MVE137 Exam Solutions

Probability and Statistical Learning Using Python

3 January 2022

Part I

1. (a) Each letter is equally likely as they are uniformly distributed. Since there are $N = 26$ different possibilities the probability for each is

$$p = \frac{1}{N} = \frac{1}{26}$$

(b) Let M be the number of letters until first "p" appears. Let X_n be letter n. Then

$$\begin{aligned}
\mathbb{E}[M] =& \mathbb{E}[M|X_1 \neq \text{“p”}]\,(1-p) + \mathbb{E}[M|X_1 = \text{“p”}]\,p \\
\mathbb{E}[M] =& (1+\mathbb{E}[M])(1-p) + 1 \cdot p \\
\Rightarrow \mathbb{E}[M] =& \frac{1}{p} = N = 26
\end{aligned}$$

(c) Now, let M be the number of letters until first "pr" appears. Similarly to the previous exercise we get

$$\begin{aligned}
\mathbb{E}[M] =& \mathbb{E}[M|X_1 \neq \text{“p”}]\,(1-p) + \mathbb{E}[M|X_1 = \text{“p”}]\,p \\
=& (1+\mathbb{E}[M])(1-p) + \mathbb{E}[M|X_1 = \text{“p”}]\,p \\
=& (1-p) + \mathbb{E}[M]\,(1-p) + \mathbb{E}[M|X_1 = \text{“p”}]\,p \\
=& (1-p) + \mathbb{E}[M]\,(1-p) + \mathbb{E}[M|X_1 = \text{“p”}, X_2 \neq \text{“r”}]\,(1-p)p + \mathbb{E}[M|X_1X_2 = \text{“pr”}]\,p^2 \\
=& (1-p) + (1-p)2p + \mathbb{E}[M]\,(1-p) + \mathbb{E}[M]\,(1-p)p + 2p^2 = 1 - p + 2p - 2p^2 + 2p^2 + \mathbb{E}[M]\,(1-p+ \\
\Rightarrow p^2\,\mathbb{E}[M] =& 1 + p \quad \Rightarrow \mathbb{E}[M] = \frac{1}{p} + \frac{1}{p^2} = N + N^2 = 702
\end{aligned}$$

(d) Finally, let M be the number of letters until the first "proof" appears

$$\begin{aligned}
\mathbb{E}[M] =& \mathbb{E}[M|X_1 \neq \text{“p”}]\,(1-p) + \mathbb{E}[M|X_1 = \text{“p”}]\,p \\
=& \ldots \\
=& (1-p)\sum_{m=1}^{5} m p^{m-1} + \mathbb{E}[M]\,(1-p)\sum_{m=1}^{5} p^{m-1} + \mathbb{E}[M|X_1X_2X_3X_4X_5 = \text{“proof”}]\,p^5 \\
=& (1-p)\sum_{k=0}^{4}(k+1)p^k + \mathbb{E}[M]\,(1-p)\sum_{k=0}^{4} p^k + 5p^5 \\
=& (1-p)(5p^4 + 4p^3 + 3p^2 + 2p + 1) + \mathbb{E}[M]\,(1-p)\frac{1-p^5}{1-p} + 5p^5 \\
\mathbb{E}[M] =& p^4 + p^3 + p^2 + p + 1 - 5p^5 + 5p^5 + \mathbb{E}[M]\,(1-p^5) \\
\Rightarrow p^5\,\mathbb{E}[M] =& p^4 + p^3 + p^2 + p + 1 \\
\Rightarrow \mathbb{E}[M] =& \frac{p^4 + p^3 + p^2 + p + 1}{p^5} = N^5 + N^4 + N^3 + N^2 + N = 12356630
\end{aligned}$$

2. (a) Let X_n be PIN/last 4 digits of classmate n. Let x_0 be your last 4 digits = PIN that you try. X_n could be any integer between 0000 and 9999, i.e. 10^4 possibilities. Since uniform

$$\mathbb{P}[X_n = x_0] = \frac{1}{10^4}$$

or

$$\mathbb{P}[X_n \neq x_0] = \left(1 - \frac{1}{10^4}\right).$$

Thus for N classmates the probability of not getting any money is

$$\mathbb{P}[\text{No money}] = \left(1 - \frac{1}{10^4}\right)^N = \left(\frac{9999}{10^4}\right)^N$$

and the condition that we are interested in is

$$\mathbb{P}[\text{No money}] < 0.5.$$

Taking $\log_2$ of both sides (which is ok since $\log_2$ is always growing and both sides in inequality is > 0) and reordering the terms results in

$$N > \frac{1}{\log_2 \frac{10^4}{9999}} \approx 6931.13 \qquad \Rightarrow N > 6931.$$

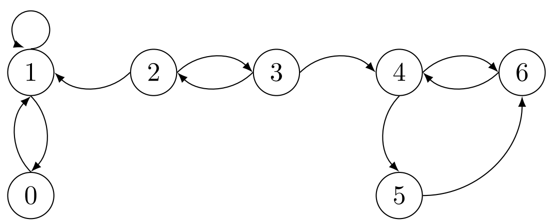
(b) Let Z = amount of money that you get. Z is almost a binomial RV, namely $Z = 200 \cdot Y$, where Y is binomial with parameters $N, p = \frac{1}{10^4}$.

$$\mathbb{E}[Z] = 200 \cdot \mathbb{E}[Y] = 200Np = N\frac{1}{50}$$

(c) Let D_n be birthday of classmate n. Then $\mathbb{P}[D_n = d] = \frac{1}{365} \forall d$. Let d be your birthday. We want the probability that two or more people have the same personal number as you. Let $y = z/200 > 0$ be the number of people with the same last 4 digits as you. Then

$$\begin{aligned}\mathbb{P}[\text{someone has same personal number as you}] &= 1 - \mathbb{P}[\text{no one has same personal number as you}] \\ &= 1 - \mathbb{P}[D_n \neq d \forall n \in \{1, ..., y\}] \\ &= 1 - \binom{y}{0}\left(1 - \frac{1}{365}\right)^y = 1 - \left(1 - \frac{1}{365}\right)^y\end{aligned}$$

3. The Markov chain is:



There are three communicating classes which have the following characteristics:

- 0,1: aperiodic, recurrent.
- 2,3: periodic, transient.
- 4,5,6: aperiodic, recurrent.

4. The CDF of exponential variable is

$$\begin{aligned} y &= 1 - e^{-x/\mu} \\ \Rightarrow x &= -\mu \ln(1 - y) \end{aligned}$$

We let Y be a uniform RV on (0,1). What is the CDF of the RV $X = -\mu \ln(1 - Y)$?

$$\begin{aligned} F_x(a) &= \mathbb{P}[X > a] \\ &= \mathbb{P}[-\mu \ln(1 - Y) > a] \\ &= ... = \mathbb{P}[Y > 1 - e^{-a/\mu}] \\ &= 1 - e^{-a/\mu} \end{aligned}$$

which we recognize as the CDF of an exponential function! Thus X is exponential with mean μ.

5.

6.

7.

Part II