STAT516: Stochastic Modeling of Scientific Data I	Fall 2021

Homework 1

Due **Due date** by 11:59pm

Instructor: Vincent Roulet *Grader: Alex Jiang*

Upload your answers to the following question on Canvas in a pdf. For the coding questions, you can use the programming language of your choice, e.g. Python or R.

First-Step Analysis

First-step analysis is a general strategy for solving many Markov chain problems by conditioning on the first step of the Markov chain. The reasoning applies more generally to discrete-time stochastic processes. Here are several exercises that illustrate the reasoning.

Though you can get the formula for the expectation of a geometric r.v. from the expression of the p.m.f., we consider here computing it by a first step analysis as a first illustration on the approach.

Exercise 1. Let $N \sim \mathrm{Geom}(p)$ be a geometric r.v. s.t. $N = k$ means that we needed to repeat k times an experiment with probability p of success before getting the first success.

Denote then $X_1, X_2, \dots$ the repeated trials of the experiment until you get the first success. To compute the expectation of N, we consider conditioning on the value of the first trial.

1. Given $X_1 = 1$, what is N?
2. If the first trial failed ($X_1 = 0$), intuitively, it's as if we were restarting the counting from X_2, i.e., we can expect $N \mid X_0 = 0 \sim N + 1$. Concretely show the following equality
$$\mathbb{P}(N = k \mid X_1 = 0) = \mathbb{P}(N + 1 = k)$$
3. Using the law of total expectation, i.e., that $\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N \mid X_1]]$, find the equation that $\mathbb{E}[N]$ must satisfy and conclude what is N.

Generally, the first-step analysis consists in conditioning the quantity of interest in a stochastic process (here it was the expectation of the first success) with respect to the values taken at the first step. Once the quantity has been expressed for any possible value of the first step, we can generally use the law of total expectation as above or the law of total probability to get an equation that the quantity of interest must respect. By solving this equation, we get the desired result.

The most classical example consists in analyzing the gambler's ruin problem presented below.

Exercise 2. Two players bet one dollar in each round. Player 1 wins with probability α and loses with probability $\beta = 1 - \alpha$. We assume that player 1 starts with a dollars and player 2 starts with b dollars. If one player does not have any dollar anymore the game naturally stops, i.e., the winner will keep the money forever. Let X_n be the fortune of player 1 after n rounds.

1. What values X_n can take? What is $p_{00} = \mathbb{P}(X_{n+1} = 0 \mid X_n = 0)$ and $p_{a+b,a+b} = \mathbb{P}(X_{n+1} = a + b \mid X_n = a + b)$?
2. More generally what are the transition probabilities p_{ij} for X_n?

Let T be the number of rounds when one of the players loses all his/her money. Because of the randomness of this model, T is also a random variable. We are interested in the probability that player 1 wins the game, which occurs when $X_T = a + b$. This probability clearly depends on the initial amount of money that player 1 has so we will denote it as

$$u(a) = P(X_T = a + b | X_0 = a).$$

We may view $u(a)$ as the probability that the chain is absorbed into the state $a + b$ when the chain starts at $X_0 = a$.

1. What is $u(0)$, $u(a+b)$?
2. Express $\mathbb{P}(X_T = a+b | X_1 = j)$ for any $j \in [0, a+b]$ in terms of u (when it is not 0)
3. Using the law of total probability, show that

$$u(a) = u(a+1)\alpha + u(a-1)\beta$$

4. For $\alpha = \beta = 1/2$, using the values $u(0), u(a+b)$, deduce $u(a)$ for any $a \in [0, a+b]$
5. (Optional) Determine $u(a)$ when $\alpha \neq \beta$. Hint: define $v(a) = u(a) - u(a-1)$, determine what recurrent equation $v(a)$ satisfies. To compute $v(1)$, express $u(a+b) - u(0)$ in terms of $v(j)$. Since you know $u(a+b)$ and $u(0)$ you'll be able to conclude for $v(a)$ and hence get $u(a)$

Here are other examples of how a problem can be solved by conditioning on the first step.

Exercise 3. A prisoner is trapped in a cell containing 3 doors. The first door leads to a tunnel that returns him to his cell after 2 days' travel. The second leads to a tunnel that returns him to his cell after 4 days' travel. The third door leads to freedom after 1 day of travel. If it is assumed that the prisoner will always select doors 1, 2, and 3 with respective probabilities 0.5, 0.3 and 0.2, what is the expected number of days until the prisoner reaches freedom?

Exercise 4. A coin that lands on heads with probability p is continually flipped. Compute the expected number of flips X that are made until a string of r heads in a row is obtained. (string of r heads means a successive sequence of r heads. For example, if $r = 2$ and you flipped the coins such that you obtained HTHTHHTTHHH, then X=6)
Hint: Condition on the time of the first occurrence of tails to obtain the equation

$$E[X] = (1-p)\sum_{i=1}^{r} p^{i-1}(i + E[X]) + (1-p)\sum_{i=r+1}^{\infty} p^{i-1} r$$

IID r.v. as Markov chains with equal rows

The following exercise is a simple generalization of what we've seen for the SIS model in the lectures.

Exercise 5. Let $(X_i)_{i\in\mathbb{N}}$ be a homogeneous Markov chain on a state space $\{1, \ldots k\}$. Assume that the t.p.m. of this Markov chain has rows all equal to $p = (p_1, \ldots, p_k)$. Assume that the initial distribution is equal to this row, i.e., $p_0 = p$. Show that the variables $X_1, \ldots, X_n$ are i.i.d.

Precipitations in Snoqualmie Falls

You'll find on Canvas a text file whose rows are the precipitation (in 1/100th of an inch, i.e. 0.254mm) each day of a given year. There are 36 rows for the years from 1948 to 1983. The data is drawn from the book Stochastic Modeling of Scientific Data by Guttorp, P. (1995). We will simply consider whether a day is dry (0 precipitation) or not.

Exercise 6. Pick a month (e.g. October) and extract the data (dry or wet) for all years for this month.

1. Plot a table whose columns are days of the month, rows are the years and each cell is either black if the day was wet or white if the day was wet.
2. Model the dry/wet days as i.i.d. Bernoulli r.v. with parameter p. Given then a 95% confidence interval for p
3. Simulate dry/wet days of this month with a Bernoulli r.v. with parameter p and plot it as in the first question.
4. Model the data as a one-step homogeneous Markov chain (one-step meaning that the distribution of a day only depends on the previous day, a two-step memory would mean that it would depend on the two previous days). Estimate the transition probabilities under this model.

5. Simulate the Markov chain by picking e.g. a uniform initial distribution and picking the next day according to the estimated transition probabilities. Plot the result as before.

6. Compute the n-step transition probability after $n = 2, 3, ..., 20$ days. After how many days would you conclude that your current information won't be informative anymore?