37161 Probability and Random Variables

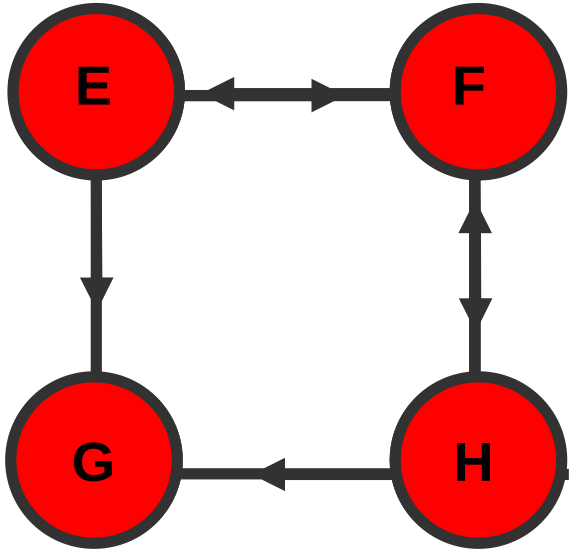
Lecture 11

Classification of Markov Chains: Absorbing

- We say that a state of a Markov Chain is **absorbing** if and only if, once entered, it cannot be left.

- That is, if state *i* is absorbing, then the transition probability $p_{ii} = 1$ and all other transition probabilities from *i*, $p_{ij} = 0$ for $i \neq j$.

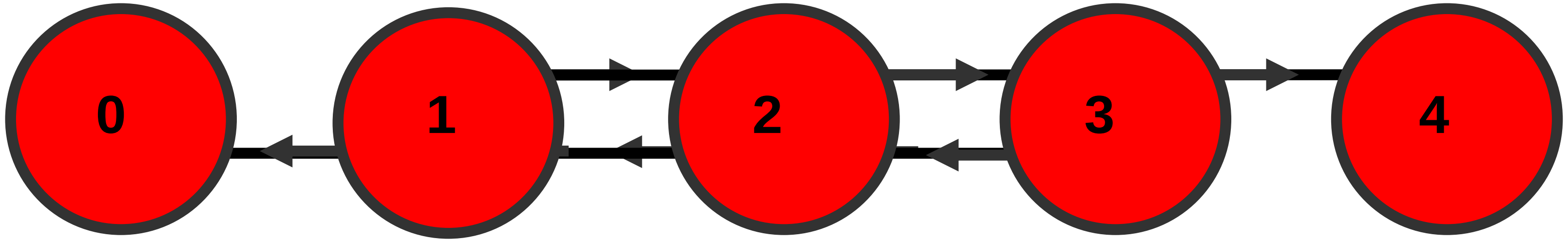
- In this example (with all arrows from a given state equally weighted), G is an absorbing state as there is no way to leave it once entered.



Classification of Markov Chains: Absorbing

- A Markov Chain can have more than one absorbing state.
- For example, consider the case of a gambler who bets on the repeated flips of a fair coin, such that he/she wins $1 each time it lands Heads and $1 each time it lands Tails. He/she bets repeatedly until he/she first reaches $0 (bankruptcy) or $4 (profit.)
- The transition matrix and state diagram for this are:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



Classification of Markov Chains: Absorbing

- We find the steady-state distribution of the Markov Chain by solving $\Pi_{eq} = \Pi_{eq} P$.

- Here, we find $\Pi_{eq} = \begin{pmatrix} A & B & C & D & E \end{pmatrix}$ such that

$$\begin{pmatrix} A & B & C & D & E \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} A & B & C & D & E \end{pmatrix}$$

- This gives $\begin{pmatrix} A + \frac{1}{2}B & \frac{1}{2}C & \frac{1}{2}B + \frac{1}{2}D & \frac{1}{2}C & \frac{1}{2}D + E \end{pmatrix} = \begin{pmatrix} A & B & C & D & E \end{pmatrix}$

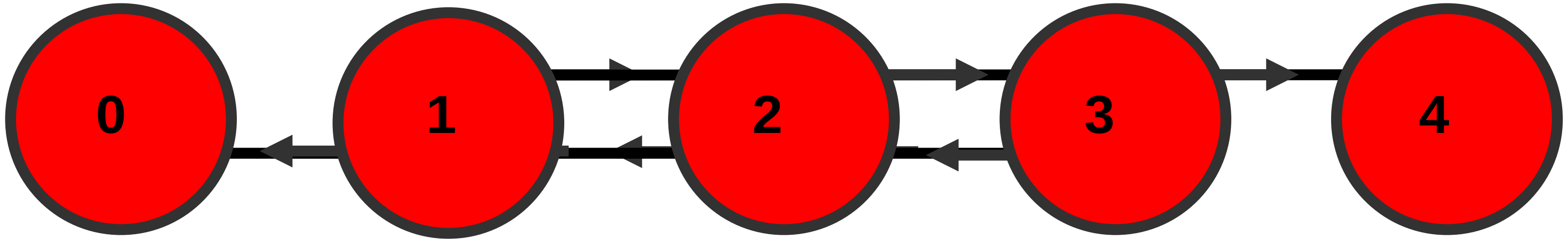
Classification of Markov Chains: Absorbing

- $\begin{pmatrix} A+\frac{1}{2}B & \frac{1}{2}C & \frac{1}{2}B+\frac{1}{2}D & \frac{1}{2}C & \frac{1}{2}D+E \end{pmatrix} = \begin{pmatrix} A & B & C & D & E \end{pmatrix} = \Pi_{eq}$ can be solved by either

  $\Pi_{eq} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}$ or $\Pi_{eq} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}$ (or any linear combination of these.)

- There is no unique absorbing state.

- This can be seen from the state diagram. The system could eventually get “stuck” in State 0 or in State 4.



Classification of Markov Chains: Communication

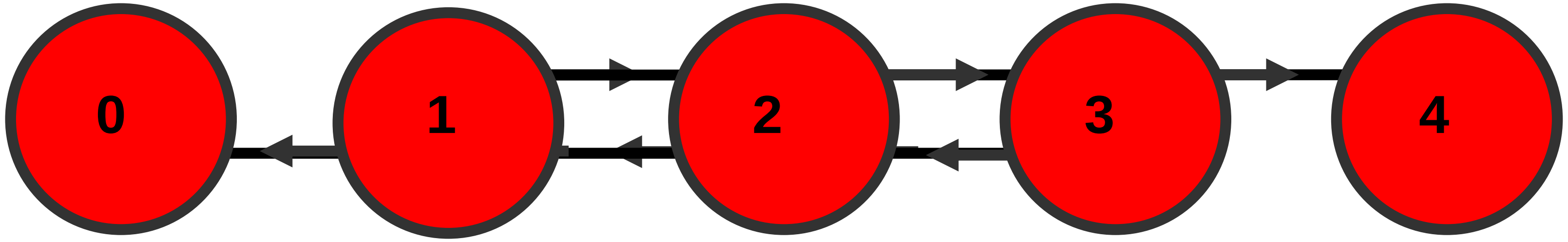
- The state *i* is said to **communicate** with state *j* (sometimes written as $i \to j$) if, with non-zero probability, a system which is in state *i* reaches state *j* at a future time. That is, for at least one possible *n*, the *n*-step transition matrix has $p^n_{ij} > 0$.

- In the case below, $1 \to 2$, $2 \to 3$ and $3 \to 4$. Also, $1 \to 4$, $1 \to 3$ and $2 \to 4$.

- Similarly, $1 \to 0$, $2 \to 1$, $3 \to 2$, $2 \to 0$, $3 \to 0$ and $3 \to 1$.

- We can also write

$1 \leftrightarrow 2$, $2 \leftrightarrow 3$, and $1 \leftrightarrow 3$.



Classification of Markov Chains: Ergodicity

- A Markov Chain is said to be **ergodic** if every state communicates with every other.

- That is, when leaving any state, it is possible (but not necessarily in a single move) to reach any other.

Classification of Markov Chains: Transience

- The state *i* is said to be **transient** if, after leaving state *i*, the probability that it is ever in state *i* again < 1.

- Clearly any state which communicates with an absorbing state is transient. For example, in the example of the fair $1 bets, the states {1, 2, 3} are transient, since there is a non-zero probability that the chain ever reaches either absorbing barrier, hence a non-zero probability that they do not return to those other states.

- States which are not transient are known as **intransient** or **persistent**.

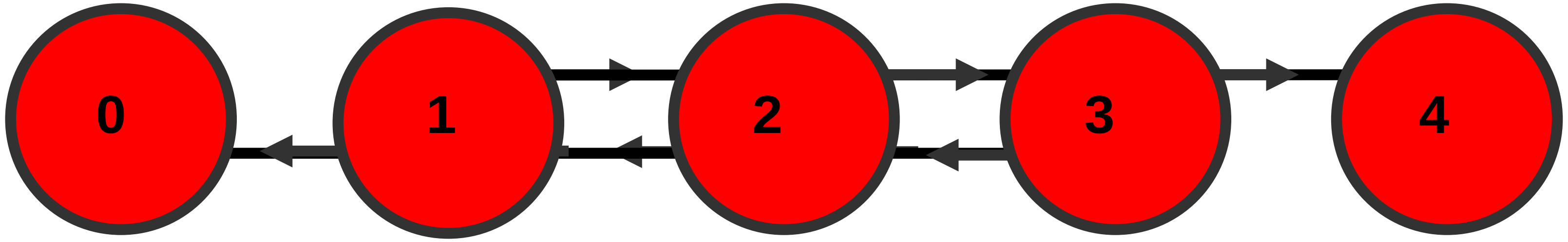
Classification of Markov Chains: Periodicity

- We say that a persistent state of a Markov Chain is **periodic** with period *d* when *d* is the largest integer such that no transitions from the state to itself are possible unless the number of moves is divisible by *d*.

- That is, when calculating an *n*-step transition matrix P^n, the state *i* is periodic with period *d* if *d* is the largest integer such that $p^n{}_{ii} = 0$ for every n not divisible by *d*.

- In the simple random walk below, states {1, 2, 3} are period 2 and {0 , 4} are period 1.

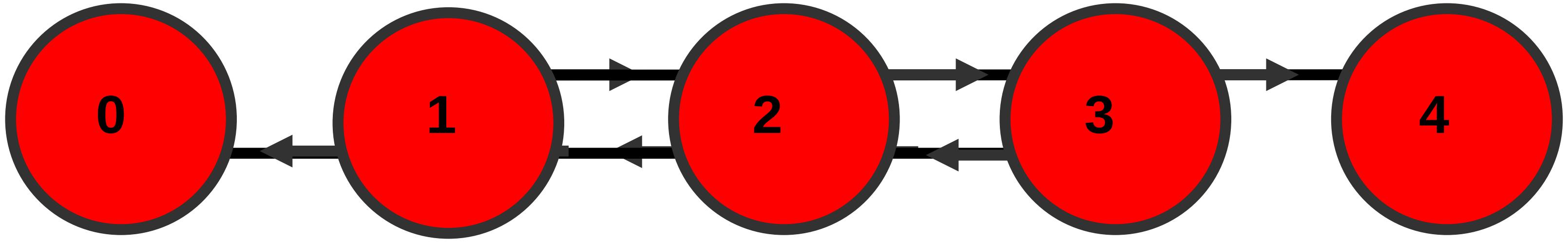
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



Classification of Markov Chains: Periodicity

- A state which has period 1 is also called **aperiodic**.
- States {1, 2, 3} are periodic with period 2, since each turn which begins in an even-numbered state moves the Markov Chain into an odd-numbered state. It therefore takes 2, 4, 6,… etc. moves before a state can be returned to.
- If the player bets from \$2, he/she moves to either \$1 or \$3 but can return to \$2 one move later.

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



Classification of Markov Chains: Periodicity

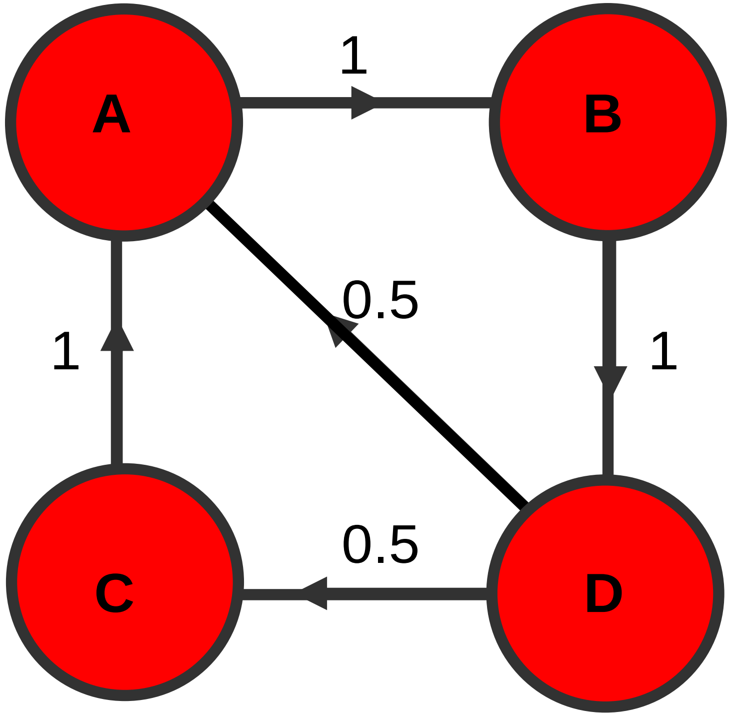
- Note, however, that a state can have period *d* but no possible return path in *d* moves.

- For example, consider transitions from State A to itself.

- One option is going A to B, B to D, D to C then C to A.
- This is four moves.

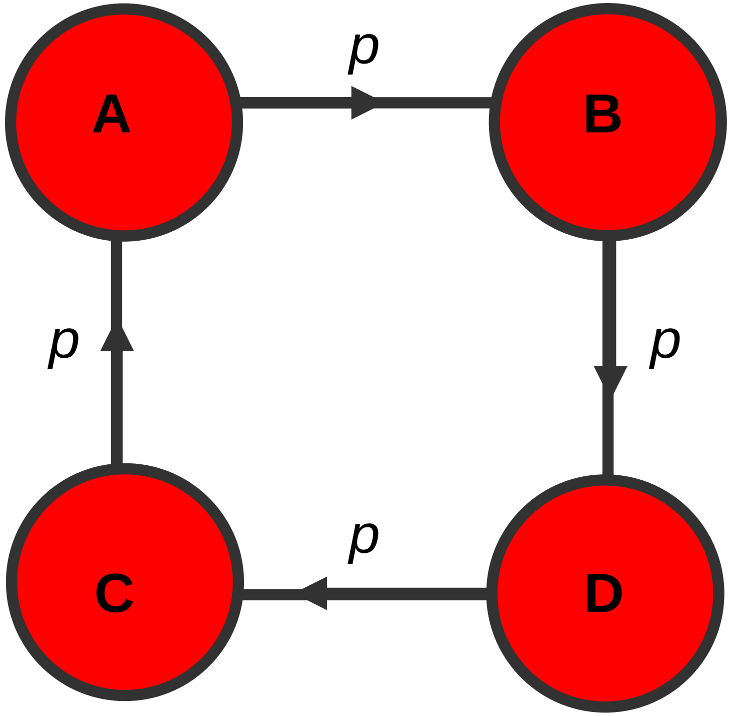
- Another option is going from A to B, B to D and D to A.
- This is three moves.

- The largest integer which is a divisor of both 3 and 4 is 1, so the state has period 1, although it is not possible to return to A in a single move.



Classification of Markov Chains

- Consider the Markov Chain represented by the state diagram below.
- What condition must hold for the chain to be ergodic (i.e. every state is reachable from every other)?
- $p \neq 0$. If this does not hold, every state is absorbing.
- If the chain is periodic (i.e. not just period 1) what is the only possible period? What condition must hold for this to happen?
- If $p = 1$, the chain has period 4, since you can only go from a state to itself in 4, 8, 12, 16 etc. moves. If $p \neq 1$, then it is aperiodic.



Urn Models

- One commonly studied class of models is that concerned with urns.

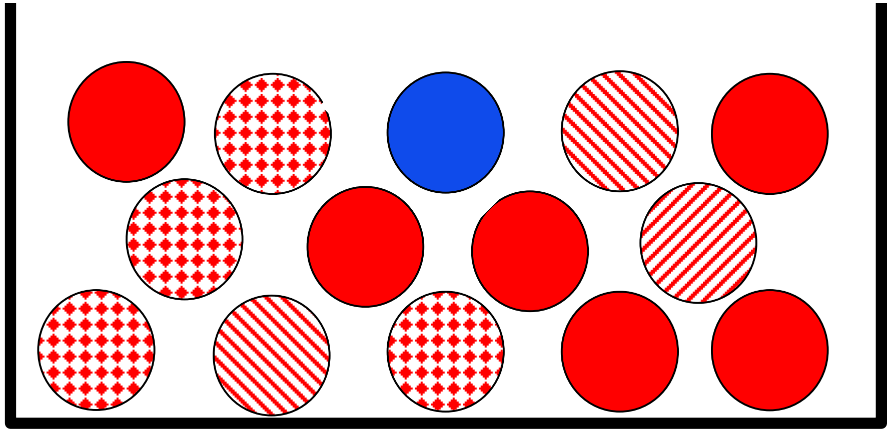
- Mathematically speaking, an urn model describes one or more urns, each containing (a possibly finite, possibly infinite) number of balls which are categorised into one or more colours.

- Balls are successively drawn from urns and may be removed, noted and replaced or replicated.

- Urn models form the basis of most major theories in population biology (the removal of a ball is a death, the replication of one is a birth and the addition of one from another urn is migration) and genetics (passing of a gene from parent to child is the replication of one ball, and a mutation is the addition of a ball from another urn.)

Urn Models

- Consider an urn modelling the competition of species for a common (limited) resource.
- In this simplified scenario, there are enough resources for 14 animals, represented as balls in the urn.
- Each ball is coloured according to the species of the animal it represents. Two balls of the same colour represent animals of the same species.
- At time intervals, one ball is selected at random and removed (‘death’).
- Its space (resources) are then taken by another ball, which is either added from a different urn (‘immigration’) or a second ball is selected from the same urn and duplicated (‘birth’).



Urn Models

- Consider the simple case of two species living in an ecosystem which can sustain five individuals.

- Initially, there are 3 of Species A and 2 of Species B.

- Both species are equally able to compete (i.e. neither is more or less likely to die than the other and neither is more or less likely to be born that the other.)

- Ignore the effects of immigration and assume that the population drifts only due to random births and deaths.

- At each timepoint, the population of Species A drops by one only if one of the Species A individuals is selected to die and then one of the Species B individuals is selected to give birth.

Urn Models

- Let A_n be the number of Species A individuals in the community at timepoint *n*.

- Say, $A_n = 3$, what is $P(A_{n+1} = 4 | A_n = 3)$?

- Species A goes from 3 to 4 individuals in a given timestep only if one of the Species B individuals is selected to die (probability 0.4, since 2 of the 5 individuals in the system are Species B) and one of the Species A individuals is selected to give birth (probability 0.6, since 3 of the 5 individuals in the system are Species A.) Thus $P(A_{n+1} = 4 | A_n = 3) = 0.4 \times 0.6 = 0.24$.

- Similarly, Species A's abundance stays at 3 if either a Species A dies and a Species A is born or if a Species B dies and a Species B is born i.e.

$$P(A_{n+1} = 3 | A_n = 3) = (0.4 \times 0.4) + (0.6 \times 0.6) = 0.52$$

Urn Models

- The transition matrix is therefore

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ (0.2\times0.8) & (0.2\times0.2)+(0.8\times0.8) & (0.2\times0.8) & 0 & 0 & 0 \\ 0 & (0.4\times0.6) & (0.4\times0.4)+(0.6\times0.6) & (0.4\times0.6) & 0 & 0 \\ 0 & 0 & (0.6\times0.4) & (0.4\times0.4)+(0.6\times0.6) & (0.6\times0.4) & 0 \\ 0 & 0 & 0 & (0.8\times0.2) & (0.8\times0.8)+(0.2\times0.2) & (0.8\times0.2) \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- This gives $P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.16 & 0.68 & 0.16 & 0 & 0 & 0 \\ 0 & 0.24 & 0.52 & 0.24 & 0 & 0 \\ 0 & 0 & 0.24 & 0.52 & 0.24 & 0 \\ 0 & 0 & 0 & 0.16 & 0.68 & 0.16 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

Urn Models: Steady-State

- As $n \to \infty$, we see that there are two steady-state solutions of $\Pi_{eq} = \Pi_{eq} P$.
- Either $\Pi_{eq} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ or $\Pi_{eq} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$.
- In other words, either Species A is monodominant and Species B is extinct, or Species B is monodominant and Species A is extinct.
- How, then, are do real ecological populations have any diversity and haven’t (through purely random drift) collapsed to be the same species everywhere?
- Adding an immigration (or “mutation”) term makes the boundaries non-absorbing. All other states are then become persistent.
- Without rare mutations, genetic diversity would collapse and all individuals within populations would be identical.

Time to Most Recent Common Ancestor

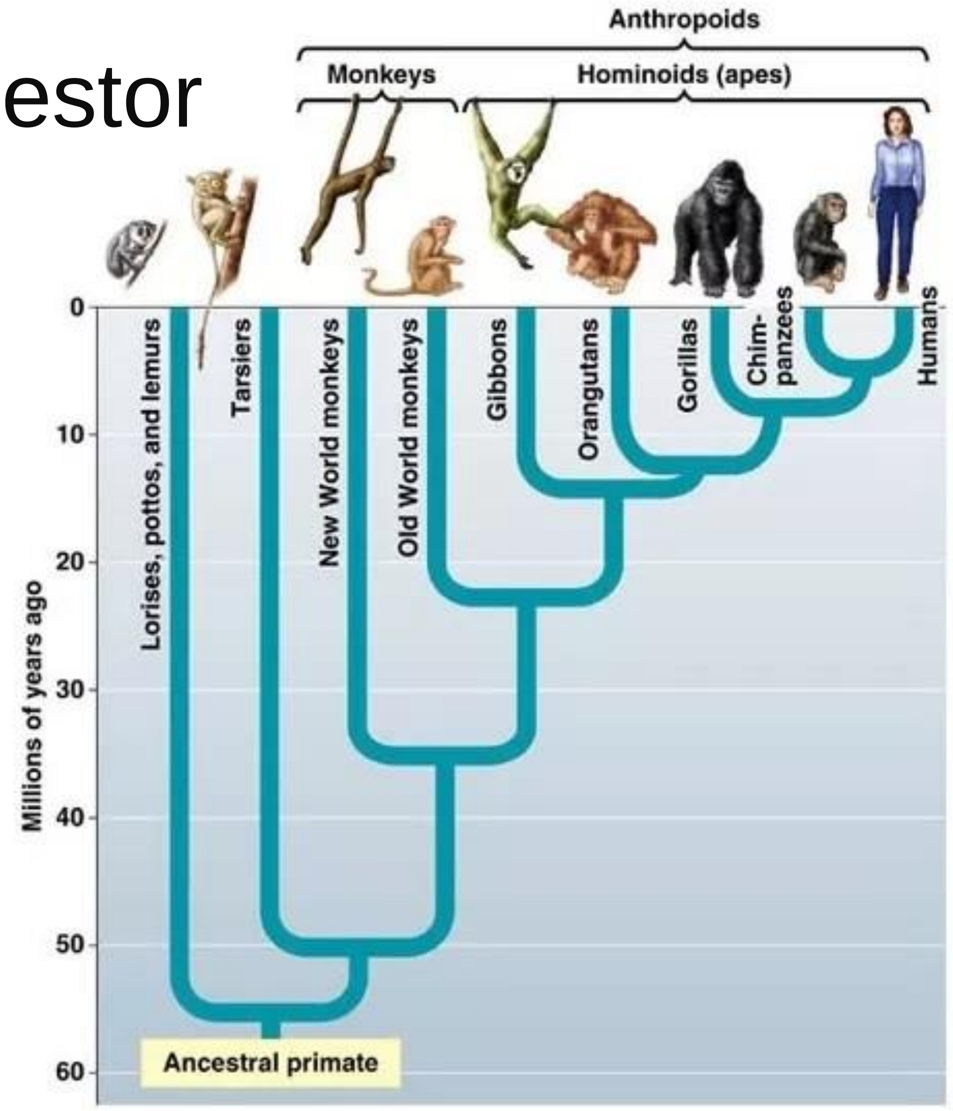
- Much work in genetics is based on analysing the expected behaviour of random walks.
- In a standard inheritance model, genetic material is passed from parent to child with almost (but not quite) 100% similarity between that of the parent and the child. Small mutations occur which randomly perturb this , however.
- Given how similar or dissimilar two individuals are can be used to assess an estimate of the **time to most recent common ancestor**.
- Analysing the Y chromosome of two brothers, for example, should reveal that their common ancestor (i.e. their father) is only one generation back.
- Analysing the Y chromosome of two male cousins, for example, should be a little less similar, as their last common ancestor would be two generations back (i.e. their grandparents.)

Time to Most Recent Common Ancestor

- By looking at how dissimilar two individuals are, and knowing the rate at which genetic lines diverge, estimates can then be made of time to most recent common ancestor.

- This can be done to reconstruct the natural history and development of different species.

- It can also be used to assess possible paternity or maternity cases.



Richard III

- King Richard III of England was the last English King to die in battle.

- He was killed at the Battle of Bosworth Field in 1485. His body was reported buried at a nearby monastery, Greyfriars, which was destroyed around 50 years later and its location was forgotten.

- In 2012, archaeologists began digging up a carpark in Leicester, in the belief that the ruins of Greyfriars would be beneath.

Richard III

- Later accounts, including Shakespeare’s play *Richard III* indicated that Richard III was “hunchbacked” or had some severe spinal deformity.
- Upon digging up the carpark, the following was found:

Richard III

- Richard III left no confirmed living descendants. His only son died aged 10.

- Historians, however, traced his sister’s family tree and located his living 14th cousin, a cabinet maker from Canada called Michael Ibsen.

- The genetic similarity of Richard III and Michael Ibsen was analysed and the dissimilarity was assessed based on how far the genetic line was expected to have diverged over 15 generations to their last common ancestor.

Richard III

- Analysing the divergence in genetic similarity down the female line gave a probability that Michael Ibsen and Richard III were related of around 99.999%.

- Combined with the spinal deformities and description of post-mortem wounds, it was concluded that the body found in 2012 was indeed Richard III. It was reburied in Leicester Cathedral in 2015.