ECS132 **Atharva Chalke, Sam Yuan, Travis Li**
Prof. Matloff Due Date: 10/29/19

Homework 2

PROBLEM A:

Think of the points 1, 2, ..., n on the number line. Each is either ”occupied” or not, with probability p, independently, except that points 1 and n are occupied for sure.

Let M denote the minimum distance between successive occupied points. E.g., if n is 6 and points 1, 5 and 6 are occupied, then M = 1.

1. Find Var(M) for the case n = 5, in terms of p.
2. Write a function with call form simA(nreps,n,p) to find Var(M) for the general case by simulation.

The answer for n = 5 and p = 0.4 is 1.46.

Solution

Part 1

Definitions
Let A_i be 1 if the point i is occupied else 0.
Let Support of M be D, Thus D = $\{1, 2, 4\}$

$$E(M) = \sum_{d \in D} d * P(M = d)$$

$$\begin{aligned} P(D = 4) &= P(A_2 = 0 \; and \; A_3 = 0 \; and \; A_4 = 0) \\ &= (1-p)^3 \end{aligned}$$

$$\begin{aligned} P(D = 2) &= P(A_2 = 0 \; and \; A_3 = 1 \; and \; A_4 = 0) \\ &= p * (1-p)^2 \end{aligned}$$

Every other combination of the 3 spaces in between results in a distance of 1. To find the probability we find the probability for each case and sum them up.

Combination	Probability
001	$p * (1-p)^2$
011	$p^2 * (1-p)$
100	$p * (1-p)^2$
101	$p^2 * (1-p)$
110	$p^2 * (1-p)$
111	p^3

$$\begin{aligned}P(D=1) &= (1-p)^2 * p + p^2 * (1-p) + p * (1-p)^2 + p^2 * (1-p) + p^2 * (1-p) + p^3 \\ &= 2 * p * (1-p)^2 + 3 * p^2 * (1-p) + p^3\end{aligned}$$

$$E(M) = 4 * (1-p)^3 + 2 * p * (1-p)^2 + 1 * (2 * p * (1-p)^2 + 3 * p^2 * (1-p) + p^3)$$

Lets find variance now.

$$\begin{aligned}Var(M) &= \sum_{d \in D} (d - E(M))^2 * P(M = d) \\ &= [(1 - E(M))^2 * (2 * p * (1-p)^2 + 3 * p^2 * (1-p) + p^3)] + [(2 - E(M))^2 * p * (1-p)^2] \\ &\quad + [(4 - E(M))^2 * (1-p)^3]\end{aligned}$$

Substituting the value for p, i.e. 0.4
Var(M) = 1.460736

Part 2
Code2.R

PROBLEM B:

In the setting of Problem B, Homework I, let Wk be the number of reported 0 bits among the first k bits. Using indicator functions as in Equations (3.98) and following (this is required), find EWk in terms of k, p and q.

For this question, I divide all the circumstances by the position of bit where the line first fails.

Let F be the position of the first fail bit. If the line doesn't fail in the first k bits, say F ¿ k.

$$\begin{aligned}
EW_k &= E(W_k \textit{ and } F = 1 + W_k \textit{ and } F = 2 + ... + W_k \textit{ and } F = k + W_k \textit{ and } F > k) \\
&= E(W_k \textit{ and } F = 1) + E(W_k \textit{ and } F = 2) + ... + E(W_k \textit{ and } F = k) \\
&= P(F = 1)E(W_k|F = 1) + P(F = 2)E(W_k|F = 2) + ... + P(F = k)E(W_k|F = k) + P(F > k)E(W_k|F \\
&= \sum_{i=1}^{k} P(F = i)E(W_k|F = i) + P(F > k)E(W_k|F > k)
\end{aligned} \tag{1}$$

The line first fails at the ith bit means that the line remains not fail in the first (i-1) bits. So

$$P(F = i) = q(1-q)^{i-1} \tag{2}$$

Also,

$$P(F > k) = (1-q)^k \tag{3}$$

Before (and excluding) the ith bit, there are (i-1) bits where we have to count in probability of not outputting 1 since the line has not failed. After (and including) the ith bit, there are (k-i+1) bits which are all 0s since the line has failed. Therefore,

$$\begin{aligned}
E(W_k|F = i) &= (1-p)(i-1) + (k-i+1) \\
&= i - 1 + k - i + 1 - p(i-1) \\
&= k - p(i-1)
\end{aligned} \tag{4}$$

Also,

$$E(W_k|F > k) = k(1-p) \tag{5}$$

Back to calculation of EW_k:

$$\begin{aligned}
EW_k &= \sum_{i=1}^{k} P(F=i)E(W_k|F=i) + P(F>k)E(W_k|F>k) \\
&= \sum_{i=1}^{k} q(1-q)^{i-1}(k-p(i-1)) + (1-q)^k k(1-p) \\
&= \sum_{i=1}^{k} q(1-q)^{i-1}k - \sum_{i=1}^{k} q(1-q)^{i-1}p(i-1) + k(1-p)(1-q)^k \\
&= qk\sum_{i=1}^{k}(1-q)^{i-1} - pq\sum_{i=1}^{k}(i-1)(1-q)^{i-1} + k(1-p)(1-q)^k \\
&= qk\frac{1-(1-q)^k}{1-(1-q)} - pq(\frac{(1-q)-(1-q)^k}{q^2} - \frac{(k-1)(1-q)^k}{q}) + k(1-q)^k - pk(1-q)^k \\
&= k(1-(1-q)^k) - \frac{p}{q}((1-q)-(1-q)^k) + p(k-1)(1-q)^k + k(1-q)^k - pk(1-q)^k \\
&= k - \cancel{k(1-q)^k} - \frac{p}{q}(1-q) + \frac{p}{q}(1-q)^k + \cancel{pk(1-q)^k} - p(1-q)^k + \cancel{k(1-q)^k} - \cancel{pk(1-q)^k} \\
&= k - \frac{p}{q}(1-q) + (\frac{p}{q}-p)(1-q)^k
\end{aligned} \tag{6}$$

PROBLEM C:

Consider some undirected, connected graph. (Starting at any vertex, we can reach any other via a finite sequence of hops.) What happens when we remove some edge? The graph may or may not still be connected. If not, it will have 2 connected components.

Write a function with call form

simConn(nreps,adjMat)
that uses simulation to find the distribution of N, the number of connected components remaining after a random edge is deleted. The return value will be a vector of length 2.

Solution

Code: Code2.R
Note:
1. The csv file should contain the matrix separated by **commas**.
2. In the vector returned, the first value is the probability of it staying connected while the second value is the probability that it doesn't stay connected.

PROBLEM D:

This problem involves Chebychev's Inequality, Equation (3.63). (You are not responsible for this material on quizzes, but it will be used here.)

Evaluate both sides of (3.63) for the following case: X has a binomial distribution with 10 trials and success probability 0.25, and c=2. Verify that the inequality does hold in this case.

Solution

Given Chebychev's Inequality, p = 0.25, c = 2

For variance:

$$\begin{aligned} V_{ar}(X) &= np(1-p) \\ &= 10 * 0.25 * (1-0.25) \\ &= 1.875 \end{aligned} \tag{7}$$

For mean:

$$\begin{aligned} \mu &= np \\ &= 10 * 0.25 = 2.5 \end{aligned} \tag{8}$$

For c = 2:

$$\frac{1}{c^2} = \frac{1}{4} \tag{9}$$

Therefore:

$$\begin{aligned} P(|X-\mu| \geq cV_{ar}) &= P(|X-2.5| \geq 2 * 1.875) \\ &= P(|X-2.5| \geq 3.75) \\ &= \sum_{i=7}^{10} P(X=i) \\ &= \sum_{i=7}^{10} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \binom{10}{7} 0.25^7 (1-0.25)^{10-7} + \binom{10}{8} 0.25^8 (1-0.25)^{10-8} \\ &\quad + \binom{10}{9} 0.25^9 (1-0.25)^{10-9} + \binom{10}{10} 0.25^{10} (1-0.25)^{10-10} \\ &= 0.0035 \leq \frac{1}{4} \end{aligned} \tag{10}$$

Hence, Chebychev's inequality does hold in this case.