STAT 515
Homework #3 WITH SOLUTIONS

1. Suppose that a population consists of a fixed number, $2m$, of genes in any generation. Each gene is one of two possible genetic types. If any generation has exactly i (of its $2m$) genes of type 1, then for any $0 \le j \le 2m$, the next generation will have exactly j genes of type 1 with binomial probability

$$\binom{2m}{j}\left(\frac{i}{2m}\right)^{j}\left(\frac{2m-i}{2m}\right)^{2m-j}.$$

Let X_n denote the number of type 1 genes in the nth generation, and assume $X_0 = m$.

(a) Find $E(X_n)$.

Solution: Conditioning shows that

$$E(X_n) = E[E(X_n \mid X_{n-1})] = E[X_{n-1}]$$

because conditional on X_{n-1}, X_n is binomial$(2m, X_{n-1}/2m)$ and therefore the conditional expectation of X_n given X_{n-1} is $2m \times (X_{n-1}/2m)$. We conclude that every X_n has the same expectation, and since X_0 is the constant m, $E(X_n)$ must be m.

(b) Suppose that $m = 6$. What is the probability that $X_n = m$ for some $n > 0$?

Solution: This problem is quite difficult without using a computer. The strategy is to first construct the 13×13 transition matrix, then define P_T to be the 11×11 submatrix that remains after removing the first and last columns and rows (leaving only the entries corresponding to moves from transient states to other transient states). Then, we calculate $S = (I - P_T)^{-1}$, whose entries give the expected number of generations spent in each transient state, conditional on the starting transient state.

```
> P <- matrix(0,13,13) # Initialize transition matrix
> for (i in 1:13) {
+   P[i,] <- dbinom(0:12, 12, (i-1)/12) # Each row uses dbinom with a different p
+ }
> id <- diag(rep(1, 11)) # Construct identity matrix
> S <- solve(id - P[2:12, 2:12])     # S has info on transient states 1 through 11
```

Using an argument developed in Section 4.6,

$$P(\text{re-entering state 6} \mid \text{starting in state 6}) = \frac{S_{66} - 1}{S_{66}},$$

So the probability in this case is given by:

```
> (S[6,6]-1)/S[6,6]
[1] 0.6208155
```

(c) If $m = 6$, what is the expected number of generations in which all genes except one are of the same type?

Solution: The states corresponding to "all genes but one of the same type" are states 1 and 11. Therefore, using the S matrix calculated in part (b), we obtain as an answer

```
> S[6,1] + S[6,11]
[1] 1.816812
```

2. A transition matrix P is called *doubly stochastic* if each of its column sums equals one.

(a) If an irreducible, aperiodic Markov chain has finitely many states and its transition matrix is doubly stochastic, prove that its limiting probability distribution is discrete uniform.

Solution: Since a finite, irreducible, aperiodic chain must be ergodic, we know that there is a unique solution to the equations $\pi_j = \sum_i \pi_i P_{ij}$. Therefore, it suffices to show that if $\pi_j = \pi_i$ for all i and j, then these equations are satisfied. However, this is immediate, since if $\pi_i = \pi_j$ then we obtain

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

because $\sum_i P_{ij} = 1$ by the assumption of double stochasticity.

(b) Find a doubly stochastic transition matrix P for a Markov chain with three states such that every entry of P is a different integer multiple of $1/12$ and such that $P_{11} = 0$ and $P_{22} = 1/2$. Calculate the matrix P^{10} (i.e., the tenth power of P). Explain why all nine entries of P^{10} should be nearly the same.

Solution: If we label the states 1, 2, and 3 (instead of 0, 1, and 2), then the only matrices that work here are P and its transpose, where

$$P = \frac{1}{12}\begin{bmatrix} 0 & 5 & 7 \\ 4 & 6 & 2 \\ 8 & 1 & 3 \end{bmatrix}.$$

(Can you argue that these two are the only possibilities?) For P^{10}, we obtain

```
> P <- matrix(c(0, 4, 8, 5, 6, 1, 7, 2, 3), 3, 3) / 12
> P2 <- P %*% P
> P4 <- P2 %*% P2
> print(P10 <- P2 %*% P4 %*% P4)
          [,1]      [,2]      [,3]
[1,] 0.3349768 0.3327671 0.3322561
[2,] 0.3330226 0.3334437 0.3335336
[3,] 0.3320006 0.3337892 0.3342103
```

We know from part (a) that $\lim_{n\to\infty} P^n$ exists and that each row of this limit equals the discrete uniform $(1/3, 1/3, 1/3)$. Here, we only take $n = 10$, but the theory tells us that P^n will get closer and closer to the limiting matrix in which each entry equals $1/3$.

3. Consider three urns, one colored red, one white, and one blue. The red urn contains 1 red and 3 blue balls; the white urn contains 3 white balls, 2 red balls, and 1 blue ball; the blue urn contains 4 white balls, 3 red balls, and 2 blue balls. At the initial stage, a ball is randomly selected from the red urn and then returned to that (red) urn. At every subsequent stage, a ball is randomly selected from the urn whose color is the same as that of the ball previously selected and is then returned to the urn from which it was drawn.

(a) Explain why this process is a Markov chain, then define an appropriate transition probability matrix to describe it.

Solution: Label the states 1 = red, 2 = white, and 3 = blue and let the stochastic process X_n be the label of the nth ball drawn, where X_0 is the first ball picked (from the red urn). Since the color of X_n has a distribution depending only on the urn from which the ball is picked, and this urn is completely determined by X_{n-1}, the Markovian property is satisfied by the process $X_0, X_1, \ldots$.

The transition probabilities are given by

$$P = \begin{bmatrix} 1/4 & 0 & 3/4 \\ 1/3 & 1/2 & 1/6 \\ 1/3 & 4/9 & 2/9 \end{bmatrix} = \frac{1}{36}\begin{bmatrix} 9 & 0 & 27 \\ 12 & 18 & 6 \\ 12 & 16 & 8 \end{bmatrix}.$$

(b) Does this process have a stationary distribution? Justify your answer.

Solution: This Markov chain is irreducible since there exists a path from i to j for any i and j. It is aperiodic because it is irreducible and $P_{11} > 0$. Finally, it is positive recurrent because it is irreducible and there are only finitely many states. So it is irreducible and ergodic, which means that it has both a limiting distribution and a stationary distribution, and these two coincide. (Technically, only the irreducibility and positive recurrence are needed to have a stationary distribution.)

(c) Explain why this process has a limiting distribution.

Solution: See part (b).

(d) In the long run, what proportion of the selected balls are red? What proportion are white? What proportion are blue?

Solution: One way to find π satisfying $\pi^\top = \pi^\top P$ is to search for a left-eigenvector whose eigenvalue equals one (i.e., search for a right-eigenvector of $P^\top$):

```
> P <- matrix(c(9, 12, 12, 0, 18, 16, 27, 6, 8), 3, 3) / 36
> print(e <- eigen(t(P)))
$values
[1]  1.00000000 -0.08333333  0.05555556

$vectors
           [,1]       [,2]          [,3]
[1,] -0.5314934  0.1860807  2.530214e-17
[2,] -0.5627577  0.5954583  7.071068e-01
[3,] -0.6331024 -0.7815391 -7.071068e-01
```

Since the first eigenvalue is one, we want the first column of the eigenvector matrix, normalized so that its entries sum to one:

```
> pi <- as.real(e$vec[,1])
> print( pi <- pi/sum(pi))
[1] 0.3076923 0.3257919 0.3665158
```

These are the long-run proportions of red, white, and blue, respectively.

(e) Simulate a Markov chain of length 100,000 using the information provided above and count the proportion of times the chain was in each of the states. Compare this to your answer.

Solution:

```
> X <- rep(0, 100000) # This is where the realizations will be stored
> pi0 <- P[1,] # The original proportions are in row 1 (the red urn's makeup)
> currentState <- sample(3, 1, prob=pi0)
> for(j in 1:100000) { # Sample according to the correct row in P
+   X[j] <- currentState <- sample(3, 1, prob=P[currentState,])
+ }
> # Now summarize the states visited by proportion:
> rbind (observed=table(X) / 100000, theoretical=pi)
                    1         2         3
observed    0.3092900 0.3235600 0.3671500
theoretical 0.3076923 0.3257919 0.3665158
```

Clearly the observed proportions are close to the theoretical stationary probabilities.

(f) Suppose you have taken 4 steps, i.e., you start with the initial distribution to obtain X_0 and use the transition probability matrix above to obtain state X_4 of the Markov chain. What proportion of times would you expect X_4 to be red, white, and blue, respectively?

Solution: Starting from $\pi_0^\top = (1/4, 0, 3/4)$ (the proportions in the red urn), we obtain

$$\pi_0^\top P^4 = \begin{bmatrix} \frac{2067201}{6718464} & \frac{2188752}{6718464} & \frac{2462511}{6718464} \end{bmatrix},$$

but of course the exact answers are really messy and the numerical values are fine here:

```
> print (pi4 <- P[1,] %*% P %*% P %*% P %*% P)
          [,1]      [,2]      [,3]
[1,] 0.3076895 0.3257816 0.3665289
```

(g) Now simulate 10,000 realizations of the random variable X_4 using the initial distribution and transition probability matrix for this process. Calculate the proportion of times in your simulations that X_4 is red, white, and blue. Compare these proportions to your theoretically obtained answers above.

Solution:

```
> X4 <- rep(0, 10000) # This is where the realizations will be stored
> pi0 <- P[1,] # The original proportions are in row 1 (the red urn's makeup)
> for (i in 1:10000) {
+   currentState <- sample(3, 1, prob=pi0)
+   for (j in 1:4) { # Sample according to the correct row in P
+     currentState <- sample(3, 1, prob=P[currentState,])
+   }
+   X4[i] <- currentState
+ }
> # Now summarize the states visited by proportion:
> rbind (observed=table(X4) / 10000, theoretical=pi4)
                 1         2         3
observed 0.3083000 0.3223000 0.3694000
         0.3076895 0.3257816 0.3665289
```

Clearly the observed proportions are close to the theoretical stationary probabilities. If we wished, it would be easy to add confidence intervals to our sample proportions since these 10,000 realizations were generated independently.

4. Suppose that in a branching process, the expected number of offspring of a given individual equals 4/5. Find the expected number of individuals that ever exist in this population, assuming that $X_0 = n$.

Solution: The results in Section 4.7 show that $EX_t = \mu^t EX_0$, which implies $EX_t = n(4/5)^t$ in this case. Therefore, the expected total number of individuals is

$$\sum_{t=0}^{\infty} EX_t = n \sum_{t=0}^{\infty} (4/5)^t = \frac{n}{1-\frac{4}{5}} = 5n.$$