Problem 2. A mouse moves along a tiled corridor with $2m$ tiles, where $m > 1$. From each tile $i \neq 1, 2m$, it moves to either tile $i - 1$ or $i + 1$ with equal probability. From tile 1 or tile $2m$, it moves to tile 2 or $2m - 1$, respectively, with probability 1. Each time the mouse moves to a tile $i \leq m$ or $i > m$, an electronic device outputs a signal L or R, respectively. Can the generated sequence of signals L and R be described as a Markov chain with states L and R?

Problem 3. Consider the Markov chain in Example 7.2, for the case where $m = 4$, as in Fig. 7.2. and assume that the process starts at any of the four states, with equal probability. Let $Y_n = 1$ whenever the Markov chain is at state 1 or 2, and $Y_n = 2$ whenever the Markov chain is at state 3 or 4. Is the process Y_n a Markov chain?

Please turn in the solution to Problem 3 (you do not need to do problem 2).
Do not simply answer yes or no; rather, prove the Markov property holds if the process is a Markov chain, or give an example to show it doesn't if it is not a Markov chain.

- The sequence $Y_1, Y_2, Y_3, \ldots$ is not a Markov chain.
- Informally speaking, we need to argue that given Y_k, knowledge of past history (i.e., information from $Y_1, \ldots, Y_{k-1}$) can still provide additional information that is helpful in predicting Y_{k+1}.
- Intuition: let $X_1, X_2, X_3, \ldots$ be the position of the fly.

  Now X_k is a Markov chain.

  But given Y_k you don't know X_k. Instead, Y_k gives you *some* information about X_k.
- Intuition: knowledge of past Y_k can give you *more* information about X_k.
- OK, but all this is vague. Now need to make a precise mathematical argument.

- Will show this by arguing that

  $P(Y_3 = 1 \mid Y_1 = 1, Y_2 = 2) \neq P(Y_3 = 1 \mid Y_1 = 2, Y_2 = 2)$.
- This shows that $Y_1, Y_2, \ldots$ is not a Markov chain. Indeed, if it was a Markov chain, the right-hand side and the left-hand side would be the same.
- How did we choose these values? Because here, on the left-hand side, knowledge of Y_1, Y_2 tells you exactly what X_1, X_2 are.
- ...but not so on the right hand side!
- Indeed, the event $Y_1 = 1, Y_2 = 2$ can only happen if $X_1 = 2$ and $X_2 = 3$.

  In this case, having $Y_3 = 1$ is equivalent to having $X_3 = 2$, which has probability 0.3

- On the other hand, $\{Y_1 = 2, Y_2 = 2\}$, which we denote by R to avoid writing it out repeatedly can actually happen in three different ways:

  $A_1 = \{X_1 = 4, X_2 = 4\}$, which has probability $1/4$

  $A_2 = \{X_1 = 3, X_2 = 4\}$, which has probability $(1/4) \cdot 0.3$

  $A_3 = \{X_1 = 3, X_2 = 3\}$, which has probability $(1/4) \cdot 0.4$

- Applying the definition of conditional probability:

  $$P(A_3 \mid R) = \frac{(1/4) \cdot 0.4}{1/4 + (1/4) \cdot 0.3 + (1/4) \cdot 0.4} = \frac{4}{17}$$

- Similarly, one can calculate $P(A_1 \mid R), P(A_2 \mid R)$ ...though, as it turns out, we won't need them.

- Using the definition of conditional probability:
$$P(Y_3 = 1 \mid R) = P(Y_3 = 1 \mid R, A_1)P(A_1 \mid R)$$
$$+P(Y_3 = 1 \mid R, A_2)P(A_2 \mid R)$$
$$+P(Y_3 = 1 \mid R, A_3)P(A_3 \mid R)$$
$$= 0 \cdot P(A_1 \mid R) + 0 \cdot P(A_2 \mid R) + 0.3 \cdot \frac{4}{17}$$
$$= \frac{0.3 \cdot 4}{17}$$

- Since $0.3 \neq 0.3 \cdot \frac{4}{17}$, we are done: we showed that Y_k is not a Markov chain.

- Let P be the probability transition matrix of a Markov chain on n states. Show that:

  (i) 1 is an eigenvalue of P.

  (ii) suppose $\sum_{i=1}^{n} p_{ij} = 1$ for every state j; in other words, every column of P adds up to one

  (for future information, a nonnegative matrix whose row and columns add up to one is called *doubly stochastic*).

  Show that if this Markov chain is initialized at the distribution where it is equally likely to be at any state, it stays at that distribution for all time.

  In other words, if the distribution at time zero is $\mathbf{p} = (1/n, \ldots, 1/n)$, then the distribution of the chain at any time k is also $(1/n, \ldots, 1/n)$.

  (iii) Show that any power of a doubly stochastic matrix is doubly stochastic.

- Part (i): Let $\mathbf{e}$ the the all-ones vector. Then the vector $P\mathbf{e}$ is just the vector of row-sums of P; so $P\mathbf{e} = \mathbf{e}$. This shows that 1 is an eigenvalue of P (with corresponding eigenvalue $\mathbf{e}$).
- Part (ii): if the chain is initialized at the distribution $(1/n)\mathbf{e}$ as assumed, then its state at the next time step is $(1/n)\mathbf{e}^T P$ (see lecture notes which explain how Markov chains propagate distributions forward by matrix multiplication).
- But $\mathbf{e}^T P$ is a vector of column sums of P, so by assumption $\mathbf{e}^T P = \mathbf{e}^T$. We thus have that $(1/n)\mathbf{e}^T P = (1/n)\mathbf{e}^T$, and at the next time step the chain once again has distribution $(1/n)\mathbf{e}$.
- Repeating this argument gives that the distribution at any future time is $(1/n)\mathbf{e}$, as the problem asks for.

- Part (iii): The statement that P has unit row sums can be written as $P\mathbf{e} = \mathbf{e}$; and the statement that P has unit column sums can be written as $\mathbf{e}^T P = \mathbf{e}^T$.
- These conditions imply

  $P^2\mathbf{e} = PP\mathbf{e} = P\mathbf{e}$

  Next,

  $P^3\mathbf{e} = P^2P\mathbf{e} = P^2\mathbf{e} = \mathbf{e}$

  and a similar argument shows that $P^k\mathbf{e} = \mathbf{e}$ for any k.
- Similarly,

  $\mathbf{e}^T P^2 = \mathbf{e}^T PP = \mathbf{e}^T P = \mathbf{e}^T$

  and

  $\mathbf{e}^T P^3 = \mathbf{e}^T PP^2 = \mathbf{e}^T P^2 = \mathbf{e}^T$

  and a similar argument shows that $\mathbf{e}^T P^k = \mathbf{e}^T$ for any k.
- The previous two bullets imply that P^k is doubly stochastic.

- Suppose X_t is a Markov chain on the state-space $\{1,\ldots,n\}$ which behaves as follows: if $X_t = i$ then X_{t+1} is a uniformly random j such that $|j - i| = 1$. For example, if $X_1 = 3$, then X_2 is either 2 or 4, both with probability 1/2. Another example: if $X_3 = n$, then $X_4 = n - 1$ with probability one.
- This Markov chain is initialized at state n. We stop the process when it reaches state 1.
- Prove: $\lim_{t\to\infty} P(\text{ process is stopped by time } t) = 1$.

- We first argue the following.

  Suppose $A_1, A_2, A_3, \ldots$ is a sequence of events such that, for all i,

  $P(A_i \,|\, A_1^c, \cdots, A_n^c) \geq p > 0.$

  Then $\lim_t P($ one of the events $A_1, \ldots, A_t$ occurs $) = 1.$

- Proof: consider the opposite event

  $$P(\text{ none of } A_1, \ldots, A_t \text{ occur }) = P(A_1^c)P(A_2^c \,|\, A_1^c)P(A_3^c \,|\, A_1^c, A_2^c) \cdots P(A_t \,|\, A_1^c, \ldots, A_{t-1}^c)$$
  $$\leq (1-p)^t$$

  by assumption; and because $0 < p \leq 1$, the last expression goes to zero as $t \to \infty$.

- Define A_i to be the event that the random walk reaches state 1 during time period $(i-1)n, (i-1)n+1, \ldots, in-1$.
- No matter where the random walk is at time $(i-1)n$, it can reach 1 by moving left until it reaches node 1. The probability of this is at least $(1/2)^{n-1}$.
- Rephrasing: even if the random walk did not visit node 1 before time $(i-1)n$, the probability it visits 1 in times $(i-1)n, (i-1)n+1, \ldots, in-1$ is at least $(1/2)^{n-1}$
- In other words, $P(A_i \mid A_1^c, \ldots, A_{i-1}^c) \geq (1/2)^{n-1}$.
- We can apply the claim on the previous slide with $p = (1/2)^{n-1}$ to get that
- $\lim_t P($ one of the events $A_1, \ldots, A_t$ occurs $) = 1$
- Since the process is stopped when any of the events A_i occur, we obtain that $\lim_{t \to \infty} P($ process is stopped by time $t) = 1$