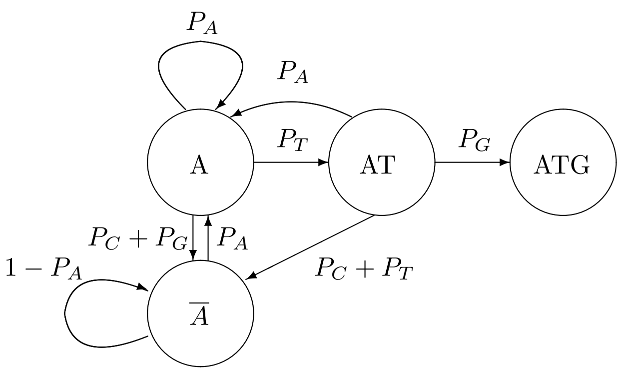
BCB BCB/GDCB/STAT/COM S 568 Spring 2011

Homework 3 Solution

February 2, 2011

1. Assume we have a random sequence S over the nucleotides A, T, C, G where $p(A) = p(B) = p(C) = p(G) = 0.25$. What is the expected waiting time to see the start codon "ATG"?.

   Solution:
   We can use first step analysis to find the expected waiting time:



$$
\begin{aligned}
\mu &= 1 + P_A\mu_A + (1 - P_A)\mu_{\overline{A}} \\
\mu_A &= 1 + P_A\mu_A + (P_C + P_G)\mu_{\overline{A}} + P_T\mu_{AT} \\
\mu_{\overline{A}} &= 1 + P_A\mu_A + (1 - P_A)\mu_{\overline{A}} \\
\mu_{AT} &= 1 + P_A\mu_A + (P_C + P_T)\mu_{\overline{A}} + P_G\mu_{ATG} \\
\mu_{ATG} &= 0
\end{aligned}
$$

   we solve these equations, and get:

$$\mu = \frac{1}{P_A P_T P_G} = \frac{1}{0.25^3}$$

2. We have a random sequence over the alphabet $\{R, Y\}$ where R and Y are drawn with equal probability 0.5 at each step. What is the expected waiting time to find the word "$YRYR$" and "$RYRR$"?
   Solution:

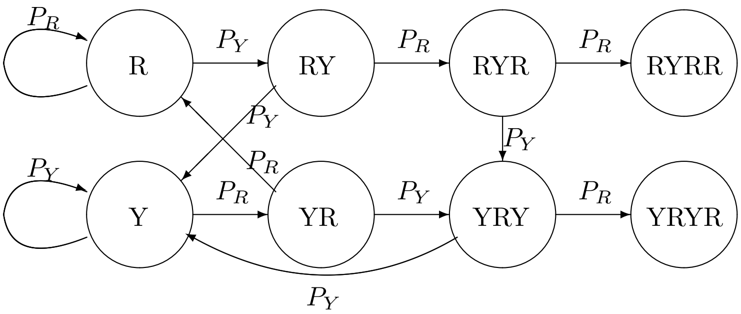
   We can use first step analysis or renewal theory to solve this problem. Here we use the renewal theory.
   Suppose $P(R) = p, P(Y) = q$
   Let u_n be the probability $YRYR$ completes at n, and μ be the expected waiting time.
   Then we have:

$$
\begin{aligned}
p^2q^2 &= u_n + u_{n-2}pq \\
&= \frac{1}{\mu} + \frac{1}{u}pq \\
&= \frac{1}{\mu}(1 + pq) \\
\mu &= \frac{1 + pq}{p^2q^2} \\
&= 20
\end{aligned}
$$

Similarly, we can find the expected waiting time of pattern $RYRR$:

$$\begin{aligned}
p^3q^2 &= u_n + u_{n-3}p^2q \\
&= \frac{1}{\mu} + \frac{1}{u}p^2q \\
&= \frac{1}{\mu}(1+p^2q) \\
\mu &= \frac{1+p^2q}{p^3q} \\
&= 18
\end{aligned}$$

3. What is the probability that the pattern $RYRR$ occurs before $YRYR$?
Solution:



Let μ_x be the probability that the pattern "RYRR" occurs before "YRYR" when the current state is x.

$$\begin{cases}
\mu_R &= P_R \times \mu_R + P_Y \times \mu_{RY} \\
\mu_{RY} &= P_R \times \mu_{RYR} + P_Y \times \mu_Y \\
\mu_{RYR} &= P_R \times \mu_{RYRR} + P_Y \times \mu_{YRY} \\
\mu_{RYRR} &= 1 \\
\mu_Y &= P_Y \times \mu_Y + P_R \times \mu_{YR} \\
\mu_{YR} &= P_Y \times \mu_{YRY} + P_R \times \mu_R \\
\mu_{YRY} &= P_Y \times \mu_Y + P_R \times \mu_{YRYR} \\
\mu_{YRYR} &= 0
\end{cases}
\Longrightarrow
\begin{cases}
\mu_R &= 1/2\mu_R + 1/2\mu_{RY} \\
\mu_{RY} &= 1/2\mu_{RYR} + 1/2\mu_Y \\
\mu_{RYR} &= 1/2\mu_{RYRR} + 1/2\mu_{YRY} \\
\mu_{RYRR} &= 1 \\
\mu_Y &= 1/2\mu_Y + 1/2\mu_{YR} \\
\mu_{YR} &= 1/2\mu_{YRY} + 1/2\mu_R \\
\mu_{YRY} &= 1/2\mu_Y + 1/2\mu_{YRYR} \\
\mu_{YRYR} &= 0
\end{cases}$$

$$\Longrightarrow
\begin{cases}
\mu_Y &= 2/7 \\
\mu_R &= 3/7 \\
\mu_{YR} &= 2/7 \\
\mu_{RY} &= 3/7 \\
\mu_{RYR} &= 4/7 \\
\mu_{YRY} &= 1/7
\end{cases}
\Longrightarrow \mu = \frac{1}{2}\mu_Y + \frac{1}{2}\mu_R = \frac{1}{2}(\frac{2}{7} + \frac{3}{7}) = \frac{5}{14}$$