Solutions

Please simplify your answers to the extent reasonable without a calculator. Show your work. Explain your answers, concisely.

1. Especially in San Diego, a good guess about tomorrow’s weather is that it will be like today’s. Let’s formalize this as there being two weather states, \( W_t \in \{ \text{sunny}, \text{rainy} \} \), where the \( t \in \mathbb{N} \) index labels days, and suppose they form a Markov chain with transition probability matrix:

\[
P = \begin{pmatrix}
0.9 & 0.1 \\
0.4 & 0.6
\end{pmatrix},
\]

where the rows and columns are labeled by the states ‘sunny’ and ‘rainy’, in that order.

a. [5 points] What is \( \Pr(W_{t+1} = \text{rainy} \mid W_t = \text{rainy}) \)?

0.6, from \( P \).

b. [10 points] What is \( \Pr(W_{t+1} = \text{rainy} \mid W_t = \text{rainy} \text{ and } W_{t+2} = \text{rainy}) \)?

Write \( R \) for ‘rainy’ and \( S \) for ‘sunny’. Then

\[
\Pr(W_{t+1} = R \mid W_t = R, W_{t+2} = R) = \frac{\Pr(W_{t+1} = R, W_{t+2} = R \mid W_t = R)}{\Pr(W_{t+2} = R \mid W_t = R)} = \frac{\Pr(W_{t+1} = R, W_{t+2} = R \mid W_t = R)}{0.6 \cdot 0.6 + 0.4 \cdot 0.1} = 0.9.
\]

so it is more likely to be sunny on a day between a rainy day and a sunny day.

c. [10 points] What is \( \Pr(W_{t+1} = \text{rainy} \mid W_t = \text{rainy} \text{ and } W_{t+2} = \text{sunny}) \)?

By the same argument as in 1.b,

\[
\Pr(W_{t+1} = R \mid W_t = R, W_{t+2} = S) = \frac{0.6 \cdot 0.4}{0.6 \cdot 0.4 + 0.4 \cdot 0.9} = 0.4,
\]

so it is more likely to be sunny on a day between a rainy day and a sunny day.

2. Suppose \( X \sim \text{Binomial}(n, P) \), where \( P \sim \text{Uniform}[0,1] \), \( i.e., \) the probability density function for \( P \) is \( f_P(p) = 1 \) if \( 0 \leq p \leq 1 \) and 0 otherwise.

a. [10 points] What is \( \mathbb{E}[X] \)?

\[
\mathbb{E}[X] = \int_0^1 \mathbb{E}[X \mid P = p] \cdot 1 \, dp = \int_0^1 np \, dp = n/2.
\]

b. [15 points] What is \( \text{Var}[X] \)?
\[ \text{Var}[X] = \mathbb{E}[(X - n/2)^2] = \int_0^1 \mathbb{E}[(X - n/2)^2 | P = p] \cdot 1 \, dp \]
\[ = \int_0^1 \mathbb{E}[(X - np + np - n/2)^2 | P = p] \, dp \]
\[ = \int_0^1 \mathbb{E}[(X - np)^2 | P = p] \, dp + n^2 \int_0^1 \mathbb{E}[(p - 1/2)^2] \, dp \]
\[ + 2n \int_0^1 (p - 1/2) \mathbb{E}[X - np | P = p] \, dp \]
\[ = \int_0^1 \text{Var}[X | P = p] \, dp + n^2 \text{Var}[P] + 2n \int_0^1 (p - 1/2) \cdot 0 \, dp \]
\[ = \int_0^1 np(1 - p) \, dp + n^2/12 = n/6 + n^2/12. \]

3. Let \( M \) be a continuous random variable with probability density function \( f_M(m) = 4me^{-2m} \) if \( 0 \leq m \) and 0 otherwise. Let \( S_t \) be the value of a stock at time \( t \) and suppose \( S_{t+1} = M_t S_t \).
   a. [15 points] Is \( \{S_t \mid t \in \mathbb{N}\} \) a Martingale? Why or why not?
   Yes, because \( \mathbb{E}[S_{t+1} | S_t = s_t] = \mathbb{E}[M_t S_t | S_t = s_t] = \mathbb{E}[M_t]s_t \), and
   \[ \mathbb{E}[M_t] = \int_0^\infty 4m^2 e^{-2m} \, dm = \int_0^\infty 4me^{-2m} \, dm = 1, \]
   where the first equality is the definition, the second equality comes from integration by parts, and the third equality follows since \( f_M(m) \) is a probability density function.
   b. [10 points] Suppose you invest $1000 in this stock and then your broker says “This stock is bound to go up. Let’s make a bet: you pay me $75 and when your stock is worth $2000, I’ll pay you an extra $100.”. Should you take this bet? Why or why not?
   No. The maximum Markov inequality tells us that the probability of a positive Martingale doubling in value is less than \( 1/2 \), so you should only even consider taking this bet if you have to pay no more than $100/2 = $50.

4. Qingwa the frog starts at 0 on the number line. Each minute she hops one unit in the positive direction with probability \( 1/2 \) and stays still otherwise. Let \( X_n \) denote her position after \( n \) minutes.
   a. [5 points] What is the transition probability matrix \( P \) for this Markov chain?
   \[
   P = \begin{pmatrix}
   0 & 1 & 2 & 3 & \ldots \\
   1/2 & 1/2 & 1/2 & 1/2 & \ldots \\
   1/2 & 1/2 & 1/2 & \ldots & \ldots \\
   \vdots & \vdots & \vdots & \ddots & \ddots 
   \end{pmatrix}
   \]
b. [10 points] What is the 2-step transition probability matrix \( P^{(2)} \)?

\[ P^{(2)} = P^2 = \frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix}. \]


c. [10 points] What is the probability that Qingwa’s position \( X_n = k \) after \( n \) minutes? These are the entries in the top row of \( P^{(n)} = P^n \), which are

\[ (P^n)_{0k} = \frac{1}{2^n} \binom{n}{k}, \]

since each sequence of \( n \) hops/no hops has the same probability \( 1/2^n \) and there are \( \binom{n}{k} \) of those with \( k \) hops.

5. [Extra credit: 15 points] Using the same information as in Problem 1, suppose there is a sequence of 16 days for which you know the weather on 5: \( S**S**R**R****R \), where \( S \) denotes ‘sunny’, \( R \) denotes ‘rainy’, and * indicates we don’t know what the weather is. Which is the most probable sequence of \( S \)s and \( R \)s? Why?

Because of the Markov property we can consider each sequence of *s surrounded by known weather days separately. The probability of any sequence is given by the product of the appropriate entries in \( P \) from 1. Since 0.9 is the biggest entry, between \( S \)s the highest probability sequence is just \( S \)s. Since the transition from \( R \) to \( R \) has probability 0.6 < 0.9, the highest probability sequence between an \( S \) and an \( R \) is also all \( S \)s. Finally, between two \( R \)s the highest probability sequence is all \( R \)s since 0.4 < 0.6, unless the sequence is long enough that it is more probable to switch to a sequence of \( S \)s. That is, \( RR^nR \) has probability 0.6\(^{n+1} \), while \( RS^nR \) has probability 0.4 \( \cdot 0.9^{n-1} \cdot 0.1 \). The latter is larger for \( n \geq 7 \), but the numbers of *s in the last two surrounded sequences are 2 and 4, respectively. So the most probable sequence is \( SSSSSSRRRRRRRRRR \).

In the somewhat more complicated setting of hidden Markov models, the Viterbi algorithm answers the analogous question. Some of you might find this worth learning about; Viterbi is one of the co-founders of Qualcomm.