Markov Chains, Reversibility and Entropy Production

Keith Johnson

March 8, 2000

1 Forward and Backward Markov Chains

A (first order) Markov chain generates a sequence by randomly choosing the next character based on the previous character. If we choose to look at the sequence in the order of it being produced from left to right, we have a forward Markov chain. If we think of the generating process as going from right to left (so that the last character produced will be the first character in the resulting sequence) we have a backward Markov chain.

Despite the one-directional nature of producing a Markov chain, it is still true that the probability of having a certain character in the i^{th} position depends on the character in both the $i-1$ and the $i+1$ positions. This will be shown first with an example, and then more formally.

1.1 Example

Consider a first-order one-directional Markov chain moving from left to right generating the sequence ... CGG ... Imagine that you forgot what the middle character was: ... C□G ...

Suppose you want to figure out what the missing character was, and you don't have time to re-run the Markov chain on the whole string. To do this correctly, you need to consider both the character before and the character after the missing location.

For example, suppose the transition probability matrix is:

	C	A	G	T
C	0	0	.5	.5
A	.25	.25	.25	.25
G	0	0	1	0
T	0	0	0	1

Then, we know that the missing character would have to be either a T or a G, since those are the only characters that can come after a C. If the character had been a T, then the following character would also have to be a T. But, the following character is a G. So, the missing character must have been a G as well.

This is an extreme case. In a more realistic distribution, we would only be able to derive the probabilities of G or T being the missing character. In this example,

$$\begin{aligned}\Pr(X_i = G \mid X_{i-1} = C, X_{i+1} = G) &= 1 \\ \Pr(X_i = T \mid X_{i-1} = C, X_{i+1} = G) &= 0\end{aligned}$$

so we can be certain that G is the missing character.

Also, notice that these probabilities are different from what we get when considering only the character to the left. These probabilites can be read from the transition probability matrix.

$$\begin{aligned}\Pr(X_i = G \mid X_{i-1} = C) &= 0.5 \\ \Pr(X_i = T \mid X_{i-1} = C) &= 0.5\end{aligned}$$

1.2 Formal Approach

Let X_l, X_m, X_r represent the three nucleotides, where X_l and X_r are known (in the previous example, X_l=C and X_r=G), and X_m is unknown.

We are interested in the probability distribution $\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i+1} = X_r)$ In particular, we are asking whether $\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i+1} = X_r) = \Pr(X_i = X_m \mid X_{i-1} = X_l)$. We will use the fact that

$$\Pr(X_{i+1} = X_r \mid X_{i-1} = X_l, X_i = X_m) = \Pr(X_{i+1} = X_r \mid X_i = X_m)$$

which is the definition of a first order Markov chain.

$$\begin{aligned}
&\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i+1} = X_r) \\
= \; & \frac{\Pr(X_i = X_m, X_{i-1} = X_l, X_{i+1} = X_r)}{\Pr(X_{i-1} = X_l, X_{i+1} = X_r)} \\
= \; & \frac{\Pr(X_{i+1} = X_r \mid X_{i-1} = X_l, X_i = X_m)\Pr(X_{i-1} = X_l, X_i = X_m)}{\Pr(X_{i-1} = X_l, X_{i+1} = X_r)} \\
= \; & \frac{\Pr(X_{i+1} = X_r \mid X_i = X_m)\Pr(X_{i-1} = X_l, X_i = X_m)}{\Pr(X_{i-1} = X_l, X_{i+1} = X_r)} \\
= \; & \frac{\Pr(X_{i+1} = X_r \mid X_i = X_m)\Pr(X_i = X_m \mid X_{i-1} = X_l)\Pr(X_{i-1} = X_l)}{\Pr(X_{i-1} = X_l, X_{i+1} = X_r)} \\
= \; & \frac{\Pr(X_{i+1} = X_r \mid X_i = X_m)\Pr(X_i = X_m \mid X_{i-1} = X_l)\Pr(X_{i-1} = X_l)}{\sum_Y \Pr(X_{i+1} = X_r \mid X_i = Y)\Pr(X_i = Y \mid X_{i-1} = X_l)\Pr(X_{i-1} = X_l)} \\
= \; & \frac{\Pr(X_{i+1} = X_r \mid X_i = X_m)\Pr(X_i = X_m \mid X_{i-1} = X_l)}{\sum_Y \Pr(X_{i+1} = X_r \mid X_i = Y)\Pr(X_i = Y \mid X_{i-1} = X_l)}
\end{aligned}$$

where the sum is taken over all possible values Y that the missing character could be.

So $\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i+1} = X_r)$ depends on $\Pr(X_{i+1} = X_r \mid X_i = X_m)$ as well as $\Pr(X_i = X_m \mid X_{i-1} = X_l)$. (Note: $\Pr(X_{i-1} = X_l)$ and $\Pr(X_{i-1} = X_l, P_{i+1} = X_r)$ are fixed because X_l and X_r are known.) Hence, the probability distribution for a given character depends both on the probability of transitioning to it, and the probability of transition away from it. In general,

$$\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i-1} = X_r) \neq \Pr(X_i = X_m \mid X_{i-1} = X_l).$$

In the earlier example, for $X_m = T$,

$$\Pr(X_{i+1} = X_r \mid X_i = X_m) = \Pr(X_{i+1} = G \mid X_i = T) = 0$$

which makes the numerator in the above expression 0, hence

$$\Pr(X_i = X_m \mid X_{i-1} = X_l, X_{i+1} = X_r) = \Pr(X_{i+1} = T \mid X_{i-1} = C, X_{i+1} = G) = 0$$

while

$$\Pr(X_i = X_m \mid X_{i-1} = X_l) = \Pr(X_i = T \mid X_{i-1} = C) = 0.5$$

and this provides our counterexample.

In the above derivation, only the properties of probability and the forward Markov chain are used. There is nothing else more used.

2 Reversibility

It can be shown that the reverse of a Markov chain is again a Markov chain (see reference 1) The forward and reverse Markov chains are indistinguishible if the transition probability matrix is symmetric. In this case, the Markov chain is said to be "reversible"

Further, the relationship between the forward and backward transition probabilities can be derived.

For a stationary discrete time Markov chain $X(t)$ with forward transition probability

$$P_{ji} = \Pr(X(t) = i \mid X(t-1) = j)$$

for any t, and stationary distribution

$$p_i = \Pr(X(t) = i)$$

for any t, the backward transition probability $P'_{ij} = \Pr(X(t) = j \mid X(t+1) = i)$ of the backward Markov chain is determined by

$$P'_{ij} = p_j P_{ji} / p_i.$$

This equation says that knowing the forward transition probabilities and the stationary distribution of a forward Markov chain determines everything about the backward transition probabilities of the reversed chain.

3 Nonreversible Chains and Entropy Production

Whether a Markov chain is reversible or not has everything to do with whether the pairwise distribution matrix is symmetric. If it is nonsymmetric, then the Markov chain is nonreversible. Since we know that our biological data does not, in general, produce a symmetric pairwise probability distribution, we know that the Markov chains that we use are nonreversible.

The term "nonreversible" is used because there is an entropy production that can be detected, and used to determine in which direction the chain was run (see Reference 2).

4 Conclusions

It is too strong a statement to say that the i^{th} character does not at all depend on the $(i+1)^{th}$ character. However, for a Markov chain with a nonsymmetric pairwise distribution (such as the Markov chains used in computational biology), the Markov chain has a nonreversible nature.

5 References

1. Bertsekas and Gallager, *Data Networks*, 2nd edition, 1992, page 215.
2. Kelly, F.P., *Reversibility and Stochastic Networks*, Wiley, New York, 1979.