sum_{m=0}^{t} \mathbb{1}_{\{T(x)=m\}} = \sum_{m=0}^{t} \mathbb{1}_{\{X_m=x\}} \prod_{i=0}^{m-1} \mathbb{1}_{\{X_i\ne x\}},$$

so $T(x)$ is a stopping time.

Examples:

We have seen above that for a sequence $(X_t)_{t\in\mathbb{Z}_+}$, the first hitting times $T(x) = \inf\{t \in \mathbb{Z}_+ : X_t = x\}$ are stopping times. The following are also stopping times:

- First strictly positive hitting times:
  $T'(x) = \inf\{t \in \mathbb{N} : X_t = x\}$.
- ith hitting times: $T_1(x) = T(x)$,
  $T_i(x) = \inf\{t > T_{i-1}(x) : X_t = x\}$.
- The maximum or minimum of two stopping times.
- A non-random time.

Something like $T - 1$ is not in general a stopping time. E.g. for a Bernoulli sequence, if T is the first time we see a 1 in the sequence then $T - 1$ is not a stopping time. Why?

Continuous stopping times:

A stopping time for a continuous-time process $(X_t)_{t\geq 0}$ is a random variable T (with values in $[0,\infty]$) such that for each $t<\infty$ there is a non-random (measureable) function h_t such that

$$\mathbb{1}_{\{T\leq t\}} = h_t((X_u)_{u\in[0,t]}).$$

As in the discrete setting, hitting times etc., are stopping times.