CSE 312: Foundations of Computing II

Section 7: Markov Chains, CLT

1. 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 random variable that denotes the state of the machine at time t. Then, define a state space $\mathcal{S}$ that includes all the possible states that the machine can be in. Lastly, for all $A, B \in \mathcal{S}$ find $\mathbb{P}(X_{t+1} = A \mid X_t = B)$ (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*.

2. Another Markov chain

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

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

(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.

3. 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 and stop flipping.

(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.

4. CLT example

Let X be the sum of 100 real numbers, and let Y be the same sum, but with each number rounded to the nearest integer before summing. If the roundoff errors are independent and uniformly distributed between -0.5 and 0.5, what is the approximate probability that $|X - Y| > 3$?

5. Tweets

A prolific twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

6. Poisson CLT practice

Suppose $X_1, \ldots, X_n$ are iid Poisson(λ) random variables, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, the sample mean. How large should we choose n to be such that $\mathbb{P}(\frac{\lambda}{2} \leq \overline{X}_n \leq \frac{3\lambda}{2}) \geq 0.99$? Use the CLT and give an answer involving $\Phi^{-1}(\cdot)$. Then evaluate it exactly when $\lambda = 1/10$ using the Φ table on the last page.

7. Bad Computer (More Faulty Machines)

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

(a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

(b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Use continuity correction.

Important: continuity correction says that if we are using the normal distribution to approximate

$$\mathbb{P}(a \leq \sum_{i=1}^{n} X_i \leq b)$$

where $a \leq b$ are integers and the X_i's are i.i.d. **discrete** random variables, then, as our approximation, we should use

$$\mathbb{P}(a - 0.5 \leq Y \leq b + 0.5)$$

where Y is the appropriate normal distribution that $\sum_{i=1}^{n} X_i$ converges to by the Central Limit Theorem.[1]

For more details see pages 209-210 in the book.

[1] The intuition here is that, to avoid a mismatch between discrete distributions (whose range is a set of integers) and continuous distributions, we get a better approximation by imagining that a discrete random variable, say W, is a continuous distribution with density function

$$f_W(x) := p_W(i) \quad \text{when } i - 0.5 \leq x < i + 0.5 \text{ and } i \text{ integer}$$

8. Waffles

A new diner specializing in waffles opens on our street. It will be open 24 hours a day, seven days a week. It is assumed that the inter-arrival times between customers will be i.i.d. Exponential random variables with mean 10 minutes. Approximate the probability that the 120th customer will arrive after the first 21 hours of operation.