STAT2003 Introduction to Applied Probability
STAT3102 Stochastic processes

Dr. Elinor Jones

Room 140, 1-19 Torrington Place (first floor)

Office hours: Tuesday 1.30-2.30pm, Friday 1.30pm-2.30pm,
(or by appointment)

E-mail: elinor.jones@ucl.ac.uk

STAT2003 Introduction to Applied Probability
STAT3102 Stochastic processes

- Course notes are available on Moodle
- Exercises will be uploaded on Moodle
- Moodle enrolment key: **stochastic**
- Books (no need to buy):
  - Ross. Introduction to Probability Models. Chapters 3 - 6
  - Stirzaker
  - Cox & Miller, Grimmett & Stirzaker

Assessment

Final mark = 10% ICA + 90% summer exam

- There is NO CHOICE OF QUESTIONS in the ICA or final written exam.
- Only College approved calculators are to be used in the exam and ICA. Please see your handbook for a list of College approved calculators.

In-course assessment

- Provisional date: Friday, 4th March, 3-4pm
- Format: Open book
- Duration: About 45 minutes
- Attendance is required in order to pass the course

Lectures

- Tuesdays 9 - 11am
- Fridays 3 - 4pm

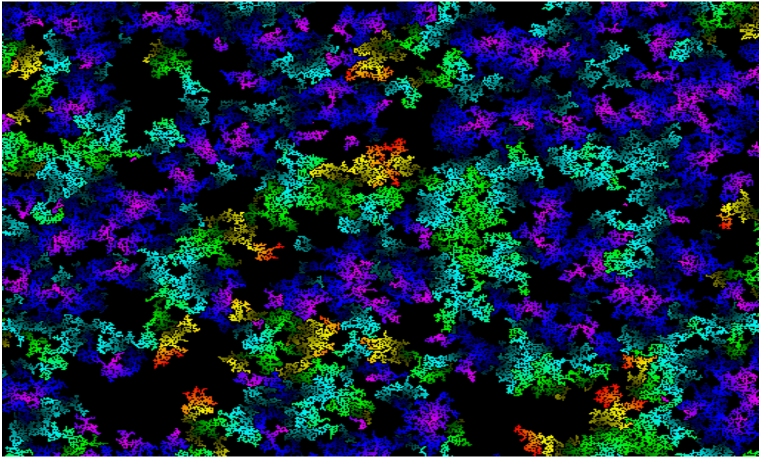
Tutorials

- Tutorial groups uploaded on Moodle.
- Put your answers in the appropriate postbox locker in the student common room (117) by Thursday 12pm. First set due Thursday January 21st.
- The tutor will return your scripts at the next tutorial, with feedback, but no grade or mark.
- Tutorials begin Monday January 25th.

Stochastic processes

- **Stochastic**: random/chance.
- **Process**: time (or space or both).

Variable evolving through time subject to random fluctuations.



The Greenberg - Hastings model

Stochastic processes

We consider some simple stochastic processes which are used to model real-life situations.

We concentrate on the mathematical tools used to study their properties.

Recurring theme

the use of conditional probability to solve problems.

Why do we need to revise probability and random variables?

Fundamental idea of course:

Study properties of stochastic (random) processes.

Cannot do this unless we're well-versed in:

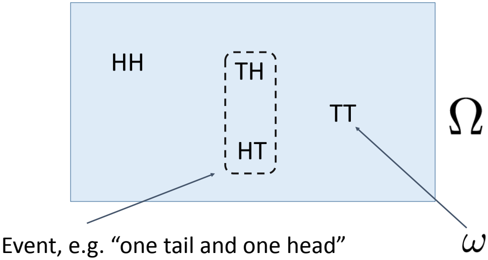
- Probability, conditional probability;
- Random variables, conditional random variables and expectation;
- Handling generating functions.

Experiments

An experiment is any situation where there is a set of possible outcomes.

- The set of possible outcomes is called the **sample space**, Ω.
- An **outcome** is denoted ω and is an element of the sample space.
- An **event** is a subset of the sample space.

e.g. tossing a coin twice:



Probability

Finite sample space $\Rightarrow$ can assign probabilities to individual sample points ω via a 'weight function', $p : \Omega \to [0, 1]$.

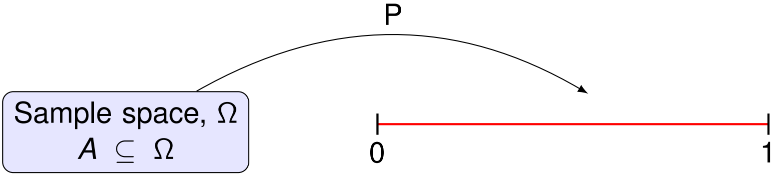
This allocates a value $p(\omega)$ for each possible outcome ω, which we interpret as the probability that ω occurs:

$$\sum_{\omega \in \Omega} p(\omega) = 1,$$

then we may calculate the probability of an event $A \subseteq \Omega$ via

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Probability



P must satisfy the following three axioms of probability:

1. $\mathrm{P}(A) \geqslant 0$ for all events A,
2. if $A_1, A_2, A_3, \ldots$ are events with $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\mathrm{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$ i.e. P is countably additive,
3. $\mathrm{P}(\Omega) = 1$.

Random variables

Make numerical measurements on outcomes ω via a **random variable** X: a function from the sample space to the reals.

$$X : \Omega \to \mathbb{R}$$

For example, let the experiment be to **throw two fair coins.** Then

$$\Omega = \{HH, HT, TH, TT\}$$

Let X be the random variable "**number of heads**".

What does X map to?

Random variables

Since

$$\Omega = \{HH, HT, TH, TT\},$$

and X is the random variable "**number of heads**", then

$$\begin{aligned} X(HH) &= 2, \\ X(HT) &= X(TH) = 1, \\ X(TT) &= 0. \end{aligned}$$

So, for example:

- Observe Tail followed by another Tail.
- $X(TT)$ represents the numerical value of the measurement when the outcome of the experiment is $\omega = TT$.

Random variables

What's the probability that X takes a value less than or equal to some value x?

$$P(X \leqslant x) = ?$$

What we really want to do is calculate the probability of the event

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \subseteq \Omega$$

So the cdf of a random variable is given, for $x \in \mathbb{R}$, by:

$$F_X(x) = \mathrm{P}(X \leqslant x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \leqslant x\}).$$

Random variables

What's the probability that X takes a value less than or equal to some value x?

$$P(X \leqslant x) = ?$$

What we really want to do is calculate the probability of the event

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \subseteq \Omega$$

So the cdf of a random variable is given, for $x \in \mathbb{R}$, by:

$$F_X(x) = \mathrm{P}(X \leqslant x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \leqslant x\}).$$

Random variables

What's the probability that X takes a value less than or equal to some value x?

$$P(X \leqslant x) = ?$$

What we really want to do is calculate the probability of the event

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \subseteq \Omega$$

So the cdf of a random variable is given, for $x \in \mathbb{R}$, by:

$$F_X(x) = \mathrm{P}(X \leqslant x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \leqslant x\}).$$

Random variables

So in our coin tossing example where

$$\Omega = \{HH, HT, TH, TT\},$$

and X is the random variable "**number of heads**", we have that

$$\mathrm{P}(\omega) = 1/4$$

for all outcomes $\omega = \{HH, HT, TH, TT\}$, and

$$F_X(1) = \mathrm{P}(X \leqslant 1) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \leqslant 1\}) = 3/4,$$

because the outcomes, ω, that satisfy $X(\omega) \leqslant 1$ are $\{TT, HT, TH\}$.

Discrete random variables

- X is a **discrete random variable** if it takes only finitely many or countably infinitely many values.
- The **probability mass function** of a discrete random variable is

$$p_X(x) = \mathrm{P}(X = x) = \mathrm{P}(\{\omega : X(\omega) = x\}).$$

Discrete random variables

The **expectation** of X is

$$E(X) = \sum_x x \, P(X = x).$$

For a function g we have

$$E(g(X)) = \sum_x g(x) \, P(X = x).$$

The variance of X is

$$\mathrm{var}(X) \;=\; E\left[(X - E(X))^2\right] \;=\; E(X^2) - [E(X)]^2 \, .$$

Continuous random variables

X is a **continuous random variable** if it has a **probability density function.**

That is, if there exists a non-negative function f_X with $\int_{-\infty}^{\infty} f_X(x)dx = 1$, such that for all x,

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du\,.$$

Then

$$f_X(x) = \frac{dF_X(x)}{dx}\,.$$

Continuous random variables

The expectation of X is

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

and the expectation of $g(X)$, for some function g, is

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx\,,$$

provided that these integrals exist.

As before, the variance of X is

$$\mathrm{var}(X) \;=\; \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \;=\; \mathrm{E}(X^2) - [\mathrm{E}(X)]^2\,.$$

What is a generating function?

[Definition taken from Grimmett and Stirzaker]

A sequence $a = \{a_i : i = 0, 1, 2, ...\}$ of real numbers may contain a lot of information.

One concise way of storing this information is to wrap up the numbers together in a 'generating function'.

For example, the (ordinary) generating function of the sequence a is the function

$$\boxed{G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}}$$

What is a generating function?

[Definition taken from Grimmett and Stirzaker]

A sequence $a = \{a_i : i = 0, 1, 2, ...\}$ of real numbers may contain a lot of information.

One concise way of storing this information is to wrap up the numbers together in a 'generating function'.

For example, the (ordinary) generating function of the sequence a is the function

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}$$

What is a generating function?

[Definition taken from Grimmett and Stirzaker]

A sequence $a = \{a_i : i = 0, 1, 2, ...\}$ of real numbers may contain a lot of information.

One concise way of storing this information is to wrap up the numbers together in a 'generating function'.

For example, the (ordinary) generating function of the sequence a is the function

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}$$

What is a generating function?

[Definition taken from Grimmett and Stirzaker]

A sequence $a = \{a_i : i = 0, 1, 2, ...\}$ of real numbers may contain a lot of information.

One concise way of storing this information is to wrap up the numbers together in a 'generating function'.

For example, the (ordinary) generating function of the sequence a is the function

$$\boxed{G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}}$$

What is a generating function?

Probability generating function (p.g.f.) of RV X which takes values in $\mathbb{N}$:

$$G(t) = \mathrm{E}[t^X] = \sum_{k=0}^{\infty} t^k P(X = k)$$

Moment generating function (m.g.f.) of RV X:

$$M(t) = \mathrm{E}[\exp(tX)] = \sum_{k} \exp(tk) P(X = k)$$

$$M(t) = \mathrm{E}[\exp(tX)] = \int_{\mathbb{R}} \exp(tk) f_X(k) dk$$

Why are generating functions useful?

They can be used to :

1. calculate moments;
2. calculate probabilities;
3. explore aspects of the distribution of sums of (independent) random variables.

Calculating moments.

Probability generating function of an **integer-valued** RV X:

$$\boxed{G_X(t) = \mathrm{E}[t^X]}$$

Equivalently:

$$G_X(t) = P(X=0) + tP(X=1) + t^2 P(X=2) + t^3 P(X=3) + \ldots$$

Then:

$$\mathrm{E}(X) = \left|\frac{\mathrm{d}G_X(t)}{\mathrm{d}t}\right|_{t=1} = G'_X(1),$$

$$\mathrm{E}[X(X-1)] = \left|\frac{\mathrm{d}^2 G_X(t)}{\mathrm{d}t^2}\right|_{t=1} = G''_X(1).$$

Calculating moments.

Probability generating function of an **integer-valued** RV X:

$$\boxed{G_X(t) = \mathrm{E}[t^X]}$$

Equivalently:

$$G_X(t) = P(X = 0) + tP(X = 1) + t^2 P(X = 2) + t^3 P(X = 3) + \ldots$$

Then:

$$\mathrm{E}(X) = \left|\frac{\mathrm{d}G_X(t)}{\mathrm{d}t}\right|_{t=1} = G'_X(1),$$

$$\mathrm{E}[X(X-1)] = \left|\frac{\mathrm{d}^2 G_X(t)}{\mathrm{d}t^2}\right|_{t=1} = G''_X(1).$$

Calculating probabilities.

Probability generating function of an **integer-valued** RV X:

$$\boxed{G_X(t) = \mathrm{E}[t^X]}$$

Equivalently:

$$G_X(t) = P(X = 0) + tP(X = 1) + t^2 P(X = 2) + t^3 P(X = 3) + \ldots$$

Then:

(a) $G_X(0) = \sum_{k=0}^{\infty} 0^k \, \mathrm{P}(X = k) = \mathrm{P}(X = 0)$.

(b) If we expand $G_X(t)$ in powers of t the coefficient of t^k is equal to $\mathrm{P}(X = t)$.

Calculating probabilities.

Probability generating function of an **integer-valued** RV X:

$$\boxed{G_X(t) = \mathrm{E}[t^X]}$$

Equivalently:

$$G_X(t) = P(X = 0) + tP(X = 1) + t^2 P(X = 2) + t^3 P(X = 3) + \dots$$

Then:

(a) $G_X(0) = \sum_{k=0}^{\infty} 0^k \,\mathrm{P}(X = k) = \mathrm{P}(X = 0)$.

(b) If we expand $G_X(t)$ in powers of t the coefficient of t^k is equal to $\mathrm{P}(X = t)$.

Calculating probabilities.

For example,

$$\begin{aligned} G_X(t) &= \frac{pt}{1-(1-p)t}, \\ &= pt\sum_{k=0}^{\infty}(1-p)^k t^k, \qquad (j = k+1) \\ &= \sum_{j=1}^{\infty} t^j (1-p)^{j-1} p. \end{aligned}$$

Therefore, $\mathrm{P}(X = j) = (1-p)^{j-1}p$, so $X \sim \mathrm{Geometric}(p)$.

The distribution of sums of random variables.

Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with common p.g.f. $G_X(t)$. The p.g.f. $G_Y(t)$ of $Y = X_1 + \cdots + X_n$ is given by

$$\begin{aligned} G_Y(t) &= \mathrm{E}\left(t^{X_1+\cdots+X_n}\right), \\ &= \prod_{i=1}^{n} \mathrm{E}\left(t^{X_i}\right), \\ &= [G_X(t)]^n. \end{aligned}$$

Also, if $Z_1 \perp\!\!\!\perp Z_2$ (that is, Z_1 and Z_2 are independent but do not necessarily have the same distribution) then

$$G_{Z_1+Z_2}(t) = G_{Z_1}(t)G_{Z_2}(t)$$

.

Calculating the p.g.f. of a random sum

(that is, the sum of a *random* number of random variables).

Suppose that $X_1, \ldots, X_N$ are i.i.d. random variables with common p.g.f. $G_X(t)$ and that N has p.g.f. $G_N(t)$.

The p.g.f. $G_Y(s)$ of $Y = X_1 + \cdots + X_N$ is given by

$$\begin{aligned} G_Y(s) &= \mathrm{E}\,\mathrm{E}\left(s^Y \mid N\right), \\ &= \mathrm{E}\left([G_X(s)]^N\right), \\ &= G_N[G_X(s)]. \end{aligned}$$

Challenge: show that the m.g.f. $M_Y(t)$ of Y is given by

$$M_Y(t) = \mathrm{E}\left(\mathrm{e}^{tY}\right) = G_N[M_X(t)].$$

Exercise

X and Y are both independent Poisson random variables with parameter 6 and 2, respectively.

What is the distribution of $X + Y$?

Exercise

The number of students entering London University Union each day follows a Poisson distribution with parameter 2,000. Of these, 40% are UCL students.

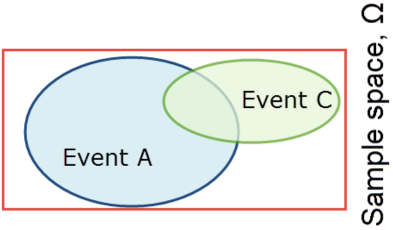
What is the distribution of the number of UCL students who enter the Union each day?

Exercise

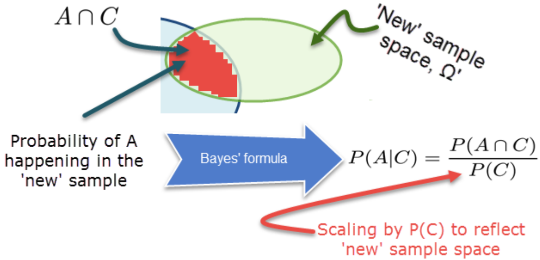
You have a deck of playing cards (52 cards in total, split equally into four suits: hearts, diamonds, spades and clubs, with the first two considered 'red' cards, and the others considered 'black' cards).

- I pick a card at random. What is the probability that the chosen card is from the diamond suit?
- I pick a card at random and tell you that it is a red card (i.e. either hearts or diamonds). What is the probability that the chosen card is from the diamond suit?

Conditional probability



If we know that Event C happened, then this becomes our new sample space...



Conditional probability

Let A and C be events with $\mathrm{P}(C) > 0$. The **conditional probability of** A **given** C is

$$\mathrm{P}(A \mid C) = \frac{\mathrm{P}(A \cap C)}{\mathrm{P}(C)} .$$

It is easy to verify that

1. $\mathrm{P}(A \mid C) \geqslant 0$,
2. if $A_1, A_2, \ldots$ are mutually exclusive events, then

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots \mid C) = \sum_i \mathrm{P}(A_i \mid C),$$

3. $\mathrm{P}(C \mid C) = 1$.

Conditional probability

- In other words, **conditional** probability obeys the axioms of probability.
    - C plays the role of Ω.
    - conditioning event C is fixed
- This means that any theorem or result that holds for probability also holds for conditional probability with a fixed conditioning event.

Conditional probability

- In other words, **conditional** probability obeys the axioms of probability.
  - C plays the role of Ω.
  - conditioning event C is fixed
- This means that any theorem or result that holds for probability also holds for conditional probability with a fixed conditioning event.

Conditional probability

- In other words, **conditional** probability obeys the axioms of probability.
  - C plays the role of Ω.
  - conditioning event C is fixed
- This means that any theorem or result that holds for probability also holds for conditional probability with a fixed conditioning event.

Conditional probability

- In other words, **conditional** probability obeys the axioms of probability.
  - C plays the role of Ω.
  - conditioning event C is fixed
- This means that any theorem or result that holds for probability also holds for conditional probability with a fixed conditioning event.

Conditional probability for discrete RVs

Joint and marginal probabilities

Now consider discrete random variables X and Y.
The **joint probability mass function** of X and Y is

$$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y).$$

The **marginal probability mass function** of X is

$$p_X(x) = \mathrm{P}(X = x) = \sum_y \mathrm{P}(X = x, Y = y).$$

The marginal pmf of Y is defined similarly.

Conditional probability for discrete RVs

The **conditional probability mass function of** X **given** $Y = y$ is

$$p_{X|Y=y}(x) \; = \; \mathrm{P}(X = x \mid Y = y) \; = \; \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)}$$

for y such that $\mathrm{P}(Y = y) > 0$.

The **conditional distribution function of** X **given** $Y = y$ is

$$F_{X|Y=y}(x) \; = \; \mathrm{P}(X \leqslant x \mid Y = y)\,.$$

Conditional probability for continuous RVs

Joint and marginal distribution function

Now consider continuous random variables X and Y.

The **joint probability density function** of X and Y is $f_{X,Y}(x,y)$.

The **marginal probability density function** of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

The marginal pdf of Y is defined similarly.

Conditional probability for continuous RVs

The **conditional probability density function of** X **given** $Y = y$ is

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f(y)}$$

for y such that $f(y) > 0$.

The **conditional distribution function of** X **given** $Y = y$ is

$$F_{X|Y=y}(x) = \int_{-\infty}^{x} f_{X|Y=y}(t)dt = P(X \leqslant x | Y = y)$$

Conditional expectation

The **conditional expectation of X given** $Y = y$ is

$$E(X|Y=y) = \begin{cases} \sum_x x P(X = x \mid Y = y) & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} x f_{X|Y=y}(x)\, \mathbf{d}x & \text{for continuous RVs} \end{cases}$$

The **conditional expectation of** $g(X)$ **given** $Y = y$**, for some function** g**,** is

$$E(g(X)|Y=y) = \begin{cases} \sum_x g(x) P(X = x \mid Y = y) & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y=y}(x)\, \mathbf{d}x & \text{for continuous RVs} \end{cases}$$

provided that these exist.

Example

$$\Omega = \{\omega_1, \omega_2, \omega_3\}$$

and

$$\mathrm{P}(\omega_1) = 1/3 \qquad \mathrm{P}(\omega_2) = 1/6 \qquad \mathrm{P}(\omega_3) = 1/2$$

Suppose also that X and Y are random variables with

$$\begin{aligned} X(\omega_1) &= 2 \qquad & Y(\omega_1) &= 2 \\ X(\omega_2) &= 3 \qquad & Y(\omega_2) &= 2 \\ X(\omega_3) &= 1 \qquad & Y(\omega_3) &= 1 \end{aligned}$$

Find the conditional pmf $p_{X|Y=2}(x)$ and the conditional expectation $\mathrm{E}(X \mid Y = 2)$.

Example

Possible values of X are 1, 2, 3.

$$p_{X|Y=2}(1) = \mathrm{P}(X = 1 \mid Y = 2) = 0.$$

$$p_{X|Y=2}(2) = \mathrm{P}(X=2 \mid Y=2) = \frac{\mathrm{P}(X=2, Y=2)}{\mathrm{P}(Y=2)}$$

$$= \frac{\mathrm{P}(\omega_1)}{\mathrm{P}(\omega_1)+\mathrm{P}(\omega_2)} = \frac{2}{3}.$$

$$p_{X|Y=2}(3) = \mathrm{P}(X=3 \mid Y=2) = \frac{\mathrm{P}(X=3, Y=2)}{\mathrm{P}(Y=2)}$$

$$= \frac{\mathrm{P}(\omega_2)}{\mathrm{P}(\omega_1)+\mathrm{P}(\omega_2)} = \frac{1}{3}.$$

Example

Therefore:

$$\begin{aligned} p_{X|Y=2}(1) &= 0; \\ p_{X|Y=2}(2) &= 2/3; \\ p_{X|Y=2}(3) &= 1/3. \end{aligned}$$

The conditional expectation is

$$\begin{aligned} \mathrm{E}(X \mid Y = 2) = &\sum_x x\, \mathrm{P}(X = x \mid Y = 2) \\ = &(1 \times 0) + (2 \times 2/3) + (3 \times 1/3) = 7/3. \end{aligned}$$

An important concept

- Notation $\mathrm{E}(X \mid Y)$ used to denote a **random variable**.
- It takes the value $\mathrm{E}(X \mid Y = y)$ with probability $\mathrm{P}(Y = y)$.
- $\mathrm{E}(X \mid Y)$ is a function of the random variable Y.

Using the previous example:

$$E(X \mid Y) = \begin{cases} E(X \mid Y = 2) & \text{w.p. } P(Y = 2) \\ E(X \mid Y = 1) & \text{w.p. } P(Y = 1) \end{cases}$$

On calculating the conditional expectations and probabilities above:

$$E(X \mid Y) = \begin{cases} \frac{7}{3} & \text{w.p. } \frac{1}{2} \\ 1 & \text{w.p. } \frac{1}{2} \end{cases}$$

Important formulae

Law of total probability

Let A be an event and Y be ANY random variable. Then

$$\mathrm{P}(A) = \begin{cases} \sum_y \mathrm{P}(A|Y=y)\mathrm{P}(Y=y) & \text{if } Y \text{ discrete} \\ \int_{-\infty}^{\infty} \mathrm{P}(A \mid Y=y) f_Y(y)\, \mathbf{d}y & \text{if } Y \text{ continuous} \end{cases}$$

Important formulae

Modification of the law of total probability

Iif Z is a third discrete random variable,

$$
\begin{aligned}
\mathrm{P}(X = x \mid Z = z) &= \sum_y \mathrm{P}(X = x, Y = y \mid Z = z) \\
&= \sum_y \frac{\mathrm{P}(X = x, Y = y, Z = z)}{\mathrm{P}(Z = z)} \\
&= \sum_y \frac{\mathrm{P}(X = x, Y = y, Z = z)}{\mathrm{P}(Y = y, Z = z)} \frac{\mathrm{P}(Y = y, Z = z)}{\mathrm{P}(Z = z)} \\
&= \sum_y \mathrm{P}(X = x \mid Y = y, Z = z) \mathrm{P}(Y = y \mid Z = z),
\end{aligned}
$$

assuming the conditional probabilities are all defined.

Important formulae

Law of conditional (iterated) expectation

$$E[X] = E[E(X \mid Y)] = \begin{cases} \sum_y E(X \mid Y = y)\, P(Y = y) & \text{if } Y \text{ discrete} \\ \int_{-\infty}^{\infty} E(X \mid Y = y) f_Y(y)\, \mathbf{d}y & \text{if } Y \text{ continuous} \end{cases}$$

and, more generally,

$$E[g(X)] \;=\; E\Big(E[g(X) \mid Y]\Big)$$

Back to the example.

Earlier we found the distribution of the random variable $\mathrm{E}(X \mid Y)$:

$$\mathrm{E}(X|Y) = \begin{cases} 1 & \text{with probability } 1/2 \\ \frac{7}{3} & \text{with probability } 1/2 \end{cases}$$

The expectation of this random variable is

$$\mathrm{E}_Y[\mathrm{E}(X|Y)] = 1 \times \frac{1}{2} + \frac{7}{3} \times \frac{1}{2} = \frac{5}{3}.$$

Also, the expectation of X is given by

$$\mathrm{E}(X) = 1 \times \frac{1}{2} + 2 \times \frac{1}{3} + 3 \times \frac{1}{6} = \frac{5}{3},$$

which verifies the law of conditional (iterated) expectation.

Example

Let X and Y have joint density

$$f_{X,Y}(x,y) = \frac{1}{y} e^{-x/y} e^{-y}, \qquad 0 < x, y < \infty .$$

Find $f_Y(y)$, $\mathrm{E}(X \mid Y)$ and hence $\mathrm{E}(X)$.

Example: find $f_Y(y)$

For any $y > 0$,

$$f_Y(y) = \int_0^\infty \frac{1}{y} e^{-x/y} e^{-y} \, \mathrm{d}x = e^{-y}[-e^{-x/y}]_0^\infty = e^{-y},$$

so Y has an exponential distribution with parameter 1. Also, for any fixed $y > 0$ and any $x > 0$,

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{e^{-x/y} e^{-y}}{y e^{-y}} = \frac{1}{y} e^{-x/y},$$

so $X \mid Y = y \sim \exp(1/y)$.

Example : find $E(X \mid Y)$ and hence $E(X)$.

It follows that $E(X \mid Y = y) = y$ and so $E(X \mid Y) = Y$.

Using the Law of Conditional Expectations we then have

$$E(X) = E[E(X \mid Y)] = E(Y) = 1\,.$$

First step decomposition

Seriously important!

Using the law of conditional (iterated) expectation to solve problems.

- Extremely useful technique for many types of problems.
- In this course, mainly interested in applying this technique by conditioning on ‘where does our random process go next?’.

First step decomposition

Seriously important!

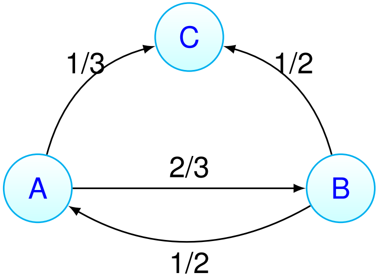
Typical questions that we will solve using this technique:

1. How long does it take, on average, to reach a particular point for the first time?
2. How often, on average, do we visit a certain point before a particular event occurs?
3. What's the probability that we ever reach/ never reach a particular point?

First step decomposition

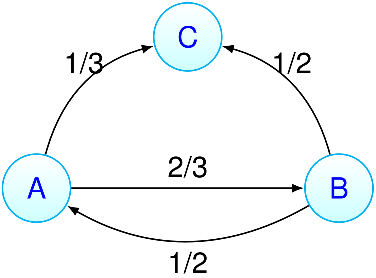
Example

The time taken to cycle along any road demonstrated below is 1 minute. Geraint (a bike courier) is at location A, and needs to get to C. However, he is a law abiding cyclist who will only follow the direction of travel permitted on each road, as indicated by the arrows. He selects any of the routes available to him at each junction with probabilities as given in the diagram.



First step decomposition

Example



(i) How long, on average, will it take Geraint to deliver the parcel?

(ii) How many times, on average, will Geraint visit location B before delivering the parcel?