Lab 8: Changing My Shape, I Feel Like an Accident

36-350, Statistical Computing

4 November 2011

A certain protein molecule has three distinct shapes: coiled, open, and stretched.[1] It is observed every microsecond, with the transition probabilities given in the figure.

1. (10) Write the down the Markov chain transition matrix. Number the three states as indicated in the figure.

2. (15) Explain what the following command does:

```
chain.transition <- function(x,q) {
  stopifnot(is.matrix(q), all(q >= 0), nrow(q)==ncol(q))
  K <- nrow(q)
  stopifnot(all.equal(rowSums(q),rep(1,K)))
  stopifnot(x>=1, x<=K, x==round(x),length(x)==1)
  new.x <- sample(1:K,size=1,prob=q[x,])
  return(new.x)
}
```

3. (15) Write a function for drawing a sequence from a Markov chain by filling in the following code. You will probably want to use a loop, and will definitely want to call `chain.transition`.

```
rmarkov <- function(n,q,x.initial) {
  x <- vector(length=n)
  x[1] <- x.initial
  #### your code here
  return(x)
}
```

4. (10) Generate a sequence of length 20 from the chain whose matrix you found earlier, starting from state 1. Plot its histogram, and use `table()` to find the distribution of the sample.

[1] The numbers which follow are based on a real but unpublished experiment.

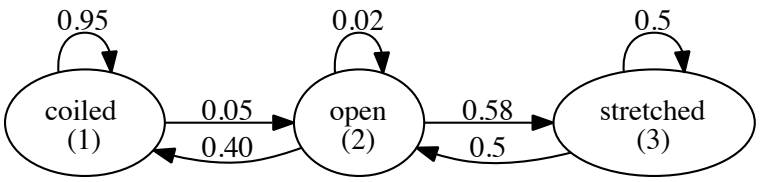

Figure 1: Observed transition rates, per microsecond, between the three main "conformations" of a certain protein.

5. (10) Generate a sequence of length 20, starting from state 2. Use a histogram and `table()`, as before, to find the distribution of this sample. Explain how, and why, it differs from the previous one.

6. (15) Generate sequences of length 100,000, starting from state 1 and state 2. Plot their histograms. Comment on the differences between these, and between these and the previous histograms.

7. (15) Use `eigen()` to find the invariant distribution. How does this relate to the histograms? (*Hint 1*: by default, `eigen()` finds the right eigenvectors of a matrix; here we want the left eigenvectors. How do we get them? *Hint 2*: You will need to re-scale the eigenvector so its entries sum to 1.)

8. (10) Use the second largest eigenvalue to estimate how many steps are needed for the distribution to come within a distance of 0.001 of the equilibrium. How long does this suggest that we have to wait for the protein to have forgotten its initial state?