# Discrete Time Markov Chains: Steady State Distributions

Suppose that bases (letters) in DNA sequences can be modeled as a Markov chain with state space (A, C, G, T) and transition matrix

```
P = rbind(
  c(0.36, 0.15, 0.20, 0.29),
  c(0.41, 0.16, 0.04, 0.39),
  c(0.30, 0.21, 0.14, 0.35),
  c(0.31, 0.19, 0.18, 0.32)
)
```

| | A | C | G | T |
|---|---|---|---|---|
| **A** | 0.36 | 0.15 | 0.20 | 0.29 |
| **C** | 0.41 | 0.16 | 0.04 | 0.39 |
| **G** | 0.30 | 0.21 | 0.14 | 0.35 |
| **T** | 0.31 | 0.19 | 0.18 | 0.32 |

1. One of the following is the unique stationary distribution. Identify which one it is and explain your reasoning conceptually without doing any calculations.

$$\boldsymbol{\pi}_1 = [0.1742, 0.3430, 0.3266, 0.1562]$$
$$\boldsymbol{\pi}_2 = [0.3430, 0.1742, 0.1562, 0.3266]$$
$$\boldsymbol{\pi}_3 = [0.4935, 0.0093, 0.0057, 0.4915]$$

2. Now use software to find the stationary distribution.

3. Find the probability that C is followed two letters later by T.

4. Find the probability that A is followed two letters later by T.

5. Find the probability that C is followed 100 letters later by T.

6. Find the probability that A is followed 100 letters later by T.

7. Find the probability that a three letter sequences spell “CAT”.