DATA SCIENCE **Stochastic Methods**

December 16, 2023

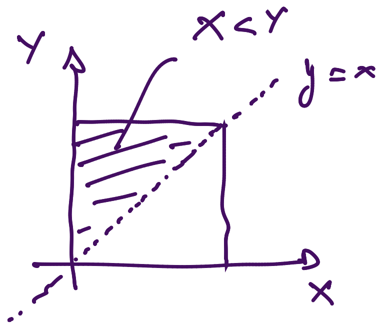
solutions

Problem 1. [9] Two friends arrange to have dinner together. Each will arrive independently at some point between 8 and 9 in the evening, wait for a maximum of 10 minutes (but not beyond 9), and if the other has not arrived within this time, they will leave. Describe the arrival times by two independent r.v.'s X and Y both $U(0,60)$, and therefore their joint density is constant on the square of vertex $(0,0)$, $(60,0)$, $(60,60)$ and $(0,60)$.

(i) Compute $\mathrm{P}[X<Y]$;

(ii) What is the probability that at least one of the two friends arrives after 8:30?

(iii) What is the probability that the two friends will have dinner together?

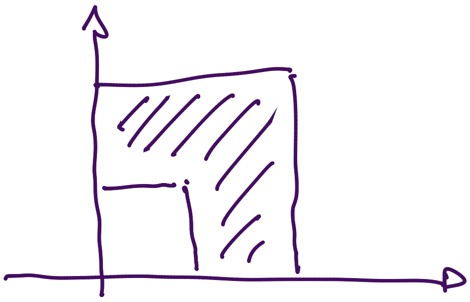


(i)

$$\mathbb{P}[X<Y] = \int_0^{60} dx \int_x^{60} \frac{1}{60^2}\, dy$$

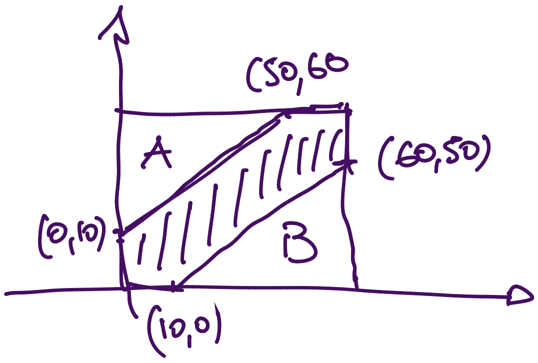
$$= 1/2$$

(ii)



$$\mathbb{P}[\max(X,Y) > 30] = 3/4$$

(iii)



$$\mathbb{P}[\,|X-Y|<10\,]$$

$$= 1 - \mathbb{P}[(X,Y)\in A] - \mathbb{P}[(X,Y)\in B]$$

$$= 1 - \frac{50^2}{60^2}\cdot\frac{1}{2} - \frac{50^2}{60^2}\cdot\frac{1}{2} =$$

$$= 1 - \frac{25}{36} = 11/36$$

Problem 2. [9] Let X be a Geometric random variable of parameter $1/2$ and Y be a Binomial random variable of parameters $(X, 1/2)$, i.e. $Y|X = n \sim Bin(n, 1/2)$.

(i) Compute $\mathrm{P}[Y = k|X = n]$ for any $k \leq n$;

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

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

(iv) Describe the support and the discrete density of the random variable $\mathrm{E}[Y|X]$.

(i) $\mathbb{P}[Y=k \mid X=n] = \binom{n}{k} \frac{1}{2}^{n}$

(ii) $h(n) = \mathbb{E}[Y \mid X=n] = n/2$ since $Y \mid X=n \sim Bin(n, 1/2)$

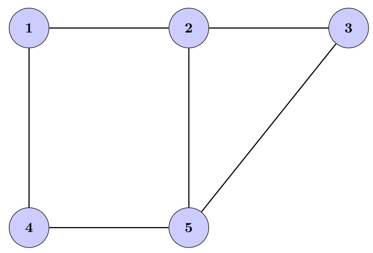
(iii) $\mathbb{E}[\,\mathbb{E}[Y \mid X]\,] = \sum_{n=1}^{\infty} h(n) \cdot \left(1-\frac{1}{2}\right)^{n-1} \cdot \frac{1}{2}$

$$= \sum_{n=1}^{\infty} \frac{n}{2} \left(\frac{1}{2}\right)^{n} = 2/2 = 1$$

(iv) $\mathbb{E}[Y \mid X] = h(X)$ is a discrete r.v. on $S = \{ n/2 : n \in \mathbb{N} \setminus \{0\} \}$ with discrete density

$$\mathbb{P}[\,h(X) = k\,] = \left(\frac{1}{2}\right)^{2k} \qquad \forall k \in S$$

Problem 3. [9] Define a simple Random Walk $\{X_n, n \geq 0\}$ on the undirected graph:



(i) Compute the probability to go from 3 to 4 in three steps.

(ii) Is the chain aperiodic?

(iii) Find the invariant distribution.

(iv) Starting from 3, what is the probability of visiting every state before visiting a state more than once?

(i) $\mathbb{P}[X_3 = 4 \mid X_0 = 3] = P_{3,2} \cdot P_{2,1} \cdot P_{1,4} + P_{3,2} \cdot P_{2,5} \cdot P_{5,1}$

$$= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{6}\left(\frac{1}{2} + \frac{1}{3}\right) = \frac{5}{36}$$

(ii) YES: $P_{33}^{(2)} > 0$ and $P_{33}^{(3)} > 0$

(iii) $\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) = \left(\frac{2}{12}, \frac{3}{12}, \frac{2}{12}, \frac{2}{12}, \frac{3}{12}\right)$

(iv) $\mathbb{P}[\text{visit every state once} \mid X_0 = 3] =$

$$= P_{32} \cdot P_{21} \cdot P_{14} \cdot P_{45} + P_{32} P_{25} \cdot P_{54} \cdot P_{41} + P_{35} \cdot P_{52} \cdot P_{21} \cdot P_{14} + P_{35} P_{54} P_{41} P_{12}$$

$$= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \frac{1}{2}$$

$$= \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{6}\left(\frac{1}{2} + \frac{1}{3}\right) = \frac{5}{36}$$

Problem 4. [9] Let $(X_i)_{1\le i\le 2n}$ be a family of i.i.d. $Exp(\lambda)$ r.v.'s and let $Y_k = X_{2k-1}+X_{2k}$, for any $k\le n$.

(i) Compute the moment generating function of Y_1;

(ii) Defined $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$, determine an exponential decay for the "upper tail" of $\overline{Y}_n - \mathrm{E}[\overline{Y}_n]$. (Hint: use the Chernoff bound proved for exponential random variables.)

(i) $$m_{X_1}(t) = \begin{cases} \dfrac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & t \geq \lambda \end{cases}$$

$$m_{Y_1}(t) = \begin{cases} \dfrac{\lambda}{\lambda - t}\cdot\dfrac{\lambda}{\lambda - t} = \dfrac{\lambda^2}{(\lambda-t)^2} & t<\lambda, \\ +\infty & t \geq \lambda \end{cases}$$

(ii) $$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{k=1}^{2n} X_k = 2\,\frac{1}{2n}\sum_{k=1}^{2n} X_k$$

$$= 2\,\overline{X}_{2n} \qquad \text{and } \mathbb{E}[Y_1] = 2/\lambda$$

We have proved the Chernoff's bound for Exp(λ) r.v.'s

$$\mathbb{P}\left[\overline{X}_n > \frac{(1+\delta)}{\lambda}\right] \leq e^{-n\,\delta^2/2(1+\delta)} \qquad \forall \delta>0$$

so

$$\mathbb{P}\left[\overline{Y}_n - \mathbb{E}[\overline{Y}_n] > \varepsilon\right] = \mathbb{P}\left[2\,\overline{X}_{2n} - 2\,\mathbb{E}[\overline{X}_{2n}] > \varepsilon\right]$$

$$= \mathbb{P}\left[\overline{X}_{2n} - 1/\lambda > \varepsilon/2\right] \qquad \text{take } \varepsilon = \frac{2\delta}{\lambda}$$

$$= \mathbb{P}\left[\overline{X}_{2n} > \delta/\lambda + 1/\lambda\right] = \mathbb{P}\left[\overline{X}_{2n} > \frac{1+\delta}{\lambda}\right]$$

$$\leq e^{-2n\cdot\delta^2/2(1+\delta)} = e^{-n\,\delta^2/(1+\delta)}$$