STA 247 — Assignment #4. Due in class on December 5.

Late assignments will be accepted only with a valid medical or other excuse.
Worth 7% of the course mark.

This assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own. Handing in work that is not your own is a serious academic offense. Fabricating results, such handing in fake output that was not actually produced by your program, is also an academic offense.

Consider a Markov model for the behaviour of a potential customer visiting an online purchasing website. The customer is modelled as being in one of four states:

1) Browsing products, looking for something they might want to buy (eg, looking at product listings or product comparisons).
2) Examining a product in detail (eg, looking at specifications or reviews of the product).
3) Ordering a product (eg, filling in shipping or billing information).
4) Submitting an order (one click, taking them to a page confirming their order has been received).
5) Leaving the site.

The variable X_0 represents the state when they first visit the website; we assume that $X_0 = 1$ always. The variables $X_1, X_2, \ldots$ represent the state after the user's first click on a link, their second click on a link, etc. The state after the ith click is assumed to depend (directly) only on the state after the previous click, with transition probabilities given as follows:

$$
\begin{aligned}
P(X_i = 1 \,|\, X_{i-1} = 1) &= 0.85 \\
P(X_i = 2 \,|\, X_{i-1} = 1) &= 0.05 \\
P(X_i = 3 \,|\, X_{i-1} = 1) &= 0.05 \\
P(X_i = 4 \,|\, X_{i-1} = 1) &= 0 \\
P(X_i = 5 \,|\, X_{i-1} = 1) &= 0.05
\end{aligned}
$$

$$
\begin{aligned}
P(X_i = 1 \,|\, X_{i-1} = 2) &= 0.2 \\
P(X_i = 2 \,|\, X_{i-1} = 2) &= 0.6 \\
P(X_i = 3 \,|\, X_{i-1} = 2) &= 0.1 \\
P(X_i = 4 \,|\, X_{i-1} = 2) &= 0 \\
P(X_i = 5 \,|\, X_{i-1} = 2) &= 0.1
\end{aligned}
$$

$$
\begin{aligned}
P(X_i = 1 \,|\, X_{i-1} = 3) &= 0.15 \\
P(X_i = 2 \,|\, X_{i-1} = 3) &= 0.05 \\
P(X_i = 3 \,|\, X_{i-1} = 3) &= 0.5 \\
P(X_i = 4 \,|\, X_{i-1} = 3) &= 0.25 \\
P(X_i = 5 \,|\, X_{i-1} = 3) &= 0.05
\end{aligned}
$$

$$
\begin{aligned}
P(X_i = 1 \,|\, X_{i-1} = 4) &= 0.3 \\
P(X_i = 2 \,|\, X_{i-1} = 4) &= 0.05 \\
P(X_i = 3 \,|\, X_{i-1} = 4) &= 0.05 \\
P(X_i = 4 \,|\, X_{i-1} = 4) &= 0 \\
P(X_i = 5 \,|\, X_{i-1} = 4) &= 0.6
\end{aligned}
$$

$$P(X_i = 5 \,|\, X_{i-1} = 5) = 1$$

The last transition probability above means that once the customer leaves the site, we model them as never coming back. (Of course, the same customer might come back another time, but we assume that the website software cannot tell that a new visitor is the same as a previous visitor, and in any case it no longer makes sense to measure time in "clicks" once they aren't clicking.)

Your task: Write a short R program to find the marginal distribution of X_i, given a positive integer i, and use it to find the marginal distributions of X_1, X_2, X_{10}, and X_{20}. Hand in your R program and the distributions you found.

Bonus [10 extra marks]: Find the expected number of items that a customer purchases — ie, the expected number of times that they are in state 4 before leaving the site. An exact answer (apart from floating point round-off error) is required; an estimate from simulation is not sufficient. Hint: Consider the formula $E(Y) = E(E(Y|X))$.

R hints: The matrix multiplication operator in R is `%*%` — the `*` operator is element-by-element multiplication, *not* matrix multiplication, and the `^` operator is element-by-element power, *not* matrix power. You can find the inverse of a matrix `M` with `solve(M)`.