DATA SCIENCE **Stochastic Methods**

December 19, 2024

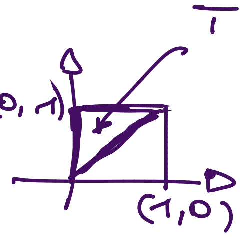
Problem 1. [9] We choose two points, X and Y, randomly, uniformly and independently on the segment [0,1].

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

[2] (ii) Compute $\mathrm{P}[\max\{X,Y\}<1/2]$;

[5] (iii) Compute the expected length of the segment with endpoints X and Y.

(i) $\mathbb{P}[X<Y] = \mathbb{P}[(X,Y)\in T]$ where



$$T = \{(x,y)\in[0,1]^2 : x<y\}.$$

$$\mathbb{P}[(X,Y)\in T] = \iint_T dx\,dy = \boxed{1/2}$$

(or $\mathbb{P}[X<Y] + \mathbb{P}[Y<X] = 1$ and $\mathbb{P}[X<Y] = \mathbb{P}[Y<X]$)

(ii) $$\mathbb{P}[\max\{X,Y\}<1/2] = \mathbb{P}[X<1/2,\ Y<1/2] =$$
$$= \mathbb{P}[X<1/2]\cdot\mathbb{P}[Y<1/2] = 1/2\cdot 1/2 = \boxed{1/4}$$

(iii) $$\mathbb{E}[|X-Y|] = \mathbb{E}[g(X,Y)] = \int_0^1\int_0^1 |x-y|\,dx\,dy$$

$$= \int_0^1 dx\left(\int_0^x (x-y)\,dy\right) + \int_0^1 dx\left(\int_x^1 (y-x)\,dy\right)$$

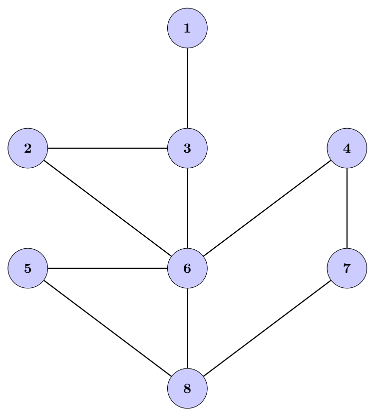
$$= \int_0^1 \left[xy - y^2/2\right]_0^x dx + \int_0^1 \left[y^2/2 - xy\right]_x^1 dx$$

$$= \int_0^1 \left(x^2 - x^2/2\right) dx + \int_0^1 \left(1/2 - x - \frac{x^2}{2} + x^2\right) dx$$

$$= \left[\frac{x^3}{6}\right]_0^1 + \left[\frac{1}{2}x - \frac{x^2}{2} + \frac{x^3}{6}\right]_0^1 = 1/6 + 1/2 - \frac{1}{2} + 1/6$$

$$= \boxed{1/3}$$

Problem 2. [9] Define a simple Random Walk $\{X_n, n \geq 0\}$ on the undirected graph:



[2] (i) Compute the probability of going from state 2 to state 8 in three steps.

[2] (ii) Is the chain irreducible? aperiodic?

[2] (iii) Find the invariant distribution.

[3] (iv) Starting from state 1, what is the probability of visiting every state before visiting any state more than once?

(i) $2 \xrightarrow{1/2} 3 \xrightarrow{1/3} 6 \xrightarrow{1/5} 8$ or $2 \xrightarrow{1/2} 6 \xrightarrow{1/5} 5 \xrightarrow{1/2} 8$

$$\frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{5} + \frac{1}{2}\cdot\frac{1}{5}\cdot\frac{1}{2} = \frac{1}{30} + \frac{1}{20} = \frac{5}{60} = \frac{1}{12}$$

(ii) YES (since the graph is connected)

YES ($P^2_{33} > 0$ and $P^3_{33} > 0 \Rightarrow$ 3 is aperiodic)

(iii) $P_1 = 1,\ P_2 = 2,\ P_3 = 3,\ P_4 = 2,\ P_5 = 2,\ P_6 = 5,\ P_7 = 2,\ P_8 = 3$

$P = \sum_{i=1}^{8} P_i = 20 \qquad \pi_i = P_i/P \qquad i = 1, \dots, 8$

(iv) The only two possible paths are:

1 — 3 — 2 — 6 — 5 — 8 — 7 — 4

1 — 3 — 2 — 6 — 4 — 7 — 8 — 5

and the probability is $1/180$

Problem 3. [9] Let X be a Binomial random variable with parameters $(2,p)$, where $0<p<1$ and define $Y=(X+1)/3$. Assume that Z is a Geometric random variable with parameter Y, i.e. $Z|Y=k\sim Geo(k)$.

[3] (i) Compute the support and the discrete density of Y;

[2] (ii) Compute $h(k)=\mathrm{E}[Z|Y=k]$ for any k in the support of Y;

[4] (iii) Compute $\mathrm{E}[Z]$.

(i) $Y \in \{1/3, 2/3, 1\}$ $\quad \mathbb{P}[Y=1/3]=(1-p)^2, \quad \mathbb{P}[Y=2/3]=2p(1-p)$

$\mathbb{P}[Y=1]=p^2$

(ii) $\boxed{h(k)=1/k \quad k\in\{1/3, 2/3, 1\},}$

(iii) $\boxed{\mathbb{E}[Z]}=\mathbb{E}[\mathbb{E}[Z|Y]]=\sum_h h(k)\cdot \mathbb{P}[Y=k]$

$= 3(1-p)^2+\frac{3}{2}\cdot 2p(1-p)+p^2$

$\boxed{=p^2-3p+3}$

Problem 4. [9] Let $(X_i)_{1 \le i \le n}$ be a family of i.i.d. $N(\mu, \sigma^2)$ r.v.'s.

[3] (i) Compute the moment generating function of X_1;

[6] (ii) Defined $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, determine an exponential decay for the "lower tail" of $\overline{X}_n - \mu$

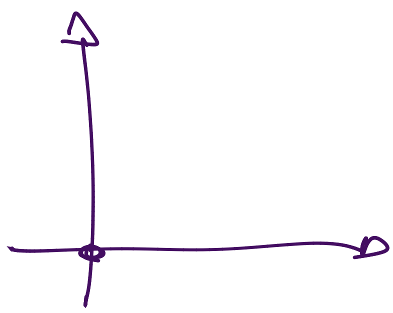
(i) $$m_X(t) = \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \boxed{e^{t\mu + \frac{\sigma^2 t^2}{2}}}$$

(ii) $$\boxed{\mathbb{P}[\overline{X}_n - \mu < -\varepsilon]} = \mathbb{P}[\overline{X}_n < \mu - \varepsilon]$$

$$= \mathbb{P}[X_1 + \dots + X_n < n(\mu - \varepsilon)]$$

$$= \mathbb{P}[e^{-t(X_1 + \dots + X_n)} > e^{-tn(\mu-\varepsilon)}] \qquad \forall t > 0$$

$$\le \frac{\mathbb{E}[e^{-t(X_1 + \dots + X_n)}]}{e^{-tn(\mu-\varepsilon)}} =$$



$$= e^{tn(\mu-\varepsilon)} (m_X(-t))^n$$

$$= e^{-n[-t(\mu-\varepsilon) + t\mu - t^2\sigma^2/2]} = e^{-n h(t)} \le \bullet$$

$$h(t) = t\varepsilon - t^2\sigma^2/2 \qquad h'(t) = \varepsilon - t\sigma^2 = 0 \qquad t^* = \varepsilon/\sigma^2$$

$$\bullet \le e^{-n\left(\varepsilon^2/\sigma^2 - \frac{\varepsilon^2\sigma^2}{2\sigma^4}\right)} = \boxed{e^{-\frac{n\varepsilon^2}{\sigma^2 2}}}$$