DATA SCIENCE **Stochastic Methods**

February 20, 2024

Solutions

Problem 1. [12] Let $X_1, X_2, \dots, X_n$ be independent, absolutely continuous uniform $[0,4]$ random variables. Define $Y_k = |X_k - 2|$ for any $k = 1, \dots, n$.

(i) Prove that $\mathrm{P}[Y_1 \le 2] = 1$ and compute $\mathrm{P}[Y_1 \le y]$ for $y \in \mathbb{R}$;

(ii) Compute $\mathrm{E}[Y_1]$;

(iii) Compute $m(t) = \mathrm{E}[e^{tY_1}]$;

(iv) Use the Heoffding's inequality to prove a Chernoff's Bound Upper tail estimate for $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$.

(i) Since $0 < X_1 < 4 \Rightarrow Y_1 = |X_1 - 2|$ is s.t. $0 \le Y_1 < 2$

$$\boxed{\mathbb{P}[Y_1 \le y] = \begin{cases} 0 & y < 0 \\ \mathbb{P}[2-y < X_1 < 2+y] = y/2 & 0 \le y < 2 \\ 1 & y \ge 2 \end{cases}}$$

$\Rightarrow Y_1 \sim \mathcal{U}(0,2)$

(ii) $\boxed{\mathbb{E}[Y_1] = 1}$

(iii) $$m(t) = \mathbb{E}[e^{tY_1}] = \int_0^2 e^{ty} \tfrac{1}{2}\, dy = \boxed{\frac{e^{2t} - 1}{2t}}$$

(iv) The Höffding's inequality for iid bounded r.v.'s in $[a,b]$ with mean μ reads

$$\mathbb{P}[\overline{Y}_n - \mu \ge \varepsilon] \le e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

In the present case

$$\boxed{\mathbb{P}[\overline{Y}_n - 1 \ge \varepsilon] \le e^{-\frac{2n\varepsilon^2}{4}} = e^{-\frac{n\varepsilon^2}{2}}}$$

Problem 2. [10] Let X be a Binomial random variable of parameters $(2, 0, 5)$ and be Y be a Geometric random variable of parameter $(X+1)/3$, i.e. $Y|X = n \sim Geom((n+1)/3)$.

(i) Compute $\mathrm{P}[Y = k|X = n]$ for any $k \in \mathbb{N}, n = 0, 1, 2$;

(ii) Compute $h(n) = \mathrm{E}[Y|X = n]$ for any n;

(iii) Compute $\mathrm{E}[\mathrm{E}[Y|X]]$.

(i) $$\mathbb{P}[Y = k \mid X = n] = \boxed{\left(1 - \frac{n+1}{3}\right)^{k-1} \cdot \frac{n+1}{3}}$$

$k \in \mathbb{N} \setminus \{0\}$
$n = 0, 1, 2$

(ii) $$h(n) = \mathbb{E}[Y \mid X = n] = \frac{1}{\frac{n+1}{3}} = \boxed{\frac{3}{n+1}}$$

(iii) $$\mathbb{E}[\mathbb{E}[Y \mid X]] = \sum_{n=0}^{2} h(n) \cdot \mathbb{P}[X = n]$$

$$= 3 \cdot \mathbb{P}[X=0] + 3/2 \; \mathbb{P}[X=1] + 1 \; \mathbb{P}[X=2]$$

$$= 3 \cdot \left(1 - \frac{1}{2}\right)^2 + \frac{3}{2} \cdot 2 \cdot \left(1 - 1/2\right) \cdot 1/2 + 1 \cdot \left(1/2\right)^2$$

$$= 3/4 + 3/4 + 1/4 = \boxed{7/4}$$

Problem 3. [10] The pattern of sunny and rainy days is a homogeneous Markov chain with two states. Every sunny day is followed by another sunny day with probability 0.7. Every rainy day is followed by another rainy day with probability 0.8.

(i) Classify the states of this Markov Chain;

(ii) Today is sunny: what is the chance of rain the day after tomorrow?

(iii) Compute approximately the probability that November 1st next year is rainy.

(iv) If today is a rainy day, on average, how long will it take to have another rainy day?

$X \in \{s, r\}$ $\quad P_{s,s} = 0.7 \Rightarrow P_{s,r} = 0.3$

$P_{r,r} = 0.8 \Rightarrow P_{r,s} = 0.2$

(i)

$$P = \begin{array}{c} s \\ r \end{array} \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

$\boxed{\text{the MC is irreducible, aperiodic and all the states are recurrent}}$

(ii) $\boxed{P^{(2)}_{sr} = 0.45}$ $\qquad P^2 = \begin{bmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{bmatrix}$

(iii) Approximately this probability is equal to π_r, where $\pi = (\pi_s, \pi_r)$ represent the invariant distribution.

By the detailed balance equation $\pi_r \cdot 0.2 = \pi_s \cdot 0.3$

$\Rightarrow \pi_r + \frac{2}{3}\pi_r = 1 \Rightarrow \boxed{\pi_r = 3/5}, \ \pi_s = 2/5$

(iv) $\boxed{m_{rr} = \frac{1}{\pi_r} = 5/3}$