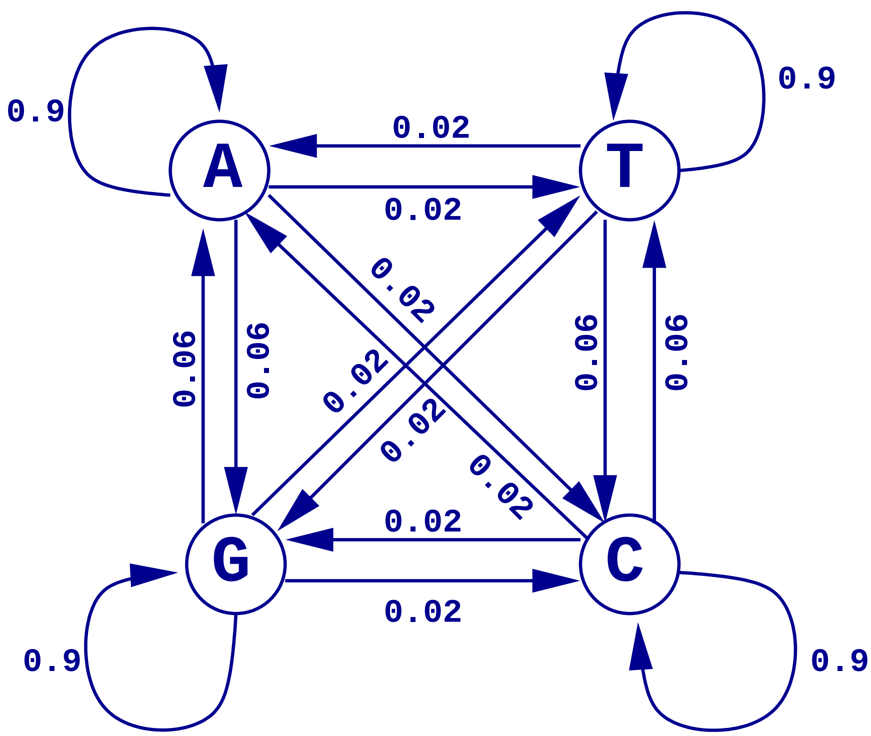

Purines A,G
Pyrimidines C,T

No Action 0.9
Transition 0.06
Transversion 0.02

Some Questions:

- If we know that $X_n = A$, what do we expect X_{n+1} to be?
- If we know that $X_{n-3} = T$, what do we expect X_{n+1} to be?
- If we know that $X_{n-3} = T$ and that$X_n = A$, what do we expect X_{n+1} to be?

- In general, past outcomes contain information on future outcomes.
- The older an outcome is the less it affects the future.
- But, if we know the present state (character) X_n then all past states (X_{n-3}) have absolutely no influence on future outcomes (X_{n+1}).
  (That is the way the experiment is designed)
- Or in other words:
  The future is independent of the past given the present.
- Stochastic processes with this property are called Markov Processes.

- The experiment is driven by the conditional probabilities $P[X_{n+1} = s | X_n = x]$. These probabilities are called transition probabilities.
- In the experiment the process could be in four different states: A,T,G or C. In general the set of possible states of a Markov Process is called its state space.
- If the state space consists of a finite or countable number of states the process is called Markov Chain.

- We assume that for all n and m

$$P[X_{n+1} = j | X_n = i] = P[X_{m+1} = j | X_m = i]$$

holds. ‘‘We do not dream up a different experiment for each step, but use the same conditional probabilities for all of them.’’

- If the state space is finite, we can enumerate the states by numbers $1, 2, \ldots n$ and summarize all transition probabilities in a $n \times n$ matrix $P = (p_{ij})$. Where

$$p_{ij} = P[X_{n+1} = j | X_n = i]$$

This matrix is called transition matrix

- Since the entries are probabilities $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$ hold.

- We need to specify with which state we want to start the chain:
- This can be done in a deterministic way by naming the state explicitly.
- Or it can be done in a stochastic way by choosing the initial state randomly.
- Let μ_i^0 denote the probability that the chain starts in state i
- The vector $(\mu_1^0, \ldots, \mu_n^0)$ is called the start distribution.

- By definition $P[X_0 = i] = \mu_i^0$ holds. We write $X_0 \sim \mu^0$. But ...
- ...what are the distributions of X_1, X_2 or X_n?

- X_1 first:
  Assume we start in state 1 ('A'): The probability of this event is
  $P[X_0 = A] = \mu_1^0$
  Now assume we go to state 2 ('C'): The probability is
  $P[X_0 = A]P[X_1 = C|X_0 = A]$
  In total the probability of having a 'C' in the second step is

$$\sum_{l \in \{A,T,C,G\}} P[X_0 = l]P[X_1 = C|X_0 = l]$$

$$= \sum_i \mu_i^0 p_{i2} =: \mu_2^1$$

- Or more general, using matrix and vector notation:

$$\mu^1 = \mu^0 P \quad \text{and} \quad X_1 \sim \mu^1$$

- What about the distribution of X_2?

- $X_2 \sim ?$

$$\begin{aligned} P[X_2 = l] &= \sum_k P[X_1 = k] P[X_2 = l | X_1 = k] \\ &= \mu^1 P \\ &= (\mu^0 P) P = \mu^0 P^2 \end{aligned}$$

- Or more general for X_n:

$$X_n \sim \mu^n = \mu^0 P^n$$

- $P(n) = P^n$ is called n-step transition matrix.

- Chapman-Kolmogorov equation:

$$\begin{aligned} p_{ij}(m+n) &= \sum_k p_{ik}(m) p_{kj}(n) \\ P(n+m) &= P(n) P(m) \end{aligned}$$

- Let us restrict our studies of Markov chains to chains with only strictly positive entries in the transition matrix.
- This means: Every transition from any state to any other state is possible in a single step.
- This is a very strong assumption, and in many typical applications of Markov chain models it does not hold.
- However, for the discussion of evolution models it is ok.
- In the literature, you will find a lot of theory on Markov chains with zeros in the transition matrix. We can skip it.

- Given that the chain starts in 'A': $\{X_0 = A\}$
- This has a strong influence on the distribution of X_1. Most likely we will have an 'A' there too and a Purine is more likely then a Pyrimidine.
- The effect on the distribution of X_2 is similar but less strong. There have been 2 random experiments that might have changed the state.
- For growing n the influence of X_0 on X_n becomes less and less.

- What happens in the limit?
- Is X_∞ independent of X_0?
- Which distribution do we get for X_∞? Can we tell?
- Which states have we observed in the meantime? And how often?
- How do transition probabilities look like for long time periods?

- The transition matrix P has strictly positive entries:
- Of course, the actual path of the Markov chain never converges. We keep on doing the random experiments and hence we will keep on changing states.
- The long term transition probabilities converge:

$$p_{ij} \longrightarrow \pi_j$$

This means the columns of P all converge to a single distribution vector π.

- The distribution of the X_n also converge to π:

$$\mu_i^n \longrightarrow \pi_i.$$

- If we examine a single path of the chain (like:AAAAGGTTTTTCCTTCCA...): The proportion of time we spent in state i converges to π_i.

  This result is one of the central results of probability theory and is know as The ergodic theorem.

- What is this distribution π?

- Let n be large, and $\{X_n = i\}$ is a shortcut for X_n is in the n'th state:

-

$$
\begin{aligned}
P[X_{n+1} = j] &= \sum_i P[X_n = i] P[X_{n+1} = j | X_n = i] \\
&= \sum_i P[X_n = i] p_{ij}
\end{aligned}
$$

- Taking the limit $n \to \infty$ we get:

$$\pi_j = \sum_i \pi_i P_{ij}$$

or

$$\pi = \pi P$$

- π is a fix-point of the linear transformation associated with P.

- If the Markov chain is in distribution π it will be in π forever.

- π is the unique solution of

$$\begin{aligned} \pi &= \pi P \\ \sum_i \pi_i &= 1. \end{aligned}$$

- π is called stationary distribution.
- If all $X_0, X_1, \ldots$ have distribution π we say that the Markov chain is in equilibrium.
- The Markov chain converges to π no matter what the start distribution was. Hence, the initial information gets lost.

- In the initial example the stationary distribution is the uniform distribution: $\pi = (1/4, 1/4, 1/4, 1/4)$.
- That is not always the case. For example:

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

leads to

$$\pi = (0.5714, 0.4286)$$

Time reversal

- Start at some time n and trace the Markov Chain backwards in time: That is, consider the sequence $X_n, X_{n-1}, X_{n-2}, \ldots$
- It turns out, that this sequence is again a Markov chain.
- Is it the same one?
- What are its transition probabilities?

- Assume the chain is in equilibrium for the rest of this lecture.
- Let us call the transition matrix of the time reversed Markov chain P^-_{ij}.
- We have

$$\begin{aligned} P^-_{ij} &= P[X_n = j | X_{n+1} = i] \\ &= \frac{P[X_n = j; X_{n+1} = i]}{P[X_{n+1} = i]} \\ &= \frac{P[X_n = j] P[X_{n+1} = i | X_n = j]}{P[X_{n+1} = i]} \\ &= \frac{\pi_j P_{ji}}{\pi_i} \end{aligned}$$

- We have $P^-_{ij} = P_{ij}$ if the detailed balance equation

$$\pi_i P_{ij} = \pi_j P_{ji}$$

holds.

- Clearly this implies also

$$\pi_i P(m)_{ij} = \pi_j P(m)_{ji}$$

- Let M^1_{ij} denote the joint distribution of two adjacent variables X_n and X_{n+1}, and M^m_{ij} the joint distribution of X_n and X_{n+m}

- $$M^m_{ij} = \pi_i P(m)_{ij}$$

  and

  $$M^m_{ji} = \pi_j P(m)_{ji}$$

- Hence, detailed balance corresponds to symmetric joint probabilities.

- If detailed balance holds we say that the chain is time reversible.

- Up to now, we have assumed that the Markov chain operates in discrete time steps. Conditional on the state X_n we perform an experiment and this experiment determines the distribution of X_{n+1}.
- Let us become a little more general now, and assume that the chain does not operate in separate steps, but that states can be changed at any time point on the half line $[0, \infty)$.
- Hence the chain is described by a infinite family of random variables $X_t \quad T >= 0$.
- The chain is a jump process:
  E.g. It starts in A, then remains in A for some time, at certain random time point it jumps to 'C' staying in 'C' for some interval of time, before jumping the next time, and so on ...

- The Markov property now reads:
  For any sequence of time points
  $t_1 <, \ldots, < t_n$ we have

$$P[X_{t_n} = j | X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{t_{n-1}}] = P[X_{t_n} = j | X_{t_{n-1}} = i_{t_{n-1}}]$$

- The future is independent of the past given the latest news.
- We now have a time continuous family of transition probabilities

$$p(t)_{ij} = P[X_s = j | X_{s-t} = i]$$

Standard Markov chains

- We assume that

$$\lim_{t \to 0} p(t)_{ii} = 1$$

and

$$\lim_{t \to 0} p(t)_{ij} = 0 \quad \texttt{for} \quad i \neq j.$$

That means, the first change of state needs at least a little bit of time.

- And all following changes are separated by each other by a maybe very small but not empty interval.

- No two events happen at the same time.

- In matrix notation:

$$P_t \longrightarrow I$$

where I is the identity matrix.

- For time discrete Markov chains we have developed the whole theory from the one-step transition matrix P.
- Problem for time continuous Markov chains, there is no natural unit of time.
- Which expression should play the role of P?

- Suppose the chain is in state $X_t = i$ at time t.
- What can happen in a short time interval $[t, t+h]$?
  1. Nothing, with probability $p(h)_{ii}$
  2. A single change to state j, with probability $p_(h)ij$
  3. More than one changes of state.
- The probability of two ore more changes of states is $o(h)$

Rates

- It turns out that $p(h)_{ij}$ is approximately linear for small h.
- Hence there are numbers q_{ij} such that

$$p(h)_{ij} \approx q_{ij}h \quad \texttt{for} i \neq j$$

  and

$$p(h)_{ii} \approx 1 + q_{ii}h$$

- Clearly $q_{ij} > 0$ and $q_{ii} < 0$
  (We still assume $p(1)_{ij} > 0$.)
- These numbers q_{ij} are called rates.
  They form the rate matrix

$$Q = (q_{ij}).$$

- The function $t \mapsto p(t)_{ij}$ is locally linear at $t = 0$.
- In other words the function is differentiable at $t = 0$.

$$Q = \lim_{h \to 0} \frac{P(h) - P(0)}{h}$$

- By moving from $t = 0$ to an arbitrary time point we get
  The forward equation and backward equation

$$\frac{d}{dt}P(t) = P(t)Q = QP(t)$$

- By solving these differential equations under the initial condition $P(0) = I$, we get

$$P(t) = \exp(tQ) = \sum_{n=0} \frac{Q^n t^n}{n!}.$$

- Hence we get a ''t-step'' rate matrix simply by tQ.

- We have $P(1) = \exp(Q)$ and hence $Q = \log(P(1))$.
- What is are the exponential and the logarithm of a matrix?
- How do we calculate them?

- Every normal ($A^T A = AA^T$) square matrix A can be decomposed into

$$A = S \begin{pmatrix} \lambda_1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \lambda_n \end{pmatrix} S^{-1}.$$

  where $\lambda_1, \ldots \lambda_n$ are the possibly complex eigenvalues of A and S consists of the orthonormal basis of eigenvectors of A.

- both P and Q are normal.

- By reversibility it follows, that the eigenvalues of P are positive real numbers.
- There is a joint orthonormal basis of P and Q, and

$$Q=\log(P(1))=S\begin{pmatrix} \log(\lambda_1) & & & & \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & \log(\lambda_n) \end{pmatrix}S^{-1}$$

where $(\lambda_1 \dots \lambda_n)$ are the positive eigenvalues of $P(1)$.

The resolvent

- For $\alpha > 0$, we define a weighted time average of $P(t)$:

$$R_\alpha = \int_0^\infty e^{-\alpha t} P(t)\, dt.$$

- R_α is called a resolvent of $P(t)$.
- The resolvent is related to the rate matrix by

$$\alpha I - R_\alpha^{-1} = Q \quad \text{for all} \quad \alpha > 0.$$