Quiz Section 9

Review

1) Maximum Likelihood Estimator (MLE): We denote the MLE of \( \theta \) as \( \hat{\theta}_{\text{MLE}} \) or simply \( \hat{\theta} \), the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

\[
\hat{\theta}_{\text{MLE}} = \arg \max _{\theta} L(x_1, \ldots, x_n \mid \theta) = \arg \max _{\theta} \ln L(x_1, \ldots, x_n \mid \theta)
\]

2) An estimator \( \hat{\theta} \) for a parameter \( \theta \) of a probability distribution is unbiased iff

\[
E[\hat{\theta}(X_1, \ldots, X_n)] = \theta
\]

3) A discrete-time stochastic process (DTSP) is a sequence of random variables \( X^{(0)}, X^{(1)}, X^{(2)}, \ldots \), where \( X^{(t)} \) is the value at time \( t \). For example, the temperature in Seattle or stock price of TESLA each day, or which node you are at after each time step on a random walk on a graph.

4) Markov Chain is a DTSP, with the additional following three properties:

(a) ...has a finite (or countably infinite) state space \( S = \{s_1, \ldots, s_n\} \) which it bounces between, so each \( X^{(t)} \in S \).

(b) ...satisfies the Markov property. A DTSP satisfies the Markov property if the future is (conditionally) independent of the past given the present. Mathematically, it means,

\[
P(X^{(t+1)} = x_{t+1} \mid X^{(0)} = x_0, X^{(1)} = x_1, \ldots, X^{(t-1)} = x_{t-1}, X^{(t)} = x_t) = P(X^{(t+1)} = x_{t+1} \mid X^{(t)} = x_t).
\]

(c) ...has fixed transition probabilities. Meaning, if we are at some state \( s_i \), we transition to another state \( s_j \) with probability independent of the current time. Due to this property and the previous, the transitions are governed by \( n^2 \) probabilities: the probability of transitioning from one of \( n \) current states to one of \( n \) next states. These are stored in a square \( n \times n \) transition probability matrix (TPM) \( M \), where \( M_{ij} = P(X^{(t+1)} = s_j \mid X^{(t)} = s_i) \) is the probability of transitioning from state \( s_i \) to state \( s_j \) for any/every value of \( t \).

5) A stationary distribution of a Markov chain is a probability distribution on states that is unchanged by taking one step of the Markov chain.

Task 1 – Mystery Dish!

A fancy new restaurant has opened up that features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: dish A with probability \( 0.5 \), dish B with probability \( \theta \), dish C with probability \( 2\theta \), and dish D with probability \( 0.5 - 3\theta \). Each diner is served a dish independently. Let \( x_A \) be the number of people who received dish A, \( x_B \) the number of people who received dish B, etc, where \( x_A + x_B + x_C + x_D = n \). Find the MLE for \( \theta, \hat{\theta} \).

Task 2 – A Red Poisson

Suppose that \( x_1, \ldots, x_n \) are i.i.d. samples from a Poisson(\( \theta \)) random variable, where \( \theta \) is unknown. In other words, they follow the distributions

\[
P(k; \theta) = \theta^k e^{-\theta}/k!, \quad k \in \mathbb{N} \quad \text{and} \quad \theta > 0 \quad \text{is a positive real number.}
\]

Find the MLE of \( \theta \).
Task 3 – A biased estimator

In class, we showed that the maximum likelihood estimate of the variance $\theta_2$ of a normal distribution (when both the true mean $\mu$ and true variance $\sigma^2$ are unknown) is what’s called the population variance. That is

$$\hat{\theta}_2 = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2\right)$$

where $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

Task 4 – Weather Forecast

A weather forecaster predicts sun with probability $\theta_1$, clouds with probability $\theta_2 - \theta_1$, rain with probability $\frac{1}{2} - \theta_2$ and snow with probability $\frac{1}{2}$. This year, there have been 55 sunny days, 100 cloudy days, 160 rainy days and 50 snowy days. What is the maximum likelihood estimator for $\theta_1$ and $\theta_2$?

Task 5 – Faulty Machines

You are trying to use a machine that only works on some days. If on a given day, the machine is working it will break down the next day with probability $0 < b < 1$, and works on the next day with probability $1 - b$. If it is not working on a given day, it will work on the next day with probability $0 < r < 1$ and not work the next day with probability $1 - r$.

a) In this problem we will formulate this process as a Markov chain. First, let $X^{(t)}$ be a variable that denotes the state of the machine at time $t$. Then, define a state space $S$ that includes all the possible states that the machine can be in. Lastly, for all $A, B \in S$ find $P(X^{(t+1)} = A | X^{(t)} = B)$ ($A$ and $B$ can be the same state).

b) Suppose that on day 1, the machine is working. What is the probability that it is working on day 3?

c) As $n \to \infty$, what does the probability that the machine is working on day $n$ converge to? To get the answer, solve for the stationary distribution.

Task 6 – Another Markov Chain

Suppose that the following is the transition probability matrix for a 4 state Markov chain (states 1,2,3,4).

$$M = \begin{bmatrix}
0 & 1/2 & 1/2 & 0 \\
1/3 & 0 & 0 & 2/3 \\
1/3 & 1/3 & 0 & 1/3 \\
1/5 & 2/5 & 2/5 & 0 \\
\end{bmatrix}$$

a) What is the probability that $X^{(2)} = 4$ given that $X^{(0)} = 4$?

b) Write down the system of equations that the stationary distribution must satisfy and solve them.

Task 7 – Three Tails

You flip a fair coin until you see three tails in a row. Model this as a Markov chain with the following states:

- $S$: start state, which we are only in before flipping any coins.
- $H$: We see a heads, which means no streak of tails currently exists.
- $T$: We’ve seen exactly one tail in a row so far.
- \( TT \): We’ve seen exactly two tails in a row so far.
- \( TTT \): We’ve accomplished our goal of seeing three tails in a row, stop flipping, and stay there.

a) Write down the transition probability matrix.

b) Write down the system of equations whose variables are \( D(s) \) for each state \( s \in \{ S, H, T, TT, TTT \} \), where \( D(s) \) is the expected number of steps until state \( TTT \) is reached starting from state \( s \). Solve this system of equations to find \( D(S) \).

c) Write down the system of equations whose variables are \( \gamma(s) \) for each state \( s \in \{ S, H, T, TT, TTT \} \), where \( \gamma(s) \) is the expected number of heads seen before state \( TTT \) is reached. Solve this system to find \( \gamma(S) \), the expected number of heads seen overall until getting three tails in a row.