1. Practice with Markov chains

Consider the two-state Markov chain $X_1, X_2, \ldots$, with $X_i \in \{0, 1\}$ and symmetric transition probabilities $P(1|0) = P(0|1) = \alpha$ and $P(0|0) = P(1|1) = 1 - \alpha$ ($\alpha < 0.5$). Suppose the initial distribution (distribution of $X_1$) is $\pi(0) = \pi(1) = 0.5$.

a) Draw the state transition diagram.

Answer:

```
1 - \alpha \quad 0 \quad \alpha \quad 1 \quad 1 - \alpha
```

b) Compute the distribution of $X_2$.

Answer:

$$P(X_2 = 0) = P(X_2 = 0|X_1 = 0)P(X_1 = 0) + P(X_2 = 0|X_1 = 1)P(X_1 = 1)$$

$$= (1 - \alpha)P(X_1 = 0) + \alpha P(X_1 = 1)$$

$$= (1 - \alpha) \times 0.5 + \alpha \times 0.5$$

$$= 0.5,$$

and $P(X_2 = 1) = 0.5$ as well.

c) What is the distribution of $X_n$ for general $n$?

Answer: Since the transition probabilities do not depend on time, we can repeat the argument with $X_2$ replacing $X_1$ and $X_3$ replacing $X_2$ and obtain $P(X_3 = 0) = P(X_3 = 1) = 0.5$. Repeating this argument, we see that $P(X_n = 1) = P(X_n = 0) = 0.5$ for all $n$.

d) Compute the joint distribution of $X_1$ and $X_3$. Are these two random variables independent?

Answer: By total probability rule,

$$P(X_1 = 1, X_3 = 0) = P(X_1 = 1, X_3 = 0, X_2 = 0) + P(X_1 = 1, X_3 = 0, X_2 = 1)$$

$$= \frac{1}{2} (P(X_3 = 0, X_2 = 0|X_1 = 1) + P(X_3 = 0, X_2 = 1|X_1 = 1))$$

$$= \frac{1}{2} ((1 - \alpha)\alpha + (1 - \alpha)\alpha)$$

$$= \alpha(1 - \alpha).$$

Similarly, $P(X_1 = 0, X_3 = 1) = \alpha(1 - \alpha)$.

$$P(X_1 = 0, X_3 = 0) = \frac{1}{2} (P(X_3 = 0, X_2 = 0|X_1 = 0) + P(X_3 = 0, X_2 = 1|X_1 = 0))$$

$$= \frac{1}{2} ((1 - \alpha)^2 + \alpha^2)$$
and \( P(X_1 = 0, X_3 = 0) = P(X_1 = 1, X_3 = 1) \).

Using part (c) we know that

\[
P(X_1 = 1)P(X_3 = 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \neq \alpha(1 - \alpha)
\]

since \( \alpha < 0.5 \). Therefore the two random variables are not independent.

e) Now suppose we have another sequence of random variables \( Y_1, Y_2, \ldots \) with \( Y_i \in \{0, 1\} \) and the \( Y_i \)'s are mutually independent conditional on the \( X_i \)'s. Further, suppose each \( X_i \) is connected to \( Y_i \) through a binary symmetric channel with crossover probability \( p < 0.5 \). Compute \( P(Y_3 = 0|Y_2 = 0) \) and \( P(Y_3 = 0|Y_1 = 0, Y_2 = 0) \). Do the \( Y_i \)'s form a Markov chain? Give an intuitive explanation of your answer.

**Answer:** First, using the total probability rule, we compute \( P(Y_2 = 0) \) and \( P(Y_2 = 0, Y_3 = 0) \) to be

\[
P(Y_2 = 0) = P(Y_2 = 0, X_2 = 0) + P(Y_2 = 0, X_2 = 1)
= \frac{1}{2}(P(Y_2 = 0|X_2 = 0) + P(Y_2 = 0|X_2 = 1))
= \frac{1}{2}((1 - p) + p)
= \frac{1}{2}
\]

\[
P(Y_2 = 0, Y_3 = 0) = P(Y_2 = 0, Y_3 = 0, X_2 = 0, X_3 = 0) + P(Y_2 = 0, Y_3 = 0, X_2 = 0, X_3 = 1)
+ P(Y_2 = 0, Y_3 = 0, X_2 = 1, X_3 = 0) + P(Y_2 = 0, Y_3 = 0, X_2 = 1, X_3 = 1)
= \frac{1}{2}(P(X_3 = 0|X_2 = 0)P(Y_2 = 0|X_2 = 0)P(Y_3 = 0|X_3 = 0)
+ P(X_3 = 1|X_2 = 0)P(Y_2 = 0|X_2 = 0)P(Y_3 = 0|X_3 = 1)
+ P(X_3 = 0|X_2 = 1)P(Y_2 = 0|X_2 = 1)P(Y_3 = 0|X_3 = 0)
+ P(X_3 = 1|X_2 = 1)P(Y_2 = 0|X_2 = 1)P(Y_3 = 0|X_3 = 1))
= \frac{1}{2}((1 - \alpha)(1 - p)^2 + 2\alpha(1 - p)p + (1 - \alpha)p^2)
\]

Therefore,

\[
P(Y_3 = 0|Y_2 = 0) = (1 - \alpha)(1 - p)^2 + 2\alpha(1 - p)p + (1 - \alpha)p^2.
\]

Now \( P(Y_1 = 0, Y_2 = 0) = P(Y_2 = 0, Y_3 = 0) \). Let \( A \) be the event that \( \{Y_1 = 0, Y_2 = 0, Y_3 = 0\} \).

We write \( P(A) \) as

\[
P(A) = \sum_{i,j,k \in \{0,1\}} P(A, X_1 = i, X_2 = j, X_3 = k)
= \frac{1}{2} \sum_{i,j,k \in \{0,1\}} P(Y_1 = 0|X_1 = i)P(Y_2 = 0|X_2 = j)P(Y_3 = 0|X_3 = k)
\cdot P(X_1 = i)P(X_2 = j|X_1 = i)P(X_3 = k|X_2 = j)
= \frac{1}{2}(1 - \alpha)^2p^3 + (1 - \alpha)^2(1 - p)^3 + 2\alpha(1 - \alpha)p^2(1 - p) + 2\alpha(1 - \alpha)(1 - p)^2p
+ \alpha^2(1 - p)^2p + \alpha^2p^2(1 - p).
\]

Finally we get that

\[
P(Y_3 = 0|Y_2 = 0, Y_1 = 0) = \frac{P(A)}{\frac{1}{2}((1 - \alpha)(1 - p)^2 + 2\alpha(1 - p)p + (1 - \alpha)p^2)}.
\]
We see that the two probabilities $P(Y_3 = 0|Y_1 = 0, Y_2 = 0)$ and $P(Y_3 = 0|Y_2 = 0)$ are not equal. Therefore the $Y_i$'s do not form a Markov chain. Intuitively, $Y_i$ does not contain all the necessary past information because it is not exactly equal to $X_i$. Hence, having further information from $Y_{i-2}, Y_{i-3}$ etc. provides more information about $X_i$.

2. Continuous observations

In class when we discussed the communication and speech recognition problems, we assumed the observed outputs $Y_i$'s are discrete random variables. For physical channels, this is an appropriate model if the $Y_i$'s are discretizations of the underlying continuous signals. Often though, one may want to model directly the continuous received signals, in which case it is natural to treat the $Y_i$'s as continuous random variables. Everything we did goes through, except that we need to replace the conditional probability of the observation given the input by the conditional pdf of the observation given the input. (You can assume this fact for the rest of the question, but you may want to think a bit why this is valid, by thinking of the continuous observation as a limit of discretization as the discretization interval $\delta$ goes to zero.)

a) In the repetition coding communication example, suppose $Y_i \sim N(A, \sigma^2)$ when $X = 0$ is transmitted, and $Y_i \sim N(-A, \sigma^2)$ when $X = 1$ is transmitted. (Physically, we are sending a signal at voltage level $A$ to represent a 0 and at voltage level $-A$ to represent a 1, and there is an additive noise corrupting the transmitted signal to yield the received signal.) Re-derive the MAP receiver (decision rule) for this channel model. Put it in the simplest form.

**Answer:** The MAP receiver rule derived in Lecture 15 was:
\[
LLR(b_1, \ldots, b_n) := \sum_{i=1}^{n} LLR(b_i) = \log \left( \frac{f(Y = b_i|X = 0)}{f(Y = b_i|X = 1)} \right) \cdot \log \left( \frac{1 - \alpha}{\alpha} \right).
\]
The difference when $Y_i$ is continuous is that
\[
LLR(b) = \log(L(b)) = \log \left( \frac{f(Y = b|X = 0)}{f(Y = b|X = 1)} \right) = \frac{1}{2\sigma^2} (b + A)^2 - (b - A)^2.
\]
Therefore the overall rule is
\[
\frac{1}{2\sigma^2} \sum_{i=1}^{n} (b_i + A)^2 - (b_i - A)^2 \cdot \log \left( \frac{1 - \alpha}{\alpha} \right).
\]

b) In the speech recognition example, suppose $Y_i \sim N(\mu_a, \sigma^2)$, when $X_i = a$. ($\mu_a$'s and $\sigma^2$ are then parameters of the model which are assumed to be known.) Compute the edge costs $d(a)$ and $d_i(a, a')$ of the Viterbi algorithm for this channel model. Be as explicit as you can.

**Answer:** From the definitions in Lecture 17,
\[
d(a) = -\log(\pi(a)Q(b|a)) = -\log(\pi(a)) - \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(b-a)^2}{2\sigma^2}} \right) = -\log(\pi(a)) - \frac{(b - \mu_a)^2}{2\sigma^2} + \log \left( \sqrt{2\pi\sigma^2} \right).
\]
and
\[
d_i(a_i, a_{i-1}) = -\log(p(a_i|a_{i-1})) = -\log(p(a_i)) - \frac{(b - \mu_{a_{i-1}})^2}{2\sigma^2} + \log \left( \sqrt{2\pi\sigma^2} \right).
\]
This edge cost is the sum of three terms. The first term depends on the likelihood of transitioning from state $a_{i-1}$ to state $a_i$. The second term depends on how close the state $a_i$ for $X_i$ fits the observation $Y_i = b_i$. The third term is a constant and can be ignored.